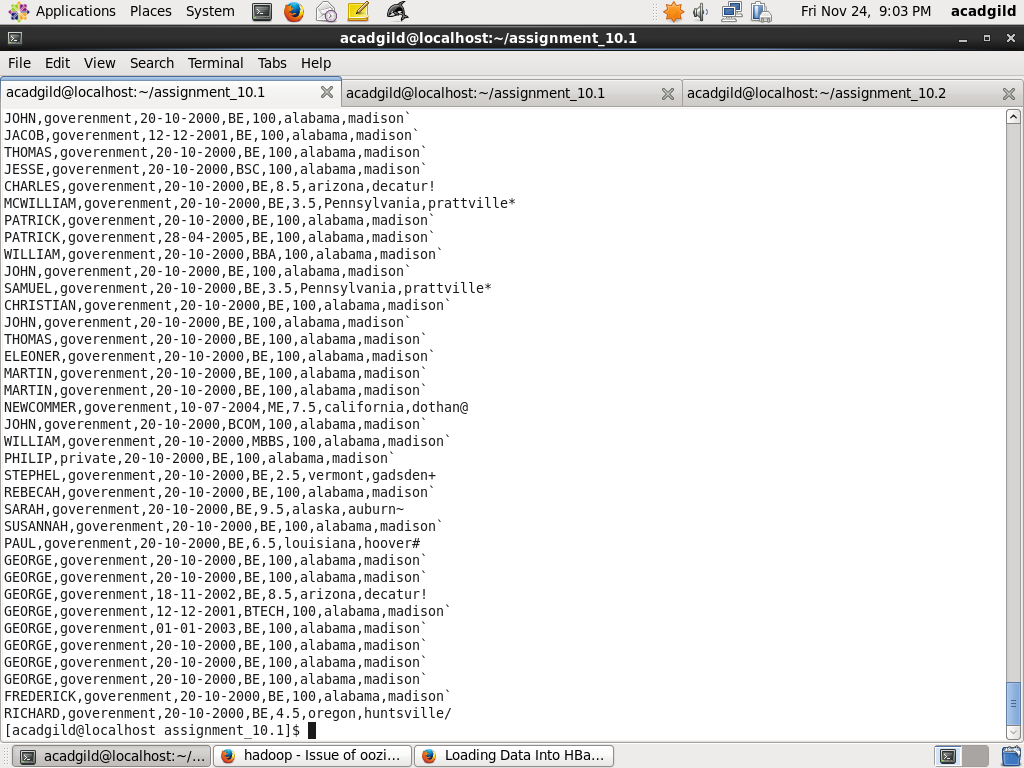
Assignment 10.2 ( Oozie and Sqoop)

Step1: Download dataset student.txt to /home/acadgild/assignment\_10.2 folder

Download the dataset to to /home/acadgild/assignment\_10.2 folder. Contents of dataset are displayed using cat student.txt

