# Amazon Elastic MapReduce

# API Reference API Version 2009-03-31



### **Amazon Elastic MapReduce: API Reference**

Copyright © 2013 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

The following are trademarks of Amazon Web Services, Inc.: Amazon, Amazon Web Services Design, AWS, Amazon CloudFront, Cloudfront, Amazon DevPay, DynamoDB, ElastiCache, Amazon EC2, Amazon Elastic Compute Cloud, Amazon Glacier, Kindle, Kindle Fire, AWS Marketplace Design, Mechanical Turk, Amazon Redshift, Amazon Route 53, Amazon S3, Amazon VPC. In addition, Amazon.com graphics, logos, page headers, button icons, scripts, and service names are trademarks, or trade dress of Amazon in the U.S. and/or other countries. Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon.

All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

### Amazon Elastic MapReduce API Reference

Welcome	
Actions	
AddInstanceGroups	
AddJobFlowSteps	
AddTags	
DescribeCluster	10
DescribeJobFlows	
DescribeStep	15
ListBootstrapActions	16
ListClusters	17
ListInstanceGroups	19
ListInstances	
ListSteps	22
ModifyInstanceGroups	24
RemoveTags	26
RunJobFlow	28
SetTerminationProtection	33
SetVisibleToAllUsers	35
TerminateJobFlows	37
Data Types	39
AddInstanceGroupsResult	40
AddJobFlowStepsResult	41
AddTagsResult	41
Application	41
BootstrapActionConfig	42
BootstrapActionDetail	43
Cluster	
ClusterStateChangeReason	
ClusterStatus	
ClusterSummary	46
ClusterTimeline	
Command	
DescribeClusterResult	
DescribeJobFlowsResult	47
DescribeStepResult	48
Ec2InstanceAttributes	
HadoopJarStepConfig	
HadoopStepConfig	50
Instance	
InstanceGroup	51
InstanceGroupConfig	
InstanceGroupDetail	
InstanceGroupModifyConfig	
InstanceGroupStateChangeReason	
InstanceGroupStatus	
InstanceGroupTimeline	
InstanceStateChangeReason	
InstanceStatus	
InstanceTimeline	
JobFlowDetail	
JobFlowExecutionStatusDetail	
JobFlowInstancesConfig	
JobFlowInstancesDetail	
KeyValue	
ListBootstrapActionsResult	
ListClustersResult	
ListInstanceGroupsResult	
ListInstancesResult	

### Amazon Elastic MapReduce API Reference

ListStepsResult	68
PlacementType	68
RemoveTagsResult	68
RunJobFlowResult	69
ScriptBootstrapActionConfig	69
Step	69
StepConfigStepDetail	70
StepDetail	71
StepExecutionStatusDetail	71
StepStateChangeReason	72
StepStatusStepSummary	73
StepSummary	73
StepTimeline	74
SupportedProductConfig	74
Tag	75
Common Parameters	76
Common Errors	78

# Welcome

This is the *Amazon Elastic MapReduce API Reference*. This guide provides descriptions and samples of the Amazon Elastic MapReduce APIs.

Amazon Elastic MapReduce (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 updated on December 5, 2013.

# **Actions**

### The following actions are supported:

- AddInstanceGroups (p. 3)
- AddJobFlowSteps (p. 5)
- AddTags (p. 8)
- DescribeCluster (p. 10)
- DescribeJobFlows (p. 12)
- DescribeStep (p. 15)
- ListBootstrapActions (p. 16)
- ListClusters (p. 17)
- ListInstanceGroups (p. 19)
- ListInstances (p. 20)
- ListSteps (p. 22)
- ModifyInstanceGroups (p. 24)
- RemoveTags (p. 26)
- RunJobFlow (p. 28)
- SetTerminationProtection (p. 33)
- SetVisibleToAllUsers (p. 35)
- TerminateJobFlows (p. 37)

# AddInstanceGroups

# **Description**

AddInstanceGroups adds an instance group to a running cluster.

### **Request Parameters**

For information about the common parameters that all actions use, see Common Parameters (p. 76).

### InstanceGroups.member.N

Instance Groups to add.

Type: InstanceGroupConfig (p. 52) list

Required: Yes

#### **JobFlowId**

Job flow in which to add the instance groups.

Type: String

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

Required: Yes

### **Response Elements**

The following elements are returned in a structure named AddInstanceGroupsResult.

#### InstanceGroupIds

Instance group IDs of the newly created instance groups.

Type: String list

#### **JobFlowId**

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

Type: String

### **Errors**

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

#### InternalServerError

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

HTTP Status Code: 500

# **Examples**

### **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, Signa
ture=4c5e7eb762ea45f292a5cd1a1cc56ed60009e19a9dba3d6e5e4e67e96d43af11
Accept: */*
    "JobFlowId": "j-3U7TSX5GZFD8Y",
    "InstanceGroups": [{
        "Name": "Task Instance Group",
        "InstanceRole": "TASK",
        "InstanceCount": 2,
        "InstanceType": "m1.small",
        "Market": "ON_DEMAND"
    } ]
}
```

```
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"
}
```

# AddJobFlowSteps

### **Description**

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

If your job flow 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, go to Add More than 256 Steps to a Job Flow in the Amazon Elastic MapReduce Developer's Guide.

A step specifies the location of a JAR file stored either on the master node of the job flow 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.

Elastic MapReduce 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 job flow that is in one of the following states: STARTING, BOOTSTRAPPING, RUNNING, or WAITING.

### **Request Parameters**

For information about the common parameters that all actions use, see Common Parameters (p. 76).

### **JobFlowId**

A string that uniquely identifies the job flow. This identifier is returned by RunJobFlow (p. 28) and can also be obtained from DescribeJobFlows (p. 12).

Type: String

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

Required: Yes Steps.member.N

A list of StepConfig (p. 70) to be executed by the job flow.

Type: StepConfig (p. 70) list

Required: Yes

# **Response Elements**

The following element is returned in a structure named AddJobFlowStepsResult.

#### **StepIds**

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

Type: String list

### **Errors**

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

#### InternalServerError

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

HTTP Status Code: 500

### **Examples**

### 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, Signa
ture=2a2393390760ae85eb74ee3a539e1d758bfdd8815a1a6d6f14d4a2fbcfdcd5b7
Accept: */*
{
    "JobFlowId": "j-3TS00IY04NFN",
    "Steps": [{
        "Name": "Example Jar Step",
        "ActionOnFailure": "CANCEL_AND_WAIT",
        "HadoopJarStep": {
            "Jar": "s3n:\\/\/elasticmapreduce\//samples\\/cloudburst\\/cloud
burst.jar",
            "Args": [
                "s3n:\\/\/elasticmapreduce\\/samples\\/cloudburst\\/in
put\\/s_suis.br",
                "s3n:\\/\/elasticmapreduce\\/samples\\/cloudburst\\/in
put\\/100k.br",
                "s3n:\\/\/examples-bucket\\/cloudburst\\/output",
                "36",
                "3",
                "0",
                "1",
                "240",
                "48",
                "24",
                "24",
                "128",
                "16"
            ]
```

```
}1
}
```

```
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
```

# **AddTags**

# **Description**

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 Tagging Amazon EMR Resources.

