Geosocial Processing using Big Data Analytics

Mrs. Mamatha A[1] Sourabh Jha[2] Ravi Prakash Verma[3] Rishi Kumar[4] Tushar Sahay[5]
[1]Assistant Professor
[1,2,3,4,5]Sapthagiri College of Engineering, Hessarghatta Main Road, Bangalore, Karnataka, India

Abstract— Geosocial Network information can be filled in as preference for the authorities to settle on consistent choices and future organizing by investigating geosocial media posts. In any case, there are innumerable Network clients who are passing on overpowering of information, called "Colossal Data" that is endeavoring to be assessed and settle on consistent choices. Thusly, in this paper, we proposed a beneficial structure for investigating Geosocial Networks while getting information and client's domain data. A framework building is recommended that strategies a plenteous proportion of different easy going gatherings' information to screen Earth occasions, scenes, helpful ailments, client precedents, and perspectives to settle on future determined choices and bolster future planning. The proposed structure contains five layers, i.e., information conglomeration, information arranging, application, correspondence, and information putting away. The Hadoop natural gathering with the end goal to run steady examinations. Twitter are bankrupt down utilizing the proposed arrangement recalling the genuine goal to perceive recent developments or fiascos, for example, seismic tremors, fires, Ebola infection, and snow. The structure is overviewed concerning ampleness while thinking about framework throughput. We showed that the framework has higher throughput and is set up for isolating massive Geosocial Network information at enduring.

Keywords: Geosocial Network; Big Data; Hadoop; Spark

I. Introduction

Geo-Social is in a general sense driving their part all around requested while making themselves from easy-going relationship to Geosocial Networks. It attracts people to make their substance open together with their geographical information. This has understood a move inside the utilization of Geosocial Networks by equipping clients with the adaptability to voice inclines, report occasions, and offer perspectives, stun, or adore while band together with others, which was awesome inside the pre-Internet age. The information the taking in the information are partaken in any media of Geosocial in light of the way that:

1) The presents have top on base substance that has a tendency to arrive information with specific territories that square measure either entered unequivocally (with selection) or superimposed obviously (by Earth empowers, similar to expansion or height).
2) The perspectives are shared by electronic casual correspondence uncover social data and reinforce relationship and correspondence.

Advances being produced have permitted the use of GPS structures in sensible telephones that made zone data a considerable measure of winning. The course of action of people posting, remarking, or trading photographs on easygoing gatherings that is recorded. In this way, by conglomerating such style of zone data from all that the structure clients and social affiliations fabricate dissemination focal points of geo-social data. Another strategy for conveying geo-social data is by swarm sourcing however surrendering self-manufactured an applications construed for various points of confinement or causes. Geosocial data from volunteers or paid clients Who offer data or information subsequently. For instance, all through the Haiyan sea tempest inside the Philippines in 2013, a virtual social affair made out of an outsized gathering of supporters, volunteers, and IT experts made on-line guides for crisis help. Such style of on-line information assembled by swarm sourcing was given the name "volunteer geographic data" (VGI). Today, a few phase and bundle are made to utilize swarm sourcing for geosocial data gathering with particular true objective to develop business, drive causes, or for elective business limits. Relate in nursing occurrence of 1 such bundle arrange is Ushahidi, that stipends report creation by arranging a picked catchphrase on a Geosocial Network that ponders to fluctuated particular domains. The report will than utilized for social care and stimulate inside the occurrence of emergencies or catastrophes. Right now, analysts square measure a great deal of fascinated by Geosocial Networks because of they take a gander at them new data resources. Punks et al. Utilized Twitter data (geo set tweets) inside their work to delineate in the us. Among 21,362 geo set tweets, the essential tweet that revealed Associate in Nursing seismic tremor seemed, by all accounts, to be one moment when the shake. Essentially, Chow and Papadimitriou engineered structures abuse geo stamped proposition while influencing a substitution social to deal with. GeoLife2.0 performed equality ID, flight examination, and suggestion decision. There square measure couple of elective structures that play out an examination for that substance of tweets, e.g., some utilization areas. Some examination has furthermore been performed to discover hotspots and hyper-near to occasions in a to an incredible degree town mishandle intensifier framework data.

Twitter, Flickr, Face book, YouTube, and so forth. The structure passes on the Hadoop framework for arranging and examination with Spark and not any more astounding it as an outcast contraption for day and age examination. we have an inclination to endeavored the framework by taking easygoing affiliations i.e.. Twitter whatever is left of the record delineates the engineered structure, in like manner the data examination, utilize, and examination.

Social network data could be beneficial for many fields if well analyzed. By analyzing the social behavior of a community in a particular area by filtering and profile matching, one can recommend people a shop, hotels, cheap markets, banking systems, advertisements, etc., based on their likes and constraints [16]. Similarly, based on people and vehicles movement, the authorities can perform better city plans [17-19] and recommend suitable traffic routes to people based on current circumstances [20, 21]. In addition, Social network data analysis, such as Twitter is used in many healthcare applications [22-24], to monitor and control fatal diseases and infections. There are several other works that have been performed using Geosocial Networks that employ limited geolocated information. However, all of the existing systems do not consider the processing aspects of real-time, high-speed geosocial traffic. There are few other high-speed big data processing system aimed at establishing healthcare system [25], processing M2M data using data fusion model [26], establishing smart city [27], etc. On the other hand, These system are not suitable to process social network data because of its geosocial nature. Using geosocial network data is not only beneficial to governments, but it can also have a major impact on human life. Geosocial Network data can provide benefits to normal citizens and business people. However, when harvesting geosocial data from networks such as Twitter or Facebook, it should be noted that these networks have millions of users who post thousands of tweets and statuses with an hour. Therefore, it can be easily reasoned that all the users of various social networks generate a significant amount of data: such data might range in the terabytes within minutes. Consequently, harvesting such real-time geo-social data is a very challenging task. We need a special computational environment and advanced computing techniques with intelligent management in order to provide in-time/real-time analysis. All the aforementioned techniques do not consider more than one social network at a time, and their analyses are scalable in terms of data size. Therefore, in order to meet these computational challenges, in this paper, we propose an advanced geosocial data analytical system that not only processes offline data efficiently within a time limit but also provides real-time data analysis for various social networks, including Twitter, Flickr, Facebook, YouTube, etc. The system deploys a Hadoop ecosystem for data processing and analysis with Spark at the top it as a third-party tool for real-time analysis. We tested the system by taking two social networks, Twitter and Flickr. The rest of the document describes the proposed system, including data analysis, implementation, and evaluation.

II. PROPOSED COMPUTING MODEL

This section describes the proposed model, including system overview, proposed architecture, application, and limitations.

A. *System Overview*

Humans are the most reliable source for reporting events, activities, and important issues. Geosocial Networks use humans as sensors for monitoring activities worldwide. When a user posts any activity that is related to some events, the user acts as a sensor that sends data to its station, i.e., the Geosocial Network that performs analysis on the data to find information about what is occurring in the world. Figure 1 shows the use of human as a sensor in the proposed system to generate and harvest data in order to monitor various Earth areas. The proposed system analyzes all the tweets generated by Twitter users, in addition to Facebook statuses, filter images, and information from other Geosocial Networks, such as YouTube and Foursquare. All such information corresponds to user locations and can provide real-time monitoring of disasters, fatal diseases, or accidents. Moreover, user data and their location information can be used to recommend various systems based on the user's current location, such as useful products, restaurants, hotels, and transportation.

