Data Ingestion from the RDS to HDFS using Sqoop

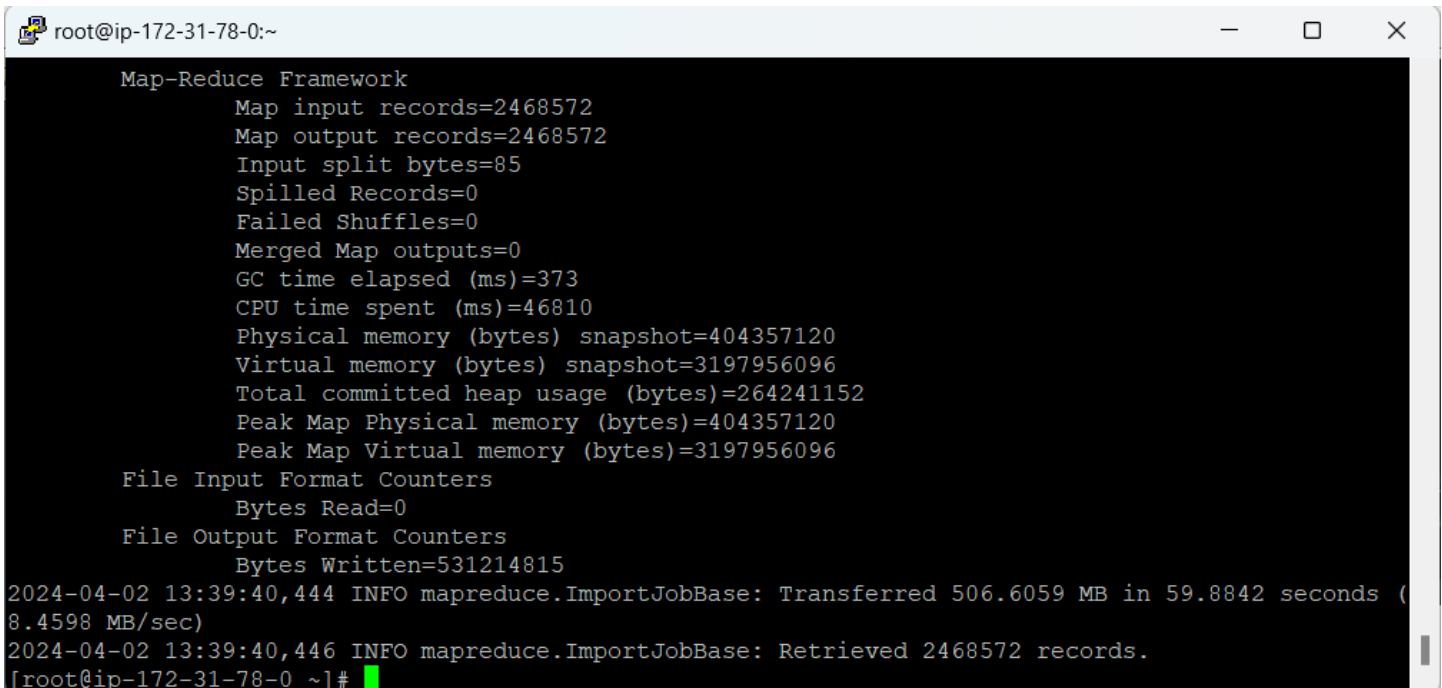**Prerequisite for executing the sqoop command and establishing a connection to RDS:**

While logged in as the root user, access the EMR instance and fetch the MySQL connector.

    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/

    sudo -i

1. **Sqoop Import command used for importing table from RDS to HDFS:**
   Command:
   sqoop import --connect jdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase --username student --password STUDENT123 --table SRC_ATM_TRANS --target-dir /user/hadoop/ETL/ATMDATA -m 1
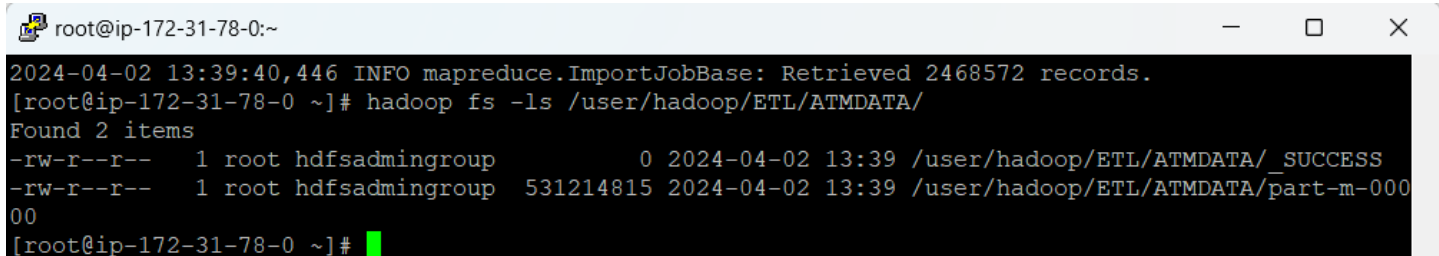

In the screenshot provided, it's evident that 2468572 records have been retrieved.