### **Request Parameters**

For information about the common parameters that all actions use, see Common Parameters (p. 76).

#### Resourceld

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

Type: String Required: No

### Tags.member.N

A list of tags to associate with a cluster and propagate to Amazon 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: Tag (p. 75) list Required: No

### **Errors**

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

### InternalServer

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

HTTP Status Code: 500

#### InvalidRequest

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

HTTP Status Code: 400

### **Examples**

### 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"
}]
```

```
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
{
}
```

# **DescribeCluster**

# **Description**

Provides cluster-level details including status, hardware and software configuration, VPC settings, and so on. For information about the cluster steps, see ListSteps (p. 22).

The following example lists all tags associated with given cluster id.

# **Request Parameters**

For information about the common parameters that all actions use, see Common Parameters (p. 76).

#### ClusterId

The identifier of the cluster to describe.

Type: String Required: No

# **Response Elements**

The following element is returned in a structure named DescribeClusterResult.

#### Cluster

This output contains the details for the requested cluster.

Type: Cluster (p. 43)

### **Errors**

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

#### InternalServer

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

HTTP Status Code: 500

### InvalidRequest

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

HTTP Status Code: 400

### **Examples**

### 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"
    }]
}
POST / HTTP/1.1
Content-Type: application/x-amz-json-1.1
X-Amz-Target: ElasticMapReduce.DescribeCluster
AUTHPARAMS
{
    "ClusterId": "j-12XY3456BB"
}
```

```
HTTP/1.1 200 OK
x-amzn-RequestId: 9da5a349-ed9e-11e2-90db-69a5154aeb8d
{\tt Content-Type: application/x-amz-json-1.1}
Content-Length: 71
Date: Mon, 15 Jul 2013 22:33:47 GMT
}
 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
  "Cluster": {
    "Id": "j-12XY3456BB",
    "Name": "Test",
    "Status": {
      "State": "RUNNING",
      "Timeline": {
        "CreationDateTime": "12345566667",
        "ReadyDateTime": "12345566667",
        "EndDateTime": "12345566667"
      }
    },
    "Tags": [{
      "Key": "stack",
      "Value": "Production"
    } ]
  }
}
```

# **DescribeJobFlows**

# **Description**

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 Elastic MapReduce can return a maximum of 512 job flow descriptions.

### **Request Parameters**

For information about the common parameters that all actions use, see Common Parameters (p. 76).

#### CreatedAfter

Return only job flows created after this date and time.

Type: DateTime

Required: No

#### CreatedBefore

Return only job flows created before this date and time.

Type: DateTime

Required: No

#### JobFlowIds.member.N

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

Type: String list

Required: No

### JobFlowStates.member.N

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

Type: String list

Required: No

# **Response Elements**

The following element is returned in a structure named DescribeJobFlowsResult.

#### **JobFlows**

A list of job flows matching the parameters supplied.

Type: JobFlowDetail (p. 59) list

### **Errors**

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

#### InternalServerError

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

HTTP Status Code: 500

### **Examples**

### 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, Signa
ture=29F98a6f44e05ad54fe1e8b3d1a7101ab08dc3ad348995f89c533693cee2bb3b
Accept: */*
{
    "JobFlowIds": ["j-ZKIY4CKQRX72"],
    "DescriptionType": "EXTENDED"
```

```
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",
                    "s3://elasticmapreduce/samples/wordcount/wordSplitter.py",
                    "-reducer",
                    "aggregate"
                ],
                "Jar": "/home/hadoop/contrib/streaming/hadoop-streaming.jar",
                "Properties": []
            "Name": "Example Streaming Step"
    }],
    "SupportedProducts": [],
    "VisibleToAllUsers": false
}]}
```

# **DescribeStep**

# **Description**

Provides more detail about the cluster step.

### **Request Parameters**

For information about the common parameters that all actions use, see Common Parameters (p. 76).

#### ClusterId

The identifier of the cluster with steps to describe.

Type: String

Required: No

#### StepId

The identifier of the step to describe.

Type: String

Required: No

# **Response Elements**

The following element is returned in a structure named DescribeStepResult.

#### Step

The step details for the requested step identifier.

Type: Step (p. 69)

### **Errors**

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

#### InternalServer

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

HTTP Status Code: 500

### InvalidRequest

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

HTTP Status Code: 400

# ListBootstrapActions

# **Description**

Provides information about the bootstrap actions associated with a cluster.

### **Request Parameters**

For information about the common parameters that all actions use, see Common Parameters (p. 76).

#### ClusterId

The cluster identifier for the bootstrap actions to list.

Type: String Required: No

#### Marker

The pagination token is a random string indicating whether there are more results to fetch. Provide the pagination token from earlier API calls to retrieve the next page of results. When the value is null, all results have been returned.

Type: String
Required: No

### **Response Elements**

The following elements are returned in a structure named ListBootstrapActionsResult.

#### **BootstrapActions**

The bootstrap actions associated with the cluster.

Type: Command (p. 47) list

### Marker

The pagination token is a random string indicating whether there are more results to fetch. Use the pagination token in later API calls to retrieve the next page of results. When the value is null, all results have been returned.

Type: String

### **Errors**

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

#### InternalServer

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

HTTP Status Code: 500

#### InvalidRequest

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

HTTP Status Code: 400

### ListClusters

# **Description**

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.

The following example lists all clusters which have a tag stack=Production:

### **Request Parameters**

For information about the common parameters that all actions use, see Common Parameters (p. 76).

#### ClusterStates.member.N

The cluster state filters to apply when listing clusters.

Type: String list

Required: No

#### CreatedAfter

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

Type: DateTime

Required: No

#### CreatedBefore

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

Type: DateTime

Required: No

#### Marker

The pagination token is a random string indicating whether there are more results to fetch. Provide the pagination token from earlier API calls to retrieve the next page of results. When the value is null, all results have been returned.

Type: String

Required: No

### **Response Elements**

The following elements are returned in a structure named ListClustersResult.

#### Clusters

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

Type: ClusterSummary (p. 46) list

#### Marker

The pagination token is a random string indicating whether there are more results to fetch. Use the pagination token in later API calls to retrieve the next page of results. When the value is null, all results have been returned.

Type: String

### **Errors**

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

#### InternalServer

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

HTTP Status Code: 500

#### InvalidRequest

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

HTTP Status Code: 400

# **Examples**

### **Sample Request**

```
POST / HTTP/1.1
Content-Type: application/x-amz-json-1.1
X-Amz-Target: ElasticMapReduce.ListClusters
AUTHPARAMS
{
    "Tags": [{
        "Key": "stack",
        "Value": "Production"
    }]
}
```

```
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
  "Clusters": [{
  "Id": "j-12XY3456BB",
  "Name": "Test",
      "Status": {
        "State": "RUNNING",
        "Timeline": {
         "CreationDateTime": "12345566667",
         "ReadyDateTime": "12345566667",
         "EndDateTime": "12345566667"
  } ]
}
```

# ListInstanceGroups

# **Description**

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

### **Request Parameters**

For information about the common parameters that all actions use, see Common Parameters (p. 76).

#### ClusterId

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

Type: String Required: No

#### Marker

The pagination token is a random string indicating whether there are more results to fetch. Provide the pagination token from earlier API calls to retrieve the next page of results. When the value is null, all results have been returned.

Type: String
Required: No

### **Response Elements**

The following elements are returned in a structure named ListInstanceGroupsResult.

#### InstanceGroups

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

Type: InstanceGroup (p. 51) list

### Marker

The pagination token is a random string indicating whether there are more results to fetch. Use the pagination token in later API calls to retrieve the next page of results. When the value is null, all results have been returned.

Type: String

### **Errors**

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

#### InternalServer

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

HTTP Status Code: 500

### InvalidRequest

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

HTTP Status Code: 400

### ListInstances

# **Description**

Provides information about the cluster instances that Amazon EMR provisions on behalf of a user when it creates the cluster. For example, this operation indicates when the EC2 instances reach the Ready state, when instances become available to Amazon EMR to use for jobs, and the IP addresses for cluster instances, etc.

### **Request Parameters**

For information about the common parameters that all actions use, see Common Parameters (p. 76).

#### ClusterId

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

Type: String Required: No

### InstanceGroupId

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

Type: String
Required: No

#### InstanceGroupTypes.member.N

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

Type: String list Required: No

#### Marker

The pagination token is a random string indicating whether there are more results to fetch. Provide the pagination token from earlier API calls to retrieve the next page of results. When the value is null, all results have been returned.

Type: String Required: No

### **Response Elements**

The following elements are returned in a structure named ListInstancesResult.

#### **Instances**

The list of instances for the cluster and given filters.

Type: Instance (p. 50) list

#### Marker

The pagination token is a random string indicating whether there are more results to fetch. Use the pagination token in later API calls to retrieve the next page of results. When the value is null, all results have been returned.

Type: String

### Amazon Elastic MapReduce API Reference Errors

# **Errors**

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

#### InternalServer

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

HTTP Status Code: 500

### InvalidRequest

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

HTTP Status Code: 400

# **ListSteps**

# **Description**

Provides a list of steps for the cluster.

### **Request Parameters**

For information about the common parameters that all actions use, see Common Parameters (p. 76).

#### ClusterId

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

Type: String Required: No

#### Marker

The pagination token is a random string indicating whether there are more results to fetch. Provide the pagination token from earlier API calls to retrieve the next page of results. When the value is null, all results have been returned.

Type: String
Required: No

### StepStates.member.N

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

Type: String list Required: No

# **Response Elements**

The following elements are returned in a structure named ListStepsResult.

#### Marker

The pagination token is a random string indicating whether there are more results to fetch. Use the pagination token in later API calls to retrieve the next page of results. When the value is null, all results have been returned.

Type: String

#### **Steps**

The filtered list of steps for the cluster.

Type: StepSummary (p. 73) list

### **Errors**

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

#### InternalServer

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

### Amazon Elastic MapReduce API Reference Errors

HTTP Status Code: 500

### InvalidRequest

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

HTTP Status Code: 400

# ModifyInstanceGroups

# **Description**

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 Parameters**

For information about the common parameters that all actions use, see Common Parameters (p. 76).

#### InstanceGroups.member.N

Instance groups to change.

Type: InstanceGroupModifyConfig (p. 55) list

Required: No

### **Errors**

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

#### InternalServerError

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

HTTP Status Code: 500

# **Examples**

### 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:
bblaf3d0c6c6ala09f2lccd7f04a0e2e6c9ce5b5810b0f6777560fe4f81bda8c
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, Signa
ture=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

# RemoveTags

# **Description**

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 Tagging Amazon EMR Resources.

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

### **Request Parameters**

For information about the common parameters that all actions use, see Common Parameters (p. 76).

#### Resourceld

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

Type: String Required: No

### TagKeys.member.N

A list of tag keys to remove from a resource.

Type: String list Required: No

### **Errors**

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

#### InternalServer

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

HTTP Status Code: 500

### InvalidRequest

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

HTTP Status Code: 400

### **Examples**

### 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"
}]

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"
    }]
}
```

```
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
{
}
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
{
}
```

### RunJobFlow

# **Description**

RunJobFlow creates and starts running a new job flow. The job flow will run the steps specified. Once the job flow completes, the cluster is stopped 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. 62) KeepJobFlowAliveWhenNoSteps parameter is set to TRUE, the job flow will transition to the WAITING state rather than shutting down once the steps have completed.

For additional protection, you can set the JobFlowInstancesConfig (p. 62) TerminationProtected parameter to TRUE to lock the job flow 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 job flow 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, go to Add More than 256 Steps to a Job Flow in the Amazon Elastic MapReduce Developer's Guide.

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

### **Request Parameters**

For information about the common parameters that all actions use, see Common Parameters (p. 76).

#### AdditionalInfo

A JSON string for selecting additional features.

Type: String

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

Required: No

#### AmiVersion

The version of the Amazon Machine Image (AMI) to use when launching Amazon EC2 instances in the job flow. The following values are valid:

- "latest" (uses the latest AMI)
- The version number of the AMI to use, for example, "2.0"

If the AMI supports multiple versions of Hadoop (for example, AMI 1.0 supports both Hadoop 0.18 and 0.20) you can use the JobFlowInstancesConfig (p. 62) HadoopVersion parameter to modify the version of Hadoop from the defaults shown above.

For details about the AMI versions currently supported by Amazon Elastic MapReduce, go to AMI Versions Supported in Elastic MapReduce in the Amazon Elastic MapReduce Developer's Guide.

Type: String

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

Required: No

### BootstrapActions.member.N

A list of bootstrap actions that will be run before Hadoop is started on the cluster nodes.

### Amazon Elastic MapReduce API Reference Request Parameters

Type: BootstrapActionConfig (p. 42) list

Required: No

#### **Instances**

A specification of the number and type of Amazon EC2 instances on which to run the job flow.

Type: JobFlowInstancesConfig (p. 62)

Required: Yes

### **JobFlowRole**

An IAM role for the job flow. The EC2 instances of the job flow assume this role. The default role is EMRJobflowDefault. In order to use the default role, you must have already created it using the CLI.

Type: String

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

Required: No

#### LogUri

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.

Required: No

#### Name

The name of the job flow.

Type: String

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

Required: Yes

#### NewSupportedProducts.member.N

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. Currently supported values are:

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

Type: SupportedProductConfig (p. 74) list

Required: No Steps.member.N

A list of steps to be executed by the job flow.

Type: StepConfig (p. 70) list

Required: No

### SupportedProducts.member.N

A list of strings that indicates third-party software to use with the job flow. For more information, go to Use Third Party Applications with Amazon EMR. Currently supported values are:

### Amazon Elastic MapReduce API Reference Response Elements

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

Type: String list Required: No

### Tags.member.N

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

Type: Tag (p. 75) list

Required: No

### VisibleToAllUsers

Whether the job flow is visible to all IAM users of the AWS account associated with the job flow. If this value is set to true, all IAM users of that AWS account can view and (if they have the proper policy permissions set) manage the job flow. If it is set to false, only the IAM user that created the job flow can view and manage it.

Type: Boolean Required: No

### **Response Elements**

The following element is returned in a structure named RunJobFlowResult.

#### **JobFlowId**

An unique identifier for the job flow.

Type: String

### **Errors**

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

### InternalServerError

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

HTTP Status Code: 500

### **Examples**

### 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, Signa
ture=71f79725c4dbe77c0e842718485f0b37fe6df69e1153c80f7748ebd9617ca2f3
Accept: */*
{
    "Name": "Development Job Flow",
    "Instances": {
        "KeepJobFlowAliveWhenNoSteps": "false",
        "TerminationProtected": "false",
        "InstanceGroups": [{
            "Name": "Master Instance Group",
            "InstanceRole": "MASTER",
            "InstanceCount": 1,
            "InstanceType": "ml.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": [],
    "VisibleToAllUsers": "false",
    "NewSupportedProducts": [],
    "AmiVersion": "latest"
POST / HTTP/1.1
Content-Type: application/x-amz-json-1.1
X-Amz-Target: ElasticMapReduce.CreateCluster
AUTHPARAMS
{
  "Cluster": [{
    "Name": "Test",
    "Tags": [{
      "Key": "stack",
      "Value": "Production"
      "Key": "hbase"
    } ]
```

```
}1
}
```

```
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"}
    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
{
    "ClusterId": "j-12XY3456BB"
}
```

## **SetTerminationProtection**

## **Description**

SetTerminationProtection locks a job flow so the Amazon 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 job flow is analogous to calling the Amazon EC2 DisableAPITermination API on all of the EC2 instances in a cluster.

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

To terminate a job flow 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, go to Protecting a Job Flow from Termination in the Amazon Elastic MapReduce Developer's Guide.

## **Request Parameters**

For information about the common parameters that all actions use, see Common Parameters (p. 76).

#### JobFlowIds.member.N

A list of strings that uniquely identify the job flows to protect. This identifier is returned by RunJobFlow (p. 28) and can also be obtained from DescribeJobFlows (p. 12).

Type: String list Required: Yes

### **TerminationProtected**

A Boolean that indicates whether to protect the job flow 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

## **Errors**

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

### InternalServerError

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

HTTP Status Code: 500

## **Examples**

## Sample Request

```
POST / HTTP/1.1
Content-Type: application/x-amz-json-1.1
{\tt 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: c362fadae0fce377aa63f04388aeb90c53cedb17a8bf
bb8cffcb10c2378137f9
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, Signa
ture=764b6aa1a38733cadff35a2e884887e9f1208a422266bc83ac77e8d0b80bd4cf
Accept: */*
    "JobFlowIds": ["j-3TS00IY04NFN"],
    "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
```

## **SetVisibleToAllUsers**

## **Description**

Sets whether all AWS Identity and Access Management (IAM) users under your account can access the specified job flows. This action works on running job flows. You can also set the visibility of a job flow when you launch it using the <code>VisibleToAllUsers</code> parameter of RunJobFlow (p. 28). The SetVisibleToAllUsers action can be called only by an IAM user who created the job flow or the AWS account that owns the job flow.

## **Request Parameters**

For information about the common parameters that all actions use, see Common Parameters (p. 76).

#### JobFlowIds.member.N

Identifiers of the job flows to receive the new visibility setting.

Type: String list Required: Yes

#### VisibleToAllUsers

Whether the specified job flows are visible to all IAM users of the AWS account associated with the job flow. If this value is set to True, all IAM users of that AWS account can view and, if they have the proper IAM policy permissions set, manage the job flows. If it is set to False, only the IAM user that created a job flow can view and manage it.

Type: Boolean Required: Yes

## **Errors**

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

#### InternalServerError

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

HTTP Status Code: 500

## **Examples**

## 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
```

### Amazon Elastic MapReduce API Reference Examples

```
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, Signa
ture=ela00b37787d9ccc43c9de32f1f0a73813b0bd6643d4db7762b62a7092d51997
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
```

## **TerminateJobFlows**

## **Description**

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

The call to TerminateJobFlows is asynchronous. Depending on the configuration of the job flow, it may take up to 5-20 minutes for the job flow to completely terminate and release allocated resources, such as Amazon EC2 instances.

## **Request Parameters**

For information about the common parameters that all actions use, see Common Parameters (p. 76).

#### JobFlowIds.member.N

A list of job flows to be shutdown.

