

Step 5: Launch Your Initial Amazon EMR Cluster

In this step, you launch your initial cluster by using “Create Cluster” in the Amazon EMR console and leaving most options to their default values.

To launch the sample Amazon EMR cluster

1. Find the EMR console page

2. Choose **Create cluster**.

3. On the **Create Cluster** page, accept the default values for most of the values except for the following fields

- Enter a Cluster name that helps you identify the cluster, for example, My First EMR Cluster. **Select Core Hadoop**.

Create cluster [Info](#)

Name and applications [Info](#)

Name

Amazon EMR release [Info](#)

A release contains a set of applications which can be installed on your cluster.

emr-6.13.0

Application bundle

Spark

Core Hadoop

Flink

HBase

Presto

Trino

Custom

Cluster configuration [Info](#)

Choose a configuration method for the primary, core, and task node groups for your cluster.

☒ Instance groups
Choose one instance type per node group

☐ Instance fleets
Choose any combination of instance types within each node group

Instance groups

Primary

Choose EC2 instance type

m4.large
2 vCore 8 GiB memory EBS only storage
On-Demand price: \$0.100 per instance/hour
Lowest Spot price: \$0.036 (us-east-2a)

Actions ▼

☐ Use multiple primary nodes
To improve cluster availability, use 3 primary nodes with the same configuration and bootstrap actions. You can not use multiple primary nodes with instance fleets.

► Node configuration - optional

Core

Choose EC2 instance type

m4.large
2 vCore 8 GiB memory EBS only storage
On-Demand price: \$0.100 per instance/hour
Lowest Spot price: \$0.036 (us-east-2a)

Actions ▼

► Node configuration - optional

- Under **Cluster configuration**, choose:

- The Instance groups.
- The Instance type as: m4.large
- The Number of instances as: 2

Task instance is optional.

- Under Cluster termination, enter the time until your cluster terminates which is recommended or manually terminate the cluster if you would like to work around more with clusters.

- Under **Security configuration and EC2 key pair** - Select the Amazon EC2 Key Pair that you created.

- Under **IAM roles** - Select EMR_DefaultRole for ServiceRole and EMR_EC2_DefaultRole for Instance profile.

Security configuration and EC2 key pair - optional [Info](#)

Security configuration

Select your cluster encryption, authentication, authorization, and instance metadata service settings.


 Choose a security configuration



Browse 

Create security configuration 

Amazon EC2 key pair for SSH to the cluster [Info](#)

 emr-key-pair



Browse

Create key pair 

Identity and Access Management (IAM) roles [Info](#)

Choose or create a service role and instance profile for the EC2 instances in your cluster.

Amazon EMR service role [Info](#)

The service role is an IAM role that Amazon EMR assumes to provision resources and perform service-level actions with other AWS services.

☒ Choose an existing service role

Select a default service role or a custom role with IAM policies attached so that your cluster can interact with other AWS services.

☐ Create a service role

Let Amazon EMR create a new service role so that you can grant and restrict access to resources in other AWS services.

Service role

EMR_DefaultRole



EC2 instance profile for Amazon EMR

The instance profile assigns a role to every EC2 instance in a cluster. The instance profile must specify a role that can access the resources for your steps and bootstrap actions.

☒ Choose an existing instance profile

Select a default role or a custom instance profile with IAM policies attached so that your cluster can interact with your resources in Amazon S3.

☐ Create an instance profile

Let Amazon EMR create a new instance profile so that you can specify a custom set of resources for it to access in Amazon S3.

Instance profile

EMR_EC2_DefaultRole



4. Then Choose **Create cluster**.

Note your cluster is ready for use when, instead of “Starting” it says “Waiting Cluster ready after last step completed.” This could sometimes take 10+ minutes, so don’t worry.

The cluster status page with the cluster Summary appears (see below). You can use this page to monitor the progress of cluster creation and view details about cluster status. As cluster creation tasks finish, items on the status page update. You may need to choose the refresh icon (circular arrow) on the right or refresh your browser to receive updates

my-first-emr-cluster Updated less than a minute ago **Actions**

Summary

Cluster info	Applications	Cluster management	Status and time
Cluster ID j-FDEAZF8TDKYH	Amazon EMR version emr-6.13.0	Log destination in Amazon S3 aws-logs-299094137703-us-east-1/elasticmapreduce	Status Starting
Cluster configuration Instance groups	Installed applications Hadoop 3.3.3, Hive 3.1.3, Hue 4.11.0, Pig 0.17.0, Tez 0.10.2	Primary node public DNS ec2-18-209-178-167.compute-1.amazonaws.com	Creation time September 10, 2023, 17:52 (UTC-05:00)
Capacity 1 Primary 2 Core 2 Task			Elapsed time 2 minutes

Properties | Bootstrap actions | Instances (Hardware) | Steps | Applications | Configurations | Monitoring | Events | Tags (0)

Operating system [Info](#)

Amazon Linux release 2.0.20230808.0

Cluster logs [Info](#)

Archive log files to Amazon S3
Turned on

Amazon S3 location
[s3://aws-logs-299094137703-us-east-1/elasticmapreduce/](#) [Info](#)

Encryption for logs
Turned off

Cluster termination [Info](#) [Edit cluster termination](#)

Termination option
Automatically terminate cluster after idle time

Termination protection
Turned on

Idle time
1 hour

Network and security [Info](#)

Network

Virtual Private Cloud (VPC)
[vpc-007da42944502a084](#) [Info](#)

Subnet(s) and Availability Zone(s) (AZ)
[subnet-0009383cd07893114](#) [Info](#) us-east-1e

► EC2 security groups (firewall)

Security configuration

Security configuration
None

EC2 key pair
key-pair

Permissions

Service role for Amazon EMR
[EMR_DefaultRole](#) [Info](#)

EC2 instance profile
[EMR_EC2_DefaultRole](#)

Auto scaling role
Not configured

Under **Network and security**, find the **Primary and Core** instance status. As soon as you see the links for Security groups for Primary and Security Groups for Core & Task (see below), you can move on to the next task, but you may want to wait until the cluster starts and is in the Waiting state. The links are blue colored identifiers starting with “sg-“ Under **EC2 security groups**.

