

## Data Ingestion from the RDS to HDFS using Sqoop

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

Following is the command us to import the tables form 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
```

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

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

### Screenshot of the imported data:

```
22/06/04 21:16:47 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: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.6.2.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]
22/06/04 21:16:47 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
22/06/04 21:16:47 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
22/06/04 21:16:47 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.
22/06/04 21:16:47 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'SRC_ATM_TRANS' AS t LIMIT 1
22/06/04 21:16:48 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'SRC_ATM_TRANS' AS t LIMIT 1
22/06/04 21:16:48 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-hadoop/compile/115ff673b0976b26481b15380d48edb/SRC_ATM_TRANS.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
22/06/04 21:16:50 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-hadoop/compile/115ff673b0976b26481b15380d48edb/SRC_ATM_TRANS.jar
22/06/04 21:16:50 INFO manager.MySQLManager: It looks like you are importing from mysql.
22/06/04 21:16:50 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
22/06/04 21:16:50 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
22/06/04 21:16:50 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
22/06/04 21:16:50 INFO mapreduce.ImportJobBase: Beginning import of SRC_ATM_TRANS
22/06/04 21:16:50 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
22/06/04 21:16:51 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
22/06/04 21:16:51 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-61-7.ec2.internal:172.31.61.7:8032
22/06/04 21:16:54 INFO jobinputformat: Using read committed transaction isolation
22/06/04 21:16:54 INFO mapreduce.JobSubmitter: number of splits:1
22/06/04 21:16:54 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1654376383354_0001
22/06/04 21:16:54 INFO mapreduce.Job: The url to track the job: http://ip-172-31-61-7.ec2.internal:28888/proxy/application_1654376383354_0001/
22/06/04 21:16:54 INFO mapreduce.Job: Running job: job_1654376383354_0001
22/06/04 21:17:01 INFO mapreduce.Job: Job job_1654376383354_0001 running in uber mode : false
22/06/04 21:17:01 INFO mapreduce.Job: map 0% reduce 0%
22/06/04 21:17:20 INFO mapreduce.Job: map 100% reduce 0%
22/06/04 21:17:25 INFO mapreduce.Job: Job job_1654376383354_0001 completed successfully
22/06/04 21:17:25 INFO mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=198538
  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)=1136960
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=29285
  Total vcore-milliseconds taken by all map tasks=29285
  Total megabyte-milliseconds taken by all map tasks=36350720
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)=170
  CPU time spent (ms)=23170
  Physical memory (bytes) snapshot=712339456
  Virtual memory (bytes) snapshot=3661213952
  Total committed heap usage (bytes)=940872672
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=531214815
22/06/04 21:17:25 INFO mapreduce.ImportJobBase: Transferred 586.6059 MB in 33.7917 seconds (14.992 MB/sec)
22/06/04 21:17:25 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[hadoop@ip-172-31-61-7 ~]$
```