
Amazon Elastic MapReduce

API Reference

API Version 2009-03-31



Amazon Elastic MapReduce: API Reference

Copyright © 2011 - 2012 Amazon Web Services LLC or its affiliates. All rights reserved.

Welcome	1
Actions	2
AddInstanceGroups	3
AddJobFlowSteps	5
DescribeJobFlows	7
ModifyInstanceGroups	11
RunJobFlow	12
SetTerminationProtection	16
SetVisibleToAllUsers	18
TerminateJobFlows	20
Data Types	22
AddInstanceGroupsResult	23
BootstrapActionConfig	23
BootstrapActionDetail	23
DescribeJobFlowsResult	24
HadoopJarStepConfig	24
InstanceGroupConfig	25
InstanceGroupDetail	25
InstanceGroupModifyConfig	27
JobFlowDetail	27
JobFlowExecutionStatusDetail	29
JobFlowInstancesConfig	29
JobFlowInstancesDetail	31
KeyValue	32
PlacementType	33
RunJobFlowResult	33
ScriptBootstrapActionConfig	33
StepConfig	34
StepDetail	34
StepExecutionStatusDetail	34
Common Query Parameters	36
Common Errors	38

Welcome

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

Amazon Elastic MapReduce is a web service that makes it easy to process large amounts of data efficiently. Elastic MapReduce 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 18, 2012.

Actions

The actions described in this guide are called by using the AWS Query protocol.

The following actions are supported:

- [AddInstanceGroups](#) (p. 3)
- [AddJobFlowSteps](#) (p. 5)
- [DescribeJobFlows](#) (p. 7)
- [ModifyInstanceGroups](#) (p. 11)
- [RunJobFlow](#) (p. 12)
- [SetTerminationProtection](#) (p. 16)
- [SetVisibleToAllUsers](#) (p. 18)
- [TerminateJobFlows](#) (p. 20)

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 Query Parameters \(p. 36\)](#).

InstanceGroups.member.N

Instance Groups to add.

Type: [InstanceGroupConfig \(p. 25\)](#) 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 `AddInstanceGroupsResult` structure.

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 \(p. 38\)](#).

InternalServerError

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

HTTP Status Code: 500

Examples

Sample Request

```
https://elasticmapreduce.amazonaws.com?Action=AddInstanceGroups
&JobFlowId=j-3UN6WX5RRO2AG
&InstanceGroups.member.1.Name="Task Instance Group"
&InstanceGroups.member.1.InstanceRole=TASK
&InstanceGroups.member.1.InstanceType=m1.small
&InstanceGroups.member.1.InstanceRequestCount=2
&AuthParams
```

Sample Response

```
<AddInstanceGroupsResponse xmlns="http://elasticmapreduce.amazonaws.com/doc/2009-
03-31">
  <InstanceGroupIds>
    <InstanceGroup>
      <InstanceGroupId>
        i-3UN6WX5RRO2AG
      </InstanceGroupId>
    </InstanceGroup>
  </InstanceGroupIds>
</AddInstanceGroupsResponse>
```

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 Query Parameters \(p. 36\)](#).

JobFlowId

A string that uniquely identifies the job flow. This identifier is returned by [RunJobFlow \(p. 12\)](#) and can also be obtained from [DescribeJobFlows \(p. 7\)](#).

Type: String

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

Required: Yes

Steps.member.N

A list of [StepConfig \(p. 34\)](#) to be executed by the job flow.

Type: [StepConfig \(p. 34\)](#) list

Required: Yes

Errors

For information about the errors that are common to all actions, see [Common Errors \(p. 38\)](#).

