What Is Service Quotas?

Service Quotas enables you to view and manage your quotas for AWS services from a central location. Quotas, also referred to as limits, are the maximum values for the resources, actions, and items in your AWS account. Each AWS service defines its quotas and establishes default values for those quotas. Depending on your business needs, you might need to increase your service quota values. Service Quotas makes it easy to quickly look up your available service quotas and to request increases.

Contents

- Service Quotas Features (p. 1)
- Introducing the Service Quotas Components (p. 2)
- Accessing Service Quotas (p. 2)
- Key Terms and Concepts in Service Quotas (p. 2)

Service Quotas Features

The following features are available.

View AWS Service Quotas

The Service Quotas console provides quick access to the AWS default quota values for your account, across all commercial Regions. When you select a service in the Service Quotas console, you'll see the quotas and whether the quota is adjustable. Applied quotas are overrides, or increases for a particular quota, over the AWS default value.

Request a Service Quota Increase

For any adjustable service quotas, you can use Service Quotas to request a quota increase. To request a quota increase, in the console simply select the service and the specific quota, and choose Request quota increase. You can also use the API or command line interface (CLI) tools to request service quota increases.

View Current Utilization

If your account has been active a while and a resource has been used, you can view a graph of your quota utilization.

Set Amazon CloudWatch Alarms for Approaching Quotas

For supported services, you can manage your quotas by configuring CloudWatch alarms to monitor usage and alert you to approaching quotas.

Control Who Manages Service Quotas

You can attach AWS Identity and Access Management (IAM) permission policies to your users, groups, and roles that grant or deny permission to manage the service quotas in your AWS account. For example, you could create a "quota administrator" who can view, manage, or request increases for any service
quotas in your account. You could also attach one policy to a group whose members need the ability to manage the quotas for a specific service.

Introducing the Service Quotas Components

For a list of terms and concepts that you need to understand to make full use of Service Quotas, see Key Terms and Concepts in Service Quotas (p. 2).

Accessing Service Quotas

You can work with Service Quotas in the following ways:

AWS Management Console

The Service Quotas console is a browser-based interface that you can use to view and manage your service quotas. You can perform almost any task that's related to your service quotas by using the console. You can access Service Quotas from any AWS console page by choosing on the top navigation bar, or by searching for Service Quotas in the AWS Management Console.

AWS Command Line Tools

The AWS command line tools let you issue commands at your system's command line to perform Service Quotas and other AWS tasks. This can be faster and more convenient than using the console. The command line tools also are useful if you want to build scripts that perform AWS tasks.

AWS provides two sets of command line tools: the AWS Command Line Interface (AWS CLI) and the AWS Tools for Windows PowerShell. For information about installing and using the AWS CLI, see the AWS Command Line Interface User Guide. For information about installing and using the Tools for Windows PowerShell, see the AWS Tools for Windows PowerShell User Guide.

AWS SDKs

The AWS SDKs consist of libraries and sample code for various programming languages and platforms (for example, Java, Python, Ruby, .NET, iOS and Android, and others). The SDKs include tasks such as cryptographically signing requests, managing errors, and retrying requests automatically. For more information about the AWS SDKs, including how to download and install them, see Tools for Amazon Web Services.

Key Terms and Concepts in Service Quotas

The following terms and concepts are important for understanding Service Quotas and how it works. In AWS, the terms quota and limit can be used interchangeably.

Service Quota

In AWS, a service quota is the maximum number of service resources or operations that apply to an account, although some service quotas apply to a Region. An example of an account-based quota is the number of AWS IAM roles per account or AWS Elastic Beanstalk applications per account. An example of a Region-based quota is the number of AWS AppSync APIs per Region. Check the service quota description to see if the limit is Region-specific.

Adjustable Value

An adjustable value is a service quota value that can be increased.
Global Quota

A global quota is a quota, or limit, applied at an account level. Global quotas are available in all Regions. You can request an increase to a global quota from any Region, and can track the status of the increase from the Region where the increase was requested. If a quota increase, for a global quota, has been requested, you can't request an increase for the same quota from a different Region until the first request has completed. Once the initial request has completed, the applied quota value will be visible across all Regions (if applied quotas are available).