2. **Command used to see the list of imported data in HDFS:**
   Command:
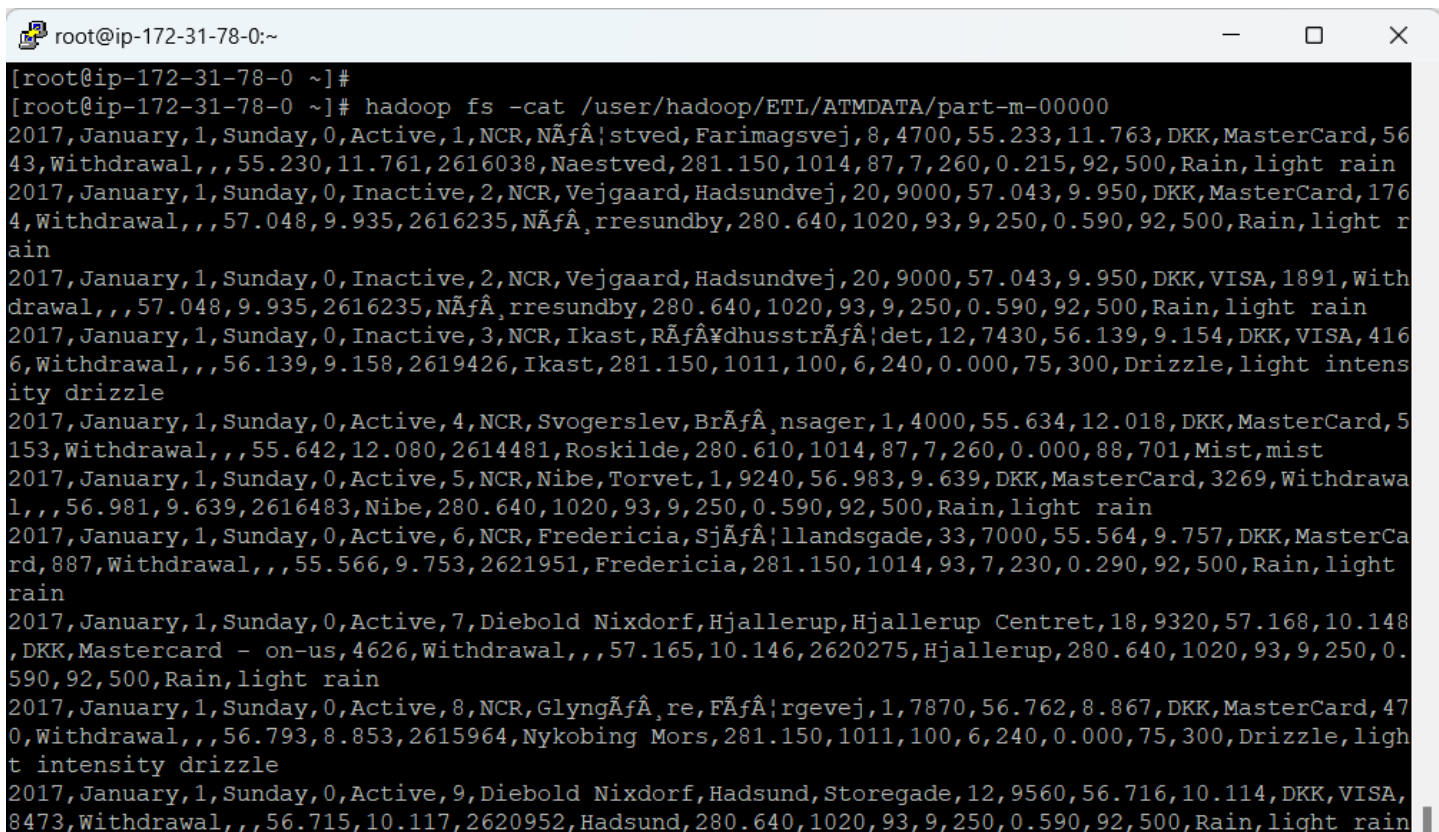   hadoop fs -ls /user/hadoop/ETL/ATMDATA

In the screenshot above, two items are visible:
- The first item is the success file, signaling the successful completion of the MapReduce job.
- The second item, labeled 'part-m-00000', is the file imported.

3. **Screenshot of the imported data:**
   Command:
   hadoop fs -cat /user/hadoop/ETL/ATMDATA/part-m-00000

The screenshot displays a segment of data extracted from the 'part-m-00000' file.