

## Data Ingestion from the RDS to HDFS using Sqoop

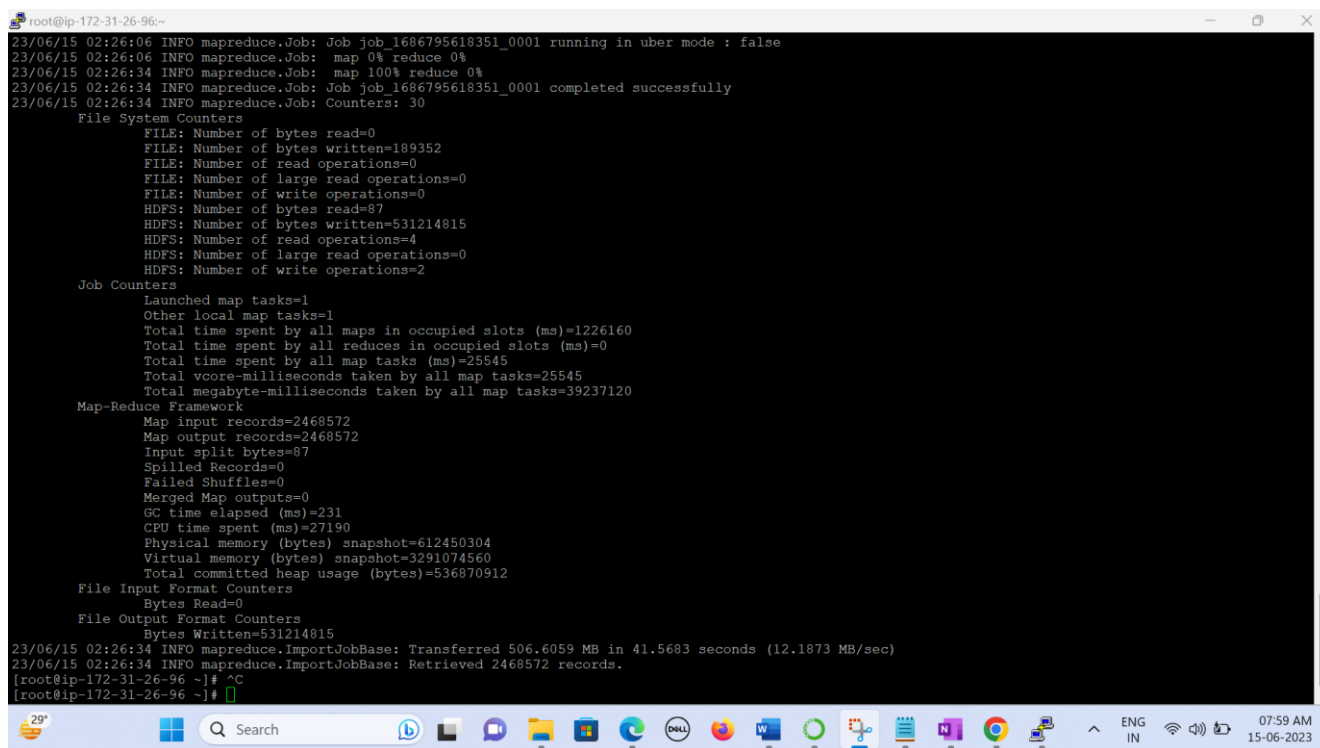
Before running the Sqoop import command, I made sure the target directory is not already created. Otherwise, the Sqoop import command would throw an error:

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

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

```
[root@ip-172-31-26-96 ~]# sqoop import \
> --connect jdbc:mysql://upgradtest.cyaieic9bmnf.us-east1.rds.amazonaws.com/testdatabase \
> --table SRC_ATM_TRANS \
> --username student --password STUDENT123 \
> --target-dir /user/root/SRC_ATM_TRANS \
> -m 1
```

In the screenshot above we can see 2468572 rows have been retrieved.



```
23/06/15 02:26:06 INFO mapreduce.Job: Job job_1686795618351_0001 running in uber mode : false
23/06/15 02:26:06 INFO mapreduce.Job: map 0% reduce 0%
23/06/15 02:26:34 INFO mapreduce.Job: map 100% reduce 0%
23/06/15 02:26:34 INFO mapreduce.Job: Job job_1686795618351_0001 completed successfully
23/06/15 02:26:34 INFO mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=189352
  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)=1226160
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=25545
  Total vcore-milliseconds taken by all map tasks=25545
  Total megabyte-milliseconds taken by all map tasks=39237120
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)=231
  CPU time spent (ms)=27190
  Physical memory (bytes) snapshot=612450304
  Virtual memory (bytes) snapshot=3291074560
  Total committed heap usage (bytes)=536870912
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=531214815
23/06/15 02:26:34 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 41.5683 seconds (12.1873 MB/sec)
23/06/15 02:26:34 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-172-31-26-96 ~]# ^C
[root@ip-172-31-26-96 ~]#
```

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

**hadoop fs -ls /user/root/SRC\_ATM\_TRANS**

```
[root@ip-172-31-26-96 ~]# hadoop fs -ls /user/root/SRC_ATM_TRANS
Found 2 items
-rw-r--r-- 1 root hadoop 0 2023-06-15 02:26 /user/root/SRC_ATM_TRANS/_SUCCESS
-rw-r--r-- 1 root hadoop 531214815 2023-06-15 02:26 /user/root/SRC_ATM_TRANS/part-m-00000
```

In the screenshot above we can see two items:

- The first file is the success file, means the MapReduce job was successful.
- The second file 'part-m-00000' is the one that I imported.

