

Problem Statement:

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

```
=====
[cloudera@quickstart ~]$ hadoop fs -copyFromLocal /home/cloudera/chhaya/student.txt /user/cloudera/student.txt
```

```
[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera/student.txt
```

```
[cloudera@quickstart ~]$ start-hbase.sh
```

```
[cloudera@quickstart ~]$ hbase shell
```

```
hbase(main):001:0> create 'studentdemo','studentdetails'
```

```
[cloudera@quickstart ~]$ pig
```

```
grunt> REGISTER /home/cloudera/Downloads/piggybank-0.15.0.jar ;
```

```
grunt> A = LOAD '/user/cloudera/student.txt'
```

```
USING PigStorage(',') AS (studentname:chararray,sector:chararray,dateofbirth:chararray,
qualification:chararray,score:float,state:chararray,randomname:chararray)
```

```
;
```

```
grunt> DESCRIBE A ;
```

```
grunt> B = RANK A;
```

```
grunt> C = FILTER B BY $0 > 1;
```

```
grunt> D = FOREACH C GENERATE studentname,sector,dateofbirth,qualification,score,state,randomname ;
```

```
grunt> E = RANK D;
```

```
grunt> STORE E INTO 'hbase://studentdemo' USING org.apache.pig.backend.hadoop.hbase.HBaseStorage(
```

```
'studentdetails:studentname
```

```
studentdetails:sector
```

```
studentdetails:dateofbirth
```

```
studentdetails:qualification
```

```
studentdetails:score
```

```
studentdetails:state
```

```
studentdetails:randomname'
```

```
)
```

```
;
```

OUTPUT OF STORE COMMAND:

```
HadoopVersion  PigVersion  UserId  StartedAt  FinishedAt  Features
2.6.0-cdh5.12.0 0.12.0-cdh5.12.0 cloudera 2017-10-15 08:41:14 2017-10-15 08:43:33 RANK,FILTER

Success!

Job Stats (time in seconds):
JobId  Maps  Reduces  MaxMapTime  MinMapTime  AvgMapTime  MedianMapTime  MaxReduceTime  MinReduceTime  AvgReduceTime  MedianReduceTime  Alias  Featu
re  Outputs
job_15080801611211_0001  1  0  15  15  15  15  n/a  n/a  n/a  n/a  A  MAP_ONLY
job_15080801611211_0002  1  0  9  9  9  9  n/a  n/a  n/a  n/a  C,D  MAP_ONLY
job_15080801611211_0003  1  0  14  14  14  14  n/a  n/a  n/a  n/a  MAP_ONLY  hbase://studentdemo,

Input(s):
Successfully read 470 records (26580 bytes) from: "/user/cloudera/student.txt"

Output(s):
Successfully stored 469 records in: "hbase://studentdemo"

Counters:
Total records written : 469
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
```

---OUTPUT VERIFICATION FOR HBASE TABLE ONCE LOAD ACTIVITY COMPLETES FROM PIG

```
hbase(main):004:0> count 'studentdemo'
```

