

CSE587 : Data Intensive Computing (Spring 2015)

Project 5 : Accumulo

Naveen Narayan

UB ID : nnarayan

UB Person # : 50134647

Aim :

The aim of this project is to use principles of Map Reduce to find the most popular and least popular for given NBA teams using Accumulo.

Introduction :

Apache Accumulo is a computer software project that developed a sorted, distributed key/value store based on the BigTable technology from Google. It is a system built on top of Apache Hadoop, Apache ZooKeeper, and Apache Thrift. Written in Java, Accumulo has cell-level access labels and server-side programming mechanisms.

Implementation :

Step 1 : Generating data files

1. Login to Twitter and create a twitter application to obtain the API keys and authentication tokens.
2. Generate data files (CSV files) for all the 30 teams using API_hashtag.py.

Step 2 : Loading Data files into Hadoop

1. Start the VM and login using the credentials given in the tutorial.
2. Change permission of the **update-script.sh** using: **chmod 775/home/guest/Desktop/update-script.sh**
3. Open terminal and run '**/home/guest/Desktop/update-script.sh**' to start Accumulo, Hadoop and ZooKeeper.
4. Create a directory in Hadoop – '**hadoop fs -mkdir /input**'
5. Copy the CSV data files into this directory.
hadoop fs -copyFromLocal /home/guest/Downloads/Python_script_and_CSV_files/*.csv /input

Step 3 : Creating Tables, Users and setting authorisations for the users in Accumulo

1. Login to the Accumulo shell
accumulo shell -u root -p acc
2. Set authorizations for the root user
setauths -s east,west -u root
3. Create 2 users, east and west, and set authorizations
createuser east
(enter password for this user)
createuser west
(enter password for this user)

setauths -s east -u east
setauths -s west -u west
4. Create a temp table, a table for Wins and a table for Losses

createtable temp
grant Table.READ -t temp -u west
grant Table.READ -t temp -u east
grant Table.READ -t temp -u west

createtable Win
setiter -class org.apache.accumulo.core.iterators.user.SummingCombiner -p 10 -t Win
-majc -minc -scan
grant Table.READ -t Win -u east
grant Table.READ -t Win -u west

createtable Lose
setiter -class org.apache.accumulo.core.iterators.user.SummingCombiner -p 10 -t
Lose -majc -minc -scan
grant Table.READ -t Lose -u east
grant Table.READ -t Lose -u west
5. Quit the accumulo shell.

Step 4 : Run the Jar file

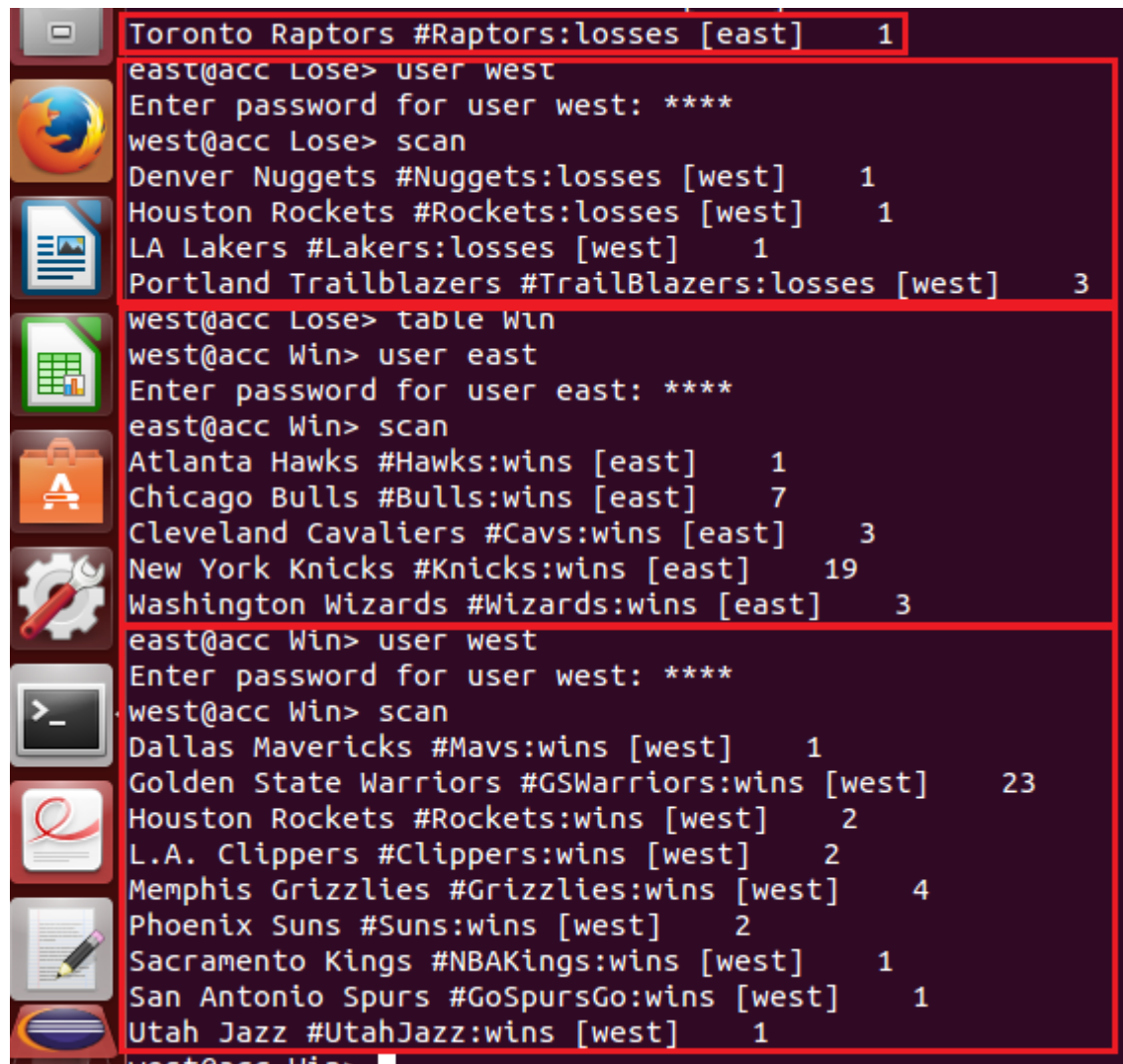
1. From /home/guest/cdse/accumulo/accumulo-1.4.2 enter the following command
bin/tool.sh /home/guest/Downloads/nbaPopularity.jar
org.apache.accumulo.examples.simple.mapreduce.WordCount acc guestvb /input
temp -u root -p acc

