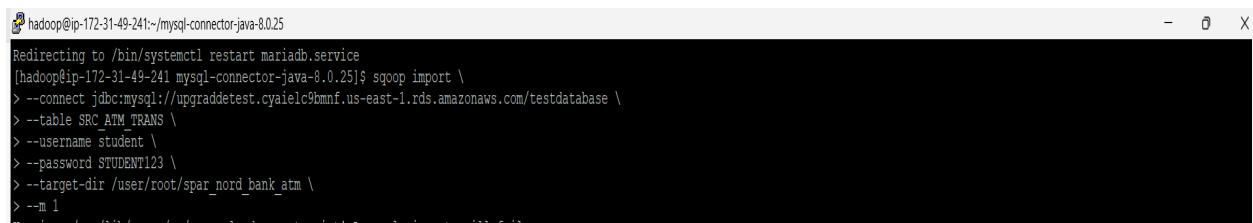


Data Ingestion from the RDS to HDFS using Sqoop

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

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

A screenshot of a terminal window with a dark background. The window title is 'hadoop@ip-172-31-49-241:~/mysql-connector-java-8.0.25'. The terminal shows the command 'sqoop import' being executed with several options: '--connect jdbc:mysql://upgraddetest.cyaieic9bmnf.us-east-1.rds.amazonaws.com/testdatabase', '--table SRC_ATM_TRANS', '--username student', '--password STUDENT123', '--target-dir /user/root/spar_nord_bank_atm', and '--m 1'. The output shows a redirection to restart the mariadb service and then the command execution starting.

Code explanation:

The code is a Sqoop import command used to transfer data from a MySQL database hosted on Amazon RDS (Relational Database Service) to HDFS (Hadoop Distributed File System)

sqoop import: This is the Sqoop command used to initiate the import process.

--connect: This parameter specifies the JDBC connection string to the MySQL database on Amazon RDS. It includes the following components:

jdbc:mysql://upgraddetest.cyaieic9bmnf.us-east-1.rds.amazonaws.com/testdatabase: This part specifies the database connection URL, which consists of the database server hostname, port, and the name of the database (testdatabase in this case).

--table: This parameter specifies the name of the table (SRC_ATM_TRANS) from which data will be imported.

--username: This parameter specifies the username (student) used to authenticate and access the MySQL database.

--password: This parameter specifies the password (STUDENT123) associated with the specified username.

--target-dir: This parameter specifies the target directory (/user/root/spar_nord_bank_atm) in HDFS where the imported data will be stored.

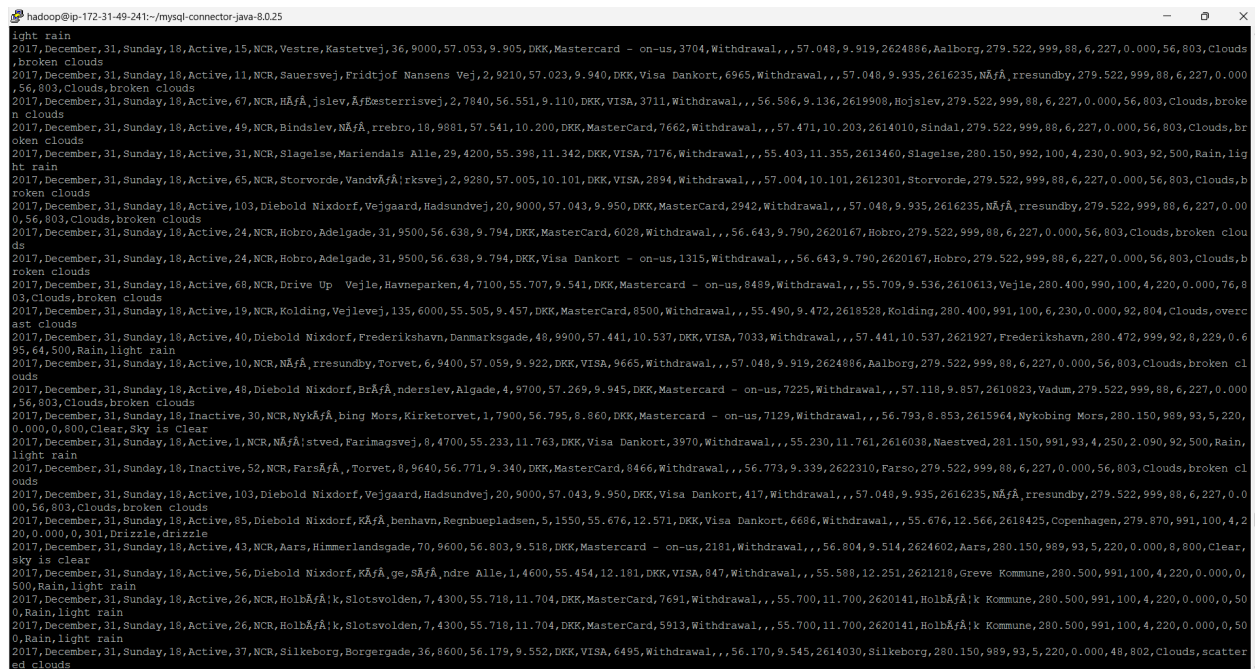
--m: This parameter specifies the number of map tasks that will be used for the import process. In this case, it's set to 1

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

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

```
hadoop@ip-172-31-49-241:~$ ls
build.xml  CHANGES  INFO_BIN  INFO_SRC  LICENSE  mysql-connector-java-8.0.25.jar  README  src  SRC_ATM_TRANS.java
hadoop@ip-172-31-49-241:~$ hadoop fs -ls /user/root/spar_nord_bank_atm
Found 2 items
-rw-r--r-- 1 hadoop hadoop 0 2023-10-02 04:19 /user/root/spar_nord_bank_atm/_SUCCESS
-rw-r--r-- 1 hadoop hadoop 531214815 2023-10-02 04:19 /user/root/spar_nord_bank_atm/part-m-00000
hadoop@ip-172-31-49-241:~$ hadoop fs -cat /user/root/spar_nord_bank_atm/part-m-00000 | wc -l
```

Screenshot of the imported data:



```
light_rain
2017,December,31,Sunday,18,Active,15,NCR,Vestre,Kastetvej,36,9000,57.053,9.905,DKK,Mastercard - on-us,3704,Withdrawal,,,57.048,9.919,2624886,Aalborg,279.522,999,88,6,227,0.000,56,803,Clouds
,broken clouds
2017,December,31,Sunday,18,Active,11,NCR,Sauersvej,Fridtjof Nansens Vej,2,9210,57.023,9.940,DKK,Visa Dankort,6965,Withdrawal,,,57.048,9.935,2616235,NÅfÅ,rrresundby,279.522,999,88,6,227,0.000
,56,803,Clouds,broken clouds
2017,December,31,Sunday,18,Active,67,NCR,HÅfÅ,jslev,ÅfBøsterrisvej,2,7840,56.551,9.110,DKK,VISA,3711,Withdrawal,,,56.586,9.136,2619908,Højslev,279.522,999,88,6,227,0.000,56,803,Clouds,broke
n clouds
2017,December,31,Sunday,18,Active,49,NCR,Bindslev,NÅfÅ,rrrebro,18,9881,57.541,10.200,DKK,MasterCard,7662,Withdrawal,,,57.471,10.203,2614010,Sindal,279.522,999,88,6,227,0.000,56,803,Clouds,br
oken clouds
2017,December,31,Sunday,18,Active,31,NCR,Slagelse,Mariendals Alle,29,4200,55.398,11.342,DKK,VISA,7176,Withdrawal,,,55.403,11.355,2613460,Slagelse,280.150,992,100,4,230,0.903,92,500,Rain,ligh
ht rain
2017,December,31,Sunday,18,Active,65,NCR,Storvorde,VandvÅfÅ,rrksvej,2,9280,57.005,10.101,DKK,VISA,2894,Withdrawal,,,57.004,10.101,2612301,Storvorde,279.522,999,88,6,227,0.000,56,803,Clouds,b
roken clouds
2017,December,31,Sunday,18,Active,103,Diebold Nixdorf,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,MasterCard,2942,Withdrawal,,,57.048,9.935,2616235,NÅfÅ,rrresundby,279.522,999,88,6,227,0.00
0,56,803,Clouds,broken clouds
2017,December,31,Sunday,18,Active,24,NCR,Hobro,Adelgade,31,9500,56.638,9.794,DKK,MasterCard,6028,Withdrawal,,,56.643,9.790,2620167,Hobro,279.522,999,88,6,227,0.000,56,803,Clouds,broken clou
ds
2017,December,31,Sunday,18,Active,24,NCR,Hobro,Adelgade,31,9500,56.638,9.794,DKK,Visa Dankort - on-us,1315,Withdrawal,,,56.643,9.790,2620167,Hobro,279.522,999,88,6,227,0.000,56,803,Clouds,b
roken clouds
2017,December,31,Sunday,18,Active,68,NCR,Drive Up Vejle,Havneparken,4,7100,55.707,9.541,DKK,Mastercard - on-us,8489,Withdrawal,,,55.709,9.536,2610613,Vejle,280.400,990,100,4,220,0.000,76,8
03,Clouds,broken clouds
2017,December,31,Sunday,18,Active,19,NCR,Kolding,Vejlevej,135,6000,55.505,9.457,DKK,MasterCard,8500,Withdrawal,,,55.490,9.472,2618528,Kolding,280.400,991,100,6,230,0.000,92,804,Clouds,overc