Default Value

A default value is the initial service quota value established by AWS.

Applied Value

An applied value is the new or latest service quota value, once the default service quota value has been increased.

Usage

In Service Quotas, usage is the number of resources or operations in use, for that service quota.

Utilization

In Service Quotas, utilization is the percentage of the service quota in use. For example, if the service quota is 200 resources, and 150 resources are in use, the utilization is 75%.
Getting Started with Service Quotas

When you open the Service Quotas console, the dashboard displays cards for up to nine services. Each card lists the number of service quotas for the service. Choosing a card opens a page that displays the quotas for the service. You can choose which services appear on the dashboard.

To modify the dashboard service cards

2. On the dashboard, choose Modify dashboard cards.
3. The services that are currently selected appear on the right. If you have selected nine services, you must remove a service before you can add a different service. For each service that you don't need on the dashboard, choose Remove.
4. To add a service to the dashboard, select it from Choose services.
5. When you have finished adding and removing services, choose Save.

Next Steps

- Viewing a Service Quota (p. 5)
- Requesting a Quota Increase (p. 6)
Viewing a Service Quota

Service Quotas makes it easy to look up the value of a particular quota, also referred to as a limit. You can also look up all quotas for a particular service.

To view the quotas for a service

2. In the navigation pane, choose AWS services.
3. Select a service from the list, or type the name of the service in the search field. For each quota, the console displays its name, applied value, default value, and whether the quota is adjustable. If the applied value is not available, the console displays a dash.
4. To view additional information about a quota, such as its description and Amazon Resource Name (ARN), choose the quota name.
Requesting a Quota Increase

You can use Service Quotas to request an increase for a quota, if the quota is adjustable. Some requests are automatically resolved, while others are submitted to AWS Support. You can track the status of a quota increase request that is submitted to AWS Support. Requests to increase service quotas do not receive priority support. If you have an urgent request, contact AWS Support.

To request a service quota increase

2. In the navigation pane, choose AWS services.
3. Select a service from the list, or type the name of the service in the search field.
4. If the quota is adjustable, you can select its radio button or its name, and then choose Request quota increase.
5. For Change quota value, enter the new value. The new value must be greater than the current value.
6. Choose Request. After the request is resolved, the Applied quota value for the quota is set to the new value.
7. To view any pending or recently resolved requests, choose Dashboard from the navigation pane. For pending requests, choose the status of the request to open the request receipt. The initial status of a request is Pending. After the status changes to Quota requested, you’ll see the case number with AWS Support. Choose the case number to open the ticket for your request.
Service Quotas and Amazon CloudWatch Alarms

You can create Amazon CloudWatch alarms on the Service Quotas console to notify you when you're close to a quota value threshold. Setting an alarm can help you know if you need to request a quota increase.

**To create a CloudWatch alarm for a quota**

2. In the navigation pane, choose AWS services and then select a service.
3. Select a quota that supports CloudWatch alarms.
   
   If you have utilization, it appears beneath the quota description. The CloudWatch alarms section appears at the bottom of the page.
4. In Amazon CloudWatch alarms, choose Create.
5. For Alarm threshold, choose a threshold.
6. For Alarm name, enter a name for the alarm. This name must be unique within the AWS account.
7. Choose Create.
8. To add a notification to the CloudWatch alarm, see Creating a CloudWatch Alarm Based on a CloudWatch Metric in the Amazon CloudWatch User Guide.

**To delete a CloudWatch alarm**

1. Choose the service quota that has the alarm.
2. Select the alarm.
3. Choose Delete.
Using Service Quotas Request Templates

A *quota request template* helps you save time when customizing quotas for new accounts in your organization. To use a template, configure the desired service quota increases for new accounts. Then, associate the template with your organization. Whenever new accounts are created in your organization, the template automatically requests quota increases for you.

**Important**
A request template can include up to 10 quota increases.

To use a request template, you must use AWS Organizations and the new accounts must be created in the same organization. Your organization must use the features set, *all features*. If you use consolidated billing features, you can't use quota request templates.

