Data Ingestion from the RDS to HDFS using Sqoop

Before running Sqoop command, ensure that target directory does not already exist in HDFS. Otherwise, the Sqoop import command will throw error.

```
hadoop fs -rm -r /user/root/spar_nord_bank_atm
```

Sqoop Import command:

```
sqoop import \
--connect jdbc:mysql://upgraddetest.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
```

```
23/02/26 11:07:15 INFO mapreduce.Job: The url to track the job: http://ip-10-0-11-23.ec2.internal:20888/proxy/application_1677408566939_0001/
23/02/26 11:07:15 INFO mapreduce.Job: Running job: job_1677408566939_0001
23/02/26 11:07:24 INFO mapreduce.Job: Job job_1677408566939_0001 running in uber mode : false
23/02/26 11:07:24 INFO mapreduce.Job:  map 0% reduce 0%
23/02/26 11:07:59 INFO mapreduce.Job:  map 100% reduce 0%
23/02/26 11:07:59 INFO mapreduce.Job: Job job_1677408566939_0001 completed successfully
23/02/26 11:07:59 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=189488
                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)=1506336
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=31382
                Total vcore-milliseconds taken by all map tasks=31382
                Total megabyte-milliseconds taken by all map tasks=48202752
        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)=250
                CPU time spent (ms)=27720
                Physical memory (bytes) snapshot=626499584
                Virtual memory (bytes) snapshot=3305517056
                Total committed heap usage (bytes)=540540928
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=531214815
23/02/26 11:07:59 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 51.1452 seconds (9.9053 MB/sec)
23/02/26 11:07:59 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[hadoop@ip-10-0-11-23 ~]$
```

In the screenshot it can be seen that 2468572 rows have been retrieved.

```
                 Bytes written=531214815
23/02/26 11:07:59 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 51.1452 seconds (9.9053 MB/sec)
23/02/26 11:07:59 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[hadoop@ip-10-0-11-23 ~]$
[hadoop@ip-10-0-11-23 ~]$
[hadoop@ip-10-0-11-23 ~]$ hadoop fs -ls /user/root/
Found 1 items
drwxr-xr-x   - hadoop hadoop          0 2023-02-26 11:07 /user/root/spar_nord_bank_atm
[hadoop@ip-10-0-11-23 ~]$ hadoop fs -ls /user/root/spar_nord_bank_atm
Found 2 items
-rw-r--r--   1 hadoop hadoop          0 2023-02-26 11:07 /user/root/spar_nord_bank_atm/_SUCCESS
-rw-r--r--   1 hadoop hadoop  531214815 2023-02-26 11:07 /user/root/spar_nord_bank_atm/part-m-00000
```

In the screenshot, it can be seen that the target directory contains 2 items:
   i)   Success file indicating that MapReduce job was successful.
   ii)   The second part-m-00000 contains imported data. Only one part file has been
         created as only one mapper was used in sqoop command (i.e. -m 1).

