

## 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.cyaielc9bmnf.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
```

```
22/04/30 07:42:51 INFO mapreduce.Job: Job job_1651302571224_0001 running in uber mode : false
22/04/30 07:42:54 INFO mapreduce.Job: map 0% reduce 0%
22/04/30 07:43:22 INFO mapreduce.Job: map 100% reduce 0%
22/04/30 07:43:23 INFO mapreduce.Job: Job job_1651302571224_0001 completed successfully
22/04/30 07:43:23 INFO mapreduce.Job: Counters: 30
    File System Counters
      FILE: Number of bytes read=0
      FILE: Number of bytes written=188956
      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)=1234272
      Total time spent by all reduces in occupied slots (ms)=0
      Total time spent by all map tasks (ms)=25714
      Total vcore-milliseconds taken by all map tasks=25714
      Total megabyte-milliseconds taken by all map tasks=39496704
    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)=251
      CPU time spent (ms)=28150
      Physical memory (bytes) snapshot=629555200
      Virtual memory (bytes) snapshot=3317964800
      Total committed heap usage (bytes)=542638080
    File Input Format Counters
      Bytes Read=0
    File Output Format Counters
      Bytes Written=531214815
22/04/30 07:43:23 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 49.8925 seconds (10.1539 MB/sec)
22/04/30 07:43:23 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-172-31-94-32 ~]#
```

In the screenshot above we can see 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
```

```
22/04/30 07:43:23 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-172-31-94-32 ~]# hadoop fs -ls /user/root/spar_nord_bank_atm
Found 2 items
-rw-r--r--  1 root hadoop          0 2022-04-30 07:43 /user/root/spar_nord_bank_atm/_SUCCESS
-rw-r--r--  1 root hadoop 531214815 2022-04-30 07:43 /user/root/spar_nord_bank_atm/part-m-00000
[root@ip-172-31-94-32 ~]#
```

In the screenshot above we can see two items:

- 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 I used only one mapper in my import command, thus the data is in a single file.

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,Godthåfsvej,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,Farsøvej,,Torvet,8,9640,56.771,9.340,DKK,Mastercard - on-us,2120,Withdrawal,,,56.773,9.339,2622310,Farsø,274.704,994,88,8,66
.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Active,64,NCR,Horsens,Gråvej,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,Næstved,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,Farsøvej,,Torvet,8,9640,56.771,9.340,DKK,VISA,249,Withdrawal,,,56.773,9.339,2622310,Farsø,274.704,994,88,8,66,0.000,92,804,C
lds,overcast clouds
17,February,23,Thursday,19,Active,10,NCR,Næstved,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,Bråvej,nderslev,Algade,4,9700,57.269,9.945,DKK,Mastercard - on-us,6167,Withdrawal,,,57.118,9.857,2610823,Vadum,274
04,994,88,8,66,0.000,92,804,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,HÅF{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 i Breum,Aakj{rsvej,1,7870,56.688,9.069,DKK,Visa Dankort,6576,Withdrawal,,,56.586,9.136,2619908,Højslev,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,HÅF{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,HÅF{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,SÅf{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,Næstved,rresundby,274.704,994,88
,66,0.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Inactive,52,NCR,Farsøvej,,Torvet,8,9640,56.771,9.340,DKK,Visa Dankort - on-us,1810,Withdrawal,,,56.773,9.339,2622310,Farsø,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
```