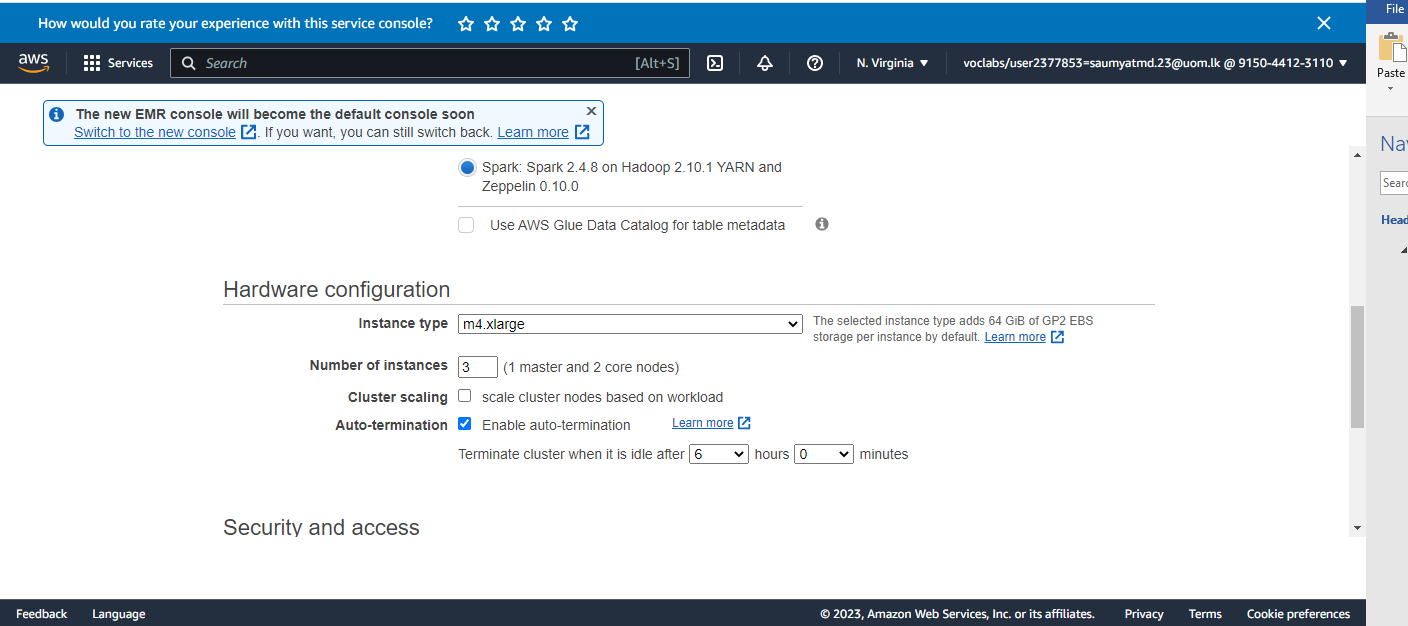
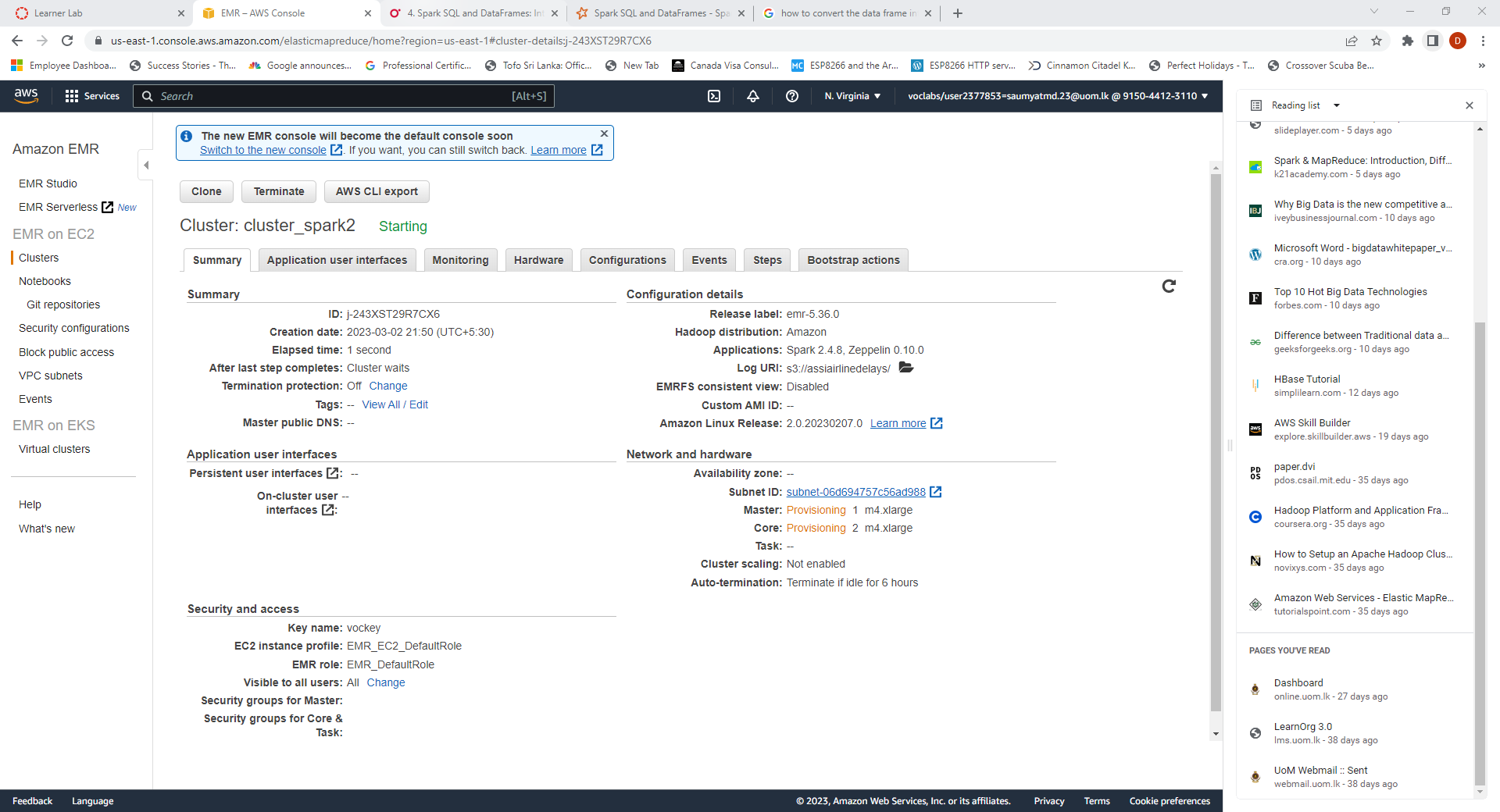
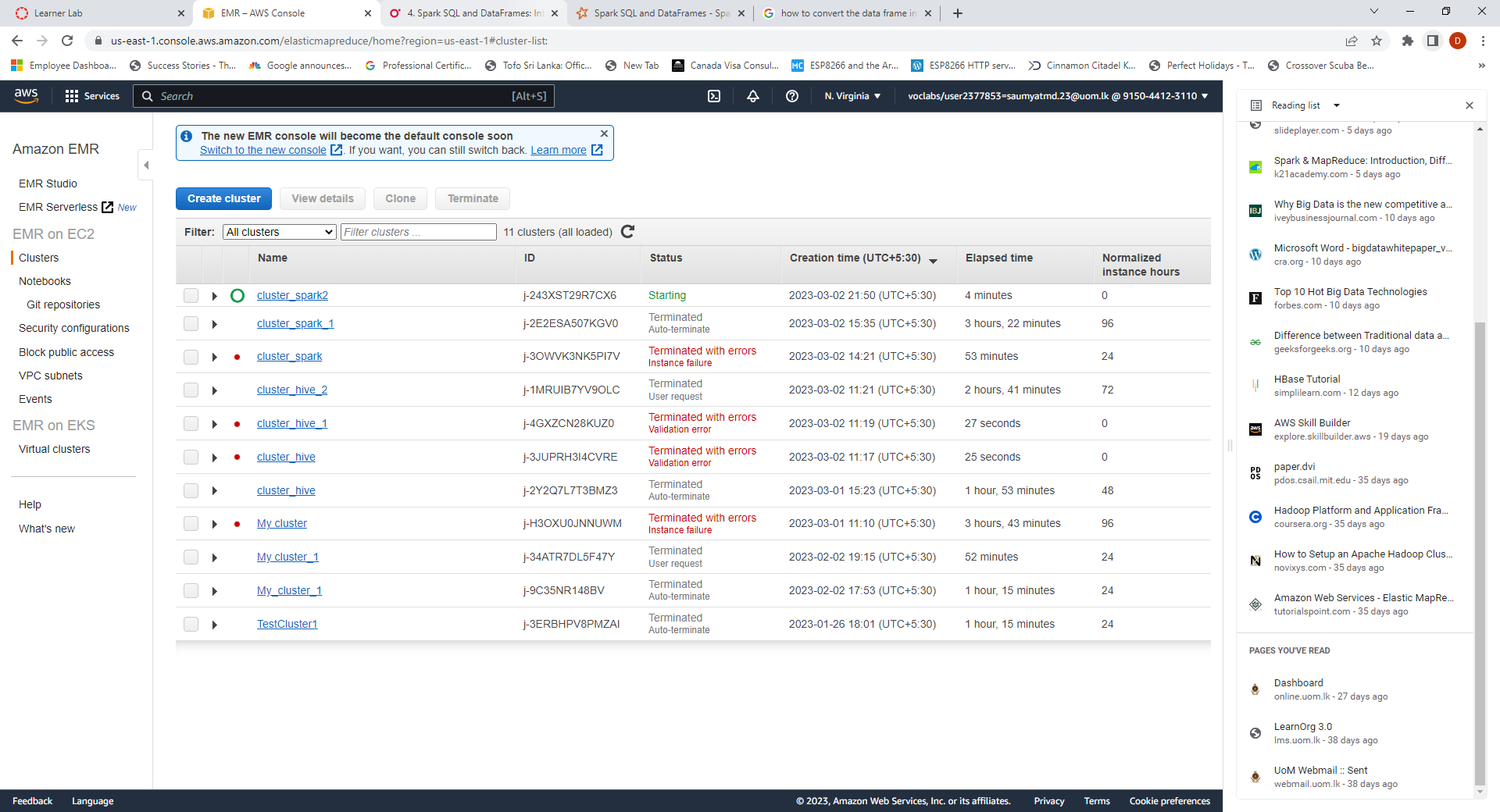
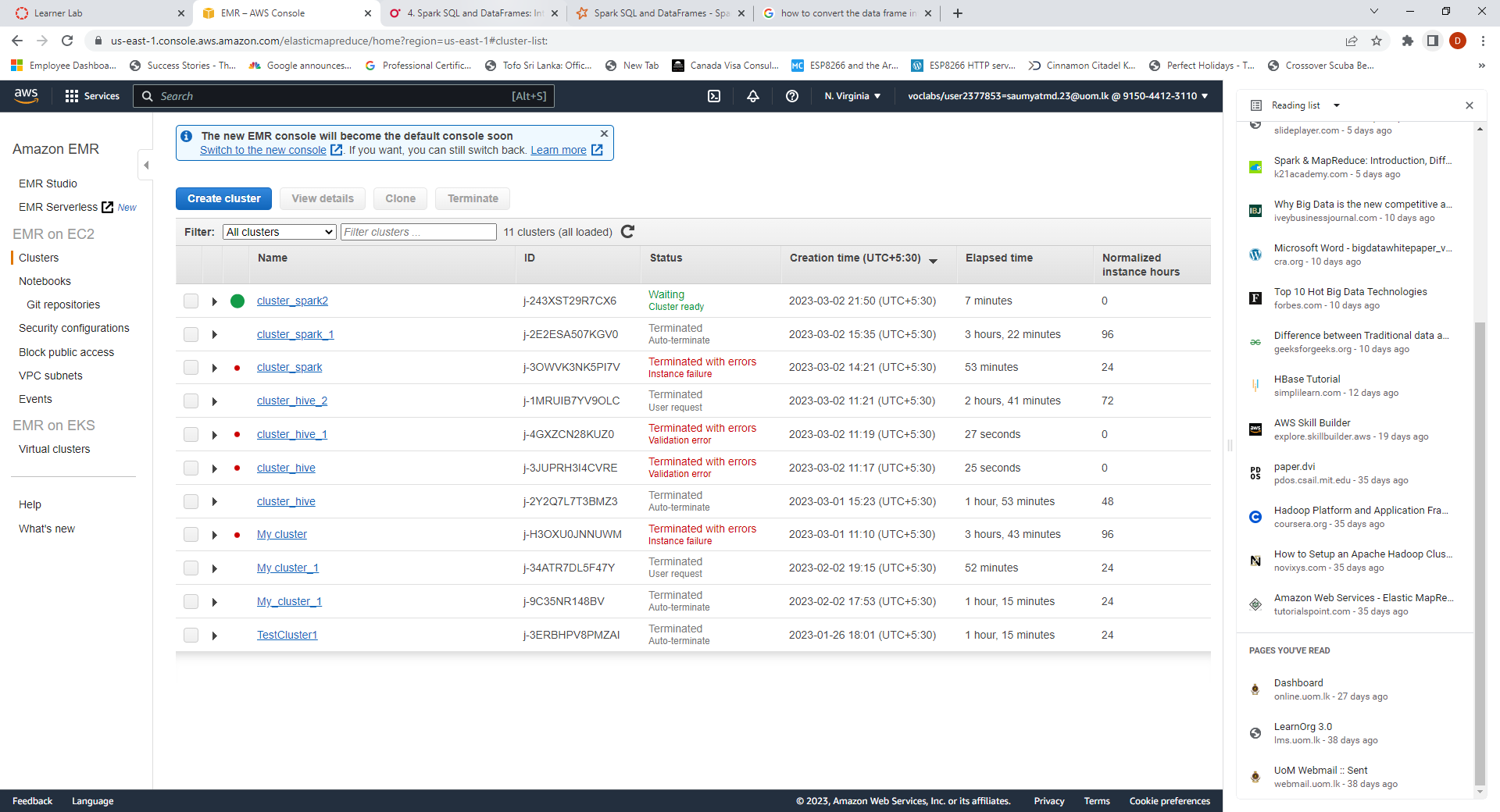
## EMR Hadoop cluster creation and to run hive queries



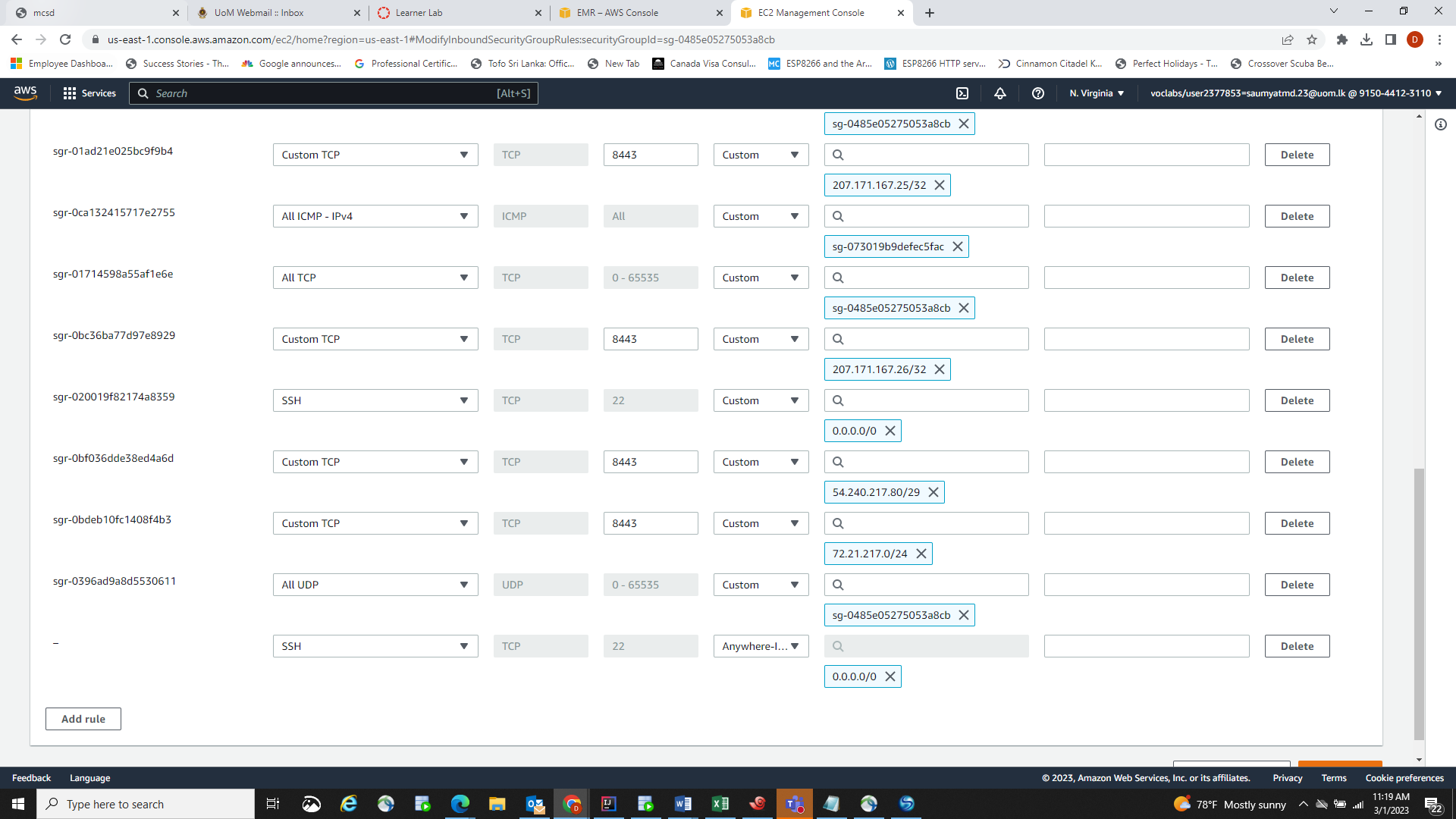




### Cluster started



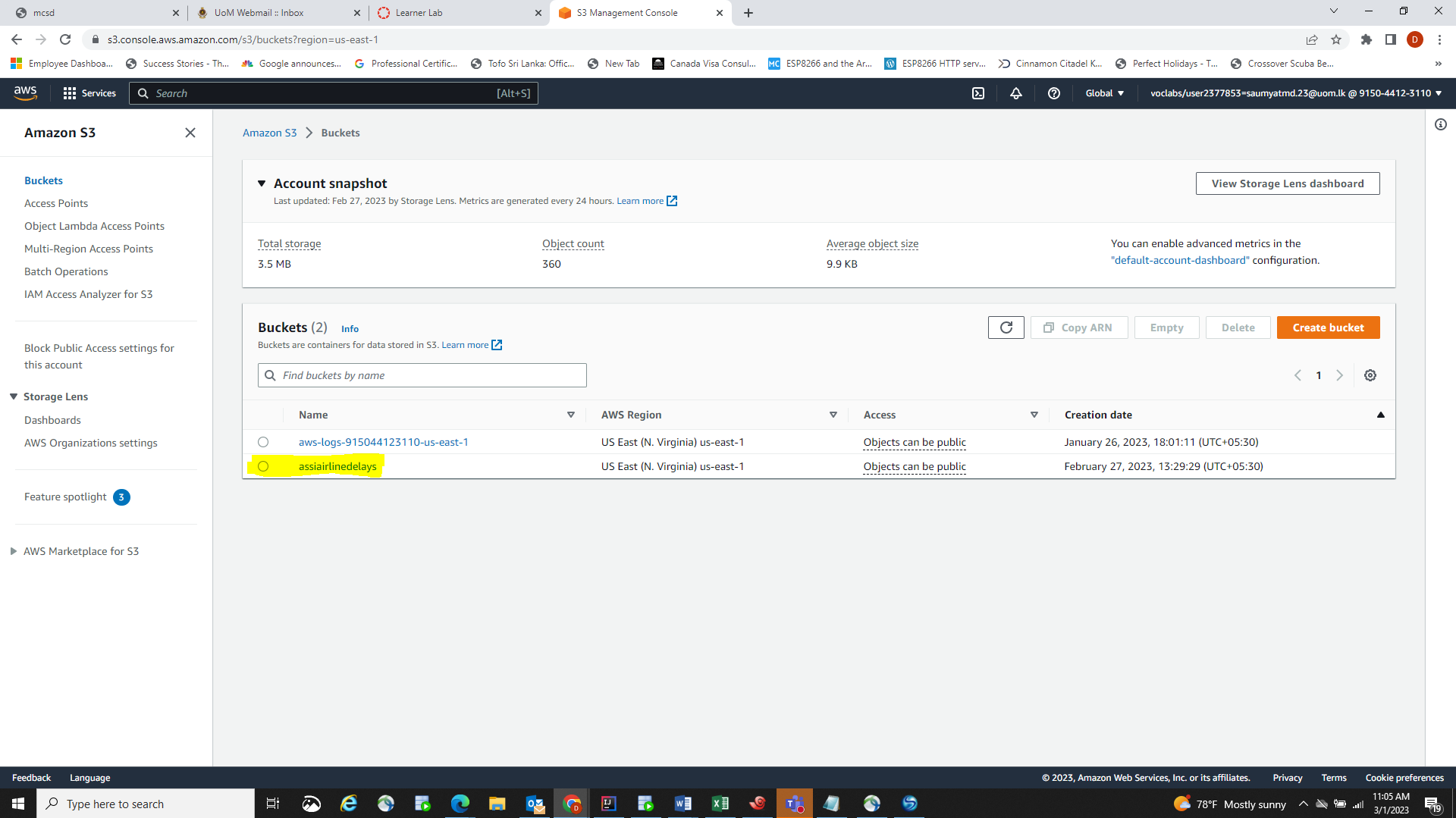
### Adding inbound rule to master rule



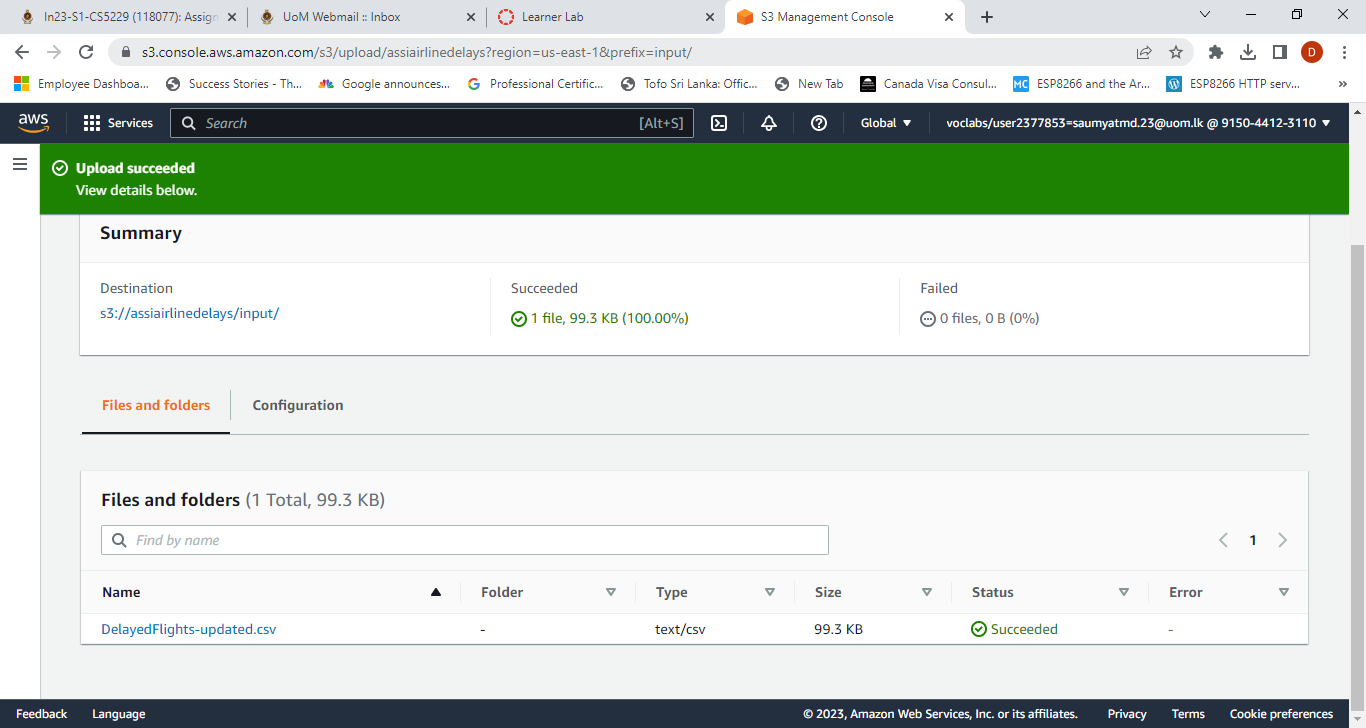
### Connecting to master node of the cluster using SSH connection using vockey key



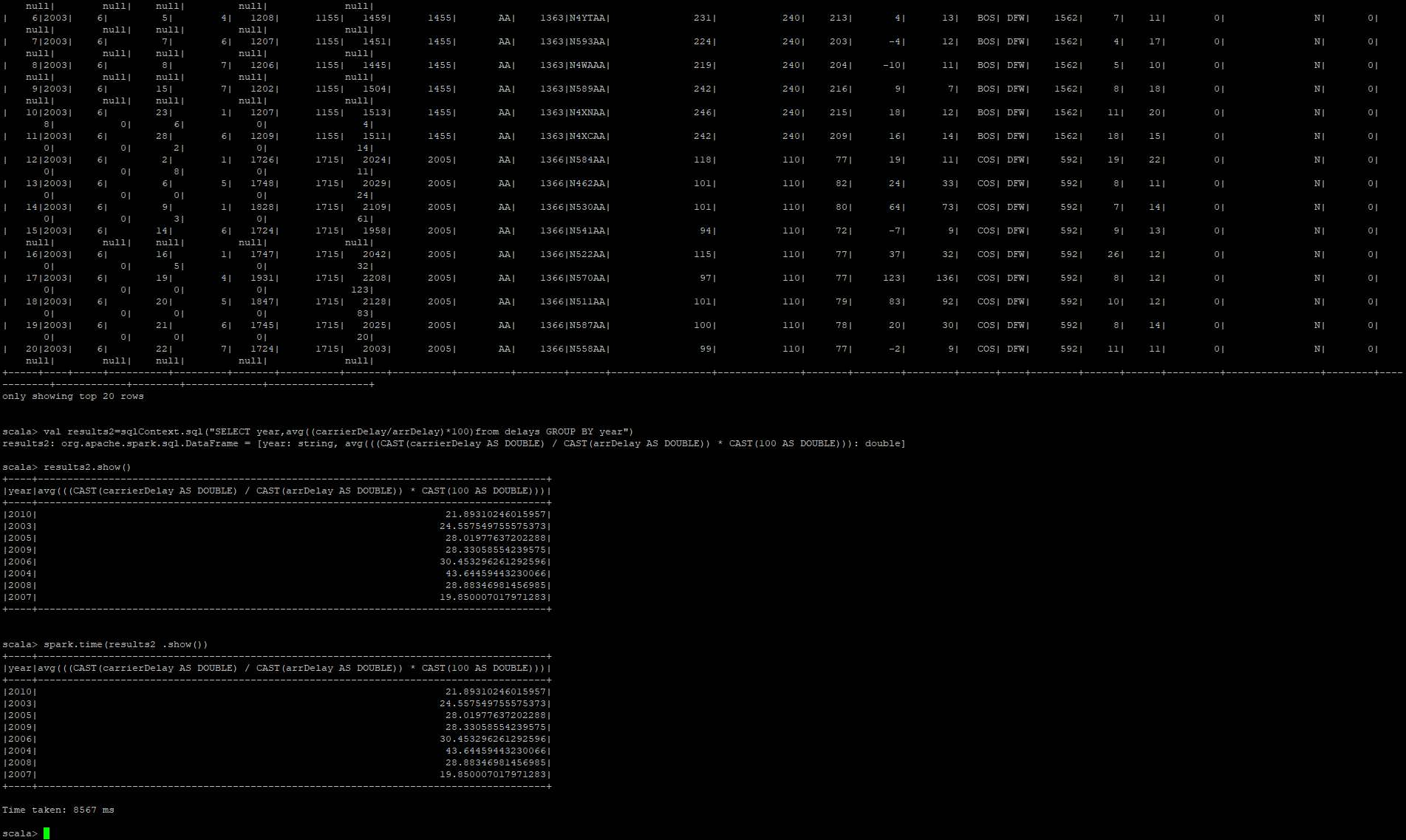
## S3 bucket created previously to run MR will be used here



## Uploading data set to S3 bucket – input folder



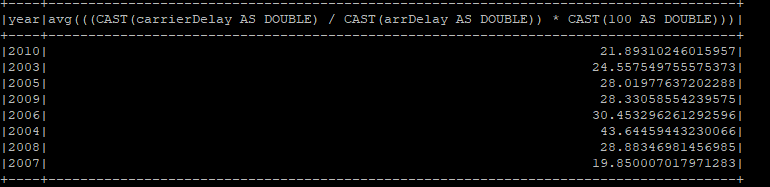




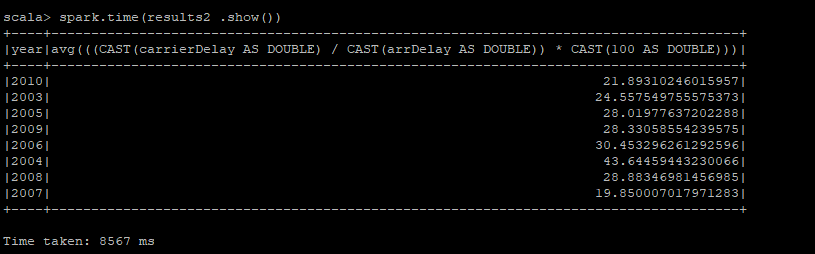
## Analysis of flight delays

### Year wise carrier delay from 2003-2010

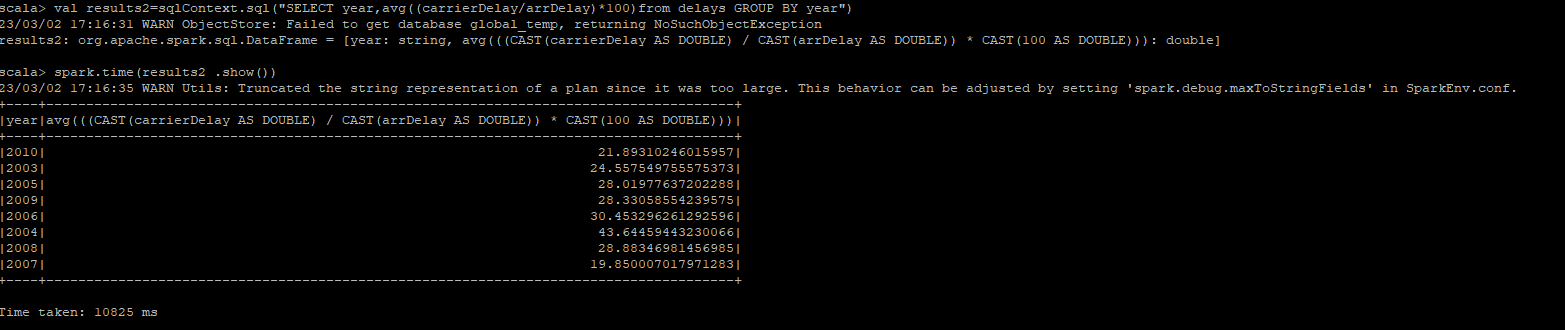
#### Output



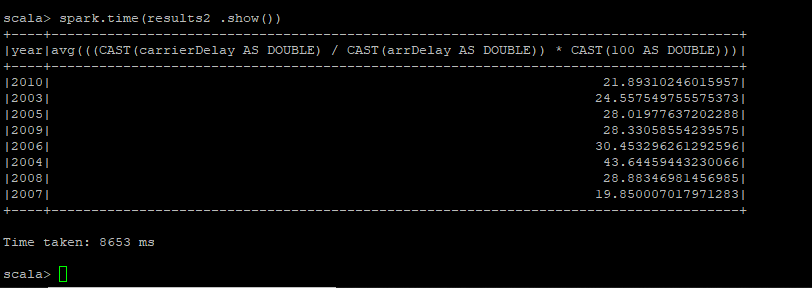
#### Iteration 1 -Spark Time Summary



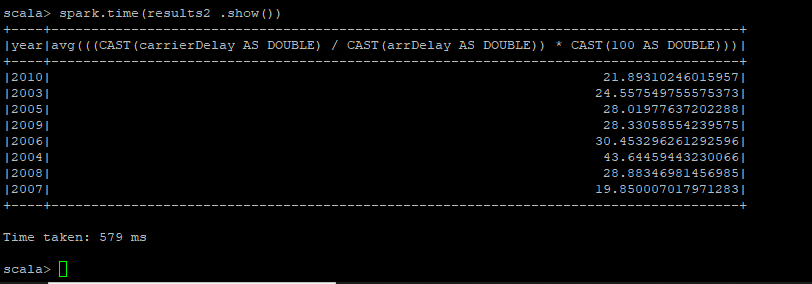
#### Iteration 2 -Spark Time Summary



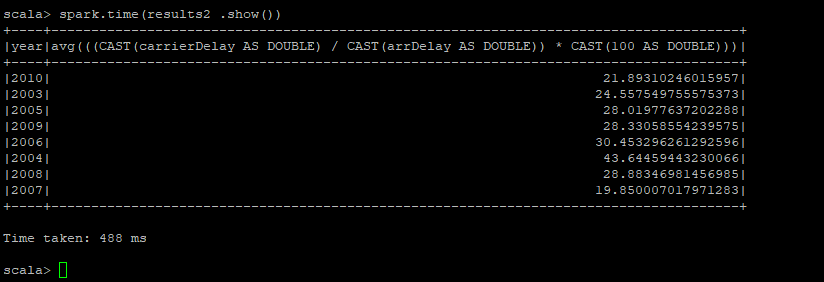
#### Iteration 3 -Spark Time Summary



#### Iteration 4 -Spark Time Summary

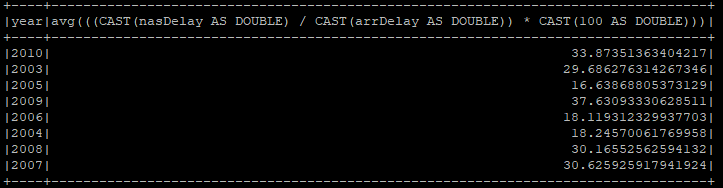


#### Iteration 5 -Spark Time Summary

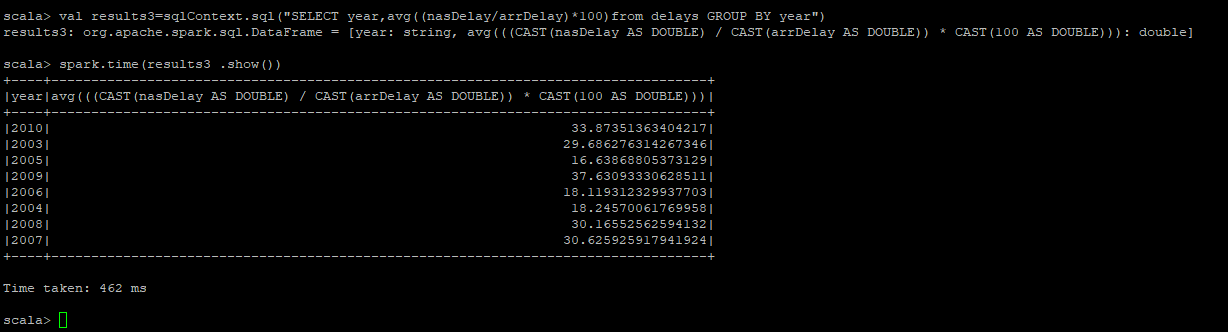


### Year wise NAS delay from 2003-2010

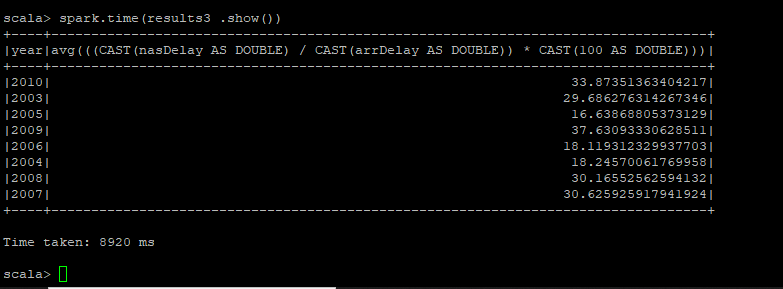
#### Output



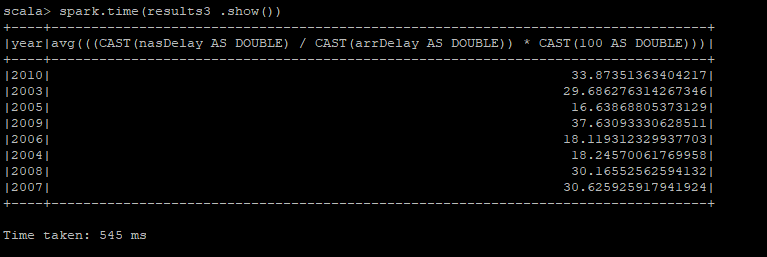
#### Iteration 1 -Spark Time Summary



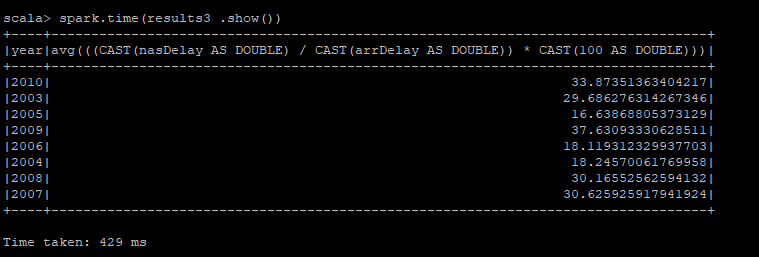
#### Iteration 2 -Spark Time Summary



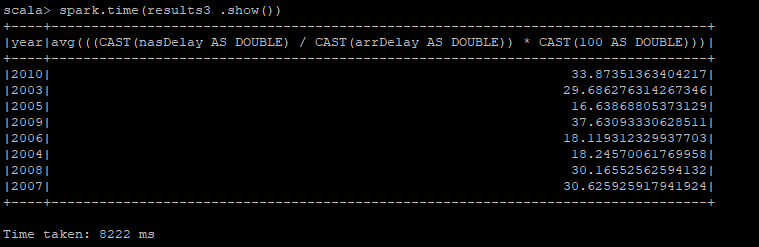
#### Iteration 3 -Spark Time Summary



#### Iteration 4 -Spark Time Summary

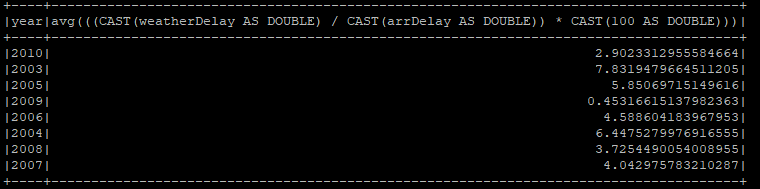


#### Iteration 5 -Spark Time Summary

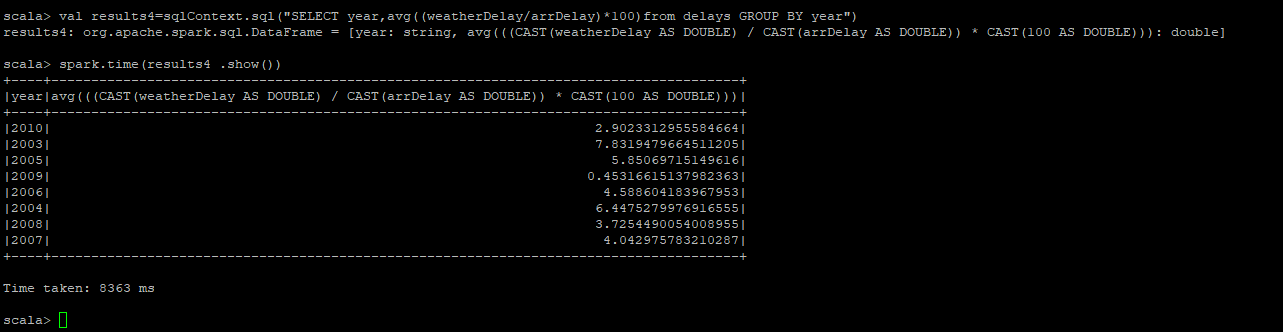


### Year wise Weather delay from 2003-2010

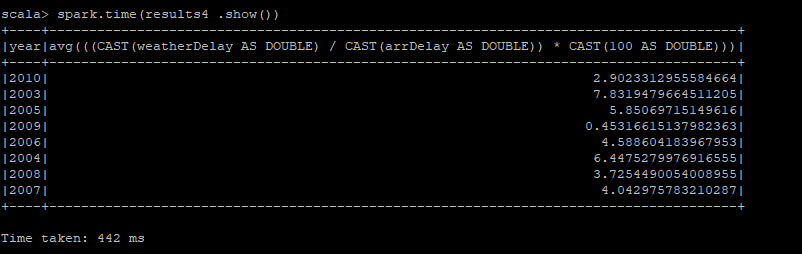
#### Output



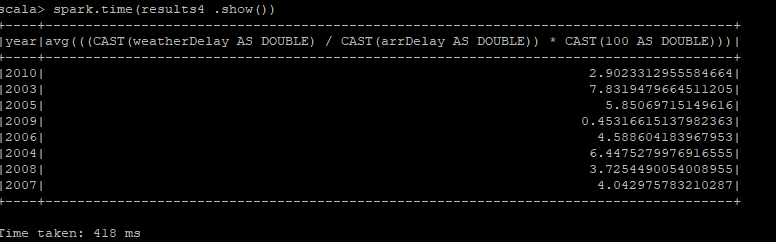
#### Iteration 1 -Spark Time Summary



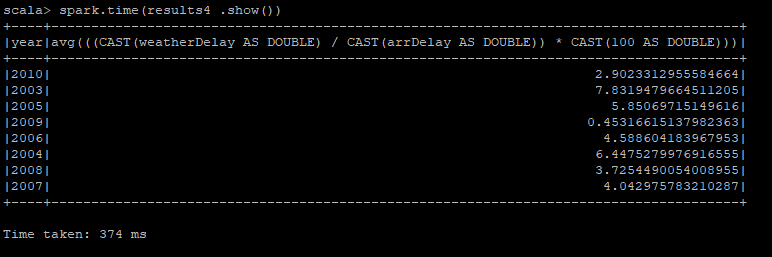
#### Iteration 2 -Spark Time Summary



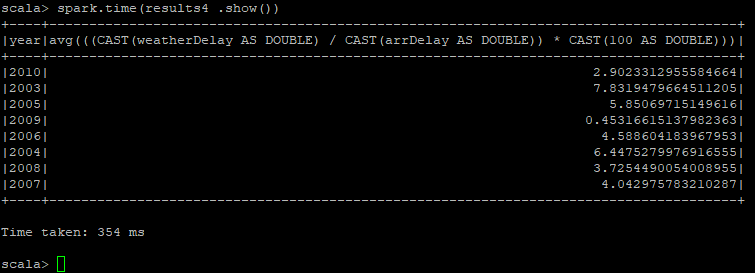
#### Iteration 3 -Spark Time Summary



#### Iteration 4 -Spark Time Summary

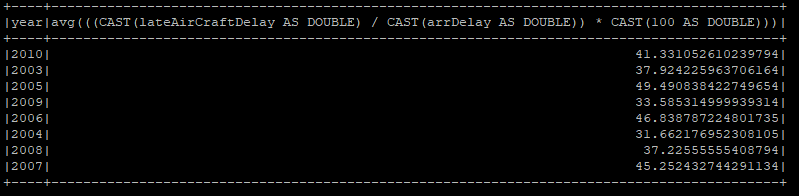


#### Iteration 5 -Spark Time Summary

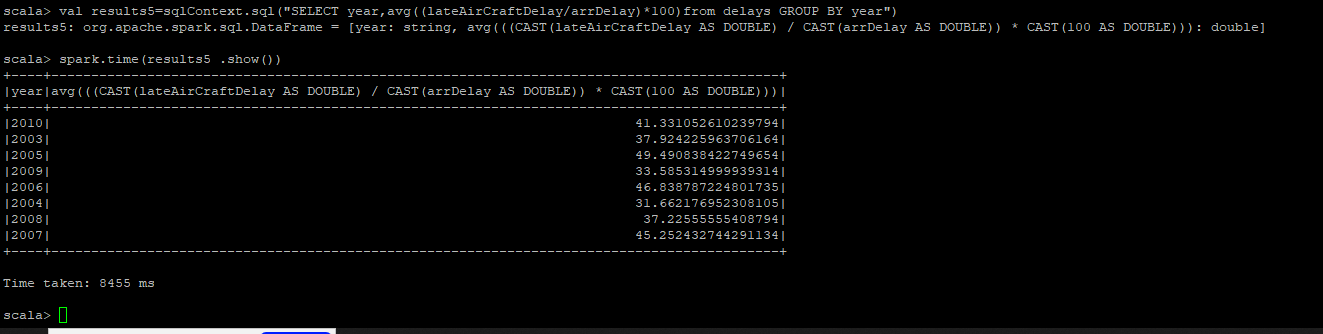


### Year wise late aircraft delay from 2003-2010

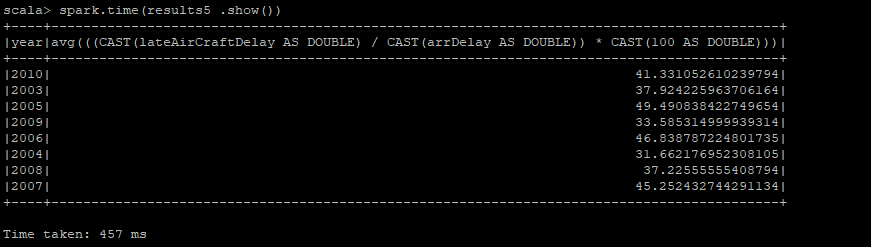
#### Output



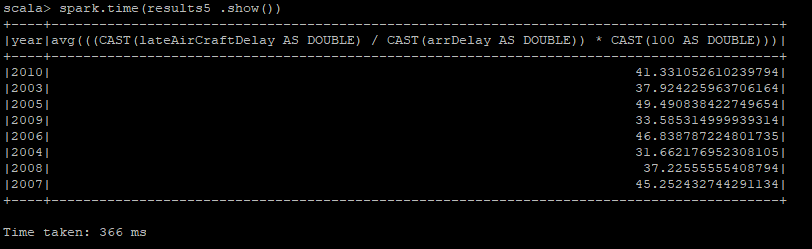
#### Iteration 1 -Spark Time Summary



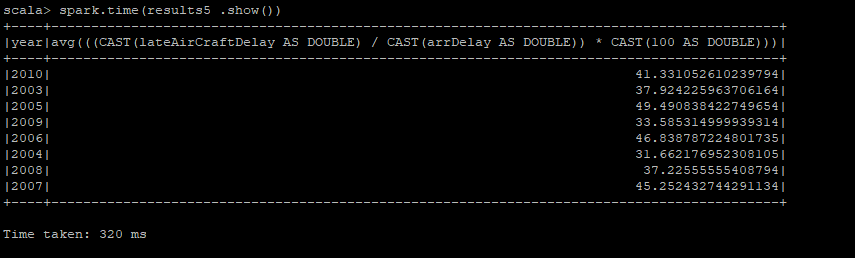
#### Iteration 2 -Spark Time Summary



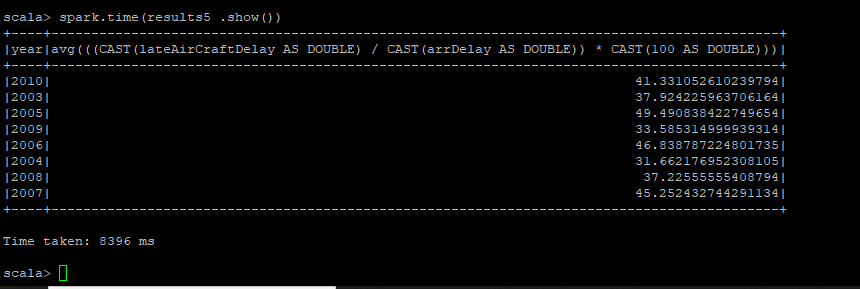
#### Iteration 3 -Spark Time Summary



#### Iteration 4 -Spark Time Summary

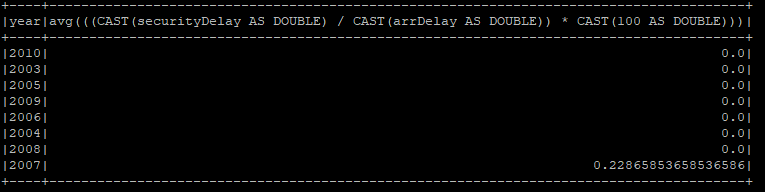


#### Iteration 5 -Spark Time Summary

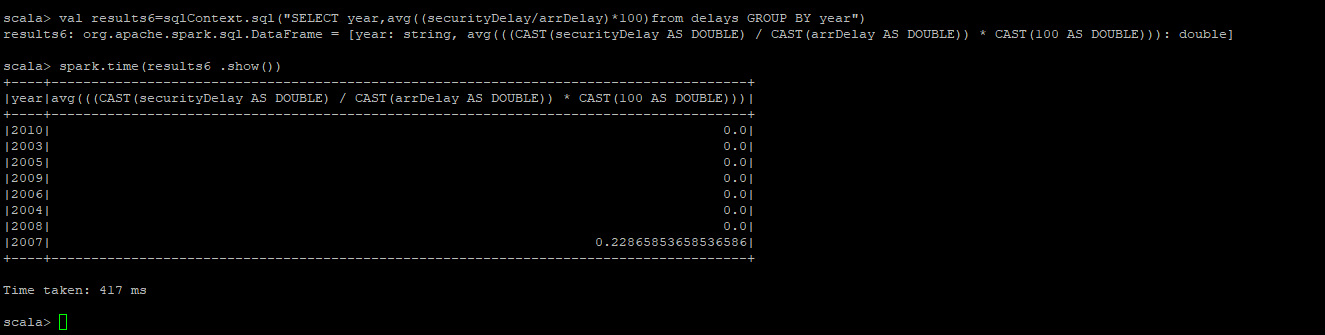


