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://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase \
> --table SRC_ATM_TRANS \
> --username student --password STUDENT123 \
> --target-dir /user/root/assignment \
> -m 1 \
> --as-parquetfile
```

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

```
hadoop fs -cat /user/root/assignment//a6d88055-0b19-4572-8e63-ce01866a70be.parquet | wc -l
hadoop fs -cat /user/root/assignment//a6d88055-0b19-4572-8e63-ce01866a70be.parquet | head
-n 10
```

Screenshot of the imported data:

```
Try --help for usage instructions.
[hadoop@ip-10-0-6-27 mysql-connector-java-8.0.25]$ sqoop import \
> --connect jdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase \
> --table SRC_ATM_TRANS \
> --username student --password STUDENT123 \
> --target-dir /user/root/assignment \
> -m 1 \
> --as-parquetfile
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.
23/06/19 19:34:49 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]
23/06/19 19:34:49 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/06/19 19:34:49 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
23/06/19 19:34:49 INFO tool.CodeGenTool: Beginning code generation
23/06/19 19:34:49 INFO tool.CodeGenTool: Will generate java class as codegen_SRC_ATM_TRANS
Loading class `com.mysql.jdbc.Driver'. This is deprecated. The new driver class is `com.mysql.cj.jdbc.Driver'. The driver is automatically registered v
nual loading of the driver class is generally unnecessary.
23/06/19 19:34:50 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_ATM_TRANS` AS t LIMIT 1
23/06/19 19:34:50 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_ATM_TRANS` AS t LIMIT 1
23/06/19 19:34:50 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-hadoop/compile/96f27ab8718abe87ad5f3245fe6dc478/codegen_SRC_ATM_TRANS.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
23/06/19 19:34:53 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-hadoop/compile/96f27ab8718abe87ad5f3245fe6dc478/codegen_SRC_ATM_TRANS.jar
23/06/19 19:34:53 WARN manager.MySQLManager: It looks like you are importing from mysql.
                FILE: Number of bytes read=0
                FILE: Number of bytes written=193713
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=41972
                HDFS: Number of bytes written=44699221
                HDFS: Number of read operations=50
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=10
        Job Counters
                Launched map tasks=1
                Other local map tasks=1
                Total time spent by all maps in occupied slots (ms)=2717520
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=56615
                Total vcore-milliseconds taken by all map tasks=56615
                Total megabyte-milliseconds taken by all map tasks=86960640
        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)=752
                CPU time spent (ms)=59190
                Physical memory (bytes) snapshot=704606208
                Virtual memory (bytes) snapshot=3311304704
                Total committed heap usage (bytes)=558366720
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=0
23/06/19 19:36:15 INFO mapreduce.ImportJobBase: Transferred 42.6285 MB in 78.9925 seconds (552.6042 KB/sec)
23/06/19 19:36:15 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[hadoop@ip-10-0-6-27 mysql-connector-java-8.0.25]$ hadoop fs -ls
[hadoop@ip-10-0-6-27 mysql-connector-java-8.0.25]$ hadoop fs -ls /user/root/assignment
Found 3 items
drwxr-xr-x   - hadoop hadoop          0 2023-06-19 19:34 /user/root/assignment/.metadata
drwxr-xr-x   - hadoop hadoop          0 2023-06-19 19:36 /user/root/assignment/.signals
-rw-r--r--   1 hadoop hadoop   44688887 2023-06-19 19:36 /user/root/assignment/a6d88055-0b19-4572-8e63-ce01866a70be.parquet
[hadoop@ip-10-0-6-27 mysql-connector-java-8.0.25]$ ^C
[hadoop@ip-10-0-6-27 mysql-connector-java-8.0.25]$  /user/root/assignment/a6d88055-0b19-4572-8e63-ce01866a70be.parquet
```