Adequate Analyzing of Social Network based on Big Data

Heena Rawal[1] Dr. Priyanka Sharma[2]
[1]M. Tech. Student (Cyber Security) [2]Professor
[1,2]Department of Information Technology
[1,2]Raksha Shakti University, Ahmedabad, India

Abstract— As a late there has been an immense increase of social network of data in modern world. This is evident within the field of social media. The amount of social data is growing fast. For a social media networking platform, this provides means of getting more information to the world. This platform gives us many useful information, thus the analysis and information extraction is of great value. The average global internet user spends two and a half hours daily on social media. This way social media users produce huge amount of data which cannot be handled with conventional data management techniques. The social media are making compelling investments in putting this data to work because it gives the analysis of media content has been central in social sciences, due to the key role that media plays in shaping public opinion etc. The main aims of this paper are impact of big data on social media and some technical issues and impact on society. As these networks can be very large the methods used to study them must scale linearly when the network size increases. Thus, an integral part of the study is to determine which social network analysis algorithms that have this scalability. Another important part of using social network analysis is to be able to interpret the results. It will include discussion on appearing and circumstantial in education, economy, advanced technology, environment and safety and role of these two contemporary technologies of Big Data and Social Media on them.

Key words: Social Networks, Social Media, Big Data, Scalability

I. INTRODUCTION

Big Data is becoming a very hot topic in many areas where data files are so large that they can no longer be handled forcefully. Big Data lead to social advantage as well as social behavior. Social media voice represents public voice is 'Very Big' involvement. It creates need of real time analysis and also need of capable cadent for social media and big data. To Analyses large portion of freely existing data from social media has created new convenience to understand and effect how people think and act.

The growing use of social networks, such as Facebook, Twitter, and Weibo, has twisted and is creating huge capacity of data. Business organizations and other firms are attentive in realizing new business insight to increase business performance.[3]

The Big Data formed by social networks can be examined by present computer technologies. MapReduce, Hadoop and NoSQL methods have supported distributed data storage, parallel data recovery and handling. Many analytical procedures and algorithms are designed for business analytics, such as K-means clustering, Association rules, Linear/logistic regression, and Time series. Many software companies have urbanized their BDA products. For example, IBM has a series of software to support BDA including InfoSphere, PureData, Cognos and SPSS modeler.[3]

II. EXPLOSION WEB DATA

A. Managing and Capitalizing on the Current Data Boom

We are living in the center of a data explosion, a true boom in databases and database technology, and the likes of which the world has never seen. The Big Data explosion by the current data boom, and how critical it is for us to be able to extract meaning from all of this data. Data boom means that data professionals hold the keys to the kingdom. Not only is data getting more immense each and every year, the rate at which data is being generated is accelerating at a super-linear rate. It's hard enough to figure out how to capture and manage all that data, but more prominently we need to identify the right set of architectural decisions, tools, and capabilities to allow our organizations to capitalize on that data. In other words, giving meaning to raw data is just as important as the collection, consistency, and management of Big Data environments.[4]
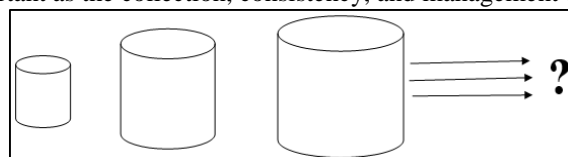

Fig. 1: The Big Data explosion

B. Focus on Database Sharding

A technique and design for horizontal *shared-nothing* splitting of database of data across independent knobs or servers. From wide experience and study, we have found that, regardless of what DBMS you use, database sharding is the only actual means

for scaling a database. This applies to all types of databases, including the outdated relational database management system (RDBMS), Big Data engines, Cloud databases, NoSQL platforms or databases of the so-called New SQL standard. If you look closely at any of these presents, particularly if they promote an accessible platform, you will see that some sort of horizontal partitioning of data across nodes is used. Thus, database sharding is the answer, used in one way or another for almost every important database platform. Figure 1.2 shows the concept of database sharding, breaking a large monolithic database into several, and smaller, shared database orders across multiple servers.[4]
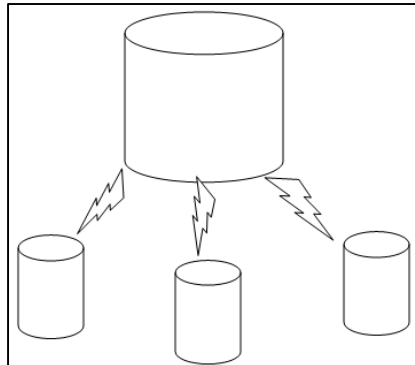

Fig. 2: Database Sharding

III. ISSUES AND CHALLENGES IN BIG DATA AND SOCIAL MEDIA

A. Analysis of Social Media Data

Social data analysis comprises two main basic parts: 1) data produced from social networking sites (or through social applications), and 2) sophisticated analysis of data, in many cases demanding real-time (or near real-time) data analytics, capacities which understand and properly weigh factors such as influence, reach, and relevancy, an understanding of the context of the data being examined, and the presence of time horizon attentions. In short, social data analytics contains the analysis of social media in order to understand and surface visions which is rooted within the data. The maximum data on social media is amorphous, so it is difficult to analysis or upholds a data. The size of data is in massive form which is collected from different birthplaces.

B. Technology Challenge in Big Data

IT becomes the data hero in today's world. It's finally IT's time to break the cycle and change from producer to enabler. IT is at the wheel of the transformation to self-service analytics at scale. IT is providing the elasticity and alertness the business needs to renovate all while balancing governance, data security, and compliance. As big data proceeds its complicated way into the enterprise, information technology (IT) experts and business sponsors alike will smash up against a number of tasks that must be addressed before any big data program can be fruitful.

Five of those challenges are:
1) Uncertainty of the Data Management Landscape – There are many challenging technologies, and within each technical area there are numerous competitors. Our first challenge is making the best choices while not introducing supplementary unknowns and risk to big data implementation.
2) The Big Data Talent Gap – The enthusiasm around big data applications appears to suggest that there is a broad community of experts available to help in enactment. However, this is not yet the case, and the talent gap poses our second task.
3) Getting Data into the Big Data Platform – The scale and diversity of data to be fascinated into a big data environment can overcome the unprepared data expert, making data user-friendliness and integration our third challenge.
4) Synchronization Across the Data Sources – As more data sets from various sources are incorporated into an analytical platform, the potential for time holdups to impact data currency and reliability becomes our fourth challenge.
5) Getting Useful Information out of the Big Data Platform – Lastly, using big data for different resolutionsreaching from storage expansion to enabling high-performance analytics is delayed if the information cannot be sufficiently provisioned back within the other modules of the enterprise information design, making big data syndication our fifth challenge.

C. Management Challenges

Social Media warehouses cover sensitive data for example personal data because in social networking sites people are connected with diversity of people and sharing private information. There are authorized and proper concerns with retrieving such data. So the data must be safe and access controlled as well as logged for reviews. The main managing challenges are data privacy and security.

D. Security Issues

An initial research area in data mining, known as privacy-concern data mining (PCDM) has been widely measured in recent years. The basic idea behind of PCDM is to change the data in such a way so that to perform data mining algorithms the alteration of the data using the data mining algorithms should not negotiate the security of sensitive information contained in the data. It is a big challenge for researcher to reduce the privacy risk carried by data mining because many unwanted expose

of sensitive information may occur at the time of data mining result. For security purpose attention should be four different types of users that are involved in data mining applications, specifically, data provider, data collector, data miner, and decision maker.

IV. RECENT STATISTICS OF BIG DATA AND SOCIAL MEDIA

As social media usage continues to grow at a steady rate. It is a fact of the internet that every click, every view and every sign-up is recorded somewhere. Here are some statistics shown:

Social media Statistics [5]

- For context, as of July 2016, total worldwide population is 7.5 billion.
- The internet has 3.17 billion users.
- There are 2.3 billion active social media users.
- 91% of retail brands use 2 or more social media networks.
- Internet operators have an average of 5.54 social media accounts.
- Social media users have risen by 176 million in the last year.
- 1 million new active mobile social operators are added every day. That's 12 each second.
- Facebook Messenger and WhatsApp handle 60 billion messages a day.

