Amazon Elastic MapReduce

API Reference

API Version 2009-03-31

Amazon Elastic MapReduce: API Reference

Table of Contents

Welcome

Topics

This is the *Amazon Elastic MapReduce API Reference*. This section describes who should read this guide, how the guide is organized, and other resources related to this web service.

Amazon Elastic MapReduce, Amazon Elastic Compute Cloud, and Amazon Simple Storage Service are sometimes referred to within this guide as "Elastic MapReduce," "EC2," and "Amazon S3," respectively. All copyrights and legal protections still apply.

Who Should Read this Guide

This reference is for developers and for the community of researchers and data analysts that need to process vast amounts of data efficiently and cost-effectively.

Required Knowledge and Skills

Elastic MapReduce offers a console, a command line interface, and an API. Each interface requires a slightly different set of skills. The console and command line interface let you execute Elastic MapReduce functionality without having to write full applications. Both of these interfaces, however, require you to write your own mapper and reducer executables in the programming language of your choice.

Of the three interfaces, the API requires the greatest programming skills and knowledge of Hadoop. Even though Elastic MapReduce hides a lot of the Hadoop configuration details, knowing the basics of Hadoop is important.

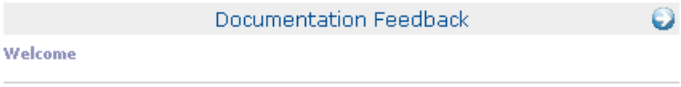
Developers using the API should be familiar with the following:

- XML (for an overview, go to the W3 Schools XML Tutorial)
- Basic understanding of web services (for an overview, go to the W3 Schools Web Services Tutorial)
- A programming language for writing mapper and reducer executables and for consuming Elastic MapReduce responses

- Hadoop (for more information, go to http://hadoop.apache.org/core/)
- Familiarity with EC2 and Amazon S3 (for more information, see the Amazon Elastic Compute Cloud Developer Guide and the Amazon Simple Storage Service Developer Guide, respectively)

How to Give Us Feedback

The online version of this guide provides a link at the top of each page that enables you to enter feedback about this guide. We strive to make our guides as complete, error free, and easy to read as possible. You can help by giving us feedback. Thank you in advance!



Amazon Elastic MapReduce Resources

The following table lists related resources that you'll find useful as you work with this service.

Resource	Description
Amazon Elastic MapReduce Developer Guide	The developer guide provides detailed descriptions of how to perform common tasks using the Amazon Elastic MapReduce API.
Amazon Elastic MapReduce Getting Started Guide	The Getting Started Guide provides a quick tutorial of the service based on a simple use case. Examples and instructions for the console are included.
Amazon Elastic MapReduce Technical FAQ	The FAQ covers the top 20 questions developers have asked about this product.
Release notes	The release notes give a high-level overview of the current release. They specifically note any new features, corrections, and known issues.
AWS Developer Resource Center	A central starting point to find documentation, code samples, release notes, and other information to help you build innovative applications with AWS.
AWS Management Console	Location of the Amazon Elastic MapReduce console.
Discussion Forums	A community-based forum for developers to discuss technical questions related to Amazon Web Services.
AWS Support Center	The home page for AWS Technical Support, including access to our Developer Forums, Technical FAQs, Service Status page, and Premium Support .
Premium Support	The primary web page for information about AWS Premium Support, a one-on-one, fast-response support channel to help you build and run applications on AWS Infrastructure Services.
Amazon Elastic MapReduce product information	The primary web page for information about Amazon Elastic MapReduce.

Resource	Description
Contact Us	A central contact point for inquiries concerning AWS billing, account, events, abuse etc.
Conditions of Use	Detailed information about the copyright and trademark usage at Amazon.com and other topics.

What's New

This What's New is associated with the 2009-03-31 version of the Amazon Elastic MapReduce web service. This guide was last updated on 2009-12-02.

Change	Description	Release Date
Technical documents reorganized	The API reference has been split out of the *Amazon Elastic MapReduce Developer Guide*. Now, on the documentation landing page, http://developer.amazonwebservices.com/connect/entry.jspa?externalID=2738&categoryID=48, you can select the document you want to view. When viewing the documents online, the links in one document will take you, when appropriate, to one of the other guides.	16 September 2009
Enhancement	Added a new endpoint for the European community. For the new WSDL location, see WSDL Locations (p. 5). For information about configuring your region in the Management Console, go to Creating a Job. For information about configuring your region in the Command Line Interface, go to the steps for editing the credentials.json file.	7 July 2009
New guide	This is the first publication of this guide.	2 April 2009

WSDL Locations

The Amazon Elastic MapReduce WSDL is available by locale.

Locale	URL
US	http://elasticmapreduce.amazonaws.com/doc/2009-03-31/ElasticMapReduce.wsdl
EU	http://eu-west-1.elasticmapreduce.amazonaws.com/doc/2009-03-31/ElasticMapReduce.wsdl

Common Parameters

Each operation in the API has its own set of input parameters. There is, however, a set of parameters that all operations use. This section describes those parameters.

The following table describes parameters that must be used in all requests.

Parameter	Description	Required
AWSAccessKeyId	A string, distributed by Amazon when you sign up to be a developer, which uniquely identifies the developer. For more information about displaying your *`AWSAccessKeyId`* and *`SecretKey`*, see the *Amazon Elastic MapReduce Getting Started Guide*. Type: String Default: None	Yes
Operation	The API operation, for example, `RunJobFlow`. Type: String: Default: None Constraint: Any operation in the API	Yes
Signature	A value calculated using the request parameters and a SHA-1 HMAC encryption algorithm to make sure the request parameters and values were not altered during the request's or response's travel across the Internet. For more information about creating a signature, see How to Create Signatures . Type: String Default: None	Yes
SignatureVersion	A value that specifies the *`Signature`* format. Type: Integer Default: None Valid Values: 1	Yes
Timestamp	A date-time value that marks the day and time the request was sent. Requests expire after a certain length of time. This characteristic prevents mischievous users from capturing requests and resubmitting them at a later time. Type: dateTime, for example, 2008-09-18T13:00:01Z Default: None	Yes

Parameter	Description	Required
Version	The version number of the WSDL to use in processing the request. Version numbers are dates, such as 2009-01-15. For a list of version numbers, go to the Amazon Resource Center at http://aws.amazon.com/resources. Type: String Default: None	Yes

Common Errors

Topics

Amazon Elastic MapReduce error results provide information about syntactical errors in your requests, as well as errors that occur during the execution of your request. If a request generates more than one error, only the first error is reported.

Query XML Error Format

Errors are composed of the following elements.

- **Type—**Specifies whether the source of the error is the request or the service
  If the source is the request, the value of *Type* is `Sender`, otherwise it's `Receiver`.
- **Code—**String that identifies the error
- **Message—**Short description of the error
- **Details—**Additional information describing the error

The error response also include a request ID so that you can match the error to the original request. To make that association, you must keep a record of your request IDs.

The following snippet shows a typical error returned from a Query request.

```
<ErrorResponse xmlns="http://elasticmapreduce.amazonaws.com/doc/2009-01-15">
  <Error>
    <Type>Sender</Type>
    <Code>ValidationError</Code>
    <Message>Job flow name must not be more than 256 bytes.</Message>
  </Error>
  <RequestId>1455d872-db80-11dd-a60e-d7c9208138e9</RequestId>
  </ErrorResponse>
```

For more information, see the data types Error (p. 28) and ErrorResponse (p. 28).

HTTP Errors and Responses

Amazon Elastic MapReduce returns the following HTTP response codes in addition to the XML errors codes described previously. The HTTP errors fall into the following categories.

- **2XX—**The request completed successfully.
- **4XX—**The web server didn't understand or couldn't process the request, the request was forbidden, or the request timed out.
  These errors do not indicate a problem with Amazon Elastic MapReduce. So, upon receiving this error, fix the request and resubmit it.
- **5XX—**Internal server error.
  These errors reflect problems with the underlying Amazon Elastic MapReduce web service. You have to wait until the web service is functioning before resubmitting the request. If errors persist, use the forum to contact us.

Amazon Elastic MapReduce Errors

All Amazon Elastic MapReduce operations can return the errors in the following table.

Error Code	Description
OptInRequired	You must subscribe to the Amazon Elastic MapReduce web service in order to submit requests to Amazon Elastic MapReduce. See the *Amazon Elastic MapReduce Getting Started Guide* to learn how to subscribe to Amazon Elastic MapReduce.
ValidationError	One or more of the parameters in your request are in error. Make sure the parameter values are valid, that the required parameters are included in the request, and that you have not misspelled or mis-capitalized the parameter names.
ValidationError	Job flow name must not be more than 256 bytes. Reduce the length of the name.
ValidationError	*EC2KeyName* must not be more than 256 bytes. Reduce the length of the key name.
ValidationError	Job flow step name must not be more than 256 bytes. Reduce the length of the job flow name.
ValidationError	Instance count must be greater than 0 Make the value 1 or greater.
ValidationError	Instance type must be among m1.small, m1.large, m1.xlarge, c1.medium, c1.xlarge. Revise entry to a valid value.
ValidationError	Specified *JobFlowId* does not exist. Revise the job flow ID.
ValidationError	Specified *JobFlowId* not valid. Revise the job flow ID.
ValidationError	A step is missing JAR information. Supply the missing step.
ValidationError	The limit for the maximum number of steps in a job flow has been reached. Reduce the length of the number of steps.

Error Code	Description
ValidationError	The given job flow state is not known. Revise the job flow state.
ValidationError	Can't modify job flow, which is shutting or has shut down There is no workaround.
ValidationError	*LogUri* is not in the specified format or the bucket name, key suffix is too long. The format must be `s3n://[bucketName]/[optionalPath]`. For more information on the naming format and maximum bucket size, see the *Amazon Simple Storage Service Developer Guide*.
ValidationError	Availability zone must be one of the following: us-east-1a,us-east-1b,us-east-1c Make the entry valid.

Operations

Topics

This section describes in detail the Elastic MapReduce operations.

AddJobFlowSteps

Description

`AddJobFlowSteps` adds new steps to a job flow already loaded on an EC2 cluster. Each step applies an algorithm to the data set, for the first step, or to the data returned by the previous step in the job flow. If the job flow isn't executing any other steps, execution begins from the first added step. The maximum number of steps in a job flow is 256.

For more information about adding steps to a job flow, see Adding Steps to a Job Flow.

Request Parameters

Parameter	Definition	Required
ActionOnFailure	Specifies the action to take on a failure. TERMINATE_JOB_FLOW stops the processing and terminates the EC2 cluster. CANCEL_AND_WAIT stops the processing but does not terminate the EC2 cluster. CONTINUE does not stop the processing. Type: ActionOnFailure (p. 27) Default: TERMINATE_JOB_FLOW Valid Values: TERMINATE_JOB_FLOW | CANCEL_AND_WAIT | CONTINUE Ancestry: Steps (p. 36).member (p. 33)	Yes
Args	Space-separated list of strings that serve as arguments for the JAR. Type: StringList (p. 36) Default: None Ancestry: Steps (p. 36).member (p. 33).HadoopJarStep (p. 28)	Yes
JobFlowId	String that uniquely identifies a job flow. Elastic MapReduce returns this value in the `RunJobFlow` response. Type: String Default: None Constraint: A valid, encrypted job flow ID	Yes
Jar	Filename of the JAR. Type: String Default: None Ancestry: Steps (p. 36).member (p. 33).HadoopJarStep (p. 28)	Yes
Key	The name of a parameter in a Hadoop step. This parameter is paired with the *`Value`* parameter. See *`Value`*. Type: String Default: None Ancestry: Steps (p. 36).member (p. 33).HadoopJarStep (p. 28).Properties (p. 33).KeyValueList (p. 33).KeyValue (p. 32)	No

Parameter	Definition	Required
MainClass	Location of the JAR's main class. Type: String Default: None Ancestry: Steps (p. 36).member (p. 33).HadoopJarStep (p. 28).Properties (p. 33)	No
Member	An argument. Type: String Default: None Ancestry: Steps (p. 36).HadoopJarStep (p. 28).Args.StringList (p. 36)	Yes
Name	Name of the step you are adding. Type: String Default: None Ancestry: Steps (p. 36).member (p. 33)	Yes
Value	Value of a parameter. This parameter is paired with the *Key* parameter. See *Key*. Type: String Default: None Ancestry: Steps (p. 36).member (p. 33).HadoopJarStep (p. 28).Properties (p. 33).KeyValueList (p. 33).KeyValue (p. 32)	No

You must also use parameters that are common to all requests that are described in Common Request Parameters (p. 6). Parameter names are case sensitive.

Response Elements

Element	Description
RequestId	String that uniquely identifies the `AddJobFlowSteps` request. If you have a problem with Amazon Elastic MapReduce, include the request ID in your correspondence to Amazon Web Services. Type: String Ancestry: ResponseMetadata (p. 34)

Special Errors

There are no special errors for this operation.

Examples

This section shows a sample request and corresponding response.

Sample Request

In this example, we use `AddJobFlowSteps`, for example, to add steps to a specified job flow ID, as follows.

```
https://elasticmapreduce.amazonaws.com?
JobFlowId=j-3UN6WX5RRO2AG&
Steps.member.1.Name=MyStep2&
Steps.member.1.ActionOnFailure=CONTINUE&
Steps.member.1.HadoopJarStep.Jar=MySecondJar&
Steps.member.1.HadoopJarStep.MainClass=MainClass&
Steps.member.1.HadoopJarStep.Args.member.1=arg1&
Operation=AddJobFlowSteps&
AWSAccessKeyId=[AWS Access Key ID]&
SignatureVersion=2&
SignatureMethod=HmacSHA256&
Timestamp=2009-01-28T21%3A51%3A51.000Z&
Signature=[calculated value]
```

Sample Response

The following XML is a snippet of the full response to the first sample request.

```
<AddJobFlowStepsResponse xmlns="http://elasticmapreduce.amazonaws.com/
doc/2009-01-15">
  <ResponseMetadata>
    <RequestId>df6f4f4a-ed85-11dd-9877-6fad448a8419</RequestId>
  </ResponseMetadata>
</AddJobFlowStepsResponse>
```

Related Operations

See also:

- RunJobFlow (p. 19)

DescribeJobFlows

Description

`DescribeJobFlows` returns extensive details about specified job flows. You specify job flows by their ID, creation date, or state. Elastic MapReduce returns descriptions of job flows that are up to two months old. Specifying an older date returns an error. The maximum number of job flow descriptions returned is 512.

Each input parameter acts as a filter so that Elastic MapReduce returns information about a more precise set of job flows with each parameter you use in your request. If you do not include parameters in a request, Elastic MapReduce returns descriptions of all job flows that have:

- Created and completed in the last two weeks
- Not ended within the last two months
  These jobs are in one of the following states: RUNNING, WAITING, SHUTTING_ DOWN, STARTING.

For more information about describing job flows, see Describing a Job Flow .

Request Parameters

Parameter	Description	Required
CreatedAfter	Returns descriptions of job flows created after this date. Type: dateTime (p. 27) (yyyy-mm-ddThh:mm:ss) Default: Two months ago Constraint: Two months ago or more recent	No
CreatedBefore	Returns descriptions of job flows created before this date. Type: dateTime (p. 27) (yyyy-mm-ddThh:mm:ss) Default: Now Constraint: Before now and after two months ago	No
JobFlowIds	Returns descriptions of job flows specified by the job flow IDs. These values are in the `RunJobFlow` response. The ID uniquely identifies the job flow. Type: StringList (p. 36) Default: None Constraint: A job flow ID returned by `RunJobFlow`	Yes
JobFlowStates	Returns descriptions of job flows specified by the job flow IDs. These values are in the `RunJobFlow` response. The ID uniquely identifies the job flow. For more information, see Job Flow States Type: JobFlowExecutionStateList (p. 30) Default: None Valid Values: One or more of the following: COMPLETED | FAILED | TERMINATED | RUNNING | SHUTTING_DOWN | STARTING | WAITING	Yes

You must also use parameters that are common to all requests that are described in Common Request Parameters (p. 6). Parameter names are case sensitive.

Response Elements

Element	Description
CreationDateTime	Specifies the time the job flow was created. Type: dateTime (p. 27) (yyyy-mm-ddThh:mm:ss)
Ec2KeyName	A KeyName maps to a public key-private key pair, which is required to launch the master node. The private key is used by SSH to authenticate the tunneling connection between the developer and the EC2 master node that was launched using the KeyName. For more information, see Using SSH to Monitor Job flow Status. Type: String
EndDateTime	Specifies the time the job flow ended. Type: dateTime (p. 27) (yyyy-mm-ddThh:mm:ss)
Id	Specifies the job flow ID, which is returned by `RunJobFlow`. Type: String
InstanceCount	Specifies the number of EC2 instances in the cluster. Type: String
JobFlowId	Specifies the string that uniquely identifies a job flow. The value is returned by `createJob`. Type: string
KeepJobAliveWhenNoSteps	Specifies whether or not to terminate the job flow when there are no more steps to complete. Type: Boolean
LogURI	The location on Amazon S3 where Elastic MapReduce stores the job flow's log files Type: String
MasterInstanceType	Specifies the master EC2 instance machine type. Type: String
MasterPublicDnsName	DNS name of master node. Type: String
MasterInstanceId	ID that uniquely identifies the master node instance. Type: String
Name	Specifies the name of the parameter required by the step type, or the step or job flow Type: String
Placement	Container for PlacementType, which describes the geographic location of the EC2 instances. Type: PlacementType (p. 33)
RequestId	Specifies the string that uniquely identifies the `DescribeJobFlows` request. If you have a problem with Amazon Elastic MapReduce, include the request ID in your correspondence to Amazon Web Services. Type: String

Element	Description
SlaveInstanceType	Specifies machine type hosting the slave EC2 instances. Type: String
State	Specifies the execution state of the step. For more information, see Step States. Type: JobFlowExecutionState Valid Values: COMPLETED | FAILED | TERMINATED | RUNNING | SHUTTING_DOWN | STARTING | WAITING
Type	Specifies the step type. Type: String Valid Values: elasticmapreduce.s3get, elasticmapreduce.s3put, elasticmapreduce.logParser
Value	Specifies the value of the parameter required by the step type Type: String

Special Errors

There are no special errors for this operation.

Examples

This section shows a sample request and corresponding response.

Sample Request

In this example, we use `DescribeJobFlows` to describe job flow j-3UN6WX5RRO2AG, as follows.

```
https://elasticmapreduce.amazonaws.com?
JobFlowIds.member.1=j-3UN6WX5RRO2AG&
Operation=DescribeJobFlows&
AWSAccessKeyId=[AWS Access Key ID]&
SignatureVersion=2&
SignatureMethod=HmacSHA256&
Timestamp=2009-01-28T21%3A49%3A59.000Z&
Signature=[calculated value]
```

Sample Response

The following XML is a snippet of the full response to the sample request.

```
<DescribeJobFlowsResponse xmlns="http://elasticmapreduce.amazonaws.com/
doc/2009-01-15">
  <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>mybucket/subdir/</LogUri>
```

```
        <Steps>
          <member>
            <ExecutionStatusDetail>
              <CreationDateTime>2009-01-28T21:49:16Z</CreationDateTime>
              <State>PENDING</State>
            </ExecutionStatusDetail>
            <StepConfig>
              <HadoopJarStep>
                <Jar>MyJarFile</Jar>
                <MainClass>MyMailClass</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>
        </Instances>
      </member>
    </JobFlows>
  </DescribeJobFlowsResult>
  <ResponseMetadata>
    <RequestId>9cea3229-ed85-11dd-9877-6fad448a8419</RequestId>
  </ResponseMetadata>
</DescribeJobFlowsResponse>
```

Related Operations

See also:

- RunJobFlow (p. 19)

RunJobFlow

Description

`RunJobFlow` creates a new job flow, an EC2 cluster, and executes the job flow steps on the cluster. When the job flow finishes, depending on the parameter values, `RunJobFlow` terminates the EC2 cluster and uploads results to a specified Amazon S3 bucket.

The maximum lifetime of a job flow is 2 weeks. The maximum number of steps allowed in a job flow is 256. For more information about running job flows, see Starting Job Flow.

Request Parameters

Parameter	Definition	Required
ActionOnFailure	Specifies the action to take on a failure. TERMINATE_JOB_FLOW stops the processing and terminates the EC2 cluster. CANCEL_AND_WAIT stops the processing but does not terminate the EC2 cluster. CONTINUE does not stop the processing. Type: ActionOnFailure (p. 27) Default: TERMINATE_JOB_FLOW Valid Values: TERMINATE_JOB_FLOW | CANCEL_AND_WAIT | CONTINUE Ancestry: Steps (p. 36).member (p. 33)	No
Args	A list of arguments used by the steps in the job flow. This is where you specify the input and output locations for your data. For more information, see Args Argument . Type: StringList (p. 36) Default: None Ancestry: Steps (p. 36).HadoopJarStep (p. 28).Properties	Yes
AvailabilityZone	Specifies the geographic location of the EC2 cluster. We discourage the use of this parameter. When *`AvailabilityZone`* is not specified we automatically pick the best *`AvailabilityZone`* for the job flow. However, you might find this parameter useful if you want to co-locate your instances with other existing running instances, and your job flow needs to read or write data from those instances. For more information, see Availability Zones . Type: String Default: None Valid Values: us-east-1a | us-east-1b | us-east-1c | eu-west-1a | eu-west-1b Ancestry: Instances (p. 29).Placement (p. 33).PlacementType (p. 33)	No

Parameter	Definition	Required
Ec2KeyName	A KeyName maps to a public key-private key pair, which is required to launch the master node. The private key is used by SSH to authenticate the tunneling connection between the developer and the EC2 master node that was launched using the KeyName. For more information, see Using SSH to Monitor Job flow Status. Type: String Default: None Constraint: 1 to 256 characters Ancestry: Instances (p. 29)	No
KeepJobFlowAlive WhenNoSteps	Specifies whether (`True`) or not (`False`) to keep the EC2 cluster engaged after all steps in the job flow complete. Only set this value to `True` while you are debugging the workflow. Type: Boolean Default: None Valid Values: True | False Ancestry: Instances (p. 29)	Yes
InstanceCount	The number of EC2 instances to use in the cluster. If the value is 1, one instance serves as the master and slave node. If the value is greater than one, one instance is the master node and the remainder are slave nodes. Type: Integer Default: None Valid Values: 1 to 20. 20 is the maximum number per account. For example, if you have two job flows running, the total number of instances running for both job flows must be 20 or less. If you need more than 20 instances, you must submit a special request. For more information, go to Request Form. Ancestry: Instances (p. 29)	Yes
Jar	Compressed file that contains the executables for steps. Type: String Default: None Ancestry: Steps (p. 36).HadoopJarStep (p. 28).Properties	Yes
Key	Name of a parameter that is paired with a value. See Value. Type: String Default: None Ancestry: Steps (p. 36).HadoopJarStep (p. 28).Properties	Yes
LogUri	URI for the log files. If you do not provide a value, logs are not generated. Type: String Default: No logs are generated Valid Form: <bucket_name>/<key_prefix> Constraint: *`bucket_name`* must be less than 256 characters and *`key_prefix`* must be less than 1025 characters.	Yes

Parameter	Definition	Required
MainClass	Location of the Main class. Type: String Default: None Ancestry: Steps (p. 36).HadoopJarStep (p. 28).Properties	Yes
MasterInstanceType	The instance type of the master instance in the EC2 cluster Type: String Default: None Valid Values: m1.small | m1.large | m1.xlarge | c1.medium | c1.xlarge Ancestry: Instances (p. 29)	Yes
Member	An argument Type: String Default: None Ancestry: Steps (p. 36).HadoopJarStep (p. 28).Args (p. 27).StringList (p. 36)	Yes
Name	Name of the job flow. Type: String, Default: None Constraint: 1 to 255 characters	Yes
Name	Name of a step. Type: String, 1 to 255 characters Default: None Constraint: 1 to 255 characters Ancestry: Steps (p. 36)	No
SlaveInstanceType	The instance type of the slave instance in the EC2 cluster Type: String Default: None Valid Values: m1.small, m1.large, m1.xlarge, c1.medium, c1.xlarge Ancestry: Instances (p. 29)	Yes
Value	Value of a parameter that is paired with a key. See Key. Type: String Default: None Ancestry: Steps (p. 36).HadoopJarStep (p. 28).Properties	No

You must also use parameters that are common to all requests that are described in Common Request Parameters (p. 6). Parameter names are case sensitive.

Response Elements

Element	Description
JobFlowId	String, returned by `RunJobFlow`, which uniquely identifies the job flow created. Type: String Ancestry: RunJobFlowResult

Element	Description
RequestId	String that uniquely identifies the `RunJobFlow` request. If you have a problem with Amazon Elastic MapReduce, include the request ID in your correspondence to Amazon Web Services. Type: String Ancestry: ResponseMetadata

Special Errors

There are no special errors for this operation.

Examples

This section shows a sample request and corresponding response.

Sample Request

In this example, we use `RunJobFlow` to start a job flow, as follows.

```
https://elasticmapreduce.amazonaws.com?
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&
Steps.member.1.Name=MyStepName&
Steps.member.1.ActionOnFailure=CONTINUE&
Steps.member.1.HadoopJarStep.Jar=MyJarFile&
Steps.member.1.HadoopJarStep.MainClass=MyMailClass&
Steps.member.1.HadoopJarStep.Args.member.1=arg1&
Steps.member.1.HadoopJarStep.Args.member.2=arg2&
Operation=RunJobFlow&ContentType=JSON&
AWSAccessKeyId=[AWS Access Key ID]&
SignatureVersion=2&
SignatureMethod=HmacSHA256&
Timestamp=2009-01-28T21%3A48%3A32.000Z&
Signature=[calculated value]
```

Sample Response

The following XML is a snippet of the full response to the first sample request.

```
<RunJobFlowResponse xmlns="http://elasticmapreduce.amazonaws.com/
doc/2009-01-15">
  <RunJobFlowResult>
    <JobFlowId>j-3UN6WX5RRO2AG</JobFlowId>
  </RunJobFlowResult>
  <ResponseMetadata>
    <RequestId>8296d8b8-ed85-11dd-9877-6fad448a8419</RequestId>
  </ResponseMetadata>
</RunJobFlowResponse>
```

Related Operations

See also:

- TerminateJobFlows (p. 24)

TerminateJobFlows

Description

`TerminateJobFlows` terminates job flow processing, uploads data from EC2 to Amazon S3, and terminates the EC2 cluster. You can use this action to terminate a single job flow or a list of job flows. Job flows that complete successfully terminate automatically unless you set the `RunJobFlows` parameter, *`KeepJobFlowAliveWhenNoSteps`*, to `True`. Job flows also terminate automatically on failure unless you set the `RunJobFlows` parameter, *`KeepJobFlowAliveWhenNoSteps`*, to `CANCEL_AND_WAIT` or `CONTINUE`.

For more information about terminating job flows, see Terminating a Job Flow .

Request Parameters

Parameter	Description	Required
JobFlowIds	One or more job flow IDs that specify the job flows to terminate. Each *`JobFlowId`* is a string returned by `RunJobFlow`. To terminate more than one job flow, use a comma-separated list of *`JobFlowIds`*. Type: StringList (p. 36) Default: None Constraint: Valid job flow ID	Yes

You must also use parameters that are common to all requests that are described in Common Request Parameters (p. 6). Parameter names are case sensitive.

Response Elements

Element	Description
RequestId	String that uniquely identifies the `TerminateJobFlows` request. If you have a problem with Amazon Elastic MapReduce, include the request ID in your correspondence to Amazon Web Services. Type: String Ancestry: ResponseMetadata (p. 34)

Special Errors

There are no special errors for this operation.

Examples

This section shows a sample request and corresponding response.

Sample Request

In this example, we use `TerminateJobFlows` to describe all job flows created on a specified date, as follows.

```
https://elasticmapreduce.amazonaws.com?
```

```
JobFlowIds.member.1=j-3UN6WX5RRO2AG&
Operation=TerminateJobFlows&
AWSAccessKeyId=[AWS Access Key ID]&
SignatureVersion=2&
SignatureMethod=HmacSHA256&
Timestamp=2009-01-28T21%3A53%3A50.000Z&
Signature=[calculated value]
```

Sample Response

The following XML is a snippet of the full response to the first sample request.

```
<TerminateJobFlowsResponse xmlns="http://elasticmapreduce.amazonaws.com/
doc/2009-01-15">
  <ResponseMetadata>
    <RequestId>2690d7eb-ed86-11dd-9877-6fad448a8419</RequestId>
  </ResponseMetadata>
  </TerminateJobFlowsResponse>
```

Related Operations

See also:

- RunJobFlow (p. 19)

DataTypes

Topics

This section describes all of the data types in Elastic MapReduce

ActionOnFailure

Element	Description
ActionOnFailure	Specifies what happens to the Amazon EC2 cluster when there is an error in the job flow. TERMINATE_JOB_FLOW cancels the job flow | stops all processing | and terminates the Amazon EC2 cluster. CANCEL_AND_WAIT stops process but does not terminate the cluster. CONTINUE proceeds with the job flow regardless of the error. Type: Enum Default: TERMINATE_JOB_FLOW Valid Values: TERMINATE_JOB_FLOW | CANCEL_AND_WAIT | CONTINUE

Args

Element	Description
Args	Container for information about arguments for the jar. Type: StringList (p. 36) Default: None

AvailabilityZone

Element	Description
AvailabilityZone	Geographical localtion of the Amazon EC2 cluster. Type: String Default: None

dateTime

Element	Description
dateTime	Date and time in the form: yyyy-mm-ddThh:mm:ss. Type: dateTime Default: None

Error

Element	Description
Code	A term that uniquely identifies an error. Type: String Default: None
Message	An explanation of the error. Type: String Default: None
Type	Identifies the sender of the request or the web service (receiver) as the source of the error. Type: String Default: None Valid Values: Receiver, Sender

ErrorResponse

Element	Description
RequestId	ID that uniquely identifies a request. Amazon keeps track of request IDs. If you have a question about a request, include the request ID in your correspondence. Type: String Default: None

HadoopJarStep

Element	Description
HadoopJarStep	Container for HadoopJarStepConfig. Type: HadoopJarStepConfig (p. 28) Default: None

HadoopJarStepConfig

Element	Description
Args	Arguments for the job flow's steps. Type: StringList (p. 36) Default: None
Jar	Java file that contains the step information. Type: String Default: None

Element	Description
MainClass	Name of the main class in the jar. Type: String Default: None
Properties	Container for KeyValueList. Type: KeyValueList (p. 33) Default: None

Instances

Element	Description
Instances	Container for information about EC2 instances. Type: JobFlowInstancesDetail (p. 31) Default: None

JobFlowDetail

Element	Description
JobFlowId	ID that uniquely identifies the job flow. Type: String Default: None
Name	Name of the job flow. Type: String Default: None
LogUri	URI for the location of the log files on Amazon S3. Type: String Default: None Constraint: Amazon S3 URIs are of the form: `bucketName]/[optionalPathToObject]`
ExecutionStatusDetail	Container for the description of the execution status. Type: JobFlowExecutionStatusDetail (p. 30) Default: None
Instances	Container for JobFlowInstancesDetail. Type: JobFlowInstancesDetail (p. 31) Default: None
Steps	A container for one or more step(s). A step is a single function applied to the data in a job flow. Type: Steps (p. 36) Default: None

JobFlowDetailList

Element	Description
JobFlowDetailList	Container for JobFlowDetail. Type: JobFlowDetail (p. 29) Default: None

JobFlowExecutionState

Element	Description
JobFlowExecutionState	The state of the job flow's execution. Type: Enum Default: None Valid Values: COMPLETED | FAILED | TERMINATED | RUNNING | SHUTTING_DOWN | STARTING | WAITING

JobFlowExecutionStateList

Element	Description
member	An array of JobFlowExecutionState. Type: JobFlowExecutionState (p. 30) Default: None

JobFlowExecutionStatusDetail

Element	Description
State	State of the job flow. Type: JobFlowExecutionState (p. 30) Default: None
CreationDateTime	Date and time the job flow was created. Type: dateTime (p. 27) Default: None
EndDateTime	Date and time the job flow ended. Type: dateTime (p. 27) Default: None
StartDateTime	Date and time the job flow started. Type: dateTime (p. 27) Default: None
LastStateChangeReason	Explains the last change of Amazon EC2 status. Type: StateChangeReason (p. 34) Default: None

JobFlowInstancesConfig

Element	Description
Ec2KeyName	Container for JobFlowInstancesConfigEc2KeyName. Type: JobFlowInstancesConfigEc2KeyName (p. 31) Default: None
Placement	Container for PlacementType. Type: PlacementType (p. 33) Default: None
MasterInstanceType	EC2 instance type of the master node. Type: String Default: None
SlaveInstanceType	EC2 instance type of the slave nodes. Type: String Default: None
InstanceCount	Number of EC2 instances in the cluster. If the value is 1, one instance serves as the master and slave node. If the value is greater than one, one instance is the master node and the remainder are slave nodes. Type: Integer Default: None
KeepJobFlowAliveWhenNoSteps	Specifies whether (true) or not (false) to keep the EC2 cluster engaged after all steps in the job flow complete. Type: Boolean Default: None

JobFlowInstancesConfigEc2KeyName

Element	Description
JobFlowInstancesConfigEc2KeyName	Keyname for master node. Type: String Constraint: 1 to 256 characters Default: None

JobFlowInstancesDetail

Element	Description
MasterInstanceType	Amazon EC2 master node instance type. Type: String Default: None
MasterPublicDnsName	DNS name of master node. Type: String Default: None

Element	Description
MasterInstanceId	ID that uniquely identifies the master node instance. Type: String Default: None
SlaveInstanceType	Amazon EC2 slave node instance type. Type: String Default: None
InstanceCount	Number of Amazon EC2 instances in the cluster. If the value is 1, one instance serves as the master and slave node. If the value is greater than one, one instance is the master node and the remainder are slave nodes. Type: Integer Default: None
Ec2KeyName	A KeyName maps to a public key-private key pair, which is required to launch the master node. The private key is used by SSH to authenticate the tunneling connection between the developer and the EC2 master node that was launched using the KeyName. For more information, see Using SSH to Monitor Job flow Status.. Type: String Default: None
Placement	Container for PlacementType. Type: PlacementType (p. 33) Default: None
KeepJobFlowAliveWhenNoSteps	Specifies whether (true) or not (false) to keep the EC2 cluster engaged after all steps in the job flow complete. Type: Boolean Default: None

KeyValue

Element	Description
Key	Name of a parameter that is paired with a value in a KeyValue pair. Type: String Default: None
Value	Value of a parameter that is paired with a key in a KeyValue pair. Type: String Default: None

KeyValueList

Element	Description
KeyValueList	Container for KeyValue. Type: KeyValue (p. 32) Default: None

Member

Element	Description
Member	Helps associate parameters related to a single step. Type: JobFlowDetail (p. 29) Default: None

Placement

Element	Description
Placement	Container for information about the placement of the Amazon EC2 instances. Type: PlacementType (p. 33) Default: None

PlacementType

Element	Description
AvailabilityZone	Geographical location of the Amazon EC2 instances. Type: String Default: None

Properties

Element	Description
Properties	Container for information about properties. Type: KeyValueList (p. 33) Default: None

RequestId

Element	Description
RequestId	ID that uniquely identifies a request. Amazon keeps track of request IDs. If you have a question about a request, include the request ID in your correspondence. Type: String Default: None

ResponseMetadata

Element	Description
RequestId	ID that uniquely identifies a request to Elastic MapReduce. Use this ID as a reference when you have questions about Elastic MapReduce. errors. Type: String Default: None

RunJobFlowInputName

Element	Description
RunJobFlowInputName	Name of the job flow.. Type: String Default: None Constraint: 1 to 256 characters

StateChangeReason

Element	Description
StateChangeReason	Describes what caused the master node to change state. Type: Enum Default: None Valid Values: MASTER_TERMINATED | START_FAILED | STEP_FAILED

StepConfig

Element	Description
Name	Name of the step. Type: StepConfigName (p. 35) Default: None

Element	Description
ActionOnFailure	Specifies what happens to the Amazon EC2 cluster when the job flow results in an error. Type: ActionOnFailure (p. 27) Default: None
HadoopJarStep	Specifies a Hadoop step. Type: HadoopJarStepConfig (p. 28) Default: None

StepConfigList

Element	Description
StepConfigList	Container for StepConfig. Type: StepConfig (p. 34) Default: None

StepConfigName

Element	Description
StepConfigName	Name of the step. Type: String Default: None Constraint: 1 to 256 characters

StepDetail

Element	Description
ExecutionStatusDetail	Description of execution status. Type: StepExecutionStatusDetail (p. 36) Default: None
StepConfig	Step configuration. Type: StepConfig (p. 34) Default: None

StepDetailList

Element	Description
StepDetailList	Container for StepDetail. Type: StepDetail (p. 35) Default: None

StepExecutionState

Element	Description
StepExecutionState	Status of the step's execution. Type: Enum Default: None Valid Values: PENDING | RUNNING | COMPLETED | CANCELLED | FAILED | INTERRUPTED

StepExecutionStatusDetail

Element	Description
State	State of the step. Type: stepFlowExecutionState (p. 30) Default: None
CreationDateTime	Date and time the step was created. Type: dateTime (p. 27) Default: None
EndDateTime	Date and time the step ended. Type: dateTime (p. 27) Default: None
StartDateTime	Date and time the step started. Type: dateTime (p. 27) Default: None
LastStateChangeReason	Explains the last change of Amazon EC2 status. Type: StateChangeReason (p. 34) Default: None

Steps

Element	Description
Steps	Container for information about steps. Type: StepDetailList (p. 35) Default: None

StringList

Element	Description
member	One or more strings. Contains information about a step, for example, `Steps.member.1.value=1`. Type: Array of strings Default: None

Glossary

AMI	Amazon Machine Instance An AMI is the equivalent of one computer in EC2. AMIs come in a variety of configurations based on the number of processors and cache size. Elastic MapReduce automatically determines the right number of AMIs and the right type of AMIs for a job flow. You can, however, use the API to override that configuration. For more information, see the *Amazon Elastic Compute Cloud Developer Guide.*
authentication	The process of proving your identity to the system.
AWS Access Key ID	A string distributed by AWS that uniquely identifies an AWS developer. The value of this ID is included in every Amazon Elastic MapReduce request.
block	A data set. Amazon Elastic MapReduce breaks large amounts of data into subsets. Each subset is called a data block. Amazon Elastic MapReduce assigns an ID to each block and uses a hash table to keep track of block processing.
bucket	A container for objects stored in Amazon Elastic MapReduce. Every object is contained within a bucket. For example, if the object named `photos/puppy.jpg` is stored in the `johnsmith` bucket, then it is addressable using the URL `http://johnsmith/S3.amazonaws.com/photos/puppy.jpg`
endpoint	The URI for Amazon Elastic MapReduce web service. The endpoint for all Amazon Elastic MapReduce requests is `https://elasticmapreduce.amazon.com.`
HMAC	Hash Message Authentication Code that is used to authenticate a message The HMAC is calculated using a standard, hash cryptographic algorithm, such as SHA-1. This algorithm uses a key value to perform the encryption. In Amazon Elastic MapReduce, that key is your *Secret Key*. For that reason, your Secret Key must remain a shared secret between you and Elastic MapReduce.

instance	Once an AMI has been launched, the resulting running system is referred to as an instance. All instances based on the same AMI start out identical and any information on them is lost when the instances are terminated or fail.
job flow	Comprised of one or more steps, which specify all of the functions to be performed on the data. A job flow specifies the complete processing of the data.
key	The unique identifier for an object within a bucket. Every object in a bucket has exactly one key. Since a bucket and key together uniquely identify each object, Amazon Elastic MapReduce can be thought of as a basic data map between "bucket + key" and the object itself. Every object in Amazon Elastic MapReduce can be uniquely addressed through the combination of the Service endpoint, bucket name, and key.
mapper	An executable that splits the raw data into key/value pairs The output of the mapper, called *intermediate results*, is used by the reducer as its input.
master node	Keeps track of processing. The master node is a process running on an AMI that keeps track of work done by slave nodes. Each slave node represents the work done by one AMI, which is processing a block of data. The master node polls the slave nodes for status, including the completion of the processing. When a slave node completes its processing, the master node terminates the AMI.
metadata	The metadata is a set of name-value pairs that describe the object. These include default metadata such as the date last modified and standard HTTP metadata such as Content-Type. The developer can also specify custom metadata at the time the Object is stored.
object	The fundamental entity stored in Amazon S3. Objects consist of object data and metadata. The data portion is opaque to Amazon S3.
reducer	An executable that uses the intermediate results produced by the mapper and processes them into the final output
Secret Access Key	A string distributed by AWS that uniquely identifies an AWS developer. The Secret Access Key is a shared secret between a developer and AWS. The Secret Key is used as the key in the HMAC algorithm that encrypts the signature.
service endpoint	Same as endpoint.
signature	A URL-encoded string composed of request parameters and their values encrypted using an HMAC algorithm. Signatures are used to authenticate and safeguard requests.
slave node	Represents an AMI that is processing a data block Amazon Elastic MapReduce loads data blocks across multiple AMIs so the data can be processed in parallel. Each AMI that processes a data block is called a slave node. Slave nodes are polled by Master nodes for status. See master node.
step	A single function applied to the data in a job flow.

	A step is a MapReduce algorithm implemented as a Java JAR or a Hadoop streaming program written in Python, Ruby, Perl or C++. Some steps have to be applied sequentially. The sum of all steps comprises a job flow.
step type	The type of work done in a step. There are a limited number of step types, including moving data from S3 to EC2 (Distcp) and moving data from EC2 to Amazon S3.
streaming	A utility that comes with Hadoop that enables you to develop MapReduce executables in languages other than Java
tuning	Selecting the number and type of AMIs to run the Hadoop job flow most efficiently.

Document Conventions

This section lists the common typographical and symbol use conventions for AWS technical publications.

Typographical Conventions

This section describes common typographical use conventions.

Convention	Description/Example
Call-outs	A call-out is a number in the body text to give you a visual reference. The reference point is for further discussion elsewhere. You can use this resource regularly. 
Code in text	Inline code samples (including XML) and commands are identified with a special font. You can use the command `java -version`.

Code blocks	Blocks of sample code are set apart from the body and marked accordingly.
	
	`# ls -l /var/www/html/index.html`
	`-rw-rw-r--  1 root root 1872 Jun 21 09:33 /var/www/html/index.html`
	`# date`
	`Wed Jun 21 09:33:42 EDT 2006`

Emphasis	Unusual or important words and phrases are marked with a special font. You *must* sign up for an account before you can use the service.
Internal cross references	References to a section in the same document are marked. See Document Conventions (p. 40).
Logical values, constants, and regular expressions, abstracta	A special font is used for expressions that are important to identify, but are not code. If the value is `null`, the returned response will be `false`.

Convention	Description/Example
Product and feature names	Named AWS products and features are identified on first use. Create an *Amazon Machine Image* (AMI).
Operations	In-text references to operations. Use the `GetHITResponse` operation.
Parameters	In-text references to parameters. The operation accepts the parameter *`AccountID`*.
Response elements	In-text references to responses. A container for one `CollectionParent` and one or more `CollectionItems`.
Technical publication references	References to other AWS publications. If the reference is hyperlinked, it is also underscored. For detailed conceptual information, see the *Amazon Mechanical Turk Developer Guide*.
User entered values	A special font marks text that the user types. At the password prompt, type **`MyPassword`**.
User interface controls and labels	Denotes named items on the UI for easy identification. On the **File** menu, click **Properties**.
Variables	When you see this style, you must change the value of the content when you copy the text of a sample to a command line. % ec2-register *`<your-s3-bucket>`*/image.manifest See also the following symbol convention.

Symbol Conventions

This section describes the common use of symbols.

Convention	Symbol	Description/Example
Mutually exclusive parameters	(Parentheses | and | vertical | bars)	Within a code description, bar separators denote options from which one must be chosen. `% data = hdfread (start | stride | edge)`

Optional parameters
XML variable text
	[square brackets]	Within a code description, square brackets denote completely optional commands or parameters.
		`% sed [-n, -quiet]`
		Use square brackets in XML examples to differentiate them from tags.
		`<CustomerId>[ID]</CustomerId>`

Variables	<arrow brackets>	Within a code sample, arrow brackets denote a variable that must be replaced with a valid value. `% ec2-register <your-s3-bucket>/image.manifest`

Index