

Alibaba Cloud Server Load Balancer

Archives

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 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 <code>cd / d C :/ windows</code> command to enter the Windows system folder.
<i>Italics</i>	It is used for parameters and variables.	<code>bae log list --instanceid <i>Instance_ID</i></code>
[] or [a b]	It indicates that it is an optional value, and only one item can be selected.	<code>ipconfig [-all -t]</code>

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

Contents

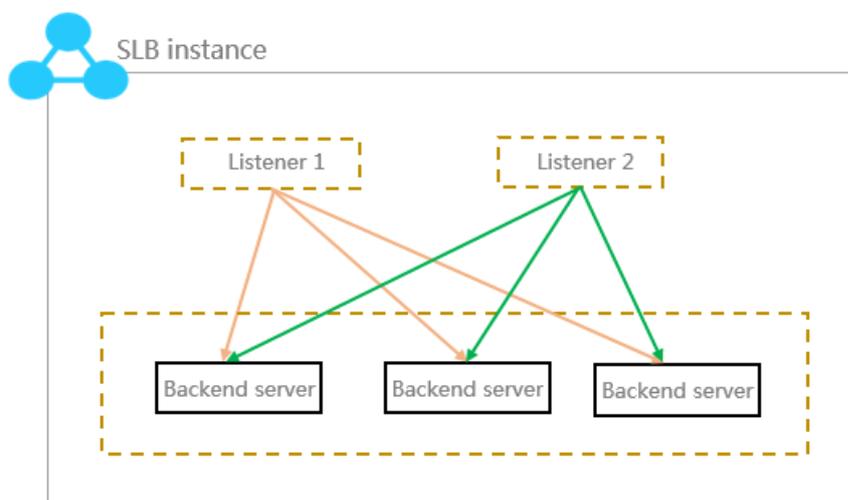
Legal disclaimer.....	I
Generic conventions.....	I
1 User Guide (Old Console).....	1
1.1 SLB instances.....	1
1.1.1 SLB instance overview.....	1
1.1.2 Guaranteed-performance instances.....	3
1.1.3 Network traffic flow.....	10
1.1.4 Create an instance.....	13
1.1.5 Manage an SLB instance.....	14
1.1.6 Change the configuration.....	16
1.2 Listener.....	17
1.2.1 Listeners overview.....	17
1.2.2 Layer-4 listeners.....	20
1.2.3 Layer-7 listeners.....	20
1.2.4 Health check.....	20
1.2.5 Shared instance bandwidth.....	20
1.3 Backend servers.....	21
1.3.1 Backend server overview.....	21
1.3.2 Add default servers.....	23
1.3.3 Create a VServer group.....	24
1.3.4 Create a master-slave server group.....	25
1.4 Certificate management.....	26
1.4.1 Certificate requirements.....	26
1.4.2 Generate CA certificates.....	29
1.4.3 Convert certificate formats.....	34
1.4.4 Upload certificates.....	34
1.4.5 Replace a certificate.....	36
1.5 Log management.....	36
1.5.1 View operation logs.....	36
1.5.2 Configure access logs.....	38
1.5.3 Authorize a RAM user to configure access logs.....	45
1.5.4 Manage health check logs.....	50
1.6 Anti-DDoS Basic.....	55
1.7 Peak limit for regional bandwidth.....	57
1.8 Multiple zone deployment.....	58
1.9 Achieve cross-regional load balancing with cloud resolution.....	60

1 User Guide (Old Console)

1.1 SLB instances

1.1.1 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.

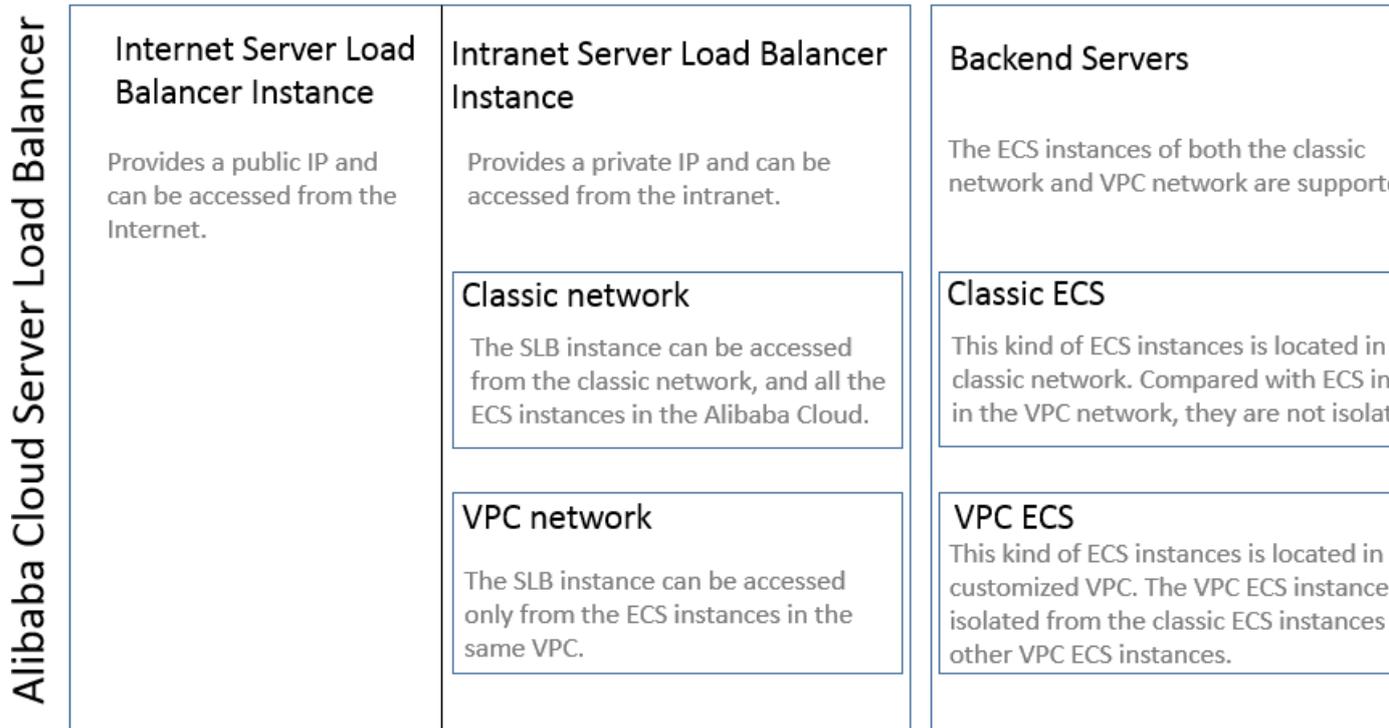


Alibaba Cloud provides two types of load balancing services, Internet load balancing service and intranet load balancing. A public or a private IP address is allocated to the SLB instance according to the instance type you select.

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:

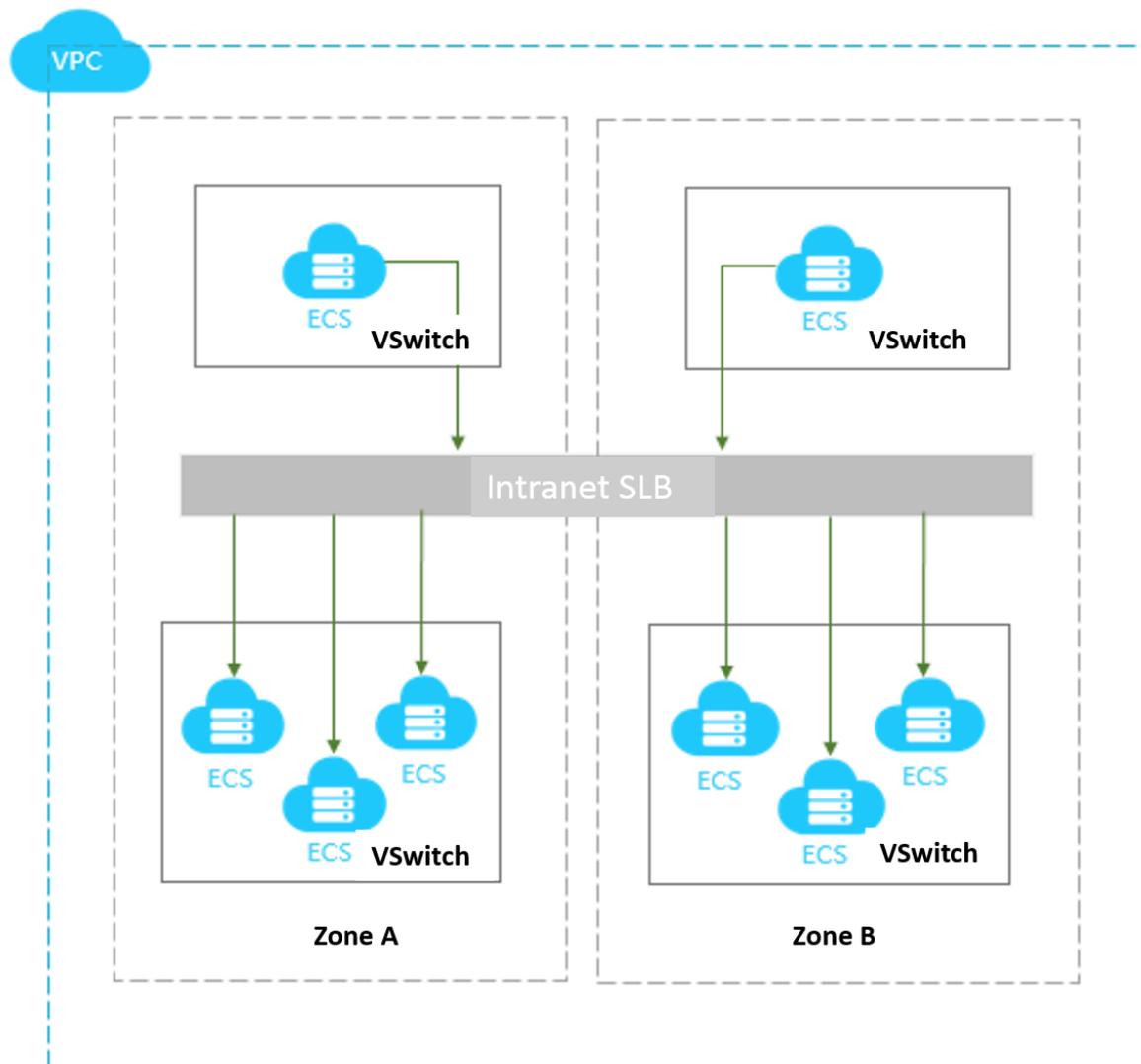
- Classic network

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.



1.1.2 Guaranteed-performance instances

Alibaba Cloud launched the guaranteed-performance instances in May 2017, and charged the capacity fee on guaranteed-performance instances on April 1, 2018.

1. [What are guaranteed-performance instances?](#)
2. [How are guaranteed-performance instances billed?](#)
3. [What is the price of each capacity?](#)
4. [How to select a guaranteed-performance instance?](#)
5. [Can I modify the capacity after the instance is created?](#)
6. [When will the guaranteed-performance instances be charged?](#)

7. After Alibaba Cloud starts to charge capacity fee on guaranteed-performance instances, will extra fees be charged on shared-performance instances?

8. Why sometimes guaranteed-performance instances cannot reach the performance limit as defined in the capacity?

9. Can I still buy shared-performance instances?

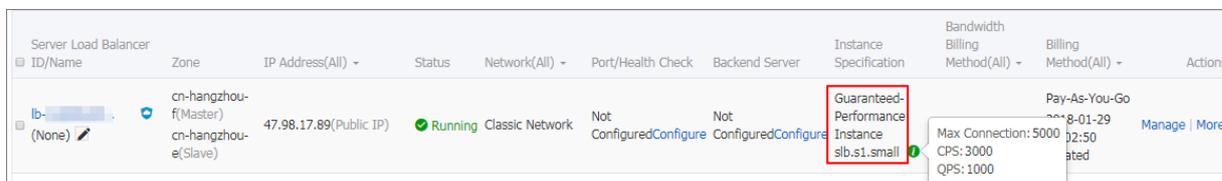
10. Will intranet SLB instances be charged for capacity fee?

1. What are guaranteed-performance instances?

A guaranteed-performance instance provides guaranteed performance metrics (performance SLA). It is opposite to a shared-performance instance. For a shared-performance instance, the performance metrics are not guaranteed and the resources are shared by all instances.

All instances are shared-performance instances before Alibaba launches guaranteed-performance instances. You can view the instance type on the console.

Hover your mouse pointer to the green icon of the target instance to view the performance metrics, as shown in the following figure.



ID/Name	Zone	IP Address(All)	Status	Network(All)	Port/Health Check	Backend Server	Instance Specification	Bandwidth Billing Method(All)	Billing Method(All)	Actions
lb- (None)	cn-hangzhou-f(Master) cn-hangzhou-e(Slave)	47.98.17.89(Public IP)	Running	Classic Network	Not Configured	Not Configured	Guaranteed-Performance Instance slb.s1.small	Pay-As-You-Go	8-01-29	Manage More

The following are three key performance metrics for guaranteed-performance instances:

- **Max Connection**

The maximum number of connections to a SLB instance. When the maximum number of connections reaches the limits of the capacity, the new connection will be dropped.

- **Connection Per Second (CPS)**

The rate at which a new connection is established per second. When the CPS reaches the limits of the capacity, the new connection will be dropped.

- **Query Per Second (QPS)**

The number of HTTP/HTTPS requests that can be processed per second. When the QPS reaches the limits of the capacity, the new connection will be dropped.

Alibaba Cloud Server Load Balancer provides the following capacities for guaranteed-performance instances:

Type	Type	Max Connection	CPS	QPS
Capacity 1	Small I (slb.s1.small)	5,000	3,000	1,000
Capacity 2	Standard I (slb.s2.small)	50,000	5,000	5,000
Capacity 3	Standard II (slb.s2.medium)	100,000	10,000	10,000
Capacity 4	Higher I (slb.s3.small)	200,000	20,000	20,000
Capacity 5	Higher II (slb.s3.medium)	500,000	50,000	30,000
Capacity 6	Super I (slb.s3.large)	1,000,000	100,000	50,000

If you want to use a larger capacity, contact your customer manager.

2. How are guaranteed-performance instances billed?

Guaranteed-performance instances are billed as follows:

Total fee (per instance) = instance fee + traffic fee + capacity fee



Note:

For intranet SLB instances, you can also choose to use guaranteed-performance instances. If guaranteed-performance instances are selected, capacity fee is collected and billed as the Internet SLB instance, but no traffic fee and instance fee are collected.

The performance guarantee instance specification fee is charged by usage, no matter what kind of specification you choose, the instance specification fee will be charged according to the specifications you actually use.

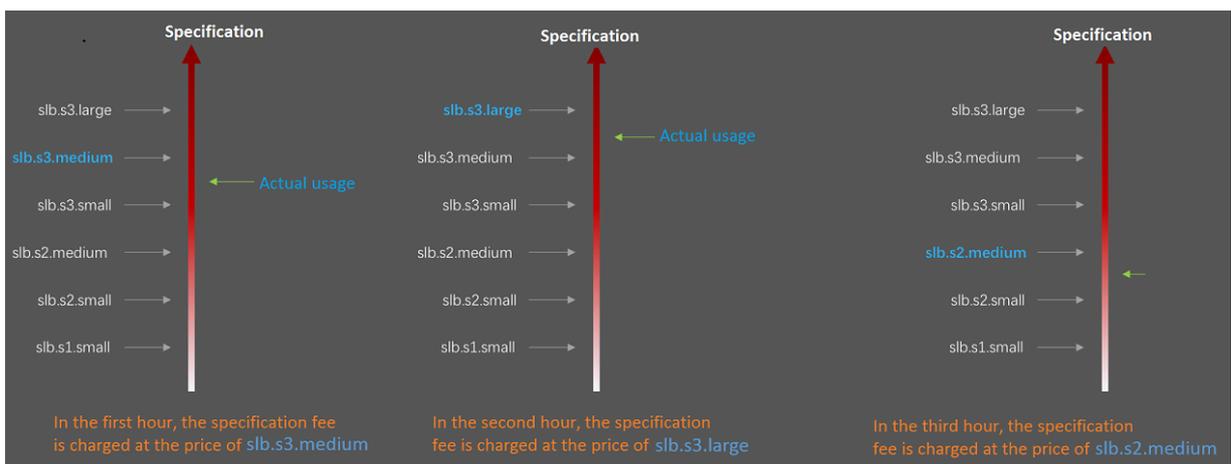
For example, if you purchase the slb.s3.large capacity (1,000,000; CPS 500,000; QPS 50,000) and the actual usage of your instance in an hour is as follow:

Max Connection	CPS	QPS
90,000	4,000	11,000

- From the perspective of Max Connection, the actual metrics 90,000 occurs between the limit 50,000 defined in the Standard I (slb.s2.small) capacity and the limit 100,000 defined in the Standard II (slb.s2.medium) capacity. Therefore, the capacity of the Max Connection metrics in this hour is Standard II (slb.s2.medium).
- From the perspective of CPS, the actual metrics 4,000 occurs between the limit 3,000 defined in the Small I (slb.s1.small) capacity and the limit 5,000 defined in the Standard I (slb.s2.small) capacity. Therefore, the capacity of the CPS metrics in this hour is Standard I (slb.s2.small).
- From the perspective of QPS, the actual metrics 11,000 occurs between the limit 10,000 defined in the Standard II (slb.s2.medium) capacity and the limit 20,000 defined in the Higher I (slb.s3.small) capacity. Therefore, the capacity of the QPS metrics in this hour is Higher I (slb.s3.small)

Comparing these three metrics, the capacity of the QPS metrics is highest, therefore, the capacity fee of the instance in this hour is charged at the price of the Higher I (slb.s3.small) capacity.

The following figure is an example showing how the capacity fee is billed for an SLB instance in the first three hours:



The billing of the guaranteed-performance instances is flexible. The capacity you select when purchasing an instance is the performance limitation of the instance. For example, if slb.s3.medium is selected, the new connections are dropped when the HTTP requests in one second reach 30,000.

3. What is the price of each capacity?

The following table lists the capacity price of each capacity. In addition to the capacity fee, you are also charged for instance fee and traffic fee. For more information, see [Billing method](#).

Region	Type	Max Connectio	CPS	QPS	Capacity fee (USD/ Hour)
China (Hangzhou)	Small I (slb.s1. small)	5,000	3,000	1,000	Free
China (Zhangjiakou)	Standard I (slb.s2. small)	50,000	5,000	5,000	0.05
China (Hohhot)	Standard II (slb.s2. medium)	100,000	10,000	10,000	0.10
China (Qingdao)	Higher I (slb.s3. small)	200,000	20,000	20,000	0.20
China (Beijing)	Higher II (slb.s3. medium)	500,000	50,000	30,000	0.31
China (Shenzhen)	Super I (slb.s3.large)	1,000,000	100,000	50,000	0.51
Singapore	Small I (slb.s1. small)	5,000	3,000	1,000	Free
Malaysia (Kuala Lumpur)	Standard I (slb.s2. small)	50,000	5,000	5,000	0.06
Indonesia (Jakarta)	Standard II (slb.s2. medium)	100,000	10,000	10,000	0.12
India (Mumbai)	Higher I (slb.s3. small)	200,000	20,000	20,000	0.24
US (Silicon Valley)	Higher II (slb.s3. medium)	500,000	50,000	30,000	0.37
China (Hong Kong)	Super I (slb.s3.large)	1,000,000	100,000	50,000	0.61

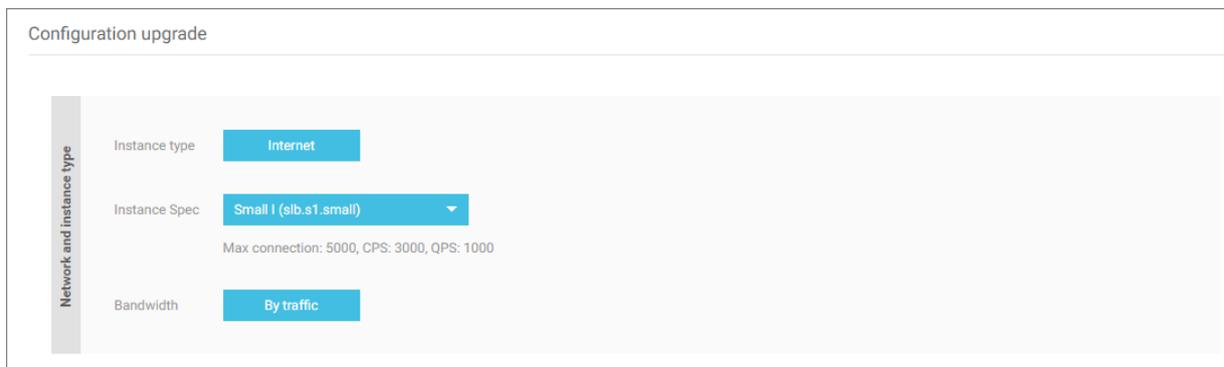
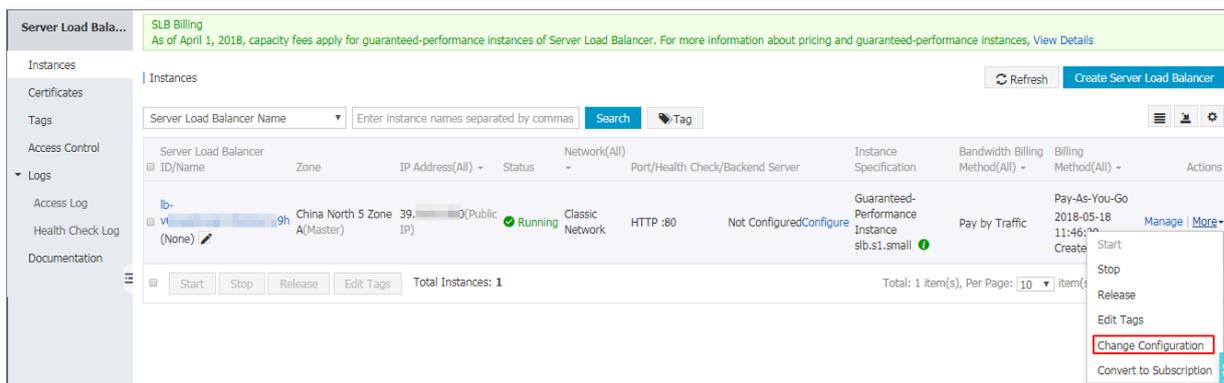
Capacity fees of guaranteed-performance instances in the international regions enjoy an 83% discount.

4. How to select a guaranteed-performance instance?

Because the capacity fee is billed based on the actual usage, we recommend that you select the largest capacity (slb.s3.large). This guarantees the business flexibility (flexibility) and will not cause extra costs. If your traffic does not reach the largest capacity, you can select a more reasonable capacity, such as slb.s3.medium.

5. Can I modify the capacity after the instance is created?

Yes. You can change the capacity at any time and the change takes effect immediately.



Note:

- After you change a shared-performance instance to a guaranteed-performance instance, you cannot change it back.
- Some SLB servers are deployed in old clusters. If 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 specification when the business is not busy.
- The IP of the SLB instance will not be changed after you changing the instance type or the capacity.

Caution
When you change the configuration of an SLB instance or 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 perform this operation when the service is not busy or after the service migrates to another SLB instance by using Global Server Load Balancer . (Changes made to the billing method and network bandwidth of the SLB instance will not affect the service.)
<input type="button" value="I Agree"/> <input type="button" value="No, Not Now"/>

6. When will the guaranteed-performance instances be charged?

Alibaba Cloud launched the guaranteed-performance instances in May 2017, and charged the capacity fee on guaranteed-performance instances on April 1, 2018.

The capacity fee takes effect in batches as follows:

- The first batch:

Time: From April 1st to April 10th

Regions: Singapore, Malaysia (Kuala Lumpur), Indonesia (Jakarta), India (Mumbai), US (Silicon Valley), US (Virginia)

- The second batch:

Time: From April 11th to April 20th

Regions: China (Hangzhou), China (Zhangjiakou), China (Hohhot), China (Hong Kong)

- The third batch:

Time: From April 21th to April 30th

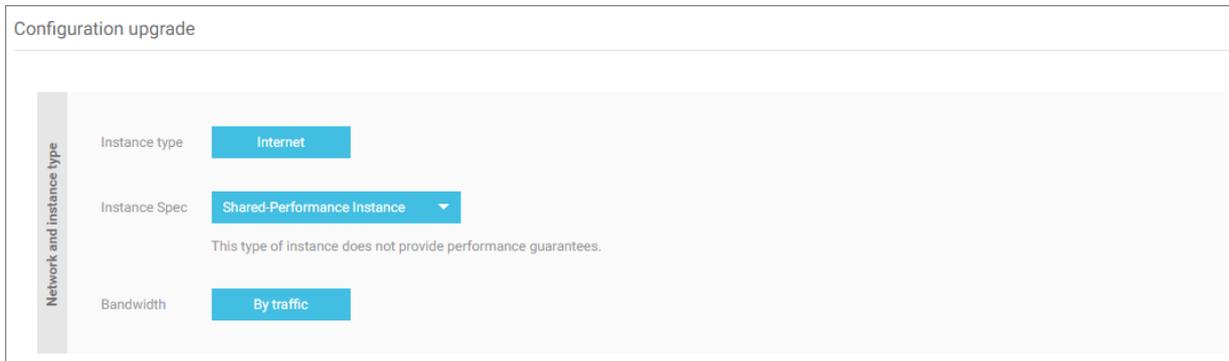
Regions: China (Qingdao), China (Beijing), China (Shanghai), China (Shenzhen)

7. After Alibaba Cloud starts to charge capacity fee on guaranteed-performance instances, will extra fees be charged on shared-performance instances?

No.

The billing of the original shared-performance instances is the same if you do not change it to a performance-guaranteed instance. You can change a shared-

performance instance to a guaranteed-performance instance. After changing to the guaranteed-performance instance, capacity fees are collected.



8. Why sometimes guaranteed-performance instances cannot reach the performance limit as defined in the capacity?

The cask theory.

Guaranteed-performance instances do not guarantee that the three metrics can reach the capacity limits at the same time. The limitation is triggered as long as a metric first reaches the limitation defined in the capacity.

For example, you have purchased a guaranteed-performance instance of the Higher I (slb.s3.small) capacity. When the QPS of the instance reaches 20,000 but the number of maximum connections does not reach 200,000, the new connections are still dropped because the QPS has reached the limitation.

9. Can I still buy shared-performance instances?

Yes.

However, shared-performance instances will be phased out in the future. Please pay attention to the official announcement.

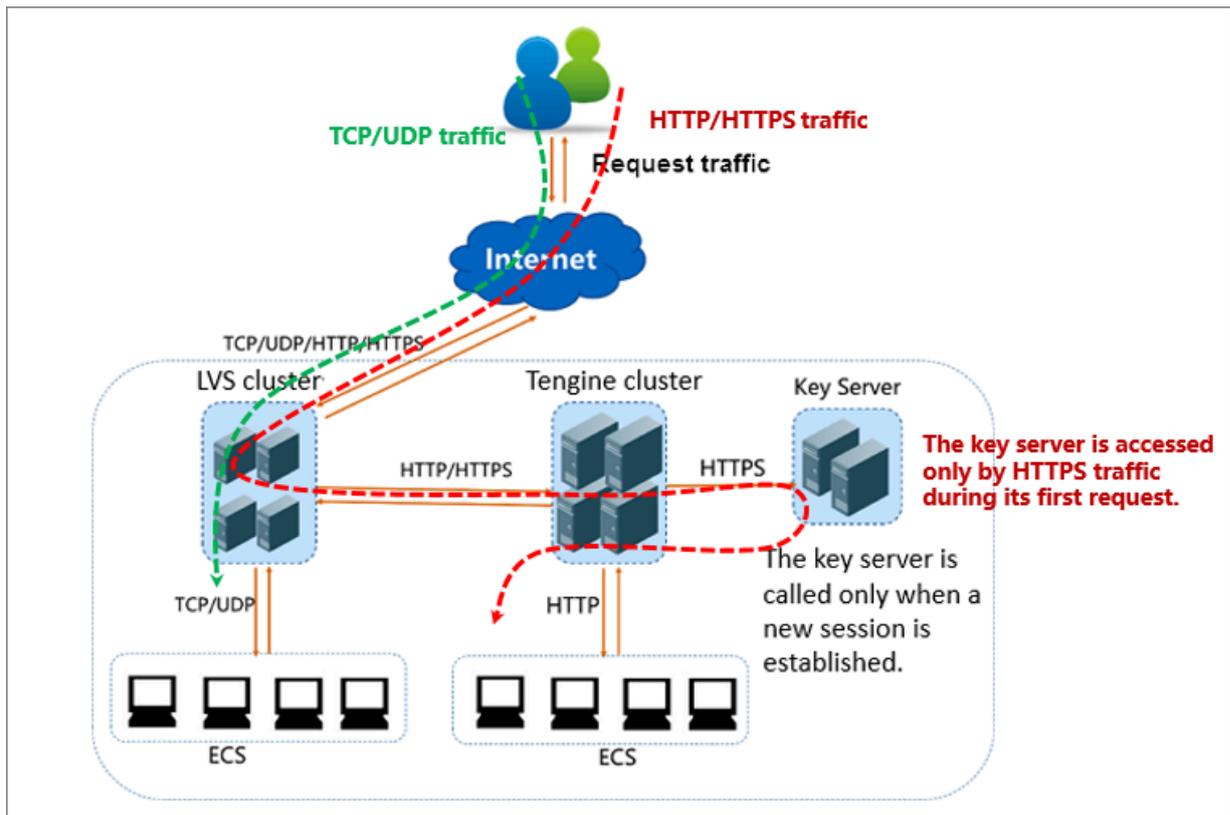
10. Will intranet SLB instances be charged for capacity fee?

If the intranet SLB instance is a shared-performance instance, no capacity fee is charged. If the intranet SLB instance is a guaranteed-performance instance, corresponding capacity fee is charged, and no other fees are charged.

1.1.3 Network traffic flow

Inbound network traffic

SLB distributes incoming traffic according to forwarding rules configured on the console or API. The following figure shows the network traffic flow.

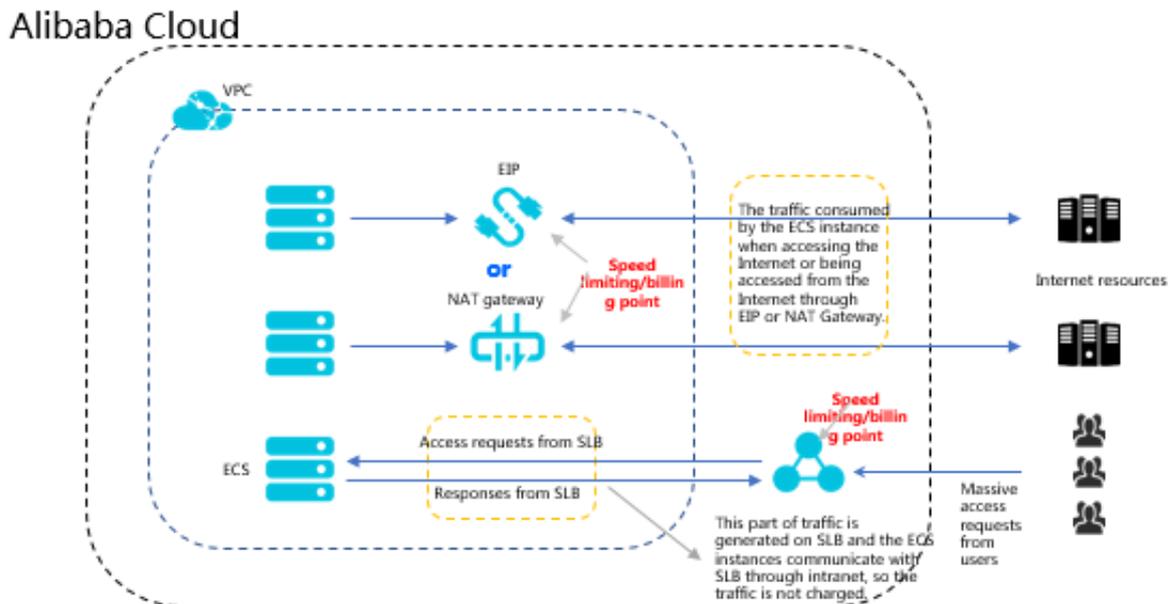


1. Regardless of TCP/UDP protocol or HTTP/HTTPS protocol, the incoming traffic must be forwarded through the LVS cluster first.
2. Numerous inbound traffic is distributed evenly among all node servers in the LVS cluster, and the node servers synchronizes session to guarantee high availability.
3. For 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.
4. 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.
5. For Layer-7 listeners (the frontend protocol is HTTPS), the request distribution is similar to the HTTP protocol. However, before distributing the requests to backend ECS instances, the system will call the Key Server to validate certificates and decrypt data packets.

Outbound network traffic

SLB communicates with backend ECS instances through the intranet. If the backend ECS instances only need to handle the traffic distributed from SLB, no public

bandwidth (EIP, NAT Gateway and public IP) is required. However, if you want to provide external services from a backend ECS instance, or the backend ECS instance needs to access the Internet, you must configure a public IP such as configuring an EIP or a NAT gateway. The following figure shows the outbound network traffic flow.



In general, the traffic goes out from where it comes in:

1. For the traffic coming from SLB, billing and speed limitation are done on SLB. You are charged by the outbound traffic and not the inbound traffic (the rule may change in the future). SLB communicates with the backend ECS instances through the intranet and no traffic fee is not charged for the internal communication.
2. For the traffic coming from the EIP or NAT Gateway, billing and speed limitation are done on EIP or NAT Gateway. If the ECS instance has configured a public IP when it is created, the billing and speed limitation are done on the ECS instance.
3. SLB only provides the function of being accessed from the Internet. That is, a backend ECS instance can only access the Internet when it responds to the request forwarded by SLB. If you want to actively access the Internet from a backend ECS instance, you must configure a public IP (configure EIP or NAT gateway) for the ECS instance.
4. A public IP (configured when you create an ECS instance), EIP, and NAT gateway can all achieve mutual Internet access (access or accessed), but they cannot forward traffic or balance traffic loads.

1.1.4 Create an instance

To use Server Load Balancer, you must first create a Server Load Balancer instance.

Prerequisites

Before creating an SLB instance, make sure that you have properly prepared the environment. For more information, see [Before you begin](#).

Procedure

1. Log on to the [SLB console](#).
2. In the left-side navigation pane, click Server Load Balancer. In the upper-left corner, click Create SLB Instance.
3. Configure the SLB instance according to the following information.

Configurat ion	Description
Region	<p>Select the region where the SLB instance is located.</p> <div style="background-color: #f0f0f0; padding: 5px;">  Note: Make sure that the region of the SLB instance is the same as that of backend ECS instances. </div>
Zone type	<p>Displays the zone type of the selected region. The zone of a cloud product refers to a set of independent infrastructure and is usually represented by Internet data centers (IDCs). Different zones have independent infrastructure (network, power supply, air-conditioning and so on). Therefore, an infrastructure fault in one zone will not affect other zones. A zone belongs to a specific region, however, a single region may have one or more zones. SLB has deployed multi-zone in most regions.</p> <ul style="list-style-type: none"> · Single-zone: The SLB instance is deployed only in one zone. · Multi-zone: The SLB instance is deployed in two zones. By default, the instance in the primary zone is used to distribute traffic. If the primary zone is faulty, the instance in the backup zone will automatically take 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 backup zone for the SLB instance. The backup zone only takes over traffic when the primary zone is unavailable.

Configuration	Description
Instance Spec	Select a performance specification for the instance. The performance metrics vary by specification. For more information, see Guaranteed-performance instance FAQ .
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 What is an SLB instance . <ul style="list-style-type: none"> · Internet: An Internet SLB instance only provides a public IP and you can access the SLB service from the Internet. · Intranet: An intranet SLB instance only provides a private IP and you can only access the SLB service from the intranet.
IP version	Select IPv4.
Network type	If the selected instance type is intranet, you have to select a network type for the instance. <ul style="list-style-type: none"> · Classic network: The IP of the instance is allocated and managed by Alibaba Cloud in a unified manner. · VPC: The IP of the instance is allocated from the VSwitch CIDR block specified by you.
Billing	Select a billing method.
Quantity	Select the number of instances to create.

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

1.1.5 Manage an SLB instance

On the **Instances** page, select a region and then you can view all the created SLB instances in the selected region. Additionally, you can:

- **Modify the name of an SLB instance**

Hover the mouse cursor to the instance ID, click the displayed pencil icon and enter the instance name.

- **Stop an SLB instance**

Select a running SLB instance, click **Stop** at the bottom of the page, or click **More > Stop**.

- **Start an SLB instance**

Select a stopped SLB instance, click **Start** at the bottom of the page, or click **More > Start**.

- **Release an SLB instance**

Select an SLB instance, and then click **Release** at the bottom of the page, or click **More > Release**. In the Release dialog box, choose whether to release the instance immediately or release the instance at a specified time.

- **Set a tag**

You can categorize and manage instances in a unified manner through tags. For more information, see [Manage tags](#).

- **Change the instance configuration**

Click **More > Change Configuration** to change the instance type.

- **View instance details**

Click the instance ID or click **Manage** to view the detailed information of the SLB instance.

- On the details page, you can click **Billing Details** to view the detailed charges of the selected SLB instance.

Basic Information	
Server Load Balancer ID: lb-1udxjes24cant50g6n7z4	Status: ● Running
Server Load Balancer Name: qqg	Region: China East 1 (Hangzhou)
Instance IP Type: Public IP	Zone: China East 1 Zone F(Master)/China East 1 Zone E(Slave)
Network Type: Classic Network	
Billing Information	
Billing Method: Pay by Traffic	Created At: 2018-06-07 16:28:18
Instance IP Address: 47.97.92.245(Public IP)	Automatic Release Time: -

- Click **Listeners** to create and view listeners. For more information, see [Listener overview](#).
- Click **Servers** to add and view backend servers. For more information, see [Backend servers](#).
- Click **Monitor** to view the monitor information and set the alarm. For more information, see [#unique_17](#).

1.1.6 Change the configuration

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

Context

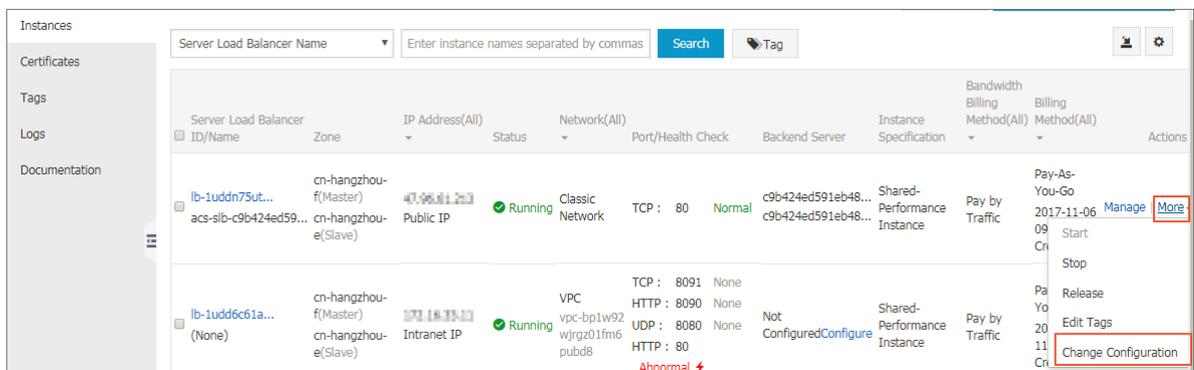
Before modifying the instance configuration, note the following:

- If 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 specification when the business is not busy.
- Once a shared-performance instance is changed to a guaranteed-performance instance, it cannot be changed back.

You can use the (slb.s1.small) capacity instead after changing to the guaranteed-performance instance.

Procedure

1. Log on to the [SLB console](#).
2. Select a region.
3. Find the target instance, and then click More > Change Configuration.



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

1.2 Listener

1.2.1 Listeners overview

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

As shown in the following figure, the listener contains listener configuration and health check configuration.

Listener configuration

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 business needs:

Protocol	Description	Scenarios
TCP	<ul style="list-style-type: none"> · A connection-oriented protocol. A reliable connection must be established with the peer end before data can be sent and received. · Source address-based session persistence. · The source address is available at the network layer. · Fast data transmission. 	<ul style="list-style-type: none"> · Applicable to scenarios with high requirements on reliability and data accuracy, but with tolerance for low speeds, such as file transmission, sending or receiving e-mails, and remote logon. · Web applications without special requirements.

Protocol	Description	Scenarios
UDP	<ul style="list-style-type: none"> • A non-connection-oriented protocol. Before sending data, UDP directly performs data packet transmission instead of making three handshakes with the other party. It does not provide error recovery and data retransmission. • Fast data transmission, however, the reliability is relatively low. 	Applicable to scenarios with preference of real-time content over reliability, such as video chats and pushes of real-time financial quotations.
HTTP	<ul style="list-style-type: none"> • An application layer protocol mainly used to package data. • Cookie-based session persistence. • Use X-Forward-For to obtain the source IP address. 	Applicable to applications that need to recognize data content, such as web applications and small-sized mobile games.
HTTPS	<ul style="list-style-type: none"> • Similar to HTTP, but with an encrypted connection that prevents unauthorized access. • Unified certificate management service. Users can upload certificates to the Server Load Balancer and the decryption operations are completed directly on the Server Load Balancer. 	Applications requiring encrypted transmission



Note:

HTTP/2 and WSS/WS protocols are available in all regions now. For more information, see [HTTP/2 FAQ](#) and [WS/WSS FAQ](#).

Health check configuration

Server Load Balancer provides health checks on the backend servers to improve service availability. For more details, see [Health check overview](#) and [Configure health check](#).

Add Listener
✕

1.Listener Configuration
2.Health Check Configuration
3.Success

Health Check Mode: ?

TCP HTTP

Health Check Port:

If no port number is specified, the backend server port will be used for health checks by default.

▾ [Collapse](#)

[Advanced](#)

[Options](#)

Response Timeout Duration:* Second(s)

Max timeout for each health check request. Enter a value from 1-300 seconds, and the default value is 5 seconds.

Health Check Interval:* Second(s)

Interval between health checks. Enter a value from 1-50 seconds, and the default value is 2 seconds.

Unhealthy Threshold:* 2 3 4 5 6 7 8 9 10

The number of consecutive health check failures on the ECS servers (from success to failure).

Healthy Threshold:* 2 3 4 5 6 7 8 9 10

The number of consecutive health check successes on the ECS servers (from failure to success).

Previous Step
Confirm
Cancel

1.2.2 Layer-4 listeners

1.2.3 Layer-7 listeners

1.2.4 Health check

1.2.5 Shared instance bandwidth

Load Balancing supports the total bandwidth of all listening shared instances under a load balancing instance that is priced by bandwidth. When you create a monitor, you can set the bandwidth peak or you can choose not to set it.

- **Configuration:** You can limit the bandwidth of the listen, however, the sum of all listen bandwidth peaks cannot exceed the instance's bandwidth peaks.
- **No limit:** if you do not limit bandwidth, listen for shared instance bandwidth under the instance.

How do I share bandwidth?

If you purchased a load balancing instance with a bandwidth peak of 10 MB, and created three listeners under this instance (Listening A, listening B, listening C). The peak bandwidth of listen a is set 4 MB, two other listeners do not have the bandwidth peak set. The bandwidth usage of the three listeners can occur in the following situations:

- If listening to a and listening to C has never been out of traffic, then listening B can run up to the remaining 6 MB bandwidth (10 MB-4 MB).
- If listening for C has never been out of traffic, and listening for B has been very large, the remaining 6 MB bandwidth is exceeded. At this point, listening for B has generated drops, and listening for a only 4 MB. The outflow of does not exceed the set bandwidth peak, so no drops are generated.
- If listen a is always running at full speed (Listening peak 4 MB), then listening for B and listening for C also have traffic and both listen for very large amounts of traffic, so listening for B and listening for C will share the remaining 6 MB. Bandwidth. At this point, listening for a traffic will not be affected by listening for B and listening for C, always up to 4 MB Reserved Peak; if listening B is the same size as listening C out of traffic, the bandwidth used by the two listeners goes closer to the same point.

Therefore, the limit to listen bandwidth is Resource Reservation, this is to ensure that the core business always has enough bandwidth. Non-core business can not set listen bandwidth values, they compete for the remaining bandwidth resources of the instance.

1.3 Backend servers

1.3.1 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.

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

SLB service virtualizes the added ECS instances in the same region into an application pool featured with high performance and high availability. By default, the backend servers are maintained in the instance, that is, all the listeners under the SLB instance can only forward the traffic to the same port of the same server.

You can also add ECS instances in the way of server groups. 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 server group is configured for a listener, the listener will forward requests to the ECS instances in the selected server group instead of the ECS instances in the backend server pool.

Master-slave server group

If you have the traditional master-backup requirements, where one backend server is used as the master server and the other is used as the slave server. When the master server works normally, requests are distributed to it; when the master server is down, the requests will be distributed to the slave server. To avoid service interruption, you can create a master-slave server group.

No health check is performed on the slave server. When the master server fails, the traffic is directly distributed to the slave sever. When the health check of the master server succeeds, the traffic will be automatically distributed to it.

The master-slave server group is available only for Layer-4 listeners. For more information, see [Create master-slave 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. For more information, see [Create VServer groups](#).

Notes

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

- SLB does not support cross-region deployment. Make sure that the region for the ECS instances and the SLB instance is the same.
- 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 the ECS. The front-end port of the listener is the application port opened on the ECS instance.
- You can specify a weight for each ECS instance. An ECS instance with the higher weight receives more requests. Set the weight based on service capacity and status of the backend ECS instance.



Note:

If you have enabled the session persistence function, the requests distributed to the backend ECS 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 as configured among the backend servers, 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. (If session persistence is enabled, you need to strip the access logs for the same IP address from the log. If the weight is configured for SLB, you need to calculate whether the percentage of access traffic recorded in the logs matches the weight ratio.)

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

1.3.2 Add default servers

Before using the SLB service, you must add at least one default server.

Prerequisites

- You have [created an SLB instance](#).
- You have created ECS instances and deploy applications to process distributed requests.

Procedure

1. Log on to the [SLB console](#).
2. On the Server Load Balancer page, select a region.
3. Click the ID of the target SLB instance.
4. Click the Default Server Group tab.
5. Click Add.
6. On the Servers Not Added page, click Add.
Then the Available Servers page is displayed.
7. Click Add next to the target ECS instance, or select multiple ECS instances and then click Add to Selected Server List.
8. In the displayed Available Servers dialog box, specify the weight of the added ECS instance and then click OK.

An ECS instance with a higher weight will receive a larger number of connection requests. You can set the weight based on the service capabilities of the ECS instances.



Notice:

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

The added ECS instances are listed on the Default Server Group page. You can remove or change the weights of the added ECS instances.

1.3.3 Create a VServer group

A virtual server group (VServer group) is a group of ECS instances. VServer groups allow you to manage and customize backend servers in the listener dimension. It allows listeners in a Server Load Balancer instance to use different backend servers. Then, different requests can be distributed to different backend servers.

Prerequisites

- You have [created an SLB instance](#).
- You have created an ECS instance for receiving the forwarded requests.

Context

If you use a VServer group when configuring a listener, the listener distributes requests to the associated VServer group. The listener no longer distribute the requests to ECS instances in the backend server pool.

For a Layer-7 listener, if you add backend servers in the server pool, configure a VServer group, and adding a forwarding rule at the same time, the requests are distributed in the following order:

- If the client requests match the configured domain name forwarding rule, the requests are distributed to the VServer group associated with the rule.
- If not, the requests are distributed to the VServer group associated with the listener.
- If no VServer group is configured for the listener, the requests are distributed to the ECS instances in the backend server pool.

When using the VServer group, note the following limitations:

- Only backend servers in the same region as listeners can be added to a VServer group.
- One ECS instance can be added to multiple VServer groups.
- One VServer group can be associated with multiple listeners.
- The VServer group consists of multiple ECS instances with different port numbers.

Procedure

1. Log on to the [SLB console](#).
2. On the Instances page, select a region.
3. Click the ID of an SLB instance.

4. In the left-side navigation pane, click **Server > VServer Group**.
5. On the **VServer Group** page, click **Create VServer Groups**.
6. In the **Create VServer Group** dialog box, complete these steps:
 - a. Enter a group name in the **Group Name** field.
 - b. Select the network type for the ECS instance you want to add.
7. In the **Available Servers**, select the ECS instances to add.

The selected instances are displayed in the **Selected Servers** list.

8. In the **Selected Servers** list, enter the port number and weight for each added ECS instance, and then click **Confirm**.

The created VServer group is displayed on the **VServer Group** page. You can delete or add ECS instances for the VServer group (click **Edit**). You can also associate this VServer Group with the instance's listeners or forwarding rules.

Group Name	Group ID	Actions
111	rsp-1udn56vcltaj3	View Edit Delete

1.3.4 Create a master-slave server group

If you have traditional active/standby requirements, where one backend server is used as the master server and the other is used as the slave server, create a master-slave server group. When the master server works normally, requests are distributed to it; when the master server is down, the requests will be distributed to the slave server to avoid service interruption.

Prerequisites

- You have [created an SLB instance](#).
- You have created ECS instances and deploy applications to process distributed requests.

Context

After a master-slave server group is configured for a listener, the listener will forward requests to the ECS instances in the server group instead of the ECS instances in the backend server pool.

**Notice:**

Only Layer-4 listeners (TCP and UDP protocols) support configuring master-slave server groups.

Procedure

1. Log on to the [SLB console](#).
2. On the Instances page, select a region.
3. Click the ID of an SLB instance.
4. In the left-side navigation pane, click **Server > Master-Slave Server Groups**.
5. On the Master-Slave Server Group page, click **Master-Slave Server Group**.
6. In the Master-Slave Server Group dialog box, complete these steps:
 - a) Enter a group name in the **Group Name** field.
 - b) Select the network type for the ECS instance you want to add.
 - c) In the **Available Servers**, select the ECS instances to add.
 - d) In the **Selected Servers** list, enter the port number and weight for each added ECS instance, and then select one ECS instance as the slave server. Click **Confirm**.

1.4 Certificate management

1.4.1 Certificate requirements

Server Load Balancer only supports certificates in the PEM format. The certificate, certificate chain, and private key must conform to the rules described in this section.

Certificates issued by a root CA

If the certificate is issued by a root CA, the received certificate is the only one that is required to upload to Server Load Balancer. The website that is configured with the certificate will be trusted by the web browser without configuring additional certificates.

The certificate format must meet the following requirements:

- The certificate content is placed between `----- BEGIN CERTIFICAT E -----`, `----- END CERTIFICAT E -----`. Include the header and footer when uploading the certificate.

e format when issuing the certificate, the certificate chain must conform to the format requirements.

The following is a sample certificate chain.

```
----- BEGIN   CERTIFICAT E -----
----- END   CERTIFICAT E -----
----- BEGIN   CERTIFICAT E -----
----- END   CERTIFICAT E -----
----- BEGIN   CERTIFICAT E -----
----- END   CERTIFICAT E -----
```

RSA private key

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

The RSA private key format must meet the following requirements:

- The key is placed between `----- BEGIN RSA PRIVATE KEY -----, -----`
`END RSA PRIVATE KEY -----`. Include the header and footer when uploading the key.
- Space is not allowed in the content. Each line except the last 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 header and footer are `----- BEGIN`
`PRIVATE KEY -----, ----- END PRIVATE KEY -----` or `----- BEGIN`
`ENCRYPTED PRIVATE KEY -----, ----- END ENCRYPTED PRIVATE`
`KEY -----`, or the private key contains `Proc - Type : 4 , ENCRYPTED` , run the following command to convert the private key before uploading it to Server Load Balancer:

```
openssl  rsa - in  old_server _key . pem - out  new_server
_key . pem
```

The following is a sample RSA private key.

```

-----BEGIN RSA PRIVATE KEY-----
MIIEpAIBAAKCAQEAzVzSSSChH67bmt8mFykAxQ1tKCYukwBiWZwk0StFEbTWHy8K
tTHSFD1u9TL6qycrHEG7cjYD4DK+kVIHU/Of/pUWj9LLnrE3W34DaVzQdKA00I3A
Xw9SgrqFJMjclVa2khNKA1+tNPSCPJoo9DDrP7wx7cQx7LbMb0dfZ8858KIoluzJ
/FD0XXyuWoqaIePZtK9Qjn957ZEPhtUpVZuhS3409DDM/tJ3Tl8aaNYWhrPBc0
jNcz0Z6XQGf1rZG/Ve520GX6rb5dUYpdcfXzN5WM6xYg8a1L7UHDHHPi4AYsatdG
z5TMPnmE8yZPUYudTLxgMVAovJr09Dq+5Dm3QIDAQABAoIBAGL68Z/nnFyRHRFi
laF6+Wen8ZvNqkm0hAMQwIjH1Vp1f174//8Qyca/EvUtuJHyB6T/2PZQoNVhxe35
cgQ93Tx424WGpCwUshSfxewfbAYGf3ur8W0xq0uU07BAxaKHNcmNG7dGyo1UowRu
S+yXLrpVzH1YkuH8TT53udd6TeTWi77r8dkGi9KSAZ0pRa19B7t+CHKIzm6ybs/2
06W/zHZ4YAxwkTYLKGHjoiEYs111ah1AJvICVgTc3+LzG2pIpM7I+K0nHC5eswvM
i5x9h/OT/ujZsyX9P0PaAyE2bqy0t080tGexM076Ssv0KVhKfVwjLUhF6WcqFCD
xqhhxkECgYEA+PftNb6eyXl+/Y/U8NM2fg3+rSCms0j9Bg+9+yZzF5GhgqHu0edU
ZXIhrJ9u6B1XE1arpijVs/WHmFhYSTm6Dbdd7S1tLy0BY4cPTRhziFTKt8AkIXMK
605u0UiWsq0Z8hn1X141ox2cW9ZQa/HC9udeyQotP4NsMJWgpBV7tC0CgYEAwvNf
0f+/jUjt0HoyxCh4SIAqk4U0o4+hBCQbWcXv5qCz4mRyTaWzFEG8/AR3Md2rhmZi
GnJ5dfde7uY+JsQfX2Q5JjwTad1BW41ed0Sa/uKRao4UzVgnYp2aJKxtuWffvVbU
+kf728ZJRA6azSLvGmA8hu/GL6bgfU3fkSkw03ECgYBpYK7TT7JvvnAerMtJf2yS
ICRkbQaB3gPSe/LCgzy1nhtaFOUbNxGeuowLAZR0wrz7X3TZqHEDcYoJ7mK346of
QhGLITyoeHkbYkAUtq038Y04EKh6S/IzMzB0frXiPKg9s8UKQzkU+GSE7ootli+a
R8Xzu835EwxT6BwNN1abpQKbgQC8TialClq1FteXQyGcNdcReLMncUHKIKcP/+xn
R3kV106MZCFAdqirAjiQWapkh9Bxbp2eHCrB81MFAWLRQSl0k79b/jVmtZMC3upd
EJ/iSWjZKPbw7hCFAeRtPhxyNTJ5idEiu9U8EQid8111giPgn0p3sE0HpDI89qZX
aaiMEQKBgQDK2bsnZE9y0ZWhtTeu94vziKmFrSkJMGH8pLaTiliw1iRhRYWJysZ9
BOIDxnrmwiPa9bCtEpK80zq28dq7axpCs9CavQRcv0Bh5Hx0yy23m9hFRzfDeQ7z
NTKh193HHF1joNM81LHFyGRFEWrr0W5gfBudR6USRnR/6iQ11xZXw==
-----END RSA PRIVATE KEY-----

```

1.4.2 Generate CA certificates

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
certificate = / root / ca / private / ca . crt
serial = / root / ca / serial
private_key = / root / ca / private / ca . key
RANDFILE = / root / ca / private /. rand
default_days = 365
default_crl_days = 30
default_md = md5
Unique_subject = No
Policy = policy_any
[ policy_any ]
countryName = match
stateOrProvinceName = match
organizationName = match
organizationalUnitName = match
localityName = optional
commonName = supplied
emailAddresses = 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@izbp1hfvivcqx1jwap3liZ:~/ca/conf# cd /root/ca
root@izbp1hfvivcqx1jwap3liZ:~/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 . cs
```



Note:

Common Name is the domain name of the SLB instance.

```
root@iZbp1hfvivcgx1jwvap31iZ:~/ca# sudo openssl req -new -key private/ca.key -out private/ca.csr
You are about to be asked to enter information that will be incorporated
into your certificate request.
What you are about to enter is what is called a Distinguished Name or a DN.
There are quite a few fields but you can leave some blank
For some fields there will be a default value,
If you enter '.', the field will be left blank.
-----
Country Name (2 letter code) [AU]:CN
State or Province Name (full name) [Some-State]:ZheJiang
Locality Name (eg, city) []:HangZhou
Organization Name (eg, company) [Internet Widgits Pty Ltd]:Alibaba
Organizational Unit Name (eg, section) []:Test
Common Name (e.g. server FQDN or YOUR name) []:mydomain
Email Address []:a@alibaba.com

Please enter the following 'extra' attributes
to be sent with your certificate request
A challenge password []:
An optional company name []:
root@iZbp1hfvivcgx1jwvap31iZ:~/ca#
```

5. Run the following command to generate a `csr` 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 - genctrl - out / root / ca / private / ca  
.crl - crldays 7 - config "/ root / ca / conf / openssl .  
conf "
```

The response is as follows:

```
Using configurat ion from / root / ca / conf / openssl . conf
```

Sign 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 the certificate signature.

```
$ 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@iZbp1hfvivcqxljwvap31iZ:~/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@iZbp1hfvivcqxljwvap31iZ:~/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
```

1.4.3 Convert certificate formats

Server Load Balancer supports PEM certificates only. Certificates in other formats must be converted to PEM 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.

- 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.

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

Convert P7B to PEM

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

Run the following command to convert the certificate format.

```
openssl pkcs7 -print_certs -in incertificate .p7b -out  
outcertificate .cer
```

Convert PFX to PEM

PFX: This format is usually used in Windows Server.

- Run the following command to extract the certificate format.

```
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
```

1.4.4 Upload certificates

Before creating HTTPS listeners, you must upload the server certificate and CA certificate (if required) to SLB. You no longer need to configure certificates on the backend servers after uploading certificates to SLB.

Prerequisites

- You have purchased a server certificate.
- You have generated a CA certificate and client certificate. For more information, see [Generate CA certificates](#).

Context

Note the following before uploading certificates:

- Certificates in SLB are regional resources. If you want to use a certificate in multiple regions, you must upload the certificate to all these regions.
- Up to 100 certificates can be uploaded per account.

Procedure

1. Log on to the [SLB console](#).
2. In the left-hand navigation pane, click Certificates.
3. Click Upload Certificate.
4. On the Upload Certificate page, upload the certificate content and then click Confirm.

Configuration	Description
Certificate Name	<p>Enter a certificate name.</p> <p>The name must be 1-80 characters in length, including letters, numbers and the following special characters:</p> <p>_.-</p>
Certificate Region	<p>Select one or more regions where the certificate is uploaded.</p> <p>The region is where the HTTPS listener is located. Certificates cannot be used across regions.</p>
Certificate Type	<p>Select a certificate type.</p> <ul style="list-style-type: none"> · Server Certificate: For one-way authentication, only server certificate is required. The client uses it to check whether the certificate sent by the server is issued by a trusted center. · CA Certificate: For two-way authentication, a CA certificate is required in addition to a server certificate. The server uses the CA certificate to authenticate the CA signature on the client certificate, as part of the authorization before launching a secure connection.

Configuration	Description
Certificate Content	<p>Paste the certificate content in the editor.</p> <p>Click Import Sample to view the valid certificate formats. Only certificates in the PEM format are supported. For more information, see Certificate requirements.</p>
Private Key	<p>Paste the private key of the server certificate in the editor.</p> <p>Click Import Sample to view the valid certificate formats. For more information, see Certificate requirements.</p> <div style="background-color: #f0f0f0; padding: 5px;">  Notice: A private key is required when uploading a server certificate. </div>

1.4.5 Replace a certificate

You must 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 \(one-way authentication\)](#) and [Add an HTTPS listener \(two-way authentication\)](#).
3. On the Certificates page, find the target certificate, and then click Delete.
4. In the displayed dialog, click Confirm.

1.5 Log management

1.5.1 View operation logs

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

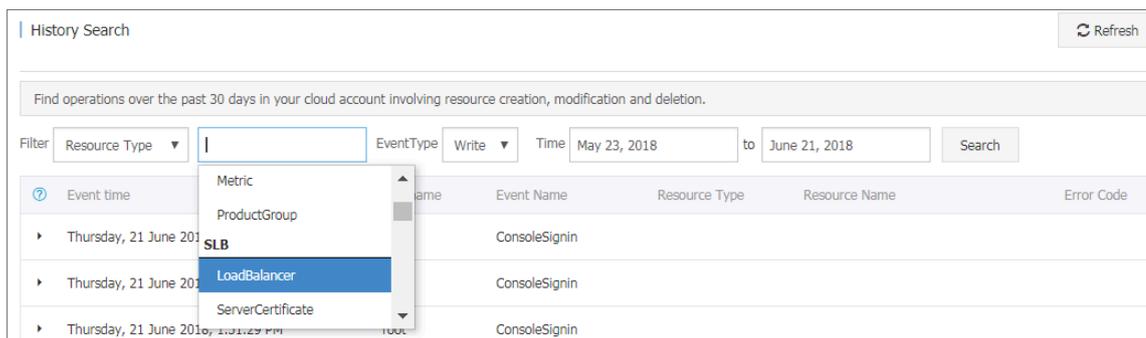
Context

The operation logs are recorded in ActionTrail. ActionTrail records the operations acting upon your Alibaba Cloud resources, you can use these records to analyze your account security, track changes made to your resources, and achieve compliance.

Procedure

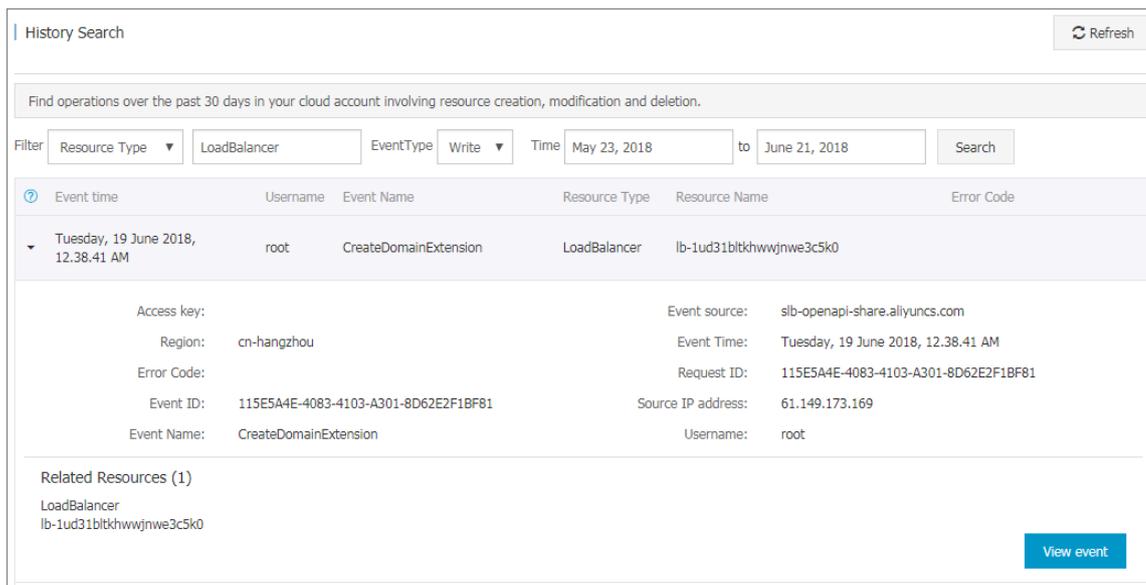
1. Log on to the [SLB console](#).
2. In the left-side navigation pane, click **Logs > Operation Log**.
3. Click **View Operation Logs**.
4. On the **History Search** page, complete these steps to view operation logs:
 - a) Select **Resource Type** as a filter.
 - b) Select the **SLB resource** of which operation logs you want to view.

In this task, LoadBalancer is selected.



- c) Select an event type.
- d) Select the time range to search.
- e) Click **Search** to view logs of operations performed on the selected resource.

Expand the resource to view more detailed information.



1.5.2 Configure access logs

By analyzing the access logs of an SLB instance with Alibaba Cloud Log Service, you can understand the behavior and geographical distribution of client users, troubleshoot problems and so on.

What are access logs

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

After enabling SLB access logging, you can store access logs in the Logstore for analysis. You can also disable access logging at any time.

There is no extra charge for SLB access logs. But corresponding fees are collected when using Log Service. If you store logs in OSS, you can save storage costs.



Note:

Only Layer-7 SLB supports configuring access logs and this function is available in all regions now.

Benefits of SLB access logs

The following are benefits of SLB access logs:

- Simple log processing

Free developers and maintenance staff from tedious and time-consuming log processing so that they can concentrate on business development and technical research.

- Cost-effective

Performance and cost problems must be taken into consideration when processing access logs, because the amount of SLB access logs is very large. Integrated with Log Service, the access log processing is faster and cost-effective than self-build 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 long time to ETL data to Hive at which a lot of work is spent on data integration. Powered by its powerful computing capability, Log Service can process and analyze access logs in seconds.

- Flexible

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

Configure access logs

Before configuring access logs, make sure:

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

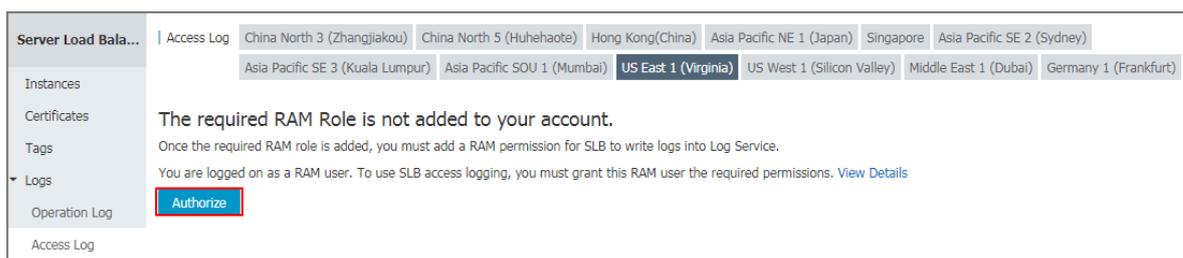
To configure access logs, complete these steps.

1. Log on to the [SLB console](#).
2. In the left-side navigation pane, click **Logs > Access Log**.
3. Click **Authorize**, and then click **Confirm Authorization Policy** to authorize SLB to write logs to Log Service.

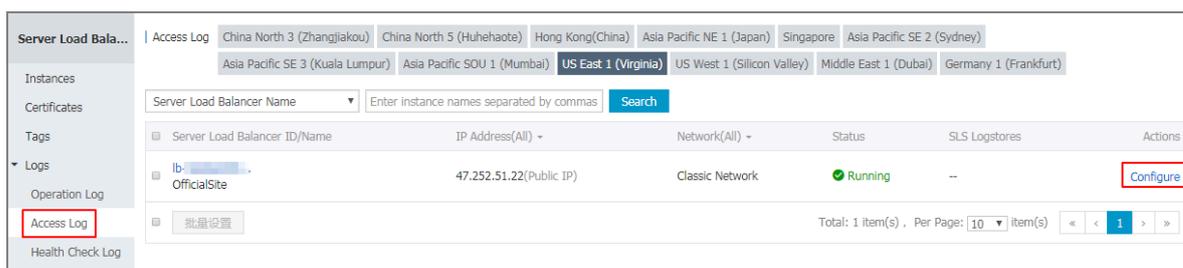


Note:

If you are a RAM user, you must be authorized to use the SLB access logging. For more information, see [Authorize a RAM user to use access logging](#).



4. On the Access Log page, find the target SLB instance and click **Configure**.



5. Select a Log Service project and Logstore, and then click Confirm.

If there is no available Logstore, click Create. Make sure that the name of the project is unique.



Note:

Make sure the Log Service project and the SLB instance are in the same region.

Log Settings ✕

Enable Layer-7 Logging

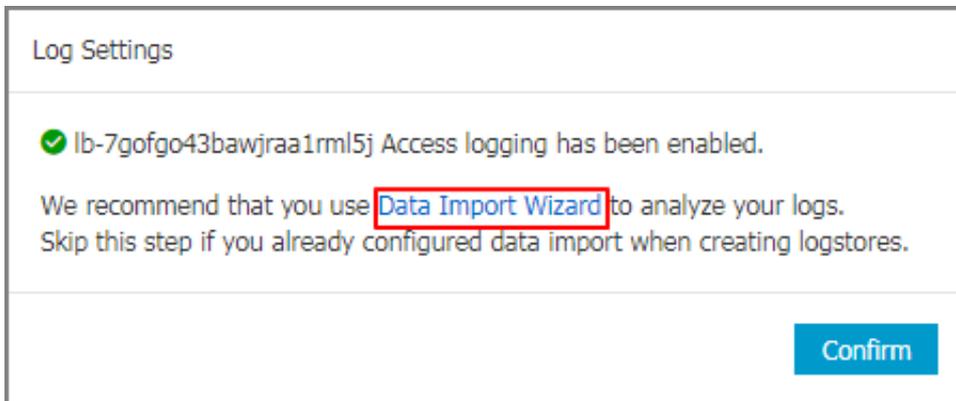
LogProject **Logstore**

Select ▼ ▼

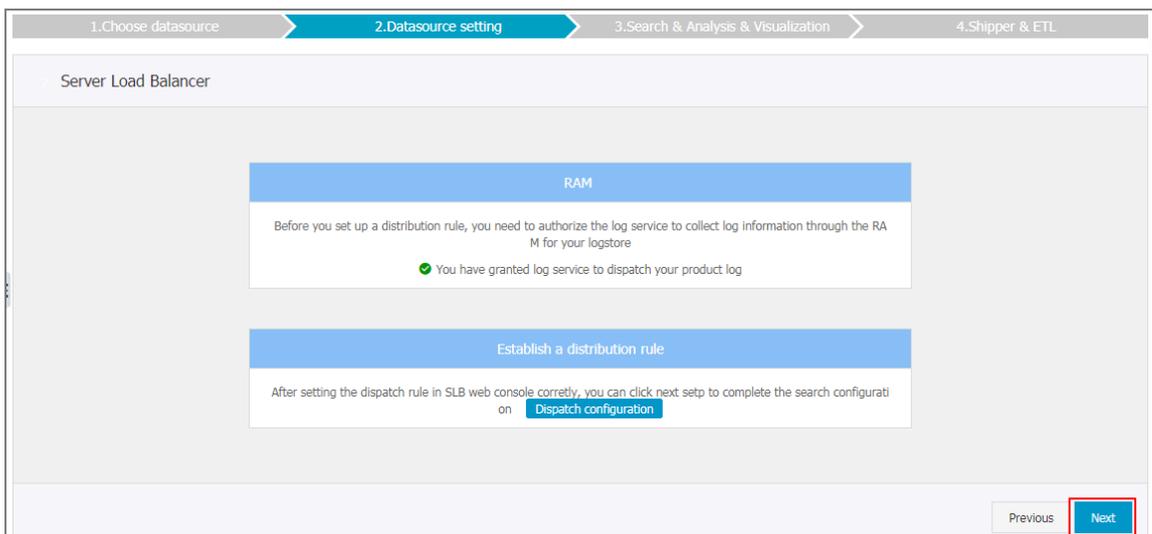
Confirm Close

6. Configure data import.

- a. Click the Data Import Wizard link to configure data import. Or click Confirm and configure data import on the Log Service console later. In this tutorial, the Data Import Wizard link is selected.



- b. Click Next.

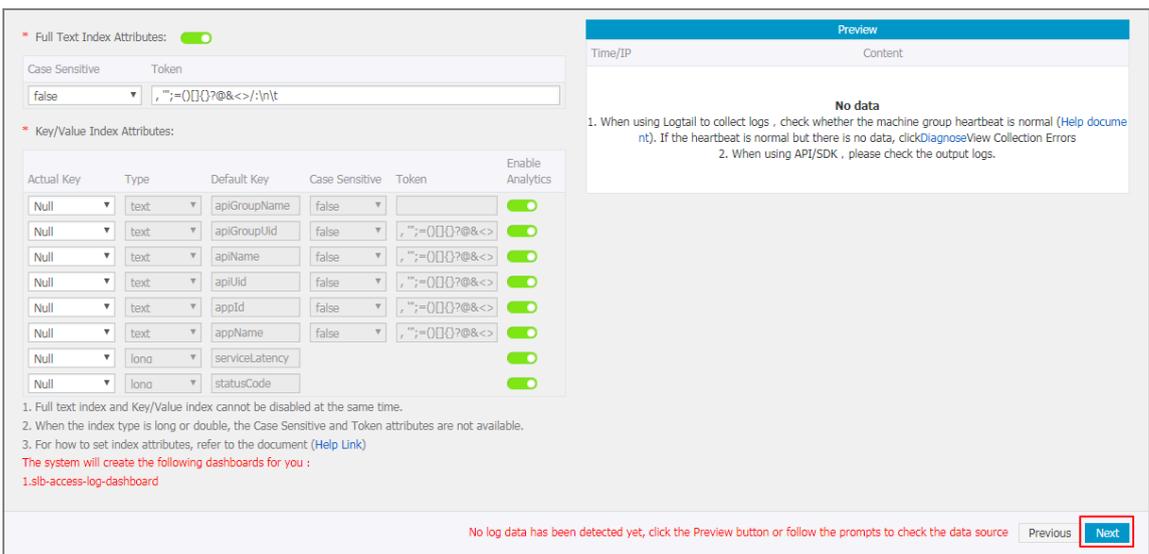


- c. Log Service has pre-configured indexing field for SLB. Click Next.

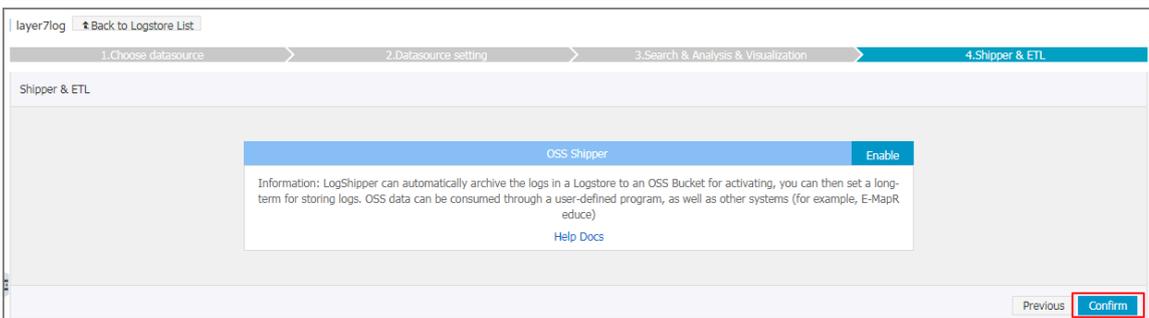


Note:

After enabling indexing search, you will be charged for indexing traffic.



d. Click Confirm to complete data import.



Search and analyze access logs

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

Field	Description
body_bytes_sent	The size of HTTP body (in byte) sent to the client.
client_ip	The client IP.
host	The host header in the request.
http_user_agent	The received http_user_agent header in the request.
request_length	The request length including startline, HTTP header and HTTP body.
request_method	The request method.
request_time	The time interval from the first request received by the SLB to the response sent by the SLB.
request_uri	The received request URI.

Field	Description
slbid	The SLB instance ID.
status	The response status code sent by the SLB.
upstream_addr	The IP address and port number of the backend server.
upstream_response_time	The time from when SLB is ready to send requests to the backend server to when SLB sends response to the client.
upstream_status	Response status code sent from backend servers.

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:

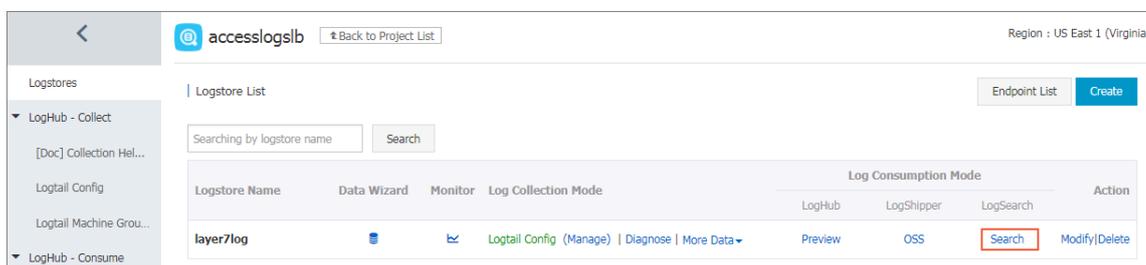
- From the SLB console:

On the Access Log page, click View Logs.

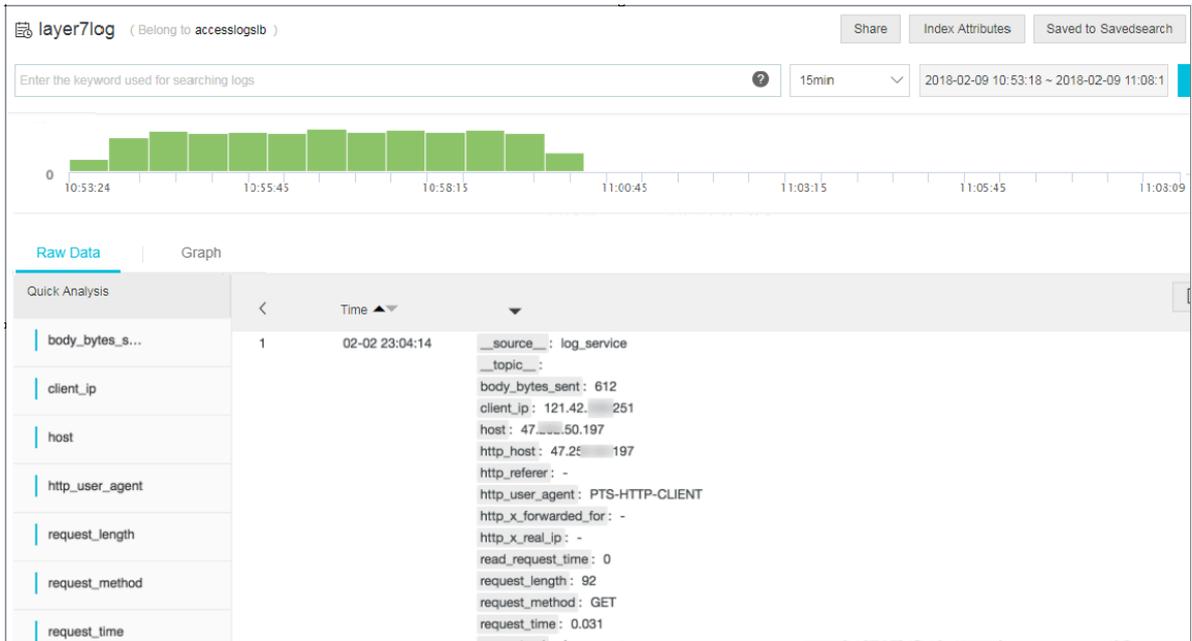


- From the Log Service console:

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



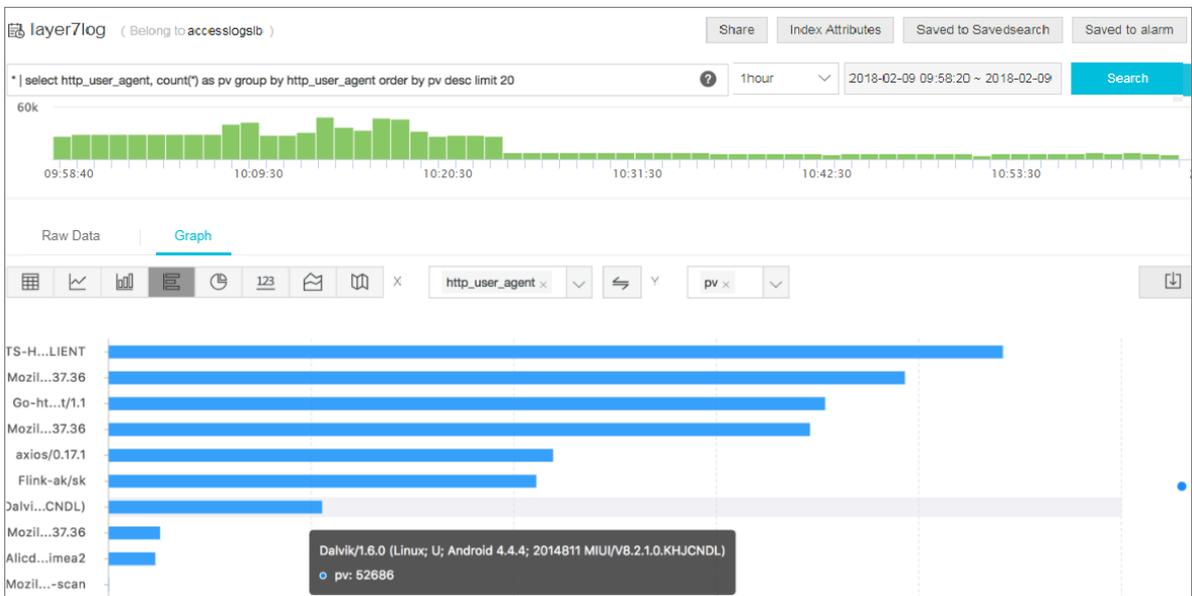
2. Click the corresponding index field to view detailed information.



3. Enter an SQL statement to query.

For example, enter the following SQL statement to query Top 20 clients.

```
* | 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
```



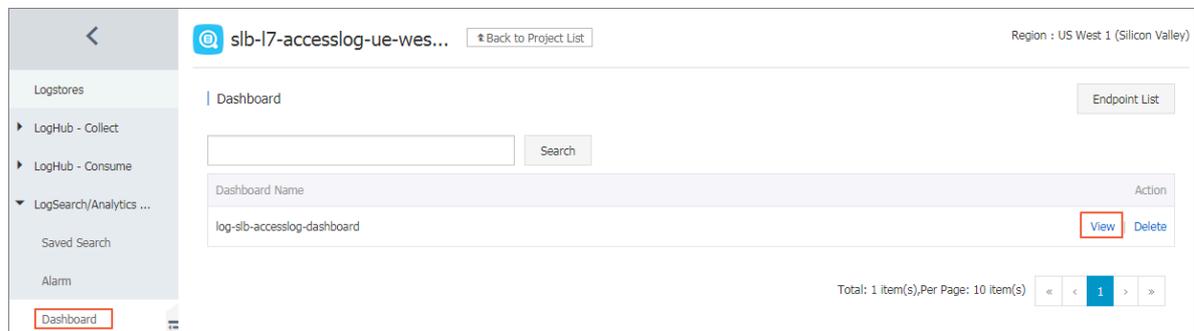
Analyze access logs

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

To analyze access logs, complete these steps:

1. On the Log Service console, click the project name of the target project.
2. In the left-side navigation pane, click Search/Analytics - Query > Dashboard, and then click View.

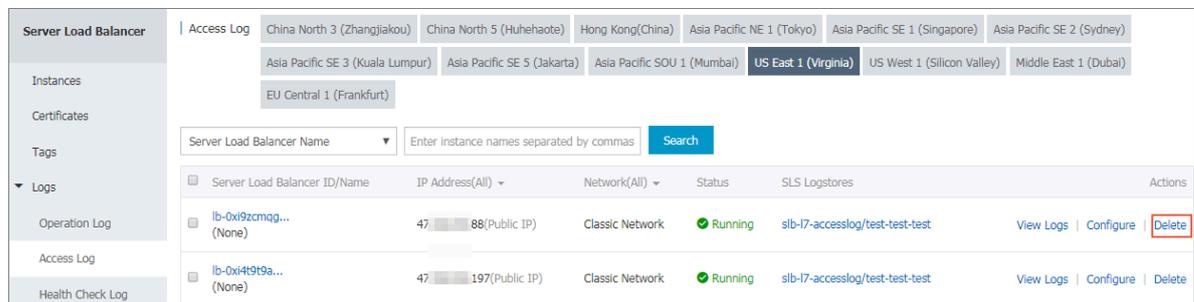
You can view information such as top clients, top hosts, status code and so on.



Disable access logging

To disable access logging, complete these steps:

1. Log on to the [SLB console](#).
2. In the left-side navigation pane, click Logs > Access Log.
3. Find the target instance, and then click Delete.



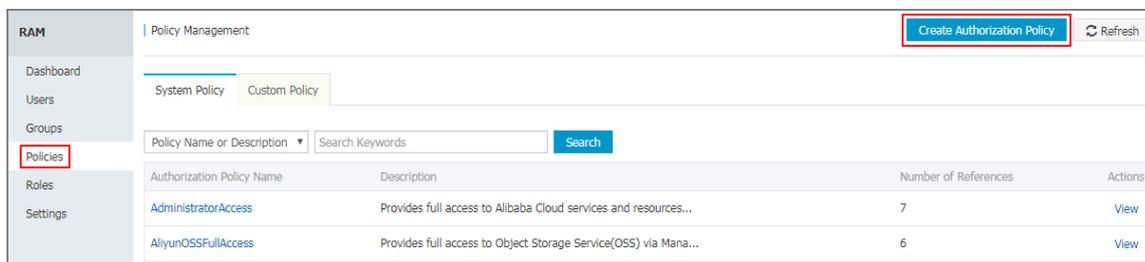
1.5.3 Authorize a RAM user to configure access logs

Before a RAM user starts to use the access logging function, the RAM user must be authorized by the primary account.

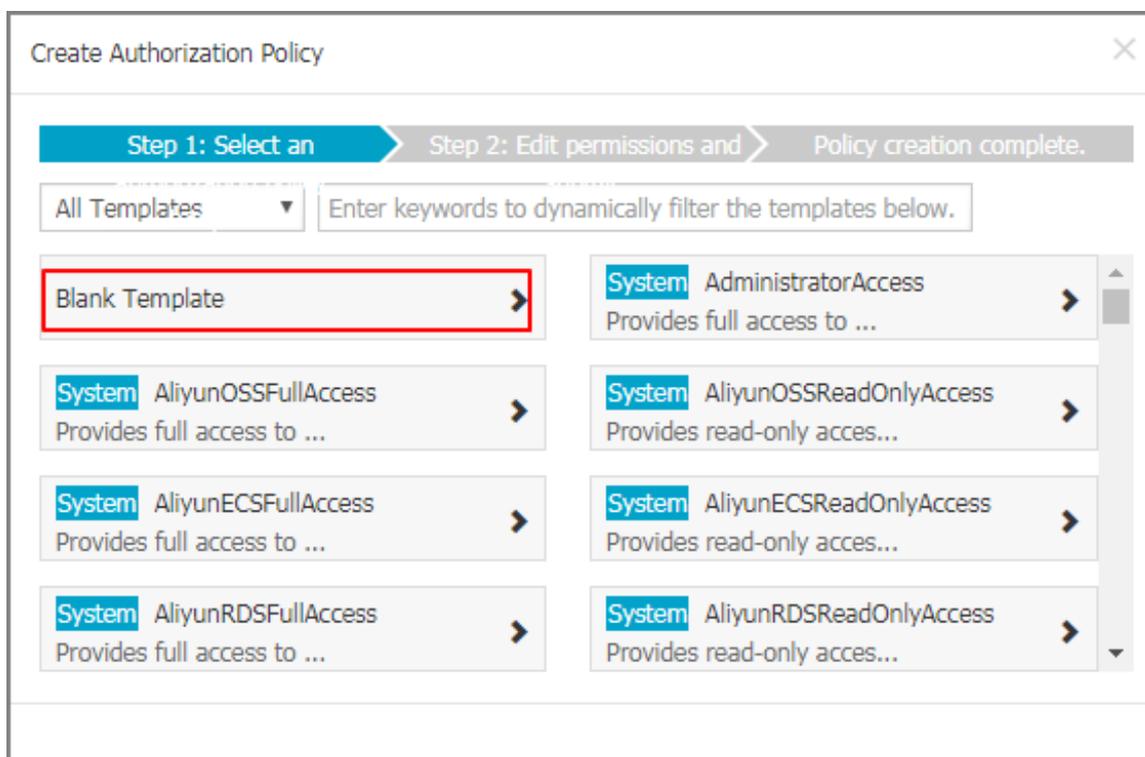
Procedure

1. Create an authorization policy:

- a) Use the primary account to log on to the RAM console.
- 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 *"
    ]
  }
}
```

```

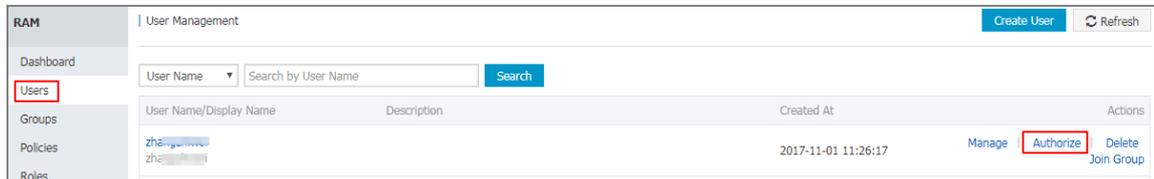
    " log : List *"
  ],
  " Effect ": " Allow ",
  " Resource ": " acs : log ::*: project /*"
},
{
  " Action ": [
    " log : Create *",
    " log : List *",
    " log : Get *",
    " log : Update *"
  ],
  " Effect ": " Allow ",
  " Resource ": " acs : log ::*: project /*/ logstore /*"
},
{
  " Action ": [
    " 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 "
}

```

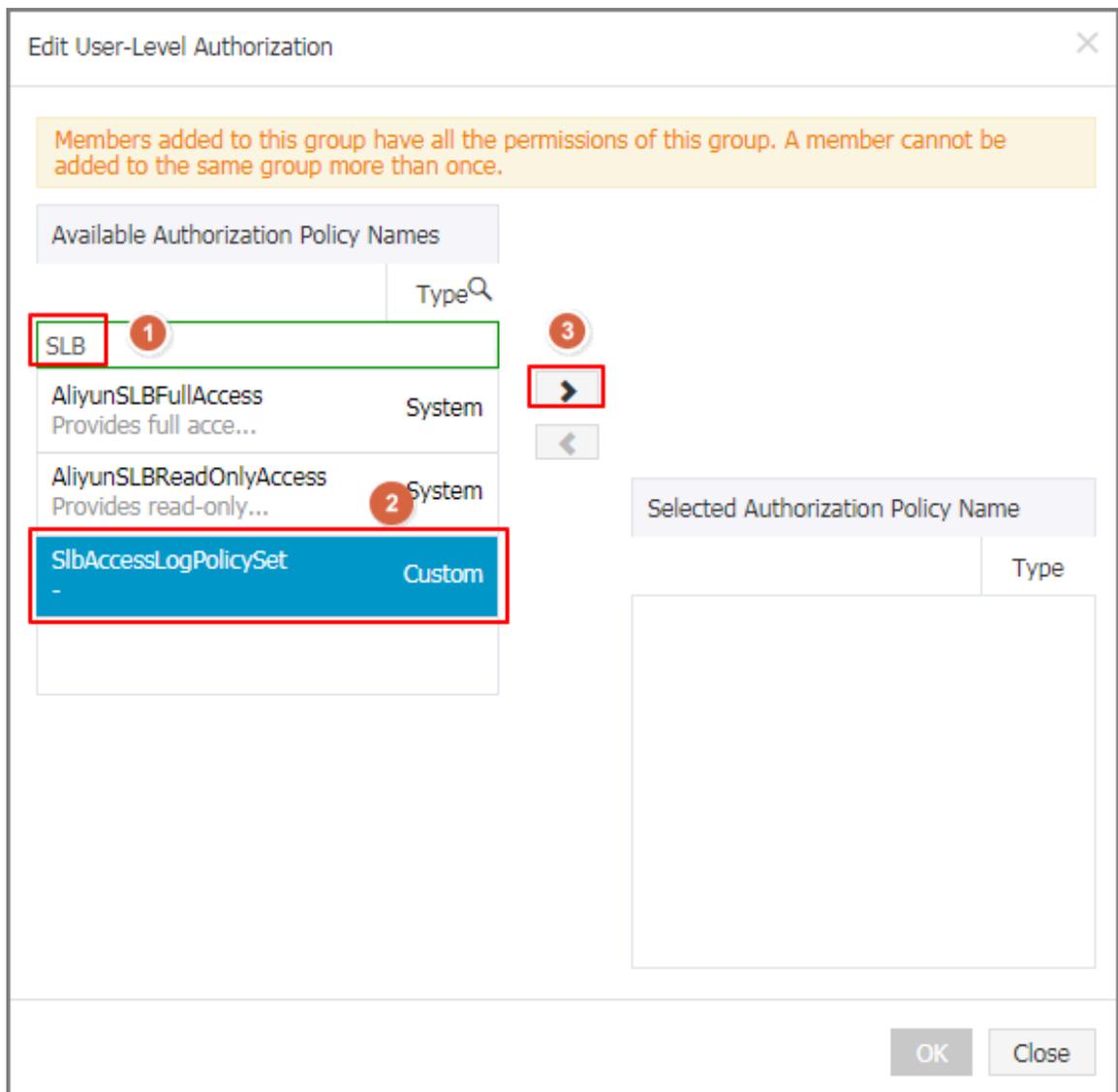
a) Click Close.

2. Attach the created policy to a RAM user

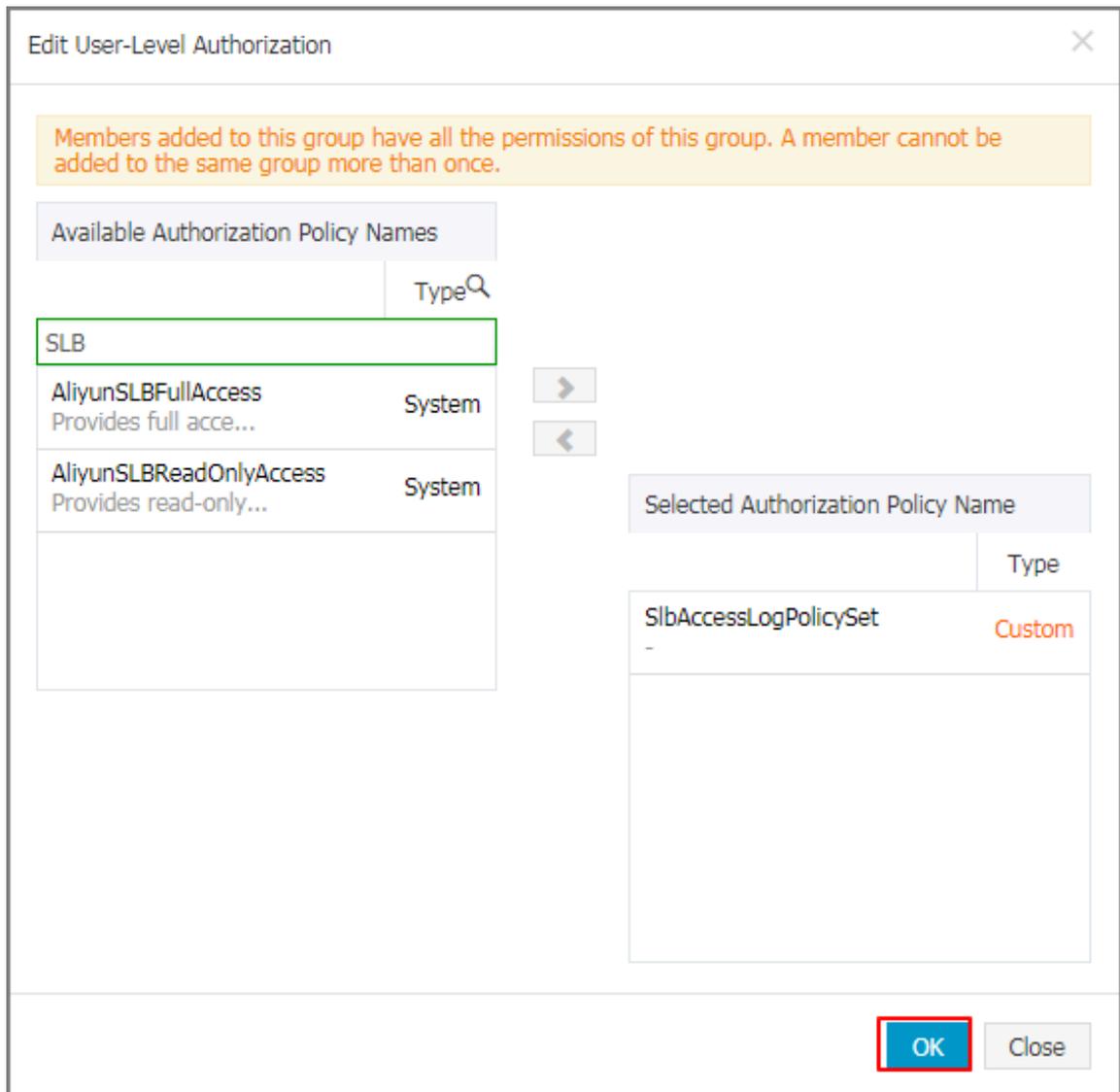
- a) In the left-side navigation pane, click Users.
- b) Find the target user (the user who uses the SLB Access Log function) and click **Authorize**.



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



- d) Click OK.



e) Go to the User Authorization Policies page to check if the policy has been attached to target RAM user.



1.5.4 Manage health check logs

You can view the health logs within three days on Health Check Log page. If you want to get the health check logs three days before or longer, you can store the health check logs to OSS. Therefore, you can download complete health check logs.

Store health check logs

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

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 the health check logs. The health logs are generated hourly 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 in a day are named after the time when they are generated. For example, the log file that is generated between 00:00-01:00, the file name will be `01.txt` and the log file that is generated between 01:00-02:00, the file name will be `02.txt`.



Note:

The health check logs are generated only when the backend server is abnormal. If no failures occur for all the backend servers in an hour, no health check logs are generated in that hour.

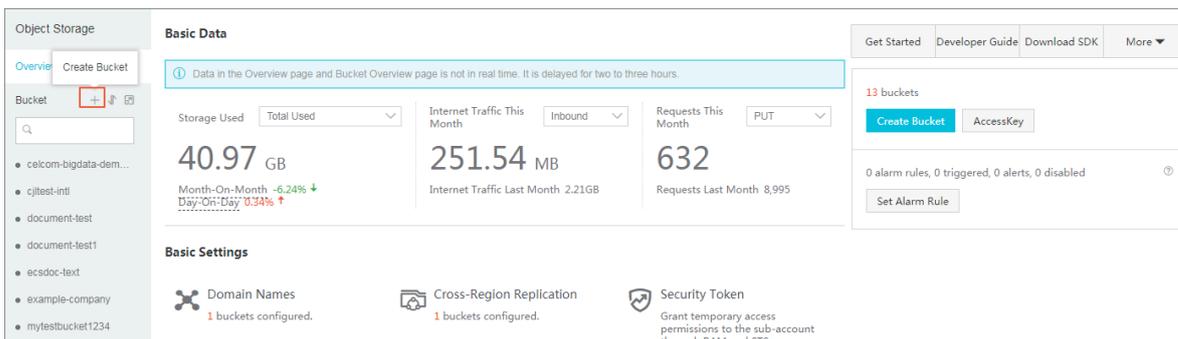
To store health check logs, complete these steps:

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.

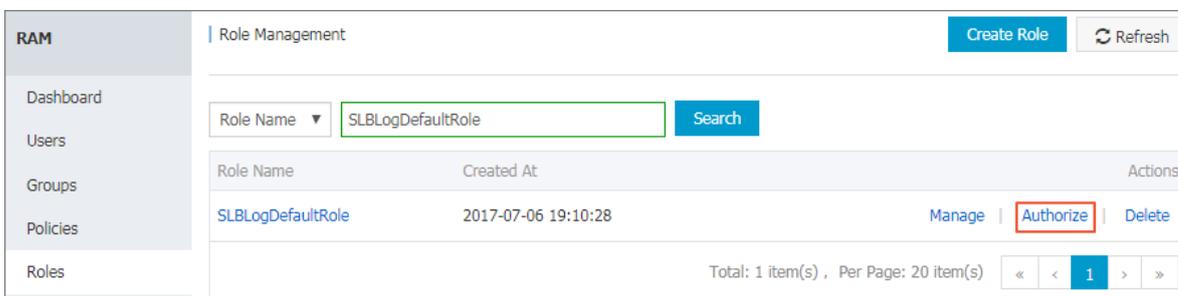
 **Note:**
 Make sure that the region of the bucket and the SLB instance are the same.

Step 2 Authorize SLB to access OSS

After creating a bucket, you have to authorize the log role (SLBLogDefaultRole) to access OSS resources.

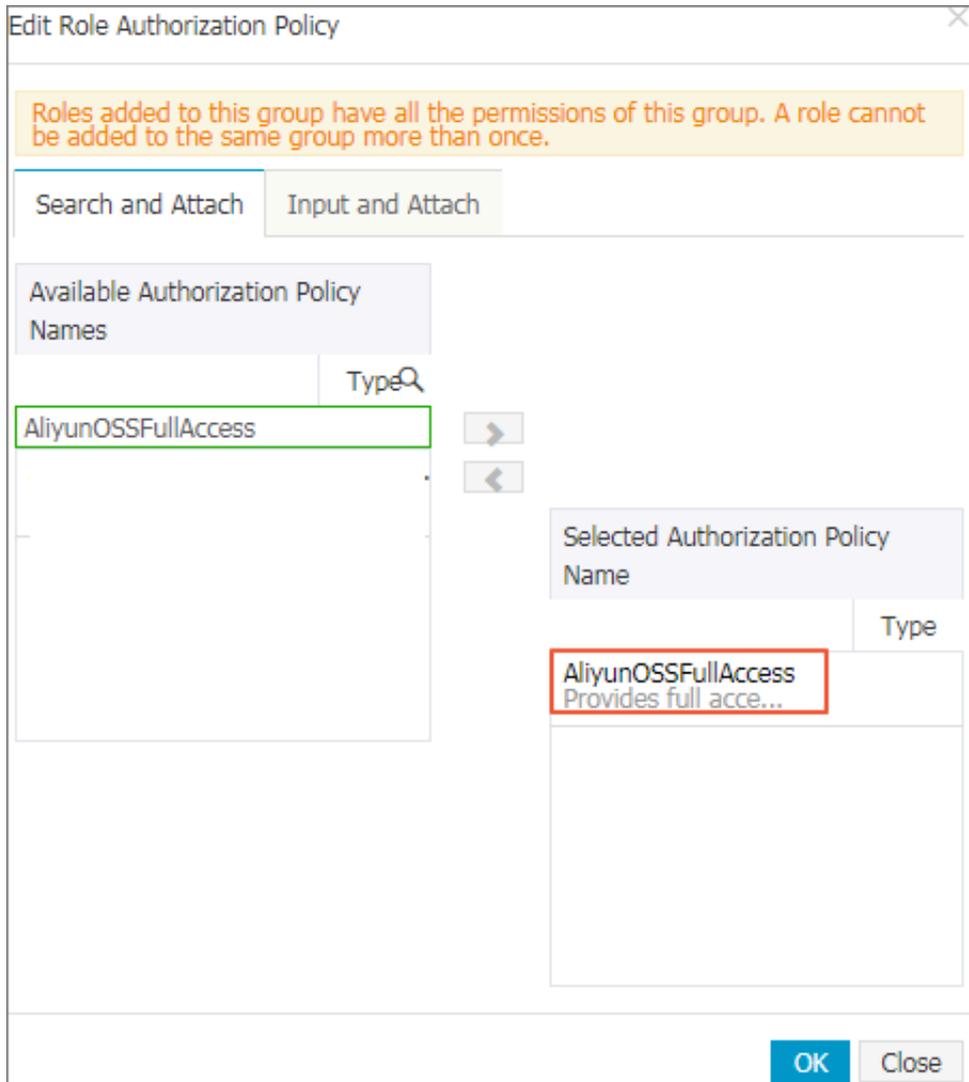
 **Notice:**
 The authorization is required only for first configuration.

1. On the SLB console, click Logs > Health Check Log.
2. Click 1. Activate OSS. if OSS has not been activated yet.
3. On the Health Check Log page, click Add Role Now in the 2. Add the RAM role to your account. 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 and find the role named SLBLogDefaultRole, and then click Authorize.

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



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

SLBLogDefaultRole Edit Authorization Policy			
Authorization Policy Name	Description	Type	Actions
AliyunOSSFullAccess	Provides full access to Object Storage Service(OSS) via Management Console.	System	View Permissions Revoke Authorization

Step 3 Configure log storage

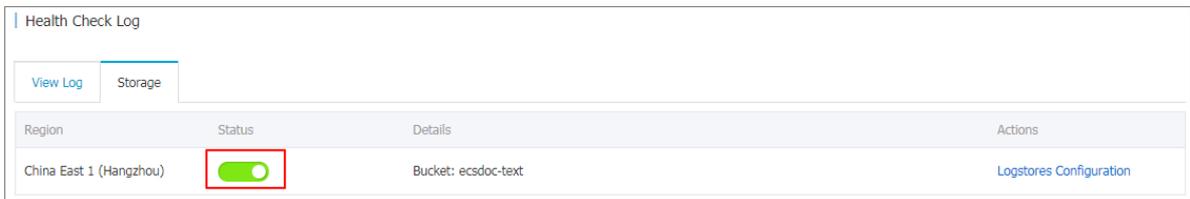
1. Log on to the [SLB console](#).
2. In the left-side navigation pane, click Logs > Health Check Log.
3. On the Health Check Log page, click Storage.

4. Click Logstores Configuration link of the target region.



5. In the Logstores Configuration dialog box, select a bucket to store health check logs, and then click Confirm.

6. Click the status switch to enable log storing.



View health check logs

To view the health check logs in three days, complete these steps:

1. Log on to the [SLB console](#).
2. Log on to the [SLB console](#).
3. In the left-side navigation pane, click Logs > Health Check Log.
4. On the Health Check Log page, click the View Log tab.



Note:

Health check logs are generated only when the health status of the backend servers is abnormal. If no failures occur for all the 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.

Health Check Log

View Log Storage

Note: Only log entries of the last three days are available. To obtain more log entries, go to Storage

Load Balancer ID **Search**

Instance ID	Created At	Details
lb-1ude5vu7cu1hvweff1niy	2018-06-08 09:41:01	[47.97.240.72]:80 to 172.16.3.120:9080 normal
lb-1ude5vu7cu1hvweff1niy	2018-06-08 09:41:01	[47.97.240.72]:80 to 172.16.3.119:9080 normal
lb-1ude5vu7cu1hvweff1niy	2018-06-08 09:41:01	[47.97.240.72]:80 to 172.16.3.120:9080 normal
lb-1ude5vu7cu1hvweff1niy	2018-06-08 09:41:01	[47.97.240.72]:80 to 172.16.3.120:9080 normal

Download health check logs

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

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 /`.

slb Type Standard Storage Region China East 1 (Hangzhou) Created At 07/06/2017, 19:13 Delete Bucket

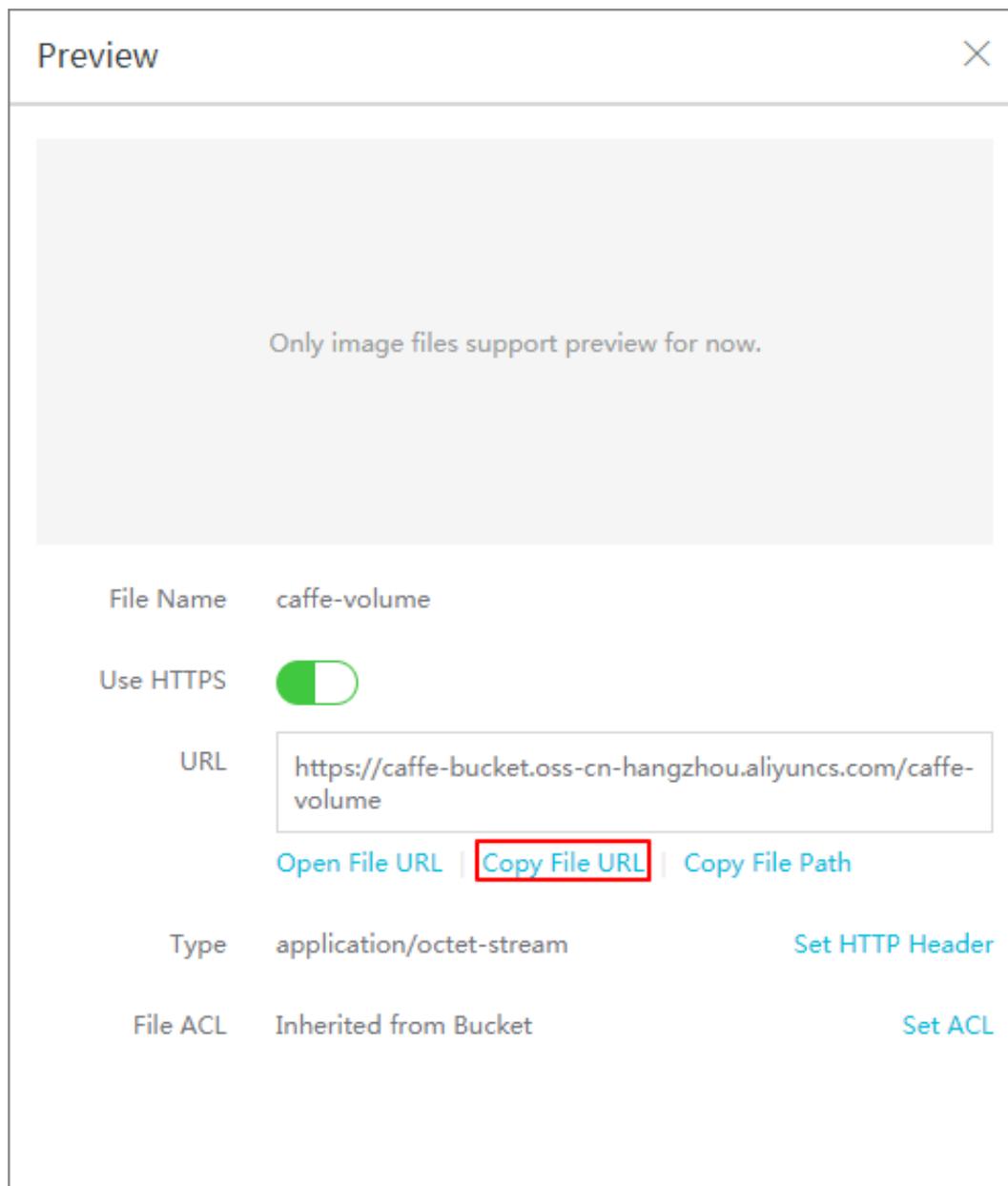
Overview | **Files** | Basic Settings | Domain Names | Image Processing | Event Notification

Basic Data | Hotspot Statistics | API Statistics | Object Access Statistics

Upload Create Directory Delete Set HTTP Header Fragments Refresh

<input type="checkbox"/>	File Name (Object Name)	File Size	Storage Class	Time Updated	Action
<input type="checkbox"/>	AliyunSLBHealthCheckLogs/				
<input type="checkbox"/>	OssAttribute	0.057KB	Standard Storage	07/25/2017, 11:22	Edit
<input type="checkbox"/>	example.jpg	21.327KB	Standard Storage	07/28/2017, 17:14	Edit

4. Click the folder of the health 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.

1.6 Anti-DDoS Basic

You can view Alibaba Cloud Security thresholds of an Internet SLB instance on the SLB console.

Introduction to Anti-DDoS Basic

Alibaba Cloud provides up to 5 Gbps basic anti-DDoS protection 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 cleans 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 cleaning threshold and blackhole threshold according to the bandwidth of the Internet SLB instance. When the inbound traffic reaches the threshold, the cleaning or blackhole is triggered:

- **Cleaning:** When the attack traffic from the Internet exceeds the cleaning threshold or matches certain attack traffic model, Alibaba Cloud Security starts cleaning the attack traffic. The cleaning operation includes packet filtration, traffic speed limitation, packet speed limitation and so on.
- **Blackhole:** When the attack traffic from the Internet exceeds the blackhole threshold, blackhole is triggered and all inbound traffic is dropped.

View thresholds

You can view the thresholds of an instance on the SLB console. If you cannot view the thresholds using a RAM account, ask your system administrator to grant the permission for you. For more information, see [Allow read-only access to Anti-DDoS Basic](#).

To view thresholds, complete these steps:

1. Log on to the [SLB console](#).
2. Select a region.
3. Hover the mouse pointer to the DDoS icon next to the target instance. You can click the link to go to the DDoS console to view more information.
 - **BPS threshold:** When the inbound traffic exceeds the BPS cleaning threshold, cleaning is triggered.
 - **PPS threshold:** When the inbound packets exceed the PPS cleaning threshold, cleaning is triggered.
 - **Blackhole threshold:** When the inbound traffic exceeds the blackhole threshold, blackhole is triggered.

Allow read-only access to Anti-DDoS Basic

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



Note:

You have to use the primary account to complete the authorization.

1. Use the primary account to log on to the RAM console.
2. In the left-side navigation pane, click Users, find the target RAM account 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 Names list. Click OK.

1.7 Peak limit for regional bandwidth

Load Balancing areas can sell maximum bandwidth per year and peak by traffic instance bandwidth as shown in the following table.



Note:

The peak bandwidth of the private network in all regions is 5 Gbps.

Region	Peak Bandwidth
China North 1 (Qingdao)	5 Gbps
East China 1 (Hangzhou)	5 Gbps
North China 2 (Beijing)	5 Gbps
East China 2 (Shanghai)	5 Gbps
South China 1 (Shenzhen)	5 Gbps
China North 3 (Zhangjiakou)	5 Gbps
China North 5 (Hohhot)	5 Gbps
cn-hongkong	2 Gbps
Eastern United States 1 (Virginia)	1 Gbps
Western United States 1 (Silicon Valley)	2 Gbps

Region	Peak Bandwidth
Asia Pacific NE 1 (Tokyo)	1 Gbps
Asia Pacific Southeast 1 (Singapore)	5 Gbps
Asia Pacific SE 2 (Sydney)	1 Gbps
Asia Pacific SE 3 (Kuala Lumpur)	5 Gbps
Middle East 1 (Dubai)	500 Mbps
Germany 1 (Frankfurt)	1 Gbps
Asia Pacific South 1, Mumbai (Bombay)	5 Gbps

1.8 Multiple zone deployment

When creating SLB instances, you can create SLB instances in the region with multiple zones to improve the availability.

What is multiple zone?

A cloud product zone refers to a set of independent infrastructures. Different zones have independent infrastructures (such as network, power supply and air-conditioning), thus an infrastructure fault 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 master zone is faulty and unavailable, SLB is able to switch to the data center in the slave zone to restore its service capabilities within 30 seconds.

Note the following about SLB master-slave zones:

- SLB supports attaching ECS instances in different zones as long as the ECS instances and the SLB instance are in the same region. SLB can distribute traffic to the ECS instances in different zones.
- Normally, the SLB instance in the slave zone is in the standby state. You cannot manually switch the active/standby state of an SLB instance. SLB will switch to the slave zone only when the data center of the master zone is unavailable such as outage. SLB will not switch to the slave zone if an SLB instance is faulty.
- SLB and ECS are deployed in different clusters. When an SLB instance is unavailable, the ECS instance is still available. Therefore, after SLB switches to the slave zone, the SLB instance in the slave zone still can distribute traffic to the added ECS instances. However, if all clusters in a zone is unavailable or the optical cable

is broken, then all the services in the zone including SLB and ECS cannot work anymore.

For more information, see [SLB high availability](#).

Master-slave zone list

The following table lists the master-slave in zones. You can call the `DescribeZones` API to obtain available master-slave zones in a region.

Regions	Zone type	Zone	Zone
China (Hangzhou)	Multi-zone	Master zone	Slave zone
		Zone B	Zone D
		Zone D	Zone E
		Zone E	Zone F
		Zone F	Zone E
China (Shanghai)	Multi-zone	Master zone	Slave zone
		Zone A	Zone B
		Zone B	Zone A or Zone D
		Zone C	Zone B
		Zone D	Zone B
China (Shenzhen)	Multi-zone	Master zone	Slave zone
		Zone A	Zone B
		Zone B	Zone A
		Zone C	Zone B
China (Qingdao)	Multi-zone	Master zone	Slave zone
		Zone B	Zone C
		Zone C	Zone B
China (Beijing)	Multi-zone	Master zone	Slave zone
		Zone A	Zone B or Zone D
		Zone B	Zone A or Zone C
		Zone C	Zone B
		Zone D	Zone A

Regions	Zone type	Zone	Zone
China (Zhangjiakou)	Single zone	Zone A	Zone A
China (Hohhot)	Single zone	Zone A	Zone A
Germany (Frankfurt)	Single zone	Zone A	Zone A
UAE (Dubai)	Single zone	Zone A	Zone A
Singapore	Multi-zone	Master zone	Slave zone
		Zone A	Zone B
		Zone B	Zone A
Australia (Sydney)	Single zone	Zone A	Zone A
Malaysia (Kuala Lumpur)	Single zone	Zone A	Zone A
Japan (Tokyo)	Single zone	Zone A	Zone A
China (Hong Kong)	Multi-zone	Master zone	Slave zone
		Zone B	Zone C
		Zone C	Zone B
US (Virginia)	Single zone	Zone A	Zone A
US (Virginia)	Multi-zone	Master zone	Slave zone
		Zone A	Zone B
		Zone B	Zone A

1.9 Achieve cross-regional load balancing with cloud resolution

Global load balancing

Load balancing is divided into local load balancing and global load balancing. Local load balancing balances loads of server groups in the same region. Global load balancing balances server groups that are in different regions and have different network structures.

Combined with cloud resolution (global load balancing edition), you can deploy cloud resolution (global load balancing edition) on the local load balancing upper layer), achieve cross-regional disaster recovery and intelligent resolution.

- **Multi-line intelligent analytical service**

Multi-line intelligent resolution is supported by cloud resolution (global load balancing, that is, based on the type of network that the site visitor is located in , the intelligent judgment provides the best access resolution address, enables access to the user for the fastest and most smooth experience.

As shown in the figure below, when judging the source of the visitor as an overseas user, the domain name is resolved to a load balancing instance located in the united states; when judging the source of the visitor as a domestic user, the domain name is resolved to the domestic load balancing instance.

- **Cross-regional disaster recovery**

Cloud resolution provides monitoring services that monitor the ip address that the host logs. According to the monitoring and detection results of the website, the fault cluster is isolated in real time and the flow rate is switched dynamically, achieve cross-regional disaster recovery.

Deploy global load balancing

This operation takes a domain name as aliyuntest. club's website (for example, the majority of its users are located in singapore and in the country), guides you through cloud resolution and load balancing for global load balancing.

Step 1 purchase and configure the cloud server

Depending on the geographic distribution of the users of your application service, purchase and configure at least two ecs under the appropriate geographic area.

In this operation, in beijing, shenzhen and singapore, two ecs were purchased separately, and set up a simple static web page on ecs.

- **Beijing regional ecs example**
- **Shenzhen regional ecs example**
- **Example of regional ecs in singapore**

Step 2: purchase and configure load balancing instances

1. Refer to creating load balancing instances in beijing, shenzhen and singapore to create three public load balancing instances.
2. Refer to configuring load balancing instances to add listening and adding the ecs configured under each zone to the back-end server pool.

- Example of load balancing in beijing
- Example of load balancing in shenzhen
- Example of regional load balancing in singapore

Step 3 configure global load balancing

1. Configure cloud resolution for global load balancing.

- a. On the global load balancing page, click buy now.
- b. Configure cloud resolution for global load balancing.

In this operation, other configurations use the default options in addition to the following:

- Configuration type, select global load balancing edition.
- Global load balancing type, select inter-line load balancing.
- Bind your site domain name. You can also bind domain names after creation.

- c. Click buy now.

- d. The vip product page of the dns management console is resolved on the cloud, view the global load balancing cloud resolution instance created.

2. Sets intelligent resolution.

- a. Log in to the domain name service management console.
- b. Locate the domain name that the cloud resolution instance is bound to, and then click the resolution setting that is listed in its actions.
- c. Sets the domain name resolution. Add three a records to point to the public network ip of the load balancing instance that was created, respectively, it also sets the-record analytical line of the singapore area to overseas.

3. Add a web site monitor.

- a. On the parse settings page, click web site monitoring.
- b. Click Add monitor to add the IP address of the three A records to the monitor list
 -
- c. Modify the monitoring configuration by clicking the settings option for each monitor record.

Set the switchover rule when the domain name record cannot be accessed to force a pause of the record resolution.

4. Turn on global load balancing.
 - a. On the parse settings page, click global load balancing.
 - b. Click the inter-line load balancing tab, select the target domain name, and then click open.
 - c. In the pop-up dialog box, click OK.

Step 4 test

Remove the back-end server for the beijing regional load balancing instance so that the service for the load balancing instance is unavailable.

Visit the website to see if the access is normal.



Note:

Cloud-resolved dns monitoring takes 1-2 minutes to aggregate judgment after your ip downtime, suppose your monitoring frequency is set to 1 minute, then the line abnormal switching effective time in 2-3 minutes; if your monitoring frequency is set to 10 minutes, then the line abnormal switching effective time in 12-13 minutes.