

ETL Project

BY ANVIKSHA SINGH

anviksha.singh0110s@gmail.com

Step 1: Data Ingestion from the RDS to HDFS using Sqoop

Data Ingestion from the RDS to HDFS using Sqoop

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

Step 1 : Install 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/
```

```
[hadoop@ip-172-31-24-211 ~]$ wget https://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz
--2022-10-23 12:44:27-- 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) ... 52.216.43.169
Connecting to de-mysql-connector.s3.amazonaws.com (de-mysql-connector.s3.amazonaws.com)|52.216.43.169|: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%[=====>] 4,079,310 --.-K/s in 0.05s

2022-10-23 12:44:28 (71.1 MB/s) - 'mysql-connector-java-8.0.25.tar.gz' saved [4079310/4079310]

[hadoop@ip-172-31-24-211 ~]$ tar -xvf mysql-connector-java-8.0.25.tar.gz
mysql-connector-java-8.0.25/
mysql-connector-java-8.0.25/src/
mysql-connector-java-8.0.25/src/build/
[hadoop@ip-172-31-24-211 mysql-connector-java-8.0.25]$ sudo cp mysql-connector-java-8.0.25.jar /usr/lib/sqoop/lib/
[hadoop@ip-172-31-24-211 mysql-connector-java-8.0.25]$ ls
build.xml  CHANGES  INFO_BIN  INFO_SRC  LICENSE  mysql-connector-java-8.0.25.jar  README  src
[hadoop@ip-172-31-24-211 mysql-connector-java-8.0.25]$ mysql_secure_installation

NOTE: RUNNING ALL PARTS OF THIS SCRIPT IS RECOMMENDED FOR ALL MariaDB
SERVERS IN PRODUCTION USE! PLEASE READ EACH STEP CAREFULLY!
```

Step 2 : Import Data Using Scoop

```
scoop 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/SRC_ATM_TRANS \
-m 1
```

```
[hadoop@ip-172-31-24-211 ~]$ sudo -i

EEEEEEEEEEEEEEEEEEEE MMMMMMMM          MMMMMMMM RRRRRRRRRRRRRRRR
E::::::::::::::::::::E M::::::::M          M::::::::M R::::::::::::R
EE::::::::EEEEEEEE::::E M::::::::M          M::::::::M R::::RRRRRR::::R
  E::::E      EEEEE M::::::::M          M::::::::M RR::::R      R::::R
  E::::E      M::::M::M          M::::M::M          R::R      R::::R
  E::::EEEEEEEE M::::M M::M M::M M::::M          R::RRRRRR::::R
  E::::::::::::E M::::M M::M::M M::::M          R:::::::::RR
  E::::EEEEEEEE M::::M M::::M M::::M          R::RRRRRR::::R
  E::::E      M::::M M::M M::::M          R::R      R::::R
  E::::E      EEEEE M::::M      MMM M::::M          R::R      R::::R
EE::::::::EEEEEEEE::::E M::::M          M::::M          R::::R      R::::R
E::::::::::::E M::::M          M::::M RR::::R          R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMMMM          MMMMMMMM RRRRRRR      RRRRRR

[root@ip-172-31-24-211 ~]# scoop 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/SRC_ATM_TRANS \
> -m 1

Bytes Written=531214815
22/10/23 12:59:20 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 46.5053 seconds (10.8935 MB/sec)
22/10/23 12:59:20 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
```

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

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

```
[root@ip-172-31-24-211 ~]# hadoop fs -ls /user/root/SRC_ATM_TRANS
Found 2 items
-rw-r--r--  1 root hadoop          0 2022-10-23 12:59 /user/root/SRC_ATM_TRANS/
_SUCCESS
-rw-r--r--  1 root hadoop  531214815 2022-10-23 12:59 /user/root/SRC_ATM_TRANS/
part-m-00000
```

Screenshot of the imported data:

hadoop fs -cat /user/root/SRC_ATM_TRANS/part-m-00000 | head -n10

```
[root@ip-172-31-24-211 ~]# hadoop fs -cat /user/root/SRC_ATM_TRANS/part-m-00000 | head -n10
2017,January,1,Sunday,0,Active,1,NCR,NÅfÅ!stved,Farimagsvej,8,4700,55.233,11.763,DKK,MasterCard,5643,Withdrawal,,,55.230,1
1.761,2616038,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,1764,Withdrawal,,,57.048,9.
935,2616235,NÅfÅ,rresundby,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,2,NCR,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,VISA,1891,Withdrawal,,,57.048,9.935,26
16235,NÅfÅ,rresundby,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,3,NCR,Ikast,RÅfÅ¸dhusstrÅfÅ!det,12,7430,56.139,9.154,DKK,VISA,4166,Withdrawal,,,56.139,9.
158,2619426,Ikast,281.150,1011,100,6,240,0.000,75,300,Drizzle,light intensity drizzle
2017,January,1,Sunday,0,Active,4,NCR,Svogerslev,BrÅfÅ¸nsager,1,4000,55.634,12.018,DKK,MasterCard,5153,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,Withdrawal,,,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,SjÅfÅ!llandsgade,33,7000,55.564,9.757,DKK,MasterCard,887,Withdrawal,,,55.5
66,9.753,2621951,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,GlyngÅfÅ¸re,FÅfÅ!rgevej,1,7870,56.762,8.867,DKK,MasterCard,470,Withdrawal,,,56.793,8.
853,2615964,Nykobing Mors,281.150,1011,100,6,240,0.000,75,300,Drizzle,light 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
```