Assignment 10.2

1. Uploaded students.txt using Hadoop

```
[acadgild@localhost sumona-hadoop]$ hadoop fs -put student.txt /user/acadgild/Assignment10
17/11/28 15:44:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes
 where applicable
[acadgild@localhost sumona-hadoop]$ hadoop fs -ls /user/acadgild/Assignment10
17/11/28 15:45:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes
 where applicable
Found 1 items
-rw-r--r--   1 acadgild supergroup       26204 2017-11-28 15:44 /user/acadgild/Assignment10/student.txt
[acadgild@localhost sumona-hadoop]$
```

2. Included few jar files of HBase to the Pig classpath.

```
[acadgild@localhost ~]$ PIG_CLASSPATH=/home/acadgild/sumona-hadoop/hbase-0.98.4-hadoop2/lib/hbase-server-0.98.4-hadoop2:/home/acad
gild/sumona-hadoop/hbase-0.98.4-hadoop2/lib/hbase-*.jar;
[acadgild@localhost ~]$
```

3. Started the HBase shell and created a table

```
[acadgild@localhost ~]$ start-hbase.sh
starting master, logging to /usr/local/hbase/logs/hbase-acadgild-master-localhost.localdomain.out
[acadgild@localhost ~]$ hbase shell
2017-11-28 16:04:31,727 INFO  [main] Configuration.deprecation: hadoop.native.lib is deprecated. Instead, use io.native.lib.availa
ble
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/StaticLo
ggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2017-11-28 16:06:06,216 WARN  [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builti
n-java classes where applicable
0 row(s) in 3.5550 seconds

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

4. Load the data from HDFS to Alias relations.

```
grunt> rawD = LOAD '/user/acadgild/Assignment10/student.txt' USING PigStorage()
>> AS (StudentName:chararray,sector:chararray,DOB:chararray,qalification:chararray,score:int,state:chararray, randomName:chararray
);
2017-11-28 16:21:13,838 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker.persist.jobstatus.hours
 is deprecated. Instead, use mapreduce.jobtracker.persist.jobstatus.hours
2017-11-28 16:21:13,838 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.heartbeats.in.second is deprecated.
 Instead, use mapreduce.jobtracker.heartbeats.in.second
2017-11-28 16:21:13,838 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - jobclient.completion.poll.interval is depr
ecated. Instead, use mapreduce.client.completion.pollinterval
2017-11-28 16:21:13,839 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.tasktracker.tasks.sleeptime-before-
sigkill is deprecated. Instead, use mapreduce.tasktracker.tasks.sleeptimebeforesigkill
2017-11-28 16:21:13,839 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker.http.address is depreca
ted. Instead, use mapreduce.jobtracker.http.address
2017-11-28 16:21:13,839 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - dfs.replication.considerLoad is deprecated
. Instead, use dfs.namenode.replication.considerLoad
2017-11-28 16:21:13,839 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - dfs.block.size is deprecated. Instead, use
 dfs.blocksize
2017-11-28 16:21:13,839 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - dfs.permissions is deprecated. Instead, us
e dfs.permissions.enabled
2017-11-28 16:21:13,839 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - topology.node.switch.mapping.impl is depre
cated. Instead, use net.topology.node.switch.mapping.impl
2017-11-28 16:21:13,839 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - dfs.access.time.precision is deprecated. I
nstead, use dfs.namenode.accesstime.precision
```

5. Now we can transfer the data inside HBase by STORE command.

We need to ensure that we give the correct name for table name created inside HBase. Also the parameters should be kept in mind to avoid mistake.

```
grunt> STORE rawD INTO 'hbase://studentAcad' USING org.apache.pig.backend.hadoop.hbase.HBaseStorage('student_data:StudentName,stud
ent_data:sector,student_data:DOB,student_data:qalification,student_data:score,student_data:state,student_data:randomName');
2017-11-28 16:31:52,622 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated
. Instead, use mapreduce.job.counters.max
2017-11-28 16:31:52,622 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Inste
ad, use dfs.bytes-per-checksum
2017-11-28 16:31:52,622 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, us
e fs.defaultFS
2017-11-28 16:31:52,898 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Inste
ad, use dfs.bytes-per-checksum
```

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

```
HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.2.0   0.14.0  acadgild        2017-11-28 16:40:31     2017-11-28 16:42:36     UNKNOWN

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceT
ime     MedianReducetime        Alias   Feature Outputs
job_1511863758374_0003  1       0       n/a     n/a     n/a     n/a     0       0       0       0       rawD    MAP_ONLY        hb
ase://studentAcad,

Input(s):
Successfully read 0 records from: "/user/acadgild/Assignment10/student.txt"

Output(s):
Successfully stored 0 records in: "hbase://studentAcad"

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

The result can be displayed through scan command.

```
hbase(main):004:0> scan 'studentAcad'
ROW                             COLUMN+CELL
 ABEDNIGO,goverenment,20-10-2000, column=student_data:DOB, timestamp=1511867458220, value=
 BBA,100,alabama,madison`
 ABEDNIGO,goverenment,20-10-2000, column=student_data:StudentName, timestamp=1511867458220, value=
 BBA,100,alabama,madison`
 ABEDNIGO,goverenment,20-10-2000, column=student_data:qalification, timestamp=1511867458220, value=
 BBA,100,alabama,madison`
 ABEDNIGO,goverenment,20-10-2000, column=student_data:score, timestamp=1511867458220, value=
 BBA,100,alabama,madison`
 ABEDNIGO,goverenment,20-10-2000, column=student_data:sector, timestamp=1511867458220, value=
 BBA,100,alabama,madison`
 ABEDNIGO,goverenment,20-10-2000, column=student_data:state, timestamp=1511867458220, value=
 BBA,100,alabama,madison`
 ABROSER,goverenment,18-11-2002,M column=student_data:DOB, timestamp=1511867458026, value=
 BBS,3.5,Pennsylvania,prattville*
 ABROSER,goverenment,18-11-2002,M column=student_data:StudentName, timestamp=1511867458026, value=
 BBS,3.5,Pennsylvania,prattville*
 ABROSER,goverenment,18-11-2002,M column=student_data:qalification, timestamp=1511867458026, value=
 BBS,3.5,Pennsylvania,prattville*
 ABROSER,goverenment,18-11-2002,M column=student_data:score, timestamp=1511867458026, value=
 BBS,3.5,Pennsylvania,prattville*
 ABROSER,goverenment,18-11-2002,M column=student_data:sector, timestamp=1511867458026, value=
 BBS,3.5,Pennsylvania,prattville*
 ABROSER,goverenment,18-11-2002,M column=student_data:state, timestamp=1511867458026, value=
 BBS,3.5,Pennsylvania,prattville*
 AGNES,goverenment,20-10-2000,BE, column=student_data:DOB, timestamp=1511867458221, value=
 100,alabama,madison`
 AGNES,goverenment,20-10-2000,BE, column=student_data:StudentName, timestamp=1511867458221, value=
 100,alabama,madison`
 AGNES,goverenment,20-10-2000,BE, column=student_data:qalification, timestamp=1511867458221, value=
 100,alabama,madison`
 AGNES,goverenment,20-10-2000,BE, column=student_data:score, timestamp=1511867458221, value=
 100,alabama,madison`
 AGNES,goverenment,20-10-2000,BE, column=student_data:sector, timestamp=1511867458221, value=
 100,alabama,madison`
 AGNES,goverenment,20-10-2000,BE, column=student_data:state, timestamp=1511867458221, value=
 100,alabama,madison`
 AGNEW,goverenment,20-10-2000,BCO column=student_data:DOB, timestamp=1511867458030, value=
```