ast clouds
2017,December,31,Sunday,18,Active,40,Diebold Nixdorf,Frederikshavn,Danmarksgade,48,9900,57.441,10.537,DKK,VISA,7033,Withdrawal,,,57.441,10.537,2621927,Frederikshavn,280.472,999,92,8,229,0.6
95,64,500,Rain,light rain
2017,December,31,Sunday,18,Active,10,NCR,NÅfÅ,rrresundby,Torvet,6,9400,57.059,9.922,DKK,VISA,9665,Withdrawal,,,57.048,9.919,2624886,Aalborg,279.522,999,88,6,227,0.000,56,803,Clouds,broken cl
ouds
2017,December,31,Sunday,18,Active,48,Diebold Nixdorf,BrÅfÅ,nderslev,Algade,4,9700,57.269,9.945,DKK,Mastercard - on-us,7225,Withdrawal,,,57.118,9.857,2610823,Vadum,279.522,999,88,6,227,0.000
,56,803,Clouds,broken clouds
2017,December,31,Sunday,18,Inactive,30,NCR,NyKÅfÅ,bing Mors,Kirketorvet,1,7900,56.795,8.860,DKK,Mastercard - on-us,7129,Withdrawal,,,56.793,8.853,2615964,Nykobing Mors,280.150,989,93,5,220,
0.000,0,800,Clear,sky is clear
2017,December,31,Sunday,18,Active,1,NCR,NÅfÅ,rrstved,Farimagvej,8,4700,55.233,11.763,DKK,Visa Dankort,3970,Withdrawal,,,55.230,11.761,2616038,Naestved,281.150,991,93,4,250,2.090,92,500,Rain,
light rain
2017,December,31,Sunday,18,Active,26,NCR,FarsÅfÅ,rrtorvet,8,9640,56.771,9.340,DKK,MasterCard,8466,Withdrawal,,,56.773,9.339,2622310,Farso,279.522,999,88,6,227,0.000,56,803,Clouds,broken cl
ouds
2017,December,31,Sunday,18,Active,103,Diebold Nixdorf,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,Visa Dankort,417,Withdrawal,,,57.048,9.935,2616235,NÅfÅ,rrresundby,279.522,999,88,6,227,0.0
00,56,803,Clouds,broken clouds
2017,December,31,Sunday,18,Active,85,Diebold Nixdorf,KÅfÅ,rrbenhavn,Regnbuepladsen,5,1550,55.676,12.571,DKK,Visa Dankort,6686,Withdrawal,,,55.676,12.566,2618425,Copenhagen,279.870,991,100,4,2
20,0.000,0,301,Drizzle,drizzle
2017,December,31,Sunday,18,Active,43,NCR,Aars,Himmerlandsgade,70,9600,56.803,9.518,DKK,Mastercard - on-us,2181,Withdrawal,,,56.804,9.514,2624602,Aars,280.150,989,93,5,220,0.000,8,800,Clear,
sky is clear
2017,December,31,Sunday,18,Active,56,Diebold Nixdorf,KÅfÅ,rrse,GrÅfÅ,ndre Alle,1,4600,55.454,12.181,DKK,VISA,847,Withdrawal,,,55.588,12.251,2621218,Greve Kommune,280.500,991,100,4,220,0.000,0,
500,Rain,light rain
2017,December,31,Sunday,18,Active,26,NCR,HolbÅfÅ,k,Slotsvolden,7,4300,55.718,11.704,DKK,MasterCard,7691,Withdrawal,,,55.700,11.700,2620141,HolbÅfÅ,k Kommune,280.500,991,100,4,220,0.000,0,50
0,Rain,light rain
2017,December,31,Sunday,18,Active,26,NCR,HolbÅfÅ,k,Slotsvolden,7,4300,55.718,11.704,DKK,MasterCard,5913,Withdrawal,,,55.700,11.700,2620141,HolbÅfÅ,k Kommune,280.500,991,100,4,220,0.000,0,50
0,Rain,light rain
2017,December,31,Sunday,18,Active,37,NCR,Silkeborg,Borgergade,36,8600,56.179,9.552,DKK,VISA,6495,Withdrawal,,,56.170,9.545,2614030,Silkeborg,280.150,989,93,5,220,0.000,48,802,Clouds,scatter
ed clouds
```

