Alibaba Cloud
Server Load Balancer

Pricing

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}`

                                                                      Issue: 20200228

Contents

1 Billing method

This topic describes the billing method of Server Load Balancer (SLB). The fees for an SLB instance are calculated by using the pay-as-you-go billing method and are calculated based on your actual traffic usage. You can release a pay-as-you-go instance at any time.

To purchase a pay-as-you-go SLB instance, go to *the purchase page***.**

Billing items

The following table details the items that are billed. Billing items vary by network type and instance type, as shown in the following table.

> **Note:**
>
> **"-" means that the corresponding item is not billed, and "#" means that the corresponding item is billed.**

Network type	Instance type	Instance fee	Traffic fee	Specification fee
Public-facing SLB instance	**Shared-performance instances**	#	#	-
	Guaranteed-performance instances	#	#	#
Internal SLB instance	**Shared-performance instances**	-	-	-
	Guaranteed-performance instances	-	-	#

Instance fee

For SLB instances that communicate over the Internet, public IP address reservatio ns incur charges.

> **Note:**

SLB instances that communicate through the internal network do not incur such charges.

Instance fees for SLB instances that use the Internet are calculated as follows:

· Instance fee = Unit price × Instance reservation time

  The reservation time is the period from the time at which the instance is created to the time at which the instance is released.

· Instance fees are billed on an hourly basis. If your period of usage is less than one hour, the billable time is rounded up to one hour.

If the price on the purchase page of the console is different from the price listed in the following table, take the price on the purchase page as the standard.

Region	Instance fee (USD/hour)
China (Hangzhou), China (Beijing), China (Shenzhen), China (Shanghai), China (Zhangjiakou)	0.003
China (Qingdao)	0.003
China (Hong Kong)	0.009
US (Silicon Valley), US (Virginia)	0.005
Singapore, Indonesia (Jakarta), India (Mumbai)	0.006
Japan (Tokyo)	0.009
Germany (Frankfurt)	0.006
UAE (Dubai)	0.009
Australia (Sydney)	0.006

Traffic fee

SLB instances that communicate through the Internet incur traffic fees based on your usage.

Note:

SLB instances that communicate through the internal network can be used free of charge.

Traffic fees for SLB instances that use the Internet are calculated as follows:

· **Internet traffic fee = Unit traffic price × Time**

The billable Internet traffic is the outbound traffic. The inbound traffic is not charged.

· **Traffic fees are billed on an hourly basis. If your period of usage is less than one hour, the billable time is rounded up to one hour.**

If the price on the purchase page of the console is different from the price listed in the following table, take the price on the purchase page as the standard.

Region	Traffic fee (USD/Gbit/s)
China (Hangzhou), China (Beijing), China (Shenzhen), China (Shanghai), China (Zhangjiakou)	0.125
China (Qingdao)	0.113
China (Hong Kong)	0.156
US (Silicon Valley), US (Virginia)	0.078
Singapore, Indonesia (Jakarta), India (Mumbai)	0.117
Japan (Tokyo)	0.120
Germany (Frankfurt)	0.070
UAE (Dubai)	0.447
Australia (Sydney)	0.096

Specification fee

The following are three key performance metrics for guaranteed-performance instances. The limits of these metrics vary according to specification. For more information, see *#unique_4***.**

· **Max Connection**

The maximum number of connections to an SLB instance. When the number of connections reaches the limit of the specification, new connections are dropped.

· **Connection Per Second (CPS)**

The number of new connections that are established per second. When the CPS reaches the limit of the specification, new connections are dropped.

· **Queries Per Second (QPS)**

The number of HTTP/HTTPS requests that can be processed per second. This metric is available only for layer-7 SLB listeners. When the QPS reaches the limit of the specification, new connections are dropped.

The specification fee of a guaranteed-performance instance is charged based on your actual usage. If the actual performance of the instance is between two specifications, the specification fee is calculated according to the higher specification.

For example, you choose the specification of slb.s3.large (Max Connection: 1,000,000; CPS: 100,000; QPS: 50,000), and the actual usage of the instance in an hour is as follows:

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

· With respect to Max Connection, the actual metric value of 90,000 lies between the limit of 50,000 defined in Standard I (slb.s2.small) and the limit of 100,000 defined in Standard II (slb.s2.medium). Therefore, the billable specification of the Max Connection metric for this hour is Standard II (slb.s2.medium).
· With respect to CPS, the actual metric value of 4,000 falls between the limit of 3,000 defined in the Small I (slb.s1.small) specification and the limit of 5,000 defined in the Standard I (slb.s2.small) specification. Therefore, the billable specification of the CPS metric for this hour is Standard I (slb.s2.small).
· With respect to QPS, the actual metric value of 11,000 falls between the limit of 10,000 defined in Standard II (slb.s2.medium) and the limit of 20,000 defined in Higher I (slb.s3.small). Therefore, the billable specification of the QPS metric for this hour is Higher I (slb.s3.small).

Out of the three metrics, the billable specification of the QPS metric is the highest. Therefore, the specification fee of the instance in this hour is charged according to the price of the Higher I (slb.s3.small) specification.

The following figure is an example showing how the specification fee is billed for an SLB instance:
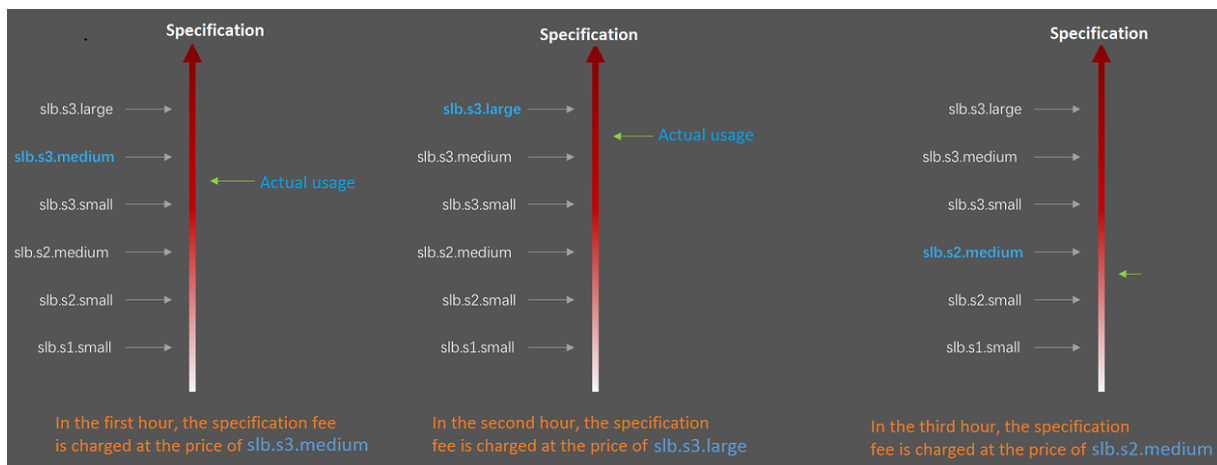
The billing is more flexible for guaranteed-performance instances. The specificat ion you select when purchasing an instance is the higher performance limit of the instance. For example, if you select Higher II (slb.s3.medium), new requests are dropped when requests reach 30,000 in one second.

The prices detailed in the following table are for reference purposes only. The price you see in the console more accurately reflect your usage.

Region	Specification	Max Connectic	CPS	QPS	Specificat ion fee (USD/ hour)
China (Hangzhou) China (Zhangjiako u) China (Hohhot) China (Qingdao) China (Beijing) China (Shanghai) China (Shenzhen)	Specification 1: Small I (slb.s1. small)	5000	3000	1000	Free of charge for a limited period
	Specification 2: Standard I (slb.s2. small)	50,000	5,000	5,000	0.05
	Specification 3: Standard II (slb.s2. medium)	100,000	10,000	10,000	0.10
	Specification 4: Higher I (slb.s3. small)	200,000	20,000	20,000	0.20
	Specification 5: Higher II (slb.s3. medium)	500,000	50,000	30,000	0.31

Region	Specification	Max Connectio	CPS	QPS	Specification fee (USD/ hour)
	Specification 6: Super I (slb.s3. large)	1,000, 000	100,000	50,000	0.51
Singapore Malaysia (Kuala Lumpur) Indonesia (Jakarta) India (Mumbai) US (Silicon Valley) US (Virginia) China (Hong Kong)	Specification 1: Small I (slb.s1. small)	5,000	3,000	1,000	Free of charge for a limited period
	Specification 2: Standard I (slb.s2. small)	50,000	5,000	5,000	0.06
	Specification 3: Standard II (slb.s2. medium)	100,000	10,000	10,000	0.12
	Specification 4: Higher I (slb.s3. small)	200,000	20,000	20,000	0.24
	Specification 5: Higher II (slb.s3. medium)	500,000	50,000	30,000	0.37
	Specification 6: Super I (slb.s3. large)	1,000, 000	100,000	50,000	0.61

2 Overdue payments

The load balancing service will not be stopped immediately after a Server Load Balancer (SLB) bill is overdue. We recommend that you pay for your SLB instances in time to avoid service interruptions.

The following will happen when a pay-as-you-go instance is overdue:

· **After a bill is overdue, the instance keeps running for 15 days. Then, the instance is locked and the service stops.**

After the instance stops running, billing is also stopped.

· **If the SLB bill is still overdue 15 days after the instance is locked, the instance is automatically released.**

The account owner will receive an email notification one day before the instance is released. The instance configuration and related data are deleted and cannot be restored after the instance is released.

3 Monitoring data and billing data

Server Load Balancer (SLB) provides a monitoring function that monitors such metrics as the inbound and outbound traffic and the number of connections. You can view real-time monitoring data in the console. Besides monitoring data, billing data is also collected, but it is collected for the calculation of fees to be charged. Monitoring data and billing data differ given the factors described as follows.

Factor	Monitoring data	Billing data
Calculation method	The SLB system collects monitoring data every minute, and reports the data to CloudMonitor. After every 15 minutes, CloudMonitor calculates the average value of data collected in that time period. The network traffic data displayed in the console is the average value calculated.	Billing data is collected every minute, and the SLB system reports the accumulated value once each hour to the billing system. Monitoring data is the calculated average for a 15-minute time period, but the billing data is the accumulated value in a billing cycle.
Latency	SLB provides real-time monitoring data. However, a short delay may inevitably occur during the process of data collection, calculation, and display. Although this delay is nearly immeasurable, it can create a certain degree of discrepancy between the monitoring and billing data.	Billing data can allow up to a three-hour delay. For example, billing data generated between 01:00–02:00 is normally reported to the billing system before 03:00. However, data may be reported up to three hours later, with the last reporting time being 05:00. As a result, there may be a discrepancy between billing data and monitoring data.
Purpose	The purpose of monitoring is to help you observe if instances are running normally. If not, you can take measures to solve problems in a timely manner.	The purpose of collecting billing data is to generate bills. Monitoring data cannot be used as billing data.

4 Server Load Balancer billing FAQ

The following are frequently asked questions about the billing of Server Load Balancer (SLB):

- *How is SLB billed?*

- *Is the inbound traffic of SLB billed?*

- *Is traffic generated by health checks billed?*

- *Is the billing method of an ECS instance changed after it is added to the backend server pool of SLB?*

- *Is traffic generated by unexpected traffic spikes, or malicious activity such as a DDoS attack, billed?*

- *If all the backend ECS instances of an SLB instance are stopped or removed, does the SLB instance continue to be billed?*

- *Is a specification fee charged to intranet SLB instances?*

- *Are the traffic fees and instance fees charged for guaranteed-performance instances the same as those charged for shared-performance instances?*

- *Is an extra fee included for shared-performance instances after Alibaba Cloud starts charging for the specification fee?*

- *Why is the monitoring data of SLB instances different from the data recorded on my bills?*

- *Why is the value of the traffic actually used by HTTPS protocol greater than the value recorded on my bills?*

How is SLB billed?

> **See** *Billing method***.**

Is the inbound traffic of SLB billed?

> **No. Only the outbound traffic of SLB is billed. For more information, see** *#unique_8***.**

Is traffic generated by health checks billed?

> **No. The traffic generated by health checks of SLB is not included in the billed traffic**
>
> **.**

Is the billing method of an ECS instance changed after it is added to the backend server pool of SLB?

> **No. The billing method of an ECS instance does not change whether or not you associate it with an SLB instance. SLB and ECS instances are billed separately based on usage.**

Is traffic generated by unexpected traffic spikes, or malicious activity such as a DDoS attack, billed?

> **Yes. However, SLB can work with Alibaba Cloud Security to protect your SLB instances and minimize the billed traffic caused by malicious attacks. A latency of several seconds may happen from the time when the attack traffic reaches the scrubbing or blackholing threshold to the time when Alibaba Cloud Security starts scrubbing or blackholing. During this period, responses may be sent to attacks and this incurs fees. Also, such attacks consume the bandwidth resources of SLB instances.**

If all the backend ECS instances of an SLB instance are stopped or removed, does the SLB instance continue to be billed?

> **Yes. Fees are charged for the following reasons:**
>
> · **Traffic-based billing**
>
>   **In the case of traffic-based billing, no traffic fee is incurred when an SLB instance is stopped, released, or not accessed.**
>
>   **SLB is a traffic distribution and control service in front of backend ECS instances . SLB provides services through its service address. If all backend ECS instances are stopped, but the SLB instance is not stopped, inbound traffic can still reach the service address of the SLB instance. In this case, the SLB instance responds when it discovers that no backend ECS instances are available after performing health checks.**
>
>   **For layer-4 SLB, only three-way handshake packets are returned. For layer-7 SLB, a Tengine 503 error page is prompted because the service is provided by Tengine. If requests keeps reaching SLB, SLB will respond continuously. The response traffic is billed.**
>
>   **This also applies to SLB instances with no ECS instances added. Therefore, we recommend that you stop the SLB instance if you no longer need it.**

Is a specification fee charged to intranet SLB instances?

- **If the intranet SLB instance is a shared-performance instance, no specification fee is charged.**
- **If the intranet SLB instance is a guaranteed-performance instance, a specification fee is charged.**

  The calculation method of specification fees for intranet SLB instances is the same as that for Internet SLB instances. No instance fee or traffic fee is charged for intranet instances.

Are the traffic fees and instance fees charged for guaranteed-performance instances the same as those charged for shared-performance instances?

  Yes.

Is an extra fee included for shared-performance instances after Alibaba Cloud starts charging for the specification fee?

  No.

  Extra fees are not charged for shared-performance instances unless you change them to guaranteed-performance instances.

Why is the monitoring data of SLB instances different from the data recorded on my bills?

- **Take traffic data as an example. The monitoring data displayed in the SLB console are average values. SLB collects monitoring data at one minute intervals and reports the data to CloudMonitor, which then calculates the average value at every 15-minute intervals. However, the data reported on your bills are accumulated values. Specifically, SLB collects traffic data at one minute intervals, and reports the accumulated values at every one hour intervals to the billing system.**

  This means that the data reported to your billing system are accumulated values calculated at one hour intervals, whereas the monitoring data displayed in the console is of average values calculated at every 15 minute interval. The different intervals used in the two calculations result in a discrepancy in monitoring data and billing data.
- **Monitoring data and billing data have different latency requirements. SLB provides real-time monitoring data. However, a short delay may occur during data collection, calculation, and display. Such a delay can cause a discrepanc**

y between monitoring and billing data. Billing data can be recorded after a maximum delay of three hours. For example, billing data generated between 01: 00 and 02:00 is normally reported to the billing system before 03:00, but due to a delay, may be reported to the billing system at 05:00, resulting in a discrepancy between billing data and monitoring data.

· **Monitoring data and billing data have different purposes. Specifically, the purpose of monitoring is to help you observe if instances are running normally . If the instances are running abnormally, you can take measures to resolve problems in a timely manner. The purpose of billing data is to generate bills based on the actual usage of resources under your account.**

Why is the value of the traffic actually used by HTTPS protocol greater than the value recorded on my bills?

Some traffic is generated due to handshakes in HTTPS transactions. Therefore, the actual traffic value is higher than the value recorded on your bills.