Alibaba Cloud

Server Load Balancer

Instance

Document Version: 20220627

⟨-⟩ Alibaba Cloud

Legal disclaimer

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

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

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

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

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

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

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

Document conventions

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

Table of Contents

1.Overview

Classic Load Balancer (CLB) instances receive requests from clients and forward requests to backend servers. To use the load balancing service, you must create a CLB instance and add listeners and backend servers to the instance.
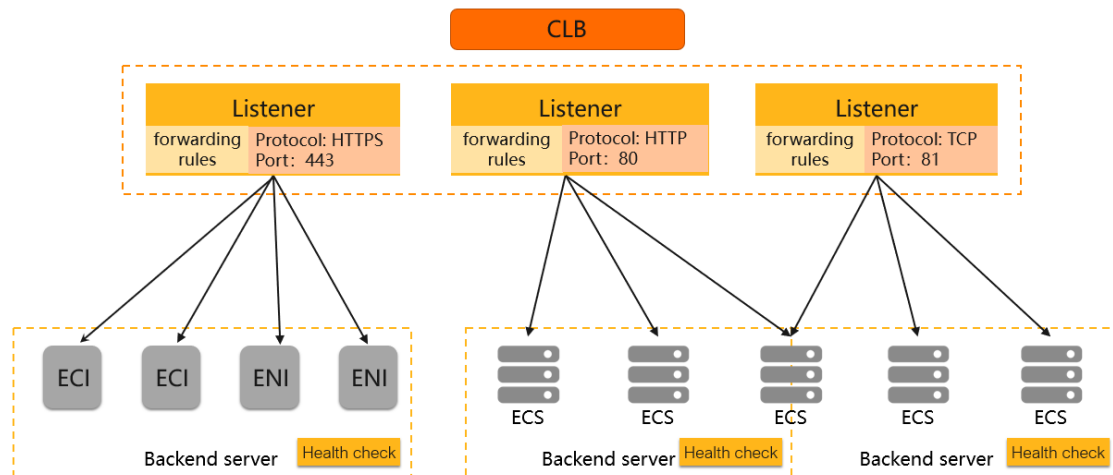
Instance types

Alibaba Cloud provides Internet-facing and internal-facing CLB instances.

Internet-facing CLB instances

When you create an Internet-facing CLB instance, the system automatically assigns a public address to the CLB instance. You can associate your domain name with the public IP address. CLB instances receive requests from clients over the Internet and forward requests to backend servers based on the forwarding rules that you specify for listeners.

An Internet-facing CLB instance has the following features:

- The system assigns a public IP address to the CLB instance. You cannot disassociate the public IP address from the CLB instance.
- When the subscription billing method is used, only the pay-by-bandwidth metering method is supported. When the pay-as-you-go billing method is used, the pay-by-bandwidth and pay-by-data-transfer metering methods are supported.

Internal-facing CLB instances

Internal-facing CLB instances provide services over private networks. Internal-facing CLB instances receive requests from internal networks and forward requests to backend servers based on listener rules.

Internal-facing CLB instances that are associated with elastic IP addresses (EIPs) can process requests from the Internet. An internal-facing CLB instance has the following features when it provides services over the Internet:

- An internal-facing CLB instance associated with an EIP can provide services over the Internet. You can disassociate the EIP from the CLB instance as needed.
- An EIP that is associated with an EIP bandwidth plan supports the pay-by-95th-percentile-bandwidth billing method in addition to the subscription and pay-as-you-go billing methods.

The network types supported by internal-facing CLB instances vary based on billing methods.

- Subscription internal-facing CLB instances support the following network types: classic network and virtual private cloud (VPC).
  - VPC

    If an internal-facing CLB instance is deployed in a VPC, the IP address of the CLB instance is allocated from the CIDR block of the vSwitch that is attached to the VPC. The internal-facing CLB instance can be accessed by only Elastic Compute Service (ECS) instances in the same VPC.
  - Classic network

    If the internal-facing CLB instance is deployed in a classic network, the private IP address of the CLB instance is allocated and managed by Alibaba Cloud. The internal-facing CLB instance can be accessed by only ECS instances in classic networks.

    > **Notice** Internal-facing CLB instances of the classic network type are no longer available.

- Pay-as-you-go internal-facing CLB instances support only the VPC network type.

  Internal-facing CLB instances support PrivateLink. Internal-facing CLB instances can receive requests from other VPCs through PrivateLink connections and distribute the requests to backend servers based on listener rules.

Instance types and specifications

> **Note** If you require a higher QPS, you can purchase Application Load Balancer (ALB) instances.

Resources are shared among all shared-resource CLB instances. Therefore, the performance of shared-resource CLB instances is not guaranteed.

> **Notice** Shared-resource CLB instances are no longer available for purchase.

Feature	High-performance CLB instance	Shared-resource CLB instance
Resource allocation	Exclusive resources	Shared resources
Service uptime guaranteed by the service-level agreement (SLA)	99.95%	Not supported
IPv6	√	-
Server Name Indication (SNI)	√	-
Blacklists and whitelists	√	-
Elastic network interface (ENI) mounting	√	-
Assigning secondary IP addresses to ENIs that are associated with ECS instances	√	-
HTTP-to-HTTPS redirection	√	-

Feature	High-performance CLB instance	Shared-resource CLB instance
Consistent hashing	√	-
TLS security policies	√	-
HTTP2	√	-
WebSocket or WebSocket Secure	√	-

? **Note**   In the preceding table, "√" indicates that a feature is supported, and "-" indicates that a feature is not supported.

2.Network traffic flow

Classic Load Balancer (CLB) uses CLB clusters to forward client requests to backend servers and receives responses from backend servers over the internal network.

Inbound network traffic flow

CLB distributes incoming traffic according to forwarding rules configured in the console or by using API operations. The following figure shows the inbound network traffic flow.

Inbound network traffic flow
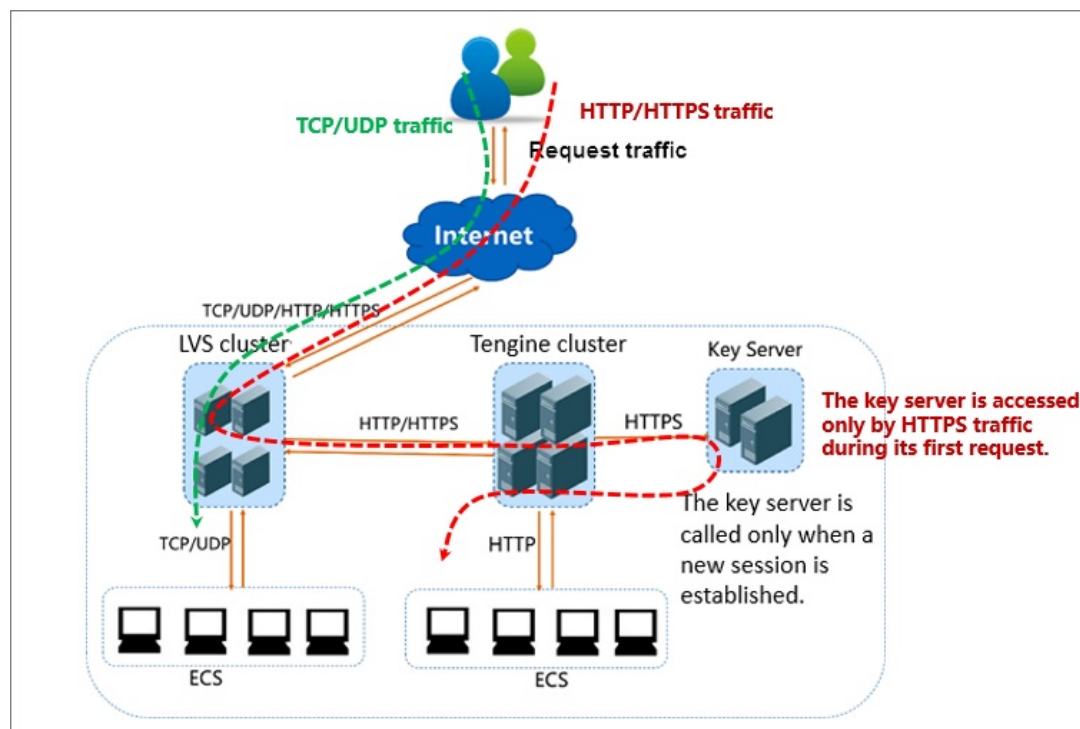
1. For TCP, UDP, HTTP, and HTTPS protocols, the incoming traffic must be forwarded through the LVS cluster first.

2. The large amounts of inbound traffic is distributed evenly among all node servers in the LVS cluster, and the node servers synchronize sessions to ensure high availability.

   ○ For Layer 4 listeners (either UDP or TCP), the node servers in the LVS cluster distribute requests directly to backend ECS instances according to the configured forwarding rules.

   ○ For Layer 7 listeners that use the HTTP frontend protocol, the node servers in the LVS cluster first distribute requests to the Tengine cluster. Then, the node servers in the Tengine cluster distribute the requests to backend ECS instances according to the configured forwarding rules.

   ○ For Layer 7 listeners that use the HTTPS frontend protocol, the request distribution is similar to the HTTP protocol except that the system calls the Key Server to validate certificates and decrypt data packets before requests are distributed to backend ECS instances.

Outbound network traffic flow

CLB communicates with backend ECS instances by using the internal network.

● If backend ECS instances handle the traffic distributed only from CLB, no public bandwidth (ECS

---

instances, EIPs, NAT gateways, and public IP addresses) is required, and you do not need to purchase public bandwidth.

> ⑦ **Note**   Previously created ECS instances are directly allocated with public IP addresses. You can view the public IP addresses by using the ipconfig command. If these ECS instances process requests only through CLB, no traffic fees are incurred for Internet traffic even if traffic statistics are read at the public network interface (NIC).

- If you want to provide external services through backend ECS instances, or backend ECS instances need to access the Internet, you must configure at least one of the following services: an ECS instance, a public IP address, an Elastic IP Address (EIP), or a NAT gateway.

The following figure shows the outbound network traffic flow.

