Cloud Assignment - 1

Name : Noopur Rajesh Kumar Kalawatia

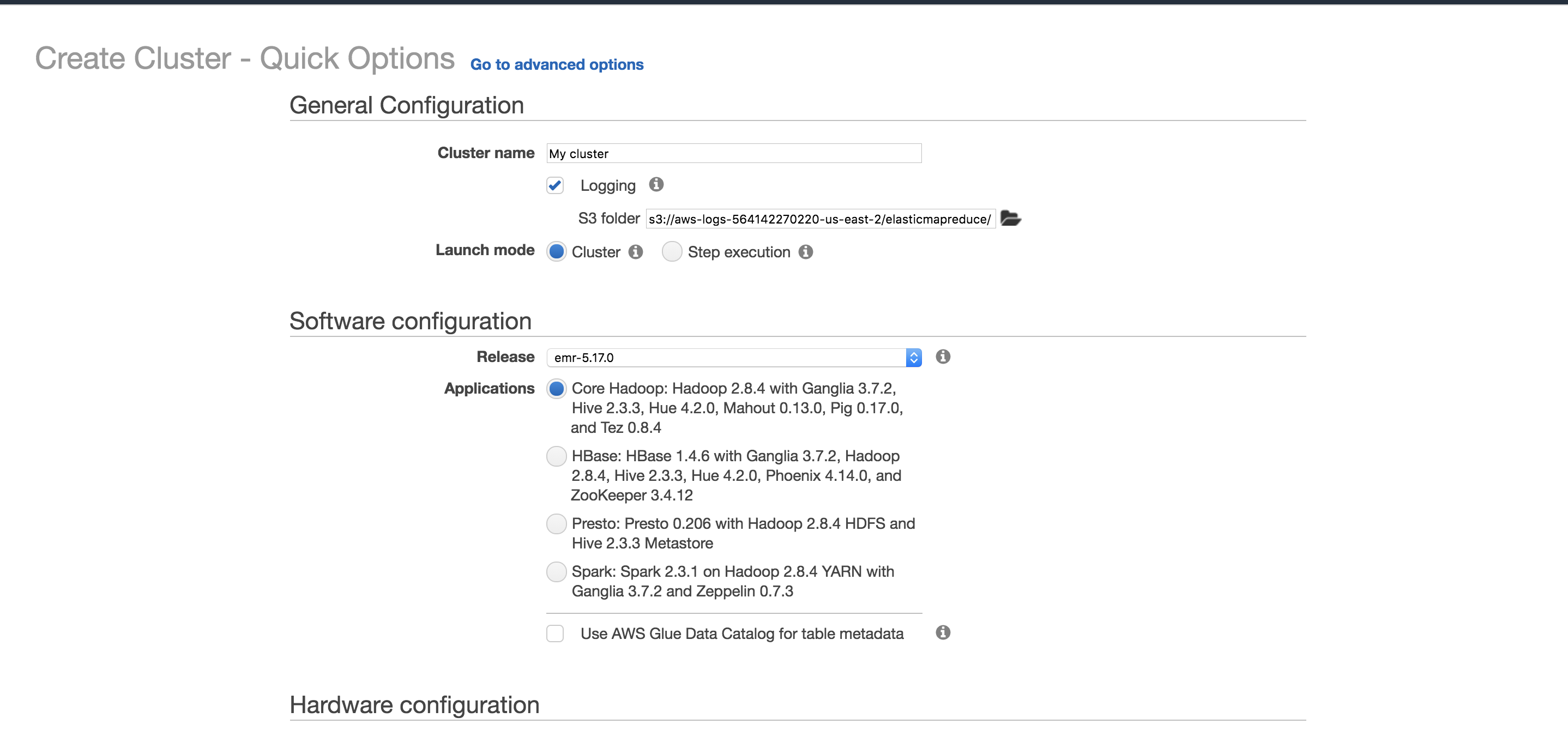
UFID : 1980 - 9834

