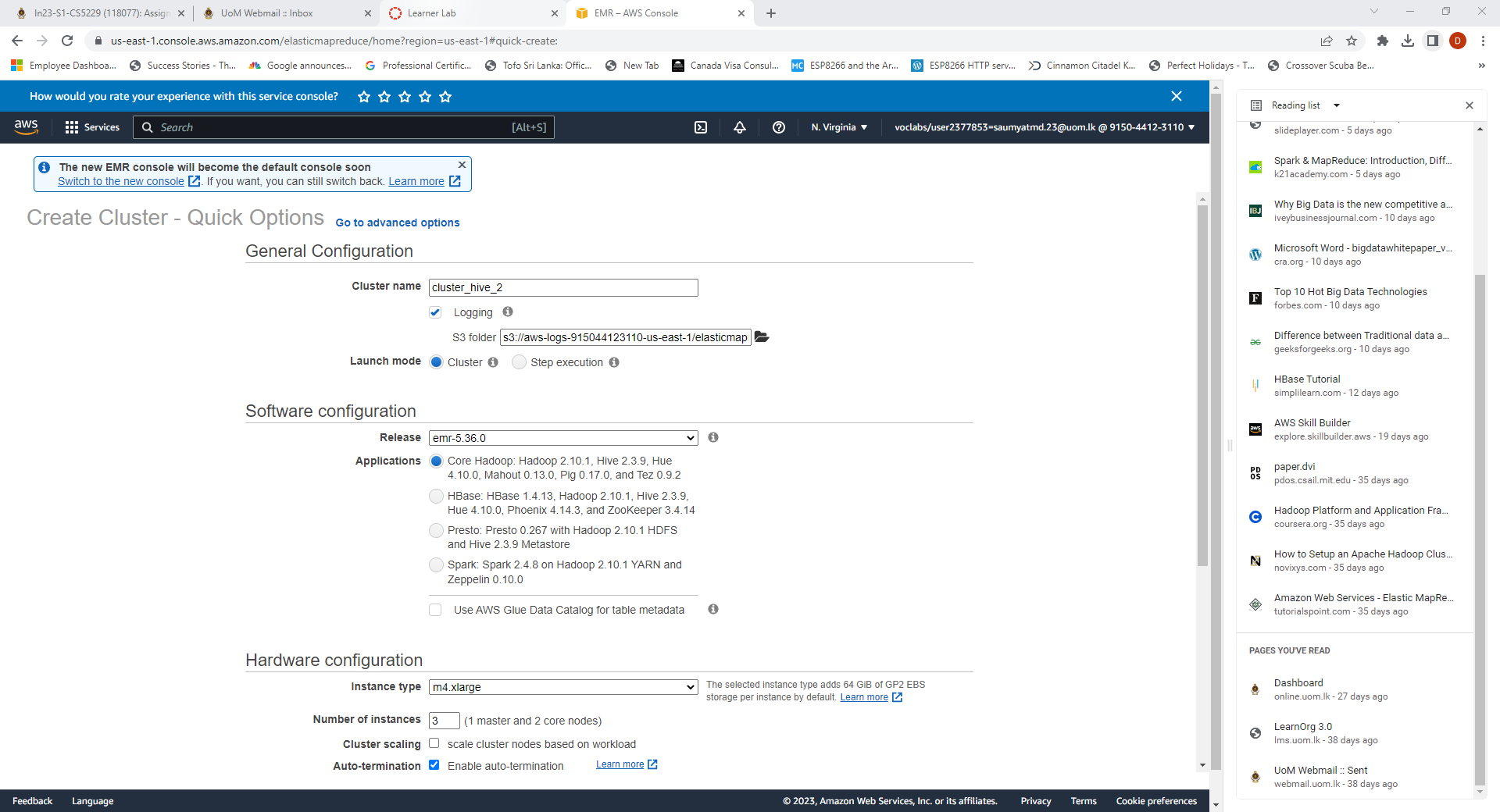
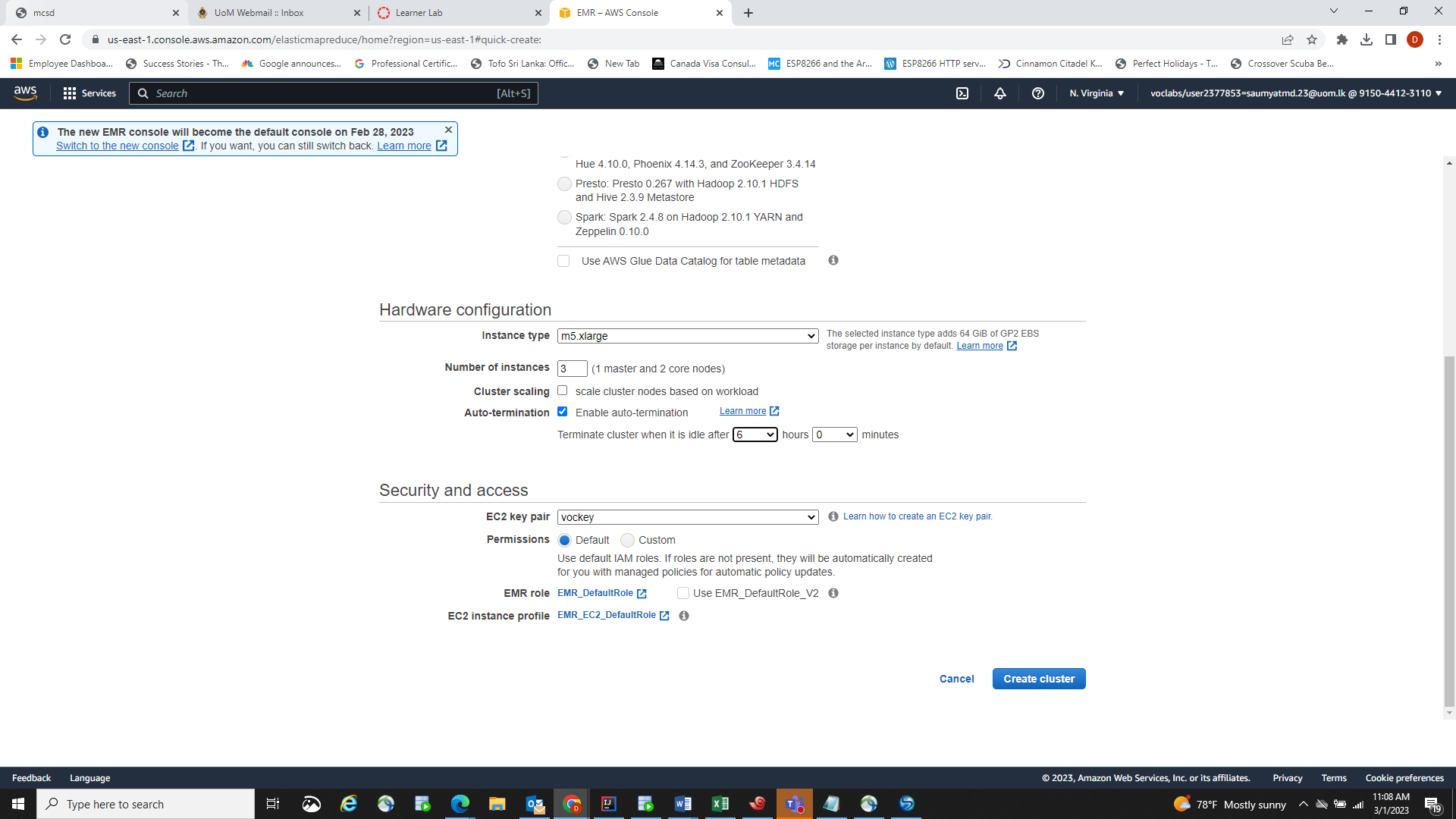
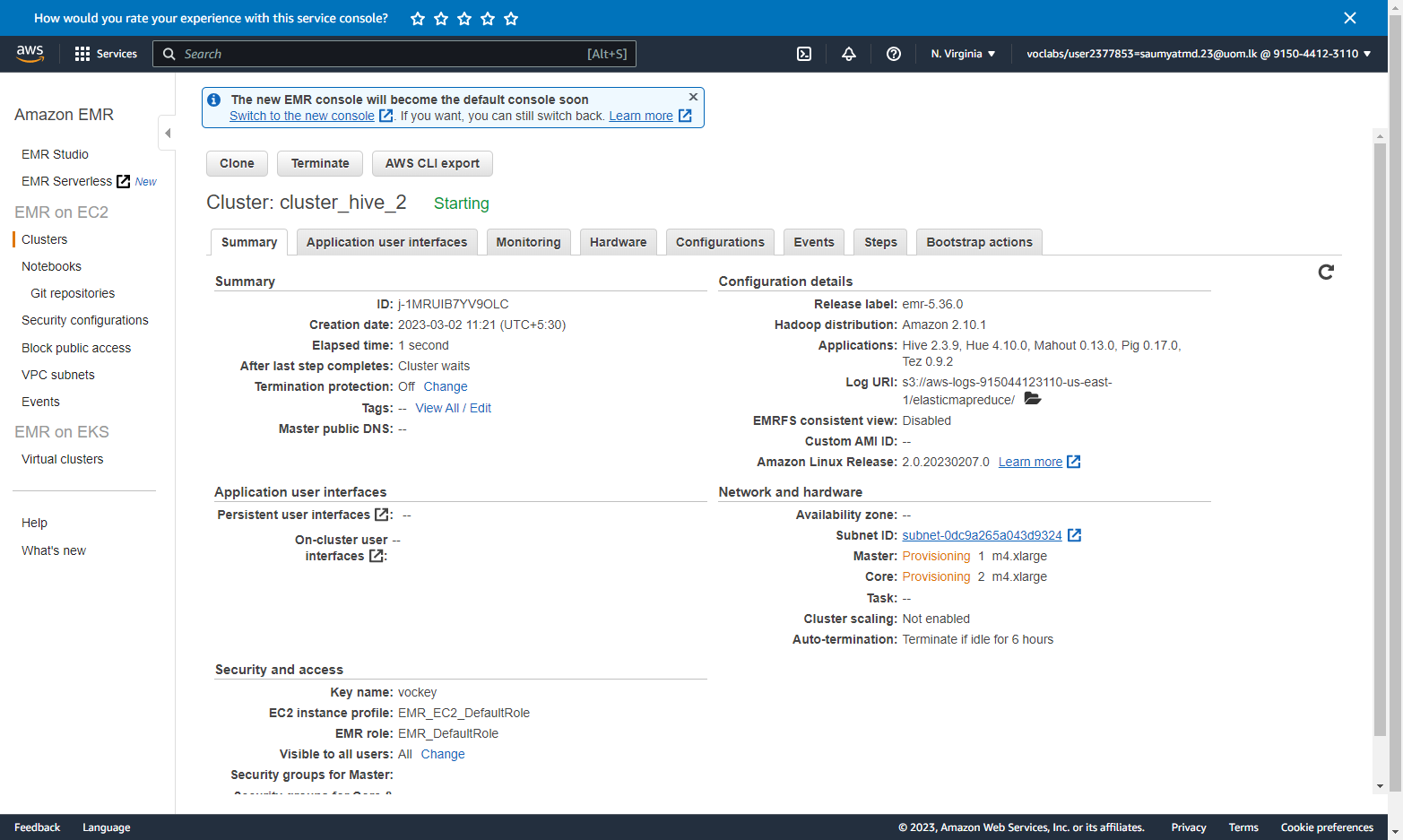
# MAP Reduce