On reading first 50 lines of the part file using the command below, it can be seen that the data
has been imported.

```
[hadoop@ip-10-0-11-23 ~]$ hadoop fs -cat /user/root/spar_nord_bank_atm/part-m-00000 | head -n 50
2017,January,1,Sunday,0,Active,1,NCR,NÃfÂ¦stved,Farimagsvej,8,4700,55.233,11.763,DKK,MasterCard,5643,Withdrawal,,,55.230,11.761,2616038,Naestved,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,NÃfÂ¸rresundby,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,NÃfÂ¸rresundby,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,3,NCR,Ikast,RÃfÂ¥dhusstrÃfÂ¦det,12,7430,56.139,9.154,DKK,VISA,4166,Withdrawal,,,56.139,9.158,2619426,Ikast,281.150,1011,100,6,240,0.000,75,300,Drizzle,light intensity drizzle
2017,January,1,Sunday,0,Active,4,NCR,Svogerslev,BrÃfÂ¸nsager,1,4000,55.634,12.018,DKK,MasterCard,5153,Withdrawal,,,55.642,12.080,2614481,Roskilde,280.610,1014,87,7,260,0.000,88,701,Mist,mist
2017,January,1,Sunday,0,Active,5,NCR,Nibe,Torvet,1,9240,56.983,9.639,DKK,MasterCard,3269,Withdrawal,,,56.981,9.639,2616483,Nibe,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,6,NCR,Fredericia,SjÃfÂ¦llandsgade,33,7000,55.564,9.757,DKK,MasterCard,887,Withdrawal,,,55.566,9.753,2621951,Fredericia,281.150,1014,93,7,230,0.290,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,7,Diebold Nixdorf,Hjallerup,Hjallerup Centret,18,9320,57.168,10.148,DKK,Mastercard - on-us,4626,Withdrawal,,,57.165,10.146,2620275,Hjallerup,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,8,NCR,GlyngÃfÂ¸re,FÃfÂ¦rgevej,1,7870,56.762,8.867,DKK,MasterCard,470,Withdrawal,,,56.793,8.853,2615964,Nykobing Mors,281.150,1011,100,6,240,0.000,75,300,Drizzle,light intensity drizzle
2017,January,1,Sunday,0,Active,9,Diebold Nixdorf,Hadsund,Storegade,12,9560,56.716,10.114,DKK,VISA,8473,Withdrawal,,,56.715,10.117,2620952,Hadsund,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,10,NCR,NÃfÂ¸rresundby,Torvet,6,9400,57.059,9.922,DKK,Dankort,953,Withdrawal,,,57.048,9.919,2624886,Aalborg,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,11,NCR,Sauersvej,Fridtjof Nansens Vej,2,9210,57.023,9.940,DKK,Visa Dankort,9346,Withdrawal,,,57.048,9.935,2616235,NÃfÂ¸rresundby,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,Mastercard - on-us,3874,Withdrawal,,,57.048,9.935,2616235,NÃfÂ¸rresundby,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,12,NCR,ÃfÆsterÃfÂ¥Â  Duus,ÃfÆsterÃfÂ¥Â¥,12,9000,57.049,9.922,DKK,Mastercard - on-us,1329,Withdrawal,,,57.048,9.919,2624886,Aalborg,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,13,NCR,SÃfÂ¦by,Vestergade,3,9300,57.334,10.515,DKK,Mastercard - on-us,5024,Withdrawal,,,57.441,10.537,2621927,Frederikshavn,281.140,1019,94,12,251,1.275,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,14,NCR,HÃfÂ¸rning,NÃfÂ¸rrealle,12,8362,56.086,10.037,DKK,Visa Dankort - on-us,1133,Withdrawal,,,56.157,10.211,2624652,Arhus,281.150,1012,87,5,250,0.000,92,300,Drizzle,light intensity drizzle
2017,January,1,Sunday,0,Active,15,NCR,Vestre,Kastetvej,36,9000,57.053,9.905,DKK,MasterCard,594,Withdrawal,,,57.048,9.919,2624886,Aalborg,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,12,NCR,ÃfÆsterÃfÂ¥Â  Duus,ÃfÆsterÃfÂ¥Â¥,12,9000,57.049,9.922,DKK,Mastercard - on-us,9570,Withdrawal,,,57.048,9.919,2624886,Aalborg,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,16,NCR,Skive,Adelgade,8,7800,56.567,9.027,DKK,Visa Dankort,6068,Withdrawal,,,56.567,9.033,2613731,Skive,281.150,1012,100,7,240,0.000,75,300,Drizzle,light intensity drizzle
2017,January,1,Sunday,0,Active,17,NCR,Randers,ÃfÆstervold,16,8900,56.462,10.038,DKK,VISA,1630,Withdrawal,,,56.715,10.117,2620952,Hadsund,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,17,NCR,Randers,ÃfÆstervold,16,8900,56.462,10.038,DKK,MasterCard,9948,Withdrawal,,,56.715,10.117,2620952,Hadsund,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,6,NCR,Fredericia,SjÃfÂ¦llandsgade,33,7000,55.564,9.757,DKK,MasterCard,4849,Withdrawal,,,55.566,9.753,2621951,Fredericia,281.150,1014,93,7,230,0.290,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,10,NCR,NÃfÂ¸rresundby,Torvet,6,9400,57.059,9.922,DKK,MasterCard,4400,Withdrawal,,,57.048,9.919,2624886,Aalborg,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,6,NCR,Fredericia,SjÃfÂ¦llandsgade,33,7000,55.564,9.757,DKK,VISA,7452,Withdrawal,,,55.566,9.753,2621951,Fredericia,281.150,1014,93,7,230,0.290,92,500,Rain,light rain
```