Data Ingestion from the RDS to HDFS using Sqoop

Sqoop Import command used for importing table from RDS to HDFS:

hadoop fs -rm -r /user/root/SRC_ATM_TRANS

```
22/04/12 21:15:12 INFO Configuration.deprecation: mapred.jar is deprecated. Inst
ead, use mapreduce.job.jar
22/04/12 21:15:12 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
22/04/12 21:15:13 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-48-248.ec2.internal/172.31.48.248:8032
22/04/12 21:15:17 INFO db.DBInputFormat: Using read commited transaction isolation
22/04/12 21:15:18 INFO mapreduce.JobSubmitter: number of splits:1
22/04/12 21:15:18 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1649793921399_0001
22/04/12 21:15:18 INFO impl.YarnClientImpl: Submitted application application_1649793921399_0001
22/04/12 21:15:19 INFO mapreduce.Job: The url to track the job: http://ip-172-31-48-248.ec2.internal:20888/proxy/application_1649793921399_0001/
22/04/12 21:15:19 INFO mapreduce.Job: Running job: job_1649793921399_0001
22/04/12 21:15:28 INFO mapreduce.Job: Job job_1649793921399_0001 running in uber mode : false
22/04/12 21:15:28 INFO mapreduce.Job:  map 0% reduce 0%
22/04/12 21:15:55 INFO mapreduce.Job:  map 100% reduce 0%
22/04/12 21:15:55 INFO mapreduce.Job: Job job_1649793921399_0001 completed successfully
22/04/12 21:15:55 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=189504
                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)=1196352
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=24924
                Total vcore-milliseconds taken by all map tasks=24924
                Total megabyte-milliseconds taken by all map tasks=38283264
        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)=217
                CPU time spent (ms)=26400
                Physical memory (bytes) snapshot=616902656
                Virtual memory (bytes) snapshot=3352891392
                Total committed heap usage (bytes)=535298048
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=531214815
22/04/12 21:15:55 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 42.5548 seconds (11.9048 MB/sec)
22/04/12 21:15:55 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[hadoop@ip-172-31-48-248 mysql-connector-java-8.0.25]$
```

```
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
```

```
[hadoop@ip-172-31-48-248 mysql-connector-java-8.0.25]$ hadoop fs -ls /user/root/SRC_ATM_TRANS
Found 2 items
-rw-r--r--   1 hadoop hadoop          0 2022-04-12 21:15 /user/root/SRC_ATM_TRANS/_SUCCESS
-rw-r--r--   1 hadoop hadoop  531214815 2022-04-12 21:15 /user/root/SRC_ATM_TRANS/part-m-00000
[hadoop@ip-172-31-48-248 mysql-connector-java-8.0.25]$ 
```

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

hadoop fs -ls /user/root/SRC_ATM_TRANS

Screenshot of the imported data: