

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 made sure the target directory is not already created. Otherwise, the Sqoop import command would throw an error:

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

Sqoop Import Command:

```
sqoop import \  
--connect jdbc:mysql://upgraddetest.cyaie1c9bmnf.us-east-1.rds.amazonaws.com/testdatabase \  
--table SRC_ATM_TRANS \  
--username student --password STUDENT123 \  
--target-dir /user/root/SRC_ATM_TRANS \  
-m 1
```

In the screenshot below, I can see that as a result of **Sqoop Import Job, 2468572 records have been retrieved** (same as the checkpoint mentioned in the Validation document):

```
hadoop@ip-172-31-81-150:~$  
2024-05-02 06:23:10,208 INFO mapreduce.Job: Job job_1714629725995_0001 running i  
n uber mode : false  
2024-05-02 06:23:10,210 INFO mapreduce.Job: map 0% reduce 0%  
2024-05-02 06:23:41,556 INFO mapreduce.Job: map 100% reduce 0%  
2024-05-02 06:23:42,569 INFO mapreduce.Job: Job job_1714629725995_0001 completed  
successfully  
2024-05-02 06:23:42,687 INFO mapreduce.Job: Counters: 33  
File System Counters  
FILE: Number of bytes read=0  
FILE: Number of bytes written=297853  
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)=1403808  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=29246  
Total vcore-milliseconds taken by all map tasks=29246  
Total megabyte-milliseconds taken by all map tasks=44921856  
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)=264  
CPU time spent (ms)=33120  
Physical memory (bytes) snapshot=630841344  
Virtual memory (bytes) snapshot=3085635584  
Total committed heap usage (bytes)=600309760  
Peak Map Physical memory (bytes)=630841344  
Peak Map Virtual memory (bytes)=3085635584  
File Input Format Counters  
Bytes Read=0  
File Output Format Counters  
Bytes Written=531214815  
2024-05-02 06:23:42,693 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in  
54.1102 seconds (9.3625 MB/sec)  
2024-05-02 06:23:42,696 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.  
[hadoop@ip-172-31-81-150 ~]$
```

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

```
hadoop fs -ls /user/root/SRC_ATM_TRANS
```

In the screenshot below, 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 my import command thus the data is in a single file.

```
hadoop@ip-172-31-81-150:~
[hadoo@ip-172-31-81-150 ~]$ hadoop fs -ls /user/root/SRC_ATM_TRANS
Found 2 items
-rw-r--r-- 1 hadoop hdfsadmin group 0 2024-05-02 06:23 /user/root/SRC_ATM_TRANS/_SUCCESS
-rw-r--r-- 1 hadoop hdfsadmin group 531214815 2024-05-02 06:23 /user/root/SRC_ATM_TRANS/part-m-00000
[hadoo@ip-172-31-81-150 ~]$
```

When I open the 'part-m-00000' file using the following command, I can see all of the data that has been imported:

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

Screenshot of the imported data:

```
hadoop@ip-172-31-81-150:~
,288.150,1007,77,8,280,0.000,40,802,Clouds,scattered clouds
2017,May,31,Wednesday,11,Active,57,NCR,Aalborg Storcenter indg. D,Hobrovej,452,9200,57.005,9.876,DKK,Visa Dankort,5167,Withdrawal,,,56.972,9.848,2612021,Sven
strup,288.150,1007,77,8,280,0.000,40,802,Clouds,scattered clouds
2017,May,31,Wednesday,11,Inactive,2,NCR,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,Mastercard - on-us,9521,Withdrawal,,,57.048,9.935,2616235,NÅfÅ,rresundby
,288.150,1007,77,8,280,0.000,40,802,Clouds,scattered clouds
2017,May,31,Wednesday,11,Active,1,NCR,NÅfÅ,istved,Farimagvej,8,4700,55.233,11.763,DKK,Visa Dankort - on-us,2101,Withdrawal,,,55.230,11.761,2616038,Naestved,2
88.150,1012,62,12,280,0.000,0,800,Clear,Sky is Clear
2017,May,31,Wednesday,11,Active,67,NCR,HÅfÅ,jalev,ÅfEøsterisvej,2,7840,56.551,9.110,DKK,VISA,1945,Withdrawal,,,56.586,9.136,2619908,Hojalev,288.150,1007,77,
8,280,0.000,40,802,Clouds,scattered clouds
2017,May,31,Wednesday,11,Active,28,NCR,LÅfÅ,gstÅfÅ,r,ÅfEøsterbrogade,8,9670,56.967,9.253,DKK,Visa Dankort - on-us,3422,Withdrawal,,,56.962,9.258,2617467,Logs
tor,288.150,1007,77,8,280,0.000,40,802,Clouds,scattered clouds
2017,May,31,Wednesday,11,Inactive,102,NCR,Aalborg Storcenter Afd,Hobrovej,452,9200,57.005,9.876,DKK,MasterCard,1274,Withdrawal,,,56.972,9.848,2612021,Svenst
rup,288.150,1007,77,8,280,0.000,40,802,Clouds,scattered clouds
2017,May,31,Wednesday,11,Inactive,90,NCR,Intern HolbÅfÅ,k,Slotsvolden,7,4300,55.718,11.704,DKK,Visa Dankort - on-us,6102,Withdrawal,,,55.700,11.700,2620141,H
olbÅfÅ,k Kommune,287.890,1012,62,11,280,0.000,75,803,Clouds,broken clouds
2017,May,31,Wednesday,11,Active,9,Diebold Nixdorf,Hadsund,Storegade,12,9560,56.716,10.114,DKK,Visa Dankort - on-us,6691,Withdrawal,,,56.715,10.117,2620952,H
adsund,288.150,1007,77,8,280,0.000,40,802,Clouds,scattered clouds
2017,May,31,Wednesday,11,Active,21,NCR,ÅfEøsterÅfÅV MÅfÅ,l1er,ÅfEøsterÅfÅV,12,9000,57.049,9.922,DKK,Visa Dankort,5147,Withdrawal,,,57.048,9.919,2624886,Aalb
org,288.150,1007,77,8,280,0.000,40,802,Clouds,scattered clouds
2017,May,31,Wednesday,11,Active,68,NCR,Drive Up Vejle,Havneparken,4,7100,55.707,9.541,DKK,VISA,5662,Withdrawal,,,55.709,9.536,2610613,Vejle,286.390,1015,71,
11,280,0.000,0,800,Clear,Sky is Clear
2017,May,31,Wednesday,11,Active,103,Diebold Nixdorf,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,Visa Dankort - on-us,2872,Withdrawal,,,57.048,9.935,2616235,
NÅfÅ,rresundby,288.150,1007,77,8,280,0.000,40,802,Clouds,scattered clouds
2017,May,31,Wednesday,11,Active,92,NCR,Intern HÅfÅ,r,ring,ÅfEøstergade,8,9800,57.459,9.988,DKK,HÅfÅ,vekort - on-us,7370,Withdrawal,,,57.464,9.982,2620214,Hj
orring,288.150,1007,77,8,280,0.000,40,802,Clouds,scattered clouds
2017,May,31,Wednesday,11,Active,17,NCR,Randers,ÅfEøstervold,16,8900,56.462,10.038,DKK,Mastercard - on-us,8237,Withdrawal,,,56.715,10.117,2620952,Hadsund,288.
150,1007,77,8,280,0.000,40,802,Clouds,scattered clouds
2017,May,31,Wednesday,11,Active,80,NCR,Menu KÅfÅ,bmand Klarup,Klarupvej,52,9270,57.013,10.046,DKK,Dankort - on-us,9074,Withdrawal,,,57.004,10.101,2612301,Sto
rvoerde,288.150,1007,77,8,280,0.000,40,802,Clouds,scattered clouds
2017,May,31,Wednesday,11,Inactive,81,NCR,Spar KÅfÅ,bmand TornhÅfÅ,j,TornhÅfÅ,jvej,4,9220,57.026,10.002,DKK,HÅfÅ,vekort - on-us,660,Withdrawal,,,56.994,9.991,
2621449,Gistrup,288.150,1007,77,8,280,0.000,40,802,Clouds,scattered clouds
2017,May,31,Wednesday,11,Active,37,NCR,Silkeborg,Borgergade,36,8600,56.179,9.552,DKK,Visa Dankort - on-us,6076,Withdrawal,,,56.170,9.545,2614030,Silkeborg,28
5.810,1014,82,12,280,0.000,75,803,Clouds,broken clouds
2017,May,31,Wednesday,11,Active,61,NCR,Esbjerg,Strandbygade,20,6700,55.468,8.440,EUR,Visa Dankort - on-us,8402,Withdrawal,,,55.467,8.450,2622447,Esbjerg,286.
490,1017,66,10,290,0.000,12,801,Clouds,few clouds
2017,May,31,Wednesday,11,Inactive,94,NCR,Intern Skagen,Sct. Laurentilvej,36,9990,57.723,10.590,DKK,CIRRUS,3781,Withdrawal,,,57.721,10.584,2613939,Skagen,288.
850,1006,82,5,250,0.000,75,803,Clouds,broken clouds
2017,May,31,Wednesday,11,Active,25,Diebold Nixdorf,Odense,FÅfÅ,lledvej,3,5000,55.394,10.370,DKK,Mastercard - on-us,3699,Withdrawal,,,55.396,10.388,2615876,Od
ense,287.660,1015,62,11,270,0.000,75,803,Clouds,broken clouds
2017,May,31,Wednesday,11,Active,22,NCR,Roskilde,KÅfÅ,benhavnsvej,65,4000,55.642,12.106,DKK,Mastercard - on-us,7150,Withdrawal,4002,Suspected malfunction,55.6
42,12.080,2614481,Roskilde,288.150,1012,62,12,280,0.000,0,800,Clear,Sky is Clear
2017,May,31,Wednesday,11,Active,19,NCR,Kolding,Vejlevej,135,6000,55.505,9.457,DKK,Mastercard - on-us,4654,Withdrawal,,,55.490,9.472,2618528,Kolding,286.670,1
017,71,11,290,0.000,75,803,Clouds,broken clouds
2017,May,31,Wednesday,11,Inactive,29,NCR,Skelagervej 15,Skelagervej,15,9000,57.023,9.891,DKK,MasterCard,1822,Withdrawal,,,57.048,9.919,2624886,Aalborg,288.15
0,1007,77,8,280,0.000,40,802,Clouds,scattered clouds
2017,May,31,Wednesday,11,Active,43,NCR,Aars,Himmerlandsgade,70,9600,56.803,9.518,EUR,Visa Dankort - on-us,5147,Withdrawal,,,56.804,9.514,2624602,Aars,286.150
,1014,82,12,280,0.000,75,803,Clouds,broken clouds
2017,May,31,Wednesday,11,Active,40,Diebold Nixdorf,Frederikshavn,Danmarksgade,48,9900,57.441,10.537,DKK,Visa Dankort - on-us,267,Withdrawal,,,57.441,10.537,2
621927,Frederikshavn,288.150,1007,77,8,280,0.000,40,802,Clouds,scattered clouds
```