

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
```

```
hadoop@ip-172-31-26-34:~$ 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
```

```
hadoop@ip-172-31-26-34:~$
```

```
FILE: Number of bytes written=189822
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)=1128672
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=23514
  Total vcore-milliseconds taken by all map tasks=23514
  Total megabyte-milliseconds taken by all map tasks=36117504
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)=218
  CPU time spent (ms)=26630
  Physical memory (bytes) snapshot=623423488
  Virtual memory (bytes) snapshot=3301744640
  Total committed heap usage (bytes)=537395200
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=531214815
22/12/19 10:37:59 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 39.2065 seconds (12.9215 MB/sec)
22/12/19 10:37:59 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
```

Total of **2468572** rows have been retrieved.

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

hadoop fs -ls /user/root/spar_nord_bank_atm

```
[hadoop@ip-172-31-26-34 ~]$ hadoop fs -ls /user/root/spar_nord_bank_atm
Found 2 items
-rw-r--r--  1 hadoop hadoop          0 2022-12-19 10:37 /user/root/spar_nord_bank_atm/ SUCCESS
-rw-r--r--  1 hadoop hadoop 531214815 2022-12-19 10:37 /user/root/spar_nord_bank_atm/part-m-00000
[hadoop@ip-172-31-26-34 ~]$
```

There are two items in the screenshot :

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

Since we used only one mapper in my import command that is why the data is in a single file.

hadoop fs -cat /user/root/spar_nord_bank_atm/part-m-00000

