

Data Ingestion from the RDS to HDFS using Sqoop

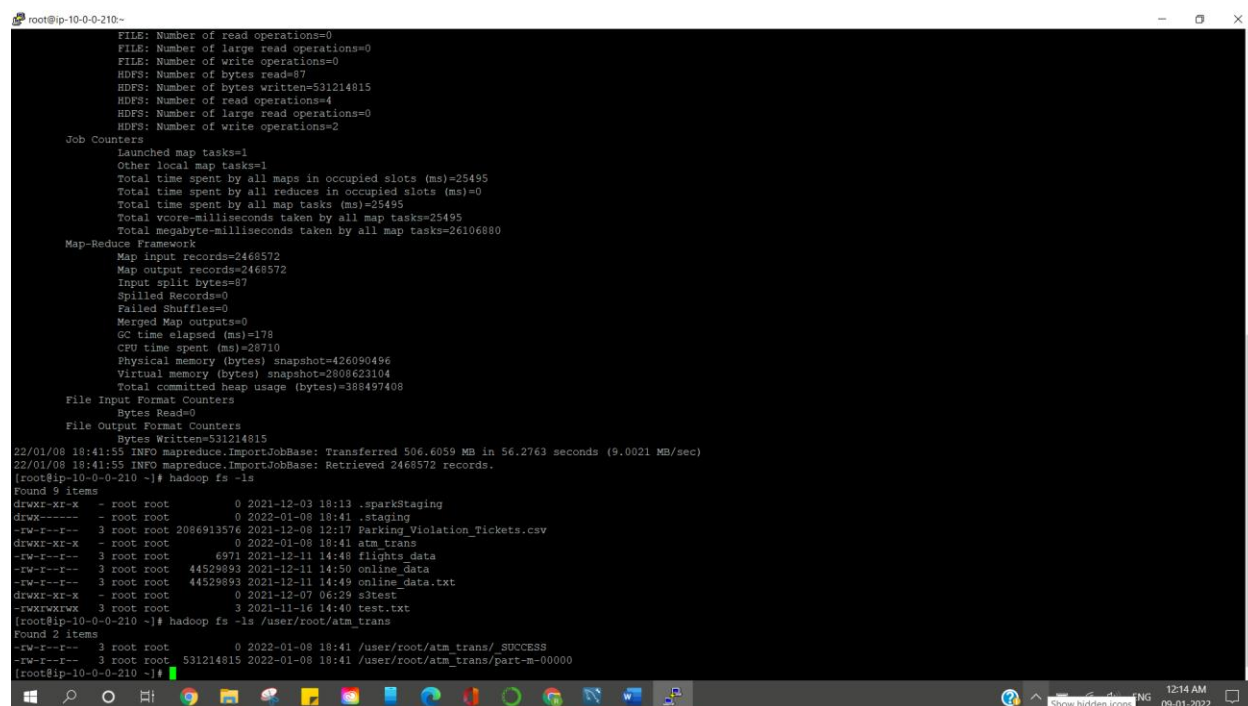
Sqoop Import command used for importing table from RDS to HDFS:

```
sqoop import \
> --connect jdbc:mysql://upgraddetest.cyaiehc9bmnf.us-east-1.rds.amazonaws.com/testdatabase \
> --table SRC_ATM_TRANS \
> --username student --password STUDENT123 \
> --target-dir /user/root/atm_trans \
> -m 1
```

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

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

Screenshot of the imported data:



```
root@ip-10-0-0-210:~$ sqoop import \
  --connect jdbc:mysql://upgraddetest.cyaiehc9bmnf.us-east-1.rds.amazonaws.com/testdatabase \
  --table SRC_ATM_TRANS \
  --username student --password STUDENT123 \
  --target-dir /user/root/atm_trans \
  -m 1

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)=25495
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=25495
  Total vcore-milliseconds taken by all map tasks=25495
  Total megabyte-milliseconds taken by all map tasks=26106880

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)=178
  CPU time spent (ms)=28710
  Physical memory (bytes) snapshot=426099486
  Virtual memory (bytes) snapshot=2808623104
  Total committed heap usage (bytes)=388497408

File Input Format Counters
  Bytes Read=0

File Output Format Counters
  Bytes Written=531214815
22/01/08 18:41:55 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 56.2763 seconds (9.0021 MB/sec)
22/01/08 18:41:55 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-10-0-0-210 ~]# hadoop fs -ls
Found 9 items
drwxr-xr-x  - root root      0 2021-12-03 18:13 .sparkStaging
drwxr-xr-x  - root root      0 2022-01-08 18:41 .staging
-rw-r--r--  3 root root 2086913576 2021-12-08 12:17 Parking_Violation_Tickets.csv
drwxr-xr-x  - root root      0 2022-01-08 18:41 atm_trans
-rw-r--r--  3 root root      6971 2021-12-11 14:48 flights_data
-rw-r--r--  3 root root 44529893 2021-12-11 14:50 online_data
-rw-r--r--  3 root root 44529893 2021-12-11 14:49 online_data.txt
drwxr-xr-x  - root root      0 2021-12-07 06:29 s3test
-rwxrwxrwx  3 root root      3 2021-11-16 14:40 test.txt
[root@ip-10-0-0-210 ~]# hadoop fs -ls /user/root/atm_trans
Found 2 items
-rw-r--r--  3 root root      0 2022-01-08 18:41 /user/root/atm_trans/_SUCCESS
-rw-r--r--  3 root root 531214815 2022-01-08 18:41 /user/root/atm_trans/part-m-00000
[root@ip-10-0-0-210 ~]#
```