

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 --table SRC_ATM_TRANS --username student --password STUDENT123 --target-dir /user/root/SRC_ATM_TRANS -m 1
```

```
hadoop@ip-172-31-21-118:~$ sqoop import --connect jdbc:mysql://upgradtest.cyaieic9bmnf.us-east-1.rds.amazonaws.com/testdatabase --table SRC_ATM_TRANS --username student --password STUDENT123 --target-dir /user/root/SRC_ATM_TRANS -m 1
Warning: /usr/lib/sqoop/../hbase does not exist! HBase imports will fail.
Please set $HBASE_HOME to the root of your HBase installation.
Warning: /usr/lib/sqoop/../hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
22/08/10 07:00:32 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/share/aws/redshift/jdbc/redshift-jdbc42-1.2.37.1061.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
22/08/10 07:00:32 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
22/08/10 07:00:32 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
22/08/10 07:00:32 INFO tool.CodeGenTool: Beginning code generation
Loading class 'com.mysql.jdbc.Driver'. This is deprecated. The new driver class is 'com.mysql.cj.jdbc.Driver'. The driver is automatically registered via the SPI and manual loading of the driver class is generally unnecessary.
22/08/10 07:00:33 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'SRC_ATM_TRANS' AS t LIMIT 1
22/08/10 07:00:33 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'SRC_ATM_TRANS' AS t LIMIT 1
22/08/10 07:00:33 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-hadoop/compile/be036e6616af124d56a021514c03940d/SRC_ATM_TRANS.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
22/08/10 07:00:35 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-hadoop/compile/be036e6616af124d56a021514c03940d/SRC_ATM_TRANS.jar
22/08/10 07:00:35 WARN manager.MySQLManager: It looks like you are importing from mysql.
22/08/10 07:00:35 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
22/08/10 07:00:35 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
22/08/10 07:00:35 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
22/08/10 07:00:36 INFO mapreduce.ImportJobBase: Beginning import of SRC_ATM_TRANS
22/08/10 07:00:36 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
22/08/10 07:00:37 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
22/08/10 07:00:37 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-21-118.ec2.internal/172.31.21.118:8032
22/08/10 07:00:39 INFO db.DBInputFormat: Using read committed transaction isolation
22/08/10 07:00:39 INFO mapreduce.JobSubmitter: number of splits:1
```

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

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

```
hadoop@ip-172-31-21-118:~$
22/08/10 07:00:39 INFO mapreduce.Job: Running job: job_1660113171877_0002
22/08/10 07:00:47 INFO mapreduce.Job: Job job_1660113171877_0002 running in uber mode : false
22/08/10 07:00:47 INFO mapreduce.Job: map 0% reduce 0%
22/08/10 07:01:15 INFO mapreduce.Job: map 100% reduce 0%
22/08/10 07:01:15 INFO mapreduce.Job: Job job_1660113171877_0002 completed successfully
22/08/10 07:01:15 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=189652
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    HDFS: 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)=1183440
    Total time spent by all reducers in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=24655
    Total vcore-milliseconds taken by all map tasks=24655
    Total megabyte-milliseconds taken by all map tasks=37870080
  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)=205
    CPU time spent (ms)=26160
    Physical memory (bytes) snapshot=624095232
    Virtual memory (bytes) snapshot=3289423872
    Total committed heap usage (bytes)=535298048
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=531214815
22/08/10 07:01:15 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 37.7407 seconds (13.4233 MB/sec)
22/08/10 07:01:15 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[hadoop@ip-172-31-21-118 ~]$
```

Screenshot of the imported data:

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

