Data Ingestion from the RDS to HDFS using Sqoop

Have mysql connector jar in place as a prerequisite for running 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

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:3306/testdatabase --username student --password STUDENT123 --table SRC_ATM_TRANS --target-dir /user/hadoop/ETL/ATMDATA -m 1

Command to see the list of imported data in HDFS:

hadoop fs -ls /user/hadoop/ETL/ATMDATA/

count of data records:

hadoop fs -cat /user/hadoop/ETL/ATMDATA/part-m-00000 | wc -l

top 3 records:

hadoop fs -cat /user/hadoop/ETL/ATMDATA/part-m-00000 | head -n 3

Screenshot of the imported data: