

Alibaba Cloud Server Load Balancer

Quick Start (New Console)

Issue: 20190415

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 Instance_ID</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 Tutorial overview.....	1
2 Before you begin.....	2
3 Create an ECS instance.....	6
4 Install static web pages.....	9
5 Create an SLB instance.....	11
6 Configure an SLB instance.....	14
7 Resolve a domain name.....	20
8 Delete an SLB instance.....	21

1 Tutorial overview

This tutorial guides you to create an Internet SLB instance to forward requests to backend ECS instances.

**Note:**

Before creating an SLB instance, you must plan your SLB service, such as the instance type, region, and more. For more information, see [Before you begin](#).

The tutorial includes the following tasks:

1. [Create an SLB instance](#)

Creates an SLB instance. An SLB instance is a running entity of Server Load Balancer.

2. [Configure listeners and add backend servers](#)

After creating an SLB instance, you have to add at least one listener, and add ECS instances as backend servers.

3. [Resolve a domain name](#)(Optional)

Use Alibaba Cloud DNS to resolve a domain name to the IP address of the SLB instance to provide external services.

4. [Delete an SLB instance](#)

If you no longer need the SLB instance, release it to avoid additional charges.

2 Before you begin

Before you use Server Load Balancer (SLB), you need to determine the instance type and region, network type, listener protocol, and backend servers according to your business needs.

Instance region

Consider the following scenarios when you select the region to which the SLB instance belongs:

- To reduce latency and increase the download speed, select a region closest to your customers.
- To provide more stable and reliable load balancing services, deploy primary and backup zones for zone-level disaster tolerance. To do this, make sure that you select a region in which primary and backup zones are available.
- SLB does not support cross-region deployment. Therefore, make sure that the region selected for the SLB instance is the same as the region for your backend ECS instances.

Network type

SLB provides Internet and intranet load balancing services. Consider the following scenarios when you select the network type of an SLB instance:

- If you want to use SLB to distribute requests from the Internet, create an Internet SLB instance.

An Internet SLB instance is provided with a public IP address to receive requests from the Internet.

- If you want to use SLB to distribute requests from the intranet, create an intranet SLB instance.

An intranet SLB instance only has a private IP address and is accessible only from a classic network or VPC.

Instance type

When you create an SLB instance, you need to choose either a guaranteed-performance or a shared-performance instance type.

The guaranteed-performance instance type provides greater flexibility in resource utilization so to guarantee service availability. For guaranteed-performance instances, Alibaba Cloud SLB provides six specifications for these instances to better meet your specific requirements. We recommend that you select the highest specification, Super I (slb.s3.large). This guarantees the running of your services and will not incur extra costs. However, if you do not require the highest specification available, you can choose a lower specification instance such as Higher II (slb.s3.medium).

Listener protocol

SLB supports Layer-4 (TCP and UDP) and Layer-7 (HTTP and HTTPS) load balancing.

- A Layer-4 listener distributes requests directly to backend servers without modifying HTTP headers. After a request arrives at a Layer-4 listener, SLB uses the backend port configured in the listener to establish a TCP connection with backend ECS instances.
- A Layer-7 listener is an implementation of reverse proxy. After a request arrives at a Layer-7 listener, SLB uses a TCP connection to transmit the data packets to backend ECS instances instead of transmitting the data packets directly.

Layer-7 listeners involve one more procedure than Layer-4 listeners for forwarding incoming requests, which can lead to longer performance times. Moreover, scenarios involving insufficient client ports or excessive connections to the backend servers also affect the performance of Layer-7 listeners. Therefore, if you require high performance, we recommend that you use Layer-4 listeners.

For more information, see [Protocols](#).

Backend servers

Before using the SLB service, you must create an ECS instance, deploy applications on it, and add it to an SLB instance.











Note the following when you create and configure an ECS instance:

- The region and zone of the ECS instance

Make sure that the region of the ECS instance is the same as that of the SLB instance. Additionally, we recommend that you deploy each ECS instance in

different zones to improve availability. For more information, see [Create an instance by using the wizard](#).

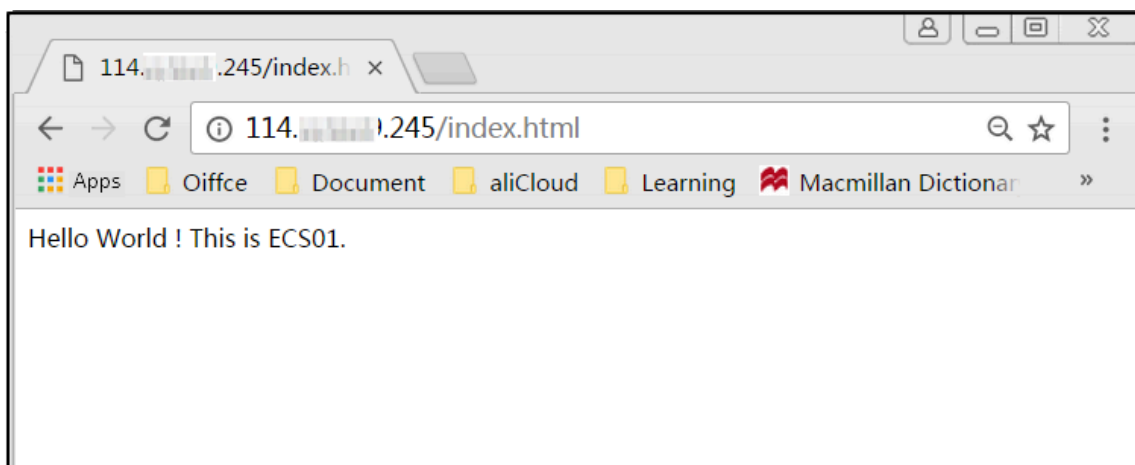
In this example, two ECS instances are created in the China (Hangzhou) region. They are named as ECS01 and ECS02 as shown in the following figure.

<input type="checkbox"/>	Instance ID/Name	Tags	Monitoring	Zone	IP Address	Status	Network Type	Configuration	Billing Method	Actions
<input type="checkbox"/>	ECS01	  		Hangzhou Zone B		 Running	VPC	4 vCPU 8 GB (I/O Optimized) ecs.n1.large 0Mbps (Peak Value)	Pay-As-You-Go February 5, 2019, 16:02 Create	Manage Connect Change Instance Type More ▼
<input type="checkbox"/>	ECS02	  		Hangzhou Zone D		 Running	VPC	4 vCPU 8 GB (I/O Optimized) ecs.n1.large 0Mbps (Peak Value)	Pay-As-You-Go February 5, 2019, 16:02 Create	Manage Connect Change Instance Type More ▼

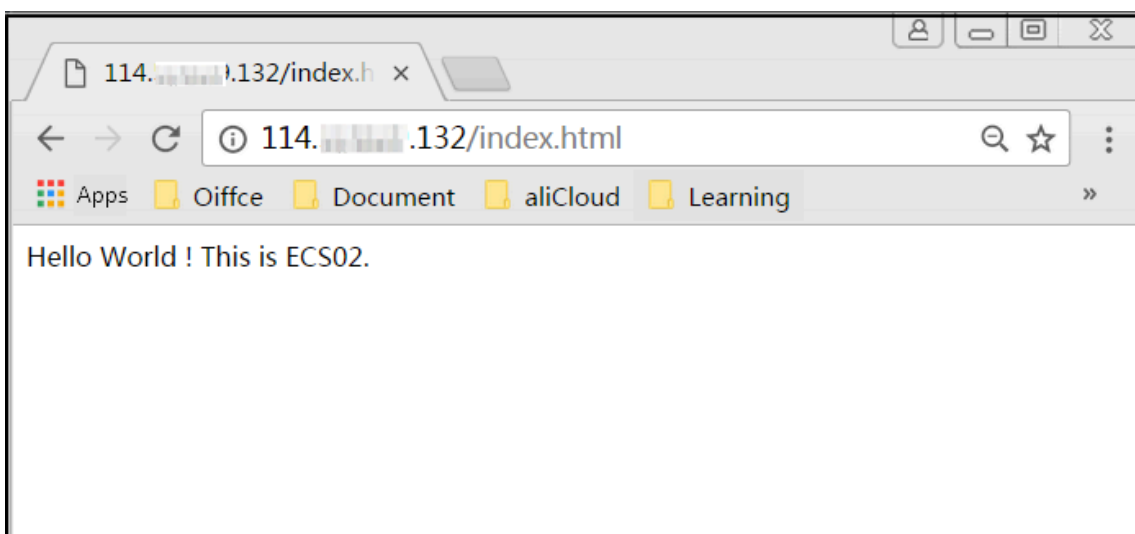
- Application configuration

In this example, a static website is deployed on ECS01 and another on ECS 02 by using Apache, as shown in the following figure.