Broken down site by site, latest number of social networks are as follows:

A. User Numbers

Websitesa	User numbers
Airbnb	50 million users
Facebook	1.86 billion users
Flickr	112 million users
Google+	300 million users
Instagram	400 million users
LinkedIn	450 million users
Myspace	50.6 million users
Periscope	10 million users
Pinterest	100 million users
Reddit	36 million accounts
Snapchat	100 million users
Twitter	320 million users
WeChat	1.12 billion users
Weibo	600 million users
WhatsApp	900 million users
YouTube	Over 1 billion users
4Chan	11 million users

Table 1: Recent Statistics of social media [5]

B. Social Media Business Statistics

- Social networks received an estimated $8.7 billion from advertising in 2017. [5]
- 38% of organizations plan to spend more than 20% of their total advertising budgets on social media channels in 2017, up from 13% a year ago. [5]
- Only 20 Prosperity 500 companies actually involve with their customers on Facebook, while 83% have a presence on Twitter. [5]
- People elderly 55-64 are more than twice as likely to engage with branded content then those 28 or younger. [5]
- 96% of the people that converse brands online, that do not follow those brands' owned profiles. [5]
- 78 percent of people who criticize to a brand via Twitter expect a reply within an hour. [5]

C. Content Statistics

- On WordPress alone, 56 million blog posts are published every month. [5]
- A 2011 study by AOL/Nielsen disclosed that 27 million bits of content were shared every day but in 2017, 2.7 billion or more content is shared between each other. [5]
- The top 3 content marketing approaches are blogging (65%); social media (64%); and case studies (64%).[5]

D. Social Video Statistics

- Facebook now perceives 8 billion average daily video views from 500 million users. [5]
- Snapchat users share 6 billion videos every day. [5]
- US adults spend an average of 1 hour, 16 minutes each day watching video on digital devices. [5]
- 78% of people watch online videos every week, 55% lookout every day. [5]

E. Google Statistics

− Google processes 100 billion searches a month. [5]
− That's an average of 40,000 search queries every second. [5]
− 89.3% of all internet searches are approved out by Google. [5]
− Those searches are approved out by 1.17 billion unique users. [5]
− Every day, 16 to 20% of that day's queries have never been asked before. [5]
− Google has answered 450 billion unique queries since 2003. [5]
− More than half of Google's searches come from mobile devices. [5]
− By 2016, Google had indexed 30,000,000,000,000 pages of the internet. [5]
− To carry out all these searches, Google's data Centre uses 0.01% of worldwide electricity.[5]
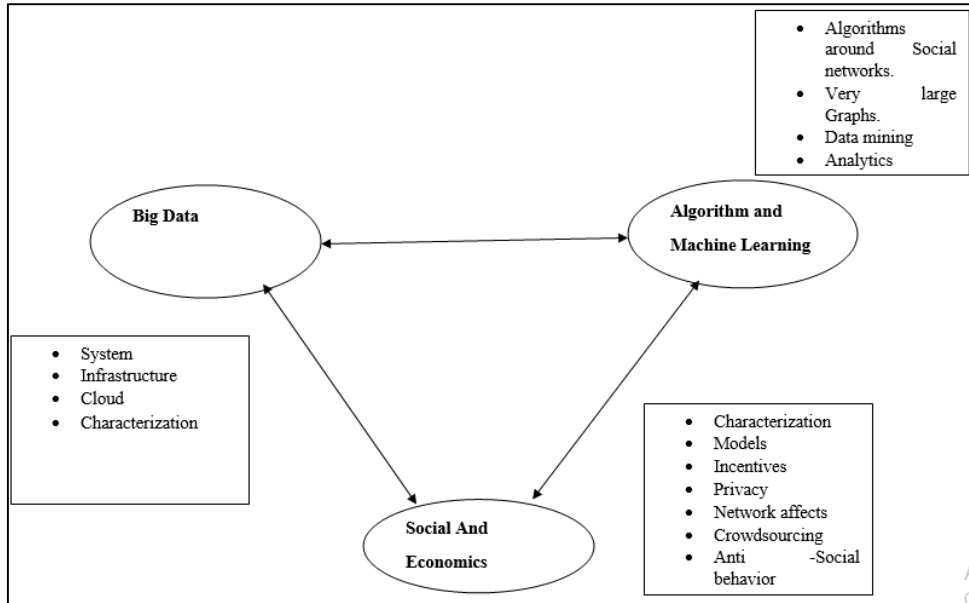
V. RESEARCH INTERESTS

