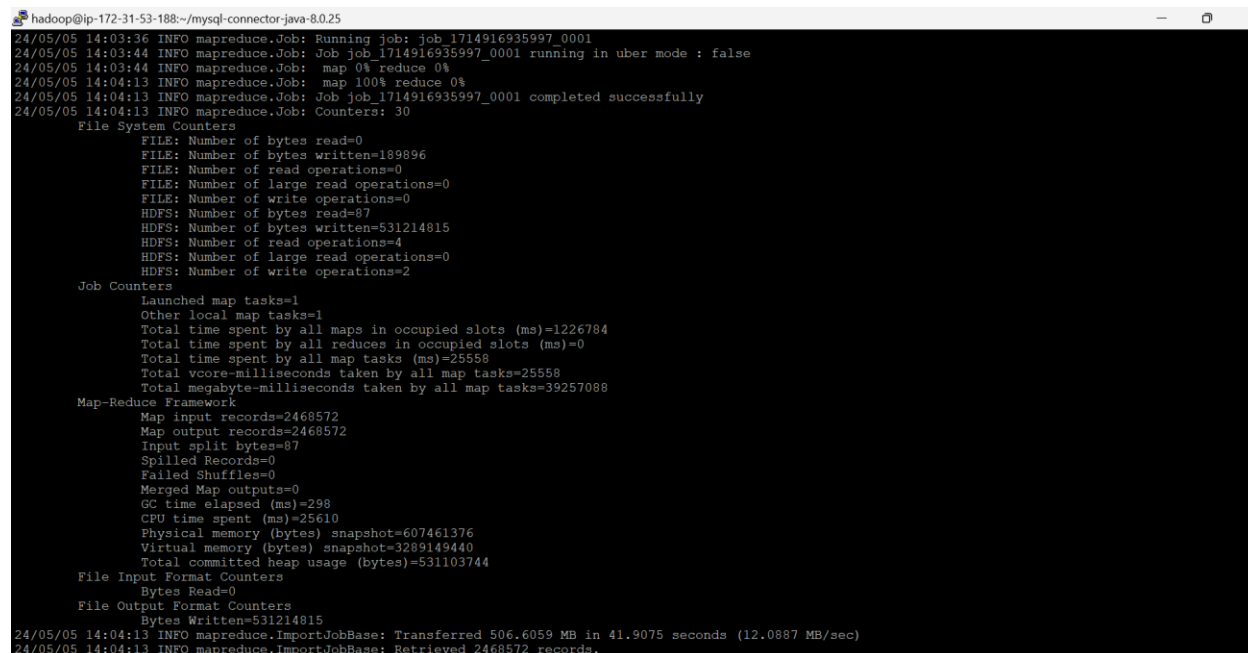


Data Ingestion from the RDS to HDFS using Sqoop

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 \
--username student \
--password STUDENT123 \
--table SRC_ATM_TRANS \
--target-dir /user/root/SRC_ATM_TRANS \
--null-string '\N' \
--null-non-string '\N' \
--num-mappers 1 \
--fetch-size 10000
```

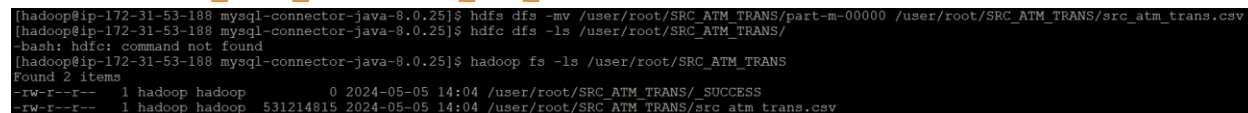


```
hadoop@ip-172-31-53-188:~/mysql-connector-java-8.0.25
24/05/05 14:03:36 INFO mapreduce.Job: Running job: job_1714916935997_0001
24/05/05 14:03:44 INFO mapreduce.Job: Job job_1714916935997_0001 running in uber mode : false
24/05/05 14:03:44 INFO mapreduce.Job: map 0% reduce 0%
24/05/05 14:04:13 INFO mapreduce.Job: map 100% reduce 0%
24/05/05 14:04:13 INFO mapreduce.Job: Job job_1714916935997_0001 completed successfully
24/05/05 14:04:13 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=189896
    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)=1226784
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=25558
    Total vcore-milliseconds taken by all map tasks=25558
    Total megabyte-milliseconds taken by all map tasks=39257088
  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)=298
    CPU time spent (ms)=25610
    Physical memory (bytes) snapshot=607461376
    Virtual memory (bytes) snapshot=3289149440
    Total committed heap usage (bytes)=531103744
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=531214815
24/05/05 14:04:13 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 41.9075 seconds (12.0887 MB/sec)
24/05/05 14:04:13 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
```

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

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