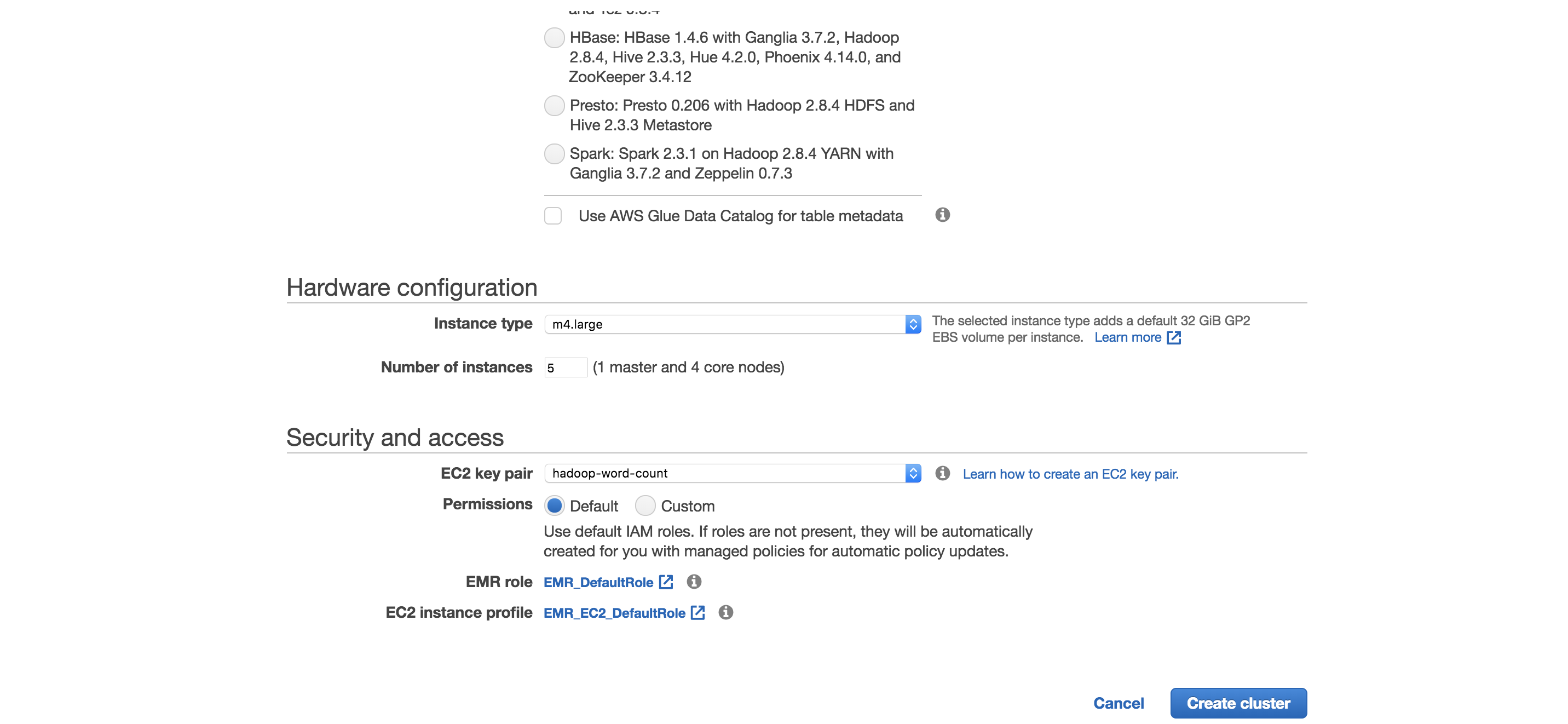
The steps that I followed to realize the problem statement are as follows,

