

Alibaba Cloud

Server Load Balancer Common Configurations

Document Version: 20220623

Legal disclaimer

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

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.
2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company or individual in any form or by any means without the prior written consent of Alibaba Cloud.
3. The content of this document may be changed because of product version upgrade, adjustment, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and an updated version of this document will be released through Alibaba Cloud-authorized channels from time to time. You should pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.
4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides this document based on the "status quo", "being defective", and "existing functions" of its products and services. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not take legal responsibility for any errors or lost profits incurred by any organization, company, or individual arising from download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, take responsibility for any indirect, consequential, punitive, contingent, special, or punitive damages, including lost profits arising from the use or trust in this document (even if Alibaba Cloud has been notified of the possibility of such a loss).
5. By law, all the contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of this document shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates.
6. Please directly contact Alibaba Cloud for any errors of this document.

Document conventions

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

Table of Contents

1.Bandwidth limits in different regions	05
2.API Inspector	07
3.Implement cross-region load balancing by using Global Traffic ...	09
4.Anti-DDoS Origin (Basic Edition)	14

1. Bandwidth limits in different regions

This topic describes the bandwidth limits of pay-as-you-go Classic Load Balancer (CLB) instances in different regions.

Note

- The bandwidth limits of internal-facing CLB instances are 5 Gbit/s in all regions.
- For each pay-by-bandwidth CLB instance, the inbound bandwidth limit is the same as the outbound bandwidth limit.
- The bandwidth limit of a pay-by-data-transfer CLB instance indicates the upper limit of data transfer rate and is used only for reference. The peak bandwidth is not guaranteed.
- The bandwidth limit of a pay-by-bandwidth CLB instance indicates the peak bandwidth and is a guaranteed value.
- When CLB instances compete for resources, the peak bandwidth of pay-by-bandwidth CLB instances is guaranteed but the peak bandwidth of pay-by-data-transfer CLB instances may be limited.

Region	Bandwidth limit
China (Qingdao)	5 Gbps
China (Hangzhou)	5 Gbps
China (Beijing)	5 Gbps
China (Shanghai)	5 Gbps
China (Shenzhen)	5 Gbps
China (Zhangjiakou)	5 Gbps
China (Hohhot)	5 Gbps
China (Hong Kong)	2 Gbps
US (Virginia)	1 Gbps
US (Silicon Valley)	2 Gbps
Japan (Tokyo)	1 Gbps
Singapore (Singapore)	5 Gbps
Australia (Sydney)	1 Gbps
Malaysia (Kuala Lumpur)	5 Gbps

Region	Bandwidth limit
UAE (Dubai)	500 Mbps
Germany (Frankfurt)	1 Gbps
India (Mumbai)	5 Gbps

2.API Inspector

API Inspector is an experimental feature that allows you to view the API call that corresponds to each operation in the console. API Inspector can automatically generate API code snippets for each programming language. You can debug the API code snippets by using Cloud Shell and OpenAPI Explorer.


Features

API Inspector is integrated with OpenAPI Explorer and Cloud Shell to provide an integrated API learning and debugging solution with the following features:

- Automatic recording: You can obtain relevant API calls by performing corresponding operations in the console. For more information, see [Automatic recording](#).
- Code generation: The system automatically generates API code snippets with preset parameters for different programming languages. You can directly run the API code snippets. For more information, see [Code generation](#).
- Online debugging: You can use OpenAPI Explorer and Cloud Shell to debug the code snippets without the need to build a developing environment. For more information, see [Online debugging with OpenAPI Explorer](#) and [Online debugging with Cloud Shell](#).


Enable API Inspector

Perform the following operations to enable API Inspector:

1. Log on to the [Classic Load Balancer \(CLB\) console](#).
2. Click  in the lower-right corner.

Automatic recording



The following example describes how to use the automatic recording feature in a scenario where you want to modify the name of a CLB instance in the console.

1. Click the ID of the CLB instance that you want to manage and click the **Instance Details** tab.
2. In the **Basic Information** section, click **Edit** to modify the name of the CLB instance.
3. Click **OK**.
4. Click  on the right of the page to view the API call that corresponds to the operation.

You can select **Hide Describe Class** to view core API operations.

Code generation

After the API call that corresponds to the operation is recorded, click the name of the API. Then, the system generates the API code snippets with preset parameters in the following formats: Python, Java, Go, Node.js, PHP, and CLI.


 **Note** Click  to copy the code snippet in a specific format. You can directly run the code snippet.


Online debugging with OpenAPI Explorer

After the API call that corresponds to the operation is recorded, click **OpenAPI Explorer** to go to the **OpenAPI Explorer** page and debug the code snippet.

 **Note** Click  to view relevant documentation and parameters.

Online debugging with Cloud Shell

After the API call that corresponds to the operation is recorded, expand the details about the API call and click  to debug the code snippet with Cloud Shell.

 **Note** When you use Cloud Shell for debugging, we recommend that you create and associate an Object Storage Service (OSS) bucket to store common scripts and files. In this case, you are charged for using OSS. You can also choose not to use OSS.

Run a command in the following format to debug CLB API with Cloud Shell:

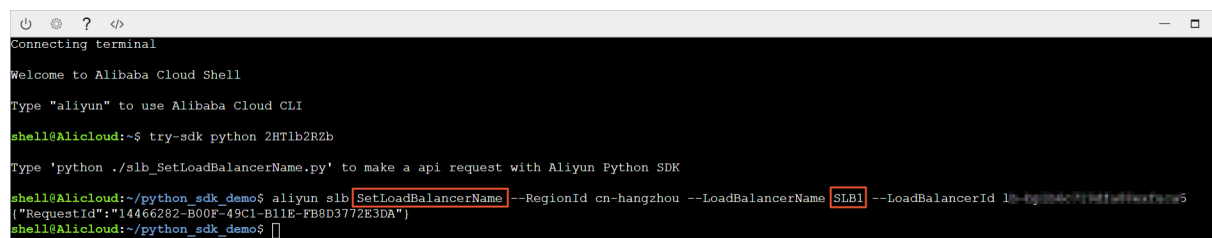
```
aliyun slb actionName --parameter1value1 --parameter2value2...
```

In this example, the API operation `SetLoadBalancerName` is performed to modify the name of the SLB1 instance. Therefore, run the following command:

```
aliyun slb SetLoadBalancerName --RegionId cn-hangzhou --LoadBalancerName SLB1 --LoadBalancerId lb-bp1b6c719dfa08exfuca5
```

The following value is returned:

```
null
```



```
Connecting terminal
Welcome to Alibaba Cloud Shell
Type "aliyun" to use Alibaba Cloud CLI
shell@AlibabaCloud:~$ try-sdk python 2HT1b2RZb
Type 'python ./slb_SetLoadBalancerName.py' to make a api request with Aliyun Python SDK
shell@AlibabaCloud:~/python_sdk_demo$ aliyun slb SetLoadBalancerName --RegionId cn-hangzhou --LoadBalancerName SLB1 --LoadBalancerId lb-bp1b6c719dfa08exfuca5
{"RequestId": "14466282-B00F-49C1-B11E-FB8D3772E3DA"}
shell@AlibabaCloud:~/python_sdk_demo$
```


3. Implement cross-region load balancing by using Global Traffic Manager

This topic describes how to manage global traffic over local load balancing services by using Global Traffic Manager (GTM) to accelerate access across regions and enable cross-region disaster recovery and intelligent resolution.

Global Traffic Manager

Server Load Balancer (SLB) provides the local and global load balancing features according to the geographical positioning of its application. The local load balancing feature balances server groups in the same region, whereas the global load balancing feature balances server groups that are in different regions and have different network architectures.

- Multi-line intelligent DNS resolution

GTM uses intelligent DNS resolution to resolve domain names and uses health checks to check the running status of application services. This way, GTM directs access requests to the most appropriate IP addresses and provides fast and smooth access for users.

- Cross-region disaster recovery

GTM allows you to add IP addresses of different regions to different address pools and perform health checks. You can set the default address pool to **Address Pool A** and the failover IP address pool to **Address Pool B** in the access policy configurations. This enables the failover of application services between primary and secondary instances.

- Accelerated access across regions

You can use GTM to direct user access from different regions to different IP address pools, which enables group-based user management and group-based access to improve user experience of application services.

Deployment of Global Traffic Manager

This section shows how to perform global load balancing by using GTM and SLB. A website with the aliyundoc.com domain name is used in the example. Most users of this website are located in Singapore and mainland China.

Step 1: Purchase and configure ECS instances

Purchase and configure at least two ECS instances for each region in which the users of the application service reside.

In this example, two ECS instances are purchased in each of the China (Beijing), China (Shenzhen), and Singapore regions. A static web page is built on each ECS instance.

