Amazon Elastic MapReduce

API Reference

API Version 2009-03-31

Amazon Elastic MapReduce: API Reference

Table of Contents

Welcome

Amazon EMR is a web service that makes it easy to process large amounts of data efficiently. Amazon EMR uses Hadoop processing combined with several AWS products to do tasks such as web indexing, data mining, log file analysis, machine learning, scientific simulation, and data warehousing.

This document was last published on May 10, 2019.

Actions

The following actions are supported:

AddInstanceFleet

Adds an instance fleet to a running cluster.

Note
The instance fleet configuration is available only in Amazon EMR versions 4.8.0 and later, excluding 5.0.x.

Request Syntax

```
{
   "ClusterId": "string",
   "InstanceFleet": {
      "InstanceFleetType": "string",
      "InstanceTypeConfigs": [
         {
            "BidPrice": "string",
            "BidPriceAsPercentageOfOnDemandPrice": number,
            "Configurations": [
               {
                  "Classification": "string",
                  "Configurations": [
                     "Configuration"
                  ],
                  "Properties": {
                     "string" : "string"
                  }
               }
            ],
            "EbsConfiguration": {
               "EbsBlockDeviceConfigs": [
                  {
                     "VolumeSpecification": {
                        "Iops": number,
                        "SizeInGB": number,
                        "VolumeType": "string"
                     },
                     "VolumesPerInstance": number
                  }
               ],
               "EbsOptimized": boolean
            },
            "InstanceType": "string",
            "WeightedCapacity": number
         }
      ],
      "LaunchSpecifications": {
         "SpotSpecification": {
            "BlockDurationMinutes": number,
            "TimeoutAction": "string",
            "TimeoutDurationMinutes": number
         }
      },
      "Name": "string",
      "TargetOnDemandCapacity": number,
      "TargetSpotCapacity": number
   }
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

ClusterId (p. 3)

The unique identifier of the cluster.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: Yes

InstanceFleet (p. 3)

Specifies the configuration of the instance fleet.

Type: InstanceFleetConfig (p. 129) object

Required: Yes

Response Syntax

```
{
   "ClusterId": "string",
   "InstanceFleetId": "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.

ClusterId (p. 4)

The unique identifier of the cluster.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

InstanceFleetId (p. 4)

The unique identifier of the instance fleet.

Type: String

Errors

For information about the errors that are common to all actions, see Common Errors (p. 202).

InternalServerException

This exception occurs when there is an internal failure in the EMR service.

HTTP Status Code: 500

InvalidRequestException

This exception occurs when there is something wrong with user input.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

AddInstanceGroups

Adds one or more instance groups to a running cluster.

Request Syntax

```
{
   "InstanceGroups": [
      {
         "AutoScalingPolicy": {
            "Constraints": {
               "MaxCapacity": number,
               "MinCapacity": number
            },
            "Rules": [
               {
                  "Action": {
                     "Market": "string",
                     "SimpleScalingPolicyConfiguration": {
                        "AdjustmentType": "string",
                        "CoolDown": number,
                        "ScalingAdjustment": number
                     }
                  },
                  "Description": "string",
                  "Name": "string",
                  "Trigger": {
                     "CloudWatchAlarmDefinition": {
                        "ComparisonOperator": "string",
                        "Dimensions": [
                           {
                              "Key": "string",
                              "Value": "string"
                           }
                        ],
                        "EvaluationPeriods": number,
                        "MetricName": "string",
                        "Namespace": "string",
                        "Period": number,
                        "Statistic": "string",
                        "Threshold": number,
                        "Unit": "string"
                     }
                  }
               }
            ]
         },
         "BidPrice": "string",
         "Configurations": [
            {
               "Classification": "string",
               "Configurations": [
                  "Configuration"
               ],
               "Properties": {
                  "string" : "string"
               }
            }
         ],
         "EbsConfiguration": {
            "EbsBlockDeviceConfigs": [
               {
                  "VolumeSpecification": {
```

```
                    "Iops": number,
                    "SizeInGB": number,
                    "VolumeType": "string"
                 },
                 "VolumesPerInstance": number
              }
           ],
           "EbsOptimized": boolean
        },
        "InstanceCount": number,
        "InstanceRole": "string",
        "InstanceType": "string",
        "Market": "string",
        "Name": "string"
     }
  ],
  "JobFlowId": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

InstanceGroups (p. 6)

Instance groups to add.

Type: Array of InstanceGroupConfig (p. 140) objects

Required: Yes

JobFlowId (p. 6)

Job flow in which to add the instance groups.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: Yes

Response Syntax

```
{
   "InstanceGroupIds": [ "string" ],
   "JobFlowId": "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.

InstanceGroupIds (p. 7)

Instance group IDs of the newly created instance groups.

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 256.

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

JobFlowId (p. 7)

The job flow ID in which the instance groups are added.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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. 202).

InternalServerError

Indicates that an error occurred while processing the request and that the request was not completed.

HTTP Status Code: 400

Example

Sample Request

```
POST / HTTP/1.1
Content-Type: application/x-amz-json-1.1
X-Amz-Target: ElasticMapReduce.AddInstanceGroups
Content-Length: 168
User-Agent: aws-sdk-ruby/1.9.2 ruby/1.9.3 i386-mingw32
Host: us-east-1.elasticmapreduce.amazonaws.com
X-Amz-Date: 20130715T223346Z
X-Amz-Content-Sha256: ac5a7193b1283898dd822a4b16ca36963879bb010d2dbe57198439973ab2a7d3
Authorization: AWS4-HMAC-SHA256 Credential=AKIAIOSFODNN7EXAMPLE/20130715/us-
east-1/elasticmapreduce/aws4_request, SignedHeaders=content-length;content-
type;host;user-agent;x-amz-content-sha256;x-amz-date;x-amz-target,
 Signature=4c5e7eb762ea45f292a5cd1a1cc56ed60009e19a9dba3d6e5e4e67e96d43af11
Accept: */*


{
    "JobFlowId": "j-3U7TSX5GZFD8Y",
    "InstanceGroups": [{
        "Name": "Task Instance Group",
        "InstanceRole": "TASK",
        "InstanceCount": 2,
        "InstanceType": "m1.small",
        "Market": "ON_DEMAND"
    }]
```

```
}
```

Sample Response

```
HTTP/1.1 200 OK
x-amzn-RequestId: 9da5a349-ed9e-11e2-90db-69a5154aeb8d
Content-Type: application/x-amz-json-1.1
Content-Length: 71
Date: Mon, 15 Jul 2013 22:33:47 GMT

{
    "InstanceGroupIds": ["ig-294A6A2KWT4WB"],
    "JobFlowId": "j-3U7TSX5GZFD8Y"
}
```

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

AddJobFlowSteps

AddJobFlowSteps adds new steps to a running cluster. A maximum of 256 steps are allowed in each job flow.

If your cluster is long-running (such as a Hive data warehouse) or complex, you may require more than 256 steps to process your data. You can bypass the 256-step limitation in various ways, including using SSH to connect to the master node and submitting queries directly to the software running on the master node, such as Hive and Hadoop. For more information on how to do this, see Add More than 256 Steps to a Cluster in the *Amazon EMR Management Guide*.

A step specifies the location of a JAR file stored either on the master node of the cluster or in Amazon S3. Each step is performed by the main function of the main class of the JAR file. The main class can be specified either in the manifest of the JAR or by using the MainFunction parameter of the step.

Amazon EMR executes each step in the order listed. For a step to be considered complete, the main function must exit with a zero exit code and all Hadoop jobs started while the step was running must have completed and run successfully.

You can only add steps to a cluster that is in one of the following states: STARTING, BOOTSTRAPPING, RUNNING, or WAITING.

Request Syntax

```
{
   "JobFlowId": "string",
   "Steps": [
      {
         "ActionOnFailure": "string",
         "HadoopJarStep": {
            "Args": [ "string" ],
            "Jar": "string",
            "MainClass": "string",
            "Properties": [
               {
                  "Key": "string",
                  "Value": "string"
               }
            ]
         },
         "Name": "string"
      }
   ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

JobFlowId (p. 10)

A string that uniquely identifies the job flow. This identifier is returned by RunJobFlow (p. 71) and can also be obtained from ListClusters (p. 39).

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: Yes

Steps (p. 10)

A list of StepConfig (p. 188) to be executed by the job flow.

Type: Array of StepConfig (p. 188) objects

Required: Yes

Response Syntax

```
{
   "StepIds": [ "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.

StepIds (p. 11)

The identifiers of the list of steps added to the job flow.

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 256.

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. 202).

InternalServerError

Indicates that an error occurred while processing the request and that the request was not completed.

HTTP Status Code: 400

Example

Sample Request

```
POST / HTTP/1.1
Content-Type: application/x-amz-json-1.1
X-Amz-Target: ElasticMapReduce.AddJobFlowSteps
```

```
Content-Length: 426
User-Agent: aws-sdk-ruby/1.9.2 ruby/1.9.3 i386-mingw32
Host: us-east-1.elasticmapreduce.amazonaws.com
X-Amz-Date: 20130716T210948Z
X-Amz-Content-Sha256: 9e5ad0a93c22224947ce98eea94f766103d91b28fa82eb60d0cb8b6f9555a6b2
Authorization: AWS4-HMAC-SHA256 Credential=AKIAIOSFODNN7EXAMPLE/20130716/us-
east-1/elasticmapreduce/aws4_request, SignedHeaders=content-length;content-
type;host;user-agent;x-amz-content-sha256;x-amz-date;x-amz-target,
 Signature=2a2393390760ae85eb74ee3a539e1d758bfdd8815a1a6d6f14d4a2fbcfdcd5b7
Accept: */*

{
    "JobFlowId": "j-3TS0OIYO4NFN",
    "Steps": [{
        "Name": "Example Jar Step",
        "ActionOnFailure": "CANCEL_AND_WAIT",
        "HadoopJarStep": {
            "Jar": "s3n:\\/\\/elasticmapreduce\\/samples\\/cloudburst\\/cloudburst.jar",
            "Args": [
                "s3n:\\/\\/elasticmapreduce\\/samples\\/cloudburst\\/input\\/s_suis.br",
                "s3n:\\/\\/elasticmapreduce\\/samples\\/cloudburst\\/input\\/100k.br",
                "s3n:\\/\\/examples-bucket\\/cloudburst\\/output",
                "36",
                "3",
                "0",
                "1",
                "240",
                "48",
                "24",
                "24",
                "128",
                "16"
            ]
        }
    }]
}
```

Sample Response

```
HTTP/1.1 200 OK
x-amzn-RequestId: 6514261f-ee5b-11e2-9345-5332e9ab2e6d
Content-Type: application/x-amz-json-1.1
Content-Length: 0
Date: Tue, 16 Jul 2013 21:05:07 GMT
```

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3

- AWS SDK for Python
- AWS SDK for Ruby V2

AddTags

Adds tags to an Amazon EMR resource. Tags make it easier to associate clusters in various ways, such as grouping clusters to track your Amazon EMR resource allocation costs. For more information, see Tag Clusters.

Request Syntax

```
{
   "ResourceId": "string",
   "Tags": [
      {
         "Key": "string",
         "Value": "string"
      }
   ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

ResourceId (p. 14)

The Amazon EMR resource identifier to which tags will be added. This value must be a cluster identifier.

Type: String

Required: Yes

Tags (p. 14)

A list of tags to associate with a cluster and propagate to EC2 instances. Tags are user-defined key/value pairs that consist of a required key string with a maximum of 128 characters, and an optional value string with a maximum of 256 characters.

Type: Array of Tag (p. 198) objects

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. 202).

InternalServerException

This exception occurs when there is an internal failure in the EMR service.

HTTP Status Code: 500

InvalidRequestException

This exception occurs when there is something wrong with user input.

HTTP Status Code: 400

Example

Sample Request

```
POST / HTTP/1.1

Content-Type: application/x-amz-json-1.1


X-Amz-Target: ElasticMapReduce.AddTags


AUTHPARAMS


{


  "ResourceId": "j-3U7TSX5GZFD8Y",


  "Tags": [{
        "Key": "stack",


      "Value": "Production"


  },


  {


      "Key": "hbase"


  }]


}
```

Sample Response

```
HTTP/1.1 200 OK

x-amzn-RequestId: 9da5a349-ed9e-11e2-90db-69a5154aeb8d


Content-Type: application/x-amz-json-1.1
```

```
Content-Length: 71

Date: Mon, 15 Jul 2013 22:33:47 GMT

{

}
```

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

CancelSteps

Cancels a pending step or steps in a running cluster. Available only in Amazon EMR versions 4.8.0 and later, excluding version 5.0.0. A maximum of 256 steps are allowed in each CancelSteps request. CancelSteps is idempotent but asynchronous; it does not guarantee a step will be canceled, even if the request is successfully submitted. You can only cancel steps that are in a `PENDING` state.

Request Syntax

```
{
   "ClusterId": "string",
   "StepIds": [ "string" ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

ClusterId (p. 17)

The `ClusterID` for which specified steps will be canceled. Use RunJobFlow (p. 71) and ListClusters (p. 39) to get ClusterIDs.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

StepIds (p. 17)

The list of `StepIDs` to cancel. Use ListSteps (p. 54) to get steps and their states for the specified cluster.

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

Response Syntax

```
{
   "CancelStepsInfoList": [
      {
         "Reason": "string",
         "Status": "string",
         "StepId": "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.

CancelStepsInfoList (p. 17)

A list of CancelStepsInfo (p. 99), which shows the status of specified cancel requests for each `StepID` specified.

Type: Array of CancelStepsInfo (p. 99) objects

Errors

For information about the errors that are common to all actions, see Common Errors (p. 202).

InternalServerError

Indicates that an error occurred while processing the request and that the request was not completed.

HTTP Status Code: 400

InvalidRequestException

This exception occurs when there is something wrong with user input.

HTTP Status Code: 400

Example

Sample Request

```
POST / HTTP/1.1
Content-Type: application/x-amz-json-1.1
X-Amz-Target: ElasticMapReduce.CancelSteps
User-Agent: aws-sdk-ruby/1.9.2 ruby/1.9.3 i386-mingw32
Host: us-east-1.elasticmapreduce.amazonaws.com
X-Amz-Date: 20160719T224800Z
X-Amz-Content-Sha256: 9e5ad0a93c22224947ce98eea94f766103d91b28fa82eb60d0cb8b6f9555a6b2
Authorization: AWS4-HMAC-SHA256 Credential=AKIAIOSFODNN7EXAMPLE/20160719/us-
east-1/elasticmapreduce/aws4_request, SignedHeaders=content-length;content-
type;host;user-agent;x-amz-content-sha256;x-amz-date;x-amz-target,
 Signature=2a2393390760ae85eb74ee3a539e1d758bfdd8815a1a6d6f14d4a2fbcfdcd5b7
Accept: */*

{
  "ClusterId": "j-2G7RS6DJZE39D",
  "StepIds":
  [
    "s-11B5G7VIKHCZQ", "s-23PUT0NR3XF6O", "s-2NUYMUZ3ADACC", "s-10O5XO5JUY9OE", "s-
CS88G2XK4N7X", "s-2M366D3KU4OTZ"
  ]
```

```
}
```

Sample Response

```
HTTP/1.1 200 OK
x-amzn-RequestId: 84931a23-4e03-11e6-b2bd-0db72d19890a
Content-Type: application/x-amz-json-1.1
Date: Tue, 19 Jul 2016 15:31:01 GMT

{
  "CancelStepsInfoList":
  [
    {"Reason": "This step cannot be cancelled.",
     "Status": "FAILED",
     "StepId": "s-11B5G7VIKHCZQ"},
    {"Reason": "Cannot cancel the step. It is already COMPLETED.",
     "Status": "FAILED",
     "StepId": "s-23PUT0NR3XF6O"},
    {"Reason": "Cannot cancel the step. It is already CANCELLED.",
     "Status": "FAILED",
     "StepId": "s-2NUYMUZ3ADACC"},
    {"Reason": "Cannot cancel the step. It is already RUNNING.",
     "Status": "FAILED",
     "StepId": "s-10O5XO5JUY9OE"},
    {"Reason": "Cannot cancel the step. It is already FAILED.",
     "Status": "FAILED",
     "StepId": "s-CS88G2XK4N7X"},
    {"Reason": "",
     "Status": "SUBMITTED",
     "StepId": "s-2M366D3KU4OTZ"}
  ]
}
```

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

CreateSecurityConfiguration

Creates a security configuration, which is stored in the service and can be specified when a cluster is created.

Request Syntax

```
{
   "Name": "string",
   "SecurityConfiguration": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

Name (p. 20)

The name of the security configuration.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: Yes

SecurityConfiguration (p. 20)

The security configuration details in JSON format. For JSON parameters and examples, see Use Security Configurations to Set Up Cluster Security in the *Amazon EMR Management Guide*.

Type: String

Required: Yes

Response Syntax

```
{
   "CreationDateTime": number,
   "Name": "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.

CreationDateTime (p. 20)

The date and time the security configuration was created.

Type: Timestamp

Name (p. 20)

The name of the security configuration.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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. 202).

InternalServerException

This exception occurs when there is an internal failure in the EMR service.

HTTP Status Code: 500

InvalidRequestException

This exception occurs when there is something wrong with user input.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

DeleteSecurityConfiguration

Deletes a security configuration.

Request Syntax

```
{
   "Name": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

Name (p. 22)

The name of the security configuration.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

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. 202).

InternalServerException

This exception occurs when there is an internal failure in the EMR service.

HTTP Status Code: 500

InvalidRequestException

This exception occurs when there is something wrong with user input.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

DescribeCluster

Provides cluster-level details including status, hardware and software configuration, VPC settings, and so on.

For information about the cluster steps, see ListSteps (p. 54).

