

All demos are kept at user\_repo/Demos path and datasets at user\_repo/Datasets

Use spark-shell to open the spark environment

Copy datasets to HDFS path

To access any file from HDFS we use below command

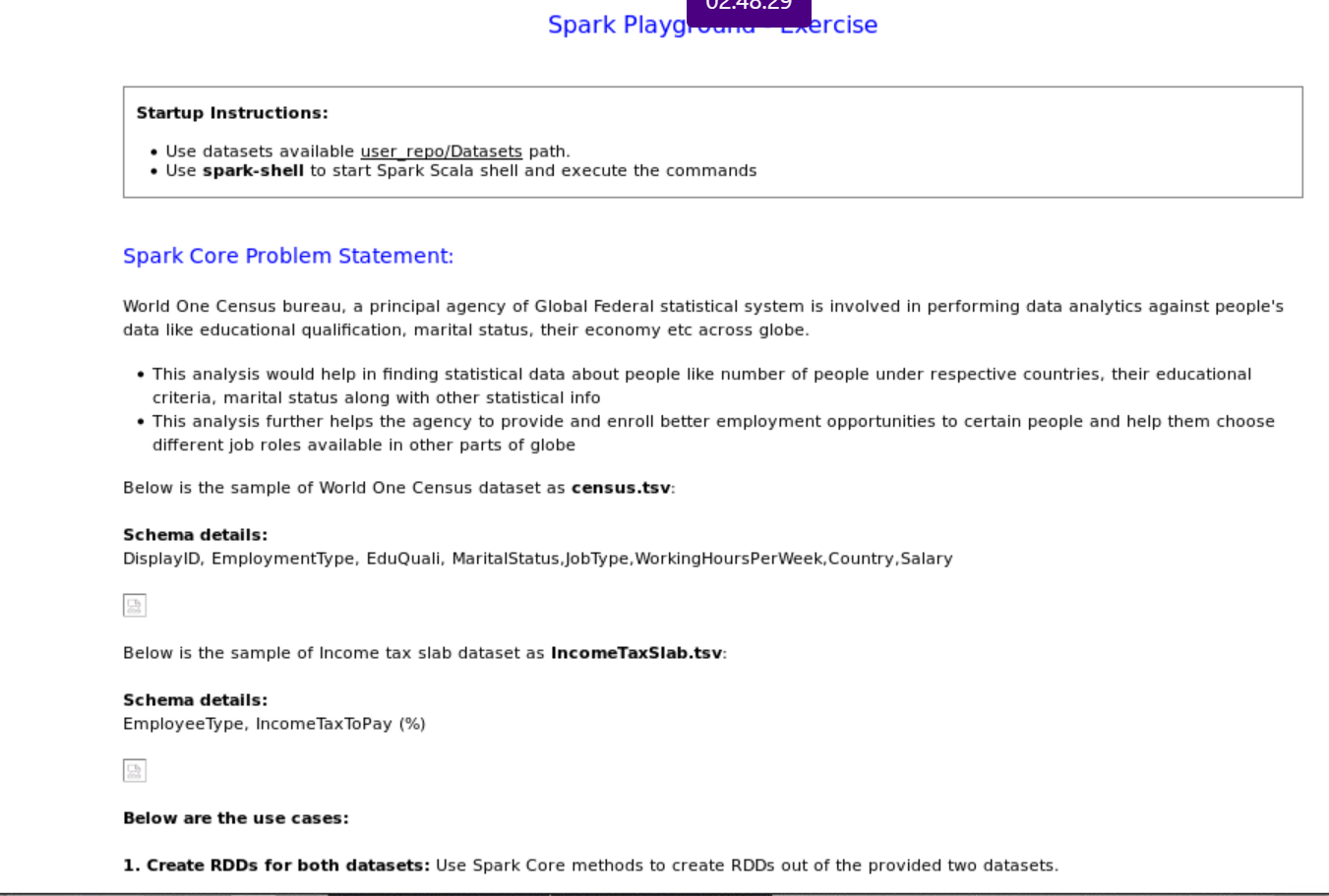
val rdd1 = sc.textFIle(“hdfs://podSpark”); # it takes hdfs path

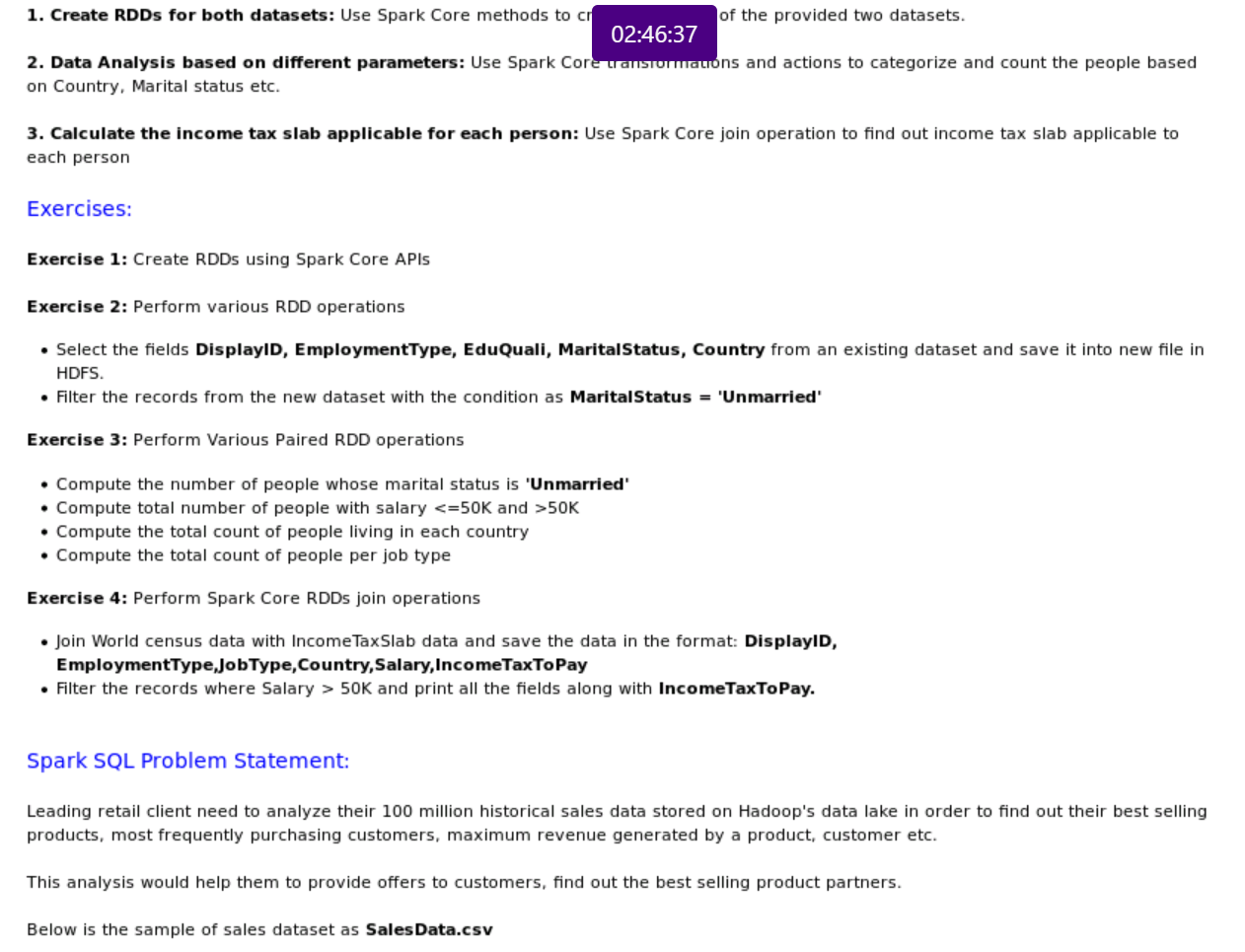
Spark Core Demos

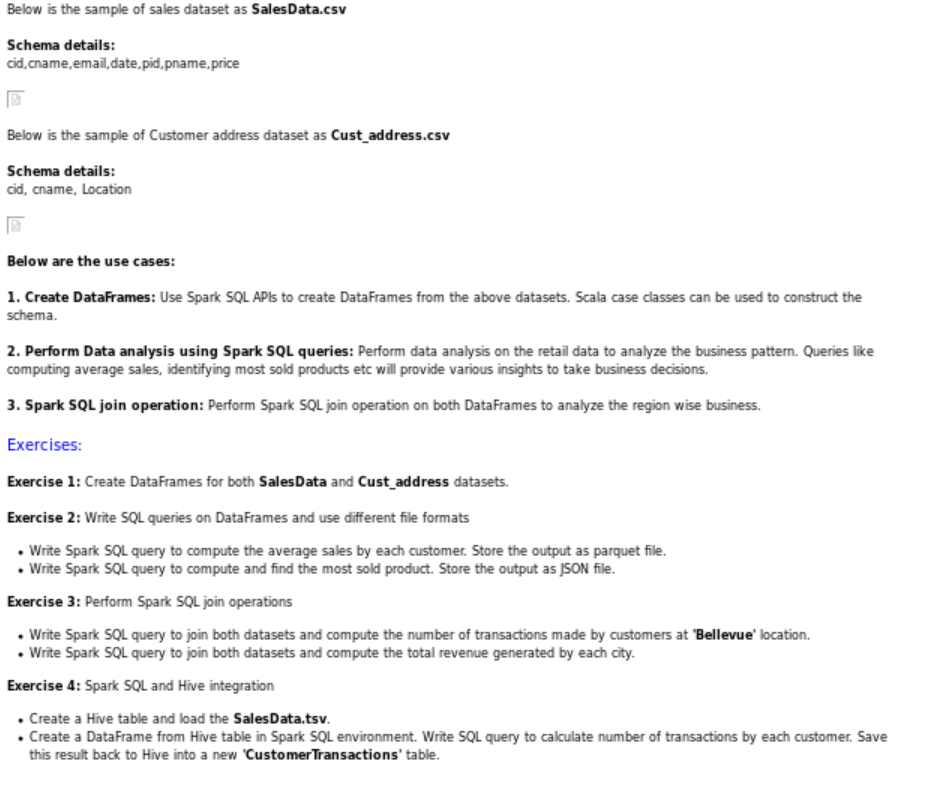
Execute the commands kept inside Desktop/user\_repo/Demos/SparkCoreDemo1,2,3