```
hadoop@ip-172-31-21-118:~$
989,100,4,200,0.138,92,500,Rain,light rain
2017,December,31,Sunday,23,Active,24,NCR,Hobro,Adelgade,31,9500,56.638,9.794,DKK,Mastercard - on-us,7780,Withdrawal,,,56.643,9.790,2620167,Hobro,277.589,999,
87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,20,NCR,Bispensgade,Bispensgade,35,9800,57.453,9.996,DKK,Mastercard - on-us,2830,Withdrawal,,,57.464,9.982,2620214,Hjorring,
278.589,999,99,8,210,0.000,44,802,Clouds,scattered clouds
2017,December,31,Sunday,23,Active,28,NCR,LÅFÅ,gstÅFÅ,r,ÅFÅesterbrogade,8,9670,56.967,9.253,DKK,VISA,812,Withdrawal,,,56.962,9.258,2617467,Logstor,277.589,999,
87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,28,NCR,LÅFÅ,gstÅFÅ,r,ÅFÅesterbrogade,8,9670,56.967,9.253,DKK,VISA,5568,Withdrawal,,,56.962,9.258,2617467,Logstor,277.589,999,
9,87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,49,NCR,Bindslev,NÅFÅ,rrebro,18,9881,57.541,10.200,DKK,MasterCard,5403,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,Active,49,NCR,Bindslev,NÅFÅ,rrebro,18,9881,57.541,10.200,DKK,MasterCard,312,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,Active,56,Diebold Nixdorf,KÅFÅ,ge,SÅFÅ,ndre Alle,1,4600,55.454,12.181,DKK,Visa Dankort - on-us,5351,Withdrawal,,,55.588,12.251,262
1218,Greve Kommune,280.150,992,100,3,190,0.000,90,804,Clouds,overcast clouds
2017,December,31,Sunday,23,Active,103,Diebold Nixdorf,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,Mastercard - on-us,5821,Withdrawal,,,57.048,9.935,2616235,
NÅFÅ,rresundby,277.589,999,99,8,210,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,1,NCR,NÅFÅ,stved,Farimagvej,8,4700,55.233,11.763,DKK,Visa Dankort,3049,Withdrawal,,,55.230,11.761,2616038,Naestved,280.150
,991,100,5,210,1.175,92,500,Rain,light rain
2017,December,31,Sunday,23,Active,85,Diebold Nixdorf,KÅFÅ,benhavn,Regnbuepladsen,5,1550,55.676,12.571,DKK,MasterCard,7785,Withdrawal,,,55.676,12.566,2618425,
Copenhagen,280.150,992,100,3,190,0.000,90,804,Clouds,overcast clouds
2017,December,31,Sunday,23,Active,15,NCR,Vestre,Kastetvej,36,9000,57.053,9.905,DKK,Mastercard - on-us,9777,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,15,NCR,Vestre,Kastetvej,36,9000,57.053,9.905,DKK,MasterCard,5531,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,34,NCR,Skipperen,Vestre Alle,2,9000,57.034,9.908,DKK,MasterCard,8323,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,20,NCR,Bispensgade,Bispensgade,35,9800,57.453,9.996,DKK,MasterCard,5023,Withdrawal,,,57.464,9.982,2620214,Hjorring,278.589,
999,99,8,210,0.000,44,802,Clouds,scattered clouds
2017,December,31,Sunday,23,Active,34,NCR,Skipperen,Vestre Alle,2,9000,57.034,9.908,DKK,MasterCard,147,Withdrawal,,,57.048,9.919,2624886,Aalborg,277.589,999,8
7,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,70,Diebold Nixdorf,Holstebro,Hostrupsvej,6,7500,56.373,8.625,DKK,MasterCard,5666,Withdrawal,,,56.360,8.616,2620046,Holstebro,
280.150,988,93,4,210,0.000,92,804,Clouds,overcast clouds
2017,December,31,Sunday,23,Active,49,NCR,Bindslev,NÅFÅ,rrebro,18,9881,57.541,10.200,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,ÅFÅesterÅFÅ Duus,ÅFÅesterÅFÅ,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,ÅFÅesterÅFÅ Duus,ÅFÅesterÅFÅ,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,NÅFÅ,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,Clouds,broken clouds
2017,December,31,Sunday,23,Active,37,NCR,Silkeborg,Borgergade,36,8600,56.179,9.552,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-172-31-21-118 ~]$
```