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.

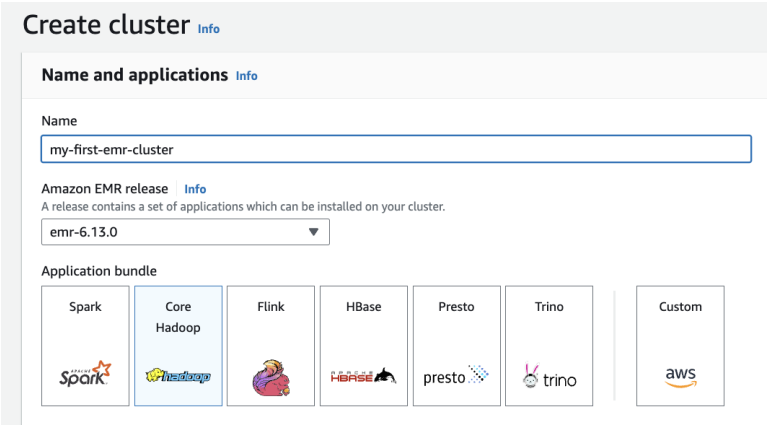
To launch the sample Amazon EMR cluster

1. Find the EMR console page

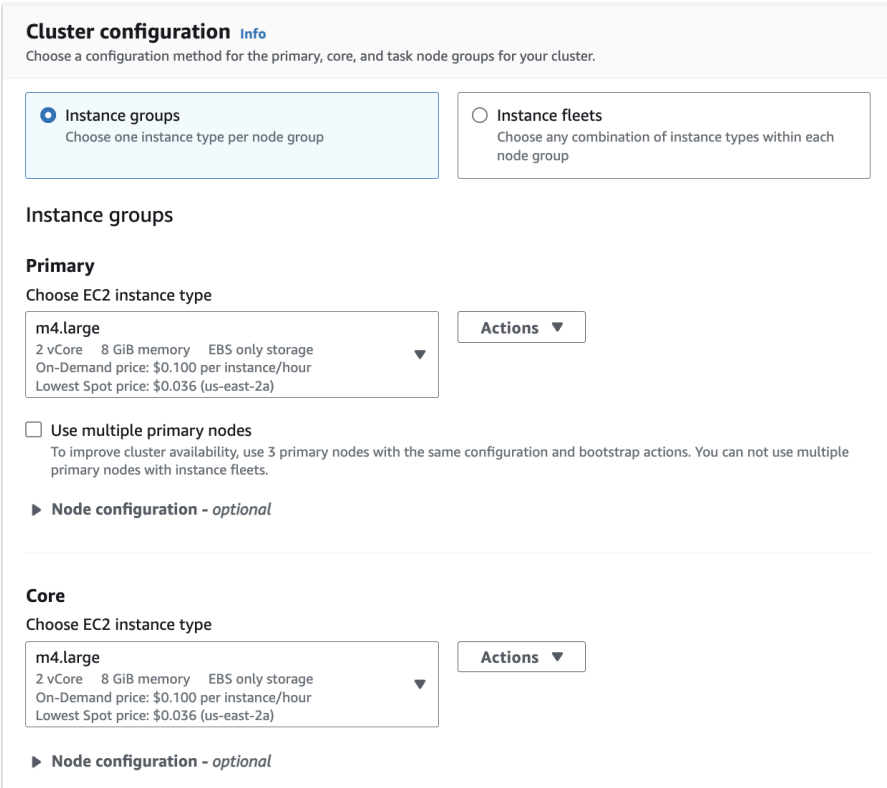
2. Choose **Create cluster**.

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**.



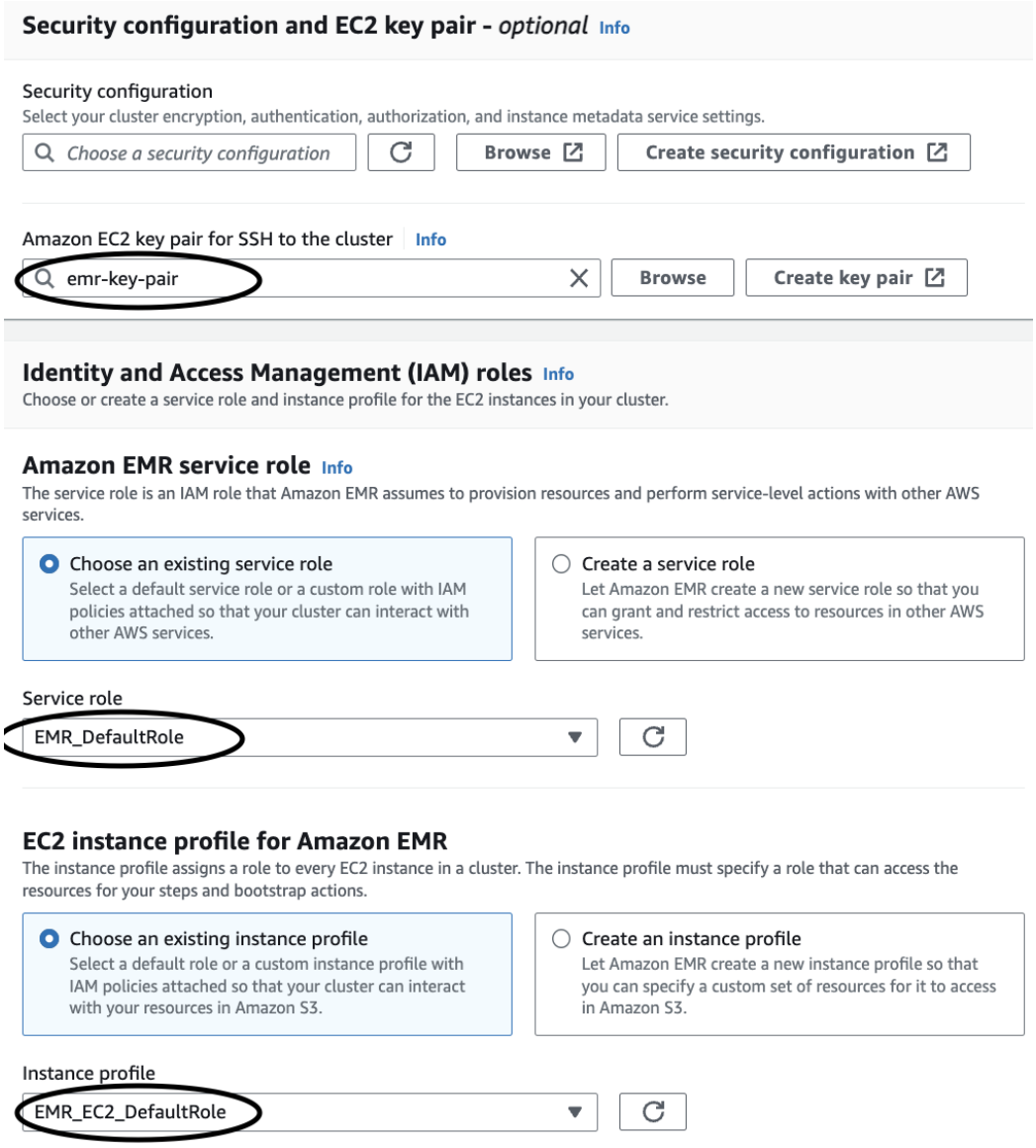
:



• 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.

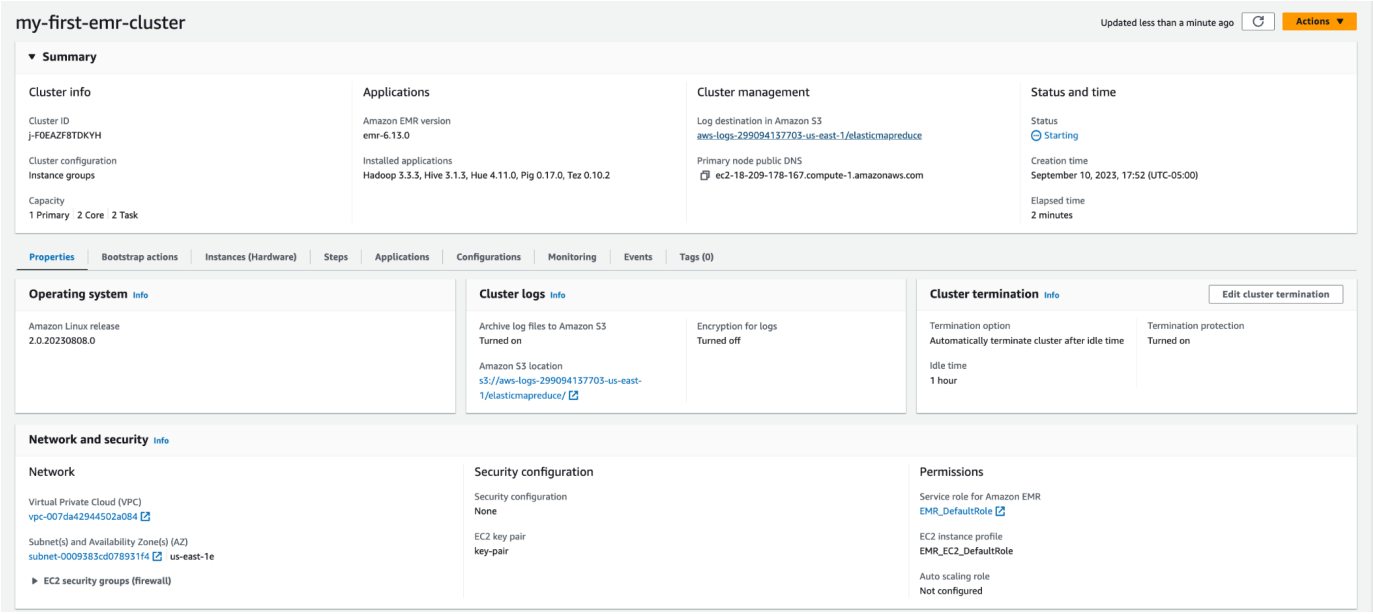
• Under **Security configuration and EC2 key pair** - Select the Amazon EC2 Key Pair that you created.
• Under **IAM roles** - Select EMR_DefaultRole for ServiceRole and EMR_EC2_DefaultRole for Instance profile.



4. Then Choose **Create cluster.**

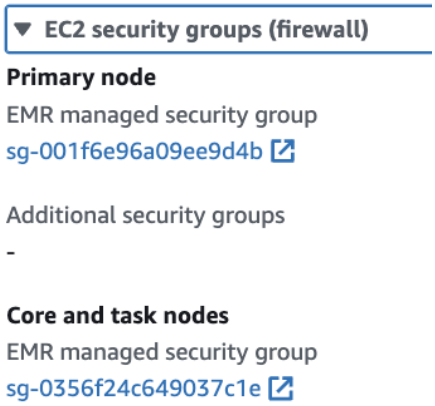
Note your cluster is ready for use when, instead of “Starting” it says “Waiting Cluster ready after last step completed.“ This could sometimes take 10+ minutes, so don’t worry.

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



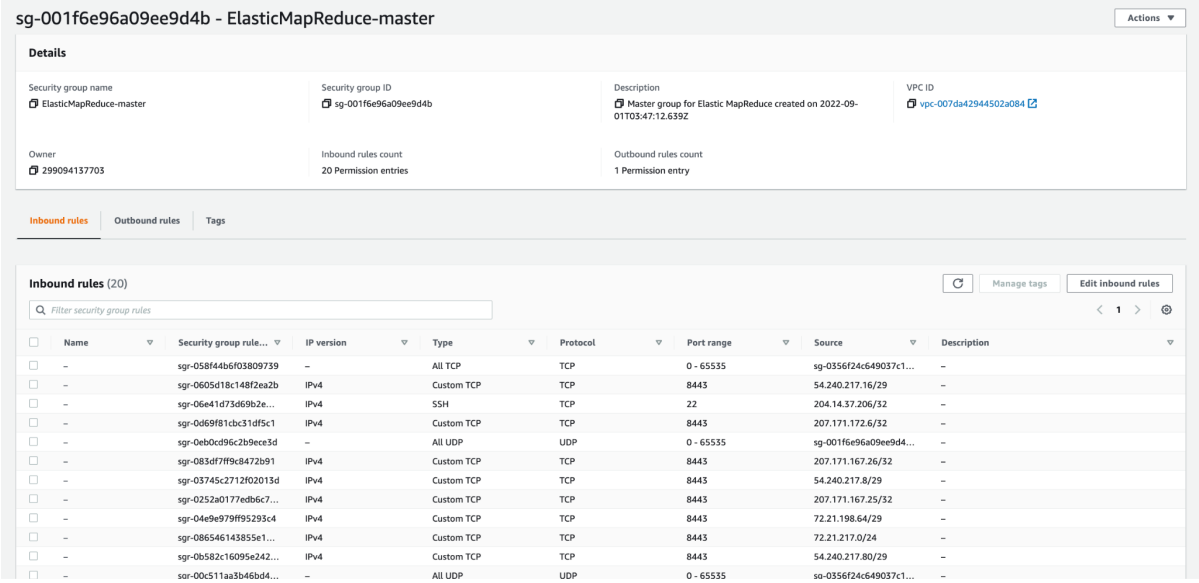
Under **Network and security**, find the **Primary and Core** instance status. As soon as you see the links for Security groups for Primary and Security Groups for Core & Task (see below), you can move on to the next task, but you may want to wait until the cluster starts and is in the Waiting state.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



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.

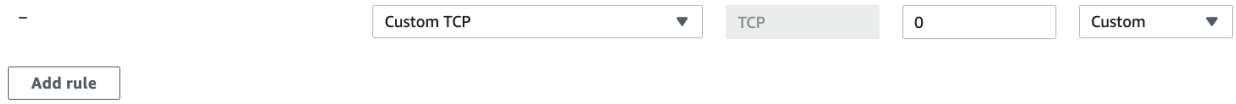


For more information about security groups, see Control Network Traffic with Security Groups and Security Groups for Your VPC in the Amazon VPC User Guide..

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:



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.

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.