

**CSE 587**  
**Data Intensive Computing**  
**Homework 5**

**Popularity Index Estimation of NBA  
Teams using Twitter Data Analysis**

Vidyadhar Reddy Annapureddy  
vannapur@buffalo.edu  
(50134358)

### **Step1:** Generating and Cleaning Data

We first run the python script that uses twitter API to generate 100 most recent tweets for the 30 NBA teams of Eastern and Western Divisions. The tweets typically consist of newline characters which ought to be removed as part of cleaning the data. There are multiple ways to remove the unnecessary characters – *strip()* in python or UNIX commands such as *tr -d "\n\r" < Celtics.csv*. I have used the tool Notepad++ (v6.7.5 on windows 8.1) regex feature to remove the (\n, \r) characters from the tweets. The cleaned data is zipped in the submission archive.

### **Step2:** Copy data to Hadoop cluster

```
hadoop fs -mkdir /data
```

```
hadoop fs -ls /home/guest/Desktop/data/*.csv /data
```

### **Step3:** Create accumulo tables, users and give them adequate permissions

```
/home/guest/cdse/accumulo/accumulo1.4.2
```

- Login to accumulo shell using root credentials

```
accumulo shell -u root -p acc
```

- granting east and west visibility to the root user

```
setauths -s east,west -u root
```

- user configuration for eastern and western conference

```
createuser east  
choose password: east  
setauths -s east -u east  
createuser west  
choose password: west  
setauths -s west -u west
```

- configure table to store win count and lose count

```
createtable win  
grant Table.READ -t win -u east  
grant Table.READ -t win -u west
```

```
setiter -class org.apache.accumulo.core.iterators.user.SummingCombiner -p 10 -t win -majc -minc -scan
```

```
TRUE <enter> <enter> STRING
```

```
createtable lose  
grant Table.READ -t lose -u east  
grant Table.READ -t lose -u west
```

```
setiter -class org.apache.accumulo.core.iterators.user.SummingCombiner -p 10 -t lose -majc -minc -scan
```

TRUE <enter> <enter> STRING

- exit the accumulo shell

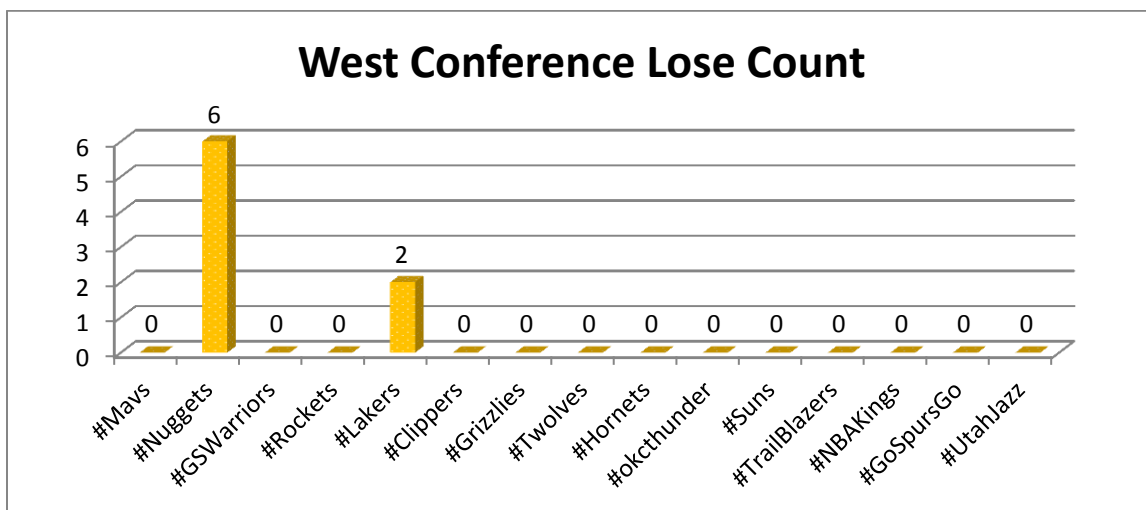
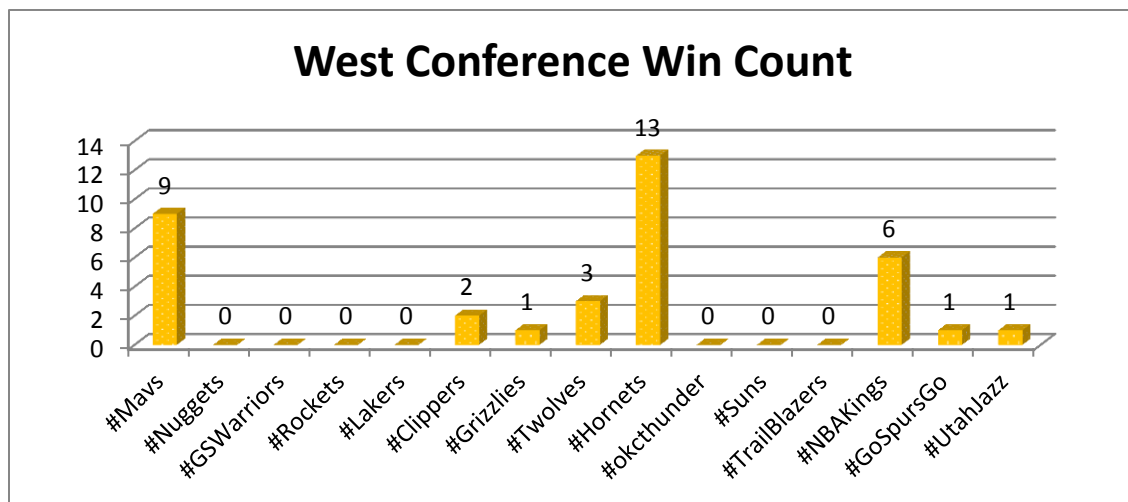
quit

**Step4:** Execute the jar file through tool.sh

```
bin/tool.sh /home/guest/Desktop/test.jar org.apache.accumulo.examples.simple.mapreduce.WordCount  
acc guestvb /data win -u root -p acc
```

### **OUTPUT & RESULTS:**

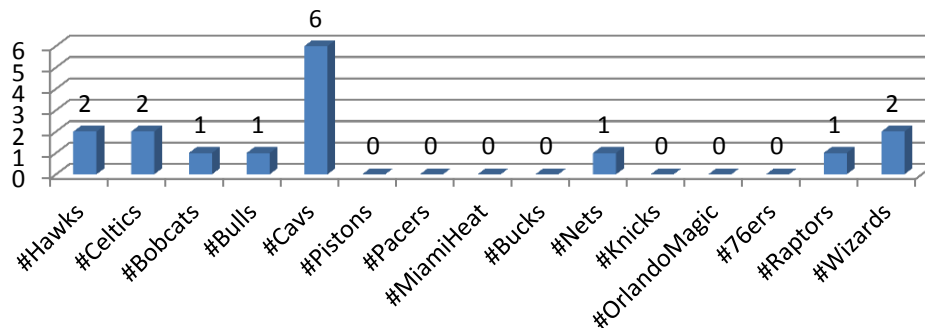
From the analysis, it can be concluded that teams with the highest popularity index in east and west conferences are Cleveland Cavaliers and New Orleans Hornets respectively. Subsequent results are given below:



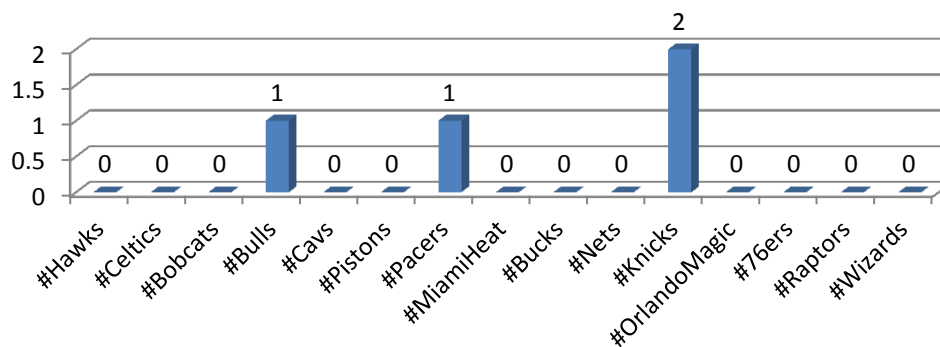
TeamName	HashTag	East/West Conf	WinCount
New Orleans Hornets	#Hornets	[west]	13
Dallas Mavericks	#Mavs	[west]	9
Sacramento Kings	#NBAKings	[west]	6
Minnesota Timberwolves	#Twolves	[west]	3
L.A.Clippers	#Clippers	[west]	2
Memphis Grizzlies	#Grizzlies	[west]	1
San Antonio Spurs	#GoSpursGo	[west]	1
Utah Jazz	#UtahJazz	[west]	1
Denver Nuggets	#Nuggets	[west]	0
Golden State Warriors	#GSWarriors	[west]	0
Houston Rockets	#Rockets	[west]	0
L A Lakers	#Lakers	[west]	0
Oklahoma City	#okcthunder	[west]	0
Phoenix Suns	#Suns	[west]	0
Portland Trailblazers	#TrailBlazers	[west]	0

TeamName	HashTag	East/West Conf	LoseCount
Denver Nuggets	#Nuggets	[west]	6
L A Lakers	#Lakers	[west]	2
New Orleans Hornets	#Hornets	[west]	0
Dallas Mavericks	#Mavs	[west]	0
Sacramento Kings	#NBAKings	[west]	0
Minnesota Timberwolves	#Twolves	[west]	0
L.A.Clippers	#Clippers	[west]	0
Memphis Grizzlies	#Grizzlies	[west]	0
San Antonio Spurs	#GoSpursGo	[west]	0
Utah Jazz	#UtahJazz	[west]	0
Golden State Warriors	#GSWarriors	[west]	0
Houston Rockets	#Rockets	[west]	0
Oklahoma City	#okcthunder	[west]	0
Phoenix Suns	#Suns	[west]	0
Portland Trailblazers	#TrailBlazers	[west]	0

## East Conference Win Count



## East Conference Lose Count



TeamName	HashTag	East/West Conf	WinCount
Cleveland Cavaliers	#Cavs	[east]	6
Atlanta Hawks	#Hawks	[east]	2
Boston Celtics	#Celtics	[east]	2
Washington Wizards	#Wizards	[east]	2
Charlotte Bobcats	#Bobcats	[east]	1
Chicago Bulls	#Bulls	[east]	1
New Jersey Nets	#Nets	[east]	1
Toronto Raptors	#Raptors	[east]	1
Detroit Pistons	#Pistons	[east]	0
Indiana Pacers	#Pacers	[east]	0
Miami Heat	#MiamiHeat	[east]	0
Milwaukee Bucks	#Bucks	[east]	0
New York Knicks	#Knicks	[east]	0
Orlando Magic	#OrlandoMagic	[east]	0
Philadelphia	#76ers	[east]	0

TeamName	HashTag	East/West Conf	LoseCount
New York Knicks	#Knicks	[east]	2
Chicago Bulls	#Bulls	[east]	1
Indiana Pacers	#Pacers	[east]	1
Cleveland Cavaliers	#Cavs	[east]	0
Atlanta Hawks	#Hawks	[east]	0
Boston Celtics	#Celtics	[east]	0
Washington Wizards	#Wizards	[east]	0
Charlotte Bobcats	#Bobcats	[east]	0
New Jersey Nets	#Nets	[east]	0
Toronto Raptors	#Raptors	[east]	0
Detroit Pistons	#Pistons	[east]	0
Miami Heat	#MiamiHeat	[east]	0
Milwaukee Bucks	#Bucks	[east]	0
Orlando Magic	#OrlandoMagic	[east]	0
Philadelphia	#76ers	[east]	0

