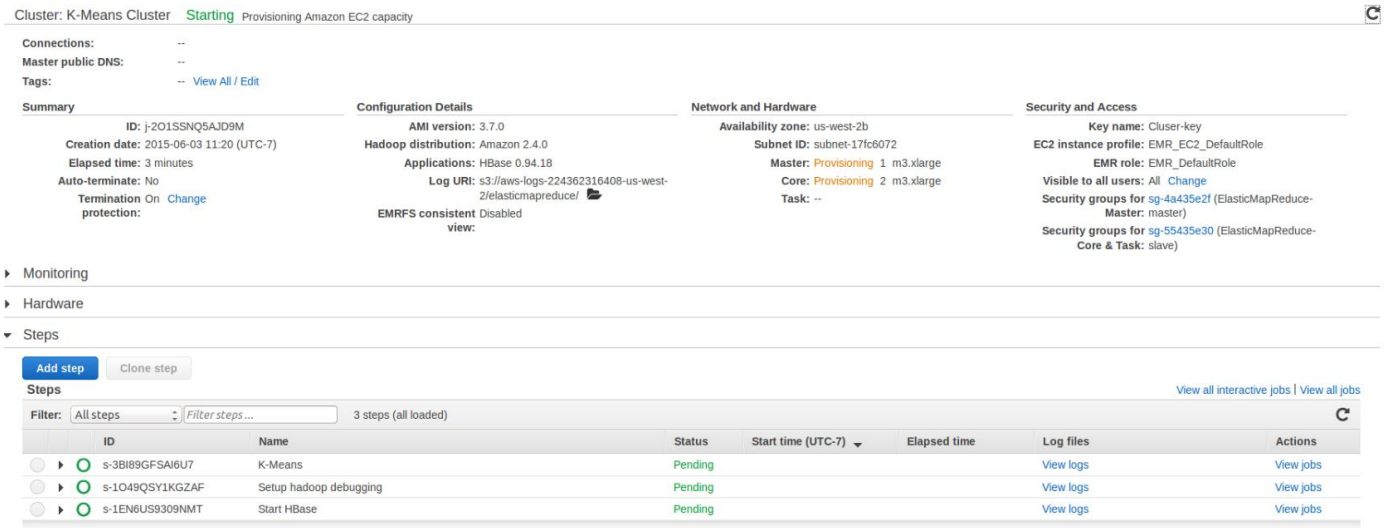
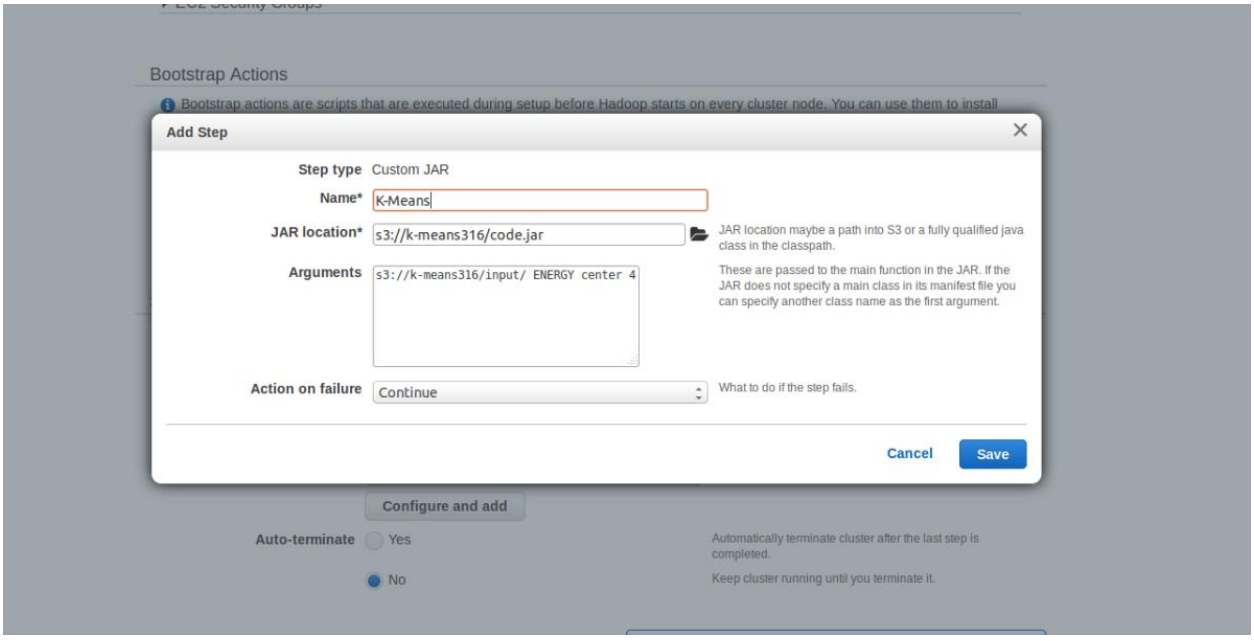
EMR Report for K-Means cluster (Project 3 - EECS 219)

Abhishek Madav Student # 86378148

The applications has been tested on EMR cluster with the settings of a regular cluster and the default HBase version available over the AWS. The cluster had been created, with custom jar settings as follows:-

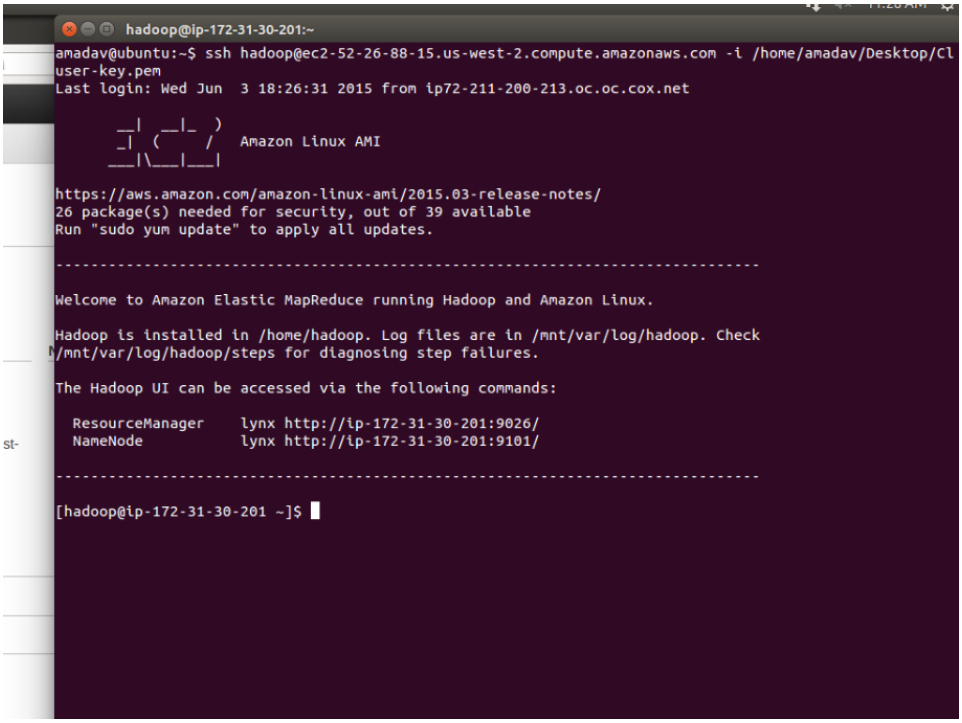




The application input four parameters as required-

1) S3 location of the dataset.txt file.
2) Table name for the input dataset.
3) Table name for the center table.
4) K- the number of clusters.

After the application has been successfully completed running, an SSH connection to the AWS master was established.



The master provides the setting for hbase shell prompt and the tables could be scanned for the values. Scan 'center' results into –

```
ROW                                              COLUMN+CELL
 center0                                         column=Area:X1, timestamp=1433368639022, value=0.8900027054976737
 center0                                         column=Area:X5, timestamp=1433368639022, value=569.623389874067
 center0                                         column=Area:X6, timestamp=1433368639022, value=306.24973874932493
 center0                                         column=Area:Y1, timestamp=1433368639022, value=131.68682556237104
 center0                                         column=Area:Y2, timestamp=1433368639022, value=6.9999992115170775
 center0                                         column=Property:X2, timestamp=1433368639022, value=3.499925657324421
 center0                                         column=Property:X3, timestamp=1433368639022, value=0.23436730994482746
 center0                                         column=Property:X4, timestamp=1433368639022, value=2.8124619369255712
 center0                                         column=Property:X7, timestamp=1433368639022, value=28.34576414701536
 center0                                         column=Property:X8, timestamp=1433368639022, value=30.49437543059147
 center1                                         column=Area:X1, timestamp=1433368491087, value=0.6402566409891939
 center1                                         column=Area:X5, timestamp=1433368491087, value=783.7587053127913
 center1                                         column=Area:X6, timestamp=1433368491087, value=342.8292448071897
 center1                                         column=Area:Y1, timestamp=1433368491087, value=220.46473025280085
 center1                                         column=Area:Y2, timestamp=1433368491087, value=3.50140423678298
 center1                                         column=Property:X2, timestamp=1433368491087, value=3.4956210738685867
 center1                                         column=Property:X3, timestamp=1433368491087, value=0.23431918875234997
 center1                                         column=Property:X4, timestamp=1433368491087, value=2.812297961850612
 center1                                         column=Property:X7, timestamp=1433368491087, value=14.569918168579095
 center1                                         column=Property:X8, timestamp=1433368491087, value=17.111189704986096
 center2                                         column=Area:X1, timestamp=1433368490962, value=0.7132145237290664
 center2                                         column=Area:X5, timestamp=1433368490962, value=710.6244708361493
 center2                                         column=Area:X6, timestamp=1433368490962, value=269.6655778328452
 center2                                         column=Area:Y1, timestamp=1433368490962, value=220.47944650165203
 center2                                         column=Area:Y2, timestamp=1433368490962, value=3.50083774498761
 center2                                         column=Property:X2, timestamp=1433368490962, value=3.5053221446271685
 center2                                         column=Property:X3, timestamp=1433368490962, value=0.2343316167774272
 center2                                         column=Property:X4, timestamp=1433368490962, value=2.8119737694300517
 center2                                         column=Property:X7, timestamp=1433368490962, value=12.112868604861262
 center2                                         column=Property:X8, timestamp=1433368490962, value=15.037270790507428
 center3                                         column=Area:X1, timestamp=1433368620112, value=0.7750120376770102
 center3                                         column=Area:X5, timestamp=1433368620112, value=649.2406887995188
 center3                                         column=Area:X6, timestamp=1433368620112, value=379.7365803608536
 center3                                         column=Area:Y1, timestamp=1433368620112, value=134.75205421933256
 center3                                         column=Area:Y2, timestamp=1433368620112, value=6.999999604723442
 center3                                         column=Property:X2, timestamp=1433368620112, value=3.5002371659348404
 center3                                         column=Property:X3, timestamp=1433368620112, value=0.23440519559978132
 center3                                         column=Property:X4, timestamp=1433368620112, value=2.8126549060597026
 center3                                         column=Property:X7, timestamp=1433368620112, value=37.13442943644058
 center3                                         column=Property:X8, timestamp=1433368620112, value=38.3221744149107
4 row(s) in 0.5580 seconds
```

The application has thus been tested over the EMR cluster for correct centroid readings.