Alibaba Cloud Server Load Balancer

User Guide

Issue: 20190705

Legal disclaimer

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

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

or trust in this document, even if Alibaba Cloud has been notified of the possibility of such a loss.

5. By law, all the content of the Alibaba Cloud website, including but not limited to works, products, images, archives, information, materials, website architecture, website graphic layout, and webpage design, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of the Alibaba Cloud website, product programs, or content shall be used, modified , reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates . The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates).
6. Please contact Alibaba Cloud directly if you discover any errors in this document.

Generic conventions

Table -1: Style conventions

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

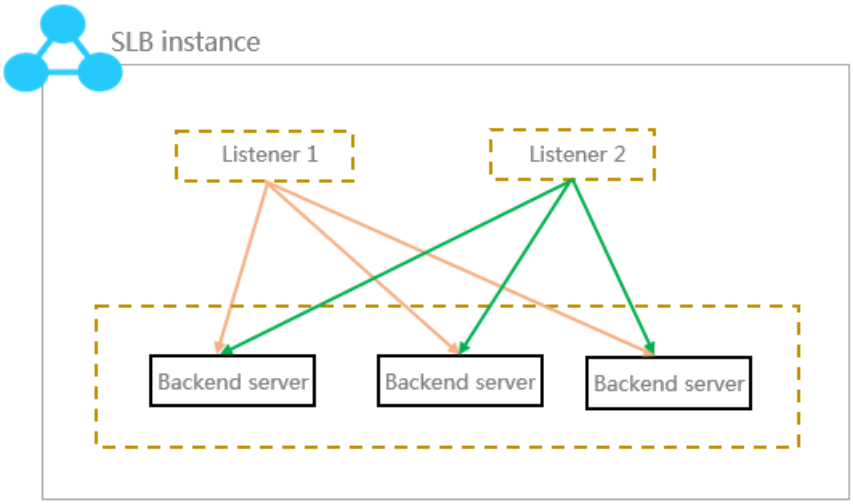
Style	Description	Example
{} or {a|b}	It indicates that it is a required value, and only one item can be selected.	`swich {stand | slave}`

Contents

1 Server Load Balancer instance

1.1 SLB instance overview

An SLB instance is a running entity of the Server Load Balancer service. To use the load balancing service, you must create an SLB instance first, and then add listeners and backend servers to it.

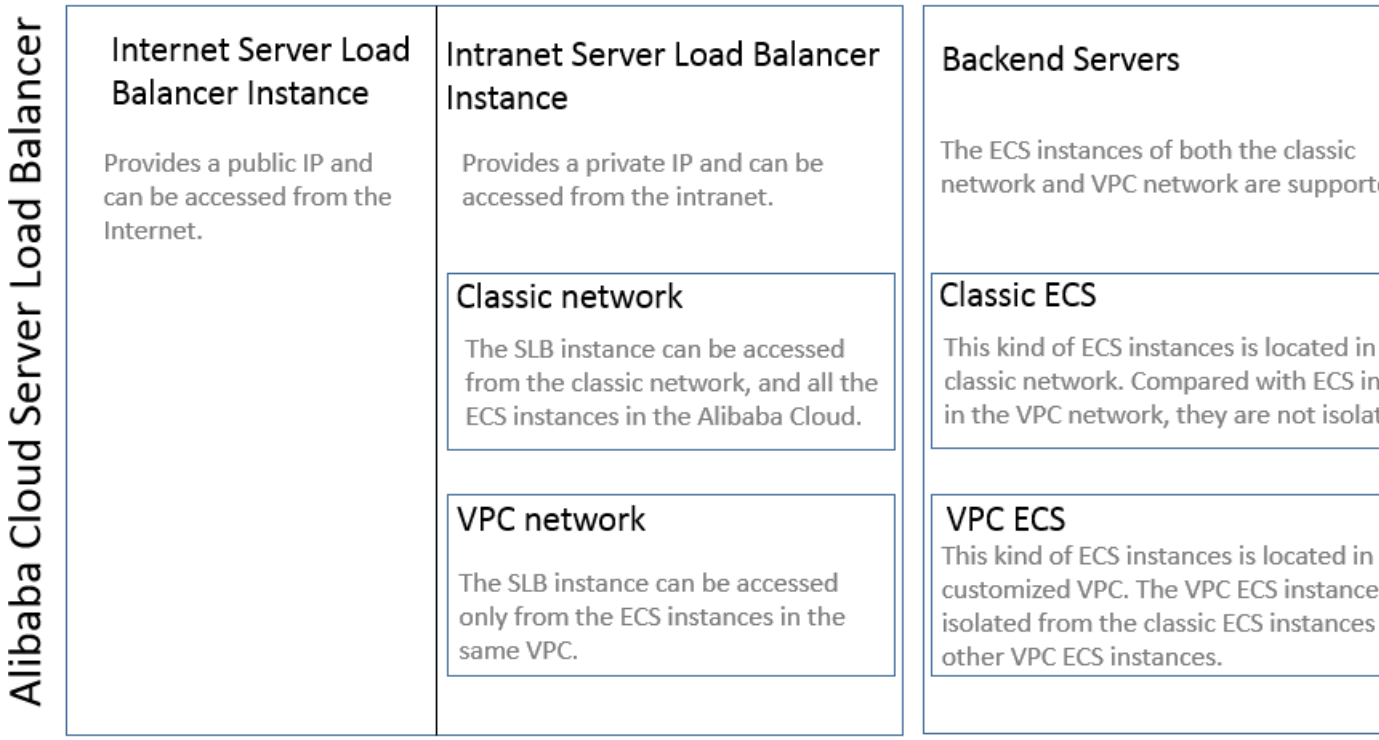


Alibaba Cloud provides Internet SLB service and intranet SLB service. A public or a private IP address is allocated to the SLB instance according to the instance type you select.

Internet SLB instances

An Internet SLB instance distributes client requests over the Internet to backend ECS servers according to configured forwarding rules.

After you create an Internet Server Load Balancer instance, the system will allocate a public IP to the instance. You can resolve a domain name to the public IP to provide public services.



Intranet SLB instances

Intranet SLB instances can only be used inside Alibaba Cloud and can only forward requests from clients that can access the intranet of SLB.

For an intranet SLB instance, you can further select the network type:

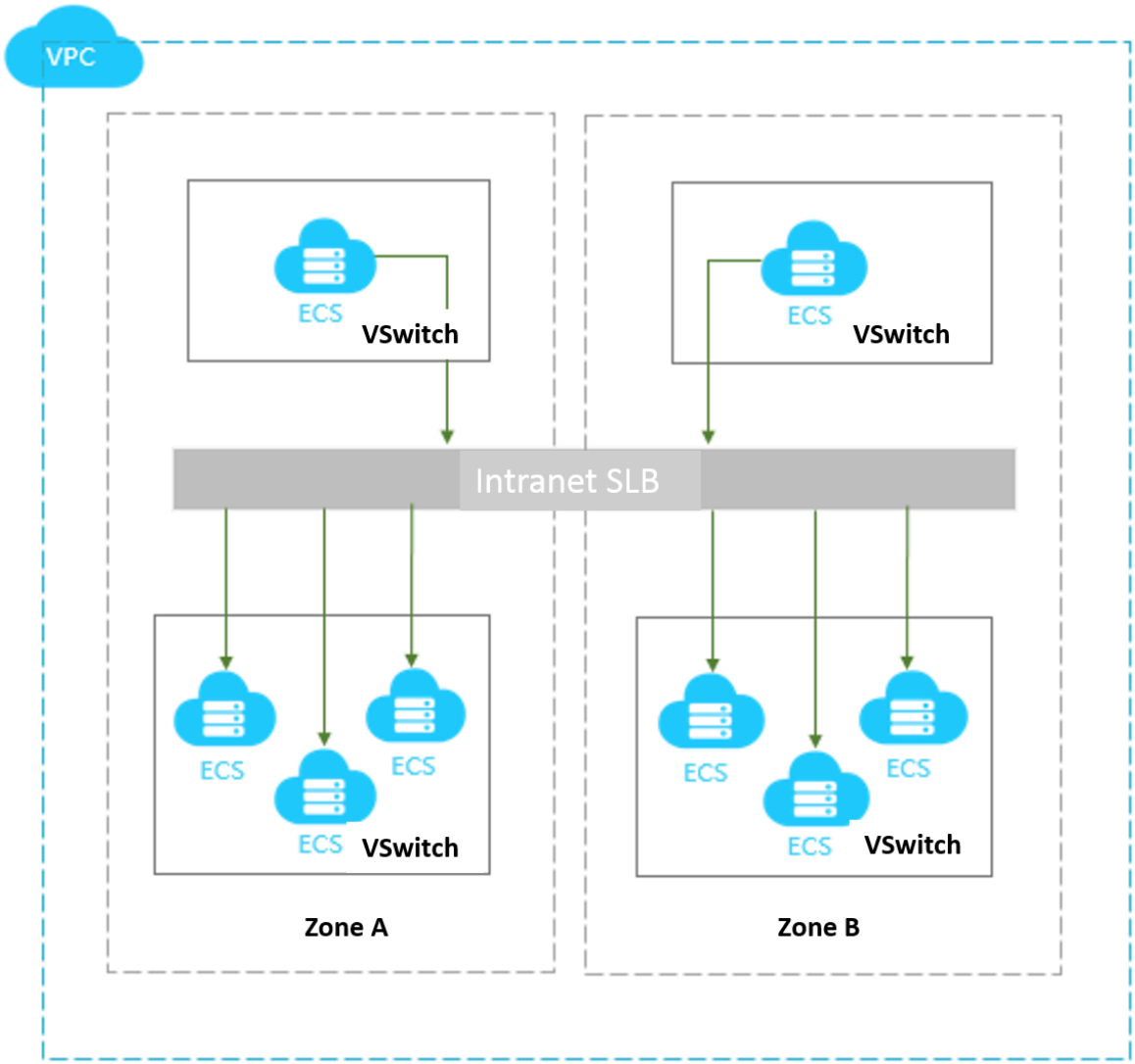
- **Classic network**

  If you choose classic network for the intranet SLB instance, the IP of the SLB instance is allocated and maintained by Alibaba Cloud. The classic SLB instance can only be accessed by the classic ECS instances.

- **VPC network**

  If you choose VPC network for the intranet SLB instance, the IP of the SLB instance is allocated from the CIDR of the VSwitch that the instance belongs to. SLB

instances of the VPC network can only be accessed by ECS instances in the same VPC.



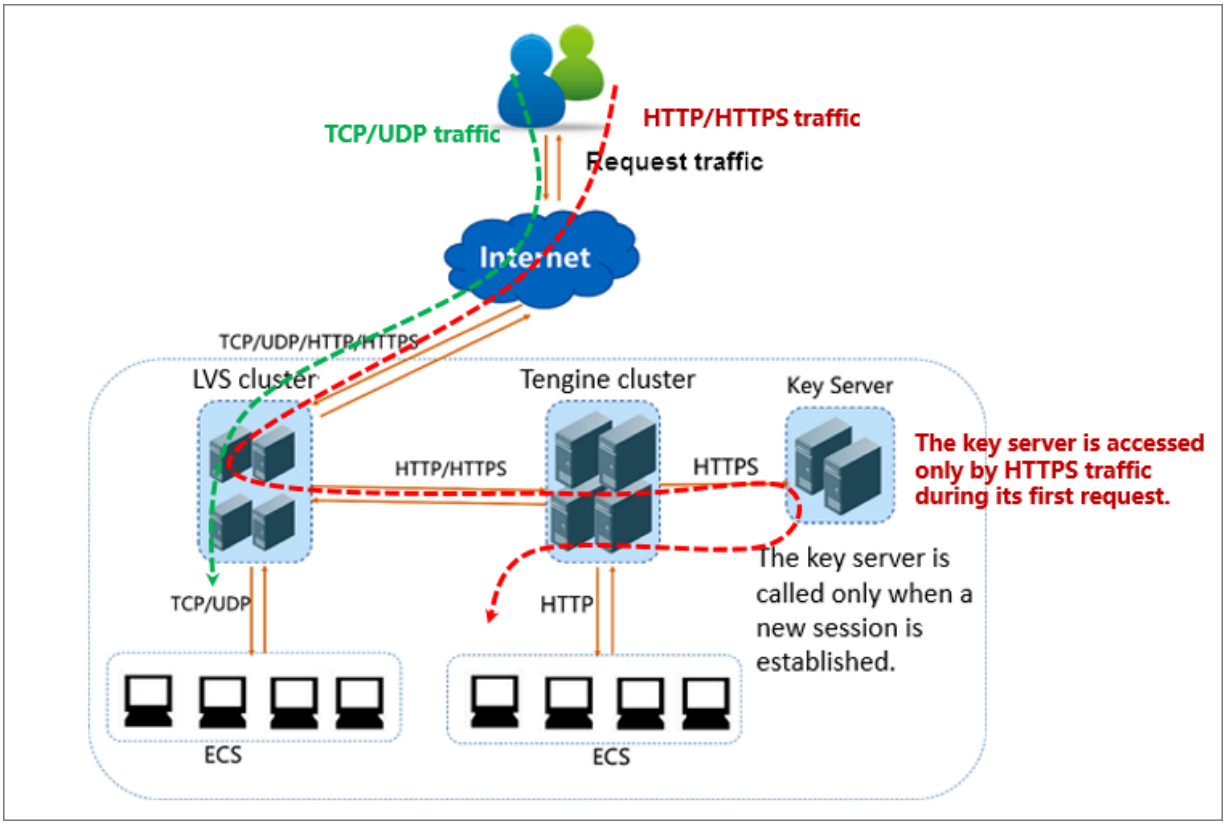
1.2 Network traffic flow

As a traffic forwarding service, SLB forwards requests from clients to backend servers through SLB clusters. Then, the backend servers return responses to SLB through the intranet.

Inbound network traffic flow

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

Figure 1-1: Inbound network traffic flow



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

2. Large amounts of access requests are evenly distributed among all servers in the LVS cluster. Servers synchronize sessions to guarantee high availability.
   - For Layer-4 listeners (the frontend protocol is 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 (the frontend protocol is HTTP), 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 (the frontend protocol is HTTPS), the request distribution is similar to the HTTP protocol. However, before distributing requests to backend ECS instances, the system calls the Key Server to validate certificates and decrypt data packets.

Outbound network traffic flow

SLB communicates with backend ECS instances through the intranet.

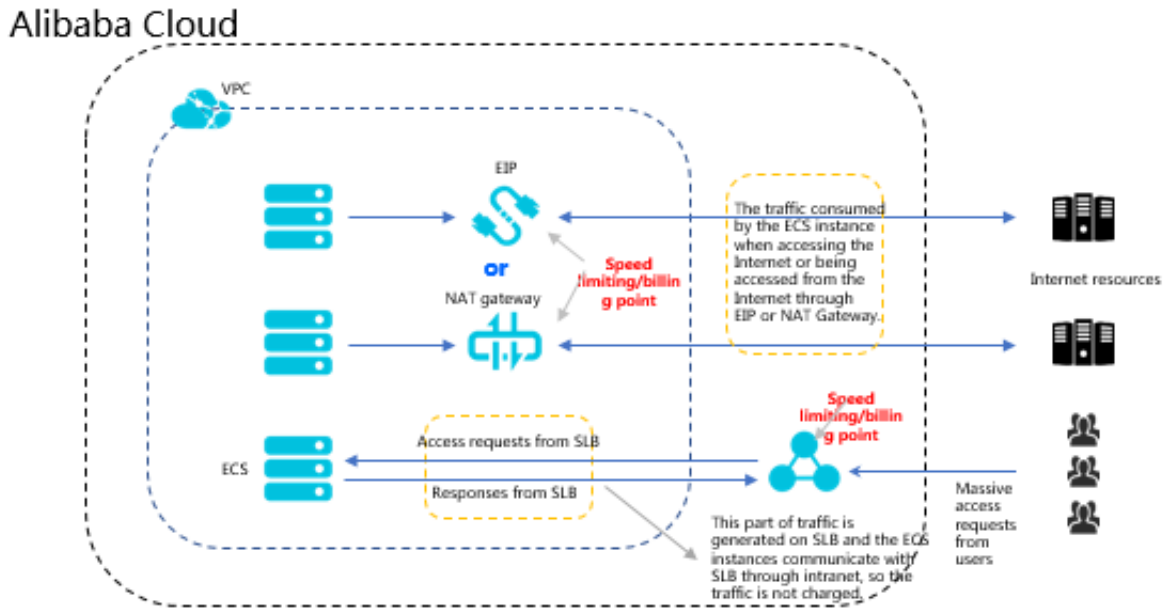
- If backend ECS instances only need to handle the traffic distributed from SLB, no public bandwidth (EIP, NAT Gateway, and public IP address) is required, and you do not need to purchase any public bandwidth.

  > **Note:**
  > ECS instances previously created are directly allocated with public IP addresses. You can view the public IP addresses by using the ifconfig command. If these ECS instances process requests only through SLB, no traffic fee is incurred for traffic sent through the Internet even 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: a public IP address, an EIP, or a NAT Gateway.

The following figure shows the outbound network traffic flow.

Figure 1-2: Outbound network traffic flow



1. For outbound traffic from SLB instances (that is, traffic transferred through the Internet), traffic is sent at speeds dependent on the current network capacity, and is charged. However, you are not charged for intranet communications, such as traffic transferred between SLB instances and backend ECS instances.
2. For outbound traffic from an EIP or from NAT Gateway (that is, traffic transferred through the Internet), traffic is sent at speeds dependent on the current network capacity, and is charged. Additionally, if an ECS instance is configured with a public IP address when it is created, the outbound traffic from this instance is also charged.
3. SLB supports dynamic access to the Internet. Specifically, if a backend ECS instance needs to access the Internet, you must first configure a public IP address for it (by using an EIP or using NAT Gateway).
4. A public IP address (configured when you create an ECS instance), EIP, and NAT gateway all allow mutual Internet access. That is, ECS instances can access the Internet or be accessed from the Internet through any of these. Note, however, that they cannot forward traffic or balance traffic loads.

1.3 Create an SLB instance

This topic describes how to create a Server Load Balancer (SLB) instance.

Prerequisites

The required instance type, instance region, network type, and listener protocol are known. The environment is prepared. For more information, see Plan and prepare.

Procedure

1. Log on to the SLB console.
2. In the left-side navigation pane, choose Instances > Server Load Balancer, and click Create SLB Instance in the upper-left corner.
3. Configure the SLB instance according to the following information.

Configuration	Description
Region	Select the region to which the SLB instance belongs. Note: Make sure that the region of the SLB instance is the same as that of backend ECS instances.
Zone Type	The zone type of the selected region is displayed. The zone of a cloud product refers to a set of independent infrastructure and is usually represented by data centers. Different zones have independent infrastructure (network, power supply, air-conditioning and so on). Therefore, an infrastructure fault in one zone does not affect other zones. A zone belongs to a specific region. A single region may have one or more zones. SLB has deployed multiple zones in most regions. · Single zone: The SLB instance is deployed only in one zone. · Multi-zone: The SLB instance is deployed in two zones. By default, the primary zone is used. If the primary zone is faulty, the secondary zone automatically takes over the load balancing service.
Primary Zone	Select the primary zone for the SLB instance. The primary zone carries traffic in normal conditions.
Backup Zone	Select the secondary zone for the SLB instance. The secondary zone only takes over traffic when the primary zone is unavailable.

Configuration	Description
Instance name	Enter a name for the SLB instance to be created. The name must be 1 to 80 characters in length and can contain letters, numbers, Chinese characters, hyphens (-), slashes (/), periods (.), and underscores (_).
Resource Group	The resource group to which the SLB instance to be created belongs.
Instance Spec	Select a performance specification for the instance. The performance metrics vary by specification. For more information, see Guaranteed-performance instances.
Instance Type	Select the instance type based on your business needs. A public or a private IP address is allocated to the SLB instance based on the instance type. For more information, see SLB instance overview. · Internet: An Internet SLB instance only provides a public IP address and you can access the SLB service from the Internet. · Intranet: An intranet SLB instance only provides a private IP address and you can access the SLB service only from the intranet.
IP version	Select the IP version of the SLB instance, which can be IPv4 or IPv6. Note: Currently, IPv6 instances are supported only in the following regions. However, the instances must be guaranteed-performance instances. · Zone E and Zone F in the China (Hangzhou) region · Zone F and Zone G in the China (Beijing) region · Zone D and Zone E in the China (Shanghai) region · Zone D and Zone E in the China (Shenzhen) region
Quantity	Select the number of SLB instances to create.

4. Click Buy Now and complete the payment.

1.4 Create an IPv6 instance

This topic describes how to create an IPv6 Server Load Balancer (SLB) instance. After an IPv6 SLB instance is created, the system allocates a public IPv6 address to the instance to forward requests from IPv6 clients.

Context

IPv6 is the abbreviation of Internet Protocol Version 6. IPv6 is the next-generation IP protocol designed by IETF (Internet Engineering Task Force) to replace the current version of IP protocol (IPv4). By extending the length of IPv4 address from 32 bits to 128 bits, it expands the address space by 79,228,162,514,264,337,593,543,950,336 times. After IPv6 is used, each grain of sand on the world can be allocated with an IP address.

 Notice:

- Currently, IPv6 instances are supported in the following zones, but the instances must be guaranteed-performance instances.
  - Zones E and F in the China (Hangzhou) region
  - Zones F and G in the China (Beijing) region
  - Zones D and E in the China (Shanghai) region
  - Zones D and E in the China (Shenzhen) region
- The Internet IPv6 network environment is still in the early stage of construction, and some links may be inaccessible. If such problem occurs, submit a ticket for technical support. SLA is not provided in the open beta test stage.
- IPv6 has a longer IP header than IPv4. Therefore, when you use a UDP listener in an IPv6 SLB instance, you must ensure that the MTU of the NIC communicating with SLB on the backend server (ECS instance) is not greater than 1480 (some applications need to synchronize their configuration files based on this MTU value). Otherwise, the packets may be discarded because they are too large.

  If you use a TCP, HTTP, or HTTPS listener, no additional configurations are required because the TCP protocol supports MSS auto-negotiation.
- HTTP listeners can use the `X-Forwarded-For` header field to obtain source IPv6 addresses of clients.

IPv6 instances have the following features:

- Smooth migration and no impact on your service

  You can directly associate ECS instances that use IPv4 addresses with an IPv6 SLB instance and smoothly migrate the service to IPv6 without modifying the original system.

  IPv6 has no impact on the original IPv4 service. If the traffic volume increases, you only need to increase backend ECS instances.
- IPv6 access control ensures more secure and reliable service deployment

  SLB supports IPv6 access control. You can configure access control lists according to your business needs.

  - A blacklist can effectively block the access of malicious addresses to the SLB service.
  - If a whitelist is configured, only addresses in the whitelist can access the SLB service.

Procedure

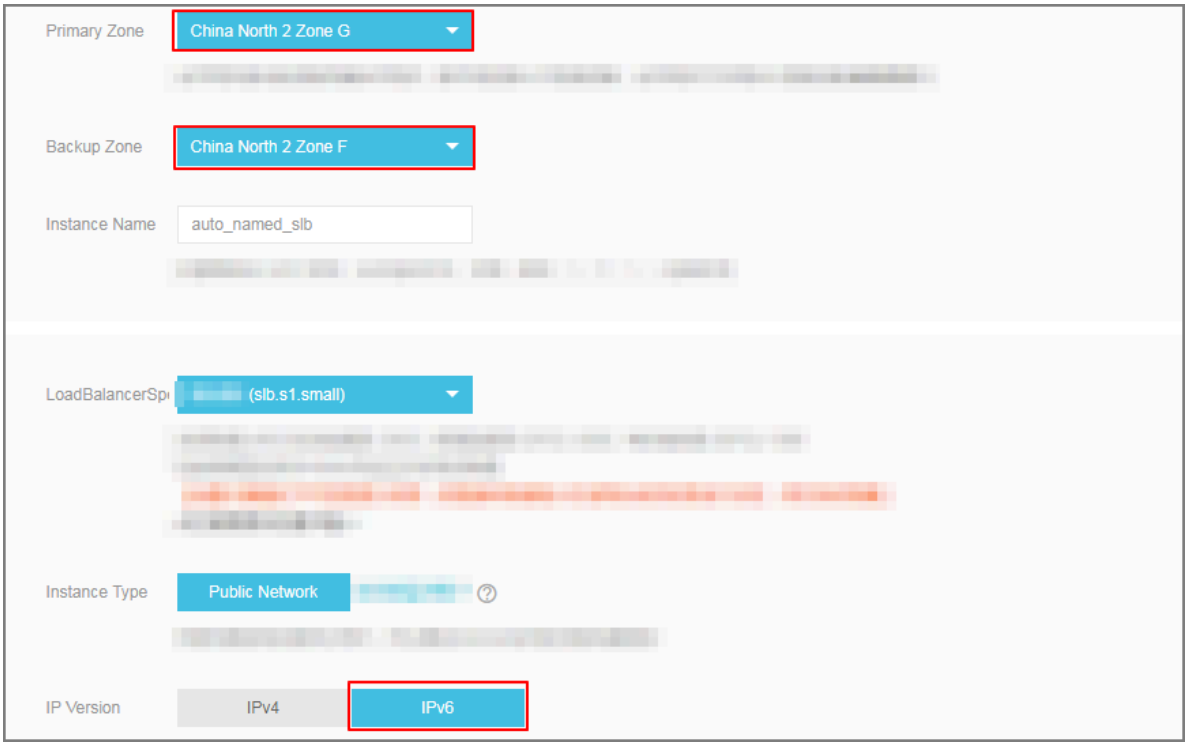
1. Log on to the SLB console.
2. In the left-side navigation pane, choose Instances > Server Load Balancer.
3. On the Server Load Balancer page, click Create SLB Instance in the upper-left corner.
4. Configure the SLB instance. For the IP version, select IPv6.

   Other configurations are the same as configurations of common instances. For more information, see SLB configurations.

    Note:
   Currently, IPv6 instances are supported in the following zones, but the instances must be guaranteed-performance instances.

   - Zones E and F in the China (Hangzhou) region
   - Zones F and G in the China (Beijing) region
   - Zones D and E in the China (Shanghai) region

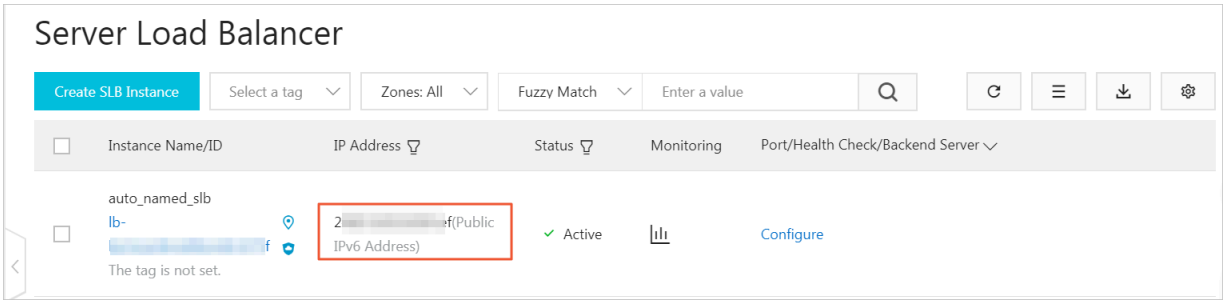
- Zones D and E in the China (Shenzhen) region



5. Go back to the Server Load Balancer page to view the created IPv6 instance.

Result

After the IPv6 instance is created, the system allocates an IPv6 address to it.



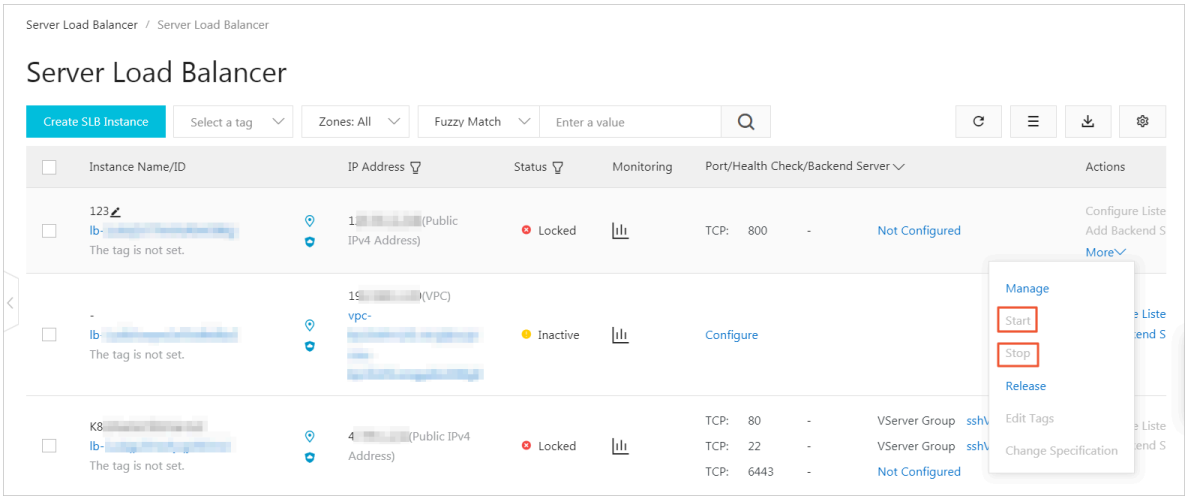
1.5 Start or stop an SLB instance

You can start or stop a Server Load Balancer (SLB) instance at any time. After being stopped, an SLB instance does not receive or forward requests any more.

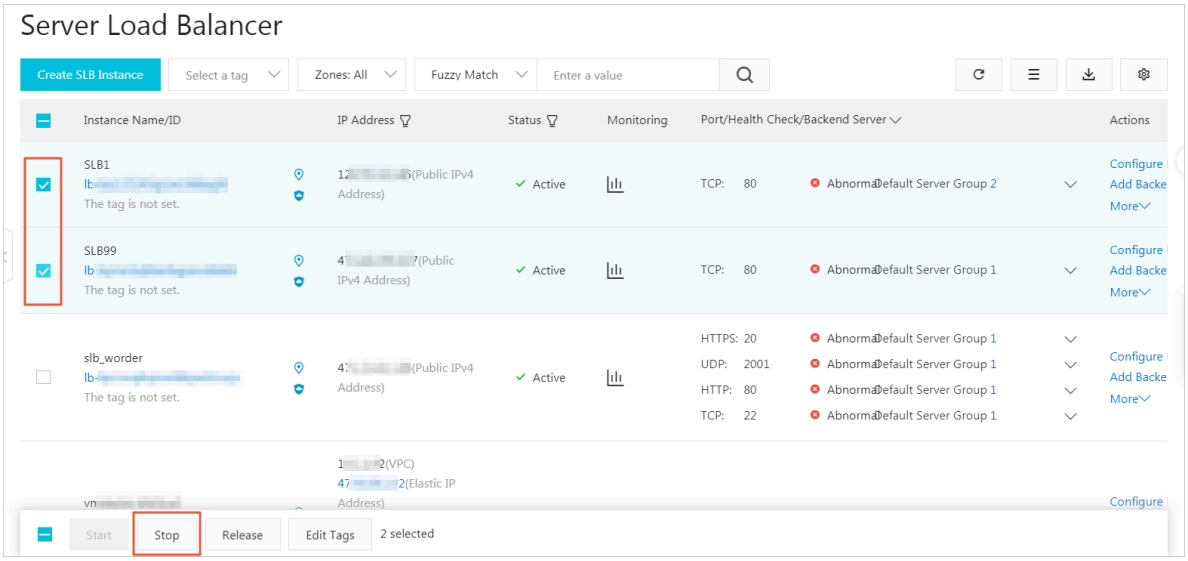
Procedure

1. Log on to the SLB console.
2. In the left-side navigation pane, choose Instances > Server Load Balancer.
3. Select the region of the target SLB instance and find the target instance.

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



5. If you want to start or stop multiple instances at a time, select the target instances and click Start or Stop at the lower part of the page.



1.6 Bind an EIP

You can bind an EIP to an SLB instance of the VPC network. After being bound to an EIP, the SLB instance can forward requests from the Internet.

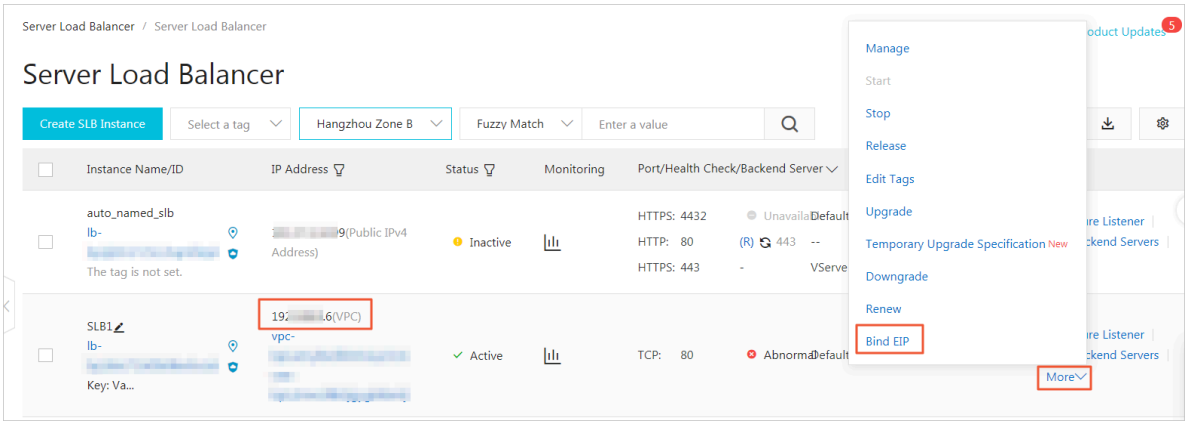
Procedure

1. Log on to the SLB console.
2. In the left-side navigation pane, click Instances > Server Load Balancer.
3. Select a region and find the target instance.

Note:

Ensure that the SLB instance is of the VPC network.

4. Click More > Bind EIP.



5. Select an EIP and click OK.

1.7 Release an SLB instance

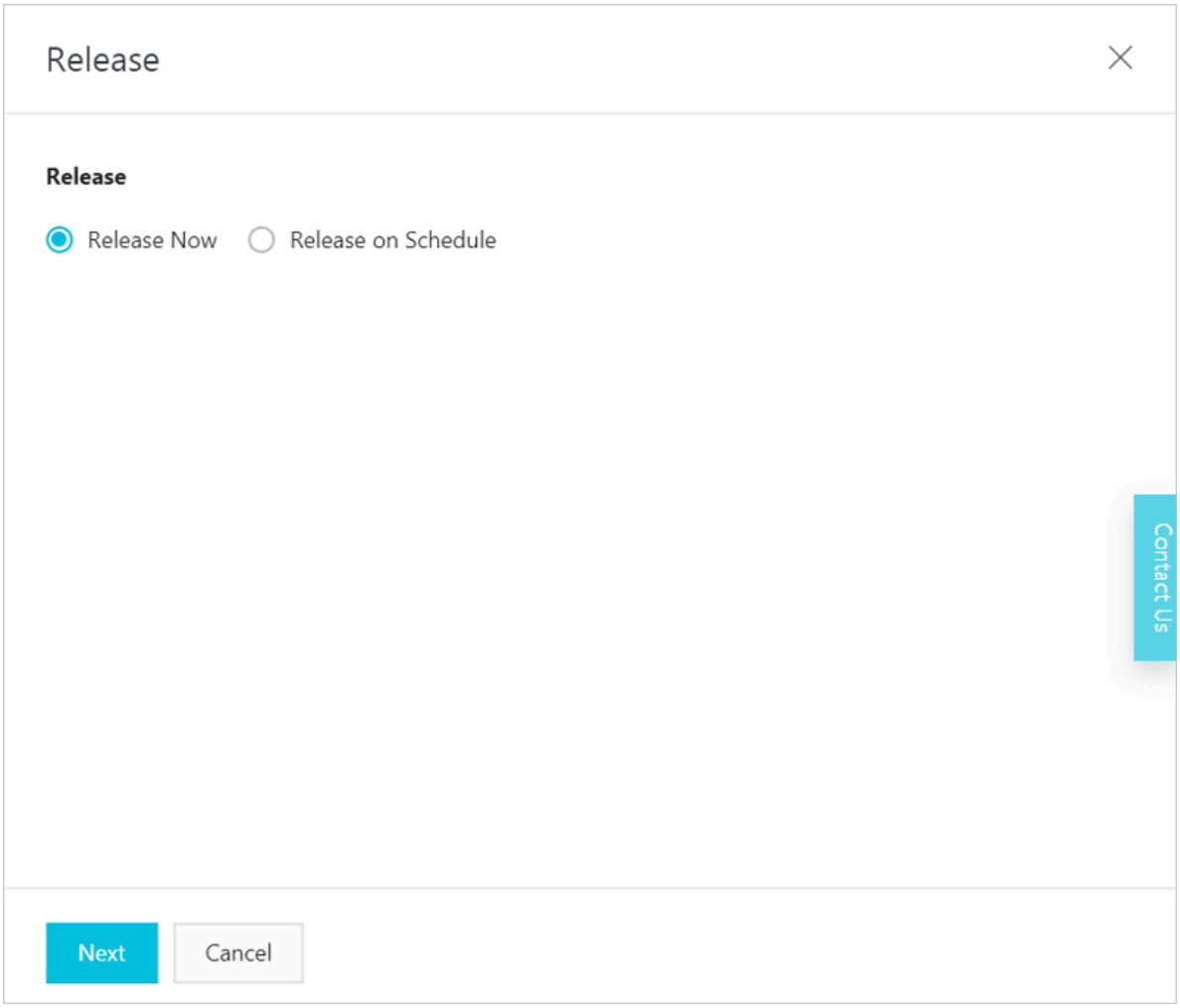
This topic describes how to release a Server Load Balancer (SLB) instance. You can release an SLB instance immediately or at a specified time.

Procedure

1. Log on to the SLB console.

2. Find the target instance and click More > Release.

   You can select multiple SLB instances at a time and click Release at the bottom of the page to release SLB instances in batches.



3. On the Release page, select Release Now or Release on Schedule.

   Note:
   While the system executes the release operation every half hour or one hour cycle, the billing of the instance is stopped immediately at the release time you set.

4. Click Next.
5. Confirm the displayed information and click OK to release the instance.

1.8 Manage tags

You can classify Server Load Balancer (SLB) instances by using tags.

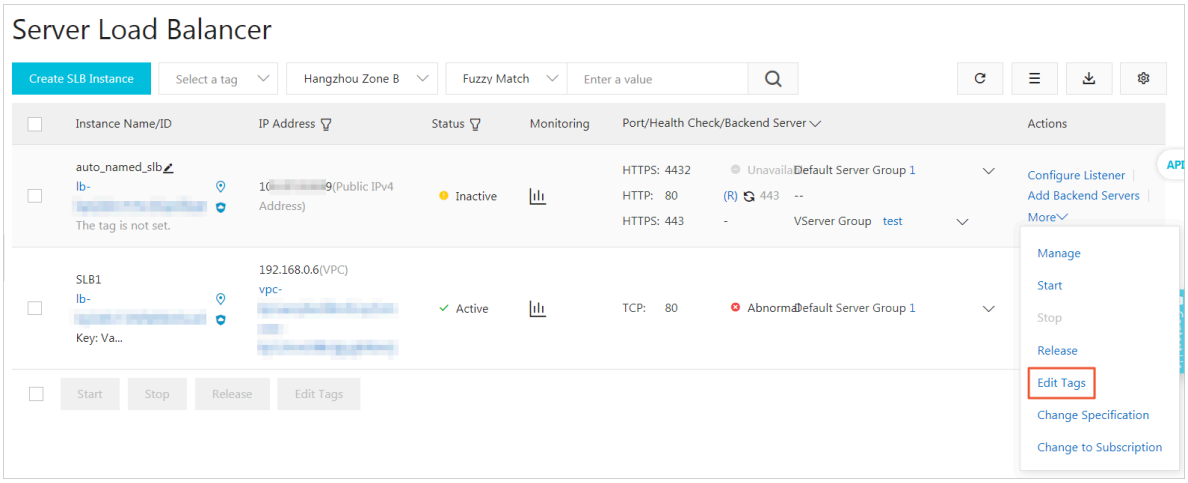
Each tag consists of a key and a value. Before you use tags, note the following limits:

- A tag cannot exist on its own and must be associated with an SLB instance.
- Up to 10 tags can be associated with an SLB instance.
- The key of each tag associated with an instance must be unique. Tags with the same key will be overwritten.
- Tags cannot be used across regions and are region-specific resources. For example , tags that belong to the China (Hangzhou) region are invisible to the China (Shanghai) region.

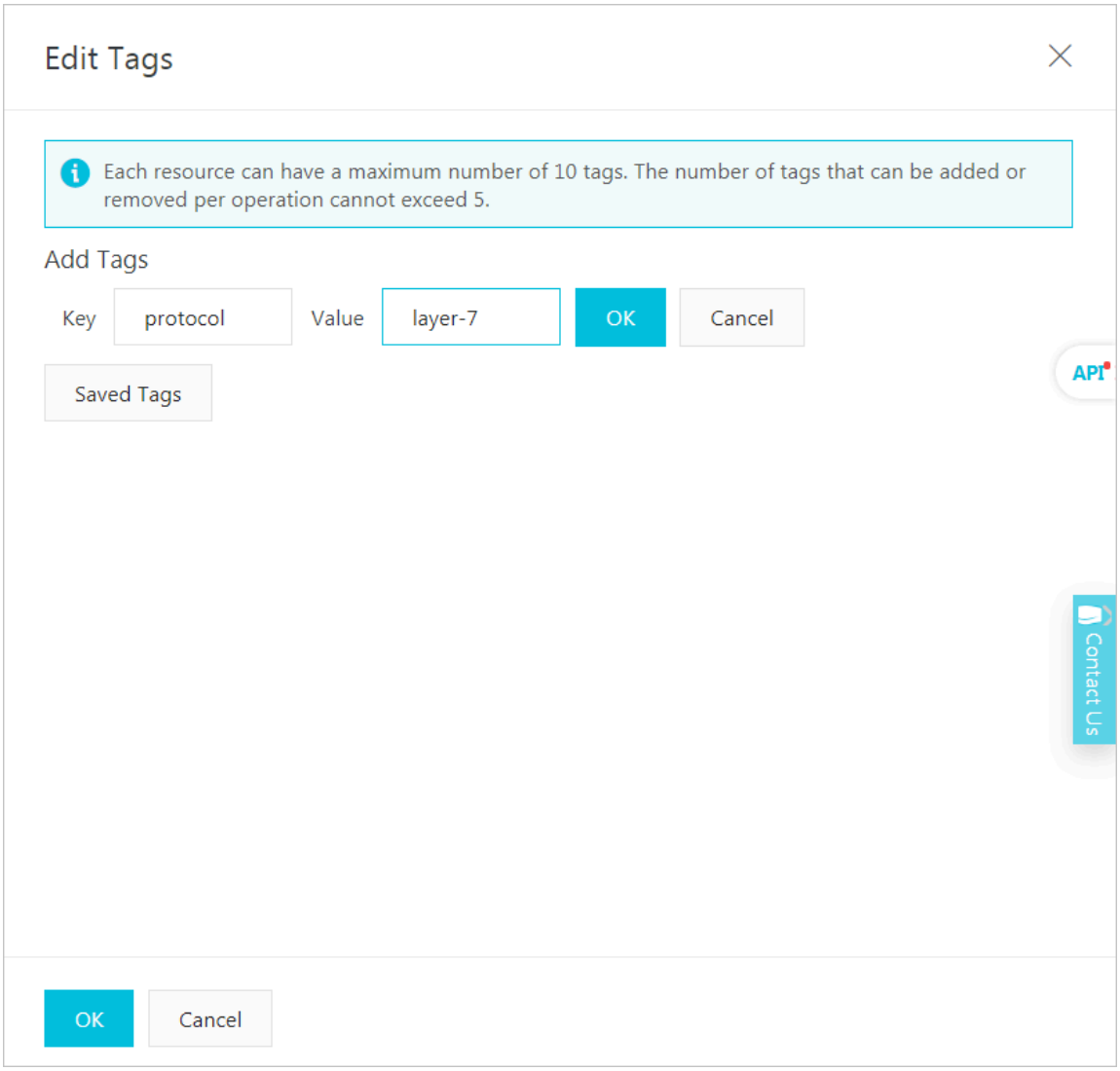
Add a tag

To add a tag, follow these steps:

1. Log on to the SLB console.
2. In the left-side navigation pane, choose Instances > Server Load Balancer.
3. Select a region and find the target SLB instance.
4. In the Actions column, choose More > Edit Tags.



5. On the Edit Tags page, complete these steps:
   a. If there are available tags, click Saved Tags and then select the tag to add.
   b. If you want to create a new tag, on the Edit Tags page, click New Tag, enter the key and value of the new tag, and click OK.



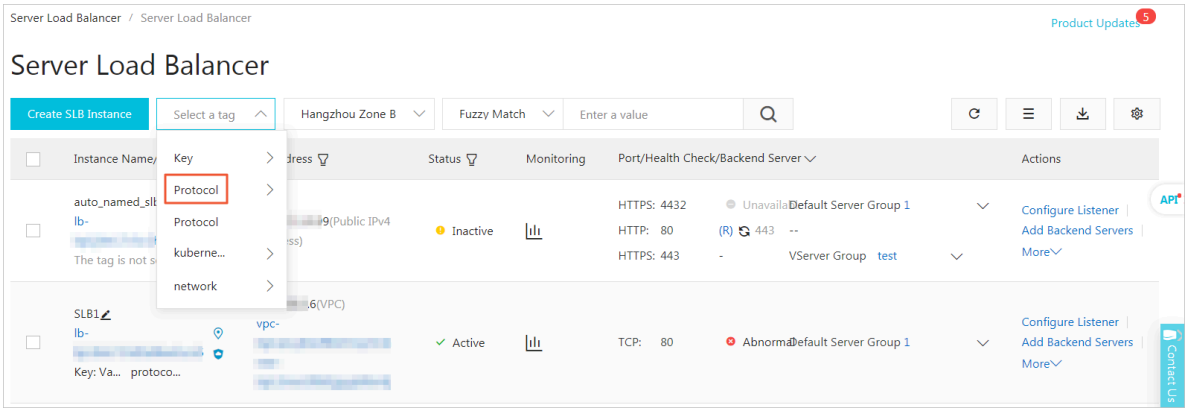
   c. Click OK.

Search instances by using a tag

To search instances by using a tag, follow these steps:

1. Log on to the SLB console.
2. In the left-side navigation pane, choose Instances > Server Load Balancer.
3. Select a region.

