Assignment 10.2

**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/**](https://acadgild.com/blog/loading-data-into-hbase-using-pig-scripts/)

To implement the concepts discussed, user is expected to have a Hadoop cluster with Pig and HBase running on it.

**Note:**You need to download the following versions of Hadoop, HBase and Pig to implement the steps discussed to load the data into HBase using Pig.

* [**Hadoop version: hadoop-2.6.0**](https://drive.google.com/open?id=0B1QaXx7tpw3SQUw5QkpYNTN2UGc)
* [**Hbase version: hbase-0.98.4-hadoop2-bin**](https://drive.google.com/open?id=0B1QaXx7tpw3SWjE3amtoVng5T3c)
* [**Pig version: pig-0.14.0**](https://drive.google.com/open?id=0B1QaXx7tpw3SdVlJYjBab0U0djA)

DataSet <https://drive.google.com/file/d/0B1QaXx7tpw3SVUlWTUQyNTMzdG8/view>

SampleData(AcadgildStudent)

StudentName,sector,DOB,qalification,score,state,randomName

ABROSER,goverenment,18-11-2002,MBBS,3.5,Pennsylvania,prattville\*

ALEXANDER,goverenment,20-10-2000,BSC,2.5,vermont,gadsden+

ALEXANDER,private,20-10-2000,BE,8.5,arizona,decatur!

ALEXANDER,goverenment,01-01-2003,BTECH,4.5,oregon,huntsville/

AGNEW,goverenment,20-10-2000,BCOM,7.5,california,dothan@

ATNEST,goverenment,20-10-2000,MTECH,8.5,arizona,decatur!

BELL,goverenment,10-07-2004,BBA,9.5,alaska,auburn~

BURR,goverenment,12-12-2001,BE,100,alabama,madison`

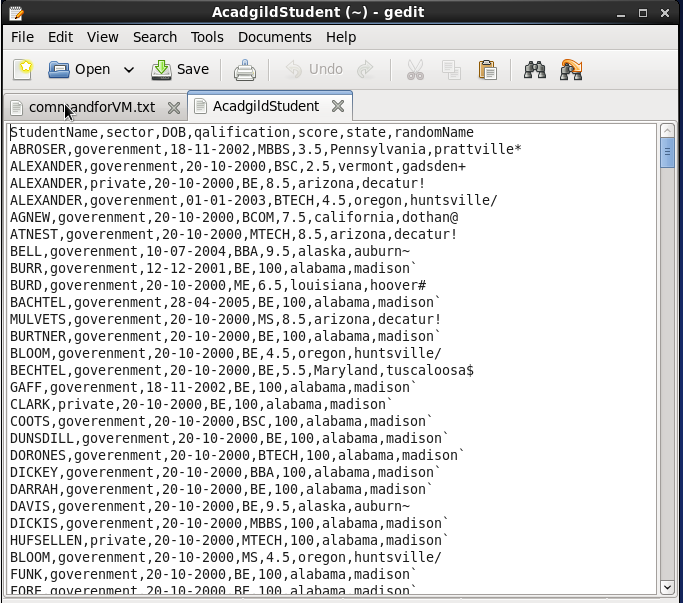
BURD,goverenment,20-10-2000,ME,6.5,louisiana,hoover#

BACHTEL,goverenment,28-04-2005,BE,100,alabama,madison`

MULVETS,goverenment,20-10-2000,MS,8.5,arizona,decatur!

BURTNER,goverenment,20-10-2000,BE,100,alabama,madison`

BLOOM,goverenment,20-10-2000,BE,4.5,oregon,huntsville/

****

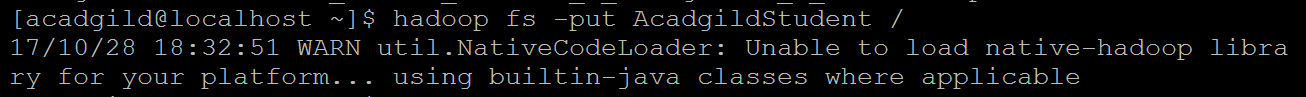
the description for the above data set containing seven columns named as:

StudentName, sector, DOB, qualification, score, state, randomName.

**Transfer into hdfs**

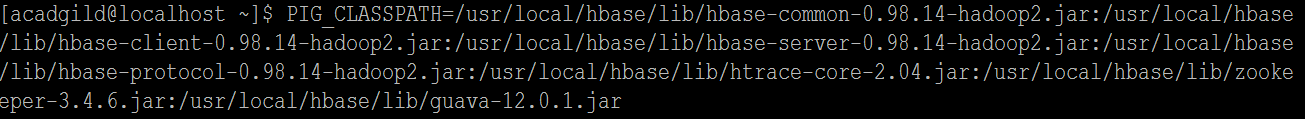
We will be copying the data set in to HDFS which will be further loaded into HBase.

hadoop fs -put AcadgildStudent /



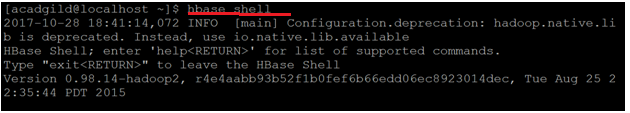
We will be including few jar files of HBase to the Pig classpath.