- Example of the ECS instance in the China (Beijing) region
- Example of the ECS instance in the China (Shenzhen) region
- Example of the ECS instance in the Singapore region

Step 2: Purchase and configure SLB instances

1. Create an Internet-facing SLB instance in each of the China (Beijing), China (Shenzhen), and Singapore regions. For more information, see [Create a CLB instance](#).
 2. Add listeners for the created SLB instances, and add the configured ECS instances to backend server groups. For more information, see [Configure a CLB instance](#).
- Example of the SLB instance in the China (Beijing) region
 - Example of the SLB instance in the China (Shenzhen) region
 - Example of the SLB instance in the Singapore region

Step 3: Configure GTM

1. Purchase a GTM instance.
 - i. Log on to the [Alibaba Cloud DNS console](#).
 - ii. In the left-side navigation pane, click **Global Traffic Manager**.
 - iii. On the **Global Traffic Manager** page, click **Create Instance**.
 - iv. Set Edition, Quantity, and Duration.
 - v. Click **Buy Now**.

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

ID/Name	CNAME Access Domain Name	Health Check Tasks	Notifications (Sent in the Current Month)	Status	Instance Package Version	Expired At (UTC+8)	Actions
gltm-xxxxxxx-25	Public Domain Name	Configured Instances:0 Instance Quota:100	Email:0 DingTalk:0 Text Message:0	Normal	Standard	2021-12-03 00:00:00	Configure Upgrade Renew

2. Configure the GTM instance.
 - i. On the **Global Traffic Manager** page, click the ID of the instance or click **Configure** in the **Actions** column.
 - ii. In the left-side navigation pane, click **Global Traffic Manager**.
 - iii. On the **Global Settings** tab, click **Edit** to configure the instance.

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

 - **Instance Name:** The instance name is used to identify the application service for which the instance is used.
 - **Primary Domain:** The primary domain name is used to access the application service. In this example, aliyundoc.com is used.
 - **Alert Group:** When an exception occurs in GTM, the system notifies contacts in an alert contact group that you configured in Cloud Monitor.
 - iv. Click **Confirm**.
3. Create address pools.
 - i. On the **Address Pool Configurations** tab, click **Create Address Pool**.

- ii. In the **Create Address Pool** panel, configure the address pool.

In this example, three address pools are created. Each address pool accommodates the addresses of one of the three SLB instances.

- **Address Pool Name:** Enter a name of the address pool. Example: China North_Beijing, China East_Shenzhen, or Singapore.
- **Address:** Enter the public IP address of the SLB instance that belongs to the specified region.

Create Address Pool

* Address Pool Name:

You must enter an address pool name.

* Address Pool Type ?

IP

* Minimum Available Addresses ?

1

Address	Mode
	Smart Return

+ New Row

Cancel Confirm

- iii. Click **Confirm**.

4. Configure health checks.

In this example, health checks are configured for the three created address pools.

- i. On the **Address Pool Configurations** tab, click **Edit** to the right of the Health Check switch.
- ii. Configure health check parameters.

The locations of monitoring nodes are displayed in the **Monitoring Node** section. Select a monitoring node based on the region of the address pool.

5. Configure access policies.

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

- i. On the **Access Policy** tab, click **Add Access Policy**.

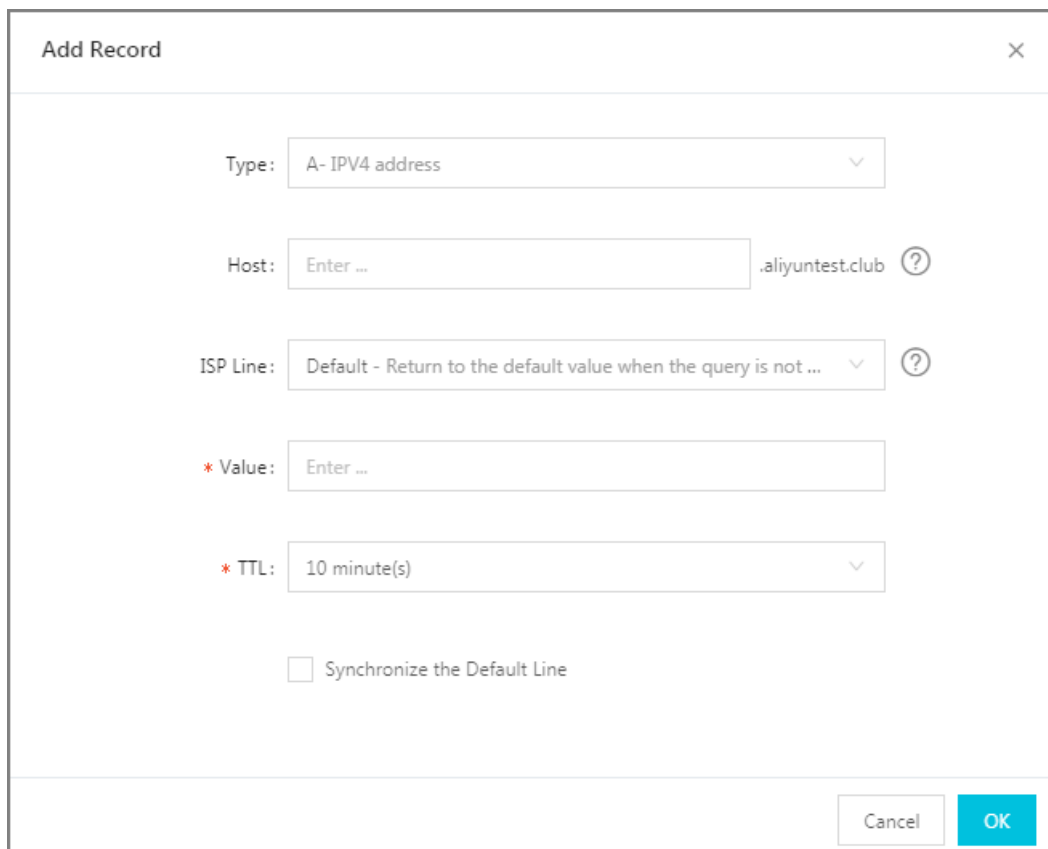
ii. In the **Add Access Policy** panel, configure the access policy.

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

The DNS Request Sources parameter of at least one access policy must be set to **Global**. Otherwise, the application service is inaccessible in some regions.

6. Configure CNAME access.

- Log on to the Alibaba Cloud DNS console.
- Find the domain name aliyundoc.com and click **Configure** in the **Actions** column.
- On the **DNS Settings** page, click **Add Record**.
- On the **Add Record** page, direct the aliyuntest.club domain name that is accessed by end users to the CNAME record of the GTM instance.




v. Click **Confirm**.

Step 4: Perform a test

Remove the backend servers that are attached to the SLB instance in the China (Beijing) region so that the SLB service becomes unavailable.

Visit the website to determine whether it can be accessed normally.

 **Note** It can take up to two minutes for GTM to make a judgment after it detects that your server is down. For example, if you set the monitoring frequency to one minute, it can take up to three minutes for failover to take effect.

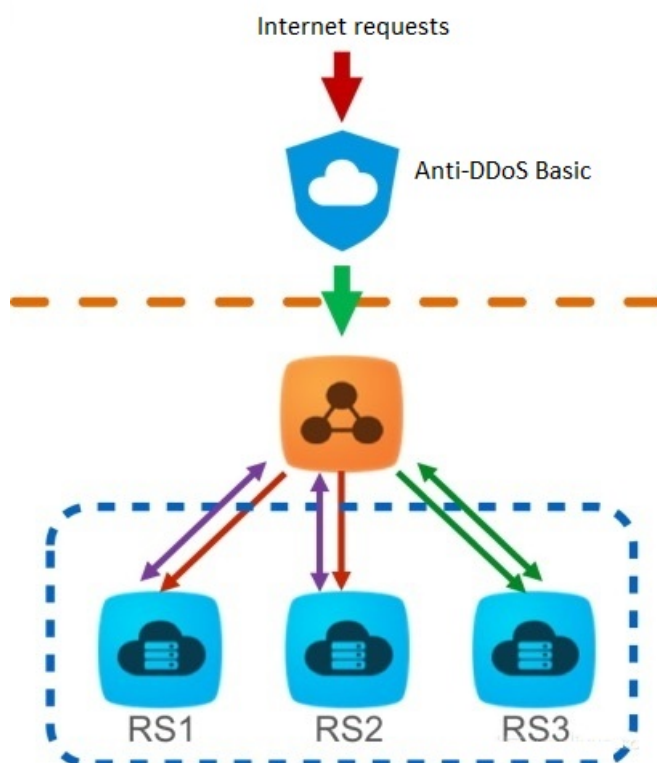
4. Anti-DDoS Origin (Basic Edition)

Anti-DDoS Origin can mitigate DDoS attacks for Elastic Compute Service (ECS), Classic Load Balancer (CLB), Web Application Firewall (WAF), and Elastic IP Address (EIP). Anti-DDoS Origin is integrated with the preceding services. You can use Anti-DDoS Origin without the need to change IP addresses. In addition, no limits of Layer 4 ports or Layer 7 domain names are imposed on Anti-DDoS Origin.

Overview

By default, Anti-DDoS Origin (Basic Edition) is enabled for CLB free of charge. Anti-DDoS Origin (Basic Edition) provides a maximum bandwidth capacity of 5 Gbit/s. All data from the Internet is filtered by Alibaba Cloud Security before the data is transferred to CLB. Alibaba Cloud Security filters out and mitigates DDoS attacks such as SYN flood, UDP flood, ACK flood, ICMP flood, and DNS flood attacks.

Anti-DDoS Origin adopts passive scrubbing as a major protection policy and active blocking as an auxiliary policy to mitigate DDoS attacks. Anti-DDoS Origin uses conventional technologies such as reverse detection, blacklists, whitelists, and packet compliance. These technologies allow protected resources to work as expected under attack. The following figure shows the network topology of Anti-DDoS Origin.



Anti-DDoS Origin (Basic Edition) sets thresholds for scrubbing and blackholing based on the bandwidth of Internet-facing CLB instances. When the inbound traffic reaches the threshold, scrubbing or blackholing is triggered:

- **Scrubbing:** When the system detects attacks that match specific models or a large number of attacks from the Internet, Alibaba Cloud Security automatically scrubs the attacks through packet filtering, traffic throttling, and packet throttling.

- Blackholing: When the system receives a large number of attacks that exceed the threshold, all requests are dropped to ensure security.

Thresholds are calculated based on the following rules:


- The outbound bandwidth of a CLB instance determines the threshold. A greater outbound bandwidth value specifies a higher threshold.
- The blackholing threshold is determined by your security credit score.

 **Note** However, your security credit score does not affect the scrubbing threshold.

Calculate the thresholds

You can perform the following steps to calculate the thresholds.

1. CLB provides a recommended threshold based on the bandwidth resources that you purchase for your CLB instances.

 **Note** If you purchase a pay-by-data-transfer CLB instance, the outbound bandwidth equals the maximum bandwidth supported by the region where the CLB instance is deployed. All regions in mainland China support a maximum bandwidth capacity of 5 Gbit/s. For more information, see [Peak bandwidth limits](#).

- The correlation between the CLB bandwidth and scrubbing threshold (bit/s)
 - When the CLB bandwidth value is less than 100 Mbit/s: Default scrubbing threshold (Mbit/s) = 120
 - When the CLB bandwidth value is greater than 100 Mbit/s: Default scrubbing threshold (Mbit/s) = CLB bandwidth value × 1.2
 - Correlation between the CLB bandwidth and scrubbing threshold (packet/s)
$$\text{Scrubbing threshold (packet/s)} = \text{CLB bandwidth value} / 500 \times 150000$$

Bandwidth values are measured in Mbit/s.
 - Correlation between the CLB bandwidth and blackholing threshold (bit/s)
 - When the CLB bandwidth value is less than 1 Gbit/s: Default blackholing threshold (Gbit/s) = 2
 - When the CLB bandwidth value is greater than 1 Gbit/s: Default blackholing threshold (Gbit/s) = $\text{Max}\{\text{CLB bandwidth value} \times 1.5, 2\}$
2. Alibaba Cloud Security calculates the final thresholds based on the recommended thresholds, security credit score, and resources in each region.
 - Alibaba Cloud Security evaluates the rules of thresholds (bit/s and packet/s).

The minimum value of the threshold is 1000 in Mbit/s and 300000 in packet/s.

 - If the threshold calculated by CLB is less than the preceding minimum value, the minimum value prevails.
 - If the threshold calculated by CLB is greater than the preceding minimum value, the threshold calculated by CLB prevails.
 - Alibaba Cloud Security determines the blackholing threshold based on your security credit score.

Grant read-only permissions to a RAM user

Perform the following steps to grant a RAM user the read-only permissions on Anti-DDoS Origin (Basic Edition).

 **Note** You must use your Alibaba Cloud account to grant the read-only permissions to a RAM user.

1. Log on to the **RAM console** with your Alibaba Cloud account.
2. In the left-side navigation pane, choose **Identities > Users**.
3. On the **Users** page, find the RAM user in the **User Logon Name/Display Name** column and click **Add Permissions**.
4. On the **System Policy** tab, select **AliyunYundunDDoSFullAccess** in the **Authorization Policy Name** column to add it to the Selected list. Then, click **OK**.

View thresholds

Perform the following steps to view thresholds:

1. Log on to the **CLB console**.
2. In the left-side navigation pane, choose **Instances > Instances**.
3. Select the region where the CLB instance is deployed and move the pointer over the Alibaba Cloud Security icon to view the scrubbing threshold (bit/s and packet/s) and blackholing threshold. **For more information, go to the Anti-DDoS console**
 - Scrubbing threshold (bit/s): When the inbound data per second exceeds this value, scrubbing is triggered.
 - Scrubbing threshold (packet/s): When the inbound packets per second exceed this value, scrubbing is triggered.
 - Blackholing threshold: When the inbound data per second exceeds this value, all requests are dropped.