

## Data Ingestion from the RDS to HDFS using Sqoop

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

#### Sql jar file import:

```
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/
```

```
[root@ip-10-0-0-103 ~]# wget https://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz
--2021-11-20 17:15:01-- https://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz
Resolving de-mysql-connector.s3.amazonaws.com (de-mysql-connector.s3.amazonaws.com)... 52.217.39.76
Connecting to de-mysql-connector.s3.amazonaws.com (de-mysql-connector.s3.amazonaws.com)|52.217.39.76|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 4079310 (3.9M) [application/x-gzip]
Saving to: 'mysql-connector-java-8.0.25.tar.gz'

100%[=====] 4079310 120 MB/s

2021-11-20 17:15:01 (120 MB/s) - 'mysql-connector-java-8.0.25.tar.gz' saved [4079310/4079310]

[root@ip-10-0-0-103 ~]# tar -xvf mysql-connector-java-8.0.25.tar.gz
```

#### Import command:

```
sqoop import --connect jdbc:mysql://upgraddetest.cyaieic9bmnf.us-east-1.rds.amazonaws.com/
testdatabase --table SRC_ATM_TRANS --username student --password STUDENT123 --null-
string '\N' --fields-terminated-by , --escaped-by \\ --enclosed-by '\"' --null-non-string '\\N' --target-
dir /user/livy/de1 -m 1
```

#### Explanation :

Along with default options, we have used the Null handling parameter to treat the null as blank instead of string 'NULL' .

We have used the escaped and enclosed parameter to handle the comma values in the Message\_text column.

Used one mapper to import the data and imported the data to /user/livy/de1 folder

```
[root@ip-10-0-0-105 mysql-connector-java-8.0.25]# sqoop import --connect jdbc:mysql://upgraddetest.cyaieic9bmnf.us-east-1.rds.amazonaws.com/testdatabase --table
SRC_ATM_TRANS --username student --password STUDENT123 --null-string '\N' --fi
elds-terminated-by , --escaped-by \\ --enclosed-by '\"' --null-non-string '\\N'
--target-dir /user/livy/de1 -m 1
```

```
21/11/23 17:20:25 INFO mapreduce.ImportJobBase: Transferred 666.7002 MB in 100.0
76 seconds (6.6619 MB/sec)
21/11/23 17:20:25 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-10-0-0-187 ~]# hadoop fs -ls /user/livy/
Found 1 items
drwxr-xr-x - root livy 0 2021-11-23 17:20 /user/livy/de1
[root@ip-10-0-0-187 ~]# hadoop fs -ls /user/livy/de1
Found 2 items
-rw-r--r-- 1 root livy 0 2021-11-23 17:20 /user/livy/de1/_SUCCESS
-rw-r--r-- 1 root livy 699085798 2021-11-23 17:20 /user/livy/de1/part-m-00000
[root@ip-10-0-0-187 ~]#
```

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

```
hadoop fs -ls /user/livy/de1
```

```
hadoop fs -cp /user/livy/de1/part-m-00000 /user/livy/
```

## Explanation:

Imported the file to /user/livy/de1 folder than moved in /user/livy/ folder for Spark access.

## Screenshot of the imported data:

```

dr rain
2017, January, 3, Tuesday, 12, Active, 32, NCR, SÅfÅdding, Tarphagevej, 59, 6710, 55.498, 8.408, DKK, MasterCard, 3114, Withdrawal, , 55.467, 8.450, 2622447, Babjerg, 279.630, 1007, 93, 11, 270, 0.000, 92, 310, Drizzle, light intensity drizzle rain
2017, January, 3, Tuesday, 12, Inactive, 95, NCR, Intern KÅfÅbenhavn, RÅfÅVdhusspladsen, 75, 1550, 55.676, 12.571, DKK, Mastercard - on-us, 8780, Withdrawal, , 55.676, 12.566, 2618425, Copenhagen, 276.800, 1005, 100, 6, 250, 0.000, 75, 310, Drizzle, light intensity drizzle rain
2017, January, 3, Tuesday, 12, Active, 10, NCR, MÅfÅ, rresundby, 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, LÅfÅ, gÅfÅfÅ, r, ÅfÅBesterbrogade, 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 Nixdorf, Frederikshavn, Danmarksgade, 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, Inactive, 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, Kastetvej, 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, MÅfÅ, rresundby, 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 Nixdorf, HelsingÅfÅ, r, Sct. Olai Gade, 39, 3000, 56.036, 12.612, DKK, Visa Dankort - on-us, 9825, Withdrawal, , 56.036, 12.614, 2620473, Helsingør, 277.470, 1003, 86, 6, 250, 0.000, 90, 310, Drizzle, light intensity drizzle rain
2017, January, 3, Tuesday, 12, Inactive, 88, NCR, Storcenter indg. A, Hobrovej, 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, Storvorde, VandvÅfÅ, rksvej, 2, 9280, 57.085, 10.101, DKK, Mastercard - on-us, 5109, Withdrawal, , 57.004, 10.101, 2612301, Storvorde, 277.150, 999, 93, 6, 220, 0.000, 75, 500, Rain, light rain
2017, January, 3, Tuesday, 12, Active, 28, NCR, LÅfÅ, gÅfÅfÅ, r, ÅfÅBesterbrogade, 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, 67, NCR, Aalborg Storcenter indg. D, Hobrovej, 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 BrÅfÅ, 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, Bindslev, MÅfÅ, rrebroy, 10, 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 Nixdorf, Skagen, Sct. Laurentilvej, 36, 9990, 57.723, 10.590, DKK, HÅfÅ, vekort - on-us, 330, Withdrawal, , 57.721, 10.584, 2613939, Skagen, 273.580, 999, 100, 5, 210, 0.000, 90, 804, Clouds, overcast clouds
2017, January, 3, Tuesday, 12, Inactive, 89, NCR, Intern Frederikshavn, Danmarksgade, 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, Kastetvej, 36, 9000, 57.053, 9.905, DKK, MasterCard, 3326, 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, MÅfÅ, rresundby, 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, Active, 87, NCR, Aalborg Storcenter indg. D, Hobrovej, 452, 9200, 57.005, 9.876, DKK, Visa Dankort - on-us, 2989, 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, 93, NCR, Intern BrÅfÅ, 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, Nykob[ro]t[ro]p-172-31-53-7 ~|# hadoop fs -cat /user/livy/de1/de_source.csv

```