Application Auto Scaling API Reference API Version 2016-02-06



Application Auto Scaling: API Reference

Copyright © 2016 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

	me	
	\$	
L	DeleteScalingPolicy	
	Request Syntax	
	Request Parameters	
	Response Elements	
	Errors	
	Example	4
	DeregisterScalableTarget	
	Request Syntax	6
	Request Parameters	6
	Response Elements	7
	Errors	7
	Example	7
	DescribeScalableTargets	9
	Request Syntax	
	Request Parameters	
	Response Syntax	
	Response Elements	
	Errors	
	Example	
г	DescribeScalingActivities	
_	Request Syntax	
	Request Parameters	
	Response Syntax	
	· · · · ·	
	Response Elements	
	Errors	
_	Example	
L	DescribeScalingPolicies	
	Request Syntax	
	Request Parameters	
	Response Syntax	
	Response Elements	
	Errors	
	Example	
F	PutScalingPolicy	
	Request Syntax	
	Request Parameters	
	Response Syntax	24
	Response Elements	. 24
	Errors	25
	Example	25
F	RegisterScalableTarget	27
	Request Syntax	
	Request Parameters	
	Response Elements	
	Errors	
	Example	
Т	ypes	_
	Alarm	
_	Contents	
_	Contents	
٥	Contents	
c		
٦	ScalingActivity	
	Contents	ა4

Application Auto Scaling API Reference

ScalingPolicy	36
Contents	36
StepAdjustment	38
Contents	
StepScalingPolicyConfiguration	
Contents	
Common Parameters	
Common Frrors	

Welcome

With Application Auto Scaling, you can automatically scale your AWS resources. The experience similar to that of Auto Scaling. You can use Application Auto Scaling to accomplish the following tasks:

- Define scaling policies to automatically scale your AWS resources
- Scale your resources in response to CloudWatch alarms
- · View the history of your scaling events

Application Auto Scaling can scale the following AWS resources:

- Amazon ECS services. For more information, see Service Auto Scaling in the *Amazon EC2 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.

For a list of supported regions, see AWS Regions and Endpoints: Application Auto Scaling in the AWS General Reference.

This document was last published on December 9, 2016.

Actions

The following actions are supported:

- DeleteScalingPolicy (p. 3)
- DeregisterScalableTarget (p. 6)
- DescribeScalableTargets (p. 9)
- DescribeScalingActivities (p. 13)
- DescribeScalingPolicies (p. 18)
- PutScalingPolicy (p. 23)
- RegisterScalableTarget (p. 27)

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. 23).

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. 40).

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 Resourceld (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-1791Y4E1L8Y10.

Type: String

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

Required: Yes

ScalableDimension (p. 3)

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.

Type: String

Application Auto Scaling API Reference Response Elements

```
Valid Values: ecs:service:DesiredCount | ec2:spot-fleet-
request:TargetCapacity | elasticmapreduce:instancegroup:InstanceCount
Required: Yes
```

ServiceNamespace (p. 3)

The namespace of the AWS service. For more information, see AWS Service Namespaces in the *Amazon Web Services General Reference*.

Type: String

Valid Values: ecs | elasticmapreduce | ec2

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. 42).

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 Put or Register API operation, which 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 Delete or Deregister API operation, this exception is thrown if the resource that is to be deleted or deregistered 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 AWS Gerneral 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
```

Application Auto Scaling API Reference Example

```
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"
}
```

DeregisterScalableTarget

Deregisters a scalable target.

Deregistering a scalable target deletes the scaling policies that are associated with it.

To create a scalable target or update an existing one, see RegisterScalableTarget (p. 27).

Request Syntax

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

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 40).

The request accepts the following data in JSON format.

Resourceld (p. 6)

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-1791Y4E1L8YIO.

Type: String

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

 $\label{lem:pattern: lambda} \textbf{Pattern: } [\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\ht]*$

Required: Yes

ScalableDimension (p. 6)

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.

Type: String

```
Valid Values: ecs:service:DesiredCount | ec2:spot-fleet-
request:TargetCapacity | elasticmapreduce:instancegroup:InstanceCount
Required: Yes
```

ServiceNamespace (p. 6)

The namespace of the AWS service. For more information, see AWS Service Namespaces in the *Amazon Web Services General Reference*.

```
Type: String
```

Valid Values: ecs | elasticmapreduce | ec2

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. 42).

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 Put or Register API operation, which 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 Delete or Deregister API operation, this exception is thrown if the resource that is to be deleted or deregistered 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 AWS Gerneral 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"
```

Application Auto Scaling API Reference Example

Exam	npie
}	
,	

DescribeScalableTargets

Provides descriptive 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. 27). If you are no longer using a scalable target, you can deregister it using DeregisterScalableTarget (p. 6).

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. 40).

The request accepts the following data in JSON format.

MaxResults (p. 9)

The maximum number of scalable target 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. 9)

The token for the next set of results.

Type: String

Required: No

Resourcelds (p. 9)

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-73fbd2ceaa30-494c-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-1791Y4E1L8YIO.

Type: array of Strings

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

Application Auto Scaling API Reference Response Syntax

 $\label{local_pattern: lambda_udcoo_udbff} $$\operatorname{Pattern: [\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBfF\uDFFf\r\n\t] *} $$$

Required: No

ScalableDimension (p. 9)

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.

Type: String

```
Valid Values: ecs:service:DesiredCount | ec2:spot-fleet-
request:TargetCapacity | elasticmapreduce:instancegroup:InstanceCount
Required: No
```

ServiceNamespace (p. 9)

The namespace of the AWS service. For more information, see AWS Service Namespaces in the *Amazon Web Services General Reference*.

```
Type: String
Valid Values: ecs | elasticmapreduce | ec2
Required: Yes
```

Response Syntax

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. 10)

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. 10)

The list of scalable targets that matches the request parameters.

Type: array of ScalableTarget (p. 32) objects

Errors

For information about the errors that are common to all actions, see Common Errors (p. 42).

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 AWS Gerneral 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
{
```

Application Auto Scaling API Reference Example

DescribeScalingActivities

Provides descriptive information about the scaling activities in the specified namespace from the previous six weeks.

You can filter the results using the ResourceId and ScalableDimension parameters.

Scaling activities are triggered by CloudWatch alarms that are associated with scaling policies. To view the scaling policies for a service namespace, see DescribeScalingPolicies (p. 18). To create a scaling policy or update an existing one, see PutScalingPolicy (p. 23).

Request Syntax

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

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 40).

The request accepts the following data in JSON format.

MaxResults (p. 13)

The maximum number of scalable target 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. 13)

The token for the next set of results.

Type: String

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

Required: No Resourceld (p. 13)

The identifier of the resource associated with the scaling activity. 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-1791Y4E1L8YIO.

Type: String

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

 $\label{lem:pattern: lambda} \textbf{Pattern: } [\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*$

Required: No

ScalableDimension (p. 13)

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.

Type: String

```
Valid Values: ecs:service:DesiredCount | ec2:spot-fleet-
request:TargetCapacity | elasticmapreduce:instancegroup:InstanceCount
```

Required: No

ServiceNamespace (p. 13)

The namespace of the AWS service. For more information, see AWS Service Namespaces in the *Amazon Web Services General Reference*.

Type: String

Valid Values: ecs | elasticmapreduce | ec2

Required: Yes

Response Syntax

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. 14)

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

Type: String

Application Auto Scaling API Reference Errors

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

ScalingActivities (p. 14)

A list of scaling activity objects.

Type: array of ScalingActivity (p. 34) objects

Errors

For information about the errors that are common to all actions, see Common Errors (p. 42).

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 AWS Gerneral 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,
```

Application Auto Scaling API Reference Example

```
"StatusCode": "InProgress",

"StatusMessage": "Successfully set desired count to 10. Waiting for change to be fulfilled by ecs."

}

]

}
```

DescribeScalingPolicies

Provides descriptive information about the scaling policies in the specified 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. 23). 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. 40).

The request accepts the following data in JSON format.

MaxResults (p. 18)

The maximum number of scalable target 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. 18)

The token for the next set of results.

Type: String

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

Required: No

PolicyNames (p. 18)

The names of the scaling policies to describe.

Type: array of Strings

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

Required: No

Resourceld (p. 18)

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.

Application Auto Scaling API Reference Response Syntax

- 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-1791Y4E1L8YIO.

Type: String

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

Required: No

ScalableDimension (p. 18)

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.

Type: String

```
Valid Values: ecs:service:DesiredCount | ec2:spot-fleet-
request:TargetCapacity | elasticmapreduce:instancegroup:InstanceCount
Required: No
```

ServiceNamespace (p. 18)

The namespace of the AWS service. For more information, see AWS Service Namespaces in the *Amazon Web Services General Reference*.

Type: String

Valid Values: ecs | elasticmapreduce | ec2

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": [
            {
```

Application Auto Scaling API Reference Response Elements

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. 19)

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. 19)

A list of scaling policy objects.

Type: array of ScalingPolicy (p. 36) objects

Errors

For information about the errors that are common to all actions, see Common Errors (p. 42).

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 currently only applies to DescribeScalingPolicies (p. 18). It 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 the proper permissions to call the CloudWatch DescribeAlarms API operation on behalf of your account.

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 AWS Gerneral 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-qt-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,
```

Application Auto Scaling API Reference Example

```
"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
                ]
            }
        }
    ]
```

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 without first registering a scalable target using RegisterScalableTarget (p. 27).

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. 18). 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
   ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 40).

The request accepts the following data in JSON format.

PolicyName (p. 23)

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. 23)

The policy type. If you are creating a new policy, this parameter is required. If you are updating a policy, this parameter is not required.

Type: String

Valid Values: StepScaling

Required: No

Resourceld (p. 23)

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-1791Y4E1L8YIO.

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. 23)

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.

Type: String

```
Valid Values: ecs:service:DesiredCount | ec2:spot-fleet-
request:TargetCapacity | elasticmapreduce:instancegroup:InstanceCount
```

Required: Yes

ServiceNamespace (p. 23)

The namespace of the AWS service. For more information, see AWS Service Namespaces in the *Amazon Web Services General Reference*.

Type: String

Valid Values: ecs | elasticmapreduce | ec2

Required: Yes

StepScalingPolicyConfiguration (p. 23)

The configuration for the step scaling policy. If you are creating a new policy, this parameter is required. If you are updating a policy, this parameter is not required. For more information, see StepScalingPolicyConfiguration (p. 39) and StepAdjustment (p. 38).

Type: StepScalingPolicyConfiguration (p. 39) object

Required: No

Response Syntax

```
{
    "PolicyARN": "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.

PolicyARN (p. 24)

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

Application Auto Scaling API Reference Errors

Type: String

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

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

Errors

For information about the errors that are common to all actions, see Common Errors (p. 42).

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

Your account exceeded a limit. This exception is thrown when 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 Put or Register API operation, which 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 Delete or Deregister API operation, this exception is thrown if the resource that is to be deleted or deregistered 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 AWS Gerneral 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 <code>web-app</code> in the <code>default</code> 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
```

Application Auto Scaling API Reference Example

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"
}
```

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 your scalable dimension.

After you register a scalable target, you can create and apply scaling policies using PutScalingPolicy (p. 23). You can view the scaling policies for a service namespace using DescribeScalableTargets (p. 9). If you are no longer using a scalable target, you can deregister it using DeregisterScalableTarget (p. 6).

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. 40).

The request accepts the following data in JSON format.

MaxCapacity (p. 27)

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

Type: Integer Required: No

MinCapacity (p. 27)

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

Type: Integer Required: No

Resourceld (p. 27)

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-lcee4EXAMPLE.
- EMR cluster The resource type is instancegroup and the unique identifier is the cluster ID and instance group ID. Example: instancegroup/j-2EEZNYKUA1NTV/ig-1791Y4E1L8YIO.

Type: String

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

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

Application Auto Scaling API Reference Response Elements

RoleARN (p. 27)

The ARN of an IAM role that allows Application Auto Scaling to modify the scalable target on your behalf. This parameter is required when you register a scalable target and optional when you update one.

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. 27)

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.

Type: String

Valid Values: ecs:service:DesiredCount | ec2:spot-fleetrequest:TargetCapacity | elasticmapreduce:instancegroup:InstanceCount

Required: Yes

ServiceNamespace (p. 27)

The namespace of the AWS service. For more information, see AWS Service Namespaces in the *Amazon Web Services General Reference*.

Type: String

Valid Values: ecs | elasticmapreduce | ec2

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. 42).

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

Your account exceeded a limit. This exception is thrown when 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 AWS Gerneral 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
}
```

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. 31)
- ScalableTarget (p. 32)
- ScalingActivity (p. 34)
- ScalingPolicy (p. 36)
- StepAdjustment (p. 38)
- StepScalingPolicyConfiguration (p. 39)

Alarm

Represents a CloudWatch alarm associated with a scaling policy.

Contents

AlarmARN

The Amazon Resource Name (ARN) of the alarm.

Type: String

 $\textbf{Pattern:} \ [\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\ht]* \\$

Required: Yes

AlarmName

The name of the alarm.

Type: String

 $\textbf{Pattern:} \ [\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\ht]* \\$

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

Resourceld

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-1791Y4E1L8YIO.

Type: String

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

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.

Type: String

Valid Values: ecs:service:DesiredCount | ec2:spot-fleet-

 $\verb|request:TargetCapacity| elastic mapreduce: \verb|instancegroup:InstanceCount| \\$

Application Auto Scaling API Reference Contents

ServiceNamespace

The namespace of the AWS service. For more information, see AWS Service Namespaces in the *Amazon Web Services General Reference*.

Type: String

Valid Values: ecs | elasticmapreduce | ec2

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

Resourceld

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-lcee4EXAMPLE.
- EMR cluster The resource type is instancegroup and the unique identifier is the cluster ID and instance group ID. Example: instancegroup/j-2EEZNYKUA1NTV/ig-1791Y4E1L8YIO.

Type: String

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

 $\label{lem:pattern: lambda} \textbf{Pattern: } [\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\ht]*$

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.

Application Auto Scaling API Reference Contents

• elasticmapreduce:instancegroup:InstanceCount - The instance count of an EMR Instance Group.

Type: String

Valid Values: ecs:service:DesiredCount | ec2:spot-fleet-

request:TargetCapacity | elasticmapreduce:instancegroup:InstanceCount

Required: Yes

ServiceNamespace

The namespace of the AWS service. For more information, see AWS Service Namespaces in the *Amazon Web Services General Reference*.

Type: String

Valid Values: ecs | elasticmapreduce | ec2

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

 $\textbf{Pattern:} \ [\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\ht]*$

Required: No

ScalingPolicy

Represents a scaling policy.

Contents

Alarms

The CloudWatch alarms associated with the scaling policy.

Type: array of Alarm (p. 31) objects

Required: No

CreationTime

The Unix timestamp for when the scaling policy was created.

Type: Timestamp Required: Yes

PolicyARN

The Amazon Resource Name (ARN) 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

PolicyName

The name of the scaling policy.

Type: String

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

Pattern: \p{Print}+

Required: Yes

PolicyType

The scaling policy type.

Type: String

Valid Values: StepScaling

Required: Yes

Resourceld

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-lcee4EXAMPLE.
- EMR cluster The resource type is instancegroup and the unique identifier is the cluster ID and instance group ID. Example: instancegroup/j-2EEZNYKUA1NTV/ig-1791Y4E1L8YIO.

Type: String

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

Pattern: [\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDFFF\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.

Application Auto Scaling API Reference Contents

• elasticmapreduce:instancegroup:InstanceCount - The instance count of an EMR Instance Group.

Type: String

Valid Values: ecs:service:DesiredCount | ec2:spot-fleet-

request:TargetCapacity | elasticmapreduce:instancegroup:InstanceCount

Required: Yes

ServiceNamespace

The namespace of the AWS service. For more information, see AWS Service Namespaces in the *Amazon Web Services General Reference*.

Type: String

Valid Values: ecs | elasticmapreduce | ec2

Required: Yes

StepScalingPolicyConfiguration

The configuration for the step scaling policy.

Type: StepScalingPolicyConfiguration (p. 39) object

Required: No

StepAdjustment

Represents a step adjustment for a StepScalingPolicyConfiguration (p. 39). 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

StepScalingPolicyConfiguration

Represents a step scaling policy configuration.

Contents

AdjustmentType

The adjustment type, which specifies how the ScalingAdjustment parameter in a StepAdjustment (p. 38) 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 <code>Cooldown</code> is in effect, the capacity that has been added by the previous scale out event that initiated the <code>Cooldown</code> 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 <code>Cooldown</code> period of 5 minutes starts. During the <code>Cooldown</code> 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 <code>Cooldown</code> 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 <code>Cooldown</code> 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. 38) objects

Required: No

Common Parameters

The following table lists the parameters that all actions use for signing Signature Version 4 requests. Any action-specific parameters are listed in the topic for that action. To view sample requests, see Examples of Signed Signature Version 4 Requests or Signature Version 4 Test Suite 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'T'HHMMSS'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. For a list of services that support AWS Security Token Service, go to Using Temporary Security Credentials to Access AWS in *Using Temporary Security Credentials*.

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 common errors that all actions return. Any action-specific errors are listed in the topic for the action.

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

Throttling

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