

## Data Ingestion from the RDS to HDFS using Sqoop

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

Command :

```
sqoop import \
--connect jdbc:mysql://upgraddetest.cyaiele9bmnf.us-east-1.rds.amazonaws.com/testdatabase \
--table SRC_ATM_TRANS \
--username student --password STUDENT123 \
--target-dir /user/root/SRC_ATM_TRANS \
-m 1
```

```
43640888181_0001
22/01/31 15:09:46 INFO impl.YarnClientImpl: Submitted application application_16
43640888181_0001
22/01/31 15:09:46 INFO mapreduce.Job: The url to track the job: http://ip-172-31-
21-158.ec2.internal:20888/proxy/application_1643640888181_0001/
22/01/31 15:09:46 INFO mapreduce.Job: Running job: job_1643640888181_0001
22/01/31 15:09:55 INFO mapreduce.Job: Job job_1643640888181_0001 running in uber
mode : false
22/01/31 15:09:55 INFO mapreduce.Job: map 0% reduce 0%
22/01/31 15:10:26 INFO mapreduce.Job: map 100% reduce 0%
22/01/31 15:10:26 INFO mapreduce.Job: Job job_1643640888181_0001 completed succe
ssfully
22/01/31 15:10:26 INFO mapreduce.Job: Counters: 30
File System Counters
FILE: Number of bytes read=0
FILE: Number of bytes written=189435
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)=1370304
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=28548
Total vcore-milliseconds taken by all map tasks=28548
Total megabyte-milliseconds taken by all map tasks=43849728
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)=322
CPU time spent (ms)=28420
Physical memory (bytes) snapshot=606138368
Virtual memory (bytes) snapshot=3290247168
Total committed heap usage (bytes)=538443776
File Input Format Counters
Bytes Read=0
File Output Format Counters
Bytes Written=531214815
22/01/31 15:10:26 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 49.92
26 seconds (10.1478 MB/sec)
22/01/31 15:10:26 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-172-31-21-158 ~]#
```

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

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

```
[root@ip-172-31-21-158 ~]# hadoop fs -ls /user/root/SRC_ATM_TRANS
Found 2 items
-rw-r--r--  1 root hadoop          0 2022-01-31 15:10 /user/root/SRC_ATM_TRANS/_SUCCESS
-rw-r--r--  1 root hadoop 531214815 2022-01-31 15:10 /user/root/SRC_ATM_TRANS/part-m-00000
[root@ip-172-31-21-158 ~]#
```

As per the above screenshot, I can see that target directory contains two items:

1. The first file is the success file, indicating that the MapReduce job was successful.
2. The second file 'part-m-00000' is the one with all of the data imported. Since in my import command i used only 1 mapper thus the data is in a single file.

### Screenshot of the imported data:

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

```
3,190,0.000,90,701,Mist,mist
2017,December,31,Sunday,23,Active,46,Diebold Nixdorf,HelsingÅr,r,Sct. Olai Gade,39,3000,56.036,12.612,DKK,Mastercard - on-us,2384,Withdrawal,,,56.036,12.614,2620473,Helsingør,279.900,989,100,4,200,0.138,92,500,Rain,light rain
2017,December,31,Sunday,23,Active,64,NCR,Horsens,GrÅfÅ,landsvej,5,8700,55.859,9.854,DKK,Visa Dankort - on-us,2475,Withdrawal,,,55.861,9.850,2619771,Horsens,279.900,989,100,4,200,0.138,92,500,Rain,light rain
2017,December,31,Sunday,23,Active,64,NCR,Horsens,GrÅfÅ,landsvej,5,8700,55.859,9.854,DKK,Visa Dankort - on-us,5223,Withdrawal,,,55.861,9.850,2619771,Horsens,279.900,989,100,4,200,0.138,92,500,Rain,light rain
2017,December,31,Sunday,23,Active,64,NCR,Horsens,GrÅfÅ,landsvej,5,8700,55.859,9.854,DKK,Visa Dankort,8690,Withdrawal,,,55.861,9.850,2619771,Horsens,279.900,989,100,4,200,0.138,92,500,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,278.589,999,99,8,210,0.000,44,802,Clouds,scattered clouds
2017,December,31,Sunday,23,Active,28,NCR,LÅfÅ,gstÅfÅ,r,ÅfEasterbrogade,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,LÅfÅ,gstÅfÅ,r,ÅfEasterbrogade,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,NÅfÅ,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,NÅfÅ,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,KÅfÅ,ge,SÅfÅ,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,NÅfÅ,rresundby,277.589,999,87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,1,NCR,NÅfÅ,stved,Farimagvej,4,4700,55.233,11.763,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,KÅfÅ,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
2017,December,31,Sunday,23,Active,34,NCR,Skipperen,Vestre Alle,2,9000,57.034,9.908,DKK,MasterCard,8323,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,20,NCR,Bispensgade,Bispensgade,35,9800,57.453,9.996,DKK,MasterCard,5023,Withdrawal,,,57.464,9.982,2620214,Hjorring,278.589,999,99,8,210,0.000,44,802,Clouds,scattered clouds
2017,December,31,Sunday,23,Active,34,NCR,Skipperen,Vestre Alle,2,9000,57.034,9.908,DKK,MasterCard,147,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,70,Diebold Nixdorf,Holstebro,Hostrupsvej,6,7500,56.373,8.625,DKK,MasterCard,5666,Withdrawal,,,56.360,8.616,2620046,Holstebro,280.150,988,93,4,210,0.000,92,804,Clouds,overcast clouds
2017,December,31,Sunday,23,Active,49,NCR,Bindslev,NÅfÅ,rrebro,18,9881,57.541,10.200,DKK,MasterCard,7886,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,Inactive,12,NCR,ÅfEasterÅfÅ, Duus,ÅfEasterÅfÅ, 12,9000,57.049,9.922,DKK,Mastercard - on-us,2436,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,Inactive,12,NCR,ÅfEasterÅfÅ, Duus,ÅfEasterÅfÅ, 12,9000,57.049,9.922,DKK,Mastercard - on-us,8519,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,10,NCR,NÅfÅ,rresundby,Torvet,6,9400,57.059,9.922,DKK,Mastercard - on-us,5286,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,37,NCR,Silkeborg,Borgergade,36,8600,56.179,9.552,DKK,VISA,876,Withdrawal,,,56.170,9.545,2614030,Silkeborg,279.800,988,93,4,210,0.000,88,804,Clouds,overcast clouds
root@in-172-31-21-158 ~#
```