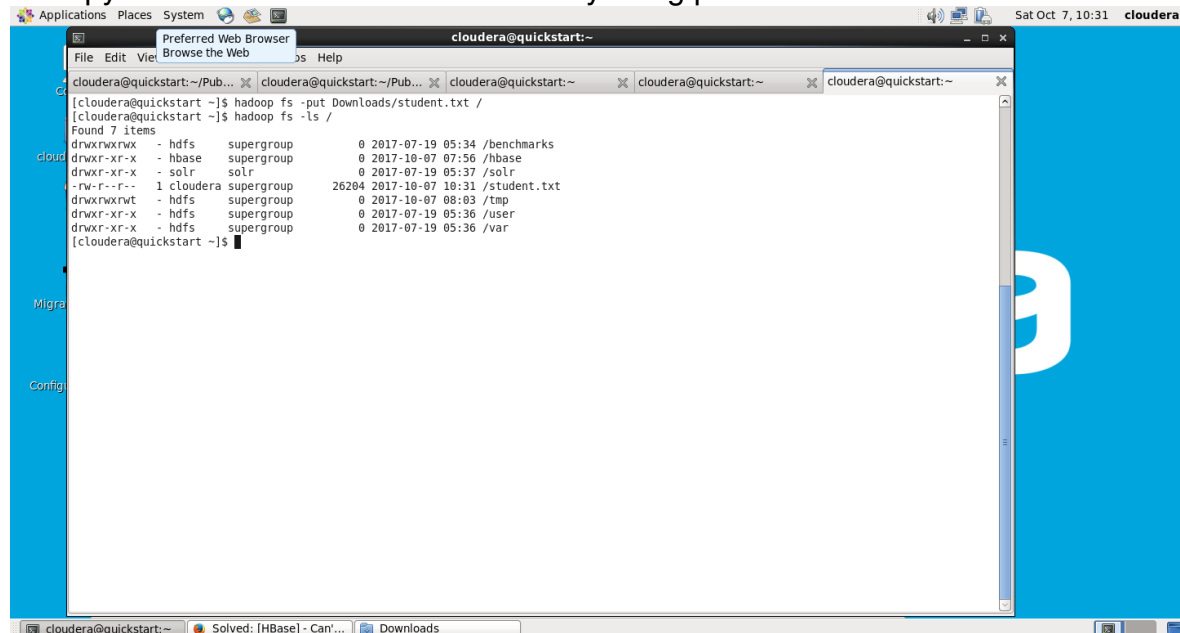


LOADING DATA INTO HBASE USING PIG SCRIPTS:

1.Data set containing student data :

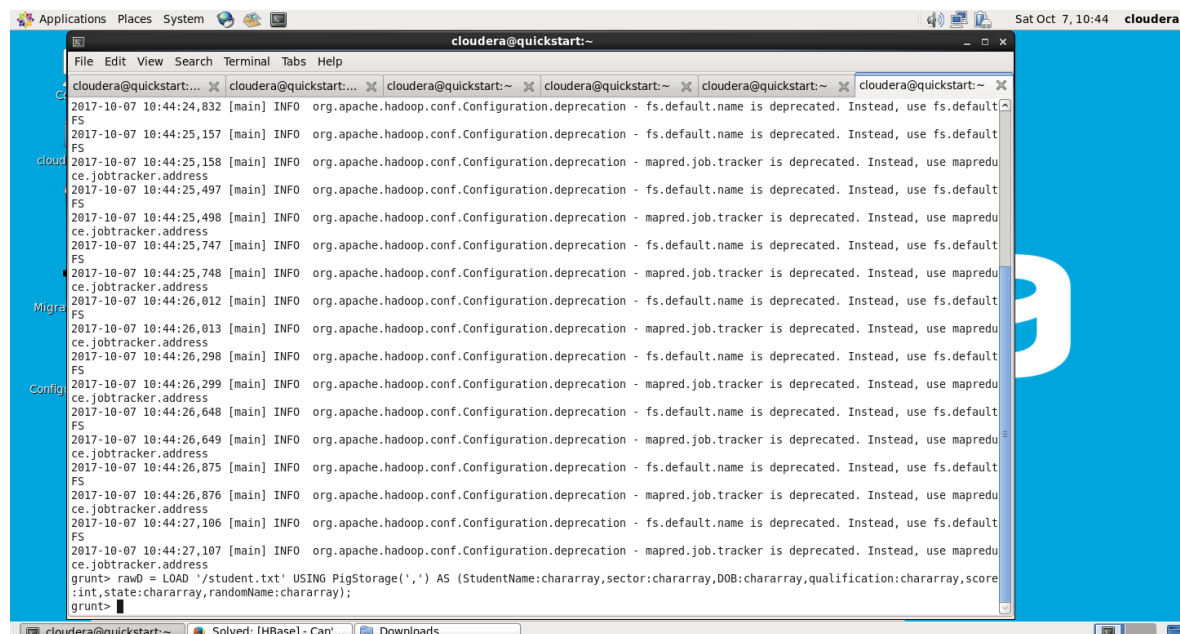
StudentName, Sector, DOB, Qualification, score, state, randomName

2. Copy data set to HDFS under / directory using put command:

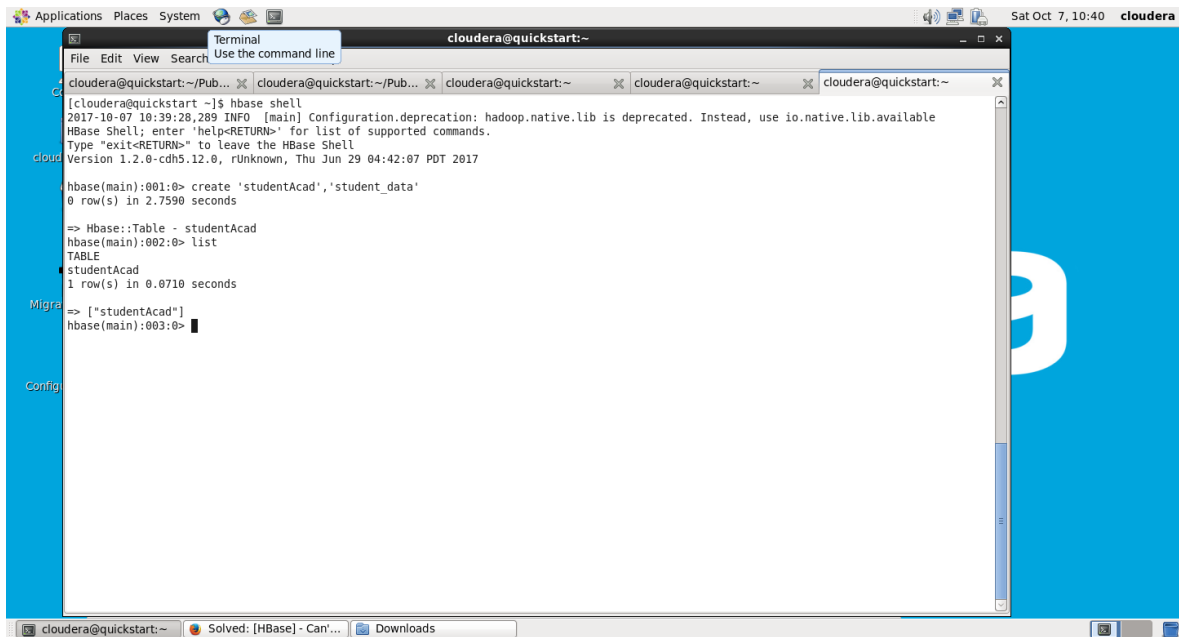


```
cloudera@quickstart:~$ hadoop fs -put Downloads/student.txt /
cloudera@quickstart:~$ hadoop fs -ls /
Found 7 items
drwxr-xr-x - hdfs supergroup          0 2017-07-19 05:34 /benchmarks
drwxr-xr-x - hbase supergroup          0 2017-10-07 07:56 /hbase
drwxr-xr-x - solr solr                  0 2017-07-19 05:37 /solr
-rw-r--r-- 1 cloudera supergroup 26204 2017-10-07 10:31 /student.txt
drwxr-xrwt - hdfs supergroup          0 2017-10-07 08:03 /tmp
drwxr-xr-x - hdfs supergroup          0 2017-07-19 05:36 /user
drwxr-xr-x - hdfs supergroup          0 2017-07-19 05:36 /var
cloudera@quickstart:~$
```

