

# Running program on EMR

4 nodes:

Create Cluster:

1. Login to AWS Console.
2. Go to EMR.
3. Click on Create Cluster.

The screenshot shows the AWS EMR console interface for creating a new cluster. The browser tabs at the top include 'EMR - AWS Console', 'S3 Management Console', 'find subnet id aws - Go...', and 'Subnets | VPC Manager'. The address bar shows the URL: <https://us-east-2.console.aws.amazon.com/elasticmapreduce/home?region=us-east-2#quick-create>. The page title is 'Create Cluster - Quick Options' with a link to 'Go to advanced options'. The 'General Configuration' section includes a 'Cluster name' field with the value 'My cluster', a checked 'Logging' checkbox, and an 'S3 folder' field with the value 's3://aws-logs-773048014446-us-east-2/elasticmapred'. The 'Launch mode' is set to 'Cluster'. The 'Software configuration' section shows the 'Release' label as 'emr-5.9.0' and the 'Applications' section with four radio button options: 'Core Hadoop: Hadoop 2.7.3 with Ganglia 3.7.2, Hive 2.3.0, Hue 4.0.1, Mahout 0.13.0, Pig 0.17.0, and Tez 0.8.4' (selected), 'HBase: HBase 1.3.1 with Ganglia 3.7.2, Hadoop 2.7.3, Hive 2.3.0, Hue 4.0.1, Phoenix 4.11.0, and ZooKeeper 3.4.10', 'Presto: Presto 0.170 with Hadoop 2.7.3 HDFS and Hive 2.3.0 Metastore', and 'Spark: Spark 2.2.0 on Hadoop 2.7.3 YARN with Ganglia 3.7.2 and Zeppelin 0.7.2'. There is also an unchecked checkbox for 'Use AWS Glue Data Catalog for table metadata'. The 'Hardware configuration' section shows the 'Instance type' as 'm4.large'. The footer includes 'Feedback', 'English (US)', and copyright information for Amazon Web Services, Inc. from 2008 to 2017.

EMR - AWS Console x S3 Management Console x find subnet id aws - Go... x Subnets | VPC Manager x

Secure | <https://us-east-2.console.aws.amazon.com/elasticmapreduce/home?region=us-east-2#quick-create>

aws Services Resource Groups

shethdivani Ohio Support

### Create Cluster - Quick Options [Go to advanced options](#)

#### General Configuration

Cluster name

☒ Logging ⓘ

S3 folder  ⓘ

Launch mode ☒ Cluster ⓘ ☐ Step execution ⓘ

#### Software configuration

Release  ⓘ

Applications

- ☒ Core Hadoop: Hadoop 2.7.3 with Ganglia 3.7.2, Hive 2.3.0, Hue 4.0.1, Mahout 0.13.0, Pig 0.17.0, and Tez 0.8.4
- ☐ HBase: HBase 1.3.1 with Ganglia 3.7.2, Hadoop 2.7.3, Hive 2.3.0, Hue 4.0.1, Phoenix 4.11.0, and ZooKeeper 3.4.10
- ☐ Presto: Presto 0.170 with Hadoop 2.7.3 HDFS and Hive 2.3.0 Metastore
- ☐ Spark: Spark 2.2.0 on Hadoop 2.7.3 YARN with Ganglia 3.7.2 and Zeppelin 0.7.2

☐ Use AWS Glue Data Catalog for table metadata ⓘ

#### Hardware configuration

Instance type  ⓘ The selected instance type adds a default 30 GiB GP2

Feedback English (US) © 2008 - 2017, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

4. Go to advanced options

# Running program on EMR

EMR - AWS Console | S3 Management Console | find subnet id aws - Go | Subnets | VPC Manager

Secure | https://us-east-2.console.aws.amazon.com/elasticmapreduce/home?region=us-east-2#create-cluster:

aws Services Resource Groups

shethdhvani Ohio Support

### Create Cluster - Advanced Options [Go to quick options](#)

**Step 1: Software and Steps**

Step 2: Hardware

Step 3: General Cluster Settings

Step 4: Security

#### Software Configuration

Release: **emr-5.9.0**

☒ Hadoop 2.7.3 ☐ Zeppelin 0.7.2 ☐ Livy 0.4.0

☐ Tez 0.8.4 ☐ Flink 1.3.2 ☐ Ganglia 3.7.2

☐ HBase 1.3.1 ☐ Pig 0.17.0 ☒ Hive 2.3.0

☐ Presto 0.184 ☐ ZooKeeper 3.4.10 ☐ Sqoop 1.4.6

☐ Mahout 0.13.0 ☐ Hue 4.0.1 ☐ Phoenix 4.11.0

☐ Oozie 4.3.0 ☐ Spark 2.2.0 ☐ HCatalog 2.3.0

Edit software settings (optional)

☒ Enter configuration ☐ Load JSON from S3

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

Add steps (optional)

Step type: **Select a step** [Configure](#)

☐ Auto-terminate cluster after the last step is completed

[Cancel](#) [Next](#)

Feedback English (US) © 2008 - 2017, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

5. Select as shown above and click on Next.

6. Select and/or edit the below values:

EMR - AWS Console | S3 Management Console | find subnet id aws - Go | Subnets | VPC Manager

Secure | https://us-east-2.console.aws.amazon.com/elasticmapreduce/home?region=us-east-2#create-cluster:

aws Services Resource Groups

shethdhvani Ohio Support

### Create Cluster - General Cluster Settings

Step 3: General Cluster Settings

Step 4: Security

#### 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-e37c4c8a (172.31.0.0/16) (default)** [Create a VPC](#)

EC2 Subnet: **subnet-0fa1a066 | Default in us-east-2a**

Root device EBS volume size: **100 GiB**

Node type	Instance type	Instance count	Purchasing option	Auto Scaling
Master Master - 1	m4.xlarge 8 vCPU, 16 GiB memory, EBS only storage EBS Storage: 32 GiB	1 Instances	<input checked="" type="radio"/> On-demand <input type="radio"/> Spot Maximum bid price: \$	Not available for Master
Core Core - 2	m4.2xlarge 16 vCPU, 32 GiB memory, EBS only storage EBS Storage: 32 GiB	4 Instances	<input checked="" type="radio"/> On-demand <input type="radio"/> Spot Maximum bid price: \$	Not enabled
Task Task - 3	m4.large 4 vCPU, 8 GiB memory, EBS only storage EBS Storage: 32 GiB	0 Instances	<input checked="" type="radio"/> On-demand <input type="radio"/> Spot Maximum bid price: \$	Not enabled

[Add task instance group](#)

Feedback English (US) © 2008 - 2017, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

7. Click on next.

# Running program on EMR

Step 1: Software and Steps  
Step 2: Hardware  
**Step 3: General Cluster Settings**  
Step 4: Security

### General Options

Cluster name:

☒ Logging ⓘ  
S3 folder:

☒ Debugging ⓘ  
☒ Termination protection ⓘ

Scale down behavior:  ⓘ

### Tags ⓘ

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

### Additional Options

☐ EMRFS consistent view ⓘ

Custom AMI ID:  ⓘ

▶ Bootstrap Actions

[Cancel](#) [Previous](#) [Next](#)

8. Give a cluster name and click on next.

Create Cluster - Advanced Options [Go to quick options](#)

Step 1: Software and Steps  
Step 2: Hardware  
Step 3: General Cluster Settings  
**Step 4: Security**

### Security Options

EC2 key pair:  ⓘ

☒ Cluster visible to all IAM users in account ⓘ

### 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](#) ⓘ

▶ Encryption Options

▶ EC2 Security Groups

[No EC2 key pair has been selected, so you will not be able to SSH to this cluster. Learn how to create an EC2 Key Pair.](#)

[Cancel](#) [Previous](#) [Create cluster](#)

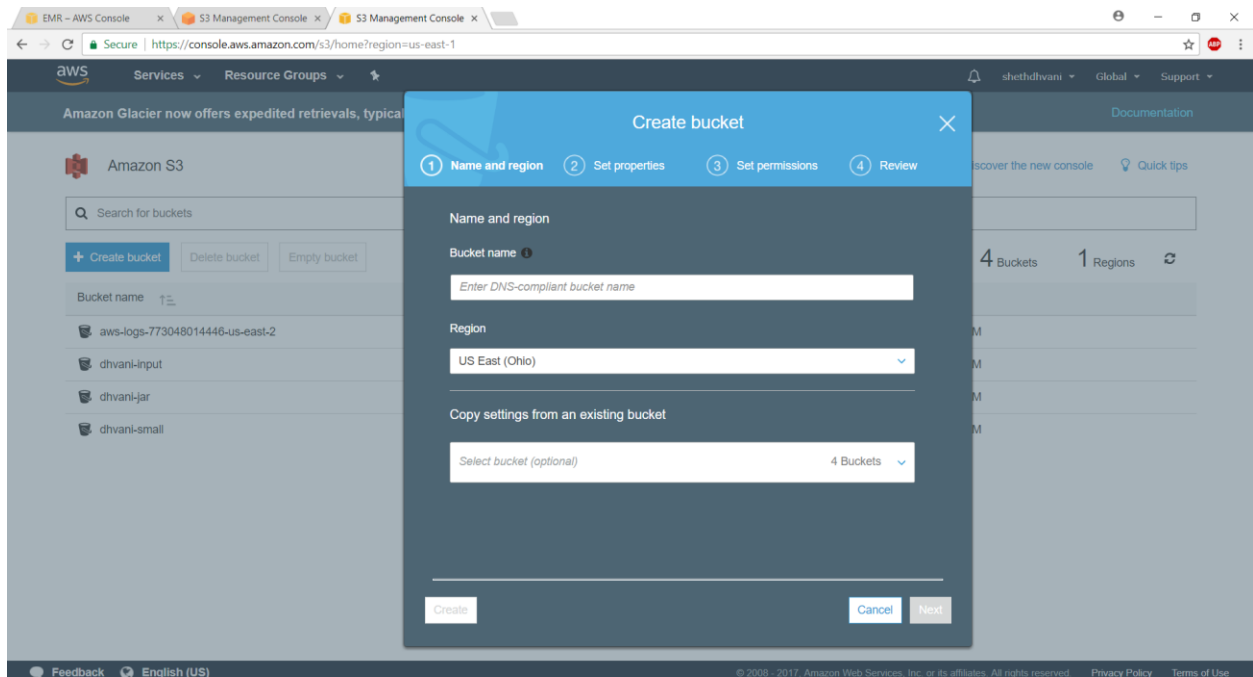
If there is an EC2 key pair already setup, select that. If not proceed without an EC2 key pair.

9. Click on create cluster.

# Running program on EMR

Add files to S3 bucket:

1. Now, go to S3 and click on create bucket.



2. Give a unique bucket name. The region should match the region in which cluster was created in the previous step.
3. Click on next till create bucket.
4. We need to create another bucket also. So one bucket is for input files and one is for jar.
5. Add the input files and jar in the respective buckets.

Running the job:

1. Go to Amazon EMR. Click on steps tab.

# Running program on EMR

The screenshot shows the AWS EMR console for a cluster named A3MR. The cluster is in a 'Waiting' state, indicating it is ready after the last step completed. The console displays various tabs for cluster management, including Summary, Application history, Monitoring, Hardware, Events, Steps, Configurations, and Bootstrap actions. The Summary tab is selected, showing details such as the cluster ID (j-3FMU64CY78RAF), creation date (2017-10-06 12:04 UTC-4), and elapsed time (10 minutes). It also lists the master public DNS, tags, and security configurations. The Configuration details section shows the release label (emr-5.9.0), Hadoop distribution (Amazon 2.7.3), and applications. The Network and hardware section shows the availability zone (us-east-2a) and subnet (subnet-0fa1a066). The Security and access section lists the key name (MRInstanceKeyPair), EC2 instance profile (EMR\_EC2\_DefaultRole), and EMR role (EMR\_DefaultRole).

Amazon EMR

Clusters

Security configurations

VPC subnets

Events

Help

Cluster: A3MR **Waiting** Cluster ready after last step completed.

Summary Application history Monitoring Hardware Events Steps Configurations Bootstrap actions

Connections: [Enable Web Connection](#) - Resource Manager ... (View All)

Master public DNS: ec2-18-221-203-65.us-east-2.compute.amazonaws.com [SSH](#)

Tags: -- [View All / Edit](#)

**Summary**

ID: j-3FMU64CY78RAF

Creation date: 2017-10-06 12:04 (UTC-4)

Elapsed time: 10 minutes

Auto-terminate: No

Termination protection: [Change](#)

**Configuration details**

Release label: emr-5.9.0

Hadoop distribution: Amazon 2.7.3

Applications: --

Log URI: s3://aws-logs-773048014446-us-east-2/elasticmapreduce/

EMRFS consistent view: Disabled

Custom AMI ID: --

**Network and hardware**

Availability zone: us-east-2a

Subnet ID: subnet-0fa1a066

Master: **Running** 1 m4.xlarge

Core: **Running** 4 m4.2xlarge

Task: --

**Security and access**

Key name: MRInstanceKeyPair

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-80bc3ee8 (ElasticMapReduce-Master: master)

Security groups for sg-f6bd3f9e (ElasticMapReduce-Core & Task: slave)

Feedback English (US) © 2008 - 2017, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

The screenshot shows the AWS EMR console for the same cluster A3MR, but with the 'Steps' tab selected. The 'Add step' button is highlighted, and a table of steps is displayed. The table has columns for ID, Name, Status, Start time (UTC-4), Elapsed time, Log files, and Actions. One step is listed with ID s-2LSPUOXKQ52EX, Name 'Setup hadoop debugging', Status 'Completed', Start time '2017-10-06 12:08 (UTC-4)', and Elapsed time '2 seconds'. The 'Log files' column has a link to 'View logs', and the 'Actions' column has a link to 'View jobs'.

Amazon EMR

Clusters

Security configurations

VPC subnets

Events

Help

Cluster: A3MR **Waiting** Cluster ready after last step completed.

Summary Application history Monitoring Hardware Events Steps Configurations Bootstrap actions

[Add step](#) [Clone step](#) [Cancel step](#)

Steps

[View all interactive jobs](#) | [View all jobs](#)

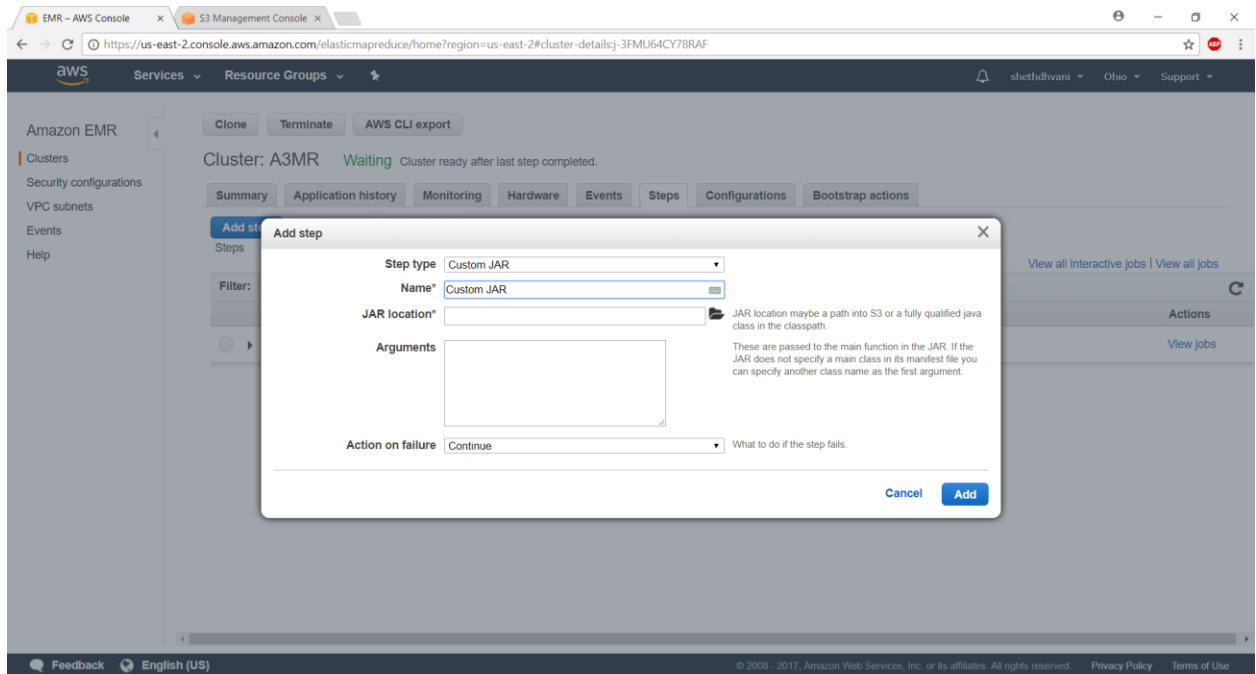
Filter: All steps  1 step (all loaded)

ID	Name	Status	Start time (UTC-4)	Elapsed time	Log files	Actions
s-2LSPUOXKQ52EX	Setup hadoop debugging	Completed	2017-10-06 12:08 (UTC-4)	2 seconds	<a href="#">View logs</a>	<a href="#">View jobs</a>

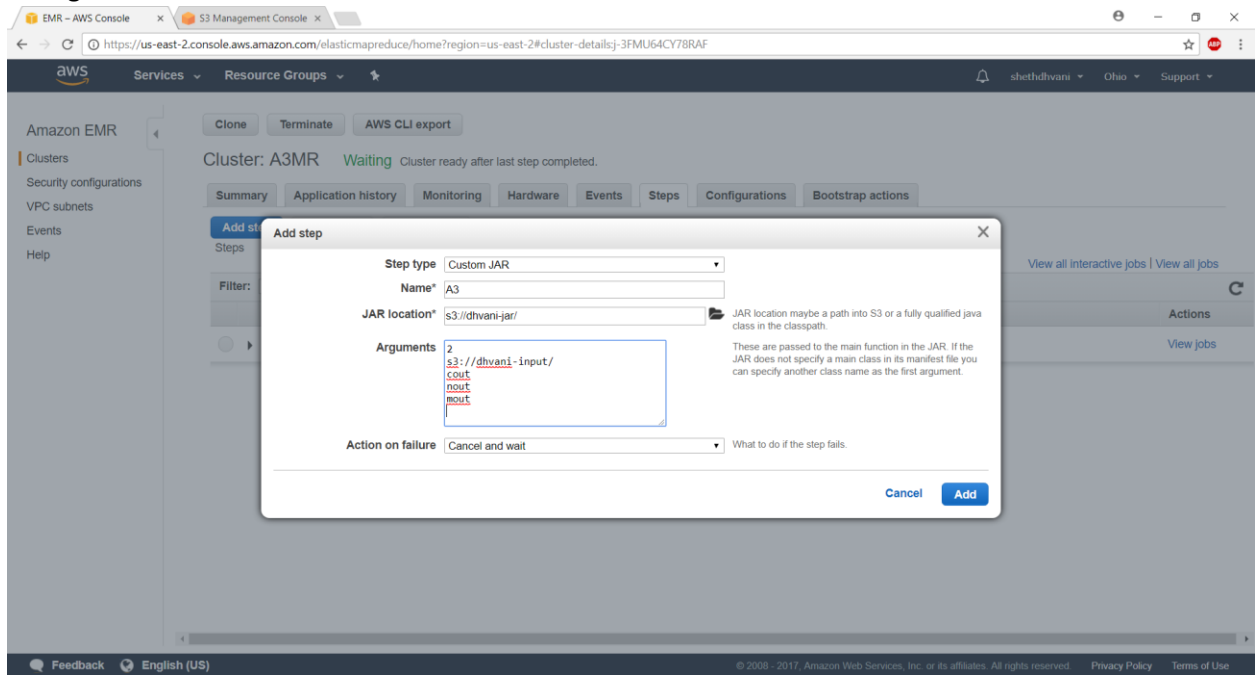
Feedback English (US) © 2008 - 2017, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

2. Click on Add step.

# Running program on EMR



3. Give a name. Select jar location from S3. Give the arguments.
4. Change Action on failure to Cancel and wait



5. Click on Add and the job will start.