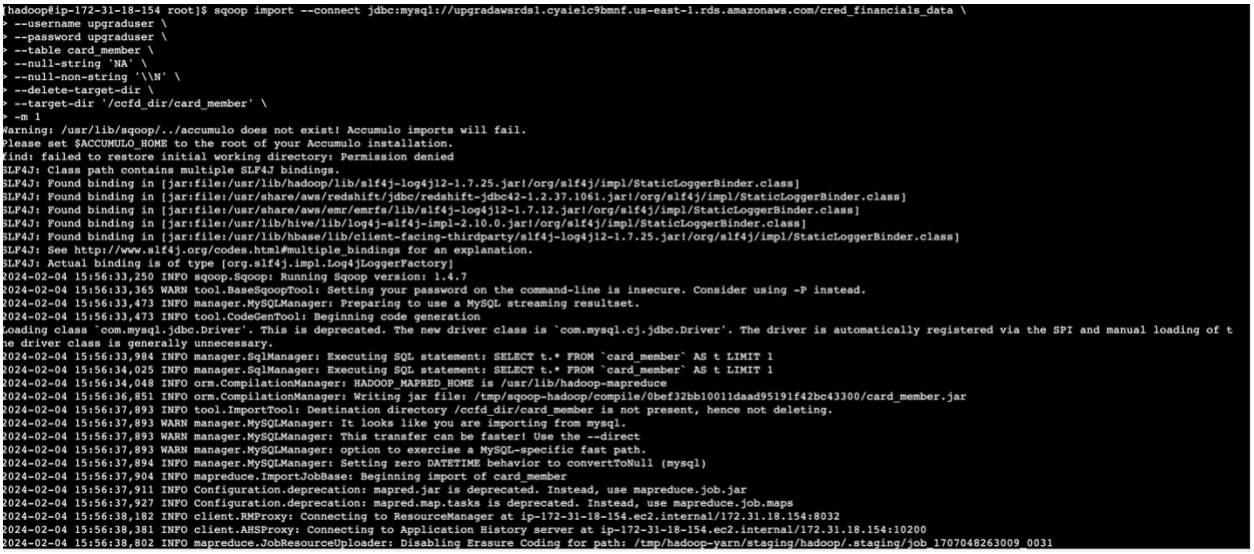
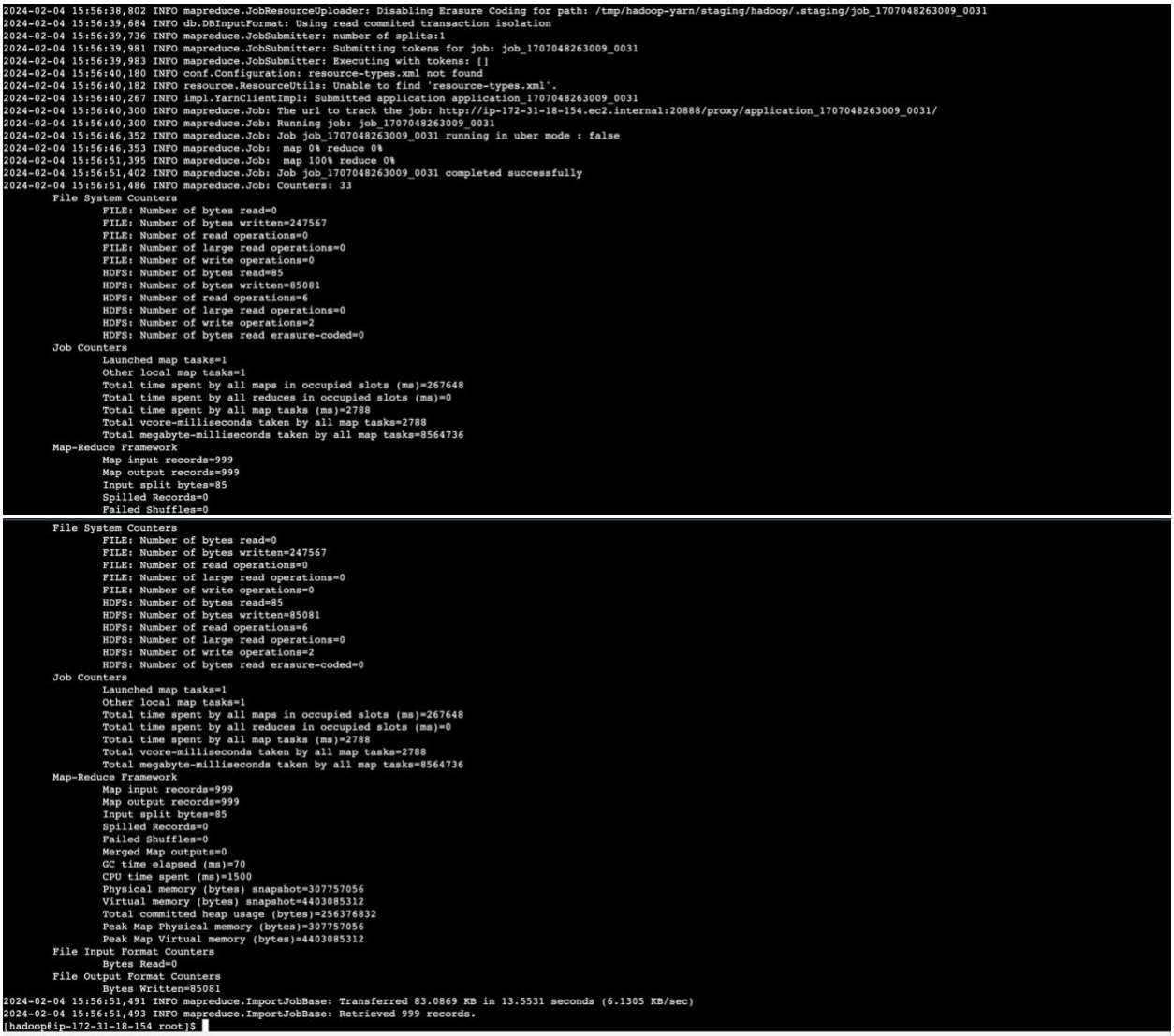
Data Ingestion from the RDS to HDFS using Sqoop

