

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://testupgrad.cyaieic9bmnf.us-east-1.rds.amazonaws.com:3306/ETL\_db \  
--table ATM_TRANS \  
--username admin --password admin123 \  
--target-dir /user/root/etl_assignment \  
-m 1 \  
--direct
```

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

```
Hadoop fs -ls /user/root/etl_assignment
```

Screenshot of the imported data:

```
Found 2 items  
drwx----- - root supergroup          0 2020-08-20 17:15 /user/root/.staging  
drwxr-xr-x - root supergroup          0 2020-08-20 17:15 /user/root/etl_assignment  
[root@ip-10-0-0-6 ~]# hadoop fs -ls /user/root/etl_assignment  
Found 2 items  
-rw-r--r--  3 root supergroup          0 2020-08-20 17:15 /user/root/etl_assignment/_SUCCESS  
-rw-r--r--  3 root supergroup 531214815 2020-08-20 17:15 /user/root/etl_assignment/part-m-00000  
[root@ip-10-0-0-6 ~]#
```

Explanation:

- 1) Scoop command imports RDS data table ATM_TRANS to a target directory located at root with a name "etl_assignment"
- 2) Once the command is executed, login to ec2 instance. Enter as root user.
- 3) Use command hadoop fs -ls which is used to list all files present.
- 4) There you see a folder etl_assignment
- 5) Open this folder, when you see _SUCCESS, it says data from RDS is loaded sucessfully.