You can update the request template by adding or deleting service quotas. You can also increase the values for any adjustable quotas. As soon as you adjust the template, those service quota values are requested for new accounts. Updating a request template does not update quota values for existing accounts.

**To configure a request template**

Use the following steps to configure the quotas request template.

2. In the navigation pane, choose *Quota request template*. If the *Quota request template* isn't visible, choose *Organization* to open it.
3. On the console, choose a *Region*, *service*, *quota*, and *quota value*, and then choose *Add*.
   - To add more quota increase requests to the template, choose *Repeat the previous step*.
4. To associate the template with your organization, choose *Associate*.

**To disassociate a request template from an organization**

If you disassociate the template from the organization, new accounts receive the AWS default quota values for all quotas. Disassociating the template from the organization doesn't delete the service quota requests from the template. You can edit the service quotas in the template.

2. In the navigation pane, choose *Quota request template*. If the *Quota request template* isn't visible, choose *Organization* to open it.
3. To disassociate the template from the organization, choose *Disassociate*.

**To delete a quota increase request from a request template**

You can remove, or delete, service quota requests from the template whether the template is associated with an organization, or not. If you reach the maximum number of service quota requests, it may be necessary to delete some service quota requests.

2. In the navigation pane, choose **Quota request template**.
3. Select the radio button for a quota increase request.
4. Choose **Remove**.
Security in Service Quotas

Cloud security at AWS is the highest priority. As an AWS customer, you benefit from a data center and network architecture that is built to meet the requirements of the most security-sensitive organizations.

Security is a shared responsibility between AWS and you. The shared responsibility model describes this as security of the cloud and security in the cloud:

- **Security of the cloud** – AWS is responsible for protecting the infrastructure that runs AWS services in the AWS Cloud. AWS also provides you with services that you can use securely. Third-party auditors regularly test and verify the effectiveness of our security as part of the AWS Compliance Programs. To learn about the compliance programs that apply to Service Quotas, see AWS Services in Scope by Compliance Program.

- **Security in the cloud** – Your responsibility is determined by the AWS service that you use. You are also responsible for other factors including the sensitivity of your data, your company’s requirements, and applicable laws and regulations.

This documentation helps you understand how to apply the shared responsibility model when using Service Quotas. The following topics show you how to configure Service Quotas to meet your security and compliance objectives. You also learn how to use other AWS services that help you to monitor and secure your Service Quotas resources.

Contents

- Data Protection in AWS Service Quotas (p. 10)
- Identity and Access Management for Service Quotas (p. 11)
- Compliance Validation for AWS Service Quotas (p. 13)
- Resilience in AWS Service Quotas (p. 14)
- Infrastructure Security in AWS Service Quotas (p. 14)

Data Protection in AWS Service Quotas

AWS Service Quotas conforms to the AWS shared responsibility model, which includes regulations and guidelines for data protection. AWS is responsible for protecting the global infrastructure that runs all the AWS services. AWS maintains control over data hosted on this infrastructure, including the security configuration controls for handling customer content and personal data. AWS customers and APN partners, acting either as data controllers or data processors, are responsible for any personal data that they put in the AWS Cloud.

For data protection purposes, we recommend that you protect AWS account credentials and set up individual user accounts with AWS Identity and Access Management (IAM), so that each user is given only the permissions necessary to fulfill their job duties. We also recommend that you secure your data in the following ways:

- Use multi-factor authentication (MFA) with each account.
- Use SSL/TLS to communicate with AWS resources.
- Set up API and user activity logging with AWS CloudTrail.
• Use AWS encryption solutions, along with all default security controls within AWS services.
• Use advanced managed security services such as Amazon Macie, which assists in discovering and securing personal data that is stored in Amazon S3.

We strongly recommend that you never put sensitive identifying information, such as your customers' account numbers, into free-form fields such as a Name field. This includes when you work with Service Quotas or other AWS services using the console, API, AWS CLI, or AWS SDKs. Any data that you enter into Service Quotas or other services might get picked up for inclusion in diagnostic logs. When you provide a URL to an external server, don't include credentials information in the URL to validate your request to that server.

For more information about data protection, see the AWS Shared Responsibility Model and GDPR blog post on the AWS Security Blog.

Identity and Access Management for Service Quotas

AWS uses security credentials to identify you and to grant you access to your AWS resources. You can use features of AWS Identity and Access Management (IAM) to allow other users, services, and applications to use your AWS resources fully or in a limited way. You can do this without sharing your security credentials.

By default, IAM users don't have permission to create, view, or modify AWS resources. To allow an IAM user to access resources such as a load balancer, and to perform tasks, you:

1. Create an IAM policy that grants the IAM user permission to use the specific resources and API actions they need.
2. Attach the policy to the IAM user or the group that the IAM user belongs to.

When you attach a policy to a user or group of users, it allows or denies the users permission to perform the specified tasks on the specified resources.

For example, you can use IAM to create users and groups under your AWS account. An IAM user can be a person, a system, or an application. Then you grant permissions to the users and groups to perform specific actions on the specified resources using an IAM policy.

Grant Permissions Using IAM Policies

When you attach a policy to a user or group of users, it allows or denies the users permission to perform the specified tasks on the specified resources.

An IAM policy is a JSON document that consists of one or more statements. Each statement is structured as shown in the following example.

```json
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "allow",
    "Action": "action",
    "Resource": "arn:aws:s3:::resource-arn",
    "Condition": {
      "condition": {
        "key": "value"
      }
    }
  }
}
```
• **Effect**— The effect can be **Allow** or **Deny**. By default, IAM users don’t have permission to use resources and API actions, so all requests are denied. An explicit allow overrides the default. An explicit deny overrides any allows.

• **Action**— The action is the specific API action for which you are granting or denying permission. For more information about specifying action, see **API Actions for Service Quotas** (p. 12).

• **Resource**— The resource that’s affected by the action. With some Service Quotas API actions, you can restrict the permissions granted or denied to a specific quota. To do so, specify its Amazon Resource Name (ARN) in this statement. Otherwise, you can use the * wildcard to specify all Service Quotas resources. For more information, see **Service Quotas Resources** (p. 12).

• **Condition**— You can optionally use conditions to control when your policy is in effect. For more information, see **Condition Keys for Service Quotas** (p. 13).

For more information, see the **IAM User Guide**.

### API Actions for Service Quotas

In the Action element of your IAM policy statement, you can specify any API action that Service Quotas offers. You must prefix the action name with the lowercase string `servicequotas:` as shown in the following example.

```
"Action": "servicequotas:GetServiceQuota"
```

To specify multiple actions in a single statement, enclose them in square brackets and separate them with a comma, as shown in the following example.

```
"Action": [ 
   "servicequotas:ListRequestedServiceQuotaChangeHistory",
   "servicequotas:ListRequestedServiceQuotaChangeHistoryByQuota"
]
```

You can also specify multiple actions using the * wildcard. The following example specifies all API action names for Service Quotas that start with Get.

```
"Action": "servicequotas:Get*"
```

To specify all API actions for Service Quotas, use the * wildcard, as shown in the following example.

```
"Action": "servicequotas:**"
```

For the list of API actions for Service Quotas, see **Service Quotas Actions**.

### Service Quotas Resources

*Resource-level permissions* refers to the ability to specify which resources users are allowed to perform actions on. For API actions that support resource-level permissions, you can control the resources that users are allowed to use with the action. To specify a resource in a policy statement, you must use its Amazon Resource Name (ARN).
The ARN for a quota has the format shown in the following example.

```
```

For API actions that don't support resource-level permissions, you must specify the resource statement shown in the following example.

```
"Resource": "*
```

**Resource-Level Permissions for Service Quotas**

The following Service Quotas actions support resource-level permissions:
- `PutServiceQuotaIncreaseRequestIntoTemplate`
- `RequestServiceQuotaIncrease`

For more information, see Actions Defined by Service Quotas in the *IAM User Guide*.

**Condition Keys for Service Quotas**

When you create a policy, you can specify the conditions that control when the policy is in effect. Each condition contains one or more key-value pairs. There are global condition keys and service-specific condition keys.

The `servicequotas:service` key is specific to Service Quotas. The following Service Quotas API actions support this key:
- `PutServiceQuotaIncreaseRequestIntoTemplate`
- `RequestServiceQuotaIncrease`

For more information about global condition keys, see AWS Global Condition Context Keys in the *IAM User Guide*.

