

## 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://upgradtest.cyaieic9bmnf.us-east-1.rds.amazonaws.com/testdatabase --username student --password STUDENT123 --table SRC_ATM_TRANS -m 1
```

After successful execution of command, data is available in user/Hadoop/SRC\_ATM\_TRANS location.

```
~/Documents/aws_pem_file -- hadoop@ip-10-0-4-212:~ -- -zsh
22/10/08 10:16:27 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'SRC_ATM_TRANS' AS t LIMIT 1
22/10/08 10:16:27 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-hadoop/compile/fd14b94244ee5e06acfe1f080384e774/SRC_ATM_TRANS.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
22/10/08 10:16:32 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-hadoop/compile/fd14b94244ee5e06acfe1f080384e774/SRC_ATM_TRANS.jar
22/10/08 10:16:32 WARN manager.MySQLManager: It looks like you are importing from mysql.
22/10/08 10:16:32 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
22/10/08 10:16:32 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
22/10/08 10:16:32 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
22/10/08 10:16:32 INFO mapreduce.ImportJobBase: Beginning import of SRC_ATM_TRANS
22/10/08 10:16:33 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
22/10/08 10:16:34 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
22/10/08 10:16:40 INFO client.RMProxy: Connecting to ResourceManager at ip-10-0-4-212.ec2.internal/10.0.4.212:8032
22/10/08 10:16:40 INFO db.DBInputFormat: Using read committed transaction isolation
22/10/08 10:16:40 INFO mapreduce.JobSubmitter: number of splits:1
22/10/08 10:16:40 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1665223789839_0001
22/10/08 10:16:40 INFO impl.YarnClientImpl: Submitted application application_1665223789839_0001
22/10/08 10:16:41 INFO mapreduce.Job: The url to track the job: http://ip-10-0-4-212.ec2.internal:20888/proxy/application_1665223789839_0001/
22/10/08 10:16:50 INFO mapreduce.Job: Job job_1665223789839_0001 running in uber mode : false
22/10/08 10:16:50 INFO mapreduce.Job: map 0% reduce 0%
22/10/08 10:17:15 INFO mapreduce.Job: map 100% reduce 0%
22/10/08 10:17:15 INFO mapreduce.Job: Job job_1665223789839_0001 completed successfully
22/10/08 10:17:15 INFO mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=190184
  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)=1114368
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=23216
  Total vcore-milliseconds taken by all map tasks=23216
  Total megabyte-milliseconds taken by all map tasks=35659776
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)=259
  CPU time spent (ms)=27060
  Physical memory (bytes) snapshot=618201088
  Virtual memory (bytes) snapshot=3290882048
  Total committed heap usage (bytes)=536870912
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=531214815
22/10/08 10:17:15 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 41.1388 seconds (12.3145 MB/sec)
22/10/08 10:17:15 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
```

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

hdfs dfs -ls SRC\_ATM\_TRANS

```
[hadoop@ip-10-0-5-106 ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x - hadoop hadoop 0 2022-10-08 05:19 SRC_ATM_TRANS
[hadoop@ip-10-0-5-106 ~]$ hdfs dfs -ls SRC_ATM_TRANS
Found 2 items
-rw-r--r-- 1 hadoop hadoop 0 2022-10-08 05:19 SRC_ATM_TRANS/SUCCESS
-rw-r--r-- 1 hadoop hadoop 531214815 2022-10-08 05:19 SRC_ATM_TRANS/part-m-000000
[hadoop@ip-10-0-5-106 ~]$ hadoop fs -cat /user/hadoop/SRC_ATM_TRANS/part-m-000000 | head
2017,January,1,Sunday,0,Active,1,NCR,NÄfÄ\stved,Farimagsvej,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,NÄfÄ,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,NÄfÄ,rresundby,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,3,NCR,Ikast,RÄfÄVdhusstrÄfÄ\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,BrÄfÄ\nsager,1,4000,55.634,12.018,DKK,MasterCard,5153,Withdrawal,,55.642,12.000,2614481,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,SjÄfÄ\llandsgade,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,GlyngÄfÄ\re,FÄfÄ\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
cat: Unable to write to output stream.
[hadoop@ip-10-0-5-106 ~]$
```

## Screenshot of the imported data:

```
[hadoop@ip-10-0-5-106 ~]$ hadoop fs -cat /user/hadoop/SRC_ATM_TRANS/part-m-000000 | head
2017,January,1,Sunday,0,Active,1,NCR,NÄfÄ\stved,Farimagsvej,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,NÄfÄ,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,NÄfÄ,rresundby,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,3,NCR,Ikast,RÄfÄVdhusstrÄfÄ\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,BrÄfÄ\nsager,1,4000,55.634,12.018,DKK,MasterCard,5153,Withdrawal,,55.642,12.000,2614481,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,SjÄfÄ\llandsgade,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,GlyngÄfÄ\re,FÄfÄ\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
cat: Unable to write to output stream.
[hadoop@ip-10-0-5-106 ~]$
```