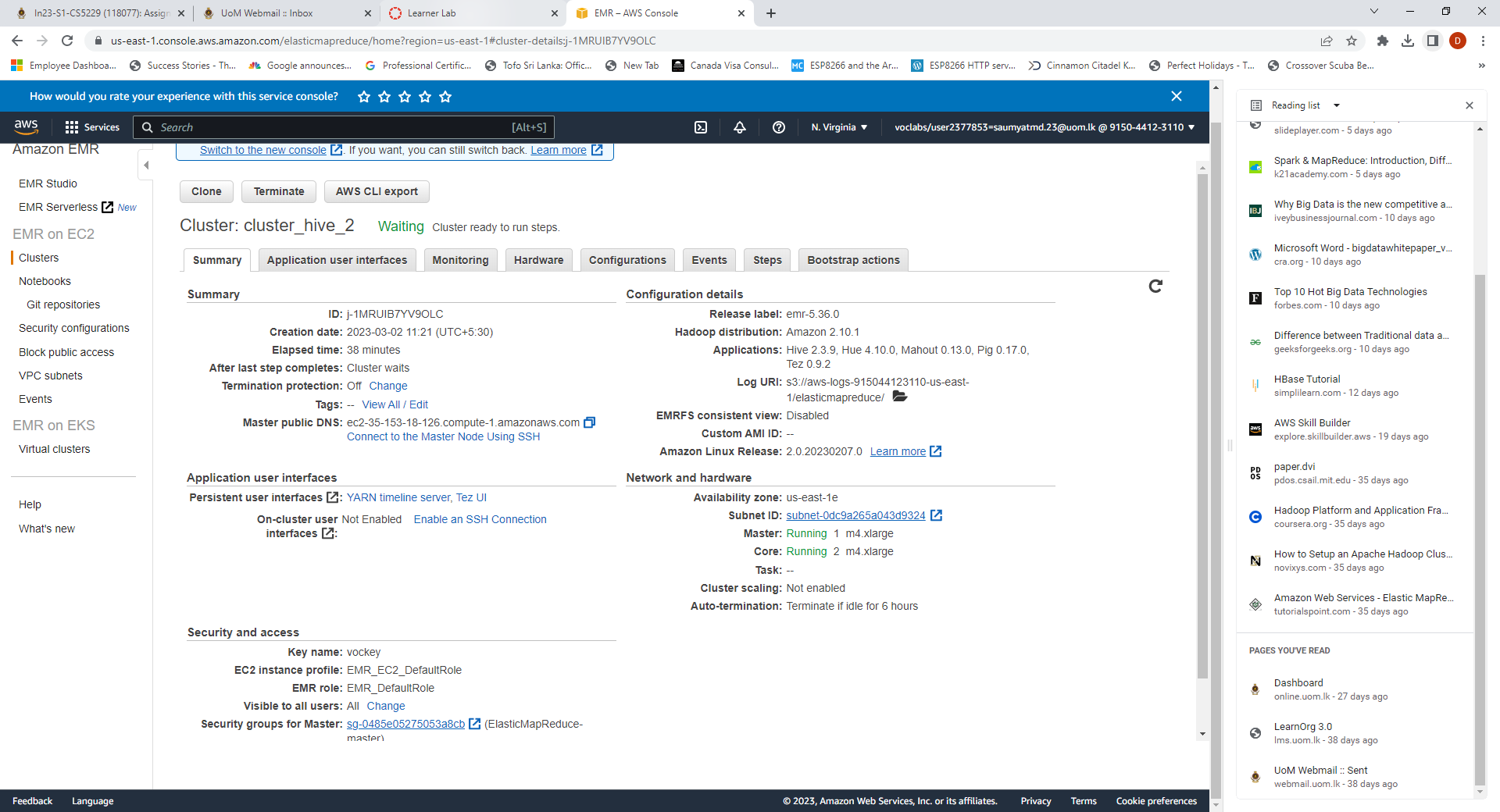
## EMR Hadoop cluster creation and to run hive queries



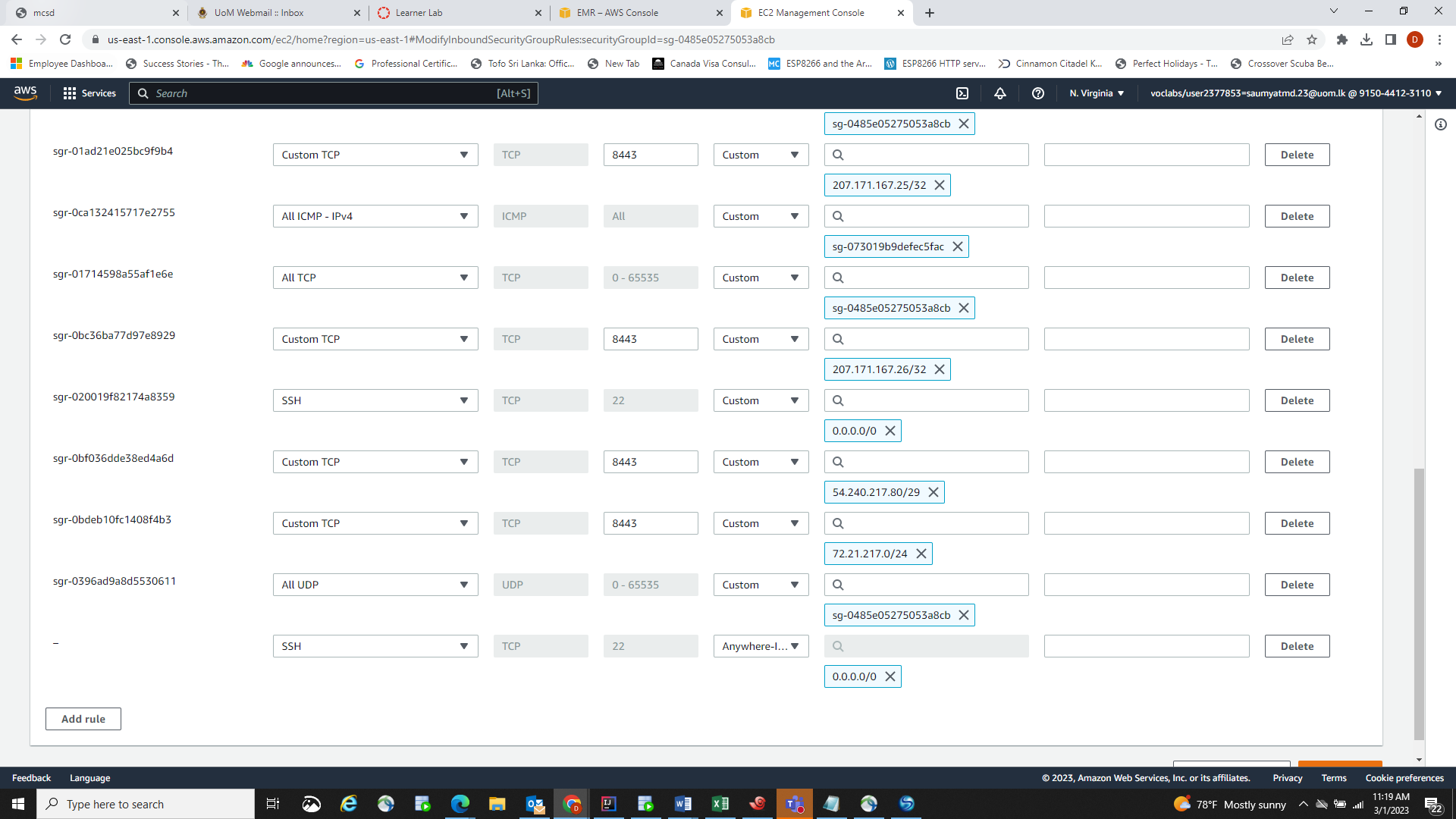




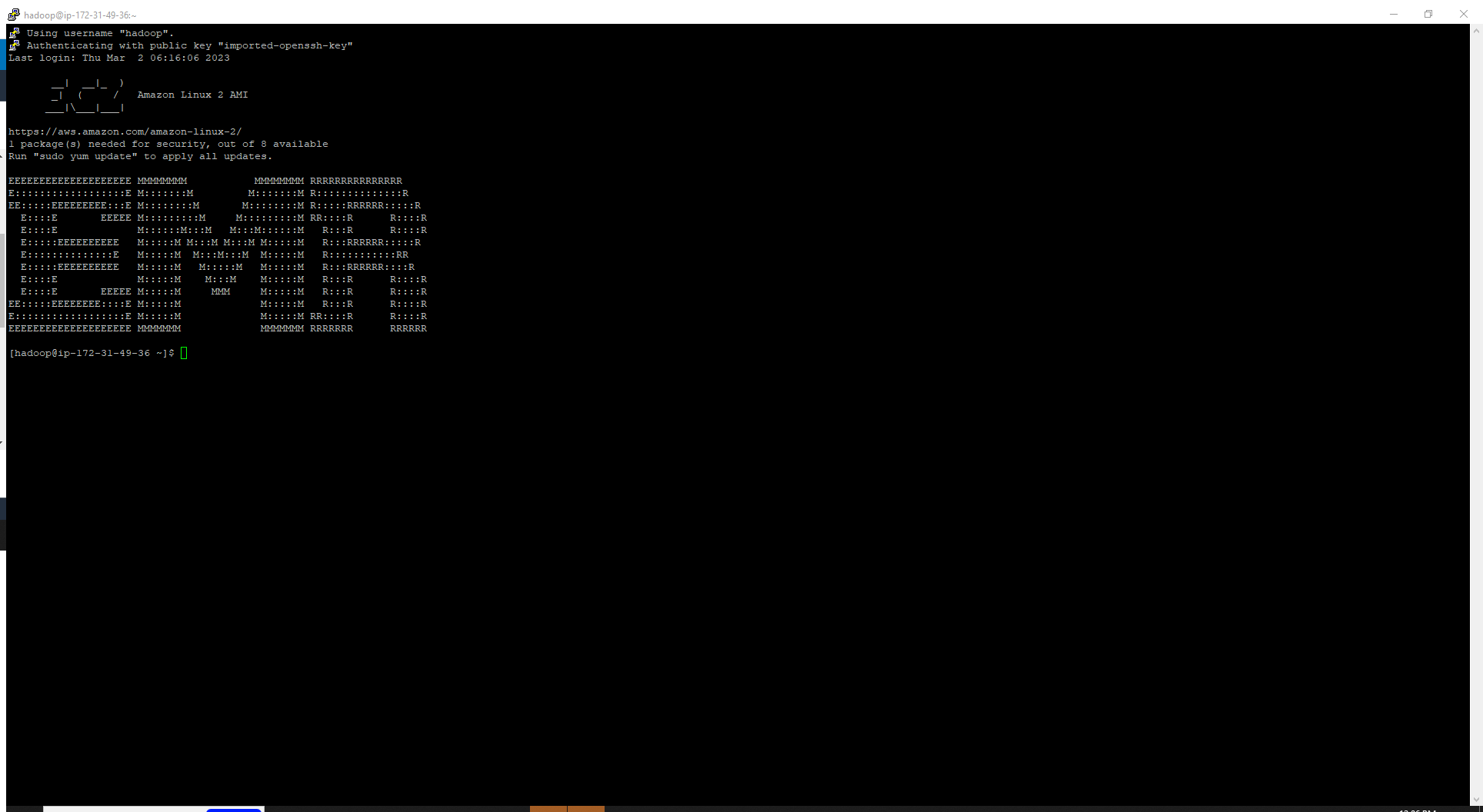
### Cluster started



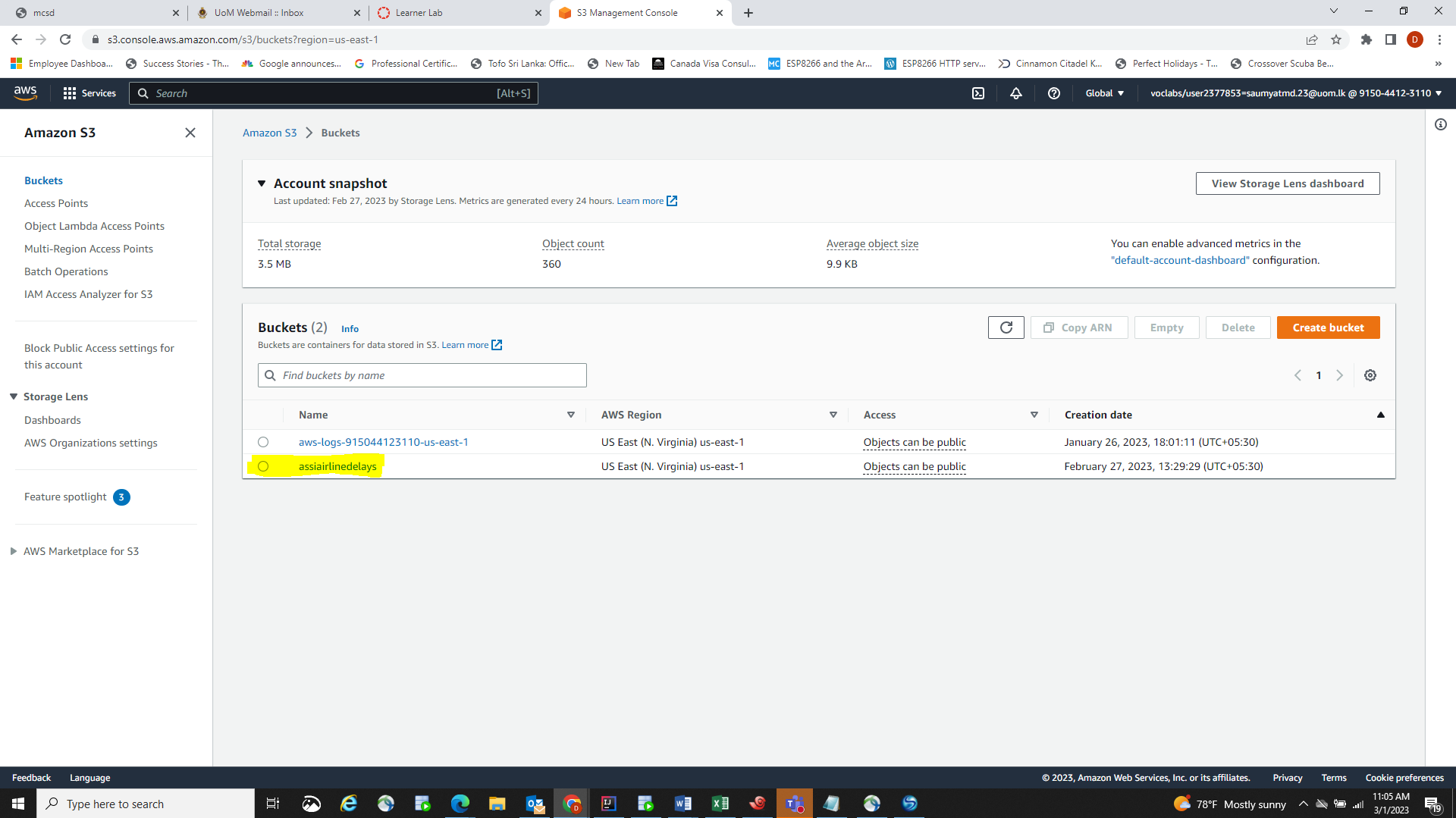
### Adding inbound rule to master rule



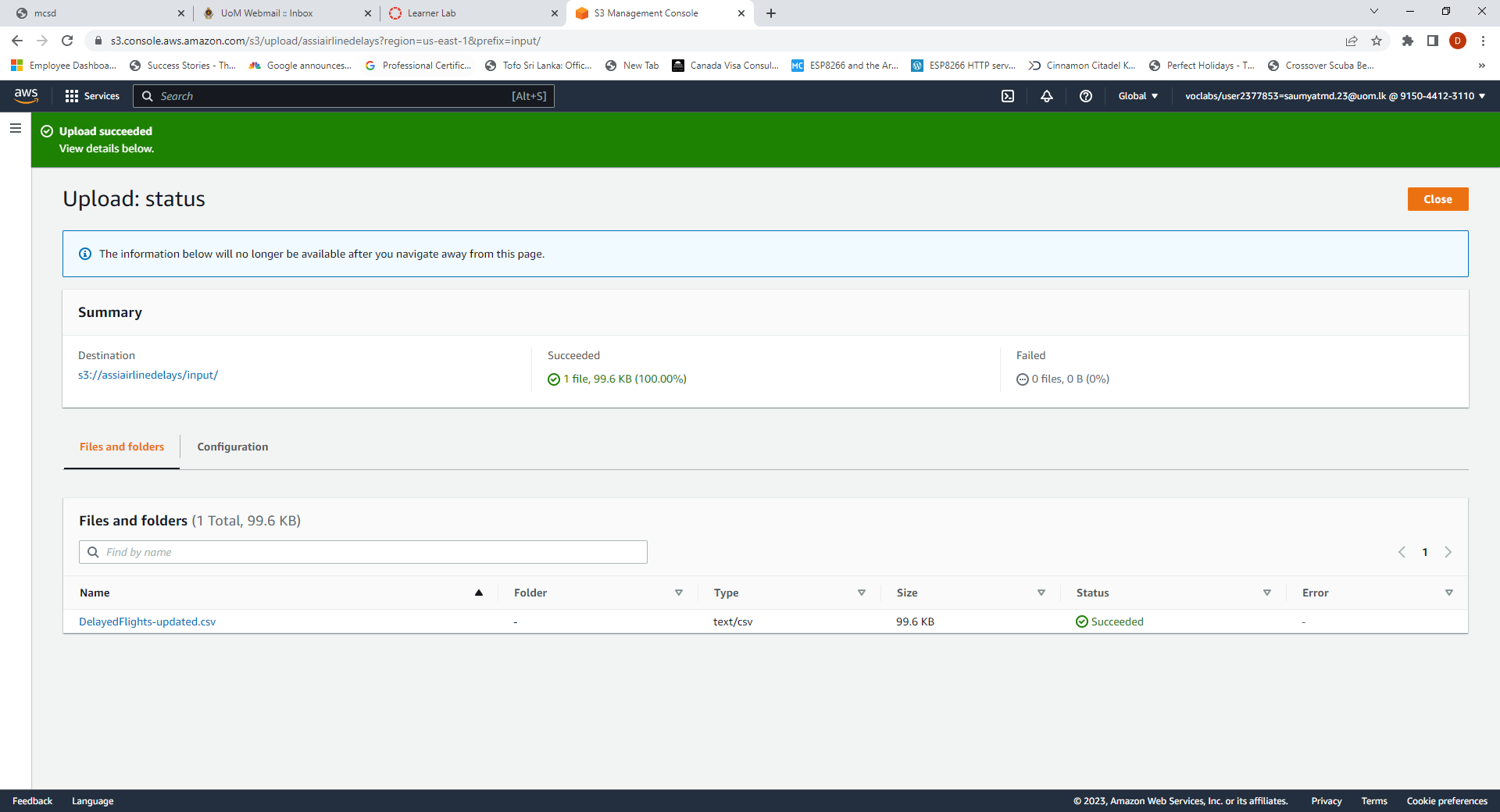
### Connecting to master node of the cluster using SSH connection using vockey key



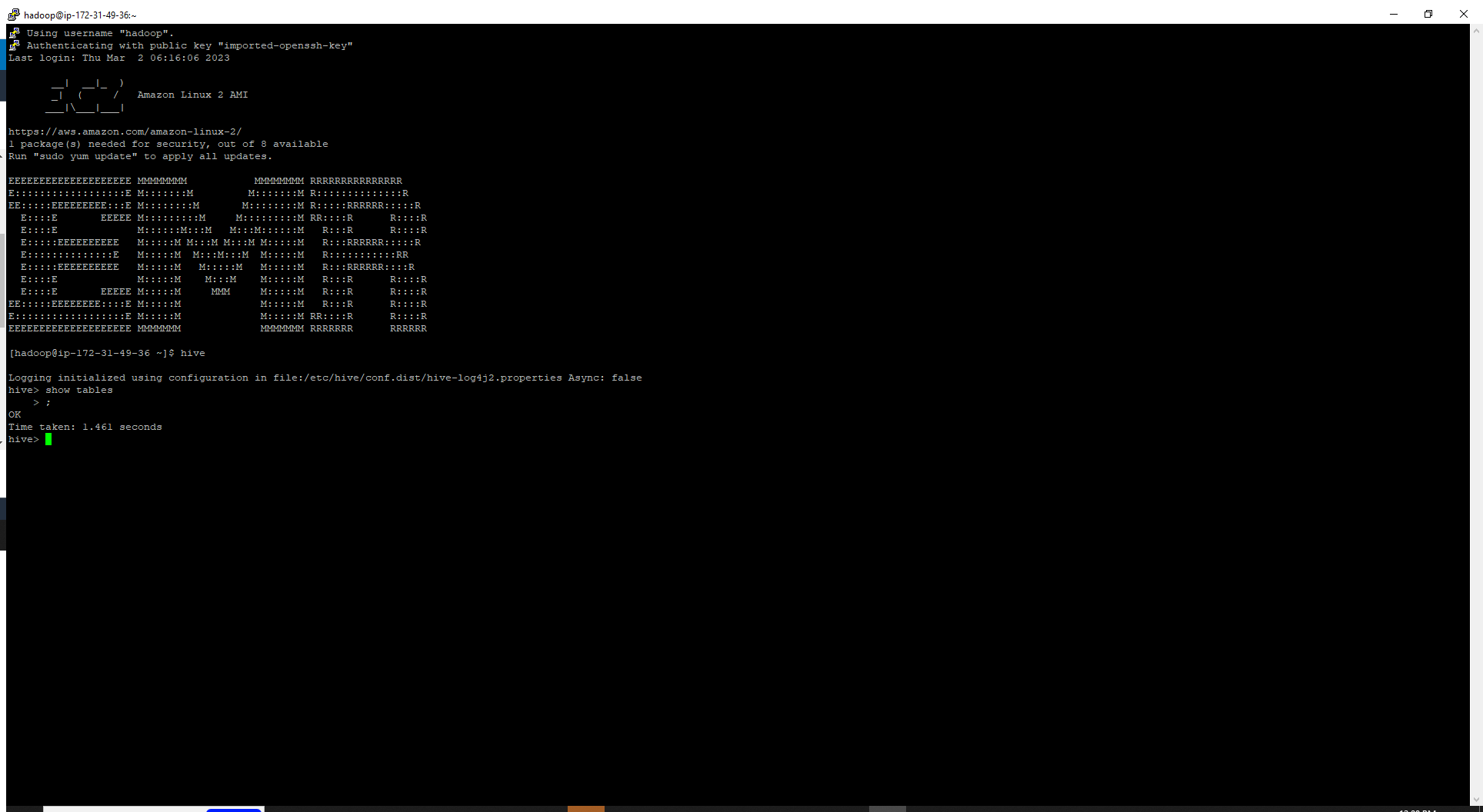
## S3 bucket creation



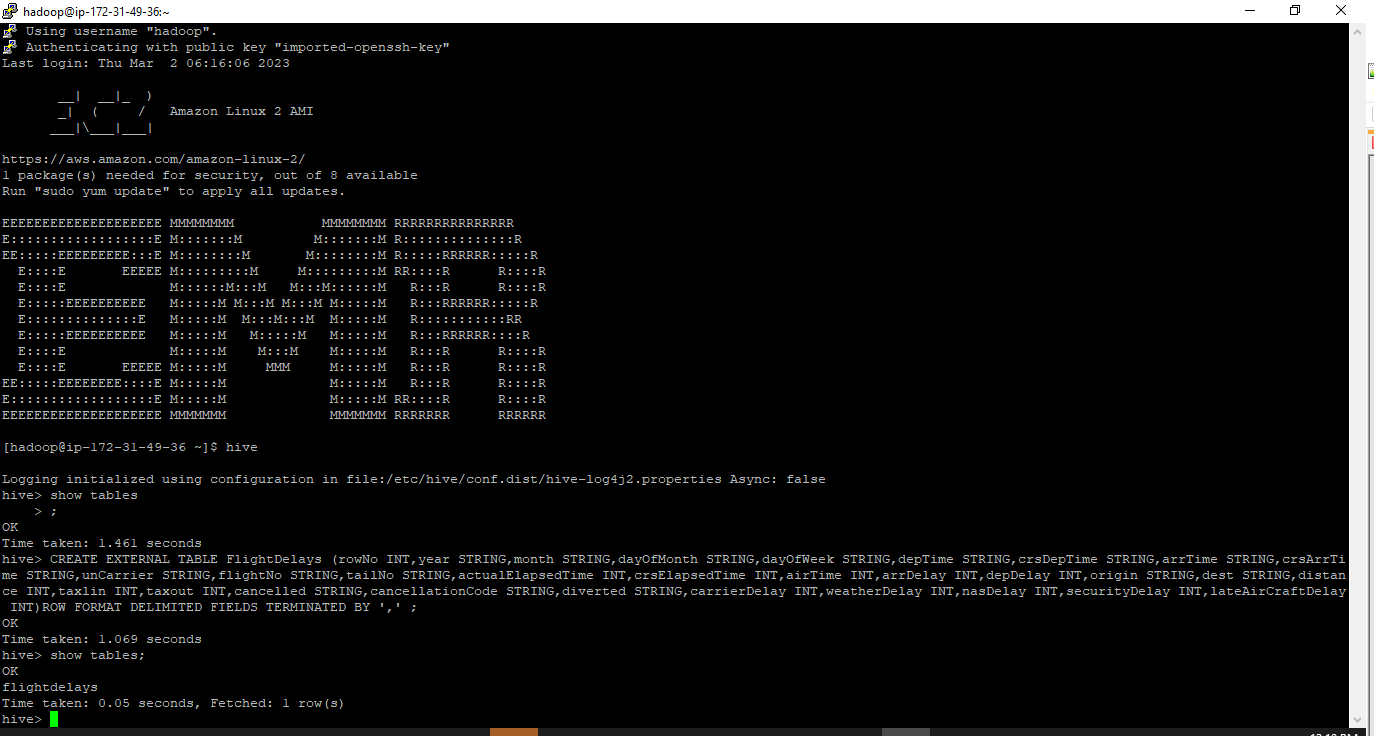
## Uploading data set to S3 bucket – input folder