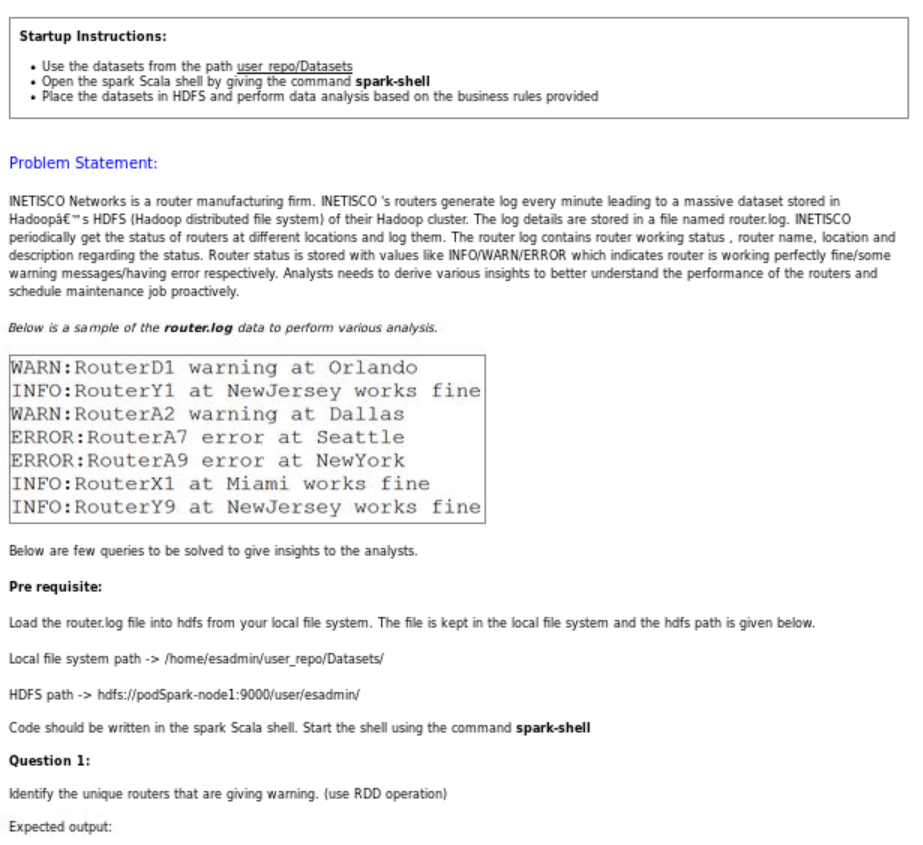
Spark SQL Demos

Desktop/user\_repo/Demos/SparkSQLDemo1,2,3



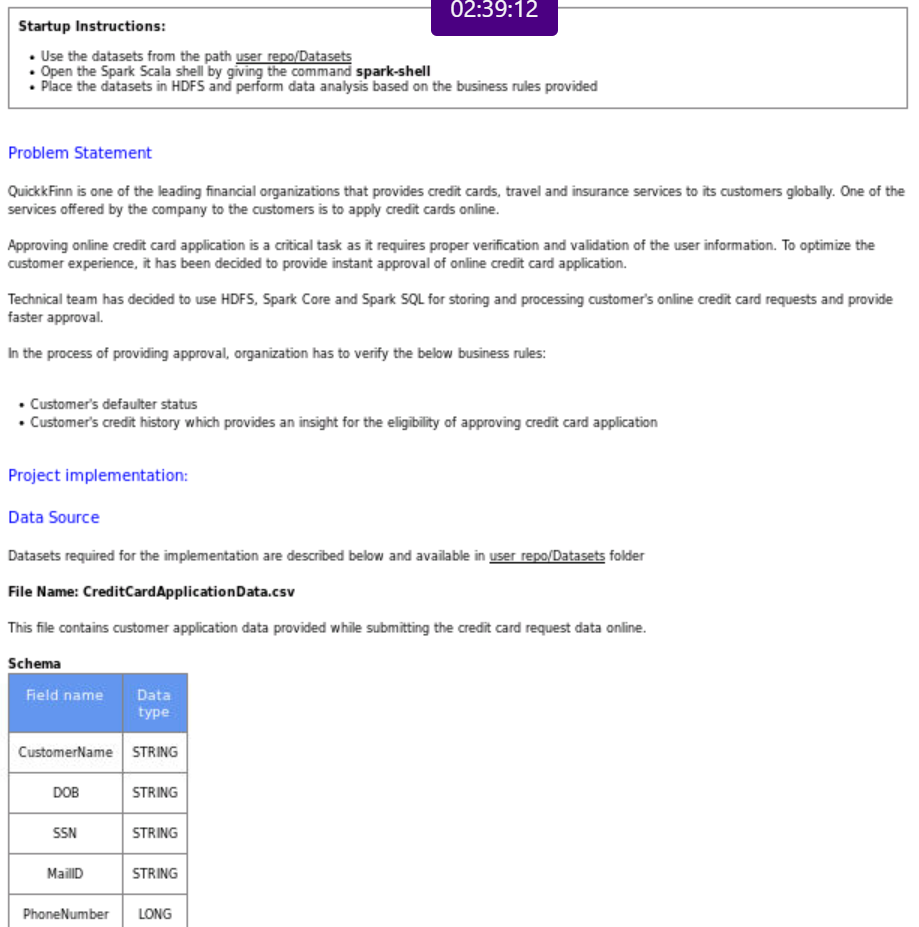


