

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 \
--table SRC_ATM_TRANS \
--username student --password STUDENT123 \
--target-dir /user/root/spar_nord_bank_atm \
-m 1
```

```
2023-07-17 15:35:12,784 INFO mapreduce.Job: The URL to track the job: http://ip-172-31-7-110.ec2.internal:20000/proxy/
2023-07-17 15:35:22,844 INFO mapreduce.Job: Running job: job_1689605148836_0001
2023-07-17 15:35:34,100 INFO mapreduce.Job: Job job_1689605148836_0001 running in uber mode : false
2023-07-17 15:35:34,101 INFO mapreduce.Job: map 0% reduce 0%
2023-07-17 15:36:04,330 INFO mapreduce.Job: map 100% reduce 0%
2023-07-17 15:36:05,342 INFO mapreduce.Job: Job job_1689605148836_0001 completed successfully
2023-07-17 15:36:05,466 INFO mapreduce.Job: Counters: 33
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=297869
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=85
    HDFS: Number of bytes written=531214815
    HDFS: Number of read operations=6
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
  Job Counters
    Launched map tasks=1
    Other local map tasks=1
    Total time spent by all maps in occupied slots (ms)=1342464
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=27968
    Total vcore-milliseconds taken by all map tasks=27968
    Total megabyte-milliseconds taken by all map tasks=42958848
  Map-Reduce Framework
    Map input records=2468572
    Map output records=2468572
    Input split bytes=85
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=255
    CPU time spent (ms)=28610
    Physical memory (bytes) snapshot=625508352
    Virtual memory (bytes) snapshot=3088277504
    Total committed heap usage (bytes)=585105408
    Peak Map Physical memory (bytes)=630480896
    Peak Map Virtual memory (bytes)=3096535040
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=531214815
2023-07-17 15:36:05,476 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 49.127 seconds (10.3122 MB/sec)
2023-07-17 15:36:05,479 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[hadoop@ip-172-31-7-110 ~]$
```

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
```

```
[hadoop@ip-172-31-7-110 ~]$
[hadoop@ip-172-31-7-110 ~]$ hadoop fs -ls /user/root/spar_nord_bank_atm
Found 2 items
-rw-r--r-- 1 hadoop hdfsadmin group 0 2023-07-17 15:36 /user/root/spar_nord_bank_atm/_SUCCESS
-rw-r--r-- 1 hadoop hdfsadmin group 531214815 2023-07-17 15:36 /user/root/spar_nord_bank_atm/part-m-00000
[hadoop@ip-172-31-7-110 ~]$
```

Screenshot of the imported data:

```
17,February,23,Thursday,19,Inactive,88,MCR,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,MCR,Svenstrup,GodthÃfÃbveje,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,MCR,FarsÃfÃ.,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,MCR,Horsens,GrÃfÃ,niandsvej,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,MCR,NÃfÃ,rrresundby,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,MCR,FarsÃfÃ.,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,MCR,NÃfÃ,rrresundby,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,BrÃfÃ,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,MCR,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,MCR,Skipperen,Vestre Alle,2,9000,57.034,9.908,DKK,HÃfÃ,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,MCR,Brugsen i Breum,AakjÃfÃ,rsvej,1,7870,56.688,9.069,DKK,Visa Dankort,6576,Withdrawal,,,56.586,9.136,2619908,Hojlslev,274.704,994
8,8,66,0.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Active,34,MCR,Skipperen,Vestre Alle,2,9000,57.034,9.908,DKK,HÃfÃ,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,MCR,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,MCR,Skipperen,Vestre Alle,2,9000,57.034,9.908,DKK,HÃfÃ,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,MCR,SÃfÃ,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,MCR,Nordkraft,Kjellerups Torv,1,9000,57.047,9.932,DKK,CIRKUS,9800,Withdrawal,,,57.048,9.935,2616235,NÃfÃ,rrresundby,274.704,994,88
,66,0.000,92,804,Clouds,overcast clouds
17,February,23,Thursday,19,Inactive,52,MCR,FarsÃfÃ.,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,MCR,Aalborg Storcenter Afd,Hobrovej,452,9200,57.005,9.876,DKK,Visa Dankort - on-us,9651,Withdrawal,,,56.972,9.848,2612021,Sve
nstrup,274.704,994,88,8,66,0.000,92,804,Clouds,overcast clouds
```