Loading Data Into HBase Using PIG Scripts.

**Dataset specification:**

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

**Sample Data:**

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!

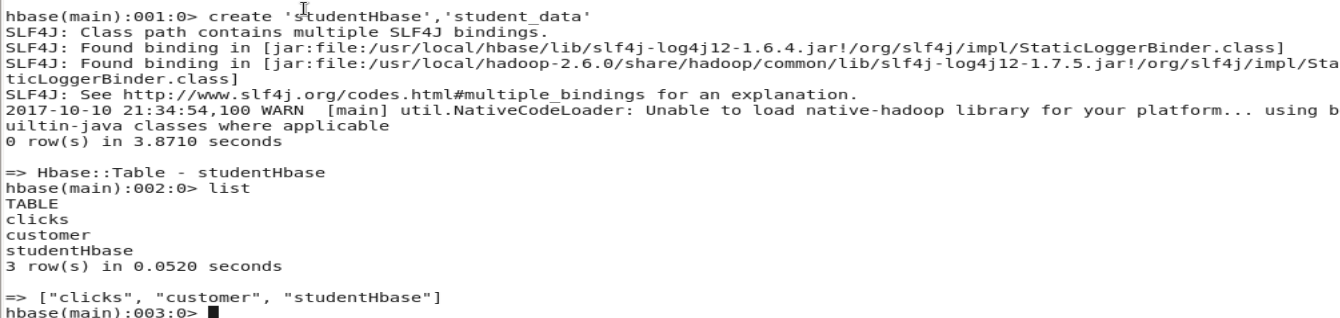
ALEXANDER,goverenment,01-01-2003,BTECH,4.5,oregon,huntsville/

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

1. **First create table in Hbase as per data specification:**

create 'studentHbase','student\_data'

--This will create a table named 'studentHbase’ and column family as 'student\_data'



1. **REGISTER all required hbase jars in pig :**

REGISTER /usr/lib/hbase/lib/hbase-common-0.96.0.2.0.6.0-76-hadoop2.jar

REGISTER /usr/lib/hbase/lib/hbase-client-0.96.0.2.0.6.0-76-hadoop2.jar

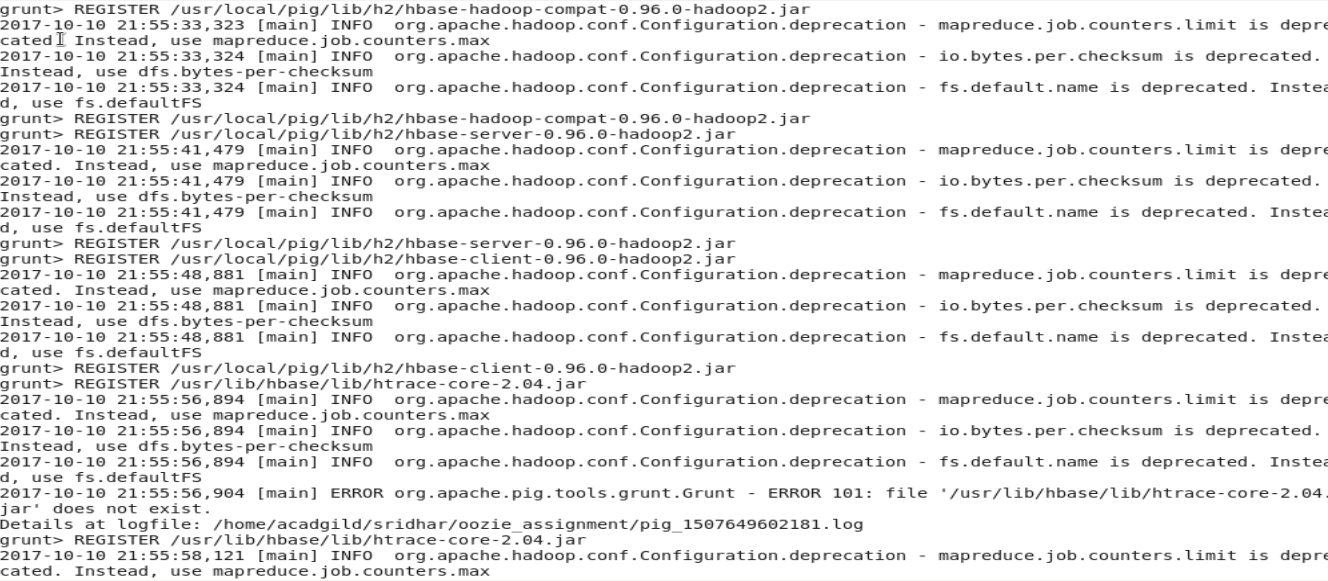
REGISTER /usr/lib/hbase/lib/hbase-server-0.96.0.2.0.6.0-76-hadoop2.jar

