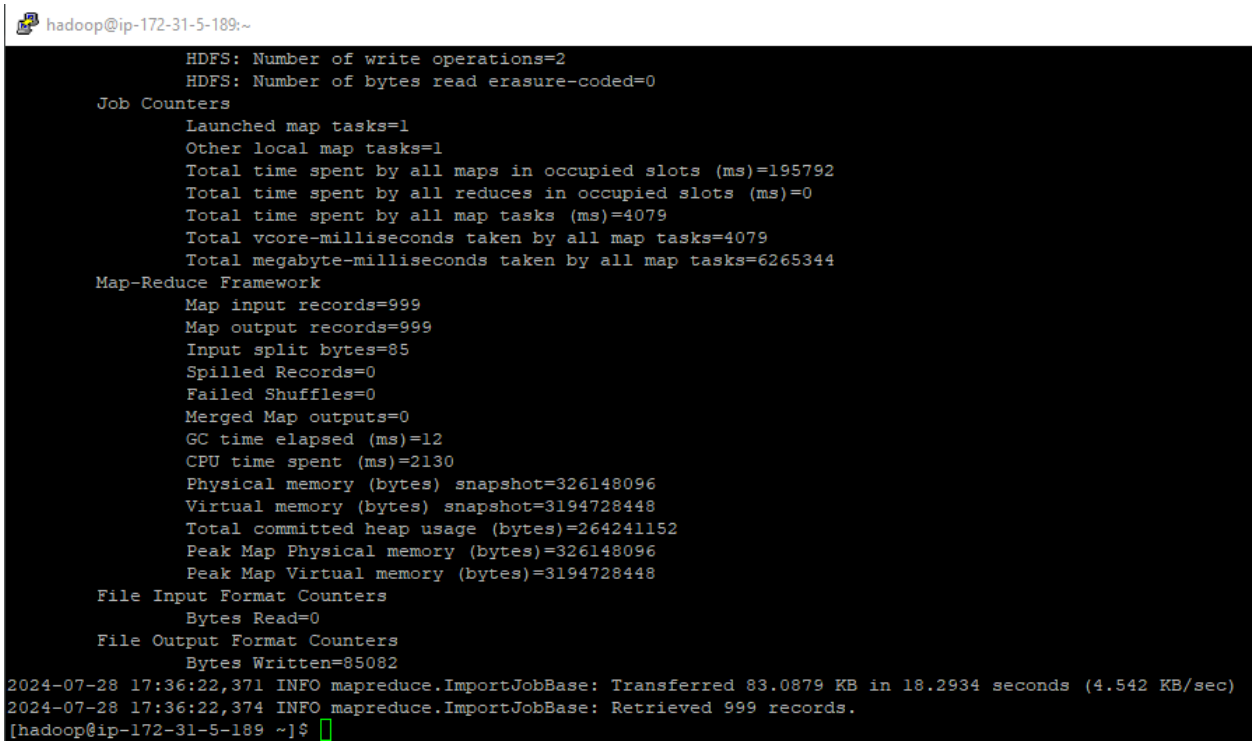
TASK 2

A PDF document **(SqoopDataIngestion.pdf)** containing the code used for ingesting data from the RDS server.

Data Ingestion from the RDS to HDFS using Sqoop

Sqoop command used for importing table from RDS to HDFS

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



As seen 999 records are retrieved

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

```
hadoop@ip-172-31-5-189:~
                HDFS: Number of write operations=2
                HDFS: Number of bytes read erasure-coded=0
        Job Counters
                Launched map tasks=1
                Other local map tasks=1
                Total time spent by all maps in occupied slots (ms)=189216
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=3942
                Total vcore-milliseconds taken by all map tasks=3942
                Total megabyte-milliseconds taken by all map tasks=6054912
        Map-Reduce Framework
                Map input records=999
                Map output records=999
                Input split bytes=85
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=7
                CPU time spent (ms)=2510
                Physical memory (bytes) snapshot=350670848
                Virtual memory (bytes) snapshot=3195916288
                Total committed heap usage (bytes)=264241152
                Peak Map Physical memory (bytes)=350670848
                Peak Map Virtual memory (bytes)=3195916288
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=19980
2024-07-28 17:37:57,246 INFO mapreduce.ImportJobBase: Transferred 19.5117 KB in 17.6081 seconds (1.1081 KB/sec)
2024-07-28 17:37:57,250 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[hadoop@ip-172-31-5-189 ~]$
```

As seen 999 records are retrieved

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

Hdfs dfs -ls capstone_project/

```
[root@ip-172-31-5-189 ~]# ls
load_hbase.py
[root@ip-172-31-5-189 ~]# hdfs dfs -ls
Found 1 items
drwxr-xr-x   - hadoop hdfsadmingroup          0 2024-07-28 17:37 capstone_project
[root@ip-172-31-5-189 ~]# hdfs dfs -ls capstone_project/
Found 2 items
drwxr-xr-x   - hadoop hdfsadmingroup          0 2024-07-28 17:36 capstone_project/card_member
drwxr-xr-x   - hadoop hdfsadmingroup          0 2024-07-28 17:37 capstone_project/member_score
[root@ip-172-31-5-189 ~]#
```

Screenshot of the imported data

Commands to view imported data

hdfs dfs -cat capstone_project/card_member/part-m-00000

```
[root@ip-172-31-5-189 ~]# hdfs dfs -ls capstone_project/card_member/
Found 2 items
-rw-r--r--   1 hadoop hdfsadmingroup          0 2024-07-28 17:36 capstone_project/card_member/_SUCCESS
-rw-r--r--   1 hadoop hdfsadmingroup      85082 2024-07-28 17:36 capstone_project/card_member/part-m-00000
[root@ip-172-31-5-189 ~]# hdfs dfs -cat capstone_project/card_member/part-m-00000
340028465709212,009250698176266,2012-02-08 06:04:13.0,05/13,United States,Barberton
340054675199675,835873341185231,2017-03-10 09:24:44.0,03/17,United States,Fort Dodge
340082915339645,512969555857346,2014-02-15 06:30:30.0,07/14,United States,Graham
340134186926007,887711945571282,2012-02-05 01:21:58.0,02/13,United States,Dix Hills
340265728490548,680324265406190,2014-03-29 07:49:14.0,11/14,United States,Rancho Cucamonga
340268219434811,929799084911715,2012-07-08 02:46:08.0,08/12,United States,San Francisco
340379737226464,089615510858348,2010-03-10 00:06:42.0,09/10,United States,Clinton
340383645652108,181180599313885,2012-02-24 05:32:44.0,10/16,United States,West New York
340803866934451,417664728506297,2015-05-21 04:30:45.0,08/17,United States,Beaverton
340889618969736,459292914761635,2013-04-23 08:40:11.0,11/15,United States,West Palm Beach
340924125838453,188119365574843,2011-04-12 04:28:47.0,12/13,United States,Scottsbluff
341005627432127,872138964937565,2013-09-08 03:16:50.0,02/17,United States,Chillum
341029651579925,974087224071871,2011-01-14 00:20:25.0,08/12,United States,Valley Station
341311317050937,561687420200207,2014-03-18 06:23:23.0,02/15,United States,Vincennes
341344252914274,695906467918552,2012-03-02 03:21:01.0,03/13,United States,Columbine
341363858179050,009190444424572,2012-02-19 05:16:44.0,04/14,United States,Cheektowaga
341519629171378,533670008048847,2013-05-13 07:59:32.0,01/15,United States,Centennial
341641153427489,230523184584316,2013-03-25 08:51:18.0,11/15,United States,Colchester
341719092861087,304847505155781,2015-12-06 08:06:35.0,11/17,United States,Vernon Hills
341722035429601,979218131207765,2015-12-22 10:46:23.0,01/17,United States,Elk Grove Village
341724964458347,210778177559185,2011-02-07 01:43:13.0,03/14,United States,Fond du Lac
```

hdfs dfs -cat capstone_project/member_score/part-m-00000

```
[root@ip-172-31-5-189 ~]# hdfs dfs -ls capstone_project/
Found 2 items
drwxr-xr-x   - hadoop hdfsadmingroup          0 2024-07-28 17:36 capstone_project/card_member
drwxr-xr-x   - hadoop hdfsadmingroup          0 2024-07-28 17:37 capstone_project/member_score
[root@ip-172-31-5-189 ~]# hdfs dfs -ls capstone_project/member_score
Found 2 items
-rw-r--r--   1 hadoop hdfsadmingroup          0 2024-07-28 17:37 capstone_project/member_score/_SUCCESS
-rw-r--r--   1 hadoop hdfsadmingroup      19980 2024-07-28 17:37 capstone_project/member_score/part-m-00000
[root@ip-172-31-5-189 ~]# hdfs dfs -cat capstone_project/member_score/part-m-00000
000037495066290,339
000117826301530,289
001147922084344,393
001314074991813,225
001739553947511,642
003761426295463,413
004494068832701,217
006836124210484,504
006991872634058,697
007955566230397,372
008732267588672,213
008765307152821,399
009136568025042,308
009190444424572,559
009250698176266,233
009873334520465,298
011716573646690,249
```