

Data Ingestion from RDS to HDFS using Sqoop

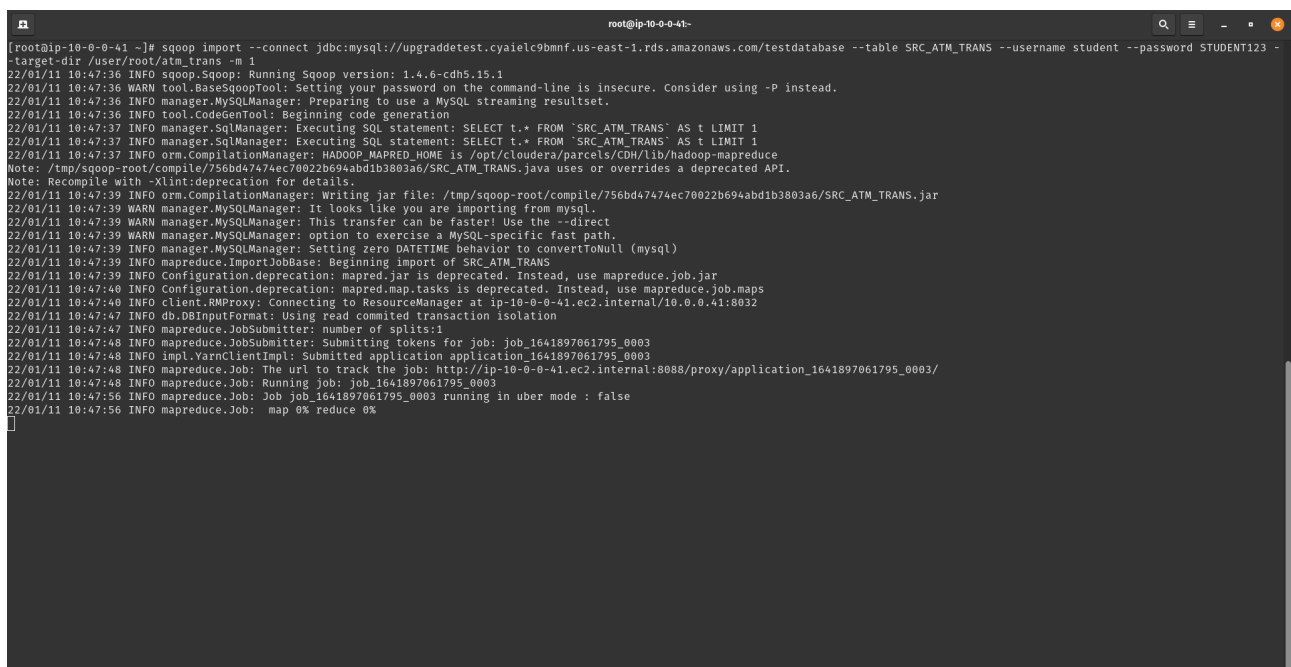
Sqoop Import command used for importing table from RDS to HDFS:

```
sqoop import --connect  
jdbc:mysql://upgraddetest.cyaie1c9bmnf.us-east-  
1.rds.amazonaws.com/testdatabase --table  
SRC_ATM_TRANS --username student --password  
STUDENT123 --target-dir /user/root/atm_trans -m 1
```

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

```
hadoop fs -ls /user/root/atm_trans  
hadoop fs -cat /user/root/atm_trans/*
```

Screenshot of the imported data:



```
root@ip-10-0-0-41:~# sqoop import --connect jdbc:mysql://upgraddetest.cyaie1c9bmnf.us-east-1.rds.amazonaws.com/testdatabase --table SRC_ATM_TRANS --username student --password STUDENT123 --target-dir /user/root/atm_trans -m 1
22/01/11 10:47:36 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.15.1
22/01/11 10:47:36 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
22/01/11 10:47:36 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
22/01/11 10:47:36 INFO tool.CodeGenTool: Beginning code generation
22/01/11 10:47:37 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_ATM_TRANS` AS t LIMIT 1
22/01/11 10:47:37 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_ATM_TRANS` AS t LIMIT 1
22/01/11 10:47:37 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce
Note: /tmp/sqoop-root/compile/756bd47474ec70022b694abd1b3803a6/SRC_ATM_TRANS.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
22/01/11 10:47:39 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/756bd47474ec70022b694abd1b3803a6/SRC_ATM_TRANS.jar
22/01/11 10:47:39 WARN manager.MySQLManager: It looks like you are importing from mysql.
22/01/11 10:47:39 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
22/01/11 10:47:39 INFO manager.MySQLManager: option to exercise a MySQL-specific fast path.
22/01/11 10:47:39 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
22/01/11 10:47:39 INFO mapreduce.ImportJobBase: Beginning import of SRC_ATM_TRANS
22/01/11 10:47:39 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
22/01/11 10:47:40 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
22/01/11 10:47:40 INFO client.RMProxy: Connecting to ResourceManager at ip-10-0-0-41.ec2.internal/10.0.0.41:8032
22/01/11 10:47:47 INFO db.DBInputFormat: Using read committed transaction isolation
22/01/11 10:47:47 INFO mapreduce.JobSubmitter: number of splits:1
22/01/11 10:47:48 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1641897061795_0003
22/01/11 10:47:48 INFO impl.YarnClientImpl: Submitted application application_1641897061795_0003
22/01/11 10:47:48 INFO mapreduce.Job: The url to track the job: http://ip-10-0-0-41.ec2.internal:8088/proxy/application_1641897061795_0003/
22/01/11 10:47:48 INFO mapreduce.Job: Running job: job_1641897061795_0003
22/01/11 10:47:56 INFO mapreduce.Job: Job job_1641897061795_0003 running in uber mode : false
22/01/11 10:47:56 INFO mapreduce.Job: map 0% reduce 0%
```

```

root@ip-10-0-0-41~
22/01/11 10:47:48 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1641897061795_0003
22/01/11 10:47:48 INFO impl.VarClientImpl: Submitted application application_1641897061795_0003
22/01/11 10:47:48 INFO mapreduce.Job: The url to track the job: http://ip-10-0-0-41.ec2.internal:8088/proxy/application_1641897061795_0003/
22/01/11 10:47:48 INFO mapreduce.Job: Running job: job_1641897061795_0003
22/01/11 10:47:56 INFO mapreduce.Job: Job job_1641897061795_0003 running in uber mode : false
22/01/11 10:47:56 INFO mapreduce.Job: map 0% reduce 0%
22/01/11 10:48:28 INFO mapreduce.Job: map 100% reduce 0%
22/01/11 10:48:28 INFO mapreduce.Job: Job job_1641897061795_0003 completed successfully
22/01/11 10:48:29 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=176625
    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)=29673
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=29673
    Total vcore-milliseconds taken by all map tasks=29673
    Total megabyte-milliseconds taken by all map tasks=30385152
  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)=278
    CPU time spent (ms)=29368
    Physical memory (bytes) snapshot=422420480
    Virtual memory (bytes) snapshot=2889896960
    Total committed heap usage (bytes)=387973120
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=531214815
22/01/11 10:48:29 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 48.744 seconds (10.3932 MB/sec)
22/01/11 10:48:29 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-10-0-0-41 ~]#

```

```

root@ip-10-0-0-41~
[root@ip-10-0-0-41 ~]# hadoop fs -ls /user/root/atm_trans
