Sqoop Data Ingestion Tasks

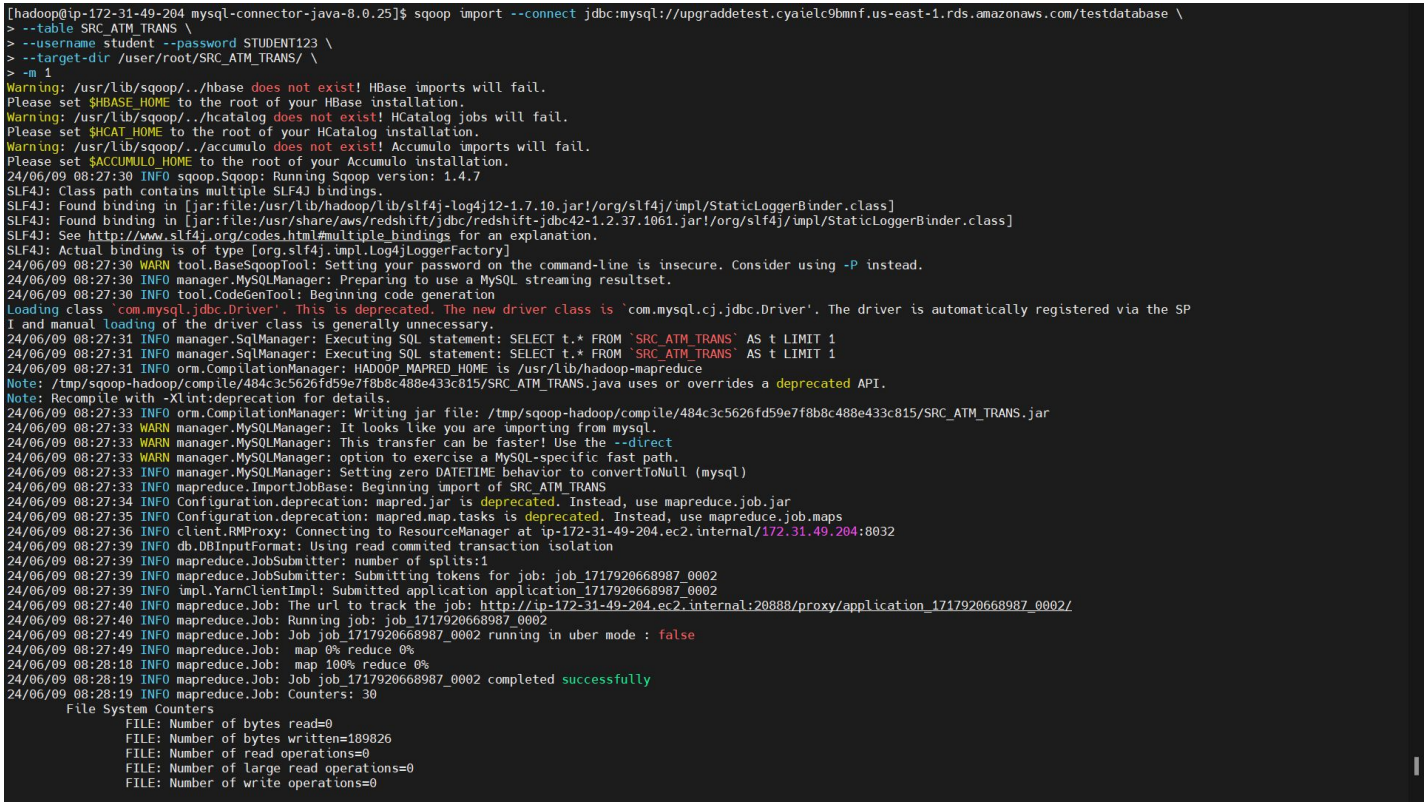
Install MySQL connector

- Switch to sudo user using “sudo -i”
- Perform the following commands.
  - wget https://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz
  - tar -xvf mysql-connector-java-8.0.25.tar.gz
  - cd mysql-connector-java-8.0.25/
  - sudo cp mysql-connector-java-8.0.25.jar /usr/lib/sqoop/lib/

Sqoop command to import data from RDS to HDFS

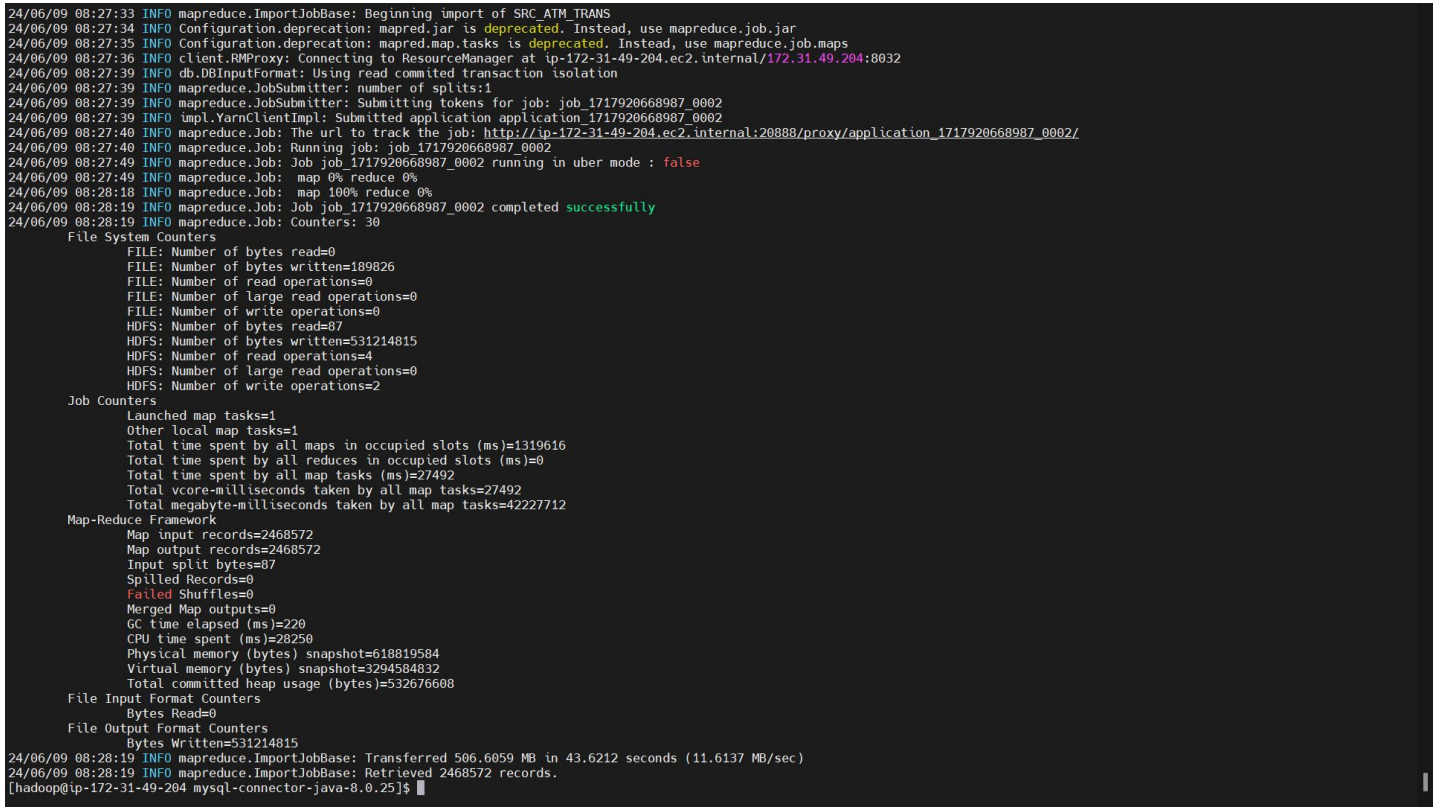
```
sqoop import --connect
jdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase \
--table SRC_ATM_TRANS \
--username student --password STUDENT123 \
--target-dir /user/root/SRC_ATM_TRANS/ \
-m 1
```