**Predefined AWS Managed Policies for Service Quotas**

The managed policies created by AWS grant the required permissions for common use cases. You can attach these policies to your IAM users, based on the access to Service Quotas that they require:

- `ServiceQuotasFullAccess` — Grants full access required to use Service Quotas features.
- `ServiceQuotasReadOnlyAccess` — Grants read-only access to Service Quotas features.

**Compliance Validation for AWS Service Quotas**

Third-party auditors assess the security and compliance of AWS Service Quotas as part of multiple AWS compliance programs. These include SOC, PCI, FedRAMP, HIPAA, and others.

For a list of AWS services in scope of specific compliance programs, see AWS Services in Scope by Compliance Program. For general information, see AWS Compliance Programs.

You can download third-party audit reports using AWS Artifact. For more information, see Downloading Reports in AWS Artifact.
Your compliance responsibility when using Service Quotas is determined by the sensitivity of your data, your company’s compliance objectives, and applicable laws and regulations. AWS provides the following resources to help with compliance:

- **Security and Compliance Quick Start Guides** – These deployment guides discuss architectural considerations and provide steps for deploying security- and compliance-focused baseline environments on AWS.
- **Architecting for HIPAA Security and Compliance Whitepaper** – This whitepaper describes how companies can use AWS to create HIPAA-compliant applications.
- **AWS Compliance Resources** – This collection of workbooks and guides might apply to your industry and location.
- **Evaluating Resources with Rules** in the **AWS Config Developer Guide** – The AWS Config service assesses how well your resource configurations comply with internal practices, industry guidelines, and regulations.
- **AWS Security Hub** – This AWS service provides a comprehensive view of your security state within AWS that helps you check your compliance with security industry standards and best practices.

### Resilience in AWS Service Quotas

The AWS global infrastructure is built around AWS Regions and Availability Zones. AWS Regions provide multiple physically separated and isolated Availability Zones, which are connected with low-latency, high-throughput, and highly redundant networking. With Availability Zones, you can design and operate applications and databases that automatically fail over between zones without interruption. Availability Zones are more highly available, fault tolerant, and scalable than traditional single or multiple data center infrastructures.

For more information about AWS Regions and Availability Zones, see [AWS Global Infrastructure](https://aws.amazon.com/global-infrastructure/).

### Infrastructure Security in AWS Service Quotas

As a managed service, AWS Service Quotas is protected by the AWS global network security procedures that are described in the [Amazon Web Services: Overview of Security Processes](https://aws.amazon.com/security/) whitepaper.

You use AWS published API calls to access Service Quotas through the network. Clients must support Transport Layer Security (TLS) 1.0 or later. We recommend TLS 1.2 or later. Clients must also support cipher suites with perfect forward secrecy (PFS) such as Ephemeral Diffie-Hellman (DHE) or Elliptic Curve Ephemeral Diffie-Hellman (ECDHE). Most modern systems such as Java 7 and later support these modes.

Additionally, requests must be signed by using an access key ID and a secret access key that is associated with an IAM principal. Or you can use the [AWS Security Token Service (AWS STS)](https://aws.amazon.com/sts/) to generate temporary security credentials to sign requests.
Service Quotas Values

This table lists the default maximum values for Service Quotas resources for your AWS account. These quota values are per Region, unless noted otherwise. You can't adjust these quota values.

Increase Requests

<table>
<thead>
<tr>
<th>Quota</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active service quota increase requests per account</td>
<td>20</td>
</tr>
<tr>
<td>Active service quota increase requests per Region</td>
<td>2</td>
</tr>
<tr>
<td>Active service quota increase requests per quota</td>
<td>1</td>
</tr>
</tbody>
</table>

