ALIBABA CLOUD

Alibaba Cloud

Elastic Compute Service

Instance

Document Version: 20201014

Alibaba Cloud

Legal disclaimer

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

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

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

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

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

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

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

Document conventions

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

Table of Contents

1.Overview

An ECS instance is a virtual machine that contains basic computing components such as the vCPU, memory, operating system, network, and disk. You can fully customize and modify all configurations of an ECS instance. After you log on to the Alibaba Cloud Management console, you can manage resources and configure the environment of your ECS instances.

Alibaba Cloud ECS server elastic computing

Instance types

ECS instances are categorized into different instance families based on business scenarios. An instance family contains different instance types based on their vCPU and memory specifications. Instance types can have different vCPU and memory specifications, such as the CPU model and clock speed. Some software or applications require instance types that use specific vCPU and memory specifications. For example, if you want to use a Windows Server image, you must purchase an instance type that is equipped with a memory of at least 512 MiB. A vCPU stands for a thread. A CPU can run multiple threads.

For more information about instance types, see Instance families and Phased-out instance types. If the instance specifications cannot meet your application requirements, you can change the instance type. For more information, see Overview of instance upgrade and downgrade.

System storage

An ECS instance must contain a system disk to store the operating system and core configurations. An image is used to initialize a system disk and determines the operating system and initial software configurations of an ECS instance. Typically, the capacity of system disks is small. If you need more storage types or space, you can add data disks. For more information, see Block Storage overview.

Network type

ECS instances support the classic network and Virtual Private Cloud (VPC). We recommend that you use VPC. ECS instances in the same network, such as in the same VPC, can communicate over the internal network through private IP addresses. If you want your instances to communicate over the Internet, you can obtain the public IP addresses to communicate with devices on the Internet. For more information, see Network types.

Billing details

For more information about the billing methods and prices of ECS instances, see Billing overview.

Usage

- For more information about the limits and quotas of ECS instances, see Limits.
- For more information about how to create ECS instances, see Creation method overview.
- For more information about the lifecycle management of ECS instances, see ECS instance lifecycle.

2.ECS instance lifecycle

The lifecycle of an ECS instance begins when the instance is created and ends when the instance is released. This topic describes the states that an ECS instance may go through during its lifecycle.

Instance states

The following table describes the states that an ECS instances may go through during its lifecycle.

State	State in an API response	State attribute	Description	Visible in the ECS console
Preparing	Pending	Transitory	After an instance is created, it is in this state before it enters the **Running** state. If the instance remains in this state for an extended period of time, an exception occurs.	No
Starting	Starting	Transitory	After you start or restart an instance by using the ECS console or by calling an API operation, the instance enters this state before it enters the **Running** state. If the instance remains in this state for an extended period of time, an exception occurs.	Yes
Running	Running	Stable	If an instance runs properly, it is in this state. **Note** An ECS instance can be externally accessed only when it is in the Running state.	Yes
Expiring	Running	Stable	A subscription instance remains in the **Expiring** state for 15 days before it expires. If your instance enters the Expiring state, we recommend that you renew the instance in a timely manner. For more information, see Overview.	Yes
Stopping	Stopping	Transitory	When you stop an instance by using the ECS console or by calling an API operation, the instance enters this state before it enters the **Stopped** state. If the instance remains in this state for an extended period of time, an exception occurs.	Yes

State	State in an API response	State attribute	Description	Visible in the ECS console
Stopped	Stopped	Stable	After an instance is stopped or after an instance is created but has not started, it is in the **Stopped** state. **Note** An ECS instance can be externally accessed only when it is in the Running state.	Yes
Expired	Stopped	Stable	When a subscription instance expires or when a pay-as-you-go instance is stopped due to overdue payments, the instance enters the **Expired** state. For more information about the changes of resource status, see Subscription and Pay-as-you-go. **Note** An ECS instance can be externally accessed only when it is in the Running state.	Yes
Locked	Stopped	Stable	If you have an overdue payment in your account or if your account is insecure, your instance enters the Locked state. You can submit a ticket to unlock the instance.	Yes
To Be Released	Stopped	Stable	If you apply for a refund for a subscription instance before the instance expires, the instance enters the To Be Released state.	Yes

Instance status in API responses

You can call the DescribeInstanceStatus or DescribeInstances operation to query the instance status. The following figure shows the transition between instance statuses in API responses.

3.Instance families

An ECS instance is the smallest unit that can provide compute capabilities and services for your business. Compute capabilities vary based on instance types. This topic describes available ECS instance families and the features, specifications, and application scenarios of each instance family.

ECS instances are categorized into different instance families based on their usage scenarios. Each instance family is divided into different instance types based on their CPU and memory specifications. *ECS instance type* defines the basic properties of an ECS instance, such as CPU, clock speed, and memory. In addition to the instance type, you must also configure the Elastic Block Storage (EBS) devices, image, and network type when you create an ECS instance.

> **Note** The available instance families and types vary based on regions. You can go to the ECS Instance Types Available for Each Region page to view the available instance types in each region.

Enterprise scenarios have high requirements for business stability. Alibaba Cloud ECS instance families are divided into enterprise-level and shared instance families based on whether the instance families are suitable for enterprise scenarios. Enterprise-level instance families offer consistent performance and dedicated resources. In enterprise-level instance families, each vCPU corresponds to a hyperthread of the Intel® Xeon® core. For more information about the differences between enterprise-level and shared instance families, see ECS instance FAQ.

You can upgrade or downgrade instance types within an instance family or across instance families. For more information, see Change instance types of instances.

For information about how to choose instance families based on scenarios, see Best practices for instance type selection.

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

Enterprise-level computing instance families based on the x86 architecture:	
Recommended intance families	Other available instance families

Enterprise-level computing instance families based on the x86 architecture:

- g6, general purpose instance family
- g6a, general purpose instance family
- g6t, security-enhanced general purpose instance family
- g6e, general purpose instance family with enhanced performance
- g5, general purpose instance family
- g5se, storage enhanced instance family
- g5ne, network enhanced instance family
- ic5, compute intensive instance family
- c6, compute optimized instance family
- c6a, compute optimized instance family
- c6t, security-enhanced compute optimized instance family
- c6e, compute optimized instance family with enhanced performance
- c5, compute optimized instance family
- r6, memory optimized instance family
- r6a, memory optimized instance family
- r6e, memory optimized instance family with enhanced performance
- re6, high memory instance family
- r5, memory optimized instance family
- d2c, compute intensive big data instance family
- d2s, storage intensive big data instance family
- d1ne, big data instance family with enhanced network performance
- i2, instance family with local SSDs
- i2g, instance family with local SSDs
- i2ne, instance family with local SSDs
- i2gne, instance family with local SSDs
- hfc7, compute optimized instance family with high clock speed
- hfc6, compute optimized instance family with high clock speed
- hfg7, general purpose instance family with high clock speed
- hfg6, general purpose instance family with high clock speed
- hfr7, memory optimized instance family with high clock speed
- hfr6, memory optimized instance family with high clock speed
- sn2ne, general purpose instance family with enhanced network performance
- sn1ne, compute optimized instance family with enhanced network performance
- re4, high memory instance family
- re4e, high memory instance family
- se1ne, memory optimized instance family with enhanced network performance
- se1, memory optimized instance family
- d1, big data instance family
- i1, instance family with local SSDs
- hfc5, compute optimized instance family with high clock speed
- hfg5, general purpose instance family with high clock speed

Enterprise-level computing instance families based on the x86 architecture: Enterprise-level heterogeneous computing instance families:	
Recommended intance families	Other available instance families
• vgn6i, lightweight GPU-accelerated compute optimized instance family • gn6i, GPU-accelerated compute optimized instance family • gn6e, GPU-accelerated compute optimized instance family • gn6v, GPU-accelerated compute optimized instance family • f3, FPGA-accelerated compute optimized instance family	• vgn5i, lightweight GPU-accelerated compute optimized instance family • gn5, GPU-accelerated compute optimized instance family • gn5i, GPU-accelerated compute optimized instance family • gn4, GPU-accelerated compute optimized instance family • f1, FPGA-accelerated compute optimized instance family

ECS Bare Metal Instance families and Super Computing Cluster (SCC) instance families:	
Recommended intance families	Other available instance families

ECS Bare Metal Instance families and Super Computing Cluster (SCC) instance families:

- ebmgn6e, GPU-accelerated compute optimized ECS Bare Metal Instance family
- ebmgn6v, GPU-accelerated compute optimized ECS Bare Metal Instance family
- ebmgn6i, GPU-accelerated compute optimized ECS Bare Metal Instance family
- ebmc6a, compute optimized ECS Bare Metal Instance family
- ebmc6e, compute optimized ECS Bare Metal Instance family with enhanced performance
- ebmc6, compute optimized ECS Bare Metal Instance family
- ebmg6a, general purpose ECS Bare Metal Instance family
- ebmg6e, general purpose ECS Bare Metal Instance family with enhanced performance
- ebmg6, general purpose ECS Bare Metal Instance family
- ebmr6a, memory optimized ECS Bare Metal Instance family
- ebmr6e, memory optimized ECS Bare Metal Instance family with enhanced performance
- ebmr6, memory optimized ECS Bare Metal Instance family
- ebmhfc6, compute optimized ECS Bare Metal Instance family with high clock speed
- ebmhfg6, general purpose ECS Bare Metal Instance family with high clock speed
- ebmhfr6, memory optimized ECS Bare Metal Instance family with high clock speed
- ebmre6p, persistent memory optimized ECS Bare Metal Instance family with enhanced performance
- ebmre6-6t, memory optimized ECS Bare Metal Instance family with enhanced performance
- scchfc6, compute optimized SCC instance family with high clock speed
- scchfg6, general purpose SCC instance family with high clock speed
- scchfr6, memory optimized SCC instance family with high clock speed
- scch5, SCC instance family with high clock speed
- sccg5, general purpose SCC instance family
- sccgn6, GPU-accelerated compute optimized SCC instance family

- ebmc5s, compute optimized ECS Bare Metal Instance family with enhanced network performance
- ebmg5s, general purpose ECS Bare Metal Instance family with enhanced network performance
- ebmr5s, memory optimized ECS Bare Metal Instance family with enhanced network performance
- ebmg5, general purpose ECS Bare Metal Instance family
- ebmhfg5, ECS Bare Metal Instance family with high clock speed
- ebmc4, compute optimized ECS Bare Metal Instance family

Shared computing instance families based on the x86 architecture:	
Recommended intance families	Other available instance families
• t6, burstable instance family	• t5, burstable instance family • v5, CPU overprovisioned instance family • xn4, n4, mn4, and e4, previous-generation shared instance families

For information about retired instance families, see Phased-out instance types.

g6, general purpose instance family

Features

- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the X-Dragon architecture.
- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.

  Note The maximum performance of disks varies based on instance families. A single g6 instance can deliver up to 200,000 IOPS. For higher storage I/O performance, we recommend that you use g5se. For more information, see g5se, memory optimized instance family with enhanced performance.

- Provides high storage I/O performance based on large computing capacity.

  Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Offers a CPU-to-memory ratio of 1:4.
- Allows you to enable or disable Hyper-Threading.

  Note By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Provides an ultra-high packet forwarding rate.

  Note The maximum network performance varies based on instance families. For higher concurrent connection capabilities, we recommend that you use g5ne. For more information, see g5ne, network enhanced instance family.

- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
- Provides high network performance based on large computing capacity.
- Supports instance type changes to c6 or r6.
- Suitable for the following scenarios:

- Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
- Enterprise-level applications of various types and sizes
- Websites and application servers
- Game servers
- Small and medium-sized database systems, caches, and search clusters
- Data analysis and computing
- Compute clusters and memory intensive data processing

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g6.large	2	8.0	None	1.0	3.0	300	Yes	Up to 250	2	2	6	10.0	1
ecs.g6.xlarge	4	16.0	None	1.5	5.0	500	Yes	Up to 250	4	3	10	20.0	1.5
ecs.g6.2xlarge	8	32.0	None	2.5	8.0	800	Yes	Up to 250	8	4	10	25.0	2
ecs.g6.3xlarge	12	48.0	None	4.0	10.0	900	Yes	Up to 250	8	6	10	30.0	2.5
ecs.g6.4xlarge	16	64.0	None	5.0	10.0	1,000	Yes	300	8	8	20	40.0	3

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g6.6xlarge	24	96.0	None	7.5	10.0	1,500	Yes	450	12	8	20	50.0	4
ecs.g6.8xlarge	32	128.0	None	10.0	None	2,000	Yes	600	16	8	20	60.0	5
ecs.g6.13xlarge	52	192.0	None	12.5	None	3,000	Yes	900	32	7	20	100.0	8
ecs.g6.26xlarge	104	384.0	None	25.0	None	6,000	Yes	1,800	32	15	20	200.0	16

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

g6a, general purpose instance family

g6a is in invitational preview. To use g6a, submit a ticket.

Features

- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the X-Dragon architecture.
- Compute:

- Equipped with 2.6 GHz AMD EPYC ™ ROME processors with a maximum turbo frequency of 3.3 GHz for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:4.
- Allows you to enable or disable Hyper-Threading.

> Note By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Storage:
  - I/O optimized.
  - Supports enhanced SSDs (ESSDs), standard SSDs, and ultra disks.
  - Provides high storage I/O performance based on large computing capacity.

> Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Network:
  - Provides an ultra-high packet forwarding rate.
  - Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Video encoding and decoding
  - Scenarios where large volumes of packets are received and transmitted
  - Websites and application servers
  - Small and medium-sized database systems, caches, and search clusters
  - Game servers
  - Test and development, such as DevOps
  - Other general purpose enterprise-level applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	ENIs (including one primary ENI)	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g6a.large	2	8.0	None	1.0	10.0	900	Yes	2	12.5	1.0
ecs.g6a.xlarge	4	16.0	None	1.5	10.0	1,000	Yes	3	20.0	1.5

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	ENIs (including one primary ENI)	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g6a.2xlarge	8	32.0	None	2.5	10.0	1,600	Yes	4	30.0	2.0
ecs.g6a.4xlarge	16	64.0	None	5.0	10.0	2,000	Yes	8	60.0	3.0
ecs.g6a.8xlarge	32	128.0	None	8.0	10.0	3,000	Yes	7	75.0	4,0
ecs.g6a.16xlarge	64	256.0	None	16.0	None	6,000	Yes	8	150.0	8.0
ecs.g6a.32xlarge	128	512.0	None	32.0	None	12,000	Yes	15	300.0	16.0
ecs.g6a.64xlarge	256	1,024.0	None	64.0	None	24,000	Yes	15	600.0	32.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

g6t, security-enhanced general purpose instance family

g6t is in invitational preview.

Features

- Implements trusted boot based on the Trusted Platform Module (TPM) chip. During a trusted boot, each module in the boot chain from the underlying hardware to the guest OS is measured and verified.
- Supports comprehensive monitoring on the IaaS layer and provides trusted capabilities of the whole IaaS layer.

- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the third-generation X-Dragon architecture. g6t improves storage performance, network performance, and computing stability by an order of magnitude through fast path acceleration of X-Dragon chips.
- Compute:
  - Equipped with 2.5 GHz Intel® Xeon® Platinum 8269 (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
  - Offers a CPU-to-memory ratio of 1:4.
  - Allows you to enable or disable Hyper-Threading.

    Note By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Storage:
  - I/O optimized.
  - Supports ESSDs only.
  - Provides high storage I/O performance based on large computing capacity.

    Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Network:
  - Provides an ultra-high packet forwarding rate.
  - Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios that require high security and enhanced trust, such as financial services, government affairs, and enterprise services
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Enterprise-level applications of various types and sizes
  - Websites and application servers
  - Game servers
  - Small and medium-sized database systems, caches, and search clusters
  - Data analysis and computing
  - Compute clusters and memory intensive data processing

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	TPM support	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g6t.large	2	8.0	None	A burstable bandwidth of up to 10.0	900	Yes	Yes	Up to 250	2	3	6	20.0	1.0
ecs.g6t.xlarge	4	16.0	None	A burstable bandwidth of up to 10.0	1,000	Yes	Yes	Up to 250	4	4	15	40.0	1.5
ecs.g6t.2xlarge	8	32.0	None	A burstable bandwidth of up to 10.0	1,600	Yes	Yes	Up to 250	8	4	15	50.0	2.0

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	TPM support	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g6t.4xlarge	16	64.0	None	A burstable bandwidth of up to 10.0	3,000	Yes	Yes	300	8	8	30	80.0	3.0
ecs.g6t.8xlarge	32	128.0	None	10.0	6,000	Yes	Yes	600	16	8	30	150.0	5.0
ecs.g6t.13xlarge	52	192.0	None	16.0	9,000	Yes	Yes	900	32	7	30	240.0	8.0
ecs.g6t.26xlarge	104	384.0	None	32.0	24,000	Yes	Yes	1,800	32	15	30	480.0	16.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
- The results for network capabilities are the maximum values obtained from single item tests. For example, when network bandwidth is tested, no stress tests are performed on the packet forwarding rate or other metrics of the network.

g6e, general purpose instance family with enhanced performance

Features

- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the third-generation X-Dragon architecture. g6e improves storage performance, network performance, and computing stability by an order of magnitude through fast path acceleration of X-Dragon chips.
- I/O optimized.
- Supports ESSDs only.
- Provides high network and storage I/O performance based on large computing capacity.

  > **Note** For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Provides an ultra-high packet forwarding rate.

  > **Note** The maximum network performance varies based on instance families. For higher concurrent connection capabilities, we recommend that you use g5ne. For more information, see g5ne, network enhanced instance family.

- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269 (Cascade) processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
- Allows you to enable or disable Hyper-Threading.

  > **Note** By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Enterprise-level applications of various types and sizes
  - Websites and application servers
  - Game servers
  - Small and medium-sized database systems, caches, and search clusters
  - Data analysis and computing
  - Compute clusters and memory intensive data processing

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g6e.large	2	8.0	None	A burstable bandwidth of up to 10.0	900	Yes	Up to 250	2	3	6	20.0	1.0
ecs.g6e.xlarge	4	16.0	None	A burstable bandwidth of up to 10.0	1,000	Yes	Up to 250	4	4	15	40.0	1.5
ecs.g6e.2xlarge	8	32.0	None	A burstable bandwidth of up to 10.0	1,600	Yes	Up to 250	8	4	15	50.0	2.0

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g6e.4xlarge	16	64.0	None	A burstable bandwidth of up to 10.0	3,000	Yes	300	8	8	30	80.0	3.0
ecs.g6e.8xlarge	32	128.0	None	10.0	6,000	Yes	600	16	8	30	150.0	5,0
ecs.g6e.13xlarge	52	192.0	None	16.0	9,000	Yes	900	32	7	30	240.0	8.0
ecs.g6e.26xlarge	104	384.0	None	32.0	24,000	Yes	1,800	32	15	30	480.0	16.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
- The results for network capabilities are the maximum values obtained from single item tests. For example, when network bandwidth is tested, no stress tests are performed on the packet forwarding rate or other metrics of the network.

g5, general purpose instance family

Features

- I/O optimized.

- Supports ESSDs, standard SSDs, and ultra disks.

  > **Note** The maximum performance of disks varies based on instance families. A single g5 instance can deliver up to 200,000 IOPS. For higher storage I/O performance, we recommend that you use g5se. For more information, see g5se, memory optimized instance family with enhanced performance.

- Offers a CPU-to-memory ratio of 1:4.
- Provides an ultra-high packet forwarding rate.

  > **Note** The maximum network performance varies based on instance families. For higher concurrent connection capabilities, we recommend that you use g5ne. For more information, see g5ne, network enhanced instance family.

- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) or 8269CY (Cascade Lake) processors for consistent computing performance.
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Enterprise-level applications of various types and sizes
  - Small and medium-sized database systems, caches, and search clusters
  - Data analysis and computing
  - Compute clusters and memory intensive data processing

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.g5.large	2	8.0	None	1.0	300	Yes	2	2	6
ecs.g5.xlarge	4	16.0	None	1.5	500	Yes	2	3	10
ecs.g5.2xlarge	8	32.0	None	2.5	800	Yes	2	4	10
ecs.g5.3xlarge	12	48.0	None	4.0	900	Yes	4	6	10

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.g5.4xlarge	16	64.0	None	5.0	1,000	Yes	4	8	20
ecs.g5.6xlarge	24	96.0	None	7.5	1,500	Yes	6	8	20
ecs.g5.8xlarge	32	128.0	None	10.0	2,000	Yes	8	8	20
ecs.g5.16xlarge	64	256.0	None	20.0	4,000	Yes	16	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

g5se, memory optimized instance family with enhanced performance

g5se is currently under invitational preview. To use g5se, you must submit a ticket.

Features:

- A single g5se instance attached with enhanced SSDs (ESSDs) can deliver a random read/write IOPS of up to 1,000,000 and a throughput of up to 32 Gbit/s.
- You can create g5se instances only on dedicated hosts.

  Note For information about other types of instances that can be created on dedicated hosts, see Dedicated host types.

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Provides strong storage I/O performance with a large compute capacity.

  Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Offers a CPU-to-memory ratio of 1:4.

- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Suitable for the following scenarios:
  - I/O-intensive scenarios, such as medium and large OLTP core databases
  - Medium and large NoSQL databases
  - Search and real-time log analytics
  - Traditional large enterprise-level commercial software, such as SAP

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g5se.large	2	8.0	None	1.0	300	2	2	6	30	1
ecs.g5se.xlarge	4	16.0	None	1.5	500	2	3	6	60	2
ecs.g5se.2xlarge	8	32.0	None	2.5	800	2	3	8	120	4
ecs.g5se.4xlarge	16	64.0	None	5.0	1,000	4	8	10	230	8
ecs.g5se.6xlarge	24	96.0	None	7.5	1,500	6	8	10	340	12
ecs.g5se.8xlarge	32	128.0	None	10.0	2,000	8	8	10	450	15
ecs.g5se.16xlarge	64	256.0	None	14.0	3,000	16	8	10	900	30

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g5se.18xlarge	70	336.0	None	16.0	4,000	16	8	10	1,000	32

Note For more information about these specifications, see Description of instance specifications.

g5ne, network enhanced instance family

Features

- Instances of the g5ne instance family significantly improve network throughput and packet forwarding rate. A single g5ne instance can deliver up to 10,000 Kpps.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) or 8269CY (Cascade Lake) processors for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:4.
- I/O optimized.
- Supports standard SSDs and ultra disks.
- Provides high network performance based on large computing capacity.

  Note We recommend that you select instance types of the g5ne instance family to deploy Data Plane Development Kit (DPDK) applications.

- Suitable for the following scenarios:
  - DPDK applications
  - Network intensive scenarios such as NFV or SD-WAN, mobile Internet, on-screen video comments, and telecom data forwarding
  - Small and medium-sized database systems, caches, and search clusters
  - Enterprise-level applications of various types and sizes
  - Big data analysis and machine learning

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g5ne.large	2	8.0	None	1.0	400	Yes	450	2	3	10	10.0	1
ecs.g5ne.xlarge	4	16.0	None	2.0	750	Yes	900	4	4	15	15.0	1
ecs.g5ne.2xlarge	8	32.0	None	3.5	1,500	Yes	1,750	8	6	15	30.0	1
ecs.g5ne.4xlarge	16	64.0	None	7.0	3,000	Yes	3,500	16	8	30	60.0	2
ecs.g5ne.8xlarge	32	128.0	None	15.0	6,000	Yes	7,000	32	8	30	120.0	4
ecs.g5ne.16xlarge	64	256.0	None	30.0	12,000	Yes	14,000	32	8	30	240.0	8
ecs.g5ne.18xlarge	72	288.0	None	33.0	13,500	Yes	16,000	32	15	50	270.0	9

> Note
>
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

ic5, compute intensive instance family

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:1.
- Provides an ultra-high packet forwarding rate.
- Equipped with 2.5 GHz Intel ® Xeon ® Platinum 8163 (Skylake) or 8269CY (Cascade Lake) processors for consistent computing performance.
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Web frontend servers
  - Data analysis, batch processing, and video encoding
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Frontend servers of MMO games

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ic5.large	2	2.0	None	1.0	300	No	2	2	6
ecs.ic5.xlarge	4	4.0	None	1.5	500	No	2	3	10
ecs.ic5.2xlarge	8	8.0	None	2.5	800	No	2	4	10
ecs.ic5.3xlarge	12	12.0	None	4.0	900	No	4	6	10

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ic5.4xlarge	16	16.0	None	5.0	1,000	No	4	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

c6, compute optimized instance family

Features

- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the X-Dragon architecture.
- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.

  Note The maximum performance of disks varies based on instance families. A single c6 instance can deliver up to 200,000 IOPS. For higher storage I/O performance, we recommend that you use g5se. For more information, see g5se, memory optimized instance family with enhanced performance.

- Provides high storage I/O performance based on large computing capacity.

  Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Allows you to enable or disable Hyper-Threading.

  Note By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Offers a CPU-to-memory ratio of 1:2.
- Provides an ultra-high packet forwarding rate.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
- Provides high network performance based on large computing capacity.
- Supports instance type changes to g6 or r6.

- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Web frontend servers
  - Frontend servers of MMO games
  - Data analysis, batch processing, and video encoding
  - High-performance scientific and engineering applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.c6.large	2	4.0	None	1.0	3.0	300	Yes	Up to 250	2	2	6	10.0	1
ecs.c6.xlarge	4	8.0	None	1.5	5.0	500	Yes	Up to 250	4	3	10	20.0	1.5
ecs.c6.2xlarge	8	16.0	None	2.5	8.0	800	Yes	Up to 250	8	4	10	25.0	2
ecs.c6.3xlarge	12	24.0	None	4.0	10.0	900	Yes	Up to 250	8	6	10	30.0	2.5
ecs.c6.4xlarge	16	32.0	None	5.0	10.0	1,000	Yes	300	8	8	20	40.0	3

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.c6.6xlarge	24	48.0	None	7.5	10.0	1,500	Yes	450	12	8	20	50.0	4
ecs.c6.8xlarge	32	64.0	None	10.0	None	2,000	Yes	600	16	8	20	60.0	5
ecs.c6.13xlarge	52	96.0	None	12.5	None	3,000	Yes	900	32	7	20	100.0	8
ecs.c6.26xlarge	104	192.0	None	25.0	None	6,000	Yes	1,800	32	15	20	200.0	16

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

c6a, compute optimized instance family

c6a is in invitational preview. To use c6a, submit a ticket.

Features

- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the X-Dragon architecture.
- Compute:

- Equipped with 2.6 GHz AMD EPYC ™ ROME processors with a maximum turbo frequency of 3.3 GHz for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:2.
- Allows you to enable or disable Hyper-Threading.

> Note By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Storage:
  - I/O optimized.
  - Supports enhanced SSDs (ESSDs), standard SSDs, and ultra disks.
  - Provides high storage I/O performance based on large computing capacity.

> Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Network:
  - Provides an ultra-high packet forwarding rate.
  - Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Video encoding and decoding
  - Scenarios where large volumes of packets are received and transmitted
  - Web frontend servers
  - Frontend servers of massively multiplayer online (MMO) games
  - Test and development, such as DevOps

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	ENIs (including one primary ENI)	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.c6a.large	2	4.0	None	1.0	10.0	900	Yes	2	12.5	1.0
ecs.c6a.xlarge	4	8.0	None	1.5	10.0	1,000	Yes	3	20.0	1.5
ecs.c6a.2xlarge	8	16.0	None	2.5	10.0	1,600	Yes	4	30.0	2.0

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	ENIs (including one primary ENI)	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.c6a.4xlarge	16	32.0	None	5.0	10.0	2,000	Yes	8	60.0	3.0
ecs.c6a.8xlarge	32	64.0	None	8.0	10.0	3,000	Yes	7	75.0	4,0
ecs.c6a.16xlarge	64	128.0	None	16.0	None	6,000	Yes	8	150.0	8.0
ecs.c6a.32xlarge	128	256.0	None	32.0	None	12,000	Yes	15	300.0	16.0
ecs.c6a.64xlarge	256	512.0	None	64.0	None	24,000	Yes	15	600.0	32.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

c6t, security-enhanced compute optimized instance family

c6t is in invitational preview.

Features

- Implements trusted boot based on the Trusted Platform Module (TPM) chip. During a trusted boot, each module in the boot chain from the underlying hardware to the guest OS is measured and verified.
- Supports comprehensive monitoring on the IaaS layer and ensures trusted capabilities of the whole IaaS layer.
- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the third-generation X-Dragon architecture. c6t improves storage performance, network performance, and computing stability by an order of magnitude through fast path acceleration of X-Dragon chips.

- Compute:
  - Equipped with 2.5 GHz Intel® Xeon® Platinum 8269 (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
  - Offers a CPU-to-memory ratio of 1:2.
  - Allows you to enable or disable Hyper-Threading.

    Note By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Storage:
  - I/O optimized.
  - Supports ESSDs only.
  - Provides high storage I/O performance based on large computing capacity.

    Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Network:
  - Provides an ultra-high packet forwarding rate.
  - Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios that require high security and enhanced trust, such as financial services, government affairs, and enterprise services
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Web frontend servers
  - Frontend servers of MMO games
  - Data analysis, batch processing, and video encoding
  - High-performance scientific and engineering applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	TPM support	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.c6t.large	2	4.0	None	A burstable bandwidth of up to 10.0	900	Yes	Yes	Up to 250	2	3	6	20.0	1.0
ecs.c6t.xlarge	4	8.0	None	A burstable bandwidth of up to 10.0	1,000	Yes	Yes	Up to 250	4	4	15	40.0	1.5
ecs.c6t.2xlarge	8	16.0	None	A burstable bandwidth of up to 10.0	1,600	Yes	Yes	Up to 250	8	4	15	50.0	2.0

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	TPM support	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.c6t.4xlarge	16	32.0	None	A burstable bandwidth of up to 10.0	3,000	Yes	Yes	300	8	8	30	80.0	3.0
ecs.c6t.8xlarge	32	64.0	None	10.0	6,000	Yes	Yes	600	16	8	30	150.0	5.0
ecs.c6t.13xlarge	52	96.0	None	16.0	9,000	Yes	Yes	900	32	7	30	240.0	8.0
ecs.c6t.26xlarge	104	192.0	None	32.0	24,000	Yes	Yes	1,800	32	15	30	480.0	16.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
- The results for network capabilities are the maximum values obtained from single item tests. For example, when network bandwidth is tested, no stress tests are performed on the packet forwarding rate or other metrics of the network.

c6e, compute optimized instance family with enhanced performance

Features

- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the third-generation X-Dragon architecture. c6e improves storage performance, network performance, and computing stability by an order of magnitude through fast path acceleration of X-Dragon chips.
- I/O optimized.
- Supports ESSDs only.
- Provides high network and storage I/O performance based on large computing capacity.

  > **Note** For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Provides an ultra-high packet forwarding rate.

  > **Note** The maximum network performance varies based on instance families. For higher concurrent connection capabilities, we recommend that you use g5ne. For more information, see g5ne, network enhanced instance family.

- Offers a CPU-to-memory ratio of 1:2.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269 (Cascade) processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
- Allows you to enable or disable Hyper-Threading.

  > **Note** By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Web frontend servers
  - Frontend servers of MMO games
  - Data analysis, batch processing, and video encoding
  - High-performance scientific and engineering applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.c6e.large	2	4.0	None	A burstable bandwidth of up to 10.0	900	Yes	Up to 250	2	3	6	20.0	1.0
ecs.c6e.xlarge	4	8.0	None	A burstable bandwidth of up to 10.0	1,000	Yes	Up to 250	4	4	15	40.0	1.5
ecs.c6e.2xlarge	8	16.0	None	A burstable bandwidth of up to 10.0	1,600	Yes	Up to 250	8	4	15	50.0	2.0

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.c6e.4xlarge	16	32.0	None	A burstable bandwidth of up to 10.0	3,000	Yes	300	8	8	30	80.0	3.0
ecs.c6e.8xlarge	32	64.0	None	10.0	6,000	Yes	600	16	8	30	150.0	5.0
ecs.c6e.13xlarge	52	96.0	None	16.0	9,000	Yes	900	32	7	30	240.0	8.0
ecs.c6e.26xlarge	104	192.0	None	32.0	24,000	Yes	1,800	32	15	30	480.0	16.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
- The results for network capabilities are the maximum values obtained from single item tests. For example, when network bandwidth is tested, no stress tests are performed on the packet forwarding rate or other metrics of the network.

c5, compute optimized instance family

Features

- I/O optimized.

- Supports ESSDs, standard SSDs, and ultra disks.

  > **Note** The maximum performance of disks varies based on instance families. A single c5 instance can deliver up to 200,000 IOPS. For higher storage I/O performance, we recommend that you use g5se. For more information, see g5se, memory optimized instance family with enhanced performance.

- Offers a CPU-to-memory ratio of 1:2.
- Provides an ultra-high packet forwarding rate.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) or 8269CY (Cascade Lake) processors for consistent computing performance.
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Web frontend servers
  - Frontend servers of MMO games
  - Data analysis, batch processing, and video encoding
  - High-performance scientific and engineering applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.c5.large	2	4.0	None	1.0	300	Yes	2	2	6
ecs.c5.xlarge	4	8.0	None	1.5	500	Yes	2	3	10
ecs.c5.2xlarge	8	16.0	None	2.5	800	Yes	2	4	10
ecs.c5.3xlarge	12	24.0	None	4.0	900	Yes	4	6	10
ecs.c5.4xlarge	16	32.0	None	5.0	1,000	Yes	4	8	20
ecs.c5.6xlarge	24	48.0	None	7.5	1,500	Yes	6	8	20

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.c5.8xlarge	32	64.0	None	10.0	2,000	Yes	8	8	20
ecs.c5.16xlarge	64	128.0	None	20.0	4,000	Yes	16	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

r6, memory optimized instance family

Features

- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the X-Dragon architecture.
- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.

  Note The maximum performance of disks varies with instance families. A single r6 instance can deliver up to 200,000 IOPS. For higher storage I/O performance, we recommend that you use g5se. For more information, see the "g5se, storage optimized instance family with enhanced performance" section in Instance families.

- Provides high storage I/O performance based on large computing capacity.

  Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Provides an ultra-high packet forwarding rate.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:8.
- Allows you to enable or disable Hyper-Threading.

> **Note** By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Provides high network performance based on large computing capacity.
- Supports changes to g6 or c6 instance families.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - High-performance databases and in-memory databases
  - Data analysis, data mining, and distributed memory caching
  - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.r6.large	2	16.0	None	1.0	3.0	300	Yes	Up to 250	2	2	6	10	1
ecs.r6.xlarge	4	32.0	None	1.5	5.0	500	Yes	Up to 250	4	3	10	20	1.5
ecs.r6.2xlarge	8	64.0	None	2.5	8.0	800	Yes	Up to 250	8	4	10	25	2
ecs.r6.3xlarge	12	96.0	None	4.0	10.0	900	Yes	Up to 250	8	6	10	30	2.5

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.r6.4xlarge	16	128.0	None	5.0	10.0	1,000	Yes	300	8	8	20	40	3
ecs.r6.6xlarge	24	192.0	None	7.5	10.0	1,500	Yes	450	12	8	20	50	4
ecs.r6.8xlarge	32	256.0	None	10.0	None	2,000	Yes	600	16	8	20	60	5
ecs.r6.13xlarge	52	384.0	None	12.5	None	3,000	Yes	900	32	7	20	100	8
ecs.r6.26xlarge	104	768.0	None	25.0	None	6,000	Yes	1,800	32	15	20	200	16

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

r6a, memory optimized instance family

r6a is in invitational preview. To use r6a, submit a ticket.

Features

- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the X-Dragon architecture.
- Compute:
  - Equipped with 2.6 GHz AMD EPYC ™ ROME processors with a maximum turbo frequency of 3.3 GHz for consistent computing performance.
  - Offers a CPU-to-memory ratio of 1:8.
  - Allows you to enable or disable Hyper-Threading.

    Note By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Storage:
  - I/O optimized.
  - Supports enhanced SSDs (ESSDs), standard SSDs, and ultra disks.
  - Provides high storage I/O performance based on large computing capacity.

    Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Network:
  - Provides an ultra high packet forwarding rate.
  - Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Video encoding and decoding
  - Scenarios where large volumes of packets are received and transmitted
  - In-memory databases
  - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications
  - Test and development, such as DevOps

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	ENIs (including one primary ENI)	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.r6a.large	2	16.0	None	1.0	10.0	900	Yes	2	12.5	1.0
ecs.r6a.xlarge	4	32.0	None	1.5	10.0	1,000	Yes	3	20	1.5

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	ENIs (including one primary ENI)	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.r6a.2xlarge	8	64.0	None	2.5	10.0	1,600	Yes	4	30	2.0
ecs.r6a.4xlarge	16	128.0	None	5.0	10.0	2,000	Yes	8	60	3.0
ecs.r6a.8xlarge	32	256.0	None	8.0	10.0	3,000	Yes	7	75	4,0
ecs.r6a.16xlarge	64	512.0	None	16.0	None	6,000	Yes	8	150	8.0
ecs.r6a.32xlarge	128	1024.0	None	32.0	None	12,000	Yes	15	300	16.0
ecs.r6a.64xlarge	256	2048.0	None	64.0	None	24,000	Yes	15	600	32.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

r6e, memory optimized instance family with enhanced performance

Features

- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the third-generation X-Dragon architecture. In addition, improves storage performance, network performance, and computing stability by an order of magnitude through fast path acceleration of X-Dragon chips.
- I/O optimized.
- Supports ESSDs.
- Provides high network and storage I/O performance based on large computing capacity.

> **Note** For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Provides an ultra high packet forwarding rate.

  > **Note** The maximum network performance varies depending on instance families. For more concurrent connections, we recommend that you use g5ne. For more information, see the "g5ne, general purpose instance family with enhanced network performance" section in Instance families.

- Offers a CPU-to-memory ratio of 1:8.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269 processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
- Allows you to enable or disable Hyper-Threading.

  > **Note** By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - High-performance databases and in-memory databases
  - Data analysis, data mining, and distributed memory caching
  - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.r6e.large	2	16.0	None	A burstable bandwidth of up to 10.0	900	Yes	Up to 250	2	3	6	20	1.0

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.r6e.xlarge	4	32.0	None	A burstable bandwidth of up to 10.0	1,000	Yes	Up to 250	4	4	15	40	1.5
ecs.r6e.2xlarge	8	64.0	None	A burstable bandwidth of up to 10.0	1,600	Yes	Up to 250	8	4	15	50	2.0
ecs.r6e.4xlarge	16	128.0	None	A burstable bandwidth of up to 10.0	3,000	Yes	300	8	8	30	80	3.0
ecs.r6e.8xlarge	32	256.0	None	10.0	6,000	Yes	600	16	8	30	150	5.0
ecs.r6e.13xlarge	52	384.0	None	16.0	9,000	Yes	900	32	7	30	240	8.0

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.r6e.26xlarge	104	768.0	None	32.0	24,000	Yes	1,800	32	15	30	480	16.0

> **Note**
>
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.
> - The results for network capabilities are the maximum values obtained from single item tests. For example, when network bandwidth is tested, no stress tests are performed on the packet forwarding rate or other metrics of the network.

re6, high memory instance family

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Optimized for high-performance databases, in-memory databases, and other memory intensive enterprise applications.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:15 and up to 3 TiB memory.
- Suitable for the following scenarios:
  - High-performance databases and in-memory databases such as SAP HANA
  - Memory intensive applications
  - Big data processing engines such as Apache Spark and Presto

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.re6.13xlarge	52	768.0	None	10.0	1,800	Yes	16	7	20	50	4
ecs.re6.26xlarge	104	1536.0	None	16.0	3,000	Yes	32	7	20	100	8
ecs.re6.52xlarge	208	3072.0	None	32.0	6,000	Yes	32	15	20	200	16

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

r5, memory optimized instance family

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.

  Note The maximum performance of disks varies with instance families. A single r5 instance can deliver up to 200,000 IOPS. For higher storage I/O performance, we recommend that you use g5se. For more information, see the "g5se, storage optimized instance family with enhanced performance" section in Instance families.

- Provides an ultra-high packet forwarding rate.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) or Intel® Xeon® Paltinum 8269CY (Cascade Lake) processors for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:8.
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:

- Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
- High-performance databases and in-memory databases
- Data analysis, data mining, and distributed memory caching
- Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.r5.large	2	16.0	None	1.0	300	Yes	2	2	6
ecs.r5.xlarge	4	32.0	None	1.5	500	Yes	2	3	10
ecs.r5.2xlarge	8	64.0	None	2.5	800	Yes	2	4	10
ecs.r5.3xlarge	12	96.0	None	4.0	900	Yes	4	6	10
ecs.r5.4xlarge	16	128.0	None	5.0	1,000	Yes	4	8	20
ecs.r5.6xlarge	24	192.0	None	7.5	1,500	Yes	6	8	20
ecs.r5.8xlarge	32	256.0	None	10.0	2,000	Yes	8	8	20
ecs.r5.16xlarge	64	512.0	None	20.0	4,000	Yes	16	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

d2c, compute intensive big data instance family

d2c is under invitational preview. To use d2c, submit a ticket.

Features:

- I/O optimized.
- Supports enhanced SSDs, standard SSDs, and ultra disks.
- High-capacity local SATA HDDs with high throughput and a maximum of 35 Gbit/s bandwidth among instances.
-
- Equipped with 2.5 GHz Intel® Xeon ® Platinum 8269CY (Cascade Lake) processors.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Big data computing and storage business scenarios that use Hadoop MapReduce, HDFS, Hive, and HBase
  - Scenarios in which EMR JindoFS and OOS are used to store hot and cold data separately and decouple storage and computing
  - Machine learning scenarios such as in-memory computing with Spark and scalable machine learning with MLlib
  - Search and log data processing scenarios that use solutions such as Elasticsearch and Kafka

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.d2c.6xlarge	24	88.0	3 × 4,000	12.0	1,600	Yes	8	8	20
ecs.d2c.12xlarge	48	176.0	6 × 4000	20.0	2,000	Yes	16	8	20
ecs.d2c.24xlarge	96	352.0	12 × 4000	35.0	4,500	Yes	16	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

d2s, storage intensive big data instance family

Features:

- I/O-optimized.
- Supports enhanced SSDs, standard SSDs, and ultra disks.
- High-capacity local SATA HDDs with high throughput and a maximum of 35 Gbit/s bandwidth among instances.
- Supports online replacement and hot swapping of damaged disks to avoid instance shutdown.

  If a local disk fails, you will receive a notification about the system event. You can respond to the system event by initiating the process to fix the damaged disk. For more information, see Overview of system events on ECS instances equipped with local disks.

  - If a backup disk is available on the physical machine, Alibaba Cloud will replace the damaged disk with the backup disk online.
  - If no backup disk is available on the physical machine, the disk hardware must be replaced manually before Alibaba Cloud can replace the damaged disk.

  > **Notice** After you have started the process to fix the damaged disk, data in the damaged disk cannot be recovered.

- Equipped with 2.5 GHz Intel ® Xeon® Platinum 8163 (Skylake) processors.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Big data computing and storage business scenarios that use Hadoop MapReduce, HDFS, Hive, and HBase
  - Machine learning scenarios such as in-memory computing with Spark and scalable machine learning with MLlib
  - Search and log data processing scenarios that use solutions such as Elasticsearch and Kafka

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.d2s.5xlarge	20	88.0	8 × 7,300	12.0	1,600	Yes	8	8	20
ecs.d2s.10xlarge	40	176.0	15 × 7,300	20.0	2,000	Yes	16	8	20
ecs.d2s.20xlarge	80	352.0	30 × 7,300	35.0	4,500	Yes	32	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

d1ne, big data instance family with enhanced network performance

Features:

- I/O optimized.
- Supports standard SSDs and ultra disks.
- High-capacity local SATA HDDs with high throughput and a maximum of 35 Gbit/s bandwidth among instances.
- Offers a CPU-to-memory ratio of 1:4, which is designed for big data scenarios.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios that use Hadoop MapReduce, HDFS, Hive, and HBase
  - Machine learning scenarios such as in-memory computing with Spark and scalable machine learning with MLlib
  - Use of solutions such as Elasticsearch for log data processing

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.d1ne.2xlarge	8	32.0	4 × 5,500	6.0	1,000	Yes	4	4	10
ecs.d1ne.4xlarge	16	64.0	8 × 5,500	12.0	1,600	Yes	4	8	20
ecs.d1ne.6xlarge	24	96.0	12 × 5,500	16.0	2,000	Yes	6	8	20
ecs.d1ne-c8d3.8xlarge	32	128.0	12 × 5,500	20.0	2,000	Yes	6	8	20

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.d1ne.8xlarge	32	128.0	16 × 5,500	20.0	2,500	Yes	8	8	20
ecs.d1ne-c14d3.14xlarge	56	160.0	12 × 5,500	35.0	4,500	Yes	14	8	20
ecs.d1ne.14xlarge	56	224.0	28 × 5,500	35.0	4,500	Yes	14	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

i2, instance family with local SSDs

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Attached with high-performance local NVMe SSDs that feature high IOPS, high I/O throughput, and low latency.
- Offers a CPU-to-memory ratio of 1:8, which is designed for high-performance databases.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Provides a fast and reliable network based on large computing capacity.
- Scenarios:
  - Online transaction processing (OLTP) and high-performance relational databases
  - NoSQL databases such as Cassandra, MongoDB, and HBase
  - Search scenarios that use solutions such as Elasticsearch

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk bandwidth (Gbit/s)
ecs.i2.xlarge	4	32.0	1 × 894	1.0	500	Yes	2	3	10	Up to 16
ecs.i2.2xlarge	8	64.0	1 × 1,788	2.0	1,000	Yes	2	4	10	Up to 16
ecs.i2.4xlarge	16	128.0	2 × 1,788	3.0	1,500	Yes	4	8	20	Up to 16
ecs.i2.8xlarge	32	256.0	4 × 1,788	6.0	2,000	Yes	8	8	20	Up to 16
ecs.i2.16xlarge	64	512.0	8 × 1,788	10.0	4,000	Yes	16	8	20	Up to 16
ecs.i2d.21xlarge	84	712.0	4 × 3,570	25.0	4,000	Yes	32	16	20	Up to 16

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
- For more information about the performance metrics of SSDs, see Local disks.

i2g, instance family with local SSDs

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Attached with high-performance local NVMe SSDs that feature high IOPS, high I/O throughput, and low latency.
- Offers a CPU-to-memory ratio of 1:4, which is designed for high-performance databases.
- Equipped with 2.5 GHz Intel ® Xeon ® Platinum 8163 (Skylake) processors.
- Provides a fast and reliable network based on large computing capacity.

- Scenarios:
  - OLTP and high-performance relational databases
  - NoSQL databases such as Cassandra, MongoDB, and HBase
  - Search scenarios that use solutions such as Elasticsearch

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.i2g.2xlarge	8	32.0	1 × 894	2.0	1,000	No	2	4	10
ecs.i2g.4xlarge	16	64.0	1 × 1,788	3.0	1,500	No	4	8	20
ecs.i2g.8xlarge	32	128.0	2 × 1,788	6.0	2,000	No	8	8	20
ecs.i2g.16xlarge	64	256.0	4 × 1,788	10.0	4,000	No	16	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
-

i2ne, instance family with local SSDs

i2ne is in invitational preview. To use i2ne, submit a ticket.

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Attached with high-performance local NVMe SSDs that feature high IOPS, high I/O throughput, and low latency.
- Provides a bandwidth of up to 20 Gbit/s.
- Offers a CPU-to-memory ratio of 1:8, which is designed for high-performance databases.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.

- Provides a fast and reliable network based on large computing capacity.
- Scenarios:
  - OLTP and high-performance relational databases
  - NoSQL databases such as Cassandra, MongoDB, and HBase
  - Search scenarios that use solutions such as Elasticsearch

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk bandwidth (Gbit/s)
ecs.i2ne.xlarge	4	32.0	1 × 894	1.5	500	Yes	2	3	10	Up to 16
ecs.i2ne.2xlarge	8	64.0	1 × 1,788	2.5	1,000	Yes	2	4	10	Up to 16
ecs.i2ne.4xlarge	16	128.0	2 × 1,788	5.0	1,500	Yes	4	8	20	Up to 16
ecs.i2ne.8xlarge	32	256.0	4 × 1,788	10.0	2,000	Yes	8	8	20	Up to 16
ecs.i2ne.16xlarge	64	512.0	8 × 1,788	20.0	4,000	Yes	16	8	20	Up to 16

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
- For more information about the performance metrics of SSDs, see Local disks.

i2gne, instance family with local SSDs

i2gne is in invitational preview. To use i2gne, submit a ticket.

Features

- I/O optimized.

- Supports standard SSDs and ultra disks.
- Attached with high-performance local NVMe SSDs that feature high IOPS, high I/O throughput, and low latency.
- Provides a bandwidth of up to 20 Gbit/s.
- Offers a CPU-to-memory ratio of 1:4, which is designed for high-performance databases.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Provides a fast and reliable network based on large computing capacity.
- Scenarios:
  - OLTP and high-performance relational databases
  - NoSQL databases such as Cassandra, MongoDB, and HBase
  - Search scenarios that use solutions such as Elasticsearch

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.i2gne.2xlarge	8	32.0	1 × 894	2.5	1,000	No	2	4	10
ecs.i2gne.4xlarge	16	64.0	1 × 1,788	5.0	1,500	No	4	8	20
ecs.i2gne.8xlarge	32	128.0	2 × 1,788	10.0	2,000	No	8	8	20
ecs.i2gne.16xlarge	64	256.0	4 × 1,788	20.0	4,000	No	16	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
-

hfc7, compute optimized instance family with high clock speed

Features

- Uses the third-generation X-Dragon architecture to provide predictable and consistent high

performance and reduce virtualization overheads.

- I/O optimized.
- Supports enhanced SSDs (ESSDs) and provides ultra-high I/O performance.
- Provides high storage I/O performance based on large computing capacity.

> **Note** For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Offers a CPU-to-memory ratio of 1:2.
- Provides an ultra-high packet forwarding rate.
- Equipped with Intel® Xeon® Platinum 8369HB (Cooper Lake) or Intel® Xeon® Platinum 8369HC (Cooper Lake) processors that deliver a maximum turbo frequency of 3.8 GHz and a minimum clock speed of 3.3 GHz for consistent computing performance.
- Allows you to enable or disable hyper-threading.

> **Note** By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are transmitted and received
  - High-performance frontend server clusters
  - Frontend servers of massively multiplayer online (MMO) games
  - Data analysis, batch processing, and video encoding
  - High-performance science and engineering applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Bursta ble bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfc7.large	2	4.0	None	1.2	10.0	900	Yes	2	2	6	20	1.0
ecs.hfc7.xlarge	4	8.0	None	2.0	10.0	1,000	Yes	4	3	15	30	1.5

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfc7.2xlarge	8	16.0	None	3.0	10.0	1,600	Yes	8	4	15	45	2.0
ecs.hfc7.3xlarge	12	24.0	None	4.5	10.0	2,000	Yes	8	6	15	60	2.5
ecs.hfc7.4xlarge	16	32.0	None	6.0	10.0	2,500	Yes	8	8	30	75	3.0
ecs.hfc7.6xlarge	24	48.0	None	8.0	10.0	3,000	Yes	12	8	30	90	4.0
ecs.hfc7.8xlarge	32	64.0	None	10.0	None	4,000	Yes	16	8	30	105	5.0
ecs.hfc7.12xlarge	48	96.0	None	16.0	None	6,000	Yes	24	8	30	150	8.0
ecs.hfc7.24xlarge	96	192.0	None	32.0	None	12,000	Yes	32	15	30	300	16.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

hfc6, compute optimized instance family with high clock speed

Features

- Uses the X-Dragon architecture to provide predictable and consistent high performance and reduce virtualization overheads.
- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Provides high storage I/O performance based on large computing capacity.

  Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Offers a CPU-to-memory ratio of 1:2.
- Provides an ultra-high packet forwarding rate.
- Equipped with 3.1 GHz Intel® Xeon® Platinum 8269 (Cascade Lake) processors that deliver a maximum turbo frequency of 3.5 GHz for consistent computing performance.

  Note The CPU of this instance family provides a 3.1 GHz clock speed. The Intel System Studio (ISS) feature may cause a lower clock speed to be displayed. Alibaba Cloud is working on to resolve this issue. This issue does not affect the actual clock speed of your instances.

  You can run the following commands separately and use the turbostat tool to view the actual clock speed.

  ```
  yum install kernel-tools
  ```

  ```
  turbostat
  ```

- Allows you to enable or disable hyper-threading.

  Note By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are transmitted and received
  - Web frontend servers

- Frontend servers of MMO games
- Data analysis, batch processing, and video encoding
- High-performance scientific and engineering applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfc6.large	2	4.0	None	1.0	3.0	300	Yes	2	2	6	10	1.0
ecs.hfc6.xlarge	4	8.0	None	1.5	5.0	500	Yes	4	3	10	20	1.5
ecs.hfc6.2xlarge	8	16.0	None	2.5	8.0	800	Yes	8	4	10	25	2.0
ecs.hfc6.3xlarge	12	24.0	None	4.0	10.0	900	Yes	8	6	10	30	2.5
ecs.hfc6.4xlarge	16	32.0	None	5.0	10.0	1,000	Yes	8	8	20	40	3.0
ecs.hfc6.6xlarge	24	48.0	None	7.5	10.0	1,500	Yes	12	8	20	50	4.0
ecs.hfc6.8xlarge	32	64.0	None	10.0	None	2,000	Yes	16	8	20	60	5.0

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfc6.10xlarge	40	96.0	None	12.5	None	3,000	Yes	32	7	20	100	8.0
ecs.hfc6.16xlarge	64	128.0	None	20.0	None	4,000	Yes	32	8	20	120	10.0
ecs.hfc6.20xlarge	80	192.0	None	25.0	None	6,000	Yes	32	15	20	200	16.0

> **Note**
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

hfg7, general purpose instance family with high clock speed

Features

- Uses the third-generation X-Dragon architecture to provide predictable and consistent high performance and reduce virtualization overheads.
- I/O optimized.
- Supports ESSDs and provides ultra-high I/O performance.
- Provides high storage I/O performance based on large computing capacity.

> **Note** For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Offers a CPU-to-memory ratio of 1:4.
- Provides an ultra-high packet forwarding rate.
- Equipped with Intel® Xeon® Platinum 8369HB (Cooper Lake) or Intel® Xeon® Platinum 8369HC (Cooper Lake) processors that deliver a maximum turbo frequency of 3.8 GHz and a minimum

clock speed of 3.3 GHz for consistent computing performance.

- Allows you to enable or disable hyper-threading.

  > **Note** By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are transmitted and received
  - Enterprise-level applications of various types and sizes
  - Game servers
  - Small and medium-sized database systems, caches, and search clusters
  - High-performance scientific computing
  - Video encoding applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfg7.large	2	8.0	None	1.2	10.0	900	Yes	2	2	6	20	1.0
ecs.hfg7.xlarge	4	16.0	None	2.0	10.0	1,000	Yes	4	3	15	30	1.5
ecs.hfg7.2xlarge	8	32.0	None	3.0	10.0	1,600	Yes	8	4	15	45	2.0
ecs.hfg7.3xlarge	12	48.0	None	4.5	10.0	2,000	Yes	8	6	15	60	2.5

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfg7.4xlarge	16	64.0	None	6.0	10.0	2,500	Yes	8	8	30	75	3.0
ecs.hfg7.6xlarge	24	96.0	None	8.0	10.0	3,000	Yes	12	8	30	90	4.0
ecs.hfg7.8xlarge	32	128.0	None	10.0	None	4,000	Yes	16	8	30	105	5.0
ecs.hfg7.12xlarge	48	192.0	None	16.0	None	6,000	Yes	24	8	30	150	8.0
ecs.hfg7.24xlarge	96	384.0	None	32.0	None	12,000	Yes	32	15	30	300	16.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about the specifications, see Description of instance specifications.

hfg6, general purpose instance family with high clock speed

Features

- Uses the X-Dragon architecture to provide predictable and consistent high performance and reduce virtualization overheads.
- I/O optimized.

- Supports ESSDs, standard SSDs, and ultra disks.
- Provides high storage I/O performance based on large compute capacity.

  Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Offers a CPU-to-memory ratio of 1:4.
- Provides an ultra-high packet forwarding rate.
- Equipped with 3.1 GHz Intel® Xeon® Platinum 8269 (Cascade Lake) processors that deliver a maximum turbo frequency of 3.5 GHz for consistent computing performance.

  Note

- Allows you to enable or disable hyper-threading.

  Note By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are transmitted and received
  - Enterprise-level applications of various types and sizes
  - Websites and application servers
  - Game servers
  - Small and medium-sized database systems, caches, and search clusters
  - Data analysis and computing
  - Compute clusters and memory intensive data processing

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfg6.large	2	8.0	None	1.0	3.0	300	Yes	2	2	6	10	1.0
ecs.hfg6.xlarge	4	16.0	None	1.5	5.0	500	Yes	4	3	10	20	1.5

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfg6.2xlarge	8	32.0	None	2.5	8.0	800	Yes	8	4	10	25	2.0
ecs.hfg6.3xlarge	12	48.0	None	4.0	10.0	900	Yes	8	6	10	30	2.5
ecs.hfg6.4xlarge	16	64.0	None	5.0	10.0	1,000	Yes	8	8	20	40	3.0
ecs.hfg6.6xlarge	24	96.0	None	7.5	10.0	1,500	Yes	12	8	20	50	4.0
ecs.hfg6.8xlarge	32	128.0	None	10.0	None	2,000	Yes	16	8	20	60	5.0
ecs.hfg6.10xlarge	40	192.0	None	12.5	None	3,000	Yes	32	7	20	100	8.0
ecs.hfg6.16xlarge	64	256.0	None	20.0	None	4,000	Yes	32	8	20	120	10.0

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfg6.20xlarge	80	384.0	None	25.0	None	6,000	Yes	32	15	20	200	16.0

> **Note**
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

hfr7, memory optimized instance family with high clock speed

Features

- Uses the third-generation X-Dragon architecture to provide predictable and consistent high performance and reduce virtualization overheads.
- I/O optimized.
- Supports ESSDs and provides ultra-high I/O performance.
- Provides high storage I/O performance based on large computing capacity.

  > **Note** For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Offers a CPU-to-memory ratio of 1:8.
- Provides an ultra-high packet forwarding rate.
- Equipped with Intel® Xeon® Platinum 8369HB (Cooper Lake) or Intel® Xeon® Platinum 8369HC (Cooper Lake) processors that deliver a maximum turbo frequency of 3.8 GHz and a minimum clock speed of 3.3 GHz for consistent computing performance.
- Allows you to enable or disable hyper-threading.

  > **Note** By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios such as on-screen video comments and telecom data forwarding where large

volumes of packets are transmitted and received
- High-performance and in-memory databases
- Data analysis, data mining, and distributed memory caching
- Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfr7.large	2	16.0	None	1.2	10.0	900	Yes	2	2	6	20	1.0
ecs.hfr7.xlarge	4	32.0	None	2.0	10.0	1,000	Yes	4	3	15	30	1.5
ecs.hfr7.2xlarge	8	64.0	None	3.0	10.0	1,600	Yes	8	4	15	45	2.0
ecs.hfr7.3xlarge	12	96.0	None	4.5	10.0	2,000	Yes	8	6	15	60	2.5
ecs.hfr7.4xlarge	16	128.0	None	6.0	10.0	2,500	Yes	8	8	30	75	3.0
ecs.hfr7.6xlarge	24	192.0	None	8.0	10.0	3,000	Yes	12	8	30	90	4.0

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfr7.8xlarge	32	256.0	None	10.0	None	4,000	Yes	16	8	30	105	5.0
ecs.hfr7.12xlarge	48	384.0	None	16.0	None	6,000	Yes	24	8	30	150	8.0
ecs.hfr7.24xlarge	96	768.0	None	32.0	None	12,000	Yes	32	15	30	300	16.0

> **Note**
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about the specifications, see Description of instance specifications.

hfr6, memory optimized instance family with high clock speed

Features

- Uses the X-Dragon architecture to provide predictable and consistent high performance and reduce virtualization overheads.
- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Provides high storage I/O performance based on large compute capacity.

> **Note** For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Offers a CPU-to-memory ratio of 1:8.
- Provides an ultra-high packet forwarding rate.
- Equipped with 3.1 GHz Intel® Xeon® Platinum 8269 (Cascade Lake) processors that deliver a maximum turbo frequency of 3.5 GHz for consistent computing performance.

Note

- Allows you to enable or disable hyper-threading.

  Note By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are transmitted and received
  - High-performance and in-memory databases
  - Data analysis, data mining, and distributed memory caching
  - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfr6.large	2	16.0	None	1.0	3.0	300	Yes	2	2	6	10	1.0
ecs.hfr6.xlarge	4	32.0	None	1.5	5.0	500	Yes	4	3	10	20	1.5
ecs.hfr6.2xlarge	8	64.0	None	2.5	8.0	800	Yes	8	4	10	25	2.0
ecs.hfr6.3xlarge	12	96.0	None	4.0	10.0	900	Yes	8	6	10	30	2.5

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfr6.4xlarge	16	128.0	None	5.0	10.0	1,000	Yes	8	8	20	40	3.0
ecs.hfr6.6xlarge	24	192.0	None	7.5	10.0	1,500	Yes	12	8	20	50	4.0
ecs.hfr6.8xlarge	32	256.0	None	10.0	None	2,000	Yes	16	8	20	60	5.0
ecs.hfr6.10xlarge	40	384.0	None	12.5	None	3,000	Yes	32	7	20	100	8.0
ecs.hfr6.16xlarge	64	512.0	None	20.0	None	4,000	Yes	32	8	20	120	10.0
ecs.hfr6.20xlarge	80	768.0	None	25.0	None	6,000	Yes	32	15	20	200	16.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about the specifications, see Description of instance specifications.

sn2ne, general purpose instance family with enhanced network performance

Features

- I/O optimized.
- Supports standard SSDs and ultra disks only.
- Offers a CPU-to-memory ratio of 1:4.
- Provides an ultra-high packet forwarding rate.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) or Platinum 8163 (Skylake) processors for consistent computing performance.
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Enterprise-level applications of various types and sizes
  - Small and medium-sized database systems, caches, and search clusters
  - Data analysis and computing
  - Compute clusters and memory intensive data processing

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.sn2ne.large	2	8.0	None	1.0	300	Yes	2	2	6
ecs.sn2ne.xlarge	4	16.0	None	1.5	500	Yes	2	3	10
ecs.sn2ne.2xlarge	8	32.0	None	2.0	1,000	Yes	4	4	10
ecs.sn2ne.3xlarge	12	48.0	None	2.5	1,300	Yes	4	6	10
ecs.sn2ne.4xlarge	16	64.0	None	3.0	1,600	Yes	4	8	20

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.sn2ne.6xlarge	24	96.0	None	4.5	2,000	Yes	6	8	20
ecs.sn2ne.8xlarge	32	128.0	None	6.0	2,500	Yes	8	8	20
ecs.sn2ne.14xlarge	56	224.0	None	10.0	4,500	Yes	14	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

sn1ne, compute optimized instance family with enhanced network performance

Features

- I/O optimized.
- Supports standard SSDs and ultra disks only.
- Offers a CPU-to-memory ratio of 1:2.
- Provides an ultra-high packet forwarding rate.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) or Platinum 8163 (Skylake) processors for consistent computing performance.
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Web frontend servers
  - Frontend servers of MMO games
  - Data analysis, batch processing, and video encoding
  - High-performance scientific and engineering applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.sn1ne.large	2	4.0	None	1.0	300	Yes	2	2	6
ecs.sn1ne.xlarge	4	8.0	None	1.5	500	Yes	2	3	10
ecs.sn1ne.2xlarge	8	16.0	None	2.0	1,000	Yes	4	4	10
ecs.sn1ne.3xlarge	12	24.0	None	2.5	1,300	Yes	4	6	10
ecs.sn1ne.4xlarge	16	32.0	None	3.0	1,600	Yes	4	8	20
ecs.sn1ne.6xlarge	24	48.0	None	4.5	2,000	Yes	6	8	20
ecs.sn1ne.8xlarge	32	64.0	None	6.0	2,500	Yes	8	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

re4, high memory instance family

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Optimized for high-performance databases, in-memory databases, and other memory intensive enterprise applications.
- Equipped with 2.2 GHz Intel® Xeon® E7 8880 v4(Broadwell) processors with a maximum turbo frequency of 2.4 GHz for consistent computing performance.

