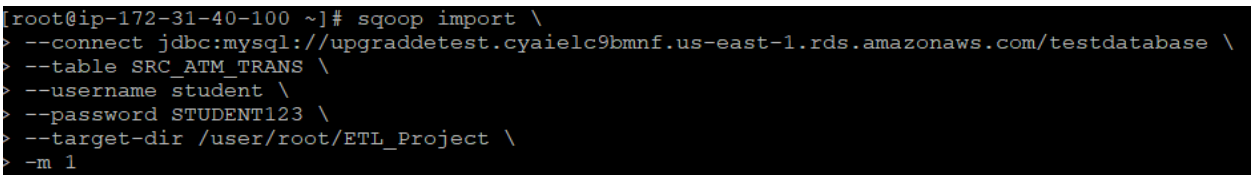
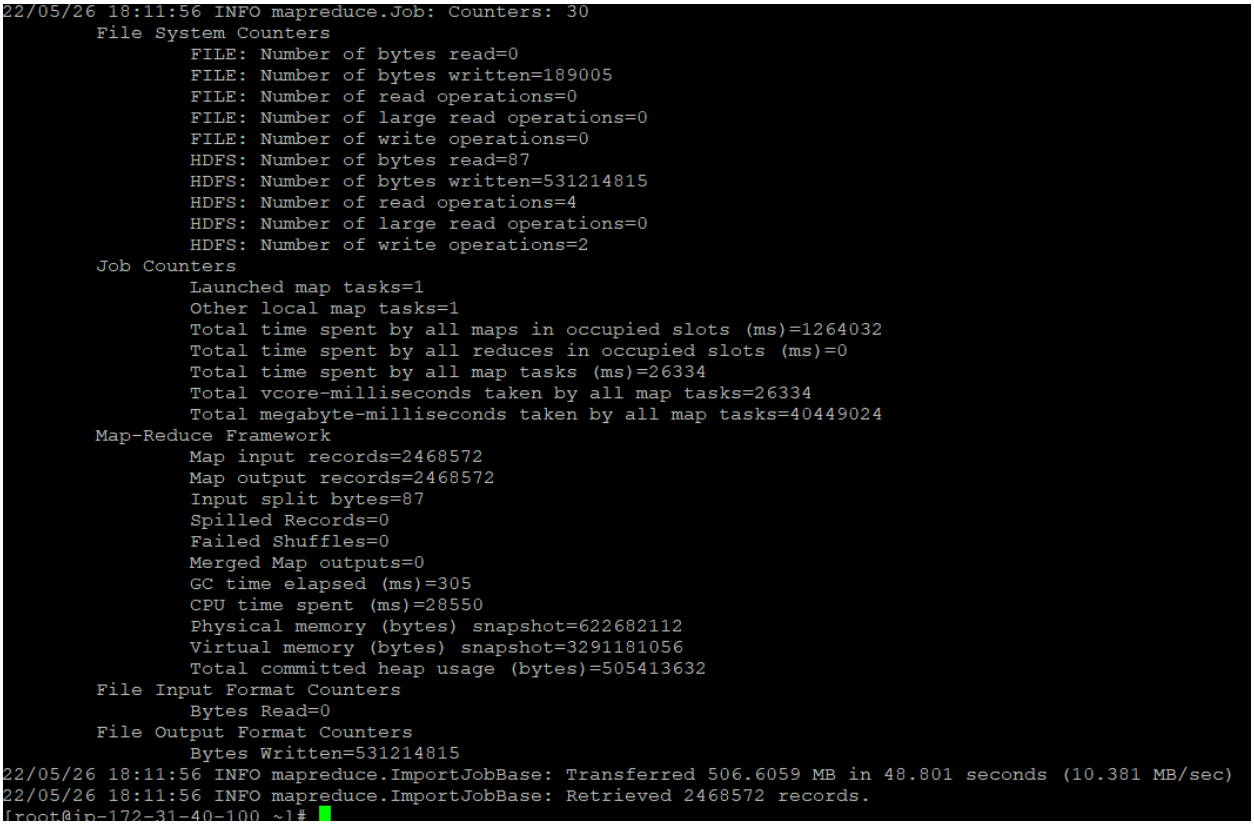
Data Ingestion from the RDS to HDFS using Sqoop

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/ETL_Project \
-m 1
```

The below screenshot shows that the data has been imported and the number of records imported is also visible





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

hadoop fs -ls /user/root/ETL_Project/

The below screenshot shows the files imported, one is success file for MapReduce job and another is the file containing data

```
[root@ip-172-31-40-100 ~]# hadoop fs -ls /user/root/ETL_Project/
Found 2 items
-rw-r--r--   1 root hadoop          0 2022-05-26 18:11 /user/root/ETL_Project/_SUCCESS
-rw-r--r--   1 root hadoop  531214815 2022-05-26 18:11 /user/root/ETL_Project/part-m-00000
[root@ip-172-31-40-100 ~]#
```

Screenshot of the imported data:

The below validation is done to check whether the record count is shown is valid
Command Used: hadoop fs – cat /user/root/ETL_Project/part-m-00000 | wc -l

```
[root@ip-172-31-40-100 ~]# hadoop fs -cat /user/root/ETL_Project/part-m-00000 | wc -l
2468572
[root@ip-172-31-40-100 ~]#
```

When the file is opened using cat command data and head condition data is present as below

Command Used: hadoop fs – cat /user/root/ETL_Project/part-m-00000 | head