ALIBABA CLOUD

Alibaba Cloud

Global Acceleration
User Guide

Document Version: 20220627

Alibaba Cloud

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

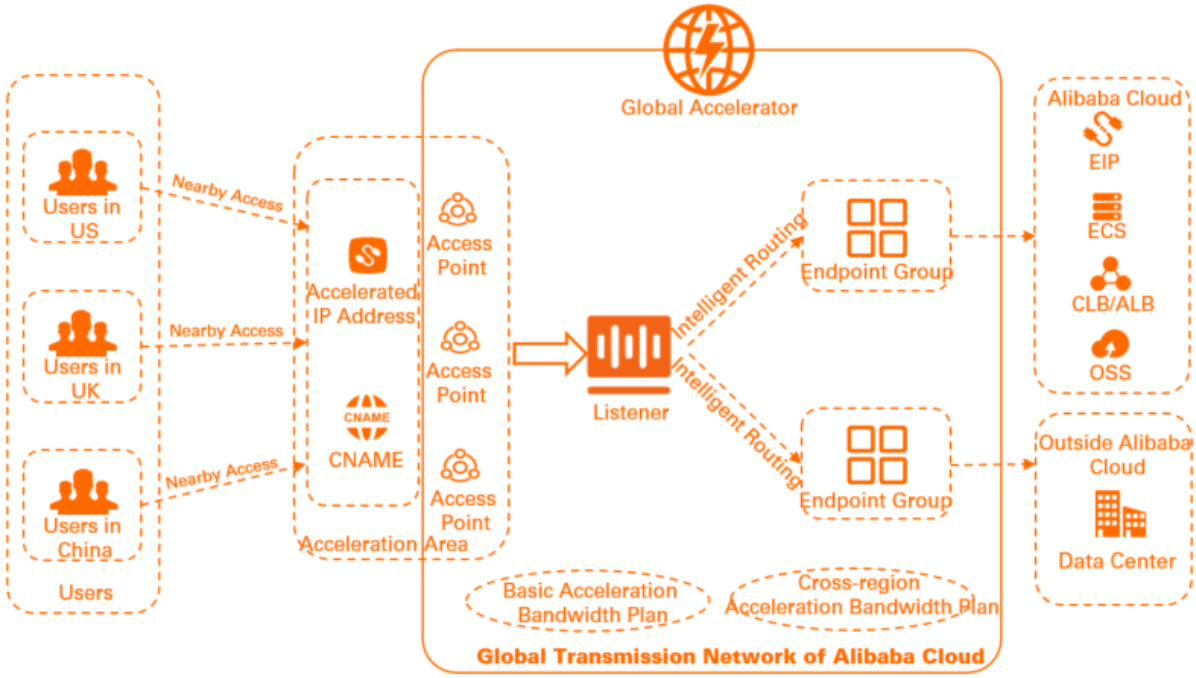
Style	Description	Example
Danger	A danger notice indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	**Danger:** Resetting will result in the loss of user configuration data.
Warning	A warning notice indicates a situation that may cause major 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 `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]`
{} or {a|b}	This format is used for a required value, where only one item can be selected.	`switch {active|stand}`

Table of Contents

1.Global Accelerator instances

1.1. Overview

Each Global Accelerator (GA) instance is an acceleration service that runs on a global scale. GA provides multiple instance specifications. Each instance specification provides different acceleration capabilities to meet your requirements in different scenarios.



When you create a GA instance, you must select the type of accelerated IP address based on the access mode that is required by your business. The following types of accelerated IP addresses are supported: Elastic IP Address (EIP) and Anycast EIP. After you create a GA instance, you must purchase a bandwidth plan, and add an acceleration area and listeners. You must add an acceleration area if you select EIP as the type of accelerated IP address.

Clients can connect to the nearest access point of the Alibaba Cloud global transmission network by sending requests to the accelerated IP address or the CNAME. GA then automatically selects routes to distribute client requests to the optimal endpoints. This helps avoid network congestion and reduce network latency.

You can specify Elastic Compute Service (ECS) instances, Classic Load Balancer (CLB) instances, Application Load Balancer (ALB) instances, Object Storage Service (OSS) buckets, Alibaba Cloud public IP addresses, custom IP addresses of origin servers, or custom domain names of origin servers as the endpoints of GA.

Types of GA instances

Type	Scenario
Basic	You can use basic GA instances to accelerate content delivery at Layer 3 (IP protocols). To implement the acceleration, you need to only specify an acceleration area and an endpoint group. For more information, see Use basic GA instances to accelerate content delivery.

Type	Scenario
Standard	You can use standard GA instances to accelerate content delivery at Layer 4 (TCP and UDP protocols) and Layer 7 (HTTP and HTTPS protocols).

Types of accelerate IP addresses

You can select the type of accelerated IP address based on the access mode that is required by your business.

> **Note**
>
> - By default, you cannot specify Anycast EIPs as accelerated IP addresses. If you want to use Anycast EIPs, submit a ticket.
> - If you use **Anycast EIPs**, the GA instances and basic bandwidth plans must meet the following requirements:
>   - GA instances: You must select standard GA instances whose specifications are Large I or higher.
>   - Basic bandwidth plans: You must select pay-by-data-transfer basic bandwidth plans whose bandwidth types are Premium. By default, you cannot use pay-by-data-transfer basic bandwidth plans. If you want to use pay-by-data-transfer basic bandwidth plans, submit a ticket.

Specifications of GA instances

GA provides the following instance specifications: Small I, Small II, Small III, Medium I, Medium II, Medium III, Large I, Large II, Large III, Large IV, Large V, Large VI, Large VII, Large VIII, Super Large I, and Super Large II. GA instances of different specifications provide different acceleration capabilities, as shown in the following table.

> **Note**
>
> - The unit price varies based on GA instance specifications.The unit price on the buy page shall prevail.
> - By default, the Large III specification and higher specifications are not available. To use these specifications, submit a ticket.

Specification	Number of acceleration regions	Bandwidth limit	Maximum number of concurrent connections	Unit price(USD/month)
Small I	1	20 Mbps	5,000	150
Small II	2	40 Mbps	10,000	300
Small III	3	60 Mbps	15,000	450
Medium I	5	100 Mbps	25,000	750

Specification	Number of acceleration regions	Bandwidth limit	Maximum number of concurrent connections	Unit price(USD/month)
Medium II	8	160 Mbps	40,000	1200
Medium III	10	200 Mbps	50,000	1500
Large I	All regions For more information about the acceleration areas and Alibaba Cloud regions that are supported by GA, see Acceleration areas and regions.	400 Mbps	100,000	3000
Large II		600 Mbps	150,000	4500
Large III		800 Mbps	200,000	6000
Large IV		1 Gbps	250,000	7500
Large V		1.2 Gbps	300,000	9000
Large VI		1.4 Gbps	350,000	10500
Large VII		1.6 Gbps	400,000	12000
Large VIII		1.8 Gbps	450,000	13500
Super Large I		2 Gbps	500,000	15000
Super Large II		4 Gbps	1,000,000	30000

Specification changes

If you want to change the specification of an existing GA instance, take note of the following items:

- You can only upgrade GA instances. The downgrade operation is not supported by default. If you want to downgrade GA instances, submit a ticket.
- You cannot change the specification of a GA instance if the acceleration region or the region where the endpoint group is deployed is a point of presence (PoP) node of Alibaba Cloud. For more information, see Modify the specification of a GA instance.

1.2. Create and manage GA instances

Global Acceleration (GA) is a global network acceleration service that features high availability and high performance. This topic describes how to create and manage a GA instance.

Create a GA instance

Before you use GA, you must create a GA instance.

1.
2. On the **Instances** page, click **Create Instance**.
3. On the buy page, set the following parameters of the instance, click **Buy Now**, and then complete the payment.

Parameter	Description
Type	Select a type of GA instance. ○ **Basic**: You can use basic GA instances to accelerate content delivery at Layer 3 (IP protocols). To implement the acceleration, you need only to specify an acceleration area and an endpoint group. For more information, see Use basic GA instances to accelerate content delivery. ○ **Standard**: You can use standard GA instances to accelerate content delivery at Layer 4 (TCP and UDP protocols) and Layer 7 (HTTP and HTTPS protocols).
Accelerated IP Address Type	Select the type of accelerated IP address. ○ **EIP** (default): If you select EIP, the custom access mode is used. You can select an access point based on your business requirements. Each access point provides a separate EIP. ○ **Anycast EIP**: If you select Anycast EIP, the automatic access mode is used. You do not need to specify an acceleration area. GA provides an Anycast EIP that is shared among multiple regions across the globe. **Note** You can select **Anycast EIP** only if you set **Type** to **Standard** and **Specification** to **Large I**. For more information, see Accelerated IP address.
Specification	Select a specification for the GA instance. You can select a specification for the GA instance only if you set **Type** to **Standard**. GA provides the following instance specifications: **Small I (Specifications Unit)**, **Small II**, **Small III**, **Medium I**, **Medium II**, **Medium III**, **Large I**, **Large II**, **Large III**, **Large IV**, **Large V**, **Large VI**, **Large VII**, **Large VIII**, **Super Large I**, and **Super Large II**. GA instances of different specifications provide different acceleration capabilities. For more information, see Specifications of GA instances.
Instance	By default, **Instance** is selected.
Subscription Duration	Select a subscription duration for the GA instance.

Change the specification of a GA instance

You can change the specifications of standard GA instances. You can only upgrade the specification of a GA instance. To downgrade a GA instance, you must apply for this feature to be enabled on your account. To enable this feature, submit a ticket.

1.
2. On the **Instances** page, find the GA instance that you want to manage and click **Upgrade** in the **Actions** column.
3. In the **Upgrade** message, confirm the information and click **OK**.

> **Note** New endpoint group IP addresses may be created after you change the specification of a GA instance. The number of newly created endpoint group IP addresses depends on the GA instance specification. You can go to the console to view the actual number. Make sure that the newly added endpoint group IP addresses are available.

4. On the **Upgrade/Downgrade** page, set the parameters, select **global accelerator Terms of Service**, and then click **Buy Now** to complete the payment.

   For more information about the acceleration capabilities provided by different specifications, see Specifications of GA instances.

2.Basic bandwidth plans

2.1. Overview

A basic bandwidth plan provides bandwidth for data transfer over the Internet and within internal networks of Alibaba Cloud. However, basic bandwidth plans are not applicable to data transfer between the Chinese mainland and areas outside the Chinese mainland. A basic bandwidth plan is required if you want to accelerate data transfer within the Chinese mainland, or between the Chinese mainland and other areas.

Bandwidth types

The following types of basic bandwidth plans are supported: basic, enhanced, and premium. The following table shows that the acceleration type, accelerated backend service, and acceleration scope of a basic bandwidth plan vary based on the bandwidth type.

Bandwidth type	Acceleration type	Accelerated backend service	Acceleration scope
Basic	Applications that are deployed on Alibaba Cloud	• Public IP addresses provided by Alibaba Cloud • Elastic Compute Service (ECS) • Classic Load Balancer (CLB) (formerly known as SLB) • Application Load Balancer (ALB) • Object Storage Service (OSS)	By default, the acceleration region and the region where the backend service is deployed are located in the Chinese mainland.
Enhanced	• Applications that are deployed on Alibaba Cloud • Applications that are not deployed on Alibaba Cloud	• Public IP addresses provided by Alibaba Cloud • ECS • CLB (formerly known as SLB) • ALB • OSS • Custom IP addresses • Custom domain names	By default, the acceleration region and the region where the backend service is deployed are located in the Chinese mainland.

Bandwidth type	Acceleration type	Accelerated backend service	Acceleration scope
Premium	• Applications that are deployed on Alibaba Cloud • Applications that are not deployed on Alibaba Cloud	• Public IP addresses provided by Alibaba Cloud • ECS • CLB (formerly known as SLB) • ALB • OSS • Custom IP addresses • Custom domain names	By default, the acceleration region and the region where the backend service is deployed are located in the areas outside the Chinese mainland. If you want to accelerate data transfer between the Chinese mainland and other areas, you must select China (Hong Kong) as the acceleration region.

Note

- You can specify ECS, CLB, and ALB instances as endpoints only if your Alibaba Cloud account is included in the whitelist. If you want to specify ECS, CLB, or ALB instances as endpoints for your GA instances, submit a ticket to upgrade the GA instances.
- If you want to specify ECS instances or CLB instances as endpoints, make sure that the instances are deployed in virtual private clouds (VPCs).
- The IP addresses of endpoint groups associated with each GA instance must be globally unique and not conflict with those of other GA instances.

Purchase a basic bandwidth plan

To purchase a basic bandwidth plan, go to the buy page.

2.2. Purchase and manage basic bandwidth plans

A basic bandwidth plan provides bandwidth for data transfer over the Internet and within Alibaba Cloud. This topic describes how to purchase and manage basic bandwidth plans.

Purchase a basic bandwidth plan

1.
2. On the **Instances** page, click **Purchase Basic Bandwidth Plan**.
3. On the buy page, set the following parameters, click **Buy Now**, and then complete the payment.

Parameter	Description
Bandwidth Type	Select a bandwidth type for the basic bandwidth plan. The following types of basic bandwidth plans are supported: basic, enhanced, and premium.

Parameter	Description
Peak Bandwidth	Select the bandwidth limit of the basic bandwidth plan.
Duration	Select a subscription duration of the basic bandwidth plan.

Associate a basic bandwidth plan

After you purchase a basic bandwidth plan, you must associate the bandwidth plan with a Global Accelerator (GA) instance. You can allocate bandwidth to an acceleration region only after you associate the basic bandwidth plan with a GA instance.

Each GA instance can be associated only with one basic bandwidth plan.

Make sure that a GA instance and a basic bandwidth plan are purchased before you associate the basic bandwidth plan with a GA instance. For more information, see Create and manage GA instances and Purchase a basic bandwidth plan.

1.
2. On the **Instances** page, find the GA instance that you want to manage and click its ID.
3. On the page that appears, click the **Bandwidth Manage** tab.
4. In the **Basic Bandwidth Plan** section, find the basic bandwidth plan that you want to manage and click **Bind** in the **Actions** column.
   After the basic bandwidth plan is associated with the GA instance, the basic bandwidth plan changes to the **In Use** state.

Replace a basic bandwidth plan

You can replace a basic bandwidth plan that is associated with a GA instance. This allows you to use the basic bandwidth plan that meets your requirements. The GA instance continues to forward network traffic when you replace the basic bandwidth plan.

After you replace the original basic bandwidth plan with the required bandwidth plan, the original one is disassociated from the GA instance and the required one is associated with the GA instance.

Make sure that the required basic bandwidth plan is purchased. The bandwidth provided by the basic bandwidth plan is equal to or more than the total bandwidth that is allocated to the specified acceleration area. For more information, see Purchase a basic bandwidth plan.

1.
2. On the **Instances** page, find the GA instance that you want to manage and click its ID.
3. On the page that appears, click the **Bandwidth Manage** tab.
4. In the **Basic Bandwidth Plan** section, find the basic bandwidth plan that you want to replace and click **Replace** in the **Actions** column.
5. In the **Replace Basic Bandwidth Plan** dialog box, select the basic bandwidth plan that you want to use and click **OK**.

   You can only select a basic bandwidth plan that is in the **Active** state.

Disassociate a basic bandwidth plan

You can disassociate a basic bandwidth plan from a GA instance. If your GA instance is associated with a basic bandwidth plan, you must disassociate the bandwidth plan before you can associate the GA instance with another basic bandwidth plan.

Make sure that no acceleration areas and listeners are configured for the GA instance from which you want to disassociate the basic bandwidth plan. Before you disassociate the basic bandwidth plan, delete all the acceleration areas and listeners that are configured. For more information, see Delete an acceleration area and Delete a listener.

1.
2. On the **Instances** page, find the GA instance that you want to manage and click its ID.
3. On the page that appears, click the **Bandwidth Manage** tab.
4. In the **Basic Bandwidth Package** section, find the bandwidth plan, and click **Unbind** in the **Actions** column.
5. In the **Unbind Bandwidth Plan** message, click **OK**.

Change specifications

You can modify the bandwidth limit of a basic bandwidth plan. The modification immediately takes effect.

Before you change the specification of a basic bandwidth plan, take note of the following information:

- You can only upgrade a basic bandwidth plan. To downgrade a basic bandwidth plan, make sure that your account is included in the whitelist. To enable this feature, submit a ticket.
- To downgrade a basic bandwidth plan, make sure that the total allocated bandwidth across all acceleration regions is no more than the bandwidth limit of the downgraded plan.
- When you upgrade or downgrade a basic bandwidth plan, make sure that the bandwidth limit of the upgraded or downgraded basic bandwidth plan does not exceed the bandwidth limit that is supported by the current GA instance. For more information about GA instance types, see Overview.

1.
2. On the **Instances** page, find the GA instance that you want to manage and click its ID.
3. On the page that appears, click the **Bandwidth Manage** tab.
4. In the **Basic Bandwidth Plan** section, find the basic bandwidth plan that you want to manage and click **Change Configurations** in the **Bandwidth Limit** column.
5. On the **Upgrade/Downgrade** page, change the bandwidth limit of the basic bandwidth plan, select **global accelerator bandwidth package Terms of Service**, and then click **Buy Now** to complete the payment.

> **Note** You can only change the bandwidth type of a basic bandwidth plan from basic to enhanced. You cannot change the enhanced bandwidth type and premium bandwidth type to other bandwidth types.

References

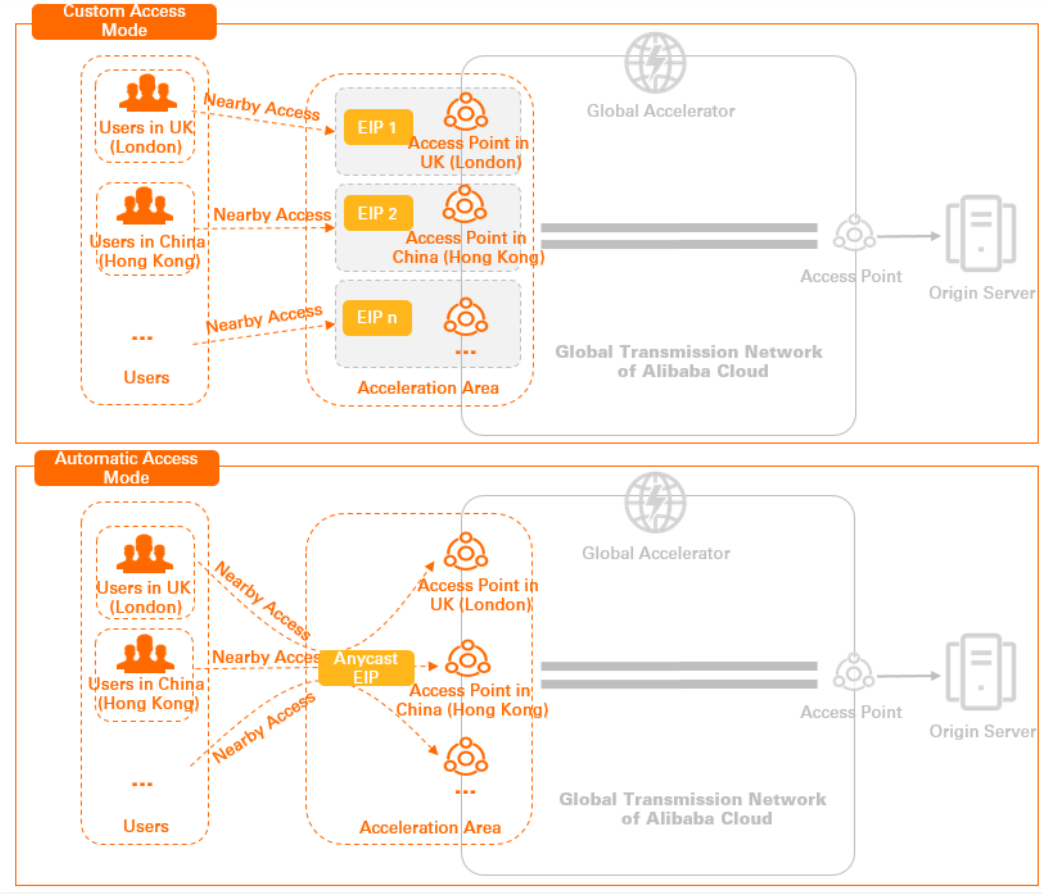
- CreateBandwidthPackage: You can call this operation to create a bandwidth plan.
- BandwidthPackageAddAccelerator: You can call this operation to associate a bandwidth plan with a GA instance.
- ReplaceBandwidthPackage: You can call this operation to replace a bandwidth plan.

- BandwidthPackageRemoveAccelerator: You can call this operation to disassociate a bandwidth plan from a GA instance.
- UpdateBandwidthPackage: You can call this operation to modify the configurations of a bandwidth plan.

3.Acceleration areas

3.1. Overview

An acceleration area is the area that requires accelerated access to your service. The access mode that is required by your business determines whether you need to specify an acceleration area.



An acceleration area is a collection of Alibaba Cloud regions. Each acceleration area contains one or more Alibaba Cloud regions. When you create a Global Accelerator (GA) instance, you must select the type of accelerated IP address based on the access mode that is required by your business. The following types of accelerated IP addresses are supported: Elastic IP Address (EIP) and Anycast EIP. The type of accelerated IP address that you select determines whether you need to specify an acceleration area.

Accelerated IP address

Clients can connect to the nearest access point of the Alibaba Cloud global transmission network by sending requests to the accelerated IP address.

Types of accelerated IP addresses

Type	Description	Supported access point	Feature	Scenario

Type	Description	Supported access point	Feature	Scenario
EIP	**The custom access mode is used. You must specify an acceleration area.** You can select an acceleration area and region based on your business requirements. GA allocates a separate EIP to each acceleration region.	For more information about the acceleration areas and Alibaba Cloud regions that are supported by GA, see Acceleration areas and regions.	• Advantages: Different accelerated IP addresses are provided for clients after the client requests are resolved by using the Alibaba Cloud DNS service. • Disadvantages: The configuration and maintenance are complex. You need to specify acceleration areas and allocate bandwidth based on your business requirements. Static IP addresses cannot be used to provide services.	You can use EIPs to accelerate applications whose users are located in specific regions. This provides a consistent experience for users that use the acceleration service. Example: SaaS applications and live streaming applications.
Anycast EIP	**The automatic access mode is used. You do not need to specify an acceleration area.** You do not need to specify an acceleration area. GA allocates an Anycast EIP to multiple regions across the globe. Users can connect to the nearest access point of the Alibaba Cloud global transmission network by sending requests to the Anycast EIP.	The acceleration service is dependent on the access points that are supported by Anycast EIP. You can use Anycast EIPs to accelerate content delivery for clients outside the Chinese mainland. To accelerate content delivery for clients in the Chinese mainland by using Anycast EIPs, you must specify China (Hong Kong) as the acceleration region. For more information, see Access point locations.	• Advantages: You do not need to specify acceleration areas and regions. Clients can automatically connect to the nearest access point, which greatly reduces O&M workloads. If you need to add or delete acceleration regions to meet business requirements, or if an acceleration region is abnormal, the accelerated IP address remains unchanged. You do not need to modify the business system. • Disadvantages: Clients can connect only to access points that are supported by Anycast EIPs. The quality of acceleration service depends on the Internet Service Provider (ISP).	Anycast EIPs are suitable for applications that use the same static IP address to provide services and do not have requirements on the regions where the clients are located. Example: online multiplayer games that use a global server architecture, cross-border e-commerce applications, and web applications.

> **Note**
>
> - By default, you cannot specify Anycast EIPs as accelerated IP addresses. If you want to use Anycast EIPs, submit a ticket.
> - If you use **Anycast EIPs**, the GA instances and basic bandwidth plans must meet the following requirements:
>   - GA instances: You must select standard GA instances whose specifications are Large I or higher.
>   - Basic bandwidth plans: You must select pay-by-data-transfer basic bandwidth plans whose bandwidth types are Premium. By default, you cannot use pay-by-data-transfer basic bandwidth plans. If you want to use pay-by-data-transfer basic bandwidth plans, submit a ticket.

IP protocols of accelerated IP addresses

You can specify an acceleration area and select the IP protocol of the accelerated IP address only if you select **EIP** as the type of accelerated IP address. If you select **Anycast EIP** as the type of accelerated IP address, only IPv4 is supported.

After you add an acceleration area, GA assigns an accelerated IP address to each acceleration region in the acceleration area based on the IP protocol that you select. Clients can connect to the nearest access point of the Alibaba Cloud global transmission network by sending requests to the accelerated IP address.

You can select one of the following IP protocols:

- **IPv4**: assigns an accelerated IPv4 address. The accelerated IPv4 address is used to accelerate IPv4 services for IPv4 clients.
- **IPv6**: assigns an accelerated IPv6 address. The accelerated IPv6 address is used to accelerate IPv4 services for IPv6 clients.

  > **Note**
  >
  > - Only IPv6 clients in the following regions can connect to GA: China (Qingdao), China (Beijing), China (Hangzhou), China (Shanghai), China (Shenzhen), China (Heyuan), China (Guangzhou), China (Chengdu), China (Hong Kong), Singapore (Singapore), US (Virginia), and Germany (Frankfurt).
  > - In the same acceleration region of a GA instance, you can select one of the following IP address protocols: IPv4 or IPv6.

Acceleration areas and regions

> **Note** By default, the following acceleration regions are unavailable: China (Heyuan), China (Nanjing), Brazil (Sao Paulo), Thailand (Bangkok), Vietnam (Ho Chi Minh), and UAS (Dubai). If you want to specify the preceding regions, submit a ticket.

3.2. Add and manage acceleration areas

After you create a Global Accelerator (GA) instance, you must add an acceleration area. An acceleration area is the area that requires accelerated access to your service.

Background information

If you specify EIP as the type of accelerated IP address, you must specify an acceleration area for a GA instance. If you specify Anycast EIP as the type of accelerated IP address, you do not need to specify an acceleration area for a GA instance.

- For more information about the types of accelerated IP addresses, see Accelerated IP address.
- For more information about how to add acceleration areas for basic GA instances, see Use basic GA instances to accelerate content delivery.

Prerequisites

- A GA instance is created. For more information, see Create and manage GA instances.
- A basic bandwidth plan is purchased and associated with the GA instance. For more information, see Purchase and manage basic bandwidth plans.

Add an acceleration area

1.
2. On the **Instances** page, find the GA instance that you want to manage and click its ID.
3. On the instance details page, click the **Acceleration Areas** tab, select the area that requires acceleration, and then click **Add Region**.
4. In the **Add Acceleration Area** dialog box, specify the following acceleration area information and click **OK**.

Parameter	Description
Region	Select the region that requires acceleration. For more information about acceleration areas and acceleration regions, see Acceleration areas and regions.

Parameter	Description
Bandwidth	Allocate bandwidth to the region. Unit: Mbit/s. **Note** ◦ You must allocate at least 2 Mbit/s of bandwidth to each acceleration region. ◦ The sum of bandwidth for all regions cannot exceed the bandwidth limit of the basic bandwidth plan that is associated with the GA instance. For example, if the bandwidth limit of your basic bandwidth plan is 10 Mbit/s and you have allocated 6 Mbit/s to the China (Qingdao) region, the available bandwidth that you can allocate is 4 Mbit/s. ◦ If you associate a pay-as-you-go basic bandwidth plan with a GA instance, you do not need to specify the bandwidth. By default, the bandwidth allocated to each acceleration region is the same as the bandwidth limit of the pay-as-you-go basic bandwidth plan.
Internet Protocol	Select the Internet protocol that is used by to connect to GA. ◦ **IPv4**: assigns an accelerated IPv4 address. The accelerated IPv4 address is used to accelerate IPv4 services for IPv4 clients. ◦ **IPv6**: assigns an accelerated IPv6 address. The accelerated IPv6 address is used to accelerate IPv4 services for IPv6 clients. **Note** ■ Only IPv6 clients in the following regions can connect to GA: China (Qingdao), China (Beijing), China (Hangzhou), China (Shanghai), China (Shenzhen), China (Heyuan), China (Guangzhou), China (Chengdu), China (Hong Kong), Singapore (Singapore), US (Virginia), and Germany (Frankfurt). ■ In the same acceleration region of a GA instance, you can select one of the following IP address protocols: IPv4 or IPv6.

You can click **Add** to add more regions and allocate bandwidth.

Note The number of regions that can be added varies based on the specification of the GA instance. For more information about the number of acceleration regions supported by each specification , see Overview.

Modify an acceleration area

You can modify the bandwidth value of an acceleration area.

1.
2. On the **Instances** page, find the GA instance that you want to manage and click its ID.

3. On the **Acceleration Areas** tab, click the tab of the acceleration area that you want to manage and click **Edit Bandwidth**.
4. In the **Edit Acceleration Area** dialog box, modify the bandwidth value and click **OK**.

Delete an acceleration area

You can delete an acceleration area. After the acceleration area is deleted, GA no longer provides acceleration services for this area.

1.
2. On the **Instances** page, find the GA instance that you want to manage and click its ID.
3. On the **Acceleration Areas** tab, find the acceleration area that you want to delete and click **Delete** in the **Actions** column.
4. In the **Delete Delete IP Addresses** message, click **OK**.

References

- CreateIpSets: You can call this API operation to create one or more acceleration regions.
- UpdateIpSet: You can call this API operation to modify a specified acceleration region in an acceleration area.
- UpdateIpSets: You can call this API operation to modify multiple acceleration regions in an acceleration area.
- DeleteIpSet: You can call this API operation to delete an acceleration region.
- DeleteIpSets: You can call this API operation to delete multiple acceleration regions.

3.3. Modify the bandwidth value of an acceleration area

This topic describes how to modify the bandwidth value of an acceleration area.

Procedure

1.
2. On the **Instances** page, find the GA instance that you want to manage and click its ID.
3. On the **Acceleration Areas** tab, click the tab of the acceleration area that you want to manage and click **Edit Bandwidth**.
4. In the **Edit Acceleration Area** dialog box, modify the bandwidth value, and click **OK**.

Related information

- UpdateIpSet
- UpdateIpSets

3.4. Delete an acceleration area

This topic describes how to delete an acceleration area. After the acceleration area is deleted, Global Accelerator (GA) will no longer provide acceleration services for this area.

Procedure

1.
2. On the **Instances** page, find the target GA instance, and click the instance ID.
3. On the **Acceleration Areas** tab, find the target acceleration area, and click **Delete** in the **Actions** column.
4. In the **Delete IP Addresses** message, click **OK**.

Related information

- DeleteIpSet
- DeleteIpSets

4.Listeners

4.1. Listener overview

After you create a Global Accelerator (GA) instance, you must configure listeners for the GA instance. A listener listens for connection requests and then distributes the requests to endpoints based on the forwarding rules that are defined by a specified scheduling algorithm.

Listener protocols

You can create 10 listeners for each GA instance. The following listener protocols are supported: TCP, UDP, HTTP, and HTTPS. You can select a protocol based on the scenario.

Protocol	Description	Scenario
TCP	• A connection-oriented protocol that provides high reliability. A logical connection must be established before data can be transmitted. • Session persistence is based on source IP addresses. • Source IP addresses are visible at the network layer. • Data is transmitted at a slow rate.	• Applicable to scenarios that require high reliability and data accuracy but can withstand a low transmission speed. These scenarios include file transmission, email sending and receiving, and remote logons. • Web applications that do not have custom requirements.
UDP	• A connectionless and unreliable protocol. Three-way handshakes are not required before UDP packets are transmitted. UDP does not provide error recovery or data retransmission. • Data is transmitted at a high rate.	Applicable to scenarios where real-time transmission outweighs reliability, such as video conferencing and real-time quote services.
HTTP	• A connection-oriented protocol that provides high reliability. A logical connection must be established before data can be transmitted. • Data is transmitted at a high rate. • Data transmission is not encrypted.	• Applicable to scenarios where HTTP websites need to be accelerated. • Applicable to scenarios where HTTP websites that contain specified domain names or paths need to be accelerated.

Protocol	Description	Scenario
HTTPS	• A connection-oriented protocol that provides high reliability. A logical connection must be established before data can be transmitted. • You can bind SSL certificates to servers. This ensures high reliability of data. Note For more information about SSL certificates, see What is Certificate Management Service?. • Data transmission is encrypted.	• Applicable to scenarios where HTTP or HTTPS websites need to be accelerated. This also ensures the network security when clients access HTTP or HTTPS websites. • Applicable to scenarios where HTTP or HTTPS websites that contain specified domain names or paths need to be accelerated.

Listener ports

Listener ports are used to receive requests and forward the requests to endpoints. Listeners consist of basic listeners and advanced listeners. Advanced listeners can listen on a large number of ports.

Note If you add listeners that use the same protocol to a GA instance, you must configure different ports for the listeners.

- Basic listeners

  The following table describes the number of ports that are supported by listeners that use different protocols. For TCP and UDP listeners, you can submit a ticket to increase the quota of **gaplus_quota_port_per_listener**. For more information, see Manage quotas.

Listener protocol	Listener port range	Listener port quota
TCP	1~65499	30. ○ Separate multiple listener ports with commas (,). Example: *80,90,8080*. ○ If you want to specify port ranges, you can use a tilde (~). For example, you can enter *80~83* to specify the ports 80, 81, 82, and 83.
UDP	1~65499	30. ○ Separate multiple listener ports with commas (,). Example: *80,90,8080*. ○ If you want to specify port ranges, you can use a tilde (~). For example, you can enter *80~83* to specify the ports 80, 81, 82, and 83.
HTTP	1~65499	1.
HTTPS	1~65499	1.

- Advanced listeners

  You can specify more than 300 consecutive listener ports for a TCP or UDP listener. Advanced listeners are listeners that each contain more than 300 consecutive listener ports. Advanced listeners have the following limits:

  - By default, you can create advanced listeners only for GA instances that are created after January 8, 2022. If your GA instances were created before this date and you want to create advanced listeners, submit a ticket to upgrade the GA instances.
  - You must specify more than 300 ports for an advanced listener. The number of ports that you specify must not exceed 65,499.
  - You can create only one advanced listener for each GA instance.
  - You can specify only consecutive ports. For example, you can set the port range to *1~350*. You cannot set the port range to *1,3~350*.
  - If the acceleration region of a GA instance is a point of presence (PoP) node of Alibaba Cloud, you cannot create an advanced listener for the GA instance.

    > **Note** If you want to check whether the acceleration region of a specified GA instance is a PoP node of Alibaba Cloud, refer to ListAvailableBusiRegions.

For example, you want to create the following listeners for a GA instance: a TCP listener whose listener ports are from 1 to 400, a TCP listener whose listener port is 443, a UDP listener whose listener ports are from 200 to 210, and a UDP listener port whose listener ports are from 230 to 240. The TCP listener whose listener ports are from 1 to 400 is an advanced listener. The following figure shows the listeners.

Listener ID/Name	Protocol	Port Number	Status	Default Endpoint Group
lsr-bp1t TCP01	TCP	1~400	Running	1
lsr-bp1f UDP01	UDP	200~210	Running	1
lsr-bp1a UDP02	UDP	230~240	Running	1
lsr-bp TCP02	TCP	443	Running	1

4.2. Add and manage listeners

After you create a Global Accelerator (GA) instance, you must configure listeners for the GA instance. A listener listens for connection requests and then distributes the requests to endpoints based on the forwarding rules that are defined by a specified scheduling algorithm.

Prerequisites

- A GA instance is created. For more information, see Create and manage GA instances.
- If you want to configure HTTPS listeners, make sure that a certificate signing request is submitted to the certificate authority (CA) and an SSL certificate is purchased. For more information, see Select and purchase certificates and Submit a certificate application.

Add a TCP or UDP listener

1. Configure the listener and protocol.
   i. Log on to the GA console.

i. Log on to the GA console.

ii. On the **Instances** page, find the GA instance that you want to manage and click **Configure Listeners** in the **Actions** column.

iii. On the **Listener** tab, click **Add Listener**.

> **Note** If this is the first time that you add a listener, or the specified GA instance is not configured with a listener, skip this step.

iv. On the **Configure Listener & Protocol** wizard page, specify the following listener information and click **Next**.

Parameter	Description
Listener Name	Enter a name for the listener. The name must be 2 to 128 characters in length, and can contain letters, digits, underscores (_), and hyphens (-). The name must start with a letter.
Protocol	Select a protocol for the listener. Valid values: ■ **TCP** ■ A connection-oriented protocol that provides high reliability. A logical connection must be established before data can be transmitted. ■ Session persistence is based on source IP addresses. ■ Source IP addresses are visible at the network layer. ■ Data is transmitted at a slow rate. ■ **UDP** ■ A connectionless and unreliable protocol. Three-way handshakes are not required before UDP packets are transmitted. UDP does not provide error recovery or data retransmission. ■ Data is transmitted at a high rate.
Port Number	Specify the listener port. The listener port is used to receive requests and forward requests to endpoints. Valid values: **1 to 65499**. You can specify at most 30 listener ports for each listener. Separate multiple listener ports with commas (,). Example: 80,90,8080. If you want to specify a port range, you can use a tilde (~). Example: 80~85. **Note** ■ If you add listeners that use the same protocol to a GA instance, you must configure different ports for the listeners. ■ You can specify more than 300 consecutive listener ports for a listener in specific regions. For more information, see Advanced listeners.

Parameter	Description
Client Affinity	Specify whether to enable client affinity. ■ If you select **Source IP Address** from the drop-down list, client affinity is enabled. After client affinity is enabled, requests from a specific client IP address are forwarded to the same endpoint. ■ If you select **Disable** from the drop-down list, client affinity is disabled. After client affinity is disabled, requests from a specific client IP address may be forwarded to different endpoints.

2. Configure endpoints.

   Each listener is associated with an endpoint group. You can associate an endpoint group with a listener by specifying the regions to which you want to distribute network traffic. After you associate an endpoint group with a listener, traffic is distributed to the optimal endpoint in the associated endpoint group.

   On the **Configure Endpoint Group** wizard page, set the following parameters and click **Next**.

   For more information about endpoint groups and endpoints, see Overview.

Parameter	Description
Endpoint Group Name	Enter a name for the endpoint group. The name must be 2 to 128 characters in length, and can contain letters, digits, underscores (_), and hyphens (-). The name must start with a letter.
Region	Select the region where you want to deploy the endpoint group.
Traffic Distribution Ratio	Set the traffic distribution ratio for the endpoint group. Unit: %. Valid values: 0 to 100. **Note** You can set **Traffic Distribution Ratio** only when you create an endpoint group for a TCP or UDP listener.
Backend Service	Select the region where you want to deploy backend servers. ○ **Alibaba Cloud**: Backend servers are deployed on Alibaba Cloud. ○ **Off Alibaba Cloud**: Backend servers are not deployed on Alibaba Cloud.
Preserve Client IP	Specify whether to preserve client IP addresses. After you enable this feature, backend servers can retrieve client IP addresses. For more information, see Preserve client IP addresses.

Parameter	Description
Endpoint	Endpoints are destinations of client requests. To add an endpoint, specify the following parameters: ◦ **Backend Service Type**: If your backend service is deployed on Alibaba Cloud, you can select **Alibaba Cloud Public IP Address**, **ECS**, **CLB**, **ALB**, or **OSS**. If your backend service is not deployed on Alibaba Cloud, you can select **Custom IP Address** or **Custom Domain Name**. Note ■ You can specify ECS, CLB, and ALB instances as endpoints only if your Alibaba Cloud account is included in the whitelist. If you want to specify ECS, CLB, or ALB instances as endpoints for your GA instances, submit a ticket to upgrade the GA instances. ■ The IP addresses of endpoint groups associated with each GA instance must be globally unique and not conflict with those of other GA instances. ■ If no service-linked role exists when you specify ECS instances, CLB instances, ALB instances, or OSS buckets as endpoints, the system automatically creates the corresponding service-linked role. For more information, see AliyunServiceRoleForGaVpcEndpoint, AliyunServiceRoleForGaAlb, and AliyunServiceRoleForGaOss. ◦ **Backend Service**: Enter the IP address, domain name, or instance ID of the backend server. ◦ **Weight**: Set a weight for the endpoint. Valid values: 0 to 255. GA distributes network traffic to endpoints based on their weights. Notice If the weight of an endpoint is set to 0, GA stops distributing network traffic to the endpoint. Proceed with caution. You can click **+ Add Endpoint** to add more endpoints. You can create at most four endpoints in each endpoint group. If you want to add more endpoints, go to the Quota Management page and increase the quota. For more information, see Manage quotas.

3. Confirm the configurations.

   On the **Confirm** wizard page, confirm the configurations of the listener and endpoint, and then click **Submit**.

   If you want to modify a specific setting, click **Modify** in the corresponding section. Then, you are redirected to the configuration page.

   Note If this is the first time you add a listener, the listener takes effect after 3 minutes. If you modify the configurations of a listener, the new configurations take effect after 1 minute.

Add an HTTP or HTTPS listener

1. Configure the listener and protocol.
   i. Log on to the GA console.

ii. On the **Instances** page, find the GA instance that you want to manage and click **Configure Listeners** in the **Actions** column.

iii. On the **Listener** tab, click **Add Listener**.

> **Note** If this is the first time that you add a listener, or the specified GA instance is not configured with a listener, skip this step.

iv. On the **Configure Listener & Protocol** wizard page, set the following parameters and click **Next**.

Parameter	Description
Listener Name	Enter a name for the listener. The name must be 2 to 128 characters in length, and can contain letters, digits, underscores (_), and hyphens (-). The name must start with a letter.

Protocol	Select a network transmission protocol for the listener. Valid values:
	■ **HTTPS**: HTTPS has the following features:
	  ■ A connection-oriented protocol that provides high reliability. A logical connection must be established before data can be transmitted.
	  ■ You can bind SSL certificates to servers. This ensures the high reliability of data.
	  ■ Data transmission is encrypted.
	■ **HTTP**: HTTP has the following features:
	  ■ A connection-oriented protocol that provides high reliability. A logical connection must be established before data can be transmitted.
	  ■ Data is transmitted at a high rate.
	  ■ Data transmission is not encrypted.

Port Number	Specify the listener port. The listener port is used to receive requests and forward requests to endpoints. Valid values: **1 to 65499**. You can configure only one listener port for each HTTP or HTTPS listener.
Client Affinity	Specify whether to enable client affinity. ■ If you select **Source IP Address** from the drop-down list, client affinity is enabled. After client affinity is enabled, requests from a specific client IP address are forwarded to the same endpoint. ■ If you select **Disable** from the drop-down list, client affinity is disabled. After client affinity is disabled, requests from a specific client IP address may be forwarded to different endpoints.

Parameter	Description
Advanced Settings	Click **Modify** and select **Add HTTP Header Fields**. ■ Add the `GA-ID` header to retrieve the ID of the GA instance. ■ Use the `GA-AP` header to retrieve the acceleration region of the GA instance. ■ Use the `GA-X-Forwarded-Proto` header to retrieve the listener protocol of the GA instance. ■ Use the `GA-X-Forwarded-Port` header to retrieve the listener port of the GA instance. ■ Use the `X-Real-IP` header to retrieve client IP addresses.

2. Optional. Configure the SSL certificate.

   You are required to configure an SSL certificate only when you add an HTTPS listener. SSL certificates ensure that data transmission over GA is encrypted.

   i. On the **Configure SSL Certificate** page, select the SSL certificate that you have purchased.

   ii. Click **Modify** to the right of **Advanced Settings** and select a TLS security policy from the **TLS Security Policies** drop-down list.

   For more information about TLS security policies, see TLS security policies.

   iii. Click **Next**.

3. Configure endpoints.

   Each listener is associated with an endpoint group. You can associate an endpoint group with a listener by specifying the regions to which you want to distribute network traffic. After you associate an endpoint group with a listener, traffic is distributed to the optimal endpoint in the associated endpoint group.

   On the **Configure Endpoint Group** wizard page, set the following parameters and click **Next**.

   For more information about endpoint groups and endpoints, see Overview.

Parameter	Description
Endpoint Group Name	Enter a name for the endpoint group. The name must be 2 to 128 characters in length, and can contain letters, digits, underscores (_), and hyphens (-). The name must start with a letter.
Region	Select the region where you want to deploy the endpoint group.
Backend Service	Select the region where you want to deploy backend servers. ○ **Alibaba Cloud**: Backend servers are deployed on Alibaba Cloud. ○ **Off Alibaba Cloud**: Backend servers are not deployed on Alibaba Cloud.
Preserve Client IP	Specify whether to preserve client IP addresses. By default, client IP address preservation is enabled for HTTP and HTTPS listeners. GA preserves the IP address of a client in the `X-Forwarded-For` HTTP header. For more information, see Preserve client IP addresses.

Parameter	Description
Endpoint	Endpoints are destinations of client requests. To add an endpoint, specify the following parameters: ○ **Backend Service Type**: If your backend service is deployed on Alibaba Cloud, you can select **Alibaba Cloud Public IP Address**, **ECS**, **CLB**, **ALB**, or **OSS**. If your backend service is not deployed on Alibaba Cloud, you can select **Custom IP Address** or **Custom Domain Name**. **Note** ■ You can specify ECS, CLB, and ALB instances as endpoints only if your Alibaba Cloud account is included in the whitelist. If you want to specify ECS, CLB, or ALB instances as endpoints for your GA instances, submit a ticket to upgrade the GA instances. ■ The IP addresses of endpoint groups associated with each GA instance must be globally unique and not conflict with those of other GA instances. ■ If no service-linked role exists when you specify ECS instances, CLB instances, ALB instances, or OSS buckets as endpoints, the system automatically creates the corresponding service-linked role. For more information, see AliyunServiceRoleForGaVpcEndpoint, AliyunServiceRoleForGaAlb, and AliyunServiceRoleForGaOss. ○ **Backend Service**: Enter the IP address, domain name, or instance ID of the backend server. ○ **Weight**: Set a weight for the endpoint. Valid values: 0 to 255. GA distributes network traffic to endpoints based on their weights. **Notice** If the weight of an endpoint is set to 0, GA stops distributing network traffic to the endpoint. Proceed with caution. You can click **+ Add Endpoint** to add more endpoints. You can create at most four endpoints in each endpoint group.
Backend Service Protocol	Select the protocol that the backend server uses. Valid values: ○ **HTTP**: This is the default value. ○ **HTTPS** **Note** ○ If the listener protocol is HTTP, this parameter is set to HTTP by default and cannot be modified. ○ You can set **Backend Service Protocol** only when you configure an endpoint group for an HTTP or HTTPS listener.

Parameter	Description
Port Mapping	If the listener port and the port that the endpoint uses to provide services are not the same, you must add a mapping between the ports. ○ **Listener Port**: Enter the listener port. ○ **Endpoint Port**: Enter the port that the endpoint uses to provide services. If the listener port and the port that the endpoint uses to provide services are the same, you do not need to add the port mapping. GA automatically distributes client requests to the listener port of the endpoint. **Note** You can set **Port Mapping** only when you configure an endpoint group for an HTTP or HTTPS listener.

4. Confirm the configurations.

   On the **Confirm** wizard page, confirm the configurations of the listener and endpoint, and then click **Submit**.

   If you want to modify a specific setting, click **Modify** in the corresponding section. Then, you are redirected to the configuration page.

   Note If this is the first time you add a listener, the listener takes effect after 3 minutes. If you modify the configurations of a listener, the new configurations take effect after 1 minute.

Note After you add an HTTP or HTTPS listener, you can configure a virtual endpoint group and a forwarding rule for the listener. Then, GA can simultaneously accelerate multiple domain names or paths to access your backend HTTP or HTTPS services. For more information, see Create and manage endpoint groups and Create and manage forwarding rules.

For more information, see Use one GA instance to accelerate multiple domain names over HTTPS.

What to do next

Operation	Description
Modify a listener	You can modify a listener to meet your business requirements. The configurations that you can modify include the basic settings, protocol, SSL certificate, and endpoint group of the listener. 1. On the **Listeners** tab, find the listener that you want to modify and click **Modify** in the **Actions** column. 2. On the **Edit Listener** page, modify the basic settings, protocol, SSL certificate, or endpoint group of the listener and then click **Next**. For more information about the basic settings, protocol, SSL certificate, and endpoint group of a listener, see Add a TCP or UDP listener or Add an HTTP or HTTPS listener.

Operation	Description
Delete a listener	You can delete a listener. After a listener is deleted, the endpoint group that is associated with the listener is also deleted. 1. On the **Listeners** tab, find the listener that you want to delete and click **Delete** in the **Actions** column. 2. In the **Delete Listener** message, click **OK**.

