Data Ingestion from the RDS to HDFS using Sqoop

sqoop import --connect jdbc:mysql://upgradawsrds1.cyaielc9bmnf.us-east-1.rds.amazonaws.com/cred_financials_data1 --username upgraduser --password upgraduser --table member_score --target-dir /user/root/cap_project/member_score -m 1 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/cap_project/card_member -m 1

```
21/11/21 04:26:46 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `card_member` AS t LIMIT 1
21/11/21 04:26:46 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `card_member` AS t LIMIT 1
21/11/21 04:26:46 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce
Note: /tmp/sqoop-root/compile/c0a7e7411263c48693f2aa4f0d0b3776/card_member.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
21/11/21 04:26:48 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/c0a7e7411263c48693f2aa4f0d0b3776/card_member.jar
21/11/21 04:26:48 WARN manager.MySQLManager: It looks like you are importing from mysql.
21/11/21 04:26:48 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
21/11/21 04:26:48 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
21/11/21 04:26:48 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
21/11/21 04:26:48 INFO mapreduce.ImportJobBase: Beginning import of card_member
21/11/21 04:26:49 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
21/11/21 04:26:51 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
21/11/21 04:26:51 INFO client.RMProxy: Connecting to ResourceManager at ip-10-0-0-32.ec2.internal/10.0.0.32:8032
21/11/21 04:27:01 INFO db.DBInputFormat: Using read committed transaction isolation
21/11/21 04:27:01 INFO mapreduce.JobSubmitter: number of splits:1
21/11/21 04:27:02 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1637467238214_0002
21/11/21 04:27:02 INFO impl.YarnClientImpl: Submitted application application_1637467238214_0002
21/11/21 04:27:03 INFO mapreduce.Job: The url to track the job: http://ip-10-0-0-32.ec2.internal:8088/proxy/application_1637467238214_0002/
21/11/21 04:27:03 INFO mapreduce.Job: Running job: job_1637467238214_0002
21/11/21 04:27:17 INFO mapreduce.Job: Job job_1637467238214_0002 running in uber mode : false
21/11/21 04:27:17 INFO mapreduce.Job:  map 0% reduce 0%
21/11/21 04:27:27 INFO mapreduce.Job:  map 100% reduce 0%
21/11/21 04:27:29 INFO mapreduce.Job: Job job_1637467238214_0002 completed successfully
21/11/21 04:27:29 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=176465
                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=85082
                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)=6362
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=6362
                Total vcore-milliseconds taken by all map tasks=6362
                Total megabyte-milliseconds taken by all map tasks=6514688
        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)=100
                CPU time spent (ms)=2190
                Physical memory (bytes) snapshot=250396672
                Virtual memory (bytes) snapshot=2804301824
                Total committed heap usage (bytes)=231735296
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=85082
21/11/21 04:27:29 INFO mapreduce.ImportJobBase: Transferred 83.0879 KB in 38.8085 seconds (2.141 KB/sec)
21/11/21 04:27:29 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[root@ip-10-0-0-32 ~]#
```

```
21/11/21 04:21:27 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `member_score` AS t LIMIT 1
21/11/21 04:21:27 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `member_score` AS t LIMIT 1
21/11/21 04:21:27 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce
Note: /tmp/sqoop-root/compile/00dd896f3c983cbaa6f53f516ff5f73f/member_score.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
21/11/21 04:21:31 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/00dd896f3c983cbaa6f53f516ff5f73f/member_score.jar
21/11/21 04:21:31 WARN manager.MySQLManager: It looks like you are importing from mysql.
21/11/21 04:21:31 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
21/11/21 04:21:31 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
21/11/21 04:21:31 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
21/11/21 04:21:31 INFO mapreduce.ImportJobBase: Beginning import of member_score
21/11/21 04:21:31 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
21/11/21 04:21:33 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
21/11/21 04:21:33 INFO client.RMProxy: Connecting to ResourceManager at ip-10-0-0-32.ec2.internal/10.0.0.32:8032
21/11/21 04:21:47 INFO db.DBInputFormat: Using read commited transaction isolation
21/11/21 04:21:48 INFO mapreduce.JobSubmitter: number of splits:1
21/11/21 04:21:48 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1637467238214_0001
21/11/21 04:21:50 INFO impl.YarnClientImpl: Submitted application application_1637467238214_0001
21/11/21 04:21:50 INFO mapreduce.Job: The url to track the job: http://ip-10-0-0-32.ec2.internal:8088/proxy/application_1637467238214_0001/
21/11/21 04:21:50 INFO mapreduce.Job: Running job: job_1637467238214_0001
21/11/21 04:22:07 INFO mapreduce.Job: Job job_1637467238214_0001 running in uber mode : false
21/11/21 04:22:07 INFO mapreduce.Job:  map 0% reduce 0%
21/11/21 04:22:16 INFO mapreduce.Job:  map 100% reduce 0%
21/11/21 04:22:24 INFO mapreduce.Job: Job job_1637467238214_0001 completed successfully
21/11/21 04:22:24 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=176415
                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=19980
                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)=7238
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=7238
                Total vcore-milliseconds taken by all map tasks=7238
                Total megabyte-milliseconds taken by all map tasks=7411712
        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)=100
                CPU time spent (ms)=1700
                Physical memory (bytes) snapshot=240119808
                Virtual memory (bytes) snapshot=2796888064
                Total committed heap usage (bytes)=246939648
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=19980
21/11/21 04:22:24 INFO mapreduce.ImportJobBase: Transferred 19.5117 KB in 51.1754 seconds (390.4222 bytes/sec)
21/11/21 04:22:24 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[root@ip-10-0-0-32 ~]#
```

