# Elastic Compute Service

## Product Introduction

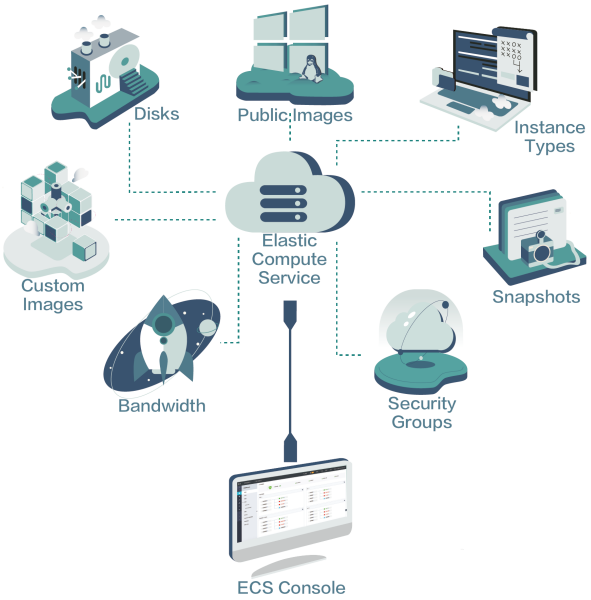
# Product Introduction

## What is ECS?

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 Management Console to configure the instance type, disks, operating system, and other affiliated resources for your ECS instance.




## Advantages

ECS has the following advantages over traditional servers and virtual hosts:

### Stability
ECS has 99.95% service availability and 99.9999999% data reliability. It also supports automatic downtime migration, data snapshot backup and rollback, and system performance alarms.

### Disaster recovery backup
A copy of each data segment is saved, guaranteeing rapid restoration should a segment be physically damaged.

### Security
ECS supports security groups, Anti-DDoS, multi-user isolation, and password cracking defense.

### Multiline access
ECS is based on the optimal routing algorithm of the Border Gateway Protocol (BGP). Multiline BGP machine rooms ensure smooth and balanced access throughout the geographic region. Backbone machine rooms ensure high output bandwidth and dedicated bandwidth.

### Low cost
Large one-time payments are not required. Flexible payment options and Pay-As-You-Go let you cope with business changes.

### Controllability
As an ECS user, you have the permission of a super administrator. This allows you to completely control the operating system of ECS instances, resolve system problems through the management terminal, and perform operations such as environment deployment and software installation.

### Ease of use
A variety of operating systems and applications are supported. Images can be deployed with the click of a button. You can quickly replicate the environment to multiple ECS instances for easy scaling. You can also create ECS instances in batches through custom images and disk snapshots.

### API
API invocation management allows configuration of access to one or multiple servers with the security group feature, making development more convenient.

## Features
ECS supports the following features:

**Flexible instance configuration**
Supports multiple instance generations, three instance type families, and dozens of instance types (ranging from 1-core 1 GiB to 56-core 480 GiB).

**Multiple regions and zones**
Allows instance creation in all regions, some of which have multiple zones.

**Abundant image resources**
Provides various image resources, including public images, custom images, and shared images, allowing quick operating system deployment and applications without installation.

**Numerous operating systems**
Supports multiple Windows and Linux operating systems.

**Multiple storage methods**
Provides three types of data storage disks (Basic Cloud Disks, Ultra Cloud Disks, and SSD Cloud Disks) and I/O-optimized instances.

**Robust network and security**

- Supports two network types (Classic Network and VPC), allowing network management in different dimensions.
- Supports two types of IP addresses (public and private IP addresses), allowing for Intranet interconnection and Internet access.
- Allows free activation of Alibaba Cloud Security products and provides network monitoring.

**Convenient management**
Provides multiple management methods, including the console, VNC, and APIs, ensuring complete control.

**Flexible payment**
Provides flexible payment methods (Subscription and Pay-As-You-Go).

# Application 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 OSS and CDN, to provide advanced solutions.

ECS can be used in applications such as:

**Official corporate websites and simple web applications**
In 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 high-availability 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.

# Instances

# Overview

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.

An ECS instance is the minimal unit that can provide computing services for your business. It provides computing capabilities at a certain specification.

# Instance type family

ECS instances can be divided into multiple specification types, based on the business and usage scenarios. In the same business scenario, you can select multiple new or old type families.

In the same type family, there are also multiple types based on the CPU and memory configuration.

# Instance type

An ECS instance type defines two basic attributes: the instance CPU and memory configuration (including CPU model and clock speed). However, the disks, images, and network attributes of the ECS instance should also be defined at the same time, for the specific service form of the instance to be determined.

# The latest type family

All the ECS instances of the latest type family are I/O-optimized. You can choose the following three types of disks for them:

- SSD cloud disk
- Ultra cloud disk
- General cloud disk

The SSD cloud disks and ultra cloud disks are recommended, to enjoy optimal storage I/O performance.

Instances can be divided into the following type families based on the business scenarios.

## For general computing scenarios

### SN1: General exclusive instances, CPU : memory = 1:2

#### Applicable scenarios

- Medium and large Web servers (high concurrency).
- Front ends of Massively Multiplayer Online (MMO) games.
- Data analysis and computing.
- Computing scenarios featuring fixed performance, such as high-precision coding and decoding, rendering, and gene computing with the CPU.

#### Features

- Stable computing performance.
- The ratio of CPU to memory is 1:2.

- 2.5 GHz Intel Xeon, E5-2682 v4 (Broadwell), or E5-2680 v3 (Haswell) processors are adopted.
- The latest DDR4 memory.
- The instance’ s network performance matches the computing type (the more advanced the instance’ s computing type, the more powerful the network performance).
- I/O-optimized by default.

**Instance types**

| Instance type | vCPU | Memory (GiB) | Ephemeral storage (GiB) | Network performance |
|---|---|---|---|---|
| ecs.sn1.medium | 2 | 4 | - | Moderate |
| ecs.sn1.large | 4 | 8 | - | Moderate |
| ecs.sn1.xlarge | 8 | 16 | - | High |
| ecs.sn1.3xlarge | 16 | 32 | - | High |
| ecs.sn1.7xlarge | 32 | 64 | - | Ultra high |

## SN2: Balanced exclusive instances, CPU : memory = 1:4

**Applicable scenarios**

- Medium and large Web servers (high concurrency).
- Front ends of Massively Multiplayer Online (MMO) games.
- Data analysis and computing.
- Computing scenarios featuring fixed performance, such as high-precision coding and decoding, rendering, gene computing, and Hadoop clusters with a CPU.

