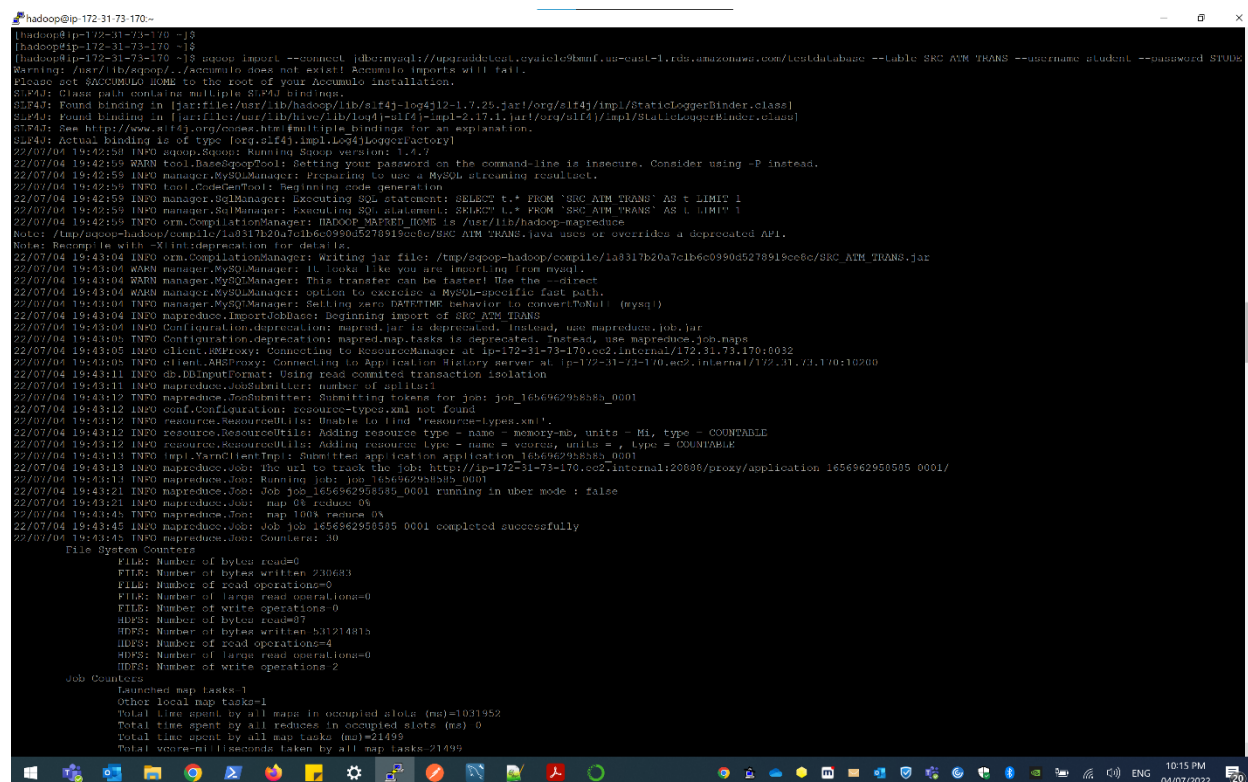


Data Ingestion from the RDS to HDFS using Sqoop

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

```
sqoop import \
--connect jdbc:mysql://upgraddetest.cyaiecl9bmnf.us-east-
1.rds.amazonaws.com/testdatabase \
--table SRC_ATM_TRANS \
--username student --password STUDENT123 \
--target-dir /user/root/SRC_ATM_TRANS \
-m 1
```



```
hadoop@ip-172-31-73-170:~$
[hadoop@ip-172-31-73-170:~]$
[hadoop@ip-172-31-73-170:~]$ sqoop import --connect jdbc:mysql://upgraddetest.cyaiecl9bmnf.us-east-1.rds.amazonaws.com/testdatabase --table SRC_ATM_TRANS --username student --password STUDENT123
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hive/lib/slf4j-log4j12-1.7.25.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/faq.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
22/07/04 19:42:58 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
22/07/04 19:42:59 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
22/07/04 19:42:59 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
22/07/04 19:42:59 INFO tool.CodeGenTool: Beginning code generation.
22/07/04 19:42:59 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'SRC_ATM_TRANS' AS t LIMIT 1
22/07/04 19:42:59 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'SRC_ATM_TRANS' AS t LIMIT 1
22/07/04 19:42:59 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-hadoop/compile/1a0317b20a7c1b6c0990d5278919cc6c/SRC_ATM_TRANS.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
22/07/04 19:43:04 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-hadoop/compile/1a0317b20a7c1b6c0990d5278919cc6c/SRC_ATM_TRANS.jar
22/07/04 19:43:04 WARN manager.MySQLManager: It looks like you are importing from mysql.
22/07/04 19:43:04 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
22/07/04 19:43:04 INFO manager.MySQLManager: option to exercise a MySQL-specific fast path.
22/07/04 19:43:04 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
22/07/04 19:43:04 INFO mapreduce.ImportJobBase: Beginning import of SRC_ATM_TRANS
22/07/04 19:43:04 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
22/07/04 19:43:05 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
22/07/04 19:43:05 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-73-170.ec2.internal:/172.31.73.170:8032