Request Syntax

```
{
   "ClusterId": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

ClusterId (p. 24)

The identifier of the cluster to describe.

Type: String

Required: Yes

Response Syntax

```
{
   "Cluster": {
      "Applications": [
         {
            "AdditionalInfo": {
               "string" : "string"
            },
            "Args": [ "string" ],
            "Name": "string",
            "Version": "string"
         }
      ],
      "AutoScalingRole": "string",
      "AutoTerminate": boolean,
      "Configurations": [
         {
            "Classification": "string",
            "Configurations": [
               "Configuration"
            ],
            "Properties": {
               "string" : "string"
            }
         }
      ],
      "CustomAmiId": "string",
      "EbsRootVolumeSize": number,
      "Ec2InstanceAttributes": {
```

```
         "AdditionalMasterSecurityGroups": [ "string" ],
         "AdditionalSlaveSecurityGroups": [ "string" ],
         "Ec2AvailabilityZone": "string",
         "Ec2KeyName": "string",
         "Ec2SubnetId": "string",
         "EmrManagedMasterSecurityGroup": "string",
         "EmrManagedSlaveSecurityGroup": "string",
         "IamInstanceProfile": "string",
         "RequestedEc2AvailabilityZones": [ "string" ],
         "RequestedEc2SubnetIds": [ "string" ],
         "ServiceAccessSecurityGroup": "string"
      },
      "Id": "string",
      "InstanceCollectionType": "string",
      "KerberosAttributes": {
         "ADDomainJoinPassword": "string",
         "ADDomainJoinUser": "string",
         "CrossRealmTrustPrincipalPassword": "string",
         "KdcAdminPassword": "string",
         "Realm": "string"
      },
      "LogUri": "string",
      "MasterPublicDnsName": "string",
      "Name": "string",
      "NormalizedInstanceHours": number,
      "ReleaseLabel": "string",
      "RepoUpgradeOnBoot": "string",
      "RequestedAmiVersion": "string",
      "RunningAmiVersion": "string",
      "ScaleDownBehavior": "string",
      "SecurityConfiguration": "string",
      "ServiceRole": "string",
      "Status": {
         "State": "string",
         "StateChangeReason": {
            "Code": "string",
            "Message": "string"
         },
         "Timeline": {
            "CreationDateTime": number,
            "EndDateTime": number,
            "ReadyDateTime": number
         }
      },
      "Tags": [
         {
            "Key": "string",
            "Value": "string"
         }
      ],
      "TerminationProtected": boolean,
      "VisibleToAllUsers": boolean
   }
}
```

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.

Cluster (p. 24)

This output contains the details for the requested cluster.

Type: Cluster (p. 102) object

Errors

For information about the errors that are common to all actions, see Common Errors (p. 202).

InternalServerException

This exception occurs when there is an internal failure in the EMR service.

HTTP Status Code: 500

InvalidRequestException

This exception occurs when there is something wrong with user input.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

DescribeJobFlows

This API is deprecated and will eventually be removed. We recommend you use ListClusters (p. 39), DescribeCluster (p. 24), ListSteps (p. 54), ListInstanceGroups (p. 45) and ListBootstrapActions (p. 37) instead.

DescribeJobFlows returns a list of job flows that match all of the supplied parameters. The parameters can include a list of job flow IDs, job flow states, and restrictions on job flow creation date and time.

Regardless of supplied parameters, only job flows created within the last two months are returned.

If no parameters are supplied, then job flows matching either of the following criteria are returned:

- Job flows created and completed in the last two weeks
- Job flows created within the last two months that are in one of the following states: `RUNNING`, `WAITING`, `SHUTTING_DOWN`, `STARTING`

Amazon EMR can return a maximum of 512 job flow descriptions.

Request Syntax

```
{
   "CreatedAfter": number,
   "CreatedBefore": number,
   "JobFlowIds": [ "string" ],
   "JobFlowStates": [ "string" ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

CreatedAfter (p. 27)

Return only job flows created after this date and time.

Type: Timestamp

Required: No

CreatedBefore (p. 27)

Return only job flows created before this date and time.

Type: Timestamp

Required: No

JobFlowIds (p. 27)

Return only job flows whose job flow ID is contained in this list.

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

JobFlowStates (p. 27)

Return only job flows whose state is contained in this list.

Type: Array of strings

Valid Values: `STARTING | BOOTSTRAPPING | RUNNING | WAITING | SHUTTING_DOWN | TERMINATED | COMPLETED | FAILED`

Required: No

Response Syntax

```
{
   "JobFlows": [
      {
         "AmiVersion": "string",
         "AutoScalingRole": "string",
         "BootstrapActions": [
            {
               "BootstrapActionConfig": {
                  "Name": "string",
                  "ScriptBootstrapAction": {
                     "Args": [ "string" ],
                     "Path": "string"
                  }
               }
            }
         ],
         "ExecutionStatusDetail": {
            "CreationDateTime": number,
            "EndDateTime": number,
            "LastStateChangeReason": "string",
            "ReadyDateTime": number,
            "StartDateTime": number,
            "State": "string"
         },
         "Instances": {
            "Ec2KeyName": "string",
            "Ec2SubnetId": "string",
            "HadoopVersion": "string",
            "InstanceCount": number,
            "InstanceGroups": [
               {
                  "BidPrice": "string",
                  "CreationDateTime": number,
                  "EndDateTime": number,
                  "InstanceGroupId": "string",
                  "InstanceRequestCount": number,
                  "InstanceRole": "string",
                  "InstanceRunningCount": number,
                  "InstanceType": "string",
                  "LastStateChangeReason": "string",
                  "Market": "string",
                  "Name": "string",
                  "ReadyDateTime": number,
                  "StartDateTime": number,
```

```
                  "State": "string"
               }
            ],
            "KeepJobFlowAliveWhenNoSteps": boolean,
            "MasterInstanceId": "string",
            "MasterInstanceType": "string",
            "MasterPublicDnsName": "string",
            "NormalizedInstanceHours": number,
            "Placement": {
               "AvailabilityZone": "string",
               "AvailabilityZones": [ "string" ]
            },
            "SlaveInstanceType": "string",
            "TerminationProtected": boolean
         },
         "JobFlowId": "string",
         "JobFlowRole": "string",
         "LogUri": "string",
         "Name": "string",
         "ScaleDownBehavior": "string",
         "ServiceRole": "string",
         "Steps": [
            {
               "ExecutionStatusDetail": {
                  "CreationDateTime": number,
                  "EndDateTime": number,
                  "LastStateChangeReason": "string",
                  "StartDateTime": number,
                  "State": "string"
               },
               "StepConfig": {
                  "ActionOnFailure": "string",
                  "HadoopJarStep": {
                     "Args": [ "string" ],
                     "Jar": "string",
                     "MainClass": "string",
                     "Properties": [
                        {
                           "Key": "string",
                           "Value": "string"
                        }
                     ]
                  },
                  "Name": "string"
               }
            }
         ],
         "SupportedProducts": [ "string" ],
         "VisibleToAllUsers": boolean
      }
   ]
}
```

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.

JobFlows (p. 28)

A list of job flows matching the parameters supplied.

Type: Array of JobFlowDetail (p. 158) objects

Errors

For information about the errors that are common to all actions, see Common Errors (p. 202).

InternalServerError

Indicates that an error occurred while processing the request and that the request was not completed.

HTTP Status Code: 400

Example

Sample Request

```
POST / HTTP/1.1
Content-Type: application/x-amz-json-1.1
X-Amz-Target: ElasticMapReduce.DescribeJobFlows
Content-Length: 62
User-Agent: aws-sdk-ruby/1.9.2 ruby/1.9.3 i386-mingw32
Host: us-east-1.elasticmapreduce.amazonaws.com
X-Amz-Date: 20130715T220330Z
X-Amz-Content-Sha256: fce83af973f96f173512aca2845c56862b946feb1de0600326f1365b658a0e39
Authorization: AWS4-HMAC-SHA256 Credential=AKIAIOSFODNN7EXAMPLE/20130715/us-
east-1/elasticmapreduce/aws4_request, SignedHeaders=content-length;content-
type;host;user-agent;x-amz-content-sha256;x-amz-date;x-amz-target,
 Signature=29F98a6f44e05ad54fe1e8b3d1a7101ab08dc3ad348995f89c533693cee2bb3b
Accept: */*

{
    "JobFlowIds": ["j-ZKIY4CKQRX72"],
    "DescriptionType": "EXTENDED"
}
```

Sample Response

```
HTTP/1.1 200 OK
x-amzn-RequestId: 634d4142-ed9a-11e2-bbba-b56d7d016ec4
Content-Type: application/x-amz-json-1.1
Content-Length: 1624
Date: Mon, 15 Jul 2013 22:03:31 GMT

{"JobFlows": [{
    "AmiVersion": "2.3.6",
    "BootstrapActions": [],
    "ExecutionStatusDetail": {
        "CreationDateTime": 1.373923429E9,
        "EndDateTime": 1.373923995E9,
        "LastStateChangeReason": "Steps completed",
        "ReadyDateTime": 1.373923754E9,
        "StartDateTime": 1.373923754E9,
        "State": "COMPLETED"
    },
    "Instances": {
        "HadoopVersion": "1.0.3",
        "InstanceCount": 1,
        "InstanceGroups": [{
            "CreationDateTime": 1.373923429E9,
            "EndDateTime": 1.373923995E9,
```

```
            "InstanceGroupId": "ig-3SRUWV3E0NB7K",
            "InstanceRequestCount": 1,
            "InstanceRole": "MASTER",
            "InstanceRunningCount": 0,
            "InstanceType": "m1.small",
            "LastStateChangeReason": "Job flow terminated",
            "Market": "ON_DEMAND",
            "Name": "Master InstanceGroup",
            "ReadyDateTime": 1.37392375E9,
            "StartDateTime": 1.373923646E9,
            "State": "ENDED"
        }],
        "KeepJobFlowAliveWhenNoSteps": false,
        "MasterInstanceId": "i-8c4fbbef",
        "MasterInstanceType": "m1.small",
        "MasterPublicDnsName": "ec2-107-20-46-140.compute-1.amazonaws.com",
        "NormalizedInstanceHours": 1,
        "Placement": {"AvailabilityZone": "us-east-1a"},
        "TerminationProtected": false
    },
    "JobFlowId": "j-ZKIY4CKQRX72",
    "Name": "Development Job Flow",
    "Steps": [{
        "ExecutionStatusDetail": {
            "CreationDateTime": 1.373923429E9,
            "EndDateTime": 1.373923914E9,
            "StartDateTime": 1.373923754E9,
            "State": "COMPLETED"
        },
        "StepConfig": {
            "ActionOnFailure": "CANCEL_AND_WAIT",
            "HadoopJarStep": {
                "Args": [
                    "-input",
                    "s3://elasticmapreduce/samples/wordcount/input",
                    "-output",
                    "s3://examples-bucket/example-output",
                    "-mapper",
                    "s3://elasticmapreduce/samples/wordcount/wordSplitter.py",
                    "-reducer",
                    "aggregate"
                ],
                "Jar": "/home/hadoop/contrib/streaming/hadoop-streaming.jar",
                "Properties": []
            },
            "Name": "Example Streaming Step"
        }
    }],
    "SupportedProducts": []
}]}
```


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 Go - Pilot
- AWS SDK for Java

- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

DescribeSecurityConfiguration

Provides the details of a security configuration by returning the configuration JSON.

Request Syntax

```
{
   "Name": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

Name (p. 33)

The name of the security configuration.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: Yes

Response Syntax

```
{
   "CreationDateTime": number,
   "Name": "string",
   "SecurityConfiguration": "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.

CreationDateTime (p. 33)

The date and time the security configuration was created

Type: Timestamp

Name (p. 33)

The name of the security configuration.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

SecurityConfiguration (p. 33)

The security configuration details in JSON format.

Type: String

Errors

For information about the errors that are common to all actions, see Common Errors (p. 202).

InternalServerException

This exception occurs when there is an internal failure in the EMR service.

HTTP Status Code: 500

InvalidRequestException

This exception occurs when there is something wrong with user input.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

DescribeStep

Provides more detail about the cluster step.

Request Syntax

```
{
   "ClusterId": "string",
   "StepId": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

ClusterId (p. 35)

The identifier of the cluster with steps to describe.

Type: String

Required: Yes

StepId (p. 35)

The identifier of the step to describe.

Type: String

Required: Yes

Response Syntax

```
{
   "Step": {
      "ActionOnFailure": "string",
      "Config": {
         "Args": [ "string" ],
         "Jar": "string",
         "MainClass": "string",
         "Properties": {
            "string" : "string"
         }
      },
      "Id": "string",
      "Name": "string",
      "Status": {
         "FailureDetails": {
            "LogFile": "string",
            "Message": "string",
            "Reason": "string"
         },
         "State": "string",
         "StateChangeReason": {
            "Code": "string",
```

```
            "Message": "string"
         },
         "Timeline": {
            "CreationDateTime": number,
            "EndDateTime": number,
            "StartDateTime": 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.

Step (p. 35)

The step details for the requested step identifier.

Type: Step (p. 186) object

Errors

For information about the errors that are common to all actions, see Common Errors (p. 202).

InternalServerException

This exception occurs when there is an internal failure in the EMR service.

HTTP Status Code: 500

InvalidRequestException

This exception occurs when there is something wrong with user input.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

ListBootstrapActions

Provides information about the bootstrap actions associated with a cluster.

Request Syntax

```
{
   "ClusterId": "string",
   "Marker": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

ClusterId (p. 37)

The cluster identifier for the bootstrap actions to list.

Type: String

Required: Yes

Marker (p. 37)

The pagination token that indicates the next set of results to retrieve.

Type: String

Required: No

Response Syntax

```
{
   "BootstrapActions": [
      {
         "Args": [ "string" ],
         "Name": "string",
         "ScriptPath": "string"
      }
   ],
   "Marker": "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.

BootstrapActions (p. 37)

The bootstrap actions associated with the cluster.

Type: Array of Command (p. 111) objects

Marker (p. 37)

The pagination token that indicates the next set of results to retrieve.

Type: String

Errors

For information about the errors that are common to all actions, see Common Errors (p. 202).

InternalServerException

This exception occurs when there is an internal failure in the EMR service.

HTTP Status Code: 500

InvalidRequestException

This exception occurs when there is something wrong with user input.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

ListClusters

Provides the status of all clusters visible to this AWS account. Allows you to filter the list of clusters based on certain criteria; for example, filtering by cluster creation date and time or by status. This call returns a maximum of 50 clusters per call, but returns a marker to track the paging of the cluster list across multiple ListClusters calls.

Request Syntax

```
{
   "ClusterStates": [ "string" ],
   "CreatedAfter": number,
   "CreatedBefore": number,
   "Marker": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

ClusterStates (p. 39)

The cluster state filters to apply when listing clusters.

Type: Array of strings

Valid Values: `STARTING | BOOTSTRAPPING | RUNNING | WAITING | TERMINATING | TERMINATED | TERMINATED_WITH_ERRORS`

Required: No

CreatedAfter (p. 39)

The creation date and time beginning value filter for listing clusters.

Type: Timestamp

Required: No

CreatedBefore (p. 39)

The creation date and time end value filter for listing clusters.

Type: Timestamp

Required: No

Marker (p. 39)

The pagination token that indicates the next set of results to retrieve.

Type: String

Required: No

Response Syntax

```
{
   "Clusters": [
      {
         "Id": "string",
         "Name": "string",
         "NormalizedInstanceHours": number,
         "Status": {
            "State": "string",
            "StateChangeReason": {
               "Code": "string",
               "Message": "string"
            },
            "Timeline": {
               "CreationDateTime": number,
               "EndDateTime": number,
               "ReadyDateTime": number
            }
         }
      }
   ],
   "Marker": "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.

Clusters (p. 40)

The list of clusters for the account based on the given filters.

Type: Array of ClusterSummary (p. 109) objects

Marker (p. 40)

The pagination token that indicates the next set of results to retrieve.

Type: String

Errors

For information about the errors that are common to all actions, see Common Errors (p. 202).

InternalServerException

This exception occurs when there is an internal failure in the EMR service.

HTTP Status Code: 500

InvalidRequestException

This exception occurs when there is something wrong with user input.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

ListInstanceFleets

Lists all available details about the instance fleets in a cluster.

Note
The instance fleet configuration is available only in Amazon EMR versions 4.8.0 and later, excluding 5.0.x versions.

Request Syntax

```
{
   "ClusterId": "string",
   "Marker": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

ClusterId (p. 42)

The unique identifier of the cluster.

Type: String

Required: Yes

Marker (p. 42)

The pagination token that indicates the next set of results to retrieve.

Type: String

Required: No

Response Syntax

```
{
   "InstanceFleets": [
      {
         "Id": "string",
         "InstanceFleetType": "string",
         "InstanceTypeSpecifications": [
            {
               "BidPrice": "string",
               "BidPriceAsPercentageOfOnDemandPrice": number,
               "Configurations": [
                  {
                     "Classification": "string",
                     "Configurations": [
                        "Configuration"
                     ],
                     "Properties": {
                        "string" : "string"
```

```
                    }
                }
            ],
            "EbsBlockDevices": [
                {
                    "Device": "string",
                    "VolumeSpecification": {
                        "Iops": number,
                        "SizeInGB": number,
                        "VolumeType": "string"
                    }
                }
            ],
            "EbsOptimized": boolean,
            "InstanceType": "string",
            "WeightedCapacity": number
        }
    ],
    "LaunchSpecifications": {
        "SpotSpecification": {
            "BlockDurationMinutes": number,
            "TimeoutAction": "string",
            "TimeoutDurationMinutes": number
        }
    },
    "Name": "string",
    "ProvisionedOnDemandCapacity": number,
    "ProvisionedSpotCapacity": number,
    "Status": {
        "State": "string",
        "StateChangeReason": {
            "Code": "string",
            "Message": "string"
        },
        "Timeline": {
            "CreationDateTime": number,
            "EndDateTime": number,
            "ReadyDateTime": number
        }
    },
    "TargetOnDemandCapacity": number,
    "TargetSpotCapacity": number
  }
 ],
 "Marker": "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.

InstanceFleets (p. 42)

The list of instance fleets for the cluster and given filters.

Type: Array of InstanceFleet (p. 126) objects

Marker (p. 42)

The pagination token that indicates the next set of results to retrieve.

Type: String

Errors

For information about the errors that are common to all actions, see Common Errors (p. 202).

InternalServerException

This exception occurs when there is an internal failure in the EMR service.

HTTP Status Code: 500

InvalidRequestException

This exception occurs when there is something wrong with user input.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

ListInstanceGroups

Provides all available details about the instance groups in a cluster.