Related topics

- CreateListener: You can call this API operation to create a listener for a GA instance.
- UpdateListener: You can call this API operation to modify a specified listener of a GA instance.
- DeleteListener: You can call this API operation to delete a specified listener of a GA instance.

4.3. Associate and manage certificates

Global Accelerator (GA) allows you to associate multiple certificates with an HTTPS listener. This topic describes how to associate multiple certificates with an HTTPS listener. This topic also describes how to use virtual endpoint groups and forwarding rules to accelerate multiple domain names over HTTPS.

Prerequisites

- A GA instance and a basic bandwidth plan are purchased. For more information, see Create and manage GA instances and Purchase and manage basic bandwidth plans.
- An acceleration area is added. For more information, see Add and manage acceleration areas.
- An Internet Content Provider (ICP) number is obtained. All websites must obtain an ICP number before they are permitted to provide services to users in the Chinese mainland. For more information, see What is an ICP filing?.
- Multiple SSL certificates are issued to you. For more information, see Select and purchase certificates and Submit a certificate application.

Manage certificates that are associated with an HTTPS listener

When you create an HTTPS listener for a GA instance, you must configure an SSL certificate for identity authentication and encrypted data transmission. You can associate multiple certificates with an HTTPS listener of a GA instance. The following types of certificates are supported:

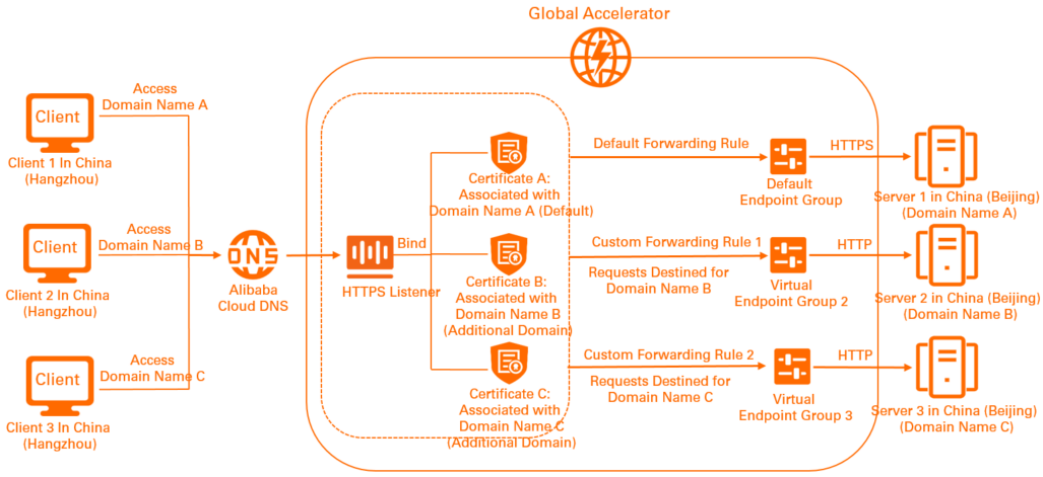
- Default certificate

  The SSL certificate that you configure when you create an HTTPS listener is used as the default certificate. You cannot delete the default certificate. You can only replace the default certificate.

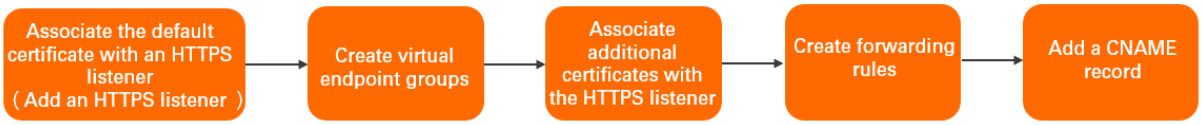
- Additional certificate

  You can associate additional certificates with an existing HTTPS listener. You can associate multiple domain names with an HTTPS listener by configuring additional certificates for the HTTPS listener. Then, you can create domain name-based forwarding rules to distribute client requests that are destined for different domain names to different endpoint groups.

Each HTTPS listener can be associated with at most three additional certificates. If you want to associate more additional certificates with an HTTPS listener, go to the Quota Management page and submit a ticket to increase the quota of **gaplus_quota_additional_certs_per_listener**. After the quota is increased, you can associate at most 10 additional certificates with an HTTPS listener. For more information, see Manage quotas.



Procedure



Step 1: Associate the default certificate with an HTTPS listener

The SSL certificate that you configure when you create an HTTPS listener is used as the default certificate. The endpoint group that you create is used as the default endpoint group. For more information about HTTPS listeners, see Add an HTTP or HTTPS listener.

1. Log on to the GA console.
2. On the **Instances** page, find the GA instance that you want to manage and click **Configure Listeners** in the **Actions** column.
3. On the **Listener** tab, click **Add Listener**.

   > **Note** If this is the first time that you add a listener, or the specified GA instance is not configured with a listener, skip this step.

4. On the **Configure Listener & Protocol** wizard page, set the required parameters, and click **Next**.
5. On the **Configure SSL Certificate** wizard page, select an SSL certificate and click **Next**.

   The certificate that you select is used as the default certificate of the HTTPS listener.

   You can also select a security policy in the **Advanced Settings** section based on your requirements. For more information about TLS security policies, see TLS security policies.
6. On the **Configure Endpoint Group** wizard page, configure the endpoint group and endpoints and click **Next**.

   The endpoint group that you configure is used as the default endpoint group of the HTTPS listener.

7. On the **Confirm** wizard page, confirm the configurations and click **Submit**.

Step 2: Create virtual endpoint groups

Create virtual endpoint groups. Each virtual endpoint group contains one of the origin servers. For more information, see Create a virtual endpoint group.

Step 3: Associate additional certificates with the HTTPS listener

1. Log on to the GA console.
2. On the **Instances** page, find the GA instance that you want to manage and click **Configure Listeners** in the **Actions** column.
3. On the **Listeners** tab, find the HTTPS listener that you want to manage and click the listener ID.
4. On the listener details page, click the **Certificates** tab.
5. On the **Certificates** tab, click **Associate Certificate** in the **Additional Certificate** section.
6. In the **Associate Certificate** dialog box, configure the additional certificate and click **OK**.
   - **Certificate**: Select the certificate that you want to associate.
   - **Associated Domain Name**: Select one or more domain names that you want to accelerate by using GA. The certificate is associated with the domain names that you select. You can select multiple domain names. Each additional certificate can be associated with at most three domain names.

   You can click **+ Add Certificate** to add multiple additional certificates at a time. Each HTTPS listener can be associated with at most three additional certificates. To associate more additional certificates with an HTTPS listener, go to the Quota Management page and submit a ticket to increase the quota of **gaplus_quota_additional_certs_per_listener**. For more information, see Manage quotas.

Step 4: Create forwarding rules

Create a domain name-based forwarding rule for each virtual endpoint group. For more information, see Create and manage forwarding rules.

Step 5: Add a CNAME record

Add CNAME records for the domain names that you want to accelerate. To forward requests from clients to GA, you must modify the DNS record to map the domain names that you want to accelerate to the CNAME of the GA instance. For more information, see 配置CNAME.

What to do next

Operation	Description
Replace the default certificate	1. On the **Listeners** tab, find the HTTPS listener that you want to manage and click the listener ID. 2. On the listener details page, click the **Certificates** tab. 3. In the **Default Server Certificate** section of the **Certificates** tab, click **Replace** in the Actions column. 4. In the Change Default Server Certificate dialog box, select the certificate that you want to use and click **OK**.

Operation	Description
Disassociate an additional certificate	You can only disassociate additional certificates from an HTTPS listener in the GA console. If you want to delete a certificate, see Delete an SSL certificate. 1. On the **Listeners** tab, find the HTTPS listener that you want to manage and click the listener ID. 2. On the listener details page, click the **Certificates** tab. 3. In the **Additional Certificate** section of the **Certificates** tab, disassociate one or more additional certificates based on the following information. ○ Disassociate one additional certificate: Find the certificate that you want to disassociate and click **Disassociate** in the **Actions** column. ○ Disassociate multiple additional certificates: Select the additional certificates that you want to disassociate and click **Batch Disassociate**. 4. In the message that appears, click **OK**.

References

- AssociateAdditionalCertificatesWithListener: You can call this API operation to associate additional certificates with an HTTPS listener.
- DissociateAdditionalCertificatesFromListener: You can call this API operation to disassociate one or more additional certificates from an HTTPS listener.
- ListListenerCertificates: You can call this API operation to query the certificates that are associated with an HTTPS listener.

4.4. TLS security policies

You can select a Transport Layer Security (TLS) security policy when you create an HTTPS listener for a Global Accelerator (GA) instance. By default, the system selects the **tls_cipher_policy_1_0** security policy. If you require higher security, you can select a TLS security policy of a higher level.

TLS security policies

A TLS security policy contains TLS protocol versions and cipher suites that are available for HTTPS. A later TLS version offers higher security but comprises compatibility with browsers. The following table describes the TLS protocol versions and cipher suites that are supported by each TLS security policy.

Security policy	Supported TLS version	Supported cipher suite

Security policy	Supported TLS version	Supported cipher suite
tls_cipher_policy_1_0	TLS 1.0, TLS 1.1, and TLS 1.2	• ECDHE-RSA-AES128-GCM-SHA256 • ECDHE-RSA-AES256-GCM-SHA384 • ECDHE-RSA-AES128-SHA256 • ECDHE-RSA-AES256-SHA384 • AES128-GCM-SHA256 • AES256-GCM-SHA384 • AES128-SHA256 • AES256-SHA256 • ECDHE-RSA-AES128-SHA • ECDHE-RSA-AES256-SHA • AES128-SHA • AES256-SHA • DES-CBC3-SHA
tls_cipher_policy_1_1	TLS 1.1 and TLS 1.2	• ECDHE-RSA-AES128-GCM-SHA256 • ECDHE-RSA-AES256-GCM-SHA384 • ECDHE-RSA-AES128-SHA256 • ECDHE-RSA-AES256-SHA384 • AES128-GCM-SHA256 • AES256-GCM-SHA384 • AES128-SHA256 • AES256-SHA256 • ECDHE-RSA-AES128-SHA • ECDHE-RSA-AES256-SHA • AES128-SHA • AES256-SHA • DES-CBC3-SHA

Security policy	Supported TLS version	Supported cipher suite
tls_cipher_policy_1_2	TLSv1.2	• ECDHE-RSA-AES128-GCM-SHA256 • ECDHE-RSA-AES256-GCM-SHA384 • ECDHE-RSA-AES128-SHA256 • ECDHE-RSA-AES256-SHA384 • AES128-GCM-SHA256 • AES256-GCM-SHA384 • AES128-SHA256 • AES256-SHA256 • ECDHE-RSA-AES128-SHA • ECDHE-RSA-AES256-SHA • AES128-SHA • AES256-SHA • DES-CBC3-SHA
tls_cipher_policy_1_2_strict	TLSv1.2	• ECDHE-RSA-AES128-GCM-SHA256 • ECDHE-RSA-AES256-GCM-SHA384 • ECDHE-RSA-AES128-SHA256 • ECDHE-RSA-AES256-SHA384 • ECDHE-RSA-AES128-SHA • ECDHE-RSA-AES256-SHA
tls_cipher_policy_1_2_strict_with_1_3	TLS 1.2 and TLS 1.3	• TLS_AES_128_GCM_SHA256 • TLS_AES_256_GCM_SHA384 • TLS_CHACHA20_POLY1305_SHA256 • TLS_AES_128_CCM_SHA256 • TLS_AES_128_CCM_8_SHA256 • ECDHE-ECDSA-AES128-GCM-SHA256 • ECDHE-ECDSA-AES256-GCM-SHA384 • ECDHE-ECDSA-AES128-SHA256 • ECDHE-ECDSA-AES256-SHA384 • ECDHE-RSA-AES128-GCM-SHA256 • ECDHE-RSA-AES256-GCM-SHA384 • ECDHE-RSA-AES128-SHA256 • ECDHE-RSA-AES256-SHA384 • ECDHE-ECDSA-AES128-SHA • ECDHE-ECDSA-AES256-SHA • ECDHE-RSA-AES128-SHA • ECDHE-RSA-AES256-SHA

Cipher suites that are supported by TLS security policies

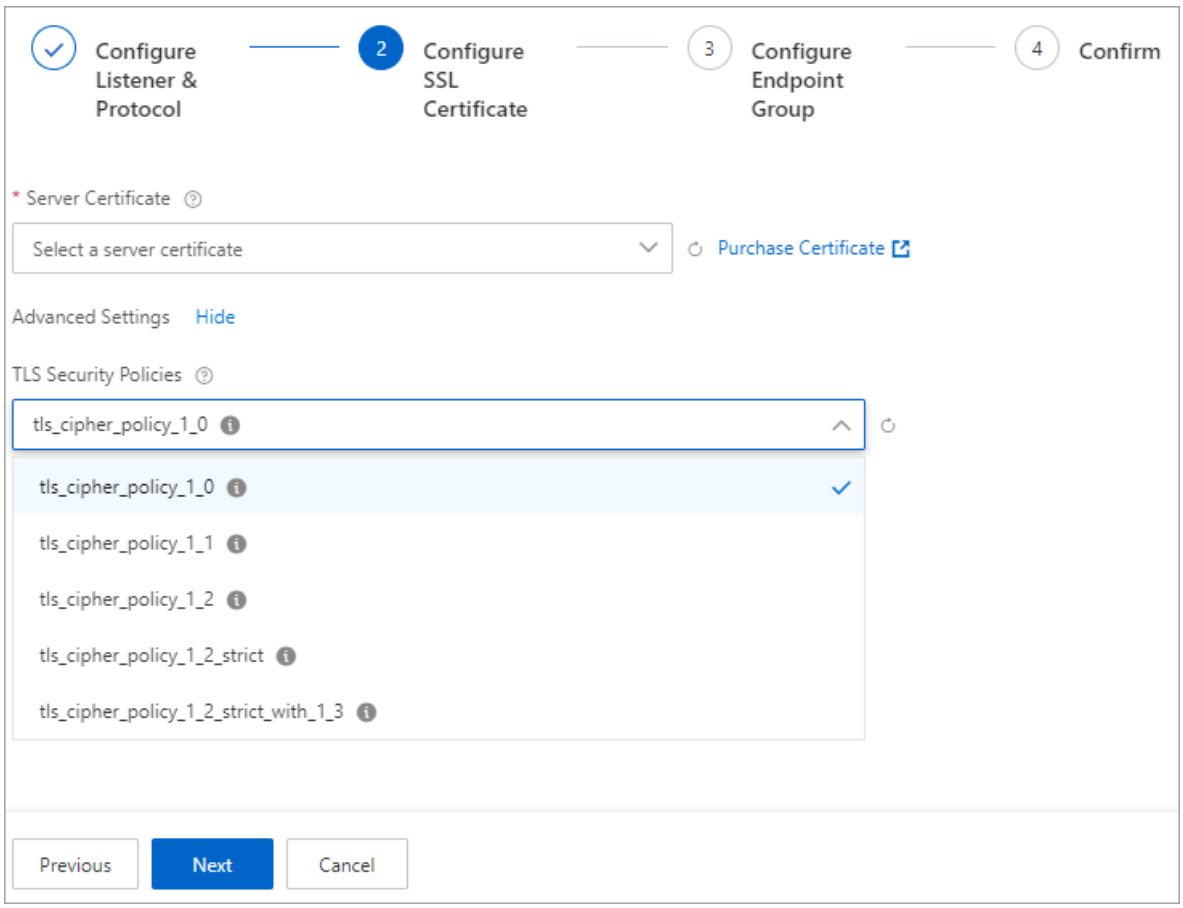
Security policy		tls_cipher_policy_1_0	tls_cipher_policy_1_1	tls_cipher_policy_1_2	tls_cipher_policy_1_2_strict	tls_cipher_policy_1_2_strict_with_1_3
TLS		1.0, 1.1, and 1.2	1.1 and 1.2	1.2	1.2	1.2 and 1.3
CIPHER	ECDHE-RSA-AES128-GCM-SHA256	✔	✔	✔	✔	✔
	ECDHE-RSA-AES256-GCM-SHA384	✔	✔	✔	✔	✔
	ECDHE-RSA-AES128-SHA256	✔	✔	✔	✔	✔
	ECDHE-RSA-AES256-SHA384	✔	✔	✔	✔	✔
	AES128-GCM-SHA256	✔	✔	✔	-	-
	AES256-GCM-SHA384	✔	✔	✔	-	-
	AES128-SHA256	✔	✔	✔	-	-
	AES256-SHA256	✔	✔	✔	-	-
	ECDHE-RSA-AES128-SHA	✔	✔	✔	✔	✔
	ECDHE-RSA-AES256-SHA	✔	✔	✔	✔	✔
	AES128-SHA	✔	✔	✔	-	-
	AES256-SHA	✔	✔	✔	-	-
	DES-CBC3-SHA	✔	✔	✔	-	-
	TLS_AES_128_GCM_SHA256	-	-	-	-	✔
	TLS_AES_256_GCM_SHA384	-	-	-	-	✔
	TLS_CHACHA20_POLY1305_SHA256	-	-	-	-	✔
	TLS_AES_128_CCM_SHA256	-	-	-	-	✔

Security policy		tls_cipher_policy_1_0	tls_cipher_policy_1_1	tls_cipher_policy_1_2	tls_cipher_policy_1_2_strict	tls_cipher_policy_1_2_strict_with_1_3
	TLS_AES_128_CCM_8_SHA256	-	-	-	-	✔
	ECDHE-ECDSA-AES128-GCM-SHA256	-	-	-	-	✔
	ECDHE-ECDSA-AES256-GCM-SHA384	-	-	-	-	✔
	ECDHE-ECDSA-AES128-SHA256	-	-	-	-	✔
	ECDHE-ECDSA-AES256-SHA384	-	-	-	-	✔
	ECDHE-ECDSA-AES128-SHA	-	-	-	-	✔
	ECDHE-ECDSA-AES256-SHA	-	-	-	-	✔

Note The √ sign in the preceding table indicates that a cipher suite is supported, while the - sign indicates that a cipher suite is not supported.

Select a TLS security policy

By default, the system selects the tls_cipher_policy_1_0 security policy when you create or configure an HTTPS listener. You can change the TLS security policy in the advanced settings. For more information, see Add an HTTP or HTTPS listener.



5.Endpoint groups and endpoints

5.1. Overview

Each listener is associated with an endpoint group, and each endpoint group contains one or more endpoints.

Endpoint groups

Each endpoint group is associated with a specific region. You can associate an endpoint group with a listener by specifying the region to which you want to distribute network traffic. After you associate an endpoint group with a listener, the system distributes network traffic to the optimal endpoints in the endpoint group.

Listeners that use different protocols support different types of endpoint groups:

- TCP or UDP listeners

  By default, you can create two default endpoint groups for each TCP or UDP listener. If you want to create more default endpoint groups, go to the Quota Management page and increase the quota of **gaplus_quota_epgs_per_listener**. For more information, see Manage quotas.

  You must deploy default endpoint groups in different regions. You can set a traffic distribution ratio for each default endpoint group. The traffic distribution ratio specifies the proportion of traffic that is distributed to a default endpoint group.

- HTTP or HTTPS listeners

  By default, you can create one default endpoint group and one virtual endpoint group for each HTTP or HTTPS listener. If you want to create multiple virtual endpoint groups, go to the Quota Management page and increase the quota of **gaplus_quota_vepg_per_listener**. For more information, see Manage quotas.

  - A default endpoint group refers to the endpoint group that you configure when you create an HTTP or HTTPS listener.
  - A virtual endpoint group refers to the endpoint group that you can create on the **Endpoint Group** page after you create a listener.

    After you create a virtual endpoint group for an HTTP or HTTPS listener, you can create a forwarding rule and associate the forwarding rule with the virtual endpoint group. Then, the HTTP or HTTPS listener forwards requests with different destination domain names or paths to the default or virtual endpoint group based on the forwarding rule. This way, you can use one GA instance to accelerate multiple domain names or paths. For more information about how to create a forwarding rule, see Create and manage forwarding rules.

Endpoints

Endpoints are destinations of client requests. You can add at most four endpoints to an endpoint group. The following table describes the backend service types of endpoints.

Backend service area	Network type	Backend service type	Backend service
Alibaba Cloud	Internet	Alibaba Cloud public IP address	• Elastic IP addresses (EIPs) • Static public IP addresses Static public IP addresses include the public IP addresses of Elastic Compute Service (ECS) instances and the public IP addresses of Internet-facing Classic Load Balancer (CLB) instances that are deployed in classic networks.
	VPC	ECS	ECS instances Only ECS instances that are deployed in virtual private clouds (VPCs) are supported.
		CLB	CLB instances Only CLB instances that are deployed in VPCs are supported.
		ALB	Application Load Balancer (ALB) instances
		OSS	Object Storage Service (OSS) buckets
Outside Alibaba Cloud	Internet	Custom IP addresses	Custom IP addresses of origin servers
		Custom domain names	Custom domain names of origin servers

> **Note**
> - You can specify ECS, CLB, and ALB instances as endpoints only if your Alibaba Cloud account is included in the whitelist. If you want to specify ECS, CLB, or ALB instances as endpoints for your GA instances, submit a ticket to upgrade the GA instances.
> - The IP addresses of endpoint groups associated with each GA instance must be globally unique and not conflict with those of other GA instances.

You can specify a weight for an endpoint. The weight specifies the proportion of traffic that is forwarded to the endpoint. GA calculates the sum of all endpoint weights in an endpoint group. Then, traffic is forwarded to endpoints based on the proportions of their weights. For more information, see What to do next.

Health checks

You can enable health checks for endpoint groups of a GA instance. This improves service reliability and availability and prevents service interruptions caused by unhealthy endpoints.

After you enable health checks for an endpoint group, GA periodically checks whether the endpoints are healthy. When GA detects an unhealthy endpoint, GA distributes new requests to other healthy endpoints. When the unhealthy endpoint recovers, GA distributes requests to the endpoint again. For more information, see Enable and manage health checks.

References

- Create and manage endpoint groups
- Create and manage forwarding rules
- Enable and manage health checks

5.2. Distribute traffic across endpoint groups in different scenarios

Global Accelerator (GA) allows you to configure multiple endpoint groups that are deployed in different regions for a TCP listener or UDP listener. You can set a traffic distribution ratio for an endpoint group to control the percentage of client requests that are forwarded to the endpoint group. You can also enable health checks to filter out unhealthy endpoint groups.

Distribute traffic across endpoint groups

Introduction to traffic distribution

GA allows you to set traffic distribution ratios for endpoint groups. You can modify the percentage of client requests that are forwarded to each endpoint group based on your business requirements. This helps improve user experience.

- Traffic distribution ratio: specifies the percentage of client requests that are distributed. Valid values: 0% to 100%. Default value: 100%. A value of 0% indicates that the endpoint group is ignored and no client request is forwarded to the endpoint group. A value of 100% indicates that all client requests are forwarded to the endpoint group.
- Endpoint group priority: The client requests that are forwarded to an endpoint group depend on the traffic distribution ratio that you set and the priority of the endpoint group. GA calculates the priority of each endpoint group based on the network latency. The network latency varies based on geographical locations and network hops. In most cases, endpoint groups that are closer to access points have fewer network hops and are assigned higher priorities. Client requests are preferably forwarded to the endpoint group whose region is closest to a specific access point.

> **Note** After you enable health checks for each endpoint group, if the endpoint group with a higher priority fails the health check, all client requests are forwarded to the endpoint group with a lower priority. The client requests are forwarded to the corresponding endpoint group regardless of the traffic distribution ratio that you set.