PIG\_CLASSPATH=/usr/local/hbase/lib/hbase-common-0.98.14-hadoop2.jar:/usr/local/hbase/lib/hbase-client-0.98.14-hadoop2.jar:/usr/local/hbase/lib/hbase-server-0.98.14-hadoop2.jar:/usr/local/hbase/lib/hbase-protocol-0.98.14-hadoop2.jar:/usr/local/hbase/lib/htrace-core-2.04.jar:/usr/local/hbase/lib/zookeeper-3.4.6.jar:/usr/local/hbase/lib/guava-12.0.1.jar

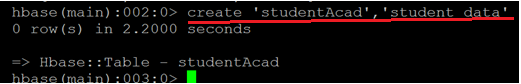


We will now start HBase shell and create a table. We only need this table as skeleton so PIG can Store data inside this by referring the table name.

hbase shell

****

create 'studentAcad','student\_data'

****

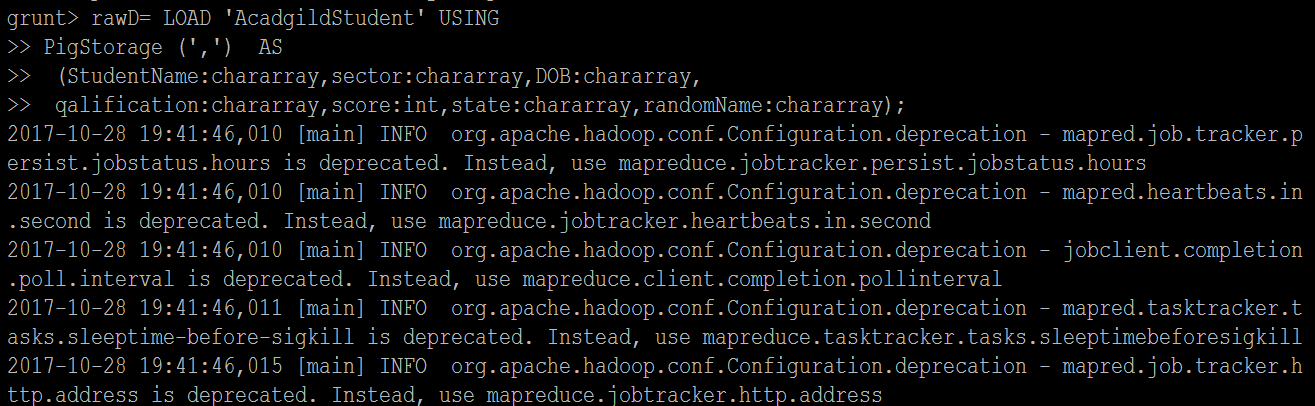
We can come out from HBase by typing exit and switch to PIG grunt shell.Once we are inside PIG mode we can load data from HDFS to Alias relation.