**Features**

- Stable computing performance.
- The ratio of CPU to memory is 1:4.
- 2.5 GHz Intel Xeon, E5-2682 v4 (Broadwell), or E5-2680 v3 (Haswell) processors are adopted.
- The latest DDR4 memory.
- The instance’ s network performance matches the computing type (the more advanced the instance’ s computing type, the more powerful the network performance).
- I/O-optimized by default.

**Instance types**

| Instance type | vCPU | Memory (GiB) | Ephemeral storage (GiB) | Network performance |
|---|---|---|---|---|
| ecs.sn2.medium | 2 | 8 | - | Moderate |
| ecs.sn2.large | 4 | 16 | - | Moderate |

| | | | | |
|---|---|---|---|---|
| ecs.sn2.xlarge | 8 | 32 | - | High |
| ecs.sn2.3xlarge | 16 | 64 | - | High |
| ecs.sn2.7xlarge | 32 | 128 | - | Ultra high |
| ecs.sn2.13xlarge | 56 | 224 | - | Ultra high |

# For memory optimization scenarios

## SE1: Memory exclusive instances, CPU : memory = 1:8

### Applicable scenarios

- Cache/Redis.
- Searches.
- Memory database.
- Databases with high I/O, such as Oracle and MongoDB.
- Hadoop clusters.
- Computing scenarios that involve massive data processing and feature fixed performance.

### Features

- Stable computing performance.
- The ratio of CPU to memory is 1:8.
- 2.5 GHz Intel Xeon, E5-2682 v4 (Broadwell), or E5-2680 v3 (Haswell) processors are adopted.
- The latest DDR4 memory.
- The instance’ s network performance matches the computing type (the more advanced the instance’ s computing type, the more powerful the network performance).
- I/O-optimized by default.

### Instance types

| Instance type | vCPU | Memory (GiB) | Ephemeral storage (GiB) | Network performance |
|---|---|---|---|---|
| ecs.se1.large | 2 | 16 | - | Moderate |
| ecs.se1.xlarge | 4 | 32 | - | Moderate |
| ecs.se1.2xlarge | 8 | 64 | - | High |
| ecs.se1.4xlarge | 16 | 128 | - | High |
| ecs.se1.8xlarge | 32 | 256 | - | Ultra high |
| ecs.se1.14xlarge | 56 | 480 | - | Ultra high |

# For computing optimization scenarios

## C4: computing-optimized instances, CPU : memory = 1:2

### Applicable scenarios

- Front ends of Massively Multiplayer Online (MMO) games.
- Video encoding.
- Distributed analysis.
- High-computing performance front-end clusters.
- High-performance Web servers.

### Features

- Stable computing performance.
- The ratio of CPU to memory is 1:2.
- 3.2 GHz Intel Xeon E5-2667 v4 (Broadwell) processors are adopted.
- The latest DDR4 memory.
- The instance’ s network performance matches the computing type (the more advanced the instance’ s computing type, the more powerful the network performance).
- I/O-optimized by default.

### Instance types

| Instance type | vCPU | Memory (GiB) | Ephemeral storage (GiB) | Network performance |
|---|---|---|---|---|
| ecs.c4.xlarge | 4 | 8 | - | High |
| ecs.c4.2xlarge | 8 | 16 | - | High |
| ecs.c4.4xlarge | 16 | 32 | - | Ultra high |

## CM4: computing-optimized instances, CPU : memory = 1:4

### Applicable scenarios

- Front ends of Massively Multiplayer Online (MMO) games.
- Medium and large Web servers (high concurrency).
- Data analysis and computing.
- Computing scenarios featuring fixed performance, such as high-precision coding/decoding, rendering, gene computing, and Hadoop clusters with a CPU.

### Features

- Stable computing performance.
- The ratio of CPU to memory is 1:4.
- 3.2 GHz Intel Xeon E5-2667 v4 (Broadwell) processors are adopted.

- The latest DDR4 memory.
- The instance’ s network performance matches the computing type (the more advanced the instance’ s computing type, the more powerful the network performance).
- I/O-optimized by default.

**Instance types**

| Instance type | vCPU | Memory (GiB) | Ephemeral storage (GiB) | Network performance |
|---|---|---|---|---|
| ecs.cm4.xlarge | 4 | 16 | - | High |
| ecs.cm4.2xlarge | 8 | 32 | - | High |
| ecs.cm4.4xlarge | 16 | 64 | - | Ultra high |
| ecs.cm4.6xlarge | 24 | 96 | - | Ultra high |

### CE4: computing-optimized instances, CPU : memory = 1:8

**Applicable scenarios**

- Front ends of Massively Multiplayer Online (MMO) games.
- Cache/Redis.
- Searches.
- Memory database.
- Databases with high I/O, such as Oracle and MongoDB.
- Hadoop clusters.
- Computing scenarios that involve massive data processing and feature fixed performance.

**Features**

- Stable computing performance.
- The ratio of CPU to memory is 1:8.
- 3.2 GHz Intel Xeon E5-2667 v4 (Broadwell) processors are adopted.
- The latest DDR4 memory.
- The instance’ s network performance matches the computing type (the more advanced the instance’ s computing type, the more powerful the network performance).
- I/O-optimized by default.

**Instance types**

| Instance type | vCPU | Memory (GiB) | Ephemeral storage (GiB) | Network performance |
|---|---|---|---|---|
| ecs.ce4.xlarge | 4 | 32 | - | High |

## For storage optimization scenarios

### I1: High-I/O ephemeral disk instances

I1 is an instance with high-I/O ephemeral disk, and is designed for business scenarios that have extremely high requirements for storage I/O performance.

#### Applicable scenarios

- Large database business scenarios that require the random IOPS read/write capabilities to reach tens of thousands to hundreds of thousands times every second with a low latency.
- Big data and concurrent computing and other large data set business scenarios that require sequential access throughput to reach as high as several GiBs.
- Online games, e-businesses, live videos, media, and customers in other industries that provide online businesses and have low latency and high I/O performance requirements on block storage for I/O-intensive applications.
- Business scenarios that have high requirements on the storage I/O performance and availability of the application layer, such as NoSQL non-relational databases, MPP data warehouses, and distributed file systems.

#### Features

- Up to hundreds of thousands of I/O reads/writes per second while the latency is maintained at the μs magnitude based on high-performance ephemeral NVMe SSD disk resources.
- Stable computing performance.
- The ratio of CPU to memory is 1:4.
- 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors are adopted.
- The latest DDR4 memory.
- The instance’ s network performance matches the computing type (the more advanced the instance’ s computing type, the more powerful the network performance).
- I/O-optimized by default.

#### Instance types

| Instance type | vCPU | Memory (GiB) | Ephemeral storage (GiB) | Network performance |
|---|---|---|---|---|
| ecs.i1.xlarge | 4 | 16 | 2 × 104 | Moderate |
| ecs.i1.2xlarge | 8 | 32 | 2 × 208 | High |
| ecs.i1.4xlarge | 16 | 64 | 2 × 416 | High |
| ecs.i1.8xlarge | 32 | 128 | 2 × 832 | Ultra high |
| ecs.i1.14xlarge | 56 | 224 | 2 × 1456 | Ultra high |

## For heterogeneous computing scenarios

### GA1: GPU-based rendering and computing instances, used in combination with

## an AMD S7150 GPU

### Applicable scenarios

- GPU-dependent non-linear editing and deep learning.
- VR field.
- 3D rendering field.
- Financial analysis, meteorological analysis, geological analysis, computational chemistry, dynamics simulation, genetic engineering, and other fields.

### Features

- The most advanced type offers four AMD S7150 GPUs, with 32 GiB GPU video memory and 8,192 concurrent processing cores, 15.08 TFLOPS of single-precision floating point operations and 1 TFLOPS of double-precision floating point operation capabilities.
- Stable computing performance.
- 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors are adopted.
- The latest DDR4 memory.
- A high-performance NVMe SSD disk as the instance storage, which features highly stable reading/writing latency and up to 230,000 IOPS.
- The instance’ s network performance matches the computing type (the more advanced the instance’ s computing type, the more powerful the network performance).
- I/O-optimized by default.
- Supporting AMD S7150 series GPU.

### Instance types

| Instance type | vCPU | Memory (GiB) | Ephemeral storage (GiB) | GPU | Network performanc e |
|---|---|---|---|---|---|
| ecs.ga1.4xlarge | 16 | 40 | 1 × 350 | 1 × AMD S7150 | High |
| ecs.ga1.8xlarge | 32 | 80 | 1 × 700 | 2 × AMD S7150 | Ultra high |
| ecs.ga1.14xlarge | 56 | 160 | 1 × 1400 | 4× AMD S7150 | Ultra high |

## GN4: GPU-based computing instances

GN4 is used in combination with the NVIDIA M40 GPU and designed for general GPU computing applications using CUDA and OpenCL.

### Applicable scenarios

- Machine learning.
- Fluid mechanics computing.

- Genomics.- Seismic analysis.
- Molecular modeling.
- Financial calculation.
- Other server-end business scenarios that require powerful concurrent floating-point operation capabilities.

**Features**

- The most advanced type offers two NVIDIA M40 GPUs, with a total of 24 GiB GPU video memory and 6,000 concurrent processing cores, and 14 TFLOPS of single-precision floating point operation capabilities.
- Stable computing performance.
- 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors are adopted.
- The latest DDR4 memory.
- The instance’ s network performance matches the computing type (the more advanced the instance’ s computing type, the more powerful the network performance).
- I/O-optimized by default.

**Instance types**

| Instance type | vCPU | Memory (GiB) | Ephemeral storage (GiB) | GPU | Network performanc e |
|---|---|---|---|---|---|
| ecs.gn4.8xla rge | 32 | 48 | - | 1 × Nvidia M40 | Ultra high |
| ecs.gn4.14xl arge | 56 | 96 | - | 2 × Nvidia M40 | Ultra high |

# For temporary usage

## XN4: Compact shared instances

**Applicable scenarios**

- Web applications of small websites.
- Small databases.
- Development or testing environments, code storage servers, and other scenarios.

**Features**

- The ratio of CPU to memory is 1:1.
- 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors are adopted.
- The latest DDR4 memory.
- I/O-optimized by default.

**Instance types**

| Instance type | vCPU | Memory (GiB) | Ephemeral storage (GiB) |
| --- | --- | --- | --- |
| ecs.xn4.small | 1 | 1 | - |

## N4: General shared instances

### Applicable scenarios

- Small and medium-sized web servers.
- Batch processing.
- Distributed analysis.
- Advertisement services.

### Features

- The ratio of CPU to memory is 1:2.
- 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors are adopted.
- The latest DDR4 memory.
- I/O-optimized by default.

### Instance types

| Instance type | vCPU | Memory (GiB) | Ephemeral storage (GiB) |
| --- | --- | --- | --- |
| ecs.n4.small | 1 | 2 | - |
| ecs.n4.large | 2 | 4 | - |
| ecs.n4.xlarge | 4 | 8 | - |
| ecs.n4.2xlarge | 8 | 16 | - |
| ecs.n4.4xlarge | 16 | 32 | - |
| ecs.n4.8xlarge | 32 | 64 | - |

## MN4: Balanced shared instances

### Applicable scenarios

- Medium-sized web servers.
- Batch processing.
- Distributed analysis.
- Advertisement services.
- Hadoop clusters.

### Features

- The ratio of CPU to memory is 1:4.

- 2.5 GHz Intel Xeon E5-2682 v4 (Broadwell) processors are adopted.
- The latest DDR4 memory.
- I/O-optimized by default.

#### Instance types

| Instance type | vCPU | Memory (GiB) | Ephemeral storage (GiB) |
|---|---|---|---|
| ecs.mn4.small | 1 | 4 | - |
| ecs.mn4.large | 2 | 8 | - |
| ecs.mn4.xlarge | 4 | 16 | - |
| ecs.mn4.2xlarge | 8 | 32 | - |
| ecs.mn4.4xlarge | 16 | 64 | - |
| ecs.mn4.8xlarge | 32 | 128 | - |

### Configuration change between different type families

- You can change the configuration among the three shared instance type families (XN4, N4, MN4), and within the same instance type family.
- You can change the configuration among the three exclusive instance type families (SN1, SN2, SE1), and within the same instance type family.
- You can change the configuration among the three computing-optimized instance type families (C4, CM4, CE4), and within the same instance type family.
- You can change the configuration within the heterogeneous computing GA1 type family.
- You can change the configuration within the heterogeneous computing GN4 type family.

## Previous generations for temporary usage

All the ECS instances of the previous generation are I/O-optimized, and you can choose the following three types of disks for them:

- SSD cloud disk
- Ultra cloud disk
- General cloud disk

### N1: General shared instances

#### Applicable scenarios

- Small and medium-sized web servers.
- Batch processing.
- Distributed analysis.
- Advertisement services.

### Features

- The ratio of CPU to memory is 1:2.
- 2.5 GHz Intel Xeon E5-2680 v3 (Haswell) processors are adopted.
- The latest DDR4 memory.
- I/O-optimized by default.

### Instance types

| Instance type | vCPU | Memory (GiB) | Ephemeral storage (GiB) |
|---|---|---|---|
| ecs.n1.tiny | 1 | 1 | - |
| ecs.n1.small | 1 | 2 | - |
| ecs.n1.medium | 2 | 4 | - |
| ecs.n1.large | 4 | 8 | - |
| ecs.n1.xlarge | 8 | 16 | - |
| ecs.n1.3xlarge | 16 | 32 | - |
| ecs.n1.7xlarge | 32 | 64 | - |

## N2: Balanced shared instances

### Applicable scenarios

- Medium-sized web servers.
- Batch processing.
- Distributed analysis.
- Advertisement services.
- Hadoop clusters.

### Features

- The ratio of CPU to memory is 1:4.
- 2.5 GHz Intel Xeon E5-2680 v3 (Haswell) processors are adopted.
- The latest DDR4 memory.
- I/O-optimized by default.

### Instance types

| Instance type | vCPU | Memory (GiB) | Ephemeral storage (GiB) |
|---|---|---|---|
| ecs.n2.small | 1 | 4 | - |
| ecs.n2.medium | 2 | 8 | - |
| ecs.n2.large | 4 | 16 | - |

| | | | |
|---|---|---|---|
| ecs.n2.xlarge | 8 | 32 | - |
| ecs.n2.3xlarge | 16 | 64 | - |
| ecs.n2.7xlarge | 32 | 128 | - |

## E3: Memory shared instances

### Applicable scenarios

- Cache/Redis.
- Searches.
- Memory database.
- Databases with high I/O, for example, Oracle and MongoDB.
- Hadoop clusters.
- Computing scenarios that involve massive data processing.

### Features

- The ratio of CPU to memory is 1:8.
- 2.5 GHz Intel Xeon E5-2680 v3 (Haswell) processors are adopted.
- The latest DDR4 memory.
- I/O-optimized by default.

### Instance types

| Instance type | vCPU | Memory (GiB) | Ephemeral storage (GiB) |
|---|---|---|---|
| ecs.e3.small | 1 | 8 | - |
| ecs.e3.medium | 2 | 16 | - |
| ecs.e3.large | 4 | 32 | - |
| ecs.e3.xlarge | 8 | 64 | - |
| ecs.e3.3xlarge | 16 | 128 | - |

## Phasing-out instance types

The phasing-out instance types include: T1, T2, S1, S2, S3, M1, M2, C1, and C2.All these instance types are legacy shared instance types. They are still categorized using the previous method (that is, by the number of cores - 1, 2, 4, 8, and 16 cores) and are not sensitive to type families.

### Features

- 2.6 GHz Intel Xeon E5-2650 v2 processors are adopted.
- The latest DDR3 memory.
- I/O-optimized and non I/O-optimized at your choice.

### I/O-optimized instance types

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

- SSD cloud disk
- Ultra cloud disk

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

### Non I/O-optimized instance types

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

| Specification types | Type code | vCPU | Memory (GiB) |
|---|---|---|---|
| Tiny | ecs.t1.small | 1 | 1 |
| Standard | ecs.s1.small | 1 | 2 |
| | ecs.s1.medium | 1 | 4 |
| | ecs.s1.large | 1 | 8 |
| | ecs.s2.small | 2 | 2 |
| | ecs.s2.large | 2 | 4 |
| | ecs.s2.xlarge | 2 | 8 |
| | ecs.s2.2xlarge | 2 | 16 |
| | ecs.s3.medium | 4 | 4 |
| | ecs.s3.large | 4 | 8 |

| | | | |
|---|---|---|---|
| High Memory | ecs.m1.medium | 4 | 16 |
| | ecs.m2.medium | 4 | 32 |
| | ecs.m1.xlarge | 8 | 32 |
| High CPU | ecs.c1.small | 8 | 8 |
| | ecs.c1.large | 8 | 16 |
| | ecs.c2.medium | 16 | 16 |
| | ecs.c2.large | 16 | 32 |
| | ecs.c2.xlarge | 16 | 64 |

### Configuration change between different type families

- You can change the configuration among the three shared instance type families (N1, N2, E3), and within the same instance type family.
- You can change the configuration among the nine shared instance type families (T1, T2, S1, S2, S3, M1, M2, C1, and C2), and within the same instance type family.

# Instance life cycle

The instance life cycle begins with creation (purchase) and ends with final release (the expiration of the yearly or monthly instance subscription). In pay-as-you-go cases, instances end as a result of unpaid fees or voluntary release.

## Inherent instance statuses

There are several inherent instance statuses in an instance life cycle, as listed in the following table.

| Status | Status Property | Description | Corresponding API Status |
|---|---|---|---|
| Preparing* | Intermediate Status | After an instance is created, it remains in this status before running. | Pending |
| Created* | Stable Status | An instance is in this status when it has been created and is awaiting start up. | Stopped |
| Starting* | Intermediate Status | An instance is in this status after it is started or restarted in the console or through API, and | Starting |

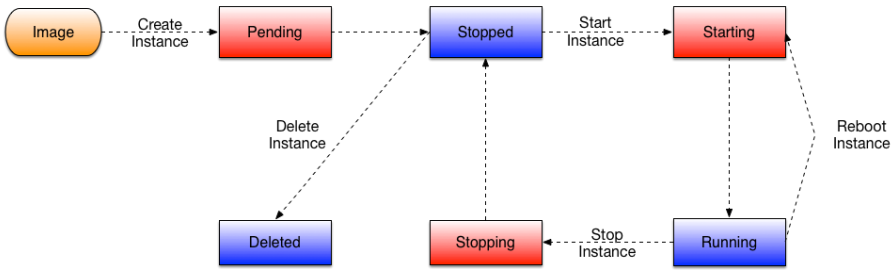
| | | | |
|---|---|---|---|
| | | before it is running. | |
| Running | Stable Status | The instance is operating properly and can accommodate your business needs. | Running |
| Stopping* | Intermediate Status | An instance is in this status after the stop operation is performed in the console or through API but before it actually stops. | Stopping |
| Stopped | Stable Status | The instance has been stopped properly. In this status, the instance cannot accommodate external services. | Stopped |
| Re-initializing* | Intermediate Status | An instance is in this status after the system disk and/or data disk is re-initialized in the console or through API and before it is running. | Stopped |
| Replacing System Disk | Intermediate Status | An instance is in this status after the operating system is replaced or another such operation is performed in the console or through API and before it is running. | Stopped |
| Expired | Stable Status | The yearly/monthly instance subscription has expired because it has not been properly renewed. The pay-as-you-go instances have expired because they are in arrears. **Note:** After expiration, both the yearly/monthly and pay-as-you-go instances will continue running for | Stopped |

| | | | |
|---|---|---|---|
| | | 15 days, and data will be retained for an extra 15 days, after which the instances will be released and the data will be removed permanently. | |

* If an instance remains in the Preparing, Created, Starting, Stopping, Re-initializing, or Replacing System Disk status for a long time, it has encountered an exception.

## API status chart

This flowchart describes the corresponding relationships between console statuses and API statuses. The API status chart is shown below.




# Disks

# Overview

An ECS disk can be used jointly or separately to meet the requirements of different application scenarios. ECS disks are categorized into ephemeral SSD disks and cloud disks. Compared with ephemeral SSD disks, cloud disks are more reliable as they use a triplicate distributed system to provide block-level data storage for ECS instances, ensuring 99.9999999% data reliability. Cloud disks are categorized as one of the following:

SSD cloud disks
Ideal for I/O-intensive applications, and provide stable and high random IOPS performance.

Ultra cloud disks
Ideal for medium I/O load application scenarios and provide a storage performance of up to 3,000 random IOPS for ECS instances.

Basic cloud disks
Ideal for least I/O-intensive application scenarios and provide an I/O performance of several hundred IOPS for ECS instances.

**Note:** For detailed instructions on attaching a disk, refer to Attach a data disk from Elastic Compute Service – User Guide.

# Disk comparison

The following table describes the features and typical application scenarios for different types of cloud disks.

| Item | SSD cloud disk | Ultra cloud disk | Basic cloud disk |
|---|---|---|---|
| Maximum capacity | 32768 GB | 32768 GB | 2000 GB |
| Maximum IOPS | 20000 | 3000 | Several hundreds |
| Maximum throughput | 256 MBps | 80 MBps | 30 MBps |
| Performance calculation formula | IOPS = min{30*capacity, 20000} Throughput = min{50 + 0.5*capacity, 256} MBps | IOPS = min{1000+6*capacity, 3000} Throughput = min {50+0.1*capacity, 80} MBps | N/A |
| Access latency | 0.5 ~ 2 ms | 1 ~ 3 ms | 5 ~ 10 ms |
| Data reliability | 99.9999999% | 99.9999999% | 99.9999999% |
| API name | cloud_ssd | cloud_efficiency | cloud |
| Price* | US$0.15/GB/month | US$0.08/GB/month | US$0.05/GB/month |
| Typical application scenarios | - I/O-intensive applications<br>- Medium/Large relational databases<br>- NoSQL databases | - Medium/Small databases<br>- Large-scale development and testing<br>- Web server logs | Infrequent access or low-I/O applications |

* Prices shown are for the US West region. For more information, see ECS Price at https://intl.aliyun.com/product/ecs#pricing.

For more information about ephemeral SSD disks, see ephemeral SSD disks.

# Methods to test disk performance

- Test random writing IOPS：
  fio -direct=1 -iodepth=128 -rw=randwrite -ioengine=libaio -bs=4k -size=1G -numjobs=1 -runtime=1000 -group_reporting -filename=iotest -name=Rand_Write_Testing
- Test random reading IOPS：
  fio -direct=1 -iodepth=128 -rw=randread -ioengine=libaio -bs=4k -size=1G -numjobs=1 -runtime=1000 -group_reporting -filename=iotest -name=Rand_Read_Testing
- Test writing throughput ：
  fio -direct=1 -iodepth=64 -rw=write -ioengine=libaio -bs=64k -size=1G -numjobs=1 -runtime=1000 -group_reporting -filename=iotest -name=Write_PPS_Testing
- Test reading throughput ：
  fio -direct=1 -iodepth=64 -rw=read -ioengine=libaio -bs=64k -size=1G -numjobs=1 -runtime=1000 -group_reporting -filename=iotest -name=Read_PPS_Testing

Descriptions of fio parameters：

| Parameter | Description |
| --- | --- |
| -direct=1 | Ignore I/O cache when testing. Data is written directly. |
| -rw=randwrite | Read and write policies. Available options: randread (random read), randwrite(random write), read(sequential read), write(sequential write) and randrw (random read and write) |
| -ioengine=libaio | Use libaio as the testing method (Linux AIO, Asynchronous I/O). Usually there are two ways for an application to use I/O: synchronous and asynchronous. Synchronous I/O only sends out one I/O request each time, and returns only after the kernel is completed. In this case, the iodepth is always less than 1 for a single job, but can be resolved by multiple concurrent jobs. Usually 16 - 32 concurrent jobs can fill up the iodepth. Asynchronous method uses libaio to submit a batch of I/O request each time, thus reduces interaction times, and makes interaction more effective. |
| -bs=4k | The size of each block for one I/O is 4k. If not specified, the default value 4k is used. |
| -size=1G | The size of the testing file is 1G. |
| -numjobs=1 | The number of testing jobs is 1. |
| -runtime=1000 | Testing time is 1000 seconds. If not specified, the test will go on with the value specified for -size, and write data in -bs each time. |
| -group_reporting | The display mode of showing the testing results. Group_reporting means sums up |

| | |
|---|---|
| | statistics of each job, instead of showing statistics by different jobs. |
| -filename=iotest | The output path and name of test files. Remember to delete the related files to free up space. |
| -name=Rand_Write_Testing | The name of the testing task. |

# Disk categories and application scenarios

## SSD cloud disks

### Product features

SSD cloud disks use a distributed, triplicate mechanism to provide high-performance storage with stable and high random I/O and high data reliability. They provide the following features:

High random I/O performance

The maximum random read/write IOPS is 20,000. Each GB of capacity provides 30 random IOPS. For example, a 100 GB SSD cloud disk can provide 3,000 IOPS, and a 334 GB SSD cloud disk can provide 10,020 IOPS.

