

## 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 \
--target-dir /user/root/SRC_ATM_TRANS \
-m 1
```

```
hadoop@ip-172-31-1-156:~$ sqoop import --connect jdbc:mysql://upgradtest.cyaieic9bmnf.us-east-1.rds.amazonaws.com/testdatabase --username student -password STUDENT123 --table SRC_ATM_TRANS --target-dir /user/root/SRC_ATM_TRANS -m 1
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hbase/lib/client-facing-thirdparty/slf4j-reload4j-1.7.33.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.Reload4jLoggerFactory]
2024-08-02 18:59:09,145 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
2024-08-02 18:59:09,198 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2024-08-02 18:59:09,370 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2024-08-02 18:59:09,370 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.
2024-08-02 18:59:10,333 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'SRC_ATM_TRANS' AS t LIMIT 1
2024-08-02 18:59:10,415 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'SRC_ATM_TRANS' AS t LIMIT 1
2024-08-02 18:59:10,449 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
2024-08-02 18:59:15,267 ERROR orm.CompilationManager: Could not rename /tmp/sqoop-hadoop/compile/9e97c20a23285916c7e542b6b5elf496/SRC_ATM_TRANS.java to /home/hadoop/.SRC_ATM_TRANS.java. Error: Destination '/home/hadoop/.SRC_ATM_TRANS.java' already exists
2024-08-02 18:59:15,267 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-hadoop/compile/9e97c20a23285916c7e542b6b5elf496/SRC_ATM_TRANS.jar
2024-08-02 18:59:15,286 WARN manager.MySQLManager: It looks like you are importing from mysql.
2024-08-02 18:59:15,286 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
2024-08-02 18:59:15,286 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
2024-08-02 18:59:15,286 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
2024-08-02 18:59:15,300 INFO mapreduce.ImportJobBase: Beginning import of SRC_ATM_TRANS
2024-08-02 18:59:15,428 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
2024-08-02 18:59:15,940 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2024-08-02 18:59:16,171 INFO client.DefaultHadoopFailoverProxyProvider: Connecting to ResourceManager at ip-172-31-1-156.ec2.internal/172.31.1.156:8032
2024-08-02 18:59:16,349 INFO client.NHSPProxy: Connecting to Application History server at ip-172-31-1-156.ec2.internal/172.31.1.156:10200
2024-08-02 18:59:16,587 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1722624095438_0001
2024-08-02 18:59:19,120 INFO db.DBInputFormat: Using read committed transaction isolation
2024-08-02 18:59:19,613 INFO mapreduce.JobSubmitter: number of splits:1
2024-08-02 18:59:19,805 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1722624095438_0001
2024-08-02 18:59:19,805 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-08-02 18:59:19,964 INFO conf.Configuration: resource-types.xml not found
2024-08-02 18:59:19,965 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-08-02 18:59:20,386 INFO impl.YarnClientImpl: Submitted application application_1722624095438_0001
2024-08-02 18:59:20,443 INFO mapreduce.Job: The url to track the job: http://ip-172-31-1-156.ec2.internal:20888/proxy/application_1722624095438_0001/
2024-08-02 18:59:20,444 INFO mapreduce.Job: Running job: job_1722624095438_0001
2024-08-02 18:59:28,633 INFO mapreduce.Job: Job job_1722624095438_0001 running in uber mode : false
2024-08-02 18:59:28,634 INFO mapreduce.Job: map 0% reduce 0%
2024-08-02 19:00:05,895 INFO mapreduce.Job: map 100% reduce 0%
```

```
hadoop@ip-172-31-1-156:~$
2024-08-02 19:00:05,895 INFO mapreduce.Job: map 100% reduce 0%
2024-08-02 19:00:06,911 INFO mapreduce.Job: Job job_1722624095438_0001 completed successfully
2024-08-02 19:00:06,985 INFO mapreduce.Job: Counters: 33
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=299304
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=85
  HDFS: Number of bytes written=531214815
  HDFS: Number of read operations=6
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Other local map tasks=1
  Total time spent by all maps in occupied slots (ms)=1676400
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=34925
  Total vcore-milliseconds taken by all map tasks=34925
  Total megabyte-milliseconds taken by all map tasks=53644800
Map-Reduce Framework
  Map input records=2468572
  Map output records=2468572
  Input split bytes=85
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=294
  CPU time spent (ms)=36640
  Physical memory (bytes) snapshot=423960576
  Virtual memory (bytes) snapshot=3194916864
  Total committed heap usage (bytes)=306184192
  Peak Map Physical memory (bytes)=430174208
  Peak Map Virtual memory (bytes)=3194916864
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=531214815
2024-08-02 19:00:06,990 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 51.0389 seconds (9.9259 MB/sec)
2024-08-02 19:00:06,992 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[hadoop@ip-172-31-1-156 ~]$
```

2024-08-02 19:00:06,990 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 51.0389 seconds (9.9259 MB/sec)

2024-08-02 19:00:06,992 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.

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

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

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

**Screenshot of the imported data:**

```
[hadoop@ip-172-31-1-156 ~]$ hadoop fs -ls /user/root/SRC_ATM_TRANS
Found 2 items
-rw-r--r-- 1 hadoop hdfsadmin group 0 2024-08-02 19:00 /user/root/SRC_ATM_TRANS/_SUCCESS
-rw-r--r-- 1 hadoop hdfsadmin group 531214815 2024-08-02 19:00 /user/root/SRC_ATM_TRANS/part-m-00000
[hadoop@ip-172-31-1-156 ~]$
```

```

hadoop@ip-172-31-1-156:~
00,0.138,92,500,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,278.589,999,87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,28,NCR,LÅfÅ,gstÅfÅ,r,ÅfEsterbrogade,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,ÅfEsterbrogade,8,9670,56.967,9.253,DKK,VISA,5568,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,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,,,55.588,12.251,2621218,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,rresun dby,277.589,999,87,6,208,0.000,76,803,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,278.589,999,87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,34,NCR,Skipperen,Vestre Alle,2,9000,57.034,9.908,DKK,MasterCard,147,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,70,Diebold Nixdorf,Holstebro,Hostrupsvej,6,7500,56.373,8.625,DKK,MasterCard,5666,Withdrawal,,,56.360,8.616,2620046,Holstebro,280.150,989,93,4,210,0.000,92,804,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,ÅfEsterÅfÅ, Duus,ÅfEsterÅfÅ,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,ÅfEsterÅfÅ, Duus,ÅfEsterÅfÅ,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-172-31-1-156 ~]$

```