

## 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.cyaieic9bmnf.us-east-1.rds.amazonaws.com/testdatabase \
--username student \
--password STUDENT123 \
--table SRC_ATM_TRANS \
--target-dir /user/root/SRC_ATM_TRANS3 -m 1
```

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

```
hadoop fs -ls /user/root/SRC_ATM_TRANS3
```

### Screenshot of the imported data:

```
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=176344
  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
  Other local map tasks=1
  Total time spent by all maps in occupied slots (ms)=29832
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=29832
  Total vcore-milliseconds taken by all map tasks=29832
  Total megabyte-milliseconds taken by all map tasks=30547968
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)=269
  CPU time spent (ms)=29850
  Physical memory (bytes) snapshot=430182400
  Virtual memory (bytes) snapshot=2818076672
  Total committed heap usage (bytes)=392167424
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=531214815
21/05/24 03:23:00 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 54.135 seconds (9.3582 MB/sec)
21/05/24 03:23:00 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-10-0-0-241 ~]# hadoop fs -ls /user/root/SRC_ATM_TRANS3
Found 2 items
-rw-r--r--  3 root supergroup          0 2021-05-24 03:22 /user/root/SRC_ATM_TRANS3/_SUCCESS
-rw-r--r--  3 root supergroup 531214815 2021-05-24 03:22 /user/root/SRC_ATM_TRANS3/part-m-00000
[root@ip-10-0-0-241 ~]# hadoop fs -ls
```