BIG DATA ANALYTICS OF GEOSOCIAL MEDIA FOR PLANNING AND REAL TIME DECISIONS

[1]Keerti Kharatmol, [2]Akshay Sadalgekar, [3]Shubham Verma, [4]Pranit Prabhakaran

[1]Assistan Professor, [2]Student, [3]Studnet, [4]Student
[1]Computer Engineering,
[1]K C College of Engineering and Management Studies and Research, Thane, India

Abstract : Geosocial network data provides information on latest trends in human, their day to day life and incidents all around humans. Hence, Geosocial Network data can be served as an asset to make real-time decisions and future planning by analyzing geosocial media posts. In this paper, we proposed an efficient system for Geosocial Networking and location information of users. A system is proposed to analyze huge amount of data from users' trend and geological events. However, there are millions of geosocial networks who generate terabytes of heterogeneous data with a variety of information day with high speed, termed as Big Data. System architecture is proposed that is capable of processing large amount of data collected from social networks to analyze and monitor Geosocial events to make decisions and draw up a plan for future. The system site Spark at the top for speeding up Hadoop computational computing process and for real time analysis. In this system we analyze Twitter for understanding and identify geosocial events. The system is capable of analyzing huge amount of real time geosocial data.

***IndexTerms* - Component,formatting,style,styling,insert.**

I. Introduction

Social media is rapidly advancing day by day changing from social networks to Geosocial Networks. It empowers users to make their posts public along with their location. This has risen in increase in Geosocial Networks allowing users to report events, share views while connecting to whole world. The information shared by users is Geosocial because user explicitly sharing a post about event occurred and the views shared on social media platforms reveals worlds' social knowledge and strengthen communication.

Advances are technology uses the GPS systems to detect location of the people. The people posting and sharing location is tracked and recorded. Thus, a data warehouse is produced by social networks. Social network data is beneficial for many fields. Based on filtering the data and well analyzing can help authorities for better city development plans.

Using geosocial data, it is not only beneficial to government but also have impact on human life. Advances are technology uses the GPS systems to detect location of the people. The people posting and sharing location is tracked and recorded.

II. Objective

The main objective of the project is to develop an efficient system for exploring Geosocial Networks while harvesting data as well as user's location information. System architecture is proposed that processes an abundant amount of various social networks' data to monitor Earth events, incidents, medical diseases, user trends, and views to make future real-time decisions and facilitate future planning.

III. Existing System

In most of the existing system, the systems have used Natural Language Processing (NLP) on social media data. It is applied for text mining particularly for Sentiment analysis, because of which Consumes large time to process when data size is in billions.

Also, in the previous systems time series data model are not supported. We have observed that most of the data collected during any natural calamity is available in huge sizes, storing and analyzing which is not an easy task the system does not support huge inflow of data. It consumes a large amount of time to perform operations on higher set of objects.

IV. Problem Statement

We have observed that most of the data collected during any natural calamity is available in huge sizes, storing and analyzing which is not an easy task. For such a condition we propose our system which has the capability of not only handling such large amount of data but also process it within seconds. This system explores Geosocial Networks while harvesting data as well as user's location information.

V. PROPOSED SYSTEM

The proposed system consists of five layers, data collection, data processing, application, communication, and data storage. The system has three basic layers, i.e., Data Collection, Data Processing, and Services and Application. Two additional layers that works with the basic layers provide communication and storage for raw data and structured data. The communication layer provides internal communication between servers through various communication technologies, such as Wi-Fi as well as external communication to the Geosocial Network servers for data harvesting using any fast Internet technology, such as 3G, and LTE.

Networks, such as Facebook, YouTube and Twitter can be used for the data collection. In this system data is collected from the twitter. Data collection layer is also responsible for the data harvesting server that obtains publicly available data from Geosocial Network servers.

Data harvesting is done on captured data and harvest that data at high speed. The data is harvested by a query method sent as an HTTP request. The query response is either JSON or XML format. Hadoop provides the Tweet information and with the help of decision server the information can be manipulated.

For Geosocial Network data analysis plays an important role with location and time and the content. Location of tweet indicated the area of the event. In Twitter, tweet posting time and date is attached with tweets as metadata.

Data processing layer is important for all type of processing and result generation. At preprocessing layer classification of data is done to produce structured data.

With the help of Flume real time data collected is stored into Hadoop’s HDFS. The decision server determines of the geosocial activity and these results can be used for town planning, citizens safety.[1]

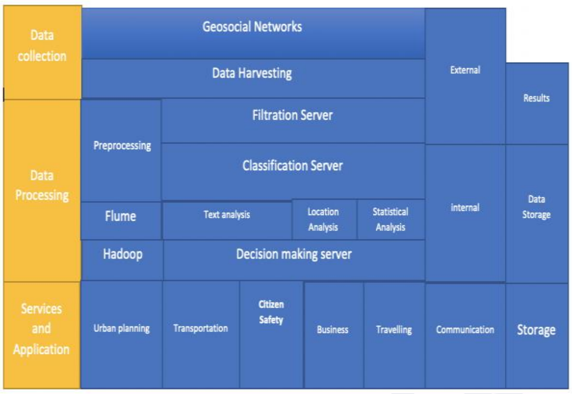
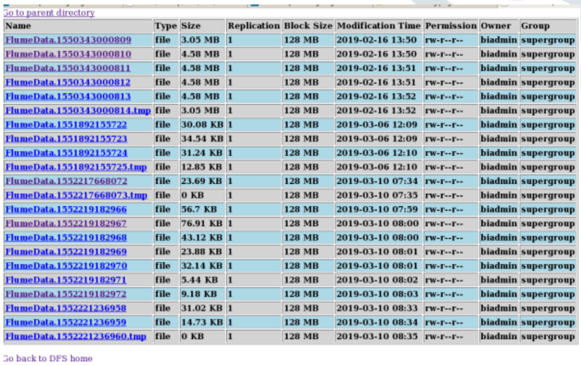


Fig. 1. Proposed System

VI. ANALYSIS AND DISCUSSION

In this section the data set details are given and discussion on the data analysis. We obtained the data from Twitter which contains tweets. The data is classified with the hashtag Earthquake. Analyses are conducted on twitter data. We analyze tweets considering event in various Earth regions such as Earthquake.



Go to parent directory

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
FlumeData.1550343000809	file	3.05 MB	1	128 MB	2019-02-16 13:50	rw-r--r--	biadmin	supergroup
FlumeData.1550343000810	file	4.58 MB	1	128 MB	2019-02-16 13:50	rw-r--r--	biadmin	supergroup
FlumeData.1550343000811	file	4.58 MB	1	128 MB	2019-02-16 13:51	rw-r--r--	biadmin	supergroup
FlumeData.1550343000812	file	4.58 MB	1	128 MB	2019-02-16 13:51	rw-r--r--	biadmin	supergroup
FlumeData.1550343000813	file	4.58 MB	1	128 MB	2019-02-16 13:52	rw-r--r--	biadmin	supergroup
FlumeData.1550343000814.tmp	file	3.05 MB	1	128 MB	2019-02-16 13:52	rw-r--r--	biadmin	supergroup
FlumeData.1551892155722	file	30.08 KB	1	128 MB	2019-03-06 12:09	rw-r--r--	biadmin	supergroup
FlumeData.1551892155723	file	34.54 KB	1	128 MB	2019-03-06 12:09	rw-r--r--	biadmin	supergroup
FlumeData.1551892155724	file	31.24 KB	1	128 MB	2019-03-06 12:10	rw-r--r--	biadmin	supergroup
FlumeData.1551892155725.tmp	file	12.85 KB	1	128 MB	2019-03-06 12:10	rw-r--r--	biadmin	supergroup
FlumeData.1552217668072	file	23.69 KB	1	128 MB	2019-03-10 07:34	rw-r--r--	biadmin	supergroup
FlumeData.1552217668073.tmp	file	0 KB	1	128 MB	2019-03-10 07:35	rw-r--r--	biadmin	supergroup
FlumeData.1552219182966	file	56.7 KB	1	128 MB	2019-03-10 07:59	rw-r--r--	biadmin	supergroup
FlumeData.1552219182967	file	76.91 KB	1	128 MB	2019-03-10 08:00	rw-r--r--	biadmin	supergroup
FlumeData.1552219182968	file	43.12 KB	1	128 MB	2019-03-10 08:00	rw-r--r--	biadmin	supergroup
FlumeData.1552219182969	file	23.88 KB	1	128 MB	2019-03-10 08:01	rw-r--r--	biadmin	supergroup
FlumeData.1552219182970	file	32.14 KB	1	128 MB	2019-03-10 08:01	rw-r--r--	biadmin	supergroup
FlumeData.1552219182971	file	5.44 KB	1	128 MB	2019-03-10 08:02	rw-r--r--	biadmin	supergroup
FlumeData.1552219182972	file	9.18 KB	1	128 MB	2019-03-10 08:03	rw-r--r--	biadmin	supergroup
FlumeData.1552221236958	file	31.02 KB	1	128 MB	2019-03-10 08:33	rw-r--r--	biadmin	supergroup
FlumeData.1552221236959	file	14.73 KB	1	128 MB	2019-03-10 08:34	rw-r--r--	biadmin	supergroup
FlumeData.1552221236960.tmp	file	0 KB	1	128 MB	2019-03-10 08:35	rw-r--r--	biadmin	supergroup

Go back to DFS home

Fig. 2.Data extracted from twitter

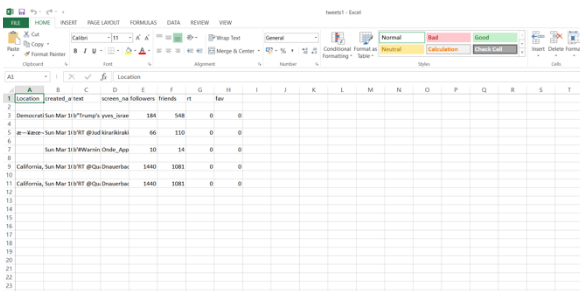


Fig. 3. Structured format of data

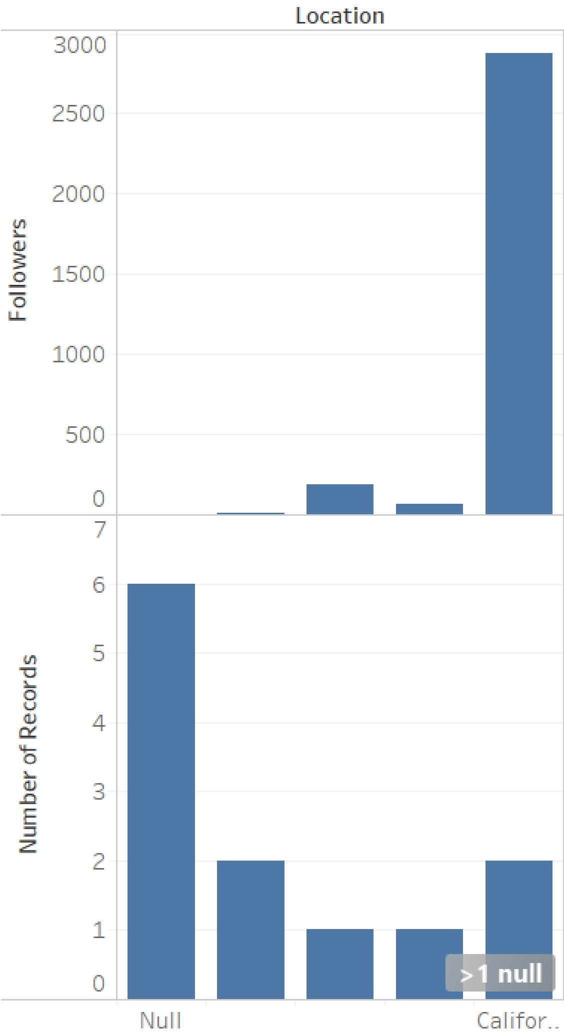


Fig. 4.Bar chart representation of tweets

VII. Conclusion

Advance technology uses the GPS systems to detect location of the people. The people posting and sharing location is tracked and recorded. The system is proposed for real time analysis which can process the data within seconds of time and can provide with high alert scenario. Such network can benefit to government body as well public sector. In this paper we developed a system for Geo social for a large amount of data at very High-speed. This system can process huge collection of data and analyze to make real time decision. This system is developed with help of Hadoop with flume.

References

[1] Big Data Analytics of Geosocial Media for Planning and Real-Time Decisions by M. Mazhar Rathore, Anand Paul, Awais Ahmad The School of Computer Science and Engineering Kyungpook National University, Daegu, Korea Department of Information and Communication Engineering Yeungnam University, Gyeongbuk, Republic of Korea Muhammad Imran3, Mohsen Guizani4 3College of Computer and Information Sciences,King Saud University, Saudi Arabia Dept. of Electrical and Computer Engineering University of Idaho, USA

[2] Big Data and Internet of Things: An Asset for Urban Planning Conference Paper · October 2015 DOI: 10.1145/2837060.2837067 Conference: Conference: The BigDas, At Jeju Island, Korea, Volume: 2015

[3] M. Zook, M. Graham, T. Shelton, and S. Gorman, "Volunteered geographic information and crowdsourcing disaster relief: A case study of the Haitian earthquake," World Medical & Health Policy, vol. 2, no. 2, pp. 7–33, 2010.