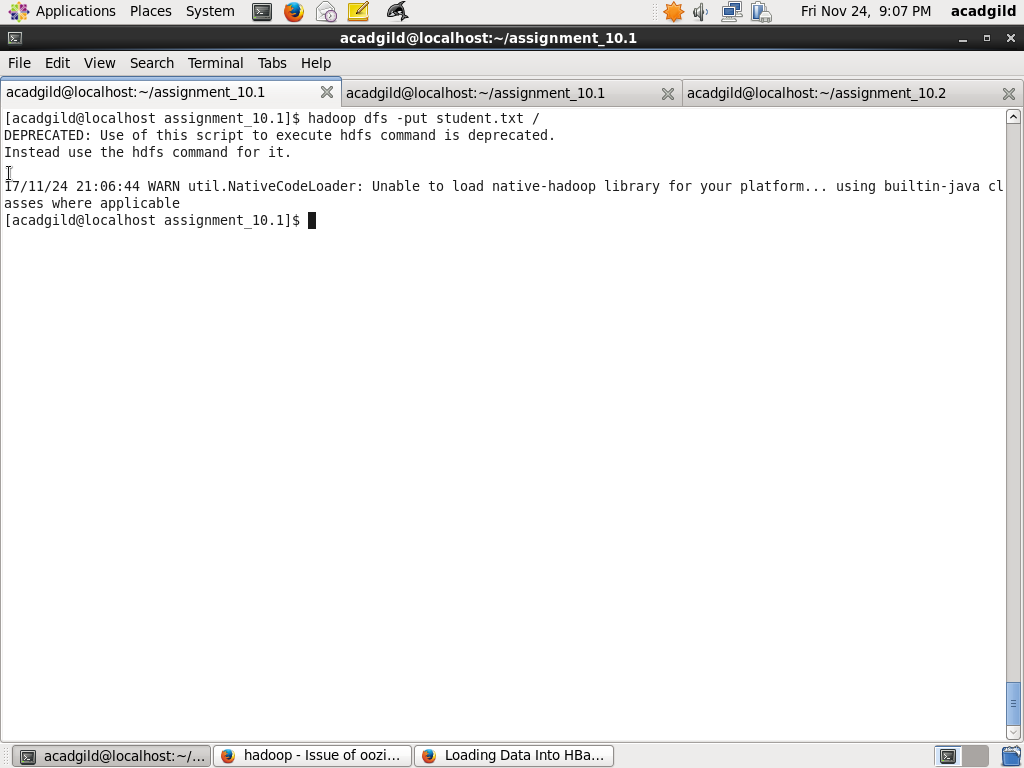
Screenshot is as below:



Step2: Copy dataset into HDFS

Load dataset to HDFS using command below

Hadoop dfs –put student.txt /

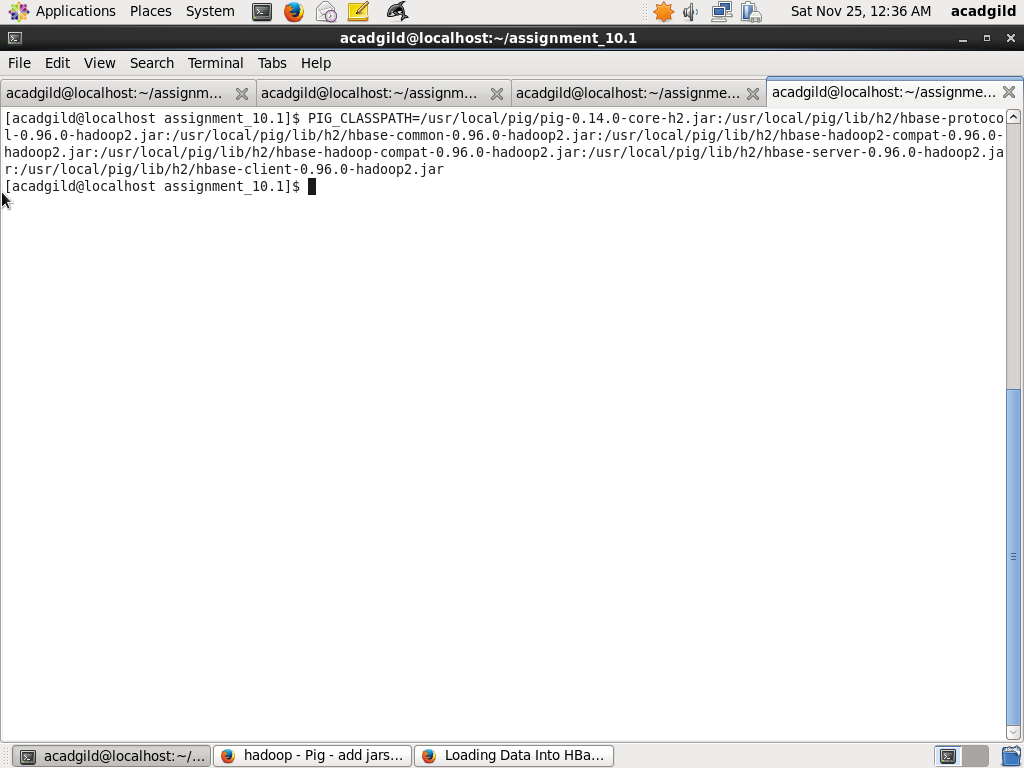


Step3 : Setup PIG CLASSPATH

Setup PIG CLASSPATH using:

PIG\_CLASSPATH=/usr/local/pig/pig-0.14.0-core-h2.jar:/usr/local/pig/lib/h2/hbase-protocol-0.96.0-hadoop2.jar:/usr/local/pig/lib/h2/hbase-common-0.96.0-hadoop2.jar:/usr/local/pig/lib/h2/hbase-hadoop2-compat-0.96.0-hadoop2.jar:/usr/local/pig/lib/h2/hbase-hadoop-compat-0.96.0-hadoop2.jar:/usr/local/pig/lib/h2/hbase-server-0.96.0-hadoop2.jar:/usr/local/pig/lib/h2/hbase-client-0.96.0-hadoop2.jar

Screenshot is as below:



Step4: Create HBASE table

First start HBASE and go to the HABASE Shell using the commands below:

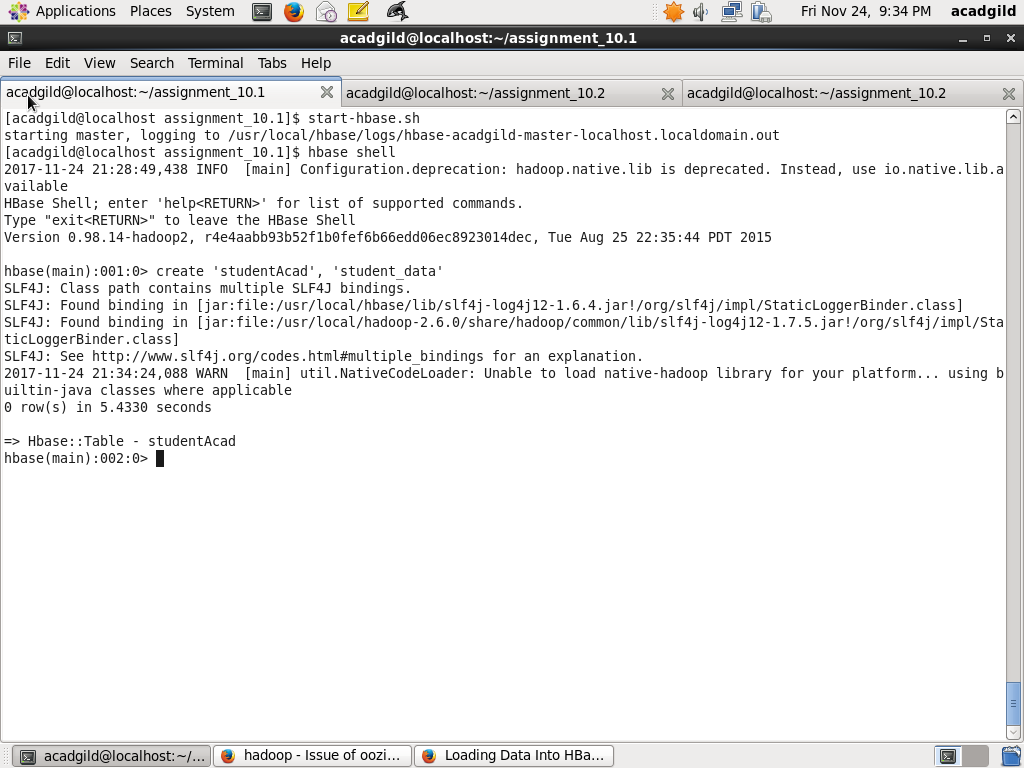
start-hbase.sh

hbase shell

Next create table studentAcad with column family student\_data using the commands below:

create 'studentAcad', 'student\_data'

