Assignment 10.2

1. **Firstly a given student Data set is loaded into HDFS as below.**
2. **Then HBase jar files are added to the PIG_CLASSPATH as mentioned below.**

```
File  Edit  View  Search  Terminal  Help
GEORGE,goverenment,20-10-2000,BE,100,alabama,madison`
GEORGE,goverenment,18-11-2002,BE,8.5,arizona,decatur!
GEORGE,goverenment,12-12-2001,BTECH,100,alabama,madison`
GEORGE,goverenment,01-01-2003,BE,100,alabama,madison`
GEORGE,goverenment,20-10-2000,BE,100,alabama,madison`
GEORGE,goverenment,20-10-2000,BE,100,alabama,madison`
GEORGE,goverenment,20-10-2000,BE,100,alabama,madison`
FREDERICK,goverenment,20-10-2000,BE,100,alabama,madison`
RICHARD,goverenment,20-10-2000,BE,4.5,oregon,huntsville/
[acadgild@localhost ~]$ hadoop dfs -put AcadStudent /
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

18/02/28 17:28:37 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ pwd
/home/acadgild
[acadgild@localhost ~]$ ls
AcadStudent  Documents  eclipse-workspace  Pictures    Templates
db.java      Downloads  install            Public      Videos
Desktop      eclipse    Music                          student.txt
[acadgild@localhost ~]$ PIG_CLASSPATH=/home/acadgild/install/hbase/hbase-1.2.6/l
ib/hbase-*.jar;
[acadgild@localhost ~]$ 
```

3. **Now the Data set is loaded as below snapshot**

```
grunt> rawD = load 'AcadStudent' using PigStorage(',') as (StudentName:chararray
,sector:chararray,DOB:chararray,qualification:chararray,score:int,state:chararra
y,randomName:chararray);
```

4. Now a HBase table is created in HBase as mentioned below

```
hbase(main):001:0> create 'studentAcad','student data'
0 row(s) in 5.6710 seconds

=> Hbase::Table - studentAcad
hbase(main):002:0> list
TABLE
a
ab
clicks
cust
customer
studentAcad
6 row(s) in 0.2820 seconds

=> ["a", "ab", "clicks", "cust", "customer", "studentAcad"]
hbase(main):003:0>
```

5. Now rawD is stored in HBase in grunt shell as below

```
nc.mapReduceLayer.MapReduceLauncher - Success:
grunt> STORE rawD into 'hbase://studentAcad' USING org.apache.pig.backend.hadoop
.hbase.HBaseStorage('student_data:StudentName,student_data:sector,student_data:D
OB,student_data:qualification,student_data:score,student_data:state,student_data
:randomName');
```

```
Total records proactively spilled: 0

Job DAG:
job_local336699627_0008


2018-02-28 18:33:53,914 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics -
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already
initialized
2018-02-28 18:33:53,938 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics -
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already
initialized
2018-02-28 18:33:53,942 [main] INFO  org.apache.hadoop.metrics.jvm.JvmMetrics -
Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already
initialized
2018-02-28 18:33:54,038 [main] WARN  org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_C
ONVERSION_FAILED 1 time(s).
2018-02-28 18:33:54,038 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
grunt> █
```

6. **Finally a dataset can be viewed in HBase using "scan 'studentAcad' " as below**

```
                           BS
    WOODS                  column=student_data:StudentName, timestamp=1519823029021,
                           value=goverenment
    WOODS                  column=student_data:qualification, timestamp=1519823029021
                           , value=100
    WOODS                  column=student_data:score, timestamp=1519823029021, value=
                           alabama
    WOODS                  column=student_data:sector, timestamp=1519823029021, value
                           =20-10-2000
    WOODS                  column=student_data:state, timestamp=1519823029021, value=
                           madison`
    WORK                   column=student_data:DOB, timestamp=1519823028839, value=BE
    WORK                   column=student_data:StudentName, timestamp=1519823028839,
                           value=goverenment
    WORK                   column=student_data:qualification, timestamp=1519823028839
                           , value=100
    WORK                   column=student_data:score, timestamp=1519823028839, value=
                           alabama
    WORK                   column=student_data:sector, timestamp=1519823028839, value
                           =28-04-2005
    WORK                   column=student_data:state, timestamp=1519823028839, value=
                           madison`
    YOUNG                  column=student_data:DOB, timestamp=1519823028841, value=MS
    YOUNG                  column=student_data:StudentName, timestamp=1519823028841,
                           value=goverenment
    YOUNG                  column=student_data:qualification, timestamp=1519823028841
                           , value=9
    YOUNG                  column=student_data:score, timestamp=1519823028841, value=
                           alaska
    YOUNG                  column=student_data:sector, timestamp=1519823028841, value
                           =20-10-2000
    YOUNG                  column=student_data:state, timestamp=1519823028841, value=
                           auburn~
230 row(s) in 10.6940 seconds

hbase(main):004:0>
```