

Data Ingestion from RDS to HDFS using Sqoop

After creating an EMR cluster connect to it through putty and then install the MySQL driver in it and then run the below command so that data from RDS is imported to the table SRC_ATM_TRANS in the HDFS.

```
[hadoop@ip-172-31-4-83 mysql-connector-java-8.0.25]$ sudo service mariadb restart
Redirecting to /bin/systemctl restart mariadb.service
[hadoop@ip-172-31-4-83 mysql-connector-java-8.0.25]$ sqoop import \
> --connect jdbc:mysql://upgraddetest.cyaiehc9bmnf.us-east-1.rds.amazonaws.com/testdatabase \
> --table SRC_ATM_TRANS \
> --username student --password STUDENT123 \
> --target-dir /user/root/SRC_ATM_TRANS \
> -m 1
```

```
23/06/12 11:15:20 INFO mapreduce.Job: Running job: job_1686567947348_0001
23/06/12 11:15:30 INFO mapreduce.Job: Job job_1686567947348_0001 running in uber mode : false
23/06/12 11:15:30 INFO mapreduce.Job: map 0% reduce 0%
23/06/12 11:15:59 INFO mapreduce.Job: map 100% reduce 0%
23/06/12 11:15:59 INFO mapreduce.Job: Job job_1686567947348_0001 completed successfully
23/06/12 11:15:59 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=189563
    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)=1270224
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=26463
    Total vcore-milliseconds taken by all map tasks=26463
    Total megabyte-milliseconds taken by all map tasks=40647168
  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)=213
    CPU time spent (ms)=28920
    Physical memory (bytes) snapshot=610881536
    Virtual memory (bytes) snapshot=3295088640
    Total committed heap usage (bytes)=536870912
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=531214815
23/06/12 11:15:59 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 48.8665 seconds (10.3671 MB/sec)
23/06/12 11:15:59 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[hadoop@ip-172-31-4-83 mysql-connector-java-8.0.25]$
```

```
[hadoop@ip-172-31-4-83 mysql-connector-java-8.0.25]$ hadoop fs -ls /user/root/SRC_ATM_TRANS
Found 2 items
-rw-r--r--  1 hadoop hadoop          0 2023-06-12 11:15 /user/root/SRC_ATM_TRANS/_SUCCESS
-rw-r--r--  1 hadoop hadoop 531214815 2023-06-12 11:15 /user/root/SRC_ATM_TRANS/part-m-00000
```

After running the above command, we 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 the data we imported. Since we used only one mapper in our import command thus the data is stored in a single file.

```
[hadoop@ip-172-31-4-83 mysql-connector-java-8.0.25]$ hadoop fs -cat /user/root/SRC ATM TRANS/part-m-00000
```

By the above command we can see the data present in the part-m-00000 file.

```
2017,January,13,Friday,11,Active,85,Diebold Nixdorf,K&A,benhavn,Ragnbuepladsen,5,1550,55.676,12.571,DKK,Mastercard - on-us,5157,Withdrawal,,,55.676,12.566,2
618425,Copenhagen,274.130,990,100,1,350,0.000,75,701,Mist,mist
2017,January,13,Friday,11,Active,39,NCR,Svenstrup,Godth&Afbavej,14,9230,56.973,9.851,DKK,Dankort - on-us,6622,Withdrawal,,,56.972,9.848,2612021,Svenstrup,27
3.150,988,88,3,180,0.000,0,800,Clear,Sky is Clear
2017,January,13,Friday,11,Active,79,NCR,Middelfart,Brogade,9,5500,55.507,9.727,DKK,Visa Dankort,7638,Withdrawal,,,55.506,9.731,2616933,Middelfart,274.330,991
,93,2,350,0.000,90,615,Snow,light rain and snow
2017,January,13,Friday,11,Active,24,NCR,Hobro,Adelgade,31,9500,56.638,9.794,DKK,Mastercard - on-us,8325,Withdrawal,,,56.643,9.790,2620167,Hobro,273.150,988,8
8,3,180,0.000,0,800,Clear,Sky is Clear
2017,January,13,Friday,11,Inactive,73,NCR,H&A,jbjerg,Rosenvangsalle,194,8270,56.119,10.192,DKK,Visa Dankort,8710,Withdrawal,,,56.157,10.211,2624652,Arhus,27
4.150,990,100,1,270,0.000,0,800,Clear,Sky is Clear
2017,January,13,Friday,11,Inactive,89,NCR,Intern Frederikshavn,Danmarksgade,48,9900,57.441,10.537,DKK,Mastercard - on-us,8576,Withdrawal,,,57.441,10.537,262
1927,Frederikshavn,273.150,988,100,3,180,0.000,0,800,Clear,Sky is Clear
2017,January,13,Friday,11,Inactive,50,NCR,Aarhus,S&A,nder Alle,11,8000,56.153,10.206,DKK,Visa Dankort - on-us,6749,Withdrawal,,,56.157,10.211,2624652,Arhus,
274.150,990,100,1,270,0.000,0,800,Clear,Sky is Clear
2017,January,13,Friday,11,Inactive,16,NCR,Skive,Adelgade,8,7800,56.567,9.027,DKK,Mastercard - on-us,8017,Withdrawal,,,56.567,9.033,2613731,Skive,274.150,989,
92,2,200,0.000,0,800,Clear,Sky is Clear
2017,January,13,Friday,11,Inactive,102,NCR,Aalborg Storcenter Afd,Hobrovej,452,9200,57.005,9.876,DKK,MasterCard,9837,Withdrawal,,,56.972,9.848,2612021,Svens
strup,273.150,988,88,3,180,0.000,0,800,Clear,Sky is Clear
2017,January,13,Friday,11,Active,100,NCR,Intern Skive,Adelgade,8,7800,56.567,9.027,DKK,Dankort - on-us,5133,Withdrawal,,,56.567,9.033,2613731,Skive,274.150,
989,92,2,200,0.000,0,800,Clear,Sky is Clear
2017,January,13,Friday,11,Active,69,NCR,Taars,Bredgade,91,9830,57.385,10.116,DKK,MasterCard,6157,Withdrawal,,,57.383,10.117,2611852,Tars,273.150,988,88,3,180
,0.000,0,800,Clear,Sky is Clear
2017,January,13,Friday,11,Active,105,Diebold Nixdorf,N&A,rresundby,Torvet,6,9400,57.059,9.922,DKK,Dankort - on-us,5384,Withdrawal,,,57.048,9.919,2624886,Aal
borg,273.150,988,88,3,180,0.000,0,800,Clear,Sky is Clear
2017,January,13,Friday,11,Active,28,NCR,L&A,gst&A,r,Af&sterbrogade,8,9670,56.967,9.253,DKK,Dankort - on-us,8451,Withdrawal,,,56.962,9.258,2617467,Logstor,
273.150,988,88,3,180,0.000,0,800,Clear,Sky is Clear
2017,January,13,Friday,11,Inactive,30,NCR,Nyk&A,bing Mors,Kirketorvet,1,7900,56.795,8.860,DKK,Mastercard - on-us,6102,Withdrawal,,,56.793,8.853,2615964,Nyko
bing Mors,275.150,990,93,3,210,0.000,20,801,Clouds,few clouds
2017,January,13,Friday,11,Active,23,Diebold Nixdorf,Vodskov,Vodskovvej,27,9310,57.104,10.027,DKK,Visa Dankort - on-us,5157,Withdrawal,,,57.109,10.022,2610095
,Vodskov,274.150,989,86,3,190,0.000,20,801,Clouds,few clouds
2017,January,13,Friday,11,Active,104,NCR,Intern Af&ster&A,AV,Af&ster&A,AV,12,9000,57.049,9.922,DKK,Mastercard - on-us,7478,Withdrawal,,,57.048,9.919,2624886
,Aalborg,273.150,988,88,3,180,0.000,0,800,Clear,Sky is Clear
2017,January,13,Friday,11,Inactive,14,NCR,H&A,rning,N&A,rrealle,12,8362,56.086,10.037,DKK,Visa Dankort - on-us,1918,Withdrawal,,,56.157,10.211,2624652,Arhu
s,274.150,990,100,1,270,0.000,0,800,Clear,Sky is Clear
2017,January,13,Friday,11,Active,60,NCR,Hirtshals,J&A,r gen Fibigersgade,4,9850,57.591,9.957,DKK,Mastercard - on-us,7157,Withdrawal,,,57.588,9.959,2620279,Hi
rtshals,273.150,988,88,3,180,0.000,0,800,Clear,Sky is Clear
2017,January,13,Friday,11,Active,9,Diebold Nixdorf,Hadsund,Storegade,12,9560,56.716,10.114,DKK,Visa Dankort - on-us,32,Withdrawal,,,56.715,10.117,2620952,Had
sund,273.150,988,88,3,180,0.000,0,800,Clear,Sky is Clear
```