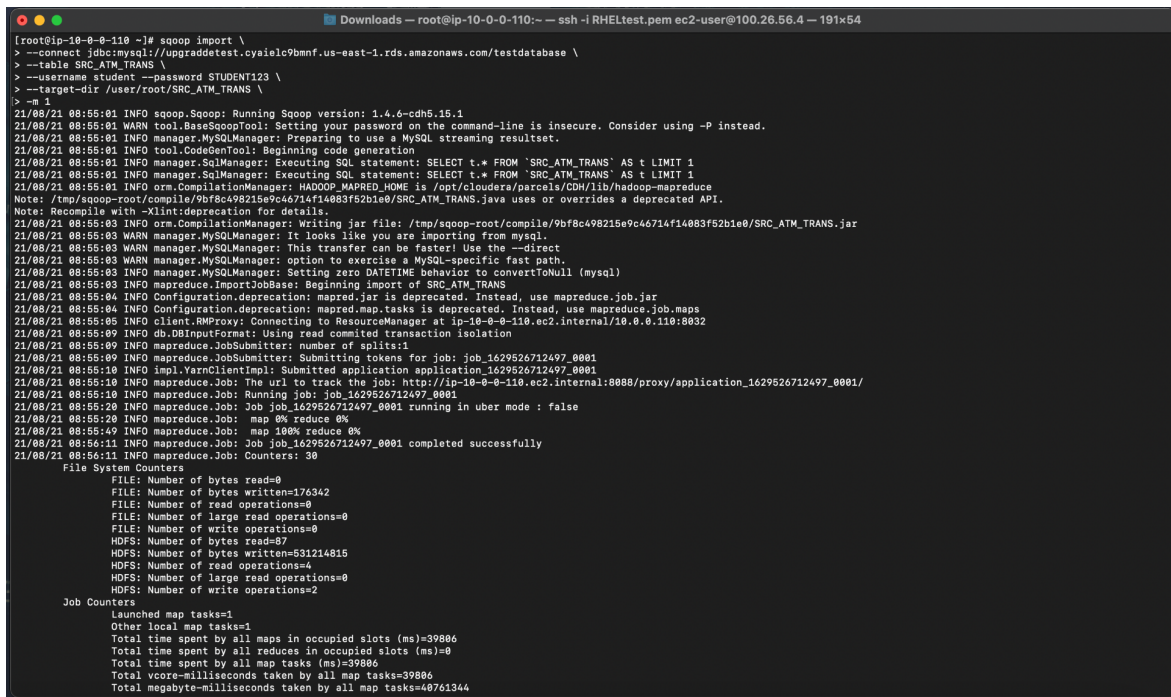


SQOOP COMMANDS:

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.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
```



```
[root@ip-10-0-0-110 ~]# sqoop import \  
> --connect jdbc:mysql://upgraddetest.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  
21/08/21 08:55:04 INFO sqoop.Sqoop: Running Sqoop version: 1.4.4-cdh5.15.1  
21/08/21 08:55:04 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.  
21/08/21 08:55:04 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.  
21/08/21 08:55:04 INFO tool.CodeGenTool: Beginning code generation  
21/08/21 08:55:04 INFO manager.MySQLManager: Executing SQL statement: SELECT t.* FROM 'SRC_ATM_TRANS' AS t LIMIT 1  
21/08/21 08:55:04 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'SRC_ATM_TRANS' AS t LIMIT 1  
21/08/21 08:55:04 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce  
Note: /tmp/sqoop-root/compile/9bf8c498215e9c46714f14083f52b1e0/SRC_ATM_TRANS.java uses or overrides a deprecated API.  
Note: Recompile with -Xlint:deprecation for details.  
21/08/21 08:55:04 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/9bf8c498215e9c46714f14083f52b1e0/SRC_ATM_TRANS.jar  
21/08/21 08:55:03 WARN manager.MySQLManager: It looks like you are importing from mysql.  
21/08/21 08:55:03 WARN manager.MySQLManager: This transfer can be faster! Use the --direct  
21/08/21 08:55:03 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.  
21/08/21 08:55:03 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)  
21/08/21 08:55:03 INFO mapreduce.ImportJobBase: Beginning import of SRC_ATM_TRANS  
21/08/21 08:55:04 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar  
21/08/21 08:55:04 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps  
21/08/21 08:55:05 INFO client.RMProxy: Connecting to ResourceManager at ip-10-0-0-110.ec2.internal/10.0.0.110:8032  
21/08/21 08:55:09 INFO db.DbInputFormat: Using read committed transaction isolation  
21/08/21 08:55:09 INFO mapreduce.JobSubmitter: number of splits:1  
21/08/21 08:55:09 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1629526712497_0001  
21/08/21 08:55:10 INFO impl.YarnClientImpl: Submitted application application_1629526712497_0001  
21/08/21 08:55:10 INFO mapreduce.Job: The url to track the job: http://ip-10-0-0-110.ec2.internal:8080/proxy/application_1629526712497_0001/  
21/08/21 08:55:10 INFO mapreduce.Job: Running job: job_1629526712497_0001  
21/08/21 08:55:20 INFO mapreduce.Job: Job job_1629526712497_0001 running in uber mode : false  
21/08/21 08:55:20 INFO mapreduce.Job: map 0% reduce 0%  
21/08/21 08:55:40 INFO mapreduce.Job: map 100% reduce 0%  
21/08/21 08:56:11 INFO mapreduce.Job: Job job_1629526712497_0001 completed successfully  
21/08/21 08:56:11 INFO mapreduce.Job: Counters: 30  
File System Counters  
FILE: Number of bytes read=0  
FILE: Number of bytes written=176342  
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)=39886  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=39886  
Total vcore-milliseconds taken by all map tasks=39886  
Total megabyte-milliseconds taken by all map tasks=40761344
```

```
21/08/21 08:55:01 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_ATM_TRANS` AS t LIMIT 1
21/08/21 08:55:01 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_ATM_TRANS` AS t LIMIT 1
21/08/21 08:55:01 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce
Note: /tmp/sqoop-root/compile/9bf8c498215e9c46714f14083f52b1e0/SRC_ATM_TRANS.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
21/08/21 08:55:03 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/9bf8c498215e9c46714f14083f52b1e0/SRC_ATM_TRANS.jar
21/08/21 08:55:03 WARN manager.MySQLManager: It looks like you are importing from mysql.
21/08/21 08:55:03 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
21/08/21 08:55:03 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
21/08/21 08:55:03 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
21/08/21 08:55:03 INFO mapreduce.ImportJobBase: Beginning import of SRC_ATM_TRANS
21/08/21 08:55:04 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
21/08/21 08:55:04 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
21/08/21 08:55:05 INFO client.RMPProxy: Connecting to ResourceManager at ip-10-0-0-110.ec2.internal/10.0.0.110:8032
21/08/21 08:55:09 INFO db.DBInputFormat: Using read committed transaction isolation
21/08/21 08:55:09 INFO mapreduce.JobSubmitter: number of splits:1
21/08/21 08:55:09 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1629526712497_0001
21/08/21 08:55:10 INFO impl.VarnClientImpl: Submitted application application_1629526712497_0001
21/08/21 08:55:10 INFO mapreduce.Job: The url to track the job: http://ip-10-0-0-110.ec2.internal:8080/proxy/application_1629526712497_0001/
21/08/21 08:55:10 INFO mapreduce.Job: Running job: job_1629526712497_0001
21/08/21 08:55:20 INFO mapreduce.Job: Job job_1629526712497_0001 running in uber mode : false
21/08/21 08:55:20 INFO mapreduce.Job: map 0% reduce 0%
21/08/21 08:55:49 INFO mapreduce.Job: map 100% reduce 0%
21/08/21 08:56:11 INFO mapreduce.Job: Job job_1629526712497_0001 completed successfully
21/08/21 08:56:11 INFO mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=176342
  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)=39806
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=39806
  Total vcore-milliseconds taken by all map tasks=39806
  Total megabyte-milliseconds taken by all map tasks=40761344
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)=183
  CPU time spent (ms)=28120
  Physical memory (bytes) snapshot=427929600
  Virtual memory (bytes) snapshot=2798385280
  Total committed heap usage (bytes)=382738240
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=531214815
21/08/21 08:56:11 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 66.9346 seconds (7.5687 MB/sec)
21/08/21 08:56:11 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-10-0-0-110 ~]#
```

```
21/08/21 08:55:04 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
21/08/21 08:55:05 INFO client.RMPProxy: Connecting to ResourceManager at ip-10-0-0-110.ec2.internal/10.0.0.110:8032
21/08/21 08:55:09 INFO db.DBInputFormat: Using read committed transaction isolation
21/08/21 08:55:09 INFO mapreduce.JobSubmitter: number of splits:1
21/08/21 08:55:09 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1629526712497_0001
21/08/21 08:55:10 INFO impl.VarnClientImpl: Submitted application application_1629526712497_0001
21/08/21 08:55:10 INFO mapreduce.Job: The url to track the job: http://ip-10-0-0-110.ec2.internal:8080/proxy/application_1629526712497_0001/
21/08/21 08:55:10 INFO mapreduce.Job: Running job: job_1629526712497_0001
21/08/21 08:55:20 INFO mapreduce.Job: Job job_1629526712497_0001 running in uber mode : false
21/08/21 08:55:20 INFO mapreduce.Job: map 0% reduce 0%
21/08/21 08:55:49 INFO mapreduce.Job: map 100% reduce 0%
21/08/21 08:56:11 INFO mapreduce.Job: Job job_1629526712497_0001 completed successfully
21/08/21 08:56:11 INFO mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=176342
  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)=39806
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=39806
  Total vcore-milliseconds taken by all map tasks=39806
  Total megabyte-milliseconds taken by all map tasks=40761344
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)=183
  CPU time spent (ms)=28120
  Physical memory (bytes) snapshot=427929600
  Virtual memory (bytes) snapshot=2798385280
  Total committed heap usage (bytes)=382738240
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=531214815
21/08/21 08:56:11 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 66.9346 seconds (7.5687 MB/sec)
21/08/21 08:56:11 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-10-0-0-110 ~]# hadoop fs -ls /user/root/SRC_ATM_TRANS \
>
Found 2 items
-rw-r--r-- 3 root supergroup 0 2021-08-21 08:56 /user/root/SRC_ATM_TRANS/_SUCCESS
-rw-r--r-- 3 root supergroup 531214815 2021-08-21 08:55 /user/root/SRC_ATM_TRANS/part-m-000000
[root@ip-10-0-0-110 ~]# hadoop fs -ls /user/root/SRC_ATM_TRANS \
> exit
-bash: /user/root/SRC_ATM_TRANS: No such file or directory
[root@ip-10-0-0-110 ~]# hadoop fs -ls /user/root/SRC_ATM_TRANS
Found 2 items
-rw-r--r-- 3 root supergroup 0 2021-08-21 08:56 /user/root/SRC_ATM_TRANS/_SUCCESS
-rw-r--r-- 3 root supergroup 531214815 2021-08-21 08:55 /user/root/SRC_ATM_TRANS/part-m-000000
[root@ip-10-0-0-110 ~]#
```