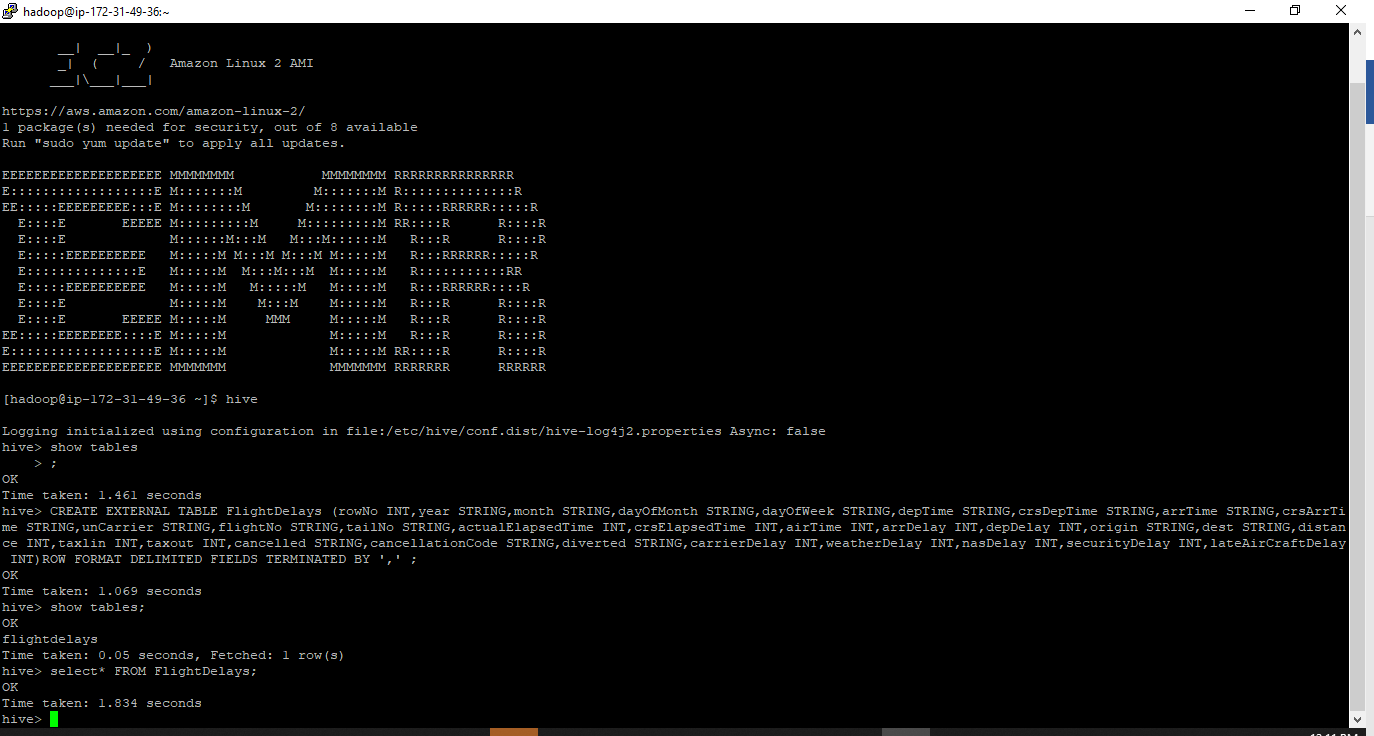


### Running hive command to see no of tables in the hive

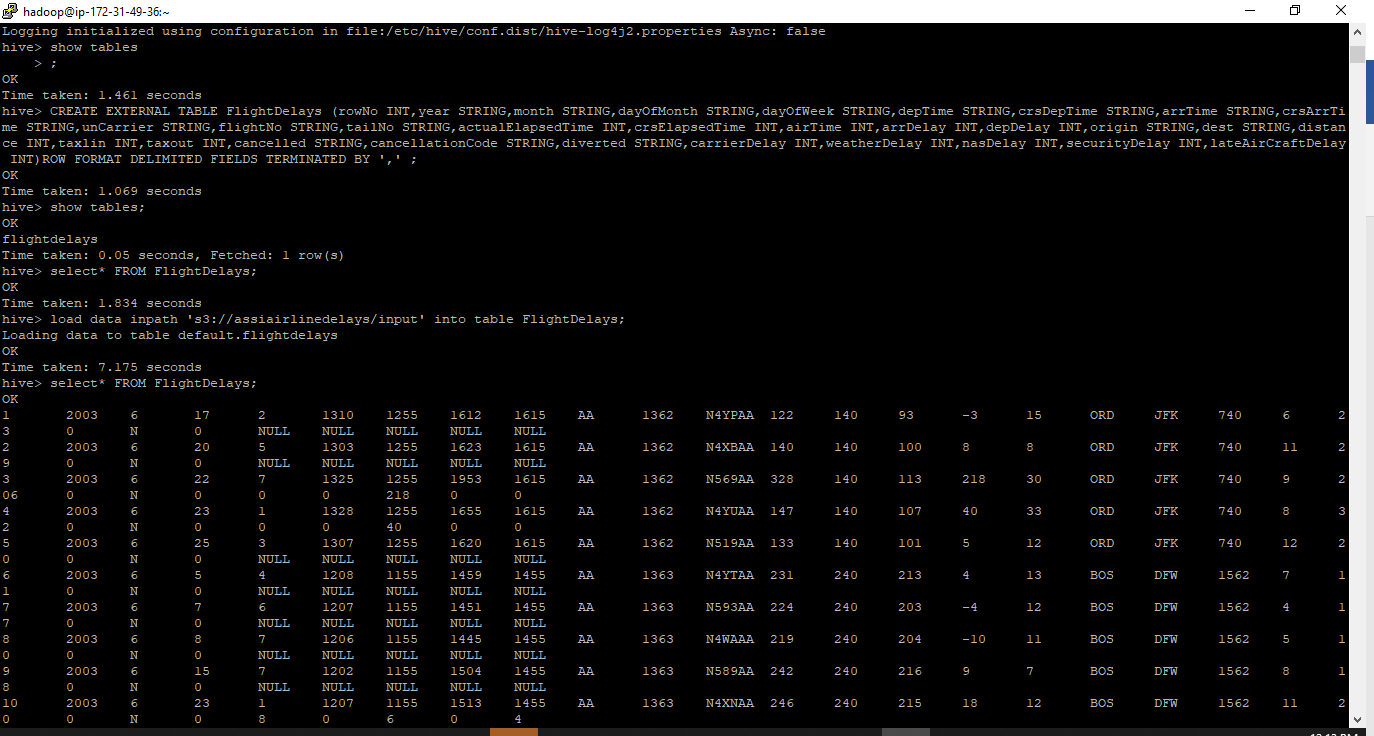


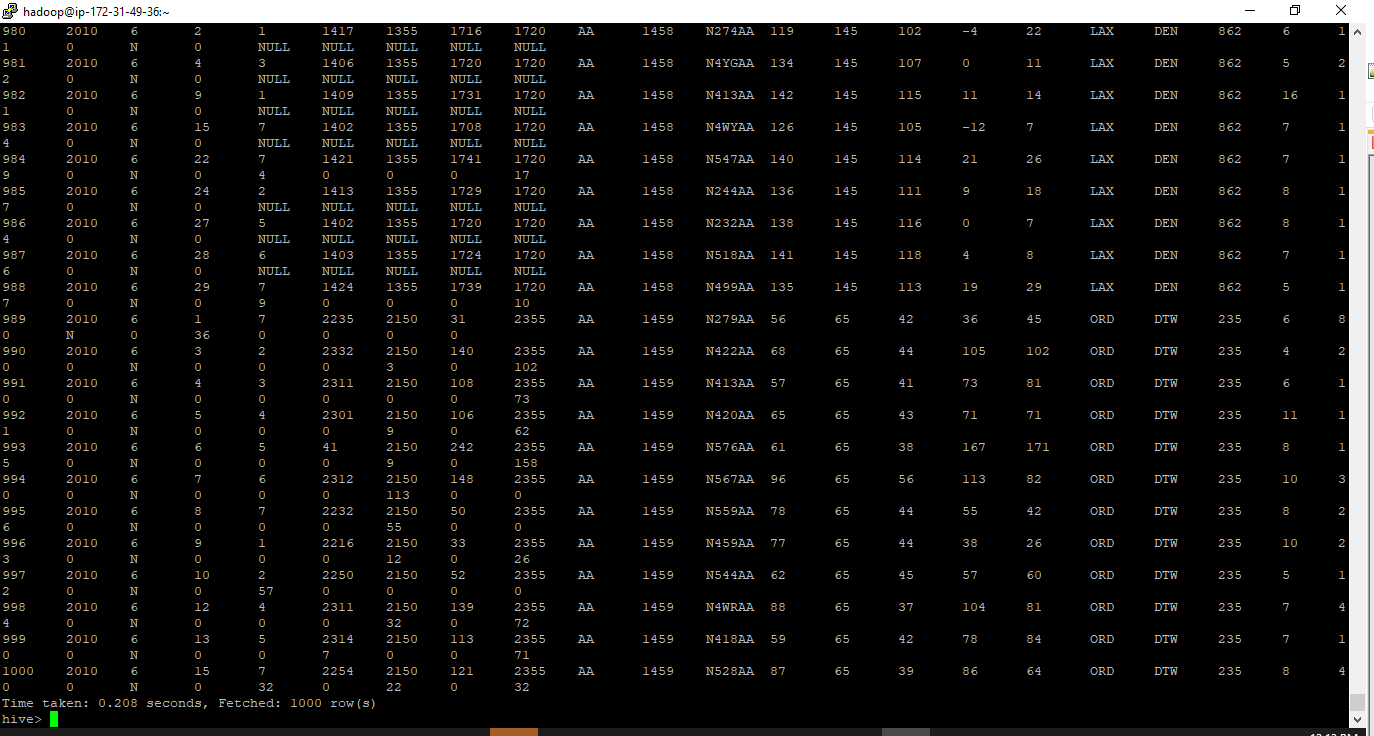
### Created local table using hive CREATE





### Load the data and ran a select all query run on table

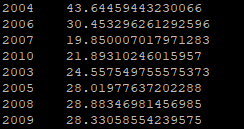




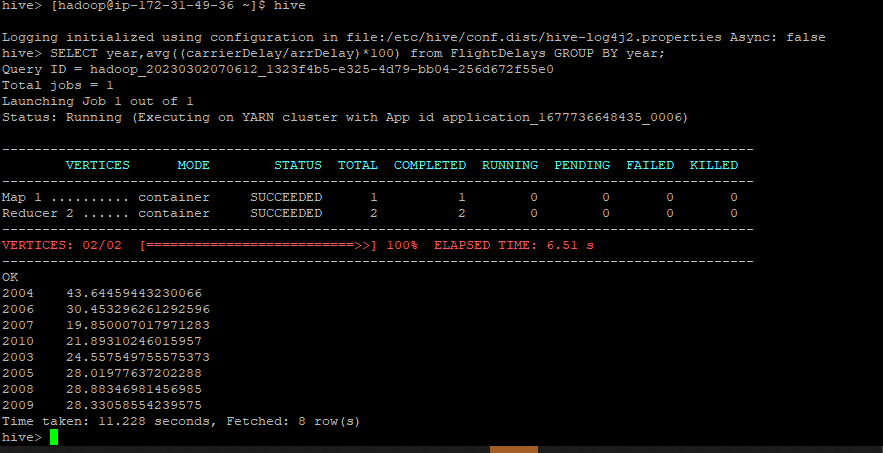
## Analysis of flight delays

### Year wise carrier delay from 2003-2010

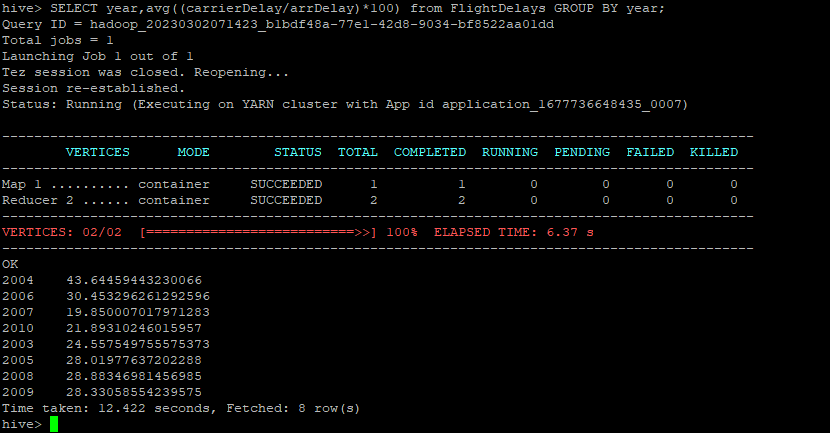
#### Output



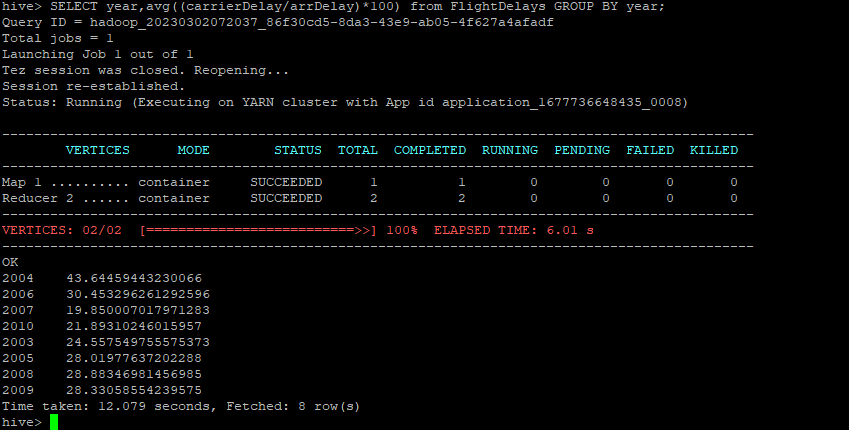
#### Iteration 1 -MR Jobs Summary



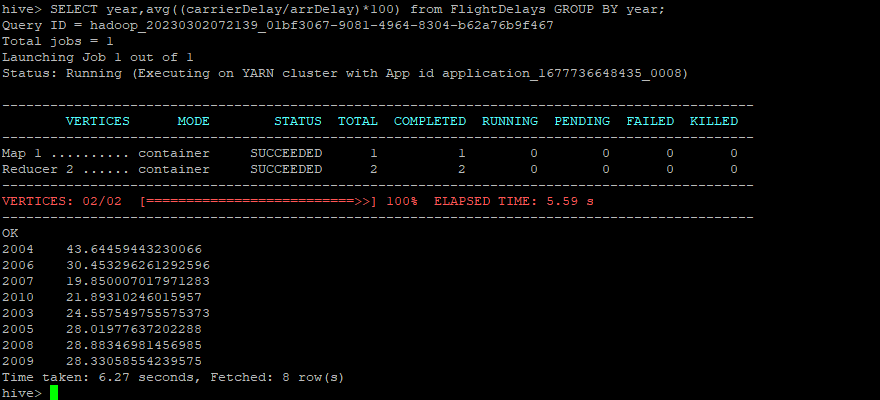
#### Iteration 2 -MR Jobs Summary



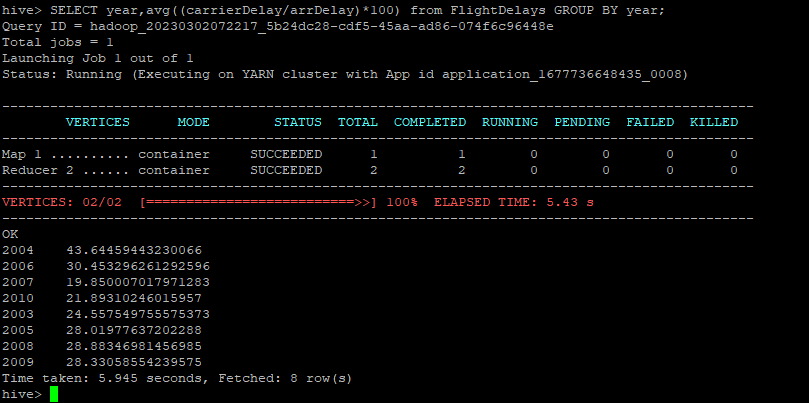
#### Iteration 3 -MR Jobs Summary



#### Iteration 4 - MR Jobs Summary

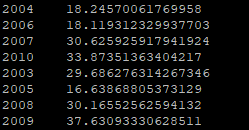


#### Iteration 5 - MR Jobs Summary

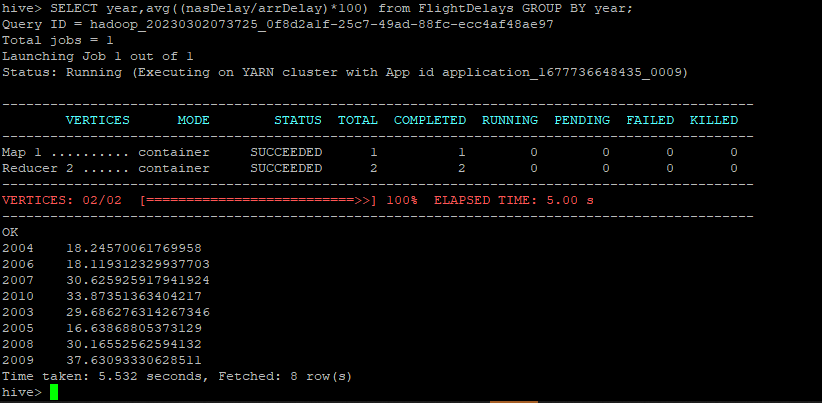


### Year wise NAS delay from 2003-2010

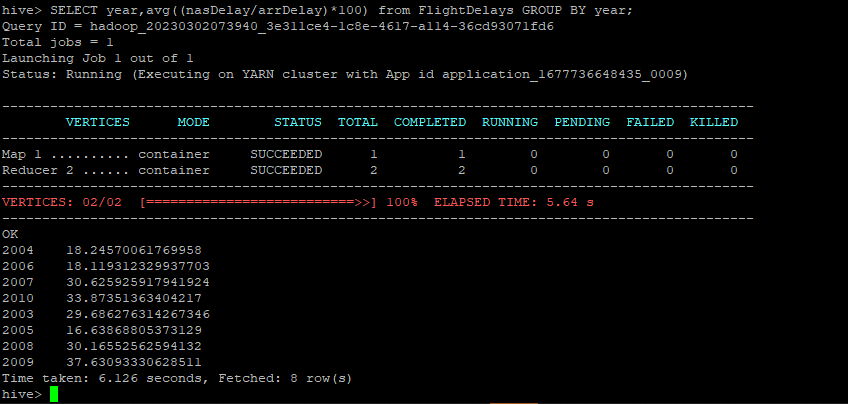
#### Output



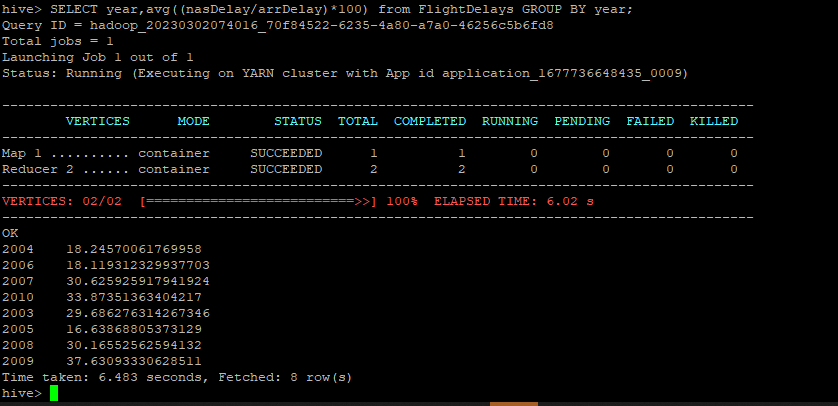
#### Iteration 1 - MR Jobs Summary



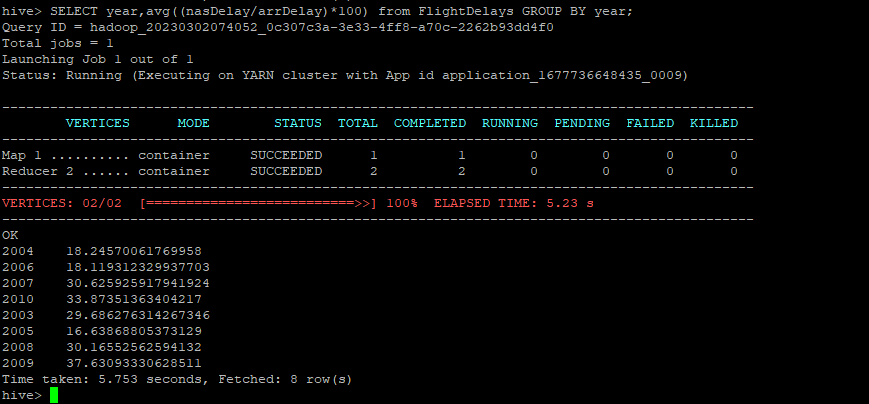
#### Iteration 2- MR Jobs Summary



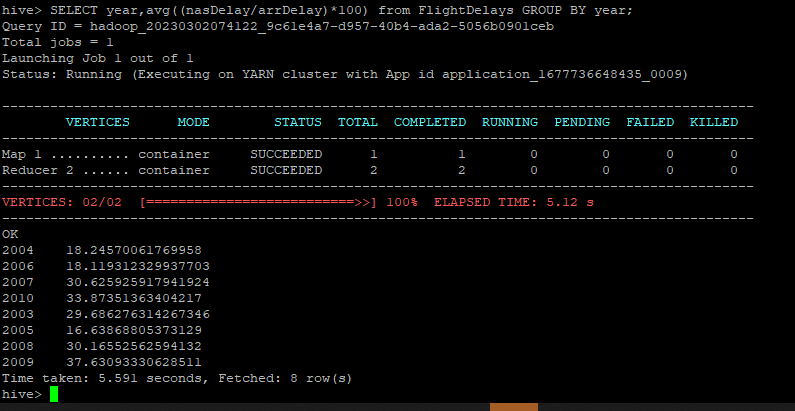
#### Iteration 3- MR Jobs Summary



#### Iteration 4 -MR Jobs Summary

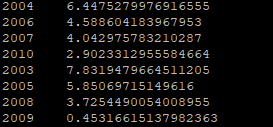