Note in the screenshot below, verified number of records imported **2468572**

```
hadoop@ip-172-31-49-241:~/mysql-connector-java-8.0.25
23/10/02 04:18:35 INFO db.DBInputFormat: Using read committed transaction isolation
23/10/02 04:18:35 INFO mapreduce.JobSubmitter: number of splits:1
23/10/02 04:18:35 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1696219123723_0001
23/10/02 04:18:36 INFO impl.YarnClientImpl: Submitted application application_1696219123723_0001
23/10/02 04:18:36 INFO mapreduce.Job: The url to track the job: http://ip-172-31-49-241.ec2.internal:20888/proxy/application_1696219123723_0001/
23/10/02 04:18:36 INFO mapreduce.Job: Running job: job_1696219123723_0001
23/10/02 04:18:47 INFO mapreduce.Job: Job job_1696219123723_0001 running in uber mode : false
23/10/02 04:18:47 INFO mapreduce.Job: map 0% reduce 0%
23/10/02 04:19:13 INFO mapreduce.Job: map 100% reduce 0%
23/10/02 04:19:14 INFO mapreduce.Job: Job job_1696219123723_0001 completed successfully
23/10/02 04:19:15 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=189938
    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)=1163472
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=24239
    Total vcore-milliseconds taken by all map tasks=24239
    Total megabyte-milliseconds taken by all map tasks=37231104
  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)=192
    CPU time spent (ms)=27200
    Physical memory (bytes) snapshot=618864640
    Virtual memory (bytes) snapshot=3288432640
    Total committed heap usage (bytes)=536870912
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=531214815
23/10/02 04:19:15 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 43.3641 seconds (11.6826 MB/sec)
23/10/02 04:19:15 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[hadoop@ip-172-31-49-241 mysql-connector-java-8.0.25]$
```