CSE 587 - DATA INTENSIVE COMPUTING
HOMEWORK ASSIGNMENT 5
PROJECT REPORT

By- Krishnakant Chavali

50132680

kchavali@buffalo.edu

PROJECT DESCRIPTION

In this project we are trying to predict the sentiment of the people through the tweets they post about their favorite NBA teams. We analyze the tweets and based on the number of positive and negative responses we get, we determine the sentiment.

Approach

We perform the following steps in sequence to compute the sentiment:

- Execute the python file and generate csv files which contain the most recent tweets about the NBA teams
- Now using this data we differentiate the teams into two zones i.e. the eastern zone and the western zone
- Now we create two user profiles East and West and we set the visibility of both the users to only view the details or sentiments about the teams from their region
- We process the files and check for the words "win" and "lose" and count the number of times they are repeated
- The counts are written into a table and then from the table they are read and the teams are divided into wins and losses and sorted according to the count of the wins and losses
- The output is produced in the hadoop home directory

P.S: please read the readme section at the end of this file for instructions to access the output.

Western region:
lose:
Portland Trailblazers,TrailBlazers,4
Houston Rockets,Rockets,2
Denver Nuggets,Nuggets,2
L.A.Clippers,Clippers,1
Oklahoma City,okcthunder,0
New Orleans Hornets,Hornets,0
Memphis Grizzlies,Grizzlies,0
Golden State Warriors,GSWarriors,0
Minnesota Timberwolves,TWolves,0
Phoenix Suns,Suns,0
Dallas Mavericks,Mavs,0
LA Lakers,Lakers,0
Sacramento Kings,NBAKings,0
Utah Jazz,UtahJazz,0
San Antonio Spurs,GoSpursGo,0

win:
Golden State Warriors,GSWarriors,18
Houston Rockets,Rockets,3
Utah Jazz,UtahJazz,2
New Orleans Hornets,Hornets,1
Dallas Mavericks,Mavs,1
Sacramento Kings,NBAKings,1
Oklahoma City,okcthunder,0
Memphis Grizzlies,Grizzlies,0
Denver Nuggets,Nuggets,0
L.A.Clippers,Clippers,0
Portland Trailblazers,TrailBlazers,0
Minnesota Timberwolves,TWolves,0
Phoenix Suns,Suns,0
LA Lakers,Lakers,0
San Antonio Spurs,GoSpursGo,0

Eastern Region:
lose
Chicago Bulls,Bulls,3
Detroit Pistons,Pistons,1
New Jersey Nets,Nets,1
Miami Heat,MiamiHeat,0
Philadelphia 76ers,76ers,0
Milwaukee Bucks,Bucks,0
Indiana Pacers,Pacers,0
Atlanta Hawks,Hawks,0
Washington Wizards,Wizards,0
Toronto Raptors,Raptors,0
Boston Celtics,Celtics,0
New York Knicks,Knicks,0
Cleveland Cavaliers,Cavs,0
Orlando Magic,OrlandoMagic,0
Charlotte Bobcats,Bobcats,0

win
New York Knicks,Knicks,5
Atlanta Hawks,Hawks,2
Miami Heat,MiamiHeat,1
Detroit Pistons,Pistons,1
Chicago Bulls,Bulls,1
New Jersey Nets,Nets,1
Washington Wizards,Wizards,1
Cleveland Cavaliers,Cavs,1
Charlotte Bobcats,Bobcats,1
Philadelphia 76ers,76ers,0
Milwaukee Bucks,Bucks,0
Indiana Pacers,Pacers,0
Toronto Raptors,Raptors,0
Boston Celtics,Celtics,0
Orlando Magic,OrlandoMagic,0

READ ME:

How to run the code:

- Copy the data into the /input directory of the hadoop system using the code:

  hadoop fs -copyFromLocal /home/guest/PythonData/ /input

- copy the jar file into the folder /home/guest/cdse/accumulo/accumulo-1.4.2/lib/ext
- run the commands: The name of the table is NBA which can be changed in the commands as per convenience

  cd /home/guest/cdse/accumulo/accumulo-1.4.2/
  bin/tool.sh lib/ext/Final.jar Main acc guestvb /input NBA -u root -p acc

- To check for the output:

  in the Browzer go to the site localhost:50070

  click on the (browze the filesystem),

  The output folders are named Eastoutput and Westoutput which contain the respective outputs for the eastern and the western zones.

Conclusion

From the experimental results we got we can conclude that Accumulo is better than Hbase as we get the option to hide or restrict the access to certain pieces of information from users.