Data Ingestion from the RDS to HDFS using Sqoop
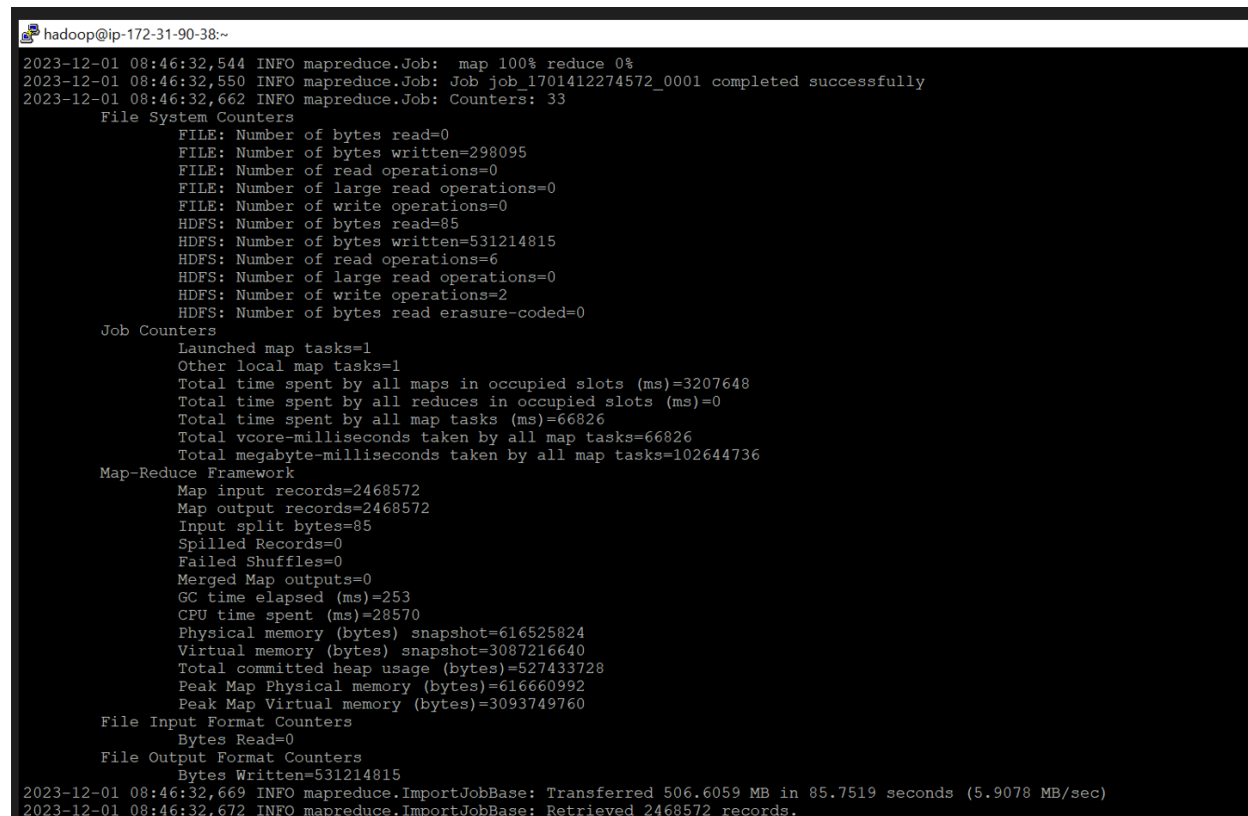
Sqoop Import command used for importing table from RDS to HDFS:

<Sqoop Import Command>

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

```
hadoop@ip-172-31-90-38:~
2023-12-01 08:46:32,544 INFO mapreduce.Job:  map 100% reduce 0%
2023-12-01 08:46:32,550 INFO mapreduce.Job: Job job_1701412274572_0001 completed successfully
2023-12-01 08:46:32,662 INFO mapreduce.Job: Counters: 33
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=298095
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=85
                HDFS: Number of bytes written=531214815
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
                HDFS: Number of bytes read erasure-coded=0
        Job Counters
                Launched map tasks=1
                Other local map tasks=1
                Total time spent by all maps in occupied slots (ms)=3207648
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=66826
                Total vcore-milliseconds taken by all map tasks=66826
                Total megabyte-milliseconds taken by all map tasks=102644736
        Map-Reduce Framework
                Map input records=2468572
                Map output records=2468572
                Input split bytes=85
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=253
                CPU time spent (ms)=28570
                Physical memory (bytes) snapshot=616525824
                Virtual memory (bytes) snapshot=3087216640
                Total committed heap usage (bytes)=527433728
                Peak Map Physical memory (bytes)=616660992
                Peak Map Virtual memory (bytes)=3093749760
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=531214815
2023-12-01 08:46:32,669 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 85.7519 seconds (5.9078 MB/sec)
2023-12-01 08:46:32,672 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
```

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

<Command used>

hadoop fs -ls /user/root/spar_nord_bank_atm

```
[hadoop@ip-172-31-90-38 ~]$ hadoop fs -ls /user/root/spar_nord_bank_atm
Found 2 items
-rw-r--r--   1 hadoop hdfsadmingroup          0 2023-12-01 08:46 /user/root/spar_nord_bank_atm/_SUCCESS
-rw-r--r--   1 hadoop hdfsadmingroup  531214815 2023-12-01 08:46 /user/root/spar_nord_bank_atm/part-m-00000
[hadoop@ip-172-31-90-38 ~]$
```

Screenshot of the imported data:

hadoop fs -cat /user/root/spar_nord_bank_atm/part-m-00000

<Screenshot of the imported data>