Alibaba Cloud
Elastic Compute Service

Product Introduction

Issue: 20180917

Legal disclaimer

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

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

secrets. No part of the Alibaba Cloud website, product programs, or content shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates).

6. Please contact Alibaba Cloud directly if you discover any errors in this document.

Generic conventions

Table -1: Style conventions

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

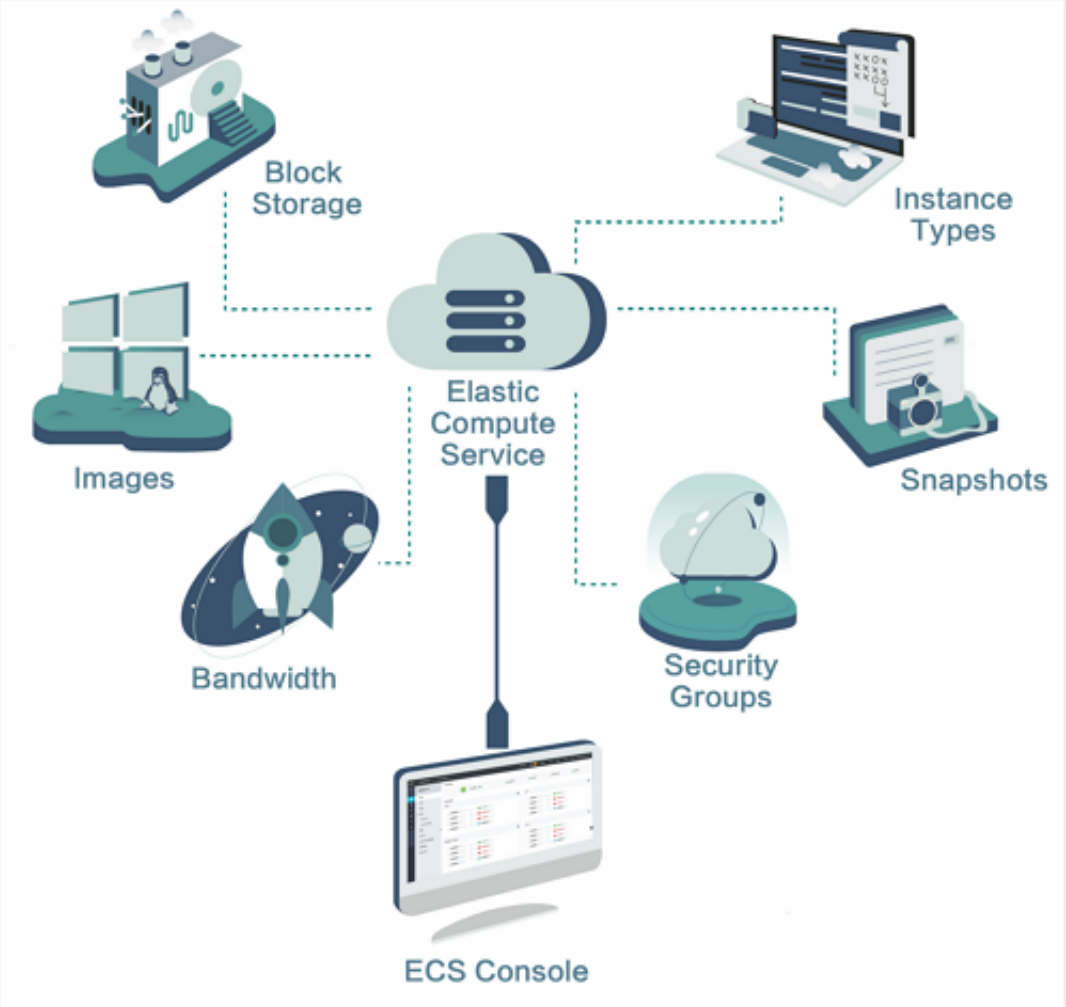
Contents

1 What is ECS?

This article gives a brief introduction to what ECS is, and the resources and services that it involves.

Elastic Compute Service (ECS) is a type of computing service that features elastic processing capabilities. ECS has a simpler and more efficient management mode than physical servers. You can create instances, change the operating system, and add or release any number of ECS instances at any time to fit your business needs. An ECS instance is a virtual computing environment that includes CPU, memory, and other basic computing components. An instance is the core component of ECS and is the actual operating entity offered by Alibaba Cloud. Other resources, such as disks, images, and snapshots, can only be used in conjunction with an ECS instance.

The following figure illustrates the concept of an ECS instance. You can use the *ECS console* to configure the instance type, disks, operating system, and other affiliated resources.



Basic concepts

It is helpful to understand the following concepts before you use ECS:

- *Region and zone*: A physical location where a data center is located.
- *ECS instance*: A virtual computing environment that includes the CPU, memory, operating system, bandwidth, disks, and other basic computing components.
- *Instance types*: The specifications of an ECS instance, including the number of vCPU cores, memory, and networking performance. The instance type of an ECS instance determines its compute capability.
- *Images*: A running environment template for ECS instances. It generally includes an operating system and preinstalled software.
- *Block storage*: Block level storage products for your ECS, including *Elastic block storage* based on the distributed storage architecture and *local disks* located on the physical server that an ECS instance is hosted on.
- *Snapshots*: A copy of the data on an elastic block storage device as it was at a specific point in time.
- *Network types*: Alibaba Cloud provides two network types, including
  - Virtual Private Cloud (VPC): A private network established in Alibaba Cloud. VPCs are logically isolated from other virtual networks in Alibaba Cloud. For more information, see *What is VPC*.
  - Classic network: A network majorly deployed in the public infrastructure of Alibaba Cloud.
- *Security group*: A logical group of instances that are in the same region and have the same security requirements and mutual trust. A security group works as a virtual firewall for the ECS instances inside it.

Related services

Alibaba Cloud marketplace is an online market. You can purchase software infrastructure, developer tools, and business software provided by third-party partners. You can become a marketplace service provider.

Auto Scaling enables you to dynamically scale your computing capacity up or down to meet the workload of your ECS instances according to scaling policies you specify. It also reduces the need of manual provision. For more information, see *What is Auto Scaling*.

Container Service enables you to manage the lifecycle of containerized applications by using Docker and Kubernetes. For more information, see *What is Container Service*.

Server Load Balancer distributes the incoming traffic among multiple ECS instances according to the configured forwarding rules. For more information, see *What is Server Load Balancer*.

CloudMonitor manages ECS instances, system disks, Internet bandwidth, and other resources. For more information, see *What is CloudMonitor*.

Server Guard (Server Security) provides real-time awareness and defense against intrusion events, which safeguards the security of your ECS instances. For more information, see *What is Server Guard*.

Anti-DDoS Basic prevents and mitigates DDoS attacks by routing traffic away from your infrastructure. Alibaba Cloud Anti-DDoS Pro safeguards your ECS instances under high volume DDoS attacks. For more information, see *What is Anti-DDoS Basic* and *What is Anti-DDoS Pro*.

Alibaba Cloud SDK enables you to access Alibaba Cloud services and to manage your applications by using the programming language of your choice. For more information, see *Developer Resources*. You can use *OpenAPI Explorer* to debug ECS API and generate the SDK Demo.

Operations

Alibaba Cloud provides an intuitive operation interface for you to manage your ECS instances. You can log on to the *ECS console* to operate ECS instances. For more information, see *User Guide*.

You can use API to manage your ECS instances. For more information, see *API References*. You can also use Alibaba Cloud CLI to call API to manage ECS instances. For more information, see *Alibaba Cloud Command Line Interface*.

ECS pricing and billing

ECS supports both Subscription and Pay-As-You-Go billing methods. For more information, see *Billing methods*.

For pricing details, see the *Pricing* page.

Learning Path

You can use the *ECS Learning Path* as a mentor to learn ECS basics or add to your knowledge.

2 Benefits of ECS

Compared with Internet Data Centers (IDCs) and server vendors, ECS has benefits in terms of availability, security, and elasticity.

Availability

Alibaba Cloud adopts more stringent IDC standards, server access standards, and O&M standards to guarantee data reliability and high availability of cloud computing infrastructure and cloud servers.

In addition, each Alibaba Cloud region consists of multiple zones. For greater fault tolerance, you can build active/standby or active/active services in multiple zones. For a finance-oriented solution with three IDCs in two regions, you can build fault tolerant systems in multiple regions and zones. Those services include disaster tolerance and backup, which are supported by the mature solutions built by Alibaba Cloud.

Switching between services is smooth within the Alibaba Cloud framework. For more information, see *E-Commerce Solutions*. Alibaba Cloud industry solutions support a variety of services, such as finance, E-commerce, and video services.

Alibaba Cloud provides you with the following support services:

- Products and services for availability improvement, including cloud servers, Server Load Balancer, multi-backup databases, and Data Transmission Services (DTS).
- Industry partners and ecosystem partners that help you build a more advanced and stable architecture and guarantee service continuity.
- Diverse training services that enable you to connect with high availability from the business end to the underlying basic service end.

Security

For cloud computing users, security and stability are priorities. Alibaba Cloud has passed a host of international information security certifications, including ISO 27001 and MTCS, which demand strict confidentiality of user data and user information, as well as user privacy protection. We recommend that you use ECS in an *Alibaba Cloud Virtual Private Cloud (VPC)*.

- **Alibaba Cloud VPC offers more business possibilities.** You only need to perform a simple configuration to connect your business environment to global IDCs, making your business more flexible, stable, and extensible.

- **Alibaba Cloud VPC can connect to your IDC** through a leased line to build a hybrid cloud architecture. You can build a more flexible business with robust networking derived from Alibaba Cloud's various hybrid cloud solutions and network products. A superior business ecosystem is made possible with Alibaba Cloud's ecosystem.
- **Alibaba Cloud VPC is more stable and secure.**

  Stable: After you build your business on VPC, you can update your network architecture and obtain new network functions on a daily basis as the network infrastructure evolves constantly, allowing your business to run steadily. You can divide, configure, and manage your network on VPC according to your needs.

  Secure: VPC features traffic isolation and attack isolation to protect your services from attack traffic on the Internet. By building your business on VPC, the first line of defense is established.

VPC provides a stable, secure, fast-deliverable, self-managed, and controllable network environment. VPC hybrid cloud brings the technical advantages of cloud computing to traditional industries, in addition to industries and enterprises that are not engaged in cloud computing.

Elasticity

Elasticity is a key benefit of cloud computing. By using Alibaba Cloud, you can have all the IT resources necessary to build an IT company of medium size provisioned within minutes. The available resources and capacities can meet the requirements of most companies, allowing their applications built on the cloud to handle a huge volume of transactions without problems.

- Elastic computing

  Elastic computing contains the following types of scalability:

  - Vertical scaling involves modifying the configurations of a server. After you purchase ECS or storage capacity of Alibaba Cloud, you can configure your server with great flexibility based on your actual transaction volume, whereas it may be difficult to change configurations in the traditional IDC model. For more information about vertical scaling, see *Change configurations*.
  - **Horizontal scaling** allows the re-division of resources between applications. For example, at peak hours for game or live video streaming apps, in the traditional IDC model, your hands may be tied if additional resources are required when you are already at full capacity. Cloud computing uses elasticity to provide additional resources to you over that period. When the period ends, you can release unnecessary resources to reduce your business costs. By using both horizontal scaling and auto-scaling, you can determine how and

when you scale your resources or apply your scaling based on business loads. For more information about horizontal scaling, see *Auto Scaling*.

- Elastic storage

  Alibaba Cloud provides elastic storage. In the traditional IDC model, if more storage space is required, you can only add servers, but the number of servers that you can add is limited. In the cloud computing model, however, the sky is the limit. Order as much storage space as you need to meet business demand. For more information about elastic storage, see *Resize a disk*.

- Elastic network

  Alibaba Cloud features elastic network as well. When you purchase Alibaba Virtual Private Cloud (VPC), you can set network configurations to be the same as those of data centers. In addition, VPC has the following benefits:

  - Interconnection between data centers
  - Separate secure domains in data centers
  - Flexible network configurations and planning within the VPC

  For more information about elastic networks, see *Virtual Private Cloud*.

Alibaba Cloud incorporates elasticity in computing, storage, network, and business architecture design. By using Alibaba Cloud, you can build your business portfolio in any way you want.

Comparison between ECS and traditional IDCs

The table lists the benefits of ECS compared with traditional IDCs.

Item	ECS	Traditional IDCs
Equipment rooms	Provides independently developed DC powered servers with low PUE.	Provides traditional AC powered servers with high PUE.
	Provides backbone equipment rooms with high outbound bandwidth and dedicated bandwidth.	Provides equipment rooms with various quality levels and shared bandwidth primarily, difficult for users to choose.
	Provides multiline BGP equipment rooms, enabling smooth and balanced access throughout the country.	Provides equipment rooms with single or dual line primarily.

Item	ECS	Traditional IDCs
Ease of operation	Provides built-in mainstream operating systems, including activated Windows OS.	Purchases and installs operating system manually.
	Switches operating systems online.	Reinstalls operating systems manually.
	Provides a Web-based console for online management.	Users must manage and maintain network manually.
	Provides mobile phone verification for password setting, increasing data security.	Has difficulty in resetting passwords, and exposes high risk of password cracking.
Disaster recovery and backup	Each data segment has multiple copies. When one copy is corrupted, the data can be quickly restored.	Users must build disaster recovery environment by themselves, and use traditional storage devices.
	Users can customize automatic snapshot policies to create automatic snapshots for data recovery.	Users must restore all corrupted data manually.
	Faults can be recovered fast and automatically.	Faults cannot be recovered automatically.
Security and reliability	Effectively prevents MAC spoofing and ARP attacks.	Fails to prevent MAC spoofing and ARP attacks.
	Effectively defends against DDoS attacks by using black holes and cleaning traffic.	Needs additional costs for devices for traffic cleaning and black hole shielding systems.
	Provides additional services, such as port scanning, Trojan scanning, and vulnerability scanning.	Typically encountered problems such as vulnerability, Trojan, and port scanning.
Flexible scalability	Activates cloud servers on demand and upgrades configurations online.	Needs a long time for server delivery.
	Adjusts outbound bandwidth as required.	One-off purchase of outbound bandwidth, unable to adjust.
	Combines with Server Load Balancer online, enabling	Uses hardware-based server load balancing, which is

Item	ECS	Traditional IDCs
	scaling up applications quickly and easily.	expensive and extremely difficult to set up.
Cost effectiveness	Low cost.	High cost.
	Small up front investment.	Large up front investment, possible waste of resources.
	Purchases on demand and pay as you go, meeting requirements for constant business changes.	Purchases up front to meet configuration requirements for peak hours.

3 Scenarios

ECS is a highly flexible solution. It can be used independently as a simple web server, or used with other Alibaba Cloud products, such as Object Storage Service (OSS) and Content Delivery Network (CDN), to provide advanced solutions.

ECS can be used in the following applications.

Official corporate websites and simple web applications

During the initial stage, corporate websites have low traffic volumes and require only low-configuration ECS instances to run applications, databases, storage files, and other resources. As your business expands, you can upgrade the ECS configuration and increase the number of ECS instances at any time. You no longer need to worry about insufficient resources during peak traffic.

Multimedia and large-traffic apps or websites

ECS can be used with OSS to store static images, videos, and downloaded packages, reducing storage fees. In addition, ECS can be used with CDN or Server Load Balancer to greatly reduce user access waiting time, reduce bandwidth fees, and improve availability.

Databases

A high-configuration I/O-optimized ECS instance can be used with an SSD cloud disk to support high I/O concurrency with higher data reliability. Alternatively, multiple lower-configuration I/O-optimized ECS instances can be used with Server Load Balancer to deliver a highly available architecture.

Apps or websites with large traffic fluctuations

Some applications may encounter large traffic fluctuations within a short period. When ECS is used with Auto Scaling, the number of ECS instances is automatically adjusted based on traffic. This feature allows you to meet resource requirements while maintaining a low cost. ECS can be used with Server Load Balancer to deliver a high availability architecture.

4 Instance type families

This article introduces the available ECS instance type families.

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

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

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

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

Note:

- If you are using sn1, sn2, t1, s1, s2, s3, m1, m2, c1, c2, c4, cm4, n1, n2, or e3, see *Phased-out instance types*.
- Upgrading instance types is supported within or between certain instance type families. For such families and corresponding upgrade rules, see *Instance type families that support upgrading instance types*.
- Upgrading instance types is not supported within or between the following instance type families: d1, d1ne, i1, i2, ga1, gn5, f1, f2, f3, ebmc4, ebmg5, sccg5, and scch5.

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

- Type families for enterprise-level computing on the x86-architecture:
  - *g5, general-purpose type family*
  - *sn2ne, general-purpose type family with enhanced network performance*

  - *ic5, intensive compute instance type family*
  - *c5, compute instance type family*
  - *sn1ne, compute optimized type family with enhanced network performance*
  - *r5, memory instance type family*
  - *re4, memory optimized type family with enhanced performance*
  - *se1ne, memory optimized type family with enhanced network performance*
  - *se1, memory optimized type family*
  - *d1ne, big data type family with enhanced network performance*
  - *d1, big data type family*
  - *i2, type family with local SSD disks*
  - *i1, type family with local SSD disks*
  - *hfc5, compute optimized type family with high clock speed*
  - *hfg5, general-purpose type family with high clock speed*
  - *ce4, compute optimized type family with high clock speed*
- Type families for enterprise-level heterogeneous computing:
  - *gn6v, compute optimized type family with GPU*
  - *gn5, compute optimized type family with GPU*
  - *gn5i, compute optimized type family with GPU*
  - *gn4, compute optimized type family with GPU*
  - *ga1, visualization compute optimized type family with GPU*
  - *f1, compute optimized type family with FPGA*
  - *f2, compute optimized type family with FPGA*
- ECS Bare Metal Instance type families and Super Computing Cluster (SCC) instance type families:
  - *ebmhfg5, ECS Bare Metal Instance type family with high clock speed*
  - *ebmc4, computing ECS Bare Metal Instance type family*
  - *ebmg5, general-purpose ECS Bare Metal Instance type family*
  - *scch5, Super Computing Cluster (SCC) instance type family with high clock speed*
  - *sccg5, geneneral-purpose Super Computing Cluster (SCC) instance type family*
- Type families for entry-level computing on the x86-architecture:
  - *t5, burstable instances*

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

g5, general-purpose type family

Features

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

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.g5. large	2	8.0	N/A	1.0	300	2	2
ecs.g5. xlarge	4	16.0	N/A	1.5	500	2	3
ecs.g5. 2xlarge	8	32.0	N/A	2.5	800	2	4
ecs.g5. 3xlarge	12	48.0	N/A	4.0	900	4	6

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.g5. 4xlarge	16	64.0	N/A	5.0	1,000	4	8
ecs.g5. 6xlarge	24	96.0	N/A	7.5	1,500	6	8
ecs.g5. 8xlarge	32	128.0	N/A	10.0	2,000	8	8
ecs.g5. 16xlarge	64	256.0	N/A	20.0	4,000	16	8

Click *here* to view other instance type families.

sn2ne, general-purpose type family with enhanced network performance

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Cloud Disks
- vCPU to memory ratio = 1:4
- Ultra high packet forwarding rate
- 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or Platinum 8163 (Skylake) processors
- Higher computing specifications matching higher network performance
- Ideal for:
  - Scenarios where a large volume of packets are received and transmitted, such as the re-transmission of telecommunication services
  - Enterprise-level applications of various types and sizes
  - Medium and small database systems, cache, and search clusters
  - Data analysis and computing
  - Computing clusters and data processing depending on memory

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.sn2ne .large	2	8.0	N/A	1.0	300	2	2
ecs.sn2ne .xlarge	4	16.0	N/A	1.5	500	2	3
ecs.sn2ne .2xlarge	8	32.0	N/A	2.0	1,000	4	4
ecs.sn2ne .3xlarge	12	48.0	N/A	2.5	1,300	4	6
ecs.sn2ne .4xlarge	16	64.0	N/A	3.0	1,600	4	8
ecs.sn2ne .6xlarge	24	96.0	N/A	4.5	2,000	6	8
ecs.sn2ne .8xlarge	32	128.0	N/A	6.0	2,500	8	8
ecs.sn2ne .14xlarge	56	224.0	N/A	10.0	4,500	14	8

Click *here* to view other instance type families.

ic5, intensive compute instance type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Cloud Disks
- vCPU to memory ratio = 1:1
- Ultra high packet forwarding rate
- 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Higher computing specifications matching higher network performance
- Ideal for:
  - Web front-end servers

- Data analysis, batch compute, and video coding
- Scenarios where a large volume of packets are received and transmitted, such as the re-transmission of telecommunication services
- Massively Multiplayer Online (MMO) game front-ends

Instance types

Instance type	Vcpu	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.ic5. large	2	2.0	N/A	1.0	300	2	2
ecs.ic5. xlarge	4	4.0	N/A	1.5	500	2	3
ecs.ic5. 2xlarge	8	8.0	N/A	2.5	800	2	4
ecs.ic5. 3xlarge	12	12.0	N/A	4.0	900	4	6
ecs.ic5. 4xlarge	16	16.0	N/A	5.0	1,000	4	8

Click *here* to view other instance type families.

c5, compute instance type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Cloud Disks
- vCPU to memory ratio = 1:2
- Ultra high packet forwarding rate
- 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Higher computing specifications matching higher network performance
- Ideal for:
  - Scenarios where a large volume of packets are received and transmitted, such as the re-transmission of telecommunication services

- Web front-end servers
- Massively Multiplayer Online (MMO) game front-ends
- Data analysis, batch compute, and video coding
- High performance science and engineering applications

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.c5. large	2	4.0	N/A	1.0	300	2	2
ecs.c5. xlarge	4	8.0	N/A	1.5	500	2	3
ecs.c5. 2xlarge	8	16.0	N/A	2.5	800	2	4
ecs.c5. 3xlarge	12	24.0	N/A	4.0	900	4	6
ecs.c5. 4xlarge	16	32.0	N/A	5.0	1,000	4	8
ecs.c5. 6xlarge	24	48.0	N/A	7.5	1,500	6	8
ecs.c5. 8xlarge	32	64.0	N/A	10.0	2,000	8	8
ecs.c5. 16xlarge	64	128.0	N/A	20.0	4,000	16	8

Click *here* to view other instance type families.

sn1ne, compute optimized type family with enhanced network performance

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Cloud Disks
- vCPU to memory ratio = 1:2

- Ultra high packet forwarding rate
- 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) or Platinum 8163 (Skylake) processors
- Higher computing specifications matching higher network performance
- Ideal for:
  - Scenarios where a large volume of packets are received and transmitted, such as the re-transmission of telecommunication services
  - Web front-end servers
  - Massively Multiplayer Online (MMO) game front-ends
  - Data analysis, batch compute, and video coding
  - High performance science and engineering applications

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.sn1ne .large	2	4.0	N/A	1.0	300	2	2
ecs.sn1ne .xlarge	4	8.0	N/A	1.5	500	2	3
ecs.sn1ne .2xlarge	8	16.0	N/A	2.0	1,000	4	4
ecs.sn1ne .3xlarge	12	24.0	N/A	2.5	1,300	4	6
ecs.sn1ne .4xlarge	16	32.0	N/A	3.0	1,600	4	8
ecs.sn1ne .6xlarge	24	48.0	N/A	4.5	2,000	6	8
ecs.sn1ne .8xlarge	32	64.0	N/A	6.0	2,500	8	8

Click *here* to view other instance type families.

r5, memory instance type family

Features

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

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.r5. large	2	16.0	N/A	1.0	300	2	2
ecs.r5. xlarge	4	32.0	N/A	1.5	500	2	3
ecs.r5. 2xlarge	8	64.0	N/A	2.5	800	2	4
ecs.r5. 3xlarge	12	96.0	N/A	4.0	900	4	6
ecs.r5. 4xlarge	16	128.0	N/A	5.0	1,000	4	8
ecs.r5. 6xlarge	24	192.0	N/A	7.5	1,500	6	8
ecs.r5. 8xlarge	32	256.0	N/A	10.0	2,000	8	8

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.r5. 16xlarge	64	512.0	N/A	20.0	4,000	16	8

Click *here* to view other instance type families.

re4, memory optimized instance type family with enhanced performance

Features

- Supports SSD Cloud Disks and Ultra Cloud Disks
- I/O-optimized
- Optimized for high performance databases, high memory databases, and other memory-intensive enterprise applications
- 2.2 GHz Intel Xeon E7 8880 v4 (Broadwell) processors, up to 2.4 GHz Turbo Boot
- vCPU to memory ratio = 1:12, up to 1920.0 GiB memory
- ecs.re4.20xlarge and ecs.re4.40xlarge have been certified by SAP HANA
- Ideal for:
  - High performance databases and high memory databases (for example, SAP HANA)
  - Memory intensive applications
  - Big Data processing engines, such as Apache spark or Presto

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.re4. 20xlarge	80	960.0	N/A	15.0	2,000	16	8
ecs.re4. 40xlarge	160	1920.0	N/A	30.0	4,500	16	8

Click *here* to view other instance type families.

se1ne, memory optimized type family with enhanced network performance

Features

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

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.se1ne .large	2	16.0	N/A	1.0	300	2	2
ecs.se1ne .xlarge	4	32.0	N/A	1.5	500	2	3
ecs.se1ne .2xlarge	8	64.0	N/A	2.0	1,000	4	4
ecs.se1ne .3xlarge	12	96.0	N/A	2.5	1,300	4	6
ecs.se1ne .4xlarge	16	128.0	N/A	3.0	1,600	4	8

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.se1ne .6xlarge	24	192.0	N/A	4.5	2,000	6	8
ecs.se1ne .8xlarge	32	256.0	N/A	6.0	2,500	8	8
ecs.se1ne .14xlarge	56	480.0	N/A	10.0	4,500	14	8

Click *here* to view other instance type families.

se1, memory optimized type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Cloud Disks
- vCPU to memory ratio = 1:8
- 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Higher computing specifications matching higher network performance
- Ideal for:
  - High performance databases and large memory databases
  - Data analysis and mining, and distributed memory cache
  - Hadoop, Spark, and other enterprise-level applications with large memory requirements

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.se1. large	2	16.0	N/A	0.5	100	1	2

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.se1.xlarge	4	32.0	N/A	0.8	200	1	3
ecs.se1.2xlarge	8	64.0	N/A	1.5	400	1	4
ecs.se1.4xlarge	16	128.0	N/A	3.0	500	2	8
ecs.se1.8xlarge	32	256.0	N/A	6.0	800	3	8
ecs.se1.14xlarge	56	480.0	N/A	10.0	1,200	4	8

Click *here* to view other instance type families.

d1ne, big data type family with enhanced network performance

Features

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

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.d1ne.2xlarge	8	32.0	4 * 5500	6.0	1,000	4	4
ecs.d1ne.4xlarge	16	64.0	8 * 5500	12.0	1,600	4	8
ecs.d1ne.6xlarge	24	96.0	12 * 5500	16.0	2,000	6	8
ecs.d1ne.8xlarge	32	128.0	16 * 5500	20.0	2,500	8	8
ecs.d1ne.14xlarge	56	224.0	28 * 5500	35.0	4,500	14	8

Note:

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

Click *here* to view other instance type families.

d1, big data type family

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Cloud Disks
- High-volume local SATA HDD disks with high I/O throughput and up to 17 Gbit/s of bandwidth for a single instance
- vCPU to memory ratio = 1:4, designed for big data scenarios
- 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Higher computing specifications matching higher network performance
- Ideal for:
  - Hadoop MapReduce, HDFS, Hive, HBase, and so on
  - Spark in-memory computing, MLlib, and so on

- Enterprises that require big data computing and storage analysis, such as those in the Internet and finance industries, to store and compute massive volumes of data
- Elasticsearch, logs, and so on

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.d1. 2xlarge	8	32.0	4 * 5500	3.0	300	1	4
ecs.d1. 3xlarge	12	48.0	16 * 5500	4.0	400	1	6
ecs.d1. 4xlarge	16	64.0	8 * 5500	6.0	600	2	8
ecs.d1. 6xlarge	24	96.0	12 * 5500	8.0	800	2	8
ecs.d1 -c8d3. 8xlarge	32	128.0	12 * 5500	10.0	1,000	4	8
ecs.d1. 8xlarge	32	128.0	16 * 5500	10.0	1,000	4	8
ecs.d1- c14d3. 14xlarge	56	160.0	12 * 5500	17.0	1,800	6	8
ecs.d1. 14xlarge	56	224.0	28 * 5500	17.0	1,800	6	8

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

Click *here* to view other instance type families.

i2, type family with local SSD disks

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Cloud Disks
- High-performance local NVMe SSD disks with high IOPS, high I/O throughput, and low latency.
- vCPU to memory ratio = 1:8, designed for high performance databases
- 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Higher computing specifications matching higher network performance
- Ideal for:
  - OLTP and high performance relational databases
  - NoSQL databases, such as Cassandra and MongoDB
  - Search applications, such as Elasticsearch

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.i2. xlarge	4	32.0	1 * 894	1.0	500	2	3
ecs.i2. 2xlarge	8	64.0	1 * 1788	2.0	1,000	2	4
ecs.i2. 4xlarge	16	128.0	2 * 1788	3.0	1,500	4	8
ecs.i2. 8xlarge	32	256.0	4 * 1788	6.0	2,000	8	8
ecs.i2. 16xlarge	64	512.0	8 * 1788	10.0	4,000	16	8

Click *here* to view other instance type families.

i1, type family with local SSD disks

Features

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

- High-performance local NVMe SSD disks with high IOPS, high I/O throughput, and low latency
- vCPU to memory ratio = 1:4, designed for big data scenarios
- 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Higher computing specifications matching higher network performance
- Ideal for:
  - OLTP and high performance relational databases
  - NoSQL databases, such as Cassandra and MongoDB
  - Search applications, such as Elasticsearch

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.i1.xlarge	4	16.0	2 * 104	0.8	200	1	3
ecs.i1.2xlarge	8	32.0	2 * 208	1.5	400	1	4
ecs.i1.3xlarge	12	48.0	2 * 312	2.0	400	1	6
ecs.i1.4xlarge	16	64.0	2 * 416	3.0	500	2	8
ecs.i1-c5d1.4xlarge	16	64.0	2 * 1456	3.0	400	2	8
ecs.i1.6xlarge	24	96.0	2 * 624	4.5	600	2	8
ecs.i1.8xlarge	32	128.0	2 * 832	6.0	800	3	8
ecs.i1-c10d1.8xlarge	32	128.0	2 * 1456	6.0	800	3	8
ecs.i1.14xlarge	56	224.0	2 * 1456	10.0	1,200	4	8

Click *here* to view other instance type families.

hfc5, compute optimized type family with high clock speed

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Cloud Disks
- Stable performance
- 3.1 GHz Intel Xeon Gold 6149 (Skylake) processors
- vCPU to memory ratio = 1:2
- Higher computing specifications matching higher network performance
- Ideal for:
  - High performance Web front-end servers
  - High performance science and engineering applications
  - Massively Multiplayer Online (MMO) games and video coding

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.hfc5. large	2	4.0	N/A	1.0	300	2	2
ecs.hfc5. xlarge	4	8.0	N/A	1.5	500	2	3
ecs.hfc5. 2xlarge	8	16.0	N/A	2.0	1,000	2	4
ecs.hfc5. 3xlarge	12	24.0	N/A	2.5	1,300	4	6
ecs.hfc5. 4xlarge	16	32.0	N/A	3.0	1,600	4	8
ecs.hfc5. 6xlarge	24	48.0	N/A	4.5	2,000	6	8

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.hfc5. 8xlarge	32	64.0	N/A	6.0	2,500	8	8

Click *here* to view other instance type families.

hfg5, general-purpose type family with high clock speed

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Cloud Disks
- Stable performance
- 3.1 GHz Intel Xeon Gold 6149 (Skylake) processors
- vCPU to memory ratio = 1:4, except for the 56 vCPU instance type
- Higher computing specifications matching higher network performance
- Ideal for:
  - High performance Web front-end servers
  - High performance science and engineering applications
  - Massively Multiplayer Online (MMO) games and video coding

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.hfg5. large	2	8.0	N/A	1.0	300	2	2
ecs.hfg5. xlarge	4	16.0	N/A	1.5	500	2	3

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.hfg5. 2xlarge	8	32.0	N/A	2.0	1,000	2	4.
ecs.hfg5. 3xlarge	12	48.0	N/A	2.5	1,300	4	6
ecs.hfg5. 4xlarge	16	64.0	N/A	3.0	1,600	4	8
ecs.hfg5. 6xlarge	24	96.0	N/A	4.5	2,000	6	8
ecs.hfg5. 8xlarge	32	128.0	N/A	6.0	2,500	8	8
ecs.hfg5. 14xlarge	56	160.0	N/A	10.0	4,000	14	8

Click *here* to view other instance type families.

ce4, compute optimized type family with high clock speed

Features

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

Instance types

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

Click *here* to view other instance type families.

gn6v, compute optimized type family with GPUs

Features

- I/O-optimized
- Supports SSD Cloud Disk and Ultra Cloud Disk
- NVIDIA V100 GPU processors
- vCPU to memory ratio = 1:4
- 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors
- Higher computing specifications matching higher network performance
- Ideal for:
  - Deep learning, autonomous vehicles, voice recognition, and other AI applications
  - Scientific computing, computational finance, genomics, and environmental analysis

Instance types

Instance types	vCPU	Memory (GiB)	Local disks (GiB)*	GPU	Bandwidt (Gbit/s)	Packet forwardin rate (Thousan pps)**	NIC queues***	ENIs****
ecs.gn6v -c8g1. 2xlarge	8	32.0	N/A	1 * NVIDIA V100	2.5	800	4	4

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

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

Click *here* to view other instance type families.

gn5, compute optimized type family with GPU

Features

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

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.gn5-c4g1.xlarge	4	30.0	440	1 * NVIDIA P100	3.0	300	1	3
ecs.gn5-c8g1.2xlarge	8	60.0	440	1 * NVIDIA P100	3.0	400	1	4
ecs.gn5-c4g1.2xlarge	8	60.0	880	2 * NVIDIA P100	5.0	1,000	2	4
ecs.gn5-c8g1.4xlarge	16	120.0	880	2 * NVIDIA P100	5.0	1,000	4	8
ecs.gn5-c28g1.7xlarge	28	112.0	440	1 * NVIDIA P100	5.0	1,000	8	8
ecs.gn5-c8g1.8xlarge	32	240.0	1760	4 * NVIDIA P100	10.0	2,000	8	8
ecs.gn5-c28g1.14xlarge	56	224.0	880	2 * NVIDIA P100	10.0	2,000	14	8
ecs.gn5-c8g1.14xlarge	54	480.0	3520	8 * NVIDIA P100	25.0	4,000	14	8

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

Click *here* to view other instance type families.

gn5i, compute optimized type family with GPU

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Cloud Disks
- NVIDIA P4 GPU processors
- vCPU to memory ratio = 1:4
- 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Higher computing specifications matching higher network performance
- Ideal for:
  - Deep learning
  - Multi-media coding and decoding, and other server-side GPU compute workloads

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.gn5i-c2g1.large	2	8.0	N/A	1 * NVIDIA P4	1.0	100	2	2
ecs.gn5i-c4g1.xlarge	4	16.0	N/A	1 * NVIDIA P4	1.5	200	2	3
ecs.gn5i-c8g1.2xlarge	8	32.0	N/A	1 * NVIDIA P4	2.0	400	4	4
ecs.gn5i-c16g1.4xlarge	16	64.0	N/A	1 * NVIDIA P4	3.0	800	4	8
ecs.gn5i-c16g1.8xlarge	32	128.0	N/A	2 * NVIDIA P4	6.0	1,200	8	8
ecs.gn5i-c28g1.14xlarge	56	224.0	N/A	2 * NVIDIA P4	10.0	2,000	14	8

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

Click *here* to view other instance type families.

gn4, compute optimized type family with GPU

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Cloud Disks
- NVIDIA M40 GPU processors
- No fixed ratio of CPU to memory
- 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Higher computing specifications matching higher network performance
- Ideal for:
  - Deep learning
  - Scientific computing, such as computational fluid dynamics, computational finance, genomics, and environmental analysis
  - High performance computing, rendering, multi-media coding and decoding, and other server-side GPU compute workloads

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.gn4-c4g1.xlarge	4	30.0	N/A	1 * NVIDIA M40	3.0	300	1	3
ecs.gn4-c8g1.2xlarge	8	60.0	N/A	1 * NVIDIA M40	3.0	400	1	4
ecs.gn4.8xlarge	32	48.0	N/A	1 * NVIDIA M40	6.0	800	3	8

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.gn4-c4g1.2xlarge	8	60.0	N/A	2 * NVIDIA M40	5.0	500	1	4
ecs.gn4-c8g1.4xlarge	16	60.0	N/A	2 * NVIDIA M40	5.0	500	1	8
ecs.gn4.14xlarge	56	96.0	N/A	2 * NVIDIA M40	10.0	1,200	4	8

Note:

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

Click *here* to view other instance type families.

ga1, visualization compute type family with GPU

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Cloud Disks
- AMD S7150 GPU processors
- vCPU to memory ratio = 1:2.5
- 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- High performance local NVMe SSD disks
- Higher computing specifications matching higher network performance
- Ideal for:
  - Rendering, multimedia coding and decoding
  - Machine learning, high-performance computing, and high performance databases
  - Other server-end business scenarios that require powerful concurrent floating-point compute capabilities

Instance types

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

Note:

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

Click *here* to view other instance type families.

f1, compute optimized type family with FPGA

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Cloud Disks
- Intel ARRIA 10 GX 1150 FPGA
- vCPU to memory ratio = 1:7.5
- 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Higher computing specifications matching higher network performance
- Ideal for:
  - Deep learning and reasoning
  - Genomics research

- Financial analysis
- Picture transcoding
- Computational workloads, such as real-time video processing and security

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	FPGA	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.f1-c8f1.2xlarge	8	60.0	N/A	Intel ARRIA 10 GX 1150	3.0	400	4	4
ecs.f2-c8f1.4xlarge	16	120.0	N/A	2 * Intel ARRIA 10 GX 1150	5.0	1,000	4	8
ecs.f1-c28f1.7xlarge	28	112.0	N/A	Intel ARRIA 10 GX 1150	5.0	2,000	8	8
ecs.f2-c28f1.14xlarge	56	224.0	N/A	2 * Intel ARRIA 10 GX 1150	10.0	2,000	14	8

Click *here* to view other instance type families.

f2, compute optimized type family with FPGA

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Cloud Disks
- Xilinx Kintex UltraScale XCKU115
- vCPU to memory ratio = 1:7.5
- 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors
- Higher computing specifications matching higher network performance

- Ideal for:
  - Deep learning and reasoning
  - Genomics research
  - Financial analysis
  - Picture transcoding
  - Computational workloads, such as real-time video processing and security

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	FPGA	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.f2-c8f1.2xlarge	8	60.0	N/A	Xilinx Kintex UltraScale XCKU115	2.0	800	4	4
ecs.f2-c8f1.4xlarge	16	120.0	N/A	2 * Xilinx Kintex UltraScale XCKU115	5.0	1,000	4	8
ecs.f2-c28f1.7xlarge	28	112.0	N/A	Xilinx Kintex UltraScale XCKU115	5.0	1,000	8	8
ecs.f2-c28f1.14xlarge	56	224.0	N/A	2 * Xilinx Kintex UltraScale XCKU115	10.0	2,000	14	8

Click *here* to view other instance type families.

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

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Cloud Disks
- vCPU to memory ratio = 1:4
- 3.7 GHz Intel Xeon E3-1240v6 (Skylake) processors, 8-core vCPU, up to 4.1 GHz Turbo Boost
- High network performance: 2 million pps packet forwarding rate
- Supports VPC network only
- Provides dedicated hardware resources and physical isolation
- Supports Intel SGX
- Ideal for:
  - Workloads that require direct access to physical resources, or scenarios where binding a license to the hardware is required
  - Gaming or financial applications featuring high performance
  - High performance Web servers
  - Enterprise-level applications, such as high performance databases

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs. ebmhfg5. 2xlarge	8	32.0	N/A	6.0	2,000	8	6

Note:
For more information about ECS Bare Metal Instance, see *ECS Bare Metal Instance and Super Computing Clusters*.

Click *here* to view other instance type families.

ebmc4, computing ECS Bare Metal Instance type family

Features

- I/O-optimized

- Supports SSD Cloud Disks and Ultra Cloud Disks
- vCPU to memory ratio = 1:2
- 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors, up to 2.9 GHz Turbo Boost
- High network performance: 4 million pps packet forwarding rate
- Supports VPC network only
- Provides dedicated hardware resources and physical isolation
- Ideal for:
  - Scenarios where a large volume of packets are received and transmitted, such as the re-transmission of telecommunication services
  - Third-party virtualization (includes but is not limited to Xen and KVM), and AnyStack (includes but is not limited to OpenStack and ZStack)
  - Containers (includes but is not limited to Docker, Clear Container, and Pouch)
  - Enterprise-level applications, such as medium and large databases
  - Video coding

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs. ebmc4. 8xlarge	32	64.0	N/A	10.0	4,000	8	12

Note:
For more information about ECS Bare Metal Instance, see *ECS Bare Metal Instance and Super Computing Clusters*.

Click *here* to view other instance type families.

ebmg5, general-purpose ECS Bare Metal Instance type family

Features

- I/O-optimized

- Supports SSD Cloud Disks and Ultra Cloud Disks
- vCPU to memory ratio = 1:4
- 2.5 GHz Intel Xeon Platinum 8163 (Skylake) processors, 96-core vCPU, up to 2.7 GHz Turbo Boost
- High network performance: 4.5 million pps packet forwarding rate
- Supports VPC network only
- Provides dedicated hardware resources and physical isolation
- Ideal for:
  - Workloads that require direct access to physical resources, or scenarios where binding a license to the hardware is required
  - Third-party virtualization (includes but is not limited to Xen and KVM), and AnyStack (includes but is not limited to OpenStack and ZStack)
  - Containers (includes but is not limited to Docker, Clear Container, and Pouch)
  - Enterprise-level applications, such as medium and large databases
  - Video encoding

Instance types

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs. ebmg5. 24xlarge	96	384.0	N/A	10.0	4,500	8	32

Note:
For more information about ECS Bare Metal Instance, see *ECS Bare Metal Instance and Super Computing Clusters*.

Click *here* to view other instance type families.

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

Features

- I/O-optimized
- Supports SSD Cloud Disks and Ultra Cloud Disks
- Supports both RoCE and VPC networks, of which RoCE is dedicated to RDMA communication
- With all features of ECS Bare Metal Instance
- 3.1 GHz Intel Xeon Gold 6149 (Skylake) processors
- vCPU to memory ratio = 1:3
- Ideal for:
  - Large-scale machine learning applications
  - Large-scale high-performance scientific and engineering applications
  - Large-scale data analysis, batch computing, video encoding

Instance types

Instance type	vCPU	Memory (GiB)	GPU	Bandwidt (Gbit/s)	Packet forwardin rate (Thousan pps)**	RoCE (Inbound / Outbounc) (Gbit/s)	NIC queues***	ENIs****
ecs. scch5. 16xlarge	64	192.0	N/A	10.0	4,500	46	8	32

Note:
For more information about SCC, see *ECS Bare Metal Instance and Super Computing Clusters*.

Click *here* to view other instance type families.

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

Features

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

- vCPU to memory ratio = 1:4
- Ideal for:
  - Large-scale machine learning applications
  - Large-scale high-performance scientific and engineering applications
  - Large-scale data analysis, batch computing, video encoding

Instance types

Instance type	vCPU	Memory (GiB)	GPU	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	RoCE (Inbound / Outbound) (Gbit/s)	NIC queues***	ENIs****
ecs.sccg5.24xlarge	96	384.0	N/A	10.0	4,500	46	8	32

Note:
For more information about SCC, see *ECS Bare Metal Instance and Super Computing Clusters*.

Click *here* to view other instance type families.

t5, burstable instances

Features

- 2.5 GHz Intel Xeon processors
- The latest DDR4 memory
- No fixed ratio of vCPU to memory
- Baseline CPU performance, burstable, but restricted by accumulated CPU credits
- Resource balance among compute, memory, and networks
- Supports VPC network only
- Ideal for:
  - Web application servers
  - Lightweight web servers
  - Development and testing environments

Instance types

Instance type	vCPU	Memory (GiB)	CPU credits/ hour	Max CPU credit balance	Avg baseline CPU performance	ENIs****
ecs.t5-lc2m1.nano	1	0.5	6	144	10%	1
ecs.t5-lc1m1.small	1	1.0	6	144	10%	1
ecs.t5-lc1m2.small	1	2.0	6	144	10%	1
ecs.t5-lc1m2.large	2	4.0	12	288	10%	1
ecs.t5-lc1m4.large	2	8.0	12	288	10%	1
ecs.t5-c1m1.large	2	2.0	18	432	15%	1
ecs.t5-c1m2.large	2	4.0	18	432	15%	1
ecs.t5-c1m4.large	2	8.0	18	432	15%	1
ecs.t5-c1m1.xlarge	4	4.0	36	864	15%	2
ecs.t5-c1m2.xlarge	4	8.0	36	864	15%	2
ecs.t5-c1m4.xlarge	4	16.0	36	864	15%	2
ecs.t5-c1m1.2xlarge	8	8.0	72	1,728	15%	2
ecs.t5-c1m2.2xlarge	8	16.0	72	1,728	15%	2
ecs.t5-c1m4.2xlarge	8	32.0	72	1,728	15%	2

Instance type	vCPU	Memory (GiB)	CPU credits/ hour	Max CPU credit balance	Avg baseline CPU performance	ENIs****
ecs.t5-c1m1 .4xlarge	16	16.0	144	3,456	15%	2
ecs.t5-c1m2 .4xlarge	16	32.0	144	3,456	15%	2

Note:
For more information about t5, see *Burstable instances*.

Click *here* to view other instance type families.

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

Features

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

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

Type family	Features	vCPU to memory ratio	Ideal for
			building servers, code repositories, microservices, and testing and staging environment • Lightweight enterprise applications
mn4	Balanced entry-level instances	1:4	• Websites and Web applications • Lightweight databases and caches • Integrated applications and lightweight enterprise services
e4	Memory entry-level instances	1:8	• Applications that require large volume of memory • Lightweight databases and cache

xn4

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.xn4. small	1	1.0	N/A	0.5	50	1	1

n4

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.n4. small	1	2.0	N/A	0.5	50	1	1
ecs.n4. large	2	4.0	N/A	0.5	100	1	1
ecs.n4. xlarge	4	8.0	N/A	0.8	150	1	2
ecs.n4. 2xlarge	8	16.0	N/A	1.2	300	1	2
ecs.n4. 4xlarge	16	32.0	N/A	2.5	400	1	2
ecs.n4. 8xlarge	32	64.0	N/A	5.0	500	1	2

mn4

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.mn4. small	1	4.0	N/A	0.5	50	1	1
ecs.mn4. large	2	8.0	N/A	0.5	100	1	1
ecs.mn4. xlarge	4	16.0	N/A	0.8	150	1	2
ecs.mn4. 2xlarge	8	32.0	N/A	1.2	300	1	2
ecs.mn4. 4xlarge	16	64.0	N/A	2.5	400	1	2

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.mn4. 8xlarge	32	128.0	N/A	5	500	2	8

e4

Instance type	vCPU	Memory (GiB)	Local disks (GiB)*	Bandwidth (Gbit/s)	Packet forwarding rate (Thousand pps)**	NIC queues***	ENIs****
ecs.e4. small	1	8.0	N/A	0.5	50	1	1
ecs.ce4. xlarge	2	16.0	N/A	0.5	100	1	1
ecs.ce4. xlarge	4	32.0	N/A	0.8	150	1	2
ecs.e4. 2xlarge	8	64.0	N/A	1.2	300	1	3
ecs.ce4. xlarge	16	128.0	N/A	2.5	400	1	8

Click *here* to view other instance type families.

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

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

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

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

5 Instances

5.1 What are ECS instances

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

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

5.2 ECS instance life cycle

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

Instance status

During this process, an ECS instance may undergo several status changes, as explained in the following table.

Status	Status attribute	Description	Corresponding API status	Viewable in the console
Preparing	Intermediate	After an instance is created, it remains in this status before running. If an instance is in this status for a long time, an exception occurs.	Pending	No
Starting	Intermediate	An instance is in this status when it is either *started* or *restarted* in the console or by using an API before it is running. If an instance is in this status for	Starting	Yes

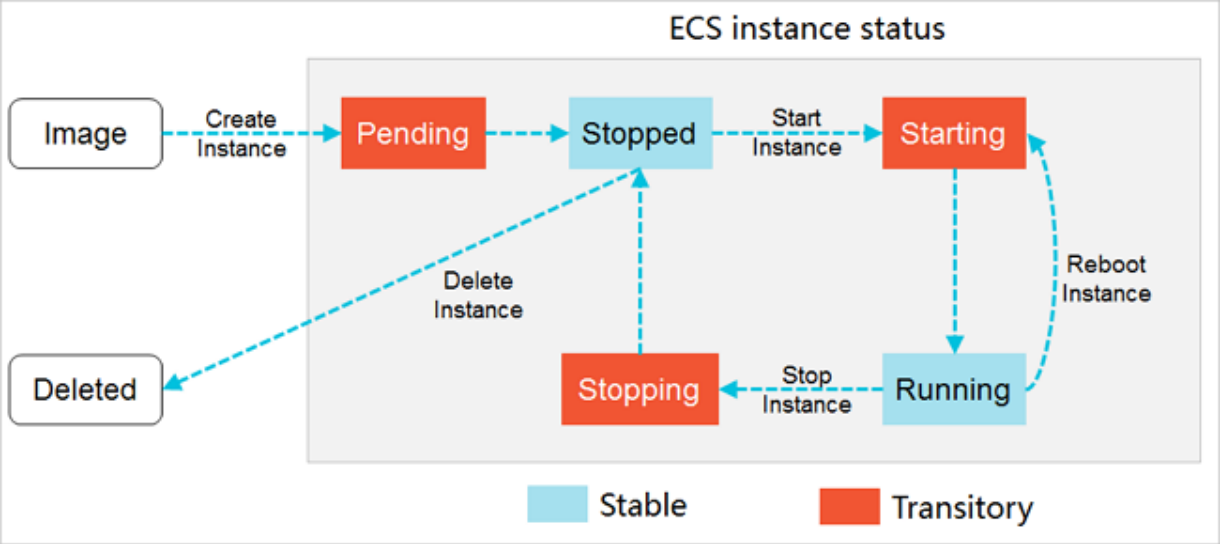
Status	Status attribute	Description	Corresponding API status	Viewable in the console
		a long time, an exception occurs.		
Running	Stable	The instance is operating normally and can accommodate your business needs.	Running	Yes
Stopping	Intermediate	An instance is in this status after the stop operation is performed in the console or when using an API but before the instance actually stops. If an instance is in this status for a long time, an exception occurs.	Stopping	Yes
Stopped	Stable	The instance has been stopped properly. In this status, the instance cannot accommodate external services.	Stopped	Yes
Expired	Stable	A yearly or monthly subscribed instance is in this status if it expires because it has not been timely renewed. A Pay-As-You-Go instance is in this status only when	Stopped	Yes

Status	Status attribute	Description	Corresponding API status	Viewable in the console
		you have an overdue payment. After an ECS instance expires, it continues running for 15 days, and the data on its disks is retained for an additional 15 days, after which the instance will be released and the data will be permanently removed. In this status, the instance cannot accommodate external services.		
Expiring	Stable	A Subscription instance is in this status for 15 days before it expires. After it is *renewed*, the instance is in the **Running**status.	Stopped	Yes
Locked	Stable	An instance is in this status because of an overdue account or security risks. To unlock the instance, *open a ticket*.	Stopped	Yes
Release pending	Stable	A Subscription instance is in this status after you	Stopped	Yes

Status	Status attribute	Description	Corresponding API status	Viewable in the console
		apply for a refund before it expires.		

API status changes

The following figure illustrates API status changes of an instance within its life cycle.

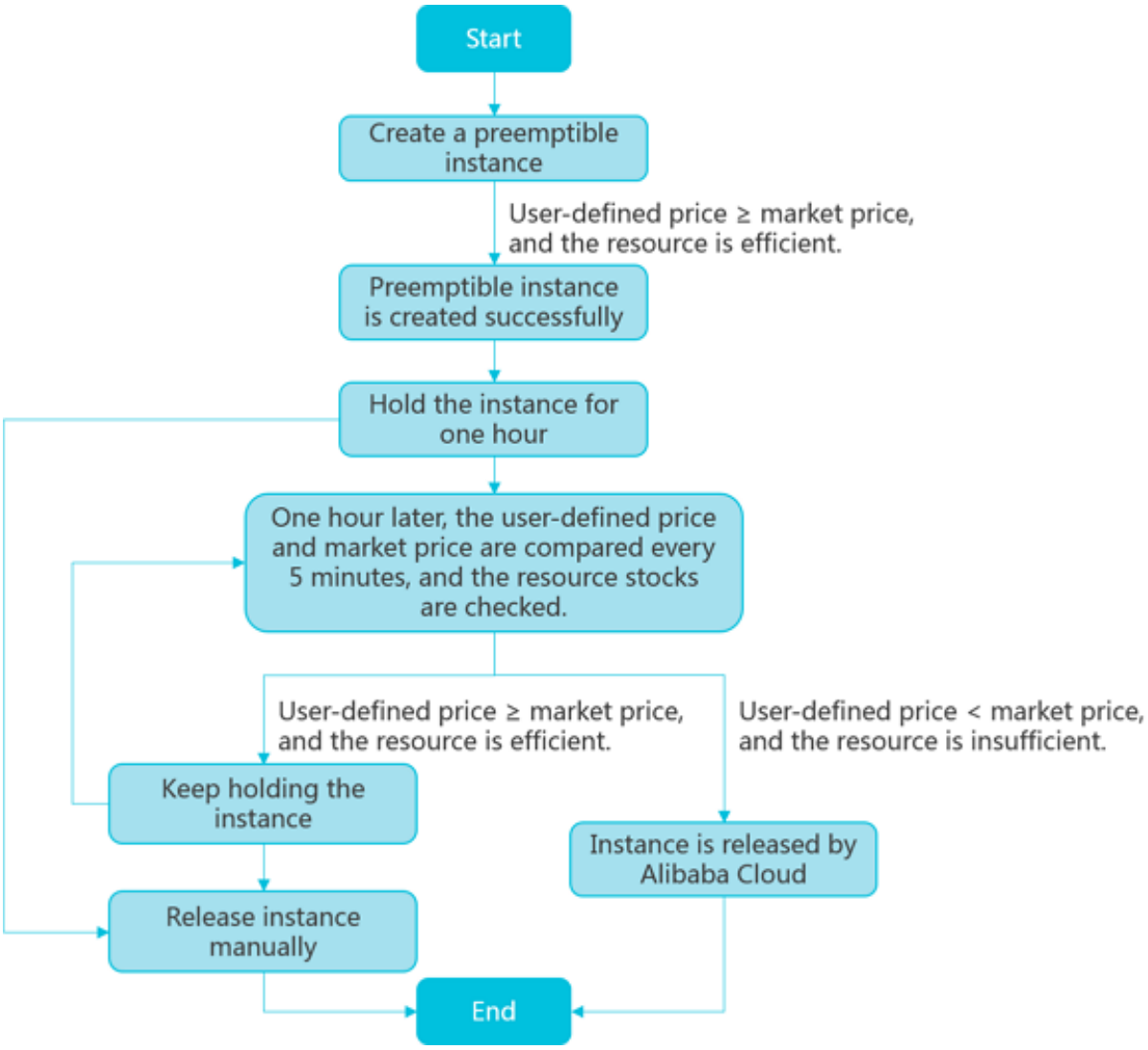


5.3 Preemptible instance

Preemptible instances are on-demand instances. They are designed to reduce your ECS costs in some cases. When you create a preemptible instance, you can set a maximum price per hour to bid for a specified instance type. If your bid is higher than or equal to the current market price, your instance is created. A preemptible instance is held without interruption for at least one hour after it is created. After one hour, your bid is compared with the market price every five minutes. When the market price exceeds your bid or the resource stock is insufficient, the instance is automatically released. The following figure shows the life cycle of a preemptible instance.

Note:

After an instance is released, its data cannot be recovered. We recommend that you *create a snapshot* to back up data before releasing an instance.



Scenarios

Preemptible instances are ideal for stateless applications, such as scalable Web services and applications for rendering figures, big data analysis, and massively parallel computing. Applicatio ns requiring higher level of distribution, scalability, and fault tolerance capability benefit from preemptible instances with respect to costs and throughput.

You can deploy the following businesses on preemptible instances:

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

- Image and media coding
- Testing

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

Bidding modes

You can bid for a preemptible instance only one time in either of the following bidding modes:

- *SpotWithPriceLimit*
- *SpotAsPriceGo*

SpotWithPriceLimit

In this mode, you must set the highest price you want to pay for a specified instance type. When creating a preemptible instance by using *RunInstances*, you can bid in this mode.

Currently, the maximum bid of a preemptible instance is the price of a Pay-As-You-Go instance of the same configuration. When creating a preemptible instance, you can set a price according to the market price history, business features, and the estimated future price fluctuation. When the market price is lower than or equal to your bid, and the resource stock is sufficient, the instance continues to run. If your estimated quote is accurate, you can hold the instance after the one hour *guaranteed duration*. Otherwise, your instance can get automatically released at any time.

SpotAsPriceGo

When creating a preemptible instance by using *RunInstances*, you can create a preemptible instance with the SpotAsPriceGo bidding mode by setting `SpotStrategy` to `SpotAsPriceGo`, which means you always set the real-time market price as the bidding price until the instance is released because of stock shortage.

Guaranteed duration

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

Price and billing

Preemptible instance price and billing considerations:

- **Price**

  The preemptible instance price applies to the instance type only, including vCPU and memory, but not to system disks, data disks, or network bandwidth. The prices for system disks, data disks, or network bandwidth are the same as for *Pay-As-You-Go* instances.

- **Billing cycle**

  Preemptible instances are billed on an hourly basis during their life cycles. You are billed for the entire hour even if your usage is less than an hour.

- **Billing duration**

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

- **Market price**

  During creation of a preemptible instance, it runs when your bid is higher than the current market price and the relevant demand and supply conditions are satisfied. The final price you pay for your instance type is based on the current market price.

The actual market price of a preemptible instance fluctuates according to the changes in the demand and supply of a given instance type, and you can take advantage of these price fluctuatio ns. If you purchase preemptible instance types at the right time, the computing costs are reduced, whereas your throughput is increased for the period the instance is held.

Quota

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

Create a preemptible instance

You can purchase a preemptible instance by using the *RunInstances* interface.

After a preemptible instance is created, it can be used in exactly the same way as a Pay-As-You-Go instance. You can also use it with other cloud products, such as cloud disks or EIP addresses.

Stop a preemptible instance

You can stop a preemptible instance in the *ECS console* or by using the *StopInstance* interface. The VPC-Connected preemptible instances support the *No fees for stopped instances (VPC-Connected)* feature.

The network type and the bidding mode of a preemptible instance determine whether it can start after it is stopped, as displayed in the following table.

Network type + Bidding mode	Stop instance	Start instance
VPC + SpotWithPriceLimit	Keep Instance, Fees Apply	During the guaranteed duration, the instance can be started successfully. After the guaranteed duration: • If your bid is not lower than the market price and the resource stock is sufficient, the instance can be started successfully. • If your bid is lower than the market price or the resource stock is insufficient, the instance cannot be started successfully.
Classic + SpotWithPriceLimit	N/A	
VPC + SpotAsPriceGo	Keep Instance, Fees Apply	During the guaranteed duration, the instance can be started successfully. After the guaranteed duration: • If the resource stock is sufficient, the instance can be started successfully. • If the resource stock is insufficient, the instance cannot be started.
Classic + SpotAsPriceGo	N/A	
VPC + SpotWithPriceLimit	Stop Instance, No Fees	During the guaranteed duration, the instance can be started successfully only if the resource stock is sufficient. After the guaranteed duration:

Network type + Bidding mode	Stop instance	Start instance
		• If your bid is not lower than the market price and the resource stock is sufficient, the instance can be started successfully. • If your bid is lower than the market price or the resource stock is insufficie nt, the instance cannot be started successfully.
VPC + SpotAsPriceGo	Stop Instance, No Fees	During the guaranteed duration, the instance can be started successfully only if the resource stock is sufficient. After the guaranteed duration: • If the resource stock is sufficient, the instance can be started successfully. • If the resource stock is insufficient, the instance cannot be started.

Release a preemptible instance

When the guaranteed period ends, we automatically release your preemptible instance because of changes in the market price or short resource stock. Additionally, you can independently *release the instance*.

When a preemptible instance is released because of market price or changes in the demand and supply of resources, the instance's status changes to **Pending Release**. Then, the instance is released in about five minutes. You can use *instance metadata* or the `OperationLocks` information returned by calling the *DescribeInstances* interface to check if an instance is in the **Pending Release** status.

Note:
Although you can check if a preemptible instance is in the **Pending Release** status by using the API and save a small amount of data while the instance is in this status, we recommend that

you design your applications so work can be properly resumed if the preemptible instance is immediately released. When you release the instance manually, you can test whether or not your application functions normally if a preemptible instance is immediately recovered.

Generally, we release preemptible instances in the order of bidding price, from low to high. If multiple preemptible instances have the same bidding price, they are randomly released.

Best practices

When using a preemptible instance, consider the following:

- Set an appropriate bidding price. In other words, you must quote a competitive price to meet your business budget and hedge against the future market price fluctuations. By using this price, your preemptible instance can be created. In addition, the price must meet your expectations based on your own business assessment.
- The image must have all the software configurations that your applications need, assuring that you can run your business immediately after the instance is created. Additionally, you can use *User-defined data* to run commands at startup.
- Store your business data on storage products that are independent from preemptible instances , such as cloud disks that are not set to release together with instances, OSS, or RDS.
- Split your tasks by using grids, Hadoop, queuing-based architecture, or check points, to facilitate store computing results frequently.
- Use the release notification to monitor the status of a preemptible instance. You can use *metadata* to check the instance status every minute. The metadata of an instance is updated five minutes before it is released automatically.
- Test your applications in advance, to make sure that they can handle events such as accidental release of an instance. To test the applications: Run the applications on a Pay-As-You-Go instance, release the instance, and then check how the applications can handle the release.

For more information, see *FAQ about preemptible instances*.

For more information about using APIs to create preemptible instances, see *Use APIs to manage preemptible instances*.

5.4 ECS Bare Metal Instance and Super Computing Clusters

ECS Bare Metal (EBM) Instance is a new type of computing product that features the elasticity of virtual machines and the performance and characteristics of physical machines. As a product completely and independently developed by Alibaba Cloud, EBM Instance is based on the next

-generation of virtualization technology. Compared with the previous generation, not only is the common virtual cloud server supported, but also is nested virtualization. The resource elasticity of common cloud servers is retained, while nested virtualization technology creates a user experience comparable to physical machines.

Super Computing Clusters (SCCs) are based on EBM Instances. With the help of the high-speed interconnectivity of RDMA (Remote Direct Memory Access) technology, SCCs greatly enhance network performance and increase the acceleration ratio of large-scale clusters. SCCs have all the advantages of EBM Instances and offer high-quality network performance featuring high bandwidth and low latency.

Advantages

EBM Instances

EBM Instances create value for customers through technological innovation. EBM Instances have the following advantages:

- **Exclusive computing resources**

  As a cloud-based elastic computing product, the EBM Instances exceed the performance and isolation of contemporary physical machines and enable exclusive computing resources without virtualization performance overheads and feature loss. EBM Instances support 8, 16, 32, and 96 CPU cores and ultrahigh frequency. An EBM Instance with 8 cores, for example, supports an ultrahigh frequency of up to 3.7 to 4.1 GHz, providing better performance and responsiveness for gaming and financial industries than similar products.

- **Encrypted compute**

  For security, the EBM Instances use a chip-level trusted execution environment (Intel® SGX) in addition to physical server isolation. This allows the instances to compute only the encrypted data in a safe and trusted environment, and provides improved security for the customer data on the cloud. This chip-level hardware security protection provides a safe box for the data of cloud users and allows users to control all the data encryption and key protection procedures.

- **Any Stack on Alibaba Cloud**

  EBM Instances combine the performance strengths and features of physical machines and *the ease-of-use and cost-effectiveness of cloud servers*. They can effectively meet your demands for high-performance computing and help you build new hybrid clouds. Due to their flexibility, elasticity, and other strengths, EBM Instances allow you to deploy any stack on them, such as Xen, KVM, and VMWare. As a result, offline private clouds can be seamlessly migrated

to Alibaba Cloud without the performance overhead issues that may arise because of nested virtualization. This facilitates a new approach for you to move businesses onto the cloud.

- **Heterogeneous instruction set processor support**

  The virtualization 2.0 technology used by EBM Instances is independently developed by Alibaba Cloud. It can zero-cost support ARM and other instruction set processors.

SCC

SCCs are based on EBM Instance, and were released by Alibaba Cloud to meet the demands of applications such as high performance computing, artificial intelligence, machine learning, scientific or engineering computing, data analysis, and audio and video processing. In the clusters, nodes are connected by Remote Direct Memory Access (RDMA) networks featuring high bandwidth and low latency, guaranteeing the highly parallel efficiency demanded by applications that require high-performance computing. Meanwhile, the RoCE (RDMA over Convergent Ethernet) rivals an Infiniband network in terms of connection speed, and supports more extensive Ethernet-based applications. The combination of SCCs built on EBM Instance and other Alibaba Cloud computing products such as ECS and GPU servers provides *the Alibaba Cloud elastic high-performance computing (E-HPC) platform* with ultimate high performance parallel computing resources, making supercomputing on the cloud a reality.

Features

EBM Instances and SCC have the following features:

- CPU specifications:
  - EBM Instances: Supports 8 cores, 16 cores, 32 cores, and 96 cores, and supports high clock speed.
  - SCC: Supports 64 cores and 96 cores, and provides support for high clock speed.
- Memory specifications:
  - EBM Instances: Supports 32 GiB to 768 GiB memory. For better computing performance, the ratio of CPU to memory is 1:2 or 1:4.
  - SCC: The ratio of CPU to memory is 1:3 or 1:4.
- Storage specifications: Supports starting from the virtual machine image and cloud disk to deliver instances in minutes.
- Network configurations:

  - Supports Virtual Private Cloud (VPC) networks, maintaining interoperability with ECS, GPU cloud servers, and other cloud products. Delivers performance and stability comparable to physical machine networks.
  - (Only for SCC) Supports RDMA communication through high-speed RoCE networks.
- Images: Supports images of Alibaba Cloud ECS.
- Security settings: Maintains the same security policies and flexibility as existing cloud server ECS instances.

The following table compares EBM Instance or SCC, physical servers, and virtual servers.

Feature type	Features	EBM Instances/ SCC	Physical servers	Virtual servers
Automated O&M	Delivery in minutes	Y	N	Y
Computing	Zero performance loss	Y	Y	N
	Zero feature loss	Y	Y	N
	Zero resource competition	Y	Y	N
Storage	Fully compatible with ECS cloud disks	Y	N	Y
	Start from cloud disks (system disks)	Y	N	Y
	System disk can be quickly reset	Y	N	Y
	Uses ECS images	Y	N	Y
	Supports cold migration between physical and virtual servers	Y	N	Y
	Requires no installation of operating system	Y	N	Y

Feature type	Features	EBM Instances/ SCC	Physical servers	Virtual servers
	Discards local RAID, and provides stronger protection of data on cloud disks	Y	N	Y
Network	Fully compatible with the ECS VPC networks	Y	N	Y
	Fully compatible with the ECS classic networks	Y	N	Y
	Free of bottlenecks for communications between physical and virtual server clusters in the VPC	Y	N	Y
Management	Fully compatible with the existing ECS management system	Y	N	Y
	Consistent user experience on VNC and other features with that of virtual servers	Y	N	Y
	Guaranteed OOB network security	Y	N	N/A

Instance type families

The type families of EBM Instances include:

- ebmg5, general purpose EBM Instance type family
- ebmhfg5, high frequency EBM Instance type family
- ebmc4, compute EBM Instance type family

The type families of SCC include:

- scch5, Super Computing Cluster (SCC) instance type family with high clock speed
- sccg5, geneneral-purpose Super Computing Cluster (SCC) instance type family

For more information, see *EBM Instance type families* and *SCC Instance type families*.

Billing methods

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

Related operations

You can *create an EBM instance* or *create an SCC server instance* in the console.

For more information, see *FAQs about EBM Instances*.

5.5 Launch templates

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

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

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

Console operations

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

API operations

- *CreateLaunchTemplate*
- *CreateLaunchTemplateVersion*
- *DescribeLaunchTemplates*
- *DescribeLaunchTemplateVersions*
- *ModifyLaunchTemplateDefaultVersion*

- *DeleteLaunchTemplate*
- *DeleteLaunchTemplateVersion*

6 Network and security

6.1 Network types

Alibaba Cloud offers Virtual Private Cloud (VPC) networks and classic networks.

Virtual Private Cloud (VPC)

VPCs are isolated networks established in Alibaba Cloud and logically isolated from each other . You can customize the topology and IP addresses in a VPC. We recommend VPC if you are skilled in network management and have higher network security requirements.

For more information about VPC, see *Virtual Private Cloud* documentation.

Classic network

A classic network is majorly deployed in the public infrastructure of Alibaba Cloud, which is responsible for its planning and management. We recommend classic networks if your business requirements are high in terms of network usability.

Note:
If you did not purchase an ECS instance before 17:00 (UTC+8) on June 14, 2017, you cannot choose the Classic network option.

VPC vs. Classic networks

The following table lists all the functional differences between the VPCs and classic networks.

Items	VPC	Classic network
Two-layer logic isolation	Supported	Not supported
Custom private network blocks	Supported	Not supported
Private IP addresses	Unique within one VPC. Replicable between VPCs.	Unique in the global Classic network
Communicate within or between private networks	Able to communicate within a VPC, but isolated between VPCs	Able to communicate in one region and under one account
Tunneling	Supported	Not supported
Custom router	Supported	Not supported
Routing table	Supported	Not supported
Switches	Supported	Not supported

Items	VPC	Classic network
SDN	Supported	Not supported
Self-built NAT gateway	Supported	Not supported
Self-built VPN	Supported	Not supported

6.2 Intranet

Currently, Alibaba Cloud servers communicate over an intranet. They use a gigabit of shared bandwidth for non I/O optimized instances, and 10 gigabits of shared bandwidth for I/O optimized instances, with no special restrictions. However, because this is a shared network, the bandwidth may fluctuate.

If you have to transmit data between two ECS instances in the same region, use an intranet connection. Intranet connections can also be used to connect any combination of ECS, RDS, SLB , and OSS if they are deployed in the same region. The Internet speed of these products is based on a gigabit of shared bandwidth.

The network types, owners, regions, and security groups affect the intranet communication of ECS instances. See the following table for details.

Network type	Owners	Regions	Security groups	How to enable intranet communication
VPC, same VPC	One account or different accounts	Same	Same	Enabled by default.
			Different	Authorize security groups for each other.
VPC, different VPCs	One account or different accounts	Same	Either the same or different	Use Express Connect. For more information, see *Application scenarios from Product Introduction to Express Connect*.
		Different	Different	
Classic	One account	Same	Same	Enabled by default.

Network type	Owners	Regions	Security groups	How to enable intranet communication
	Different accounts		Either the same or different	Authorize security groups for each other. For more information, see *Scenarios of security groups*.

Private IP addresses are used for intranet communication. You cannot *change the private IP address* of an instance of the Classic network type, but you can change the private IP address of a VPC-Connected ECS instance. Private and public addresses of ECS instances do not support virtual IP (VIP) configuration.

By default, instances of different network types cannot communicate with one another in one intranet. VPC provides the *ClassicLink* function, which allows you to link an ECS instance in the classic network to cloud resources in a VPC through the intranet.

6.3 IP addresses of a classic network-connected ECS instance

IP addresses are mainly used for remote access to your instance or to the services deployed on your instance. Currently, for ECS instances of the classic network type, IP addresses are distributed in a unified way and divided into public and private IP addresses.

Intranet IP addresses

Each classic network-connected ECS instance is assigned an intranet IP address.

Scenarios

Intranet IP addresses can be used in the following scenarios:

- Load balancing
- Mutual intranet access between ECS instances
- Mutual intranet access between ECS instances and other cloud services, such as OSS and RDS

Communication traffic through intranet IP addresses within an intranet is free of charge. For more information, see *Intranet*.

Modify an intranet IP address

Once a classic network-connected ECS instance is created, you cannot change its intranet IP address.

> **Note:**
> Do not change an intranet IP address within a guest operating system. Otherwise, communication within an intranet is interrupted.

Public IP addresses

If you purchase bandwidth for Internet access, a public IP address is assigned to your classic network-connected ECS instance. You cannot change the public IP address once it is assigned.

Scenarios

A public IP address is used in the following scenarios:

- Mutual access between an ECS instance and the Internet
- Mutual Internet access between ECS instances and other Alibaba Cloud services

Assign a public IP address

When you create an ECS instance, a public IP address is assigned to it if **Assign public IP** is selected.

For a Subscription instance with no public IP address, you can use the *Upgrade Configuration* or the *Renew for Configuration Downgrade* feature to purchase public network bandwidth.

> **Note:**
> - For a Pay-As-You-Go classic network-connected ECS instance with no public IP address, you cannot assign a public IP address after the instance is created.
> - For a classic network-connected ECS instance, you cannot unbind or release its public IP address once the IP address is assigned. If you set the bandwidth to 0 Mbit/s when renewing an instance for configuration downgrade, in the next purchase cycle, the public IP address is retained, but the instance cannot access the Internet.

Billing

You are billed for usage of Internet outbound traffic only. For more information, see *Billing of network bandwidth*.

Multicast and broadcast

Intranet IP addresses cannot be used for multicast or broadcast.

6.4 IP addresses of VPC-Connected ECS instances

Each VPC-Connected ECS instance can communicate within an intranet by using a private IP address or over the Internet by using a public IP address.

Private IP addresses

Each VPC-Connected ECS instance is assigned a private IP address when it is created. That address is determined by the VPC and the CIDR block of the VSwitch to which the instance is connected.

Scenarios

A private IP address can be used in the following scenarios:

- Load balancing
- Communication among ECS instances within an intranet
- Communication between an ECS instance and other cloud products (such as OSS and RDS) within an intranet

For more information, see *Intranet*.

Modify a private IP address

To meet your business needs, you can modify the private IP address of a VPC-Connected ECS instance in the ECS console. For more information, see *Change the private IP of an ECS instance*.

Public IP addresses

VPC-Connected ECS instances support two public IP address types:

- NatPublicIp, which is assigned to a VPC-Connected ECS instance, can be released only, and cannot be unbound from the instance.
- Elastic public IP (EIP). For more information, see *What is an EIP address*.

When a VPC-Connected ECS instance accesses the Internet, its public IP address is mapped to its private IP address through network address translation (NAT).

You cannot find a network interface for Internet access by running commands within the operating system.

Scenarios

NatPublicIp and EIP are applicable to different scenarios:

- NatPublicIp: If you want to assign a public IP address to a VPC-Connected ECS instance when creating the instance and do not want to retain the public IP address when the instance is released, you can use a NatPublicIp address.
- EIP: If you want to keep a public IP address and bind it to any of your VPC-Connected ECS instances in the same region, you can use an EIP address.

Obtain a public IP address

- NatPublicIp: When creating a VPC-Connected ECS instance, if you select **Assign a public IP**, a NatPublicIp is assigned to the instance when it is created.
- EIP: You can apply for an EIP address and bind it to a VPC-Connected ECS instance. In this case, do not assign a NatPublicIp to an instance. For more information, see *Apply for an EIP address*.

Release a public IP address

- NatPublicIp: When a NatPublicIp address is assigned to an instance, you can only release the IP address, but cannot unbind it. Only a NatPublicIp address that is assigned to a Subscription instance can be released. For more information, see *Renew for configuration downgrade*.
- EIP: If you do not need an EIP address, unbind it from a VPC-Connected ECS instance and release it in the EIP console. For more information, see *Unbind and release an EIP address*.

Billing

You are billed for outbound Internet traffic usage only. For more information, see *Billing of network bandwidth*.

6.5 Multi-queue for NICs

A single CPU is not sufficient for handling network interruptions. Therefore, you should route NIC interruptions in the ECS instances to different CPUs. Results of network PPS and bandwidth tests show that a solution that uses two queues instead of one queue can enhance network performance by 50% to 100%. A solution that uses four queues can bring further significant increases in network performance.

ECS instance types supporting multi-queue

See *Instance type families* to find instance types supporting multi-queue and the number of queues that are supported.

Images supporting multi-queue

The following public images officially provided by Alibaba Cloud support multi-queue:

> **Note:**
> Whether an image supports multi-queue is not related to the memory address width of the operating system.

- CentOS 6.8/6.9/7.2/7.3/7.4
- Ubuntu 14.04/16.04
- Debian 8.9
- SUSE Linux Enterprise Server 12 SP1
- Windows 2012 R2 and Windows 2016: You may be invited to test this feature in the future.

The SUSE Linux Enterprise Server 12 SP2 edition will be available soon.

Configure multi-queue support for NICs on a Linux ECS instance

We recommend that you use one of the latest Linux distributions, such as CentOS 7.2, to configure multi-queue for the NICs.

Here we take CentOS 7.2 as an example to illustrate how to configure multi-queue for the NIC. In this example, we want to configure two queues, and the NIC name is eth0.

- To check whether the NIC supports multi-queue, run the command: `ethtool -l eth0`.
- To enable multi-queue for the NIC, run the command: `ethtool -L eth0 combined 2`.
- If you are using more than one NIC, configure each NIC.

```
    [root@localhost ~]# ethtool -l eth0
    Channel parameters for eth0:
    Pre-set maximums:
    RX: 0
    TX: 0
    Other: 0
    Combined: 2 # This line indicates that a maximum of two queues
can be configured
    Current hardware settings:
    RX: 0
    TX: 0
    Other: 0
    Combined: 1 #It indicates that one queue is currently taking
effect
```

```
    [root@localhost ~]# ethtool -L eth0 combined 2 # It sets eth0 to
use two queues currently
```

- We recommend that you enable the irqbalance service so that the system can automatically adjust the allocation of the NIC interrupts on multiple CPU cores. Run the command: `systemctl start irqbalance` (this feature is enabled by default in CentOS 7.2).
- If the network performance is not improved as expected after the multi-queue feature is enabled, you can enable the RPS feature. See the following Shell script.

```
#!/bin/bash
cpu_num=$(grep -c processor /proc/cpuinfo)
quotient=$((cpu_num/8))
if [ $quotient -gt 2 ]; then
    quotient=2
elif [ $quotient -lt 1 ]; then
    quotient=1
fi
for i in $(seq $quotient)
do
    cpuset="${cpuset}f"
done
for rps_file in $(ls /sys/class/net/eth*/queues/rx-*/rps_cpus)
do
    echo $cpuset > $rps_file
done
```

Configure multi-queue support for NICs on a Windows ECS instance

> **Note:**
> We are inviting Windows users to test the performance improvement. Windows systems see improved network performance after using multi-queue for NICs, but the improvement is not as much as for Linux systems.

If you are using a Windows instance, you must install the driver to use the multi-queue feature for NICs.

To install the driver for Windows systems, follow these steps:

1. *Open a ticket* to request and download the driver installation package.
2. Unzip the driver installation package. For Windows 2012/2016 systems, use the driver in the Win8/amd64 folder.
3. Upgrade the NIC driver:
   a. Select **Device Manager** > **Network adapters**.
   b. Right click **Red Hat VirtIO Ethernet Adapter** and select **Update Driver**.

c. Select the Win8/admin64 directory of the driver directory that you have unzipped, and update the driver.

4. Recommended: Restart the Windows system after the driver is upgraded.

The multi-queue feature for NICs is now ready to use.

6.6 Elastic network interfaces

An Elastic Network Interface (ENI) is a virtual network interface that can be attached to an ECS instance in a VPC. By using ENIs, you can build high-availability clusters, implement failover at a lower cost, and achieve refined network management. The ENI feature is available in all regions.

Scenarios

ENIs can be used in the following scenarios:

- **Deploying a high-availability cluster**

  An ENI can meet the demands of a high-availability architecture for multiple network interfaces on a single instance.

- **Providing a low-cost failover solution**

  You can detach an ENI from a failed ECS instance and then attach it to another ECS instance to quickly redirect the failed instance's traffic to a backup instance. This action recovers the service immediately.

- **Managing the network with refined controls**

  You can configure multiple ENIs for an instance. For example, you can use some ENIs for internal management and other ENIs for Internet business access, so as to isolate managerial data from business data. You can also configure precisely-targeted security group rules for each ENI based on the source IP address, protocols, ports, and more to achieve secured traffic control.

ENI types

ENIs are classified into two types:

- **Primary ENI**

  The ENI created by default upon the creation of an instance in a VPC is called the **primary ENI**. The lifecycle of the primary ENI is the same as that of the instance and you are not allowed to remove the primary ENI from the instance.

- **Secondary ENI**

You can create a secondary ENI and attach it to an instance or detach it from the instance. Multiple private IPs are supported for one secondary ENI. The maximum number of ENIs that you can attach to one instance varies with the instance type. For more information, see *Instance type families*.

ENI attributes

The following table displays ENI attributes.

Attribute	Quantity
Primary private IP addresses	1
Secondary private IP addresses	Depends on the instance type.
MAC address	1
Security group	Min. 1, and Max. 5
Description	1
ENI name	1

Limits

ENIs have the following limits:

- By default, one account can own up to 100 ENIs in one region. The quota increases with the membership level. If you require a higher quota, *open a ticket*.
- The ECS instance must be in the same zone of the same region as the ENI, but they do not have to be in the same VSwitch.
- The number of ENIs that can be attached to an ECS instance is determined by the instance type. For more information, see *Instance type families*.
- Only I/O optimized instance types support ENIs.
- You cannot increase the instance bandwidth capability by attaching multiple ENIs.

Note:

The instance bandwidth capability varies with the instance type.

Related operations

For images that cannot identify ENIs, you can log on to the instance to *configure the ENI*.

Console operations

You can complete the following operations in the ECS console:

- *Attach an ENI when creating an instance*
- *Create an ENI*
- *Delete an ENI*
- *Attach an ENI to an instance*The instance must be in a **Stopped** or **Running** status.
- *Detach an ENI from an instance*The instance must be in a **Stopped** or **Running** status.
- *Modify attributes of an ENI*You can modify attributes of an ENI, including its name, security group, and description.
- When an ENI is attached to an instance, you can view the information of the ENI on the instance details page and the network interfaces page.

API operations

You can complete the following operations by using APIs:

- *Create an ENI*
- *Delete an ENI*
- *Query ENI list*
- *Attach an ENI to an instance*: The instance must be in a **Stopped** or **Running** status.
- *Detach an ENI from an instance*: The instance must be in a **Stopped** or **Running** status.
- *Modify attributes of an ENI*: You can modify attributes of an ENI, including its name, its security group, and its description.
- You can use the DescribeInstances *DescribeInstances* interface to query the information of an ENI when the ENI is attached to an instance.

6.7 Security group

A security group is a virtual firewall that provides Stateful Packet Inspection (SPI). Security groups are used to set network access control for one or more ECS instances. As an important means of security isolation, security groups are used to divide security domains on the cloud.

A security group is a logical group that contains instances in the same region with the same security requirements and mutual trust. Each instance belongs to at least one security group , which must be specified at the time of creation. Instances in the same security group can communicate through the intranet network, but instances in different security groups cannot communicate by default. However, mutual access between two security groups can be authorized.

Security group restrictions

- There is a maximum limit for the number of security groups you can have for a region. The limit depends on your level of experience with Alibaba Cloud. For new users, the limit is 100 security groups. For more experienced users, the limit is higher. To raise the upper limit, you can *open a ticket*.
- Each Elastic Network Interface (ENI) of an instance can join to up to five security groups by default. You can *open a ticket* to raise the upper limit to a maximum of 16.
- Security groups have two network types: classic network and Virtual Private Cloud (VPC).
  - Classic network instances can join security groups on classic networks in the same region.

    A single security group on a classic network cannot contain more than 1,000 instances. If you require mutual intranet access between more than 1,000 instances, you can allocate them to different security groups and authorize mutual access.
  - VPC instances can join security groups on the same VPC.

    A single security group on a VPC cannot contain more than 2,000 private IP addresses (shared by the primary and secondary ENIs). If you require mutual intranet access between more than 2,000 private IP addresses, you can allocate the relevant instances to different security groups and authorize mutual access.
- Adjusting security groups will not affect the continuity of user service.
- Security groups are stateful. If an outbound packet is permitted, inbound packets correspond ing to this connection will also be permitted.

For more information, see *Security groups*.

Security group rules

Security group rules can be set that permit or forbid ECS instances in a security group from accessing a public network or intranet in the inbound and outbound directions.

You can create or delete security group rules at any time. Once changes are made, the updated security group rules are automatically applied to ECS instances in the security group.

When setting security group rules, make sure they are concise. If you add an ECS instance to multiple security groups, hundreds of rules may apply to the instance, which may cause connection errors when you access the instance.

Security group rule restrictions :

- Each security group can have a maximum of 100 security group rules in total, including both inbound and outbound rules.
- Each ENI of an instance can have a maximum of 500 security group rules.

6.8 SSH key pairs

What is an SSH key pair?

An SSH key pair, or key pair for short, is a secure authentication method offered by Alibaba Cloud for remote log-on to your Linux instance. It is an alternative to authentication using a username and password.

The cryptography feature uses the**public key** to encrypt data, and then the local client uses the **private key** to decrypt the data. Together, the public and private keys are known as a key pair.

The Linux ECS instance stores the public key. You use the private key to connect to your instance by entering SSH commands or using other tools, and you no longer need to remember a username and password to log on. Username and password authentication is disabled by ECS once the SSH key pair is enabled to guarantee security.

Benefits

Compared with typical username and password authentication, SSH key pair has the following benefits:

High security

Using an SSH key pair to log on to a Linux instance is more secure and reliable.

- A key pair prevents brute force password-cracking attacks.
- It is impossible to deduce the private key even if the public key is maliciously acquired.

Ease of use

- You can remotely log on to the instance by configuring the key pair in the ECS console and on the local client. You do not have to enter the password every time you log on.
- We recommend this method if you maintain multiple ECS instances.

Limits

Using an SSH key pair has the following restrictions:

- Applies only to Linux instances.
- Alibaba Cloud only supports the creation of 2048-bit RSA key pairs.

  - Alibaba Cloud holds the public key of the key pair.
  - After the key pair is created, you must download and keep the private key for further use.
  - The private key is in the unencrypted PEM-encoded PKCS#8 format.
- An Alibaba Cloud account can have a maximum of 500 key pairs in a region.
- A Linux instance can be only bound to one SSH key pair. If a key pair has already been bound to your instance, the new key pair replaces the old one.
- During the lifecycle of a Linux instance, you can bind or unbind an SSH key pair at any time. After you bind or unbind a key pair, you must *restart the instance* for the change to take effect.
- All instances of any *instance type family*, except for the I/O optimized instances of Generation I, support SSH key pairs.

Create an SSH key pair

To create an SSH key pair, you can use either of the following methods:

- *Create an SSH key pair* in the ECS console.

  Note:
  Once you create a key pair in the ECS console, you must immediately download and keep the private key for further use. If SSH key pair authentication is enabled for an ECS instance, you cannot log on to the ECS instance without the private key of the key pair.
- Create an SSH key pair by using other key pair builders and *import it* to ECS.

  The following key types are supported:

  - rsa
  - dsa
  - ssh-rsa
  - ssh-dss
  - ecdsa
  - ssh-rsa-cert-v00@openssh.com
  - ssh-dss-cert-v00@openssh.com
  - ssh-rsa-cert-v01@openssh.com
  - ssh-dss-cert-v01@openssh.com
  - ecdsa-sha2-nistp256-cert-v01@openssh.com
  - ecdsa-sha2-nistp384-cert-v01@openssh.com

- ecdsa-sha2-nistp521-cert-v01@openssh.com

If your key pair is generated by Alibaba Cloud, you must download the private key and keep it safe . When a key pair is bound to an ECS instance, you cannot log on to that ECS instance if you do not have the private key.

Related operations

- If you do not have an SSH key pair, you can *create an SSH key pair*.
- If you have created an SSH key pair by using another tool, you can *import an SSH key pair*.
- If you do not need a key pair, you can *delete an SSH key pair*.
- If you want to enable or disable SSH key pair authentication for logging on to a Linux ECS instance, you can *bind or unbind an SSH key pair*.
- You can allocate an SSH key pair when *creating an ECS instance*.
- You can *log on to an instance by using an SSH key pair*.

6.9 Anti-DDoS Basic

Anti-DDoS Basic is a free Distributed Denial of Service (DDoS) protection service that safeguards data and applications on your ECS instance. As a global service from Alibaba Cloud Security, Anti-DDoS Basic offers a mitigation capacity of 5 Gbit/s against common DDoS attacks. When the inbound traffic of an ECS instance exceeds its limits, which is determined by the ECS instance type, Alibaba Cloud Security enables throttling to maintain stable performance.

How Anti-DDoS Basic works

When Anti-DDoS Basic is enabled, Alibaba Cloud Security monitors the inbound traffic in real time. When massive or abnormal traffic involving DDoS attacks is monitored, Alibaba Cloud Security redirects the traffic, drops malicious traffic, and passes clean traffic back to the ECS instance. This process is called **flow cleaning**. For more information, see *How Anti-DDoS Basic works*.

Note:

If Anti-DDoS Basic is enabled for an ECS instance, when the inbound traffic from Internet is higher than 5 Gbit/s, to secure the global cluster, Alibaba Cloud Security triggers a black hole to receive such traffic. For more information, see *Alibaba Cloud black hole policies*.

Flow cleaning is triggered in the following situations:

- When specified attacks are identified in the inbound traffic.

- When the inbound traffic to an ECS instance exceeds the specified threshold.

Methods to clean the flow include filtering ICMP packets, limiting bit rate, and limiting the packet forwarding rate.

Therefore, when using Anti-DDoS Basic, you must set the following thresholds:

- BPS threshold: When the inbound traffic exceeds the BPS threshold, flow cleaning is triggered.
- PPS threshold: When the inbound packet forwarding rate exceeds the PPS threshold, flow cleaning is triggered.

Cleaning thresholds of each instance type

The configuration of each instance type determines its maximum flow cleaning threshold. The following table lists the cleaning thresholds of some *available* and *phased-out* instance types.

Instance type	Maximum BPS threshold (Mbit/s)	Maximum PPS threshold (PPS)
ecs.g5.16xlarge	20,000	4,000,000
ecs.g5.22xlarge	30,000	4,500,000
ecs.g5.2xlarge	2,500	800,000
ecs.g5.4xlarge	5,000	1,000,000
ecs.g5.6xlarge	7,500	1,500,000
ecs.g5.8xlarge	10,000	2,000,000
ecs.g5.large	1,000	300,000
ecs.g5.xlarge	1,500	500,000
ecs.sn2ne. 14xlarge	10,000	4,500,000
ecs.sn2ne. 2xlarge	2,000	1,000,000
ecs.sn2ne. 4xlarge	3,000	1,600,000
ecs.sn2ne. 8xlarge	6,000	2,500,000
ecs.sn2ne.large	1,000	300,000
ecs.sn2ne.xlarge	1,500	500,000
ecs.c5.16xlarge	20,000	4,000,000
ecs.c5.2xlarge	2,500	800,000
ecs.c5.4xlarge	5,000	1,000,000
ecs.c5.6xlarge	7,500	1,500,000

Instance type	Maximum BPS threshold (Mbit/s)	Maximum PPS threshold (PPS)
ecs.c5.8xlarge	10,000	2,000,000
ecs.c5.large	1,000	300,000
ecs.c5.xlarge	1,500	500,000
ecs.sn1ne. 2xlarge	2,000	1,000,000
ecs.sn1ne. 4xlarge	3,000	1,600,000
ecs.sn1ne. 8xlarge	6,000	2,500,000
ecs.sn1ne.large	1,000	300,000
ecs.sn1ne.xlarge	1,500	500,000
ecs.r5.16xlarge	20,000	4,000,000
ecs.r5.22xlarge	30,000	4,500,000
ecs.r5.2xlarge	2,500	800,000
ecs.r5.4xlarge	5,000	1,000,000
ecs.r5.6xlarge	7,500	1,500,000
ecs.r5.8xlarge	10,000	2,000,000
ecs.r5.large	1,000	300,000
ecs.r5.xlarge	1,500	500,000
ecs.re4.20xlarge	15,000	2,000,000
ecs.re4.40xlarge	30,000	4,000,000
ecs.se1ne. 14xlarge	10,000	4,500,000
ecs.se1ne. 2xlarge	2,000	1,000,000
ecs.se1ne. 4xlarge	3,000	1,600,000
ecs.se1ne. 8xlarge	6,000	2,500,000
ecs.se1ne.large	1,000	300,000
ecs.se1ne.xlarge	1,500	500,000
ecs.se1.14xlarge	10,000	1,200,000
ecs.se1.2xlarge	1,500	400,000
ecs.se1.4xlarge	3,000	500,000
ecs.se1.8xlarge	6,000	800,000

Instance type	Maximum BPS threshold (Mbit/s)	Maximum PPS threshold (PPS)
ecs.se1.large	500	100,000
ecs.d1ne. 2xlarge	6,000	1,000,000
ecs.d1ne. 4xlarge	12,000	1,600,000
ecs.d1ne. 6xlarge	16,000	2,000,000
ecs.d1ne. 8xlarge	20,000	2,500,000
ecs.d1ne. 14 x large	35,000	4,500,000
ecs.d1.2xlarge	3,000	300,000
ecs.d1.4xlarge	6,000	600,000
ecs.d1.6xlarge	8,000	800,000
ecs.d1.8xlarge	10,000	1,000,000
ecs.d1-c8d3.8xlarge	10,000	1,000,000
ecs.d1.14xlarge	17,000	1,800,000
ecs.d1-c14d3.14xlarge	17,000	1,400,000
ecs.i2.xlarge	1,000	500,000
ecs.i2.2xlarge	2,000	1,000,000
ecs.i2.4xlarge	3,000	1,500,000
ecs.i2.8xlarge	6,000	2,000,000
ecs.i2.16xlarge	10,000	4,000,000
ecs.i1.xlarge	800	200,000
ecs.i1.2xlarge	1,500	400,000
ecs.i1.4xlarge	3,000	500,000
ecs.i1-c10d1.8xlarge	6,000	800,000
ecs.i1-c5d1.4xlarge	3,000	400,000
ecs.i1.14xlarge	10,000	1,200,000
ecs.hfc5.large	1,000	300,000
ecs.hfc5.xlarge	1,500	500,000
ecs.hfc5.2xlarge	2,000	1,000,000
ecs.hfc5.4xlarge	3,000	1,600,000

Instance type	Maximum BPS threshold (Mbit/s)	Maximum PPS threshold (PPS)
ecs.hfc5.6xlarge	4,500	2,000,000
ecs.hfc5.8xlarge	6,000	2,500,000
ecs.hfg5.large	1,000	300,000
ecs.hfg5.xlarge	1,500	500,000
ecs.hfg5.2xlarge	2,000	1,000,000
ecs.hfg5.4xlarge	3,000	1,600,000
ecs.hfg5.6xlarge	4,500	2,000,000
ecs.hfg5.8xlarge	6,000	2,500,000
ecs.hfg5.14xlarge	10,000	4,000,000
ecs.c4.2xlarge	3,000	400,000
ecs.c4.4xlarge	6,000	800,000
ecs.c4.xlarge	1,500	200,000
ecs.ce4.xlarge	1,500	200,000
ecs.cm4.4xlarge	6,000	800,000
ecs.cm4.6xlarge	10,000	1,200,000
ecs.cm4.xlarge	1,500	200,000
ecs.gn5-c28g1.14xlarge	10,000	4,500,000
ecs.gn5-c4g1.xlarge	3,000	300,000
ecs.gn5-c4g1.2xlarge	5,000	1,000,000
ecs.gn5-c8g1.2xlarge	3,000	400,000
ecs.gn5-c8g1.4xlarge	5,000	1,000,000
ecs.gn5-c28g1.7xlarge	5,000	2,250,000
ecs.gn5-c8g1.8xlarge	10,000	2,000,000
ecs.gn5-c8g1.14xlarge	25,000	4,000,000
ecs.gn5i-c2g1.large	1,000	100,000
ecs.gn5i-c4g1.xlarge	1,500	200,000
ecs.gn5i-c8g1.2xlarge	2,000	400,000
ecs.gn5i-c16g1.4xlarge	3,000	800,000

Instance type	Maximum BPS threshold (Mbit/s)	Maximum PPS threshold (PPS)
ecs.gn5i-c28g1.14xlarge	10,000	2,000,000
ecs.gn4-c4g1.xlarge	3,000	300,000
ecs.gn4-c8g1.2xlarge	3,000	400,000
ecs.gn4-c4g1.2xlarge	5,000	500,000
ecs.gn4-c8g1.4xlarge	5,000	500,000
ecs.gn4.8xlarge	6,000	800,000
ecs.gn4.14xlarge	10,000	1,200,000
ecs.ga1.xlarge	1,000	200,000
ecs.ga1.2xlarge	1,500	300,000
ecs.ga1.4xlarge	3,000	500,000
ecs.ga1.8xlarge	6,000	800,000
ecs.ga1.14xlarge	10,000	1,200,000
ecs.f1-c28f1.7xlarge	5,000	2,000,000
ecs.f1-c8f1.2xlarge	2,000	800,000
ecs.f2-c28f1.14xlarge	10,000	2,000,000
ecs.f2-c28f1.7xlarge	5,000	1,000,000
ecs.f2-c8f1.2xlarge	2,000	400,000
ecs.f2-c8f1.4xlarge	5,000	1,000,000
ecs.t5-c1m1.2xlarge	1,200	400,000
ecs.t5-c1m1.large	500	100,000
ecs.t5-c1m1.xlarge	800	200,000
ecs.t5-c1m2.2xlarge	1,200	400,000
ecs.t5-c1m2.large	500	100,000
ecs.t5-c1m2.xlarge	800	200,000
ecs.t5-c1m4.2xlarge	1,200	400,000
ecs.t5-c1m4.large	500	100,000
ecs.t5-c1m4.xlarge	800	200,000
ecs.t5-lc1m1.small	200	60,000

Instance type	Maximum BPS threshold (Mbit/s)	Maximum PPS threshold (PPS)
ecs.t5-lc1m2.large	400	100,000
ecs.t5-lc1m2.small	200	60,000
ecs.t5-lc1m4.large	400	100,000
ecs.t5-lc2m1.nano	100	40,000
ecs.ebmg4.8xlarge	10,000	4,500,000
ecs.ebmg5.24xlarge	10,000	4,500,000
ecs.sccg5.24xlarge	10,000	4,500,000
ecs.xn4.small	500	50,000
ecs.mn4.small	500	50,000
ecs.mn4.large	500	100,000
ecs.mn4.xlarge	800	150,000
ecs.mn4.2xlarge	1,200	300,000
ecs.mn4.4xlarge	2,500	400,000
ecs.n4.small	500	50,000
ecs.n4.large	500	100,000
ecs.n4.xlarge	800	150,000
ecs.n4.2xlarge	1,200	300,000
ecs.n4.4xlarge	2,500	400,000
ecs.n4.8xlarge	5,000	500,000
ecs.e4.small	500	50,000
ecs.sn1.medium	500	100,000
ecs.sn1.large	800	200,000
ecs.sn1.xlarge	1,500	400,000
ecs.sn1.3xlarge	3,000	500,000
ecs.sn1.7xlarge	6,000	800,000
ecs.sn2.medium	500	100,000
ecs.sn2.large	800	200,000
ecs.sn2.xlarge	1,500	400,000

Instance type	Maximum BPS threshold (Mbit/s)	Maximum PPS threshold (PPS)
ecs.sn2.3xlarge	3,000	500,000
ecs.sn2.7xlarge	6,000	800,000
ecs.sn2.13xlarge	10,000	120,000

Related operations

By default, Anti-DDoS Basic is enabled for an ECS instance after it is created. You can do the following:

- Set a threshold for flow cleaning. After an ECS instance is created, the maximum threshold for the instance type is used for Anti-DDoS Basic by default. However, the maximum BPS threshold for some instance types is excessive for security. Therefore, you must set a threshold according to your business needs. For more information, see *DDoS basic protection configuration* in the Anti-DDoS Basic documentation.
- Cancel flow cleaning, which is not recommended. When the inbound traffic to an ECS instance exceeds the cleaning threshold, the traffic, including normal business traffic, is cleaned. To avoid business interruptions, you can cancel flow cleaning. For more information, see *How to cancel flow cleaning*Cite LeftCite Right.

> **Warning:**
> If you cancel flow cleaning, when the inbound traffic to an ECS instance exceeds 5 Gbit/s, all traffic is routed to a black hole.

7 Images

An image is a running environment template for ECS instances. It generally includes an operating system and preinstalled software. You can use an image to create an ECS instance or change the system disk of an ECS instance. It works as a file copy that includes data from one more multiple disks. These disks can be a single system disk, or the combination of the system disk and data disks.

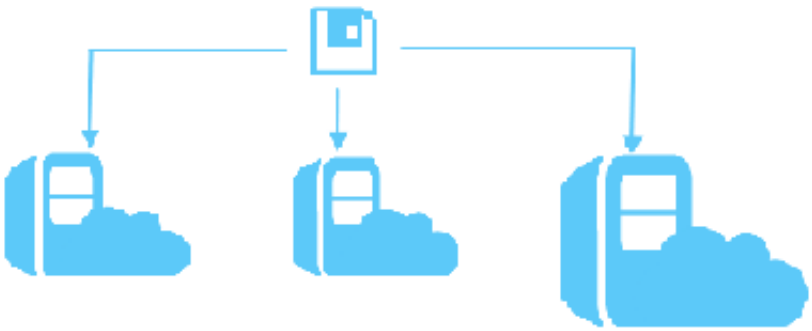


Image types

ECS provides a diverse types of images for you to easily access image resources.

Type	Description	Source
Public image	Public images officially provided by Alibaba Cloud support nearly all main Windows and Linux versions. These images are of high stability and are licensed. You can customize your application environment based on a public image.	Officially provided by Alibaba Cloud
Custom image	Custom images created based on your existing physical server, virtual machine, or cloud host. These images are flexible to meet your personalized needs.	• You can create it based on an existing instance. • You can also import one from the on-premises environment into the corresponding region.
Cloud Marketplace	Provided by third-party service providers (ISV, independent software) Vendor). The image of the Marketplace includes not only the operating system required for the application, but also the configuration environment. It saves you complicated deployment process and deploy the environment with one-click.	*Alibaba Cloud Marketplace*

Type	Description	Source
Shared image	Shared by other Alibaba Cloud users.	A custom image shared by other Alibaba Cloud users.

Public images

Alibaba Cloud provides authorized and certificated public images, which cover nearly all the trending and popular platforms. Note that the available public images are different when you select different instances. The following images are offered by Alibaba Cloud ECS for public use:

Platform	Version of public images
Windows Server	• Windows Server 2008 R2 Enterprise Edition 64 bit Chinese Edition • Windows Server 2008 R2 Enterprise Edition 64 bit English Edition • Windows Server 2012 R2 Data Center Edition 64 bit Chinese Edition • Windows Server 2012 R2 Data Center Edition 64 bit English Edition • Windows Server 2016 R2 Data Center Edition 64 bit Chinese Edition • Windows Server 2016 R2 Data Center Edition 64 bit English Edition • Windows Server Version 1709 Data Center Edition 64 bit Chinese Edition • Windows Server Version 1709 Data Center Edition 64 bit English Edition
CentOS	• CentOS 6.8 64bit • CentOS 6.8 32bit • CentOS 6.9 64bit • CentOS 7.2 64bit • CentOS 7.3 64bit • CentOS 7.4 64bit • CentOS 7.5 64bit
Ubuntu	• Ubuntu 14.04 64bit • Ubuntu 14.04 32bit • Ubuntu 16.04 64bit • Ubuntu 16.04 32bit
Debian	• Debian 8.9 64bit • Debian 9.2 64bit • Debian 9.5 64bit
Red Hat	• Red Hat Enterprise Linux 7.5 64bit • Red Hat Enterprise Linux 7.4 64bit • Red Hat Enterprise Linux 6.9 64bit

Platform	Version of public images
SUSE Linux	• SUSE Linux Enterprise Server 11 SP4 64bit • SUSE Linux Enterprise Server 12 SP4 64bit
OpenSUSE	OpenSUSE 42.3 64bit
Aliyun Linux	Aliyun Linux 17.1 64bit
CoreOS	CoreOS 1465.8.0 64bit
FreeBSD	FreeBSD 11.1 64bit

Image format

Currently, ECS supports VHD, qcow2, and RAW. You must convert other formats before using them in ECS.

Billing details

Billing details of the images are as follows:

Type	Description
Public image	Only Windows Server and Red Hat Enterprise Linux public images involve billing calculation, which are included in the bill when an instance is created. The public Windows Sever images or Red Hat Enterprise Linux images are provided with certificated license and authorized support from Microsoft or Red Hat, you do have to purchase additional license. • Red Hat Enterprise Linux: Billing is related to the instance type. • Windows Server: Free of charge in Alibaba Cloud regions that are in mainland China. For international regions, public Windows Server images are charging services. Other public images are free of charge to use.
Custom image	Free. Potential costs include: • If you use a snapshot to create a custom image: — If the image used by the system disk snapshot comes from the Marketplace, the following cost may incur: the fees for the image, and the fees for snapshot capacity. — If the image used by the system disk snapshot does not come from the Marketplace, the following cost may incur: the fees for snapshot capacity. Currently, snapshot is commercialized.

Type	Description
	• If you use an instance to create a custom image, and the image is from the Marketplace, comply to the billing policies from the ISV.
Alibaba Cloud Marketplace	Subject to ISV policies.
Shared image	If the origin of the shared image is from the Marketplace, it is subject to the ISV policies.

For more information, see *Pricing overview*.

Limits

Custom images, marketplace images, and shared images vary depending on the region. For more information about regions and zones, see *Cite LeftGeneral ReferenceCite Right Regions and zones*.

Related operations

Console operations

- You can *create instances by using existing images*.
- You can change the system disk in any of the following ways:
  - *By using a public image*.
  - *By using other images other than public ones*.
- You can obtain custom images in the following ways:
  - *By creating a custom image by using a snapshot*.
  - *By creating a custom image by using a an instance*.
  - *By importing a custom image*.
- After creating custom images, you can perform the following operations:
  - *Copy your custom images to other regions*.
  - *Share your custom images with other Alibaba Cloud users*.
  - *Export custom images* to local testing environments or your private cloud environments.

API operations

You can view the *APIs about images* in the Developer Guide.