- Enter the EIP address attached to ECS01 in the browser:



- Enter the EIP address attached to ECS02 in the browser:



No additional configuration is required after you deploy applications on the ECS instances. However, if you want to use Layer-4 listeners, and the ECS instances use a Linux operating system, make sure that the values of the following parameters in the `net . ipv4 . conf` file in `/ etc / sysctl . conf` are set to `0` :

```
net . ipv4 . conf . default . rp_filter = 0
net . ipv4 . conf . all . rp_filter = 0
net . ipv4 . conf . eth0 . rp_filter = 0
```

3 Create an ECS instance

Before using Server Load Balancer, you must create at least two ECS instances and deploy corresponding applications. You can add the ECS instances to an SLB instance so that they can receive client requests as backend servers.

Context

Follow the instructions in this document to create two ECS instances, ECS01 and ECS02.

Procedure

1. Log on to the ECS console.
2. In the left-side navigation pane, click Instances and then click Create Instance.

3. On the Elastic Compute Services (ECS) page, configure the ECS instance.

The following are ECS settings used in this tutorial. You can change the configuration according to your needs.







- **Region:** Server Load Balancer does not support cross-region deployment. The region must be the same for the Server Load Balancer instance and the ECS instances. In this tutorial, select China East 1.
- **Network Type:** In this tutorial, select VPC. Use the default VPC and VSwitch.
- **Image:** In this tutorial, select Ubuntu 16.04 64 bit.
- **Target:** Set the purchase quantity to 2 and the system automatically creates two ECS instances with the same configurations.
- **Assign public IP:** Select to automatically allocate a public IP address to the ECS instance.
- **Bandwidth Pricing:** Select billing by bandwidth and set the bandwidth to 1 Mbps.
- **Security Group:** The configured security group rules must include Port 22 and Port 80 in the inbound direction.
 - Port 22 is the SSH remote port used for logging on to the ECS instance.
 - Port 80 is the web service default port used for accessing the static page built by Apache in [Install static web pages](#).

The screenshot displays the 'Networking' configuration step in the ECS console. It shows the selection of a VPC ('test_nfs_hzb') and a VSwitch ('test'). The 'Network Billing Method' section is expanded, showing 'Pay-By-Traffic' selected with a bandwidth of 26 Mbps. The total cost is calculated as \$0.036 USD per hour. Navigation buttons for 'Prev: Basic Configurations', 'Next: System Configurations', and 'Preview' are visible at the bottom.

4. Click Create Order to complete the creation.

5. Go back to the instances page and click China (Hangzhou). The two newly created ECS instances are displayed. Hover the mouse pointer over one instance name and

click the displayed pencil icon to change the instance name to ECS01. Then change the other instance name to ECS02.

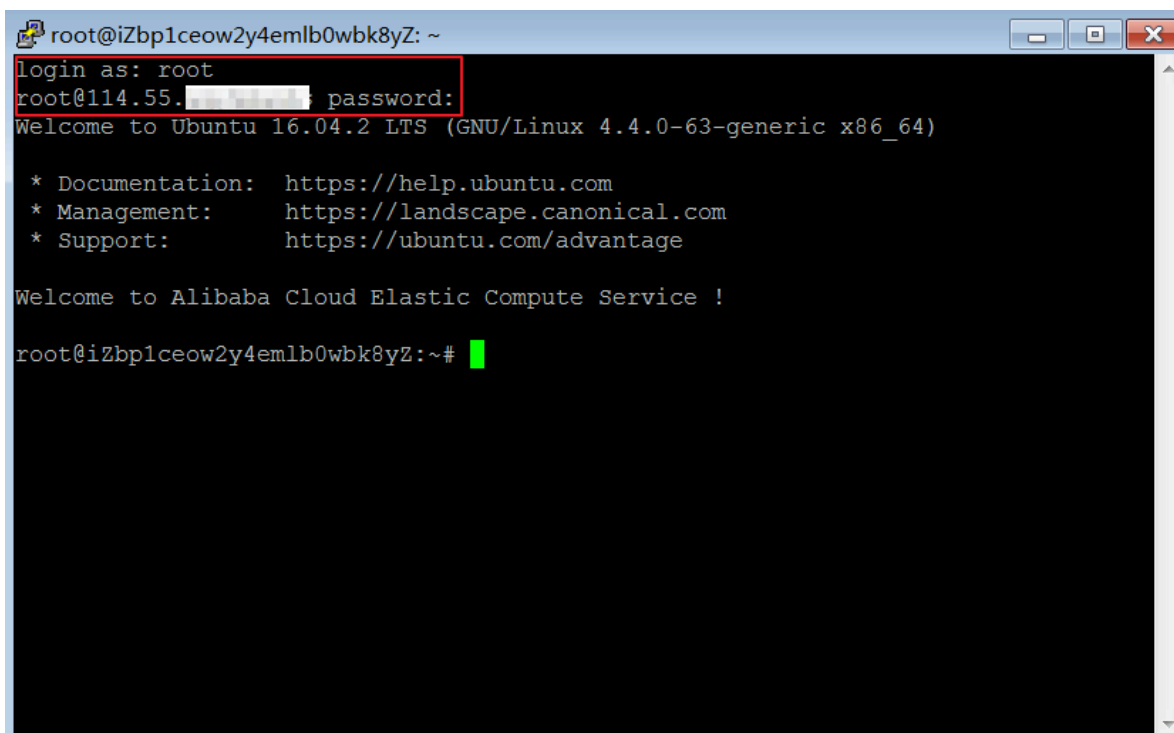
Instance ID/Name	IP Address	Status(All) ▾	Network Type(All) ▾	Billing Method(All) ▾	Action
 i-bp1s- ECS01	 172.17.0.1 (Private IP Address)	 Running	VPC	Pay-As-You-Go 17-07-23 17:23 created	Manage Connect More ▾
 i-bp1s- ECS02	 172.17.0.2 (Private IP Address)	 Running	VPC	Pay-As-You-Go 17-07-23 17:23 created	Manage Connect More ▾

4 Install static web pages

After you create the ECS instances, deploy applications on them. In this tutorial, two static web pages are deployed on the ECS instances using Apache.

Procedure

1. Log on to the ECS instance.



2. Run the following command to update the installation package.

```
sudo apt - get update
```

3. Run the following command to install the Apache server.

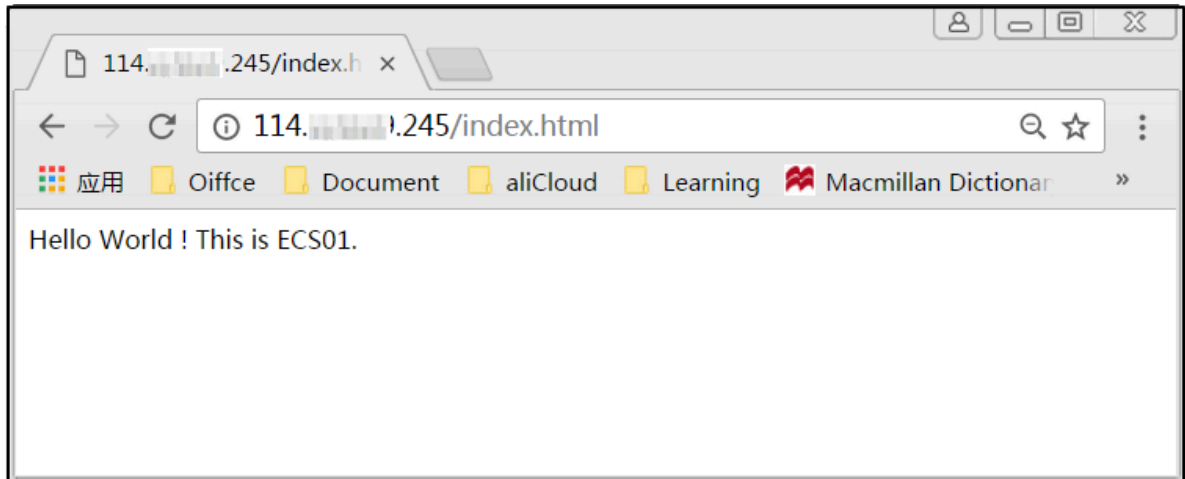
```
sudo apt - get install apache2
```

4. Run the following command to modify the contents of the `index . html` file.

```
cd / var / www / html
```

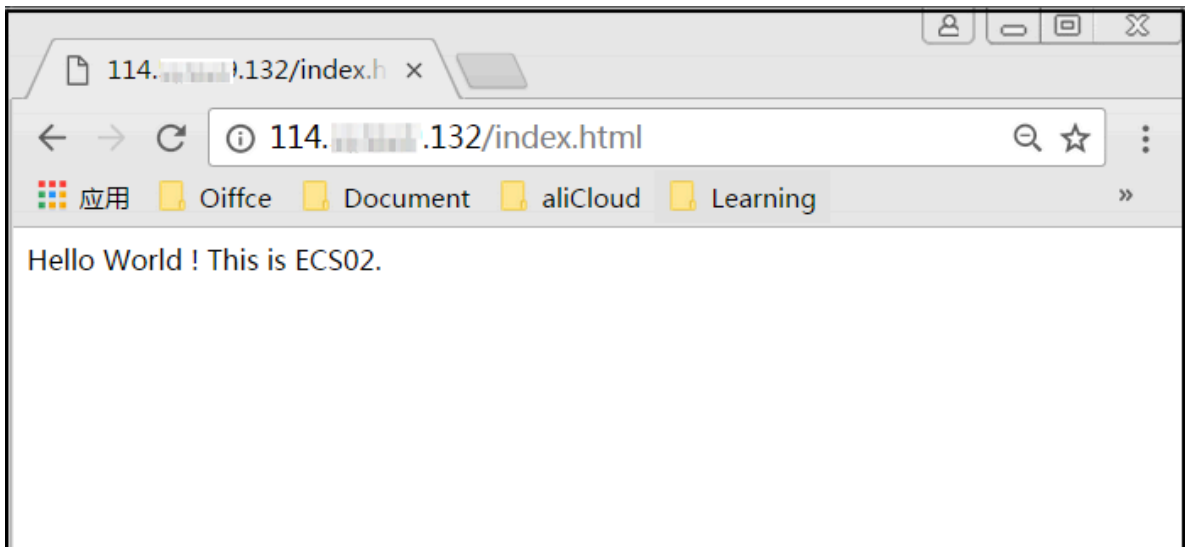
```
echo " Hello World ! This is ECS01 ." > index . html
```

After modifying the content, enter the Elastic IP of the ECS instance in the web browser, you will see the following content.



5. Repeat the preceding steps to create a web page on the other ECS instance and change the content to `Hello World ! This is ECS02 ..`

After modifying the content, enter the EIP of the ECS instance in the web browser, you will see the following content.



5 Create an SLB instance

