SPARNOD BANK ETL PROJECT

Sqoop Data Ingestion

Submitted By:

Prerna Gupta (prerna.gupta1792@gmail.com)

Nikhil Mahakal (nikhilmahakal07@gmail.com)

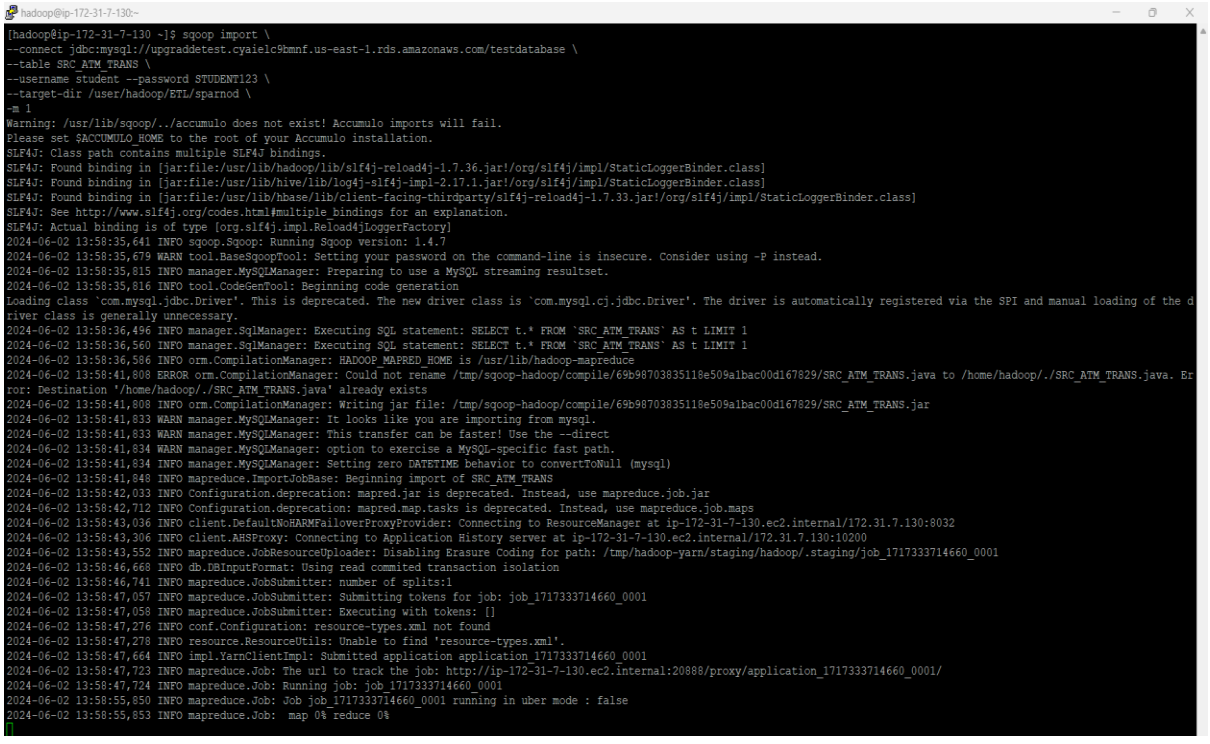
Pradeep Singh Negi (pradeep.negidba@gmail.com)

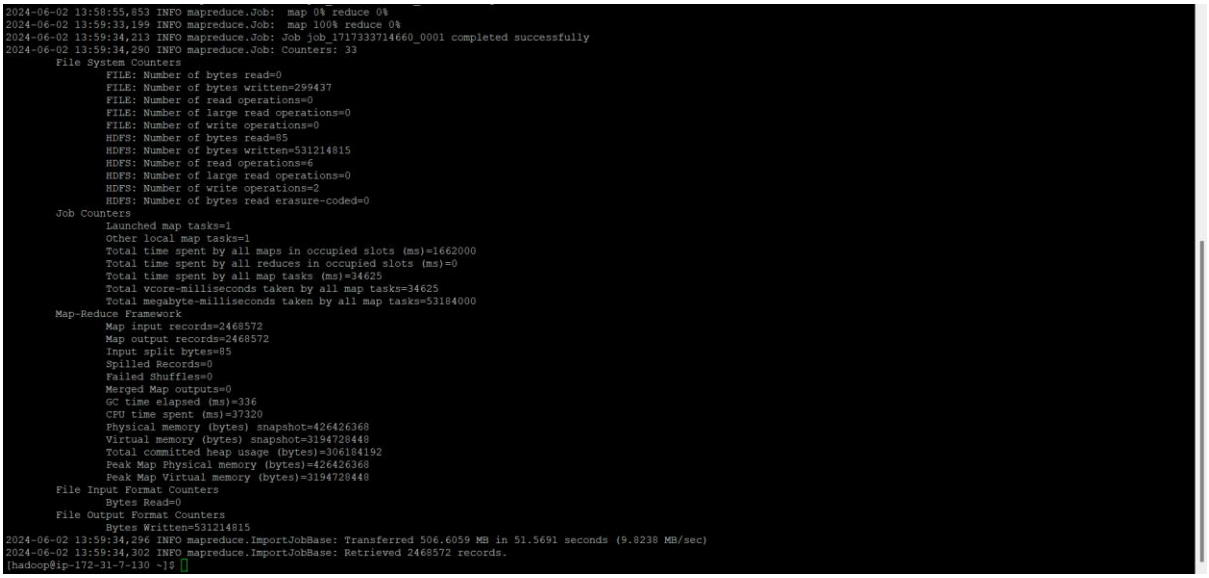
Data Ingestion from the RDS to HDFS using Sqoop

1. Installed mysql connector on the cluster using below command to run sqoop.
   sudo -i
   wget https://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/

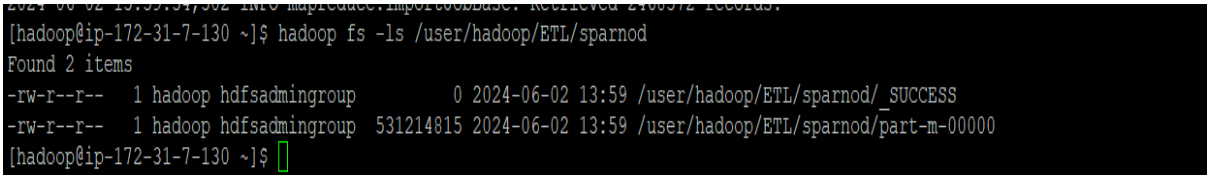
2. Run the below sqoop command to ingest data from SRC_ATM_TRANS table in testdatabase RDS to HDFS

   **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/hadoop/ETL/sparnod **
   -m 1





3. Run below command to check the files on HDFS after completion of import
hadoop fs -ls /user/hadoop/ETL/sparnod



4. Run below command to check the number of records exported from RDS to HDFS file.
hadoop fs -cat /user/hadoop/ETL/sparnod/part-m-00000| wc -l

Above screenshot shows **2468572** rows imported.

5. Screenshot of imported data.