

BigData Assignment 4.2

Loading Data Into HBase Using PIG Scripts

Solution -

- Copying the dataset into hdfs.
hdfs dfs -put AcadgildStudent.txt /

```
cloudera@quickstart:~/Downloads
File Edit View Search Terminal Help
[cloudera@quickstart Downloads]$ ls
AcadgildStudent.txt hbase-0.98.4-hadoop2 hbase-0.98.4-hadoop2-bin.tar.gz
[cloudera@quickstart Downloads]$ hdfs dfs -put AcadgildStudent.txt /
[cloudera@quickstart Downloads]$ hdfs dfs -ls /
Found 7 items
-rw-r--r-- 1 cloudera supergroup 26204 2018-07-10 12:22 /AcadgildStudent.txt
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase supergroup 0 2018-07-10 10:07 /hbase
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxrwxrwt - hdfs supergroup 0 2018-07-06 06:09 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
[cloudera@quickstart Downloads]$
```

- Adding jar files of HBase to the Pig classpath and loading the data into a relation.

REGISTER 'usr/lib/hbase/*.jar';

data = LOAD 'hdfs://localhost:8020/student.txt' USING

PigStorage(',')

AS(studentname:chararray,sector:chararray,dob:chararray,qualification:chararray,score:int,state:chararray,randonName:chararray);

```
cloudera@quickstart:~
File Edit View Search Terminal Help
grunt> REGISTER 'usr/lib/hbase/*.jar';
grunt> data = LOAD 'hdfs://localhost:8020/AcadgildStudent.txt' USING PigStorage(',') AS(studentname:chararray,sector:chararra
2018-07-10 12:28:53,062 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
2018-07-10 12:28:53,062 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Ins
tead, use mapreduce.jobtracker.address
2018-07-10 12:28:53,062 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated.
Instead, use dfs.bytes-per-checksum
grunt>
```

- Cheking the data is there in the relation
dump data;

```
File Edit View Search Term
grunt> dump data;
```

```
cloudera@quickstart:~
File Edit View Search Terminal Help
2018-07-10 12:37:12,154 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.Task - Task 'attempt_local41119371_0003_m_0000000_0' done.
2018-07-10 12:37:12,154 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.LocalJobRunner - Finishing task: attempt_local41119371_0003_m_0000000_0
2018-07-10 12:37:12,154 [Thread-198] INFO org.apache.hadoop.mapred.LocalJobRunner - map task executor complete.
2018-07-10 12:37:16,269 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2018-07-10 12:37:16,269 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
2018-07-10 12:37:22,270 [main] WARN org.apache.pig.tools.pigstats.PigStatsUtil - Failed to get RunningJob for job job_local41119371_0003
2018-07-10 12:37:22,271 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2018-07-10 12:37:22,271 [main] INFO org.apache.pig.tools.pigstats.SimplePigStats - Detected Local mode. Stats reported below may be incomplete
2018-07-10 12:37:22,301 [main] INFO org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.6.0-cdh5.15.0 0.12.0-cdh5.15.0 cloudera 2018-07-10 12:37:08 2018-07-10 12:37:22 UNKNOWN

Success!

Job Stats (time in seconds):
JobId Alias Feature Outputs
job_local41119371_0003 data MAP_ONLY hbase://studentAcad,

Input(s):
Successfully read records from: "hdfs://localhost:8020/AcadgildStudent.txt"

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

Job DAG:
job_local41119371_0003

2018-07-10 12:37:28,303 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt>
```

```
cloudera@quickstart:~
File Edit View Search Terminal Help
(SARAH,governement,20-10-2000,BE,100,alabama,madison`)
(WILLIAM,governement,20-10-2000,BE,100,alabama,madison`)
(JOHN,governement,20-10-2000,BE,100,alabama,madison`)
(JACOB,governement,12-12-2001,BE,100,alabama,madison`)
(THOMAS,governement,20-10-2000,BE,100,alabama,madison`)
(JESSE,governement,20-10-2000,BSC,100,alabama,madison`)
(CHARLES,governement,20-10-2000,BE,8,arizona,decatur!)
(MCWILLIAM,governement,20-10-2000,BE,3,Pennsylvania,prattville*)
(PATRICK,governement,20-10-2000,BE,100,alabama,madison`)
(PATRICK,governement,28-04-2005,BE,100,alabama,madison`)
(WILLIAM,governement,20-10-2000,BBA,100,alabama,madison`)
(JOHN,governement,20-10-2000,BE,100,alabama,madison`)
(SAMUEL,governement,20-10-2000,BE,3,Pennsylvania,prattville*)
(CHRISTIAN,governement,20-10-2000,BE,100,alabama,madison`)
(JOHN,governement,20-10-2000,BE,100,alabama,madison`)
(THOMAS,governement,20-10-2000,BE,100,alabama,madison`)
(ELEONER,governement,20-10-2000,BE,100,alabama,madison`)
(MARTIN,governement,20-10-2000,BE,100,alabama,madison`)
(MARTIN,governement,20-10-2000,BE,100,alabama,madison`)
(NEWCOMMER,governement,10-07-2004,ME,7,california,dothan@)
(JOHN,governement,20-10-2000,BCOM,100,alabama,madison`)
(WILLIAM,governement,20-10-2000,MBBS,100,alabama,madison`)
(PHILIP,private,20-10-2000,BE,100,alabama,madison`)
(STEPHEL,governement,20-10-2000,BE,2,vermont,gadsden+)
(REBECAH,governement,20-10-2000,BE,100,alabama,madison`)
(SARAH,governement,20-10-2000,BE,9,alaska,auburn~)
(SUSANNAH,governement,20-10-2000,BE,100,alabama,madison`)
(PAUL,governement,20-10-2000,BE,6,louisiana,hoover#)
(GEORGE,governement,20-10-2000,BE,100,alabama,madison`)
(GEORGE,governement,20-10-2000,BE,100,alabama,madison`)
(GEORGE,governement,18-11-2002,BE,8,arizona,decatur!)
(GEORGE,governement,12-12-2001,BTECH,100,alabama,madison`)
(GEORGE,governement,01-01-2003,BE,100,alabama,madison`)
(GEORGE,governement,20-10-2000,BE,100,alabama,madison`)
(GEORGE,governement,20-10-2000,BE,100,alabama,madison`)
(GEORGE,governement,20-10-2000,BE,100,alabama,madison`)
(FREDERICK,governement,20-10-2000,BE,100,alabama,madison`)
(RICHARD,governement,20-10-2000,BE,4,oregon,huntsville/)
grunt>
```

- Creating a table in hbase
create 'studentAcad' , 'studentData'

```

cloudera@qu
File Edit View Search Terminal Help
hbase(main):002:0> create 'studentAcad','studentData'
0 row(s) in 1.2650 seconds

=> Hbase::Table - studentAcad
hbase(main):003:0> █

```

- Transfer the data inside HBase by STORE command.

STORE data 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');

```

cloudera@quickstart:~
File Edit View Search Terminal Help
grunt> STORE data INTO 'hbase://studentAcad' USING org.apache.pig.backend.hadoop.hbase.HBaseStorage('studentData:studentname,studentData:sector,studentData:dob,studentData:qualification,studentData:score,studentData:state,studentData:randomName'); █

```

```

cloudera@quickstart:~
File Edit View Search Terminal Help
shut down
2018-07-10 13:05:26,940 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.Task - Task:attempt_local624658582_0005_m_000000_0 is done. And is in the process of committing
2018-07-10 13:05:26,979 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.LocalJobRunner - map
2018-07-10 13:05:26,979 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.Task - Task 'attempt_local624658582_0005_m_000000_0' done.
2018-07-10 13:05:26,979 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.LocalJobRunner - Finishing task: attempt_local624658582_0005_m_000000_0
2018-07-10 13:05:26,979 [Thread-346] INFO org.apache.hadoop.mapred.LocalJobRunner - map task executor complete.
2018-07-10 13:05:38,078 [main] WARN org.apache.pig.tools.pigstats.PigStatsUtil - Failed to get RunningJob for job job_local624658582_0005
2018-07-10 13:05:38,079 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2018-07-10 13:05:38,079 [main] INFO org.apache.pig.tools.pigstats.SimplePigStats - Detected Local mode. Stats reported below may be incomplete
2018-07-10 13:05:38,079 [main] INFO org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion  PigVersion  UserId  StartedAt  FinishedAt  Features
2.6.0-cdh5.15.0 0.12.0-cdh5.15.0 cloudera 2018-07-10 13:05:24 2018-07-10 13:05:38 UNKNOWN

Success!

Job Stats (time in seconds):
JobId  Alias  Feature  Outputs
job_local624658582_0005 data  MAP_ONLY  hbase://studentAcad,

Input(s):
Successfully read records from: "hdfs://localhost:8020/AcadgildStudent.txt"

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

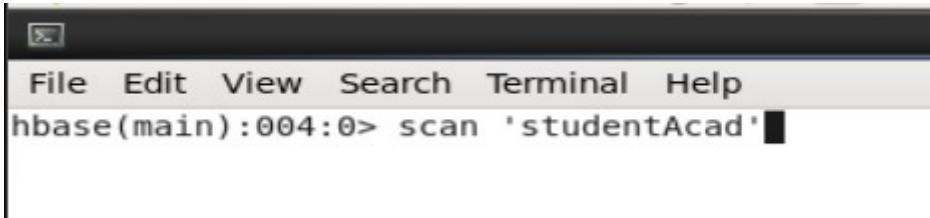
Job DAG:
job_local624658582_0005

2018-07-10 13:05:44,080 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> █

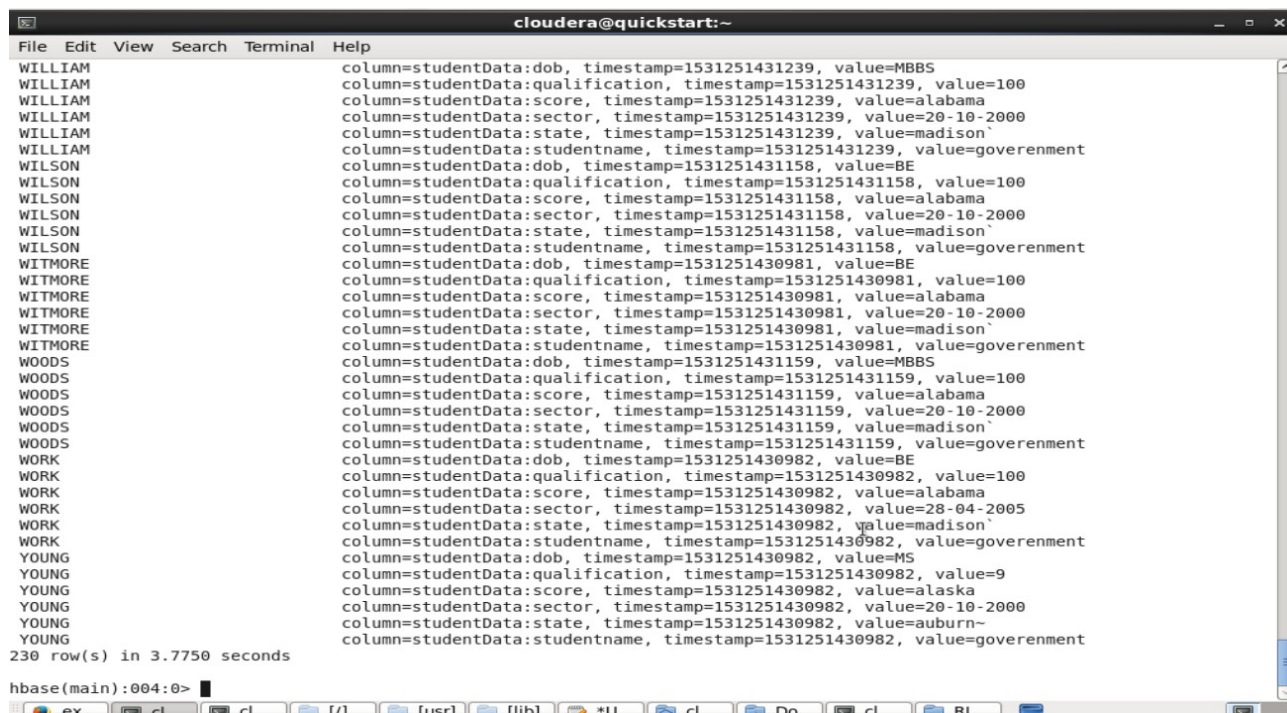
```

