Alibaba Cloud
Elastic Compute Service

Product Introduction

Issue: 20200520
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 contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of this document shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates.

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

<table>
<thead>
<tr>
<th>Style</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>🚨</td>
<td>A danger notice indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.</td>
<td>🚨 <strong>Danger:</strong> Resetting will result in the loss of user configuration data.</td>
</tr>
<tr>
<td>⚠️</td>
<td>A warning notice indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.</td>
<td>⚠️ <strong>Warning:</strong> Restarting will cause business interruption. About 10 minutes are required to restart an instance.</td>
</tr>
<tr>
<td>🔄</td>
<td>A caution notice indicates warning information, supplementary instructions, and other content that the user must understand.</td>
<td>🔄 <strong>Notice:</strong> If the weight is set to 0, the server no longer receives new requests.</td>
</tr>
<tr>
<td>📝</td>
<td>A note indicates supplemental instructions, best practices, tips, and other content.</td>
<td>📝 <strong>Note:</strong> You can use Ctrl + A to select all files.</td>
</tr>
<tr>
<td>&gt;</td>
<td>Closing angle brackets are used to indicate a multi-level menu cascade.</td>
<td>Click <strong>Settings &gt; Network &gt; Set network type.</strong></td>
</tr>
<tr>
<td><strong>Bold</strong></td>
<td>Bold formatting is used for buttons, menus, page names, and other UI elements.</td>
<td>Click <strong>OK.</strong></td>
</tr>
<tr>
<td><strong>Courier font</strong></td>
<td>Courier font is used for commands.</td>
<td>Run the <code>cd /d C:/window</code> command to enter the Windows system folder.</td>
</tr>
<tr>
<td><strong>Italic</strong></td>
<td>Italic formatting is used for parameters and variables.</td>
<td><code>bae log list --instanceid Instance_ID</code></td>
</tr>
<tr>
<td>[] or [a</td>
<td>b]</td>
<td>This format is used for an optional value, where only one item can be selected.</td>
</tr>
<tr>
<td>Style</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>{} or {a</td>
<td>b}</td>
<td>This format is used for a required value, where only one item can be selected.</td>
</tr>
<tr>
<td>Contents</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>Legal disclaimer</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>Document conventions</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>1 What is ECS?</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2 Benefits of ECS</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>3 Scenarios</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>4 Regions and zones</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>5 Limits</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>6 Usage notes</td>
<td>27</td>
<td></td>
</tr>
</tbody>
</table>
1 What is ECS?

Elastic Compute Service (ECS) is a high-performance, stable, reliable, and scalable IaaS-level service provided by Alibaba Cloud. ECS eliminates the need to invest in IT hardware up front and allows you to quickly scale computing resources on demand, making ECS more convenient and efficient than physical servers. ECS provides a variety of instance types that suit different business needs and help boost business growth.

Why ECS?

ECS provides the following benefits:

- You do not have to purchase hardware or construct data centers.
- Instances can be delivered within minutes, enabling rapid deployment and reducing time to market.
- You can make use of resources in data centers and BGP servers around the world.
- You can scale resources up or down based on your actual business needs at a transparent, easy-to-understand cost.
- x86 architecture-based ECS instances, ECS Bare Metal instances, and heterogenous computing ECS instances such as those with GPUs and FPGAs are provided.
- You can use ECS to access other Alibaba Cloud services over the internal network, reducing Internet traffic costs.
- A host of security solutions such as virtual firewalls, role-based permission control, internal network isolation, virus protection, and traffic throttling are provided.
- ECS comes with a performance monitoring framework and active O&M system.
- A standardized API is provided to improve ease-of-use and applicability.

For more benefits, see Benefits of ECS and Scenarios.

Architecture

ECS comprises the following major components:

- **Instance**: An ECS instance is a virtual server that includes basic computing components such as CPU, memory, operating system, bandwidth, and disks. The computing performance, memory specifications, and applicable scenarios of an instance are
determined by its instance type. Each instance type has particular specifications, including the number of vCPUs, memory capacity, and network performance.

- **Image**: Images provide operating systems, initial application data, and pre-installed software for instances. Multiple Linux distributions and Windows Server operating systems are supported.
- **Block Storage**: Block Storage devices feature high performance and low latency. ECS comes with distributed storage architecture-based cloud disks and physical storage-based local disks.
- **Snapshot**: A snapshot is a stateful data file of a cloud disk at a certain point in time. Snapshots are often used to back up and restore data or to create custom images.
- **Security group**: A security group is a logical grouping of instances located within the same region that have the same security requirements and require access to each other. A security group works as a virtual firewall for the ECS instances inside it.
- **Network**:
  - **Virtual private cloud (VPC)**: A VPC is a logically isolated private cloud network. You can configure a private IP address range, a routing table, and a gateway for a VPC.
  - **Classic network**: All classic network-type instances are built on a shared infrastructure network that is centrally planned and managed by Alibaba Cloud.

For more information, see the Elastic Compute Service page.

The following figure shows the architecture of ECS components. For more information about the functional components in the figure, see the ECS documentation.
Pricing

ECS supports multiple billing methods such as subscription and pay-as-you-go and purchasing options such as reserved instances and preemptible instances. For more information, see Billing overview and the Pricing tab of the Elastic Compute Service page.

Management tools

After registering an Alibaba Cloud account, you can create, use, or release ECS instances in any region by using one of the following methods provided by Alibaba Cloud:

- ECS console: a web service page used to manage ECS instances. For more information, see Quick reference.

- ECS API: an RPC API that supports GET and POST requests. For more information, see API Reference. The following developer tools can be used to call ECS API operations:
  - CLI: a flexible and scalable management tool based on Alibaba Cloud APIs. You can use CLI to encapsulate Alibaba Cloud native APIs to develop custom features.
  - OpenAPI Explorer: allows you to retrieve API operations, call API operations, and dynamically generate SDK sample code.
  - Alibaba Cloud SDK: SDKs for a variety of programming languages such as Java, Python, and PHP are provided.
- **Resource Orchestration Service (ROS):** automatically creates and configures Alibaba Cloud resources based on user-defined templates.

- **Terraform:** an open-source tool that uses configuration files to call computing resources of Alibaba Cloud and other platforms that support Terraform. Terraform also implements version control.

### Deployment suggestions

Before you purchase an ECS instance, consider the following factors:

- **Region and zone**

  A region represents an Alibaba Cloud data center. The region and zone determine the physical location of an ECS instance. After an instance is created, its metadata is established and its region cannot be changed. You can obtain metadata only of the ECS instances located within VPCs. Select a region and zone based on the target geographical location, availability of Alibaba Cloud services, application availability requirements, and whether internal network communication is required. For example, if you want to access both ECS and ApsaraDB for RDS, the ApsaraDB for RDS instance and ECS instance must be within the same region. For more information, see [Regions and zones](#).

- **High availability**

  To ensure business consistency and continuity, we recommend that you use snapshots to back up data, and use multi-zone deployment, deployment sets, and Server Load Balancer (SLB) for disaster recovery.

- **Network planning**

  We recommend that you use VPC to plan your own private IP addresses. VPC supports all new features and new instance types. VPC also supports business system isolation and multi-region system deployment. For more information, see [What is a VPC?](#).
• Security solutions

You can use ECS security groups to control inbound and outbound access policies and the port listening status of ECS instances. For applications deployed on ECS instances, Anti-DDoS Basic are provided for free. Alibaba Cloud Security is also available.

- For example, you can use Anti-DDoS Pro to ensure the stability and reliability of source sites. For more information, see What is Anti-DDoS Pro.
- Security Center can be used to safeguard the security of ECS instances. For more information, see What is Security Center?

Related services

Together with ECS, you can select the following Alibaba Cloud services:

• Auto Scaling: automatically adjusts the number of ECS instances based on business and policy changes. For more information, see What is Auto Scaling?
• Dedicated Host (DDH): allows you to deploy ECS instances on a dedicated host to have dedicated use of its physical resources. DDH also allows you to migrate your businesses to the cloud at minimal costs while meeting compliance requirements. For more information, see What is Dedicated Host (DDH)?.
• Container Service for Kubernetes: manages application lifecycles on groups of ECS instances. For more information, see What is Container Service for Kubernetes?
• Server Load Balancer (SLB): distributes traffic among multiple ECS instances. For more information, see What is Server Load Balancer?
• CloudMonitor: develops monitoring solutions for instances, system disks, and public bandwidth. For more information, see Overview.
• ApsaraDB for RDS: provides database services accessible over internal networks to ECS instances, reduces network latency and access fees, and delivers top-notch performance. ApsaraDB for RDS supports multiple database engines, including MySQL, SQL Server, PostgreSQL, PPAS, and MariaDB. For more information, see What is ApsaraDB for RDS?
• Alibaba Cloud Marketplace: a platform where third-party partners provide software infrastructure, business software, and various software and services related to website construction, hosted O&M, security, data, APIs, and solutions. You can also provide software and services as a service provider in Alibaba Cloud Marketplace.

For more solutions, see Alibaba Cloud Solutions.
2 Benefits of ECS

ECS provides better availability, security, and elasticity than integrated data centers (IDCs) and cloud server providers.

High Availability

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

Each Alibaba Cloud region has multiple zones. You can create active/standby or active/active ECS instances in multiple zones to achieve higher availability. You can build fault tolerant systems across multiple regions and zones to implement a financial-grade solution that spans three data centers across two regions. Alibaba Cloud provides mature solutions for fault tolerant services such as disaster recovery.

The Alibaba Cloud framework allows you to seamlessly switch between services. For more information about industry solutions, see Solutions by industry. 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, such as Elastic Compute Service (ECS), Server Load Balancer (SLB), Relational Database Service (RDS), and Data Transmission Service (DTS).
- Industry partners and ecosystem partners that help you build a more advanced and stable architecture and guarantee service continuity.
- Diverse training services that help you achieve high availability from businesses to underlying services.

Security

Alibaba Cloud has passed a host of international information security certifications, such as ISO 27001 and MTCS, which demands strict confidentiality of user data and user information, as well as user privacy protection.

We recommend that you use ECS in a Virtual Private Cloud (VPC). VPCs provide a stable, secure, controllable network environment that can be delivered in a short period of time. The capability and architecture of VPC hybrid cloud bring the technical advantages of
cloud computing to enterprises in traditional industries that have not implemented cloud computing.

• Breadth of network products
  
  You only need to perform a simple configuration to connect your business environment to global IDCs, making your business more flexible, stable, and extensible.

• Interconnection with your IDC
  
  You can use Express Connect to connect Alibaba Cloud VPC to your IDC to build a hybrid cloud architecture. You can use a variety of hybrid cloud architectures to provide network services and robust networking.

• Stability
  
  After constructing your VPC, you can update your network architecture and obtain new network functions daily to constantly evolve your network infrastructure and ensure your business is always running steadily.

• Security
  
  VPC features traffic isolation and attack isolation to protect your services from cyber attacks. After you build your business in a VPC, the first line of defense is established.

Elasticity

Elasticity is a key benefit of cloud computing. Alibaba Cloud is capable of providing IT resources required by a medium-sized Internet enterprise within a few minutes. In this way, most enterprises that build business on the cloud can process huge business volumes.

Alibaba Cloud provides elastic computing, storage, networking, and business architecture planning and allows you to combine your businesses as needed.

• Elastic computing
  
  - Vertical scaling
    
    Vertical scaling is the process where the configurations of an ECS instance are modified. After you purchase an ECS instance or storage capacity from Alibaba Cloud, you can configure the instance based on your transaction volume, whereas it may
be difficult to change the configurations of a server in a traditional IDC. For more information about vertical scaling, see Overview of configuration changes.

- Horizontal scaling

Horizontal scaling allows the re-division of resources between applications. A traditional IDC may not be able to immediately provide sufficient resources for online gaming or live video streaming applications during peak hours. The elasticity of cloud computing makes it possible to provide the resources required during peak hours. When the load returns to normal levels, you can release unnecessary resources to reduce operation costs. The combination of ECS vertical and horizontal elasticity enables you to scale resources up and down by specific quantities as scheduled or against business load. For more information about horizontal scaling, see #unique_22.

- Elastic storage

In a traditional IDC, you must add servers to increase the storage space. However, the number of servers that you can add is limited. Alibaba Cloud provides unlimited storage capacity and allows you to order as much storage space based on the business requirements. For more information about elastic storage, see Resize a disk.

- Elastic network

Alibaba Cloud provides network elasticity. When purchasing Alibaba Cloud VPCs, you can configure the VPCs in the same way as the IDCs. In addition, VPCs have the following benefits: interconnection between data centers, separate secure domains in data centers, and flexible network configurations and planning within a VPC. For more information about elastic networks, see Virtual Private Cloud.

**Comparison between ECS and traditional IDCs**

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

<table>
<thead>
<tr>
<th>Item</th>
<th>ECS</th>
<th>Traditional IDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equipment room deployment</td>
<td>Provides independently developed DC-powered servers with low PUE.</td>
<td>Provides traditional AC-powered servers with high PUE.</td>
</tr>
<tr>
<td></td>
<td>Provides backbone equipment rooms with high outbound bandwidth and dedicated bandwidth.</td>
<td>Provides equipment rooms with various quality levels and shared bandwidth primarily, difficult for users to choose from.</td>
</tr>
<tr>
<td>Item</td>
<td>ECS</td>
<td>Traditional IDC</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>-----------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Ease of operation</strong></td>
<td>Provides mainstream OSs, including activated Windows OS.</td>
<td>Requires users to purchase and install OSs manually.</td>
</tr>
<tr>
<td></td>
<td>Easily switches between OSs online.</td>
<td>OSs have to be manually reinstalled.</td>
</tr>
<tr>
<td></td>
<td>Provides a Web-based console for online management.</td>
<td>Users must manually perform management and maintenance operations.</td>
</tr>
<tr>
<td></td>
<td>Provides mobile phone verification for password setting, increasing data security.</td>
<td>Brings difficulty in resetting passwords, and exposes high risk of password cracking.</td>
</tr>
<tr>
<td><strong>Disaster recovery and backup</strong></td>
<td>Stores three copies of each piece of data. When one copy is corrupted, the data can be quickly restored.</td>
<td>Users must build a disaster recovery environment by themselves, and use traditional storage devices.</td>
</tr>
<tr>
<td></td>
<td>Users can customize automatic snapshot policies to create automatic snapshots for data recovery.</td>
<td>Does not support automatic recovery because the snapshot function is not provided.</td>
</tr>
<tr>
<td></td>
<td>Hardware failures can be recovered quickly and automatically.</td>
<td>Users must restore corrupted data manually.</td>
</tr>
<tr>
<td><strong>Security and reliability</strong></td>
<td>Effectively prevents MAC spoofing and ARP attacks.</td>
<td>Fails to prevent MAC spoofing and ARP attacks.</td>
</tr>
<tr>
<td></td>
<td>Effectively defends against DDoS attacks by using black holes and traffic scrubbing.</td>
<td>Needs additional costs for devices for traffic scrubbing and black hole shielding systems.</td>
</tr>
<tr>
<td></td>
<td>Provides additional services, such as port scanning, trojan scanning, and vulnerability scanning.</td>
<td>Typically encounters problems such as port scanning, trojan scanning, and vulnerability scanning.</td>
</tr>
<tr>
<td><strong>Flexible scalability</strong></td>
<td>Activates cloud servers on demand and upgrades configurations online.</td>
<td>Needs a long time for server delivery.</td>
</tr>
<tr>
<td></td>
<td>Adjusts outbound bandwidth as required.</td>
<td>Requires one-off purchase of outbound bandwidth that cannot be adjusted.</td>
</tr>
<tr>
<td></td>
<td>Combines with Server Load Balancer online, enabling scaling up applications quickly and easily.</td>
<td>Uses hardware-based server load balancing, which is expensive and difficult to set up.</td>
</tr>
<tr>
<td>Item</td>
<td>ECS</td>
<td>Traditional IDC</td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------------------------</td>
<td>------------------------------------------------------</td>
</tr>
<tr>
<td>Cost effectiveness</td>
<td>Low cost.</td>
<td>High cost.</td>
</tr>
<tr>
<td></td>
<td>Small up-front investment.</td>
<td>Large up-front investment, causing serious waste of resources.</td>
</tr>
<tr>
<td></td>
<td>Provides pay-as-you-go and flexible payment options to allow you to flexibly respond to business changes.</td>
<td>Requires users to purchase up front to meet configuration requirements during peak hours.</td>
</tr>
</tbody>
</table>
3 Scenarios

ECS is a highly flexible solution that can be used independently as a simple web server or an application server, or used with other Alibaba Cloud products to deliver advanced solutions.