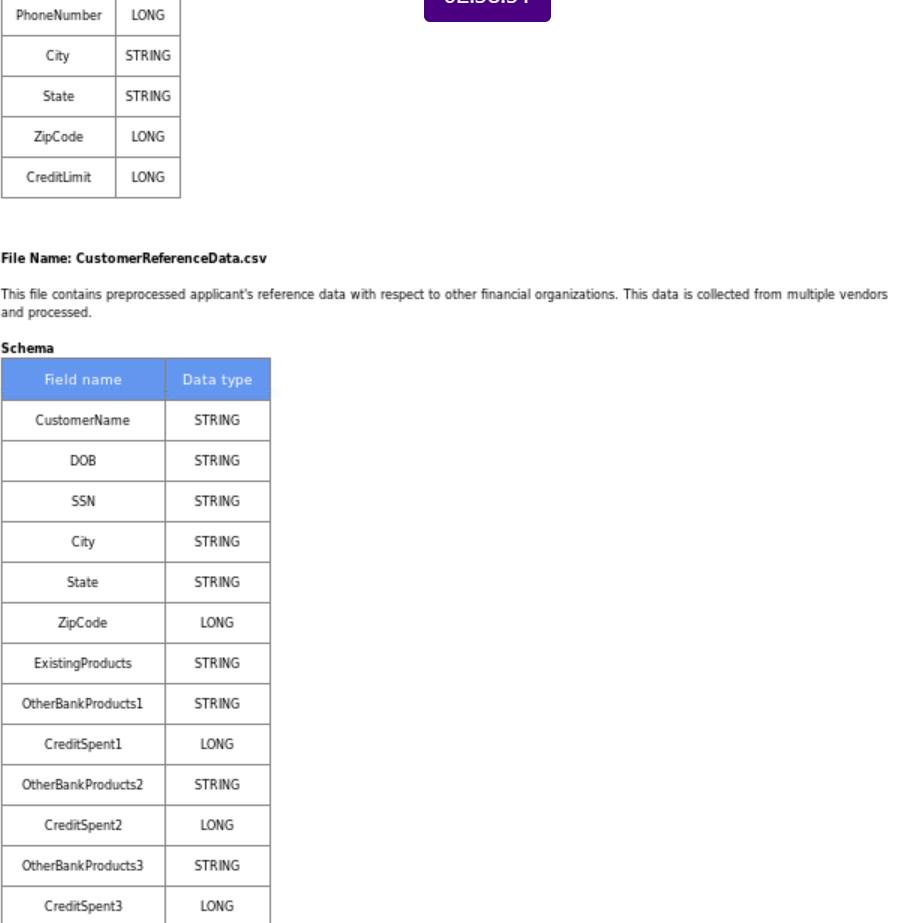
Cust

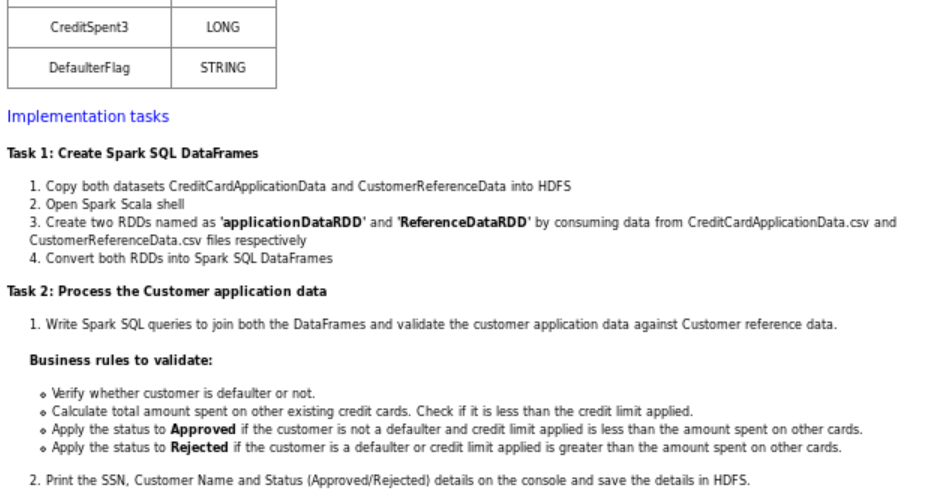




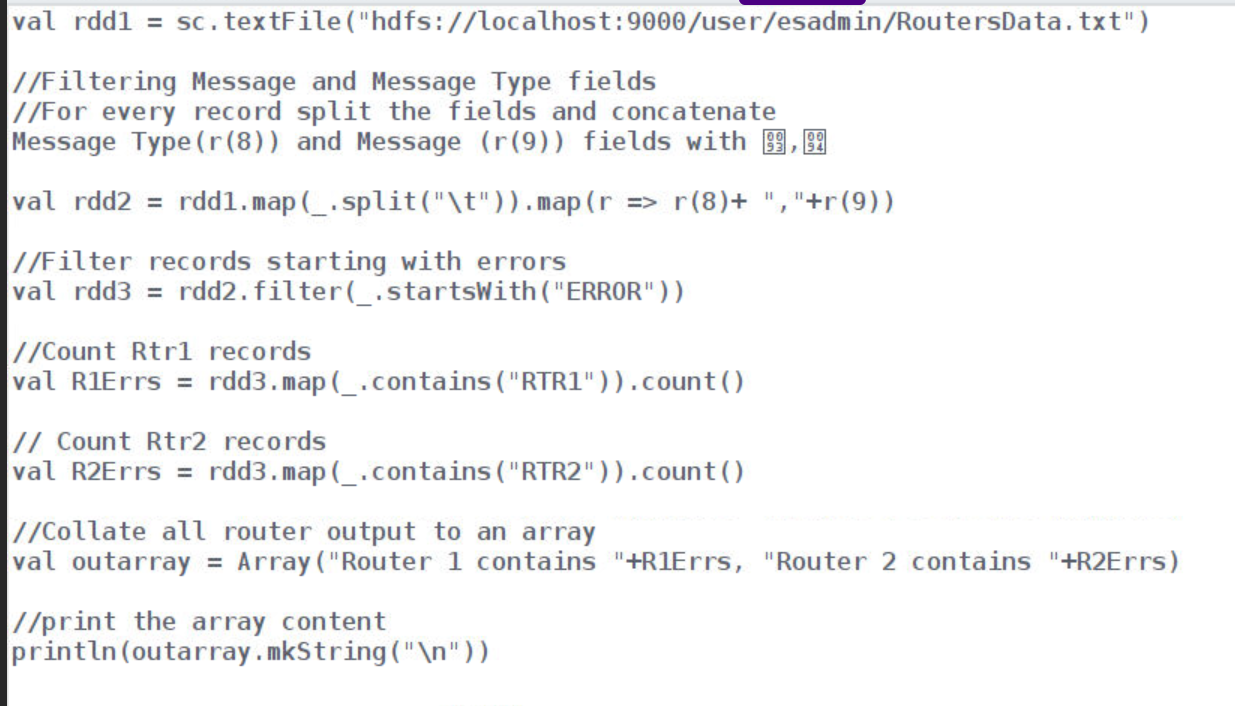




