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.cyaielc9bmnf.us-east-
1.rds.amazonaws.com/testdatabase \
--table SRC_ATM_TRANS \
--username student --password STUDENT123 \
--target-dir /user/root/ SRC_ATM_TRANS \
-m 1
```

```
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=189595
                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)=1132128
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=23586
                Total vcore-milliseconds taken by all map tasks=23586
                Total megabyte-milliseconds taken by all map tasks=36228096
        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)=270
                CPU time spent (ms)=27470
                Physical memory (bytes) snapshot=633417728
                Virtual memory (bytes) snapshot=3300995072
                Total committed heap usage (bytes)=532676608
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=531214815
22/12/14 13:27:45 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 41.9952 seconds (12.0634 MB/sec)
22/12/14 13:27:45 INFO mapreduce.ImportJobBase: Retrieved 2468572 records
```

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/SRC_ATM_TRANS
```

```
[hadoop@ip-172-31-90-60 ~]$ hadoop fs -ls /user/root
Found 1 items
drwxr-xr-x   - hadoop hadoop          0 2022-12-14 13:27 /user/root/SRC_ATM_TRANS
[hadoop@ip-172-31-90-60 ~]$ hadoop fs -ls /user/root/SRC_ATM_TRANS
Found 2 items
-rw-r--r--   1 hadoop hadoop          0 2022-12-14 13:27 /user/root/SRC_ATM_TRANS/_SUCCESS
-rw-r--r--   1 hadoop hadoop  531214815 2022-12-14 13:27 /user/root/SRC_ATM_TRANS/part-m-00000
```

Command used to view the imported data in HDFS:

```
hadoop fs -cat /user/root/SRC_ATM_TRANS/part-m-00000
```

989,100,4,200,0.138,92,500,Rain,light rain
2017,December,31,Sunday,23,Active,24,NCR,Hobro,Adelgade,31,9500,56.638,9.794,DKK,Mastercard - on-us,7780,Withdrawal,,,56.643,9.790,2620167,Hobro,277.589,999,87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,20,NCR,Bispensgade,Bispensgade,35,9800,57.453,9.996,DKK,Mastercard - on-us,2830,Withdrawal,,,57.464,9.982,2620214,Hjorring,278.589,998,99,8,210,0.000,44,802,Clouds,scattered clouds
2017,December,31,Sunday,23,Active,28,NCR,LÃƒÂ¸gstÃƒÂ¸r,ÃƒËœsterbrogade,8,9670,56.967,9.253,DKK,VISA,812,Withdrawal,,,56.962,9.258,2617467,Logstor,277.589,999,87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,28,NCR,LÃƒÂ¸gstÃƒÂ¸r,ÃƒËœsterbrogade,8,9670,56.967,9.253,DKK,VISA,5568,Withdrawal,,,56.962,9.258,2617467,Logstor,277.589,999,87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,49,NCR,Bindslev,NÃƒÂ¸rrebro,18,9881,57.541,10.200,DKK,MasterCard,5403,Withdrawal,,,57.471,10.203,2614010,Sindal,277.589,999,87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,49,NCR,Bindslev,NÃƒÂ¸rrebro,18,9881,57.541,10.200,DKK,MasterCard,312,Withdrawal,,,57.471,10.203,2614010,Sindal,277.589,999,87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,56,Diebold Nixdorf,KÃƒÂ¸ge,SÃƒÂ¸ndre Alle,1,4600,55.454,12.181,DKK,Visa Dankort - on-us,5351,Withdrawal,,,55.588,12.251,2621218,Greve Kommune,280.150,992,100,3,190,0.000,90,804,Clouds,overcast clouds
2017,December,31,Sunday,23,Active,103,Diebold Nixdorf,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,Mastercard - on-us,5821,Withdrawal,,,57.048,9.935,2616235,NÃƒÂ¸rresundby,277.589,999,87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,1,NCR,NÃƒÂ¦stved,Farimagsvej,8,4700,55.233,11.763,DKK,Visa Dankort,3049,Withdrawal,,,55.230,11.761,2616038,Naestved,280.150,991,100,5,210,1.175,92,500,Rain,light rain
2017,December,31,Sunday,23,Active,85,Diebold Nixdorf,KÃƒÂ¸benhavn,Regnbuepladsen,5,1550,55.676,12.571,DKK,MasterCard,7785,Withdrawal,,,55.676,12.566,2618425,Copenhagen,280.150,992,100,3,190,0.000,90,804,Clouds,overcast clouds
2017,December,31,Sunday,23,Active,15,NCR,Vestre,Kastetvej,36,9000,57.053,9.905,DKK,Mastercard - on-us,9777,Withdrawal,,,57.048,9.919,2624886,Aalborg,277.589,999,87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,15,NCR,Vestre,Kastetvej,36,9000,57.053,9.905,DKK,MasterCard,5531,Withdrawal,,,57.048,9.919,2624886,Aalborg,277.589,999,87,6,208,0.000,76,803,Clouds,broken clouds