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:

This document contains the step to execute a **Loading/Transferring of Data into HBase using Pig Scripts**. To achieve this Hadoop cluster is needed with Pig and HBase running on it.

In this example, **Hadoop version: hadoop-2.6.0, Hbase version: hbase-0.98.4-hadoop2-bin and Pig version: pig-0.14.0** is used.

The student.txt data set has below column names

Columns Names: StudentName, sector, DOB, qalification, score, state, randomName.

- The 'student.txt' data set will be copied in to HDFS which will be further loaded into HBase.

- Setting the Class Path for Pig

```
[acadgild@localhost ~]$ PIG_CLASSPATH=/usr/local/hbase/lib/hbase-server-0.98.14-
hadoop2:/usr/local/hbase/lib/hbase-*.jar;
```

- Starting HBase shell and create a table

```
hbase(main):005:0> create 'student','studentdoc'
0 row(s) in 0.8140 seconds

=> Hbase::Table - student
hbase(main):006:0>
```

- Coming out from HBase by typing exit and switch to PIG grunt shell. Loading data from HDFS to Alias relation.Transfering the data inside HBase by STORE command.

```
grunt> studemo = LOAD '/home/acadgild/Downloads/student.txt' USING PigStorage(',') AS(StudentName:chararray,sector:chararray,
DOB:chararray,qalification:chararray,score:int,state:chararray,randomName:chararray);
2017-11-17 17:23:22,725 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is depre
cated. Instead, use mapreduce.job.counters.max
2017-11-17 17:23:22,727 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated.
Instead, use dfs.bytes-per-checksum
2017-11-17 17:23:22,728 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
grunt> STORE studemo INTO 'hbase://student' USING org.apache.pig.backend.hadoop.hbase.HBaseStorage('studentdoc:StudentName,st
udentdoc:sector,studentdoc:DOB,studentdoc:qalification,studentdoc:score,studentdoc:state,studentdoc:randomName');
2017-11-17 17:23:37,360 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is depre
cated. Instead, use mapreduce.job.counters.max
2017-11-17 17:23:37,366 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated.
Instead, use dfs.bytes-per-checksum
2017-11-17 17:23:37,366 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
```

- Once the success message comes , it is confirmed data is loaded inside HBase.

```
Success!

Job Stats (time in seconds):
JobId       Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgRe
duceTime        MedianReducetime        Alias   Feature Outputs
job_local1068106496_0004        1       0       n/a     n/a     n/a     n/a     0       0       0       0       studemo MAP_O
NLY     hbase://student,

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

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

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_local1068106496_0004


2017-11-17 17:23:49,029 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam
e=JobTracker, sessionId= - already initialized
2017-11-17 17:23:49,047 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam
e=JobTracker, sessionId= - already initialized
2017-11-17 17:23:49,065 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam
e=JobTracker, sessionId= - already initialized
2017-11-17 17:23:49,198 [main] WARN  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encount
ered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 1 time(s).
2017-11-17 17:23:49,198 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success
!
grunt>                                          acadgild@localhost:~
```

- CHECKING THE DATA LOADED IN HBASE

```
hbase(main):002:0> count 'student'
230 row(s) in 0.5060 seconds

=> 230
hbase(main):003:0> scan 'student'
```

```
 WILLIAM             column=studentdoc:sector, timestamp=1510919625409, value=20-10-2000
 WILLIAM             column=studentdoc:state, timestamp=1510919625409, value=madison`
 WILSON              column=studentdoc:DOB, timestamp=1510919625283, value=BE
 WILSON              column=studentdoc:StudentName, timestamp=1510919625283, value=goverement
 WILSON              column=studentdoc:qalification, timestamp=1510919625283, value=100
 WILSON              column=studentdoc:score, timestamp=1510919625283, value=alabama
 WILSON              column=studentdoc:sector, timestamp=1510919625283, value=20-10-2000
 WILSON              column=studentdoc:state, timestamp=1510919625283, value=madison`
 WITMORE             column=studentdoc:DOB, timestamp=1510919624999, value=BE
 WITMORE             column=studentdoc:StudentName, timestamp=1510919624999, value=goverement
 WITMORE             column=studentdoc:qalification, timestamp=1510919624999, value=100
 WITMORE             column=studentdoc:score, timestamp=1510919624999, value=alabama
 WITMORE             column=studentdoc:sector, timestamp=1510919624999, value=20-10-2000
 WITMORE             column=studentdoc:state, timestamp=1510919624999, value=madison`
 WOODS               column=studentdoc:DOB, timestamp=1510919625290, value=MBBS
 WOODS               column=studentdoc:StudentName, timestamp=1510919625290, value=goverement
 WOODS               column=studentdoc:qalification, timestamp=1510919625290, value=100
 WOODS               column=studentdoc:score, timestamp=1510919625290, value=alabama
 WOODS               column=studentdoc:sector, timestamp=1510919625290, value=20-10-2000
 WOODS               column=studentdoc:state, timestamp=1510919625290, value=madison`
 WORK                column=studentdoc:DOB, timestamp=1510919625009, value=BE
 WORK                column=studentdoc:StudentName, timestamp=1510919625009, value=goverement
 WORK                column=studentdoc:qalification, timestamp=1510919625009, value=100
 WORK                column=studentdoc:score, timestamp=1510919625009, value=alabama
 WORK                column=studentdoc:sector, timestamp=1510919625009, value=28-04-2005
 WORK                column=studentdoc:state, timestamp=1510919625009, value=madison`
 YOUNG               column=studentdoc:DOB, timestamp=1510919625012, value=MS
 YOUNG               column=studentdoc:StudentName, timestamp=1510919625012, value=goverement
 YOUNG               column=studentdoc:qalification, timestamp=1510919625012, value=9
 YOUNG               column=studentdoc:score, timestamp=1510919625012, value=alaska
 YOUNG               column=studentdoc:sector, timestamp=1510919625012, value=20-10-2000
 YOUNG               column=studentdoc:state, timestamp=1510919625012, value=auburn~
230 row(s) in 6.2130 seconds

hbase(main):004:0>                                                                          25
    acadgild@localhost:~      acadgild@localhost:~      acadgild@localhost:~
```