Note:
Typical scenarios of ECS are included but not limited to the applications described in this topic.

Official websites and simple web applications

A new website has low traffic and requires only low-configuration ECS instances to run applications, databases, storage files, and other resources. As your business expands, you can upgrade the ECS configuration or add ECS instances at any time without worrying about resource insufficiency during peak traffic.

Multimedia and high-traffic apps or websites

ECS can be used with OSS to store static images, videos, and downloaded packages, reducing the storage fees. In addition, it can work with CDN and Load Balancer to significantly reduce your waiting time, cut down public bandwidth costs, and improve service availability.

Databases with high I/O requirements

ECS supports databases with high I/O requirements, such as OLTP and NoSQL databases. A high-configuration I/O-optimized ECS instance can be used with the ESSD to achieve high level of I/O concurrency with higher data reliability. Alternatively, multiple lower-configuration I/O-optimized ECS instances can be used with Server Load Balancer (SLB) to deliver a high availability architecture.

Apps and websites with sharp traffic fluctuations

Some applications may experience sharp 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 way, you can meet the changing resource requirements at a low cost. ECS can also work with SLB to deliver a high availability architecture.
Big data and real-time online and offline analysis

ECS allows you to use big data instance type families to support Hadoop distributed computing, log processing, and large data warehouses. By adopting the local storage architecture, the big data instance type families can deliver better network performance for Hadoop and Spark clusters while providing abundant storage space with higher storage performance.

AI applications

ECS allows you to use compute optimized and GPU-equipped instance type families to build AI applications with TensorFlow. Such instance type families have lower requirements on the computing capacity of clients and are suitable for image processing, and real-time online rendering for cloud gaming and AR/VR applications.

Customer case

For more scenarios of ECS, see Alibaba Cloud solutions.
4 Regions and zones

This topic provides a complete list of Alibaba Cloud regions and zones.

Regions in Alibaba Cloud are independent. Zones in Alibaba Cloud are completely isolated. However, zones in the same region are connected through links with low latency. The following figure shows the relationship between regions and zones in Alibaba Cloud.

Regions

Alibaba Cloud regions are physical locations (data centers) that spread all over the world to reduce the network latency. The region cannot be changed once a resource is created in it. The following table lists all Alibaba Cloud regions, corresponding cities, and Region IDs.

Note: The available regions vary according to the product. To view the list of regions available for each product, see Alibaba Cloud Global Infrastructure.

- Regions in Mainland China

<table>
<thead>
<tr>
<th>Region name</th>
<th>City</th>
<th>Region ID</th>
<th>Number of zones</th>
</tr>
</thead>
<tbody>
<tr>
<td>China (Qingdao)</td>
<td>Qingdao</td>
<td>cn-qingdao</td>
<td>2</td>
</tr>
<tr>
<td>China (Beijing)</td>
<td>Beijing</td>
<td>cn-beijing</td>
<td>8</td>
</tr>
<tr>
<td>China (Zhangjiakou)</td>
<td>Zhangjiakou</td>
<td>cn-zhangjiakou</td>
<td>3</td>
</tr>
<tr>
<td>China (Hohhot)</td>
<td>Hohhot</td>
<td>cn-huhehaote</td>
<td>2</td>
</tr>
<tr>
<td>China (Hangzhou)</td>
<td>Hangzhou</td>
<td>cn-hangzhou</td>
<td>8</td>
</tr>
<tr>
<td>China (Shanghai)</td>
<td>Shanghai</td>
<td>cn-shanghai</td>
<td>7</td>
</tr>
</tbody>
</table>
Region name | City | Region ID | Number of zones |
---|---|---|---|
China (Shenzhen) | Shenzhen | cn-shenzhen | 5 |
China (Heyuan) | Heyuan | cn-heyan | 2 |
China (Chengdu) | Chengdu | cn-chengdu | 2 |

- Regions outside Mainland China

| Region name | City | Region ID | Number of zones |
---|---|---|---|
China (Hong Kong) | Hong Kong | cn-hongkong | 2 |
Singapore | Singapore | ap-southeast-1 | 3 |
Australia (Sydney) | Sydney | ap-southeast-2 | 2 |
Malaysia (Kuala Lumpur) | Kuala Lumpur | ap-southeast-3 | 2 |
Indonesia (Jakarta) | Jakarta | ap-southeast-5 | 2 |
India (Mumbai) | Mumbai | ap-south-1 | 2 |
Japan (Tokyo) | Tokyo | ap-northeast-1 | 2 |
US (Silicon Valley) | Silicon Valley | us-west-1 | 2 |
US (Virginia) | Virginia | us-east-1 | 2 |
Germany (Frankfurt) | Frankfurt | eu-central-1 | 2 |
UK (London) | London | eu-west-1 | 2 |
UAE (Dubai) | Dubai | me-east-1 | 1 |

When choosing a region, consider the following factors:

- Geographic location

Choose a region based on the geographic locations of you and your target users.