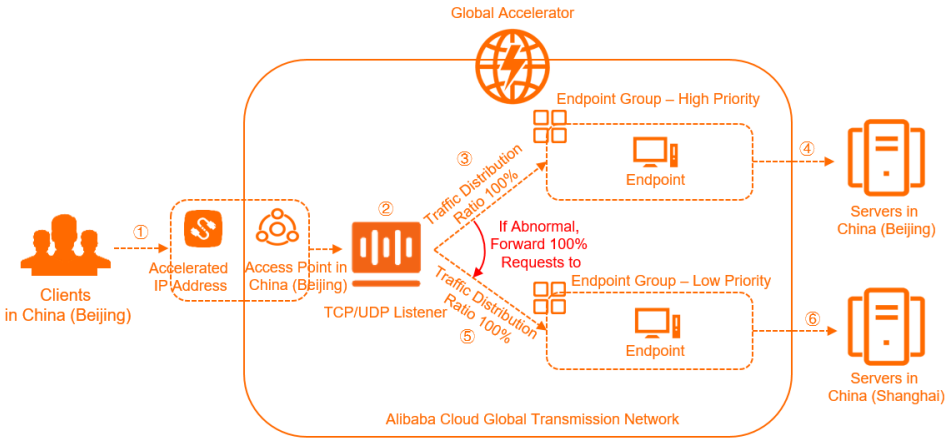
Traffic distribution formula

The following examples show how traffic distribution works:

- One acceleration region with multiple endpoint groups

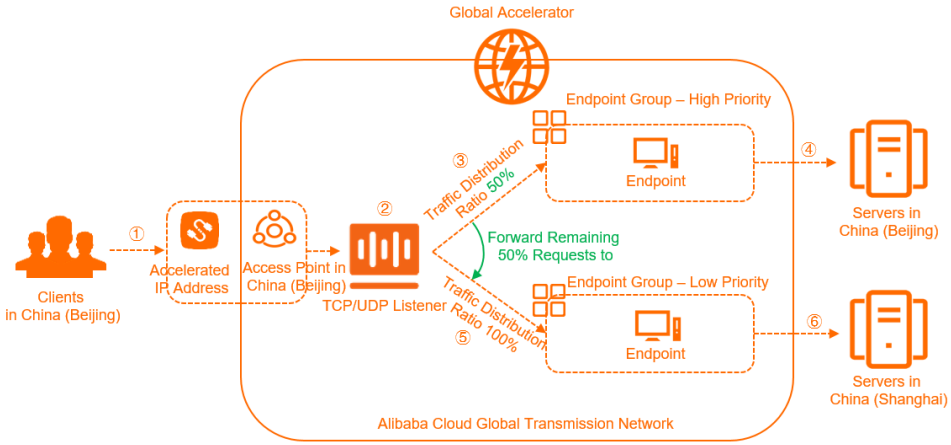
  Clients in the China (Beijing) region want to access an application. The servers that host the application are deployed in the China (Beijing) and China (Shanghai) regions. You specify China (Beijing) as the acceleration region and create an endpoint group in the China (Beijing) region and the China (Shanghai) region. You want to forward client requests that are sent to the China (Beijing) region and the China (Shanghai) region based on your requirements.

- Set the traffic distribution ratio of each endpoint group to 100%.



No.	Description
①	Client requests are scheduled to the nearest access point in the China (Beijing) region and then forwarded to the Alibaba Cloud global transmission network.
②	The listener of the GA instance checks the connection requests from clients based on the protocol and port that are configured and forwards the client requests to endpoint groups based on their priorities and traffic distribution ratios.
③	The priority of the endpoint group in the China (Beijing) region is higher than that of the endpoint group in the China (Shanghai) region. The endpoint group in the China (Beijing) region passes the health check and the traffic distribution ratio of the endpoint group is set to 100%. All client requests are forwarded to the endpoint group in the China (Beijing) region.
④	Client requests are processed by servers in the China (Beijing) region.
⑤	If the endpoint group in the China (Beijing) region fails the health check but the endpoint group in the China (Shanghai) region passes the health check, the listener forwards all client requests to the endpoint group with a lower priority in the China (Shanghai) region.
⑥	Client requests are processed by servers in the China (Shanghai) region.

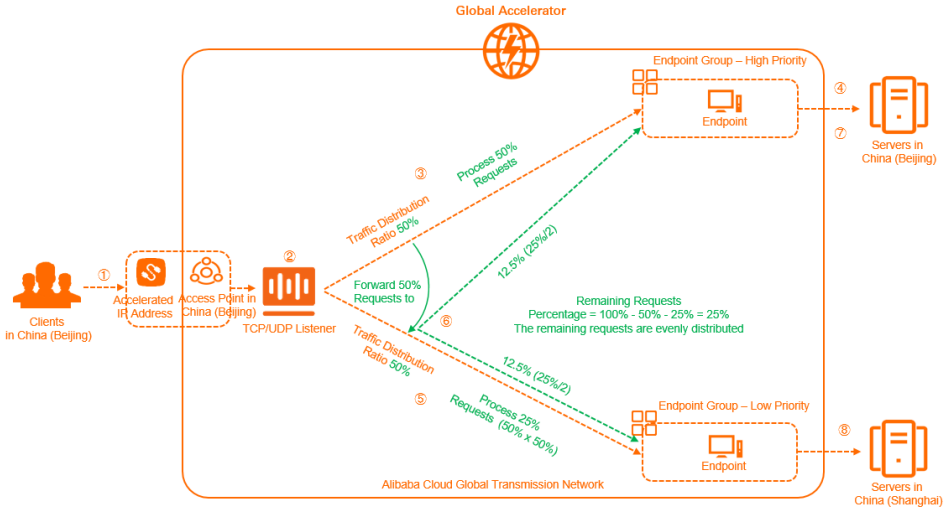
- Set the traffic distribution ratio to 50% for the endpoint group in the China (Beijing) region and set the traffic distribution ratio to 100% for the endpoint group in the China (Shanghai) region. You can change the traffic distribution ratio based on your business requirements.



This scenario is similar to the scenario in which you set the traffic distribution ratio to 100% for both endpoint groups. Requests from clients in the China (Beijing) region are preferably forwarded to the endpoint group in the China (Beijing) region. After you set the traffic distribution ratio to 50% for the endpoint group in the China (Beijing) region, 50% of client requests are forwarded to the endpoint group in the China (Beijing) region and the remaining 50% of client requests are forwarded to the endpoint group in the China (Shanghai) region. If you set the traffic distribution ratio to 30% for the endpoint group in the China (Beijing) region, 30% of client requests are forwarded to the endpoint group in the China (Beijing) region and 70% of client requests are forwarded to the endpoint group in the China (Shanghai) region.

If you set the traffic distribution ratio to 100% for the endpoint group in the China (Shanghai) region, all the remaining client requests are forwarded to the endpoint group in the China (Shanghai) region. In the preceding two examples, 50% and 70% of client requests are forwarded to the endpoint group in the China (Shanghai) region.

- Set the traffic distribution ratio to 50% for both endpoint groups. You can change the traffic distribution ratio based on your business requirements.

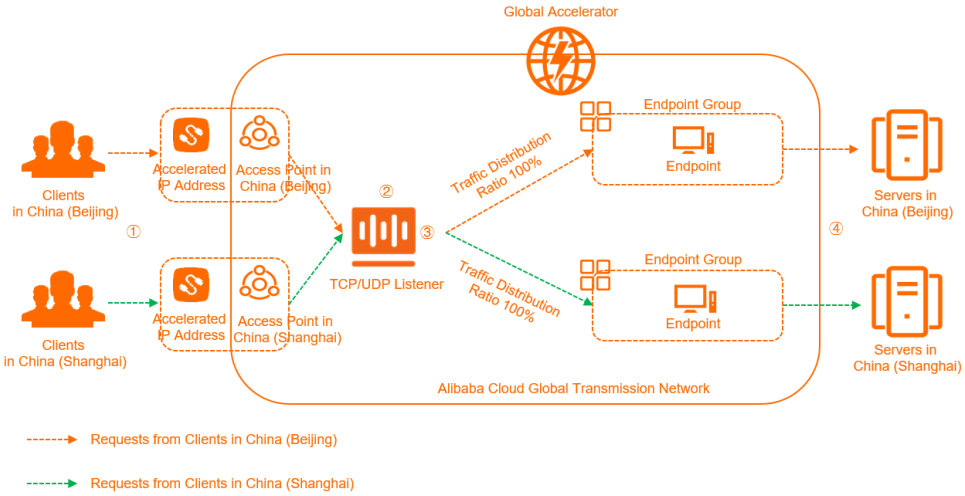


No.	Description
①	Client requests are scheduled to the nearest access point in the China (Beijing) region and then forwarded to the Alibaba Cloud global transmission network.
②	The listener of the GA instance checks the connection requests from clients based on the protocol and port that are configured and forwards the client requests to endpoint groups based on their priorities and traffic distribution ratios.
③	The priority of the endpoint group in the China (Beijing) region is higher than that of the endpoint group in the China (Shanghai) region. The endpoint group in the China (Beijing) region passes the health check and the traffic distribution ratio of the endpoint group is set to 50%. 50% of client requests are forwarded to the endpoint group in the China (Beijing) region.
④	Servers in the China (Beijing) region process 50% of client requests.
⑤	The remaining 50% of client requests are first forwarded to the endpoint group in the China (Shanghai) region. The percentage of client requests that are received by the endpoint group in the China (Shanghai) region is 25% based on the following formula: 50% × 50% = 25%. The endpoint group in the China (Beijing) region receives 50% of client requests and the endpoint group in the China (Shanghai) region receives 25% of client requests. The remaining 25% of client requests are not received.
⑥	GA evenly distributes the remaining client requests to each endpoint group. The remaining 25% of client requests are evenly distributed to each endpoint group. This indicates that each endpoint group in the China (Beijing) region and the China (Shanghai) region receives 12.5% of client requests.
⑦	Servers in the China (Beijing) region process 12.5% of client requests.
⑧	Servers in the China (Shanghai) region process 37.5% of client requests based on the following formula: 25% + 12.5% = 37.5%.

- Multiple acceleration regions with multiple endpoint groups

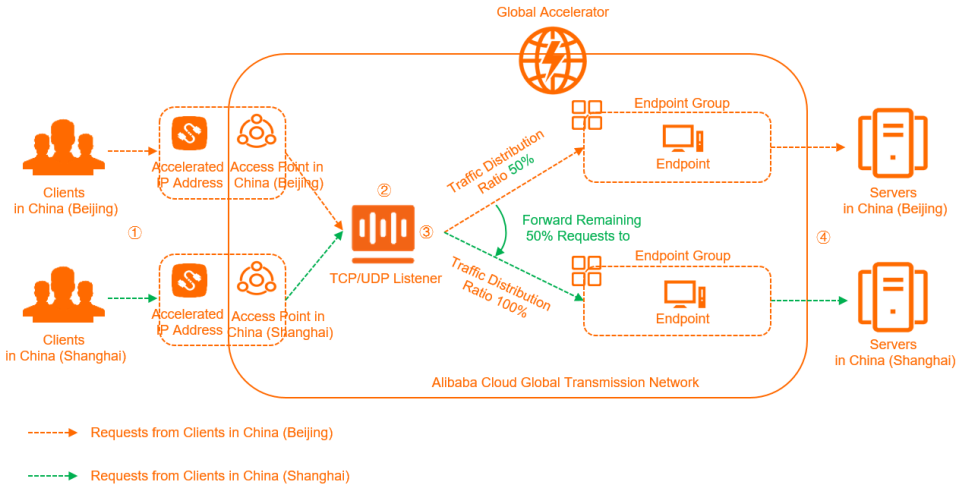
  If you specify multiple acceleration regions for clients that are located in multiple regions, the clients can connect to the nearest access points of the Alibaba Cloud global transmission network by sending requests to the accelerated IP addresses. Then, the client requests are forwarded to the endpoint groups that are closest to the access points.

  - Set the traffic distribution ratio of each endpoint group to 100%.



No.	Description
①	Requests from clients in the China (Beijing) region are forwarded to the nearest access point in the China (Beijing) region. Requests from clients in the China (Shanghai) region are forwarded to the nearest access point in the China (Shanghai) region. Then, the client requests are forwarded to the Alibaba Cloud global transmission network.
②	The listener of the GA instance checks the connection requests from clients based on the protocol and port that are configured and forwards the client requests to endpoint groups based on their priorities and traffic distribution ratios.
③	GA distributes client requests from different regions based on the traffic distribution ratio. ■ Forward client requests from the China (Beijing) region The priority of the endpoint group in the China (Beijing) region is higher than that of the endpoint group in the China (Shanghai) region. The endpoint group in the China (Beijing) region passes the health check and the traffic distribution ratio of the endpoint group is set to 100%. All client requests from the China (Beijing) region are forwarded to the endpoint group in the China (Beijing) region. ■ Forward client requests from the China (Shanghai) region The priority of the endpoint group in the China (Shanghai) region is higher than that of the endpoint group in the China (Beijing) region. The endpoint group in the China (Shanghai) region passes the health check and the traffic distribution ratio of the endpoint group is set to 100%. All client requests from the China (Shanghai) region are forwarded to the endpoint group in the China (Shanghai) region.
④	Servers in the China (Beijing) region and the China (Shanghai) region process the client requests that they receive.

- Set the traffic distribution ratio to 50% for the endpoint group in the China (Beijing) region and set the traffic distribution ratio to 100% for the endpoint group in the China (Shanghai) region. You can change the traffic distribution ratio based on your business requirements.

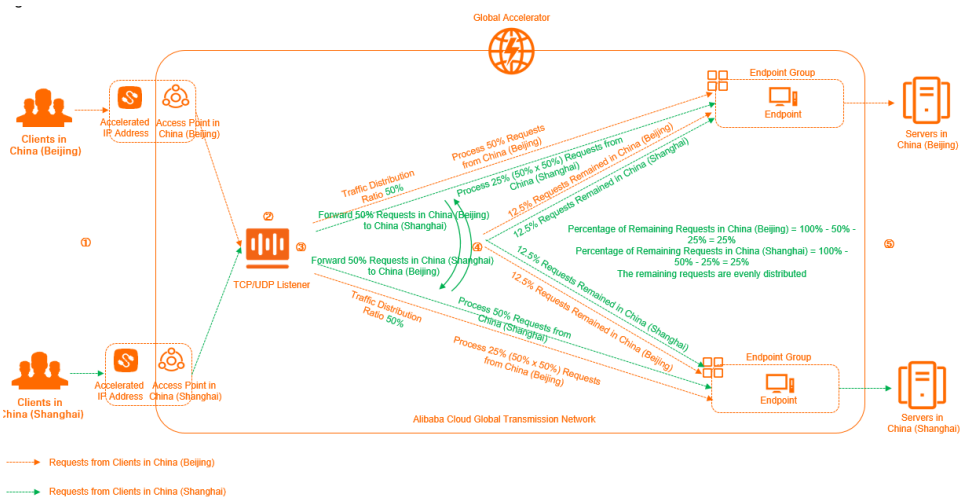


This scenario is similar to the scenario in which you set the traffic distribution ratio to 100% for both endpoint groups. Requests from clients in the China (Beijing) region are preferably forwarded to the endpoint group in the China (Beijing) region. After you set the traffic distribution ratio to 50% for the endpoint group in the China (Beijing) region, 50% of client requests are forwarded to the endpoint group in the China (Beijing) region and the remaining 50% of client requests are forwarded to the endpoint group in the China (Shanghai) region. If you set the traffic distribution ratio to 30% for the endpoint group in the China (Beijing) region, 30% of client requests are forwarded to the endpoint group in the China (Beijing) region and 70% of client requests are forwarded to the endpoint group in the China (Shanghai) region.

All requests from clients in the China (Shanghai) region are forwarded to the endpoint group in the China (Shanghai) region. This is because you set the traffic distribution ratio to 100% for the endpoint group in the China (Shanghai) region.

In this scenario, the endpoint group in the China (Beijing) region receives 50% of requests from clients in the China (Beijing) region. The endpoint group in the China (Shanghai) receives 100% of requests from clients in the China (Shanghai) region and 50% of requests from clients in the China (Beijing) region.

- Set the traffic distribution ratio to 50% for both endpoint groups. You can change the traffic distribution ratio based on your business requirements.



No.	Description
①	Requests from clients in the China (Beijing) region are forwarded to the nearest access point in the China (Beijing) region. Requests from clients in the China (Shanghai) region are forwarded to the nearest access point in the China (Shanghai) region. Then, the client requests are forwarded to the Alibaba Cloud global transmission network.
②	The listener of the GA instance checks the connection requests from clients based on the protocol and port that are configured and forwards the client requests to endpoint groups based on their priorities and traffic distribution ratios.
③	GA distributes client requests from different regions based on the traffic distribution ratio. ■ Forward client requests from the China (Beijing) region The priority of the endpoint group in the China (Beijing) region is higher than that of the endpoint group in the China (Shanghai) region. The endpoint group in the China (Beijing) region passes the health check and the traffic distribution ratio of the endpoint group is set to 50%. 50% of client requests are forwarded to the endpoint group in the China (Beijing) region. The remaining 50% of client requests are forwarded to the endpoint group in the China (Shanghai) region. The percentage of client requests that are received by the endpoint group in the China (Shanghai) region is 25% based on the following formula: 50% × 50% = 25%. The requests from clients in the China (Beijing) region that are not received is 25% based on the following formula: 100% - 50% - 25% = 25%. ■ Forward client requests from the China (Shanghai) region The priority of the endpoint group in the China (Shanghai) region is higher than that of the endpoint group in the China (Beijing) region. The endpoint group in the China (Shanghai) region passes the health check and the traffic distribution ratio of the endpoint group is set to 50%. 50% of client requests are forwarded to the endpoint group in the China (Shanghai) region. The remaining 50% of client requests are forwarded to the endpoint group in the China (Beijing) region. The percentage of client requests that are received by the endpoint group in the China (Beijing) region is 25% based on the following formula: 50% × 50% = 25%. The percentage of requests from clients in the China (Shanghai) region that are not received is 25% based on the following formula: 100% - 50% - 25% = 25%.
④	GA evenly forwards the remaining client requests to each endpoint group. The remaining 25% of requests from clients in the China (Beijing) region are evenly distributed to each endpoint group. This indicates that each endpoint group in the China (Beijing) region and the China (Shanghai) region receives 12.5% of client requests. Each endpoint group in the China (Beijing) region and the China (Shanghai) region receives 12.5% of requests from clients in the China (Shanghai) region.
⑤	Servers in the China (Beijing) region and the China (Shanghai) region process the client requests that they receive.

Scenarios

Overview

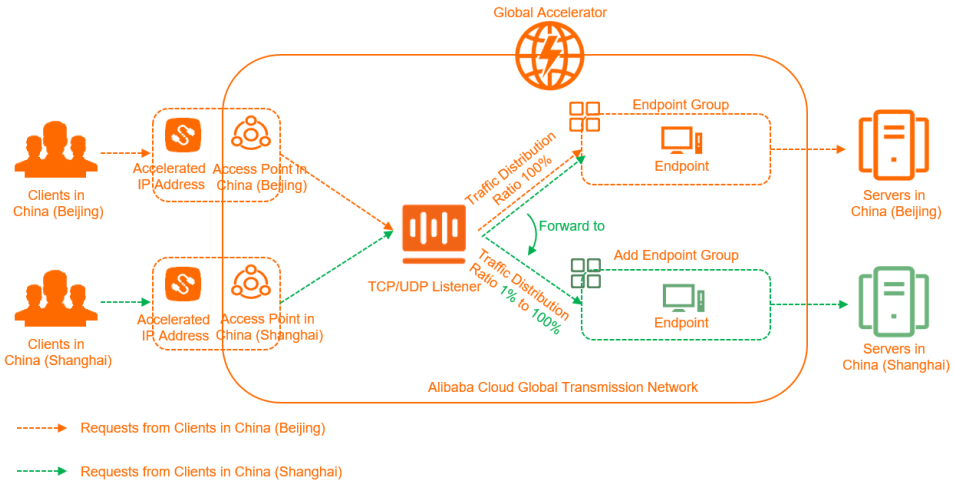
Scenario	Description
Deploy an application in multiple regions	The servers do not meet the requirements of an application or users in specific regions have poor network experience. For example, users in different regions share the same acceleration region or multiple acceleration regions share one endpoint group. In this case, you can deploy the application in another region.
Forward client requests across regions	If you deploy a service in a single region, a large number of client requests may be sent to the service and the servers that host the service may become overloaded. To resolve the issues, you can deploy the service across regions and add an endpoint group in each region. Then, you can use the traffic distribution feature to change the percentage of client requests that are forwarded to each region to reduce the loads on the servers in a region.
Cross-region disaster recovery for applications	If you have requirements for service continuity and high availability, you can deploy the service across regions, specify the backend service in different regions as the endpoint group, and enable health checks for the endpoint groups. If the service in a region cannot be accessed, you can enable GA to forward client requests to healthy endpoint groups in other regions. This meets the requirements of disaster recovery.
Unpublish or update a service based on regions	You want to adjust your business in a region. For example, if you want to smoothly unpublish a service that receives low traffic in a region or update a service in a region, you can set the traffic distribution ratio for the endpoint group in the region to migrate the service in a flexible manner.

Deploy an application in multiple regions

If you want to scale out your business and the servers do not meet the requirements of the application or users in specific regions have poor network experience, you can deploy the application in another region. You can add endpoint groups or acceleration regions for the GA instance to improve user experience.

- Add endpoint groups to improve the traffic processing capabilities of the application.

The scenario in the following figure is used as an example. An application is deployed on servers in the China (Beijing) region. Clients in the China (Beijing) region connect to the access point in the China (Beijing) region. Clients in the China (Shanghai) region connect to the access point in the China (Shanghai) region. All client requests are processed by the servers in the endpoint group in the China (Beijing) region. As the number of clients increases, the loads on the servers also increase.



In this case, you can add an endpoint group in the China (Shanghai) region and forward requests from clients in the China (Shanghai) region to the servers in the endpoint group in the China (Shanghai) region. This improves the availability of your application. To add the endpoint group, perform the following steps:

i. Deploy servers in the China (Shanghai) region.

ii. Add an endpoint group in the China (Shanghai) region for the listener of a GA instance. For more information, see Create a default endpoint group.

When you add the endpoint group in the China (Shanghai) region, you can set the traffic distribution ratio to a lower value for testing. For example, you can set the value to 1%.

iii. Check how requests from clients in the China (Shanghai) region are distributed.

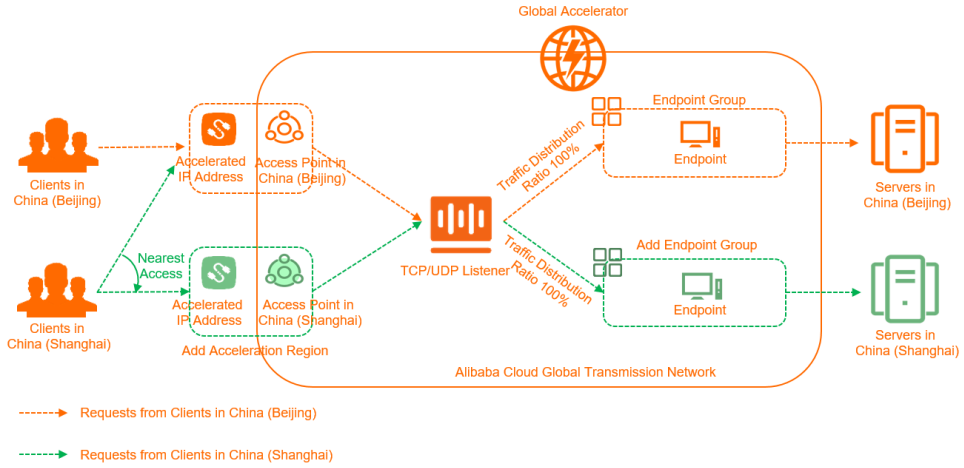
Requests from clients in the China (Beijing) region are processed by the servers in the endpoint group in the China (Beijing) region and 1% of requests from clients in the China (Shanghai) region are processed by the servers in the endpoint group in the China (Shanghai) region. The remaining 99% of client requests are processed by the servers in the endpoint group in the China (Beijing) region.

iv. After the testing is completed, change the traffic distribution ratio of the endpoint group in the China (Shanghai) region to 100%.

This way, all requests from clients in the China (Shanghai) region are forwarded to the servers in the endpoint group in the China (Shanghai) region. The servers in the endpoint group in the China (Beijing) region do not process requests from clients in the China (Shanghai) region. For more information, see Set the traffic distribution ratio for an endpoint group.

- Add an acceleration region to improve user experience

The scenario in the following figure is used as an example. An application is deployed on servers in the China (Beijing) region. Clients in the China (Beijing) region and the China (Shanghai) region connect to the Alibaba Cloud global transmission network by sending requests to the access point in the China (Beijing) region. All client requests are processed by the servers in the endpoint group in the China (Beijing) region. When clients in the China (Shanghai) region access the application, network issues such as network latency and network jitter frequently occur.

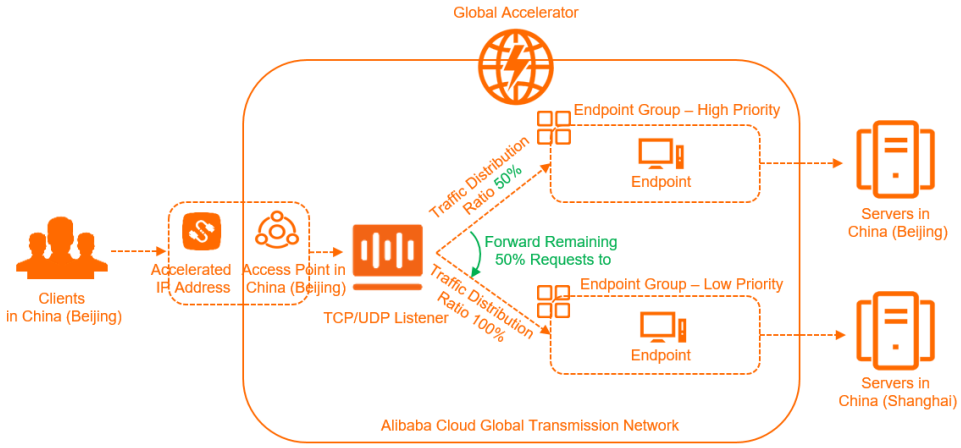


You can deploy the application on servers in the China (Shanghai) region, add China (Shanghai) as the acceleration region, and create an endpoint group in the China (Shanghai) region for the GA instance. Requests from clients in the China (Shanghai) region are forwarded to the nearest access point in the China (Shanghai) region. The listener then checks the connection requests and forwards the requests to the endpoint group that is close to the access point in the China (Shanghai) region. This improves experience for clients in the China (Shanghai) region. For more information, see Add and manage acceleration areas and Create a default endpoint group.

Forward client requests across regions

You can use the traffic distribution feature to forward client requests from a specific acceleration region to multiple endpoint groups that are deployed in different regions. This reduces the loads on the servers in the endpoint group of the acceleration region.

The scenario in the following figure is used as an example. An application is deployed on servers in the China (Beijing) region and the China (Shanghai) region. The clients are located in the China (Beijing) region. You added the China (Beijing) acceleration region, an endpoint group in the China (Beijing) region, and an endpoint group in the China (Shanghai) region in the GA console. By default, GA forwards all requests from clients in the China (Beijing) region to the servers in the endpoint group that is deployed in the China (Beijing) region. A large number of requests are sent from clients in the China (Beijing) region. This causes the servers in the endpoint group that is deployed in the China (Beijing) region to become overloaded. Network latency and packet loss occur when clients access the application.

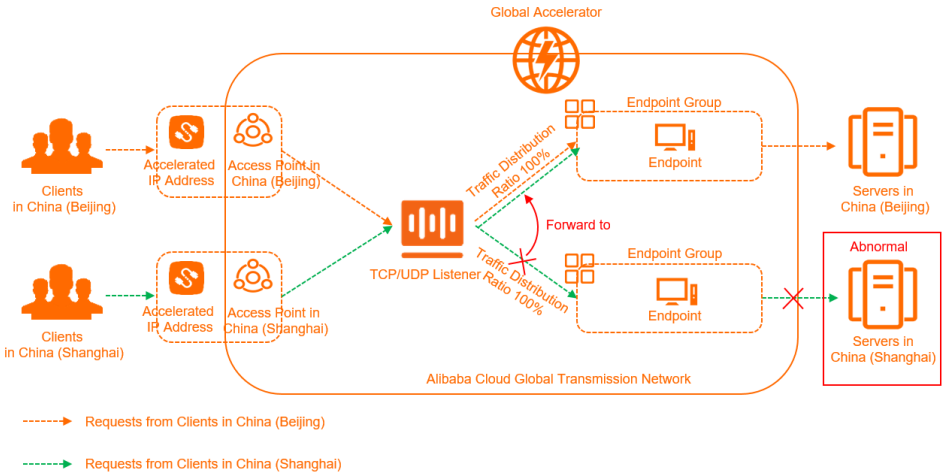


You can change the traffic distribution ratios for the endpoint groups in the China (Beijing) region and the China (Shanghai) region. For example, you can change the traffic distribution ratio for the endpoint group in the China (Beijing) region from 100% to 50%. This way, 50% of requests from clients in the China (Beijing) region are processed by the servers in the endpoint group in the China (Beijing) region. The remaining 50% of client requests are processed by the servers in the endpoint group in the China (Shanghai) region. This way, you can properly allocate client requests in the China (Beijing) region and reduce the loads on the servers in the endpoint group that is deployed in the China (Beijing) region. For more information about how to change the traffic distribution ratios for endpoint groups, see Set the traffic distribution ratio for an endpoint group.

Cross-region disaster recovery for applications

You can add multiple endpoint groups that are deployed in different regions for a GA instance and enable health checks for the endpoint groups. This achieves cross-region disaster recovery for applications.

The scenario in the following figure is used as an example. An application is deployed on servers in the China (Beijing) region and the China (Shanghai) region. You added China (Beijing) and China (Shanghai) as acceleration regions and an endpoint group to each acceleration region in the GA console. In most cases, requests from clients in the China (Beijing) region and the China (Shanghai) region are forwarded to the nearest acceleration region. The listener then checks the client requests and forwards the client requests to the corresponding endpoint group based on the traffic distribution ratio and priority. To ensure that the application can provide continuous and stable services, you must make sure that client requests can be forwarded to a healthy acceleration region if errors occur on the application in one of the acceleration regions.

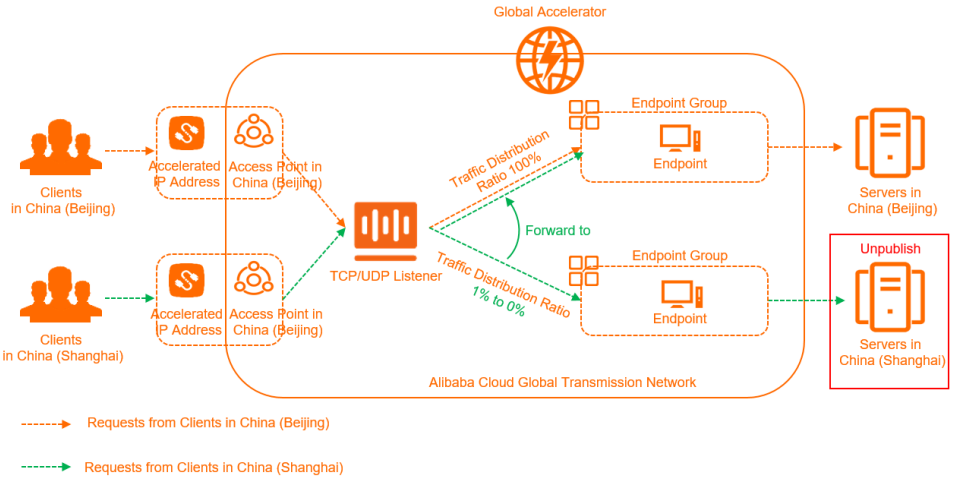


You can enable health checks for endpoint groups in the China (Beijing) region and the China (Shanghai) regions. If the endpoint group in the China (Shanghai) region fails the health check, the listener automatically forwards client requests to the healthy endpoint group in the China (Beijing) region. If the endpoint group in the China (Shanghai) region passes the health check, the listener automatically forwards requests from clients in the China (Shanghai) region to the endpoint group in the China (Shanghai) region. For more information about how to configure health checks, see Enable and manage health checks.

Unpublish or update a service based on regions

You can use the traffic distribution feature to unpublish or update a service based on regions. This reduces the impact on clients.

The scenario in the following figure is used as an example. A service is deployed on servers in the China (Beijing) region and the China (Shanghai) region. You added China (Beijing) and China (Shanghai) as acceleration regions and an endpoint group to each acceleration region in the GA console. You want to unpublish the service that is deployed in the China (Shanghai) region because a small number of client requests are sent to the service. When you unpublish the service, you must make sure that clients in the China (Shanghai) region can access the service as normal.



You can set the traffic distribution ratio to a lower value, such as 1%, for the endpoint group in the China (Shanghai) region and distribute 99% of client requests to the endpoint group in the China (Beijing) region. After the client requests that are sent to the service in the China (Shanghai) region are less than you expected, you can set the traffic distribution ratio to 0% for the endpoint group in the China (Shanghai) region. This way, you can unpublish the service that is deployed in the China (Shanghai) region.

If you want to update the service that is deployed in the China (Shanghai) region, you can change the traffic distribution ratio based on the preceding information when you unpublish the service. After you set the traffic distribution ratio to 0%, requests from clients in the China (Shanghai) region are forwarded to the endpoint group in the China (Beijing) region. After you update the service, set the traffic distribution ratio to 100% for the endpoint group in the China (Shanghai) region. This way, all requests from clients in the China (Shanghai) region are forwarded to the endpoint group in the China (Shanghai) region.

5.3. Create and manage endpoint groups

To associate a listener with an endpoint group, you can specify the region to which you want to distribute network traffic. Then, the system distributes network traffic to the optimal endpoint in the endpoint group.

Prerequisites

A Global Accelerator (GA) instance is created. For more information, see Create and manage GA instances.

Context

Each endpoint group is associated with a specific region. You can associate an endpoint group with a listener by specifying the region to which you want to distribute network traffic. After you associate an endpoint group with a listener, the system distributes network traffic to the optimal endpoints in the endpoint group.

Listeners that use different protocols support different types of endpoint groups:

- TCP or UDP listeners

  By default, you can create two default endpoint groups for each TCP or UDP listener. If you want to create more default endpoint groups, go to the Quota Management page and increase the quota of **gaplus_quota_epgs_per_listener**. For more information, see Manage quotas.

  You must deploy default endpoint groups in different regions. You can set a traffic distribution ratio for each default endpoint group. The traffic distribution ratio specifies the proportion of traffic that is distributed to a default endpoint group.

- HTTP or HTTPS listeners

  By default, you can create one default endpoint group and one virtual endpoint group for each HTTP or HTTPS listener. If you want to create multiple virtual endpoint groups, go to the Quota Management page and increase the quota of **gaplus_quota_vepg_per_listener**. For more information, see Manage quotas.

  - A default endpoint group refers to the endpoint group that you configure when you create an HTTP or HTTPS listener.
  - A virtual endpoint group refers to the endpoint group that you can create on the **Endpoint Group** page after you create a listener.

    After you create a virtual endpoint group for an HTTP or HTTPS listener, you can create a forwarding rule and associate the forwarding rule with the virtual endpoint group. Then, the HTTP or HTTPS listener forwards requests with different destination domain names or paths to the default or virtual endpoint group based on the forwarding rule. This way, you can use one GA instance to accelerate multiple domain names or paths. For more information about how to create a forwarding rule, see Create and manage forwarding rules.

Create a default endpoint group

1.
2. On the **Instances** page, find the GA instance that you want to manage and click **Configure Listeners** in the **Actions** column.
3. On the **Listener** tab, click **Add Listener**.

   > **Note** If this is your first time you create an endpoint group, skip this step.

4. On the **Configure Listener & Protocol** wizard page, set the required parameters, and click **Next**.

   If you want to create an endpoint group for an HTTPS listener, you must also configure SSL certificates. For more information, see Add and manage listeners.

5. On the **Configure Endpoint Group** wizard page, set the following parameters.

Parameter	Description

Parameter	Description
Endpoint Group Name	Enter a name for the endpoint group. The name must be 2 to 128 characters in length, and can contain letters, digits, underscores (_), and hyphens (-). The name must start with a letter.
Region	Select the region where you want to deploy the endpoint group.
Traffic Distribution Ratio	Set the traffic distribution ratio for the endpoint group. Unit: %. Valid values: 0 to 100. **Note** You can set **Traffic Distribution Ratio** only when you create an endpoint group for a TCP or UDP listener.
Backend Service	Specify whether backend servers are deployed on Alibaba Cloud. ○ **Alibaba Cloud**: Backend servers are deployed on Alibaba Cloud. ○ **Off Alibaba Cloud**: Backend servers are not deployed on Alibaba Cloud.
Preserve Client IP	Specify whether to preserve client IP addresses. After you enable this feature, backend servers can retrieve client IP addresses. For more information, see Preserve client IP addresses.

Parameter	Description
Endpoint	Endpoints are destinations of client requests. To add an endpoint, specify the following parameters: ◦ **Backend Service Type**: If your backend service is deployed on Alibaba Cloud, you can select **Alibaba Cloud Public IP Address**, **ECS**, **CLB**, **ALB**, or **OSS**. If your backend service is not deployed on Alibaba Cloud, you can select **Custom IP Address** or **Custom Domain Name**. **Note** ■ You can specify ECS, CLB, and ALB instances as endpoints only if your Alibaba Cloud account is included in the whitelist. If you want to specify ECS, CLB, or ALB instances as endpoints for your GA instances, submit a ticket to upgrade the GA instances. ■ The IP addresses of endpoint groups associated with each GA instance must be globally unique and not conflict with those of other GA instances. ■ If no service-linked role exists when you specify ECS instances, CLB instances, ALB instances, or OSS buckets as endpoints, the system automatically creates the corresponding service-linked role. For more information, see AliyunServiceRoleForGaVpcEndpoint, AliyunServiceRoleForGaAlb, and AliyunServiceRoleForGaOss. ◦ **Backend Service**: Enter the IP address, domain name, or instance ID of the backend server. ◦ **Weight**: Set a weight for the endpoint. Valid values: 0 to 255. GA distributes network traffic to endpoints based on their weights. **Notice** If the weight of an endpoint is set to 0, GA stops distributing network traffic to the endpoint. Proceed with caution. You can click **+ Add Endpoint** to add more endpoints. You can create at most four endpoints in each endpoint group. If you want to add more endpoints, go to the Quota Management page and increase the quota. For more information, see Manage quotas.

Parameter	Description
Backend Service Protocol	Select the protocol that is used by the backend service. Valid values: ○ **HTTP** (default) ○ **HTTPS** **Note** ○ If the listener protocol is HTTP, this parameter is set to HTTP by default and cannot be modified. ○ You can set **Backend Service Protocol** only when you configure an endpoint group for an HTTP or HTTPS listener.
Port Mapping	If the listener port and the port that the endpoint uses to provide services are not the same, you must add a mapping between the ports. ○ **Listener Port**: Enter the listener port. ○ **Endpoint Port**: Enter the port that the endpoint uses to provide services. If the listener port and the port that the endpoint uses to provide services are the same, you do not need to add the port mapping. GA automatically distributes client requests to the listener port of the endpoint. **Note** You can set **Port Mapping** only when you configure an endpoint group for an HTTP or HTTPS listener.
Health Check	Specify whether to enable or disable the health check feature. After you enable this feature, you can use health checks to check the status of endpoints. For more information about the health check feature, see Enable and manage health checks. **Note** If your GA instance uses UDP listeners, you can enable the health check feature for an endpoint only if the endpoint is associated with a TCP, HTTP, or HTTPS service. Otherwise, the endpoint is marked as unhealthy.
Health Check Protocol	Select the protocol that you want to use for health checks. Valid values: TCP, HTTP, and HTTPS. ○ A TCP health check probes whether a server port is healthy at the network layer by sending SYN packets to the port. ○ An HTTP health check probes whether an endpoint is healthy by simulating HTTP GET requests sent from a browser.
Port	Set the port of the endpoint to which probe packets are sent for health checks. Valid values: 1 to 65535.

Parameter	Description
Health Check Interval	Set the interval between two consecutive health checks. Unit: seconds. Valid values: 1 to 50. Default value: 2.
URI	Specify the URI for health checks. The URI must be 1 to 80 characters in length and start with a forward slash (/). The URI can contain letters, digits, hyphens (-), forward slashes (/), periods (.), percent signs (%), question marks (?), number signs (#), and ampersands (&). The URI can also contain the following extended characters: `_ ; ~ ! ( ) * [ ] @ $ ^ : ' , +` . By default, GA sends a GET request to the default homepage of the backend service. If you do not want to use the default homepage for health checks, you can manually specify a URI. **Note** This parameter is supported only for HTTP and HTTPS health checks.
Healthy Threshold	The number of consecutive health check failures that must occur before a healthy endpoint is considered unhealthy, or the number of consecutive health check successes that must occur before an unhealthy endpoint is considered healthy. Valid values: 2 to 10. Default value: 3.

6. (Optional)Click **+ Add Endpoint Group** to add multiple endpoint groups based on the preceding information.

   Note
   - You can add multiple endpoint groups only for TCP and UDP listeners.
   - By default, you can add two default endpoint groups for a TCP or UDP listener. If you want to add more endpoint groups, go to the Quota Management page and increase the quota of **gaplus_quota_epgs_per_listener**. For more information, see Manage quotas.

7. Click **Next**.
8. On the **Confirm** wizard page, check the configurations and click **Submit**.

   To modify a specific setting, click **Modify** in the corresponding section.

Create a virtual endpoint group

Before you create a virtual endpoint group, take note of the following limits:

- You can create a virtual endpoint group only for an HTTP or HTTPS listener.
- Before you can create a virtual endpoint group, you must create a default endpoint group.

1.
2. On the **Instances** page, find the GA instance that you want to manage and click **Configure Listeners** in the **Actions** column.
3. On the **Listeners** tab, click the endpoint group ID or number in the **Default Endpoint Group**

ID/Name column.

4. On the **Endpoint Group** tab, click **Add Virtual Endpoint Group** in the **Virtual Endpoint Group** section.
5. In the **Create Virtual Endpoint Group** dialog box, set the parameters and click **Create**.

   For more information, see Create a default endpoint group.

What to do next

Operation	Description
Modify an endpoint group	1. On the **Listeners** tab, find the listener that you want to manage and click the endpoint group ID or number in the **Default Endpoint Group ID/Name** column. 2. On the **Endpoint Group** tab, find the default endpoint group or virtual endpoint group that you want to modify and click **Modify** in the **Actions** column. 3. In the **Modify Default Endpoint Group** or **Modify Virtual Endpoint Group** dialog box, modify the name and endpoint configuration, and then click **Save**. For more information about the configurations of the default endpoint group, see Create a default endpoint group. **Note** You can configure and modify virtual endpoint groups only for HTTP and HTTPS listeners. For more information about virtual endpoint groups, see Overview.
Set the traffic distribution ratio for an endpoint group	You can set the proportion of traffic that is distributed to different endpoint groups. 1. On the **Listeners** tab, find the listener and click **Edit Endpoint Group** in the Actions column. 2. On the **Configure Endpoint Group** wizard page, find the endpoint group that you want to manage, set the traffic distribution ratio, and then click **Next**. Valid values of the traffic distribution ratio: 0 to 100. Unit: %. 3. Confirm the information of the endpoint group and click **Submit**. **Note** You can set traffic distribution ratios only for TCP and UDP listeners.

Operation	Description
Set the weight of an endpoint	You can set the weight of an endpoint. The weight specifies the proportion of traffic that GA distributes to an endpoint in the endpoint group. GA calculates the sum of all endpoint weights in an endpoint group. Then, traffic is forwarded to endpoints based on the proportions of their weights. For example, if you want to distribute 1/3 of the network traffic to Endpoint 1 and 2/3 of the network traffic to Endpoint 2, you can set the weight of Endpoint 1 to 1 and the weight of Endpoint 2 to 2. To disable GA from distributing network traffic to an endpoint, set the weight of the endpoint to 0. 1. On the **Listeners** tab, find the listener that you want to manage and click the endpoint group ID or number in the **Default Endpoint Group ID/Name** column. 2. On the **Endpoint Group** tab, find the endpoint group that contains the endpoint for which you want to set the weight and click **Modify** in the **Actions** column. 3. In the **Modify Default Endpoint Group** or **Modify Virtual Endpoint Group** dialog box, find and set the weight of the endpoint in the **Endpoint** section and click **Save**. Valid values of the weight: 0 to 255.
Delete an endpoint group	You can delete an endpoint group that you no longer need. After you delete an endpoint group, GA stops forwarding requests to the endpoint group. 1. On the **Listeners** tab, find the listener that you want to manage and click the endpoint group ID or number in the **Default Endpoint Group ID/Name** column. 2. On the **Endpoint Group** tab, find the default endpoint group or virtual endpoint group that you want to delete and click **Delete** in the **Actions** column. 3. In the message that appears, click **OK**. **Note** If a listener is associated with only one endpoint group and you delete the endpoint group, the listener becomes unavailable.
Delete an endpoint	You can delete an endpoint that you no longer need. After you delete an endpoint, GA stops forwarding requests to the endpoint. If an endpoint group contains only one endpoint, you cannot delete the endpoint. 1. On the **Listeners** tab, find the listener that you want to manage and click the endpoint group ID or number in the **Default Endpoint Group ID/Name** column. 2. On the **Endpoint Group** tab, find the default endpoint group or virtual endpoint group to which the endpoint that you want to delete belongs and click **Modify** in the **Actions** column. 3. In the **Modify Default Endpoint Group** or **Modify Virtual Endpoint Group** dialog box, find the endpoint in the **Endpoint** section, click **Delete** in the **Actions** column, and then click **Save**.

References

- CreateEndpointGroup: You can call this API operation to create an endpoint group.
- CreateEndpointGroups: You can call this API operation to create multiple endpoint groups.
- UpdateEndpointGroup: You can call this API operation to modify an endpoint group.

- DeleteEndpointGroup: You can call this API operation to delete an endpoint group.

5.4. Create and manage forwarding rules

Only HTTP and HTTPS listeners support domain name-based or path-based forwarding rules. After an HTTP or HTTPS listener receives a request, the listener forwards the request to a specific endpoint group if the destination domain name or path of the request matches a forwarding rule.

Prerequisites

- Only HTTP and HTTPS listeners support forwarding rules. Make sure that you have created an HTTP or HTTPS listener. For more information, see Add and manage listeners.
- A virtual endpoint group is created. For more information, see Create a virtual endpoint group.

Context

Forwarding rules are classified into default forwarding rules and custom forwarding rules:

- Default forwarding rules: After you create an HTTP or HTTPS listener, the system automatically creates a default forwarding rule and associates it with the default endpoint group. A listener contains only one default forwarding rule. You cannot modify or delete the default forwarding rule.
- Custom forwarding rules: After you create an HTTP or HTTPS listener, you can create custom forwarding rules based on your business requirements. You can create multiple custom forwarding rules for a listener.

Each forwarding rule consists of the following components:

- Forwarding conditions: A request is forwarded to the specified endpoint group only if the request matches a forwarding condition. You can configure a forwarding condition in the following ways:
  - Specify a domain name: You can specify a domain name as the forwarding condition in a forwarding rule. If a request matches the specified domain name, the request is forwarded to the specified endpoint group.
  - Specify paths: You can specify multiple paths as the forwarding condition in a forwarding rule. If a request matches one of the specified paths, the request is forwarded to the specified endpoint group.
  - Specify a domain name and multiple paths: If a request matches the specified domain name or one of the specified paths, the request is forwarded to the specified endpoint group.
- Forwarding action: forwards the request that matches the forwarding condition to a specific endpoint group. Each forwarding rule can point only to one endpoint group.

A listener can contain one default forwarding rule and multiple custom forwarding rules. The system attempts to match a request with a forwarding rule in the following ways:

- Method 1: If the request contains a domain name, the system attempts to match the request with a forwarding rule based on the domain name.

- If the domain name matches a forwarding rule, the system attempts to match the path of the request with the forwarding rule.

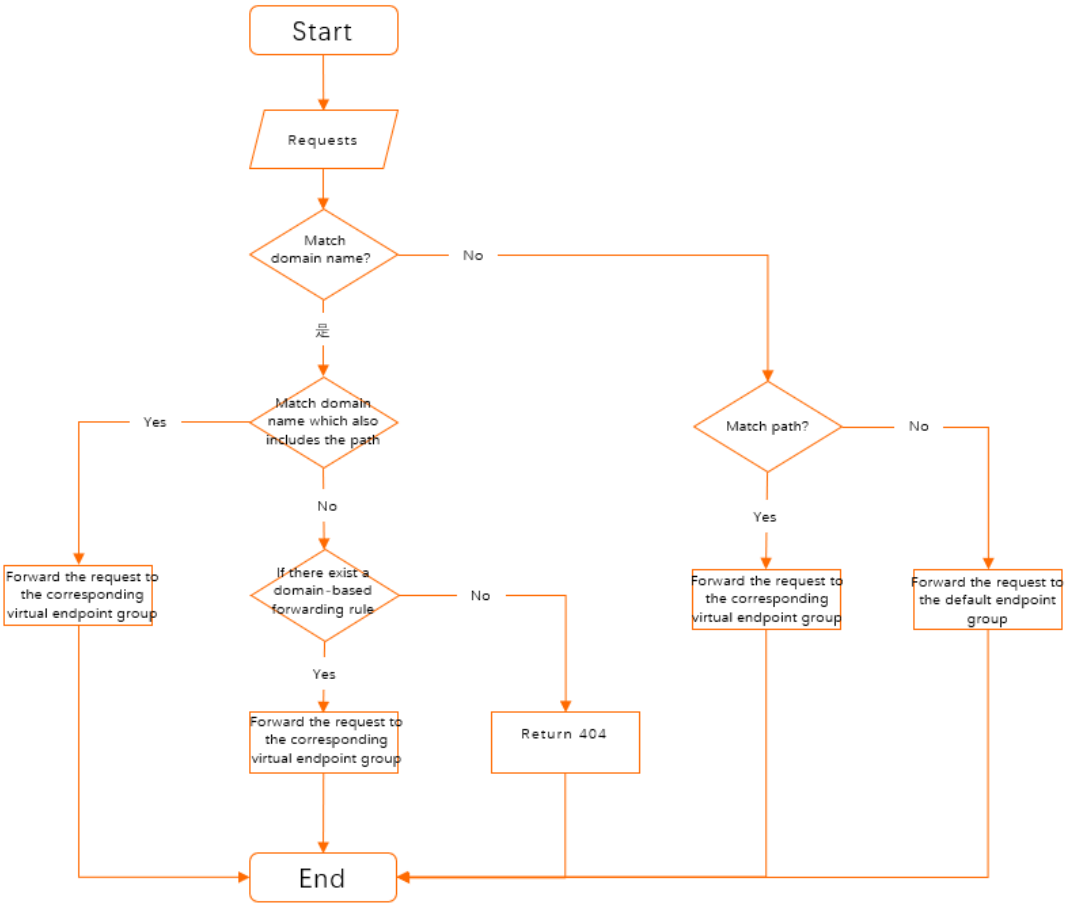
  If the path also matches the forwarding rule, the request is forwarded to the specified endpoint group. If the path does not match the forwarding rule, the request is forwarded based on a domain name-based forwarding rule. The domain name of the request is specified as the forwarding condition of the domain name-based forwarding rule and no path is specified.

  If such a domain name-based forwarding rule is not configured for the listener, an HTTP 404 status code is returned to the client.

- If the domain name of the request does not match a forwarding rule, the request is forwarded by using Method 2.

- Method 2: If a request does not contain a domain name or the listener does not contain a forwarding rule that matches the domain name, the system attempts to match the request with a path-based forwarding rule. Only paths are specified as the forwarding condition of the path-based forwarding rule and no domain name is specified.

If the system matches a request by using one of the preceding methods, the request is forwarded to the specified endpoint group. If no forwarding rule matches the request, the request is matched with the default forwarding rule and forwarded to the default endpoint group.



Create a forwarding rule

After you create an HTTP or HTTPS listener, the system automatically creates a default forwarding rule and associates it with the default endpoint group. You can perform the following steps to create a custom forwarding rule and forward requests that match the custom forwarding rule to the specified virtual endpoint group.

1.
2. On the **Instances** page, find the Global Accelerator (GA) instance that you want to manage and click **Configure Listeners** in the **Actions** column.
3. On the **Listeners** tab, find the listener that you want to manage and click the ID of the listener.
4. On the listener details page, click the **Forwarding Rule** tab.
5. On the **Forwarding Rule** tab, click **Add Forwarding Rule**, configure the following parameters, and then click **OK**.

Parameter	Description
If (Matching All Conditions)	Configure the forwarding condition. ○ **Domain Name** The domain name must be 3 to 128 characters in length and can contain letters, digits, hyphens (-), and periods (.). Supported wildcard characters are asterisks (*) and question marks (?). ○ **Path** The path must be 1 to 128 characters in length and must start with a forward slash (/). The path can contain letters, digits, dollar signs ($), hyphens (-), underscores (_), periods (.), plus signs (+), forward slashes (/), ampersands (&), tildes (~), at signs (@), colons (:), and apostrophes ('). Supported wildcard characters are asterisks (*) and question marks (?).
Forward to Virtual Endpoint Group	Select the virtual endpoint group to which a matched request is forwarded.

Modify a forwarding rule

1.
2. On the **Instances** page, find the Global Accelerator (GA) instance that you want to manage and click **Configure Listeners** in the **Actions** column.
3. On the **Listeners** tab, find the listener that you want to manage and click the ID of the listener.
4. On the listener details page, click the **Forwarding Rule** tab.
5. On the **Forwarding Rule** tab, find the forwarding rule that you want to modify, click  in the upper-right corner, modify the forwarding rule, and then click **Save**.

Delete a forwarding rule

1.
2. On the **Instances** page, find the Global Accelerator (GA) instance that you want to manage and click **Configure Listeners** in the **Actions** column.
3. On the **Listeners** tab, find the listener that you want to manage and click the ID of the listener.

4. On the listener details page, click the **Forwarding Rule** tab.
5. On the **Forwarding Rule** tab, find the forwarding rule that you want to delete and click in the upper-right corner.
6. In the message that appears, confirm the ID of the forwarding rule and click **OK**.

5.5. Enable and manage health checks

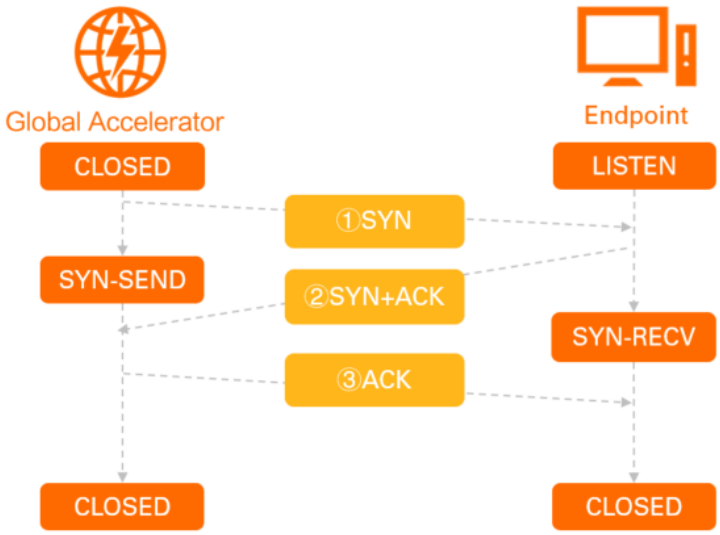
Global Accelerator (GA) performs health checks to test the status of endpoints. Health checks improve service reliability and availability and prevent service interruptions caused by unhealthy endpoints.

Introduction to health checks

You can enable health checks for endpoint groups of a GA instance. After you enable health checks, GA periodically checks whether the endpoints are healthy. When GA detects an unhealthy endpoint, GA distributes new requests to other healthy endpoints. When the unhealthy endpoint recovers, GA distributes requests to the endpoint again.

GA supports health checks that use the following protocols: TCP, HTTP, and HTTPS.

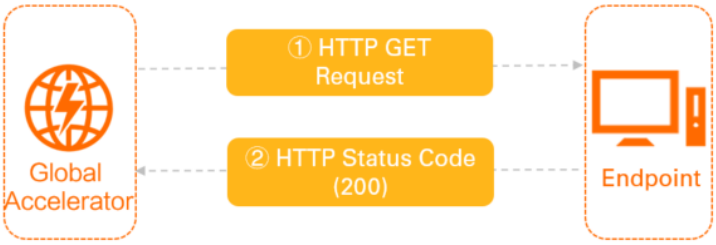
A TCP health check probes whether a server port is healthy at the network layer by sending SYN packets to the port. The following figure shows the process of TCP health checks.



No.	Description
1	A GA instance sends a TCP SYN packet to the IP address and port of an endpoint based on the health check configurations of the TCP listener.

No.	Description
2	The GA instance verifies the health status of the endpoint based on whether the endpoint can return an SYN-ACK packet within the specified timeout period. • If the GA instance receives an SYN-ACK packet from the endpoint within the specified timeout period (3 seconds), the endpoint is considered healthy. • If the GA instance receives an RST packet from the endpoint within the specified timeout period (3 seconds), the endpoint is considered unhealthy. • If the GA instance does not receive an SYN-ACK packet from the endpoint within the specified timeout period (3 seconds), the GA instance considers that the endpoint cannot be reached or respond. As a result, the endpoint is considered unhealthy. **Note** The response timeout period specifies the maximum amount of time to wait for a health check response. If an endpoint does not respond within the specified timeout period, the endpoint fails to pass the health check. By default, the timeout period is set to 3 seconds and cannot be changed.
3	After the GA instance receives an SYN-ACK packet from the endpoint, the GA instance sends an ACK packet to establish a TCP session.

An HTTP health check probes whether an endpoint is healthy by simulating HTTP GET requests sent from a browser. The following figure shows the process of HTTP health checks.



No.	Description
1	A GA instance sends an HTTP GET request to an endpoint based on the health check configurations of the listener. The HTTP GET request is sent to an address in the following format: the IP address of the endpoint + health check port + health check path.

No.	Description
2	After the endpoint receives the request, the endpoint checks the status of the service and returns a relevant HTTP status code. • If the GA instance receives the `200` status code from the endpoint within the specified timeout period (3 seconds), the endpoint is considered healthy. • If the GA instance receives a status code other than the `200` status code from the endpoint within the specified timeout period (3 seconds), the endpoint is considered unhealthy. • If the GA instance does not receive a status code from the endpoint within the specified timeout period (3 seconds), the GA instance considers that the endpoint cannot be reached or respond. As a result, the endpoint is considered unhealthy. **Note** The response timeout period specifies the maximum amount of time to wait for a health check response. If an endpoint does not respond within the specified timeout period, the endpoint fails to pass the health check. By default, the timeout period is set to 3 seconds and cannot be changed.

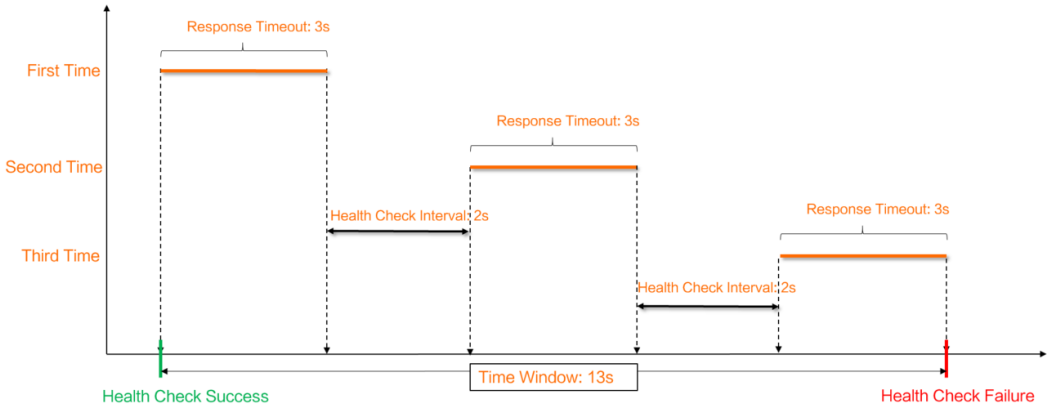
Health checks improve the availability of your services. However, frequent failovers caused by unhealthy endpoints may affect system availability. Health check time windows are introduced to control failovers. A failover is performed only if an endpoint consecutively passes or fails a specific number of health checks within a time window. The health check time window is determined by the following factors:

- Health check interval: the interval at which health checks are performed.
- Response timeout: the amount of time to wait for a response.
- Healthy threshold: the number of consecutive successes or failures of health checks.

The health check time window is calculated based on the following formula:

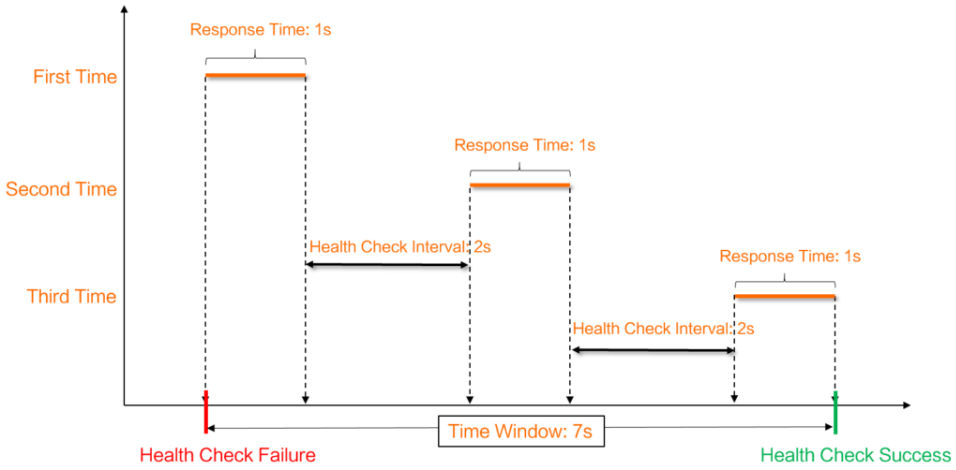
- Time window for health check failures = Response timeout × Healthy threshold + Health check interval × (Healthy threshold - 1)

  The following figure shows an example in which the response timeout is 3 seconds, the health check interval is 2 seconds, and the healthy threshold is 3 times. Therefore, the time window for health check failures is 13 seconds based on the formula 3 × 3 + 2 × (3 - 1).



- Time window for health check successes = (Response time of a successful health check × Healthy threshold) + Heath check interval × (Healthy threshold - 1)

The following figure shows an example in which the response time is 1 second, the health check interval is 2 seconds, and the healthy threshold is 3 times. Therefore, the time window for health check successes is 7 seconds based on the formula 1 × 3 + 2 × (3 - 1).



If your GA instance uses UDP listeners, you can enable health checks for an endpoint only if the endpoint is associated with a TCP, HTTP, or HTTPS service. Otherwise, the endpoint is marked as abnormal.

TCP health checks

HTTP and HTTPS health checks

Health check time window

Limits

Enable health checks

1.
2. On the **Instances** page, find the GA instance that you want to manage and click **Configure Listeners** in the **Actions** column.
3. On the **Listeners** tab, find the listener that you want to manage and click **Modify** in the **Actions** column.
4. On the **Edit Listener** page, click **Next**.
5. In the **Health Check** section of the **Configure Endpoint Group** wizard page, enable the **health check** feature and set the following parameters.

Parameter	Description
Health Check Protocol	Select the protocol that you want to use for health checks. Valid values: TCP, HTTP, and HTTPS. ○ A TCP health check probes whether a server port is healthy at the network layer by sending SYN packets to the port. ○ An HTTP health check probes whether an endpoint is healthy by simulating HTTP GET requests sent from a browser.

Parameter	Description
Port	Set the port of the endpoint to which probe packets are sent for health checks. Valid values: 1 to 65535.
Health Check Interval	Set the interval between two consecutive health checks. Unit: seconds. Valid values: 1 to 50. Default value: 2.
URI	Specify the URI for health checks. The URI must be 1 to 80 characters in length and start with a forward slash (/). The URI can contain letters, digits, hyphens (-), forward slashes (/), periods (.), percent signs (%), question marks (?), number signs (#), and ampersands (&). The URI can also contain the following extended characters: `_ ; ~ ! ( ) * [ ] @ $ ^ : ' , +` . By default, GA sends a GET request to the default homepage of the backend service. If you do not want to use the default homepage for health checks, you can manually specify a URI. **Note** This parameter is supported only for HTTP and HTTPS health checks.
Healthy Threshold	The number of consecutive health check failures that must occur before a healthy endpoint is considered unhealthy, or the number of consecutive health check successes that must occur before an unhealthy endpoint is considered healthy. Valid values: 2 to 10. Default value: 3.

6. Click **Next**. On the **Confirm** wizard page, confirm the health check configurations and click **Submit**.

What to do next

Operation	Description
Modify health check configurations	1. On the **Listeners** tab, find the listener and click **Edit Endpoint Group** in the **Actions** column. 2. In the **Health Check** section of the **Configure Endpoint Group** wizard page, modify the health check protocol, port, and health check interval and click **Next**. For more information, see Enable health checks. 3. On the **Confirm** wizard page, click **Next**.
Disable health checks	1. On the **Listeners** tab, find the listener and click **Edit Endpoint Group** in the **Actions** column. 2. In the **Health Check** section of the **Configure Endpoint Group** wizard page, disable the **health check** feature and click **Next**. 3. On the **Confirm** wizard page, click **Next**.

Related topics

- CreateEndpointGroup: Creates an endpoint group. You can configure health checks when you create an endpoint group.
- UpdateEndpointGroup: Modifies an endpoint group. You can configure health checks when you modify an endpoint group.
- GetHealthStatus: Queries health check information about an endpoint.

5.6. Examples on how to configure the traffic distribution feature for multiple endpoint groups

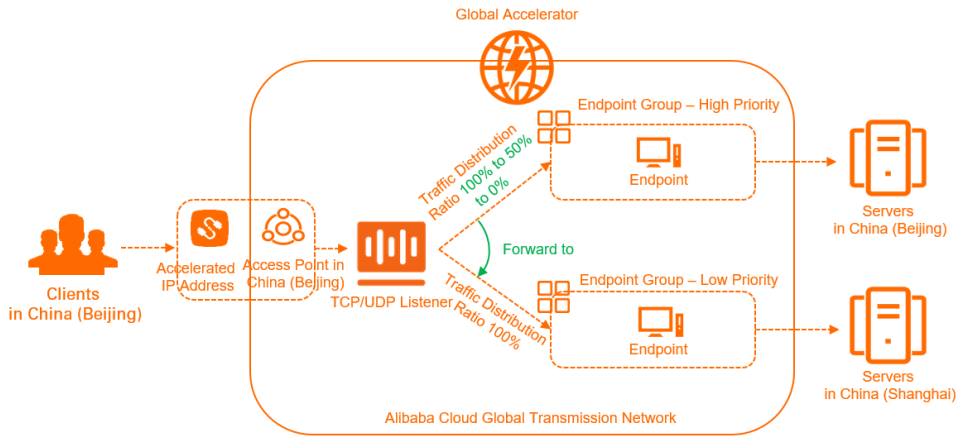
Distribute traffic for an application that is deployed across regions

This topic describes how to use the traffic distribution feature to control the percentage of client requests that are forwarded to endpoint groups in different regions.

Scenarios

A company deploys a service on servers in the China (Beijing) and China (Shanghai) regions. The TCP protocol is used and port 80 is open. The clients are located in the China (Beijing) region. The company specifies China (Beijing) as the acceleration region and creates an endpoint group in the China (Beijing) and China (Shanghai) regions in the Global Accelerator (GA) console. By default, GA forwards all requests from clients in the China (Beijing) region to the servers in the endpoint group that is deployed in the China (Beijing) region. The endpoint group in the China (Shanghai) region serves as the secondary endpoint group. If the endpoint group in the China (Beijing) region is abnormal, client requests are forwarded to the endpoint group in the China (Shanghai) region. Due to business development, the company wants to forward requests from clients in the China (Beijing) region to the servers in the endpoint group that is deployed in the China (Shanghai) region. The company also wants to ensure that clients can access the service as normal during the switchover process.

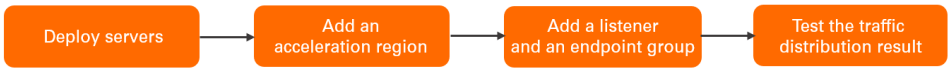
You can change the traffic distribution ratio for the endpoint group in the China (Beijing) region. For example, you can change the traffic distribution ratio from 100% to 50%. This way, 50% of requests from clients in the China (Beijing) region are forwarded to the servers in the endpoint group in the China (Shanghai) region. If clients can access the service as normal, change the traffic distribution ratio to 0%. This way, all requests from clients in the China (Beijing) region are forwarded to the servers in the endpoint group in the China (Shanghai) region. This ensures the seamless switchover of traffic from clients in the China (Beijing) region.



Prerequisites

A GA instance and a basic bandwidth plan are purchased. For more information, see Select and purchase GA resources.

Procedure



Step 1: Deploy servers

The servers in this example run the Alibaba Cloud Linux 3.2104 64-bit operating system. The command that is used to run the test may vary based on the operating system. For more information, refer to the user guide of the operating system.

1. Deploy servers in the China (Beijing) and China (Shanghai) regions, and specify the TCP protocol and open port 80 for the servers.
2. Open the command prompt on a client in the China (Beijing) region and run the **curl** command to access the servers in the China (Beijing) region and the China (Shanghai) region.

   ```
   curl <Origin server IP address>
   ```

   The following figures show the region information that is returned.

   Access a server in the China (Beijing) region

   Access a server in the China (Shanghai) region

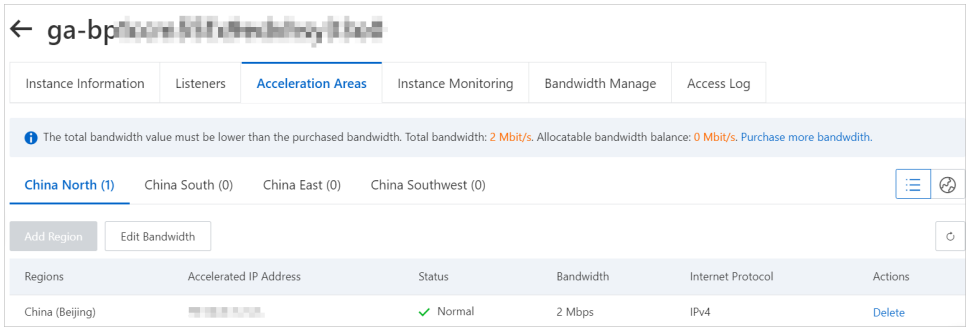
Step 2: Add an acceleration region

1.
2. On the **Instances** page, find the GA instance that you created and click its ID.
3. Click the **Acceleration Areas** tab and then click **Add Region** on the **China North** tab.
4. In the **Add Acceleration Area** dialog box, set the following parameters and click **OK**.

Parameter	Description
Region	Select the region where the users that require the acceleration service are located. In this example, **China (Beijing)** is selected.
Bandwidth	Allocate bandwidth to the region. In this example, *2* Mbit/s of bandwidth is allocated.

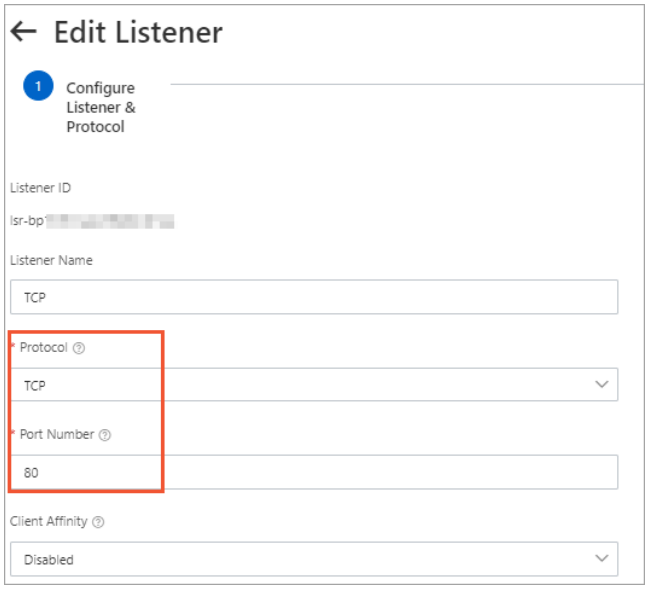
Parameter	Description
Internet Protocol	Select the Internet protocol that is used by the users to connect to GA. In this example, **IPv4** is selected.

After you add the region, the system assigns an accelerated IP address to the region that is added to the GA instance. This accelerated IP address is used to accelerate data transfer from users in the specified region to the specified backend servers through GA.



