

CSE 587 HW # 5

Name: Daniel Nazareth

UbitName: dnazaret

Email: dnazaret@buffalo.edu

SUMMARY: The project uses the Accumulo database layered on Hadoop to gain an understanding of the popularity of NBA(National Basketball Association) teams. We use a pre-provided VM with all necessary software installed, mine for tweets relating to the NBA teams using a provided python script, then store the resulting data into csv files and view popularity in terms of the 'win' and 'lose' tweets for each team using accumulo

RATIONALE FOR USING ACCUMULO: The biggest advantage of Accumulo is its strong emphasis on security and visibility which is not seen in other rival BigTable based databases. Thus in this project we create two users east and west for each conference and can see in real-time how accumulo does not allow one user to see the data of the other.

To summarise, Accumulo offer the following benefits:

- Highly secure data, only visible to certain users
- Easy ability to manipulate data as well as permissions and views of that data.
- Highly scalable way of storing very sparse data

RESULTS: The results are summarised in tabular form below for each of the conference. This can be verified by running the attached code in the submission as given in the README file attached to this submission

EAST RESULTS

Username: east

East Conference(15 Teams)-WIN Tweets	
Charlotte Bobcats	4
Philadelphia 76ers	2
Washington Wizards	2
Cleveland Cavaliers	1
Indiana Pacers	1

Orlando Magic	1
Milwaukee Bucks	1

East Conference(15 Teams)-LOSE Tweets	
Detroit Pistons	3
Indiana Pacers	2
Boston Celtics	1

Username:west

West Conference(15 Teams)-WIN Tweets	
LA Clippers	5
Golden State Warriors	4
San Antonio Spurs	2
Houston Rockets	2
Phoenix Suns	1
Sacramento Kings	1
Utah JAZZ	1
Minnesota Timberwolves	1

West Conference(15 Teams)-LOSE Tweets	
Portland TrailBlazers	6
LA Clippers	3
Golden State Warriors	1

Denver Nuggets	1
Minnesota Timberwolves	1

NOTES/ISSUES FACED:

- Any team not mentioned in either of the results for any conference had ZERO tweets related to it, hence omitted for brevity
- 100 tweets is very small number to examine for specific words like win/lose, hence the small numbers obtained.

REFERENCES

<https://accumulo.apache.org/>