

Data Ingestion from the RDS to HDFS using Sqoop

Prerequisite for executing the sqoop command and establishing a connection to RDS:

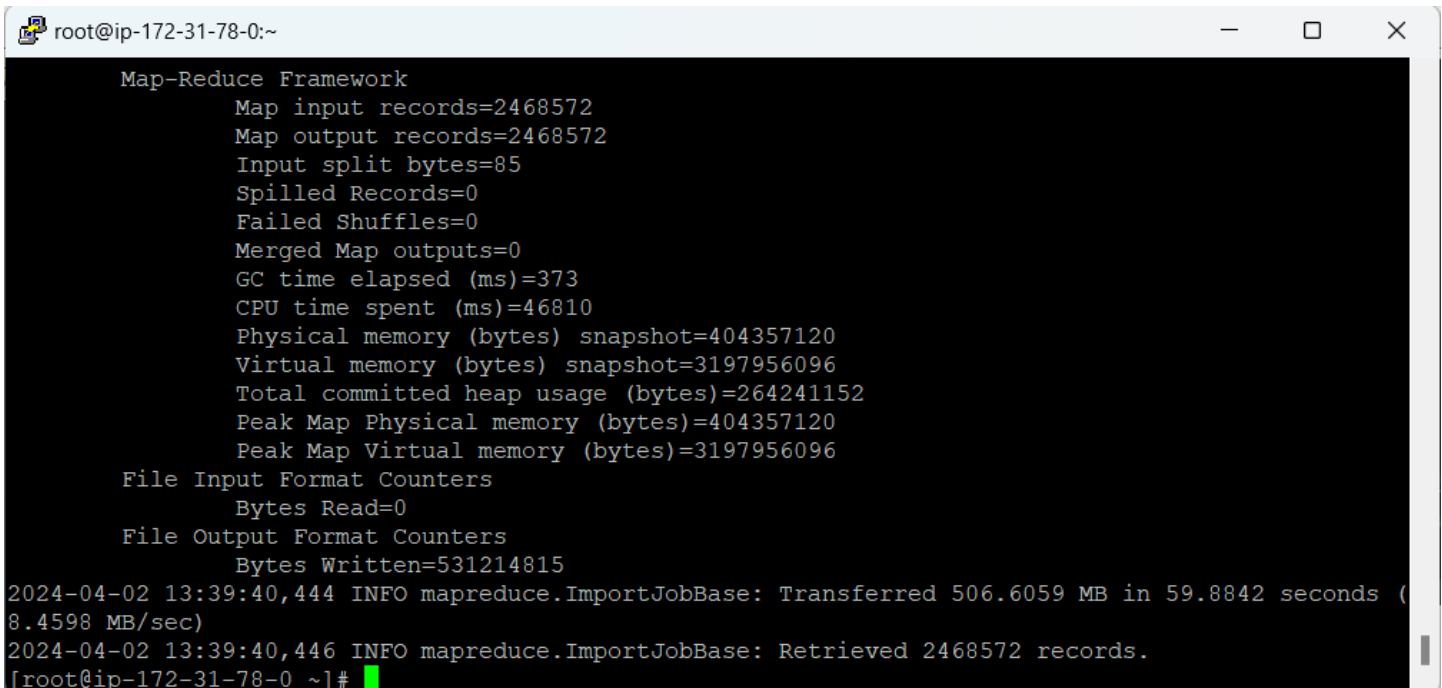
While logged in as the root user, access the EMR instance and fetch the MySQL connector.

```
sudo -i  
  
wget https://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz  
  
tar -xvf mysql-connector-java-8.0.25.tar.gz  
  
cd mysql-connector-java-8.0.25/  
  
sudo cp mysql-connector-java-8.0.25.jar /usr/lib/sqoop/lib/  
  
sudo -i
```

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

Command:

```
sqoop import --connect jdbc:mysql://upgraddetest.cyaieic9bmnf.us-east-1.rds.amazonaws.com/testdatabase --username student --password STUDENT123 --table SRC_ATM_TRANS --target-dir /user/hadoop/ETL/ATMDATA -m 1
```



