Problem Statement:

Implement the concept given in below blog link and share the complete steps along with

screenshots.

https://acadgild.com/blog/loading-data-into-hbase-using-pig-scripts/

Solution:

- Copying the data set in to HDFS which will be further loaded into HBase

```
put: /home/acadgild/hadoop: No such file or directory
[acadgild@localhost ~]$ hadoop fs -put student.txt /user/acadgild/hadoop
17/11/24 03:11:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
applicable
```

```
[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/hadoop
17/11/24 03:12:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
applicable
Found 15 items
drwxr-xr-x   - acadgild supergroup          0 2017-10-20 02:44 /user/acadgild/hadoop/InvalidFilter
drwxr-xr-x   - acadgild supergroup          0 2017-10-20 06:41 /user/acadgild/hadoop/TotalUnitOutput
drwxr-xr-x   - acadgild supergroup          0 2017-10-20 07:55 /user/acadgild/hadoop/TotalUnitResult
-rwxrwx---   1 acadgild supergroup         84 2017-10-12 00:07 /user/acadgild/hadoop/max-temp.txt
drwxr-xr-x   - acadgild supergroup          0 2017-09-29 14:05 /user/acadgild/hadoop/maxout
drwxr-xr-x   - acadgild supergroup          0 2017-10-19 20:48 /user/acadgild/hadoop/output
drwxr-xr-x   - acadgild supergroup          0 2017-10-20 01:46 /user/acadgild/hadoop/output1
drwxr-xr-x   - acadgild supergroup          0 2017-10-20 02:37 /user/acadgild/hadoop/output3
drwxr-xr-x   - acadgild supergroup          0 2017-10-20 03:50 /user/acadgild/hadoop/output4
drwxr-xr-x   - acadgild supergroup          0 2017-10-20 03:57 /user/acadgild/hadoop/output5
-rw-r--r--   1 acadgild supergroup      21007 2017-09-29 14:02 /user/acadgild/hadoop/sample_temperature_dataset.csv
-rw-r--r--   1 acadgild supergroup      26204 2017-11-24 03:11 /user/acadgild/hadoop/student.txt
-rw-r--r--   1 acadgild supergroup        700 2017-10-19 20:48 /user/acadgild/hadoop/television.txt
-rw-r--r--   1 acadgild supergroup          0 2017-10-04 00:30 /user/acadgild/hadoop/word-count
-rwxrwx---   1 acadgild supergroup         47 2017-10-04 00:33 /user/acadgild/hadoop/word-count.txt
[acadgild@localhost ~]$
```
Activate Windows
Go to Settings to activate Windows.

- Including few jar files of HBase to the Pig classpath.

PIG_CLASSPATH=/home/hadoop/HADOOP/hbase-0.98.4-hadoop2/lib/hbase-server-0.98.4-
hadoop2:/home/hadoop/HADOOP/hbase-0.98.4-hadoop2/lib/hbase-*.jar;

- Start HBase shell and create a table.

- We only need this table as skeleton so PIG can Store data inside this by referring the table name.

```
[acadgild@localhost ~]$ PIG_CLASSPATH=/home/hadoop/HADOOP/hbase-0.98.4-hadoop2/lib/hbase-server-0.98.4-hadoop2:/home/hadoop/HADOOP/hbase-
0.98.4-hadoop2/lib/hbase-*.jar;
[acadgild@localhost ~]$ hbase shell
2017-11-24 03:20:49,523 INFO  [main] Configuration.deprecation: hadoop.native.lib is deprecated. Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 0.98.14-hadoop2, r4e4aabb93b52f1b0fef6b66edd06ec8923014dec, Tue Aug 25 22:35:44 PDT 2015

hbase(main):001:0> create 'StudentAcad','Student_data'
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBin
der.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2017-11-24 03:22:22,950 WARN  [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
0 row(s) in 3.7450 seconds

=> Hbase::Table - StudentAcad
hbase(main):002:0>
```

Activate Windows
Go to Settings to activate Windows.

- We can come out from HBase by typing exit and switch to PIG grunt shell.

- Once we are inside PIG mode we can load data from HDFS to Alias relation.

- we can transfer the data inside HBase by STORE command.

```
2017-11-24 03:49:48,499 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Failed!
grunt> rawD = LOAD '/home/acadgild/hadoop/student.txt' USING PigStorage (',') AS (StudentName:chararray,sector:chararray,DOB:chararray,qu
alification:chararray,score:int,state:chararray,randomName:chararray);
2017-11-24 03:50:29,947 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Inste
ad, use mapreduce.job.counters.max
2017-11-24 03:50:29,947 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use
 dfs.bytes-per-checksum
2017-11-24 03:50:29,947 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.de
faultFS
grunt>  STORE rawD INTO 'hbase://StudentAcad' USING org.apache.pig.backend.hadoop.hbase.HBaseStorage('Student_data:StudentName,Student_da
ta:sector,Student_data:DOB,Student_data:qualification,Student_data:score,Student_data:state,Student_data:randomName');
2017-11-24 03:50:37,903 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Inste
ad, use mapreduce.job.counters.max
2017-11-24 03:50:37,903 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use
 dfs.bytes-per-checksum
2017-11-24 03:50:37,903 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.de
faultFS
2017-11-24 03:50:38,121 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use
 dfs.bytes-per-checksum
2017-11-24 03:50:38,122 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Inste
```

- Once the success message comes as shown below , it is confirmed our data is loaded inside HBase

```
Input(s):
Successfully read 470 records from: "/home/acadgild/hadoop/student.txt"

Output(s):
Successfully stored 470 records in: "hbase://StudentAcad"

Counters:
Total records written : 470
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1722380335_0005


2017-11-24 03:50:41,560 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker
, sessionId= - already initialized
2017-11-24 03:50:41,562 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker
, sessionId= - already initialized
2017-11-24 03:50:41,562 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker
, sessionId= - already initialized
2017-11-24 03:50:41,579 [main] WARN  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning
 FIELD_DISCARDED_TYPE_CONVERSION_FAILED 1 time(s).
2017-11-24 03:50:41,579 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> scan 'StudentAcad'
```

- The result can be displayed through scan command followed by table name inside quotes(' ').

```
2. 127.0.0.1 (acadgild)          4. 127.0.0.1 (acadgild)

hbase(main):001:0> scan 'StudentAcad'
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBin
der.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2017-11-24 04:11:05,318 WARN  [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
ROW                            COLUMN+CELL
 ABEDNIGO                       column=Student_data:DOB, timestamp=1511475640362, value=BBA
 ABEDNIGO                       column=Student_data:StudentName, timestamp=1511475640362, value=goverement
 ABEDNIGO                       column=Student_data:qualification, timestamp=1511475640362, value=100
 ABEDNIGO                       column=Student_data:score, timestamp=1511475640362, value=alabama
 ABEDNIGO                       column=Student_data:sector, timestamp=1511475640362, value=20-10-2000
 ABEDNIGO                       column=Student_data:state, timestamp=1511475640362, value=madison`
 ABROSER                        column=Student_data:DOB, timestamp=1511475640195, value=MBBS
 ABROSER                        column=Student_data:StudentName, timestamp=1511475640195, value=goverement
 ABROSER                        column=Student_data:qualification, timestamp=1511475640195, value=3
 ABROSER                        column=Student_data:score, timestamp=1511475640195, value=Pennsylvania
 ABROSER                        column=Student_data:sector, timestamp=1511475640195, value=18-11-2002
 ABROSER                        column=Student_data:state, timestamp=1511475640195, value=prattville*
 AGNES                          column=Student_data:DOB, timestamp=1511475640362, value=BE
 AGNES                          column=Student_data:StudentName, timestamp=1511475640362, value=goverement
 AGNES                          column=Student_data:qualification, timestamp=1511475640362, value=100
 AGNES                          column=Student_data:score, timestamp=1511475640362, value=alabama
 AGNES                          column=Student_data:sector, timestamp=1511475640362, value=20-10-2000
 AGNES                          column=Student_data:state, timestamp=1511475640362, value=madison`
 AGNEW                          column=Student_data:DOB, timestamp=1511475640196, value=BCOM
 AGNEW                          column=Student_data:StudentName, timestamp=1511475640196, value=goverement
 AGNEW                          column=Student_data:qualification, timestamp=1511475640196, value=7
 AGNEW                          column=Student_data:score, timestamp=1511475640196, value=california
 AGNEW                          column=Student_data:sector, timestamp=1511475640196, value=20-10-2000
 AGNEW                          column=Student_data:state, timestamp=1511475640196, value=dothan@
 ALEXANDER                      column=Student_data:DOB, timestamp=1511475640363, value=MBBS
 ALEXANDER                      column=Student_data:StudentName, timestamp=1511475640363, value=goverement
 ALEXANDER                      column=Student_data:qualification, timestamp=1511475640363, value=100
 ALEXANDER                      column=Student_data:score, timestamp=1511475640363, value=alabama
 ALEXANDER                      column=Student_data:sector, timestamp=1511475640363, value=20-10-2000
```

Activate Windows
Go to Settings to activate Windows.

Submitted By:

Shishir Kumar Jha