API Gateway

User Guide for Providers

User Guide for Providers

API Gateway provides high-performance and high-availability API hosting service to help users to publish or access to the APIs on Alibaba Cloud products such as ECS and Container Service. It manages the entire API lifecycle from release and management to maintenance. You can quickly open data or services at low costs and risks through simple operations.

API Gateway provides the following features:

API management

You can manage the lifecycle of an API, including creation, testing, release, deprecation, and version switching.

Easy data conversion

You can configure a mapping rule to convert the calling request into the format required by the backend.

Presetting of request verification

You can preset the verification of the parameter type and values (range, enumeration, regular expression, and JSON Schema) for gateway to preclude the illegal requests, reduce the utilization rate of your backend.

Flexible traffic control

You can set traffic control for APIs, users, and APPs by minute, hour, or day.

In addition, you can also specialize some users or APPs with the independent traffic control.

Easy security protection

API Gateway supports AppKey authentication and HMAC (SHA-1,SHA-256) signature.

API Gateway supports SSL/TSL encryption and uses Alibaba Cloud Security to prevent viruses and attacks.

Comprehensive monitoring and warning

API Gateway provides visualized API monitoring in real time, including the calling traffic, calling method, response time, and error rate, and supports query of historical records for comprehensive analysis. You can also configure and subscribe to the warning method (SMS or email) to check the API running status in real time.

Lower cost of publication

API Gateway automatically generates API documentation and SDKs (service end and mobile end), reducing the cost of publication of API.

Create an API

API creation is a process to define an API request. When creating an API, you need to define the format of API call requests, the format of requests sent from the gateway to backend services, the format of returned results, the parameter verification rules and so on.

To Define Basic Request Information

Basic API information includes the API group, API name, description and API type.

1. Select an API group when creating an API. An API group is a management unit of APIs with a corresponding region and domain name (For details about the API group, refer to **Group and Domain Name** described below). APIs in an API group share the same region and domain name. Once selected, the group cannot be changed.
2. The API name shall be unique in the group and cannot be changed.
3. There are two types of APIs: public and private. Both of them share no substantial difference at the public beta stage.

Define Backend Service Information of the Request

Information of API backend services includes the type, address and timeout time of backend services.

1. Backend service type. It only supports the HTTP service at present, and will support Sigma, Mock and other types of services in the future.
2. Backend service address. It refers to the complete IP address used by the API Gateway to call underlying services, which includes a domain name/IP+Path without Query parameter. It may contain dynamic parameters, such as username (written as **username**), could be obtained only through the path input by the caller. Therefore, do not omit these dynamic parameters when defining the final path.
3. Backend timeout time. It refers to the response time for beckend service to return the results after receiving the requests from gateway. The timeout time shall not exceed 30 seconds.

Define the Format of API Request

The API request format definition includes protocol and method definition, path definition, input parameter definition, system parameter definition, parameter mapping, and parameter verification definition.

Protocol and method definition. HTTP/HTTPS protocols are supported for API calling. Methods include PUT, GET, POST, DELETE, HEAD and MULITIPART.

Path definition. It refers to the path used by the caller to call the API available to external resources. The gateway stores the corresponding relations and locates the corresponding paths. The path may differ from that in the backend service address. You have to map the parameters when defining the path if they are in the backend service address.

Input parameter definition. The parameters to input conprise header, query and body. You need to define the **name** of the input parameter of the user request; choose the **required parameters**; and provide the **example value**, **default value** and **description**. The types of parameters include **String**, **Number**, **Boolean** and **JSON**. The transmission mode of the body parameter may be **transparent transmission**.

Parameter verification definition. When defining the input parameters, you can click **More** to set verification for the parameter, including verification of the **enumeration value**, the **string length**, and **the maximum and minimum values of the number**. The gateway will intercept illegal requests, relieving burdens of your backend services.

Parameter mapping. To guise your auctual parameter name of your backend service, you can configure a backend parameter mapping for each parameter when defining the parameter.

System parameter definition. System parameters are invisible to API callers. There are two types of system parameter. One transmitted by gateway to you is described in the table below:

Name	Meaning
CaClientIp	IP address of the client sending the request
CaDomain	Domain name sending the request
CaRequestHandleTime	Request time (Greenwich mean time)
CaAppId	ID of the APP sending the request
CaRequestId	RequestId
CaApiName	API name

CaHttpSchema	Protocol used by the user to call the API, which may be HTTP or HTTPS
CaProxy	Proxy (AliCloudApiGateway)

When creating an API, you need to configure this system parameter and select the **parameter position** and **backend parameter name**.

