ALIBABA CLOUD

Alibaba Cloud

Global Acceleration

Quick Start

Document Version: 20211118

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.Get started with GA

This topic describes the workflow and basic scenarios to help you get started with Global Accelerator (GA).

> **Note** First-time users can apply for a one-month free trial of GA with their Alibaba Cloud accounts from April 1, 2021 to April 1, 2022. The Alibaba Cloud accounts must have passed enterprise real-name verification. To apply for a free trial, visit the GA Free Trial Application page. For more information about the free trial and supported instance types, see GA free trial.

Workflow

Basic scenarios

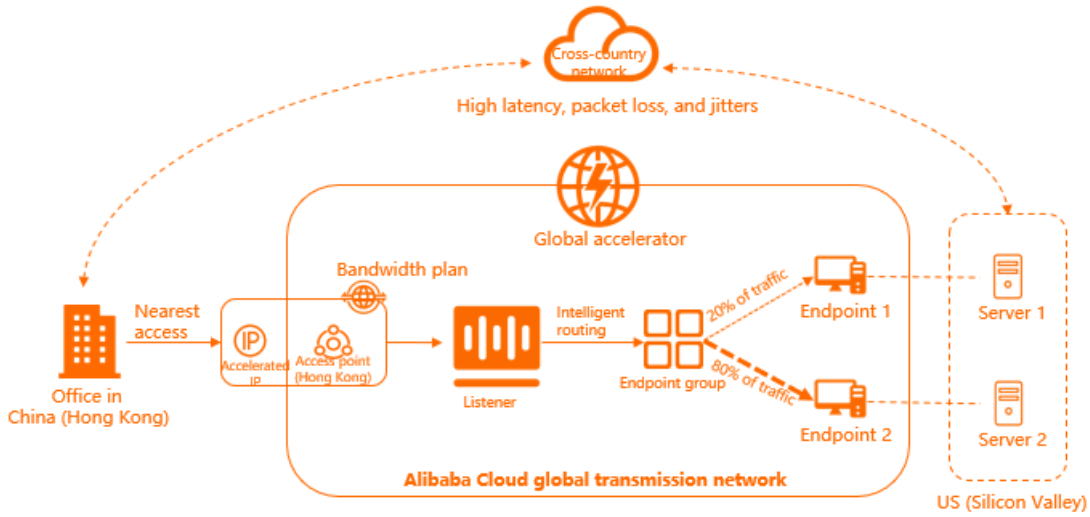
- Accelerate transmission of network traffic destined for a specified IP address
- Accelerate a specified domain name
- Accelerate transmission of network traffic between mainland China and other areas

2.Accelerate transmission of network traffic destined for a specified IP address

This topic describes how to use Global Accelerator (GA) to accelerate connections to backend servers that have specific IP addresses.

Scenario

The following scenario is used as an example in this topic. A company has built two on-premises backend servers Server 1 and Server 2 in the US (Silicon Valley) region and deployed enterprise applications on the servers. Server 1 processes up to 20% of the total workloads. Server 2 processes up to 80% of the total workloads. Unstable network performance may cause issues such as increased latency, network jitter, and packet loss. These issues arise when users from the office in the China (Hong Kong) region connect to the enterprise applications deployed in the US (Silicon Valley) region over the Internet.



To accelerate connections to the backend servers, you can create a GA instance that provides an access point in China (Hong Kong). When users in the China (Hong Kong) region send requests to the servers, the access point in the China (Hong Kong) region receives the requests and forwards the requests to the endpoints in the US (Silicon Valley) region through intelligent routing. The system uses the endpoints to distribute 20% of the requests to Server 1 and 80% of the requests to Server 2.

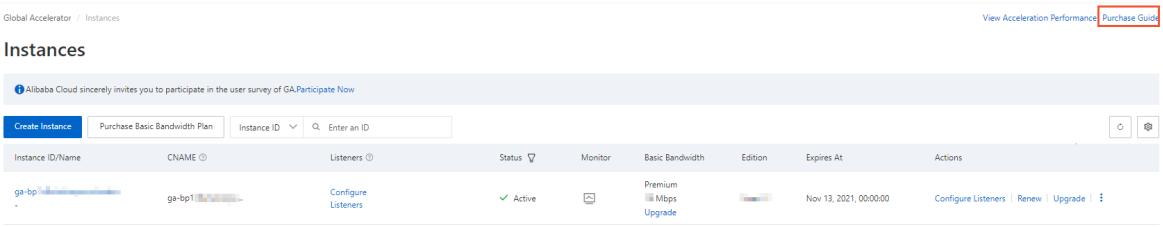
Procedure

Step 1: Purchase a GA service bundle

You can enter the information about the web service in the GA console. After you enter the information, the system generates a list of recommended services. The list includes a GA instance and a basic bandwidth plan.

1. Log on to the Global Accelerator console.
2. In the upper-right corner of the **Instances** page, click **Purchase Guide**.

   > **Note** If this is the first time that you use the GA service, skip this step.

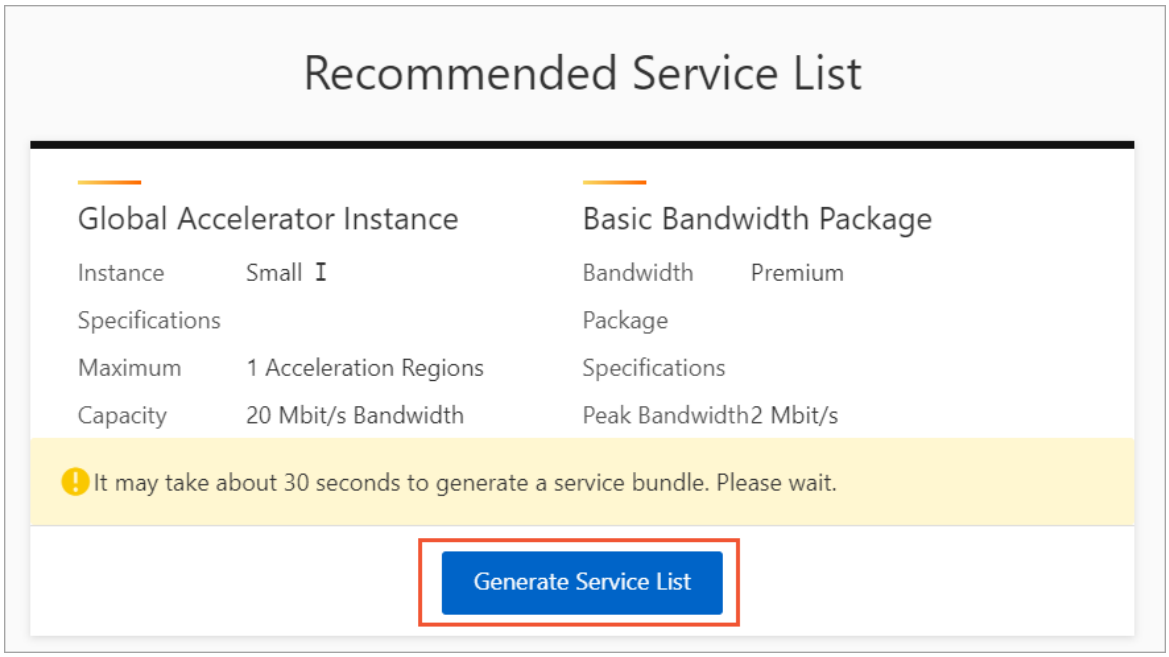
   

3. In the **Enter the required information to generate a list of recommended services** section, enter the required information and click **Generate Service List**.

Parameter	Description
Acceleration Area	Select the region that requires acceleration. In this example, **China (Hong Kong)** is selected.
Service Region	Select the region where the backend servers are deployed. In this example, **US (Silicon Valley)** is selected.
ICP Filing	Specify whether you have applied for an Internet Content Provider (ICP) number for the domain name of the web service. In this example, **No** is selected. **Note** All websites must obtain an ICP number before they are allowed to provide services to users in mainland China. For more information, see What is an ICP filing?.
Server Area	Specify whether the web service is deployed on Alibaba Cloud. In this example, **Off Alibaba Cloud** is selected.
Peak Bandwidth Range	Enter the bandwidth required during peak hours. Unit: Mbit/s. In this example, *2* is entered.
Maximum Concurrent Connections	The maximum number of concurrent connections that a GA instance supports. When the number of existing concurrent connections reaches the upper limit, new connection requests are dropped. In this example, **5 Thousand** is selected.

4. In the **Recommended Service List** section, click **Generate Service List** after you confirm the

information.



5. On the buy page, set the following parameters and click **Buy Now** to complete the payment.

Parameter	Description
Term	Select the subscription duration.
Specification	Select a specification for the GA instance. In this example, **Small I (Specification Unit)** is selected.
Bandwidth Type	Select a bandwidth type for the basic bandwidth plan. In this example, **Premium** is selected.
Peak Bandwidth	Select the bandwidth limit of the basic bandwidth plan. In this example, **2** Mbit/s is selected.

Step 2: Add an acceleration area

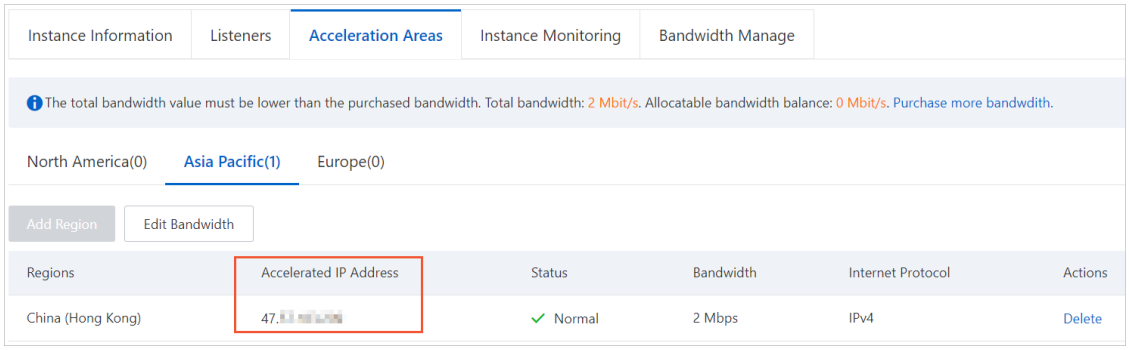
After you purchase a GA instance, you can add an acceleration area, specify the region where users are located, and then allocate bandwidth resources to the region.

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

Parameter	Description
Regions	Select the region where the users are located. In this example, **China (Hong Kong)** is selected.

Parameter	Description
Bandwidth	Specify a bandwidth value for the acceleration region. In this example, *2* Mbit/s is entered.
Internet Protocol	Select the IP address version used to access GA. In this example, **IPv4** is selected.

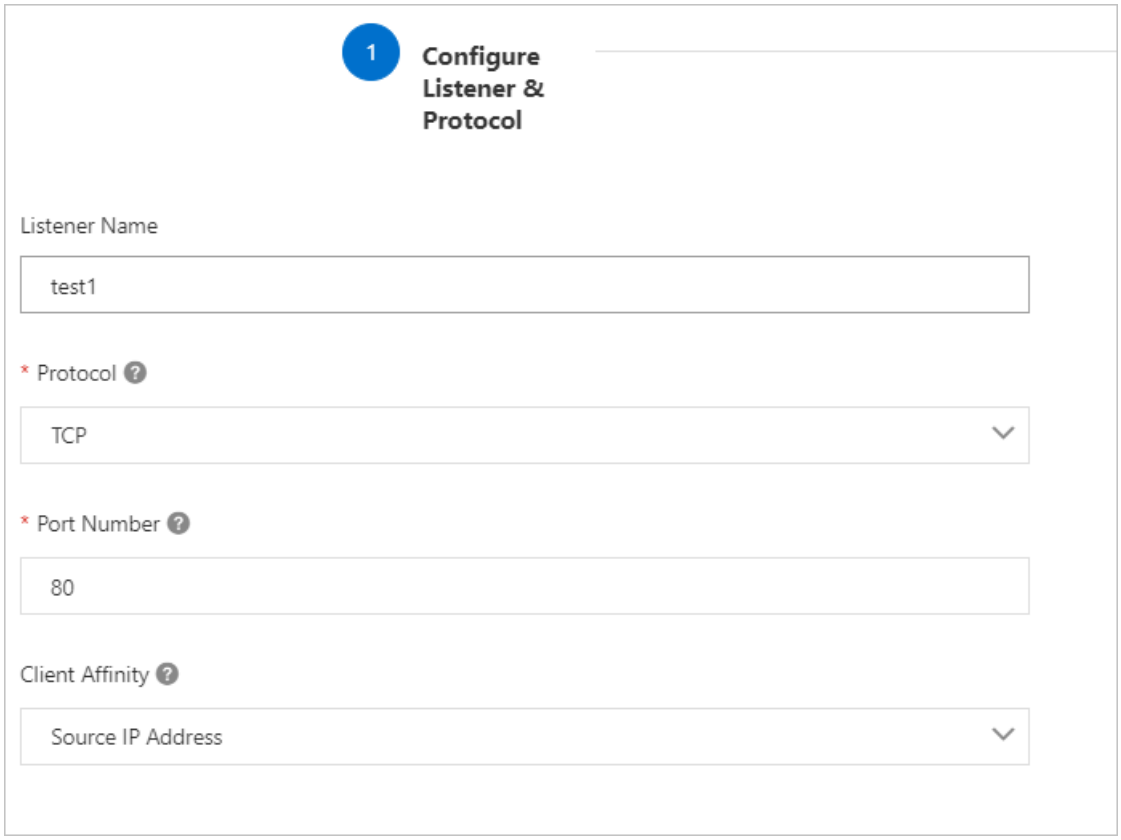
After you add the region, the system assigns an accelerated IP address to the region that is added to the GA instance. This accelerated IP address is used to accelerate data transfer from users in the specified region to the specified backend servers through GA.



Step 3: Add a listener and an endpoint group

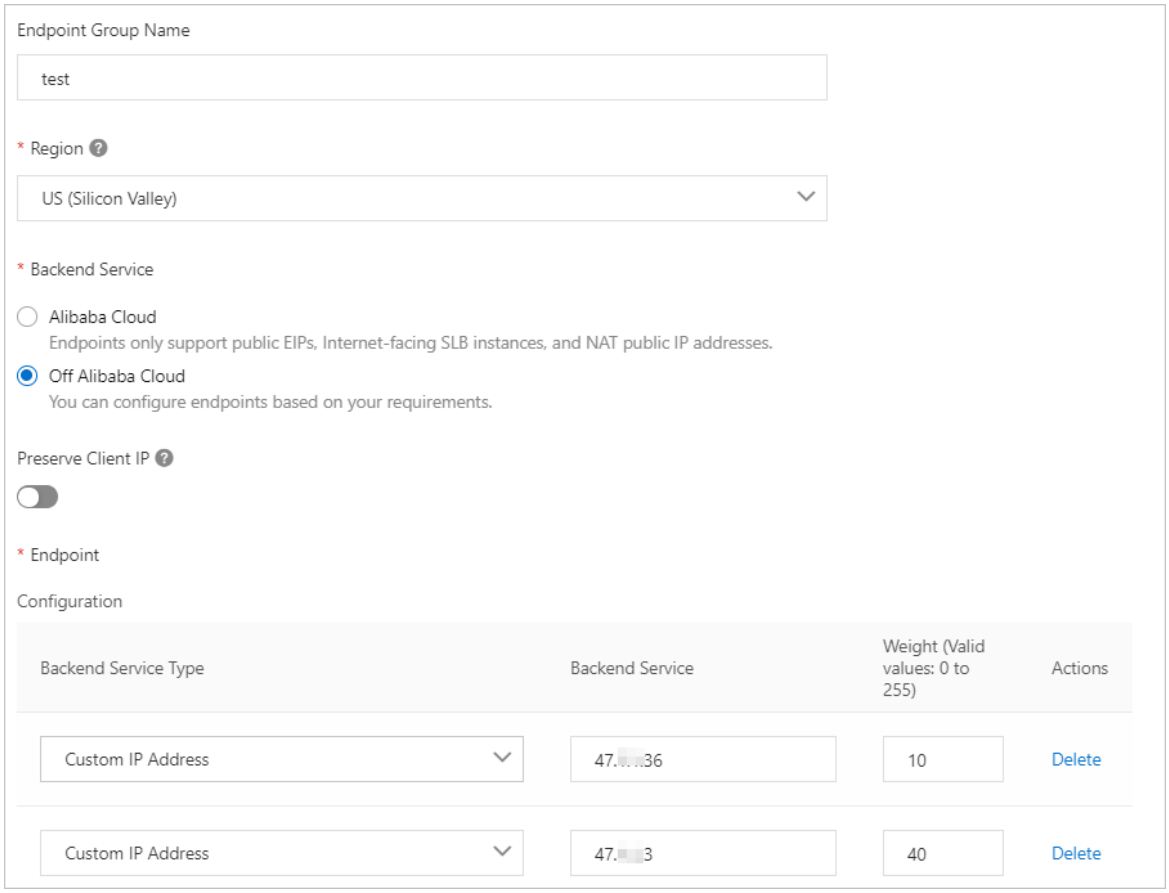
A listener checks for connection requests and then distributes the requests to backend servers based on the specified protocol and ports. Each listener is associated with an endpoint group. You can associate an endpoint group with a listener by specifying the region to which you want to distribute network traffic. After you associate an endpoint group with a listener, traffic is distributed to the optimal endpoint in the associated endpoint group.

1. On the instance details page, click the **Listeners** tab and then click **Add Listener**.
2. 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 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, **Source IP Address** is selected.

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



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 digits, underscores (_), and hyphens (-). The name must start with a letter.
Region	Select the region to which the endpoint group belongs, which is the region to which the destination servers belong. In this example, **US (Silicon Valley)** is selected.
Backend Service	Specify whether the backend service is deployed on Alibaba Cloud. **Off Alibaba Cloud** is selected in this example.
Preserve Client IP	Specify whether to preserve client IP addresses. After the feature is enabled, backend servers can obtain source IP addresses of clients. In this example, this feature is disabled.

Parameter	Description
Endpoint	An endpoint serves as the destination that a client requests to access. Set the following parameters to configure an endpoint: i. To configure Server 1 as an endpoint, set the following parameters: ■ **Backend Service Type**: Select **Custom IP Address** from the drop-down list. ■ **Backend Service**: Enter the IP address that you want to accelerate. In this example, the IP address of Server 1 is specified. ■ **Weight**: Specify a weight for the endpoint. Valid values: 0 to 255. GA routes network traffic to endpoints based on the specified weights. In this example, the weight of Server 1 is set to **10**. ii. Click **Add Endpoint** to configure Server 2 as an endpoint. Set the following parameters: ■ **Backend Service Type**: Select **Custom IP Address** from the drop-down list. ■ **Backend Service**: Enter the IP address of Server 2. ■ **Weight**: Set the weight of Server 2 to **40**. **Notice** If you set the weight of an endpoint to 0, GA does not route network traffic to the endpoint. Proceed with caution.

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

Step 4: Test the acceleration performance

Note If you specify UDP as the protocol when you add a listener, you can verify the acceleration performance by using UDPing. For more information, see Test the acceleration performance a GA instance that uses a UDP listener.