### Year wise security delay from 2003-2010

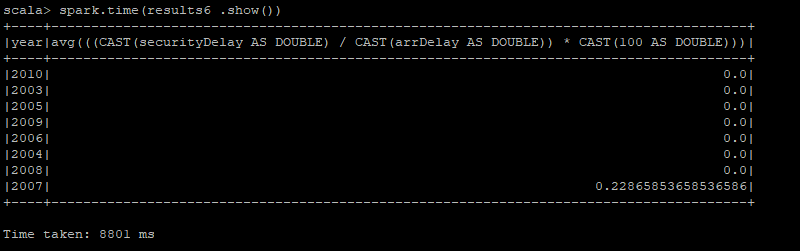
#### Output



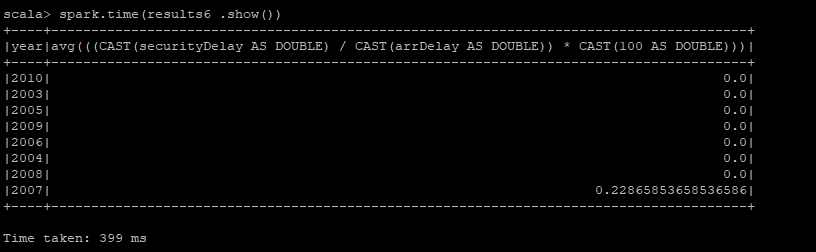
#### Iteration 1 -Spark Time Summary



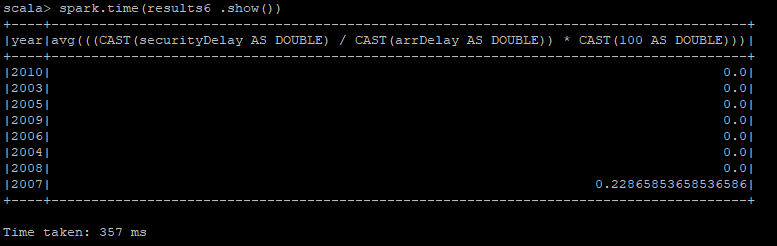
#### Iteration 2 -Spark Time Summary



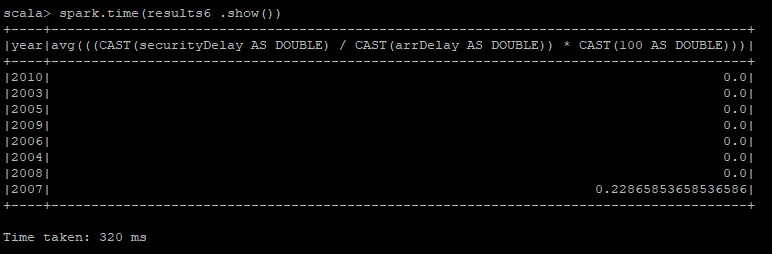
#### Iteration 3 -Spark Time Summary



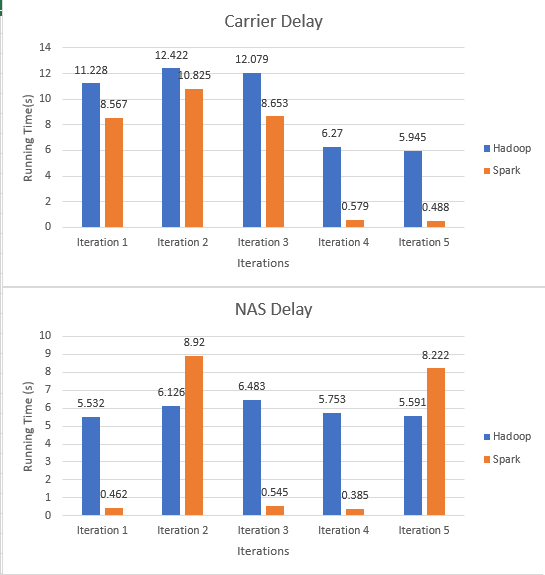
#### Iteration 4 -Spark Time Summary

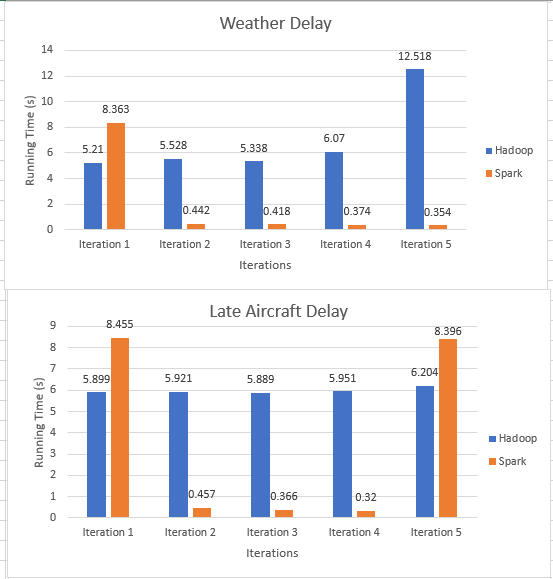


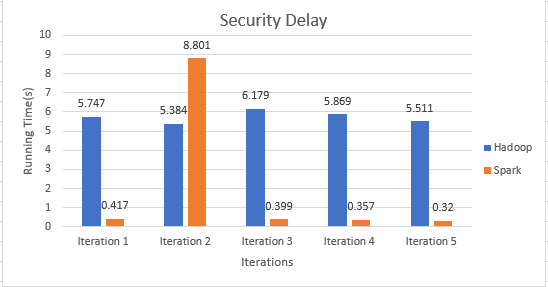
#### Iteration 5 -Spark Time Summary



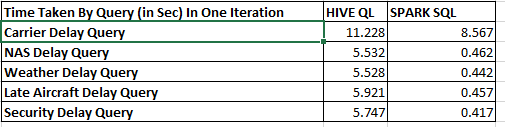
## E.  Hadoop and Spark for 5 times relates to each delay type







## 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

