

---

# AWS Auto Scaling

## User Guide



## **AWS Auto Scaling: User Guide**

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

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

## Table of Contents

What Is AWS Auto Scaling? .....	1
Scalable Resources .....	1
How AWS Auto Scaling Works .....	1
How to Get Started .....	1
Getting Started .....	2
Requirements .....	2
Step 1: Select Your Application .....	2
Step 2: Configure Your Scaling Plan .....	2
Step 3: Create Your Scaling Plan .....	2
Step 4: Delete Your Scaling Plan (Optional) .....	2
Authentication and Access Control .....	3
AWS Auto Scaling Actions .....	3
AWS Auto Scaling Resources .....	3
AWS Auto Scaling Keys .....	3
Example Policies .....	3
Limits .....	6
Document History .....	7

# What Is AWS Auto Scaling?

AWS Auto Scaling enables you to configure automatic scaling for the scalable AWS resources for your application in a matter of minutes. AWS Auto Scaling uses the Auto Scaling and Application Auto Scaling services to configure scaling policies for your scalable AWS resources.

For more information about the benefits of this service, see [AWS Auto Scaling](#).

## Scalable Resources

Use AWS Auto Scaling to automatically scale the following resources for your applications:

- Aurora DB clusters
- Auto Scaling groups
- ECS services
- DynamoDB global secondary indexes
- DynamoDB tables
- Spot Fleet requests

## How AWS Auto Scaling Works

With AWS Auto Scaling, you create a scaling plan with a set of instructions used to configure dynamic scaling for the scalable resources in your application. AWS Auto Scaling creates target tracking scaling policies for the scalable resources in your scaling plan. Target tracking scaling policies adjust the capacity of your scalable resource as required to maintain resource utilization at the target value that you specified. For more information, see [Target Tracking Scaling Policies](#) in the *Application Auto Scaling User Guide*.

You can create one scaling plan per application. You can add each scalable resource to one scaling plan. If you have already configured scaling policies for a scalable resource in your application, AWS Auto Scaling keeps the existing scaling policies instead of creating additional scaling policies for the resource.

## How to Get Started

To get started, create a scaling plan. For more information, see [Getting Started with AWS Auto Scaling \(p. 2\)](#).

# Getting Started with AWS Auto Scaling

This tutorial provides a hands-on introduction to AWS Auto Scaling through the AWS Management Console, a web-based interface. To create your first scaling plan, complete the following steps.

## Tasks

- [Requirements](#) (p. 2)
- [Step 1: Select Your Application](#) (p. 2)
- [Step 2: Configure Your Scaling Plan](#) (p. 2)
- [Step 3: Create Your Scaling Plan](#) (p. 2)
- [Step 4: Delete Your Scaling Plan \(Optional\)](#) (p. 2)

## Requirements

AWS Auto Scaling uses a AWS CloudFormation stack to discover the AWS resources that power your application. You need the name of your AWS CloudFormation stack before you can create a scaling plan.

You can create one scaling plan per application and add each scalable resource to one scaling plan.

## Step 1: Select Your Application

Open the AWS Auto Scaling console at <https://console.aws.amazon.com/autoscaling/>. From the welcome page, choose **Get started**. Select your AWS CloudFormation stack and choose **Next**.

## Step 2: Configure Your Scaling Plan

On the **Configure scaling plan** page, for **Scaling plan details**, type a name for your scaling plan. For each type of scalable resource, select a strategy and a metric. To omit a type of scalable resource from your scaling plan, clear **Include in plan**. When you are finished, choose **Next**.

## Step 3: Create Your Scaling Plan

Review your scaling plan and choose **Create scaling plan**.

## Step 4: Delete Your Scaling Plan (Optional)

When you are finished with a scaling plan, you can delete it. Deleting a scaling plan deletes the target tracking policies that AWS Auto Scaling created on your behalf. Deleting a scaling plan does not delete your AWS CloudFormation stack or your application resources.

# Authentication and Access Control for AWS Auto Scaling

By default, IAM users don't have permission to create or modify AWS resources. To grant IAM users permission to create or modify AWS resources, you must create policies using AWS Identity and Access Management (IAM). IAM policies grant permissions to specific resources and API actions. You attach an IAM policy to the IAM users or groups that require the permissions it grants. For more information, see [Access Management](#) in the *IAM User Guide*.

## AWS Auto Scaling Actions

You can specify any and all AWS Auto Scaling actions in an IAM policy. Use the following prefix with the name of the action: `autoscaling-plans:`. For example:

```
"Action": "autoscaling-plans:DescribeScalingPlans"
```

You can also use wildcards. For example, use `autoscaling-plans:*` to specify all AWS Auto Scaling actions.

```
"Action": "autoscaling-plans:*"
```

Use `Describe:*` to specify all actions whose names start with `Describe`.

```
"Action": "autoscaling-plans:Describe*"
```

For a list of actions, see [AWS Auto Scaling Actions](#).

## AWS Auto Scaling Resources

When writing an IAM policy to control access to AWS Auto Scaling actions, you must use `**` as the resource. There are no supported Amazon Resource Names (ARNs) for AWS Auto Scaling resources.

## AWS Auto Scaling Keys

For a list of context keys supported by each AWS service and a list of AWS-wide policy keys, see [AWS Service Actions and Condition Context Keys](#) and [Global and IAM Condition Context Keys](#) in the *IAM User Guide*.

## Example Policies

To create a scaling plan, users must have permission to use the actions in the following example policy.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "autoscaling-plans:*",
        "cloudwatch:PutMetricAlarm",
        "cloudwatch>DeleteAlarms",
        "cloudwatch:DescribeAlarms",
        "cloudformation:ListStackResources"
      ],
      "Resource": "*"
    }
  ]
}
```

Users must have additional permissions for each type of scalable resource they must add to a scaling plan.

### Auto Scaling groups

- `autoscaling:UpdateAutoScalingGroups`
- `autoscaling:DescribeAutoScalingGroups`
- `autoscaling:PutScalingPolicy`
- `autoscaling:DescribePolicies`
- `autoscaling>DeletePolicy`

### Resource types other than Auto Scaling groups

- `application-autoscaling:RegisterScalableTarget`
- `application-autoscaling:DescribeScalableTargets`
- `application-autoscaling:DeregisterScalableTarget`
- `application-autoscaling:PutScalingPolicy`
- `application-autoscaling:DescribeScalingPolicies`
- `application-autoscaling>DeleteScalingPolicy`
- `iam:CreateServiceLinkedRole`

### ECS services

- `ecs:DescribeServices`
- `ecs:UpdateServices`

### Spot Fleet requests

- `ec2:DescribeSpotFleetRequests`
- `ec2:ModifySpotFleetRequest`

### DynamoDB tables or global indexes

- `dynamodb:DescribeTable`
- `dynamodb:UpdateTable`

### **Aurora DB clusters**

- `rds:AddTagsToResource`
- `rds:CreateDBInstance`
- `rds>DeleteDBInstance`
- `rds:DescribeDBClusters`
- `rds:DescribeDBInstances`



# AWS Auto Scaling Limits

Your AWS account has the following limits related to AWS Auto Scaling. To request a limit increase, use the [Auto Scaling Limits form](#).

- Scaling plans: 100
- Target tracking configurations per instruction: 10
- Target tracking configurations per scaling plan: 500

# Document History

The following table describes important additions to the documentation for AWS Auto Scaling.

<b>Change</b>	<b>Description</b>	<b>Date</b>
New service	Initial release of AWS Auto Scaling	January 16, 2018