

## 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/etl -m 1
```

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

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

Screenshot of the imported data:

```
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)=1073376
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=22362
  Total vcore-milliseconds taken by all map tasks=22362
  Total megabyte-milliseconds taken by all map tasks=34348032
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)=187
  CPU time spent (ms)=25550
  Physical memory (bytes) snapshot=643862528
  Virtual memory (bytes) snapshot=3311247360
  Total committed heap usage (bytes)=526909440
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=531214815
22/10/29 02:49:37 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 42.0564 seconds (12.0459 MB/sec)
22/10/29 02:49:37 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-172-31-45-144 mysql-connector-java-8.0.25]# hadoop fs -ls /user/root/etl
Found 2 items
-rw-r--r-- 1 root hdfsadmingroup 0 2022-10-29 02:49 /user/root/etl/_SUCCESS
-rw-r--r-- 1 root hdfsadmingroup 531214815 2022-10-29 02:49 /user/root/etl/part-m-00000
[root@ip-172-31-45-144 mysql-connector-java-8.0.25]# hadoop fs -cat /user/root/etl/part-m-00000 | head -5
2017,January,1,Sunday,0,Active,1,NCR,NÃfÃstved,Farimagsvej,8,4700,55.233,11.763,DKK,MasterCard,5643,Withdrawal,,55.230,11.761,2616038,Naest
ved,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,NÃfÃ,rr
esundby,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,NÃfÃ,rresundb
y,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,3,NCR,Ikast,RÃfÃVdhusstrÃfÃdet,12,7430,56.139,9.154,DKK,VISA,4166,Withdrawal,,56.139,9.158,2619426,Ikast,2
81.150,1011,100,6,240,0.000,75,300,Drizzle,light intensity drizzle
2017,January,1,Sunday,0,Active,4,NCR,Svogerslev,BrÃfÃnsager,1,4000,55.634,12.018,DKK,MasterCard,5153,Withdrawal,,55.642,12.000,2614481,Rosk
ilde,280.610,1014,87,7,260,0.000,88,701,Mist,mist
cat: Unable to write to output stream.
[root@ip-172-31-45-144 mysql-connector-java-8.0.25]#
```