Alibaba Cloud
Server Load Balancer

Common Configurations

Issue: 20190815

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 informatio n, 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]*

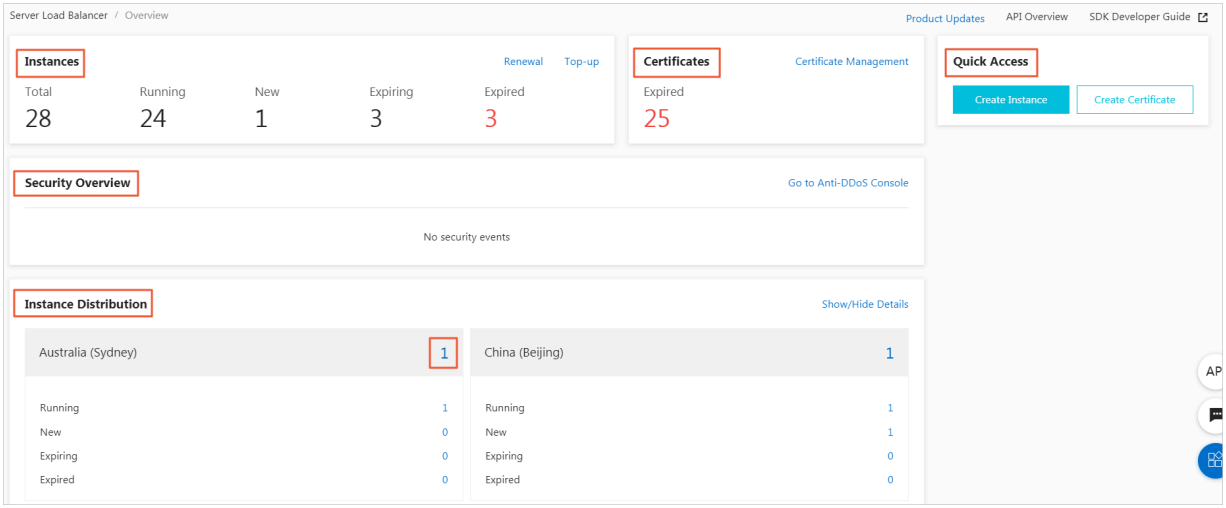
Style	Description	Example
{} or {a|b}	It indicates that it is a required value, and only one item can be selected.	`swich` *{stand | slave}*

Contents

1 Overview page

This topic describes the overview page in the Server Load Balancer (SLB) console. On the overview page, you can see all the SLB instances under the current account, the associated certificates, the security status of the SLB instances, and the regional distribution of the SLB instances.

To view the overview page of the SLB console, log on to the SLB console. In the left-side navigation pane, click Overview.



The following table describes the functions of the overview page.

Section	Description
Instances	Displays the number of SLB instances in different states under the current account. You can click Renewal to renew expiring SLB instances or click Top-up to add funds to the current account.
Certificates	Displays the number of certificates in different states under the current account. You can click Certificate Management to view certificate details and manage certificates.

Section	Description
Security Overview	Displays the security information of SLB instances.
Instance Distribution	Displays the number of SLB instances in different regions under the current account. You can click a number to go to the instance management page of the corresponding SLB instance.
Quick Access	Provides the entry to common SLB operations, including creating an SLB instance and creating a certificate.

2 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 

.

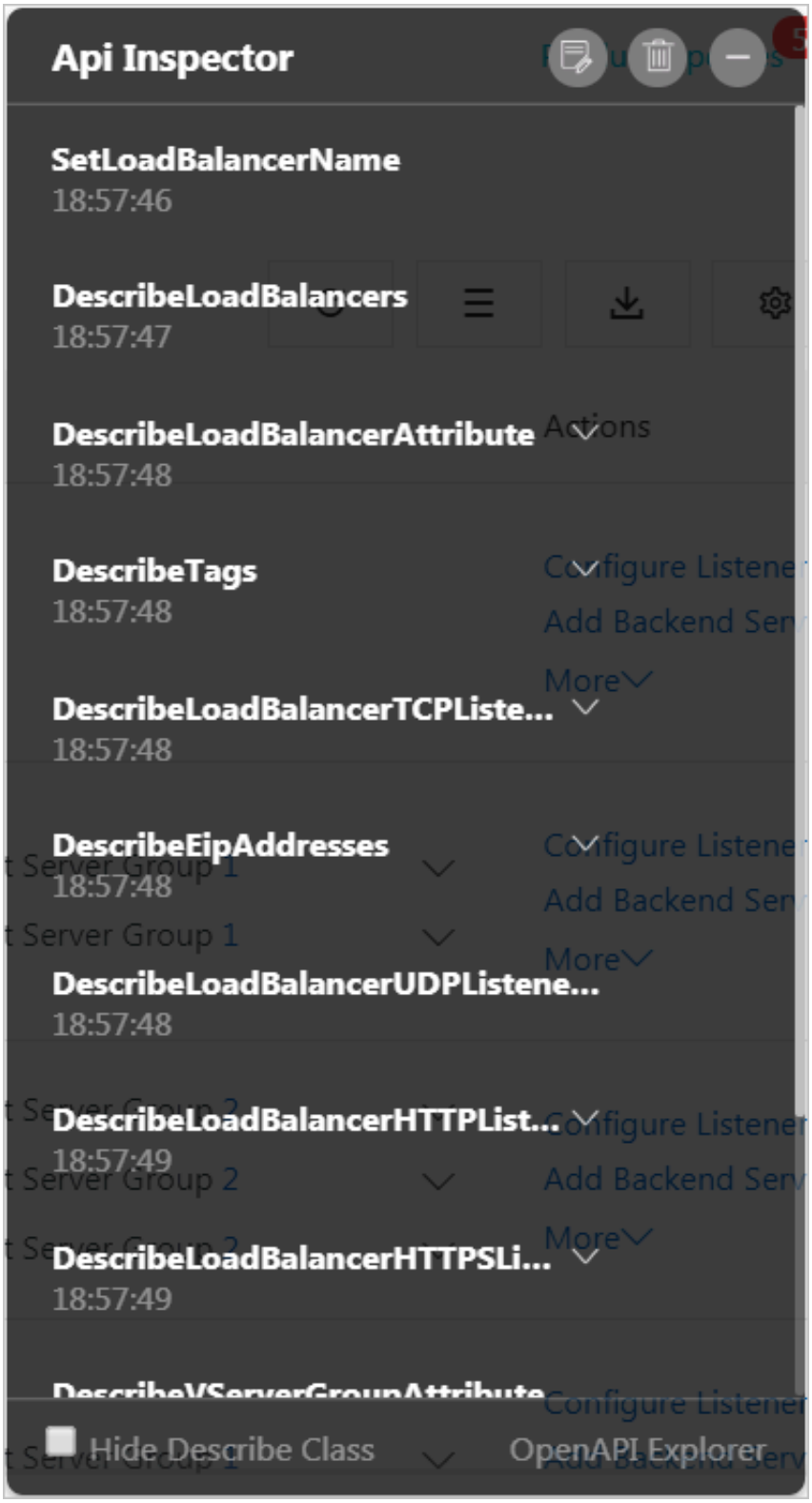
Automatically record API calls

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

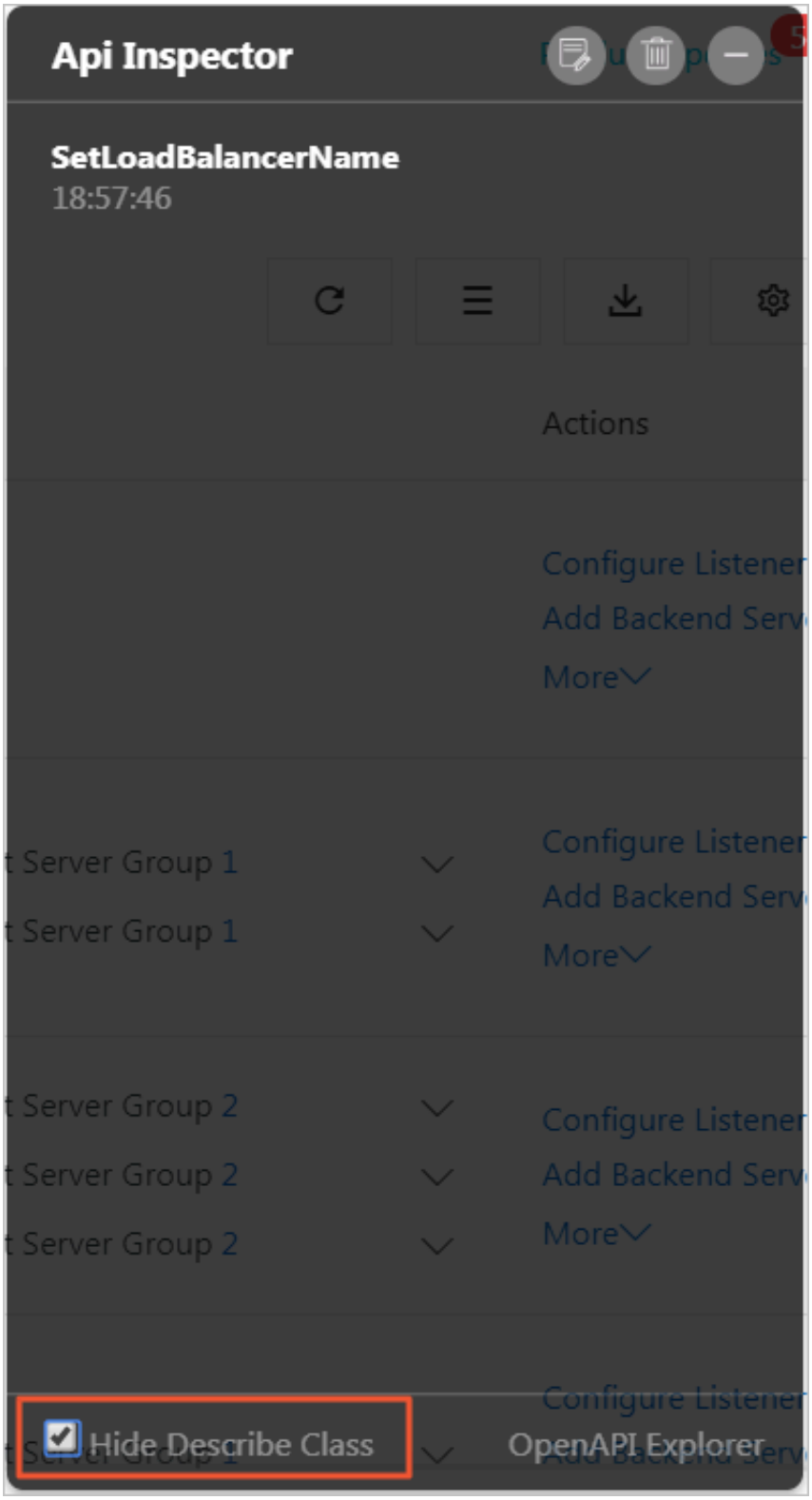
1. Choose Instances > Server Load Balancer.
2. Modify the name of an SLB instance to SLB1.
3. Click OK.

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`.

