

Ingestion data into Hbase table from RDS table and using a python script

Step 1: Login to EMR cluster.

Command:

ssh -i EMR-cluster-key.pem hadoop@ec2-54-236-6-90.compute-1.amazonaws.com

hbase shell # To enter Hbase shell

```
[root@ip-172-31-27-81 ~]# hbase shell
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>
```

Step 2: Create a directory named 'hdfs-map' under hdfs '/user/

Commands:

hadoop fs -ls / # to list under / file system

hadoop fs -ls /user # to list under /user

hadoop fs -mkdir /user/hdfs-map # To create directory under /user

hadoop fs -ls /user/hdfs-map # to list under /user/hdfs-map

```
[hadoop@ip-172-31-27-81 root]$
[hadoop@ip-172-31-27-81 root]$
[hadoop@ip-172-31-27-81 root]$ hadoop fs -ls /
Found 4 items
drwxr-xr-x   - hdfs hadoop          0 2023-05-07 15:26 /apps
drwxrwxrwt   - hdfs hadoop          0 2023-05-07 15:26 /tmp
drwxr-xr-x   - hdfs hadoop          0 2023-05-07 15:56 /user
drwxr-xr-x   - hdfs hadoop          0 2023-05-07 15:26 /var
[hadoop@ip-172-31-27-81 root]$
[hadoop@ip-172-31-27-81 root]$ hadoop fs -ls /user
Found 5 items
drwxrwxrwx   - hadoop hadoop        0 2023-05-07 15:26 /user/hadoop
drwxrwxr-x   - hbase hbase          0 2023-05-07 15:27 /user/hbase
drwxr-xr-x   - mapred mapred        0 2023-05-07 15:26 /user/history
drwxrwxrwx   - hdfs hadoop          0 2023-05-07 15:26 /user/hive
drwxrwxrwx   - root hadoop          0 2023-05-07 15:26 /user/root
[hadoop@ip-172-31-27-81 root]$ hadoop fs -mkdir /user/hdfs-map
[hadoop@ip-172-31-27-81 root]$ hadoop fs -ls /user/hdfs-map
[hadoop@ip-172-31-27-81 root]$
```

Step 3: Checking the connectivity between RDS and EMR cluster using jdbc and listing all the databases in RDS.

Command:

```
sqoop-list-databases --connect jdbc:mysql://database-1.czxy6rglkiuf.us-east-1.rds.amazonaws.com:3306/ --username admin --password 123456789
```

```
hadoop@ip-172-31-27-81 ~$ sqoop-list-databases --connect jdbc:mysql://database-1.czxy6rglkiuf.us-east-1.rds.amazonaws.com:3306/ --username admin --password 123456789
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
23/05/07 16:13:27 INFO Sqoop.Sqoop: Running Sqoop version: 1.4.7
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/share/aws/redshift/jdbc/redshift-jdbc42-1.2.37.1061.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.4.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
23/05/07 16:13:27 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/05/07 16:13:27 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
Loading class 'com.mysql.jdbc.Driver'. This is deprecated. The new driver class is 'com.mysql.cj.jdbc.Driver'. The driver is automatically registered via the SPI and manual loading of the driver class is generally unnecessary.
mysql
information_schema
performance_schema
sys
trip
hadoop@ip-172-31-27-81 ~$
```

Step 4: Listing the tables in Hbase.

Command:

list # to list existing tables

```
[root@ip-172-31-27-81 mapreduce-assignment]#
[root@ip-172-31-27-81 mapreduce-assignment]# hbase shell
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> list
TABLE
0 row(s) in 0.2480 seconds

=> []
hbase(main):002:0>
```

Step 5: Importing data from RDS table to EMR Hbase.

Command:

```
sqoop import --connect "jdbc:mysql://database-1.czxy6rglkiuf.us-east-1.rds.amazonaws.com:3306/trip" --username admin --password 123456789 --table yello --columns
```

```
"VendorID,tpep_pickup_datetime,tpep_dropoff_datetime,passenger_count,trip_distance,Ra
tecodeID,store_and_fwd_flag,PULocationID,DOLocationID,payment_type,fare_amount,extra
_mta_tax,tip_amount,tolls_amount,improvement_surcharge,total_amount,Airport_fee" --
hbase-create-table --hbase-table yello --column-family trip_details --hbase-row-key
VendorID,tpep_pickup_datetime,tpep_dropoff_datetime --split-by tpep_dropoff_datetime -
m 8
```

```
[root@ip-172-31-16-199 mysql-connector-java-8.0.25]#
[root@ip-172-31-16-199 mysql-connector-java-8.0.25]#
[root@ip-172-31-16-199 mysql-connector-java-8.0.25]# sqoop import --connect 'jdbc:mysql://database-1.crxydrglkuf.us-east-1.rds.amazonaws.com:3306/trip' --username admin --password 123456789 --table yellow --column
na "VendorID, tpep_pickup_datetime, tpep_dropoff_datetime, passenger_count, trip_distance, RatecodeID, store_and_fwd_flag, PULocationID, DOLocationID, payment_type, fare_amount, extra, mta_tax, tip, amount, tolls_amount, improve
ment_surcharge, total_amount, Airport_fee" --hbase-create-table --hbase-table yellow --column-family trip_details --hbase-row-key VendorID, tpep_pickup_datetime, tpep_dropoff_datetime
--e
Warning: /usr/lib/sqoop/.accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
23/05/08 06:28:09 INFO SqoopRunner: Running Sqoop version: 1.4.7
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.10.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/lib/hive/lib/log4j-slf4j-impl-2.6.2.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]
23/05/08 06:28:09 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/05/08 06:28:09 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
23/05/08 06:28:09 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 the driver class is generally
unnecessary.
23/05/08 06:28:09 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'yellow' AS t LIMIT 1
23/05/08 06:28:09 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'yellow' AS t LIMIT 1
23/05/08 06:28:09 INFO com.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-root/compile/A98abdf48ce8e34fc08db326f738b10/yellow.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
23/05/08 06:28:13 INFO com.CompilationManager: writing jar file: /tmp/sqoop-root/compile/A98abdf48ce8e34fc08db326f738b10/yellow.jar
23/05/08 06:28:13 WARN manager.MySQLManager: It looks like you are importing from mysql.
23/05/08 06:28:13 INFO manager.MySQLManager: This transfer can be faster! Use the --direct
23/05/08 06:28:13 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
23/05/08 06:28:13 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
23/05/08 06:28:13 INFO mapreduce.ImportJobBase: Beginning import of yellow
23/05/08 06:28:14 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
23/05/08 06:28:14 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
23/05/08 06:28:17 INFO mapreduce.HBaseImportJob: Creating missing HBase table yellow
23/05/08 06:28:18 WARN mapreduce.TableMapReduceUtil: The addDependencyJars(Configuration, Class<?>...) method has been deprecated since it is easy to use incorrectly. Most users should rely on addDependencyJars()
ob instead. See HBASE-8386 for more details.
23/05/08 06:28:20 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-16-199.ec2.internal:172.31.16.199:8032
23/05/08 06:28:32 INFO db.DatabaseInputFormat: Using read committed transaction isolation
23/05/08 06:28:32 INFO db.DatabaseInputFormat: BoundingValueQuery: SELECT MIN('tpep_dropoff_datetime'), MAX('tpep_dropoff_datetime') FROM 'yellow'
23/05/08 06:28:33 INFO db.IntegralSplitter: Split size: 877977625; Num splits: 8 from: 1481380818000 to: 148841834000
23/05/08 06:28:33 INFO mapreduce.JobSubmitter: number of splits:8
23/05/08 06:28:34 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1683526383858_0001
23/05/08 06:28:34 INFO impl.VarClientImpl: Submitted application application_1683526383858_0001
23/05/08 06:28:34 INFO mapreduce.Job: The url to track the job: http://ip-172-31-16-199.ec2.internal:20888/proxy/application_1683526383858_0001/
23/05/08 06:28:34 INFO mapreduce.Job: Running job: job_1683526383858_0001
23/05/08 06:28:34 INFO mapreduce.Job: Job job_1683526383858_0001 running in uber mode : false
23/05/08 06:28:34 INFO mapreduce.Job: map 0% reduce 0%
23/05/08 06:28:34 INFO mapreduce.Job: map 13% reduce 0%
23/05/08 06:28:34 INFO mapreduce.Job: map 26% reduce 0%
23/05/08 06:28:35 INFO mapreduce.Job: map 38% reduce 0%
```

```
Warning: /usr/lib/sqoop/.accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