Before using Server Load Balancer, you must create a Server Load Balancer instance. You can add multiple listeners and backend servers to a Server Load Balancer instance. This tutorial provides step-by-step guidance on how to create an Internet SLB instance. After an Internet SLB instance is created, a public IP is allocated to it. You can resolve a domain to this IP.

Procedure

1. Log on to the [SLB](#) console.
2. On the Server Load Balancer page, click Create Server Load Balancer.
3. Configure the instance according to [Create an SLB instance](#).

The configurations for the Server Load Balancer instance in this tutorial are as follows:

- **Region:** Server Load Balancer does not support cross-region deployment. The region must be the same for the Server Load Balancer instance and ECS instances. In this tutorial, select China (Hangzhou).
- **Zone Type:** Multiple zones have been deployed in most regions for better disaster tolerance. Server Load Balancer can switch to the backup zone to provide the load balancing service when the primary zone is unavailable, and

will automatically switch back to the primary zone when the primary zone is recovered.

In this tutorial, select China East 1 Zone B as the primary zone and China East 1 Zone D as the backup zone.

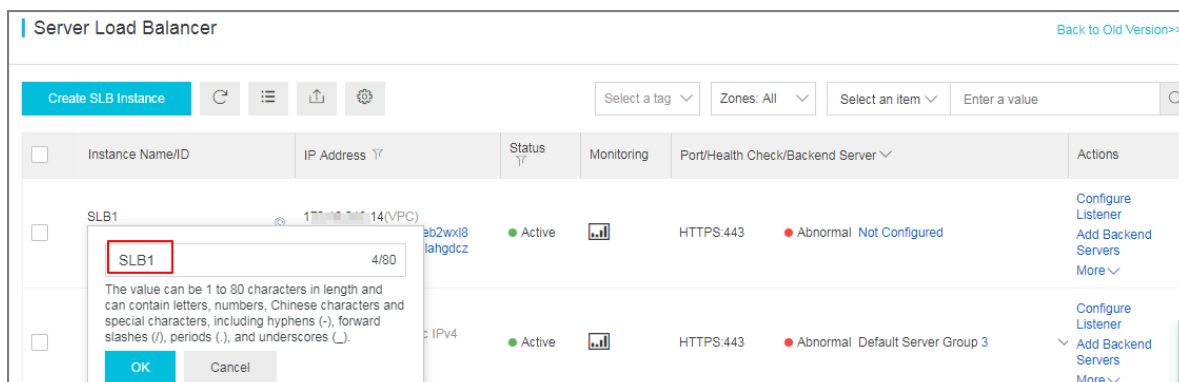
- Instance Type: Select Internet.