#### Iteration 5 - MR Jobs Summary

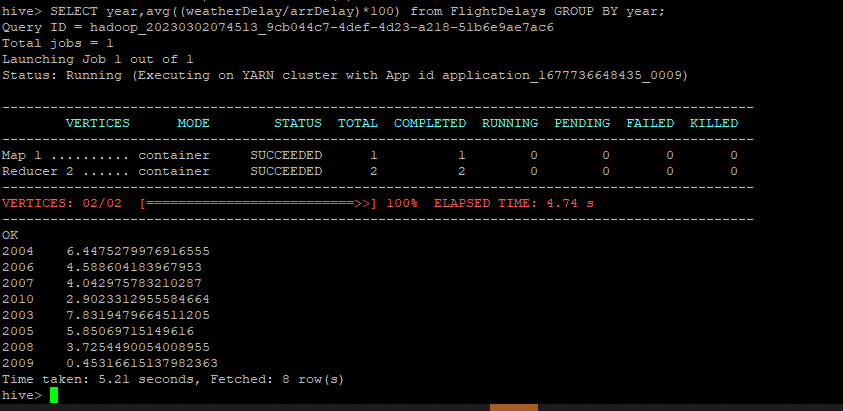


### Year wise Weather delay from 2003-2010

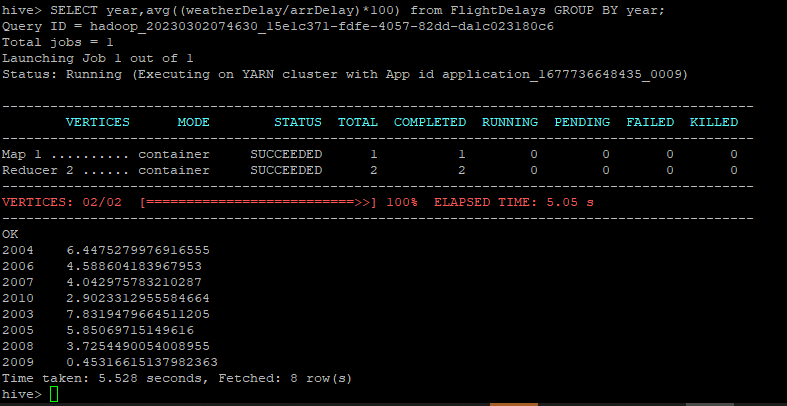
#### Output



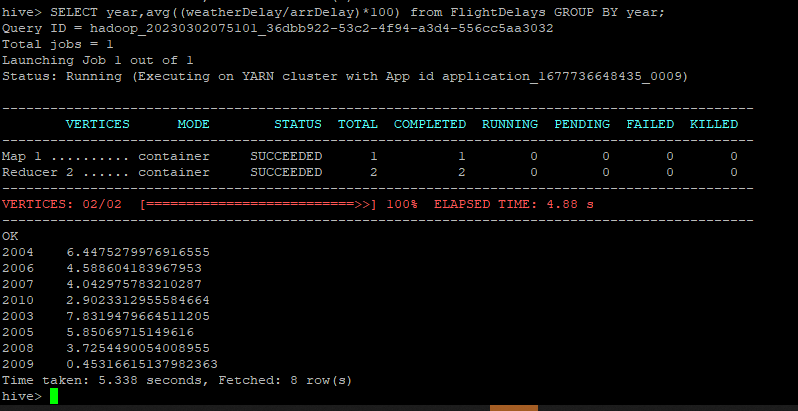
#### Iteration 1 - MR Jobs Summary



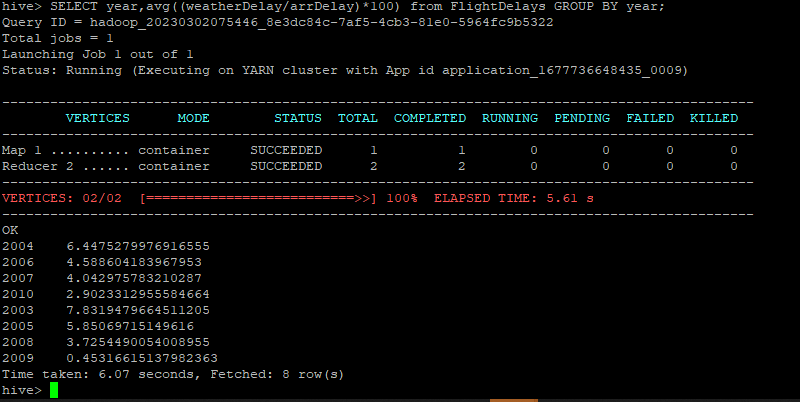
#### Iteration 2- MR Jobs Summary



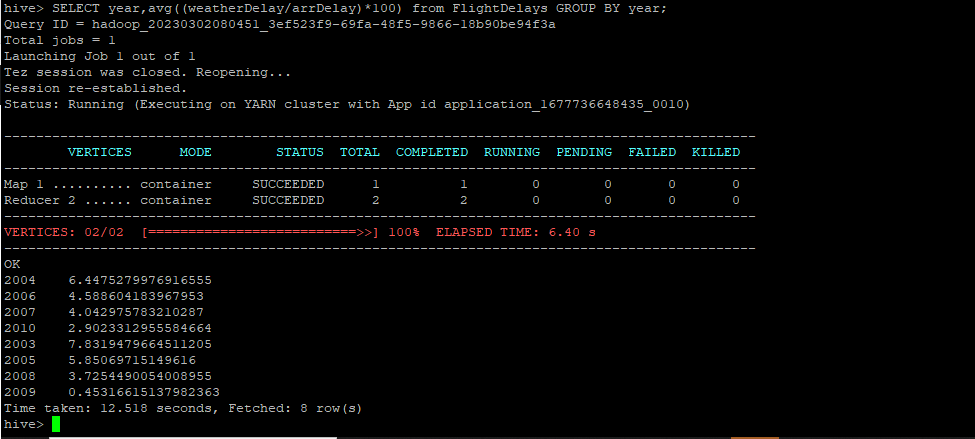
#### Iteration 3- MR Jobs Summary



#### Iteration 4 -MR Jobs Summary

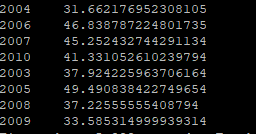


#### Iteration 5 - MR Jobs Summary

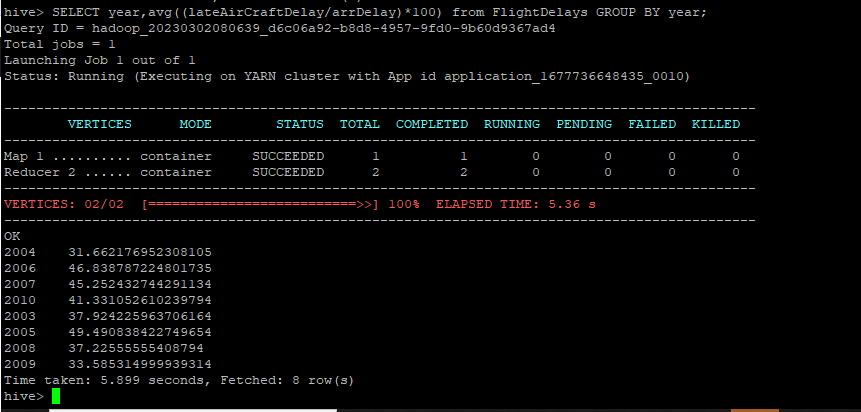


### Year wise late aircraft delay from 2003-2010

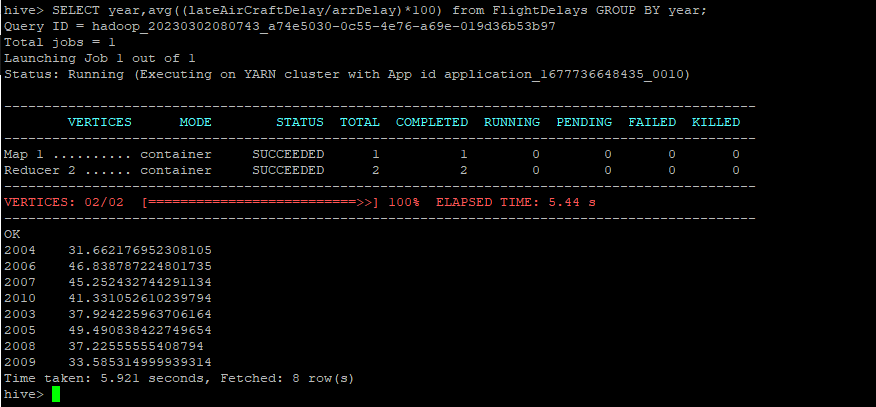
#### Output



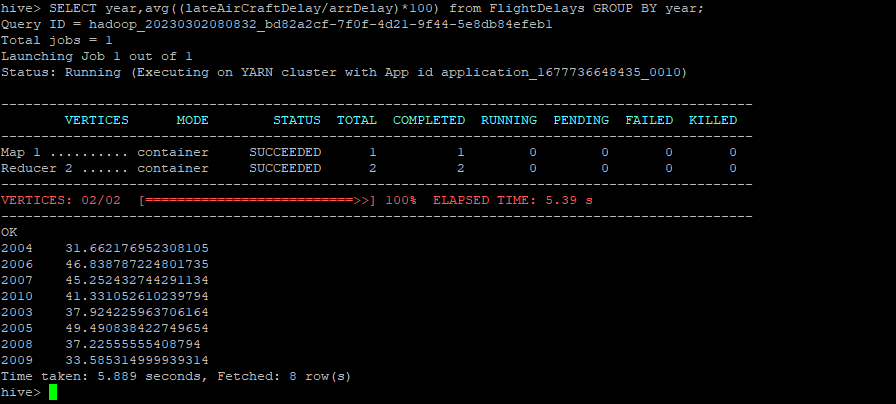
#### Iteration 1 - MR Jobs Summary



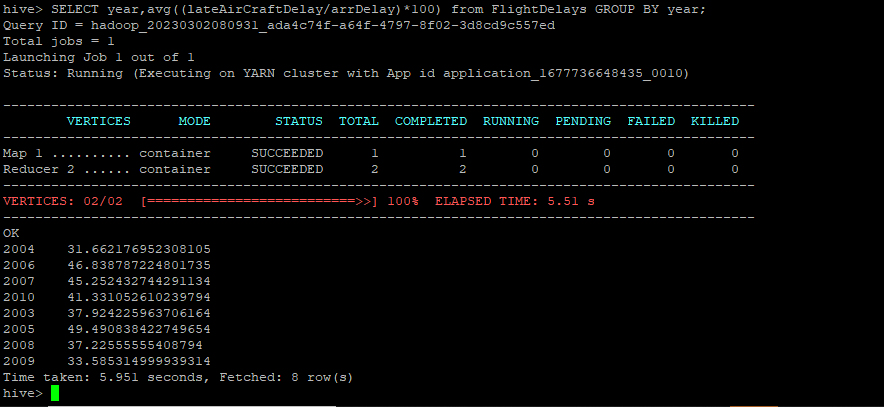
#### Iteration 2- MR Jobs Summary



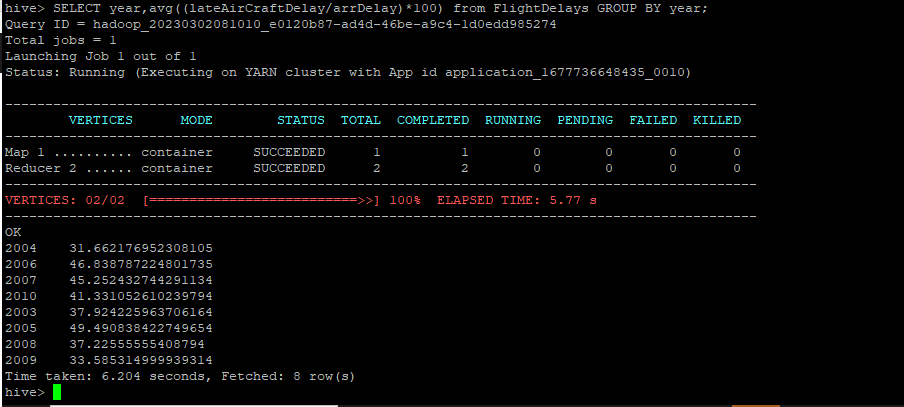
#### Iteration 3- MR Jobs Summary



#### Iteration 4 -MR Jobs Summary

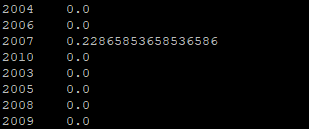


