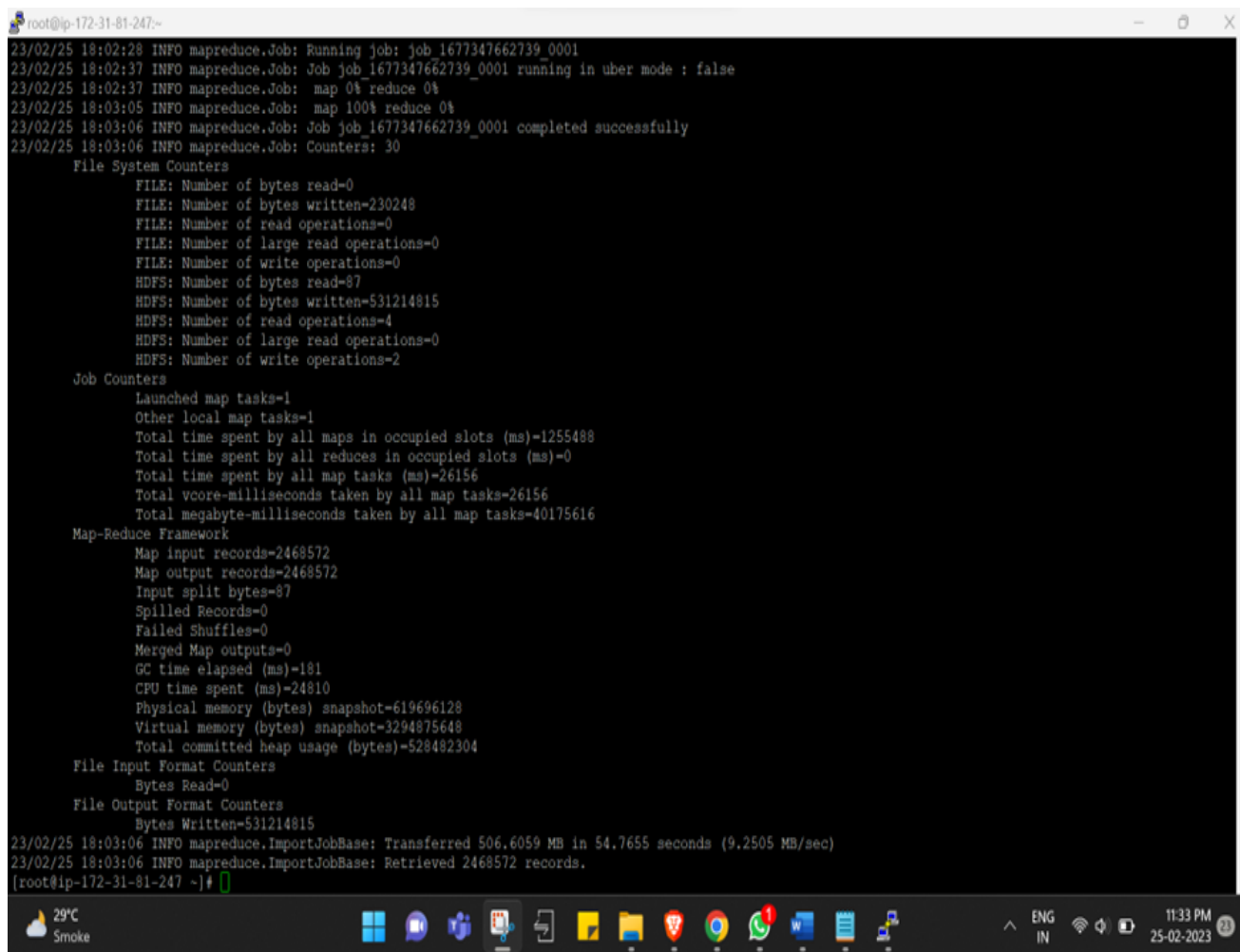


Data Ingestion from the RDS to HDFS using Sqoop

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

```
sqoop import \
--connect jdbc:mysql://upgradtest.cyaieic9bmnf.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
```



```
root@ip-172-31-81-247:~#
23/02/25 18:02:28 INFO mapreduce.Job: Running job: job_1677347662739_0001
23/02/25 18:02:37 INFO mapreduce.Job: Job job_1677347662739_0001 running in uber mode : false
23/02/25 18:02:37 INFO mapreduce.Job: map 0% reduce 0%
23/02/25 18:03:05 INFO mapreduce.Job: map 100% reduce 0%
23/02/25 18:03:06 INFO mapreduce.Job: Job job_1677347662739_0001 completed successfully
23/02/25 18:03:06 INFO mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=230248
  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)=1255488
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=26156
  Total vcore-milliseconds taken by all map tasks=26156
  Total megabyte-milliseconds taken by all map tasks=40175616
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)=181
  CPU time spent (ms)=24810
  Physical memory (bytes) snapshot=619696128
  Virtual memory (bytes) snapshot=3294875648
  Total committed heap usage (bytes)=528482304
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=531214815
23/02/25 18:03:06 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 54.7655 seconds (9.2505 MB/sec)
23/02/25 18:03:06 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-172-31-81-247 ~]#
```

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

```
hadoop fs -ls /user/root/spar_nord_bank_atm
```

Screenshot of the imported data:

