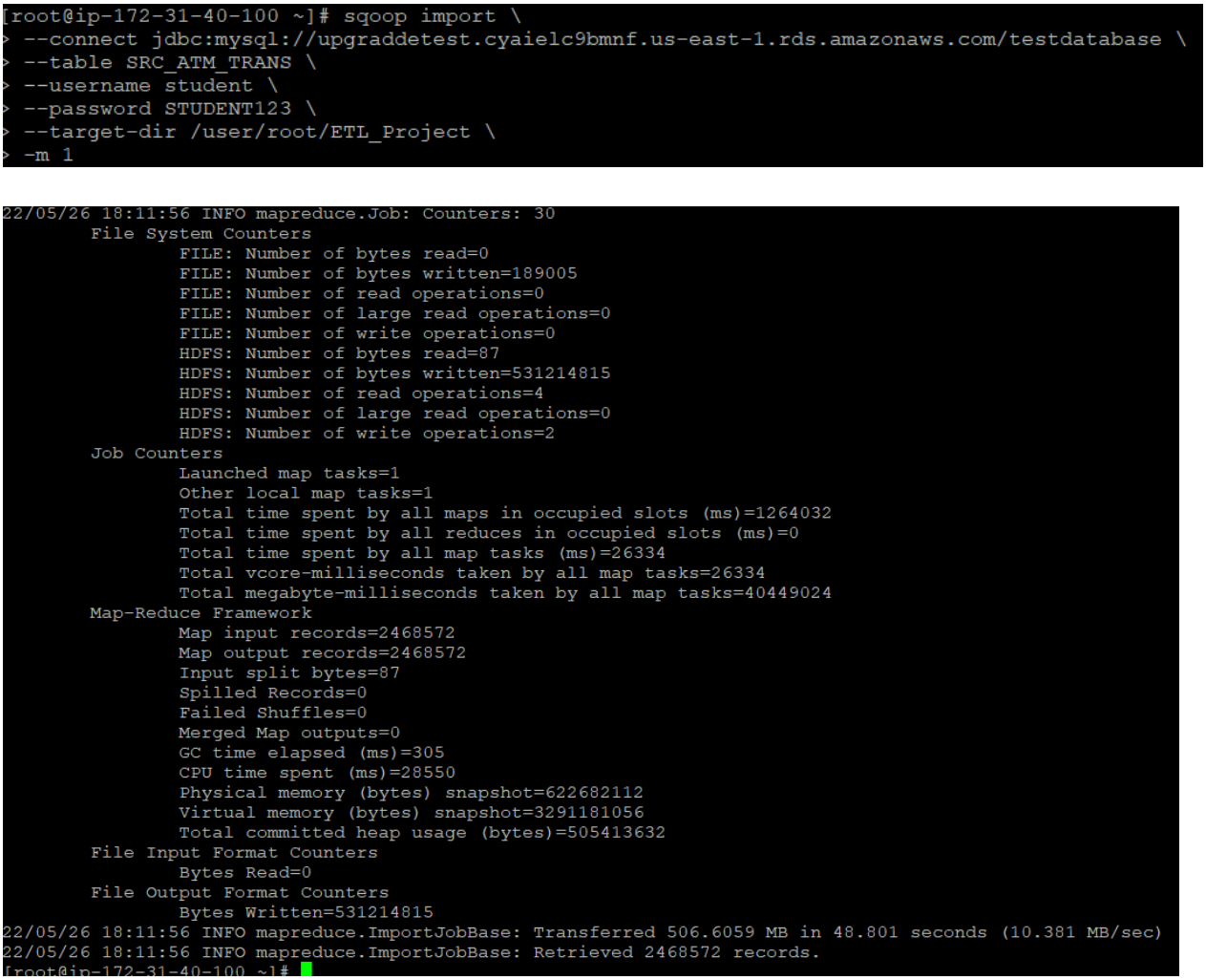
Data Ingestion from the RDS to HDFS using Sqoop

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

```
                Physical memory (bytes) snapshot=622194688
                Virtual memory (bytes) snapshot=3286917120
                Total committed heap usage (bytes)=539492352
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=531214815
22/05/29 11:57:38 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 46.6339 seconds (10.8635 MB/sec)
22/05/29 11:57:38 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-172-31-32-22 ~]# hadoop fs -ls
Found 1 items
drwxr-xr-x   - root hadoop          0 2022-05-29 11:57 ETL_Project
[root@ip-172-31-32-22 ~]# hadoop fs -ls /user/root/ETL_Project/
Found 2 items
-rw-r--r--   1 root hadoop          0 2022-05-29 11:57 /user/root/ETL_Project/_SUCCESS
-rw-r--r--   1 root hadoop  531214815 2022-05-29 11:57 /user/root/ETL_Project/part-m-00000
[root@ip-172-31-32-22 ~]# hadoop fs -cat /user/root/ETL_Project/part-m-00000 | head
2017,January,1,Sunday,0,Active,1,NCR,NÃƒÂ¦stved,Farimagsvej,8,4700,55.233,11.763,DKK,MasterCard,5643,Withdrawal,,,55.230,11.761,26160
8,Naestved,281.150,1014,87,7,260,0.215,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,2,NCR,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,MasterCard,1764,Withdrawal,,,57.048,9.935,2616235
,NÃƒÂ,rresundby,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,2,NCR,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,VISA,1891,Withdrawal,,,57.048,9.935,2616235,NÃƒÂ
resundby,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,3,NCR,Ikast,RÃƒÂ¥dhusstrÃƒÂ¦det,12,7430,56.139,9.154,DKK,VISA,4166,Withdrawal,,,56.139,9.158,261942
Ikast,281.150,1011,100,6,240,0.000,75,300,Drizzle,light intensity drizzle
2017,January,1,Sunday,0,Active,4,NCR,Svogerslev,BrÃƒÂ¸nsager,1,4000,55.634,12.018,DKK,MasterCard,5153,Withdrawal,,,55.642,12.080,261
81,Roskilde,280.610,1014,87,7,260,0.000,88,701,Mist,mist
2017,January,1,Sunday,0,Active,5,NCR,Nibe,Torvet,1,9240,56.983,9.639,DKK,MasterCard,3269,Withdrawal,,,56.981,9.639,2616483,Nibe,280.
0,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,6,NCR,Fredericia,SjÃƒÂ¦llandsgade,33,7000,55.564,9.757,DKK,MasterCard,887,Withdrawal,,,55.566,9.753,2
1951,Fredericia,281.150,1014,93,7,230,0.290,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,7,Diebold Nixdorf,Hjallerup,Hjallerup Centret,18,9320,57.168,10.148,DKK,Mastercard - on-us,4626,Withd
wal,,,57.165,10.146,2620275,Hjallerup,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,8,NCR,GlyngÃƒÅ¸re,FÃƒÂ¦rgevej,1,7870,56.762,8.867,DKK,MasterCard,470,Withdrawal,,,56.793,8.853,2615964
Nykobing Mors,281.150,1011,100,6,240,0.000,75,300,Drizzle,light intensity drizzle
2017,January,1,Sunday,0,Active,9,Diebold Nixdorf,Hadsund,Storegade,12,9560,56.716,10.114,DKK,VISA,8473,Withdrawal,,,56.715,10.117,262
952,Hadsund,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
cat: Unable to write to output stream.
[root@ip-172-31-32-22 ~]#
```