Type: String list Required: Yes

## **Errors**

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

#### InternalServerError

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

HTTP Status Code: 500

## **Examples**

## 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, Signa ture=9791416eaf09f36aa753a324b0de27ff5cc7084b8548cc748487a2bcb3439d58
Accept: */*

{"JobFlowIds": ["j-3TS00IY04NFN"]}
```

### Amazon Elastic MapReduce API Reference Examples

## **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

# **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 the response is not guaranteed. Applications should not assume a particular order.

The following data types are supported:

- AddInstanceGroupsResult (p. 40)
- AddJobFlowStepsResult (p. 41)
- AddTagsResult (p. 41)
- Application (p. 41)
- BootstrapActionConfig (p. 42)
- BootstrapActionDetail (p. 43)
- Cluster (p. 43)
- ClusterStateChangeReason (p. 44)
- ClusterStatus (p. 45)
- ClusterSummary (p. 46)
- ClusterTimeline (p. 46)
- Command (p. 47)
- DescribeClusterResult (p. 47)
- DescribeJobFlowsResult (p. 47)
- DescribeStepResult (p. 48)
- Ec2InstanceAttributes (p. 48)
- HadoopJarStepConfig (p. 49)
- HadoopStepConfig (p. 50)
- Instance (p. 50)
- InstanceGroup (p. 51)
- InstanceGroupConfig (p. 52)
- InstanceGroupDetail (p. 53)
- InstanceGroupModifyConfig (p. 55)
- InstanceGroupStateChangeReason (p. 56)

### Amazon Elastic MapReduce API Reference AddInstanceGroupsResult

- InstanceGroupStatus (p. 57)
- InstanceGroupTimeline (p. 57)
- InstanceStateChangeReason (p. 58)
- InstanceStatus (p. 58)
- InstanceTimeline (p. 59)
- JobFlowDetail (p. 59)
- JobFlowExecutionStatusDetail (p. 61)
- JobFlowInstancesConfig (p. 62)
- JobFlowInstancesDetail (p. 63)
- KeyValue (p. 65)
- ListBootstrapActionsResult (p. 66)
- ListClustersResult (p. 66)
- ListInstanceGroupsResult (p. 67)
- ListInstancesResult (p. 67)
- ListStepsResult (p. 68)
- PlacementType (p. 68)
- RemoveTagsResult (p. 68)
- RunJobFlowResult (p. 69)
- ScriptBootstrapActionConfig (p. 69)
- Step (p. 69)
- StepConfig (p. 70)
- StepDetail (p. 71)
- StepExecutionStatusDetail (p. 71)
- StepStateChangeReason (p. 72)
- StepStatus (p. 73)
- StepSummary (p. 73)
- StepTimeline (p. 74)
- SupportedProductConfig (p. 74)
- Tag (p. 75)

# AddInstanceGroupsResult

## **Description**

Output from an AddInstanceGroups call.

## **Contents**

#### InstanceGroupIds

Instance group IDs of the newly created instance groups.

Type: String list

Required: No

### **JobFlowId**

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

### Amazon Elastic MapReduce API Reference AddJobFlowStepsResult

Type: String

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

Required: No

# AddJobFlowStepsResult

# **Description**

The output for the AddJobFlowSteps (p. 5) operation.

## **Contents**

### StepIds

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

Type: String list

Required: No

# AddTagsResult

# **Description**

This output indicates the result of adding tags to a resource.

## **Contents**

# **Application**

## **Description**

An 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. For more information, see Launch a Job Flow on the MapR Distribution for Hadoop. Currently supported values are:

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

## **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: String list

Required: No

#### Name

The name of the application.

Type: String Required: No

#### Version

The version of the application.

Type: String Required: No

# **BootstrapActionConfig**

# **Description**

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.

Required: Yes

#### ScriptBootstrapAction

The script run by the bootstrap action.

Type: ScriptBootstrapActionConfig (p. 69)

Required: Yes

# **BootstrapActionDetail**

# **Description**

Reports the configuration of a bootstrap action in a job flow.

## **Contents**

#### **BootstrapActionConfig**

A description of the bootstrap action.

Type: BootstrapActionConfig (p. 42)

Required: No

## Cluster

## **Description**

The detailed description of the cluster.

## **Contents**

#### **Applications**

The applications installed on this cluster.

Type: Application (p. 41) list

Required: No

AutoTerminate

Specifies whether the cluster should terminate after completing all steps.

Type: Boolean Required: No

#### Ec2InstanceAttributes

Provides information about the EC2 instances in a cluster grouped by category. For example, EC2 Key Name, Subnet Id, Instance Profile, and so on.

Type: Ec2InstanceAttributes (p. 48)

Required: No

ld

The unique identifier for the cluster.

Type: String Required: No

LoaUri

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

Type: String

### Amazon Elastic MapReduce API Reference ClusterStateChangeReason

Required: No

Name

The name of the cluster.

Type: String

Required: No

### RequestedAmiVersion

The AMI version requested for this cluster.

Type: String Required: No

### RunningAmiVersion

The AMI version running on this cluster. This differs from the requested version only if the requested version is a meta version, such as "latest".

Type: String Required: No

#### Status

The current status details about the cluster.

Type: ClusterStatus (p. 45)

Required: No

#### Tags

A list of tags associated with a cluster.

Type: Tag (p. 75) list

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 job flow is visible to all IAM users of the AWS account associated with the job flow. If this value is set to true, all IAM users of that AWS account can view and manage the job flow if they have the proper policy permissions set. If this value is false, only the IAM user that created the cluster can view and manage it. This value can be changed using the SetVisibleToAllUsers (p. 35) action.

Type: Boolean Required: No

# ClusterStateChangeReason

# **Description**

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 | BOOTSTRAP\_FAILURE | USER\_REQUEST | STEP\_FAILURE | ALL\_STEPS\_COMPLETED

Required: No

#### Message

The descriptive message for the state change reason.

Type: String Required: No

## **ClusterStatus**

# **Description**

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. 44)

Required: No

#### **Timeline**

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

Type: ClusterTimeline (p. 46)

Required: No

# **ClusterSummary**

# **Description**

The summary description of the cluster.

## **Contents**

ld

The unique identifier for the cluster.

Type: String

Required: No

Name

The name of the cluster.

Type: String

Required: No

**Status** 

The details about the current status of the cluster.

Type: ClusterStatus (p. 45)

Required: No

## ClusterTimeline

# **Description**

Represents the timeline of the cluster's lifecycle.

## **Contents**

#### CreationDateTime

The creation date and time of the cluster.

Type: DateTime

Required: No

**EndDateTime** 

The date and time when the cluster was terminated.

Type: DateTime

Required: No

ReadyDateTime

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

Type: DateTime

Required: No

## **Command**

## **Description**

An entity describing an executable that runs on a cluster.

## **Contents**

### Args

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

Type: String list

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

## **DescribeClusterResult**

# **Description**

This output contains the description of the cluster.

## **Contents**

### Cluster

This output contains the details for the requested cluster.

Type: Cluster (p. 43)

Required: No

## **DescribeJobFlowsResult**

# **Description**

The output for the DescribeJobFlows (p. 12) operation.

## **Contents**

### **JobFlows**

A list of job flows matching the parameters supplied.

Type: JobFlowDetail (p. 59) list

Required: No

# **DescribeStepResult**

## **Description**

This output contains the description of the cluster step.

## **Contents**

### Step

The step details for the requested step identifier.

Type: Step (p. 69)

Required: No

## Ec2InstanceAttributes

## **Description**

Provides information about the EC2 instances in a cluster grouped by category. For example, EC2 Key Name, Subnet Id, Instance Profile, and so on.

## **Contents**

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

### Amazon Elastic MapReduce API Reference HadoopJarStepConfig

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 job flow launched in a VPC.

Type: String

Required: No

lamInstanceProfile

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

assume this role.

Type: String Required: No

# **HadoopJarStepConfig**

# **Description**

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: String list

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.

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.

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: KeyValue (p. 65) list

Required: No

# **HadoopStepConfig**

# **Description**

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: String list

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

## **Instance**

## **Description**

Represents an EC2 instance provisioned as part of cluster.

## **Contents**

#### Ec2InstanceId

The unique identifier of the instance in Amazon EC2.

Type: String

Required: No

ld

The unique identifier for the instance in Amazon EMR.

### Amazon Elastic MapReduce API Reference InstanceGroup

Type: String

Required: No **PrivateDnsName** 

The private DNS name of the instance.

Type: String

Required: No **PrivatelpAddress** 

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. 58)

Required: No

# InstanceGroup

# **Description**

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**

### **BidPrice**

The bid price for each EC2 instance in the instance group when launching nodes as Spot Instances, expressed in USD.

Type: String

Required: No

ld

The identifier of the instance group.

Type: String

### Amazon Elastic MapReduce API Reference InstanceGroupConfig

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.

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

**Status** 

The current status of the instance group.

Type: InstanceGroupStatus (p. 57)

Required: No

# InstanceGroupConfig

# **Description**

Configuration defining a new instance group.

## **Contents**

### **BidPrice**

Bid price for each Amazon EC2 instance in the instance group when launching nodes as Spot Instances, expressed in USD.

Type: String

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

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

Type: String

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

Required: Yes

### Market

Market type of the Amazon 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.

Required: No

# InstanceGroupDetail

# **Description**

Detailed information about an instance group.

# Amazon Elastic MapReduce API Reference Contents

## **Contents**

#### **BidPrice**

Bid price for EC2 Instances when launching nodes as Spot Instances, expressed in USD.

Type: String

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

Required: No CreationDateTime

The date/time the instance group was created.

Type: DateTime

Required: Yes

**EndDateTime** 

The date/time the instance group was terminated.

Type: DateTime

Required: No InstanceGroupId

Unique identifier for the instance group.

Type: String

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

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

Amazon EC2 Instance type.

Type: String

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

Required: Yes

### Amazon Elastic MapReduce API Reference InstanceGroupModifyConfig

#### LastStateChangeReason

Details regarding the state of the instance group.

Type: String

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

Required: No

#### Market

Market type of the Amazon 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.

Required: No

### ReadyDateTime

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

Type: DateTime

Required: No

#### **StartDateTime**

The date/time the instance group was started.

Type: DateTime

Required: No

### State

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

Type: String

Valid Values: PROVISIONING | BOOTSTRAPPING | RUNNING | RESIZING | SUSPENDED | TERMINATING | TERMINATED | ARRESTED | SHUTTING\_DOWN | ENDED

Required: Yes

# InstanceGroupModifyConfig

# **Description**

Modify an instance group size.

## **Contents**

#### EC2InstanceIdsToTerminate

The EC2 InstanceIds to terminate. For advanced users only. Once you terminate the instances, the instance group will not return to its original requested size.

Type: String list

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.

Required: Yes

# InstanceGroupStateChangeReason

## **Description**

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

# **InstanceGroupStatus**

# **Description**

The details of the instance group status.

## **Contents**

#### State

The current state of the instance group. The following values are deprecated: ARRESTED, SHUTTING\_DOWN, and ENDED. Use SUSPENDED, TERMINATING, and TERMINATED instead, respectively.

Type: String

Valid Values: PROVISIONING | BOOTSTRAPPING | RUNNING | RESIZING | SUSPENDED | TERMINATING | TERMINATED | ARRESTED | SHUTTING\_DOWN | ENDED

Required: No StateChangeReason

The status change reason details for the instance group.

Type: InstanceGroupStateChangeReason (p. 56)

Required: No

**Timeline** 

The timeline of the instance group status over time.

Type: InstanceGroupTimeline (p. 57)

Required: No

# InstanceGroupTimeline

## **Description**

The timeline of the instance group lifecycle.

## **Contents**

### CreationDateTime

The creation date and time of the instance group.

Type: DateTime

Required: No

### **EndDateTime**

The date and time when the instance group terminated.

Type: DateTime Required: No

### Amazon Elastic MapReduce API Reference InstanceStateChangeReason

### ReadyDateTime

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

Type: DateTime Required: No

# InstanceStateChangeReason

# **Description**

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

## **InstanceStatus**

# **Description**

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. 58)

### Amazon Elastic MapReduce API Reference InstanceTimeline

Required: No

#### **Timeline**

The timeline of the instance status over time.

Type: InstanceTimeline (p. 59)

Required: No

# **InstanceTimeline**

# **Description**

The timeline of the instance lifecycle.

## **Contents**

### CreationDateTime

The creation date and time of the instance.

Type: DateTime

Required: No

#### **EndDateTime**

The date and time when the instance was terminated.

Type: DateTime

Required: No

### ReadyDateTime

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

Type: DateTime

Required: No

# **JobFlowDetail**

## **Description**

A description of a job flow.

## **Contents**

### **AmiVersion**

The version of the AMI used to initialize Amazon EC2 instances in the job flow. For a list of AMI versions currently supported by Amazon ElasticMapReduce, go to AMI Versions Supported in Elastic MapReduce in the Amazon Elastic MapReduce Developer's Guide.

Type: String

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

# Amazon Elastic MapReduce API Reference Contents

Required: No **BootstrapActions** 

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

Type: BootstrapActionDetail (p. 43) list

Required: No

**ExecutionStatusDetail** 

Describes the execution status of the job flow.

Type: JobFlowExecutionStatusDetail (p. 61)

Required: Yes

Instances

Describes the Amazon EC2 instances of the job flow.

Type: JobFlowInstancesDetail (p. 63)

Required: Yes

**JobFlowId** 

The job flow identifier.

Type: String

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

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.

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.

Required: No

Name

The name of the job flow.

Type: String

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

Required: Yes

Steps

A list of steps run by the job flow.

Type: StepDetail (p. 71) list

Required: No

### Amazon Elastic MapReduce API Reference JobFlowExecutionStatusDetail

#### **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: String list Required: No VisibleToAllUsers

Specifies whether the job flow is visible to all IAM users of the AWS account associated with the job flow. If this value is set to true, all IAM users of that AWS account can view and (if they have the proper policy permissions set) manage the job flow. If it is set to false, only the IAM user that created the job flow can view and manage it. This value can be changed using the SetVisibleToAllUsers (p. 35) action.

Type: Boolean Required: No

## **JobFlowExecutionStatusDetail**

# **Description**

Describes the status of the job flow.

## **Contents**

### CreationDateTime

The creation date and time of the job flow.

Type: DateTime Required: Yes

### **EndDateTime**

The completion date and time of the job flow.

Type: DateTime Required: No

### LastStateChangeReason

Description of the job flow last changed state.

Type: String

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

Required: No ReadyDateTime

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

Type: DateTime Required: No

**StartDateTime** 

The start date and time of the job flow.

### Amazon Elastic MapReduce API Reference JobFlowInstancesConfig

Type: DateTime

Required: No

**State** 

The state of the job flow.

Type: String

Valid Values: STARTING | BOOTSTRAPPING | RUNNING | WAITING | SHUTTING DOWN |

TERMINATED | COMPLETED | FAILED

Required: Yes

# **JobFlowInstancesConfig**

# **Description**

A description of the Amazon EC2 instance running the job flow.

## **Contents**

### Ec2KeyName

The name of the Amazon 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.

Required: No

### Ec2SubnetId

To launch the job flow in Amazon Virtual Private Cloud (Amazon VPC), set this parameter to the identifier of the Amazon VPC subnet where you want the job flow to launch. If you do not specify this value, the job flow is launched in the normal Amazon Web Services cloud, outside of an Amazon 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 job flow launched in a Amazon VPC.

Type: String

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

Required: No

#### **HadoopVersion**

The Hadoop version for the job flow. Valid inputs are "0.18", "0.20", or "0.20.205". 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.

Required: No

### Amazon Elastic MapReduce API Reference JobFlowInstancesDetail

#### InstanceCount

The number of Amazon EC2 instances used to execute the job flow.

Type: Integer
Required: No
InstanceGroups

Configuration for the job flow's instance groups.

Type: InstanceGroupConfig (p. 52) list

Required: No

### KeepJobFlowAliveWhenNoSteps

Specifies whether the job flow should terminate 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.

Required: No

#### **Placement**

The Availability Zone the job flow will run in.

Type: PlacementType (p. 68)

Required: No SlaveInstanceType

The EC2 instance type of the slave nodes.

Type: String

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

Required: No

### **TerminationProtected**

Specifies whether to lock the job flow 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

## **JobFlowInstancesDetail**

## **Description**

Specify the type of Amazon EC2 instances to run the job flow on.

# Amazon Elastic MapReduce API Reference Contents

## **Contents**

### Ec2KeyName

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

Type: String

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

Required: No

#### Ec2SubnetId

For job flows launched within Amazon Virtual Private Cloud, this value specifies the identifier of the subnet where the job flow was launched.

Type: String

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

Required: No **HadoopVersion** 

The Hadoop version for the job flow.

Type: String

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

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 slave node. If the value is greater than 1, one instance is the master node and all others are slave nodes.

Type: Integer

Required: Yes

### InstanceGroups

Details about the job flow's instance groups.

Type: InstanceGroupDetail (p. 53) list

Required: No

### KeepJobFlowAliveWhenNoSteps

Specifies whether the job flow should terminate 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.

Required: No MasterInstanceType

The Amazon EC2 master node instance type.

### Amazon Elastic MapReduce API Reference KeyValue

Type: String

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

Required: Yes

### **MasterPublicDnsName**

The DNS name of the master node.

Type: String

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

Required: No

### NormalizedInstanceHours

An approximation of the cost of the job flow, represented in m1.small/hours. This value is incremented once for every hour 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 job flow.

Type: PlacementType (p. 68)

Required: No SlaveInstanceType

The Amazon EC2 slave node instance type.

Type: String

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

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

# **KeyValue**

## **Description**

A key value pair.

## **Contents**

Kev

The unique identifier of a key value pair.

### Amazon Elastic MapReduce API Reference ListBootstrapActionsResult

Type: String

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

Required: No

Value

The value part of the identified key.

Type: String

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

Required: No

# ListBootstrapActionsResult

## **Description**

This output contains the bootstrap actions detail.

## **Contents**

### **BootstrapActions**

The bootstrap actions associated with the cluster.

Type: Command (p. 47) list

Required: No

### Marker

The pagination token is a random string indicating whether there are more results to fetch. Use the pagination token in later API calls to retrieve the next page of results. When the value is null, all results have been returned.

Type: String Required: No

## ListClustersResult

## **Description**

This contains a ClusterSummaryList with the cluster details; for example, the cluster IDs, names, and status.

## **Contents**

#### **Clusters**

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

Type: ClusterSummary (p. 46) list

Required: No

### Amazon Elastic MapReduce API Reference ListInstanceGroupsResult

#### Marker

The pagination token is a random string indicating whether there are more results to fetch. Use the pagination token in later API calls to retrieve the next page of results. When the value is null, all results have been returned.

Type: String Required: No

# ListInstanceGroupsResult

# **Description**

This input determines which instance groups to retrieve.

## **Contents**

### InstanceGroups

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

Type: InstanceGroup (p. 51) list

Required: No

#### Marker

The pagination token is a random string indicating whether there are more results to fetch. Use the pagination token in later API calls to retrieve the next page of results. When the value is null, all results have been returned.

Type: String Required: No

## ListInstancesResult

## **Description**

This output contains the list of instances.

## **Contents**

#### Instances

The list of instances for the cluster and given filters.

Type: Instance (p. 50) list

Required: No

#### Marker

The pagination token is a random string indicating whether there are more results to fetch. Use the pagination token in later API calls to retrieve the next page of results. When the value is null, all results have been returned.

Type: String

Required: No

# **ListStepsResult**

## **Description**

This output contains the list of steps.

## **Contents**

#### Marker

The pagination token is a random string indicating whether there are more results to fetch. Use the pagination token in later API calls to retrieve the next page of results. When the value is null, all results have been returned.

Type: String

Required: No

### **Steps**

The filtered list of steps for the cluster.

Type: StepSummary (p. 73) list

Required: No

# **PlacementType**

## **Description**

The Amazon EC2 location for the job flow.

## **Contents**

### AvailabilityZone

The Amazon EC2 Availability Zone for the job flow.

Type: String

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

Required: Yes

# RemoveTagsResult

## **Description**

This output indicates the result of removing tags from a resource.

## **Contents**

## RunJobFlowResult

## **Description**

The result of the RunJobFlow (p. 28) operation.

## **Contents**

### **JobFlowId**

An unique identifier for the job flow.

Type: String

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

Required: No

# ScriptBootstrapActionConfig

# **Description**

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: String list