4. Click Select a tag, and select the tag to be used as the search criteria.



5. To clear the filter criteria, you can click the delete icon next to the selected tag.

Delete a tag

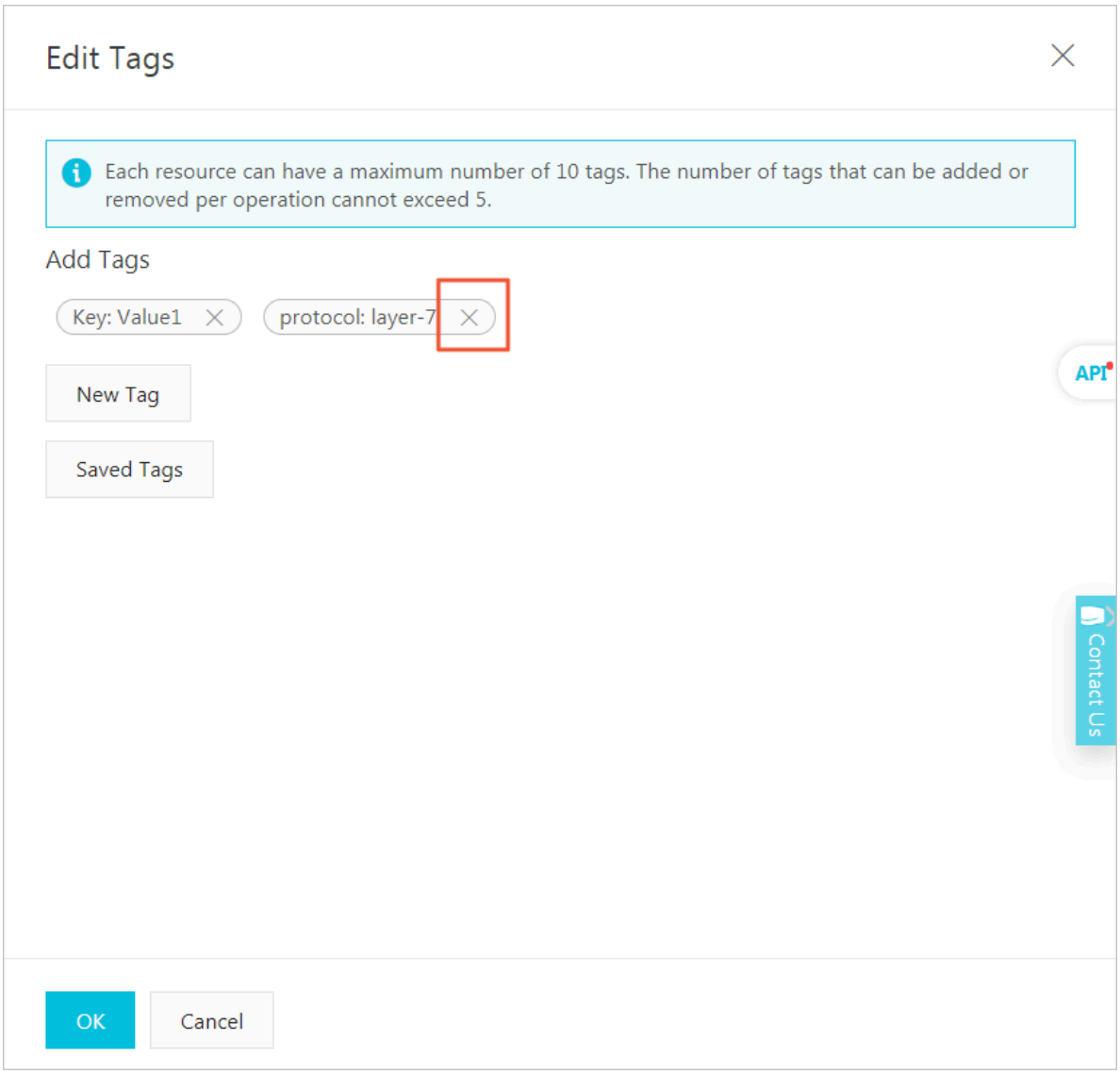
SLB does not support deleting tags of multiple instances in batches. You can remove the tags of only one instance at a time.

To delete a tag, follow these steps:

1. Log on to the SLB console.
2. In the left-side navigation pane, choose Instances > Server Load Balancer.
3. Select a region and find the target instance.
4. In the Actions column, choose More > Edit Tags.
5. On the Edit Tags page, click the delete icon next to the tag to be removed, and then click OK.

Note:

If a tag is removed from one instance and is not associated with any other instances, the tag is removed from the system.



1.9 Renew an expiring instance

This topic describes how to renew an expiring Server Load Balancer (SLB) instance. If an SLB instance has an overdue payment, it is added to the list of expiring instances and, if not handled, released.

Context

If you do not renew an expiring instance, the process by which the instance is released is as follows:

1. The SLB instance runs normally for 24 hours after an overdue payment is detected.

2. If after 24 hours the payment is not settled, the SLB instance is stopped and locked , and added to the list of expiring instances, but not released.
3. If after seven days the payment is not settled, the SLB instance is released.

Procedure

1. Log on to the SLB console.
2. Choose Instances > Expiring Instances.
3. View detailed information of overdue instances.
4. Click Renew in the Actions column of the target SLB instance, then the instance will be added back to the Server Load Balancer list.

1.10 Change the instance specification

You can change a shared-performance instance to a guaranteed-performance instance, or modify the specification of a guaranteed-performance instance.

Before you modify the instance specification, note the following:

- When you change a shared-performance instance to a guaranteed-performance instance, a brief disconnection of service may occur for 10 to 30 seconds.

  We recommend that you change the instance type in a low-traffic period, or use DNS to schedule services to other SLB instances before you change the instance type.
- After you change a shared-performance instance to a guaranteed-performance instance, you cannot change it back.

  You can use the simple I (slb. s1.small) specification after you change a shared-performance instance to a guaranteed-performance instance. This specification is free of charge.
- Intranet SLB instances only support traffic-based billing and cannot be changed to instances that are billed based on bandwidth.

Change the specification of a Pay-As-You-Go instance

To change the specification of a Pay-As-You-Go instance, follow these steps:

1. Log on to the SLB console.
2. Select the region of the target SLB instance.
3. Find the target instance, choose More > Change Specification.

4. In the Configuration upgrade section, select a new specification, and complete the payment.

1.11 Manage idle instances

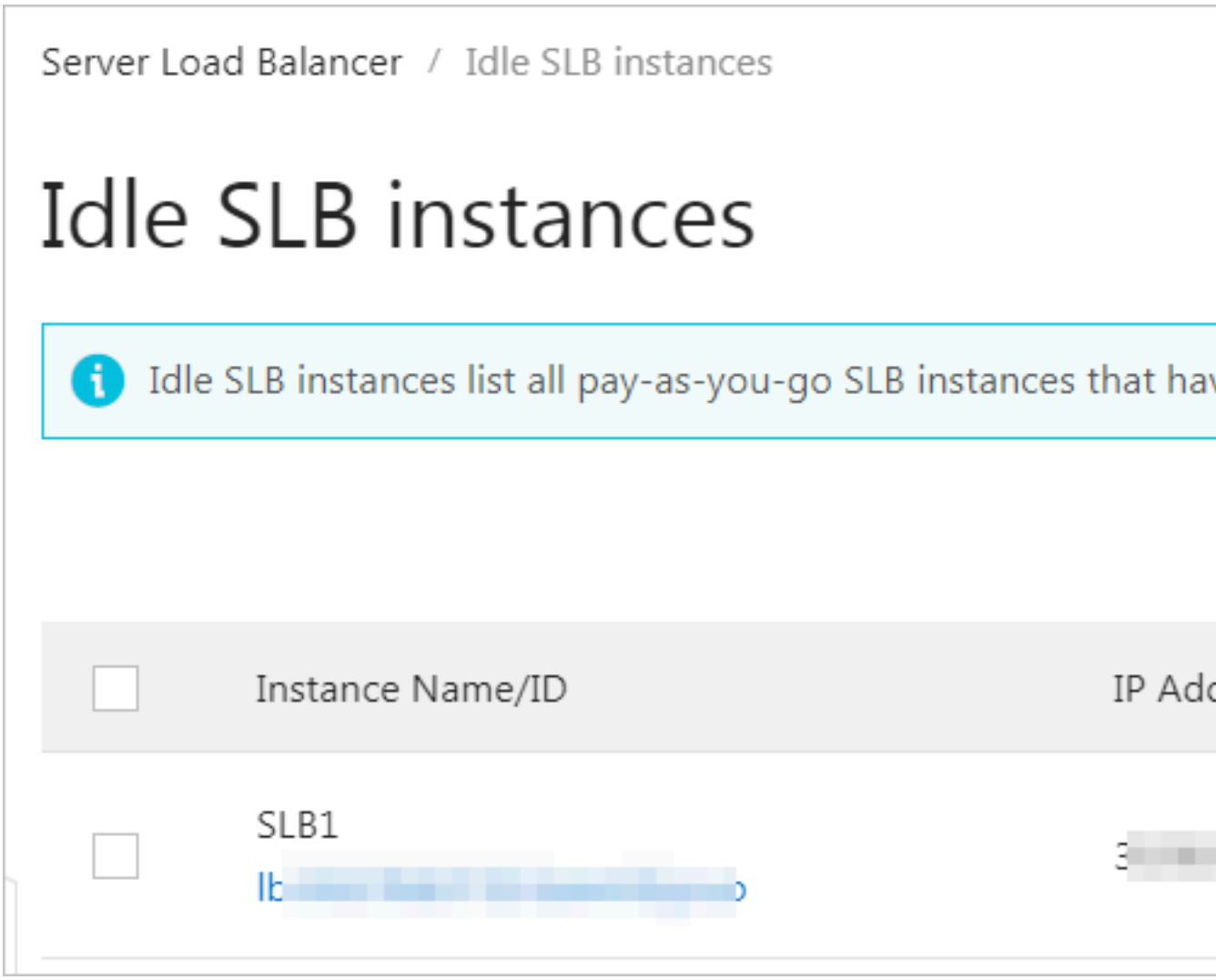
This topic describes how to use the Server Load Balancer (SLB) console to display the Pay-As-You-Go instances that have been idle for more than seven days.

Procedure

1. Log on to the Server Load Balancer console.
2. In the left-side navigation pane, choose SLB Lab > Idle SLB instances.
3. On the Idle SLB instance 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.
4. To release an idle instance, click Release from the Actions column to immediately release the instance.

 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.



2 Listeners

2.1 Listener overview

After creating a Server Load Balancer (SLB) instance, you need to configure a listener for it. The listener checks connection requests and then distributes the requests to backend servers according to configured rules.

Alibaba Cloud provides Layer-4 (TCP and UDP protocols) and Layer-7 (HTTP and HTTPS protocols) load balancing services. Select the protocol based on your needs.

Protocol	Description	Scenario
TCP	· A connection-oriented protocol. A reliable connection must be established before data can be sent and received. · Source address-based session persistence. · The source address is available at the network layer. · Fast data transmission.	· Applicable to scenarios where high transmission reliability and data accuracy are required, but some flexibility regarding network latency is permitted, such as file transmission, sending or receiving emails, and remote logons. · Web applications that have no special requirements. For more information, see Add a TCP listener.
UDP	· A non-connection-oriented protocol. UDP directly transmits data packets instead of making a three-way handshake with the other party before sending data. It does not provide error recovery and data re-transmission. · Fast data transmission, but the reliability is relatively low.	Applicable to scenarios with preference to real-time content over reliability, such as video chats and real-time financial quotations. For more information, see Add a UDP listener.

Protocol	Description	Scenario
HTTP	· An application layer protocol mainly used to package data. · Cookie-based session persistence. · Use X-Forward-For to obtain source IP addresses.	Applicable to applications that need to recognize data content , such as web applications and small-sized mobile games. For more information, see Add an HTTP listener.
HTTPS	· Encrypted data transmission that prevents unauthorized access. · Unified certificate management service. You can upload certificates to SLB and decryption operations are completed directly on SLB.	Applications that require encrypted transmission. For more information, see Add an HTTPS listener.

Note:
HTTP/2 and WSS/WS protocols are supported by all regions now. For more information, see HTTP/2 support FAQ and WS and WSS support FAQs.

2.2 Add a TCP listener

This topic describes how to add a TCP listener to a Server Load Balancer (SLB) instance. The TCP protocol is applicable to scenarios with high requirements on reliability and data accuracy but with tolerance for low speed, such as file transmission, sending or receiving emails, and remote logons. You can add a TCP listener to forward requests from the TCP protocol.

Prerequisites

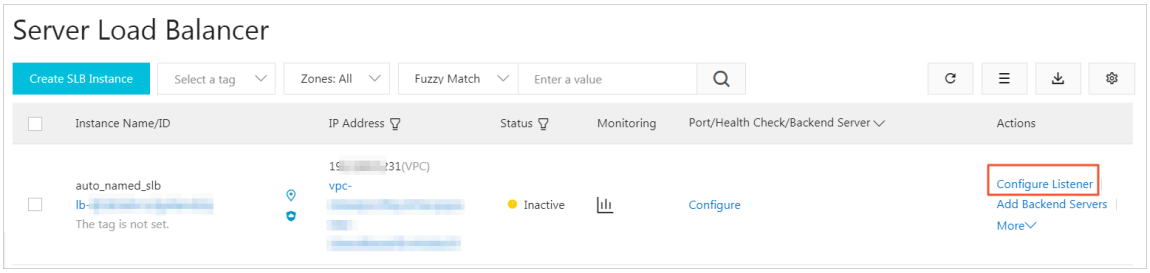
Create an SLB instance.

Step 1 Open the listener configuration wizard

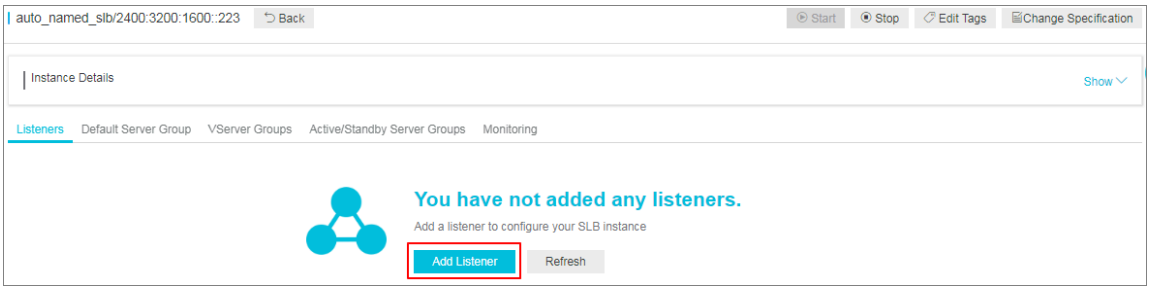
To open the listener configuration wizard, follow these steps:

1. Log on to the SLB console.
2. In the left-side navigation pane, choose Instances > Server Load Balancer.
3. Select the region of the target SLB instance.

4. Select one of the following methods to open the listener configuration wizard:
   - On the Server Load Balancer page, find the target SLB instance and then click Configure Listener in the Actions column.



   - On the Server Load Balancer page, click the ID of the target SLB instance. On the Listeners page, click Add Listener.



Step 2 Configure the TCP listener

To configure the TCP listener, follow these steps:

1. On the Protocol and Listener page, configure the TCP listener according to the following information.

Configuration	Description
Select Listener Protocol	Select the protocol type of the listener. In this topic, select TCP.

Configuration	Description
Listening Port	The listening port used to receive requests and forward the requests to backend servers. The port number is in the range of 1 to 65535. Note: UDP and TCP listener port numbers can be the same in the following regions. However, you must first apply for the privilege to use the beta function of configuring the same ports in TCP/UDP listeners on the Quota Management page of the SLB console. In other cases, the listener port numbers must be unique. · UAE (Dubai) · Australia (Sydney) · UAE (Dubai) · UK (London) · Germany (Frankfurt) · US (Silicon Valley) · USA (Virginia) · Indonesia (Jakarta) · Japan (Tokyo) · India (Mumbai) · Singapore · Malaysia (Kuala Lumpur) · Hong Kong · China (Shenzhen) · China (Hohhot) · China (Qingdao) · China (Chengdu) · China (Zhangjiakou) · China (Shanghai)
Advanced configurations	

Configuration	Description
Scheduling Algorithm	SLB supports four scheduling algorithms: round robin, weighted round robin (WRR), weighted least connections (WLC), and consistent hash. · Weighted Round-Robin (WRR): Backend servers with higher weights receive more requests. · Round-Robin (RR): Requests are evenly and sequentially distributed to backend servers. · Weighted Least Connections (WLC): A server with a higher weight receives more requests. When the weight values of two backend servers are the same, the backend server with a smaller number of connections is more likely to be polled. · Consistent Hash (CH): - Source IP: the consistent hash based on source IP addresses. Requests from the same source IP address are scheduled to the same backend server. - Tuple: the consistent hash based on four factors: source IP address + destination IP address + source port + destination port. The same streams are scheduled to the same backend server. Note: Currently, the Consistent Hash (CH) algorithm is only supported in the following regions: - Japan (Tokyo) - Australia (Sydney) - Malaysia (Kuala Lumpur) - Indonesia (Jakarta) - Germany (Frankfurt) - US (Silicon Valley) - US (Virginia) - UAE (Dubai) - China (Hohhot)

Configuration	Description
Enable Session Persistence	Select whether to enable session persistence. If you enable session persistence, all session requests from the same client are sent to the same backend server. For TCP listeners, session persistence is based on IP addresses. Requests from the same IP address are forwarded to the same backend server.
Enable Access Control	Select whether to enable the access control function.
Access Control Method	Select an access control method after you enable the access control function: · Whitelist: Only requests from IP addresses or CIDR blocks in the selected access control list are forwarded. It applies to scenarios where the application only allows access from some specific IP addresses. Enabling a whitelist poses some risks to your services. After a whitelist is configured, only the IP addresses in the list can access the SLB listener . If you enable the whitelist without adding any IP addresses in the corresponding access control list, all requests are forwarded. · Blacklist: Requests from IP addresses or CIDR blocks in the selected access control list are not forwarded. It applies to scenarios where the application only denies access from some specific IP addresses. If you enable a blacklist without adding any IP addresses in the corresponding access control list, all requests are forwarded.

Configuration	Description
Access Control List	Select an access control list as the whitelist or the blacklist. Note: An IPv6 instance can only be associated with IPv6 access control lists and an IPv4 instance can only be associated with IPv4 access control lists. For more information, see Configure an access control list.
Enable Peak Bandwidth Limit	Select whether to configure the listening bandwidth. If the SLB instance is charged based on bandwidth , you can set different peak bandwidth values for different listeners to limit the traffic passing through the listeners. The sum of the peak bandwidth values of all listeners under an SLB instance cannot exceed the bandwidth value of that SLB instance. By default, all listeners share the bandwidth of the SLB instance. Note: SLB instances billed by traffic have no peak bandwidth limit by default.
Idle Timeout	Specify the idle connection timeout period. Value range: 10 to 900. Unit: seconds.
Listener Name	Enter a name for the TCP listener to be added.
Get Client Source IP Address	Backend servers of a Layer-4 listener can directly obtain the source IP addresses of clients.
Automatically Enable Listener after Creation	Choose whether to start the listener after the listener is configured. The listener is started by default.

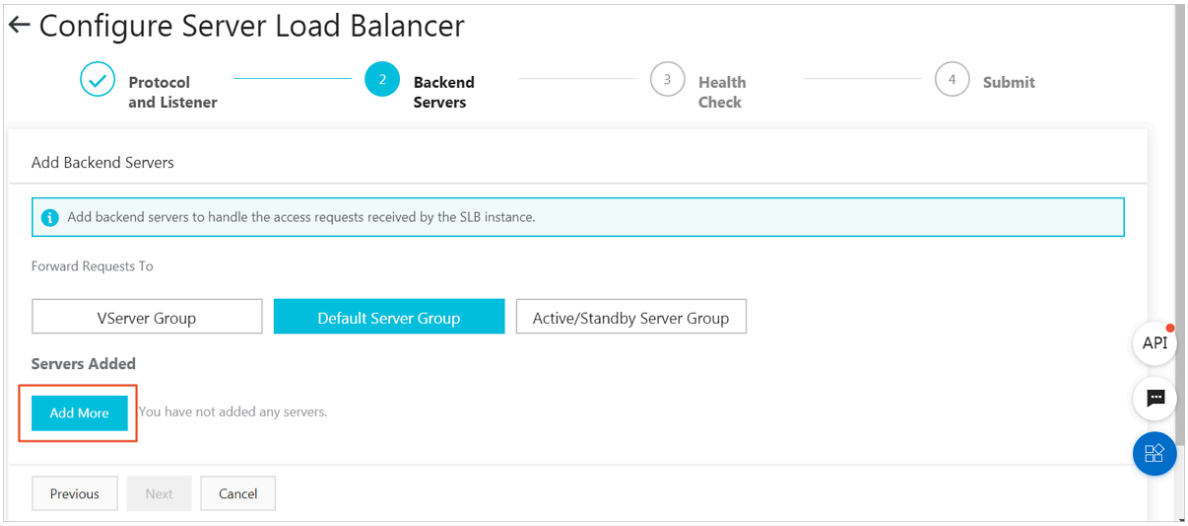
2. Click Next.

Step 3 Add backend servers

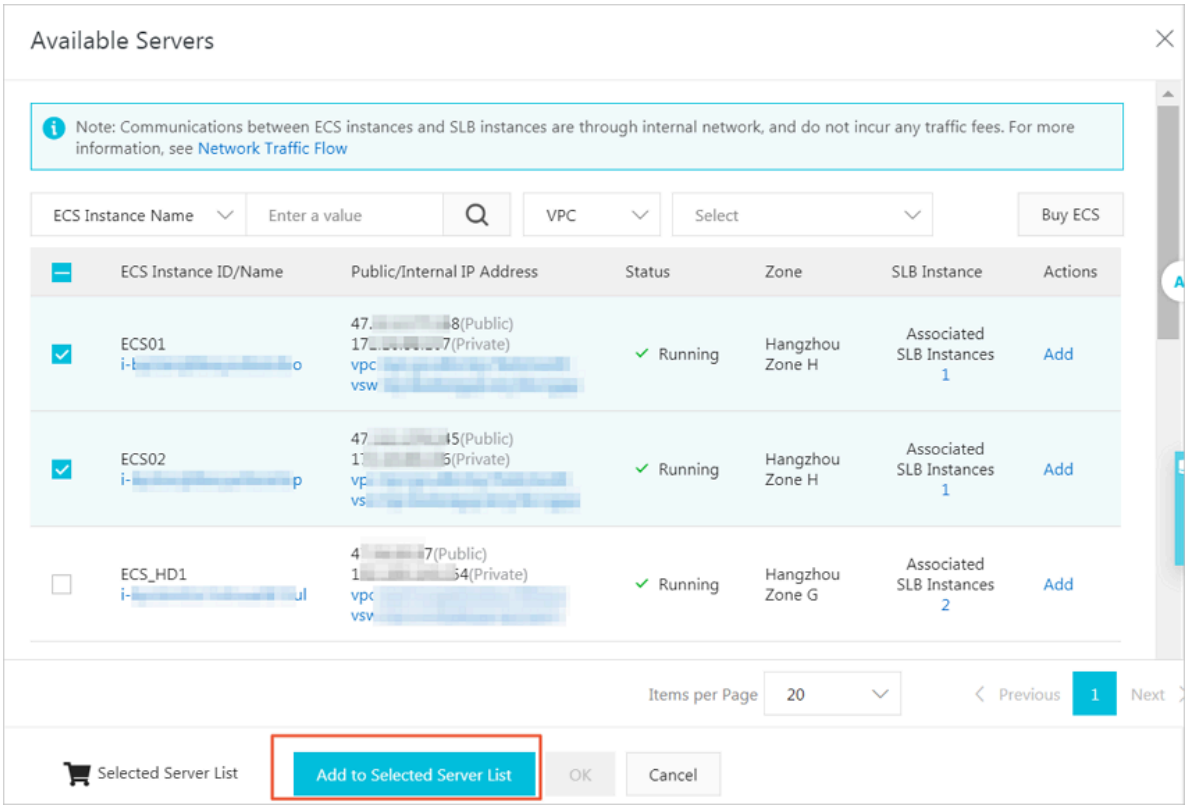
After configuring the listener, you need to add backend servers to process requests. You can use the default server group configured for the SLB instance, or configure a VServer group or an active/standby server group for the listener. For more information, see Backend server overview.

In this topic, use the default server group.

1. Select Default Server Group and then click Add More.



2. Select the ECS instances to add, and then click Next: Set Weight and Port.



3. Configure ports and weights for the added backend servers.

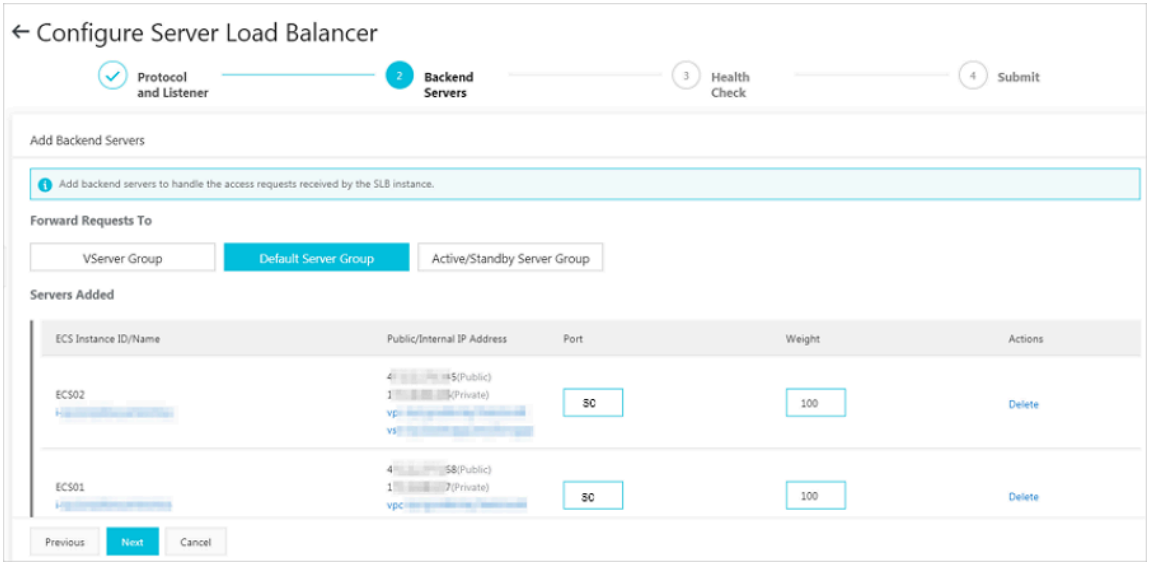
   - Port

     The port opened on the backend server (ECS instance) to receive requests. The port number is in the range of 1 to 65535. Ports of backend servers can be the same in an SLB instance.

   - Weight

     The weight of the backend server (ECS instance). An ECS instance with a higher weight receives more requests.

     > **Note:**
     > If the weight is set to 0, no requests will be sent to the ECS instance.



4. Click Next.

Step 4 Configure health checks

SLB checks the service availability of backend servers by performing health checks. The health check function improves the overall availability of your services and avoids the impact of backend server failures. Click Modify to change health check configurations. For more information, see Configure health checks.

Step 5 Submit the configurations

To confirm the listener configurations, follow these steps:

1. On the Submit page, check the listener configurations. You can click Modify to change the configurations.

2. Click Submit.
3. On the Submit page, click OK after the configurations are successful.

After the configurations are successful, you can view the created listener on the Listeners page.

Related operations

- Configure health checks
- Manage a default server group
- Manage a VServer group
- Manage an active/standby server group
- Configure access control

2.3 Add a UDP listener

This topic describes how to add a UDP listener to a Server Load Balancer (SLB) instance. You can add a UDP listener to forward requests from the UDP protocol.

Limits

Note the following before you add a UDP listener:

- Currently, ports 250, 4789, and 4790 are reserved.
- Currently, fragmented packets are not supported.
- UDP listeners of an SLB instance of the classic network do not support viewing source IP addresses.
- The following operations require five minutes to take effect if they are performed in a UDP listener:
  - Remove backend ECS instances.
  - Set the weight of a backend server to 0 after the backend server is declared as unhealthy.
- Because IPv6 has a longer IP header than IPv4, when you use a UDP listener on an IPv6 SLB instance, you must ensure that the MTU of the NIC on the backend server (ECS instance) communicating with the SLB instance is not greater than 1480 (some applications need to synchronizing its configuration files based on this MTU value). Otherwise, packets may be discarded because they are too large.

  If you use a TCP/HTTP/HTTPS listener, no additional configurations are required because the TCP protocol supports MSS auto-negotiation.

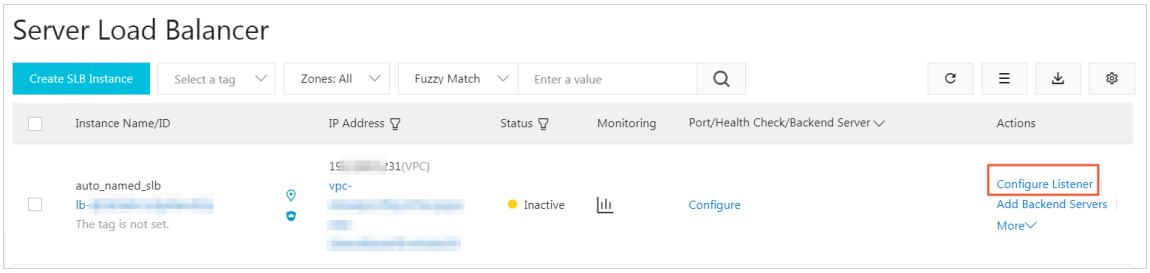
Prerequisites

Create an SLB instance.

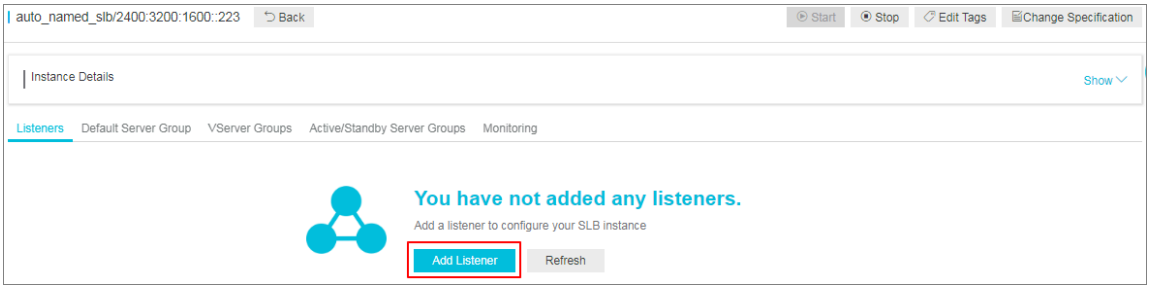
Step 1 Open the listener configuration wizard

To open the listener configuration wizard, follow these steps:

1. Log on to the SLB console.
2. In the left-side navigation pane, choose Instances > Server Load Balancer.
3. Select the region of the target SLB instance.
4. Select one of the following methods to open the listener configuration wizard:
   - On the Server Load Balancer page, find the target SLB instance and then click Configure Listener in the Actions column.



   - On the Server Load Balancer page, click the ID of the target SLB instance. On the Listeners page, click Add Listener.



Step 2 Configure the UDP listener

To configure the UDP listener, follow these steps:

1. On the Protocol and Listener page, configure the UDP listener according to the following information.

Configuration	Description
Select Listener Protocol	Select the protocol type of the listener. In this topic, select UDP.

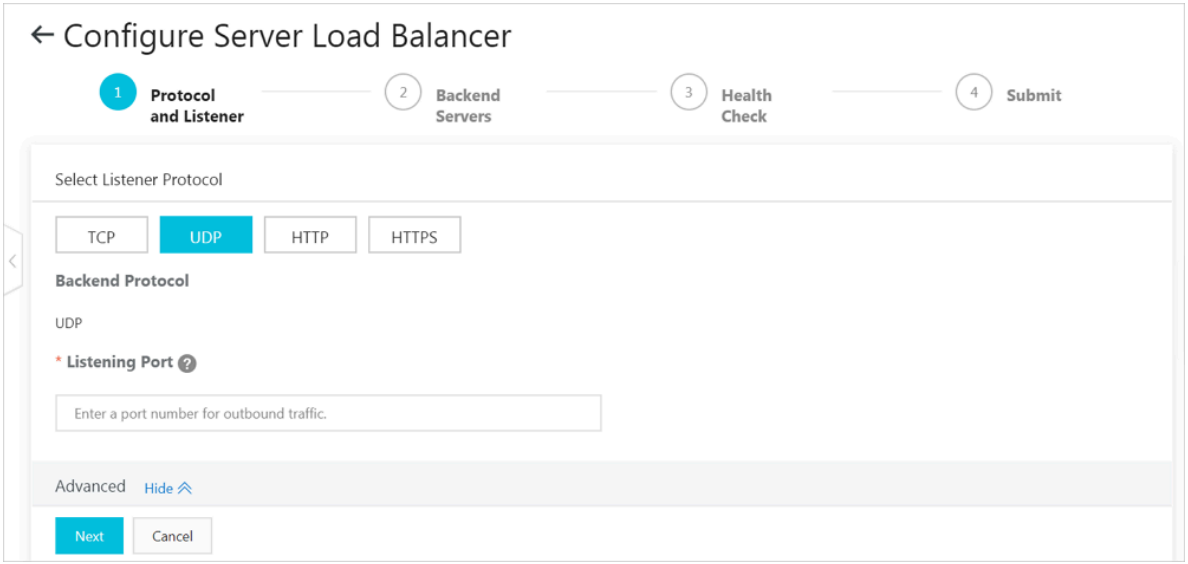
Configuration	Description
Listening Port	The listening port used to receive requests and forward the requests to backend servers. The port number is in the range of 1 to 65535. Note: UDP and TCP listener port numbers can be the same in the following regions. However, you must first apply for the privilege to use the beta function of configuring the same ports in TCP/UDP listeners on the Quota Management page of the SLB console. In other cases, the listener port numbers must be unique. · UAE (Dubai) · Australia (Sydney) · UAE (Dubai) · UK (London) · Germany (Frankfurt) · US (Silicon Valley) · USA (Virginia) · Indonesia (Jakarta) · Japan (Tokyo) · India (Mumbai) · Singapore · Malaysia (Kuala Lumpur) · Hong Kong · China (Shenzhen) · China (Hohhot) · China (Qingdao) · China (Chengdu) · China (Zhangjiakou) · China (Shanghai)
Advanced configurations	

Configuration	Description
Scheduling Algorithm	SLB supports four scheduling algorithms: round robin, weighted round robin (WRR), weighted least connections (WLC), and consistent hash. · Weighted Round-Robin (WRR): Backend servers with higher weights receive more requests. · Round-Robin (RR): Requests are evenly and sequentially distributed to backend servers. · Weighted Least Connections (WLC): A server with a higher weight receives more requests. When the weight values of two backend servers are the same, the backend server with a smaller number of connections is more likely to be polled. · Consistent Hash (CH): - Source IP: the consistent hash based on source IP addresses. Requests from the same source IP address are scheduled to the same backend server. - Tuple: the consistent hash based on four factors: source IP address + destination IP address + source port + destination port. The same streams are scheduled to the same backend server. - QUIC ID: the consistent hash based on the QUIC Connection ID. The same QUIC Connection IDs are scheduled to the same backend server. Notice: The QUIC protocol is rapidly evolving. The algorithm is based on draft-ietf-quic-transport-10 and does not guarantee the compatibility of all QUIC versions. We recommend that you do enough tests before using it for the production environment. Note: Currently, the Consistent Hash (CH) algorithm is only supported in the following regions: - Japan (Tokyo) - Australia (Sydney) - Malaysia (Kuala Lumpur) - Indonesia (Jakarta) - Germany (Frankfurt)

Configuration	Description
Enable Access Control	Select whether to enable the access control function.
Access Control Method	Select an access control method after you enable the access control function: · Whitelist: Only requests from IP addresses or CIDR blocks in the selected access control list are forwarded. It applies to scenarios where the application only allows access from some specific IP addresses. Enabling a whitelist poses some risks to your services. After a whitelist is configured, only the IP addresses in the list can access the listener. If you enable the whitelist without adding any IP addresses in the corresponding access control list, all requests are forwarded. · Blacklist: Requests from IP addresses or CIDR blocks in the selected access control list are not forwarded. It applies to scenarios where the application only denies access from some specific IP addresses. If you enable a blacklist without adding any IP addresses in the corresponding access control list, all requests are forwarded.
Access Control List	Select an access control list as the whitelist or the blacklist.  Note: An IPv6 instance can only be associated with IPv6 access control lists and an IPv4 instance can only be associated with IPv4 access control lists. For more information, see Configure an access control list.

Configuration	Description
Enable Peak Bandwidth Limit	Select whether to configure the listener bandwidth. If the SLB instance is charged based on bandwidth, you can set different peak bandwidth values for different listeners to limit the traffic passing through the listeners. The sum of the peak bandwidth values of all listeners under an SLB instance cannot exceed the bandwidth value of that SLB instance. By default, all listeners share the bandwidth of the SLB instance. Note: SLB instances billed by traffic have no bandwidth peak limit by default.
Get Client Source IP Address	Backend servers of a UDP listener can directly obtain source IP addresses of clients. Note: UDP listeners of an SLB instance of the classic network do not support viewing source IP addresses.
Automatically Enable Listener After Creation	Choose whether to start the listener after the listener is configured. The listener is started by default.

2. Click Next.

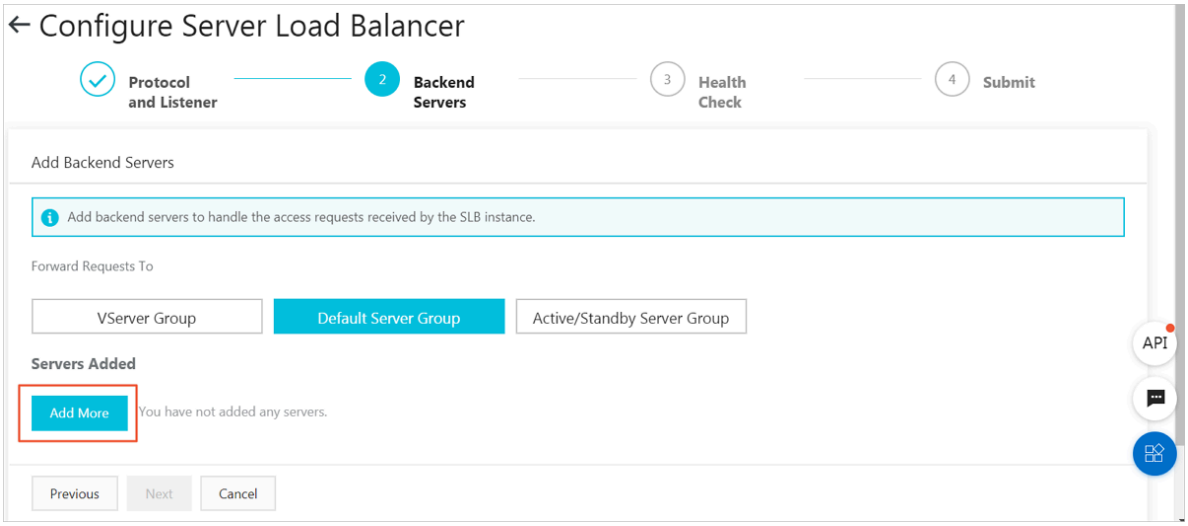


Step 3 Add backend servers

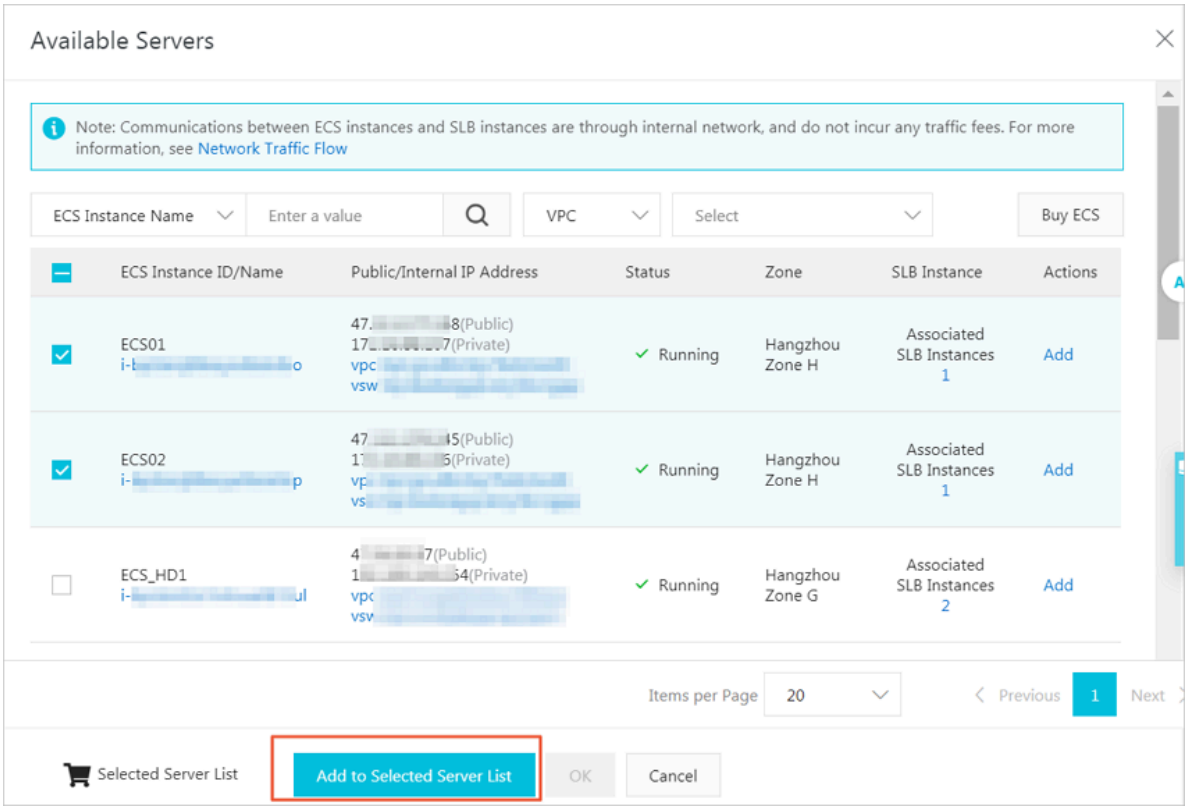
After configuring the listener, you need to add backend servers to process requests. You can use the default server group configured for the SLB instance, or configure a VServer group or an active/standby server group for the listener. For more information, see Backend server overview.

In this topic, use the default server group.

1. Select Default Server Group and then click Add More.



2. Select the ECS instances to add, and then click Next: Set Weight and Port.



3. Configure ports and weights for the added backend servers.

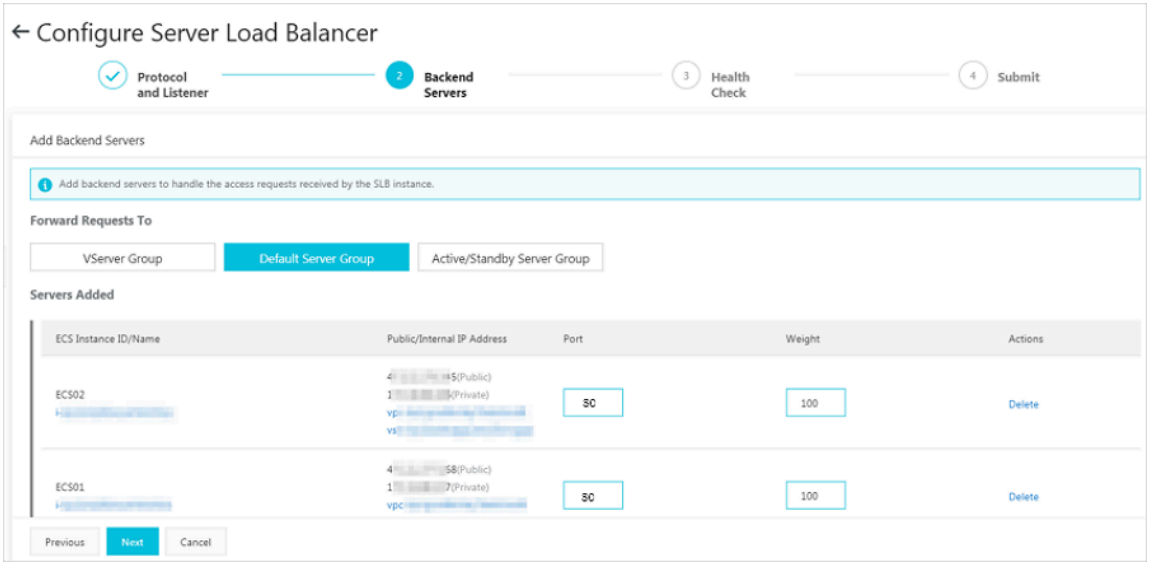
   - Port

     The port opened on the backend server (ECS instance) to receive requests. The port number is in the range of 1 to 65535. Ports of backend servers can be the same in an SLB instance.

   - Weight

     The weight of the backend server (ECS instance). An ECS instance with a higher weight receives more requests.

     > **Note:**
     > If the weight is set to 0, no requests will be sent to the ECS instance.



4. Click Next.

Step 4 Configure health checks

SLB checks the service availability of backend servers by performing health checks. The health check function improves the overall availability of your services and avoids the impact of backend server failures. Click Modify to change health check configurations. For more information, see Configure health checks.

Step 5 Submit the configurations

To confirm the listener configurations, follow these steps:

1. On the Submit page, check the listener configurations. You can click Modify to change the configurations.

2. Click Submit.
3. On the Submit page, click OK after the configurations are successful.

After the configurations are successful, you can view the created listener on the Listeners page.

Related operations

- Configure health checks
- Manage a default server group
- Manage a VServer group
- Manage an active/standby server group
- Configure access control

2.4 Add an HTTP listener

This topic describes how to add an HTTP listener to a Server Load Balancer (SLB) instance. You can add an HTTP listener to forward requests from the HTTP protocol.

Prerequisites

Create an SLB instance.