**Screenshot of the imported data:**

A portion of data read from part-m-00000 file:

```
17,February,23,Thursday,19,Inactive,88,NCR,Storcenter indg. A,Hobrovej,452,9200,57.005,9.876,DKK,VISA,4776,Withdrawal,,,56.972,9.848,2612021,Svenstrup,274.704,994,88,
66,0.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Active,39,NCR,Svenstrup,GodthåfjÅvsvej,14,9230,56.973,9.851,DKK,Visa Dankort,7290,Withdrawal,,,56.972,9.848,2612021,Svenstrup,274.704,994,8
8,66,0.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Inactive,52,NCR,FarsÅfÅ.,Torvet,8,9640,56.771,9.340,DKK,Mastercard - on-us,2120,Withdrawal,,,56.773,9.339,2622310,Farso,274.704,994,88,8,66
.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Active,64,NCR,Horsens,GrÅfÅ,nlandsvej,5,8700,55.859,9.854,DKK,Mastercard - on-us,8729,Withdrawal,,,55.861,9.850,2619771,Horsens,273.830,978
00,6,70,0.000,75,600,Snow,light snow
17,February,23,Thursday,19,Active,10,NCR,NÅfÅ,rresundby,Torvet,6,9400,57.059,9.922,DKK,Mastercard - on-us,7287,Withdrawal,,,57.048,9.919,2624886,Aalborg,274.704,994,8
8,66,0.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Inactive,52,NCR,FarsÅfÅ.,Torvet,8,9640,56.771,9.340,DKK,VISA,249,Withdrawal,,,56.773,9.339,2622310,Farso,274.704,994,88,8,66,0.000,92,804,C
uds,overcast clouds
17,February,23,Thursday,19,Active,10,NCR,NÅfÅ,rresundby,Torvet,6,9400,57.059,9.922,DKK,Mastercard - on-us,9383,Withdrawal,,,57.048,9.919,2624886,Aalborg,274.704,994,8
8,66,0.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Active,48,Diebold Nixdorf,BrÅfÅ,nderslev,Algade,4,9700,57.269,9.945,DKK,Mastercard - on-us,6167,Withdrawal,,,57.118,9.857,2610823,Vadum,274
04,984,88,8,66,0.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Active,33,NCR,Vadum,Ellehammersvej,43,9430,57.118,9.861,DKK,Visa Dankort - on-us,2685,Withdrawal,,,57.118,9.857,2610823,Vadum,274.704,994,8
8,66,0.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Active,34,NCR,Skipperen,Vestre Alle,2,9000,57.034,9.908,DKK,HÅfÅ|vekort - on-us,4926,Withdrawal,,,57.048,9.919,2624886,Aalborg,274.704,994,
,8,66,0.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Active,63,NCR,Brugsen 1 Breum,AakjÅfÅ|rsvej,1,7870,56.688,9.069,DKK,Visa Dankort,6576,Withdrawal,,,56.586,9.136,2619908,Hojlslev,274.704,994
8,8,66,0.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Active,34,NCR,Skipperen,Vestre Alle,2,9000,57.034,9.908,DKK,HÅfÅ|vekort - on-us,8103,Withdrawal,,,57.048,9.919,2624886,Aalborg,274.704,994,
,8,66,0.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Inactive,102,NCR,Aalborg Storcenter Afd,Hobrovej,452,9200,57.005,9.876,EUR,Visa Dankort - on-us,785,Withdrawal,,,56.972,9.848,2612021,Sven
strup,274.704,994,88,8,66,0.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Active,34,NCR,Skipperen,Vestre Alle,2,9000,57.034,9.908,DKK,HÅfÅ|vekort - on-us,9617,Withdrawal,,,57.048,9.919,2624886,Aalborg,274.704,994,
,8,66,0.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Active,13,NCR,SÅfÅ|by,Vestergade,3,9300,57.334,10.515,DKK,Visa Dankort - on-us,5730,Withdrawal,,,57.441,10.537,2621927,Frederikshavn,275.15
997,99,9,59,0.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Active,75,NCR,Nordkraft,Kjellerups Torv,1,9000,57.047,9.932,DKK,CIRRUS,9800,Withdrawal,,,57.048,9.935,2616235,NÅfÅ,rresundby,274.704,994,88
,66,0.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Inactive,52,NCR,FarsÅfÅ.,Torvet,8,9640,56.771,9.340,DKK,Visa Dankort - on-us,1810,Withdrawal,,,56.773,9.339,2622310,Farso,274.704,994,88,8
,0.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Inactive,102,NCR,Aalborg Storcenter Afd,Hobrovej,452,9200,57.005,9.876,DKK,Visa Dankort - on-us,9651,Withdrawal,,,56.972,9.848,2612021,Sve
strup,274.704,994,88,8,66,0.000,92,804,Clouds,overcast clouds
```