

Data Ingestion from the RDS to HDFS using Sqoop

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

This following command is used on the terminal to import data from RDS to HDFS. The data is saved into HDFS (path /user/livy).

```
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' --null-non-string '\N' \  
--target-dir /user/livy/atm_transaction \  
-m 1
```

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

Because the data lies on HDFS, the following command is used to list the imported data.

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

Screenshot of the imported data:

```
[hadoop@ip-172-31-84-63 mysql-connector-java-8.0.25]$ hadoop fs -ls /user/livy/atm_transaction  
Found 2 items  
-rw-r--r--  1 hadoop livy          0 2022-11-18 14:32 /user/livy/atm_transaction/_SUCCESS  
-rw-r--r--  1 hadoop livy 531214815 2022-11-18 14:32 /user/livy/atm_transaction/part-m-00000  
[hadoop@ip-172-31-84-63 mysql-connector-java-8.0.25]$
```

```

22/12/13 14:01:31 INFO db.DBInputFormat: Using read committed transaction isolation
22/12/13 14:01:31 INFO mapreduce.JobSubmitter: number of splits:1
22/12/13 14:01:31 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1670928975516_0001
22/12/13 14:01:32 INFO impl.YarnClientImpl: Submitted application application_1670928975516_0001
22/12/13 14:01:32 INFO mapreduce.Job: The url to track the job: http://ip-172-31-89-193.ec2.internal:20888/proxy/application_1670928975516_0001/
22/12/13 14:01:32 INFO mapreduce.Job: Running job: job_1670928975516_0001
22/12/13 14:01:42 INFO mapreduce.Job: Job job_1670928975516_0001 running in uber mode : false
22/12/13 14:01:42 INFO mapreduce.Job: map 0% reduce 0%
22/12/13 14:02:08 INFO mapreduce.Job: map 100% reduce 0%
22/12/13 14:02:09 INFO mapreduce.Job: Job job_1670928975516_0001 completed successfully
22/12/13 14:02:09 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=189653
    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)=1164528
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=24261
    Total vcore-milliseconds taken by all map tasks=24261
    Total megabyte-milliseconds taken by all map tasks=37264896
  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)=231
    CPU time spent (ms)=28870
    Physical memory (bytes) snapshot=626229248
    Virtual memory (bytes) snapshot=3307827200
    Total committed heap usage (bytes)=535298048
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=531214815
22/12/13 14:02:09 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 42.6864 seconds (11.8681 MB/sec)
22/12/13 14:02:09 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[hadoop@ip-172-31-89-193 mysql-connector-java-8.0.25]$

```

```

[hadoop@ip-172-31-89-193 mysql-connector-java-8.0.25]$ sudo hadoop fs -tail /user/liwy/atm_transaction/part-m-00000
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,ÄfEsterÄfÄV Duus,ÄfEsterÄfÄV,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,ÄfEsterÄfÄV Duus,ÄfEsterÄfÄV,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,NÄfÄ, 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,Cl
ouds,broken clouds
2017,December,31,Sunday,23,Active,37,NCR,Silkeborg,Borgergade,36,9600,56.179,9.552,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-172-31-89-193 mysql-connector-java-8.0.25]$

```