The other type is custom system parameter required by your API backend service, may be a constant parameter. The configuration includes the **parameter value**, **backend parameter name** and the **parameter’ s position** in the request.

Returned result definition. It refers to the type and example of returned results. Currently, the gateway does not process returned results.

NOTE: You shall enter the following parameters: the **dynamic parameters** in the path, **headers parameter**, **query parameter**, **body parameter** (non-binary), **constant parameter**, and **system parameter**. The parameter name shall be globally unique. It is not allowed to enter a parameter named “name” in the headers and query at the same time.

So far, you have created an API. Next, you can test and release API, authorize permissions to your customers, bind a signature key and traffic control policy to the API, and perform other security configurations.

Enable API services

This section provides information that you need to understand for the API group and domain name before enabling API services.

API group

An API group is the management unit of APIs. You need to create a group before creating an API. The group consists of four attributes: name, description, region, and domain name.

The group region is fixed once selected.

Each account can have up to 50 API groups and each API group can have up to 200 APIs.

3. When you create a group, the system assigns the group a second-level domain name to test your API. To enable the API service, you need to bind the group to an independent domain name filed on Alibaba Cloud’ s system and resolve the CNAME of the independent

domain name to the group’ s second-level domain name. At most five independent domain names can be bound to a group.

Domain name and certificate

API Gateway locates the unique API group through the domain name, and the unique API through the Path+HTTPMethod. Before enabling API services, you need to know the second-level domain name and independent domain name.

1. The unique and fixed second-level domain name is assigned by the system during group creation. By default, a second-level domain name is used to call the API only in the test environment under a small amount of traffic.

   An independent domain name is used for enabling API services. You can bind up to five independent domain names to a group. When configuring independent domain names, pay attention to the following points:

   a. Independent domain names are only available when they are filed on the Alibaba Cloud filing system or a system connected to the Alibaba Cloud filing system. For more information, refer to Filing and Connection Process.

   b. Resolve the CNAME of an independent domain name to the API group’ s second-level domain name before binding the API group and domain name.

   c. Verify the domain name within one day. Otherwise, the unprocessed binding request will be automatically withdrawn by the system.

   d. If a domain name is already bound to another group, resolve the domain name to the to-be-bound group’ s second-level domain name before binding. Otherwise, the binding will fail.

   If your API supports the HTTPS protocol, you need to upload the SSL certificate of the domain name by filling in the parameters. These parameters include the name, content, and private key on the **Group Details** page.

Test, production, and authorization

To test or enable the API, authorization is indispensable. Authorization means granting an app the permission to call an API.

1. You can authorize the created app and access the second-level domain name to call the API.
2. You can authorize your customers’ apps to access the independent domain name to call your API service.

3. Only an authorized app can call the API.

Now you have successfully enabled your API service. In the process from API creation to enabling, you can create, modify, delete, view, test, release, or deprecate an API, authorize an API or revoke the authorization, view the release history, and switch the version.

API definitions refer to the definitions related to the API request structure when you create an API. You can view, edit, delete, create, or copy an API definition on the console. Pay attention to the following points when you are working with API definitions:

1. Editing the definition of a released API does not affect the definition in the production environment unless you release and synchronize it to the production environment.
2. It is not allowed to directly delete the API definition. Deprecate the API definition before deleting it.
3. You can copy the definition from the test/production environment to overwrite the latest definition, and then, if needed, click **Edit** to modify the definition.

API release management

You can release or deprecate an API in a test or production environment with the attentions below:

1. You can access the second-level domain name or independent domain name to call the API that is released to the test or production environment.
2. The latest released version of an API overwrites the preceding version in the test/production environment and takes effect in real time.
3. When you deprecate an API in the test/production environment, the binding policy, keys, app, and authorization persists are automatically deprecated unless the API is released to production again. To revoke this relationship, you need to delete it.

