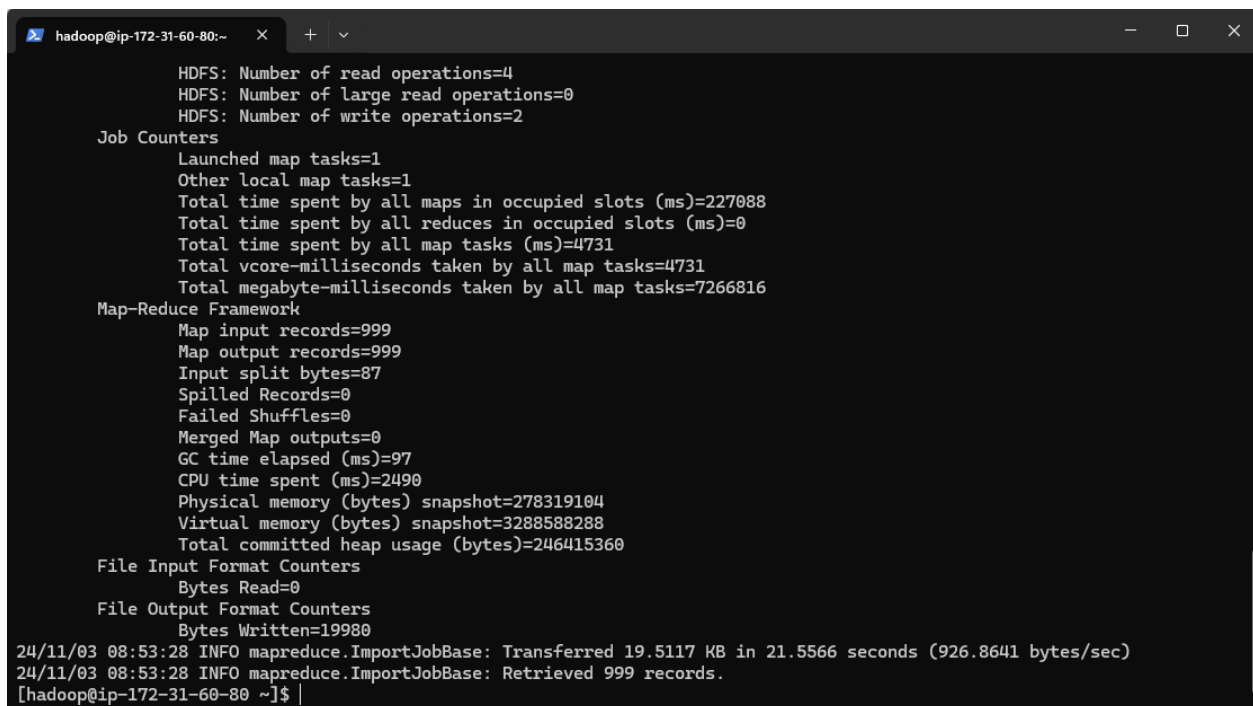# Data Ingestion from the RDS to HDFS using Sqoop