- Checking the content of hbase table

scan 'studentAcad'



```
File Edit View Search Terminal Help
hbase(main):004:0> scan 'studentAcad'
```



```
cloudera@quickstart:~
File Edit View Search Terminal Help
column=studentData:dob, timestamp=1531251431239, value=MBBS
WILLIAM column=studentData:qualification, timestamp=1531251431239, value=100
WILLIAM column=studentData:score, timestamp=1531251431239, value=alabama
WILLIAM column=studentData:sector, timestamp=1531251431239, value=20-10-2000
WILLIAM column=studentData:state, timestamp=1531251431239, value=madison
WILLIAM column=studentData:studentname, timestamp=1531251431239, value=governement
WILSON column=studentData:dob, timestamp=1531251431158, value=BE
WILSON column=studentData:qualification, timestamp=1531251431158, value=100
WILSON column=studentData:score, timestamp=1531251431158, value=alabama
WILSON column=studentData:sector, timestamp=1531251431158, value=20-10-2000
WILSON column=studentData:state, timestamp=1531251431158, value=madison
WILSON column=studentData:studentname, timestamp=1531251431158, value=governement
WITMORE column=studentData:dob, timestamp=1531251430981, value=BE
WITMORE column=studentData:qualification, timestamp=1531251430981, value=100
WITMORE column=studentData:score, timestamp=1531251430981, value=alabama
WITMORE column=studentData:sector, timestamp=1531251430981, value=20-10-2000
WITMORE column=studentData:state, timestamp=1531251430981, value=madison
WITMORE column=studentData:studentname, timestamp=1531251430981, value=governement
WOODS column=studentData:dob, timestamp=1531251431159, value=MBBS
WOODS column=studentData:qualification, timestamp=1531251431159, value=100
WOODS column=studentData:score, timestamp=1531251431159, value=alabama
WOODS column=studentData:sector, timestamp=1531251431159, value=20-10-2000
WOODS column=studentData:state, timestamp=1531251431159, value=madison
WOODS column=studentData:studentname, timestamp=1531251431159, value=governement
WORK column=studentData:dob, timestamp=1531251430982, value=BE
WORK column=studentData:qualification, timestamp=1531251430982, value=100
WORK column=studentData:score, timestamp=1531251430982, value=alabama
WORK column=studentData:sector, timestamp=1531251430982, value=28-04-2005
WORK column=studentData:state, timestamp=1531251430982, value=madison
WORK column=studentData:studentname, timestamp=1531251430982, value=governement
YOUNG column=studentData:dob, timestamp=1531251430982, value=MS
YOUNG column=studentData:qualification, timestamp=1531251430982, value=9
YOUNG column=studentData:score, timestamp=1531251430982, value=alaska
YOUNG column=studentData:sector, timestamp=1531251430982, value=20-10-2000
YOUNG column=studentData:state, timestamp=1531251430982, value=auburn
YOUNG column=studentData:studentname, timestamp=1531251430982, value=governement
230 row(s) in 3.7750 seconds
hbase(main):004:0>
```