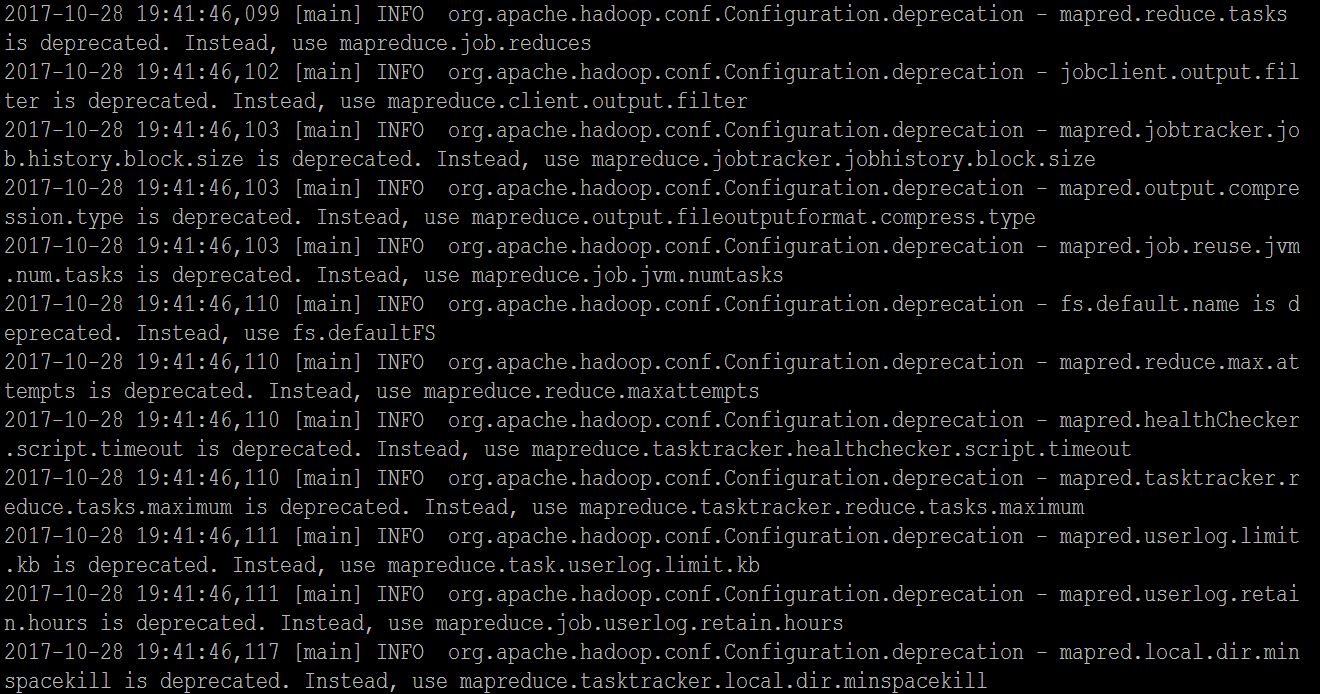
rawD= LOAD 'AcadgildStudent' USING

PigStorage (',') AS

(StudentName:chararray,sector:chararray,DOB:chararray,

qalification:chararray,score:int,state:chararray,randomName:chararray);





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.

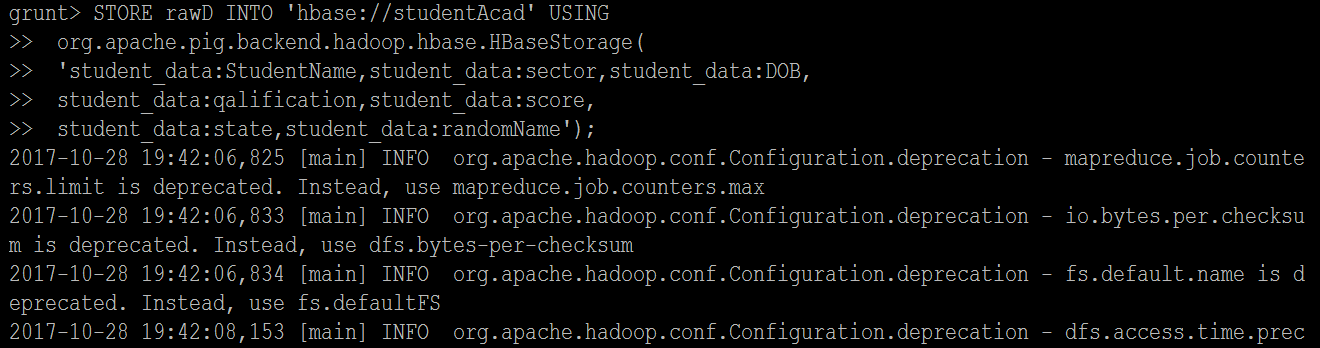
STORE rawD INTO 'hbase://studentAcad' USING