<Command to see the list of imported data in HDFS> hadoop fs -ls /user/root/cap_project/

```
[root@ip-10-0-0-32 ~]# hadoop fs -ls /user/root/cap_project/
Found 2 items
drwxr-xr-x   - root supergroup          0 2021-11-21 04:27 /user/root/cap_project/card_member
drwxr-xr-x   - root supergroup          0 2021-11-21 04:22 /user/root/cap_project/member_score
[root@ip-10-0-0-32 ~]#
```

Renaming the files : hadoop fs -mv /user/root/cap_project/card_member/part-m-00000 /user/root/ cap_project/card_member/card_member.csv hadoop fs -mv /user/root/cap_project/member_score/part-m-00000 /user/root/

cap_project/member_score/member_score.csv Command to check the data imported: hadoop fs -cat /user/root/cap_project/card_member/card_member.csv hadoop fs -cat /user/root/cap_project/member_score/member_score.csv

```
[[root@ip-10-0-0-32 ~]# clear

[[root@ip-10-0-0-32 ~]# hadoop -fs -cat /user/root/cap_project/card_member/card_member.csv
Error: No command named '-fs' was found. Perhaps you meant 'hadoop fs'
[[root@ip-10-0-0-32 ~]# hadoop fs -cat /user/root/cap_project/card_member/card_member.csv
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
341151629171378,533670000048847,2013-05-13 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
341920010925925,185524819205415,2010-11-15 00:57:15.0,10/16,United States,Roosevelt
342016727481601,329679840450294,2015-02-15 06:56:32.0,10/16,United States,Oakdale
342048672911139,409346767454057,2011-03-03 10:41:02.0,08/12,United States,Winter Springs
342106756753728,439669104078192,2013-11-25 02:05:46.0,03/14,United States,Pittsburg
342136378319429,756971727303005,2016-01-09 01:19:50.0,10/17,United States,Superior
342341653931950,569294535026544,2014-09-06 12:16:48.0,10/14,United States,Woodstock
342400733369183,557238536043272,2016-03-01 07:30:53.0,08/17,United States,Derry
342409040326783,093291208532654,2015-05-02 10:06:45.0,04/16,United States,Key West
342541202178377,228647326620185,2013-12-28 05:34:26.0,04/15,United States,Taylors
342659393657862,075642075917789,2016-02-10 07:27:55.0,07/17,United States,Norman
342679959520647,328988632419725,2017-06-20 06:30:36.0,07/17,United States,Oklahoma City
342684196338179,329884605707231,2011-11-11 03:06:15.0,08/13,United States,Lisle
342853996970884,683933065094215,2011-11-04 12:32:30.0,02/14,United States,Bradley
342894449456021,645003407215198,2015-07-15 01:27:46.0,04/17,United States,Olive Branch
342917304431089,110913209929562,2015-01-02 09:32:57.0,03/16,United States,Arvada
342966201267661,015777120121252,2010-06-10 07:32:59.0,10/14,United States,Branford
343021795545168,034436901002830,2010-07-24 06:53:37.0,06/13,United States,North Bel Air
343107037888023,979274357975688,2014-07-05 01:41:46.0,12/16,United States,Inglewood-Finn Hill
343252340741735,055787011798644,2014-12-24 03:29:32.0,02/15,United States,Kissimmee
343330859464824,051518771748227,2011-02-09 04:11:31.0,11/14,United States,Murfreesboro
343478616471867,918968215840378,2013-08-11 10:10:44.0,09/17,United States,Pinewood
343753962914526,319436604377423,2017-02-28 08:59:13.0,07/17,United States,Cahokia
343794320536223,083447760133750,2015-02-28 09:56:55.0,10/15,United States,Alice
343802190729995,103489181611635,2015-11-19 04:09:19.0,12/16,United States,Laplace
343803912797603,374775668614336,2010-03-18 09:22:55.0,09/12,United States,McDonough
343902075851032,447204392852035,2016-04-04 09:12:08.0,01/17,United States,Lakeville
343923094022124,233154406228727,2013-05-02 08:50:27.0,11/16,United States,Lacey
343953209773343,495110335914510,2014-07-01 00:42:06.0,10/17,United States,Englewood
343962656588624,874267591301724,2017-09-13 04:32:55.0,09/17,United States,Ramona
343991629633818,372225917645777,2017-07-09 04:03:45.0,07/17,United States,Lake Shore
344000005693030,067585562947546,2010-09-08 03:40:42.0,01/14,United States,Wellesley
344002520206946,313129704156102,2012-04-08 12:06:09.0,12/14,United States,Burien
344413554194093,605980287528081,2016-07-03 01:16:26.0,07/16,United States,Savage
344209859893876,145224809882579,2012-06-05 01:24:25.0,04/15,United States,Trenton
344250415187512,787250659693155,2013-01-10 02:44:57.0,02/15,United States,Walnut Park
344261255055849,255246229177887,2013-10-12 01:36:31.0,08/16,United States,Hopatcong
```