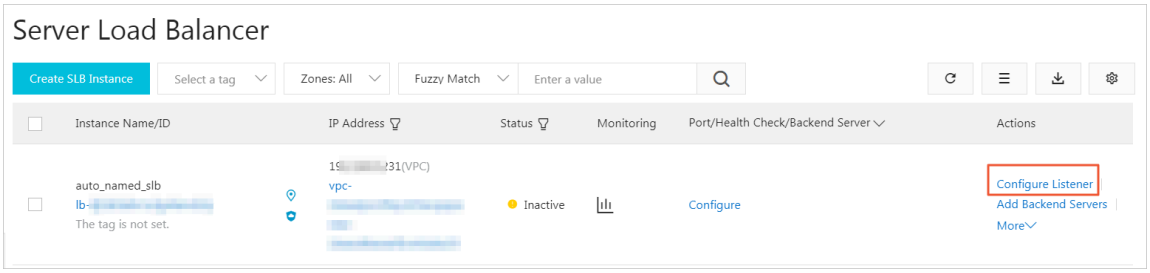
Step 1 Open the listener configuration wizard

To open the listener configuration wizard, follow these steps:

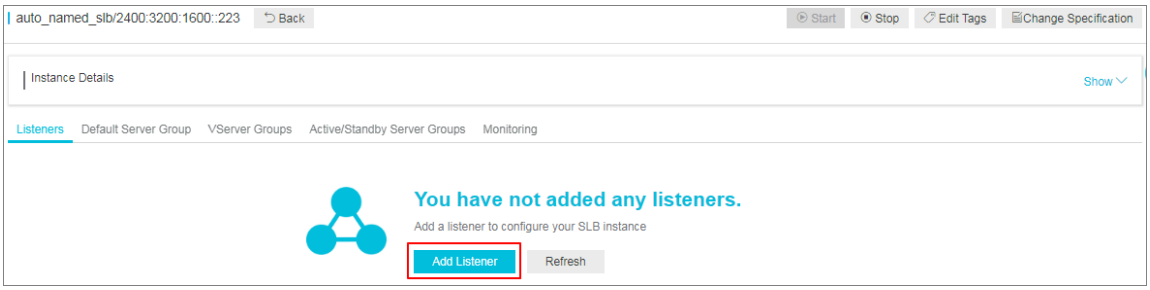
1. Log on to the SLB console.
2. In the left-side navigation pane, choose Instances > Server Load Balancer.
3. Select the region of the target SLB instance.

4. Select one of the following methods to open the listener configuration wizard:

   - On the Server Load Balancer page, find the target SLB instance and then click Configure Listener in the Actions column.



   - On the Server Load Balancer page, click the ID of the target SLB instance. On the Listeners page, click Add Listener.



Step 2 Configure the HTTP listener

To configure the HTTP listener, follow these steps:

1. On the Protocol and Listener page, configure the HTTP listener according to the following information:

Configuration	Description
Select Listener Protocol	Select the protocol type of the listener. In this topic, select HTTP.
Listening Port	The listening port used to receive requests and forward the requests to backend servers. The port number is in the range of 1 to 65535. **Note:** The listening port must be unique in an SLB instance.
Advanced configurations	

Configuration	Description
Scheduling Algorithm	SLB supports three scheduling algorithms: round robin, weighted round robin (WRR), and weighted least connections (WLC). · Weighted Round-Robin (WRR): Backend servers with higher weights receive more requests. · Round-Robin (RR): Requests are evenly and sequentially distributed to backend servers. · Weighted Least Connections (WLC): A backend server with a higher weight receives more requests. When the weight values of two backend servers are the same, the backend server with a smaller number of connections is more likely to be polled.
Redirection	Select whether to forward traffic of the HTTP listener to an HTTPS listener. Note: Before you enable the redirection function, make sure that you have created an HTTPS listener.

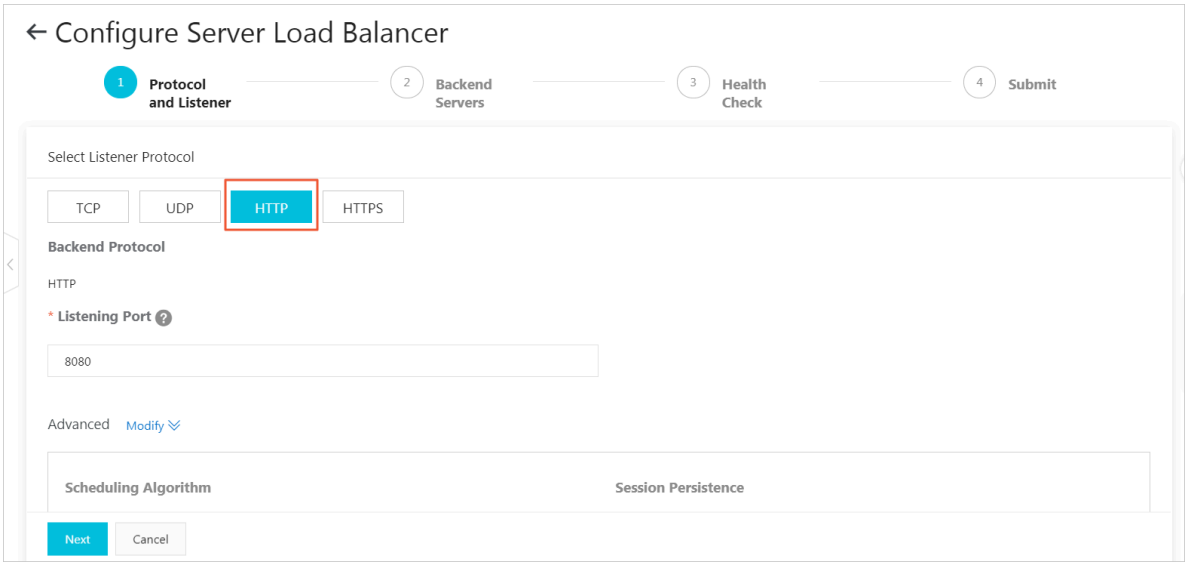
Configuration	Description
Session Persistence	Select whether to enable session persistence. After you enable session persistence, all session requests from the same client are sent to the same backend server. HTTP session persistence is based on cookies. The following two methods are supported: · Insert cookie: You only need to specify the cookie timeout period. SLB adds a cookie to the first response from the backend server (inserts SERVERID in the HTTP and HTTPS response packet). The next request will contain the cookie and the listener will distribute the request to the same backend server. · Rewrite cookie: You can set the cookie to insert to the HTTP or HTTPS response according to your needs. You must maintain the timeout period and lifecycle of the cookie on the backend server. SLB will overwrite the original cookie when it discovers that a new cookie is set. The next time the client carries the new cookie to access SLB, the listener will distribute the request to the recorded backend server. For more information, see Session persistence.
Enable Access Control	Select whether to enable the access control function.

Configuration	Description
Access Control Method	Select an access control method after you enable the access control function: · Whitelist: Only requests from IP addresses or CIDR blocks in the selected access control list are forwarded. It applies to scenarios where the application only allows access from some specific IP addresses. Enabling a whitelist poses some risks to your services. After a whitelist is configured, only the IP addresses in the list can access the listener. If you enable the whitelist without adding any IP addresses in the corresponding access control list, all requests are forwarded. · Blacklist: Requests from IP addresses or CIDR blocks in the selected access control list are not forwarded. It applies to scenarios where the application only denies access from some specific IP addresses. If you enable a blacklist without adding any IP addresses in the corresponding access control list, all requests are forwarded.
Access Control List	Select an access control list as the whitelist or the blacklist. Note: An IPv6 instance can only be associated with IPv6 access control lists and an IPv4 instance can only be associated with IPv4 access control lists. For more information, see Configure an access control list.

Configuration	Description
Enable Peak Bandwidth Limit	Select whether to configure the listener bandwidth. If the SLB instance is billed by bandwidth, you can set different peak bandwidth values for different listeners to limit the traffic passing through the listeners. The sum of the peak bandwidth values of all listeners under an SLB instance cannot exceed the bandwidth value of that SLB instance. By default, all listeners share the bandwidth of the SLB instance. Note: SLB instances billed by traffic have no peak bandwidth limit by default.
Idle Timeout	Specify the idle connection timeout period. Value range: 1 to 60. Unit: seconds. If no request is received during the specified timeout period, SLB will close the connection and restart the connection when the next request comes. This function is available in all regions.
Request Timeout	Specify the request timeout period. Value range: 1 to 180. Unit: seconds. If no response is received from the backend server during the specified timeout period, SLB will stop waiting and send an HTTP 504 error code to the client . Currently, this function is available in all regions.
Gzip Compression	Choose whether to enable Gzip compression to compress files of specific formats. Now Gzip supports the following file types: text/xml, text/plain, text/css, application/javascript, application/x-javascript, application/rss+xml, application/atom+xml, and application/xml.

Configuration	Description
Add HTTP Header Fields	Select the custom HTTP headers that you want to add: · Use the `X-Forwarded-For` field to retrieve client source IP addresses. · Use the `X-Forwarded-Proto` field to retrieve the listener protocol used by the SLB instance. · Use the `SLB-IP` field to retrieve the public IP address of the SLB instance. · Use the `SLB-ID` field to retrieve the ID of the SLB instance.
Get Client Source IP Address	HTTP listeners use X-Forwarded-For to obtain real IP addresses of clients.
Automatically Enable Listener After Creation	Choose whether to start the listener after the listener is configured. This function is enabled by default.

2. Click Next.

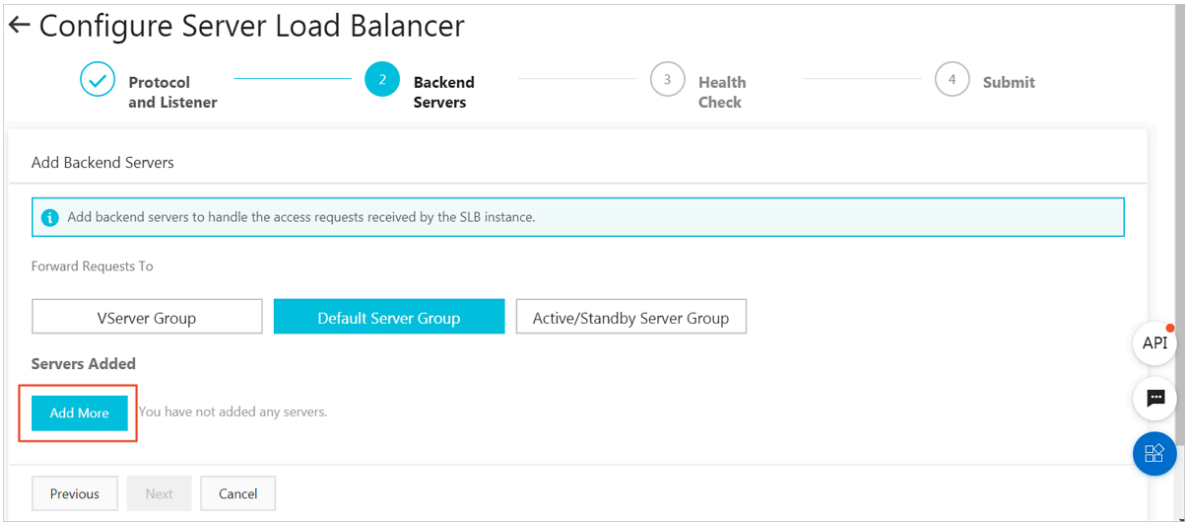


Step 3 Add backend servers

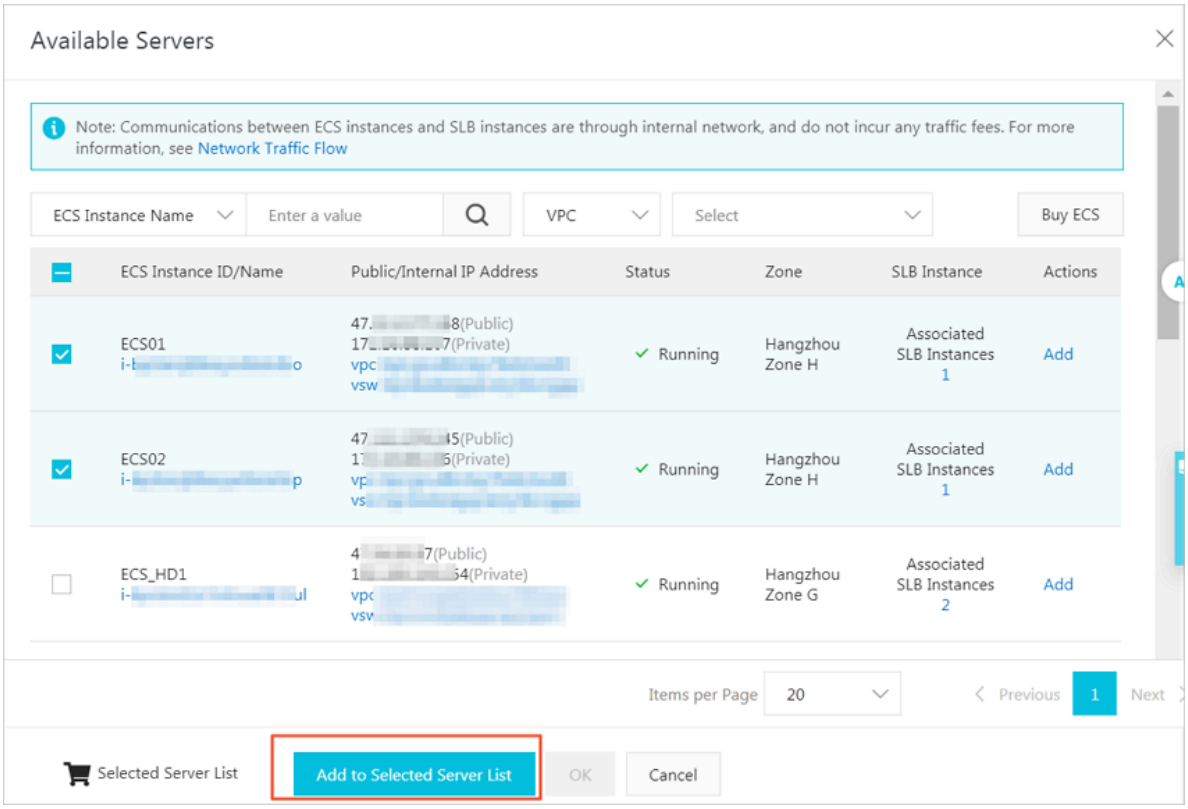
After configuring the listener, you need to add backend servers to process requests. You can use the default server group configured for the SLB instance, or configure a VServer group or an active/standby server group for the listener. For more information, see Backend server overview.

In this topic, use the default server group.

1. Select Default Server Group and then click Add More.



2. Select the ECS instances to add, and then click Next: Set Weight and Port.



3. Configure ports and weights for the added backend servers.

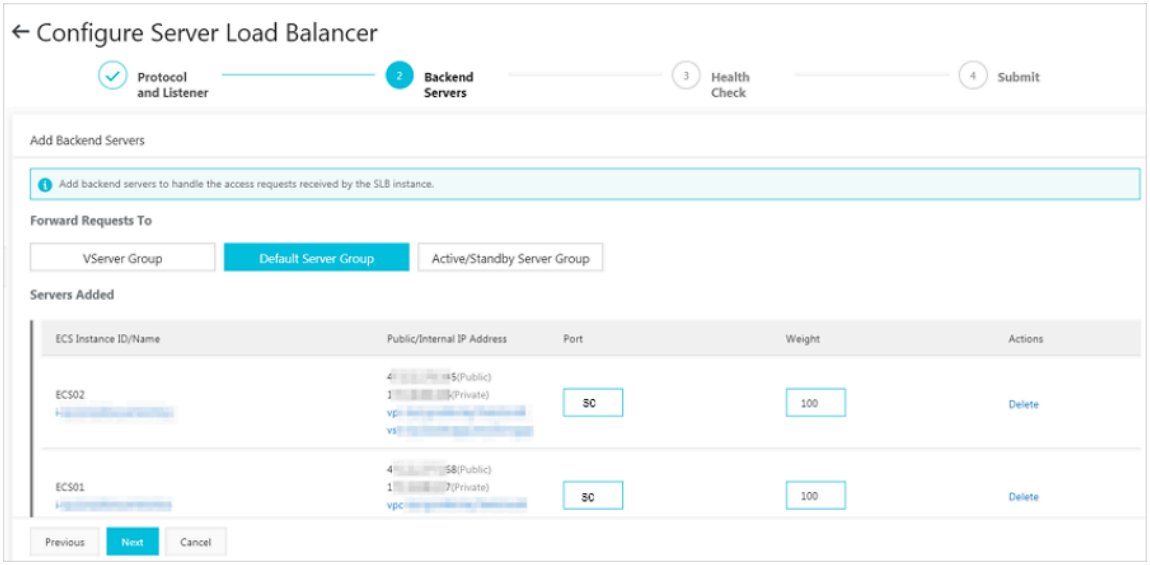
   - Port

     The port opened on the backend server (ECS instance) to receive requests. The port number is in the range of 1 to 65535. Ports of backend servers can be the same in an SLB instance.

   - Weight

     The weight of the backend server (ECS instance). An ECS instance with a higher weight receives more requests.

     > **Note:**
     > If the weight is set to 0, no requests will be sent to the ECS instance.



4. Click Next.

Step 4 Configure health checks

SLB checks the service availability of backend servers by performing health checks. The health check function improves the overall availability of your services and avoids the impact of backend server failures. Click Modify to change health check configurations. For more information, see Configure health checks.

Step 5 Submit the configurations

To confirm the listener configurations, follow these steps:

1. On the Submit page, check the listener configurations. You can click Modify to change the configurations.

2. Click Submit.
3. On the Submit page, click OK after the configurations are successful.

After the configurations are successful, you can view the created listener on the Listeners page.

Related operations

- Configure health checks
- Manage a default server group
- Manage a VServer group
- Manage an active/standby server group
- Configure access control
- Add domain-name based or URL-based forwarding rules
- Manage a domain name extension

2.5 Add an HTTPS listener

This topic describes how to add an HTTP listener to a Server Load Balancer (SLB) instance. You can add an HTTPS listener to forward requests from the HTTPS protocol.

Prerequisites

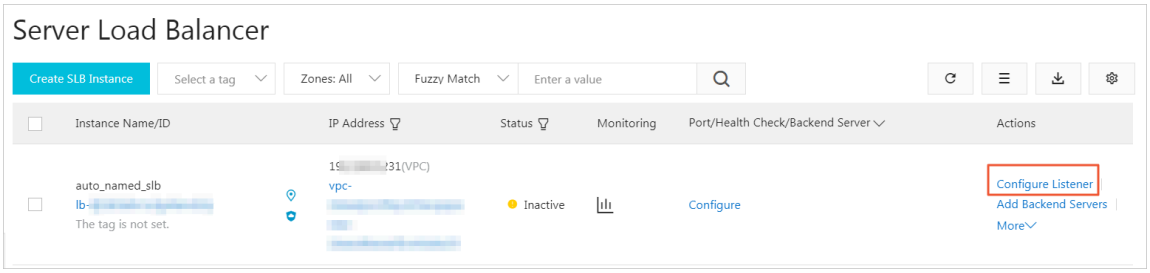
Create an SLB instance.

Step 1 Open the listener configuration wizard

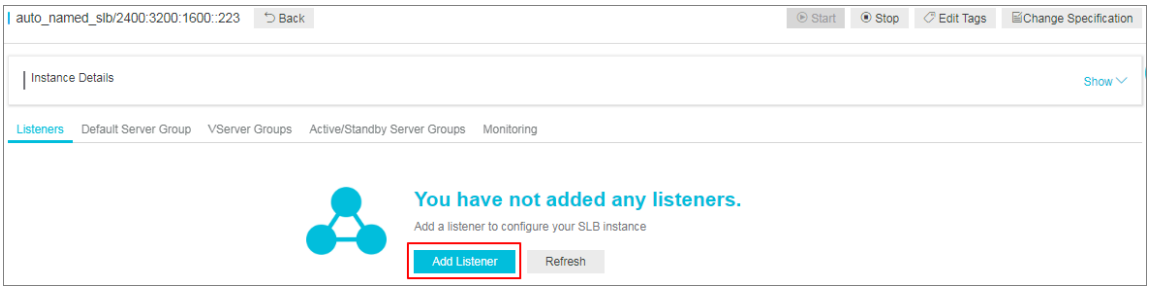
To open the listener configuration wizard, follow these steps:

1. Log on to the SLB console.
2. In the left-side navigation pane, choose Instances > Server Load Balancer.
3. Select the region of the target SLB instance.

4. Select one of the following methods to open the listener configuration wizard:
   - On the Server Load Balancer page, find the target SLB instance and then click Configure Listener in the Actions column.



   - On the Server Load Balancer page, click the ID of the target SLB instance. On the Listeners page, click Add Listener.



Step 2 Configure the HTTPS listener

To configure the HTTPS listener, follow these steps:

1. On the Protocol and Listener page, configure the HTTPS listener according to the following information:

Configuration	Description
Select Listener Protocol	Select the protocol type of the listener. In this topic, select HTTPS.
Listening Port	The listening port used to receive requests and forward the requests to backend servers. The port number is in the range of 1 to 65535. **Note:** The listening port must be unique in an SLB instance.
Advanced configurations	

Configuration	Description
Scheduling Algorithm	SLB supports three scheduling algorithms: round robin, weighted round robin (WRR), and weighted least connections (WLC). · Weighted Round-Robin (WRR): Backend servers with higher weights receive more requests. · Round-Robin (RR): Requests are evenly and sequentially distributed to backend servers. · Weighted Least Connections (WLC): A server with a higher weight receives more requests. When the weight values of two backend servers are the same, the backend server with a smaller number of connections is more likely to be polled.

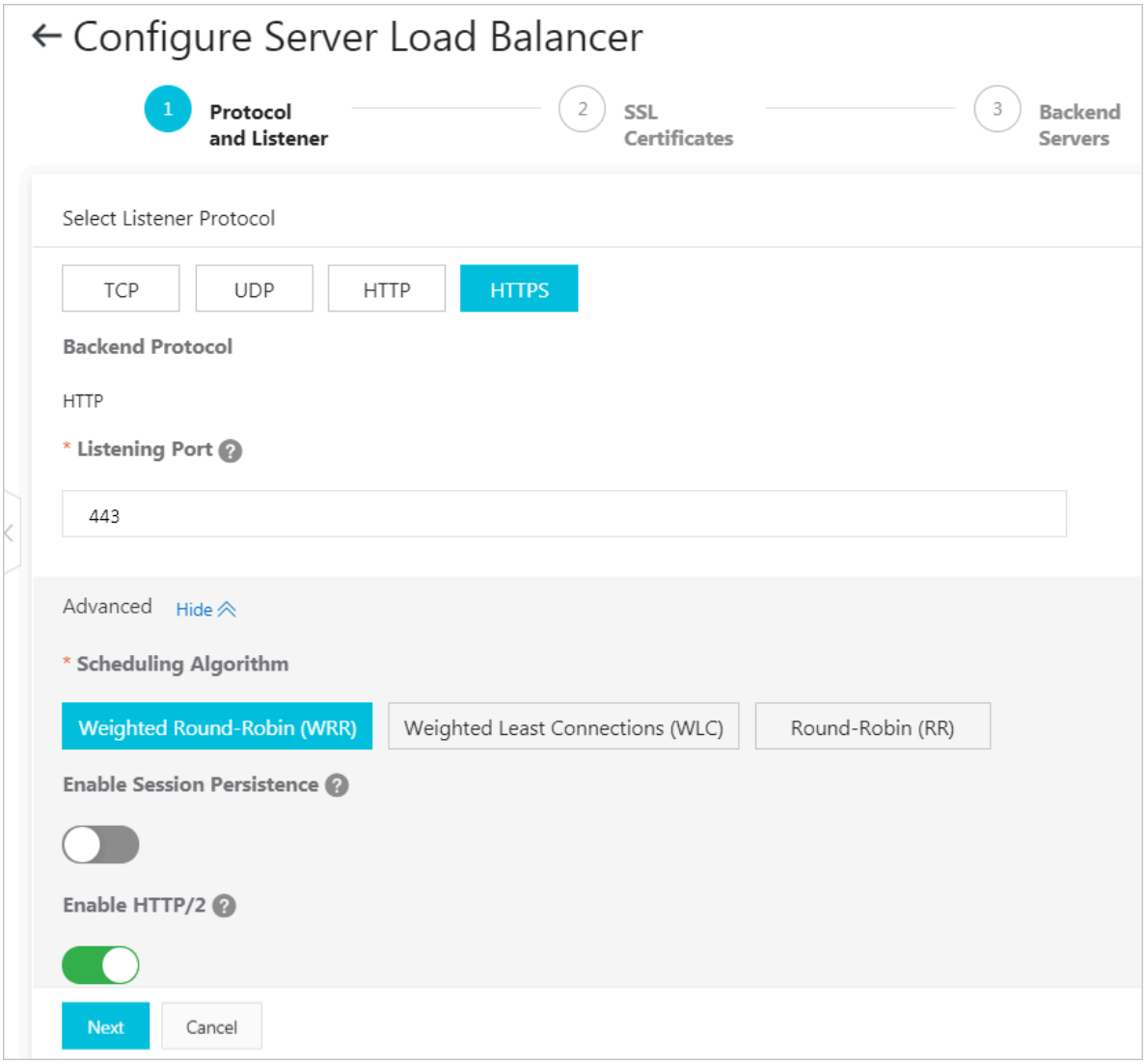
Configuration	Description
Enable Session Persistence	Select whether to enable session persistence. After you enable session persistence, all session requests from the same client are sent to the same backend server. HTTP session persistence is based on cookies. The following two methods are supported: · Insert cookie: You only need to specify the cookie timeout period. SLB adds a cookie to the first response from the backend server (inserts SERVERID in the HTTP and HTTPS response packet). The next request will contain the cookie and the listener will distribute the request to the same backend server. · Rewrite cookie: You can set the cookie to insert to the HTTP or HTTPS response according to your needs. You must maintain the timeout period and lifecycle of the cookie on the backend server. SLB will overwrite the original cookie when it discovers that a new cookie is set. The next time the client carries the new cookie to access SLB, the listener will distribute the request to the recorded backend server. For more information, see Configure session persistence.
Enable HTTP/2	Select whether to enable HTTP 2.0.
Enable Access Control	Select whether to enable the access control function.

Configuration	Description
Access Control Method	Select an access control method after you enable the access control function: · Whitelist: Only requests from IP addresses or CIDR blocks in the selected access control list are forwarded. It applies to scenarios where the application only allows access from some specific IP addresses. Enabling a whitelist poses some risks to your services. After a whitelist is configured, only the IP addresses in the list can access the listener. If you enable the whitelist without adding any IP addresses in the corresponding access control list, all requests are forwarded. · Blacklist: Requests from IP addresses or CIDR blocks in the selected access control list are not forwarded. It applies to scenarios where the application only denies access from some specific IP addresses. If you enable a blacklist without adding any IP addresses in the corresponding access control list, all requests are forwarded.
Access Control List	Select an access control list as the whitelist or the blacklist. Note: An IPv6 instance can only be associated with IPv6 access control lists and an IPv4 instance can only be associated with IPv4 access control lists. For more information, see Configure an access control list.

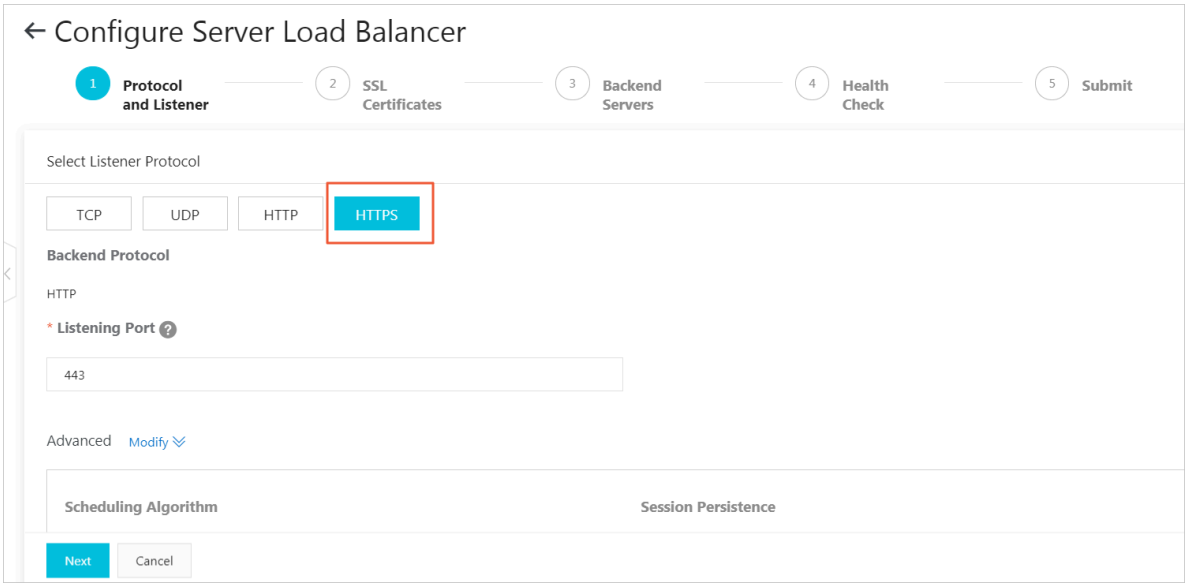
Configuration	Description
Enable Peak Bandwidth Limit	Select whether to configure the listening bandwidth. If the SLB instance is billed by bandwidth, you can set different peak bandwidth values for different listeners to limit the traffic passing through the listeners. The sum of the peak bandwidth values of all listeners under an SLB instance cannot exceed the bandwidth value of that SLB instance. By default, all listeners share the bandwidth of the SLB instance. Note: SLB instances billed by traffic have no peak bandwidth limit by default.
Idle Timeout	Specify the idle connection timeout period. Value range: 1 to 60. Unit: seconds. If no request is received during the specified timeout period, SLB will close the connection and restart the connection when the next request comes. This function is available in all regions.
Request Timeout	Specify the request timeout period. Value range: 1 to 180. Unit: seconds. If no response is received from the backend server during the specified timeout period, SLB will stop waiting and send an HTTP 504 error code to the client . Currently, this function is available in all regions.
TLS Security Policy	Only guaranteed-performance instances support selecting the TLS security policy. The TLS security policy contains available TLS protocol versions and supported cipher suites. For more information, see Manage TLS security policies.

Configuration	Description
Enable Gzip Compression	Choose whether to enable Gzip compression to compress files of specific formats. Now Gzip supports the following file types: text/xml, text/plain, text/css, application/javascript, application/x-javascript, application/rss+xml, application/atom+xml, and application/xml.
Add HTTP Header Fields	Select the custom HTTP headers that you want to add: · Use the `X-Forwarded-For` field to retrieve client source IP addresses. · Use the `X-Forwarded-Proto` field to retrieve the listener protocol used by the SLB instance. · Use the `SLB-IP` field to retrieve the public IP address of the SLB instance. · Use the `SLB-ID` field to retrieve the ID of the SLB instance.
Get Client Source IP Address	HTTP listeners use X-Forwarded-For to obtain real IP addresses of clients.

Configuration	Description
Automatically Enable Listener After Creation	Choose whether to start the listener after the listener is configured. The listener is started by default.



2. Click Next.



Step 3 Configure the SSL certificate

To add an HTTPS listener, you must upload a server certificate or CA certificate, as shown in the following table.

Certificate	Description	Required for one-way authentication?	Required for mutual authentication?
Server certificate	Used to identify a server. The client uses it to check whether the certificate sent by the server is issued by a trusted center.	Yes You need to upload the server certificate to the certificate management system of SLB.	Yes You need to upload the server certificate to the certificate management system of SLB.
Client certificate	Used to identify a client. The client user can prove its true identity when communicating with the server. You can sign a client certificate with a self-signed CA certificate.	No	Yes You need to install the client certificate on the client.

Certificate	Description	Required for one-way authentication?	Required for mutual authentication?
CA certificate	The server uses the CA certificate to authenticate the signature on the client certificate, as part of the authorization before launching a secure connection. If the authentication fails, the connection is rejected.	No	Yes You need to upload the CA certificate to the certificate management system of SLB.

Note the following before you upload a certificate:

- The uploaded certificate must be in the PEM format. For more information, see Certificate requirements.
- After the certificate is uploaded to SLB, SLB can manage the certificate and you do not need to associate the certificate with backend ECS instances.
- It usually takes one to three minutes to activate the HTTPS listener because the uploading, loading, and validation of certificates take some time. Normally it takes effect in one minute and it will definitely take effect in three minutes.
- The ECDHE algorithm cluster used by HTTPS listeners supports forward secrecy, but does not support uploading security enhancement parameter files required by the DHE algorithm cluster, such as strings containing the `BEGIN DH PARAMETERS` field in the PEM certificate file. For more information, see Certificate requirements.
- Currently, SLB HTTPS listeners do not support SNI (Server Name Indication). You can use TCP listeners instead, and then configure SNI on backend ECS instances.
- The session ticket timeout period of HTTPS listeners is 300 seconds.
- The actual amount of traffic is larger than the billed traffic amount because some traffic is used for protocol handshaking.
- In the case of a large number of new connections, HTTPS listeners consume more traffic.

To configure the SSL certificate, follow these steps:

1. Select the server certificate that has been uploaded, or click Create Server Certificate to upload a server certificate.

   For more information, see Create a certificate.
2. If you want to enable HTTPS mutual authentication or set a TLS security policy, click Modify.

3. Select an uploaded CA certificate, or click Create CA Certificate to upload a CA certificate.

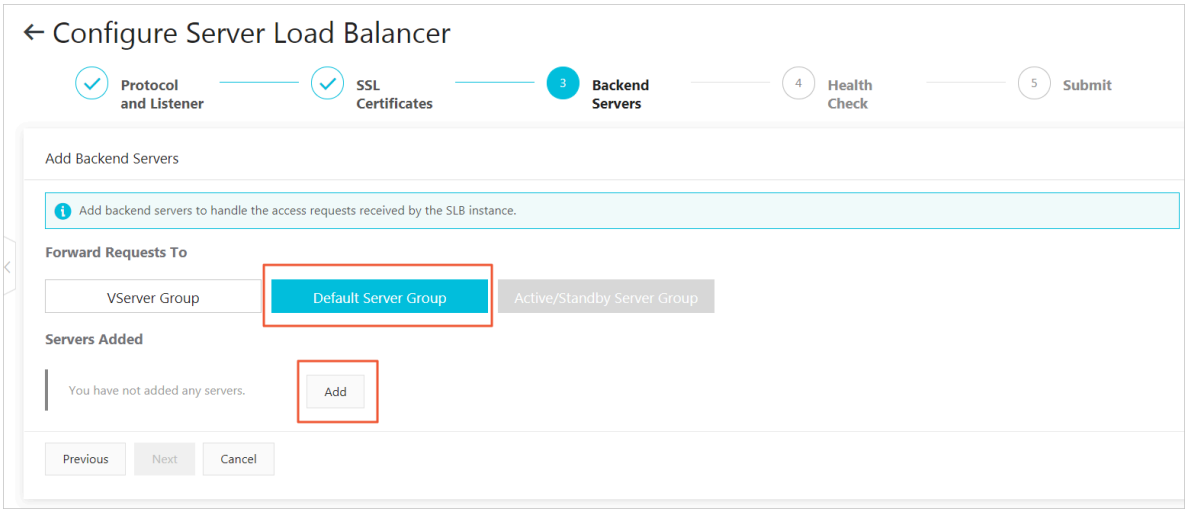
   You can use a self-signed CA certificate. For more information, see Generate a CA certificate.
4. Select a TLS security policy. For more information, see Manage TLS security policies.

Step 4 Add backend servers

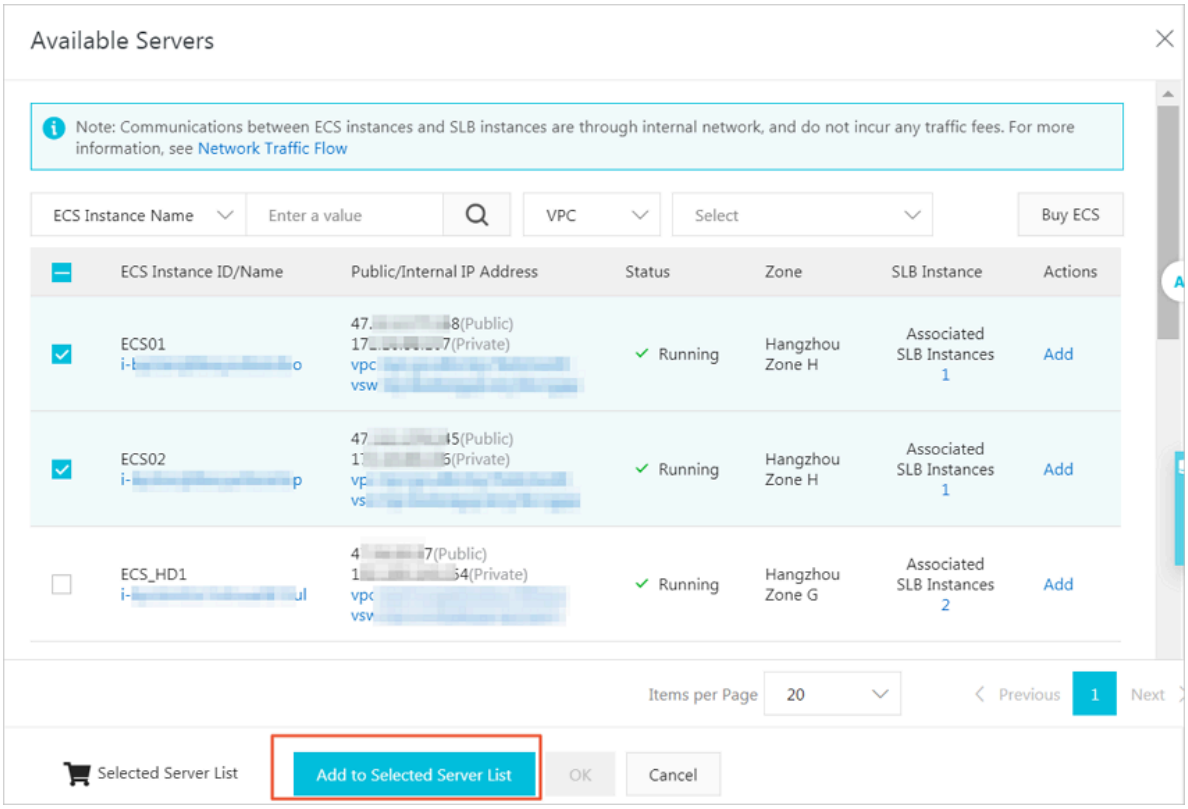
You need to add backend servers to process requests. You can use the default server group configured for the SLB instance, or configure a VServer group or an active/standby server group for the listener. For more information, see Backend server overview.

In this topic, use the default server group.

1. Select Default Server Group and then click Add More.



2. Select the ECS instances to add, and then click Next: Set Weight and Port.



3. Configure ports and weights for the added backend servers.

   - Port

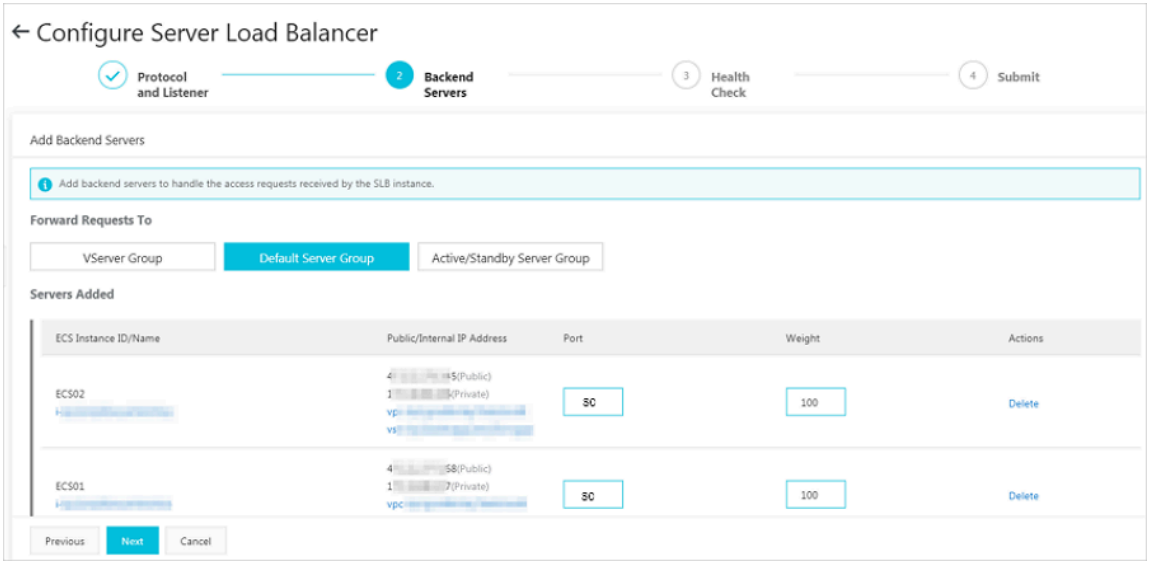
     The port opened on the backend server (ECS instance) to receive requests. The port number is in the range of 1 to 65535. Ports of backend servers can be the same in an SLB instance.

   - Weight

     The weight of the backend server (ECS instance). An ECS instance with a higher weight receives more requests.

     Note:

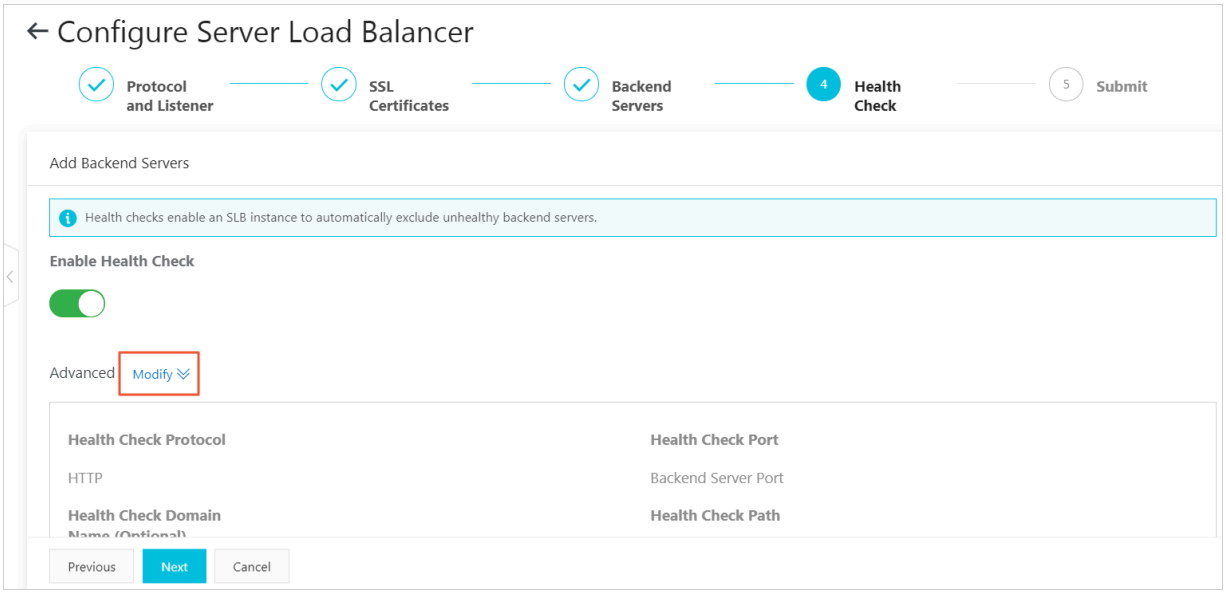
If the weight is set to 0, no requests are sent to the ECS instance.



4. Click Next.

Step 5 Configure health checks

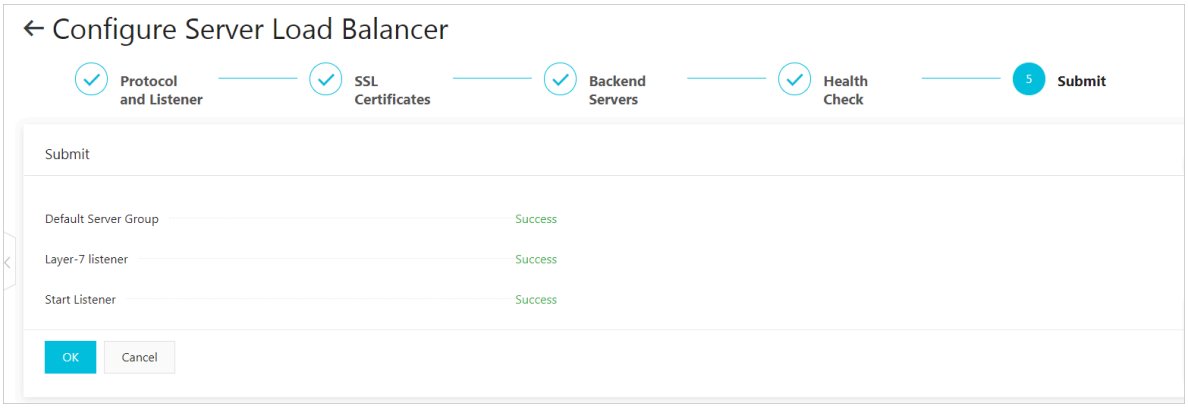
SLB checks the service availability of backend servers (ECS instances) by performing health checks. The health check function improves the overall availability of your services and avoids the impact of backend server failures. Click Modify to change health check configurations. For more information, see Configure health checks.



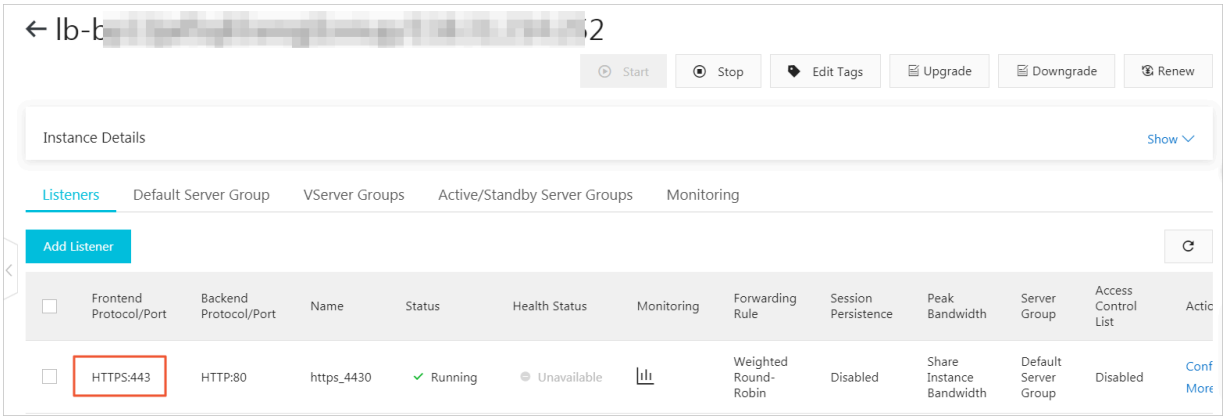
Step 6 Submit the configurations

To confirm the listener configurations, follow these steps:

1. On the Submit page, check the listener configurations. You can click Modify to change the configurations. Click Submit.
2. On the Submit page, click OK after the configurations are successful.



After the configurations are submitted, you can view the created listener on the Listeners page.