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.

Required: Yes

# **Step**

# **Description**

This represents a step in a cluster.

## **Contents**

#### **ActionOnFailure**

This specifies what action to take when the cluster step fails. TERMINATE\_JOB\_FLOW is deprecated, use 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. 50)

Required: No

ld

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. 73)

Required: No

# **StepConfig**

## **Description**

Specification of a job flow step.

## **Contents**

#### **ActionOnFailure**

The action to take if the job flow step fails.

Type: String

Valid Values: TERMINATE\_JOB\_FLOW | TERMINATE\_CLUSTER | CANCEL\_AND\_WAIT | CONTINUE

Required: No

### **HadoopJarStep**

The JAR file used for the job flow step.

### Amazon Elastic MapReduce API Reference StepDetail

Type: HadoopJarStepConfig (p. 49)

Required: Yes

Name

The name of the job flow step.

Type: String

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

Required: Yes

# **StepDetail**

# **Description**

Combines the execution state and configuration of a step.

## **Contents**

#### **ExecutionStatusDetail**

The description of the step status.

Type: StepExecutionStatusDetail (p. 71)

Required: Yes

**StepConfig** 

The step configuration.

Type: StepConfig (p. 70)

Required: Yes

# **StepExecutionStatusDetail**

# **Description**

The execution state of a step.

## **Contents**

### CreationDateTime

The creation date and time of the step.

Type: DateTime

Required: Yes

#### **EndDateTime**

The completion date and time of the step.

Type: DateTime

### Amazon Elastic MapReduce API Reference StepStateChangeReason

Required: No

### LastStateChangeReason

A description of the step's current state.

Type: String

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

Required: No StartDateTime

The start date and time of the step.

Type: DateTime

Required: No

State

The state of the job flow step.

Type: String

Valid Values: PENDING | RUNNING | CONTINUE | COMPLETED | CANCELLED | FAILED |

INTERRUPTED

Required: Yes

# **StepStateChangeReason**

## **Description**

The details of the step state change reason.

## **Contents**

### Code

The programmable code for the state change reason.

Type: String

Valid Values: NONE

Required: No

### Message

The descriptive message for the state change reason.

Type: String

Required: No

# **StepStatus**

# **Description**

The execution status details of the cluster step.

## **Contents**

#### State

The execution state of the cluster step.

Type: String

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

Required: No

StateChangeReason

The reason for the step execution status change.

Type: StepStateChangeReason (p. 72)

Required: No

**Timeline** 

The timeline of the cluster step status over time.

Type: StepTimeline (p. 74)

Required: No

# **StepSummary**

# **Description**

The summary of the cluster step.

## **Contents**

ld

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. 73)

Required: No

# **StepTimeline**

## **Description**

The timeline of the cluster step lifecycle.

## **Contents**

### CreationDateTime

The date and time when the cluster step was created.

Type: DateTime

Required: No

### **EndDateTime**

The date and time when the cluster step execution completed or failed. This can display a time that pre-dates a call to DescribeStep that indicates the step is running, due to delays in step status reporting.

Type: DateTime

Required: No

### **StartDateTime**

The date and time when the cluster step execution started. Due to delays in step status reporting, this can display a time which pre-dates a previous call to DescribeStep that indicated the step was not yet running.

Type: DateTime

Required: No

# **SupportedProductConfig**

## **Description**

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: String list

Required: No

### Name

The name of the product configuration.

### Amazon Elastic MapReduce API Reference Tag

Type: String

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

Required: No

# **Tag**

# **Description**

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 Tagging Amazon EMR Resources.

## **Contents**

### Key

A user-defined key, which is the minimum required information for a valid tag. For more information, see Tagging Amazon EMR Resources.

Type: String

Required: No

#### Value

A user-defined value, which is optional in a tag. For more information, see Tagging Amazon EMR Resources.

Type: String

Required: No

# **Common Parameters**

This section lists the request parameters that all actions use. Any action-specific parameters are listed in the topic for the action.

### Action

The action to be performed.

Default: None

Type: string

Required: Yes

#### **AuthParams**

The parameters that are required to authenticate a Conditional request. Contains:

- AWSAccessKeyID
- SignatureVersion
- Timestamp
- Signature

Default: None

Required: Conditional

### **AWSAccessKeyId**

The access key ID that corresponds to the secret access key that you used to sign the request.

Default: None

Type: string

Required: Yes

### **Expires**

The date and time when the request signature expires, expressed in the format YYYY-MM-DDThh:mm:ssZ, as specified in the ISO 8601 standard.

Condition: Requests must include either *Timestamp* or *Expires*, but not both.

Default: None

Type: string

Required: Conditional

#### SecurityToken

The temporary security token that was obtained through a call to AWS Security Token Service. For a list of services that support AWS Security Token Service, go to Using Temporary Security Credentials to Access AWS in **Using Temporary Security Credentials**.

Default: None

Type: string

Required: No

### **Signature**

The digital signature that you created for the request. For information about generating a signature, go to the service's developer documentation.

Default: None

Type: string

Required: Yes

### **SignatureMethod**

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

Default: None

Type: string

Valid Values: HmacSHA256 | HmacSHA1

Required: Yes

### **SignatureVersion**

The signature version you use to sign the request. Set this to the value that is recommended for your service.

Default: None

Type: string

Required: Yes

#### **Timestamp**

The date and time when the request was signed, expressed in the format YYYY-MM-DDThh:mm:ssZ, as specified in the ISO 8601 standard.

Condition: Requests must include either *Timestamp* or *Expires*, but not both.

Default: None

Type: string

Required: Conditional

### Version

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

Default: None

Type: string

Required: Yes

# **Common Errors**

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

### IncompleteSignature

The request signature does not conform to AWS standards.

HTTP Status Code: 400

### InternalFailure

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

HTTP Status Code: 500

### InvalidAction

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

HTTP Status Code: 400

#### InvalidClientTokenId

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

HTTP Status Code: 403

### InvalidParameterCombination

Parameters that must not be used together were used together.

HTTP Status Code: 400

#### **InvalidParameterValue**

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

HTTP Status Code: 400

#### InvalidQueryParameter

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

HTTP Status Code: 400

### MalformedQueryString

The guery string contains a syntax error.

HTTP Status Code: 404

#### MissingAction

The request is missing an action or a required parameter.

HTTP Status Code: 400

### MissingAuthenticationToken

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

HTTP Status Code: 403

#### MissingParameter

A required parameter for the specified action is not supplied.

HTTP Status Code: 400

### **OptInRequired**

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

HTTP Status Code: 403

### RequestExpired

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

HTTP Status Code: 400

### ServiceUnavailable

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

HTTP Status Code: 503

### **Throttling**

The request was denied due to request throttling.

HTTP Status Code: 400

### ValidationError

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

HTTP Status Code: 400