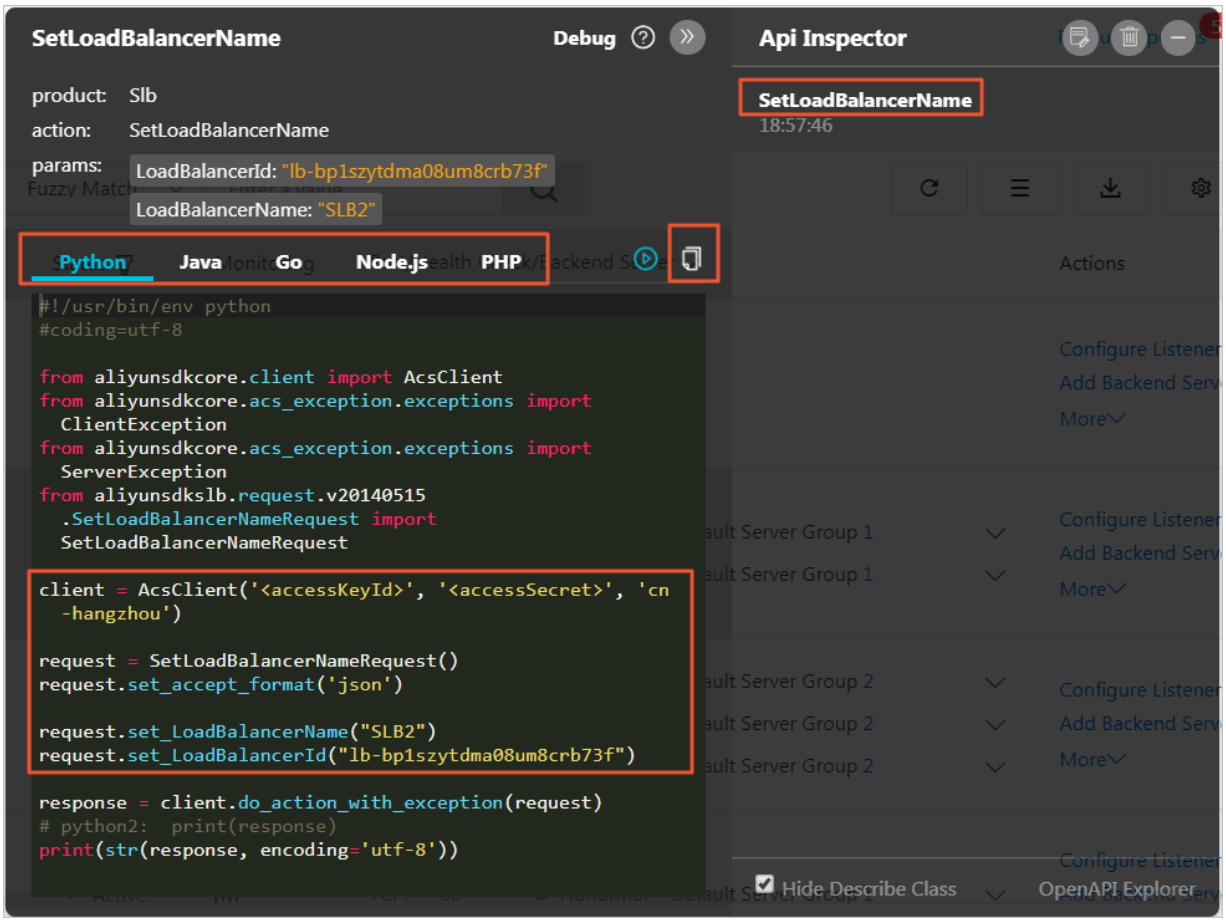


Generate API codes with one click

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:

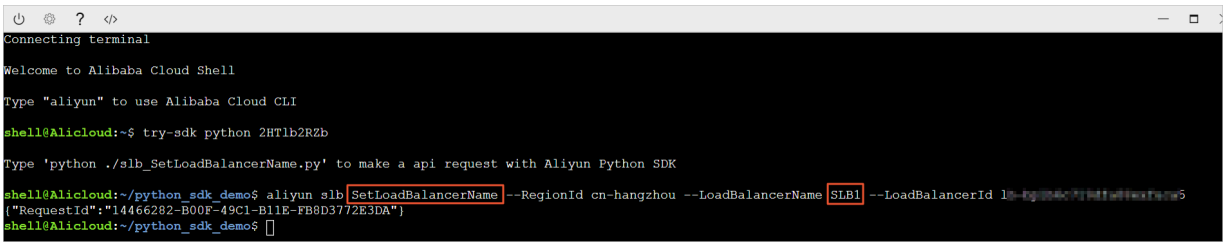
```
aliyun slb actionName --parameter1value1 --paramter2value2...
```

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

```
aliyun slb SetLoadBalancerName --RegionId cn-hangzhou
--LoadBalancerName SLB1 --LoadBalancerId lb-bp1b6c719d
fa08exfuca5
```

The returned value is:

```
{"RequestId":"14466282-B00F-49C1-B11E-FB8D3772E3DA"}
```



3 Multi-zone deployment

You can create Server Load Balancer (SLB) instances in a region with multiple zones to improve the service availability.

What is multi-zone deployment?

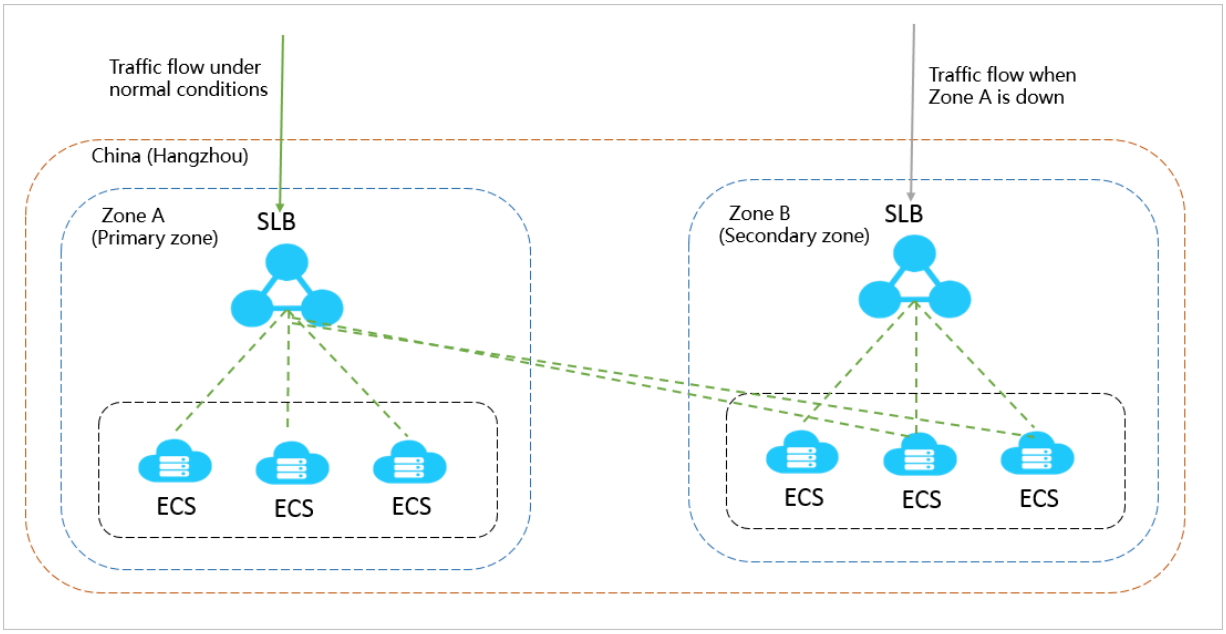
A cloud product zone is a set of independent infrastructures. Different zones have independent infrastructures (such as network, power supply, and air-conditioning). Therefore, infrastructure faults in one zone does not affect other zones.

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	Zone	
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	Zone	
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	Zone	
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	Zone	
		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
China (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

4 Achieve cross-region load balancing through Global Traffic Manager

By using Global Traffic Manager (GTM), 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

  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.

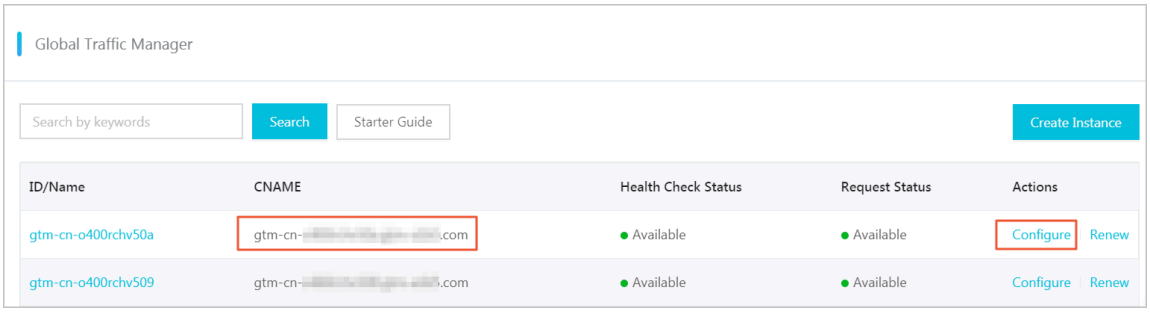
Step 2 Purchase and configure SLB instances

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, click Edit to set the parameters of the GTM instance.

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

      - 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.

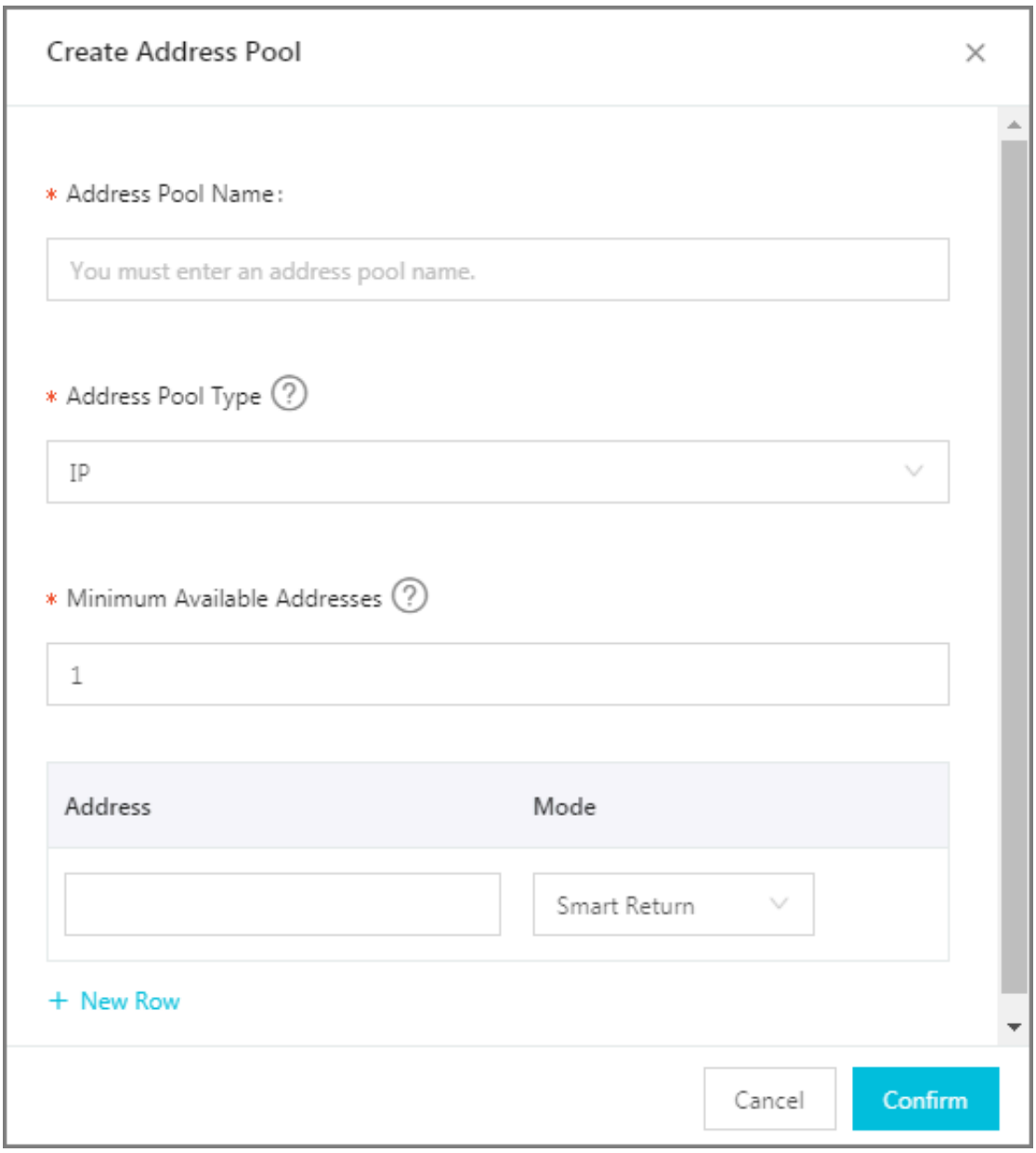
3. Configure address pools.

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

   b. On the Create Address Pool page, configure the IP 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 checks.

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

   a. On the Address Pool Configurations tab, click Edit next to Health Check in the Settings section.
   b. Configure 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 policies.

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

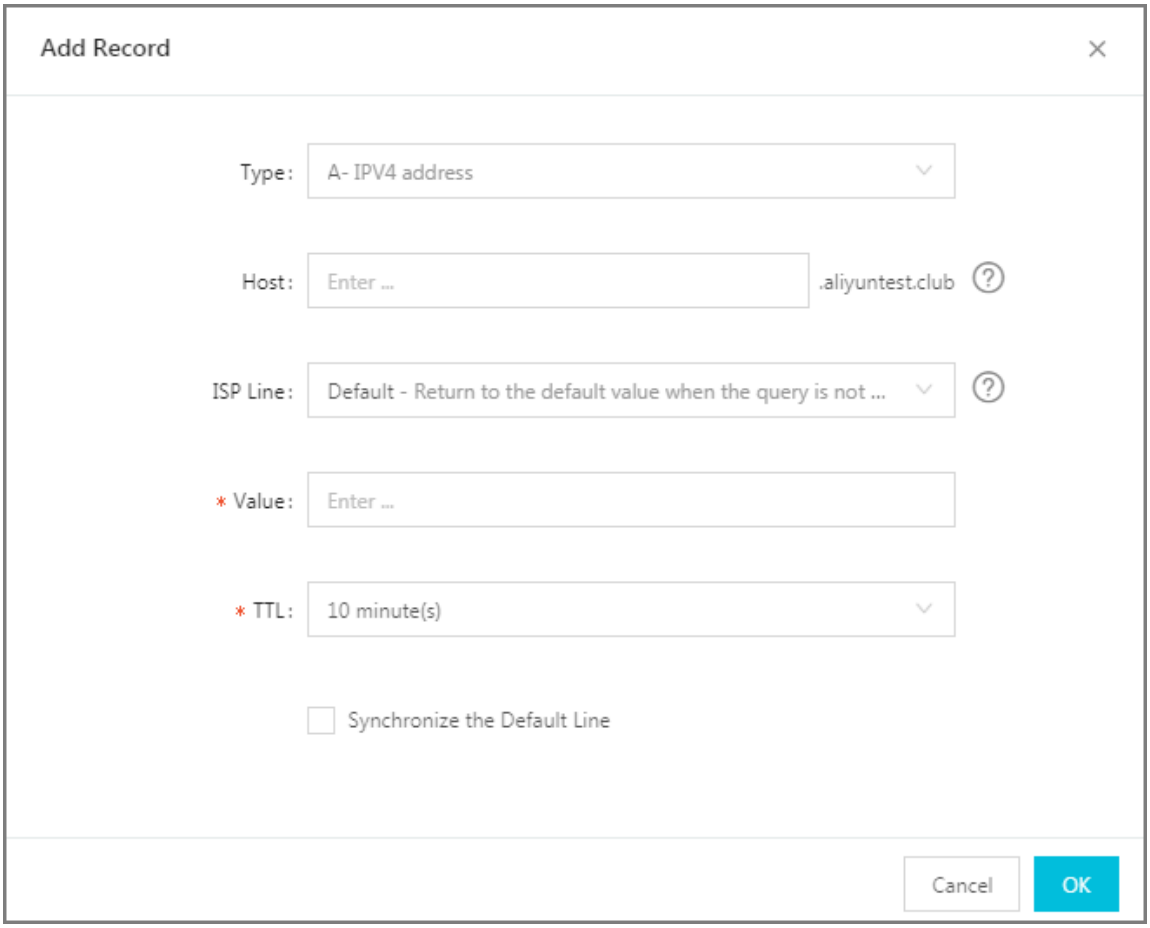
   a. On the Access Policy tab, click Add Access Policy.
   b. On the Add Access Policy page, configure the access policy.

      - 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. 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 the Beijing region so that the SLB service becomes unavailable.

Visit the website to see if the access is normal.

Note:

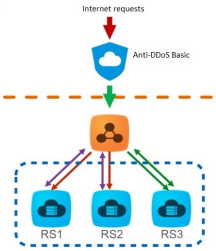
It takes one to two minutes for GTM to make judgment after it detects that your IP address is down. If you set the monitoring frequency to 1 minute, it takes two to three minutes for the failover to take effect.

5 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_11.

   - 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.

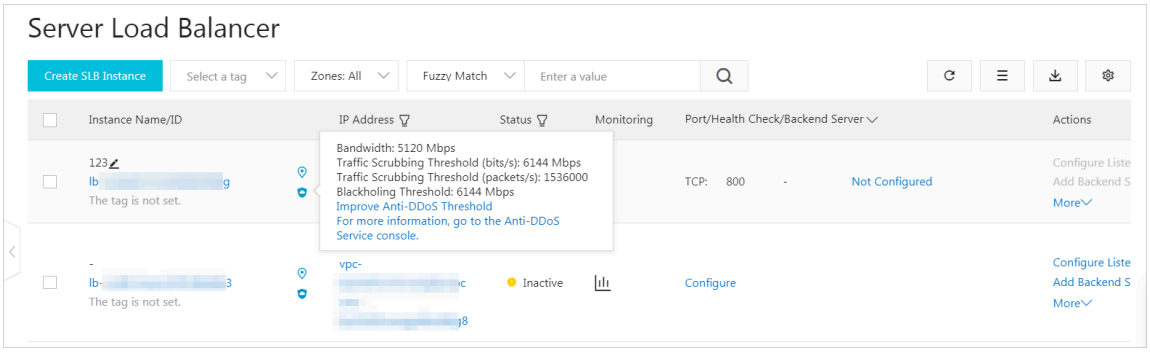
View thresholds

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

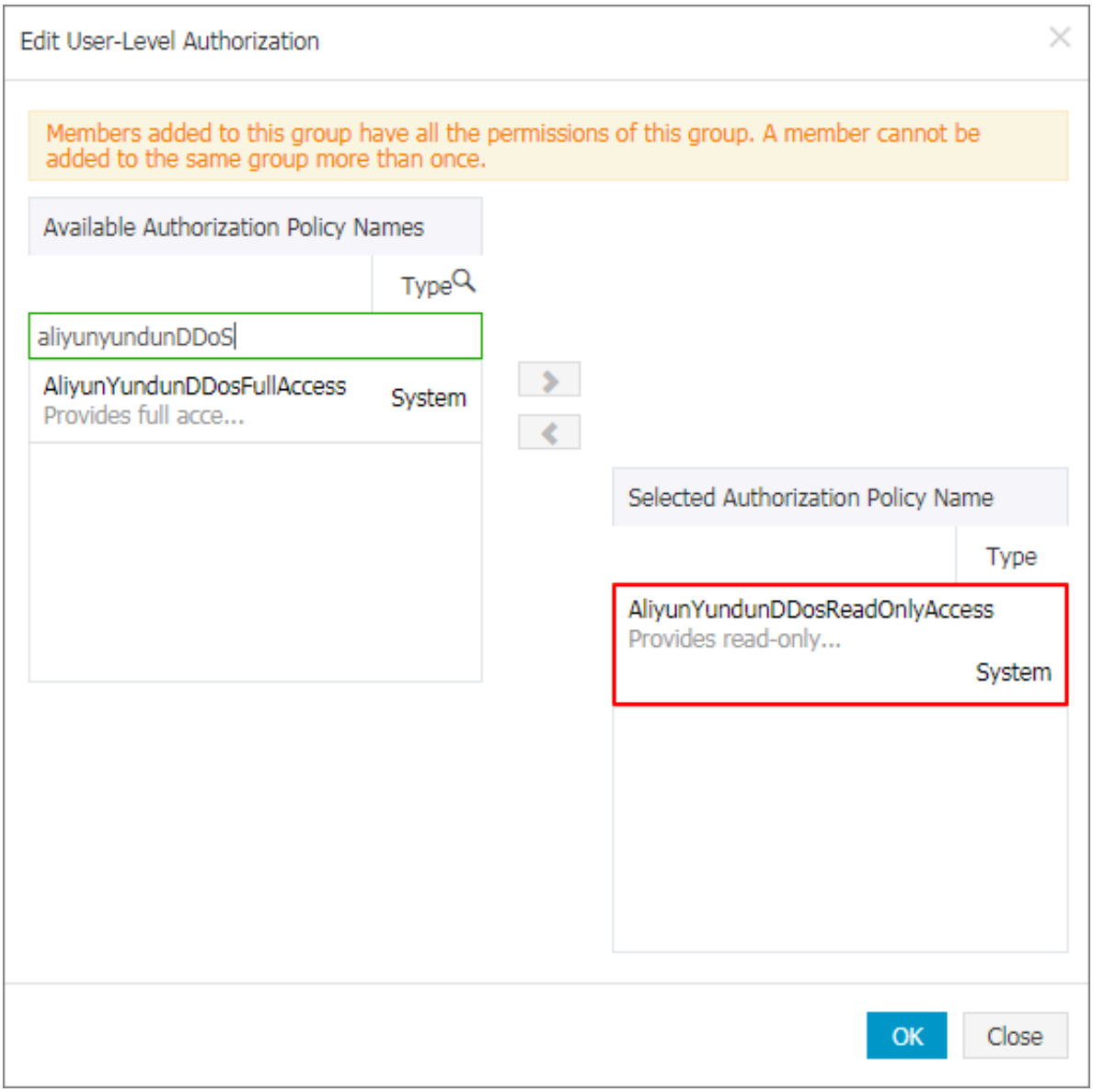
To allow read-only access to Anti-DDoS Basic, follow these steps:

> **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.