

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.cyaieic9bmnf.us-east-1.rds.amazonaws.com/testdatabase \  
--table SRC_ATM_TRANS \  
--username student --password STUDENT123 \  
--target-dir /user/root/spar_nord_bank_atm \  
--m 1
```

```
root@ip-172-31-67-120:~  
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)=1181760  
  Total time spent by all reduces in occupied slots (ms)=0  
  Total time spent by all map tasks (ms)=24620  
  Total vcore-milliseconds taken by all map tasks=24620  
  Total megabyte-milliseconds taken by all map tasks=37816320  
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)=306  
  CPU time spent (ms)=26660  
  Physical memory (bytes) snapshot=615636992  
  Virtual memory (bytes) snapshot=3293876224  
  Total committed heap usage (bytes)=542113792  
File Input Format Counters  
  Bytes Read=0  
File Output Format Counters  
  Bytes Written=531214815  
23/05/01 18:10:25 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 39.0193 seconds (12.9835 MB/sec)  
23/05/01 18:10:25 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.  
[root@ip-172-31-67-120 ~]#
```

As per the above screenshot, 2468572 rows retrieved.

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

Command used:

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

Screenshot:

```
[root@ip-172-31-67-120 ~]# hadoop fs -ls /user/root/spar_nord_bank_atm
Found 2 items
-rw-r--r-- 1 root hadoop 0 2023-05-01 18:10 /user/root/spar_nord_bank_atm/_SUCCESS
-rw-r--r-- 1 root hadoop 531214815 2023-05-01 18:10 /user/root/spar_nord_bank_atm/part-m-00000
[root@ip-172-31-67-120 ~]#
```

Above screenshot shows couple of files listed:

- The first file is the `_SUCCESS` file, indicating that the MapReduce job was successful.
- The second file 'part-m-00000' is the one that we imported.

Note: One mapper in import command used, thus the data is in a single file.

Screenshot of the imported data:

```
root@ip-172-31-67-120:~
ing Mors,279.150,997,75,11,290,0.000,0.800,Clear,sky is Clear
2017,January,3,Tuesday,19,Active,1,NCR,NÅfÅ;stved,Farimagvej,8,4700,55.233,11.763,DKK,Mastercard - on-us,5436,Withdrawal,,,55.230,11.76
1,2616038,Naestved,279.150,995,87,11,290,0.000,92,520,Rain,light intensity shower rain
2017,January,3,Tuesday,19,Active,31,NCR,Slagelse,Mariendals Alle,29,4200,55.398,11.342,DKK,Visa Dankort,7700,Withdrawal,,,55.403,11.355,
2613460,Slagelse,279.150,995,87,11,290,0.000,92,520,Rain,light intensity shower rain
2017,January,3,Tuesday,19,Active,8,NCR,GlyngÅfÅ;re,FÅfÅ;rgevej,1,7870,56.762,8.867,DKK,VISA,2191,Withdrawal,,,56.793,8.853,2615964,Nykob
ing Mors,279.150,997,75,11,290,0.000,0.800,Clear,sky is Clear
2017,January,3,Tuesday,19,Active,48,Diebold Nixdorf,BrÅfÅ;nderslev,Algade,4,9700,57.269,9.945,DKK,Visa Dankort - on-us,7853,Withdrawal,,,
,57.118,9.857,2610823,Vadum,278.069,1004,83,15,296,0.000,0.800,Clear,sky is Clear
2017,January,3,Tuesday,19,Active,49,NCR,Bindslev,NÅfÅ;rrebro,18,9881,57.541,10.200,DKK,Visa Dankort,2693,Withdrawal,,,57.471,10.203,2614
010,Sindal,278.069,1004,83,15,296,0.000,0.800,Clear,sky is Clear
2017,January,3,Tuesday,19,Active,46,Diebold Nixdorf,HelsingÅfÅ;r,Sct. Olai Gade,39,3000,56.036,12.612,DKK,Mastercard - on-us,9907,Withdr
awal,,,56.036,12.614,2620473,Helsingor,279.480,991,81,11,260,0.000,20,520,Rain,light intensity shower rain
2017,January,3,Tuesday,19,Active,10,NCR,NÅfÅ;rresundby,Torvet,6,9400,57.059,9.922,DKK,Mastercard - on-us,1456,Withdrawal,,,57.048,9.919,
2624886,Aalborg,278.069,1004,83,15,296,0.000,0.800,Clear,sky is Clear
2017,January,3,Tuesday,19,Active,1,NCR,NÅfÅ;stved,Farimagvej,8,4700,55.233,11.763,DKK,Visa Dankort,7561,Withdrawal,,,55.230,11.761,2616
038,Naestved,279.150,995,87,11,290,0.000,92,520,Rain,light intensity shower rain
2017,January,3,Tuesday,19,Active,45,NCR,Abildgaard,HjÅfÅ;rtingvej,144,9900,57.447,10.506,DKK,Visa Dankort,3434,Withdrawal,,,57.441,10.53
7,2621927,Frederikshavn,280.094,1002,99,17,297,0.000,0.800,Clear,sky is Clear
2017,January,3,Tuesday,19,Inactive,50,NCR,Aarhus,SÅfÅ;nder Alle,11,8000,56.153,10.206,DKK,Mastercard - on-us,4486,Withdrawal,,,56.157,10
.211,2624652,Aarhus,279.150,995,90,10,280,0.000,88,804,Clouds,overcast clouds
2017,January,3,Tuesday,19,Active,17,NCR,Randers,ÅfÅ;stervold,16,8900,56.462,10.038,DKK,Visa Dankort,2131,Withdrawal,,,56.715,10.117,2620
952,Hadsund,278.069,1004,83,15,296,0.000,0.800,Clear,sky is Clear
2017,January,3,Tuesday,19,Active,19,NCR,Kolding,Vejlevej,135,6000,55.505,9.457,EUR,Visa Dankort,7197,Withdrawal,,,55.490,9.472,2618528,K
olding,280.560,1000,81,9,280,0.000,92,520,Rain,light intensity shower rain
2017,January,3,Tuesday,19,Active,68,NCR,Drive Up Vejle,Havneparken,4,7100,55.707,9.541,DKK,Visa Dankort,9590,Withdrawal,,,55.709,9.536,
2610613,Vejle,280.410,999,81,12,280,0.000,92,520,Rain,light intensity shower rain
```

Case Study submitted by:

Susil Patro, Vivek Agrawal & Harshal Sapkade

