Get Started: Count Words with Amazon EMR

Now that you know what Amazon EMR can do, let's walk through a tutorial using mapper and reducer functions to analyze data in a streaming cluster. In this example, you'll use Amazon EMR to count the frequency of words in a text file. The mapper logic is written as a Python script and you'll use the built-in `aggregator` function provided by Hadoop as the reducer. Using the Amazon EMR console, you'll launch a cluster of virtual servers into a cluster to process the data in a distributed fashion, according to the logic in the Python script and the `aggregator` function.

In addition to the console used in this tutorial, Amazon EMR provides a command-line client, a REST-like API set, and several SDKs that you can use to launch and manage clusters. For more information about these interfaces, see What Tools are Available for Amazon EMR? (p. 10).

For console access, use your IAM user name and password to sign in to the AWS Management Console using the IAM sign-in page. IAM lets you securely control access to AWS services and resources in your AWS account. For more information about creating access keys, see How Do I Get Security Credentials? in the *AWS General Reference*.

How Much Does it Cost to Run this Tutorial?

The AWS service charges incurred by working through this tutorial are the cost of running an Amazon EMR cluster containing three m1.small instances for one hour. These prices vary by region and storage used. If you are a new customer, within your first year of using AWS, the Amazon S3 storage charges are potentially waived, given you have not used the capacity allowed in the Free Usage Tier. Amazon EMR charges are not included in the Free Usage Tier.

AWS service pricing is subject to change. For current pricing information, see the AWS Service Pricing Overview and use the AWS Simple Monthly Calculator to estimate your bill.

Topics

Sign up for the Service

If you do not have an AWS account, use the following procedure to create one.

To sign up for AWS

1. Open http://aws.amazon.com and click **Sign Up**.
2. Follow the on-screen instructions.

AWS notifies you by email when your account is active and available for you to use. Your AWS account gives you access to all services, but you are charged only for the resources that you use. For this example walk-through, the charges will be minimal.

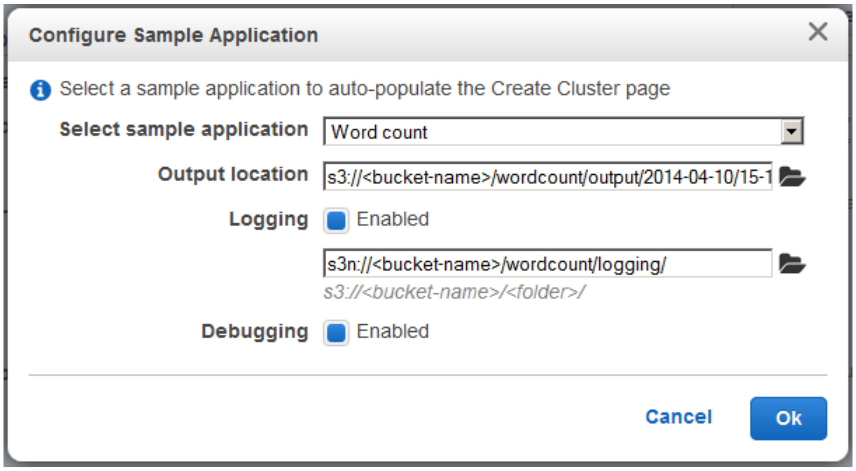
For console access, use your IAM user name and password to sign in to the AWS Management Console using the IAM sign-in page. IAM lets you securely control access to AWS services and resources in your AWS account. For more information about creating access keys, see How Do I Get Security Credentials? in the *AWS General Reference*.

Launch the Cluster

The next step is to launch the cluster. When you do, Amazon EMR provisions EC2 instances (virtual servers) to perform the computation. These EC2 instances are preloaded with an Amazon Machine Image (AMI) that has been customized for Amazon EMR and which has Hadoop and other big data applications preloaded.

To launch the Amazon EMR cluster

1. Open the Amazon Elastic MapReduce console at https://console.aws.amazon.com/elasticmapreduce/.
2. Click **Create cluster**.
3. In the **Create Cluster** page, click **Configure sample application**.



Configure Sample Application

Select a sample application to auto-populate the Create Cluster page

Select sample application: Word count

Output location: s3://<bucket-name>/wordcount/output/2014-04-10/15-1

Logging: ☑ Enabled

s3n://<bucket-name>/wordcount/logging/

s3://<bucket-name>/<folder>/

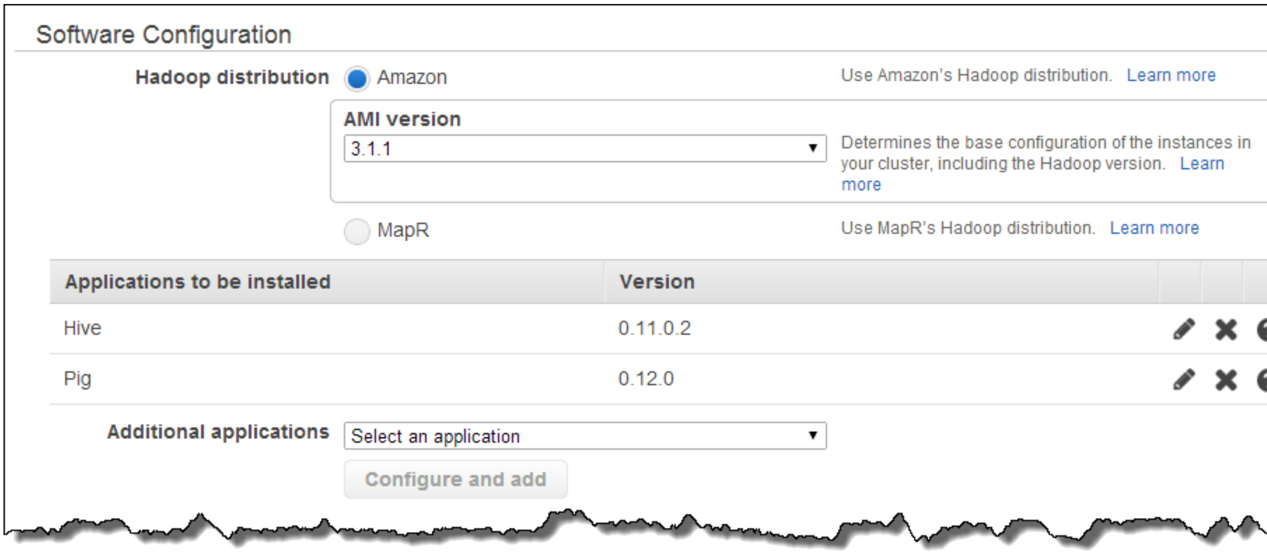
Debugging: ☑ Enabled

Field	Action
Select sample application	Select **Word count**.

Field	Action
Output location	Type the path of an Amazon S3 bucket to store your output. If the bucket does not exist, the Amazon EMR console creates it for you.
Logging	Choose **Enabled**. This determines whether Amazon EMR captures detailed log data to Amazon S3. When logging is enabled, you need to enter the output location. For more information, see View Log Files (p. 449).
Debugging	Check the box to enable debugging. This option creates a debug log index in Amazon SimpleDB (additional charges apply) to enable detailed debugging in the Amazon EMR console. You can only set this when the cluster is created. For more information about Amazon SimpleDB, go to the Amazon SimpleDB product description page.

When you have finished configuring the sample Word Count application, click **OK**.

4. In the **Software Configuration** section, verify the fields according to the following table.



Software Configuration

Hadoop distribution ● Amazon — Use Amazon's Hadoop distribution. Learn more

AMI version: 3.1.1 — Determines the base configuration of the instances in your cluster, including the Hadoop version. Learn more

○ MapR — Use MapR's Hadoop distribution. Learn more

Applications to be installed	Version
Hive	0.11.0.2
Pig	0.12.0

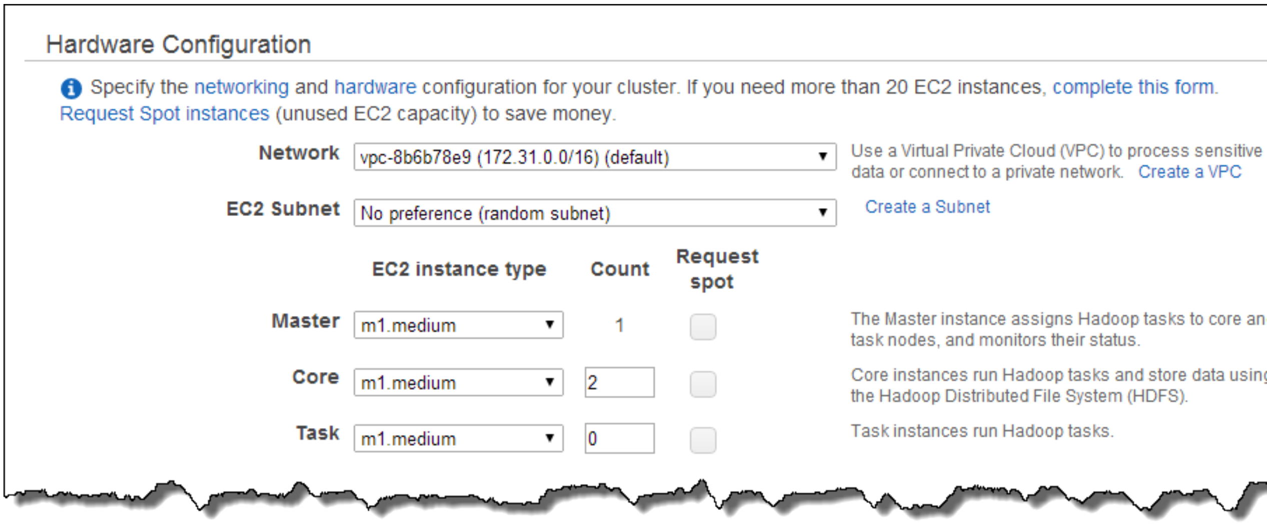
Additional applications: Select an application

Field	Action
Hadoop distribution	Choose **Amazon**. This determines which distribution of Hadoop to run on your cluster. You can choose to run the Amazon distribution of Hadoop or one of several MapR distributions. For more information, see Using the MapR Distribution for Hadoop (p. 181).

Field	Action
AMI version	Choose the latest Hadoop 2.x AMI or the latest Hadoop 1.x AMI from the list. The AMI you choose determines the specific version of Hadoop and other applications such as Hive or Pig to run on your cluster. For more information, see Choose an Amazon Machine Image (AMI) (p. 53).

5. In the **Hardware Configuration** section, verify the fields according to the following table.

   Note
   Twenty is the default maximum number of nodes per AWS account. For example, if you have two clusters, the total number of nodes running for both clusters must be 20 or less. Exceeding this limit results in cluster failures. If you need more than 20 nodes, you must submit a request to increase your Amazon EC2 instance limit. Ensure that your requested limit increase includes sufficient capacity for any temporary, unplanned increases in your needs. For more information, go to the Request to Increase Amazon EC2 Instance Limit Form.



Hardware Configuration

Specify the networking and hardware configuration for your cluster. If you need more than 20 EC2 instances, complete this form. Request Spot instances (unused EC2 capacity) to save money.

Network: vpc-8b6b78e9 (172.31.0.0/16) (default) — Use a Virtual Private Cloud (VPC) to process sensitive data or connect to a private network. Create a VPC

EC2 Subnet: No preference (random subnet) — Create a Subnet

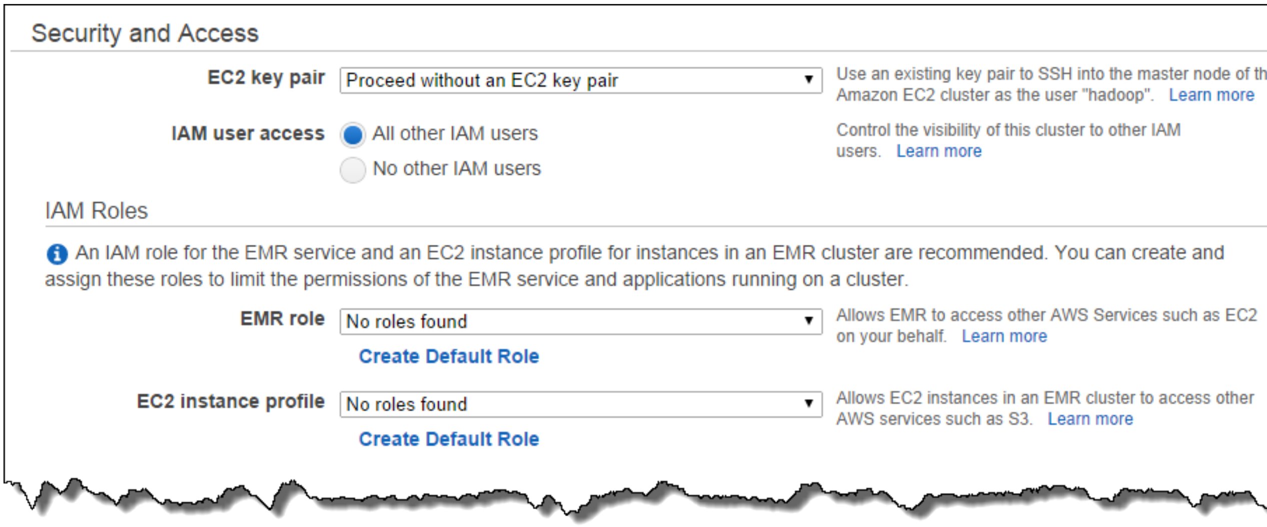
	EC2 instance type	Count	Request spot	
Master	m1.medium	1	□	The Master instance assigns Hadoop tasks to core an task nodes, and monitors their status.
Core	m1.medium	2	□	Core instances run Hadoop tasks and store data using the Hadoop Distributed File System (HDFS).
Task	m1.medium	0	□	Task instances run Hadoop tasks.

Field	Action
Network	Choose the default VPC. For more information about the default VPC, see Your Default VPC and Subnets in the *guide-vpc-user;*. Optionally, if you have created additional VPCs, you can choose your preferred VPC subnet identifier from the list to launch the cluster in that Amazon VPC. For more information, see Select a Amazon VPC Subnet for the Cluster (Optional) (p. 166).
EC2 Availability Zone	Choose **No preference**. Optionally, you can launch the cluster in a specific EC2 Availability Zone. For more information, see Regions and Availability Zones in the *Amazon EC2 User Guide for Linux Instances*.

Field	Action
Master	Accept the default instance type. The master node assigns Hadoop tasks to core and task nodes, and monitors their status. There is always one master node in each cluster. This specifies the EC2 instance type to use for the master node. The default instance type is m1.medium for Hadoop 2.x. This instance type is suitable for testing, development, and light workloads. For more information on instance types supported by Amazon EMR, see Virtual Server Configurations. For more information on Amazon EMR instance groups, see Instance Groups (p. 34). For information about mapping legacy clusters to instance groups, see Mapping Legacy Clusters to Instance Groups (p. 515).
Request Spot Instances	Leave this box unchecked. This specifies whether to run master nodes on Spot Instances. For more information, see Lower Costs with Spot Instances (Optional) (p. 37).
Core	Accept the default instance type. A core node is an EC2 instance that runs Hadoop map and reduce tasks and stores data using the Hadoop Distributed File System (HDFS). Core nodes are managed by the master node. This specifies the EC2 instance types to use as core nodes. The default instance type is m1.medium for Hadoop 2.x. This instance type is suitable for testing, development, and light workloads. For more information on instance types supported by Amazon EMR, see Virtual Server Configurations. For more information on Amazon EMR instance groups, see Instance Groups (p. 34). For information about mapping legacy clusters to instance groups, see Mapping Legacy Clusters to Instance Groups (p. 515).
Count	Choose **2**.
Request Spot Instances	Leave this box unchecked. This specifies whether to run core nodes on Spot Instances. For more information, see Lower Costs with Spot Instances (Optional) (p. 37).
Task	Accept the default instance type. Task nodes only process Hadoop tasks and don't store data. You can add and remove them from a cluster to manage the EC2 instance capacity your cluster uses, increasing capacity to handle peak loads and decreasing it later. Task nodes only run a TaskTracker Hadoop daemon. This specifies the EC2 instance types to use as task nodes. For more information on instance types supported by Amazon EMR, see Virtual Server Configurations. For more information on Amazon EMR instance groups, see Instance Groups (p. 34). For information about mapping legacy clusters to instance groups, see Mapping Legacy Clusters to Instance Groups (p. 515).
Count	Choose **0**.

Field	Action
Request Spot Instances	Leave this box unchecked. This specifies whether to run task nodes on Spot Instances. For more information, see Lower Costs with Spot Instances (Optional) (p. 37).

6. In the **Security and Access** section, complete the fields according to the following table.



Security and Access

EC2 key pair: Proceed without an EC2 key pair ▾ — Use an existing key pair to SSH into the master node of th Amazon EC2 cluster as the user "hadoop". Learn more

IAM user access: ● All other IAM users ○ No other IAM users — Control the visibility of this cluster to other IAM users. Learn more

IAM Roles

An IAM role for the EMR service and an EC2 instance profile for instances in an EMR cluster are recommended. You can create and assign these roles to limit the permissions of the EMR service and applications running on a cluster.

EMR role: No roles found ▾ — Allows EMR to access other AWS Services such as EC2 on your behalf. Learn more

Create Default Role

EC2 instance profile: No roles found ▾ — Allows EC2 instances in an EMR cluster to access other AWS services such as S3. Learn more

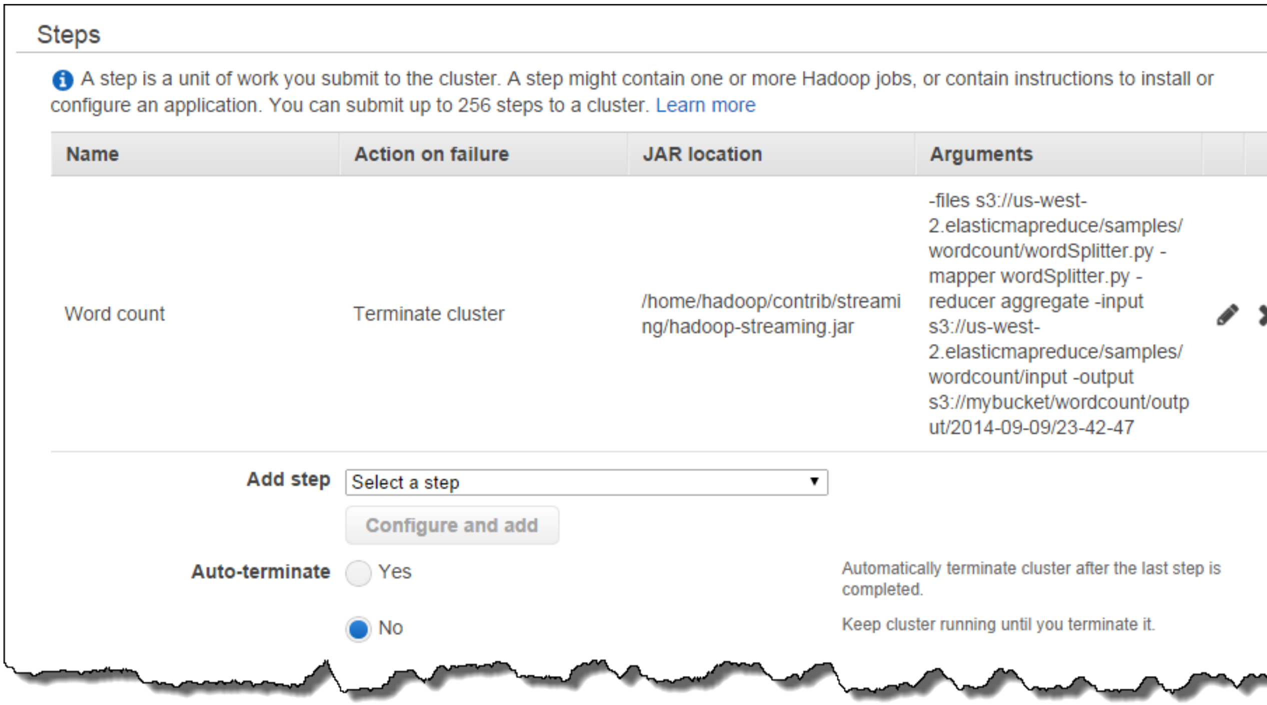
Create Default Role

Field	Action
EC2 key pair	Choose your Amazon EC2 key pair from the list. For more information, see Create an Amazon EC2 Key Pair and PEM File (p. 142). Optionally, choose **Proceed without an EC2 key pair**. If you do not enter a value in this field, you cannot use SSH to connect to the master node. For more information, see Connect to the Cluster (p. 481).
IAM user access	Choose **All other IAM users** to make the cluster visible and accessible to all IAM users on the AWS account. For more information, see Configure IAM User Permissions (p. 144). Alternatively, choose **No other IAM users** to restrict access to the current IAM user.
EMR role	Accept the default option - **No roles found**. Alternatively, click **Create Default Role > Create Role** to generate a default EMR role. Allows Amazon EMR to access other AWS services on your behalf. For more information, see Configure IAM Roles for Amazon EMR (p. 150).

Field	Action
EC2 instance profile	You can proceed without choosing an instance profile by accepting the default option - **No roles found**. Alternatively, click **Create Default Role > Create Role** to generate a default EMR role. This controls application access to the Amazon EC2 instances in the cluster. For more information, see Configure IAM Roles for Amazon EMR (p. 150).

7. In the **Bootstrap Actions** section, there are no bootstrap actions necessary for this sample configuration.

   Optionally, you can use bootstrap actions, which are scripts that can install additional software and change the configuration of applications on the cluster before Hadoop starts. For more information, see Create Bootstrap Actions to Install Additional Software (Optional) (p. 110).
8. In the **Steps** section, note the step that Amazon EMR configured for you by choosing the sample application. You can modify these settings to meet your needs. Complete the fields according to the following table.



Steps

A step is a unit of work you submit to the cluster. A step might contain one or more Hadoop jobs, or contain instructions to install or configure an application. You can submit up to 256 steps to a cluster. Learn more

Name	Action on failure	JAR location	Arguments
Word count	Terminate cluster	/home/hadoop/contrib/streaming/hadoop-streaming.jar	-files s3://us-west-2.elasticmapreduce/samples/wordcount/wordSplitter.py -mapper wordSplitter.py -reducer aggregate -input s3://us-west-2.elasticmapreduce/samples/wordcount/input -output s3://mybucket/wordcount/output/2014-09-09/23-42-47

Add step: Select a step ▾

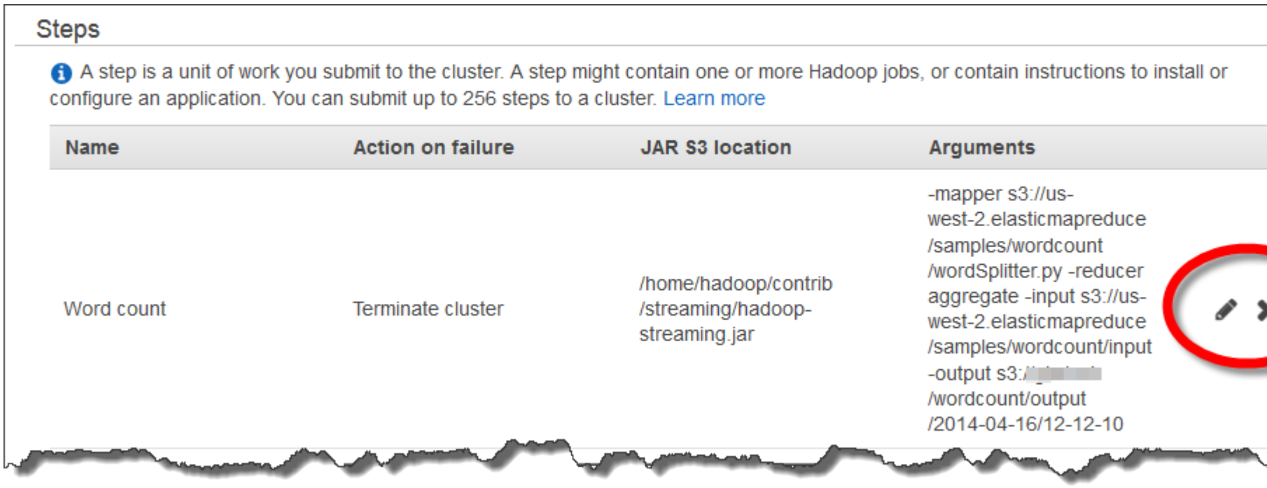
Configure and add

Auto-terminate: ○ Yes — Automatically terminate cluster after the last step is completed.

● No — Keep cluster running until you terminate it.

Field	Action
Add step	Leave this option set to **Select a step**. For more information, see Steps (p. 8).

Field	Action
Auto-terminate	Choose **No**. This determines what the cluster does after its last step. **Yes** means the cluster auto-terminates after the last step completes. **No** means the cluster runs until you manually terminate it. Remember to terminate the cluster when it is done so you do not continue to accrue charges on an idle cluster.



Steps

A step is a unit of work you submit to the cluster. A step might contain one or more Hadoop jobs, or contain instructions to install or configure an application. You can submit up to 256 steps to a cluster. Learn more

Name	Action on failure	JAR S3 location	Arguments
Word count	Terminate cluster	/home/hadoop/contrib/streaming/hadoop-streaming.jar	-mapper s3://us-west-2.elasticmapreduce/samples/wordcount/wordSplitter.py -reducer aggregate -input s3://us-west-2.elasticmapreduce/samples/wordcount/input -output s3://[redacted]/wordcount/output/2014-04-16/12-12-10

The preceding image shows the step details section with a red circle around the edit step (pencil) and delete step (X) buttons. If you click the edit button, you can edit the following settings.

Field	Action
Mapper	Set this field to **`s3n://elasticmapreduce/samples/wordcount/wordSplitter.py`**.
Reducer	Set this field to **`aggregate`**.
Input S3 location	Set this field to **`s3n://elasticmapreduce/samples/wordcount/input`**.
Output S3 location	Set this field to **`s3://`*`example-bucket`*`/wordcount/output/2013-11-11/11-07-05`**.
Arguments	Leave this field blank.
Action on failure	Set this field to **Terminate cluster**.

9. Review your configuration and if you are satisfied with the settings, click **Create Cluster**.
10. When the cluster starts, the console displays the **Cluster Details** page.

Next, Amazon EMR begins to count the words in the text of the CIA World Factbook, which is pre-configured in an Amazon S3 bucket as the input data for demonstration purposes. When the cluster is finished processing the data, Amazon EMR copies the word count results into the output Amazon S3 bucket that you chose in the previous steps.

Monitor the Cluster

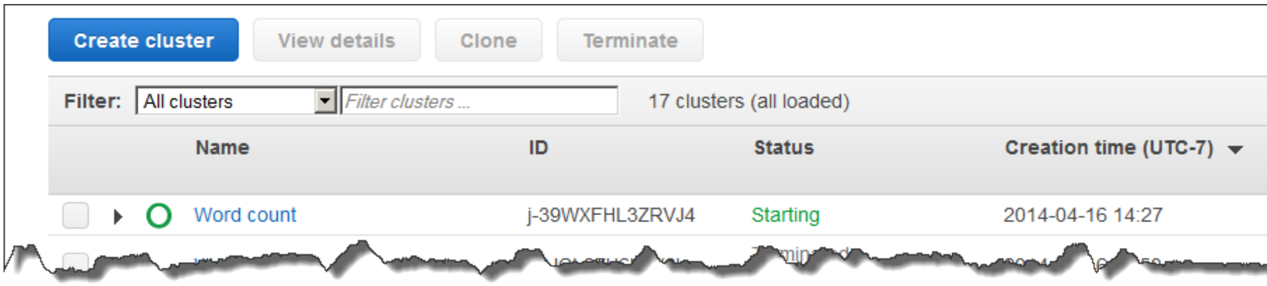
There are several ways to gain information about your cluster while it is running.