InternalServerError

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

HTTP Status Code: 500

Examples

Sample Request

```
https://elasticmapreduce.amazonaws.com?Action=AddJobFlowSteps
&JobFlowId=j-3UN6WX5RRO2AG
&Steps.member.1.Name=MyStep2
&Steps.member.1.ActionOnFailure=CONTINUE
&Steps.member.1.HadoopJarStep.Jar=MyJar
&Steps.member.1.HadoopJarStep.MainClass=MainClass
&Steps.member.1.HadoopJarStep.Args.member.1=arg1
&AuthParams
```

Sample Response

```
<AddJobFlowStepsResponse xmlns="http://elasticmapreduce.amazonaws.com/doc/2009-
03-31">
  <ResponseMetadata>
    <RequestId>
      df6f4f4a-ed85-11dd-9877-6fad448a8419
    </RequestId>
  </ResponseMetadata>
</AddJobFlowStepsResponse>
```

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 Query Parameters \(p. 36\)](#).

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 `DescribeJobFlowsResult` structure.

JobFlows

A list of job flows matching the parameters supplied.

Type: [JobFlowDetail \(p. 27\)](#) list

Errors

For information about the errors that are common to all actions, see [Common Errors \(p. 38\)](#).

InternalServerError

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

HTTP Status Code: 500

Examples

Sample Request

```
https://elasticmapreduce.amazonaws.com?Action=DescribeJobFlows
&JobFlowIds.member.1=j-3UN6WX5RRO2AG
&AuthParams
```

Sample Response

```
<DescribeJobFlowsResponse xmlns="http://elasticmapreduce.amazonaws.com/doc/2009-03-31">
  <DescribeJobFlowsResult>
    <JobFlows>
      <member>
        <ExecutionStatusDetail>
          <CreationDateTime>
            2009-01-28T21:49:16Z
          </CreationDateTime>
          <StartDateTime>
            2009-01-28T21:49:16Z
          </StartDateTime>
          <State>
            STARTING
          </State>
        </ExecutionStatusDetail>
        <Name>
          MyJobFlowName
        </Name>
        <LogUri>
          myawsbucket/subdir/
        </LogUri>
        <Steps>
          <member>
            <ExecutionStatusDetail>
              <CreationDateTime>
                2009-01-28T21:49:16Z
              </CreationDateTime>
              <State>
                PENDING
              </State>
            </ExecutionStatusDetail>
            <StepConfig>
              <HadoopJarStep>
```

Amazon Elastic MapReduce API Reference Examples

```
        <Jar>
          MyJarFile
        </Jar>
        <MainClass>
          MyMainClass
        </MainClass>
        <Args>
          <member>
            arg1
          </member>
          <member>
            arg2
          </member>
        </Args>
        <Properties/>
      </HadoopJarStep>
      <Name>
        MyStepName
      </Name>
      <ActionOnFailure>
        CONTINUE
      </ActionOnFailure>
    </StepConfig>
  </member>
</Steps>
<JobFlowId>
  j-3UN6WX5RRO2AG
</JobFlowId>
<Instances>
  <Placement>
    <AvailabilityZone>
      us-east-1a
    </AvailabilityZone>
  </Placement>
  <SlaveInstanceType>
    m1.small
  </SlaveInstanceType>
  <MasterInstanceType>
    m1.small
  </MasterInstanceType>
  <Ec2KeyName>
    myec2keyname
  </Ec2KeyName>
  <InstanceCount>
    4
  </InstanceCount>
  <KeepJobFlowAliveWhenNoSteps>
    true
  </KeepJobFlowAliveWhenNoSteps>
  <TerminationProtected>
    true
  </TerminationProtected>
</Instances>
</member>
</JobFlows>
</DescribeJobFlowsResult>
<ResponseMetadata>
  <RequestId>
```

Amazon Elastic MapReduce API Reference Examples

```
    9cea3229-ed85-11dd-9877-6fad448a8419
  </RequestId>
</ResponseMetadata>
</DescribeJobFlowsResponse>
```

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 Query Parameters \(p. 36\)](#).

InstanceGroups.member.N

Instance groups to change.

Type: [InstanceGroupModifyConfig \(p. 27\)](#) list

Required: No

Errors

For information about the errors that are common to all actions, see [Common Errors \(p. 38\)](#).

