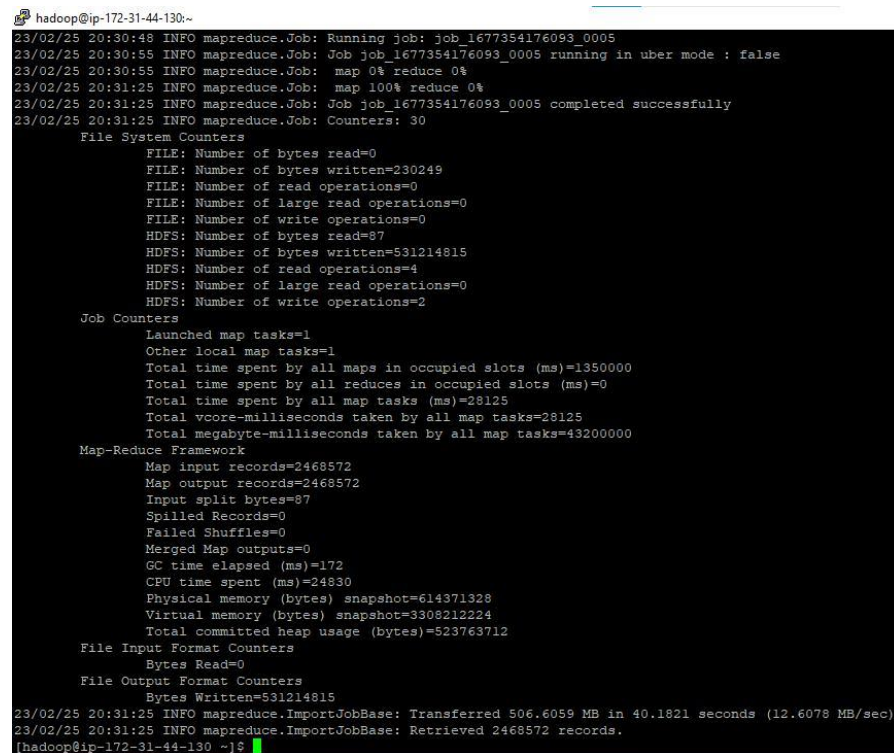


Data Ingestion from the RDS to HDFS using Sqoop

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

```
sqoop import \  
--connect jdbc:mysql://upgraddetest.cyaieic9bmnf.us-east-1.rds.amazonaws.com/testdatabase \  
--table SRC_ATM_TRANS \  
--username student \  
--password STUDENT123 \  
--target-dir /user/root/SRC_ATM_TRANS \  
--m 1
```

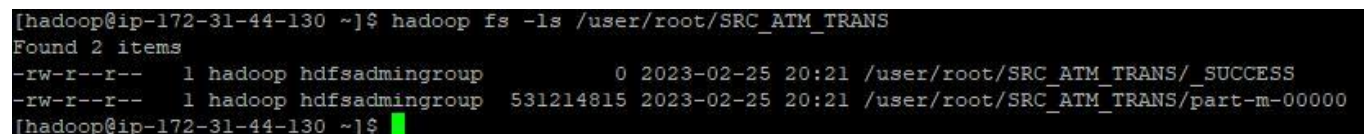


```
hadoop@ip-172-31-44-130:~  
23/02/25 20:30:48 INFO mapreduce.Job: Running job: job_1677354176093_0005  
23/02/25 20:30:55 INFO mapreduce.Job: Job job_1677354176093_0005 running in uber mode : false  
23/02/25 20:30:55 INFO mapreduce.Job: map 0% reduce 0%  
23/02/25 20:31:25 INFO mapreduce.Job: map 100% reduce 0%  
23/02/25 20:31:25 INFO mapreduce.Job: Job job_1677354176093_0005 completed successfully  
23/02/25 20:31:25 INFO mapreduce.Job: Counters: 30  
File System Counters  
  FILE: Number of bytes read=0  
  FILE: Number of bytes written=230249  
  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)=1350000  
  Total time spent by all reduces in occupied slots (ms)=0  
  Total time spent by all map tasks (ms)=28125  
  Total vcore-milliseconds taken by all map tasks=28125  
  Total megabyte-milliseconds taken by all map tasks=43200000  
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)=172  
  CPU time spent (ms)=24830  
  Physical memory (bytes) snapshot=614371328  
  Virtual memory (bytes) snapshot=3308212224  
  Total committed heap usage (bytes)=523763712  
File Input Format Counters  
  Bytes Read=0  
File Output Format Counters  
  Bytes Written=531214815  
23/02/25 20:31:25 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 40.1821 seconds (12.6078 MB/sec)  
23/02/25 20:31:25 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.  
[hadoop@ip-172-31-44-130 ~]$
```

We can see that 2468572 records have been retrieved as an output of Sqoop Import Job.

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

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



```
[hadoop@ip-172-31-44-130 ~]$ hadoop fs -ls /user/root/SRC_ATM_TRANS  
Found 2 items  
-rw-r--r--  1 hadoop hdfsadmin 0 2023-02-25 20:21 /user/root/SRC_ATM_TRANS/_SUCCESS  
-rw-r--r--  1 hadoop hdfsadmin 531214815 2023-02-25 20:21 /user/root/SRC_ATM_TRANS/part-m-00000  
[hadoop@ip-172-31-44-130 ~]$
```

Screenshot of the imported data:

```
hadoop fs -cat /user/root/SRC_ATM_TRANS/part-m-00000
```

[illegible]