1. Open the CLI on an on-premises machine in the China (Hong Kong) region.
2. Run the following command to query the network latency:

```
curl -o /dev/null -s -w "time_connect: %{time_connect}\ntime_starttransfer: %{time_starttransfer}\ntime_total: %{time_total}\n" "http[s]://<accelerated IP address>[:<port>]"
```

In the command:

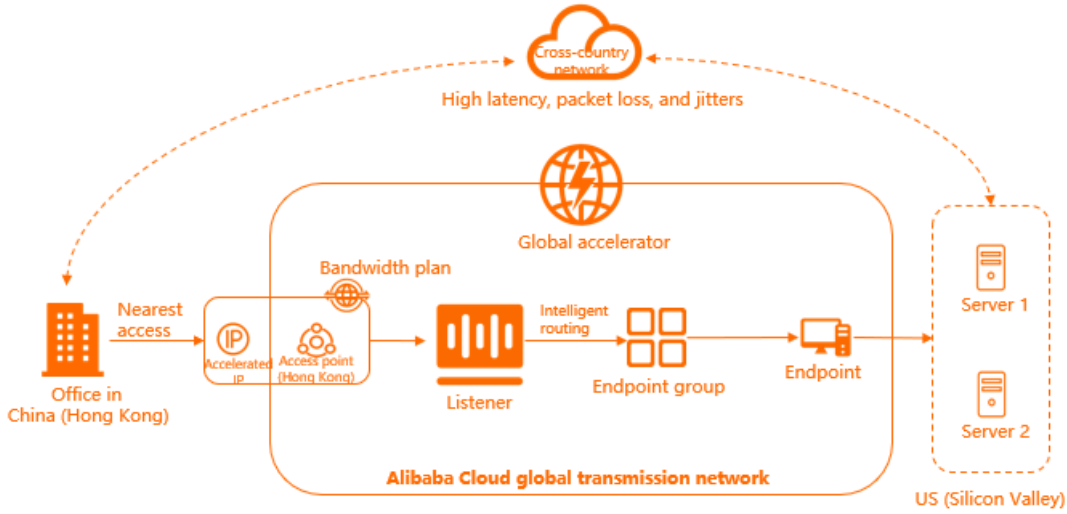
- time_connect: the period of time that it takes to establish a TCP connection. Unit: seconds.
- time_starttransfer: the start time of data transfer. The start time refers to the amount of time from when the client sends a request to the backend server to when the first byte is sent to the client. Unit: seconds.
- time_total: the total connection time. The total connection time refers to the amount of time from when the client sends a request to when the client receives the last byte from the backend server. Unit: seconds.

3.Accelerate a specified domain name

This topic describes how to use Global Accelerator (GA) to accelerate a specified domain name.

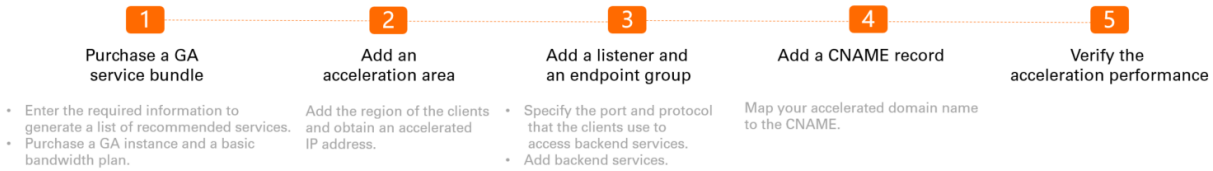
Scenario

The following scenario is used as an example in this topic. The headquarters of a company is located in the US (Silicon Valley) region. A web service is deployed on two self-managed servers. The web service can be accessed through domain name www.example.com and port 80. When users from the office in the China (Hong Kong) region access the web service deployed in the US (Silicon Valley) region over the Internet, the network condition is unstable. Issues such as network latency, network jitter, and packet loss may occur.



You can purchase a GA instance to accelerate network traffic from users in the China (Hong Kong) region to servers in the US (Silicon Valley) region through an accelerated IP address. GA forwards user requests to the global transmission network of Alibaba Cloud and then uses intelligent routing to route the requests to the endpoints in the US (Silicon Valley) region.

Procedure

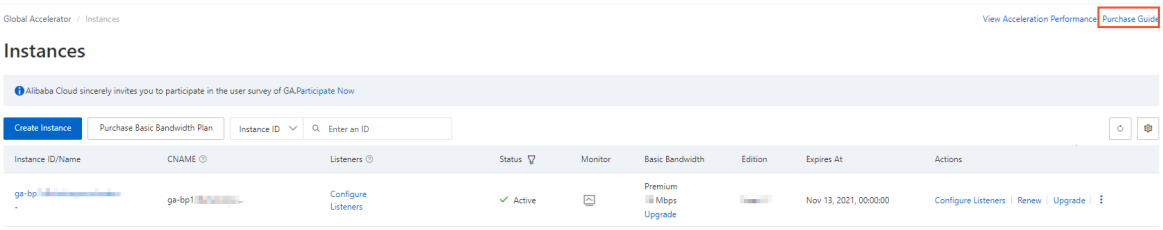


Step 1: Purchase a GA service bundle

You can enter the information about the web service in the GA console. After you enter the information, the system generates a list of recommended services. The list includes a GA instance and a basic bandwidth plan.

1. Log on to the Global Accelerator console.
2. In the upper-right corner of the **Instances** page, click **Purchase Guide**.

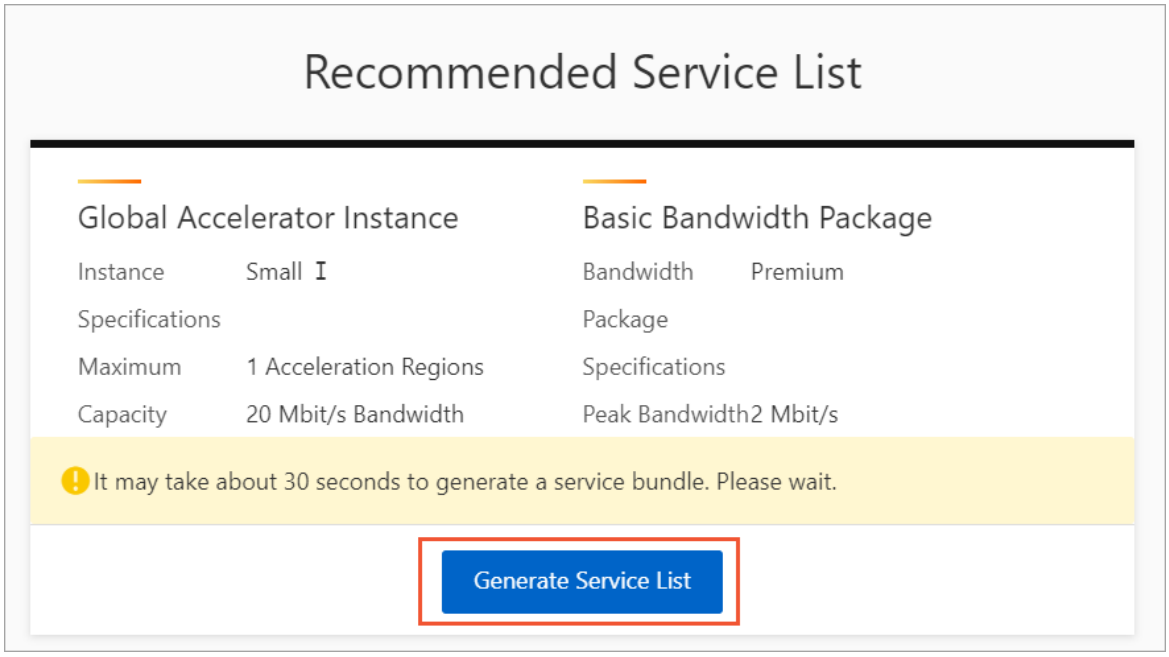
> **Note** If this is the first time that you use the GA service, skip this step.



3. In the **Enter the required information to generate a list of recommended services** section, enter the required information and click **Generate Service List**.

Parameter	Description
Acceleration Area	Select the region that requires acceleration. In this example, **China (Hong Kong)** is selected.
Service Region	Select the region where the backend servers are deployed. In this example, **US (Silicon Valley)** is selected.
ICP Filing	Specify whether you have applied for an Internet Content Provider (ICP) number for the domain name of the web service. In this example, **No** is selected. **Note** All websites must obtain an ICP number before they are allowed to provide services to users in mainland China. For more information, see What is an ICP filing?.
Server Area	Specify whether the web service is deployed on Alibaba Cloud. In this example, **Off Alibaba Cloud** is selected.
Peak Bandwidth Range	Enter the bandwidth required during peak hours. Unit: Mbit/s. In this example, *2* is entered.
Maximum Concurrent Connections	The maximum number of concurrent connections that a GA instance supports. When the number of existing concurrent connections reaches the upper limit, new connection requests are dropped. In this example, **5 Thousand** is selected.

4. In the **Recommended Service List** section, click **Generate Service List** after you confirm the information.



5. On the buy page, set the following parameters and click **Buy Now** to complete the payment.

Parameter	Description
Term	Select the subscription duration.
Specification	Select a specification for the GA instance. In this example, **Small I (Specification Unit)** is selected.
Bandwidth Type	Select a bandwidth type for the basic bandwidth plan. In this example, **Premium** is selected.
Peak Bandwidth	Select the bandwidth limit of the basic bandwidth plan. In this example, **2** Mbit/s is selected.

Step 2: Add an acceleration area

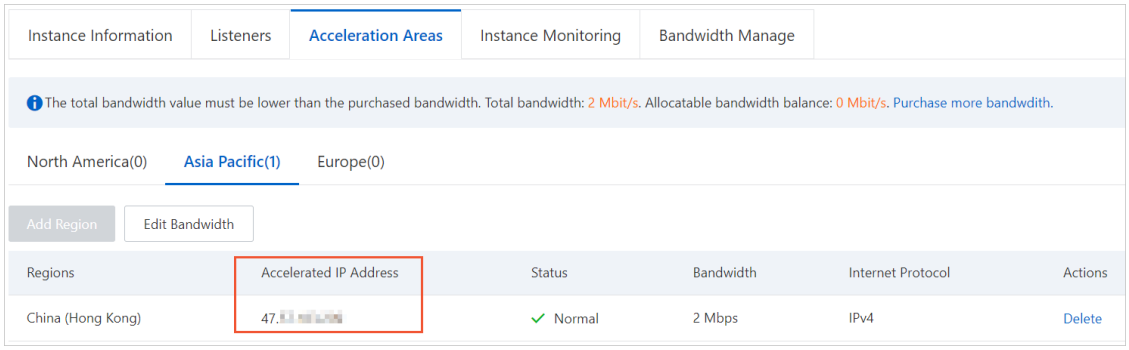
After you purchase a GA instance, you can add an acceleration area, specify the region where users are located, and then allocate bandwidth resources to the region.

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

Parameter	Description
Regions	Select the region where the users are located. In this example, **China (Hong Kong)** is selected.

Parameter	Description
Bandwidth	Specify a bandwidth value for the acceleration region. In this example, *2* Mbit/s is entered.
Internet Protocol	Select the IP address version used to access GA. In this example, **IPv4** is selected.

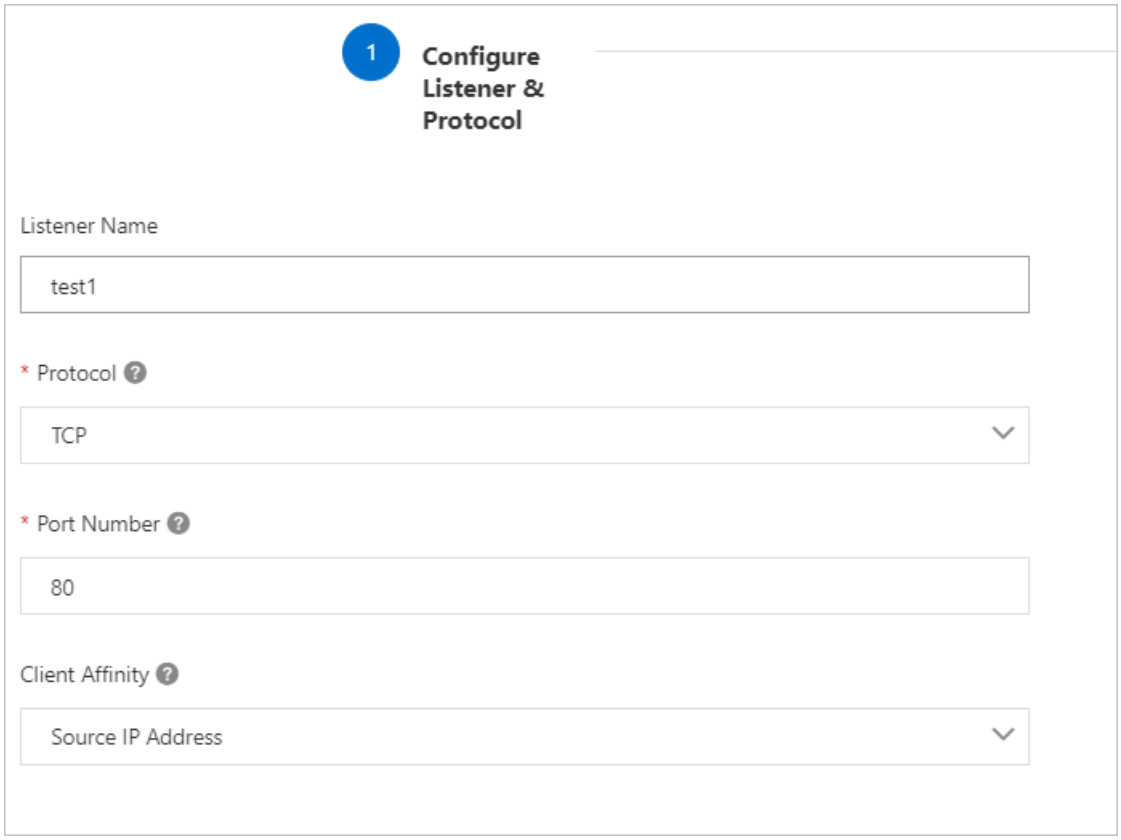
After you add the region, the system assigns an accelerated IP address to the region that is added to the GA instance. This accelerated IP address is used to accelerate data transfer from users in the specified region to the specified backend servers through GA.



Step 3: Add a listener and an endpoint group

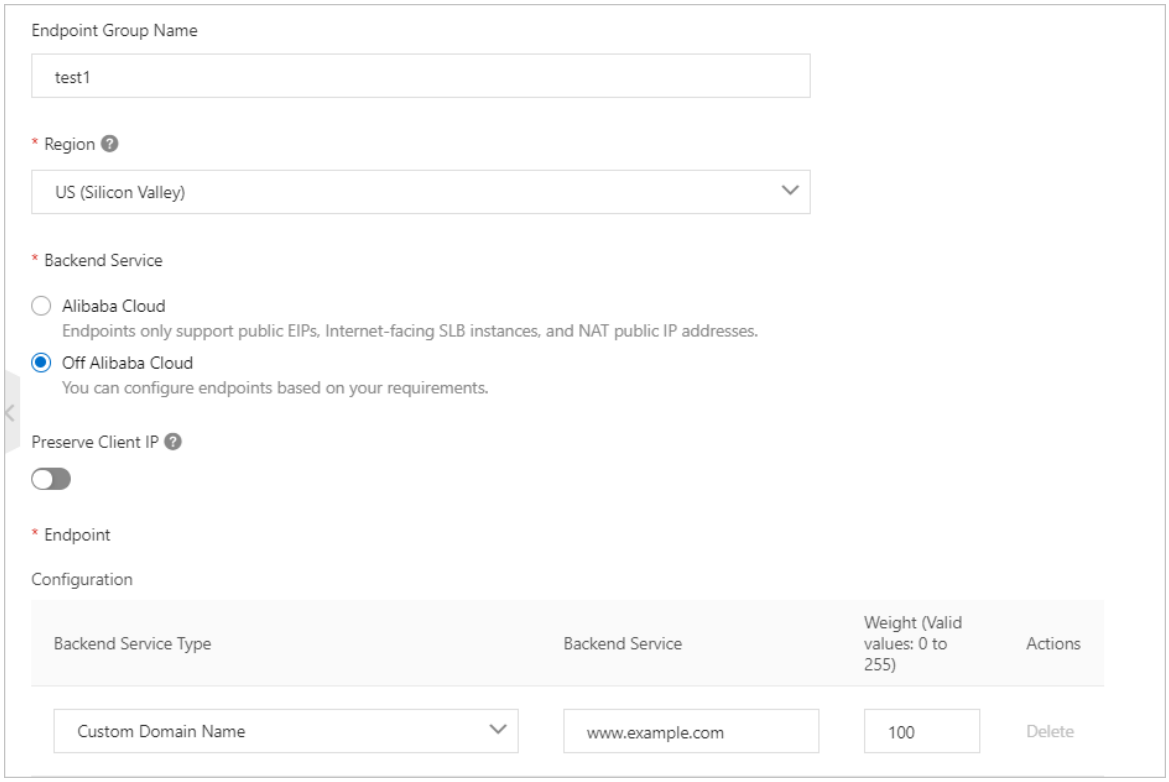
A listener checks for connection requests and then distributes the requests to backend servers based on the specified protocol and ports. Each listener is associated with an endpoint group. You can associate an endpoint group with a listener by specifying the region to which you want to distribute network traffic. After you associate an endpoint group with a listener, traffic is distributed to the optimal endpoint in the associated endpoint group.

1. On the instance details page, click the **Listeners** tab and then click **Add Listener**.
2. 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 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, **Source IP Address** is selected.

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



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 to which the endpoint group belongs, which is the region to which the backend servers belong. In this example, **US (Silicon Valley)** is selected.
Backend Service	Specify whether the web 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 this feature is enabled, backend servers can retrieve client IP addresses. In this example, this feature is disabled.

Parameter	Description
Endpoint	Endpoints are backend servers that receive and handle client requests. Set the following parameters to configure an endpoint: ○ **Backend Service Type**: Select **Custom Domain Name** from the drop-down list. ○ **Backend Service**: Enter the domain name that you want to accelerate. ○ **Weight**: Enter a weight for the endpoint. Valid values: **0 to 255**. GA routes network traffic to the endpoints based on their weights. **Notice** If you set the weight of an endpoint to 0, GA stops routing network traffic to the endpoint. Proceed with caution.

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

Step 4: Add a CNAME record

To enable the acceleration service based on the CNAME assigned to your GA instance, you must configure a CNAME record to map your accelerated domain name to the CNAME.

1. Log on to the Alibaba Cloud DNS console.
2. On the **Manage DNS** page, find the domain name and click **Configure** in the **Actions** column to go to the DNS Settings page.
3. Click **Add Record**, set the following parameters and click **Confirm**.

Parameter	Description
Type	Select **CNAME** from the drop-down list.
Host	Enter the prefix of the accelerated domain name. ○ If the accelerated domain name is `www.aliyun.com`, set the prefix to `www`. ○ If the accelerated domain name is `aliyun.com`, set the prefix to `@`. ○ If the accelerated domain name is `*.aliyun.com`, set the prefix to `*`. ○ If the accelerated domain name is `mail.aliyun.com`, set the prefix to `mail`.
ISP Line	Select **Default** from the drop-down list.
Value	Enter the CNAME value that is allocated to your GA instance. You can find the CNAME value on the **Instances** page.
TTL	Select **10 minute(s)** from the drop-down list.

> **Note**
>
> - New CNAME records immediately take effect. If you modify the CNAME record, the record takes effect within 72 hours after it is modified.
> - After you add a CNAME record, it requires about 10 minutes for the system to update the status in the console. The message "You must add the CNAME record" may appear on the Domain Names page.

Step 5: Verify the acceleration performance

> **Note** If you specify UDP as the protocol when you add a listener, you can verify the acceleration performance by using UDPing. For more information, see Test the acceleration performance a GA instance that uses a UDP listener.

1. Open the CLI on an on-premises machine in the China (Hong Kong) region.
2. Run the following command to query the network latency:

```
curl -o /dev/null -s -w "time_connect: %{time_connect}\ntime_starttransfer: %{time_starttransfer}\ntime_total: %{time_total}\n" ""http[s]://<Domain name of the endpoint>[:<Port>]""
```

In the command:

- time_connect: the period of time that it takes to establish a TCP connection. Unit: seconds.
- time_starttransfer: the start time of data transfer. The start time refers to the amount of time from when the client sends a request to the backend server to when the first byte is sent to the client. Unit: seconds.
- time_total: the total connection time. The total connection time refers to the amount of time from when the client sends a request to when the client receives the last byte from the backend server. Unit: seconds.

4.Access IPv4 services from IPv6 clients by using Global Accelerator

This topic describes how to access backend IPv4 services from IPv6 clients by using Global Accelerator (GA).

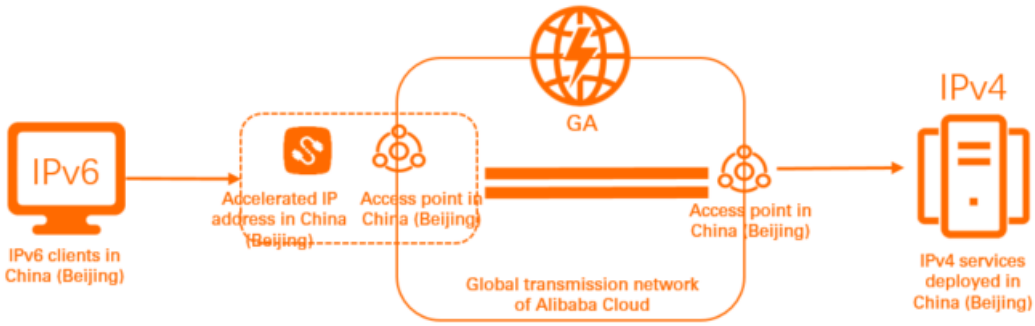
Prerequisites

To connect IPv6 clients to GA, the IPv6 clients must belong to the following regions: China (Beijing), China (Hangzhou), China (Shanghai), China (Shenzhen), and China (Hong Kong). In addition, make sure that the IPv6 clients belong to the preceding regions and are added to the whitelist. To use this feature, submit a ticket.

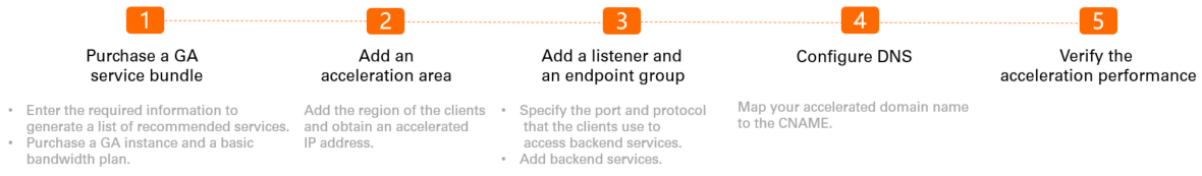
Scenarios

An enterprise has deployed services in data centers that use IPv4 networks. The enterprise needs to provide services to IPv6 clients due to business growth. However, it is complex and time-consuming for the enterprise to enable the current service system to support IPv6 requests. In this scenario, the enterprise can use GA to enable the service system to support IPv6 requests without changing the service system. After GA processes requests from IPv6 clients, the requests are routed to backend IPv4 services. If some enterprises have their services deployed on the cloud, the enterprises can also use GA to enable the services to support requests from IPv6 clients.

The following example describes how to deploy a GA instance to enable IPv4 services to support IPv6 requests. An enterprise in Beijing has deployed the service system in a data center that uses IPv4 networks. The enterprise plans to use GA to provide IPv4 web services to IPv6 clients due to business growth.



Procedure

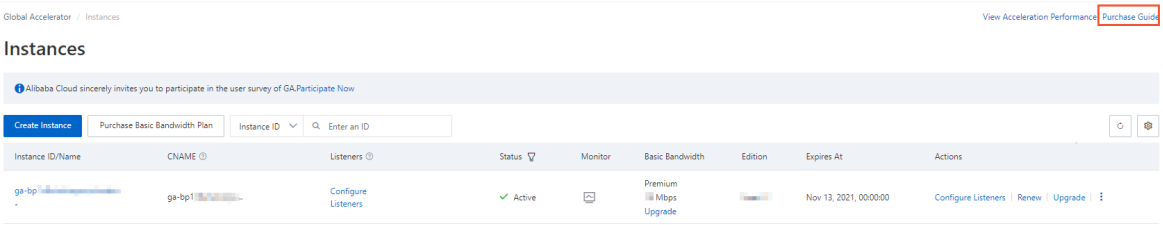


Step 1: Purchase a GA service bundle

You can enter the information about the web service in the GA console. After you enter the information, the system generates a list of recommended services. The list includes a GA instance and a basic bandwidth plan.

1. Log on to the Global Accelerator console.
2. In the upper-right corner of the **Instances** page, click **Purchase Guide**.

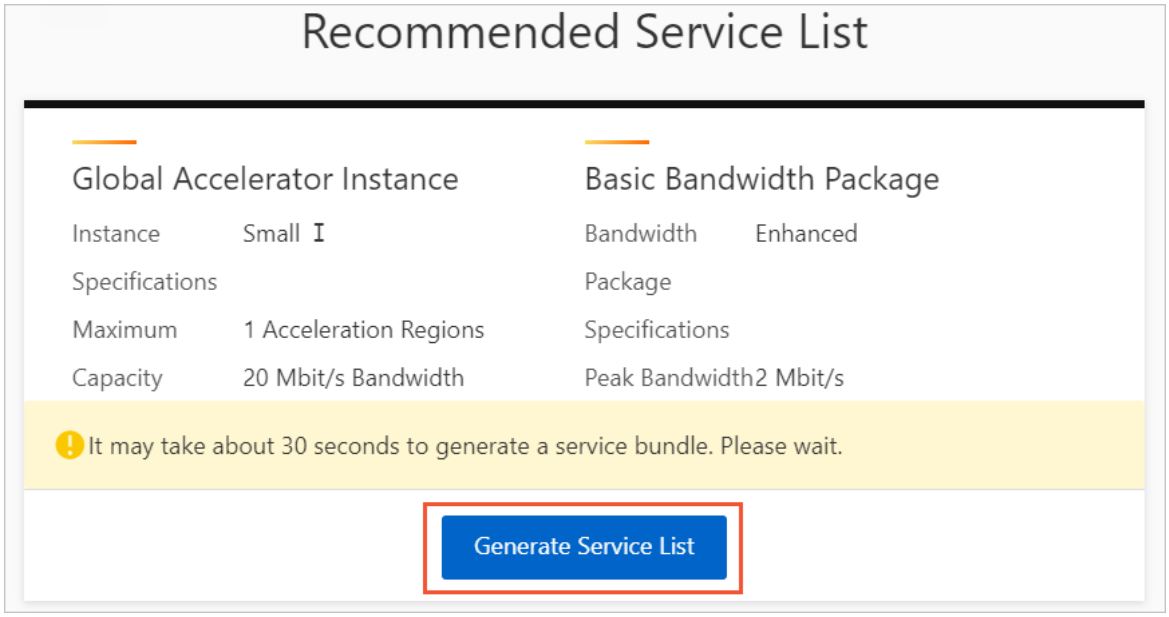
   > **Note** If this is the first time that you use the GA service, skip this step.

3. In the **Enter the required information to generate a list of recommended services** section, enter the required information and click **Generate Service List**.

Parameter	Description
Acceleration Area	Select the region that requires acceleration. **China (Beijing)** is selected in this example.
Service Region	Select the region where the backend servers are deployed. **China (Beijing)** is selected in this example.
ICP Filing	Specify whether you have applied for an Internet Content Provider (ICP) number for the domain name of the web service. **Yes** is selected in this example. **Note** All websites must obtain an ICP number before they are permitted to provide services to users in mainland China. For more information, see What is an ICP filing?.
Server Area	Specify whether the web service is deployed on Alibaba Cloud. In this example, **Off Alibaba Cloud** is selected.
Peak Bandwidth Range	Enter the bandwidth required during peak hours. Unit: Mbit/s. In this example, *2* is entered.
Maximum Concurrent Connections	The maximum number of concurrent connections that a GA instance supports. When the number of existing concurrent connections reaches the upper limit, new connection requests are dropped. In this example, **5 Thousand** is selected.

4. In the **Recommended Service List** section, click **Generate Service List** after you confirm the information.



5. On the buy page, set the following parameters and click **Buy Now** to complete the payment.

Parameter	Description
Term	Select the subscription duration.
Specification	Select a specification for the GA instance. In this example, **Small I (Specification Unit)** is selected.
Bandwidth Type	Select a bandwidth type for the basic bandwidth plan. In this example, **Enhanced** is selected.
Peak Bandwidth	Select the bandwidth cap of the basic bandwidth plan. In this example, **2** Mbit/s is selected.

Step 2: Add an acceleration area

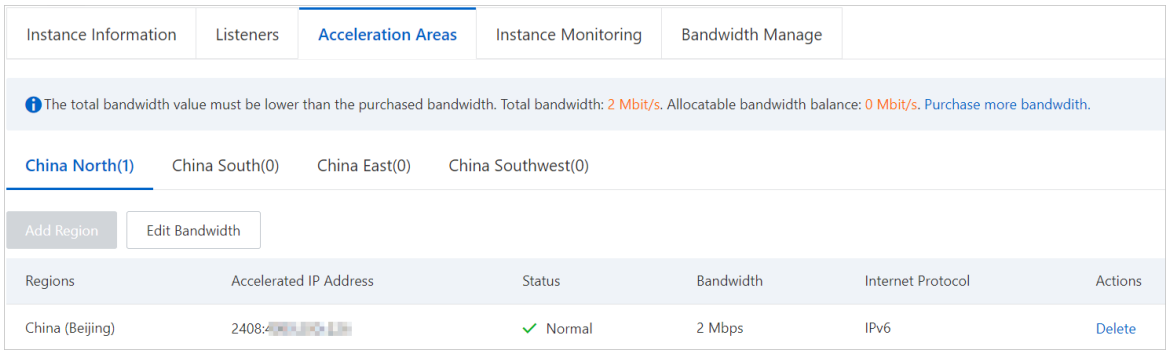
After you purchase a GA instance, you can add an acceleration area, specify the region where users are located, and then allocate bandwidth resources to the region.

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

Parameter	Description
Regions	Select the region where the clients are located. **China (Beijing)** is selected in this example.

Parameter	Description
Bandwidth	Specify a bandwidth value for the acceleration region. In this example, *2* Mbit/s is entered.
Internet Protocol	Select the IP version used by the clients. **IPv6** is selected in this example.

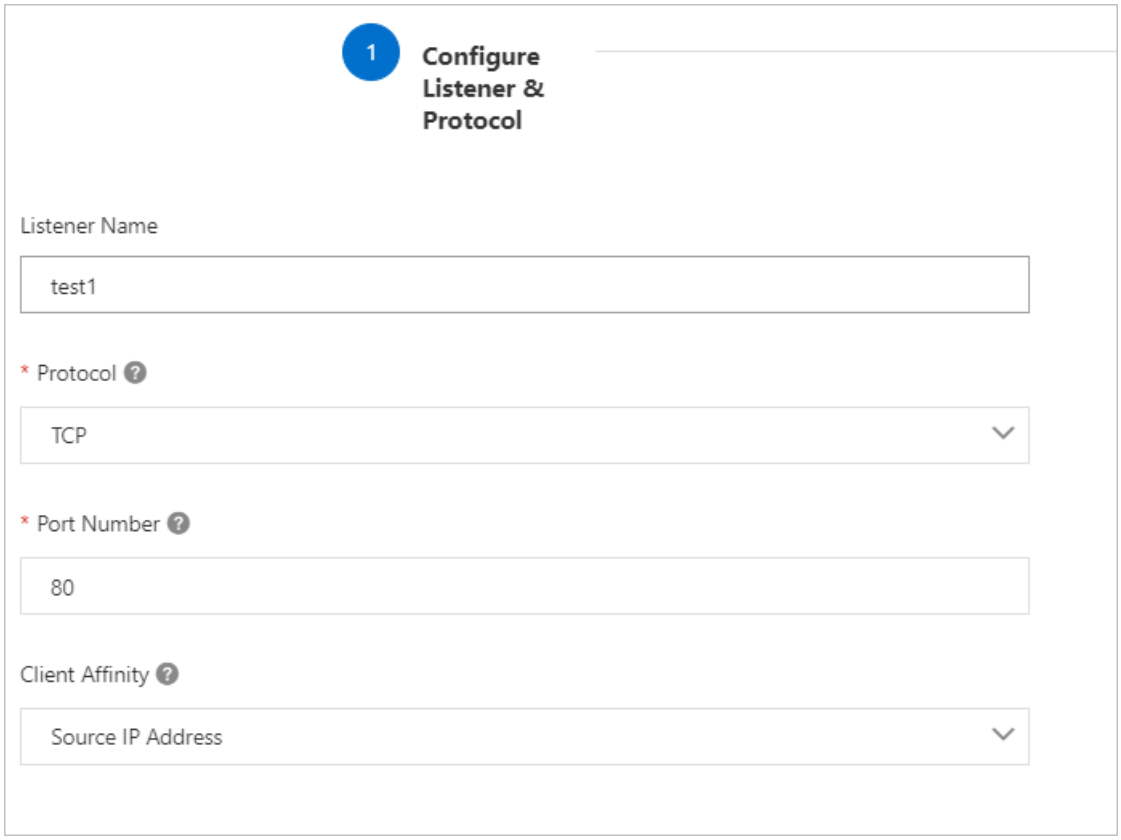
After you add a region, the system allocates an accelerated IPv6 address to the region. The IPv6 address is used to accelerate user access.



Step 3: Add a listener and an endpoint group

A listener checks for connection requests and then distributes the requests to backend servers based on the specified protocol and ports. Each listener is associated with an endpoint group. 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, traffic is distributed to the optimal endpoint in the associated endpoint group.

1. On the instance details page, click the **Listeners** tab and then click **Add Listener**.
2. 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 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, **Source IP Address** is selected.

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

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 (-). It must start with a letter.
Region	Select the region to which the endpoint group (destination servers) belongs. **China (Beijing)** is selected in this example.
Backend Service	Specify whether the backend service is deployed on Alibaba Cloud. **Off Alibaba Cloud** is selected in this example.
Preserve Client IP	Specify whether to preserve client IP addresses. After the feature is enabled, backend servers can retrieve client IP addresses. In this example, this feature is disabled.
Endpoint	An endpoint serves as the destination that a client requests to access. Set the following parameters to configure an endpoint: ○ **Backend Service Type**: Select **Custom IP Address** from the drop-down list. ○ **Backend Service**: Enter the public IP address of the IPv4 web service. ○ **Weight**: Specify a weight for the endpoint. Valid values: 0 to 255. GA routes network traffic to each endpoint in proportion based on the weights of the endpoints. **Notice** If you set the weight of an endpoint to 0, GA does not route network traffic to the endpoint. Proceed with caution.

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

Step 4: Configure DNS

You must configure DNS to route requests destined for the IPv4 service to GA. Select one of the following methods to configure DNS:

- Add an AAAA record that maps the domain name for the IPv4 service to the IPv6 address allocated by GA.
- Add a CNAME record that maps the domain name for the IPv4 service to the CNAME allocated by GA.

1. Log on to the Alibaba Cloud DNS console.
2. On the **Manage DNS** page, find the domain name and click **Configure** in the **Actions** column to go to the DNS Settings page.
3. Click **Add Record**, set the following parameters, and then click **Confirm**.
   - Add an AAAA record

Parameter	Description
Type	Select **AAAA**. The AAAA record is used to map the domain name for the IPv4 service to the IPv6 address allocated by GA.
Host	Enter the prefix of the accelerated domain name. ■ If the domain name is **www.aliyun.com**, set the prefix to **www**. ■ If the domain name is **aliyun.com**, set the prefix to **@**. ■ If the domain name is ***.aliyun.com**, set the prefix to *. ■ If the domain name is **mail.aliyun.com**, set the prefix to **mail**.
ISP Line	Select **Default** from the drop-down list.
Value	Enter the IPv6 address that is allocated by GA for acceleration.
TTL	Select **10 minute(s)** from the drop-down list. The time-to-live (TTL) period of the DNS record on the DNS server.

- Add a CNAME record

Parameter	Description
Type	Select **CNAME** from the drop-down list. The CNAME record is used to map the domain name for the IPv4 service to the CNAME allocated by GA.
Host	Enter the prefix of the accelerated domain name. ■ If the domain name is **www.aliyun.com**, set the prefix to **www**. ■ If the domain name is **aliyun.com**, set the prefix to **@**. ■ If the domain name is ***.aliyun.com**, set the prefix to *. ■ If the domain name is **mail.aliyun.com**, set the prefix to **mail**.
ISP Line	Select **Default** from the drop-down list.
Value	Enter the CNAME that is allocated by GA. You can find the CNAME on the **Instances** page.
TTL	Select **10 minute(s)** from the drop-down list. The TTL period of the DNS record on the DNS server.

> **Note**
>
> - New CNAME records immediately take effect. If you modify the CNAME record, the record takes effect within 72 hours after it is modified.
> - After you add a CNAME record, it requires about 10 minutes for the system to update the status in the console. The message "You must add the CNAME record" may appear on the Domain Names page.

Step 5: Test the connectivity

1. Open the CLI of an IPv6 client in the acceleration region. An IPv6 client in the China (Beijing) region is used in this example.
2. Run the following command to test whether the IPv6 client can access the IPv4 service:

```
curl -6 -g http://[<IPv6 address assigned by GA>]
```

> **Note** In this example, the test is performed on the CentOS operating system. The command that is used to verify the connectivity varies based on the operating system that you use. For more information, see the user guide of your operating system.

The test result shows that the IPv6 client can access the IPv4 web service by using the accelerated IP address.

```
[root@GA-IPV6 ~]# ifconfig
eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 192.1[redacted]68.0.255
        inet6 fe8[redacted]:455f  prefixlen 64  scopeid 0x20<link>
        inet6 240[redacted]ef:ee78:c5ce:b34c  prefixlen 64  scopeid 0x0<global>
        ether 00:16:3e:06:45:5f  txqueuelen 1000  (Ethernet)
        RX packets 448894  bytes 378053918 (360.5 MiB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 172865  bytes 18398232 (17.5 MiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

lo: flags=73<UP,LOOPBACK,RUNNING>  mtu 65536
        inet 127.0.0.1  netmask 255.0.0.0
        inet6 ::1  prefixlen 128  scopeid 0x10<host>
        loop  txqueuelen 1000  (Local Loopback)
        RX packets 2563  bytes 321964 (314.4 KiB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 2563  bytes 321964 (314.4 KiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

[root@GA-IPV6 ~]# curl -6 -g http://[240b[redacted]:1c2]
This is ipv6 access ipv4 web test.
[root@GA-IPV6 ~]#
```

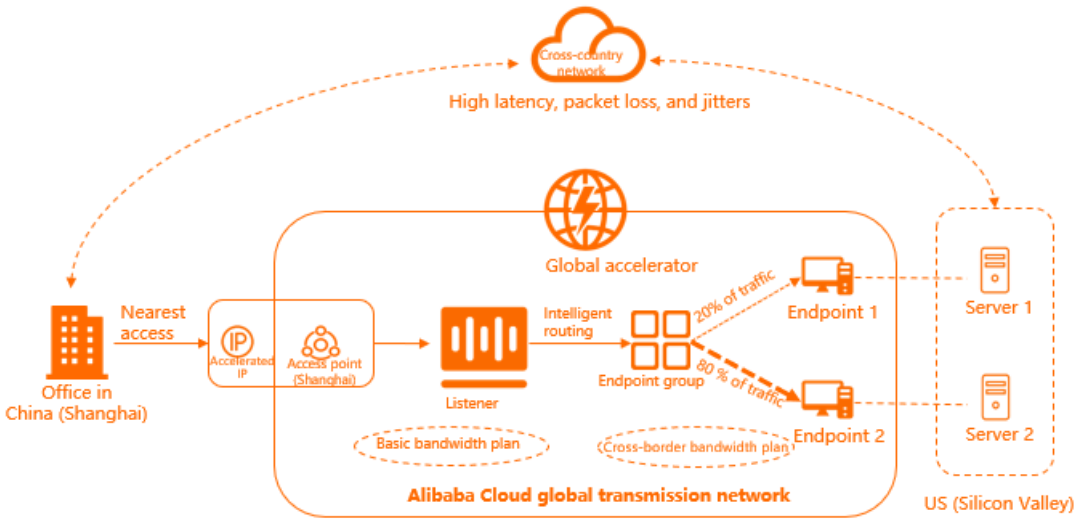
5.Accelerate transmission of network traffic between mainland China and other areas

This topic describes how to use Global Accelerator (GA) to accelerate transmission of network traffic between mainland China and other areas. In the following example, GA is used to accelerate connections from mainland China to applications deployed in other areas.

Context

The following scenario is introduced as an example. A multinational company has built two on-premises origin servers including Server 1 and Server 2 in the US (Silicon Valley) region and deployed enterprise applications on the servers. Server 1 processes only up to 20% of the total workloads. Server 2 processes up to 80% of the total workloads. Unstable network performance may cause issues such as latency, network jitters, and packet loss when users from the office in the China (Shanghai) region connect to the enterprise applications deployed in US (Silicon Valley) over the Internet.

To accelerate connections to the origin servers, you can create a GA instance that provides an access point in China (Shanghai). We recommend that you purchase a cross-border bandwidth plan to optimize acceleration performance. When users in China (Shanghai) send requests to the servers, the access point in China (Shanghai) receives the requests and forwards the requests to the endpoints in US (Silicon Valley) through intelligent routing. The system uses the endpoints to distribute 20% of the requests to Server 1 and 80% of the requests to Server 2.



Procedure

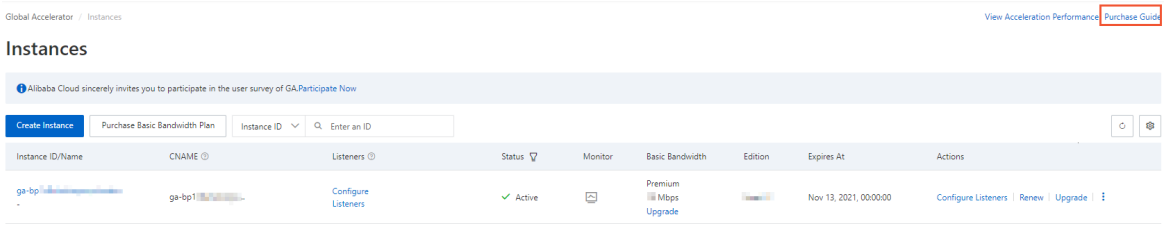
Step 1: Enter the required information about the accelerated application

You can enter the required information about the accelerated application in the GA console. The system recommends the GA instance, basic bandwidth plan, and cross-border bandwidth plan to fit your service needs.

To enter the required information about the accelerated application, follow these steps:

1. Log on to the Global Accelerator console.
2. On the **Instances** page, click **Purchase Guide** in the upper-right corner.

   > **Note** If this is your first time purchasing the GA service, skip this step.

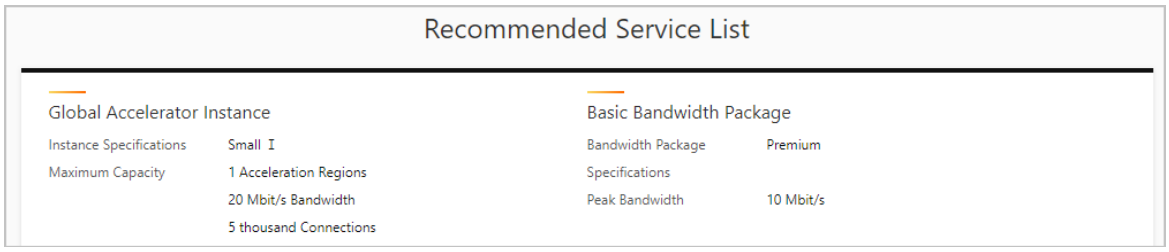
   

3. In the **Enter the required information to generate a list of recommended services** section, enter the required information in the following fields:
   - **Acceleration Area**: Select the area that requires acceleration. **China (Shanghai)** is selected in this example.
   - **Service Region**: Select the region where origin servers are located. **US (Silicon Valley)** is selected in this example.
   - **ICP Filing**: To accelerate a web service, specify whether you have applied for ICP filing. If the service to accelerate is not a web service, select **No**. In this example, **No** is selected.

     > **Note** Therefore, all websites must obtain an ICP number before they are permitted to operate in mainland China. For more information, see What is an ICP filing.

   - **Server Area**: Specify whether the origin servers are deployed on Alibaba Cloud or in the environments outside Alibaba Cloud. **Off Alibaba Cloud** is selected in this example.
   - **Peak Bandwidth Range**: Enter the bandwidth required during peak hours. Unit: Mbit/s. The value is set to **10** in this example.
   - **Maximum Concurrent Connections**: Specify the maximum number of concurrent connections to the GA instance. If the number of existing concurrent connections reaches the upper limit, no extra connections are established. The value is set to **5 thousand** in this example.
4. Click **Generate Service List**.

   You can check the list of recommended services in the Recommended Service List section.

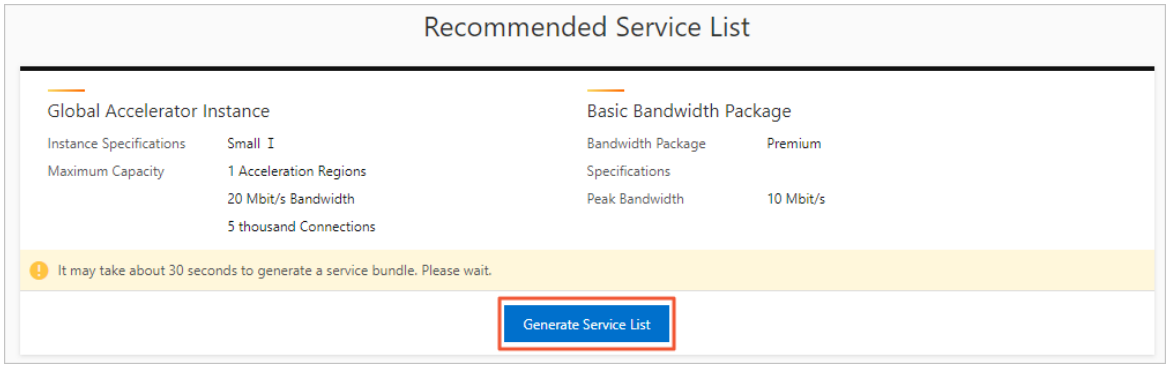
   

Step 2: Purchase a GA service bundle

You can purchase a GA service bundle that includes the recommended one GA instance, basic bandwidth plan, and cross-border bandwidth plan.

Follow these steps:

1. Click **Generate Service List** at the bottom of the Recommended Service List section.



2. On the buy page, set the following parameters to purchase the required GA instance.
   - **Term**: Select a subscription duration.

     > **Note** The subscription duration takes effect on the instance and bandwidth plans specified in this service bundle. For example, if you set Term to 1 Year, the subscription duration of the specified GA instance, basic bandwidth plan, and cross-border bandwidth plan are valid for one year.

   - **Specification**: Select a specification of the GA instance. **Small I** is selected in this example.

     GA supports the following types of instance specifications: Small I, Small II, Small III, Medium I, Medium II, and Medium III. The acceleration performance can vary based on the instance specification.

Instance specification	Number of acceleration regions	Peak bandwidth	Maximum number of concurrent connections
Small I	1	20 Mbit/s	5,000
Small II	2	40 Mbit/s	10,000
Small III	3	60 Mbit/s	15,000
Medium I	5	100 Mbit/s	25,000
Medium II	8	160 Mbit/s	40,000
Medium III	10	200 Mbit/s	50,000

   - **Bandwidth Type**: Select a bandwidth type for the basic bandwidth plan. **Enhanced** is selected in this example.

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

Bandwidth type	Workload type	Accelerated object	Acceleration scope
Basic bandwidth	Applications that are deployed on Alibaba Cloud	■ Elastic Compute Service (ECS) ■ Server Load Balancer (SLB) ■ Alibaba Cloud public IP address **Note** If ECS instances and SLB instances run in classic networks, both types of instances are not supported.	By default, networking within mainland China is accelerated. You can also purchase a cross-border bandwidth plan. This allows you to optimize the acceleration of networking between mainland China and other areas.
Enhanced bandwidth	■ Applications that are deployed on Alibaba Cloud ■ Applications that are not deployed on Alibaba Cloud	■ ECS ■ SLB ■ Alibaba Cloud public IP address ■ Custom IP address ■ Custom domain name	By default, networking within mainland China is accelerated. You can also purchase a cross-border bandwidth plan. This allows you to optimize the acceleration of networking between mainland China and other areas.
Premium bandwidth	■ Applications that are deployed on Alibaba Cloud ■ Applications that are not deployed on Alibaba Cloud	■ ECS ■ SLB ■ Alibaba Cloud public IP address ■ Custom IP address ■ Custom domain name	By default, network connections are accelerated on a global scale. Network traffic transmitted from mainland China to areas outside China is accelerated in the China (Hong Kong) region. If you also purchase a cross-border bandwidth plan, the acceleration of network connections between mainland China and areas outside China are reinforced.

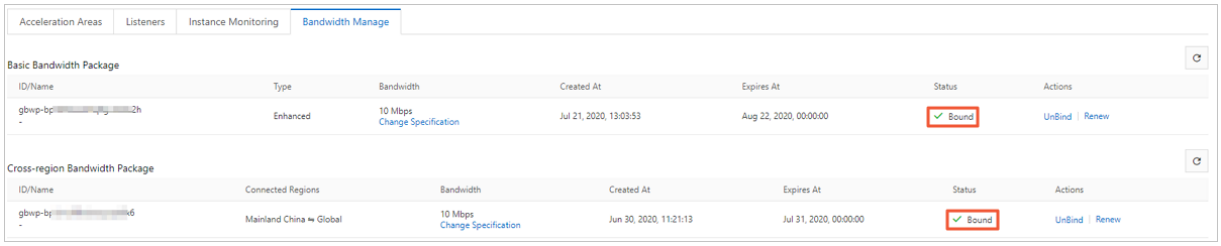
> **Note**
>
> - You can specify ECS or SLB as the backend service type only if your account is added to the whitelist. To specify ECS or SLB as the backend service type, submit a ticket.
> - Only an ECS instance or SLB instance in a virtual private cloud (VPC) can be specified as an endpoint.
> - The IP addresses of endpoint groups associated with each GA instance must be globally unique and not conflict with those of other GA instances.

- **Peak Bandwidth**: Specify a value as the peak bandwidth for the basic bandwidth plan. The value is set to **10Mb** in this example.
- **Area A**: Specify one of the areas to interconnect. **Mainland China** is selected in this example.
- **Area B**: Specify one of the areas to interconnect. **Global** is selected in this example.
- **Bandwidth**: Specify a value as the bandwidth of the cross-border bandwidth plan.