- Regions in Mainland China

In general, we recommend that you choose the data center closest to your target users so that they have fast access. Alibaba Cloud’s data centers of the regions in Mainland China are similar to each other in terms of infrastructure, BGP network quality, service quality, ECS operation and configuration. The regions offer BGP
backbone network lines covering all provinces, autonomous regions, and municipalities to guarantee fast, stable access within Mainland China.

- Regions outside Mainland China

Regions outside Mainland China offer access at international bandwidth, targeting areas outside Mainland China. Users in Mainland China that access services deployed in these regions may experience high latency. Therefore, we do not recommend deploying services in regions outside Mainland China if users from Mainland China require access.

- If you have business operations in China (Hong Kong) or Southeast Asia, choose the region China (Hong Kong), Singapore, Malaysia (Kuala Lumpur), or Indonesia (Jakarta).
- If you have business operations in Japan and Korea, choose the region Japan (Tokyo).
- If you have business operations in India, choose the region India (Mumbai).
- If you have business operations in Australia, choose the region Australia (Sydney).
- If you have business operations in America, choose the region US (Silicon Valley) or US (Virginia).
- If you have business operations in Continental Europe, choose the region Germany (Frankfurt).
- If you have business operations in the Middle East, choose the region UAE (Dubai).

- Intranet communication between Alibaba Cloud products across regions

Intranet communication between Alibaba Cloud products that are not in the same region is not supported, which means that ECS instances and other products, such as ApsaraDB for RDS and OSS instances, cannot communicate with one another over the intranet if they are in different regions.

- Resource pricing

The price of resources may vary according to the region. For details, see the product pricing page.
• Business license filing

When you choose a region, consider the special requirements of some areas. If you have purchased an ECS instance in Mainland China and used it for a web server, complete business license filing first.

To complete business license filing, pay attention to the following points:

- If your company is located in Beijing, choose the **China (Beijing)** region for the ECS instance you have purchased.

- If your company is located in Guangdong, choose the **China (Shenzhen)** region for the ECS instance you have purchased.

**Note:**
The communications administrations in different provinces, autonomous regions, and municipalities have different approval requirements for business license filing. In case of any change, the information published on the business license filing website of your local administration prevails.

**Zones**

Zones are physical areas with independent power grids and networks within one region. The network latency for instances within the same zone is lower.

Intranet communication can take place between instances in different zones of the same region, and fault isolation can be achieved between zones. Whether or not instances can be deployed in the same zone depends on the requirements for disaster recovery and network latency.

- If your applications require high disaster recovery capabilities, we recommend that you deploy your instances in different zones of the same region.

- If your applications require low network latency between instances, we recommend that you create your instances in the same zone.

For more information about regions and zones, see the **Alibaba Cloud Global Infrastructure**.
5 Limits

This topic describes the limits of ECS and how to apply for exceptions to some limits.

Overview

ECS has the following limits:

- You cannot install virtualization software such as VMware Workstation, or use it for secondary virtualization. Only ECS Bare Metal Instances and Super Computing Clusters (SCCs) support secondary virtualization.
- ECS does not support sound card applications.
- External hardware devices such as hardware dongles, USB flash drives, external hard disks, and hardware tokens, cannot be directly attached to ECS. Software verification methods such as two-factor authentication and dynamic passwords can be used.
- ECS does not support multicast protocols. We recommend that you use unicast transmission instead.
- Log Service does not support 32-bit Linux ECS instances.

For information about the ECS instances that support Log Service, see #unique_32.

- If you need to apply for ICP filing for websites that are deployed on your ECS instance, the instance must meet certain purchase requirements. You can apply for a limited number of ICP filing service numbers for websites that are deployed on each ECS instance. For more information, see #unique_33. For information about how to apply for ICP filing, see #unique_34.

View quotas

You can click Privileges & Quota on the Overview page of the ECS console, and select a region to view the resource usage and quotas in that region. If the quota of a resource is insufficient, submit a ticket to apply for an increased quota. For more information about how to view privileges and quotas, see #unique_35 or #unique_36.
## Instance limits

<table>
<thead>
<tr>
<th>Item</th>
<th>Limit</th>
<th>How to apply for an exception to the limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Permissions to create an ECS instance</td>
<td>To create an ECS instance in mainland China regions, you must first complete real-name verification.</td>
<td>None</td>
</tr>
<tr>
<td>Instance types whose instances can be used as pay-as-you-go instances</td>
<td>Instance types with less than 16 vCPUs</td>
<td>Submit a ticket</td>
</tr>
<tr>
<td>Quota for vCPUs of all pay-as-you-go instances in each region for an account</td>
<td>50 vCPUs</td>
<td>Submit a ticket</td>
</tr>
<tr>
<td>Quota for vCPUs of all preemptible instances in each region for an account</td>
<td>You must submit a ticket to apply for the quota, which is up to 50 vCPUs.</td>
<td>Submit a ticket</td>
</tr>
<tr>
<td>Total number of instance launch templates in each region for an account</td>
<td>30</td>
<td>None</td>
</tr>
<tr>
<td>Total number of versions of an instance launch template</td>
<td>30</td>
<td>None</td>
</tr>
<tr>
<td>Switching of the billing method from pay-as-you-go to subscription</td>
<td>Not supported by the following instance families: t1, s1, s2, s3, c1, c2, m1, m2, n1, n2, and e3.</td>
<td>None</td>
</tr>
<tr>
<td>Switching of the billing method from subscription to pay-as-you-go</td>
<td>• Whether this feature is supported depends on your ECS usage.</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>• 5,000 vCPUs × Hours per month.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Switching the billing method may result in a refund. Each account is limited to a maximum monthly refund amount. For more information, see Switch the billing method from subscription to pay-as-you-go.</td>
<td></td>
</tr>
</tbody>
</table>
Reserved instance limits

