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;

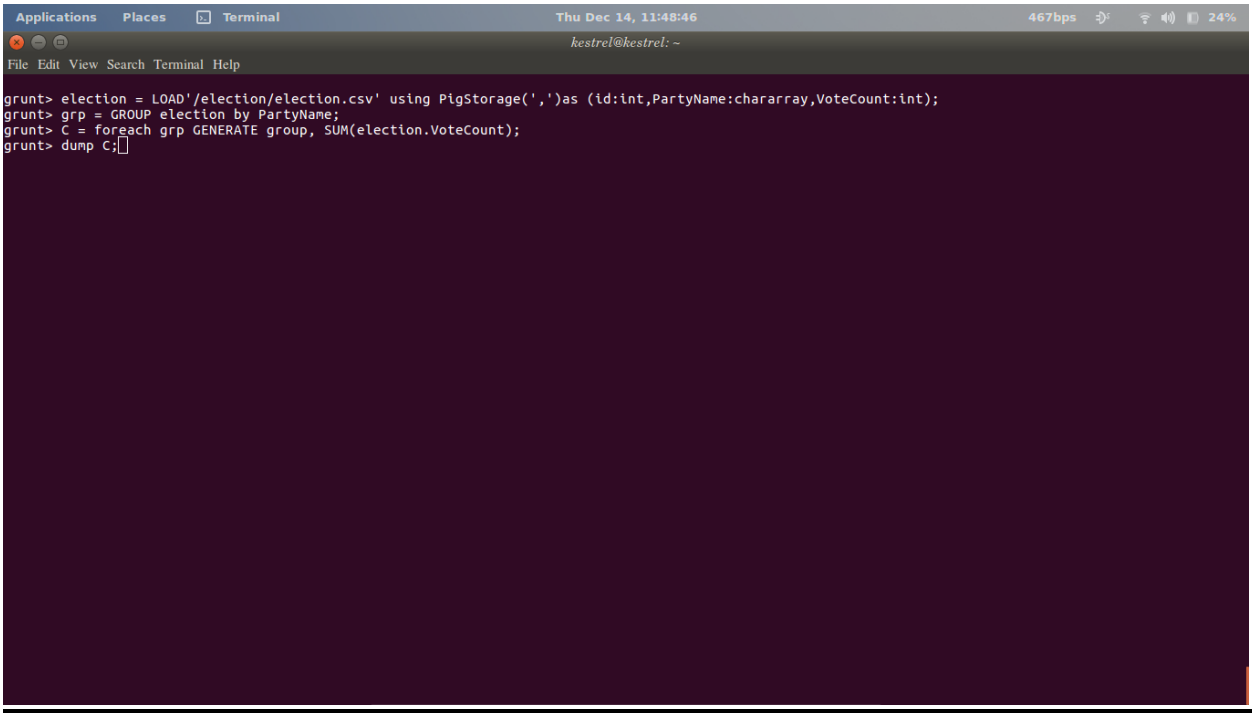
The result is, *(Democratic,231198)*

Hence Democratic is the winner.!

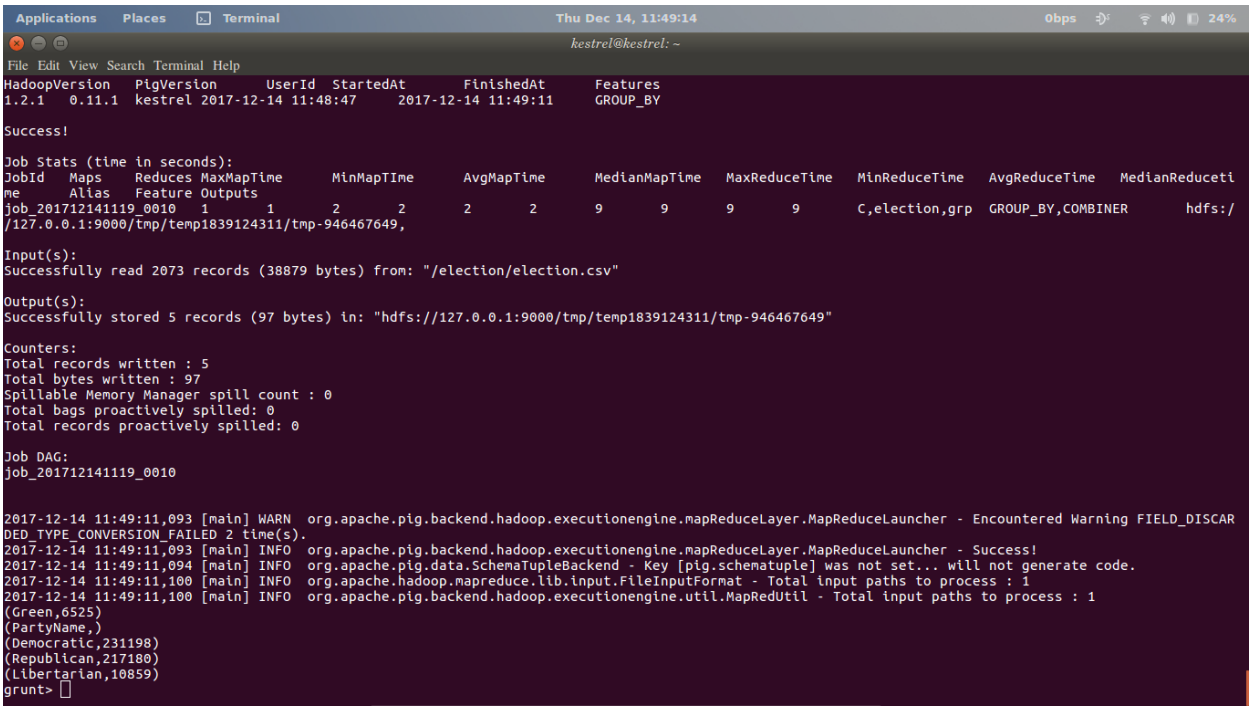
By using the above commands, the final output will be displayed as below the final screenshot.

Screenshots:

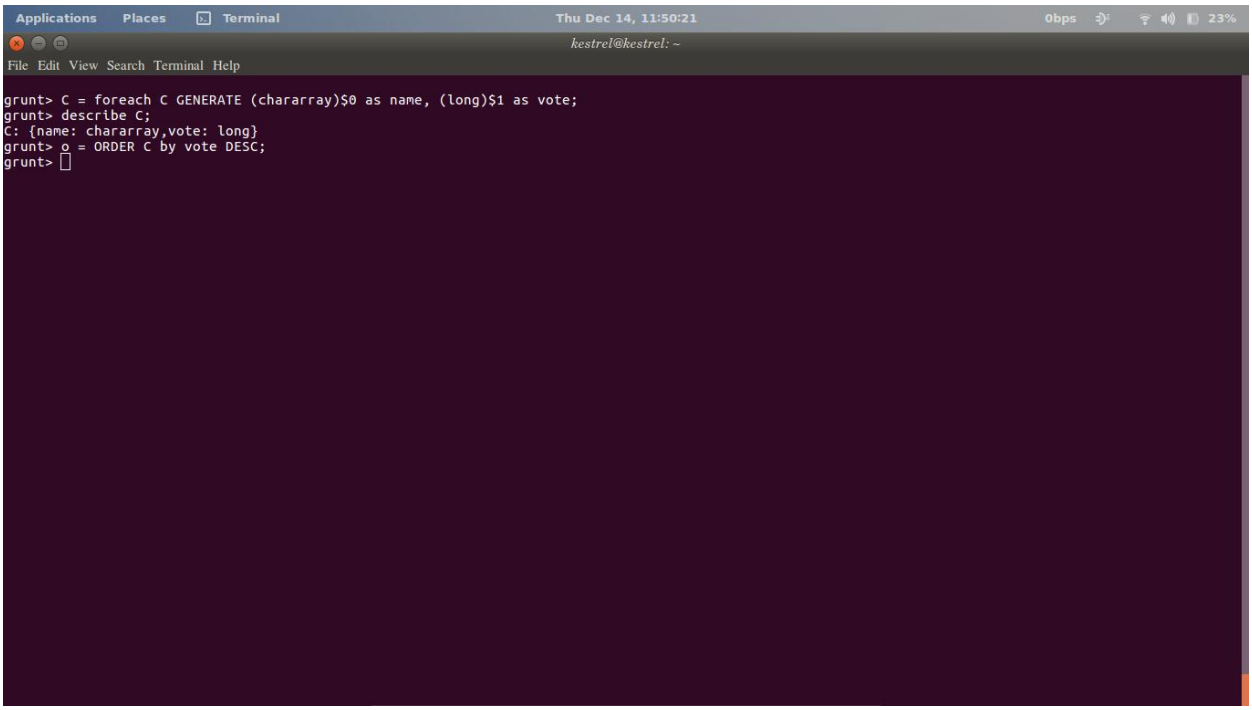
1. Loading file from HDFS to pig variable and grouping it using Party name:

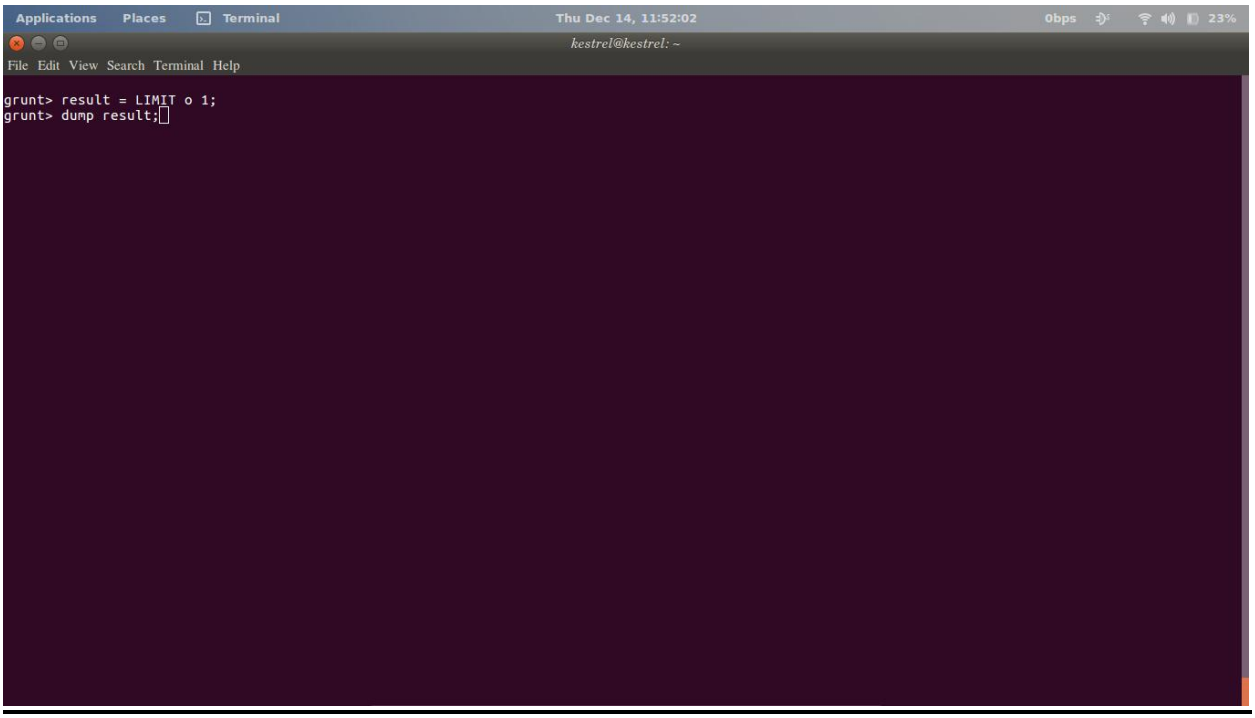


2. The dump Result for variable C:

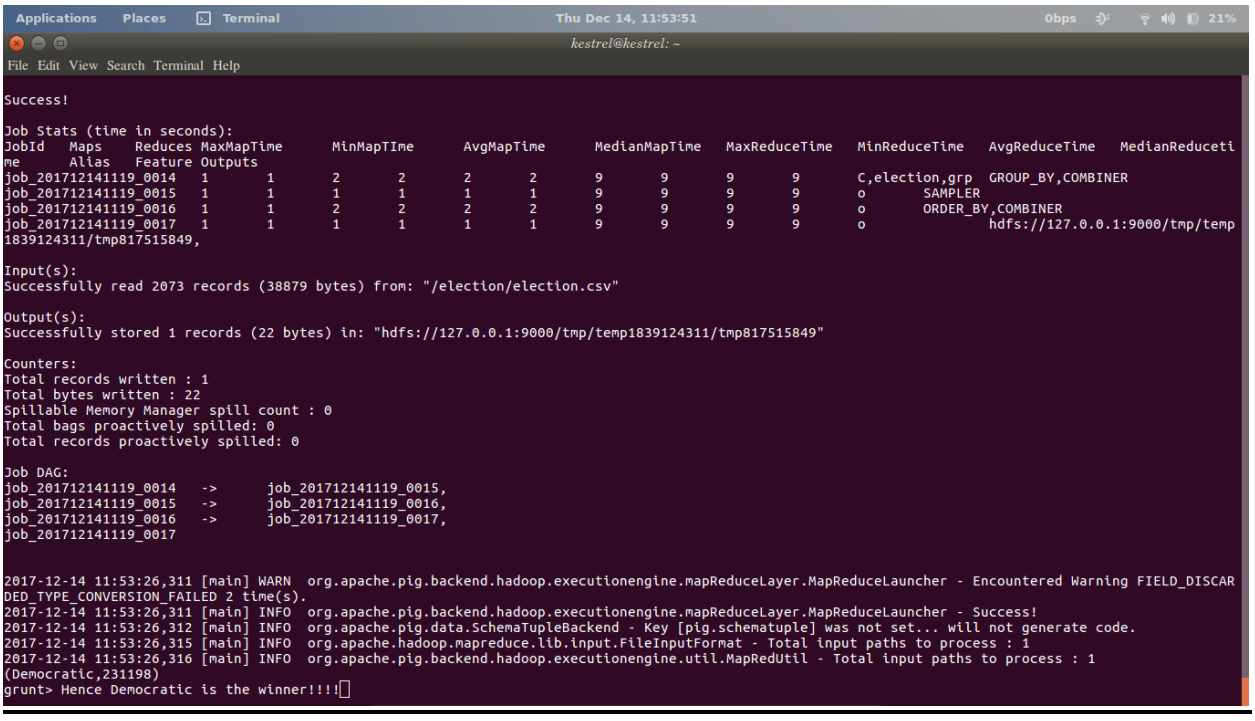


3. Ordering it by descending order:





4. The final Result is:



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.