Data Ingestion from the RDS to HDFS using Sqoop

We Need MySQL connector jar file in our Sqoop lib directory before running Sqoop jobs

- Switch to root user
  sudo -i

- Download the tar file on to your cluster
  wget https://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz

- Extract the file
  tar -xvf mysql-connector-java-8.0.25.tar.gz

- Copy the file and move it to /usr/lib/sqoop/lib/
  sudo cp mysql-connector-java-8.0.25/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/spar_bank_data/ \
--m 1
```

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

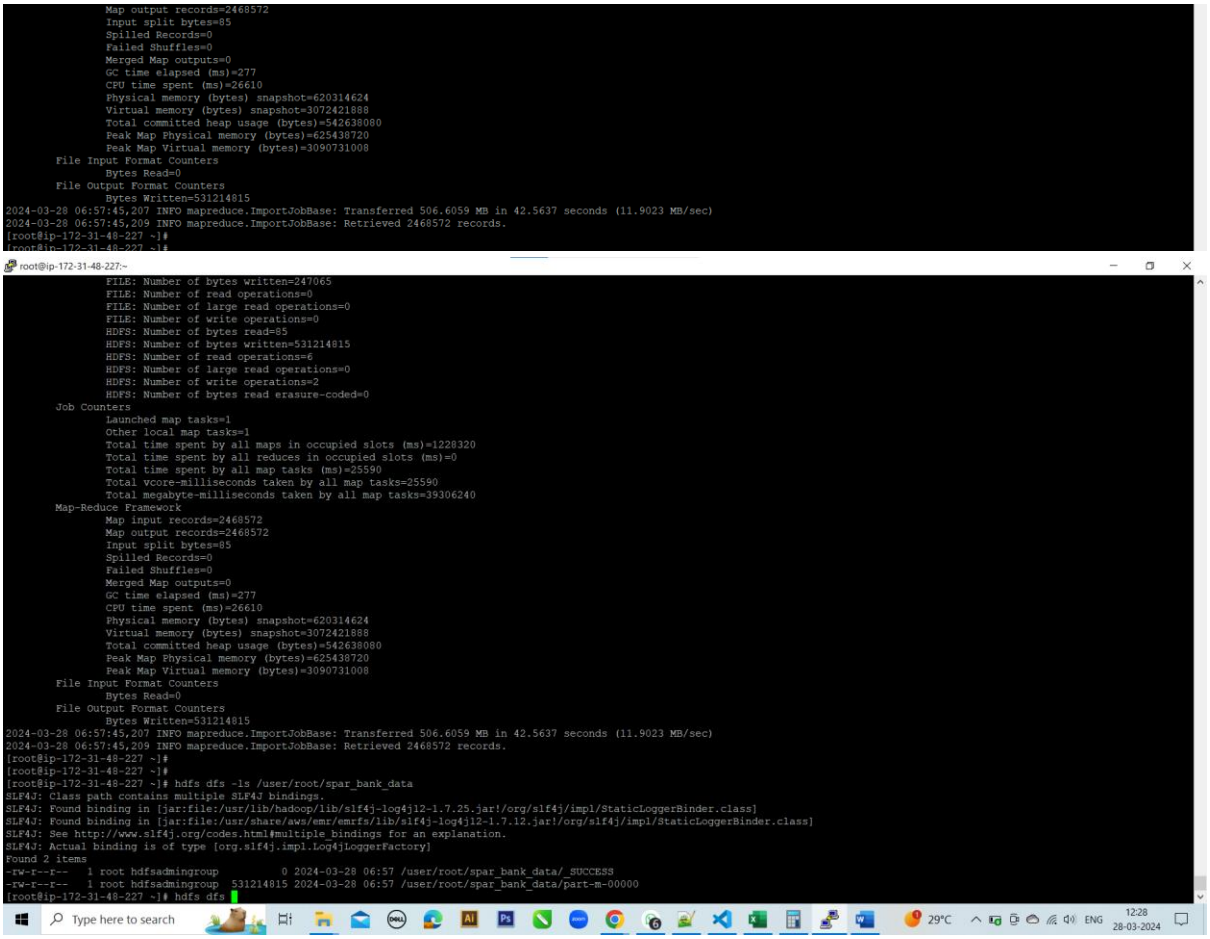
- To check the files present in destination directory.
  hdfs dfs -ls /user/root/spar_bank_data

- To check the first three records
  hdfs dfs -cat /user/root/spar_bank_data/part-m-00000 | head -n 3

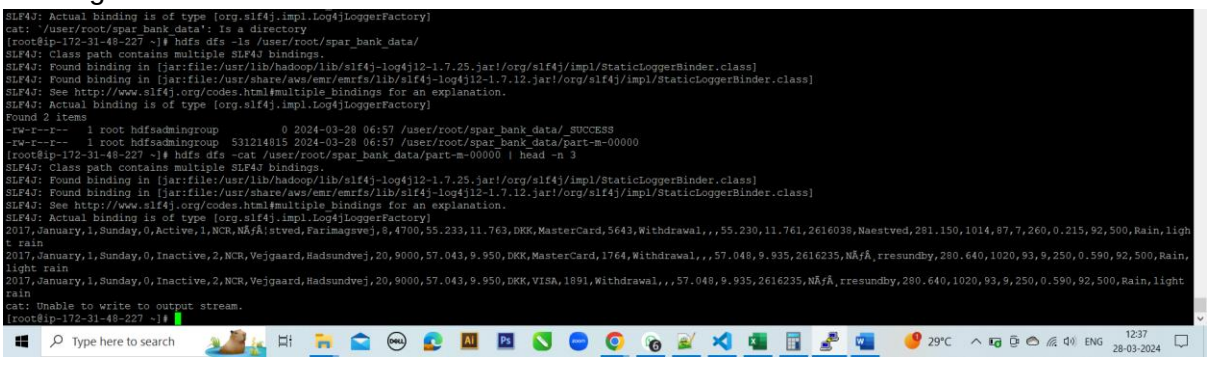
- To check the total count of the records
  hdfs dfs -cat /user/root/spar_bank_data/part-m-00000 | wc -l

Screenshot of the imported data:

Successful loading of data with 2 files present in the destination. Success file and the part file.



Viewing the first 3 records in the file



Checking the total record count in the part file.