

Data Ingestion from the RDS to HDFS using Sqoop

Before running the Sqoop import command, verification of non-existence of target directory was undertaken to avoid error generation by Sqoop import command.

```
[hadoop@ip-172-31-46-141 ~]$ hadoop fs -rm -r /user/root/SRC_ATM_TRANS
rm: `/user/root/SRC_ATM_TRANS': No such file or directory
[hadoop@ip-172-31-46-141 ~]$
```

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

```
sqoop import \
--connect jdbc:mysql://upgradtest.cyaiele9bmnf.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-172-31-46-141:~
22/07/24 22:05:53 INFO mapreduce.ImportJobBase: Beginning import of SRC_ATM_TRANS
22/07/24 22:05:53 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
22/07/24 22:05:54 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
22/07/24 22:05:54 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-46-141.ec2.internal/172.31.46.141:8032
22/07/24 22:05:59 INFO db.DBInputFormat: Using read committed transaction isolation
22/07/24 22:05:59 INFO mapreduce.JobSubmitter: number of splits:1
22/07/24 22:06:00 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1658696203632_0001
22/07/24 22:06:00 INFO impl.YarnClientImpl: Submitted application application_1658696203632_0001
22/07/24 22:06:00 INFO mapreduce.Job: The url to track the job: http://ip-172-31-46-141.ec2.internal:20888/proxy/application_1658696203632_0001/
22/07/24 22:06:00 INFO mapreduce.Job: Running job: job_1658696203632_0001
22/07/24 22:06:11 INFO mapreduce.Job: Job job_1658696203632_0001 running in uber mode : false
22/07/24 22:06:11 INFO mapreduce.Job: map 0% reduce 0%
22/07/24 22:06:37 INFO mapreduce.Job: map 100% reduce 0%
22/07/24 22:06:37 INFO mapreduce.Job: Job job_1658696203632_0001 completed successfully
22/07/24 22:06:37 INFO mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=109873
  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)=1118592
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=23304
  Total vcore-milliseconds taken by all map tasks=23304
  Total megabyte-milliseconds taken by all map tasks=35794944
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)=220
  CPU time spent (ms)=25180
  Physical memory (bytes) snapshot=610910208
  Virtual memory (bytes) snapshot=3289710592
  Total committed heap usage (bytes)=534249472
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=531214815
22/07/24 22:06:37 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 43.2333 seconds (11.718 MB/sec)
22/07/24 22:06:37 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[hadoop@ip-172-31-46-141 ~]$ ^C
[hadoop@ip-172-31-46-141 ~]$
```

As a result of Sqoop import job, 2468572 records have been imported which is in consonance with the Validation Document.

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

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

```
[hadoop@ip-172-31-46-141 ~]$ hadoop fs -ls /user/root/SRC_ATM_TRANS
Found 2 items
-rw-r--r-- 1 hadoop hadoop 0 2022-07-24 22:06 /user/root/SRC_ATM_TRANS/S/_SUCCESS
-rw-r--r-- 1 hadoop hadoop 531214815 2022-07-24 22:06 /user/root/SRC_ATM_TRANS/S/part-m-00000
[hadoop@ip-172-31-46-141 ~]$
```

Screenshot of the imported data:

```
hadoop fs -cat /user/root/SRC_ATM_TRANS/part-m-00000
```