22/07/04 19:43:05 INFO client.AHSProxy: Connecting to Application History server at ip-172-31-73-170.ec2.internal/172.31.73.170:10200
22/07/04 19:43:11 INFO db.InputFormat: Using read committed transaction isolation
22/07/04 19:43:11 INFO mapreduce.JobSubmitter: number of splits:1
22/07/04 19:43:12 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_166662956505_0001
22/07/04 19:43:12 INFO conf.Configuration: resource-types.xml not found
22/07/04 19:43:12 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
22/07/04 19:43:12 INFO resource.ResourceUtils: Adding resource type = name = memory-mb, units = M, type = COUNTABLE
22/07/04 19:43:12 INFO resource.ResourceUtils: Adding resource type = name = vcores, units = v, type = COUNTABLE
22/07/04 19:43:13 INFO (api.YarnClientImpl): Submitted application application_166662956505_0001
22/07/04 19:43:13 INFO mapreduce.Job: The url to track the job: http://ip-172-31-73-170.ec2.internal:20806/proxy/application_166662956505_0001/
22/07/04 19:43:13 INFO mapreduce.Job: Running job: job_166662956505_0001
22/07/04 19:43:21 INFO mapreduce.Job: Job job_166662956505_0001 running in uber mode : false
22/07/04 19:43:21 INFO mapreduce.Job: map 0% reduce 0%
22/07/04 19:43:45 INFO mapreduce.Job: map 100% reduce 0%
22/07/04 19:43:45 INFO mapreduce.Job: Job job_166662956505_0001 completed successfully
22/07/04 19:43:45 INFO mapreduce.Job: Counters: 30
File System Counters
FILE: Number of bytes read=0
FILE: Number of bytes written=236683
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=0
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)=1031952
Total time spent by all reducers in occupied slots (ms)=0
Total time spent by all map tasks (ms)=21499
Total time spent by all map tasks (ms)=21499
Total time spent by all map tasks (ms)=21499
```

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

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

```
hadoop@ip-172-31-73-170:~$ hadoop fs -ls /user/root/SRC_ATM_TRANS
Found 2 items
-rw-r--r-- 1 hadoop hdfsadmin group 0 2022-07-04 19:43 /user/root/SRC_ATM_TRANS/ SUCCESS
-rw-r--r-- 1 hadoop hdfsadmin group 531214815 2022-07-04 19:43 /user/root/SRC_ATM_TRANS/part-m-00000
hadoop@ip-172-31-73-170:~$
hadoop@ip-172-31-73-170:~$
hadoop@ip-172-31-73-170:~$
```

© Copyright. upGrad Education Pvt. Ltd. All rights reserved