Request Syntax

```
{
   "ClusterId": "string",
   "Marker": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

ClusterId (p. 45)

The identifier of the cluster for which to list the instance groups.

Type: String

Required: Yes

Marker (p. 45)

The pagination token that indicates the next set of results to retrieve.

Type: String

Required: No

Response Syntax

```
{
   "InstanceGroups": [
      {
         "AutoScalingPolicy": {
            "Constraints": {
               "MaxCapacity": number,
               "MinCapacity": number
            },
            "Rules": [
               {
                  "Action": {
                     "Market": "string",
                     "SimpleScalingPolicyConfiguration": {
                        "AdjustmentType": "string",
                        "CoolDown": number,
                        "ScalingAdjustment": number
                     }
                  },
                  "Description": "string",
                  "Name": "string",
                  "Trigger": {
```

```
                  "CloudWatchAlarmDefinition": {
                     "ComparisonOperator": "string",
                     "Dimensions": [
                        {
                           "Key": "string",
                           "Value": "string"
                        }
                     ],
                     "EvaluationPeriods": number,
                     "MetricName": "string",
                     "Namespace": "string",
                     "Period": number,
                     "Statistic": "string",
                     "Threshold": number,
                     "Unit": "string"
                  }
               }
            }
         ],
         "Status": {
            "State": "string",
            "StateChangeReason": {
               "Code": "string",
               "Message": "string"
            }
         }
      },
      "BidPrice": "string",
      "Configurations": [
         {
            "Classification": "string",
            "Configurations": [
               "Configuration"
            ],
            "Properties": {
               "string" : "string"
            }
         }
      ],
      "ConfigurationsVersion": number,
      "EbsBlockDevices": [
         {
            "Device": "string",
            "VolumeSpecification": {
               "Iops": number,
               "SizeInGB": number,
               "VolumeType": "string"
            }
         }
      ],
      "EbsOptimized": boolean,
      "Id": "string",
      "InstanceGroupType": "string",
      "InstanceType": "string",
      "LastSuccessfullyAppliedConfigurations": [
         {
            "Classification": "string",
            "Configurations": [
               "Configuration"
            ],
            "Properties": {
               "string" : "string"
            }
         }
      ],
      "LastSuccessfullyAppliedConfigurationsVersion": number,
```

```
         "Market": "string",
         "Name": "string",
         "RequestedInstanceCount": number,
         "RunningInstanceCount": number,
         "ShrinkPolicy": {
            "DecommissionTimeout": number,
            "InstanceResizePolicy": {
               "InstancesToProtect": [ "string" ],
               "InstancesToTerminate": [ "string" ],
               "InstanceTerminationTimeout": number
            }
         },
         "Status": {
            "State": "string",
            "StateChangeReason": {
               "Code": "string",
               "Message": "string"
            },
            "Timeline": {
               "CreationDateTime": number,
               "EndDateTime": number,
               "ReadyDateTime": number
            }
         }
      }
   ],
   "Marker": "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.

InstanceGroups (p. 45)

The list of instance groups for the cluster and given filters.

Type: Array of InstanceGroup (p. 137) objects

Marker (p. 45)

The pagination token that indicates the next set of results to retrieve.

Type: String

Errors

For information about the errors that are common to all actions, see Common Errors (p. 202).

InternalServerException

This exception occurs when there is an internal failure in the EMR service.

HTTP Status Code: 500

InvalidRequestException

This exception occurs when there is something wrong with user input.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

ListInstances

Provides information for all active EC2 instances and EC2 instances terminated in the last 30 days, up to a maximum of 2,000. EC2 instances in any of the following states are considered active: AWAITING_FULFILLMENT, PROVISIONING, BOOTSTRAPPING, RUNNING.

Request Syntax

```
{
   "ClusterId": "string",
   "InstanceFleetId": "string",
   "InstanceFleetType": "string",
   "InstanceGroupId": "string",
   "InstanceGroupTypes": [ "string" ],
   "InstanceStates": [ "string" ],
   "Marker": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

ClusterId (p. 49)

The identifier of the cluster for which to list the instances.

Type: String

Required: Yes

InstanceFleetId (p. 49)

The unique identifier of the instance fleet.

Type: String

Required: No

InstanceFleetType (p. 49)

The node type of the instance fleet. For example MASTER, CORE, or TASK.

Type: String

Valid Values: `MASTER | CORE | TASK`

Required: No

InstanceGroupId (p. 49)

The identifier of the instance group for which to list the instances.

Type: String

Required: No

InstanceGroupTypes (p. 49)

The type of instance group for which to list the instances.

Type: Array of strings

Valid Values: `MASTER | CORE | TASK`

Required: No

InstanceStates (p. 49)

A list of instance states that will filter the instances returned with this request.

Type: Array of strings

Valid Values: `AWAITING_FULFILLMENT | PROVISIONING | BOOTSTRAPPING | RUNNING | TERMINATED`

Required: No

Marker (p. 49)

The pagination token that indicates the next set of results to retrieve.

Type: String

Required: No

Response Syntax

```
{
   "Instances": [
      {
         "EbsVolumes": [
            {
               "Device": "string",
               "VolumeId": "string"
            }
         ],
         "Ec2InstanceId": "string",
         "Id": "string",
         "InstanceFleetId": "string",
         "InstanceGroupId": "string",
         "InstanceType": "string",
         "Market": "string",
         "PrivateDnsName": "string",
         "PrivateIpAddress": "string",
         "PublicDnsName": "string",
         "PublicIpAddress": "string",
         "Status": {
            "State": "string",
            "StateChangeReason": {
               "Code": "string",
               "Message": "string"
            },
            "Timeline": {
               "CreationDateTime": number,
               "EndDateTime": number,
               "ReadyDateTime": number
            }
         }
      }
   ],
   "Marker": "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.

Instances (p. 50)

The list of instances for the cluster and given filters.

Type: Array of Instance (p. 124) objects

Marker (p. 50)

The pagination token that indicates the next set of results to retrieve.

Type: String

Errors

For information about the errors that are common to all actions, see Common Errors (p. 202).

InternalServerException

This exception occurs when there is an internal failure in the EMR service.

HTTP Status Code: 500

InvalidRequestException

This exception occurs when there is something wrong with user input.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

ListSecurityConfigurations

Lists all the security configurations visible to this account, providing their creation dates and times, and their names. This call returns a maximum of 50 clusters per call, but returns a marker to track the paging of the cluster list across multiple ListSecurityConfigurations calls.

Request Syntax

```
{
   "Marker": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

Marker (p. 52)

The pagination token that indicates the set of results to retrieve.

Type: String

Required: No

Response Syntax

```
{
   "Marker": "string",
   "SecurityConfigurations": [
      {
         "CreationDateTime": number,
         "Name": "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.

Marker (p. 52)

A pagination token that indicates the next set of results to retrieve. Include the marker in the next ListSecurityConfiguration call to retrieve the next page of results, if required.

Type: String

SecurityConfigurations (p. 52)

The creation date and time, and name, of each security configuration.

Type: Array of SecurityConfigurationSummary (p. 180) objects

Errors

For information about the errors that are common to all actions, see Common Errors (p. 202).

InternalServerException

This exception occurs when there is an internal failure in the EMR service.

HTTP Status Code: 500

InvalidRequestException

This exception occurs when there is something wrong with user input.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

ListSteps

Provides a list of steps for the cluster in reverse order unless you specify stepIds with the request.

Request Syntax

```
{
   "ClusterId": "string",
   "Marker": "string",
   "StepIds": [ "string" ],
   "StepStates": [ "string" ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

ClusterId (p. 54)

The identifier of the cluster for which to list the steps.

Type: String

Required: Yes

Marker (p. 54)

The pagination token that indicates the next set of results to retrieve.

Type: String

Required: No

StepIds (p. 54)

The filter to limit the step list based on the identifier of the steps.

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

StepStates (p. 54)

The filter to limit the step list based on certain states.

Type: Array of strings

Valid Values: `PENDING | CANCEL_PENDING | RUNNING | COMPLETED | CANCELLED | FAILED | INTERRUPTED`

Required: No

Response Syntax

```
{
   "Marker": "string",
   "Steps": [
      {
         "ActionOnFailure": "string",
         "Config": {
            "Args": [ "string" ],
            "Jar": "string",
            "MainClass": "string",
            "Properties": {
               "string" : "string"
            }
         },
         "Id": "string",
         "Name": "string",
         "Status": {
            "FailureDetails": {
               "LogFile": "string",
               "Message": "string",
               "Reason": "string"
            },
            "State": "string",
            "StateChangeReason": {
               "Code": "string",
               "Message": "string"
            },
            "Timeline": {
               "CreationDateTime": number,
               "EndDateTime": number,
               "StartDateTime": 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.

Marker (p. 55)

The pagination token that indicates the next set of results to retrieve.

Type: String

Steps (p. 55)

The filtered list of steps for the cluster.

Type: Array of StepSummary (p. 194) objects

Errors

For information about the errors that are common to all actions, see Common Errors (p. 202).

InternalServerException

This exception occurs when there is an internal failure in the EMR service.

HTTP Status Code: 500

InvalidRequestException

This exception occurs when there is something wrong with user input.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

ModifyInstanceFleet

Modifies the target On-Demand and target Spot capacities for the instance fleet with the specified InstanceFleetID within the cluster specified using ClusterID. The call either succeeds or fails atomically.

Note
The instance fleet configuration is available only in Amazon EMR versions 4.8.0 and later, excluding 5.0.x versions.

Request Syntax

```
{
   "ClusterId": "string",
   "InstanceFleet": {
      "InstanceFleetId": "string",
      "TargetOnDemandCapacity": number,
      "TargetSpotCapacity": number
   }
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

ClusterId (p. 57)

The unique identifier of the cluster.

Type: String

Required: Yes

InstanceFleet (p. 57)

The unique identifier of the instance fleet.

Type: InstanceFleetModifyConfig (p. 131) object

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. 202).

InternalServerException

This exception occurs when there is an internal failure in the EMR service.

HTTP Status Code: 500

InvalidRequestException

This exception occurs when there is something wrong with user input.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

ModifyInstanceGroups

ModifyInstanceGroups modifies the number of nodes and configuration settings of an instance group. The input parameters include the new target instance count for the group and the instance group ID. The call will either succeed or fail atomically.

Request Syntax

```
{
   "ClusterId": "string",
   "InstanceGroups": [
      {
         "Configurations": [
            {
               "Classification": "string",
               "Configurations": [
                  "Configuration"
               ],
               "Properties": {
                  "string" : "string"
               }
            }
         ],
         "EC2InstanceIdsToTerminate": [ "string" ],
         "InstanceCount": number,
         "InstanceGroupId": "string",
         "ShrinkPolicy": {
            "DecommissionTimeout": number,
            "InstanceResizePolicy": {
               "InstancesToProtect": [ "string" ],
               "InstancesToTerminate": [ "string" ],
               "InstanceTerminationTimeout": number
            }
         }
      }
   ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

ClusterId (p. 59)

The ID of the cluster to which the instance group belongs.

Type: String

Required: No

InstanceGroups (p. 59)

Instance groups to change.

Type: Array of InstanceGroupModifyConfig (p. 145) objects

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. 202).

InternalServerError

Indicates that an error occurred while processing the request and that the request was not completed.

HTTP Status Code: 400

Example

Sample Request

```
POST / HTTP/1.1
Content-Type: application/x-amz-json-1.1
X-Amz-Target: ElasticMapReduce.ModifyInstanceGroups
Content-Length: 77
User-Agent: aws-sdk-ruby/1.9.2 ruby/1.9.3 i386-mingw32
Host: us-east-1.elasticmapreduce.amazonaws.com
X-Amz-Date: 20130716T205843Z
X-Amz-Content-Sha256: bb1af3d0c6c6a1a09f21ccd7f04a0e2e6c9ce5b5810b0f6777560fe4f81bda8c
Authorization: AWS4-HMAC-SHA256 Credential=AKIAIOSFODNN7EXAMPLE/20130716/us-
east-1/elasticmapreduce/aws4_request, SignedHeaders=content-length;content-
type;host;user-agent;x-amz-content-sha256;x-amz-date;x-amz-target,
 Signature=17bbbb4448a1f47a14d5657445e9de5cadf16bed58b850585f80865882133b33
Accept: */*

{"InstanceGroups": [{
    "InstanceGroupId": "ig-1S8NWT31S2OVG",
    "InstanceCount": 5
}]}
```

Sample Response

```
HTTP/1.1 200 OK
x-amzn-RequestId: 80a74808-ee5a-11e2-90db-69a5154aeb8d
Content-Type: application/x-amz-json-1.1
Content-Length: 0
Date: Tue, 16 Jul 2013 20:58:44 GMT
```

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

PutAutoScalingPolicy

Creates or updates an automatic scaling policy for a core instance group or task instance group in an Amazon EMR cluster. The automatic scaling policy defines how an instance group dynamically adds and terminates EC2 instances in response to the value of a CloudWatch metric.

Request Syntax

```
{
   "AutoScalingPolicy": {
      "Constraints": {
         "MaxCapacity": number,
         "MinCapacity": number
      },
      "Rules": [
         {
            "Action": {
               "Market": "string",
               "SimpleScalingPolicyConfiguration": {
                  "AdjustmentType": "string",
                  "CoolDown": number,
                  "ScalingAdjustment": number
               }
            },
            "Description": "string",
            "Name": "string",
            "Trigger": {
               "CloudWatchAlarmDefinition": {
                  "ComparisonOperator": "string",
                  "Dimensions": [
                     {
                        "Key": "string",
                        "Value": "string"
                     }
                  ],
                  "EvaluationPeriods": number,
                  "MetricName": "string",
                  "Namespace": "string",
                  "Period": number,
                  "Statistic": "string",
                  "Threshold": number,
                  "Unit": "string"
               }
            }
         }
      ]
   },
   "ClusterId": "string",
   "InstanceGroupId": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

AutoScalingPolicy (p. 62)

Specifies the definition of the automatic scaling policy.

Type: AutoScalingPolicy (p. 93) object

Required: Yes

ClusterId (p. 62)

Specifies the ID of a cluster. The instance group to which the automatic scaling policy is applied is within this cluster.

Type: String

Required: Yes

InstanceGroupId (p. 62)

Specifies the ID of the instance group to which the automatic scaling policy is applied.

Type: String

Required: Yes

Response Syntax

```
{
   "AutoScalingPolicy": {
      "Constraints": {
         "MaxCapacity": number,
         "MinCapacity": number
      },
      "Rules": [
         {
            "Action": {
               "Market": "string",
               "SimpleScalingPolicyConfiguration": {
                  "AdjustmentType": "string",
                  "CoolDown": number,
                  "ScalingAdjustment": number
               }
            },
            "Description": "string",
            "Name": "string",
            "Trigger": {
               "CloudWatchAlarmDefinition": {
                  "ComparisonOperator": "string",
                  "Dimensions": [
                     {
                        "Key": "string",
                        "Value": "string"
                     }
                  ],
                  "EvaluationPeriods": number,
                  "MetricName": "string",
                  "Namespace": "string",
                  "Period": number,
                  "Statistic": "string",
                  "Threshold": number,
                  "Unit": "string"
               }
            }
         }
      ],
      "Status": {
         "State": "string",
```

```
         "StateChangeReason": {
            "Code": "string",
            "Message": "string"
         }
      }
   },
   "ClusterId": "string",
   "InstanceGroupId": "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.

AutoScalingPolicy (p. 63)

The automatic scaling policy definition.

Type: AutoScalingPolicyDescription (p. 94) object

ClusterId (p. 63)

Specifies the ID of a cluster. The instance group to which the automatic scaling policy is applied is within this cluster.

Type: String

InstanceGroupId (p. 63)

Specifies the ID of the instance group to which the scaling policy is applied.

Type: String

Errors

For information about the errors that are common to all actions, see Common Errors (p. 202).

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

RemoveAutoScalingPolicy

Removes an automatic scaling policy from a specified instance group within an EMR cluster.

Request Syntax

```
{
   "ClusterId": "string",
   "InstanceGroupId": "string"
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

ClusterId (p. 65)

Specifies the ID of a cluster. The instance group to which the automatic scaling policy is applied is within this cluster.

Type: String

Required: Yes

InstanceGroupId (p. 65)

Specifies the ID of the instance group to which the scaling policy is applied.

Type: String

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. 202).

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 Go - Pilot
- AWS SDK for Java

- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

RemoveTags

Removes tags from an Amazon EMR resource. Tags make it easier to associate clusters in various ways, such as grouping clusters to track your Amazon EMR resource allocation costs. For more information, see Tag Clusters.

The following example removes the stack tag with value Prod from a cluster:

Request Syntax

```
{
   "ResourceId": "string",
   "TagKeys": [ "string" ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

ResourceId (p. 67)

The Amazon EMR resource identifier from which tags will be removed. This value must be a cluster identifier.

Type: String

Required: Yes

TagKeys (p. 67)

A list of tag keys to remove from a resource.

Type: Array of strings

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. 202).

InternalServerException

This exception occurs when there is an internal failure in the EMR service.

HTTP Status Code: 500

InvalidRequestException

This exception occurs when there is something wrong with user input.

HTTP Status Code: 400

Examples

Example

Sample Request

```
POST / HTTP/1.1

Content-Type: application/x-amz-json-1.1


X-Amz-Target: ElasticMapReduce.RemoveTags


AUTHPARAMS


{


  "ResourceId": "j-3U7TSX5GZFD8Y",


  "Tags": [{


      "Key": "stack",


      "Value": "Prod"


  }]


}
```

Sample Response

```
HTTP/1.1 200 OK

x-amzn-RequestId: 9da5a349-ed9e-11e2-90db-69a5154aeb8d


Content-Type: application/x-amz-json-1.1


Content-Length: 71


Date: Mon, 15 Jul 2013 22:33:47 GMT


{


}
```

Example

The following example removes the stack and hbase tags from a cluster:

Sample Request

```
POST / HTTP/1.1

Content-Type: application/x-amz-json-1.1


X-Amz-Target: ElasticMapReduce.RemoveTags


AUTHPARAMS


{


  "ResourceId": "j-3U7TSX5GZFD8Y",


  "Tags": [{


      "Key": "stack"


  },


  {


      "Key": "hbase"


  }]


}
```

Sample Response

```
HTTP/1.1 200 OK

x-amzn-RequestId: 9da5a349-ed9e-11e2-90db-69a5154aeb8d


Content-Type: application/x-amz-json-1.1


Content-Length: 71
```

```
Date: Mon, 15 Jul 2013 22:33:47 GMT


{


}
```

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

RunJobFlow

RunJobFlow creates and starts running a new cluster (job flow). The cluster runs the steps specified. After the steps complete, the cluster stops and the HDFS partition is lost. To prevent loss of data, configure the last step of the job flow to store results in Amazon S3. If the JobFlowInstancesConfig (p. 163) `KeepJobFlowAliveWhenNoSteps` parameter is set to `TRUE`, the cluster transitions to the WAITING state rather than shutting down after the steps have completed.

