

SqoopDataIngestion

Query for importing data from Upgrad database to HDFS and storing it as Parquet File.

For this step, the RDS connection string and credentials are as follows:

RDS Connection String -

```
jdbc:mysql://upgraddetest.cyaieic9bmnf.us-east-1.rds.amazonaws.com/testdatabase
```

Username - student

Password - STUDENT123

Table Name - SRC_ATM_TRANS

```
sqoop import --connect jdbc:mysql://upgraddetest.cyaieic9bmnf.us-east-1.rds.amazonaws.com/testdatabase \
```

```
-table SRC_ATM_TRANS \
```

```
-username student -password STUDENT123 \
```

```
-target-dir /user/root/atm_trans \
```

$$-m \quad 1 \quad \backslash$$

-as-parquetfile

EEEEEEEEEEEEEEEEEEEE	MMMMMMMM	MMMMMMMM	RRRRRRRRRRRRRRRR			
E:::EEEEEEEEEEEE::E	M:::M	M:::M	R:::R			
EE:::EEEEEEEE::E	M:::M	M:::M	R:::RRRRRR:::R			
E:::E	EEEE	M:::M	M:::M	RR:::R	R:::R	
E:::E	M:::M	M:::M	M:::M	R:::R	R:::R	
E:::EEEEEEEE	M:::M	M:::M	M:::M	M:::M	R:::RRRRRR:::R	
E:::EEEEEEEE::E	M:::M	M:::M	M:::M	M:::M	R:::RR	
E:::EEEEEEEE	M:::M	M:::M	M:::M	M:::M	R:::RRRRRR:::R	
E:::E	M:::M	M:::M	M:::M	M:::M	R:::R	R:::R
E:::E	EEEE	M:::M	MMM	M:::M	R:::R	R:::R
EE:::EEEEEEEE:::E	M:::M	M:::M	M:::M	R:::R	R:::R	
E:::EEEEEEEE::E	M:::M	M:::M	M:::M	RR:::R	R:::R	
EEEEEEEEEEEEEEEEEEEE	MMMMMMMM	MMMMMMMM	RRRRRRR	RRRRRR		

```
[root@ip-172-31-78-181 ~]# sqoop import --connect jdbc:mysql://upgraddetest.cyai
elc9bmnf.us-east-1.rds.amazonaws.com/testdatabase \
> --table SRC_ATM_TRANS \
> --username student --password STUDENT123 \
> --target-dir /user/livy/atm_trans_parquet \
> -m 1 \
> --as-parquetfile
```

SqoopDataIngestion

Query Result for importing data from Upgrad database to HDFS and storing it as Parquet File.

The file was stored in the following path:

```
[root@ip-172-31-78-181 ~]# hadoop fs -ls /user/livy/atm_trans_parquet
```

Found 3 items

```
drwxr-xr-x - root livy      0 2022-04-13 03:29 /user/livy/atm_trans_parquet/.metadata
```

```
drwxr-xr-x - root livy      0 2022-04-13 03:30 /user/livy/atm_trans_parquet/.signals
```

```
-rw-r--r-- 1 root livy 44688901 2022-04-13 03:30  
/user/livy/atm_trans_parquet/bf3733f1-005b-4b88-950a-e228d9a7096c.parquet
```

```
2022-04-13 03:30:37 INFO mapreduce.ImportJobBase: Transferred 42.6285 MB in 86.519  
3 seconds (504.53 KB/sec)  
2022-04-13 03:30:37 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.  
[root@ip-172-31-78-181 ~]# hadoop fs -ls /user/livy/atm_trans_parquet  
Found 3 items  
drwxr-xr-x - root livy      0 2022-04-13 03:29 /user/livy/atm_trans_parquet/.metadata  
drwxr-xr-x - root livy      0 2022-04-13 03:30 /user/livy/atm_trans_parquet/.signals  
-rw-r--r-- 1 root livy 44688901 2022-04-13 03:30 /user/livy/atm_trans_parquet/bf3733f1-005b-4b88-950a-e228d9a7096c.parquet  
[root@ip-172-31-78-181 ~]#
```

```
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)=787  
  CPU time spent (ms)=59980  
  Physical memory (bytes) snapshot=705257472  
  Virtual memory (bytes) snapshot=3300474880  
  Total committed heap usage (bytes)=573046784  
File Input Format Counters  
  Bytes Read=0  
File Output Format Counters  
  Bytes Written=0
```