```
hadoop@ip-172-31-26-34:~$ hadoop fs -cat /user/root/spar_nord_bank_atm/part-m-00000
93,2,280,0.000,92,500,Rain,light rain
2017,January,1,Sunday,13,Active,28,NCR,LÅfÅ,gstÅfÅ,r,ÅfEasterbrogade,8,9670,56.967,9.253,DKK,MasterCard,8201,Withdrawal,,,56.962,9.258,2617467,Logstor,278.521,1017,91,6
,305,0.000,68,803,Clouds,broken clouds
2017,January,1,Sunday,13,Active,67,NCR,HÅfÅ,jÅlev,ÅfEsterriavej,2,7840,56.551,9.110,DKK,Mastercard - on-us,3295,Withdrawal,,,56.586,9.136,2619908,Højslev,278.521,1017,
91,6,305,0.000,68,803,Clouds,broken clouds
2017,January,1,Sunday,13,Active,25,Diebold Nixdorf,Odense,FÅfÅ,lledvej,3,5000,55.394,10.370,DKK,MasterCard,1873,Withdrawal,,,55.396,10.388,2615876,Odense,279.150,1011,9
3,2,280,0.000,92,500,Rain,light rain
2017,January,1,Sunday,13,Active,28,NCR,LÅfÅ,gstÅfÅ,r,ÅfEasterbrogade,8,9670,56.967,9.253,DKK,MasterCard,9479,Withdrawal,,,56.962,9.258,2617467,Logstor,278.521,1017,91,6
,305,0.000,68,803,Clouds,broken clouds
2017,January,1,Sunday,13,Active,78,Diebold Nixdorf,Nyborg,Vestergade,35,5800,55.318,10.781,DKK,Visa Dankort - on-us,1778,Withdrawal,,,55.060,10.607,2612045,Svendborg,27
9.150,1011,93,2,280,0.000,92,500,Rain,light rain
2017,January,1,Sunday,13,Active,37,NCR,Silkeborg,Borgergade,36,8600,56.179,9.552,DKK,MasterCard,450,Withdrawal,,,56.170,9.545,2614030,Silkeborg,278.840,1009,93,3,270,0.
000,0,800,Clear,Sky is Clear
2017,January,1,Sunday,13,Inactive,3,NCR,Ikast,BÅfÅvhusstrÅfÅ;det,12,7430,56.139,9.154,DKK,Visa Dankort,6918,Withdrawal,,,56.139,9.158,2619426,Ikast,278.680,1009,86,4,2
80,0.000,0,800,Clear,Sky is Clear
2017,January,1,Sunday,13,Active,24,NCR,Hobro,Adelgade,31,9500,56.638,9.794,DKK,Dankort - on-us,3238,Withdrawal,,,56.643,9.790,2620167,Hobro,278.521,1017,91,6,305,0.000,
68,803,Clouds,broken clouds
2017,January,1,Sunday,13,Active,60,NCR,Hirtshals,JÅfÅ,rge Fibigersgade,4,9850,57.591,9.957,DKK,Mastercard - on-us,5437,Withdrawal,,,57.588,9.959,2620279,Hirtshals,278.
521,1017,91,6,305,0.000,68,803,Clouds,broken clouds
2017,January,1,Sunday,13,Active,15,NCR,Vestre,Mastervej,36,9000,57.053,9.905,DKK,Mastercard - on-us,7470,Withdrawal,,,57.048,9.919,2624886,Aalborg,278.521,1017,91,6,305
,0.000,68,803,Clouds,broken clouds
2017,January,1,Sunday,13,Active,34,NCR,Skipperen,Vestre Alle,2,9000,57.034,9.908,DKK,Dankort - on-us,1038,Withdrawal,,,57.048,9.919,2624886,Aalborg,278.521,1017,91,6,30
5,0.000,68,803,Clouds,broken clouds
2017,January,1,Sunday,13,Active,65,NCR,Storvorde,VandvÅfÅ;rkvsvej,2,9280,57.005,10.101,DKK,MasterCard,2781,Withdrawal,,,57.004,10.101,2612301,Storvorde,278.521,1017,91,6
,305,0.000,68,803,Clouds,broken clouds
2017,January,1,Sunday,13,Active,11,NCR,Sauersvej,Fridtjof Nansens Vej,2,9210,57.023,9.940,DKK,Mastercard - on-us,790,Withdrawal,,,57.048,9.935,2616235,NÅfÅ,rresundby,27
8.521,1017,91,6,305,0.000,68,803,Clouds,broken clouds
2017,January,1,Sunday,13,Active,78,Diebold Nixdorf,Nyborg,Vestergade,35,5800,55.318,10.781,DKK,Visa Dankort,5661,Withdrawal,,,55.060,10.607,2612045,Svendborg,279.150,101
1,93,2,280,0.000,92,500,Rain,light rain
2017,January,1,Sunday,13,Active,11,NCR,Sauersvej,Fridtjof Nansens Vej,2,9210,57.023,9.940,DKK,Mastercard - on-us,1488,Withdrawal,,,57.048,9.935,2616235,NÅfÅ,rresundby,2
78.521,1017,91,6,305,0.000,68,803,Clouds,broken clouds
2017,January,1,Sunday,13,Inactive,29,NCR,Skelagervej 15,Skelagervej,15,9000,57.023,9.891,DKK,Visa Dankort - on-us,458,Withdrawal,,,57.048,9.919,2624886,Aalborg,278.521,
1017,91,6,305,0.000,68,803,Clouds,broken clouds
2017,January,1,Sunday,13,Active,35,NCR,Aabybro,ÅfEsterbrogade,6,9440,57.162,9.730,DKK,Visa Dankort - on-us,7824,Withdrawal,,,57.158,9.737,2625037,Aabybro,278.521,1017,91,6
,305,0.000,68,803,Clouds,broken clouds
2017,January,1,Sunday,13,Active,44,NCR,Lyngby,Jernbanevej,6,2800,55.772,12.500,DKK,Visa Dankort,7748,Withdrawal,,,55.770,12.504,2618468,Kongens Lyngby,279.150,1009,93,4
,260,0.000,75,803,Clouds,broken clouds
2017,January,1,Sunday,13,Active,13,NCR,SÅfÅ;by,Vestergade,3,9300,57.334,10.515,DKK,Visa Dankort - on-us,5268,Withdrawal,,,57.441,10.537,2621927,Frederikshavn,279.646,10
17,100,7,310,0.000,56,803,Clouds,broken clouds
2017,January,1,Sunday,13,Active,22,NCR,Roskilde,FÅfÅ,benhavnsvej,65,4000,55.642,12.106,DKK,MasterCard,3095,Withdrawal,,,55.642,12.080,2614481,Roskilde,279.150,1009,93,4
,270,1.925,92,500,Rain,light rain
2017,January,1,Sunday,13,Active,45,NCR,Abildgaard,HjÅfÅ,rtingvej,144,9900,57.447,10.506,DKK,Visa Dankort - on-us,9669,Withdrawal,,,57.441,10.537,2621927,Frederikshavn,2
```