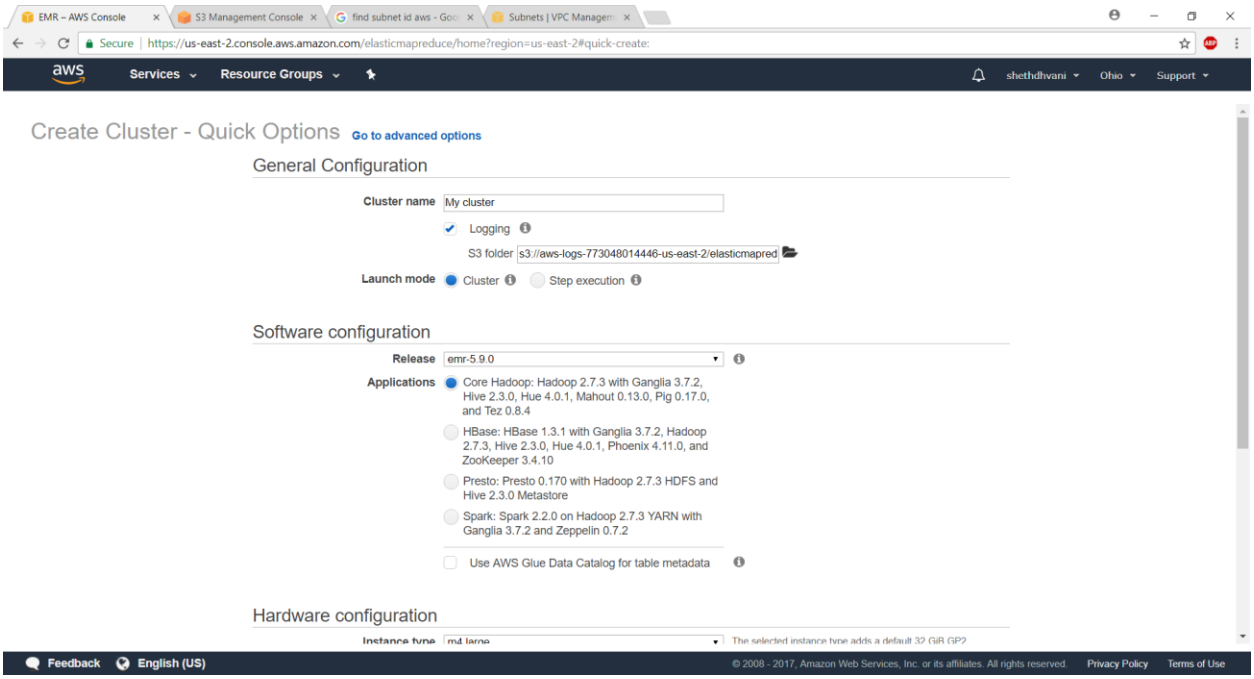
Running program on EMR

4 nodes:

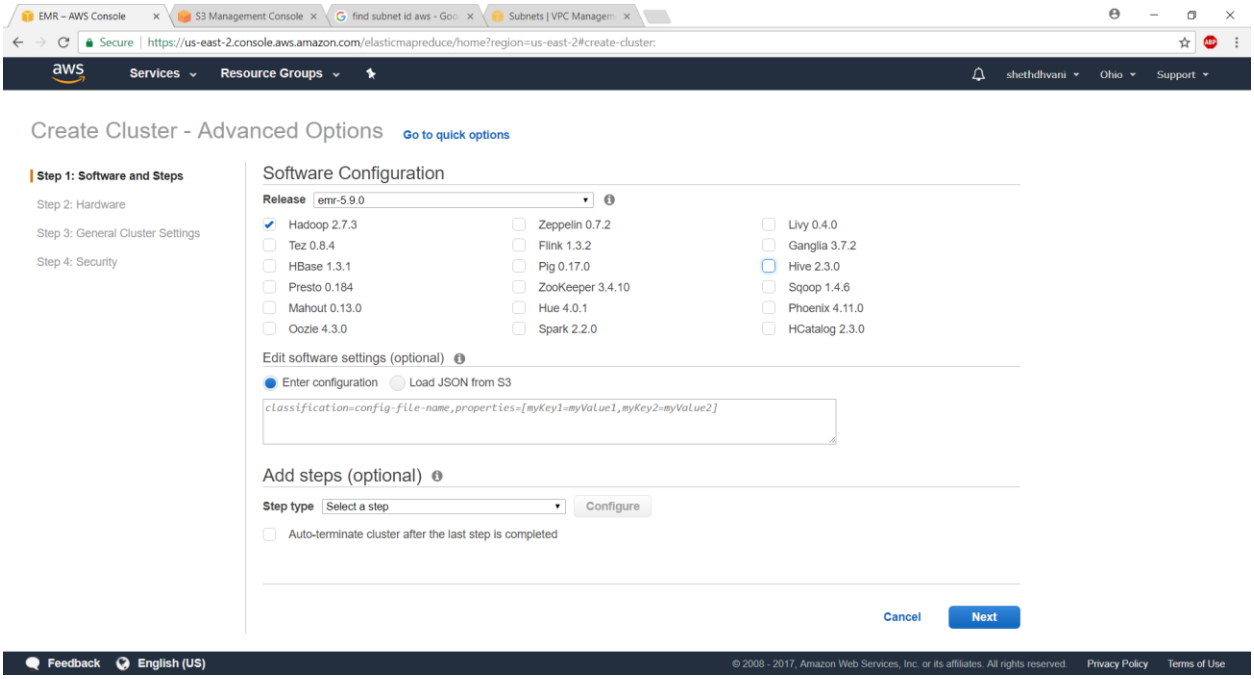
Create Cluster:

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



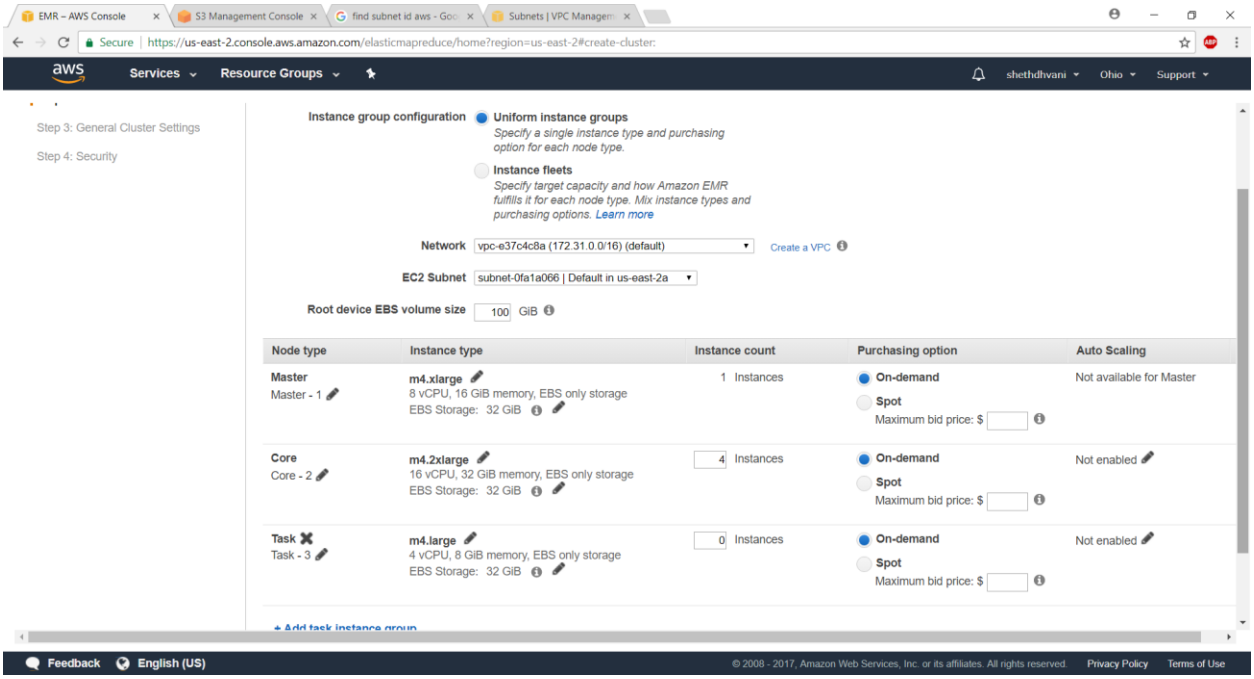
4. Go to advanced options

Running program on EMR



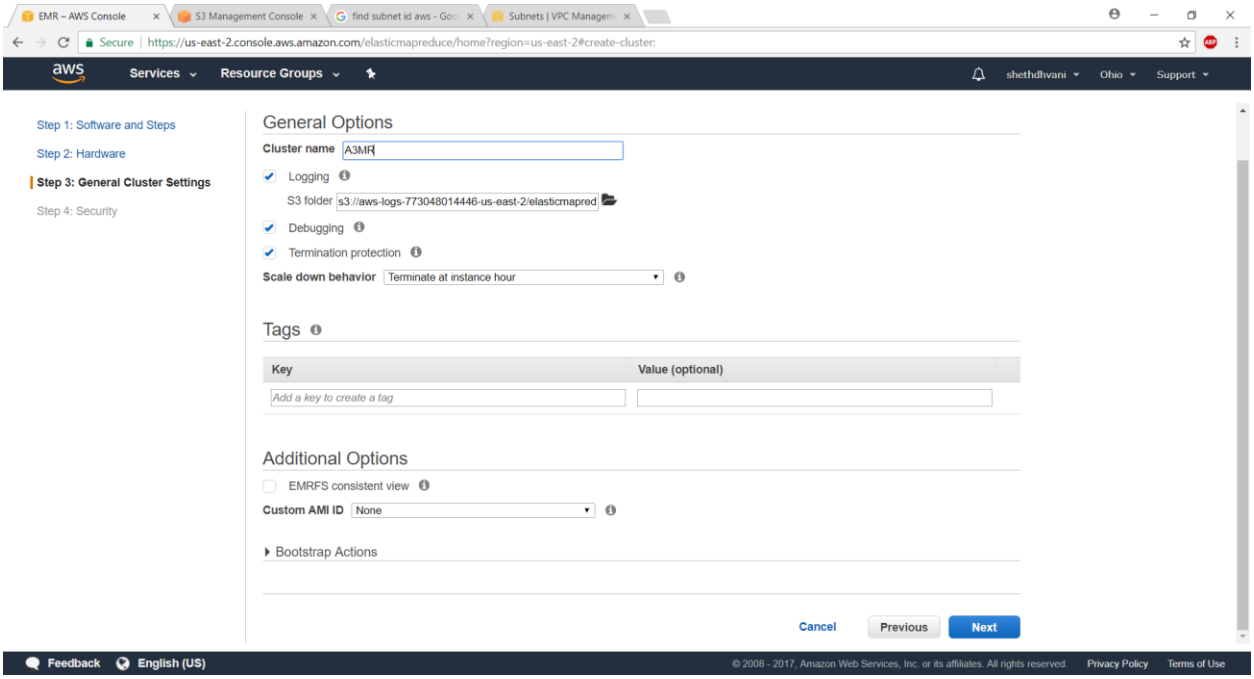
5. Select as shown above and click on Next.

6. Select and/or edit the below values:

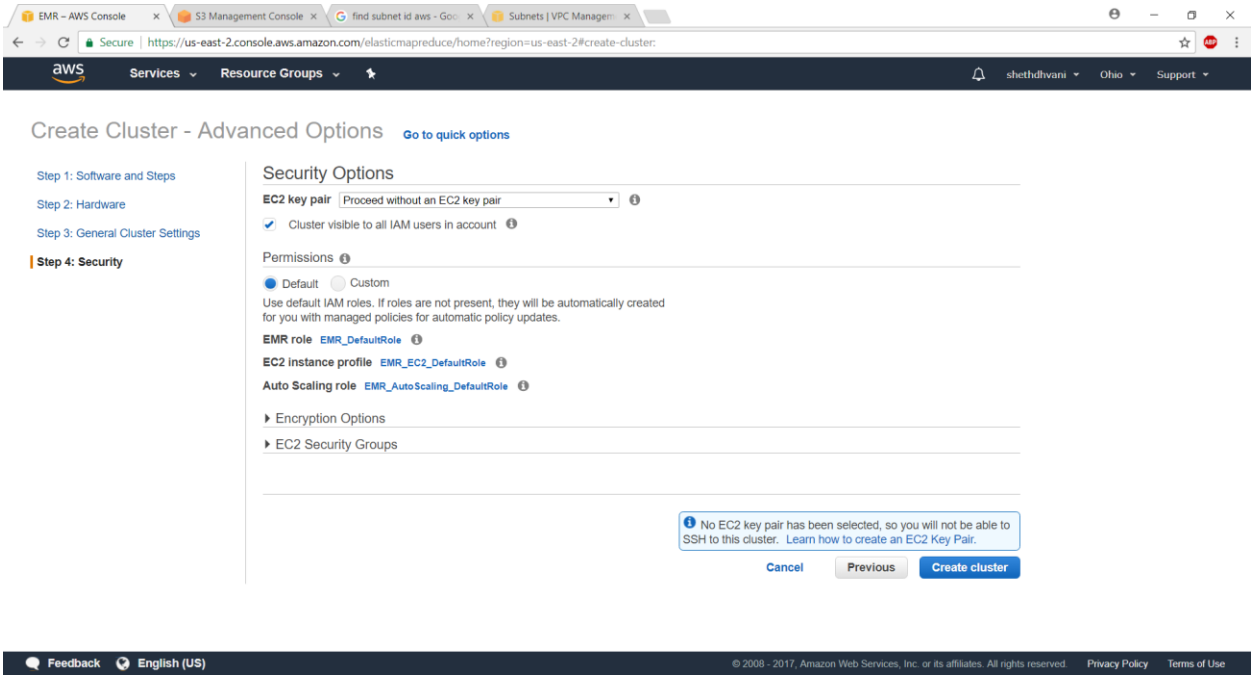


7. Click on next.

Running program on EMR



8. Give a cluster name and click on next.



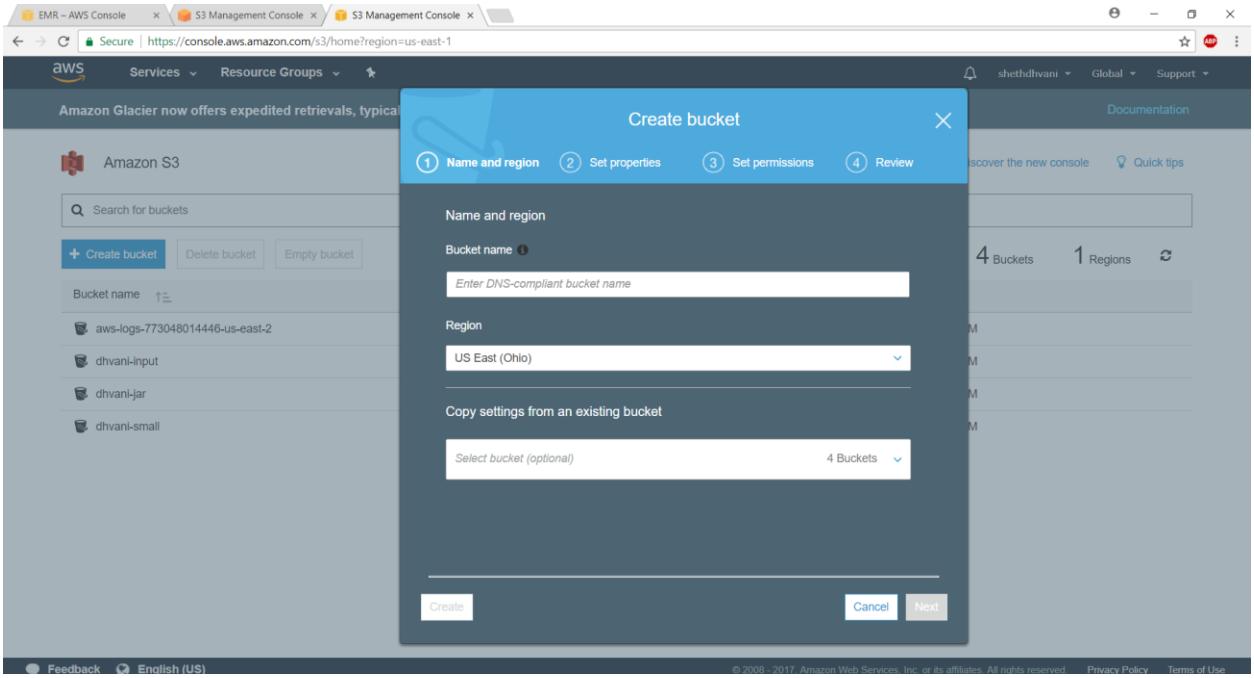
If there is an EC2 key pair already setup, select that. If not proceed without an EC2 key pair.

9. Click on create cluster.

Running program on EMR

Add files to S3 bucket:

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

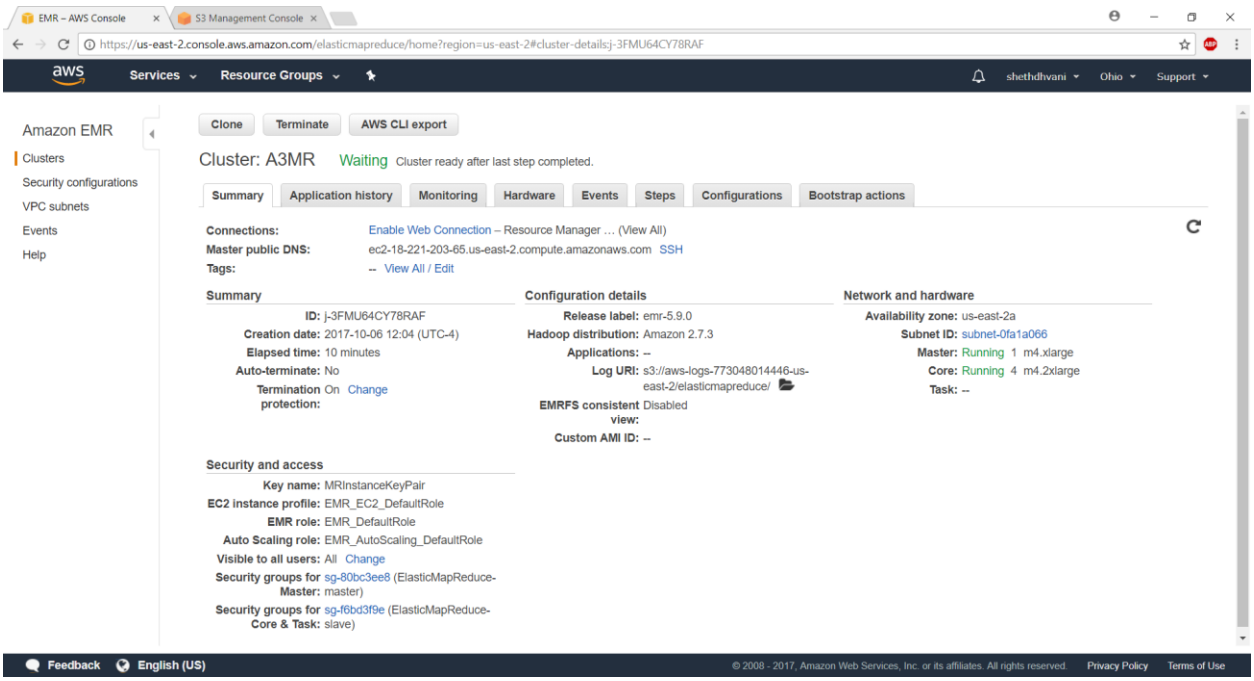


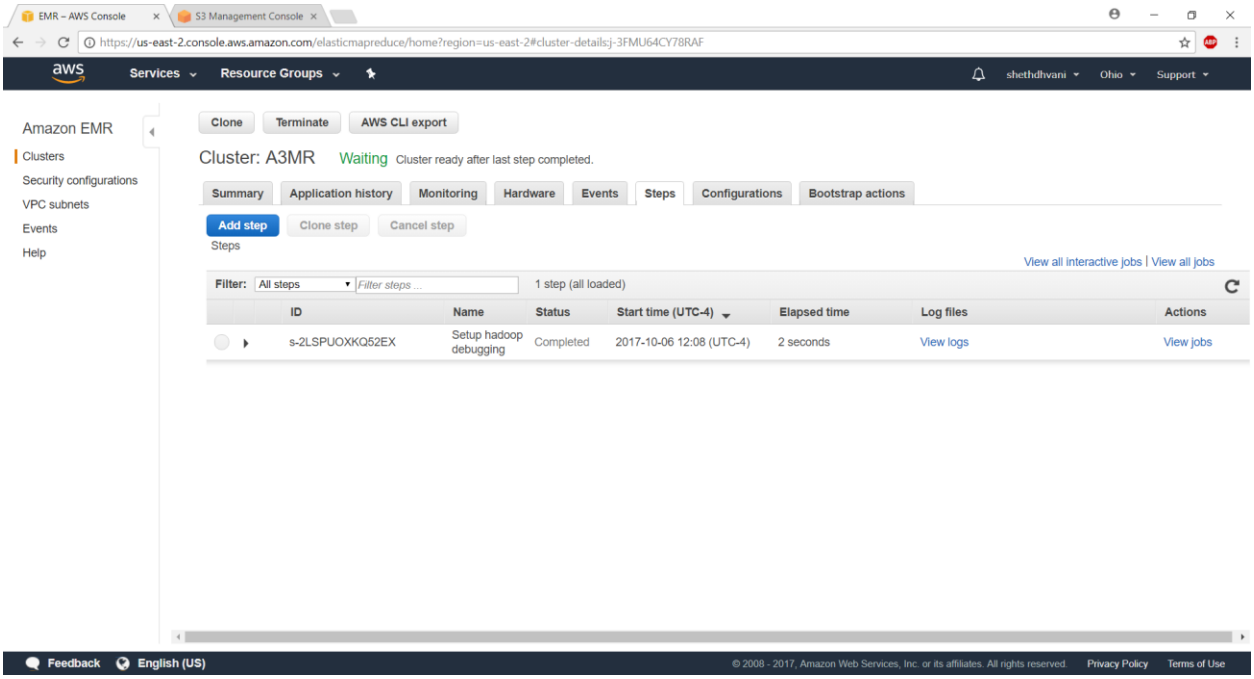
2. Give a unique bucket name. The region should match the region in which cluster was created in the previous step.
3. Click on next till create bucket.
4. We need to create another bucket also. So one bucket is for input files and one is for jar.
5. Add the input files and jar in the respective buckets.

Running the job:

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

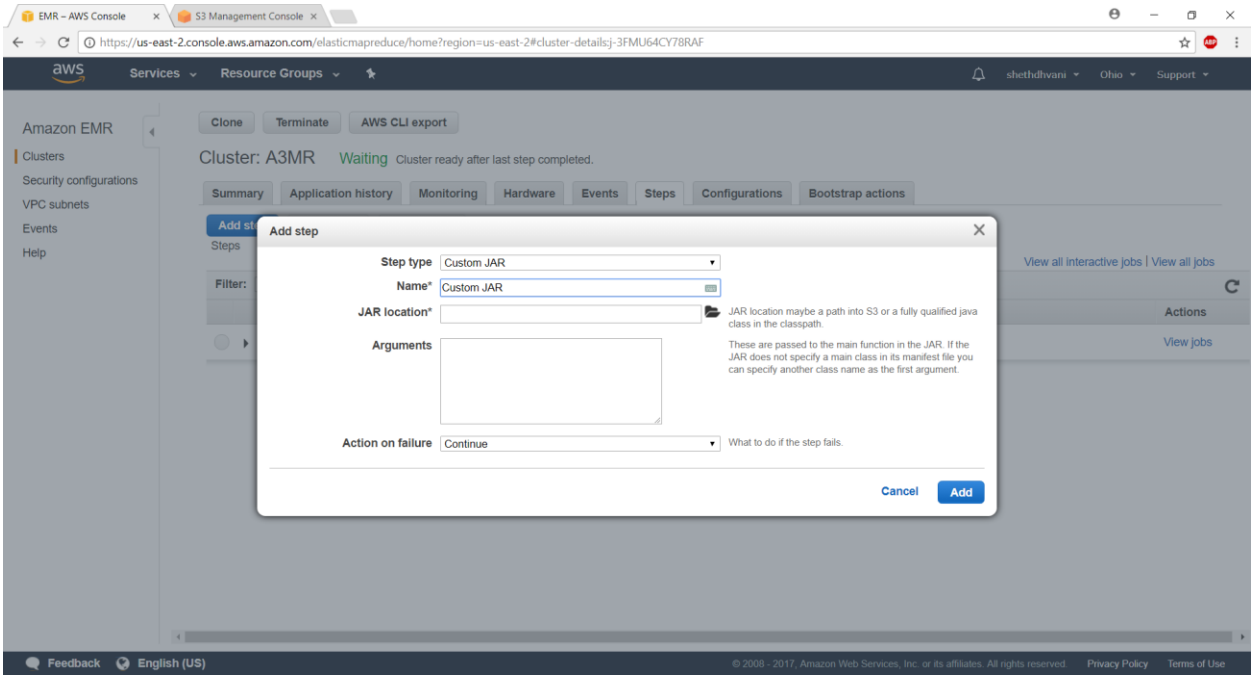
Running program on EMR



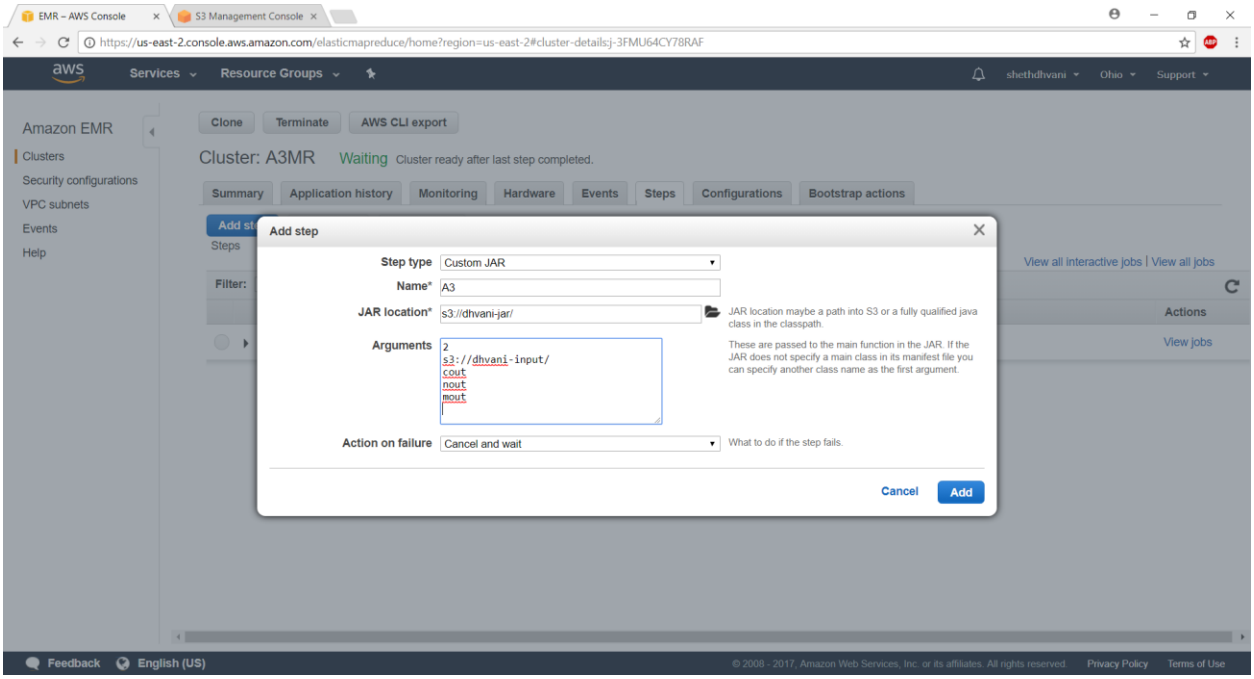


2. Click on Add step.

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.