Session 10: Oozie and Sqoop

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/>

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



Step1:

Copying the data set into HDFS,

**hadoop dfs -put student.txt /user/acadgild/hadoop**



**Step2:**

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

**PIG\_CLASSPATH=/home/acadgild/hbase-1.0.3/lib/hbase-server-1.0.3-hadoop2:/lib/hbase-server-1.0.3-hadoop2:/home/acadgild/hbase-1.0.3-hadoop2:/lib/hbase-\*.jar;**



**Step3:**

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.

**create 'studentAcad','student data'**



**Step4:**

Once we are inside PIG mode we can load data from HDFS to Alias relation.

**rawD = LOAD '/home/acadgild/hadoop/student.txt' USING PigStorage(',') AS (StudentName:chararray, sector:chararray, DOB:chararray, qualification:chararray,**



**DUMP rawD;**



**Step5:**

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:qualification,student data:score,student data:state,student data:randomName');**