For more information about reading the cluster summary, see **View Cluster Status and Details**.

▼ EC2 security groups (firewall)

Primary node

EMR managed security group

[sg-001f6e96a09ee9d4b](#) 

Additional security groups

-

Core and task nodes

EMR managed security group







[sg-0356f24c649037c1e](#) 

Allow SSH Connections to the Cluster from Your Client Security groups act as virtual firewalls to control inbound and outbound traffic to your cluster. When you create your first cluster, Amazon EMR creates the default Amazon EMR-managed security group associated with the master instance, ElasticMapReduce-master, and the security group associated with core and task nodes, ElasticMapReduce-slave.


To reach ElasticMapReduce-master just click on the blue link associated with the **Security group for Primary node** and you should then see something like the following.

sg-001f6e96a09ee9d4b - ElasticMapReduce-master Actions ▼

Details

Security group name  ElasticMapReduce-master	Security group ID  sg-001f6e96a09ee9d4b	Description  Master group for Elastic MapReduce created on 2022-09-01T03:47:12.639Z	VPC ID  vpc-007da42944502a084 vpc-007da42944502a084 
Owner  299094137703	Inbound rules count 20 Permission entries	Outbound rules count 1 Permission entry	

Inbound rules Outbound rules Tags

Inbound rules (20)  Manage tags Edit inbound rules

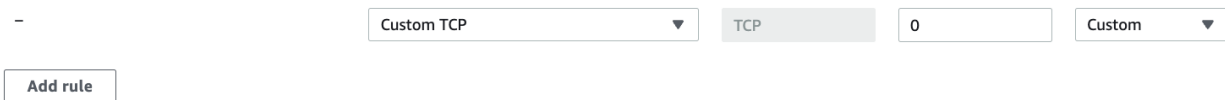
<input type="checkbox"/>	Name	Security group rule...	IP version	Type	Protocol	Port range	Source	Description
<input type="checkbox"/>	-	sg-058f44b6f03809739	-	All TCP	TCP	0 - 65535	sg-0356f24c649037c1...	-
<input type="checkbox"/>	-	sg-0605d18c148f2ea2b	IPv4	Custom TCP	TCP	8443	54.240.217.16/29	-
<input type="checkbox"/>	-	sg-06e41d73d69b2e...	IPv4	SSH	TCP	22	204.14.37.206/32	-
<input type="checkbox"/>	-	sg-0d69f81cbc31dff5c1	IPv4	Custom TCP	TCP	8443	207.171.172.6/32	-
<input type="checkbox"/>	-	sg-0eb0cd96c2b9ece3d	-	All UDP	UDP	0 - 65535	sg-001f6e96a09ee9d4...	-
<input type="checkbox"/>	-	sg-083df7f9c8472b91	IPv4	Custom TCP	TCP	8443	207.171.167.26/32	-
<input type="checkbox"/>	-	sg-03745c2712f02013d	IPv4	Custom TCP	TCP	8443	54.240.217.8/29	-
<input type="checkbox"/>	-	sg-0252a0177ed86c7...	IPv4	Custom TCP	TCP	8443	207.171.167.25/32	-
<input type="checkbox"/>	-	sg-04e9e979f995293c4	IPv4	Custom TCP	TCP	8443	72.21.198.64/29	-
<input type="checkbox"/>	-	sg-086546143855e1...	IPv4	Custom TCP	TCP	8443	72.21.217.0/24	-
<input type="checkbox"/>	-	sg-0b582c16095e242...	IPv4	Custom TCP	TCP	8443	54.240.217.80/29	-
<input type="checkbox"/>	-	sg-00c511aa3b46bd4...	-	All UDP	UDP	0 - 65535	sa-0356f24c649037c1...	-

For more information about security groups, see Control Network Traffic with Security Groups and Security Groups for Your VPC in the Amazon VPC User Guide..

Select the “Edit Inbound rules” option to the top right corner of the Inbound rules section.

A new pane will appear allowing you to modify access rules. Scroll down to the bottom of the list where you will see the “Add rule” button. Select it.

A line for you to enter a new access rule will appear:



The screenshot shows a configuration interface for a new access rule. It includes a dropdown menu with 'Custom TCP' selected, a 'TCP' button, a text input field with '0', and another dropdown menu with 'Custom' selected. Below these fields is an 'Add rule' button.

1. Select the field with label “Custom TCP” which pops up a list of options, select “SSH”. When you do the next field to its left will display the value “TCP” and the next field to the left of that will show “22”.
2. Now select the next field showing the value “Custom” which pops up a list from which you should select “My IP” which causes your IP to be the only one allowed to access your EMR cluster via SSH (or SCP). Scroll down a bit more, if needed, and click on the “Save rules” button.



The screenshot shows the completed configuration for the access rule. The dropdown menu now shows 'SSH', the 'TCP' button is still present, the text input field now shows '22', and the next dropdown menu shows 'My IP'. Below these fields is a search bar with '208.59.145.16/32' entered and a 'Delete' button. At the bottom of the interface are 'Cancel', 'Preview changes', and 'Save rules' buttons. The 'Save rules' button is highlighted with an orange border.

Note, once you have set up this rule, in most cases when you create a new cluster, it will use the same security group, so you likely will not need to set up this rule again. But it is always good to check.