For additional protection, you can set the JobFlowInstancesConfig (p. 163) `TerminationProtected` parameter to `TRUE` to lock the cluster and prevent it from being terminated by API call, user intervention, or in the event of a job flow error.

A maximum of 256 steps are allowed in each job flow.

If your cluster is long-running (such as a Hive data warehouse) or complex, you may require more than 256 steps to process your data. You can bypass the 256-step limitation in various ways, including using the SSH shell to connect to the master node and submitting queries directly to the software running on the master node, such as Hive and Hadoop. For more information on how to do this, see Add More than 256 Steps to a Cluster in the *Amazon EMR Management Guide*.

For long running clusters, we recommend that you periodically store your results.

Note
The instance fleets configuration is available only in Amazon EMR versions 4.8.0 and later, excluding 5.0.x versions. The RunJobFlow request can contain InstanceFleets parameters or InstanceGroups parameters, but not both.

Request Syntax

```
{
   "AdditionalInfo": "string",
   "AmiVersion": "string",
   "Applications": [
      {
         "AdditionalInfo": {
            "string" : "string"
         },
         "Args": [ "string" ],
         "Name": "string",
         "Version": "string"
      }
   ],
   "AutoScalingRole": "string",
   "BootstrapActions": [
      {
         "Name": "string",
         "ScriptBootstrapAction": {
            "Args": [ "string" ],
            "Path": "string"
         }
      }
   ],
   "Configurations": [
      {
         "Classification": "string",
         "Configurations": [
            "Configuration"
         ],
         "Properties": {
            "string" : "string"
```

```
        }
    }
],
"CustomAmiId": "string",
"EbsRootVolumeSize": number,
"Instances": {
   "AdditionalMasterSecurityGroups": [ "string" ],
   "AdditionalSlaveSecurityGroups": [ "string" ],
   "Ec2KeyName": "string",
   "Ec2SubnetId": "string",
   "Ec2SubnetIds": [ "string" ],
   "EmrManagedMasterSecurityGroup": "string",
   "EmrManagedSlaveSecurityGroup": "string",
   "HadoopVersion": "string",
   "InstanceCount": number,
   "InstanceFleets": [
      {
         "InstanceFleetType": "string",
         "InstanceTypeConfigs": [
            {
               "BidPrice": "string",
               "BidPriceAsPercentageOfOnDemandPrice": number,
               "Configurations": [
                  {
                     "Classification": "string",
                     "Configurations": [
                        "Configuration"
                     ],
                     "Properties": {
                        "string" : "string"
                     }
                  }
               ],
               "EbsConfiguration": {
                  "EbsBlockDeviceConfigs": [
                     {
                        "VolumeSpecification": {
                           "Iops": number,
                           "SizeInGB": number,
                           "VolumeType": "string"
                        },
                        "VolumesPerInstance": number
                     }
                  ],
                  "EbsOptimized": boolean
               },
               "InstanceType": "string",
               "WeightedCapacity": number
            }
         ],
         "LaunchSpecifications": {
            "SpotSpecification": {
               "BlockDurationMinutes": number,
               "TimeoutAction": "string",
               "TimeoutDurationMinutes": number
            }
         },
         "Name": "string",
         "TargetOnDemandCapacity": number,
         "TargetSpotCapacity": number
      }
   ],
   "InstanceGroups": [
      {
         "AutoScalingPolicy": {
            "Constraints": {
```

```
            "MaxCapacity": number,
            "MinCapacity": number
         },
         "Rules": [
            {
               "Action": {
                  "Market": "string",
                  "SimpleScalingPolicyConfiguration": {
                     "AdjustmentType": "string",
                     "CoolDown": number,
                     "ScalingAdjustment": number
                  }
               },
               "Description": "string",
               "Name": "string",
               "Trigger": {
                  "CloudWatchAlarmDefinition": {
                     "ComparisonOperator": "string",
                     "Dimensions": [
                        {
                           "Key": "string",
                           "Value": "string"
                        }
                     ],
                     "EvaluationPeriods": number,
                     "MetricName": "string",
                     "Namespace": "string",
                     "Period": number,
                     "Statistic": "string",
                     "Threshold": number,
                     "Unit": "string"
                  }
               }
            }
         ]
      },
      "BidPrice": "string",
      "Configurations": [
         {
            "Classification": "string",
            "Configurations": [
               "Configuration"
            ],
            "Properties": {
               "string" : "string"
            }
         }
      ],
      "EbsConfiguration": {
         "EbsBlockDeviceConfigs": [
            {
               "VolumeSpecification": {
                  "Iops": number,
                  "SizeInGB": number,
                  "VolumeType": "string"
               },
               "VolumesPerInstance": number
            }
         ],
         "EbsOptimized": boolean
      },
      "InstanceCount": number,
      "InstanceRole": "string",
      "InstanceType": "string",
      "Market": "string",
      "Name": "string"
```

```
            }
         ],
         "KeepJobFlowAliveWhenNoSteps": boolean,
         "MasterInstanceType": "string",
         "Placement": {
            "AvailabilityZone": "string",
            "AvailabilityZones": [ "string" ]
         },
         "ServiceAccessSecurityGroup": "string",
         "SlaveInstanceType": "string",
         "TerminationProtected": boolean
      },
      "JobFlowRole": "string",
      "KerberosAttributes": {
         "ADDomainJoinPassword": "string",
         "ADDomainJoinUser": "string",
         "CrossRealmTrustPrincipalPassword": "string",
         "KdcAdminPassword": "string",
         "Realm": "string"
      },
      "LogUri": "string",
      "Name": "string",
      "NewSupportedProducts": [
         {
            "Args": [ "string" ],
            "Name": "string"
         }
      ],
      "ReleaseLabel": "string",
      "RepoUpgradeOnBoot": "string",
      "ScaleDownBehavior": "string",
      "SecurityConfiguration": "string",
      "ServiceRole": "string",
      "Steps": [
         {
            "ActionOnFailure": "string",
            "HadoopJarStep": {
               "Args": [ "string" ],
               "Jar": "string",
               "MainClass": "string",
               "Properties": [
                  {
                     "Key": "string",
                     "Value": "string"
                  }
               ]
            },
            "Name": "string"
         }
      ],
      "SupportedProducts": [ "string" ],
      "Tags": [
         {
            "Key": "string",
            "Value": "string"
         }
      ],
      "VisibleToAllUsers": boolean
   }
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

AdditionalInfo (p. 71)

A JSON string for selecting additional features.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

AmiVersion (p. 71)

Applies only to Amazon EMR AMI versions 3.x and 2.x. For Amazon EMR releases 4.0 and later, `ReleaseLabel` is used. To specify a custom AMI, use `CustomAmiID`.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

Applications (p. 71)

Applies to Amazon EMR releases 4.0 and later. A case-insensitive list of applications for Amazon EMR to install and configure when launching the cluster. For a list of applications available for each Amazon EMR release version, see the Amazon EMR Release Guide.

Type: Array of Application (p. 92) objects

Required: No

AutoScalingRole (p. 71)

An IAM role for automatic scaling policies. The default role is `EMR_AutoScaling_DefaultRole`. The IAM role provides permissions that the automatic scaling feature requires to launch and terminate EC2 instances in an instance group.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

BootstrapActions (p. 71)

A list of bootstrap actions to run before Hadoop starts on the cluster nodes.

Type: Array of BootstrapActionConfig (p. 97) objects

Required: No

Configurations (p. 71)

For Amazon EMR releases 4.0 and later. The list of configurations supplied for the EMR cluster you are creating.

Type: Array of Configuration (p. 112) objects

Required: No

CustomAmiId (p. 71)

Available only in Amazon EMR version 5.7.0 and later. The ID of a custom Amazon EBS-backed Linux AMI. If specified, Amazon EMR uses this AMI when it launches cluster EC2 instances. For more information about custom AMIs in Amazon EMR, see Using a Custom AMI in the *Amazon EMR Management Guide*. If omitted, the cluster uses the base Linux AMI for the `ReleaseLabel` specified. For Amazon EMR versions 2.x and 3.x, use `AmiVersion` instead.

For information about creating a custom AMI, see Creating an Amazon EBS-Backed Linux AMI in the *Amazon Elastic Compute Cloud User Guide for Linux Instances*. For information about finding an AMI ID, see Finding a Linux AMI.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

EbsRootVolumeSize (p. 71)

The size, in GiB, of the EBS root device volume of the Linux AMI that is used for each EC2 instance. Available in Amazon EMR version 4.x and later.

Type: Integer

Required: No

Instances (p. 71)

A specification of the number and type of Amazon EC2 instances.

Type: JobFlowInstancesConfig (p. 163) object

Required: Yes

JobFlowRole (p. 71)

Also called instance profile and EC2 role. An IAM role for an EMR cluster. The EC2 instances of the cluster assume this role. The default role is `EMR_EC2_DefaultRole`. In order to use the default role, you must have already created it using the CLI or console.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

KerberosAttributes (p. 71)

Attributes for Kerberos configuration when Kerberos authentication is enabled using a security configuration. For more information see Use Kerberos Authentication in the *EMR Management Guide*.

Type: KerberosAttributes (p. 170) object

Required: No

LogUri (p. 71)

The location in Amazon S3 to write the log files of the job flow. If a value is not provided, logs are not created.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

Name (p. 71)

The name of the job flow.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: Yes

NewSupportedProducts (p. 71)

> **Note**
> For Amazon EMR releases 3.x and 2.x. For Amazon EMR releases 4.x and later, use Applications.

A list of strings that indicates third-party software to use with the job flow that accepts a user argument list. EMR accepts and forwards the argument list to the corresponding installation script as bootstrap action arguments. For more information, see "Launch a Job Flow on the MapR Distribution for Hadoop" in the Amazon EMR Developer Guide. Supported values are:

- "mapr-m3" - launch the cluster using MapR M3 Edition.
- "mapr-m5" - launch the cluster using MapR M5 Edition.
- "mapr" with the user arguments specifying "--edition,m3" or "--edition,m5" - launch the job flow using MapR M3 or M5 Edition respectively.
- "mapr-m7" - launch the cluster using MapR M7 Edition.
- "hunk" - launch the cluster with the Hunk Big Data Analtics Platform.
- "hue"- launch the cluster with Hue installed.
- "spark" - launch the cluster with Apache Spark installed.
- "ganglia" - launch the cluster with the Ganglia Monitoring System installed.

Type: Array of SupportedProductConfig (p. 197) objects

Required: No

ReleaseLabel (p. 71)

The Amazon EMR release label, which determines the version of open-source application packages installed on the cluster. Release labels are in the form `emr-x.x.x`, where x.x.x is an Amazon EMR release version, for example, `emr-5.14.0`. For more information about Amazon EMR release versions and included application versions and features, see https://docs.aws.amazon.com/emr/latest/ReleaseGuide/. The release label applies only to Amazon EMR releases versions 4.x and later. Earlier versions use `AmiVersion`.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

RepoUpgradeOnBoot (p. 71)

Applies only when `CustomAmiID` is used. Specifies which updates from the Amazon Linux AMI package repositories to apply automatically when the instance boots using the AMI. If omitted, the default is `SECURITY`, which indicates that only security updates are applied. If `NONE` is specified, no updates are applied, and all updates must be applied manually.

Type: String

Valid Values: `SECURITY | NONE`

Required: No

ScaleDownBehavior (p. 71)

Specifies the way that individual Amazon EC2 instances terminate when an automatic scale-in activity occurs or an instance group is resized. `TERMINATE_AT_INSTANCE_HOUR` indicates that Amazon EMR terminates nodes at the instance-hour boundary, regardless of when the request to terminate the instance was submitted. This option is only available with Amazon EMR 5.1.0 and later and is the default for clusters created using that version. `TERMINATE_AT_TASK_COMPLETION` indicates that Amazon EMR blacklists and drains tasks from nodes before terminating the Amazon EC2 instances, regardless of the instance-hour boundary. With either behavior, Amazon EMR removes the least active nodes first and blocks instance termination if it could lead to HDFS corruption. `TERMINATE_AT_TASK_COMPLETION` available only in Amazon EMR version 4.1.0 and later, and is the default for versions of Amazon EMR earlier than 5.1.0.

Type: String

Valid Values: `TERMINATE_AT_INSTANCE_HOUR | TERMINATE_AT_TASK_COMPLETION`

Required: No

SecurityConfiguration (p. 71)

The name of a security configuration to apply to the cluster.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

ServiceRole (p. 71)

The IAM role that will be assumed by the Amazon EMR service to access AWS resources on your behalf.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

Steps (p. 71)

A list of steps to run.

Type: Array of StepConfig (p. 188) objects

Required: No

SupportedProducts (p. 71)

> **Note**
> For Amazon EMR releases 3.x and 2.x. For Amazon EMR releases 4.x and later, use Applications.

A list of strings that indicates third-party software to use. For more information, see the Amazon EMR Developer Guide. Currently supported values are:

- "mapr-m3" - launch the job flow using MapR M3 Edition.
- "mapr-m5" - launch the job flow using MapR M5 Edition.

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

Tags (p. 71)

A list of tags to associate with a cluster and propagate to Amazon EC2 instances.

Type: Array of Tag (p. 198) objects

Required: No

VisibleToAllUsers (p. 71)

A value of `true` indicates that all IAM users in the AWS account can perform cluster actions if they have the proper IAM policy permissions. This is the default. A value of `false` indicates that only the IAM user who created the cluster can perform actions.

Type: Boolean

Required: No

Response Syntax

```
{
   "JobFlowId": "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.

JobFlowId (p. 79)

An unique identifier for the job flow.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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. 202).

InternalServerError

Indicates that an error occurred while processing the request and that the request was not completed.

HTTP Status Code: 400

Example

Example 1

Sample Request

```
POST / HTTP/1.1
Content-Type: application/x-amz-json-1.1
X-Amz-Target: ElasticMapReduce.RunJobFlow
Content-Length: 734
User-Agent: aws-sdk-ruby/1.9.2 ruby/1.9.3 i386-mingw32
Host: us-east-1.elasticmapreduce.amazonaws.com
X-Amz-Date: 20130715T210803Z
X-Amz-Content-Sha256: 8676d21986e4628a89fb1232a1344063778d4ffc23d10be02b437e0d53a24db3
Authorization: AWS4-HMAC-SHA256 Credential=AKIAIOSFODNN7EXAMPLE/20130715/us-
east-1/elasticmapreduce/aws4_request, SignedHeaders=content-length;content-
type;host;user-agent;x-amz-content-sha256;x-amz-date;x-amz-target,
 Signature=71f79725c4dbe77c0e842718485f0b37fe6df69e1153c80f7748ebd9617ca2f3
Accept: */*


{
    "Name": "Development Job Flow",
    "Instances": {
        "KeepJobFlowAliveWhenNoSteps": "false",
        "TerminationProtected": "false",
        "InstanceGroups": [{
            "Name": "Master Instance Group",
            "InstanceRole": "MASTER",
            "InstanceCount": 1,
            "InstanceType": "m1.small",
            "Market": "ON_DEMAND"
        }]
    },
    "Steps": [{
        "Name": "Example Streaming Step",
        "ActionOnFailure": "CANCEL_AND_WAIT",
        "HadoopJarStep": {
            "Jar": "/home/hadoop/contrib/streaming/hadoop-streaming.jar",
            "Args": [
                "-input",
                "s3://elasticmapreduce/samples/wordcount/input",
                "-output",
                "s3://examples-bucket/example-output",
                "-mapper",
                "s3://elasticmapreduce/samples/wordcount/wordSplitter.py",
                "-reducer",
                "aggregate"
            ]
```

```
        }
    }],
    "BootstrapActions": [],
    "NewSupportedProduct": [],
    "AmiVersion": "3.8.0"
}
```

Sample Response

```
HTTP/1.1 200 OK
x-amzn-RequestId: a4406d6b-ed92-11e2-9787-192218ecb460
Content-Type: application/x-amz-json-1.1
Content-Length: 31
Date: Mon, 15 Jul 2013 21:08:05 GMT

{"JobFlowId": "j-ZKIY4CKQRX72"}
```

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

SetTerminationProtection

SetTerminationProtection locks a cluster (job flow) so the EC2 instances in the cluster cannot be terminated by user intervention, an API call, or in the event of a job-flow error. The cluster still terminates upon successful completion of the job flow. Calling `SetTerminationProtection` on a cluster is similar to calling the Amazon EC2 `DisableAPITermination` API on all EC2 instances in a cluster.

`SetTerminationProtection` is used to prevent accidental termination of a cluster and to ensure that in the event of an error, the instances persist so that you can recover any data stored in their ephemeral instance storage.

To terminate a cluster that has been locked by setting `SetTerminationProtection` to `true`, you must first unlock the job flow by a subsequent call to `SetTerminationProtection` in which you set the value to `false`.

For more information, seeManaging Cluster Termination in the *Amazon EMR Management Guide*.

Request Syntax

```
{
   "JobFlowIds": [ "string" ],
   "TerminationProtected": boolean
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

JobFlowIds (p. 82)

A list of strings that uniquely identify the clusters to protect. This identifier is returned by RunJobFlow (p. 71) and can also be obtained from DescribeJobFlows (p. 27) .

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: Yes

TerminationProtected (p. 82)

A Boolean that indicates whether to protect the cluster and prevent the Amazon EC2 instances in the cluster from shutting down due to API calls, user intervention, or job-flow error.

Type: Boolean

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. 202).

InternalServerError

Indicates that an error occurred while processing the request and that the request was not completed.

HTTP Status Code: 400

Example

Sample Request

```
POST / HTTP/1.1
Content-Type: application/x-amz-json-1.1
X-Amz-Target: ElasticMapReduce.SetTerminationProtection
Content-Length: 61
User-Agent: aws-sdk-ruby/1.9.2 ruby/1.9.3 i386-mingw32
Host: us-east-1.elasticmapreduce.amazonaws.com
X-Amz-Date: 20130716T211420Z
X-Amz-Content-Sha256: c362fadae0fce377aa63f04388aeb90c53cedb17a8bfbb8cffcb10c2378137f9
Authorization: AWS4-HMAC-SHA256 Credential=AKIAIOSFODNN7EXAMPLE/20130716/us-
east-1/elasticmapreduce/aws4_request, SignedHeaders=content-length;content-
type;host;user-agent;x-amz-content-sha256;x-amz-date;x-amz-target,
 Signature=764b6aa1a38733cadff35a2e884887e9f1208a422266bc83ac77e8d0b80bd4cf
Accept: */*

{
    "JobFlowIds": ["j-3TS0OIYO4NFN"],
    "TerminationProtected": true
}
```

Sample Response

```
HTTP/1.1 200 OK
x-amzn-RequestId: af23b1db-ee5c-11e2-9787-192218ecb460
Content-Type: application/x-amz-json-1.1
Content-Length: 0
Date: Tue, 16 Jul 2013 21:14:21 GMT
```

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 Go - Pilot
- AWS SDK for Java

- AWS SDK for JavaScript
- AWS SDK for PHP V3
- AWS SDK for Python
- AWS SDK for Ruby V2

SetVisibleToAllUsers

Sets the Cluster:VisibleToAllUsers (p. 105) value, which determines whether the cluster is visible to all IAM users of the AWS account associated with the cluster. Only the IAM user who created the cluster or the AWS account root user can call this action. The default value, `true`, indicates that all IAM users in the AWS account can perform cluster actions if they have the proper IAM policy permissions. If set to `false`, only the IAM user that created the cluster can perform actions. This action works on running clusters. You can override the default `true` setting when you create a cluster by using the `VisibleToAllUsers` parameter with `RunJobFlow`.

Request Syntax

```
{
   "JobFlowIds": [ "string" ],
   "VisibleToAllUsers": boolean
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

JobFlowIds (p. 85)

The unique identifier of the job flow (cluster).

