Load data from AWS RDS to Hadoop

Command to import data from AWS RDS to Hadoop

Command:
sqoop import \
--connect jdbc:mysql://#######.######.us-east-1.rds.amazonaws.com/testdatabase \
--table bookings \
--username student \
--password ######## \
--target-dir /user/hadoop/bookings_data \
-m 1

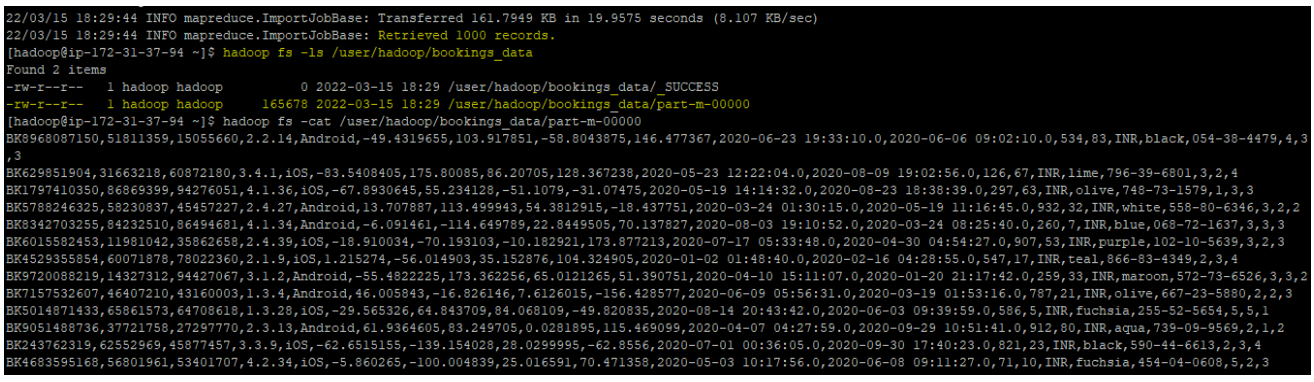
Explanation: This is the sqoop import command used to import the data from the AWS RDS to the HDFS using the RDS connection string, username, password and table name provided in the problem statement of the project. I have also provided the target directory where the data will be copied in the HDFS (/user/hadoop/bookings_data).

Command to view the imported data

Command:
hadoop fs -ls /user/hadoop/bookings_data

Explanation: This is the command used to list the imported data in the HDFS. This command will show the files present in the target directory in the HDFS (/user/hadoop/bookings_data). The command will show two files named _**SUCCESS** and **part-m-00000** in the output.

Screenshot of the data



Explanation: Above is the screenshot of the data imported from the AWS RDS to the HDFS using Sqoop import command.