Basic Configuration	Region	Singapore	Australia (Sydney)	Malaysia (Kuala Lumpur)	Indonesia (Jakarta)	Japan (Tokyo)
		India (Mumbai)	Hong Kong	US (Virginia)	US (Silicon Valley)	China (Hangzhou)
		China (Shanghai)	China (Shenzhen)	China (Qingdao)	China (Beijing)	China (Zhangjiakou)
		China (Hohhot)	Germany (Frankfurt)	UAE (Dubai)		
	Zone type	Multi-zone				
Primary zone	China East 1 Zone B					
Backup zone	China East 1 Zone D					
Instance name	<input type="text"/> <p>The length must be to 1-80 characters, allowing letters, numbers, and '-', '/', ':', '_'.</p>					
work and instance type	Instance type	<input checked="" type="radio"/> Internet <input type="radio"/> Intranet				
	Instance Spec	<input checked="" type="radio"/> Small I (slb.s1.small)				
		Max connection: 5000, CPS: 3000, QPS: 1000				

4. Click Buy Now and complete the payment.

5. Go back to the SLB console.

6. On the Server Load Balancer page, select the China (Hangzhou) region. Hover the mouse pointer to the instance name area and then click the pencil icon. Enter SLB1 as the name of the instance, click OK.



What's next

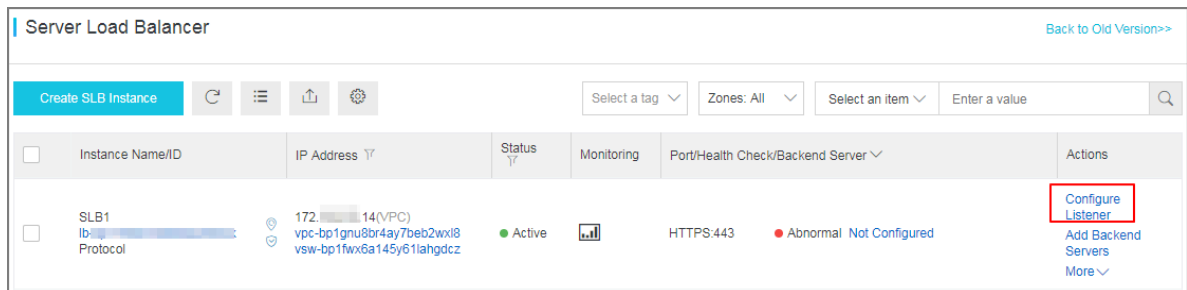
[Resolve a domain name](#)

6 Configure an SLB instance

After creating an SLB instance, you must add at least one listener and a group of backend servers to it. In this tutorial, we will add one TCP listener and two ECS instances to the created SLB instance.

Procedure

1. Log on to the [SLB console](#).
2. On the Server Load Balancer page, locate the target instance and click **Configure Listener**.



3. In the Protocol and Listener tab, configure the listening rule according to the following information and use the default values for the remaining configurations.

- **Select Listener Protocol:** In this tutorial, select TCP.
- **Listening Port:** The frontend protocol and port used to receive requests and forward the requests to backend servers. The frontend ports in an SLB instance must be unique.

In this tutorial, set the port number to 80.

- **Enable Peak Bandwidth Limit:** You can set a peak bandwidth to limit the service capabilities that applications on the ECS instances can provide.

In this tutorial, you do not need to set the peak bandwidth because the instance is billed by traffic.

- **Scheduling Algorithm:** Server Load Balancer supports the following scheduling algorithms. In this tutorial, Round Robin is selected.
 - **Weighted round robin (WRR):** Distribute requests according to the weights of backend servers. Servers with higher weights receive more requests than those with lower weights.
 - **Weighted least connections (WLC):** In addition to the weight set to each backend ECS server, the number of connections to the client is also considered. A server with a higher weight value will receive a larger percentage of live connections at any one time. If the weights are the same,

the system directs network connections to the server with the least number of established connections.

- Round robin (RR): Requests are distributed evenly across the group of backend ECS servers sequentially.

Configure Server Load Balancer [Back](#)

Protocol and Listener Backend Servers Health Check Submit

Select Listener Protocol

Listening Port [?](#)

80

Advanced [Modify](#) [v](#)

Scheduling Algorithm	Weighted Round-Robin	Session Persistence	Disabled
Access Control	Disabled	Peak Bandwidth	No Limit

[Next](#) [Cancel](#)

4. Click Next. In the Backend Servers tab, click Default Server Group, and then click Add.
 - a) On the Available Servers page, select the created ECS instances, and then click Add to Selected Server List.
 - b) Click OK.
 - c) Configure ports and weights for the added backend servers.
 - The ports are backend ports opened on ECS instances to receive requests and can be the same in an SLB instance. In this tutorial, set the backend port numbers to 80 .
 - An ECS instance with a higher weight will receive a larger number of requests . The default value is 100 and we recommend that you use the default value.

Protocol and Listener | **Backend Servers** | Health Check | Submit

| Add Backend Servers

① Add backend servers to handle the access requests received by the SLB instance.

Forward Requests To: **Default Server Group** | VServer Group | Active/Standby Server

Servers Added

ECS Instance ID/Name	Public/Internal IP Address	Port	Weight	Actions
ECS01_KT	47.110.172.182(Public) vpc-bp1w08 vsw-bp1w08 bp1w08	80	100	Delete
ECS02_KT	47.110.172.183(Public) vpc-bp1w08 vsw-bp1w08 bp1w08	80	100	Delete

0 servers have been added. 2 servers are to be added, and 0 servers are to be deleted. [Add More](#)

[Previous](#) [Next](#) [Cancel](#)

5. Click Next to configure health check settings. In this tutorial, default configurations are used.

With health check enabled, when an ECS instance is declared as unhealthy, Server Load Balancer will distribute requests to other healthy ECS instances and restore service to it when it becomes healthy.

6. Click Next. On the Submit page, click Submit.

Configure Server Load Balancer [Back](#)

Protocol and Listener > Backend Servers > Health Check > **Submit**

Submit

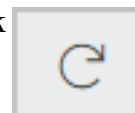
Default Server Group Success

Layer-4 listener Success

Start Listener Success

[OK](#) [Cancel](#)

7. Click OK. Go back to the Server Load Balancer page and click



When the health check status of the backend server is Normal, it indicates that the backend server can process forwarded client requests.

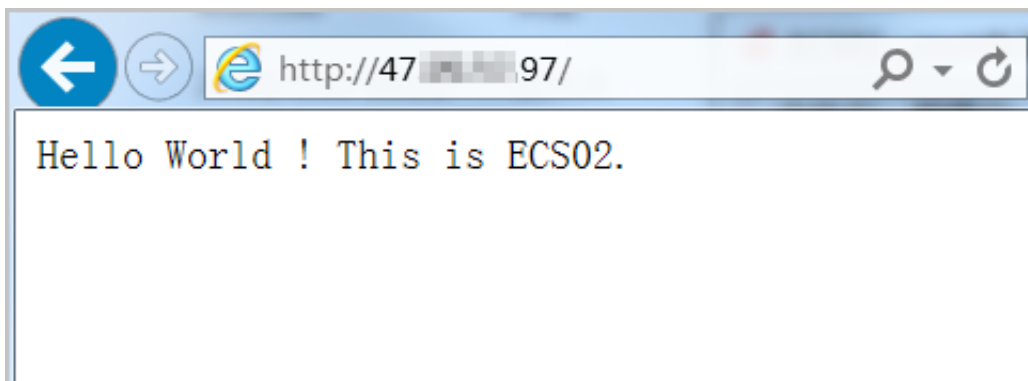
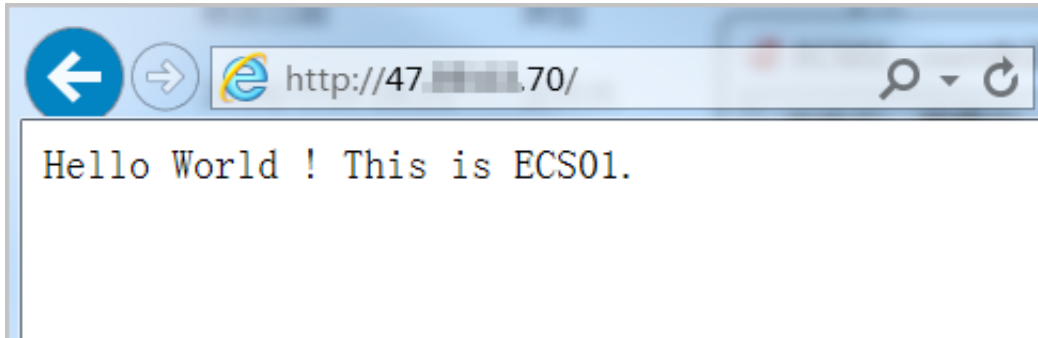
Server Load Balancer [Back to Old Version>>](#)

[Create SLB Instance](#) [Refresh](#) [List](#) [Add](#) [Settings](#)

Select a tag Zones: All Select an item Enter a value

<input type="checkbox"/>	Instance Name/ID	IP Address	Status	Monitoring	Port/Health Check/Backend Server	Actions
<input type="checkbox"/>	SLB1 lb-xxxxxx Protocol	172.16.0.14(VPC) vpc-xxxxxx vs-xxxxxx	Active		TCP: 80 HTTPS:443 Unavailable Default Server Group 2 Abnormal Default Server Group 2	Configure Listener Add Backend Servers More
<input type="checkbox"/>	- lb-xxxxxx The tag is not set.	118.252.252(Public IPv4 Address)	Active		HTTPS:443 Abnormal Default Server Group 3	Configure Listener Add Backend Servers More
<input type="checkbox"/>	- lb-xxxxxx The tag is not set.	128.242(Public IPv4 Address)	Active		HTTPS:143 Normal Default Server Group 2	Configure Listener Add Backend Servers More

8. In the web browser, enter the IP address of the Server Load Balancer instance to test the service.



7 Resolve a domain name

You can resolve a domain name to the public address of an SLB instance.

Context

For example, the domain name of your website is `www.abc.com` and the website is running on an ECS instance with the public IP `1.1.1.1`. After creating a Server Load Balancer instance, a public IP `2.2.2.2` is allocated to the instance. You have to add the ECS instance hosting the website to the backend server pool and resolve the domain name `www.abc.com` to `2.2.2.2`. We recommend that you add an A record resolution (resolve a domain name to an IP address).

Procedure

1. Log on to the Alibaba Cloud DNS console.
2. Click Add Domain Name to add a domain name.
3. On the Basic DNS page, click Configure in the Actions column of the target domain name, and complete the DNS configuration.

8 Delete an SLB instance

Delete the SLB instance when you no longer need the load balancing service to avoid additional charges. Deleting the Server Load Balancer instance does not delete or affect backend ECS instances.

Context

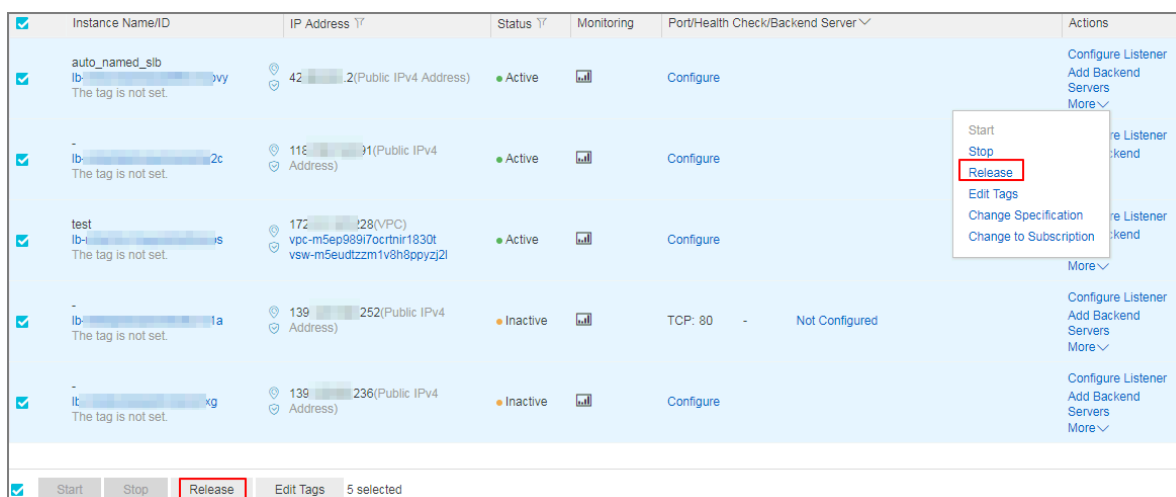


Note:

- If you have resolved a domain name to the SLB endpoint, resolve it to another IP address first to avoid service interruption.
- Only Pay-As-You-Go SLB instances can be released. Subscription SLB instances are automatically released if they are not renewed timely.
- The backend ECS instances are still running after the SLB instance is released. You can release the backend ECS instances if you do not need them anymore.

Procedure

1. Log on to the [SLB console](#).
2. On the Instances page, select the region where the instance is located.
3. Locate the target instance, click Release at the bottom of the list or click More > Release in the actions column.



4. In the Release dialog box, select Release Now or Release on Schedule.

If you select Release on Schedule, set a release time.

5. Click Next.

6. Click OK to release the SLB instance.