Screenshot of the Sqoop process - 1



```
[hadoop@ip-172-31-49-204 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/SRC_ATM_TRANS/ \
> -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 $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.
24/06/09 08:27:30 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]
24/06/09 08:27:30 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
24/06/09 08:27:30 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
24/06/09 08:27:30 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 SP
I and manual loading of the driver class is generally unnecessary.
24/06/09 08:27:31 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_ATM_TRANS` AS t LIMIT 1
24/06/09 08:27:31 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_ATM_TRANS` AS t LIMIT 1
24/06/09 08:27:31 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-hadoop/compile/484c3c5626fd59e7f8b8c488e433c815/SRC_ATM_TRANS.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
24/06/09 08:27:33 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-hadoop/compile/484c3c5626fd59e7f8b8c488e433c815/SRC_ATM_TRANS.jar
24/06/09 08:27:33 WARN manager.MySQLManager: It looks like you are importing from mysql.
24/06/09 08:27:33 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
24/06/09 08:27:33 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
24/06/09 08:27:33 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
24/06/09 08:27:33 INFO mapreduce.ImportJobBase: Beginning import of SRC_ATM_TRANS
24/06/09 08:27:34 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
24/06/09 08:27:35 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
24/06/09 08:27:36 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-49-204.ec2.internal/172.31.49.204:8032
24/06/09 08:27:39 INFO db.DBInputFormat: Using read commited transaction isolation
24/06/09 08:27:39 INFO mapreduce.JobSubmitter: number of splits:1
24/06/09 08:27:39 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1717920668987_0002
24/06/09 08:27:39 INFO impl.YarnClientImpl: Submitted application application_1717920668987_0002
24/06/09 08:27:40 INFO mapreduce.Job: The url to track the job: http://ip-172-31-49-204.ec2.internal:20888/proxy/application_1717920668987_0002/
24/06/09 08:27:40 INFO mapreduce.Job: Running job: job_1717920668987_0002
24/06/09 08:27:49 INFO mapreduce.Job: Job job_1717920668987_0002 running in uber mode : false
24/06/09 08:27:49 INFO mapreduce.Job:  map 0% reduce 0%
24/06/09 08:28:18 INFO mapreduce.Job:  map 100% reduce 0%
24/06/09 08:28:19 INFO mapreduce.Job: Job job_1717920668987_0002 completed successfully
24/06/09 08:28:19 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=189826
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
```

Screenshot of the Sqoop process - 2



Command 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 | wc -l

hadoop fs -cat /user/root/SRC_ATM_TRANS/part-m-00000 | head -n 5

Screenshot of the imported data

```
[hadoop@ip-172-31-49-204 mysql-connector-java-8.0.25]$ hadoop fs -ls /user/root/SRC_ATM_TRANS/
Found 2 items
-rw-r--r--   1 hadoop hadoop          0 2024-06-09 08:28 /user/root/SRC_ATM_TRANS/_SUCCESS
-rw-r--r--   1 hadoop hadoop  531214815 2024-06-09 08:28 /user/root/SRC_ATM_TRANS/part-m-00000
[hadoop@ip-172-31-49-204 mysql-connector-java-8.0.25]$ hadoop fs -cat /user/root/SRC_ATM_TRANS/part-m-00000 | wc -l
2468572
[hadoop@ip-172-31-49-204 mysql-connector-java-8.0.25]$ hadoop fs -cat /user/root/SRC_ATM_TRANS/part-m-00000 | head -n 5
2017,January,1,Sunday,0,Active,1,NCR,NÃƒÂ¦stved,Farimagsvej,8,4700,55.233,11.763,DKK,MasterCard,5643,Withdrawal,,,55.230,11.761,2616038,Naestved,281.150,1014,87
,7,260,0.215,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,2,NCR,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,MasterCard,1764,Withdrawal,,,57.048,9.935,2616235,NÃƒÂ¸rresundby,280.640,102
0,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,2,NCR,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,VISA,1891,Withdrawal,,,57.048,9.935,2616235,NÃƒÂ¸rresundby,280.640,1020,93,9
,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,3,NCR,Ikast,RÃƒÂ¥dhusstrÃƒÂ¦det,12,7430,56.139,9.154,DKK,VISA,4166,Withdrawal,,,56.139,9.158,2619426,Ikast,281.150,1011,100,6,2
40,0.000,75,300,Drizzle,light intensity drizzle
2017,January,1,Sunday,0,Active,4,NCR,Svogerslev,BrÃƒÂ¸nsager,1,4000,55.634,12.018,DKK,MasterCard,5153,Withdrawal,,,55.642,12.080,2614481,Roskilde,280.610,1014,8
7,7,260,0.000,88,701,Mist,mist
```