Session 10: Oozie and Sqoop

Assignment 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/>

Solution:-

**Loading Data Into HBase Using PIG Scripts.**

We are taking sample data set of student which will be loaded into HBase. We have attached snapshot with every step for better understanding.

You can download this sample data set for your own practice from the below link.

[DATASET](https://drive.google.com/open?id=0B1QaXx7tpw3SVUlWTUQyNTMzdG8)



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

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





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



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

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.



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.

**rawD2 = LOAD '/Assignment10\_2/student.txt' using PigStorage (',') AS (StudentName:chararray, sector:chararray, DOB:chararray, qualification: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 rawD2 INTO 'hbase://studentAcadgild' 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');**

****

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