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: Yes

VisibleToAllUsers (p. 85)

A value of `true` indicates that all IAM users in the AWS account can perform cluster actions if they have the proper IAM policy permissions. This is the default. A value of `false` indicates that only the IAM user who created the cluster can perform actions.

Type: Boolean

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. 202).

InternalServerError

Indicates that an error occurred while processing the request and that the request was not completed.

HTTP Status Code: 400

Example

Sample Request

```
POST / HTTP/1.1
Content-Type: application/x-amz-json-1.1
X-Amz-Target: ElasticMapReduce.SetVisibleToAllUsers
Content-Length: 58
User-Agent: aws-sdk-ruby/1.9.2 ruby/1.9.3 i386-mingw32
Host: us-east-1.elasticmapreduce.amazonaws.com
X-Amz-Date: 20130715T221616Z
X-Amz-Content-Sha256: 2ff32d11eab2383d764ffcb97571454e798689ecd09a7b1bb2327e22b0b930d4
Authorization: AWS4-HMAC-SHA256 Credential=AKIAIOSFODNN7EXAMPLE/20130715/us-
east-1/elasticmapreduce/aws4_request, SignedHeaders=content-length;content-
type;host;user-agent;x-amz-content-sha256;x-amz-date;x-amz-target,
 Signature=e1a00b37787d9ccc43c9de32f1f0a73813b0bd6643d4db7762b62a7092d51997
Accept: */*

{
    "JobFlowIds": ["j-ZKIY4CKQRX72"],
    "VisibleToAllUsers": true
}
```

Sample Response

```
HTTP/1.1 200 OK
x-amzn-RequestId: 2be9cde9-ed9c-11e2-82b6-2351cde3f33f
Content-Type: application/x-amz-json-1.1
Content-Length: 0
Date: Mon, 15 Jul 2013 22:16:18 GMT
```

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for JavaScript
- AWS SDK for PHP V3

- AWS SDK for Python
- AWS SDK for Ruby V2

TerminateJobFlows

TerminateJobFlows shuts a list of clusters (job flows) down. When a job flow is shut down, any step not yet completed is canceled and the EC2 instances on which the cluster is running are stopped. Any log files not already saved are uploaded to Amazon S3 if a LogUri was specified when the cluster was created.

The maximum number of clusters allowed is 10. The call to `TerminateJobFlows` is asynchronous. Depending on the configuration of the cluster, it may take up to 1-5 minutes for the cluster to completely terminate and release allocated resources, such as Amazon EC2 instances.

Request Syntax

```
{
   "JobFlowIds": [ "string" ]
}
```

Request Parameters

For information about the parameters that are common to all actions, see Common Parameters (p. 200).

The request accepts the following data in JSON format.

JobFlowIds (p. 88)

A list of job flows to be shutdown.

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

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. 202).

InternalServerError

Indicates that an error occurred while processing the request and that the request was not completed.

HTTP Status Code: 400

Example

Sample Request

```
POST / HTTP/1.1
Content-Type: application/x-amz-json-1.1
X-Amz-Target: ElasticMapReduce.TerminateJobFlows
Content-Length: 33
User-Agent: aws-sdk-ruby/1.9.2 ruby/1.9.3 i386-mingw32
Host: us-east-1.elasticmapreduce.amazonaws.com
X-Amz-Date: 20130716T211858Z
X-Amz-Content-Sha256: ab64713f61e066e80a6083844b9249b6c6362d34a7ae7393047aa46d38b9e315
Authorization: AWS4-HMAC-SHA256 Credential=AKIAIOSFODNN7EXAMPLE/20130716/us-
east-1/elasticmapreduce/aws4_request, SignedHeaders=content-length;content-
type;host;user-agent;x-amz-content-sha256;x-amz-date;x-amz-target,
 Signature=9791416eaf09f36aa753a324b0de27ff5cc7084b8548cc748487a2bcb3439d58
Accept: */*

{"JobFlowIds": ["j-3TS0OIYO4NFN"]}
```

Sample Response

```
HTTP/1.1 200 OK
x-amzn-RequestId: 5551a7c9-ee5d-11e2-9542-25296c300ff0
Content-Type: application/x-amz-json-1.1
Content-Length: 0
Date: Tue, 16 Jul 2013 21:18:59 GMT
```

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 Go - Pilot
- 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 Amazon Elastic MapReduce 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:

Application

With Amazon EMR release version 4.0 and later, the only accepted parameter is the application name. To pass arguments to applications, you use configuration classifications specified using configuration JSON objects. For more information, see Configuring Applications.

With earlier Amazon EMR releases, the application is any Amazon or third-party software that you can add to the cluster. This structure contains a list of strings that indicates the software to use with the cluster and accepts a user argument list. Amazon EMR accepts and forwards the argument list to the corresponding installation script as bootstrap action argument.

Contents

AdditionalInfo

This option is for advanced users only. This is meta information about third-party applications that third-party vendors use for testing purposes.

Type: String to string map

Required: No

Args

Arguments for Amazon EMR to pass to the application.

Type: Array of strings

Required: No

Name

The name of the application.

Type: String

Required: No

Version

The version of the application.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

AutoScalingPolicy

An automatic scaling policy for a core instance group or task instance group in an Amazon EMR cluster. An automatic scaling policy defines how an instance group dynamically adds and terminates EC2 instances in response to the value of a CloudWatch metric. See PutAutoScalingPolicy (p. 62).

Contents

Constraints

The upper and lower EC2 instance limits for an automatic scaling policy. Automatic scaling activity will not cause an instance group to grow above or below these limits.

Type: ScalingConstraints (p. 176) object

Required: Yes

Rules

The scale-in and scale-out rules that comprise the automatic scaling policy.

Type: Array of ScalingRule (p. 177) objects

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

AutoScalingPolicyDescription

An automatic scaling policy for a core instance group or task instance group in an Amazon EMR cluster. The automatic scaling policy defines how an instance group dynamically adds and terminates EC2 instances in response to the value of a CloudWatch metric. See PutAutoScalingPolicy (p. 62).

Contents

Constraints

The upper and lower EC2 instance limits for an automatic scaling policy. Automatic scaling activity will not cause an instance group to grow above or below these limits.

Type: ScalingConstraints (p. 176) object

Required: No

Rules

The scale-in and scale-out rules that comprise the automatic scaling policy.

Type: Array of ScalingRule (p. 177) objects

Required: No

Status

The status of an automatic scaling policy.

Type: AutoScalingPolicyStatus (p. 96) 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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

AutoScalingPolicyStateChangeReason

The reason for an AutoScalingPolicyStatus (p. 96) change.

Contents

Code

The code indicating the reason for the change in status.`USER_REQUEST` indicates that the scaling policy status was changed by a user. `PROVISION_FAILURE` indicates that the status change was because the policy failed to provision. `CLEANUP_FAILURE` indicates an error.

Type: String

Valid Values: `USER_REQUEST | PROVISION_FAILURE | CLEANUP_FAILURE`

Required: No

Message

A friendly, more verbose message that accompanies an automatic scaling policy state change.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

AutoScalingPolicyStatus

The status of an automatic scaling policy.

Contents

State

Indicates the status of the automatic scaling policy.

Type: String

Valid Values: `PENDING | ATTACHING | ATTACHED | DETACHING | DETACHED | FAILED`

Required: No

StateChangeReason

The reason for a change in status.

Type: AutoScalingPolicyStateChangeReason (p. 95) 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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

BootstrapActionConfig

Configuration of a bootstrap action.

Contents

Name

The name of the bootstrap action.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: Yes

ScriptBootstrapAction

The script run by the bootstrap action.

Type: ScriptBootstrapActionConfig (p. 179) object

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

BootstrapActionDetail

Reports the configuration of a bootstrap action in a cluster (job flow).

Contents

BootstrapActionConfig

A description of the bootstrap action.

Type: BootstrapActionConfig (p. 97) 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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

CancelStepsInfo

Specification of the status of a CancelSteps request. Available only in Amazon EMR version 4.8.0 and later, excluding version 5.0.0.

Contents

Reason

The reason for the failure if the CancelSteps request fails.

Type: String

Required: No

Status

The status of a CancelSteps Request. The value may be SUBMITTED or FAILED.

Type: String

Valid Values: `SUBMITTED | FAILED`

Required: No

StepId

The encrypted StepId of a step.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

CloudWatchAlarmDefinition

The definition of a CloudWatch metric alarm, which determines when an automatic scaling activity is triggered. When the defined alarm conditions are satisfied, scaling activity begins.

Contents

ComparisonOperator

Determines how the metric specified by `MetricName` is compared to the value specified by `Threshold`.

Type: String

Valid Values: `GREATER_THAN_OR_EQUAL | GREATER_THAN | LESS_THAN | LESS_THAN_OR_EQUAL`

Required: Yes

Dimensions

A CloudWatch metric dimension.

Type: Array of MetricDimension (p. 173) objects

Required: No

EvaluationPeriods

The number of periods, expressed in seconds using `Period`, during which the alarm condition must exist before the alarm triggers automatic scaling activity. The default value is `1`.

Type: Integer

Required: No

MetricName

The name of the CloudWatch metric that is watched to determine an alarm condition.

Type: String

Required: Yes

Namespace

The namespace for the CloudWatch metric. The default is `AWS/ElasticMapReduce`.

Type: String

Required: No

Period

The period, in seconds, over which the statistic is applied. EMR CloudWatch metrics are emitted every five minutes (300 seconds), so if an EMR CloudWatch metric is specified, specify `300`.

Type: Integer

Required: Yes

Statistic

The statistic to apply to the metric associated with the alarm. The default is `AVERAGE`.

Type: String

Valid Values: `SAMPLE_COUNT | AVERAGE | SUM | MINIMUM | MAXIMUM`

Required: No

Threshold

The value against which the specified statistic is compared.

Type: Double

Valid Range: Minimum value of 0.0.

Required: Yes

Unit

The unit of measure associated with the CloudWatch metric being watched. The value specified for `Unit` must correspond to the units specified in the CloudWatch metric.

Type: String

Valid Values: `NONE | SECONDS | MICRO_SECONDS | MILLI_SECONDS | BYTES | KILO_BYTES | MEGA_BYTES | GIGA_BYTES | TERA_BYTES | BITS | KILO_BITS | MEGA_BITS | GIGA_BITS | TERA_BITS | PERCENT | COUNT | BYTES_PER_SECOND | KILO_BYTES_PER_SECOND | MEGA_BYTES_PER_SECOND | GIGA_BYTES_PER_SECOND | TERA_BYTES_PER_SECOND | BITS_PER_SECOND | KILO_BITS_PER_SECOND | MEGA_BITS_PER_SECOND | GIGA_BITS_PER_SECOND | TERA_BITS_PER_SECOND | COUNT_PER_SECOND`

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

Cluster

The detailed description of the cluster.

Contents

Applications

The applications installed on this cluster.

Type: Array of Application (p. 92) objects

Required: No

AutoScalingRole

An IAM role for automatic scaling policies. The default role is `EMR_AutoScaling_DefaultRole`. The IAM role provides permissions that the automatic scaling feature requires to launch and terminate EC2 instances in an instance group.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

AutoTerminate

Specifies whether the cluster should terminate after completing all steps.

Type: Boolean

Required: No

Configurations

Applies only to Amazon EMR releases 4.x and later. The list of Configurations supplied to the EMR cluster.

Type: Array of Configuration (p. 112) objects

Required: No

CustomAmiId

Available only in Amazon EMR version 5.7.0 and later. The ID of a custom Amazon EBS-backed Linux AMI if the cluster uses a custom AMI.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

EbsRootVolumeSize

The size, in GiB, of the EBS root device volume of the Linux AMI that is used for each EC2 instance. Available in Amazon EMR version 4.x and later.

Type: Integer

Required: No

Ec2InstanceAttributes

Provides information about the EC2 instances in a cluster grouped by category. For example, key name, subnet ID, IAM instance profile, and so on.

Type: Ec2InstanceAttributes (p. 117) object

Required: No

Id

The unique identifier for the cluster.

Type: String

Required: No

InstanceCollectionType

> **Note**
> The instance fleet configuration is available only in Amazon EMR versions 4.8.0 and later, excluding 5.0.x versions.

The instance group configuration of the cluster. A value of `INSTANCE_GROUP` indicates a uniform instance group configuration. A value of `INSTANCE_FLEET` indicates an instance fleets configuration.

Type: String

Valid Values: `INSTANCE_FLEET | INSTANCE_GROUP`

Required: No

KerberosAttributes

Attributes for Kerberos configuration when Kerberos authentication is enabled using a security configuration. For more information see Use Kerberos Authentication in the *EMR Management Guide*.

Type: KerberosAttributes (p. 170) object

Required: No

LogUri

The path to the Amazon S3 location where logs for this cluster are stored.

Type: String

Required: No

MasterPublicDnsName

The DNS name of the master node. If the cluster is on a private subnet, this is the private DNS name. On a public subnet, this is the public DNS name.

Type: String

Required: No

Name

The name of the cluster.

Type: String

Required: No

NormalizedInstanceHours

An approximation of the cost of the cluster, represented in m1.small/hours. This value is incremented one time for every hour an m1.small instance runs. Larger instances are weighted more, so an EC2 instance that is roughly four times more expensive would result in the normalized instance hours being incremented by four. This result is only an approximation and does not reflect the actual billing rate.

Type: Integer

Required: No

ReleaseLabel

The Amazon EMR release label, which determines the version of open-source application packages installed on the cluster. Release labels are in the form `emr-x.x.x`, where x.x.x is an Amazon EMR release version, for example, `emr-5.14.0`. For more information about Amazon EMR release versions and included application versions and features, see https://docs.aws.amazon.com/emr/latest/ReleaseGuide/. The release label applies only to Amazon EMR releases versions 4.x and later. Earlier versions use `AmiVersion`.

Type: String

Required: No

RepoUpgradeOnBoot

Applies only when `CustomAmiID` is used. Specifies the type of updates that are applied from the Amazon Linux AMI package repositories when an instance boots using the AMI.

Type: String

Valid Values: `SECURITY | NONE`

Required: No

RequestedAmiVersion

The AMI version requested for this cluster.

Type: String

Required: No

RunningAmiVersion

The AMI version running on this cluster.

Type: String

Required: No

ScaleDownBehavior

The way that individual Amazon EC2 instances terminate when an automatic scale-in activity occurs or an instance group is resized. `TERMINATE_AT_INSTANCE_HOUR` indicates that Amazon EMR terminates nodes at the instance-hour boundary, regardless of when the request to terminate the instance was submitted. This option is only available with Amazon EMR 5.1.0 and later and is the default for clusters created using that version. `TERMINATE_AT_TASK_COMPLETION` indicates that Amazon EMR blacklists and drains tasks from nodes before terminating the Amazon EC2 instances, regardless of the instance-hour boundary. With either behavior, Amazon EMR removes

the least active nodes first and blocks instance termination if it could lead to HDFS corruption. `TERMINATE_AT_TASK_COMPLETION` is available only in Amazon EMR version 4.1.0 and later, and is the default for versions of Amazon EMR earlier than 5.1.0.

Type: String

Valid Values: `TERMINATE_AT_INSTANCE_HOUR | TERMINATE_AT_TASK_COMPLETION`

Required: No

SecurityConfiguration

The name of the security configuration applied to the cluster.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

ServiceRole

The IAM role that will be assumed by the Amazon EMR service to access AWS resources on your behalf.

Type: String

Required: No

Status

The current status details about the cluster.

Type: ClusterStatus (p. 108) object

Required: No

Tags

A list of tags associated with a cluster.

Type: Array of Tag (p. 198) objects

Required: No

TerminationProtected

Indicates whether Amazon EMR will lock the cluster to prevent the EC2 instances from being terminated by an API call or user intervention, or in the event of a cluster error.

Type: Boolean

Required: No

VisibleToAllUsers

Indicates whether the cluster is visible to all IAM users of the AWS account associated with the cluster. The default value, `true`, indicates that all IAM users in the AWS account can perform cluster actions if they have the proper IAM policy permissions. If this value is `false`, only the IAM user that created the cluster can perform actions. This value can be changed on a running cluster by using the SetVisibleToAllUsers (p. 85) action. You can override the default value of `true` when you create a cluster by using the `VisibleToAllUsers` parameter of the `RunJobFlow` action.

Type: Boolean

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

ClusterStateChangeReason

The reason that the cluster changed to its current state.

Contents

Code

The programmatic code for the state change reason.

Type: String

Valid Values: `INTERNAL_ERROR | VALIDATION_ERROR | INSTANCE_FAILURE | INSTANCE_FLEET_TIMEOUT | BOOTSTRAP_FAILURE | USER_REQUEST | STEP_FAILURE | ALL_STEPS_COMPLETED`

Required: No

Message

The descriptive message for the state change reason.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

ClusterStatus

The detailed status of the cluster.

Contents

State

The current state of the cluster.

Type: String

Valid Values: `STARTING | BOOTSTRAPPING | RUNNING | WAITING | TERMINATING | TERMINATED | TERMINATED_WITH_ERRORS`

Required: No

StateChangeReason

The reason for the cluster status change.

Type: ClusterStateChangeReason (p. 107) object

Required: No

Timeline

A timeline that represents the status of a cluster over the lifetime of the cluster.

Type: ClusterTimeline (p. 110) 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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

ClusterSummary

The summary description of the cluster.

Contents

Id

The unique identifier for the cluster.

Type: String

Required: No

Name

The name of the cluster.

Type: String

Required: No

NormalizedInstanceHours

An approximation of the cost of the cluster, represented in m1.small/hours. This value is incremented one time for every hour an m1.small instance runs. Larger instances are weighted more, so an EC2 instance that is roughly four times more expensive would result in the normalized instance hours being incremented by four. This result is only an approximation and does not reflect the actual billing rate.

Type: Integer

Required: No

Status

The details about the current status of the cluster.

Type: ClusterStatus (p. 108) 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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

ClusterTimeline

Represents the timeline of the cluster's lifecycle.

Contents

CreationDateTime

The creation date and time of the cluster.

Type: Timestamp

Required: No

EndDateTime

The date and time when the cluster was terminated.

Type: Timestamp

Required: No

ReadyDateTime

The date and time when the cluster was ready to execute steps.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

Command

An entity describing an executable that runs on a cluster.

Contents

Args

Arguments for Amazon EMR to pass to the command for execution.

Type: Array of strings

Required: No

Name

The name of the command.

Type: String

Required: No

ScriptPath

The Amazon S3 location of the command script.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

Configuration

Note
Amazon EMR releases 4.x or later.

An optional configuration specification to be used when provisioning cluster instances, which can include configurations for applications and software bundled with Amazon EMR. A configuration consists of a classification, properties, and optional nested configurations. A classification refers to an application-specific configuration file. Properties are the settings you want to change in that file. For more information, see Configuring Applications.

Contents

Classification

The classification within a configuration.

Type: String

Required: No

Configurations

A list of additional configurations to apply within a configuration object.

Type: Array of Configuration (p. 112) objects

Required: No

Properties

A set of properties specified within a configuration classification.

Type: String to string map

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

EbsBlockDevice

Configuration of requested EBS block device associated with the instance group.

Contents

Device

The device name that is exposed to the instance, such as /dev/sdh.

Type: String

Required: No

VolumeSpecification

EBS volume specifications such as volume type, IOPS, and size (GiB) that will be requested for the EBS volume attached to an EC2 instance in the cluster.

Type: VolumeSpecification (p. 199) 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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

EbsBlockDeviceConfig

Configuration of requested EBS block device associated with the instance group with count of volumes that will be associated to every instance.

Contents

VolumeSpecification

EBS volume specifications such as volume type, IOPS, and size (GiB) that will be requested for the EBS volume attached to an EC2 instance in the cluster.

Type: VolumeSpecification (p. 199) object

Required: Yes

VolumesPerInstance

Number of EBS volumes with a specific volume configuration that will be associated with every instance in the instance group

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

EbsConfiguration

The Amazon EBS configuration of a cluster instance.

Contents

EbsBlockDeviceConfigs

An array of Amazon EBS volume specifications attached to a cluster instance.

Type: Array of EbsBlockDeviceConfig (p. 114) objects

Required: No

EbsOptimized

Indicates whether an Amazon EBS volume is EBS-optimized.

Type: Boolean

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

EbsVolume

EBS block device that's attached to an EC2 instance.

Contents

Device

The device name that is exposed to the instance, such as /dev/sdh.

Type: String

Required: No

VolumeId

The volume identifier of the EBS volume.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

Ec2InstanceAttributes

Provides information about the EC2 instances in a cluster grouped by category. For example, key name, subnet ID, IAM instance profile, and so on.

Contents

AdditionalMasterSecurityGroups

A list of additional Amazon EC2 security group IDs for the master node.

Type: Array of strings

Required: No

AdditionalSlaveSecurityGroups

A list of additional Amazon EC2 security group IDs for the core and task nodes.

Type: Array of strings

Required: No

Ec2AvailabilityZone

The Availability Zone in which the cluster will run.

Type: String

Required: No

Ec2KeyName

The name of the Amazon EC2 key pair to use when connecting with SSH into the master node as a user named "hadoop".

Type: String

Required: No

Ec2SubnetId

To launch the cluster in Amazon VPC, set this parameter to the identifier of the Amazon VPC subnet where you want the cluster to launch. If you do not specify this value, the cluster is launched in the normal AWS cloud, outside of a VPC.

Amazon VPC currently does not support cluster compute quadruple extra large (cc1.4xlarge) instances. Thus, you cannot specify the cc1.4xlarge instance type for nodes of a cluster launched in a VPC.

Type: String

Required: No

EmrManagedMasterSecurityGroup

The identifier of the Amazon EC2 security group for the master node.

Type: String

Required: No

EmrManagedSlaveSecurityGroup

The identifier of the Amazon EC2 security group for the core and task nodes.

Type: String

Required: No

IamInstanceProfile

The IAM role that was specified when the cluster was launched. The EC2 instances of the cluster assume this role.

Type: String

Required: No

RequestedEc2AvailabilityZones

Applies to clusters configured with the instance fleets option. Specifies one or more Availability Zones in which to launch EC2 cluster instances when the EC2-Classic network configuration is supported. Amazon EMR chooses the Availability Zone with the best fit from among the list of `RequestedEc2AvailabilityZones`, and then launches all cluster instances within that Availability Zone. If you do not specify this value, Amazon EMR chooses the Availability Zone for you. `RequestedEc2SubnetIDs` and `RequestedEc2AvailabilityZones` cannot be specified together.

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

RequestedEc2SubnetIds

Applies to clusters configured with the instance fleets option. Specifies the unique identifier of one or more Amazon EC2 subnets in which to launch EC2 cluster instances. Subnets must exist within the same VPC. Amazon EMR chooses the EC2 subnet with the best fit from among the list of `RequestedEc2SubnetIds`, and then launches all cluster instances within that Subnet. If this value is not specified, and the account and region support EC2-Classic networks, the cluster launches instances in the EC2-Classic network and uses `RequestedEc2AvailabilityZones` instead of this setting. If EC2-Classic is not supported, and no Subnet is specified, Amazon EMR chooses the subnet for you. `RequestedEc2SubnetIDs` and `RequestedEc2AvailabilityZones` cannot be specified together.

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

ServiceAccessSecurityGroup

The identifier of the Amazon EC2 security group for the Amazon EMR service to access clusters in VPC private subnets.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

FailureDetails

The details of the step failure. The service attempts to detect the root cause for many common failures.

Contents

LogFile

The path to the log file where the step failure root cause was originally recorded.

Type: String

Required: No

Message

The descriptive message including the error the EMR service has identified as the cause of step failure. This is text from an error log that describes the root cause of the failure.

Type: String

Required: No

Reason

The reason for the step failure. In the case where the service cannot successfully determine the root cause of the failure, it returns "Unknown Error" as a reason.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

HadoopJarStepConfig

A job flow step consisting of a JAR file whose main function will be executed. The main function submits a job for Hadoop to execute and waits for the job to finish or fail.

Contents

Args

A list of command line arguments passed to the JAR file's main function when executed.

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

Jar

A path to a JAR file run during the step.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: Yes

MainClass

The name of the main class in the specified Java file. If not specified, the JAR file should specify a Main-Class in its manifest file.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

Properties

A list of Java properties that are set when the step runs. You can use these properties to pass key value pairs to your main function.

Type: Array of KeyValue (p. 172) 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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

HadoopStepConfig

A cluster step consisting of a JAR file whose main function will be executed. The main function submits a job for Hadoop to execute and waits for the job to finish or fail.

Contents

Args

The list of command line arguments to pass to the JAR file's main function for execution.

Type: Array of strings

Required: No

Jar

The path to the JAR file that runs during the step.

Type: String

Required: No

MainClass

The name of the main class in the specified Java file. If not specified, the JAR file should specify a main class in its manifest file.

Type: String

Required: No

Properties

The list of Java properties that are set when the step runs. You can use these properties to pass key value pairs to your main function.

Type: String to string map

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

Instance

Represents an EC2 instance provisioned as part of cluster.

Contents

EbsVolumes

The list of EBS volumes that are attached to this instance.

Type: Array of EbsVolume (p. 116) objects

Required: No

Ec2InstanceId

The unique identifier of the instance in Amazon EC2.

Type: String

Required: No

Id

The unique identifier for the instance in Amazon EMR.

Type: String

Required: No

InstanceFleetId

The unique identifier of the instance fleet to which an EC2 instance belongs.

Type: String

Required: No

InstanceGroupId

The identifier of the instance group to which this instance belongs.

Type: String

Required: No

InstanceType

The EC2 instance type, for example `m3.xlarge`.

Type: String

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

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

Required: No

Market

The instance purchasing option. Valid values are `ON_DEMAND` or `SPOT`.

Type: String

Valid Values: `ON_DEMAND | SPOT`

Required: No

PrivateDnsName

The private DNS name of the instance.

Type: String

Required: No

PrivateIpAddress

The private IP address of the instance.

Type: String

Required: No

PublicDnsName

The public DNS name of the instance.

Type: String

Required: No

PublicIpAddress

The public IP address of the instance.

Type: String

Required: No

Status

The current status of the instance.

Type: InstanceStatus (p. 152) 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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceFleet

Describes an instance fleet, which is a group of EC2 instances that host a particular node type (master, core, or task) in an Amazon EMR cluster. Instance fleets can consist of a mix of instance types and On-Demand and Spot instances, which are provisioned to meet a defined target capacity.

Note
The instance fleet configuration is available only in Amazon EMR versions 4.8.0 and later, excluding 5.0.x versions.

Contents

Id

The unique identifier of the instance fleet.

Type: String

Required: No

InstanceFleetType

The node type that the instance fleet hosts. Valid values are MASTER, CORE, or TASK.

Type: String

Valid Values: `MASTER | CORE | TASK`

Required: No

InstanceTypeSpecifications

The specification for the instance types that comprise an instance fleet. Up to five unique instance specifications may be defined for each instance fleet.

Type: Array of InstanceTypeSpecification (p. 156) objects

Required: No

LaunchSpecifications

Describes the launch specification for an instance fleet.

Type: InstanceFleetProvisioningSpecifications (p. 132) object

Required: No

Name

A friendly name for the instance fleet.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

ProvisionedOnDemandCapacity

The number of On-Demand units that have been provisioned for the instance fleet to fulfill `TargetOnDemandCapacity`. This provisioned capacity might be less than or greater than `TargetOnDemandCapacity`.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

ProvisionedSpotCapacity

The number of Spot units that have been provisioned for this instance fleet to fulfill `TargetSpotCapacity`. This provisioned capacity might be less than or greater than `TargetSpotCapacity`.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

Status

The current status of the instance fleet.

Type: InstanceFleetStatus (p. 134) object

Required: No

TargetOnDemandCapacity

The target capacity of On-Demand units for the instance fleet, which determines how many On-Demand instances to provision. When the instance fleet launches, Amazon EMR tries to provision On-Demand instances as specified by InstanceTypeConfig (p. 154). Each instance configuration has a specified `WeightedCapacity`. When an On-Demand instance is provisioned, the `WeightedCapacity` units count toward the target capacity. Amazon EMR provisions instances until the target capacity is totally fulfilled, even if this results in an overage. For example, if there are 2 units remaining to fulfill capacity, and Amazon EMR can only provision an instance with a `WeightedCapacity` of 5 units, the instance is provisioned, and the target capacity is exceeded by 3 units. You can use InstanceFleet:ProvisionedOnDemandCapacity (p. 126) to determine the Spot capacity units that have been provisioned for the instance fleet.

> **Note**
> If not specified or set to 0, only Spot instances are provisioned for the instance fleet using `TargetSpotCapacity`. At least one of `TargetSpotCapacity` and `TargetOnDemandCapacity` should be greater than 0. For a master instance fleet, only one of `TargetSpotCapacity` and `TargetOnDemandCapacity` can be specified, and its value must be 1.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

TargetSpotCapacity

The target capacity of Spot units for the instance fleet, which determines how many Spot instances to provision. When the instance fleet launches, Amazon EMR tries to provision Spot instances as specified by InstanceTypeConfig (p. 154). Each instance configuration has a specified `WeightedCapacity`. When a Spot instance is provisioned, the `WeightedCapacity` units count toward the target capacity. Amazon EMR provisions instances until the target capacity is totally fulfilled, even if this results in an overage. For example, if there are 2 units remaining to fulfill capacity, and Amazon EMR can only provision an instance with a `WeightedCapacity` of 5 units, the instance is provisioned, and the target capacity is exceeded by 3 units. You can use InstanceFleet:ProvisionedSpotCapacity (p. 127) to determine the Spot capacity units that have been provisioned for the instance fleet.

Note
If not specified or set to 0, only On-Demand instances are provisioned for the instance fleet. At least one of `TargetSpotCapacity` and `TargetOnDemandCapacity` should be greater than 0. For a master instance fleet, only one of `TargetSpotCapacity` and `TargetOnDemandCapacity` can be specified, and its value must be 1.

Type: Integer

Valid Range: Minimum value of 0.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceFleetConfig

The configuration that defines an instance fleet.

Note
The instance fleet configuration is available only in Amazon EMR versions 4.8.0 and later, excluding 5.0.x versions.

Contents

InstanceFleetType

The node type that the instance fleet hosts. Valid values are MASTER,CORE,and TASK.

Type: String

Valid Values: `MASTER | CORE | TASK`

Required: Yes

InstanceTypeConfigs

The instance type configurations that define the EC2 instances in the instance fleet.

Type: Array of InstanceTypeConfig (p. 154) objects

Required: No

LaunchSpecifications

The launch specification for the instance fleet.

Type: InstanceFleetProvisioningSpecifications (p. 132) object

Required: No

Name

The friendly name of the instance fleet.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

TargetOnDemandCapacity

The target capacity of On-Demand units for the instance fleet, which determines how many On-Demand instances to provision. When the instance fleet launches, Amazon EMR tries to provision On-Demand instances as specified by InstanceTypeConfig (p. 154). Each instance configuration has a specified `WeightedCapacity`. When an On-Demand instance is provisioned, the `WeightedCapacity` units count toward the target capacity. Amazon EMR provisions instances until the target capacity is totally fulfilled, even if this results in an overage. For example, if there are 2 units remaining to fulfill capacity, and Amazon EMR can only provision an instance with a `WeightedCapacity` of 5 units, the instance is provisioned, and the target capacity is exceeded by 3 units.

Note
If not specified or set to 0, only Spot instances are provisioned for the instance fleet using `TargetSpotCapacity`. At least one of `TargetSpotCapacity` and

`TargetOnDemandCapacity` should be greater than 0. For a master instance fleet, only one of `TargetSpotCapacity` and `TargetOnDemandCapacity` can be specified, and its value must be 1.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

TargetSpotCapacity

The target capacity of Spot units for the instance fleet, which determines how many Spot instances to provision. When the instance fleet launches, Amazon EMR tries to provision Spot instances as specified by InstanceTypeConfig (p. 154). Each instance configuration has a specified `WeightedCapacity`. When a Spot instance is provisioned, the `WeightedCapacity` units count toward the target capacity. Amazon EMR provisions instances until the target capacity is totally fulfilled, even if this results in an overage. For example, if there are 2 units remaining to fulfill capacity, and Amazon EMR can only provision an instance with a `WeightedCapacity` of 5 units, the instance is provisioned, and the target capacity is exceeded by 3 units.

Note
If not specified or set to 0, only On-Demand instances are provisioned for the instance fleet. At least one of `TargetSpotCapacity` and `TargetOnDemandCapacity` should be greater than 0. For a master instance fleet, only one of `TargetSpotCapacity` and `TargetOnDemandCapacity` can be specified, and its value must be 1.

Type: Integer

Valid Range: Minimum value of 0.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceFleetModifyConfig

Configuration parameters for an instance fleet modification request.

Note
The instance fleet configuration is available only in Amazon EMR versions 4.8.0 and later, excluding 5.0.x versions.

Contents

InstanceFleetId

A unique identifier for the instance fleet.

Type: String

Required: Yes

TargetOnDemandCapacity

The target capacity of On-Demand units for the instance fleet. For more information see InstanceFleetConfig:TargetOnDemandCapacity (p. 129).

Type: Integer

Valid Range: Minimum value of 0.

Required: No

TargetSpotCapacity

The target capacity of Spot units for the instance fleet. For more information, see InstanceFleetConfig:TargetSpotCapacity (p. 130).

Type: Integer

Valid Range: Minimum value of 0.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceFleetProvisioningSpecifications

The launch specification for Spot instances in the fleet, which determines the defined duration and provisioning timeout behavior.

Note
The instance fleet configuration is available only in Amazon EMR versions 4.8.0 and later, excluding 5.0.x versions.

Contents

SpotSpecification

The launch specification for Spot instances in the fleet, which determines the defined duration and provisioning timeout behavior.

Type: SpotProvisioningSpecification (p. 184) object

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceFleetStateChangeReason

Provides status change reason details for the instance fleet.

Note
The instance fleet configuration is available only in Amazon EMR versions 4.8.0 and later, excluding 5.0.x versions.

Contents

Code

A code corresponding to the reason the state change occurred.

Type: String

Valid Values: `INTERNAL_ERROR | VALIDATION_ERROR | INSTANCE_FAILURE | CLUSTER_TERMINATED`

Required: No

Message

An explanatory message.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceFleetStatus

The status of the instance fleet.

Note
The instance fleet configuration is available only in Amazon EMR versions 4.8.0 and later, excluding 5.0.x versions.

Contents

State

A code representing the instance fleet status.

- `PROVISIONING`—The instance fleet is provisioning EC2 resources and is not yet ready to run jobs.
- `BOOTSTRAPPING`—EC2 instances and other resources have been provisioned and the bootstrap actions specified for the instances are underway.
- `RUNNING`—EC2 instances and other resources are running. They are either executing jobs or waiting to execute jobs.
- `RESIZING`—A resize operation is underway. EC2 instances are either being added or removed.
- `SUSPENDED`—A resize operation could not complete. Existing EC2 instances are running, but instances can't be added or removed.
- `TERMINATING`—The instance fleet is terminating EC2 instances.
- `TERMINATED`—The instance fleet is no longer active, and all EC2 instances have been terminated.

Type: String

Valid Values: `PROVISIONING | BOOTSTRAPPING | RUNNING | RESIZING | SUSPENDED | TERMINATING | TERMINATED`

Required: No

StateChangeReason

Provides status change reason details for the instance fleet.

Type: InstanceFleetStateChangeReason (p. 133) object

Required: No

Timeline

Provides historical timestamps for the instance fleet, including the time of creation, the time it became ready to run jobs, and the time of termination.

Type: InstanceFleetTimeline (p. 136) 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 Go - Pilot

- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceFleetTimeline

Provides historical timestamps for the instance fleet, including the time of creation, the time it became ready to run jobs, and the time of termination.

Note
The instance fleet configuration is available only in Amazon EMR versions 4.8.0 and later, excluding 5.0.x versions.

Contents

CreationDateTime

The time and date the instance fleet was created.

Type: Timestamp

Required: No

EndDateTime

The time and date the instance fleet terminated.

Type: Timestamp

Required: No

ReadyDateTime

The time and date the instance fleet was ready to run jobs.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceGroup

This entity represents an instance group, which is a group of instances that have common purpose. For example, CORE instance group is used for HDFS.

Contents

AutoScalingPolicy

An automatic scaling policy for a core instance group or task instance group in an Amazon EMR cluster. The automatic scaling policy defines how an instance group dynamically adds and terminates EC2 instances in response to the value of a CloudWatch metric. See PutAutoScalingPolicy.

Type: AutoScalingPolicyDescription (p. 94) object

Required: No

BidPrice

The maximum Spot price your are willing to pay for EC2 instances.

If `BidPrice` is specified, Amazon EMR uses Spot Instances for the instance group. Specified in USD. Alternatively, a value of `OnDemandPrice` indicates that the maximum Spot price is set equal to the On-Demand price.

Type: String

Required: No

Configurations

Note
Amazon EMR releases 4.x or later.

The list of configurations supplied for an EMR cluster instance group. You can specify a separate configuration for each instance group (master, core, and task).

Type: Array of Configuration (p. 112) objects

Required: No

ConfigurationsVersion

The version number of the requested configuration specification for this instance group.

Type: Long

Required: No

EbsBlockDevices

The EBS block devices that are mapped to this instance group.

Type: Array of EbsBlockDevice (p. 113) objects

Required: No

EbsOptimized

If the instance group is EBS-optimized. An Amazon EBS-optimized instance uses an optimized configuration stack and provides additional, dedicated capacity for Amazon EBS I/O.

Type: Boolean

Required: No

Id

The identifier of the instance group.

Type: String

Required: No

InstanceGroupType

The type of the instance group. Valid values are MASTER, CORE or TASK.

Type: String

Valid Values: `MASTER | CORE | TASK`

Required: No

InstanceType

The EC2 instance type for all instances in the instance group.

Type: String

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

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

Required: No

LastSuccessfullyAppliedConfigurations

A list of configurations that were successfully applied for an instance group last time.

Type: Array of Configuration (p. 112) objects

Required: No

LastSuccessfullyAppliedConfigurationsVersion

The version number of a configuration specification that was successfully applied for an instance group last time.

Type: Long

Required: No

Market

The marketplace to provision instances for this group. Valid values are ON_DEMAND or SPOT.

Type: String

Valid Values: `ON_DEMAND | SPOT`

Required: No

Name

The name of the instance group.

Type: String

Required: No

RequestedInstanceCount

The target number of instances for the instance group.

Type: Integer

Required: No

RunningInstanceCount

The number of instances currently running in this instance group.

Type: Integer

Required: No

ShrinkPolicy

Policy for customizing shrink operations.

Type: ShrinkPolicy (p. 181) object

Required: No

Status

The current status of the instance group.

Type: InstanceGroupStatus (p. 148) 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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceGroupConfig

Configuration defining a new instance group.

Contents

AutoScalingPolicy

An automatic scaling policy for a core instance group or task instance group in an Amazon EMR cluster. The automatic scaling policy defines how an instance group dynamically adds and terminates EC2 instances in response to the value of a CloudWatch metric. See PutAutoScalingPolicy (p. 62).

Type: AutoScalingPolicy (p. 93) object

Required: No

BidPrice

The maximum Spot price your are willing to pay for EC2 instances.

If `BidPrice` is specified, Amazon EMR uses Spot Instances for the instance group. Specified in USD. Alternatively, a value of `OnDemandPrice` indicates that the maximum Spot price is set equal to the On-Demand price.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

Configurations

Note
Amazon EMR releases 4.x or later.

The list of configurations supplied for an EMR cluster instance group. You can specify a separate configuration for each instance group (master, core, and task).

Type: Array of Configuration (p. 112) objects

Required: No

EbsConfiguration

EBS configurations that will be attached to each EC2 instance in the instance group.

Type: EbsConfiguration (p. 115) object

Required: No

InstanceCount

Target number of instances for the instance group.

Type: Integer

Required: Yes

InstanceRole

The role of the instance group in the cluster.

Type: String

Valid Values: `MASTER | CORE | TASK`

Required: Yes

InstanceType

The EC2 instance type for all instances in the instance group.

Type: String

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

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

Required: Yes

Market

Market type of the EC2 instances used to create a cluster node.

Type: String

Valid Values: `ON_DEMAND | SPOT`

Required: No

Name

Friendly name given to the instance group.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceGroupDetail

Detailed information about an instance group.

Contents

BidPrice

The maximum Spot price your are willing to pay for EC2 instances.

If `BidPrice` is specified, Amazon EMR uses Spot Instances for the instance group. Specified in USD. Alternatively, a value of `OnDemandPrice` indicates that the maximum Spot price is set equal to the On-Demand price.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

CreationDateTime

The date/time the instance group was created.

Type: Timestamp

Required: Yes

EndDateTime

The date/time the instance group was terminated.

Type: Timestamp

Required: No

InstanceGroupId

Unique identifier for the instance group.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

InstanceRequestCount

Target number of instances to run in the instance group.

Type: Integer

Required: Yes

InstanceRole

Instance group role in the cluster

Type: String

Valid Values: `MASTER | CORE | TASK`

Required: Yes

InstanceRunningCount

Actual count of running instances.

Type: Integer

Required: Yes

InstanceType

EC2 instance type.

Type: String

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

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

Required: Yes

LastStateChangeReason

Details regarding the state of the instance group.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

Market

Market type of the EC2 instances used to create a cluster node.

Type: String

Valid Values: `ON_DEMAND | SPOT`

Required: Yes

Name

Friendly name for the instance group.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

ReadyDateTime

The date/time the instance group was available to the cluster.

Type: Timestamp

Required: No

StartDateTime

The date/time the instance group was started.

Type: Timestamp

Required: No

State

State of instance group. The following values are deprecated: STARTING, TERMINATED, and FAILED.

Type: String

Valid Values: `PROVISIONING | BOOTSTRAPPING | RUNNING | RECONFIGURING | RESIZING | SUSPENDED | TERMINATING | TERMINATED | ARRESTED | SHUTTING_DOWN | ENDED`

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceGroupModifyConfig

Modify the size or configurations of an instance group.

Contents

Configurations

A list of new or modified configurations to apply for an instance group.

Type: Array of Configuration (p. 112) objects

Required: No

EC2InstanceIdsToTerminate

The EC2 InstanceIds to terminate. After you terminate the instances, the instance group will not return to its original requested size.

Type: Array of strings

Required: No

InstanceCount

Target size for the instance group.

Type: Integer

Required: No

InstanceGroupId

Unique ID of the instance group to expand or shrink.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: Yes

ShrinkPolicy

Policy for customizing shrink operations.

Type: ShrinkPolicy (p. 181) 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 Go - Pilot
- AWS SDK for Java

- AWS SDK for Ruby V2

InstanceGroupStateChangeReason

The status change reason details for the instance group.

Contents

Code

The programmable code for the state change reason.

Type: String

Valid Values: `INTERNAL_ERROR | VALIDATION_ERROR | INSTANCE_FAILURE | CLUSTER_TERMINATED`

Required: No

Message

The status change reason description.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceGroupStatus

The details of the instance group status.

Contents

State

The current state of the instance group.

Type: String

Valid Values: `PROVISIONING | BOOTSTRAPPING | RUNNING | RECONFIGURING | RESIZING | SUSPENDED | TERMINATING | TERMINATED | ARRESTED | SHUTTING_DOWN | ENDED`

Required: No

StateChangeReason

The status change reason details for the instance group.

Type: InstanceGroupStateChangeReason (p. 147) object

Required: No

Timeline

The timeline of the instance group status over time.

Type: InstanceGroupTimeline (p. 149) 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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceGroupTimeline

The timeline of the instance group lifecycle.

Contents

CreationDateTime

The creation date and time of the instance group.

Type: Timestamp

Required: No

EndDateTime

The date and time when the instance group terminated.

Type: Timestamp

Required: No

ReadyDateTime

The date and time when the instance group became ready to perform tasks.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceResizePolicy

Custom policy for requesting termination protection or termination of specific instances when shrinking an instance group.

Contents

InstancesToProtect

Specific list of instances to be protected when shrinking an instance group.

Type: Array of strings

Required: No

InstancesToTerminate

Specific list of instances to be terminated when shrinking an instance group.

Type: Array of strings

Required: No

InstanceTerminationTimeout

Decommissioning timeout override for the specific list of instances to be terminated.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceStateChangeReason

The details of the status change reason for the instance.

Contents

Code

The programmable code for the state change reason.

Type: String

Valid Values: `INTERNAL_ERROR | VALIDATION_ERROR | INSTANCE_FAILURE | BOOTSTRAP_FAILURE | CLUSTER_TERMINATED`

Required: No

Message

The status change reason description.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceStatus

The instance status details.

Contents

State

The current state of the instance.

Type: String

Valid Values: `AWAITING_FULFILLMENT | PROVISIONING | BOOTSTRAPPING | RUNNING | TERMINATED`

Required: No

StateChangeReason

The details of the status change reason for the instance.

Type: InstanceStateChangeReason (p. 151) object

Required: No

Timeline

The timeline of the instance status over time.

Type: InstanceTimeline (p. 153) 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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceTimeline

The timeline of the instance lifecycle.

Contents

CreationDateTime

The creation date and time of the instance.

Type: Timestamp

Required: No

EndDateTime

The date and time when the instance was terminated.

Type: Timestamp

Required: No

ReadyDateTime

The date and time when the instance was ready to perform tasks.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceTypeConfig

An instance type configuration for each instance type in an instance fleet, which determines the EC2 instances Amazon EMR attempts to provision to fulfill On-Demand and Spot target capacities. There can be a maximum of 5 instance type configurations in a fleet.

Note
The instance fleet configuration is available only in Amazon EMR versions 4.8.0 and later, excluding 5.0.x versions.

Contents

BidPrice

The bid price for each EC2 Spot instance type as defined by `InstanceType`. Expressed in USD. If neither `BidPrice` nor `BidPriceAsPercentageOfOnDemandPrice` is provided, `BidPriceAsPercentageOfOnDemandPrice` defaults to 100%.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

BidPriceAsPercentageOfOnDemandPrice

The bid price, as a percentage of On-Demand price, for each EC2 Spot instance as defined by `InstanceType`. Expressed as a number (for example, 20 specifies 20%). If neither `BidPrice` nor `BidPriceAsPercentageOfOnDemandPrice` is provided, `BidPriceAsPercentageOfOnDemandPrice` defaults to 100%.

Type: Double

Valid Range: Minimum value of 0.0.

Required: No

Configurations

A configuration classification that applies when provisioning cluster instances, which can include configurations for applications and software that run on the cluster.

Type: Array of Configuration (p. 112) objects

Required: No

EbsConfiguration

The configuration of Amazon Elastic Block Storage (EBS) attached to each instance as defined by `InstanceType`.

Type: EbsConfiguration (p. 115) object

Required: No

InstanceType

An EC2 instance type, such as `m3.xlarge`.

Type: String

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

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

Required: Yes

WeightedCapacity

The number of units that a provisioned instance of this type provides toward fulfilling the target capacities defined in InstanceFleetConfig (p. 129). This value is 1 for a master instance fleet, and must be 1 or greater for core and task instance fleets. Defaults to 1 if not specified.

Type: Integer

Valid Range: Minimum value of 0.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

InstanceTypeSpecification

The configuration specification for each instance type in an instance fleet.

Note
The instance fleet configuration is available only in Amazon EMR versions 4.8.0 and later, excluding 5.0.x versions.

Contents

BidPrice

The bid price for each EC2 Spot instance type as defined by `InstanceType`. Expressed in USD.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

BidPriceAsPercentageOfOnDemandPrice

The bid price, as a percentage of On-Demand price, for each EC2 Spot instance as defined by `InstanceType`. Expressed as a number (for example, 20 specifies 20%).

Type: Double

Valid Range: Minimum value of 0.0.

Required: No

Configurations

A configuration classification that applies when provisioning cluster instances, which can include configurations for applications and software bundled with Amazon EMR.

Type: Array of Configuration (p. 112) objects

Required: No

EbsBlockDevices

The configuration of Amazon Elastic Block Storage (EBS) attached to each instance as defined by `InstanceType`.

Type: Array of EbsBlockDevice (p. 113) objects

Required: No

EbsOptimized

Evaluates to `TRUE` when the specified `InstanceType` is EBS-optimized.

Type: Boolean

Required: No

InstanceType

The EC2 instance type, for example `m3.xlarge`.

Type: String

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

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

Required: No

WeightedCapacity

The number of units that a provisioned instance of this type provides toward fulfilling the target capacities defined in InstanceFleetConfig (p. 129). Capacity values represent performance characteristics such as vCPUs, memory, or I/O. If not specified, the default value is 1.

Type: Integer

Valid Range: Minimum value of 0.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

JobFlowDetail

A description of a cluster (job flow).

Contents

AmiVersion

Applies only to Amazon EMR AMI versions 3.x and 2.x. For Amazon EMR releases 4.0 and later, `ReleaseLabel` is used. To specify a custom AMI, use `CustomAmiID`.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

AutoScalingRole

An IAM role for automatic scaling policies. The default role is `EMR_AutoScaling_DefaultRole`. The IAM role provides a way for the automatic scaling feature to get the required permissions it needs to launch and terminate EC2 instances in an instance group.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

BootstrapActions

A list of the bootstrap actions run by the job flow.

Type: Array of BootstrapActionDetail (p. 98) objects

Required: No

ExecutionStatusDetail

Describes the execution status of the job flow.

Type: JobFlowExecutionStatusDetail (p. 161) object

Required: Yes

Instances

Describes the Amazon EC2 instances of the job flow.

Type: JobFlowInstancesDetail (p. 167) object

Required: Yes

JobFlowId

The job flow identifier.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: Yes

JobFlowRole

The IAM role that was specified when the job flow was launched. The EC2 instances of the job flow assume this role.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

LogUri

The location in Amazon S3 where log files for the job are stored.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

Name

The name of the job flow.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: Yes

ScaleDownBehavior

The way that individual Amazon EC2 instances terminate when an automatic scale-in activity occurs or an instance group is resized. `TERMINATE_AT_INSTANCE_HOUR` indicates that Amazon EMR terminates nodes at the instance-hour boundary, regardless of when the request to terminate the instance was submitted. This option is only available with Amazon EMR 5.1.0 and later and is the default for clusters created using that version. `TERMINATE_AT_TASK_COMPLETION` indicates that Amazon EMR blacklists and drains tasks from nodes before terminating the Amazon EC2 instances, regardless of the instance-hour boundary. With either behavior, Amazon EMR removes the least active nodes first and blocks instance termination if it could lead to HDFS corruption. `TERMINATE_AT_TASK_COMPLETION` available only in Amazon EMR version 4.1.0 and later, and is the default for versions of Amazon EMR earlier than 5.1.0.

Type: String

Valid Values: `TERMINATE_AT_INSTANCE_HOUR | TERMINATE_AT_TASK_COMPLETION`

Required: No

ServiceRole

The IAM role that will be assumed by the Amazon EMR service to access AWS resources on your behalf.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

Steps

A list of steps run by the job flow.

Type: Array of StepDetail (p. 189) objects

Required: No

SupportedProducts

A list of strings set by third party software when the job flow is launched. If you are not using third party software to manage the job flow this value is empty.

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

VisibleToAllUsers

Indicates whether the cluster is visible to all IAM users of the AWS account associated with the cluster. The default value, `true`, indicates that all IAM users in the AWS account can perform cluster actions if they have the proper IAM policy permissions. If this value is `false`, only the IAM user that created the cluster can perform actions. This value can be changed on a running cluster by using the SetVisibleToAllUsers (p. 85) action. You can override the default value of `true` when you create a cluster by using the `VisibleToAllUsers` parameter of the `RunJobFlow` action.

Type: Boolean

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

JobFlowExecutionStatusDetail

Describes the status of the cluster (job flow).

Contents

CreationDateTime

The creation date and time of the job flow.

Type: Timestamp

Required: Yes

EndDateTime

The completion date and time of the job flow.

Type: Timestamp

Required: No

LastStateChangeReason

Description of the job flow last changed state.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

ReadyDateTime

The date and time when the job flow was ready to start running bootstrap actions.

Type: Timestamp

Required: No

StartDateTime

The start date and time of the job flow.

Type: Timestamp

Required: No

State

The state of the job flow.

Type: String

Valid Values: `STARTING | BOOTSTRAPPING | RUNNING | WAITING | SHUTTING_DOWN | TERMINATED | COMPLETED | FAILED`

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

JobFlowInstancesConfig

A description of the Amazon EC2 instance on which the cluster (job flow) runs. A valid JobFlowInstancesConfig must contain either InstanceGroups or InstanceFleets, which is the recommended configuration. They cannot be used together. You may also have MasterInstanceType, SlaveInstanceType, and InstanceCount (all three must be present), but we don't recommend this configuration.

Contents

AdditionalMasterSecurityGroups

A list of additional Amazon EC2 security group IDs for the master node.

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

AdditionalSlaveSecurityGroups

A list of additional Amazon EC2 security group IDs for the core and task nodes.

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

Ec2KeyName

The name of the EC2 key pair that can be used to ssh to the master node as the user called "hadoop."

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

Ec2SubnetId

Applies to clusters that use the uniform instance group configuration. To launch the cluster in Amazon Virtual Private Cloud (Amazon VPC), set this parameter to the identifier of the Amazon VPC subnet where you want the cluster to launch. If you do not specify this value, the cluster launches in the normal Amazon Web Services cloud, outside of an Amazon VPC, if the account launching the cluster supports EC2 Classic networks in the region where the cluster launches.

Amazon VPC currently does not support cluster compute quadruple extra large (cc1.4xlarge) instances. Thus you cannot specify the cc1.4xlarge instance type for clusters launched in an Amazon VPC.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

Ec2SubnetIds

Applies to clusters that use the instance fleet configuration. When multiple EC2 subnet IDs are specified, Amazon EMR evaluates them and launches instances in the optimal subnet.

Note
The instance fleet configuration is available only in Amazon EMR versions 4.8.0 and later, excluding 5.0.x versions.

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

EmrManagedMasterSecurityGroup

The identifier of the Amazon EC2 security group for the master node.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

EmrManagedSlaveSecurityGroup

The identifier of the Amazon EC2 security group for the core and task nodes.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

HadoopVersion

Applies only to Amazon EMR release versions earlier than 4.0. The Hadoop version for the cluster. Valid inputs are "0.18" (deprecated), "0.20" (deprecated), "0.20.205" (deprecated), "1.0.3", "2.2.0", or "2.4.0". If you do not set this value, the default of 0.18 is used, unless the `AmiVersion` parameter is set in the RunJobFlow call, in which case the default version of Hadoop for that AMI version is used.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

InstanceCount

The number of EC2 instances in the cluster.

Type: Integer

Required: No

InstanceFleets

> **Note**
> The instance fleet configuration is available only in Amazon EMR versions 4.8.0 and later, excluding 5.0.x versions.

Describes the EC2 instances and instance configurations for clusters that use the instance fleet configuration.

Type: Array of InstanceFleetConfig (p. 129) objects

Required: No

InstanceGroups

Configuration for the instance groups in a cluster.

Type: Array of InstanceGroupConfig (p. 140) objects

Required: No

KeepJobFlowAliveWhenNoSteps

Specifies whether the cluster should remain available after completing all steps.

Type: Boolean

Required: No

MasterInstanceType

The EC2 instance type of the master node.

Type: String

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

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

Required: No

Placement

The Availability Zone in which the cluster runs.

Type: PlacementType (p. 174) object

Required: No

ServiceAccessSecurityGroup

The identifier of the Amazon EC2 security group for the Amazon EMR service to access clusters in VPC private subnets.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

SlaveInstanceType

The EC2 instance type of the core and task nodes.

Type: String

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

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

Required: No

TerminationProtected

Specifies whether to lock the cluster to prevent the Amazon EC2 instances from being terminated by API call, user intervention, or in the event of a job-flow error.

Type: Boolean

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

JobFlowInstancesDetail

Specify the type of Amazon EC2 instances that the cluster (job flow) runs on.

Contents

Ec2KeyName

The name of an Amazon EC2 key pair that can be used to ssh to the master node.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

Ec2SubnetId

For clusters launched within Amazon Virtual Private Cloud, this is the identifier of the subnet where the cluster was launched.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

HadoopVersion

The Hadoop version for the cluster.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

InstanceCount

The number of Amazon EC2 instances in the cluster. If the value is 1, the same instance serves as both the master and core and task node. If the value is greater than 1, one instance is the master node and all others are core and task nodes.

Type: Integer

Required: Yes

InstanceGroups

Details about the instance groups in a cluster.

Type: Array of InstanceGroupDetail (p. 142) objects

Required: No

KeepJobFlowAliveWhenNoSteps

Specifies whether the cluster should remain available after completing all steps.

Type: Boolean

Required: No

MasterInstanceId

The Amazon EC2 instance identifier of the master node.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

MasterInstanceType

The Amazon EC2 master node instance type.

Type: String

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

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

Required: Yes

MasterPublicDnsName

The DNS name of the master node. If the cluster is on a private subnet, this is the private DNS name. On a public subnet, this is the public DNS name.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

NormalizedInstanceHours

An approximation of the cost of the cluster, represented in m1.small/hours. This value is incremented one time for every hour that an m1.small runs. Larger instances are weighted more, so an Amazon EC2 instance that is roughly four times more expensive would result in the normalized instance hours being incremented by four. This result is only an approximation and does not reflect the actual billing rate.

Type: Integer

Required: No

Placement

The Amazon EC2 Availability Zone for the cluster.

Type: PlacementType (p. 174) object

Required: No

SlaveInstanceType

The Amazon EC2 core and task node instance type.

Type: String

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

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

Required: Yes

TerminationProtected

Specifies whether the Amazon EC2 instances in the cluster are protected from termination by API calls, user intervention, or in the event of a job-flow error.

Type: Boolean

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

KerberosAttributes

Attributes for Kerberos configuration when Kerberos authentication is enabled using a security configuration. For more information see Use Kerberos Authentication in the *EMR Management Guide*.

Contents

ADDomainJoinPassword

The Active Directory password for `ADDomainJoinUser`.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

ADDomainJoinUser

Required only when establishing a cross-realm trust with an Active Directory domain. A user with sufficient privileges to join resources to the domain.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

CrossRealmTrustPrincipalPassword

Required only when establishing a cross-realm trust with a KDC in a different realm. The cross-realm principal password, which must be identical across realms.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: No

KdcAdminPassword

The password used within the cluster for the kadmin service on the cluster-dedicated KDC, which maintains Kerberos principals, password policies, and keytabs for the cluster.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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

Required: Yes

Realm

The name of the Kerberos realm to which all nodes in a cluster belong. For example, `EC2.INTERNAL`.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

KeyValue

A key value pair.

Contents

Key

The unique identifier of a key value pair.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

Value

The value part of the identified key.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

MetricDimension

A CloudWatch dimension, which is specified using a `Key` (known as a `Name` in CloudWatch), `Value` pair. By default, Amazon EMR uses one dimension whose `Key` is `JobFlowID` and `Value` is a variable representing the cluster ID, which is `${emr.clusterId}`. This enables the rule to bootstrap when the cluster ID becomes available.

Contents

Key

The dimension name.

Type: String

Required: No

Value

The dimension value.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

PlacementType

The Amazon EC2 Availability Zone configuration of the cluster (job flow).

Contents

AvailabilityZone

The Amazon EC2 Availability Zone for the cluster. `AvailabilityZone` is used for uniform instance groups, while `AvailabilityZones` (plural) is used for instance fleets.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

AvailabilityZones

When multiple Availability Zones are specified, Amazon EMR evaluates them and launches instances in the optimal Availability Zone. `AvailabilityZones` is used for instance fleets, while `AvailabilityZone` (singular) is used for uniform instance groups.

Note
The instance fleet configuration is available only in Amazon EMR versions 4.8.0 and later, excluding 5.0.x versions.

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 256.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

ScalingAction

The type of adjustment the automatic scaling activity makes when triggered, and the periodicity of the adjustment.

Contents

Market

Not available for instance groups. Instance groups use the market type specified for the group.

Type: String

Valid Values: `ON_DEMAND | SPOT`

Required: No

SimpleScalingPolicyConfiguration

The type of adjustment the automatic scaling activity makes when triggered, and the periodicity of the adjustment.

Type: SimpleScalingPolicyConfiguration (p. 182) object

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

ScalingConstraints

The upper and lower EC2 instance limits for an automatic scaling policy. Automatic scaling activities triggered by automatic scaling rules will not cause an instance group to grow above or below these limits.

Contents

MaxCapacity

The upper boundary of EC2 instances in an instance group beyond which scaling activities are not allowed to grow. Scale-out activities will not add instances beyond this boundary.

Type: Integer

Required: Yes

MinCapacity

The lower boundary of EC2 instances in an instance group below which scaling activities are not allowed to shrink. Scale-in activities will not terminate instances below this boundary.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

ScalingRule

A scale-in or scale-out rule that defines scaling activity, including the CloudWatch metric alarm that triggers activity, how EC2 instances are added or removed, and the periodicity of adjustments. The automatic scaling policy for an instance group can comprise one or more automatic scaling rules.

Contents

Action

The conditions that trigger an automatic scaling activity.

Type: ScalingAction (p. 175) object

Required: Yes

Description

A friendly, more verbose description of the automatic scaling rule.

Type: String

Required: No

Name

The name used to identify an automatic scaling rule. Rule names must be unique within a scaling policy.

Type: String

Required: Yes

Trigger

The CloudWatch alarm definition that determines when automatic scaling activity is triggered.

Type: ScalingTrigger (p. 178) object

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

ScalingTrigger

The conditions that trigger an automatic scaling activity.

Contents

CloudWatchAlarmDefinition

The definition of a CloudWatch metric alarm. When the defined alarm conditions are met along with other trigger parameters, scaling activity begins.

Type: CloudWatchAlarmDefinition (p. 100) object

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

ScriptBootstrapActionConfig

Configuration of the script to run during a bootstrap action.

Contents

Args

A list of command line arguments to pass to the bootstrap action script.

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

Path

Location of the script to run during a bootstrap action. Can be either a location in Amazon S3 or on a local file system.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

SecurityConfigurationSummary

The creation date and time, and name, of a security configuration.

Contents

CreationDateTime

The date and time the security configuration was created.

Type: Timestamp

Required: No

Name

The name of the security configuration.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

ShrinkPolicy

Policy for customizing shrink operations. Allows configuration of decommissioning timeout and targeted instance shrinking.

Contents

DecommissionTimeout

The desired timeout for decommissioning an instance. Overrides the default YARN decommissioning timeout.

Type: Integer

Required: No

InstanceResizePolicy

Custom policy for requesting termination protection or termination of specific instances when shrinking an instance group.

Type: InstanceResizePolicy (p. 150) 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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

SimpleScalingPolicyConfiguration

An automatic scaling configuration, which describes how the policy adds or removes instances, the cooldown period, and the number of EC2 instances that will be added each time the CloudWatch metric alarm condition is satisfied.

Contents

AdjustmentType

The way in which EC2 instances are added (if `ScalingAdjustment` is a positive number) or terminated (if `ScalingAdjustment` is a negative number) each time the scaling activity is triggered. `CHANGE_IN_CAPACITY` is the default. `CHANGE_IN_CAPACITY` indicates that the EC2 instance count increments or decrements by `ScalingAdjustment`, which should be expressed as an integer. `PERCENT_CHANGE_IN_CAPACITY` indicates the instance count increments or decrements by the percentage specified by `ScalingAdjustment`, which should be expressed as an integer. For example, 20 indicates an increase in 20% increments of cluster capacity. `EXACT_CAPACITY` indicates the scaling activity results in an instance group with the number of EC2 instances specified by `ScalingAdjustment`, which should be expressed as a positive integer.

Type: String

Valid Values: `CHANGE_IN_CAPACITY | PERCENT_CHANGE_IN_CAPACITY | EXACT_CAPACITY`

Required: No

CoolDown

The amount of time, in seconds, after a scaling activity completes before any further trigger-related scaling activities can start. The default value is 0.

Type: Integer

Required: No

ScalingAdjustment

The amount by which to scale in or scale out, based on the specified `AdjustmentType`. A positive value adds to the instance group's EC2 instance count while a negative number removes instances. If `AdjustmentType` is set to `EXACT_CAPACITY`, the number should only be a positive integer. If `AdjustmentType` is set to `PERCENT_CHANGE_IN_CAPACITY`, the value should express the percentage as an integer. For example, -20 indicates a decrease in 20% increments of cluster capacity.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

SpotProvisioningSpecification

The launch specification for Spot instances in the instance fleet, which determines the defined duration and provisioning timeout behavior.

Note
The instance fleet configuration is available only in Amazon EMR versions 4.8.0 and later, excluding 5.0.x versions.

Contents

BlockDurationMinutes

The defined duration for Spot instances (also known as Spot blocks) in minutes. When specified, the Spot instance does not terminate before the defined duration expires, and defined duration pricing for Spot instances applies. Valid values are 60, 120, 180, 240, 300, or 360. The duration period starts as soon as a Spot instance receives its instance ID. At the end of the duration, Amazon EC2 marks the Spot instance for termination and provides a Spot instance termination notice, which gives the instance a two-minute warning before it terminates.

Type: Integer

Valid Range: Minimum value of 0.

Required: No

TimeoutAction

The action to take when `TargetSpotCapacity` has not been fulfilled when the `TimeoutDurationMinutes` has expired; that is, when all Spot instances could not be provisioned within the Spot provisioning timeout. Valid values are `TERMINATE_CLUSTER` and `SWITCH_TO_ON_DEMAND`. SWITCH_TO_ON_DEMAND specifies that if no Spot instances are available, On-Demand Instances should be provisioned to fulfill any remaining Spot capacity.

Type: String

Valid Values: `SWITCH_TO_ON_DEMAND | TERMINATE_CLUSTER`

Required: Yes

TimeoutDurationMinutes

The spot provisioning timeout period in minutes. If Spot instances are not provisioned within this time period, the `TimeOutAction` is taken. Minimum value is 5 and maximum value is 1440. The timeout applies only during initial provisioning, when the cluster is first created.

Type: Integer

Valid Range: Minimum value of 0.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

Step

This represents a step in a cluster.

Contents

ActionOnFailure

The action to take when the cluster step fails. Possible values are TERMINATE_CLUSTER, CANCEL_AND_WAIT, and CONTINUE. TERMINATE_JOB_FLOW is provided for backward compatibility. We recommend using TERMINATE_CLUSTER instead.

Type: String

Valid Values: `TERMINATE_JOB_FLOW | TERMINATE_CLUSTER | CANCEL_AND_WAIT | CONTINUE`

Required: No

Config

The Hadoop job configuration of the cluster step.

Type: HadoopStepConfig (p. 123) object

Required: No

Id

The identifier of the cluster step.

Type: String

Required: No

Name

The name of the cluster step.

Type: String

Required: No

Status

The current execution status details of the cluster step.

Type: StepStatus (p. 193) 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 Go - Pilot
- AWS SDK for Java

- AWS SDK for Ruby V2

StepConfig

Specification of a cluster (job flow) step.

Contents

ActionOnFailure

The action to take when the cluster step fails. Possible values are TERMINATE_CLUSTER, CANCEL_AND_WAIT, and CONTINUE. TERMINATE_JOB_FLOW is provided for backward compatibility. We recommend using TERMINATE_CLUSTER instead.

Type: String

Valid Values: `TERMINATE_JOB_FLOW | TERMINATE_CLUSTER | CANCEL_AND_WAIT | CONTINUE`

Required: No

HadoopJarStep

The JAR file used for the step.

Type: HadoopJarStepConfig (p. 121) object

Required: Yes

Name

The name of the step.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

StepDetail

Combines the execution state and configuration of a step.

Contents

ExecutionStatusDetail

The description of the step status.

Type: StepExecutionStatusDetail (p. 190) object

Required: Yes

StepConfig

The step configuration.

Type: StepConfig (p. 188) object

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

StepExecutionStatusDetail

The execution state of a step.

Contents

CreationDateTime

The creation date and time of the step.

Type: Timestamp

Required: Yes

EndDateTime

The completion date and time of the step.

Type: Timestamp

Required: No

LastStateChangeReason

A description of the step's current state.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

StartDateTime

The start date and time of the step.

Type: Timestamp

Required: No

State

The state of the step.

Type: String

Valid Values: `PENDING | RUNNING | CONTINUE | COMPLETED | CANCELLED | FAILED | INTERRUPTED`

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 Go - Pilot

- AWS SDK for Java
- AWS SDK for Ruby V2

StepStateChangeReason

The details of the step state change reason.

Contents

Code

The programmable code for the state change reason. Note: Currently, the service provides no code for the state change.

Type: String

Valid Values: `NONE`

Required: No

Message

The descriptive message for the state change reason.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

StepStatus

The execution status details of the cluster step.

Contents

FailureDetails

The details for the step failure including reason, message, and log file path where the root cause was identified.

Type: FailureDetails (p. 120) object

Required: No

State

The execution state of the cluster step.

Type: String

Valid Values: `PENDING | CANCEL_PENDING | RUNNING | COMPLETED | CANCELLED | FAILED | INTERRUPTED`

Required: No

StateChangeReason

The reason for the step execution status change.

Type: StepStateChangeReason (p. 192) object

Required: No

Timeline

The timeline of the cluster step status over time.

Type: StepTimeline (p. 196) 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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

StepSummary

The summary of the cluster step.

Contents

ActionOnFailure

The action to take when the cluster step fails. Possible values are TERMINATE_CLUSTER, CANCEL_AND_WAIT, and CONTINUE. TERMINATE_JOB_FLOW is available for backward compatibility. We recommend using TERMINATE_CLUSTER instead.

Type: String

Valid Values: `TERMINATE_JOB_FLOW | TERMINATE_CLUSTER | CANCEL_AND_WAIT | CONTINUE`

Required: No

Config

The Hadoop job configuration of the cluster step.

Type: HadoopStepConfig (p. 123) object

Required: No

Id

The identifier of the cluster step.

Type: String

Required: No

Name

The name of the cluster step.

Type: String

Required: No

Status

The current execution status details of the cluster step.

Type: StepStatus (p. 193) 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 Go - Pilot
- AWS SDK for Java

- AWS SDK for Ruby V2

StepTimeline

The timeline of the cluster step lifecycle.

Contents

CreationDateTime

The date and time when the cluster step was created.

Type: Timestamp

Required: No

EndDateTime

The date and time when the cluster step execution completed or failed.

Type: Timestamp

Required: No

StartDateTime

The date and time when the cluster step execution started.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

SupportedProductConfig

The list of supported product configurations which allow user-supplied arguments. EMR accepts these arguments and forwards them to the corresponding installation script as bootstrap action arguments.

Contents

Args

The list of user-supplied arguments.

Type: Array of strings

Length Constraints: Minimum length of 0. Maximum length of 10280.

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

Required: No

Name

The name of the product configuration.

Type: String

Length Constraints: Minimum length of 0. Maximum length of 256.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

Tag

A key/value pair containing user-defined metadata that you can associate with an Amazon EMR resource. Tags make it easier to associate clusters in various ways, such as grouping clusters to track your Amazon EMR resource allocation costs. For more information, see Tag Clusters.

Contents

Key

A user-defined key, which is the minimum required information for a valid tag. For more information, see Tag .

Type: String

Required: No

Value

A user-defined value, which is optional in a tag. For more information, see Tag Clusters.

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 Go - Pilot
- AWS SDK for Java
- AWS SDK for Ruby V2

VolumeSpecification

EBS volume specifications such as volume type, IOPS, and size (GiB) that will be requested for the EBS volume attached to an EC2 instance in the cluster.

Contents

Iops

The number of I/O operations per second (IOPS) that the volume supports.

Type: Integer

Required: No

SizeInGB

The volume size, in gibibytes (GiB). This can be a number from 1 - 1024. If the volume type is EBS-optimized, the minimum value is 10.

Type: Integer

Required: Yes

VolumeType

The volume type. Volume types supported are gp2, io1, standard.

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 Go - Pilot
- 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'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 (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