

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/>

To implement the concepts discussed, user is expected to have a Hadoop cluster with Pig and HBase running on it.

Transfer into hdfs

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

`hadoop fs -put AcadgildStudent /`

```
[acadgild@localhost Downloads]$ hadoop fs -put student /
17/12/18 08:58:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin
applicable
[acadgild@localhost Downloads]$ hadoop fs -ls /
17/12/18 08:58:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin
applicable
Found 6 items
drwxr-xr-x - acadgild supergroup          0 2017-12-18 04:21 /hbasestorage
drwxr-xr-x - acadgild supergroup          0 2017-11-28 15:23 /pig_data
-rw-r--r-- 1 acadgild supergroup    26205 2017-12-18 08:58 /student
drwxrwxr-x - acadgild supergroup          0 2017-12-10 20:02 /tmp
drwxr-xr-x - acadgild supergroup          0 2015-11-17 01:56 /user
drwxr-xr-x - acadgild supergroup          0 2015-11-05 12:56 /zookeeper
[acadgild@localhost Downloads]$ hadoop fs -cat /student
17/12/18 08:59:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin
applicable
StudentName,sector,DOB,qualification,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@
ATNEST,goverenment,20-10-2000,MTECH,8.5,arizona,decatur!
BELL,goverenment,10-07-2004,BBA,9.5,alaska,auburn~
BURR,goverenment,12-12-2001,BE,100,alabama,madison`
BURD,goverenment,20-10-2000,ME,6.5,louisiana,hooover#
BACHTEL,goverenment,28-04-2005,BE,100,alabama,madison`
MULVETS,goverenment,20-10-2000,MS,8.5,arizona,decatur!
BURTNER,goverenment,20-10-2000,BE,100,alabama,madison`
BLOOM,goverenment,20-10-2000,BE,4.5,oregon,huntsville/
BECHTEL,goverenment,20-10-2000,BE,5.5,Maryland,tuscaloosa$
GAFF,goverenment,18-11-2002,BE,100,alabama,madison`
CLARK,private,20-10-2000,BE,100,alabama,madison`
COOTS,goverenment,20-10-2000,BSC,100,alabama,madison`
DUNSDILL,goverenment,20-10-2000,BE,100,alabama,madison`
DORONES,goverenment,20-10-2000,BTECH,100,alabama,madison`
DICKEY,goverenment,20-10-2000,BBA,100,alabama,madison`
DARRAH,goverenment,20-10-2000,BE,100,alabama,madison`
```

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

`PIG_CLASSPATH=/usr/local/hbase/lib/hbase-common-0.98.14-
hadoop2.jar:/usr/local/hbase/lib/hbase-client-0.98.14-`

hadoop2.jar:/usr/local/hbase/lib/hbase-server-0.98.14-
hadoop2.jar:/usr/local/hbase/lib/hbase-protocol-0.98.14-
hadoop2.jar:/usr/local/hbase/lib/htrace-core-2.04.jar:/usr/local/hbase/lib/zookeeper-
3.4.6.jar:/usr/local/hbase/lib/guava-12.0.1.jar

```
[acadgild@localhost Downloads]$ PIG_CLASSPATH=/home/hadoop/HADOOP/hbase-0.98.4-hadoop2/lib/hbase-server-0.98.4-hadoop2:/home/hadoop/HADOOP/hbase-0.98.4-hadoop2/lib/hbase-*.jar;
[acadgild@localhost Downloads]$ hbase shell
2017-12-18 09:02:48,853 INFO [main] Configuration.deprecation: hadoop.native.lib is deprecated. Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 0.98.14-hadoop2, r4e4aabb93b52f1b0fef6b66edd06ec8923014dec, Tue Aug 25 22:35:44 PDT 2015
```

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.

```
hbase(main):010:0> create 'studentAcad','student_data'
0 row(s) in 2.5910 seconds

=> Hbase::Table - studentAcad
hbase(main):011:0> █
```

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 and then 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 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');
```

