

## Data Ingestion from the RDS to HDFS using Sqoop

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

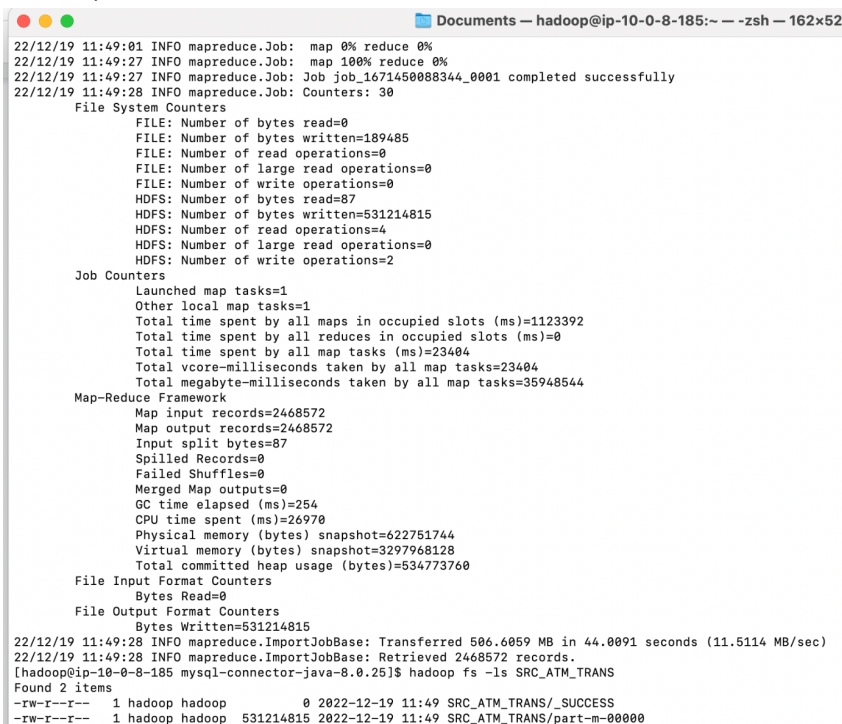
```
sqoop import --connect jdbc:mysql://upgradtest.cyaieic9bmfn.us-east-1.rds.amazonaws.com/testdatabase --username student --password STUDENT123 --table SRC_ATM_TRANS -m 1
```

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

```
hadoop fs -ls SRC_ATM_TRANS
```

### Screenshot of the imported data:

Following screenshots shows that 2468572 records have been imported into the HDFS, and a parquet file has been created for the imported data.



```
Documents — hadoop@ip-10-0-8-185:~ — zsh — 162x52
22/12/19 11:49:01 INFO mapreduce.Job: map 0% reduce 0%
22/12/19 11:49:27 INFO mapreduce.Job: map 100% reduce 0%
22/12/19 11:49:27 INFO mapreduce.Job: Job job_1671450888344_0001 completed successfully
22/12/19 11:49:28 INFO mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=189485
  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)=1123392
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=23404
  Total vcore-milliseconds taken by all map tasks=23404
  Total megabyte-milliseconds taken by all map tasks=35948544
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)=254
  CPU time spent (ms)=26970
  Physical memory (bytes) snapshot=622751744
  Virtual memory (bytes) snapshot=3297968128
  Total committed heap usage (bytes)=534773760
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=531214815
22/12/19 11:49:28 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 44.0091 seconds (11.5114 MB/sec)
22/12/19 11:49:28 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[hadoop@ip-10-0-8-185 mysql-connector-java-8.0.25]$ hadoop fs -ls SRC_ATM_TRANS
Found 2 items
-rw-r--r-- 1 hadoop hadoop 0 2022-12-19 11:49 SRC_ATM_TRANS/_SUCCESS
-rw-r--r-- 1 hadoop hadoop 531214815 2022-12-19 11:49 SRC_ATM_TRANS/part-m-00000
```