Platform used : Amazon AWS

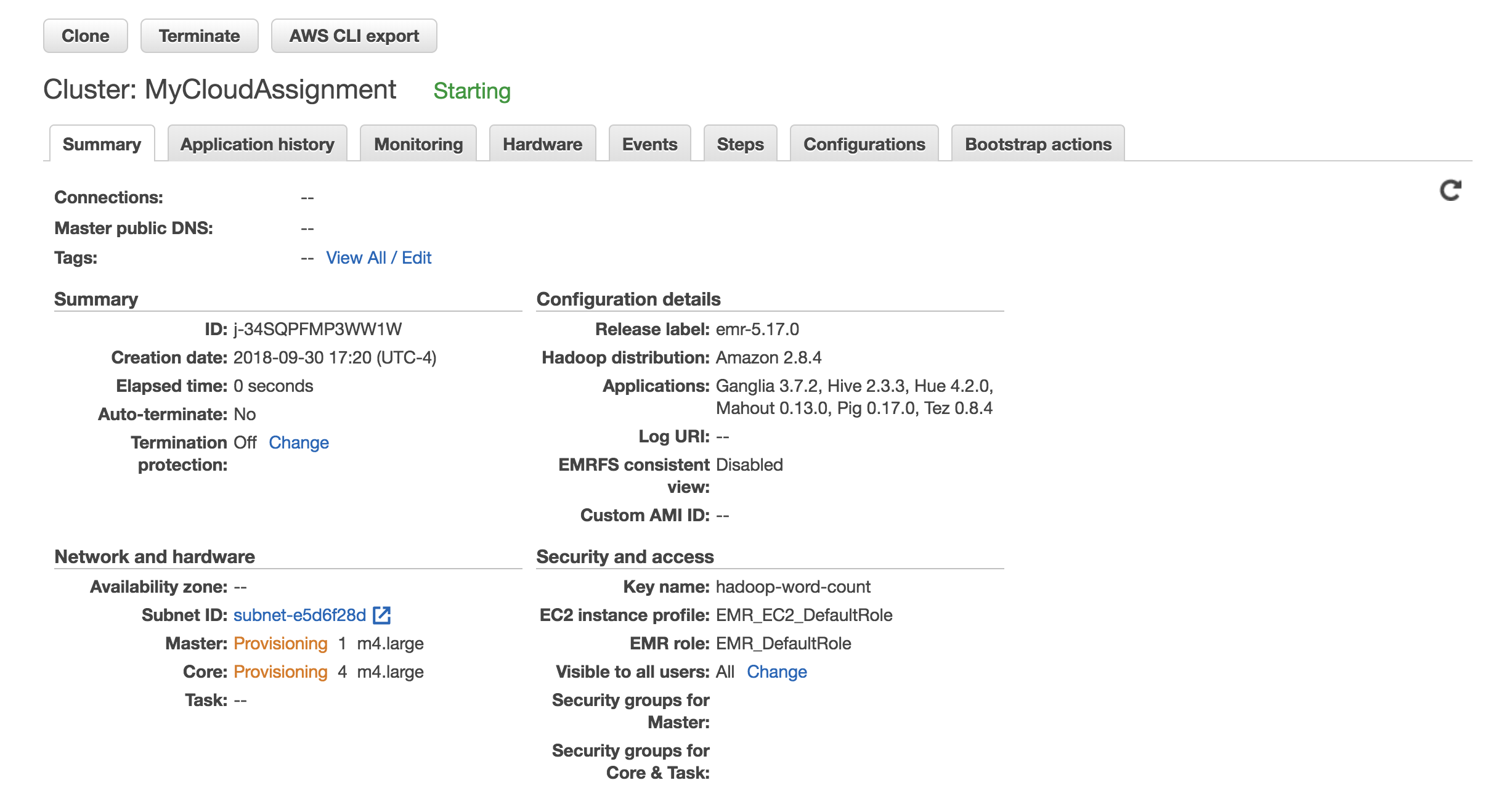
Product used : Amazon EMR

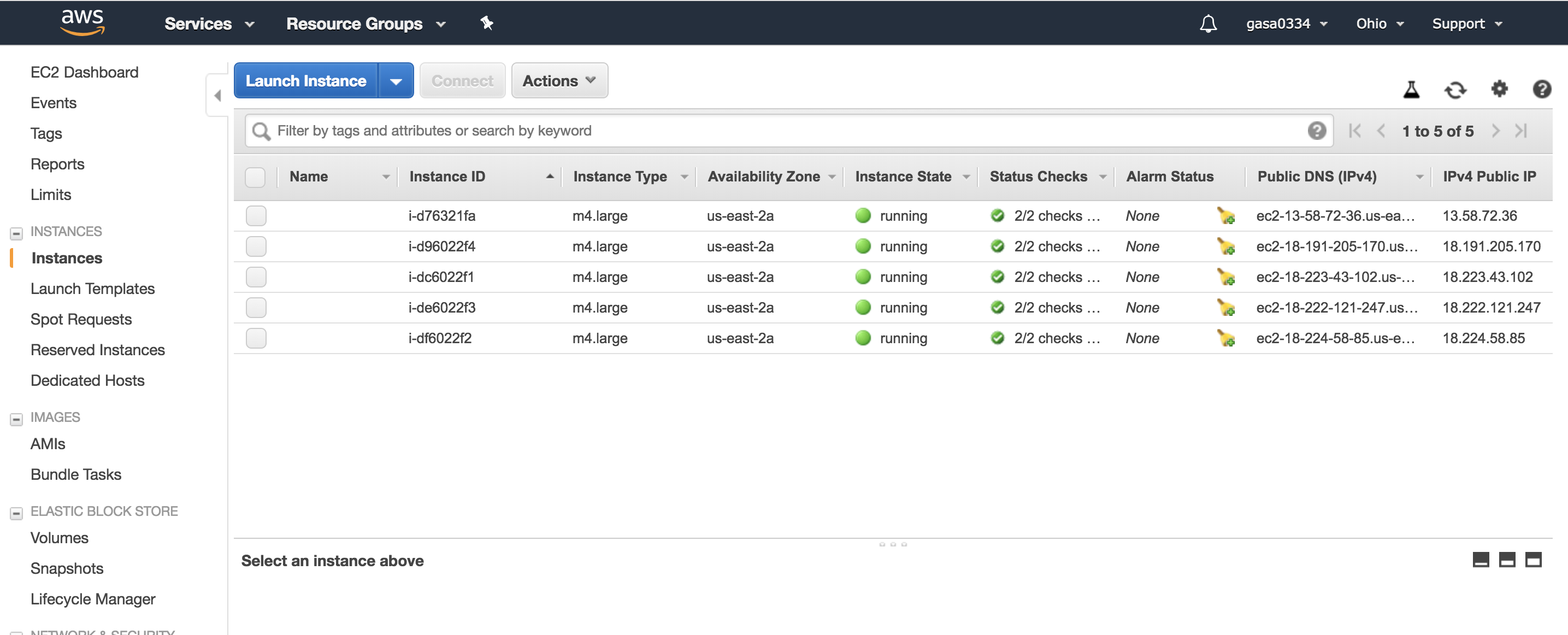
1. **Create a cluster in Amazon cluster.**

a. Input the desired name fo the cluster.

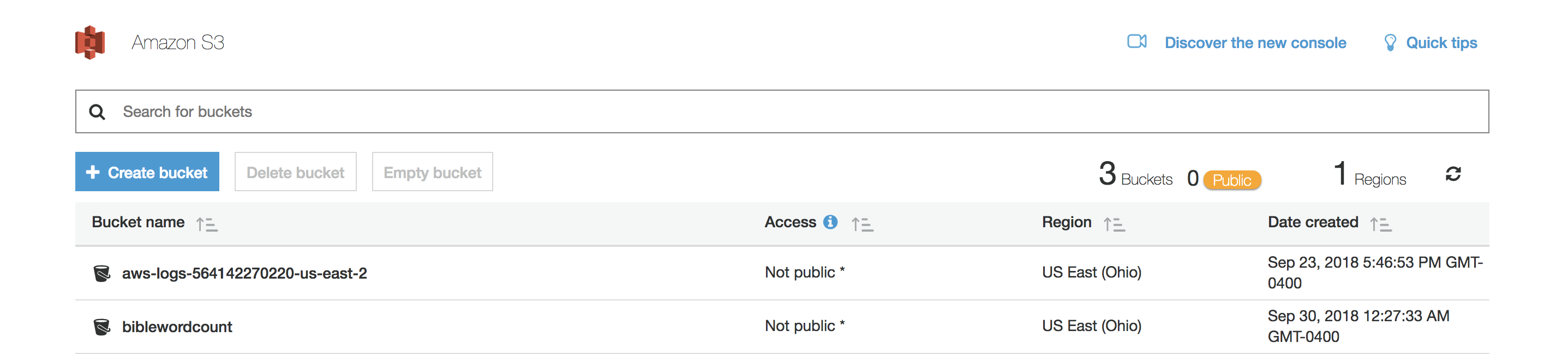
For this assignment we have selected the core Hadoop type of cluster.

The field - Number of instances in the above picture gives us an indication of the number of EC2 instances that was will create for us.

For this assignment , I have selected 5, which translates to 1 master node and 4 slave nodes.

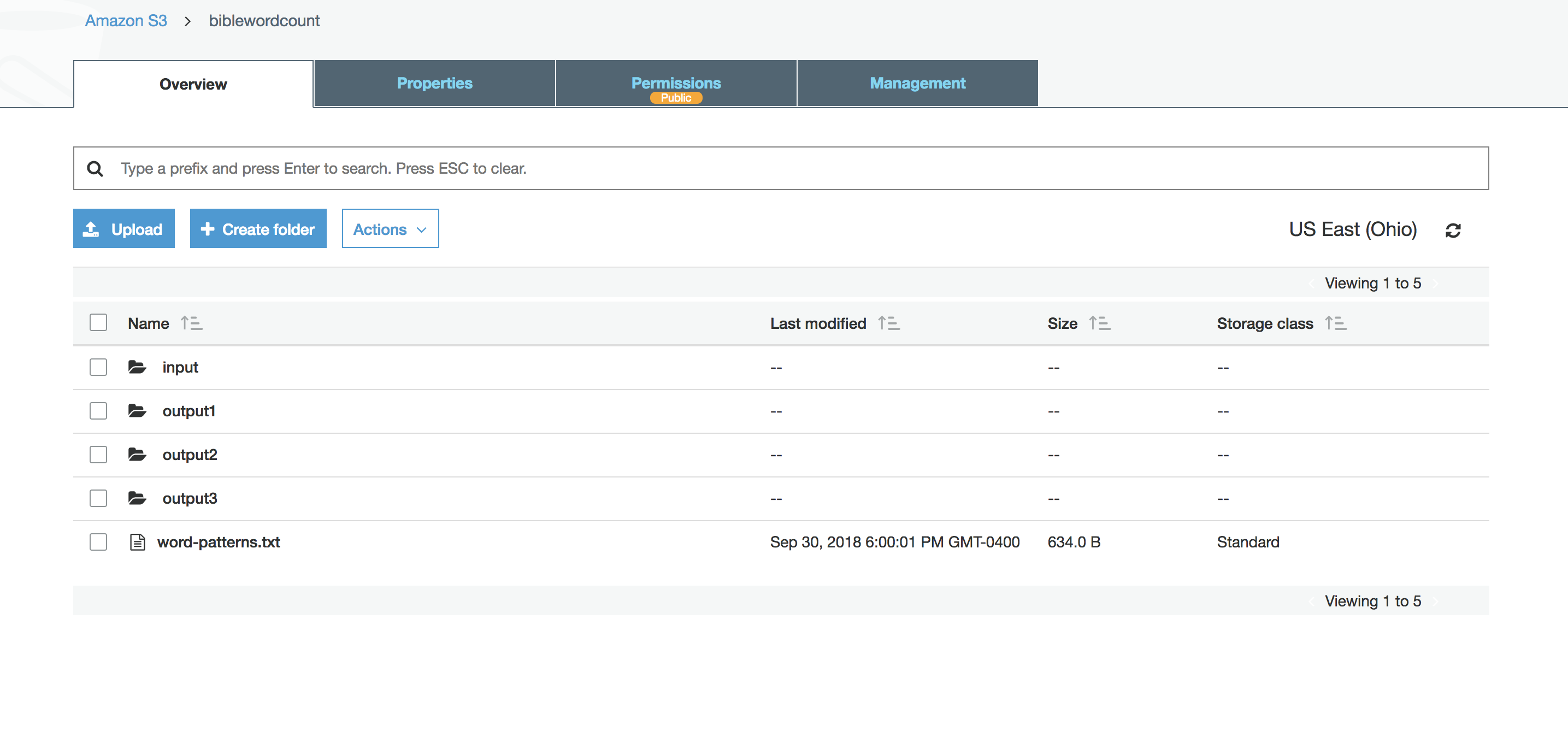
Thus after you have created the cluster you can go to your EC2 dashboard and observe the instances that are currently,

**2. Create a S3 bucket that will serve as input and output storage for your assignment.**

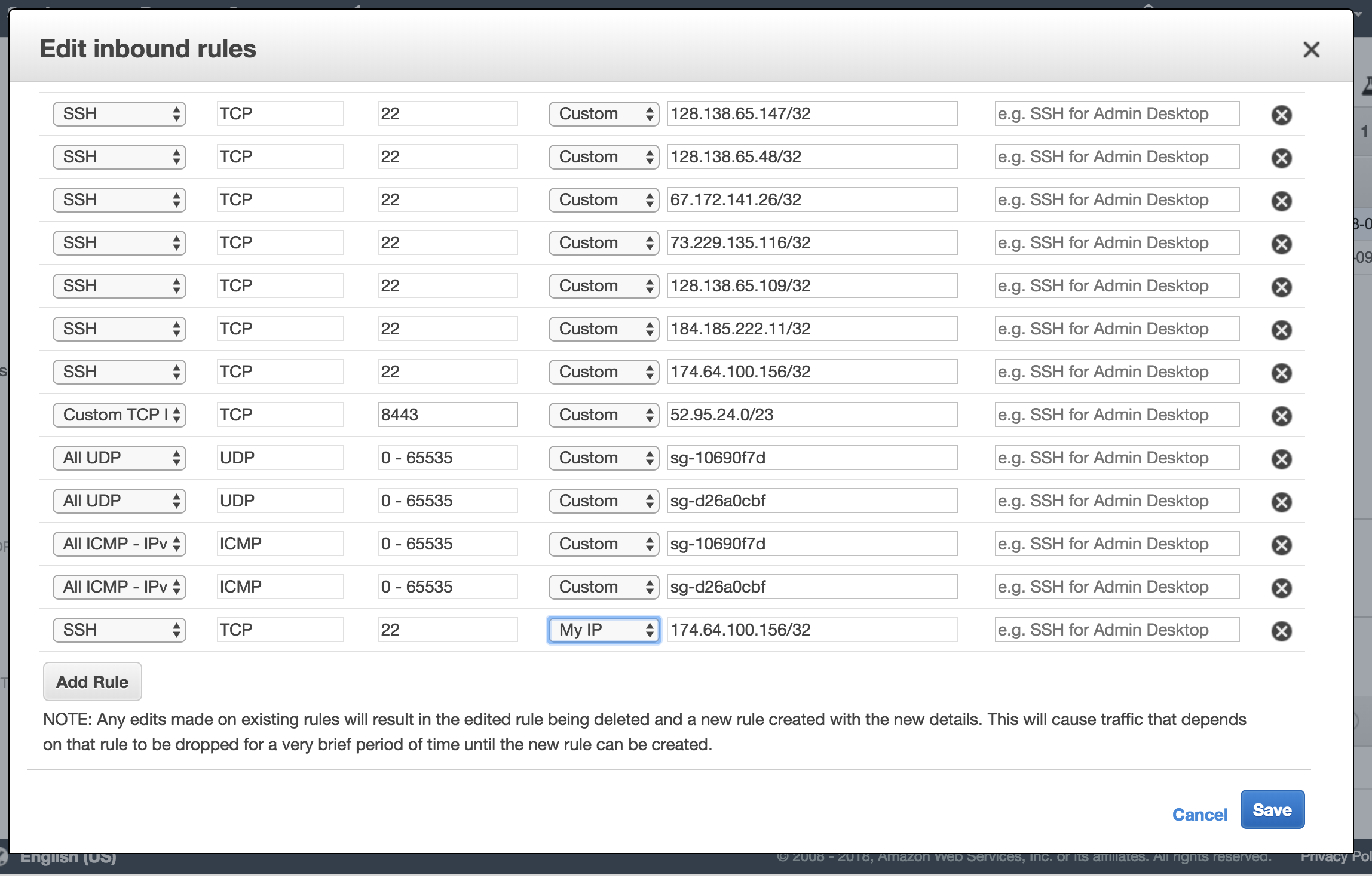
a. Find the product - S3, in AWS. Create a bucket with the desired name.

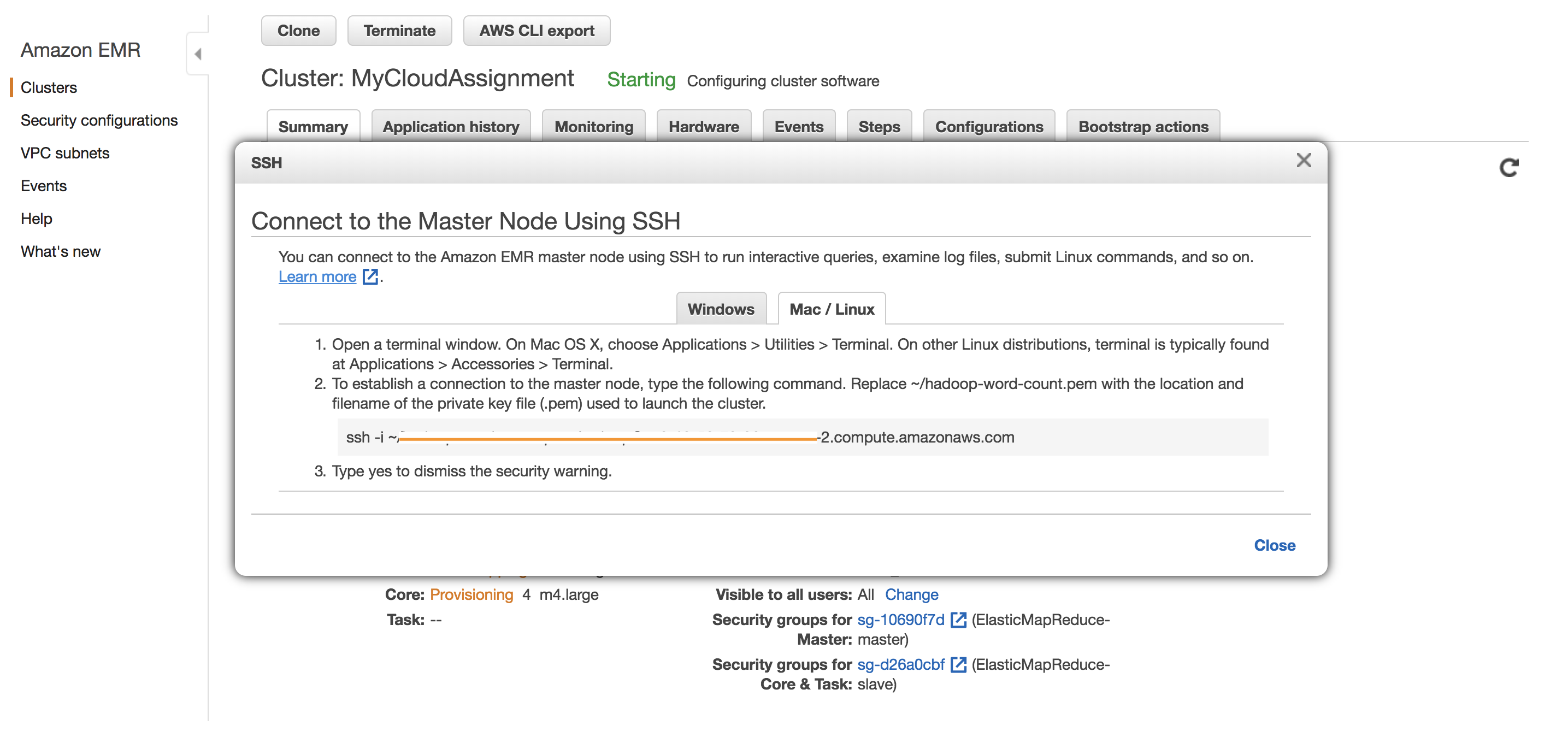
The bucket for our project is biblewordcount.

The input and output are arranged in the above fashion (Note that : when you begin with creating the s3 bucket only the input is uploaded, the folder structure for output is created after the map reduce job is successfully executed).



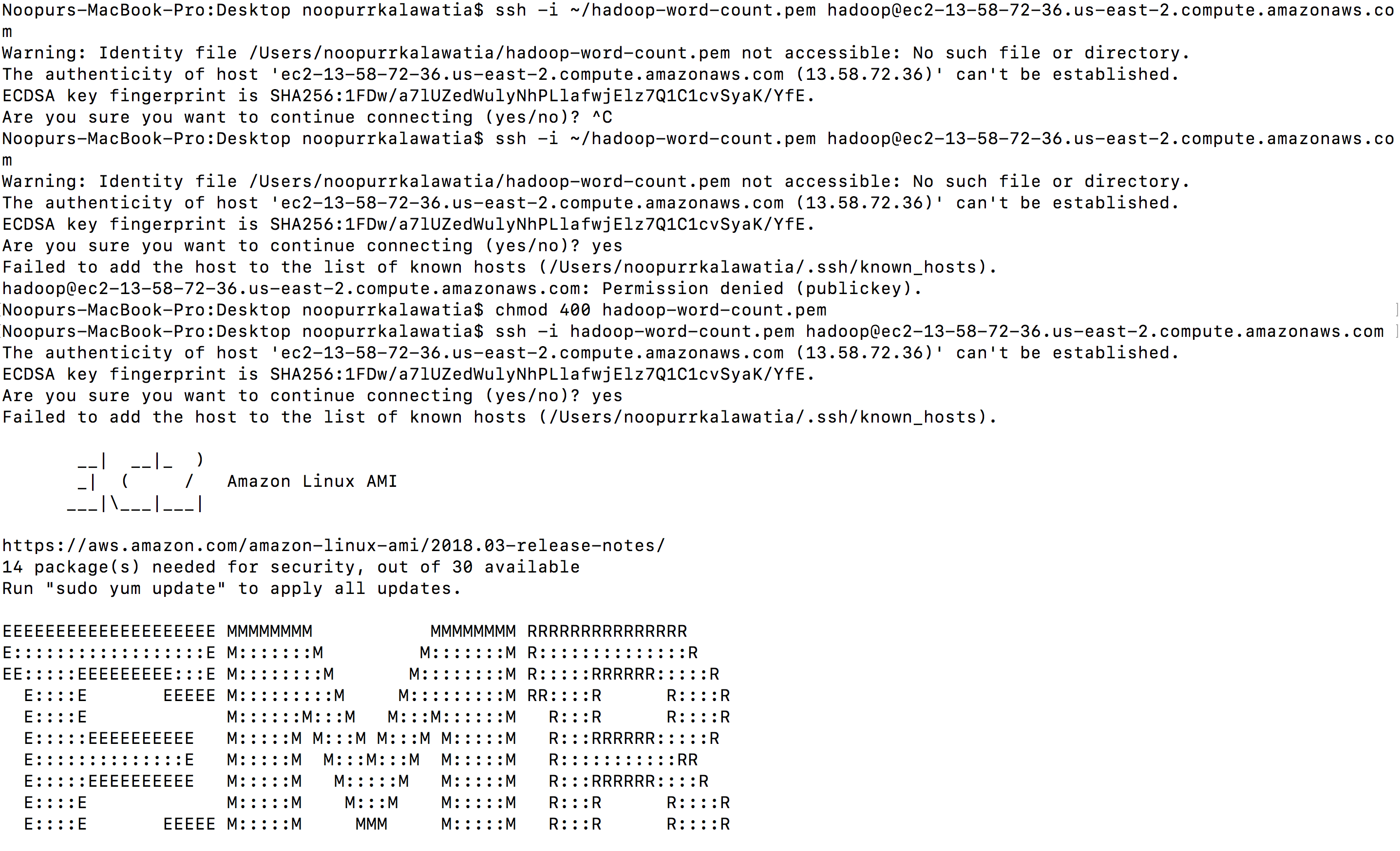
**3. Establish the SSH connection from your local machine to EMR. To accomplish this we need to configure the inbound for the master node of your EMR.**

The command to run can be found from the EMR cluster configuration 



This will establish your SSH connection to the master node and thus you can run your map reduce program.