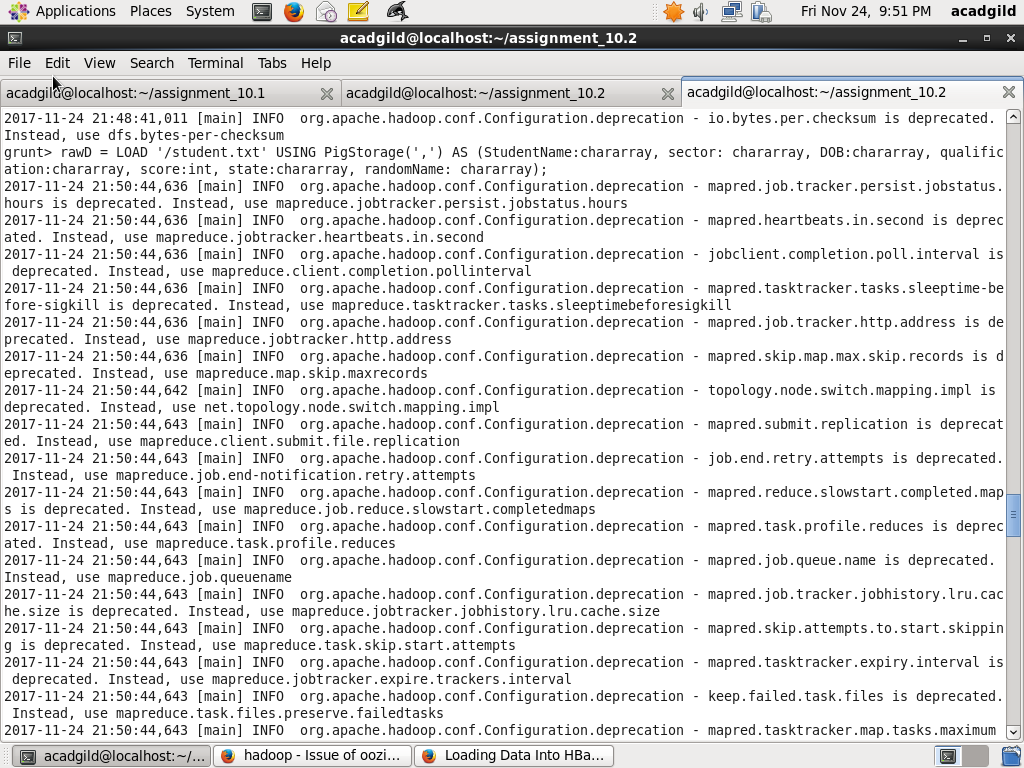
Screenshot is as below:



Step5: Load dataset to PIG

Start pig and Load dataset from HDFS into the PIG having schema (StudentName, sector, DOB,qualification, score, state, randomName) using the command below:

rawD = LOAD '/student.txt' USING PigStorage(',') AS (StudentName:chararray, sector: chararray, DOB:chararray, qualification:chararray, score:int, state:chararray, randomName: chararray);