2017, February, 21, Tuesday, 13, Active, 19, NCR, Molding, Vejlevej, 136,6000,55,606,9,487,DKK, MasterCard, 9180, Withdrawal, ,55,490,9,472,2618598, Molding, 279,945,1018,93,10,291,0,043,20,801,Clouds, few clouds
2017, February, 21, Tuesday, 13, Active, 19, NCR, Intern Skagen, Sct. Laurentievej, 36,9990,57,723,10,590,DKK, Visa Dankort - on-us,4673,Withdrawal, ,57,721,10,584,1639939, Skagen,279,087,1008,78,5,294,0,000,
0,800,Clear,sky is Clear
2017, February, 21, Tuesday, 13, Active, 26, NCR, Holbåfå: k, Slotsvolden, 7,4300,55,718,11,704,DKK, MasterCard - on-us,5826,Withdrawal, ,55,700,11,700,2620141, Holbåfå: k Kommune, 278,087,1012,91,11,291,0,290,92,
500,Rain,light rain
2017, February, 21, Tuesday, 13, Active, 13, NCR, SÅfå: by, Vestergade, 3,9300,57,334,10,515,DKK, Visa Dankort - on-us,5109,Withdrawal, ,57,441,10,537,2621927, Frederikshavn, 277,678,1009,80,10,288,0,000,0,800,C,
Clear,sky is Clear
2017, February, 21, Tuesday, 13, Active, 10, NCR, NÅfå, rresundby, Torvet, 6,9400,57,059,9,922,DKK, MasterCard - on-us,8066,Withdrawal, ,57,048,9,919,2624886, Aalborg, 279,312,1012,88,11,293,0,005,0,800,Clear,sk y is Clear
2017, February, 21, Tuesday, 13, Active, 23, Diebold Nixdorf, Vodskov, Vodskovvej, 27,9310,57,104,10,027,DKK, Visa Dankort - on-us,5006,Withdrawal, ,57,109,10,022,2610095, Vodskov, 279,312,1012,88,11,293,0,005,
0,800,Clear,sky is Clear
2017, February, 21, Tuesday, 13, Inactive, 2, NCR, Vejgaard, Hadsundvej, 20,9000,57,043,9,950,DKK, Visa Dankort - on-us,831,Withdrawal, ,57,048,9,935,2616235, NÅfå, rresundby, 279,312,1012,88,11,293,0,005,0,800,
Clear,sky is Clear
2017, February, 21, Tuesday, 13, Inactive, 82, NCR, Skallerup Klir, Nordre Klirvej, 21,9800,57,494,9,838,DKK, Visa Dankort, 9136, Withdrawal, ,57,464,9,982,2620214, Hjørring, 279,487,1010,84,10,289,0,000,0,800,Ci,
ear,sky is Clear
2017, February, 21, Tuesday, 13, Active, 17, NCR, Randers, Åfæsterovold, 16,8900,56,462,10,038,DKK, Visa Dankort, 3188, Withdrawal, ,56,715,10,117,2620952, Hadsund, 279,312,1012,88,11,293,0,005,0,800,Clear,sky is
Clear
2017, February, 21, Tuesday, 13, Active, 101, NCR, Bryggen Vejle, SÅfå, nderbrogade, 2,7100,55,705,9,532,DKK, MasterCard - on-us,8532, Withdrawal, ,55,709,9,536,2610613, Vejle, 279,153,1016,91,10,293,0,053,40,80
Clouds, scattered clouds
2017, February, 21, Tuesday, 13, Inactive, 89, NCR, Intern Frederikshavn, Danmarksgade, 48,9900,57,441,10,537,DKK, Dankort - on-us,3095,Withdrawal, ,57,441,10,537,2621927, Frederikshavn, 277,678,1009,80,10,288,
0,000,0,800,Clear,sky is Clear
2017, February, 21, Tuesday, 13, Active, 26, NCR, Holbåfå: k, Slotsvolden, 7,4300,55,718,11,704,DKK, MasterCard, 9878, Withdrawal, ,55,700,11,700,2620141, Holbåfå: k Kommune, 278,087,1012,91,11,291,0,290,92,500,Rai,
n,light rain
2017, February, 21, Tuesday, 13, Active, 103, Diebold Nixdorf, Vejgaard, Hadsundvej, 20,9000,57,043,9,950,DKK, Visa Dankort - on-us,108,Withdrawal, ,57,048,9,935,2616235, NÅfå, rresundby, 279,312,1012,88,11,293,
0,005,0,800,Clear,sky is Clear
2017, February, 21, Tuesday, 13, Inactive, 30, NCR, Nykåfå, Bøgh Marskirtorvet, 1,7900,56,795,8,860,DKK, Visa Dankort, 904, Withdrawal, ,56,793,8,853,2615964, Nykøbing Mors, 279,212,1013,80,11,296,0,000,2,800,C,
Clear,sky is Clear
2017, February, 21, Tuesday, 13, Active, 39, NCR, Svenstrup, Gind HåfÅbvej, 14,9230,56,973,9,851,DKK, Visa Dankort, 4195, Withdrawal, ,56,972,9,848,2612021, Hjørring, 279,312,1012,88,11,293,0,005,0,800,Clear,sk y is Clear
2017, February, 21, Tuesday, 13, Active, 25, Diebold Nixdorf, Odense, FÅfå, lledvej, 3,5000,55,394,10,370,DKK, Visa Dankort - on-us,8266, Withdrawal, ,55,396,10,388,2615876, Odense, 279,087,1017,88,10,290,0,050,3
6,802,Clouds, scattered clouds
2017, February, 21, Tuesday, 13, Active, 101, NCR, Bryggen Vejle, SÅfå, nderbrogade, 2,7100,55,705,9,532,DKK, MasterCard - on-us,8743, Withdrawal, ,55,709,9,536,2610613, Vejle, 279,153,1016,91,10,293,0,053,40,80
2,Clouds, scattered clouds
2017, February, 21, Tuesday, 13, Inactive, 81, NCR, Spar KÅfå, lmand Tornhåfå, j, Tornhåfå, jvej, 4,9220,57,026,10,002,DKK, MasterCard, 8916, Withdrawal, ,56,994,9,991,2621449, Gistrup, 279,312,1012,88,11,293,0,005,
0,800,Clear,sky is Clear
2017, February, 21, Tuesday, 13, Inactive, 88, NCR, Storcenter indg. A, Høbrovej, 452,9200,57,005,9,876,DKK, Visa Dankort, 8353, Withdrawal, ,56,972,9,848,2620121, Svenstrup, 279,312,1012,88,11,293,0,005,0,800,Ci,
ear,sky is Clear
2017, February, 21, Tuesday, 13, Active, 17, NCR, Randers, Åfæsterovold, 16,8900,56,462,10,038,DKK, Visa Dankort, 5016, Withdrawal, ,56,715,10,117,2620952, Hadsund, 279,312,1012,88,11,293,0,005,0,800,Clear,sky is
Clear
2017, February, 21, Tuesday, 13, Active, 21, NCR, ÅfæsterÅfåfå, HÅfå, ller, ÅfæsterÅfåfå, 9,922,DKK, MasterCard - on-us,23,Withdrawal, ,57,048,9,919,2624886, Aalborg, 279,312,1012,88,11,293,0,005,
0,800,Clear,sky is Clear
2017, February, 21, Tuesday, 13, Active, 36, NCR, HjÅfå, ring, Åfæstergade, 8,9800,57,459,9,988,DKK, MasterCard - on-us,5066, Withdrawal, ,57,464,9,982,2620214, Hjørring, 279,487,1010,84,10,289,0,000,0,800,Clea,
r,sky is Clear
2017, February, 21, Tuesday, 13, Active, 91, NCR, Skive Lobby, Adelgade, 8,7800,56,567,9,027,DKK, HÅfå, vekort - on-us,5262, Withdrawal, ,56,567,9,033,2613791, Skive, 279,503,1015,91,10,293,0,013,64,803,Clouds,br
2017, February, 21, Tuesday, 13, Active, 70, Diebold Nixdorf, Holstebro, Høstrupsvej, 6,7500,56,373,8,625,DKK, Visa Dankort - on-us,1112, Withdrawal, ,56,360,8,616,2620046, Holstebro, 279,020,1016,93,10,291,0,00
0,16,801,Clouds, few clouds
2017, February, 21, Tuesday, 13, Active, 101, NCR, Bryggen Vejle, SÅfå, nderbrogade, 2,7100,55,705,9,532,DKK, MasterCard - on-us,9774, Withdrawal, ,55,709,9,536,2610613, Vejle, 279,153,1016,91,10,293,0,053,40,80
2,Clouds, scattered clouds
2017, February, 21, Tuesday, 13, Inactive, 12, NCR, ÅfæsterÅfåfå, Duus, ÅfæsterÅfåfå, 12,9000,57,049,9,922,DKK, Visa Dankort - on-us,5532, Withdrawal, ,57,048,9,919,2624886, Aalborg, 279,312,1012,88,11,293,0,005
0,800,Clear,sky is Clear
2017, February, 21, Tuesday, 13, Inactive, 29, NCR, Skelagervej 15, Skelagervej, 15,9000,57,023,9,891,DKK, Visa Dankort - on-us,8340, Withdrawal, ,57,048,9,919,2624886, Aalborg, 279,312,1012,88,11,293,0,005,0,80
0,Clear,sky is Clear
2017, February, 2