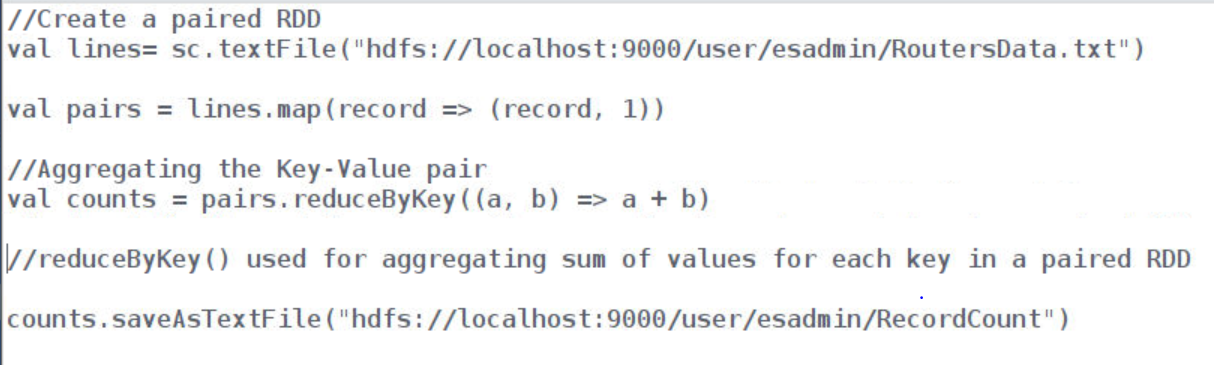




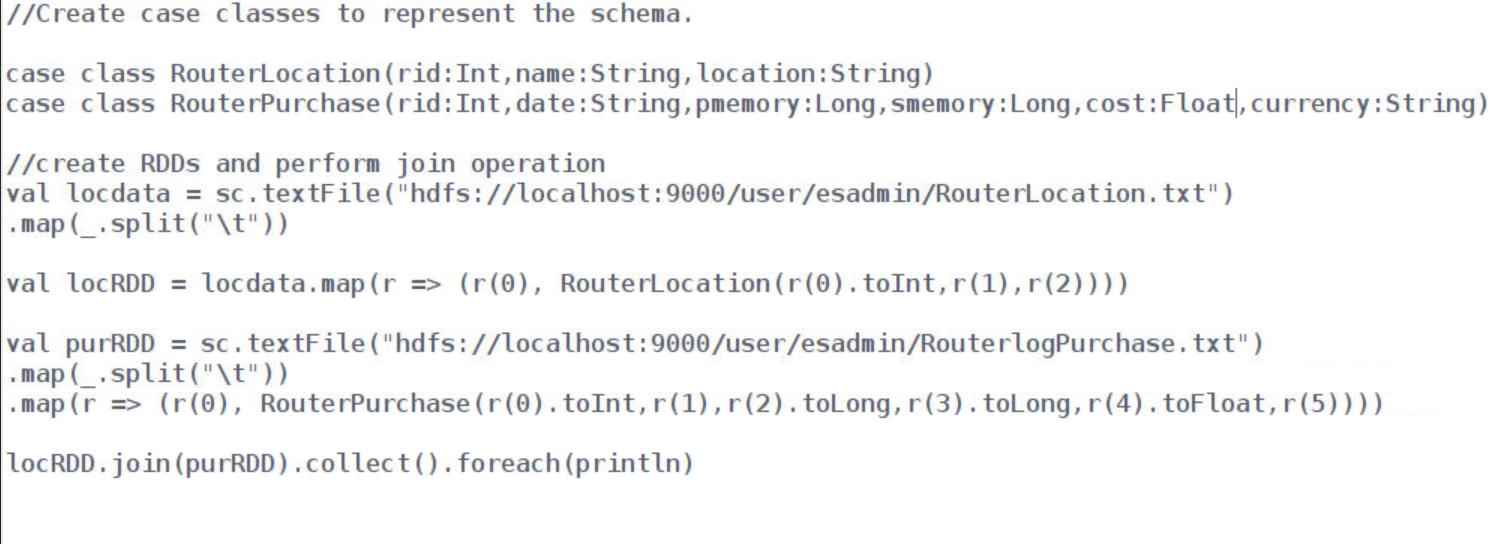
Spark Core DEMO-1



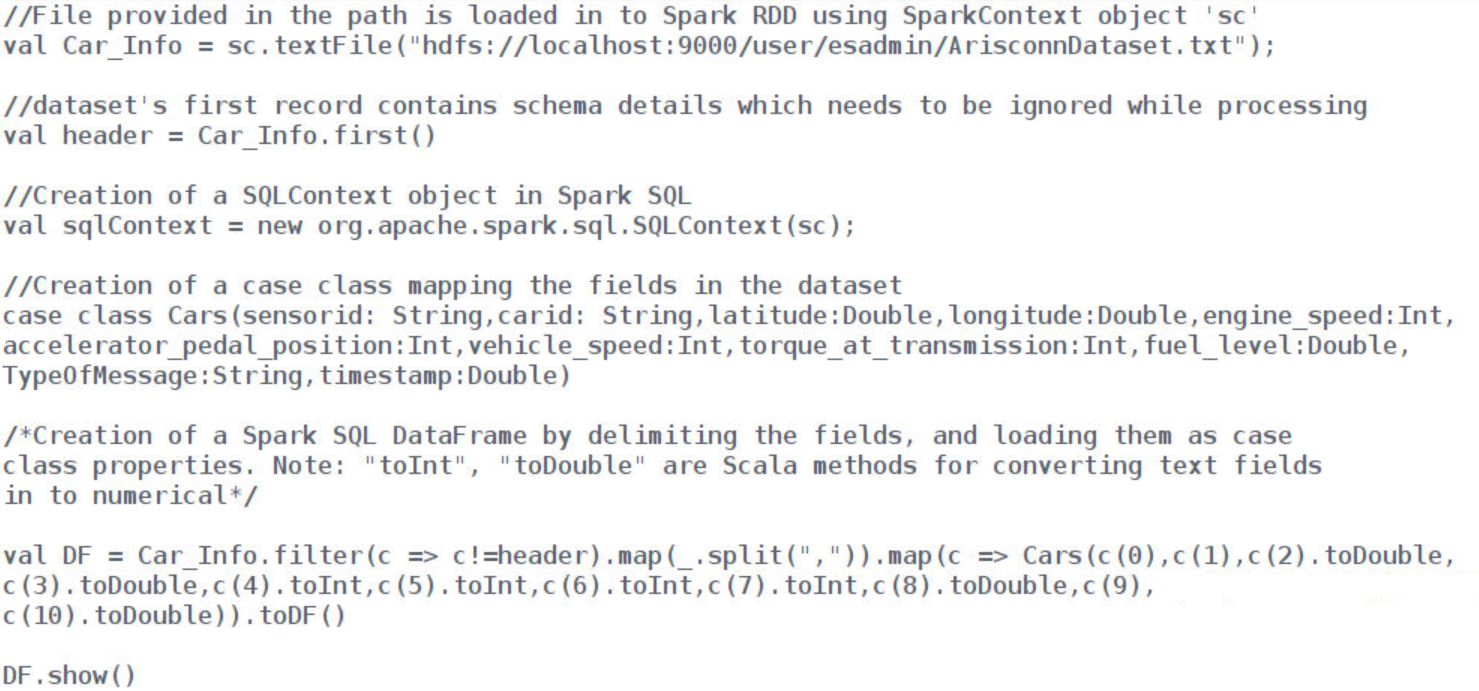
Spark Core DEMO -2



