

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 source -m 1
```

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

```
hadoop fs -ls source/
```

Screenshot of the imported data:

```
[root@ip-10-0-0-254 ~]# sqoop import --connect jdbc:mysql://upgradtest.cyaieic9bmnf.us-east-1.rds.amazonaws.com/testdatabase --username student --password STUDENT123 --table SRC_ATM_TRANS --target-dir source -m 1
Warning: /opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/bin/./lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
21/07/02 09:46:01 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.15.1
21/07/02 09:46:01 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
21/07/02 09:46:02 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
21/07/02 09:46:02 INFO tool.CodeGenTool: Beginning code generation
21/07/02 09:46:02 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_ATM_TRANS` AS t LIMIT 1
21/07/02 09:46:03 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_ATM_TRANS` AS t LIMIT 1
21/07/02 09:46:03 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce
Note: /tmp/sqoop-root/compile/4ff9dfa9d0185d5fc0375983556a2af9/SRC_ATM_TRANS.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
21/07/02 09:46:08 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/4ff9dfa9d0185d5fc0375983556a2af9/SRC_ATM_TRANS.jar
21/07/02 09:46:08 WARN manager.MySQLManager: It looks like you are importing from mysql.
21/07/02 09:46:08 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
21/07/02 09:46:08 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
21/07/02 09:46:08 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
21/07/02 09:46:08 INFO mapreduce.ImportJobBase: Beginning import of SRC_ATM_TRANS
21/07/02 09:46:08 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
21/07/02 09:46:10 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
21/07/02 09:46:10 INFO client.RMProxy: Connecting to ResourceManager at ip-10-0-0-254.ec2.internal/10.0.0.254:8032
21/07/02 09:46:22 INFO db.DBInputFormat: Using read committed transaction isolation
21/07/02 09:46:23 INFO mapreduce.JobSubmitter: number of splits:1
21/07/02 09:46:23 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625218999121_0001
21/07/02 09:46:24 INFO impl.YarnClientImpl: Submitted application application_1625218999121_0001
21/07/02 09:46:24 INFO mapreduce.Job: The url to track the job: http://ip-10-0-0-254.ec2.internal:8088/proxy/application_1625218999121_0001/
```

```
[root@ip-10-0-0-254 ~]# sqoop import --connect jdbc:mysql://upgradtest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase
--username student --password STUDENT123 --table SRC_ATM_TRANS --target-dir source -m 1
Warning: /opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/bin/../lib/sqoop/./accumulo does not exist! Accumulo imports will
fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
21/07/02 09:46:01 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.15.1
21/07/02 09:46:01 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
21/07/02 09:46:02 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
21/07/02 09:46:02 INFO tool.CodeGenTool: Beginning code generation
21/07/02 09:46:02 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_ATM_TRANS` AS t LIMIT 1
21/07/02 09:46:03 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_ATM_TRANS` AS t LIMIT 1
21/07/02 09:46:03 INFO orm.CompilationManager: HADOOP MAPRED_HOME is /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce
Note: /tmp/sqoop-root/compile/4ff9dfa9d0185d5fc0375983556a2af9/SRC_ATM_TRANS.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
21/07/02 09:46:08 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/4ff9dfa9d0185d5fc0375983556a2af9/SRC_
ATM_TRANS.jar
21/07/02 09:46:08 WARN manager.MySQLManager: It looks like you are importing from mysql.
21/07/02 09:46:08 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
21/07/02 09:46:08 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
21/07/02 09:46:08 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
21/07/02 09:46:08 INFO mapreduce.ImportJobBase: Beginning import of SRC_ATM_TRANS
21/07/02 09:46:08 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
21/07/02 09:46:10 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
21/07/02 09:46:10 INFO client.RMProxy: Connecting to ResourceManager at ip-10-0-0-254.ec2.internal/10.0.0.254:8032
21/07/02 09:46:22 INFO db.DBInputFormat: Using read committed transaction isolation
21/07/02 09:46:23 INFO mapreduce.JobSubmitter: number of splits:1
21/07/02 09:46:23 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625218999121_0001
21/07/02 09:46:24 INFO impl.YarnClientImpl: Submitted application application_1625218999121_0001
21/07/02 09:46:24 INFO mapreduce.Job: The url to track the job: http://ip-10-0-0-254.ec2.internal:8088/proxy/application_16252
18999121_0001/
```

```
[root@ip-10-0-0-254 ~]# sqoop import --connect jdbc:mysql://upgradtest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase
--username student --password STUDENT123 --table SRC_ATM_TRANS --target-dir source -m 1
Warning: /opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/bin/../lib/sqoop/./accumulo does not exist! Accumulo imports will
fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
21/07/02 09:46:01 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.15.1
21/07/02 09:46:01 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
21/07/02 09:46:02 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
21/07/02 09:46:02 INFO tool.CodeGenTool: Beginning code generation
21/07/02 09:46:02 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_ATM_TRANS` AS t LIMIT 1
21/07/02 09:46:03 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_ATM_TRANS` AS t LIMIT 1
21/07/02 09:46:03 INFO orm.CompilationManager: HADOOP MAPRED_HOME is /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce
Note: /tmp/sqoop-root/compile/4ff9dfa9d0185d5fc0375983556a2af9/SRC_ATM_TRANS.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
21/07/02 09:46:08 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/4ff9dfa9d0185d5fc0375983556a2af9/SRC_
ATM_TRANS.jar
21/07/02 09:46:08 WARN manager.MySQLManager: It looks like you are importing from mysql.
21/07/02 09:46:08 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
21/07/02 09:46:08 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
21/07/02 09:46:08 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
21/07/02 09:46:08 INFO mapreduce.ImportJobBase: Beginning import of SRC_ATM_TRANS
21/07/02 09:46:08 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
21/07/02 09:46:10 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
21/07/02 09:46:10 INFO client.RMProxy: Connecting to ResourceManager at ip-10-0-0-254.ec2.internal/10.0.0.254:8032
21/07/02 09:46:22 INFO db.DBInputFormat: Using read committed transaction isolation
21/07/02 09:46:23 INFO mapreduce.JobSubmitter: number of splits:1
21/07/02 09:46:23 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625218999121_0001
21/07/02 09:46:24 INFO impl.YarnClientImpl: Submitted application application_1625218999121_0001
21/07/02 09:46:24 INFO mapreduce.Job: The url to track the job: http://ip-10-0-0-254.ec2.internal:8088/proxy/application_16252
18999121_0001/
```