API authorization management

You can establish or revoke the authorization relationship between an API and an app. API Gateway verifies the permission relationship. During authorization, pay attention to the following points:

1. You can authorize one or more APIs to one or more apps. We recommend that you do not operate APIs in multiple groups at the same time during batch operation.
2. During batch operation, select an API and related environment. For example, if an API has been released to both the test and production environments, but only the test environment is chosen, only the API in the test environment will be authorized.
3. You can locate an app based on the AppID or Alibaba Mail account provided by the customer.
4. When you need to revoke the authorization for an app under an API, you can view the API authorization list and delete the app from the list.

Release history and version switching

You can view the release history of each of you APIs, including the version number, notes, test/production, and time of each release.

When viewing the release history, you can select a version and switch to it. The new version directly overwrites the previous one and takes effect in real time.

What Is a Signature Key

A signature key is the Key-Secret pair you create, based on which the backend service verifies the request received from the gateway. Pay attention to the following points:

1. An unchangable region shall be selected during key creation. The key can only be bound to APIs in the same region.
2. One API can be bound with only one key. The key can be replaced, modified, bound to, or unbound from the API.
3. After binding a key to an API, the signature information will be added to all the requests sent from the gateway to the API at your service backend. You need to resolve the signature information through symmetric calculation at the backend to verify the gateway’ s identity. For details about adding signature to the HTTP service, refer to Backend HTTP Service Signature.

Modify or Replace the Leaked Key

To modify the Key-Secret pair once a key is leaked or to substitute a key bound to an API with another key, proceed the following steps:

1. Configure the backend to support two keys: the original key and to-be-modified or replaced key, so that the request during the switching process can pass signature verification regardless the key modification or replacement.
2. After the backend is configured, modify the key. Verify that the new Key and Secret take effect and delete the leaked or obsolete key.

What Is Traffic Control Policy

You can set traffic control for APIs, users, and APPs by minute, hour, or day; or sort out the specific users or APPs with designated traffic control policy. The traffic control policy is described below:

API traffic limit	The call times within per time unit for the API bound by the policy shall not exceed the set value. The time unit may be minute, hour or day, for example, 5,000

	times per minute.
APP traffic limit	The call times called by each APP within per time unit for an API bound to the policy shall not exceed the set value, for example, 50,000 times per hour.
User traffic limit	The call times called by each Alibaba Cloud account within per time unit shall not exceed the set value. An Alibaba Cloud account may have multiple APPs. The traffic limit for an Alibaba Cloud account refers to the limit on the total traffic of all APPs in this account. For example, the traffic may be 500,000 times per day.

In addition, you can set an additional threshold value as the traffic limit value （ not allowed to exceed the value of **API traffic limit** ）for the special APP or user. However, the **basic APP traffic limit** and **user traffic limit settings** in the traffic control policy are no longer applicable to the special APP or user.

An unchangable region shall be selected for the traffic control policy, and the traffic control policy can only be applied to APIs in the same region.

The traffic of a single IP address is restricted within 100 QPS regarding with the value of **API traffic limit**.

A traffic control policy can be bound to multiple APIs, with the limit value and special object settings appliable to each API separately. The lattest policy bound to the API will overwrite the previous one and take effect forthwith.

To add a special APP or user, you need to obtain the APP’ s ID (AppID) or Alibaba Mail account.

On the API Gateway console, you can create, modify, delete, view, bind, and unbind a traffic control policy.

- The API Gateway console provides visualized API monitoring and warning in real time. You can obtain the calling status of an API, including the calling traffic, calling method, response time, and error rate. API Gateway displays data statistics on the calling status from multiple dimensions in multiple time units, and supports query of historical data for comprehensive analysis.

- You can also configure the warning method (SMS or email) and subscribe to warning information to know the API running status in real time.

Limits on API Gateway products and business.

Restrictions	Description
User restrictions on activating the API Gateway service.	To activate the service, you need to complete the real-name authentication.
Restrictions on the number of API groups created by a user.	Each account can have at most 50 API groups.
Restrictions on the number of APIs created by a user.	At most 200 APIs can be created in each API group. That is, at most 10,000 (50 * 200) APIs can be created in each account.
Restrictions on the number of independent domain names bound to an API group.	At most five independent domain names can be bound to a group.
Restrictions on the traffic for calling an API.	The traffic of a single IP address of a single user used for calling each API made available by you shall not exceed 100 QPS.
The limit of the official subdomain.	When the API group is created successfully, the API gateway will issue a secondary domain name for that group. You can test the API in the group by accessing the domain name, and the gateway restricts the number of visits to 1000 times per day. Please do not use the secondary domain name to provide API service directly.
Restrictions on parameter size.	The parameters of the body location (including Form and Form other forms) can not exceed 2 Mb, and other locations (including Header and Query) can not exceed 128 Kb.

Overview

API Gateway provides the backend HTTP service signature verification function. To enable backend signature, you need to create a signature key and bind the key to the corresponding API. (keep this key properly. API Gateway encrypts and stores the key to ensure the security of the key.) After backend signature is enabled, API Gateway will add signature information to the request destined to the backend HTTP service. The backend HTTP service reads the signature string of API Gateway and performs local signature calculation on the received request to check whether the gateway signature and local signature result are consistent.

All the parameters you have defined will be added to the signature, including the service parameters

you have entered, and constant system parameters and API Gateway system parameters (such as CaClientIp) you have defined.

How to Read the API Gateway Signature

- Save the signature calculated by the gateway in the header of the request. The Header name is X-Ca-Signature.

How to Add a Signature at the Backend HTTP Service

For details about the demo (Java) of signature calculation, refer to https://github.com/aliyun/api-gateway-demo-sign-backend-java.

The signature calculation procedure is as follows:

Organize Data Involved in Signature Adding

```
String stringToSign=
HTTPMethod + "\n" + // All letters in the HTTPMethod should be capitalized.
Content-MD5 + "\n" + // Check whether Content-MD5 is empty. If yes, add a linefeed "\n".
Headers + // If Headers is empty, "\n" is not required. The specified Headers includes "\n". For details, refer to the headers organization method described below.
Url
```

Calculate the Signature

```
Mac hmacSha256 = Mac.getInstance("HmacSHA256");
byte[] keyBytes = secret.getBytes("UTF-8");
hmacSha256.init(new SecretKeySpec(keyBytes, 0, keyBytes.length, "HmacSHA256"));
String sign = new String(Base64.encodeBase64(Sha256.doFinal(stringToSign.getBytes("UTF-8")),"UTF-8"));
```

secret is the signature key bound to an API.

Description

Content-MD5

Content-MD5 indicates the MD5 value of the body. MD5 is calculated only when HTTPMethod is **PUT**

or **POST** and the body is not a form. The calculation method is as follows:

```
String content-MD5 = Base64.encodeBase64(MD5(bodyStream.getbytes("UTF-8")));
```

Headers

Headers indicates the keys and values of the headers involved in signature calculation. Read the keys of all headers involved in signature calculation from the header of the request. The key is X-Ca-Proxy-Signature-Headers. Multiple keys are separated by commas.

Headers Organization Method

Rank the keys of all headers involved in signature calculation in lexicographic order, and change all uppercase letters in the key of the header to lowercase, and splice the keys in the following method:

```
String headers =
HeaderKey1.toLowerCase() + ":" + HeaderValue1 + "\n"\+
HeaderKey2.toLowerCase() + ":" + HeaderValue2 + "\n"\+
...
HeaderKeyN.toLowerCase() + ":" + HeaderValueN + "\n"
```

Url

URL indicates the Form parameter in the Path + Query + Body. The organization method is as follows:If Query or Form is not empty, add a **?**, rank the keys of Query+Form in lexicographic order, and then splice them in the following method. If Query or Form is empty, then URL is equal to Path.

```
String url =
Path +
"?" +
Key1 + "=" + Value1 +
"&" + Key2 + "=" + Value2 +
...
"&" + KeyN + "=" + ValueN
```

Note that Query or Form may have multiple values. If there are multiple values, use the first value for signature calculation.

Debug Mode

To access and debug the backend signature conveniently, you can enable the Debug mode. The debugging procedure is as follows:

Add **X-Ca-Request-Mode = debug** to the **header** of the request destined to API Gateway.

The backend service can just read **X-Ca-Proxy-Signature-String-To-Sign** from the **header** because the linefeed is not allowed in the HTTP Header and thereby is replaced with "|".

NOTE: **X-Ca-Proxy-Signature-String-To-Sign** is not involved in backend signature calculation.

Verify the Time Stamp

When the backend verifies the time stamp of the request, the system parameter **CaRequestHandleTime** is selectable in API definition and its value is the Greenwich mean time when the gateway receives the request.