

Alibaba Cloud

Server Load Balancer Limits

Document Version: 20220628

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 because of product version upgrade, adjustment, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and an updated version of this document will be released through Alibaba Cloud-authorized channels from time to time. You should 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 this document based on the "status quo", "being defective", and "existing functions" of its products and services. 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 take legal responsibility for any errors or lost profits incurred by any organization, company, or individual arising from download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, take responsibility for any indirect, consequential, punitive, contingent, 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 intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of 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 directly contact Alibaba Cloud for any errors of this document.

Document conventions

Style	Description	Example
 Danger	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.
 Warning	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.
 Notice	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.
 Note	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 <code>cd /d C:/window</code> command to enter the Windows system folder.
<i>Italic</i>	Italic formatting is used for parameters and variables.	<code>bae log list --instanceid</code> <i>Instance_ID</i>
[] or [a b]	This format is used for an optional value, where only one item can be selected.	<code>ipconfig [-all -t]</code>
{ } or {a b}	This format is used for a required value, where only one item can be selected.	<code>switch {active stand}</code>

Table of Contents

1. Peak bandwidth limits	05
2. Manage quotas	06

1. Peak bandwidth limits

Alibaba Cloud provides pay-by-bandwidth Server Load Balancer (SLB) instances and pay-by-data-transfer SLB instances. This topic describes the peak bandwidth limits of the two types of SLB instances.

Peak bandwidth limits on SLB instances that use different billing methods

- The maximum bandwidth of a pay-by-data-transfer instance is used as a reference. It indicates the upper limit of the peak bandwidth. This upper limit is not a guaranteed value.
- For a pay-by-bandwidth SLB instance, its inbound peak bandwidth equals its outbound peak bandwidth. You are charged only for the outbound bandwidth that is consumed to transmit data from Alibaba Cloud to the Internet.
- When SLB instances compete for resources, the peak bandwidth of pay-by-bandwidth SLB instances is guaranteed. However, the peak bandwidth of pay-by-data-transfer SLB instances may be limited.
- The sum of the peak bandwidth of all pay-by-data-transfer SLB instances in each region under the same account cannot exceed 5 Gbit/s. If your business requires larger peak bandwidth, you can purchase pay-by-bandwidth SLB instances. You can purchase subscription SLB instances that are billed on a pay-by-bandwidth basis, or purchase pay-as-you-go SLB instances, and pay for the amount of bandwidth that you use. In addition, you can associate elastic IP addresses (EIPs) with SLB instances. In this case, the SLB instances are billed on a pay-by-95-percentile basis.

The maximum bandwidth values of pay-by-data-transfer SLB instances that are displayed in the SLB console shall prevail.

To view the maximum bandwidth of a pay-by-data-transfer SLB instance, perform the following operations:

1. Log on to the [SLB console](#).
2. In the left-side navigation pane, click **Instances**.
3. On the **Instances** page, click the ID of the SLB instance that you want to manage.
4. On the **Instance Details** page, you can view the maximum bandwidth.

The screenshot shows the 'Instance Details' page for an SLB instance. The instance name is 'vpctf_update_'. The 'Basic Information' section shows the instance is 'Running' and has a 'Public Network' address type. The 'Billing Information' section shows the bandwidth is '32 Mbps'.

Basic Information		
Name	vpctf_update_	ID
Status	Running	Network Type
Address Type	Public Network	Zone
Deletion Protection	Enabled	Configuration Read-only Mode

Billing Information		
Bandwidth Billing Method	Pay-As-You-Go	Billing Method
Instance Type	Guaranteed-Performance slb.s2.small	IP Address
Creation Time	Nov 24, 2020, 03:40:53	Bandwidth

2. Manage quotas

This topic describes how to query the quota usage of a cloud resource in the Server Load Balancer (SLB) console. If the remaining quota of a resource cannot meet your business requirements, you can apply for a quota increase.

Apply for a quota increase

- 1.
- 2.
3. In the left-side navigation pane, choose **SLB Lab > Quota Management**.
4. On the **Quota Management** page, you can view the usage of the SLB resources and the privileges of your account.
5. To apply for a quota increase, click **Submit Application** in the **Actions** column.
 - **Requested Value**: the number of resources that you require. The number must be greater than the current quota. For more information about the quotas of SLB resources, see [Limits](#).
 - **Reason**: the reasons for the application, including business scenarios and necessity.
 - **Email**: the email address of the applicant.

Apply

Note: We recommend that you apply for resources based on your actual needs.

Quota Name

slb_quota_domain_extensions_num

The maximum number of domain extensions for each listener.

Current Quota

* Requested Value

* Reason

Enter your reason for application in details, such as business scenarios and why.

Mobile/Landline Phone Number

Email

OK

Cancel

6. Click **OK**.

The system reviews your application. You can check whether your application is approved based on the application status: **Rejected** or **Approved**. After your application is approved, the quota is raised to the specified quantity.

You can click **History** in the **Actions** column to view the application history.

What do I do if the number of instances is insufficient?

If you need more guaranteed-performance instances but the number of instances has reached the quota, you can apply for a quota increase.

1. Click **Submit Application** in the **Actions** column corresponding to the `slb_privilege_allow_more_guaranteed_performance_instances` privilege.

This privilege allows you to apply for more guaranteed-performance instances, but you cannot apply for more shared-performance instances.