Session 10: Oozie and Sqoop

Assignment 2 Question

Problem Statement

Implemented the steps given on blogs. Followed the given link: -

https://acadgild.com/blog/loading-data-into-hbase-using-pig-scripts/

3. Output

[cloudera@quickstart ~]$ hadoop fs -put /home/cloudera/Downloads/student.txt

```
[acadgild@localhost ~]$ hadoop fs -ls /
17/10/26 20:00:19 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 19 items
-rw-r--r--   1 acadgild supergroup        416 2015-11-16 23:43 /Wc.txt
drwxr-xr-x   - acadgild supergroup          0 2017-10-14 11:57 /flume_import
drwxr-xr-x   - acadgild supergroup          0 2015-11-09 19:21 /hbasestorage
-rw-r--r--   1 acadgild supergroup         28 2015-11-17 01:49 /inp
drwxr-xr-x   - acadgild supergroup          0 2015-11-17 02:03 /out23
drwxr-xr-x   - acadgild supergroup          0 2015-11-17 01:47 /sample-mr
drwxr-xr-x   - acadgild supergroup          0 2017-10-24 17:19 /sqoop_sql
drwxr-xr-x   - acadgild supergroup          0 2015-11-05 13:46 /sqoopout
-rw-r--r--   1 acadgild supergroup      26204 2017-10-24 11:03 /student.txt
```

[acadgild@localhost ~]$ PIG_CLASSPATH=/usr/local/hbase/lib/hbase.*;

```
[acadgild@localhost ~]$ PIG_CLASSPATH=/usr/local/hbase/lib/hbase.*;
[acadgild@localhost ~]$ 
```

hbase(main):001:0> create 'studentAcad','student_data'

0 row(s) in 3.9250 seconds

=> Hbase::Table – studentAcad

```
hbase(main):001:0> create 'studentAcad','student_data'
0 row(s) in 3.9250 seconds

=> Hbase::Table - studentAcad
```

 grunt> rawD = Load '/student.txt' USING PigStorage(',') AS
(StudentName:chararray,sector:chararray,DOB:chararray,qualification:chararray,score:i
nt,State:chararray,randomName:chararray);

grunt> dump rawD;

```
grunt> rawD = Load '/student.txt' USING PigStorage(',') AS (StudentName:chararray,sector:chararray,DOB:chararray,qualification:chararra
y,score:int,State:chararray,randomName:chararray);
grunt> dump rawD;
```

Note:- As suggested by Prateek some issue with this version unable to get output

```
Job DAG:
job_1508819263566_0001


2017-10-23 23:08:03,253 [main] WARN  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warni
ng FIELD_DISCARDED_TYPE_CONVERSION_FAILED 1 time(s).
2017-10-23 23:08:03,253 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-10-23 23:08:03,266 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.
defaultFS
2017-10-23 23:08:03,266 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use
mapreduce.jobtracker.address
2017-10-23 23:08:03,267 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate co
de.
2017-10-23 23:08:03,345 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-23 23:08:03,345 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(StudentName,sector,DOB,qalification,,state,randomName)
(ABROSER,goverenment,18-11-2002,MBBS,3,Pennsylvania,prattville*)
(ALEXANDER,goverenment,20-10-2000,BSC,2,vermont,gadsden+)
(ALEXANDER,private,20-10-2000,BE,8,arizona,decatur!)
(ALEXANDER,goverenment,01-01-2003,BTECH,4,oregon,huntsville/)
(AGNEW,goverenment,20-10-2000,BCOM,7,california,dothan@)
(ATNEST,goverenment,20-10-2000,MTECH,8,arizona,decatur!)
(BELL,goverenment,10-07-2004,BBA,9,alaska,auburn~)
(BURR,goverenment,12-12-2001,BE,100,alabama,madison`)
(BURD,goverenment,20-10-2000,ME,6,louisiana,hoover#)
(BACHTEL,goverenment,28-04-2005,BE,100,alabama,madison`)
(MULVETS,goverenment,20-10-2000,MS,8,arizona,decatur!)
(BURTNER,goverenment,20-10-2000,BE,100,alabama,madison`)
(BLOOM,goverenment,20-10-2000,BE,4,oregon,huntsville/)
(BECHTEL,goverenment,20-10-2000,BE,5,Maryland,tuscaloosa$)
(GAFF,goverenment,18-11-2002,BE,100,alabama,madison`)
(CLARK,private,20-10-2000,BE,100,alabama,madison`)
```

grunt> STORE rawD INTO 'hbase://studentAcad' USING
org.apache.pig.backend.hadoop.hbase.HbaseStorage('student_data:StudentName',student_da
ta:sector,student_data:DOB,student_data:qualification,student_data:score,student_data:state,
student_data:randomName');

```
grunt> 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');
>> ;
>>
```