- Query Amazon EMR using the console, command-line interface (CLI), or programmatically.
- Amazon EMR automatically reports metrics about the cluster to CloudWatch. These metrics are provided free of charge. You can access them either through the CloudWatch interface or in the Amazon EMR console. For more information, see Monitor Metrics with CloudWatch (p. 455).
- Create an SSH tunnel to the master node and view the Hadoop web interfaces. Creating an SSH tunnel requires that you specify a value for **Amazon EC2 Key Pair** when you launch the cluster. For more information, see View Web Interfaces Hosted on Amazon EMR Clusters (p. 487).
- Run a bootstrap action when you launch the cluster to install the Ganglia monitoring application. You can then create an SSH tunnel to view the Ganglia web interfaces. Creating an SSH tunnel requires that you specify a value for **Amazon EC2 Key Pair** when you launch the cluster. For more information, see Monitor Performance with Ganglia (p. 472).
- Use SSH to connect to the master node and browse the log files. Creating an SSH connection requires that you specify a value for **Amazon EC2 Key Pair** when you launch the cluster.
- View the archived log files on Amazon S3. This requires that you specify a value for **Amazon S3 Log Path** when you create the cluster. For more information, see View Log Files (p. 449).

In this tutorial, you'll monitor the cluster using the Amazon EMR console.

To monitor the cluster using the Amazon EMR console

1. Click **Cluster List** in the Amazon EMR console. This shows a list of clusters to which your account has access and the status of each. In this example, see a cluster in the Starting status. There are other possible status messages, for example **Running**, **Waiting**, **Terminated (All steps completed)**, **Terminated (User request)**, **Terminated with errors (Validation error)**, etc.

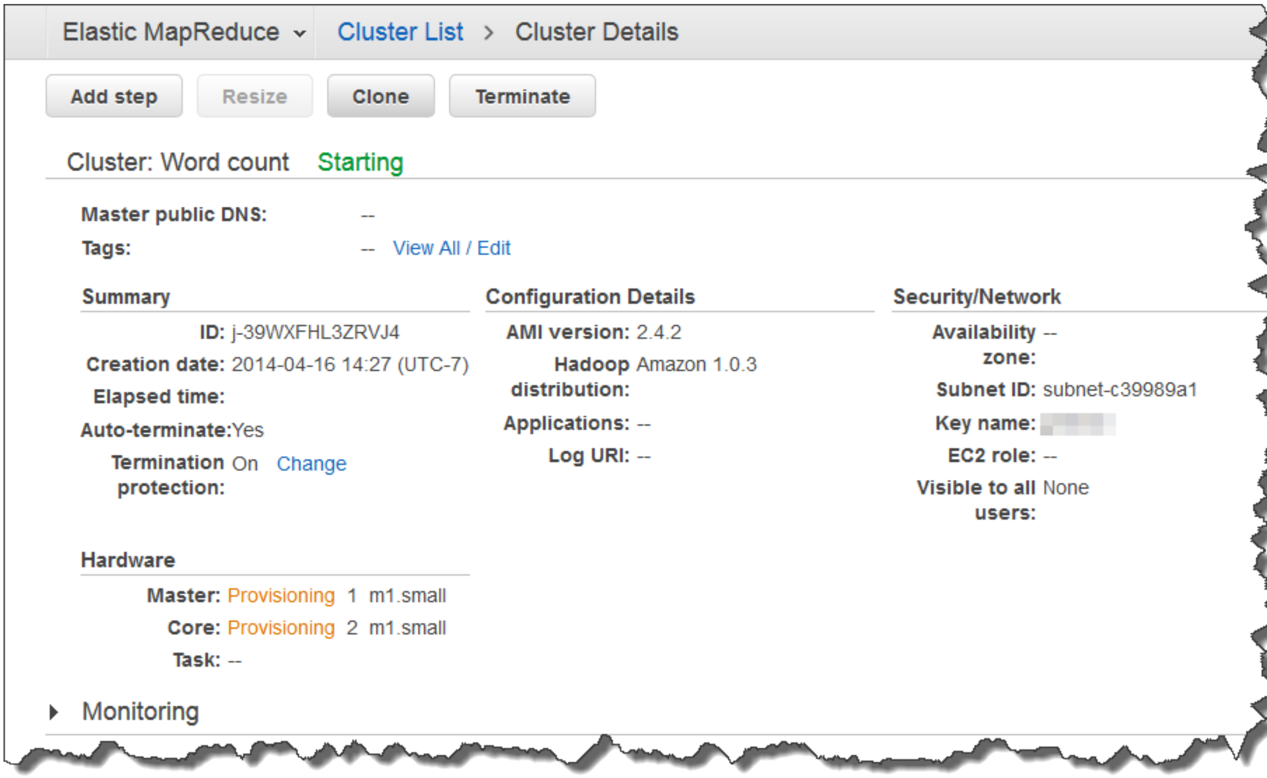


Filter: All clusters | Filter clusters ... | 17 clusters (all loaded)

Name	ID	Status	Creation time (UTC-7)
Word count	j-39WXFHL3ZRVJ4	Starting	2014-04-16 14:27

2. Click the details icon next to your cluster to see the cluster details page.

   In this example, the cluster is in **Starting** status, provisioning the compute resources needed for the Word Count application. When the cluster finishes, it will sit idle in the Waiting status because we did not configure the cluster to terminate automatically. Remember to terminate your cluster to avoid additional charges.



Cluster: Word count Starting

Master public DNS: --

Tags: --

Summary

ID: j-39WXFHL3ZRVJ4

Creation date: 2014-04-16 14:27 (UTC-7)

Elapsed time:

Auto-terminate: Yes

Termination protection: On

Configuration Details

AMI version: 2.4.2

Hadoop distribution: Amazon 1.0.3

Applications: --

Log URI: --

Security/Network

Availability zone: --

Subnet ID: subnet-c39989a1

Key name: [redacted]

EC2 role: --

Visible to all users: None

Hardware

Master: Provisioning 1 m1.small

Core: Provisioning 2 m1.small

Task: --

3. The **Monitoring** section displays metrics about the cluster. These metrics are also reported to CloudWatch, and can also be viewed from the CloudWatch console. The charts track various cluster statistics over time, for example:

   - Number of jobs the cluster is running
   - Status of each node in the cluster
   - Number of remaining map and reduce tasks
   - Number of Amazon S3 and HDFS bytes read/written

   Note
   The statistics in the **Monitoring** section may take several minutes to populate. In addition, the Word Count sample application runs very quickly and may not generate highly detailed runtime information.