- login to Accumulo shell to check results:

*accumulo shell -u root -p acc*

a) East Conference Win Count

*user east*

*table win*

*scan*

```

guest@guestvb: ~/cdse/accumulo/accumulo-1.4.2
root@acc win> user east
Enter password for user east: ****
east@acc win> scan
Atlanta Hawks #Hawks: [east]    2
Boston Celtics #Celtics: [east]  2
Charlotte Bobcats #Bobcats: [east] 1
Chicago Bulls #Bulls: [east]    1
Cleveland Cavaliers #Cavs: [east] 6
Detroit Pistons #Pistons: [east] 0
Indiana Pacers #Pacers: [east]   0
Miami Heat #MiamiHeat: [east]   0
Milwaukee Bucks #Bucks: [east]   0
New Jersey Nets #Nets: [east]    1
New York Knicks #Knicks: [east]  0
Orlando Magic #OrlandoMagic: [east] 0
Philadelphia 76ers #76ers: [east] 0
Toronto Raptors #Raptors: [east] 1
Washington Wizards #Wizards: [east] 2
east@acc win> 

```

b) East Conference Lose Count

*table lose*  
*scan*

```
guest@guestvb: ~/cdse/accumulo/accumulo-1.4.2
east@acc lose> scan
Atlanta Hawks #Hawks: [east]    0
Boston Celtics #Celtics: [east]  0
Charlotte Bobcats #Bobcats: [east] 0
Chicago Bulls #Bulls: [east]    1
Cleveland Cavaliers #Cavs: [east] 0
Detroit Pistons #Pistons: [east] 0
Indiana Pacers #Pacers: [east]   1
Miami Heat #MiamiHeat: [east]   0
Milwaukee Bucks #Bucks: [east]  0
New Jersey Nets #Nets: [east]   0
New York Knicks #Knicks: [east]  2
Orlando Magic #OrlandoMagic: [east] 0
Philadelphia 76ers #76ers: [east] 0
Toronto Raptors #Raptors: [east] 0
Washington Wizards #Wizards: [east] 0
east@acc lose> █
```

