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 --username student --password STUDENT123 --table SRC_ATM_TRANS --split-by day --direct --warehouse-dir /user/root/testdatabase --compress --delete-target-dir

```
22/09/01 05:05:44 INFO mapreduce.Job: Job job_1662008401963_0001 running in uber mode : false
22/09/01 05:05:44 INFO mapreduce.Job:  map 0% reduce 0%
22/09/01 05:06:06 INFO mapreduce.Job:  map 100% reduce 0%
22/09/01 05:06:32 INFO mapreduce.Job: Job job_1662008401963_0001 completed successfully
22/09/01 05:06:33 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=760696
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=406
                HDFS: Number of bytes written=39077529
                HDFS: Number of read operations=16
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=8
        Job Counters
                Launched map tasks=4
                Other local map tasks=4
                Total time spent by all maps in occupied slots (ms)=8529360
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=177695
                Total vcore-milliseconds taken by all map tasks=177695
                Total megabyte-milliseconds taken by all map tasks=272939520
        Map-Reduce Framework
                Map input records=4
                Map output records=2468572
                Input split bytes=406
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=1917
                CPU time spent (ms)=86130
                Physical memory (bytes) snapshot=2341957632
                Virtual memory (bytes) snapshot=13183541248
                Total committed heap usage (bytes)=2087714816
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=39077529
22/09/01 05:06:33 INFO mapreduce.ImportJobBase: Transferred 37.2672 MB in 77.3766 seconds (493.1934 KB/sec)
22/09/01 05:06:33 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
```

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

hadoop fs -ls /user/root/testdatabase/SRC_ATM_TRANS/

Screenshot of the imported data:

```
[hadoop@ip-172-31-94-68 mysql-connector-java-8.0.25]$ hadoop fs -ls /user/root/testdatabase/SRC_ATM_TRANS/
Found 5 items
-rw-r--r--   1 hadoop hadoop          0 2022-09-01 05:06 /user/root/testdatabase/SRC_ATM_TRANS/_SUCCESS
-rw-r--r--   1 hadoop hadoop   11167588 2022-09-01 05:06 /user/root/testdatabase/SRC_ATM_TRANS/part-m-00000.gz
-rw-r--r--   1 hadoop hadoop    9760577 2022-09-01 05:06 /user/root/testdatabase/SRC_ATM_TRANS/part-m-00001.gz
-rw-r--r--   1 hadoop hadoop    8385138 2022-09-01 05:06 /user/root/testdatabase/SRC_ATM_TRANS/part-m-00002.gz
-rw-r--r--   1 hadoop hadoop    9764226 2022-09-01 05:06 /user/root/testdatabase/SRC_ATM_TRANS/part-m-00003.gz
```