### Data Ingestion from the RDS to HDFS using Sqoop

wget hppt://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz

tar -xvf mysql-connector-java-8.0.25.tar.gz

cd mysql-connector-java-8.0.25 /

sudo cp mysql-connector-java-8.0.25.jar /usr/lib/sqoop/lib

**Sqoop Import command used for importing table from RDS to HDFS:**

sqoop import \

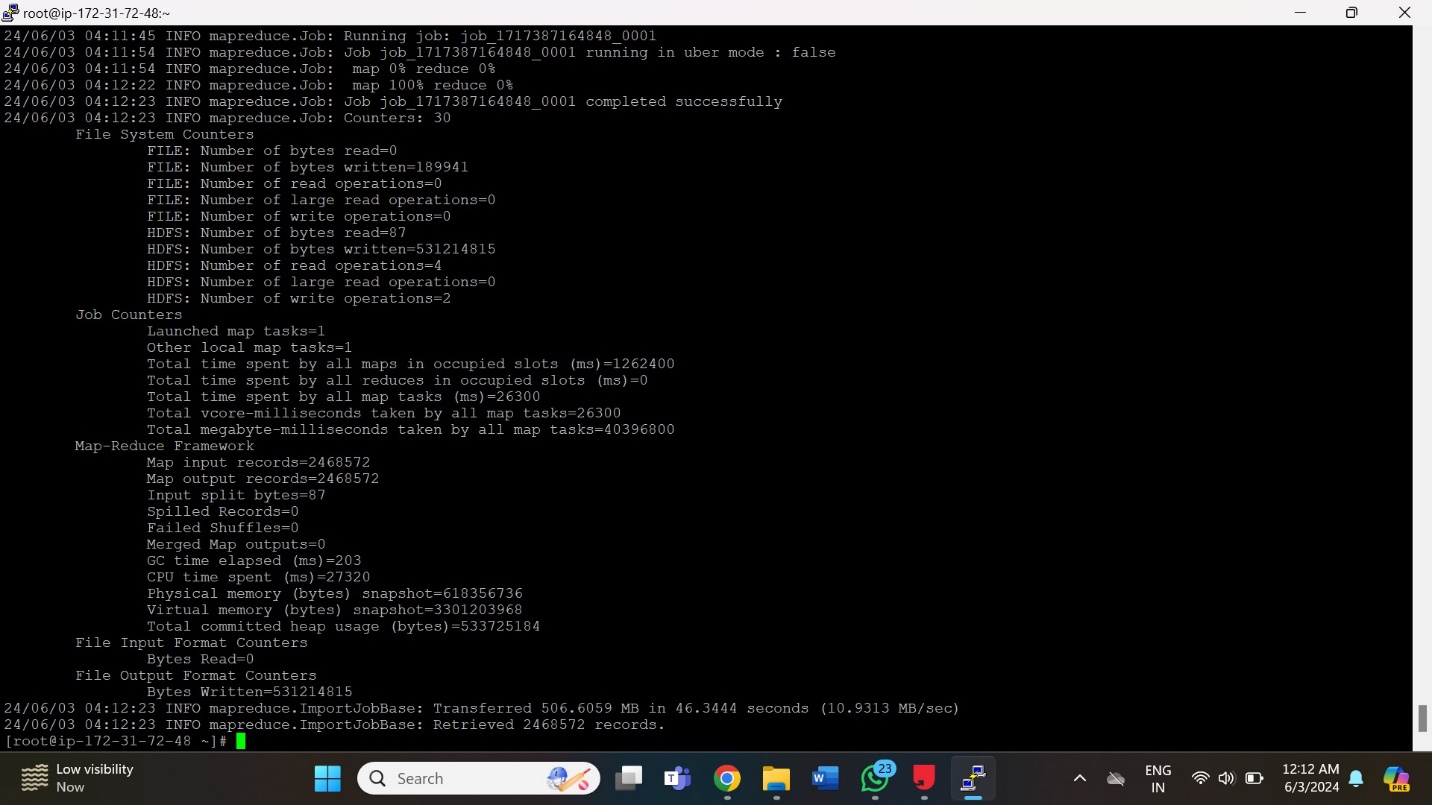
> --connect jdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase \

> --table SRC\_ATM\_TRANS \

> --username student --password STUDENT123 \

> --target-dir /user/root/etl\_spar\_nord\_atm \

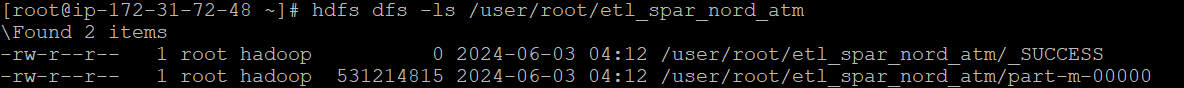
> --m 1



We can see in this screenshot that 506.6059 MB of data was imported. Sqoop command retrieved 2468572 records in total.

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

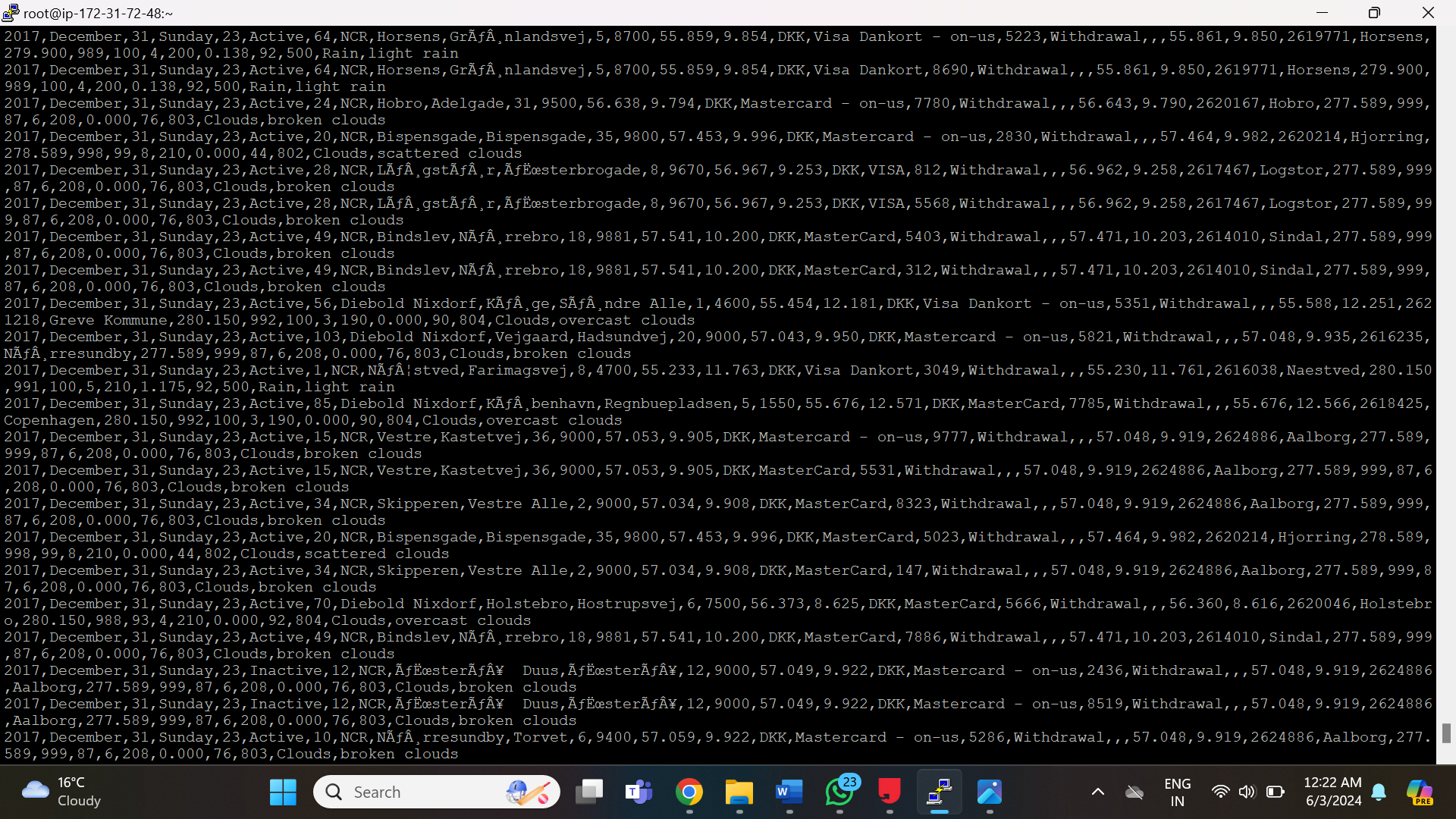
hdfs dfs -ls /user/root/etl\_spar\_nord\_atm



Since, I used only one mapper, data is imported in only one file which is **“part-m-0000”.**

**Screenshot of the imported data:**

1. hdfs dfs -cat /user/root/etl\_spar\_nord\_atm/part-m-0000



Screenshot of the portion of the data read from **“part-m-0000”.**

1. hdfs dfs -cat /user/root/etl\_spar\_nord\_atm/part-m-00000 | wc -l