org.apache.pig.backend.hadoop.hbase.HBaseStorage(

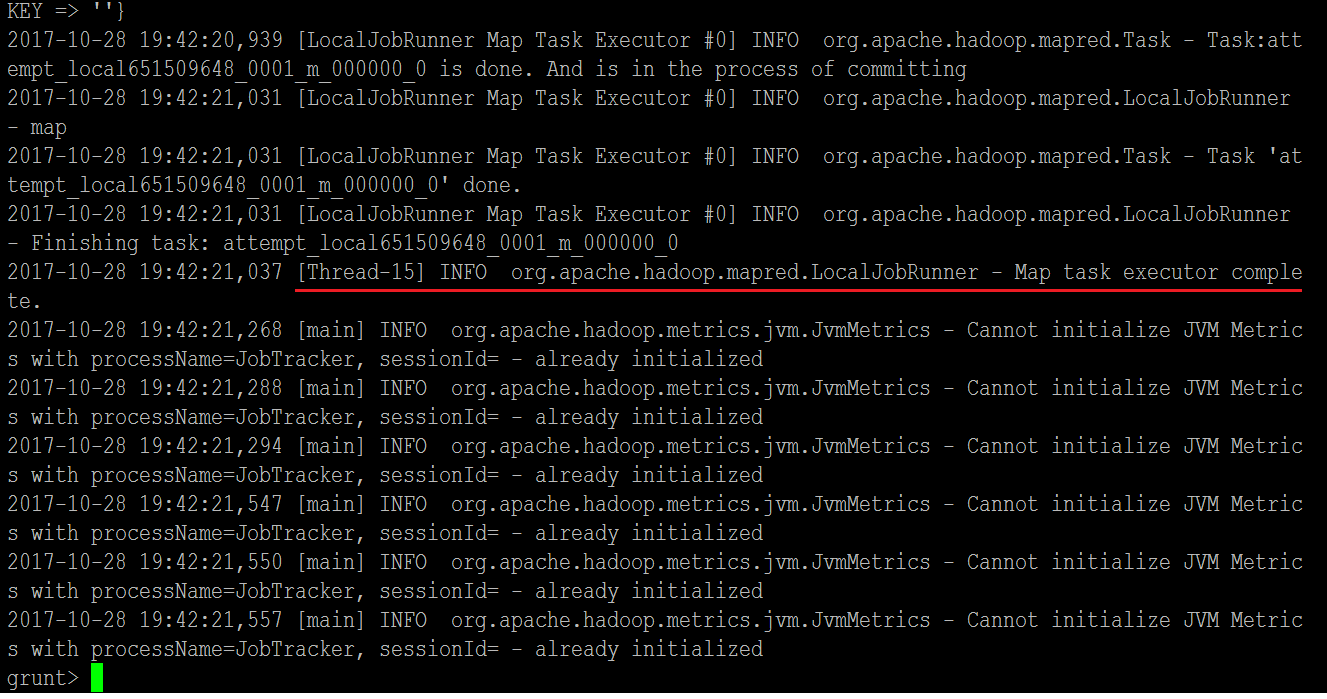
'student\_data:StudentName,student\_data:sector,student\_data:DOB,

student\_data:qalification,student\_data:score,

student\_data:state,student\_data:randomName');

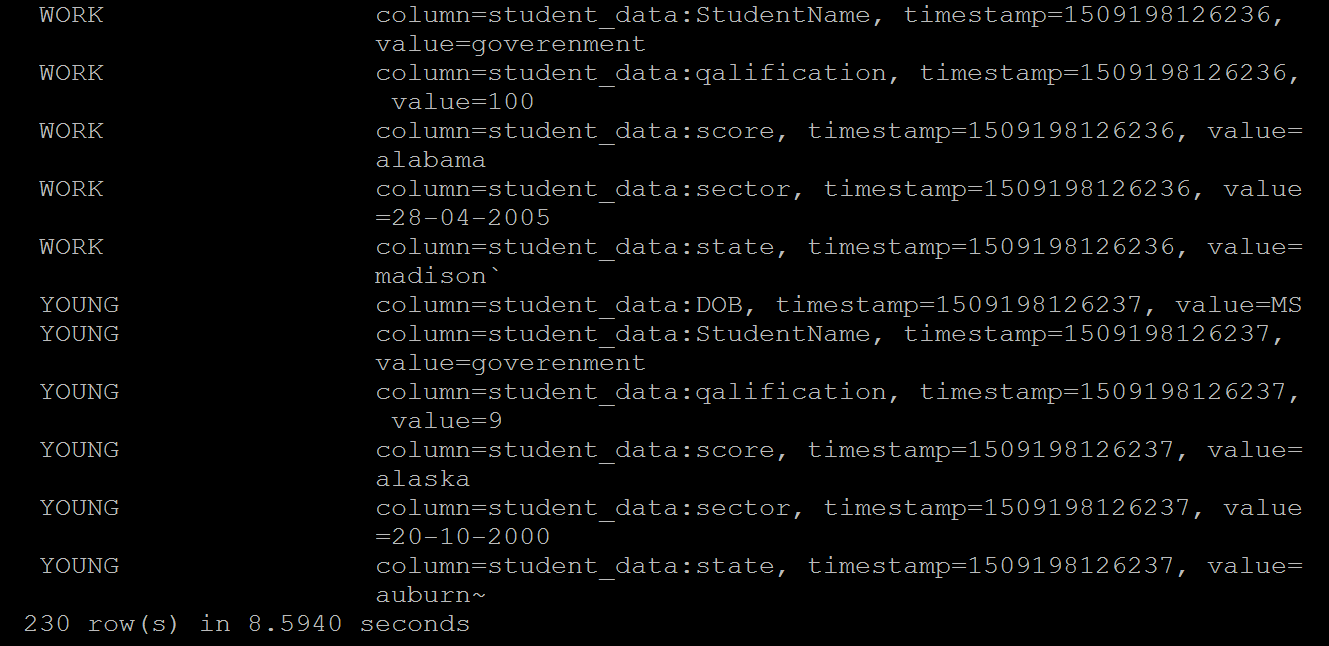


Once the success message comes as shown below , it is confirmed our data is loaded inside HBase.



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

scan 'studentAcad'

****