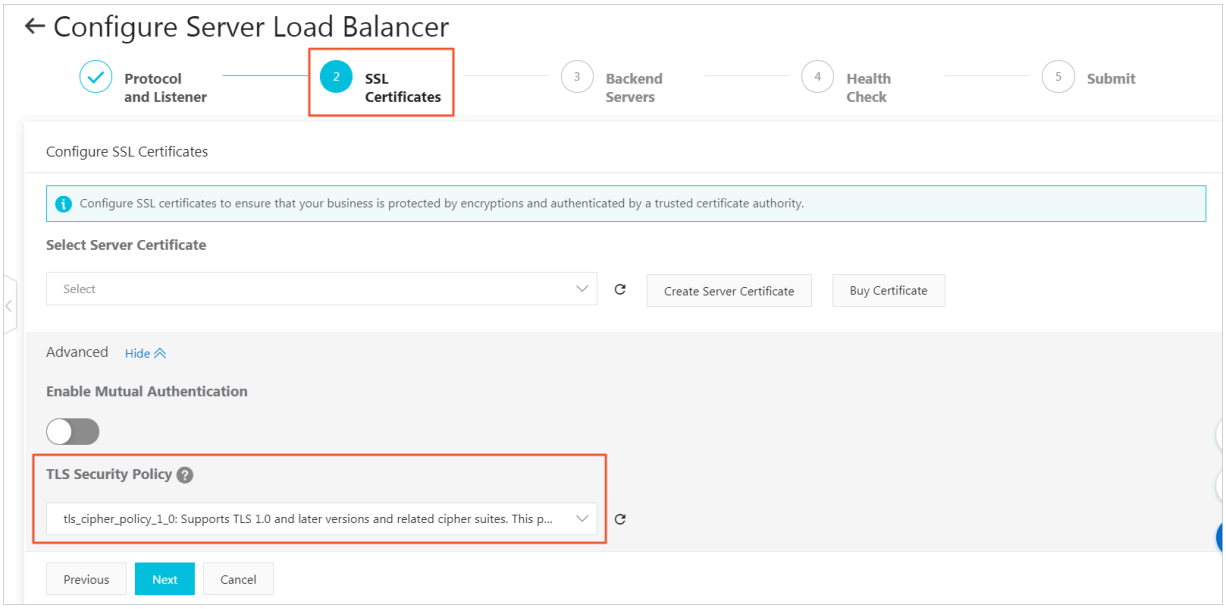
Related operations

- Configure health checks
- Manage a default server group
- Manage a VServer group
- Manage an active/standby server group
- Generate a CA certificate
- Create a certificate
- Configure access control
- Add domain-name based or URL-based forwarding rules
- Manage a domain name extension

2.6 Manage TLS security policies

When you add or configure an HTTPS listener for a guaranteed-performance Server Load Balancer (SLB) instance, you can select from a variety of TLS security policies and apply one according to your requirements.

You can select the TLS security policy when you set advanced configurations of SSL Certificates for an HTTPS listener. For more information, see Add an HTTPS listener.



The TLS security policy contains supported TLS protocol versions and cipher suites.

TLS security policy

Security policy	Features	Supported TLS versions	Supported cipher suites
tls_cipher_policy_1_0	Optimal compatibility and with basic security	TLSv1.0、TLSv1.1和TLSv1.2	ECDHE-RSA-AES128-GCM-SHA256, ECDHE-RSA-AES256-GCM-SHA384, ECDHE-RSA-AES128-SHA256, ECDHE-RSA-AES256-SHA384, AES128-GCM-SHA256, AES256-GCM-SHA384, AES128-SHA256, AES256-SHA256, ECDHE-RSA-AES128-SHA, ECDHE-RSA-AES256-SHA, AES128-SHA, AES256-SHA, and DES-CBC3-SHA

Security policy	Features	Supported TLS versions	Supported cipher suites
tls_cipher_policy_1_1	Compatible and with standard security	TLSv1.1 and TLSv1.2	ECDHE-RSA-AES128-GCM-SHA256, ECDHE-RSA-AES256-GCM-SHA384, ECDHE-RSA-AES128-SHA256, ECDHE-RSA-AES256-SHA384, AES128-GCM-SHA256, AES256-GCM-SHA384, AES128-SHA256, AES256-SHA256, ECDHE-RSA-AES128-SHA, ECDHE-RSA-AES256-SHA, AES128-SHA, AES256-SHA and DES-CBC3-SHA.
tls_cipher_policy_1_2	Compatible and with advanced security	TLSv1.2	ECDHE-RSA-AES128-GCM-SHA256, ECDHE-RSA-AES256-GCM-SHA384, ECDHE-RSA-AES128-SHA256, ECDHE-RSA-AES256-SHA384, AES128-GCM-SHA256, AES256-GCM-SHA384, AES128-SHA256, AES256-SHA256, ECDHE-RSA-AES128-SHA, ECDHE-RSA-AES256-SHA, AES128-SHA, AES256-SHA, and DES-CBC3-SHA
tls_cipher_policy_1_2_strict	Supports only perfect forward secrecy (PFS) cipher suites and offers premium security.	TLSv1.2	ECDHE-RSA-AES128-GCM-SHA256, ECDHE-RSA-AES256-GCM-SHA384, ECDHE-RSA-AES128-SHA256, ECDHE-RSA-AES256-SHA384, ECDHE-RSA-AES128-SHA, and ECDHE-RSA-AES256-SHA

Security policy	Features	Supported TLS versions	Supported cipher suites
tls_cipher_policy_1_2_strict_with_1_3 Note: Currently, only the UK (London) region supports TLS1.3.	Supports only perfect forward secrecy (PFS) cipher suites and offers premium security.	TLS1.2 and TLS1.3	TLS_AES_128_GCM_SHA256, TLS_AES_256_GCM_SHA384, TLS_CHACHA20_POLY1305_SHA256 , TLS_AES_128_CCM_SHA256, TLS_AES_128_CCM_8_SHA256, ECDHE -ECDSA-AES128-GCM-SHA256, ECDHE -ECDSA-AES256-GCM-SHA384, ECDHE-ECDSA-AES128-SHA256, ECDHE-ECDSA -AES256-SHA384, ECDHE-RSA-AES128 -GCM-SHA256, ECDHE-RSA-AES256 -GCM-SHA384, ECDHE-RSA-AES128-SHA256, ECDHE-RSA-AES256-SHA384 , ECDHE-ECDSA-AES128-SHA, ECDHE -ECDSA-AES256-SHA, ECDHE-RSA-AES128-SHA, and ECDHE-RSA-AES256-SHA

Algorithm support of different TLS security policies

Security Policy		tls_cipher_policy_1_0	tls_cipher_policy_1_1	tls_cipher_policy_1_2	tls_cipher_policy_1_2_strict	tls_cipher_policy_1_2_strict_with_1_3
TLS		1.2/1.1/1.0	1.2/1.1	1.2	1.2	1.2 and 1.3
CIPHER	ECDHE-RSA-AES128-GCM-SHA256	#	#	#	#	#
	ECDHE-RSA-AES256-GCM-SHA384	#	#	#	#	#
	ECDHE-RSA-AES128-SHA256	#	#	#	#	#
	ECDHE-RSA-AES256-SHA384	#	#	#	#	#
	AES128-GCM-SHA256	#	#	#		
	AES256-GCM-SHA384	#	#	#		

Security Policy		tls_cipher_policy_1_0	tls_cipher_policy_1_1	tls_cipher_policy_1_2	tls_cipher_policy_1_2_strict	tls_cipher_policy_1_2_strict_with_1_3
	AES128-SHA256	#	#	#		
	AES256-SHA256	#	#	#		
	ECDHE-RSA-AES128-SHA	#	#	#	#	#
	ECDHE-RSA-AES256-SHA	#	#	#	#	#
	AES128-SHA	#	#	#		
	AES256-SHA	#	#	#		
	DES-CBC3-SHA	#	#	#		
	TLS_AES_128_GCM_SHA256					#
	TLS_AES_256_GCM_SHA384					#
	TLS_CHACHA20_POLY1305_SHA256					#
	TLS_AES_128_CCM_SHA256					#
	TLS_AES_128_CCM_8_SHA256					#
	ECDHE-ECDSA-AES128-GCM-SHA256					#
	ECDHE-ECDSA-AES256-GCM-SHA384					#
	ECDHE-ECDSA-AES128-SHA256					#
	ECDHE-ECDSA-AES256-SHA384					#
	ECDHE-ECDSA-AES128-SHA					#

Security Policy		tls_cipher_policy_1_0	tls_cipher_policy_1_1	tls_cipher_policy_1_2	tls_cipher_policy_1_2_strict	tls_cipher_policy_1_2_strict_with_1_3
	ECDHE-ECDSA-AES256-SHA					#

2.7 Manage a domain name extension

HTTPS listeners of guaranteed-performance Server Load Balancer (SLB) instances support configuring multiple certificates, allowing you to forward requests with different domain names to different backend servers.

Introduction to SNI

Server Name Indication (SNI) is an extension to the SSL/TLS protocol, allowing a server to install multiple SSL certificates on the same IP address. When a client accesses SLB, the certificate configured for the domain name is used by default. If no certificate is configured for the domain name, the certificate configured for the HTTPS listener is used.

Note:
Only guaranteed-performance SLB instances support SNI.

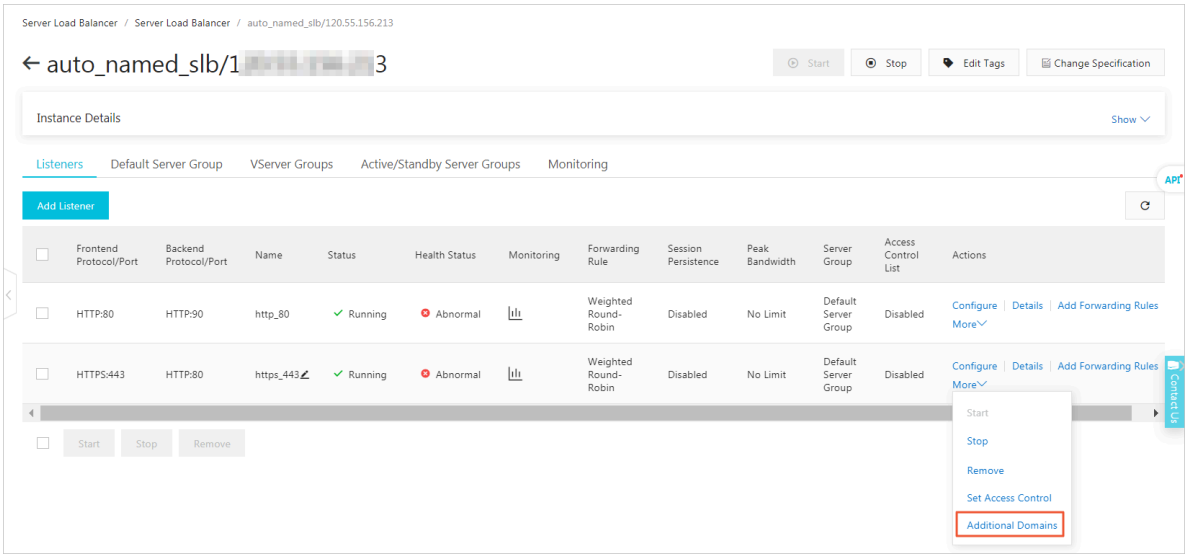
If you want to resolve multiple domain names to the IP address of an SLB instance, distribute requests from different domains to different backend servers, and at the same time use HTTPS encrypted access, you can use the domain name extension function.

The domain name extension function is available in all regions.

Add a domain name extension

1. Log on to the SLB console.
2. Select the region of the target SLB instance.
3. Find the target SLB instance and click the instance ID.
4. Click the Listeners tab.

5. On the Listeners tab page, find the target HTTPS listener, and choose More > Additional Domains in the Actions column.



6. Click Add Additional Domain and configure the domain name:

   a. Enter a domain name. The domain name can only contain letters, numbers, hyphens (-), and periods (.), and must start with a letter or a number. To check if the domain name you enter is valid, you can use the Alibaba Cloud domain name check tool.

      Domain name-based forwarding rules support exact match and wildcard match.

      - Exact domain name: www.aliyun.com
      - Wildcard domain name (generic domain name): *.aliyun.com, *.market.aliyun.com

      When a request matches multiple forwarding rules, exact match takes precedence over small-scale wildcard match and small-scale wildcard match takes precedence over large-scale wildcard match, as shown in the following table.

Type	Request URL	Request URL		
		www.aliyun.com	*.aliyun.com	*.market.aliyun.com
Exact match	www.aliyun.com	√	×	×
Wildcard match	market.aliyun.com	×	√	×

Type	Request URL	Request URL		
		www.aliyun.com	*.aliyun.com	*.market.aliyun.com
Wildcard match	info.market.aliyun.com	×	×	√

b. Select the certificate associated with the domain name.

> Note:
> The domain name in the certificate must be the same as the added domain name extension.

c. Click OK.

7. On the Listeners page, find the target HTTPS listener and click Add Forwarding Rules in the Actions column.
8. On the Add Forwarding Rules page, configure the forwarding rule and click Add Forwarding Rules.
9. For more information, see Traffic forwarding based on domain names or URLs.

> Note:
> Make sure that the domain name configured in the forwarding rule is the same as the added domain name extension.

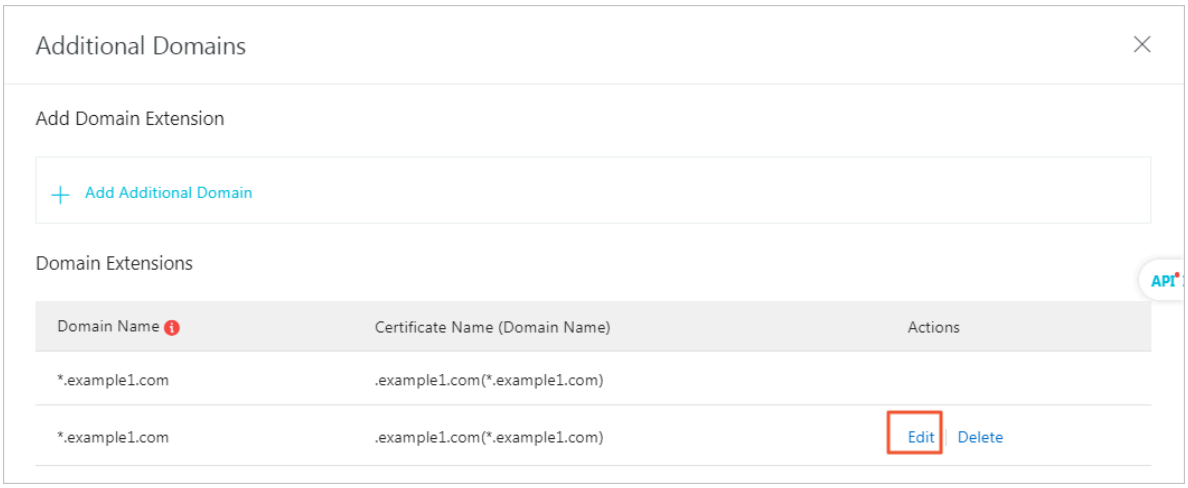
Edit a domain name extension

You can replace the certificate used by an added domain name extension.

To edit a domain name extension, follow these steps:

1. Log on to the SLB console.
2. Select the region of the target SLB instance.
3. Find the target SLB instance and click the instance ID.
4. Click the Listeners tab.
5. On the Listeners page, find the created HTTPS listener, and then choose More > Additional Domains in the Actions column.
6. Find the target domain name extension and then click Edit.

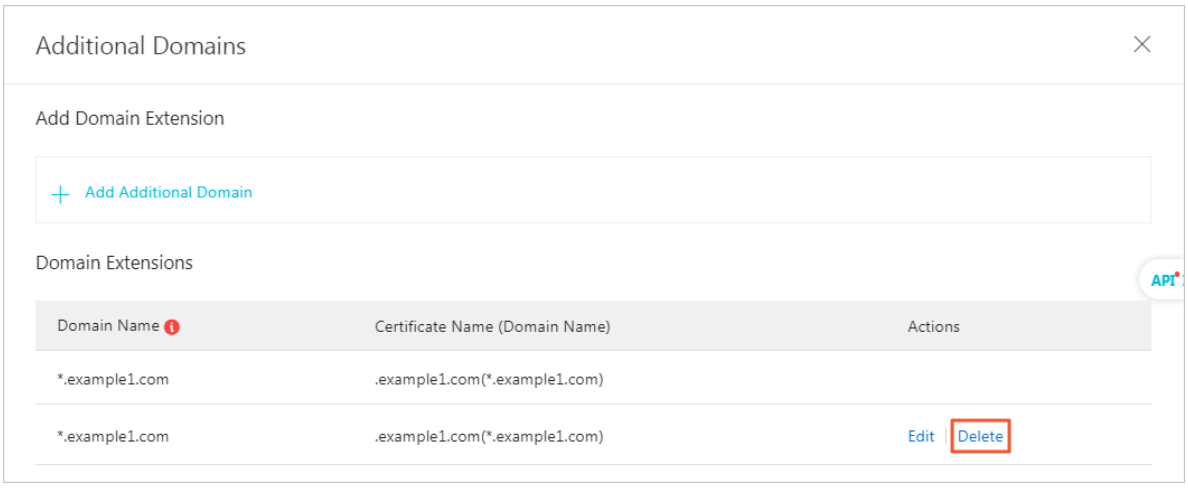
7. In the Edit Additional Domain dialog box, select a new certificate and then click OK.



Delete a domain name extension

To delete a domain name extension, follow these steps:

1. Log on to the SLB console.
2. Select the region of the target SLB instance.
3. Find the target SLB instance and click the instance ID.
4. Click the Listeners tab.
5. On the Listeners page, find the created HTTPS listener, and then choose More > Additional Domains in the Actions column.
6. Find the target domain name extension and click Delete.



7. In the displayed dialog box, click OK.

2.8 Redirect HTTP requests to HTTPS

HTTPS is the secure version of HTTP. With HTTPS, the data sent between the browser and the server is encrypted. Server Load Balancer (SLB) supports redirecting HTTP requests to HTTPS to facilitate whole-site HTTPS deployment. Redirecting HTTP requests to HTTPS is supported in all regions.

Prerequisites

An HTTPS listener is created. For more information, see Add an HTTPS listener.

Context

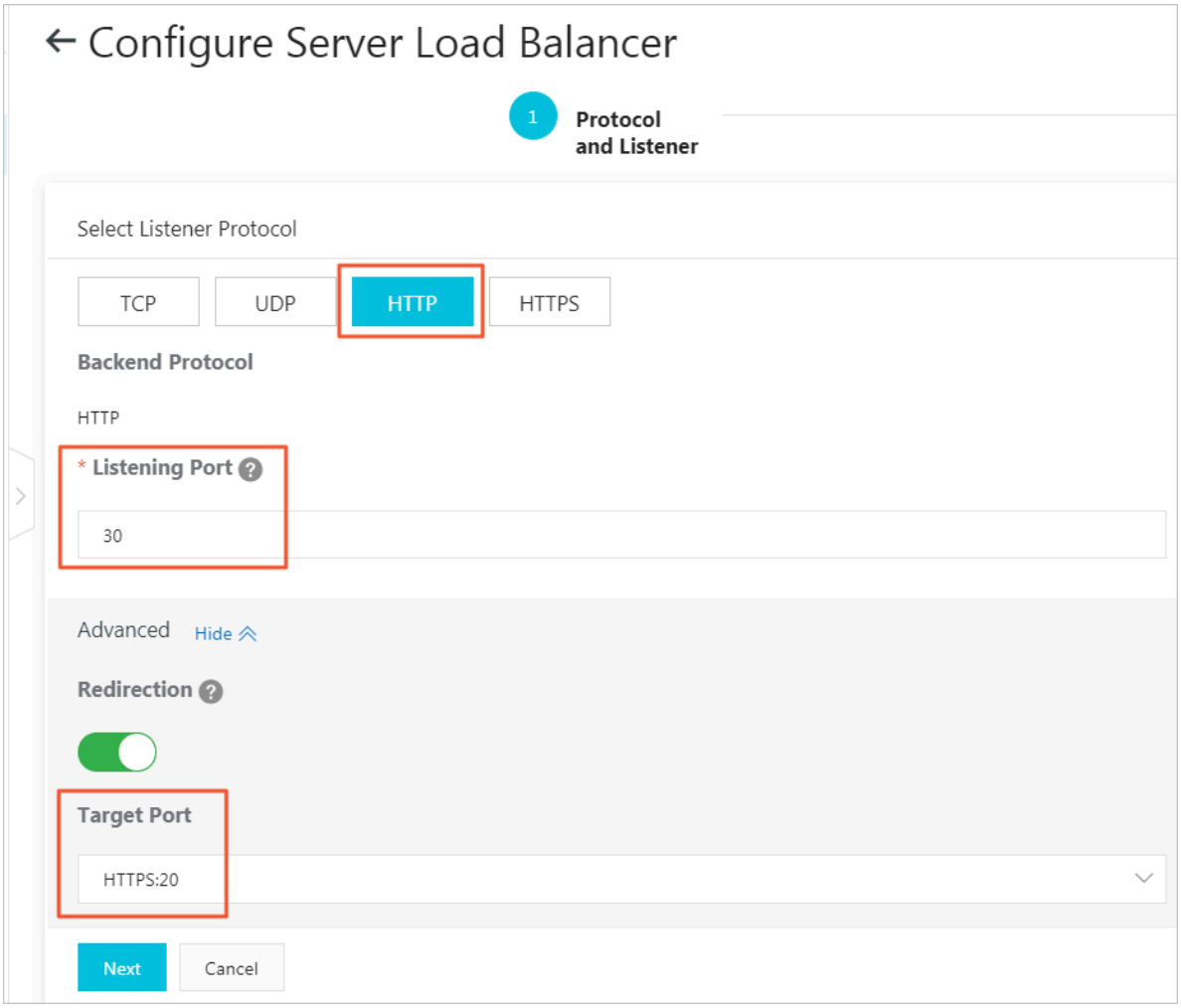
The redirection function is supported only by the new version SLB console.

Procedure

1. Log on to the SLB console.
2. Select the region of the target SLB instance.
3. On the Server Load Balancer page, click the ID of the target SLB instance.
4. Click the Listeners tab and then click Add Listener.
5. In the Configure Server Load Balancer dialog box, select HTTP as the listener protocol and configure the listening port.

6. In the Advanced section, turn on Redirection and select the target HTTPS listener.

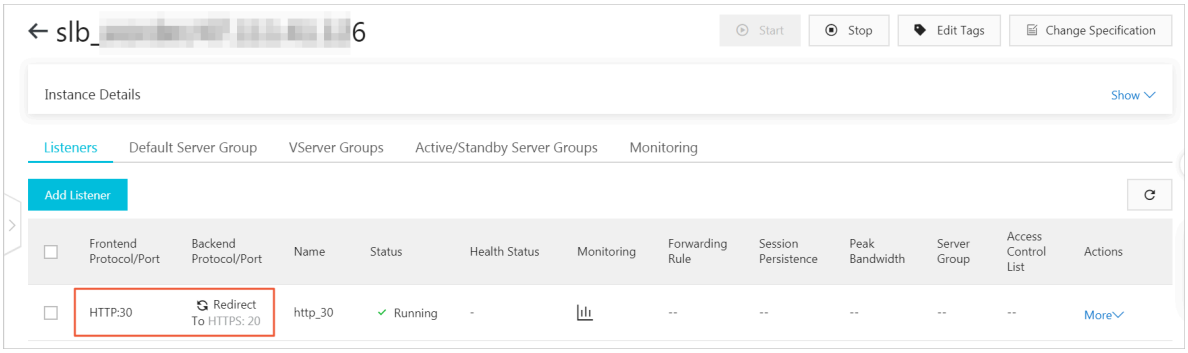
   The target listener can be an HTTPS listener with any port in the SLB instance.



7. Click Next.
8. Check the configurations and click Submit.

9. Click OK.

   After the redirection function is enabled, all HTTP requests will be redirected to the selected HTTPS listener and distributed according to the listener configurations of the HTTPS listener.



3 Health check

3.1 Health check overview

Server Load Balancer checks the service availability of the backend servers (ECS instances) by performing health checks. Health check improves the overall availability of the front-end service, and avoids impact on the entire service caused by exceptions of the backend ECS instances.

After enabling the health check function, SLB stops distributing requests to the instance that is discovered as unhealthy and restarts forwarding requests to the instance only when it is declared healthy.

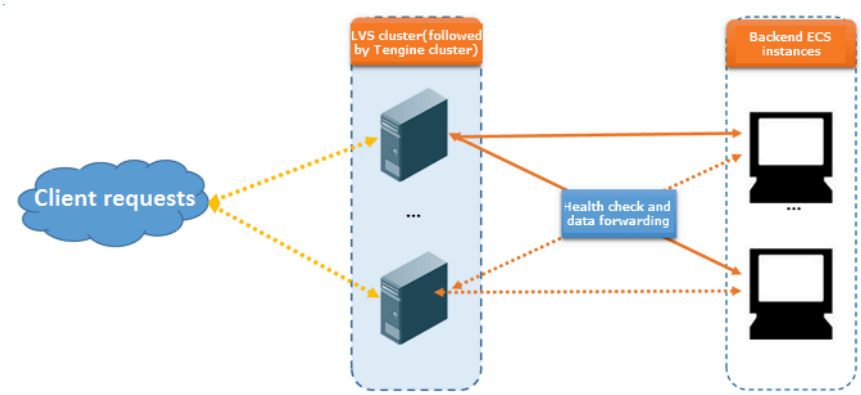
If your business is highly sensitive to traffic load, frequent health checks may impact normal service. You can reduce this impact by reducing the frequency of health checks, increasing the health check interval, or changing the HTTP health check to TCP health check. To guarantee the service availability, we do not recommend disabling all health checks.

Health check process

Server Load Balancer is deployed in clusters. Data forwarding and health checks are handled at the same time by the node servers in the LVS cluster and Tengine cluster.

The node servers in the cluster independently perform health checks in parallel, according to the health check configuration. If a node server discovers an ECS instance is unhealthy, the node server will stop distributing requests to the ECS instance. This operation is synchronized through all node servers.

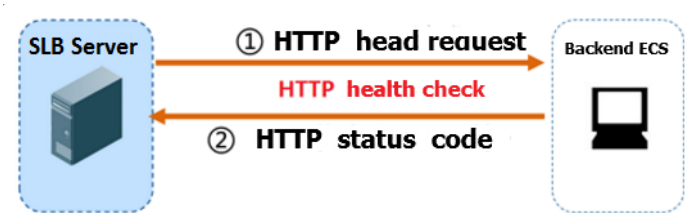
The IP address range used to perform the health check is 100.64.0.0/10. The backend servers cannot block this CIDR block. You do not need to additionally configure a security group rule to allow access from this CIDR block. However, if you have configured security rules such as iptables, allow access from this CIDR block (100.64.0.0/10 is reserved by Alibaba Cloud, and other users cannot use any IP address in this CIDR block, so there is no security risk).



Health check of HTTP/HTTPS listeners

For Layer-7 (HTTP or HTTPS) listeners, SLB detects the status of backend servers by sending HTTP HEAD requests, as shown in the following figure.

For HTTPS listeners, certificates are managed in SLB. Data exchange (including health check data and service interaction data) between SLB and backend ECS instances is not transmitted over HTTPS to improve the system performance.

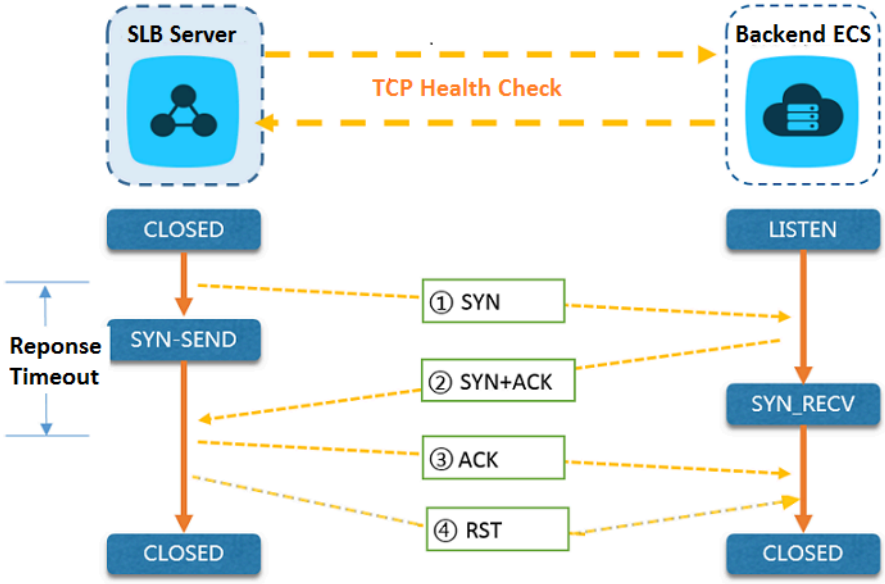


The health check process of a Layer-7 listener is as follows:

1. The Tengine node server sends an HTTP HEAD request to the intranet IP +Health Check Port+Health Check Path of the ECS instance according to the health check settings.
2. After receiving the request, the backend server returns an HTTP status code based on the running status.
3. If the Tengine node server does not receive the response from the backend server within the Response Timeout period, the ECS instance is declared unhealthy.
4. If the Tengine node server receives the response from the backend ECS instance within the Response Timeout period, it compares the returned status code with the status code specified in the listener configuration. If the status code is the same, the backend server is declared healthy. Otherwise, the backend server is declared unhealthy.

Health check of TCP listeners

For TCP listeners, SLB detects the status of backend servers by sending TCP detections, as the following figure shows.



The health check process of a TCP listener is as follows:

1. The LVS node server sends a TCP SYN packet to the intranet IP + Health Check Port of the backend ECS instance.
2. After receiving the request, the backend server returns a TCP SYN and ACK packet if the corresponding port is listening normally.
3. If the LVS node server does not receive the required data packet from the backend server within the Response Timeout period, the ECS instance is declared unhealthy . Then, the LVS node server sends an RST data packet to the backend server to terminate the TCP connection.
4. If the LVS node server receives the data packet from the backend ECS instance within the Response Timeout period, the ECS instance is declared healthy. Then, the LVS node server sends an RST data packet to the backend server to terminate the TCP connection.

Note:
In general, TCP three-way handshakes are conducted to establish a TCP connection. After the LVS node server receives an SYN + ACK data packet from the backend ECS instance, the LVS node server sends an ACK data packet, and then immediately sends an RST data packet to terminate the TCP connection.

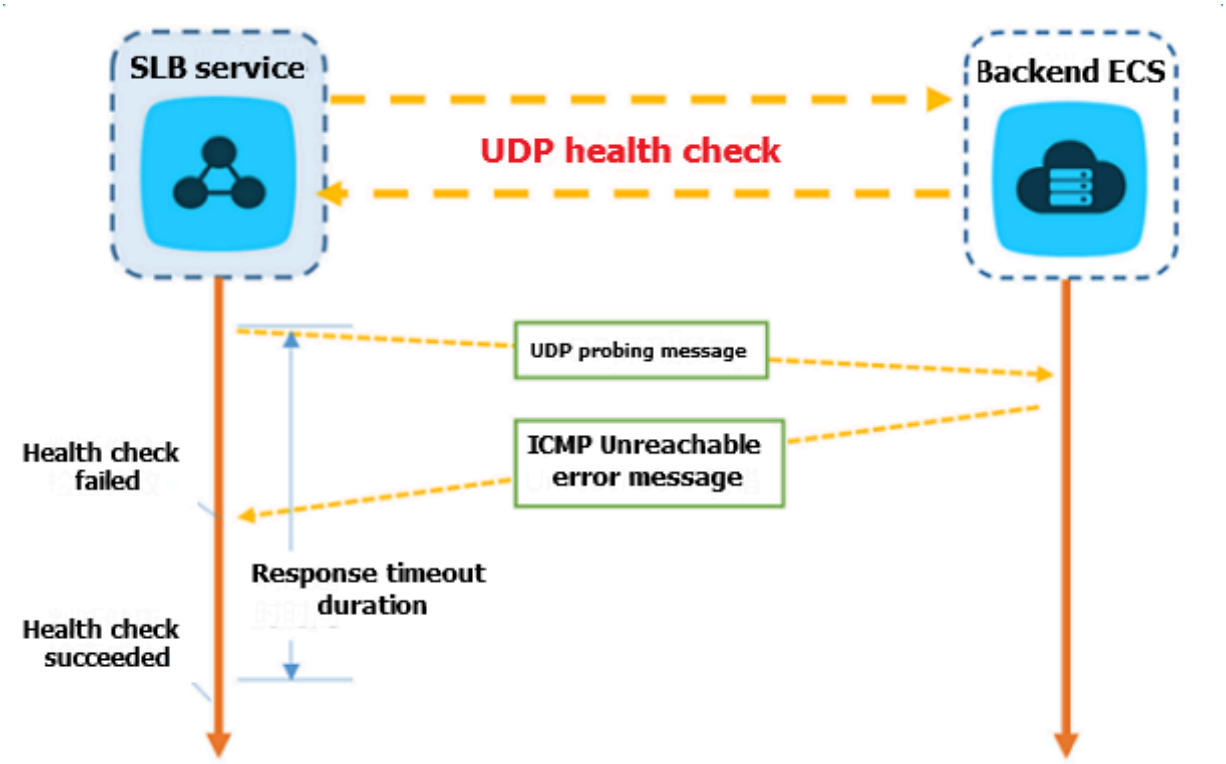
This process may make backend server think an error (such as an abnormal exit) occurred in the TCP connection and then throw a corresponding error message, such as `Connection reset by peer`.

Resolution:

- Use the HTTP health check.
- If obtaining real IP is enabled, ignore the connection errors caused by access of the SLB IP address.

Health check of UDP listeners

For UDP listeners, Server Load Balancer detects the status of the backend servers through UDP packet detection, as shown in the following figure.



The health check process of a UDP listener is as follows:

1. The LVS node server sends a UDP packet to the intranet IP + Health Check Port of the ECS instance according to health check configurations.
2. If the corresponding port of the ECS instance is not listening normally, the system will return an ICMP error message, such as `port XX unreachable`. Otherwise, no message is sent.

3. If the LVS node server receives the ICMP error message within the Response Timeout period, the ECS instance is declared unhealthy.
4. If the LVS node server does not receive any messages within the Response Timeout period, the ECS instance is declared healthy.

 Note:

For UDP health checks, the real status of the backend server and the health check result may not be the same in the following situation:

If the ECS instance uses a Linux operating system, the speed of sending ICMP messages in high-concurrency scenarios is limited due to the anti-ICMP attack protection in Linux. In this case, even if an exception occurs in the ECS instance, SLB may declare the backend server healthy because the error message `port XX unreachabl e` is not returned. As a result, the actual service status is different from the health check result.

Resolution:

Set a pair of custom request and response for the UDP health check. If the custom response is returned, the ECS instance is considered healthy. Otherwise, the ECS instance is considered unhealthy. To achieve this, you must add corresponding configurations for the client.

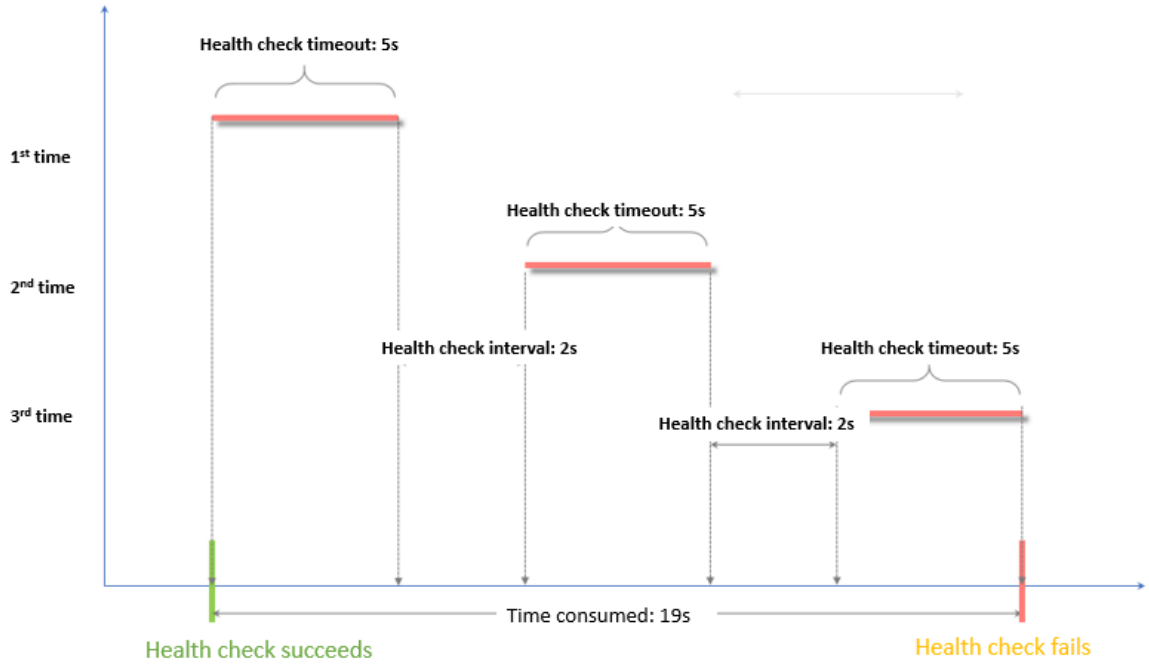
Health check time window

The health check function has effectively improved the availability of your business services. However, to reduce impact on the system availability caused by frequent system switches because of health check failure, SLB declares an ECS instance healthy or unhealthy only after successive successes or failures within a specified timeframe. The health check time window is determined by the following three factors:

- Health check interval (How often the health check is performed.)
- Response timeout (The amount of time to wait for the response.)
- Health check threshold (The number of consecutive successful or failed health checks.)

The health check time window is calculated as follows:

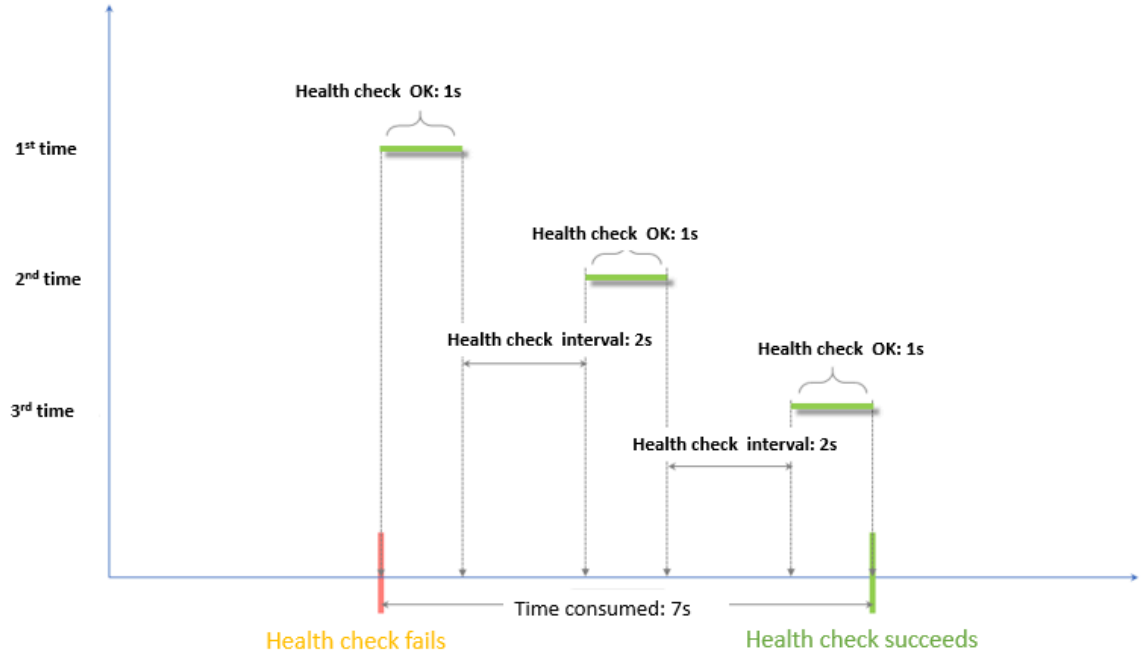
- Health check failure time window = Response Timeout x Unhealthy Threshold + Health Check Interval X (Unhealthy Threshold) -1)



- Health check success time window = (Response Time of a Successful Health Check X Healthy Threshold) + Health Check Interval X (Healthy Threshold-1)

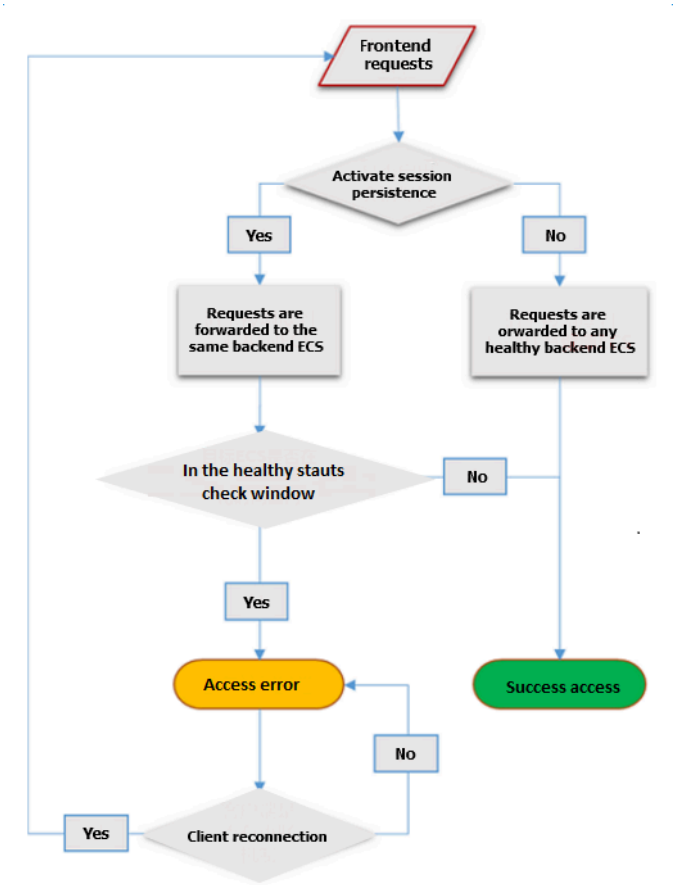
> Note:
> The success response time of a health check is the duration from the time when the health check request is sent to the time when the response is received. For TCP health check, the time is very short and almost negligible because TCP health check only detects whether the port is alive. For HTTP health check, the time

depends on the performance and load of the application server and is generally within seconds.



The health check result has the following impact on the requests forwarding:

- If the health check of the target ECS instance fails, new requests will not be distributed to the ECS instance. Therefore, there is no impact on the client access.
- If the health check of the target ECS instance succeeds, new requests will be distributed to it. The client access is normal.
- If a request arrives during a health check failure window, the request is still sent to the ECS instance because the ECS instance is being checked and has not been declared unhealthy. As a result, the client access fails.



3.2 Configure health checks

You can configure the health check function when you add a listener. Generally, the default settings can meet your requirements.

Configure health checks

You can configure the health check function of a listener through the console or APIs. For more information, see Health check overview and Health check FAQ.

To configure the health check function, follow these steps:

1. Log on to the SLB console.
2. Select the region of the target SLB instance.
3. Find the target SLB instance and click the instance ID.
4. On the Instance Details page, click the Listeners tab.
5. Click Add Listener, or find the target listener and click Configure in the Actions column.

6. On the Health Check page, configure the health check function.

   We recommend that you use the default values when you configure the health check function.

   Table 3-1: Health check configurations

Configuration	Description
Health Check Protocol	For TCP listeners, both TCP health checks and HTTP health checks are supported. · TCP health checks are based on network layer detection. · HTTP health checks are performed by sending HEAD requests.
Health Check Method (HTTP and HTTPS health checks only)	Health checks of Layer-7 listeners (HTTP and HTTPS listeners) support both the HEAD and the GET request methods. The HEAD request method is used by default. Therefore, if your backend servers do not support the HEAD request method or the HEAD request method is disabled, health checks may fail. To resolve this issue, you can choose to use the GET request method for health checks. However, only the India (Mumbai) region supports the GET request method. Support for other regions is in development . Note: When the GET method is used, if the response length exceeds 8 KB, it is truncated, but the health check result is not affected.

Configuration	Description
Health Check Path and Domain Name (HTTP health checks only)	By default, SLB sends an HTTP HEAD request to the default homepage configured on the application server through the intranet IP address of the backend ECS instance to do health checks. If you do not use the default homepage of the application server to do health checks, you must specify the URL for health checks. Some application servers verify the host field in a request. Therefore, the request header must contain the host field. If a domain name is configured in the health check function, SLB adds the domain name to the host field when forwarding the request to a backend server. If not, the health check request is denied by the server and the health check may fail . Therefore, if your application server verifies the host field in the request, you must configure a domain name to make sure the health check works.
Normal Status Code (HTTP health checks only)	Select the HTTP status code that indicates normal health checks. The default values are http_2xx and http_3xx.
Health Check Port	The detection port used by health checks to access backend ECS instances. By default, the backend port configured in the listener is used. Note: If a VServer group or an active/standby server group is configured for the listener, and the ECS instances in the group use different ports, leave this parameter empty. SLB uses the backend port of each ECS instance to do health checks.

Configuration	Description
Response Timeout	The length of time to wait for the response from a health check. If the backend ECS instance does not send a correct response within the specified time, the health check fails. Value range: 1 to 300. Unit: seconds. Default value for UDP listeners: 10. Default value for HTTP, HTTPS, and TCP listeners: 5.
Health Check Interval	The time interval between two consecutive health checks. All node servers in the LVS cluster independently and concurrently perform health checks on backend ECS instances according to the interval. The statistics from a health check request on a single ECS instance cannot reflect the health check interval because the health check time of each node server is not synchronized. Value range: 1 to 50. Unit: seconds. Default value for UDP listeners: 5. Default value for HTTP, HTTPS, and TCP listeners: 2.
Unhealthy Threshold	The number of consecutive failures of health checks performed by the same LVS node server on the same ECS instance before the ECS instance is declared as unhealthy (from success to failure). Value range: 2 to 10. Default value: 3.
Healthy Threshold	The number of consecutive successes of health checks performed by the same LVS node server on the same ECS instance before the ECS instance is declared as healthy (from failure to success). Value range: 2 to 10. Default value: 3.

Configuration	Description
Health Check Requests and Results	When you configure health checks for UDP listeners, you can enter the request contents (such as youraccountID) in Health Check Request and the expected response (such as slb123) in Health Check Response. Add the corresponding health check response logic to the application logic of the backend server. For example, return slb123 when youraccountID is received. If SLB receives the expected response from the backend server, the health check succeeds. Otherwise, the health check fails. This method can guarantee the reliability of health checks.

Example of health check response timeout and health check interval

Take the following health check configurations as the example:

- Response Timeout: 5 seconds
- Health Check Interval: 2 seconds
- Healthy Threshold: 3 times
- Unhealthy Threshold: 3 times

Health check failure time window = Response Timeout × Unhealthy Threshold + Health Check Interval × (Unhealthy Threshold - 1). That is, 5 × 3 + 2 × (3 - 1) = 19s.

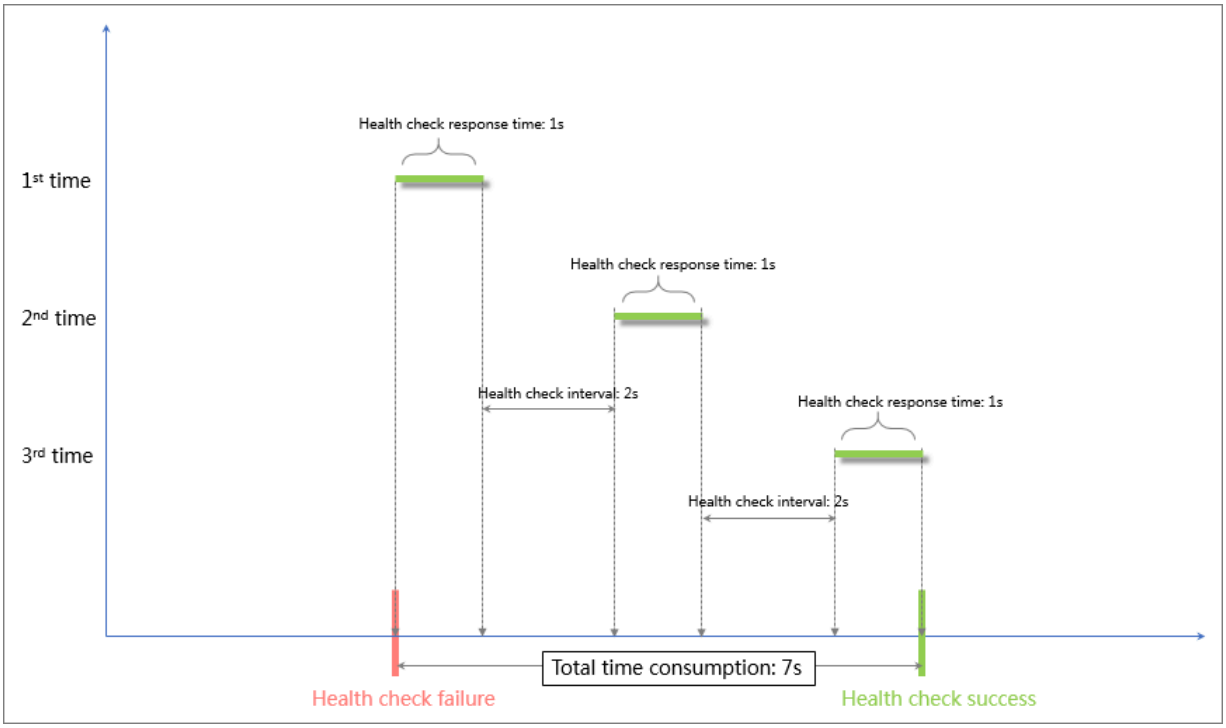
The following figure shows the process to declare an unhealthy backend server:

Health check success time window = Health check response time × Healthy Threshold + Health Check Interval × (Healthy Threshold - 1). That is, (1 × 3) + 2 × (3 - 1) = 7s.

Note:
Health check response time is the duration from the time when the health check request is sent to the time when the response is received. When the TCP health check is used, the time is very short and almost negligible because the health check only detects whether the port is alive. When the HTTP health check is used, the response

time depends on the performance and load of the application server and is usually within seconds.

The following figure shows the process to declare a healthy backend server (Assume that it takes one second for the backend server to respond to the health check request):



Configure a domain name in HTTP health checks

When the HTTP health check is used, you can set a domain name for the health check, but it is not required. Some application servers verify the host field in the request. Therefore, the request header must contain the host field. If a domain name is configured in the health check function, SLB adds the domain name to the host field when forwarding the request to the backend server. If not, the health check request will be denied by the server and the health check may fail. Therefore, if your application server verifies the host field in the request, you must configure a domain name to make sure the health check works.

3.3 Disable the health check function

If you disable the health check function, requests may be distributed to unhealthy ECS instances, resulting in disruption to your services. Therefore, we recommend that you enable the health check function.

Context

> Note:
> You can only disable the health check function for HTTP and HTTPS listeners. The health check function for UDP and TCP listeners cannot be disabled.

Procedure

1. Log on to the SLB console.
2. Select the region of the target SLB instance. On the Server Load Balancer page, find the target SLB instance and click the instance ID.
3. On the Listeners tab, find the target listener and click Configure in the Actions column.
4. On the Configure Listener page, click Next until the Health Check tab is displayed.
5. Turn off Enable Health Check, click Next, click Submit, and then click OK.

4 Backend servers

4.1 Backend server overview

Before using the load balancing service, you must add one or more ECS instances as the backend servers to an SLB instance to process the distributed client requests.

SLB service virtualizes the added ECS instances in the same region into an applicatio n pool featured with high performance and high availability. You can also manage backend servers through a VServer group. Different listeners can be associated with different server groups so that different listeners of an SLB instance can forward requests to the backend servers with different ports.

Note:

After a VServer group is configured for a listener, the listener will forward requests to the ECS instances in the associated VServer group instead of the ECS instances in the default server group.

You can increase or decrease the number of the backend ECS instances at any time and specify the ECS instances that receive requests. However, we recommend that you enable the health check function, and there must be at least one normal ECS to maintain service stability.

When adding ECS instances to an SLB instance, note the following:

- SLB does not support cross-region deployment. Make sure that the ECS instances and the SLB instance are in the same region.
- SLB does not limit the operating system used in the ECS instances as long as the applications deployed in the ECS instances are the same, and the data is consistent . However, we recommend that you use the same operating system for better management and maintenance.
- Up to 50 listeners can be added to an SLB instance. Each listener corresponds to an application deployed on backend ECS instances. The listening port of an SLB instance corresponds to the application port opened on the ECS instance.
- You can specify a weight for each ECS instance in the backend server pool. An ECS instance with a higher weight will receive a larger number of connection requests.

- If you have enabled the session persistence function, the requests distributed to the backend ECS instances may be imbalanced. If so, we recommend that you disable the session persistence function to check if the problem persists.

  When the traffic is not distributed evenly, troubleshoot as follows:

  1. Collect the access logs of the web service within a period of time.
  2. Check if the number of logs of multiple ECS instances are different according to SLB configurations. If session persistence is enabled, you need to strip the access logs for the same IP address. If weights are configured for SLB, you need to calculate whether the percentage of access logs matches the weight ratio.)

- When an ECS instance is undergoing live migration, the persistent connection s of the SLB may be interrupted and can be restored by reconnecting them. Be prepared for the reconnection.

Default server group

A default server group contains ECS instances that receive requests. If a listener is not associated with a VServer group or an active/standby server group, requests are forwarded to ECS instances in the default server group by default.

See Manage a default server group to create a default server group.

Active/standby server group

An active/standby server group only contains two ECS instances. One acts as the active server and the other acts as the standby server. No health check is performed on the standby server. When the active server is declared as unhealthy, the system forwards traffic to the standby server. When the active server is declared as healthy and restores service, the traffic is forwarded to the active server again.

See Manage an active/standby server group to create an active/standby server group.

> Note:
> Only Layer-4 listeners (TCP and UDP protocols) support configuring active/standby server groups.

VServer group

When you need to distribute different requests to different backend servers, or you want to configure domain name or URL based forwarding rules, you can use VServer groups.

See Manage a VServer group to create a VServer group.

4.2 Manage a default server group

Before you use the Server Load Balancer (SLB) service, you must add at least one default server to the default server group to receive client requests forwarded by SLB.

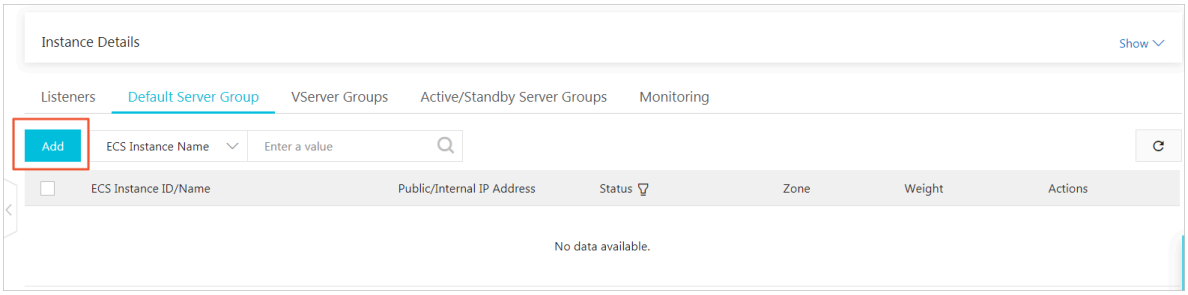
Add default servers

Before you add ECS instances to the default server group, make sure the following conditions are met:

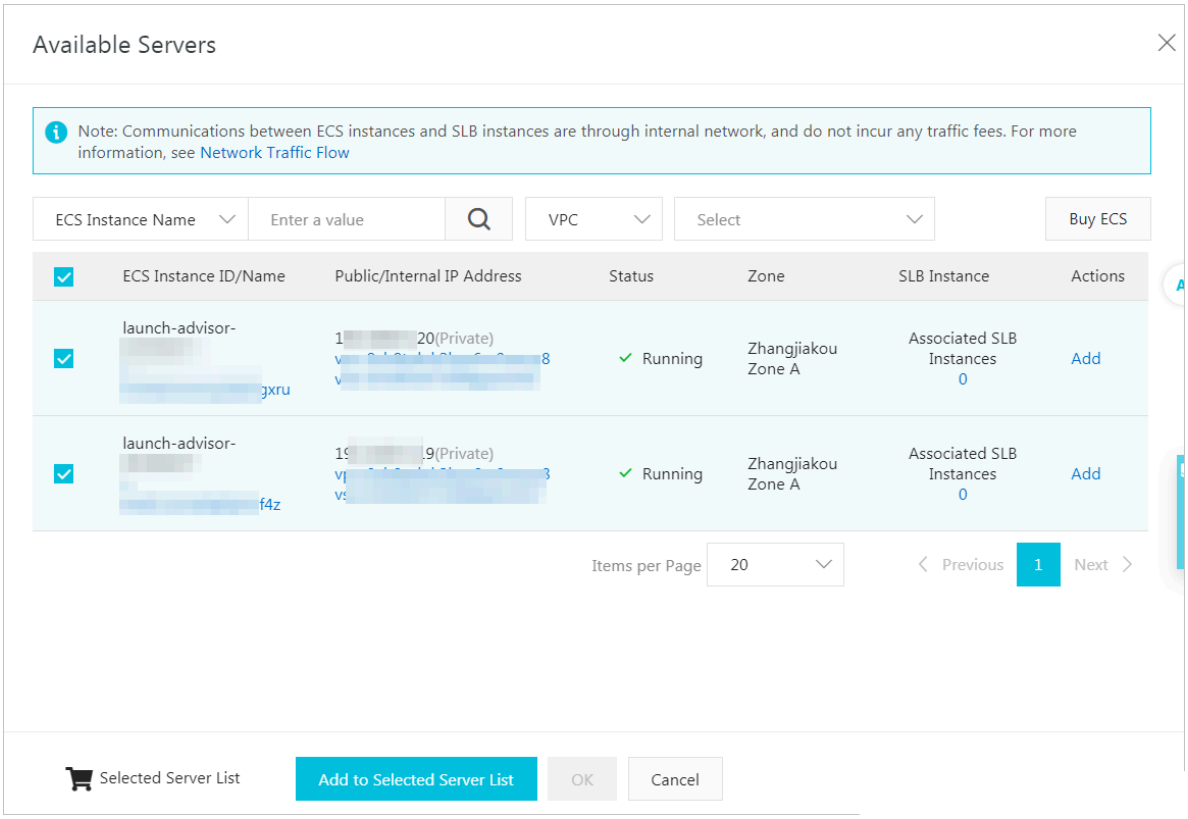
- An SLB instance is created. For more information, see Create an SLB instance.
- ECS instances are created and applications are deployed on the ECS instances to process distributed requests.

To add backend servers, follow these steps:

1. Log on to the SLB console.
2. On the Server Load Balancer page, select the region of the target SLB instance.
3. Find the target SLB instance and click the instance ID.
4. Click the Default Server Group tab.
5. Click Add.



6. On the Available Servers page, select the ECS instances to add to the default server group.



7. Click Next: Set Weight and Port.

8. On the Available Servers page, set the weights and ports of added ECS instances, and click OK.

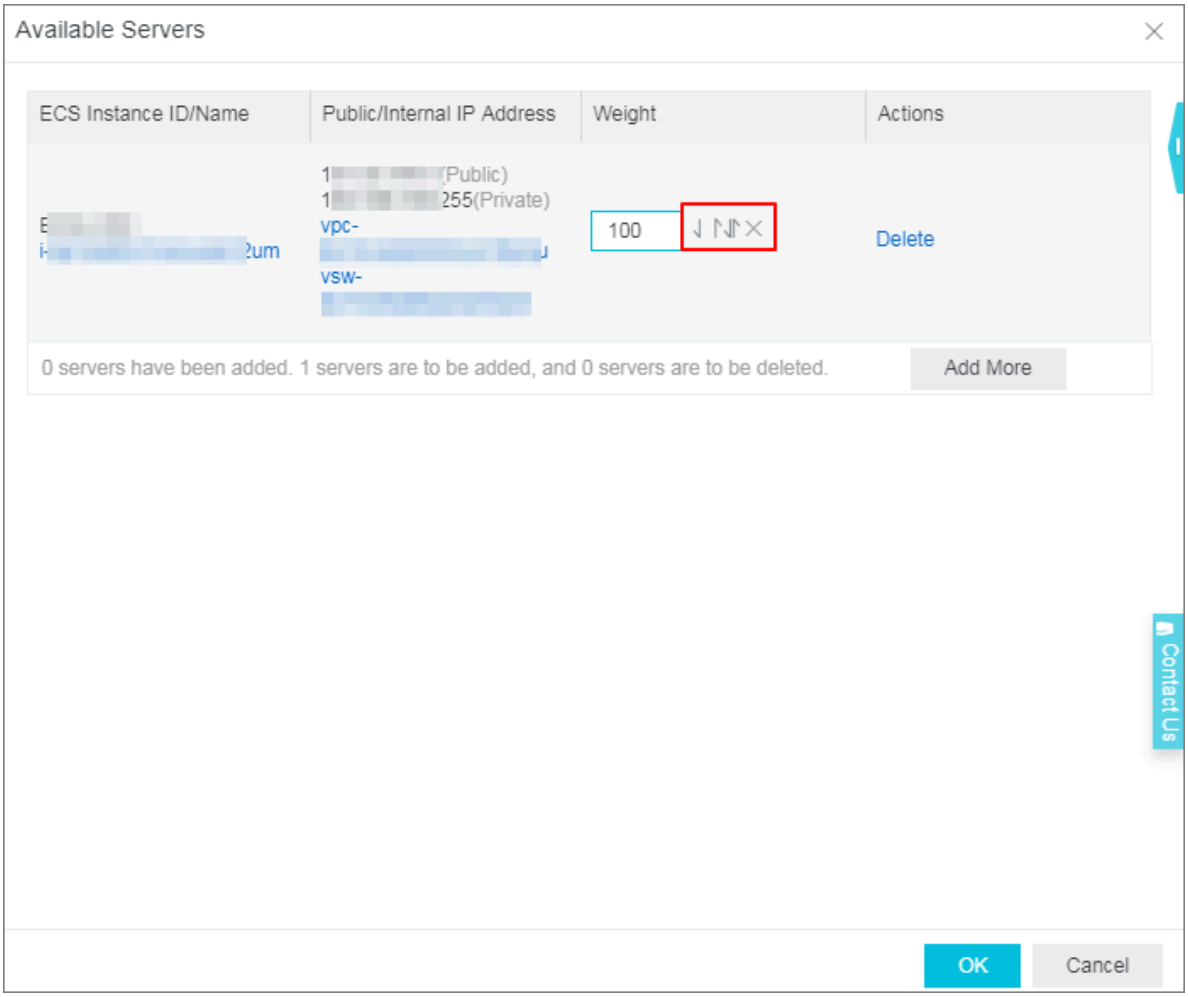
   Weight: An ECS instance with a higher weight receivers more requests.

   You can modify server weights in batches:

   - Click  : Duplicate to below. If you modify the weight of the current server, the weights of all servers blow are also changed.
   - Click  : Duplicate to above. If you modify the weight of the current server, the weights of all servers above are also changed.
   - Click  : Duplicate to all. If you modify the weight of the current server, the weights of all servers in the default server group are also changed.
   - Click  : Clear all. If you clear the weight of the current server, the weights of all servers in the default server group are also cleared.

    Notice:

If the weight is set to 0, the server no longer receives new requests.



9. Click OK.

Edit the weight of a backend server

To edit the weight of a backend server, follow these steps:

1. Log on to the SLB console.
2. On the Server Load Balancer page, select the region of the target SLB instance.
3. Find the target SLB instance and click the instance ID.
4. Click the Default Server Group tab.
5. Rest the pointer over the weight value of the target backend server, and then click the displayed pencil icon.

6. Modify the weight and then click OK.

   An ECS instance with a higher weight receives more requests.

   > Notice:
   > If the weight is set to 0, no requests are sent to the ECS instance.

Remove a backend server

To remove a backend server, follow these steps:

1. Log on to the SLB console.
2. On the Server Load Balancer page, select the region of the target SLB instance.
3. Find the target SLB instance and click the instance ID.
4. Click the Default Server Group tab.
5. Find the target backend server and click Remove in the Actions column.

4.3 Manage a VServer group

A virtual server group (VServer group) is a group of ECS instances. If you associate a VServer group with a listener, the listener distributes requests to the associated VServer group instead of other backend servers.

For Layer-7 listeners, the following algorithm is used to determine whether requests are forwarded to default backend server groups, or VServer groups, and whether forwarding rules are applied:

- If the requests match a forwarding rule, the requests are distributed to the VServer group associated with the rule.
- If no forwarding rule is matched and a VServer group is configured on the listener, the requests are distributed to the VServer group associated with the listener.
- If no VServer group is configured on the listener, the requests are forwarded to ECS instances in the default server group.

Create a VServer group

Before you create a VServer group, make sure the following conditions are met:

- A Server Load Balancer (SLB) instance is created. For more information, see Create an SLB instance.
- ECS instances are created and applications are deployed on the ECS instances to process distributed requests.

Note the following when you create a VServer group:

- The ECS instances added to a VServer group and the SLB instance must belong to the same region.
- One ECS instance can be added to multiple VServer groups.
- One VServer group can be associated with multiple listeners of an SLB instance.
- A VServer group consists of ECS instances and application ports.

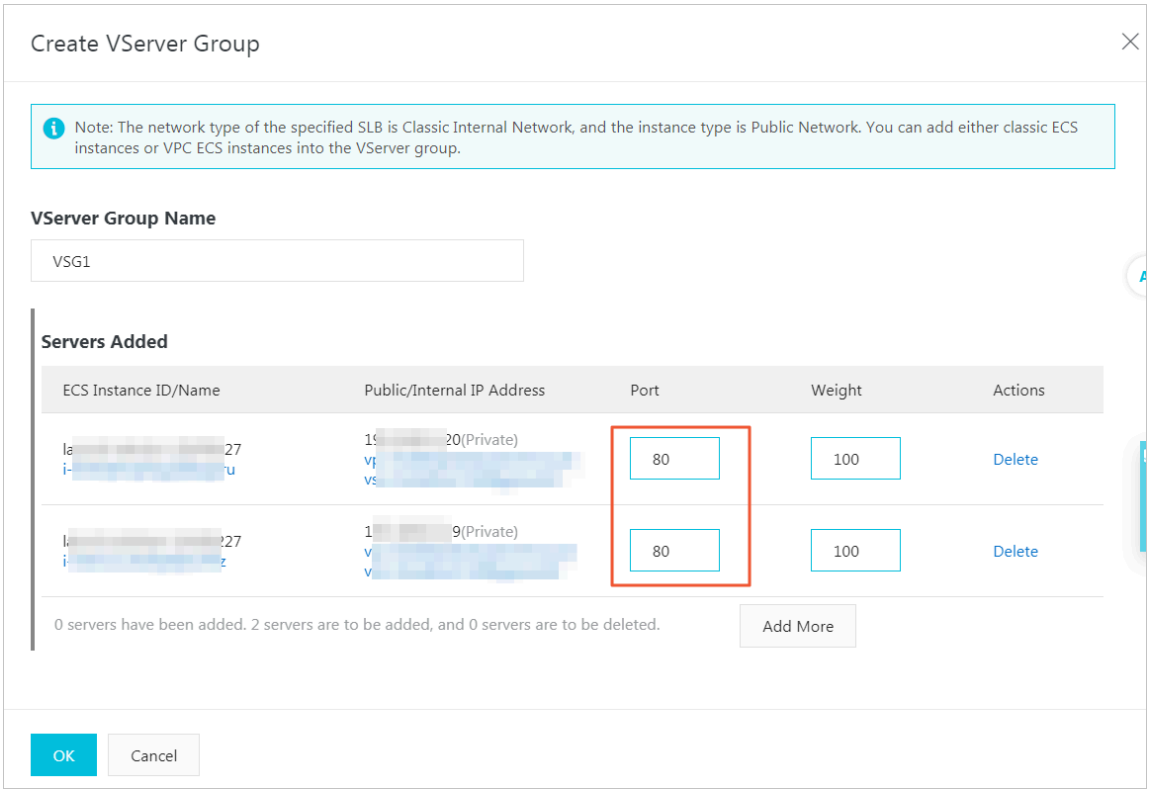
To add ECS instances, follow these steps:

1. Log on to the SLB console.
2. On the Server Load Balancer page, select the region of the target SLB instance.
3. Find the target SLB instance and click the instance ID.
4. Click the VServer Groups tab.
5. On the VServer Groups page, click Create VServer Group.
6. On the Create VServer Group page, complete these steps:
   a. In the VServer Group Name field, enter a name for the VServer group to be created.
   b. Click Add and on the Available Servers page, select the servers to add.
   c. Click Next: Set weight and Port.
   d. Enter the port and weight of each ECS instance, and click OK.
      - Port: The backend port opened on the ECS instance to receive requests.

        The backend ports in an SLB instance can be the same.
      - Weight: An ECS instance with a higher weight receivers more requests.

         Notice:

If the weight is set to 0, no requests are sent to the ECS instance.



You can modify the ports and weights of added servers in batches.

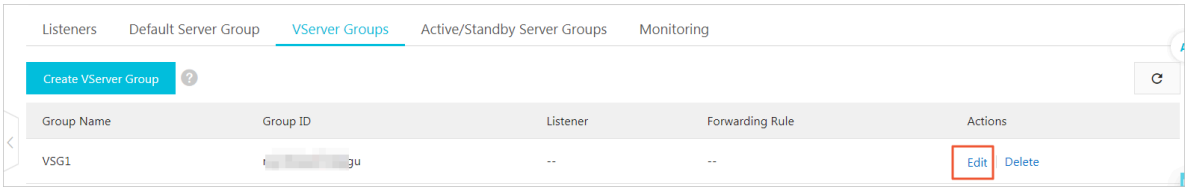
- Click : Duplicate to below. If you modify the port or weight of the current server, the ports or weights of all servers blow are also changed.
- Click : Duplicate to above. If you modify the port or weight of the current server, the ports or weights of all servers above are also changed.
- Click : Duplicate to all. If you modify the port or weight of the current server, the ports or weights of all servers in the VServer group are also changed.
- Click : Clear all. If you clear the port or weight of the current server, the ports or weights of all servers in the VServer group are also cleared.

Edit a VServer group

To modify the ECS instance configuration in a VServer group, follow these steps:

1. Log on to the SLB console.
2. On the Server Load Balancer page, select the region of the target instance.

3. Find the target SLB instance and click the instance ID.
4. Click the VServer Groups tab.
5. Find the target VServer group, and then click Edit in the Actions column.

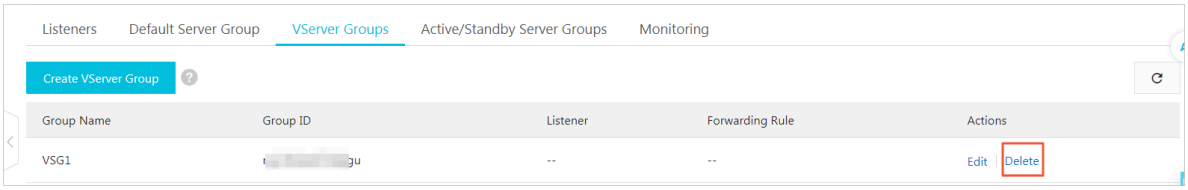


6. Modify the ports and weights of ECS instances or click Delete to remove ECS instances from the VServer group, and then click OK.

Delete a VServer group

To delete a VServer group, follow these steps:

1. Log on to the SLB console.
2. On the Server Load Balancer page, select the region of the target instance.
3. Find the target SLB instance and click the instance ID.
4. Click the VServer Groups tab.
5. Find the target VServer group, and then click Delete in the Actions column.



6. In the displayed dialog box, click OK.

4.4 Manage an active/standby server group

If you need active/standby failover configurations, where one backend server is used as the active server and the other as the standby server, you can create an active/standby server group. When the active server works normally, requests are distributed to the active server. If the active server is down, requests are distributed to the standby server to avoid service interruptions.

An active/standby server group only contains two ECS instances. One acts as the active server and the other acts as the standby server. No health check is performed on the standby server. When the active server is declared as unhealthy, the system forwards traffic to the standby server. When the active server is declared as healthy and restores service, the traffic is forwarded to the active server again.

> **Notice:**
> Only Layer-4 listeners (TCP and UDP protocols) support active/standby server groups.

Create an active/standby server group

Before you create an active/standby server group, make sure the following conditions are met:

- A Server Load Balancer (SLB) instance is created. For more information, see Create an SLB instance.
- ECS instances are created and applications are deployed on the ECS instances to process distributed requests.

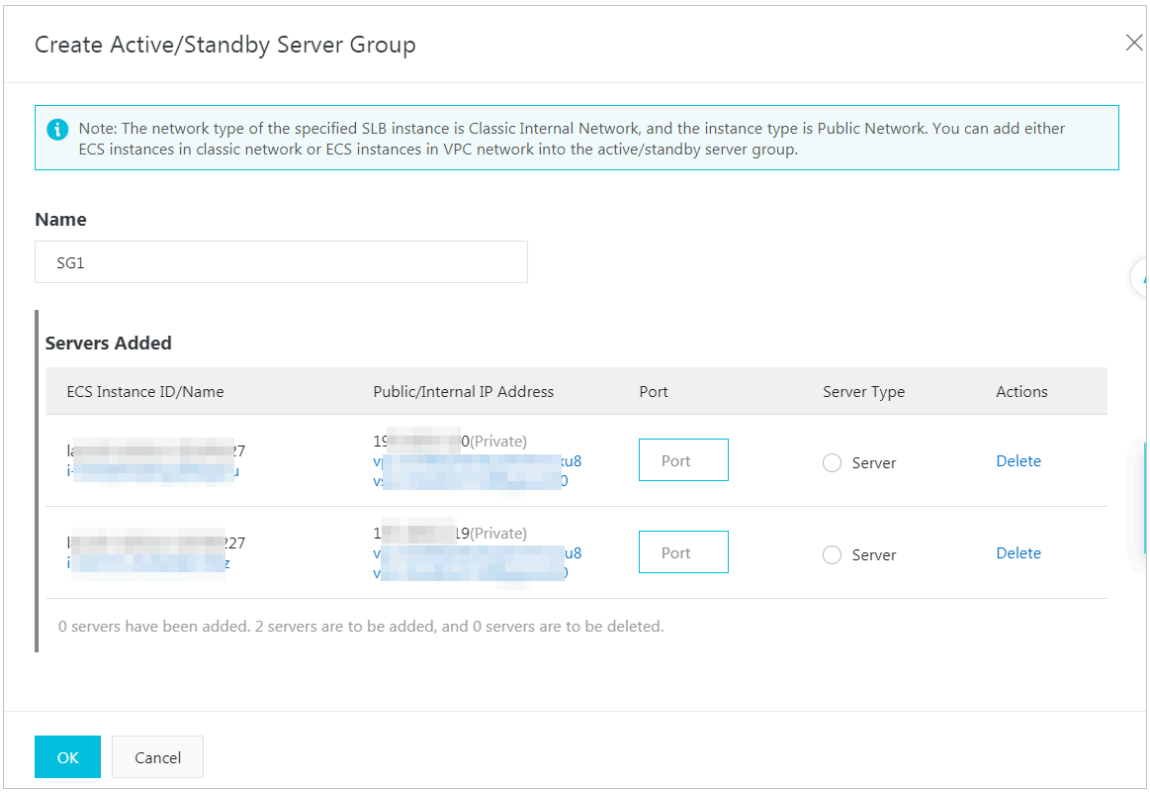
To create an active/standby server group, follow these steps:

1. Log on to the SLB console.
2. On the Server Load Balancer page, select the region of the target SLB instance.
3. Find the target SLB instance and click the instance ID.
4. Click the Active/Standby Server Groups tab.
5. On the Active/Standby Server Groups tab, click Create Active/Standby Server Group.

6. On the Create Active/Standby Server Group page, complete these steps:
   a. In the Name filed, enter a name for the active/standby server group to be created.
   b. Click Add and on the Available Servers page, select the servers to add.

      You can add up to two ECS instances to an active/standby server group.
   c. Click Next: Set Weight and Port.
   d. In the Servers Added section, set the port, select an active server, and click OK.
      - Port: The backend port opened on the ECS instance to receive requests.

        The backend ports in an SLB instance can be the same.
      - Server: Select a server to act as the active server.



Delete an active/standby server group

To delete an active/standby server group, follow these steps:

1. Log on to the SLB console.
2. On the Server Load Balancer page, select the region of the target SLB instance.
3. Click the ID of the target SLB instance.
4. Click the Active/Standby Server Groups tab.

5. Find the target active/standby server group and click Delete in the Actions column.

6. In the displayed dialog box, click OK.

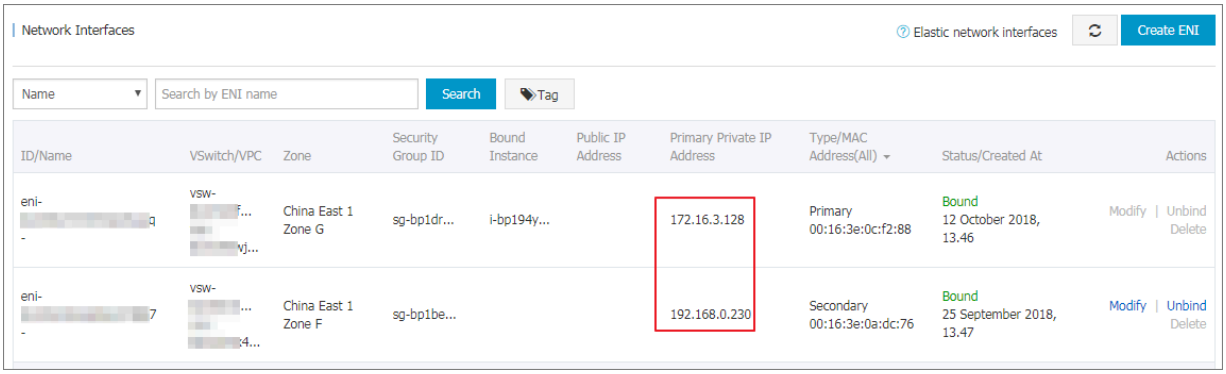
4.5 Add private IP addresses of ENIs to backend servers

An Elastic Network Interface (ENI) is a virtual network interface that can be attached to an ECS instance in a VPC. When you add backend servers to a guaranteed-performance Server Load Balancer (SLB) instance, you can choose to add the primary and secondary private IP addresses of ENIs if the ENIs are associated with ECS instances.

Prerequisites

The ECS instances are associated with ENIs.

For more information about how to associate an ENI with an ECS instance, see Attach an ENI.



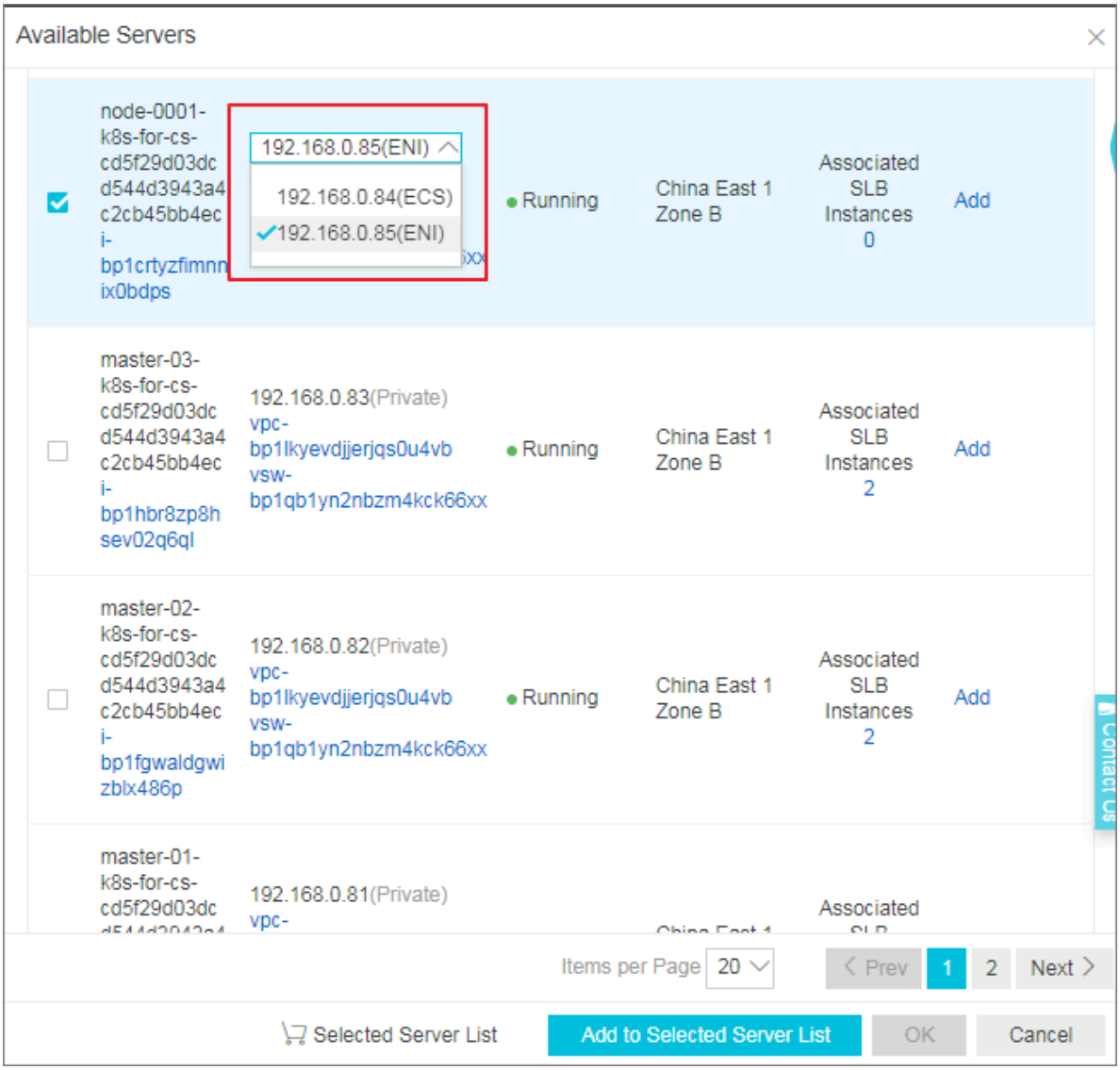
Note:

Only guaranteed-performance SLB instances support adding the primary and secondary private IP addresses of ENIs to backend servers.

Procedure

1. Log on to the SLB console.
2. In the left-side navigation pane, click Server Load Balancer. On the Server Load Balancer page, click the ID of the target SLB instance.

3. Select the backend server group type by clicking the corresponding tab. Default server groups, VServer groups, and active/standby server groups all support adding the primary and secondary private IP addresses of ENIs. In this topic, click the Default Server Group tab and then click Add.
4. On the Available Servers page, turn on Advanced Mode and click to select ENIs and its secondary private IP addresses.



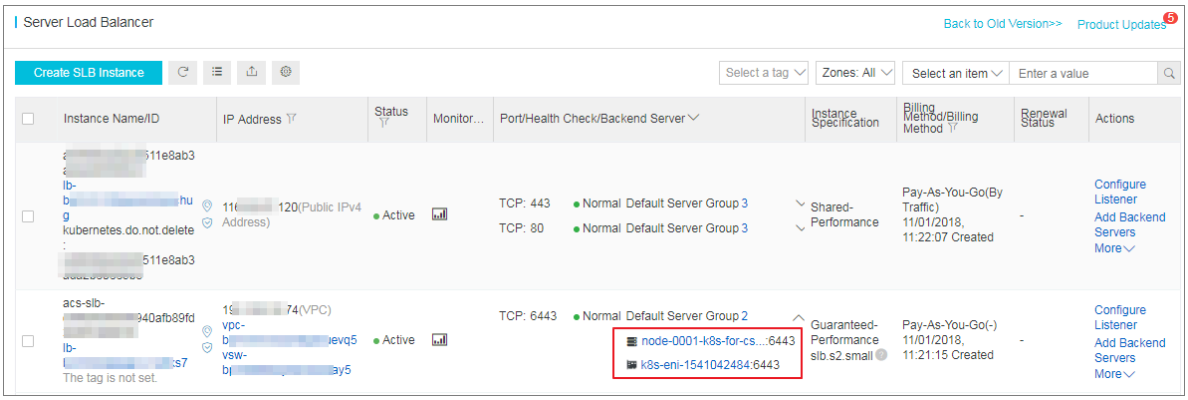
5. Click Next: Set Weight and Port to set the weights and port numbers of the added backend servers.

6. Click OK and you can see the added ENIs and its private IP addresses on the Default Server Group tab.

   If the default server group is added for a listener, you can see the backend servers added with ENIs and secondary private IP addresses on the Server Load Balancer page as follows:

   where,

   -  : Represents an ECS instance.
   -  : Represents an ENI and its secondary private IP address.



5 Certificate management

5.1 Certificate requirements

Server Load Balancer (SLB) only supports certificates in the PEM format. Before you upload a certificate, make sure that the certificate content, certificate chain, and private key conform to the corresponding format requirements.

Certificates issued by a root CA

If the certificate is issued by a root CA, the received certificate is the only one required to be uploaded to SLB. In this case, the website that is configured with the certificate will be regarded as a trusted website and does not require additional certificates.

The certificate format must meet the following format requirements:

- The certificate must start with `----- BEGIN CERTIFICAT E -----`, and end with `----- END CERTIFICAT E -----`, and both parts must be uploaded together.
- Each line except the last line must contain exactly 64 characters. The last line can contain 64 or fewer characters.
- Spaces are not allowed in the certificate content.

The following is a sample certificate issued by a root CA.

Certificates issued by an intermediate CA

If the certificate is issued by an intermediate CA, you are able to obtain multiple intermediate certificates. You must upload both the server certificate and the required number of immediate certificates to SLB.

The format of the certificate chain must meet the following requirements:

- Paste the server certificate content first, and then paste the content of the one or more required intermediate certificates directly underneath without any blank lines in between the certificates.
- Spaces are not allowed in the content.
- Blank lines are not allowed in the content. Each line must contain exactly 64 characters. For more information, see RFC1421.
- The certificate must conform to the corresponding format requirements. Generally , the intermediate CA will provide instructions about the certificate format when issuing the certificate. The certificate chain must conform to the format requirements.

The following is a sample certificate chain.

```
-----BEGIN CERTIFICATE-----
-----END CERTIFICATE-----
-----BEGIN CERTIFICATE-----
-----END CERTIFICATE-----
-----BEGIN CERTIFICATE-----
-----END CERTIFICATE-----
```

RSA private key

When you upload the server certificate, you also need to upload the private key of the certificate.

The RSA private key format must meet the following requirements:

- The private key must start with `-----BEGIN RSA PRIVATE KEY-----`, and end with `-----END RSA PRIVATE KEY-----`, and both parts must be uploaded together.
- Blank lines are not allowed in the content. Each line except the last line must contain exactly 64 characters. The last line can contain 64 or fewer characters. For more information, see RFC1421.

If your private key is encrypted (for example, the content at the beginning and end of the private key is `-----BEGIN PRIVATE KEY-----`, `-----END PRIVATE KEY-----` or `-----BEGIN ENCRYPTED PRIVATE KEY-----`, `-----END ENCRYPTED PRIVATE KEY-----`, or the private key contains `Proc-Type:4,ENCRYPTED`), you must first run the following command to convert the private key:

```
openssl rsa -in old_server_key.pem -out new_server_key.pem
```

The following is a sample RSA private key.

EC private key

> **Note:**
> Currently, EC private keys are supported only in the UK (London) region.

When you upload the server certificate, you also need to upload the private key of the certificate.

The EC private key format must meet the following requirements:

- The private key must start with `----- BEGIN EC PARAMETERS -----`, and end with `----- END EC PARAMETERS -----`, and both parts must be uploaded together.
- Blank lines are not allowed in the content. Each line except the last line must contain exactly 64 characters. The last line can contain 64 or fewer characters. For more information, see RFC1421.

If your private key is encrypted (for example, the content at the beginning and end of the private key is `----- BEGIN EC PRIVATE KEY -----`, `----- END`

`EC PRIVATE KEY -----`, or the private key contains `Proc-Type: 4, ENCRYPTED`), you must first run the following command to convert the private key:

```
openssl rsa -in old_server_key.pem -out new_server_key.pem
```

The following is a sample EC private key.

```
-----BEGIN EC PARAMETERS-----
Bggq**********Bw==
-----END EC PARAMETERS-----
-----BEGIN EC PRIVATE KEY-----
MHcCAQEEIC o9b+vQUhqFUWgW jE0YY4h0b3 bE/udcubxVwcV
Y99MuoAoGC CqGSM49
AwEHoUQDQg AEgpla3Bj9 rX**********4xz0SHsuQc/7XBmgmrMpA
mE80c0DR
5HcMHFxRPt GLv22T62e5 KqN1W3uN9H plgg==
-----END EC PRIVATE KEY-----
```

5.2 Create a certificate

To configure an HTTPS listener, you can directly use a certificate from Alibaba Cloud SSL Certificate Service or upload a third-party server certificate and CA certificate to Server Load Balancer (SLB). After you upload the certificate to SLB, you do not need to configure certificates on backend servers.

SLB supports certificates from the following two sources:

- Certificates issued or hosted by Alibaba Cloud SSL Certificate Service: You can select the required certificate from Alibaba Cloud SSL Certificate Service. When the certificate is about to expire, Alibaba Cloud will send alerts notifying you to renew the certificate to ensure its validity.

  Currently client CA certificates are not supported.
- Third-party certificates: To upload a third-party certificate, you must have the public key and private key files of the certificate.

  HTTPS server certificates and client CA certificates are supported.

Before you create a certificate, note the following:

- If you need to use a certificate in multiple regions, you must select all the required regions when creating the certificate.
- Each Alibaba Cloud account can create up to 100 certificates.

Select a certificate from SSL Certificate Service

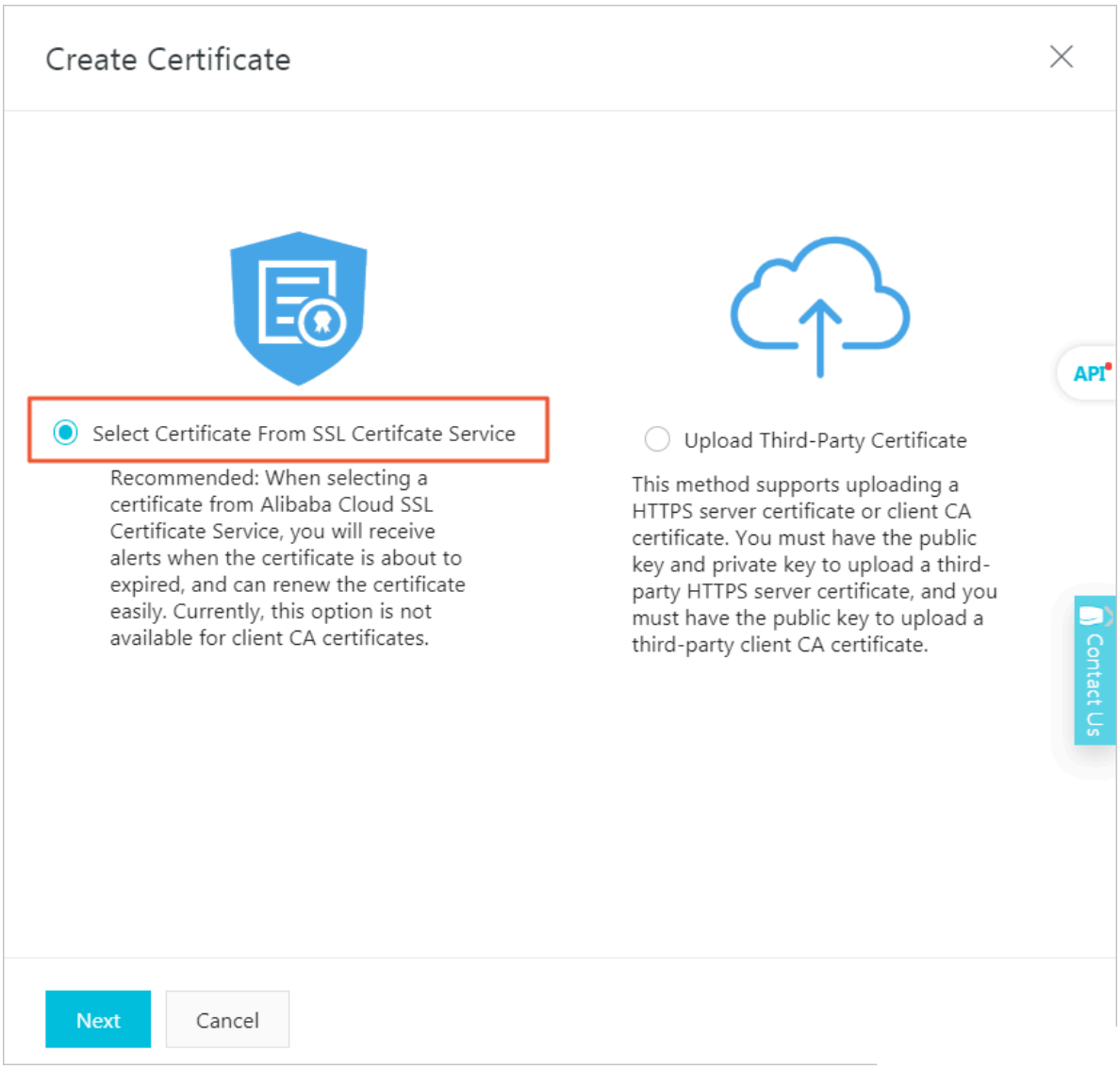
Alibaba Cloud SSL Certificate Service issues digital certificates of a variety of authorities to provide HTTPS services. Additionally, Alibaba Cloud SSL Certificate Service can uniformly manage the life cycles of certificates to simplify certificate deployment. For more information, see SSL certificate service.

To use a certificate in SSL Certificate Service, you must log on to the SSL Certificate console to buy a certificate or upload a third-party certificate to SSL Certificate Service.

To use a certificate from SSL Certificate Service, follow these steps:

1. Log on to the SLB console.
2. In the left-side navigation pane, click Certificates.

3. Click Create Certificate. On the Create Certificate page, select Select Certificate From SSL Certificate Service.



4. Click Next. On the Select Certificate From SSL Certificate Service page, select the region to deploy the certificate and then select the SSL certificate to use from the certificate list.

   A certificate cannot be used across regions. If you need to use a certificate in multiple regions, you must select all the required regions.

5. Click OK.

Upload a third-party certificate

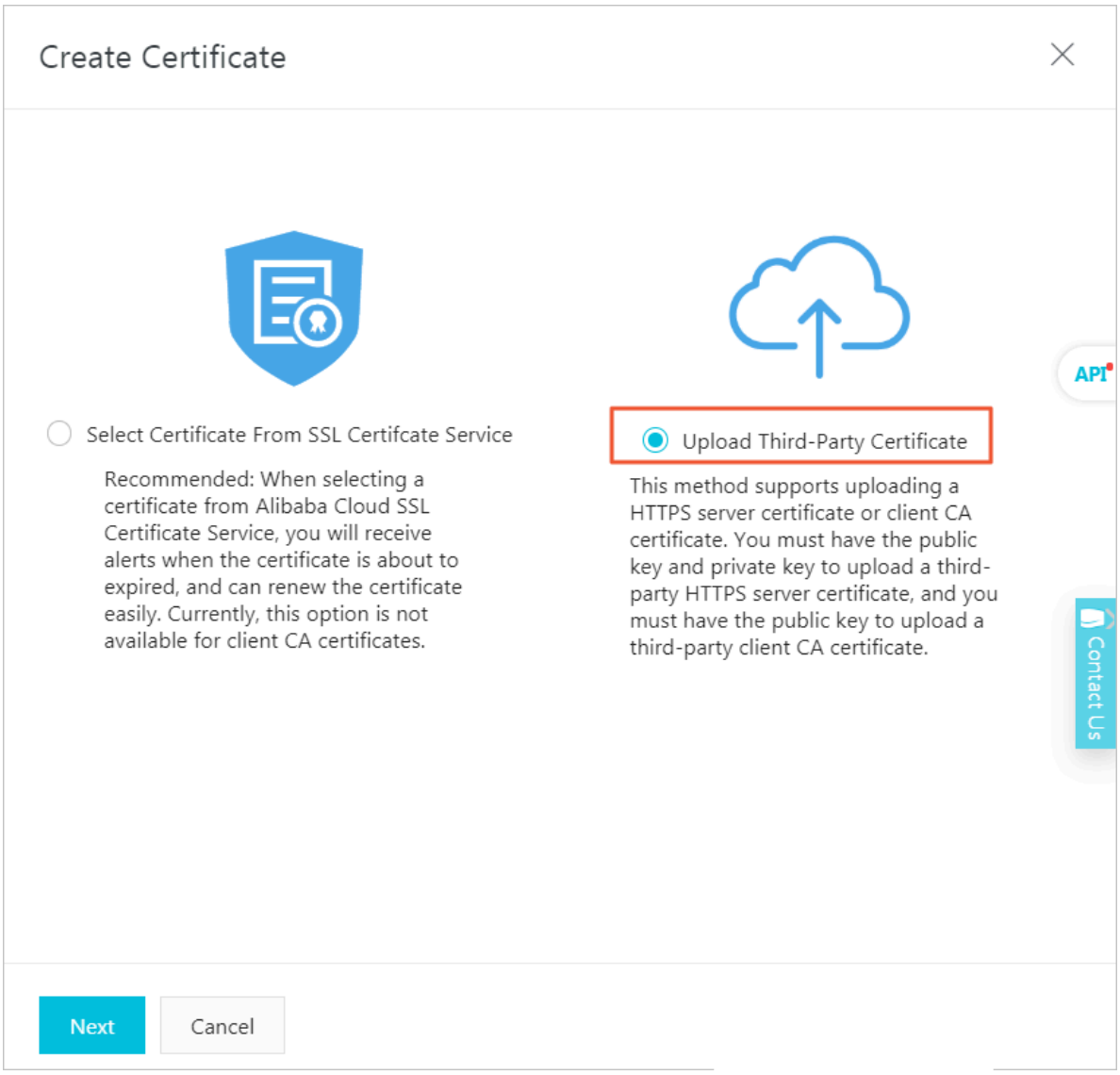
Before you upload a third-party certificate, make sure that the following conditions are met:

- A server certificate is purchased.

- A CA certificate and a client certificate are generated. For more information, see Generate a CA certificate.

To upload a third-party certificate to SLB, follow these steps:

1. Log on to the SLB console.
2. In the left-side navigation pane, click Certificates.
3. Click Create Certificate.
4. On the Create Certificate page, select Upload Third-Party Certificate.



5. Click Next. On the Upload Third-Party Certificate page, upload the certificate content.

Configuration	Description
Certificate Name	Enter a name for the certificate to be uploaded. The name must be 1 to 80 characters in length, and can contain letters, numbers, and the following special characters: _ / . -
Regions	Select one or more regions to which the certificate to be uploaded belongs. A certificate cannot be used across regions. If you need to use a certificate in multiple regions, you must select all the required regions.
Certificate Type	Select the type of the certificate to be uploaded: · Server Certificate: For HTTPS one-way authentication, only the server certificate and the private key are required. · CA Certificate: For HTTPS mutual authentication, both the server certificate and the CA certificate are required.
Certificate Content	Paste the certificate content into the text editor. Click View Sample Certificate to view the valid certificate formats. For more information, see Certificate requirements.

Configuration	Description

Private Key	Paste the private key of the server certificate into the text editor.
	Click View Sample Certificate to view the valid certificate formats. For more information, see Certificate requirements.
	SLB supports the following two private key formats:
	`----- BEGIN RSA PRIVATE KEY -----`
	`Private key content ( BASE64 encoding )`
	`----- END RSA PRIVATE KEY -----`
	or
	`----- BEGIN EC PARAMETERS -----`
	`Private key content ( BASE64 encoding )`
	`----- END EC PARAMETERS -----`
	`----- BEGIN EC PRIVATE KEY -----`
	`Private key content ( BASE64 encoding )`
	`----- END EC PRIVATE KEY -----`
	Notice:
	· A private key is required only when you upload a server certificate.
	· Currently, EC private keys are supported only in the UK (London) region.

6. Click OK.

5.3 Generate a CA certificate

When configuring HTTPS listeners, you can use self-signed CA certificates. Follow the instructions in this document to generate a CA certificate and use the CA certificate to sign a client certificate.

Generate a CA certificate by using Open SSL

1. Run the following commands to create a `ca` folder in the `/root` directory and then create four sub folders under the `ca` folder.

```
$ sudo mkdir ca
$ cd ca
```

```
$ sudo mkdir newcerts private conf server
```

- *newcerts* is used to store the digit certificate signed by a CA certificate.
- *private* is used to store the private key of the CA certificate.
- *conf* is used to store the configuration files.
- *server* is used to store the server certificate.

2. Create an *openssl . conf* file that contains the following information in the *conf* directory.

```
[ ca ]
 default_ca = foo
[ foo ]
 dir = / root / ca
 database = / root / ca / index . txt
 new_certs_ dir = / root / ca / newcerts
 certificat e = / root / ca / private / ca . crt
 serial = / root / ca / serial
 private_ke y = / root / ca / private / ca . key
 RANDFILE = / root / ca / private /. rand
 default_da ys = 365
 default_cr l_days = 30
 default_md = md5
 unique_sub ject = no
 policy = policy_any
[ policy_any ]
 countryNam e = match
 stateOrPro vinceName = match
 organizati onName = match
 organizati onalUnitNa me = match
 localityNa me = optional
 commonName = supplied
 emailAddre ss = optional
```

3. Run the following command to generate a private key.

```
$ cd / root / ca
 $ sudo openssl genrsa - out private / ca . key
```

The following figure is an example of key generation.

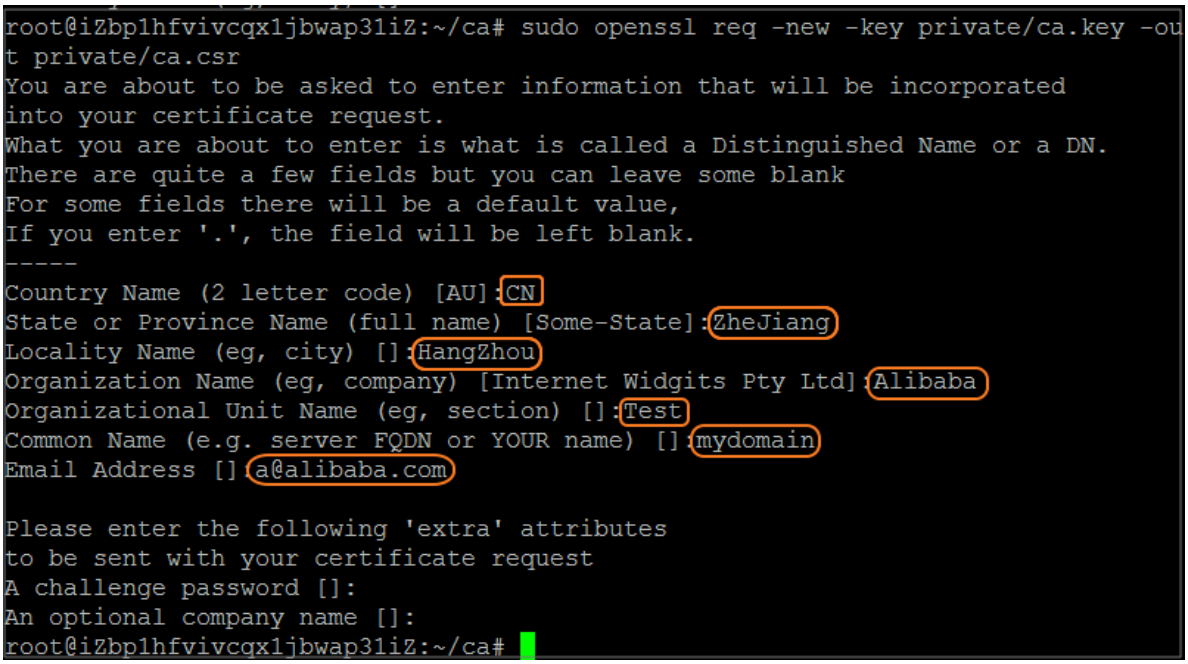
```
root@iZbp1hfvivcqx1jbwap31iZ:~/ca/conf# cd /root/ca
root@iZbp1hfvivcqx1jbwap31iZ:~/ca# sudo openssl genrsa -out private/ca.key
Generating RSA private key, 2048 bit long modulus
.....+++
..+++
e is 65537 (0x10001)
```

4. Run the following command and input the required information according to the prompts. Press Enter to generate a *csr* file.

```
$ sudo openssl req -new -key private/ca.key -out
 private/ca.csr
```

Note:

Common Name is the domain name of the SLB instance.



5. Run the following command to generate a *crt* file.

```
$ sudo openssl x509 -req -days 365 -in private/ca.csr -signkey private/ca.key -out private/ca.crt
```

6. Run the following command to set the start sequence number for the private key, which can be any four characters.

```
$ sudo echo FACE > serial
```

7. Run the following command to create a CA key library.

```
$ sudo touch index.txt
```

8. Run the following command to create a certificate revocation list for removing the client certificate.

```
$ sudo openssl ca -gencrl -out /root/ca/private/ca.crl -crldays 7 -config "/root/ca/conf/openssl.conf"
```

The response is as follows:

```
Using configuration from /root/ca/conf/openssl.conf
```

Sign the client certificate

1. Run the following command to generate a *users* folder under the *ca* directory to store the client key.

```
$ sudo mkdir users
```

2. Run the following command to create a key for the client certificate.

```
$ sudo openssl genrsa -des3 -out /root/ca/users/client.key 1024
```

Note:

Enter a pass phrase when creating the key. It is the password to protect the private key from unauthorized access. The pass phrase entered is the password for this key.

3. Run the following command to create a *csr* file for requesting certificate sign.

```
$ sudo openssl req -new -key /root/ca/users/client.key -out /root/ca/users/client.csr
```

Enter the pass phrase set in the previous step when prompted.

> **Note:**
> A challenge password is the password of the client certificate. Note that it is not the password of the client key.

4. Run the following command to sign the client key.

```
$ sudo openssl ca -in /root/ca/users/client.csr -cert /root/ca/private/ca.crt -keyfile /root/ca/private/ca.key -out /root/ca/users/client.crt -config "/root/ca/conf/openssl.conf"
```

Enter y twice when prompted.

```
root@iZbp1hfvivcqx1jbwap31iZ:~/ca# sudo openssl ca -in /root/ca/users/client.csr
 -cert /root/ca/private/ca.crt -keyfile /root/ca/private/ca.key -out /root/ca/us
ers/client.crt -config "/root/ca/conf/openssl.conf"
Using configuration from /root/ca/conf/openssl.conf
Check that the request matches the signature
Signature ok
The Subject's Distinguished Name is as follows
countryName           :PRINTABLE:'CN'
stateOrProvinceName   :ASN.1 12:'ZheJiang'
localityName          :ASN.1 12:'HangZhou'
organizationName      :ASN.1 12:'Alibaba'
organizationalUnitName:ASN.1 12:'Test'
commonName            :ASN.1 12:'mydomain'
emailAddress          :IA5STRING:'a@alibaba.com'
Certificate is to be certified until Jun  4 15:28:55 2018 GMT (365 days)
Sign the certificate? [y/n]:y


1 out of 1 certificate requests certified, commit? [y/n]y
Write out database with 1 new entries
Data Base Updated
root@iZbp1hfvivcqx1jbwap31iZ:~/ca#
```

5. Run the following command to convert the certificate to a *PKCS12* file.

```
$ sudo openssl pkcs12 -export -clcerts -in /root/ca/users/client.crt -inkey /root/ca/users/client.key -out /root/ca/users/client.p12
```

Enter the password of the client key when prompted. Then, enter the password used for exporting the client certificate. This password is used to protect the client certificate, which is required when installing the client certificate.

6. Run the following command to view the generated client certificate.

```
cd users
ls
```

5.4 Convert the certificate format

Server Load Balancer supports PEM certificates only. Certificates in other formats must be converted to the PEM format before they can be uploaded to Server Load Balancer. We recommend that you use Open SSL for conversion.

Convert DER to PEM

DER: This format is usually used on a Java platform. The certificate file suffix is generally .*der*, .*cer*, or .*crt*.

- Run the following command to convert the certificate format:

```
openssl x509 -inform der -in certificate.cer -out certificate.pem
```

-

- Run the following command to convert the private key:

```
openssl rsa -inform DER -outform PEM -in privatekey.der -out privatekey.pem
```

Convert P7B to PEM

P7B: This format is usually used in a Windows server and Tomcat.

Run the following command to convert the certificate format:

```
openssl pkcs7 -print_cert s -in incertific ate.p7b -out outcertifi cate.cer
```

Convert PFX to PEM

PFX: This format is usually used in a Windows server.

- Run the following command to extract the certificate:

```
openssl pkcs12 -in certname.pfx -nokeys -out cert.pem
```

- Run the following command to extract the private key:

```
openssl pkcs12 -in certname.pfx -nocerts -out key.pem -nodes
```

5.5 Replace a certificate

To avoid the impact of certificate expiration on your service, replace the certificate before the certificate expires.

Procedure

1. Create and upload a new certificate.

   For more information, see Upload certificates and Generate certificates.

2. Configure the new certificate in HTTPS listener configuration.

   For more information, see Add an HTTPS listener.

3. On the Certificates page, find the target certificate, and then click Delete.
4. In the displayed dialog box, click OK.

6 Log management

6.1 View operation logs

You can view the logs of operations performed on SLB instances, HTTP listeners and server certificates in the past one month.

Context

The operation logs are recorded in ActionTrail. ActionTrail records the operations acted upon your Alibaba Cloud resources. You can query operation records and store the records to OSS.

Procedure

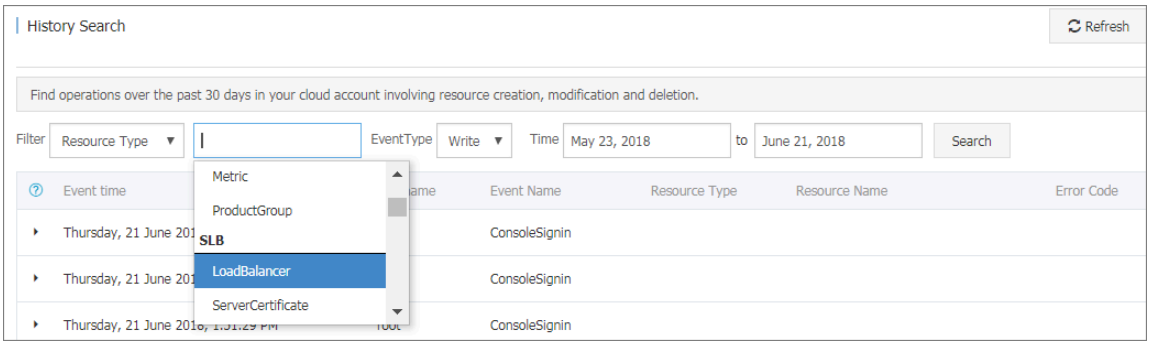
1. Log on to the SLB console.
2. In the left-side navigation pane, click Logs > Operation Log.
3. Click View Operation Logs.

4. On the History Search page, complete these steps to view operation logs:

   a) Select Resource Type as a filter.

   b) Select the SLB resource of which operation logs you want to view.

   In this tutorial, LoadBalancer is selected.



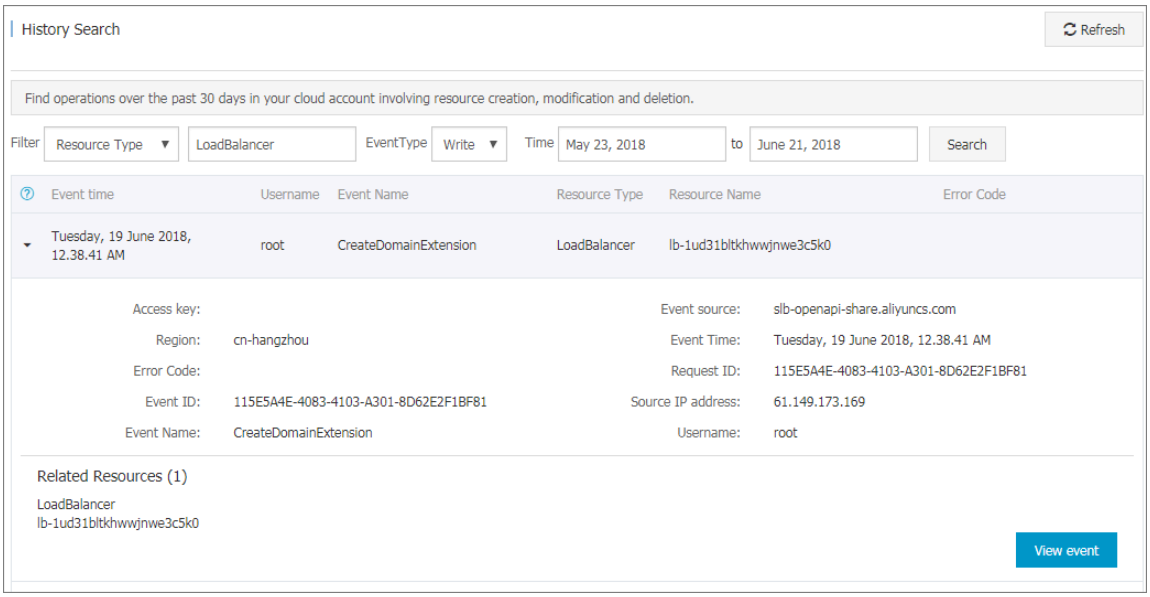
   c) Select an event type.

   In this tutorial, All is selected.

   d) Select the time range to search.

   e) Click Search to view logs of operations performed on the selected resource.

   Expand a record to view more detailed information.



6.2 Manage health check logs

You can view the health logs of Server Load Balancer (SLB) within three days on the Health Check Logs page. If you want to get health check logs generated three days

or longer before, you can store the health check logs to OSS and download complete health check logs.

Store health check logs

You can view the health check logs of backend servers by using the health check log function of SLB. Currently, logs in the past three days are provided. If you want to view more logs, store the health check logs to OSS buckets.

You can enable and disable the storage function at any time. After the storage function is enabled, SLB will create a folder named `AliyunSLBHealthCheckLogs` in the selected bucket to store health check logs of SLB. Health check logs are generated on an hourly basis and the system will create a subfolder named after the date to store the log files generated in that day, for example, `20170707`.

The log files generated in each hour of a day are named after the time when they are generated. For example, the file name of a log file generated between 00:00-01:00 is `01.txt` and the file name of a log file generated between 01:00-02:00 is `02.txt`.

 Note:
Health check logs are generated only when backend servers are abnormal. If no failures occur to backend servers in an hour, no health check logs are generated in that hour.

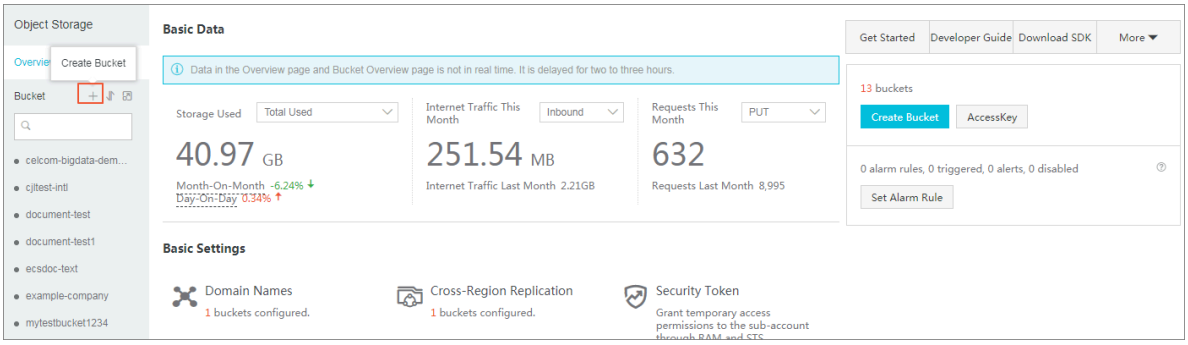
To store health check logs, follow these steps:

1. Create a bucket
2. Authorize SLB to access OSS
3. Configure log storage

Step 1 Create a bucket

1. Open the OSS product page and click Buy Now to activate the OSS service.
2. Log on to the OSS console.

3. Click Create Bucket.



4. In the Create Bucket dialog box, configure the bucket and click OK.

> **Note:**
> Make sure that the bucket and the SLB instance belong to the same region.

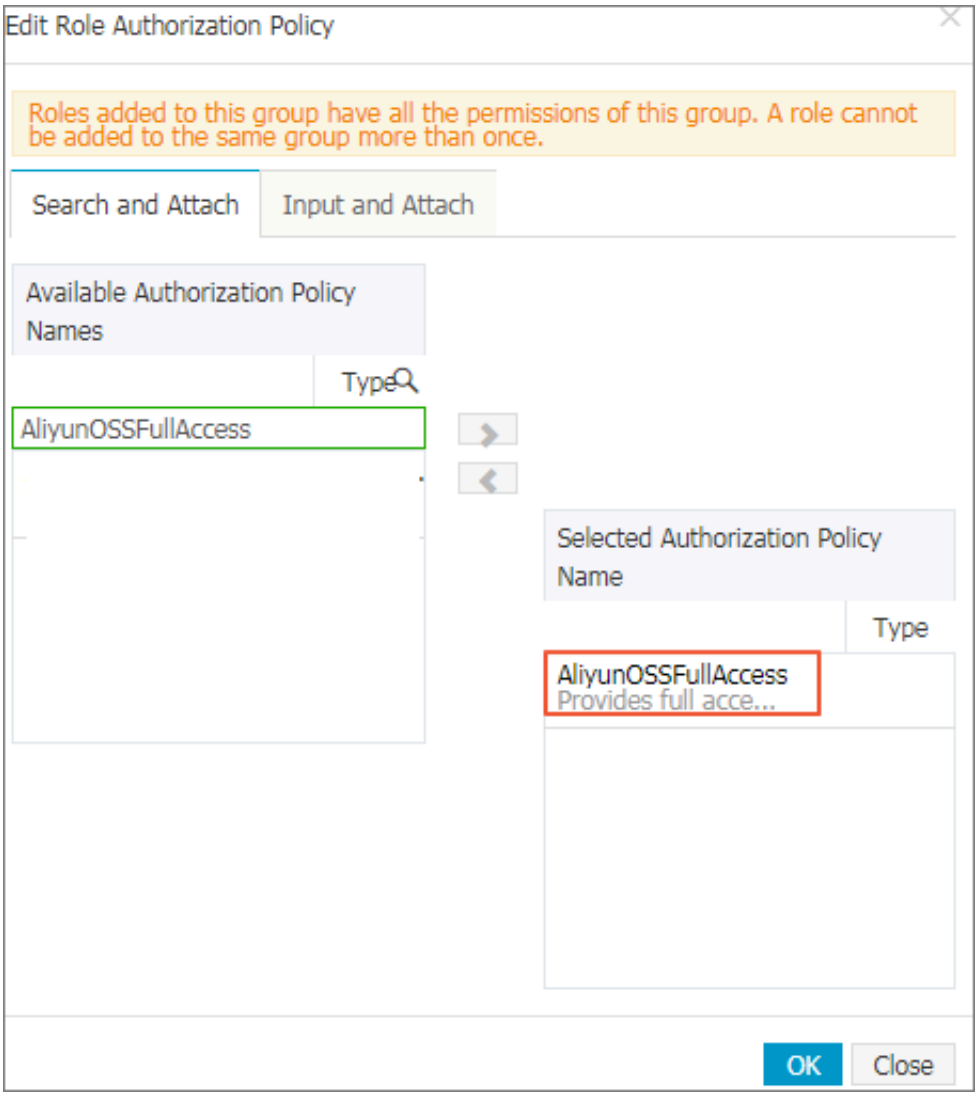
Step 2 Authorize SLB to access OSS

After creating a bucket, you must authorize the role (`SLBLogDefa ultRole`) to access OSS resources.

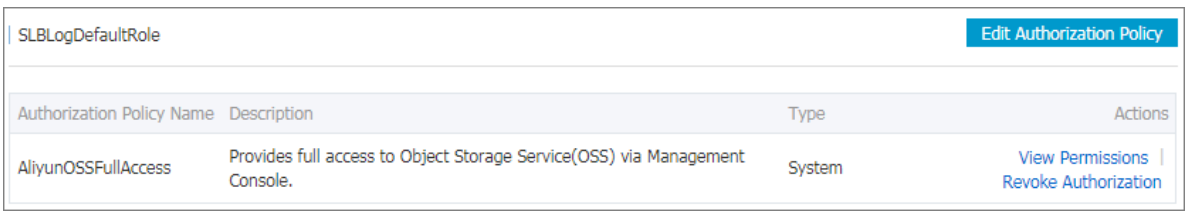
> **Notice:**
> The authorization is required only for the first time.

1. In the left-side navigation pane of the SLB console, click Logs > Health Check Logs.
2. Click 1. Activate OSS. if OSS has not been activated yet.
3. On the Health Check Logs page, click Authorize Now in the 2. Authorize the required RAM role. section.
4. Read the authorization description, and then click Confirm Authorization Policy.
5. Log on to the RAM console.
6. In the left-side navigation pane, click Roles, find the role named SLBLogDefaultRole, and then click Authorize.

7. In the Edit Role Authorization Policy dialog box, find the AliyunOSSFullAccess policy, click the policy, and then click OK.



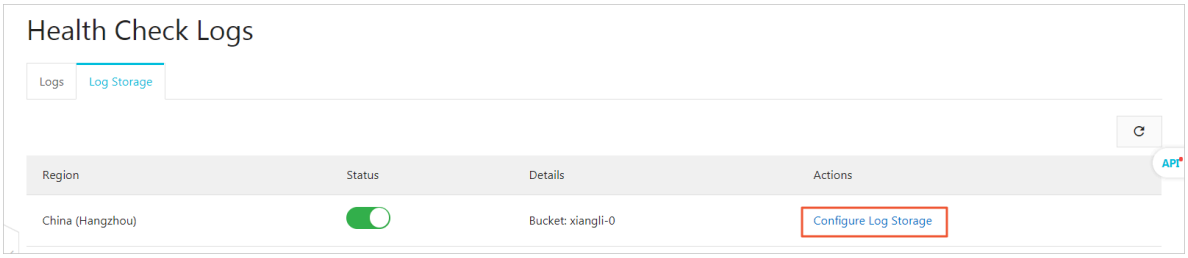
After the authorization, click the role name of SLBLogDefaultRole, and then click the Role Authorization Policies tab to view the attached policy.



Step 3 Configure log storage

1. Log on to the SLB console.
2. In the left-side navigation pane, click Logs > Health Check Logs.
3. On the Health Check Logs page, click the Log Storage tab.

4. Find the target region and click Configure Log Storage.



5. In the Configure Log Storage dialog box, select a bucket to store health check logs, and then click OK.
6. Turn on the status switch to enable log storage.

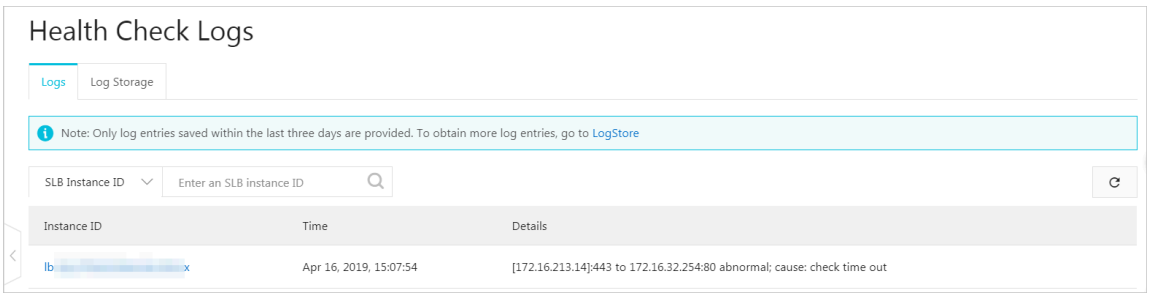
View health check logs

To view the health check logs generated in the past three days, follow these steps:

1. Log on to the SLB console.
2. In the left-side navigation pane, click Logs > Health Check Logs.
3. On the Health Check Logs page, click the Logs tab.

> **Note:**
> Health check logs are generated only when the health status of a backend server is abnormal. Health check logs are generated every one hour. If no failure occurs to backend servers in an hour, no health check logs are generated in that hour.

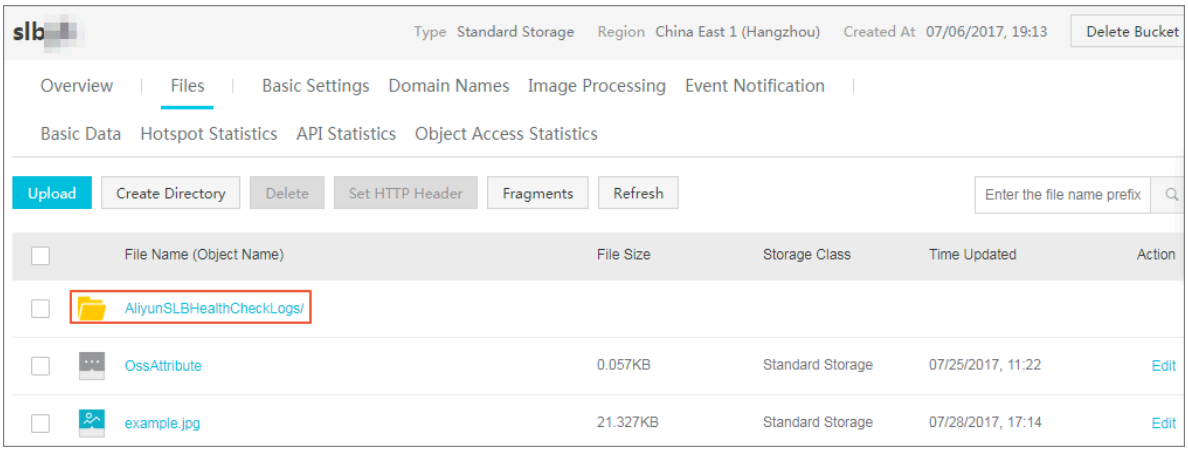
- The `SLB_instance_IP:port to Added_ECS_instance_IP:port abnormal; cause: XXX` log message indicates that the backend server is abnormal. Troubleshoot according to the detailed error message.
- The `SLB_instance_IP:port to Added_ECS_instance_IP:port normal` log message indicates that the backend server becomes normal again.



Download health check logs

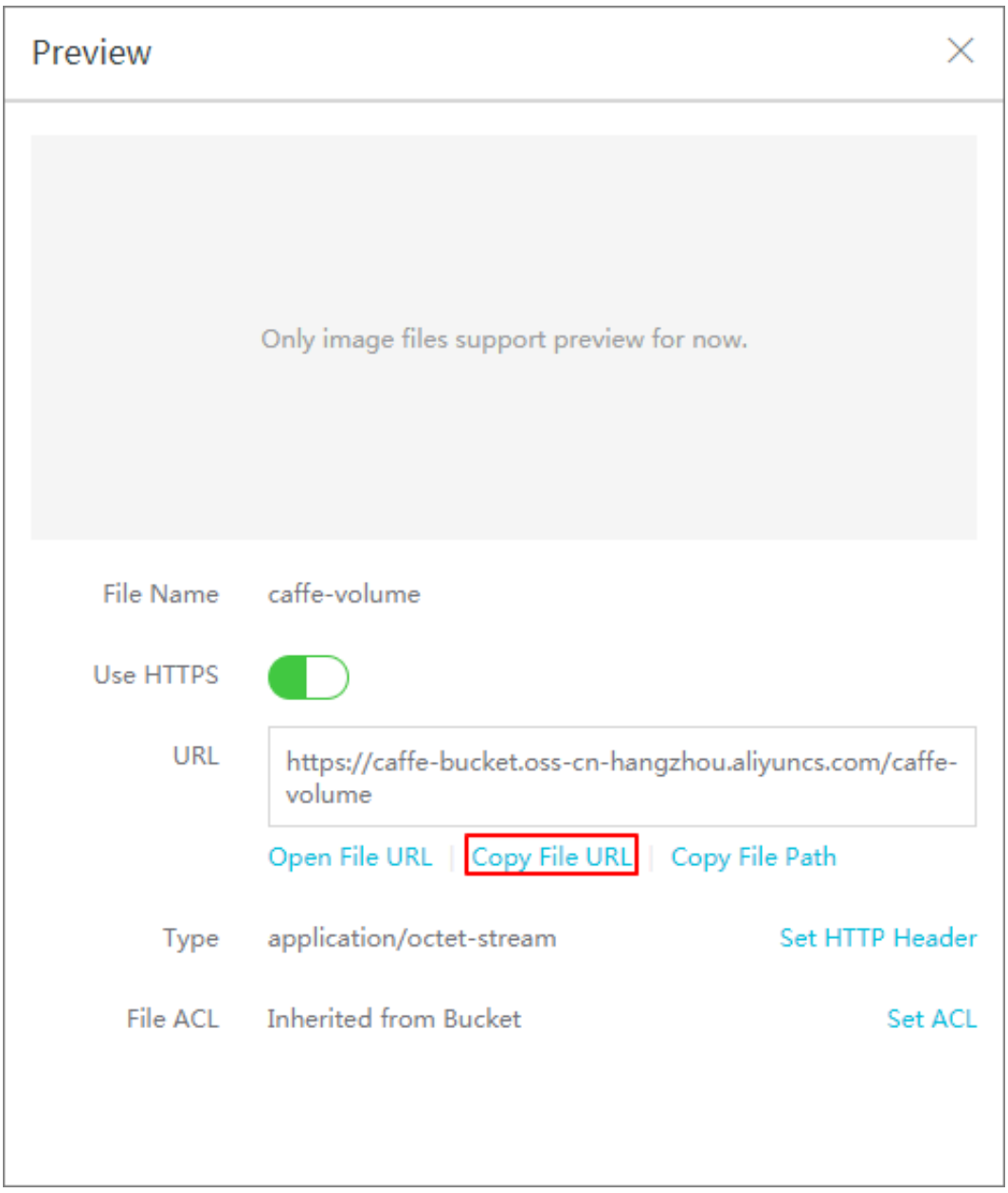
You can download the completed health check logs stored in OSS buckets.

1. Log on to the OSS console.
2. On the Overview page, click the target bucket and then click Files.
3. On the Files page, click `AliyunSLBH ealthCheck Logs /`.



4. Click the folder of the heath logs to download.

5. Click Edit of the target folder. Then, click Copy File URL in the displayed page.



6. Enter the copied URL in the web browser to download the logs.

6.3 Authorize a RAM user to use access logs

Before a RAM user starts to use the access log function, the RAM user must be authorized by the corresponding Alibaba Cloud account.

Prerequisites

The account has enabled the access log function.

1. Log on to the RAM console by using the credentials of your account.

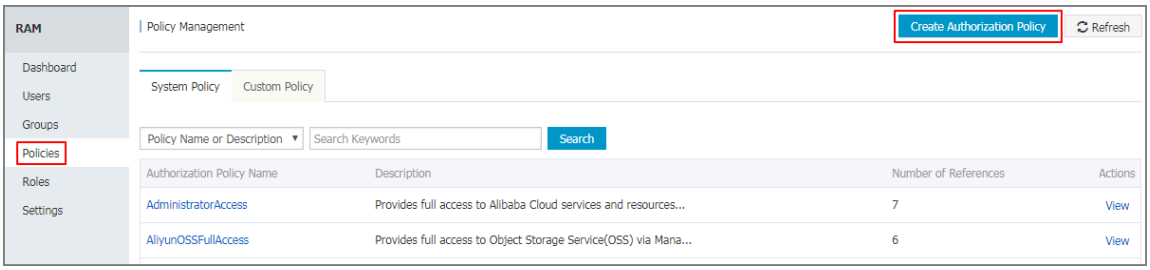
2. Click Roles to see whether the account has the AliyunLogArchiveRole.

   If the account does not have this role, log on to the SLB console by using the credentials of the account, select Logs > Access Logs, click Authorize. In the displayed dialog box, click Confirm Authorization Policy.

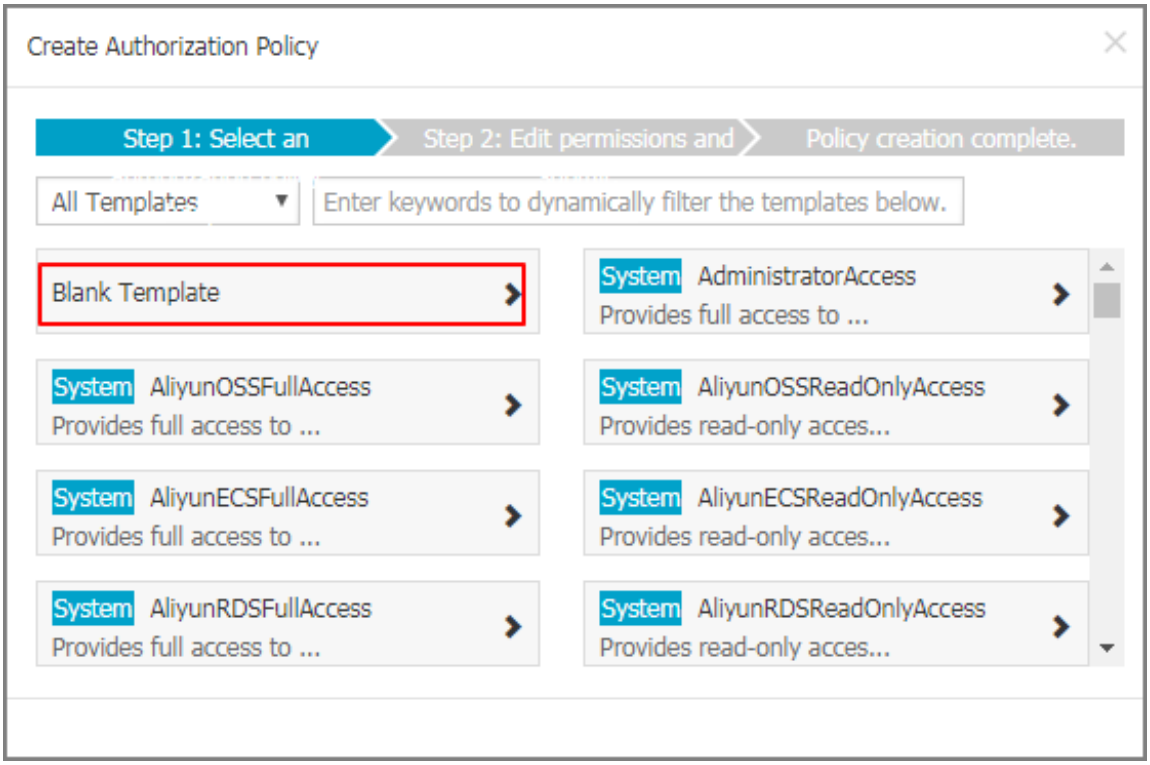
    Note:
   This operation is required only at the first time.

Procedure

1. Create an authorization policy:
   a) Log on to the RAM console by using the credentials of your account.
   b) In the left-side navigation pane, click Policies, and then click Create Authorization Policy.



   c) Click Blank Template.



   d) Enter a policy name, such as SlbAccessLogPolicySet, and then enter the following policy. Click Create Authorization Policy.

```
{
" Statement  ":[
 {
   " Action  ":[
     " slb : Create *",
     " slb : List *"
   ],
   " Effect ": " Allow ",
   " Resource ": " acs : log :*:*: project /*"
 },
 {
   " Action ": [
     " log : Create *",
```

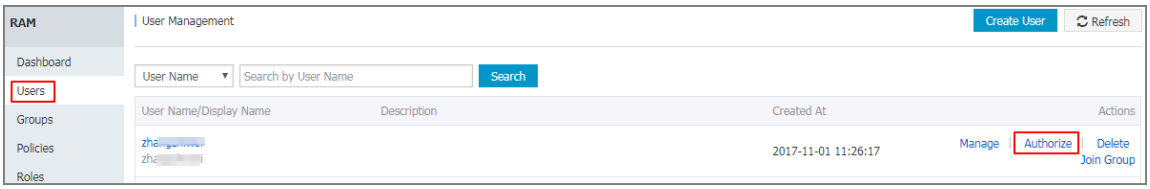
```
      " log : List *"
    ],
    " Effect ": " Allow ",
    " Resource ": " acs : log :*:*: project /*"
  },
  {
    " Action ": [
      " log : Create *",
      " log : List *",
      " log : Get *",
      " log : Update *"
    ],
    " Effect ": " Allow ",
    " Resource ": " acs : log :*:*: project /*/ logstore /*"
  },
  {
    " Action ": [
      " log : Create *",
      " log : List *",
      " log : Get *",
      " log : Update *"
    ],
    " Effect ": " Allow ",
    " Resource ": " acs : log :*:*: project /*/ dashboard /*"
  },
  {
    " Action ": " cms : QueryMetri  c *",
    " Resource ": "*",
    " Effect ": " Allow "
  },
  {
    " Action ": [
      " slb : Describe *",
      " slb : DeleteAcce  ssLogsDown  loadAttrib  ute ",
      " slb : SetAccessL  ogsDownloa  dAttribute ",
      " slb : DescribeAc  cessLogsDo  wnloadAttr  ibute "
    ],
    " Resource ": "*",
    " Effect ": " Allow "
  },
  {
    " Action ": [
      " ram : Get *",
      " ram : ListRoles "
    ],
    " Effect ": " Allow ",
    " Resource ": "*"
  }
],
" Version ": " 1 "
}
```

e) Click Close.

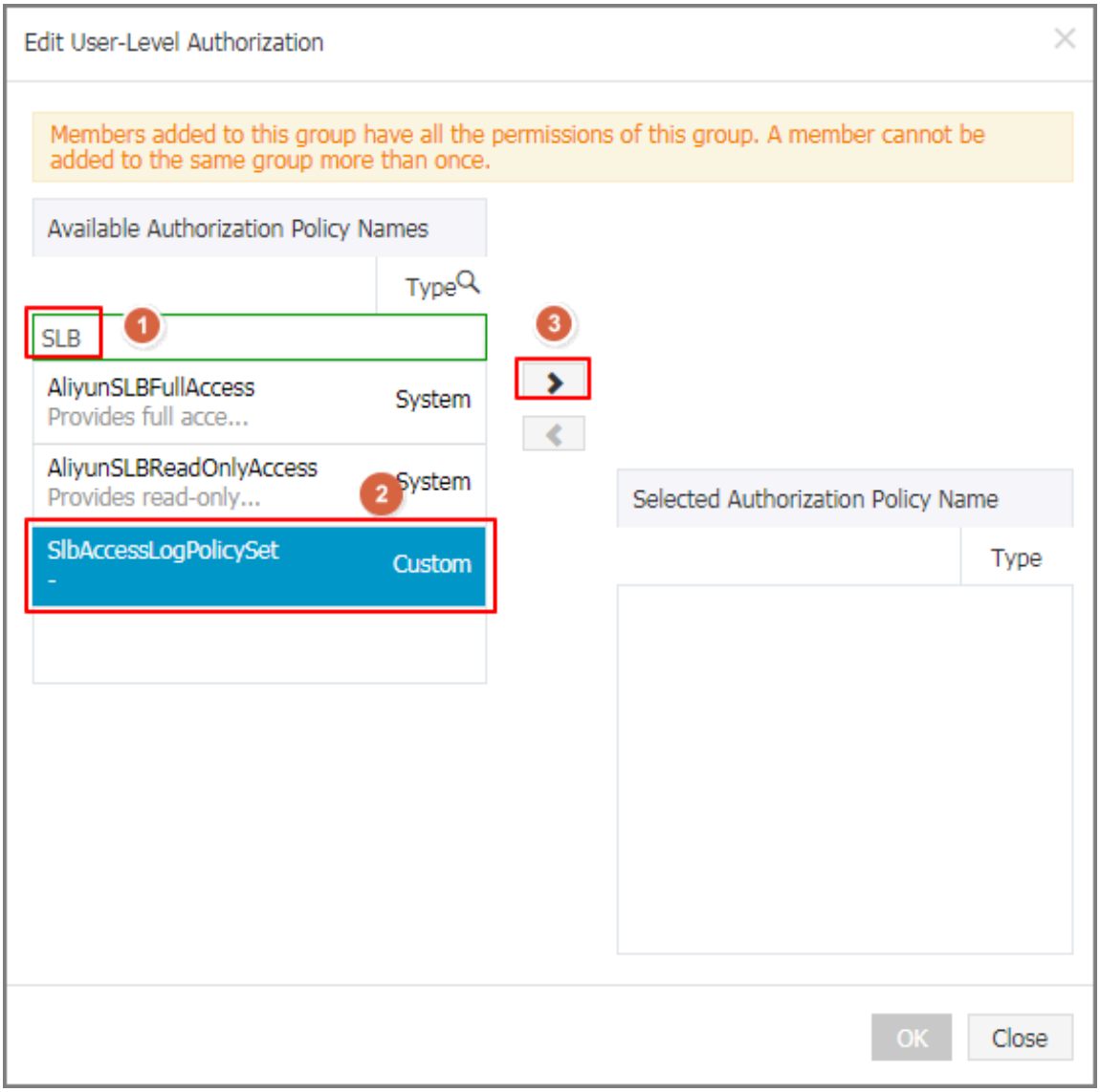
2. Attach the created policy to the RAM user:

   a) In the left-side navigation pane, click Users.