  We recommend that you specify the same bandwidth value for the cross-border acceleration bandwidth plan and the basic bandwidth plan. The value is set to **10Mb** in this example.

3. Click **Buy Now** and complete the payment.

After you purchase the service bundle, the basic bandwidth plan and cross-border bandwidth plan are automatically bound to the GA instance and changed to the **Bound** state.



Step 3: Add an acceleration area

After you purchase a service bundle, you must add one or more acceleration areas where users are located, and allocate bandwidth to these areas.

To specify China East as an acceleration area and China (Shanghai) as a region, follow these steps:

1. Log on to the Global Accelerator console.
2. On the **Instances** page, find the GA instance that you want to manage and click the instance ID to show the details of the instance.
3. On the **Acceleration Areas** tab, click **Add Acceleration Area**.
4. In the **Add Acceleration Area** dialog box, configure the acceleration area and regions.
   - **Acceleration Area**: Select the area where users are located. **China East** is selected in this example.
   - **Regions**: Select the region where users are located and allocate bandwidth to the region.

     In this example, **China (Shanghai)** is selected and **10Mbps** is specified as the bandwidth.
5. Click **OK**.

Step 4: Add a listener

A listener monitors connection requests from clients and forwards the requests based on the specified protocol and port.

To add a listener to the GA instance, follow these steps:

1. On the instance details page, click the **Listeners** tab. Then, click **Add Listener**.
2. On the **Configure Listener & Protocol** wizard, set the following parameters.
   - **Listener Name**: Enter a name for the listener to be created.

     The name must be 2 to 128 characters in length and can contain letters, Chinese characters, digits, underscores (_), and hyphens (-). It must start with a letter or Chinese character.
   - **Protocol**: Select a protocol of the listener from the drop-down list.
     - **TCP**: Transmission Control Protocol. TCP-based communication has the following characteristics:
       - TCP is connection-oriented and reliable. A secure connection must be established with the peer before data can be transmitted.
       - Session persistence is based on source IP addresses.
       - The source IP address is visible at the network layer.
       - Data transmission is slow.
     - **UDP**: User Datagram Protocol. UDP-based communication has the following characteristics:
       - UDP is not connection-oriented and is unreliable. UDP packets are transmitted without three handshakes. Fault tolerance and retransmission are not supported.
       - Data transmission is efficient.

     TCP is selected in this example.
   - **Port Number**: Enter a port or port range for receiving and forwarding requests to the endpoints. Valid values: 1 to 65535.

     The value is set to **80** in this example.
   - **Client Affinity**: Enable or disable client affinity.
     - If you select **Source IP Address** from the drop-down list, client affinity is enabled. Then, requests from a specific client IP address are always forwarded to the same endpoint. The client IP address is regarded as the source IP address.
     - If you select **Disabled**, client affinity is disabled. Then, requests from a specific client IP address are not always forwarded to the same endpoint.

     Disabled is selected in this example.
3. Click **Next**.

Step 5: Configure an endpoint group

Each listener is associated with an endpoint group. You can associate an endpoint group with listeners by specifying the regions to which you want to distribute network traffic. After a listener forwards network traffic to the associated endpoint group based on the association rules, GA distributes network traffic to the optimal endpoints of the endpoint group.

To create an endpoint group, follow these steps:

1. In the **Configure Endpoint Group** wizard, set the following parameters of the endpoint group.

- **Name**: Enter the name of the endpoint.

  The name must be 2 to 128 characters in length and can contain letters, Chinese characters, digits, underscores (_), and hyphens (-). It must start with a letter or Chinese character.

- **Region**: Select the region to be associated with the endpoint group. **US (Silicon Valley)** is selected in this example.
- **Backend Service**: Specify whether the origin servers are deployed on Alibaba Cloud or in the environments outside Alibaba Cloud. **Off Alibaba Cloud** is selected in this example.
- **Reserve Client IP**: Enable or disable the origin servers to reserve client IP addresses in the specified region. After this feature is enabled, the origin servers can obtain the IP addresses of the clients that send requests. In this example, this feature is disabled.

  > **Note** To use client IP address reservation, submit a ticket.

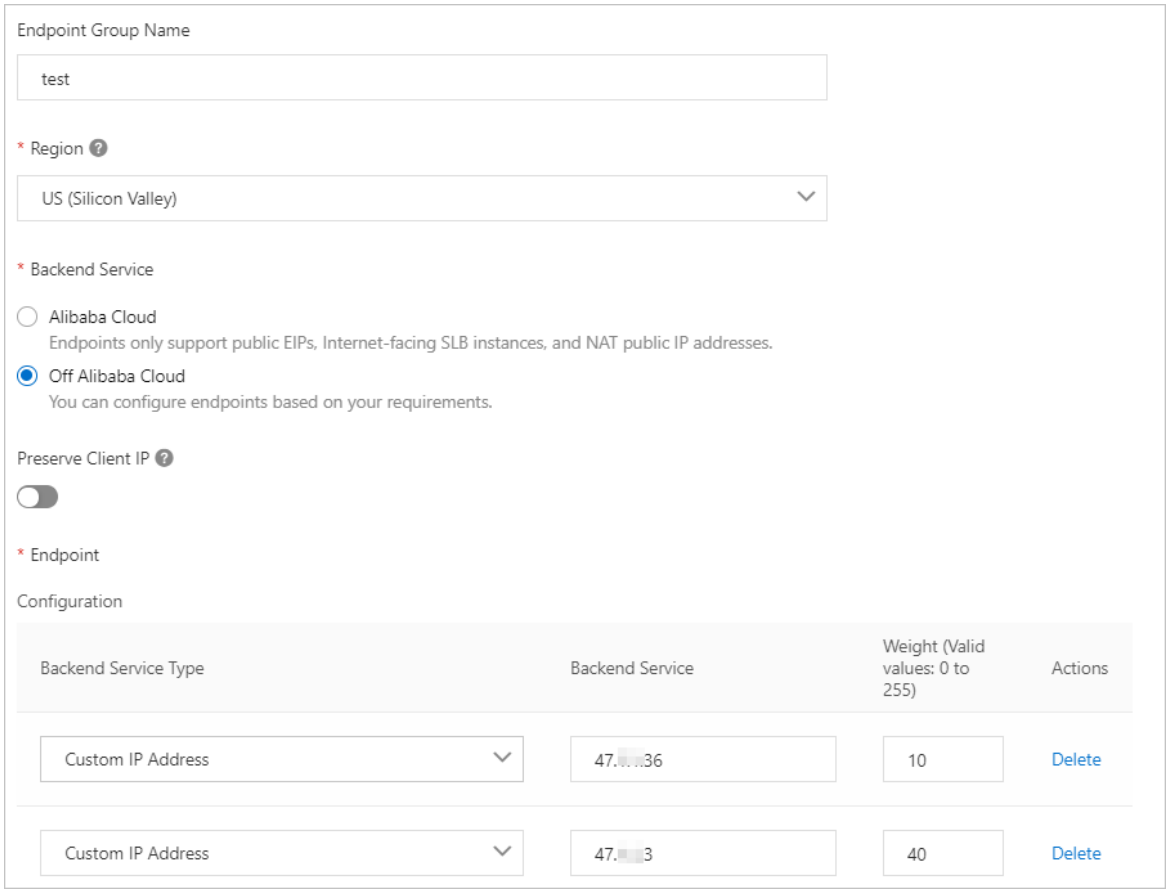
- **Endpoint**: An endpoint is the origin server to which client requests are forwarded. In this example, Server 1 and Server 2 are specified as two endpoints.

  Specify Server 1 as an endpoint:

  - **Backend Service Type**: The Custom IP Address and Custom Domain Name options are available. **Custom IP Address** is selected in this example.
  - **Backend Service**: Enter the public IP address of Server 1.
  - **Weight**: Enter a number from 0 to 255 as the weight for the endpoint. GA forwards network traffic to the endpoint based on the specified weight. In this example, the weight of Server 1 is set to **10**.

  Specify Server 2 as an endpoint:

  - **Backend Service Type**: **Custom IP Address** is selected in this example.
  - **Backend Service**: Enter the public IP address of Server 2.
  - **Weight**: Enter the weight of the endpoint. Valid values: 0 to 255. GA forwards network traffic to the endpoint based on the specified weight. In this example, the weight of Server 2 is set to **40**.



2. Click **Next**.
3. In the **Check Configurations** wizard, check the configurations and click **Next**.

Step 6: Test the acceleration of TCP traffic

To test the acceleration of TCP traffic, follow these steps:

1. Open Command Prompt on the client located in China (Shanghai).
2. Run the following commands to check the latency of data packet transmission.

   ```
   curl -o /dev/null -s -w "time_connect: %{time_connect}\ntime_starttransfer: %{time_starttransfer}\ntime_total: %{time_total}\n" "http[s]://<IP address of the origin server>[:<Listening port>]"
   ```

   Where:

   - time_connect: the period between the time when a request is sent and when a TCP connection is established.
   - time_starttransfer: the period between the time when a request is sent and when the client receives the first byte returned by the origin server. After this period, the system starts to transmit data packets.
   - time_total: the period between the time when a request is sent and when the client receives the last byte returned by the origin server. This is the total time consumed for the communication between the client and the origin server.

   The following figures show that GA has reduced the latency of the application.

   Connection performance before GA is used

```
C:\Users\25513>curl -o /dev/null -s -w "time_connect: %{time_connect}\ntime_starttra
nsfer: %{time_starttransfer}\ntime_total: %{time_total}\n" "47.[redacted]"
time_connect: 3.172000
time_starttransfer: 3.344000
time_total: 3.344000
```

Connection performance after GA is used

```
C:\Users\25513>curl -o /dev/null -s -w "time_connect: %{time_connect}\ntime_starttra
nsfer: %{time_starttransfer}\ntime_total: %{time_total}\n" "47.[redacted]"
time_connect: 0.203000
time_starttransfer: 0.422000
time_total: 0.422000
```

Note If you have added a UDP listener to your GA instance, you can use UDPing to test the acceleration of UDP traffic. For more information, see Test the acceleration performance a GA instance that uses a UDP listener.