Sqoop command used for importing table from RDS to HDFS

• Sqoop command to import tables from RDS to HDFS:
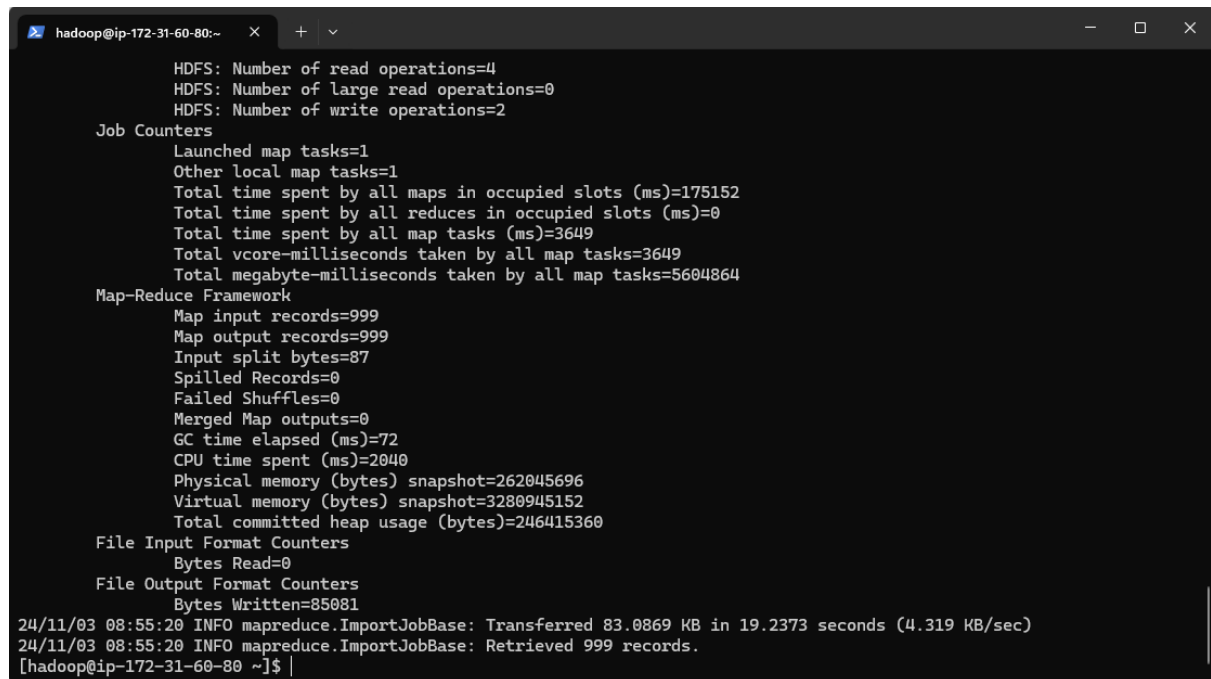
Table 1: Command for member_score table:

```
sqoop import \
--connect jdbc:mysql://upgradawsrds1.cyaielc9bmnf.us-east-
1.rds.amazonaws.com/cred_financials_data \
--table member_score \
--username upgraduser --password upgraduser \
--target-dir /user/root/cap_project/member_score \
-m 1
```

```
hadoop@ip-172-31-60-80:~

                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)=227088
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=4731
                Total vcore-milliseconds taken by all map tasks=4731
                Total megabyte-milliseconds taken by all map tasks=7266816
        Map-Reduce Framework
                Map input records=999
                Map output records=999
                Input split bytes=87
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=97
                CPU time spent (ms)=2490
                Physical memory (bytes) snapshot=278319104
                Virtual memory (bytes) snapshot=3288588288
                Total committed heap usage (bytes)=246415360
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=19980
24/11/03 08:53:28 INFO mapreduce.ImportJobBase: Transferred 19.5117 KB in 21.5566 seconds (926.8641 bytes/sec)
24/11/03 08:53:28 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[hadoop@ip-172-31-60-80 ~]$
```

Table 2: Command for card_member table:

```
sqoop import \
--connect jdbc:mysql://upgradawsrds1.cyaielc9bmnf.us-east-
1.rds.amazonaws.com/cred_financials_data \
--table card_member \
--username upgraduser --password upgraduser \
--target-dir /user/root/cap_project/card_member \
-m 1
```

```
                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)=175152
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=3649
                Total vcore-milliseconds taken by all map tasks=3649
                Total megabyte-milliseconds taken by all map tasks=5604864
        Map-Reduce Framework
                Map input records=999
                Map output records=999
                Input split bytes=87
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=72
                CPU time spent (ms)=2040
                Physical memory (bytes) snapshot=262045696
                Virtual memory (bytes) snapshot=3280945152
                Total committed heap usage (bytes)=246415360
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=85081
24/11/03 08:55:20 INFO mapreduce.ImportJobBase: Transferred 83.0869 KB in 19.2373 seconds (4.319 KB/sec)
24/11/03 08:55:20 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[hadoop@ip-172-31-60-80 ~]$
```