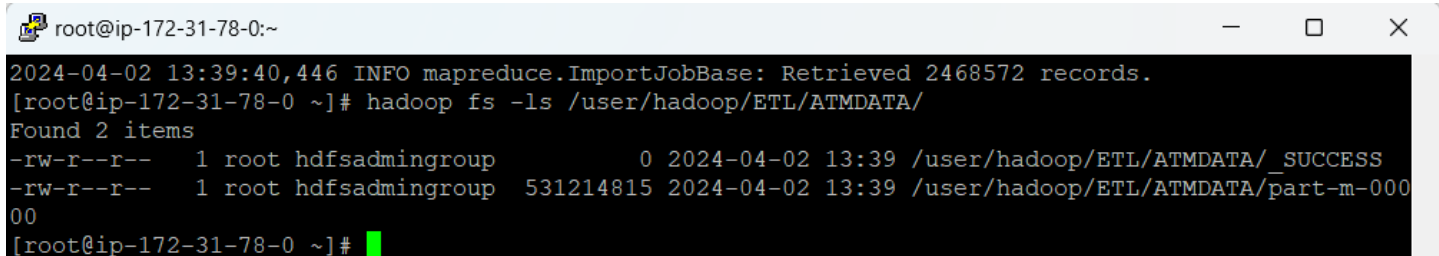
```
root@ip-172-31-78-0:~  
  
Map-Reduce Framework  
  Map input records=2468572  
  Map output records=2468572  
  Input split bytes=85  
  Spilled Records=0  
  Failed Shuffles=0  
  Merged Map outputs=0  
  GC time elapsed (ms)=373  
  CPU time spent (ms)=46810  
  Physical memory (bytes) snapshot=404357120  
  Virtual memory (bytes) snapshot=3197956096  
  Total committed heap usage (bytes)=264241152  
  Peak Map Physical memory (bytes)=404357120  
  Peak Map Virtual memory (bytes)=3197956096  
  
File Input Format Counters  
  Bytes Read=0  
  
File Output Format Counters  
  Bytes Written=531214815  
  
2024-04-02 13:39:40,444 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 59.8842 seconds (8.4598 MB/sec)  
2024-04-02 13:39:40,446 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.  
[root@ip-172-31-78-0 ~]#
```

In the screenshot provided, it's evident that 2468572 records have been retrieved.

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

Command:

```
hadoop fs -ls /user/hadoop/ETL/ATMDATA
```



```
root@ip-172-31-78-0:~
2024-04-02 13:39:40,446 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-172-31-78-0 ~]# hadoop fs -ls /user/hadoop/ETL/ATMDATA/
Found 2 items
-rw-r--r--  1 root hdfsadmingroup          0 2024-04-02 13:39 /user/hadoop/ETL/ATMDATA/_SUCCESS
-rw-r--r--  1 root hdfsadmingroup 531214815 2024-04-02 13:39 /user/hadoop/ETL/ATMDATA/part-m-00000
[root@ip-172-31-78-0 ~]#
```

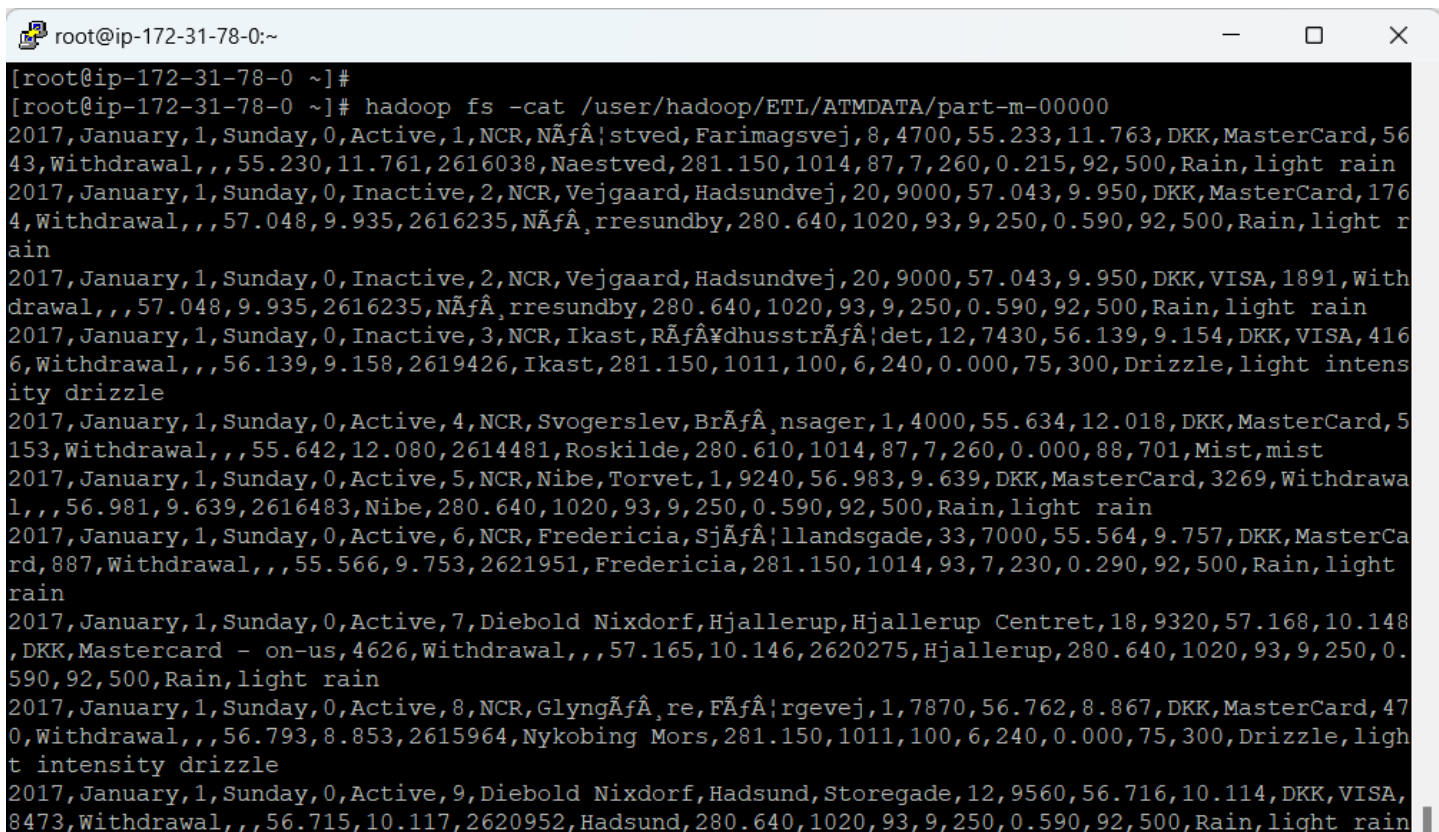
In the screenshot above, two items are visible:

- The first item is the success file, signaling the successful completion of the MapReduce job.
- The second item, labeled 'part-m-00000', is the file imported.

3. Screenshot of the imported data:

Command:

```
hadoop fs -cat /user/hadoop/ETL/ATMDATA/part-m-00000
```



```
root@ip-172-31-78-0:~
[root@ip-172-31-78-0 ~]#
[root@ip-172-31-78-0 ~]# hadoop fs -cat /user/hadoop/ETL/ATMDATA/part-m-00000
2017,January,1,Sunday,0,Active,1,NCR,NÃfÃ;stved,Farimagsvej,8,4700,55.233,11.763,DKK,MasterCard,56
43,Withdrawal,,55.230,11.761,2616038,Naestved,281.150,1014,87,7,260,0.215,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,2,NCR,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,MasterCard,176
4,Withdrawal,,57.048,9.935,2616235,NÃfÃ;rresundby,280.640,1020,93,9,250,0.590,92,500,Rain,light r
ain
2017,January,1,Sunday,0,Inactive,2,NCR,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,VISA,1891,With
drawal,,57.048,9.935,2616235,NÃfÃ;rresundby,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,3,NCR,Ikast,RÃfÃ;dhustrÃfÃ;det,12,7430,56.139,9.154,DKK,VISA,416
6,Withdrawal,,56.139,9.158,2619426,Ikast,281.150,1011,100,6,240,0.000,75,300,Drizzle,light intens
ity drizzle
2017,January,1,Sunday,0,Active,4,NCR,Svogerslev,BrÃfÃ;nsager,1,4000,55.634,12.018,DKK,MasterCard,5
153,Withdrawal,,55.642,12.080,2614481,Roskilde,280.610,1014,87,7,260,0.000,88,701,Mist,mist
2017,January,1,Sunday,0,Active,5,NCR,Nibe,Torvet,1,9240,56.983,9.639,DKK,MasterCard,3269,Withdrawa
l,,56.981,9.639,2616483,Nibe,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,6,NCR,Fredericia,SjÃfÃ;lilandsgade,33,7000,55.564,9.757,DKK,MasterCa
rd,887,Withdrawal,,55.566,9.753,2621951,Fredericia,281.150,1014,93,7,230,0.290,92,500,Rain,light
rain
2017,January,1,Sunday,0,Active,7,Diebold Nixdorf,Hjallerup,Hjallerup Centret,18,9320,57.168,10.148
,DKK,Mastercard - on-us,4626,Withdrawal,,57.165,10.146,2620275,Hjallerup,280.640,1020,93,9,250,0.
590,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,8,NCR,GlyngÃfÃ;re,FÃfÃ;rgevej,1,7870,56.762,8.867,DKK,MasterCard,47
0,Withdrawal,,56.793,8.853,2615964,Nykobing Mors,281.150,1011,100,6,240,0.000,75,300,Drizzle,ligh
t intensity drizzle
2017,January,1,Sunday,0,Active,9,Diebold Nixdorf,Hadsund,Storegade,12,9560,56.716,10.114,DKK,VISA,
8473,Withdrawal,,56.715,10.117,2620952,Hadsund,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
```

The screenshot displays a segment of data extracted from the 'part-m-00000' file.