InternalServerError

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

HTTP Status Code: 500

Examples

Sample Request

```
https://elasticmapreduce.amazonaws.com?Action=ModifyInstanceGroups
&InstanceGroups.member.1.InstanceGroupId=i-3UN6WX5RRO2AG
&InstanceGroups.member.1.InstanceRequestCount=2
&AuthParams
```

Sample Response

```
<ModifyInstanceGroupsResponse xmlns="http://elasticmapreduce.amazon
aws.com/doc/2009-03-31">
  <ResponseMetadata>
    <RequestId>
      2690d7eb-ed86-11dd-9877-6fad448a8419
    </RequestId>
  </ResponseMetadata>
</ModifyInstanceGroupsResponse>
```

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 \(p. 29\)](#) `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 \(p. 29\)](#) `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 Query Parameters \(p. 36\)](#).

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 this value is not specified, the job flow uses the default of (AMI 1.0, Hadoop 0.18).

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 \(p. 29\)](#) `HadoopVersion` parameter to modify the version of Hadoop from the defaults shown above.

For details about the 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.

Required: No

BootstrapActions.member.N

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

Type: [BootstrapActionConfig \(p. 23\)](#) list

Required: No

Instances

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

Type: [JobFlowInstancesConfig \(p. 29\)](#)

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

Specifies 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

Steps.member.N

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

Type: [StepConfig \(p. 34\)](#) 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:

- "karmasphere-enterprise-utility" - tag the job flow for management by [Karmasphere](#).
- "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

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 `RunJobFlowResult` structure.

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 \(p. 38\)](#).

InternalServerError

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

HTTP Status Code: 500

Examples

Sample Request

```
https://elasticmapreduce.amazonaws.com?Action=RunJobFlow
&Name=MyJobFlowName
&LogUri=s3n%3A%2F%2Fmybucket%2Fsubdir
&Instances.MasterInstanceType=m1.small
&Instances.SlaveInstanceType=m1.small
&Instances.InstanceCount=4
&Instances.Ec2KeyName=myec2keyname
&Instances.Placement.AvailabilityZone=us-east-1a
&Instances.KeepJobFlowAliveWhenNoSteps=true
&Instances.TerminationProtected=true
&Steps.member.1.Name=MyStepName
&Steps.member.1.ActionOnFailure=CONTINUE
&Steps.member.1.HadoopJarStep.Jar=MyJarFile
&Steps.member.1.HadoopJarStep.MainClass=MyMainClass
&Steps.member.1.HadoopJarStep.Args.member.1=arg1
&Steps.member.1.HadoopJarStep.Args.member.2=arg2
&AuthParams
```

Sample Response

```
<RunJobFlowResponse xmlns="http://elasticmapreduce.amazonaws.com/doc/2009-03-31">
  <RunJobFlowResult>
```

Amazon Elastic MapReduce API Reference Examples

```
<JobFlowId>
  j-3UN6WX5RRO2AG
</JobFlowId>
</RunJobFlowResult>
<ResponseMetadata>
  <RequestId>
    8296d8b8-ed85-11dd-9877-6fad448a8419
  </RequestId>
</ResponseMetadata>
</RunJobFlowResponse>
```

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 Query Parameters \(p. 36\)](#).

JobFlowIds.member.N

A list of strings that uniquely identify the job flows to protect. This identifier is returned by [RunJobFlow \(p. 12\)](#) and can also be obtained from [DescribeJobFlows \(p. 7\)](#).

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 \(p. 38\)](#).

InternalServerError

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

HTTP Status Code: 500

Examples

Sample Request

```
https://elasticmapreduce.amazonaws.com?Action=SetTerminationProtection
&JobFlowIds.member.1=j-3UN6WX5RRO2AG
&TerminationProtected=true
&AuthParams
```

Sample Response

```
<SetTerminationProtection xmlns="http://elasticmapreduce.amazonaws.com/doc/2009-
03-31">
  <ResponseMetadata>
    <RequestId>
      df6f4f4a-ed85-11dd-9877-6fad448a8419
    </RequestId>
  </ResponseMetadata>
</SetTerminationProtectionResponse>
```

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 `VisibleToAllUsers` parameter of [RunJobFlow](#) (p. 12). 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 Query Parameters](#) (p. 36).