REGISTER /usr/lib/hbase/lib/hbase-protocol-0.96.0.2.0.6.0-76-hadoop2.jar

REGISTER /usr/lib/hbase/lib/htrace-core-2.01.jar

REGISTER /usr/lib/hbase/lib/zookeeper.jar

REGISTER /usr/lib/hbase/lib/guava-12.0.1.jar



1. **In Pig, pig –x local (local mode)**

**Load data into relation using PigStorage , Here I am loading the file from local file system,**

student\_data = load '/home/acadgild/sridhar/oozie\_assignment/student.txt' using PigStorage(',') AS(

StudentName:chararray,

sector:chararray,

DOB:chararray,

qualification:chararray,

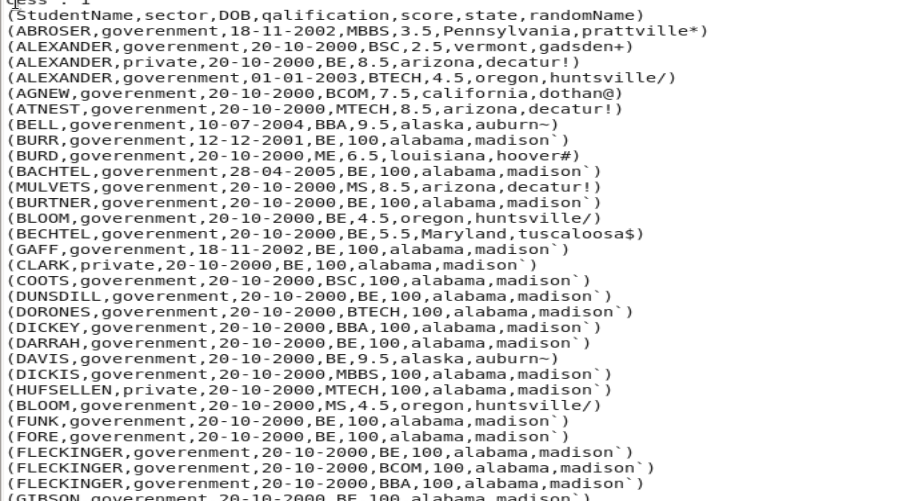
score:chararray,

state:chararray,

randomName:chararray

);

**Data got loaded into relation student\_Data:**



1. **Store Data from pig to Hbase table using HBaseStorage load function here ‘studentHbase' is the hbase table name**

STORE student\_data INTO

'hbase://studentHbase'

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

'student\_data:StudentName,

student\_data:sector,

student\_data:DOB,

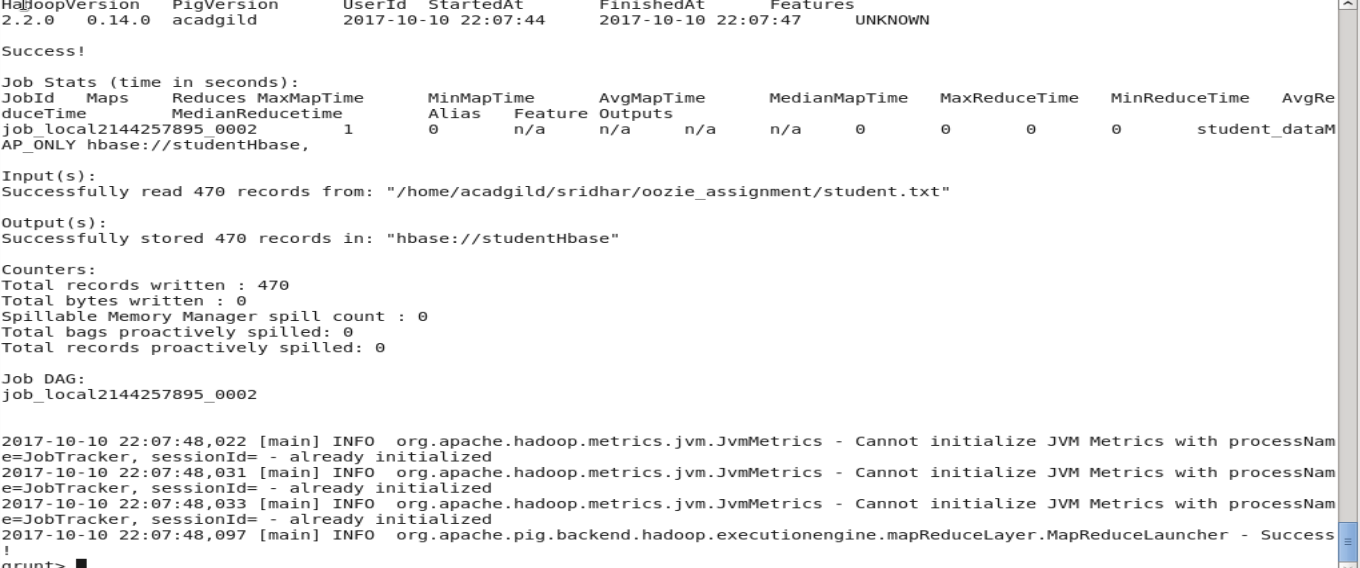
student\_data:qualification,

student\_data:score,

student\_data:state,

student\_data:randomName'

);



**5) scan 'studentHbase'**

**--To display data from habse table**