```
hdfs dfs -mv /user/root/SRC_ATM_TRANS/part-m-00000
/user/root/SRC_ATM_TRANS/src_atm_trans.csv
```



```
[hadoop@ip-172-31-53-188 mysql-connector-java-8.0.25]$ hdfs dfs -mv /user/root/SRC_ATM_TRANS/part-m-00000 /user/root/SRC_ATM_TRANS/src_atm_trans.csv
[hadoop@ip-172-31-53-188 mysql-connector-java-8.0.25]$ hdfs dfs -ls /user/root/SRC_ATM_TRANS/
-bash: hdfs: command not found
[hadoop@ip-172-31-53-188 mysql-connector-java-8.0.25]$ hadoop fs -ls /user/root/SRC_ATM_TRANS
Found 2 items
-rw-r--r-- 1 hadoop hadoop 0 2024-05-05 14:04 /user/root/SRC_ATM_TRANS/_SUCCESS
-rw-r--r-- 1 hadoop hadoop 531214815 2024-05-05 14:04 /user/root/SRC_ATM_TRANS/src_atm_trans.csv
```

Screenshot of the imported data:

hadoop fs -cat /user/root/SRC_ATM_TRANS/src_atm_trans.csv | head -10

```
[hadoop@ip-172-31-53-188 mysql-connector-java-8.0.25]$ hadoop fs -cat /user/root/SRC_ATM_TRANS/src_atm_trans.csv | wc -l
2468572
[hadoop@ip-172-31-53-188 mysql-connector-java-8.0.25]$ hadoop fs -cat /user/root/SRC_ATM_TRANS/src_atm_trans.csv | head -10
2017,January,1,Sunday,0,Active,1,NCR,NÄfÄ:stved,Farimagvej,8,4700,55.233,11.763,DKK,MasterCard,5643,Withdrawal,,,55.230,11.761,2616038,Naestved,281.150,1014
,87,7,260,0.215,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,2,NCR,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,MasterCard,1764,Withdrawal,,,57.048,9.935,2616235,NÄfÄ:rresundby,280.640,
1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,2,NCR,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,VISA,1891,Withdrawal,,,57.048,9.935,2616235,NÄfÄ:rresundby,280.640,1020,9
3,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,3,NCR,Ikast,RÄfÄVdhusstrÄfÄ:det,12,7430,56.139,9.154,DKK,VISA,4166,Withdrawal,,,56.139,9.158,2619426,Ikast,281.150,1011,100,
6,240,0.000,75,300,Drizzle,light intensity drizzle
2017,January,1,Sunday,0,Active,4,NCR,Svogerslev,BrÄfÄ:nsager,1,4000,55.634,12.018,DKK,MasterCard,5153,Withdrawal,,,55.642,12.080,2614481,Roskilde,280.610,101
4,87,7,260,0.000,88,701,Mist,mist
2017,January,1,Sunday,0,Active,5,NCR,Nibe,Torvet,1,9240,56.983,9.639,DKK,MasterCard,3269,Withdrawal,,,56.981,9.639,2616483,Nibe,280.640,1020,93,9,250,0.590,9
2,500,Rain,light rain
2017,January,1,Sunday,0,Active,6,NCR,Fredericia,SjÄfÄ:llandsgade,33,7000,55.564,9.757,DKK,MasterCard,887,Withdrawal,,,55.566,9.753,2621951,Fredericia,281.150
,1014,93,7,230,0.290,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,7,Diebold Nixdorf,Hjallerup,Hjallerup Centret,18,9320,57.168,10.148,DKK,Mastercard - on-us,4626,Withdrawal,,,57.165,10.146,262
0275,Hjallerup,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,8,NCR,GlyngÄfÄ:re,FÄfÄ:rgesvej,1,7870,56.762,8.867,DKK,MasterCard,470,Withdrawal,,,56.793,8.853,2615964,Nykobing Mors,281.150,1
011,100,6,240,0.000,75,300,Drizzle,light intensity drizzle
2017,January,1,Sunday,0,Active,9,Diebold Nixdorf,Hadsund,Storegade,12,9560,56.716,10.114,DKK,VISA,8473,Withdrawal,,,56.715,10.117,2620952,Hadsund,280.640,102
0,93,9,250,0.590,92,500,Rain,light rain
cat: Unable to write to output stream.
[hadoop@ip-172-31-53-188 mysql-connector-java-8.0.25]$
```