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. 38).

InternalServerError

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

HTTP Status Code: 500

Examples

Sample Request

```
https://elasticmapreduce.amazonaws.com?Action=SetVisibleToAllUsers
&VisibleToAllUsers=true
&JobFlowIds.member.1=j-3UN6WX5RRO2AG
&AuthParams
```

Sample Response

```
<SetVisibleToAllUsersResponse xmlns="http://elasticmapreduce.amazon
aws.com/doc/2009-03-31">
  <ResponseMetadata>
    <RequestId>
      2690d7eb-ed86-11dd-9877-6fad448a8419
    </RequestId>
  </ResponseMetadata>
</SetVisibleToAllUsersResponse>
```

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 Query Parameters \(p. 36\)](#).

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 \(p. 38\)](#).

InternalServerError

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

HTTP Status Code: 500

Examples

Sample Request

```
https://elasticmapreduce.amazonaws.com?Action=TerminateJobFlows
&JobFlowIds.member.1=j-3UN6WX5RRO2AG
&AuthParams
```

Sample Response

```
<TerminateJobFlowsResponse xmlns="http://elasticmapreduce.amazonaws.com/doc/2009-
03-31">
  <ResponseMetadata>
    <RequestId>
      2690d7eb-ed86-11dd-9877-6fad448a8419
    </RequestId>
  </ResponseMetadata>
```

```
</TerminateJobFlowsResponse>
```


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. 23)
- [BootstrapActionConfig](#) (p. 23)
- [BootstrapActionDetail](#) (p. 23)
- [DescribeJobFlowsResult](#) (p. 24)
- [HadoopJarStepConfig](#) (p. 24)
- [InstanceGroupConfig](#) (p. 25)
- [InstanceGroupDetail](#) (p. 25)
- [InstanceGroupModifyConfig](#) (p. 27)
- [JobFlowDetail](#) (p. 27)
- [JobFlowExecutionStatusDetail](#) (p. 29)
- [JobFlowInstancesConfig](#) (p. 29)
- [JobFlowInstancesDetail](#) (p. 31)
- [KeyValue](#) (p. 32)
- [PlacementType](#) (p. 33)
- [RunJobFlowResult](#) (p. 33)
- [ScriptBootstrapActionConfig](#) (p. 33)
- [StepConfig](#) (p. 34)
- [StepDetail](#) (p. 34)
- [StepExecutionStatusDetail](#) (p. 34)

AddInstanceGroupsResult

Description

Output from an AddInstanceGroups call.

Contents

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

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

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.

ScriptBootstrapAction

The script run by the bootstrap action.

Type: [ScriptBootstrapActionConfig](#) (p. 33)

BootstrapActionDetail

Description

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

Contents

BootstrapActionConfig

A description of the bootstrap action.

Type: [BootstrapActionConfig](#) (p. 23)

DescribeJobFlowsResult

Description

The output for the [DescribeJobFlows](#) (p. 7) operation.

Contents

JobFlows

A list of job flows matching the parameters supplied.

Type: [JobFlowDetail](#) (p. 27) list

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

Jar

A path to a JAR file run during the step.

Type: String

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

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.

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. 32) list

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.

InstanceCount

Target number of instances for the instance group.

Type: Integer

InstanceRole

The role of the instance group in the cluster.

Type: String

Valid Values: MASTER | CORE | TASK

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.

Market

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

Type: String

Valid Values: ON_DEMAND | SPOT

Name

Friendly name given to the instance group.

Type: String

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

InstanceGroupDetail

Description

Detailed information about an instance group.

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.

CreationDateTime

The date/time the instance group was created.

Type: DateTime

EndDateTime

The date/time the instance group was terminated.

Type: DateTime

InstanceGroupID

Unique identifier for the instance group.

