



EMR Cluster Creation

Presenter, Team

Date

Creation Options



AWS
Management
Console



AWS CLI



AWS
SDKs



Compute

EC2
Lightsail [↗](#)
ECR
ECS
EKS
Lambda
Batch
Elastic Beanstalk
Serverless Application Repository



Storage

S3
EFS
FSx
S3 Glacier
Storage Gateway
AWS Backup



Database

RDS
DynamoDB



Robotics

AWS RoboMaker



Blockchain

Amazon Managed Blockchain



Satellite

Ground Station



Management & Governance

AWS Organizations
CloudWatch
AWS Auto Scaling
CloudFormation
CloudTrail
Config
OpsWorks
Service Catalog
Systems Manager
Trusted Advisor
Managed Services



Analytics

Athena
EMR
CloudSearch
Elasticsearch Service
Kinesis
QuickSight [↗](#)
Data Pipeline
AWS Glue
MSK



Security, Identity, & Compliance

IAM
Resource Access Manager
Cognito
Secrets Manager
GuardDuty
Inspector
Amazon Macie [↗](#)
AWS Single Sign-On
Certificate Manager
Key Management Service
CloudHSM



Business Applications

Alexa for Business
Amazon Chime [↗](#)
WorkMail



End User Computing

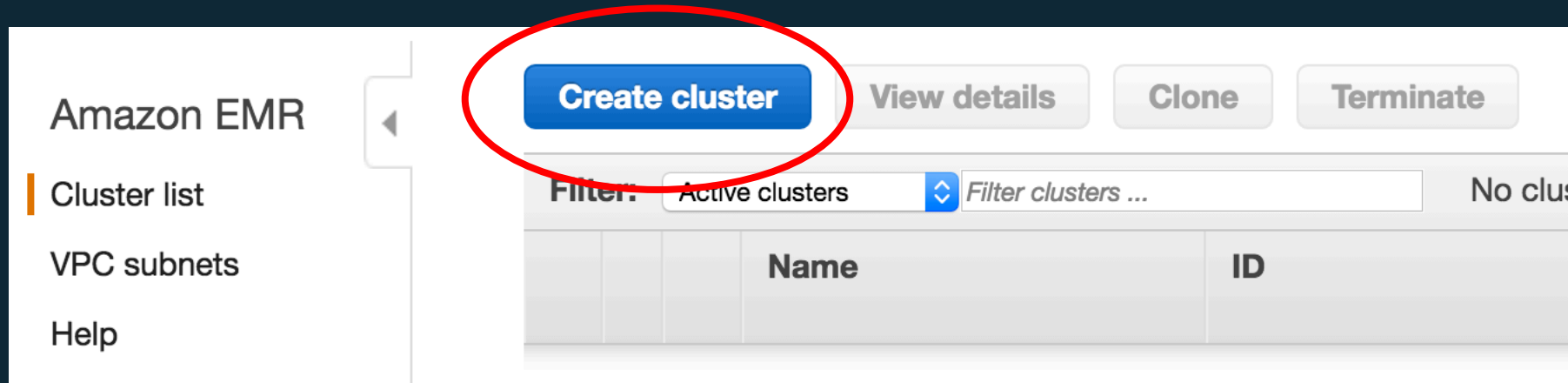
WorkSpaces
AppStream 2.0
WorkDocs
WorkLink



Internet Of Things

IoT Core
Amazon FreeRTOS
IoT 1-Click
IoT Analytics
IoT Device Defender
IoT Device Management
IoT Events
IoT Greengrass
IoT SiteWise
IoT Things Graph

Create Cluster – Console



Advanced
Options

General Configuration

Cluster name

☒ Logging ⓘ

S3 folder

Launch mode ☒ Cluster ⓘ ☐ Step execution ⓘ

Software configuration

Release ⓘ

Applications

- ☒ Core Hadoop: Hadoop 2.8.5 with Ganglia 3.7.2, Hive 2.3.4, Hue 4.4.0, Mahout 0.13.0, Pig 0.17.0, and Tez 0.9.1
- ☐ HBase: HBase 1.4.9 with Ganglia 3.7.2, Hadoop 2.8.5, Hive 2.3.4, Hue 4.4.0, Phoenix 4.14.1, and ZooKeeper 3.4.13
- ☐ Presto: Presto 0.219 with Hadoop 2.8.5 HDFS and Hive 2.3.4 Metastore
- ☐ Spark: Spark 2.4.2 on Hadoop 2.8.5 YARN with Ganglia 3.7.2 and Zeppelin 0.8.1

☐ Use AWS Glue Data Catalog for table metadata ⓘ

Hardware configuration

Instance type

Number of instances (1 master and 2 core nodes)

Security and access

EC2 key pair ⓘ [Learn how to create an EC2 key pair.](#)

Permissions ☒ Default ☐ Custom

Use default IAM roles. If roles are not present, they will be automatically created for you with managed policies for automatic policy updates.

EMR role [EMR_DefaultRole](#) ⓘ

EC2 instance profile [EMR_EC2_DefaultRole](#) ⓘ

Quick Create

Cancel

Create cluster



Step 1: Software and Steps

Step 2: Hardware

Step 3: General Cluster Settings

Step 4: Security

Software Configuration

Release



☒ Hadoop 2.8.5

☐ JupyterHub 0.9.6

☐ Ganglia 3.7.2

☒ Hive 2.3.4

☐ MXNet 1.4.0

☒ Hue 4.4.0

☐ Spark 2.4.2

☐ Zeppelin 0.8.1

☐ Tez 0.9.1

☐ HBase 1.4.9

☐ Presto 0.219

☐ Sqoop 1.4.7

☐ Phoenix 4.14.1

☐ HCatalog 2.3.4

☐ Livy 0.6.0

☐ Flink 1.8.0

☒ Pig 0.17.0

☐ ZooKeeper 3.4.13

☐ Mahout 0.13.0

☐ Oozie 5.1.0

☐ TensorFlow 1.12.0

Multi-master support

☐ Enable multi-master support

AWS Glue Data Catalog settings (optional)

☐ Use for Hive table metadata

Edit software settings

☒ Enter configuration ☐ Load JSON from S3

```
classification=config-file-name,properties={myKey1=myValue1,myKey2=myValue2}
```

Add steps (optional)

Step type



Configure

☐ Auto-terminate cluster after the last step is completed

Cancel

Next

Step 1: Software and Steps

Step 2: Hardware

Step 3: General Cluster Settings

Step 4: Security

Software Configuration

Release



☒ Hadoop 2.8.5

☐ JupyterHub 0.9.6

☐ Ganglia 3.7.2

☒ Hive 2.3.4

☐ MXNet 1.4.0

☒ Hue 4.4.0

☐ Spark 2.4.2

☐ Zeppelin 0.8.1

☐ Tez 0.9.1

☐ HBase 1.4.9

☐ Presto 0.219

☐ Sqoop 1.4.7

☐ Phoenix 4.14.1

☐ HCatalog 2.3.4

☐ Livy 0.6.0

☐ Flink 1.8.0

☒ Pig 0.17.0

☐ ZooKeeper 3.4.13

☐ Mahout 0.13.0

☐ Oozie 5.1.0

☐ TensorFlow 1.12.0

Multi-master support

☐ Enable multi-master support

AWS Glue Data Catalog settings (optional)

☐ Use for Hive table metadata

Edit software settings

☒ Enter configuration ☐ Load JSON from S3

`classification=config-file-name,properties={myKey1=myValue1,myKey2=myValue2}`

Add steps (optional)

Step type



Configure

☐ Auto-terminate cluster after the last step is completed

Cancel

Next

EMR Multi-Master

- When selected EMR multi-master will create a cluster with three master nodes
- Multi-master grants support for high availability for Hbase, YARN Resource Manager, HDFS Name Node, Spark, Hive, and Ganglia
- With multi-master Amazon EMR automatically fails over to a standby master node if the primary master node fails

Step 1: Software and Steps

Step 2: Hardware

Step 3: General Cluster Settings

Step 4: Security

Software Configuration

Release

☒ Hadoop 2.8.5

☐ JupyterHub 0.9.6

☐ Ganglia 3.7.2

☒ Hive 2.3.4

☐ MXNet 1.4.0

☒ Hue 4.4.0

☒ Spark 2.4.2

☐ Zeppelin 0.8.1

☐ Tez 0.9.1

☐ HBase 1.4.9

☒ Presto 0.219

☐ Sqoop 1.4.7

☐ Phoenix 4.14.1

☐ HCatalog 2.3.4

☐ Livy 0.6.0

☐ Flink 1.8.0

☒ Pig 0.17.0

☐ ZooKeeper 3.4.13

☐ Mahout 0.13.0

☐ Oozie 5.1.0

☐ TensorFlow 1.12.0

Multi-master support

☐ Enable multi-master support

AWS Glue Data Catalog settings (optional)

☐ Use for Hive table metadata

☐ Use for Presto table metadata

☐ Use for Spark table metadata

Edit software settings

☒ Enter configuration ☐ Load JSON from S3

```
classification=config-file-name,properties={myKey1=myValue1,myKey2=myValue2}
```

Add steps (optional)

Step type

Configure

☐ Auto-terminate cluster after the last step is completed

Cancel

Next

Glue Data Catalog Settings

- Specify the Glue data catalog as your external metastore supported in Hive, Spark, and Presto
- This configuration option is useful when you require a persistent metastore or a metastore shared by different clusters, services, applications, or AWS accounts

Step 1: Software and Steps

Step 2: Hardware

Step 3: General Cluster Settings

Step 4: Security

Software Configuration

Release

- | | | |
|--|---|--|
| <input checked="" type="checkbox"/> Hadoop 2.8.5 | <input type="checkbox"/> Zeppelin 0.8.1 | <input type="checkbox"/> Livy 0.6.0 |
| <input type="checkbox"/> JupyterHub 0.9.6 | <input type="checkbox"/> Tez 0.9.1 | <input type="checkbox"/> Flink 1.8.0 |
| <input type="checkbox"/> Ganglia 3.7.2 | <input type="checkbox"/> HBase 1.4.9 | <input checked="" type="checkbox"/> Pig 0.17.0 |
| <input checked="" type="checkbox"/> Hive 2.3.4 | <input type="checkbox"/> Presto 0.219 | <input type="checkbox"/> ZooKeeper 3.4.13 |
| <input type="checkbox"/> MXNet 1.4.0 | <input type="checkbox"/> Sqoop 1.4.7 | <input type="checkbox"/> Mahout 0.13.0 |
| <input checked="" type="checkbox"/> Hue 4.4.0 | <input type="checkbox"/> Phoenix 4.14.1 | <input type="checkbox"/> Oozie 5.1.0 |
| <input type="checkbox"/> Spark 2.4.2 | <input type="checkbox"/> HCatalog 2.3.4 | <input type="checkbox"/> TensorFlow 1.12.0 |

Multi-master support

- ☐ Enable multi-master support ⓘ

AWS Glue Data Catalog settings (optional)

- ☐ Use for Hive table metadata ⓘ

Edit software settings ⓘ

- ☒ Enter configuration ☐ Load JSON from S3

```
classification=config-file-name,properties={myKey1=myValue1,myKey2=myValue2}
```

Add steps (optional) ⓘ

Step type

- ☐ Auto-terminate cluster after the last step is completed

[Cancel](#)

[Next](#)

Step 1: Software and Steps

Step 2: Hardware

Step 3: General Cluster Settings

Step 4: Security

Software Configuration

Release



- | | | |
|--|--|--|
| <input checked="" type="checkbox"/> Hadoop 2.8.5 | <input type="checkbox"/> Zeppelin 0.8.1 | <input type="checkbox"/> Livy 0.6.0 |
| <input type="checkbox"/> JupyterHub 0.9.6 | <input type="checkbox"/> Tez 0.9.1 | <input type="checkbox"/> Flink 1.8.0 |
| <input type="checkbox"/> Ganglia 3.7.2 | <input type="checkbox"/> HBase 1.4.9 | <input checked="" type="checkbox"/> Pig 0.17.0 |
| <input checked="" type="checkbox"/> Hive 2.3.4 | <input checked="" type="checkbox"/> Presto 0.219 | <input type="checkbox"/> ZooKeeper 3.4.13 |
| <input type="checkbox"/> MXNet 1.4.0 | <input type="checkbox"/> Sqoop 1.4.7 | <input type="checkbox"/> Mahout 0.13.0 |
| <input checked="" type="checkbox"/> Hue 4.4.0 | <input type="checkbox"/> Phoenix 4.14.1 | <input type="checkbox"/> Oozie 5.1.0 |
| <input checked="" type="checkbox"/> Spark 2.4.2 | <input type="checkbox"/> HCatalog 2.3.4 | <input type="checkbox"/> TensorFlow 1.12.0 |

Multi-master support

- ☐ Enable multi-master support

AWS Glue Data Catalog settings (optional)

- ☐ Use for Hive table metadata
- ☐ Use for Presto table metadata
- ☐ Use for Spark table metadata

Edit software settings

- ☒ Enter configuration ☐ Load JSON from S3

```
classification=config-file-name,properties=[myKey1=myValue1,myKey2=myValue2]
```

Add steps (optional)

Step type



Configure

- ☐ Auto-terminate cluster after the last step is completed

Cancel

Next

Add Steps

- A step is a unit of work you submit to the cluster
- Can choose to auto-terminate cluster upon completion
- Preconfigured options for streaming, Hive, Pig, Spark, and customer JARs

Add steps (optional) ⓘ

Step type ☐ Auto-terminate

- ✓ Select a step
- Streaming program
- Hive program
- Pig program
- Spark application
- Custom JAR

Hardware Configuration

Hardware Configuration ⓘ

If you need more than 20 EC2 instances, [see this topic](#).

Instance group configuration



Uniform instance groups

Specify a single instance type and purchasing option for each node type.



Instance fleets

Specify target capacity and how Amazon EMR fulfills it for each node type. Mix instance types and purchasing options. [Learn more](#)

Network

vpc-227bc846 (172.31.0.0/16) (default) | private



[Create a VPC](#) ⓘ

EC2 Subnet

subnet-30d2ee1b | Default in us-east-1a



Root device EBS volume size

10

GiB ⓘ

Choose the instance type, number of instances, and a purchasing option. You can choose to use On-Demand Instances, Spot Instances, or both. The instance type and purchasing option apply to all EC2 instances in each instance group, and you can only specify these options for an instance group when you create it. [Learn more about instance purchasing options](#)

Node type	Instance type	Instance count	Purchasing option	Auto Scaling	
Master Master - 1	m3.xlarge 8 vCore, 15 GiB memory, 80 SSD GB storage EBS Storage: none Add configuration settings	1 Instances	<input checked="" type="radio"/> On-demand ⓘ <input type="radio"/> Spot ⓘ Use on-demand as max price	Not available for Master	?
Core Core - 2	m3.xlarge 8 vCore, 15 GiB memory, 80 SSD GB storage EBS Storage: none Add configuration settings	2 Instances	<input checked="" type="radio"/> On-demand ⓘ <input type="radio"/> Spot ⓘ Use on-demand as max price	Not enabled	?
Task ✕ Task - 3	m3.xlarge 8 vCore, 15 GiB memory, 80 SSD GB storage EBS Storage: none Add configuration settings	0 Instances	<input checked="" type="radio"/> On-demand ⓘ <input type="radio"/> Spot ⓘ Use on-demand as max price	Not enabled	?

Instance Group Configuration – Instance fleets

- Provision from a list of instance types with Spot and On-Demand
- Launch in the most optimal Availability Zone based on capacity/price
- Spot block support

Instance group configuration



Uniform instance groups

Specify a single instance type and purchasing option for each node type.



Instance fleets

Specify target capacity and how Amazon EMR fulfills it for each node type. Mix instance types and purchasing options. [Learn more](#)

If you need more than 20 EC2 instances, [see this topic](#)

- Instance group configuration

☐ Uniform instance groups

Specify a single instance type and purchasing option for each node type.
- ☒ Instance fleets

Specify target capacity and how Amazon EMR fulfills it for each node type. Mix instance types and purchasing options. [Learn more](#)

Network

vpc-227bc846 (172.31.0.0/16) (default) | private

[Create a VPC](#) ⓘ

EC2 Subnet

subnet-a0a8089d | Default in us-east-1e

subnet-aad933dc | Default in us-east-1b

subnet-bbc55ab7 | Default in us-east-1f

subnet-d90d9ebc | Default in us-east-1d

Root device EBS volume size GiB ⓘ

Networking – VPC & Subnets

Node types
Instance types
Instance count

For each fleet, specify up to five instance types. For core and task fleets, enter target capacities for on-demand and spot instances. Amazon EMR launches instances from among the types you specify to fulfill the targets. For the master fleet, the target is always one. For each instance type, choose a maximum spot price. The advanced Spot options for each fleet determine Spot provisioning behavior. [Learn more](#)

Node type	Fleet instance types	Target capacity	Advanced Spot options	Auto Scaling
Master Master - 1	m3.xlarge 8 vCore, 15 GiB memory, 80 SSD GB storage EBS Storage: none Maximum Spot price: % On-Demand 100 Add / remove instance types to fleet	<input checked="" type="radio"/> On-demand <input type="radio"/> Spot <i>The master fleet consists of one EC2 instance</i>		Not available for instance fleets
Core Core - 2	m3.xlarge 8 vCore, 15 GiB memory, 80 SSD GB storage EBS Storage: none Maximum Spot price: % On-Demand 100 Each instance counts as <input type="text" value="8"/> units ⓘ Add / remove instance types to fleet	<div><input type="text" value="0"/> On-demand units</div> <div><input type="text" value="0"/> Spot units</div> <div>0 Total units</div>	<div>Defined duration ⓘ Not set </div> <div>Provisioning timeout ⓘ Terminate cluster </div> <div>after <input type="text" value="60"/> min. of Spot unavailability</div>	Not available for instance fleets

Hardware Configuration – Spot Integration

Node type	Instance type	Instance count	Purchasing option
Master Master - 1	m3.xlarge 8 vCore, 15 GiB memory, 80 SSD GB storage EBS Storage: none Add configuration settings	1 Instances	<input type="radio"/> On-demand <input checked="" type="radio"/> Spot Set max price
Core Core - 2	m3.xlarge 8 vCore, 15 GiB memory, 80 SSD GB storage EBS Storage: none Add configuration settings	2 Instances	<input checked="" type="radio"/> Spot Use on-demand as max price
Task Task - 3	m3.xlarge 8 vCore, 15 GiB memory, 80 SSD GB storage EBS Storage: none Add configuration settings	0 Instances	<input checked="" type="radio"/> Spot Use on-demand as max price

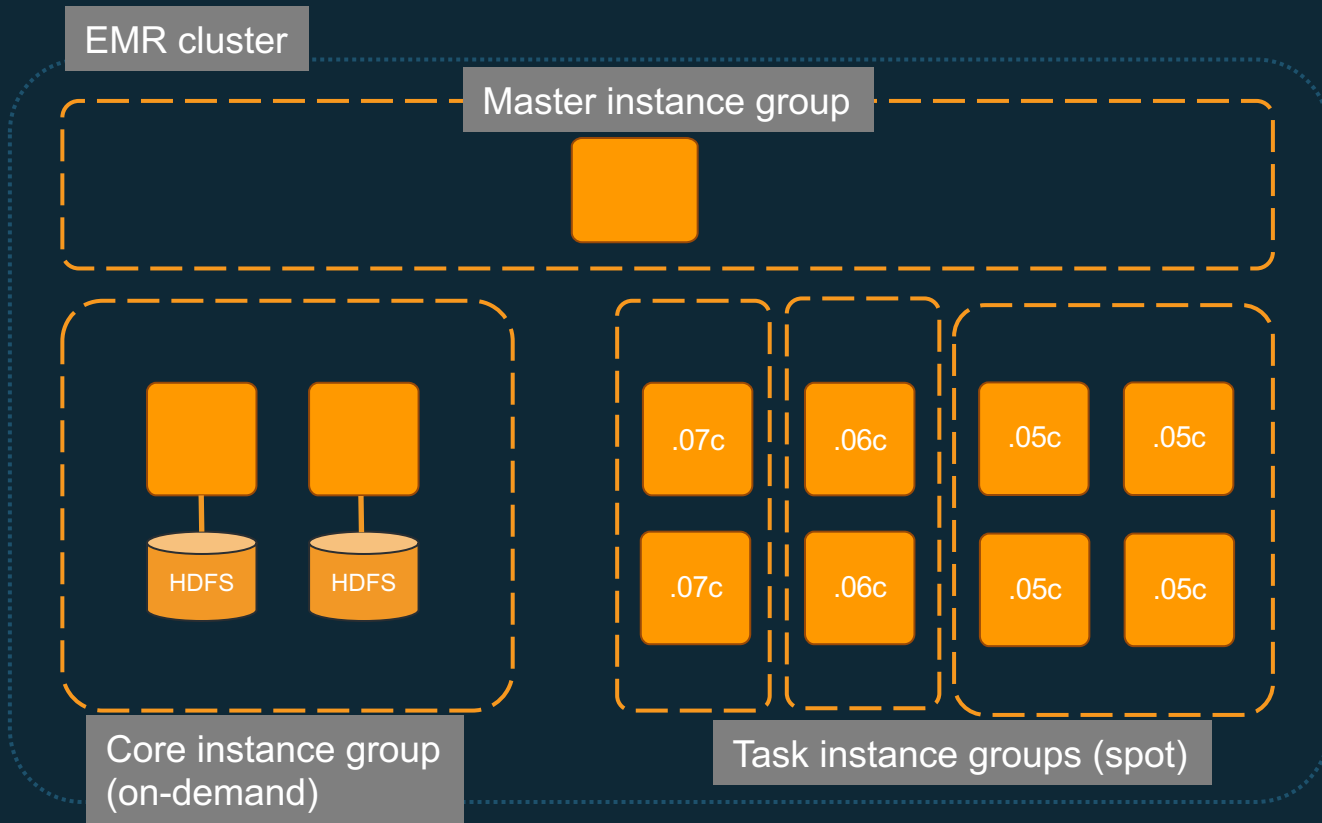
Current spot price

Availability zone	Price	
us-east-1a	\$0.062	lowest
us-east-1b	\$0.065	
us-east-1c	\$0.062	lowest
us-east-1e	\$0.062	lowest

0.266

Current spot
bid price

Hardware Configuration – Multiple Task Groups



Hardware Configuration – Auto Scaling

Node type	Instance type	Instance count	Purchasing option	Auto Scaling
Master Master - 1	m3.xlarge 8 vCore, 15 GiB memory, 80 SSD GB storage EBS Storage: none Add configuration settings	1 Instances	<input checked="" type="radio"/> On-demand <input type="radio"/> Spot Set max price per instance/hr \$ 0.266	Not available for Master
Core Core - 2	m3.xlarge 8 vCore, 15 GiB memory, 80 SSD GB storage EBS Storage: none Add configuration settings	2 Instances	<input checked="" type="radio"/> On-demand <input type="radio"/> Spot Use on-demand as max price	Not enabled
Task Task - 3	m3.xlarge 8 vCore, 15 GiB memory, 80 SSD GB storage EBS Storage: none Add configuration settings	0 Instances	<input type="radio"/> On-demand <input checked="" type="radio"/> Spot Set max price per instance/hr \$ 0.266	Not enabled

Auto scaling option

Maximum instances: ⓘ

Minimum instances: ⓘ

☒ Scale out

Default-scale-out-1: Add instance if YARNMemoryAvailablePercentage is less than for five-minute period with a cooldown of seconds



Default-scale-out-2: Add instance if ContainerPendingRatio is greater than for five-minute period with a cooldown of seconds



[+ Add rule](#)

☒ Scale in

Default-scale-in: Terminate instance if YARNMemoryAvailablePercentage is greater than for five-minute period with a cooldown of seconds



[+ Add rule](#)

Done

General Options and Tags

General Options

Cluster name



☒ Logging 

S3 folder 

☒ Debugging 

☒ Termination protection 

Tags

Key	Value (optional)	
<input type="text" value="Environment"/>	<input type="text" value="Dev"/>	
<input type="text" value="Team"/>	<input type="text" value="Software"/>	
<input type="text" value="Add a key to create a tag"/>	<input type="text"/>	

General Options – Logging

- By default, logs written to master node - /mnt/var/log
 - Step logs
 - Hadoop and YARN component logs
 - Bootstrap Action logs
 - Instance State logs
- If Logging box is checked, logs are also written to S3
 - 5 minute intervals
 - On by default for console
 - Off by default for CLI

General Options – Debugging

- When you enable debugging on a cluster, Amazon EMR archives the log files to Amazon S3 and then indexes those files
- You can then use the console to browse the step, job, task, and task-attempt logs for the cluster
- Logs for debugging are also pushed to S3 at 5 minutes intervals

General Options – Termination Protection

- Prevents accidental termination of the cluster
 - On by default in Console
 - Off by default for CLI

Terminate cluster ✕

This cluster has Termination Protection on. You must turn off termination protection to proceed.

Termination protection: ☒ On ☐ Off ✓ ✗

Any pending work or data residing on the cluster will be lost, such as data stored in HDFS. This action is irreversible.

Cancel Terminate

General Options – Additional Options

Additional Options

☒ EMRFS consistent view ⓘ

EMRFS metadata store

EmrFSMetadata ⓘ

Number of retries

5 ⓘ

Retry period (in seconds)

10 ⓘ

Custom AMI ID

None ⓘ

▼ Bootstrap Actions

Bootstrap actions are scripts that are executed during setup before Hadoop starts on every cluster node. You can use them to install additional software and customize your applications. [Learn more](#)

Add bootstrap action

Select a bootstrap action ▼

Configure and add

Additional Options – EMRFS Consistent View

- S3 is eventually consistent
- EMRFS consistent view:
 - Uses DynamoDB as a file registry
 - Allows EMR clusters to check for list and read-after-write consistency for Amazon S3 objects written by or synced with EMRFS
- You can configure:
 - Number of times EMFRS calls S3 after finding inconsistency
 - Amount of time until the first retry. Subsequent retries use an exponential back-off

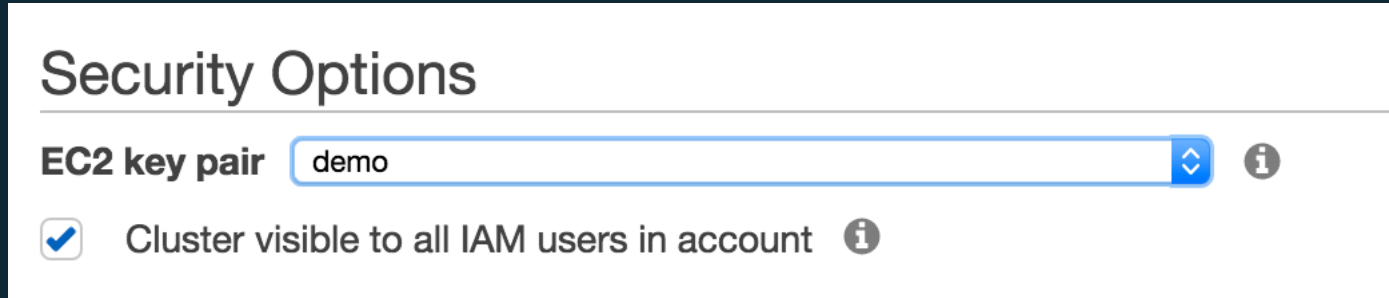
Additional Options – Custom AMI

- Benefits
 - Can reduce cluster start time by pre-installing applications and perform other customizations instead of using bootstrap actions
 - Prevent unexpected bootstrap action failures
 - Support for Amazon EBS root volume encryption
- Requirements
 - Must be an *Amazon Linux* AMI. Amazon Linux 2 AMIs are not supported
 - Must be an HVM and an EBS-backed AMI
 - Must be a 64-bit AMI
 - Must not have users with the same name as applications (example: `hadoop`, `hdfs`, `yarn`, or `spark`)

Additional Options – Bootstrap Actions (BA)

- Scripts executed before Hadoop starts on each node
- Typically used for installing additional software
- Can run up to 16 BAs
- Run If
 - Conditional on an instance-specific value found in the instance.json or job-flow.json i.e. IsMaster=true
- Custom
 - Run custom script. i.e. Copy a file from S3 to each node

Security Options – EC2 Key Pair & Visibility



The screenshot shows the 'Security Options' section of the AWS Management Console. It features a header 'Security Options' followed by a horizontal line. Below the line, there is a label 'EC2 key pair' followed by a dropdown menu showing 'demo'. To the right of the dropdown is an information icon (i). Below this, there is a checked checkbox followed by the text 'Cluster visible to all IAM users in account' and another information icon (i).

- EC2 key pair
 - Need to attach key pair to be able to SSH into Master node
- Cluster visible
 - If unchecked, only creator can see Cluster in CLI and Console

Security Options – Permissions

Permissions ⓘ

☒ Default ☐ Custom

Use default IAM roles. If roles are not present, they will be automatically created for you with managed policies for automatic policy updates.

EMR role [EMR_DefaultRole](#) ⓘ

EC2 instance profile [EMR_EC2_DefaultRole](#) ⓘ

Auto Scaling role [EMR_AutoScaling_DefaultRole](#) ⓘ

Security Options – Permissions

- EMR uses Roles to access AWS resources
 - EMR Role
 - Allows EMR to access resources such as EC2
 - EC2 Instance Profile
 - Allows EC2 instances in cluster to access resources such as S3
 - Auto Scaling Role
 - Allows Auto Scaling to add and terminate instances
- Can use default or choose your own

Security Options – Encryption & Security Groups

▼ Encryption Options

Security configuration

Dev Config



▼ EC2 Security Groups

An EC2 security group acts as a virtual firewall for your cluster nodes to control inbound and outbound traffic. There are two types of security groups you can configure, [EMR managed security groups](#) and [additional security groups](#). EMR will [automatically update](#) the rules in the EMR managed security groups in order to launch a cluster. [Learn more](#).

Type	EMR managed security groups EMR will automatically update the selected group	Additional security groups EMR will not modify the selected groups
Master	Default: sg-f712b29a (ElasticMapReduce-master)	No security groups selected
Core & Task	Default: sg-f112b29c (ElasticMapReduce-slave)	No security groups selected

[Create a security group](#)

Create security configuration

Needs to be created
before you start cluster
creation

Name

☒ At-rest encryption

Enable and choose options for at-rest data encryption features in Amazon EMR, including Amazon S3 with EMRFS, local volumes attached to cluster instances, and block-transfer encryption for HDFS. [Learn more](#)

S3 encryption ⓘ

Encryption mode ⓘ

AWS KMS Key ⓘ

Local disk encryption ⓘ

Key provider type

AWS KMS Key ⓘ

☒ In-transit encryption

Enable and choose options for open-source encryption features that apply to in-transit data for specific applications. Available encryption options may vary by Amazon EMR release. [Learn more](#)

TLS certificate provider

Certificate provider type ⓘ

S3 object ⓘ

EMR Notebooks

- Based on Jupyter to analyze data interactively
- Create and attach notebooks to Amazon EMR clusters running Hadoop, Spark, and Livy. Notebooks are saved in Amazon S3 independently of clusters

Notebooks

Use EMR notebooks based on Jupyter to analyze data interactively with live code, narrative text, visualizations, and more. Create and attach notebooks to Amazon EMR clusters running Hadoop, Spark, and Livy. Notebooks run free of charge and are saved in Amazon S3 independently of clusters. Standard billing for clusters and Amazon S3 apply. [Learn more](#)

Create notebook

View details

Open

Start

Stop

Delete

Filter: All notebooks



x

Showing 0 of 1 notebooks



Name	Status	Cluster	Creation time (UTC-4)	Last modified
Showing 0 of 1 notebooks				



Showing 0 of 1 notebooks

Name and configure your notebook

Name your notebook, choose a cluster or create one, and customize configuration options if desired. [Learn more](#) 

Notebook name*


Names may only contain letters (a-z), numbers (0-9), hyphens (-), or underscores (_).

Description



256 characters max.

Cluster* ☒ Choose an existing cluster


☐ Create a cluster 

Security groups ☒ Use default security groups 

☐ Choose security groups

AWS service role*  

Notebook location* Choose an S3 location where files for this notebook are saved.

☐ Use a location that EMR creates 

☒ Choose an existing S3 location in us-east-1



► **Tags** 

Cluster Details

Cluster: My cluster **Waiting** Cluster ready after last step completed.

Summary

Application history

Monitoring

Hardware

Events

Steps

Configurations

Bootstrap actions

Connections: [Enable Web Connection](#) – Hue, Spark History Server, Resource Manager ... (View All)

Master public DNS: ec2-54-89-173-61.compute-1.amazonaws.com [SSH](#)

Tags: Name = My Cluster [View All / Edit](#)

Summary

ID: j-3OQFQVKDBEL82

Creation date: 2018-06-26 10:14 (UTC-5)

Elapsed time: 10 minutes

Auto-terminate: No


Termination protection: On [Change](#)

Configuration details

Release label: emr-5.14.0

Hadoop distribution: Amazon 2.8.3

Applications: Hive 2.3.2, Pig 0.17.0, Hue 4.1.0,
Spark 2.3.0, Livy 0.4.0, Oozie 4.3.0,
Sqoop 1.4.7

Log URI: s3://tm-app-logs/emr/ 

EMRFS consistent view: Disabled

Custom AMI ID: --

Network and hardware

Availability zone: us-east-1a

Subnet ID: [subnet-2fac0165](#)

Master: **Running** 1 m3.xlarge

Core: **Running** 2 m3.xlarge

Task: **Running** 2 m3.xlarge

Security and access

Key name: dev-virginia

EC2 instance profile: EMR_EC2_DefaultRole

EMR role: EMR_DefaultRole

Auto Scaling role: EMR_AutoScaling_DefaultRole

Visible to all users: All [Change](#)

Security groups for [sg-8ebea5fb](#) (emr-master)

Master:

Security groups for [sg-d3b8a3a6](#) (emr-slave)

Core & Task:

Cluster Details

Summary

Application history

Monitoring

Hardware

Events

Steps

Configurations

Bootstrap actions

Connections: [Enable Web Connection](#) – Hue, Spark History Server, Resource Manager ... (View All)

Master public DNS: [ec2-54-89-173-61.compute-1.amazonaws.com](#) [SSH](#)

Tags: Name = My Cluster [View All / Edit](#)

Master node DNS

Summary

ID: j-3OQFQVKDBEL82

Creation date: 2018-06-26 10:14 (UTC-5)

Elapsed time: 10 minutes

Auto-terminate: No


Termination protection: On [Change](#)

Configuration details

Release label: emr-5.14.0

Hadoop distribution: Amazon 2.8.3

Applications: Hive 2.3.2, Pig 0.17.0, Hue 4.1.0,
Spark 2.3.0, Livy 0.4.0, Oozie 4.3.0,
Sqoop 1.4.7

Log URI: s3://tm-app-logs/emr/ 

EMRFS consistent view: Disabled

Custom AMI ID: --

Network and hardware

Availability zone: us-east-1a

Subnet ID: [subnet-2fac0165](#)

Master: Running 1 m3.xlarge

Core: Running 2 m3.xlarge

Task: Running 2 m3.xlarge

Cluster Details

Summary Application history Monitoring Hardware Events Steps Configurations Bootstrap

Connections:

[Enable Web Connection](#) – Hue, Spark History Server, Resource Manager ... (View All)

Master public DNS:

ec2-54-89-173-61.compute-1.amazonaws.com [SSH](#)

Tags:

Name = My Cluster [View All](#) / [Edit](#)

Information about the software you are running, logs, and features

Summary

ID: j-3OQFQVKDBEL82

Creation date: 2018-06-26 10:14 (UTC-5)

Elapsed time: 10 minutes

Auto-terminate: No

Termination protection: [Change](#)

Configuration details

Release label: emr-5.14.0

Hadoop distribution: Amazon 2.8.3

Applications: Hive 2.3.2, Pig 0.17.0, Hue 4.1.0, Spark 2.3.0, Livy 0.4.0, Oozie 4.3.0, Sqoop 1.4.7

Log URI: s3://tm-app-logs/emr/ 

EMRFS consistent view: Disabled

Custom AMI ID: --

Network and hardware

Availability zone: us-east-1a

Subnet ID: [subnet-2fac0165](#)

Master: Running 1 m3.xlarge

Core: Running 2 m3.xlarge

Task: Running 2 m3.xlarge

Cluster Details

Summary Application history Monitoring Hardware Events Steps Configurations Bootstrap actions


Connections: [Enable Web Connection](#) – Hue, S
Master public DNS: ec2-54-89-173-61.compute-1.am
Tags: Name = My Cluster [View All / Edit](#)

Information on the
infrastructure for this
cluster

Summary

ID: j-3OQFQVKDBEL82
Creation date: 2018-06-26 10:14 (UTC-5)
Elapsed time: 10 minutes
Auto-terminate: No
Termination protection: On [Change](#)

Configuration details

Release label: emr-5.14.0
Hadoop distribution: Amazon 2.8.3
Applications: Hive 2.3.2, Pig 0.17.0, Hue 4.1.0,
Spark 2.3.0, Livy 0.4.0, Oozie 4.3.0,
Sqoop 1.4.7
Log URI: s3://tm-app-logs/emr/ 
EMRFS consistent view: Disabled
Custom AMI ID: --

Network and hardware

Availability zone: us-east-1a
Subnet ID: [subnet-2fac0165](#)
Master: Running 1 m3.xlarge
Core: Running 2 m3.xlarge
Task: Running 2 m3.xlarge

Cluster Details

Summary

ID: j-3OQFQVKDBEL82

Creation date: 2018-06-26 10:14 (UTC-5)

Elapsed time: 10 minutes

Auto-terminate: No

Termination protection: On [Change](#)

Security and access

Key name: dev-virginia

EC2 instance profile: EMR_EC2_DefaultRole

EMR role: EMR_DefaultRole

Auto Scaling role: EMR_AutoScaling_DefaultRole

Visible to all users: All [Change](#)

Security groups for [sg-8ebea5fb](#) (emr-master)
Master:


Security groups for [sg-d3b8a3a6](#) (emr-slave)
Core & Task:

Configuration details

Release label: emr-5.14.0

Hadoop distribution: Amazon 2.8.3

Applications: Hive 2.3.2, Pig 0.17.0, Hue 4.1.0,
Spark 2.3.0, Livy 0.4.0, Oozie 4.3.0,
Sqoop 1.4.7

Log URI: s3://tm-app-logs/emr/ 

EMRFS consistent view: Disabled

Custom AMI ID: --

Information on security
groups and IAM roles