High throughput

The maximum throughput is 256 MBps. The throughput of an SSD cloud disk can be determined using the equation min {50 + 0.5*disk_size, 256} MBps.

High data reliability

SSD cloud disks adopt a distributed, triplicate mechanism to provide 99.9999999% data reliability.

Large storage capacity

A single SSD cloud disk provides up to 32,768 GB storage space.

Independent attaching

SSD cloud disks can be attached to any ECS instance in the same zone.

**Note:** Expected IOPS performance can be achieved only when the SSD cloud disk is attached to an I/O-optimized instance. An SSD cloud disk attached to a non I/O-optimized instance cannot achieve the expected IOPS performance.

## Performance baselines

| Block size | Maximum IOPS | Maximum throughput |
| --- | --- | --- |
| 4/8 KB | 20,000 | N/A |
| 16 KB | 16,300 | 256 MBps |
| 32 KB | 8,150 | 256 MBps |
| 64 KB | 4,100 | N/A |

## Application Scenarios

SSD cloud disks have stable and high random I/O performance, and high data reliability. They are applicable to the following scenarios:

- PostgreSQL, MySQL, Oracle, SQL Server, and other medium/large relational database applications.
- Medium to large development and testing environments with high data reliability requirements.

# Ultra Cloud Disks

## Product Features

Ultra cloud disks adopt the hybrid media of SSD and HDD as the storage media. They provide the following features:

High random I/O performance

The maximum random read/write IOPS is 3,000. The random read/write IOPS is initially 1,000 and increases by 6 IOPS for each GB. For example, a 250 GB ultra cloud disk features 2,500 random read/write IOPS.

High throughput

The maximum throughput is 80 MBps. The throughput is initially 50 MBps and increases by 0.1 MBps for each GB. For example, a 250 GB ultra cloud disk features a throughput of 75 MBps.

High data reliability

Ultra cloud disks adopt a distributed, triplicate mechanism to provide 99.9999999% data reliability.

Large storage capacity

A single ultra cloud disk provides up to 32768 GB storage space.

Independent attaching

Ultra cloud disks can be attached to any ECS instance in the same zone.

## Application Scenarios

Ultra cloud disks are applicable to the following scenarios:

- MySQL, SQL Server, PostgreSQL, and other small or medium relational database applications.
- Medium or large development and testing environments with high data reliability requirements and intermediate performance requirements.

# Basic Cloud Disks

## Product Features

Basic cloud disks adopt HDDs as the storage medium and use a distributed, triplicate mechanism to provide high data reliability. They provide the following features:

High random I/O performance

The maximum random read/write IOPS is of several hundreds.

High throughput

The maximum throughput is 30 ~ 40 MBps.

High data reliability

Disks adopt a distributed, triplicate mechanism provides 99.9999999% data reliability.

Large storage capacity

A single basic cloud disk provides up to 2,000 GB storage space.

Independent attaching

Basic cloud disks can be attached to any ECS instance in the same zone.

## Application Scenarios

Basic cloud disks are applicable to the following scenarios:

- Applicable to scenarios in which data is not frequently accessed, or which have low I/O loads. If an application requires higher I/O performance, using an SSD cloud disk is recommended.
- Application environments that require low costs and have random I/O reading/writing.

# Introduction to triplicate technology

The Alibaba Cloud Distributed File System provides stable, efficient, and reliable random data access capabilities for ECS.

## Chunks

When ECS users perform read and write operations onto virtual disks, the operations are translated into corresponding processes on the files stored in the Alibaba Cloud Distributed File System. Alibaba Cloud provides a flat storage space, in which the linear addresses are divided into chunks, also referred to as slices. Alibaba Cloud employs a certain strategy to create three copies for each chunk and stores these copies on different nodes, ensuring the reliability of user data.

## Principles of Triplicate Technology

The Alibaba Cloud data storage system consists of three roles: Master, Chunk Server, and Client. The write operation of an ECS user goes through several conversions and is executed by the Client. The procedure is as follows:

1. The Client calculates the chunk corresponding to a given writing operation.
2. The Client sends a request to the Master for the storage location of the three copies of the chunk.
3. The Client sends writing requests to the three Chunk Servers according to the results returned from the Master.
4. The Client returns a message to the user indicating whether the operation was successful.

The distribution strategy of the Master is decided based on an overall consideration of the following:

- Disk use conditions of all Chunk Servers in the cluster.
- Distribution of the Chunk Servers under different kinds of switch racks.
- The power supply.
- The instrument load.

This strategy ensures that all the copies of a Chunk are distributed on different Chunk Servers on different racks. This can effectively prevent data unavailability caused by the failure of a Chunk Server or rack.

## Data protection mechanism

When some data nodes are corrupted, or some hard drives on a certain data node fail, the number of valid copies of some Chunks in the cluster will be less than three. If this occurs, the Master initiates the copy mechanism to copy data between Chunk Servers, making three valid copies of all Chunks in the cluster.

In sum, for the data on the cloud disk, all user operations and data addition or modification will be synchronized to the three copies. This mode ensures the reliability and consistency of user data.

To prevent data losses caused by virus infection or cyber-attacks, you are recommended to use the triplicate technology with other protection methods, such as taking snapshots.

# Ephemeral SSDs

Ephemeral SSD disks use the local storage of physical machines at which the instances are located. This type of storage provides block-level data access capabilities to instances. It has low latency, high random IOPS, and high throughput I/O capability.

Pay attention to the following issues when using ephemeral SSDs:

- The storage space provided by ephemeral SSDs of servers has potential single point of failure (SPOF) risks. It is recommended to implement data redundancy at the application layer to ensure data availability.
- Users cannot upgrade or downgrade the CPU, memory, and ephemeral SSD disks after purchasing.
- Since the ephemeral SSD uses the local disk of a physical server, it does not support independent attaching/detaching.

## Product features

Ephemeral SSD disks provide the following features:

**Low latency**
Under normal conditions, the access latency is in microseconds.

**High random I/O performance**
The maximum random IOPS is 12,000.

**High throughput**
The maximum I/O throughput is 300 MBps.

**Large storage capacity**
A single ephemeral SSD provides a maximum storage space of 800 GB.

## Application Scenarios

Ephemeral SSD disks are applicable to distributed, I/O-intensive applications with redundancy and scenarios that do not require high data reliability, such as the following:

**Distributed applications**

NoSQL and MPP data warehouse, distributed file systems, and other I/O-intensive applications have their own distributed data redundancy. Ephemeral SSD disks can provide low latency, high random I/O, and high throughput I/O performance.

**Logs for large online applications**

Large online applications can produce large amounts of log data and require high-performance storage. At the same time, the log data does not require highly reliable storage.

**As the swap partition of an instance**

When the memory required by an application exceeds the memory actually allocated, a swap space can be used in Linux. When the swap space is enabled, the Linux system can frequently swap in-use memory pages from the physical memory to the swap space (no matter whether it is a dedicated partition of the existing file system or a swap file). In addition, it can free up space for memory pages that require a high access speed.

# Network and security

# Intranet

Currently, Alibaba Cloud servers communicate through the 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 speed may fluctuate.

If you need to transmit data between two ECS instances in the same region, you should use an intranet connection. Intranet connections can also be used to connect RDS, Server Load Balancer, and OSS instances. The internet speed of these instances is based on a gigabit shared bandwidth environment. At present, you can also use a direct intranet connection to link RDS, Server Load Balancer, and OSS instances with ECS instances in the same region.

For ECS instances in the intranet:

For instances of Classic network:

- Intranet communication is by default used only for instances in the same security group of the same account in the same region.
- An intranet communication can also be used for instances in the same security group of the same account and region but of different zones, even if the intranet IP addresses are in different network segments.
- For intranet communication between instances in the same region but of different accounts, you can use security groups. For more information, see Application scenarios of security group from ECS User Guide.

For instances of VPC network:

- Intranet communication is by default used only for instances in the same security group of the same account and same VPC network in the same region.
- An intranet communication can also be used for instances of the same account and region but of different VPC networks only if you use ExpressConnect to authorize their intranet communication. For more information, see Application scenarios from Product Introduction to ExpressConnect.

The intranet IP addresses of instances cannot be modified or changed.

Intranet and Internet addresses of instances do not support virtual IP (VIP) configuration.

Instances of different network types cannot communicate with each other in intranet.

# IP addresses for Classic network

IP addresses are an important means for users to access ECS instances, and for ECS instances to provide external services. Currently, classic IP addresses are uniformly distributed by Alibaba Cloud. They are divided into public and private IP addresses.

## Private IP addresses

An instance is allocated with a private network card and bound to a specific private IP address. Private IP addresses are required and cannot be modified.

If a private IP address is changed independently in an operating system, communication in the private network will be interrupted.

Communication traffic through private IP addresses between instances in the same region is free. Private IP addresses can be used in the following scenarios:

- Load balancing of the Server Load Balancer.
- Intranet mutual access between ECS instances.
- Intranet mutual access between an ECS instance and another cloud service (such as OSS and RDS).

## Public IP addresses

Each instance is by default configured with a public network interface card. Unlike private IP addresses, public IP addresses are optional. If you select a public network bandwidth greater than 0 Mbps when purchasing an instance, a public IP address will be allocated during creation of the instance.

Regardless of your selected payment method, you must select a public network bandwidth limit. The bandwidth limit you select will determine the limit of the outgoing bandwidth for the public network card.

Public network traffic will be charged. Public IP addresses can be used in the following scenarios:

- Mutual access between an ECS instance and the Internet.
- Mutual access between an ECS instance and another cloud service

## Multicast and Broadcast

ECS does not support multicast or broadcast.

# Security groups

A security group is a logical group that groups 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 network, but instances in different security groups by default cannot communicate through an intranet. However, mutual access can be authorized between two security groups.

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 ECSs. As an important means of security isolation, security groups are used to divide security domains on the cloud.

## Security group restrictions

A single security group cannot contain more than 1,000 instances. If you require intranet mutual access between more than 1,000 instances, you can allocate them to different security groups and permit mutual access through mutual authorization.

- Each instance can join up to five security groups.
- Each user can have up to 100 security groups.
- Adjusting security groups will not affect the continuity of user service.
- Security groups are stateful. If an outbound packet is permitted, inbound packets corresponding to this connection will also be permitted.
- Security groups have two network types: classic network and Virtual Private Cloud (VPC).
    - Classic Network type instances can join security groups on classic networks in the same region.
    - VPC type instances can join security groups on the same VPC.

## Security group rules

Security group rules can be set to permit or forbid ECS instances associated with security groups to access a public network or an intranet from inbound and outbound directions.

You can authorize or delete security group rules at any time. Security group rules you have changed will automatically apply to ECS instances associated with security groups.

When setting security group rules, make sure security group rules are simple. If you associate an ECS instance with multiple security groups, up to 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.

# Images

# 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.

ECS allows you to easily obtain an image in the following ways:

- Choosing a public image officially provided by Alibaba Cloud (multiple Windows and Linux versions are available).
- Creating a custom image based on an existing ECS instance.
- Choosing an image shared by another Alibaba Cloud account.

You can import an offline image file into an ECS cluster to generate a custom image.

You can also copy a custom image to another region to maintain a consistent environment and application deployment across multiple regions.

# Snapshots

# Overview

A snapshot is a copy of data on a disk at a certain point in time. Scheduled creation of disk snapshots ensures continuous operation of your business. Snapshot is a simple and efficient data protection method, and is recommended for the following scenarios:

**Routine backup of system and data disks**

You can back up business-critical data at regular intervals using snapshots to prevent data loss from misoperations, attacks, and viruses.

### OS replacement

Before important operations such as upgrading application software or migrating business data, you need to create one or more snapshots. In case of any issues occurring during the upgrade or migration, you can restore timely to normal status using the snapshots.

### Use of multiple copies of production data

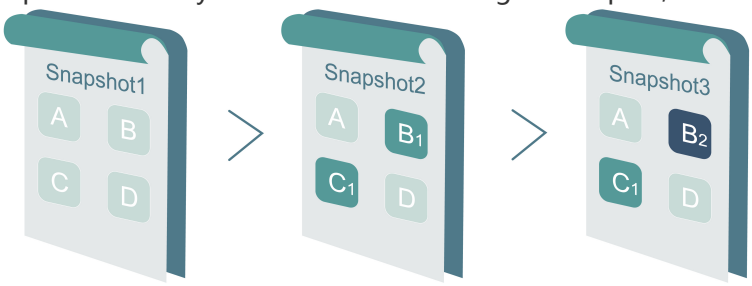
You can take snapshots of production data to provide close-to-real-time production data for data mining, report queries, and developing and testing applications. You can also take snapshots to reuse data on a disk as basic data for another disk.

### Restoring data stored on cloud disks

Cloud disks provide a secure storage method to ensure that your stored content will not be lost. However, if the data stored on a cloud disk is incorrect due to an application error, or the data is maliciously tampered by hackers through an application vulnerability, snapshots ensure that your data can be restored to the desired state.

# Incremental snapshot mechanism

Snapshots on Alibaba Cloud are taken using an incremental method. In this method, two snapshots are compared and only the data that has changed is copied, as shown in the following image:




In the preceding figure, Snapshot 1, Snapshot 2, and Snapshot 3 are the first, second, and third snapshots of a disk. The file system checks the disk data by blocks. When a snapshot is created, only the blocks with changed data will be copied to the snapshot. In this example:

1. In Snapshot 1, all data on the disk is copied since it is the first disk snapshot.
2. Snapshot 2 only copies the changed data blocks B1 and C1. Data blocks, A and D, are referenced from Snapshot 1.
3. Snapshot 3 copies the changed data block B2 but references data blocks, A, D, and C1.

When the disk needs to be restored to the status at the time of Snapshot 3, you can perform snapshot rollback to copy data blocks A, B2, C1, and D to the disk.

If Snapshot 2 is deleted, data block B1 in the snapshot will be deleted but data block C1 will not be deleted. In this way, when the disk is restored to the status at the time of Snapshot 3, data block C1 can also be restored.

Snapshot creation time varies depending on actual volume. For a frame of reference, it typically takes several minutes to manually create a 40 GB snapshot.

Snapshots are stored on the Object Storage Service (OSS), but they are invisible to users and will not be computed in the OSS space occupied by the users' buckets. Snapshot operations can only be performed through the ECS console or APIs.

# ECS Snapshot 2.0

Built on original basic snapshot features, ECS Snapshot 2.0 data backup service provides a higher snapshot quota and more flexible automatic task policies, further reducing its impact on business IO. The features of ECS Snapshot 2.0 are described in the following table.

| Feature | Original snapshot specifications | Snapshot 2.0 specifications | User benefit |
| --- | --- | --- | --- |
| Snapshot quota | (Number of disks)*6+6 | 64 snapshots for each disk | Longer protection circle<br>Smaller protection granularity |
| Automatic task policy | Hardcoded, triggered once daily, and unmodifiable | Customizable weekly snapshot day, time of day, and snapshot retention period<br>Query-able disk quantity and related details associated with an automatic snapshot policy | More flexible protection policy |
| Implementation principle | COW (Copy-on-write) | ROW (Redirect-on-write) | Mitigated performance impact of the snapshot task on business IO write |

The implementation of ECS Snapshot 2.0 features is described in the following table.

| Feature | Implementation |
| --- | --- |
| Snapshot quota | Snapshot backup of a data disk for non-core businesses occurs at 00:00 every day. This backup data is retained for over 2 months.<br>Snapshot backup of a data disk for core businesses occurs every 4 hours. This backup data is retained for over 10 days. |

| | |
|---|---|
| Automatic task policy | A user can take snapshots on the hour and for several times in a day.<br>A user can choose any day as the recurring day for taking weekly snapshots.<br>A user can specify the snapshot retention period or choose to retain it permanently (When the maximum number of automatic snapshots has been reached, the oldest automatic snapshot will be deleted). |
| Implementation principle | The implementation principle is not made visible to users, allowing snapshots to be taken at any time of day without affecting user experience. |

# ECS Snapshot 2.0 vs. traditional storage products

Alibaba Cloud ECS Snapshot 2.0 has many advantages compared with the snapshot feature of traditional storage products, as described in the following table.

| Comparison item | ECS Snapshot 2.0 | Snapshot feature of traditional storage products |
|---|---|---|
| Capacity limit | Unlimited capacity, meeting data protection needs for extra-large businesses. | Capacity limited by initial storage device capacity, merely meeting data protection needs for a few core services. |
| Scalability | One-click auto scaling, allowing you to scale up and down according to their business scale, in mere seconds. | Poor scalability, restrained by factors such as production and storage performance, available capacity, and vendor support capabilities. Scaling typically takes 1 ~ 2 weeks. |
| Cost | Billed based on the actual amount of data changed in your business and snapshot size. | Large, inefficient upfront investment involving software licenses, reserved space, and upgrade and maintenance expenses. |
| Usability | 24x7 online post-sales support. | Complex operations, greatly restrained by vendor support capabilities. |

# Change history

| Description | Date |
| --- | --- |
| Germany data center went live. | November 2016 |
| Instance Generation III went live. | November 2016 |
| Japan data center went live. | November 2016 |
| The system disk was resized. | January 2016 |
| Instance Generation II went live. | November 2015 |
| The security group feature went live. | November 2015 |
| West-USA Zone 1B went live. | October 2015 |
| The image market was commercialized. | September 2015 |
| Singapore data center went live. | September 2015 |
| The efficient cloud disk went live. | September 2015 |
| The tag grouping feature went live. | August 2015 |
| The Virtual Private Cloud (VPC) went live. | August 2015 |
| The image of Windows Server 2003 was deprecated. | June 2015 |
| The shared image went live. | May 2015 |
| Disk resizing went live. | April 2015 |
| The ephemeral SSD was officially commercialized. | December 2014 |
| Deployment of the Docker container application was allowed. | October 2014 |
| Shenzhen data center went live. | August 2014 |
| The independent cloud disk feature went live. | August 2014 |
| The zone feature went live. | July 2014 |
| The automatic snapshot feature went live. | June 2014 |
| Hong Kong data center went live. | May 2014 |
| The image market went live. | May 2014 |
| Beijing data center went live. | April 2014 |
| ECS API was officially launched. | April 2014 |
| ECS’ s brand new user-defined image feature went live. | July 2013 |
| The official website of Alibaba Cloud was successfully launched, and sales of ECS to | July 2011 |

| | |
|---|---|
| external customers began. | |