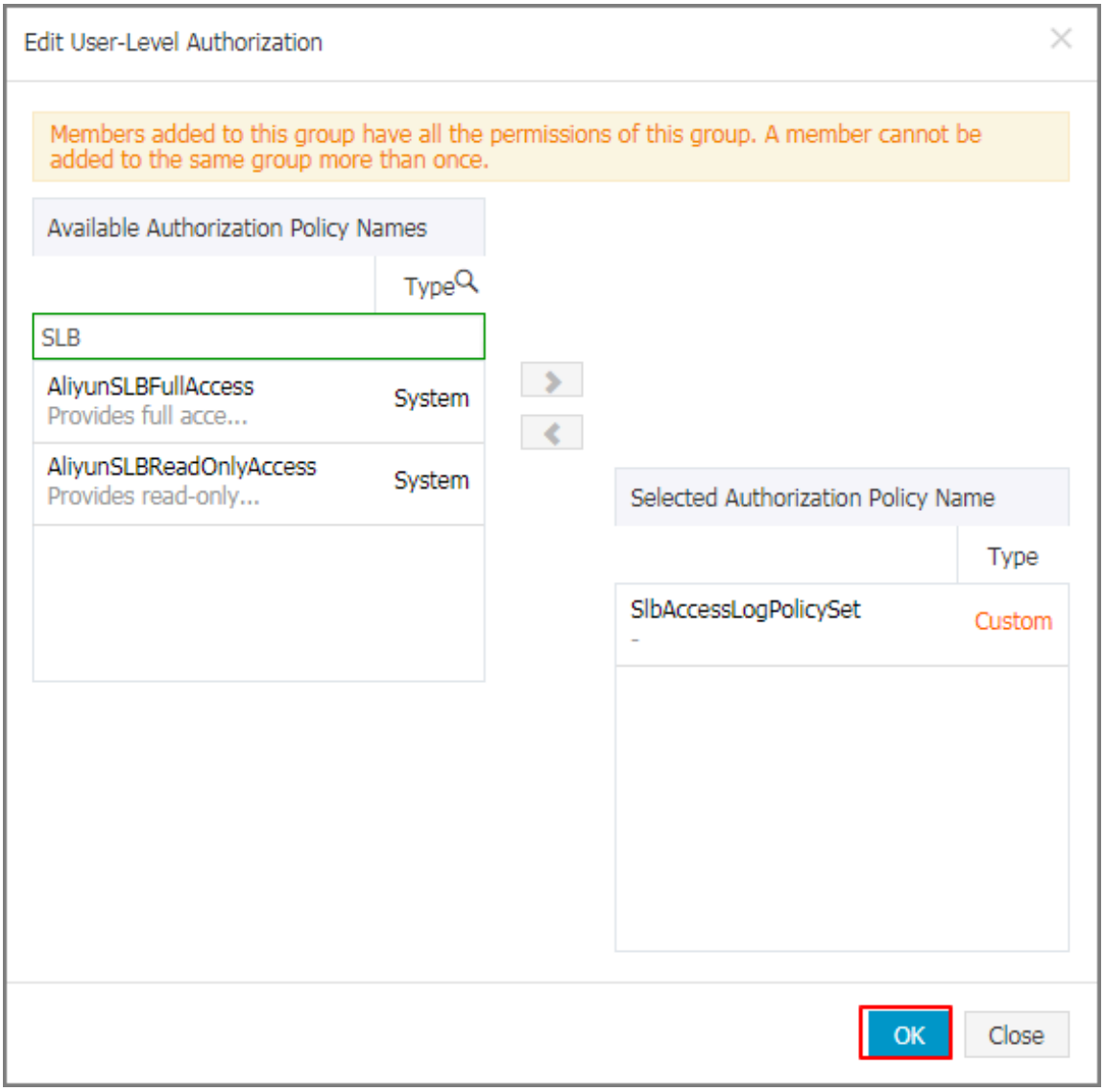
   b) Find the target RAM user (the user who uses the SLB access log function) and click Authorize.



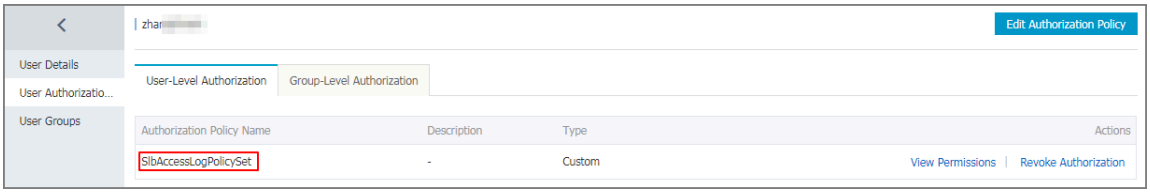
   c) Search the created authorization policy and attach the policy to the RAM user.



   d) Click OK.



e) Go back to the user details page to check whether the policy has been attached to the target RAM user.



6.4 Configure access logs

This topic describes how to configure access logs. By using Alibaba Cloud Log Service, you can analyze the access logs of a Server Load Balancer (SLB) instance

to understand the behavior and geographical distribution of client users and troubleshoot problems.

What are access logs?

Access logs collect detailed information of all requests sent to an SLB instance, including the request time, client IP address, latency, request URL, and server response. As the entry of Internet access, SLB receives massive client requests. You can use access logs to analyze user behavior and geographical distribution, and troubleshoot problems.

After you enable the SLB access log feature, you can store access logs in the Logstore of Log Service to collect and analyze the access logs. You can also disable the access log feature at any time.

SLB access logs can be used free of charge. You only need to pay for fees incurred by the use of Log Service.

 Note:

- Only Layer-7 SLB supports access logs and the access log function is available in all regions.
- Make sure that the HTTP header value does not contain `||`. Otherwise, the exported logs may be misplaced.

Benefits

The following are benefits of SLB access logs:

- Easy to use

  The access log function frees developers and maintenance staff from tedious and time-consuming log processing so that they can concentrate on business development and technical research.
- Cost-effective

  Access logs are typically massive. Processing access logs takes a lot of time and consumes a lot of resources. With Log Service, the processing of access logs is faster and cost-effective than self-built open-source solutions. Log Service can analyze one hundred million logs in one second.

- Real-time

  Scenarios such as DevOps, monitoring, and alerting require real-time log data . Traditional data storage and analysis tools cannot meet this requirement. For example, it takes a long time to ETL data to Hive where a lot of time is spent on data integration. Powered by its powerful big data computing capability, Log Service can process and analyze access logs in seconds.

- Flexible

  You can enable or disable the SLB access log feature according to the instance specification. Additionally, you can set the storage period (1 to 365 days) as needed and the Logstore's capacity is scaleable to meet increasing service demands.

Configure access logs

Before you configure access logs, make sure that:

- A Layer-7 listener is added.
- Log Service is activated.

To configure access logs, complete these steps:

1. Log on to the SLB console.
2. In the left-side navigation pane, choose Logs > Access Logs.
3. Select a region.
4. Click Authorize, and then click Confirm Authorization Policy to authorize SLB to write logs to Log Service.

   If you are a RAM user, you must obtain permissions from the corresponding account. For more information, see Authorize a RAM user to use access logs.

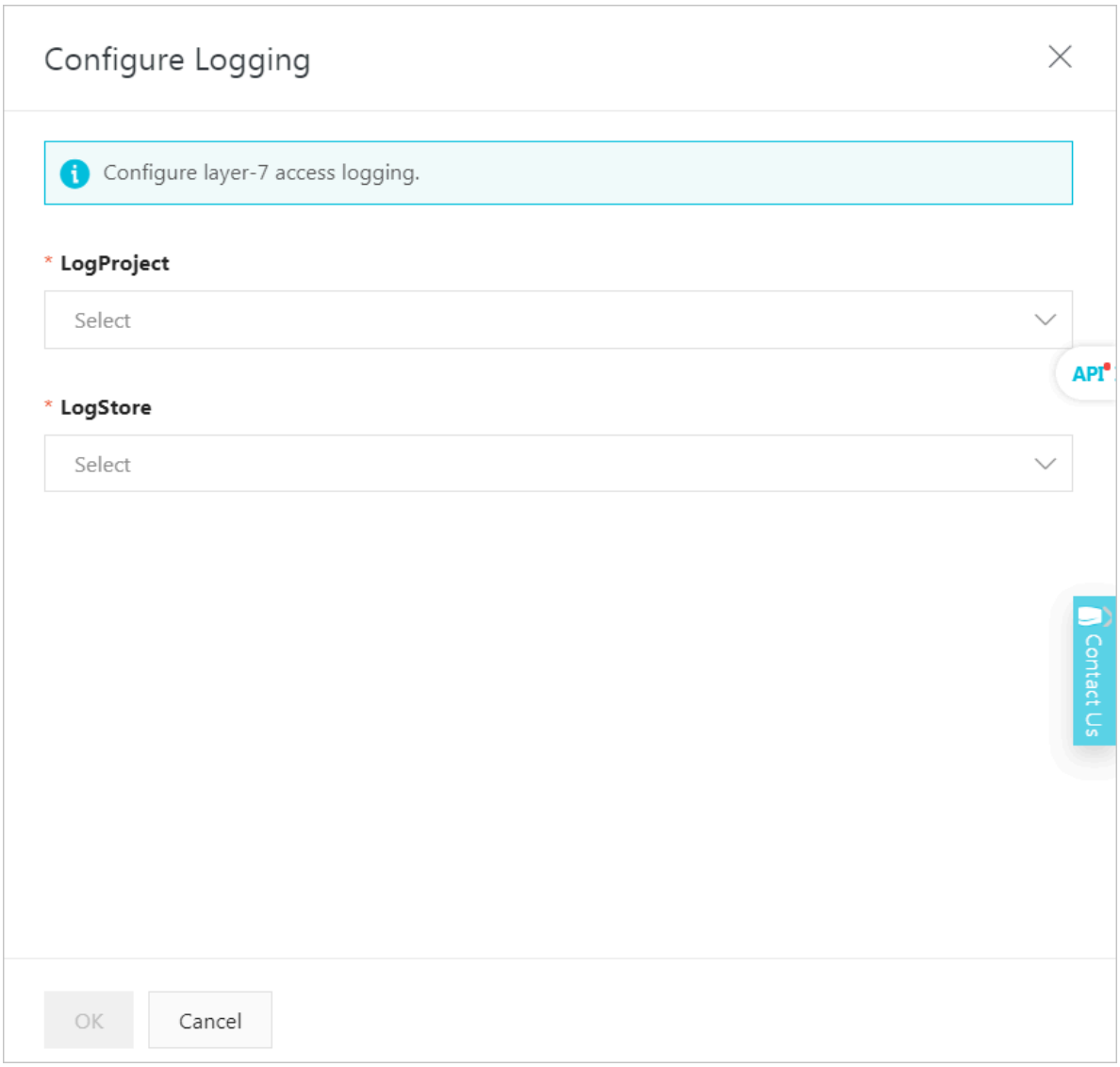
   > Note:
   > This step is required only at the first time.

5. On the Access Logs page, find the target SLB instance and click Configure Logging.
6. Select the LogProject and LogStore and then click OK.

   If there is no available LogStore, click Log Service console to create log projects.

   > Note:

Make sure that the name of the LogProject is globally unique and the region of the LogProject is the same as that of the SLB instance.



Search and analyze access logs

After configuring SLB access logs, you can search and view logs by using the following indexing fields.

Field	Description
body_bytes_sent	The size of HTTP body (in byte) sent to the client.
client_ip	The client IP address.
host	The host header in the request.
http_user_agent	The received http_user_agent header in the request.
request_length	The length of the request including startline, HTTP header, and HTTP body.

Field	Description
request_method	The request method.
request_time	The interval of time from when SLB receives the first request to the time when SLB returns a response.
request_uri	The URL of the received request.
Slbid	The ID of the SLB instance.
status	The status of the SLB response.
Upstream_addr	The IP address and port number of the backend server.
upstream_response_time	The interval of time from when SLB establishes a connection with the backend server to the time when SLB receives the last byte of the response.
upstream_status	The response status code of the backend server received by SLB.

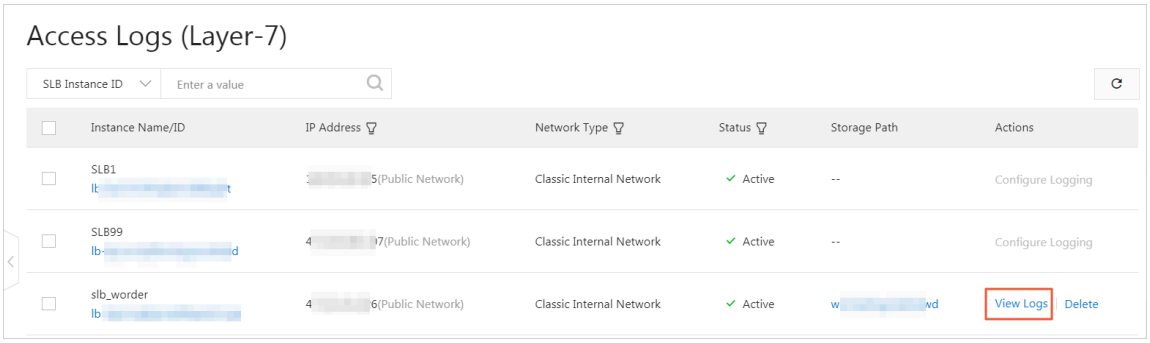
Search access logs

To search access logs, complete these steps:

1. Go to the log search page. You can navigate to the search page from the SLB console or the Log Service Console:

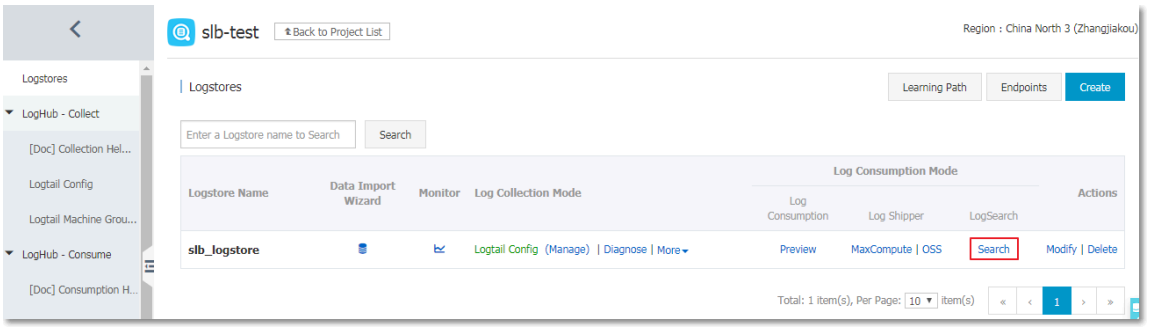
   - From the SLB console:

     On the Access Logs page, click View Logs.



   - From the Log Service Console:

     On the Logstores page, click Search of the target Logstore.

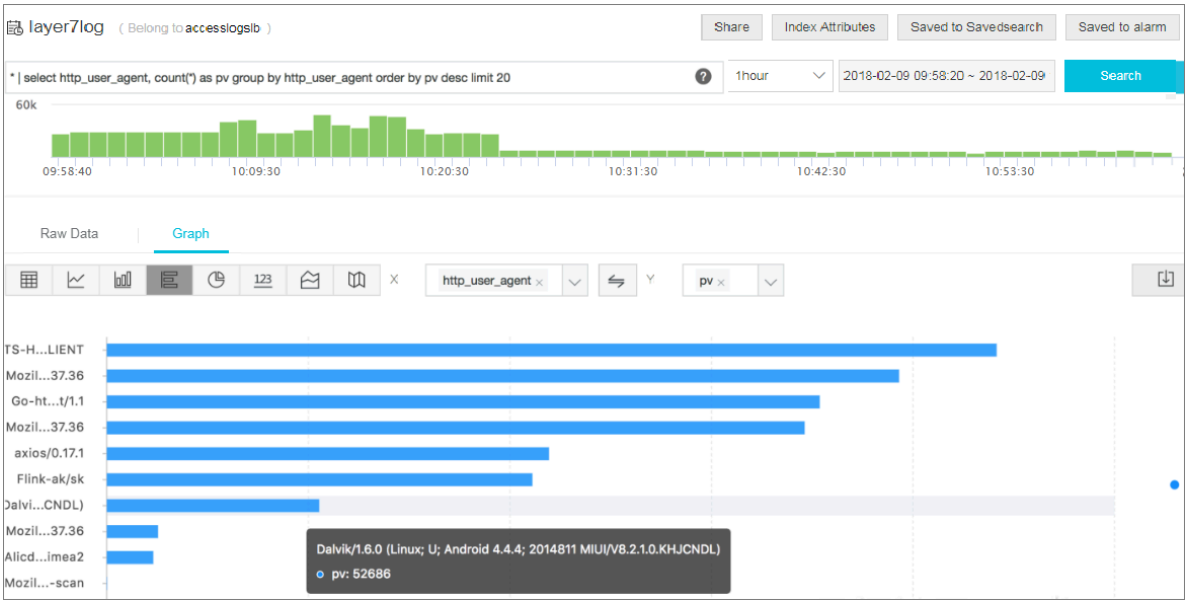


2. Click the target log field to view detailed information.
3. Enter an SQL statement to query access logs.

   For example, enter the following SQL statement to query the Top20 clients, which is used for analyzing the request source to assist business decision-making.

```
* | select ip_to_prov ince ( client_ip ) as client_ip_
 province , count (*) as pv group by
```

```
client_ip_ province order by pv desc limit 50
```

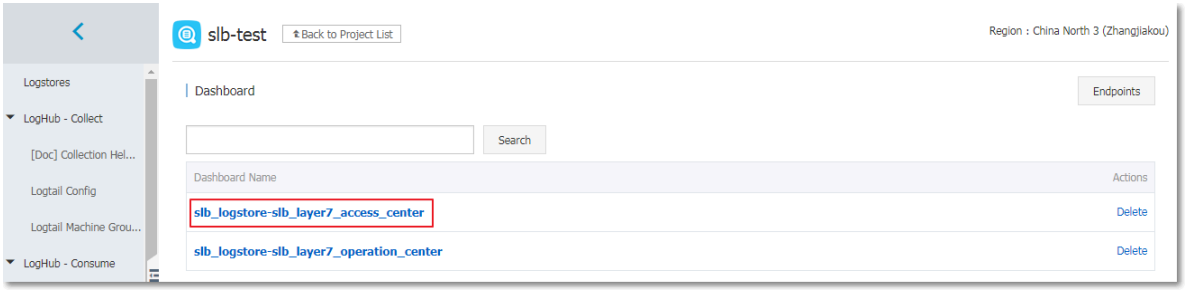


Analyze access logs

You can analyze access logs through the dashboard, which provides rich graphic information.

To analyze access logs, complete these steps:

1. In the Log Service console, click the project link of the SLB instance.
2. In the left-side navigation pane, choose LogSearch/Analytics - Query > Dashboard, and then click the name of the access log.

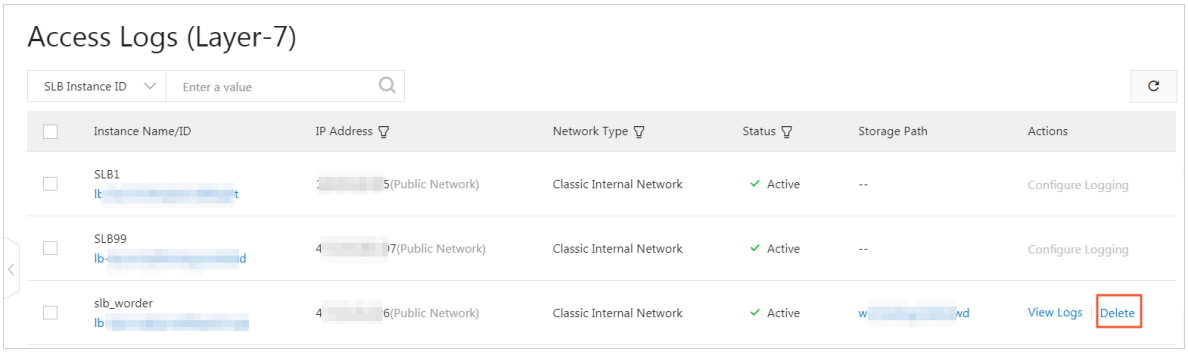


Disable the access log function

To disable the access log function, complete these steps:

1. Log on to the SLB console.
2. In the left-side navigation pane, choose Logs > Access Logs.
3. Select the region of the target SLB instance.

4. On the Access Logs page, find the target instance and click Delete.



5. In the displayed dialog box, click OK.

7 Access control

7.1 Configure an access control list

Server Load Balancer (SLB) provides you with an access control function. You can configure different access control rules (whitelist or blacklist) for different listeners. Before configuring the access control function for a listener, you must first configure an access control list.

You can create multiple access control lists. Each list contains multiple IP addresses or CIDR blocks. Limits on access control lists are shown in the following table.

Resource	Limit
The maximum number of access control lists per region	50
The maximum number of IP entries added each time	50
The maximum number of IP entries per access control list	300
The maximum number of listeners with which an access control list can be associated	50

Create an access control list

To create an access control list, follow these steps:

1. Log on to the SLB console.
2. Select a region.
3. In the left-side navigation pane, click Access Control.
4. Click Create Access Control List, enter an access control list name, select the IP version, and select the resource group.
5. Click OK.

Add IP entries

To add IP entries to the access control list, follow these steps:

1. Log on to the SLB console.

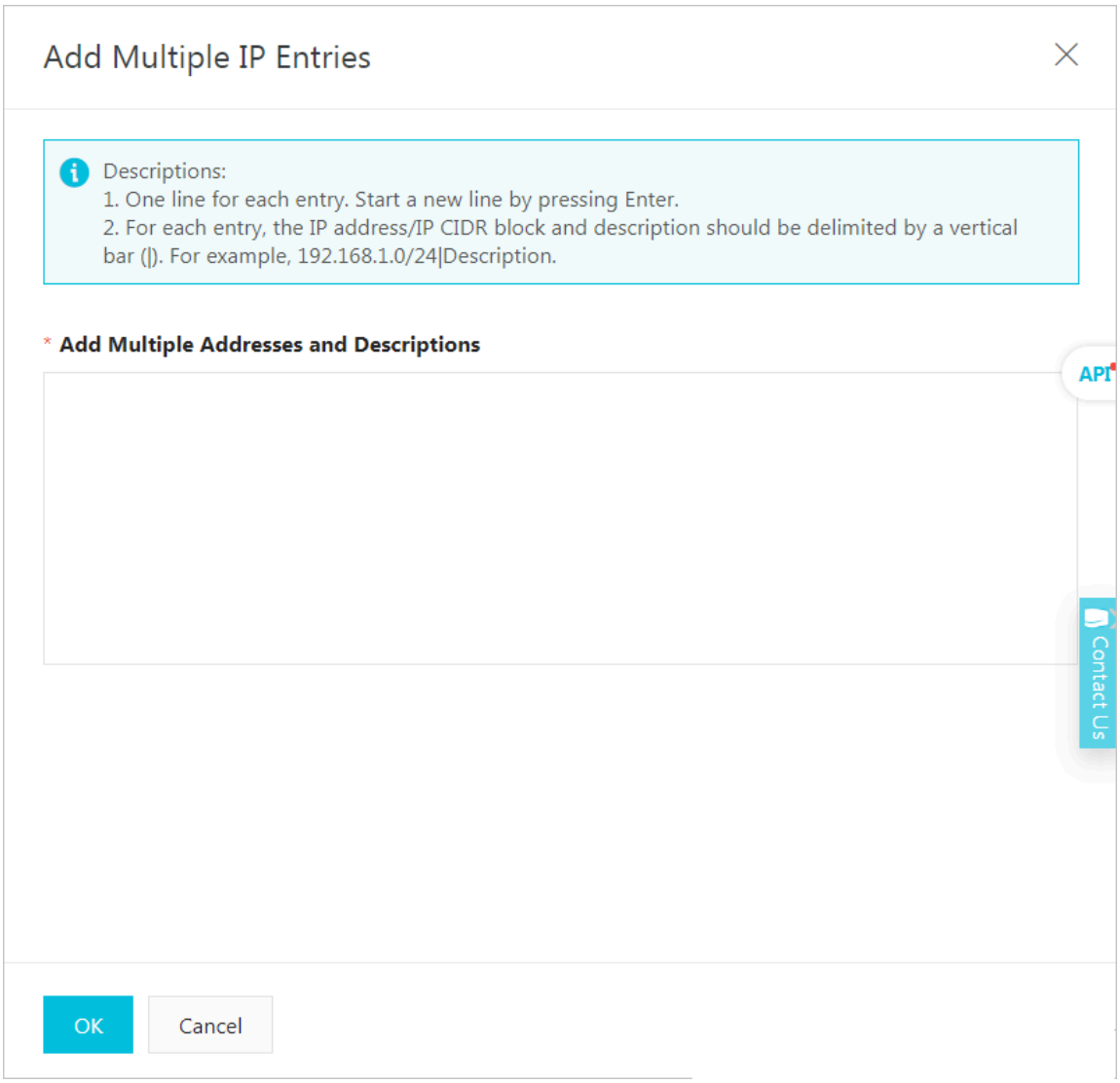
2. Select a region.
3. In the left-side navigation pane, click Access Control.
4. Find the target access control list and click Manage.

5. Add IP entries:

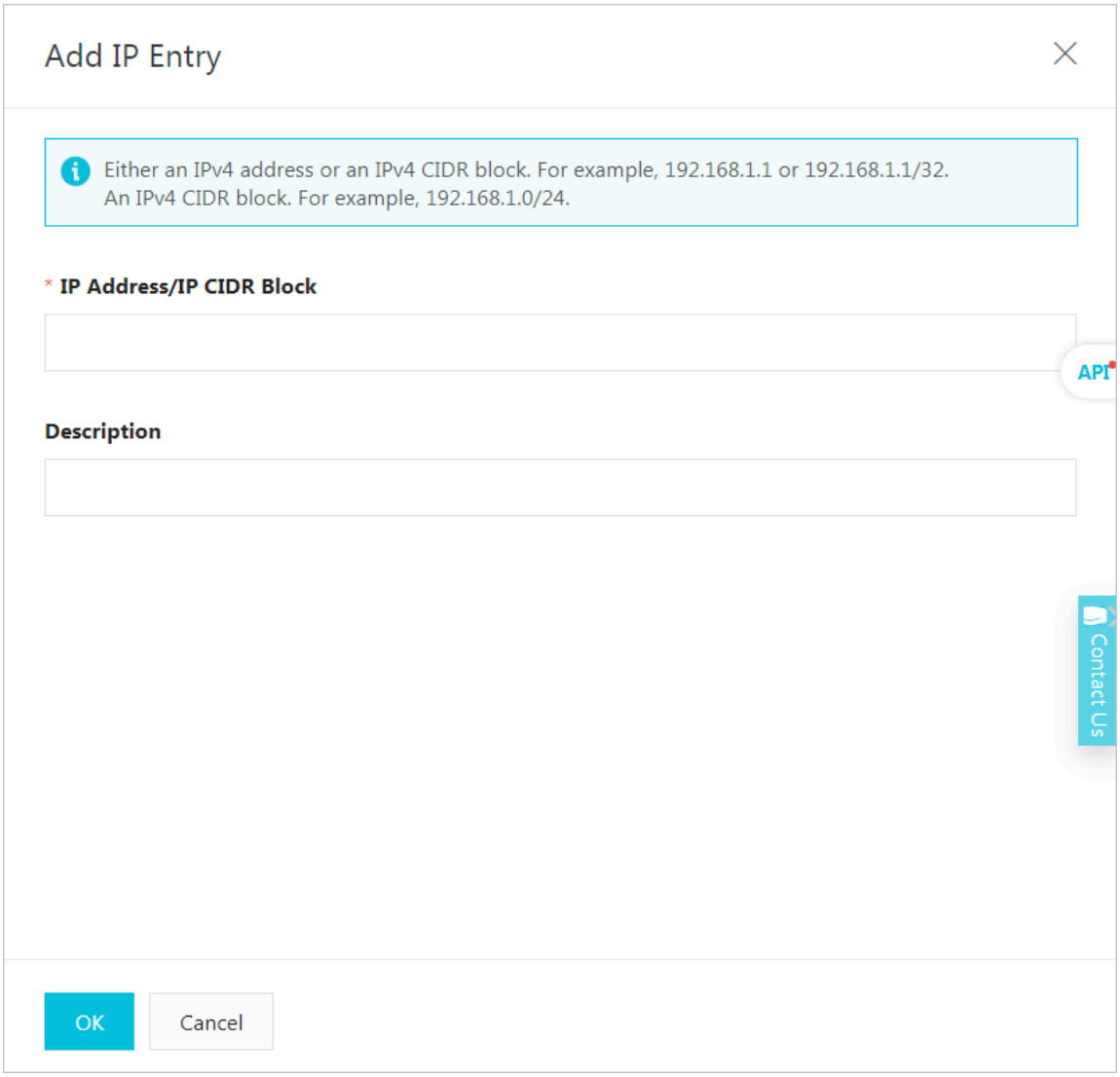
   · Click Add Multiple Entries. In the displayed dialog box, add IP addresses or CIDR blocks and click OK.

     Note the following when you add IP entries:

     - Each line should include only one IP entry. Use the Enter key to break lines.
     - Use a vertical bar (|) to separate an IP address or a CIDR block with the description, for example, 192.168.1.0/24|description.



   · Click Add Entry. In the displayed dialog box, add an IP address or a CIDR block and the description, and click OK.



Delete IP entries

To delete IP entries, follow these steps:

1. Log on to the SLB console.
2. Select a region.
3. In the left-side navigation pane, click Access Control.
4. Find the target access control list and click Manage.
5. Click Delete in the Actions column of the target IP entry, or select multiple IP entries and click Delete at the bottom of the list.
6. In the displayed dialog box, click OK.

7.2 Configure access control

Server Load Balancer (SLB) provides an access control function for listeners. You can configure different whitelists or blacklists for different listeners.

You can configure access control when you create a listener or change access control configurations after a listener is created.

This topic describes how to configure access control after a listener is created.

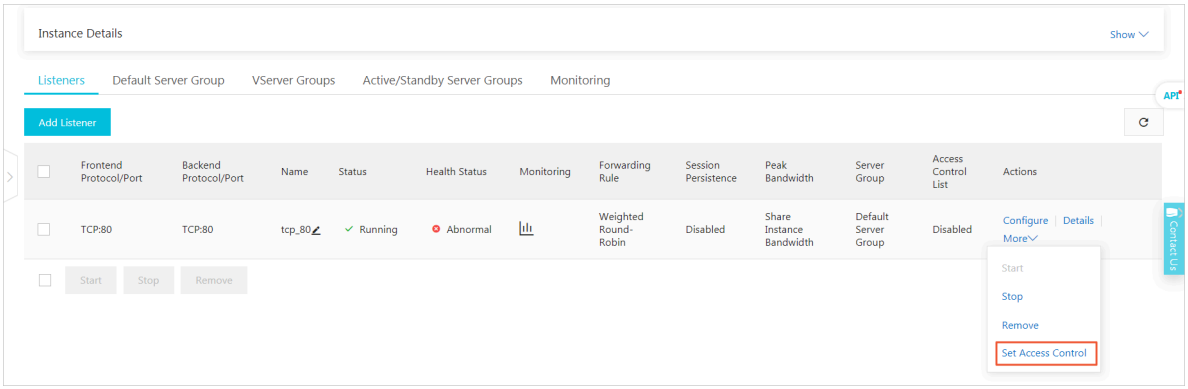
Enable access control

Before you enable access control, make sure:

- An access control list is created. For more information, see Configure an access control list.
- A listener is created.

To enable access control, follow these steps:

1. Log on to the SLB console.
2. Select the region of the target SLB instance.
3. Locate the target SLB instance and click the instance ID.
4. On the Instance Details page, click the Listeners tab.
5. Locate the target listener, and then choose More > Set Access Control.



6. On the Access Control Settings page, enable access control, select an access control method and an access control list, and click OK.
   - Whitelist: Only requests from IP addresses or CIDR blocks in the selected access control list are forwarded. It applies to scenarios where the application only allows access from some specific IP addresses.

     Enabling a whitelist poses some business risks. After a whitelist is configured , only the IP addresses in the list can access the listener. If you enable the whitelist without adding any IP addresses in the corresponding access control list, no requests are forwarded.
   - Blacklist: Requests from IP addresses or CIDR blocks in the selected access control list are not forwarded. It applies to scenarios where the application only denies access from some specific IP addresses.

     If you enable a blacklist without adding any IP addresses in the corresponding access control list, all requests are forwarded.

Disable access control

To disable access control, follow these steps:

1. Log on to the SLB console.
2. Select the region of the target SLB instance.
3. Locate the target SLB instance and click the instance ID.
4. On the Instance Details page, click the Listeners tab.
5. Locate the target listener, and then click More > Set Access Control.
6. On the Access Control Settings page, disable access control and click OK.

7.3 Migrate to the new access control

If you have already configured a whitelist for a listener, Server Load Balancer can automatically add the IP addresses or CIDR blocks in the whitelist to an access control list and apply the list to the listener.

Migrate a whitelist to an access control list

To migrate a previously configured whitelist to an access control list, complete these steps:

1. Log on to the SLB console.

2. Select the region where the SLB instance is located, and then click the ID of the target SLB instance.
3. Click the Listeners tab.
4. Find the target listener, select More > Set Access Control.
5. Click Use New Access Control Features.
6. Enter a name of the access control list and click Create Access Control List.
7. Click Apply to apply the list to the listener as a whitelist.

> Note:
> If you do not apply the list to a listener, the whitelist does not take effect.

View the migrated access control list

To view the migrated access control list, complete these steps:

1. Log on to the SLB console.
2. Select a region.
3. In the left-side navigation pane, click Access Control.
4. Find the created access control list and view the associated listener. You can click Manage to manage IP entries.

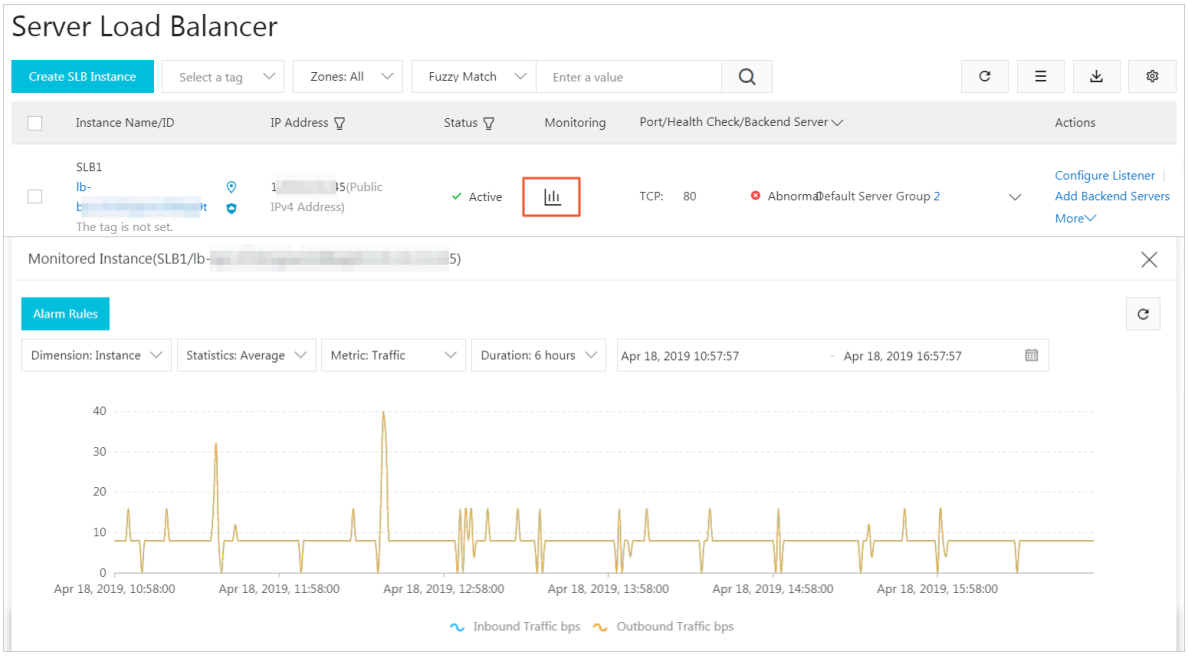
8 Monitoring

8.1 View monitoring data

This topic describes how to view the monitoring data of a Server Load Balancer (SLB) instance.

Procedure

1. Log on to the SLB console.
2. Select the region of the target SLB instance.
3. Click the monitoring icon  next to the target SLB instance.
4. Select the monitoring metrics that you want to view.



The following metrics are monitored for SLB instances.

Metric	Description
Traffic	· Inbound Traffic: the traffic sent from an external network to SLB · Outbound Traffic: the traffic sent from SLB

Metric	Description
Packets	· RX Packets Count: the number of request packets received per second · TX Packets Count: the number of response packets sent per second
Concurrent Connections	· Active Connections Count: the number of established TCP connections. If persistent connections are used, a connection can transfer multiple file requests at one time. · Inactive Connections Count: the number of TCP connections not in the established state. You can use the `netstat -an` command to view the connections for both Windows and Linux servers. · Max Concurrent Connections Count: the total number of TCP connections.
Average Connection Requests Count	The average number of new TCP connections established between clients and SLB in a statistical period
Dropped Traffic	· Dropped Inbound Traffic: the amount of inbound traffic dropped per second · Dropped Outbound Traffic: the amount of outbound traffic dropped per second
Dropped Packets	· Dropped RX Packets: the number of inbound packets dropped per second · Dropped TX Packets: the number of outbound packets dropped per second
Dropped Connections Count	The number of TCP connections dropped per second
The following metrics are specific to Layer-7 listeners.	
Layer-7 Protocol QPS	The number of HTTP/HTTPS requests that can be handled per second
Response Time (Listener)	The average response time of SLB
HTTP Status Code 2xx/3xx/4xx/5xx/Others (Listener)	The average number of HTTP response codes generated by listeners
Response Code 4xx/5xx (Server)	The average number of HTTP response codes generated by backend servers

Metric	Description
Response Time (Server)	The average response time of backend servers

8.2 Configure alarm rules

After activating the CloudMonitor service, you can configure alarm rules for SLB instances on the CloudMonitor console.

Context

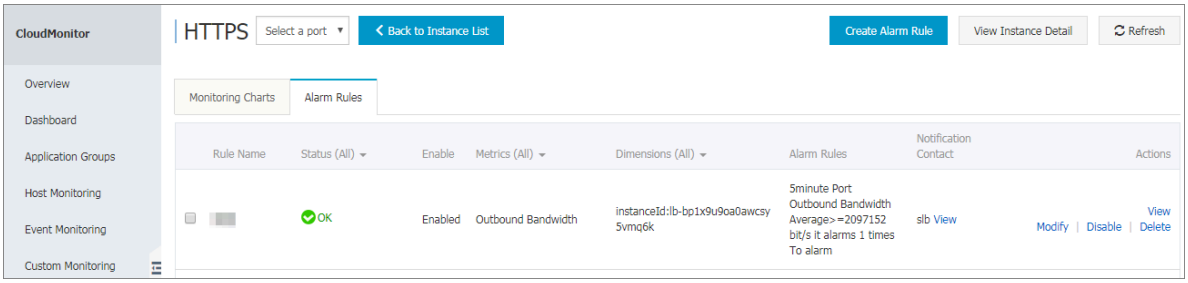
> Note:
> If a listener or an SLB instance is deleted, its alarm settings are deleted correspondingly.

Procedure

1. Log on to the SLB console.
2. Select the region where the SLB instance is located.
3. Find the target instance and click  .

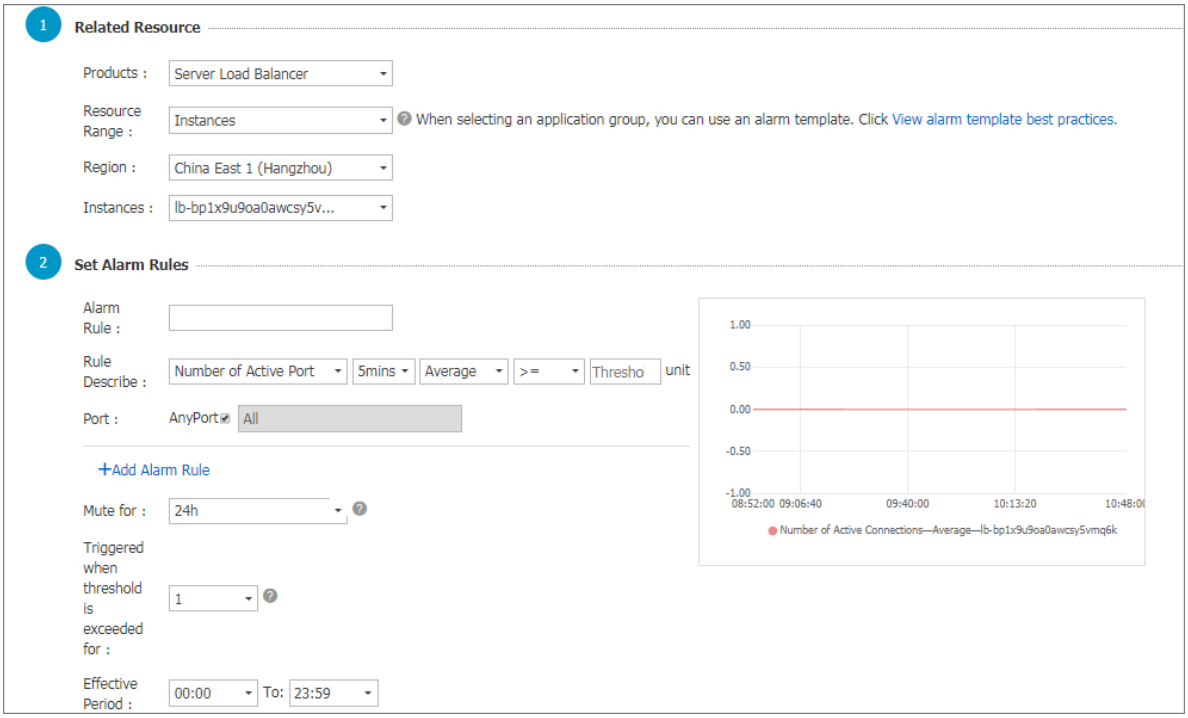
> Notice:
> Make sure the instance has configured with listeners and enabled health check.

4. Click Alarm Rules. You are then directed to CloudMonitor console.



5. Click Create Alarm Rule.

6. Configure the alarm rule.



9 API Inspector

API Inspector is an experimental feature. With API Inspector, you can view the API calls behind each operation in the console, and automatically generate API code of different languages. You can debug online through Cloud Shell or OpenAPI Explorer.

Features

API Inspector, OpenAPI Explorer, and Cloud Shell form an integrated solution for you to learn and debug APIs. API Inspector has the following features:

- Automatic recording: To obtain related API calls, you only need to perform operations in the console. For more information, see Automatically record API calls.
- Code generating with one click: API code scripts in different languages with pre-filled parameters are generated and can be run directly. For more information, see Generate API codes with one click.
- Online debugging: When API Inspector is used together with OpenAPI Explorer and Cloud Shell, one-click online debugging can be implemented and you do not need to build the development environment. What you see is what you get. For more information, see Debug online through OpenAPI Explorer and Debug online through Cloud Shell.

Enable API Inspector

To enable API Inspector, follow these steps:

1. Log on to the SLB console.
2. In the lower-right corner of the page, click 

.

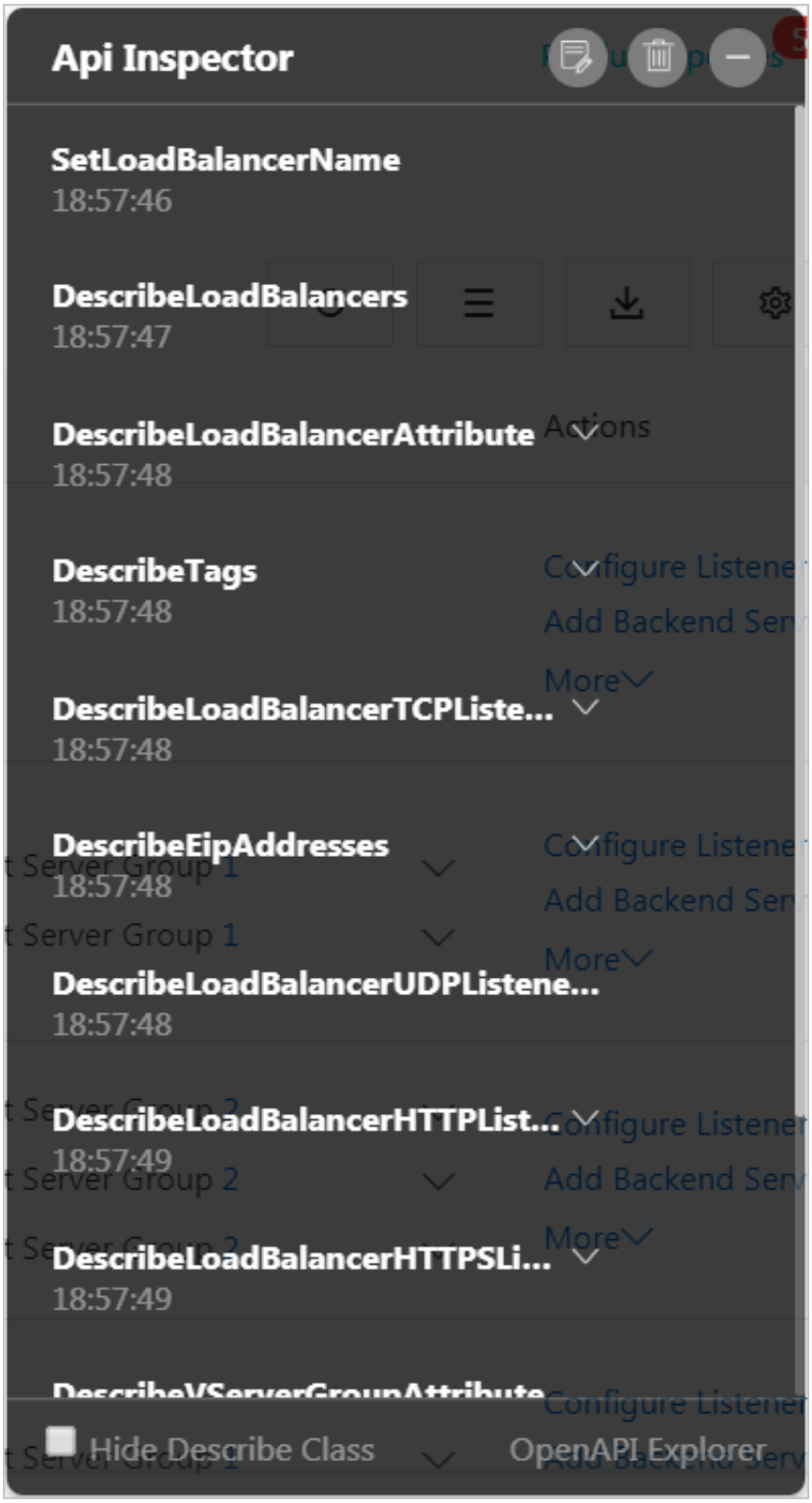
Automatically record API calls

In this topic, modifying the name of an SLB instance is taken as an example to demonstrate the automatic recording function of API Inspector.