**3. Perform the Hadoop map reduce job**

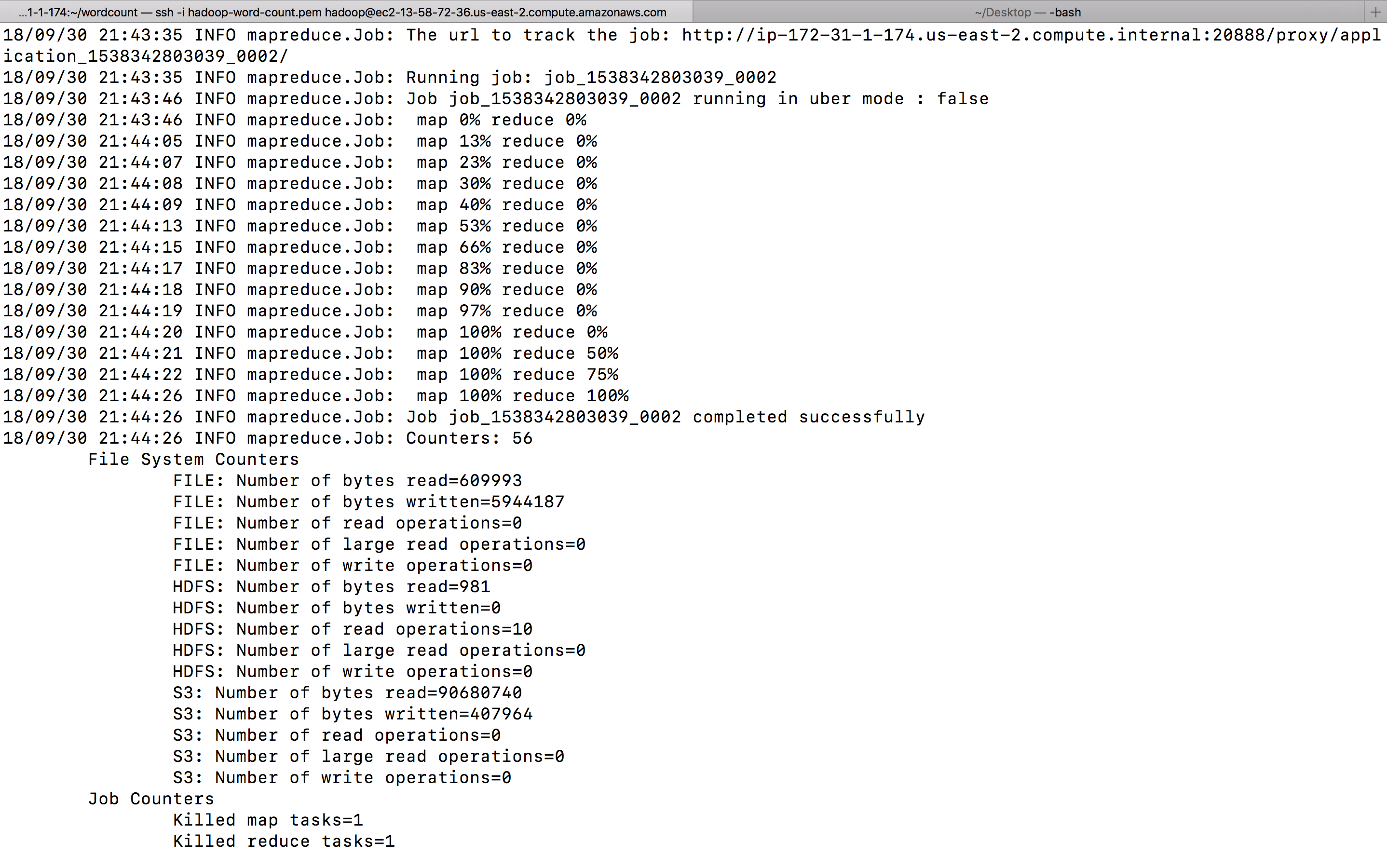


You are basically in the Hadoop instance of your EMR master node. To upload your jar to perform the Hadoop map reduce job you can use the following command,

**ssh -i <certificate> <file name> <hadoop master node name and path where you want to place your jar file>**

Now that the jar file is uploaded you can run the following command to perform your Hadoop map reduce job.

**hadoop jar <your jar file> <the class that deduces the calculation in your jar> <name of your s3 bucket and the input path> <<s3://output> path>**



Similarly for all the questions we upload the required jar files and repeat the Hadoop command to acquire our results.

**Git profile link :**

**<https://github.com/noopurrkalawatia/Cloud-Computing-and-Storage.git>**

**Result format :**

**For all the three questions, the output file is going to look**

**<word pattern as specified by the problem statement> <count as received after reducing>**

Results for the Cloud Assignment - 1 : **I have configured 4 slave nodes for my assignment.**

For question - 1 :

<https://s3.us-east-2.amazonaws.com/biblewordcount/output1/_SUCCESS>

<https://s3.us-east-2.amazonaws.com/biblewordcount/output1/part-r-00000>

<https://s3.us-east-2.amazonaws.com/biblewordcount/output1/part-r-00001>

<https://s3.us-east-2.amazonaws.com/biblewordcount/output1/part-r-00002>

<https://s3.us-east-2.amazonaws.com/biblewordcount/output1/part-r-00003>

For question - 2 :

<https://s3.us-east-2.amazonaws.com/biblewordcount/output2/_SUCCESS>

<https://s3.us-east-2.amazonaws.com/biblewordcount/output2/part-r-00000>

<https://s3.us-east-2.amazonaws.com/biblewordcount/output2/part-r-00001>

<https://s3.us-east-2.amazonaws.com/biblewordcount/output2/part-r-00002>

<https://s3.us-east-2.amazonaws.com/biblewordcount/output2/part-r-00003>

For question - 3:

<https://s3.us-east-2.amazonaws.com/biblewordcount/output3/_SUCCESS>

<https://s3.us-east-2.amazonaws.com/biblewordcount/output3/part-r-00000>

<https://s3.us-east-2.amazonaws.com/biblewordcount/output3/part-r-00001>

<https://s3.us-east-2.amazonaws.com/biblewordcount/output3/part-r-00002>

<https://s3.us-east-2.amazonaws.com/biblewordcount/output3/part-r-00003>

x-x-x