

Data Ingestion from the RDS to HDFS using Sqoop

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

Command:

```
sqoop import \  
--connect jdbc:mysql://#####.us-east-1.rds.amazonaws.com/testdatabase \  
--table SRC_ATM_TRANS \  
--username ##### \  
--password ##### \  
--target-dir /user/root/atm_data \  
-m 1
```

Explanation: This is the sqoop import command used to import the dataset/table from the RDS to the HDFS using the RDS connection string, username, password and table name provided in the problem statement of the project. I have also provided the target directory where the data will be copied in the HDFS (/user/root/atm_data).

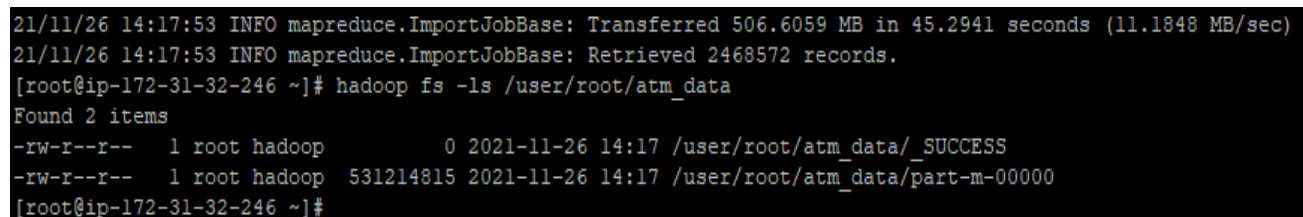
Command used to see the list of imported data in HDFS:

Command:

```
hadoop fs -ls /user/root/atm_data
```

Explanation: This is the command used to list the imported data in the HDFS. This command will show the files present in the target directory in the HDFS (/user/root/atm_data). The command will show two files named **_SUCCESS** and **part-m-00000** in the output.

Screenshot of the imported data:



```
21/11/26 14:17:53 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 45.2941 seconds (11.1848 MB/sec)  
21/11/26 14:17:53 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.  
[root@ip-172-31-32-246 ~]# hadoop fs -ls /user/root/atm_data  
Found 2 items  
-rw-r--r--  1 root hadoop          0 2021-11-26 14:17 /user/root/atm_data/_SUCCESS  
-rw-r--r--  1 root hadoop 531214815 2021-11-26 14:17 /user/root/atm_data/part-m-00000  
[root@ip-172-31-32-246 ~]#
```

Explanation: Above is the screenshot of the data/files imported from the RDS to the HDFS using sqoop.