<table>
<thead>
<tr>
<th>Item</th>
<th>Limit</th>
<th>How to apply for an exception to the limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of regional reserved instances of an account</td>
<td>20</td>
<td>Submit a ticket</td>
</tr>
<tr>
<td>Total number of zonal reserved instances of an account in a zone</td>
<td>20</td>
<td>Submit a ticket</td>
</tr>
<tr>
<td>Instance types that support reserved instances</td>
<td>Instance types in the following instance families: sn1ne, sn2ne, se1ne, ic5, c5, g5, r5, c6, g6, r6, i2, i2g, hfc5, hfg5, and t5.</td>
<td>None</td>
</tr>
</tbody>
</table>

**Note:**
The t5 instance family supports only zonal reserved instances.

Note:
For more information, see the "Limits" section in Reserved instance overview.

Block Storage limits

<table>
<thead>
<tr>
<th>Item</th>
<th>Limit</th>
<th>How to apply for an exception to the limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Permissions to create a pay-as-you-go disk</td>
<td>You must complete real-name verification before you can create a disk in mainland China regions.</td>
<td>None</td>
</tr>
<tr>
<td>Total number of pay-as-you-go disks in all regions for an account</td>
<td>Number of pay-as-you-go instances in all regions for an account × 5. You can create at least 10 pay-as-you-go disks for an account.</td>
<td>Submit a ticket</td>
</tr>
<tr>
<td>Item</td>
<td>Limit</td>
<td>How to apply for an exception to the limit</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>Capacity of pay-as-you-go disks that are used as data disks for an account</td>
<td>This limit is subject to ECS resource usage, region, and disk category. You can go to the Privileges &amp; Quotas page of the ECS console to view this information. For more information, see #unique_35.</td>
<td>Submit a ticket</td>
</tr>
<tr>
<td>Total number of system disks for each instance</td>
<td>1</td>
<td>None</td>
</tr>
<tr>
<td>Total number of data disks for each instance</td>
<td>16 (including cloud disks and Shared Block Storage devices)</td>
<td>None</td>
</tr>
<tr>
<td>Number of instances to which a Shared Block Storage device can be attached</td>
<td>8</td>
<td>None</td>
</tr>
<tr>
<td>Total number of Shared Block Storage devices in all regions for an account</td>
<td>10</td>
<td>Submit a ticket</td>
</tr>
<tr>
<td>Capacity of a basic disk</td>
<td>5 GiB to 2,000 GiB</td>
<td>None</td>
</tr>
<tr>
<td>Capacity of a standard SSD</td>
<td>20 GiB to 32,768 GiB</td>
<td>None</td>
</tr>
<tr>
<td>Capacity of an ultra disk</td>
<td>20 GiB to 32,768 GiB</td>
<td>None</td>
</tr>
<tr>
<td>Capacity of an enhanced SSD</td>
<td>20 GiB to 32,768 GiB</td>
<td>None</td>
</tr>
<tr>
<td>Capacity of a local SSD</td>
<td>5 GiB to 800 GiB</td>
<td>None</td>
</tr>
<tr>
<td>Total capacity of all local SSDs for an instance</td>
<td>1,024 GiB</td>
<td>None</td>
</tr>
<tr>
<td>Capacity of a local NVMe SSD</td>
<td>1,456 GiB</td>
<td>None</td>
</tr>
<tr>
<td>Total capacity of all local NVMe SSDs for an instance</td>
<td>2,912 GiB</td>
<td>None</td>
</tr>
<tr>
<td>Capacity of a local SATA HDD</td>
<td>5,500 GiB</td>
<td>None</td>
</tr>
<tr>
<td>Total capacity of all local SATA HDDs for an instance</td>
<td>154,000 GiB</td>
<td>None</td>
</tr>
<tr>
<td>Capacity of a Shared SSD Block Storage device</td>
<td>32,768 GiB</td>
<td>None</td>
</tr>
<tr>
<td>Item</td>
<td>Limit</td>
<td>How to apply for an exception to the limit</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>--------------------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>Total capacity of all Shared SSD Block Storage devices for an instance</td>
<td>128 TiB</td>
<td>None</td>
</tr>
<tr>
<td>Capacity of a Shared Ultra Block Storage device</td>
<td>32,768 GiB</td>
<td>None</td>
</tr>
<tr>
<td>Total capacity of all Shared Ultra Block Storage devices for an instance</td>
<td>128 TiB</td>
<td>None</td>
</tr>
</tbody>
</table>
| Capacity of a system disk                                           | • Windows Server: 40 GiB to 500 GiB  
• CoreOS and FreeBSD: 30 GiB to 500 GiB  
• Linux systems excluding CoreOS: 20 GiB to 500 GiB | None                                       |
| Attachment of new local disks to instances that are equipped with local disks | Not allowed        | None                                       |
| Change of specifications of instances that are equipped with local disks | Only bandwidth can be changed | None                                       |
| System disk mount points                                            | /dev/vda           | None                                       |
| Data disk mount points                                              | /dev/vd[b-z]       | None                                       |