Fig. 3: Life cycle of Social Network, Algorithms And big Data [2]
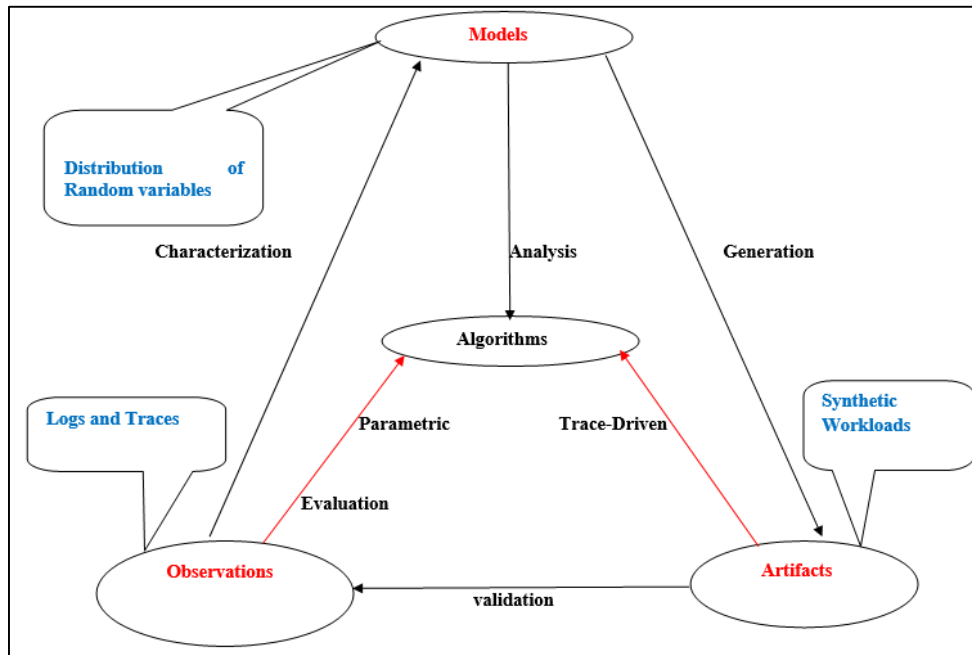
A. Experimental Methodology

Fig. 4: Cycle of Experiment methodology [2]

B. Challenges in Online Social Networking Research

− Volatile growth in size, density, and unstructured data;

− Enabled by several investigational methods: observational studies, simulations, huge amount of data;
− It is "big data," the immense sets of information collected by investigators at companies like, Google, Facebook, and Microsoft from outlines of cell phone calls, text messages and Internet clicks by millions of users around the world. Companies often refuse to make such information public, sometimes for reasonable reasons and sometimes to protect customers' privacy.

C. Enablers of Big Data

Hardware Capability	Applications And Algorithms
Storage capacity	Online social networking
Network Bandwidth	Algorithmic breakthroughs: machine learning and data mining
Exponentially increasing capability at constant cost	Cloud: Cost reductions and scalability improvements in computing.
Processing capacity	Sensor Everywhere

Table 2: Enablers of Big Data [2]

VI. CONCLUSION AND FUTURE WORK

In this paper I have addressed two contemporary useful technologies viz. Big Data and Social Media along with their main issues and con Early in this paper I have talked two existing useful technologies Big Data and Social Media along with their main problems and concern in the area of research. For various reasons, Hadoop was chosen as the system to investigate further. This could reduce the amount of data that needs to be sent between computers as well as improve the performance of any combiner in use. The scalability of algorithms is provide more information to their patterns in social networks. This might be helpful in terms of the machine learning algorithms as well, as it is more likely to find relations to the training set classification if a larger number of attributes is used.

REFERENCES

[1] Jonathan Magnusson UPPSALA University, "Social Network Analysis Utilizing Big Data Technology"(https://www.diva-portal.org/smash/get/diva2:509757/FULLTEXT01.pdf)

[2] Virgilio @dcc INWEB (National science and technology Institute for web) - Fedral University of MinasGerais–UFMG, "Exploring Big Data in social Networks" (http://wic.litislab.fr/2013/slides/virgilio-keynote-www2013.pdf)

[3] O.Liu, K.L.Man, W. Chong and C.O.Chan , Proceedings of the International Multi Conference of Engineers and Computer Scientists 2016 Vol II, IMECS 2016, March 16 - 18, 2016, Hong Kong. "Social Network Analysis Using Big Data" (http://www.iaeng.org/publication/IMECS2016/IMECS2016_pp639-640.pdf)

[4] Cory Isaacson, Under Standing Big Data Scalability, Part-1, (http://www.informit.com/store/understanding-big-data-scalability-big-data-scalability 9780133598704?w_ptgrevartcl=Introduction+to+Understanding+Big+Data+Scalability_2238298)

[5] Marketing: 96 Amazing Social Media Statistics and Facts: (https://www.brandwatch.com/blog/96-amazing-social-media-statistics-and-facts-for-2016/