

Data Ingestion from the RDS to HDFS using Sqoop

Sqoop command used for importing table from RDS to HDFS

<Sqoop Command>

```
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 '/capstone_project/member_score' -m1
```

```
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 '/capstone_project/card_member' -m 1
```

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

<Command used>

```
describe 'card_transactions_hive'
```

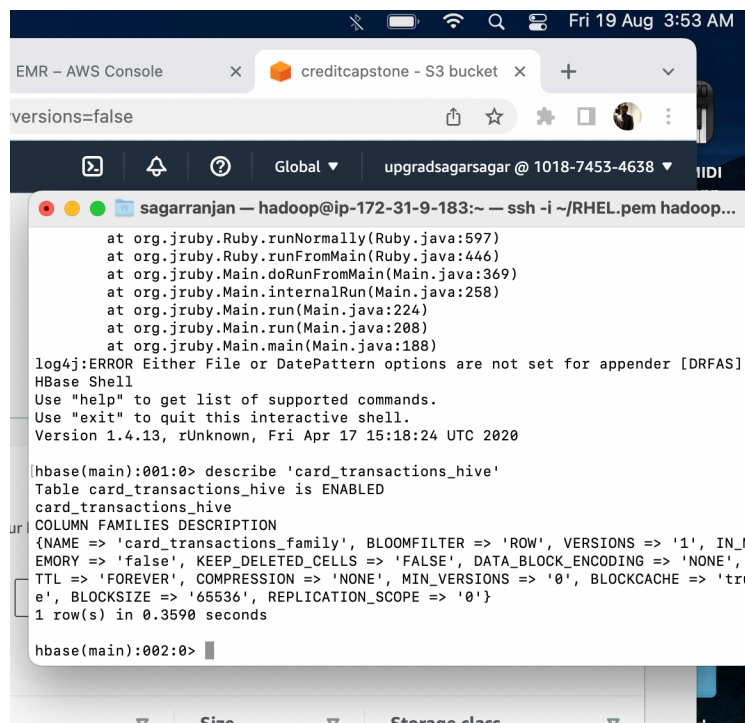
```
COUNT 'card_transactions_hive'
```

```
describe 'lookup_data_hive'
```

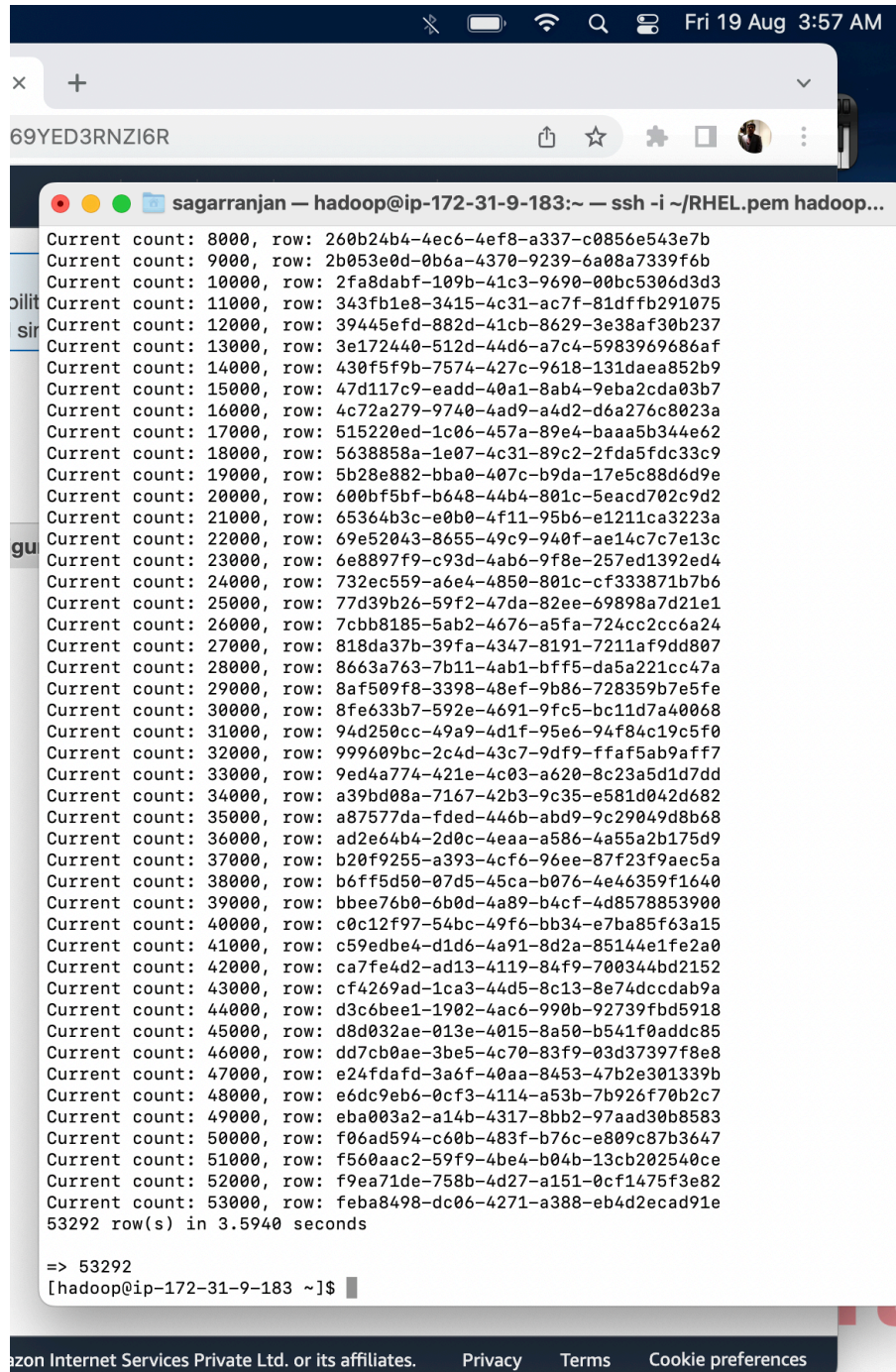
Screenshot of the imported data

<Screenshot of the imported data>

1>>



2>>



The screenshot shows a terminal window with a title bar indicating an SSH session to a Hadoop node. The command executed is `hadoop fs -ls /`, which lists the contents of the root directory of the Hadoop file system. The output shows a long list of files, each with its permissions, size, and name. The files are listed in ascending order of size, with the largest files at the top. The terminal window also shows the prompt `=> 53292` and the command `[hadoop@ip-172-31-9-183 ~]$`.

```

sagarranjan — hadoop@ip-172-31-9-183:~ — ssh -i ~/RHEL.pem hadoop...
Current count: 8000, row: 260b24b4-4ec6-4ef8-a337-c0856e543e7b
Current count: 9000, row: 2b053e0d-0b6a-4370-9239-6a08a7339f6b
Current count: 10000, row: 2fa8dabf-109b-41c3-9690-00bc5306d3d3
Current count: 11000, row: 343fb1e8-3415-4c31-ac7f-81dfffb291075
Current count: 12000, row: 39445efd-882d-41cb-8629-3e38af30b237
Current count: 13000, row: 3e172440-512d-44d6-a7c4-5983969686af
Current count: 14000, row: 430f5f9b-7574-427c-9618-131daea852b9
Current count: 15000, row: 47d117c9-eadd-40a1-8ab4-9eba2cda03b7
Current count: 16000, row: 4c72a279-9740-4ad9-a4d2-d6a276c8023a
Current count: 17000, row: 515220ed-1c06-457a-89e4-baaa5b344e62
Current count: 18000, row: 5638858a-1e07-4c31-89c2-2fda5fdc33c9
Current count: 19000, row: 5b28e882-bba0-407c-b9da-17e5c88d6d9e
Current count: 20000, row: 600bf5bf-b648-44b4-801c-5eacd702c9d2
Current count: 21000, row: 65364b3c-e0b0-4f11-95b6-e1211ca3223a
Current count: 22000, row: 69e52043-8655-49c9-940f-ae14c7c7e13c
Current count: 23000, row: 6e8897f9-c93d-4ab6-9f8e-257ed1392ed4
Current count: 24000, row: 732ec559-a6e4-4850-801c-cf333871b7b6
Current count: 25000, row: 77d39b26-59f2-47da-82ee-69898a7d21e1
Current count: 26000, row: 7cbb8185-5ab2-4676-a5fa-724cc2cc6a24
Current count: 27000, row: 818da37b-39fa-4347-8191-7211af9dd807
Current count: 28000, row: 8663a763-7b11-4ab1-bff5-da5a221cc47a
Current count: 29000, row: 8af509f8-3398-48ef-9b86-728359b7e5fe
Current count: 30000, row: 8fe633b7-592e-4691-9fc5-bc11d7a40068
Current count: 31000, row: 94d250cc-49a9-4d1f-95e6-94f84c19c5f0
Current count: 32000, row: 999609bc-2c4d-43c7-9df9-ffaf5ab9aff7
Current count: 33000, row: 9ed4a774-421e-4c03-a620-8c23a5d1d7dd
Current count: 34000, row: a39bd08a-7167-42b3-9c35-e581d042d682
Current count: 35000, row: a87577da-fded-446b-abd9-9c29049d8b68
Current count: 36000, row: ad2e64b4-2d0c-4eaa-a586-4a55a2b175d9
Current count: 37000, row: b20f9255-a393-4cf6-96ee-87f23f9aec5a
Current count: 38000, row: b6ff5d50-07d5-45ca-b076-4e46359f1640
Current count: 39000, row: bbee76b0-6b0d-4a89-b4cf-4d8578853900
Current count: 40000, row: c0c12f97-54bc-49f6-bb34-e7ba85f63a15
Current count: 41000, row: c59edbe4-d1d6-4a91-8d2a-85144e1fe2a0
Current count: 42000, row: ca7fe4d2-ad13-4119-84f9-700344bd2152
Current count: 43000, row: cf4269ad-1ca3-44d5-8c13-8e74dcccab9a
Current count: 44000, row: d3c6bee1-1902-4ac6-990b-92739fbd5918
Current count: 45000, row: d8d032ae-013e-4015-8a50-b541f0addc85
Current count: 46000, row: dd7cb0ae-3be5-4c70-83f9-03d37397f8e8
Current count: 47000, row: e24fdafd-3a6f-40aa-8453-47b2e301339b
Current count: 48000, row: e6dc9eb6-0cf3-4114-a53b-7b926f70b2c7
Current count: 49000, row: eba003a2-a14b-4317-8bb2-97aad30b8583
Current count: 50000, row: f06ad594-c60b-483f-b76c-e809c87b3647
Current count: 51000, row: f560aac2-59f9-4be4-b04b-13cb202540ce
Current count: 52000, row: f9ea71de-758b-4d27-a151-0cf1475f3e82
Current count: 53000, row: feba8498-dc06-4271-a388-eb4d2ecad91e
53292 row(s) in 3.5940 seconds

=> 53292
[hadoop@ip-172-31-9-183 ~]$

```

3>>

```

Fri 19 Aug 3:59 AM
sails-j-269YED3RNZI6R
sagarranjan — hadoop@ip-172-31-9-183:~ — ssh -i ~/RHEL.pem hadoop...
at org.apache.log4j.LogManager.<clinit>(LogManager.java:127)
at org.apache.log4j.Logger.getLogger(Logger.java:104)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccess
sorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.jruby.javasupport.JavaMethod.invokeDirectWithExceptionHandling(Ja
vaMethod.java:450)
at org.jruby.javasupport.JavaMethod.invokeStaticDirect(JavaMethod.java:3
62)
at org.jruby.java.invokers.StaticMethodInvoker.call(StaticMethodInvoker.
java:58)
at org.jruby.runtime.callsite.CachingCallSite.cacheAndCall(CachingCallSi
te.java:312)
at org.jruby.runtime.callsite.CachingCallSite.call(CachingCallSite.java:
169)
at usr.lib.hbase.bin.hirb.__file__(/usr/lib/hbase/bin/hirb.rb:129)
at usr.lib.hbase.bin.hirb.load(/usr/lib/hbase/bin/hirb.rb)
at org.jruby.Ruby.runScript(Ruby.java:697)
at org.jruby.Ruby.runScript(Ruby.java:690)
at org.jruby.Ruby.runNormally(Ruby.java:597)
at org.jruby.Ruby.runFromMain(Ruby.java:446)
at org.jruby.Main.doRunFromMain(Main.java:369)
at org.jruby.Main.internalRun(Main.java:258)
at org.jruby.Main.run(Main.java:224)
at org.jruby.Main.run(Main.java:208)
at org.jruby.Main.main(Main.java:188)
log4j:ERROR Either File or DatePattern options are not set for appender [DRFAS].
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
Version 1.4.13, rUnknown, Fri Apr 17 15:18:24 UTC 2020

[hbase(main):001:0> describe 'lookup_data_hive'
Table lookup_data_hive is ENABLED
lookup_data_hive
COLUMN FAMILIES DESCRIPTION
{NAME => 'lookup_card_family', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY
=> 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL =>
'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BL
OCKSIZE => '65536', REPLICATION_SCOPE => '0'}
{NAME => 'lookup_transaction_family', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_
MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE',
TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'tr
ue', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}
2 row(s) in 0.3240 seconds

hbase(main):002:0>

```


4>>

```

FILE: Number of write operations=0
HDFS: Number of bytes read=87
HDFS: Number of bytes written=85081
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)=232560
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=4845
  Total vcore-milliseconds taken by all map tasks=4845
  Total megabyte-milliseconds taken by all map tasks=7441920
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)=83
  CPU time spent (ms)=2800
  Physical memory (bytes) snapshot=286519296
  Virtual memory (bytes) snapshot=3303858176
  Total committed heap usage (bytes)=242745344
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=85081
22/08/19 21:46:00 INFO mapreduce.ImportJobBase: Transferred 83.0869 KB in 20.0173 seconds (4.1508 KB/sec)
22/08/19 21:46:00 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[hadoop@ip-172-31-2-146 mysql-connector-java-8.0.25]$

```