

# 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://*****.us-east-1.rds.amazonaws.com/testdatabase \  
--table SRC_ATM_TRANS \  
--username student --password ***** \  
--target-dir /user/root/atm \  
-m 1
```

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

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

## Screenshot of the imported data:

```
[hadoop@ip-172-31-43-144 mysql-connector-java-8.0.25]$ sqoop import --connect jdbc:mysql://ap-172-31-43-144.ec2.internal:3306/testdatabase --table SRC_ATM_TRANS --username student --password STUDENT123 --target-dir /user/root/atm -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 HCATALOG_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.
24/04/23 14:20:24 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/awx/redshift-jdbc/redshift-jdbc02-1.2.37-1001.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]
24/04/23 14:20:24 WARN ToolBaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
24/04/23 14:20:25 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
24/04/23 14:20:25 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.
24/04/23 14:20:25 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'SRC_ATM_TRANS' AS t LIMIT 1
24/04/23 14:20:25 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'SRC_ATM_TRANS' AS t LIMIT 1
24/04/23 14:20:25 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-hadoop/compile/b04a4b4e6986ee90848ae937821f9c/SRC_ATM_TRANS.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
24/04/23 14:20:26 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-hadoop/compile/b04a4b4e6986ee90848ae937821f9c/SRC_ATM_TRANS.jar
24/04/23 14:20:26 WARN manager.MySQLManager: It looks like you are importing from mysql.
24/04/23 14:20:26 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
24/04/23 14:20:26 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
24/04/23 14:20:26 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
24/04/23 14:20:26 INFO mapreduce.ImportJobBase: Beginning import of SRC_ATM_TRANS
24/04/23 14:20:26 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
24/04/23 14:20:26 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
24/04/23 14:20:26 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-43-144.ec2.internal/172.31.43.144:8032
24/04/23 14:20:26 INFO db.DatabaseInputFormat: Using read committed transaction isolation
24/04/23 14:20:26 INFO mapreduce.JobSubmitter: number of splits:1
24/04/23 14:20:26 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1713881166918_0001
24/04/23 14:20:40 INFO impl.YarnClientImpl: Submitted application application_1713881166918_0001
24/04/23 14:20:41 INFO mapreduce.Job: The url to track the job: http://ap-172-31-43-144.ec2.internal:20888/proxy/application_1713881166918_0001/
24/04/23 14:20:41 INFO mapreduce.Job: Running job: job_1713881166918_0001
24/04/23 14:20:41 INFO mapreduce.Job: Job job_1713881166918_0001 running in uber mode : false
24/04/23 14:20:41 INFO mapreduce.Job: map 0% reduce 0%
24/04/23 14:21:17 INFO mapreduce.Job: map 100% reduce 0%
24/04/23 14:21:20 INFO mapreduce.Job: Job job_1713881166918_0001 completed successfully
24/04/23 14:21:20 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=190526
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=7
    HDFS: Number of bytes written=531214815
    HDFS: Number of read operations=1
    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)=1574016
    Total time spent by all reducers in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=32792
    Total vcore-millisecsd taken by all map tasks=32792
    Total megabyte-millisecsd taken by all map tasks=50368512
  Map-Reduce Framework
    Map input records=3468572
    Map output records=3468572
    Input split bytes=7
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=256
    CPU time spent (ms)=35130
    Physical memory (bytes) snapshot=4068922144
    Virtual memory (bytes) snapshot=3201768448
    Total committed heap usage (bytes)=530055168
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=531214815
24/04/23 14:21:20 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 50.4623 seconds (0.6655 MB/sec)
24/04/23 14:21:20 INFO mapreduce.ImportJobBase: Retrieved 3468572 records.
[hadoop@ip-172-31-43-144 mysql-connector-java-8.0.25]$
```

```
[hadoop@ip-172-31-43-144 mysql-connector-java-8.0.25]$ hadoop fs -ls /user/root/atm
Found 2 items
-rw-r--r-- 1 hadoop hadoop 0 2024-04-23 14:21 /user/root/atm/_SUCCESS
-rw-r--r-- 1 hadoop hadoop 531214815 2024-04-23 14:21 /user/root/atm/part-m-000000
```