Step 3: Add a listener and an endpoint group

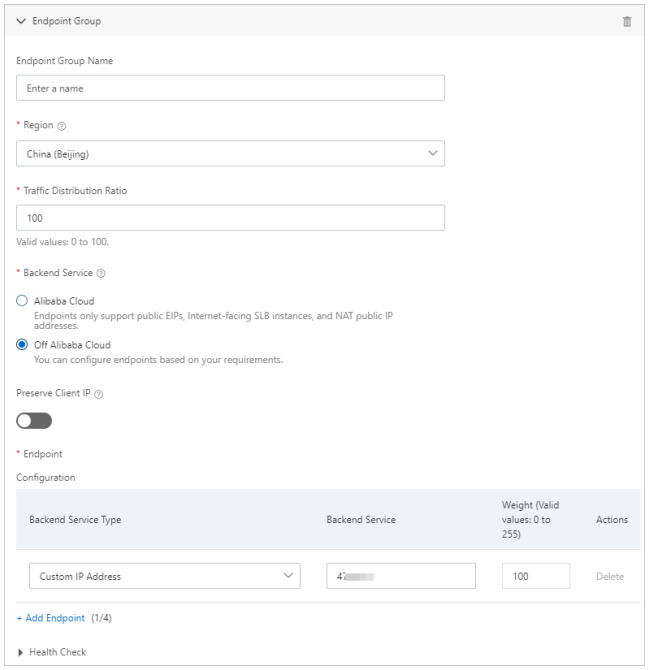
1.
2. On the **Configure Listener & Protocol** wizard page, specify the following listener information and click **Next**.



Parameter	Description
Listener Name	Enter a name for the listener. The name must be 2 to 128 characters in length, and can contain letters, digits, underscores (_), and hyphens (-). The name must start with a letter.

Parameter	Description
Protocol	Select the protocol of the listener. In this example, **TCP** is selected.
Port Number	Specify a listener port. The port is used to receive and forward requests to endpoints. Valid values: **1 to 65499**. In this example, the value is set to *80*.
Client Affinity	Specify whether to enable client affinity. If client affinity is enabled, requests from the same client are forwarded to the same endpoint when the client connects to a stateful application. In this example, **Disable** is selected.

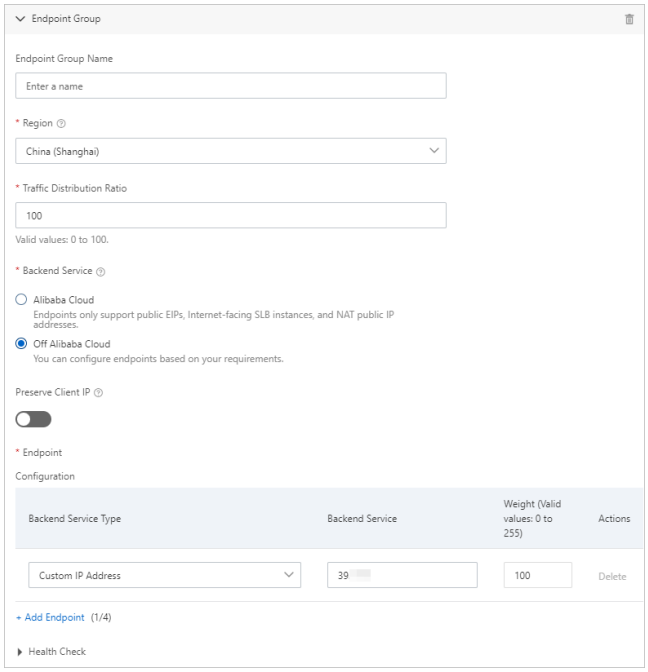
3. On the **Configure Endpoint Group** wizard page, set the following parameters for the endpoint group that is deployed in the China (Beijing) region.



Parameter	Description
Endpoint Group Name	Enter a name for the endpoint group.
Region	Select the region where you want to create the endpoint group. The server that the clients want to access must be deployed in the specified region. In this example, **China (Beijing)** is selected.

Parameter	Description
Traffic Distribution Ratio	Set the traffic distribution ratio for the endpoint group. Unit: %. Valid values: 0 to 100. In this example, the default value **100** is used. **Note** You can set **Traffic Distribution Ratio** only if you create an endpoint group for a TCP or UDP listener.
Backend Service	Specify whether the backend service is deployed on Alibaba Cloud. In this example, **Off Alibaba Cloud** is selected.
Preserve Client IP	Specify whether to preserve client IP addresses. After you enable this feature, backend servers can retrieve client IP addresses. In this example, client IP address preservation is disabled.
Endpoint	Endpoints are destinations of client requests. To add an endpoint, specify the following parameters: ○ **Backend Service Type**: In this example, **Custom IP Address** is selected. ○ **Backend Service**: Enter the public IP address of the backend server. ○ **Weight**: Enter the weight of the endpoint. Valid values: 0 to 255. GA distributes network traffic to endpoints based on their weights. **Notice** If the weight of an endpoint is set to 0, GA stops distributing network traffic to the endpoint. Proceed with caution.
Health Check	Specify whether to enable or disable the health check feature. After you enable this feature, you can use health checks to check the status of endpoints. For more information about how to configure health checks, see Enable and manage health checks. In this example, the health check feature is enabled.

4. Click **+ Add Endpoint Group** to add another endpoint group in the China (Shanghai) region, configure the endpoint group based on the parameter description in Substep , and then click **Next**.



5.

Step 4: Test the traffic distribution result

In this example, the following command is used to simulate client requests to test the traffic distribution result.

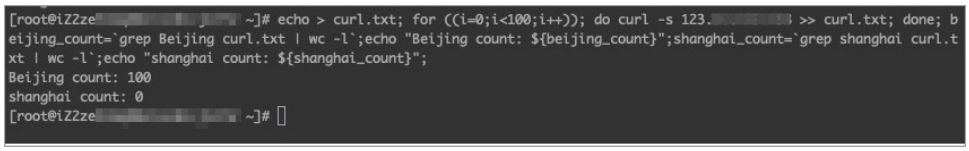
```
echo > curl.txt; for ((i=0;i<<Number of requests>;i++)); do curl -s <Accelerated IP address > >> curl.txt; done; beijing_count=`grep Beijing curl.txt | wc -l`;echo "Beijing count: ${beijing_count}";shanghai_count=`grep Shanghai curl.txt | wc -l`;echo "shanghai count: ${shanghai_count}";
```

Parameter description:

- `Number of requests`: The number of client requests that are simulated. For example, if you set `Number of requests` to 100, 100 requests are sent from the client.
- `Accelerated IP address`: The accelerated IP address assigned by GA.
- `Beijing count`: The number of requests processed by the servers in the China (Beijing) region.
- `Shanghai count`: The number of requests processed by the servers in the China (Shanghai) region.

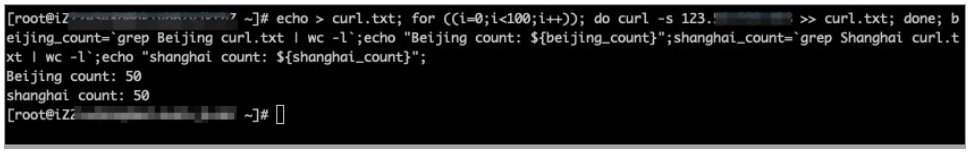
1. Check how client requests are scheduled when you set the traffic distribution ratio to 100% for the endpoint group that is assigned a higher priority in the China (Beijing) region.

   Open the command prompt on a client in the China (Beijing) region and send 100 requests. Then, check the number of requests that are processed by the servers in the China (Beijing) region and the number of requests that are processed by the servers in the China (Shanghai) region.

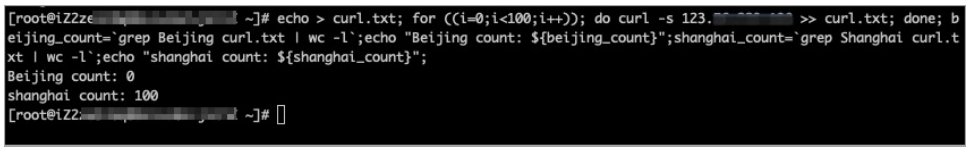
   The result indicates that all requests from the client in the China (Beijing) region are forwarded to the endpoint group in the China (Beijing) region.

2. Check how client requests are scheduled when you set the traffic distribution ratio to 50% for the endpoint group that is assigned a higher priority in the China (Beijing) region.
   i. Change the traffic distribution ratio to 50% for the endpoint group in the China (Beijing) region. For more information, see Set the traffic distribution ratio for an endpoint group.
   ii. Send 100 requests from a client in the China (Beijing) region and check the number of requests that are processed by the servers in the China (Beijing) region and the number of requests that are processed by the servers in the China (Shanghai) region.



   The result indicates that each endpoint group in the China (Beijing) region and the China (Shanghai) region processes 50 requests.

3. Check how client requests are scheduled when you set the traffic distribution ratio to 0% for the endpoint group that is assigned a higher priority in the China (Beijing) region.
   i. Change the traffic distribution ratio to 0% for the endpoint group in the China (Beijing) region. For more information, see Set the traffic distribution ratio for an endpoint group.
   ii. Send 100 requests from a client in the China (Beijing) region and check the number of requests that are processed by the servers in the China (Beijing) region and the number of requests that are processed by the servers in the China (Shanghai) region.



   The result indicates that all requests from the client in the China (Beijing) region are forwarded to and processed by the servers in the China (Shanghai) region.

6.Access control

This topic describes how to configure access control for a listener. You can configure different access control modes and access control lists (ACLs) for different listeners of a Global Accelerator (GA) instance.

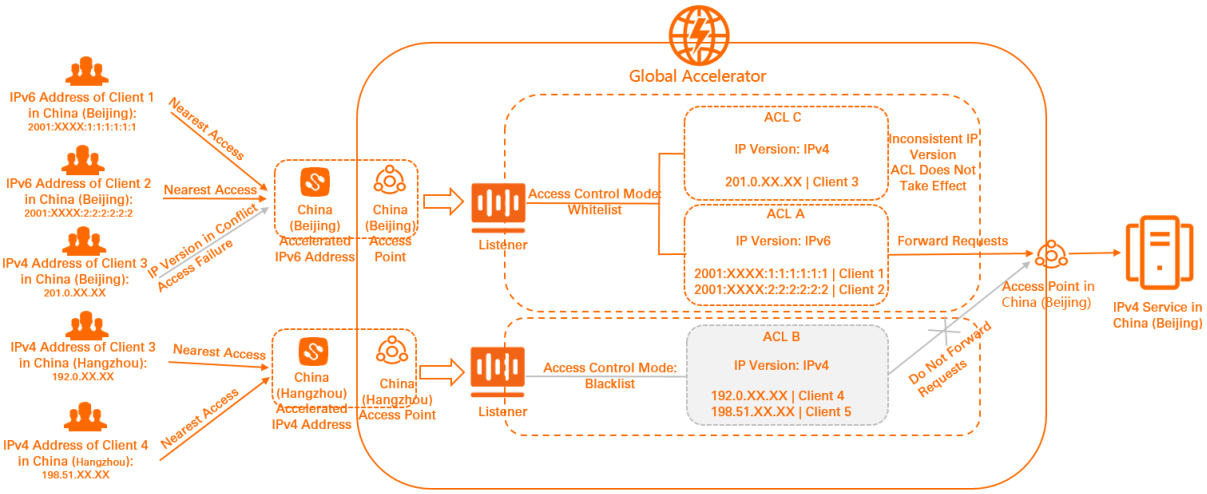
Introduction

The access control feature consists of access control modes and access control lists (ACLs). Access control modes include the whitelist mode and blacklist mode. An ACL can contain multiple IP addresses or CIDR blocks. You can set whitelists or blacklists for different listeners:

- Whitelist: Only the requests from the IP addresses or CIDR blocks in the specified ACL are forwarded. If you want to allow access from specific IP addresses, you can configure a whitelist.
- Blacklist: All requests from the IP addresses or CIDR blocks in the specified ACL are denied. If you want to block access from specific IP addresses, you can configure a blacklist.

> **Notice**
>
> - Risks may arise if the whitelist is improperly configured. After you configure a whitelist for a listener, only requests from the IP addresses that are added to the whitelist are forwarded by the listener. If the whitelist is enabled but no IP addresses are added to the ACL, the listener denies all requests.
> - If the blacklist is enabled but no IP addresses are added to the ACL, the listener forwards all requests.

When you create an ACL, you can select IPv4 or IPv6 as the supported IP version. When you configure access control for a listener, you can select an ACL that uses the same IP version as the accelerated IP address of the access point.



Limits

- The total number of IP addresses and CIDR blocks in the ACLs that are associated with a listener cannot exceed 200. Each IP address and CIDR block must be unique.
- An ACL can be associated with up to 10 listeners.
- A listener can be associated with at most two ACLs. If you associate two ACLs with a listener, one ACL

must be based on IPv4 and the other must be based on IPv6.

- If you associate an IPv4 ACL and an IPv6 ACL with a listener, only the ACL that matches the IP version of the accelerated IP address is applied.

Procedure

The following figure shows how to configure access control for a listener.

To configure an ACL for a listener, perform the following steps:

1. Create an ACL: Before you enable access control, you must create an ACL.
2. Add IP addresses or CIDR blocks to the ACL: You can add multiple IP addresses or CIDR blocks to the ACL.
3. Enable access control for a listener.: Enable access control for a listener. Then, set the access control mode and select an ACL.

Create an ACL

Before you enable access control for a listener, you must create an ACL.

1.
2. In the left-side navigation pane, click **Access Control**.
3. On the **Access Control** page, click **Create ACL**. In the Create ACL dialog box, set **ACL Name** and **IP Version**.

   Select **IPv4** or **IPv6** based on your business requirements.

   - If you select **IPv4**, the ACL is applied only in acceleration regions that use accelerated IPv4 addresses.
   - If you select **IPv6**, the ACL is applied only in acceleration regions that use accelerated IPv6 addresses.
4. Click **OK**.

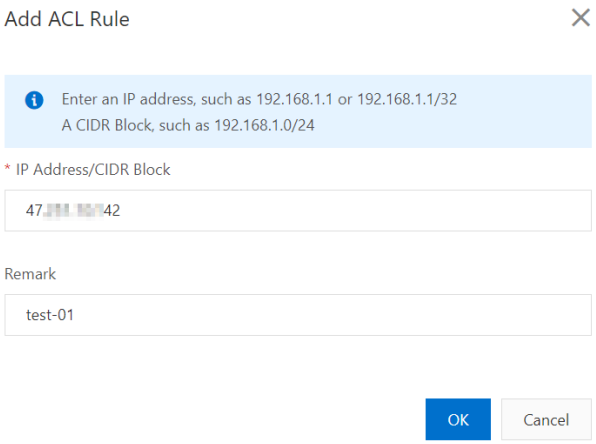
Add IP addresses or CIDR blocks to the ACL

After the ACL is created, you can add multiple IP addresses or CIDR blocks to the ACL. This way, you can enable a listener to allow or block access from the specified IP addresses or CIDR blocks.

1.
2.
3. Find the ACL that you want to manage and click **Manage ACL** in the **Actions** column.
4. Add IP addresses or CIDR blocks to the ACL.
   - Add one IP address or CIDR block to the ACL

     On the ACL Details page, click **Add Rule**. In the **Add ACL Rule** dialog box, enter an IP address or a CIDR block, enter remarks, and then click **OK**.

The remarks must be 2 to 256 characters in length, and can contain letters, digits, hyphens (-), forward slashes (/), periods (.), underscores (_), commas (,), semicolons (;), and at signs (@).

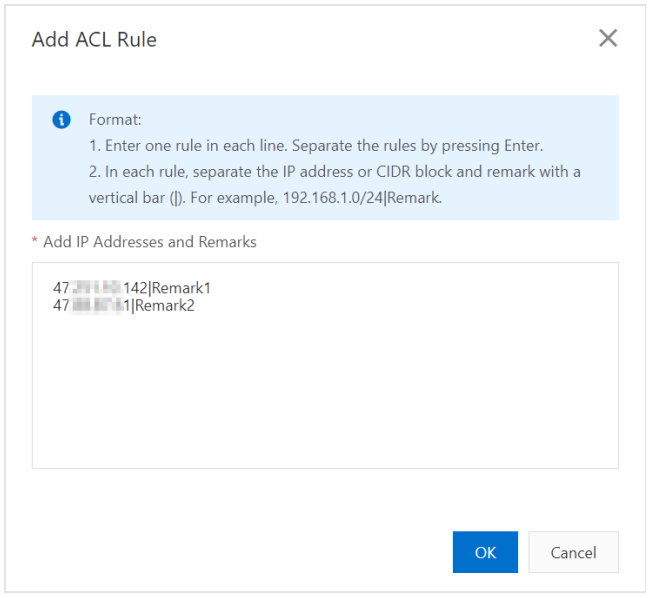


- Add multiple IP addresses or CIDR blocks at a time

  On the ACL Details page, click **Add Multiple Rules**. In the **Add ACL Rules** dialog box, enter multiple IP addresses or CIDR blocks, enter remarks, and then click **OK**.

  Take note of the following items:

  - Enter one entry per line. Press the Enter key to start a new line.
  - Separate an IP address or CIDR block and the remarks with a vertical bar (|). For example, 47.57.XX.XX|remarks.
  - The remarks must be 2 to 256 characters in length, and can contain letters, digits, hyphens (-), forward slashes (/), periods (.), underscores (_), commas (,), semicolons (;), and at signs (@).



Enable access control for a listener.

GA allows you to configure access control for a listener. You can configure whitelists or blacklists for different listeners.

Before you enable access control, make sure that a listener is created. For more information, see Add and manage listeners.

1.
2. On the **Instances** page, find the GA instance that you want to manage and click **Configure Listeners** in the **Actions** column.
3. On the **Listeners** tab, click the ID of the listener for which you want to enable access control.
4. On the **Listener Details** tab, turn on **Access Control**.
5. In the **Enable Access Control** dialog box, set the following parameters and click **OK**.

Parameter	Description
Access Control Mode	Select an access control mode. Valid values: ○ **Whitelist**: After you associate an ACL with the listener, the listener forwards only requests from IP addresses or CIDR blocks that are added to the ACL. ○ **Blacklist**: After you associate an ACL with the listener, the listener denies requests from IP addresses or CIDR blocks that are added to the ACL. **Notice** ○ Risks may arise if the whitelist is improperly configured. After you configure a whitelist for a listener, only requests from the IP addresses that are added to the whitelist are forwarded by the listener. If the whitelist is enabled but no IP addresses are added to the ACL, the listener denies all requests. ○ If the blacklist is enabled but no IP addresses are added to the ACL, the listener forwards all requests.
Select ACL	Select an ACL.

Remove IP addresses or CIDR blocks from the ACL

You can remove IP addresses or CIDR blocks from the ACL.

1.
2.
3. Find the ACL that you want to manage and click **Manage ACL** in the **Actions** column.
4. Find the IP address or CIDR block that you want to remove from the ACL and click **Delete** in the **Actions** column. To remove multiple IP addresses or CIDR blocks at a time, select the IP addresses or CIDR blocks that you want to remove and click **Delete** below the list.
5. In the message that appears, click **OK**.

Disable access control

If a listener no longer requires access control, you can disable access control for the listener.

1.
2. On the **Instances** page, find the GA instance that you want to manage and click **Configure Listeners** in the **Actions** column.

3. On the **Listeners** tab, click the ID of the listener for which you want to disable access control.
4. On the **Listener Details** tab, turn off **Access Control**.

7.Log management

Context

1.

7.1. Query operations logs

All operations that you perform on a Global Accelerator (GA) instance are recorded in operations logs. You can query and search log information based on the event time, users, and relevant resources. This allows you to keep track of a GA instance.

Procedure

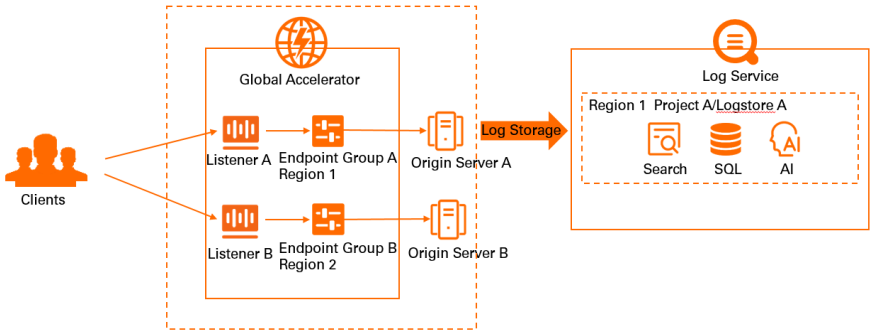
1.
2. In the left-side navigation pane, choose **Log Management > Operations Log** .
3. On the **Operations Log** page, select a query condition and click.
   - **Service Name**: By default, **Global Accelerator(Ga)** is selected.
   - Select an event type: You can select **Read/Write type**, **Username**, and **Resource Type**.
   - Select a time range: You can select a default or custom time range.
4. Find the operations log that you want to view and click + to view details.

7.2. Work with access logs

Global Accelerator (GA) can create access logs to record the traffic information of endpoints. You can analyze the traffic information to verify Access Control List (ACL) rules and troubleshoot network errors.

Introduction to access logs

You can configure GA to create access logs for one or more endpoint groups of a GA instance. The collected log data is delivered to the Logstores provided by Log Service in the regions where the endpoint groups are deployed. An access log contains the following information: the source IP address, source port, destination IP address, destination port, and acceleration region.



- Troubleshooting

  You can troubleshoot issues based on the information in an access log.

For example, you can check whether GA returns an expected response based on the **status** parameter in an access log and then locate the cause.

- Business planning

  You can analyze an access log to make informed business decisions.

  For example, you can upgrade bandwidth plans in advance to meet your business requirements based on the traffic trend in the acceleration region. You can also view the hosts that access your application within a specified time period and prepare for application upgrades based on the **http_host** parameter in the access log.

You are not charged additional fees for using the access log feature. You need only to pay for Log Service. For more information, see Billing of Log Service.

- The access log feature is supported only in regions where Log Service is available. For more information, see Supported regions.
- Only standard GA instances support the access log feature. Basic GA instances do not support the access log feature. In this topic, a standard GA instance is used as an example.
- You cannot collect the access log of an endpoint group if the endpoint group is deployed on a point of presence (PoP) node of Alibaba Cloud.
- You cannot query the domain names of endpoints.
- The access log feature is automatically enabled for GA instances that are created after January 8, 2022. If you want to enable the access log feature for GA instances that are created before January 8, 2022, submit a ticket to upgrade the GA instances.

Click here to view more information about access logs.

The following table describes the access log information that you can query in the Log Service console.

Parameter	Description
accelerator_region	The acceleration region.
client_ip	The IP address of the client, which is the source IP address.
client_port	The port of the client, which is the source port.
egress_bytes	The outbound traffic during the time period when traffic information is collected.
endpoint_group_id	The ID of the endpoint group.
endpoint_group_region	The region where the endpoint group is deployed.
endpoint_ip	The IP address of the endpoint, which is the destination IP address.
endpoint_port	The port of the endpoint, which is the destination port.
ga_id	The ID of the GA instance.
ingress_bytes	The inbound traffic during the time period when traffic information is collected.

Parameter	Description
listener_id	The ID of the listener.
protocol	The network transmission protocol that is used by the listener.
status	The status of the response packet that is sent by GA.
time	The time when the log entry is generated.
session_time	The duration of the session, which starts from the time when GA receives the request and ends at the time when the last byte is sent to the client.
end_time	The time when the session ends.
epg_region	The region where the endpoint group is deployed.

The following parameters are available when HTTP and HTTPS listeners are used.

Parameter	Description
http_host	The Host header of the request.
http_referer	The HTTP referer header of the request.
request_method	The request method.
request_uri	The URI of the request that is received by GA.

Scenarios

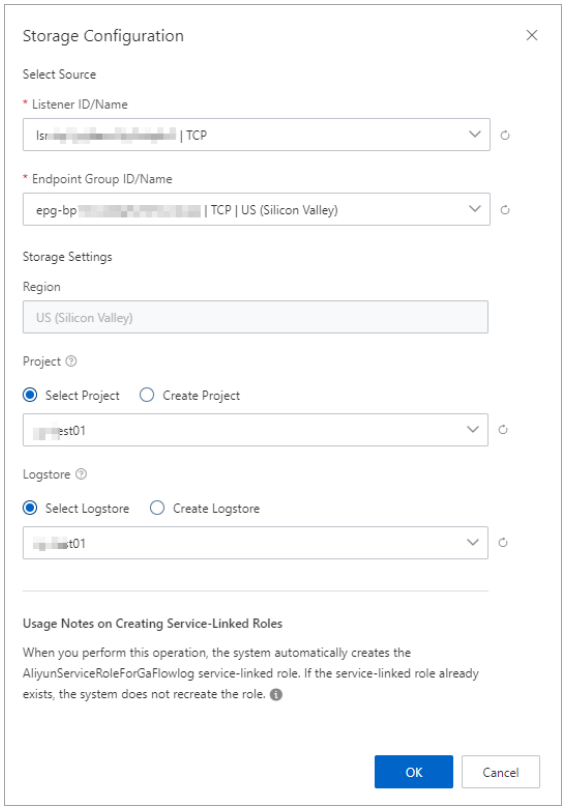
Billing

Limits

Create an access log

Before you create an access log for a GA instance, make sure that you have added listeners and endpoint groups for the GA instance. For more information, see Add and manage listeners.