Sqoop command used for importing table from RDS to HDFS.

- Scoop command to import data into card_member table.

```
sqoop import --connect jdbc:mysql://upgradawsrds1.cyaielc9bmnf.us-east-1.rds.amazonaws.com/cred_financials_data \
--username upgraduser \
--password upgraduser \
--table card_member \
--null-string 'NA' \
--null-non-string '\\N' \
--delete-target-dir \
--target-dir '/ccfd_dir/card_member' \
-m 1
```





- Scoop command to import data into member_score table.

```
sqoop import --connect jdbc:mysql://upgradawsrds1.cyaielc9bmnf.us-east-
1.rds.amazonaws.com/cred_financials_data \
--username upgraduser \
--password upgraduser \
--table member_score \
--null-string 'NA' \
--null-non-string '\\N' \
--delete-target-dir \
--target-dir '/ccfd_dir/member_score' \
-m 1
```

```
[hadoop@ip-172-31-18-154 root]$ sqoop import --connect jdbc:mysql://upgradawsrds1.cyaielc9bmnf.us-east-1.rds.amazonaws.com/cred_financials_data \
> --username upgraduser \
> --password upgraduser \
> --table member_score \
> --null-string 'NA' \
> --null-non-string '\\N' \
> --delete-target-dir \
> --target-dir '/ccfd_dir/member_score' \
> -m 1
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
find: failed to restore initial working directory: Permission denied
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/share/aws/redshift/jdbc/redshift-jdbc42-1.2.37.1061.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/share/aws/emr/emrfs/lib/slf4j-log4j12-1.7.12.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hbase/lib/client-facing-thirdparty/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2024-02-04 15:58:28,527 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
2024-02-04 15:58:28,605 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2024-02-04 15:58:28,707 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2024-02-04 15:58:28,707 INFO tool.CodeGenTool: Beginning code generation
Loading class `com.mysql.jdbc.Driver'. This is deprecated. The new driver class is `com.mysql.cj.jdbc.Driver'. The driver is automatically registered via the SPI and manual loading of t
he driver class is generally unnecessary.
2024-02-04 15:58:29,165 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `member_score` AS t LIMIT 1
2024-02-04 15:58:29,206 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `member_score` AS t LIMIT 1
2024-02-04 15:58:29,228 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
2024-02-04 15:58:31,505 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-hadoop/compile/a036d96e4ea0155877542947e704e405/member_score.jar
2024-02-04 15:58:32,941 INFO tool.ImportTool: Destination directory /ccfd_dir/member_score is not present, hence not deleting.
2024-02-04 15:58:32,942 WARN manager.MySQLManager: It looks like you are importing from mysql.
2024-02-04 15:58:32,942 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
2024-02-04 15:58:32,942 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
2024-02-04 15:58:32,942 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
2024-02-04 15:58:32,953 INFO mapreduce.ImportJobBase: Beginning import of member_score
2024-02-04 15:58:32,959 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
2024-02-04 15:58:32,977 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2024-02-04 15:58:33,259 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-18-154.ec2.internal/172.31.18.154:8032
2024-02-04 15:58:33,514 INFO client.AHSProxy: Connecting to Application History server at ip-172-31-18-154.ec2.internal/172.31.18.154:10200
2024-02-04 15:58:34,026 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1707048263009_0032
2024-02-04 15:58:34,026 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1707048263009_0032
2024-02-04 15:58:34,863 INFO db.DBInputFormat: Using read commited transaction isolation
2024-02-04 15:58:34,915 INFO mapreduce.JobSubmitter: number of splits:1
2024-02-04 15:58:35,171 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1707048263009_0032
2024-02-04 15:58:35,172 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-02-04 15:58:35,342 INFO conf.Configuration: resource-types.xml not found
2024-02-04 15:58:35,343 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-02-04 15:58:35,412 INFO impl.YarnClientImpl: Submitted application application_1707048263009_0032
2024-02-04 15:58:35,456 INFO mapreduce.Job: The url to track the job: http://ip-172-31-18-154.ec2.internal:20888/proxy/application_1707048263009_0032/
2024-02-04 15:58:35,457 INFO mapreduce.Job: Running job: job_1707048263009_0032
2024-02-04 15:58:41,511 INFO mapreduce.Job: Job job_1707048263009_0032 running in uber mode : false
2024-02-04 15:58:41,511 INFO mapreduce.Job:  map 0% reduce 0%
2024-02-04 15:58:46,560 INFO mapreduce.Job:  map 100% reduce 0%
2024-02-04 15:58:46,567 INFO mapreduce.Job: Job job_1707048263009_0032 completed successfully
2024-02-04 15:58:46,673 INFO mapreduce.Job: Counters: 33
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=247514
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=85
                HDFS: Number of bytes written=19980
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                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)=270432
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=2817
                Total vcore-milliseconds taken by all map tasks=2817
                Total megabyte-milliseconds taken by all map tasks=8653824
        Map-Reduce Framework
                Map input records=999
                Map output records=999
                Input split bytes=85
                Spilled Records=0
                Failed Shuffles=0
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=247514
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=85
                HDFS: Number of bytes written=19980
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                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)=270432
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=2817
                Total vcore-milliseconds taken by all map tasks=2817
                Total megabyte-milliseconds taken by all map tasks=8653824
        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)=79
                CPU time spent (ms)=1400
                Physical memory (bytes) snapshot=310341632
                Virtual memory (bytes) snapshot=4397051904
                Total committed heap usage (bytes)=330825728
                Peak Map Physical memory (bytes)=310341632
                Peak Map Virtual memory (bytes)=4397051904
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=19980
2024-02-04 15:58:46,679 INFO mapreduce.ImportJobBase: Transferred 19.5117 KB in 13.6907 seconds (1.4252 KB/sec)
2024-02-04 15:58:46,682 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[hadoop@ip-172-31-18-154 root]$
```

- Command to create external table to store values from card_member table.

```
CREATE EXTERNAL TABLE IF NOT EXISTS card_member_ext (
  card_id STRING,
  member_id STRING,
  member_joining_dt TIMESTAMP,
  card_purchase_dt STRING,
  country STRING,
  city STRING
 )
 ROW FORMAT DELIMITED
 FIELDS TERMINATED BY ','
 LOCATION '/ccfd_dir/card_member';
```

- Command to create external table to store values from member_score table.

```
CREATE EXTERNAL TABLE IF NOT EXISTS member_score_ext (
  member_id STRING,
  score INT
 )
 ROW FORMAT DELIMITED
 FIELDS TERMINATED BY ','
 LOCATION '/ccfd_dir/member_score';
```

```
hive> CREATE EXTERNAL TABLE IF NOT EXISTS member_score_ext (
    >   member_id STRING,
    >   score INT
    > )
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ','
    > LOCATION '/ccfd_dir/member_score';
OK
Time taken: 0.074 seconds
```

- Command to create and load data to CARD_MEMBER_ORC table in ORC format for better performance.

```
CREATE TABLE IF NOT EXISTS CARD_MEMBER_ORC(
      `CARD_ID` STRING,
      `MEMBER_ID` STRING,
      `MEMBER_JOINING_DT` TIMESTAMP,
      `CARD_PURCHASE_DT` STRING,
      `COUNTRY` STRING,
      `CITY` STRING
   )
   STORED AS ORC
   TBLPROPERTIES ("orc.compress"="SNAPPY");

INSERT OVERWRITE TABLE CARD_MEMBER_ORC
   SELECT
     CARD_ID,
     MEMBER_ID,
     MEMBER_JOINING_DT,
     CARD_PURCHASE_DT,
     COUNTRY,
     CITY
   FROM
   CARD_MEMBER_EXT;
```

```
hive>
    > CREATE TABLE IF NOT EXISTS CARD_MEMBER_ORC(
    >     `CARD_ID` STRING,
    >     `MEMBER_ID` STRING,
    >     `MEMBER_JOINING_DT` TIMESTAMP,
    >     `CARD_PURCHASE_DT` STRING,
    >     `COUNTRY` STRING,
    >     `CITY` STRING
    > )
    > STORED AS ORC
    > TBLPROPERTIES ("orc.compress"="SNAPPY");
OK
Time taken: 0.806 seconds
hive> INSERT OVERWRITE TABLE CARD_MEMBER_ORC
    > SELECT
    >     CARD_ID,
    >     MEMBER_ID,
    >     MEMBER_JOINING_DT,
    >     CARD_PURCHASE_DT,
    >     COUNTRY,
    >     CITY
    > FROM
    >     CARD_MEMBER_EXT;
Query ID = hadoop_20240204173045_a612fc73-7ade-4e69-92a9-7cdd62d4bf07
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1707048263009_0036)

--------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      1          1        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 4.50 s
--------------------------------------------------------------------------------
Loading data to table default.card_member_orc
OK
Time taken: 9.118 seconds
hive> select count(1) from CARD_MEMBER_ORC;
OK
999
Time taken: 0.483 seconds, Fetched: 1 row(s)
```

```
CREATE TABLE IF NOT EXISTS MEMBER_SCORE_ORC(
      `MEMBER_ID` STRING,
    `SCORE` INT )
    STORED AS ORC
    TBLPROPERTIES ("orc.compress"="SNAPPY");

INSERT OVERWRITE TABLE MEMBER_SCORE_ORC
    SELECT
    MEMBER_ID,
    SCORE FROM MEMBER_SCORE_EXT;
```

```
hive> CREATE TABLE IF NOT EXISTS MEMBER_SCORE_ORC(
    >     `MEMBER_ID` STRING,
    > `SCORE` INT )
    > STORED AS ORC
    > TBLPROPERTIES ("orc.compress"="SNAPPY");
OK
Time taken: 0.051 seconds
hive> INSERT OVERWRITE TABLE MEMBER_SCORE_ORC
    > SELECT
    > MEMBER_ID,
    > SCORE FROM
    >     MEMBER_SCORE_EXT;
Query ID = hadoop_20240204173317_7231eaf3-6ec5-4ceb-abc6-f0ac07479f27
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1707048263009_0036)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      1          1        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 3.76 s
----------------------------------------------------------------------------------------------
Loading data to table default.member_score_orc
OK
Time taken: 4.791 seconds
hive> select count(1) from MEMBER_SCORE_ORC;
OK
999
Time taken: 0.137 seconds, Fetched: 1 row(s)
```

Command to see the list of imported data in HDFS

- Command to see first 10 values from CARD_MEMBER_ORC table.

SELECT * FROM CARD_MEMBER_ORC LIMIT 10;

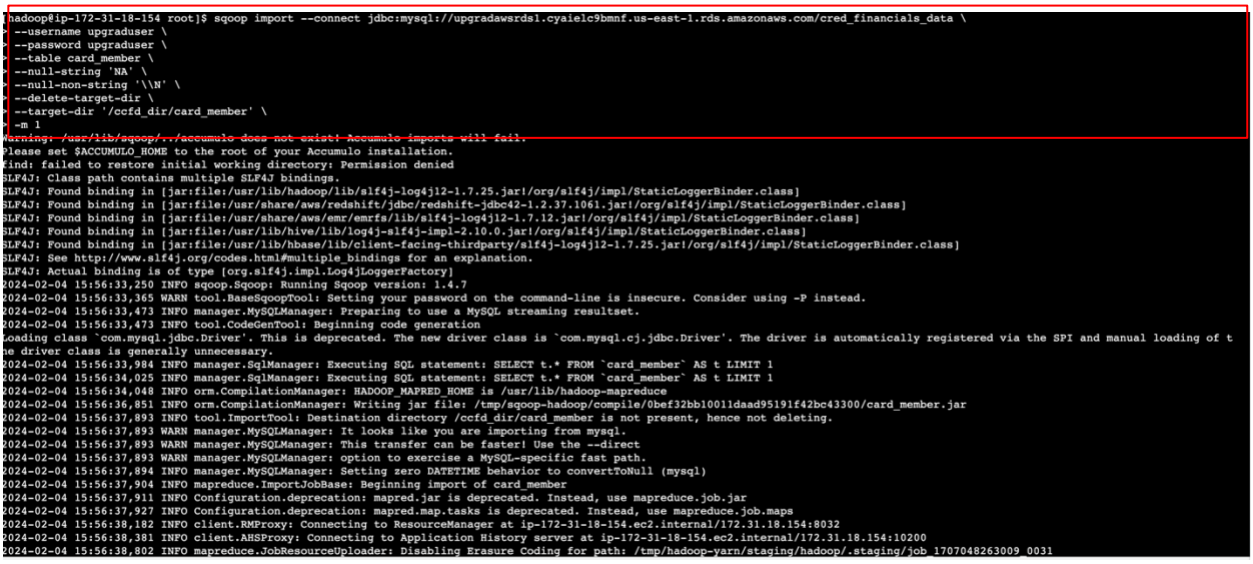
```
Time taken: 0.137 seconds, Fetched: 1 row(s)
hive> SELECT * FROM CARD_MEMBER_ORC LIMIT 10;
OK
340028465709212 009250698176266 2012-02-08 06:04:13     05/13   United States   Barberton
340054675199675 835873341185231 2017-03-10 09:24:44     03/17   United States   Fort Dodge
340082915339645 512969555857346 2014-02-15 06:30:30     07/14   United States   Graham
340134186926007 887711945571282 2012-02-05 01:21:58     02/13   United States   Dix Hills
340265728490548 680324265406190 2014-03-29 07:49:14     11/14   United States   Rancho Cucamonga
340268219434811 929799084911715 2012-07-08 02:46:08     08/12   United States   San Francisco
340379737226464 089615510858348 2010-03-10 00:06:42     09/10   United States   Clinton
340383645652108 181180599313885 2012-02-24 05:32:44     10/16   United States   West New York
340803866934451 417664728506297 2015-05-21 04:30:45     08/17   United States   Beaverton
340889618969736 459292914761635 2013-04-23 08:40:11     11/15   United States   West Palm Beach
Time taken: 0.131 seconds, Fetched: 10 row(s)
```

- Command to see the first 10 values from MEMBER_SCORE_ORC table.

SELECT * FROM MEMBER_SCORE_ORC LIMIT 10;

```
hive> SELECT * FROM MEMBER_SCORE_ORC LIMIT 10;
OK
000037495066290 339
000117826301530 289
001147922084344 393
001314074991813 225
001739553947511 642
003761426295463 413
004494068832701 217
006836124210484 504
006991872634058 697
007955566230397 372
Time taken: 0.117 seconds, Fetched: 10 row(s)
```

Screenshot of the imported data



```
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=247567
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=85
                HDFS: Number of bytes written=85081
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                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)=267648
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=2788
                Total vcore-milliseconds taken by all map tasks=2788
                Total megabyte-milliseconds taken by all map tasks=8564736
        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)=70
                CPU time spent (ms)=1500
                Physical memory (bytes) snapshot=307757056
                Virtual memory (bytes) snapshot=4403085312
                Total committed heap usage (bytes)=256376832
                Peak Map Physical memory (bytes)=307757056
                Peak Map Virtual memory (bytes)=4403085312
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=85081
2024-02-04 15:56:51,491 INFO mapreduce.ImportJobBase: Transferred 83.0869 KB in 13.5531 seconds (6.1305 KB/sec)
2024-02-04 15:56:51,493 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[hadoop@ip-172-31-18-154 root]$
```

```
[hadoop@ip-172-31-18-154 root]$ sqoop import --connect jdbc:mysql://upgradawsrds1.cyaielc9bmnf.us-east-1.rds.amazonaws.com/cred_financials_data \
> --username upgraduser \
> --password upgraduser \
> --table member_score \
> --null-string 'NA' \
> --null-non-string '\\N' \
> --delete-target-dir \
> --target-dir '/ccfd_dir/member_score' \
> -m 1
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
find: failed to restore initial working directory: Permission denied
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/share/aws/redshift/jdbc/redshift-jdbc42-1.2.37.1061.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/share/aws/emr/emrfs/lib/slf4j-log4j12-1.7.12.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hbase/lib/client-facing-thirdparty/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2024-02-04 15:58:28,527 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
2024-02-04 15:58:28,605 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2024-02-04 15:58:28,707 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2024-02-04 15:58:28,707 INFO tool.CodeGenTool: Beginning code generation
Loading class `com.mysql.jdbc.Driver'. This is deprecated. The new driver class is `com.mysql.cj.jdbc.Driver'. The driver is automatically registered via the SPI and manual loading of t
he driver class is generally unnecessary.
2024-02-04 15:58:29,165 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `member_score` AS t LIMIT 1
2024-02-04 15:58:29,206 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `member_score` AS t LIMIT 1
2024-02-04 15:58:29,228 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
2024-02-04 15:58:31,505 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-hadoop/compile/a036d96e4ea0155877542947e704e405/member_score.jar
2024-02-04 15:58:32,941 INFO tool.ImportTool: Destination directory /ccfd_dir/member_score is not present, hence not deleting.
2024-02-04 15:58:32,942 WARN manager.MySQLManager: It looks like you are importing from mysql.
2024-02-04 15:58:32,942 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
2024-02-04 15:58:32,942 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
2024-02-04 15:58:32,942 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
2024-02-04 15:58:32,953 INFO mapreduce.ImportJobBase: Beginning import of member_score
2024-02-04 15:58:32,959 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
2024-02-04 15:58:32,977 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2024-02-04 15:58:33,259 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-18-154.ec2.internal/172.31.18.154:8032
2024-02-04 15:58:33,514 INFO client.AHSProxy: Connecting to Application History server at ip-172-31-18-154.ec2.internal/172.31.18.154:10200
2024-02-04 15:58:34,026 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1707048263009_0032
```

```
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=247514
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=85
                HDFS: Number of bytes written=19980
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                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)=270432
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=2817
                Total vcore-milliseconds taken by all map tasks=2817
                Total megabyte-milliseconds taken by all map tasks=8653824
        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)=79
                CPU time spent (ms)=1400
                Physical memory (bytes) snapshot=310341632
                Virtual memory (bytes) snapshot=4397051904
                Total committed heap usage (bytes)=330825728
                Peak Map Physical memory (bytes)=310341632
                Peak Map Virtual memory (bytes)=4397051904
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=19980
2024-02-04 15:58:46,679 INFO mapreduce.ImportJobBase: Transferred 19.5117 KB in 13.6907 seconds (1.4252 KB/sec)
2024-02-04 15:58:46,682 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[hadoop@ip-172-31-18-154 root]$
```

The final row count is **999** for both the tables as mentioned in the project description.

```
Time taken: 0.137 seconds, Fetched: 1 row(s)
hive> SELECT * FROM CARD_MEMBER_ORC LIMIT 10;
OK
340028465709212 009250698176266 2012-02-08 06:04:13     05/13   United States   Barberton
340054675199675 835873341185231 2017-03-10 09:24:44     03/17   United States   Fort Dodge
340082915339645 512969555857346 2014-02-15 06:30:30     07/14   United States   Graham
340134186926007 887711945571282 2012-02-05 01:21:58     02/13   United States   Dix Hills
340265728490548 680324265406190 2014-03-29 07:49:14     11/14   United States   Rancho Cucamonga
340268219434811 929799084911715 2012-07-08 02:46:08     08/12   United States   San Francisco
340379737226464 089615510858348 2010-03-10 00:06:42     09/10   United States   Clinton
340383645652108 181180599313885 2012-02-24 05:32:44     10/16   United States   West New York
340803866934451 417664728506297 2015-05-21 04:30:45     08/17   United States   Beaverton
340889618969736 459292914761635 2013-04-23 08:40:11     11/15   United States   West Palm Beach
Time taken: 0.131 seconds, Fetched: 10 row(s)
```

```
hive> SELECT * FROM MEMBER_SCORE_ORC LIMIT 10;
OK
000037495066290 339
000117826301530 289
001147922084344 393
001314074991813 225
001739553947511 642
003761426295463 413
004494068832701 217
006836124210484 504
006991872634058 697
007955566230397 372
Time taken: 0.117 seconds, Fetched: 10 row(s)
```