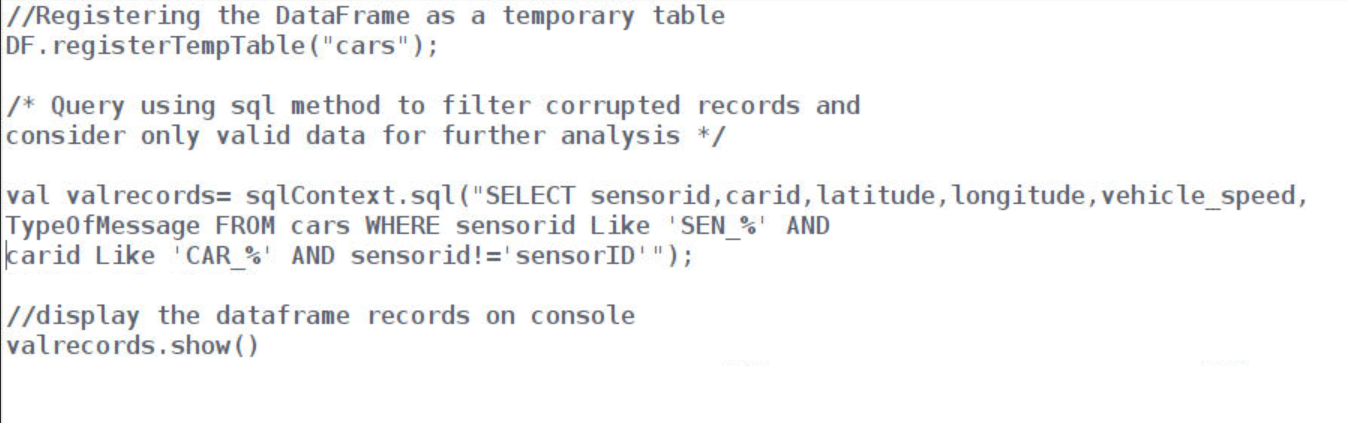
Spark Core DEMO-3



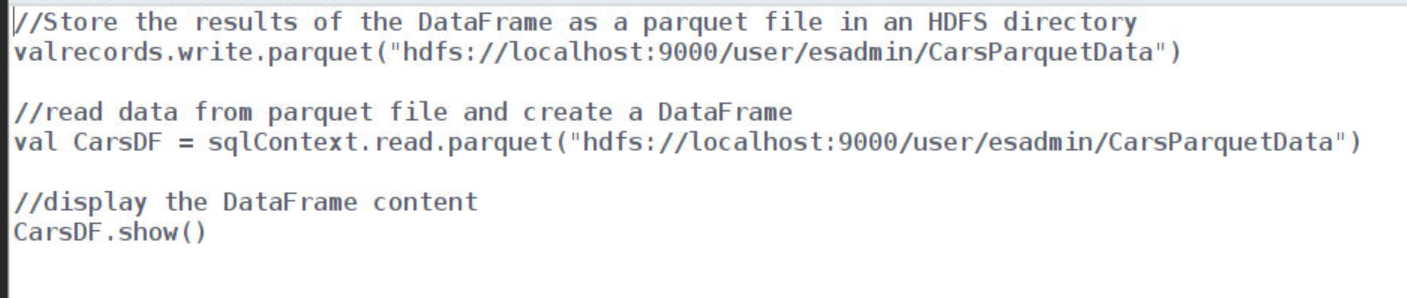
Spark-SQL DEMO -1



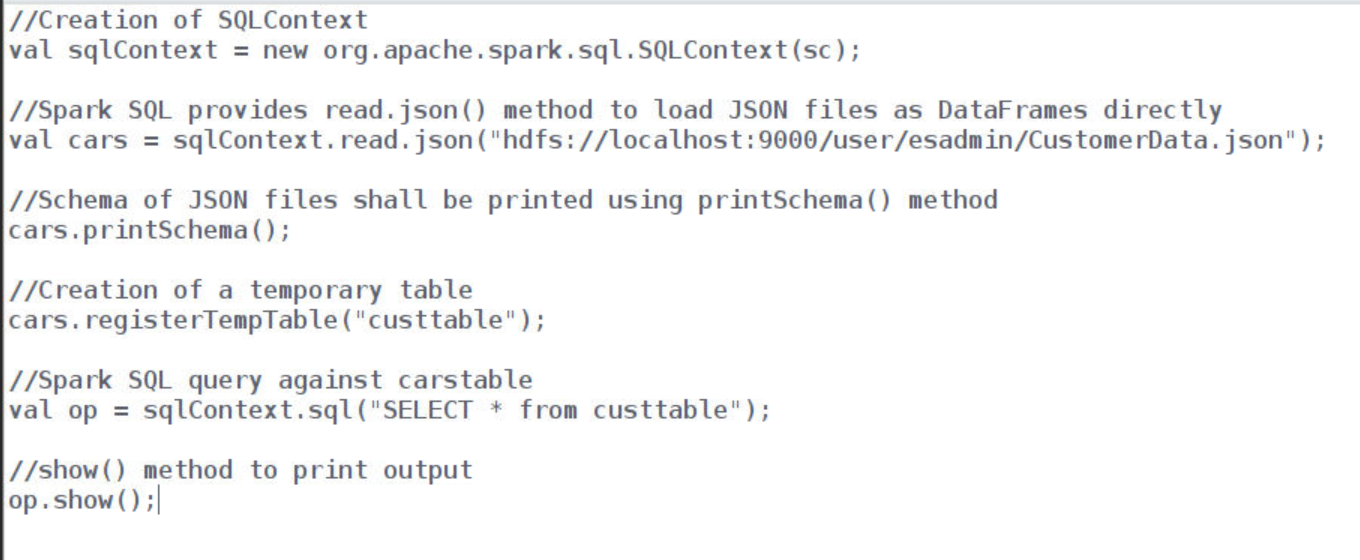
Spark-SQL DEMO -2



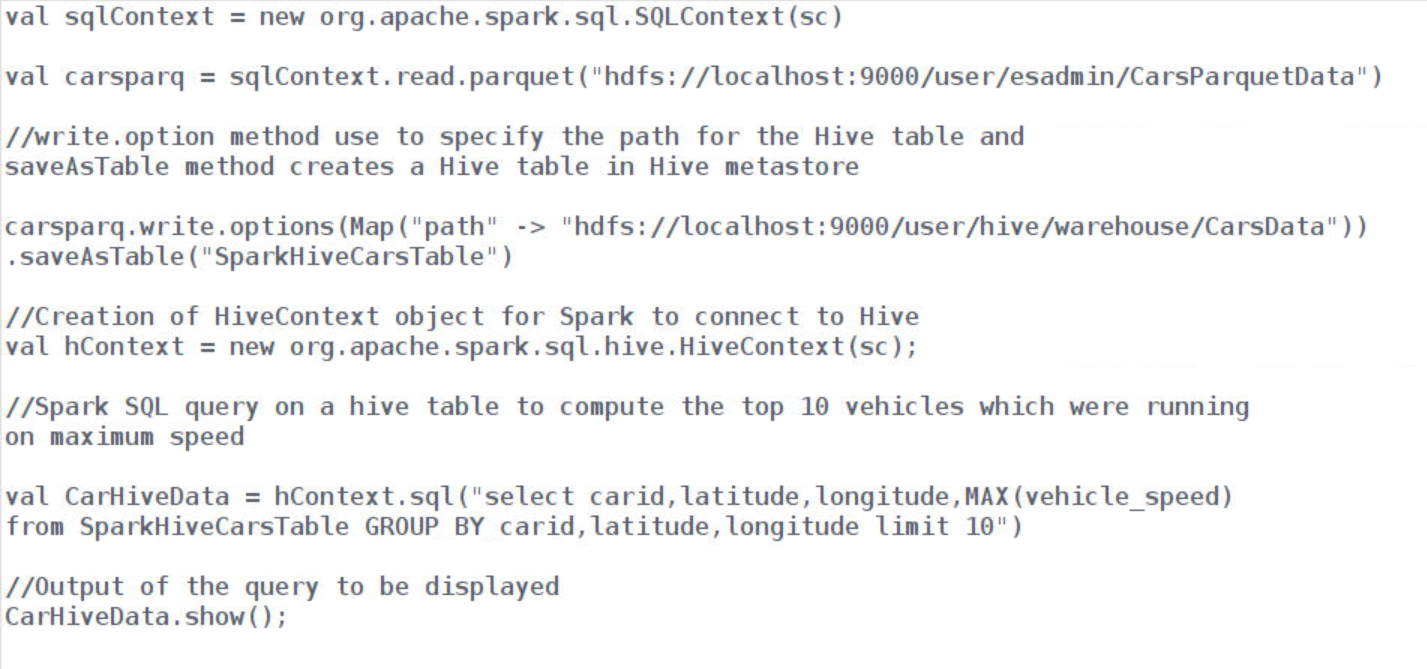
Spark-SQL DEMO -3

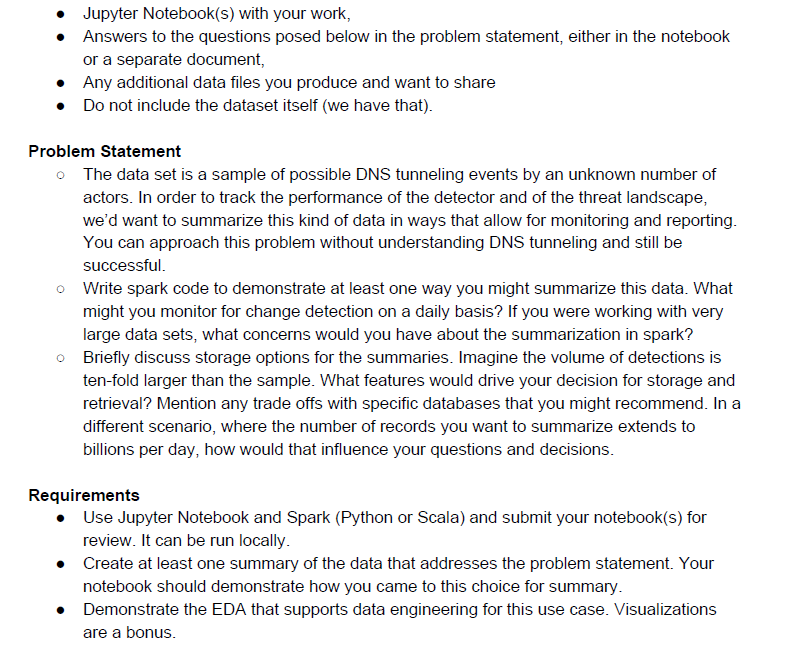


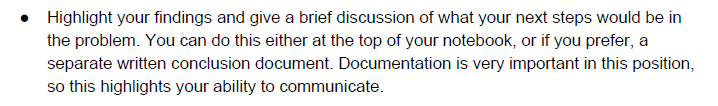
Spark-SQL DEMO -4



Spark-SQL DEMO -5







Pyspark

Partion\_by

Hive historical data and sql for reporting

