

## Data Ingestion from RDS to HDFS using Sqoop

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

```
[root@ip-10-0-0-165 ~]# sqoop import --connect jdbc:mysql://upgradetest1.cyaieic9bmnf.us-east-1.rds.amazonaws.com/testdatabase --table SRC_ATM_TRANS --username student --password STUDENT123 --target-dir /user/root/SRC_ATM_TRANS -m 1
```

```
20/12/03 16:15:07 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=176683
    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=522660639
    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)=29412
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=29412
    Total vcore-milliseconds taken by all map tasks=29412
    Total megabyte-milliseconds taken by all map tasks=30117888
  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)=211
    CPU time spent (ms)=30660
    Physical memory (bytes) snapshot=422649856
    Virtual memory (bytes) snapshot=2819510272
    Total committed heap usage (bytes)=384827392
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=522660639
20/12/03 16:15:07 INFO mapreduce.ImportJobBase: Transferred 498.448 MB in 53.4809 seconds (9.3201 MB/sec)
20/12/03 16:15:07 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-10-0-0-165 ~]#
```

Retrieved **2468572** records.

Commands used to see the list of imported data in HDFS

```
[root@ip-10-0-0-165 ~]# hadoop fs -ls /user/root/SRC_ATM_TRANS
```

Found 2 items

```
[root@ip-10-0-0-165 ~]# hadoop fs -ls /user/root/SRC_ATM_TRANS
Found 2 items
-rw-r--r--  3 root supergroup          0 2020-12-03 16:15 /user/root/SRC_ATM_TRANS/_SUCCESS
-rw-r--r--  3 root supergroup 522660639 2020-12-03 16:15 /user/root/SRC_ATM_TRANS/part-m-00000
[root@ip-10-0-0-165 ~]#
```

Screenshot of the imported data

```
[root@ip-10-0-0-165 ~]# hadoop fs -cat /user/root/SRC_ATM_TRANS/part-m-00000 | wc -l
```

**2468572**

```
[root@ip-10-0-0-165 ~]# hadoop fs -cat /user/root/SRC_ATM_TRANS/part-m-00000 | wc -l
2468572
[root@ip-10-0-0-165 ~]#
```