Data Ingestion from the RDS to HDFS using Sqoop

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_nord_bank_atm \
-m 1
```

```
22/04/30 07:42:54 INFO mapreduce.Job:  map 0% reduce 0%
22/04/30 07:43:22 INFO mapreduce.Job:  map 100% reduce 0%
22/04/30 07:43:23 INFO mapreduce.Job: Job job_1651302571224_0001 completed successfully
22/04/30 07:43:23 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=188956
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=87
                HDFS: Number of bytes written=531214815
                HDFS: Number of read operations=4
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Other local map tasks=1
                Total time spent by all maps in occupied slots (ms)=1234272
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=25714
                Total vcore-milliseconds taken by all map tasks=25714
                Total megabyte-milliseconds taken by all map tasks=39496704
        Map-Reduce Framework
                Map input records=2468572
                Map output records=2468572
                Input split bytes=87
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=251
                CPU time spent (ms)=28150
                Physical memory (bytes) snapshot=629555200
                Virtual memory (bytes) snapshot=3317964800
                Total committed heap usage (bytes)=542638080
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=531214815
22/04/30 07:43:23 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 49.8925 seconds (10.1539 MB/sec)
22/04/30 07:43:23 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-172-31-94-32 ~]#
```

In the screenshot above we can see 2468572 rows have been retrieved.

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

hadoop fs -ls /user/root/spar_nord_bank_atm

In the screenshot above we can see two items:
- The first file is the success file, indicating that the MapReduce job was successful.
- The second file 'part-m-00000' is the one that I imported. Since I used only one mapper in my import command, thus the data is in a single file.

Screenshot of the imported data:

A portion of data read from part-m-00000 file: