

US ELECTION RESULT (using PIG)

Submitted by
S. MOHANRAJ,
B-TECH IT,
INSTITUTE OF ROAD AND TRANSPORT TECHNOLOGY,
ERODE.

Abstract:

The aim of this paper is to propose a way to conduct elections in better and predict the result accurately. The analysis is done using pig tool in bigdata. The csv file has been downloaded from internet and it is stored in HDFS. Then it is analyzed using pig then display the final result.

Steps Performed and Commands used:

The csv file I have downloaded is [linked here](#). I had made some changes in column field for the easy analysis.

Commands:

In terminal after starting Hadoop servers,

1. pig -x mapreduce
2. grunt> election = LOAD '/election/election.csv' using PigStorage(',') as (id:int,PartyName:chararray,VoteCount:int);

3. grp = GROUP election by PartyName;

4. C = foreach grp GENERATE group, SUM(election.VoteCount);

5. dump C;

(Green,6525)

(PartyName,)

(Democratic,231198)

(Republican,217180)

(Libertarian,10859)

6. C = foreach C GENERATE (chararray)\$0 as name, (long)\$1 as vote;

7. D = ORDER C by vote DESC;

(Democratic,231198)

(Republican,217180)

(Libertarian,10859)

(Green,6525)

(PartyName,)

8. result = LIMIT D 1;

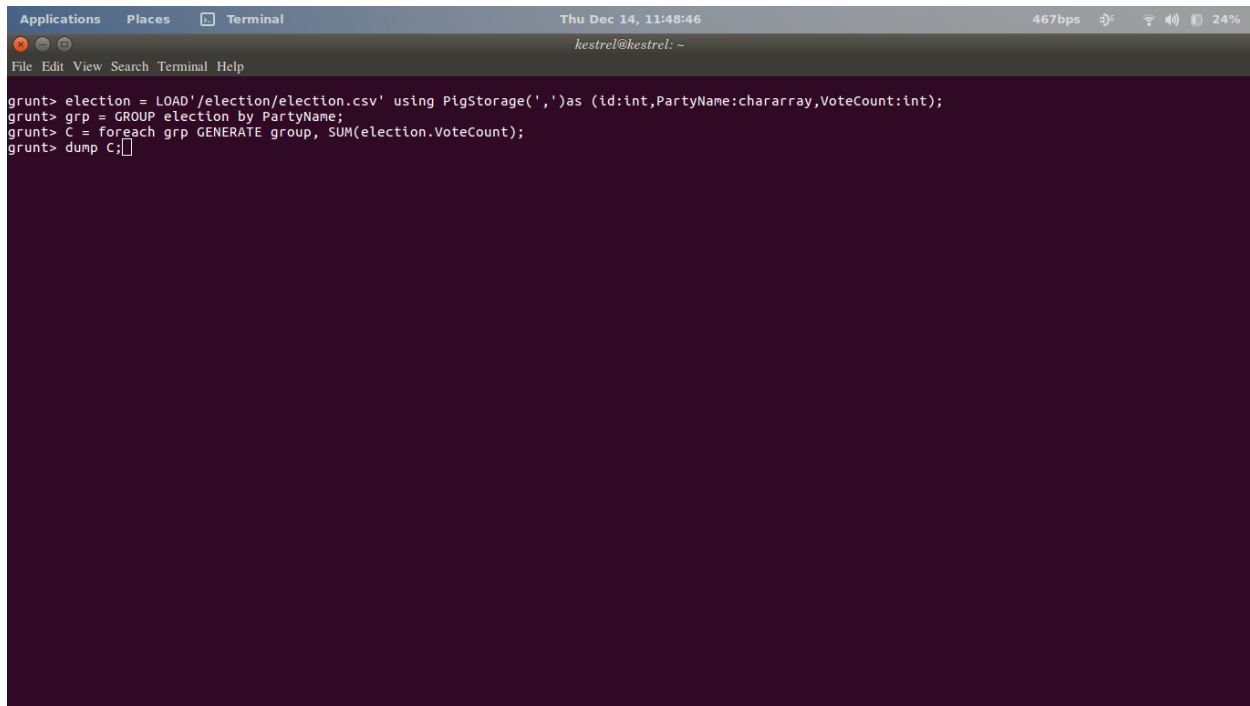
The result is, (*Democratic,231198*)

Hence Democratic is the winner.!

By using the above commands, the final output will be displayed as below the final screenshot.

Screenshots:

1. Loading file from HDFS to pig variable and grouping it using Party name:



```
Applications  Places  Terminal  Thu Dec 14, 11:48:46  467bps  24%
kestrel@kestrel: ~
File Edit View Search Terminal Help

grunt> election = LOAD '/election/election.csv' using PigStorage(',') as (id:int,PartyName:chararray,VoteCount:int);
grunt> grp = GROUP election by PartyName;
grunt> C = foreach grp GENERATE group, SUM(election.VoteCount);
grunt> dump C;
```

2. The dump Result for variable C:

```
Applications  Places  Terminal  Thu Dec 14, 11:49:14  0bps  24%
kestrel@kestrel: ~

File Edit View Search Terminal Help

HadoopVersion  PigVersion  UserId  StartedAt  FinishedAt  Features
1.2.1  0.11.1  kestrel  2017-12-14 11:48:47  2017-12-14 11:49:11  GROUP_BY

Success!

Job Stats (time in seconds):
JobId  Maps  Reduces  MaxMapTime  MinMapTime  AvgMapTime  MedianMapTime  MaxReduceTime  MinReduceTime  AvgReduceTime  MedianReduceTime
Job_201712141119_0010  1  1  2  2  2  2  9  9  9  9
/127.0.0.1:9000/tmp/temp1839124311/tmp-946467649,  C,election,grp  GROUP_BY,COMBINER  hdfs:/

Input(s):
Successfully read 2073 records (38879 bytes) from: "/election/election.csv"

Output(s):
Successfully stored 5 records (97 bytes) in: "hdfs://127.0.0.1:9000/tmp/temp1839124311/tmp-946467649"

Counters:
Total records written : 5
Total bytes written : 97
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
Job_201712141119_0010

2017-12-14 11:49:11,093 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 2 time(s).
2017-12-14 11:49:11,093 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-12-14 11:49:11,094 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2017-12-14 11:49:11,100 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-12-14 11:49:11,100 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(PartyName,)
(Democratic,231198)
(Republican,217180)
(Libertarian,10859)
grunt> 
```

3. Ordering it by descending order:

```
Applications  Places  Terminal  Thu Dec 14, 11:50:21  0bps  23%
kestrel@kestrel: ~

File Edit View Search Terminal Help

grunt> C = foreach C GENERATE (chararray)$0 as name, (long)$1 as vote;
grunt> describe C;
C: {name: chararray,vote: long}
grunt> o = ORDER C by vote DESC;
grunt> 
```

```
Applications  Places  Terminal  Thu Dec 14, 11:52:02  0bps  23%
kestrel@kestrel: ~

File Edit View Search Terminal Help

grunt> result = LIMIT o 1;
grunt> dump result;
```

4. The final Result is:

```
Applications  Places  Terminal  Thu Dec 14, 11:53:51  0bps  21%
kestrel@kestrel: ~

File Edit View Search Terminal Help

Success!

Job Stats (time in seconds):
JobId  Maps  Reduces  MaxMapTime  MinMapTime  AvgMapTime  MedianMapTime  MaxReduceTime  MinReduceTime  AvgReduceTime  MedianReduceTime
me  Alias  Feature  Outputs
job_201712141119_0014  1  1  2  2  2  2  9  9  9  9  C,election,grp  GROUP_BY,COMBINER
job_201712141119_0015  1  1  1  1  1  1  9  9  9  9  o  SAMPLER
job_201712141119_0016  1  1  2  2  2  2  9  9  9  9  o  ORDER_BY,COMBINER
job_201712141119_0017  1  1  1  1  1  1  9  9  9  9  o  hdfs://127.0.0.1:9000/tmp/temp
1839124311/tmp817515849,

Input(s):
Successfully read 2073 records (38879 bytes) from: "/election/election.csv"

Output(s):
Successfully stored 1 records (22 bytes) in: "hdfs://127.0.0.1:9000/tmp/temp1839124311/tmp817515849"

Counters:
Total records written : 1
Total bytes written : 22
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_201712141119_0014  ->  job_201712141119_0015,
job_201712141119_0015  ->  job_201712141119_0016,
job_201712141119_0016  ->  job_201712141119_0017,
job_201712141119_0017

2017-12-14 11:53:26,311 [main] WARN  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 2 time(s).
2017-12-14 11:53:26,311 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-12-14 11:53:26,312 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2017-12-14 11:53:26,315 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-12-14 11:53:26,316 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Democratic,231198)
grunt> Hence Democratic is the winner!!!!
```

Conclusion:

Thus, the data csv file is downloaded and analyzed the data using pig tool. The data is grouped by party name and the total vote count is summed and the maximum vote Party name is displayed with its total vote count.