```

grunt>
grunt> rawD = load '/student' using PigStorage(',') as (StudentName:chararray,sector:chararray,DOB:chararray,qualification:chararray,score:int,state:chararray,randomName:chararray);
2017-12-18 09:35:28,630 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-12-18 09:35:28,630 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-12-18 09:35:28,632 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-12-18 09:35:28,916 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-12-18 09:35:28,917 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-12-18 09:35:28,917 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
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');
2017-12-18 09:36:29,210 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max

```

```

2017-12-18 09:37:18,455 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-12-18 09:37:18,461 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-12-18 09:37:18,579 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-12-18 09:37:18,585 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-12-18 09:37:18,675 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8032
2017-12-18 09:37:18,686 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-12-18 09:37:18,796 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregation.
2017-12-18 09:37:18,796 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt>

```

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

scan 'studentAcad'

```

hbase(main):011:0>
hbase(main):011:0> scan 'studentAcad'
ROW                                COLUMN+CELL
ABEDNIGO                          column=student_data:DOB, timestamp=1513570021894, value=BBA
ABEDNIGO                          column=student_data:StudentName, timestamp=1513570021894, value=goverenment
ABEDNIGO                          column=student_data:qualification, timestamp=1513570021894, value=100
ABEDNIGO                          column=student_data:score, timestamp=1513570021894, value=alabama
ABEDNIGO                          column=student_data:sector, timestamp=1513570021894, value=20-10-2000
ABEDNIGO                          column=student_data:state, timestamp=1513570021894, value=madison`
ABROSER                           column=student_data:DOB, timestamp=1513570021763, value=MBBS
ABROSER                           column=student_data:StudentName, timestamp=1513570021763, value=goverenment
ABROSER                           column=student_data:qualification, timestamp=1513570021763, value=3
ABROSER                           column=student_data:score, timestamp=1513570021763, value=Pennsylvania
ABROSER                           column=student_data:sector, timestamp=1513570021763, value=18-11-2002
ABROSER                           column=student_data:state, timestamp=1513570021763, value=prattville*
AGNES                             column=student_data:DOB, timestamp=1513570021894, value=BE
AGNES                             column=student_data:StudentName, timestamp=1513570021894, value=goverenment
AGNES                             column=student_data:qualification, timestamp=1513570021894, value=100
AGNES                             column=student_data:score, timestamp=1513570021894, value=alabama
AGNES                             column=student_data:sector, timestamp=1513570021894, value=20-10-2000
AGNES                             column=student_data:state, timestamp=1513570021894, value=madison`
AGNEW                             column=student_data:DOB, timestamp=1513570021764, value=BCOM
AGNEW                             column=student_data:StudentName, timestamp=1513570021764, value=goverenment
AGNEW                             column=student_data:qualification, timestamp=1513570021764, value=7
AGNEW                             column=student_data:score, timestamp=1513570021764, value=california
AGNEW                             column=student_data:sector, timestamp=1513570021764, value=20-10-2000
AGNEW                             column=student_data:state, timestamp=1513570021764, value=dothan@
ALEXANDER                        column=student_data:DOB, timestamp=1513570021895, value=MBBS
ALEXANDER                        column=student_data:StudentName, timestamp=1513570021895, value=goverenment
ALEXANDER                        column=student_data:qualification, timestamp=1513570021895, value=100
ALEXANDER                        column=student_data:score, timestamp=1513570021895, value=alabama
ALEXANDER                        column=student_data:sector, timestamp=1513570021895, value=20-10-2000
ALEXANDER                        column=student_data:state, timestamp=1513570021895, value=madison`
ANDREW                           column=student_data:DOB, timestamp=1513570021895, value=BE
ANDREW                           column=student_data:StudentName, timestamp=1513570021895, value=goverenment
ANDREW                           column=student_data:qualification, timestamp=1513570021895, value=100
ANDREW                           column=student_data:score, timestamp=1513570021895, value=alabama
ANDREW                           column=student_data:sector, timestamp=1513570021895, value=20-10-2000

```