Alibaba Cloud Server Load Balancer

FAQ

Issue: 20200203

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, consequent**

ial, 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 contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, are intellectu al property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of this document shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates.**
6. **Please contact Alibaba Cloud directly if you discover any errors in this document .**

Document conventions

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

Style	Description	Example
{} or {a|b}	**This format is used for a required value, where only one item can be selected.**	`switch {active|stand}`

Contents

1 Why am I unable to access the SLB instance?

This topic describes possible reasons and resolutions for failing to access a Server Load Balancer (SLB) instance from a client.

Note:
In this example, the frontend port of the SLB instance is 80, the port of the backend ECS instance is 80, and the internal IP address of the ECS instance is 10.11.192.1. You must configure the port and internal IP address according to your actual situation.

No.	Cause	Resolution
1	SLB cannot be accessed by backend servers. For a layer-4 SLB service, a backend ECS instance cannot directly provide services for clients and function as the backend server of the SLB service at the same time.	None
2	Health check exceptions.	For more information, see *How do I troubleshoot health check exceptions of a layer-4 (TCP/UDP) listener?* and *How do I troubleshoot a health check exception of a layer-7 (HTTP/HTTPS) listener?*.
3	Using FTP, TFTP, H.323, and SIP protocols through SLB is not supported.	· For a Linux system, you can configure the forwarding of port 22 and use SFTP to connect and transmit data. · You can associate an Elastic IP Address (EIP) with an FTP server in the cut-through mode to provide external FTP service. For more information, see *#unique_6*.

No.	Cause	Resolution

4	The internal firewall of the server does not allow port 80.	You can run the following command to temporarily disable the firewall to do a test.
		· For a Windows server, run:
		`firewall.cpl`
		· For a Linux server, run:
		`/etc/init.d/iptables stop`

5	Backend port exceptions.	· For a layer-4 SLB service, you can perform a Telnet test. If you receive a response, the backend port is normal. Example: Use `telnet 10.11.192.1 80` to perform a Telnet test. · For a layer-7 SLB service, you can check the HTTP status code returned. The status code must be a status code that indicates a normal condition, such as 200. The test methods are as follows: - Windows: Access the internal IP address of the ECS instance directly from the ECS instance to check if access is normal. Example: `http://10.11.192.1` - Linux: Run the `curl -I` command and check if the status is `HTTP/1.1 200 OK`. Example: `curl -I 10.11.192.1`

6	The rp_filter feature conflicts with the policy-based route of the LVS of SLB.	1. Log on to the ECS instance that is added to the SLB instance. The ECS instance runs a Linux system.
		2. Edit the */etc/sysctl.conf* file and set the following three parameters in the system configuration file to `0`.
		`net.ipv4.conf.default.rp_filter = 0`
		`net.ipv4.conf.all.rp_filter = 0`
		`net.ipv4.conf.eth0.rp_filter = 0`
		3. Run the `sysctl -p` command to make the configurations take effect.

No.	Cause	Resolution

7	Listener exceptions.	Run the following command on the server. If you can see the monitoring information of 10.1.1.192.1: 80, or the monitoring information of 0.0.0.0: 80, the listening function of the ports is normal.
		· For a Windows server, run:
		`netstat -ano | findstr :80`
		· For a Linux server, run:
		`netstat -anp | grep :80`

8	No listener is added to the SLB instance.	Configure a listener. For more information, see *#unique_7*.
9	SLB cannot be accessed through the domain name. This may be caused by an error in domain name resolution.	None
10	Exceptions of the local network of the client or exceptions of the intermediate link of the service provider.	Perform access tests on the service port of SLB in different regions and network environments. If the exception only occurs when the SLB instance is accessed from the local network, it can be determined that the problem is caused by a network exception. Then you can do further troubleshooting and analysis through Ping tests or MTR route tracing.

No.	Cause	Resolution
11	The client IP address is blocked by Alibaba Cloud Security.	1. Visit `http://ip.taobao.com` in the network environment of the client and obtain the public IP address of the client network. 2. Add the IP address to the SLB whitelist to allow access from the IP address.  **Note:** This operation may pose security risks. Make sure that the IP address does not incur malicious attacks on SLB.
12	After you switch to Anti-DDoS Basic from Anti-DDoS Pro , the whitelist is not disabled.	Disable the whitelist.
If the problem persists, open a ticket and submit the following details: · The ID of the SLB instance or the IP address of the SLB service. · The public IP address of the client obtained when you visit `ip.taobao.com`. · Screenshots of ping and MTR route tracing tests performed by the client by using the IP address of the SLB service.		

2 Why is the traffic among my ECS instances unevenly distributed?

Causes

Traffic may be unevenly distributed due to the following reasons:

- **Only a small number of requests are being received by ECS instances.**
- **The target ECS instances have different network capacities.**

  > **Note:**
  > **The memory usage of ECS instances does not indicate whether requests are evenly distributed.**

- **Session persistence is enabled.**

  If session persistence is enabled, it will cause traffic imbalance when few clients are accessing the Server Load Balancer (SLB) instance. This is especially common when a small number of clients are used to test the SLB instance. For example, session persistence (based on source IP addresses) is enabled for a TCP listener and a client is used to test the load balancing service.

- **The ECS instance status is abnormal.**

  Backend servers with abnormal heath status can also lead to an imbalance especially during a stress test. If the health check for a backend ECS instance fails or the health status of a backend ECS instance changes frequently, this will cause an imbalance.

- **TCP Keepalive is enabled.**

  When some backend ECS instances enable TCP Keepalive and others do not, the connections will accumulate on the ECS instances with TCP Keepalive enabled. This scenario will cause an imbalance.

Troubleshooting

- **Check whether the weights of backend ECS instances are the same.**
- **Check whether health checks of backend ECS instances fail or whether the health status is unstable in a specified period. Check whether the health check is correctly configured with the status code.**

- **Check whether both the WLC scheduling algorithm and session persistence are enabled. If so, change the scheduling algorithm to WRR.**

3 How do I obtain real IP addresses of clients?

Server Load Balancer (SLB) supports obtaining real client IP addresses.

Overview

Support for obtaining real IP addresses in SLB is enabled by default.

- **For layer-4 SLB service (TCP protocol), listeners distribute client requests to backend ECS servers without modifying the request headers. Therefore, you can obtain real client IP addresses directly.**
- **For layer-7 SLB service (HTTP and HTTPS protocols), you need to configure application servers, and then use the `X-Forwarded-For` header to obtain real IP addresses of clients.**

  Real client IP addresses are put in the X-Forwarded-For fields of HTTP headers in the following format:

  ```
  X-Forwarded-For: the real IP address of the user, the proxy server 1
  -IP, the proxy server 2-IP, ...
  ```

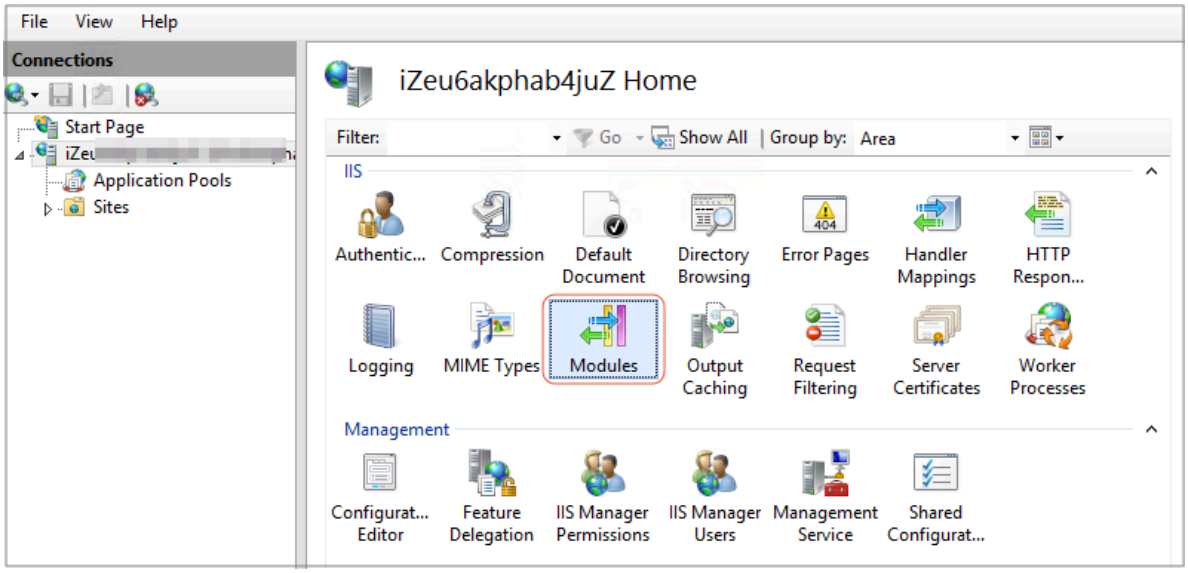
  When you use the X-Forwarded-For header to obtain the real IP address of a client, the first IP address obtained is the real IP address.

  Note:
  For the HTTPS SLB service, SSL certificates are configured in frontend listeners, and the backend still uses the HTTP protocol. Therefore, the configurations on application servers for obtaining real client IP addresses are the same for HTTP and HTTPS protocols.

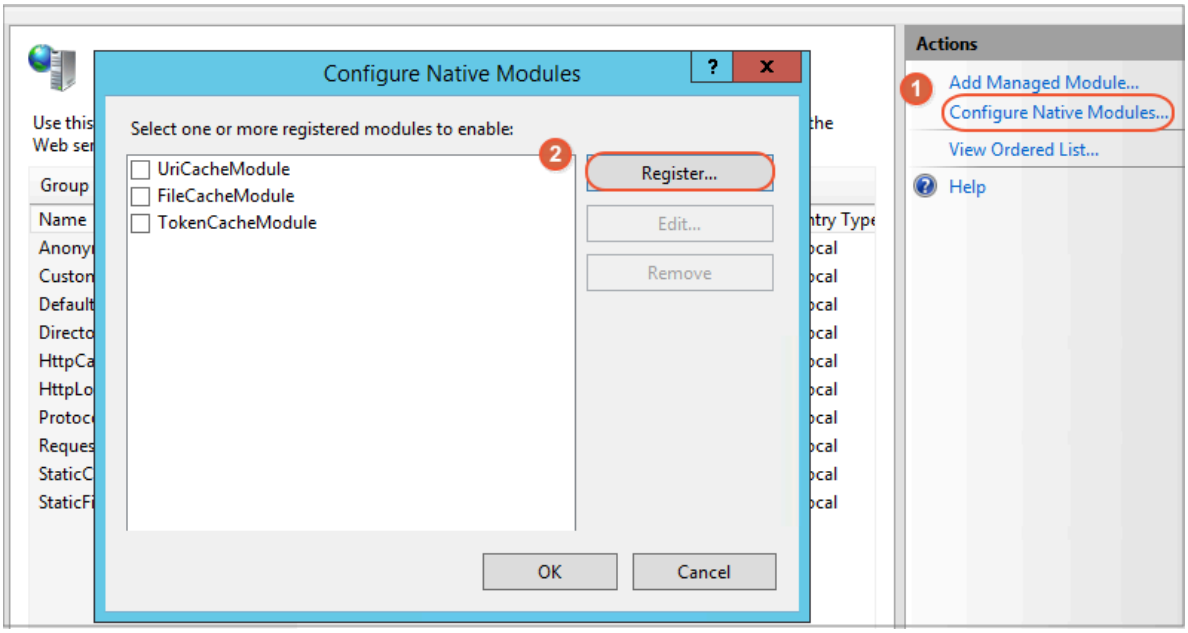
Configure IIS7/IIS8

1. *Download* **and extract** *F5XForwardedFor*.
2. **Copy the** *F5XFFHttpModule.dll* **and** *F5XFFHttpModule.ini* **files from the** *x86\Release* **or** *x64\Release* **directory (depending on the operating system version) of your server to a directory, such as** *C:\F5XForwardedFor*. **Make sure that the IIS process has write access to this directory.**

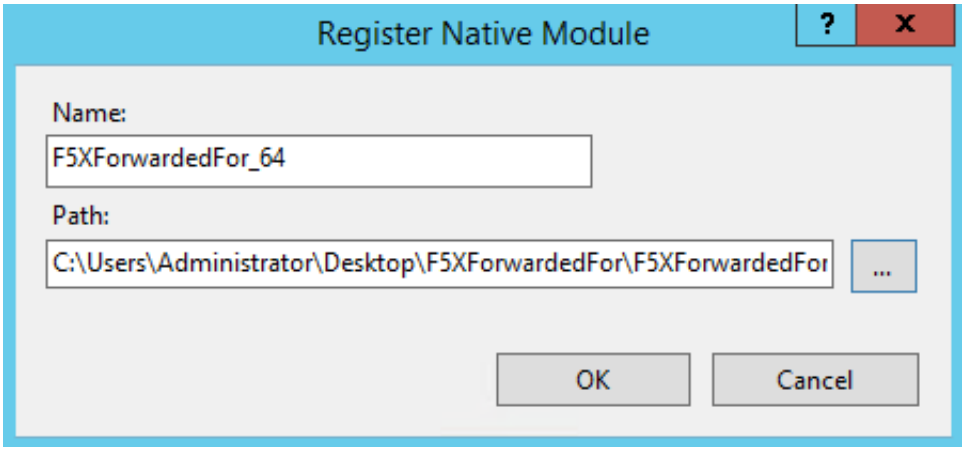
3. **Open IIS Manager and double-click the Modules function.**



4. **Click Configure Native Modules, and then click Register in the displayed dialog box.**

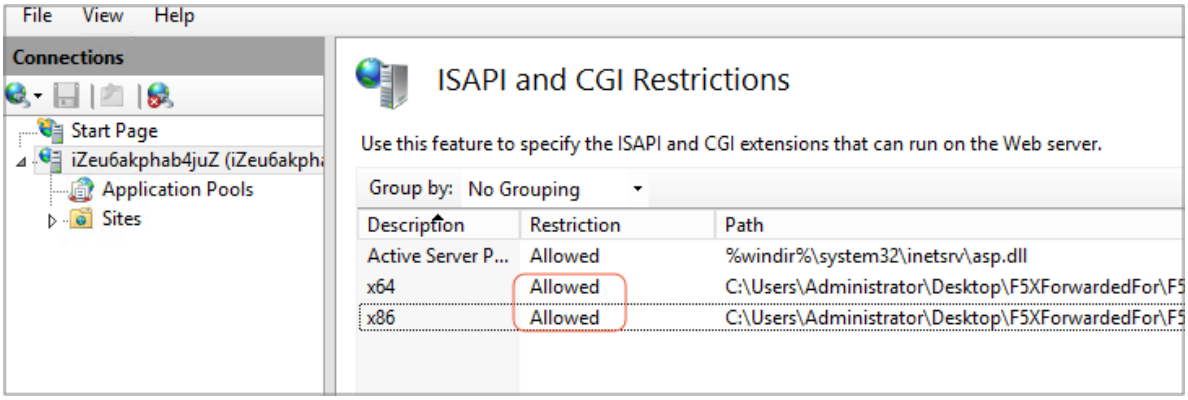


5. **Add the downloaded `.dll` file.**



6. **Add the ISAPI and CGI restrictions for the added files and set the restrictions to Allowed.**

> **Note:**
> **Make sure that you have installed the ISAPI and CGI applications.**



7. **Restart IIS Manager.**

Configure Apache

1. **Run the following command to install the mod_rpaf module:**

```
wget https://github.com/gnif/mod_rpaf/archive/v0.6.0.tar.gz
tar zxvf mod_rpaf-0.6.tar.gz
cd mod_rpaf-0.6
/alidata/server/httpd/bin/apxs -i -c -n mod_rpaf-2.0.so mod_rpaf-2.
0.c
```

2. **Open the `/alidata/server/httpd/conf/httpd.conf` file and add the following information at the end of the content:**

```
LoadModule rpaf_module modules/mod_rpaf-2.0.so
RPAFenable On
RPAFsethostname On
RPAFproxy_ips  <IP_address>
```

```
RPAFheader X-Forwarded-For
```

Note:

To obtain the IP address of the proxy server, add the CIDR block of the proxy server to `RPAFproxy_ips <IP_address>`, such as the IP address range of SLB 100.64.0.0/10 (100.64.0.0/10 is reserved by Alibaba Cloud. No other users can use the IP addresses in this CIDR block and therefore no security risks exist.) and the address range of Anti-DDoS Pro. Separate multiple CIDR blocks by using commas (,).

3. **Restart Apache.**

```
/alidata/server/httpd/bin/apachectl restart
```

Configure Nginx

1. **Run the following command to install http_realip_module.**

```
 wget http://nginx.org/download/nginx-1.0.12.tar.gz
 tar zxvf nginx-1.0.12.tar.gz
 cd nginx-1.0.12
 ./configure --user=www --group=www --prefix=/alidata/server/
nginx --with-http_stub_status_module --without-http-cache --with-
http_ssl_module --with-http_realip_module
 make
 make install
 kill -USR2 `cat /alidata/server/nginx/logs/nginx.pid`
 kill -QUIT `cat /alidata/server/nginx/logs/ nginx.pid.oldbin`
```

2. **Open the *`nginx.conf`* file.**

```
vi /alidata/server/nginx/conf/nginx.conf
```

3. **Add new configuration fields and information at the end of the following configuration information:**

```
 fastcgi connect_timeout 300;
 fastcgi send_timeout 300;
 fastcgi read_timeout 300;
 fastcgi buffer_size 64k;
 fastcgi buffers 4 64k;
 fastcgi busy_buffers_size 128k;
 fastcgi temp_file_write_size 128k;
```

The configuration fields and information that need to be added are:

```
 set_real_ip_from IP_address
 real_ip_header X-Forwarded-For;
```

Note:

To obtain the IP address of the proxy server, add the CIDR block of the proxy server to `set_real_ip_from <IP_address>`, such as the IP address range of SLB 100.64.0.0/10 (100.64.0.0/10 is reserved by Alibaba Cloud. No other users can use the IP addresses in this CIDR block and therefore no security risks exist.) and the address arrange of Anti-DDos Pro. Separate multiple CIDR blocks by using commas (,).

4. **Restart Nginx.**

```
/alidata/server/nginx/sbin/nginx -s reload
```

4 What can I do if health checks generate an excessive number of logs?

SLB can automatically save health check logs generated in three days. If too many health check logs are generated and affect your maintenance, you can reduce health check logs or prevent certain logs from being generated through the following methods.

> **Note:**
> **If you reduce health check logs, SLB faults may be missed. Therefore, we recommend that you consider the risks of the following methods and use the methods with caution.**

- *Get access logs*
- *Adjust health check frequency*
- *Close Layer-7 health checks*
- *Change Layer-7 SLB to Layer-4 SLB*
- *Disable application logs on the health check page*

Get access logs

HTTP health checks use the HEAD request method by default (the GET method will be supported later). Therefore, you can obtain access logs by filtering out HEAD requests.

Adjust health check frequency

You can increase the interval between two health checks to reduce the health check frequency and generated logs.

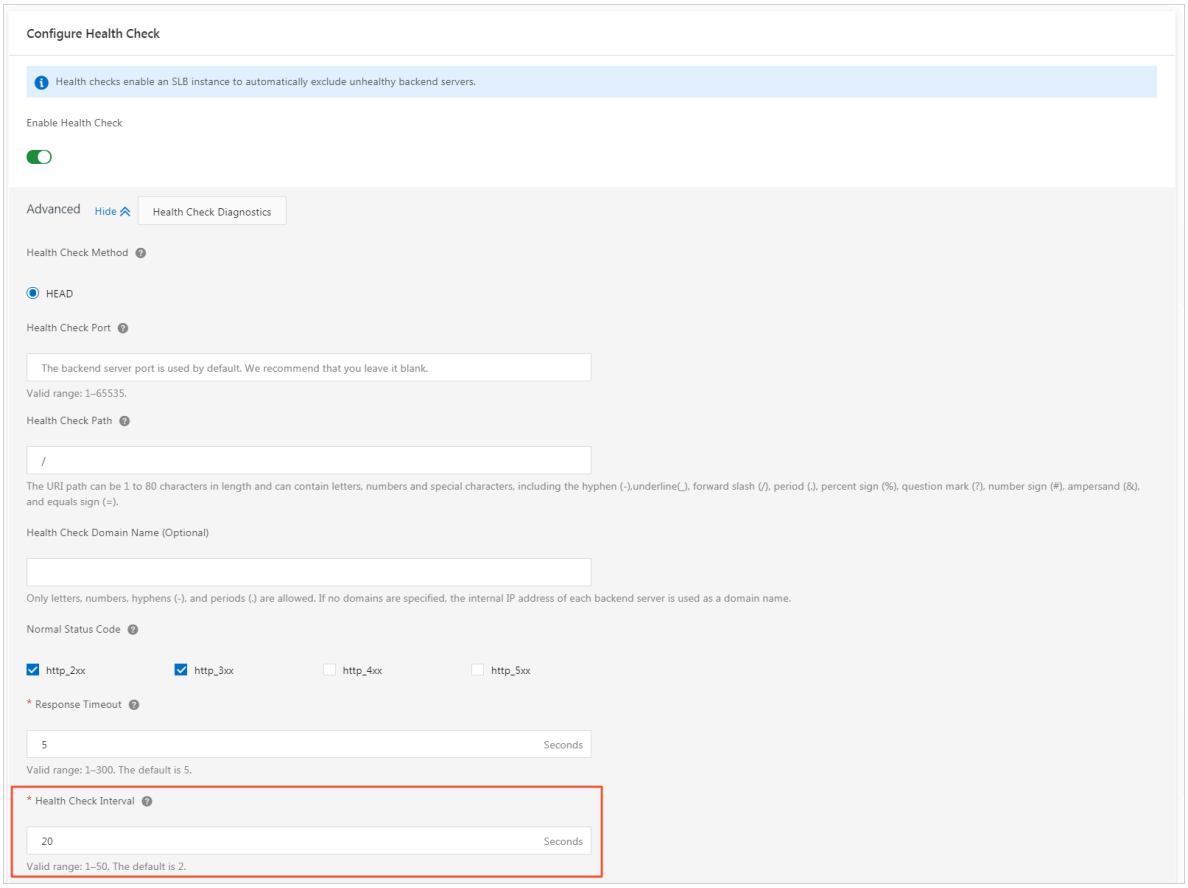
Potential risks

After you increase the interval, if a backend ECS instance fails, the time needed for SLB to detect the faulty ECS instance is increased accordingly.

Procedure

1. **Log on to the** *SLB console*.
2. **On the Server Load Balancer page, click the ID of the target SLB instance.**

3. **Click the Listeners tab, find the target listener, and click Configure in the Actions column.**
4. **On the Configure Listener page, click Next and then click Next again to go to the Health Check tab.**
5. **Adjust the Health Check Interval. Value range: 1 to 50. Unit: seconds. The greater the interval is, the lower the health check frequency is, and the fewer logs are generated by backend servers. Modify the interval according to your actual situation.**



6. **Click OK.**

Close Layer-7 health checks

When Layer-7 (HTTP or HTTPS) SLB is used, health checks are performed through HTTP HEAD requests. Application logs of backend servers record the health check requests, leading to a large number of logs.

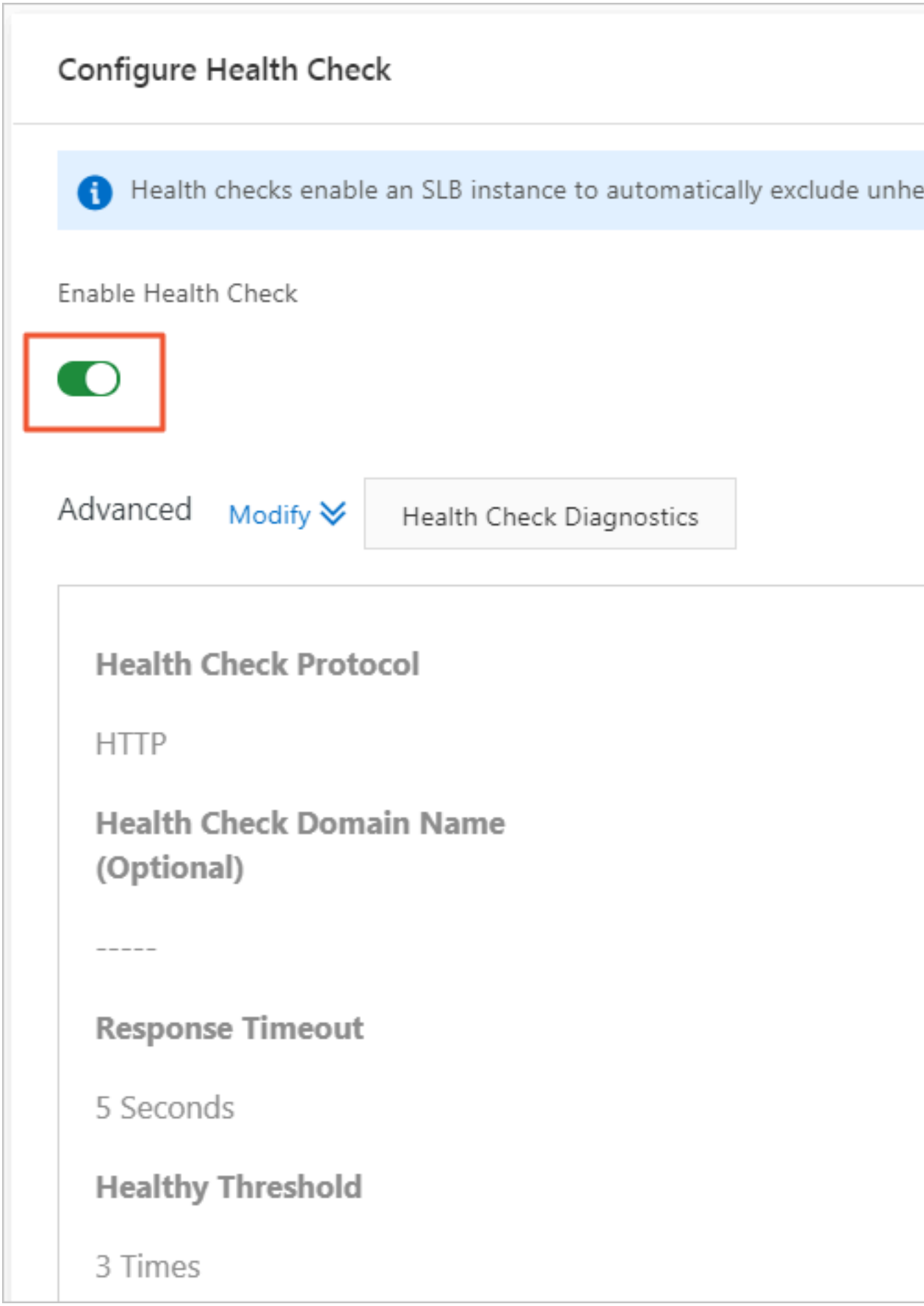
Potential risks

After you close HTTP/HTTPS health checks, SLB does not check backend servers . If a backend server fails, the traffic cannot be automatically forwarded to other normal backend servers.

Procedure

1. **Log on to the *SLB console*.**
2. **On the Server Load Balancer page, click the ID of the target SLB instance.**
3. **Click the Listeners tab, find the target listener, and click Configure in the Actions column.**
4. **On the Configure Listener page, click Next and then click Next again to go to the Health Check tab.**

5. **Turn off Enable Health Check.**


Configure Health Check
Health checks enable an SLB instance to automatically exclude unhe
Enable Health Check
Advanced
Modify
Health Check Diagnostics
Health Check Protocol
HTTP
Health Check Domain Name (Optional)

Response Timeout
5 Seconds
Healthy Threshold
3 Times

6. Click OK.

Change Layer-7 SLB to Layer-4 SLB

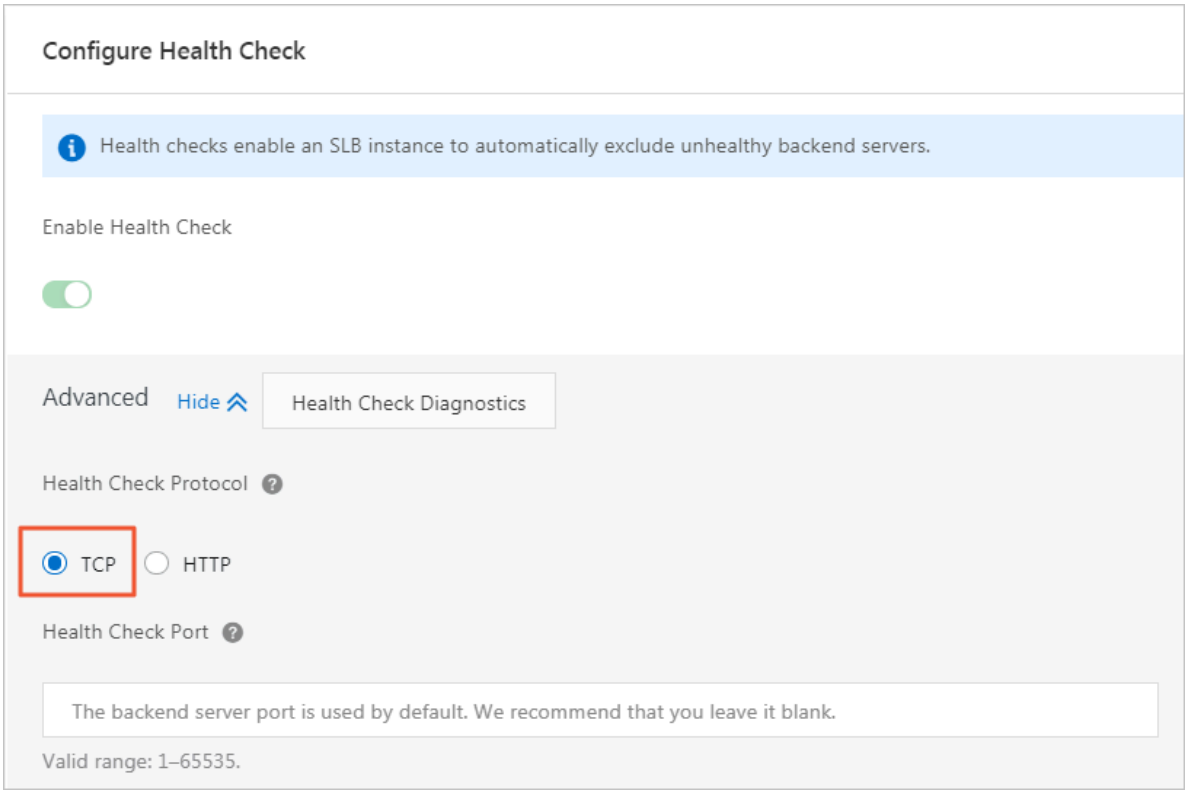
Layer-4 health checks are preformed through TCP three-way handshakes and generate no application logs. If you change Layer-7 SLB to Layer-4 SLB, the number of application logs can be reduced.

Potential risks

After you change the Layer-7 SLB to Layer-4 SLB, SLB checks only the status of the listener port and does not check the HTTP status. In this way, SLB cannot detect the exceptions occurring to HTTP applications in real time.

Procedure

1. Log on to the *SLB console*.
2. On the Server Load Balancer page, click the ID of the target SLB instance.
3. Click the Listeners tab, find the target listener, and click Configure in the Actions column.
4. On the Configure Listener page, click Next and then click Next again to go to the Health Check tab.
5. Change the Health Check Protocol to TCP.



6. Click OK.

Disable application logs on the health check page

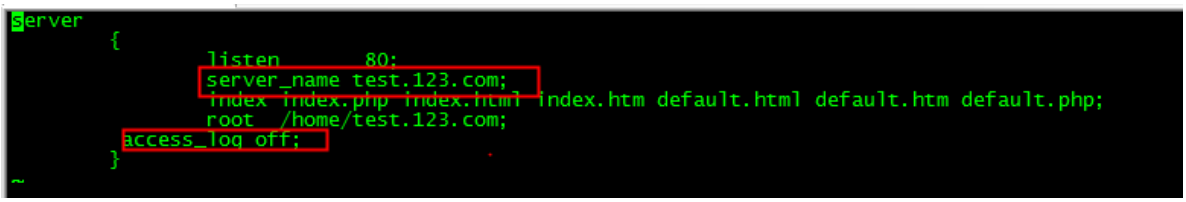
You can configure an independent site for health checks and disable application logs of this site. This method can also reduce the number of health checks. For example, the service site is abc.123.com. You can use test.123.com as the health check site and disable logs of test.123.com.

Potential risks

If the health check site is running normally, but an exception occurs to the service site, health checks cannot detect the exception of the service site.

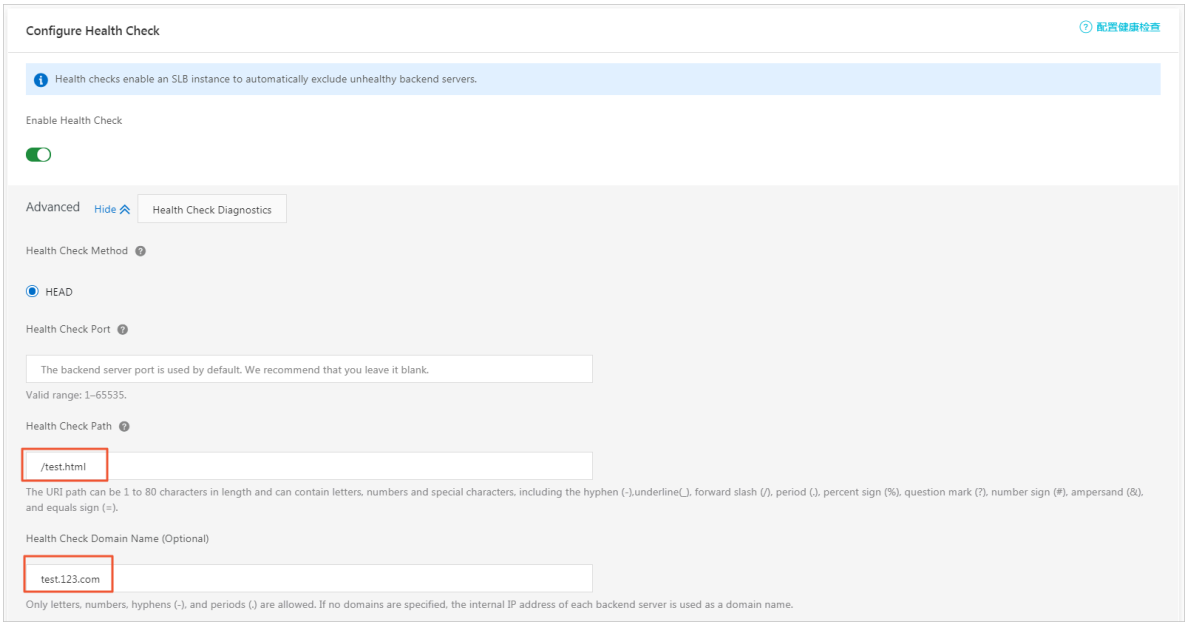
Procedure

1. Create a new health check site and health check page on the backend server and disable logs. In this example, NGINX is used.



2. Log on to the *SLB console*.
3. On the Server Load Balancer page, click the ID of the target SLB instance.
4. Click the Listeners tab, find the target listener, and click Configure in the Actions column.
5. On the Configure Listener page, click Next and then click Next again to go to the Health Check tab.

6. In the Health Check Domain Name field, enter the domain name of the health check site. In the Health Check Path field, enter the path of the health check page.



7. Click OK.

5 What can I do if my ECS instance is declared unhealthy after I enable health checks for Server Load Balancer?

After you enable health checks of Server Load Balancer (SLB), when one backend ECS instance is declared as unhealthy, requests are forwarded to other normal ECS instances. When the faulty ECS instance becomes normal, SLB forwards requests to the ECS instance again.

For layer-7 SLB service, when an ECS instance is declared as unhealthy, you can troubleshoot problems from the following aspects:

- **Make sure you can directly access your service through the ECS instance.**
- **Make sure the backend port you configured in the listener is opened on the backend server.**
- **Check whether the backend ECS instance has installed a firewall or other security protection software. This type of software may block the local IP address of the SLB service, and thus disable the communication between the SLB service and the backend server.**
- **Check whether the SLB health check parameters are correctly set. We recommend that you use default health check settings.**
- **We recommend that you use a static page for health checks. If the static page you use is not the default health check page of the backend ECS instance, you must set this page as the health check page in health check configurations. We recommend that you use a simple HTML page for health checks and use the page only for checking health check responses. We do not recommend that you use dynamic scripting languages such as php.**
- **Check whether the backend ECS instance has high loads, which can slow the response speed of the ECS instance.**

Besides, because the layer-7 SLB service communicates with the backend ECS instance through an internal network, the ECS instance must listen to the internal network or all-network ports. You can check the ECS instance with the following methods:

1. **Check whether the listening function is normal.**

   Assume that the frontend port of SLB and backend port of the ECS instance are both 80. The ECS internal IP address is 10.11.192.1. Run the following command on the server. If you can see the monitoring information of 10.1.1.192.1: 80, or the monitoring information of 0.0.0.0: 80, the listening function of the ports is normal.

   - **Windows server:** `netstat -ano | findstr :80`
   - **Linux server:** `netstat -anp | grep :80`

2. **Check whether the internal network firewall of the server allows port 80. You can disable the firewall temporarily to do a test. Enter the following command to disable the firewall.**

   - **Windows:** `firewall.cpl`
   - **Linux:** `/etc/init.d/iptables stop`

3. **Check whether the backend port is normal.**

   - **For layer-4 SLB service, you can perform a telnet test. If you receive responses, the backend port is normal. Example: Use** `telnet 10.11.192.1 80` **to test.**
   - **For layer-7 SLB service, you can determine whether the port is normal by checking the HTTP status code received. The HTTP status code must be a status code that indicates a normal condition, such as 200. The test methods are as follows:**

     - **Windows: Access the internal IP address of the ECS instance. In this example, access** `http://10.11.192.1`**.**
     - **Linux: Run the** `curl -I` **command and check whether the status is HTTP/1.1 200 OK. In this example, run** `curl -I 10.11.192.1`**.**

6 How do I troubleshoot health check exceptions of a layer-4 (TCP/UDP) listener?

This topic describes how to troubleshoot a health check exception of a layer-4 (TCP/UDP) listener. The health check function is used to determine whether your backend servers are healthy. When a health check exception occurs, it generally means that your backend server is unhealthy. The exception may also be caused by incorrect health check configurations.

Procedure

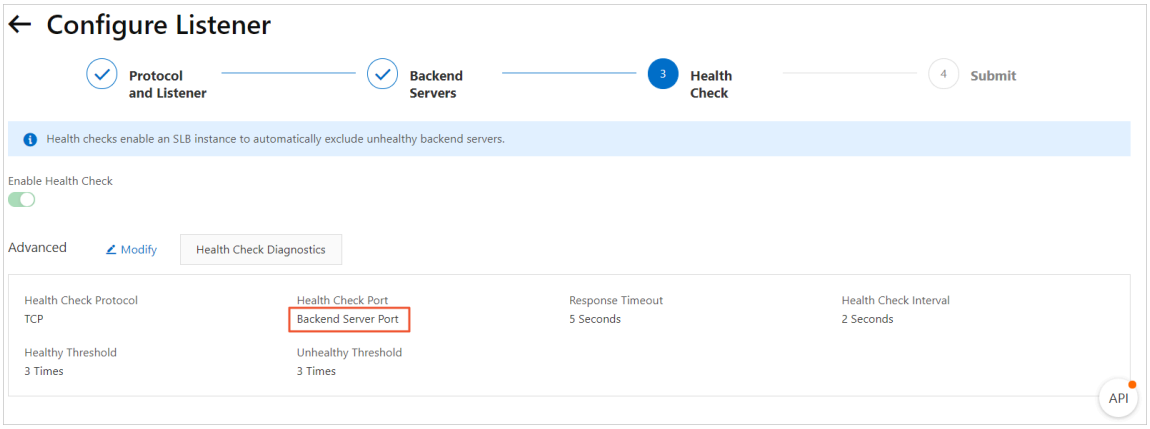
1. **Make sure that the backend server does not block the CIDR block 100.64.0.0/10 through iptables or other third-party firewalls or security software.**

   Server Load Balancer (SLB) communicates with backend servers by using IP addresses in the reserved CIDR block 100.64.0.0/10. If the CIDR block is blocked, health check exceptions occur and SLB cannot work normally.

2. Run the `telnet` command to test the backend server.

   a) Log on to the *SLB console* and check the health check configurations.

   By default, the port of the backend server is used as the Health Check Port. You can also set the port manually. In this example, the port of the backend server, namely port 80, is used.



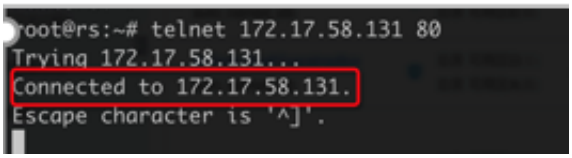
   b) Run the following command to connect the health check port. The health check port configured on the SLB instance must be the same as the listening port on the backend server.

```
telnet 172.17.58.131 80
```

   In this example, *`172.17.58.131`* is the internal IP address of the backend server, and *`80`* is the health check port. By default, the port of the backend server is used as the health check port. You can configure the health check port according to your actual situation.

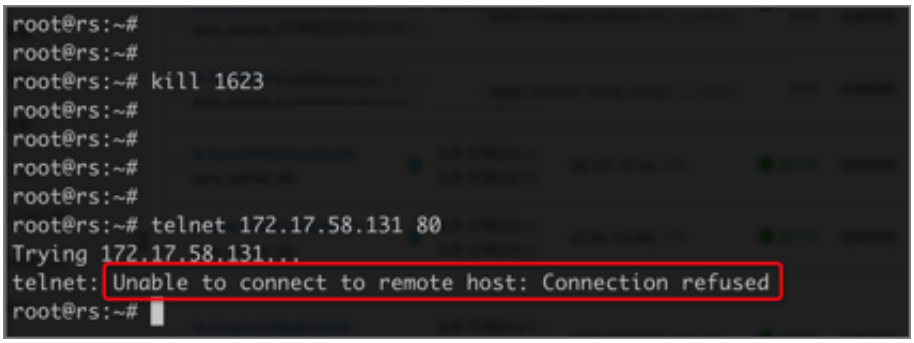
   - In normal conditions, `Connected to xxx.xxx.xxx.xxx` is returned.

     This indicates that the port on the backend server is working (listening) normally and the health check succeeds, as shown in the following figure.



   - Exception example: Assume you do not change the listener configurations of the SLB instance but stop the listening process of port 80 on the backend server. Then, if you run the `telnet` command, the system prompts that the

host cannot be connected. This means that a health check exception occurs if the listening process of port 80 stops, as shown in the following figure.



3. **Optional: Layer-4 listeners support HTTP health checks. If you use HTTP health checks, see** *How do I troubleshoot a health check exception of a layer-7 (HTTP/HTTPS) listener?* **for troubleshooting.**

7 How do I troubleshoot a health check exception of a layer-7 (HTTP/HTTPS) listener?

This topic describes how to troubleshoot a health check exception of a layer-7 (HTTP/HTTPS) listener. The health check function is used to determine whether your backend servers are healthy. If a health check exception occurs, it generally means that your backend server is unhealthy. The exception may also be caused by incorrect health check configurations.

Procedure

1. **Make sure that the backend server does not block the CIDR block 100.64.0.0/10 through iptables or other third-party firewalls or security software.**

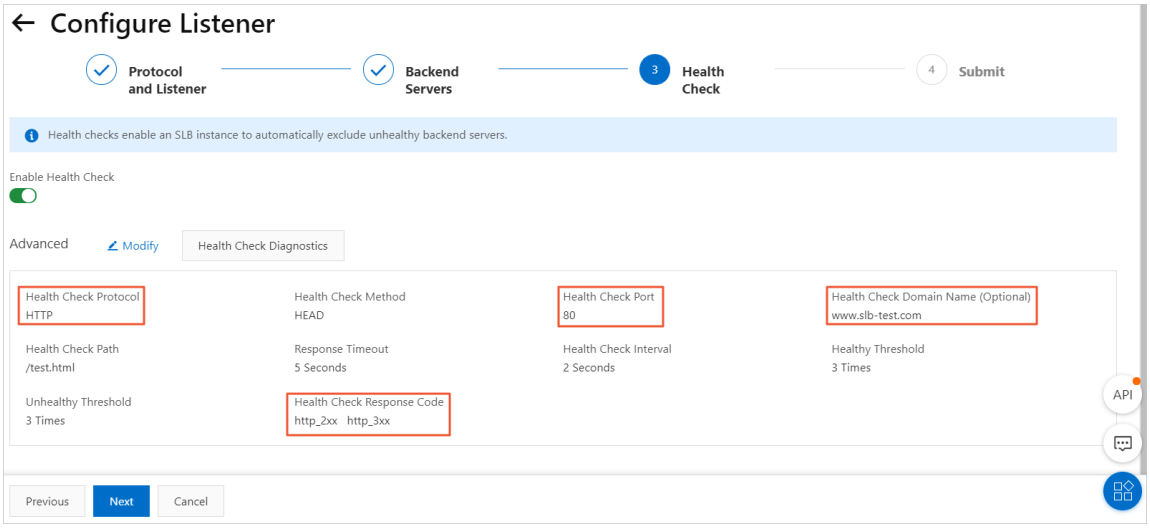
   Server Load Balancer (SLB) communicates with backend servers by using IP addresses in the reserved CIDR block 100.64.0.0/10. If the CIDR block is blocked, health check exceptions occur and SLB cannot work normally.

2. **Access the HTTP service on the backend server from the backend server to check whether the HTTP service works normally.**

   a) **Log on to the SLB console and check the health check configurations on the listener details page.**

   In this example, an HTTP listener is used and the internal IP address of the backend server with the health check exception is 10.0.0.2. Other health check configurations are as follows:

   - **Health Check Port: 80**
   - **Health Check Domain Name (Optional):** `www.slb-test.com`
   - **Health Check Path:** `/test.html`



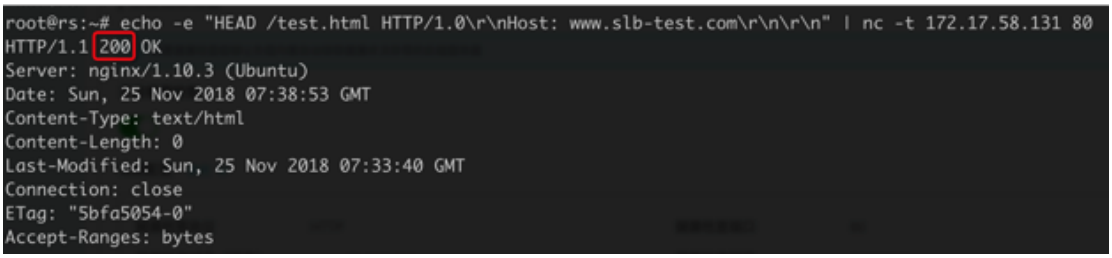
   b) **For a Linux server, run the** `nc` **or** `curl` **command to test the HTTP service on the backend server. Make sure that the configurations of health check path, health**

check port, and health check domain name are the same for the HTTP service and the backend server. Otherwise, a health check exception occurs.

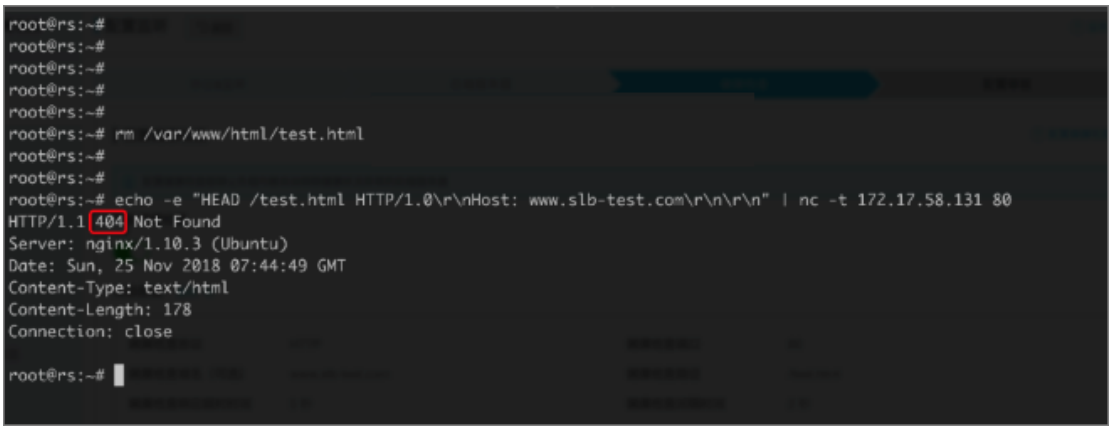
In this example, the `nc` command is used. Configure the health check path, health check domain name, public IP address, and health check port according to your actual situation.

```
echo -e "HEAD /test.html HTTP/1.0\r\nHost: www.slb-test.com\r\n\r
\n" | nc -t 172.17.58.131 80
```

- **In normal conditions, `200` or `2xx/3xx` status codes are returned, as shown in the following figure.**



- **Assume you do not change the listener configurations of the SLB instance but delete the /test.html page on the backend server. Then, when you run the `nc` command, the error code 404, instead of 2xx or 3xx, is returned, indicating a health check exception, as shown in the following figure.**



8 How do I troubleshoot HTTP 5xx errors?

After a Server Load Balancer (SLB) instance is configured, errors such as 500 Internal Server Error, 502 Bad Gateway, and 504 Gateway Timeout may occur. These errors can be caused by the blockage of the service provider, Alibaba Cloud blockage caused by abnormal client activities, wrong configurations of the SLB instance, health check failures, or failures in accessing web applications on the backend ECS instances.

This topic lists the causes, resolutions, and troubleshooting steps of these problems .

1. *Possible causes and resolutions*
   - *The source site domain name is not put on record or it is not configured with any Layer-7 forwarding rule in Anti-DDoS Pro or security network.*
   - *The source IP address of the client is blocked by Alibaba Cloud Security.*
   - *The source IP address is blocked by the security protection software of the backend ECS instance.*
   - *Parameters of the Linux kernel of the backend ECS instance are configured wrong.*
   - *The performance of the backend ECS instance reaches a bottleneck.*
   - *SLB reports 502 errors due to health check failures.*
   - *The health check is normal but the web application reports 502 errors.*
   - *The HTTP header is too long.*
2. *Troubleshooting*
3. *Open a ticket*

Potential causes and resolutions

1. **The source site domain name is not put on record or it is not configured with any layer-7 forwarding rule in Anti-DDoS Pro or security network.**

   Resolution: Put the domain name on record. If the SLB instance is in Anti-DDoS Pro or security network, configure corresponding domain name-based forwarding rules.

2. **The source IP address of the client is blocked by Alibaba Cloud Security.**

   Test if the same problem occurs to clients of other service providers. If not, the problem is generally caused by the blockage of the service provider.

   Resolution: Open a ticket to Alibaba Cloud who decides if blockage has occurred through packet capture. If blockage exists, contact the service provider to solve the problem.

3. **The source IP address is blocked by the security protection software of the backend ECS instance.**

   100.64.0.0/10 (100.64.0.0/10 is reserved by Alibaba Cloud. No other users can use it and security risks are avoided.) is the IP address range of SLB servers and is mainly used for health checks and request forwarding. If security software or a firewall inside the system is applied, add the IP address range to the whitelist to avoid 500 or 502 errors.

   Resolution: Add the SLB IP address range to the whitelist of the antivirus or firewall software, or unload the software to test if the problem is caused by the blockage of the software.

4. **Parameters of the Linux kernel of the backend ECS instance are configured wrong.**

   If the backend ECS instance uses the Linux system, you can disable the rp_filter parameters in system kernel when you change the layer-7 listener to a layer-4 listener.

   Resolution: Set the values of the following parameters in the system configuration file `/etc/sysctl.conf` **to zero, and then run** `sysctl -P`**.**

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

5. **The performance of the backend ECS instance reaches a bottleneck.**

   High CPU utilization or no extra bandwidth may cause access exceptions.

   Resolution: Check the performance of the backend ECS instance and solve performance bottlenecks. If the overall system capacity is insufficient, you can increase the number of backend ECS instances.

6. **SLB reports 502 errors due to health check failures.**

   For more information about health check failures, see *Resolve health check failures*.

   Also, 502 errors occur if the health check function of SLB is disabled and the web service in the backend server cannot process HTTP requests.

7. **The health check is normal but the web application reports 502 errors.**

   If the 502 Bad Gateway error message indicates that SLB can forward requests from the client to backend servers, but the web application in the backend ECS instance cannot process the requests, you must check the configurations and running status of the web application in the backend server. For example, the time used by the web application to process HTTP requests exceeds the timeout value of SLB.

   For layer-7 listeners, if the time used by the backend server to process PHP requests exceeds proxy_read_timeout of 60 seconds, SLB reports 504 Gateway Timeout. For layer-4 listeners, the timeout value is 900 seconds.

   Resolution: Make sure that the web service and related services run normally. Check if PHP requests are processed properly, and optimize the processing of PHP requests by the backend server. Take Nginx+php-fpm as an example:

   a. **The number of PHP requests being processed has reached the limit.**

      If the total number of PHP requests being processed in the server has reached the limit set by max_children in php-fpm, and more PHP requests are being sent to the server, 502 or 504 errors may occur:

      - **If existing PHP requests in the backend server are processed timely and new PHP requests can be processed, no error occurs.**
      - **If existing PHP requests are not processed timely and new PHP requests must wait to be processed, when the value of fastcgi_read_timeout of Nginx is exceeded, a 504 Gateway Timeout error occurs.**
      - **If existing PHP requests are not processed in a timely manner and new PHP requests must wait to be processed, when the value of request_terminate_timeout in Nginx is exceeded, a 502 Bad Gateway error occurs.**

   b. **If the PHP script execution time exceeds the limit, namely, the time used by php-fpm to process PHP scripts exceeds the value of**

request_terminate_timeout in Nginx, a 502 error occurs and the following error log is shown in Nginx logs:

```
[error] 1760#0: *251777 recv() failed (104: Connection reset by
peer) while reading response header from upstream, client: xxx.xxx
.xxx.xxx, server: localhost, request: “GET /timeoutmore.php HTTP/1
.1”, upstream: “fastcgi://127.0.0.1:9000”
```

c. The health check is performed on static pages. Errors occur when exceptions are detected in the process handling dynamic requests. For example, php-fpm is not running.

8. The HTTP header is too long.

   An HTTP header that is too long may make SLB unable to process relevant data, resulting in 502 errors.

   Resolution: Decrease the amount of data transmitted by the header or switch to the TCP listener.

9. The service access logic is inappropriate.

   Make sure that no backend ECS instance in SLB accesses the public IP address of SLB. If the backend server accesses its own port through the IP address of SLB, the requests may be scheduled to the server itself based on the scheduling rules of SLB. This can lead to an infinite loop, thus resulting in 500 or 502 errors.

   Resolution: Make sure that SLB is correctly used and no backend ECS instance is accessing the public IP address of SLB.

Troubleshooting

- Check the screenshot of 500, 502, or 504 error to determine the cause of the error. The cause of the error could be with SLB, Anti-DDoS or security network, or backend ECS instance configurations.
- If Anti-DDoS or security network is used, make sure that the layer-7 forwarding rules are correctly configured.
- Check whether the problem occurs in all clients. If not, check whether the client indicating an error has been blocked by Alibaba Cloud Security. Also, check whether the domain name or IP address of SLB is intercepted by the service provider.
- Check the status of SLB and whether the health check failed in a backend ECS instances. If the health check of a backend ECS instance failed, resolve the detected health check failure.

- **Associate the service address of SLB with the IP address of the backend server by using the hosts file on the client. If a 5xx error occurs at intervals, the error is probably caused by inappropriate configurations of a backend ECS instance.**
- **Change the layer-7 SLB instance to a layer-4 SLB instance to see whether the problem occurs again.**
- **Check the performance of backend ECS servers and whether performance bottlenecks of the CPU, memory, disk, or bandwidth exist.**
- **If it is determined that the error is due to the backend server, check the web server logs of the backend server. Check whether the web service is running normally and whether the web access logic is correct. Test the server by uninstalling anti-virus software on the server and restarting the server.**
- **Check whether the TCP kernel parameters of the Linux system on the backend ECS instance are correctly configured.**

Open a ticket

Perform the preceding troubleshooting procedures step by step and record the test results in detail. Provide the test results when you open a ticket so that the technical support can help you solve the problem as soon as possible.

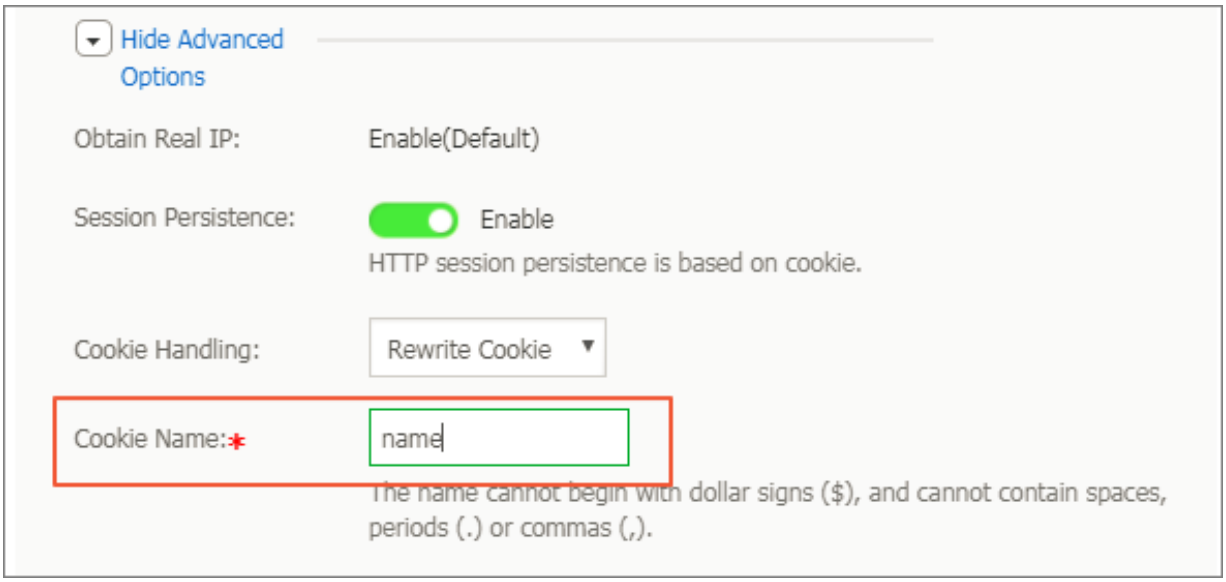
If the problem persists, consult Alibaba Cloud after-sales technical support.

9 Configure cookie in the backend server

Server Load Balancer provides session persistence function. With session persistence enabled, Server Load Balancer can distribute requests from the same client to the same backend server during the session period.

For layer-4 listeners, session persistence is based on the IP address. The listener of Server Load Balancer forwards requests from the same IP address to the same backend server. For layer-7 listeners, session persistence is based on cookies.

If you choose to rewrite the cookie, you must configure the cookie on the backend server. Suppose there are two domain names under your Server Load Balancer service: vip.a.com and img.a.com. If you want to configure session persistence for vip.a.com, you can set the cookie name to name, and set a cookie of which the key is name for vip.a.com on the backend server.



Follow the instructions in this section to set cookies on a backend server.

Apache

1. **Open the** *httpd.conf* **file and make sure that the following line is not commented.**

```
LoadModule usertrack_module modules/mod_usertrack.so
```

2. **Add the following configurations in the VirtualHost file.**

```
CookieName name
CookieExpires "1 days"
```

```
  CookieStyle Cookie
  CookieTracking on
```

Nginx

Configure the cookie as follows.

```
server {
    listen 8080;
    server_name wqwq.example.com;
    location / {
      add_header Set-Cookie name=xxxx;
        root html;
        index index.html index.htm;
    }
}
```

Lighttpd

Configure the cookie as follows.

```
    server.modules  = ( "mod_setenv" )
    $HTTP["host"] == "test.example.com" {
          server.document-root = "/var/www/html/"
          setenv.add-response-header = ( "Set-Cookie" => "name=XXXXXX"
      }
    }
```

10 Session persistence FAQ

1. What is session persistence?

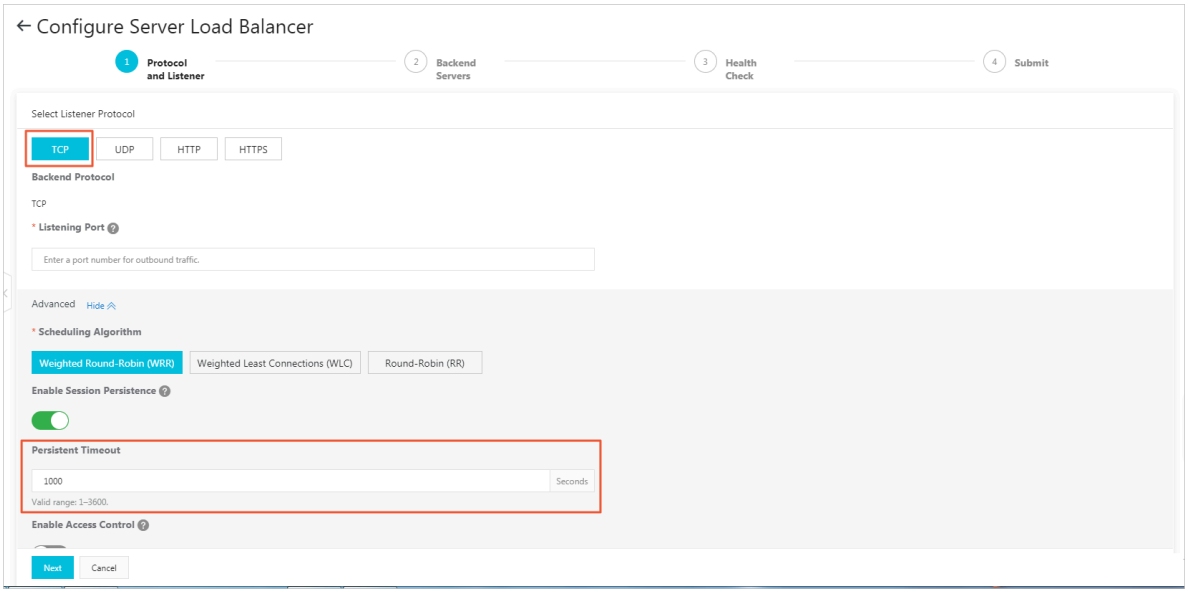
   Session persistence serves to forward session requests from the same client to a specified backend server for processing.

2. How can I enable session persistence?

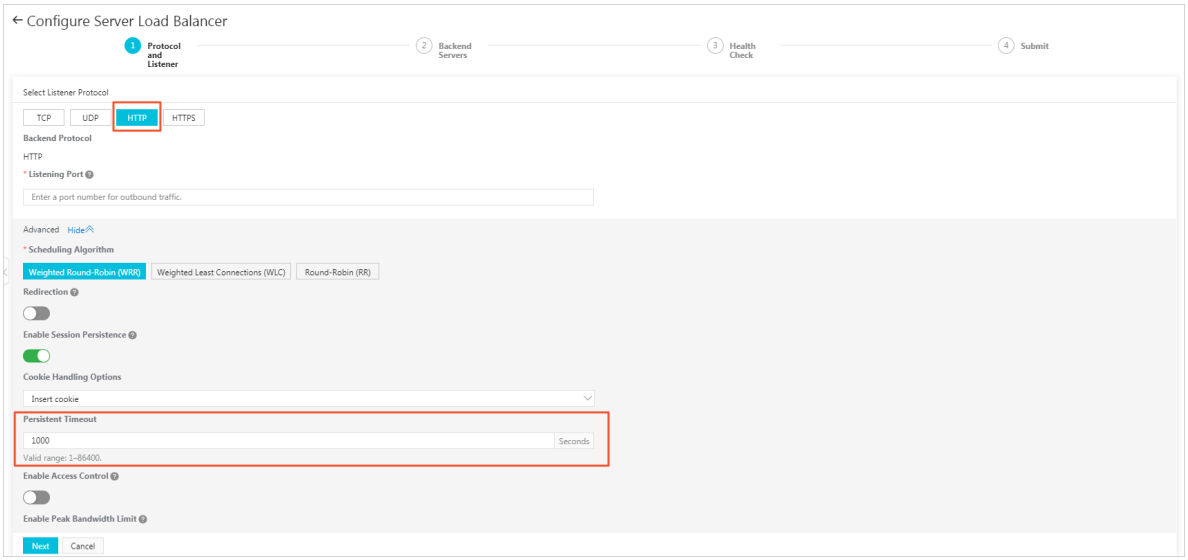
   You can choose whether to enable the session persistence function when configuring listeners. You can configure different session persistence policies for different listeners. The maximum session persistence duration is 86,400 seconds (24 hours).

3. What type of session persistence does SLB support?

   - **For Layer-4 (TCP protocol) services, session persistence is based on source IP addresses. The maximum duration of Layer-4 session persistence is 3,600 seconds.**



- **For Layer-7 (HTTP or HTTPS) services, session persistence is based on cookies. The maximum duration of session persistence based on cookie inserting is 86,400 seconds (24 hours).**



4. What kind of cookie configurations are supported?

HTTP/HTTPS listeners can use cookie inserting and rewriting methods to achieve session persistence.

- **Cookie inserting: When this method is used, you only need to specify the cookie timeout. For the first access by a client, SLB inserts a cookie (inserts a SERVERID string in the HTTP/HTTPS response message) in the response. The next request from the client will contain this cookie and SLB will forward the request to the same ECS instance.**
- **Cookie rewriting: When this method is used, you can specify the cookie to be inserted in the HTTPS/HTTP response as needed. You must maintain the timeout and TTL of the cookie in the backend ECS instance. SLB rewrites the original cookie when it discovers a customized cookie. The next request from the client will contain this rewritten cookie and SLB will forward the request to the same ECS instance. For more information, see** *Configure cookie in the backend server*.

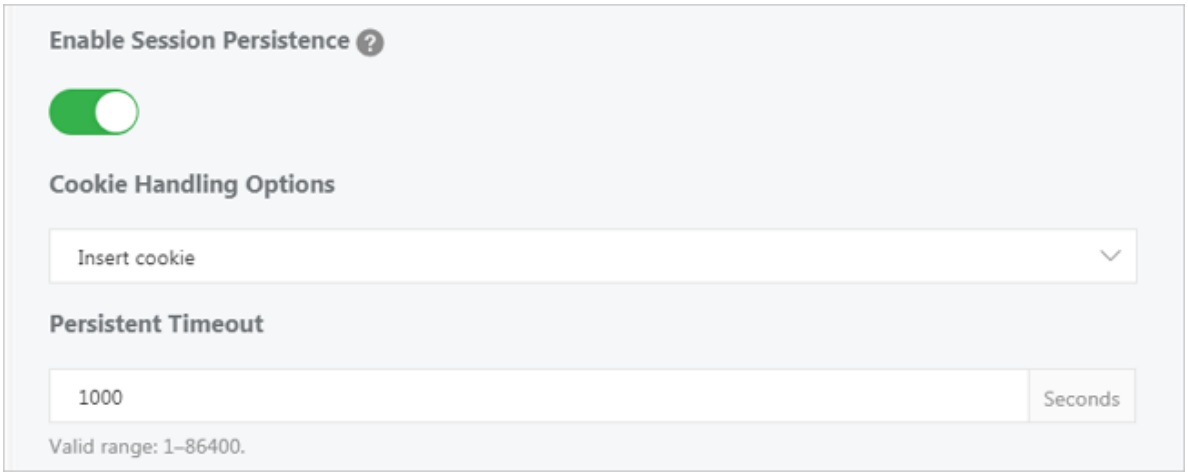
5. Can I configure different session persistence rules for different domain names?

Yes.

You can configure different session persistence rules by using the cookie rewriting method.

6. What timeout value should I set for a cookie?

- **For cookie inserting, you can set a timeout value from 1 to 86400 seconds on the SLB console.**



- **For cookie rewriting, you must maintain the timeout value on the backend ECS instance.**

7. How to check the session persistence string?

You can use developer tools in the browser to view whether the response message contains the SERVERID string or a user-specified keyword, or you can run `curl www.xxx.com -c /tmp/cookie123` to save the cookie and then run `curl www.xxx.com -b /tmp/cookie123` to initiate the access.

8. Why the session persistence does not work sometimes?

- **Check whether the session persistence function has been enabled in the listener configuration.**
- **HTTP/HTTPS listeners cannot insert session persistence cookies to the response messages containing 4xx response codes that are returned by backend servers.**

  Resolution: Use TCP listeners instead. TCP listeners achieve session persistence based on the source IP address of the client. Additionally, you can configure cookies for the backend ECS instances and add cookie detection logic to guarantee that session persistence works.

- **302 redirection changes the SERVERID string.**

  If a 302 redirection packet is returned by the backend servers when SLB is inserting a cookie, the SERVERID string in session persistence will be changed. Therefore, session persistence fails to work.

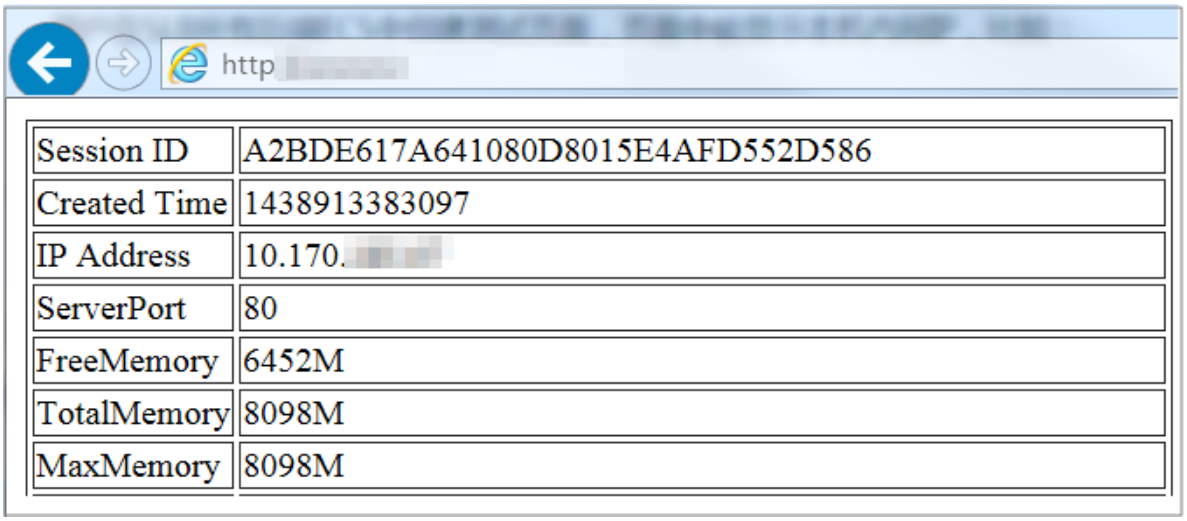
  Troubleshoot: Capture the request and returned response on the browser or use a tool to capture packets to check whether a 302 response message is returned. Then, compare the SERVERID strings in the packets to see if they are different.

  Resolution: Use TCP listeners instead. TCP listeners achieve session persistence based on the source IP address of the client. Additionally, you can configure cookies for the backend ECS instances and add cookie detection logic to guarantee that session persistence works.
- **Too short session persistence duration will also cause session persistence failure.**

9. How can I use Linux curl to test the session persistence?

1. **Create test pages.**

   Create test pages on all the backend ECS instances. The local intranet IP address is displayed, as shown in the following figure. The intranet IP address is used to verify the backend server to which client requests are distributed. Observe the consistency of this IP address to check whether session persistence works.



2. **Perform curl test in a Linux environment.**

   Assume that the IP address of an SLB instance is 1.1.1.1, and the URL of the created test page is *http://1.1.1.1/check.jsp*.

   a. **Log on to the Linux server used for test.**

   b. **Run the following command to obtain the cookie.**

   ```
   curl -c test.cookie http://1.1.1.1/check.jsp
   ```

   Note:

   The default session persistence method of SLB is cookie inserting, and the curl test does not save or send a cookie. Therefore you must save the cookie for test first. Otherwise, the curl test result is random. As a result, you will consider that session persistence does not work by mistake.

   c. **Run the following command to test session persistence.**

   ```
   for ((a=1;a<=30;a++)); 2="" do="" curl="" -b="" 1.cookie=""
              check.jsp="">/dev/null | grep '10.170.*';sleep 1; done
   `
   ```

   Note:

   a<=30 is the number of tests to do, you can change the number as needed. grep ‘10.170. *' is the IP address to display, you can change it ccording to the intranet IP address of the backend ECS instance.

   d. **Observe the IP addresses returned in the preceding tests. If they are the intranet IP address of the same ECS instance, then session persistence works; otherwise, there is something wrong with the SLB session persistence.**