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 --query 'select * from testdatabase.SRC_ATM_TRANS where $CONDITIONS' --username "student" --password "STUDENT123" --target-dir "/user/root/bank_atm" --m 1

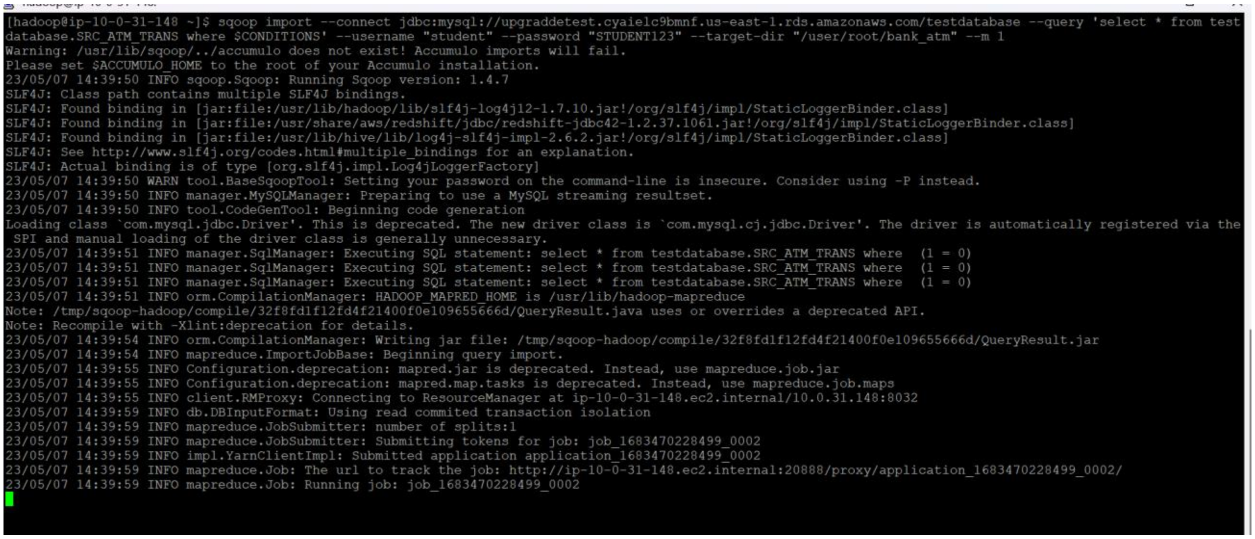
Command used to see the list of imported data in HDFS:

hadoop fs -mv /user/root/bank_atm/part-m-00000 /user/livy

hadoop fs -ls /user/livy/part-m-00000

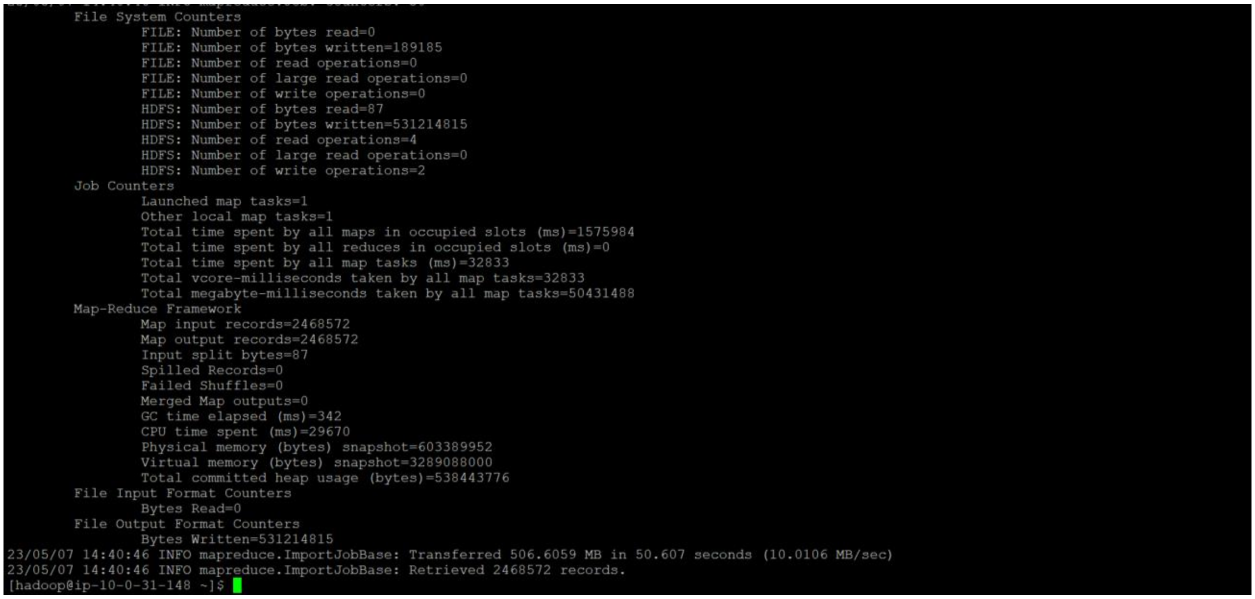
Screenshot of the imported data:

1. **Screenshot of data import sqoop command**



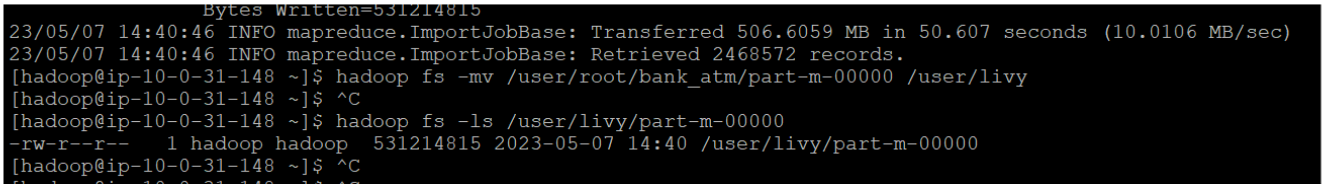
```
[hadoop@ip-10-0-31-148 ~]$ sqoop import --connect jdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase --query 'select * from test
database.SRC_ATM_TRANS where $CONDITIONS' --username "student" --password "STUDENT123" --target-dir "/user/root/bank_atm" --m 1
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
23/05/07 14:39:50 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: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.6.2.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]
23/05/07 14:39:50 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/05/07 14:39:50 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
23/05/07 14:39:50 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
 SPI and manual loading of the driver class is generally unnecessary.
23/05/07 14:39:51 INFO manager.SqlManager: Executing SQL statement: select * from testdatabase.SRC_ATM_TRANS where  (1 = 0)
23/05/07 14:39:51 INFO manager.SqlManager: Executing SQL statement: select * from testdatabase.SRC_ATM_TRANS where  (1 = 0)
23/05/07 14:39:51 INFO manager.SqlManager: Executing SQL statement: select * from testdatabase.SRC_ATM_TRANS where  (1 = 0)
23/05/07 14:39:51 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-hadoop/compile/32f8fd1f12fd4f21400f0e109655666d/QueryResult.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
23/05/07 14:39:54 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-hadoop/compile/32f8fd1f12fd4f21400f0e109655666d/QueryResult.jar
23/05/07 14:39:54 INFO mapreduce.ImportJobBase: Beginning query import.
23/05/07 14:39:55 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
23/05/07 14:39:55 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
23/05/07 14:39:55 INFO client.RMProxy: Connecting to ResourceManager at ip-10-0-31-148.ec2.internal/10.0.31.148:8032
23/05/07 14:39:59 INFO db.DBInputFormat: Using read commited transaction isolation
23/05/07 14:39:59 INFO mapreduce.JobSubmitter: number of splits:1
23/05/07 14:39:59 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1683470228499_0002
23/05/07 14:39:59 INFO impl.YarnClientImpl: Submitted application application_1683470228499_0002
23/05/07 14:39:59 INFO mapreduce.Job: The url to track the job: http://ip-10-0-31-148.ec2.internal:20888/proxy/application_1683470228499_0002/
23/05/07 14:39:59 INFO mapreduce.Job: Running job: job_1683470228499_0002
```

2. Screenshot of the data imported



```
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=189185
                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)=1575984
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=32833
                Total vcore-milliseconds taken by all map tasks=32833
                Total megabyte-milliseconds taken by all map tasks=50431488
        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)=342
                CPU time spent (ms)=29670
                Physical memory (bytes) snapshot=603389952
                Virtual memory (bytes) snapshot=3289088000
                Total committed heap usage (bytes)=538443776
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=531214815
23/05/07 14:40:46 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 50.607 seconds (10.0106 MB/sec)
23/05/07 14:40:46 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[hadoop@ip-10-0-31-148 ~]$
```

3. Screen shot of the data moved to livy location



```
                Bytes Written=531214815
23/05/07 14:40:46 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 50.607 seconds (10.0106 MB/sec)
23/05/07 14:40:46 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[hadoop@ip-10-0-31-148 ~]$ hadoop fs -mv /user/root/bank_atm/part-m-00000 /user/livy
[hadoop@ip-10-0-31-148 ~]$ ^C
[hadoop@ip-10-0-31-148 ~]$ hadoop fs -ls /user/livy/part-m-00000
-rw-r--r--   1 hadoop hadoop  531214815 2023-05-07 14:40 /user/livy/part-m-00000
[hadoop@ip-10-0-31-148 ~]$ ^C
```