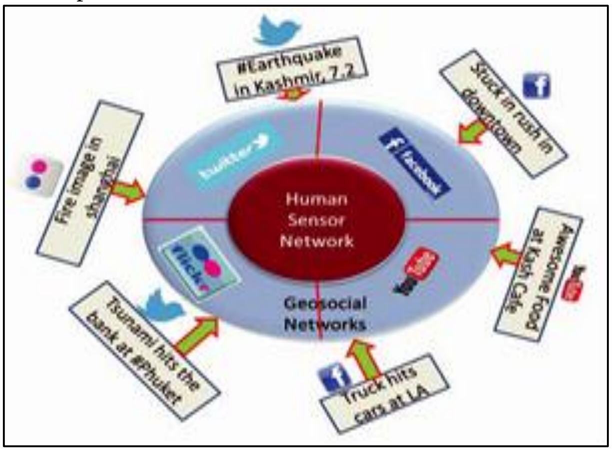

Fig. 1: Humans working as sensors in Geosocial Networks

Our system harvests user data from Geosocial Networks as a tuple (l, t, u) of location l, time t, and user u. The tuple describes, what user u posted on a Geosocial Network at time t from location l. Obtaining user location is also a challenging task that has become easier because of advanced remote smart devices and GPS systems. Smart devices send location coordinates by default to the network while the user posts any activity or event. The IP-based technique uses the "at" tag of Facebook; check-in mechanism and friend tags are also used to determine user location in case the location information is missing with the post. The location indicates the particular area in order to determine the events or activities that are occurring at different time variants based on the Geosocial Network being used. Such location is attached with the post, and it is used to find the authenticity and relevance of the activity or event that corresponds to the location. The posts, texts, tweets, comments, statuses, and smiley are analyzed using text analytics, statistical analytics, complex machine learning, and data mining techniques in order to monitor and determine what is occurring, where, and why. Such type of data can be used to predict future events based on the current user trends that correspond to various areas. Therefore, businesspeople, entrepreneurs, government agencies, and citizens can make plans based on current trends. However, the major challenge is how to provide analysis for such a huge amount of data generated by many social networks. The proposed system uses advanced computing technologies to meet the challenge. A detailed description of the proposed system, including the proposed architecture, implementation model, application, issues, and challenges are given in the following sections.

B. *Proposed System Architecture*

Figure 2 shows the system architecture. The system has three basic top to bottom layers, i.e., Data Collection, Data Processing, and Services and Application. Two additional layers work side by side with the basic layers. Such additional layers provide communication and storage for raw and structured data. The communication layer provides internal communication between various servers through different communication technologies, such as Wi-Fi and Ethernet, as well as external communication to the Geosocial Network servers for data harvesting using any fast Internet technology, such as WiMAX, 3G, and LTE. The storage layer manages services for storing data that can be structured, unstructured, or raw for future use or planning. Instead of working with all three basic layers, as is done by the communication layer, the storage layer only works side by side with the last two basic layers, i.e., data processing and services and application. The data are stored in the database after classification, and the results are stored after analysis and decision-making. Because the data sources are Geosocial.

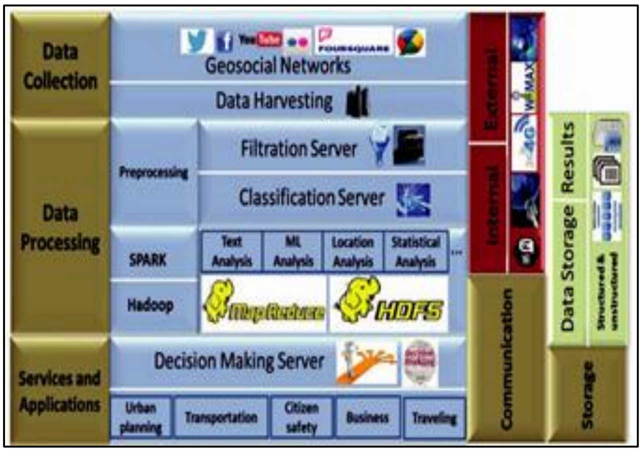

Fig. 2: Proposed system architecture

Networks, such as Twitter, Facebook, YouTube, Flickr, and Foursquare, the data collection layer is responsible for the data harvesting server that obtains data or content from Geosocial Network servers. (Recall that such data or content are distinguished by tuples (l, t, u) with location l, posting time t, and user u.) The important item for harvesting data is the central software plus hardware platform that works as a switch/hub. We call this central platform a “data harvesting server” that provides an intermediate point between the Geosocial Network and analysis building.

Data harvesting is one of the main and challenging tasks of the system because data are captured and harvested at high-speed in real time. There are two possible methods for obtaining data from social networks. First, the data can easily be obtained from friends’ Facebook walls, joined groups, liked pages, and the user’s account. However, with this method, the user is required to have many friends, liked-pages, groups, etc. with a diverse location in order to obtain a huge amount of heterogeneous data with diverse topics. Second, we can harvest data that is publically available to anyone by the social networks. However, this method cannot guarantee of harvesting the real-time data for all the social networks. With either method, data are harvested using two types of APIs: streaming and state transfer. Streaming APIs are used to continuously obtain real-time data with minimum delay, whereas state transfer API allows obtaining data with static time slots, such as after 5 minutes, 1 hour, or 1 day.

In any Geosocial Network, the data is harvested by a query method sent as an HTTP request. The query response might be in either XML or JASON format. The query is always sent and data are received based on the given location, time, particular user ID, or content. In the case of Twitter, two APIs are provided to obtain tweeter data, i.e., 1) Streaming, used to continuously obtain real- time data without delay by giving the topic/content, location, and maximum size limit and 2) RESTful, which can work similarly to the state transfer API, and is used to obtain data in discrete delays, such as after 5 minutes, 1 hour, or 1 data [28]. In both cases, the received data contain the actual data and metadata. In the case of Flickr images, the response received contains user information, time, and location with the attached image. Twitter’s response also contains metadata with the tweet or comment.

In any Geosocial Network data analysis, location and time are the two most important items, along with content, because time is used to determine the relevancy and authenticity of the post, whereas location indicates the area where various events occur and activities are performed. However, gaining and extracting metadata also require technical aspects. Location can be identified by various IP techniques [29,30]. Moreover, with the current advancement in the GPS technology, all smart devices send geolocated information, such as Earth coordinates, along with the data to Geosocial Networks. In the case of Flickr, location is given as metadata that is either precise or descriptive. Actual location can be obtained from EXIF records of the image. A Flickr image contains two types of time attached to the image post, i.e., 1) time taken, which is the actual time the image was captured with a camera, and 2) update time, which is the time the image was uploaded to, or updated on, Flickr. The time information is encapsulated in an EXIF record. In the case of Twitter, the tweet posting time is attached as metadata with the tweets. After a few seconds (2–5), the posted tweet becomes public/online.

The data processing layer has two subparts: 1) preprocessing and computations and 2) analysis. This layer is responsible for all types of complex processing, analysis, and results generation. Because the system manages a significant amount of social network data, it is not possible to process each and every byte, regardless of its usefulness. Thus, filtration is performed in order to reduce data size by discarding unnecessary data and decrease computational overhead. The unnecessary metadata, irrelevant posts with respect to content, area, and time, are discarded. At the preprocessing level, classification is performed in order to further reduce computational overhead by forming an organized structure for data, which makes data access very fast. The data are classified based on location, time, and content. Various content types are already notified, such as earthquakes, fires, and Ebola virus, based on the requirements for classifying the data. The next phase of data processing is the analysis of the classified data. Given that such analysis requires significant data computation, we need a powerful hardware and software system with the ability to manage such large amount of social network data. Thus, we deployed a Hadoop ecosystem with a strong distributed file system (HDFS) that can store data across multiple nodes with higher security, reliability, and fast access. The Hadoop ecosystem can perform parallel processing on the same data stored on HDFS nodes using its parallel programming paradigm, i.e., Map Reduce. Hadoop was initially developed for batch processing, but here we need real-time processing on continuously incoming geosocial data. Therefore, in order to simultaneously perform real-time analysis and getting parallel processing benefits of Hadoop, we deployed Spark at the top of the Hadoop system for efficient processing of real-time data. Various machine learning algorithms, statistical analysis, text analytics, and content-based analysis are performed on the processing layer to generate results for decision-making. In our case, we simply provide basic analysis to determine various disaster events, such as earthquakes, fires, and diseases.

Finally, we have the service and application layer that is responsible for decision-making and using the results for various applications based on needs. For example, in the case of earthquake detection in various areas, Hadoop provides all the computation and statistical analyses of the tweets that contain earthquake information at time t. Hadoop can only provide the results, which can be manipulated by the decision server. The decision server then identifies the location based on the results, such as time, how many tweets are received, and authenticity of the results. The decision server determines whether there is an earthquake or any other event at a particular location. Later, these results can be used for many applications, such as urban planning, transportation, citizen safety, business, and traveling.

III. Analysis and Discussion

In this section, we provide the geosocial data analysis details aimed at detecting various events and disasters. First, the dataset details are given, and then a discussion on data analysis is provided.

A. Dataset Description

We obtained data from Twitter's stream grab [31] for analysis; such data contain tweets from 2010 to 2015. Each dataset is more than 40 GB in size, has more than a month of tweets, and is classified by date. We also obtained data from MAPD [32] for simple analysis; in this case, the data contain more than 60000000 tweets from November 1, 2014, to March 01, 2015. The data can be classified with respect to popular hashtags. The tweets are heterogeneous and were made in multiple languages.

B. Discussion

Analyses are conducted on Twitter data obtained from November 2014 to March 2015. The overall tweet map is shown in Figure 3. The map reflects that most of the tweets are in English by showing more tweets in the USA, UK, and other European countries, followed by Indonesia, Japan, and other Asian countries.

We also analyze all tweets by considering several events and disasters in various Earth regions, such as fires, earthquakes, snow, and Ebola virus; the analysis is conducted using the corresponding hashtags and content. Figure 4 shows the number of tweets for each event/disaster. The tweets that contain #snow, #Ebola, and #earthquake are almost close to 5000, whereas the numbers for "fire" and "snow" are quite higher. The high number for "snow" can be explained by the fact that it snows in winter in most of the countries analyzed. Similarly, the keyword "fire" cannot be used only for actual fire events given that it is also used for hot topics, anger, etc. Therefore, the number of tweets in the case of fire is also higher. We also have to perform content analysis for such types of keywords that can be used for several purposes.

In the case of earthquake analysis, we found that a majority of the tweets are from the USA and Japan. The overall tweet map with the #earthquake tag is shown in Figure 5, whereas the map shown in Figure 6 indicates the earthquake tweets in various area of Japan, including its ocean area. Figure 7 also shows the countries with the maximum number of tweets with the #earthquake tag. Moreover, we found that most tweets are posted within few minutes of an earthquake, and the tweets also contain the magnitude of the earthquake. Similarly, the fire and #fire tweet map for the entire world is shown in Figure 8, where it is obvious that most of the fire events occur in the USA and

Fig. 3: Overall tweet map from November 2014 to March 2015

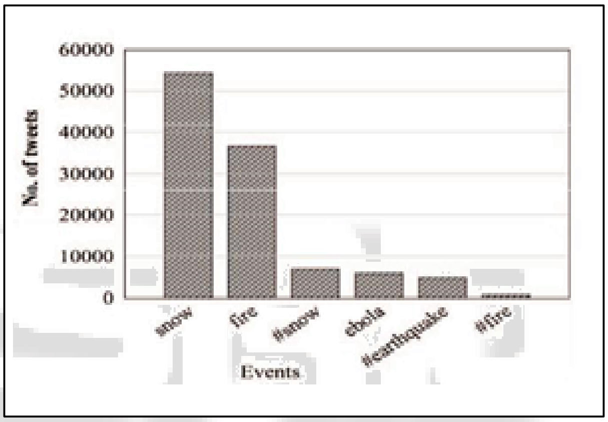

Fig. 4: Various world events and number of tweets

Fig. 5: Tweets found with #earthquake tag in various world areas

Fig. 6. Tweets found with #earthquake tag in Japan

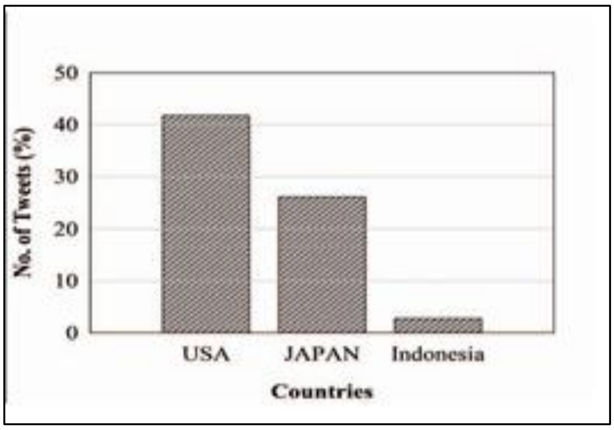

Fig. 7: Countries with maximum number of #earthquake tag

UK. The locations of fire and #fire tweets within the USA are shown in Figure 9, where it can be seen that most tweets are from the East and West coasts. The number of countries where most of the fire events occur is depicted in Figure 10.