23/05/08 16:21:26 INFO tool.CodeGenTool: Beginning code generation
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.10.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/lib/hive/lib/log4j-slf4j-impl-2.6.2.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]
Enter password:
23/05/08 16:21:26 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
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 the driver class is generally unnecessary.
23/05/08 16:21:27 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'yellow' AS t LIMIT 1
23/05/08 16:21:27 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'yellow' AS t LIMIT 1
23/05/08 16:21:27 INFO com.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-hadoop/compile/796c4ab577d6df4d77ecb83772808/yellow.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
23/05/08 16:21:38 INFO com.CompilationManager: writing jar file: /tmp/sqoop-hadoop/compile/796c4ab577d6df4d77ecb83772808/yellow.jar
23/05/08 16:21:38 WARN manager.MySQLManager: It looks like you are importing from mysql.
23/05/08 16:21:38 INFO manager.MySQLManager: This transfer can be faster! Use the --direct
23/05/08 16:21:38 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
23/05/08 16:21:38 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
23/05/08 16:21:38 INFO mapreduce.ImportJobBase: Beginning import of yellow
23/05/08 16:21:38 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
23/05/08 16:21:38 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
23/05/08 16:21:31 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-19-244.ec2.internal:172.31.19.244:8032
23/05/08 16:21:37 INFO db.DatabaseInputFormat: BoundingValueQuery: SELECT MIN('tpep_dropoff_datetime'), MAX('tpep_dropoff_datetime') FROM 'yellow'
23/05/08 16:21:37 INFO db.IntegralSplitter: Split size: 877977625; Num splits: 8 from: 1481380818000 to: 148841834000
23/05/08 16:21:34 INFO mapreduce.JobSubmitter: number of splits:8
23/05/08 16:21:34 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1683561592377_0003
23/05/08 16:21:34 INFO impl.VarClientImpl: Submitted application application_1683561592377_0003
23/05/08 16:21:34 INFO mapreduce.Job: The url to track the job: http://ip-172-31-19-244.ec2.internal:20888/proxy/application_1683561592377_0003/
23/05/08 16:21:34 INFO mapreduce.Job: Running job: job_1683561592377_0003
23/05/08 16:21:34 INFO mapreduce.Job: Job job_1683561592377_0003 running in uber mode : false
23/05/08 16:21:34 INFO mapreduce.Job: map 0% reduce 0%
23/05/08 16:21:34 INFO mapreduce.Job: map 13% reduce 0%
23/05/08 16:21:34 INFO mapreduce.Job: map 26% reduce 0%
23/05/08 16:21:34 INFO mapreduce.Job: map 38% reduce 0%
23/05/08 16:21:34 INFO mapreduce.Job: map 50% reduce 0%
23/05/08 16:21:34 INFO mapreduce.Job: map 63% reduce 0%
23/05/08 16:21:34 INFO mapreduce.Job: map 75% reduce 0%
23/05/08 16:21:34 INFO mapreduce.Job: map 88% reduce 0%
23/05/08 16:21:34 INFO mapreduce.Job: map 100% reduce 0%
23/05/08 16:21:34 INFO mapreduce.Job: Job job_1683561592377_0003 completed successfully
23/05/08 16:21:39 INFO mapreduce.Job: Counters: 31
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=1517248
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=1481
  HDFS: Number of bytes written=180799616
  HDFS: Number of read operations=32
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=16
Job Counters
  Killed map tasks=3
  Launched map tasks=11
  Other local map tasks=11
  Total time spent by all maps in occupied slots (ms)=6411568
  Total time spent by all reducers in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=1175241
  Total vcore-millisecsd taken by all map tasks=1175241
```

```
23/05/08 16:24:16 INFO impl.YarnClientImpl: Submitted application application_1683561592377_0003
23/05/08 16:24:16 INFO mapreduce.Job: The url to track the job: http://ip-172-31-19-244.ec2.internal:20888/proxy/application_1683561592377_0003/
23/05/08 16:24:16 INFO mapreduce.Job: Running job: job_1683561592377_0003
23/05/08 16:24:25 INFO mapreduce.Job: Job job_1683561592377_0003 running in uber mode : false
23/05/08 16:24:25 INFO mapreduce.Job: map 0% reduce 0%
23/05/08 16:26:16 INFO mapreduce.Job: map 13% reduce 0%
23/05/08 16:26:17 INFO mapreduce.Job: map 25% reduce 0%
23/05/08 16:26:45 INFO mapreduce.Job: map 38% reduce 0%
23/05/08 16:26:47 INFO mapreduce.Job: map 50% reduce 0%
23/05/08 16:28:38 INFO mapreduce.Job: map 63% reduce 0%
23/05/08 16:28:44 INFO mapreduce.Job: map 75% reduce 0%
23/05/08 16:29:25 INFO mapreduce.Job: map 88% reduce 0%
23/05/08 16:29:28 INFO mapreduce.Job: map 100% reduce 0%
23/05/08 16:29:29 INFO mapreduce.Job: Job job_1683561592377_0003 completed successfully
23/05/08 16:29:29 INFO mapreduce.Job: Counters: 31
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=1517248
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=1461
    HDFS: Number of bytes written=1890799616
    HDFS: Number of read operations=32
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=16
  Job Counters
    Killed map tasks=3
    Launched map tasks=11
    Other local map tasks=11
    Total time spent by all maps in occupied slots (ms)=56411568
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=1175241
    Total vcore-milliseconds taken by all map tasks=1175241
    Total megabyte-milliseconds taken by all map tasks=1885170176
  Map-Reduce Framework
    Map input records=18880595
    Map output records=18880595
    Input split bytes=1461
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=2940
    CPU time spent (ms)=241760
    Physical memory (bytes) snapshot=3915235328
    Virtual memory (bytes) snapshot=26396659712
    Total committed heap usage (bytes)=3311403008
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=1890799616
23/05/08 16:29:29 INFO mapreduce.ImportJobBase: Transferred 1.7609 GB in 357.9347 seconds (5.0378 MB/sec)
23/05/08 16:29:29 INFO mapreduce.ImportJobBase: Retrieved 18880595 records.
[hadoop@ip-172-31-19-244 ~]$
```



```

[root@ip-172-31-26-225 ~]# hbase shell
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> list
TABLE
yello_trip
1 row(s) in 0.5770 seconds

=> ["yello_trip"]
hbase(main):002:0> count 'yello_trip'
Current count: 1000, row: 1_2017-01-01 00:11:36.0_2017-01-01 00:24:48.0
Current count: 2000, row: 1_2017-01-01 00:18:15.0_2017-01-01 00:26:33.0
Current count: 3000, row: 1_2017-01-01 00:23:58.0_2017-01-01 00:42:06.0
Current count: 4000, row: 1_2017-01-01 00:29:40.0_2017-01-01 00:46:41.0
Current count: 5000, row: 1_2017-01-01 00:35:28.0_2017-01-01 00:48:58.0
Current count: 6000, row: 1_2017-01-01 00:41:06.0_2017-01-01 01:00:39.0
Current count: 7000, row: 1_2017-01-01 00:46:43.0_2017-01-01 01:06:23.0
Current count: 8000, row: 1_2017-01-01 00:52:09.0_2017-01-01 00:59:09.0
Current count: 9000, row: 1_2017-01-01 00:57:37.0_2017-01-01 01:11:55.0
Current count: 10000, row: 1_2017-01-01 01:02:58.0_2017-01-01 01:09:05.0
Current count: 11000, row: 1_2017-01-01 01:08:31.0_2017-01-01 01:14:57.0
Current count: 12000, row: 1_2017-01-01 01:13:50.0_2017-01-01 01:25:21.0
Current count: 13000, row: 1_2017-01-01 01:19:34.0_2017-01-01 01:27:16.0
Current count: 14000, row: 1_2017-01-01 01:25:05.0_2017-01-01 01:36:39.0
Current count: 15000, row: 1_2017-01-01 01:30:50.0_2017-01-01 01:39:13.0
Current count: 16000, row: 1_2017-01-01 01:36:48.0_2017-01-01 01:42:39.0
Current count: 17000, row: 1_2017-01-01 01:42:38.0_2017-01-01 01:51:12.0
Current count: 18000, row: 1_2017-01-01 01:48:15.0_2017-01-01 02:03:10.0
Current count: 19000, row: 1_2017-01-01 01:54:15.0_2017-01-01 02:09:29.0
Current count: 20000, row: 1_2017-01-01 02:00:02.0_2017-01-01 02:33:57.0
Current count: 21000, row: 1_2017-01-01 02:06:23.0_2017-01-01 02:10:00.0
Current count: 22000, row: 1_2017-01-01 02:12:21.0_2017-01-01 02:25:16.0
Current count: 23000, row: 1_2017-01-01 02:18:02.0_2017-01-01 02:39:10.0
Current count: 24000, row: 1_2017-01-01 02:24:12.0_2017-01-01 02:33:02.0
Current count: 25000, row: 1_2017-01-01 02:30:08.0_2017-01-01 02:37:26.0
Current count: 26000, row: 1_2017-01-01 02:36:28.0_2017-01-01 03:06:36.0
Current count: 27000, row: 1_2017-01-01 02:42:39.0_2017-01-01 03:07:05.0
Current count: 28000, row: 1_2017-01-01 02:48:47.0_2017-01-01 03:18:16.0
Current count: 29000, row: 1_2017-01-01 02:55:12.0_2017-01-01 03:20:59.0
Current count: 30000, row: 1_2017-01-01 03:01:54.0_2017-01-01 03:25:22.0
Current count: 31000, row: 1_2017-01-01 03:08:42.0_2017-01-01 03:29:51.0
Current count: 32000, row: 1_2017-01-01 03:15:29.0_2017-01-01 03:23:57.0
Current count: 33000, row: 1_2017-01-01 03:22:43.0_2017-01-01 03:32:34.0
Current count: 34000, row: 1_2017-01-01 03:29:54.0_2017-01-01 03:54:30.0
Current count: 35000, row: 1_2017-01-01 03:37:23.0_2017-01-01 03:47:56.0
Current count: 36000, row: 1_2017-01-01 03:45:15.0_2017-01-01 04:06:54.0
Current count: 37000, row: 1_2017-01-01 03:53:03.0_2017-01-01 04:10:03.0
Current count: 38000, row: 1_2017-01-01 04:01:01.0_2017-01-01 04:14:34.0
Current count: 39000, row: 1_2017-01-01 04:08:58.0_2017-01-01 04:26:40.0
Current count: 40000, row: 1_2017-01-01 04:16:57.0_2017-01-01 04:31:43.0
Current count: 41000, row: 1_2017-01-01 04:26:03.0_2017-01-01 04:31:49.0
Current count: 42000, row: 1_2017-01-01 04:36:03.0_2017-01-01 05:00:44.0
Current count: 43000, row: 1_2017-01-01 04:48:14.0_2017-01-01 04:59:50.0
Current count: 44000, row: 1_2017-01-01 05:02:45.0_2017-01-01 05:11:55.0

