Data Ingestion from the RDS to HDFS using Sqoop
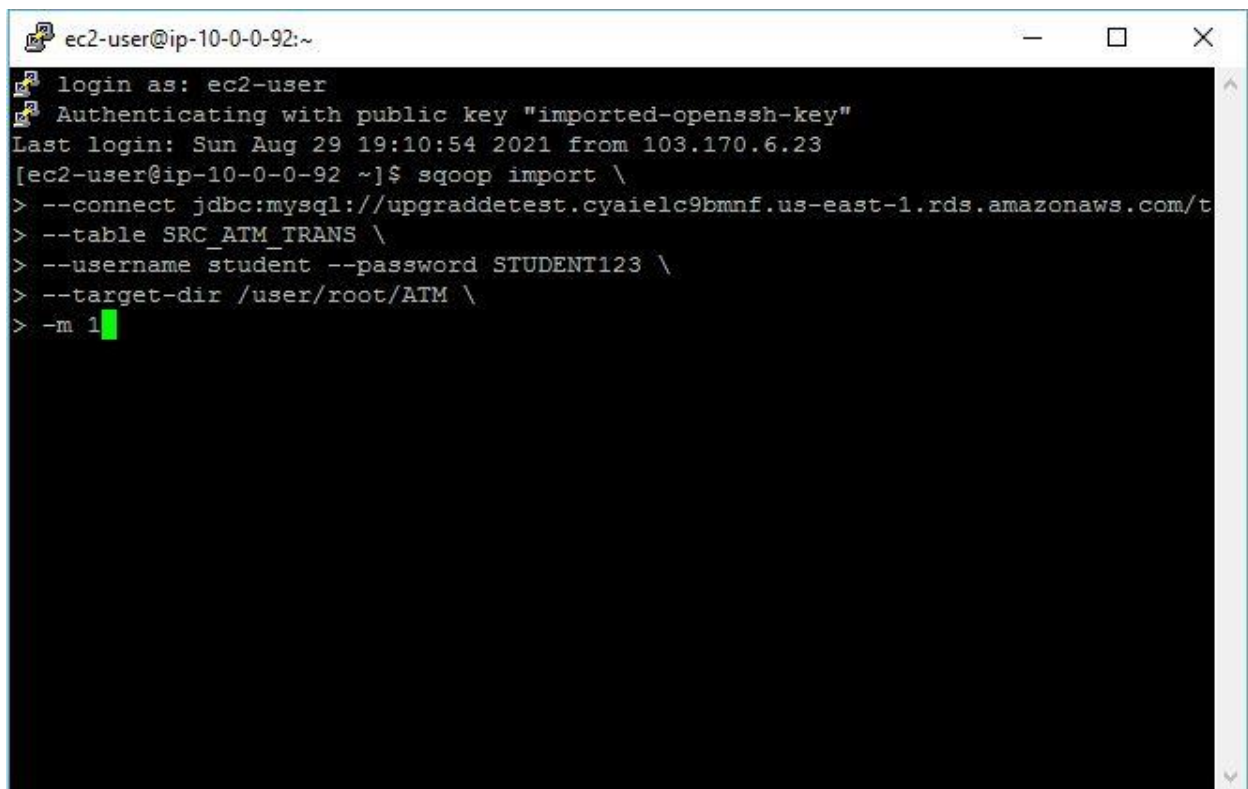
Sqoop Import command used for importing table from RDS to HDFS:

sqoop import \
--connect jdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase \
--table SRC_ATM_TRANS \
--username student --password STUDENT123 \
--target-dir /user/root/ATM \
–m 1

Command used to see the list of imported data in HDFS:

hadoop fs –ls /user/root

Screenshot of the imported data:

```
                    Input split bytes=87
                    Spilled Records=0
                    Failed Shuffles=0
                    Merged Map outputs=0
                    GC time elapsed (ms)=329
                    CPU time spent (ms)=33110
                    Physical memory (bytes) snapshot=416985088
                    Virtual memory (bytes) snapshot=2808754176
                    Total committed heap usage (bytes)=385351680
        File Input Format Counters
                    Bytes Read=0
        File Output Format Counters
                    Bytes Written=531214815
21/08/29 19:45:38 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 69.16
21/08/29 19:45:38 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-10-0-0-92 ~]# hadoop fs -ls /user/root
Found 4 items
drwx------    - root supergroup          0 2021-08-29 19:45 /user/root/.staging
drwxr-xr-x    - root supergroup          0 2021-08-29 19:45 /user/root/ATM
-rw-r--r--    3 hdfs supergroup       6971 2021-07-25 06:58 /user/root/flights_da
ta
-rw-r--r--    3 hdfs supergroup   44529893 2021-07-25 06:58 /user/root/online_dat
a
[root@ip-10-0-0-92 ~]#
```