- Offers a CPU-to-memory ratio of 1:12 and up to 1,920 GiB memory.
- The ecs.re4.20xlarge and ecs.re4.40xlarge instance types are SAP HANA-certified.
- Suitable for the following scenarios:
  - High-performance databases and in-memory databases such as SAP HANA
  - Memory intensive applications
  - Big data processing engines such as Apache Spark and Presto

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.re4.20xlarge	80	960.0	None	15.0	2,000	Yes	16	8	20
ecs.re4.40xlarge	160	1920.0	None	30.0	4,500	Yes	16	8	20

> **Note**
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

re4e, high memory instance family

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Optimized for high-performance databases, in-memory databases, and other memory intensive enterprise applications.
- Equipped with 2.2 GHz Intel® Xeon® E7 8880 v4(Broadwell) processors with a maximum turbo frequency of 2.4 GHz for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:24 and up to 3,840 GiB memory.
- Suitable for the following scenarios:
  - High-performance databases and in-memory databases such as SAP HANA
  - Memory intensive applications
  - Big data processing engines such as Apache Spark and Presto

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.re4e.40xlarge	160	3840.0	None	30.0	4,500	Yes	16	15	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

se1ne, memory optimized instance family with enhanced network performance

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Offers a CPU-to-memory ratio of 1:8.
- Provides an ultra-high packet forwarding rate.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) or Intel® Xeon® Platinum 8163 (Skylake) processors for consistent computing performance.
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - High-performance databases and in-memory databases
  - Data analysis, data mining, and distributed memory caching
  - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.se1ne.large	2	16.0	None	1.0	300	Yes	2	2	6

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.se1ne.xlarge	4	32.0	None	1.5	500	Yes	2	3	10
ecs.se1ne.2xlarge	8	64.0	None	2.0	1,000	Yes	4	4	10
ecs.se1ne.3xlarge	12	96.0	None	2.5	1,300	Yes	4	6	10
ecs.se1ne.4xlarge	16	128.0	None	3.0	1,600	Yes	4	8	20
ecs.se1ne.6xlarge	24	192.0	None	4.5	2,000	Yes	6	8	20
ecs.se1ne.8xlarge	32	256.0	None	6.0	2,500	Yes	8	8	20
ecs.se1ne.14xlarge	56	480.0	None	10.0	4,500	Yes	14	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

se1, memory optimized instance family

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Offers a CPU-to-memory ratio of 1:8.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors for consistent computing performance.
- Provides high network performance based on large computing capacity.

- Suitable for the following scenarios:
  - High-performance databases and in-memory databases
  - Data analysis, data mining, and distributed memory caching
  - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.se1.large	2	16.0	None	0.5	100	No	1	2	6
ecs.se1.xlarge	4	32.0	None	0.8	200	No	1	3	10
ecs.se1.2xlarge	8	64.0	None	1.5	400	No	1	4	10
ecs.se1.4xlarge	16	128.0	None	3.0	500	No	2	8	20
ecs.se1.8xlarge	32	256.0	None	6.0	800	No	3	8	20
ecs.se1.14xlarge	56	480.0	None	10.0	1,200	No	4	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

d1, big data instance family

Features:

- I/O optimized.
- Supports standard SSDs and ultra disks.
- High-capacity local SATA HDDs with high throughput and up to 17 Gbit/s of bandwidth among instances.

- Offers a CPU-to-memory ratio of 1:4, which is designed for big data scenarios.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios that use Hadoop MapReduce, HDFS, Hive, and HBase
  - Machine learning scenarios such as in-memory computing with Spark and scalable machine learning with MLlib
  - Suitable for customers in Internet, finance, and other industries that need to compute, store, and analyze big data
  - Use of solutions such as Elasticsearch for log data processing

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.d1.2xlarge	8	32.0	4 × 5,500	3.0	300	No	1	4	10
ecs.d1.3xlarge	12	48.0	6 × 5,500	4.0	400	No	1	6	10
ecs.d1.4xlarge	16	64.0	8 × 5,500	6.0	600	No	2	8	20
ecs.d1.6xlarge	24	96.0	12 × 5,500	8.0	800	No	2	8	20
ecs.d1-c8d3.8xlarge	32	128.0	12 × 5,500	10.0	1,000	No	4	8	20
ecs.d1.8xlarge	32	128.0	16 × 5,500	10.0	1,000	No	4	8	20
ecs.d1-c14d3.14xlarge	56	160.0	12 × 5,500	17.0	1,800	No	6	8	20

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.d1.14xlarge	56	224.0	28 × 5,500	17.0	1,800	No	6	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

i1, instance family with local SSDs

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Attached with high-performance local NVMe SSDs that feature high IOPS, high I/O throughput, and low latency.
- Offers a CPU-to-memory ratio of 1:4, which is designed for high-performance databases.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors.
- Provides a fast and reliable network based on large computing capacity.
- Scenarios:
  - OLTP and high-performance relational databases
  - NoSQL databases such as Cassandra and MongoDB
  - Search scenarios that use solutions such as Elasticsearch

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.i1.xlarge	4	16.0	2 × 104	0.8	200	No	1	3	10
ecs.i1.2xlarge	8	32.0	2 × 208	1.5	400	No	1	4	10

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.i1.3xlarge	12	48.0	2 × 312	2.0	400	No	1	6	10
ecs.i1.4xlarge	16	64.0	2 × 416	3.0	500	No	2	8	20
ecs.i1-c5d1.4xlarge	16	64.0	2 × 1,456	3.0	400	No	2	8	20
ecs.i1.6xlarge	24	96.0	2 × 624	4.5	600	No	2	8	20
ecs.i1.8xlarge	32	128.0	2 × 832	6.0	800	No	3	8	20
ecs.i1-c10d1.8xlarge	32	128.0	2 × 1,456	6.0	800	No	3	8	20
ecs.i1.14xlarge	56	224.0	2 × 1,456	10.0	1,200	No	4	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
-

hfc5, compute optimized instance family with high clock speed

Features

- I/O optimized.
- Supports standard SSDs and ultra disks only.
- Provides consistent computing performance.
- Equipped with 3.1 GHz Intel® Xeon® Gold 6149 (Skylake) processors.

- Offers a CPU-to-memory ratio of 1:2.
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - High-performance web frontend servers
  - High-performance scientific and engineering applications
  - MMO gaming and video encoding

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.hfc5.large	2	4.0	None	1.0	300	No	2	2	6
ecs.hfc5.xlarge	4	8.0	None	1.5	500	No	2	3	10
ecs.hfc5.2xlarge	8	16.0	None	2.0	1,000	No	2	4	10
ecs.hfc5.3xlarge	12	24.0	None	2.5	1,300	No	4	6	10
ecs.hfc5.4xlarge	16	32.0	None	3.0	1,600	No	4	8	20
ecs.hfc5.6xlarge	24	48.0	None	4.5	2,000	No	6	8	20
ecs.hfc5.8xlarge	32	64.0	None	6.0	2,500	No	8	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about the specifications, see Description of instance specifications.

hfg5, general purpose instance family with high clock speed

Features

- I/O optimized.
- Supports standard SSDs and ultra disks only.
- Provides consistent computing performance.
- Equipped with 3.1 GHz Intel® Xeon® Gold 6149 (Skylake) processors.
- Offers a CPU-to-memory ratio of 1:4 (excluding the instance type with 56 vCPUs).
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - High-performance web frontend servers
  - High-performance scientific and engineering applications
  - MMO gaming and video encoding

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.hfg5.large	2	8.0	None	1.0	300	No	2	2	6
ecs.hfg5.xlarge	4	16.0	None	1.5	500	No	2	3	10
ecs.hfg5.2xlarge	8	32.0	None	2.0	1,000	No	2	4	10
ecs.hfg5.3xlarge	12	48.0	None	2.5	1,300	No	4	6	10
ecs.hfg5.4xlarge	16	64.0	None	3.0	1,600	No	4	8	20
ecs.hfg5.6xlarge	24	96.0	None	4.5	2,000	No	6	8	20
ecs.hfg5.8xlarge	32	128.0	None	6.0	2,500	No	8	8	20

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.hfg5.14xlarge	56	160.0	None	10.0	4,000	No	14	8	20

> **Note**
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

vgn6i, lightweight compute optimized instance family with GPU capabilities

vgn6i is under invitational preview. To use vgn6i, submit a ticket.

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Uses NVIDIA T4 GPU computing accelerators.
- Contains virtual GPUs (vGPUs), which are the result of GPU virtualization with mediated pass-through.
  - Supports the 1/8, 1/4, and 1/2 computing capacity of NVIDIA® Tesla® T4 GPUs.
  - Supports 2, 4, and 8 GB of GPU memory.
- Offers a CPU-to-memory ratio of 1:5.
- Equipped with 2.5 GHz Intel ® Xeon ® Platinum 8163 (Skylake) processors.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Real-time rendering for cloud gaming
  - Real-time rendering for AR and VR applications
  - AI (deep learning and machine learning) inference for the elastic deployment of Internet services
  - Educational environment of deep learning
  - Modeling experiment environment of deep learning

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.vgn6i-m4.xlarge	4	23.0	None	1/4 × T4	4	3.0	500	Yes	2	4	10
ecs.vgn6i-m8.2xlarge	10	46.0	None	1/2 × T4	8	4.0	800	Yes	4	5	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

gn6i, compute optimized instance family with GPU capabilities

Features

- I/O optimized.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Supports enhanced SSDs (ESSDs) that deliver millions of IOPS, standard SSDs, and ultra disks.
- Uses NVIDIA T4 GPU computing accelerators.
  - Powered by the new NVIDIA Turing architecture.
  - 16 GB GPU memory (320 GB/s bandwidth).
  - 2,560 CUDA cores per GPU.
  - Up to 320 Turing Tensor cores.
  - Mixed-precision Tensor cores support 65 FP16 TFLOPS, 130 INT8 TOPS, and 260 INT4 TOPS.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - AI (deep learning and machine learning) inference for computer vision, speech recognition, speech synthesis, natural language processing (NLP), machine translation, and recommendation systems
  - Real-time rendering for cloud gaming

- Real-time rendering for AR and VR applications
- Graphics workstations or overloaded graphics computing
- GPU-accelerated databases
- High-performance computing

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.gn6i-c4g1.xlarge	4	15.0	None	1 × T4	16	4.0	500	Yes	2	2	10
ecs.gn6i-c8g1.2xlarge	8	31.0	None	1 × T4	16	5.0	800	Yes	2	2	10
ecs.gn6i-c16g1.4xlarge	16	62.0	None	1 × T4	16	6.0	1,000	Yes	4	3	10
ecs.gn6i-c24g1.6xlarge	24	93.0	None	1 × T4	16	7.5	1,200	Yes	6	4	10
ecs.gn6i-c24g1.12xlarge	48	186.0	None	2 × T4	32	15.0	2,400	Yes	12	6	10
ecs.gn6i-c24g1.24xlarge	96	372.0	None	4 × T4	64	30.0	4,800	Yes	24	8	10

> **Note**
>
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

gn6e, compute optimized instance family with GPU capabilities

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Uses NVIDIA V100 (32 GB NVLink) GPU processors.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Uses NVIDIA V100 GPU computing accelerators that have the SXM2 module.
  - Powered by the new NVIDIA Volta architecture.
  - 32 GB HBM2 GPU memory (900 GB/s bandwidth).
  - 5,120 CUDA cores per GPU.
  - 640 Tensor cores per GPU.
  - Supports up to six NVLink connections and a total bandwidth of 300 GB/s (25 GB/s per connection).
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Deep learning applications such as training and inference applications of AI algorithms used in image classification, autonomous vehicles, and speech recognition
  - Scientific computing applications, such as fluid dynamics, finance, molecular dynamics, and environmental analysis

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.gn6e-c12g1.3xlarge	12	92.0	None	1 × V100	32	5.0	800	Yes	8	6	10

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.gn6e-c12g1.12xlarge	48	368.0	None	4 × V100	128	16.0	2,400	Yes	8	8	20
ecs.gn6e-c12g1.24xlarge	96	736.0	None	8 × V100	256	32.0	4,800	Yes	16	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

gn6v, compute optimized instance family with GPU capabilities

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Uses NVIDIA V100 GPU processors.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Uses NVIDIA V100 GPU computing accelerators that have the SXM2 module.
  - Powered by the new NVIDIA Volta architecture.
  - 16 GB HBM2 GPU memory (900 GB/s bandwidth).
  - 5,120 CUDA cores per GPU.
  - 640 Tensor cores per GPU.
  - Supports up to six NVLink connections and a total bandwidth of 300 GB/s (25 GB/s per connection).
- Provides a fast and reliable network based on large computing capacity.

- Suitable for the following scenarios:
  - Deep learning applications such as training and inference applications of AI algorithms used in image classification, autonomous vehicles, and speech recognition
  - Scientific computing applications, such as fluid dynamics, finance, molecular dynamics, and environmental analysis

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.gn6v-c8g1.2xlarge	8	32.0	None	1 × NVIDIA V100	1 × 16	2.5	800	Yes	4	4	10
ecs.gn6v-c8g1.8xlarge	32	128.0	None	4 × NVIDIA V100	4 × 16	10.0	2,000	Yes	8	8	20
ecs.gn6v-c8g1.16xlarge	64	256.0	None	8 × NVIDIA V100	8 × 16	20.0	2,500	Yes	16	8	20
ecs.gn6v-c10g1.20xlarge	82	336.0	None	8 × NVIDIA V100	8 × 16	32.0	4,500	Yes	16	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

f3, compute optimized instance family with FPGAs

Features:

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Uses Xilinx 16nm Virtex UltraScale+ VU9P FPGAs.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Deep learning and inference
  - Genomics research
  - Database acceleration
  - Image transcoding such as conversion of JPEG images to WebP images
  - Real-time video processing such as H.265 video compression

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	FPGAs	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.f3-c4f1.xlarge	4	16.0	None	1 × Xilinx VU9P	1.5	300	No	2	3	10
ecs.f3-c8f1.2xlarge	8	32.0	None	1 × Xilinx VU9P	2.5	500	No	4	4	10
ecs.f3-c16f1.4xlarge	16	64.0	None	1 × Xilinx VU9P	5.0	1,000	No	4	8	20
ecs.f3-c16f1.8xlarge	32	128.0	None	2 × Xilinx VU9P	10.0	2,000	No	8	8	20

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	FPGAs	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.f3-c16f1.16xlarge	64	256.0	None	4 × Xilinx VU9P	20.0	2,500	No	16	8	20
ecs.f3-c22f1.22xlarge	88	336.0	None	4 × Xilinx VU9P	30.0	4,500	No	16	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

vgn5i, lightweight compute optimized instance family with GPU capabilities

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Uses NVIDIA P4 GPU computing accelerators.
- Contains virtual GPUs (vGPUs), which are the result of GPU virtualization with mediated pass-through.
  - Supports the 1/8, 1/4, 1/2, and 1/1 computing capacity of NVIDIA® Tesla® P4 GPUs.
  - Supports 1, 2, 4, and 8 GB of GPU memory.
- Offers a CPU-to-memory ratio of 1:3.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Real-time rendering for cloud gaming
  - Real-time rendering for AR and VR applications
  - AI (deep learning and machine learning) inference for the elastic deployment of Internet services

- Educational environment of deep learning
- Modeling experiment environment of deep learning

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.vgn5i-m1.large	2	6.0	None	1/8 × P4	1	1.0	300	Yes	2	2	6
ecs.vgn5i-m2.xlarge	4	12.0	None	1/4 × P4	2	2.0	500	Yes	2	3	10
ecs.vgn5i-m4.2xlarge	8	24.0	None	1/2 × P4	4	3.0	800	Yes	2	4	10
ecs.vgn5i-m8.4xlarge	16	48.0	None	1 × P4	8	5.0	1,000	Yes	4	5	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

gn5, compute optimized instance family with GPU capabilities

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Uses NVIDIA P100 GPU processors.
- Offers multiple CPU-to-memory ratios.
- Attached with high-performance local NVMe SSDs.

- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Deep learning
  - Scientific computing applications, such as fluid dynamics, finance, genomics, and environmental analysis
  - Server-side GPU compute workloads such as high-performance computation, rendering, and multi-media coding and decoding

Instance types

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

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

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

gn5i, compute optimized instance family with GPU capabilities

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Uses NVIDIA P4 GPU processors.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Deep learning and inference
  - Server-side GPU compute workloads such as multimedia encoding and decoding

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.gn5i-c2g1.large	2	8.0	None	1 × NVIDIA P4	1 × 8	1.0	100	Yes	2	2	6
ecs.gn5i-c4g1.xlarge	4	16.0	None	1 × NVIDIA P4	1 × 8	1.5	200	Yes	2	3	10
ecs.gn5i-c8g1.2xlarge	8	32.0	None	1 × NVIDIA P4	1 × 8	2.0	400	Yes	4	4	10
ecs.gn5i-c16g1.4xlarge	16	64.0	None	1 × NVIDIA P4	1 × 8	3.0	800	Yes	4	8	20
ecs.gn5i-c16g1.8xlarge	32	128.0	None	2 × NVIDIA P4	2 × 8	6.0	1,200	Yes	8	8	20
ecs.gn5i-c28g1.14xlarge	56	224.0	None	2 × NVIDIA P4	2 × 8	10.0	2,000	Yes	14	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

gn4, compute optimized family with GPU capabilities

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Uses NVIDIA M40 GPU processors.
- Offers multiple CPU-to-memory ratios.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Deep learning
  - Scientific computing applications, such as fluid dynamics, finance, genomics, and environmental analysis
  - Server-side GPU compute workloads such as high-performance computation, rendering, and multi-media coding and decoding

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.gn4-c4g1.xlarge	4	30.0	None	1 × NVIDIA M40	1 × 12	3.0	300	No	1	3	10
ecs.gn4-c8g1.2xlarge	8	30.0	None	1 × NVIDIA M40	1 × 12	3.0	400	No	1	4	10
ecs.gn4.8xlarge	32	48.0	None	1 × NVIDIA M40	1 × 12	6.0	800	No	3	8	20
ecs.gn4-c4g1.2xlarge	8	60.0	None	2 × NVIDIA M40	2 × 12	5.0	500	No	1	4	10

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.gn4-c8g1.4xlarge	16	60.0	None	2 × NVIDIA M40	2 × 12	5.0	500	No	1	8	20
ecs.gn4.14xlarge	56	96.0	None	2 × NVIDIA M40	2 × 12	10.0	1,200	No	4	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

f1, compute optimized instance family with FPGAs

Features:

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Uses Intel® Arria® 10 GX 1150 FPGAs.
- Offers a CPU-to-memory ratio of 1:7.5.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Deep learning and inference
  - Genomics research
  - Financial analysis
  - Image transcoding
  - Computational workloads such as real-time video processing and security management

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	FPGAs	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.f1-c8f1.2xlarge	8	60.0	None	Intel Arria 10 GX 1150	3.0	400	Yes	4	4	10
ecs.f1-c8f1.4xlarge	16	120.0	None	2 × Intel Arria 10 GX 1150	5.0	1,000	Yes	4	8	20
ecs.f1-c28f1.7xlarge	28	112.0	None	Intel Arria 10 GX 1150	5.0	2,000	Yes	8	8	20
ecs.f1-c28f1.14xlarge	56	224.0	None	2 × Intel Arria 10 GX 1150	10.0	2,000	Yes	14	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmgn6e, GPU-accelerated compute optimized ECS Bare Metal Instance family

ebmgn6e is in invitational preview. To use ebmgn6e, submit a ticket.

Features

- Provides flexible and powerful software-defined compute based on the X-Dragon architecture.
- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Uses NVIDIA V100 (32 GB NVLink) GPU processors.
- Offers a CPU-to-memory ratio of 1:8.

- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Uses NVIDIA V100 GPU computing accelerators (SXM2-based).
  - Powered by the new NVIDIA Volta architecture.
  - Equipped with 32 GB HBM2 GPU memory (900 GB/s bandwidth) per GPU.
  - Equipped with 5,120 CUDA cores per GPU.
  - Equipped with 640 Tensor cores per GPU.
  - Supports up to six NVLink connections for a total bandwidth of 300 GB/s (25 GB/s per connection).
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Deep learning applications such as training and inference applications of AI algorithms used in image classification, autonomous vehicles, and speech recognition
  - Scientific computing applications such as computational fluid dynamics, computational finance, molecular dynamics, and environmental analysis

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmgn6e.24xlarge	96	768.0	None	V100*8	256	32.0	4,800	Yes	16	15	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmgn6v, GPU-accelerated compute optimized ECS Bare Metal Instance family

Features

- Provides flexible and powerful software-defined compute based on the X-Dragon architecture.
- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.

- Uses NVIDIA V100 GPU processors.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Uses NVIDIA V100 GPU computing accelerators (SXM2-based).
  - Powered by the new NVIDIA Volta architecture.
  - Equipped with 16 GB HBM2 GPU memory (900 GB/s bandwidth) per GPU.
  - Equipped with 5,120 CUDA cores per GPU.
  - Equipped with 640 Tensor cores per GPU.
  - Supports up to six NVLink connections for a total bandwidth of 300 GB/s (25 GB/s per connection).
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Deep learning applications such as training and inference applications of AI algorithms used in image classification, autonomous vehicles, and speech recognition
  - Scientific computing applications such as computational fluid dynamics, computational finance, molecular dynamics, and environmental analysis

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmgn6v.24xlarge	96	384.0	None	V100*8	128	30.0	4,500	Yes	8	32	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmgn6i, GPU-accelerated compute optimized ECS Bare Metal Instance family

Features

- Provides flexible and powerful software-defined compute based on the X-Dragon architecture.

- I/O optimized.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Supports ESSDs that deliver millions of IOPS, standard SSDs, and ultra disks.
- Uses NVIDIA T4 GPU computing accelerators.
  - Powered by the new NVIDIA Turing architecture.
  - Equipped with 16 GB memory (320 GB/s bandwidth) per GPU.
  - Equipped with 2,560 CUDA cores per GPU.
  - Euipped with up to 320 Turing Tensor cores per GPU.
  - Mixed-precision Tensor cores support 65 FP16 TFLOPS, 130 INT8 TOPS, and 260 INT4 TOPS.
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - AI (deep learning and machine learning) inference for computer vision, speech recognition, speech synthesis, natural language processing (NLP), machine translation, and recommendation systems
  - Real-time rendering for cloud gaming
  - Real-time rendering for AR and VR applications
  - Graphics workstations or overloaded graphics computing
  - GPU-accelerated databases
  - High-performance computing

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmgn6i.24xlarge	96	384.0	None	T4*4	64	30.0	4,500	Yes	8	32	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmc6a, compute optimized ECS Bare Metal Instance family

ebmc6a is in invitational preview. To use ebmc6a, submit a ticket.

Features

- Provides predictable and consistent ultra-high computing, storage, and network performance with the use of the fast path acceleration based on the third-generation X-Dragon architecture.
- Provides dedicated hardware resources and physical isolation.
- Equipped with 2.6 GHz AMD EPYC ™ ROME processors with a maximum turbo frequency of 3.3 GHz for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:2.
- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Provides ultra-high network performance with a packet forwarding rate of 24,000 Kpps.
- Supports VPCs only.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Video encoding, decoding, and rendering
  - Data analysis and computing

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	ENIs (including one primary ENI)	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.ebmc6a.64xlarge	256	512.0	None	64.0	24,000	Yes	31	480	32.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmc6e, compute optimized ECS Bare Metal Instance family with enhanced performance

Features

- Provides predictable and consistent ultra-high computing, storage, and network performance with the use of the fast path acceleration based on the third-generation X-Dragon architecture.
- Provides dedicated hardware resources and physical isolation.
- I/O optimized.
- Supports ESSDs.
- Offers a CPU-to-memory ratio of 1:2.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz.
- Provides ultra-high network performance with a packet forwarding rate of 24,000 Kpps.
- Supports VPCs only.
- Suitable for the following scenarios:
  - Scenarios that have high security and regulatory requirements, such as deploying core database services
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Web frontend servers
  - Frontend servers of massively multiplayer online (MMO) games
  - Data analysis, batch processing, and video encoding
  - High-performance scientific and engineering applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.ebmc6e.26xlarge	104	192.0	None	30.0	24,000	Yes	1,800	16	31	10	480	20.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmc6, compute optimized ECS Bare Metal Instance family

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:1.8.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz.
- Provides high network performance with a packet forwarding rate of 6,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Video encoding, decoding, and rendering
  - Frontend servers of MMO games
  - High-performance scientific and engineering applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmc6.26xlarge	104	192.0	None	30.0	6,000	Yes	8	32	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmg6a, general purpose ECS Bare Metal Instance family

ebmg6a is in invitational preview. To use ebmg6a, submit a ticket.

Features

- Provides predictable and consistent ultra-high computing, storage, and network performance with the use of the fast path acceleration based on the third-generation X-Dragon architecture.
- Provides dedicated hardware resources and physical isolation.
- Equipped with 2.6 GHz AMD EPYC ™ ROME processors with a maximum turbo frequency of 3.3 GHz for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:4.
- I/O optimized.
- Supports enhanced SSDs (ESSDs), standard SSDs, and ultra disks.
- Provides ultra-high network performance with a packet forwarding rate of 24,000 Kpps.
- Supports VPCs only.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Video encoding, decoding, and rendering
  - Computing clusters and memory intensive data processing
  - Data analysis and computing

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	ENIs (including one primary ENI)	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.ebmg6a.64xlarge	256	1024.0	None	64.0	24,000	Yes	31	480	32.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmg6e, general purpose ECS Bare Metal Instance family with enhanced performance

Features

- Provides predictable and consistent ultra-high computing, storage, and network performance with the use of the fast path acceleration based on the third-generation X-Dragon architecture.
- Provides dedicated hardware resources and physical isolation.
- I/O optimized.
- Supports ESSDs.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel ® Xeon ® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz.
- Provides ultra-high network performance with a packet forwarding rate of 24,000 Kpps.
- Supports VPCs only.
- Suitable for the following scenarios:
  - Scenarios that have high security and regulatory requirements, such as deploying core database services
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Enterprise-level applications of various types and sizes
  - Websites and application servers
  - Game servers
  - Small and medium-sized database systems, caches, and search clusters
  - Data analysis and computing
  - Compute clusters and memory intensive data processing
  - High-performance scientific and engineering applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.ebmg6e.26xlarge	104	384.0	None	30.0	24,000	Yes	1,800	16	31	10	480	20.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmg6, general purpose ECS Bare Metal Instance family

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:3.7.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz.
- Provides high network performance with a packet forwarding rate of 6,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Video encoding, decoding, and rendering
  - Enterprise-level applications such as large and medium-sized databases
  - Computing clusters and memory intensive data processing
  - Data analysis and computing

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmg6.26xlarge	104	384.0	None	30.0	6,000	Yes	8	32	10

> **Note**
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

ebmr6a, memory optimized ECS Bare Metal Instance family

ebmr6a is in invitational preview. To use ebmr6a, submit a ticket.

Features

- Provides predictable and consistent ultra-high computing, storage, and network performance with the use of the fast path acceleration based on the third-generation X-Dragon architecture.
- Provides dedicated hardware resources and physical isolation.
- Equipped with 2.6 GHz AMD EPYC ™ ROME processors with a maximum turbo frequency of 3.3 GHz for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:8.
- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Provides ultra-high network performance with a packet forwarding rate of 24,000 Kpps.
- Supports VPCs only.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - In-memory databases
  - Data analysis, data mining, and distributed memory caching
  - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	ENIs (including one primary ENI)	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.ebmr6a.64xlarge	256	2048.0	None	64.0	24,000	Yes	31	480	32.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmr6e, memory optimized ECS Bare Metal Instance family with enhanced performance

Features

- Provides predictable and consistent ultra-high computing, storage, and network performance with the use of the fast path acceleration based on the third-generation X-Dragon architecture.
- Provides dedicated hardware resources and physical isolation.
- I/O optimized.
- Supports ESSDs.
- Offers a CPU-to-memory ratio of 1:8.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz.
- Provides ultra-high network performance with a packet forwarding rate of 24,000 Kpps.
- Supports VPCs only.
- Suitable for the following scenarios:
  - Scenarios that have high security and regulatory requirements, such as deploying core database services
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - High-performance databases and in-memory databases
  - Data analysis, data mining, and distributed memory caching

- Hadoop clusters, Spark clusters, and other memory intensive enterprise applications
- High-performance scientific and engineering applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.ebmr6e.26xlarge	104	768.0	None	30.0	24,000	Yes	1,800	16	31	10	480	20.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmr6, memory optimized ECS Bare Metal Instance family

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:7.4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz.
- Provides high network performance with a packet forwarding rate of 6,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - High-performance databases and in-memory databases
  - Data analysis, data mining, and distributed memory caching

- Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmr6.26xlarge	104	768.0	None	30.0	6,000	Yes	8	32	10

> **Note**
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

ebmhfc6, compute optimized ECS Bare Metal Instance family with high clock speed

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:2.4.
- Equipped with 3.1 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.5 GHz.
- Provides high network performance with a packet forwarding rate of 6,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Video encoding, decoding, and rendering

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmhfc6.20xlarge	80	192.0	None	30.0	6,000	Yes	8	32	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmhfg6, general purpose ECS Bare Metal Instance family with high clock speed

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:4.8.
- Equipped with 3.1 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.5 GHz.
- Provides high network performance with a packet forwarding rate of 6,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Enterprise-level applications such as large and medium-sized databases
  - Video encoding, decoding, and rendering

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmhfg6.20xlarge	80	384.0	None	30.0	6,000	Yes	8	32	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmhfr6, memory optimized ECS Bare Metal Instance family with high clock speed

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:9.6.
- Equipped with 3.1 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.5 GHz.
- Provides high network performance with a packet forwarding rate of 6,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - High-performance databases and in-memory databases
  - Data analysis, data mining, and distributed memory caching
  - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmhfr6.20xlarge	80	768.0	None	30.0	6,000	Yes	8	32	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmre6p, persistent memory optimized ECS Bare Metal Instance family with enhanced performance

To use ebmre6p, submit a ticket.

Features

- Uses Intel® Optane™ non-volatile memory.
- Cost-effective due to end-to-end optimization for ApsaraDB for Redis scenarios.
- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
- Supports a maximum of 1,920 GiB memory (384 GiB DRAM memory + 1,536 GiB Intel® Optane™ non-volatile memory), offers a CPU-to-memory ratio of about 1:20, and can meet the needs of memory intensive applications.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - In-memory databases such as ApsaraDB for Redis
  - High-performance databases such as SAP HANA
  - Other memory intensive applications such as AI applications and smart search applications

Instance types

Instance type	vCPUs	DRAM (GiB)	AEP memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmre6p.26xlarge	104	384.0	1536.0	None	30.0	6,000	Yes	16	31	10

> **Note**
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

ebmre6-6t, memory optimized ECS Bare Metal Instance family with enhanced performance

To use ebmre6-6t, submit a ticket.

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:30.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269 (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz.
- Provides high network performance with a packet forwarding rate of 6,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - High-performance databases and in-memory databases such as SAP HANA
  - Memory intensive applications
  - Big data processing engines such as Apache Spark and Presto

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmre6-6t.52xlarge	208	6144.0	None	30.0	6,000	Yes	8	32	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

scchfc6, compute optimized SCC instance family with high clock speed

To use this instance family, submit a ticket.

Features

- I/O optimized.
- Supports enhanced SSDs (ESSDs), standard SSDs, and ultra disks.
- Supports both RoCE and VPCs, of which RoCE is dedicated to RDMA communication.
- Provides all the features of ECS Bare Metal Instance.
- Equipped with 3.1 GHz Intel® Xeon® Platinum 8269 (Cascade Lake) processors with a maximum turbo frequency of 3.5 GHz.
- Offers a CPU-to-memory ratio of 1:2.4.
- Scenarios:
  - Large-scale machine learning training
  - Large-scale high performance scientific computing and simulations
  - Large-scale data analysis, batch processing, and video encoding

Instance types

Instance type	vCPU	Physical cores	Memory (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	RoCE (Gbit/s)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.scchfc6.20xlarge	80	40	192.0	None	30.0	6,000	50	Yes	8	32	10

> **Note**
>
> - ecs.scchfc6.20xlarge provides 80 logical processors on 40 physical cores.
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

scchfg6, general purpose SCC instance family with high clock speed

To use this instance family, submit a ticket.

Features

- I/O optimized.
- Supports enhanced SSDs, standard SSDs, and ultra disks.
- Supports both RoCE and VPCs, of which RoCE is dedicated to RDMA communication.
- Provides all the features of ECS Bare Metal Instance.
- Equipped with 3.1 GHz Intel® Xeon® Platinum 8269 (Cascade Lake) processors with a maximum turbo frequency of 3.5 GHz.
- Offers a CPU-to-memory ratio of 1:4.8.
- Scenarios:
  - Large-scale machine learning training
  - Large-scale high performance scientific computing and simulations
  - Large-scale data analysis, batch processing, and video encoding

Instance types

Instance type	vCPU	Physical cores	Memory (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	RoCE (Gbit/s)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.scchfg6.20xlarge	80	40	384.0	None	30.0	6,000	50	Yes	8	32	10

> **Note**
> - ecs.scchfg6.20xlarge provides 80 logical processors on 40 physical cores.
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

scchfr6, memory optimized SCC instance family with high clock speed

To use this instance family, submit a ticket.

Features

- I/O optimized.
- Supports enhanced SSDs, standard SSDs, and ultra disks.
- Supports both RoCE and VPCs, of which RoCE is dedicated to RDMA communication.
- Provides all the features of ECS Bare Metal Instance.
- Equipped with 3.1 GHz Intel® Xeon® Platinum 8269 (Cascade Lake) processors with a maximum turbo frequency of 3.5 GHz.
- Offers a CPU-to-memory ratio of 1:9.6.
- Scenarios:
  - Large-scale machine learning training
  - Large-scale high performance scientific computing and simulations
  - Large-scale data analysis, batch processing, and video encoding

Instance types

Instance type	vCPU	Physical cores	Memory (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	RoCE (Gbit/s)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.scchfr6.20xlarge	80	40	768.0	None	30.0	6,000	50	Yes	8	32	10

> **Note**
>
> - ecs.scchfr6.20xlarge provides 80 logical processors on 40 physical cores.
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

scch5, SCC instance family with high clock speed

Features

- I/O optimized.
- Supports only standard SSDs and ultra disks.
- Supports both RoCE and VPCs, of which RoCE is dedicated to RDMA communication.
- Provides all the features of ECS Bare Metal Instance.
- Equipped with 3.1 GHz Intel® Xeon® Gold 6149 (Skylake) processors.
- Offers a CPU-to-memory ratio of 1:3.
- Scenarios:
  - Large-scale machine learning training
  - Large-scale high performance scientific computing and simulations
  - Large-scale data analysis, batch processing, and video encoding

Instance types

Instance type	vCPU	Physical cores	Memory (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	RoCE (Gbit/s)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.scch5.16xlarge	64	32	192.0	None	10.0	4,500	25 × 2	No	8	32	10

> **Note**
> - ecs.scch5.16xlarge provides 64 logical processors on 32 physical cores.
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

sccg5, general purpose SCC instance family

Features

- I/O optimized.
- Supports only standard SSDs and ultra disks.
- Supports both RoCE and VPCs, of which RoCE is dedicated to RDMA communication.
- Provides all the features of ECS Bare Metal Instance.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Offers a CPU-to-memory ratio of 1:4.
- Scenarios:
  - Large-scale machine learning training
  - Large-scale high performance scientific computing and simulations
  - Large-scale data analysis, batch processing, and video encoding

Instance types

Instance type	vCPU	Physical cores	Memory (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	RoCE (Gbit/s)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.sccg5.24xlarge	96	48	384.0	None	10.0	4,500	25 × 2	No	8	32	10

> **Note**
> - ecs.sccg5.24xlarge provides 96 logical processors on 48 physical cores.
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

sccgn6, compute optimized SCC instance family with GPU capabilities

Features

- I/O optimized.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Provides all the features of ECS Bare Metal Instance.
- Storage:
  - Supports enhanced SSDs, standard SSDs, and ultra disks
  - Supports a high performance CPFS
- Networking:
  - Supports VPCs
  - Supports the RoCE v2 network, which is dedicated to low-latency RDMA communication
- Uses NVIDIA V100 GPU processors that have the SXM2 module:
  - Powered by the new NVIDIA Volta architecture
  - Offers a 16 GB HBM2 GPU memory
  - CUDA Cores 5120
  - Tensor Cores 640
  - Offers a GPU memory bandwidth of up to 900 GB/s
  - Supports up to six NVLink connections and total bandwidth of 300 GB/s (25 GB/s per connection)
- Scenarios:

- Ultra-large-scale machine learning training on a distributed GPU cluster
- Large-scale high performance scientific computing and simulations
- Large-scale data analysis, batch processing, and video encoding

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	RoCE (Gbit/s)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.sccgn6.24xlarge	96	384.0	None	V100*8	30	4,500	25 × 2	Yes	8	32	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

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

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:2.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors with a maximum turbo frequency of 2.7 GHz.
- Provides high network performance with a packet forwarding rate of 4,500 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack

- Containers including Docker, Clear Containers, and Pouch
- Video encoding, decoding, and rendering

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmc5s.24xlarge	96	192.0	None	30.0	4,500	No	8	32	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

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

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors with a maximum turbo frequency of 2.7 GHz.
- Provides high network performance with a packet forwarding rate of 4,500 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Enterprise-level applications such as large and medium-sized databases
  - Video encoding

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmg5s.24xlarge	96	384.0	None	30.0	4,500	No	8	32	10

> **Note**
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

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

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:8.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors with a maximum turbo frequency of 2.7 GHz.
- Provides high network performance with a packet forwarding rate of 4,500 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - High-performance databases and in-memory databases
  - Data analysis, data mining, and distributed memory caching
  - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmr5s.24xlarge	96	768.0	None	30.0	4,500	No	8	32	10

> **Note**
>
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

ebmg5, general purpose ECS Bare Metal Instance family

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors with a maximum turbo frequency of 2.7 GHz.
- Provides high network performance with a packet forwarding rate of 4,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Enterprise-level applications such as large and medium-sized databases
  - Video encoding

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmg5.24xlarge	96	384.0	None	10.0	4,000	No	8	32	10

> **Note**
>
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

ebmhfg5, ECS Bare Metal Instance family with high clock speed

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 3.7 GHz Intel® Xeon® E3-1240v6 (Skylake) processors with a maximum turbo frequency of 4.1 GHz.
- Provides high network performance with a packet forwarding rate of 2,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Disabled failover by default.

  You can call the ModifyInstanceMaintenanceAttributes operation to modify the maintenance action. Set ActionOnMaintenance to *AutoRedeploy* to enable failover.
- Supports Intel® SGX.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Gaming and finance applications that require high performance
  - High-performance web servers
  - Enterprise-level applications such as high-performance databases

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmhfg5.2xlarge	8	32.0	None	6.0	2,000	No	8	6	8

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmc4, compute optimized ECS Bare Metal Instance family

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Offers a CPU-to-memory ratio of 1:2.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors with a maximum turbo frequency of 2.9 GHz.
- Provides high network performance with a packet forwarding rate of 4,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Enterprise-level applications such as large and medium-sized databases
  - Video encoding

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmc4.8xlarge	32	64.0	None	10.0	4,000	No	8	12	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

t6, burstable instance family

Features:

- More cost-effective when compared with the t5 burstable instance family.
- Equipped with 2.5 GHz Intel® Xeon® Cascade Lake processors that deliver a maximum turbo frequency of 3.2 GHz.
- Provides baseline CPU performance and is burstable, but limited by accumulated CPU credits.
- Paired with the DDR4 memory.
- Supports enhanced SSDs (ESSD), standard SSDs, and ultra disks.

  Note PL2 and PL3 ESSDs cannot provide maximum performance due to the specification limits of burstable instances. We recommend that you use enterprise-level instances or ESSDs that are of lower performance levels.

- Delivers a bandwidth of up to 4 Gbit/s.
- Supports VPCs only.
- Suitable for the following scenarios:
  - Web application servers
  - Lightweight applications and microservices
  - Development and testing environments

Instance types

Instance type	vCPUs	Memory (GiB)	Baseline CPU computing performance	CPU credits per hour	Max CPU credit balance	Local storage (GiB)	Base bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.t6-c4m1.large	2	0.5	5%	6	144	None	0.08	40	Yes	1	2	2
ecs.t6-c2m1.large	2	1.0	10%	12	288	None	0.08	60	Yes	1	2	2
ecs.t6-c1m1.large	2	2.0	20%	24	576	None	0.08	100	Yes	1	2	2
ecs.t6-c1m2.large	2	4.0	20%	24	576	None	0.08	100	Yes	1	2	2
ecs.t6-c1m4.large	2	8.0	30%	36	864	None	0.08	100	Yes	1	2	2
ecs.t6-c1m4.xlarge	4	16.0	40%	96	2304	None	0.16	200	Yes	1	2	6
ecs.t6-c1m4.2xlarge	8	32.0	40%	192	4608	None	0.32	400	Yes	1	2	6

> **Note**
>
> - When you bind an ENI to or unbind an ENI from an instance of the following instance types, the instance must be in the Stopped state: ecs.t6-c1m1.large, ecs.t6-c1m2.large, ecs.t6-c1m4.large, ecs.t6-c2m1.large, and ecs.t6-c4m1.large.
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

t5, burstable instance family

Features:

- Equipped with 2.5 GHz Intel® Xeon® processors.
- Paired with the DDR4 memory.
- Offers multiple CPU-to-memory ratios.
- Provides baseline CPU performance and is burstable, but limited by accumulated CPU credits.
- Offers a balance of compute, memory, and network resources.
- Supports VPCs only.
- Suitable for the following scenarios:
  - Web application servers
  - Lightweight applications and microservices
  - Development and testing environments

Instance types

Instance type	vCPUs	Memory (GiB)	Baseline CPU computing performance	CPU credits per hour	Max CPU credit balance	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.t5-lc2m1.nano	1	0.5	20%	12	288	None	0.1	40	Yes	1	2	2
ecs.t5-lc1m1.small	1	1.0	20%	12	288	None	0.2	60	Yes	1	2	2

Instance type	vCPUs	Memory (GiB)	Baseline CPU computing performance	CPU credits per hour	Max CPU credit balance	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.t5-lc1m2.small	1	2.0	20%	12	288	None	0.2	60	Yes	1	2	2
ecs.t5-lc1m2.large	2	4.0	20%	24	576	None	0.4	100	Yes	1	2	2
ecs.t5-lc1m4.large	2	8.0	20%	24	576	None	0.4	100	Yes	1	2	2
ecs.t5-c1m1.large	2	2.0	25%	30	720	None	0.5	100	Yes	1	2	2
ecs.t5-c1m2.large	2	4.0	25%	30	720	None	0.5	100	Yes	1	2	2
ecs.t5-c1m4.large	2	8.0	25%	30	720	None	0.5	100	Yes	1	2	2
ecs.t5-c1m1.xlarge	4	4.0	25%	60	1440	None	0.8	200	Yes	1	2	6

Instance type	vCPUs	Memory (GiB)	Baseline CPU computing performance	CPU credits per hour	Max CPU credit balance	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.t5-c1m2.xlarge	4	8.0	25%	60	1440	None	0.8	200	Yes	1	2	6
ecs.t5-c1m4.xlarge	4	16.0	25%	60	1440	None	0.8	200	Yes	1	2	6
ecs.t5-c1m1.2xlarge	8	8.0	25%	120	2880	None	1.2	400	Yes	1	2	6
ecs.t5-c1m2.2xlarge	8	16.0	25%	120	2880	None	1.2	400	Yes	1	2	6
ecs.t5-c1m4.2xlarge	8	32.0	25%	120	2880	None	1.2	400	Yes	1	2	6
ecs.t5-c1m1.4xlarge	16	16.0	25%	240	5760	None	1.2	600	Yes	1	2	6

Instance type	vCPUs	Memory (GiB)	Baseline CPU computing performance	CPU credits per hour	Max CPU credit balance	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.t5-c1m2.4xlarge	16	32.0	25%	240	5760	None	1.2	600	Yes	1	2	6

> **Note**
>
> - When you bind an ENI to or unbind an ENI from an instance of the following instance types, the instance must be in the Stopped state: ecs.t5-lc2m1.nano, ecs.t5-c1m1.large, ecs.t5-c1m2.large, ecs.t5-c1m4.large, ecs.t5-lc1m1.small, ecs.t5-lc1m2.large, ecs.t5-lc1m2.small, and ecs.t5-lc1m4.large.
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

v5, CPU overprovisioned instance family

Features:

- Supports a range of CPU-to-memory ratios, such as 1:1, 1:2, 1:4, and 1:8.
- You can create v5 instances only on dedicated hosts.

> **Note** For information about other types of instances that can be created on dedicated hosts, see Dedicated host types.

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Suitable for the following scenarios:
  - Migration from offline virtualization environments to Alibaba Cloud
  - Services that generate low, medium, or burstable CPU loads

Instance types

Instance type	vCPUs	Memory (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.v5-c1m1.large	2	2.0	2.0	300	Yes	2	2	2
ecs.v5-c1m1.xlarge	4	4.0	2.0	300	Yes	2	2	6
ecs.v5-c1m1.2xlarge	8	8.0	3.0	400	Yes	2	3	6
ecs.v5-c1m1.3xlarge	12	12.0	3.0	400	Yes	4	3	6
ecs.v5-c1m1.4xlarge	16	16.0	4.0	500	Yes	4	4	6
ecs.v5-c1m1.8xlarge	32	32.0	4.0	500	Yes	8	4	6
ecs.v5-c1m2.large	2	4.0	2.0	300	Yes	2	2	2
ecs.v5-c1m2.xlarge	4	8.0	2.0	300	Yes	2	2	6
ecs.v5-c1m2.2xlarge	8	16.0	3.0	400	Yes	2	3	6
ecs.v5-c1m2.3xlarge	12	24.0	3.0	400	Yes	4	3	6
ecs.v5-c1m2.4xlarge	16	32.0	4.0	500	Yes	4	4	6
ecs.v5-c1m2.8xlarge	32	64.0	4.0	500	Yes	8	4	6

Instance type	vCPUs	Memory (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.v5-c1m4.large	2	8.0	2.0	300	Yes	2	2	2
ecs.v5-c1m4.xlarge	4	16.0	2.0	300	Yes	2	2	6
ecs.v5-c1m4.2xlarge	8	32.0	3.0	400	Yes	2	3	6
ecs.v5-c1m4.3xlarge	12	48.0	3.0	400	Yes	4	3	6
ecs.v5-c1m4.4xlarge	16	64.0	4.0	500	Yes	4	4	6
ecs.v5-c1m4.8xlarge	32	128.0	4.0	500	Yes	8	4	6
ecs.v5-c1m8.large	2	16.0	2.0	300	Yes	2	2	2
ecs.v5-c1m8.xlarge	4	32.0	2.0	300	Yes	2	2	6
ecs.v5-c1m8.2xlarge	8	64.0	3.0	400	Yes	2	3	6
ecs.v5-c1m8.3xlarge	12	96.0	3.0	400	Yes	4	3	6
ecs.v5-c1m8.4xlarge	16	128.0	4.0	500	Yes	4	4	6
ecs.v5-c1m8.8xlarge	32	256.0	4.0	500	Yes	8	4	6

Note For more information about these specifications, see Description of instance specifications.

Previous-generation shared instance families xn4, n4, mn4, and e4

Features:

- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors.
- Paired with the DDR4 memory.
- Offer multiple CPU-to-memory ratios.

Instance family	Description	vCPU-to-memory ratio	Scenario
xn4	Shared compact type	1:1	• Frontend web applications • Lightweight applications and microservices • Applications for development or testing environments
n4	Shared compute type	1:2	• Websites and web applications • Development environments, servers, code repositories, microservices, and testing and staging environments • Lightweight enterprise applications
mn4	Shared balanced type	1:4	• Websites and web applications • Lightweight databases and caches • Integrated applications and lightweight enterprise services

Instance family	Description	vCPU-to-memory ratio	Scenario
e4	Shared memory type	1:8	• Applications that require a large amount of memory • Lightweight databases and caches

Instance types of xn4

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.xn4.small	1	1.0	None	0.5	50	No	1	2	2

> **Note**
>
> - When you bind an ENI to or unbind an ENI from an instance of the ecs.xn4.small instance type, the instance must be in the Stopped state.
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

Instance types of n4

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.n4.small	1	2.0	None	0.5	50	No	1	2	2
ecs.n4.large	2	4.0	None	0.5	100	No	1	2	2
ecs.n4.xlarge	4	8.0	None	0.8	150	No	1	2	6
ecs.n4.2xlarge	8	16.0	None	1.2	300	No	1	2	6

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.n4.4xlarge	16	32.0	None	2.5	400	No	1	2	6
ecs.n4.8xlarge	32	64.0	None	5.0	500	No	1	2	6

Note

- When you bind an ENI to or unbind an ENI from an instance of the ecs.n4.small and ecs.n4.large instance types, the instance must be in the Stopped state.
- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

Instance types of mn4

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.mn4.small	1	4.0	None	0.5	50	No	1	2	2
ecs.mn4.large	2	8.0	None	0.5	100	No	1	2	2
ecs.mn4.xlarge	4	16.0	None	0.8	150	No	1	2	6
ecs.mn4.2xlarge	8	32.0	None	1.2	300	No	1	2	6
ecs.mn4.4xlarge	16	64.0	None	2.5	400	No	1	2	6

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.mn4.8xlarge	32	128.0	None	5	500	No	2	8	6

> **Note**
>
> - When you bind an ENI to or unbind an ENI from an instance of the ecs.mn4.small and ecs.mn4.large instance types, the instance must be in the Stopped state.
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

Instance types of e4

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.e4.small	1	8.0	None	0.5	50	No	1	2	2
ecs.e4.large	2	16.0	None	0.5	100	No	1	2	2
ecs.e4.xlarge	4	32.0	None	0.8	150	No	1	2	6
ecs.e4.2xlarge	8	64.0	None	1.2	300	No	1	3	6
ecs.e4.4xlarge	16	128.0	None	2.5	400	No	1	8	6

Note

- When you bind an ENI to or unbind an ENI from an instance of the ecs.e4.small and ecs.e4.large instance types, the instance must be in the Stopped state.
- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

Description of instance specifications

Specification	Descripition
Local storage	Local storage, also called cache disks or local disks, refers to the disks attached to the physical servers where ECS instances are hosted. Local storage provides temporary block storage for instances. Local storage capacity is measured in GiB. Data stored on local disks may be lost when the compute resources (vCPUs and memory) of an instance are released or when an instance is failed over to a normal physical server upon a physical server failure. For more information, see Local disks.
Bandwidth	The maximum bandwidth in one direction. Inbound bandwidth and outbound bandwidth are calculated separately. Note Each instance specification is obtained through verification in a test environment. In actual scenarios, the performance of an instance may vary depending on other factors such as instance load. We recommend that you perform business stress tests on instances to choose appropriate instance types.
Packet forwarding rate	The maximum sum of inbound and outbound packet forwarding rates. For more information about how to test the packet forwarding rate, see Test network performance. Note Each instance specification is obtained through verification in a test environment. In actual scenarios, the performance of an instance may vary depending on other factors such as instance load. We recommend that you perform business stress tests on instances to choose appropriate instance types.
Connections	Connections, also called sessions, are the process of establishing connections and transferring data between a client and a server. A connection is uniquely defined by the network communication quintuple that consists of a source IP address, a destination IP address, a source port, a destination port, and a protocol. Connections of an ECS instance include TCP, UDP, and ICMP connections.

Specification	Descripition
NIC queues	The maximum number of NIC queues that the primary NIC of an instance supports. If your instance type is not a member of an ECS Bare Metal Instance family, the maximum number of NIC queues supported by the secondary NIC is the same as that supported by the primary NIC.

4.Instance type families

4.1. General purpose instance families

This topic describes the features of general purpose instance families of ECS and lists the instance types of each family.

- Recommended instance families
  - g6a, general purpose instance family
  - g6t, security-enhanced general purpose instance family
  - g6e, general purpose instance family with enhanced performance
  - g6, general purpose instance family
  - g5, general purpose instance family
  - g5ne, network enhanced instance family
- Other available instance families

  sn2ne, general purpose instance family with enhanced network performance

g6a, general purpose instance family

g6a is in invitational preview. To use g6a, submit a ticket.

Features

- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the X-Dragon architecture.
- Compute:
  - Equipped with 2.6 GHz AMD EPYC ™ ROME processors with a maximum turbo frequency of 3.3 GHz for consistent computing performance.
  - Offers a CPU-to-memory ratio of 1:4.
  - Allows you to enable or disable Hyper-Threading.

    Note By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.
- Storage:
  - I/O optimized.
  - Supports enhanced SSDs (ESSDs), standard SSDs, and ultra disks.
  - Provides high storage I/O performance based on large computing capacity.

    Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.
- Network:
  - Provides an ultra-high packet forwarding rate.
  - Provides high network performance based on large computing capacity.

- Suitable for the following scenarios:
  - Video encoding and decoding
  - Scenarios where large volumes of packets are received and transmitted
  - Websites and application servers
  - Small and medium-sized database systems, caches, and search clusters
  - Game servers
  - Test and development, such as DevOps
  - Other general purpose enterprise-level applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	ENIs (including one primary ENI)	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g6a.large	2	8.0	None	1.0	10.0	900	Yes	2	12.5	1.0
ecs.g6a.xlarge	4	16.0	None	1.5	10.0	1,000	Yes	3	20.0	1.5
ecs.g6a.2xlarge	8	32.0	None	2.5	10.0	1,600	Yes	4	30.0	2.0
ecs.g6a.4xlarge	16	64.0	None	5.0	10.0	2,000	Yes	8	60.0	3.0
ecs.g6a.8xlarge	32	128.0	None	8.0	10.0	3,000	Yes	7	75.0	4,0
ecs.g6a.16xlarge	64	256.0	None	16.0	None	6,000	Yes	8	150.0	8.0
ecs.g6a.32xlarge	128	512.0	None	32.0	None	12,000	Yes	15	300.0	16.0
ecs.g6a.64xlarge	256	1,024.0	None	64.0	None	24,000	Yes	15	600.0	32.0

> Note
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

g6t, security-enhanced general purpose instance family

g6t is in invitational preview.

Features

- Implements trusted boot based on the Trusted Platform Module (TPM) chip. During a trusted boot, each module in the boot chain from the underlying hardware to the guest OS is measured and verified.
- Supports comprehensive monitoring on the IaaS layer and provides trusted capabilities of the whole IaaS layer.
- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the third-generation X-Dragon architecture. g6t improves storage performance, network performance, and computing stability by an order of magnitude through fast path acceleration of X-Dragon chips.
- Compute:
  - Equipped with 2.5 GHz Intel® Xeon® Platinum 8269 (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
  - Offers a CPU-to-memory ratio of 1:4.
  - Allows you to enable or disable Hyper-Threading.

    > Note By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Storage:
  - I/O optimized.
  - Supports ESSDs only.
  - Provides high storage I/O performance based on large computing capacity.

    > Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Network:
  - Provides an ultra-high packet forwarding rate.
  - Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios that require high security and enhanced trust, such as financial services, government affairs, and enterprise services

- Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
- Enterprise-level applications of various types and sizes
- Websites and application servers
- Game servers
- Small and medium-sized database systems, caches, and search clusters
- Data analysis and computing
- Compute clusters and memory intensive data processing

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	TPM support	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g6t.large	2	8.0	None	A burstable bandwidth of up to 10.0	900	Yes	Yes	Up to 250	2	3	6	20.0	1.0
ecs.g6t.xlarge	4	16.0	None	A burstable bandwidth of up to 10.0	1,000	Yes	Yes	Up to 250	4	4	15	40.0	1.5

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	TPM support	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g6t.2xlarge	8	32.0	None	A burstable bandwidth of up to 10.0	1,600	Yes	Yes	Up to 250	8	4	15	50.0	2.0
ecs.g6t.4xlarge	16	64.0	None	A burstable bandwidth of up to 10.0	3,000	Yes	Yes	300	8	8	30	80.0	3.0
ecs.g6t.8xlarge	32	128.0	None	10.0	6,000	Yes	Yes	600	16	8	30	150.0	5.0
ecs.g6t.13xlarge	52	192.0	None	16.0	9,000	Yes	Yes	900	32	7	30	240.0	8.0
ecs.g6t.26xlarge	104	384.0	None	32.0	24,000	Yes	Yes	1,800	32	15	30	480.0	16.0

Note
- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
- The results for network capabilities are the maximum values obtained from single item tests. For example, when network bandwidth is tested, no stress tests are performed on the packet forwarding rate or other metrics of the network.

g6e, general purpose instance family with enhanced performance

Features

- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the third-generation X-Dragon architecture. g6e improves storage performance, network performance, and computing stability by an order of magnitude through fast path acceleration of X-Dragon chips.
- I/O optimized.
- Supports ESSDs only.
- Provides high network and storage I/O performance based on large computing capacity.

  Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.
- Provides an ultra-high packet forwarding rate.

  Note The maximum network performance varies based on instance families. For higher concurrent connection capabilities, we recommend that you use g5ne. For more information, see g5ne, network enhanced instance family.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269 (Cascade) processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
- Allows you to enable or disable Hyper-Threading.

  Note By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Enterprise-level applications of various types and sizes
  - Websites and application servers
  - Game servers
  - Small and medium-sized database systems, caches, and search clusters
  - Data analysis and computing

- Compute clusters and memory intensive data processing

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g6e.large	2	8.0	None	A burstable bandwidth of up to 10.0	900	Yes	Up to 250	2	3	6	20.0	1.0
ecs.g6e.xlarge	4	16.0	None	A burstable bandwidth of up to 10.0	1,000	Yes	Up to 250	4	4	15	40.0	1.5
ecs.g6e.2xlarge	8	32.0	None	A burstable bandwidth of up to 10.0	1,600	Yes	Up to 250	8	4	15	50.0	2.0

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g6e.4xlarge	16	64.0	None	A burstable bandwidth of up to 10.0	3,000	Yes	300	8	8	30	80.0	3.0
ecs.g6e.8xlarge	32	128.0	None	10.0	6,000	Yes	600	16	8	30	150.0	5,0
ecs.g6e.13xlarge	52	192.0	None	16.0	9,000	Yes	900	32	7	30	240.0	8.0
ecs.g6e.26xlarge	104	384.0	None	32.0	24,000	Yes	1,800	32	15	30	480.0	16.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
- The results for network capabilities are the maximum values obtained from single item tests. For example, when network bandwidth is tested, no stress tests are performed on the packet forwarding rate or other metrics of the network.

g6, general purpose instance family

Features

- Provides predictable and consistent high performance and reduces virtualization overheads

with the use of the X-Dragon architecture.

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.

  > **Note** The maximum performance of disks varies based on instance families. A single g6 instance can deliver up to 200,000 IOPS. For higher storage I/O performance, we recommend that you use g5se. For more information, see g5se, memory optimized instance family with enhanced performance.

- Provides high storage I/O performance based on large computing capacity.

  > **Note** For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Offers a CPU-to-memory ratio of 1:4.
- Allows you to enable or disable Hyper-Threading.

  > **Note** By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Provides an ultra-high packet forwarding rate.

  > **Note** The maximum network performance varies based on instance families. For higher concurrent connection capabilities, we recommend that you use g5ne. For more information, see g5ne, network enhanced instance family.

- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
- Provides high network performance based on large computing capacity.
- Supports instance type changes to c6 or r6.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Enterprise-level applications of various types and sizes
  - Websites and application servers
  - Game servers
  - Small and medium-sized database systems, caches, and search clusters
  - Data analysis and computing
  - Compute clusters and memory intensive data processing

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g6.large	2	8.0	None	1.0	3.0	300	Yes	Up to 250	2	2	6	10.0	1
ecs.g6.xlarge	4	16.0	None	1.5	5.0	500	Yes	Up to 250	4	3	10	20.0	1.5
ecs.g6.2xlarge	8	32.0	None	2.5	8.0	800	Yes	Up to 250	8	4	10	25.0	2
ecs.g6.3xlarge	12	48.0	None	4.0	10.0	900	Yes	Up to 250	8	6	10	30.0	2.5
ecs.g6.4xlarge	16	64.0	None	5.0	10.0	1,000	Yes	300	8	8	20	40.0	3
ecs.g6.6xlarge	24	96.0	None	7.5	10.0	1,500	Yes	450	12	8	20	50.0	4
ecs.g6.8xlarge	32	128.0	None	10.0	None	2,000	Yes	600	16	8	20	60.0	5

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g6.13xlarge	52	192.0	None	12.5	None	3,000	Yes	900	32	7	20	100.0	8
ecs.g6.26xlarge	104	384.0	None	25.0	None	6,000	Yes	1,800	32	15	20	200.0	16

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

g5, general purpose instance family

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.

  Note The maximum performance of disks varies based on instance families. A single g5 instance can deliver up to 200,000 IOPS. For higher storage I/O performance, we recommend that you use g5se. For more information, see g5se, memory optimized instance family with enhanced performance.

- Offers a CPU-to-memory ratio of 1:4.
- Provides an ultra-high packet forwarding rate.

  Note The maximum network performance varies based on instance families. For higher concurrent connection capabilities, we recommend that you use g5ne. For more information, see g5ne, network enhanced instance family.

- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) or 8269CY (Cascade Lake)

processors for consistent computing performance.

- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Enterprise-level applications of various types and sizes
  - Small and medium-sized database systems, caches, and search clusters
  - Data analysis and computing
  - Compute clusters and memory intensive data processing

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.g5.large	2	8.0	None	1.0	300	Yes	2	2	6
ecs.g5.xlarge	4	16.0	None	1.5	500	Yes	2	3	10
ecs.g5.2xlarge	8	32.0	None	2.5	800	Yes	2	4	10
ecs.g5.3xlarge	12	48.0	None	4.0	900	Yes	4	6	10
ecs.g5.4xlarge	16	64.0	None	5.0	1,000	Yes	4	8	20
ecs.g5.6xlarge	24	96.0	None	7.5	1,500	Yes	6	8	20
ecs.g5.8xlarge	32	128.0	None	10.0	2,000	Yes	8	8	20
ecs.g5.16xlarge	64	256.0	None	20.0	4,000	Yes	16	8	20

Note
- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

g5ne, network enhanced instance family

Features

- Instances of the g5ne instance family significantly improve network throughput and packet forwarding rate. A single g5ne instance can deliver up to 10,000 Kpps.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) or 8269CY (Cascade Lake) processors for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:4.
- I/O optimized.
- Supports standard SSDs and ultra disks.
- Provides high network performance based on large computing capacity.

  Note We recommend that you select instance types of the g5ne instance family to deploy Data Plane Development Kit (DPDK) applications.

- Suitable for the following scenarios:
  - DPDK applications
  - Network intensive scenarios such as NFV or SD-WAN, mobile Internet, on-screen video comments, and telecom data forwarding
  - Small and medium-sized database systems, caches, and search clusters
  - Enterprise-level applications of various types and sizes
  - Big data analysis and machine learning

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g5ne.large	2	8.0	None	1.0	400	Yes	450	2	3	10	10.0	1

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.g5ne.xlarge	4	16.0	None	2.0	750	Yes	900	4	4	15	15.0	1
ecs.g5ne.2xlarge	8	32.0	None	3.5	1,500	Yes	1,750	8	6	15	30.0	1
ecs.g5ne.4xlarge	16	64.0	None	7.0	3,000	Yes	3,500	16	8	30	60.0	2
ecs.g5ne.8xlarge	32	128.0	None	15.0	6,000	Yes	7,000	32	8	30	120.0	4
ecs.g5ne.16xlarge	64	256.0	None	30.0	12,000	Yes	14,000	32	8	30	240.0	8
ecs.g5ne.18xlarge	72	288.0	None	33.0	13,500	Yes	16,000	32	15	50	270.0	9

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

sn2ne, general purpose instance family with enhanced network performance

Features

- I/O optimized.
- Supports standard SSDs and ultra disks only.
- Offers a CPU-to-memory ratio of 1:4.
- Provides an ultra-high packet forwarding rate.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) or Platinum 8163 (Skylake) processors for consistent computing performance.
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Enterprise-level applications of various types and sizes
  - Small and medium-sized database systems, caches, and search clusters
  - Data analysis and computing
  - Compute clusters and memory intensive data processing

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.sn2ne.large	2	8.0	None	1.0	300	Yes	2	2	6
ecs.sn2ne.xlarge	4	16.0	None	1.5	500	Yes	2	3	10
ecs.sn2ne.2xlarge	8	32.0	None	2.0	1,000	Yes	4	4	10
ecs.sn2ne.3xlarge	12	48.0	None	2.5	1,300	Yes	4	6	10
ecs.sn2ne.4xlarge	16	64.0	None	3.0	1,600	Yes	4	8	20

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.sn2ne.6xlarge	24	96.0	None	4.5	2,000	Yes	6	8	20
ecs.sn2ne.8xlarge	32	128.0	None	6.0	2,500	Yes	8	8	20
ecs.sn2ne.14xlarge	56	224.0	None	10.0	4,500	Yes	14	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

References

- Instance families
- Create an instance by using the provided wizard

4.2. Compute optimized instance families

This topic describes the features of compute optimized instance families of ECS and lists the instance types of each family.

- Recommended instance families
  - c6a, compute optimized instance family
  - c6t, security-enhanced compute optimized instance family
  - c6e, compute optimized instance family with enhanced performance
  - c6, compute optimized instance family
  - c5, compute optimized instance family
  - ic5, compute intensive instance family
- Other available instance families

  sn1ne, compute optimized instance family with enhanced network performance

c6a, compute optimized instance family

c6a is in invitational preview. To use c6a, submit a ticket.

Features

- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the X-Dragon architecture.
- Compute:
  - Equipped with 2.6 GHz AMD EPYC ™ ROME processors with a maximum turbo frequency of 3.3 GHz for consistent computing performance.
  - Offers a CPU-to-memory ratio of 1:2.
  - Allows you to enable or disable Hyper-Threading.

    Note By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Storage:
  - I/O optimized.
  - Supports enhanced SSDs (ESSDs), standard SSDs, and ultra disks.
  - Provides high storage I/O performance based on large computing capacity.

    Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Network:
  - Provides an ultra-high packet forwarding rate.
  - Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Video encoding and decoding
  - Scenarios where large volumes of packets are received and transmitted
  - Web frontend servers
  - Frontend servers of massively multiplayer online (MMO) games
  - Test and development, such as DevOps

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	ENIs (including one primary ENI)	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.c6a.large	2	4.0	None	1.0	10.0	900	Yes	2	12.5	1.0

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	ENIs (including one primary ENI)	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.c6a.xlarge	4	8.0	None	1.5	10.0	1,000	Yes	3	20.0	1.5
ecs.c6a.2xlarge	8	16.0	None	2.5	10.0	1,600	Yes	4	30.0	2.0
ecs.c6a.4xlarge	16	32.0	None	5.0	10.0	2,000	Yes	8	60.0	3.0
ecs.c6a.8xlarge	32	64.0	None	8.0	10.0	3,000	Yes	7	75.0	4,0
ecs.c6a.16xlarge	64	128.0	None	16.0	None	6,000	Yes	8	150.0	8.0
ecs.c6a.32xlarge	128	256.0	None	32.0	None	12,000	Yes	15	300.0	16.0
ecs.c6a.64xlarge	256	512.0	None	64.0	None	24,000	Yes	15	600.0	32.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

c6t, security-enhanced compute optimized instance family

c6t is in invitational preview.

Features

- Implements trusted boot based on the Trusted Platform Module (TPM) chip. During a trusted boot, each module in the boot chain from the underlying hardware to the guest OS is

measured and verified.

- Supports comprehensive monitoring on the IaaS layer and ensures trusted capabilities of the whole IaaS layer.
- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the third-generation X-Dragon architecture. c6t improves storage performance, network performance, and computing stability by an order of magnitude through fast path acceleration of X-Dragon chips.
- Compute:
  - Equipped with 2.5 GHz Intel® Xeon® Platinum 8269 (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
  - Offers a CPU-to-memory ratio of 1:2.
  - Allows you to enable or disable Hyper-Threading.

    Note By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Storage:
  - I/O optimized.
  - Supports ESSDs only.
  - Provides high storage I/O performance based on large computing capacity.

    Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Network:
  - Provides an ultra-high packet forwarding rate.
  - Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios that require high security and enhanced trust, such as financial services, government affairs, and enterprise services
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Web frontend servers
  - Frontend servers of MMO games
  - Data analysis, batch processing, and video encoding
  - High-performance scientific and engineering applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	TPM support	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.c6t.large	2	4.0	None	A burstable bandwidth of up to 10.0	900	Yes	Yes	Up to 250	2	3	6	20.0	1.0
ecs.c6t.xlarge	4	8.0	None	A burstable bandwidth of up to 10.0	1,000	Yes	Yes	Up to 250	4	4	15	40.0	1.5
ecs.c6t.2xlarge	8	16.0	None	A burstable bandwidth of up to 10.0	1,600	Yes	Yes	Up to 250	8	4	15	50.0	2.0

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	TPM support	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.c6t.4xlarge	16	32.0	None	A burstable bandwidth of up to 10.0	3,000	Yes	Yes	300	8	8	30	80.0	3.0
ecs.c6t.8xlarge	32	64.0	None	10.0	6,000	Yes	Yes	600	16	8	30	150.0	5.0
ecs.c6t.13xlarge	52	96.0	None	16.0	9,000	Yes	Yes	900	32	7	30	240.0	8.0
ecs.c6t.26xlarge	104	192.0	None	32.0	24,000	Yes	Yes	1,800	32	15	30	480.0	16.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
- The results for network capabilities are the maximum values obtained from single item tests. For example, when network bandwidth is tested, no stress tests are performed on the packet forwarding rate or other metrics of the network.

c6e, compute optimized instance family with enhanced performance

Features

- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the third-generation X-Dragon architecture. c6e improves storage performance, network performance, and computing stability by an order of magnitude through fast path acceleration of X-Dragon chips.
- I/O optimized.
- Supports ESSDs only.
- Provides high network and storage I/O performance based on large computing capacity.

  Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Provides an ultra-high packet forwarding rate.

  Note The maximum network performance varies based on instance families. For higher concurrent connection capabilities, we recommend that you use g5ne. For more information, see g5ne, network enhanced instance family.

- Offers a CPU-to-memory ratio of 1:2.
- Equipped with 2.5 GHz Intel ® Xeon ® Platinum 8269 (Cascade) processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
- Allows you to enable or disable Hyper-Threading.

  Note By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Web frontend servers
  - Frontend servers of MMO games
  - Data analysis, batch processing, and video encoding
  - High-performance scientific and engineering applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.c6e.large	2	4.0	None	A burstable bandwidth of up to 10.0	900	Yes	Up to 250	2	3	6	20.0	1.0
ecs.c6e.xlarge	4	8.0	None	A burstable bandwidth of up to 10.0	1,000	Yes	Up to 250	4	4	15	40.0	1.5
ecs.c6e.2xlarge	8	16.0	None	A burstable bandwidth of up to 10.0	1,600	Yes	Up to 250	8	4	15	50.0	2.0

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.c6e.4xlarge	16	32.0	None	A burstable bandwidth of up to 10.0	3,000	Yes	300	8	8	30	80.0	3.0
ecs.c6e.8xlarge	32	64.0	None	10.0	6,000	Yes	600	16	8	30	150.0	5.0
ecs.c6e.13xlarge	52	96.0	None	16.0	9,000	Yes	900	32	7	30	240.0	8.0
ecs.c6e.26xlarge	104	192.0	None	32.0	24,000	Yes	1,800	32	15	30	480.0	16.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
- The results for network capabilities are the maximum values obtained from single item tests. For example, when network bandwidth is tested, no stress tests are performed on the packet forwarding rate or other metrics of the network.

c6, compute optimized instance family

Features

- Provides predictable and consistent high performance and reduces virtualization overheads

with the use of the X-Dragon architecture.

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.

  > **Note** The maximum performance of disks varies based on instance families. A single c6 instance can deliver up to 200,000 IOPS. For higher storage I/O performance, we recommend that you use g5se. For more information, see g5se, memory optimized instance family with enhanced performance.

- Provides high storage I/O performance based on large computing capacity.

  > **Note** For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Allows you to enable or disable Hyper-Threading.

  > **Note** By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Offers a CPU-to-memory ratio of 1:2.
- Provides an ultra-high packet forwarding rate.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
- Provides high network performance based on large computing capacity.
- Supports instance type changes to g6 or r6.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Web frontend servers
  - Frontend servers of MMO games
  - Data analysis, batch processing, and video encoding
  - High-performance scientific and engineering applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.c6.large	2	4.0	None	1.0	3.0	300	Yes	Up to 250	2	2	6	10.0	1
ecs.c6.xlarge	4	8.0	None	1.5	5.0	500	Yes	Up to 250	4	3	10	20.0	1.5
ecs.c6.2xlarge	8	16.0	None	2.5	8.0	800	Yes	Up to 250	8	4	10	25.0	2
ecs.c6.3xlarge	12	24.0	None	4.0	10.0	900	Yes	Up to 250	8	6	10	30.0	2.5
ecs.c6.4xlarge	16	32.0	None	5.0	10.0	1,000	Yes	300	8	8	20	40.0	3
ecs.c6.6xlarge	24	48.0	None	7.5	10.0	1,500	Yes	450	12	8	20	50.0	4
ecs.c6.8xlarge	32	64.0	None	10.0	None	2,000	Yes	600	16	8	20	60.0	5

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.c6.13xlarge	52	96.0	None	12.5	None	3,000	Yes	900	32	7	20	100.0	8
ecs.c6.26xlarge	104	192.0	None	25.0	None	6,000	Yes	1,800	32	15	20	200.0	16

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

c5, compute optimized instance family

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.

  Note The maximum performance of disks varies based on instance families. A single c5 instance can deliver up to 200,000 IOPS. For higher storage I/O performance, we recommend that you use g5se. For more information, see g5se, memory optimized instance family with enhanced performance.

- Offers a CPU-to-memory ratio of 1:2.
- Provides an ultra-high packet forwarding rate.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) or 8269CY (Cascade Lake) processors for consistent computing performance.
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:

- Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
- Web frontend servers
- Frontend servers of MMO games
- Data analysis, batch processing, and video encoding
- High-performance scientific and engineering applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.c5.large	2	4.0	None	1.0	300	Yes	2	2	6
ecs.c5.xlarge	4	8.0	None	1.5	500	Yes	2	3	10
ecs.c5.2xlarge	8	16.0	None	2.5	800	Yes	2	4	10
ecs.c5.3xlarge	12	24.0	None	4.0	900	Yes	4	6	10
ecs.c5.4xlarge	16	32.0	None	5.0	1,000	Yes	4	8	20
ecs.c5.6xlarge	24	48.0	None	7.5	1,500	Yes	6	8	20
ecs.c5.8xlarge	32	64.0	None	10.0	2,000	Yes	8	8	20
ecs.c5.16xlarge	64	128.0	None	20.0	4,000	Yes	16	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ic5, compute intensive instance family

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:1.
- Provides an ultra-high packet forwarding rate.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) or 8269CY (Cascade Lake) processors for consistent computing performance.
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Web frontend servers
  - Data analysis, batch processing, and video encoding
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Frontend servers of MMO games

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ic5.large	2	2.0	None	1.0	300	No	2	2	6
ecs.ic5.xlarge	4	4.0	None	1.5	500	No	2	3	10
ecs.ic5.2xlarge	8	8.0	None	2.5	800	No	2	4	10
ecs.ic5.3xlarge	12	12.0	None	4.0	900	No	4	6	10
ecs.ic5.4xlarge	16	16.0	None	5.0	1,000	No	4	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

sn1ne, compute optimized instance family with enhanced network performance

Features

- I/O optimized.
- Supports standard SSDs and ultra disks only.
- Offers a CPU-to-memory ratio of 1:2.
- Provides an ultra-high packet forwarding rate.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) or Platinum 8163 (Skylake) processors for consistent computing performance.
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Web frontend servers
  - Frontend servers of MMO games
  - Data analysis, batch processing, and video encoding
  - High-performance scientific and engineering applications

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.sn1ne.large	2	4.0	None	1.0	300	Yes	2	2	6
ecs.sn1ne.xlarge	4	8.0	None	1.5	500	Yes	2	3	10
ecs.sn1ne.2xlarge	8	16.0	None	2.0	1,000	Yes	4	4	10

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.sn1ne.3xlarge	12	24.0	None	2.5	1,300	Yes	4	6	10
ecs.sn1ne.4xlarge	16	32.0	None	3.0	1,600	Yes	4	8	20
ecs.sn1ne.6xlarge	24	48.0	None	4.5	2,000	Yes	6	8	20
ecs.sn1ne.8xlarge	32	64.0	None	6.0	2,500	Yes	8	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

References

- Instance families
- Create an instance by using the provided wizard

4.3. Memory optimized instance families

This topic describes the features of memory optimized instance families and lists the instance types of each family.

- Recommended instance families
  - r6a, memory optimized instance family
  - r6e, memory optimized instance family with enhanced performance
  - r6, memory optimized instance family
  - re6, high memory instance family
  - r5, memory optimized instance family
- Other available instance families
  - re4, high memory instance family
  - re4e, high memory instance family

- se1ne, memory optimized instance family with enhanced network performance
- se1, memory optimized instance family

r6a, memory optimized instance family

r6a is in invitational preview. To use r6a, submit a ticket.

Features

- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the X-Dragon architecture.
- Compute:
  - Equipped with 2.6 GHz AMD EPYC ™ ROME processors with a maximum turbo frequency of 3.3 GHz for consistent computing performance.
  - Offers a CPU-to-memory ratio of 1:8.
  - Allows you to enable or disable Hyper-Threading.

    Note By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Storage:
  - I/O optimized.
  - Supports enhanced SSDs (ESSDs), standard SSDs, and ultra disks.
  - Provides high storage I/O performance based on large computing capacity.

    Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Network:
  - Provides an ultra high packet forwarding rate.
  - Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Video encoding and decoding
  - Scenarios where large volumes of packets are received and transmitted
  - In-memory databases
  - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications
  - Test and development, such as DevOps

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	ENIs (including one primary ENI)	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.r6a.large	2	16.0	None	1.0	10.0	900	Yes	2	12.5	1.0
ecs.r6a.xlarge	4	32.0	None	1.5	10.0	1,000	Yes	3	20	1.5
ecs.r6a.2xlarge	8	64.0	None	2.5	10.0	1,600	Yes	4	30	2.0
ecs.r6a.4xlarge	16	128.0	None	5.0	10.0	2,000	Yes	8	60	3.0
ecs.r6a.8xlarge	32	256.0	None	8.0	10.0	3,000	Yes	7	75	4,0
ecs.r6a.16xlarge	64	512.0	None	16.0	None	6,000	Yes	8	150	8.0
ecs.r6a.32xlarge	128	1024.0	None	32.0	None	12,000	Yes	15	300	16.0
ecs.r6a.64xlarge	256	2048.0	None	64.0	None	24,000	Yes	15	600	32.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

r6e, memory optimized instance family with enhanced performance

Features

- Provides predictable and consistent high performance and reduces virtualization overheads with the use of the third-generation X-Dragon architecture. In addition, improves storage

performance, network performance, and computing stability by an order of magnitude through fast path acceleration of X-Dragon chips.

- I/O optimized.
- Supports ESSDs.
- Provides high network and storage I/O performance based on large computing capacity.

  > **Note** For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Provides an ultra high packet forwarding rate.

  > **Note** The maximum network performance varies depending on instance families. For more concurrent connections, we recommend that you use g5ne. For more information, see the "g5ne, general purpose instance family with enhanced network performance" section in Instance families.

- Offers a CPU-to-memory ratio of 1:8.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269 processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
- Allows you to enable or disable Hyper-Threading.

  > **Note** By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - High-performance databases and in-memory databases
  - Data analysis, data mining, and distributed memory caching
  - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.r6e.large	2	16.0	None	A burstable bandwidth of up to 10.0	900	Yes	Up to 250	2	3	6	20	1.0
ecs.r6e.xlarge	4	32.0	None	A burstable bandwidth of up to 10.0	1,000	Yes	Up to 250	4	4	15	40	1.5
ecs.r6e.2xlarge	8	64.0	None	A burstable bandwidth of up to 10.0	1,600	Yes	Up to 250	8	4	15	50	2.0

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.r6e.4xlarge	16	128.0	None	A burstable bandwidth of up to 10.0	3,000	Yes	300	8	8	30	80	3.0
ecs.r6e.8xlarge	32	256.0	None	10.0	6,000	Yes	600	16	8	30	150	5.0
ecs.r6e.13xlarge	52	384.0	None	16.0	9,000	Yes	900	32	7	30	240	8.0
ecs.r6e.26xlarge	104	768.0	None	32.0	24,000	Yes	1,800	32	15	30	480	16.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
- The results for network capabilities are the maximum values obtained from single item tests. For example, when network bandwidth is tested, no stress tests are performed on the packet forwarding rate or other metrics of the network.

r6, memory optimized instance family

Features

- Provides predictable and consistent high performance and reduces virtualization overheads

with the use of the X-Dragon architecture.

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.

  > **Note** The maximum performance of disks varies with instance families. A single r6 instance can deliver up to 200,000 IOPS. For higher storage I/O performance, we recommend that you use g5se. For more information, see the "g5se, storage optimized instance family with enhanced performance" section in Instance families.

- Provides high storage I/O performance based on large computing capacity.

  > **Note** For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Provides an ultra-high packet forwarding rate.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:8.
- Allows you to enable or disable Hyper-Threading.

  > **Note** By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Provides high network performance based on large computing capacity.
- Supports changes to g6 or c6 instance families.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - High-performance databases and in-memory databases
  - Data analysis, data mining, and distributed memory caching
  - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.r6.large	2	16.0	None	1.0	3.0	300	Yes	Up to 250	2	2	6	10	1
ecs.r6.xlarge	4	32.0	None	1.5	5.0	500	Yes	Up to 250	4	3	10	20	1.5
ecs.r6.2xlarge	8	64.0	None	2.5	8.0	800	Yes	Up to 250	8	4	10	25	2
ecs.r6.3xlarge	12	96.0	None	4.0	10.0	900	Yes	Up to 250	8	6	10	30	2.5
ecs.r6.4xlarge	16	128.0	None	5.0	10.0	1,000	Yes	300	8	8	20	40	3
ecs.r6.6xlarge	24	192.0	None	7.5	10.0	1,500	Yes	450	12	8	20	50	4
ecs.r6.8xlarge	32	256.0	None	10.0	None	2,000	Yes	600	16	8	20	60	5

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.r6.13xlarge	52	384.0	None	12.5	None	3,000	Yes	900	32	7	20	100	8
ecs.r6.26xlarge	104	768.0	None	25.0	None	6,000	Yes	1,800	32	15	20	200	16

> **Note**
>
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

re6, high memory instance family

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Optimized for high-performance databases, in-memory databases, and other memory intensive enterprise applications.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:15 and up to 3 TiB memory.
- Suitable for the following scenarios:
  - High-performance databases and in-memory databases such as SAP HANA
  - Memory intensive applications
  - Big data processing engines such as Apache Spark and Presto

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.re6.13xlarge	52	768.0	None	10.0	1,800	Yes	16	7	20	50	4
ecs.re6.26xlarge	104	1536.0	None	16.0	3,000	Yes	32	7	20	100	8
ecs.re6.52xlarge	208	3072.0	None	32.0	6,000	Yes	32	15	20	200	16

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

r5, memory optimized instance family

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.

  Note The maximum performance of disks varies with instance families. A single r5 instance can deliver up to 200,000 IOPS. For higher storage I/O performance, we recommend that you use g5se. For more information, see the "g5se, storage optimized instance family with enhanced performance" section in Instance families.

- Provides an ultra-high packet forwarding rate.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) or Intel® Xeon® Paltinum 8269CY (Cascade Lake) processors for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:8.
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:

- Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
- High-performance databases and in-memory databases
- Data analysis, data mining, and distributed memory caching
- Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.r5.large	2	16.0	None	1.0	300	Yes	2	2	6
ecs.r5.xlarge	4	32.0	None	1.5	500	Yes	2	3	10
ecs.r5.2xlarge	8	64.0	None	2.5	800	Yes	2	4	10
ecs.r5.3xlarge	12	96.0	None	4.0	900	Yes	4	6	10
ecs.r5.4xlarge	16	128.0	None	5.0	1,000	Yes	4	8	20
ecs.r5.6xlarge	24	192.0	None	7.5	1,500	Yes	6	8	20
ecs.r5.8xlarge	32	256.0	None	10.0	2,000	Yes	8	8	20
ecs.r5.16xlarge	64	512.0	None	20.0	4,000	Yes	16	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

re4, high memory instance family

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Optimized for high-performance databases, in-memory databases, and other memory intensive enterprise applications.
- Equipped with 2.2 GHz Intel® Xeon® E7 8880 v4(Broadwell) processors with a maximum turbo frequency of 2.4 GHz for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:12 and up to 1,920 GiB memory.
- The ecs.re4.20xlarge and ecs.re4.40xlarge instance types are SAP HANA-certified.
- Suitable for the following scenarios:
  - High-performance databases and in-memory databases such as SAP HANA
  - Memory intensive applications
  - Big data processing engines such as Apache Spark and Presto

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.re4.20xlarge	80	960.0	None	15.0	2,000	Yes	16	8	20
ecs.re4.40xlarge	160	1920.0	None	30.0	4,500	Yes	16	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

re4e, high memory instance family

Features

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

- Equipped with 2.2 GHz Intel® Xeon® E7 8880 v4(Broadwell) processors with a maximum turbo frequency of 2.4 GHz for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:24 and up to 3,840 GiB memory.
- Suitable for the following scenarios:
  - High-performance databases and in-memory databases such as SAP HANA
  - Memory intensive applications
  - Big data processing engines such as Apache Spark and Presto

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.re4e.40xlarge	160	3840.0	None	30.0	4,500	Yes	16	15	20

> **Note**
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

se1ne, memory optimized instance family with enhanced network performance

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Offers a CPU-to-memory ratio of 1:8.
- Provides an ultra-high packet forwarding rate.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) or Intel® Xeon® Platinum 8163 (Skylake) processors for consistent computing performance.
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - High-performance databases and in-memory databases
  - Data analysis, data mining, and distributed memory caching
  - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.se1ne.large	2	16.0	None	1.0	300	Yes	2	2	6
ecs.se1ne.xlarge	4	32.0	None	1.5	500	Yes	2	3	10
ecs.se1ne.2xlarge	8	64.0	None	2.0	1,000	Yes	4	4	10
ecs.se1ne.3xlarge	12	96.0	None	2.5	1,300	Yes	4	6	10
ecs.se1ne.4xlarge	16	128.0	None	3.0	1,600	Yes	4	8	20
ecs.se1ne.6xlarge	24	192.0	None	4.5	2,000	Yes	6	8	20
ecs.se1ne.8xlarge	32	256.0	None	6.0	2,500	Yes	8	8	20
ecs.se1ne.14xlarge	56	480.0	None	10.0	4,500	Yes	14	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

se1, memory optimized instance family

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Offers a CPU-to-memory ratio of 1:8.

- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors for consistent computing performance.
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - High-performance databases and in-memory databases
  - Data analysis, data mining, and distributed memory caching
  - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.se1.large	2	16.0	None	0.5	100	No	1	2	6
ecs.se1.xlarge	4	32.0	None	0.8	200	No	1	3	10
ecs.se1.2xlarge	8	64.0	None	1.5	400	No	1	4	10
ecs.se1.4xlarge	16	128.0	None	3.0	500	No	2	8	20
ecs.se1.8xlarge	32	256.0	None	6.0	800	No	3	8	20
ecs.se1.14xlarge	56	480.0	None	10.0	1,200	No	4	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

References

- Instance families
- Create an instance by using the provided wizard

4.4. Instance families with local SSDs

This topic describes the features of ECS instance families with local SSDs and lists the instance types of each family.

Alibaba Cloud ecs server elastic computing

- Recommended instance families
  - i2, instance family with local SSDs
  - i2g, instance family with local SSDs
  - i2ne, instance family with local SSDs
  - i2gne, instance family with local SSDs
- Other available instance families

  i1, instance family with local SSDs

Overview

Instances with local SSDs provide high I/O performance. They are suitable for scenarios that impose high demands on storage I/O performance and high availability architecture at the application level. For example, they are suitable for NoSQL databases, massively parallel processing (MPP) data warehouses, and distributed file systems.

Instances with local SSDs are suitable for enterprises that provide online services such as online gaming, e-commerce, live video streaming, and media. Instances with local SSDs can satisfy the high requirements of I/O-intensive applications that require low latency and high I/O performance of block storage.

Instances with local SSDs have the following features:

- Provide up to hundreds of thousands of low latency random read/write IOPS for large databases.
- Offer a maximum throughput of several gibibytes per second for sequential read/write operations in big data, parallel computing, and other large dataset scenarios.
- Use local NVMe SSDs to deliver hundreds of thousands of random read/write IOPS with only several microseconds of latency.

When you use instances with local SSDs, take note of the following items:

- Instances with local SSDs do not support changes in instance types, bandwidth, and billing methods, and do not support failover.
- The associations of instances to local disks are fixed. The number and capacity of local disks of an instance are based on the instance type. Instances with local SSDs are bound to their local disks. You cannot attach additional local disks to these instances or detach local disks from these instances and attach the disks to another instance.
- You cannot create snapshots for local disks. If you need to create an image for the system disk and data disks of an instance with local SSDs, we recommend that you create an image by using the snapshots of both the system disk and data disks (data disks must be non-local disks).
- You cannot create images that contain system disks and data disks based on instance IDs.
- You can attach a standard SSD to an instance with local SSDs. The capacity of the standard SSD is scalable.

- Local disks are attached to a single physical server, which increases the risk of single point of failure (SPOF). The reliability of data stored on local disks depends on the reliability of the physical server.

  Warning For example, data stored on local disks may be lost when a hardware failure occurs. We recommend that you do not use local disks for long-term data storage.

  - To ensure data availability, we recommend that you implement data redundancy at the application layer. You can use deployment sets to distribute ECS instances across multiple physical machines to achieve high availability and disaster recovery. For more information, see Create a deployment set.
  - If your applications do not have data reliability architecture, we recommend that you use cloud disks or the backup service in your ECS instances for data reliability. For more information, see Disk overview.

- Operations on an instance with local SSDs may affect the data stored on the local disks. For more information, see Impacts of instance operations on the data stored on local disks.

i2, instance family with local SSDs

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Attached with high-performance local NVMe SSDs that feature high IOPS, high I/O throughput, and low latency.
- Offers a CPU-to-memory ratio of 1:8, which is designed for high-performance databases.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Provides a fast and reliable network based on large computing capacity.
- Scenarios:
  - Online transaction processing (OLTP) and high-performance relational databases
  - NoSQL databases such as Cassandra, MongoDB, and HBase
  - Search scenarios that use solutions such as Elasticsearch

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk bandwidth (Gbit/s)
ecs.i2.xlarge	4	32.0	1 × 894	1.0	500	Yes	2	3	10	Up to 16
ecs.i2.2xlarge	8	64.0	1 × 1,788	2.0	1,000	Yes	2	4	10	Up to 16

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk bandwidth (Gbit/s)
ecs.i2.4xlarge	16	128.0	2 × 1,788	3.0	1,500	Yes	4	8	20	Up to 16
ecs.i2.8xlarge	32	256.0	4 × 1,788	6.0	2,000	Yes	8	8	20	Up to 16
ecs.i2.16xlarge	64	512.0	8 × 1,788	10.0	4,000	Yes	16	8	20	Up to 16
ecs.i2d.21xlarge	84	712.0	4 × 3,570	25.0	4,000	Yes	32	16	20	Up to 16

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
- For more information about the performance metrics of SSDs, see Local disks.

i2g, instance family with local SSDs

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Attached with high-performance local NVMe SSDs that feature high IOPS, high I/O throughput, and low latency.
- Offers a CPU-to-memory ratio of 1:4, which is designed for high-performance databases.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Provides a fast and reliable network based on large computing capacity.
- Scenarios:
  - OLTP and high-performance relational databases
  - NoSQL databases such as Cassandra, MongoDB, and HBase
  - Search scenarios that use solutions such as Elasticsearch

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.i2g.2xlarge	8	32.0	1 × 894	2.0	1,000	No	2	4	10
ecs.i2g.4xlarge	16	64.0	1 × 1,788	3.0	1,500	No	4	8	20
ecs.i2g.8xlarge	32	128.0	2 × 1,788	6.0	2,000	No	8	8	20
ecs.i2g.16xlarge	64	256.0	4 × 1,788	10.0	4,000	No	16	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
- For more information about the performance metrics of SSDs, see Local disks.

i2ne, instance family with local SSDs

i2ne is in invitational preview. To use i2ne, submit a ticket.

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Attached with high-performance local NVMe SSDs that feature high IOPS, high I/O throughput, and low latency.
- Provides a bandwidth of up to 20 Gbit/s.
- Offers a CPU-to-memory ratio of 1:8, which is designed for high-performance databases.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Provides a fast and reliable network based on large computing capacity.
- Scenarios:
  - OLTP and high-performance relational databases
  - NoSQL databases such as Cassandra, MongoDB, and HBase
  - Search scenarios that use solutions such as Elasticsearch

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk bandwidth (Gbit/s)
ecs.i2ne.xlarge	4	32.0	1 × 894	1.5	500	Yes	2	3	10	Up to 16
ecs.i2ne.2xlarge	8	64.0	1 × 1,788	2.5	1,000	Yes	2	4	10	Up to 16
ecs.i2ne.4xlarge	16	128.0	2 × 1,788	5.0	1,500	Yes	4	8	20	Up to 16
ecs.i2ne.8xlarge	32	256.0	4 × 1,788	10.0	2,000	Yes	8	8	20	Up to 16
ecs.i2ne.16xlarge	64	512.0	8 × 1,788	20.0	4,000	Yes	16	8	20	Up to 16

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
- For more information about the performance metrics of SSDs, see Local disks.

i2gne, instance family with local SSDs

i2gne is in invitational preview. To use i2gne, submit a ticket.

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Attached with high-performance local NVMe SSDs that feature high IOPS, high I/O throughput, and low latency.
- Provides a bandwidth of up to 20 Gbit/s.
- Offers a CPU-to-memory ratio of 1:4, which is designed for high-performance databases.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.

- Provides a fast and reliable network based on large computing capacity.
- Scenarios:
  - OLTP and high-performance relational databases
  - NoSQL databases such as Cassandra, MongoDB, and HBase
  - Search scenarios that use solutions such as Elasticsearch

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.i2gne.2xlarge	8	32.0	1 × 894	2.5	1,000	No	2	4	10
ecs.i2gne.4xlarge	16	64.0	1 × 1,788	5.0	1,500	No	4	8	20
ecs.i2gne.8xlarge	32	128.0	2 × 1,788	10.0	2,000	No	8	8	20
ecs.i2gne.16xlarge	64	256.0	4 × 1,788	20.0	4,000	No	16	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
- For more information about the performance metrics of SSDs, see Local disks.

i1, instance family with local SSDs

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Attached with high-performance local NVMe SSDs that feature high IOPS, high I/O throughput, and low latency.
- Offers a CPU-to-memory ratio of 1:4, which is designed for high-performance databases.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors.
- Provides a fast and reliable network based on large computing capacity.

- Scenarios:
  - OLTP and high-performance relational databases
  - NoSQL databases such as Cassandra and MongoDB
  - Search scenarios that use solutions such as Elasticsearch

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.i1.xlarge	4	16.0	2 × 104	0.8	200	No	1	3	10
ecs.i1.2xlarge	8	32.0	2 × 208	1.5	400	No	1	4	10
ecs.i1.3xlarge	12	48.0	2 × 312	2.0	400	No	1	6	10
ecs.i1.4xlarge	16	64.0	2 × 416	3.0	500	No	2	8	20
ecs.i1-c5d1.4xlarge	16	64.0	2 × 1,456	3.0	400	No	2	8	20
ecs.i1.6xlarge	24	96.0	2 × 624	4.5	600	No	2	8	20
ecs.i1.8xlarge	32	128.0	2 × 832	6.0	800	No	3	8	20
ecs.i1-c10d1.8xlarge	32	128.0	2 × 1,456	6.0	800	No	3	8	20
ecs.i1.14xlarge	56	224.0	2 × 1,456	10.0	1,200	No	4	8	20

Note
- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.
- For more information about the performance metrics of SSDs, see Local disks.

4.5. Instance families with high clock speeds

This topic describes the features of instance families with high clock speeds and lists the instance types of each instance family.

- Recommended instance families
  - hfc7, compute optimized instance family with high clock speed
  - hfc6, compute optimized instance family with high clock speed
  - hfg7, general purpose instance family with high clock speed
  - hfg6, general purpose instance family with high clock speed
  - hfr7, memory optimized instance family with high clock speed
  - hfr6, memory optimized instance family with high clock speed
- Other available instance families
  - hfc5, compute optimized instance family with high clock speed
  - hfg5, general purpose instance family with high clock speed

hfc7, compute optimized instance family with high clock speed

Features

- Uses the third-generation X-Dragon architecture to provide predictable and consistent high performance and reduce virtualization overheads.
- I/O optimized.
- Supports enhanced SSDs (ESSDs) and provides ultra-high I/O performance.
- Provides high storage I/O performance based on large computing capacity.

  Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Offers a CPU-to-memory ratio of 1:2.
- Provides an ultra-high packet forwarding rate.
- Equipped with Intel® Xeon® Platinum 8369HB (Cooper Lake) or Intel® Xeon® Platinum 8369HC (Cooper Lake) processors that deliver a maximum turbo frequency of 3.8 GHz and a minimum clock speed of 3.3 GHz for consistent computing performance.
- Allows you to enable or disable hyper-threading.

> **Note** By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are transmitted and received
  - High-performance frontend server clusters
  - Frontend servers of massively multiplayer online (MMO) games
  - Data analysis, batch processing, and video encoding
  - High-performance science and engineering applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfc7.large	2	4.0	None	1.2	10.0	900	Yes	2	2	6	20	1.0
ecs.hfc7.xlarge	4	8.0	None	2.0	10.0	1,000	Yes	4	3	15	30	1.5
ecs.hfc7.2xlarge	8	16.0	None	3.0	10.0	1,600	Yes	8	4	15	45	2.0
ecs.hfc7.3xlarge	12	24.0	None	4.5	10.0	2,000	Yes	8	6	15	60	2.5
ecs.hfc7.4xlarge	16	32.0	None	6.0	10.0	2,500	Yes	8	8	30	75	3.0

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfc7.6xlarge	24	48.0	None	8.0	10.0	3,000	Yes	12	8	30	90	4.0
ecs.hfc7.8xlarge	32	64.0	None	10.0	None	4,000	Yes	16	8	30	105	5.0
ecs.hfc7.12xlarge	48	96.0	None	16.0	None	6,000	Yes	24	8	30	150	8.0
ecs.hfc7.24xlarge	96	192.0	None	32.0	None	12,000	Yes	32	15	30	300	16.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

hfc6, compute optimized instance family with high clock speed

Features

- Uses the X-Dragon architecture to provide predictable and consistent high performance and reduce virtualization overheads.
- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Provides high storage I/O performance based on large computing capacity.

> **Note** For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Offers a CPU-to-memory ratio of 1:2.
- Provides an ultra-high packet forwarding rate.
- Equipped with 3.1 GHz Intel® Xeon® Platinum 8269 (Cascade Lake) processors that deliver a maximum turbo frequency of 3.5 GHz for consistent computing performance.

  > **Note** The CPU of this instance family provides a 3.1 GHz clock speed. The Intel System Studio (ISS) feature may cause a lower clock speed to be displayed. Alibaba Cloud is working on to resolve this issue. This issue does not affect the actual clock speed of your instances.
  >
  > You can run the following commands separately and use the turbostat tool to view the actual clock speed.

  ```
  yum install kernel-tools
  ```

  ```
  turbostat
  ```

- Allows you to enable or disable hyper-threading.

  > **Note** By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are transmitted and received
  - Web frontend servers
  - Frontend servers of MMO games
  - Data analysis, batch processing, and video encoding
  - High-performance scientific and engineering applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfc6.large	2	4.0	None	1.0	3.0	300	Yes	2	2	6	10	1.0
ecs.hfc6.xlarge	4	8.0	None	1.5	5.0	500	Yes	4	3	10	20	1.5
ecs.hfc6.2xlarge	8	16.0	None	2.5	8.0	800	Yes	8	4	10	25	2.0
ecs.hfc6.3xlarge	12	24.0	None	4.0	10.0	900	Yes	8	6	10	30	2.5
ecs.hfc6.4xlarge	16	32.0	None	5.0	10.0	1,000	Yes	8	8	20	40	3.0
ecs.hfc6.6xlarge	24	48.0	None	7.5	10.0	1,500	Yes	12	8	20	50	4.0
ecs.hfc6.8xlarge	32	64.0	None	10.0	None	2,000	Yes	16	8	20	60	5.0
ecs.hfc6.10xlarge	40	96.0	None	12.5	None	3,000	Yes	32	7	20	100	8.0

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfc6.16xlarge	64	128.0	None	20.0	None	4,000	Yes	32	8	20	120	10.0
ecs.hfc6.20xlarge	80	192.0	None	25.0	None	6,000	Yes	32	15	20	200	16.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

hfg7, general purpose instance family with high clock speed

Features

- Uses the third-generation X-Dragon architecture to provide predictable and consistent high performance and reduce virtualization overheads.
- I/O optimized.
- Supports ESSDs and provides ultra-high I/O performance.
- Provides high storage I/O performance based on large computing capacity.

  Note For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Offers a CPU-to-memory ratio of 1:4.
- Provides an ultra-high packet forwarding rate.
- Equipped with Intel® Xeon® Platinum 8369HB (Cooper Lake) or Intel® Xeon® Platinum 8369HC (Cooper Lake) processors that deliver a maximum turbo frequency of 3.8 GHz and a minimum clock speed of 3.3 GHz for consistent computing performance.
- Allows you to enable or disable hyper-threading.

> **Note** By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are transmitted and received
  - Enterprise-level applications of various types and sizes
  - Game servers
  - Small and medium-sized database systems, caches, and search clusters
  - High-performance scientific computing
  - Video encoding applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfg7.large	2	8.0	None	1.2	10.0	900	Yes	2	2	6	20	1.0
ecs.hfg7.xlarge	4	16.0	None	2.0	10.0	1,000	Yes	4	3	15	30	1.5
ecs.hfg7.2xlarge	8	32.0	None	3.0	10.0	1,600	Yes	8	4	15	45	2.0
ecs.hfg7.3xlarge	12	48.0	None	4.5	10.0	2,000	Yes	8	6	15	60	2.5
ecs.hfg7.4xlarge	16	64.0	None	6.0	10.0	2,500	Yes	8	8	30	75	3.0

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfg7.6xlarge	24	96.0	None	8.0	10.0	3,000	Yes	12	8	30	90	4.0
ecs.hfg7.8xlarge	32	128.0	None	10.0	None	4,000	Yes	16	8	30	105	5.0
ecs.hfg7.12xlarge	48	192.0	None	16.0	None	6,000	Yes	24	8	30	150	8.0
ecs.hfg7.24xlarge	96	384.0	None	32.0	None	12,000	Yes	32	15	30	300	16.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about the specifications, see Description of instance specifications.

hfg6, general purpose instance family with high clock speed

Features

- Uses the X-Dragon architecture to provide predictable and consistent high performance and reduce virtualization overheads.
- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Provides high storage I/O performance based on large compute capacity.

> **Note** For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Offers a CPU-to-memory ratio of 1:4.
- Provides an ultra-high packet forwarding rate.
- Equipped with 3.1 GHz Intel® Xeon® Platinum 8269 (Cascade Lake) processors that deliver a maximum turbo frequency of 3.5 GHz for consistent computing performance.

  > **Note** The CPU of this instance family provides a 3.1 GHz clock speed. The Intel System Studio (ISS) feature may cause a lower clock speed to be displayed. Alibaba Cloud is working on to resolve this issue. This issue does not affect the actual clock speed of your instances.
  >
  > You can run the following commands separately and use the turbostat tool to view the actual clock speed.

  ```
  yum install kernel-tools
  ```

  ```
  turbostat
  ```

- Allows you to enable or disable hyper-threading.

  > **Note** By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are transmitted and received
  - Enterprise-level applications of various types and sizes
  - Websites and application servers
  - Game servers
  - Small and medium-sized database systems, caches, and search clusters
  - Data analysis and computing
  - Compute clusters and memory intensive data processing

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfg6.large	2	8.0	None	1.0	3.0	300	Yes	2	2	6	10	1.0
ecs.hfg6.xlarge	4	16.0	None	1.5	5.0	500	Yes	4	3	10	20	1.5
ecs.hfg6.2xlarge	8	32.0	None	2.5	8.0	800	Yes	8	4	10	25	2.0
ecs.hfg6.3xlarge	12	48.0	None	4.0	10.0	900	Yes	8	6	10	30	2.5
ecs.hfg6.4xlarge	16	64.0	None	5.0	10.0	1,000	Yes	8	8	20	40	3.0
ecs.hfg6.6xlarge	24	96.0	None	7.5	10.0	1,500	Yes	12	8	20	50	4.0
ecs.hfg6.8xlarge	32	128.0	None	10.0	None	2,000	Yes	16	8	20	60	5.0
ecs.hfg6.10xlarge	40	192.0	None	12.5	None	3,000	Yes	32	7	20	100	8.0

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfg6.16xlarge	64	256.0	None	20.0	None	4,000	Yes	32	8	20	120	10.0
ecs.hfg6.20xlarge	80	384.0	None	25.0	None	6,000	Yes	32	15	20	200	16.0

> **Note**
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

hfr7, memory optimized instance family with high clock speed

Features

- Uses the third-generation X-Dragon architecture to provide predictable and consistent high performance and reduce virtualization overheads.
- I/O optimized.
- Supports ESSDs and provides ultra-high I/O performance.
- Provides high storage I/O performance based on large computing capacity.

  > **Note** For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Offers a CPU-to-memory ratio of 1:8.
- Provides an ultra-high packet forwarding rate.
- Equipped with Intel® Xeon® Platinum 8369HB (Cooper Lake) or Intel® Xeon® Platinum 8369HC (Cooper Lake) processors that deliver a maximum turbo frequency of 3.8 GHz and a minimum clock speed of 3.3 GHz for consistent computing performance.
- Allows you to enable or disable hyper-threading.

> **Note** By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are transmitted and received
  - High-performance and in-memory databases
  - Data analysis, data mining, and distributed memory caching
  - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfr7.large	2	16.0	None	1.2	10.0	900	Yes	2	2	6	20	1.0
ecs.hfr7.xlarge	4	32.0	None	2.0	10.0	1,000	Yes	4	3	15	30	1.5
ecs.hfr7.2xlarge	8	64.0	None	3.0	10.0	1,600	Yes	8	4	15	45	2.0
ecs.hfr7.3xlarge	12	96.0	None	4.5	10.0	2,000	Yes	8	6	15	60	2.5
ecs.hfr7.4xlarge	16	128.0	None	6.0	10.0	2,500	Yes	8	8	30	75	3.0

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfr7.6xlarge	24	192.0	None	8.0	10.0	3,000	Yes	12	8	30	90	4.0
ecs.hfr7.8xlarge	32	256.0	None	10.0	None	4,000	Yes	16	8	30	105	5.0
ecs.hfr7.12xlarge	48	384.0	None	16.0	None	6,000	Yes	24	8	30	150	8.0
ecs.hfr7.24xlarge	96	768.0	None	32.0	None	12,000	Yes	32	15	30	300	16.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about the specifications, see Description of instance specifications.

hfr6, memory optimized instance family with high clock speed

Features

- Uses the X-Dragon architecture to provide predictable and consistent high performance and reduce virtualization overheads.
- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Provides high storage I/O performance based on large compute capacity.

> **Note** For more information about the storage I/O performance of the new generation of enterprise-level instance families, see Storage I/O performance.

- Offers a CPU-to-memory ratio of 1:8.
- Provides an ultra-high packet forwarding rate.
- Equipped with 3.1 GHz Intel® Xeon® Platinum 8269 (Cascade Lake) processors that deliver a maximum turbo frequency of 3.5 GHz for consistent computing performance.

  > **Note** The CPU of this instance family provides a 3.1 GHz clock speed. The Intel System Studio (ISS) feature may cause a lower clock speed to be displayed. Alibaba Cloud is working on to resolve this issue. This issue does not affect the actual clock speed of your instances.
  >
  > You can run the following commands separately and use the turbostat tool to view the actual clock speed.
  >
  > ```
  > yum install kernel-tools
  > ```
  >
  > ```
  > turbostat
  > ```

- Allows you to enable or disable hyper-threading.

  > **Note** By default, Hyper-Threading is enabled on ECS instances. For more information, see Customize CPU options.

- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Scenarios such as on-screen video comments and telecom data forwarding where large volumes of packets are transmitted and received
  - High-performance and in-memory databases
  - Data analysis, data mining, and distributed memory caching
  - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfr6.large	2	16.0	None	1.0	3.0	300	Yes	2	2	6	10	1.0

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfr6.xlarge	4	32.0	None	1.5	5.0	500	Yes	4	3	10	20	1.5
ecs.hfr6.2xlarge	8	64.0	None	2.5	8.0	800	Yes	8	4	10	25	2.0
ecs.hfr6.3xlarge	12	96.0	None	4.0	10.0	900	Yes	8	6	10	30	2.5
ecs.hfr6.4xlarge	16	128.0	None	5.0	10.0	1,000	Yes	8	8	20	40	3.0
ecs.hfr6.6xlarge	24	192.0	None	7.5	10.0	1,500	Yes	12	8	20	50	4.0
ecs.hfr6.8xlarge	32	256.0	None	10.0	None	2,000	Yes	16	8	20	60	5.0
ecs.hfr6.10xlarge	40	384.0	None	12.5	None	3,000	Yes	32	7	20	100	8.0
ecs.hfr6.16xlarge	64	512.0	None	20.0	None	4,000	Yes	32	8	20	120	10.0

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Base bandwidth (Gbit/s)	Burstable bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.hfr6.20xlarge	80	768.0	None	25.0	None	6,000	Yes	32	15	20	200	16.0

> **Note**
>
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about the specifications, see Description of instance specifications.

hfc5, compute optimized instance family with high clock speed

Features

- I/O optimized.
- Supports standard SSDs and ultra disks only.
- Provides consistent computing performance.
- Equipped with 3.1 GHz Intel® Xeon® Gold 6149 (Skylake) processors.
- Offers a CPU-to-memory ratio of 1:2.
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - High-performance web frontend servers
  - High-performance scientific and engineering applications
  - MMO gaming and video encoding

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.hfc5.large	2	4.0	None	1.0	300	No	2	2	6

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.hfc5.xlarge	4	8.0	None	1.5	500	No	2	3	10
ecs.hfc5.2xlarge	8	16.0	None	2.0	1,000	No	2	4	10
ecs.hfc5.3xlarge	12	24.0	None	2.5	1,300	No	4	6	10
ecs.hfc5.4xlarge	16	32.0	None	3.0	1,600	No	4	8	20
ecs.hfc5.6xlarge	24	48.0	None	4.5	2,000	No	6	8	20
ecs.hfc5.8xlarge	32	64.0	None	6.0	2,500	No	8	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about the specifications, see Description of instance specifications.

hfg5, general purpose instance family with high clock speed

Features

- I/O optimized.
- Supports standard SSDs and ultra disks only.
- Provides consistent computing performance.
- Equipped with 3.1 GHz Intel® Xeon® Gold 6149 (Skylake) processors.
- Offers a CPU-to-memory ratio of 1:4 (excluding the instance type with 56 vCPUs).
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - High-performance web frontend servers

- High-performance scientific and engineering applications
- MMO gaming and video encoding

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.hfg5.large	2	8.0	None	1.0	300	No	2	2	6
ecs.hfg5.xlarge	4	16.0	None	1.5	500	No	2	3	10
ecs.hfg5.2xlarge	8	32.0	None	2.0	1,000	No	2	4	10
ecs.hfg5.3xlarge	12	48.0	None	2.5	1,300	No	4	6	10
ecs.hfg5.4xlarge	16	64.0	None	3.0	1,600	No	4	8	20
ecs.hfg5.6xlarge	24	96.0	None	4.5	2,000	No	6	8	20
ecs.hfg5.8xlarge	32	128.0	None	6.0	2,500	No	8	8	20
ecs.hfg5.14xlarge	56	160.0	None	10.0	4,000	No	14	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

4.6. Compute optimized type family with GPU

4.6.1. Compute optimized instance families with GPU capabilities

This topic describes the features of compute optimized instance families with GPU capabilities and lists the instance types of each family.

Alibaba Cloud ECS server elastic computing

- Recommended instance families
  - vgn6i, lightweight compute optimized instance family with GPU capabilities
  - gn6i, compute optimized instance family with GPU capabilities
  - gn6e, compute optimized instance family with GPU capabilities
  - gn6v, compute optimized instance family with GPU capabilities
- Other available instance families
  - vgn5i, lightweight compute optimized instance family with GPU capabilities
  - gn5, compute optimized instance family with GPU capabilities
  - gn5i, compute optimized instance family with GPU capabilities
  - gn4, compute optimized family with GPU capabilities

vgn6i, lightweight compute optimized instance family with GPU capabilities

vgn6i is under invitational preview. To use vgn6i, submit a ticket.

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Uses NVIDIA T4 GPU computing accelerators.
- Contains virtual GPUs (vGPUs), which are the result of GPU virtualization with mediated pass-through.
  - Supports the 1/8, 1/4, and 1/2 computing capacity of NVIDIA® Tesla® T4 GPUs.
  - Supports 2, 4, and 8 GB of GPU memory.
- Offers a CPU-to-memory ratio of 1:5.
- Equipped with 2.5 GHz Intel ® Xeon ® Platinum 8163 (Skylake) processors.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Real-time rendering for cloud gaming
  - Real-time rendering for AR and VR applications
  - AI (deep learning and machine learning) inference for the elastic deployment of Internet services

- Educational environment of deep learning
- Modeling experiment environment of deep learning

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.vgn6i-m4.xlarge	4	23.0	None	1/4 × T4	4	3.0	500	Yes	2	4	10
ecs.vgn6i-m8.2xlarge	10	46.0	None	1/2 × T4	8	4.0	800	Yes	4	5	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

gn6i, compute optimized instance family with GPU capabilities

Features

- I/O optimized.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Supports enhanced SSDs (ESSDs) that deliver millions of IOPS, standard SSDs, and ultra disks.
- Uses NVIDIA T4 GPU computing accelerators.
  - Powered by the new NVIDIA Turing architecture.
  - 16 GB GPU memory (320 GB/s bandwidth).
  - 2,560 CUDA cores per GPU.
  - Up to 320 Turing Tensor cores.
  - Mixed-precision Tensor cores support 65 FP16 TFLOPS, 130 INT8 TOPS, and 260 INT4 TOPS.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:

- AI (deep learning and machine learning) inference for computer vision, speech recognition, speech synthesis, natural language processing (NLP), machine translation, and recommendation systems
- Real-time rendering for cloud gaming
- Real-time rendering for AR and VR applications
- Graphics workstations or overloaded graphics computing
- GPU-accelerated databases
- High-performance computing

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.gn6i-c4g1.xlarge	4	15.0	None	1 × T4	16	4.0	500	Yes	2	2	10
ecs.gn6i-c8g1.2xlarge	8	31.0	None	1 × T4	16	5.0	800	Yes	2	2	10
ecs.gn6i-c16g1.4xlarge	16	62.0	None	1 × T4	16	6.0	1,000	Yes	4	3	10
ecs.gn6i-c24g1.6xlarge	24	93.0	None	1 × T4	16	7.5	1,200	Yes	6	4	10
ecs.gn6i-c24g1.12xlarge	48	186.0	None	2 × T4	32	15.0	2,400	Yes	12	6	10

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.gn6i-c24g1.24xlarge	96	372.0	None	4 × T4	64	30.0	4,800	Yes	24	8	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

gn6e, compute optimized instance family with GPU capabilities

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Uses NVIDIA V100 (32 GB NVLink) GPU processors.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Uses NVIDIA V100 GPU computing accelerators that have the SXM2 module.
  - Powered by the new NVIDIA Volta architecture.
  - 32 GB HBM2 GPU memory (900 GB/s bandwidth).
  - 5,120 CUDA cores per GPU.
  - 640 Tensor cores per GPU.
  - Supports up to six NVLink connections and a total bandwidth of 300 GB/s (25 GB/s per connection).
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Deep learning applications such as training and inference applications of AI algorithms used in image classification, autonomous vehicles, and speech recognition
  - Scientific computing applications, such as fluid dynamics, finance, molecular dynamics, and environmental analysis

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.gn6e-c12g1.3xlarge	12	92.0	None	1 × V100	32	5.0	800	Yes	8	6	10
ecs.gn6e-c12g1.12xlarge	48	368.0	None	4 × V100	128	16.0	2,400	Yes	8	8	20
ecs.gn6e-c12g1.24xlarge	96	736.0	None	8 × V100	256	32.0	4,800	Yes	16	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

gn6v, compute optimized instance family with GPU capabilities

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Uses NVIDIA V100 GPU processors.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Uses NVIDIA V100 GPU computing accelerators that have the SXM2 module.
  - Powered by the new NVIDIA Volta architecture.
  - 16 GB HBM2 GPU memory (900 GB/s bandwidth).

- 5,120 CUDA cores per GPU.
- 640 Tensor cores per GPU.
- Supports up to six NVLink connections and a total bandwidth of 300 GB/s (25 GB/s per connection).

- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Deep learning applications such as training and inference applications of AI algorithms used in image classification, autonomous vehicles, and speech recognition
  - Scientific computing applications, such as fluid dynamics, finance, molecular dynamics, and environmental analysis

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.gn6v-c8g1.2xlarge	8	32.0	None	1 × NVIDIA V100	1 × 16	2.5	800	Yes	4	4	10
ecs.gn6v-c8g1.8xlarge	32	128.0	None	4 × NVIDIA V100	4 × 16	10.0	2,000	Yes	8	8	20
ecs.gn6v-c8g1.16xlarge	64	256.0	None	8 × NVIDIA V100	8 × 16	20.0	2,500	Yes	16	8	20
ecs.gn6v-c10g1.20xlarge	82	336.0	None	8 × NVIDIA V100	8 × 16	32.0	4,500	Yes	16	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

vgn5i, lightweight compute optimized instance family with GPU capabilities

Features

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

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.vgn5i-m1.large	2	6.0	None	1/8 × P4	1	1.0	300	Yes	2	2	6

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.vgn5i-m2.xlarge	4	12.0	None	1/4 × P4	2	2.0	500	Yes	2	3	10
ecs.vgn5i-m4.2xlarge	8	24.0	None	1/2 × P4	4	3.0	800	Yes	2	4	10
ecs.vgn5i-m8.4xlarge	16	48.0	None	1 × P4	8	5.0	1,000	Yes	4	5	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

gn5, compute optimized instance family with GPU capabilities

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Uses NVIDIA P100 GPU processors.
- Offers multiple CPU-to-memory ratios.
- Attached with high-performance local NVMe SSDs.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Deep learning
  - Scientific computing applications, such as fluid dynamics, finance, genomics, and environmental analysis

- Server-side GPU compute workloads such as high-performance computation, rendering, and multi-media coding and decoding

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.gn5-c4g1.xlarge	4	30.0	440	1 × NVIDIA P100	1 × 16	3.0	300	No	1	3	10
ecs.gn5-c8g1.2xlarge	8	60.0	440	1 × NVIDIA P100	1 × 16	3.0	400	No	1	4	10
ecs.gn5-c4g1.2xlarge	8	60.0	880	2 × NVIDIA P100	2 × 16	5.0	1,000	No	2	4	10
ecs.gn5-c8g1.4xlarge	16	120.0	880	2 × NVIDIA P100	2 × 16	5.0	1,000	No	4	8	20
ecs.gn5-c28g1.7xlarge	28	112.0	440	1 × NVIDIA P100	1 × 16	5.0	1,000	No	8	8	20
ecs.gn5-c8g1.8xlarge	32	240.0	1,760	4 × NVIDIA P100	4 × 16	10.0	2,000	No	8	8	20

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.gn5-c28g1.14xlarge	56	224.0	880	2 × NVIDIA P100	2 × 16	10.0	2,000	No	14	8	20
ecs.gn5-c8g1.14xlarge	54	480.0	3,520	8 × NVIDIA P100	8 × 16	25.0	4,000	No	14	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

gn5i, compute optimized instance family with GPU capabilities

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Uses NVIDIA P4 GPU processors.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Deep learning and inference
  - Server-side GPU compute workloads such as multimedia encoding and decoding

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.gn5i-c2g1.large	2	8.0	None	1 × NVIDIA P4	1 × 8	1.0	100	Yes	2	2	6
ecs.gn5i-c4g1.xlarge	4	16.0	None	1 × NVIDIA P4	1 × 8	1.5	200	Yes	2	3	10
ecs.gn5i-c8g1.2xlarge	8	32.0	None	1 × NVIDIA P4	1 × 8	2.0	400	Yes	4	4	10
ecs.gn5i-c16g1.4xlarge	16	64.0	None	1 × NVIDIA P4	1 × 8	3.0	800	Yes	4	8	20
ecs.gn5i-c16g1.8xlarge	32	128.0	None	2 × NVIDIA P4	2 × 8	6.0	1,200	Yes	8	8	20
ecs.gn5i-c28g1.14xlarge	56	224.0	None	2 × NVIDIA P4	2 × 8	10.0	2,000	Yes	14	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

gn4, compute optimized family with GPU capabilities

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Uses NVIDIA M40 GPU processors.
- Offers multiple CPU-to-memory ratios.
- Equipped with 2.5 GHz Intel ® Xeon ® E5-2682 v4 (Broadwell) processors.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Deep learning
  - Scientific computing applications, such as fluid dynamics, finance, genomics, and environmental analysis
  - Server-side GPU compute workloads such as high-performance computation, rendering, and multi-media coding and decoding

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.gn4-c4g1.xlarge	4	30.0	None	1 × NVIDIA M40	1 × 12	3.0	300	No	1	3	10
ecs.gn4-c8g1.2xlarge	8	30.0	None	1 × NVIDIA M40	1 × 12	3.0	400	No	1	4	10
ecs.gn4.8xlarge	32	48.0	None	1 × NVIDIA M40	1 × 12	6.0	800	No	3	8	20
ecs.gn4-c4g1.2xlarge	8	60.0	None	2 × NVIDIA M40	2 × 12	5.0	500	No	1	4	10

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.gn4-c8g1.4xlarge	16	60.0	None	2 × NVIDIA M40	2 × 12	5.0	500	No	1	8	20
ecs.gn4.14xlarge	56	96.0	None	2 × NVIDIA M40	2 × 12	10.0	1,200	No	4	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

4.6.2. Create an instance equipped with NVIDIA GPUs

Instances equipped with NIVIDIA GPUs must have drivers installed to use the GPUs. The drivers consist of GPU drivers and GRID drivers. You can select the automatic installation method when you create an instance or you can manually install a driver after the instance is created.

Prerequisites

You must complete the following preparations to create an ECS instance:

1. Create an account and complete the account information.
   - Create an Alibaba Cloud account.
   - Bind your credit card or PayPal account.
   - To purchase ECS instances in mainland China regions, you must complete real-name verification.
2. Alibaba Cloud provides a default VPC. If you do not want to use the default VPC, you can create a VPC and a VSwitch in the specified region. For more information, see Create an IPv4 VPC network.
3. Alibaba Cloud provides a default security group. If you do not want to use the default security group, you can create a security group in the region where the instance is created. For more information, see Create a security group.

If you need other extended features, you must complete corresponding preparations:

- To specify an SSH key pair when you create a Linux instance, you must create an SSH key pair in the corresponding region. For more information, see Create an SSH key pair.
- To configure user data, you must first prepare user data. For more information about how to prepare user data, see Prepare user data.
- To associate an ECS instance with an instance RAM role, you must create the RAM role, assign a permission policy to the role, and then bind the role to the instance. For more information, see Bind an instance RAM role.

Context

The following drivers are involved:

- GPU drivers: used to drive physical GPUs. You can install GPU drivers only on GPU-accelerated instances that are not equipped with vGPUs.
- GRID drivers: used to grant graphics acceleration capabilities to instances. You can install GRID drivers on GPU-accelerated instances to grant graphics acceleration capabilities to the instances, regardless of whether the instances are equipped with vGPUs. Instances of the vgn6i and vgn5i instance families are equipped with vGPUs.

The following table describes whether the preceding drivers can be installed on different types of GPU-accelerated instances.

Driver	GPU-accelerated instance equipped with vGPUs (vgn6i and vgn5i)	GPU-accelerated instance not equipped with vGPUs
GPU driver	Not supported	Supported
GRID driver	Supported	Supported

Procedure

The following procedure focuses on the configurations related to NVIDIA GPUs when an instance equipped with NVIDIA GPUs is created. For more information about general configurations, see Create an instance by using the provided wizard.

1. Go to the Custom Launch tab in the ECS console.
2. Complete the settings in the Basic Configurations step.

   > **Note** GPU-accelerated ECS instance types are available only in specific regions and zones. For more information, see ECS Instance Types Available for Each Region. Select a billing method and enter an instance type name to search for the target instance type.

   Instances of the vgn6i and vgn5i instance families are equipped with vGPUs that are generated from GPU slice virtualization. Therefore, you can install only GRID drivers on these instances. Configure the instance and image based on the instance type.

   - Create a GPU-accelerated instance equipped with vGPUs (vgn6i and vgn5i)
     - **Instance Type**: Click **Heterogeneous Computing**, click **Visualization Compute Optimized Type with GPU**, and then select the instance type that you need.

- **Image**: The operating system type determines the installation method of the GRID driver.
  - Windows: Search for the **Windows Server 2016 Chinese Version with GRID Driver** paid image in Alibaba Cloud Marketplace. This image is pre-installed with a GRID driver that has a license activated and you do not need to manually install the driver.

    > **Note** More paid Windows Server images will be available soon.

  - Linux: You must manually purchase a license for a GRID driver and manually install the GRID driver after the instance is created. For more information, see Install NVIDIA GRID drivers on vgn6i or vgn5i Linux instances.

- Create a GPU-accelerated instance not equipped with vGPUs
  - **Instance Type**: Click **Heterogeneous Computing**, click **Compute Optimized Type with GPU**, and then select the instance type that you need.
  - **Image**: You can install GPU drivers on GPU-accelerated instances that are not equipped with vGPUs. Perform the following steps to install the drivers.

    > **Note** If you select Shared Image or Custom Image, make sure that the selected image is pre-installed with the required GPU driver and relevant software.

- Select Auto-install GPU Driver.

  Public images are base images provided by Alibaba Cloud or third-party vendors. The following Linux images support the automatic installation of GPU drivers:

  - CentOS 64-bit images that are applied and tested by Alibaba Cloud
  - Ubuntu 16.04 64-bit images
  - Ubuntu 18.04 64-bit images
  - SUSE Linux Enterprise Server 12 SP2 64-bit images
  - Alibaba Cloud Linux 64-bit images

  If you select an image that supports the automatic installation of a GPU driver, select **Auto-install GPU Driver**, and then select the versions of the GPU driver, CUDA, and cuDNN library. If you want to create an instance for a new business system, we recommend that you select the latest versions.

  After you select **Auto-install GPU Driver**, you can select whether to automatically install the GPU cloud accelerator. Apsara AI Accelerator (AIACC) is the GPU cloud accelerator provided by Alibaba Cloud. It enables you to build a deep learning training system that provides a high-performance distributed architecture and accelerate AI training performance. For more information, see Use the GPU cloud accelerator AIACC.

  > **Note** The GPU cloud accelerator is not supported in CentOS 8, CentOS 6, SUSE Linux, and Alibaba Cloud Linux.

  If you select an image that supports the automatic installation of a GPU driver and do not select **Auto-install GPU Driver**, you can still configure the installation script of the GPU driver in the **User Data** module. For more information, see the Automatic installation script V3.1 section.

  > **Note** If you call the RunInstances operation to create an instance equipped with NVIDIA GPUs, you must use the UserData parameter to upload the installation script. The script content must be Base64-encoded.

- Select an image that is pre-installed with a GPU driver and relevant software in Alibaba Cloud Marketplace.

  Alibaba Cloud Marketplace provides images that have operating systems, application environments, and various software pre-installed. Alibaba Cloud Marketplace images are reviewed by Alibaba Cloud to ensure quality and stability. You can use these images to deploy ECS instances without additional configurations. Alibaba Cloud Marketplace provides images that support deep learning and machine learning.

  - To create an instance equipped with NVIDIA GPUs for deep learning scenarios, you can select an image pre-installed with deep learning frameworks. You can use the keyword *deep learning* to search for available images in Alibaba Cloud Marketplace. Only CentOS 7.3 is supported.
  - The NVIDIA GPU Cloud VM Image is an optimized environment for running deep learning frameworks, HPC applications, and HPC visualization tools available from the NVIDIA GPU Cloud (NGC) container registry. For more information, see Deploy an NGC environment on instances with GPU capabilities.

- Manually install a GRID driver after an instance is created. For more information, see Manually install a GPU driver.

You can also install GRID drivers for GPU-accelerated instances that are not equipped with vGPUs. The operating system type determines the installation method of the GRID driver.

- Windows: Search for the **Windows Server 2016 Chinese Version with GRID Driver** paid image in Alibaba Cloud Marketplace. This image is pre-installed with a GRID driver that has a license activated and you do not need to manually install the driver.

  > **Note** More paid Windows Server images will be available soon.

- Linux: You must purchase a license for a GRID driver and manually install the GRID driver after the instance is created. For more information, see Install NVIDIA GRID drivers on GPU-accelerated Linux instances.

3. Complete the settings in the Networking step. When you configure the parameters, take note of the following items:
   - **Network Type**: Select **VPC**.
   - **Public IP Address**: Select a bandwidth based on your actual needs.

     > **Notice** If you select an image of Windows 2008 R2 or earlier in the **Basic Configurations** step, you cannot connect to the instance equipped with NVIDIA GPUs by using VNC after the GPU driver is installed and activated. When you connect to the instance, a black screen or the startup interface persists. You must select **Assign Public IP Address** in the Public IP Address section, or associate an elastic IP address (EIP) after you create the instance to connect to the instance over other protocols, such as Remote Desktop in Windows (RDP), PCOIP, and XenDesktop HDX 3D. RDP does not support applications such as DirectX and OpenGL. You must install the VNC server and client on your own.

4. Complete the settings in the System Configurations step. When you configure the parameters, take note of the following items:
   - **Logon Credentials**: We recommend that you select **Key Pair** or **Password**. If you select **Set Later** and want to log on to the instance by using VNC, you must bind an SSH key pair or reset the password and then restart the instance for the modification to take effect. If the GPU driver is being installed, restarting the instance will cause the installation to fail.
   - **User Data**:
     - If you select **Auto-install GPU Driver** in the **Image** section in the **Basic Configurations** step, the precautions and shell script content for automatically installing the GPU driver and CUDA will be displayed in the User Data section. The automatic installation script has been updated to V3.1.
     - If you do not select **Auto-install GPU Driver**, you can configure the installation script in the **User Data** section. For information about a script example, see Automatic installation script V3.1.

5. Configure the parameters in the Grouping step, confirm the configurations in the Preview step, and then click Create Order or Create Instance.

If you configure the automatic installation script, the system automatically installs the GPU driver after the instance is started. After the installation is complete, the instance will be automatically restarted for the GPU driver to operate properly.

> **Note** The GPU driver is more stable in persistence mode. The script automatically enables the persistence mode for the GPU driver. Then, the script adds the corresponding command as a Linux system service to ensure that the persistence mode is enabled for the GPU driver by default after the instance restarts.

The automatic installation process may take 10 to 20 minutes based on the internal bandwidth and CPU cores of different instance types. During the process, the GPU is not available. To prevent installation failures and instance unavailability, do not perform operations on the instance or install any other GPU-related software until the installation is complete. You can connect to the instance and view the installation progress and result in the installation logs.

- If the installation is in progress, you can see the installation progress bar.
- If the installation succeeds, **ALL INSTALL OK** is displayed in the result.
- If the installation fails, **INSTALL FAIL** is displayed in the result.
- You can check the detailed installation log entries in the */root/auto_install/auto_install.log* file.

> **Note** If you change the operating system after you create the instance, make sure that you use an image that can have a GPU driver and CUDA automatically installed to prevent failures in automatic installation.

Script for automatically installing a GPU driver

When the instance is started for the first time, cloud-init automatically runs the shell script to install the GPU driver, CUDA, and cuDNN library.

- The following table lists the available versions of the GPU driver, CUDA, and cuDNN library when you select **Auto-install GPU Driver**.

CUDA	GPU driver	cuDNN	Supported public image version (only images applied and tested by Alibaba Cloud)	Supported instance family
10.2.89	440.64.00	7.6.5	◦ Alibaba Cloud Linux 2 ◦ Ubuntu 18.04 and Ubuntu 16.04 ◦ CentOS 8.x, CentOS 7.x, and CentOS 6.x	◦ gn6v, gn6i, gn6e, gn5, gn5i, and gn4 ◦ ebmgn6v, ebmgn6i, ebmgn6e, and ebmgn5i

CUDA	GPU driver	cuDNN	Supported public image version (only images applied and tested by Alibaba Cloud)	Supported instance family
10.1.168	○ 440.64.00 ○ 418.126.02	○ 7.6.5 ○ 7.5.0	○ Ubuntu 18.04 and Ubuntu 16.04 ○ CentOS 7.x and CentOS 6.x	○ gn6v, gn6i, gn6e, gn5, gn5i, and gn4 ○ ebmgn6v, ebmgn6i, ebmgn6e, and ebmgn5i
10.0.130	○ 440.64.00 ○ 418.126.02	○ 7.6.5 ○ 7.5.0 ○ 7.4.2 ○ 7.3.1	○ Ubuntu 18.04 and Ubuntu 16.04 ○ CentOS 7.x and CentOS 6.x	○ gn6v, gn6i, gn6e, gn5, gn5i, and gn4 ○ ebmgn6v, ebmgn6i, ebmgn6e, and ebmgn5i
9.2.148	○ 440.64.00 ○ 418.126.02 ○ 390.116	○ 7.6.5 ○ 7.5.0 ○ 7.4.2 ○ 7.3.1 ○ 7.1.4	○ Ubuntu 16.04 ○ CentOS 7.x and CentOS 6.x	○ gn6v, gn6e, gn5, gn5i, and gn4 ○ ebmgn6v, ebmgn6e, and ebmgn5i
9.0.176	○ 440.64.00 ○ 418.126.02 ○ 390.116	○ 7.6.5 ○ 7.5.0 ○ 7.4.2 ○ 7.3.1 ○ 7.1.4 ○ 7.0.5	○ Ubuntu 16.04 ○ CentOS 7.x and CentOS 6.x ○ SUSE 12 SP2	○ gn6v, gn6e, gn5, gn5i, and gn4 ○ ebmgn6v, ebmgn6e, and ebmgn5i
8.0.61	○ 440.64.00 ○ 418.126.02 ○ 390.116	○ 7.1.3 ○ 7.0.5	○ Ubuntu 16.04 ○ CentOS 7.x and CentOS 6.x	○ gn5, gn5i, and gn4 ○ ebmgn5i

- If you configure the installation script in the **User Data** section, you can see the script in the Automatic installation script V3.1 section.

  The automatic installation script V3.1 has the following benefits:

  - Provides the latest GPU driver, CUDA, and cuDNN library.
  - After you log on to the instance, if the GPU driver is being installed, you can see the installation progress bar. If the installation succeeds, the instance is automatically restarted. After you log on to the instance again, **ALL INSTALL OK** is displayed in the result. If the installation fails, **INSTALL FAIL** is displayed in the result.

To use the automatic installation script V3.1, you must modify the version parameters of the GPU driver, CUDA, and cuDNN library in the installation script, and specify whether to install AIACC. If you do not install AIACC, you must change the value of IS_INSTALL_PERSEUS to FALSE. Example:

```
IS_INSTALL_PERSEUS="FALSE"
DRIVER_VERSION="440.64.00"
CUDA_VERSION="10.2.89"
CUDNN_VERSION="7.6.5"
```

> **Note** If the image is a CentOS or SUSE image, the installation script uses the *.run* installation package. If the image is an Ubuntu image, the installation script uses the *.deb* installation package.

Automatic installation script V3.1

```
#! /bin/sh

#Please input version to install
IS_INSTALL_PERSEUS=""
DRIVER_VERSION=""
CUDA_VERSION=""
CUDNN_VERSION=""
IS_INSTALL_RAPIDS="FALSE"

INSTALL_DIR="/root/auto_install"

#using .deb to install driver and cuda on ubuntu OS
#using .run to install driver and cuda on ubuntu OS
auto_install_script="auto_install.sh"

script_download_url=$(curl http://100.100.100.200/latest/meta-data/source-address | head -1)"/opsx/ecs/linux/binary/script/${auto_install_script}"
echo $script_download_url

mkdir $INSTALL_DIR && cd $INSTALL_DIR
wget -t 10 --timeout=10 $script_download_url && sh ${INSTALL_DIR}/${auto_install_script} $DRIVER_VERSION $CUDA_VERSION $CUDNN_VERSION $IS_INSTALL_PERSEUS $IS_INSTALL_RAPIDS
```

Related information

References

- RunInstances
- Manually install a GPU driver
- Install GRID drivers on instances with GPU capabilities
- Manually uninstall the GPU driver
- GPU monitoring

4.6.3. Manually install a GPU driver

If you do not select to automatically install a GPU driver when you create an instance, you must manually install the GPU driver after the instance is created. This topic describes how to manually install a GPU driver on a GPU-accelerated instance.

Context

You can install only GRID drivers on vgn6i or vgn5i instances. Therefore, the procedure in this topic does not apply to vgn6i or vgn5i instances. For more information about how to install GRID drivers, see Create an instance equipped with NVIDIA GPUs and Install NVIDIA GRID drivers on vgn6i or vgn5i Linux instances.

Procedure

1. Go to the NVIDIA official website.
2. Manually search for the suitable driver.
   i. Select a product type, series, and product based on the GPU with which your instance type is equipped. The following table lists the information of GPUs with which different instance types are equipped.

Item	gn4	gn5	gn5i	gn6v	gn6i
Product type	Tesla	Tesla	Tesla	Tesla	Tesla
Product series	M-Class	P-Series	P-Series	V-Series	T-Series
Product	M40	Tesla P100	Tesla P4	Tesla V100	Tesla T4

   ii. Select an operating system based on the image of the instance. If your instance runs Debian, select **Linux 64-bit**. If the drop-down list does not contain the operating system that you instance uses, click **Show All Operating Systems** in the lower part of the drop-down list.
   iii. Select a CUDA Toolkit version.
   iv. Select a language.
   v. Click **SEARCH**.
3. Confirm the driver information, and click **DOWNLOAD**. If your instance runs a Linux operating system, do not close the download page. When you install the GPU driver, you may need to refer to the installation steps in the **ADDITIONAL INFORMATION** section.
4. Install the GPU driver.
   a. Install the GPU driver on a Linux instance:

- Install the GPU driver on a Linux instance:
  a. Download and install the kernel-devel and kernel-header packages based on your kernel version.

     > **Note** The mismatch between kernel and kernel-devel versions results in a driver compilation error when the driver is installed from the .rpm file. You can run the rpm -qa | grep kernel command on an instance to check whether the versions match. Make sure that the versions match and re-install the driver.

  b. Run the **sudo rpm -qa | grep $(uname -r)** command to check whether the kernel-devel and kernel-header packages are downloaded and installed.

     If information similar to the following content is displayed, the kernel-devel and kernel-header packages are installed. CentOS 7.3 is used in this example.

     ```
     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
     ```

  c. Perform the following steps in the **ADDITIONAL INFORMATION** section on the download page to install the GPU driver.

     The steps in the **ADDITIONAL INFORMATION** section are shown in the following figure. Linux Ubuntu 14.04 64-bit is used in this example.
- Install the GPU driver on a Windows instance:

  Double-click the package and follow the prompts to complete the installation.

  > **Note** On Windows instances where the installed GPU drivers have taken effect, Windows Remote Desktop Protocol (RDP) may not support DirectX- and OpenGL-based applications. In this case, you must install the Virtual Network Computing (VNC) service and client or use other protocols that support these applications, such as PC over IP (PCoIP) and XenDesktop HDX 3D.

4.6.4. Manually uninstall the GPU driver

You can manually uninstall the GPU driver. The uninstall command varies depending on the installation method of the GPU driver and the image type.

Context

The `root` account is used in this topic. If you are a common user, run a `sudo` command to obtain the permissions of a `root` user before you proceed.

If you have installed the GPU driver by using the automatic installation feature, the uninstall method of the GPU driver varies depending on the operating system type:

- The run mode is used for CentOS and SUSE.

- The deb mode is used for Ubuntu.

Uninstall the GPU driver in Ubuntu

If you have used the deb package to install the GPU driver, we recommend that you perform the following operations to uninstall the GPU driver. Driver 410.104, CUDA 10.0.130, and cuDNN 7.5.0 are used in the examples.

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

```
apt-get remove --purge nvidia-*
```

2. Run the following commands to uninstall CUDA and the cuDNN library:

```
apt autoremove --purge cuda-10-0
rm -rf /usr/local/cuda-10.0
```

3. Run the following command to restart the instance:

```
reboot
```

If you have used the run package to install the GPU driver, we recommend that you perform the following operations to uninstall the GPU driver:

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

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

2. Run the following commands to uninstall CUDA and the cuDNN library:

```
/usr/local/cuda/bin/cuda-uninstaller
rm -rf /usr/local/cuda-10.0
```

> **Note** The uninstall command may differ between CUDA versions. If the *cuda-uninstaller* file does not exist, check whether a file whose name starts with *uninstall_cuda* exists in the */usr/local/cuda/bin/* directory. If yes, replace *cuda-uninstaller* in the command with the file name that starts with uninstall_cuda.

3. Run the following command to restart the instance:

```
reboot
```

Uninstall the GPU driver in CentOS

If you have used the run package to install the GPU driver, we recommend that you perform the following operations to uninstall the GPU driver. Driver 410.104, CUDA 10.0.130, and cuDNN 7.5.0 are used in the examples.

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

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

2. Run the following commands to uninstall CUDA and the cuDNN library:

```
/usr/local/cuda/bin/cuda-uninstaller
rm -rf /usr/local/cuda-10.0
```

> **Note** The uninstall command may differ between CUDA versions. If the *cuda-uninstaller* file does not exist, check whether a file whose name starts with *uninstall_cuda* exists in the */usr/local/cuda/bin/* directory. If yes, replace *cuda-uninstaller* in the command with the file name that starts with *uninstall_cuda*.

3. Run the following command to restart the instance:

```
reboot
```

If you have used the RPM package to install the GPU driver in CentOS 7, we recommend that you perform the following operations to uninstall the GPU driver:

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

```
yum remove xorg-x11-drv-nvidia nvidia-kmod cuda-drivers
yum remove nvidia-diag-driver-local-repo-rhel7-410.104
```

2. Run the following commands to uninstall CUDA and the cuDNN library:

```
yum remove /usr/local/cuda-10.0
rm -rf /usr/local/cuda-10.0
```

3. Run the following command to restart the instance:

```
reboot
```

If you have used the RPM package to install the GPU driver in CentOS 6, we recommend that you perform the following operations to uninstall the GPU driver:

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

```
yum remove xorg-x11-drv-nvidia nvidia-kmod cuda-drivers
yum remove nvidia-diag-driver-local-repo-rhel6-410.104
```

2. Run the following command to uninstall CUDA:

```
yum remove /usr/local/cuda-10.0
```

3. Run the following command to restart the instance:

```
reboot
```

Uninstall the GPU driver in SUSE

If you have used the run package to install the GPU driver, we recommend that you perform the following operations to uninstall the GPU driver. CUDA 9.0.176 is used in the examples.

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

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

2. Run the following commands to uninstall CUDA and the cuDNN library:

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

3. Run the following command to restart the instance:

```
reboot
```

4.6.5. Install NVIDIA GRID drivers on GPU-accelerated Linux instances

This topic describes how to install a NVIDIA GRID driver and build a desktop environment on a GPU-accelerated Linux instance.

Prerequisites

- A GPU-accelerated instance that can access the Internet is created.

  > **Note** This topic describes how to install NVIDIA GRID drivers on GPU-accelerated Linux instances. For GPU-accelerated Windows instances, you can select paid images with NVIDIA GRID drivers pre-installed when you create the instances. For more information, see Create an instance equipped with NVIDIA GPUs.

  When you create an instance, we recommend that you select an image from **Public Image**. If you select an image pre-installed with the NVIDIA GRID driver from **Marketplace Image**, you must disable the nouveau driver after you create the instance.

  nouveau is an open source driver. It must be disabled before you can install another driver. You can create a *nouveau.conf* file under the */etc/modprobe.d* directory and write `blacklist nouveau` to the file to disable nouveau.

- A VNC application is installed. VNC Viewer is used in this example.
- A GRID license is obtained from NVIDIA. You must build a license server. You can purchase an ECS instance and build a license server by following the tutorial on the official NVIDIA website.

Context

You must install a NVIDIA GRID driver if your GPU-accelerated instances require Open Graphics Library (OpenGL). By default, the NVIDIA GRID license that is granted to NVIDIA GPUs such as P100, P4, and V100 is not activated. You can activate the license to use OpenGL.

> **Note** Only NVIDIA partners can download the driver from the official NVIDIA website. This topic describes how to obtain the NVIDIA GRID driver package from Alibaba Cloud.

This topic describes how to install NVIDIA GRID drivers on GPU-accelerated instances that are not equipped with vGPUs. For information about how to install NVIDIA GRID drivers on vgn6i or vgn5i GPU-accelerated instances that are equipped with vGPUs, see Install NVIDIA GRID drivers on vgn6i or vgn5i Linux instances.

Procedure

Perform the following steps to install a NVIDIA GRID driver:

- Ubuntu 16.04 64-bit:
  i. Install a NVIDIA GRID driver on a Linux instance that runs Ubuntu 16.04 64-bit
  ii. Test NVIDIA GRID drivers installed on Linux instances that run Ubuntu 16.04 64-bit
- CentOS 7.3 64-bit:
  i. Install a NVIDIA GRID driver on a Linux instance that runs CentOS 7.3 64-bit
  ii. Test NVIDIA GRID drivers installed on Linux instances that run CentOS 7.3 64-bit

Install a NVIDIA GRID driver on a Linux instance that runs Ubuntu 16.04 64-bit

1. Connect to the Linux instance. For more information, see Connect to a Linux instance by using a username and password.
2. Run the following commands in sequence to upgrade the system and install KDE:

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

3. Run the `reboot` command to restart the system.
4. Connect to the Linux instance again. 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, select *NVIDIA-Linux-x86_64-410.39-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 and follow the on-screen tips 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 the `nvidia-smi` command to test whether the NVIDIA GRID driver is installed.

   If a command output similar to the following one is displayed, the NVIDIA GRID driver is installed:

7. Add a license server and activate the license.
   i. Run the `cd /etc/nvidia` command to go to the */etc/nvidia* directory.

ii. Run the `cp gridd.conf.template gridd.conf` command to create a file named *gridd.conf*.

iii. Add license server information to the *gridd.conf* file.

```
ServerAddress=<IP address of the license server>
ServerPort=<Port of the license server (Default port: 7070)>
FeatureType=2
EnableUI=TRUE
```

8. Run the following command to install x11vnc:

```
apt-get install x11vnc
```

9. Run the `lspci | grep NVIDIA` command to query the GPU BusID.

   In this example, the GPU BusID is `00:07.0`.

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

    i. Run the `nvidia-xconfig --enable-all-gpus --separate-x-screens` command.

    ii. Add your GPU BusID to `Section "Device"` in the */etc/X11/xorg.conf* file. In this example, `BusID "PCI:0:7:0"` is added.

    iii. Run the `reboot` command to restart the system.

Test NVIDIA GRID drivers installed on Linux instances that run Ubuntu 16.04 64-bit

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

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

2. Run the `startx` command to start X Server.

   - If the `startx` command is unavailable, run the `apt-get install xinit` command to install the GLX application.
   - If you run the `startx` command, the `hostname: Name or service not known` error may be reported. This error does not affect the startup of X Server. You can run the `hostname` command to obtain the hostname of the instance, and modify the */etc/hosts* file by replacing the `hostname` that follows `127.0.0.1` with the actual hostname 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 a command output similar to the following one is displayed, x11vnc is started. In this case, 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 security group to which the instance belongs, to allow inbound traffic on TCP port 5900. For more information, see Add security group rules.
5. On the local machine, start VNC Viewer and enter `<Public IP address of the instance>:5900` to connect to the instance and go to KDE.
6. Run the `glxinfo` command to view the configurations supported by the current NVIDIA GRID driver.
   i. Start a new terminal session of the SSH client.
   ii. Run the `export DISPLAY=:1` command.
   iii. Run the `glxinfo -t` command to list the configurations supported by the current NVIDIA GRID driver.
7. Run the `glxgears` command to test the NVIDIA GRID driver.
   i. On KDE, right-click the desktop and select **Run Command**.
   ii. Run the `glxgears` command to start the testing application.
   If a window similar to the following one is displayed, the NVIDIA GRID driver functions properly:

Install a NVIDIA GRID driver on a Linux instance that runs CentOS 7.3 64-bit

1. Connect to the Linux instance. For more information, see Connect to a Linux instance by using a username and password.
2. Run the following commands in sequence to upgrade the system and install KDE:

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

3. Run the `reboot` command to restart the system.
4. Connect to the Linux instance again. Run the following command to download and decompress the NVIDIA GRID driver package.

   The NVIDIA GRID driver package contains the drivers for various operating systems. For Linux, select *NVIDIA-Linux-x86_64-410.39-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:
   i. Run the `vim /etc/modprobe.d/blacklist.conf` command and add `blacklist nouveau` to the file.
   ii. Run the `vim /lib/modprobe.d/dist-blacklist.conf` command and add the following content:

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

iii. Run the `mv /boot/initramfs-$(uname -r).img /boot/initramfs-$(uname -r)-nouveau.img` command.

iv. Run the `dracut /boot/initramfs-$(uname -r).img $(uname -r)` command.

6. Run the `reboot` command to restart the system.
7. Run the following commands in sequence and follow the on-screen tips 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 the `nvidia-smi` command to test whether the NVIDIA GRID driver is installed.
   If a command output similar to the following one is displayed, the NVIDIA GRID driver is installed:
9. Add a license server and activate the license.
   i. Run the `cd /etc/nvidia` command to go to the */etc/nvidia* directory.
   ii. Run the `cp gridd.conf.template gridd.conf` command to create a file named *gridd.conf*.
   iii. Add license server information to the *gridd.conf* file.

```
ServerAddress=<IP address of the license server>
ServerPort=<Port of the license server (Default port: 7070)>
FeatureType=2
EnableUI=TRUE
```

10. Run the following command to install x11vnc:

```
yum install x11vnc
```

11. Run the `lspci | grep NVIDIA` command to query the GPU BusID.In this example, the GPU BusID is `00:07.0`.
12. Configure the X Server environment.
    i. Run the `nvidia-xconfig --enable-all-gpus --separate-x-screens` command.
    ii. Add your GPU BusID to `Section "Device"` in the */etc/X11/xorg.conf* file. In this example, `BusID "PCI:0:7:0"` is added.
13. Run the `reboot` command to restart the system.

Test NVIDIA GRID drivers installed on Linux instances that run CentOS 7.3 64-bit

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

```
x11vnc -display :0
```

If a command output similar to the following one is displayed, x11vnc is started: In this case, 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 security group to which the instance belongs, to allow inbound traffic on TCP port 5900. For more information, see Add security group rules.
4. On the local machine, start VNC Viewer and enter `<Public IP address of the instance>:5900` to connect to the instance and go to KDE.
5. Run the `glxinfo` command to view the configurations supported by the current NVIDIA GRID driver.
   i. Start a new terminal session of the SSH client.
   ii. Run the `export DISPLAY=:0` command.
   iii. Run the `glxinfo -t` command to list the configurations supported by the current NVIDIA GRID driver.
6. Run the `glxgears` command to test the NVIDIA GRID driver.
   i. On KDE, right-click the desktop and select **Run Command**.
   ii. Run the `glxgears` command to start the testing application.
   If a window similar to the following one is displayed, the NVIDIA GRID driver functions properly:

4.6.6. Install NVIDIA GRID drivers on vgn6i or vgn5i Linux instances

You must install an NVIDIA GRID driver if your GPU-accelerated instances require Open Graphics Library (OpenGL). By default, the NVIDIA GRID license granted to NVIDIA GPUs is not activated. You must purchase and activate the license to use OpenGL. This topic describes how to install NVIDIA GRID drivers and activate the GRID license. vgn6i or vgn5i lightweight GPU-accelerated instances that are running the Ubuntu 16.04 64-bit operating system are used in the example.

Prerequisites

- A vgn6i or vgn5i instance that can access the Internet is created. We recommend that you select an image from **Public Image** when you create an instance.

  > **Note** This topic describes how to install GRID drivers on Linux instances. For Windows instances, you can select paid images with GRID drivers pre-installed when you create the instances. For more information, see Create an instance equipped with NVIDIA GPUs.

- A remote connection tool such as VNC Viewer is installed on your local machine.
- A GRID license is obtained from NVIDIA. You must build a license server. You can purchase an ECS instance and build a license server by following the tutorial on the official NVIDIA website.

Context

This topic describes how to install GRID drivers on vgn6i or vgn5i GPU-accelerated instances that are equipped with vGPUs. For information about how to install GRID drivers on GPU-accelerated instances that are not equipped with vGPUs, see Install NVIDIA GRID drivers on GPU-accelerated Linux instances.

Procedure

1. Disable nouveau.nouveau is an open source driver. It must be disabled before you can install another driver.
   i. Connect to the Linux instance. For more information, see Overview.
   ii. Check whether the *blacklist-nouveau.conf* file exists.

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

   iii. If the *blacklist-nouveau.conf* file exists and contains the following content, skip this step. If the file does not exist, run the `vim /etc/modprobe.d/blacklist-nouveau.conf` command to create the file. Then, add the following content to the file to disable nouveau:

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

   iv. Generate kernel initramfs.

   ```
   rmmod nouveau
   update-initramfs -u
   ```

   v. Restart the instance.

   ```
   reboot
   ```

2. Download the NVIDIA GRID driver package.
   i. Connect to the Linux instance. For more information, see Overview.
   ii. Download the NVIDIA GRID driver package.
   - vgn5i GRID Guest driver package:

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

   - vgn6i GRID Guest driver package:

     ```
     wget http://grid-9-2.oss-cn-hangzhou.aliyuncs.com/NVIDIA-Linux-x86_64-430.63-grid.run
     ```

3. Install the NVIDIA GRID driver.
   - vgn5i

   ```
   chmod +x NVIDIA-Linux-x86_64-418.70-grid.run
   . /NVIDIA-Linux-x86_64-418.70-grid.run
   ```

- vgn6i

```
chmod +x NVIDIA-Linux-x86_64-430.63-grid.run
. /NVIDIA-Linux-x86_64-430.63-grid.run
```

4. Test whether the NVIDIA GRID driver is installed.

```
nvidia-smi
```

If a command output similar to the following one is displayed, the NVIDIA GRID driver is installed:

5. Add a license server.
   i. Go to the */etc/nvidia* directory.

```
cd /etc/nvidia
```

   ii. Create a file named *gridd.conf*.

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

   iii. Add license server information to the *gridd.conf* file.

```
ServerAddress=<IP address of the license server>
ServerPort=<Port of the license server (Default port: 7070)>
FeatureType=1
```

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

```
reboot
```

7. Check whether the license is activated.
   i. Connect to the Linux instance. For more information, see Overview.
   ii. Check the license status.

```
systemctl status nvidia-gridd
```

If **License acquired successfully** is displayed, the license is activated.

4.6.7. Use the GPU cloud accelerator AIACC

Apsara AI Accelerator (AIACC) is an AI accelerator developed by Alibaba Cloud. It consists of a training accelerator (AIACC-Training) and an inference accelerator (AIACC-Inference), and can accelerate mainstream AI frameworks such as TensorFlow, PyTorch, MxNet, and Caffe. AIACC can automatically configure the Python conda environment that contains AIACC-Training.

Context

Conda is an open source cross-platform software package and environment management system. In the configured Python conda environment, you can install and switch the deep learning framework, and significantly improve the training performance by using AIACC-Training.

AIACC-Training has the following acceleration features:

- Uses the gradient fusion communication method to support adaptive multi-stream fusion and adaptive gradient fusion and improves the training performance of bandwidth-intensive network models by up to 300%.
- Uses the decentralized gradient-based negotiation mechanism to reduce the traffic of gradient-based negotiation on large-scale nodes by up to two orders of magnitude.
- Uses the hierarchical Allreduce algorithm to support FP16 gradient compression and mixed precision compression.
- Allows you to enable the NaN check during the training process and determine which gradient the NaN comes from when the graphics card architecture is SM60 or later.
- Provides API extensions for MXNet to support data parallelism and model parallelism of the InsightFace type.
- Provides deep optimization for RDMA networks.

Automatically install AIACC-Training

AIACC depends on the GPU driver, CUDA, and cuDNN. When you create a GPU-accelerated instance, select Auto-install GPU Driver and then select **GPU Cloud Accelerator**. After a GPU-accelerated instance is created, the Python conda environment that contains AIACC-Training V1.3.0 is configured based on the CUDA version that you selected. For more information about how to create a GPU-accelerated instance, see Create an instance equipped with NVIDIA GPUs.

The Python conda environment contains dependency packages such as AIACC-Training and OpenMPI, but does not contain deep learning frameworks. For more information about how to install a deep learning framework, see Install the deep learning framework.

The CUDA version determines the versions of supported deep learning frameworks. The following table describes the mappings.

CUDA version	Default conda environment	Version of the supported deep learning framework
CUDA 10.1	tf2.1_cu10.1_py36	Tensorflow 2.1
CUDA 10.0	tf1.15_tr1.4.0_mx1.5.0_cu10.0_py36	• Tensorflow 1.15 + Pytorch 1.4.0 + MXNet 1.5.0 • Tensorflow 1.14 + Pytorch 1.3.0 + MXNet 1.4.0
CUDA 9.0	tf1.12_tr1.3.0_mx1.5.0_cu9.0_py36	Tensorflow 1.12 + Pytorch 1.3.0 + MXNet 1.5.0

Install the deep learning framework

1. Connect to a Linux instance from the console.
2. View the automatically activated conda environment. View the version number before the username, as shown in the following figure.

   

   tf2.1_cu10.1_py36 indicates:

- Tensorflow 2.1
- CUDA 10.1
- Python 3.6

3. (Optional)If you do not need to use the automatically activated conda environment, activate a different conda environment.
   i. Run the following command to view all conda environments:

   ```
   conda env list
   ```

   The following figure shows an example command output.

   

   ii. Run the following command to activate the required conda environment:

   ```
   conda activate [version number]
   ```

   The following figure shows an example command output.

   

4. Run the following command to install the deep learning framework:

   ```
   install_frameworks.sh
   ```

   The following figure shows an example command output.

   

5. Test the demo. The TensorFlow demo is tested in this example.
   - For TensorFlow 2.1, perform the following operations:
     a. Decompress the demo test package.

     ```
     tar -xvf ali-perseus-demos.tgz
     ```

     b. Go to the directory of the TensorFlow demo.

     ```
     cd ali-perseus-demos/tensorflow2-examples
     ```

     c. Run the test script in the directory.

     Sample command:

     ```
     python tensorflow2_keras_mnist_perseus.py
     ```

     This demo uses the Modified National Institute of Standards and Technology (MNIST) dataset for training. This can ensure the same precision as your benchmark code while improving the training performance. The following figure shows an example training result.

     

   - For TensorFlow 1.14, perform the following operations:

a. Decompress the demo test package.

```
tar -xvf ali-perseus-demos.tgz
```

b. Go to the directory of the TensorFlow demo.

```
cd ali-perseus-demos/tensorflow-benchmarks
```

c. View the test command in *README.txt*.

d. Go to the directory where the test script of the corresponding version resides.

Sample command:

```
cd benchmarks-tf1.14
```

e. Modify and run the test command based on the number of GPUs that the specified instance type is equipped with.

Sample command:

```
mpirun --allow-run-as-root --bind-to none -np 1 -npernode 1 \
    --mca btl_tcp_if_include eth0 \
    --mca orte_keep_fqdn_hostnames t \
    -x NCCL_SOCKET_IFNAME=eth0 \
    -x LD_LIBRARY_PATH \
    ./config-fp16-tf.sh
```

This demo uses synthetic data for training to test the training speed. The following figure shows an example training result.

4.7. Compute optimized type family with FPGA

4.7.1. Compute optimized instance families with FPGAs

This topic describes the features of compute optimized instance families with Field Programmable Gate Arrays (FPGAs) and lists the instance types of each family.

- Recommended instance families

  f3, compute optimized instance family with FPGAs

- Other available instance families

  f1, compute optimized instance family with FPGAs

f3, compute optimized instance family with FPGAs

Features:

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Uses Xilinx 16nm Virtex UltraScale+ VU9P FPGAs.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Deep learning and inference
  - Genomics research
  - Database acceleration
  - Image transcoding such as conversion of JPEG images to WebP images
  - Real-time video processing such as H.265 video compression

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	FPGAs	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.f3-c4f1.xlarge	4	16.0	None	1 × Xilinx VU9P	1.5	300	No	2	3	10
ecs.f3-c8f1.2xlarge	8	32.0	None	1 × Xilinx VU9P	2.5	500	No	4	4	10
ecs.f3-c16f1.4xlarge	16	64.0	None	1 × Xilinx VU9P	5.0	1,000	No	4	8	20
ecs.f3-c16f1.8xlarge	32	128.0	None	2 × Xilinx VU9P	10.0	2,000	No	8	8	20
ecs.f3-c16f1.16xlarge	64	256.0	None	4 × Xilinx VU9P	20.0	2,500	No	16	8	20

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	FPGAs	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.f3-c22f1.22xlarge	88	336.0	None	4 × Xilinx VU9P	30.0	4,500	No	16	8	20

> **Note**
>
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

f1, compute optimized instance family with FPGAs

Features:

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Uses Intel® Arria® 10 GX 1150 FPGAs.
- Offers a CPU-to-memory ratio of 1:7.5.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Deep learning and inference
  - Genomics research
  - Financial analysis
  - Image transcoding
  - Computational workloads such as real-time video processing and security management

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	FPGAs	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.f1-c8f1.2xlarge	8	60.0	None	Intel Arria 10 GX 1150	3.0	400	Yes	4	4	10
ecs.f1-c8f1.4xlarge	16	120.0	None	2 × Intel Arria 10 GX 1150	5.0	1,000	Yes	4	8	20
ecs.f1-c28f1.7xlarge	28	112.0	None	Intel Arria 10 GX 1150	5.0	2,000	Yes	8	8	20
ecs.f1-c28f1.14xlarge	56	224.0	None	2 × Intel Arria 10 GX 1150	10.0	2,000	Yes	14	8	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

4.7.2. Create an f1 instance

This topic describes how to create an f1 instance.

Prerequisites

An image pre-installed with the Intel development environment is obtained. Currently, the image is available only as a shared image. To obtain the image, submit a ticket.

Context

This topic describes the parameters for creating an f1 instance. For more information about other common parameters, see Create an instance by using the provided wizard.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. Click **Create Instance**.
4. In the Basic Configurations step, configure the parameters and click **Next: Networking**.When you configure parameters, note the following items:
   - **Region**: Select a region from the following regions where f1 instances are available. Note that whether you can purchase a specific instance type is subject to the available instance types displayed on the buy page.
     - China (Hangzhou)
     - China (Shenzhen)
     - China (Beijing)

     You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.

     > **Note** If you can view subscription resources but not pay-as-you-go resources when you purchase an instance, access the FAQ, as instructed in Why are some instance types not available on the instance buy page when I attempt to purchase a pay-as-you-go instance?

   - **Instance Type**: Choose **Heterogeneous Computing > Compute Optimized Type with FPGA**.
   - **Image**: Click **Shared Image** and select an image that is pre-installed with the Intel development environment.

     > **Note** Currently, images pre-installed with the Intel development environment can be available only as shared images. The FaaS F1 basic image is pre-installed with quartus17.0, vcs2017.3, and dcp sdk. You can view the files in the *opt* directory.

5. In the Networking step, configure the parameters and click **Next: System Configurations**.Only the network type of VPC is supported.
6. In the System Configurations (Optional) step, configure the parameters and click **Next: Grouping**.
7. In the Grouping (Optional) step, configure the parameters and click **Next: Preview**.
8. Confirm the order and click **Create Order**.

What's next

After the f1 instance is created, you can connect to the instance and run the following command to check whether the license is configured. For more information about remote connections, see Overview.

```
echo $LM_LICENSE_FILE #Whether the variable is configured.
```

If the variable is configured, the actual value is displayed. Otherwise, no values are displayed.

Related information

References

- Use OpenCL on an f1 instance
- Use RTL Compiler on an f1 instance

4.7.3. Create an f3 instance

This topic describes how to create an f3 instance.

Context

This topic describes the parameters for creating an f3 instance. For more information about other common parameters, see Create an instance by using the provided wizard.

To facilitate testing, Alibaba Cloud provides an image pre-installed with the Xilinx development environment. Currently, the image is available only as a shared image. To obtain the image, submit a ticket.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. Click **Create Instance**.
4. In the Basic Configurations step, configure the parameters and click **Next: Networking**. When you configure parameters, note the following items:
   - **Region**: Select a region from the following regions where f3 instances are available. Note that whether you can purchase a specific instance type is subject to the available instance types displayed on the buy page.
     - China (Shanghai)
     - China (Beijing)
     - China (Zhangjiakou-Beijing Winter Olympic)

     You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.

     > **Note** If you can view subscription resources but not pay-as-you-go resources when you purchase an instance, access the FAQ, as instructed in Why are some instance types not available on the instance buy page when I attempt to purchase a pay-as-you-go instance?.

   - **Instance Type**: Choose **Heterogeneous Computing > Compute Optimized Type with FPGA**.
   - **Image**: Click **Shared Image** and select an image that is pre-installed with the Xilinx development environment.
   - **Storage**: Images that are pre-installed with the Xilinx development environment occupy some storage space. We recommend that you select an ultra disk of 200 GiB as the system disk.
5. In the Networking step, configure the parameters and click **Next: System Configurations**. Only the network type of VPC is supported.
6. In the System Configurations (Optional) step, configure the parameters and click **Next:**

Grouping.

7. In the Grouping (Optional) step, configure the parameters and click **Next: Preview**.
8. Confirm the order and click **Create Order**.

Related information

References

- Use OpenCL on an f3 instance
- Use the RTL design on an f3 instance

4.8. ECS bare metal instance type family

4.8.1. ECS Bare Metal Instances

This topic describes the features of ECS Bare Metal Instance families and lists the instance types of each family.

- Recommended instance families
  - General purpose instance families:
    - ebmg6a, general purpose ECS Bare Metal Instance family
    - ebmg6e, general purpose ECS Bare Metal Instance family with enhanced performance
    - ebmg6, general purpose ECS Bare Metal Instance family
  - Compute optimized instance families:
    - ebmc6a, compute optimized ECS Bare Metal Instance family
    - ebmc6e, compute optimized ECS Bare Metal Instance family with enhanced performance
    - ebmc6, compute optimized ECS Bare Metal Instance family
  - Memory optimized instance families:
    - ebmr6a, memory optimized ECS Bare Metal Instance family
    - ebmr6e, memory optimized ECS Bare Metal Instance family with enhanced performance
    - ebmr6, memory optimized ECS Bare Metal Instance family
    - ebmre6p, persistent memory optimized ECS Bare Metal Instance family with enhanced performance
    - ebmre6-6t, memory optimized ECS Bare Metal Instance family with enhanced performance
  - Instance families with high clock speed:
    - ebmhfg6, general purpose ECS Bare Metal Instance family with high clock speed
    - ebmhfc6, compute optimized ECS Bare Metal Instance family with high clock speed
    - ebmhfr6, memory optimized ECS Bare Metal Instance family with high clock speed
  - GPU-accelerated compute optimized instance families:
    - ebmgn6e, GPU-accelerated compute optimized ECS Bare Metal Instance family
    - ebmgn6v, GPU-accelerated compute optimized ECS Bare Metal Instance family
    - ebmgn6i, GPU-accelerated compute optimized ECS Bare Metal Instance family
- Other available instance families

- ebmg5s, general purpose ECS Bare Metal Instance family with enhanced network performance
- ebmg5, general purpose ECS Bare Metal Instance family
- ebmc5s, compute optimized ECS Bare Metal Instance family with enhanced network performance
- ebmc4, compute optimized ECS Bare Metal Instance family
- ebmr5s, memory optimized ECS Bare Metal Instance family with enhanced network performance
- ebmhfg5, ECS Bare Metal Instance family with high clock speed

Overview

ECS Bare Metal Instance is a compute service that combines the elasticity of virtual machines and the performance and features of physical machines. ECS Bare Metal Instance is designed based on the state-of-the-art virtualization 2.0 technology developed by Alibaba Cloud. The virtualization technology used by ECS Bare Metal Instance is optimized to support common ECS instances and nested virtualization. It maintains the elastic performance of ECS instances and the performance and features of physical machines.

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

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

When you use ECS Bare Metal Instance, take note of the following items:

- ECS Bare Metal Instance does not support instance type changes.
- When the physical machine that hosts an ECS bare metal instance fails, the system fails the instance over to another physical machine. Data is retained within the data disks of the instance.

Benefits

ECS Bare Metal Instance provides the following benefits based on technological innovations:

- Exclusive computing resources

  As a cloud-based elastic computing service, ECS Bare Metal Instance delivers the same performance and resource isolation as standard physical machines and can ensure the exclusivity of computing resources without virtualization overheads or feature loss. ECS Bare Metal Instance supports 8, 32, 80, 96, or 104 vCPUs and high clock speeds. An ECS bare metal instance with eight vCPUs can provide a core frequency of 3.7 GHz to 4.1 GHz for better performance and faster response for gaming and finance businesses than peer services.

- Chip-level security

  In addition to physical machine isolation, ECS Bare Metal Instance uses a chip-level trusted execution environment of Intel® SGX to ensure that encrypted data can be computed only in a secure and trusted environment. This chip-level hardware security protection provides a safe box for your data in the cloud and allows you to control all data encryption and key protection processes. For more information, see Install SGX.

- Compatible with multiple private clouds

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

- Support for heterogeneous instruction set processors

  Virtualization 2.0 used by ECS Bare Metal Instance is developed independently by Alibaba Cloud, and supports instruction set processors such as ARM at no additional cost.

Comparison of ECS bare metal instances, physical machines, and virtual machines

Compared with a physical machine with the same configurations, an ECS bare metal instance delivers greater performance. During Double 11, ECS bare metal instances delivered robust computing capabilities with millions of vCPUs to handle traffic spikes.

The following table describes a comparison among the features of ECS bare metal instances, physical machines, and virtual machines. In this table, Y means supported, N means not supported, and N/A means not applicable.

Feature type	Feature	ECS bare metal instance	Physical machine	Virtual machine
Automated O&M	Delivery within minutes	Y	N	Y
Compute	Zero performance loss	Y	Y	N
	Zero feature loss	Y	Y	N
	Zero resource preemption	Y	Y	N
	Compatible with ECS disks	Y	N	Y
	Startup from system disks	Y	N	Y
	Quick reset of system disks	Y	N	Y

Feature type	Feature	ECS bare metal instance	Physical machine	Virtual machine
Storage	Use of ECS images	Y	N	Y
	Cold migration between physical and virtual machines	Y	N	Y
	No need to install the operating system	Y	N	Y
	No need of local Redundant Arrays of Independent Disks (RAID), and better protection of data in disks	Y	N	Y
Networking	Compatible with VPCs	Y	N	Y
	Compatible with the classic network	Y	N	Y
	No communication bottlenecks between physical and virtual machine clusters in VPCs	Y	N	Y
Management	Compatible with existing ECS management systems	Y	N	Y
	Consistent user experiences on features such as VNC with that of virtual machines	Y	N	Y
	Out-of-band (OOB) network security	Y	N	N/A

ebmg6a, general purpose ECS Bare Metal Instance family

ebmg6a is in invitational preview. To use ebmg6a, submit a ticket.

Features

- Provides predictable and consistent ultra-high computing, storage, and network performance with the use of the fast path acceleration based on the third-generation X-Dragon architecture.
- Provides dedicated hardware resources and physical isolation.
- Equipped with 2.6 GHz AMD EPYC ™ ROME processors with a maximum turbo frequency of 3.3 GHz for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:4.
- I/O optimized.
- Supports enhanced SSDs (ESSDs), standard SSDs, and ultra disks.
- Provides ultra-high network performance with a packet forwarding rate of 24,000 Kpps.
- Supports VPCs only.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Video encoding, decoding, and rendering
  - Computing clusters and memory intensive data processing
  - Data analysis and computing

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	ENIs (including one primary ENI)	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.ebmg6a.64xlarge	256	1024.0	None	64.0	24,000	Yes	31	480	32.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmg6e, general purpose ECS Bare Metal Instance family with enhanced performance

Features

- Provides predictable and consistent ultra-high computing, storage, and network performance with the use of the fast path acceleration based on the third-generation X-Dragon architecture.
- Provides dedicated hardware resources and physical isolation.
- I/O optimized.
- Supports ESSDs.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel ® Xeon ® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz.
- Provides ultra-high network performance with a packet forwarding rate of 24,000 Kpps.
- Supports VPCs only.
- Suitable for the following scenarios:
  - Scenarios that have high security and regulatory requirements, such as deploying core database services
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Enterprise-level applications of various types and sizes
  - Websites and application servers
  - Game servers
  - Small and medium-sized database systems, caches, and search clusters
  - Data analysis and computing
  - Compute clusters and memory intensive data processing
  - High-performance scientific and engineering applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.ebmg6e.26xlarge	104	384.0	None	30.0	24,000	Yes	1,800	16	31	10	480	20.0

> **Note**
>
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

ebmg6, general purpose ECS Bare Metal Instance family

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:3.7.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz.
- Provides high network performance with a packet forwarding rate of 6,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Video encoding, decoding, and rendering
  - Enterprise-level applications such as large and medium-sized databases
  - Computing clusters and memory intensive data processing
  - Data analysis and computing

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmg6.26xlarge	104	384.0	None	30.0	6,000	Yes	8	32	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmc6a, compute optimized ECS Bare Metal Instance family

ebmc6a is in invitational preview. To use ebmc6a, submit a ticket.

Features

- Provides predictable and consistent ultra-high computing, storage, and network performance with the use of the fast path acceleration based on the third-generation X-Dragon architecture.
- Provides dedicated hardware resources and physical isolation.
- Equipped with 2.6 GHz AMD EPYC ™ ROME processors with a maximum turbo frequency of 3.3 GHz for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:2.
- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Provides ultra-high network performance with a packet forwarding rate of 24,000 Kpps.
- Supports VPCs only.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Video encoding, decoding, and rendering
  - Data analysis and computing

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	ENIs (including one primary ENI)	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.ebmc6a.64xlarge	256	512.0	None	64.0	24,000	Yes	31	480	32.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmc6e, compute optimized ECS Bare Metal Instance family with enhanced performance

Features

- Provides predictable and consistent ultra-high computing, storage, and network performance with the use of the fast path acceleration based on the third-generation X-Dragon architecture.
- Provides dedicated hardware resources and physical isolation.
- I/O optimized.
- Supports ESSDs.
- Offers a CPU-to-memory ratio of 1:2.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz.
- Provides ultra-high network performance with a packet forwarding rate of 24,000 Kpps.
- Supports VPCs only.
- Suitable for the following scenarios:
  - Scenarios that have high security and regulatory requirements, such as deploying core database services
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Web frontend servers
  - Frontend servers of massively multiplayer online (MMO) games

- Data analysis, batch processing, and video encoding
- High-performance scientific and engineering applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.ebmc6e.26xlarge	104	192.0	None	30.0	24,000	Yes	1,800	16	31	10	480	20.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmc6, compute optimized ECS Bare Metal Instance family

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:1.8.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz.
- Provides high network performance with a packet forwarding rate of 6,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Video encoding, decoding, and rendering

- Frontend servers of MMO games
- High-performance scientific and engineering applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmc6.26xlarge	104	192.0	None	30.0	6,000	Yes	8	32	10

> **Note**
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

ebmr6a, memory optimized ECS Bare Metal Instance family

ebmr6a is in invitational preview. To use ebmr6a, submit a ticket.

Features

- Provides predictable and consistent ultra-high computing, storage, and network performance with the use of the fast path acceleration based on the third-generation X-Dragon architecture.
- Provides dedicated hardware resources and physical isolation.
- Equipped with 2.6 GHz AMD EPYC ™ ROME processors with a maximum turbo frequency of 3.3 GHz for consistent computing performance.
- Offers a CPU-to-memory ratio of 1:8.
- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Provides ultra-high network performance with a packet forwarding rate of 24,000 Kpps.
- Supports VPCs only.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - In-memory databases
  - Data analysis, data mining, and distributed memory caching

- Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	ENIs (including one primary ENI)	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.ebmr6a.64xlarge	256	2048.0	None	64.0	24,000	Yes	31	480	32.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmr6e, memory optimized ECS Bare Metal Instance family with enhanced performance

Features

- Provides predictable and consistent ultra-high computing, storage, and network performance with the use of the fast path acceleration based on the third-generation X-Dragon architecture.
- Provides dedicated hardware resources and physical isolation.
- I/O optimized.
- Supports ESSDs.
- Offers a CPU-to-memory ratio of 1:8.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz.
- Provides ultra-high network performance with a packet forwarding rate of 24,000 Kpps.
- Supports VPCs only.
- Suitable for the following scenarios:
  - Scenarios that have high security and regulatory requirements, such as deploying core database services
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch

- Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
- High-performance databases and in-memory databases
- Data analysis, data mining, and distributed memory caching
- Hadoop clusters, Spark clusters, and other memory intensive enterprise applications
- High-performance scientific and engineering applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	Connections (K)	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI	Disk IOPS (K)	Disk bandwidth (Gbit/s)
ecs.ebmr6e.26xlarge	104	768.0	None	30.0	24,000	Yes	1,800	16	31	10	480	20.0

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmr6, memory optimized ECS Bare Metal Instance family

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:7.4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz.
- Provides high network performance with a packet forwarding rate of 6,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack

- Containers including Docker, Clear Containers, and Pouch
- High-performance databases and in-memory databases
- Data analysis, data mining, and distributed memory caching
- Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmr6.26xlarge	104	768.0	None	30.0	6,000	Yes	8	32	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmre6p, persistent memory optimized ECS Bare Metal Instance family with enhanced performance

To use ebmre6p, submit a ticket.

Features

- Uses Intel® Optane™ non-volatile memory.
- Cost-effective due to end-to-end optimization for ApsaraDB for Redis scenarios.
- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz for consistent computing performance.
- Supports a maximum of 1,920 GiB memory (384 GiB DRAM memory + 1,536 GiB Intel® Optane™ non-volatile memory), offers a CPU-to-memory ratio of about 1:20, and can meet the needs of memory intensive applications.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - In-memory databases such as ApsaraDB for Redis
  - High-performance databases such as SAP HANA
  - Other memory intensive applications such as AI applications and smart search applications

Instance types

Instance type	vCPUs	DRAM (GiB)	AEP memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmre6p.26xlarge	104	384.0	1536.0	None	30.0	6,000	Yes	16	31	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmre6-6t, memory optimized ECS Bare Metal Instance family with enhanced performance

To use ebmre6-6t, submit a ticket.

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:30.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8269 (Cascade Lake) processors with a maximum turbo frequency of 3.2 GHz.
- Provides high network performance with a packet forwarding rate of 6,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - High-performance databases and in-memory databases such as SAP HANA
  - Memory intensive applications
  - Big data processing engines such as Apache Spark and Presto

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmre6-6t.52xlarge	208	6144.0	None	30.0	6,000	Yes	8	32	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmhfg6, general purpose ECS Bare Metal Instance family with high clock speed

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:4.8.
- Equipped with 3.1 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.5 GHz.
- Provides high network performance with a packet forwarding rate of 6,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Enterprise-level applications such as large and medium-sized databases
  - Video encoding, decoding, and rendering

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmhfg6.20xlarge	80	384.0	None	30.0	6,000	Yes	8	32	10

> **Note**
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

ebmhfc6, compute optimized ECS Bare Metal Instance family with high clock speed

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:2.4.
- Equipped with 3.1 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.5 GHz.
- Provides high network performance with a packet forwarding rate of 6,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Video encoding, decoding, and rendering

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmhfc6.20xlarge	80	192.0	None	30.0	6,000	Yes	8	32	10

> **Note**
>
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

ebmhfr6, memory optimized ECS Bare Metal Instance family with high clock speed

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:9.6.
- Equipped with 3.1 GHz Intel® Xeon® Platinum 8269CY (Cascade Lake) processors with a maximum turbo frequency of 3.5 GHz.
- Provides high network performance with a packet forwarding rate of 6,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - High-performance databases and in-memory databases
  - Data analysis, data mining, and distributed memory caching
  - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmhfr6.20xlarge	80	768.0	None	30.0	6,000	Yes	8	32	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmgn6e, GPU-accelerated compute optimized ECS Bare Metal Instance family

ebmgn6e is in invitational preview. To use ebmgn6e, submit a ticket.

Features

- Provides flexible and powerful software-defined compute based on the X-Dragon architecture.
- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Uses NVIDIA V100 (32 GB NVLink) GPU processors.
- Offers a CPU-to-memory ratio of 1:8.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Uses NVIDIA V100 GPU computing accelerators (SXM2-based).
  - Powered by the new NVIDIA Volta architecture.
  - Equipped with 32 GB HBM2 GPU memory (900 GB/s bandwidth) per GPU.
  - Equipped with 5,120 CUDA cores per GPU.
  - Equipped with 640 Tensor cores per GPU.
  - Supports up to six NVLink connections for a total bandwidth of 300 GB/s (25 GB/s per connection).
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Deep learning applications such as training and inference applications of AI algorithms used in image classification, autonomous vehicles, and speech recognition
  - Scientific computing applications such as computational fluid dynamics, computational finance, molecular dynamics, and environmental analysis

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmgn6e.24xlarge	96	768.0	None	V100*8	256	32.0	4,800	Yes	16	15	20

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmgn6v, GPU-accelerated compute optimized ECS Bare Metal Instance family

Features

- Provides flexible and powerful software-defined compute based on the X-Dragon architecture.
- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Uses NVIDIA V100 GPU processors.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Uses NVIDIA V100 GPU computing accelerators (SXM2-based).
  - Powered by the new NVIDIA Volta architecture.
  - Equipped with 16 GB HBM2 GPU memory (900 GB/s bandwidth) per GPU.
  - Equipped with 5,120 CUDA cores per GPU.
  - Equipped with 640 Tensor cores per GPU.
  - Supports up to six NVLink connections for a total bandwidth of 300 GB/s (25 GB/s per connection).
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:
  - Deep learning applications such as training and inference applications of AI algorithms used in image classification, autonomous vehicles, and speech recognition

- Scientific computing applications such as computational fluid dynamics, computational finance, molecular dynamics, and environmental analysis

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmgn6v.24xlarge	96	384.0	None	V100*8	128	30.0	4,500	Yes	8	32	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmgn6i, GPU-accelerated compute optimized ECS Bare Metal Instance family

Features

- Provides flexible and powerful software-defined compute based on the X-Dragon architecture.
- I/O optimized.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Supports ESSDs that deliver millions of IOPS, standard SSDs, and ultra disks.
- Uses NVIDIA T4 GPU computing accelerators.
  - Powered by the new NVIDIA Turing architecture.
  - Equipped with 16 GB memory (320 GB/s bandwidth) per GPU.
  - Equipped with 2,560 CUDA cores per GPU.
  - Euipped with up to 320 Turing Tensor cores per GPU.
  - Mixed-precision Tensor cores support 65 FP16 TFLOPS, 130 INT8 TOPS, and 260 INT4 TOPS.
- Provides high network performance based on large computing capacity.
- Suitable for the following scenarios:

- AI (deep learning and machine learning) inference for computer vision, speech recognition, speech synthesis, natural language processing (NLP), machine translation, and recommendation systems
- Real-time rendering for cloud gaming
- Real-time rendering for AR and VR applications
- Graphics workstations or overloaded graphics computing
- GPU-accelerated databases
- High-performance computing

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmgn6i.24xlarge	96	384.0	None	T4*4	64	30.0	4,500	Yes	8	32	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

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

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors with a maximum turbo frequency of 2.7 GHz.
- Provides high network performance with a packet forwarding rate of 4,500 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:

- Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
- Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
- Containers including Docker, Clear Containers, and Pouch
- Enterprise-level applications such as large and medium-sized databases
- Video encoding

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmg5s.24xlarge	96	384.0	None	30.0	4,500	No	8	32	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

ebmg5, general purpose ECS Bare Metal Instance family

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors with a maximum turbo frequency of 2.7 GHz.
- Provides high network performance with a packet forwarding rate of 4,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Enterprise-level applications such as large and medium-sized databases

- Video encoding

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmg5.24xlarge	96	384.0	None	10.0	4,000	No	8	32	10

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

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

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:2.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors with a maximum turbo frequency of 2.7 GHz.
- Provides high network performance with a packet forwarding rate of 4,500 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Scenarios where large volumes of packets are received and transmitted, such as on-screen video comments and telecom data forwarding
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Video encoding, decoding, and rendering

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmc5s.24xlarge	96	192.0	None	30.0	4,500	No	8	32	10

> **Note**
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

ebmc4, compute optimized ECS Bare Metal Instance family

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Offers a CPU-to-memory ratio of 1:2.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors with a maximum turbo frequency of 2.9 GHz.
- Provides high network performance with a packet forwarding rate of 4,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - Enterprise-level applications such as large and medium-sized databases
  - Video encoding

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmc4.8xlarge	32	64.0	None	10.0	4,000	No	8	12	10

> **Note**
>
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

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

Features

- I/O optimized.
- Supports ESSDs, standard SSDs, and ultra disks.
- Offers a CPU-to-memory ratio of 1:8.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors with a maximum turbo frequency of 2.7 GHz.
- Provides high network performance with a packet forwarding rate of 4,500 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Third-party virtualization including Xen and KVM, and AnyStack including OpenStack and ZStack
  - Containers including Docker, Clear Containers, and Pouch
  - High-performance databases and in-memory databases
  - Data analysis, data mining, and distributed memory caching
  - Hadoop clusters, Spark clusters, and other memory intensive enterprise applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmr5s.24xlarge	96	768.0	None	30.0	4,500	No	8	32	10

> **Note**
>
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

ebmhfg5, ECS Bare Metal Instance family with high clock speed

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 3.7 GHz Intel® Xeon® E3-1240v6 (Skylake) processors with a maximum turbo frequency of 4.1 GHz.
- Provides high network performance with a packet forwarding rate of 2,000 Kpps.
- Supports VPCs only.
- Provides dedicated hardware resources and physical isolation.
- Disabled failover by default.

  You can call the ModifyInstanceMaintenanceAttributes operation to modify the maintenance action. Set ActionOnMaintenance to *AutoRedeploy* to enable failover.
- Supports Intel® SGX.
- Suitable for the following scenarios:
  - Workloads that require direct access to physical resources or scenarios that require a license to be bound to the hardware
  - Gaming and finance applications that require high performance
  - High-performance web servers
  - Enterprise-level applications such as high-performance databases

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ebmhfg5.2xlarge	8	32.0	None	6.0	2,000	No	8	6	8

Note

- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

Billing methods

ECS Bare Metal Instance supports pay-as-you-go and subscription billing methods. For more information, see Comparison of billing methods.

4.8.2. Install SGX

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

Prerequisites

The instance is of the ebmhfg5 instance family.

Context

Intel SGX is an architecture extension developed by Intel. SGX protects selected code and data from malicious disclosure or modification by using enclaves, which are encrypted areas of execution in memory.

SGX sets aside one or more ranges of physical memory as Enclave Page Caches (EPCs) and uses the Memory Encryption Engine (MEE) to encrypt data stored within the EPCs. Data stored within EPCs is decrypted only inside the CPU. SGX offers CPU-based security controls. Data remains protected even if the operating system, Virtual Machine Manager (VMM), or Basic Input/Output System (BIOS) becomes compromised.

You can encrypt sensitive data, pass the encrypted data to an enclave in the cloud, and provide the corresponding key to the enclave by using remote attestation. You can then compute over the encrypted data protected by the CPU and have an encrypted result returned. With this feature, you can make use of the powerful cloud computing capabilities with low risks of data disclosure.

Enclave Definition Language (EDL) is the fundamental part of SGX. It defines all enclave interface functions. During the compilation process, the Edger8r tool generates trusted and untrusted proxy/bridge functions based on the functions defined in EDL and performs security checks.

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

- ECALL: A call made 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 illustration purposes only.
    // It should not be used in a real enclave,
    // unless it is during the development phase
    // for debugging purposes.
    void dump_secret([in] const secret_t* secret);
    };
};
```

Install SGX by using the installer file

Before you use the installer file to install SGX, you must first install corresponding Linux kernel header files. The installer file includes the SGX driver, Platform Software (PSW), and Software Development Kit (SDK).

> **Note** The default installation directory of Makefile is */opt/intel/* in this example.

1. Download the applicable SGX installer file. Click here to download the installer file.
2. Install SGX. For more information about how to install SGX, visit Installation Guide.

Install SGX by using the source code

Before you use the source code to install SGX, you must first install corresponding Linux kernel header files. The installer file includes the SGX driver, PSW, and SDK.

> **Note** The default installation directory of Makefile is */opt/intel/* in this example.

1. Download the source code from GitHub.

2. Compile the source code by performing the operations described in the *README.md* file.

4.9. Super Computing Cluster instance type family

4.9.1. Overview

This topic describes the features of Super Computing Cluster (SCC) instance families and lists the instance types of each family.

Alibaba Cloud ecs server elastic computing

- scchfc6, compute optimized SCC instance family with high clock speed
- scchfg6, general purpose SCC instance family with high clock speed
- scchfr6, memory optimized SCC instance family with high clock speed
- scch5, SCC instance family with high clock speed
- sccg5, general purpose SCC instance family
- sccgn6, compute optimized SCC instance family with GPU capabilities

Introduction

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

SCCs are used to meet the demands of applications such as high performance computing, artificial intelligence, machine learning, scientific and engineering computing, data analysis, and audio and video processing. In the clusters, nodes are connected over RDMA networks that feature high bandwidth and low latency. This ensures the parallel efficiency of applications in areas such as high performance computing, artificial intelligence, and machine learning. The RDMA over Converged Ethernet (RoCE) rivals an InfiniBand network in terms of connection speed and can support more Ethernet-based applications.

The combination of SCCs and other Alibaba Cloud computing services such as ECS and Elastic GPU Service provides E-HPC with the ultimate high performance parallel computing resources, making supercomputing on the cloud possible.

Comparison of SCCs, physical machines, and virtual machines

The following table compares the features of SCCs, physical machines, and virtual machines. In this table, Y means supported, N means not supported, and N/A means not applicable.

Feature type	Feature	SCC	Physical machine	Virtual machine
Automated O&M	Delivery within minutes	Y	N	Y
	Zero performance loss	Y	Y	N

Feature type	Feature	SCC	Physical machine	Virtual machine
Compute	Zero feature loss	Y	Y	N
	Zero resource preemption	Y	Y	N
Storage	Compatible with ECS disks	Y	N	Y
	Startup from system disks	Y	N	Y
	Quick reset of system disks	Y	N	Y
	Use of ECS images	Y	N	Y
	Cold migration between physical and virtual machines	Y	N	Y
	No need to install the operating system	Y	N	Y
	No need of local Redundant Arrays of Independent Disks (RAID), and better protection of data in disks	Y	N	Y
Network	Compatible with ECS VPCs	Y	N	Y
	Compatible with the ECS classic network	Y	N	Y
	No communication bottlenecks between physical and virtual machine clusters in VPCs	Y	N	Y

Feature type	Feature	SCC	Physical machine	Virtual machine
Management	Compatible with existing ECS management systems	Y	N	Y
	Consistent user experience on features such as VNC with that on virtual machines	Y	N	Y
	Out-of-band (OOB) network security	Y	N	N/A

scchfc6, compute optimized SCC instance family with high clock speed

To use this instance family, submit a ticket.

Features

- I/O optimized.
- Supports enhanced SSDs (ESSDs), standard SSDs, and ultra disks.
- Supports both RoCE and VPCs, of which RoCE is dedicated to RDMA communication.
- Provides all the features of ECS Bare Metal Instance.
- Equipped with 3.1 GHz Intel® Xeon® Platinum 8269 (Cascade Lake) processors with a maximum turbo frequency of 3.5 GHz.
- Offers a CPU-to-memory ratio of 1:2.4.
- Scenarios:
  - Large-scale machine learning training
  - Large-scale high performance scientific computing and simulations
  - Large-scale data analysis, batch processing, and video encoding

Instance types

Instance type	vCPU	Physical cores	Memory (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	RoCE (Gbit/s)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.scchfc6.20xlarge	80	40	192.0	None	30.0	6,000	50	Yes	8	32	10

Note

- ecs.scchfc6.20xlarge provides 80 logical processors on 40 physical cores.
- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

scchfg6, general purpose SCC instance family with high clock speed

To use this instance family, submit a ticket.

Features

- I/O optimized.
- Supports enhanced SSDs, standard SSDs, and ultra disks.
- Supports both RoCE and VPCs, of which RoCE is dedicated to RDMA communication.
- Provides all the features of ECS Bare Metal Instance.
- Equipped with 3.1 GHz Intel® Xeon® Platinum 8269 (Cascade Lake) processors with a maximum turbo frequency of 3.5 GHz.
- Offers a CPU-to-memory ratio of 1:4.8.
- Scenarios:
  - Large-scale machine learning training
  - Large-scale high performance scientific computing and simulations
  - Large-scale data analysis, batch processing, and video encoding

Instance types

Instance type	vCPU	Physical cores	Memory (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	RoCE (Gbit/s)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.scchfg6.20xlarge	80	40	384.0	None	30.0	6,000	50	Yes	8	32	10

Note

- ecs.scchfg6.20xlarge provides 80 logical processors on 40 physical cores.
- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

scchfr6, memory optimized SCC instance family with high clock speed

To use this instance family, submit a ticket.

Features

- I/O optimized.
- Supports enhanced SSDs, standard SSDs, and ultra disks.
- Supports both RoCE and VPCs, of which RoCE is dedicated to RDMA communication.
- Provides all the features of ECS Bare Metal Instance.
- Equipped with 3.1 GHz Intel® Xeon® Platinum 8269 (Cascade Lake) processors with a maximum turbo frequency of 3.5 GHz.
- Offers a CPU-to-memory ratio of 1:9.6.
- Scenarios:
  - Large-scale machine learning training
  - Large-scale high performance scientific computing and simulations
  - Large-scale data analysis, batch processing, and video encoding

Instance types

Instance type	vCPU	Physical cores	Memory (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	RoCE (Gbit/s)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.scchfr6.20xlarge	80	40	768.0	None	30.0	6,000	50	Yes	8	32	10

Note

- ecs.scchfr6.20xlarge provides 80 logical processors on 40 physical cores.
- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

scch5, SCC instance family with high clock speed

Features

- I/O optimized.
- Supports only standard SSDs and ultra disks.
- Supports both RoCE and VPCs, of which RoCE is dedicated to RDMA communication.
- Provides all the features of ECS Bare Metal Instance.
- Equipped with 3.1 GHz Intel® Xeon® Gold 6149 (Skylake) processors.
- Offers a CPU-to-memory ratio of 1:3.
- Scenarios:
  - Large-scale machine learning training
  - Large-scale high performance scientific computing and simulations
  - Large-scale data analysis, batch processing, and video encoding

Instance types

Instance type	vCPU	Physical cores	Memory (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	RoCE (Gbit/s)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.scch5.16xlarge	64	32	192.0	None	10.0	4,500	25 × 2	No	8	32	10

Note

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

sccg5, general purpose SCC instance family

Features

- I/O optimized.
- Supports only standard SSDs and ultra disks.
- Supports both RoCE and VPCs, of which RoCE is dedicated to RDMA communication.
- Provides all the features of ECS Bare Metal Instance.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Offers a CPU-to-memory ratio of 1:4.
- Scenarios:
  - Large-scale machine learning training
  - Large-scale high performance scientific computing and simulations
  - Large-scale data analysis, batch processing, and video encoding

Instance types

Instance type	vCPU	Physical cores	Memory (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	RoCE (Gbit/s)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.sccg5.24xlarge	96	48	384.0	None	10.0	4,500	25 × 2	No	8	32	10

Note

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

sccgn6, compute optimized SCC instance family with GPU capabilities

Features

- I/O optimized.
- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel® Xeon® Platinum 8163 (Skylake) processors.
- Provides all the features of ECS Bare Metal Instance.
- Storage:
  - Supports enhanced SSDs, standard SSDs, and ultra disks
  - Supports a high performance CPFS

- Networking:
  - Supports VPCs
  - Supports the RoCE v2 network, which is dedicated to low-latency RDMA communication
- Uses NVIDIA V100 GPU processors that have the SXM2 module:
  - Powered by the new NVIDIA Volta architecture
  - Offers a 16 GB HBM2 GPU memory
  - CUDA Cores 5120
  - Tensor Cores 640
  - Offers a GPU memory bandwidth of up to 900 GB/s
  - Supports up to six NVLink connections and total bandwidth of 300 GB/s (25 GB/s per connection)
- Scenarios:
  - Ultra-large-scale machine learning training on a distributed GPU cluster
  - Large-scale high performance scientific computing and simulations
  - Large-scale data analysis, batch processing, and video encoding

Instance types

Instance type	vCPU	Memory (GiB)	Local storage (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	RoCE (Gbit/s)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.sccgn6.24xlarge	96	384.0	None	V100*8	30	4,500	25 × 2	Yes	8	32	10

> **Note**
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

Billing methods

SCCs support pay-as-you-go and subscription billing methods. For more information, see Comparison of billing methods.

4.9.2. Create an SCC instance

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

Context

This topic describes the configuration items you must pay attention to when you create an SCC instance. For more information about other general configuration items, see Create an instance by using the provided wizard.

To use remote direct memory access (RDMA), HPC schedulers, and cluster scaling services, log on to the E-HPC console, create an SCC cluster, and create an SCC instance. For more information, see Create an E-HPC cluster .

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images** > **Instances**.
3. Click **Create Instance**.
4. Complete basic configurations. Click **Next: Networking**. When you configure parameters, note that:
   - **Region**: Select the region and zone in which SCC instance families are available based on the Regions and zones table. Note that the purchase page displays the latest region and zone information, which may differ from information provided in this topic.

     > **Note** If your business needs to be deployed in a region or zone that is not included in the table, submit a ticket for special deployment.

Instance family	Regions and zones
scch5	■ China (Hangzhou) Zone H ■ China (Shanghai) Zone D and Zone B ■ China (Qingdao) Zone C ■ China (Zhangjiakou-Beijing Winter Olympics) Zone A
sccg5	■ China (Hangzhou) Zone H ■ China (Shanghai) Zone B ■ China (Qingdao) Zone C
sccgn6	■ China (Shanghai) Zone G ■ China (Zhangjiakou-Beijing Winter Olympics) Zone A

     You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.

> **Note** If you can view subscription resources but not pay-as-you-go resources when you purchase an instance, access the FAQ, as instructed in Why are some instance types not available on the instance buy page when I attempt to purchase a pay-as-you-go instance?.

- **Instance Type**: Select **Super Computing Cluster**. Currently, CPU-based instance families scch5 and sccg5 and GPU-based instance family sccgn6 are available. After selecting the SCC instance type, you must specify the SCC name and description as required.
- **Image**: Select **Public Image**. Currently, only the custom Linux CentOS 7.5 image for SCC is supported.

  > **Note** The customized image supports the RDMA over Converged Ethernet (RoCE)-compliant driver and OpenFabrics Enterprise Distribution (OFED) stack. You can use SCC RDMA through the IB Verbs API and manage RDMA communication through the MPI.

- **Storage**: A maximum of 16 data disks can be attached. You can add a data disk when you create an SCC instance or attach an instance after the instance is created. For more information, see Create a disk and Attach a data disk.

5. Complete networking configurations. Click **Next: System Configurations (Optional)**. Only the network type of VPC is supported.
6. Complete system configurations. Click **Next: Grouping (Optional)**.
7. Complete grouping configurations. Click **Next: Preview**.
8. Confirm the order. Click **Create Order**.

4.10. Burstable performance instances

4.10.1. Overview

Burstable instances are an economical instance type that is intended to cope with burstable performance requirements in entry-level computing scenarios. This topic describes the features, application scenarios, specific instance families and instance types, baseline performance, CPU credits, and performance modes of burstable instances.

What are burstable instances?

Burstable instances use CPU credits to ensure computing performance and are suited for scenarios where CPU usage is typically low but bursts in CPU usage occur on occasion. You can accumulate CPU credits that can be used to increase the computing performance of burstable instances when required by your workloads. This consumption pattern does not affect the environments or applications that are running on your instances. Burstable instances are more flexible and less expensive than other types of instances in CPU usage.

The CPU credit mechanism allows you to minimize the consumption of resources during off-peak hours, and scale resources out during peak hours at no extra cost. If you have unplanned performance requirements, you can enable the unlimited mode for your burstable instances.

Burstable instances are of the following instance families:

- t6, burstable instance family

- t5, burstable instance family

> **Note** Burstable instances are special shared instances. For information about other shared instance families, see Shared instance families.

The following table describes the baseline performance, CPU credits, and performance modes of burstable instances.

Term	Description	Reference
baseline performance	The amount of vCPU capacity that is continuously provisioned to a burstable instance. Baseline performance varies based on the instance type.	Baseline performance
initial CPU credit	The CPU credits that are provisioned when you create a burstable instance. You can earn 30 initial credits for each vCPU. These credits cannot be replenished after they are used up.	CPU credits
CPU credit balance	The net credits that are accrued when the earned CPU credits exceed the consumed credits. You can use these credits to run instances above their baseline performance.	CPU credits
max CPU credit balance	The maximum amount of CPU credits that can be earned by a burstable instance within a 24 hour period. Your CPU credit balance is valid for 24 hours to ensure the availability of CPU credits. A specific instance type earns CPU credits at a fixed rate. Therefore, its CPU credit balance is limited.	CPU credits
performance mode	A burstable instance can run in standard or unlimited mode. • A burstable instance in standard mode runs below its baseline performance if its CPU credits are depleted. • A burstable instance in unlimited mode allows you to overdraw or pay for additional CPU credits and utilize the CPU above its baseline performance at any time. In this case, you may be charged for using these additional CPU credits.	Performance mode
advance CPU credit	The CPU credits that you can pay for in advance and can receive within 24 hours after you make the payment. You can use advance CPU credits only when you enable the unlimited mode.	Performance mode
overdrawn CPU credit	The CPU credits that you use when you have consumed all advance CPU credits to ensure that the instance is running above the baseline performance. You are charged for using these credits. You can use overdrawn CPU credits only when you enable the unlimited mode.	Performance mode

Scenarios

When you purchase an enterprise-level instance, its vCPUs are reserved for you exclusively. In this case, you are charged for the vCPU resources regardless of whether you fully utilize the performance of the vCPUs. Even if you require high levels of computing power only for a specific period during a day, you are charged for the remainder of the entire day. To avoid this situation, you can use burstable instances to better meet your business requirements.

Burstable instances apply to scenarios where you require higher-than-normal performance for a specific period, such as stress testing service applications, lightweight applications, microservices, and web application servers. We recommend that you evaluate your business requirements to determine the performance levels required during off-peak and peak hours before you make a purchase. The baseline performance of the instances that you purchase must meet your business requirements during off-peak hours. This way, you can enjoy the required performance at significantly lower costs.

> **Note** If the purchased burstable instances cannot meet your requirements, you can change the configurations. For more information, see Change configurations.

Baseline performance

Baseline performance is the amount of vCPU capacity that is continuously provisioned to a burstable instance. Baseline performance varies based on the instance type. If the vCPUs of your burstable instances are running at the baseline performance, the CPU credits that you earn are equal to the CPU credits that you consume. The *Baseline CPU computing performance* column in the instance type table indicates the baseline performance of a single vCPU.

CPU credits

CPU credits can be described as computing resources that are available. These computing resources determine the computing performance of your burstable instances. The following concepts are essential to understand CPU credits:

- Initial CPU credits

  When you create a burstable instance, 30 CPU credits are provisioned for each vCPU of the instance, which are *initial CPU credits*. These credits enable you to complete deployment tasks after you start the instance.

  For example, an ecs.t5-lc1m2.large instance has two vCPUs. You earn 60 initial CPU credits when you create an instance of this instance type. An ecs.t5-c1m1.xlarge instance has four vCPUs. You earn 120 initial CPU credits when you create an instance of this instance type.

- Rate of earning CPU credits

  When a burstable instance is started, it starts to consume CPU credits to maintain its computing performance. At the same time, it also earns CPU credits at a fixed rate that is determined by the instance type. The amount of CPU credits that a vCPU can earn per hour is based on its baseline performance. The *CPU credits per hour* column in the instance type tables indicates the CPU credits that all of the vCPUs of an instance can earn per hour.

  For example, 25% baseline performance of an ecs.t5-c1m1.large instance indicates that the CPU credits that a vCPU of the instance earns per hour can keep the vCPU running at 25% utilization for an hour or at 100% utilization for 15 minutes (60 × 25%). In response to its baseline performance, each vCPU earns 15 CPU credits per hour. Therefore, an ecs.t5-c1m1.large instance that has two vCPUs earns 30 CPU credits per hour.

- CPU credit balance

  If the CPU credits earned exceed the credits consumed, the net credits are accrued as *CPU credit balance*. Your CPU credit balance is valid for 24 hours to ensure the availability of CPU credits. A specific instance type earns CPU credits at a fixed rate. Therefore, its CPU credit balance is limited. The maximum CPU credit balance of a specific instance type is the number of CPU credits that the instance can earn within 24 hours. For more information, see the *Max CPU credit balance* column in the instance type table.

  For example, an ecs.t5-c1m1.large instance can earn 30 CPU credits per hour. The maximum CPU credit balance that the instance can earn is 720 (30 × 24) credits.

- Rate of consuming CPU credits

  The rate of CPU credit consumption of a burstable instance is based on the number of vCPUs, CPU utilization, and operating hours. For example, you consume one CPU credit in the following scenarios:

  - One vCPU runs at 100% utilization for a minute.
  - One vCPU runs at 50% utilization for two minutes.
  - Two vCPUs run at 25% utilization for two minutes.

  When a burstable instance is started, it starts to consume CPU credits to maintain its computing performance. Initial credits that cannot be replenished are consumed first. When the initial credits are used up, the instance consumes the accumulated CPU credits.

  - When your vCPUs run below the baseline performance, the earned credits are greater than the consumed credits. In this case, the CPU credit balance increases.
  - When your vCPUs run at the baseline performance, the earned credits are equal to the consumed credits. In this case, the CPU credit balance remains unchanged.
  - When your vCPUs run above the baseline performance, the earned credits are less than the consumed credits. In this case, the CPU credit balance decreases.

In different scenarios, the operation of stopping your instances may have different impacts on your CPU credits:

- If a pay-as-you-go instance is stopped while the No fees for Stopped Instances (VPC-Connected) feature is disabled, the current CPU credit balance of the instance is retained and the instance continues to earn CPU credits.
- If a pay-as-you-go instance is stopped while the No fees for Stopped Instances (VPC-Connected) feature is enabled, the current CPU credit balance of the instance becomes invalid and the instance cannot continue to earn credits. When you restart the instance, the instance receives initial credits and starts to earn credits again.
- If a pay-as-you-go instance is stopped due to overdue payments, its current CPU credit balance is retained, but the instance cannot continue to earn credits until you complete the payment.
- If a subscription instance expires and stops, its current CPU credit balance is retained, but the instance cannot continue to earn credits. When the instance is restarted, it starts to earn credits again.

Performance mode

A burstable instance can run in either standard or unlimited mode.

- Standard mode

The performance of a burstable instance in standard mode is based on the availability of CPU credits. After the instance consumes all of its initial credits and accrued CPU credits, the instance cannot run above its baseline performance. When the CPU credit balance is low, the instance gradually reduces performance to its baseline performance within 15 minutes. This way, the instance does not experience a sharp performance drop-off when its accrued CPU credit balance is depleted.

The standard mode applies to scenarios such as lightweight web servers, development and testing environments, and databases with low and medium performance. In these scenarios, you have stable workloads, do not need instances to run above the baseline performance for an extended period of time, and may occasionally need burst performance.

- Unlimited mode

  The performance of a burstable instance in unlimited mode is not limited by the availability of CPU credits. You can overdraw or pay for additional CPU credits to obtain performance boosts at any time. If your instances continue to run above the baseline performance after the initial CPU credits and accrued credits are consumed, the CPU credits change as shown in the following figure. The following section describes the two concepts involved in the figure:

  - Advance CPU credits: Advance CPU credits are credits that you can pay for in advance and can receive within 24 hours after you make the payment.
  - Overdrawn CPU credits: When you have consumed all the advance CPU credits, overdrawn CPU credits are used to ensure that the instance is running above the baseline performance. You are charged for using overdrawn CPU credits.

  > **Note** For more information about the billing of burstable instances, see Impact of performance modes on billing.

  Advance CPU credits and overdrawn CPU credits

  You can enable the unlimited mode for your burstable instances if you need to consume advance CPU credits or overdrawn CPU credits in addition to your credit balance to meet burstable performance requirements. Examples:

  - Some events, such as new feature releases, e-commerce promotions, and website promotions, cause a substantial increase to your workloads. High CPU performance is required during this period of time. In this case, you can enable the unlimited mode for your burstable instances. You can disable the mode to save costs when the workload peak ends.
  - Some web applications may require CPU bursts for a specific period during a day, but the daily average CPU utilization is below the baseline CPU utilization. In this case, you can enable the unlimited mode for your instances during peak hours to ensure a positive customer experience. If the CPU credits that you have earned during off-peak hours can offset the advance CPU credits that you consumed during peak hours, you can ensure customer satisfaction without paying additional charges.

By default, when you create a burstable instance, the standard mode is enabled. For more information about how to enable the unlimited mode, see Enable the unlimited mode.

For more information about how the CPU credits change when an instance is running in different performance modes, see CPU credit change examples.

Change configurations

During monitoring, you may find that the vCPUs are running above or below their baseline performance on a constant basis. This indicates that the instance type is not suitable for your business. We recommend that you re-evaluate the instance type to decide whether to select another burstable or enterprise-level instance type. For more information, see Instance families that support instance type changes.

The operation of changing configurations varies based on the billing method. For more information, see Overview of instance upgrade and downgrade.

t6, burstable instance family

Features:

- More cost-effective when compared with the t5 burstable instance family.
- Equipped with 2.5 GHz Intel® Xeon® Cascade Lake processors that deliver a maximum turbo frequency of 3.2 GHz.
- Provides baseline CPU performance and is burstable, but limited by accumulated CPU credits.
- Paired with the DDR4 memory.
- Supports enhanced SSDs (ESSD), standard SSDs, and ultra disks.

  > **Note** PL2 and PL3 ESSDs cannot provide maximum performance due to the specification limits of burstable instances. We recommend that you use enterprise-level instances or ESSDs that are of lower performance levels.

- Delivers a bandwidth of up to 4 Gbit/s.
- Supports VPCs only.
- Suitable for the following scenarios:
  - Web application servers
  - Lightweight applications and microservices
  - Development and testing environments

Instance types

Instance type	vCPUs	Memory (GiB)	Baseline CPU computing performance	CPU credits per hour	Max CPU credit balance	Local storage (GiB)	Base bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.t6-c4m1.large	2	0.5	5%	6	144	None	0.08	40	Yes	1	2	2

Instance type	vCPUs	Memory (GiB)	Baseline CPU computing performance	CPU credits per hour	Max CPU credit balance	Local storage (GiB)	Base bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.t6-c2m1.large	2	1.0	10%	12	288	None	0.08	60	Yes	1	2	2
ecs.t6-c1m1.large	2	2.0	20%	24	576	None	0.08	100	Yes	1	2	2
ecs.t6-c1m2.large	2	4.0	20%	24	576	None	0.08	100	Yes	1	2	2
ecs.t6-c1m4.large	2	8.0	30%	36	864	None	0.08	100	Yes	1	2	2
ecs.t6-c1m4.xlarge	4	16.0	40%	96	2304	None	0.16	200	Yes	1	2	6
ecs.t6-c1m4.2xlarge	8	32.0	40%	192	4608	None	0.32	400	Yes	1	2	6

> **Note**
>
> - When you bind an ENI to or unbind an ENI from an instance of the following instance types, the instance must be in the Stopped state: ecs.t6-c1m1.large, ecs.t6-c1m2.large, ecs.t6-c1m4.large, ecs.t6-c2m1.large, and ecs.t6-c4m1.large.
> - You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
> - For more information about these specifications, see Description of instance specifications.

t5, burstable instance family

Features:

- Equipped with 2.5 GHz Intel® Xeon® processors.
- Paired with the DDR4 memory.
- Offers multiple CPU-to-memory ratios.
- Provides baseline CPU performance and is burstable, but limited by accumulated CPU credits.
- Offers a balance of compute, memory, and network resources.
- Supports VPCs only.
- Suitable for the following scenarios:
  - Web application servers
  - Lightweight applications and microservices
  - Development and testing environments

Instance types

Instance type	vCPUs	Memory (GiB)	Baseline CPU computing performance	CPU credits per hour	Max CPU credit balance	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.t5-lc2m1.nano	1	0.5	20%	12	288	None	0.1	40	Yes	1	2	2
ecs.t5-lc1m1.small	1	1.0	20%	12	288	None	0.2	60	Yes	1	2	2

Instance type	vCPUs	Memory (GiB)	Baseline CPU computing performance	CPU credits per hour	Max CPU credit balance	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.t5-lc1m2.small	1	2.0	20%	12	288	None	0.2	60	Yes	1	2	2
ecs.t5-lc1m2.large	2	4.0	20%	24	576	None	0.4	100	Yes	1	2	2
ecs.t5-lc1m4.large	2	8.0	20%	24	576	None	0.4	100	Yes	1	2	2
ecs.t5-c1m1.large	2	2.0	25%	30	720	None	0.5	100	Yes	1	2	2
ecs.t5-c1m2.large	2	4.0	25%	30	720	None	0.5	100	Yes	1	2	2
ecs.t5-c1m4.large	2	8.0	25%	30	720	None	0.5	100	Yes	1	2	2
ecs.t5-c1m1.xlarge	4	4.0	25%	60	1440	None	0.8	200	Yes	1	2	6

Instance type	vCPUs	Memory (GiB)	Baseline CPU computing performance	CPU credits per hour	Max CPU credit balance	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.t5-c1m2.xlarge	4	8.0	25%	60	1440	None	0.8	200	Yes	1	2	6
ecs.t5-c1m4.xlarge	4	16.0	25%	60	1440	None	0.8	200	Yes	1	2	6
ecs.t5-c1m1.2xlarge	8	8.0	25%	120	2880	None	1.2	400	Yes	1	2	6
ecs.t5-c1m2.2xlarge	8	16.0	25%	120	2880	None	1.2	400	Yes	1	2	6
ecs.t5-c1m4.2xlarge	8	32.0	25%	120	2880	None	1.2	400	Yes	1	2	6
ecs.t5-c1m1.4xlarge	16	16.0	25%	240	5760	None	1.2	600	Yes	1	2	6

Instance type	vCPUs	Memory (GiB)	Baseline CPU computing performance	CPU credits per hour	Max CPU credit balance	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.t5-c1m2.4xlarge	16	32.0	25%	240	5760	None	1.2	600	Yes	1	2	6

Note

- When you bind an ENI to or unbind an ENI from an instance of the following instance types, the instance must be in the Stopped state: ecs.t5-lc2m1.nano, ecs.t5-c1m1.large, ecs.t5-c1m2.large, ecs.t5-c1m4.large, ecs.t5-lc1m1.small, ecs.t5-lc1m2.large, ecs.t5-lc1m2.small, and ecs.t5-lc1m4.large.
- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

4.10.2. Benefits

Burstable instances are the only instances that use CPU credits for maximum cost efficiency. This topic describes the benefits of burstable instances.

Lower cost

The cost of a burstable instance is 10% to 48% lower than that of a shared instance with the same configurations.

Note The actual performance of a burstable instance depends on the accumulation and usage of CPU credits. Before you purchase a burstable instance, make sure that you fully understand the relevant terms such as baseline performance and CPU credits of burstable instances. For more information, see Overview.

Finer-granularity specifications

ecs.t5-lc2m1.nano is an instance type of burstable instances, which has the baseline level specifications of one vCPU and 0.5 GiB memory. The instance specifications of burstable instances allow you to flexibly combine instance types to meet finer-granularity requirements. For more information, see t5, burstable instance family.

Predictable CPU performance

Burstable instances provide a baseline level of CPU performance (*baseline performance*). You can predict the burstable performance of a burstable instance based on its baseline performance and accrued CPU credits, and select appropriate compute capabilities.

Higher network performance

Burstable instances use the most advanced network-based technologies to reduce network latency to one-third of that of the last-generation instances.

4.10.3. CPU credit change examples

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

Background information

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

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

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

Standard mode

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

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

> **Note** The ecs.t5-lc1m2.large instance has the following features:
>
> - It is configured with 2 vCPUs and 4 GiB memory.
> - It receives 60 launch CPU credits upon creation (30 CPU credits for each vCPU).
> - It has a baseline performance of 10%.
> - It earns 12 CPU credits per hour and accrues a maximum of 288 CPU credits over a 24-hour period. For more information, see t5 instance type family.
> - It consumes 12 CPU credits per hour when running at baseline performance.

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

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

- 0–24 hours

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

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

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

- 25–48 hours

  i. Phase B: The CPU usage is 10%, which is equal to the baseline performance. Launch CPU credits are preferentially used and will not be recovered after being used up. The instance consumes 12 CPU credits per hour when it starts to run.

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

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

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

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

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

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

- 48–72 hours

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Unlimited mode

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

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

> **Note** The ecs.t5-c1m1.xlarge instance has the following features:
>
> - It is configured with 4 vCPUs and 4 GiB memory.
> - It receives 120 launch CPU credits upon creation (30 CPU credits for each vCPU).
> - It has a baseline performance of 15%.
> - It earns 36 CPU credits per hour and accrues a maximum of 864 CPU credits over a 24-hour period. For more information, see t5 instance type family.
> - It consumes 36 CPU credits per hour when running at baseline performance.

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

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

- 0–24 hours

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

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

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

- 25–48 hours

  i. Phase B: The CPU usage is 15%, which is equal to the baseline performance. Launch CPU credits are preferentially used and will not be recovered after being used up. The instance consumes 36 CPU credits per hour when it starts to run.

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

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

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

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

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

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

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

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

  v. Phase F: The instance runs at full capacity and consumes 240 CPU credits per hour. The baseline performance cannot meet requirements, and the instance starts to consume advance CPU credits. For more information, see Overview.

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

  vi. Phase G: The instance runs at full capacity and consumes 240 CPU credits per hour. The

advance CPU credits are used up and the instance starts to consume overdrawn CPU credits, which incur additional fees. For more information, see Overview.

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

- 48-72 hours

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

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

- 73-96 hours

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

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

4.10.4. Billing

The actual performance of burstable instances is determined based on CPU credits. This helps you control costs when you purchase computing capabilities. If your burstable instances consume additional CPU credits, you are charged additional fees. Therefore, the fees for a burstable instance contains the instance purchase fee and additional fees.

Instance purchase fee

Same as other instance families, burstable instances support pay-as-you-go and subscription billing methods. For more information about the difference between pay-as-you-go and subscription, see Comparison of billing methods. For information about the prices of burstable instances, see the Pricing page.

Burstable instances can be preemptible instances. For more information, see Overview.

If the billing method of your burstable instances is pay-as-you-go, you can purchase reserved instances to offset bills of the pay-as-you-go instances. For more information, see Reserved instance overview. However, the following limits apply if the instance types of your reserved instances belong to the gn6i and t5 instance families:

- You can purchase only zonal reserved instances.
- You cannot split or merge reserved instances.

Impact of performance modes on billing

When you purchase or use a burstable instance, you can choose whether to enable the unlimited mode for the burstable instance. The billing of a burstable instance is related to its performance mode. The following table describes the differences between the billing of burstable instances in different performance modes.

Performance mode	Instance purchase fee	Additional fee

Performance mode	Instance purchase fee	Additional fee
Standard mode	The purchase fee is determined by the billing method, not by the performance mode of the instances. For more information, see Instance purchase fee.	You are not charged additional fees when you use burstable instances.
Unlimited mode	The purchase fee is determined by the billing method, not by the performance mode of the instances. For more information, see Instance purchase fee.	You are charged additional fees in the following cases: • Overdrawn CPU credits are used. You will be billed and charged by hour. • Advance CPU credits are used and the performance mode is changed to the standard mode before the advance CPU credits are replenished. You will be charged a lump sum and the CPU credit balance will remain unchanged. • Advance CPU credits are used and the instance is stopped or released before the advance CPU credits are replenished. You will be charged a lump sum. For more information about the impact of stopping an instance on CPU credits, see the "Earn CPU credits" section in Overview.

The following table describes how additional fees are charged in the unlimited mode.

Region	Windows instance (USD/Credit)	Linux instance (USD/Credit)
Mainland China	0.0008	0.0008
Outside mainland China	0.0016	0.0008

Cost comparison between burstable instances and enterprise-level instances

Overdrawn CPU credits are available only when the unlimited mode is enabled. You will be charged for using overdrawn CPU credits for burstable instances. If the number of overdrawn CPU credits that are used by a burstable instance reaches a specific level, the total costs of the burstable instance will catch up with or even exceed the costs of an enterprise-level instance with the same configurations. For more information, see the "Unlimited mode" section in Overview.

> **Note** In this example, the prices on April 30, 2019 are used for illustration purposes. For the latest prices, see Pricing overview.

An ecs.t5-lc1m2.large burstable instance and an ecs.c5.large enterprise-level instance are used in this example. Both instances have two vCPUs and 4 GiB memory. Assume that the ecs.t5-lc1m2.large burstable instance has consumed all of the initial CPU credits, CPU credit balance, and advance CPU credits. The following figure shows the cost trends for two instances when overdrawn CPU credits are used.

In the preceding figure, the hourly pay-as-you-go price is calculated based on the configurations of the ecs.t5-lc1m2.large burstable instance (such as baseline CPU performance of 10% and two vCPUs) and the prices of the overdrawn CPU credits (such as for a Linux instance in the China (Beijing) region). For other instance types, the calculation method is the same, but the specific CPU utilization threshold is different.

You can learn from the figure that when the burstable instance starts to consume the overdrawn CPU credits:

- If the average CPU utilization is less than 57.83%, the cost of the ecs.t5-lc1m2.large burstable instance is lower than that of the ecs.c5.large enterprise-level instance.
- If the average CPU utilization is equal to 57.83%, the cost of the ecs.t5-lc1m2.large burstable instance is equal to that of the ecs.c5.large enterprise-level instance.
- If the average CPU utilization is greater than 57.83%, the cost of the ecs.t5-lc1m2.large burstable instance is higher than that of the ecs.c5.large enterprise-level instance.

Therefore, you can choose instance types based on the expected CPU utilization threshold to ensure service availability and control your costs.

The following figures show the cost trends of several popular burstable instance types. To learn more about other burstable instance types, submit a ticket.

4.10.5. Create a burstable instance

Burstable instances can use CPU credits to burst performance when required by your workloads. You can enable the unlimited mode for burstable instances. The minimum memory of burstable instances can be 0.5 GiB. This topic describes the configuration items of which you must take note when you create a burstable instance.

Context

For information about other general configuration items, see Create an instance by using the provided wizard.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. Click **Create Instance**.
4. Configure the parameters in the Basic Configurations step. Click **Next: Networking**.When you

configure parameters, note the following parameters:

- **Region**: Instance types that are available to your account in each region are displayed on the instance creation page.

  > Note You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.

- **Instance Type**: Select **x86-Architecture** > **Entry-Level (Shared)**. You can select **Enable Unlimited Mode for t5 Instances** when you create a burstable instance or after the instance is created. For more information, see Enable the unlimited mode.

- **Image**: Burstable instance types have a minimum memory of 0.5 GiB. They support only Linux and Windows Server Version 1809, and do not support operating systems that require more than 0.5 GiB memory, such as Windows Server 2016. For more information, see Select an image.

5. Configure the network and security group parameters. Click **Next: System Configurations**.Only the network type of VPC is supported.
6. Configure the system parameters. Click **Next: Grouping**.
7. Configure the grouping parameters. Click **Next: Preview**.
8. Read and select *ECS Service Terms*, and click Create Order.

Related information

- RunInstances

4.10.6. Switch the performance mode of a burstable instance

A burstable instance can run in standard or unlimited mode. This topic describes how to query and switch the performance mode of a burstable instance.

Context

If a burstable instance in standard mode does not have initial credits and exhausts its CPU credit balance, the instance will be unable to burst beyond its performance baseline. If a burstable instance in unlimited mode depletes its CPU credit balance, the instance can spend advance or overdraft CPU credits to burst beyond its performance baseline. In this case, you may be billed for the consumption of these CPU credits. For more information, see Overview.

- By default, the standard mode is enabled for newly created burstable instances.
- If a burstable instance is in the **Stopped** state and the No Fees for Stopped Instances (VPC-Connected) feature is enabled, the instance runs in the standard mode by default after it is started.
- If a burstable instance is in the **Stopped** state and the No Fees for Stopped Instances (VPC-Connected) feature is not enabled, the performance mode used before the instance is stopped will continue to take effect after the instance is started.
- If your account has overdue payments, the unlimited mode will be disabled for bustable instances automatically and will not be re-enabled until you settle the payments.

Query the performance mode of a burstable instance

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. (Optional)If the **Unlimited Mode** column is not displayed on the **Instances** page, configure the column to be displayed.
   i. Click the icon in the upper-right corner.
   ii. In the Column Filters dialog box that appears, select **Unlimited Mode** and click **OK**.
4. Find the target burstable instance and view the value in the **Unlimited Mode** column.
   - **Disabled**: indicates that the instance is in standard mode.
   - **Enabled**: indicates that the instance is in unlimited mode.

Enable the unlimited mode

Prerequisites:

- The target burstable instance is in the **Running** state.
- The burstable instance is in standard mode.

To enable the unlimited mode for a burstable instance, perform the following operations:

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. Find the target burstable instance and use one of the following methods to enable the unlimited mode for the instance:
   - To enable the unlimited mode for a single burstable instance, choose **More > Instance Settings > Enable Unlimited Mode** from the **Actions** column corresponding to the instance.
   - To enable the unlimited mode for one or more burstable instances, select the target instances and choose **More > Instance Settings > Enable Unlimited Mode** at the lower-left corner of the Instances page.
4. In the **Enable Unlimited Mode** message that appears, click **OK**.

Disable the unlimited mode

Prerequisites:

- The target burstable instance is in the **Running** state.
- The burstable instance is in unlimited mode.

To disable the unlimited mode for a burstable instance, perform the following operations:

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. Find the target burstable instance and use one of the following methods to disable the unlimited mode for the instance:
   - To disable the unlimited mode for a single burstable instance, choose **More > Instance Settings > Disable Unlimited Mode** from the **Actions** column corresponding to the instance.

- To disable the unlimited mode for one or more burstable instances, select the target instances and choose **More > Instance Settings > Disable Unlimited Mode** at the lower-left corner of the Instances page.

4. In the **Disable Unlimited Mode** message that appears, click **OK**.

Related information

- ModifyInstanceAttribute

4.10.7. Monitor burstable instances

This topic describes how to query the CPU utilization and credits of a burstable instance in the ECS console, and how to set CPU credit alert rules in the Cloud Monitor console.

Prerequisites

To set the contacts who receive notifications, you must create a contact group in advance. For more information about how to create a contact group, see Create an alert contact or alert group in *Cloud Monitor documentation*.

Context

Changes to the CPU credits of a burstable instance directly affect the CPU utilization and load performance of the instance. You can set monitoring alert rules for one or more burstable instances in the Cloud Monitor console. The following items can be monitored: consumed CPU credits, banked CPU credits, overdrawn CPU credits, and advance CPU credits. The following table describes the monitoring metrics of CPU credits for burstable instances.

Monitor metric	Description
Burst Credit	Displays changes in consumed CPU credits. Consumption trends are consistent with CPU utilization. For more information, see Overview.
Total Credit	Displays changes in banked CPU credits. Banked CPU credits are used to maintain CPU utilization. For more information, see CPU credits.
Notpaid Surplus Credit	Displays changes in overdrawn CPU credits. Overdrawn CPU credits can be used only when the unlimited mode is enabled. For more information, see Overview.
AdvanceCredit	Displays changes in advance CPU credits. Advance CPU credits can be used only when the unlimited mode is enabled. For more information, see Overview.

Query CPU credit usage information

You can perform the following steps to query the real-time credit trend of a burstable instance in the ECS console.

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

4. Find the burstable instance that you want to query and click the instance ID.
5. On the **Instance Details** page that appears, view the CPU credit usage and CPU utilization information of the instance.

Create CPU credit alert rules

You can perform the following steps to create alert rules for **Total Credit** and **Notpaid Surplus Credit** in the Cloud Monitor console. Two modes are available: standard and unlimited.

- In standard mode, if a burstable instance does not have any available CPU credits, the CPU utilization cannot exceed the baseline performance. When the **Total Credit** item is monitored, you can receive notifications when the instance performance is limited to decide whether to enable the unlimited mode.
- In unlimited mode, if a burstable instance has consumed all of its advance CPU credits, overdrawn CPU credits are consumed and billed on an hourly basis. This ensures that the CPU utilization exceeds the baseline performance. When the **Notpaid Surplus Credit** item is monitored, you can receive notifications when overdrawn CPU credits are billed to determine whether to disable the unlimited mode.

1. Log on to the CloudMonitor console.
2. In the left-side navigation pane, choose **Alarms > Alarm Rules**.
3. On the **Alarm Rules** page, click **Create Alarm Rule**.
4. On the **Create Alarm Rule** page, complete the following configurations:
   i. Configure parameters in the **Related Resource** section.
      - **Product**: Select **ECS** from the drop-down list.
      - **Resource Range**: Select **Instances** from the drop-down list.
      - **Instances**: Select one or more burstable instances.
   ii. Configure parameters in the **Set Alarm Rules** section.
      - **Alarm Rule**: Customize an alert rule name.

- **Rule Description:** Set alert rules and judgment standards.
  - Monitoring of **Total Credit**: Select **Total Credit** to monitor banked CPU credits. The values of 1Minute cycle, Continue for 1 periods, Average, <, and 1 are used in this example. If the average value of **Total Credit** is less than 1 and the status lasts for at least 1 minute, an alert is triggered.

    > **Note** In standard mode, if the banked CPU credits are less than 1, the CPU utilization of the burstable instance cannot exceed the baseline performance. In unlimited mode, if the CPU utilization exceeds the baseline performance, the burstable instance will consume advance CPU credits. If all advance CPU credits are consumed, the burstable instance will consume overdrawn CPU credits. You can also configure the average value for multiple consecutive periods as the alert triggering condition based on your actual requirements on CPU performance.

  - Monitoring of **Notpaid Surplus Credit**: Select **Notpaid Surplus Credit** to monitor overdrawn CPU credits. The values of 1Minute cycle, Continue for 1 periods, Average, >, and 0 are used in this example. If the average value of **Notpaid Surplus Credit** is greater than 0 and the status lasts for at least 1 minute, an alert is triggered.

    > **Note** If overdrawn CPU credits are greater than 0, overdrawn CPU credits are being used and billed. You can also configure the average value for multiple consecutive periods as the alert triggering condition based on your actual requirements on billing of overdrawn CPU credits.

  - Monitoring of **Burst Credit**: Select **Burst Credit** to monitor consumed CPU credits.
  - Monitoring of **AdvanceCredit**: Select **AdvanceCredit** to monitor advance CPU credits.
- **Mute for:** Select the interval at which notifications are pushed. The value of **10 min** is used in this example.
- **Effective Period:** Select the time range during which you can receive notifications.

iii. Set parameters in the **Notification Method** section.

- **Notification Contact**: Select a contact group to receive notifications.
- **Notification Methods**: Select **Email + DingTalk (Info)** or other methods as needed.

5. Click **Confirm**.

What's next

Switch the performance mode of a burstable instance

Related information

- Create a threshold-triggered alert rule
- Add charts

4.10.8. View bills of a burstable instance

After you purchase a burstable instance, additional fees will be charged to your account when you enable the unlimited mode and use overdrawn CPU credits. This topic describes how to query whether your burstable instances incur additional fees.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. Find the burstable instance that you want to query and click its ID.
4. In the **Payment Information** section, choose **More > View Fees**.
5. On the **Bills** page, click **Detail** in the **Actions** column corresponding to the target instance.
6. In the **Summary of billing items** section, view **SurplusCredit**.

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

4.11. Shared instance families

This topic describes the features of shared instance families and lists the instance types of each instance family.

- Previous-generation shared instance families xn4, n4, mn4, and e4

Shared instances use a CPU-unbound scheduling scheme. Each vCPU is randomly allocated to an idle CPU hyperthread. vCPUs of different instances compete for CPU resources, which causes computing performance to fluctuate when traffic loads are heavy. Shared instances can guarantee availability but cannot guarantee the performance that may be required in the SLA. Unlike enterprise-level instances that have exclusive resources, shared instances share resources. Therefore, shared instances cannot guarantee consistent computing performance but they are able to operate at a lower cost.

> **Note** Burstable instances are also shared instances. For more information, see Overview.

Previous-generation shared instance families xn4, n4, mn4, and e4

Features:

- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors.
- Paired with the DDR4 memory.
- Offer multiple CPU-to-memory ratios.

Instance family	Description	vCPU-to-memory ratio	Scenario

Instance family	Description	vCPU-to-memory ratio	Scenario
xn4	Shared compact type	1:1	• Frontend web applications • Lightweight applications and microservices • Applications for development or testing environments
n4	Shared compute type	1:2	• Websites and web applications • Development environments, servers, code repositories, microservices, and testing and staging environments • Lightweight enterprise applications
mn4	Shared balanced type	1:4	• Websites and web applications • Lightweight databases and caches • Integrated applications and lightweight enterprise services
e4	Shared memory type	1:8	• Applications that require a large amount of memory • Lightweight databases and caches

Instance types of xn4

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.xn4.small	1	1.0	None	0.5	50	No	1	2	2

Note

- When you bind an ENI to or unbind an ENI from an instance of the ecs.xn4.small instance type, the instance must be in the Stopped state.
- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

Instance types of n4

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.n4.small	1	2.0	None	0.5	50	No	1	2	2
ecs.n4.large	2	4.0	None	0.5	100	No	1	2	2
ecs.n4.xlarge	4	8.0	None	0.8	150	No	1	2	6
ecs.n4.2xlarge	8	16.0	None	1.2	300	No	1	2	6
ecs.n4.4xlarge	16	32.0	None	2.5	400	No	1	2	6
ecs.n4.8xlarge	32	64.0	None	5.0	500	No	1	2	6

Note

- When you bind an ENI to or unbind an ENI from an instance of the ecs.n4.small and ecs.n4.large instance types, the instance must be in the Stopped state.
- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

Instance types of mn4

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.mn4.small	1	4.0	None	0.5	50	No	1	2	2
ecs.mn4.large	2	8.0	None	0.5	100	No	1	2	2
ecs.mn4.xlarge	4	16.0	None	0.8	150	No	1	2	6
ecs.mn4.2xlarge	8	32.0	None	1.2	300	No	1	2	6
ecs.mn4.4xlarge	16	64.0	None	2.5	400	No	1	2	6
ecs.mn4.8xlarge	32	128.0	None	5	500	No	2	8	6

Note

- When you bind an ENI to or unbind an ENI from an instance of the ecs.mn4.small and ecs.mn4.large instance types, the instance must be in the Stopped state.
- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

Instance types of e4

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.e4.small	1	8.0	None	0.5	50	No	1	2	2
ecs.e4.large	2	16.0	None	0.5	100	No	1	2	2
ecs.e4.xlarge	4	32.0	None	0.8	150	No	1	2	6
ecs.e4.2xlarge	8	64.0	None	1.2	300	No	1	3	6
ecs.e4.4xlarge	16	128.0	None	2.5	400	No	1	8	6

Note

- When you bind an ENI to or unbind an ENI from an instance of the ecs.e4.small and ecs.e4.large instance types, the instance must be in the Stopped state.
- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

References

- Instance families
- Create an instance by using the provided wizard

5.Instance purchasing options

5.1. Subscription

Subscription is a billing method that allows you to use resources only after you pay for them. By using the subscription billing method, you can reserve resources in advance and reduce costs with discounted rates. This topic describes the billing rules for subscription ECS resources.

Overview

Before you use subscription resources, you must create a subscription ECS instance. The following figure shows subscription durations that you can choose for your ECS instance.

During instance creation, resources are billed separately to generate the total price. You can use subscription resources only after you pay the total price. For more information about how prices are calculated, see Billing.

After a subscription ECS instance is created, you can change its configurations or resize subscription cloud disks attached to the instance. For more information, see Overview of instance upgrade and downgrade and Overview.

After a subscription period expires, you can renew your ECS instance to continue to use it. For more information, see Overview.

Applicable resources

When you create an ECS instance, you must configure the instance type, Block Storage devices, image, and network type. The following table describes the ECS resources that support the subscription billing method.

Resource	Description
ECS instance	When you create an ECS instance, you must specify whether to use the subscription billing method.
Image	You can use images that use the subscription billing method to create subscription ECS instances.
Cloud disk	Images created along with subscription ECS instances also use the subscription billing method. After a subscription ECS instance is created, you can create subscription disks for the instance or attach pay-as-you-go disks that are separately created to the instance. For more information, see Create a subscription disk and Attach a data disk.
Public bandwidth (pay-by-bandwidth)	If you select pay-by-bandwidth as the billing method for network usage when you create a subscription ECS instance, the bandwidth is also billed on a subscription basis. For more information, see Billing methods of public bandwidth.

You can view the total price of the preceding resources in the lower-left part of the instance buy page, as shown in the following figure.

- **Total** price is the price of the following resources:
  - ECS instance
  - Cloud disk
  - Public bandwidth (pay-by-bandwidth)

    > **Note** If you select pay-by-traffic as the billing method for network usage, the total price does not include the price for public bandwidth. For more information, see Billing methods of public bandwidth.

- **Marketplace Image Fees** indicates that you selected a paid image.

Billing

The billing cycle of a subscription instance is the time commitment you made when you purchased the instance (based on UTC+8:00). The cycle starts from the time when the purchased subscription resource is activated or renewed (accurate to the second) and ends at 00:00:00 the next day after the expiration date.

Assume that you created a subscription ECS instance at 13:00:00 on August 9, 2019. Related resources include the ECS instance, image, and cloud disk (system disk). You select a subscription duration of one month and manually renew the instance with another one-month commitment. The following billing cycles apply:

- The first billing cycle is from 13:00:00 on August 9, 2019 to 00:00:00 on September 10, 2019.
- The second billing cycle is from 00:00:00 on September 10, 2019 to 00:00:00 on October 10, 2019.

ECS resources are billed separately. You must pay these fees before you can use the resources. You can calculate the total price based on the configurations that you choose. The following table describes the formulas used to calculate the fee of each ECS resource.

Resource	Formula	Unit price
ECS instance	Unit price of an instance type × Subscription duration	For more information, see the **Instance** tab on the ECS Pricing page. **Note** Local disks attached to an instance are billed together with the instance.
Image	Unit price of an image × Subscription duration	You can obtain the price on the buy page or in Alibaba Cloud Marketplace.

Resource	Formula	Unit price
Cloud disk (system disk)	Unit price of a disk × Disk capacity × Subscription duration	For more information, see the **System Cloud Disk Fee** section on the **Storage** tab of the ECS Pricing page. **Note** The price for a pay-as-you-go disk on the page is displayed in the USD/100 GiB/hour format. Divide it by 100 to obtain the unit price per GiB.
Cloud disk (data disk)	Unit price of a disk × Disk capacity × Subscription duration	For more information, see the **Data Cloud Disk Fee** section on the **Storage** tab of the ECS Pricing page. **Note** The price for a pay-as-you-go disk on the page is displayed in the USD/100 GiB/hour format. Divide it by 100 to obtain the unit price per GiB.
Public bandwidth (pay-by-bandwidth)	Unit price of bandwidth × Bandwidth value × Subscription duration For more information, see Billing methods of public bandwidth.	A tiered billing model is used for bandwidth. You can select a bandwidth value on the buy page to view the changes in fees.

Assume that you created a subscription ECS instance in China (Qingdao) with a subscription duration of three months. The following figure shows the process for calculating the price of the subscription ECS instance.

Note The price is only for reference. For more information about the actual price, see the links in the preceding table.

Changes in resource status after expiration

Notice After an ECS instance expires, the instance may be stopped. The system may send you notifications that you have overdue payments. In this case, renew your instance to ensure service availability. If you still have problems, submit a ticket.

If the auto-renewal feature is not enabled for a subscription ECS instance, the instance stops providing services at some point from 00:00:00 on the expiration date to 00:00:00 on the next day.

> **Note** You cannot enable the auto-renewal feature for an expired subscription ECS instance.

The following table describes the resource status of an ECS instance after expiration.

Resource	Within 15 days after expiration	More than 15 days after expiration
ECS instance	The ECS instance is retained but cannot be used. > **Note** After an ECS instance is stopped, you cannot connect to the instance or access websites deployed on the instance, and service errors may occur.	The ECS instance is released.
Image	The image is unavailable.	The image is unavailable.
Block Storage	• Cloud disks and data on them are retained, but the cloud disks cannot be used. • Local disks and data on them are retained, but the local disks cannot be used.	• Subscription disks are released and data on them cannot be restored. > **Note** If you manually attach a pay-as-you-go cloud disk to a subscription instance and do not set the release mode to Release with Instance, the pay-as-you-go cloud disk stops to work. • Local disks are released and data on them cannot be restored.

Public IP address	• In the classic network: The public IP address is retained.
	• In a VPC:
	  ◦ The public IP address is retained.
	  ◦ The EIP associated with the instance remains unchanged.
		• In the classic network: The public IP address is released.
		• In a VPC:
		  ◦ The public IP address is released.
		  ◦ The EIP is disassociated from the instance.

If the auto-renewal feature is enabled for a subscription ECS instance but the renewal fails, the instance automatically stops providing services at some point from 00:00:00 on the 15th day after expiration to 00:00:00 on the 16th day after expiration.

The following table describes the resource status of an ECS instance after expiration.

Resource	Within 15 days after expiration	More than 15 days after expiration	More than 30 days after expiration
ECS instance	The ECS instance is retained and works properly. Note When an ECS instance works properly, you can start or stop the instance, and connect to the instance by using the ECS console or other connection methods.	The ECS instance is retained but it cannot be used. Note After an ECS instance is stopped, you cannot connect to the instance or access websites deployed on the instance, and service errors may occur.	The ECS instance is released.
Image	The image is available.	The image is unavailable.	The image is unavailable.
Block Storage	• Cloud disks and data on them are retained. The cloud disks can work properly. • Local disks and data on them are retained. The local disks can work properly.	• Cloud disks and data on them are retained, but the cloud disks cannot be used. • Local disks and data on them are retained, but the local disks cannot be used.	• Subscription disks are released and data on them cannot be restored. Note If you manually attach a pay-as-you-go cloud disk to a subscription instance and do not set the release mode to Release with Instance, the pay-as-you-go cloud disk stops to work. • Local disks are released and data on them cannot be restored.

Resource	Within 15 days after expiration	More than 15 days after expiration	More than 30 days after expiration

Public IP address	• In the classic network: The public IP address is retained.
	• In a VPC:
	  ○ The public IP address is retained.
	  ○ The EIP associated with the instance remains unchanged.
		• In the classic network: The public IP address is retained.
		• In a VPC:
		  ○ The public IP address is retained.
		  ○ The EIP associated with the instance remains unchanged.
			• In the classic network: The public IP address is released.
			• In a VPC:
			  ○ The public IP address is released.
			  ○ The EIP is disassociated from the instance.

Overdue payments

If you have overdue payments in your account, subscription ECS resources can be used properly, but you cannot purchase, renew, or upgrade instances.

5.2. Pay-as-you-go

This topic describes the billing and settlement rules for pay-as-you-go ECS resources. Pay-as-you-go allows you to pay only for the amount of resources you actually use. You can purchase and release resources on demand and scale up as your business grows. Pay-as-you-go resources can help reduce costs by 30% to 80% compared to traditional hosts.

Overview

Pay-as-you-go resources are billed based on the billing cycle and paid each hour. If you have a quota agreement with Alibaba Cloud, fees are deducted when the cumulative consumption of your account exceeds the quota. You must complete the payment at least once a month.

After you create a pay-as-you-go ECS resource, you can change its configurations. For more information, see Change the instance type of a pay-as-you-go instance and Change the Internet bandwidth of a pay-as-you-go instance.

You can change the billing method of your pay-as-you-go ECS resources. For more information, see Switch the billing method from pay-as-you-go to subscription.

You can use one of the following methods to view your bills:

- To view the fee calculation method, see Billing.
- To understand how ECS resource status affects the billing duration, see Billing duration.

  > **Note** If you stop an ECS instance but do not release related resources, you continue to be charged for the resources.

- For information about settlement, see Settlement period.

Applicable resources

The pay-as-you-go billing method is available for the following ECS resources:

- ECS instances

- Images
- Cloud disks
- Public bandwidth (pay-by-bandwidth)
- Snapshots

In addition to the instance type, you must also configure the Block Storage, image, and network type when you create an instance. The images and the cloud disks created along with the pay-as-you-go ECS instance also use the pay-as-you-go billing method. However, you can select the billing method for network usage.

> **Note** After you create a pay-as-you-go ECS instance, you can attach a separately created pay-as-you-go cloud disk to the instance. For more information, see Attach a data disk.

You are charged for created snapshots.

You can view the total price of the preceding resources in the lower-left part of the instance creation page.

- **Total** price is the price of the following resources:
  - ECS instance
  - Cloud disk
  - Public bandwidth (pay-by-bandwidth)

    > **Note** If you select pay-by-traffic as the billing method for network usage, the total price does not include the price for public bandwidth. For more information, see Billing methods of public bandwidth.

- **Marketplace Image Fees** indicates that you selected a paid image.

Billing

You are charged for pay-as-you-go ECS resources after they are created. Resources are billed separately. You can calculate the total fee that you need to pay for a certain period of time based on the configurations you choose. The following table describes the billing cycle of each ECS resource and the formula used to calculate the fees.

Resource	Billing cycle	Formula	Unit price
ECS instance	Second	Unit price of an instance type × Billing duration	For more information, see the **Instance** tab on the ECS Pricing page. **Note** Local disks attached to an instance are billed together with the instance.

Resource	Billing cycle	Formula	Unit price
Image	Second	Unit price of an image × Billing duration	You can view the price on the instance buy page and in Alibaba Cloud Marketplace.
Cloud disk (system disk)	Second	Unit price of a cloud disk × Disk capacity × Billing duration	For more information, see the **System Cloud Disk Fee** section on the **Storage** tab of the ECS Pricing page. **Note** The price on the page is displayed in the USD/100 GiB/hour format. Divide it by 100 to obtain the unit price per GiB.
Cloud disk (data disk)	Second	Unit price of a cloud disk × Disk capacity × Billing duration	For more information, see the **Data Cloud Disk Fee** section on the **Storage** tab of the ECS Pricing page. **Note** The price on the page is displayed in the USD/100 GiB/hour format. Divide it by 100 to obtain the unit price per GiB.
Public bandwidth (pay-by-bandwidth)	Second	Unit price of bandwidth × Bandwidth value × Billing duration For more information, see Billing methods of public bandwidth.	A tiered billing model is used for bandwidth. You can select a bandwidth value on the buy page to view the changes in fees.
Snapshot	Hour	Unit price of a snapshot × Snapshot capacity × Billing duration For more information, see Snapshot billing.	For more information, see the **Snapshot** tab on the ECS Pricing page.

> **Note**
> - If the billing cycle is one second, the fee generated each second is added on to the bill. If an hourly price is displayed, you can divide the price by 3,600 to obtain the price per second.
> - If the billing cycle is one hour, the fee generated every hour is added on to the bill. A usage duration of less than an hour is calculated as an hour.

Assume that you created a pay-as-you-go ECS instance in the China (Qingdao) region and the resource was in use from 11:00:00 to 12:00:00 on August 8, 2019. The following figure shows the process for calculating the price of the instance.

> **Note** The price is for reference only. For more information about the exact price, visit the links in the preceding table.

Billing duration

If a pay-as-you-go ECS instance is automatically stopped due to an overdue payment, the billing of its resources is suspended. The billing resumes after you settle the overdue payment and reactivate the instance.

The following table describes the billing duration of each resource type on the prerequisite that you have no overdue payments.

Resource	Billing duration

ECS instance	The billing duration is affected by the network type of the instance.
	• For an ECS instance in the classic network, billing starts when the instance is created and stops when the instance is released.
	• For an ECS instance in a VPC, the billing duration depends on whether the No Fees for Stopped Instances (VPC-Connected) feature is enabled.
	  ◦ If this feature is not enabled, billing starts when the instance is created and stops when the instance is released.
	  ◦ If this feature is enabled, billing starts when the instance is created or is restarted in the ECS console and stops when the instance is stopped in the ECS console or when the instance is released. For more information, see No Fees for Stopped Instances (VPC-Connected).
	Notice If you stop the instance by shutting down its operating system, you cannot benefit from the No Fees for Stopped Instances (VPC-Connected) feature.
	You can purchase reserved instances to reduce your costs. For more information, see Reserved instance overview.

Resource	Billing duration
Image	Billing starts when the instance is created and stops when the instance is released.
Cloud disk (system disk)	Billing starts when the instance is created and stops when the instance is released.
Cloud disk (data disk)	Billing starts when the data disk is created and stops when the data disk is released.
Public bandwidth (pay-by-bandwidth)	Billing starts when public bandwidth (pay-by-bandwidth) is enabled and stops when public bandwidth is disabled or when the ECS instance is released. For information about how to disable public bandwidth, see Change the Internet bandwidth of a pay-as-you-go instance.
Snapshot	Billing starts when the snapshot is created and stops when the snapshot is deleted.

Note If a pay-as-you-go ECS instance incurs less than USD 0.01 in charges during its entire lifecycle, USD 0.01 is charged.

To avoid unexpected fees incurred when a pay-as-you-go instance is kept for longer than expected, we recommend that you enable the automatic release feature. If automatic release is enabled, billing stops when the resources are released. The automatic release time is accurate to the second.

Settlement period

The fees for pay-as-you-go resources are paid each hour. These fees are paid together with the fees incurred by subscription products in your account. If you have a quota agreement with Alibaba Cloud, fees are deducted when the cumulative consumption of your account exceeds the quota. If the cumulative monthly consumption of your account is less than the quota, fees are deducted on the first day of the following month.

- If your default payment method is bank card, the quota is USD 1,000.
- If your default payment method is PayPal or Paytm (India), the quota depends on your ECS resource usage.

Fee deduction is attempted three times: on the due date (T), T+7, and T+14. If fee deduction fails on the due date (T), the system will attempt to deduct fees again on the day T+7 and day T+14. If fee deduction fails for these three times, the instance is stopped on the day T+15 and the billing of its resources also stops.

The following changes occur to the resource status in the event of an overdue payment:

1. Within 15 days after the payment becomes overdue, you can use existing ECS resources but cannot purchase new ECS instances, upgrade instance configurations, or renew instances.
2. Within 15 days after the instance is stopped, you must submit a ticket to complete the overdue payment and then reactivate the instance. Otherwise, the instance will be released. For information about resource status, see Resource status when an ECS instance

is out of service.

> **Note** Reactivating and restarting an instance are different.
>
> - *Reactivate* means that you resume a pay-as-you-go ECS instance that has entered the Expired or Expired and Being Recycled state due to an overdue payment. For more information, see Reactivate an instance.
> - *Restart* means that you restart an ECS instance that is in the Running state.

3. More than 15 days after the payment becomes overdue, pay-as-you-go instances are released.

Resource status when an ECS instance is out of service

> **Notice** If you have overdue payments in your account, your ECS instances may be stopped. The system will send you notifications about the overdue payments. We recommend that you settle the overdue payment in time to avoid impacts on your services. If you have any other questions, submit a ticket.

If all three deductions fail, the ECS instance goes out of service. The following table describes the resource status after your instance is stopped due to an overdue payment.

Resource	Within 15 days after the instance is taken out of service	More than 15 days after the instance is taken out of service
ECS instance	The ECS instance is retained but is stopped and taken out of service. When a pay-as-you-go ECS instance is automatically *stopped due to an overdue payment*, it enters the Expired state and the billing stops. After the instance stops providing services, you cannot connect to the instance or access websites deployed on the instance. Service errors may occur.	The ECS instance is released. **Note** If your ECS instance is released due to an overdue payment, Alibaba Cloud will send you an email notification.
Image	The image is unavailable.	The image is unavailable.

Resource	Within 15 days after the instance is taken out of service	More than 15 days after the instance is taken out of service
Block Storage	• Cloud disks and data on them are retained, but the cloud disks cannot be used. • Local disks and data on them are retained, but the local disks cannot be used. Note After cloud disks or local disks become unusable, they cannot properly process I/O read and write requests. This affects the normal running of ECS instances to which these disks are attached, such as excessive time needed to perform operations and unpredictable power-off or restart failures for some operating systems.	• Cloud disks are released and data on them cannot be restored. Note Cloud disks (data disks) created with pay-as-you-go ECS instances and pay-as-you-go cloud disks (data disks) created separately on the Disks page in the ECS console are released, regardless of whether they are attached to the instance. • Local disks are released and data on them cannot be restored.
IP address	• In the classic network: The public IP address is retained. • In a VPC: ◦ The public IP address is retained. Notice If the No Fees for Stopped Instances (VPC-Connected) feature is enabled before the instance is stopped, the public IP address of the instance may be recycled after the instance is stopped due to an overdue payment. The public IP address may change when the instance is restarted. ◦ The EIP associated with the instance remains unchanged.	• In the classic network: The public IP address is released. • In a VPC: ◦ The public IP address is released. ◦ The EIP is disassociated from the instance.
Snapshot	All snapshots are retained but automatic snapshots cannot be created.	All snapshots are deleted except for those used to create cloud disks or custom images.

5.3. Preemptible instances

5.3.1. Overview

Preemptible instances are a type of on-demand instance. You can use this type of instances to minimize ECS instance costs in a variety of scenarios.

Introduction

The market price of a preemptible instance fluctuates based on changes to the supply and demand of its instance type. When you create a preemptible instance, you must specify a maximum price per hour to bid for a specific instance type. If your bid is higher than the current market price and the stock of the instance type is sufficient, your instance is created and charged at the current market price. 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 services such as cloud disks or elastic IP address.

A preemptible instance has a protected period of one hour after it is created. During this period, the instance is not released even if your bid is lower than the current market price, which ensures that the instance can run services normally. One hour later, the system checks the market price and the stock of the instance type every five minutes. If the market price is higher than your bid or if the stock of the instance type is insufficient, your preemptible instance is released. For more information about preemptible instances, see ECS instance FAQ.

> **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 the instance is released. For more information, see Create a normal snapshot.

The following figure shows the lifecycle of a preemptible instance.

After a preemptible instance is created, you can release the instance at any time. For more information, see Release an instance. When the market price exceeds your bid or when the stock of the instance type is insufficient, your instance will enter the **Recycling** state. After five minutes, the instance is automatically released. You can check whether the instance enters the **Recycling** state based on the instance metadata or the `OperationLocks` information returned by the DescribeInstances operation.

You can check whether your preemptible instance enters the **Recycling** state and store a small amount of data while you wait for the release of the instance. However, we recommend that you optimize the application design to ensure that the application runs properly after your preemptible instance is released. You can try to manually release your preemptible instance to check whether the application runs properly after the preemptible instance is released.

Typically, the system first releases the instance with the lowest bid price. If multiple preemptible instances have the same bid price, the system randomly determines the order in which the instances are released.

Limits

- Whether you can purchase a preemptible instance depends on your ECS instance resource usage.
- Preemptible instances cannot be converted to subscription instances.
- The instance types of preemptible instances cannot be changed.
- For information about the quota of preemptible instances, see the "Instance limits" section in Limits.

Scenarios

Preemptible instances are ideal for stateless applications, such as scalable web services, image rendering, big data analysis, and large-scale parallel computing. Preemptible instances are applicable to applications that require a high level of distribution, scalability, and fault tolerance capabilities. Preemptible instances help save costs and increase throughput of these applications.

You can use preemptible instances for the following businesses:

- Real-time analysis
- Big data
- Geospatial surveys
- Image and media coding
- Scientific computing
- Scalable websites and web crawlers
- Tests

Preemptible instances are not suitable for stateful applications such as databases, because when a preemptible instance is released due to a failed bid or other reasons, it is difficult to store application state data.

Pricing and billing

- Prices

  The price of a preemptible instance covers only the price of the instance type (including vCPUs and memory) and excludes the prices of resources such as system disks, data disks, and network bandwidth.

  - System disks and data disks are charged on a pay-as-you-go basis. For more information, see Pay-as-you-go.
  - Network bandwidth is charged based on the bandwidth billing method of pay-as-you-go instances. For more information, see Billing methods of public bandwidth.

- Market prices

  The market price of a preemptible instance fluctuates based on changes to the supply and demand of its instance type. When you attempt to create a preemptible instance, the creation is successful if your bid price is higher than the current market price of the specified instance type and the stock of the instance type is sufficient.

  Within the first hour after the instance is created, the instance is charged based on the market price at the time of bidding. After one hour, the instance is charged based on the real-time market price.

  We recommend that you evaluate the market price fluctuations to minimize computing costs and increase throughput when you purchase preemptible instances.

- Billing methods

  Preemptible instances are charged by second. The market price of a preemptible instance is an hourly price. You can divide the hourly price by 3,600 to get the price per second.

  The cost incurred by a preemptible instance from creation to release is accurate to two decimal places. Accrued costs of less than USD 0.01 are not charged.

- Billing duration

  A preemptible instance is charged based on its actual usage period, which lasts from the creation to the release of the instance. If you stop an instance only by calling the StopInstance operation or by using the ECS console, the instance continues to be charged. When you no longer need a preemptible instance, we recommend that you create snapshots to back up your data and environment and then release the instance. You can purchase new preemptible instances at any time.

5.3.2. Create a preemptible instance

This topic describes how to create a preemptible instance in the ECS console.

Context

To create and use a preemptible instance, take note of the following items:

- Set an appropriate bid which is comparable to but higher than the fluctuating market price. This way, your preemption request can be accepted and processed, and the created preemptible instance will not be released due to a low bid. The bid must also meet your business expectations.
- Use an image that contains the configurations of all the required software. This can ensure the instance can be started any time after it is created. You can also use user data of the instance to run commands at instance startup. For more information, see Prepare user data.
- Save important data in storage media that are not affected when preemptible instances are released, such as separately created cloud disks, OSS buckets, or ApsaraDB for RDS instances. This can prevent data loss caused by instance release.
- Break down your jobs into small tasks by using grids, Hadoop, queue-based architecture, or use checkpoints to save calculation results.
- Monitor the status of a preemptible instance by checking the instance release notifications from ECS. ECS updates the instance metadata five minutes before ECS releases a preemptible instance. You can obtain the status of a preemptible instance every minute by checking instance metadata. For more information, see Metadata.
- Run your applications on a pay-as-you-go instance and release the instance to verify whether your applications can automatically adjust themselves in response to the release of preemptible instances.

You can use developer tools such as Alibaba Cloud CLI, OpenAPI Explorer, and Alibaba Cloud SDK to call the RunInstances operation and create a preemptible instance.

> **Note** You can set the SpotStrategy parameter to *SpotAsPriceGo* to use the actual market price. Alternatively, you can set the SpotStrategy parameter to *SpotWithPriceLimit* to use the acceptable maximum price.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. On the **Instances** page, click **Create Instance**.
4. Set **Billing Method** to **Preemptible Instance**.

5. In the **Maximum Price for Instance Type** section, specify your bid in the spin box. The preemptible instance you request is created at the market price only if your bid is higher than or equal to the market price and resources are sufficient. You can bid for a preemptible instance only once. The following bidding modes are supported:
   - **Use Automatic Bid**: The real-time market price is used as the bidding price.
   - **Set Maximum Price**: You must set the highest price you are willing to pay for the instance type.

     > **Note** In the displayed price range, the maximum price is equal to the price for the pay-as-you-go instance of the same instance type. Your bid must be based on the displayed price range, your business needs, and the estimated future price fluctuation. If you take into account the estimated future price fluctuation, you can continue to hold the instance even after the one-hour guaranteed duration expires. Otherwise, after the guaranteed duration expires, your instance may be automatically released.
6. Select or enter the quantity of instances you want to purchase.
7. Complete other settings. For more information about other parameters, see Create an instance by using the provided wizard.
8. Preview the configurations you selected, and then click **Create Instance**.

Result

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** in the Billing Method column. Click the instance ID or instance name to go to the Instance Details page. You can view the bidding policy configured during instance creation in the **Payment Information** section.

Related information

- RunInstances

5.3.3. Query interruption events of preemptible instances

If a preemptible instance is forcibly recycled due to a market price change or insufficient resources, an interruption event is triggered. Before the preemptible instance is recycled, it enters the locked state and you are prompted that the instance will be recycled. This topic describes how to query interruption events of preemptible instances. You can automate the instance interruption and recycle processes based on the recycle status of instances.

Query interruption events of preemptible instances by using Cloud Monitor SDK

This section describes how to query interruption events of preemptible instances by using Cloud Monitor SDK for Java.

1. Access Cloud Monitor SDK for Java. For more information, see Java SDK.
2. Query interruption events of preemptible instances by using the SDK.

```
import com.aliyuncs.DefaultAcsClient;
import com.aliyuncs.IAcsClient;
import com.aliyuncs.exceptions.ClientException;
import com.aliyuncs.exceptions.ServerException;
import com.aliyuncs.profile.DefaultProfile;
import com.google.gson.Gson;
import java.util.*;
import com.aliyuncs.cms.model.v20190101.*;

public class DescribeSystemEventAttribute {
   public static void main(String[] args) {

      DefaultProfile profile = DefaultProfile.getProfile("cn-hangzhou", "<accessKeyId>", "<accessSe
cret>");
      IAcsClient client = new DefaultAcsClient(profile);

      DescribeSystemEventAttributeRequest request = new DescribeSystemEventAttributeRequest
();
      request.setRegionId("cn-hangzhou");
      request.setProduct("ECS");
      request.setEventType("StatusNotification");
      request.setName("Instance:PreemptibleInstanceInterruption");
      try {
         DescribeSystemEventAttributeResponse response = client.getAcsResponse(request);
         System.out.println(new Gson().toJson(response));
      } catch (ServerException e) {
         e.printStackTrace();
      } catch (ClientException e) {
         System.out.println("ErrCode:" + e.getErrCode());
         System.out.println("ErrMsg:" + e.getErrMsg());
         System.out.println("RequestId:" + e.getRequestId());
      }
   }
}
```

3. Check interruption events of preemptible instances based on the responses.

   The following code shows the event notification in the JSON format:

```
{
 "ver": "1.0",
 "id": "2256A988-0B26-4E2B-820A-8A********E5",
 "product": "ECS",
 "resourceId": "acs:ecs:cn-hangzhou:169070********30:instance/i-bp1ecr********5go2go",
 "level": "INFO",
 "name": "Instance:PreemptibleInstanceInterruption",
 "userId": "169070********30",
 "eventTime": "20190409T121826.922+0800",
 "regionId": "cn-hangzhou",
 "content": {
  "instanceId": "i-bp1ecr********5go2go",
  "action": "delete"
 }
}
```

The following table describes parameters in the content field.

Parameter	Description	Example
instanceId	The ID of the preemptible instance.	*i-bp1ecr********5go2go*
action	The operation on the preemptible instance. If the parameter is set to *delete*, the interrupted preemptible instance will be forcibly recycled.	*delete*

Query interruption events of preemptible instances by using instance metadata

1. Remotely connect to an ECS instance. For more information, see Overview.
2. Run the following command to query instance metadata:

```
curl 'http://100.100.100.200/latest/meta-data/instance/spot/termination-time'
```

   - If the response is empty, the instance can still be used.
   - If information in the `2015-01-05T18:02:00Z` (UTC+0) format is returned, the preemptible instance will be recycled at that point in time.

Query interruption events of preemptible instances by calling an API operation

This section describes how to call the DescribeInstances operation to determine whether the instance has entered the *to be recycled* state based on the returned OperationLocks parameter.

The following content shows sample code of DescribeInstancesSample.java:

```
public class DescribeInstancesSample {
  public static void main(String[] args) throws InterruptedException {
    OpenApiCaller caller = new OpenApiCaller();
    JSONArray allInstances = new JSONArray();
    allInstances.addAll(Arrays.asList("i-bp18hgfai8ekoqwo0***", "i-bp1ecbyds24ij63w1***"));
    while (! allInstances.isEmpty()) {
      DescribeInstancesRequest request = new DescribeInstancesRequest();
      request.setRegionId("cn-hangzhou");
      request.setInstanceIds(allInstances.toJSONString());// Specify the ID of the instance and query
only the instance whose ID has been specified.
      DescribeInstancesResponse response = caller.doAction(request);
      List<DescribeInstancesResponse.Instance> instanceList = response.getInstances();
      if (instanceList ! = null && ! instanceList.isEmpty()) {
        for (DescribeInstancesResponse.Instance instance : instanceList) {
          System.out.println("result:instance:" + instance.getInstanceId() + ",az:" + instance.getZoneI
d());
          if (instance.getOperationLocks() ! = null) {
            for (DescribeInstancesResponse.Instance.LockReason lockReason : instance.getOperatio
nLocks()) {
              System.out.println("instance:" + instance.getInstanceId() + "-->lockReason:" + lockReas
on.getLockReason() + ",vmStatus:" + instance.getStatus());
              if ("Recycling".equals(lockReason.getLockReason())) {
                //do your action
                System.out.println("spot instance will be recycled immediately, instance id:" + instanc
e.getInstanceId());
                allInstances.remove(instance.getInstanceId());
              }
            }
          }
        }
        System.out.print("try describeInstances again later ...") ;
        Thread.sleep(2 * 60 * 1000);
      } else {
        break;
      }
    }
  }
}
```

The following result is returned if the recycle is triggered:

```
instance:i-bp1ecbyds24ij63w****-->lockReason:Recycling,vmStatus:Stopped
spot instance will be recycled immediately, instance id:i-bp1ecbyds24ij63w****
```

5.3.4. View bills of a preemptible instance

This topic describes how to view the bills of a preemptible instance.

Context

The price of each pay-as-you-go instance is fixed at the time of purchase, whereas the price of each preemptible instance fluctuates with the supply and demand changes.

Procedure

1. Log on to the ECS console.
2. Choose **Billing > User Center**.
3. In the left-side navigation pane, choose **Spending Summary > Instance Spending Detail**.
4. Enter the ID of the preemptible instance whose bills you want to view to search for the instance. Find the instance in the searching results and click **Detail** in the **Action** column. You can also search for instance bills by product family, product name, or billing method.

5.3.5. Stop a preemptible instance

This topic describes how to stop a preemptible instance and whether the stopped instance can be restarted under different conditions.

Context

Only VPC-type preemptible instances support the No Fees for Stopped Instances (VPC-Connected) feature. For more information, see No Fees for Stopped Instances (VPC-Connected). The following table describes the restart of stopped instances that are of different network types and use different bidding modes.

Network type	Bidding mode	Stop mode	Restart of stopped instance
Classic network	SpotWithPriceLimit	Keep Stopped Instances and Continue Billing	The stopped instance can be restarted within the protection period. After the protection period ends: • If your bid price for the instance type is greater than or equal to the prevailing market price and if the stock of the instance type is sufficient, the stopped instance can be restarted. • If your bid price is lower than the prevailing market price or if the stock of the instance type is insufficient, the stopped instance cannot be restarted.

Network type	Bidding mode	Stop mode	Restart of stopped instance
	SpotAsPriceGo	Keep Stopped Instances and Continue Billing	The stopped instance can be restarted within the protection period. After the protection period ends: • If the stock of the instance type is sufficient, the stopped instance can be restarted. • If the stock of the instance type is insufficient, the stopped instance cannot be restarted.
VPC	SpotWithPriceLimit	Keep Stopped Instances and Continue Billing	The stopped instance can be restarted within the protection period. After the protection period ends: • If your bid price for the instance type is greater than or equal to the prevailing market price and if the stock of the instance type is sufficient, the stopped instance can be restarted. • If your bid price is lower than the prevailing market price or if the stock of the instance type is insufficient, the stopped instance cannot be restarted.
		No Fees for Stopped Instances (VPC-Connected)	The stopped instance can be restarted within the protection period as long as the stock of the instance type is sufficient. After the protection period ends: • If your bid price for the instance type is greater than or equal to the prevailing market price and if the stock of the instance type is sufficient, the stopped instance can be restarted. • If your bid price is lower than the prevailing market price or if the stock of the instance type is insufficient, the stopped instance cannot be restarted.
	SpotAsPriceGo	Keep Stopped Instances and Continue Billing	The stopped instance can be restarted within the protection period. After the protection period ends: • If the stock of the instance type is sufficient, the stopped instance can be restarted. • If the stock of the instance type is insufficient, the stopped instance cannot be restarted.

Network type	Bidding mode	Stop mode	Restart of stopped instance
		No Fees for Stopped Instances (VPC-Connected)	The stopped instance can be restarted within the protection period as long as the stock of the instance type is sufficient. After the protection period ends: • If the stock of the instance type is sufficient, the stopped instance can be restarted. • If the stock of the instance type is insufficient, the stopped instance cannot be restarted.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances.**
3. On the **Instances** page, find the preemptible instance that you want to stop, and choose **More > Instance Status > Stop** in the Actions column.
4. In the **Stop Instance** dialog box, click **OK.**

Related information

- StopInstance

5.4. Reserved Instances

5.4.1. Reserved instance overview

A reserved instance is a discount coupon that can be automatically applied to one or more pay-as-you-go instances, excluding preemptible instances. A reserved instance can also be used to reserve instance resources. A combination of reserved instances and pay-as-you-go instances provides similar cost-effectiveness to subscription instances but with a higher degree of flexibility.

Comparison between reserved instances, pay-as-you-go instances, and subscription instances

The following table lists differences between reserved instances, pay-as-you-go instances, and subscription instances.

Item	Reserved instance	Pay-as-you-go instance	Subscription instance
Form	A discount coupon. Reserved instances are classified into regional and zonal reserved instances.	An instance that uses the pay-as-you-go billing method. A pay-as-you-go instance is equivalent to a virtual machine. For more information, see Pay-as-you-go.	An instance that uses the subscription billing method. A subscription instance is equivalent to a virtual machine. For more information, see Subscription.

Item	Reserved instance	Pay-as-you-go instance	Subscription instance
Purpose	Reserved instances cannot be used independently. They must match pay-as-you-go instances to offset bills.	Pay-as-you-go instances can be used independently. They can be used as simple web servers, or used in combination with other Alibaba Cloud services to deliver powerful solutions.	Subscription instances can be used independently. They can be used as simple web servers, or used in combination with other Alibaba Cloud services to deliver powerful solutions.

Scenarios

The following table lists scenarios where a combination of reserved instances and pay-as-you-go instances is the optimal solution.

Scenario	Combination of reserved instances and pay-as-you-go instances	Pay-as-you-go instance	Subscription instance
You may need to change the region for your business. You must release ECS instances in the original zone and create ECS instances in zones of the destination region.	A combination of reserved instances and pay-as-you-go instances has the following benefits: • After you purchase reserved instances, you make commitments to using pay-as-you-go instances for a period of time. Reserved instances provide significant discounts compared with pay-as-you-go pricing. • Reserved instances that you purchase deliver computing power instead of instances. Reserved instances can match pay-as-you-go instances that meet the matching requirements, and are more flexible than subscription instances.	The unit prices of pay-as-you-go instances are higher than those of subscription instances. You may fail to create pay-as-you-go instances if resources are insufficient. However, pay-as-you-go instances are easier to manage. For example, after you configure automated O&M, ECS instances	Bills are associated with ECS instances. You may need to pay service fees when you apply for refunds.
During automated O&M, the number of ECS instances needs to be automatically adjusted.			Refunds must be manually implemented.
You use Auto Scaling to manage ECS instances and own a large number of pay-as-you-go instances. You want to lower your costs.			You must manually change pay-as-you instances to subscription instances. This process is inefficient and prone to errors.
You want to simplify the operations of managing the lifecycle of subscription instances, such as renewing, releasing, and synchronizing the expiration dates of ECS instances.			You must perform a large number of operations.

Scenario	Combination of reserved instances and pay-as-you-go instances	Pay-as-you-go instance	Subscription instance
You want to pay for resources by installments to mitigate financial pressure.	• You can split or merge reserved instances to offset bills for pay-as-you-go instances of different instance types. • You can modify the zone of a reserved instance anytime. A regional reserved instance can be used to offset bills of pay-as-you-go instances across zones. • You can pay by hour by selecting the partial upfront or no upfront payment option to avoid financial pressure caused by one-time payment.	are automatically created and released. No manual refunds are involved after ECS instances are released. Pay-as-you-go instances can be used together with Auto Scaling.	The unit prices are lower. However, one-time payment is required.

Attributes

The following figure shows the key attributes of a reserved instance.

The following table describes the key attributes of the preceding reserved instance.

Section	Attribute	Description

Section	Attribute	Description
①	Instance type	The following instance families support reserved instances: • General purpose instance families: g6e, g6, g5, and sn2ne • Compute optimized instance families: c6e, c6, c5, ic5, and sn1ne • Memory optimized instance families: r6e, r6, r5, and se1ne • Big data instance family: d2s • Instance families with local SSDs: i2 and i2g • Instance families with high clock speed: hfc6, hfc5, hfg6, hfg5, and hfr6 • GPU-accelerated compute optimized instance families: gn6i and gn6e • ECS Bare Metal Instance families: ebmc6, ebmg6, ebmr6, ebmhfc6, ebmhfg6, and ebmhfr6 • Burstable instance families: t6 and t5
②	Region and zone	You only need to specify a region for a regional reserved instance. Regional reserved instances provide zone flexibility and instance size flexibility. You must specify a region and a zone for a zonal reserved instance. Resource reservation is supported. The matching conditions are determined by the type of a reserved instance. For more information, see Match between reserved instances and pay-as-you-go instances. You can modify the region and zone of a reserved instance after you purchase it. For more information, see Modify a reserved instance.
③	Operating system	Reserved instances are classified into Linux and Windows reserved instances. Windows reserved instances can be used to offset image bills of Windows pay-as-you-go instances.
④	Normalization factor	A normalization factor indicates the performance level of an instance type and also the computing power. Normalization factors are determined by the number of vCPUs. For information about detailed specifications, see View normalization factors.

Section	Attribute	Description
⑤	Instance quantity	The instance quantity is used for the following purposes: • Calculates the computing power of reserved instances. • Specifies the number of reserved resources for zonal reserved instances. For example, in the preceding figure, the instance quantity is two and the used instance type is ecs.r6e.xlarge. This indicates that two pay-as-you-go instances of the ecs.r6e.xlarge instance type are reserved.
⑥	Computing power	Reserved instances deliver computing power in advance. Pay-as-you-go instances consume the computing power. Computing power of a reserved instance = Normalization factor of an instance type × Instance quantity. The computing power is used for the following purposes: • Evaluates whether the computing power is the same before and after you split or merge reserved instances. • Evaluates the usage of a reserved instance when the size of a regional reserved instance is different from that of the matched pay-as-you-go instance.
⑦	Term	You must specify the term when you purchase a reserved instance. After you split, merge, and modify reserved instances, terms of the original and new reserved instances are changed accordingly. For more information, see Split a reserved instance, Merge reserved instances, and Modify a reserved instance. Note You cannot shorten the term of a reserved instance to deliver more computing power. A reserved instance takes effect and expires on the hour. For example, you purchased a reserved instance with a term of one year at 13:45 of May 29, 2020. The reserved instance takes effect at 13:00 of May 29, 2020 and expires at 24:00 of May 30, 2021. If you have pay-as-you-go instances that match the reserved instance, the reserved instance is applied to offset bills of pay-as-you-go instances from 13:00 of May 29, 2020 by hour until it expires. Expired reserved instances cannot continue to offset bills of pay-as-you-go instances. However, the pay-as-you-go instances are not released. This ensures your service continuity. Note Make sure that your account balance is sufficient to ensure service availability of your pay-as-you-go instances.

Limits

Limited object	Limited item	Description
Reserved instance	Maximum number of reserved instances	The maximum number is subject to the type of reserved instances. • Maximum number of regional reserved instances: Each account can have up to 20 regional reserved instances in all regions. • Maximum number of zonal reserved instances: Each account can have up to 20 zonal reserved instances in each zone. For example, you can purchase 10 regional reserved instances in China (Hangzhou) and 10 regional reserved instances in China (Qingdao), but the total number of regional reserved instances cannot exceed 20. You can purchase 20 zonal reserved instances in Hangzhou Zone B and 20 zonal reserved instances in Hangzhou Zone H. If you need more reserved instances, submit a ticket.
ECS instance	Billing method	• Reserved instances can match only pay-as-you-go instances. Reserved instances cannot match preemptible instances. • Reserved instances can be used to offset charges for compute resources of pay-as-you-go instances, and cannot be used to offset charges for network and storage resources of pay-as-you-go instances.
	Instance family	The gn6i and t5 instance families do not support regional reserved instances, and support only zonal reserved instances. gn6i and t5 reserved instances cannot be split or merged.

Billing management

Reserved instances support the All Upfront, Partial Upfront, and No Upfront payment options. Billing standards vary depending on the payment option. For more information, see Reserved instance billing.

> **Note** Whether you can use the No Upfront payment option depends on your ECS instance resource usage. If you want to use the No Upfront payment option, submit a ticket.

References

- Match between reserved instances and pay-as-you-go instances
- Purchase reserved instances
- Split a reserved instance
- Merge reserved instances
- Modify a reserved instance

- ECS instance FAQ

5.4.2. Match between reserved instances and pay-as-you-go instances

After you purchase reserved instances, you make commitments to using pay-as-you-go instances for a period of time. Reserved instances can be used to offset bills of pay-as-you-go instances only when they match the pay-as-you-go instances. If you do not have any pay-as-you-go instances in your account, the reserved instances are idle and continue to be charged. This topic describes the matching rules of reserved instances and provides some examples.

Matching rules

The match between reserved instances and pay-as-you-go instances cannot be manually managed. After you purchase a reserved instance, the reserved instance automatically matches one or more pay-as-you-go instances that have some attributes within the term. After the reserved instance is matched, it checks for eligible pay-as-you-go bills on an hourly basis and deducts fees based on its computing power. You can view pay-as-you-go instances that match the reserved instance. For more information, see View matched pay-as-you-go instances.

The matching conditions of a regional reserved instance are different from those of a zonal reserved instance. The following table describes the attributes that determine the match between reserved instances and pay-as-you-go instances.

Attribute	Regional reserved instance	Zonal reserved instance
Region and zone	A regional reserved instance can match one or more pay-as-you-go instances across zones within a specific region.	A zonal reserved instance can match one or more pay-as-you-go instances only in the same zone.
Instance type	The following section describes the instance size flexibility and resource reservation status: • Within the same instance family, regional reserved instances of small instance types can match pay-as-you-go instances of large instance types or regional reserved instances of large instance types can match pay-as-you-go instances of small instance types. • Resource reservation is not supported.	The following section describes the instance size flexibility and resource reservation status: • A zonal reserved instance can match only pay-as-you-go instances of the same instance type. • Resource reservation is supported. A specific number of pay-as-you-go instances of the specified instance types are reserved within the term. This ensures that you can create pay-as-you-go instances anytime in the specified zone.
Operating system	A regional reserved instance can match only pay-as-you-go instances that have the same operating system.	A zonal reserved instance can match only one or more pay-as-you-go instances that have the same operating system.

Examples of regional reserved instances

Regional reserved instances and pay-as-you-go instances must meet the following requirements to match:

- They must reside within the same region.
- They must use instances types that belong to the same instance family. Within the same instance family, regional reserved instances of small instance types can match pay-as-you-go instances of large instance types or regional reserved instances of large instance types can match pay-as-you-go instances of small instance types.

  > **Note** You can evaluate the usage of reserved instances based on the computing power when the instance sizes of the reserved instance and the pay-as-you-go instance are different.

- They must run the same operating system.

The following table describes examples of regional reserved instances.

Scenario	Pay-as-you-go instance	Regional reserved instance	How bills are offset
Regional reserved instances of small instance types match pay-as-you-go instances of large instance types	You have a pay-as-you-go instance that has the following configurations: • Qingdao Zone B • ecs.g5.4xlarge • Linux	You have two active regional reserved instances that have the following attributes: • All zones in China (Qingdao) • ecs.g5.xlarge • Linux • Instance quantity: 2	The following section describes the normalization factors: • The normalization factor of ecs.g5.4xlarge is 16. • The normalization factor of ecs.g5.xlarge is 4. Comparison between the delivered and consumed computing powers: • Pay-as-you-go instance: The pay-as-you-go instance consumes 16 units of computing power per hour (1 instance × 16). • Reserved instance: The reserved instance delivers 8 units of computing power per hour (2 instances × 4). Therefore, one reserved instance offsets 50% of the pay-as-you-go instance bills per hour, and two reserved instances offset 100% of the pay-as-you-go instance bills per hour.

Scenario	Pay-as-you-go instance	Regional reserved instance	How bills are offset
Regional reserved instance of large instance types match pay-as-you-go instances of small instance types	You have six pay-as-you-go instances. • The following section describes the configurations of one of the pay-as-you-go instances: ○ Qingdao Zone B ○ ecs.g5.xlarge ○ Linux • The following section describes the configurations of the other five pay-as-you-go instances: ○ Qingdao Zone C ○ ecs.g5.xlarge ○ Linux	You have two active regional reserved instances. • The following section describes the attributes of one of the reserved instances: ○ All zones in China (Qingdao) ○ ecs.g5.4xlarge ○ Linux ○ Instance quantity: 1 • The following section describes the attributes of the other reserved instance: ○ All zones in China (Qingdao) ○ ecs.g5.2xlarge ○ Linux ○ Instance quantity: 1	The following section describes the normalization factors: • The normalization factor of ecs.g5.xlarge is 4. • The normalization factor of ecs.g5.2xlarge is 8. Comparison between the delivered and consumed computing powers: • Pay-as-you-go instance: The six pay-as-you-go instances consume 24 units of computing power per hour (6 instances × 4). • Reserved instance: One reserved instance delivers 16 units of computing power (1 instance × 16) per hour and the other reserved instance delivers 8 units of computing power (1 instance × 8) per hour. Therefore, the two reserved instances offset 100% of bills of the six pay-as-you-go instances per hour. Note The pay-as-you-go instances belong to two zones. However, the regional reserved instances support the zone flexibility and can offset bills of pay-as-you-go instances across zones.

Scenario	Pay-as-you-go instance	Regional reserved instance	How bills are offset

Failed to match	You have two pay-as-you-go instances.
	• The following section describes the configurations of one of the pay-as-you-go instances:
	  ○ Qingdao Zone B
	  ○ ecs.g5.xlarge
	  ○ Windows
	• The following section describes the configurations of the other pay-as-you-go instance:
	  ○ Hangzhou Zone B
	  ○ ecs.c5.xlarge
	  ○ Linux
		You have two active regional reserved instances.
		• The following section describes the attributes of one of the reserved instances:
		  ○ All zones in China (Qingdao)
		  ○ ecs.g5.4xlarge
		  ○ Linux
		  ○ Instance quantity: 1
		• The following section describes the attributes of the other reserved instance:
		  ○ All zones in China (Qingdao)
		  ○ ecs.g5.xlarge
		  ○ Linux
		  ○ Instance quantity: 1
			The reserved instances failed to match the pay-as-you-go instances due to the following causes:
			• The operating system of one of the pay-as-you-go instances is Windows.
			• One of the pay-as-you-go instances resides in China (Hangzhou) and uses an instance type that belongs to the c5 instance family.
			Therefore, the reserved instances remain idle and continue to be charged. The bills of pay-as-you-go instances are paid by account balance.

Examples of zonal reserved instances

Zonal reserved instances and pay-as-you-go instances must meet the following requirements to match:

- They must reside within the same region and zone.
- They must use instance types that belong to the same instance family and have the same size.
- They must run the same operating system.

The following table describes examples of zonal reserved instances.

Scenario	Pay-as-you-go instance	Zonal reserved instance	How bills are offset
Matched	You have five pay-as-you-go instances that have the following configurations: • Qingdao Zone B • ecs.g5.xlarge • Windows	You have an active zonal reserved instance that has the following attributes: • Qingdao Zone B • ecs.g5.xlarge • Windows • Instance quantity: 5	The reserved instance matches the pay-as-you-go instances. The reserved instance offsets 100% of bills of the five pay-as-you-go instances per hour.

Scenario	Pay-as-you-go instance	Zonal reserved instance	How bills are offset
Resource reservation	You have no pay-as-you-go instances.	You have an active zonal reserved instance that has the following attributes: • Qingdao Zone B • ecs.g5.2xlarge • Linux • Instance quantity: 10	The reserved instance is idle and continues to be charged. However, 10 pay-as-you-go instances of the ecs.g5.2xlarge instance type are reserved for you within the term of the reserved instance. This ensures that you can create pay-as-you-go instances anytime in Qingdao Zone B.
Failed to match	You have two pay-as-you-go instances. • The following section describes the configurations of one of the pay-as-you-go instances: ◦ Qingdao Zone B ◦ ecs.g5.xlarge ◦ Windows • The following section describes the configurations of the other pay-as-you-go instance: ◦ Qingdao Zone C ◦ ecs.g5.4xlarge ◦ Linux	You have two active zonal reserved instances. • The following section describes the attributes of one of the reserved instances: ◦ Qingdao Zone B ◦ ecs.g5.xlarge ◦ Linux ◦ Instance quantity: 1 • The following section describes the attributes of the other reserved instance: ◦ Qingdao Zone B ◦ ecs.g5.xlarge ◦ Linux ◦ Instance quantity: 1	The reserved instances failed to match the pay-as-you-go instances due to the following causes: • The operating system of one of the pay-as-you-go instances is Windows. • One of the pay-as-you-go instances resides in Qingdao Zone C and uses the ecs.g5.4xlarge instance type. Therefore, the reserved instances remain idle and continue to be charged. The bills of pay-as-you-go instances are paid by account balance.

Match between a single reserved instance and multiple pay-as-you-go instances

A single regional or zonal reserved instance can match multiple pay-as-you-go instances. However, you cannot shorten the term of a reserved instance to deliver more computing power.

The following table describes an example that involves six pay-as-you-go instances and an active reserved instance.

Six pay-as-you-go instances	An active reserved instance
The following section describes the configurations of each pay-as-you-go instance: • Qingdao Zone B • ecs.g5.6xlarge • Linux • Normalization factor of the instance type: 24	The following section describes the attributes of the reserved instance: • Qingdao Zone B • ecs.g5.6xlarge • Linux • Normalization factor of the instance type: 24 • Instance quantity: 1 • Term: one year

The six pay-as-you-go instances all match the reserved instance. The following table describes examples of how the reserved instances offset bills of the pay-as-you-go instances based on how long each pay-as-you-go instance exists.

Six pay-as-you-go instances	A reserved instance	How bills are offset
The following section describes the consumed computing power: • Duration of each pay-as-you-go instance: one hour (each consumes 24 units of computing power) • Computing power consumed by the six pay-as-you-go instances per hour: 144 (6 instances × 24)	The following section describes the delivered computing power: Computing power delivered by the reserved instance per hour: 24 (1 instance × 24)	The computing power delivered by the reserved instance is equal to that consumed by a pay-as-you-go instance. The reserved instance offsets the bills of a pay-as-you-go instance randomly. You cannot shorten the term of the reserved instance to two months to offset bills of the six pay-as-you-go instances at the same time.
The following section describes the consumed computing power: • Duration of each pay-as-you-go instance: 10 minutes (each consumes 24/6 units of computing power) • Computing power consumed by the six pay-as-you-go instances per hour: 24 (6 instances × 24/6)	The following section describes the delivered computing power: Computing power delivered by the reserved instance per hour: 24 (1 instance × 24)	The computing power delivered by the reserved instance is equal to that consumed by the six pay-as-you-go instances. The reserved instance offsets bills of the six pay-as-you-go instances.

Six pay-as-you-go instances	A reserved instance	How bills are offset
The following section describes the consumed computing power: • Duration of each pay-as-you-go instance: 15 minutes (each consumes 24/4 units of computing power) • Computing power consumed by the six pay-as-you-go instances per hour: 36 (6 instances × 24/4)	The following section describes the delivered computing power: Computing power delivered by the reserved instance per hour: 24 (1 instance × 24)	The computing power delivered by the reserved instance is less than that consumed by the six pay-as-you-go instances. The reserved instance offsets bills of the six pay-as-you-go instances for an hour. The deducted amount of each pay-as-you-go instance is random.

5.4.3. Purchase reserved instances

This topic describes how to purchase reserved instances in the ECS console.

Prerequisites

- Before you purchase reserved instances, make sure that the pay-as-you-go instances that you want to match meet the requirements for applying reserved instances. For more information, see Reserved instance overview.
- You cannot manually manage how reserved instances and pay-as-you-go instances are matched. Make sure that you understand the matching rules for reserved instances. For more information, see Match between reserved instances and pay-as-you-go instances.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Reserved Instances**.
3. In the top navigation bar, select a region.
4. Click **Purchase Reserved Instance**.
5. Configure region-related parameters.
   i. Select a region.
   ii. Set Resource Reservation.

   > **Note** Only zonal reserved instances support resource reservation. Regional reserved instances apply to pay-as-you-go instances of different sizes in different zones within the same region.

   iii. Select a zone.
6. Configure instance-related parameters.

i. Select an instance type.

> **Note** You must select an instance type when you purchase a regional reserved instance. The regional reserved instance can match any pay-as-you-go instances of the specified instance family within the specified region regardless of size.

ii. Set Operating System Platform.You can select **Linux** or **Windows**.

> **Note** The reserved instance matches only pay-as-you-go instances that use the selected type of operating system. The operating system of an reserved instance cannot be changed after you purchase the reserved instance.

To apply a reserved instance to pay-as-you-go instances created from Bring Your Own License (BYOL) images, submit a ticket.

iii. Set Payment Option.**All Upfront**, **Partial Upfront**, and **No Upfront** are available. For more information, see Reserved instance billing.

7. Configure purchase-related parameters.
   i. (Optional)Set Name.
   ii. Set Term.You can select **1 Month**, **1 Year**, **3 Years**, or **5 Years**.
   iii. Set Quantity. The reserved instance can match the specified number of pay-as-you-go instances of the specified instance type. For example, if the instance type is ecs.g5.large and Quantity is set to 3, the reserved instance can match three pay-as-you-go instances of the ecs.g5.large instance type.
8. Configure tags.You can add tags for multiple types of Alibaba Cloud resources. You can use tags to perform cost sharing, monitoring by group, and automated O&M by group. For more information, see Overview.
9. Read and select *ECS Terms of Service* and then click **Purchase**.
10. In the message that appears, confirm that the parameters are correct and click **Create Order**.
11. Read the payment information and then click **Subscribe**.

Result

After you purchase a reserved instance, you can immediately use it to offset bills when the reserved instance matches one or more pay-as-you-go instances. You can also manage the reserved instance in response to pay-as-you-go instance changes.

Related information

- PurchaseReservedInstancesOffering
- Split a reserved instance
- Merge reserved instances
- Modify a reserved instance

5.4.4. Split a reserved instance

You can split a single reserved instance into multiple reserved instances to match smaller pay-as-you-go instances. For ease of description, the reserved instance to be split is referred to as the original reserved instance. The resulting reserved instances are referred to as the destination reserved instances.

Prerequisites

- The original reserved instance is in the **Active** state.
- No ongoing requests for splitting, merging, or modifying reserved instances exist.

Context

When you split a reserved instance, take note of the following items:

- Reserved instances that belong to the gn6i and t5 instance families cannot be split.
- You can change the instance type of a reserved instance. However, you cannot change the instance family of a reserved instance.
- You cannot change the zone or region of a reserved instance.
- The total computing power of destination reserved instances must be equal to that of the original reserved instance. For more information about computing power, see Match between reserved instances and pay-as-you-go instances. The following figure shows an example of splitting a reserved instance.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images** > **Reserved Instances**.
3. On the **Reserved Instances** page, find the original reserved instance and click **Split** in the **Actions** column.
4. In the **Split Reserved Instance** pane, configure the names, instance types, and quantities of the destination reserved instances.
5. Click **OK**.

Result

After you submit the request for instance splitting, the original reserved instance enters the **Updating** state, and the destination reserved instances in the Creating state are displayed. You cannot cancel the ongoing request for splitting a reserved instance. If you want to roll back the change made by the splitting operation, you can merge the destination reserved instances to obtain the original reserved instance.

After the request for splitting a reserved instance is processed, you can obtain one of the following results:

- If the reserved instance is split:
  - The original reserved instance enters the **Inactive** state and expires on the hour when it is split. The price becomes USD 0.
  - The destination reserved instances enter the **Active** state and take effect on the hour when the original reserved instance is split. If the destination reserved instances are zonal reserved instances, the type of reserved resources is updated automatically.

  - If the destination reserved instances match pay-as-you-go instances, the billing discounts provided by these reserved instances are applied to the matched pay-as-you-go instances starting from the hour when the destination reserved instances take effect.

- If the original reserved instance fails to be split, this reserved instance remains active.

Assume that you split an ecs.g5.2xlarge zonal reserved instance into two ecs.g5.xlarge zonal reserved instances at 20:30 of May 28, 2020. The following figure shows the time when the original reserved instance expires and the time when the destination reserved instances take effect.

Related information

- ModifyReservedInstances
- DescribeReservedInstances

5.4.5. Merge reserved instances

This topic describes how to merge reserved instances. If traffic to your instances increases, you can merge multiple reserved instances into a single reserved instance to match larger pay-as-you-go instances. For ease of description, the reserved instances to be merged are referred to as the original reserved instances. The resulting merged reserved instance is referred to as the destination reserved instance.

Prerequisites

- The original reserved instances are purchased by using the same currency.
- The original reserved instances are in the **Active** state.
- The original reserved instances belong to the same region if they are regional reserved instances. The original reserved instances belong to the same zone if they are zonal reserved instances.
- The original reserved instances have the same operating system and end time.

  > **Note** The end time of a reserved instance is determined when you purchase the reserved instance, and is not necessarily the time when the reserved instance becomes inactive. For example, after you merge the reserved instances, the original reserved instances become inactive. However, the end time of the original reserved instances remains unchanged.

- No ongoing requests for splitting, merging, or modifying reserved instances exist.

Context

When you merge reserved instances, take note of the following items:

- Reserved instances that belong to the gn6i and t5 instance families cannot be merged.
- You can change the instance type of a reserved instance. However, you cannot change the instance family of a reserved instance.
- You cannot change the zone or region of a reserved instance.
- The number of pay-as-you-go instances to which the destination reserved instance can be applied cannot exceed 100.

- The total computing power of the destination reserved instance must be equal to that of the original reserved instances. For more information about computing power, see Match between reserved instances and pay-as-you-go instances. The following figure shows an example of merging reserved instances.

  

- The instance type of the destination reserved instance must be valid. For example, you cannot obtain an ecs.g5.5xlarge reserved instance because the ecs.g5.5xlarge instance type is not included in the g5 instance family, as shown in the following figure.

  

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images** > **Reserved Instances**.
3. On the **Reserved Instances** page, find an original reserved instance and click **Merge** in the **Actions** column.
4. In the **Merge Reserved Instances** pane, select the reserved instances that you want to merge with the current reserved instance. Configure the name, instance type, and quantity of the destination reserved instance.
5. Click **OK**.

Result

After you submit the request for instance merging, the original reserved instances enter the **Updating** state, and the destination reserved instance in the Creating state is displayed. You cannot cancel the ongoing request for merging reserved instances. If you want to roll back the change made by the merging operation, you can split the merged reserved instance to obtain the original reserved instances.

After the request for merging reserved instances is processed, you can obtain one of the following results:

- If the reserved instances are merged:
  - The original reserved instances enter the **Inactive** state and expires on the hour when they are merged. The prices become USD 0.
  - The destination reserved instance enters the **Active** state and takes effect on the hour when the original reserved instances are merged. If the destination reserved instance is a zonal reserved instance, the type of reserved resources is updated automatically.
  - If the destination reserved instance matches pay-as-you-go instances, the billing discount provided by the reserved instance is applied to the matched pay-as-you-go instances starting from the hour when the destination reserved instance takes effect.
- If the original reserved instances fail to be merged, they remain active.

Assume that you merged two ecs.g5.xlarge regional reserved instances into an ecs.g5.2xlarge regional reserved instance at 08:30 of May 29, 2020. The following figure shows the time when the original reserved instances expire and the time when the destination reserved instance takes effect.

Related information

- ModifyReservedInstances
- DescribeReservedInstances

5.4.6. Modify a reserved instance

You can modify the name of a reserved instance. You can modify the scope (zone or region) of a reserved instance to meet your service requirements. For ease of description, the reserved instance to be modified is referred to as the original reserved instance. The resulting modified reserved instance is referred to as the target reserved instance.

region zone resource reservation applicable scope

Prerequisites

- The original reserved instance is in the **Active** state.
- No ongoing requests for splitting, merging, or modifying reserved instances exist.

Context

When you modify the zone or region of a reserved instance, you can make the following changes:

- From a zonal reserved instance to a regional one within the same region
- From a regional reserved instance to a zonal one within the same region
- From one zone to another within the same region

The following figure shows the modified reserved instances.

You cannot modify the scope of a reserved instance across regions. The following figure shows reserved instances that fail to be modified.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Reserved Instances**.
3. On the **Reserved Instances** page, find the original reserved instance and click **Modify** in the **Actions** column.
4. In the **Modify Reserved Instance** pane that appears, reset the name, region, and zone.

   > **Note** If instance types used by reserved instances are insufficient in the target zone, you cannot select this zone.

5. Click **OK**.

Result

After you submit the modification request, the original reserved instance enters the **Updating** state, and the target reserved instance in the Creating state is displayed. You cannot cancel the ongoing request for modifying a reserved instance. If you want to roll back the change made by the modification operation, you can modify the reserved instance again.

After the request for modifying a reserved instance is processed, you can obtain one of the following results:

- If the reserved instance is modified:
  - The original reserved instance enters the **Inactive** state and expires at the top of the hour when it is modified. The price becomes USD 0.
  - The target reserved instance enters the **Active** state and takes effect at the top of the hour when the original reserved instance is modified. If the target reserved instance is a zonal reserved instance, the zone where resources are reserved will also be updated automatically.
  - If the target reserved instance matches one or more pay-as-you-go instances, the billing discount provided by this reserved instance is applied to the matched pay-as-you-go instances starting from the hour when the target reserved instance takes effect.
- If the original reserved instance fails to be modified, it remains active.

Assume that you modified a zonal reserved instance with reserved resources to a regional reserved instance without reserved resources at 10:50 of May 29, 2020. The following figure shows the time when the original reserved instance expires and the time when the target reserved instance takes effect.

Related information

- ModifyReservedInstances
- DescribeReservedInstances

5.4.7. View the usage details of a reserved instance

After you purchase a reserved instance (RI), you can view its matching pay-as-you-go instances and usage details in the ECS console. The bills of pay-as-you-go instances that are offset by RIs are refreshed every hour. This topic describes how to view the usage details of an RI.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images** > **Reserved Instances**.
3. Find the target RI and click **View Bill** in the **Actions** column.
4. On the page that appears, set the search conditions and click **Search**.

5.4.8. View normalization factors

Reserved instances deliver computing power in advance. Pay-as-you-go instances consume the computing power. A normalization factor indicates the performance level of an instance type and also the computing power. This topic describes how to view and download the normalization factor table.

Context

Computing power of a reserved instance = Normalization factor of an instance type × Instance quantity.

The computing power of a reserved instance can be used for the following purposes:

- Evaluates whether the computing power is the same before and after you split or merge reserved instances.
- Evaluates the usage of a reserved instance when the size of a regional reserved instance is different from that of the matched pay-as-you-go instance.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images** > **Reserved Instances**.
3. In the upper-right corner of the **Reserved Instances** page, click **View Normalization Factor Table**.
4. View the normalization factor of each instance type. You can view normalization factors by instance family. You can also click **Download** to save the normalization factor table to your computer for later use.

5.4.9. View matched pay-as-you-go instances

Reserved instances allow you to view the purchased pay-as-you-go instances that the reserved instances match. If pay-as-you-go instances are displayed in the matched instance list of a reserved instance, these pay-as-you-go instances match the reserved instance. However, the reserved instance is not necessarily applied to all these pay-as-you-go instances. You can check your bills for usage details of the reserved instance.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images** > **Reserved Instances**.
3. In the upper-left corner of the **Reserved Instances** page, click **View Reserved Instances**.
4. Select the region and zone where the target reserved instance resides.
5. Find the reserved instance and click **View the matching instances** that corresponds to the reserved instance.

> **Note** The pay-as-you-go instances created by E-MapReduce (EMR) or Alibaba Cloud Container Service for Kubernetes (ACK) are not displayed in the matching list of reserved instances. However, these pay-as-you-go instances can also benefit from billing discounts of reserved instances.

Result

You are directed to the Instances page where the matched pay-as-you-go instances are displayed.

5.5. Switch billing method

5.5.1. Switch the billing method from Pay-As-You-Go to Subscription

This topic describes how to switch the billing method of your instance from Pay-As-You-Go to Subscription in the ECS console. After you create a Pay-As-You-Go instance, you can convert its billing method to Subscription to pay only for the reserved resources.

Subscription Pay-As-You-Go Switch the billing method

Prerequisites

The ECS instance you want to switch the billing method for must meet the following requirements:

- The instance belongs to your account.
- The following instance types are not applicable:
  - Generation I: t1, s1, s2, s3, m1, m2, c1, or c2
  - Generation n1, n2, or e3

  > **Note** For more information about these instance types, see Phased-out instance types.

- The instance cannot be a preemptible instance.
- There is no unpaid switch order for the instance.

  If there is an unpaid switch order, you must cancel the unpaid order and then place another order to switch the billing method.
- Auto release is not set for the instance.

  If auto release has been set for an instance, you must disable the auto release configuration and then switch the billing method. For more information, see Release an instance.
- The instance is in the **Running** or **Stopped** state.

  Example: An order to switch the billing method has been placed when the ECS instance is in the Running or Stopped state. However, the instance status has changed when payment is attempted for the order. The preceding requirement is not met. The order fails and the billing method remains unchanged. You can go to the billing center and pay for the order when the instance is in the Running or Stopped state again.

Context

You can switch a maximum of 20 Pay-As-You-Go instances to Subscription instances each time.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images** > **Instances**.
3. In the top navigation bar, select a region.
4. Select one or more Pay-As-You-Go instances. Click **Switch to Subscription** under the instance list.

5. On the Switch to Subscription page, click **Batch Change**.
6. In the dialog box that appears, set the subscription plan, including:
   i. Duration: You can set the length of service time for the Subscription instance. Instances whose billing methods are converted at the same time must have the same length of service time.
   ii. Data disk: If Pay-As-You-Go data disks are attached to the selected instances, you can set whether to also switch their billing method to Subscription.
7. Click **OK**.
8. Complete the payment as prompted.

5.5.2. Change the billing method of instances from subscription to pay-as-you-go

This topic describes how to change the billing method of an instance from subscription to pay-as-you-go. After you create a subscription instance, you can change its billing method to pay-as-you-go if you want to pay for only the actual usage of your resources. After the change, make sure that you have sufficient funds in your account. Otherwise, overdue payments will affect the running of your services.

Prerequisites

- The instance for which you want to change the billing method is in the **Running** or **Stopped** state.
- The availability of this feature is dynamically determined based on your ECS usage.

Context

The following content shows the billing methods of related resources after the change:

- The billing method of the following resources is changed to pay-as-you-go:
  - The instance
  - The system disk
  - Data disks attached to the instance
- The billing method for network usage remains unchanged.
- After the billing method is changed, the subscription duration that was offered for reasons such as the ICP filing, failure, or migration from on-premises data centers is automatically invalidated.

> **Note** If the No Fees for Stopped Instances (VPC-Connected) feature is enabled and the instance is in the Stopped state before the change, the instance will not automatically enter the No Fees for Stopped Instances (VPC-Connected) state after its billing method is changed to pay-as-you-go. You must manually restart the instance and then stop it for the instance to enter the No Fees for Stopped Instances (VPC-Connected) state.

The following content lists the refund rules for changing the billing method from subscription to pay-as-you-go:

- A refund generated from the billing method change will consume the refund quota. If your

account has reached the refund quota, you cannot apply for another refund until the refund record is cleared on the first day of next month. For more information about the refund quota, see Limits.

The refund amount resulting from the billing method change can be calculated based on the number of vCPUs and the remaining hours in the current billing cycle. Example: 1 refund unit = `1 vCPU × 1 hour`.

Assume that you have purchased a subscription instance equipped with four vCPUs for six months. Four months later, you change the billing method to pay-as-you-go. In this case, the refund amount for this instance is calculated based on the following formula: `4 (vCPUs) × 60 (remaining days) × 24 (hours/day) = 5760 (refund units)`.

- If the instance involves renewal or upgrade orders that have not taken effect, a full refund is made. If the orders have already taken effect, only a partial refund is made.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select a region.
4. Choose one of the following methods to change the billing method of an instance from subscription to pay-as-you-go:
   - Change the billing method of a single instance: Find the target instance, choose **More > Configuration Change > Switch to Pay-As-You-Go** in the **Actions** column.
   - Change the billing method of multiple instances: Select the instances and choose **More > Configuration Change > Switch to Pay-As-You-Go** in the lower part of the page.
5. Read the notes. Read and select *ECS Service Terms* and then click **Switch**.

Result

After the billing method is changed, you can go to the ECS console to view the billing method of the instance:

- On the **Instances** page, the billing method of the instance has been changed to **Pay-As-You-Go** in the **Billing Method** column.
- Click the instance ID to go to the **Instance Details** page. In the left-side navigation pane, click **Disks**. The billing method of the system disk and data disks (if any) is changed to **Pay-As-You-Go** in the **Billing Method** column.

What's next

You can set the auto-release time for the instance to automatically release when you no longer need the instance to stop it from incurring costs. For more information, see Release an instance.

Related information

- ModifyInstanceChargeType

6.Create an instance

6.1. Creation method overview

This topic describes several methods of creating an ECS instance, from basic creation operations to advanced customization operations.

You can create an ECS instance and choose configurations by following the instructions prompted by the wizard on the buy page. For more information, see Create an instance by using the provided 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. For more information, see Purchase an ECS instance of the same configuration.

You can also create a launch template in advance, and then use it to create an instance in one click. For more information, see Create an instance by using a launch template. For more information about launch templates, see Launch templates.

6.2. Create an instance by using the provided wizard

This topic describes how to create an ECS instance by using the wizard in the ECS console. To create an ECS instance, you must specify the instance type, image, storage, network, and security group. The wizard provides a variety of extended configuration features to meet your custom deployment and management requirements.

Prerequisites

You must complete the following preparations to create an ECS instance:

1. Create an account and complete the account information.
   - Create an Alibaba Cloud account.
   - Bind your credit card or PayPal account.
   - To purchase ECS instances in mainland China regions, you must complete real-name verification.
2. Alibaba Cloud provides a default VPC. If you do not want to use the default VPC, you can create a VPC and a VSwitch in the specified region. For more information, see Create an IPv4 VPC network.
3. Alibaba Cloud provides a default security group. If you do not want to use the default security group, you can create a security group in the region where the instance is created. For more information, see Create a security group.

If you need other extended features, you must complete corresponding preparations:

- To specify an SSH key pair when you create a Linux instance, you must create an SSH key pair in the corresponding region. For more information, see Create an SSH key pair.
- To configure user data, you must first prepare user data. For more information about how to prepare user data, see Prepare user data.
- To associate an ECS instance with an instance RAM role, you must create the RAM role, assign a permission policy to the role, and then bind the role to the instance. For more information, see Bind an instance RAM role.

Procedure

1. Go to the Custom Launch tab.
2. Perform the following operations in the **Basic Configurations** step:
   i. Set **Billing Method** to **Subscription**, **Pay-As-You-Go**, or Preemptible Instance. For more information, see Preemptible instances.

   > Note For information about how to create a preemptible instance, see Create a preemptible instance.

   ii. Select a region and zone.

   By default, a zone is randomly assigned by the system. You can select a zone based on your business requirements. For more information about how to select a region and zone, see Regions and zones.

   > Note You cannot change the region or zone after the instance is created.

   iii. Select an instance type and specify the number of instances.

   Instance types that are available are determined by the selected region. You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region. For information about the scenarios for each instance type, see Instance families.

   > Note
   > - The quota of pay-as-you-go or preemptible instances for your account is displayed on the page.
   > - To use elastic network interfaces (ENIs), select an enterprise-level instance type equipped with no less than two vCPUs, or an entry-level instance type equipped with no less than four vCPUs. For more information about the maximum number of ENIs that can be bound to instances of each instance type, see Instance families.
   > - To use a standard SSD, select an I/O optimized instance.

iv. Select an image. You can select an image from Public Image, Custom Image, Shared Image, or Marketplace Image.

> Note
>
> - To use an SSH key pair, you must select a Linux image.
> - To configure user data, you can select only specified images. For more information, see Prepare user data.
> - To use a Red Hat public image, you must make sure that the instance family supports Red Hat images. For more information, see the Which instance families do Red Hat Enterprise Linux (RHEL) images support? section in Image FAQ.
> - Public images contain only initial system environments. You can find more images in Alibaba Cloud Marketplace.

v. Select a storage space.

- **System Disk**: required. You must create a system disk for the operating system. Select a disk category and specify the size for the system disk.
  - Disk category: Categories are available based on the selected region and instance type.
  - Size: The default size of the system disk is 40 GiB. If the selected image file is greater than 40 GiB, the disk size is increased to support the image file. The minimum size of the system disk is related to the image. The actual size is displayed on the buy page.

Image	System disk capacity (GiB)
Linux (excluding CoreOS and Red Hat)	[max{20, image file size}, 500]
FreeBSD	[max{30, image file size}, 500]
CoreOS	[max{30, image file size}, 500]
Red Hat	[max{40, image file size}, 500]
Windows	[max{40, image file size}, 500]

- **Data Disk**: optional. To create a data disk while you are creating an instance, you must select the disk type, and specify the size and quantity of the disk. You must also determine whether to encrypt the disk. For more information, see Encryption overview. You can create an empty data disk or create a data disk from a snapshot. You can add up to 16 data disks.

  > **Note**
  >
  > An data disk that is created together with the instance has the following features:
  >
  > - The billing method of the data disk is the same as that of the instance.
  > - A subscription data disk must be released together with the instance. A pay-as-you-go data disk can be released either separately or together with the instance.

- **NAS File System**: optional. You can select from existing file systems. Up to five file systems can be specified. For more information, see Mount a NAS file system when you purchase an ECS instance.
- If you have selected an instance family that is equipped with local disks (such as i1, d1, or d1ne), the local disk information is displayed. You cannot specify the quantity or category of local disks because these settings depend on the selected instance type. For more information about the local disks that are supported by different instance types, see Instance families.

3. Click **Next: Networking** to configure networking and security groups for the instance.
   i. Select the network type.
      - **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**: If you purchase an ECS instance for the first time after 12:00, June 16, 2016 (UTC+8), you can no longer select the classic network.
   ii. Set the public bandwidth.
      - To assign a public IP address to the instance, you must select **Assign Public IP Address**. Then, select **Pay-By-Traffic** or **Pay-By-Bandwidth** as the billing method for network usage and specify the bandwidth. Public IP addresses that are assigned this way cannot be disassociated from the instance. For more information about the billing methods for network usage, see Billing methods for network usage.
      - If your instance does not need to access the Internet or your VPC-type instance uses an elastic IP address (EIP) to access the Internet, you do not need to assign a public IP address. You can associate an EIP with or disassociate an EIP from an instance at any time.
   iii. Select a security group.If you have not created a security group, you can use the default security group. For more information about rules of the default security group, see Overview.

iv. Add an ENI.If the instance type that you selected supports ENIs, you can add an ENI and specify a VSwitch.

> Note By default, the added ENI is released together with the instance. You can use the ECS console or call the DetachNetworkInterface operation to unbind the ENI from the instance.

4. (Optional)Click **Next: System Configurations** to complete the following configurations:
   i. Select and set logon credentials.Select a credential based on the image:
      - Linux: You can select a password or an SSH key pair as the logon credential.
      - Windows: You can select only a password as the logon credential.

      You can also set the logon credential for an instance after the instance is created. For more information, see Reset the logon password of an instance.
   ii. Specify the instance name that you want to display in the ECS console, and the hostname that can be obtained from within the guest operating system.
   iii. Set advanced options.
      - RAM Role: Assign a RAM role to the instance.
      - User Data: Customize the startup behavior of the instance or pass data into the instance.
5. (Optional)Click **Next: Grouping** to group the instance you have created.
   i. Add tags.If you have created multiple instances, you can use tags to facilitate management. For more information, see Overview.
   ii. Select a deployment set.Deployment sets are designed to manage the deployment of instances. Instances in the same deployment set can be assigned to different physical servers. This policy can ensure high availability of services and is used to support the disaster recovery capability of the infrastructure. For more information, see Create a deployment set.
6. Confirm the order.
   i. In the **Configurations Selected** section, confirm all the configurations. You can also click the Edit icon to change the configurations.
      - Optional. Click **Save as Launch Template** to save your configurations as a launch template that can be used later. For more information, see Launch templates.
      - Optional. Click **View Open API** to view best-practice API scripts. On the left side of the dialog box, the **API Workflow** section describes the API operations related to the current operation and lists the request parameters and their values. On the right side of the dialog box, programming language-specific SDK examples are provided. **Java** and **Python** examples are available. For more information, see API Introduction.
      - Optional. Click **Save as ROS Template** to save your configurations as a ROS template that can be used to create stacks. For more information, see Create a stack.
   ii. (Optional)If the billing method is **Pay-As-You-Go**, you can select **Automatic Release**.
   iii. (Optional)If the billing method is **Subscription**, you can specify the duration and specify whether to select **Enable Auto-renewal**.

iv. Confirm the configuration fees.The following table lists the billing methods for instances and network usage that are used to calculate the fees that you must pay.

Instance billing method	Billing method for network usage	Billed item
Pay-as-you-go or preemptible instance	Pay-by-traffic	Internet traffic fee and configuration fee. The configuration fee consists of fees for the instance type (vCPUs and memory), the system disk, data disks (if any), and local disks (if any).
	Pay-by-bandwidth	The configuration fee consists of fees for the instance type (vCPUs and memory), the system disk, data disks (if any), local disks (if any), and the public bandwidth.
Subscription	Pay-by-bandwidth	The configuration fee consists of fees for the instance type (vCPUs and memory), the system disk, data disks (if any), local disks (if any), and the public bandwidth.
	Pay-by-traffic	Internet traffic fee and configuration fee. The configuration fee consists of fees for the instance type (vCPUs and memory), the system disk, data disks (if any), and local disks (if any).

v. Read and confirm **ECS Terms of Service**.

7. Confirm to create the instance based on the instance billing method.
   - Subscription instance: Click **Create Order**.
   - Pay-as-you-go instance: Click **Create Instance**.

Result

After the instance is activated, click **Console** to view the instance details in the ECS console. On the **Instances** page, you can view the information of the new instance, such as the instance name, public IP address, internal IP address, and private IP address.

What's next

- You can build an FTP site on the ECS instance to upload local files to the instance. For more information, see Manually build an FTP site on a Windows instance.
- To secure your instance after you create the instance, we recommend that you perform security compliance inspection and configuration:
  - For Linux instances, see Harden operating system security for Linux in *Security Advisories*.
  - For Windows instances, see Harden operating system security for Windows in *Security Advisories*.
- If you have added data disks when you create the instance, you must format the partitions before you can use the data disks. For more information, see Format a data disk for a Windows ECS instance or Format a data disk for a Linux instance.

6.3. Create an ECS instance by using a custom image

This topic describes how to use a custom image to create an ECS instance. You can use a custom image to create an ECS instance that has the same operating system, applications, and data as those of the custom image to improve efficiency.

Alibaba Cloud ECS server elastic computing

Prerequisites

A custom image is created under the account and region where you want to create an instance.

Context

If you do not have a custom image in the account and region where you want to create an instance, you can use one of the following solutions.

Scenario	Solution
You have an image on the local device.	Import the local image to Alibaba Cloud as a custom image. For more information, see Image import procedure.
You do not have custom images but have an instance as a template.	Create a custom image from an instance.
You do not have custom images but have a snapshot as a template.	Create a custom image from a snapshot.
You have a custom image in another region.	Copy the custom image to the region where you want to create an instance. For more information, see Copy custom images.
You have a custom image in another account.	Share the custom image to the account under which you want to create an instance. For more information, see Share or unshare custom images.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Images**.
3. In the top navigation bar, select a region.
4. Based on the image source, use one of the following methods to go to the Images page:
   - Custom image created or exported: Go to the **Custom Images** tab.
   - Custom image obtained by copying: Go to the **Custom Images** tab.
   - Custom image obtained by sharing: Go to the **Shared Images** tab.
5. Find the image that you want to use. Click **Create Instance** in the **Actions** column.
6. Configure the parameters and create the instance. Information of the region and image sections is automatically filled. Configure other parameters based on your needs. For more

information, see Create an instance by using the provided wizard.

> **Note** If the selected custom image contains one or more data disk snapshots, an equal number of data disks are automatically created from these snapshots. Each disk has the same size as the snapshot from which the disk is created. You can increase the size of a data disk but cannot decrease it.

Related information

- RunInstances

6.4. Purchase an ECS instance of the same configuration

You can purchase an ECS instance that has the same configuration as that of your existing instance. This allows you to improve the efficiency of creating ECS instances.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select a region.
4. Find the target instance and choose **More > Buy Same Type** in the **Actions** column.
5. Confirm the configuration that is automatically selected.You can modify the configuration that is automatically selected. After you modified the configuration, you can click **Compare the original configuration** in the lower part of the page to view the same and different configurations.
6. Specify time options based on the billing method.
   - Subscription instance: Specify the duration and whether to enable auto-renewal.
   - Pay-as-you-go instance: Specify whether to enable automatic release. If you enabled automatic release, you must specify when to release the instance.
7. Read and select *ECS Terms of Service*.
8. Confirm instance creation based on the billing method.
   - Subscription instance: Click **Create Order**.
   - Pay-as-you-go instance: Click **Create Instance**.

Related information

- RunInstances

6.5. Create an instance by using a launch template

You can use an existing launch template to quickly create ECS instances.

Alibaba Cloud ECS server elastic computing

Prerequisites

Prerequisites

A launch template or a new version of an existing launch template is created. For more information, see Create a launch template and 创建实例启动模板的新版本.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Deployment & Elasticity > Launch Templates**.
3. Find the launch template or version that you want to use. Click **Create Instance** in the **Actions** column.
4. On the **Custom Launch** tab that appears, select the template and version. Check all the configurations after they are loaded.

   > **Note** If you want to modify the parameters, or if the selected template does not contain the required parameters, you can modify the configurations by clicking the Edit icon.

5. Create an instance.
   - To create a subscription instance, set Duration and select ECS Terms of Service and Product Terms of Service. Click **Create Order**.
   - To create a pay-as-you-go instance, select ECS Terms of Service and Product Terms of Service. Click **Create Instance**.

   After the instance is created, you can view its details in the ECS console.

Related information

- RunInstances

6.6. Instructions for purchase

This topic describes the information that you must know before you purchase ECS instances.

Resource upgrades

For information about how to upgrade the configurations of ECS instances, see Overview of instance upgrade and downgrade. Before you upgrade ECS instances, take note of the following items:

- With the exception of ECS instances that use local storage, ECS instances allow their CPU and memory resources to be scaled and their bandwidths to be upgraded while the instances are running. You can also downgrade ECS instances based on your business needs.
- A maximum of 16 data disks can be attached to each ECS instance. You cannot reduce the size of a data disk after it has been extended.
- The bandwidth of each ECS instance is measured in Mbit/s and can range from 0 Mbit/s to 200 Mbit/s. You can downgrade the configurations of an ECS instance when you renew it.

References

The following topics describe the basic terms of ECS and how to use ECS:

- For the terms and services related to ECS, see What is ECS?

- For information about how to select ECS instance families, see Instance families.
- For information about how to select images, see Select an image.
- For the performance of Elastic Block Storage devices, see EBS performance.
- For usage precautions on ECS instances, see Usage notes.
- For the features of pay-as-you-go resources, see Pay-as-you-go.
- For the features of subscription resources, see Subscription.
- For the limits that apply to ECS, see Limits.
- If you want to select instance types based on their benchmark data for compute performance, contact the Service Manager or submit a ticket to contact Alibaba Cloud after-sales support.

7.Connect to instances

7.1. Overview

You can choose from a variety of tools such as VNC and third-party client tools to connect to ECS instances. Select a suitable tool to connect to ECS instances based on the operating systems of the instances, the operating system of your local machine, and the operations that you want to perform.

Connection methods

Instance operating system	Operating system of your local machine	Connection method
Linux	Windows	• Use VNC For more information, see Connect to a Linux instance by using VNC. • Use client tools such as PuTTY ◦ For information about how to connect to an instance by using an SSH key pair as the credential, see Use an SSH key pair to connect to a Linux instance from a Windows device. ◦ For information about how to connect to an instance by using a password as the credential, see Use a username and password for authentication on a Windows device.
	Unix-like operating systems including Linux and macOS	• Use VNC For more information, see Connect to a Linux instance by using VNC. • Use SSH commands ◦ For information about how to connect to an instance by using an SSH key pair as the credential, see Use SSH key pairs in operating systems that support SSH commands (configure the private key file by using commands). ◦ For information about how to connect to an instance by using a password as the credential, see Use a username and password for authentication on a Linux or Mac OS X device.
	Operating systems of mobile devices, including iOS and Android	Use an app such as SSH Control Lite or JuiceSSH For more information, see Connect to a Linux instance from a mobile device.

Instance operating system	Operating system of your local machine	Connection method
Windows	Windows	• Use VNC For more information, see Connect to a Windows instance by using VNC. • Use a client tool such as Remote Desktop Connection (formerly called MSTSC) For more information, see Connect from a local client that runs a Windows operating system.
	Linux	• Use VNC For more information, see Connect to a Windows instance by using VNC. • Use a client tool such as rdesktop For more information, see Connect from a local client that runs a Linux operating system.
	macOS	• Use VNC For more information, see Connect to a Windows instance by using VNC. • Use a client tool such as Microsoft Remote Desktop Connection for Mac For more information, visit Get started with the macOS client.
	Operating systems of mobile devices, including iOS and Android	Use an app such as Microsoft Remote Desktop For more information, see Connect to a Windows instance from a mobile device.

Note

- Except for VNC, all connection tools require that the target instances have public IP addresses or Elastic IP addresses.
- After a Windows instance is created, it takes 2 to 3 minutes to initialize its operating system. Do not restart the instance during initialization. After a non-I/O optimized Windows instance is created, it takes 10 minutes to initialize its operating system. Do not connect to the instance during initialization.

Comparison of connection tools

The following table compares the advantages of VNC and other third-party client tools.

Item	VNC	Third-party client tool
Allocation of a public IP address or an Elastic IP address to the target instance	Optional. VNC can be used to troubleshoot exceptions including network misconfigurations, such as firewall being enabled by mistake.	Required.
Enabling services including SSH on the target instance	Optional. VNC can be used to troubleshoot exceptions including SSH service exceptions, such as SSHD being disabled.	Required.
Logon through the ECS console	Supported.	Not supported. The local client must be installed.
Independence of the target instance operating system	VNC can be used to connect to both Linux and Windows instances.	Third-party client tools can be used to connect to Linux or Windows instances, depending on what client tool is used.
Simultaneous logons by multiple operating system users to a single instance	Not supported.	Depends on the client tool.
Ease of interaction	VNC does not support directly copying and pasting text. To copy and paste text, use the feature of copying long commands.	Depends on the client tool.
Visually viewing Linux system file resources	Not supported.	Depends on the client tool.
Permissions to control and modify hardware	Supported. VNC can be used to manage resources including BIOS and troubleshoot exceptions including system startup failure.	Not supported.
Terminal configurability	Not supported.	Supported, but depends on the client tool.

8.Manage instances

8.1. Start ECS instances

This topic describes how to start instances in the ECS console.

Prerequisites

The instances that you want to start are in the **Stopped** state.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images** > **Instances**.
3. Start instances.
   - To start a single instance, choose **More** > **Instance Status** > **Start** in the **Actions** column corresponding to the target instance.
   - To start multiple instances at a time, select the instances and click **Start** in the lower-left corner of the Instances page.
4. Verify the information and click **OK**.

Result

After instances are started, they enter the **Running** state.

Related information

- StartInstance

8.2. Stop an instance

This topic describes how to stop an instance in the ECS console and introduces operations related to the No Fees for Stopped Instances (VPC-Connected) feature.

stop an instance No Charges After Instance Is Stopped

Prerequisites

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

Context

If you stop a subscription instance, the billing of the instance is not affected.

If you stop a pay-as-you-go instance, the billing of the instance may be affected based on the network type of the instance and the **No Fees for Stopped Instances (VPC-Connected)** feature setting.

- Pay-as-you-go instances in the classic network do not support the **No Fees for Stopped Instances (VPC-Connected)** feature. A pay-as-you-go instance in the classic network continues to incur fees after the instance is stopped. The billing stops only when the instance is released. For more information, see Release an instance.
- Pay-as-you-go instances in VPCs support the **No Fees for Stopped Instances (VPC-Connected)**

feature.

- If the **No Fees for Stopped Instances (VPC-Connected)** feature is not enabled, the billing of a pay-as-you-go instance continues after the instance is stopped.
- If the **No Fees for Stopped Instances (VPC-Connected)** feature is enabled, you can use the Stop Mode parameter to configure whether to retain and bill an instance after the instance is stopped. If you set Stop Mode to No Charges After Instance Is Stopped in the Stop Instance dialog box, billing of the vCPUs, memory, and public IP address stops after the instance is stopped. However, you are still charged for other resources. For more information, see No Fees for Stopped Instances (VPC-Connected).

> **Note** Services that are running on an instance will be interrupted if you stop the instance. Exercise caution when you perform this operation.

Stop a subscription instance

To stop a subscription instance, perform the following steps:

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select a region.
4. Use a suitable method to stop instances.
   - To stop a single instance, find the instance and choose **More > Instance Status > Stop** in the **Actions** column.
   - To stop multiple instances at a time, select the instances and click **Stop** in the lower part of the Instances page.
5. In the Stop Instance dialog box that appears, set **Stopped By**.
   - **Stop**: stops the instance by shutting it down properly.
   - **Force Stop**: forcibly stops the instance. Forcible stop is equivalent to a physical shutdown, and may cause data loss if instance data has not been written to disks.
6. Click **OK**.

Stop a pay-as-you-go instance

To stop a pay-as-you-go instance, perform the following steps:

> **Note** Services that are running on an instance will be interrupted if you stop the instance. Exercise caution when you perform this operation. The procedure to stop a preemptible instance is the same as that of a pay-as-you-go instance. For more information, see Stop a preemptible instance.

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select a region.
4. Use a suitable method to stop instances.
   - To stop a single instance, find the instance and choose **More > Instance Status > Stop** in the **Actions** column.

- To stop multiple instances at a time, select the instances and click **Stop** in the lower part of the Instances page.

5. Configure required parameters based on the instance network type and the **No Fees for Stopped Instances (VPC-Connected)** feature setting.
   - If the network type is classic network or if **No Fees for Stopped Instances (VPC-Connected)** is not enabled:
     a. In the Stop Instance dialog box that appears, set **Stopped By**.
        - **Stop**: stops the instance by shutting it down properly.
        - **Force Stop**: forcibly stops the instance. Forcible stop is equivalent to a physical shutdown, and may cause data loss if instance data has not been written to disks.
     b. Click **OK**.
   - If **No Fees for Stopped Instances (VPC-Connected)** is enabled:
     a. Move the pointer over the icon next to **No Charges After Instance Is Stopped** and read description of the No Fees for Stopped Instances (VPC-Connected) feature.
     b. In the Stop Instance dialog box that appears, set **Stopped By**.
        - **Stop**: stops the instance by shutting it down properly.
        - **Force Stop**: forcibly stops the instance. Forcible stop is equivalent to a physical shutdown, and may cause data loss if instance data has not been written to disks.
     c. Set **Stop Mode**.
        - **Retain Instance and Continue Charging After Instance Is Stopped**: After the instance is stopped, resources of the instance are retained and continue to be charged.
        - **No Charges After Instance Is Stopped**: After the instance is stopped, computing resources such as vCPUs and memory of the instance are not retained or charged. The cloud disks such as system and data disks, Elastic IP address, and bandwidth continue to be charged. The public IP address is reclaimed and the private IP address is retained.
     d. Click **OK**.

> **Note** For information about how to disable No Fees for Stopped Instances (VPC-Connected), see Disable the No Fees for Stopped Instances (VPC-Connected) feature.

Result

The instance enters the **Stopped** state after it is stopped.

Related information

- StopInstance

8.3. Restart an instance

This topic describes how to restart an instance by using the ECS console.

Limits

- Only instances in the **Running** state can be restarted.
- Restarting an instance will stop the instance. As a result, services provided by the instance are disrupted.

Procedure

1.
2.
3.
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**.

You can also call the RebootInstance API action to complete this task.

8.4. Release an instance

Only pay-as-you-go instances (including preemptible instances) and expired subscription instances can be released. This topic describes how to manually and automatically release a pay-as-you-go instance.

release an instance delete an instance remove an instance Alibaba Cloud ECS deleteinstance release a pay-as-you-go instance

Prerequisites

After an instance is released, its data cannot be recovered. We recommend that you create a snapshot to back up data before releasing an instance. For more information, see Create a snapshot.

Context

- For a subscription instance, you can manually release the instance after it expires. If you do not renew the instance after it expires, the instance is automatically released.
- For a pay-as-you-go instance, if the No Fees for Stopped Instances (VPC-Connected) feature is not enabled, charges continue to incur until the instance is released.
- You can enable instance release protection for a pay-as-you-go instance to prevent irreversible data loss resulting from accidental or incorrect operations during a manual release. For more information, see Enable and disable instance release protection.
- If the Release Disk with Instance feature is disabled for the disk attached to an instance, the disk is automatically converted to a pay-as-you-go data disk and retained when the instance is released. For more information, see Release a disk.

Manually release an instance

You can release a pay-as-you-go instance immediately in the console.

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

3. In the top navigation bar, select a region.
4. Configure the release.
   - To release only one instance, find the instance that you want to release and choose **More > Instance Status > Release** in the **Actions** column.

     
   - If you want to release multiple instances, find the pay-as-you-go instances based on the **Billing Method**, select the instances to be released, and click **Release** at the bottom of the list.

     
5. In the dialog box that appears, select **Release Now**.
6. Click **Next**, and then click **OK**.

Enable automatic release

You can enable the automatic release function and set a time to automatically release an instance. If you set the automatic release time multiple times, the latest setting works.

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select a region.
4. Configure the release.
   - To release only one instance, find the instance that you want to release and choose **More > Instance Status > Release** in the **Actions** column.
   - If you want to release multiple instances, find the pay-as-you-go instances based on the **Billing Method**, select the instances to be released, and click **Release** at the bottom of the list.
5. In the dialog box that appears, select **Scheduled Release**.
6. Turn on the automatic release switch, and specify the release date and time.

   > **Note** The automatic release time must be at least 30 minutes later than the current time.

   
7. Click **Next**, and then click **OK**.

Disable automatic release

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select a region.
4. Release configuration.
   - To disable the automatic release function for only one instance, find the instance that you want to release and choose **More > Instance Status > Release** in the **Actions** column.
   - If you want to disable the automatic release function for multiple instances, find the pay-as-you-go instances based on the **Billing Method**, select the instances for which you want

to disable the automatic release function, and click **Release** at the bottom of the list.

5. In the dialog box that appears, select **Scheduled Release**.
6. Turn off the automatic release switch.
7. Click **Next**, and then click **OK**.

Related information

References

- DeleteInstance
- ModifyInstanceAutoReleaseTime

8.5. Reactivate an instance

This topic describes how to reactivate an instance. After paying the overdue bill of a Pay-As-You-Go instance, you must reactivate the instance. Otherwise, the instance will be released.

Precautions

For a Pay-As-You-Go instance, if the due date of an overdue payment (T) is not settled within 15 days after the due date (T+15), the instance is stopped due to overdue payment and its status changes to **Expired**. You must submit a ticket to settle the payment and reactivate your instance within 30 days after the due date (T+30). Otherwise, the instance is released and the data cannot be recovered.

> **Note** If you fail to reactivate the ECS instance within 30 days after the due date (T+30), the instance is automatically released 30 days after the due date and the data cannot be recovered.

Prerequisites

The Pay-As-You-Go instance is in the **Expired** or **Expired and Being Recycled** state.

You have settled the payment by opening a ticket.

Procedure

To reactivate an instance in the ECS console, follow these steps:

1.
2.
3.
4. Select the instance to be reactivated, and then choose **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** state after about 10 minutes.

Alternatively, you can complete this task by calling the ReactivateInstances API action through the Alibaba Cloud CLI, OpenAPI Explorer, or SDK.

8.6. View instance information

This topic describes how to view an overview of instances and details of a single instance under your account.

Alibaba Cloud ECS polling view instance information before submitting a ticket

View instance resources on the Overview page

When you log on to the ECS console, the **Overview** page appears.

On the **Overview** page, you can view the following information of ECS instances under your account, including:

- **Pending Events**: lists all pending events and instances that are associated with the events.
- **My Resources**: lists ECS instances and other resources in each region.

View information of instances on the Instances page

Perform the following steps to go to the **Instances** page:

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images** > **Instances**.
3. In the top navigation bar, select a region.
4. View the information of all ECS instances in the specified region, such as **Instance ID/Name**, **Zone**, **IP Address**, **Status**, **Network Type**, **Billing Method**, and **Actions**. You can use **Column Filters** to customize your columns:
   i. In the upper-right corner of the **Instances** page, click the

   icon.
   ii. In the **Column Filters** dialog box that appears, select the instance information that you want to view and click **OK**.

   

View information of a single instance on the Instance Details page

You can perform the following steps to view the information of a single instance on the **Instance Details** page:

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images** > **Instances**.
3. In the top navigation bar, select a region.
4. Find the ECS instance that you want to view and click the instance ID or **Manage** in the corresponding **Actions** column.

Note The following table describes the information of an instance on the Instance Details page.

Information	Description
Basic Information	The information that is related to instance identification, such as instance ID, name, zone, region, instance type, instance family, image ID, and key pair name (applicable only to Linux instances).
Configuration Information	The information that is related to instance configuration, such as vCPU, memory, instance type, operating system, IP address, bandwidth billing method, current bandwidth, and VPC (applicable only to VPC instances).
Payment Information	The information that is related to billing, such as billing method, stop mode, creation time, and automatic release time (applicable only to pay-as-you-go instances).
Monitoring Information	The information that is related to instance running, such as CPU and network usage.

You can switch from the **Instance Details** page to the **Disks**, **Snapshots**, or **Security Groups** page to view other types of resources of the instance.

Related information

- DescribeInstances

8.7. Modify the properties of an instance

After an instance is created, you can modify its name, hostname, and description. If the instance is a pay-as-you-go instance, you can also enable or disable release protection for the instance.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select a region.
4. Find the target instance, and choose **More > Instance Settings > Modify Instance Properties** in the **Actions** column.
5. Modify the properties of the instance.
6. Click **OK**.
7. If the hostname of the instance is modified, restart the instance for the new hostname to take effect.

Notice To restart an ECS instance, you must use the ECS console or call the RebootInstance operation. You cannot restart an instance from within the operating system.

Related information

- ModifyInstanceAttribute

8.8. Customize CPU options

The CPU options of an Elastic Compute Service (ECS) instance include the number of physical CPU cores and the number of threads per core. For some ECS instance types, you can customize these options when you use the RunInstances operation to create an instance.

CPU and vCPU

CPUs are central processing units. A single CPU can contain several physical cores. Hyper-Threading (HT) Technology can be used to create two virtual processing cores for each physical core that is present in a CPU. Virtual CPUs (vCPUs) are virtual processing cores of ECS instances.

Alibaba Cloud ECS supports multi-threading based on Intel® HT Technology. HT Technology enables two threads to run concurrently on a single physical core. Each thread can be considered as a vCPU. For more information, visit Intel® Hyper-Threading Technology.

The following table describes the CPU options for ECS instances.

CPU option	API parameter	Function	Scenario	Applicable instance type
Number of physical CPU cores	CpuOptions.Core	Determines the number of physical CPU cores to use.	You can use a smaller number of physical CPU cores to improve the CPU-to-memory ratio of the instance. This reduces the number of chargeable items and software licensing costs.	For more information, see Valid values for the number of physical CPU cores and the number of threads per core.
Number of threads per core	CpuOptions.ThreadsPerCore	Determines whether to enable Hyper-Threading on the CPU. Number of vCPUs = Number of physical CPU cores × Number of threads per core.	In most cases, the default configuration of an ECS instance type provides sufficient performance. You can disable Hyper-Threading in the following scenarios: • High-performance computing (HPC) scenarios. In these scenarios, you can improve the performance of instances by disabling Hyper-Threading. • Memory intensive business scenarios. You can disable Hyper-Threading to reduce the number of vCPUs and increase the CPU-to-memory ratio. This can also decrease the number of chargeable items and reduce software licensing costs.	

Billing methods

You can customize CPU options at no additional costs.

Limits

- The following instance families support custom CPU options. For more information about the CPU options of each instance family, see Valid values for the number of physical CPU cores and the number of threads per core.
  - hfg7, hfc7, and hfr7
  - g6e, c6e, and r6e
  - g6, c6, and r6
  - hfg6, hfc6, and hfr6
- CPU options can be customized only when you create an ECS instance. You cannot modify CPU options after the instance is created.
- If you upgrade or downgrade the configurations of an instance, the custom CPU options are changed to the default CPU options of the new instance type.
- The instance type of an instance determines the number of physical cores available in the instance. You can specify the number of physical CPU cores to be enabled within the defined range.

Enable or disable Hyper-Threading

You can call the RunInstances operation to customize the CPU options of an ECS instance. If you want to use an Alibaba Cloud ECS SDK, upgrade the SDK to the latest version.

- By default, Hyper-Threading is enabled on ECS instances. You can enable Hyper-Threading in Alibaba Cloud CLI, as shown in the following sample request:

```
aliyun ecs RunInstances --RegionId cn-hangzhou --CpuOptions.Core 2 --CpuOptions.ThreadsPerCore
2 --ImageId ubuntu_18_04_64_20G_alibase_20190624.vhd --InstanceType ecs.g6.6xlarge --SecurityGro
upId sg-bp67acfmxazb4ph*** --VSwitchId vsw-bp1s5fnvk4gn2tws03*** --Amount 1 --SystemDisk.Auto
SnapshotPolicyId sp-bp67acfmxazb4ph***
```

- To disable Hyper-Threading, set the CpuOptions. ThreadsPerCore parameter to *1* in Alibaba Cloud CLI, as shown in the following sample request:

```
aliyun ecs RunInstances --RegionId cn-hangzhou --CpuOptions.Core 2 --CpuOptions.ThreadsPerCore
1 --ImageId ubuntu_18_04_64_20G_alibase_20190624.vhd --InstanceType ecs.g6.6xlarge --SecurityGro
upId sg-bp67acfmxazb4ph*** --VSwitchId vsw-bp1s5fnvk4gn2tws03*** --Amount 1 --SystemDisk.Auto
SnapshotPolicyId sp-bp67acfmxazb4ph***
```

For example, the ecs.g6.xlarge instance type provides two physical CPU cores by default.

- To enable Hyper-Threading, you can set the number of threads per core to 2. Then, the number of vCPUs is 4, which is the product of the number of CPU cores and the number of threads per core. Hyper-Threading is enabled by default for this instance type.
- To disable Hyper-Threading, you can set the number of threads per core to 1. Then, the number of vCPUs is 2.

View CPU options

You can call the DescribeInstances operation to view the CPU options of an ECS instance. If you want to use an Alibaba Cloud ECS SDK, upgrade the SDK to the latest version.

- Sample request in Alibaba Cloud CLI:

```
aliyun ecs DescribeInstances --InstanceIds '["i-bp19rxmzeocge2z57***"]' --output cols=CpuOptions rows=Instances.Instance[]
```

Sample response:

```
CpuOptions
----------
map[CoreCount:1 Numa: ThreadsPerCore:2]
```

- Sample code by using the shell command **lscpu**:

```
shell@ecshost:~$ lscpu
Architecture:       x86_64
Byte Order:         Little Endian
CPU(s):             1   # Indicates the number of physical CPU cores.
On-line CPU(s) list: 0
Thread(s) per core:  2   # Indicates the number of threads per core.
Core(s) per socket:  1
Socket(s):          1
Vendor ID:          GenuineIntel
CPU family:         6
Model:              85
Model name:         Intel(R) Xeon(R) Platinum 8163 CPU @ 2.50GHz
......
```

Valid values for the number of physical CPU cores and the number of threads per core

The following tables list the default values and valid values for the number of physical CPU cores (CpuOptions.Core) and the number of threads per core (CpuOptions.ThreadsPerCore). Instance types that are not listed in the tables do not support custom CPU options.

Default values and valid values for the number of physical CPU cores and the number of threads per core of the hfg7 instance family

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.hfg7.large	2	1	2	1, 2
ecs.hfg7.xlarge	4	2	2	1, 2
ecs.hfg7.2xlarge	8	2, 4	2	1, 2
ecs.hfg7.3xlarge	12	2, 4, 6	2	1, 2

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.hfg7.4xlarge	16	2, 4, 6, 8	2	1, 2
ecs.hfg7.6xlarge	24	2, 4, 6, 8, 10, 12	2	1, 2
ecs.hfg7.8xlarge	32	2, 4, 6, 8, 10, 12, 14, 16	2	1, 2
ecs.hfg7.12xlarge	48	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24	2	1, 2
ecs.hfg7.24xlarge	96	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48	2	1, 2

Default values and valid values for the number of physical CPU cores and the number of threads per core of the hfc7 instance family

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.hfc7.large	2	1	2	1, 2
ecs.hfc7.xlarge	4	2	2	1, 2
ecs.hfc7.2xlarge	8	2, 4	2	1, 2
ecs.hfc7.3xlarge	12	2, 4, 6	2	1, 2
ecs.hfc7.4xlarge	16	2, 4, 6, 8	2	1, 2
ecs.hfc7.6xlarge	24	2, 4, 6, 8, 10, 12	2	1, 2
ecs.hfc7.8xlarge	32	2, 4, 6, 8, 10, 12, 14, 16	2	1, 2
ecs.hfc7.12xlarge	48	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24	2	1, 2

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.hfc7.24xlarge	96	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48	2	1, 2

Default values and valid values for the number of physical CPU cores and the number of threads per core of the hfr7 instance family

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.hfr7.large	2	1	2	1, 2
ecs.hfr7.xlarge	4	2	2	1, 2
ecs.hfr7.2xlarge	8	2, 4	2	1, 2
ecs.hfr7.3xlarge	12	2, 4, 6	2	1, 2
ecs.hfr7.4xlarge	16	2, 4, 6, 8	2	1, 2
ecs.hfr7.6xlarge	24	2, 4, 6, 8, 10, 12	2	1, 2
ecs.hfr7.8xlarge	32	2, 4, 6, 8, 10, 12, 14, 16	2	1, 2
ecs.hfr7.12xlarge	48	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24	2	1, 2
ecs.hfr7.24xlarge	96	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48	2	1, 2

Default values and valid values for the number of physical CPU cores and the number of threads per core of the g6e instance family

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.g6e.large	2	1	2	1, 2
ecs.g6e.xlarge	4	2	2	1, 2
ecs.g6e.2xlarge	8	2, 4	2	1, 2
ecs.g6e.4xlarge	16	2, 4, 6, 8	2	1, 2
ecs.g6e.8xlarge	32	2, 4, 6, 8, 10, 12, 14, 16	2	1, 2
ecs.g6e.13xlarge	52	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26	2	1, 2
ecs.g6e.26xlarge	104	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52	2	1, 2

Default values and valid values for the number of physical CPU cores and the number of threads per core of the c6e instance family

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.c6e.large	2	1	2	1, 2
ecs.c6e.xlarge	4	2	2	1, 2
ecs.c6e.2xlarge	8	2, 4	2	1, 2
ecs.c6e.4xlarge	16	2, 4, 6, 8	2	1, 2
ecs.c6e.8xlarge	32	2, 4, 6, 8, 10, 12, 14, 16	2	1, 2
ecs.c6e.13xlarge	52	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26	2	1, 2

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.c6e.26xlarge	104	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52	2	1, 2

Default values and valid values for the number of physical CPU cores and the number of threads per core of the r6e instance family

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.r6e.large	2	1	2	1, 2
ecs.r6e.xlarge	4	2	2	1, 2
ecs.r6e.2xlarge	8	2, 4	2	1, 2
ecs.r6e.4xlarge	16	2, 4, 6, 8	2	1, 2
ecs.r6e.8xlarge	32	2, 4, 6, 8, 10, 12, 14, 16	2	1, 2
ecs.r6e.13xlarge	52	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26	2	1, 2
ecs.r6e.26xlarge	104	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52	2	1, 2

Default values and valid values for the number of physical CPU cores and the number of threads per core of the g6 instance family

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.g6.large	2	1	2	1, 2
ecs.g6.xlarge	4	2	2	1, 2

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.g6.2xlarge	8	2, 4	2	1, 2
ecs.g6.3xlarge	12	2, 4, 6	2	1, 2
ecs.g6.4xlarge	16	2, 4, 6, 8	2	1, 2
ecs.g6.6xlarge	24	2, 4, 6, 8, 10, 12	2	1, 2
ecs.g6.8xlarge	32	2, 4, 6, 8, 10, 12, 14, 16	2	1, 2
ecs.g6.13xlarge	52	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26	2	1, 2
ecs.g6.26xlarge	104	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52	2	1, 2

Default values and valid values for the number of physical CPU cores and the number of threads per core of the c6 instance family

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.c6.large	2	1	2	1, 2
ecs.c6.xlarge	4	2	2	1, 2
ecs.c6.2xlarge	8	2, 4	2	1, 2
ecs.c6.3xlarge	12	2, 4, 6	2	1, 2
ecs.c6.4xlarge	16	2, 4, 6, 8	2	1, 2
ecs.c6.6xlarge	24	2, 4, 6, 8, 10, 12	2	1, 2
ecs.c6.8xlarge	32	2, 4, 6, 8, 10, 12, 14, 16	2	1, 2
ecs.c6.13xlarge	52	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26	2	1, 2

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.c6.26xlarge	104	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52	2	1, 2

Default values and valid values for the number of physical CPU cores and the number of threads per core of the r6 instance family

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.r6.large	2	1	2	1, 2
ecs.r6.xlarge	4	2	2	1, 2
ecs.r6.2xlarge	8	2, 4	2	1, 2
ecs.r6.3xlarge	12	2, 4, 6	2	1, 2
ecs.r6.4xlarge	16	2, 4, 6, 8	2	1, 2
ecs.r6.6xlarge	24	2, 4, 6, 8, 10, 12	2	1, 2
ecs.r6.8xlarge	32	2, 4, 6, 8, 10, 12, 14, 16	2	1, 2
ecs.r6.13xlarge	52	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26	2	1, 2
ecs.r6.26xlarge	104	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52	2	1, 2

Default values and valid values for the number of physical CPU cores and the number of threads per core of the hfg6 instance family

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.hfg6.large	2	1	2	1, 2
ecs.hfg6.xlarge	4	2	2	1, 2
ecs.hfg6.2xlarge	8	2, 4	2	1, 2
ecs.hfg6.3xlarge	12	2, 4, 6	2	1, 2
ecs.hfg6.4xlarge	16	2, 4, 6, 8	2	1, 2
ecs.hfg6.6xlarge	24	2, 4, 6, 8, 10, 12	2	1, 2
ecs.hfg6.8xlarge	32	2, 4, 6, 8, 10, 12, 14, 16	2	1, 2
ecs.hfg6.10xlarge	40	2, 4, 6, 8, 10, 12, 14, 16, 18, 20	2	1, 2
ecs.hfg6.16xlarge	64	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32	2	1, 2
ecs.hfg6.20xlarge	80	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40	2	1, 2

Default values and valid values for the number of physical CPU cores and the number of threads per core of the hfc6 instance family

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.hfc6.large	2	1	2	1, 2
ecs.hfc6.xlarge	4	2	2	1, 2
ecs.hfc6.2xlarge	8	2, 4	2	1, 2
ecs.hfc6.3xlarge	12	2, 4, 6	2	1, 2
ecs.hfc6.4xlarge	16	2, 4, 6, 8	2	1, 2
ecs.hfc6.6xlarge	24	2, 4, 6, 8, 10, 12	2	1, 2

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.hfc6.8xlarge	32	2, 4, 6, 8, 10, 12, 14, 16	2	1, 2
ecs.hfc6.10xlarge	40	2, 4, 6, 8, 10, 12, 14, 16, 18, 20	2	1, 2
ecs.hfc6.16xlarge	64	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32	2	1, 2
ecs.hfc6.20xlarge	80	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40	2	1, 2

Default values and valid values for the number of physical CPU cores and the number of threads per core of the hfr6 instance family

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.hfr6.large	2	1	2	1, 2
ecs.hfr6.xlarge	4	2	2	1, 2
ecs.hfr6.2xlarge	8	2, 4	2	1, 2
ecs.hfr6.3xlarge	12	2, 4, 6	2	1, 2
ecs.hfr6.4xlarge	16	2, 4, 6, 8	2	1, 2
ecs.hfr6.6xlarge	24	2, 4, 6, 8, 10, 12	2	1, 2
ecs.hfr6.8xlarge	32	2, 4, 6, 8, 10, 12, 14, 16	2	1, 2
ecs.hfr6.10xlarge	40	2, 4, 6, 8, 10, 12, 14, 16, 18, 20	2	1, 2
ecs.hfr6.16xlarge	64	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32	2	1, 2

Instance type	Default value for the number of vCPUs	Valid values for the number of physical CPU cores	Default value for the number of threads per core	Valid values for the number of threads per core
ecs.hfr6.20xlarge	80	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40	2	1, 2

8.9. Modify instance maintenance attributes

By default, Alibaba Cloud automatically restarts an ECS instance if an automatic recovery event occurs for the instance, such as when the instance goes down unexpectedly or when an active O&M plan is carried out. You can customize the automatic recovery mode for an instance by modifying its maintenance attribute.

Context

For more information about automatic recovery and the applicable scope and impact of each maintenance attribute, see Automatic recovery events of instances.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select a region.
4. Find the target instance and use one of the following methods to modify the instance maintenance attribute.
   - Method 1: Choose **More > Operations and Troubleshooting > Modify Instance Maintenance Attribute** in the **Actions** column corresponding to the instance.

     
   - Method 2: Click the instance ID to go to the instance details page. In the **Basic Information** section, choose **More > Modify Instance Maintenance Attribute**.

     

5. In the **Modify Instance Maintenance Attribute** dialog box that appears, modify the instance maintenance attribute. Click **OK**.
   - If only cloud disks are attached to the instance, you can select one of the following options for Maintenance Action:
     - Automatically Restart
     - Stop
   - If local disks are also attached to the instance, you can select one of the following options for Maintenance Action:
     - Automatically Restart

- Stop
- Automatically Redeploy

6. In the **Basic Information** section of the **instance details** page, confirm the setting that you made for **Maintenance Attribute**.

Related information

- ModifyInstanceMaintenanceAttributes
- DescribeInstanceMaintenanceAttributes

8.10. Reset the logon password of an instance

Change a password

This topic describes how to reset the logon password of an ECS instance. This topic is applicable to scenarios such as when you forget the password or when you did not set a password during instance creation.

the credential does not work when you connect to the instance forget the password change the password retrieve the password the username and password of the remote desktop

Prerequisites

The instance is in a stable state, such as **Stopped** or **Running**. For more information about the status of an instance, see ECS instance lifecycle.

Context

- After you reset the logon password of an instance that is in the **Running** state, you must restart the instance to make the new password take effect. Restarting the instance may affect your services. We recommend that you reset the logon password during off-peak hours to avoid service disruption.
- If the instance is a Linux instance, you can log on to the instance by using the password or the key pair. If you only use the password for authentication, this authentication method becomes invalid after you attach a key pair to your instance. If you want to use both methods to log on to the instance, you must reset its logon password.
- You can also change the logon password of an instance by connecting to the instance. The change takes effect immediately and you do not need to restart the instance. For more information, see Change the logon password of an instance by connecting to the instance.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select a region.
4. Based on the number of instances for which you want to change passwords, you can perform different operations.
   - To reset the password of a single instance, find the target instance and choose **More > Password/Key Pair > Reset Password** in the **Actions** column.

- To reset the password of multiple instances, select the target instances and click **Reset Password** below the instance list.

5. In the **Reset Password** dialog box that appears, enter a valid new password and click **Submit**.
6. Based on the instance status, you can perform one of the following operations to make the new password take effect:
   - If the instance is in the **Running** state, click **Restart Now**.
   - If the instance is in the **Stopped** state, click **Cancel** and manually restart the instance.

     If you click **Restart Now**, an error will be displayed, indicating that this operation is not supported while the instance is in the current state. However, the password is already reset and will take effect the next time the instance starts.

Related information

- ModifyInstanceAttribute

8.11. Change the logon password of an instance by connecting to the instance

You can change the logon password of an instance without going to the console when you operate on the instance. The change takes effect immediately and you do not need to restart the instance. This topic describes how to change the logon password of an ECS Linux instance and an ECS Windows instance by connecting to the instance.

Change the logon password of a Linux instance

This topic takes a Linux instance that runs CentOS 7.6 as an example. Perform the following steps to 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 that you can use to connect to a Linux instance, see Overview.
2. Run the **passwd username** command (for example, **passwd root**).
3. Enter a new password.
4. Enter the new password again to confirm the password.

Change the logon password of a Windows instance

This topic uses a Windows instance that runs Windows Server 2012 as an example. Perform the following steps to 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 that you can use to connect to a Windows instance, see Overview.
2. Choose **Start** > **Run**, type **compmgmt.msc**, and press the Enter key.
3. In the **Computer Management** window, choose **System Tools** > **Local Users and Groups** > **Users**.
4. Right-click the username (for example, **Administrator**) for which you will change the password.

5. Click **Set Password**.
6. In the **Set Password for Administrator** dialog box, click **Proceed**. In the dialog box that appears, enter a new password in the **New password** and **Confirm password** fields, and then click **OK**.

Related information

- Restart an instance
- RebootInstance

8.12. Enable or disable release protection for ECS instances

You can enable release protection for pay-as-you-go instances to prevent potential irreversible consequences arising from accidental manual instance release. This topic describes how to enable and disable release protection for ECS instances, how to check whether release protection is enabled, and how release protection is implemented.

Prerequisites

The target instance is a pay-as-you-go instance.

Context

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

- A payment in your account is overdue for more than 15 days.
- The automatic release time that you set for the instance has been reached.
- The instance does not comply with the applicable security compliance policies.
- The instance was automatically created by Auto Scaling and is removed by subsequent scale-in events.

The following examples show how release protection is implemented:

- When you attempt to manually release instances in the ECS console, instances with release protection enabled are automatically skipped.

  

- When you attempt to manually release instances in the ECS console, the selected instances cannot be released if they all have release protection enabled.

  

- The `InvalidOperation.DeletionProtection` error code is returned if you attempt to call the DeleteInstance operation to release an instance with release protection enabled.

Enable release protection when you create an instance

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

1. Log on to the ECS console.

2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select a region.
4. On the **Instances** page, click **Create Instance**.
5. In the **Basic Configurations** step, set **Billing Method** to **Pay-As-You-Go** and complete the remaining configurations. Click **Next: Networking**.
6. In the **Networking** step, complete all configurations. Click **Next: System Configurations**.
7. In the **System Configurations** step, select **Prevent users from releasing the instance inadvertently by using the console or API** and complete the remaining configurations. Click **Next: Grouping**.

   
8. Complete the remaining configurations until the instance is created.

When you call the RunInstances or CreateInstance operation to create an instance, you can enable or disable release protection for the instance by setting the DeletionProtection parameter.

Change the release protection settings

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

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select a region.
4. On the **Instances** page, use one of the following methods to change the release protection settings of instances:
   - Change the release protection setting of a single instance: Find the instance for which you want to change the release protection setting, and choose **More > Instance Settings > Change Release Protection Setting** in the **Actions** column.
   - Change the release protection settings of one or more instances: Select multiple instances and choose **More > Instance Settings > Change Release Protection Setting** in the lower part of the Instances page.
5. In the Change Release Protection Setting dialog box that appears, turn on or turn off **Release Protection**.
6. Click **OK**.

When you call the ModifyInstanceAttribute operation to modify the attributes of an instance, you can enable or disable release protection for the instance by setting the DeletionProtection parameter.

Check whether release protection is enabled

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select a region.
4. On the **Instances** page, use one of the following methods to view details of an instance:
   - In the **Instance ID/Name** column, click the ID of the instance.

- Find the instance and click **Manage** in the **Actions** column.

5. On the **Instance Details** page, check whether release protection is enabled in the **Enable release protection** line of the **Payment Information** section.

Related information

References

- DeleteInstance
- RunInstances
- CreateInstance
- ModifyInstanceAttribute

8.13. Run remote commands

You can run remote commands to perform O&M operations on an instance without having to log on to the instance.

Prerequisites

- The target instance is in the **Running** state.
- The Cloud Assistant client is started. For more information, see Start or stop the Cloud Assistant client.

Context

- Remote commands can only be run on a single instance at a time.
- You must use Cloud Assistant to run remote commands, and remote commands consume the quota of Cloud Assistant commands. For more information about Cloud Assistant and its limits, see Overview.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select a region.
4. Click the ID of the target instance to go to the details page of the instance.
5. In the left-side navigation pane, click **Remote Commands**.
6. On the page that appears, click **Send Remote Commands**.
7. In the **Send Remote Commands** dialog box that appears, perform the following operations:
   i. Select a command type.
      - Linux instances: **Shell** is selected by default.
      - Windows instances: Select **Bat** or **PowerShell**.

ii. Specify whether to retain commands.

> Note You can view the retained commands on the Cloud Assistant page and run these commands repeatedly. For more information about how to use Cloud Assistant to run remote commands on ECS instances, see Run commands.

iii. Enter a command in the **Command Content** section.

> Note
>
> - The command must be able to return results of a single execution. Interactions with returned information are not allowed.
> - For more information about shell commands, see View instance configurations.

iv. Click **Run**.

- You can click **Stop Task** to cancel running a command during execution.
- After the execution is complete, you can view the response in the **Command Output** section.

> Note After an execution is complete, you can enter another command in the **Command Content** section to run it.

8.14. User data

8.14.1. Prepare user data

User data allows you to customize the startup behaviors of ECS instances and pass data into instances.

Context

Both Windows and Linux instances support the user data feature. You can use the feature in the following ways:

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

Prepare the user data of a Linux instance

You can configure the user data of a Linux instance by using multiple types of scripts such as the User-Data script, Cloud Config script, Include file, Gzip-compressed script, and Upstart Job script. User data of Linux instances adopts the open-source cloud-init architecture and takes metadata of instances as data sources to automatically configure attributes of Linux instances. For more information, visit User-Data Formats in the cloud-init documentation.

If you are creating an Include file or Gzip-compressed script, you must upload script files to available storage services, obtain the links, and set the validity period for the links. We recommend that you use Alibaba Cloud OSS to create links. For more information, see Upload objects and Configure lifecycle rules in the OSS documentation.

Perform the following operations to prepare the user data of a Linux instance:

1. Use an editor to create a text file, such as Notepad++.
2. Edit the script that is related to the user data in the text file. The first line of a script must meet the format requirements of this script type. The following section describes the format requirements of different scripts and the format examples:
   - User-Data script

     User-Data scripts can be shell scripts. The following section describes the characteristics of a User-Data script:

     - The first line must start with `#!`, such as `#! /bin/sh`.
     - The script size cannot exceed 16 KB before the script is encoded in Base64.
     - The script is run only when the instance is started for the first time.

     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 is created, start and connect to the instance. Run the `cat [file]` command to check the execution result of the script.

     ```
     [root@XXXXX2z ~]# cat output10.txt
     Hello World. The time is now Mon, 24 Jul 2017 13:03:19 +0800!
     ```

   - Cloud Config script

     Cloud Config scripts are a user-friendly and efficient way to implement user data. You can use Cloud Config scripts to configure services such as updating YUM sources, importing SSH keys, and installing dependency packages. The following section describes the characteristics of a Cloud Config script:

     - The first line must start with `#cloud-config`, and the header cannot have spaces.
     - The script must be a valid YAML file.
     - The script size cannot exceed 16 KB before the script is encoded in Base64.
     - The running frequency of the user data varies based on the service that you configured.

     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 is created, start and connect to the instance. Check the execution result of the script.

- Include file

  An Include file consists of script links. Each line consists of a link. When an ECS instance is started, cloud-init reads user data from script links of the Include file. If an error occurs on a line, the instance stops reading the user data. The following section describes the characteristics of an Include file:

  - The first line must start with `#include`, and the header cannot have spaces.
  - The file size cannot exceed 16 KB before the file is encoded in Base64.
  - The running frequency of the user data varies based on the type of the script that is configured in the Include file.

  Include file example:

  ```
  #include
  http://ecs-image-test.oss-cn-hangzhou.aliyuncs.com/UserData/myscript.sh
  ```

  After the instance is created, start and connect to the instance. Check the execution result of the script.

- Gzip-compressed script

  When you use the User-Data script, Cloud Config script, or Include file, make sure that the script size is no more than 16 KB before the script is encoded in Base64. If your script size may exceed 16 KB, you can use the Gzip tool to compress the script. Compress the script files into script links. The script links constitute an Include file. The following section describes the characteristics of a Gzip-compressed script:

  - The first line must start with `#include`, and the header cannot start with spaces.
  - The running frequency of the user data varies based on the script type.

  Gzip-compressed script example:

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

  If you are making a Gzip-compressed script file, you must compress the script file into the `.gz` format.

- Upstart Job

  If you want to use an Upstart Job script, you must install the upstart service as the init system. The upstart service uses the CentOS 6, Ubuntu 10/12/14, or Debian 6/7 operating system. Upstart Job scripts put your user data in the */etc/init* directory. The following section describes the characteristics of an Upstart Job script:

  - The first line must start with `#upstart-job`, and the header cannot start with spaces.
  - The script is run every time the instance is started.

  Upstart Job script example:

```
#upstart-job
description "upstart test"
start on runlevel [2345]
stop on runlevel [! 2345]
exec echo "Hello World. The time is now $(date -R)!" | tee /root/output.txt
```

3. Debug the script to make sure that the user data is correct.

Prepare the user data of a Windows instance

User Data is a feature developed by Alibaba Cloud and enables Windows instances to run initialized scripts. You can use the Bat and PowerShell scripts to configure the user data of a Windows instance.

Perform the following operations to prepare the user data of a Windows instance:

1. Use an editor to create a text file, such as Notepad++.
2. Edit the script that is related to the user data in the text file. The first line of a script must meet the format requirements of this script type. The following section describes the format requirements of different scripts and the format examples:
   - The following section describes the characteristics of a Bat script:
     - The first line must start with `[bat]`, and the header cannot start with spaces.
     - The script size cannot exceed 16 KB before the script is encoded in Base64.
     - Only half-width letters can be entered, and no additional characters are allowed.

     Bat script example:

```
[bat]
echo "bat test" > c:\1.txt
```

     After the instance is created, start and connect to the instance. Check the execution result of the script. You can find that the *1.txt* text file is added to the *C:*\ drive.

   - The following section describes the characteristics of a PowerShell script:
     - The first line must start with `[powershell]`, and the header cannot start with spaces.
     - The script size cannot exceed 16 KB before the script is encoded in Base64.
     - Only half-width letters can be entered, and no additional characters are allowed.

PowerShell script example:

```
[powershell]
write-output "Powershell Test" | Out-File C:\2.txt
```

3. Debug the script to make sure that the user data is correct.

What's next

After the script is prepared, you can use the script to configure the user data. For more information, see Configure user data.

8.14.2. Configure user data

This topic describes how to configure user data by using the console.

Context

When you configure user data, take note of the following points:

- You can configure user data only for VPC-type instances.
- If an instance is of a phased-out instance type, the instance must be I/O optimized. This limit does not apply to instance types that are not phased-out. For more information, see Phased-out instance types and Instance families.
- The size of user data cannot exceed 16 KB before it is encoded in Base64.

  Note The user data must be Base64-encoded before it can be configured. If you do not use the Base64 encoding service provided by the ECS console, you must manually convert the user data to Base64-encoded data before you configure it.

- The instance must use a public image or a custom image derived from a public image. The following table lists the operating systems that are supported.

Platform	Operating system
Windows	Windows Server 2008 R2 and later
Linux	○ CentOS ○ Ubuntu ○ SUSE Linux Enterprise ○ OpenSUSE ○ Debian ○ Alibaba Cloud Linux

Procedure

1. Create a Linux instance. For more information, see Create an instance by using the provided wizard.When you create the instance, enter user data in the **User Data** field of the **Advanced (based on instance RAM roles or cloud-init)** section. If your user data is already Base64-encoded, select **Enter Based64 Encoded Information**.

2. After the instance is started, connect to the instance. For more information, see Overview.

   When the instance is in the **Running** state, the system will run the user data by using the root permission. Then the system will run the initialization scripts in the `/etc/init` folder.

3. View execution results based on your configured user data.

   If a failure occurs, you must check the relevant log files. The following figure shows an example result generated from user data configured by using the Upstart Job script on a CentOS instance.

   In the preceding output, the *part-001.conf* startup job file is generated in the */etc/init* folder.

8.14.3. View user data

This topic describes how to view the existing user data of an instance.

Prerequisites

User data is configured for an instance. For more information, see Configure user data.

Context

You can use the specified server at IP Address 100.100.100.200 to view user data.

Procedure

1. Connect to an instance. For more information about the procedure, see Overview.
2. Run one of the following commands to view user data of the instance:
   - For a Linux instance:

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

   - For a Windows instance:

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

8.14.4. Modify user data

This topic describes how to modify existing user data of an instance in the console.

Prerequisites

- User data is configured for an instance. For more information, see Configure user data.
- The instance is stopped. For more information about how to stop an instance, see Stop an instance.

Context

After you modify the user data, whether the instance runs the modified user data depends on the script and module type. Examples:

- If you use a shell script as user data, such as the User-Data script, the instance does not run

the modified user data.

- If you use user data to configure modules such as Byobu, Set Hostname, and Set Passwords, the instance does not run the modified user data.
- If you use user data to configure modules such as bootcmd, update_etc_hosts, and yum_add_repo, the instance runs the modified user data.

> **Note** For a pay-as-you-go instance of the VPC type, if you want to restart the instance after you modify its user data, we recommend that you disable the No Fees for Stopped Instances (VPC-Connected) feature.

For more information, visit the Module frequency section of Modules in the cloud-init documentation.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, choose **Instances & Images > Instances**.
3. In the top navigation bar, select a region.
4. Find the instance for which you want to modify the user data and choose **More > Instance Settings > Set User Data** in the **Actions** column.
5. In the Set User Data dialog box that appears, enter information and click **OK**.

Related information

References

- ModifyInstanceAttribute

8.15. 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
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+gAwIBAgIEZmbRhzANBgkqhkiG9w0BAQsFADBsMRAwDgYDVQQGEwdV
bmtub3duMRAwDgYDVQQIEwdVbmtub3duMRAwDgYDVQQHEwdVbmtub3duMRAwDgYD
VQQKEwdVbmtub3duMRAwDgYDVQQLEwdVbmtub3duMRAwDgYDVQQDEwdVbmtub3du
MB4XDTE4MDIyMzAxMjkzOFoXDTM4MDIxODAxMjkzOFowbDEQMA4GA1UEBhMHVW5r
bm93bjEQMA4GA1UECBMHVW5rbm93bjEQMA4GA1UEBxMHVW5rbm93bjEQMA4GA1UE
ChMHVW5rbm93bjEQMA4GA1UECxMHVW5rbm93bjEQMA4GA1UEAxMHVW5rbm93bjCC
ASIwDQYJKoZIhvcNAQEBBQADggEPADCCAQoCggEBAIJwy5sbZDiNyX4mvdP32pqM
YMK4k7+5lRnVR2Fky/5uwyGSPbddNXaXzwEm+u4wIsJiaAN3OZgJpYIoCGik+9lG
5gVAIr0+/3rZ61IbeVE+vDenDd8g/m/YIdYBfC2IbzgS9EVGAf/gJdtDODXrDfQj
Fk2rQsvpftVOUs3Vpl9O+jeCQLoRbZYm0c5v7jP/L2lK0MjhiywPF2kpDeisMtnD
/ArkSPIlg1qVYm3F19v3pa6ZioM2hnwXg5DibYlgVvsIBGhvYqdQ1KosNVcVGGQa
HCUuVGdS7vHJYp3byH0vQYYygzxUJT2TqvK7pD57eYMN5drc7e19oyRQvbPQ3kkC
AwEAAaMhMB8wHQYDVR0OBBYEFAwwrnHlRgFvPGo+UD5zS1xAkC91MA0GCSqGSIb3
DQEBCwUAA4IBAQBBLhDRgezd/OOppuYEVNB9+XiJ9dNmcuHUhjNTnjiKQWVk/YDA
v+T2V3t9yl8L8o61tRIVKQ++lDhjlVmur/mbBN25/UNRpJllfpUH6oOaqvQAze4a
nRgyTnBwVBZkdJ0d1sivL9NZ4pKelJF3Ylw6rp0YMqV+cwkt/vRtzRJ31ZEeBhs7
vKh7F6BiGCHL5ZAwEUYe8O3akQwjgrMUcfuiFs4/sAeDMnmgN6Uq8DFEBXDpAxVN
sV/6Hockdfinx85RV2AUwJGfClcVcu4hMhOvKROpcH27xu9bBIeMuY0vvzP2VyOm
DoJeqU7qZjyCaUBkPimsz/1eRod6d4P5qxTj
-----END CERTIFICATE-----
```

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_signature_without_audience(){
curl 100.100.100.200/latest/dynamic/instance-identity/document > document
echo "-----BEGIN CERTIFICATE-----" > signature
curl 100.100.100.200/latest/dynamic/instance-identity/pkcs7 >> signature
echo "" >> signature
echo "-----END CERTIFICATE-----" >> signature
openssl smime -verify -in signature -inform PEM -content document -certfile cert.cer -noverify > /d
ev/null
}
verify_signature_without_audience
```

5. Once the response result shows `Verification successful`, remove the restriction and run the image in the instance.

Sample 2. Audience in the signature

Assume that you published an image in the image market. The following example shows you how to grant access to the instances of your customers by specifying an audience parameter during the process of validation. To make sure that the instance identity is not maliciously acquired and used, you can implement the access control at the application server by combining your audience parameter. We recommend the value of the audience parameter is a random string, timestamp, regularly changed data, or some output generated by a specific algorithm.

1. Connect to the target instance.
2. Verify whether the image used by the instance is from the Alibaba Cloud Marketplace, or another source, by calling the metadata items of product-code and charge-type.

```
curl http://100.100.100.200/latest/meta-data/image/market-place/product-code
curl http://100.100.100.200/latest/meta-data/image/market-place/charge-type
```

3. Create a temporary file cert.cer in the working directory and save the public certificate to the file.
4. Determine the identity of the instance by running the following script.

```
#! /usr/bin/bash
function verify_signature_with_specified_audience(){
audience='your audience' #Here is your audience parameter.
document=$(curl 100.100.100.200/latest/dynamic/instance-identity/document)
audience_json=',"audience":""${audience}""}'
echo -n ${document%?} ${audience_json} > document
echo "-----BEGIN CERTIFICATE-----" > signature
curl 100.100.100.200/latest/dynamic/instance-identity/pkcs7? audience=${audience} >> signature
echo "" >> signature
echo "-----END CERTIFICATE-----" >> signature
openssl smime -verify -in signature -inform PEM -content document -certfile cert.cer -noverify > /d
ev/null
}
verify_signature_with_specified_audience
```

5. Once the response result shows `Verification successful`, remove the restriction and run the image in the instance.

8.16. Manage software on Linux instances

8.17. Configure time

8.17.1. Alibaba Cloud NTP server

This topic describes Alibaba Cloud Network Time Protocol (NTP) servers. Alibaba Cloud provides internal and public NTP servers to synchronize the local time of ECS instances in each network.

Alibaba Cloud ECS time calibration time zone

Internal and public NTP servers

NTP is used to synchronize the computer time in a network.

The consistency of time and time zone is important to ECS instances because it can affect the task execution results. For example, when you update a database or analyze logs, the time sequence has a significant impact on the results. When you run business on ECS instances, you must standardize the time zone of the involved instances to avoid problems such as logical confusion and network request errors. In addition, you can use NTP servers to synchronize the local time of all ECS instances in a network.

ECS provides a high-precision NTP server for your convenience. The `ntp.cloud.aliyuncs.com` server offers a globally distributed NTP service that uses Stratum 1 servers. Stratum 1 servers are suitable for industries that require precise timing, such as finance, communication, scientific research, and astronomy. The NTP service is also used to synchronize the local time between ECS instances and other cloud services. The following table describes the domain names of Alibaba Cloud NTP servers in various networks.

Classic network (internal network)	VPC (internal network)	Internet
N/A	ntp.cloud.aliyuncs.com	ntp.aliyun.com
ntp1.cloud.aliyuncs.com	ntp7.cloud.aliyuncs.com	ntp1.aliyun.com
ntp2.cloud.aliyuncs.com	ntp8.cloud.aliyuncs.com	ntp2.aliyun.com
ntp3.cloud.aliyuncs.com	ntp9.cloud.aliyuncs.com	ntp3.aliyun.com
ntp4.cloud.aliyuncs.com	ntp10.cloud.aliyuncs.com	ntp4.aliyun.com
ntp5.cloud.aliyuncs.com	ntp11.cloud.aliyuncs.com	ntp5.aliyun.com
ntp6.cloud.aliyuncs.com	ntp12.cloud.aliyuncs.com	ntp6.aliyun.com
N/A	N/A	ntp7.aliyun.com

Other Alibaba Cloud public services

The following table describes other public services provided by Alibaba Cloud.

Public service	Description
Public DNS: 223.5.5.5/223.6.6.6	Domain name: `http://www.alidns.com`
Public images: `https://developer.aliyun.com/mirror`	Update frequency: The image files are updated every day from 02:00 to 04:00 (UTC+8). The images contain a wide collection of Linux distributions and open source software.

References

- Configure the NTP service for Windows instances
- Time setting: Synchronize NTP servers and change time zone for Linux instances

8.17.2. Configure the NTP service for Windows instances

This topic describes how to enable and configure the NTP service for a Windows instance to ensure the local system time is synchronized precisely.

Alibaba Cloud ECS time calibration time zone

Context

ECS instances in all Alibaba Cloud regions use UTC+8 by default. You can also configure or modify time zones for your instances as needed.

This topic uses Windows Server 2012 R2 Datacenter Edition 64-bit as an example to describe how to use the NTP service to synchronize the local system time for Windows instances.

Enable the NTP service

By default, the Windows Time service is enabled on Windows Server operating systems. Therefore, the NTP service must be enabled for Windows instances to synchronize the local system time. To check and enable the NTP service, perform the following operations:

1. Connect to a Windows instance.
2. Click **Start** and open the **Run** dialog box. Run the `services.msc` command in the dialog box.
3. In the **Services** dialog box that appears, find and double-click **Windows Time**.
4. In the **Windows Time Properties (Local Computer)** dialog box that appears, perform the following operations:
   i. Select **Automatic** for **Startup type**.
   ii. Ensure that the value of **Service status** is **Running**. Otherwise, click **Start**.
   iii. Click **Apply** and click **OK**.

Modify the default NTP server address

By default, Windows Server operating systems use the Microsoft NTP server (time.windows.com), but synchronization errors may occur. When you use a Windows instance, you can replace the default NTP server with the intranet NTP server provided by Alibaba Cloud. To modify the default NTP server address, perform the following operations:

1. Connect to a Windows instance.
2. In the notification area of the taskbar, click the date and time and click **Change date and time settings**.
3. In the **Date and Time** dialog box that appears, click the **Internet Time** tab and click **Change settings**.
4. In the **Internet Time Settings** dialog box that appears, select **Synchronize with an Internet time server** and enter an Alibaba Cloud intranet NTP server address. Click **Update now**. For more information, see Alibaba Cloud NTP server.

Modify the NTP synchronization interval

The default NTP synchronization interval is five minutes, you can modify it as needed. To modify the NTP synchronization interval, perform the following operations:

1. Connect to a Windows instance.
2. Click **Start** and open the **Run** dialog box. Run the `regedit` command in the dialog box.
3. In the left-side navigation pane of the **Registry Editor**, choose **HKEY_LOCAL_MACHINE** > **SYSTEM** > **CurrentControlSet** > **Services** > **W32Time** > **TimeProviders** > **NtpClient** and double-click **SpecialPollInterval**.
4. In the **Edit DWORD (32-bit) Value** dialog box that appears, select **Decimal** in the **Base** section and enter the required value in the **Value data** field. The required value is the new synchronization interval. Unit: seconds.
5. Click **OK**.

Related information

- Alibaba Cloud NTP server
- Time setting: Synchronize NTP servers and change time zone for Linux instances

8.17.3. 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.

9.Instance families that do not support advanced VPC features

This topic describes instance families that do not support advanced VPC features.

Advanced VPC features include network ACLs, routing tables, and flow logs. For more information, see Overview, Overview, and Flow log overview.

Restrictions of advanced VPC features on ECS instances:

- Instance creation: If advanced features are enabled on a VPC, you cannot create instances of instance families in the following table within the VPC.
- Instance upgrade or downgrade: If the instance type of an ECS instance supports advanced features and those advanced features are enabled on the corresponding VPC, the instance cannot be changed to an instance that belongs to instance families in the following table. These methods to change instances include upgrading, downgrading, renewal and downgrading, and changing instance types.

When you enable advanced features on a VPC, the system automatically detects the instances within the VPC that need to be upgraded or downgraded. Then you can upgrade or downgrade these instances. For more information, see Change the instance type of a pay-as-you-go instance or Upgrade configurations of subscription instances.

Only the following instance families do not support advanced VPC features.

Instance family type	Instance family
General purpose instance families	sn2 (phased-out)
Compute optimized instance families	sn1 (phased-out)
Memory optimized instance families	se1
Big data instance families	d1
Instance families with local SSDs	i1
Instance families with high clock speed	• c4 (phased-out) • ce4 (phased-out) • cm4 (phased-out)
Compute optimized instance families with GPU capabilities	• gn4 • gn5
Graphic-accelerated instance families with GPU capabilities	ga1 (phased-out)

Instance family type	Instance family
Shared instance families	• n1 (phased-out) • n2 (phased-out) • e3 (phased-out) • xn4 • n4 • mn4 • e4
Generation I instance types	• t1 (phased-out) • s1 (phased-out) • s2 (phased-out) • s3 (phased-out) • m1 (phased-out) • m2 (phased-out) • c1 (phased-out) • c2 (phased-out)

10.Phased-out instance types

This topic describes all phased-out instance types on the China site (aliyun.com). However, instance types sn1, sn2, n1, n2, and e3 are still available for purchase on the International site (alibabacloud.com).

Instructions for specifications

- The packet forwarding rate refers to the total number of inbound and outbound packets that can be processed by a network per second. For more information about how to test the packet forwarding rate, see Test network performance.
- NIC queues refer to the maximum number of NIC queues supported by the current instance type. By default, a CentOS 7.3 image uses the maximum number of NIC queues.
- Enterprise-level instance types that are equipped with two or more vCPUs support Elastic Network Interfaces (ENIs). Shared instance types that are equipped with four or more vCPUs support ENIs. For more information, see ENI overview.

Instance type upgrade

For information about the phased-out instance types that can be upgraded to new instance types, see Instance families that support instance type changes.

sn2, general purpose instance family

Features

- Offers a CPU-to-memory ratio of 1:4.
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or E5-2680 v3 (Haswell) processors with consistent computing performance.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Enterprise-level applications of various types and scales
  - Small and medium-sized database systems, caches, and search clusters
  - Data analysis and computing

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
ecs.sn2.medium	2	8.0	None	0.5	100	1	2
ecs.sn2.large	4	16.0	None	0.8	200	1	3
ecs.sn2.xlarge	8	32.0	None	1.5	400	1	4

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
ecs.sn2.3xlarge	16	64.0	None	3.0	500	2	8
ecs.sn2.7xlarge	32	128.0	None	6.0	800	3	8
ecs.sn2.13xlarge	56	224.0	None	10.0	1,200	4	8

sn1, compute optimized instance family

Features

- Offers a CPU-to-memory ratio of 1:2.
- Equipped with 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or E5-2680 v3 (Haswell) processors with consistent computing performance.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Web frontend servers
  - Frontend servers of massively multiplayer online (MMO) games
  - Data analysis, batch processing, and video encoding
  - High-performance scientific and engineering applications

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
ecs.sn1.medium	2	4.0	None	0.5	100	1	2
ecs.sn1.large	4	8.0	None	0.8	200	1	3
ecs.sn1.xlarge	8	16.0	None	1.5	400	1	4
ecs.sn1.3xlarge	16	32.0	None	3.0	500	2	8
ecs.sn1.7xlarge	32	64.0	None	6.0	800	3	8

c4, ce4, and cm4, compute optimized instance families with high clock speed

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Provides consistent computing performance.
- Equipped with 3.2 GHz Intel Xeon E5-2667 v4 (Broadwell) processors.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - High-performance web frontend servers
  - High-performance scientific and engineering applications
  - MMO gaming and video encoding

c4

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
ecs.c4.xlarge	4	8.0	None	1.5	200	1	3
ecs.c4.2xlarge	8	16.0	None	3.0	400	1	4
ecs.c4.3xlarge	12	24.0	None	4.5	600	2	6
ecs.c4.4xlarge	16	32.0	None	6.0	800	2	8

ce4

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
ecs.ce4.xlarge	4	32.0	None	1.5	200	1	3
ecs.ce4.2xlarge	8	64.0	None	3.0	400	1	3

cm4

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	NIC queues	ENIs (including one primary ENI)
ecs.cm4.xlarge	4	16.0	None	1.5	200	1	3
ecs.cm4.2xlarge	8	32.0	None	3.0	400	1	4
ecs.cm4.3xlarge	12	48.0	None	4.5	600	2	6
ecs.cm4.4xlarge	16	64.0	None	6.0	800	2	8
ecs.cm4.6xlarge	24	96.0	None	10.0	1,200	4	8

ga1, visualization and compute optimized instance family with GPU capabilities

Features

- I/O optimized.
- Supports standard SSDs and ultra disks.
- Uses AMD S7150 GPUs.
- Offers a CPU-to-memory ratio of 1:2.5.
- Equipped with 2.5 GHz Intel® Xeon® E5-2682 v4 (Broadwell) processors.
- Equipped with high-performance local NVMe SSDs.
- Provides a fast and reliable network based on large computing capacity.
- Suitable for the following scenarios:
  - Rendering and multimedia encoding and decoding
  - Machine learning, high performance computing, and high-performance databases
  - Server-side workloads that require powerful concurrent floating-point compute capabilities

Instance types

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	GPUs	GPU memory (GB)	Bandwidth (Gbit/s)	Packet forwarding rate (Kpps)	IPv6 support	NIC queues	ENIs (including one primary ENI)	Private IP addresses per ENI
ecs.ga1.xlarge	4	10.0	1 × 87	0.25 × AMD S7150	2	1.0	200	No	1	3	10
ecs.ga1.2xlarge	8	20.0	1 × 175	0.5 × AMD S7150	4	1.5	300	No	1	4	10
ecs.ga1.4xlarge	16	40.0	1 × 350	1 × AMD S7150	8	3.0	500	No	2	8	20
ecs.ga1.8xlarge	32	80.0	1 × 700	2 × AMD S7150	2 × 8	6.0	800	No	3	8	20
ecs.ga1.14xlarge	56	160.0	1 × 1,400	4 × AMD S7150	4 × 8	10.0	1,200	No	4	8	20

Note

- For more information, see Create a ga1 instance.
- You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.
- For more information about these specifications, see Description of instance specifications.

n1, n2, and e3, shared instance families

Features

- Equipped with 2.5 GHz Intel Xeon E5-2680 v3 (Haswell) processors.
- Provides a fast and reliable network based on large computing capacity.
- I/O optimized.
- Supports the following categories of disks:
  - Standard SSDs

- Ultra disks

Instance family	Feature	CPU-to-memory ratio	Scenario
n1	Shared compute optimized instances	1:2	• Small and medium-sized web servers • Batch processing • Distributed analysis • Advertisement services
n2	Shared general purpose instances	1:4	• Medium-sized web servers • Batch processing • Distributed analysis • Advertisement services • Hadoop clusters
e3	Shared memory optimized instances	1:8	• Cache and Redis • Search applications • In-memory databases • Databases with high I/O requirements, such as Oracle and MongoDB • Hadoop clusters • Large-volume data processing

n1

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	ENIs (including one primary ENI)
ecs.n1.tiny	1	1.0	None	1
ecs.n1.small	1	2.0	None	1
ecs.n1.medium	2	4.0	None	1
ecs.n1.large	4	8.0	None	2
ecs.n1.xlarge	8	16.0	None	2
ecs.n1.3xlarge	16	32.0	None	2

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	ENIs (including one primary ENI)
ecs.n1.7xlarge	32	64.0	None	2

n2

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	ENIs (including one primary ENI)
ecs.n2.small	1	4.0	None	1
ecs.n2.medium	2	8.0	None	1
ecs.n2.large	4	16.0	None	2
ecs.n2.xlarge	8	32.0	None	2
ecs.n2.3xlarge	16	64.0	None	2
ecs.n2.7xlarge	32	128.0	None	2

e3

Instance type	vCPUs	Memory (GiB)	Local storage (GiB)	ENIs (including one primary ENI)
ecs.e3.small	1	8.0	None	1
ecs.e3.medium	2	16.0	None	1
ecs.e3.large	4	32.0	None	2
ecs.e3.xlarge	8	64.0	None	2
ecs.e3.3xlarge	16	128.0	None	2

Generation I instance families

Generation I instance families include t1, s1, s2, s3, m1, m2, c1, and c2. All these instance families are legacy shared instance families. They are categorized based on the number of vCPUs such as 1, 2, 4, 8, or 16.

Features

- Equipped with Intel Xeon E5-2420 processors with the clock speed no less than 1.9 GHz.
- Uses the DDR3 memory.
- Provides optional I/O optimized performance or non-I/O optimized performance.

I/O optimized instance types

I/O optimized instances support the following categories of disks:

- Standard SSDs

- Ultra disks

Category	Instance type	vCPUs	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 can only be equipped with basic disks.

Category	Instance type	vCPUs	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

Category	Instance type	vCPUs	Memory (GiB)
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

11.ECS instance FAQ

This topic provides answers to commonly asked questions about ECS instances.

- FAQ about purchasing instances
  - How do I check which instance resources are available for purchase in a specific region or zone?
  - What can I do if a resource is sold out when I attempt to purchase an instance within a specific region or zone?
  - How do I select an instance type that is suitable for my business?
  - How long does it take to create an ECS instance?
  - I paid for an ECS instance but no ECS instance was created. Why?
  - Why are some instance types not available on the instance buy page when I attempt to purchase a pay-as-you-go instance?
- FAQ about enterprise-level instances
  - What are enterprise-level instances? What are shared instances?
  - What are the differences between enterprise-level and shared instances?
  - Which instance families are enterprise-level instance families? Which instance families are shared instance families?
  - In what business scenarios do I need to purchase enterprise-level instances?
  - How is the network performance of enterprise-level instances?
  - Which disk categories do enterprise-level instances support?
  - Which image types do enterprise-level instances support?
  - What are the limits on changing the instance types of enterprise-level instances?
  - Can I upgrade a shared instance to an enterprise-level instance?
- FAQ about GPU-accelerated instances
  - After the NVIDIA driver was installed, nvidia-smi did not work and an error message was returned indicating that the NVIDIA driver had not been installed. Why?
  - Why doesn't Windows Graphics support DirectX-based functions?
  - Do GPU-accelerated instances support Android emulators?
  - Can I change the configurations of GPU-accelerated instances?
  - Do pay-as-you-go GPU-accelerated instances support the No Fees for Stopped Instances (VPC-Connected) feature?
  - How do I view GPU monitoring data?
- FAQ about ECS Bare Metal Instances
  - What are the differences between ECS bare metal instances, traditional cloud hosts (virtual machines), and traditional physical machines?
  - How is the network performance of ECS bare metal instances?
  - Which disk categories do ECS bare metal instances support? How many data disks can be attached to a single ECS bare metal instance?
  - Can ECS bare metal instances have their configurations upgraded or downgraded? Do they support failover?

- FAQ about SCC
  - How do I create an SCC instance?
  - How are SCC instances billed?
  - How do I create an SCC by using E-HPC?
  - How do I use SCC RDMA?
- FAQ about preemptible instances
  - I do not have overdue payments in my account. Why have my preemptible instances been released?
  - Will I be notified when my preemptible instances are about to be released? How will I be notified?
  - Can the data of a preemptible instance be automatically retained when the instance is released?
  - Can I cancel or reschedule the automatic release of a preemptible instance?
  - Can preemptible instances be converted into subscription instances?
  - To what resources are the prices of preemptible instances applicable?
  - How do I bid for a preemptible instance?
  - What is the relationship between the user-defined maximum hourly price and the spot price of a preemptible instance?
  - Will I be charged the same price for all preemptible instances started at the same time?
  - For the first hour of a preemptible instance running, will I be charged at a price fluctuating with the spot price?
  - Can I view the spot price of an instance type when I purchase a preemptible instance?
  - Can I view the historical prices of a preemptible instance type? How?
  - Will preemptible instances continue to be billed after they are stopped?
  - How many preemptible instances can be purchased by a single account?
  - How do I increase my vCPU-based quotas?
  - Can the instance type of a preemptible instance be changed?
  - Which instance families support preemptible instances?
  - In which regions can I create preemptible instances?
  - Why is the Preemptible Instance option not available on the instance buy page when I attempt to purchase an ECS instance?
- FAQ about reserved instances
  - What are reserved instances?
  - Will a reserved instance provide a resource reservation?
  - What operating systems do reserved instances support?
  - Which instance families do reserved instances support?
  - Can reserved instances be applied to preemptible instances?
  - Can the instance families of reserved instances be changed?
  - To what scenarios are zonal reserved instances applicable?
  - To what scenarios are regional reserved instances applicable?
  - How is the zone flexibility of reserved instances applied?

- FAQ about instance limits
  - What limits apply to the transfer and change of public IP addresses of ECS instances?
  - Can I access amazon.com from my ECS instance?
  - Why am I unable to access a website hosted outside mainland China after I log on to my ECS instance?
  - I cannot purchase more pay-as-you-go instances. What can I do?
  - How can I query resource quotas?
- FAQ about instance billing
  - Will I still be charged for a pay-as-you-go instance after it is stopped either manually or due to an overdue payment?
  - What can I do if an order cannot be placed to change the billing method of an instance from pay-as-you-go to subscription?
  - How long after an order is paid does it take to change the billing method of an instance from pay-as-you-go to subscription?
  - What can I do if the billing method of an instance cannot be changed from pay-as-you-go to subscription?
  - When I change the billing method of an instance from pay-as-you-go to subscription, will the billing method for network usage of the instance also change?
  - I have an unpaid order to change the billing method of an instance from pay-as-you-go to subscription. If I upgrade the configurations of the instance, will the order still be valid?
  - What can I do if the billing method of an instance cannot be changed from subscription to pay-as-you-go?
  - When I try to change the billing method of a disk in an ECS instance, an error message is returned indicating that I have already changed the billing method three times. What does this mean?
  - Why am I unable to change a pay-as-you-go instance to a subscription one?
  - How do I view the expiration time of a subscription instance?

How do I check which instance resources are available for purchase in a specific region or zone?

You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.

What can I do if a resource is sold out when I attempt to purchase an instance within a specific region or zone?

If the resource of a specific instance type is sold out when you attempt to purchase an instance of that instance type within a specific region or zone, take one of the following measures:

- Select another region.
- Select another zone.
- Select another instance type.

If the resource that you want to purchase is still not available after you take all of the preceding measures, try again later. Instance resources are dynamic. When resources are insufficient, Alibaba Cloud makes sure to replenish them as soon as possible.

You can also use the arrival notice feature to receive notifications when resources become available.

How do I select an instance type that is suitable for my business?

Consider the following factors when you select an instance type. For more information about how to select an instance type, see Best practices for instance type selection.

- Your business needs
- Your website type
- The average number of page views per day on your website
- The size of your homepage
- The data capacity of your website

How long does it take to create an ECS instance?

It takes a minute or two to create an ECS instance. After the instance is created:

- For an ECS instance that runs the Linux system, you can connect to the instance without any other configurations. For more information, see Connect to an ECS instance.
- For an ECS instance that runs the Windows system, you must use the Sysprep tool to initialize the operating system. Do not restart the instance during initialization. After initialization, you can connect to the instance. For more information, see Connect to an ECS instance. The amount of time required for initialization is determined by the type of ECS instances.
  - For an I/O optimized instance that runs the Windows system, two to three minutes are required for initialization.
  - For a non-I/O optimized instance that runs the Windows system, ten minutes are required for initialization.

> **Note** If an error occurs during instance creation, submit a ticket.

I paid for an ECS instance but no ECS instance was created. Why?

If resources within the specified zone are insufficient to create an instance of your selected instance type, the instance will fail to be created. Your account will automatically be refunded the cost of the instance. If you do not receive a refund within half an hour, submit a ticket.

You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.

Why are some instance types not available on the instance buy page when I attempt to purchase a pay-as-you-go instance?

Some of the instance types equipped with 16 or more vCPUs are available only to authorized accounts when the pay-as-you-go billing method is selected. If an instance type is available on the instance buy page when the subscription billing method is selected but is not available when the pay-as-you-go billing method is selected, submit a ticket to check whether your account is authorized to purchase pay-as-you-go instances of that instance type.

If your account is authorized but the instance type is still not available, the current region may not have sufficient resources for you to create pay-as-you-go resources of that instance type. For more information, see What can I do if a resource is sold out when I attempt to purchase an instance within a specific region or zone?.

What are enterprise-level instances? What are shared instances?

Enterprise-level instances belong to a series of instance families released by Alibaba Cloud in September 2016. Enterprise-level instances provide high performance, consistent computing power, and balanced network performance. Enterprise-level instances have exclusive and consistent computing, storage, and network resources, and are suitable for enterprise scenarios with high business stability requirements.

Shared instances belong to a series of instance families that are targeted at individuals or small and medium-sized websites. Shared instances share resources, in contrast with enterprise-level instances that each have their own resources exclusively. As a result, shared instances do not provide consistent computing performance, but cost less.

What are the differences between enterprise-level and shared instances?

Enterprise-level instances use a CPU-bound scheduling scheme. Each vCPU is bound to a CPU hyperthread. Instances do not compete for CPU resources and do provide consistent computing performance as guaranteed in the SLA.

Shared instances use a CPU-unbound scheduling scheme. Each vCPU is randomly allocated to an idle CPU hyperthread. vCPUs of different instances compete for CPU resources, which causes computing performance to fluctuate when traffic loads are heavy. Shared instances can guarantee only availability but cannot guarantee the performance that may be required in the SLA.

Which instance families are enterprise-level instance families? Which instance families are shared instance families?

Among the instance families that are available for purchase, t6, t5, s6, n4, mn4, xn4, and e4 are shared instance families, and the rest are enterprise-level instance families.

In what business scenarios do I need to purchase enterprise-level instances?

For business scenarios to which different enterprise-level instances are applicable, see Instance families and Best practices for instance type selection.

How is the network performance of enterprise-level instances?

The network performance of enterprise-level instances depends on their specifications. The higher their specifications are, the higher network performance the instances can deliver. For more information about the network performance of different instance specifications, see Instance families.

Which disk categories do enterprise-level instances support?

For the disk categories that enterprise-level instances support, see Disk categories.

Which image types do enterprise-level instances support?

For the public images that enterprise-level instances support, see Overview.

You can also import custom images. For more information, see Import custom images.

What are the limits on changing the instance types of enterprise-level instances?

For the limits on changing the instance types of enterprise-level instances, see Change instance types of instances.

Can I upgrade a shared instance to an enterprise-level instance?

Yes, you can upgrade a shared instance to an enterprise-level instance. For more information, see Change instance types of instances.

After the NVIDIA driver was installed, nvidia-smi did not work and an error message was returned indicating that the NVIDIA driver had not been installed. Why?

Cause: The mismatch between the kernel and kernel-devel package versions resulted in a driver compilation error when the driver was installed from the .rpm file.

Solution: Check the kernel version and download the corresponding kernel-devel package version. Then, run the `rpm -qa | grep kernel` command on the instance to check whether the kernel-devel package version matches the kernel version. Make sure that they match and then install the driver again.

Why doesn't Windows Graphics support DirectX-based functions?

Problem description: On Windows instances where the installed GPU driver has taken effect, Windows Remote Desktop Protocol (RDP) may not support DirectX- or OpenGL-based applications.

Solution: Install the VNC service and client or use other protocols that support these applications, such as PC over IP (PCoIP) and XenDesktop HDX 3D.

Do GPU-accelerated instances support Android emulators?

No, GPU-accelerated instances do not support Android emulators.

Can I change the configurations of GPU-accelerated instances?

You cannot change the configurations of the following GPU-accelerated instances:

- GPU-accelerated instances equipped with local storage.
- GPU-accelerated instances with virtual GPUs

For information about the GPU-accelerated instance families that support instance type changes, see Change instance types of instances.

Do pay-as-you-go GPU-accelerated instances support the No Fees for Stopped Instances (VPC-Connected) feature?

Instance families that are equipped with local storage, such as gn5, do not support the No Fees for Stopped Instances (VPC-Connected) feature. For more information, see No Fees for Stopped Instances (VPC-Connected).

How do I view GPU monitoring data?

You can log on to the CloudMonitor console or call the DescribeMetricList operation to view GPU monitoring data. For more information, see GPU monitoring.

What are the differences between ECS bare metal instances, traditional cloud hosts (virtual machines), and traditional physical machines?

For information about the differences, see ECS Bare Metal Instances.

How is the network performance of ECS bare metal instances?

The network performance of ECS bare metal instances depends on their specifications. The higher their specifications are, the higher network performance the instances can deliver. For more information about the network performance of different ECS bare metal instance specifications, see Instance families.

Which disk categories do ECS bare metal instances support? How many data disks can be attached to a single ECS bare metal instance?

ECS bare metal instances support ultra disks and standard SSDs. Up to 16 data disks can be attached to a single ECS bare metal instance.

Can ECS bare metal instances have their configurations upgraded or downgraded? Do they support failover?

ECS bare metal instances cannot have their configurations upgraded or downgraded. They do support failover. When the physical machine that hosts an ECS bare metal instance fails, the system fails the instance over to another physical machine. Data is retained in the data disks of the instance.

How do I create an SCC instance?

You can create an SCC instance in one of the following ways:

- If you only want to use RDMA, log on to the ECS console. Create an SCC instance as described in Create an SCC instance.
- If you want to use the HPC scheduler and cluster resizing service in addition to RDMA, log on to the E-HPC console. Create an SCC and then create an SCC instance.

How are SCC instances billed?

SCC instances can be billed on a weekly, monthly, or yearly subscription basis.

How do I create an SCC by using E-HPC?

You can log on to the E-HPC console or call the CreateCluster operation to create an SCC.

How do I use SCC RDMA?

If you select an image that is customized for SCC and supports RDMA RoCE driver and OFED stack to create an SCC instance, you can use SCC RDMA by using the IB verbs API or implement RDMA communication by using the MPI.

I do not have overdue payments in my account. Why have my preemptible instances been released?

Each newly created preemptible instance has a protection period of one hour. During the protection period, the preemptible instance cannot be released due to insufficient resources or fluctuations in the spot price (the current price for the preemptible instance per hour). If the spot price exceeds the specified maximum hourly price or if resources are insufficient after the protection period ends, the preemptible instance will be automatically released.

Will I be notified when my preemptible instances are about to be released? How will I be notified?

Yes, you will be notified when your preemptible instances are about to be released. When your preemptible instance needs to be released due to a spot price change or insufficient resources, the instance will first enter the To Be Released state and then be automatically released in five minutes.

You can use CloudMonitor to subscribe to notifications for interrupted preemptible instances. For more information, see Set event notifications.

To check whether an instance is in the To Be Released state, you can view the instance metadata, or call the DescribeInstances operation and view the returned OperationLocks data.

Can the data of a preemptible instance be automatically retained when the instance is released?

No, the data of a preemptible instance cannot be automatically retained when the instance is released. When a preemptible instance is no longer needed, we recommend that you back up your data and environment by following the instructions in Create a snapshot and then release the instance. If necessary, you can purchase new preemptible instances.

Can I cancel or reschedule the automatic release of a preemptible instance?

Yes, you can cancel or reschedule the automatic release of your preemptible instance at any time.

Can preemptible instances be converted into subscription instances?

No, preemptible instances cannot be converted into subscription instances.

To what resources are the prices of preemptible instances applicable?

The prices of preemptible instances are applicable only to instance types. Other resources such as system disks, data disks, and network bandwidth are billed at the same prices as those of pay-as-you-go instances.

How do I bid for a preemptible instance?

To create a preemptible instance, you can specify a maximum hourly price to bid for an instance type. If the spot price is less than or equal to your specified maximum hourly price, the preemtible instance that you request is created and billed based on the spot price. For more information, see Create a preemptible instance.

What is the relationship between the user-defined maximum hourly price and the spot price of a preemptible instance?

If the spot price is less than or equal to the user-defined maximum hourly price, the preemtible instance will be created and billed based on the spot price. Each newly created preemptible instance has a protection period of one hour. During the protection period, the instance cannot be automatically released due to insufficient resources or fluctuations in the spot price.

After the protection period ends, the system checks the spot price and resource availability of the instance type every five minutes. If the spot price exceeds the user-defined maximum hourly price or resources of the instance type are insufficient, the preemptible instance that is in the Running state will be released.

Will I be charged the same price for all preemptible instances started at the same time?

Yes, you will be charged the same price for all preemptible instances started at the same time.

For the first hour of a preemptible instance running, will I be charged at a price fluctuating with the spot price?

No, the hourly price of the first instance hour for a preemptible instance is set at the start of the hour and remains in effect for the whole hour.

Can I view the spot price of an instance type when I purchase a preemptible instance?

Yes, when you create a preemptible instance in the ECS console, you can view the spot price and historical prices of each selected instance type. The total instance price (including the instance type, storage, and bandwidth fees) is displayed in the lower part of the instance buy page. The instance type fee is the spot price of the selected instance type.

Can I view the historical prices of a preemptible instance type? How?

Yes, you can select an instance type to view its historical prices when you attempt to create a preemptible instance in the ECS console. You can also call the DescribeSpotPriceHistory operation to view the historical prices of a preemptible instance type.

Will preemptible instances continue to be billed after they are stopped?

Yes, preemptible instances will continue to be billed after they are stopped. When a preemptible instance is no longer needed, we recommend that you back up your data and environment by following the instructions in Create a snapshot and then release the instance. If necessary, you can purchase new preemptible instances.

> **Note** Preemptible instances will continue to be billed after you stop them from the ECS console or by calling the StopInstance operation.

How many preemptible instances can be purchased by a single account?

vCPU-based quotas, instead of instance count-based quotas, apply to preemptile instances in each account. When you create a preemptible instance, you can view the available vCPU-based quota after you select an instance type. For more information, see the "Instance limits" section in Limits.

How do I increase my vCPU-based quotas?

To increase your vCPU-based quotas, submit a ticket.

Can the instance type of a preemptible instance be changed?

No, the instance types of preemptible instances cannot be changed.

Which instance families support preemptible instances?

Instance families that support the pay-as-you-go billing method also support preemptible instances. If a preemptible instance of a specific instance type cannot be created due to insufficient resources, try a different instance type.

In which regions can I create preemptible instances?

Preemptible instances can be created within all Alibaba Cloud regions. If preemptible instances cannot be created within a specific region due to insufficient resources, try a different region.

Why is the Preemptible Instance option not available on the instance buy page when I attempt to purchase an ECS instance?

Whether the Preemptible Instance option is available depends on your ECS usage.

What are reserved instances?

Reserved instances automatically match pay-as-you-go instances (excluding preemptible instances) in your account to provide a billing discount. Reserved instances can also be used to reserve resources for pay-as-you-go instances. A combination of reserved instances and pay-as-you-go instances provides higher cost-effectiveness and a higher degree of flexibility as compared to subscription instances.

Will a reserved instance provide a resource reservation?

Zonal reserved instances provide resource reservations, but regional reserved instances do not.

What operating systems do reserved instances support?

Reserved instances support both Windows and Linux operating systems. For example, a reserved Linux instance can be applied to any pay-as-you-go Linux instances that match its attributes regardless of the image type (public images, custom images, shared images, and Alibaba Cloud Marketplace images).

To apply a reserved instance to pay-as-you-go instances created from Bring Your Own License (BYOL) images, submit a ticket.

Which instance families do reserved instances support?

Reserved instances support the following instance families:

- General purpose instance families: g6e, g6, g5, and sn2ne
- Compute optimized instance families: c6e, c6, c5, ic5, and sn1ne

- Memory optimized instance families: r6e, r6, r5, and se1ne
- Big data instance family: d2s
- Instance families with local SSDs: i2 and i2g
- Instance families with high clock speed: hfc6, hfc5, hfg6, hfg5, and hfr6
- GPU-accelerated compute optimized instance families: gn6i and gn6e
- ECS Bare Metal Instance families: ebmc6, ebmg6, ebmr6, ebmhfc6, ebmhfg6, and ebmhfr6
- Burstable instance families: t6 and t5

Can reserved instances be applied to preemptible instances?

No, reserved instances cannot be applied to preemptible instances.

Can the instance families of reserved instances be changed?

No, the instance families of reserved instances cannot be changed.

To what scenarios are zonal reserved instances applicable?

We recommend that you purchase zonal reserved instances when you have clear requirements to reserve resources.

To what scenarios are regional reserved instances applicable?

We recommend that you purchase regional reserved instances if you want to have a higher degree of zone flexibility or instance size flexibility.

How is the zone flexibility of reserved instances applied?

Only regional reserved instances provide zone flexibility. Example:

You are running the following pay-as-you-go instance:

One ecs.c5.xlarge Linux instance in Qingdao Zone B. The instance is named C5PAYG-b.

You have purchased the following reserved instance:

One regional ecs.c5.xlarge reserved instance in the China (Qingdao) region. The reserved instance is named C5RI.

C5RI is matched to C5PAYG-b.

You release C5PAYG-b and start another Linux instance of the same instance type named C5PAYG-c in Qingdao Zone C. C5RI is then matched to C5PAYG-c.

How is the instance size flexibility of reserved instances applied?

Only regional reserved instances provide instance size flexibility. Example:

Assume that you have one regional ecs.g5.4xlarge reserved instance. It can be applied to one ecs.g5.4xlarge pay-as-you-go instance, two ecs.g5.2xlarge pay-as-you-go instances, or four ecs.g5.xlarge pay-as-you-go instances.

Assume that you have a one-year regional ecs.g5.xlarge reserved instance. It can be applied to one ecs.g5.xlarge pay-as-you-go instance running for one year, or applied to one ecs.g5.2xlarge pay-as-you-go instance running for six months.

Do zonal reserved instances provide instance size flexibility?

No, zonal reserved instances do not provide instance size flexibility. A zonal reserved instance can be applied only to pay-as-you-go instances of the same instance type as it.

Do zonal reserved instances provide zone flexibility?

No, zonal reserved instances do not provide zone flexibility. A zonal reserved instance can be applied only to pay-as-you-go instances in the same zone as it.

Can a zonal reserved instance be changed into a regional one?

Yes, a zonal reserved instance can be changed into a regional one. You can change the scope of a reserved instance you purchased in the following ways:

- From a zone to a region
- From a region to a zone
- From one zone to another within the same region for a zonal reserved instance

Can the scope of a reserved instance be changed from one region to another?

No, the scope of a reserved instance cannot be changed from one region to another. For example, if you have a zonal reserved instance in Hangzhou Zone B, you can change the instance scope to another zone within the China (Hangzhou) region or change the instance into a regional reserved instance scoped to the China (Hangzhou) region. However, you cannot change the scope of the zonal reserved instance to a zone of another region, or change the instance into a regional reserved instance scoped to another region.

Can reserved instances be used across accounts?

No, reserved instances cannot be used across accounts.

Can reserved instances be used to cover the storage and network usage charges of pay-as-you-go instances?

No, reserved instances cannot be used to cover the storage or network usage charges of pay-as-you-go instances. Reserved instances cover the vCPU and memory costs of pay-as-you-go instances. For pay-as-you-go Windows instances, reserved instances also cover the image costs.

Can I configure a reserved instance to be applied to a specific pay-as-you-go instance?

No, you cannot configure a reserved instance to be applied to a specific pay-as-you-go instance. When multiple pay-as-you-go instances match the attributes of a reserved instance, the reserved instance is applied based on the optimized matching scheme.

How are reserved instances billed?

Reserved instances are billed separately and support the All Upfront, Partial Upfront, and No Upfront payment options.

The term of a reserved instance starts immediately after purchase. You are charged based on the payment option that you selected regardless of whether the reserved instance is matched to pay-as-you-go instances. The All Upfront option is most cost-effective.

When does a reserved instance take effect after it is purchased?

A reserved instance takes effect and is billed starting from the top of the hour of your purchase. The reserved instance expires at 00:00:00 of the day after its term end date. For example, if you purchase a reserved instance with a term of one year at 2019-02-26 13:45:00, the reserved instance will take effect and be billed starting from 2019-02-26 13:00:00. The reserved instance will expire at 2020-02-27 00:00:00. If you already have eligible pay-as-you-go instances when you purchase a reserved instance, the reserved instance will be applied to offset the bills generated by the pay-as-you-go instances starting from the hour of 13:00 to 14:00 on February 26, 2019 until the reserved instance expires.

After I modify, split, or merge a reserved instance, when will the operation take effect?

When reserved instances are modified, split, or merged, new reserved instances are generated and the original ones become invalid. The new reserved instances take effect and the original reserved instances become invalid. Both occur at the top of the hours of the modification, splitting, or merging operations. Assume that you split the ecs.g5.2xlarge zonal reserved instance RI1 into two zonal ecs.g5.xlarge reserved instances RI2 and RI3 at 13:45:00 on February 26, 2019. At 13:00:00 on February 26, 2019, RI1 becomes invalid and RI2 and RI3 take effect. Starting from 13:00:00 on February 26, 2019, the eligible reserved instance type is ecs.g5.xlarge, not ecs.g5.2xlarge any more. If RI2 and RI3 are matched to pay-as-you-go instances immediately after they take effect, RI2 and RI3 are also applied to offset the hourly bills of ecs.g5.xlarge pay-as-you-go instances starting from 13:00:00 on February 26, 2019.

Why isn't the No Upfront payment option displayed on the buy page?

The availability of this option depends on your ECS usage.

Can the payment option of a reserved instance be changed?

No, the payment options of reserved instances cannot be changed.

Can reserved instances be resold?

No, reserved instances cannot be resold.

Can I use reserved instances to cover the image costs of pay-as-you-go Windows instances?

Yes, reserved instances can be used to cover the image costs of pay-as-you-go Windows instances. This is because reserved Windows instances already include Windows images at no additional cost.

Can reserved instances be used to cover the image costs of pay-as-you-go Linux instances?

No, reserved instances cannot be used to cover the image costs of pay-as-you-go Linux instances.

Are the consumption details of reserved instances refreshed each hour?

Yes, the consumption details of reserved instances are refreshed each hour.

Can a reserved instance be applied to more than one pay-as-you-go instance at the same time?

Yes, a reserved instance can be applied to more than one pay-as-you-go instance at the same time. The reserved instance checks for eligible pay-as-you-go bills on an hourly basis and deducts fees based on its computing power.

> **Note** The computing power and term of each reserved instance are fixed. You cannot increase the computing power of a reserved instance by shortening its term.

Assume that you have a reserved instance with the following attributes:

- c5.large instance type
- Instances: 1 (indicating that the reserved instance can match one pay-as-you-go instance of the specified instance type)
- One-year term

The following examples demonstrate how the reserved instance is applied based on the pay-as-you-go instances that exist:

- Six c5.large pay-as-you-go instances exist for one hour each. Each of these pay-as-you-go instances consumes one hour of computing power equal to that delivered by the c5.large reserved instance every hour. The reserved instance will be applied randomly to one of the pay-as-you-go instances. You cannot configure the reserved instance to be applied to all six pay-as-you-go instances by shortening the term of the reserved instance to two months.
- Six c5.large pay-as-you-go instances exist for 10 minutes each. The six instances consume 10 minutes of computing power each and in total consume the amount of computing power that the c5.large reserved instance can deliver every hour. The reserved instance will be applied to all six pay-as-you-go instances.
- Six c5.large pay-as-you-go instances exist for 15 minutes each. The six instances consume 15 minutes of computing power each, in total exceeding the amount of computing power that the reserved instance can deliver every hour. The reserved instance will be applied randomly to the pay-as-you-go instances to offset the charges for one hour of computing power.

Does VNC allow multiple users to log on simultaneously?

No, VNC allows a single user to log on at a time.

What can I do if I forget the remote connection password?

You can reset your remote connection password. For more information, see Modify the VNC password.

Why am I unable to connect to VNC even after I reset my remote connection password?

If the instance to which you are connecting is not I/O optimized, you must restart the instance from the ECS console or by calling the RebootInstance operation for the new password to take effect.

> **Note** If you restart the instance from within the instance itself, the new password will not take effect.

I was prompted with an authentication failure when I attempted to connect to VNC. What can I do?

Authentication will fail if the entered password is not correct. Perform the following troubleshooting operations:

1. Enter the correct remote connection password.
2. If you forget your password, you can reset it and try again. For more information, see Modify the VNC password.

> Note If the instance to which you are connecting is not I/O optimized, you must restart the instance from the ECS console or by calling the RebootInstance operation for the new password to take effect.

What can I do if a black screen appears while I am connected to VNC?

A black screen indicates that the instance is in sleep mode. Perform the following operations based on your operating system:

- For a Linux instance, click your mouse or press any key to activate the instance.
- For a Windows instance, you can choose **Send Remote Call** > **CTRL+ALT+DELETE** in the upper-left corner to go to the logon interface.

What can I do if VNC cannot be accessed?

You can use a browser to access VNC for troubleshooting. For example, use Google Chrome to access VNC, and press the F12 key to open the developer tools pane. Then click the **Console** tab and identify errors based on the information displayed.

Why am I unable to use Internet Explorer 8 to access VNC?

VNC supports Internet Explorer 10 and later.

We recommend that you use Google Chrome because it is most compatible with the Alibaba Cloud Management console.

When I use Firefox to access VNC, an error message is returned indicating that the secure connection has failed. What can I do?

This problem occurs if the encryption algorithm used by your version of Firefox is different from that used by VNC.

We recommend that you use Google Chrome because it is most compatible with the Alibaba Cloud Management console.

How do I remotely log on to a Linux instance?

Linux instances support SSH for remote connection. You can use one of the following methods to remotely log on to a Linux instance:

- Connect to a Linux instance by using VNC.
- Connect to a Linux instance by using a username and password.
- Connect to a Linux instance by using an SSH key pair.
- Connect to a Linux instance from a mobile device.

What are the default username and password for remote logon to the operating system of an ECS instance?

The default username varies by operating system.

- For a Windows instance, the default username is `administrator`.
- For a Linux instance, the default username is `root`.

The password for remote logon to the operating system of an instance is set by you when you create the instance. For more information, see Create an instance by using the provided wizard. If you forget the password, you can reset it. For more information, see Reset the logon password of an instance.

> **Note** This password is used to remotely log on to the instance operating system, not to VNC.

What can I do if I fail to deploy a Data Plane Development Kit (DPDK) application on an ECS instance?

We recommend that you deploy DPDK applications on g5ne instances. For more information about the g5ne instance types, see Instance families.

If you deploy DPDK applications on ECS instances of the sixth generation and later instance types (such as g6, c6, and r6 instance types), the bound igb_uio port may not be detected when you use Pktgen-DPDK to perform tests, and the following error is reported: `EAL:`

`eal_parse_sysfs_value(): cannot open sysfs value`

`/sys/bus/pci/devices/0000:00:09.0/uio/uio0/portio/port0/start`

If you are using DPDK 19.05, you can troubleshoot the problem by installing patches. You can submit a ticket to obtain the patches and run the following commands to install them:

```
cd dpdk
copy virtio0.95_bar0_memebar.diff to dpdk
git apply virtio0.95_bar0_memebar.diff
```

Can I upgrade the instance types and configurations of subscription instances?

Yes, you can upgrade the instance types and configurations of subscription instances. For more information, see Upgrade configurations of subscription instances.

Can I upgrade the instance types and configurations of pay-as-you-go instances?

Yes, but you must stop pay-as-you-go instances before you can upgrade their instance types and configurations. You can upgrade the instance types and configurations of pay-as-you-go instances by following the instructions in Change the instance type of a pay-as-you-go instance or by calling the ModifyInstanceSpec operation.

How long does it take to upgrade the instance type of an instance?

- Subscription instances do not need to be stopped for their instance types to be upgraded. It takes about 15 minutes to upgrade the instance type of a subscription instance.

- Pay-as-you-go instances must be stopped before their instance types can be upgraded. It also takes about 15 minutes to upgrade the instance type of a pay-as-you-go instance.
- You can upgrade the bandwidths of instances without stopping the instances. The upgrade process takes about five minutes.

How is the fee incurred for an instance upgrade calculated?

The incurred fee and its calculation method are displayed in the ECS console when you upgrade the instance type and configurations of an instance. You can also view the billing details on the Account Overview page.

Does upgrading ECS instances affect my cloud service configurations?

Pay-as-you-go instances must be stopped before their instance type can be upgraded. After you upgrade the instance type of a subscription instance, you must restart the instance for the new configurations to take effect. The upgrade operation will interrupt the services running on the instance for a short period of time. We recommend that you upgrade instances during off-peak hours. Instances can seamlessly resume services after upgrades without environment reconfiguration.

How do I upgrade ECS resources?

For information about how to upgrade ECS resources, see Overview of instance upgrade and downgrade.

- With the exception of ECS instances that use local storage, ECS instances allow their CPU and memory resources to be scaled and their bandwidths to be upgraded while the instances are running. You can also downgrade the configurations of ECS instances.
- A maximum of 16 data disks can be attached to each ECS instance. You cannot reduce the size of a data disk after it has been extended.
- The bandwidth of each ECS instance is measured in Mbit/s and can range from 0 Mbit/s to 200 Mbit/s. You can also temporarily upgrade the bandwidth of an ECS instance, or downgrade the configurations of the instance when you renew it.

I have upgraded the configurations of an instance but no changes have taken effect. Why?

After you upgrade the configurations of an instance, you must restart it from the ECS console or by calling an API operation for the new configurations to take effect.

My ECS instance was stuck in the Starting state, and AliyunService was disabled or deleted. What can I do?

Problem description: After an ECS instance was started, it stayed in the Starting state for an extended period of time and then automatically stopped. You logged on to the instance and found that AliyunService had been deleted or disabled in the system services.

Solution:

- If AliyunService was disabled:
  i. Change the state of AliyunService to automatic.
  ii. Restart the instance.
- If AliyunService was deleted:

i. Run the following command to add AliyunService to the instance:

```
sc create AliyunService type= "own" start= "auto" binPath= "C:\Program Files\AliyunService\AliyunService.exe -d" tag= "no" DisplayName= "AliyunService"
```

> **Note** Make sure that you leave a space after the equal sign `(=)`.

ii. Find the registry key *HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\services\AliyunService*, and change `c:\Program Files\AliyunService\AliyunService.exe -d` to `"c:\Program Files\AliyunService\AliyunService.exe" -d`.

iii. Restart the instance.

How do I use f1 instances?

After you create an f1 instance, Alibaba Cloud shares an FPGA development image to you. Currently, only CentOS 7u2 images are supported. The FPGA development image includes the complete Intel Quartus development suite and the f1 instance constraint files to provide a complete cloud development environment.

> **Note** You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.

The following section describes the basic workflow for using an f1 instance:

1. After development is completed, you can generate an intermediate QAR file during the compilation stage and upload the file to an OSS bucket. Currently, you can only upload this file to an OSS bucket within the China (Hangzhou) region. Then, you can register the QAR file information with Alibaba Cloud by calling an API operation.

   We recommend that you use the free Intel Quartus development suite to complete development, compilation, and simulation operations on the cloud.

2. Alibaba Cloud will verify your registration request for the QAR file, and then send you an email including an FPGA image ID if the verification succeeds.
3. To deploy the image, you can call an API operation with the f1 instance ID and FPGA image ID specified to associate the instance with the image.

   You can initiate the association operation in all scenarios where ECS API is available.

   - If the f1 instance has never been associated with any FPGA image, initiate the association operation.
   - If the f1 instance was previously associated with an FPGA image and had the image loaded, erase the FPGA image from the f1 instance before you initiate the association operation.

4. After you associate the FPGA image with the instance, call an API operation to load the image.

   You must initiate the load operation from the f1 instance. Then the underlying service of Alibaba Cloud will burn the associated FPGA image to the corresponding FPGA on the instance.

If you want to restore the f1 instance to its initial state, call an API operation to erase the burned FPGA image from the f1 instance.

For more information about how to manage f1 instances, see the following topics:

- Create an f1 instance
- Use OpenCL on an f1 instance
- Use RTL Compiler on an f1 instance

How do I upload files by using the FTP tool in macOS?

Method 1: Upload files by using the Terminal of macOS

Open the Terminal in macOS or iTerm2 for Mac (click here to download iTerm2). Make sure that you select the correct destination path.

1. Connect to the FTP server.
2. Access the destination directory. In Windows, use the working directory as the destination directory. In Linux, switch to *htdocs* and use the htdocs directory as the destination directory.
3. Run the **put** command to upload files.

Method 2: Upload files by using a third-party tool

1. Download Yummy FTP.
2. Install Yummy FTP.
3. Enter the server IP address, username, and password. Set Protocol to Standard (FTP). Set Port to 21 or a different port number that you are using, and leave the SSH key field unselected.
4. Click **Connect**.
5. In the right-side pane, select the destination directory. In Windows, use the current working directory. In Linux, select the htdocs directory. In the left-side pane, select files and click the Upload icon to upload the files.

> **Note** If you try to install Yummy FTP and are prompted with a message similar to "Your security preferences only allow the installation of applications from the Mac App Store and authorized developers", perform the following steps:
>
> i. Choose **System Preference** > **Security and Privacy**.
>
> ii. Click the security lock in the lower-left corner of the window and enter the administrator password.
>
> iii. Set **Allow apps download from** to **Anywhere**.
>
> Then, you can upload files by using Yummy FTP.

If you have further questions, submit a ticket.

How do I apply for an ICP filing for my domain name after I purchase an ECS instance?

For each ECS instance, you can apply for a limited quantity of service identification numbers for ICP filings. For more information, see 备案服务器（接入信息）准备与检查.

For information about how to apply for an ICP filing, see ICP filing application overview.

An ECS instance fails to load the kernel to start up. What can I do?

Problem description: The system does not respond when you select an option from the GRUB menu at system startup. After you have mounted the LiveCD image to the ECS instance, you are logged on to the instance and confirm that the file system privileges are correct and that the message logs show no exceptions.

Cause: The system has been attacked by ransomware.

Solution: Back up your data and re-initialize the system.

How do I change the logon password within an instance?

For information about how to change the logon password within an instance, see Change the logon password of an instance by connecting to the instance.

Why am I unable to add sound or video cards to ECS instances?

The servers that Alibaba Cloud ECS provides are not multimedia servers and do not provide sound or video card components. Sound or video cards cannot be added to ECS instances.

Can I transfer the unused time of an ECS instance to another ECS instance?

No, the unused time of an ECS instance cannot be transferred to others. If you want both higher flexibility and cost-effectiveness, we recommend that you use a combination of reserved instances and pay-as-you-go instances. For more information, see Reserved instance overview.

Do ECS instances provide databases by default?

No, ECS instances do not provide databases by default. To use database services, perform the following operations:

- Deploy your own database.
- Purchase ApsaraDB for RDS provided by Alibaba Cloud.
- Use a database image provided in Alibaba Cloud Marketplace.

Can I build a database on an ECS instance?

Yes, you can use the image copy feature to migrate an ECS instance from a region inside mainland China to a region outside mainland China. you can install database software and configure a database environment on an ECS instance. You can also purchase ApsaraDB for RDS separately.

Do ECS instances support Oracle databases?

Yes, ECS instances support Oracle databases. Before you install an Oracle database, we recommend that you perform a performance stress test on the target ECS instance to ensure that the instance satisfies the read/write requirements of the database.

Are public IP addresses and private IP addresses independent? Can I specify or add IP addresses?

In the classic network, public and private IP addresses are independent of each other. Private IP addresses in the classic network are used for communication between ECS instances and between ECS instances and OSS or ApsaraDB for RDS instances. If the public bandwidth of an ECS instance is set to 0 Mbit/s, no public IP address is assigned to the instance. In normal cases, public and private IP addresses in the classic network do not change. You cannot specify, select, or add IP addresses in the classic network.

In VPCs, NAT gateways map public IP addresses to private IP addresses. You can add IP addresses by specifying or automatically assigning secondary private IP addresses to ENIs.

Can load balancing be implemented for a single ECS instance?

Both Linux and Windows ECS instances can be load-balanced. You must ensure that the configurations of ECS instances serving as web servers meet the requirements for website code to run. Load balancing can be implemented for as few as one ECS instance in an account. However, we recommend you implement load balancing for two or more ECS instances.

Can I change the region of an ECS instance?

No, the regions of purchased ECS instances cannot be changed. To change the region of an ECS instance, you can use the `ACS-ECS-CloneInstancesAcrossRegion` public template provided by Operation Orchestration Service (OOS) to copy the instance to another region. The new instance and the original instance have identical disk data but different IP addresses.

Can I adjust the partition size of a purchased disk?

For system security and stability purposes, system disks cannot be repartitioned on neither Windows nor Linux instances. If you use a third-party tool to perform repartitioning, unknown exceptions such as system failure and data loss may occur.

Repartitioning data disks may cause data loss. We recommend that you do not repartition data disks.

How do I view subscription ECS instances within all regions in my account?

You can go to the Renewal page to view subscription ECS instances within all regions in your account.

1. Log on to the ECS console.
2. In the top navigation bar, choose **Billing** > **Renew**.

When can I forcibly stop an ECS instance? What are the consequences?

If an instance cannot be stopped by a proper shutdown procedure, you can forcibly stop it. A forced stop is equivalent to disconnecting the power from the instance, and can result in loss of unsaved data.

Why am I unable to reactivate my ECS instance?

An instance may fail to be reactivated due to one of the following reasons:

- Your account has overdue payments. Pay the outstanding bills and try again.

- The system is busy. Try again later.
- Resources of the specified instance type are sold out.

> Note You can go to the ECS Instance Types Available for Each Region page to view the instance types available in each region.

Why has an ECS instance with release protection enabled been automatically released from a scaling group?

Auto Scaling can automatically release an ECS instance created by a scale-out event even if you have enabled release protection for the instance from the ECS console or by calling the ModifyInstanceAttribute operation.

To prevent the ECS instance from being automatically released, you must change its state to Protected in the Auto Scaling console. For more information, see Put an ECS instance into the Protected state.

What is the AliVulfix process in an ECS instance?

The AliVulfix process is an Alibaba Cloud Security program that scans ECS instances for vulnerabilities.

How do I protect ECS instances against attacks?

ECS instances use Alibaba Cloud Security to defend against DDoS attacks. CloudMonitor deployed on ECS instances can automatically detect network attacks and scrub suspicious traffic. Alibaba Cloud implements black hole filtering to protect ECS instances against high-volume attacks. To strengthen security protection, we recommend that you install security software and disable ports that are not commonly used.

What security services does Alibaba Cloud provide?

Alibaba Cloud Security is powered by the robust data analysis capabilities of the Alibaba Cloud cloud computing platform to provide a comprehensive set of security services such as security vulnerability detection, website trojan detection, host intrusion detection, and anti-DDoS protection.

I have already renewed an expired Linux instance but I am still unable to access the website hosted on it. What can I do?

Problem description: A Linux ECS instance is in the Stopped state after it has expired. After you have renewed and restarted the instance, you still cannot access the website hosted on it.

Cause: This may be because the website service has not been started.

Solution:

1. Connect to the instance and run the following command to check whether the website service has been started:

   ```
   # netstat -nltp //Check whether port 80 on the instance is being listened on.
   ```

2. If no information about port 80 is displayed in the command output, the website service has not been started. Run a command to manually start the website service and relevant

services.

In Linux, websites are typically developed based on PHP and MySQL.

- In Apache, you only need to start the website service and MySQL.

```
#/etc/init.d/httpd start       //Start the website service. This command is applicable to Apache.
#/etc/init.d/mysqld start      //Start MySQL.
```

- In NGINX, you must start the website service, PHP, and MySQL.

```
#/etc/init.d/nginx start       //Start the website service. This command is applicable to NGINX.
#/etc/init.d/php-fpm start     //Start PHP.
#/etc/init.d/mysqld start      //Start MySQL.
```

3. Check again whether the website service has been started.

```
#netstat -nltp //Check whether port 80 on the instance is being listened on.
```

4. After the website service has been started, access the website again.

If the problem persists, submit a ticket.

How do I activate a Windows ECS instance within a VPC?

To activate a Windows ECS instance within a VPC, you must use a specific KMS domain name. For more information, see How to activate the VPC-Connected Windows instances using KMS servers.

How do I query, partition, and format the disks of a Linux instance?

You can run the **df -h** command to check the capacity and usage of disks, and run the **fdisk -l** command to view disk information. For information about how to partition and format disks in Linux instances, see Format a data disk for a Linux instance.

How do I upload files to a Linux instance?

You can use the FTP service to upload files to a Linux instance.

How do I change the owner and owner group of directories and files on a Linux instance?

If the file or directory permissions are not correctly configured on the web server, a 403 error occurs when you access a website hosted on the instance. Before you adjust a file or directory, you must identify the identity under which the file or directory process is running.

You can run the **ps** and **grep** commands to query the identities under which processes are running.

You can run the **ls -l** command to query the owners and owner groups of files and directories.

To change the owners and owner groups, run the chown command. For example, you can run the **chown -R www.www /alidata/www/phpwind/** command to change the owner and owner group of all files and directories under the */alidata/www/phpwind* directory to Account www.

How do I update the software repository for Linux instances?

You can use an automatic tool to update the software repository for Linux instances. For more information, see Automatic source updating tool for Linux ECS.

What limits apply to the transfer and change of public IP addresses of ECS instances?

The following limits apply to the transfer and change of public IP addresses of ECS instances:

- In the classic network:
  - You cannot transfer public IP addresses across accounts.
  - The public IP address of an ECS instance can be changed within six hours after the instance is created, and can be changed a maximum of three times. For more information, see Change the public IP address of an ECS instance.
  - If Anti-DDoS Pro is deployed, you can change the IP address of an ECS instance a maximum of 10 times from the Anti-DDoS Pro console. For more information, see the Change ECS IP section in Anti-DDoS Pro User Guide.
- In VPCs:
  - You cannot transfer public IP addresses or EIPs across accounts.
  - If no public IP address has been assigned to your instance, you can associate an EIP with the instance. You can replace the public IP address of your instance with an EIP.
  - If a public IP address has been assigned to your instance during instance creation:
    - This public IP address can be changed within six hours after the instance is created, and can be changed a maximum of three times. For more information, see Change the public IP address of an ECS instance.
    - You can convert this public IP address to an EIP and then replace the EIP. For more information, see Convert the public IP address of a VPC-type instance to an Elastic IP address.

If the problem persists, submit a ticket.

Can I access amazon.com from my ECS instance?

You can access amazon.com from your ECS instance if the instance can connect to the Internet properly.

Why am I unable to access a website hosted outside mainland China after I log on to my ECS instance?

You can access a website hosted outside mainland China from your ECS instance only when the website complies with the laws, regulations, and regulatory requirements of the country or region where your instance is located. Ensure that your ECS instance can connect to the Internet properly and that the website complies with the preceding laws, regulations, and regulatory requirements.

I cannot purchase more pay-as-you-go instances. What can I do?

If you have reached the maximum number of pay-as-you-go instances that you can purchase, you cannot purchase more pay-as-you-go instances. For more information, see the "Instance limits" section in Limits. You can log on to the ECS console and click Privileges in the upper-right corner of the Overview page to view your resource quotas. For more information, see View quotas (old version).

How can I view the resource quota?

For more information about how to view the limits and quotas of resources, see Limits.

Will I still be charged for a pay-as-you-go instance after it is stopped either manually or due to an overdue payment?

Stopped due to an overdue payment: A pay-as-you-go instance will be automatically stopped when a payment becomes overdue for it, and stop incurring charges. Instances will not always stay in the Stopped state after they are stopped due to overdue payments. For more information, see Pay-as-you-go.

Manually stopped: You can stop a running pay-as-you-go instance from the ECS console or by calling the StopInstance operation. When the instance is stopped, its state changes to **Stopped**. Billing of stopped pay-as-you-go instances depends on their network types.

- VPC: You can enable the No Fees for Stopped Instances (VPC-Connected) feature.
  - After the feature is enabled, billing of pay-as-you-go instances in VPCs will start when the instances are created, stop when the instances enter the **Stopped** state, and resume when the instances are started again. When a pay-as-you-go instance enters the Stopped state, the No Fees for Stopped Instances (VPC-Connected) feature stops only the vCPUs, memory, and public IP address of the instance from incurring charges. Other resources such as disks and EIP of the instance will continue to incur charges. For more information, No Fees for Stopped Instances (VPC-Connected).
  - If the No Fees for Stopped Instances (VPC-Connected) feature is disabled, pay-as-you-go instances continue to incur charges when they are stopped.
- Classic network: ECS instances in the classic network will be billed regardless of whether they are in the **Stopped** state.

What can I do if an order cannot be placed to change the billing method of an instance from pay-as-you-go to subscription?

You may be unable to place the order due to one of the following reasons:

- The instance is in a state that does not support changes to the billing method. For example, the instance has an unpaid order.
- Changes to the billing method are not allowed due to an upcoming scheduled automatic release.
- Changes to the billing method are not allowed because the instance information has changed.
- A previous order to change the billing method of the instance has not been paid.

If one of the preceding errors is reported, adjust the instance accordingly.

How long after an order is paid does it take to change the billing method of an instance from pay-as-you-go to subscription?

The billing method of your ECS instance will be changed after the order is paid. It can take up to four seconds to change the billing method of 20 instances. After the change is completed, you can see that the billing method of your instance has been changed to **Subscription** in the console.

What can I do if the billing method of an instance cannot be changed from pay-as-you-go to subscription?

Submit a ticket.

When I change the billing method of an instance from pay-as-you-go to subscription, will the billing method for network usage of the instance also change?

No, the billing method for network usage of the instance will not change. Only the billing method of instances and disks can be changed from pay-as-you-go to subscription. For information about how to change the billing method for network usage, see Overview of instance upgrade and downgrade.

I have an unpaid order to change the billing method of an instance from pay-as-you-go to subscription. If I upgrade the configurations of the instance, will the order still be valid?

An order is created when you change the billing method of your instance from pay-as-you-go to subscription. You must pay for the order to complete the change. If you upgrade the configurations of the instance before the order is paid, the order payment cannot be completed because the instance components are different and the original order no longer matches. If you still want to change the billing method, you must cancel the unpaid order and place a new order.

What can I do if the billing method of an instance cannot be changed from subscription to pay-as-you-go?

You may be unable to change the billing method of an instance from subscription to pay-as-you-go due to one of the following reasons:

- The instance is in a state that does not support changes to the billing method. For example, the instance has an unpaid order.
- The instance is in the **Expired** state.
- The instance information has changed. For example, the bandwidth of the instance has been temporarily upgraded.

If one of the preceding errors is reported, adjust the instance accordingly. If the problem persists, submit a ticket.

When I try to change the billing method of a disk in an ECS instance, an error message is returned indicating that I have already changed the billing method three times. What does this mean?

Each ECS instance can only be downgraded a maximum of three times. Downgrade operations include downgrades of instance specifications, bandwidth downgrades, and the change of the disk billing method from subscription to pay-as-you-go.

Why am I unable to change a pay-as-you-go instance to a subscription one?

The ECS instance whose billing method you want to change must meet the following requirements:

- The instance belongs to your account.
- The instance is of none of the following instance types:
  - Instance types in Generation I instance families: t1, s1, s2, s3, m1, m2, c1, and c2.
  - Instance types in the n1, n2, and e3 instance families.

  > **Note** For more information about these instance types, see Phased-out instance types.

- The instance cannot be a preemptible instance.
- You have no unpaid order to switch the billing method of the instance.

  If you have an unpaid order to switch the billing method of the instance, you must cancel the unpaid order and then place another order to switch the billing method.

- Automatic release is not set for the instance.

  If automatic release has been set for an instance, you must disable the automatic release configuration and then switch the billing method. For more information, see Disable automatic release.

- The instance is in the **Running** or **Stopped** state.

  Example: An order to switch the billing method has been placed when the ECS instance is in the Running or Stopped state. However, the instance status changed when the payment was attempted. The preceding requirement is not met. The order fails and the billing method does not change. You can go to the Billing Management console and pay for the order when the instance is in the Running or Stopped state again.

How do I view the expiration time of a subscription instance?

You can log on to the ECS console and go to the **Instances** page. Then, you can view the expiration time of your subscription instance in the **Billing Method** column.

> **Note** If the **Billing Method** column is not displayed, click the icon in the upper-right corner. In the dialog box that appears, select **Billing Method** and click **OK**.