

Data Ingestion from the RDS to HDFS - sqoop

Purpose of document: This is a pdf document which contains the code used for ingesting data from RDS server. (RDS to Hadoop.)

Sqoop command used for importing table from RDS to HDFS

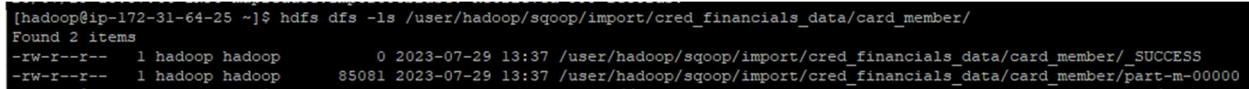
```
sqoop import --connect "jdbc:mysql://upgradawsrds1.cyaiehc9bmnf.us-east-1.rds.amazonaws.com/cred_financials_data?connectionCollation=latin1_swedish_ci" --table card_member --target-dir /user/hadoop/sqoop/import/cred_financials_data/card_member --username upgraduser -m 1 --password upgraduser
```

```
sqoop import --connect "jdbc:mysql://upgradawsrds1.cyaiehc9bmnf.us-east-1.rds.amazonaws.com/cred_financials_data?connectionCollation=latin1_swedish_ci" --table member_score --target-dir /user/hadoop/sqoop/import/cred_financials_data/member_score --username upgraduser -m 1 --password upgraduser
```

Command to see the list of imported data in HDFS:

```
hdfs dfs -ls /user/hadoop/sqoop/import/cred_financials_data/card_member  
hdfs dfs -cat /user/hadoop/sqoop/import/cred_financials_data/card_member/part-m-00000
```

Screenshot of the imported data:



```
[hadoop@ip-172-31-64-25 ~]$ hdfs dfs -ls /user/hadoop/sqoop/import/cred_financials_data/card_member/  
Found 2 items  
-rw-r--r-- 1 hadoop hadoop 0 2023-07-29 13:37 /user/hadoop/sqoop/import/cred_financials_data/card_member/_SUCCESS  
-rw-r--r-- 1 hadoop hadoop 85081 2023-07-29 13:37 /user/hadoop/sqoop/import/cred_financials_data/card_member/part-m-00000
```

```
[hadoop@ip-172-31-64-25 ~]$ hdfs dfs -cat /user/hadoop/sqoop/import/cred_financials_data/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
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
342679959520647,328980632419725,2017-06-20 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
342917304031089,110913209929562,2015-01-02 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
343923894022124,233154406228727,2013-05-02 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
```

Member Score data:

Command:

```
hdfs dfs -ls /user/hadoop/sqoop/import/cred_financials_data/member_score
```

```
hdfs dfs -cat /user/hadoop/sqoop/import/cred_financials_data/member_score/part-m-00000
```

Screenshot:

```
[hadoop@ip-172-31-64-25 ~]$ hdfs dfs -ls /user/hadoop/sqoop/import/cred_financials_data/member_score/
Found 2 items
-rw-r--r-- 1 hadoop hadoop 0 2023-07-29 13:39 /user/hadoop/sqoop/import/cred_financials_data/member_score/_SUCCESS
-rw-r--r-- 1 hadoop hadoop 19980 2023-07-29 13:39 /user/hadoop/sqoop/import/cred_financials_data/member_score/part-m-00000
[hadoop@ip-172-31-64-25 ~]$

[hadoop@ip-172-31-64-25 ~]$ hdfs dfs -cat /user/hadoop/sqoop/import/cred_financials_data/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
011877954983420,497
012390918683920,407
012731668664932,612
012871894714576,689
015235737327750,244
015509173543086,176
015582765997171,345
015777120121252,464
019834203768714,670
019999364464243,604
023329987813357,268
028379321561729,507
028632455787521,636
030262674475086,425
034436901002830,464
035217612427824,587
036571958569825,312
037011200807810,296
038888888888888,188
```

Capstone project – Credit Card Fraud Detection (mid submission)

by

Susil Patro, Krishna Mohan & Ashmeet Singh Deol