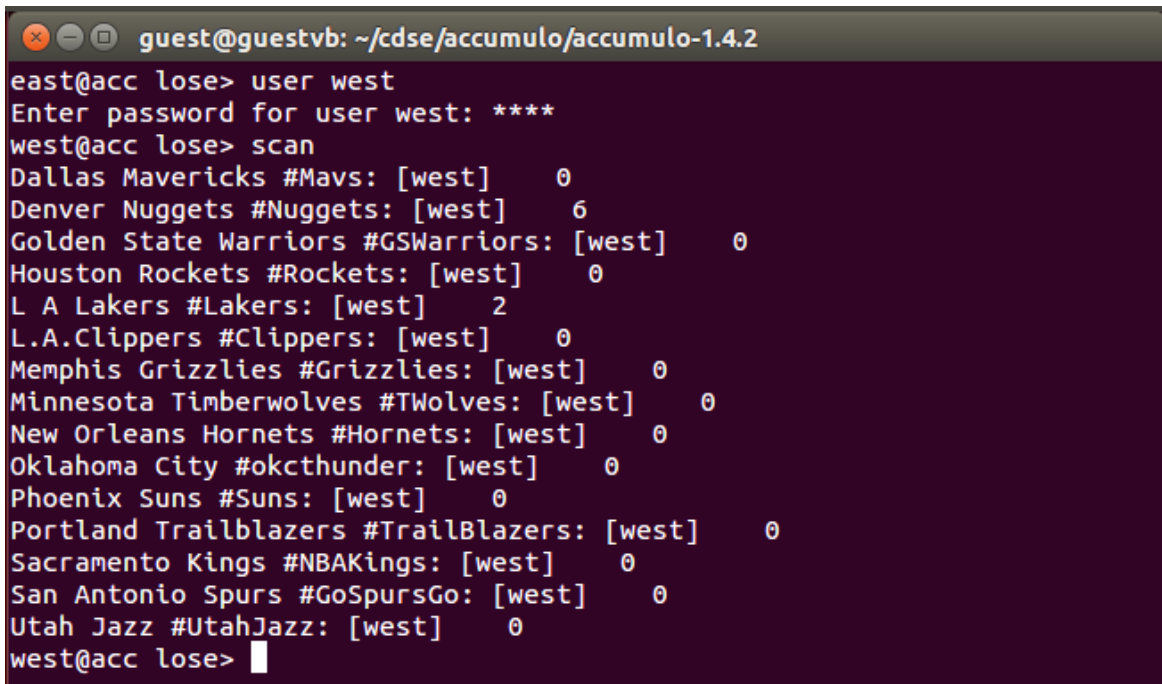
c) West Conference Win Count

*user west*  
*table win*  
*scan*

```
guest@guestvb: ~/cdse/accumulo/accumulo-1.4.2
west@acc win> scan
Dallas Mavericks #Mavs: [west]   9
Denver Nuggets #Nuggets: [west]  0
Golden State Warriors #GSWarriors: [west] 0
Houston Rockets #Rockets: [west] 0
L A Lakers #Lakers: [west]    0
L.A.Clippers #Clippers: [west]  2
Memphis Grizzlies #Grizzlies: [west] 1
Minnesota Timberwolves #TWolves: [west] 3
New Orleans Hornets #Hornets: [west] 13
Oklahoma City #okcthunder: [west] 0
Phoenix Suns #Suns: [west]    0
Portland Trailblazers #TrailBlazers: [west] 0
Sacramento Kings #NBAKings: [west] 6
San Antonio Spurs #GoSpursGo: [west] 1
Utah Jazz #UtahJazz: [west]    1
west@acc win> █
```

d) West Conference Lose Count

*table lose  
scan*



```
guest@guestvb: ~/cdse/accumulo/accumulo-1.4.2
east@acc lose> user west
Enter password for user west: ****
west@acc lose> scan
Dallas Mavericks #Mavs: [west]    0
Denver Nuggets #Nuggets: [west]   6
Golden State Warriors #GSWarriors: [west]  0
Houston Rockets #Rockets: [west]   0
L A Lakers #Lakers: [west]    2
L.A.Clippers #Clippers: [west]   0
Memphis Grizzlies #Grizzlies: [west]  0
Minnesota Timberwolves #TWolves: [west]  0
New Orleans Hornets #Hornets: [west]  0
Oklahoma City #okcthunder: [west]  0
Phoenix Suns #Suns: [west]    0
Portland Trailblazers #TrailBlazers: [west]  0
Sacramento Kings #NBAKings: [west]  0
San Antonio Spurs #GoSpursGo: [west]  0
Utah Jazz #UtahJazz: [west]   0
west@acc lose> 
```

e) Admin/Root observation:

Since the root user has both east and west authorization on win and lose tables, he can view all the 30 teams' win and lose count by scanning the table.