   For more information about these metrics and how to interpret them, see Monitor Metrics with CloudWatch (p. 455).
4. In the **Software Configuration** section, you can see details about the software configuration of the cluster; for example:

   - The AMI version of the nodes in the cluster
   - The Hadoop Distribution
   - The Log URI to store output logs

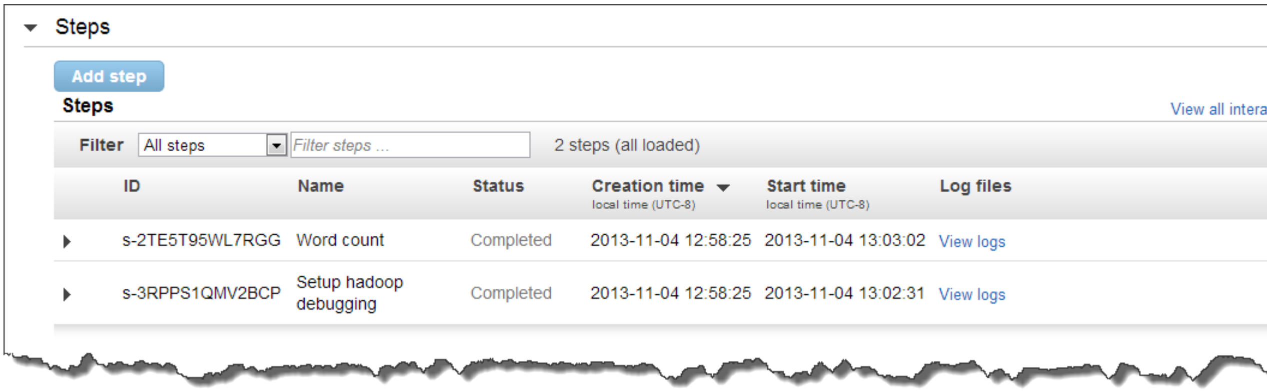
5. In the **Hardware Configuration** section, you can see details about the hardware configuration of the cluster, for example:

   - The Availability Zone the cluster runs within

- The number of master, core, and task nodes including their instance sizes and status

In addition, you can control Termination Protection and resize the cluster.

6. In the **Steps** section, you can see details about each step in the cluster. In addition, you can add steps to the cluster.



Steps

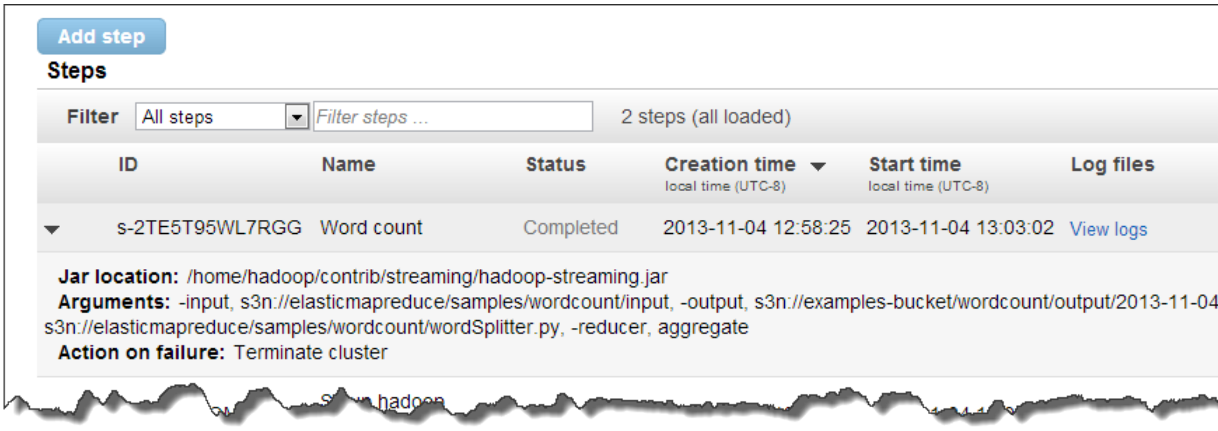
Filter: All steps | Filter steps ... | 2 steps (all loaded)

	ID	Name	Status	Creation time local time (UTC-8)	Start time local time (UTC-8)	Log files
▶	s-2TE5T95WL7RGG	Word count	Completed	2013-11-04 12:58:25	2013-11-04 13:03:02	View logs
▶	s-3RPPS1QMV2BCP	Setup hadoop debugging	Completed	2013-11-04 12:58:25	2013-11-04 13:02:31	View logs

In this example, you can see that the cluster had two steps: the Word count step (streaming step) and the Setup hadoop debugging step (script-runner step). If you enable debugging, Amazon EMR automatically adds the Setup hadoop debugging step to copy logs from the cluster to Amazon S3.

For more information about how steps are used in a cluster, see Life Cycle of a Cluster (p. 9).

- Click the arrow next to the Word Count step to see more information about the step.



Steps

Filter: All steps | Filter steps ... | 2 steps (all loaded)

	ID	Name	Status	Creation time local time (UTC-8)	Start time local time (UTC-8)	Log files
▼	s-2TE5T95WL7RGG	Word count	Completed	2013-11-04 12:58:25	2013-11-04 13:03:02	View logs

Jar location: /home/hadoop/contrib/streaming/hadoop-streaming.jar
Arguments: -input, s3n://elasticmapreduce/samples/wordcount/input, -output, s3n://examples-bucket/wordcount/output/2013-11-04
s3n://elasticmapreduce/samples/wordcount/wordSplitter.py, -reducer, aggregate
Action on failure: Terminate cluster

In this example, you can determine the following:

- The step uses a streaming JAR located on the cluster
- The input are files in an Amazon S3 location
- The output writes to an Amazon S3 location
- The mapper is a Python script named wordSplitter.py
- The final output compiles using the aggregate reducer
- The cluster will terminate if it encounters an error

7. Lastly, the **Bootstrap Actions** section lists the bootstrap actions run by the cluster, if any. In this example, the cluster has not run any bootstrap actions to initialize the cluster. For more information about how to use bootstrap actions in a cluster, see Create Bootstrap Actions to Install Additional Software (Optional) (p. 110).

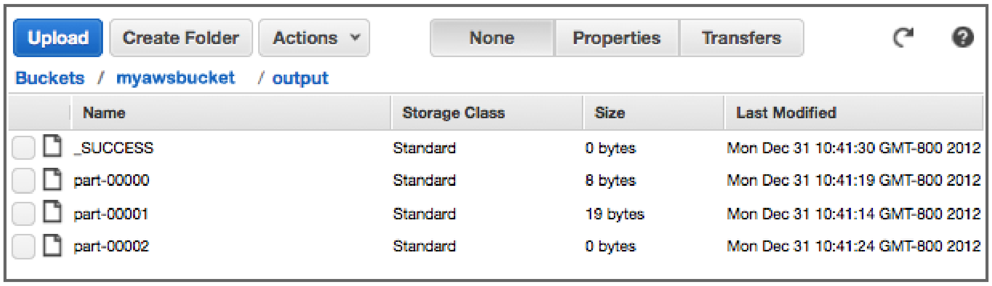
View the Results

After the cluster is complete, the results of the word frequency count are stored in the folder you specified on Amazon S3 when you launched the cluster.

To view the output of the cluster

1. From the Amazon S3 console, select the bucket you used for the output location when you configured the sample application.
2. Select the `output` folder, click **Actions**, and then select **Open**.

   The results of running the cluster are stored in text files. The first file in the listing is an empty file titled according to the result of the cluster. In this case, it is titled "_SUCCESS" to indicate that the cluster succeeded.



	Name	Storage Class	Size	Last Modified
□	_SUCCESS	Standard	0 bytes	Mon Dec 31 10:41:30 GMT-800 2012
□	part-00000	Standard	8 bytes	Mon Dec 31 10:41:19 GMT-800 2012
□	part-00001	Standard	19 bytes	Mon Dec 31 10:41:14 GMT-800 2012
□	part-00002	Standard	0 bytes	Mon Dec 31 10:41:24 GMT-800 2012

3. To download each file, right-click on it and select **Download**.
4. Open the text files using a text editor such as Notepad (Windows), TextEdit (Mac OS), or gEdit (Linux). In the output files, you should see a column that displays each word found in the source text followed by a column that displays the number of times that word was found.

The other output generated by the cluster are log files which detail the progress of the cluster. Viewing the log files can provide insight into the workings of the cluster, and can help you troubleshoot any problems that arise.

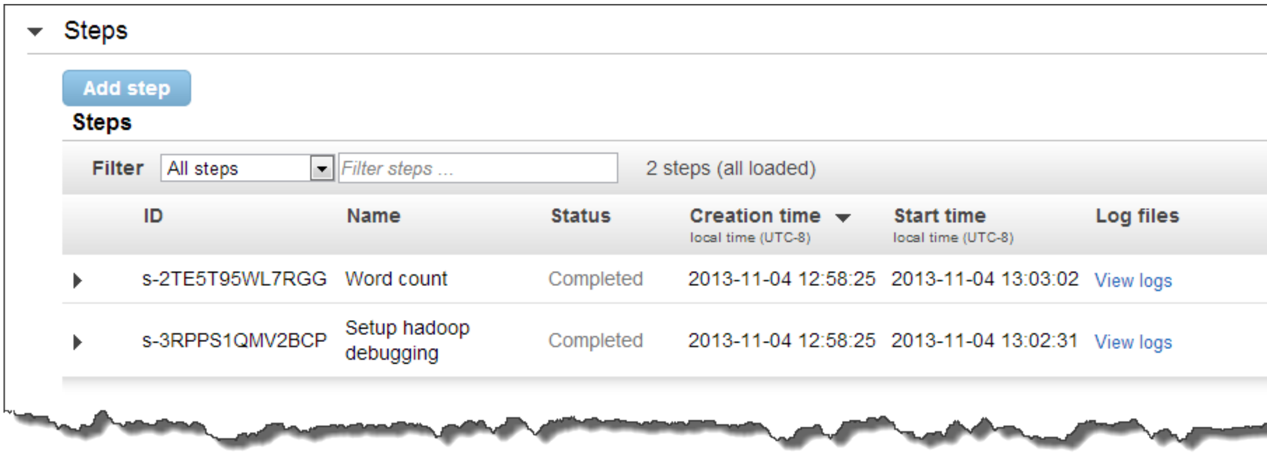
View the Debug Logs (Optional)

If you encounter any errors, you can use the debug logs to gather more information and troubleshoot the problem.

To view cluster logs using the console

1. Open the Amazon Elastic MapReduce console at https://console.aws.amazon.com/elasticmapreduce/.
2. From the **Cluster List** page, click the details icon next to the cluster you want to view.

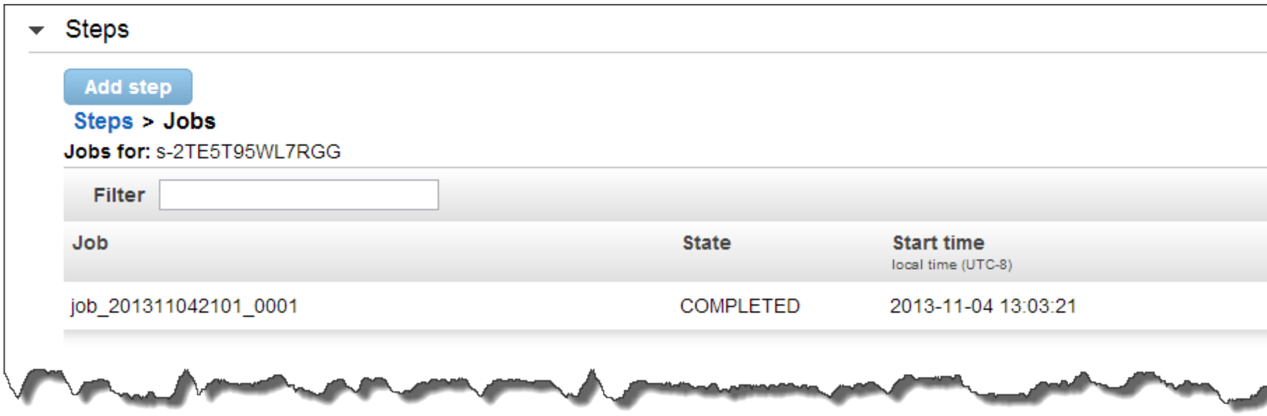
   This brings up the **Cluster Details** page. In the **Steps** section, the links to the right of each step display the various types of logs available for the step. These logs are generated by Amazon EMR.
3. To view a list of the Hadoop jobs associated with a given step, click the **View Jobs** link to the right of the step.



2 steps (all loaded)

	ID	Name	Status	Creation time local time (UTC-8)	Start time local time (UTC-8)	Log files
▸	s-2TE5T95WL7RGG	Word count	Completed	2013-11-04 12:58:25	2013-11-04 13:03:02	View logs
▸	s-3RPPS1QMV2BCP	Setup hadoop debugging	Completed	2013-11-04 12:58:25	2013-11-04 13:02:31	View logs

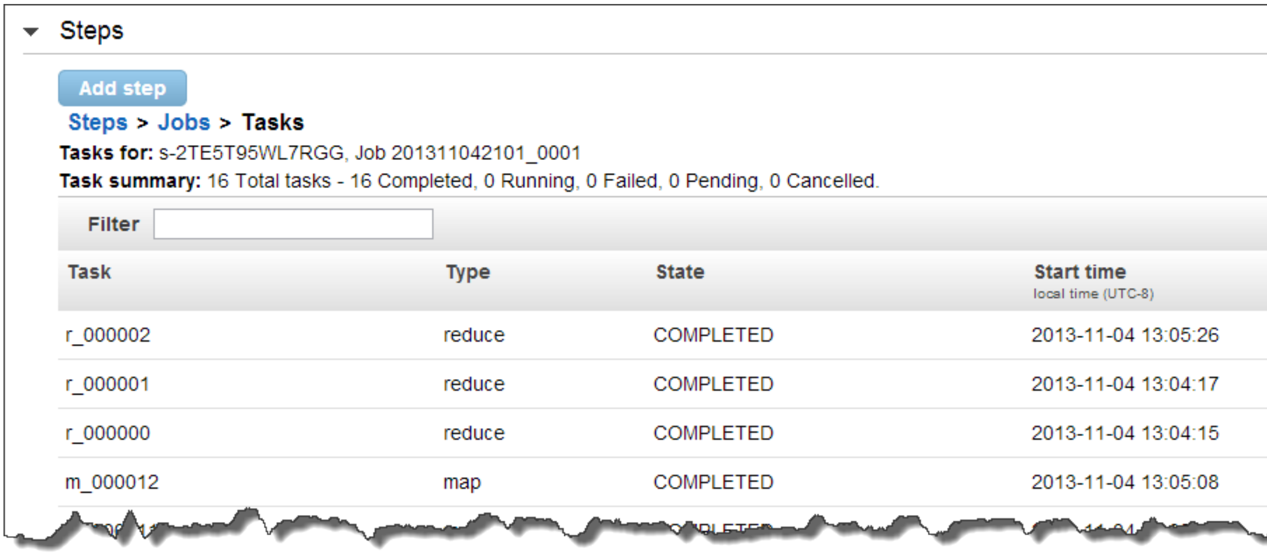
4. To view a list of the Hadoop tasks associated with a given job, click the **View Tasks** link to the right of the job.



Jobs for: s-2TE5T95WL7RGG

Job	State	Start time local time (UTC-8)
job_201311042101_0001	COMPLETED	2013-11-04 13:03:21

5. To view a list of the attempts a given task has run while trying to complete, click the **View Attempts** link to the right of the task.



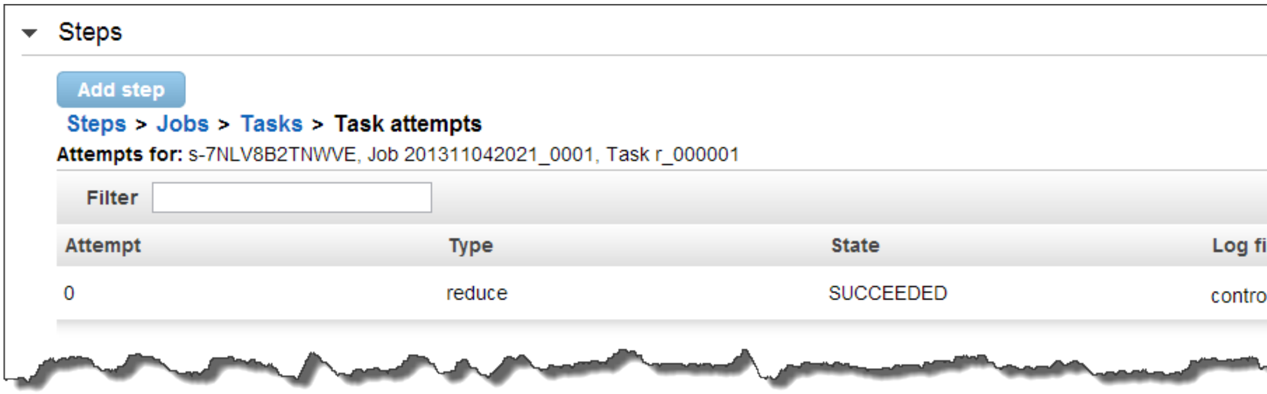
Tasks for: s-2TE5T95WL7RGG, Job 201311042101_0001

Task summary: 16 Total tasks - 16 Completed, 0 Running, 0 Failed, 0 Pending, 0 Cancelled.

Filter

Task	Type	State	Start time local time (UTC-8)
r_000002	reduce	COMPLETED	2013-11-04 13:05:26
r_000001	reduce	COMPLETED	2013-11-04 13:04:17
r_000000	reduce	COMPLETED	2013-11-04 13:04:15
m_000012	map	COMPLETED	2013-11-04 13:05:08

6. To view the logs generated by a task attempt, click the **stderr**, **stdout**, and **syslog** links to the right of the task attempt.



Attempts for: s-7NLV8B2TNWVE, Job 201311042021_0001, Task r_000001

Filter

Attempt	Type	State	Log fi
0	reduce	SUCCEEDED	contro

Clean Up

Now that you've completed the tutorial, you should delete the Amazon S3 bucket that you created to ensure that your account does not accrue additional storage charges.

You do not need to delete the completed cluster. After a cluster ends, it terminates the associated EC2 instances and no longer accrues Amazon EMR maintenance charges. Amazon EMR preserves metadata information about completed clusters for your reference, at no charge, for two months. The console does not provide a way to delete completed clusters from the console; these are automatically removed for you after two months.

Buckets with objects in them cannot be deleted. Before deleting a bucket, all objects within the bucket must be deleted.

You should also disable logging for your Amazon S3 bucket. Otherwise, logs might be written to your bucket immediately after you delete your bucket's objects.

To disable logging

1. Open the Amazon S3 console at https://console.aws.amazon.com/s3/.
2. Right-click your bucket and select **Properties**.
3. Click the **Logging** tab.
4. Deselect the **Enabled** check box to disable logging.

To delete an object

1. Open the Amazon S3 console at https://console.aws.amazon.com/s3/.
2. Click the bucket where the objects are stored.
3. Right-click the object to delete.

   Tip
   You can use the `SHIFT` and `CRTL` keys to select multiple objects and perform the same action on them simultaneously.
4. Click **Delete**.
5. Confirm the deletion when the console prompts you.

To delete a bucket, you must first delete all of the objects in it.

To delete a bucket

1. Right-click the bucket to delete.
2. Click **Delete**.
3. Confirm the deletion when the console prompts you.

You have now deleted your bucket and all its contents.

The next step is optional. It deletes two security groups created for you by Amazon EMR when you launched the cluster. You are not charged for security groups. If you are planning to explore Amazon EMR further, you should retain them.

To delete Amazon EMR security groups

1. In the Amazon EC2 console **Navigation** pane, click **Security Groups**.
2. In the **Security Groups** pane, click**ElasticMapReduce-slave**.
3. In the details pane for the ElasticMapReduce-slave security group, delete all rules that reference ElasticMapReduce. Click **Apply Rule Changes**.
4. In the right pane, select**ElasticMapReduce-master**.
5. In the details pane for the ElasticMapReduce-master security group, delete all rules that reference Amazon EMR. Click **Apply Rule Changes**.
6. With ElasticMapReduce-master security group still selected in the **Security Groups** pane, click **Delete**. Click **Yes, Delete** to confirm.
7. In the **Security Groups**pane, click **ElasticMapReduce-slave**, and then click **Delete**. Click **Yes, Delete** to confirm.