3. Now start hbase shell and create a studentAcad table using create command and list to check table is created or not:



```
cloudera@quickstart:~$ hbase shell
2017-10-07 10:44:24,832 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default
FS
2017-10-07 10:44:25,157 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default
FS
2017-10-07 10:44:25,158 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred
ce.jobtracker.address
2017-10-07 10:44:25,497 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default
FS
2017-10-07 10:44:25,498 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred
ce.jobtracker.address
2017-10-07 10:44:25,747 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default
FS
2017-10-07 10:44:25,748 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred
ce.jobtracker.address
2017-10-07 10:44:26,012 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default
FS
2017-10-07 10:44:26,013 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred
ce.jobtracker.address
2017-10-07 10:44:26,298 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default
FS
2017-10-07 10:44:26,299 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred
ce.jobtracker.address
2017-10-07 10:44:26,648 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default
FS
2017-10-07 10:44:26,649 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred
ce.jobtracker.address
2017-10-07 10:44:26,875 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default
FS
2017-10-07 10:44:26,876 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred
ce.jobtracker.address
2017-10-07 10:44:27,107 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred
ce.jobtracker.address
2017-10-07 10:44:27,107 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred
ce.jobtracker.address
grunt> rawD = LOAD 'student.txt' USING PigStorage(',') AS (StudentName:chararray,sector:chararray,DOB:chararray,qualification:chararray,score
:int,state:chararray,randomName:chararray);
grunt>
```



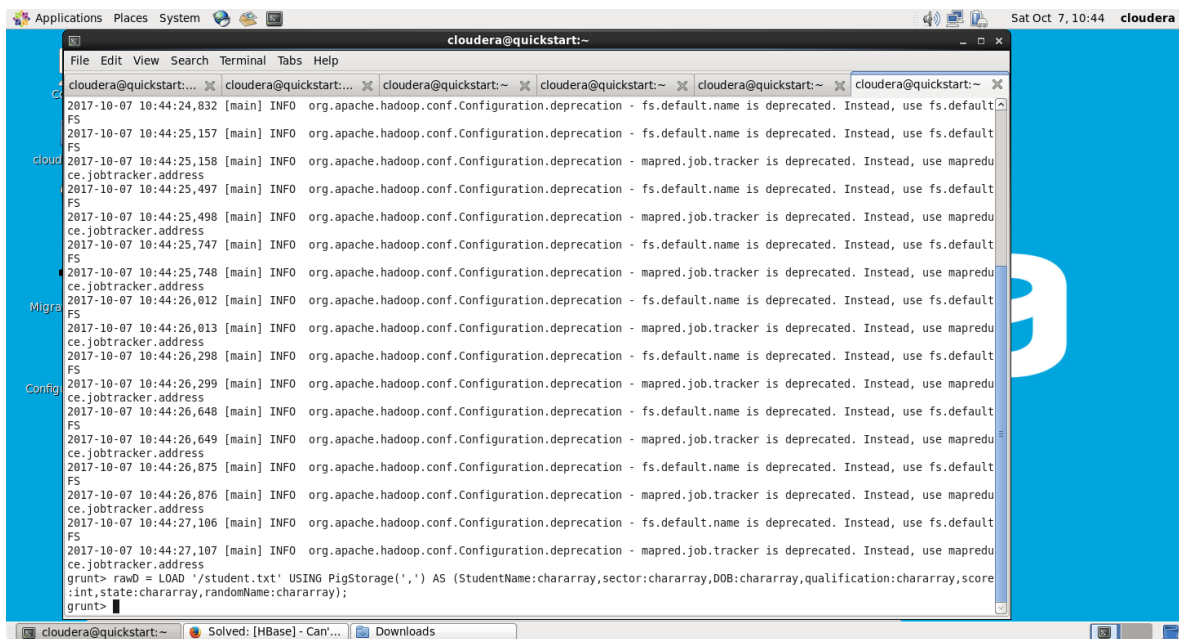
```
cloudera@quickstart:~$ hbase shell
2017-10-07 10:39:28,289 INFO [main] Configuration.deprecation: hadoop.native.lib is deprecated. Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.0-cdh5.12.0, rUnknown, Thu Jun 29 04:42:07 PDT 2017

hbase(main):001:0> create 'studentAcad','student_data'
0 row(s) in 2.7590 seconds

=> Hbase::Table - studentAcad
hbase(main):002:0> list
TABLE
studentAcad
1 row(s) in 0.0710 seconds

Migra => ["studentAcad"]
hbase(main):003:0>
```

3. login to grunt shell and load student.txt data from HDFS to pig variable using LOAD command:



```
cloudera@quickstart:~$ grunt
2017-10-07 10:44:24,832 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default
FS
2017-10-07 10:44:25,157 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default
FS
2017-10-07 10:44:25,158 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred
ce.jobtracker.address
2017-10-07 10:44:25,497 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default
FS
2017-10-07 10:44:25,498 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred
ce.jobtracker.address
2017-10-07 10:44:25,747 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default
FS
2017-10-07 10:44:25,748 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred
ce.jobtracker.address
2017-10-07 10:44:26,012 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default
FS
2017-10-07 10:44:26,013 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred
ce.jobtracker.address
2017-10-07 10:44:26,298 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default
FS
2017-10-07 10:44:26,299 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred
ce.jobtracker.address
2017-10-07 10:44:26,648 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default
FS
2017-10-07 10:44:26,649 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred
ce.jobtracker.address
2017-10-07 10:44:26,875 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default
FS
2017-10-07 10:44:26,876 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred
ce.jobtracker.address
2017-10-07 10:44:27,106 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default
FS
2017-10-07 10:44:27,107 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred
ce.jobtracker.address
grunt> rawD = LOAD '/student.txt' USING PigStorage(',') AS (StudentName:chararray,sector:chararray,D0B:chararray,qualification:chararray,score
:int,state:chararray,randomName:chararray);
grunt>
```

4. Once loading is done, check the data using dump command:

5. Transfer data inside HBASE by store command :

```
cloudera@quickstart:~  
File Edit View Search Terminal Tabs Help  
cloudera@quickstart:~  
2017-10-07 10:44:25,157 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default  
2017-10-07 10:44:25,158 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred  
ce.jobtracker.address  
2017-10-07 10:44:25,497 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default  
2017-10-07 10:44:25,498 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred  
ce.jobtracker.address  
2017-10-07 10:44:25,747 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default  
2017-10-07 10:44:25,748 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred  
ce.jobtracker.address  
2017-10-07 10:44:26,012 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default  
2017-10-07 10:44:26,013 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred  
ce.jobtracker.address  
2017-10-07 10:44:26,298 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default  
2017-10-07 10:44:26,299 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred  
ce.jobtracker.address  
2017-10-07 10:44:26,648 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default  
2017-10-07 10:44:26,649 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred  
ce.jobtracker.address  
2017-10-07 10:44:26,875 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default  
2017-10-07 10:44:26,876 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred  
ce.jobtracker.address  
2017-10-07 10:44:27,106 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.default  
2017-10-07 10:44:27,107 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapred  
ce.jobtracker.address  
grunt> rawD = LOAD '/student.txt' USING PigStorage(',') AS (StudentName:chararray,sector:chararray,D0B:chararray,qualification:chararray,score  
:int,state:chararray,randomName:chararray);  
grunt> STORE rawD INTO 'hbase://studentAcad' USING org.apache.pig.backend.hadoop.hbase.HBaseStorage('student_data:studentName,student_data:sec  
tor,student_data:D0B,student_data:qualification,student_data:score,student_data:state,student_data:randomName');
```

6. Confirm data is loaded inside HABSE by checking success message:

```
cloudera@quickstart:~  
File Edit View Se Send email Help  
cloudera@quickstart:~  
2017-10-07 10:54:00,562 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.reduce.tasks is deprecated. Instead, use mapred  
uce.job.reduces  
2017-10-07 10:54:00,765 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete  
2017-10-07 10:54:00,780 [main] INFO org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:  
  
HadoopVersion PigVersion UserId StartedAt FinishedAt Features  
2.6.0-cdh5.12.0 0.12.0-cdh5.12.0 cloudera 2017-10-07 10:52:17 2017-10-07 10:54:00 UNKNOWN  
  
Success!  
  
Job Stats (time in seconds):  
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime Median  
Reductime Alias Feature Outputs  
Job_1507287275802_0006 1 0 25 25 25 25 n/a n/a n/a n/a rawD MAP_ONLY hbase://studen  
tAcad,  
  
Input(s):  
Successfully read 470 records (26566 bytes) from: "/student.txt"  
  
Output(s):  
Successfully stored 470 records in: "hbase://studentAcad"  
  
Counters:  
Total records written : 470  
Total bytes written : 0  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0  
  
Job DAG:  
job_1507287275802_0006  
  
2017-10-07 10:54:01,045 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIEL  
D DISCARDED TYPE CONVERSION FAILED 1 time(s).  
2017-10-07 10:54:01,045 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
grunt>
```

7. Check data in the base using scan command on StudentAcad table :

```
Applications Places System Sat Oct 7, 10:57 cloudera
cloudera@quickstart:~
File Edit View Search Terminal Tabs Help
cloudera@quickstart:~ cloudera@quickstart:~ cloudera@quickstart:~ cloudera@quickstart:~ cloudera@quickstart:~ cloudera@quickstart:~
cloud WILLIAM column=student_data:score, timestamp=1507398829130, value=alabama
cloud WILLIAM column=student_data:sector, timestamp=1507398829130, value=20-10-2000
cloud WILLIAM column=student_data:state, timestamp=1507398829130, value=madison
cloud WILLIAM column=student_data:studentName, timestamp=1507398829130, value=governement
cloud WILSON column=student_data:DOB, timestamp=1507398829039, value=BE
cloud WILSON column=student_data:qualification, timestamp=1507398829039, value=100
cloud WILSON column=student_data:score, timestamp=1507398829039, value=alabama
cloud WILSON column=student_data:sector, timestamp=1507398829039, value=20-10-2000
cloud WILSON column=student_data:state, timestamp=1507398829039, value=madison
cloud WILSON column=student_data:studentName, timestamp=1507398829039, value=governement
cloud WITMORE column=student_data:DOB, timestamp=1507398828748, value=BE
cloud WITMORE column=student_data:qualification, timestamp=1507398828748, value=100
cloud WITMORE column=student_data:score, timestamp=1507398828748, value=alabama
cloud WITMORE column=student_data:sector, timestamp=1507398828748, value=20-10-2000
cloud WITMORE column=student_data:state, timestamp=1507398828748, value=madison
cloud WITMORE column=student_data:studentName, timestamp=1507398828748, value=governement
cloud WOODS column=student_data:DOB, timestamp=1507398829041, value=MBBS
cloud WOODS column=student_data:qualification, timestamp=1507398829041, value=100
cloud WOODS column=student_data:score, timestamp=1507398829041, value=alabama
cloud WOODS column=student_data:sector, timestamp=1507398829041, value=20-10-2000
cloud WOODS column=student_data:state, timestamp=1507398829041, value=madison
cloud WOODS column=student_data:studentName, timestamp=1507398829041, value=governement
cloud WORK column=student_data:DOB, timestamp=1507398828749, value=BE
cloud WORK column=student_data:qualification, timestamp=1507398828749, value=100
cloud WORK column=student_data:score, timestamp=1507398828749, value=alabama
cloud WORK column=student_data:sector, timestamp=1507398828749, value=28-04-2005
cloud WORK column=student_data:state, timestamp=1507398828749, value=madison
cloud WORK column=student_data:studentName, timestamp=1507398828749, value=governement
cloud YOUNG column=student_data:DOB, timestamp=1507398828750, value=MS
cloud YOUNG column=student_data:qualification, timestamp=1507398828750, value=9
cloud YOUNG column=student_data:score, timestamp=1507398828750, value=alaska
cloud YOUNG column=student_data:sector, timestamp=1507398828750, value=20-10-2000
cloud YOUNG column=student_data:state, timestamp=1507398828750, value=auburn-
cloud YOUNG column=student_data:studentName, timestamp=1507398828750, value=governement
230 row(s) in 8.8170 seconds
hbase(main):004:0>
```

Total 230 rows copied from student.txt to HBase table.