```

Step 6: Importing data into hbase table from yellow_tripdata_2017-03.csv and yellow_tripdata_2017-04.csv using python batch file.

Command:

python batch_ingest.py

```
[root@ip-172-31-21-63 mapreduce-assignment]# python batch_ingest.py
Starting batch insert
batch insert started for file: yellow_tripdata_2017-03.csv
inside batch:
0 Lines loaded
1000 Lines loaded
2000 Lines loaded
3000 Lines loaded
4000 Lines loaded
5000 Lines loaded
6000 Lines loaded
7000 Lines loaded
8000 Lines loaded
9000 Lines loaded
10000 Lines loaded
11000 Lines loaded
12000 Lines loaded
13000 Lines loaded
14000 Lines loaded
15000 Lines loaded
16000 Lines loaded
17000 Lines loaded
18000 Lines loaded
19000 Lines loaded
20000 Lines loaded
21000 Lines loaded
22000 Lines loaded
23000 Lines loaded
24000 Lines loaded
25000 Lines loaded
26000 Lines loaded
27000 Lines loaded
28000 Lines loaded
29000 Lines loaded
30000 Lines loaded
31000 Lines loaded
32000 Lines loaded
33000 Lines loaded
34000 Lines loaded
35000 Lines loaded
36000 Lines loaded
37000 Lines loaded
38000 Lines loaded
39000 Lines loaded
40000 Lines loaded
41000 Lines loaded
42000 Lines loaded
43000 Lines loaded
44000 Lines loaded
45000 Lines loaded
46000 Lines loaded
47000 Lines loaded
48000 Lines loaded
49000 Lines loaded
50000 Lines loaded
51000 Lines loaded
52000 Lines loaded
53000 Lines loaded
```