

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:3306/testdatabase --table SRC_ATM_TRANS --username student --password STUDENT123 --target-dir /home/hadoop/ATM_dataset --m 1
```

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

```
hadoop fs -ls /home/hadoop/ATM_dataset
```

<Screenshot of the imported data>

```
~/Downloads -- hadoop@ip-172-31-33-87:~ -- zsh
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Other local map tasks=1
  Total time spent by all maps in occupied slots (ms)=1483200
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=30900
  Total vcore-milliseconds taken by all map tasks=30900
  Total megabyte-milliseconds taken by all map tasks=47462400
Map-Reduce Framework
  Map input records=2468572
  Map output records=2468572
  Input split bytes=85
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=288
  CPU time spent (ms)=28688
  Physical memory (bytes) snapshot=603213824
  Virtual memory (bytes) snapshot=3076116480
  Total committed heap usage (bytes)=591396864
  Peak Map Physical memory (bytes)=605679616
  Peak Map Virtual memory (bytes)=3076116480
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=531214815
2023-11-01 01:21:00,568 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 52.9899 seconds (9.5604 MB/sec)
2023-11-01 01:21:00,563 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[hadoop@ip-172-31-33-87 ~]$ ls
SRC_ATM_TRANS.java  mysql-connector-java-8.0.25  mysql-connector-java-8.0.25.tar.gz
[hadoop@ip-172-31-33-87 ~]$ hadoop fs -ls /home/hadoop
Found 1 items
drwxr-xr-x  - hadoop hdfsadmin group 0 2023-11-01 01:20 /home/hadoop/ATM_dataset
[hadoop@ip-172-31-33-87 ~]$ hadoop fs -ls /home/hadoop/ATM_dataset
Found 2 items
-rw-r--r--  1 hadoop hdfsadmin group 0 2023-11-01 01:20 /home/hadoop/ATM_dataset/_SUCCESS
-rw-r--r--  1 hadoop hdfsadmin group 531214815 2023-11-01 01:20 /home/hadoop/ATM_dataset/part-m-000000
[hadoop@ip-172-31-33-87 ~]$ hadoop fs -cp /home/hadoop/ATM_dataset/part-m-000000 /user/livy
[hadoop@ip-172-31-33-87 ~]$ hadoop fs -ls /user/livy
Found 1 items
-rw-r--r--  1 hadoop livy 531214815 2023-11-01 01:23 /user/livy/part-m-000000
[hadoop@ip-172-31-33-87 ~]$ hadoop fs -cat /user/livy/part-m-000000/head
2017, January, 1, Sunday, 0, Active, 1, NCR, NÅFÅstved, Farimagvej, 8, 4700, 55.233, 11.763, DKK, MasterCard, 5643, Withdrawal, , 55.230, 11.761, 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, 2616235, NÅFÅ, rresundby, 280.640, 1020, 93, 9, 250, 0.590, 92, 500, Rain, light rain
2017, January, 1, Sunday, 0, Inactive, 3, NCR, Ikast, RÅFÅdhustrÅFÅdet, 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, BrÅFÅnsager, 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, SjøFÅllandsgade, 33, 7000, 55.564, 9.757, DKK, MasterCard, 887, Withdrawal, , 55.566, 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, FÅFÅrgevej, 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
cat: Unable to write to output stream.
[hadoop@ip-172-31-33-87 ~]$ client_loop: send disconnect: Broken pipe
admin@Admins-MacBook-Pro Downloads %
```