**Note:**
Block storage capacity is measured in binary units. 1 KiB equals 1,024 bytes. 1 GiB equals 1,024 MiB.

**Snapshot limits**

<table>
<thead>
<tr>
<th>Item</th>
<th>Limit</th>
<th>How to apply for an exception to the limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of manual snapshots that can be created for each disk or Shared Block Storage device</td>
<td>256</td>
<td>None</td>
</tr>
</tbody>
</table>
### Limits

<table>
<thead>
<tr>
<th>Item</th>
<th>Limit</th>
<th>How to apply for an exception to the limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of automatic snapshots that can be created for each disk or Shared Block Storage device</td>
<td>1,000</td>
<td>None</td>
</tr>
<tr>
<td>Number of automatic snapshot policies that can be created in a region for an account</td>
<td>100</td>
<td>None</td>
</tr>
</tbody>
</table>

### Image limits

<table>
<thead>
<tr>
<th>Item</th>
<th>Limit</th>
<th>How to apply for an exception to the limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of custom images that can be created in each region for an account</td>
<td>100</td>
<td>Submit a ticket</td>
</tr>
<tr>
<td>Maximum number of users to whom a single image can be shared</td>
<td>50</td>
<td>Submit a ticket</td>
</tr>
<tr>
<td>Restrictions on images and instance types</td>
<td>Instance types that have 4 GiB or larger memory do not support 32-bit images.</td>
<td>None</td>
</tr>
</tbody>
</table>

### SSH key pair limits

<table>
<thead>
<tr>
<th>Item</th>
<th>Limit</th>
<th>How to apply for an exception to the limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of SSH key pairs in each region for an account</td>
<td>500</td>
<td>None</td>
</tr>
<tr>
<td>Instance types that support SSH key pairs</td>
<td>All instance types except non-I/O optimized instance types in Generation I</td>
<td>None</td>
</tr>
<tr>
<td>Images that support SSH key pairs</td>
<td>Linux images only</td>
<td>None</td>
</tr>
</tbody>
</table>
### Public bandwidth limits

<table>
<thead>
<tr>
<th>Item</th>
<th>Limit</th>
<th>How to apply for an exception to the limit</th>
</tr>
</thead>
</table>
| Peak inbound bandwidth              | • If the purchased peak outbound bandwidth is less than or equal to 10 Mbit/s, Alibaba Cloud will allocate an inbound bandwidth of 10 Mbit/s.  
• If the purchased peak outbound bandwidth is greater than 10 Mbit/s, Alibaba Cloud will allocate an inbound bandwidth that is equal to the purchased peak outbound bandwidth. | None                                      |
| Peak outbound bandwidth             | • Subscription: 200 Mbit/s  
• Pay-as-you-go: 100 Mbit/s | None                                      |
| Change in the assigned public IP address for an instance | The public IP address can be changed within six hours after the instance is created, and can be changed a maximum of three times. | None                                      |

### Security group limits

<table>
<thead>
<tr>
<th>Item</th>
<th>Basic security group</th>
<th>Advanced security group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of security groups that can be created in each region for an account</td>
<td>100</td>
<td>Same as the limit on basic security groups.</td>
</tr>
<tr>
<td>Number of classic network-type ECS instances that can be included in a classic network-type security group</td>
<td>1,000*</td>
<td>The classic network is not supported.</td>
</tr>
<tr>
<td>Item</td>
<td>Basic security group</td>
<td>Advanced security group</td>
</tr>
<tr>
<td>---------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td>Number of VPC-type ECS instances that can be included in a VPC-type security group</td>
<td>Depends on the number of private IP addresses that can be included in the VPC-type security group.</td>
<td>No limit</td>
</tr>
<tr>
<td>Number of security groups to which an ECS instance can belong</td>
<td>5</td>
<td>Same as the limit on basic security groups.</td>
</tr>
<tr>
<td>Number of security groups to which an ENI of an ECS instance can belong</td>
<td>5</td>
<td>Same as the limit on basic security groups.</td>
</tr>
<tr>
<td>Maximum number of rules (both inbound and outbound) in a security group</td>
<td>200***</td>
<td>Same as the limit on basic security groups.</td>
</tr>
<tr>
<td>Maximum number of rules (both inbound and outbound) in all security groups to which an ENI belongs</td>
<td>1,000</td>
<td>Same as the limit on basic security groups.</td>
</tr>
<tr>
<td>Number of private IP addresses that can be included in a VPC-type security group</td>
<td>2,000**</td>
<td>65,536</td>
</tr>
<tr>
<td>Public network access port</td>
<td>The default SMTP port for outbound traffic is port 25, which is disabled by default. It cannot be enabled by security group rules.</td>
<td>Same as the limit on basic security groups.</td>
</tr>
</tbody>
</table>

* If more than 1,000 classic network-type instances need mutual access over an internal network, you can assign them to multiple security groups and authorize mutual access among these security groups.

** If more than 2,000 instances with private IP addresses need mutual access over an internal network, you can assign them to multiple security groups and authorize mutual access among these security groups.

*** If you increase the number of security groups to which an ECS instance can belong, the maximum number of rules in each security group will decrease. The product of the number of security groups to which an ECS instance belongs and the maximum number of rules in each security group must be less than or equal to 1,000. For example, if the number
of security groups to which the ECS instance belongs is 5, 10, or 16, the corresponding maximum number of rules in each security group is 200, 100, or 60, because $5 \times 200 = 1000$, $10 \times 100 = 1000$, and $16 \times 60 \leq 1000$.

**Deployment set limits**

<table>
<thead>
<tr>
<th>Item</th>
<th>Limit</th>
<th>How to apply for an exception to the limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of deployment sets in each region for an account</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>Total number of instances that can be included in a deployment set</td>
<td>A maximum of seven instances are allowed in a zone. The number of instances allowed in a region is calculated as follows: $7 \times$ Number of zones in the region.</td>
<td>None</td>
</tr>
<tr>
<td>Instance types that can be created in a deployment set</td>
<td>Instance types in the following instance families: c5, d1, d1ne, g5, hfc5, hfg5, i2, ic5, r5, se1ne, sn1ne, and sn2ne.</td>
<td>None</td>
</tr>
</tbody>
</table>

**Cloud Assistant limits**

<table>
<thead>
<tr>
<th>Item</th>
<th>Limit</th>
<th>How to apply for an exception to the limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of Cloud Assistant scripts that can be created in each region for an account</td>
<td>100</td>
<td>Submit a ticket</td>
</tr>
<tr>
<td>Total number of Cloud Assistant scripts that can be used by an account each day in a region</td>
<td>5,000</td>
<td>Submit a ticket</td>
</tr>
</tbody>
</table>
### ENI limits

<table>
<thead>
<tr>
<th>Item</th>
<th>Limit</th>
<th>How to apply for an exception to the limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quota of ENIs in each region for an account</td>
<td>100</td>
<td>Submit a ticket</td>
</tr>
</tbody>
</table>

### Tag limits

<table>
<thead>
<tr>
<th>Item</th>
<th>Limit</th>
<th>How to apply for an exception to the limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of tags that can be bound to an instance</td>
<td>20</td>
<td>None</td>
</tr>
</tbody>
</table>

### API call limits

<table>
<thead>
<tr>
<th>Item</th>
<th>Limit</th>
<th>How to apply for an exception to the limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of CreateInstance calls</td>
<td>200 calls per minute</td>
<td>Submit a ticket</td>
</tr>
</tbody>
</table>

**Note:**
For more information about the limits on VPCs, see #unique_41.
6 Usage notes

This topic describes the general limits for the usage of ECS instances and limits that are specific to the installed OS.

General limits

- After you create an ECS instance, only you have administrator permissions on this instance.
- You are not allowed to resell or sublicense any bandwidth assigned to your ECS instances without authorization. Failure to comply will result in a suspension or termination of your right to use any portions or all Alibaba Cloud services.
- Do not use your ECS instances for malicious, fraudulent, or unlawful activities, such as click farming or fraudulent transactions, on e-commerce websites such as Taobao. Failure to comply will result in suspension or termination of your account.
- Do not uninstall relevant hardware drivers.
- Do not modify the MAC address of ENIs unless necessary.
- Do not enable SELinux.
- An ECS instance with more than 4 GiB RAM must use a 64-bit operating system (OS). The 64-bit OSs currently supported are as follows:
  - Aliyun Linux 64-bit
  - CoreOS 64-bit
  - CentOS 64-bit
  - Debian 64-bit
  - FreeBSD 64-bit
  - OpenSUSE 64-bit
  - SUSE Linux 64-bit
  - Ubuntu 64-bit
  - Red Hat 64-bit
  - Windows 64-bit

For the latest available OSs, go to the instance purchase page in the ECS console.
- To guarantee service continuity, we recommend that you enable automatic startup upon instance boot for relevant software. If service applications are connected to databases, we recommend that you enable automatic re-connection for these service applications.
• We recommend that you do not upgrade the kernel and the operating system. If you need to upgrade the kernel, see How to avoid Linux instance startup failure after kernel upgrade.

**Windows limits**

• Do not stop the built-in AliyunService or shutdownmon.exe process. Otherwise, ECS instances may not be stopped or restarted properly.
• Do not modify the hostname of the domain controller.
• We recommend that you do not create custom images by using a virtual machine that acts as a domain controller.
• Do not rename, delete, or disable the administrator account.
• We recommend that you do not use the virtual memory if Basic Cloud Disks are used. For ultra disks, standard SSDs, or ESSDs, you can use the virtual memory as needed.
• Exercise caution when you use the administrator account to resize cloud disks, operate spanned volumes or the registry, and update the system. Failure to comply can result in data loss. For more information, see Resize a cloud disk.
• Windows 32-bit OSs support up to four CPU cores.
• Ensure that a minimum of 2 GiB RAM is available when you build a website or deploy a Web environment on a Windows instance. Instances with 1-core CPU and 1 GiB RAM cannot start MySQL.
• For more information see Select an image.

**Linux limits**

• Do not modify the content of the default /etc/issue file on Linux instances. Otherwise, if you create a custom image of such an instance and then use the image to create a new instance, the new instance cannot start properly because the operating system edition cannot be recognized.
• Do no modify permissions of the directories in the root partition, especially /etc, /sbin, /bin, /boot, /dev, /usr, and /lib. Improper modifications of permissions may cause system errors.
• Do not rename, delete, or disable the root account.
• Do not compile the Linux kernel or perform any other operations on it.
• We recommend that you do not use the swap partition if Basic Cloud Disks are used. However, for ultra disks, standard SSDs, or ESSDs, you can use the swap partition as needed.
• Exercise caution when you use the root account to run fio, mkfs, or fsck commands or resize cloud disks. Failure to comply can result in data loss.

• For more information see Select an image.