

---

# Application Auto Scaling

## API Reference

**API Version 2016-02-06**



## **Application Auto Scaling: API Reference**

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

Welcome .....	1
Actions .....	2
DeleteScalingPolicy .....	3
Request Syntax .....	3
Request Parameters .....	3
Response Elements .....	5
Errors .....	5
Example .....	5
See Also .....	6
DeleteScheduledAction .....	7
Request Syntax .....	7
Request Parameters .....	7
Response Elements .....	9
Errors .....	9
See Also .....	9
DeregisterScalableTarget .....	10
Request Syntax .....	10
Request Parameters .....	10
Response Elements .....	11
Errors .....	11
Example .....	12
See Also .....	12
DescribeScalableTargets .....	14
Request Syntax .....	14
Request Parameters .....	14
Response Syntax .....	16
Response Elements .....	16
Errors .....	16
Example .....	17
See Also .....	18
DescribeScalingActivities .....	19
Request Syntax .....	19
Request Parameters .....	19
Response Syntax .....	21
Response Elements .....	21
Errors .....	22
Example .....	22
See Also .....	24
DescribeScalingPolicies .....	25
Request Syntax .....	25
Request Parameters .....	25
Response Syntax .....	27
Response Elements .....	28
Errors .....	28
Example .....	29
See Also .....	30
DescribeScheduledActions .....	32
Request Syntax .....	32
Request Parameters .....	32
Response Syntax .....	34
Response Elements .....	34
Errors .....	35
See Also .....	35
PutScalingPolicy .....	37

Request Syntax .....	37
Request Parameters .....	38
Response Syntax .....	40
Response Elements .....	40
Errors .....	40
Example .....	41
See Also .....	42
PutScheduledAction .....	43
Request Syntax .....	43
Request Parameters .....	43
Response Elements .....	46
Errors .....	46
See Also .....	46
RegisterScalableTarget .....	48
Request Syntax .....	48
Request Parameters .....	48
Response Elements .....	50
Errors .....	50
Example .....	51
See Also .....	51
Data Types .....	52
Alarm .....	53
Contents .....	53
See Also .....	53
CustomizedMetricSpecification .....	54
Contents .....	54
See Also .....	54
MetricDimension .....	55
Contents .....	55
See Also .....	55
PredefinedMetricSpecification .....	56
Contents .....	56
See Also .....	56
ScalableTarget .....	57
Contents .....	57
See Also .....	59
ScalableTargetAction .....	60
Contents .....	60
See Also .....	60
ScalingActivity .....	61
Contents .....	61
See Also .....	63
ScalingPolicy .....	64
Contents .....	64
See Also .....	66
ScheduledAction .....	67
Contents .....	67
See Also .....	69
StepAdjustment .....	70
Contents .....	70
See Also .....	71
StepScalingPolicyConfiguration .....	72
Contents .....	72
See Also .....	73
TargetTrackingScalingPolicyConfiguration .....	74
Contents .....	74
See Also .....	75

Common Parameters .....	76
Common Errors .....	78
Logging API Calls with CloudTrail .....	80
Application Auto Scaling Information in CloudTrail .....	80
Understanding Application Auto Scaling Log File Entries .....	81
.....	81

# Welcome

With Application Auto Scaling, you can configure automatic scaling for your scalable resources. You can use Application Auto Scaling to accomplish the following tasks:

- Define scaling policies to automatically scale your AWS or custom resources
- Scale your resources in response to CloudWatch alarms
- Schedule one-time or recurring scaling actions
- View the history of your scaling events

Application Auto Scaling can scale the following resources:

- Amazon ECS services. For more information, see [Service Auto Scaling](#) in the *Amazon Elastic Container Service Developer Guide*.
- Amazon EC2 Spot fleets. For more information, see [Automatic Scaling for Spot Fleet](#) in the *Amazon EC2 User Guide*.
- Amazon EMR clusters. For more information, see [Using Automatic Scaling in Amazon EMR](#) in the *Amazon EMR Management Guide*.
- AppStream 2.0 fleets. For more information, see [Fleet Auto Scaling for Amazon AppStream 2.0](#) in the *Amazon AppStream 2.0 Developer Guide*.
- Provisioned read and write capacity for Amazon DynamoDB tables and global secondary indexes. For more information, see [Managing Throughput Capacity Automatically with DynamoDB Auto Scaling](#) in the *Amazon DynamoDB Developer Guide*.
- Amazon Aurora Replicas. For more information, see [Using Amazon Aurora Auto Scaling with Aurora Replicas](#).
- Amazon SageMaker endpoint variants. For more information, see [Automatically Scaling Amazon SageMaker Models](#).
- Custom resources provided by your own applications or services. More information is available in our [GitHub repository](#).

To learn more about Application Auto Scaling, see the [Application Auto Scaling User Guide](#).

To configure automatic scaling for multiple resources across multiple services, use AWS Auto Scaling to create a scaling plan for your application. For more information, see the [AWS Auto Scaling User Guide](#).

This document was last published on November 19, 2018.

# Actions

The following actions are supported:

- [DeleteScalingPolicy](#) (p. 3)
- [DeleteScheduledAction](#) (p. 7)
- [DeregisterScalableTarget](#) (p. 10)
- [DescribeScalableTargets](#) (p. 14)
- [DescribeScalingActivities](#) (p. 19)
- [DescribeScalingPolicies](#) (p. 25)
- [DescribeScheduledActions](#) (p. 32)
- [PutScalingPolicy](#) (p. 37)
- [PutScheduledAction](#) (p. 43)
- [RegisterScalableTarget](#) (p. 48)

# DeleteScalingPolicy

Deletes the specified Application Auto Scaling scaling policy.

Deleting a policy deletes the underlying alarm action, but does not delete the CloudWatch alarm associated with the scaling policy, even if it no longer has an associated action.

To create a scaling policy or update an existing one, see [PutScalingPolicy](#) (p. 37).

## Request Syntax

```
{
  "PolicyName": "string",
  "ResourceId": "string",
  "ScalableDimension": "string",
  "ServiceNamespace": "string"
}
```

## Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](#) (p. 76).

The request accepts the following data in JSON format.

### **PolicyName** (p. 3)

The name of the scaling policy.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1600.

Pattern: `[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*`

Required: Yes

### **ResourceId** (p. 3)

The identifier of the resource associated with the scalable target. This string consists of the resource type and unique identifier.

- ECS service - The resource type is `service` and the unique identifier is the cluster name and service name. Example: `service/default/sample-webapp`.
- Spot fleet request - The resource type is `spot-fleet-request` and the unique identifier is the Spot fleet request ID. Example: `spot-fleet-request/sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE`.
- EMR cluster - The resource type is `instancegroup` and the unique identifier is the cluster ID and instance group ID. Example: `instancegroup/j-2EEZNYKUA1NTV/ig-1791Y4E1L8YI0`.
- AppStream 2.0 fleet - The resource type is `fleet` and the unique identifier is the fleet name. Example: `fleet/sample-fleet`.
- DynamoDB table - The resource type is `table` and the unique identifier is the resource ID. Example: `table/my-table`.
- DynamoDB global secondary index - The resource type is `index` and the unique identifier is the resource ID. Example: `table/my-table/index/my-table-index`.
- Aurora DB cluster - The resource type is `cluster` and the unique identifier is the cluster name. Example: `cluster:my-db-cluster`.





Valid Values: `ecs` | `elasticmapreduce` | `ec2` | `appstream` | `dynamodb` | `rds` | `sagemaker` | `custom-resource`

Required: Yes

## Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

## Errors

For information about the errors that are common to all actions, see [Common Errors \(p. 78\)](#).

### **ConcurrentUpdateException**

Concurrent updates caused an exception, for example, if you request an update to an Application Auto Scaling resource that already has a pending update.

HTTP Status Code: 400

### **InternalServiceException**

The service encountered an internal error.

HTTP Status Code: 400

### **ObjectNotFoundException**

The specified object could not be found. For any operation that depends on the existence of a scalable target, this exception is thrown if the scalable target with the specified service namespace, resource ID, and scalable dimension does not exist. For any operation that deletes or deregisters a resource, this exception is thrown if the resource cannot be found.

HTTP Status Code: 400

### **ValidationException**

An exception was thrown for a validation issue. Review the available parameters for the API request.

HTTP Status Code: 400

## Example

If you plan to create requests manually, you must replace the Authorization header contents in the examples (AUTHPARAMS) with a signature. For more information, see [Signature Version 4 Signing Process](#) in the *Amazon Web Services General Reference*. If you plan to use the [AWS CLI](#) or one of the [AWS SDKs](#), these tools sign the requests for you.

## Example

The following example deletes a scaling policy for the Amazon ECS service web-app running in the default cluster.

### Sample Request

```
POST / HTTP/1.1
Host: autoscaling.us-west-2.amazonaws.com
```

```
Accept-Encoding: identity
Content-Length: 152
X-Amz-Target: AnyScaleFrontendService.DeleteScalingPolicy
X-Amz-Date: 20160506T205712Z
User-Agent: aws-cli/1.10.23 Python/2.7.11 Darwin/15.4.0 botocore/1.4.8
Content-Type: application/x-amz-json-1.1
Authorization: AUTHPARAMS

{
  "PolicyName": "web-app-cpu-lt-25",
  "ServiceNamespace": "ecs",
  "ScalableDimension": "ecs:service:DesiredCount",
  "ResourceId": "service/default/web-app"
}
```

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for JavaScript](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V2](#)



### ScalableDimension (p. 7)

The scalable dimension. This string consists of the service namespace, resource type, and scaling property.

- `ecs:service:DesiredCount` - The desired task count of an ECS service.
- `ec2:spot-fleet-request:TargetCapacity` - The target capacity of a Spot fleet request.
- `elasticmapreduce:instancegroup:InstanceCount` - The instance count of an EMR Instance Group.
- `appstream:fleet:DesiredCapacity` - The desired capacity of an AppStream 2.0 fleet.
- `dynamodb:table:ReadCapacityUnits` - The provisioned read capacity for a DynamoDB table.
- `dynamodb:table:WriteCapacityUnits` - The provisioned write capacity for a DynamoDB table.
- `dynamodb:index:ReadCapacityUnits` - The provisioned read capacity for a DynamoDB global secondary index.
- `dynamodb:index:WriteCapacityUnits` - The provisioned write capacity for a DynamoDB global secondary index.
- `rds:cluster:ReadReplicaCount` - The count of Aurora Replicas in an Aurora DB cluster. Available for Aurora MySQL-compatible edition.
- `sagemaker:variant:DesiredInstanceCount` - The number of EC2 instances for an Amazon SageMaker model endpoint variant.
- `custom-resource:ResourceType:Property` - The scalable dimension for a custom resource provided by your own application or service.

Type: String

Valid Values: `ecs:service:DesiredCount` | `ec2:spot-fleet-request:TargetCapacity` | `elasticmapreduce:instancegroup:InstanceCount` | `appstream:fleet:DesiredCapacity` | `dynamodb:table:ReadCapacityUnits` | `dynamodb:table:WriteCapacityUnits` | `dynamodb:index:ReadCapacityUnits` | `dynamodb:index:WriteCapacityUnits` | `rds:cluster:ReadReplicaCount` | `sagemaker:variant:DesiredInstanceCount` | `custom-resource:ResourceType:Property`

Required: No

### ScheduledActionName (p. 7)

The name of the scheduled action.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1600.

Pattern: `[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*`

Required: Yes

### ServiceNamespace (p. 7)

The namespace of the AWS service that provides the resource or `custom-resource` for a resource provided by your own application or service. For more information, see [AWS Service Namespaces](#) in the *Amazon Web Services General Reference*.

Type: String

Valid Values: `ecs` | `elasticmapreduce` | `ec2` | `appstream` | `dynamodb` | `rds` | `sagemaker` | `custom-resource`

Required: Yes

## Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

## Errors

For information about the errors that are common to all actions, see [Common Errors \(p. 78\)](#).

### **ConcurrentUpdateException**

Concurrent updates caused an exception, for example, if you request an update to an Application Auto Scaling resource that already has a pending update.

HTTP Status Code: 400

### **InternalServiceException**

The service encountered an internal error.

HTTP Status Code: 400

### **ObjectNotFoundException**

The specified object could not be found. For any operation that depends on the existence of a scalable target, this exception is thrown if the scalable target with the specified service namespace, resource ID, and scalable dimension does not exist. For any operation that deletes or deregisters a resource, this exception is thrown if the resource cannot be found.

HTTP Status Code: 400

### **ValidationException**

An exception was thrown for a validation issue. Review the available parameters for the API request.

HTTP Status Code: 400

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for JavaScript](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V2](#)



### ScalableDimension (p. 10)

The scalable dimension associated with the scalable target. This string consists of the service namespace, resource type, and scaling property.

- `ecs:service:DesiredCount` - The desired task count of an ECS service.
- `ec2:spot-fleet-request:TargetCapacity` - The target capacity of a Spot fleet request.
- `elasticmapreduce:instancegroup:InstanceCount` - The instance count of an EMR Instance Group.
- `appstream:fleet:DesiredCapacity` - The desired capacity of an AppStream 2.0 fleet.
- `dynamodb:table:ReadCapacityUnits` - The provisioned read capacity for a DynamoDB table.
- `dynamodb:table:WriteCapacityUnits` - The provisioned write capacity for a DynamoDB table.
- `dynamodb:index:ReadCapacityUnits` - The provisioned read capacity for a DynamoDB global secondary index.
- `dynamodb:index:WriteCapacityUnits` - The provisioned write capacity for a DynamoDB global secondary index.
- `rds:cluster:ReadReplicaCount` - The count of Aurora Replicas in an Aurora DB cluster. Available for Aurora MySQL-compatible edition.
- `sagemaker:variant:DesiredInstanceCount` - The number of EC2 instances for an Amazon SageMaker model endpoint variant.
- `custom-resource:ResourceType:Property` - The scalable dimension for a custom resource provided by your own application or service.

Type: String

Valid Values: `ecs:service:DesiredCount` | `ec2:spot-fleet-request:TargetCapacity` | `elasticmapreduce:instancegroup:InstanceCount` | `appstream:fleet:DesiredCapacity` | `dynamodb:table:ReadCapacityUnits` | `dynamodb:table:WriteCapacityUnits` | `dynamodb:index:ReadCapacityUnits` | `dynamodb:index:WriteCapacityUnits` | `rds:cluster:ReadReplicaCount` | `sagemaker:variant:DesiredInstanceCount` | `custom-resource:ResourceType:Property`

Required: Yes

### ServiceNamespace (p. 10)

The namespace of the AWS service that provides the resource or `custom-resource` for a resource provided by your own application or service. For more information, see [AWS Service Namespaces](#) in the *Amazon Web Services General Reference*.

Type: String

Valid Values: `ecs` | `elasticmapreduce` | `ec2` | `appstream` | `dynamodb` | `rds` | `sagemaker` | `custom-resource`

Required: Yes

## Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

## Errors

For information about the errors that are common to all actions, see [Common Errors \(p. 78\)](#).



### **ConcurrentUpdateException**

Concurrent updates caused an exception, for example, if you request an update to an Application Auto Scaling resource that already has a pending update.

HTTP Status Code: 400

### **InternalServiceException**

The service encountered an internal error.

HTTP Status Code: 400

### **ObjectNotFoundException**

The specified object could not be found. For any operation that depends on the existence of a scalable target, this exception is thrown if the scalable target with the specified service namespace, resource ID, and scalable dimension does not exist. For any operation that deletes or deregisters a resource, this exception is thrown if the resource cannot be found.

HTTP Status Code: 400

### **ValidationException**

An exception was thrown for a validation issue. Review the available parameters for the API request.

HTTP Status Code: 400

## Example

If you plan to create requests manually, you must replace the Authorization header contents in the examples (AUTHPARAMS) with a signature. For more information, see [Signature Version 4 Signing Process](#) in the *Amazon Web Services General Reference*. If you plan to use the [AWS CLI](#) or one of the [AWS SDKs](#), these tools sign the requests for you.

## Example

The following example deregisters a scalable target for an Amazon ECS service called `web-app` that is running in the `default` cluster.

### Sample Request

```
POST / HTTP/1.1
Host: autoscaling.us-west-2.amazonaws.com
Accept-Encoding: identity
Content-Length: 117
X-Amz-Target: AnyScaleFrontendService.DeregisterScalableTarget
X-Amz-Date: 20160506T210150Z
User-Agent: aws-cli/1.10.23 Python/2.7.11 Darwin/15.4.0 botocore/1.4.8
Content-Type: application/x-amz-json-1.1
Authorization: AUTHPARAMS

{
  "ResourceId": "service/default/web-app",
  "ServiceNamespace": "ecs",
  "ScalableDimension": "ecs:service:DesiredCount"
}
```

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for JavaScript](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V2](#)

# DescribeScalableTargets

Gets information about the scalable targets in the specified namespace.

You can filter the results using the `ResourceIds` and `ScalableDimension` parameters.

To create a scalable target or update an existing one, see [RegisterScalableTarget \(p. 48\)](#). If you are no longer using a scalable target, you can deregister it using [DeregisterScalableTarget \(p. 10\)](#).

## Request Syntax

```
{
  "MaxResults": number,
  "NextToken": "string",
  "ResourceIds": [ "string" ],
  "ScalableDimension": "string",
  "ServiceNamespace": "string"
}
```

## Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters \(p. 76\)](#).

The request accepts the following data in JSON format.

### MaxResults (p. 14)

The maximum number of scalable targets. This value can be between 1 and 50. The default value is 50.

If this parameter is used, the operation returns up to `MaxResults` results at a time, along with a `NextToken` value. To get the next set of results, include the `NextToken` value in a subsequent call. If this parameter is not used, the operation returns up to 50 results and a `NextToken` value, if applicable.

Type: Integer

Required: No

### NextToken (p. 14)

The token for the next set of results.

Type: String

Pattern: [ \u0020-\uD7FF\uE000-\uFFFF\uD800\uDC00-\uDBFF\uDFFF\r\n\t]\*

Required: No

### ResourceIds (p. 14)

The identifier of the resource associated with the scalable target. This string consists of the resource type and unique identifier. If you specify a scalable dimension, you must also specify a resource ID.

- ECS service - The resource type is `service` and the unique identifier is the cluster name and service name. Example: `service/default/sample-webapp`.
- Spot fleet request - The resource type is `spot-fleet-request` and the unique identifier is the Spot fleet request ID. Example: `spot-fleet-request/sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE`.

- EMR cluster - The resource type is `instancegroup` and the unique identifier is the cluster ID and instance group ID. Example: `instancegroup/j-2EEZNYKUA1NTV/ig-1791Y4E1L8YI0`.
- AppStream 2.0 fleet - The resource type is `fleet` and the unique identifier is the fleet name. Example: `fleet/sample-fleet`.
- DynamoDB table - The resource type is `table` and the unique identifier is the resource ID. Example: `table/my-table`.
- DynamoDB global secondary index - The resource type is `index` and the unique identifier is the resource ID. Example: `table/my-table/index/my-table-index`.
- Aurora DB cluster - The resource type is `cluster` and the unique identifier is the cluster name. Example: `cluster:my-db-cluster`.
- Amazon SageMaker endpoint variants - The resource type is `variant` and the unique identifier is the resource ID. Example: `endpoint/my-end-point/variant/KMeansClustering`.
- Custom resources are not supported with a resource type. This parameter must specify the `OutputValue` from the CloudFormation template stack used to access the resources. The unique identifier is defined by the service provider.

Type: Array of strings

Length Constraints: Minimum length of 1. Maximum length of 1600.

Pattern: `[ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*`

Required: No

#### ScalableDimension (p. 14)

The scalable dimension associated with the scalable target. This string consists of the service namespace, resource type, and scaling property. If you specify a scalable dimension, you must also specify a resource ID.

- `ecs:service:DesiredCount` - The desired task count of an ECS service.
- `ec2:spot-fleet-request:TargetCapacity` - The target capacity of a Spot fleet request.
- `elasticmapreduce:instancegroup:InstanceCount` - The instance count of an EMR Instance Group.
- `appstream:fleet:DesiredCapacity` - The desired capacity of an AppStream 2.0 fleet.
- `dynamodb:table:ReadCapacityUnits` - The provisioned read capacity for a DynamoDB table.
- `dynamodb:table:WriteCapacityUnits` - The provisioned write capacity for a DynamoDB table.
- `dynamodb:index:ReadCapacityUnits` - The provisioned read capacity for a DynamoDB global secondary index.
- `dynamodb:index:WriteCapacityUnits` - The provisioned write capacity for a DynamoDB global secondary index.
- `rds:cluster:ReadReplicaCount` - The count of Aurora Replicas in an Aurora DB cluster. Available for Aurora MySQL-compatible edition.
- `sagemaker:variant:DesiredInstanceCount` - The number of EC2 instances for an Amazon SageMaker model endpoint variant.
- `custom-resource:ResourceType:Property` - The scalable dimension for a custom resource provided by your own application or service.

Type: String

Valid Values: `ecs:service:DesiredCount` | `ec2:spot-fleet-request:TargetCapacity` | `elasticmapreduce:instancegroup:InstanceCount` | `appstream:fleet:DesiredCapacity` | `dynamodb:table:ReadCapacityUnits` | `dynamodb:table:WriteCapacityUnits` | `dynamodb:index:ReadCapacityUnits` | `dynamodb:index:WriteCapacityUnits` | `rds:cluster:ReadReplicaCount`

| sagemaker:variant:DesiredInstanceCount | custom-  
resource:ResourceType:Property

Required: No

#### ServiceNamespace (p. 14)

The namespace of the AWS service that provides the resource or custom-resource for a resource provided by your own application or service. For more information, see [AWS Service Namespaces](#) in the *Amazon Web Services General Reference*.

Type: String

Valid Values: ecs | elasticmapreduce | ec2 | appstream | dynamodb | rds |  
sagemaker | custom-resource

Required: Yes

## Response Syntax

```
{
  "NextToken": "string",
  "ScalableTargets": [
    {
      "CreationTime": number,
      "MaxCapacity": number,
      "MinCapacity": number,
      "ResourceId": "string",
      "RoleARN": "string",
      "ScalableDimension": "string",
      "ServiceNamespace": "string"
    }
  ]
}
```

## Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

#### NextToken (p. 16)

The token required to get the next set of results. This value is null if there are no more results to return.

Type: String

Pattern: [ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]\*

#### ScalableTargets (p. 16)

The scalable targets that match the request parameters.

Type: Array of [ScalableTarget \(p. 57\)](#) objects

## Errors

For information about the errors that are common to all actions, see [Common Errors \(p. 78\)](#).

### **ConcurrentUpdateException**

Concurrent updates caused an exception, for example, if you request an update to an Application Auto Scaling resource that already has a pending update.

HTTP Status Code: 400

### **InternalServiceException**

The service encountered an internal error.

HTTP Status Code: 400

### **InvalidNextTokenException**

The next token supplied was invalid.

HTTP Status Code: 400

### **ValidationException**

An exception was thrown for a validation issue. Review the available parameters for the API request.

HTTP Status Code: 400

## Example

If you plan to create requests manually, you must replace the Authorization header contents in the examples (AUTHPARAMS) with a signature. For more information, see [Signature Version 4 Signing Process](#) in the *Amazon Web Services General Reference*. If you plan to use the [AWS CLI](#) or one of the [AWS SDKs](#), these tools sign the requests for you.

## Example

The following example describes the scalable targets for the ecs service namespace.

### Sample Request

```
POST / HTTP/1.1
Host: autoscaling.us-west-2.amazonaws.com
Accept-Encoding: identity
Content-Length: 27
X-Amz-Target: AnyScaleFrontendService.DescribeScalableTargets
X-Amz-Date: 20160506T184921Z
User-Agent: aws-cli/1.10.23 Python/2.7.11 Darwin/15.4.0 botocore/1.4.8
Content-Type: application/x-amz-json-1.1
Authorization: AUTHPARAMS

{
  "ServiceNamespace": "ecs"
}
```

### Sample Response

```
HTTP/1.1 200 OK
x-amzn-RequestId: 3f10dab0-13bb-11e6-a873-676fff004c09
Content-Type: application/x-amz-json-1.1
Content-Length: 272
Date: Fri, 06 May 2016 18:49:21 GMT
```

```
{
  "ScalableTargets": [
    {
      "CreationTime": 1462558906.199,
      "MaxCapacity": 10,
      "MinCapacity": 1,
      "ResourceId": "service/default/web-app",
      "RoleARN": "arn:aws:iam::012345678910:role/ApplicationAutoscalingECSSRole",
      "ScalableDimension": "ecs:service:DesiredCount",
      "ServiceNamespace": "ecs"
    }
  ]
}
```

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for JavaScript](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V2](#)







```
| appstream:fleet:DesiredCapacity | dynamodb:table:ReadCapacityUnits  
| dynamodb:table:WriteCapacityUnits | dynamodb:index:ReadCapacityUnits  
| dynamodb:index:WriteCapacityUnits | rds:cluster:ReadReplicaCount  
| sagemaker:variant:DesiredInstanceCount | custom-  
resource:ResourceType:Property
```

Required: No

### ServiceNamespace (p. 19)

The namespace of the AWS service that provides the resource or custom-resource for a resource provided by your own application or service. For more information, see [AWS Service Namespaces](#) in the *Amazon Web Services General Reference*.

Type: String

Valid Values: ecs | elasticmapreduce | ec2 | appstream | dynamodb | rds |  
sagemaker | custom-resource

Required: Yes

## Response Syntax

```
{  
  "NextToken": "string",  
  "ScalingActivities": [  
    {  
      "ActivityId": "string",  
      "Cause": "string",  
      "Description": "string",  
      "Details": "string",  
      "EndTime": number,  
      "ResourceId": "string",  
      "ScalableDimension": "string",  
      "ServiceNamespace": "string",  
      "StartTime": number,  
      "StatusCode": "string",  
      "StatusMessage": "string"  
    }  
  ]  
}
```

## Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

### NextToken (p. 21)

The token required to get the next set of results. This value is null if there are no more results to return.

Type: String

Pattern: [ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]\*

### ScalingActivities (p. 21)

A list of scaling activity objects.

Type: Array of [ScalingActivity](#) (p. 61) objects

## Errors

For information about the errors that are common to all actions, see [Common Errors](#) (p. 78).

### **ConcurrentUpdateException**

Concurrent updates caused an exception, for example, if you request an update to an Application Auto Scaling resource that already has a pending update.

HTTP Status Code: 400

### **InternalServiceException**

The service encountered an internal error.

HTTP Status Code: 400

### **InvalidNextTokenException**

The next token supplied was invalid.

HTTP Status Code: 400

### **ValidationException**

An exception was thrown for a validation issue. Review the available parameters for the API request.

HTTP Status Code: 400

## Example

If you plan to create requests manually, you must replace the Authorization header contents in the examples (AUTHPARAMS) with a signature. For more information, see [Signature Version 4 Signing Process](#) in the *Amazon Web Services General Reference*. If you plan to use the [AWS CLI](#) or one of the [AWS SDKs](#), these tools sign the requests for you.

## Example

The following example describes the scaling activities for an Amazon ECS service called `web-app` that is running in the `default` cluster, and it limits the returned results to 2.

### Sample Request

```
POST / HTTP/1.1
Host: autoscaling.us-west-2.amazonaws.com
Accept-Encoding: identity
Content-Length: 117
X-Amz-Target: AnyScaleFrontendService.DescribeScalingActivities
X-Amz-Date: 20160506T224112Z
User-Agent: aws-cli/1.10.26 Python/2.7.11 Darwin/15.4.0 botocore/1.4.8
Content-Type: application/x-amz-json-1.1
Authorization: AUTHPARAMS

{
  "ResourceId": "service/default/web-app",
  "ServiceNamespace": "ecs",
  "ScalableDimension": "ecs:service:DesiredCount"
```

```
}
```

## Sample Response

```
HTTP/1.1 200 OK
x-amzn-RequestId: a2704130-13db-11e6-9fca-039a3edb2541
Content-Type: application/x-amz-json-1.1
Content-Length: 1784
Date: Fri, 06 May 2016 22:41:12 GMT

{
  "ScalingActivities": [
    {
      "ActivityId": "0b812df9-a093-4074-9064-8a8f6c0521f5",
      "Cause": "monitor alarm web-app-cpu-gt-75 in state ALARM triggered policy web-app-cpu-gt-75",
      "Description": "Setting desired count to 3.",
      "ResourceId": "service/default/web-app",
      "ScalableDimension": "ecs:service:DesiredCount",
      "ServiceNamespace": "ecs",
      "StartTime": 1462568034.684,
      "StatusCode": "Pending"
    },
    {
      "ActivityId": "4d759079-a31f-4d0c-8468-504c56e2eecf",
      "Cause": "monitor alarm web-app-cpu-gt-75 in state ALARM triggered policy web-app-cpu-gt-75",
      "Description": "Setting desired count to 3.",
      "EndTime": 1462574276.686,
      "ResourceId": "service/default/web-app",
      "ScalableDimension": "ecs:service:DesiredCount",
      "ServiceNamespace": "ecs",
      "StartTime": 1462574194.658,
      "StatusCode": "Successful",
      "StatusMessage": "Successfully set desired count to 3. Change successfully fulfilled by ecs."
    },
    {
      "ActivityId": "90aff0eb-dd6a-443c-889b-b809e78061c1",
      "Cause": "monitor alarm web-app-cpu-gt-75 in state ALARM triggered policy web-app-cpu-gt-75",
      "Description": "Setting desired count to 9.",
      "EndTime": 1462574333.492,
      "ResourceId": "service/default/web-app",
      "ScalableDimension": "ecs:service:DesiredCount",
      "ServiceNamespace": "ecs",
      "StartTime": 1462574254.223,
      "StatusCode": "Successful",
      "StatusMessage": "Successfully set desired count to 9. Change successfully fulfilled by ecs."
    },
    {
      "ActivityId": "ee381679-5079-46b5-ac1a-418253981efd",
      "Cause": "monitor alarm web-app-cpu-gt-75 in state ALARM triggered policy web-app-cpu-gt-75",
      "Description": "Setting desired count to 10.",
      "ResourceId": "service/default/web-app",
      "ScalableDimension": "ecs:service:DesiredCount",
      "ServiceNamespace": "ecs",
      "StartTime": 1462574434.077,
      "StatusCode": "InProgress",
      "StatusMessage": "Successfully set desired count to 10. Waiting for change to be fulfilled by ecs."
    }
  ]
}
```

```
]
}
```

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for JavaScript](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V2](#)

# DescribeScalingPolicies

Describes the scaling policies for the specified service namespace.

You can filter the results using the `ResourceId`, `ScalableDimension`, and `PolicyNames` parameters.

To create a scaling policy or update an existing one, see [PutScalingPolicy \(p. 37\)](#). If you are no longer using a scaling policy, you can delete it using [DeleteScalingPolicy \(p. 3\)](#).

## Request Syntax

```
{
  "MaxResults": number,
  "NextToken": "string",
  "PolicyNames": [ "string" ],
  "ResourceId": "string",
  "ScalableDimension": "string",
  "ServiceNamespace": "string"
}
```

## Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters \(p. 76\)](#).

The request accepts the following data in JSON format.

### MaxResults (p. 25)

The maximum number of scalable targets. This value can be between 1 and 50. The default value is 50.

If this parameter is used, the operation returns up to `MaxResults` results at a time, along with a `NextToken` value. To get the next set of results, include the `NextToken` value in a subsequent call. If this parameter is not used, the operation returns up to 50 results and a `NextToken` value, if applicable.

Type: Integer

Required: No

### NextToken (p. 25)

The token for the next set of results.

Type: String

Pattern: [ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t ]\*

Required: No

### PolicyNames (p. 25)

The names of the scaling policies to describe.

Type: Array of strings

Length Constraints: Minimum length of 1. Maximum length of 1600.

Pattern: [ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t ]\*

Required: No

#### **ResourceId (p. 25)**

The identifier of the resource associated with the scaling policy. This string consists of the resource type and unique identifier. If you specify a scalable dimension, you must also specify a resource ID.

- ECS service - The resource type is `service` and the unique identifier is the cluster name and service name. Example: `service/default/sample-webapp`.
- Spot fleet request - The resource type is `spot-fleet-request` and the unique identifier is the Spot fleet request ID. Example: `spot-fleet-request/sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE`.
- EMR cluster - The resource type is `instancegroup` and the unique identifier is the cluster ID and instance group ID. Example: `instancegroup/j-2EEZNYKUA1NTV/ig-1791Y4E1L8YI0`.
- AppStream 2.0 fleet - The resource type is `fleet` and the unique identifier is the fleet name. Example: `fleet/sample-fleet`.
- DynamoDB table - The resource type is `table` and the unique identifier is the resource ID. Example: `table/my-table`.
- DynamoDB global secondary index - The resource type is `index` and the unique identifier is the resource ID. Example: `table/my-table/index/my-table-index`.
- Aurora DB cluster - The resource type is `cluster` and the unique identifier is the cluster name. Example: `cluster:my-db-cluster`.
- Amazon SageMaker endpoint variants - The resource type is `variant` and the unique identifier is the resource ID. Example: `endpoint/my-end-point/variant/KMeansClustering`.
- Custom resources are not supported with a resource type. This parameter must specify the `OutputValue` from the CloudFormation template stack used to access the resources. The unique identifier is defined by the service provider.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1600.

Pattern: `[ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*`

Required: No

#### **ScalableDimension (p. 25)**

The scalable dimension. This string consists of the service namespace, resource type, and scaling property. If you specify a scalable dimension, you must also specify a resource ID.

- `ecs:service:DesiredCount` - The desired task count of an ECS service.
- `ec2:spot-fleet-request:TargetCapacity` - The target capacity of a Spot fleet request.
- `elasticmapreduce:instancegroup:InstanceCount` - The instance count of an EMR Instance Group.
- `appstream:fleet:DesiredCapacity` - The desired capacity of an AppStream 2.0 fleet.
- `dynamodb:table:ReadCapacityUnits` - The provisioned read capacity for a DynamoDB table.
- `dynamodb:table:WriteCapacityUnits` - The provisioned write capacity for a DynamoDB table.
- `dynamodb:index:ReadCapacityUnits` - The provisioned read capacity for a DynamoDB global secondary index.
- `dynamodb:index:WriteCapacityUnits` - The provisioned write capacity for a DynamoDB global secondary index.
- `rds:cluster:ReadReplicaCount` - The count of Aurora Replicas in an Aurora DB cluster. Available for Aurora MySQL-compatible edition.
- `sagemaker:variant:DesiredInstanceCount` - The number of EC2 instances for an Amazon SageMaker model endpoint variant.

- `custom-resource:ResourceType:Property` - The scalable dimension for a custom resource provided by your own application or service.

Type: String

Valid Values: `ecs:service:DesiredCount` | `ec2:spot-fleet-request:TargetCapacity` | `elasticmapreduce:instancegroup:InstanceCount` | `appstream:fleet:DesiredCapacity` | `dynamodb:table:ReadCapacityUnits` | `dynamodb:table:WriteCapacityUnits` | `dynamodb:index:ReadCapacityUnits` | `dynamodb:index:WriteCapacityUnits` | `rds:cluster:ReadReplicaCount` | `sagemaker:variant:DesiredInstanceCount` | `custom-resource:ResourceType:Property`

Required: No

#### **ServiceNamespace (p. 25)**

The namespace of the AWS service that provides the resource or `custom-resource` for a resource provided by your own application or service. For more information, see [AWS Service Namespaces](#) in the *Amazon Web Services General Reference*.

Type: String

Valid Values: `ecs` | `elasticmapreduce` | `ec2` | `appstream` | `dynamodb` | `rds` | `sagemaker` | `custom-resource`

Required: Yes

## Response Syntax

```
{
  "NextToken": "string",
  "ScalingPolicies": [
    {
      "Alarms": [
        {
          "AlarmARN": "string",
          "AlarmName": "string"
        }
      ],
      "CreationTime": number,
      "PolicyARN": "string",
      "PolicyName": "string",
      "PolicyType": "string",
      "ResourceId": "string",
      "ScalableDimension": "string",
      "ServiceNamespace": "string",
      "StepScalingPolicyConfiguration": {
        "AdjustmentType": "string",
        "Cooldown": number,
        "MetricAggregationType": "string",
        "MinAdjustmentMagnitude": number,
        "StepAdjustments": [
          {
            "MetricIntervalLowerBound": number,
            "MetricIntervalUpperBound": number,
            "ScalingAdjustment": number
          }
        ]
      },
      "TargetTrackingScalingPolicyConfiguration": {
```



```
    "CustomizedMetricSpecification": {
      "Dimensions": [
        {
          "Name": "string",
          "Value": "string"
        }
      ],
      "MetricName": "string",
      "Namespace": "string",
      "Statistic": "string",
      "Unit": "string"
    },
    "DisableScaleIn": boolean,
    "PredefinedMetricSpecification": {
      "PredefinedMetricType": "string",
      "ResourceLabel": "string"
    },
    "ScaleInCooldown": number,
    "ScaleOutCooldown": number,
    "TargetValue": number
  }
}
```

## Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

### NextToken (p. 27)

The token required to get the next set of results. This value is null if there are no more results to return.

Type: String

Pattern: [ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]\*

### ScalingPolicies (p. 27)

Information about the scaling policies.

Type: Array of [ScalingPolicy \(p. 64\)](#) objects

## Errors

For information about the errors that are common to all actions, see [Common Errors \(p. 78\)](#).

### ConcurrentUpdateException

Concurrent updates caused an exception, for example, if you request an update to an Application Auto Scaling resource that already has a pending update.

HTTP Status Code: 400

### FailedResourceAccessException

Failed access to resources caused an exception. This exception is thrown when Application Auto Scaling is unable to retrieve the alarms associated with a scaling policy due to a client error,

for example, if the role ARN specified for a scalable target does not have permission to call the CloudWatch [DescribeAlarms](#) on your behalf.

HTTP Status Code: 400

#### **InternalServiceException**

The service encountered an internal error.

HTTP Status Code: 400

#### **InvalidNextTokenException**

The next token supplied was invalid.

HTTP Status Code: 400

#### **ValidationException**

An exception was thrown for a validation issue. Review the available parameters for the API request.

HTTP Status Code: 400

## Example

If you plan to create requests manually, you must replace the Authorization header contents in the examples (AUTHPARAMS) with a signature. For more information, see [Signature Version 4 Signing Process](#) in the *Amazon Web Services General Reference*. If you plan to use the [AWS CLI](#) or one of the [AWS SDKs](#), these tools sign the requests for you.

## Example

The following example describes the scaling policies for the `ecs` service namespace.

### Sample Request

```
POST / HTTP/1.1
Host: autoscaling.us-west-2.amazonaws.com
Accept-Encoding: identity
Content-Length: 27
X-Amz-Target: AnyScaleFrontendService.DescribeScalingPolicies
X-Amz-Date: 20160506T194435Z
User-Agent: aws-cli/1.10.23 Python/2.7.11 Darwin/15.4.0 botocore/1.4.8
Content-Type: application/x-amz-json-1.1
Authorization: AUTHPARAMS

{
  "ServiceNamespace": "ecs"
}
```

### Sample Response

```
HTTP/1.1 200 OK
x-amzn-RequestId: f662c515-13c2-11e6-add4-41b78770ca43
Content-Type: application/x-amz-json-1.1
Content-Length: 1363
Date: Fri, 06 May 2016 19:44:35 GMT

{
  "ScalingPolicies": [
```

```
{
  "Alarms": [
    {
      "AlarmARN": "arn:aws:cloudwatch:us-west-2:012345678910:alarm:web-app-
cpu-gt-75",
      "AlarmName": "web-app-cpu-gt-75"
    }
  ],
  "CreationTime": 1462561899.23,
  "PolicyARN": "arn:aws:autoscaling:us-
west-2:012345678910:scalingPolicy:6d8972f3-efc8-437c-92d1-6270f29a66e7:resource/ecs/
service/default/web-app:policyName/web-app-cpu-gt-75",
  "PolicyName": "web-app-cpu-gt-75",
  "PolicyType": "StepScaling",
  "ResourceId": "service/default/web-app",
  "ScalableDimension": "ecs:service:DesiredCount",
  "ServiceNamespace": "ecs",
  "StepScalingPolicyConfiguration": {
    "AdjustmentType": "PercentChangeInCapacity",
    "Cooldown": 60,
    "StepAdjustments": [
      {
        "MetricIntervalLowerBound": 0,
        "ScalingAdjustment": 200
      }
    ]
  }
},
{
  "Alarms": [
    {
      "AlarmARN": "arn:aws:cloudwatch:us-west-2:012345678910:alarm:web-app-
cpu-lt-25",
      "AlarmName": "web-app-cpu-lt-25"
    }
  ],
  "CreationTime": 1462562575.099,
  "PolicyARN": "arn:aws:autoscaling:us-
west-2:012345678910:scalingPolicy:6d8972f3-efc8-437c-92d1-6270f29a66e7:resource/ecs/
service/default/web-app:policyName/web-app-cpu-lt-25",
  "PolicyName": "web-app-cpu-lt-25",
  "PolicyType": "StepScaling",
  "ResourceId": "service/default/web-app",
  "ScalableDimension": "ecs:service:DesiredCount",
  "ServiceNamespace": "ecs",
  "StepScalingPolicyConfiguration": {
    "AdjustmentType": "PercentChangeInCapacity",
    "Cooldown": 1,
    "StepAdjustments": [
      {
        "MetricIntervalUpperBound": 0,
        "ScalingAdjustment": -50
      }
    ]
  }
}
]
```

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)

- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for JavaScript](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V2](#)

# DescribeScheduledActions

Describes the scheduled actions for the specified service namespace.

You can filter the results using the `ResourceId`, `ScalableDimension`, and `ScheduledActionNames` parameters.

To create a scheduled action or update an existing one, see [PutScheduledAction \(p. 43\)](#). If you are no longer using a scheduled action, you can delete it using [DeleteScheduledAction \(p. 7\)](#).

## Request Syntax

```
{
  "MaxResults": number,
  "NextToken": "string",
  "ResourceId": "string",
  "ScalableDimension": "string",
  "ScheduledActionNames": [ "string" ],
  "ServiceNamespace": "string"
}
```

## Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters \(p. 76\)](#).

The request accepts the following data in JSON format.

### MaxResults (p. 32)

The maximum number of scheduled action results. This value can be between 1 and 50. The default value is 50.

If this parameter is used, the operation returns up to `MaxResults` results at a time, along with a `NextToken` value. To get the next set of results, include the `NextToken` value in a subsequent call. If this parameter is not used, the operation returns up to 50 results and a `NextToken` value, if applicable.

Type: Integer

Required: No

### NextToken (p. 32)

The token for the next set of results.

Type: String

Pattern: `[ \u0020-\uD7FF\uE000-\uFFFF\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*`

Required: No

### ResourceId (p. 32)

The identifier of the resource associated with the scheduled action. This string consists of the resource type and unique identifier. If you specify a scalable dimension, you must also specify a resource ID.

- ECS service - The resource type is `service` and the unique identifier is the cluster name and service name. Example: `service/default/sample-webapp`.

- Spot fleet request - The resource type is `spot-fleet-request` and the unique identifier is the Spot fleet request ID. Example: `spot-fleet-request/sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE`.
- EMR cluster - The resource type is `instancegroup` and the unique identifier is the cluster ID and instance group ID. Example: `instancegroup/j-2EEZNYKUA1NTV/ig-1791Y4E1L8YI0`.
- AppStream 2.0 fleet - The resource type is `fleet` and the unique identifier is the fleet name. Example: `fleet/sample-fleet`.
- DynamoDB table - The resource type is `table` and the unique identifier is the resource ID. Example: `table/my-table`.
- DynamoDB global secondary index - The resource type is `index` and the unique identifier is the resource ID. Example: `table/my-table/index/my-table-index`.
- Aurora DB cluster - The resource type is `cluster` and the unique identifier is the cluster name. Example: `cluster:my-db-cluster`.
- Amazon SageMaker endpoint variants - The resource type is `variant` and the unique identifier is the resource ID. Example: `endpoint/my-end-point/variant/KMeansClustering`.
- Custom resources are not supported with a resource type. This parameter must specify the `OutputValue` from the CloudFormation template stack used to access the resources. The unique identifier is defined by the service provider.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1600.

Pattern: `[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\x\r\n\t]*`

Required: No

#### ScalableDimension (p. 32)

The scalable dimension. This string consists of the service namespace, resource type, and scaling property. If you specify a scalable dimension, you must also specify a resource ID.

- `ecs:service:DesiredCount` - The desired task count of an ECS service.
- `ec2:spot-fleet-request:TargetCapacity` - The target capacity of a Spot fleet request.
- `elasticmapreduce:instancegroup:InstanceCount` - The instance count of an EMR Instance Group.
- `appstream:fleet:DesiredCapacity` - The desired capacity of an AppStream 2.0 fleet.
- `dynamodb:table:ReadCapacityUnits` - The provisioned read capacity for a DynamoDB table.
- `dynamodb:table:WriteCapacityUnits` - The provisioned write capacity for a DynamoDB table.
- `dynamodb:index:ReadCapacityUnits` - The provisioned read capacity for a DynamoDB global secondary index.
- `dynamodb:index:WriteCapacityUnits` - The provisioned write capacity for a DynamoDB global secondary index.
- `rds:cluster:ReadReplicaCount` - The count of Aurora Replicas in an Aurora DB cluster. Available for Aurora MySQL-compatible edition.
- `sagemaker:variant:DesiredInstanceCount` - The number of EC2 instances for an Amazon SageMaker model endpoint variant.
- `custom-resource:ResourceType:Property` - The scalable dimension for a custom resource provided by your own application or service.

Type: String

Valid Values: `ecs:service:DesiredCount` | `ec2:spot-fleet-request:TargetCapacity` | `elasticmapreduce:instancegroup:InstanceCount` | `appstream:fleet:DesiredCapacity` | `dynamodb:table:ReadCapacityUnits`

```
| dynamodb:table:WriteCapacityUnits | dynamodb:index:ReadCapacityUnits  
| dynamodb:index:WriteCapacityUnits | rds:cluster:ReadReplicaCount  
| sagemaker:variant:DesiredInstanceCount | custom-  
resource:ResourceType:Property
```

Required: No

#### **ScheduledActionNames (p. 32)**

The names of the scheduled actions to describe.

Type: Array of strings

Length Constraints: Minimum length of 1. Maximum length of 1600.

Pattern: [ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]\*

Required: No

#### **ServiceNamespace (p. 32)**

The namespace of the AWS service that provides the resource or custom-resource for a resource provided by your own application or service. For more information, see [AWS Service Namespaces](#) in the *Amazon Web Services General Reference*.

Type: String

Valid Values: ecs | elasticmapreduce | ec2 | appstream | dynamodb | rds |  
sagemaker | custom-resource

Required: Yes

## Response Syntax

```
{  
  "NextToken": "string",  
  "ScheduledActions": [  
    {  
      "CreationTime": number,  
      "EndTime": number,  
      "ResourceId": "string",  
      "ScalableDimension": "string",  
      "ScalableTargetAction": {  
        "MaxCapacity": number,  
        "MinCapacity": number  
      },  
      "Schedule": "string",  
      "ScheduledActionARN": "string",  
      "ScheduledActionName": "string",  
      "ServiceNamespace": "string",  
      "StartTime": number  
    },  
    ...  
  ]  
}
```

## Response Elements

If the action is successful, the service sends back an HTTP 200 response.

The following data is returned in JSON format by the service.

### **NextToken (p. 34)**

The token required to get the next set of results. This value is `null` if there are no more results to return.

Type: String

Pattern: `[ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*`

### **ScheduledActions (p. 34)**

Information about the scheduled actions.

Type: Array of [ScheduledAction \(p. 67\)](#) objects

## **Errors**

For information about the errors that are common to all actions, see [Common Errors \(p. 78\)](#).

### **ConcurrentUpdateException**

Concurrent updates caused an exception, for example, if you request an update to an Application Auto Scaling resource that already has a pending update.

HTTP Status Code: 400

### **InternalServiceException**

The service encountered an internal error.

HTTP Status Code: 400

### **InvalidNextTokenException**

The next token supplied was invalid.

HTTP Status Code: 400

### **ValidationException**

An exception was thrown for a validation issue. Review the available parameters for the API request.

HTTP Status Code: 400

## **See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for JavaScript](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V2](#)





# PutScalingPolicy

Creates or updates a policy for an Application Auto Scaling scalable target.

Each scalable target is identified by a service namespace, resource ID, and scalable dimension. A scaling policy applies to the scalable target identified by those three attributes. You cannot create a scaling policy until you register the scalable target using [RegisterScalableTarget](#) (p. 48).

To update a policy, specify its policy name and the parameters that you want to change. Any parameters that you don't specify are not changed by this update request.

You can view the scaling policies for a service namespace using [DescribeScalingPolicies](#) (p. 25). If you are no longer using a scaling policy, you can delete it using [DeleteScalingPolicy](#) (p. 3).

## Request Syntax

```
{
  "PolicyName": "string",
  "PolicyType": "string",
  "ResourceId": "string",
  "ScalableDimension": "string",
  "ServiceNamespace": "string",
  "StepScalingPolicyConfiguration": {
    "AdjustmentType": "string",
    "Cooldown": number,
    "MetricAggregationType": "string",
    "MinAdjustmentMagnitude": number,
    "StepAdjustments": [
      {
        "MetricIntervalLowerBound": number,
        "MetricIntervalUpperBound": number,
        "ScalingAdjustment": number
      }
    ]
  },
  "TargetTrackingScalingPolicyConfiguration": {
    "CustomizedMetricSpecification": {
      "Dimensions": [
        {
          "Name": "string",
          "Value": "string"
        }
      ],
      "MetricName": "string",
      "Namespace": "string",
      "Statistic": "string",
      "Unit": "string"
    },
    "DisableScaleIn": boolean,
    "PredefinedMetricSpecification": {
      "PredefinedMetricType": "string",
      "ResourceLabel": "string"
    },
    "ScaleInCooldown": number,
    "ScaleOutCooldown": number,
    "TargetValue": number
  }
}
```

## Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](#) (p. 76).

The request accepts the following data in JSON format.

### **PolicyName** (p. 37)

The name of the scaling policy.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: `\p{Print}+`

Required: Yes

### **PolicyType** (p. 37)

The policy type. This parameter is required if you are creating a policy.

For DynamoDB, only `TargetTrackingScaling` is supported. For Amazon ECS, Spot Fleet, and Amazon RDS, both `StepScaling` and `TargetTrackingScaling` are supported. For any other service, only `StepScaling` is supported.

Type: String

Valid Values: `StepScaling` | `TargetTrackingScaling`

Required: No

### **ResourceId** (p. 37)

The identifier of the resource associated with the scaling policy. This string consists of the resource type and unique identifier.

- ECS service - The resource type is `service` and the unique identifier is the cluster name and service name. Example: `service/default/sample-webapp`.
- Spot fleet request - The resource type is `spot-fleet-request` and the unique identifier is the Spot fleet request ID. Example: `spot-fleet-request/sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE`.
- EMR cluster - The resource type is `instancegroup` and the unique identifier is the cluster ID and instance group ID. Example: `instancegroup/j-2EEZNYKUA1NTV/ig-1791Y4E1L8YI0`.
- AppStream 2.0 fleet - The resource type is `fleet` and the unique identifier is the fleet name. Example: `fleet/sample-fleet`.
- DynamoDB table - The resource type is `table` and the unique identifier is the resource ID. Example: `table/my-table`.
- DynamoDB global secondary index - The resource type is `index` and the unique identifier is the resource ID. Example: `table/my-table/index/my-table-index`.
- Aurora DB cluster - The resource type is `cluster` and the unique identifier is the cluster name. Example: `cluster:my-db-cluster`.
- Amazon SageMaker endpoint variants - The resource type is `variant` and the unique identifier is the resource ID. Example: `endpoint/my-end-point/variant/KMeansClustering`.
- Custom resources are not supported with a resource type. This parameter must specify the `OutputValue` from the CloudFormation template stack used to access the resources. The unique identifier is defined by the service provider.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1600.

Pattern: [ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]\*

Required: Yes

#### ScalableDimension (p. 37)

The scalable dimension. This string consists of the service namespace, resource type, and scaling property.

- `ecs:service:DesiredCount` - The desired task count of an ECS service.
- `ec2:spot-fleet-request:TargetCapacity` - The target capacity of a Spot fleet request.
- `elasticmapreduce:instancegroup:InstanceCount` - The instance count of an EMR Instance Group.
- `appstream:fleet:DesiredCapacity` - The desired capacity of an AppStream 2.0 fleet.
- `dynamodb:table:ReadCapacityUnits` - The provisioned read capacity for a DynamoDB table.
- `dynamodb:table:WriteCapacityUnits` - The provisioned write capacity for a DynamoDB table.
- `dynamodb:index:ReadCapacityUnits` - The provisioned read capacity for a DynamoDB global secondary index.
- `dynamodb:index:WriteCapacityUnits` - The provisioned write capacity for a DynamoDB global secondary index.
- `rds:cluster:ReadReplicaCount` - The count of Aurora Replicas in an Aurora DB cluster. Available for Aurora MySQL-compatible edition.
- `sagemaker:variant:DesiredInstanceCount` - The number of EC2 instances for an Amazon SageMaker model endpoint variant.
- `custom-resource:ResourceType:Property` - The scalable dimension for a custom resource provided by your own application or service.

Type: String

Valid Values: `ecs:service:DesiredCount` | `ec2:spot-fleet-request:TargetCapacity` | `elasticmapreduce:instancegroup:InstanceCount` | `appstream:fleet:DesiredCapacity` | `dynamodb:table:ReadCapacityUnits` | `dynamodb:table:WriteCapacityUnits` | `dynamodb:index:ReadCapacityUnits` | `dynamodb:index:WriteCapacityUnits` | `rds:cluster:ReadReplicaCount` | `sagemaker:variant:DesiredInstanceCount` | `custom-resource:ResourceType:Property`

Required: Yes

#### ServiceNamespace (p. 37)

The namespace of the AWS service that provides the resource or `custom-resource` for a resource provided by your own application or service. For more information, see [AWS Service Namespaces](#) in the *Amazon Web Services General Reference*.

Type: String

Valid Values: `ecs` | `elasticmapreduce` | `ec2` | `appstream` | `dynamodb` | `rds` | `sagemaker` | `custom-resource`

Required: Yes

#### StepScalingPolicyConfiguration (p. 37)

A step scaling policy.

Type: [StepScalingPolicyConfiguration](#) (p. 72) object

## TargetTrackingScalingPolicyConfiguration (p. 37)

This parameter is required if you are creating a policy and the policy type is `TargetTrackingScaling`.

Required: No

```
{
  "Alarms": [
    {
      "AlarmARN": "string",
      "AlarmName": "string"
    }
  ],
  "PolicyARN": "string"
}
```

If the action is successful, the service sends back an HTTP 200 response.

## Alarms (p. 40)

Type: Array of [Alarm](#) (p. 53) objects

The Amazon Resource Name (ARN) of the resulting scaling policy.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1600.

**Pattern:** [\u0020-\uD7FF\uE000-\uFFFF\uD800\uDC00-\uDBFF\uDFFF\r\n\t]\*

For information about the errors that are common to all actions, see [Common Errors \(p. 78\)](#).

Concurrent updates caused an exception, for example, if you request an update to an Application Auto Scaling resource that already has a pending update.

HTTP Status Code: 400

#### **FailedResourceAccessException**

Failed access to resources caused an exception. This exception is thrown when Application Auto Scaling is unable to retrieve the alarms associated with a scaling policy due to a client error, for example, if the role ARN specified for a scalable target does not have permission to call the CloudWatch [DescribeAlarms](#) on your behalf.

HTTP Status Code: 400

#### **InternalServiceException**

The service encountered an internal error.

HTTP Status Code: 400

#### **LimitExceededException**

A per-account resource limit is exceeded. For more information, see [Application Auto Scaling Limits](#).

HTTP Status Code: 400

#### **ObjectNotFoundException**

The specified object could not be found. For any operation that depends on the existence of a scalable target, this exception is thrown if the scalable target with the specified service namespace, resource ID, and scalable dimension does not exist. For any operation that deletes or deregisters a resource, this exception is thrown if the resource cannot be found.

HTTP Status Code: 400

#### **ValidationException**

An exception was thrown for a validation issue. Review the available parameters for the API request.

HTTP Status Code: 400

## Example

If you plan to create requests manually, you must replace the Authorization header contents in the examples (AUTHPARAMS) with a signature. For more information, see [Signature Version 4 Signing Process](#) in the *Amazon Web Services General Reference*. If you plan to use the [AWS CLI](#) or one of the [AWS SDKs](#), these tools sign the requests for you.

## Example

The following example applies a scaling policy to an Amazon ECS service called `web-app` in the `default` cluster. The policy increases the desired count of the service by 200%, with a cool down period of 60 seconds.

### Sample Request

```
POST / HTTP/1.1
Host: autoscaling.us-west-2.amazonaws.com
Accept-Encoding: identity
Content-Length: 358
X-Amz-Target: AnyScaleFrontendService.PutScalingPolicy
X-Amz-Date: 20160506T191138Z
User-Agent: aws-cli/1.10.23 Python/2.7.11 Darwin/15.4.0 botocore/1.4.8
Content-Type: application/x-amz-json-1.1
Authorization: AUTHPARAMS
```

```
{
  "PolicyName": "web-app-cpu-gt-75",
  "ScalableDimension": "ecs:service:DesiredCount",
  "ResourceId": "service/default/web-app",
  "StepScalingPolicyConfiguration": {
    "Cooldown": 60,
    "StepAdjustments": [
      {
        "ScalingAdjustment": 200,
        "MetricIntervalLowerBound": 0
      }
    ],
    "AdjustmentType": "PercentChangeInCapacity"
  },
  "PolicyType": "StepScaling",
  "ServiceNamespace": "ecs"
}
```

## Sample Response

```
HTTP/1.1 200 OK
x-amzn-RequestId: 5bc8d06e-13be-11e6-a468-37acb4b5a1b2
Content-Type: application/x-amz-json-1.1
Content-Length: 175
Date: Fri, 06 May 2016 19:11:38 GMT

{
  "PolicyARN": "arn:aws:autoscaling:us-west-2:012345678910:scalingPolicy:6d8972f3-efc8-437c-92d1-6270f29a66e7:resource/ecs/service/default/web-app:policyName/web-app-cpu-gt-75"
}
```

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for JavaScript](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V2](#)

# PutScheduledAction

Creates or updates a scheduled action for an Application Auto Scaling scalable target.

Each scalable target is identified by a service namespace, resource ID, and scalable dimension. A scheduled action applies to the scalable target identified by those three attributes. You cannot create a scheduled action until you register the scalable target using [RegisterScalableTarget](#) (p. 48).

To update an action, specify its name and the parameters that you want to change. If you don't specify start and end times, the old values are deleted. Any other parameters that you don't specify are not changed by this update request.

You can view the scheduled actions using [DescribeScheduledActions](#) (p. 32). If you are no longer using a scheduled action, you can delete it using [DeleteScheduledAction](#) (p. 7).

## Request Syntax

```
{
  "EndTime": number,
  "ResourceId": "string",
  "ScalableDimension": "string",
  "ScalableTargetAction": {
    "MaxCapacity": number,
    "MinCapacity": number
  },
  "Schedule": "string",
  "ScheduledActionName": "string",
  "ServiceNamespace": "string",
  "StartTime": number
}
```

## Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters](#) (p. 76).

The request accepts the following data in JSON format.

### [EndTime](#) (p. 43)

The date and time for the scheduled action to end.

Type: Timestamp

Required: No

### [ResourceId](#) (p. 43)

The identifier of the resource associated with the scheduled action. This string consists of the resource type and unique identifier.

- ECS service - The resource type is `service` and the unique identifier is the cluster name and service name. Example: `service/default/sample-webapp`.
- Spot fleet request - The resource type is `spot-fleet-request` and the unique identifier is the Spot fleet request ID. Example: `spot-fleet-request/sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE`.
- EMR cluster - The resource type is `instancegroup` and the unique identifier is the cluster ID and instance group ID. Example: `instancegroup/j-2EEZNYKUA1NTV/ig-1791Y4E1L8YI0`.





Required: No

## ScalableTargetAction (p. 43)

The new minimum and maximum capacity. You can set both values or just one. During the scheduled time, if the current capacity is below the minimum capacity, Application Auto Scaling scales out to the minimum capacity. If the current capacity is above the maximum capacity, Application Auto Scaling scales in to the maximum capacity.

Type: [ScalableTargetAction](#) (p. 60) object

Required: No

**Schedule (p. 43)**

The schedule for this action. The following formats are supported:

- At expressions - `at(yyyy-mm-ddThh:mm:ss)`
- Rate expressions - `rate(value unit)`
- Cron expressions - `cron(fields)`

At expressions are useful for one-time schedules. Specify the time, in UTC.

For rate expressions, *value* is a positive integer and *unit* is minute | minutes | hour | hours | day | days.

For more information about cron expressions, see [Cron Expressions](#) in the *Amazon CloudWatch Events User Guide*.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1600.

Pattern: `[ \u0020-\u0D7FF\u0E000-\uFFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*`

Required: No

## ScheduledActionName (p. 43)

The name of the scheduled action.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: `(?!((^[ ]+.*)|(.*(["\u0000-\u001f"]|["\u007f-\u009f"]|[:/|])+.*)|(.*[ ]+$)))+`

Required: Yes

## ServiceNamespace (p. 43)

The namespace of the AWS service that provides the resource or custom-resource for a resource provided by your own application or service. For more information, see [AWS Service Namespaces](#) in the *Amazon Web Services General Reference*.

Type: String

Valid Values: ecs | elasticmapreduce | ec2 | appstream | dynamodb | rds |  
sagemaker | custom-resource

Required: Yes

### StartTime (p. 43)

The date and time for the scheduled action to start.

Type: Timestamp

Required: No

## Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

## Errors

For information about the errors that are common to all actions, see [Common Errors \(p. 78\)](#).

### **ConcurrentUpdateException**

Concurrent updates caused an exception, for example, if you request an update to an Application Auto Scaling resource that already has a pending update.

HTTP Status Code: 400

### **InternalServiceException**

The service encountered an internal error.

HTTP Status Code: 400

### **LimitExceededException**

A per-account resource limit is exceeded. For more information, see [Application Auto Scaling Limits](#).

HTTP Status Code: 400

### **ObjectNotFoundException**

The specified object could not be found. For any operation that depends on the existence of a scalable target, this exception is thrown if the scalable target with the specified service namespace, resource ID, and scalable dimension does not exist. For any operation that deletes or deregisters a resource, this exception is thrown if the resource cannot be found.

HTTP Status Code: 400

### **ValidationException**

An exception was thrown for a validation issue. Review the available parameters for the API request.

HTTP Status Code: 400

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for JavaScript](#)
- [AWS SDK for PHP V3](#)

- [AWS SDK for Python](#)
- [AWS SDK for Ruby V2](#)

# RegisterScalableTarget

Registers or updates a scalable target. A scalable target is a resource that Application Auto Scaling can scale out or scale in. After you have registered a scalable target, you can use this operation to update the minimum and maximum values for its scalable dimension.

After you register a scalable target, you can create and apply scaling policies using [PutScalingPolicy \(p. 37\)](#). You can view the scaling policies for a service namespace using [DescribeScalableTargets \(p. 14\)](#). If you no longer need a scalable target, you can deregister it using [DeregisterScalableTarget \(p. 10\)](#).

## Request Syntax

```
{
  "MaxCapacity": number,
  "MinCapacity": number,
  "ResourceId": "string",
  "RoleARN": "string",
  "ScalableDimension": "string",
  "ServiceNamespace": "string"
}
```

## Request Parameters

For information about the parameters that are common to all actions, see [Common Parameters \(p. 76\)](#).

The request accepts the following data in JSON format.

### MaxCapacity (p. 48)

The maximum value to scale to in response to a scale out event. This parameter is required if you are registering a scalable target.

Type: Integer

Required: No

### MinCapacity (p. 48)

The minimum value to scale to in response to a scale in event. This parameter is required if you are registering a scalable target.

Type: Integer

Required: No

### ResourceId (p. 48)

The identifier of the resource associated with the scalable target. This string consists of the resource type and unique identifier.

- ECS service - The resource type is `service` and the unique identifier is the cluster name and service name. Example: `service/default/sample-webapp`.
- Spot fleet request - The resource type is `spot-fleet-request` and the unique identifier is the Spot fleet request ID. Example: `spot-fleet-request/sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE`.
- EMR cluster - The resource type is `instancegroup` and the unique identifier is the cluster ID and instance group ID. Example: `instancegroup/j-2EEZNYKUA1NTV/ig-1791Y4E1L8YI0`.



- `sagemaker:variant:DesiredInstanceCount` - The number of EC2 instances for an Amazon SageMaker model endpoint variant.
- `custom-resource:ResourceType:Property` - The scalable dimension for a custom resource provided by your own application or service.

Type: String

Valid Values: `ecs:service:DesiredCount` | `ec2:spot-fleet-request:TargetCapacity` | `elasticmapreduce:instancegroup:InstanceCount` | `appstream:fleet:DesiredCapacity` | `dynamodb:table:ReadCapacityUnits` | `dynamodb:table:WriteCapacityUnits` | `dynamodb:index:ReadCapacityUnits` | `dynamodb:index:WriteCapacityUnits` | `rds:cluster:ReadReplicaCount` | `sagemaker:variant:DesiredInstanceCount` | `custom-resource:ResourceType:Property`

Required: Yes

#### [ServiceNamespace \(p. 48\)](#)

The namespace of the AWS service that provides the resource or `custom-resource` for a resource provided by your own application or service. For more information, see [AWS Service Namespaces](#) in the *Amazon Web Services General Reference*.

Type: String

Valid Values: `ecs` | `elasticmapreduce` | `ec2` | `appstream` | `dynamodb` | `rds` | `sagemaker` | `custom-resource`

Required: Yes

## Response Elements

If the action is successful, the service sends back an HTTP 200 response with an empty HTTP body.

## Errors

For information about the errors that are common to all actions, see [Common Errors \(p. 78\)](#).

### **ConcurrentUpdateException**

Concurrent updates caused an exception, for example, if you request an update to an Application Auto Scaling resource that already has a pending update.

HTTP Status Code: 400

### **InternalServiceException**

The service encountered an internal error.

HTTP Status Code: 400

### **LimitExceededException**

A per-account resource limit is exceeded. For more information, see [Application Auto Scaling Limits](#).

HTTP Status Code: 400

### **ValidationException**

An exception was thrown for a validation issue. Review the available parameters for the API request.

HTTP Status Code: 400

## Example

If you plan to create requests manually, you must replace the Authorization header contents in the examples (AUTHPARAMS) with a signature. For more information, see [Signature Version 4 Signing Process](#) in the *Amazon Web Services General Reference*. If you plan to use the [AWS CLI](#) or one of the [AWS SDKs](#), these tools sign the requests for you.

## Example

The following example registers an Amazon ECS service with Application Auto Scaling.

### Sample Request

```
POST / HTTP/1.1
Host: autoscaling.us-west-2.amazonaws.com
Accept-Encoding: identity
Content-Length: 229
X-Amz-Target: AnyScaleFrontendService.RegisterScalableTarget
X-Amz-Date: 20160506T182145Z
User-Agent: aws-cli/1.10.23 Python/2.7.11 Darwin/15.4.0 botocore/1.4.8
Content-Type: application/x-amz-json-1.1
Authorization: AUTHPARAMS

{
  "ScalableDimension": "ecs:service:DesiredCount",
  "ResourceId": "service/default/web-app",
  "RoleARN": "arn:aws:iam::012345678910:role/ApplicationAutoscalingECSRole",
  "MinCapacity": 1,
  "ServiceNamespace": "ecs",
  "MaxCapacity": 10
}
```

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS Command Line Interface](#)
- [AWS SDK for .NET](#)
- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for JavaScript](#)
- [AWS SDK for PHP V3](#)
- [AWS SDK for Python](#)
- [AWS SDK for Ruby V2](#)



# Data Types

The Application Auto Scaling API contains several data types that various actions use. This section describes each data type in detail.

**Note**

The order of each element in a data type structure is not guaranteed. Applications should not assume a particular order.

The following data types are supported:

- [Alarm](#) (p. 53)
- [CustomizedMetricSpecification](#) (p. 54)
- [MetricDimension](#) (p. 55)
- [PredefinedMetricSpecification](#) (p. 56)
- [ScalableTarget](#) (p. 57)
- [ScalableTargetAction](#) (p. 60)
- [ScalingActivity](#) (p. 61)
- [ScalingPolicy](#) (p. 64)
- [ScheduledAction](#) (p. 67)
- [StepAdjustment](#) (p. 70)
- [StepScalingPolicyConfiguration](#) (p. 72)
- [TargetTrackingScalingPolicyConfiguration](#) (p. 74)

# Alarm

Represents a CloudWatch alarm associated with a scaling policy.

## Contents

### **AlarmARN**

The Amazon Resource Name (ARN) of the alarm.

Type: String

Pattern: `[ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*`

Required: Yes

### **AlarmName**

The name of the alarm.

Type: String

Pattern: `[ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*`

Required: Yes

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for Ruby V2](#)

# CustomizedMetricSpecification

Configures a customized metric for a target tracking policy.

## Contents

### Dimensions

The dimensions of the metric.

Type: Array of [MetricDimension](#) (p. 55) objects

Required: No

### MetricName

The name of the metric.

Type: String

Required: Yes

### Namespace

The namespace of the metric.

Type: String

Required: Yes

### Statistic

The statistic of the metric.

Type: String

Valid Values: `Average` | `Minimum` | `Maximum` | `SampleCount` | `Sum`

Required: Yes

### Unit

The unit of the metric.

Type: String

Required: No

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for Ruby V2](#)

# MetricDimension

Describes the dimension of a metric.

## Contents

### Name

The name of the dimension.

Type: String

Required: Yes

### Value

The value of the dimension.

Type: String

Required: Yes

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for Ruby V2](#)

# PredefinedMetricSpecification

Configures a predefined metric for a target tracking policy.

## Contents

### PredefinedMetricType

The metric type. The `ALBRequestCountPerTarget` metric type applies only to Spot fleet requests and ECS services.

Type: String

Valid Values: `DynamoDBReadCapacityUtilization` | `DynamoDBWriteCapacityUtilization` | `ALBRequestCountPerTarget` | `RDSReaderAverageCPUUtilization` | `RDSReaderAverageDatabaseConnections` | `EC2SpotFleetRequestAverageCPUUtilization` | `EC2SpotFleetRequestAverageNetworkIn` | `EC2SpotFleetRequestAverageNetworkOut` | `SageMakerVariantInvocationsPerInstance` | `ECSServiceAverageCPUUtilization` | `ECSServiceAverageMemoryUtilization`

Required: Yes

### ResourceLabel

Identifies the resource associated with the metric type. You can't specify a resource label unless the metric type is `ALBRequestCountPerTarget` and there is a target group attached to the Spot fleet request or ECS service.

The format is `app/<load-balancer-name>/<load-balancer-id>/targetgroup/<target-group-name>/<target-group-id>`, where:

- `app/<load-balancer-name>/<load-balancer-id>` is the final portion of the load balancer ARN
- `targetgroup/<target-group-name>/<target-group-id>` is the final portion of the target group ARN.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1023.

Required: No

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for Ruby V2](#)

# ScalableTarget

Represents a scalable target.

## Contents

### CreationTime

The Unix timestamp for when the scalable target was created.

Type: Timestamp

Required: Yes

### MaxCapacity

The maximum value to scale to in response to a scale out event.

Type: Integer

Required: Yes

### MinCapacity

The minimum value to scale to in response to a scale in event.

Type: Integer

Required: Yes

### ResourceId

The identifier of the resource associated with the scalable target. This string consists of the resource type and unique identifier.

- ECS service - The resource type is `service` and the unique identifier is the cluster name and service name. Example: `service/default/sample-webapp`.
- Spot fleet request - The resource type is `spot-fleet-request` and the unique identifier is the Spot fleet request ID. Example: `spot-fleet-request/sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE`.
- EMR cluster - The resource type is `instancegroup` and the unique identifier is the cluster ID and instance group ID. Example: `instancegroup/j-2EEZNYKUA1NTV/ig-1791Y4E1L8YI0`.
- AppStream 2.0 fleet - The resource type is `fleet` and the unique identifier is the fleet name. Example: `fleet/sample-fleet`.
- DynamoDB table - The resource type is `table` and the unique identifier is the resource ID. Example: `table/my-table`.
- DynamoDB global secondary index - The resource type is `index` and the unique identifier is the resource ID. Example: `table/my-table/index/my-table-index`.
- Aurora DB cluster - The resource type is `cluster` and the unique identifier is the cluster name. Example: `cluster:my-db-cluster`.
- Amazon SageMaker endpoint variants - The resource type is `variant` and the unique identifier is the resource ID. Example: `endpoint/my-end-point/variant/KMeansClustering`.
- Custom resources are not supported with a resource type. This parameter must specify the `OutputValue` from the CloudFormation template stack used to access the resources. The unique identifier is defined by the service provider.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1600.

Pattern: [ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]\*

Required: Yes

#### **RoleARN**

The ARN of an IAM role that allows Application Auto Scaling to modify the scalable target on your behalf.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1600.

Pattern: [ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]\*

Required: Yes

#### **ScalableDimension**

The scalable dimension associated with the scalable target. This string consists of the service namespace, resource type, and scaling property.

- `ecs:service:DesiredCount` - The desired task count of an ECS service.
- `ec2:spot-fleet-request:TargetCapacity` - The target capacity of a Spot fleet request.
- `elasticmapreduce:instancegroup:InstanceCount` - The instance count of an EMR Instance Group.
- `appstream:fleet:DesiredCapacity` - The desired capacity of an AppStream 2.0 fleet.
- `dynamodb:table:ReadCapacityUnits` - The provisioned read capacity for a DynamoDB table.
- `dynamodb:table:WriteCapacityUnits` - The provisioned write capacity for a DynamoDB table.
- `dynamodb:index:ReadCapacityUnits` - The provisioned read capacity for a DynamoDB global secondary index.
- `dynamodb:index:WriteCapacityUnits` - The provisioned write capacity for a DynamoDB global secondary index.
- `rds:cluster:ReadReplicaCount` - The count of Aurora Replicas in an Aurora DB cluster. Available for Aurora MySQL-compatible edition.
- `sagemaker:variant:DesiredInstanceCount` - The number of EC2 instances for an Amazon SageMaker model endpoint variant.
- `custom-resource:ResourceType:Property` - The scalable dimension for a custom resource provided by your own application or service.

Type: String

Valid Values: `ecs:service:DesiredCount` | `ec2:spot-fleet-request:TargetCapacity` | `elasticmapreduce:instancegroup:InstanceCount` | `appstream:fleet:DesiredCapacity` | `dynamodb:table:ReadCapacityUnits` | `dynamodb:table:WriteCapacityUnits` | `dynamodb:index:ReadCapacityUnits` | `dynamodb:index:WriteCapacityUnits` | `rds:cluster:ReadReplicaCount` | `sagemaker:variant:DesiredInstanceCount` | `custom-resource:ResourceType:Property`

Required: Yes

#### **ServiceNamespace**

The namespace of the AWS service that provides the resource or `custom-resource` for a resource provided by your own application or service. For more information, see [AWS Service Namespaces](#) in the *Amazon Web Services General Reference*.

Type: String

Valid Values: `ecs` | `elasticmapreduce` | `ec2` | `appstream` | `dynamodb` | `rds` | `sagemaker` | `custom-resource`

Required: Yes

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for Ruby V2](#)



# ScalableTargetAction

Represents the minimum and maximum capacity for a scheduled action.

## Contents

### **MaxCapacity**

The maximum capacity.

Type: Integer

Required: No

### **MinCapacity**

The minimum capacity.

Type: Integer

Required: No

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for Ruby V2](#)

# ScalingActivity

Represents a scaling activity.

## Contents

### ActivityId

The unique identifier of the scaling activity.

Type: String

Pattern: [ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t ] \*

Required: Yes

### Cause

A simple description of what caused the scaling activity to happen.

Type: String

Pattern: [ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t ] \*

Required: Yes

### Description

A simple description of what action the scaling activity intends to accomplish.

Type: String

Pattern: [ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t ] \*

Required: Yes

### Details

The details about the scaling activity.

Type: String

Pattern: [ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t ] \*

Required: No

### EndTime

The Unix timestamp for when the scaling activity ended.

Type: Timestamp

Required: No

### ResourceId

The identifier of the resource associated with the scaling activity. This string consists of the resource type and unique identifier.

- ECS service - The resource type is `service` and the unique identifier is the cluster name and service name. Example: `service/default/sample-webapp`.
- Spot fleet request - The resource type is `spot-fleet-request` and the unique identifier is the Spot fleet request ID. Example: `spot-fleet-request/sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE`.

- Type: String

Pattern: `[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*`

Required: Yes

The scalable dimension. This string consists of the service namespace, resource type, and scaling property.

- Type: String

API Version 2016-02-06

| `sagemaker:variant:DesiredInstanceCount` | `custom-resource:ResourceType:Property`

Required: Yes

### **ServiceNamespace**

The namespace of the AWS service that provides the resource or `custom-resource` for a resource provided by your own application or service. For more information, see [AWS Service Namespaces](#) in the *Amazon Web Services General Reference*.

Type: String

Valid Values: `ecs` | `elasticmapreduce` | `ec2` | `appstream` | `dynamodb` | `rds` | `sagemaker` | `custom-resource`

Required: Yes

### **StartTime**

The Unix timestamp for when the scaling activity began.

Type: Timestamp

Required: Yes

### **StatusCode**

Indicates the status of the scaling activity.

Type: String

Valid Values: `Pending` | `InProgress` | `Successful` | `Overridden` | `Unfulfilled` | `Failed`

Required: Yes

### **StatusMessage**

A simple message about the current status of the scaling activity.

Type: String

Pattern: `[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*`

Required: No

## **See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for Ruby V2](#)





```
| dynamodb:table:WriteCapacityUnits | dynamodb:index:ReadCapacityUnits  
| dynamodb:index:WriteCapacityUnits | rds:cluster:ReadReplicaCount  
| sagemaker:variant:DesiredInstanceCount | custom-  
resource:ResourceType:Property
```

Required: Yes

### **ServiceNamespace**

The namespace of the AWS service that provides the resource or custom-resource for a resource provided by your own application or service. For more information, see [AWS Service Namespaces](#) in the *Amazon Web Services General Reference*.

Type: String

Valid Values: ecs | elasticmapreduce | ec2 | appstream | dynamodb | rds |  
sagemaker | custom-resource

Required: Yes

### **StepScalingPolicyConfiguration**

A step scaling policy.

Type: [StepScalingPolicyConfiguration](#) (p. 72) object

Required: No

### **TargetTrackingScalingPolicyConfiguration**

A target tracking policy.

Type: [TargetTrackingScalingPolicyConfiguration](#) (p. 74) object

Required: No

## **See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for Ruby V2](#)

# ScheduledAction

Represents a scheduled action.

## Contents

### CreationTime

The date and time that the scheduled action was created.

Type: Timestamp

Required: Yes

### EndTime

The date and time that the action is scheduled to end.

Type: Timestamp

Required: No

### ResourceId

The identifier of the resource associated with the scaling policy. This string consists of the resource type and unique identifier.

- ECS service - The resource type is `service` and the unique identifier is the cluster name and service name. Example: `service/default/sample-webapp`.
- Spot fleet request - The resource type is `spot-fleet-request` and the unique identifier is the Spot fleet request ID. Example: `spot-fleet-request/sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE`.
- EMR cluster - The resource type is `instancegroup` and the unique identifier is the cluster ID and instance group ID. Example: `instancegroup/j-2EEZNYKUA1NTV/ig-1791Y4E1L8YI0`.
- AppStream 2.0 fleet - The resource type is `fleet` and the unique identifier is the fleet name. Example: `fleet/sample-fleet`.
- DynamoDB table - The resource type is `table` and the unique identifier is the resource ID. Example: `table/my-table`.
- DynamoDB global secondary index - The resource type is `index` and the unique identifier is the resource ID. Example: `table/my-table/index/my-table-index`.
- Aurora DB cluster - The resource type is `cluster` and the unique identifier is the cluster name. Example: `cluster:my-db-cluster`.
- Amazon SageMaker endpoint variants - The resource type is `variant` and the unique identifier is the resource ID. Example: `endpoint/my-end-point/variant/KMeansClustering`.
- Custom resources are not supported with a resource type. This parameter must specify the `OutputValue` from the CloudFormation template stack used to access the resources. The unique identifier is defined by the service provider.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1600.

Pattern: `[ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*`

Required: Yes

### ScalableDimension

The scalable dimension. This string consists of the service namespace, resource type, and scaling property.



- `ecs:service:DesiredCount` - The desired task count of an ECS service.
- `ec2:spot-fleet-request:TargetCapacity` - The target capacity of a Spot fleet request.
- `elasticmapreduce:instancegroup:InstanceCount` - The instance count of an EMR Instance Group.
- `appstream:fleet:DesiredCapacity` - The desired capacity of an AppStream 2.0 fleet.
- `dynamodb:table:ReadCapacityUnits` - The provisioned read capacity for a DynamoDB table.
- `dynamodb:table:WriteCapacityUnits` - The provisioned write capacity for a DynamoDB table.
- `dynamodb:index:ReadCapacityUnits` - The provisioned read capacity for a DynamoDB global secondary index.
- `dynamodb:index:WriteCapacityUnits` - The provisioned write capacity for a DynamoDB global secondary index.
- `rds:cluster:ReadReplicaCount` - The count of Aurora Replicas in an Aurora DB cluster. Available for Aurora MySQL-compatible edition.
- `sagemaker:variant:DesiredInstanceCount` - The number of EC2 instances for an Amazon SageMaker model endpoint variant.
- `custom-resource:ResourceType:Property` - The scalable dimension for a custom resource provided by your own application or service.

Type: String

Valid Values: `ecs:service:DesiredCount` | `ec2:spot-fleet-request:TargetCapacity` | `elasticmapreduce:instancegroup:InstanceCount` | `appstream:fleet:DesiredCapacity` | `dynamodb:table:ReadCapacityUnits` | `dynamodb:table:WriteCapacityUnits` | `dynamodb:index:ReadCapacityUnits` | `dynamodb:index:WriteCapacityUnits` | `rds:cluster:ReadReplicaCount` | `sagemaker:variant:DesiredInstanceCount` | `custom-resource:ResourceType:Property`

Required: No

### ScalableTargetAction

The new minimum and maximum capacity. You can set both values or just one. During the scheduled time, if the current capacity is below the minimum capacity, Application Auto Scaling scales out to the minimum capacity. If the current capacity is above the maximum capacity, Application Auto Scaling scales in to the maximum capacity.

Type: [ScalableTargetAction](#) (p. 60) object

Required: No

### Schedule

The schedule for this action. The following formats are supported:

- At expressions - `at(yyyy-mm-ddThh:mm:ss)`
- Rate expressions - `rate(value unit)`
- Cron expressions - `cron(fields)`

At expressions are useful for one-time schedules. Specify the time, in UTC.

For rate expressions, *value* is a positive integer and *unit* is `minute` | `minutes` | `hour` | `hours` | `day` | `days`.

For more information about cron expressions, see [Cron Expressions](#) in the *Amazon CloudWatch Events User Guide*.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1600.

Pattern: `[ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*`

Required: Yes

#### **ScheduledActionARN**

The Amazon Resource Name (ARN) of the scheduled action.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 1600.

Pattern: `[ \u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*`

Required: Yes

#### **ScheduledActionName**

The name of the scheduled action.

Type: String

Length Constraints: Minimum length of 1. Maximum length of 256.

Pattern: `(?!((^[ ]+.*)|(.*([\u0000-\u001f]|[\u007f-\u009f]|[:/])+.*)|(.*[ ]+$)))).+`

Required: Yes

#### **ServiceNamespace**

The namespace of the AWS service that provides the resource or `custom-resource` for a resource provided by your own application or service. For more information, see [AWS Service Namespaces](#) in the *Amazon Web Services General Reference*.

Type: String

Valid Values: `ecs | elasticmapreduce | ec2 | appstream | dynamodb | rds | sagemaker | custom-resource`

Required: Yes

#### **StartTime**

The date and time that the action is scheduled to begin.

Type: Timestamp

Required: No

## **See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for Ruby V2](#)

# StepAdjustment

Represents a step adjustment for a [StepScalingPolicyConfiguration](#) (p. 72). Describes an adjustment based on the difference between the value of the aggregated CloudWatch metric and the breach threshold that you've defined for the alarm.

For the following examples, suppose that you have an alarm with a breach threshold of 50:

- To trigger the adjustment when the metric is greater than or equal to 50 and less than 60, specify a lower bound of 0 and an upper bound of 10.
- To trigger the adjustment when the metric is greater than 40 and less than or equal to 50, specify a lower bound of -10 and an upper bound of 0.

There are a few rules for the step adjustments for your step policy:

- The ranges of your step adjustments can't overlap or have a gap.
- At most one step adjustment can have a null lower bound. If one step adjustment has a negative lower bound, then there must be a step adjustment with a null lower bound.
- At most one step adjustment can have a null upper bound. If one step adjustment has a positive upper bound, then there must be a step adjustment with a null upper bound.
- The upper and lower bound can't be null in the same step adjustment.

## Contents

### **MetricIntervalLowerBound**

The lower bound for the difference between the alarm threshold and the CloudWatch metric. If the metric value is above the breach threshold, the lower bound is inclusive (the metric must be greater than or equal to the threshold plus the lower bound). Otherwise, it is exclusive (the metric must be greater than the threshold plus the lower bound). A null value indicates negative infinity.

Type: Double

Required: No

### **MetricIntervalUpperBound**

The upper bound for the difference between the alarm threshold and the CloudWatch metric. If the metric value is above the breach threshold, the upper bound is exclusive (the metric must be less than the threshold plus the upper bound). Otherwise, it is inclusive (the metric must be less than or equal to the threshold plus the upper bound). A null value indicates positive infinity.

The upper bound must be greater than the lower bound.

Type: Double

Required: No

### **ScalingAdjustment**

The amount by which to scale, based on the specified adjustment type. A positive value adds to the current scalable dimension while a negative number removes from the current scalable dimension.

Type: Integer

Required: Yes

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for Ruby V2](#)

# StepScalingPolicyConfiguration

Represents a step scaling policy configuration.

## Contents

### AdjustmentType

The adjustment type, which specifies how the `ScalingAdjustment` parameter in a [StepAdjustment](#) (p. 70) is interpreted.

Type: String

Valid Values: `ChangeInCapacity` | `PercentChangeInCapacity` | `ExactCapacity`

Required: No

### Cooldown

The amount of time, in seconds, after a scaling activity completes where previous trigger-related scaling activities can influence future scaling events.

For scale out policies, while the cooldown period is in effect, the capacity that has been added by the previous scale out event that initiated the cooldown is calculated as part of the desired capacity for the next scale out. The intention is to continuously (but not excessively) scale out. For example, an alarm triggers a step scaling policy to scale out an Amazon ECS service by 2 tasks, the scaling activity completes successfully, and a cooldown period of 5 minutes starts. During the Cooldown period, if the alarm triggers the same policy again but at a more aggressive step adjustment to scale out the service by 3 tasks, the 2 tasks that were added in the previous scale out event are considered part of that capacity and only 1 additional task is added to the desired count.

For scale in policies, the cooldown period is used to block subsequent scale in requests until it has expired. The intention is to scale in conservatively to protect your application's availability. However, if another alarm triggers a scale out policy during the cooldown period after a scale-in, Application Auto Scaling scales out your scalable target immediately.

Type: Integer

Required: No

### MetricAggregationType

The aggregation type for the CloudWatch metrics. Valid values are `Minimum`, `Maximum`, and `Average`.

Type: String

Valid Values: `Average` | `Minimum` | `Maximum`

Required: No

### MinAdjustmentMagnitude

The minimum number to adjust your scalable dimension as a result of a scaling activity. If the adjustment type is `PercentChangeInCapacity`, the scaling policy changes the scalable dimension of the scalable target by this amount.

Type: Integer

Required: No

### **StepAdjustments**

A set of adjustments that enable you to scale based on the size of the alarm breach.

Type: Array of [StepAdjustment](#) (p. 70) objects

Required: No

## **See Also**

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for Ruby V2](#)

# TargetTrackingScalingPolicyConfiguration

Represents a target tracking scaling policy configuration.

## Contents

### CustomizedMetricSpecification

A customized metric.

Type: [CustomizedMetricSpecification \(p. 54\)](#) object

Required: No

### DisableScaleIn

Indicates whether scale in by the target tracking policy is disabled. If the value is `true`, scale in is disabled and the target tracking policy won't remove capacity from the scalable resource. Otherwise, scale in is enabled and the target tracking policy can remove capacity from the scalable resource. The default value is `false`.

Type: Boolean

Required: No

### PredefinedMetricSpecification

A predefined metric.

Type: [PredefinedMetricSpecification \(p. 56\)](#) object

Required: No

### ScaleInCooldown

The amount of time, in seconds, after a scale in activity completes before another scale in activity can start.

The cooldown period is used to block subsequent scale in requests until it has expired. The intention is to scale in conservatively to protect your application's availability. However, if another alarm triggers a scale out policy during the cooldown period after a scale-in, Application Auto Scaling scales out your scalable target immediately.

Type: Integer

Required: No

### ScaleOutCooldown

The amount of time, in seconds, after a scale out activity completes before another scale out activity can start.

While the cooldown period is in effect, the capacity that has been added by the previous scale out event that initiated the cooldown is calculated as part of the desired capacity for the next scale out. The intention is to continuously (but not excessively) scale out.

Type: Integer

Required: No

### TargetValue

The target value for the metric. The range is 8.515920e-109 to 1.174271e+108 (Base 10) or 2e-360 to 2e360 (Base 2).

Type: Double

Required: Yes

## See Also

For more information about using this API in one of the language-specific AWS SDKs, see the following:

- [AWS SDK for C++](#)
- [AWS SDK for Go](#)
- [AWS SDK for Java](#)
- [AWS SDK for Ruby V2](#)



# Common Parameters

The following list contains the parameters that all actions use for signing Signature Version 4 requests with a query string. Any action-specific parameters are listed in the topic for that action. For more information about Signature Version 4, see [Signature Version 4 Signing Process](#) in the *Amazon Web Services General Reference*.

**Action**

The action to be performed.

Type: string

Required: Yes

**Version**

The API version that the request is written for, expressed in the format YYYY-MM-DD.

Type: string

Required: Yes

**X-Amz-Algorithm**

The hash algorithm that you used to create the request signature.

Condition: Specify this parameter when you include authentication information in a query string instead of in the HTTP authorization header.

Type: string

Valid Values: `AWS4-HMAC-SHA256`

Required: Conditional

**X-Amz-Credential**

The credential scope value, which is a string that includes your access key, the date, the region you are targeting, the service you are requesting, and a termination string ("aws4\_request"). The value is expressed in the following format: `access_key/YYYYMMDD/region/service/aws4_request`.

For more information, see [Task 2: Create a String to Sign for Signature Version 4](#) in the *Amazon Web Services General Reference*.

Condition: Specify this parameter when you include authentication information in a query string instead of in the HTTP authorization header.

Type: string

Required: Conditional

**X-Amz-Date**

The date that is used to create the signature. The format must be ISO 8601 basic format (YYYYMMDD'THHMMSS'Z'). For example, the following date time is a valid X-Amz-Date value: `20120325T120000Z`.

Condition: X-Amz-Date is optional for all requests; it can be used to override the date used for signing requests. If the Date header is specified in the ISO 8601 basic format, X-Amz-Date is

not required. When X-Amz-Date is used, it always overrides the value of the Date header. For more information, see [Handling Dates in Signature Version 4](#) in the *Amazon Web Services General Reference*.

Type: string

Required: Conditional

#### **X-Amz-Security-Token**

The temporary security token that was obtained through a call to AWS Security Token Service (AWS STS). For a list of services that support temporary security credentials from AWS Security Token Service, go to [AWS Services That Work with IAM](#) in the *IAM User Guide*.

Condition: If you're using temporary security credentials from the AWS Security Token Service, you must include the security token.

Type: string

Required: Conditional

#### **X-Amz-Signature**

Specifies the hex-encoded signature that was calculated from the string to sign and the derived signing key.

Condition: Specify this parameter when you include authentication information in a query string instead of in the HTTP authorization header.

Type: string

Required: Conditional

#### **X-Amz-SignedHeaders**

Specifies all the HTTP headers that were included as part of the canonical request. For more information about specifying signed headers, see [Task 1: Create a Canonical Request For Signature Version 4](#) in the *Amazon Web Services General Reference*.

Condition: Specify this parameter when you include authentication information in a query string instead of in the HTTP authorization header.

Type: string

Required: Conditional

# Common Errors

This section lists the errors common to the API actions of all AWS services. For errors specific to an API action for this service, see the topic for that API action.

**AccessDeniedException**

You do not have sufficient access to perform this action.

HTTP Status Code: 400

**IncompleteSignature**

The request signature does not conform to AWS standards.

HTTP Status Code: 400

**InternalFailure**

The request processing has failed because of an unknown error, exception or failure.

HTTP Status Code: 500

**InvalidAction**

The action or operation requested is invalid. Verify that the action is typed correctly.

HTTP Status Code: 400

**InvalidClientTokenId**

The X.509 certificate or AWS access key ID provided does not exist in our records.

HTTP Status Code: 403

**InvalidParameterCombination**

Parameters that must not be used together were used together.

HTTP Status Code: 400

**InvalidParameterValue**

An invalid or out-of-range value was supplied for the input parameter.

HTTP Status Code: 400

**InvalidQueryParameter**

The AWS query string is malformed or does not adhere to AWS standards.

HTTP Status Code: 400

**MalformedQueryString**

The query string contains a syntax error.

HTTP Status Code: 404

**MissingAction**

The request is missing an action or a required parameter.

HTTP Status Code: 400

**MissingAuthenticationToken**

The request must contain either a valid (registered) AWS access key ID or X.509 certificate.

HTTP Status Code: 403

**MissingParameter**

A required parameter for the specified action is not supplied.

HTTP Status Code: 400

**OptInRequired**

The AWS access key ID needs a subscription for the service.

HTTP Status Code: 403

**RequestExpired**

The request reached the service more than 15 minutes after the date stamp on the request or more than 15 minutes after the request expiration date (such as for pre-signed URLs), or the date stamp on the request is more than 15 minutes in the future.

HTTP Status Code: 400

**ServiceUnavailable**

The request has failed due to a temporary failure of the server.

HTTP Status Code: 503

**ThrottlingException**

The request was denied due to request throttling.

HTTP Status Code: 400

**ValidationError**

The input fails to satisfy the constraints specified by an AWS service.

HTTP Status Code: 400

# Logging Application Auto Scaling API Calls with AWS CloudTrail

Application Auto Scaling is integrated with AWS CloudTrail, a service that provides a record of actions taken by a user, role, or an AWS service using the Application Auto Scaling API. CloudTrail captures all API calls for Application Auto Scaling as events. The calls captured include calls from the AWS Management Console and code calls to the Application Auto Scaling API. If you create a trail, you can enable continuous delivery of CloudTrail events to an Amazon S3 bucket, including events for Application Auto Scaling. If you don't configure a trail, you can still view the most recent events in the CloudTrail console in **Event history**. Using the information collected by CloudTrail, you can determine the request that was made to Application Auto Scaling, the IP address from which the request was made, who made the request, when it was made, and additional details.

To learn more about CloudTrail, see the [AWS CloudTrail User Guide](#).

## Application Auto Scaling Information in CloudTrail

CloudTrail is enabled on your AWS account when you create the account. When Application Auto Scaling activity occurs, that activity is recorded in a CloudTrail event along with other AWS service events in **Event history**. You can view, search, and download recent events in your AWS account. For more information, see [Viewing Events with CloudTrail Event History](#).

For an ongoing record of events in your AWS account, including events for Application Auto Scaling, create a trail. A *trail* enables CloudTrail to deliver log files to an Amazon S3 bucket. By default, when you create a trail in the console, the trail applies to all AWS Regions. The trail logs events from all Regions in the AWS partition and delivers the log files to the Amazon S3 bucket that you specify. Additionally, you can configure other AWS services to further analyze and act upon the event data collected in CloudTrail logs. For more information, see the following:

- [Overview for Creating a Trail](#)
- [CloudTrail Supported Services and Integrations](#)
- [Configuring Amazon SNS Notifications for CloudTrail](#)
- [Receiving CloudTrail Log Files from Multiple Regions](#) and [Receiving CloudTrail Log Files from Multiple Accounts](#)

All Application Auto Scaling actions are logged by CloudTrail and are documented in the [Application Auto Scaling API Reference](#). For example, calls to the `PutScalingPolicy`, `DeleteScalingPolicy`, and `DescribeScalingPolicies` actions generate entries in the CloudTrail log files.

Every event or log entry contains information about who generated the request. The identity information helps you determine the following:

- Whether the request was made with root or AWS Identity and Access Management (IAM) user credentials.
- Whether the request was made with temporary security credentials for a role or federated user.
- Whether the request was made by another AWS service.

For more information, see the [CloudTrail `userIdentity` Element](#).

# Understanding Application Auto Scaling Log File Entries

A trail is a configuration that enables delivery of events as log files to an Amazon S3 bucket that you specify. CloudTrail log files contain one or more log entries. An event represents a single request from any source and includes information about the requested action, the date and time of the action, request parameters, and so on. CloudTrail log files aren't an ordered stack trace of the public API calls, so they don't appear in any specific order.

The following example shows a CloudTrail log entry that demonstrates the `DescribeScalableTargets` action.

```
{
  "eventVersion": "1.05",
  "userIdentity": {
    "type": "Root",
    "principalId": "123456789012",
    "arn": "arn:aws:iam::123456789012:root",
    "accountId": "123456789012",
    "accessKeyId": "AKIAIOSFODNN7EXAMPLE",
    "sessionContext": {
      "attributes": {
        "mfaAuthenticated": "false",
        "creationDate": "2018-08-21T17:05:42Z"
      }
    }
  },
  "eventTime": "2018-08-16T23:20:32Z",
  "eventSource": "autoscaling.amazonaws.com",
  "eventName": "DescribeScalableTargets",
  "awsRegion": "us-west-2",
  "sourceIPAddress": "72.21.196.68",
  "userAgent": "EC2 Spot Console",
  "requestParameters": {
    "serviceNamespace": "ec2",
    "scalableDimension": "ec2:spot-fleet-request:TargetCapacity",
    "resourceIds": [
      "spot-fleet-request/sfr-05ceaf79-3ba2-405d-e87b-612857f1357a"
    ]
  },
  "responseElements": null,
  "additionalEventData": {
    "service": "application-autoscaling"
  },
  "requestID": "0737e2ea-fb2d-11e3-bfd8-99133058e7bb",
  "eventID": "3fcfb182-98f8-4744-bd45-b38835ab61cb",
  "eventType": "AwsApiCall",
  "recipientAccountId": "123456789012"
}
```