#### Iteration 5 - MR Jobs Summary

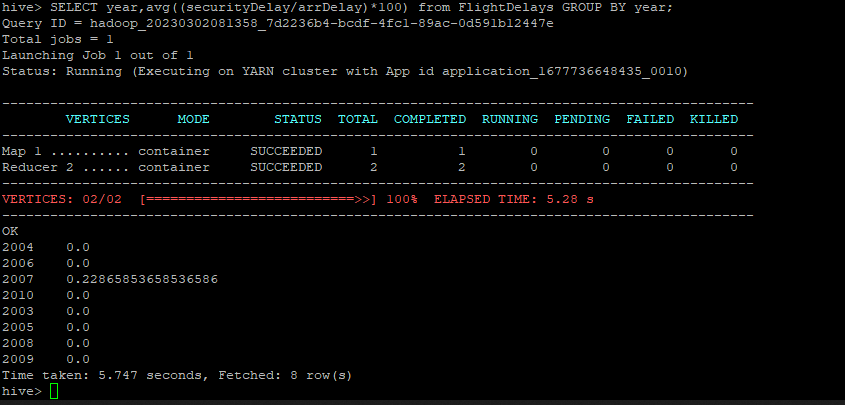


### Year wise security delay from 2003-2010

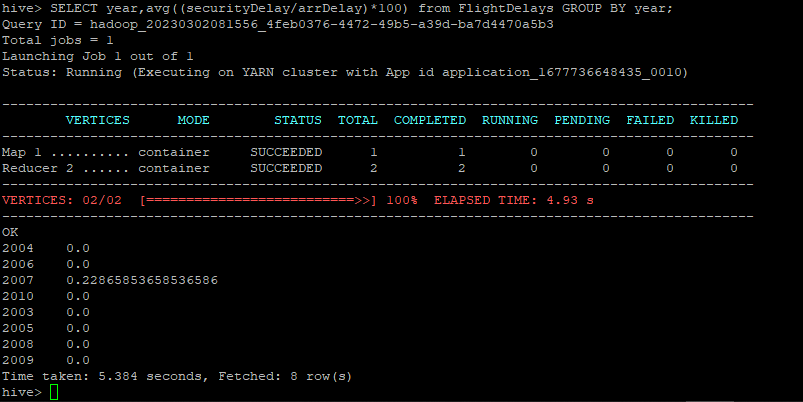
#### Output



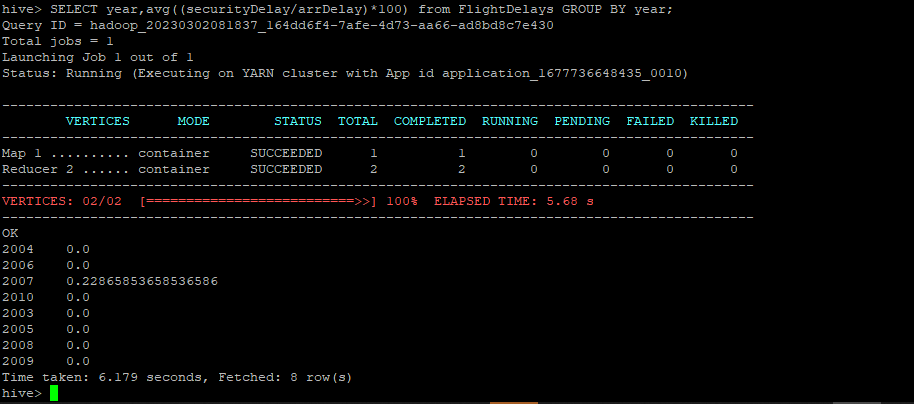
#### Iteration 1 - MR Jobs Summary



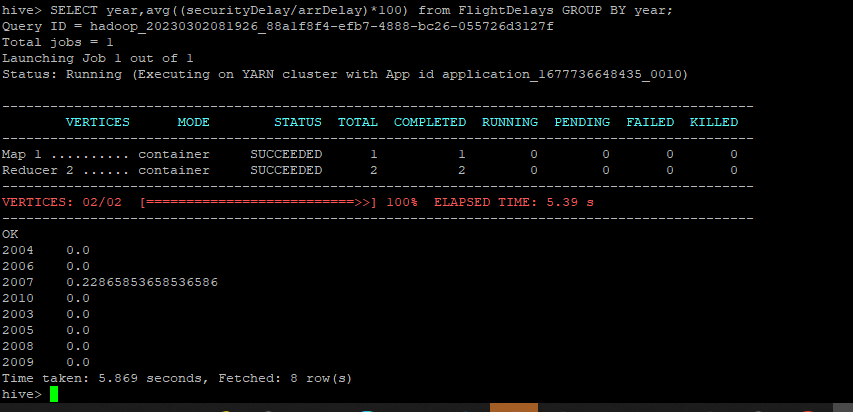
#### Iteration 2- MR Jobs Summary



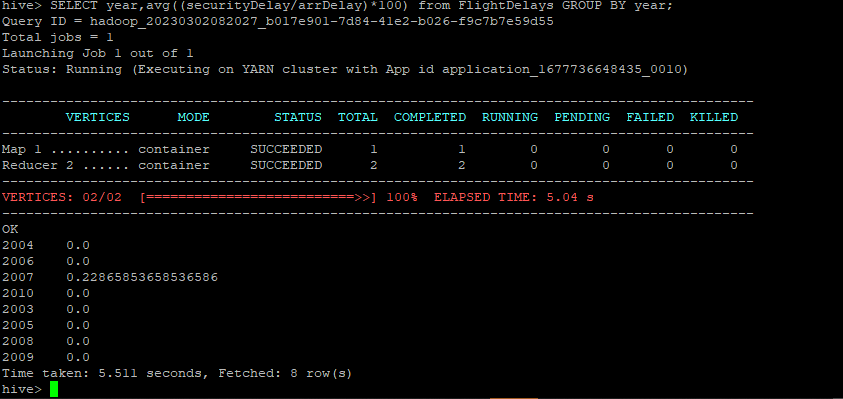
#### Iteration 3- MR Jobs Summary



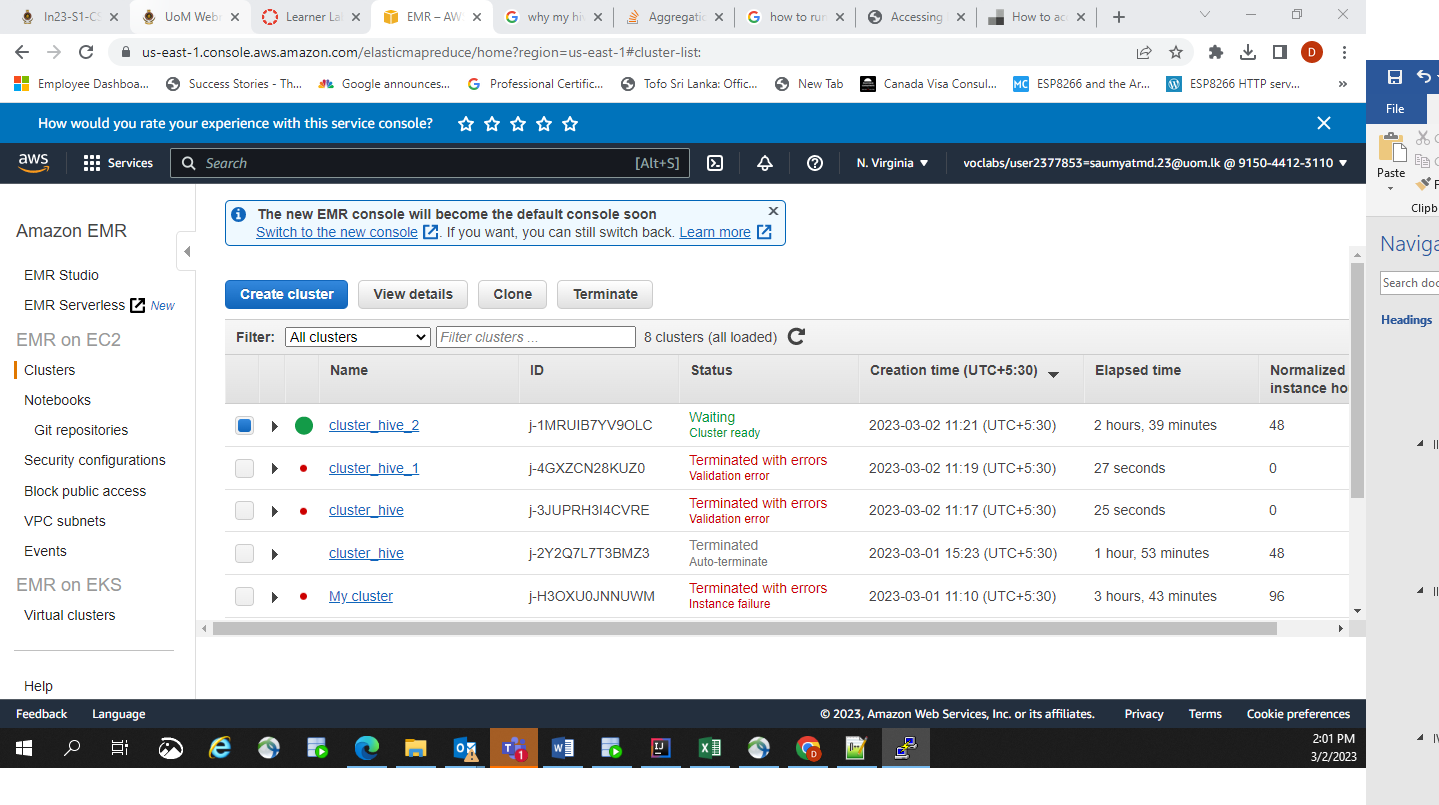
#### Iteration 4 -MR Jobs Summary

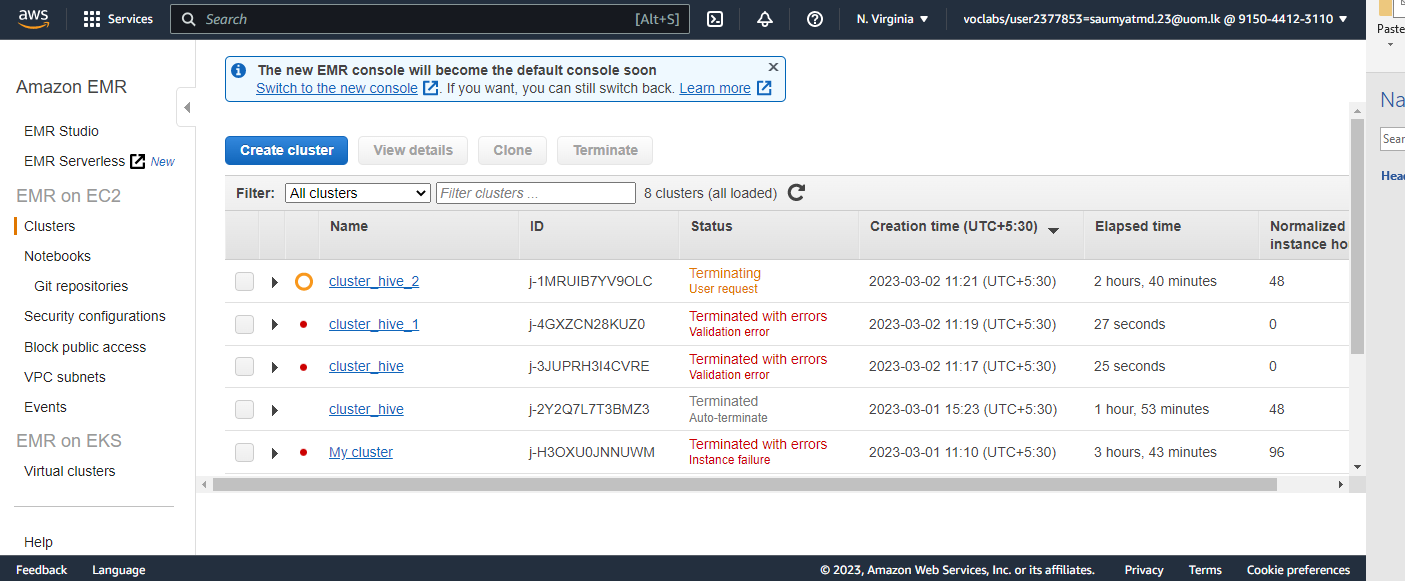


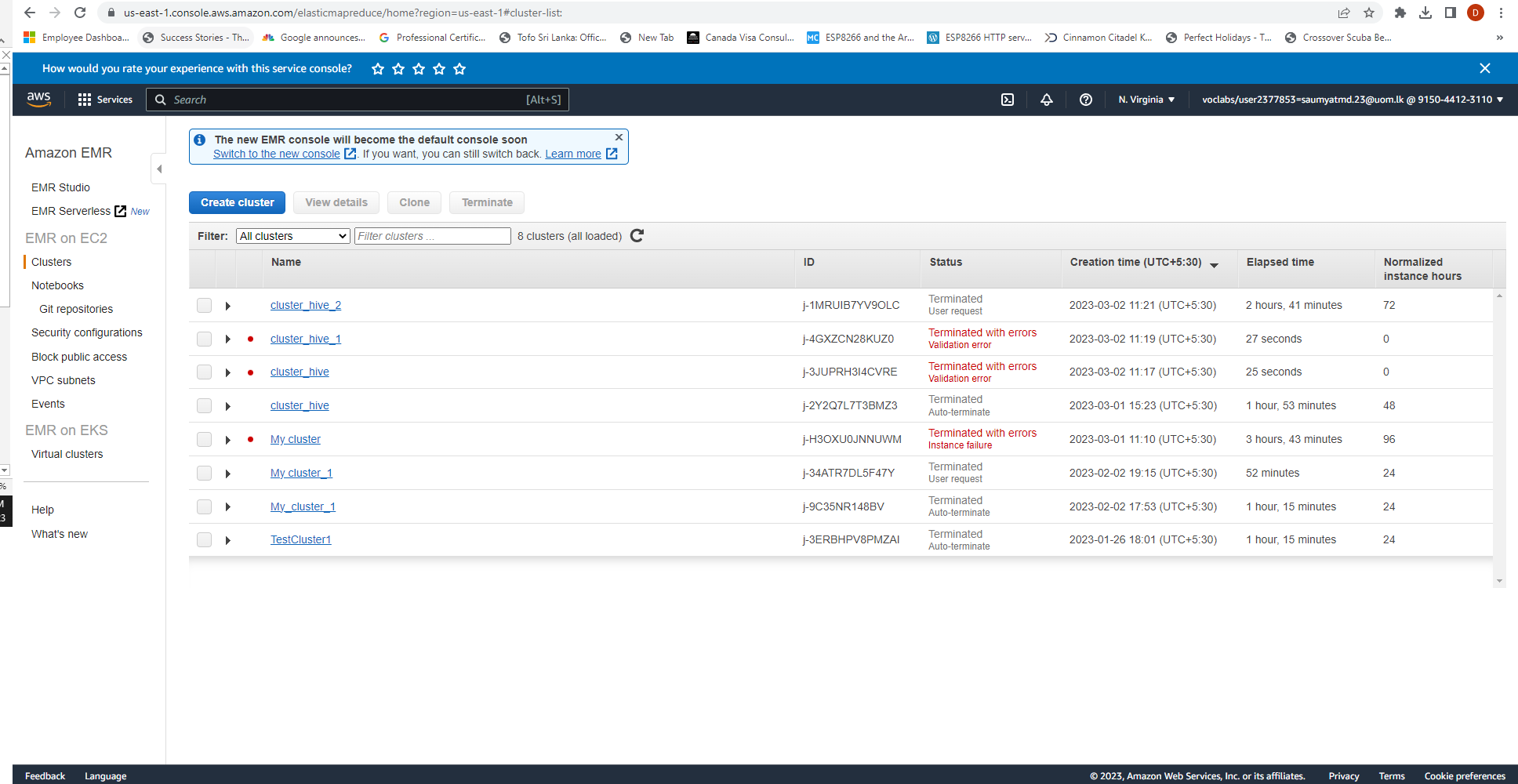
#### Iteration 5 - MR Jobs Summary



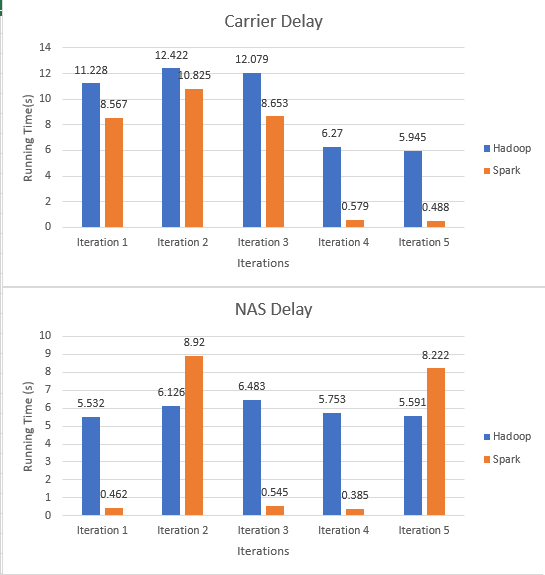
### Terminating the created cluster

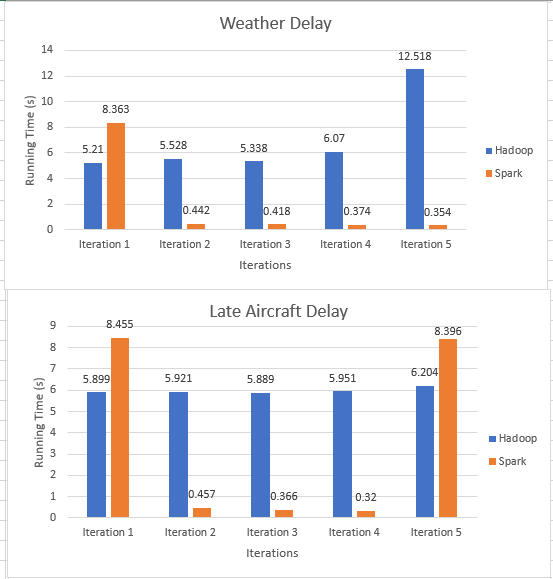


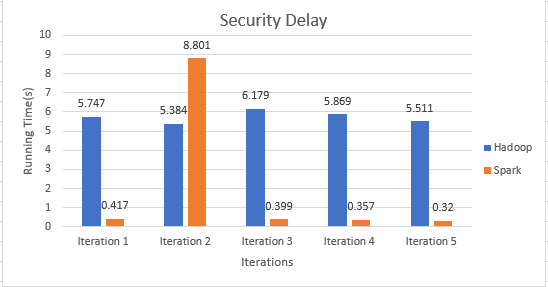




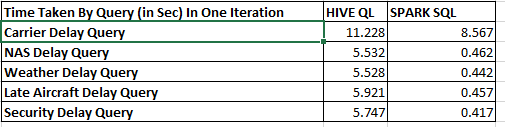
## E.  Hadoop and Spark for 5 times relates to each delay type







## F. Table below shows the comparison for all the queries with sample data for an iteration



## G. Average time taken by MapReduce and Spark for each query