Step 5 : Login to the Accumulo shell and verify the results

```
root@acc> table Win
root@acc Win> scan
Atlanta Hawks #Hawks:wins [east]      1
Chicago Bulls #Bulls:wins [east]      7
Cleveland Cavaliers #Cavs:wins [east]  3
Dallas Mavericks #Mavs:wins [west]    1
Golden State Warriors #GSWarriors:wins [west] 23
Houston Rockets #Rockets:wins [west]  2
L.A. Clippers #Clippers:wins [west]   2
Memphis Grizzlies #Grizzlies:wins [west] 4
New York Knicks #Knicks:wins [east]   19
Phoenix Suns #Suns:wins [west]        2
Sacramento Kings #NBAKings:wins [west] 1
San Antonio Spurs #GoSpursGo:wins [west] 1
Utah Jazz #UtahJazz:wins [west]       1
Washington Wizards #Wizards:wins [east] 3

root@acc Win> table Lose
root@acc Lose> scan
Atlanta Hawks #Hawks:losses [east]    2
Boston Celtics #Celtics:losses [east] 2
Chicago Bulls #Bulls:losses [east]    1
Denver Nuggets #Nuggets:losses [west] 1
Houston Rockets #Rockets:losses [west] 1
LA Lakers #Lakers:losses [west]       1
New York Knicks #Knicks:losses [east]  3
Portland Trailblazers #TrailBlazers:losses [west] 3
Toronto Raptors #Raptors:losses [east] 1

root@acc Lose> user east
Enter password for user east: ****
east@acc Lose> scan
Atlanta Hawks #Hawks:losses [east]    2
Boston Celtics #Celtics:losses [east]  2
Chicago Bulls #Bulls:losses [east]    1
New York Knicks #Knicks:losses [east]  3
```

A terminal window with a dark background and red text. On the left side, there is a vertical column of application icons: a folder, Firefox, a document, a spreadsheet, a shopping bag, a gear, a terminal, a red 'Q' logo, a document with a pencil, and a blue oval. The terminal text shows a sequence of commands and outputs for a database application.

```
Toronto Raptors #Raptors:losses [east] 1
east@acc Lose> user west
Enter password for user west: ****
west@acc Lose> scan
Denver Nuggets #Nuggets:losses [west] 1
Houston Rockets #Rockets:losses [west] 1
LA Lakers #Lakers:losses [west] 1
Portland Trailblazers #TrailBlazers:losses [west] 3
west@acc Lose> table Win
west@acc Win> user east
Enter password for user east: ****
east@acc Win> scan
Atlanta Hawks #Hawks:wins [east] 1
Chicago Bulls #Bulls:wins [east] 7
Cleveland Cavaliers #Cavs:wins [east] 3
New York Knicks #Knicks:wins [east] 19
Washington Wizards #Wizards:wins [east] 3
east@acc Win> user west
Enter password for user west: ****
west@acc Win> scan
Dallas Mavericks #Mavs:wins [west] 1
Golden State Warriors #GSWarriors:wins [west] 23
Houston Rockets #Rockets:wins [west] 2
L.A. Clippers #Clippers:wins [west] 2
Memphis Grizzlies #Grizzlies:wins [west] 4
Phoenix Suns #Suns:wins [west] 2
Sacramento Kings #NBAKings:wins [west] 1
San Antonio Spurs #GoSpursGo:wins [west] 1
Utah Jazz #UtahJazz:wins [west] 1
```

References :

http://en.wikipedia.org/wiki/Apache_Accumulo