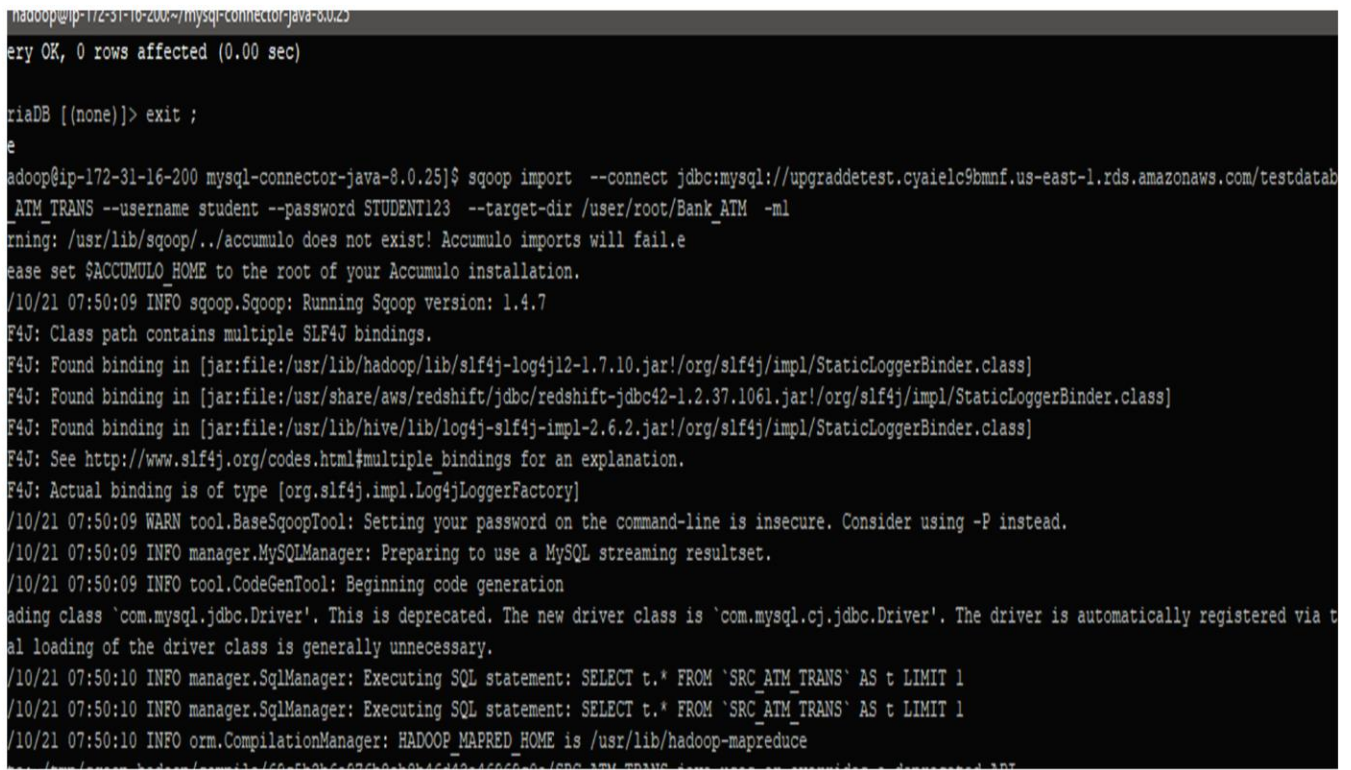


## Data Ingestion from the RDS to HDFS using Sqoop

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

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



```
mysql> query OK, 0 rows affected (0.00 sec)

mysql> exit ;
mysql>
[upgrad@ip-172-31-16-200 ~]$ sqoop import --connect jdbc:mysql://upgradtest.cyaieic9bmnf.us-east-1.rds.amazonaws.com/testdatabase \
--table SRC_ATM_TRANS --username student --password STUDENT123 --target-dir /user/root/Bank_ATM -m1
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
/10/21 07:50:09 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
F4J: Class path contains multiple SLF4J bindings.
F4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
F4J: Found binding in [jar:file:/usr/share/aws/redshift/jdbc/redshift-jdbc42-1.2.37.1061.jar!/org/slf4j/impl/StaticLoggerBinder.class]
F4J: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
F4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
F4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
/10/21 07:50:09 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
/10/21 07:50:09 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
/10/21 07:50:09 INFO tool.CodeGenTool: Beginning code generation
Generating class 'com.mysql.jdbc.Driver'. This is deprecated. The new driver class is 'com.mysql.cj.jdbc.Driver'. The driver is automatically registered via the driver class. The driver class is generally unnecessary.
/10/21 07:50:10 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_ATM_TRANS` AS t LIMIT 1
/10/21 07:50:10 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_ATM_TRANS` AS t LIMIT 1
/10/21 07:50:10 INFO orm.CompilationManager: HADOOP MAPRED HOME is /usr/lib/hadoop-mapreduce
/10/21 07:50:10 INFO orm.CompilationManager: HADOOP MAPRED HOME is /usr/lib/hadoop-mapreduce
```

This screenshot is command of data ingestion RDS to HDFS from Sqoop .

```

hadoop@ip-172-31-16-200:~/mysql-connector-java-8.0.25
22/10/21 07:51:07 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=189904
    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)=1187088
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=24731
    Total vcore-milliseconds taken by all map tasks=24731
    Total megabyte-milliseconds taken by all map tasks=37986816
  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)=305
    CPU time spent (ms)=25580
    Physical memory (bytes) snapshot=629608448
    Virtual memory (bytes) snapshot=3302236160
    Total committed heap usage (bytes)=537395200
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=531214815
22/10/21 07:51:07 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 51.3347 seconds (9.8687 MB/sec)
22/10/21 07:51:07 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.

```

I can see the result of sqoop import job 2468572 records have been retrived (same as the checkpoint mentioned in the Validation document):

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

Command used

**Hadoop fs -ls /user/root/Bank\_ATM**

```
hadoop@ip-172-31-16-200:~/mysql-connector-java-8.0.25
22/10/21 07:51:07 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=189904
    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)=1187088
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=24731
    Total vcore-milliseconds taken by all map tasks=24731
    Total megabyte-milliseconds taken by all map tasks=37986816
  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)=305
    CPU time spent (ms)=25580
    Physical memory (bytes) snapshot=629608448
    Virtual memory (bytes) snapshot=3302236160
    Total committed heap usage (bytes)=537395200
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=531214815
22/10/21 07:51:07 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 51.3347 seconds (9.8687 MB/sec)
22/10/21 07:51:07 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[hadoop@ip-172-31-16-200 mysql-connector-java-8.0.25]$ hadoop fs -ls
[hadoop@ip-172-31-16-200 mysql-connector-java-8.0.25]$ hadoop fs -ls /user/root/Bank_ATM
Found 2 items
-rw-r--r-- 1 hadoop hadoop      0 2022-10-21 07:51 /user/root/Bank_ATM/_SUCCESS
-rw-r--r-- 1 hadoop hadoop 531214815 2022-10-21 07:51 /user/root/Bank_ATM/part-m-00000
[hadoop@ip-172-31-16-200 mysql-connector-java-8.0.25]$
```

In the screenshot above we can see two items: - The first file is the success file, indicating that the MapReduce job was successful. - The second file 'part-m-00000' is the one that I imported. Since I used only one mapper in my import command, thus the data is in a single file.

### +Screenshot of the imported data:

```
Hadoop fs -cat /user/root/Bank_ATM/part-m-00000
```

<Screenshot of the imported data>

```
hadoop@ip-172-31-23-76:~/mysql-connecter-java-8.0.25
00,0.138,92,500,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,278.589,998,99,8,210,0.000,44,802,Clouds,scattered clouds
2017,December,31,Sunday,23,Active,28,NCR,LÃfÃ.gstÃfÃ.r,ÃfÃesterbrogade,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,LÃfÃ.gstÃfÃ.r,ÃfÃesterbrogade,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,NÃfÃ.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,NÃfÃ.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,KÃfÃ.ge,SÃfÃ.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,NÃfÃ.rresundby,277.589,999,87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,1,NCR,NÃfÃ.stved,Farimagvej,8,4700,55.233,11.763,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,KÃfÃ.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,278.589,998,99,8,210,0.000,44,802,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,280.150,998,93,4,210,0.000,92,804,Clouds,overcast clouds
2017,December,31,Sunday,23,Active,49,NCR,Bindslev,NÃfÃ.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,ÃfÃesterÃfÃ.Duus,ÃfÃesterÃfÃ,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,ÃfÃesterÃfÃ.Duus,ÃfÃesterÃfÃ,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,NÃfÃ.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
hadoop@ip-172-31-23-76:~/mysql-connecter-java-8.0.25
```