Data Ingestion from the RDS to HDFS using Sqoop

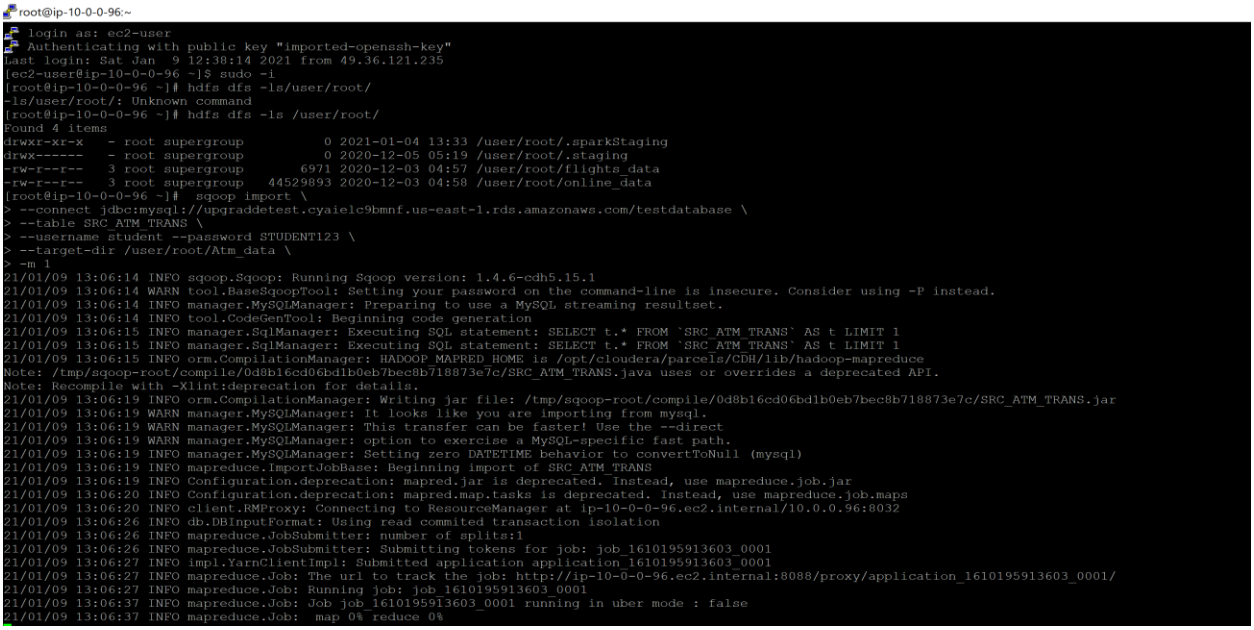
Sqoop Import command used for importing table from RDS to HDFS:

```
sqoop import \
--connect jdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase \
--table SRC_ATM_TRANS \
--username student --password STUDENT123 \
--target-dir /user/root/Atm_data \
-m 1
```

We have imported the data from RDS to HDFS using sqoop import, where the jdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase is the connects string to connect to RDS table, source table name is SRC_ATM_TRANS , target directory on hdfs is SRC_ATM_TRANS and number of mappers used is 1.
It will take default field delimiter as ‘,’.

Screenshot of the imported data:

Sqoop command submission:



© Copyright 2020. upGrad Education Pvt. Ltd. All rights reserved

Imported Data: 2468572 rows imported

```
21/01/09 13:06:19 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
21/01/09 13:06:19 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
21/01/09 13:06:19 INFO mapreduce.ImportJobBase: Beginning import of SRC_ATM_TRANS
21/01/09 13:06:19 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
21/01/09 13:06:20 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
21/01/09 13:06:20 INFO client.RMProxy: Connecting to ResourceManager at ip-10-0-0-96.ec2.internal/10.0.0.96:8032
21/01/09 13:06:26 INFO db.DBInputFormat: Using read commited transaction isolation
21/01/09 13:06:26 INFO mapreduce.JobSubmitter: number of splits:1
21/01/09 13:06:26 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1610195913603_0001
21/01/09 13:06:27 INFO impl.YarnClientImpl: Submitted application application_1610195913603_0001
21/01/09 13:06:27 INFO mapreduce.Job: The url to track the job: http://ip-10-0-0-96.ec2.internal:8088/proxy/application_1610195913603_0001/
21/01/09 13:06:27 INFO mapreduce.Job: Running job: job_1610195913603_0001
21/01/09 13:06:37 INFO mapreduce.Job: Job job_1610195913603_0001 running in uber mode : false
21/01/09 13:06:37 INFO mapreduce.Job:  map 0% reduce 0%
21/01/09 13:07:07 INFO mapreduce.Job:  map 100% reduce 0%
21/01/09 13:07:09 INFO mapreduce.Job: Job job_1610195913603_0001 completed successfully
21/01/09 13:07:09 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=176278
                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)=28181
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=28181
                Total vcore-milliseconds taken by all map tasks=28181
                Total megabyte-milliseconds taken by all map tasks=28857344
        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)=207
                CPU time spent (ms)=29080
                Physical memory (bytes) snapshot=445526016
                Virtual memory (bytes) snapshot=2841772032
                Total committed heap usage (bytes)=386924544
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=531214815
21/01/09 13:07:09 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 48.7176 seconds (10.3988 MB/sec)
21/01/09 13:07:09 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-10-0-0-96 ~]#
```

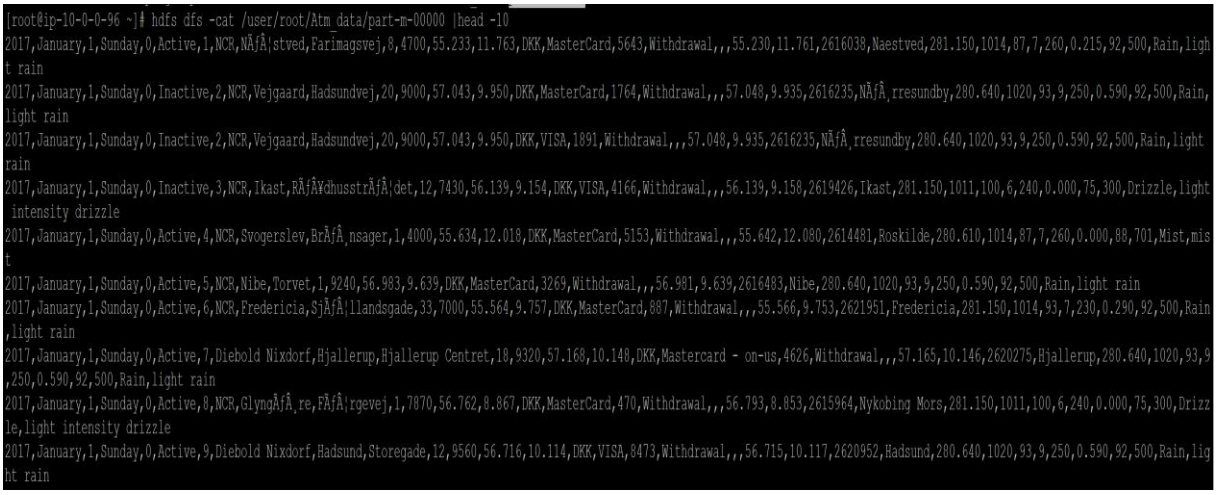
Command used to see the list of imported data in HDFS:

hadoop fs -ls /user/root/Atm_data/

```
[root@ip-10-0-0-96 ~]# hadoop fs -ls /user/root/Atm_data/
Found 2 items
-rw-r--r--   3 root supergroup          0 2021-01-09 13:07 /user/root/Atm_data/_SUCCESS
-rw-r--r--   3 root supergroup  531214815 2021-01-09 13:07 /user/root/Atm_data/part-m-00000
[root@ip-10-0-0-96 ~]#
```

Viewing the file content:
hdfs dfs -cat /user/root/Atm_data/part-m-00000 |head -10

© Copyright 2020. upGrad Education Pvt. Ltd. All rights reserved



© Copyright 2020. upGrad Education Pvt. Ltd. All rights reserved