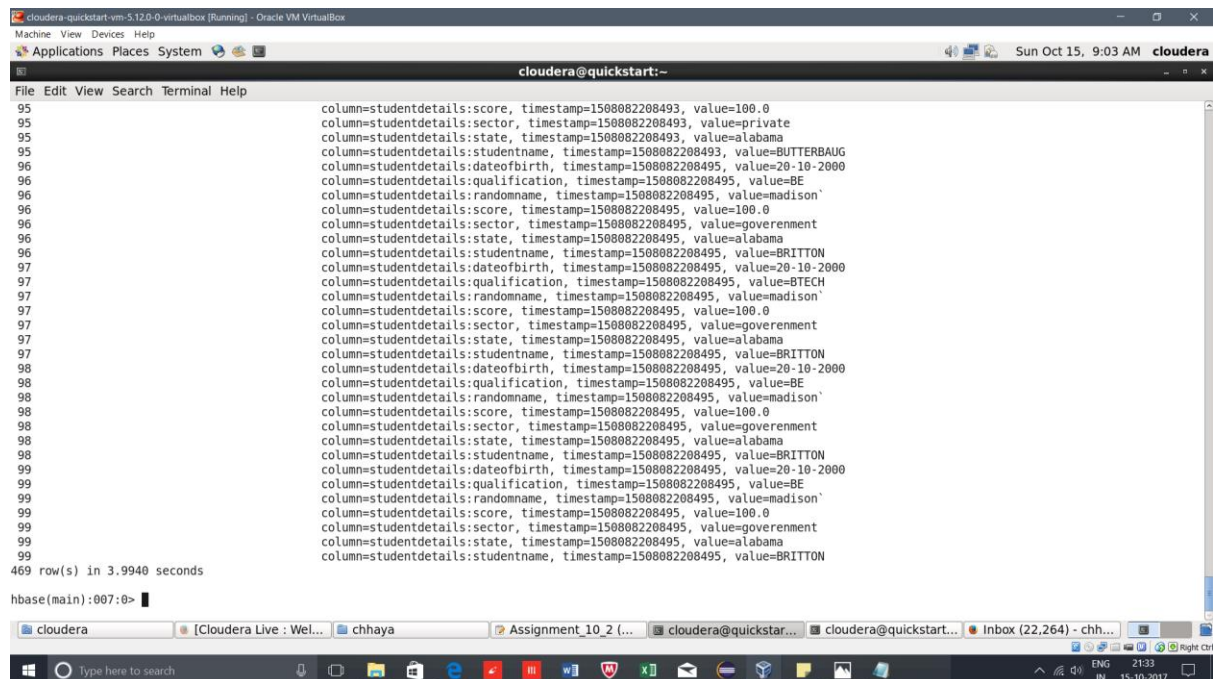
```
hbase(main):005:0> scan 'studentdemo'
```

```
hbase(main):006:0> exit;
```

```
hbase(main):004:0> create 'studentdemo', 'studentdetails'
0 row(s) in 1.4680 seconds

=> Hbase::Table - studentdemo
hbase(main):005:0> count 'studentdemo'
469 row(s) in 0.1960 seconds

=> 469
hbase(main):006:0> scan 'studentdemo'
```



```
cloudera-quickstart-vm:5.12.0-0-virtualbox (Running) - Oracle VM VirtualBox
Machine View Devices Help
Applications Places System
cloudera@quickstart:~$
File Edit View Search Terminal Help
95 column=studentdetails:score, timestamp=1508082208493, value=100.0
95 column=studentdetails:sector, timestamp=1508082208493, value=private
95 column=studentdetails:state, timestamp=1508082208493, value=alabama
95 column=studentdetails:studentname, timestamp=1508082208493, value=BUTTERBAUG
96 column=studentdetails:dateofbirth, timestamp=1508082208495, value=20-10-2000
96 column=studentdetails:qualification, timestamp=1508082208495, value=BE
96 column=studentdetails:randomname, timestamp=1508082208495, value=madison
96 column=studentdetails:score, timestamp=1508082208495, value=100.0
96 column=studentdetails:sector, timestamp=1508082208495, value=government
96 column=studentdetails:state, timestamp=1508082208495, value=alabama
96 column=studentdetails:studentname, timestamp=1508082208495, value=BRITTON
97 column=studentdetails:dateofbirth, timestamp=1508082208495, value=20-10-2000
97 column=studentdetails:qualification, timestamp=1508082208495, value=BTECH
97 column=studentdetails:randomname, timestamp=1508082208495, value=madison
97 column=studentdetails:score, timestamp=1508082208495, value=100.0
97 column=studentdetails:sector, timestamp=1508082208495, value=government
97 column=studentdetails:state, timestamp=1508082208495, value=alabama
97 column=studentdetails:studentname, timestamp=1508082208495, value=BRITTON
98 column=studentdetails:dateofbirth, timestamp=1508082208495, value=20-10-2000
98 column=studentdetails:qualification, timestamp=1508082208495, value=BE
98 column=studentdetails:randomname, timestamp=1508082208495, value=madison
98 column=studentdetails:score, timestamp=1508082208495, value=100.0
98 column=studentdetails:sector, timestamp=1508082208495, value=government
98 column=studentdetails:state, timestamp=1508082208495, value=alabama
98 column=studentdetails:studentname, timestamp=1508082208495, value=BRITTON
99 column=studentdetails:dateofbirth, timestamp=1508082208495, value=20-10-2000
99 column=studentdetails:qualification, timestamp=1508082208495, value=BE
99 column=studentdetails:randomname, timestamp=1508082208495, value=madison
99 column=studentdetails:score, timestamp=1508082208495, value=100.0
99 column=studentdetails:sector, timestamp=1508082208495, value=government
99 column=studentdetails:state, timestamp=1508082208495, value=alabama
99 column=studentdetails:studentname, timestamp=1508082208495, value=BRITTON
469 row(s) in 3.9940 seconds
hbase(main):007:0>
```