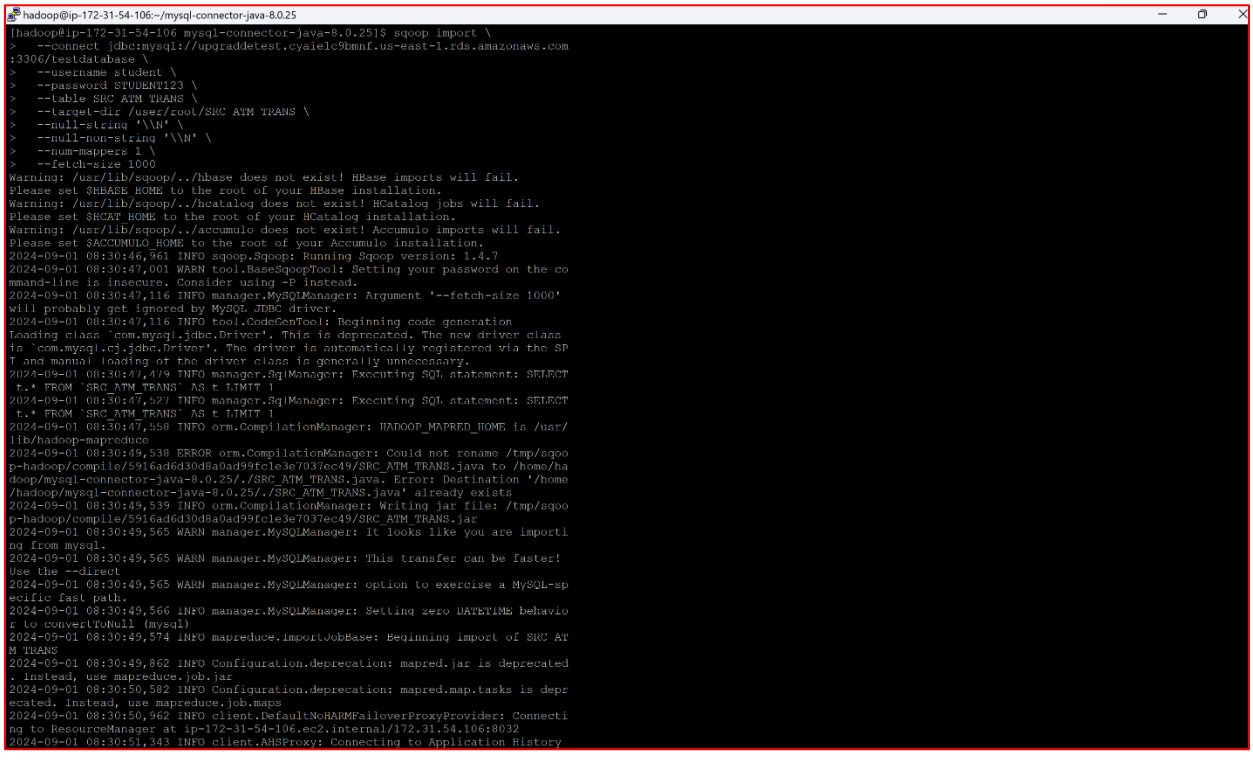
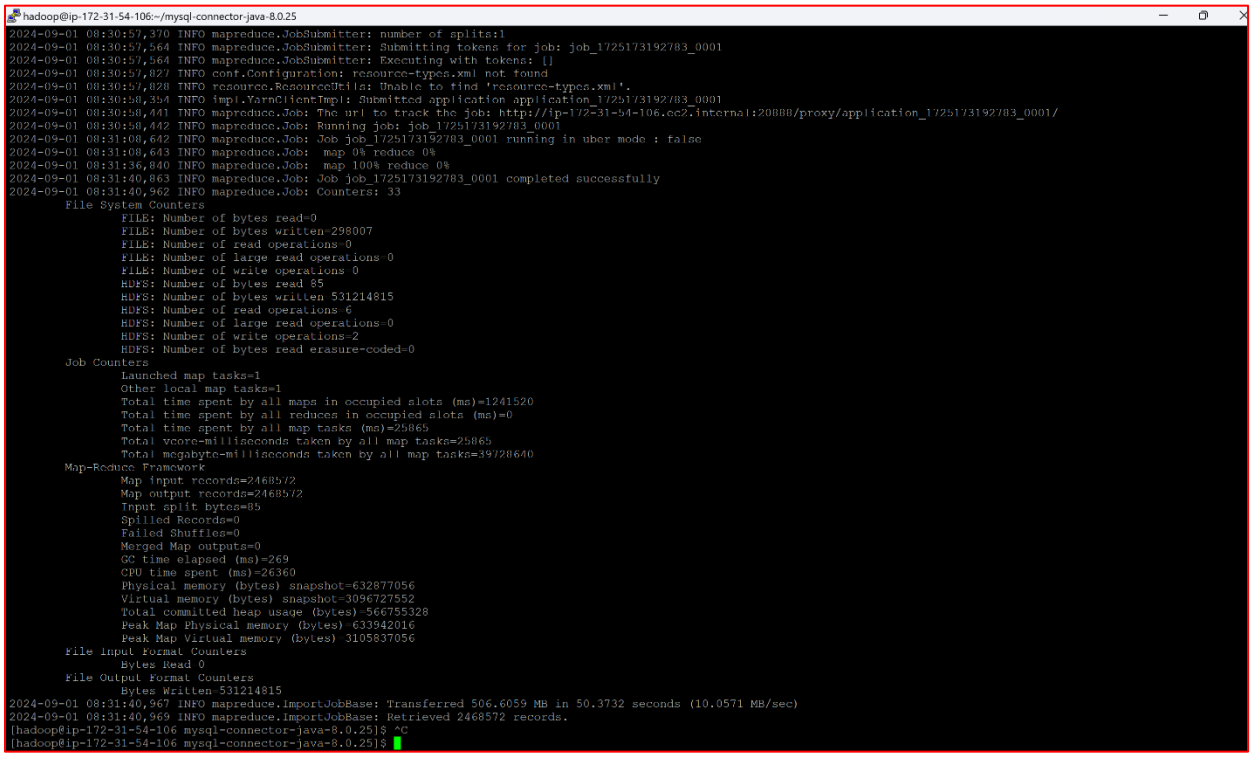
Data Ingestion from the RDS to HDFS using Sqoop

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 \
--username student \
--password STUDENT123 \
--table SRC_ATM_TRANS \
--target-dir /user/root/SRC_ATM_TRANS \
--null-string '\\N' \
--null-non-string '\\N' \
--num-mappers 1 \
--fetch-size 10000
```

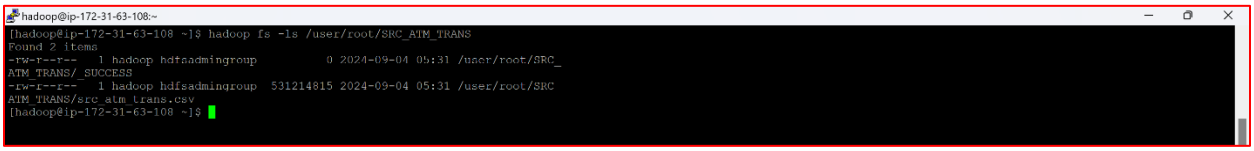




In the screenshot above we can see 2468572 rows have been retrieved

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

hadoop fs -ls /user/root/SRC_ATM_TRANS
hdfs dfs -mv /user/root/SRC_ATM_TRANS/part-m-00000
/user/root/SRC_ATM_TRANS/src_atm_trans.csv



In the screenshot above we can see two items:

- The first file is the success file, indicating that the MapReduce job was successful.

- The second file ‘part-m-00000’ since only one mapper was used in the import command, as a result the data is in a single file

<u>Command used to see the first 10 rows of the imported data</u>

hadoop fs -cat /user/root/SRC_ATM_TRANS/src_atm_trans.csv | head -1