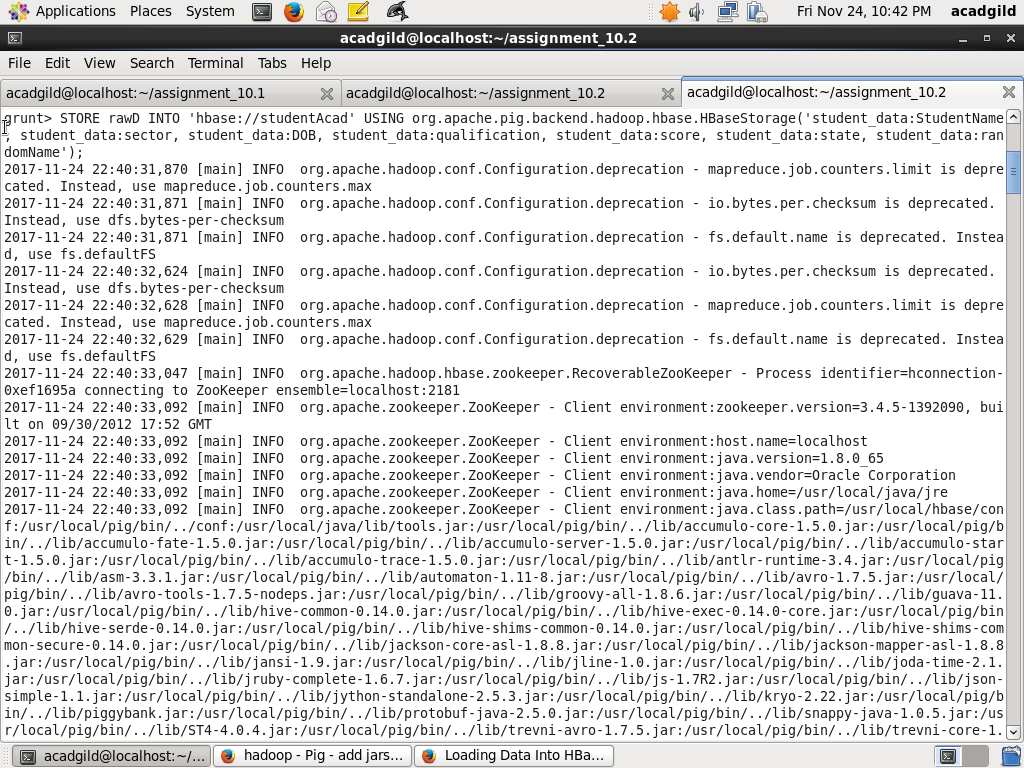


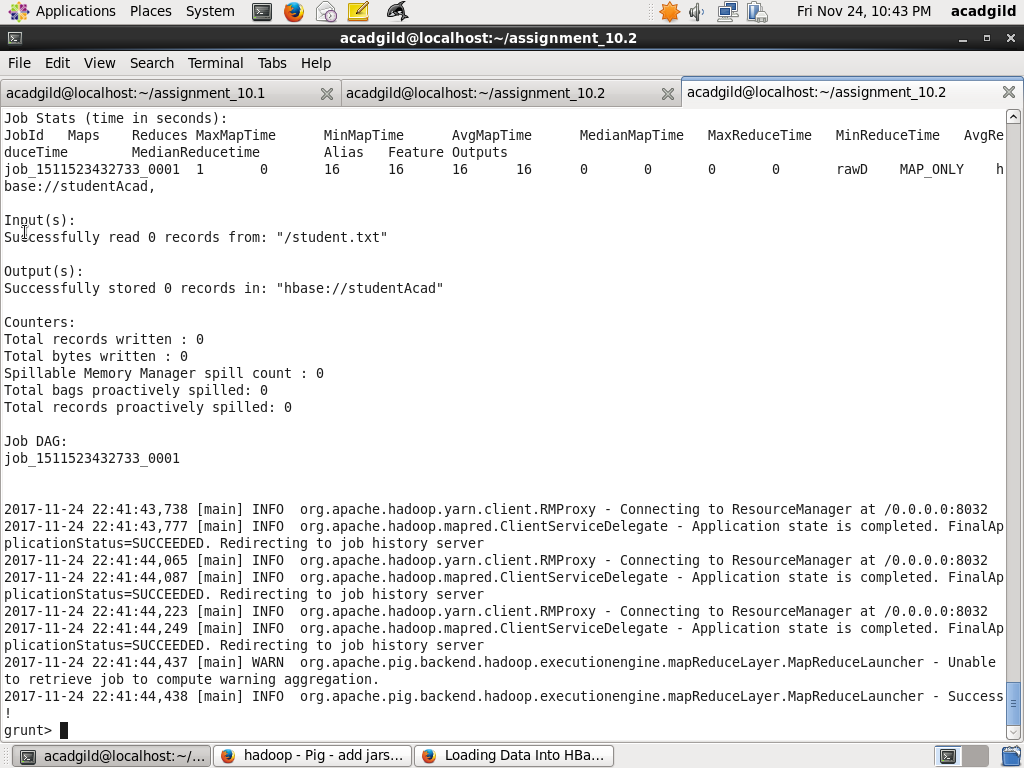
Step6: Transfer data to HBase

Use store command from PIG to HBase using command below:

STORE rawD INTO 'hbase://studentAcad' USING org.apache.pig.backend.hadoop.hbase.HBaseStorage('student\_data:StudentName, student\_data:sector, student\_data:DOB, student\_data:qualification, student\_data:score, student\_data:state, student\_data:randomName');

Screenshot is as below:





Step7: Display records in HBase

Use scan command below to display records in HBase

Screenshots are as below:

