



AWS Auto Scaling ▾

Overview

Features

Pricing

Resources

FAQs

General

Q. What is AWS Auto Scaling?

AWS Auto Scaling is a new AWS service that helps you optimize the performance of your applications while lowering infrastructure costs by easily and safely scaling multiple AWS resources. It simplifies the scaling experience by allowing you to scale collections of related resources that support your application with just a few clicks. AWS Auto Scaling helps you configure consistent and congruent scaling policies across the full infrastructure stack backing your application. AWS Auto Scaling will automatically scale resources as needed to align to your selected scaling strategy, so you maintain performance and pay only for the resources you actually need.

[Show less](#)

Q. What are the benefits of AWS Auto Scaling?

AWS Auto Scaling is a fast, easy way to optimize the performance and costs of your applications.

- **Setup scaling quickly:** AWS Auto Scaling provides a unified scaling experience for all of the scalable resources powering your application. You can see the average utilization for all of your scalable resources and quickly define target utilization levels for each group of like resources from a single, intuitive interface.



AWS Auto Scaling ▾

Overview

Features

Pricing

Resources

FAQs

resource capacity so you avoid overspending.

[Show less](#)

Q. When should I use AWS Auto Scaling?

You should use AWS Auto Scaling if you have an application that uses one or more scalable resources and experiences variable load. A good example would be an e-commerce web application that receives variable traffic through the day. It follows a standard three tier architecture with Elastic Load Balancing for distributing incoming traffic, Amazon EC2 for the compute layer, and DynamoDB for the data layer. In this case, AWS Auto Scaling will scale one or more EC2 Auto Scaling groups and DynamoDB tables that are powering the application in response to the demand curve.

[Show less](#)

Q. How can I get started with AWS Auto Scaling?

AWS Auto Scaling allows you to select your applications based on resource tags or AWS CloudFormation stacks. In just a few clicks, you can create a scaling plan for your application, which defines how each of the resources in your application should be scaled. For each resource, AWS Auto Scaling creates a target tracking scaling policy with the most popular metric for that resource type and keeps it at a target value based on your selected scaling strategy. To set target values for your resource metrics, you can choose from three predefined scaling recommendations that optimize availability, optimize costs, or balance the two. Or, if you prefer, you can define your own target values. AWS Auto Scaling also automatically sets the min/max values for the resources.



AWS Auto Scaling ▾

Overview

Features

Pricing

Resources

FAQs

To scale a resource other than EC2, you can use the [Application Auto Scaling API](#), which allows you to define scaling policies to automatically scale your AWS resources or schedule one-time or recurring scaling actions. Application Auto Scaling can scale Amazon ECS services, Amazon EC2 Spot fleets, Amazon EMR clusters, Amazon AppStream 2.0 fleets, provisioned read and write capacity for Amazon DynamoDB tables and global secondary indexes, Amazon Aurora Replicas, and Amazon SageMaker endpoint variants.

To configure automatic scaling for multiple resources across multiple services, use AWS Auto Scaling to create a scaling plan for the resources underlying your application.

[Show less](#)

Q. When should I use AWS Auto Scaling vs. Amazon EC2 Auto Scaling?

You should use AWS Auto Scaling to manage scaling for multiple resources across multiple services. AWS Auto Scaling lets you define dynamic scaling policies for multiple EC2 Auto Scaling groups or other resources using predefined scaling strategies. Using AWS Auto Scaling to configure scaling policies for all of the scalable resources in your application is faster than managing scaling policies for each resource via its individual service console. It's also easier, as AWS Auto Scaling includes predefined scaling strategies that simplify the setup of scaling policies.

You should use EC2 Auto Scaling if you only need to scale Amazon EC2 Auto Scaling groups, or if you are only interested in maintaining the health of your EC2 fleet. You should also use EC2 Auto Scaling if you need to create or configure Amazon EC2 Auto Scaling groups, or if you need to set up scheduled or step scaling policies (as AWS Auto Scaling supports only target tracking scaling policies).

AWS Auto Scaling

▼

- Overview
- Features
- Pricing
- Resources

FAQs

individual service consoles, Auto Scaling API, or Application Auto Scaling API to scale individual AWS services. You should also use the individual consoles or API if you want to setup step scaling policies or scheduled scaling, as AWS Auto Scaling creates target tracking scaling policies only.

[Show less](#)

Q. How is AWS Auto Scaling different than the scaling capabilities for individual services?

The following table provides a comparison of AWS scaling options.

	AWS Auto Scaling	Amazon EC2 Auto Scaling	Auto Scaling for Other Services
Resources you can scale	EC2 Auto Scaling groups EC2 Spot Fleets ECS services DynamoDB provisioned capacity for tables & GSIs Aurora Replicas	EC2 Auto Scaling groups	EC2 Spot Fleets ECS services DynamoDB provisioned capacity for tables & GSIs Aurora Replicas EMR clusters Appstream 2.0 fleet Sagemaker endpoint variants
Scaling method	Application-wide scaling using a unified interface	One Auto Scaling group at a time	One resource at a time
Automatic discovery of all scalable resources in your application	Yes	No	No

AWS Auto Scaling ▾

Overview

Features

Pricing

Resources

FAQs

target tracking scaling policies*	yes	yes	yes
Setup scheduled scaling actions	No	Yes	Yes
Setup step scaling policies	No	Yes	Yes
Configure a scaling policy with different metrics and thresholds for each resource	No	Yes	Yes

* Recommended versus step scaling policies

[Show less](#)

Features

Q. What can I scale with AWS Auto Scaling?

You can use AWS Auto Scaling to setup scaling for the following resources in your application through a single, unified interface:

- [Amazon EC2](#) Auto Scaling groups
- [Amazon Elastic Container Service \(ECS\)](#) services (currently ECS services cannot be discovered using resource tags)
- [Amazon EC2 Spot](#) Fleets
- [Amazon DynamoDB](#) throughput capacity



AWS Auto Scaling ▾

[Overview](#)[Features](#)[Pricing](#)[Resources](#)[FAQs](#)

Q. How do I select an application stack within AWS Auto Scaling?

You can either select an AWS CloudFormation stack or select resources based on common resource tag(s). Please note that currently, ECS services cannot be discovered using tags.

[Show less](#)

Q. How does AWS Auto Scaling discover what resources can scale?

AWS Auto Scaling will scan your selected AWS CloudFormation stack or resources with the specified tags to identify the supported AWS resource types that can be scaled. Please note that currently, ECS services cannot be discovered using tags.

[Show less](#)

Availability and Pricing

Q. Which regions is AWS Auto Scaling available in?

AWS Auto Scaling is available in US East (Northern Virginia), US East (Ohio), US West (Oregon), EU (Ireland), and Asia Pacific (Singapore) public AWS regions, with more regions to follow.

[Show less](#)

Q. How much does AWS Auto Scaling cost?



AWS Auto Scaling ▾

- Overview
- Features
- Pricing
- Resources
- FAQs**

Ready to get started?
Sign up

Have more questions?
Contact us

Sign In to the Console

- Twitter
- Facebook
- Podcast
- Twitch
- AWS Blog
- RSS News Feed
- Email Updates

- AWS & Cloud Computing**
- What is Cloud Computing?
 - What is Caching?
 - What is NoSQL?
 - What is DevOps?
 - What is Docker?
 - Products & Services
 - Customer Success
 - Economics Center
 - Architecture Center
 - Security Center
 - What's New
 - Whitepapers



AWS Auto Scaling

Overview

Features

Pricing

Resources

FAQs

[DevOps](#)

[Serverless Computing](#)

[Big Data](#)

[High Performance Computing](#)

[Mobile Services](#)

[Digital Marketing](#)

[Game Development](#)

[Digital Media](#)

[Government & Education](#)

[Health](#)

[Financial Services](#)

[Windows on AWS](#)

[Retail](#)

[Power & Utilities](#)

[Oil & Gas](#)

[Automotive](#)

[Blockchain](#)

[Manufacturing](#)

Resources & Training

[Developers](#)

[Java on AWS](#)

[JavaScript on AWS](#)

[Mobile on AWS](#)

[PHP on AWS](#)

[Python on AWS](#)

[Ruby on AWS](#)

[.NET on AWS](#)

[SDKs & Tools](#)

[AWS Marketplace](#)

[User Groups](#)



AWS Auto Scaling ▾

[Overview](#)

[Features](#)

[Pricing](#)

[Resources](#)

[FAQs](#)

[AWS Identity & Access Management](#)

[Security Credentials](#)

[Request Service Limit Increases](#)

[Contact Us](#)

Amazon Web Services is Hiring.

Amazon Web Services (AWS) is a dynamic, growing business unit within Amazon.com. We are currently hiring Software Development Engineers, Product Managers, Account Managers, Solutions Architects, Support Engineers, System Engineers, Designers and more. Visit our [careers](#) page to learn more.

Amazon.com is an Equal Opportunity-Affirmative Action Employer – Minority / Female / Disability / Veteran / Gender Identity / Sexual Orientation.

Language [Bahasa Indonesia](#) | [Deutsch](#) | [English](#) | [Español](#) | [Français](#) | [Italiano](#) | [Português](#) | [Tiếng Việt](#) | [Türkçe](#)
[Русский](#) | [ไทย](#) | [日本語](#) | [한국어](#) | [中文 \(简体\)](#) | [中文 \(繁體\)](#)

[Site Terms](#) | [Privacy](#)

© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.