Type: String

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

InstanceRequestCount

Target number of instances to run in the instance group.

Type: Integer

InstanceRole

Instance group role in the cluster

Type: String

Valid Values: MASTER | CORE | TASK

InstanceRunningCount

Actual count of running instances.

Type: Integer

InstanceType

Amazon EC2 Instance type.

Type: String

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

LastStateChangeReason

Details regarding the state of the instance group.

Type: String

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

Market

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

Type: String

Valid Values: ON_DEMAND | SPOT

Name

Friendly name for the instance group.

Type: String

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

ReadyDateTime

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

Type: DateTime

StartDateTime

The date/time the instance group was started.

Type: DateTime

State

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

Type: String

Valid Values: PROVISIONING | STARTING | BOOTSTRAPPING | RUNNING | RESIZING | ARRESTED | SHUTTING_DOWN | TERMINATED | FAILED | ENDED

InstanceGroupModifyConfig

Description

Modify an instance group size.

Contents

InstanceCount

Target size for the instance group.

Type: Integer

InstanceGroupid

Unique ID of the instance group to expand or shrink.

Type: String

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

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.

BootstrapActions

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

Type: [BootstrapActionDetail \(p. 23\)](#) list

ExecutionStatusDetail

Describes the execution status of the job flow.

Type: [JobFlowExecutionStatusDetail \(p. 29\)](#)

Instances

Describes the Amazon EC2 instances of the job flow.

Type: [JobFlowInstancesDetail \(p. 31\)](#)

JobFlowId

The job flow identifier.

Type: String

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

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.

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.

Name

The name of the job flow.

Type: String

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

Steps

A list of steps run by the job flow.

Type: [StepDetail \(p. 34\)](#) list

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

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 \(p. 18\)](#) action.

Type: Boolean

JobFlowExecutionStatusDetail

Description

Describes the status of the job flow.

Contents

CreationDateTime

The creation date and time of the job flow.

Type: DateTime

EndDateTime

The completion date and time of the job flow.

Type: DateTime

LastStateChangeReason

Description of the job flow last changed state.

Type: String

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

ReadyDateTime

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

Type: DateTime

StartDateTime

The start date and time of the job flow.

Type: DateTime

State

The state of the job flow.

Type: String

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

JobFlowInstancesConfig

Description

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

Contents

Ec2KeyName

Specifies 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.

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.

HadoopVersion

Specifies 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.

InstanceCount

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

Type: Integer

InstanceGroups

Configuration for the job flow's instance groups.

Type: [InstanceGroupConfig \(p. 25\)](#) list

KeepJobFlowAliveWhenNoSteps

Specifies whether the job flow should terminate after completing all steps.

Type: Boolean

MasterInstanceType

The EC2 instance type of the master node.

Type: String

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

Placement

Specifies the Availability Zone the job flow will run in.

Type: [PlacementType \(p. 33\)](#)

SlaveInstanceType

The EC2 instance type of the slave nodes.

Type: String

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

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

JobFlowInstancesDetail

Description

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

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.

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.

HadoopVersion

Specifies the Hadoop version for the job flow.

Type: String

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

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

InstanceGroups

Details about the job flow's instance groups.

Type: [InstanceGroupDetail \(p. 25\)](#) list

KeepJobFlowAliveWhenNoSteps

Specifies whether or not the job flow should terminate after completing all steps.

Type: Boolean

MasterInstanceId

The Amazon EC2 instance identifier of the master node.

Type: String

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

MasterInstanceType

The Amazon EC2 master node instance type.

Type: String

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

MasterPublicDnsName

The DNS name of the master node.

Type: String

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

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

Placement

Specifies the Amazon EC2 Availability Zone for the job flow.

Type: [PlacementType](#) (p. 33)

SlaveInstanceType

The Amazon EC2 slave node instance type.

Type: String

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

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

KeyValue

Description

A key value pair.

Contents

Key

The unique identifier of a key value pair.

Type: String

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

Value

The value part of the identified key.

Type: String

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

PlacementType

Description

Specifies 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.

RunJobFlowResult

Description

The result of the [RunJobFlow](#) (p. 12) operation.

Contents

JobFlowId

An unique identifier for the job flow.

Type: String

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

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

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.

StepConfig

Description

Specification of a job flow step.

Contents

ActionOnFailure

Specifies the action to take if the job flow step fails.

Type: String

Valid Values: `TERMINATE_JOB_FLOW` | `CANCEL_AND_WAIT` | `CONTINUE`

HadoopJarStep

Specifies the JAR file used for the job flow step.

Type: [HadoopJarStepConfig](#) (p. 24)

Name

The name of the job flow step.

Type: String

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

StepDetail

Description

Combines the execution state and configuration of a step.

Contents

ExecutionStatusDetail

The description of the step status.

Type: [StepExecutionStatusDetail](#) (p. 34)

StepConfig

The step configuration.

Type: [StepConfig](#) (p. 34)

StepExecutionStatusDetail

Description

Specifies the execution state of a step.

Contents

CreationDateTime

The creation date and time of the step.

Type: DateTime

EndDateTime

The completion date and time of the step.

Type: DateTime

LastStateChangeReason

A description of the step's current state.

Type: String

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

StartDateTime

The start date and time of the step.

Type: DateTime

State

The state of the job flow step.

Type: String

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

Common Query Parameters

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

Parameter Name	Description	Required
<i>Action</i>	The action to be performed. Default: None Type: String	Yes
<i>AuthParams</i>	The parameters that are required to authenticate a Conditional request. Contains: AWSAccessKeyId SignatureVersion Timestamp Signature Default: None	Conditional
<i>AWSAccessKeyId</i>	The access key ID that corresponds to the secret access key that you used to sign the request. Default: None Type: String	Yes
<i>Expires</i>	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 <i>Timestamp</i> or <i>Expires</i> , but not both. Default: None Type: String	Conditional

Parameter Name	Description	Required
<i>SecurityToken</i>	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 <i>Using Temporary Security Credentials</i> . Default: None Type: String	
<i>Signature</i>	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	Yes
<i>SignatureMethod</i>	The hash algorithm that you used to create the request signature. Default: None Valid Values: HmacSHA256 HmacSHA1. Type: String	Yes
<i>SignatureVersion</i>	The signature version you use to sign the request. Set this to the value that is recommended for your service. Default: None Type: String	Yes
<i>Timestamp</i>	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 <i>Timestamp</i> or <i>Expires</i> , but not both. Default: None Type: String	Conditional
<i>Version</i>	The API version that the request is written for, expressed in the format YYYY-MM-DD. Default: None Type: String	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.

Error	Description	HTTP Status Code
<code>IncompleteSignature</code>	The request signature does not conform to AWS standards.	400
<code>InternalFailure</code>	The request processing has failed because of an unknown error, exception or failure.	500
<code>InvalidAction</code>	The action or operation requested is invalid. Verify that the action is typed correctly.	400
<code>InvalidClientTokenId</code>	The X.509 certificate or AWS access key ID provided does not exist in our records.	403
<code>InvalidParameterCombination</code>	Parameters that must not be used together were used together.	400
<code>InvalidParameterValue</code>	An invalid or out-of-range value was supplied for the input parameter.	400
<code>InvalidQueryParameter</code>	AWS query string is malformed, does not adhere to AWS standards.	400
<code>MalformedQueryString</code>	The query string contains a syntax error.	404
<code>MissingAction</code>	The request is missing an action or a required parameter.	400
<code>MissingAuthenticationToken</code>	Request must contain either a valid (registered) AWS access key ID or X.509 certificate.	403
<code>MissingParameter</code>	A required parameter for the specified action is not supplied.	400
<code>OptInRequired</code>	The AWS access key ID needs a subscription for the service.	403

Error	Description	HTTP Status Code
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.	400
ServiceUnavailable	The request has failed due to a temporary failure of the server.	503
Throttling	Request was denied due to request throttling.	400