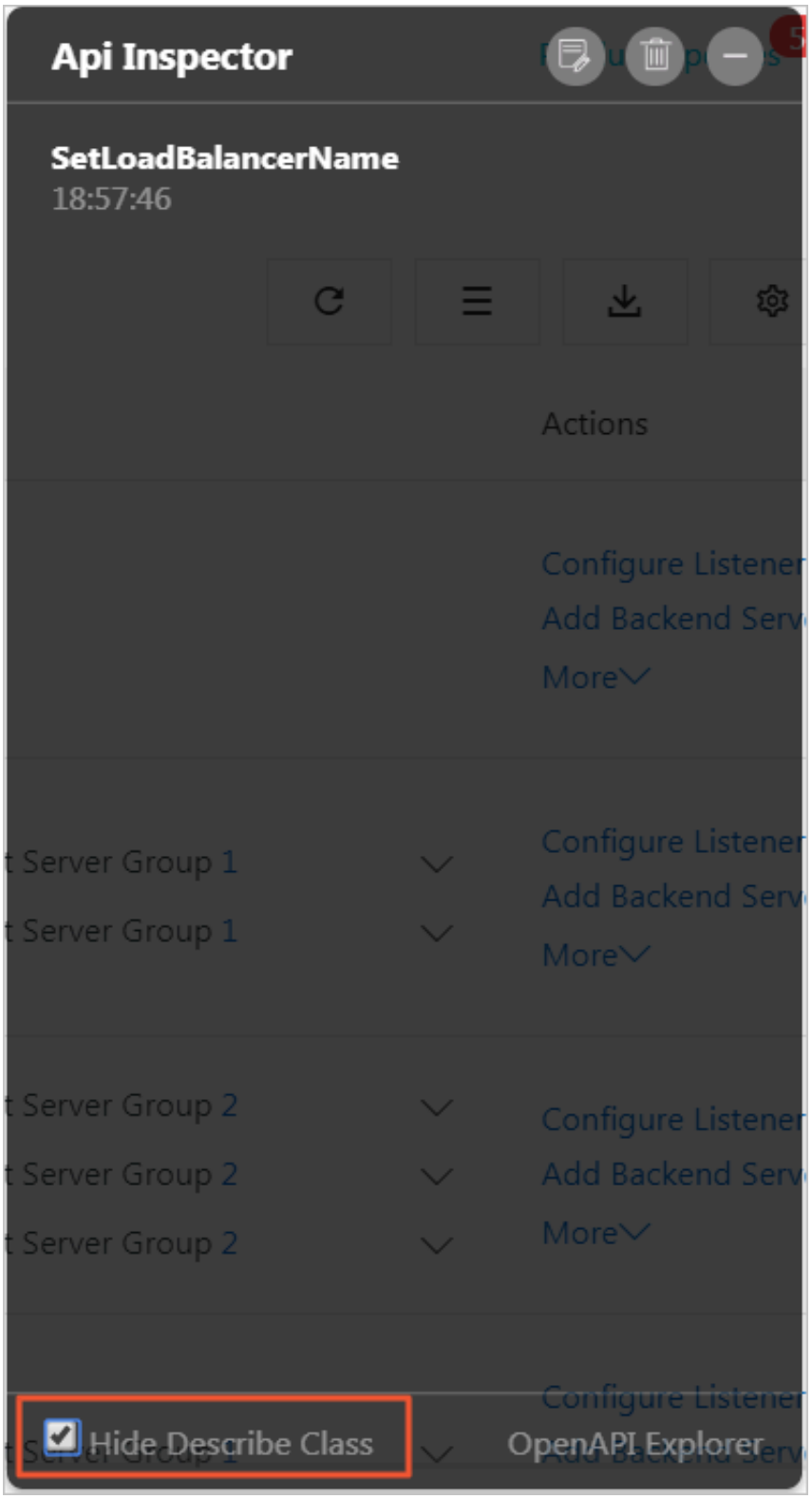
1. Choose Instances > Server Load Balancer.
2. Modify the name of an SLB instance to SLB1.
3. Click OK.

4. Click 

on the right side of the page. Then you can see all API calls related to the preceding operation.



5. You can click Hide Describe Class to view core APIs. In this example, the core API is `SetLoadBalancerName`.

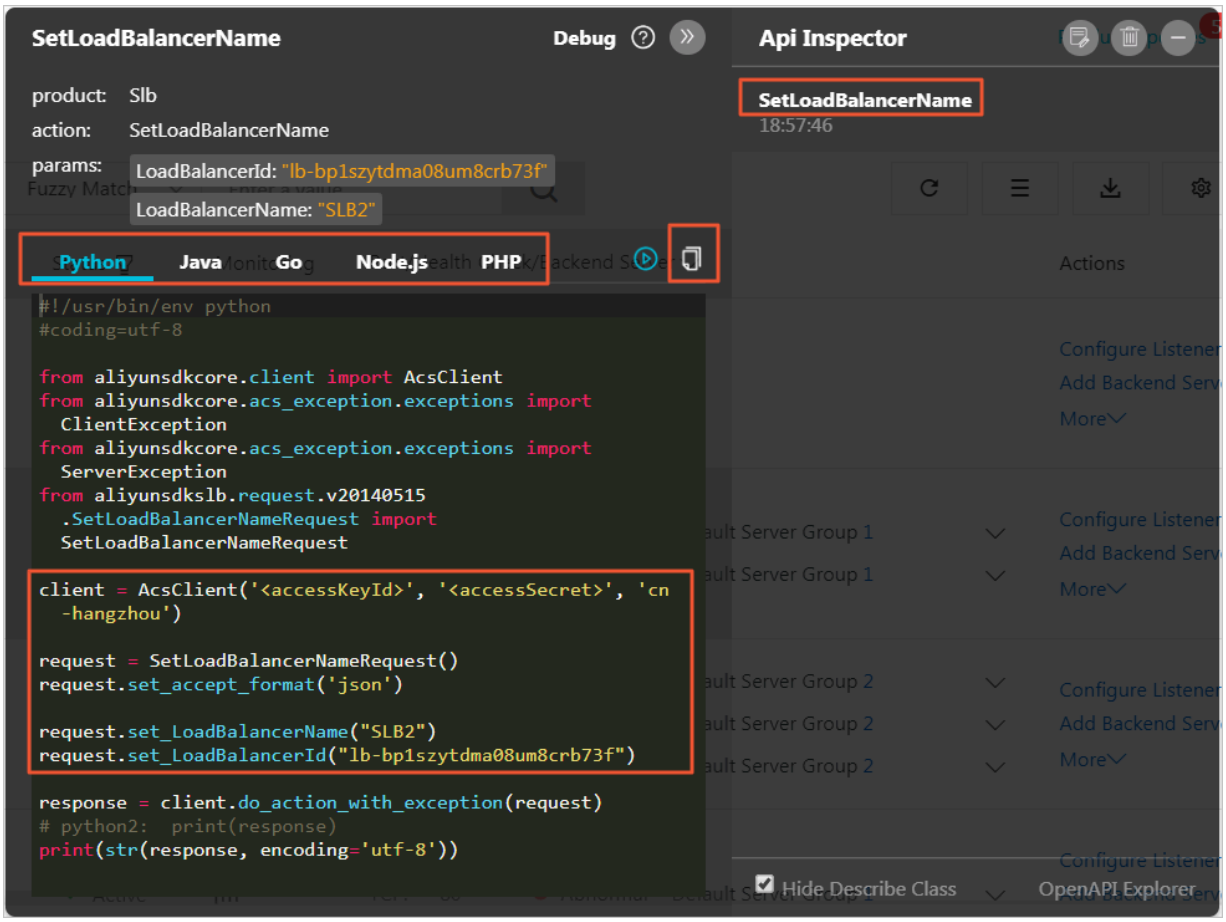


Generate API codes with one click

After API recording is completed, click the API name to generate API code scripts in Python, Java, Go, Node.js, and PHP, with pre-filled parameters.

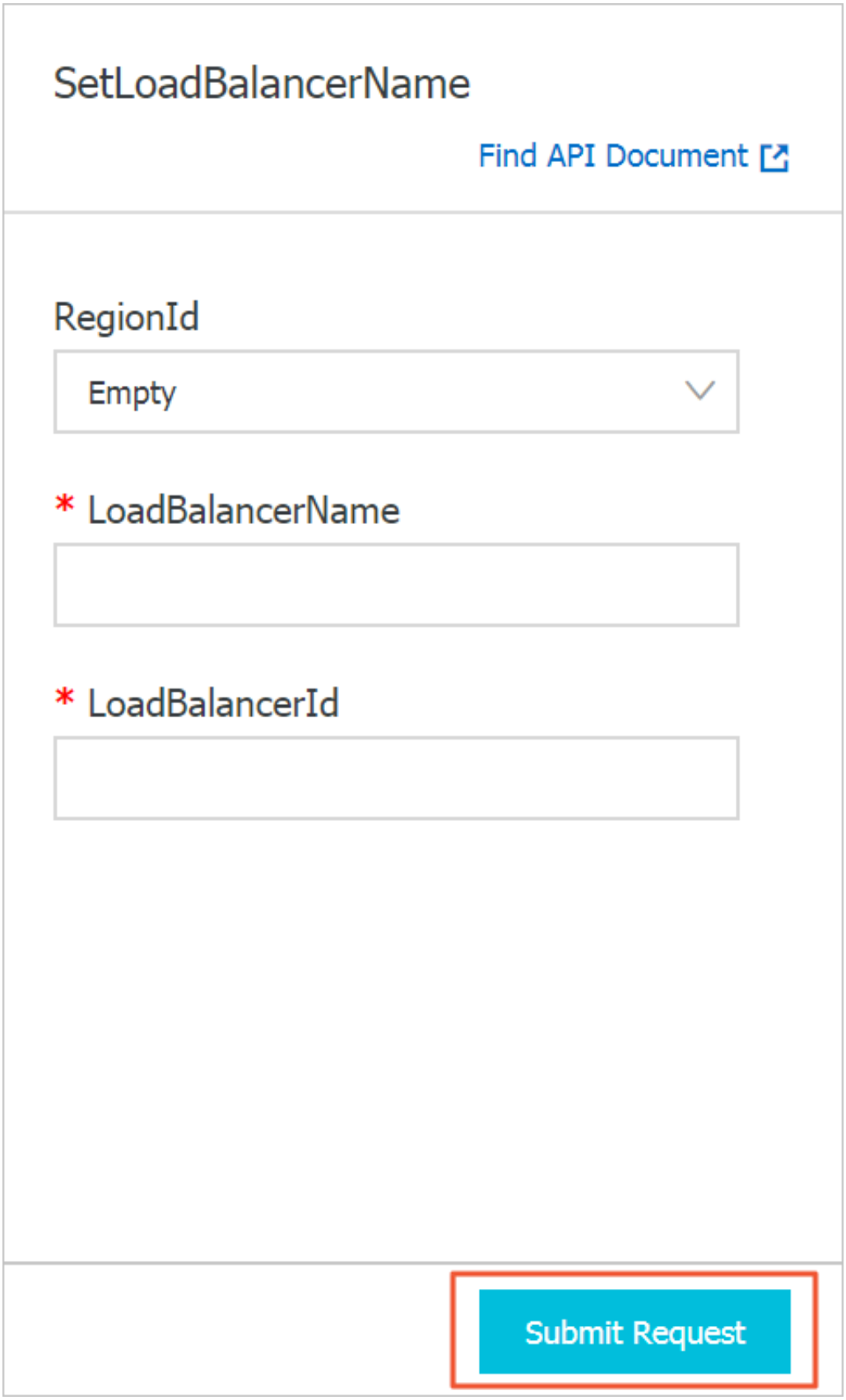
Note:

Click to copy the code scripts of the corresponding format, which can be run directly.



Debug online through OpenAPI Explorer

After the API recording is completed, click OpenAPI Explorer or to go to the OpenAPI Explorer console to debug the corresponding function. The API parameter values have been automatically generated according to operations in the console.


SetLoadBalancerName
Find API Document
RegionId
Empty
* LoadBalancerName
* LoadBalancerId
Submit Request

Note:

Click  to view the document describing parameter details of the called API.

Debug online through Cloud Shell

After API recording, unfold the API calling details and click  to use the online one-click debugging function of Cloud Shell.

 **Note:**

If you use the one-click debugging function of Cloud Shell, we recommend that you create and associate an OSS bucket to store your frequently used scripts and files. However, some OSS fees will be generated. You can also choose not to create an OSS bucket.

The command format for the Cloud Shell debugging of SLB is as follows:

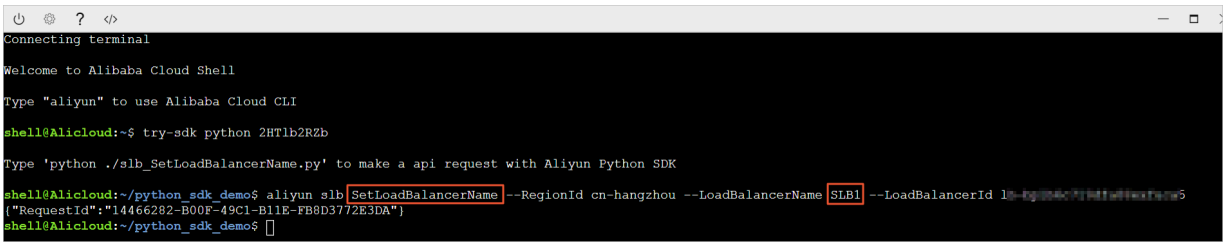
```
aliyun slb actionName --parameter1value1 --paramter2value2...
```

In this example, the called `SetLoadBalancerName` API modifies the name of the SLB instance to SLB1. The corresponding command is:

```
aliyun slb SetLoadBal ancerName -- RegionId cn - hangzhou
-- LoadBalanc erName SLB1 -- LoadBalanc erId lb - bp1b6c719d
fa08exfuca 5
```

The returned value is:

```
{" RequestId ":" 14466282 - B00F - 49C1 - B11E - FB8D3772E3 DA "}
```



10 Multiple-zone deployment

You can create SLB instances in a region with multiple zones to improve the service availability.

What is multiple-zone deployment?

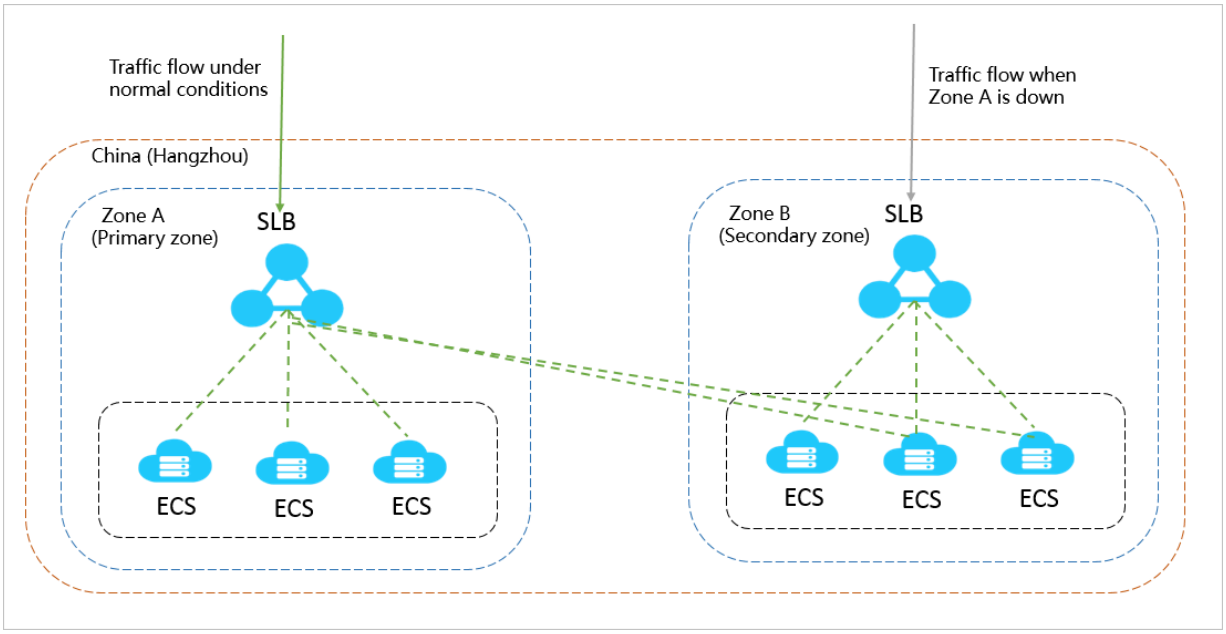
A cloud product zone is a set of independent infrastructures. Different zones have independent infrastructures (such as network, power supply, and air-conditioning). Therefore, infrastructure faults in one zone does not affect other zones.

To provide more reliable services, SLB has deployed multiple zones in most regions to achieve disaster recovery across data centers. When the data center in the primary zone is faulty and unavailable, SLB is able to switch to the data center in the secondary zone to restore its service within 30 seconds. When the primary zone becomes available again, SLB will switch back to the primary zone.

Note the following about SLB primary/secondary zones:

- SLB supports ECS instances in different zones. However, the ECS instances and the SLB instance must belong to the same region. SLB can distribute traffic to the ECS instances in different zones.
- Normally, the SLB instance in the secondary zone is in the standby state. You cannot manually switch to the secondary zone. SLB switches to the secondary zone only when the primary zone is unavailable due to reasons such as data center power outage and exit cable failures. SLB does not switch to the secondary zone when an SLB instance in the primary zone is faulty.
- SLB and ECS are deployed in different clusters. When an SLB instance in Zone A is unavailable, the ECS instances in Zone A are not necessarily unavailable. Therefore, after SLB switches to the secondary zone due to SLB cluster faults, the SLB instance in the secondary zone still can distribute traffic to the ECS instances in different zones. However, if power outrage or optical cable failures occur to all clusters in a zone, all services (including but not limited to SLB instances and ECS instances) in the zone cannot work anymore.

For more information, see SLB high availability.



Primary/secondary zone list

The following table lists the primary/secondary zones in different regions. You can call the `DescribeZones` API to query available primary/secondary zones in a region.

Region	Zone type	Zones	
China (Hangzhou)	Multi-zone	Primary zone	Secondary zone
		Zone B	Zone D Zone G
		Zone D	Zone E
		Zone E	Zone D Zone F
		Zone F	Zone E
		Zone G	Zone B Zone H
		Zone H	Zone G

Region	Zone type	Zones	
China (Shanghai)	Multi-zone	Primary zone	Secondary zone
		Zone A	Zone B
		Zone B	Zone A Zone C Zone D
		Zone C	Zone B
		Zone D	Zone B Zone E
		Zone E	Zone D Zone F
		Zone F	Zone E
China (Shenzhen)	Multi-zone	Primary zone	Secondary zone
		Zone A	Zone B
		Zone B	Zone A Zone C
		Zone C	Zone B Zone D
		Zone D	Zone C Zone E
		Zone E	Zone D
China (Qingdao)	Multi-zone	Primary zone	Secondary zone
		Zone B	Zone C
		Zone C	Zone B

Region	Zone type	Zones	
China (Beijing)	Multi-zone	Primary zone	Secondary zone
		Zone A	Zone B Zone D Zone E
		Zone B	Zone C
		Zone C	Zone E
		Zone D	Zone A
		Zone E	Zone C Zone F
		Zone F	Zone E Zone G
		Zone G	Zone F
China (Zhangjiakou)	Multi-zone	Primary zone	Secondary zone
		Zone A	Zone B
		Zone B	Zone A
China (Hohhot)	Multi-zone	Primary zone	Secondary zone
		Zone A	Zone B
		Zone B	Zone A
Germany (Frankfurt)	Multi-zone	Primary zone	Secondary zone
		Zone A	Zone B
		Zone B	Zone A
UK (London)	Multi-zone	Primary zone	Secondary zone
		Zone A	Zone B
		Zone B	Zone A
UAE (Dubai)	Single-zone	Zone A	
Singapore	Multi-zone	Primary zone	Secondary zone
		Zone A	Zone B

Region	Zone type	Zones	
		Zone B	Zone A
		Zone C	Zone B
Australia (Sydney)	Multi-zone	Primary zone	Secondary zone
		Zone A	Zone B
		Zone B	Zone A
Malaysia (Kuala Lumpur)	Multi-zone	Primary zone	Secondary zone
		Zone A	Zone B
		Zone B	Zone A
Indonesia (Jakarta)	Multi-zone	Primary zone	Secondary zone
		Zone A	Zone B
		Zone B	Zone A
India (Mumbai)	Multi-zone	Primary zone	Secondary zone
		Zone A	Zone B
		Zone B	Zone A
Japan (Tokyo)	Multi-zone	Primary zone	Secondary zone
		Zone A	Zone B
		Zone B	Zone A
Hong Kong	Multi-zone	Primary zone	Secondary zone
		Zone B	Zone C
		Zone C	Zone B
US (Virginia)	Multi-zone	Primary zone	Secondary zone
		Zone A	Zone B
		Zone B	Zone A
US (Silicon Valley)	Multi-zone	Primary zone	Secondary zone
		Zone A	Zone B
		Zone B	Zone A

11 Achieve cross-region load balancing through Global Traffic Manager

By using Global Traffic Manager, you can apply global traffic balancing management on a higher plane than the level of local traffic balancing to achieve cross-region disaster tolerance, accelerate access across different regions, and achieve intelligent DNS resolution.

Global traffic management

Server Load Balancer (SLB) provides local load balancing and global load balancing functions according to the geographical positioning of its application. Specifically, the local load balancing function balances a number of server groups in the same region, whereas the global load balancing function balances server groups that are in different regions and have different network requirements.

- Multi-line intelligent resolution

  Global Traffic Manager (GTM) uses DNS intelligent resolution to resolve domain names and health checks to check the running status of application servers so that it can direct access requests to the most appropriate IP addresses, helping users experience the fastest and smoothest access.
- Cross-region disaster tolerance

  With GTM, you can add IP addresses of different regions to different address pools and perform health checks. In access policy configurations, by setting the address pool A as the default IP address pool and address pool B as the failover IP address pool, you can realize disaster tolerance of IP addresses.
- Accelerate access across different regions

  By using GTM, you can direct user access requests from different regions to different IP address pools, thus achieving grouped user and access management, and improving user experience.

Deploy global traffic management

This topic takes the website aliyuntest.club as an example (most users of the website are from Singapore and China) to show you how to achieve global load balancing through GTM and SLB.

Step 1 Purchase and configure ECS instances

Purchase and configure at least two ECS instances in each region where the users of the application service are located.

In this example, two ECS instances are purchased in Beijing, Shenzhen, and Singapore separately, and a simple static web page is built on each ECS instance.

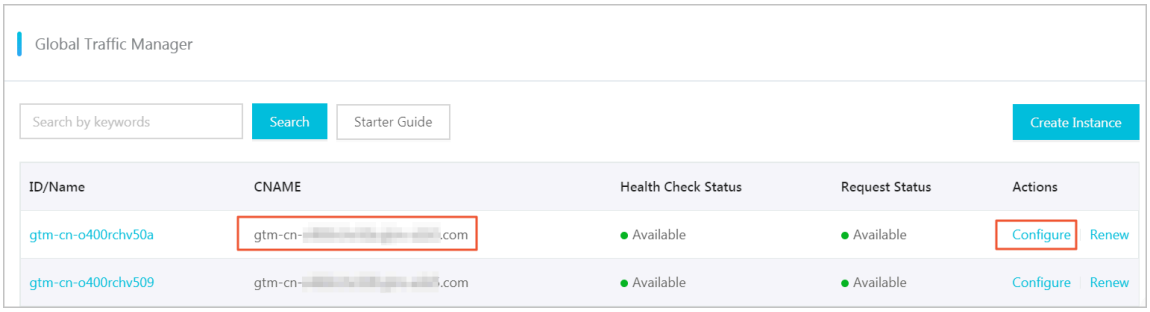
Step 2 Purchase and configure SLB instances

1. Create one Internet SLB instance in each of the region Beijing, Shenzhen, and Singapore. For more information about how to create an Internet SLB instance, see Create an SLB instance.
2. Add listeners for the created SLB instances, and add the configured ECS instances to backend server groups. For more information, see Configure an SLB instance.

Step 3 Configure GTM

1. Purchase a GTM instance.
   a. Log on to the Alibaba Cloud DNS console.
   b. In the left-side navigation pane, click Global Traffic Manager.
   c. On the Global Traffic Manager page, click Create Instance.
   d. Select the Version, Quantity, and Service Time.
   e. Click Buy Now.

   After the instance is successfully purchased, the system automatically allocates a CNAME record.



2. Configure the GTM instance.
   a. On the Global Traffic Manager page, click the target GTM instance ID or click Configure in the Actions column.
   b. In the left-side navigation pane, click Configurations.
   c. On the Global Settings tab page, click Edit to set the parameters of the GTM instance.

      Configure the following parameters and use the default values for the remaining parameters.

      - Instance Name: It is used to help you identify which application this instance is created for. Enter a customized name.
      - Primary Domain: It is the domain name you use to access the application. In this example, enter aliyuntest.club.
      - Alert Group: Select an alarm contact group you configured in CloudMonitor. When an exception occurs, the contact group is notified.

   d. Click Confirm.

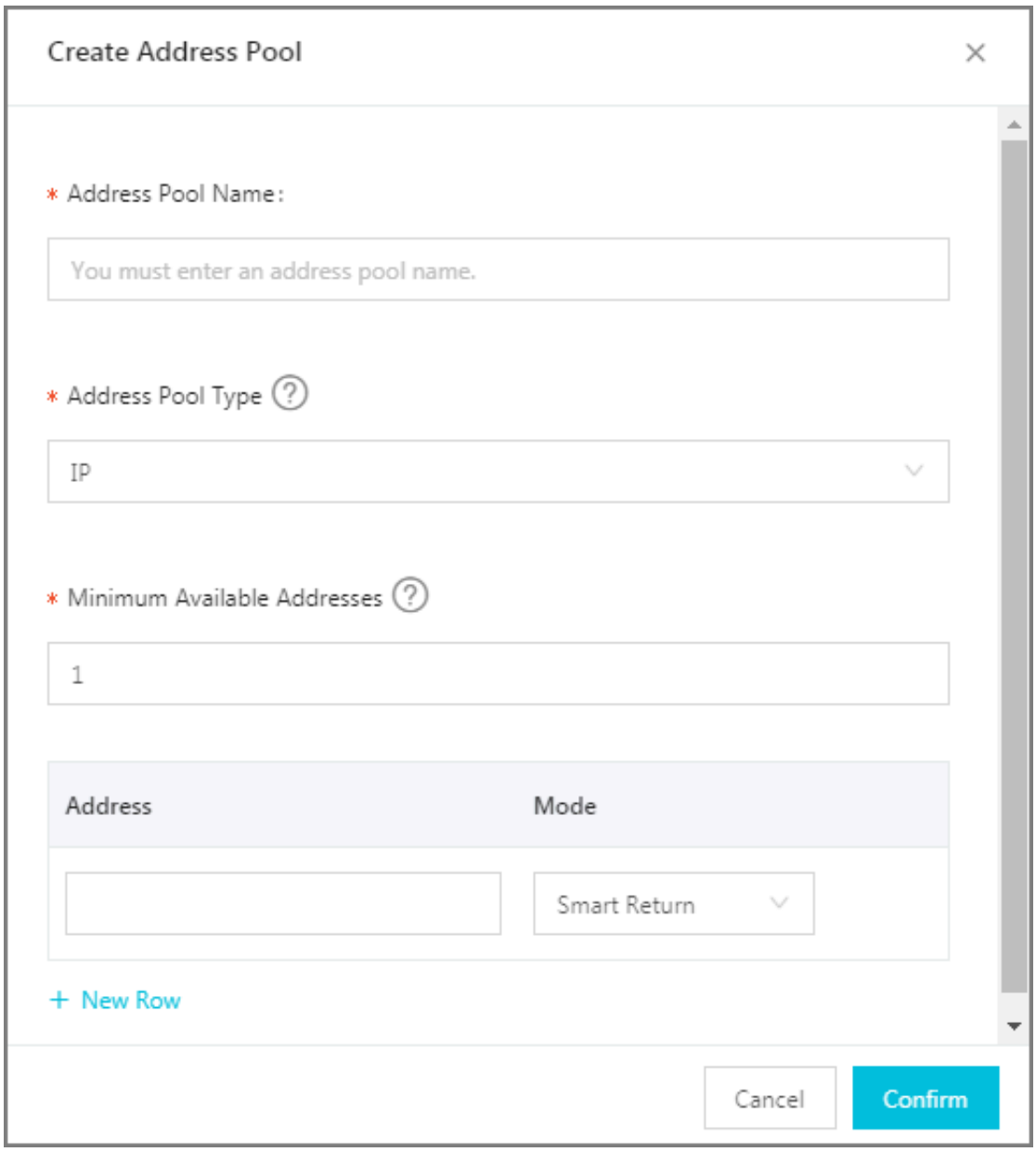
3. Configure address pool.

   a. On the Address Pool Configurations tab page, click Create Address Pool.

   b. On the Create Address Pool page, configure the address pool.

      In this example, create three address pools and each address pool accommodates the addresses of one of the three SLB instances.

      - Address Pool Name: Enter a name, for example, China North_Beijing, China East_Shenzhen, and Singapore.
      - Address: Enter the public IP address of the Internet SLB instance that belongs to the region in the address pool name.



   c. Click Confirm.

4. Configure health check.

   In this example, configure health checks for the three address pools separately.

   a. On the Address Pool tab page, click the target address pool.
   b. In the Settings area, click Add next to Health Check.
   c. Set health check parameters.

      Monitoring Node shows the locations of monitoring nodes. Select the monitoring node according to the region of the address pool.

5. Configure access policy.

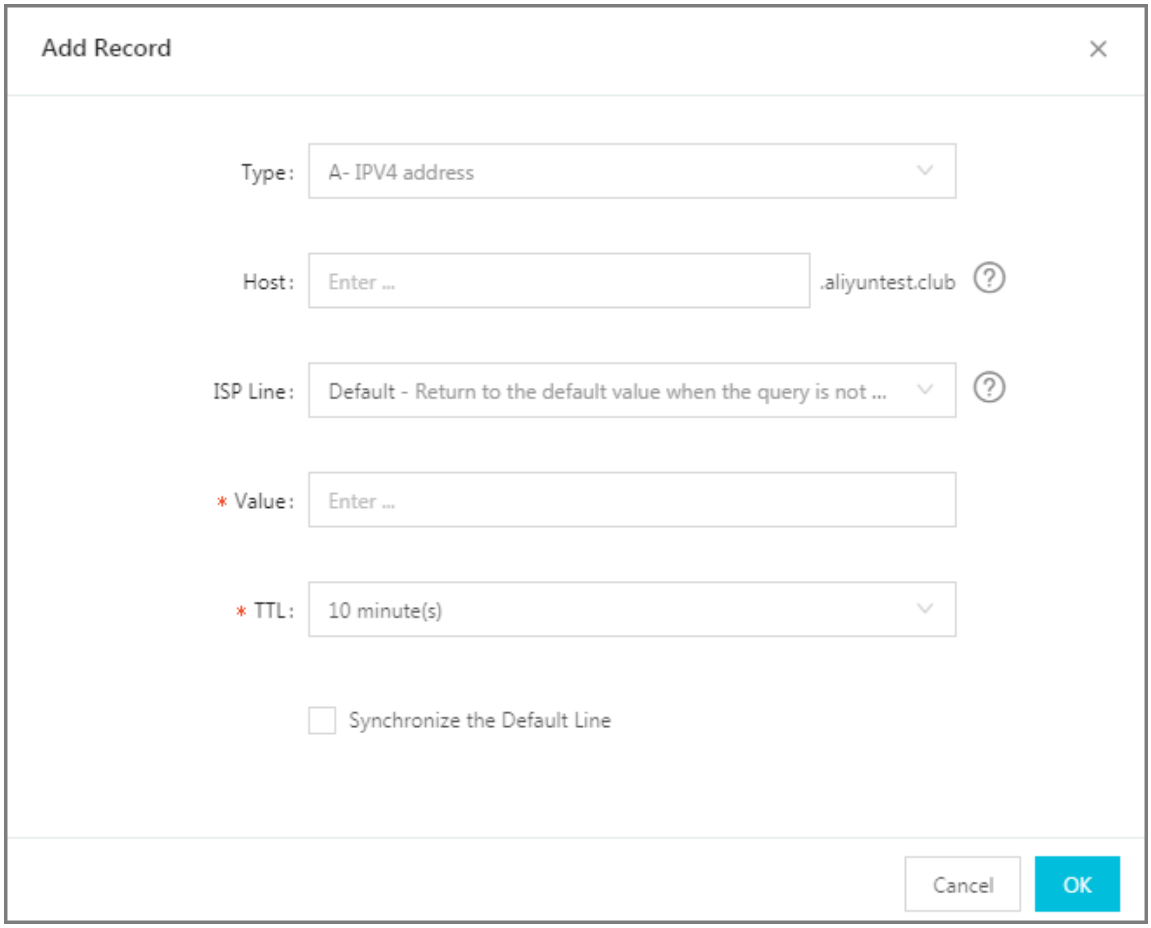
   In this example, add different access policies for the three different regions.

   a. On the Access Policy tab page, click Add Access Policy.
   b. On the Add Access Policy page, configure the access policy.

      - Configure corresponding default address pools for different access regions, and set an address pool of another region as the failover address pool.
      - Select the access region. When users in this region access the application, the address pool configured in the access policy is matched.

        There must be an access policy with Global selected as the access region. Otherwise some regions cannot access the application.

6. Configure CNAME access.
   a. Log on to the Alibaba Cloud DNS console.
   b. Find the domain name aliyuntest.club and click Configure in the Actions column.
   c. On the DNS Settings page, click Add Record.
   d. On the Add Record page, direct the domain name that is accessed by end users, aliyuntest.club in this example, to the CNAME record of the GTM instance.



   e. Click OK.

Step 4 Test

Remove the ECS instances of the SLB instance in Beijing so that the SLB service becomes unavailable.

Visit the website to see if the access is normal.

Note:

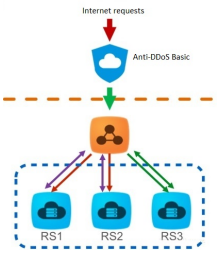
It takes one to two minutes for GTM to make judgment after it detects that your IP address is down. If you set the monitoring frequency to 1 minute, it takes two to three minutes for the failover to take effect.

12 Anti-DDoS Basic

You can view Alibaba Cloud Security thresholds of an Internet Server Load Balancer (SLB) instance through the SLB console.

Introduction to Anti-DDoS Basic

Alibaba Cloud provides up to 5 Gbit/s Anti-DDoS Basic for SLB. As shown in the following figure, all traffic from the Internet must first go through Alibaba Cloud Security before arriving at SLB. Anti-DDoS Basic scrubs and filters common DDoS attacks and protects your services against attacks such as SYN flood, UDP flood, ACK flood, ICMP flood, and DNS Query flood.



Anti-DDoS Basic sets the scrubbing threshold and blackholing threshold according to the bandwidth of the Internet SLB instance. When the inbound traffic reaches the threshold, scrubbing or blackholing is triggered:

- Scrubbing: When the attack traffic from the Internet exceeds the scrubbing threshold or matches certain attack traffic model, Alibaba Cloud Security starts scrubbing the attack traffic. The scrubbing includes packet filtration, traffic speed limitation, packet speed limitation and more.
- Blackholing: When the attack traffic from the Internet exceeds the blackholing threshold, blackholing is triggered and all inbound traffic is dropped.

The thresholds are calculated based on the following principles:

- The thresholds are determined by the bandwidth of the SLB instance, that is, the outbound bandwidth of the SLB instance. The thresholds are high when the bandwidth of the instance is high and vise versa.
- The blackholing threshold is determined by the security credit score of your account.

Note:

The security credit score only influences the blackholing threshold and does not influence the scrubbing threshold.

Complete these steps to calculate the threshold:

1. The SLB backstage provides the recommended threshold value that can ensure normal running of the instance according to the purchased bandwidth.

   Note:
   The outbound bandwidth of a Pay-As-You-Go instance is the peak bandwidth in the region. Currently the peak bandwidth in Mainland China is 5 Gbit/s. For more information, see #unique_91.

   - The relationship between SLB bandwidth and traffic scrubbing threshold (bit/s)
     - When the SLB bandwidth < 100 Mbit/s, the default traffic scrubbing threshold (bit/s) = 120 Mbit/s
     - When the SLB bandwidth > 100 Mbit/s, the default traffic scrubbing threshold (bit/s) = bandwidth × 1.2
   - The relationship between SLB bandwidth and traffic scrubbing threshold (packet/s)

     Traffic scrubbing threshold (packet/s) = (SLB bandwidth/500) × 150000

     The SLB bandwidth is in Mbit/s.
   - The relationship between SLB bandwidth and blackholing threshold (bit/s)
     - When the SLB bandwidth < 1 Gbit/s, the default blackholing threshold (bit/s) = 2 Gbit/s
     - When the SLB bandwidth > 1 Gbit/s, the default blackholing threshold (bit/s) = MAX (SLB bandwidth × 1.5, 2 Gbit/s)

2. Alibaba Cloud Security calculates the threshold according to the recommended value, the security credit score and the resource conditions in different regions.
   - Rules for determining the traffic scrubbing threshold (bit/s) and the traffic scrubbing threshold (packet/s)

     The minimum traffic scrubbing threshold (bit/s) is 1,000 M and the minimum traffic scrubbing threshold (packet/s) is 300,000.

     - If the threshold recommended by SLB is lower than the minimum cleaning threshold, the minimum threshold is used.
     - If the threshold recommended by SLB is higher than the minimum cleaning threshold, the recommended threshold is used.
   - Alibaba Cloud Security determines the blackholing threshold according to the security credit score of your account.

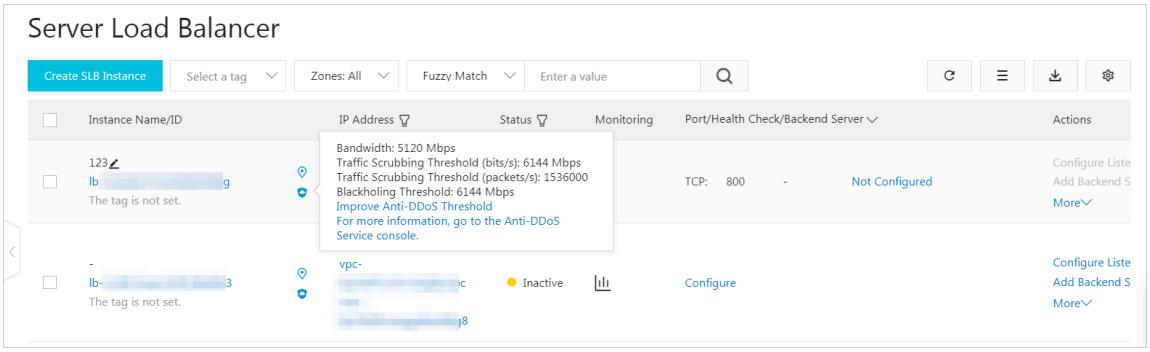
View thresholds

You can view the thresholds of an instance in the SLB console as a RAM user. If not, you must authorize the RAM account first. For more information, see Allow read-only access to Anti-DDoS Basic.

To view thresholds, follow these steps:

1. Log on to the SLB console.
2. Select the region of the target SLB instance.

3. Rest the pointer over the DDoS icon next to the target SLB instance to view the following thresholds. You can click the link to go to the DDoS console to view more information.
   - Traffic Scrubbing Threshold (bit/s): When the inbound traffic exceeds this value , scrubbing is triggered.
   - Traffic Scrubbing Threshold (packet/s): When the inbound packets exceed this value, scrubbing is triggered.
   - Blackholing Threshold: When the inbound traffic exceeds this value, blackholing is triggered.

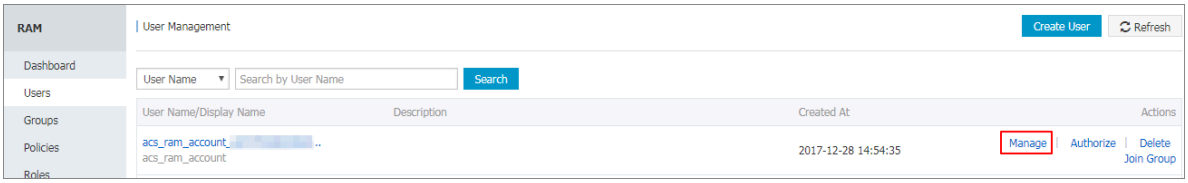


Allow read-only access to Anti-DDoS Basic

To allow read-only access to Anti-DDoS Basic, follow these steps:

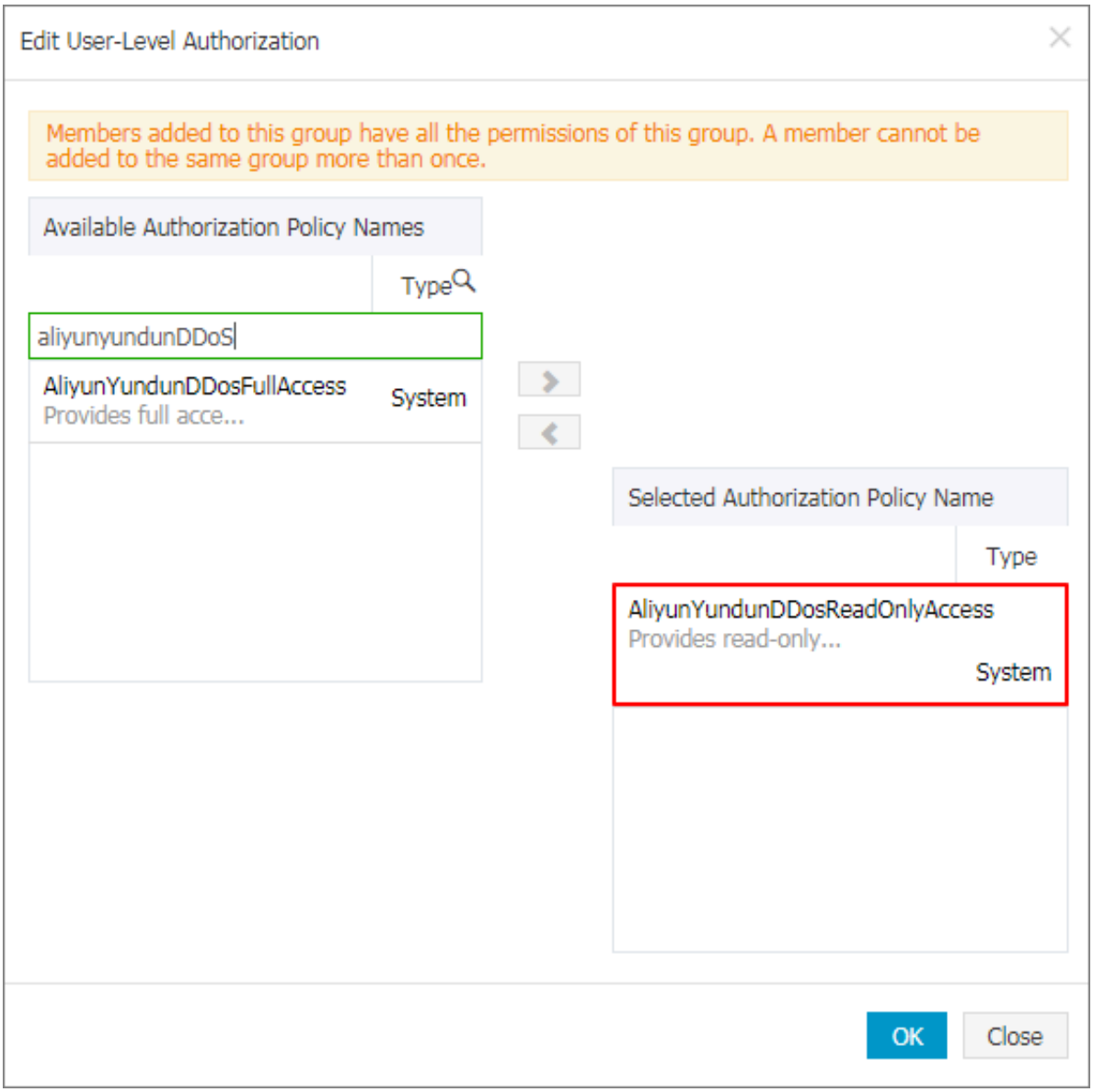
> **Note:**
> Use the Alibaba Cloud account to complete the authorization.

1. Use the Alibaba Cloud account to log on to the RAM console.
2. In the left-side navigation pane, click Users, find the target RAM user and click Manage.



3. Click User Authorization Policies, and then click Edit Authorization Policy.

4. In the displayed dialog box, search AliyunYundunDDosReadOnlyAccess, and then add it to the Selected Authorization Policy Name list. Click OK.



View the security credit score

The security credit score is provided by Alibaba Cloud based on your attack history , purchase history, account activity, security level, expectation and more. With a higher security credit score, you can have a higher free blackholing threshold and a shorter blackholing duration (how long the blackholing status lasts).

To view the security credit score, follow these steps:

1. Log on to the Anti-DDoS Basic console.
2. Select Anti-DDoS Basic > Instances.

3. Click the Security Credibility link to view the security credit score of the account.

> **Note:**
> Security credit scores are region-based.