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/SRC_ATM_TRANS \
-m 1
```

```
[root@ip-172-31-34-110 ~]# 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/SRC_ATM_TRANS \
> -m 1
```

```
22/09/02 07:01:09 INFO mapreduce.Job: Job job_1662099862612_0001 running in uber mode : false
22/09/02 07:01:09 INFO mapreduce.Job:  map 0% reduce 0%
22/09/02 07:01:37 INFO mapreduce.Job:  map 100% reduce 0%
22/09/02 07:01:38 INFO mapreduce.Job: Job job_1662099862612_0001 completed successfully
22/09/02 07:01:38 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=189411
                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)=1207776
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=25162
                Total vcore-milliseconds taken by all map tasks=25162
                Total megabyte-milliseconds taken by all map tasks=38648832
        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)=208
                CPU time spent (ms)=28390
                Physical memory (bytes) snapshot=631590912
                Virtual memory (bytes) snapshot=3291594752
                Total committed heap usage (bytes)=535298048
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=531214815
22/09/02 07:01:38 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 45.3657 seconds (11.1672 MB/sec)
22/09/02 07:01:38 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
```

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

hadoop fs -ls /user/root/SRC_ATM_TRANS

Screenshot of the imported data: