

Server Load Balancer

Troubleshoot

Troubleshoot

Resolve access failures

If an error occurs accessing Server Load Balancer (SLB), potential causes and steps you can take to resolve the issues are as follows:

Cause: No listeners are configured after creating an SLB instance.

Resolution: You must configure listeners for your instance. To configure listeners, see [Listeners overview](#). If you do not configure listeners for a created SLB instance, the client cannot access the SLB instance.

Cause: Incorrect backend Linux ECS kernel configuration of the Layer-4 SLB instance.

The `rp_filter` feature of the Linux kernel must be disabled for the backend Linux ECS instances added to a Layer-4 SLB instance. Otherwise, telnet to the service port of SLB from the frontend may fail, while the instance is indicated as healthy.

The `rp_filter` feature of the Linux kernel is used for implementing Unicast Reverse Path Forwarding (URPF). The `rp_filter` verifies the direction of reverse-path data packets to avoid attacks using a forged IP address. However, this feature may conflict with the underlying Linux Virtual Server (LVS) routing policy of SLB and result in access exceptions.

Resolution: Make sure that the values of the following three parameters in the system configuration file of the Linux ECS server are set to zero. Edit `/etc/sysctl.conf` and then run `sysctl -p` to make the configuration take effect.

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

Cause : Incorrect backend Windows ECS configuration of the Layer-4 SLB instance.

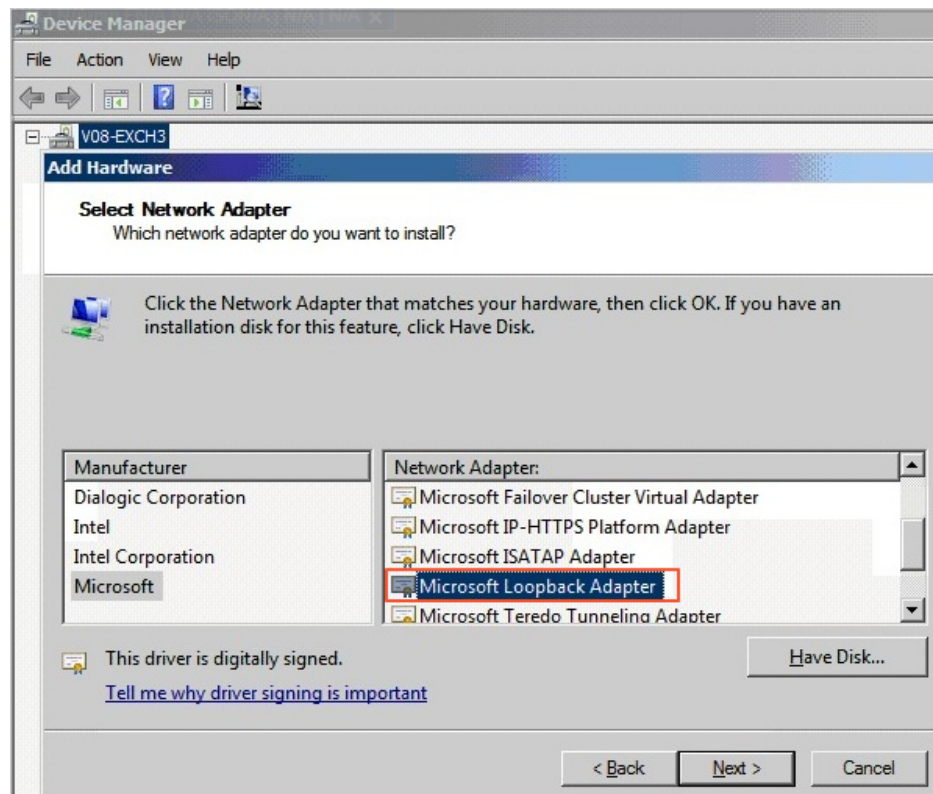
For the Layer-4 SLB, a backend ECS instance cannot act as both the real server and the client

to access the SLB instance.

This may cause the related access requests to be forwarded to the same ECS instance, resulting in a data access loop. The ECS instance will fail to access the SLB instance.

Resolution:

Install Windows Loopback Adapter: Right-click **Computer** > **Properties**. On the **Control Panel** page, click **Device Manager** > **Add Hardware** > **Install hardware that I manually select from the list** > **Show All Devices**. Then select and install the device shown in the following figure.



Enable Weak Host Model and run the following command to check the Idx of all network interfaces.

```
netsh interface ipv4 show interface
```

Configure weakhostsend=enabled, and weakhostsend=enabled for all network interfaces. For example, configure the adapter with the Idx of 12 as follows:

```
netsh interface ipv4 set interface 12 weakhostsend=enabled
netsh interface ipv4 set interface 12 weakhostreceive=enabled
```

Cause : Exceptions in the local network of the client or the intermediate link of the carrier.

For the Internet SLB, exception in the network of the client or in the network of the carrier may also lead to the failure of client access.

Resolution: Test the network access of SLB in different regions and different network environments. If the access fails only when accessing SLB from your local network, the access failure may be caused by your local network. You can do further troubleshooting and analysis by using methods such as continuous ping tests or MTR route tracking.

Resolve 500/502/504 errors

If errors 500 Internal Server Error, 502 Bad Gateway, and 504 Gateway Timeout occur when attempting to access a service through SLB (SLB), steps you can take to resolve the issues are as follows:

Potential causes and resolutions

- Blocked by the security protection software of the backend ECS instance
- Parameter error of the Linux kernel of the backend ECS instance
- Performance bottleneck of the backend ECS instance
- SLB reports 502 error due to health check failure
- The health check is normal but the web application reports 502 error
- The HTTP header is too long
- Problem of service access logic

2. Troubleshooting procedures

3. Open a ticket

Potential causes and resolutions

Cause: Blocked by security protection software of the backend ECS instance.

The IP address ranges 100.64.0.0/10, 10.158.0.0/16, 10.159.0.0/16 and 10.49.0.0/16 are used by SLB to perform health checks and forward requests.

Resolution: You can add these IP address ranges to the firewall exceptions to prevent 500 or 502 errors.

Cause: Parameter error of the Linux kernel of the backend ECS instance

Resolution: If the backend ECS instance is using the Linux system, disable the `rp_filter` feature in the system kernel parameters when changing the Layer-7 listener to the Layer-4 listener.

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
```

Cause: Performance bottleneck of the backend ECS instance

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

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

Cause: SLB reports 502 error due to health check failure.

A 502 error may occur if the health check function of SLB is disabled. Then the web service in the backend server cannot process HTTP requests. For more information, see [Resolve health check failures](#).

Cause: The health check is normal but the web application reports 502 errors.

The 502 Bad Gateway error message indicates that SLB can forward requests from the client to the backend servers, but the web application in the backend ECS instance cannot process the requests. Therefore, 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 the `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. For example with the web server of Nginx and php-fpm:

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, then 502 or 504 errors may occur:

If existing PHP requests in the backend server are processed timely, new PHP requests can be processed successively.

If the existing PHP requests are not processed timely, new PHP requests will remain in a waiting mode. If the value of `fastcgi_read_timeout` of Nginx is exceeded, a 504 Gateway Time-out error occurs.

If the existing PHP requests are not processed in a timely manner, new PHP requests will remain in a waiting mode. If the value of `request_terminate_timeout` in Nginx is exceeded, a 502 Bad Gateway error occurs.

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"
```

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.

Cause: 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 change the Layer-7 listener to the Layer-4 listener.

Cause: Problem of service access logic.

Make sure that no backend ECS instance in SLB accesses the public IP of SLB. When 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 will lead to an infinite loop, thus resulting in 500 or 502 error for the requests.

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

Troubleshooting

Troubleshoot 500 errors as follows:

Check the screenshot of 500/502/504 error to determine the cause of the error. The cause of the error could be with SLB, of Anti-DDoS or QuickShield, and/or a problem with backend ECS instance configurations.

If Anti-DDoS or QuickShield 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 of SLB is intercepted by the carrier.

Check the status of SLB and whether there are any health check failures in any backend ECS instances. If so, resolve the detected health check failure.

Bind the service address of SLB to the IP address of the backend server by using the hosts file on the client. If a 5XX error occurs at intervals, it is possible that a backend ECS server is not correctly configured.

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 there is performance bottleneck of the CPU, memory, disk, or bandwidth.

If it is determined that the error is due to the backend server, check whether there are any related errors in web server logs of the backend ECS instance. Check whether the web service is running normally and whether the web access logic is correct. Test 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 troubleshooting procedures step by step and record the test results in detail. Provide the test results when you open the ticket so that our after-sales technical support can help you solve the problem as soon as possible.

Perform stress test

Stress test overview

The Layer-4 Server Load Balancer (SLB) uses open source software Linux Virtual Server (LVS) and Keepalived to achieve load balancing. The Layer-7 SLB uses Tengine to achieve load balancing. For the Layer-4 SLB, requests directly reach backend servers after going through LVS. For the Layer-7 SLB, requests reach backend servers after going through LVS and then through Tengine. The Layer-7 SLB has one more procedure than the Layer-4 SLB when forwarding incoming requests. Due to this additional procedure, the performance of the Layer-7 SLB is less efficient to that of the Layer-4 SLB.

When performing stress tests on a Layer-7 SLB instance, you may find that the performance is much lower than expected. Also, you may experience the performance of the SLB instance with two backend ECS instances is slower to one single backend ECS instance of the same specification. The following are additional possible causes:

Cause: Insufficient client ports.

Lack of client ports may lead to a connection failure especially during stress testing. The SLB erases the timestamp attribute of TCP connection by default, and the `tw_reuse` feature of Linux protocol stack (reuse time_wait connection) cannot take effect. As a result, the `time_wait` connections pile up, leading to insufficient client ports.

Resolution: Use persistent connection instead of short-lived connection on the client. Use the RST message, instead of sending a FIN packet, to disconnect (set the `SO_LINGER` attribute for the socket).

Cause: The accept queue of the backend server is full.

The accept queue of the backend server is full, so that the backend server does not return syn_ack packets and the client times out.

Resolution: The default value of net.core.somaxconn is 128. Run `sysctl -w net.core.somaxconn=1024` to change its value, and restart the application on the backend server.

Cause: Too many connections to the backend server.

Persistent connections change to short-lived connections after going through Tengine when using the Layer-7 SLB due to the architecture design. Therefore, numerous connections are sent to the backend server, resulting in poor performance during stress test.

Cause: Poor application performance.

After SLB forwards requests to the backend server, the load of the backend server is normal, but the performance is not as good as expected because the applications deployed on the backend server depend on other applications such as databases. If the database reaches its performance bottleneck, this may cause poor performance.

Cause: Abnormal health check status of backend servers.

It is easy to ignore the health check status of backend servers when doing stress testing. Health check failure of a backend server or frequent change of the health check status (frequent changes from healthy to unhealthy or from unhealthy to healthy) also results in poor performance.

Suggestions on stress testing

Note the following during the stress testing:

Use short-lived connections to test the forwarding capability of SLB.

In addition to testing the session persistence and balancing test, the main goal of stress testing is to test the forwarding capability of SLB. We recommend that you use short-lived connections to test the processing capabilities of SLB and backend servers. Note the insufficient client ports problem when using the short-lived connections to do the stress testing.

Use persistent connections to test the throughput of SLB. Persistent connection is used to

test the bandwidth limit or special services.

We recommend that you set a smaller timeout value (5 seconds) for the stress test tool. If the timeout value is too large, the test result shows an increased average RT, making it difficult to determine whether the throughput limit of SLB has been reached. When the timeout value is small, the test result shows the success rate, making it easier to quickly determine the throughput limit of SLB.

Use static web pages for stress testing to prevent the application logic cost.

Use the following listener configurations for stress testing:

Do not enable session persistence. This may cause the stress to be concentrated on a particular backend server.

Disable health check of the listeners to reduce the health check requests to the backend servers.

Perform the stress testing on multiple (more than five) clients with dispersed source IP addresses to simulate the actual online situation.

Do not use Apache ab as the stress testing tool.

In high-concurrency scenarios, Apache ab may experience tiered interruptions at 3s, 6s, and 9s. Apache ab determines whether a request is successful based on the content length. When multiple backend servers are added to SLB, the returned content length will be inconsistent thereby the test result is adversely affected.

Resolve health check failures

If a backend ECS instance is declared as unhealthy by the health check, Server Load Balancer (SLB) stops distributing requests to the unhealthy ECS instance. SLB will then distribute requests to other healthy ECS instances and will again distribute requests to this ECS instance when it becomes healthy.

For the Layer-7 SLB, use these techniques to troubleshoot the backend ECS instance if its health check is detected abnormal:

Check whether you can access the application service through the backend ECS instance

directly.

Check whether the port opened on the backend server is the same as that configured in the listener.

Check whether the backend ECS instance has installed a firewall or other security protection software, blocking the IP address of SLB. The SLB system cannot communicate with the backend server if its IP address is blocked.

Check whether the health check configurations of SLB are correct. We recommend that you use the default values.

Check whether the web page used for the health check is a static page. If it is not the default home page of the backend ECS instance, you have to specify the URL of the web page in health check configurations. We recommend that you use a simple HTML page for health checks used only for checking the returned results. We recommend that you do not use dynamic scripting languages such as php.

Check whether there is a large load on the backend ECS instance. A large load may lower the response speed of the instance.

Follow these steps to check if any backend ports are open:

Check whether a port is being listened.

Assume that the frontend port of SLB is 80, the backend port on the ECS instance is 80, and the intranet IP address of the ECS instance is 10.11.192.1. Run the following command on the server. If you can see the listening information of 10.11.192.1:80 or 0.0.0.0:80, then the port is being listened.

For Windows system: `netstat -ano | findstr :80`

For Linux system: `netstat -anp | grep :80`

Check whether the intranet firewall of the server allows port 80. You can temporarily close the firewall for testing. Run the following command to close the firewall:

For Windows system: `firewall.cpl`

For Linux system: `/etc/init.d/iptables stop`

Check whether the backend port is normal.

For Layer-4 SLB, use telnet to test. The port is normal if there is a response. For example, run telnet 10.11.192.1 80.

For Layer-7 SLB, the HTTP status code must be a status code indicating the normal status, such as 200. Follow these methods to check:

For Windows systems: Enter the intranet IP address in the browser of the backend ECS instance to check whether the backend ECS instance can provide services over the intranet. In this example, the intranet IP address is: `http://10.11.192.1`.

For Linux systems: Use the curl -I command to check whether the response is HTTP/1.1 200 OK. In this example, the curl -I command is: `curl -I 10.11.192.1`.

Manage excessive health check logs

Server Load Balancer (SLB) automatically stores health check logs for the most recent three days. The quantity of health check logs may present concerns to your Operations & Maintenance (O&M). You can reduce the number of health check logs or prevent the logs from generating in certain scenarios by using the following methods.

Note: If the number of health check logs is reduced, you may be unable to identify problems that occurred. Evaluate the risks and configure as necessary.

- Filter HEAD requests
- Adjust the health check frequency
- Disable the health check of Layer-7 SLB
- Change Layer-7 SLB to Layer-4 SLB
- Disable application logs on the health check page

Filter HEAD requests

HTTP health check uses HEAD method by default, filter out all HEAD requests, you can get the access

logs.

Adjust the health check frequency

You can increase the health check intervals to reduce the health check frequency and the number of logs generated.

Potential risks

If you increase the health check intervals, and a backend ECS instance fails, it may take a longer time for SLB to detect the faulty backend ECS instance.

Procedure

Log on the Server Load Balancer console.

On the **Instances** page, find the target SLB instance and click **Manage**.

lb-3c111e7a...	cn-hangzhou-b(Master)	114.58(Public IP)	Running	Classic Network	TCP : 80	Normal	k8s-for-cs-c3da...	Shared-Performance Instance	Pay by Traffic	Pay-As-You-Go	2017-12-18 15:04:28 Created	Manage	More+
----------------	-----------------------	-------------------	---------	-----------------	----------	--------	--------------------	-----------------------------	----------------	---------------	-----------------------------	--------	-------

Click **Listeners**. Find the target listener and click **Configure**.

Front-end Protocol/Port	Backend Protocol/Port	Status	Forwarding Rules	Session Persistence	Health Check	Peak Bandwidth	Server Group	Actions
TCP: 80	TCP: 32000Normal	Running	Weighted Round Robin	Disable	Enable	50 Mbps	-	Configure Details More+

Click **Next** in the **Listener Configuration** dialog box to configure the health check.

Adjust the **Health Check Interval**, in the range of 1-50 seconds. When you set a longer interval, the health check frequency becomes lower and the number of logs generated by backend ECS instances are reduced accordingly. Change the health check interval as necessary.

Add Listener

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

Health Check Method: ☒ TCP ☐ HTTP

Health Check Port: Port range is 1-65535.
By default, the backend server's port is used for health checks.

☐ Hide Advanced Options

Response Timeout: 5 Second(s)
The amount of time to wait for the response from a health check. The range is 1-300 seconds. The default is 5 seconds.

Health Check Interval: 2 Second(s)
The amount of time between two consecutive health checks. The range is 1-50 seconds. The default is 2 seconds.

Unhealthy Threshold: 3
The number of consecutive health check failures on the ECS instances (from success to failure).

Healthy Threshold: 3
The number of consecutive health check successes on the ECS instances (from failure to success).

Back **Confirm** Cancel

Click **Confirm** to complete the modification.

Disable the health check of Layer-7 SLB

For Layer-7 (HTTP/HTTPS) SLB, the health check is achieved by HTTP Head requests. Application logs of backend ECS instances record health check requests and a large amount of logs may be generated.



Potential risks

After you disable the health check of a Layer-7 SLB instance, SLB does not do health check on backend ECS instances. In this situation, if one backend ECS instance fails, the requests cannot be forwarded to other normal backend ECS instances.

Procedure

Log on to the Server Load Balancer console.

On the **Instances** page, find the target SLB instance and click **Manage**.

	cn-hangzhou-b(Master) cn-hangzhou-d(Slave)	120 135(Public IP)	 Running	Classic Network	HTTP :25 TCP : 8080	Abnormal Normal	k8s-for-cs-c3da...	Shared-Performance Instance	Pay by Traffic	Pay-As-You-Go 2017-12-18 14:32:49 Created	Manage More
---	---	-----------------------	---	-----------------	------------------------	--------------------	--------------------	-----------------------------	----------------	--	--------------------

Click **Listeners**. Find the target listener and click **Configure**.

Front-end Protocol/Port	Backend Protocol/Port	Status	Forwarding Rules	Session Persistence	Health Check	Peak Bandwidth	Server Group	Actions
HTTP: 25	HTTP: 25	Abnormal	Weighted Round Robin	Disable	Enable	No Limits	-	Configure Details Add Forwarding Rules More

Click **Next** in the **Listener Configuration** dialog box to configure the health check.

Disable **Health Check**.

Configure Listener

1.Listener Configuration 2.Health Check 3.Success

Health Check: Enable

Domain Name: It must be 1-80 characters
Only letters a-z, numbers 0-9, hyphens (-), and periods (.) are allowed. If no domain is specified, the intranet IP addresses of the ECS instances added in the backend server pool are used.

Health Check Port: Port range is 1-65535.
By default, the backend server's port is used for health checks.

Health Check Path: /
The URI of the file page that is used to do the health check. It is recommended using a static page. The URI must be 1-80 characters long, and only the letters a-z, numbers 0-9, and the characters '-' '/' '.' '%' '?' '#' '&' and '=' are allowed.

Normal Status Code: ☒ http_2xx ☒ http_3xx ☐ http_4xx ☐ http_5xx
Status code used for successful health check

Show Advanced Options

Back Confirm Cancel

Click **Confirm** to complete the modification.

Change Layer-7 SLB to Layer-4 SLB

The health check of a Layer-4 SLB instance is achieved by using the three-way handshake of TCP and does not generate application logs. If possible, you can change the Layer-7 SLB to the Layer-4 SLB to reduce the number of the application logs.

Potential risks

After you change the Layer-7 SLB to the Layer-4 SLB, SLB checks only the status of the listener port, but does not check the HTTP status. SLB cannot detect HTTP application exceptions in real time.

Procedure

Log on to the Server Load Balancer console.

On the **Instances** page, find the target SLB instance and click **Manage**.

	cn-hangzhou-b(Master) cn-hangzhou-d(Slave)	114.58(Public IP)		Classic Network	TCP : 80	Normal	k8s-for-cs-c3da...	Shared-Performance Instance	Pay by Traffic	Pay-As-You-Go 2017-12-18 15:04:28 Created	Manage	More
--	---	-------------------	--	-----------------	----------	--------	--------------------	-----------------------------	----------------	--	------------------------	----------------------

Click **Listeners**. Find the target listener and click **Configure**.

Front-end Protocol/Port	Backend Protocol/Port	Status	Forwarding Rules	Session Persistence	Health Check	Peak Bandwidth	Server Group	Actions
TCP : 80	TCP: 3200Normal		Weighted Round Robin	Disable	Enable	50 Mbps	-	Configure Details More

Click **Next** in the **Listener Configuration** dialog box to configure the health check.

Change the **Health Check Method** to **TCP**.

Configure Listener

1.Listener Configuration

2.Health Check

3.Success

Health Check Method: ?

☐ TCP
 ☒ HTTP

Domain Name:

It must be 1-80 characters

Only letters a-z, numbers 0-9, hyphens (-), and periods (.) are allowed. If no domain is specified, the intranet IP addresses of the ECS instances added in the backend server pool are used.

Health Check Port:

Port range is 1-65535.

By default, the backend server's port is used for health checks.

Health Check Path:

/

The URI of the file page that is used to do the health check. It is recommended using a static page. The URI must be 1-80 characters long, and only the letters a-z, numbers 0-9, and the characters '-', '/', '.', '%', '?', '#', '&' and '=' are allowed.

Normal Status Code:

☒ http_2xx
 ☒ http_3xx
 ☐ http_4xx
 ☐ http_5xx

Status code used for successful health check

☐ Show Advanced Options

Back

Confirm

Cancel

Click **Confirm** to complete the modification.

Disable application logs on the health check page

Configure a health check site that is independent from the service site. Disable application logs of the health check page to reduce the number of health check logs. For example, if the service site is abc.123.com, 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 the service site has an exception, the health check cannot detect the exception of the service site.

Procedure

Create a new health check site and health check page on the backend ECS instance and disable logs. The following steps take Nginx as the example.

Log on to the Server Load Balancer console.

On the **Instances** page, find the target SLB instance and click **Manage**.

lb-3c111e7a...	cn-hangzhou-b(Master)	114.58(Public IP)	Running	Classic Network	TCP : 80	Normal	k8s-for-cs-c3da...	Shared-Performance Instance	Pay by Traffic	2017-12-18 15:04:28 Created	Manage	More+
----------------	-----------------------	-------------------	---------	-----------------	----------	--------	--------------------	-----------------------------	----------------	-----------------------------	--------	-------

Click **Listeners**. Find the target listener and click **Configure**.

Front-end Protocol/Port	Backend Protocol/Port	Status	Forwarding Rules	Session Persistence	Health Check	Peak Bandwidth	Server Group	Actions
TCP: 80	TCP: 3200Normal	Running	Weighted Round Robin	Disable	Enable	50 Mbps	-	Configure Details More+

Click **Next** in the **Listener Configuration** dialog box to configure the health check.

Enter the domain name of the health check site in **Domain Name** and the relative path of the health check page in **Health Check Path**.

Configure Listener

1.Listener Configuration

2.Health Check

3.Success

Health Check Method: ?

☐ TCP

☒ HTTP

Domain Name:

test.123.com

Only letters a-z, numbers 0-9, hyphens (-), and periods (.) are allowed. If no domain is specified, the intranet IP addresses of the ECS instances added in the backend server pool are used.

Health Check Port:

Port range is 1-65535.
By default, the backend server's port is used for health checks.

Health Check Path:

/test.html

The URI of the file page that is used to do the health check. It is recommended using a static page. The URI must be 1-80 characters long, and only the letters a-z, numbers 0-9, and the characters '-' '/' '.' '%' '?' '#' '&' and '=' are allowed.

Normal Status Code:

☒ http_2xx

☒ http_3xx

☐ http_4xx

☐ http_5xx

Status code used for successful health check

Show Advanced Options

Back

Confirm

Cancel

Click **Confirm** to complete the modification.