Fig. 8: Tweets found with #fire and "fire" in various world areas

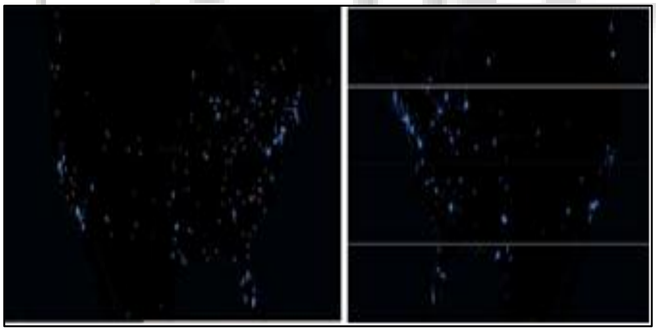

Fig. 9: Tweets found with #fire tag and "fire" keyword in the USA

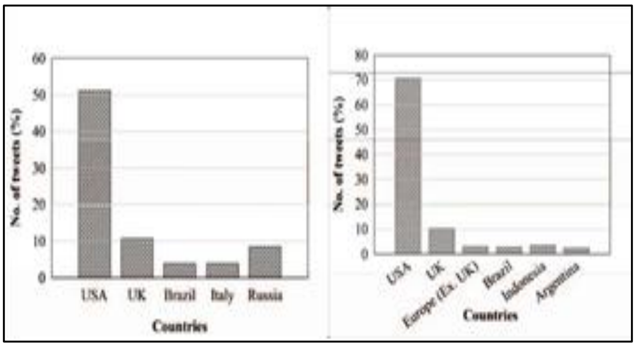

Fig. 10: Countries with maximum number of a) #fire tag b) fire keyword

The snow event is also important for monitoring floods and water resources. Therefore, we also consider the snow keyword in Twitter to find snow events. Figures 11 and Figure 12 depict the snow event analysis, where the USA is again the country with the most tweets. Because we only considered English keywords and most Twitter users are located in the USA, it is possible for the USA to be the top country in most cases.

IV. SYSTEM IMPLEMENTATION AND EVALUATION

For offline data analysis, the system is implemented using a Hadoop single node cluster at Ubuntu 14.04 LTS core TMi5 machine with 3.2 GHz × four processors and 4 GB memory. The offline Twitter data is obtained by using RESTful API [28] from Twitter's stream grab [31]. We preferred Hadoop ecosystem as it is very powerful, efficient, fault tolerant platform equipped with Hadoop Distributed File System (HDFS) and parallel processing programming mechanism, i.e., MapReduce. The data is divided into chunks and stored at multiple data nodes, which make data be processed in parallel. On the other hand, the main disadvantage of the Hadoop is its unsuitability to process real-time data. To overcome this limitation, we used Apache Spark for real-time streaming data processing and deployed it over the Hadoop ecosystem to gain the advantages of Hadoop.

Real-time implementation of the system is done by using Spark Streaming API of Apache Spark that has higher throughput and more fault tolerant. Spark Streaming collects real-time streaming data from a various system such as Twitter, Kafka, Flume, etc. It also has the ability to take data from various file systems, such as HDFS, NFS, and S3. Also, it can capture real-time data from TCP sockets. Additionally, it also has the ability to harvest the data from multiple sources in parallel using its distributed streaming (DStream) facility. For the proposed system implementation, we directly collect the data from Twitter by creating a Twitter application with the access token, key, and token secret. The real-time Twitter data is harvested continuously in chunks of few seconds (we collected in 5 seconds chunks) duration by Spark Streaming and stored them in HDFS. These data chunks called resilient distributed datasets (RDDs). Later, the Spark Engine processed each incoming chunk, performed analysis, and stored the results into either in-memory database or offline database/HDFS. The in-memory database contains the intermediate results that are used for processing and analyzing the next chunk of data. At filtration level, we just provide the simple filtration algorithm to filter tweets with given hash (#) tags and keywords. All other data is discarded. Moreover, the data is also processed in groups of chunks by sliding window when needed.

Fig. 11: Tweets found with #snow tag in various world areas

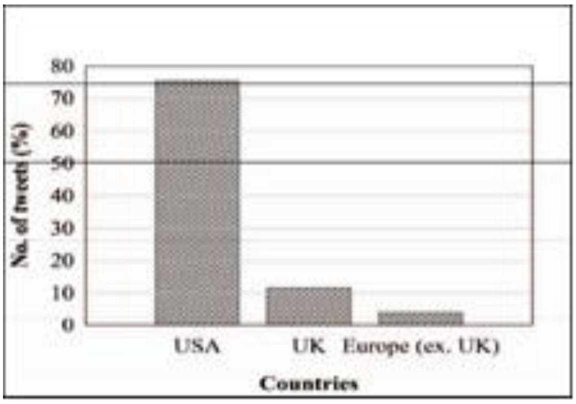

Fig. 12: Countries with maximum number of "snow" keywords

Efficiency is the main consideration when we talk about Big Data analytics. Therefore, the system is evaluated with respect to efficiency in terms of throughput (in megabytes/sec, or Mbps). The throughput analysis result is shown in Figure 13. It is obvious from the graph that the throughput is directly proportional to data size because of the parallel processing nature of the Hadoop system. The major achievement of our system is that with an increase in data size, throughput also increases. When data size increases, the average number of switches between the Map and Reduce functions decreases, thus resulting in rising throughput.

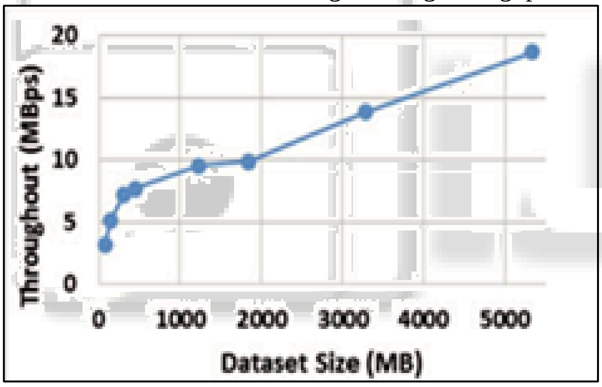

Fig. 13: Throughput of datasets based on data size

V. CONCLUSION

Geosocial Networks can be an asset for governments in terms of providing facilities and safety from disasters through proper management and reduction of the fear of the spread of any infections. Similarly, such networks can benefit to common citizens by providing recommended systems, transport safety, healthcare, etc., and to entrepreneurs for launching new products in various areas by monitoring the geosocial data of a particular area. However, such benefits can only be derived with better analytics that employs a significant amount of data generated from various Geosocial Networks. This is possible with advanced technology and better analytics, and a system with high computing capabilities. Therefore, in this paper, we proposed a system that uses geosocial data for better planning, safety from disasters, and proper management, awareness, etc., based on various geolocations. The system not only can harvest a large amount of data at high-speed from Geosocial Networks, but it can also process, analyze, and make decisions in real time. We analyzed Twitter data for various events using the proposed system. The system was developed using a Hadoop ecosystem with Spark. The system was more efficient when processing a lot of datasets, and showed the advantage of increased throughput with an increase in data volume.

ACKNOWLEDGEMENT

This study was supported by the Brain Korea 21 Plus project (SW Human Resource Development Program for Supporting Smart Life) funded by Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, Korea (21A20131600005).

REFERENCES

[1] Hern, "Online volunteers map Philippines after typhoon Haiyan," The Guardian, November 15, 2013. http://www.theguardian.com/technology/2013/nov/15/online-volunteers-map-philippines-after-typhoon-haiyan." [Accessed: October 12, 2016].
[2] M. Haklay, "How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets," Environment and planning. B, Planning & design,vol. 37, no. 4, p. 682, 2010.
[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.
[4] Hootsuite, "hootsuite.com", accessed on 19/02/2016
[5] 140kit,"https://github.com/WebEcologyProject/140kit," accessed on 19/02/2016
[6] Stefanidis, A. Crooks, and J. Radzikowski, "Harvesting ambient geospatial information from social media feeds," GeoJournal, pp. 1–20, 2012.
[7] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui, "Exploring millions of footprints in location sharing services.," ICWSM, vol. 2011, pp. 81–88, 2011.
[8] Crooks, A. Croitoru, A. Stefanidis, and J. Radzikowski, "#Earthquake: Twitter as a distributed sensor system," Transactions in GIS, 17(1), pp.124-147, 2012.
[9] Chow, J. Bao, and M. Mokbel, "Towards location-based social networking services," 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks, pp. 31-38, ACM, 2010.
[10] Papadimitriou, P. Symeonidis, and Y. Manolopoulos, "Geo-social recommendations," ACM Recommender Systems 2011 (RecSys) Workshop on Personalization in Mobile Applications, 2011.
[11] Y. Zheng, Y. Chen, X. Xie, and W. Ma, "GeoLife 2.0: A location-based social networking service," IEEE Tenth International Conference on Mobile Data Management: Systems, Services and Middleware, 2009 (MDM'09), pp. 357-358, 2009.
[12] O'Connor, M. Krieger, and D. Ahn, "Tweetmotif: Exploratory search and topic summarization for twitter," In ICWSM, 2010.
[13] L. Ferrari, A. Rosi, M. Mamei, and F. Zambonelli, "Extracting urban patterns from location-based social networks," In Proc. of the 3rd ACM LBSN, 2011.

[14] Xia, R. Schwartz, K. E. Xie, A. Krebs, A. Langdon, J. Ting, and M. Naaman “CityBeat: real-time social media visualization of hyper-local city data,” In Proc. of WWW Conference, 2014.
[15] Paul, Anand, et al. "Smartbuddy: defining human behaviors using big data analytics in social internet of things." IEEE Wireless Communications 23.5 (2016): 68-74.
[16] J. Bobadilla, F. Ortega, A. Hernando, and A. Guti_errez, “Recommender systems survey,” Knowledge-Based Systems, 46:109-132, 2013.
[17] R. A. Becker, R. Caceres, K. Hanson, J. M. Loh, S. Urbanek, A. Varshavsky, and C. Volinsky, “A tale of one city: Using cellular network data for urban planning, “IEEE Pervasive Computing, 10(4):18-26, 2011.
[18] Y. Kanza, E. Kravi, and U. Motchan,” City nexus: Discovering pairs of jointly-visited locations based on geo-tagged posts in social networks,” SIGSPATIAL'14, pages 597-600, Dallas, TX, 2014.
[19] Ratti, S. Williams, D. Frenchman, and R. Pulselli, “Mobile landscapes: using location data from cell phones for urban analysis,” Environment and Planning B Planning and Design, 33(5):727, 2006.
[20] Hefez, Y. Kanza, and R. Levin, “Tarsius: A system for traffic-aware route search under conditions of uncertainty,” SIGSPATIAL'11, pages 517-520, Chicago, 2011
[21] R. Levin and Y. Kanza, “Tars: Traffic-aware route search,” Geoinformatica, 18(3):461-500, 2014.
[22] R. Lan, M. D. Adelfio, and H. Samet, “Spatio-temporal disease tracking using news article,” HealthGIS'14, pages 31-38, Dallas, TX, 2014.
[23] R. Lan, M. D. Lieberman, and H. Samet, “The picture of health: map-based, collaborative spatio-temporal disease tracking,” HealthGIS'12, pages 27-35, Redondo Beach, CA, 2012.
[24] L. A. Waller and C. A. Gotway, “Applied Spatial Statistics for Public Health Data,” Book: John Wiley & Sons, Hoboken, New Jersey, volume 368. 2004.
[25] Rathore, M. Mazhar, et al. "Real-time Medical Emergency Response System: Exploiting IoT and Big Data for Public Health." Journal of medical systems 40.12 (2016): 283.
[26] Ahmad, Awais, et al. "An Efficient Multidimensional Big Data Fusion Approach in Machine-to-Machine Communication." ACM Transactions on Embedded Computing Systems (TECS) 15.2 (2016): 39.
[27] Rathore, M. Mazhar, et al. "Exploiting encrypted and tunneled multimedia calls in high-speed big data environment." Multimedia Tools and Applications (2017): 1-26.
[28] Fielding, R.T., “Architectural styles and the design of network-based software architectures,” Thesis (PhD). University of California, Irvine, CA, 2000.
[29] Eriksson, Brian, Paul Barford, Joel Sommers, and Robert Nowak, "A learning-based approach for IP geolocation," In International Conference on Passive and Active Network Measurement, pp. 171-180. Springer Berlin Heidelberg, 2010.
[30] Poese I, Uhlig S, Kaafar M A, Donnet B, and Gueye B, “IP Geolocation databases: Unreliable?,” Computer Communication Review 4 (2): 53–56, 2011.
[31] Twitter stream, “Archive Team: The Twitter Stream Grab,” https://archive.org/details/twitterstream, [Accessed on 19/02/2016].
[32] TWEET MAP, “MAPD Twitter,” http://tweetmap.mapd.com, [Accessed on 19/02/2016].