Service Quota API Request Rates

<table>
<thead>
<tr>
<th>Quota</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>GetAWSDefaultServiceQuota requests per second</td>
<td>5</td>
</tr>
<tr>
<td>Additional GetAWSDefaultServiceQuota requests per second sent in one burst</td>
<td>5</td>
</tr>
<tr>
<td>GetRequestedServiceQuotaChange requests per second</td>
<td>5</td>
</tr>
<tr>
<td>Additional GetRequestedServiceQuotaChange requests per second sent in one burst</td>
<td>5</td>
</tr>
<tr>
<td>GetServiceQuota requests per second</td>
<td>5</td>
</tr>
<tr>
<td>Additional GetServiceQuota requests per second sent in one burst</td>
<td>5</td>
</tr>
<tr>
<td>ListAWSDefaultServiceQuotas requests per second</td>
<td>10</td>
</tr>
<tr>
<td>Additional ListAWSDefaultServiceQuotas requests per second sent in one burst</td>
<td>10</td>
</tr>
<tr>
<td>ListRequestedServiceQuotaChangeHistory requests per second</td>
<td>5</td>
</tr>
<tr>
<td>Additional ListRequestedServiceQuotaChangeHistory requests per second sent in one burst</td>
<td>5</td>
</tr>
<tr>
<td>ListRequestedServiceQuotaChangeHistoryByQuota requests per second</td>
<td>5</td>
</tr>
<tr>
<td>Additional ListRequestedServiceQuotaChangeHistoryByQuota requests per second sent in one burst</td>
<td>5</td>
</tr>
<tr>
<td>ListServiceQuotas requests per second</td>
<td>10</td>
</tr>
<tr>
<td>Additional ListServiceQuotas requests per second sent in one burst</td>
<td>10</td>
</tr>
<tr>
<td>ListServices requests per second</td>
<td>10</td>
</tr>
<tr>
<td>Additional ListServices requests per second sent in one burst</td>
<td>10</td>
</tr>
<tr>
<td>RequestServiceQuotasIncrease requests per second</td>
<td>3</td>
</tr>
</tbody>
</table>
## Quota Request Template API Request Rates

<table>
<thead>
<tr>
<th>Quota</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>ServiceQuotaIncrease requests per second sent in one burst</td>
<td>3</td>
</tr>
<tr>
<td>AssociateQuotaTemplate requests per second</td>
<td>1</td>
</tr>
<tr>
<td>Additional AssociateQuotaTemplate requests per second sent in one burst</td>
<td>1</td>
</tr>
<tr>
<td>DeleteServiceQuotaTemplate requests per second</td>
<td>2</td>
</tr>
<tr>
<td>Additional DeleteServiceQuotaTemplate requests per second sent in one burst</td>
<td>1</td>
</tr>
<tr>
<td>DisassociateQuotaTemplate requests per second</td>
<td>1</td>
</tr>
<tr>
<td>Additional DisassociateQuotaTemplate requests per second sent in one burst</td>
<td>1</td>
</tr>
<tr>
<td>GetAssociationForQuotaTemplate requests per second</td>
<td>2</td>
</tr>
<tr>
<td>Additional GetAssociationForQuotaTemplate requests per second sent in one burst</td>
<td>2</td>
</tr>
<tr>
<td>GetServiceQuotaTemplateIncreaseRequestFromTemplate requests per second</td>
<td>2</td>
</tr>
<tr>
<td>Additional GetServiceQuotaTemplateIncreaseRequestFromTemplate requests per second sent in one burst</td>
<td>1</td>
</tr>
<tr>
<td>ListServiceQuotaTemplateIncreaseRequestsInTemplate requests per second</td>
<td>2</td>
</tr>
<tr>
<td>Additional ListServiceQuotaTemplateIncreaseRequestsInTemplate requests per second sent in one burst</td>
<td>1</td>
</tr>
<tr>
<td>PutServiceQuotaTemplateIncreaseRequestIntoTemplate requests per second</td>
<td>1</td>
</tr>
<tr>
<td>Additional PutServiceQuotaTemplateIncreaseRequestIntoTemplate per second sent in one burst</td>
<td>1</td>
</tr>
</tbody>
</table>
Document History for Service Quotas

The following table describes the releases for Service Quotas.

<table>
<thead>
<tr>
<th>Change</th>
<th>Description</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial release</td>
<td>This release introduces Service Quotas</td>
<td>June 24, 2019</td>
</tr>
</tbody>
</table>