*user root  
table win  
scan  
table lose  
scan*



```
guest@guestvb: ~/cdse/accumulo/accumulo-1.4.2
root@acc> table win
root@acc win> scan
Atlanta Hawks #Hawks: [east]    2
Boston Celtics #Celtics: [east]  2
Charlotte Bobcats #Bobcats: [east]  1
Chicago Bulls #Bulls: [east]    1
Cleveland Cavaliers #Cavs: [east]  6
Dallas Mavericks #Mavs: [west]   9
Denver Nuggets #Nuggets: [west]  0
Detroit Pistons #Pistons: [east]  0
Golden State Warriors #GSWarriors: [west]  0
Houston Rockets #Rockets: [west]  0
Indiana Pacers #Pacers: [east]   0
L A Lakers #Lakers: [west]      0
L.A.Clippers #Clippers: [west]   2
Memphis Grizzlies #Grizzlies: [west]  1
Miami Heat #MiamiHeat: [east]    0
Milwaukee Bucks #Bucks: [east]   0
Minnesota Timberwolves #TWolves: [west]  3
New Jersey Nets #Nets: [east]    1
New Orleans Hornets #Hornets: [west] 13
New York Knicks #Knicks: [east]   0
Oklahoma City #okcthunder: [west]  0
Orlando Magic #OrlandoMagic: [east]  0
Philadelphia 76ers #76ers: [east]  0
Phoenix Suns #Suns: [west]       0
Portland Trailblazers #TrailBlazers: [west]  0
Sacramento Kings #NBAKings: [west]  6
San Antonio Spurs #GoSpursGo: [west]  1
Toronto Raptors #Raptors: [east]   1
Utah Jazz #UtahJazz: [west]       1
Washington Wizards #Wizards: [east]  2
root@acc win> 
```

```
guest@guestvb: ~/cdse/accumulo/accumulo-1.4.2
root@acc win> table lose
root@acc lose> scan
Atlanta Hawks #Hawks: [east] 0
Boston Celtics #Celtics: [east] 0
Charlotte Bobcats #Bobcats: [east] 0
Chicago Bulls #Bulls: [east] 1
Cleveland Cavaliers #Cavs: [east] 0
Dallas Mavericks #Mavs: [west] 0
Denver Nuggets #Nuggets: [west] 6
Detroit Pistons #Pistons: [east] 0
Golden State Warriors #GSWarriors: [west] 0
Houston Rockets #Rockets: [west] 0
Indiana Pacers #Pacers: [east] 1
L A Lakers #Lakers: [west] 2
L.A.Clippers #Clippers: [west] 0
Memphis Grizzlies #Grizzlies: [west] 0
Miami Heat #MiamiHeat: [east] 0
Milwaukee Bucks #Bucks: [east] 0
Minnesota Timberwolves #TWolves: [west] 0
New Jersey Nets #Nets: [east] 0
New Orleans Hornets #Hornets: [west] 0
New York Knicks #Knicks: [east] 2
Oklahoma City #okcthunder: [west] 0
Orlando Magic #OrlandoMagic: [east] 0
Philadelphia 76ers #76ers: [east] 0
Phoenix Suns #Suns: [west] 0
Portland Trailblazers #TrailBlazers: [west] 0
Sacramento Kings #NBAKings: [west] 0
San Antonio Spurs #GoSpursGo: [west] 0
Toronto Raptors #Raptors: [east] 0
Utah Jazz #UtahJazz: [west] 0
Washington Wizards #Wizards: [east] 0
root@acc lose> 
```

Apache Accumulo is a computer software project that developed a sorted, distributed key/value store. It extends the BigTable data model, adding a new element to the key called Column Visibility which stores a logical combination of security labels that must be satisfied at query time in order for the key and value to be returned as part of user-request. This allows data of varying security requirements to be stored in the same table, and allows users to see only those keys and values for which they are authorized. In the implementation, we have used "east" and "west" as the column visibility depending upon the tweet for rows inserted in "win" and "lose" tables authorization for words corresponding to 'Win' or 'win' or 'lose' or 'Lose'. We set appropriate authorizations for each user 'east' and 'west' to access only the permitted rows in each table. The root user, however, has access to entries corresponding to all 30 teams in the two tables.

The final result is then exported to a file using **exporttable** command, which **IS NOT a feature of accumulo 1.4.2** (version of accumulo provided in the VM image). An alternative to that is the **flush** command. For ease of purpose, the entries of win and lose tables are consolidated in a single csv file

included in the submission. Explicit screenshots depicting the results for the specific set of data are mentioned for reference in the report.