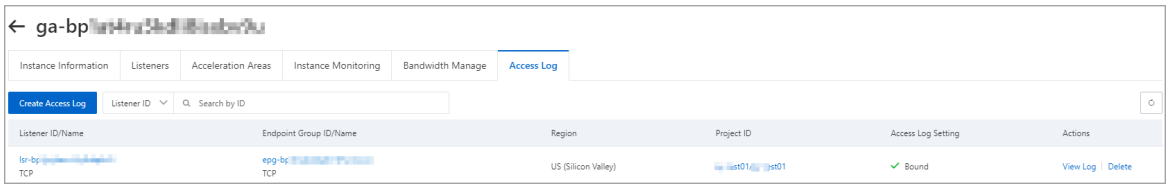
1.
2. On the **Instances** page, click the ID of the GA instance that you want to manage.
3. On the instance details page, click the **Access Log** tab.
4. On the **Access Log** tab, click **Create Access Log**. In the **Storage Configuration** dialog box, set the following parameters and click **OK**.



Parameter		Description
Select Source	**Listener ID/Name**	Select a listener.
	Endpoint Group ID/Name	Select a destination endpoint group.
Storage Settings	**Region**	By default, the region where the endpoint group resides is selected.
	Project	Log Service projects are used to isolate and manage resources. You can click **Select Project** and select an existing project. You can also click **Create Project** and create a project.
	Logstore	Log Service Logstores are used to collect, store, and query log data. You can click **Select Logstore** and select an existing Logstore. You can also click **Create Logstore** and create a Logstore.

> **Note** When you perform this operation, the system checks whether the service-linked role AliyunServiceRoleForGaFlowlog is assigned to GA.
>
> - If the service-linked role AliyunServiceRoleForGaFlowlog does not exist, the system automatically creates the service-linked role and attaches the permission policy AliyunServiceRolePolicyForGaFlowlog to the service-linked role. This allows GA to access Log Service and deliver flow logs to Log Service.
> - If the service-linked role AliyunServiceRoleForGaFlowlog is assigned to GA, the system does not create it again.
>
> For more information, see AliyunServiceRoleForGaFlowlog.

After you create the access log, you can find it on the **Access Log** tab.



What to do next

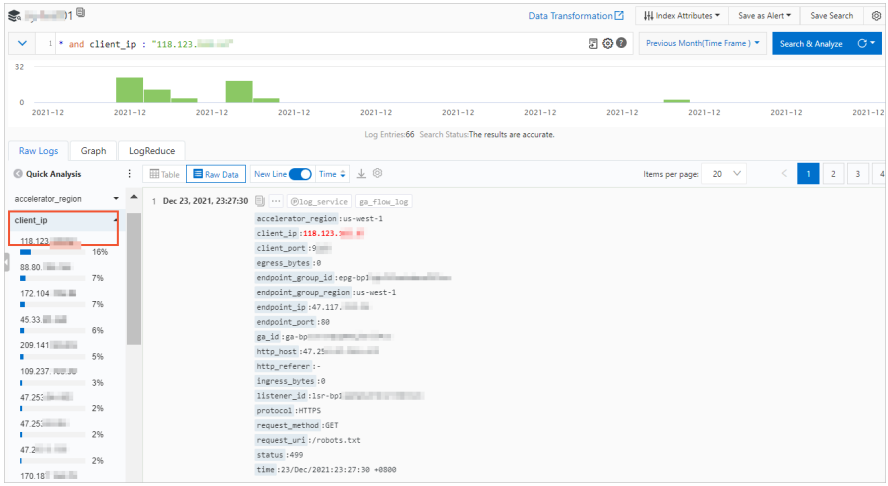
Operation	Description
View access logs	1. On the **Access Log** tab, find the access log that you created and click **View Log** in the **Actions** column to go to the Log Service console. 2. You can view and analyze the access log. For more information, see Examples.
Delete an access log	1. On the **Access Log** tab, find the access log that you want to delete and click **Delete** in the **Actions** column. 2. In the **Delete Log** message, click **OK**.

After Log Service collects an access log, you can download, deliver, and process the access log. You can also create alerts for the access log. For more information, see Common operations on logs of Alibaba Cloud services.

Examples

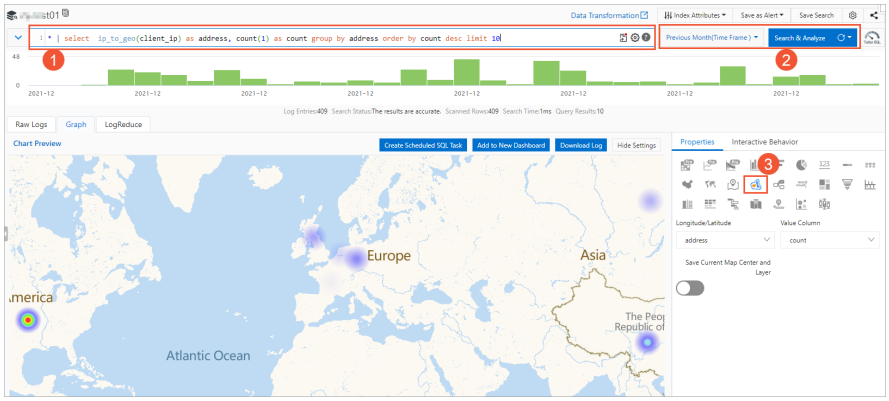
On the **Raw Logs** tab of the Logstore page, you can view information about raw logs.

For example, you can click **client_ip** to view information about client IP addresses.



On the Logstore page, enter an SQL statement in the Search & Analyze search box to search for a specified access log.

For example, you can query the distribution of client IP addresses based on the order in the following figure.



No.	Description
1	Enter the following SQL statement to query the heat map of client IP addresses and view the top 10 regions where the clients are distributed. This helps you plan your business. `* | select ip_to_geo(client_ip) as address, count(1) as count group by address order by count desc limit 10`
2	Select a time range during which access logs are generated and click **Search & Analyze**.
3	On the **Graph** tab, click the **Properties** tab and then click the icon to view the distribution of client IP addresses.

View a raw access log

Query a specified access log

8.Manage quotas

This topic describes how to manage quotas of Global Accelerator (GA). If the quota of a cloud resource is insufficient, you can apply for a quota increase.

Procedure

1.
2. In the left-side navigation pane, click **Quota Management**.
3. On the **Quota Management** page, view the quota usage of GA resources for the current Alibaba Cloud account.

Quota Name	Description	Type	Used quota/Total quota	Actions
gaplus_quota_accelerator	The maximum number of global accelerator instances for each user.	Quota	10	Submit Application
gaplus_quota_endpoint	The maximum number of endpoints for each endpoint group.	Quota	4	Submit Application
gaplus_quota_listener_per_accelerator	The maximum number of listeners for each instance.	Quota	10	Submit Application

4. To increase a quota, click **Submit Application** in the **Actions** column, set the following parameters, and then click **OK**.
   - **Requested Value**: Specify the requested value. You must enter a number that is greater than the current quota. For more information about default quota limits, see Limits.
   - **Reason for Application**: Enter the detailed reason for the application, including the scenarios and necessity.
   - **Mobile/Landline Phone Number**: Enter the mobile or landline phone number of the applicant.
   - **Email**: Enter the email address of the applicant.

Result

After you submit the application, you can click **History** in the **Actions** column to view the application status.

The system automatically assesses whether to approve your application.

- If the requested value exceeds the upper limit, the system automatically rejects the application and the application status changes to **Rejected**.

  If your application is rejected, reduce the requested value and submit the application again.
- If the requested value falls within the expected range, the system automatically approves the application, the application status changes to **Approved**, and the requested value immediately takes effect.

9.Permission management

9.1. Service-linked role

9.1.1. AliyunServiceRoleForGaVpcEndpoint

You can specify an Elastic Compute Service (ECS) instance or a Classic Load Balancer (CLB) instance (formerly known as an SLB instance) as an endpoint for a Global Accelerator (GA) instance. In this case, if your GA instance does not have the service-linked role AliyunServiceRoleForGaVpcEndpoint, the system automatically creates the service-linked role.

Overview

AliyunServiceRoleForGaVpcEndpoint is a service-linked role of GA. If you want to specify an ECS instance or a CLB instance as an endpoint, make sure that your GA instance has the service-linked role AliyunServiceRoleForGaVpcEndpoint.

> **Note** A service-linked role is a Resource Access Management (RAM) role that is associated with an Alibaba Cloud service. In some cases, to use a feature of a cloud service, you must first acquire the permissions to access other cloud services. Service-linked roles simplify the authorization process and avoid user errors. For more information, see Service-linked roles.

Permissions required to create the service-linked role

By default, an Alibaba Cloud account is authorized to create the service-linked role AliyunServiceRoleForGaVpcEndpoint. RAM users must be granted the following permissions to create the service-linked role:

```
{
      "Action": "ram:CreateServiceLinkedRole",
      "Resource": "*",
      "Effect": "Allow",
      "Condition": {
        "StringEquals": {
          "ram:ServiceName": "vpcendpoint.ga.aliyuncs.com"
        }
      }
}
```

You can authorize a RAM user to create the service-linked role by using one of the following methods:

- Attach the administrator permission policy AliyunGlobalAccelerationFullAccess to the RAM user. For more information, see Grant permissions to a RAM role.

  > **Note** The permissions required to create the service-linked role AliyunServiceRoleForGaVpcEndpoint are included in the administrator permission policy AliyunGlobalAccelerationFullAccess. You can attach the administrator permission policy to a RAM user. This way, the RAM user can create the service-linked role AliyunServiceRoleForGaVpcEndpoint.

- Attach a custom permission policy to a RAM user. The following code block shows the content of the custom permission policy:

```
{
      "Action": "ram:CreateServiceLinkedRole",
      "Resource": "*",
      "Effect": "Allow",
      "Condition": {
        "StringEquals": {
          "ram:ServiceName": "vpcendpoint.ga.aliyuncs.com"
        }
      }
}
```

  For more information, see Create a custom policy and Grant permissions to a RAM role.

Create the service-linked role

When you specify an ECS instance or a CLB instance as an endpoint for a GA instance, the system checks whether the GA instance has the service-linked role AliyunServiceRoleForGaVpcEndpoint. In this case, the following rules apply to the GA instance:

- If the GA instance does not have the service-linked role AliyunServiceRoleForGaVpcEndpoint, the system automatically creates the service-linked role and attaches the permission policy AliyunServiceRoleForGaVpcEndpoint to the service-linked role. This allows GA to access ECS and CLB. The following code block shows the content of the permission policy:

```
{
  "Version": "1",
  "Statement": [
    {
      "Effect": "Allow",
      "Resource": "*",
      "Action": [
        "ecs:CreateNetworkInterface",
        "ecs:DeleteNetworkInterface",
        "ecs:DescribeNetworkInterfaces",
        "ecs:ModifyNetworkInterfaceAttribute",
        "ecs:DescribeSecurityGroups",
        "ecs:CreateSecurityGroup",
        "ecs:AuthorizeSecurityGroup",
        "ecs:AuthorizeSecurityGroupEgress",
        "ecs:RevokeSecurityGroup",
        "ecs:RevokeSecurityGroupEgress",
        "ecs:JoinSecurityGroup",
        "ecs:LeaveSecurityGroup",
        "ecs:DeleteSecurityGroup",
        "ecs:DescribeSecurityGroupAttribute",
        "ecs:DescribeSecurityGroups",
        "ecs:DescribeSecurityGroupReferences",
        "ecs:ModifySecurityGroupAttribute",
        "ecs:ModifySecurityGroupEgressRule",
        "ecs:ModifySecurityGroupPolicy",
        "ecs:ModifySecurityGroupRule",
        "vpc:DescribeVSwitches"
      ]
    },
    {
      "Action": "ram:DeleteServiceLinkedRole",
      "Resource": "*",
      "Effect": "Allow",
      "Condition": {
        "StringEquals": {
          "ram:ServiceName": "vpcendpoint.ga.aliyuncs.com"
        }
      }
    }
  ]
}
```

- If the GA instance has the service-linked role AliyunServiceRoleForGaVpcEndpoint, the system does not create the service-linked role again.

Delete the service-linked role

The system does not automatically delete the service-linked role AliyunServiceRoleForGaVpcEndpoint. To delete the service-linked role, you must first delete the ECS instance or CLB instance that serves as an endpoint. For more information, see the following topics:

1. Delete an endpoint
2. Delete a service-linked role

9.1.2. AliyunServiceRoleForGaFlowlog

This topic describes the scenarios of the service-linked role AliyunServiceRoleForGaFlowlog and how to create and delete the service-linked role.

Overview

AliyunServiceRoleForGaFlowlog is a service-linked role of Global Accelerator (GA). After you create AliyunServiceRoleForGaFlowlog, GA can access your Log Service and deliver logs to Log Service.

> **Note** A service-linked role is a Resource Access Management (RAM) role that is associated with an Alibaba Cloud service. In some scenarios, to use a feature of a cloud service, you must obtain the permissions to access other cloud services. Service-linked roles simplify the authorization process and avoid risks caused by user errors. For more information, see Service-linked roles.

Permissions required to create AliyunServiceRoleForGaFlowlog

You can use an Alibaba Cloud account to create AliyunServiceRoleForGaFlowlog. If you want to create AliyunServiceRoleForGaFlowlog as a RAM user, the RAM user must first obtain the following permissions:

```
{
      "Action": "ram:CreateServiceLinkedRole",
      "Resource": "*",
      "Effect": "Allow",
      "Condition": {
        "StringEquals": {
          "ram:ServiceName": "flowlog.ga.aliyuncs.com"
        }
      }
}
```

You can grant the RAM user the required permissions in one of the following ways:

- Attach the administrator permission policy AliyunGlobalAccelerationFullAccess to the RAM user. For more information, see Grant permissions to a RAM role.

  > **Note** The permission to create a service-linked role is included in AliyunGlobalAccelerationFullAccess. Therefore, you can create a service-linked role as a RAM user after you attach AliyunGlobalAccelerationFullAccess to the RAM user.

- Create a custom permission policy and attach it to the RAM user. The following code block shows the content of the custom permission policy:

```
{
      "Action": "ram:CreateServiceLinkedRole",
      "Resource": "*",
      "Effect": "Allow",
      "Condition": {
        "StringEquals": {
          "ram:ServiceName": "flowlog.ga.aliyuncs.com"
        }
      }
}
```

For more information, see Create a custom policy and Grant permissions to a RAM role.

Create AliyunServiceRoleForGaFlowlog

After you enable the log delivery feature of flow logs for GA, the system automatically creates the service-linked role AliyunServiceRoleForGaFlowlog, and attaches a permission policy named AliyunServiceRolePolicyForGaFlowlog to it. The permission policy allows GA to access flow logs. The following code block shows the content of the permission policy:

```
{
  "Version": "1",
  "Statement": [
    {
      "Action": [
        "log:PostLogStoreLogs"
      ],
      "Resource": "*",
      "Effect": "Allow"
    },
    {
      "Action": "ram:DeleteServiceLinkedRole",
      "Resource": "*",
      "Effect": "Allow",
      "Condition": {
        "StringEquals": {
          "ram:ServiceName": "flowlog.ga.aliyuncs.com"
        }
      }
    }
  ]
}
```

Delete AliyunServiceRoleForGaFlowlog

The system cannot automatically delete the service-linked role AliyunServiceRoleForGaFlowlog of GA. To manually delete AliyunServiceRoleForGaFlowlog, delete all GA instances first. For more information, see Delete a service-linked role.

9.1.3. AliyunServiceRoleForGaAlb

When you specify an Application Load Balancer (ALB) instance as an origin server, your GA instance must assume the service-linked role AliyunServiceRoleForGaAlb. If your GA instance does not assume the service-linked role, the system automatically creates the role for your GA instance.

AliyunServiceRoleForGaAlb

AliyunServiceRoleForGaAlb is a service-linked role of GA. To specify an ALB instance as an origin server, your GA instance must assume the service-linked role AliyunServiceRoleForGaAlb.

> **Note** A service-linked role is a Resource Access Management (RAM) role that is associated with an Alibaba Cloud service. In some cases, to use a feature of a cloud service, you must first acquire the permissions to access other cloud services. Service-linked roles simplify the authorization process and avoid user errors. For more information, see Service-linked roles.

Permissions required to create AliyunServiceRoleForGaAlb

By default, an Alibaba Cloud account is authorized to create the service-linked role AliyunServiceRoleForGaAlb. If a RAM user wants to create the service-linked role, you must first use the Alibaba Cloud account to grant the following permissions to the RAM user:

```
{
      "Action": "ram:CreateServiceLinkedRole",
      "Resource": "*",
      "Effect": "Allow",
      "Condition": {
        "StringEquals": {
          "ram:ServiceName": "alb.ga.aliyuncs.com"
        }
      }
}
```

You can grant the RAM user the required permissions by using one of the following methods:

- Attach the administrator permission policy AliyunGlobalAccelerationFullAccess to the RAM user. For more information, see Grant permissions to a RAM role.

  > **Note** The permissions required to create the service-linked role AliyunServiceRoleForGaAlb are included in the administrator permission policy AliyunGlobalAccelerationFullAccess. Therefore, after you attach the administrator permission policy to a RAM user, the RAM user can create the service-linked role AliyunServiceRoleForGaAlb.

- Attach a custom permission policy to a RAM user. The following code block shows the content of the custom permission policy:

```
{
      "Action": "ram:CreateServiceLinkedRole",
      "Resource": "*",
      "Effect": "Allow",
      "Condition": {
        "StringEquals": {
          "ram:ServiceName": "alb.ga.aliyuncs.com"
        }
      }
}
```

For more information, see Create a custom policy and Grant permissions to a RAM role.

Create the service-linked role AliyunServiceRoleForGaAlb

When you specify an ALB instance as an origin server, the system checks whether your GA instance assumes the service-linked role AliyunServiceRoleForGaAlb.

- If your GA instance does not assume the service-linked role AliyunServiceRoleForGaAlb, the system automatically creates the service-linked role and attaches the permission policy AliyunServiceRoleForGaAlb to the service-linked role. This allows GA to access ALB. The following code block shows the content of the permission policy:

```
{
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "alb:GetLoadBalancerAttribute",
      "Resource": "*"
    },
    {
      "Action": "ram:DeleteServiceLinkedRole",
      "Resource": "*",
      "Effect": "Allow",
      "Condition": {
        "StringEquals": {
          "ram:ServiceName": "alb.ga.aliyuncs.com"
        }
      }
    }
  ],
  "Version": "1"
}
```

- If your GA instance assumes the service-linked role AliyunServiceRoleForGaAlb, the system does not create the service-linked role again.

Delete the service-linked role AliyunServiceRoleForGaAlb

The system does not automatically delete the service-linked role AliyunServiceRoleForGaAlb. To delete the service-linked role, you must first disassociate the ALB instance from your GA instance. Then, you can delete the service-linked role. For more information, see:

1. Delete an endpoint
2. Delete a service-linked role

9.1.4. AliyunServiceRoleForGaOss

When you specify an Object Storage Service (OSS) instance as an origin server, your GA instance must assume the service-linked role AliyunServiceRoleForGaOss. If your GA instance does not assume the service-linked role, the system automatically creates the role for your GA instance.

AliyunServiceRoleForGaOss

AliyunServiceRoleForGaOss is a service-linked role of GA. To specify an OSS instance as an origin server, your GA instance must assume the service-linked role AliyunServiceRoleForGaVpcEndpoint.

> **Note** A service-linked role is a Resource Access Management (RAM) role that is associated with an Alibaba Cloud service. In some cases, to use a feature of a cloud service, you must first acquire the permissions to access other cloud services. Service-linked roles simplify the authorization process and avoid risks caused by user errors. For more information, see Service-linked roles.

Permissions required to create AliyunServiceRoleForGaOss

By default, an Alibaba Cloud account is authorized to create the service-linked role AliyunServiceRoleForGaOss. If a RAM user wants to create the service-linked role, you must first use the Alibaba Cloud account to grant the following permissions to the RAM user:

```
{
      "Action": "ram:CreateServiceLinkedRole",
      "Resource": "*",
      "Effect": "Allow",
      "Condition": {
        "StringEquals": {
          "ram:ServiceName": "oss.ga.aliyuncs.com"
        }
      }
}
```

You can grant the RAM user the required permissions by using one of the following methods:

- Attach the administrator permission policy AliyunGlobalAccelerationFullAccess to the RAM user. For more information, see Grant permissions to a RAM role.

  > **Note** The permissions required to create the service-linked role AliyunServiceRoleForGaOss are included in the administrator permission policy AliyunGlobalAccelerationFullAccess. Therefore, after you attach the administrator permission policy to a RAM user, the RAM user can create the service-linked role AliyunServiceRoleForGaOss.

- Attach a custom permission policy to a RAM user. The following code block shows the content of the custom permission policy:

```
{
      "Action": "ram:CreateServiceLinkedRole",
      "Resource": "*",
      "Effect": "Allow",
      "Condition": {
        "StringEquals": {
          "ram:ServiceName": "oss.ga.aliyuncs.com"
        }
      }
}
```

For more information, see Create a custom policy and Grant permissions to a RAM role.

Create the service-linked role AliyunServiceRoleForGaOss

When you specify an ALB instance as an origin server, the system checks whether your GA instance assumes the service-linked role AliyunServiceRoleForGaOss.

- If your GA instance does not assume the service-linked role AliyunServiceRoleForGaOss, the system automatically creates the service-linked role and attaches the permission policy AliyunServiceRoleForGaOss to the service-linked role. This allows GA to access OSS. The following code block shows the content of the permission policy:

```
{
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "oss:getBucketInfo",
      "Resource": "*"
    },
    {
      "Action": "ram:DeleteServiceLinkedRole",
      "Resource": "*",
      "Effect": "Allow",
      "Condition": {
        "StringEquals": {
          "ram:ServiceName": "oss.ga.aliyuncs.com"
        }
      }
    }
  ],
  "Version": "1"
}
```

- If your GA instance assumes the service-linked role AliyunServiceRoleForGaOss, the system does not create the service-linked role again.

Delete the service-linked role AliyunServiceRoleForGaOss

The system does not automatically delete the service-linked role AliyunServiceRoleForGaOss. To delete the service-linked role, you must first disassociate the OSS instance from your GA instance. Then, you can delete the service-linked role. For more information, see:

1. Delete an endpoint
2. Delete a service-linked role

9.2. Grant permissions to a RAM user

By default, Resource Access Management (RAM) users cannot create Global Accelerator (GA) resources, or access or manage GA resources created by Alibaba Cloud accounts. If you want to access or manage GA resources as a RAM user, you must first grant the required permissions to the RAM user.

Prerequisites

A RAM user is created. For more information, see Create a RAM user.

Procedure

1. Log on to the RAM console with your Alibaba Cloud account.
2. In the left-side navigation pane, choose **Identities > Users**.
3. On the **Users** page, find the RAM user and click **Add Permissions** in the **Actions** column.
4. In the **Add Permissions** panel, set the following parameters and click **OK**.

Parameter	Description
Authorized Scope	The authorization scope. Valid values: ○ **Alibaba Cloud Account**: The authorization takes effect on the current Alibaba Cloud account. ○ **Specific Resource Group**: The authorization takes effect on a specified resource group.
Principal	The system automatically specifies the RAM user created in Step as the principal.
Select Policy	Select **System Policy** and then select permission policies that you want to attach to the RAM user. You can attach the following system policies of GA to a RAM user: ○ **AliyunGlobalAccelerationReadOnlyAccess**: Grants the RAM user read-only permissions on GA. ○ **AliyunGlobalAccelerationFullAccess**: Grants the RAM user full permissions on GA.

5. Confirm the authorization scope and permission policies and click **Complete**.