Data Ingestion from the RDS to HDFS using Sqoop

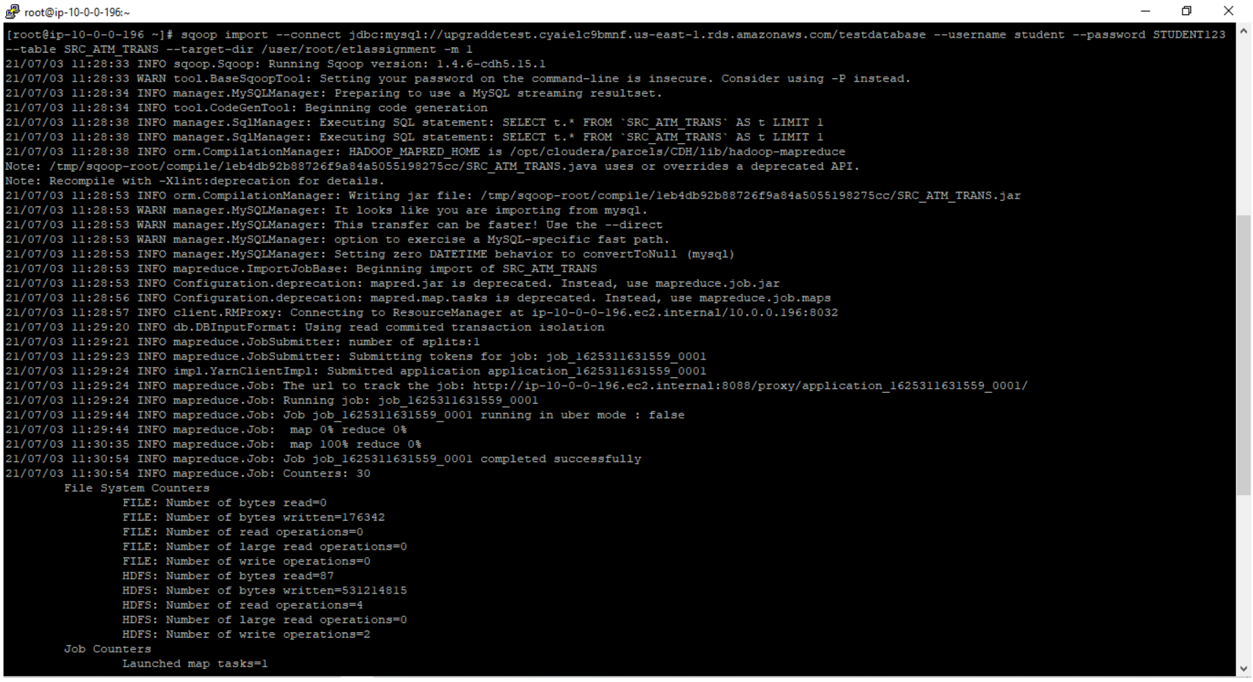
Sqoop Import command used for importing table from RDS to HDFS:

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

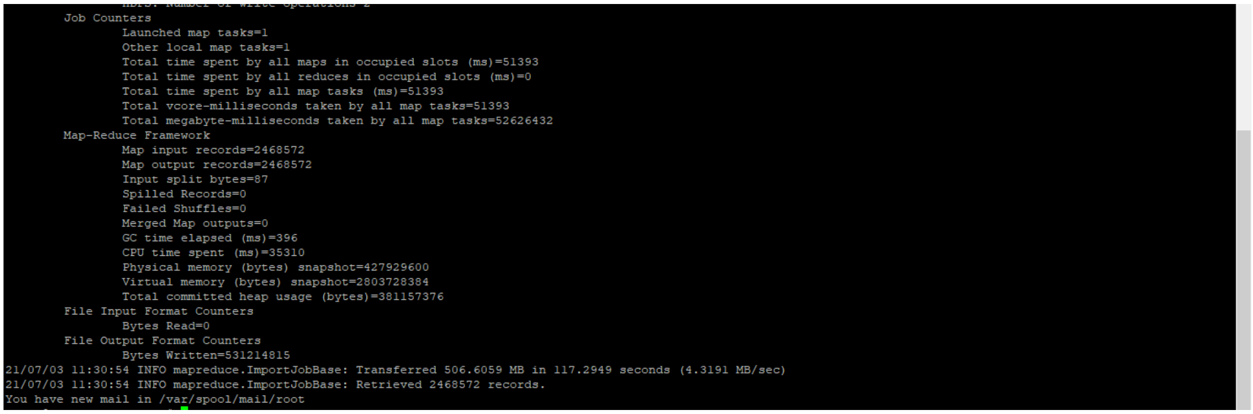
Command used to see the list of imported data in HDFS:

```
hdfs dfs -ls
```

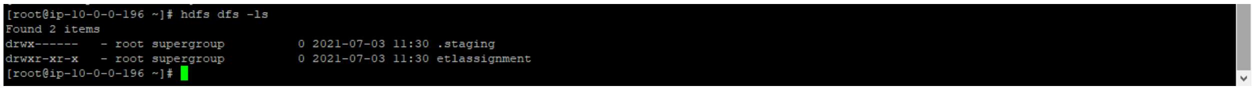
Screenshot of the imported data:



```
[root@ip-10-0-0-196 ~]# sqoop import --connect jdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase --username student --password STUDENT123
--table SRC_ATM_TRANS --target-dir /user/root/etlassignment -m 1
21/07/03 11:28:33 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.15.1
21/07/03 11:28:33 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
21/07/03 11:28:34 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
21/07/03 11:28:34 INFO tool.CodeGenTool: Beginning code generation
21/07/03 11:28:38 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_ATM_TRANS` AS t LIMIT 1
21/07/03 11:28:38 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `SRC_ATM_TRANS` AS t LIMIT 1
21/07/03 11:28:38 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce
Note: /tmp/sqoop-root/compile/1eb4db92b88726f9a84a5055198275cc/SRC_ATM_TRANS.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
21/07/03 11:28:53 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/1eb4db92b88726f9a84a5055198275cc/SRC_ATM_TRANS.jar
21/07/03 11:28:53 WARN manager.MySQLManager: It looks like you are importing from mysql.
21/07/03 11:28:53 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
21/07/03 11:28:53 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
21/07/03 11:28:53 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
21/07/03 11:28:53 INFO mapreduce.ImportJobBase: Beginning import of SRC_ATM_TRANS
21/07/03 11:28:53 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
21/07/03 11:28:56 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
21/07/03 11:28:57 INFO client.RMProxy: Connecting to ResourceManager at ip-10-0-0-196.ec2.internal/10.0.0.196:8032
21/07/03 11:29:20 INFO db.DBInputFormat: Using read commited transaction isolation
21/07/03 11:29:21 INFO mapreduce.JobSubmitter: number of splits:1
21/07/03 11:29:23 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625311631559_0001
21/07/03 11:29:24 INFO impl.YarnClientImpl: Submitted application application_1625311631559_0001
21/07/03 11:29:24 INFO mapreduce.Job: The url to track the job: http://ip-10-0-0-196.ec2.internal:8088/proxy/application_1625311631559_0001/
21/07/03 11:29:24 INFO mapreduce.Job: Running job: job_1625311631559_0001
21/07/03 11:29:44 INFO mapreduce.Job: Job job_1625311631559_0001 running in uber mode : false
21/07/03 11:29:44 INFO mapreduce.Job:  map 0% reduce 0%
21/07/03 11:30:35 INFO mapreduce.Job:  map 100% reduce 0%
21/07/03 11:30:54 INFO mapreduce.Job: Job job_1625311631559_0001 completed successfully
21/07/03 11:30:54 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
```



```
                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)=51393
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=51393
                Total vcore-milliseconds taken by all map tasks=51393
                Total megabyte-milliseconds taken by all map tasks=52626432
        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)=396
                CPU time spent (ms)=35310
                Physical memory (bytes) snapshot=427929600
                Virtual memory (bytes) snapshot=2803728384
                Total committed heap usage (bytes)=381157376
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=531214815
21/07/03 11:30:54 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 117.2949 seconds (4.3191 MB/sec)
21/07/03 11:30:54 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
You have new mail in /var/spool/mail/root
```



```
[root@ip-10-0-0-196 ~]# hdfs dfs -ls
Found 2 items
drwx------   - root supergroup          0 2021-07-03 11:30 .staging
drwxr-xr-x   - root supergroup          0 2021-07-03 11:30 etlassignment
[root@ip-10-0-0-196 ~]#
```