Found 2 items
-rw-r--r-- 3 root supergroup          0 2022-01-11 10:48 /user/root/atm_trans/_SUCCESS
-rw-r--r-- 3 root supergroup 531214815 2022-01-11 10:48 /user/root/atm_trans/part-m-00000
[root@ip-10-0-0-41 ~]#

```

```
root@ip-10-0-0-41:~#
2017, January, 2, Monday, 17, Active, 58, NCR, StÅfÅ_vring, Baunebakken, 4, 9530, 56.890, 9.836, DKK, Visa Dankort, 2871, Withdrawal, , , 56.885, 9.838, 2612500, Stovring, 272.167, 1025, 69, 2, 277, 0.000, 0, 800, Clear, Sky is Clear
2017, January, 2, Monday, 17, Inactive, 88, NCR, Storcenter indg. A, Hobrovej, 452, 9200, 57.005, 9.876, DKK, Mastercard - on-us, 7538, Withdrawal, , , 56.972, 9.848, 2612021, Svenstrup, 272.167, 1025, 69, 2, 277, 0.000, 0, 800, Clear, Sky is Clear
2017, January, 2, Monday, 17, Active, 19, NCR, Kolding, Vejlevej, 135, 6000, 55.505, 9.457, DKK, VISA, 9078, Withdrawal, , , 55.490, 9.472, 2618528, Kolding, 274.320, 1015, 100, 3, 240, 0.000, 90, 520, Rain, light intensity shower rain
2017, January, 2, Monday, 17, Active, 68, NCR, Drive Up Vejle, Havneparken, 4, 7100, 55.707, 9.541, DKK, Visa Dankort, 2776, Withdrawal, , , 55.709, 9.536, 2610613, Vejle, 273.950, 1015, 99, 3, 240, 0.000, 90, 520, Rain, light intensity shower rain
2017, January, 2, Monday, 17, Active, 58, NCR, StÅfÅ_vring, Baunebakken, 4, 9530, 56.890, 9.836, DKK, Visa Dankort, 6647, Withdrawal, , , 56.885, 9.838, 2612500, Stovring, 272.167, 1025, 69, 2, 277, 0.000, 0, 800, Clear, Sky is Clear
2017, January, 2, Monday, 17, Active, 9, Diebold Nixdorf, Hadsund, Storegade, 12, 9560, 56.716, 10.114, DKK, Visa Dankort, 4907, Withdrawal, , , 56.715, 10.117, 2620952, Hadsund, 272.167, 1025, 69, 2, 277, 0.000, 0, 800, Clear, Sky is Clear
2017, January, 2, Monday, 17, Active, 26, NCR, HolbÅfÅ_k, Slotsvolden, 7, 4300, 55.718, 11.704, DKK, Mastercard - on-us, 4594, Withdrawal, , , 55.700, 11.700, 2620141, HolbÅfÅ_k Kommune, 272.150, 1014, 100, 2, 260, 0.000, 0, 800, Clear, Sky is Clear
2017, January, 2, Monday, 17, Active, 10, NCR, NÅfÅ_rresundby, Torvet, 6, 9400, 57.059, 9.922, DKK, HÅfÅ'vekort - on-us, 8250, Withdrawal, , , 57.048, 9.919, 2624886, Aalborg, 272.167, 1025, 69, 2, 277, 0.000, 0, 800, Clear, Sky is Clear
2017, January, 2, Monday, 17, Active, 46, Diebold Nixdorf, HelsingÅfÅ_r, Sct. Olai Gade, 39, 3000, 56.036, 12.612, DKK, Mastercard - on-us, 7465, Withdrawal, , , 56.036, 12.614, 2620473, Helsingør, 271.840, 1012, 79, 5, 289, 0.000, 20, 801, Clouds, few clouds
2017, January, 2, Monday, 17, Active, 48, Diebold Nixdorf, BrÅfÅ_nderslev, Algade, 4, 9700, 57.269, 9.945, DKK, Visa Dankort - on-us, 2577, Withdrawal, , , 57.118, 9.857, 2610823, Vadum, 272.167, 1025, 69, 2, 277, 0.000, 0, 800, Clear, Sky is Clear
2017, January, 2, Monday, 17, Inactive, 81, NCR, Spar KÅfÅ_bmand TornhÅfÅ_j, TornhÅfÅ_jvej, 4, 9220, 57.026, 10.002, DKK, MasterCard, 490, Withdrawal, , , 56.994, 9.991, 2621449, Gistrup, 272.167, 1025, 69, 2, 277, 0.000, 0, 800, Clear, Sky is Clear
2017, January, 2, Monday, 17, Active, 11, NCR, Sauersvej, Fridtjof Nansens Vej, 2, 9210, 57.023, 9.948, DKK, VISA, 4719, Withdrawal, , , 57.048, 9.935, 2616235, NÅfÅ_rresundby, 272.167, 1025, 69, 2, 277, 0.000, 0, 800, Clear, Sky is Clear
2017, January, 2, Monday, 17, Active, 33, NCR, Vadum, Ellehammersvej, 43, 9430, 57.118, 9.861, DKK, Visa Dankort - on-us, 2125, Withdrawal, , , 57.118, 9.857, 2610823, Vadum, 272.167, 1025, 69, 2, 277, 0.000, 0, 800, Clear, Sky is Clear
2017, January, 2, Monday, 17, Active, 28, NCR, LÅfÅ_gstÅfÅ_r, Åfemsterbrogade, 8, 9670, 56.967, 9.253, DKK, Visa Dankort - on-us, 6467, Withdrawal, , , 56.962, 9.258, 2617467, Logstor, 272.167, 1025, 69, 2, 277, 0.000, 0, 800, Clear, Sky is Clear
2017, January, 2, Monday, 17, Inactive, 16, NCR, Skive, Adelgade, 8, 7800, 56.567, 9.027, DKK, HÅfÅ'vekort - on-us, 5962, Withdrawal, , , 56.567, 9.033, 2613731, Skive, 274.150, 1014, 100, 3, 240, 0.000, 32, 520, Rain, light intensity shower rain
2017, January, 2, Monday, 17, Active, 17, NCR, Randers, Åfemstervold, 16, 8900, 56.462, 10.038, DKK, Visa Dankort, 406, Withdrawal, , , 56.715, 10.117, 2620952, Hadsund, 272.167, 1025, 69, 2, 277, 0.000, 0, 800, Clear, Sky is Clear
2017, January, 2, Monday, 17, Active, 20, NCR, Bispensgade, Bispensgade, 35, 9800, 57.453, 9.996, DKK, Dankort - on-us, 4147, Withdrawal, , , 57.464, 9.982, 2620214, Hjørring, 274.917, 1025, 100, 3, 298, 0.000, 0, 800, Clear, Sky is Clear
2017, January, 2, Monday, 17, Active, 42, NCR, Vinderup, SÅfÅ_ndergade, 5, 7830, 56.481, 8.779, DKK, Visa Dankort - on-us, 9517, Withdrawal, , , 56.475, 8.791, 2610196, Vinderup, 272.167, 1025, 69, 2, 277, 0.000, 0, 800, Clear, Sky is Clear
2017, January, 2, Monday, 17, Active, 46, Diebold Nixdorf, HelsingÅfÅ_r, Sct. Olai Gade, 39, 3000, 56.036, 12.612, DKK, Mastercard - on-us, 5145, Withdrawal, , , 56.036, 12.614, 2620473, Helsingør, 271.840, 1012, 79, 5, 289, 0.000, 20, 801, Clouds, few clouds
2017, January, 2, Monday, 17, Active, 84, NCR, Svendborg, Sankt Nicolai Gade, 1, 5700, 55.058, 10.609, DKK, Mastercard - on-us, 7173, Withdrawal, , , 55.060, 10.607, 2612045, Svendborg, 274.400, 1015, 100, 4, 260, 0.000, 0, 800, Clear, Sky is Clear
2017, January, 2, Monday, 17, Active, 37, NCR, Silkeborg, Borgergade, 36, 8600, 56.179, 9.552, DKK, VISA, 428, Withdrawal, , , 56.170, 9.545, 2614030, Silkeborg, 273.090, 1014, 100, 3, 240, 0.000, 0, 520, Rain, light intensity shower rain
2017, January, 2, Monday, 17, Active, 67, NCR, HÅfÅ_jslev, Åfemsterrisvej, 2, 7840, 56.551, 9.110, DKK, Visa Dankort, 624, Withdrawal, , , 56.586, 9.136, 2619908, Højslev, 272.167, 1025, 69, 2, 277, 0.000, 0, 800, Clear, Sky is Clear
2017, January, 2, Monday, 17, Active, 15, NCR, Vestre, Kastetvej, 36, 9000, 57cat: Filesystem closed
[root@ip-10-0-0-41 ~]# ^C
[root@ip-10-0-0-41 ~]#
```