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 --table SRC_ATM_TRANS --username student --password STUDENT123 --target-dir /user/hadoop/etl_project -m 1

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

hadoop fs -ls /user/hadoop/etl_project

Screenshot of the imported data:

```
[hadoop@ip-172-31-88-180 ~]$ hadoop fs -ls /user/hadoop/
Found 1 items
drwxr-xr-x   - hadoop hadoop          0 2021-11-17 03:25 /user/hadoop/etl_project
[hadoop@ip-172-31-88-180 ~]$ hadoop fs -ls /user/hadoop/etl_project
Found 2 items
-rw-r--r--   1 hadoop hadoop          0 2021-11-17 03:25 /user/hadoop/etl_project/_SUCCESS
-rw-r--r--   1 hadoop hadoop  531214815 2021-11-17 03:25 /user/hadoop/etl_project/part-m-00000
```

Steps Involved:
1) Set an EMR Cluster with Sqoop and login to the EMR cluster via putty.
2) Once logged in, execute the sqoop import command given above. It will import the data from RDS table "SRC_ATM_TRANS" and and store it in a file under "/user/hadoop/etl_project" directory. Below is a screenshot depicting the success of sqoop import.

```
21/11/17 03:25:14 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 52.3869 seconds (9.6705 MB/sec)
21/11/17 03:25:14 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
```

3) Use command "hadoop fs -ls /user/hadoop/etl_project" given above to list the file created via sqoop import command
4) The _SUCCESS indicates that data from RDS has been imported successfully.