<Command to see the list of imported data in HDFS>

```
hadoop fs -ls cap_project command is used list the imported files

#commands used to view imported data
hadoop fs -cat cap_project/card_member
hadoop fs -cat cap_project/member_score
```

Screenshot of the imported data

• member_score

```
65625105449485881,659982919406634,2015-07-04 06:39:15.0,11/16,United States,Mount Vernon
6566718331696237,335526117403543,2013-09-19 10:34:17.0,08/14,United States,Sand Springs
6568438963325843,183038102441540,2017-07-06 05:25:55.0,07/17,United States,North Las Vegas
6570061543816002,722794052951029,2017-01-10 03:01:36.0,02/17,United States,Sedalia
6573615549671948,985619195348046,2015-09-29 08:21:37.0,12/16,United States,Belleville
6574255180086418,891702243060747,2012-07-19 07:47:35.0,07/15,United States,Selma
6574656060675651,555127039474442,2010-06-02 00:11:01.0,11/10,United States,Coral Terrace
6577397607792660,479331531880555,2013-06-13 00:02:59.0,02/16,United States,Norco
6578328793401039,149907932138330,2016-03-21 02:51:11.0,11/16,United States,Wasco
6580194731591489,218337397923739,2018-01-09 00:32:14.0,01/18,United States,Oxford
6582135370631324,389050970153068,2016-07-17 10:26:54.0,08/16,United States,Roy
6582950745622331,185289866247329,2012-04-17 03:37:04.0,03/13,United States,Hutchinson
6583283909793527,398607904778893,2015-08-27 03:19:32.0,01/17,United States,Galesburg
6585229076432173,887283455220105,2015-03-06 00:53:06.0,05/16,United States,Security-Widefield
6587114664896609,623414156260264,2015-10-05 06:45:35.0,03/17,United States,Hobart
6588637504808417,678537053778506,2012-05-30 04:37:45.0,04/14,United States,Concord
6589539191695876,832393339334706,2014-01-26 11:59:35.0,09/15,United States,Roseville
6589855658684769,193386908982058,2010-02-07 09:52:17.0,11/15,United States,Westfield
6589894320960430,276119478113309,2011-08-08 02:25:29.0,11/15,United States,Blaine
6590907016354002,849289511810887,2016-10-28 07:32:23.0,07/17,United States,Reading
6591175617713393,135615957283320,2015-05-16 03:21:15.0,08/16,United States,Arvin
6592184145413632,714853788321318,2010-03-12 08:11:17.0,08/16,United States,Mehlville
6594248319343442,846162645575666,2016-02-15 02:32:58.0,12/16,United States,Lyndhurst
6595638658736751,377918213614639,2015-08-13 10:58:24.0,09/17,United States,Mahwah
6595814135833988,236864426408837,2011-02-18 06:45:46.0,06/14,United States,Johnston
6595928469079750,422173403853722,2012-02-10 02:19:20.0,02/14,United States,Raleigh
6597703848279563,657991752750272,2017-02-16 07:45:34.0,03/17,United States,Pinewood
6598830758632447,660350842993890,2013-07-26 02:22:17.0,12/16,United States,Tampa
6599900931314251,928036864799687,2017-09-09 03:31:54.0,10/17,United States,Loves Park
[hadoop@ip-172-31-60-80 ~]$
```

• card_member

```
973974062665904,565
974087224071871,682
975111321017949,306
976362362966586,210
976740397894598,303
978465390240911,270
978786807247400,652
979218131207765,552
979274357975688,353
981612675682175,265
982171321477033,683
983248478716258,473
983390343419665,183
985095728641698,325
985619195348046,310
987837606639725,623
988661034778977,496
988666952112281,531
989174793157941,533
989548647998824,506
990571384923260,469
990740662283435,202
991784667284668,621
991958482478439,215
992104321998571,658
992552823055811,694
994983851226493,687
996411635289270,335
997128952368160,683
[hadoop@ip-172-31-60-80 ~]$
```