Data Ingestion from the RDS to HDFS using Sqoop

Sqoop Import command used for importing table from RDS to HDFS:

Before running the Sqoop import command, I checked that the target directory is not already created by with the following command:

hadoop fs -rm -r /user/root/SRC_ATM_TRANS

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/SRC_ATM_TRANS \

-m 1

After running the above Sqoop import command, I have retrieved 2468572 records that matched the validation document.. Below is the screenshot for the retrieved records using Sqoop import.

```
0/03 09:50:26 INFO mapreduce.Job: Job job_1664789117511_0001 running in uber mode : false
0/03 09:50:27 INFO mapreduce.Job:  map 0% reduce 0%
0/03 09:50:54 INFO mapreduce.Job:  map 100% reduce 0%
0/03 09:50:54 INFO mapreduce.Job: Job job_1664789117511_0001 completed successfully
0/03 09:50:54 INFO mapreduce.Job: Counters: 30
    File System Counters
            FILE: Number of bytes read=0
            FILE: Number of bytes written=189240
            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)=1187280
            Total time spent by all reduces in occupied slots (ms)=0
            Total time spent by all map tasks (ms)=24735
            Total vcore-milliseconds taken by all map tasks=24735
            Total megabyte-milliseconds taken by all map tasks=37992960
    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)=295
            CPU time spent (ms)=28340
            Physical memory (bytes) snapshot=633749504
            Virtual memory (bytes) snapshot=3311742976
            Total committed heap usage (bytes)=541065216
    File Input Format Counters
            Bytes Read=0
    File Output Format Counters
            Bytes Written=531214815
0/03 09:50:54 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 46.7782 seconds (10.83 MB/sec)
0/03 09:50:54 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
```

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

hadoop fs -ls /user/root/SRC_ATM_TRANS

After running the above command, I can see that the target directory contains 2 items:

- The first file is the success file, indicating that the MapReduce job was successful.
- The second file 'part-m-00000' is the one with all of the data I imported. Since I used only one mapper in import command thus the data is in a single file.

```
[root@ip-10-0-8-225 ~]# hadoop fs -ls /user/root/SRC_ATM_TRANS
Found 2 items
-rw-r--r--    1 root hadoop           0 2022-10-03 09:50 /user/root/SRC_ATM_TRANS/_SUCCESS
-rw-r--r--    1 root hadoop   531214815 2022-10-03 09:50 /user/root/SRC_ATM_TRANS/part-m-00000
```

Screenshot of the imported data:

By using the below command I opened the 'part-m-00000' file. I can see all of the data that has been imported:

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

```
,76,500,Rain,light rain
2017,February,16,Thursday,10,Active,77,NCR,Aarhus Lufthavn,Ny Lufthavnsvej,24,8560,56.308,10.627,DKK,VISA,8137,Withdrawal,,,56.361,10.584,2618515,Kolinc

2017,February,16,Thursday,10,Active,42,NCR,Vinderup,SÃfÂ ndergade,5,7830,56.481,8.779,DKK,Mastercard - on-us,1558,Withdrawal,,,56.475,8.791,2610196,Vinc
00,Rain,light rain
2017,February,16,Thursday,10,Inactive,90,NCR,Intern HolbÃfÂ¦k,Slotsvolden,7,4300,55.718,11.704,DKK,Visa Dankort - on-us,6121,Withdrawal,,,55.700,11.700,
1,98,5,233,0.000,68,803,Clouds,broken clouds
2017,February,16,Thursday,10,Active,87,NCR,Aalborg Storcenter indg. D,Hobrovej,452,9200,57.005,9.876,DKK,VISA,5931,Withdrawal,,,56.972,9.848,2612021,Sve
,500,Rain,light rain
2017,February,16,Thursday,10,Active,60,NCR,Hirtshals,JÃfÂ rgen Fibigersgade,4,9850,57.591,9.957,DKK,Dankort - on-us,1294,Withdrawal,,,57.588,9.959,26202
270,76,500,Rain,light rain
2017,February,16,Thursday,10,Active,35,NCR,Aabybro,ÃfËœstergade,6,9440,57.162,9.730,DKK,Visa Dankort - on-us,8678,Withdrawal,,,57.158,9.737,2625037,Abyl
,Rain,light rain
2017,February,16,Thursday,10,Active,45,NCR,Abildgaard,HjÃfÂ rringvej,144,9900,57.447,10.506,DKK,Visa Dankort - on-us,9505,Withdrawal,,,57.441,10.537,262
209,0.225,76,500,Rain,light rain
2017,February,16,Thursday,10,Active,79,NCR,Middelfart,Brogade,9,5500,55.507,9.727,DKK,Mastercard - on-us,1494,Withdrawal,,,55.506,9.731,2616933,Middelfa
,Rain,light rain
2017,February,16,Thursday,10,Active,20,NCR,Bispensgade,Bispensgade,35,9800,57.453,9.996,DKK,Dankort - on-us,8954,Withdrawal,,,57.464,9.982,2620214,Hjor
0,Rain,light rain
2017,February,16,Thursday,10,Inactive,90,NCR,Intern HolbÃfÂ¦k,Slotsvolden,7,4300,55.718,11.704,DKK,MasterCard,287,Withdrawal,,,55.700,11.700,2620141,Hol
0.000,68,803,Clouds,broken clouds
2017,February,16,Thursday,10,Active,11,NCR,Sauersvej,Fridtjof Nansens Vej,2,9210,57.023,9.940,DKK,Dankort - on-us,4573,Withdrawal,,,57.048,9.935,261623
,0.270,76,500,Rain,light rain
2017,February,16,Thursday,10,Inactive,93,NCR,Intern  BrÃfÂ nderslev,Algade,4,9700,57.269,9.945,DKK,Visa Dankort - on-us,2005,Withdrawal,,,57.118,9.857,2
270,76,500,Rain,light rain
2017,February,16,Thursday,10,Active,104,NCR,Intern  ÃfËœsterÃfÂ¥,ÃfËœsterÃfÂ¥,12,9000,57.049,9.922,DKK,Dankort - on-us,6306,Withdrawal,,,57.048,9.919,2€
.270,76,500,Rain,light rain
2017,February,16,Thursday,10,Active,35,NCR,Aabybro,ÃfËœstergade,6,9440,57.162,9.730,DKK,Mastercard - on-us,5515,Withdrawal,,,57.158,9.737,2625037,Abybro
ain,light rain
2017,February,16,Thursday,10,Active,57,NCR,HillerÃfÂ d,KÃfÂ benhavnsvej,31,3400,55.933,12.314,DKK,MasterCard,8655,Withdrawal,,,55.767,12.433,2621394,Gla
0.000,68,803,Clouds,broken clouds
2017,February,16,Thursday,10,Active,55,NCR,Arden,Vestergade,6,9510,56.769,9.859,DKK,Visa Dankort,6732,Withdrawal,,,56.885,9.838,2612500,Stovring,276.864
t rain
2017,February,16,Thursday,10,Active,34,NCR,Skipperen,Vestre Alle,2,9000,57.034,9.908,DKK,Visa Dankort,7697,Withdrawal,,,57.048,9.919,2624886,Aalborg,27€
light rain
2017,February,16,Thursday,10,Active,20,NCR,Bispensgade,Bispensgade,35,9800,57.453,9.996,DKK,HÃfÂ¦vekort - on-us,8285,Withdrawal,,,57.464,9.982,2620214,E
6,500,Rain,light rain
2017,February,16,Thursday,10,Active,59,Diebold Nixdorf,NykÃfÂ bing Mors Lobby,Kirketorvet,1,7900,56.795,8.860,EUR,Mastercard - on-us,8336,Withdrawal,,,
.914,1031,90,8,229,0.695,92,500,Rain,light rain
2017,February,16,Thursday,10,Inactive,81,NCR,Spar KÃfÂ bmand TornhÃfÂ j,TornhÃfÂ jvej,4,9220,57.026,10.002,DKK,Visa Dankort,6826,Withdrawal,,,56.994,9.9
217,0.270,76,500,Rain,light rain
2017,February,16,Thursday,10,Active,60,NCR,Hirtshals,JÃfÂ rgen Fibigersgade,4,9850,57.591,9.957,DKK,Visa Dankort - on-us,9123,Withdrawal,,,57.588,9.959,
17,0.270,76,500,Rain,light rain
2017,February,16,Thursday,10,Active,62,Diebold Nixdorf,Terndrup,Bymidten,2,9575,56.815,10.057,DKK,Visa Dankort - on-us,5135,Withdrawal,,,56.715,10.117,2
```