Outbound network traffic flow
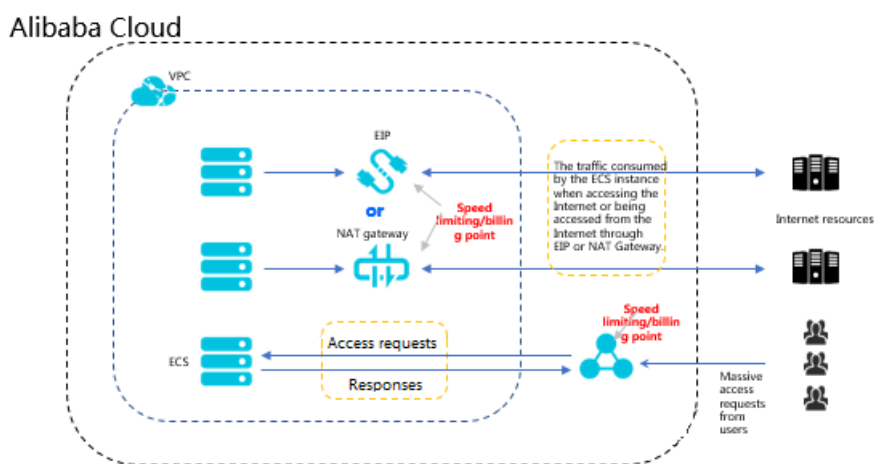
General principle: The traffic goes out from where it comes in.

- Outbound traffic from CLB instances is charged, and the speeds at which it is sent depend on the current network capacity. However, you are not charged for inbound traffic and internal communications between CLB instances and backend ECS instances.

- Traffic from an EIP or from a NAT gateway incurs charges. The speed at which EIP or NAT gateway traffic is sent depends on the current network capacity. If you configure public bandwidth when you purchase an ECS instance, traffic is sent at speeds dependent on the current network capacity, and is charged.

- CLB supports dynamic access to the Internet. Specifically, if a backend ECS instance needs to access the Internet, you must first configure public bandwidth for it, and purchase an EIP or a NAT gateway.

- The public bandwidth that you configured when you create ECS instances, EIPs, and NAT gateways allow ECS instances to access the Internet or be accessed from the Internet, but they cannot forward traffic or balance traffic loads.

3.Overview of IPv6 CLB instances

You can create IPv6 instances. After you create an IPv6 CLB instance, the system allocates a public IPv6 address to the CLB instance. The public IPv6 address is used to receive requests from IPv6 clients.

Context

IPv6 is short for Internet Protocol Version 6. IPv6 is the next-generation Internet Protocol (IP) designed by Internet Engineering Task Force (IETF) to replace IPv4, which is the current version of IP. IPv6 extends the length of an IPv4 address from 32 bits to 128 bits. This means that the address space is expanded to 2^96 times the original size. Therefore, IPv6 ensures that each instance can be allocated an IP address.

Features of IPv6 instances

- Switching from IPv4 to IPv6 without negative impacts on your business

  - You can add IPv4 Elastic Compute Service (ECS) instances to an IPv6 instance to receive IPv6 requests without the need to modify your business system.

  - When your existing IPv4 CLB instances are unable to withstand the growing user traffic, you can create IPv6 CLB instances. This way, clients can access your CLB instances by using IPv4 or IPv6. In addition, this does not adversely affect your business system.

- More secure and reliable workload deployment

  - You can configure a blacklist for CLB to block network traffic from malicious IP addresses.

  - You can also configure a whitelist for CLB to allow network traffic only from specified IP addresses.

Limits

- The IPv6 network environment of Alibaba Cloud is in the initial stage of development. If you cannot connect to an IPv6 service, submit a ticket for technical support. Alibaba Cloud does not provide service level agreement (SLA) guarantees during the public preview.

- IPv6 packets have longer IP headers than IPv4 packets. When an IPv6 instance uses a UDP listener, make sure that the following requirement is met: The maximum transmission unit (MTU) supported by the elastic network interface (ENI) that each backend server uses to communicate with does not exceed 1,200 bytes. Otherwise, oversized packets may be discarded. You must modify the MTU settings in the configuration files of some applications accordingly. TCP supports the Maximum Segment Size (MSS) announcement. Therefore, you do not need to modify the configurations of applications if a CLB instance uses a TCP listener.

- HTTP listeners can use the `X-Forwarded-For` header field to obtain IPv6 addresses of clients.

Related information

- Create a CLB instance

4.Create a CLB instance

A Classic Load Balancer (CLB) instance is an entity that provides load balancing services. To use the service, you must first create a instance.

Prerequisites

Before you create a instance, make sure that you have made the required preparations. For more information, see Preparations.

Procedure

1.
2. On the **Instances** page, click **Create CLB**.
3. Configure the CLB instance based on the following information.

Parameter	Description
Region	Select the region where you want to create the CLB instance. ⊘ **Note** Make sure that the CLB instance and the ECS instances that act as backend servers are deployed in the same region.
Zone Type	Select a zone type for the instance. A zone represents a data center that contains a set of independent infrastructure resources. The infrastructure resources (such as networks, power supply, and air-conditioning) in different zones are independent of each other. Therefore, when the infrastructure resources in one zone are down, the other zones are not affected. Each zone belongs to a specific region. A region may contain one or more zones. Multi-zone instances are available in most regions. ○ Single zone: The CLB instance is deployed in only one zone. ○ Multi-zone: The CLB instance is deployed across two zones. By default, the primary zone is used to distribute network traffic. When the primary zone is down, the system automatically switches to the secondary zone and continues to provide load balancing services. This ensures high availability of the service.
Primary Zone	Select a primary zone for the CLB instance. The primary zone distributes traffic in normal conditions.
Backup Zone	Select a secondary zone for the CLB instance. By default, the secondary zone distributes traffic only if the primary zone is down.
Instance name	Enter a name for the CLB instance. The name must be 1 to 80 characters in length, and can contain letters, digits, hyphens (-), forward slashes (/), periods (.), and underscores (_).
LoadBalancerSpec	Select a specification for the CLB instance. The performance metrics of CLB instances vary based on the specifications. For more information, see Overview.

Parameter	Description
Instance Type	You can create an Internet-facing CLB instance or an internal-facing CLB instance based on your business requirements. The system allocates a public or private IP address to the CLB instance based on the specified instance type. **Public Network** is selected in this example. ○ Public Network: If you create an Internet-facing CLB instance, a public IP address is allocated to the CLB instance. In this case, the CLB instance can provide services over the Internet. ○ Internal Network: If you create an internal-facing CLB instance, a private IP address is allocated to the CLB instance. You can access the CLB instance only within the networks of Alibaba Cloud. You cannot access the CLB instance over the Internet. For more information, see Overview.
IP Version	Select the IP protocol that is used by the CLB instance. You can select IPv4 or IPv6. For more information about the regions and zones that support IPv6 CLB instances, see Overview of IPv6 CLB instances.
Billing Method	Select a billing method.
Quantity	Specify the number of CLB instances that you want to purchase.
Resource Group	Select the resource group to which the cloud resource belongs.

4. Click **Buy Now** and complete the payment.

Related information
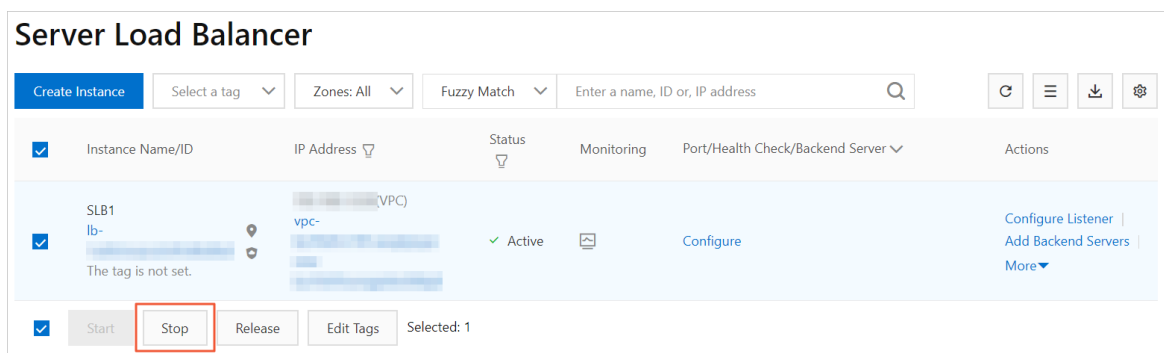
- CreateLoadBalancer

5.Start or stop an SLB instance

This topic describes how to start or stop a Server Load Balancer (SLB) instance. You can start or stop SLB instances at any time. A stopped SLB instance does not receive or forward client traffic.

Procedure

1. Log on to the SLB console.

2. In the left-side navigation pane, choose **Instances > Instances**.

3. Select a region and find the SLB instance that you want to start or stop.

4. In the corresponding **Actions** column, choose **More > Start** or **More > Stop**.

5. To start or stop multiple instances at a time, select the instances and click **Start** or **Stop** at the lower part of the page.

Related information

- SetLoadBalancerStatus

6.Associate an EIP with a CLB instance

This topic describes how to associate an elastic IP address (EIP) with an internal-facing instance deployed in a virtual private cloud (VPC). After you associate an EIP with an internal-facing instance, the CLB instance can forward requests from the Internet.

Context

An EIP is a public IP address that you can purchase and use as an independent resource. You can associate an EIP with an Elastic Compute Service (ECS) instance, an internal-facing instance in a VPC, or an Internet NAT gateway. For more information, see What is an EIP?.

Procedure

1.

2. In the top navigation bar, select the region where the instance that you want to manage is deployed and find the CLB instance.

> 🔊 **Notice** The instance to be associated with an EIP must meet the following requirements:
>   ○ The instance must be deployed in a VPC.
>   ○ The instance and the EIP must belong to the same region.
>   ○ You can associate a instance with only one EIP.

3. In the **Actions** column of the instance, choose ⋮ **> Bind EIP**.

4. In the **Associate EIP** panel, set the following parameters and click **OK**.

Parameter	Description
IP Address Type	Select **EIP**.
IP address	Select an EIP from the drop-down list. If no EIP is available, click **Purchase EIP** and follow the instructions on the buy page to purchase an EIP. For more information, see Apply for an EIP.

> ⑦ **Note** When you associate an EIP with an internal-facing instance, your backend service associated with the CLB instance will be temporarily interrupted. Therefore, we recommend that you perform this operation during off-peak hours or associate the service with another CLB instance first.

7.Tags

7.1. Tag overview

This topic provides an overview of tags in SLB. SLB provides the tag management feature that allows you to classify SLB instances by using tags.

Each tag consists of a key and a value. Before you use tags, note the following limits:

- Tags must be added to SLB instances.
- Each SLB instance can have a maximum of ten tags.
- The key of each tag added to an SLB instance must be unique. If a tag with the same key already exists, the tag is overwritten with the new value.

7.2. Add tags

This topic describes how to add tags to a CLB instance.
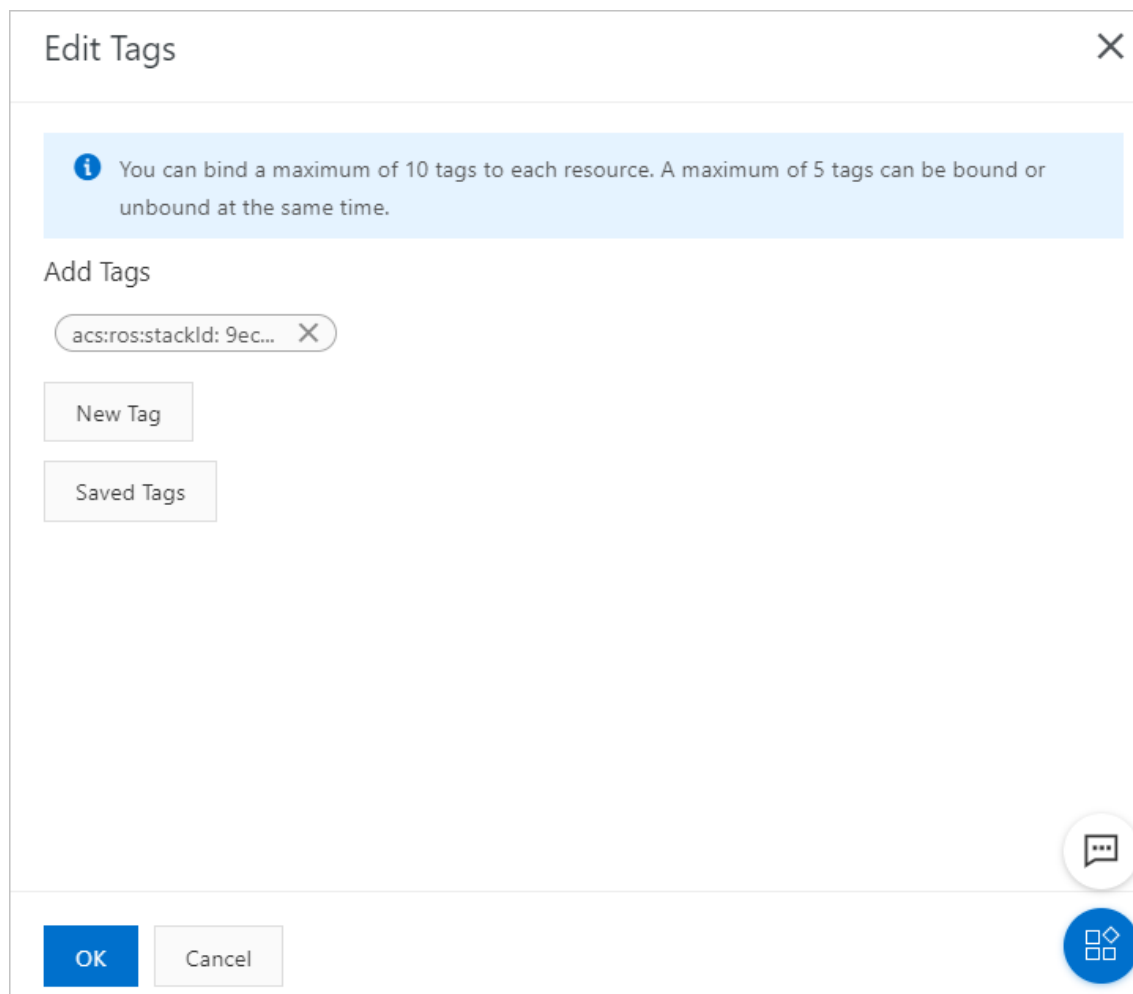
Procedure

1.

2.

3. In the left-side navigation pane, choose **Instances > Server Load Balancers**.

4. Select the region of the CLB instance that you want to manage and find the CLB instance.

5. In the **Actions** column, choose ⋮ **> Edit Tags**.
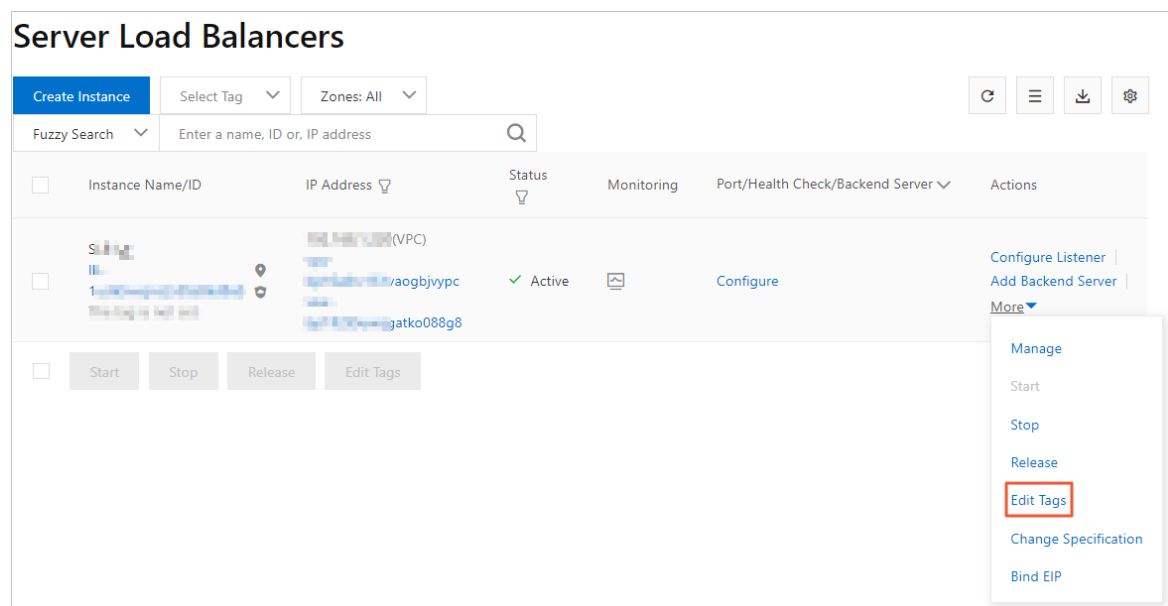
6. Edit tags in the **Edit Tags** dialog box.

   To add a tag, perform the following operations:

   - To add an existing tag, click **Saved Tags** and then select a tag.

   - To create and add a new tag, click **New Tag** in the **Edit Tags** dialog box, enter the key and

value of the new tag, and then click **OK** next to the value.

7. Click **OK**.

7.3. Query CLB instances by tag

This topic describes how to use tags to query CLB instances.

Procedure

1.

2. In the left-side navigation pane, choose **Instances > Server Load Balancers**.

3. Select the region of the CLB instance that you want to manage and find the CLB instance.

4. Click **Select a tag** and select a tag to filter instances.

> ⑦ **Note**   To clear the search condition, move the pointer over the selected tag and click the displayed deletion icon next to it.

7.4. Remove tags

This topic describes how to remove tags from an SLB instance. You can only remove tags for one SLB instance at a time.

Procedure

1.

2.

3. In the left-side navigation pane, choose **Instances > Server Load Balancers**.

4. Select the region of the target SLB instance and find the target SLB instance.

5. In the **Actions** column, choose ⋮ > **Edit Tags**.

6. In the **Edit Tags** dialog box, click the deletion icon next to the tags to be removed, and then click **OK**.

> ⑦ **Note**   If a tag is removed from an SLB instance and is not added to any other instances, the tag is deleted from the system.

8.Enable deletion protection

This topic describes how to enable deletion protection for pay-as-you-go Classic Load Balancer (CLB) instances. Deletion protection prevents CLB instances from being mistakenly deleted.

Procedure

1.

2. On the **Instances** page, find the CLB instance and click its ID or choose ⋮ **> Manage** in the
   Actions column.

3. Click the **Instance Details** tab and turn on **Deletion Protection**.

   ⑦ **Note**   You must disable **Deletion Protection** before you can delete the CLB instance.
   Otherwise, the system reports an error.

Related information

- SetLoadBalancerDeleteProtection: enables or disables deletion protection for a CLB instance.
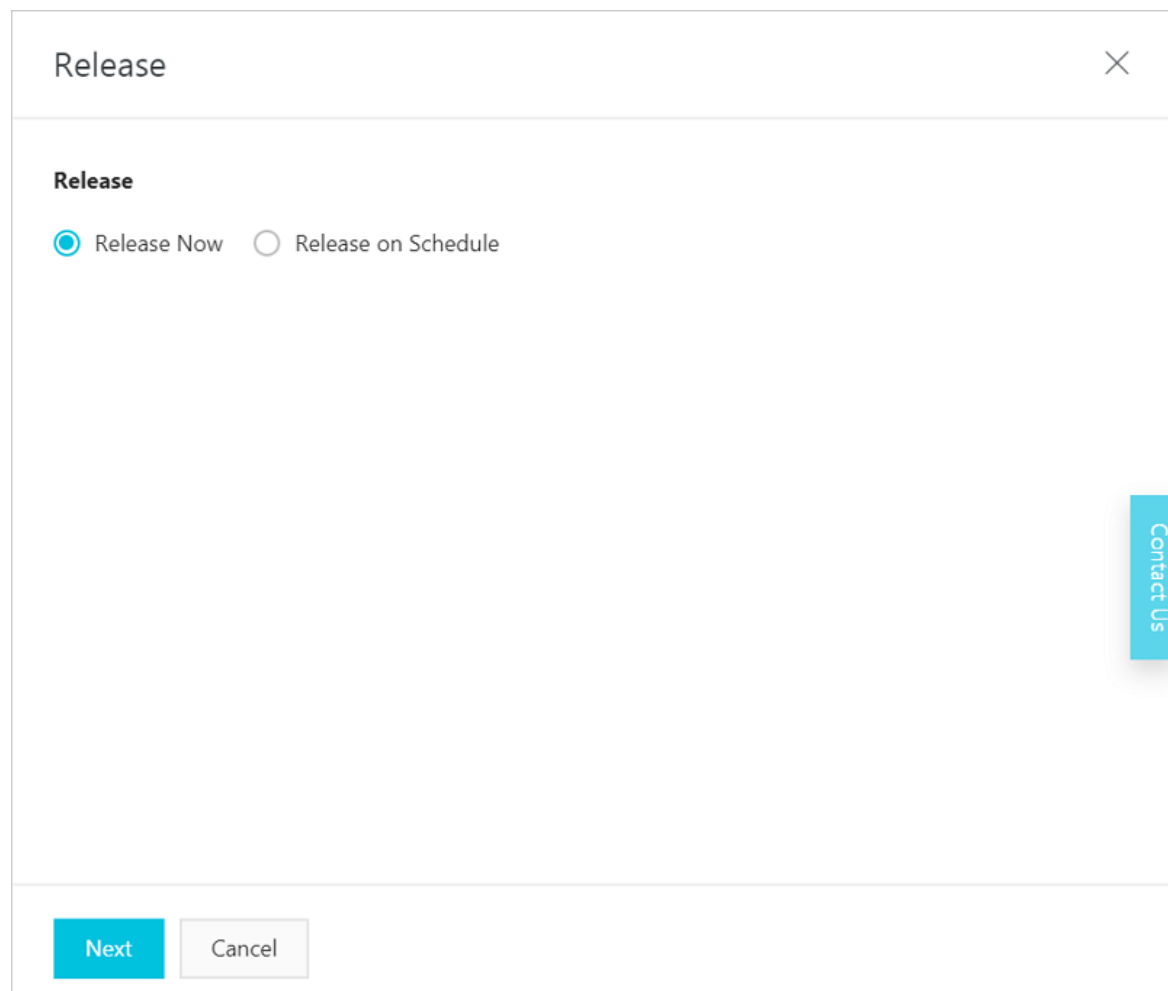
9.Release an SLB instance

This topic describes how to release a Server Load Balancer (SLB) instance. You can immediately release pay-as-you-go SLB instances or release them on schedule.

Context

Procedure

1. Log on to the SLB console.

2. In the left-side navigation pane, choose **Instances > Instances**.

3. Find the SLB instance that you want to release and choose ⋮ > **Release** in the Actions column.

   To release multiple SLB instances at a time, select the instances and click **Release** at the bottom of the page.

4. In the **Release** panel, select Release Now or Release on Schedule.

   > ⑦ **Note**   The system releases the SLB instances at 30-minute and hour marks. However, billing for the SLB instance is stopped at the specified release time.

5. Click **Next**.

6. Confirm the displayed information and click **OK** to release the instance.

10.Expired instances

This topic describes how to manage instances that are expired due to overdue payments.

Procedure

1.

2. In the left-side navigation pane, choose **CLB (Formerly Known as SLB) > Expired Instances**.

3. View details of the expired Classic Load Balancer (CLB) instances.

4. Find the CLB instance that you want to renew and click **Renew** in the **Actions** column. After the CLB instance is renewed, the CLB instance is removed from the Expired Instances page and listed on the Instances page.

11.Manage idle instances

This topic describes how to use the Server Load Balancer (SLB) console to display the pay-as-you-go instances that have been idle for more than seven days.

Procedure

1.

2.

3. In the left-side navigation pane, choose **CLB (FKA SLB) > Idle SLB instances**.

4. On the **Idle SLB instances** page, view all the pay-as-you-go instances that have not been used for more than seven days. You can click [⚙] to customize the display of IP Address and Idle Cause.

5. To release an idle instance, click **Release** in the **Actions** column to immediately release the instance.

> ⑦ **Note**   Given that the data of idle instances has a one-day cache period, make sure that the instance you want to release is not in use to prevent possible exceptions.

12.FAQ
12.1. FAQ about CLB instances

This topic provides answers to some frequently asked questions about instances.

- Can I change a shared-resource CLB instance to a high-performance CLB instance?
- What are the specifications of shared-resource CLB instances?
- How do I choose between different specifications for a high-performance CLB instance?
- Can I change the specification of a high-performance CLB instance?

Can I change a shared-resource CLB instance to a high-performance CLB instance?

Yes, you can change a shared-resource CLB instance to a high-performance CLB instance.

-
- After you change a shared-resource CLB instance to a high-performance CLB instance, you are charged a specification fee.

What are the specifications of shared-resource CLB instances?

The performance of shared-resource CLB instances is not guaranteed. Therefore, you cannot specify the specification of a shared-resource CLB instance.

How do I choose between different specifications for a high-performance CLB instance?

- You can select the highest specification if your high-performance CLB instance is billed on a pay-as-you-go basis. Pay-as-you-go CLB instances are billed based on the actual resource usage.

Can I change the specification of a high-performance CLB instance?

Yes, you can change the specification of a high-performance CLB instance.

- You can upgrade or downgrade the specification of a high-performance CLB instance if the CLB instance is billed on a pay-as-you-go basis. For more information, see Upgrade or downgrade a pay-as-you-go CLB instance.

> ⑦ Note
> -
> -

12.2. FAQ about high-performance CLB instances

High-performance Classic Load Balancer (CLB) instances provide guaranteed performance. This topic provides answers to some frequently asked questions about high-performance CLB instances.

- What is a high-performance CLB instance?
- How are high-performance CLB instances billed?
- What is the optimal specification for a high-performance CLB instance?
- Can I change the specification of a high-performance CLB instance?
- Are specification fees charged by existing shared-resource CLB instances?
- Why is a high-performance CLB instance unable to reach the performance limit defined in the specification?
- Are shared-resource CLB instances still available for purchase?
- Is a specification fee charged by an internal-facing CLB instance?
- What do I do if I need more high-performance CLB instances?

What is a high-performance CLB instance?

A high-performance CLB instance provides guaranteed performance. In comparison, shared-resource CLB instances share resources with each other, which means that performance cannot be guaranteed.

Before Alibaba Cloud released high-performance CLB instances, all instances were shared-resource instances. You can view the instance type in the CLB console.

> 🔊 **Notice**   Shared-resource CLB instances are no longer available.

The following table describes the differences between shared-resource CLB instances and high-performance CLB instances.

Feature	Shared-resource CLB instance	High-performance CLB instance
Resource allocation	Shared resources	Exclusive resources
Service uptime guaranteed by terms of service level agreement	Not supported	99.95%
IPv6	×	√
Support for Server Name Indication (SNI) certificates	×	√
Blacklists and whitelists	×	√
Associating with elastic network interfaces (ENIs)	×	√
Adding secondary IP addresses of ENIs that are associated with Elastic Compute Service (ECS) instances	×	√
HTTP-to-HTTPS redirection	×	√
Consistent hashing	×	√
Support for TLS security policies	×	√

Feature	Shared-resource CLB instance	High-performance CLB instance
HTTP2	×	√
Websocket(S)	×	√

You can move the pointer over the question mark that corresponds to the specification of a high-performance CLB instance to view the performance metrics, as shown in the following figure.

The following section lists three key metrics for high-performance CLB instances:

- Max Connection

  The maximum number of concurrent connections that a CLB instance supports. When the number of concurrent connections reaches the specified limit, new connection requests are dropped.

- Connections Per Second (CPS)

  The number of new connections that are established per second. When the CPS value reaches the specified limit, new connection requests are dropped.

- Queries Per Second (QPS)

  The number of HTTP or HTTPS queries (requests) that can be processed per second. This metric is specific to Layer 7 listeners. When the QPS value reaches the specified limit, new connection requests are dropped.

The following table lists the specifications of high-performance instances provided by Alibaba Cloud. The specifications available for purchase vary by region. You can go to the buy page of the CLB console to view the specifications available for purchase in your region.

> ? Note

How are high-performance CLB instances billed?

High-performance CLB instances are billed based on the following formula:

Total fee (per instance) = Instance fee + Data transfer fee or bandwidth fee + Specification fee

The specification fee of a high-performance CLB instance is charged based on the actual usage no matter which specification you choose.

Assume that you purchase a CLB instance of the Super I (slb.s3.large) specification (Max Connection: 1,000,000; CPS: 500,000; QPS: 50,000). The following table lists the highest metric values of your CLB instance within an hour.

Max Connection	CPS	QPS
90000	4000	11000

- The actual Max Connection value of 90,000 is between the limit of 50,000 defined in Standard I (slb.s2.small) and the limit of 100,000 defined in Standard II (slb.s2.medium). Therefore, the specification based on the Max Connection metric for this hour is Standard II (slb.s2.medium).

- The actual CPS value of 4,000 is between the limit of 3,000 defined in Small I (slb.s1.small) and the limit of 5,000 defined in Standard I (slb.s2.small). Therefore, the specification based on the CPS metric for this hour is Standard I (slb.s2.small).

- The actual QPS value of 11,000 lies between the limit of 10,000 defined in Standard II (slb.s2.medium) and the limit of 20,000 defined in Higher I (slb.s3.small). Therefore, the specification based on the QPS metric for this hour is Higher I (slb.s3.small).

  The QPS metric uses the highest specification (slb.s3.small) among three metrics. Therefore, the instance is charged based on the Higher I (slb.s3.small) specification for this hour.

Billing is more flexible for pay-as-you-go high-performance CLB instances. The specification that you select when you purchase a CLB instance represents the upper performance limit of the instance. For example, if you select Higher II (slb.s3.medium), the instance performance cannot be higher than Higher II (slb.s3.medium).

What is the optimal specification for a high-performance CLB instance?

The specification fee is charged based on your actual usage. We recommend that you select the highest specification (slb.s3.large) available to ensure service flexibility at no extra cost. However, if the capacity of Super I (slb.s3.large) significantly exceeds your business demand, you can select a more appropriate specification, such as Higher II (slb.s3.medium).

Can I change the specification of a high-performance CLB instance?

Yes, you can change the specification of a high-performance CLB instance in the console.

> ? **Note**
> -
>   - Changing a shared-resource CLB instance to a high-performance CLB instance does not negatively affect your workloads, or change the IP address of the CLB instance.
>
>   However, we recommend that you change a shared-resource CLB instance to a high-performance CLB instance during off-peak hours. You can also configure DNS resolution to replace your current load balancing service before you change the CLB instance.
>
>   - The IP addresses of CLB instances are not affected when you change instance specifications.

Are specification fees charged by existing shared-resource CLB instances?

No.

Specification fees are not charged by existing shared-resource instances unless you change them to high-performance CLB instances. You can change a shared-resource CLB instance to a high-performance CLB instance by changing the specification. When you change a shared-resource CLB instance to a high-performance CLB instance, a specification fee is charged.

Why is a high-performance CLB instance unable to reach the performance limit defined in the specification?

This issue can be explained by the cask theory.

High-performance CLB instances do not guarantee that the three metrics can reach the upper limits of a specification at the same time. The instance performance is limited if one of the metrics reaches the upper limit.

Assume that you have purchased a high-performance CLB instance of the Higher I (slb.s3.small) specification. When the QPS of the instance reaches 20,000 but the number of maximum connections does not reach 200,000, new connection requests are dropped because the QPS limit is reached.

Are shared-resource CLB instances still available for purchase?

No.

Is a specification fee charged by an internal-facing CLB instance?

If the internal-facing CLB instance that you use is a shared-resource CLB instance, no specification fee is charged. If the internal-facing CLB instance is a high-performance CLB instance, a specification fee is charged.

The specification fees for internal-facing CLB instances are charged in the same way as Internet-facing CLB instances. No instance fees or data transfer fees are charged by internal-facing CLB instances.

What do I do if I need more high-performance CLB instances?

If you need more high-performance CLB instances but the number of the instances has reached the quota, you can apply for the **slb_privilege_allow_more_guaranteed_performance_instances** privilege for a quota increase. This privilege allows you to own more high-performance CLB instances, but does not allow you to own more shared-resource CLB instances. For more information, see Manage quotas.