

Data Ingestion from the RDS to HDFS using Sqoop

Connecting to mysql-connector -

- `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/`

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 --username student --password STUDENT123 --table SRC_ATM_TRANS --target-dir /user/root/nord_bank_atm --m 1
```

```
[hadoop@ip-172-31-54-94 mysql-connector-java-8.0.25]$ sqoop import --connect jdbc:mysql://upgradtest.cyaieic9bmnf.us-east-1.rds.amazonaws.com/testdatabase --username student --password STUDENT123 --table SRC_ATM_TRANS --target-dir /user/root/nord_bank_atm --m 1
```

```
Total megabyte-milliseconds taken by all map tasks=40648704
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)=182
  CPU time spent (ms)=25430
  Physical memory (bytes) snapshot=620044288
  Virtual memory (bytes) snapshot=3303927808
  Total committed heap usage (bytes)=518520832
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=531214815
23/06/18 13:57:47 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 43.16
16 seconds (11.7374 MB/sec)
23/06/18 13:57:47 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
```

Explanation-

- Using the above sqoop command we can confirm that a total of 2468572 records were created.
- For connection we are using the jdbc connection, username, password and the table provided to us.
- For this task we are using only 1 mapper to move data to user/root/nord_bank_atm

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

- `hadoop fs -ls /user/root/nord_bank_atm`
- `hadoop fs -cp /user/root/nord_bank_atm/part-m-00000 /user/livy/`

```
[hadoop@ip-172-31-54-94 mysql-connector-java-8.0.25]$ hadoop fs -ls /user/root/nord_bank_atm
Found 2 items
-rw-r--r-- 1 hadoop hdfsadmingroup 0 2023-06-18 13:57 /user/root/nord_bank_atm/_SUCCESS
-rw-r--r-- 1 hadoop hdfsadmingroup 531214815 2023-06-18 13:57 /user/root/nord_bank_atm/part-m-00000
```

Explanation –

- We can confirm from the first command that a part file was created
- Using the 2nd command we are moving the data to /livy , so spark can access it

Screenshot of the imported data:



```
2017, January, 3, Tuesday, 12, Active, 32, NCR, RÅfÅrdsing, Torphagevej, 59, 6710, 55.498, 0.408, DKK, MasterCard, 3114, Withdrawal, , 55.467, 0.450, 2622447, Rahjerg, 279.630, 1007, 93, 11, 270, 0.000, 92, 310, Drizzle, light intensity drizzle rain
2017, January, 3, Tuesday, 12, Inactive, 95, NCR, Intern KÅfÅr, benhavn, RÅfÅrdspladsen, 75, 1550, 55.676, 12.571, DKK, MasterCard - on-us, 8780, Withdrawal, , 55.676, 12.566, 2618425, Copenhagen, 276.000, 1005, 100, 6, 250, 0.000, 75, 310, Drizzle, light intensity drizzle rain
2017, January, 3, Tuesday, 12, Active, 10, NCR, RÅfÅr, resundby, Torvet, 6, 9400, 57.059, 9.922, DKK, MasterCard - on-us, 4557, Withdrawal, , 57.048, 9.919, 2624886, Aalborg, 277.150, 999, 93, 6, 220, 0.000, 75, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Active, 28, NCR, LÅfÅr, gÅfÅr, c, ÅfÅr, Bøsterbrogade, 8, 9670, 56.967, 9.253, DKK, Visa Dankort - on-us, 6447, Withdrawal, , 56.962, 9.258, 2617467, Logstor, 277.150, 999, 93, 6, 220, 0.000, 75, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Active, 40, Diebold Mixdorf, Frederikshavn, Danmarksvej, 48, 9900, 57.441, 10.537, DKK, Visa Dankort, 8563, Withdrawal, , 57.441, 10.537, 2621927, Frederikshavn, 277.150, 999, 93, 6, 220, 0.000, 75, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Active, 19, NCR, Kolding, Vejlevej, 135, 6000, 55.505, 9.457, DKK, Visa Dankort, 3472, Withdrawal, , 55.490, 9.472, 2618528, Kolding, 279.820, 1007, 93, 13, 260, 0.000, 75, 310, Drizzle, light intensity drizzle rain
2017, January, 3, Tuesday, 12, Active, 15, NCR, Vestre, Mastetvej, 36, 9000, 57.053, 9.905, DKK, MasterCard - on-us, 1674, Withdrawal, , 57.048, 9.919, 2624886, Aalborg, 277.150, 999, 93, 6, 220, 0.000, 75, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Inactive, 16, NCR, Skive, Adelgade, 8, 7800, 56.567, 9.027, DKK, Visa Dankort - on-us, 7951, Withdrawal, , 56.567, 9.033, 2613731, Skive, 281.150, 1003, 93, 10, 270, 2.145, 92, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Active, 10, NCR, RÅfÅr, resundby, Torvet, 6, 9400, 57.059, 9.922, DKK, MasterCard, 4736, Withdrawal, , 57.048, 9.919, 2624886, Aalborg, 277.150, 999, 93, 6, 220, 0.000, 75, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Active, 46, Diebold Mixdorf, Helsingør, t, Sct. Olai Gade, 39, 3000, 56.036, 12.612, DKK, Visa Dankort - on-us, 9625, Withdrawal, , 56.036, 12.614, 2620473, Helsingør, 277.470, 1003, 94, 6, 250, 0.000, 90, 310, Drizzle, light intensity drizzle rain
2017, January, 3, Tuesday, 12, Inactive, 88, NCR, Storcenter indg. D, Høbrovej, 452, 9200, 57.005, 9.876, DKK, Visa Dankort - on-us, 4917, Withdrawal, , 56.972, 9.848, 2612021, Svenstrup, 277.150, 999, 93, 6, 220, 0.000, 75, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Active, 65, NCR, Storvoerde, VandvÅfÅr, rksvej, 2, 9280, 57.005, 10.101, DKK, MasterCard - on-us, 5109, Withdrawal, , 57.004, 10.101, 2612301, Storvoerde, 277.150, 999, 93, 6, 220, 0.000, 75, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Active, 20, NCR, LÅfÅr, gÅfÅr, c, ÅfÅr, Bøsterbrogade, 8, 9670, 56.967, 9.253, DKK, Visa Dankort - on-us, 795, Withdrawal, , 56.962, 9.258, 2617467, Logstor, 277.150, 999, 93, 6, 220, 0.000, 75, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Active, 87, NCR, Aalborg Storcenter indg. D, Høbrovej, 452, 9200, 57.005, 9.876, DKK, Visa Dankort, 8648, Withdrawal, , 56.972, 9.848, 2612021, Svenstrup, 277.150, 999, 93, 6, 220, 0.000, 75, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Inactive, 16, NCR, Skive, Adelgade, 8, 7800, 56.567, 9.027, DKK, Visa Dankort - on-us, 853, Withdrawal, , 56.567, 9.033, 2613731, Skive, 281.150, 1003, 93, 10, 270, 2.145, 92, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Inactive, 93, NCR, Intern BÅfÅr, nderslev, Algade, 4, 9700, 57.269, 9.945, DKK, Visa Dankort - on-us, 8324, Withdrawal, , 57.118, 9.857, 2610823, Vadum, 277.150, 999, 93, 6, 220, 0.000, 75, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Active, 49, NCR, Bindlev, RÅfÅr, rrebro, 18, 9881, 57.541, 10.200, DKK, MasterCard - on-us, 9060, Withdrawal, , 57.471, 10.203, 2614010, Sindal, 277.150, 999, 93, 6, 220, 0.000, 75, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Active, 41, Diebold Mixdorf, Skagen, Sct. Laurentiivej, 34, 9990, 57.723, 10.590, DKK, RÅfÅr, vekort - on-us, 330, Withdrawal, , 57.721, 10.584, 2613939, Skagen, 273.580, 999, 100, 5, 210, 0.000, 90, 904, Clouds, overcast clouds
2017, January, 3, Tuesday, 12, Inactive, 89, NCR, Intern Frederikshavn, Danmarksvej, 48, 9900, 57.441, 10.537, DKK, Visa Dankort - on-us, 4467, Withdrawal, , 57.441, 10.537, 2621927, Frederikshavn, 277.150, 999, 93, 6, 220, 0.000, 75, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Active, 91, NCR, Skive Lobby, Adelgade, 8, 7800, 56.567, 9.027, DKK, Visa Dankort - on-us, 1345, Withdrawal, , 56.567, 9.033, 2613731, Skive, 281.150, 1003, 93, 10, 270, 2.145, 92, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Active, 15, NCR, Vestre, Mastetvej, 36, 9000, 57.053, 9.905, DKK, MasterCard, 1326, Withdrawal, , 57.048, 9.919, 2624886, Aalborg, 277.150, 999, 93, 6, 220, 0.000, 75, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Active, 10, NCR, RÅfÅr, resundby, Torvet, 6, 9400, 57.059, 9.922, DKK, Dankort - on-us, 2593, Withdrawal, , 57.048, 9.919, 2624886, Aalborg, 277.150, 999, 93, 6, 220, 0.000, 75, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Inactive, 93, NCR, Intern BÅfÅr, nderslev, Algade, 4, 9700, 57.269, 9.945, DKK, Visa Dankort - on-us, 7165, Withdrawal, , 57.118, 9.857, 2610823, Vadum, 277.150, 999, 93, 6, 220, 0.000, 75, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Inactive, 54, NCR, Durup, Torvet, 4, 7870, 56.745, 8.949, DKK, MasterCard - on-us, 6859, Withdrawal, , 56.793, 8.853, 2615964, Nykøb, [root@ip-172-31-53-7 ~]# hadoop fs -cat /user/livy/del/de_source.csv
```