



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

Big Data Mining: Challenges and Opportunities to Forecast Future Scenario

Poonam G. Sawant, Dr. B.L.Desai

Assist. Professor, Dept. of MCA, SIMCA, Savitribai Phule Pune University, Pune, Maharashtra, India

Manager, BIM Learning & Development and Big Data Academy, Capgemini, Pune, Maharashtra India

ABSTRACT: Big Data is a hot research topic in today's Internet world. Big Data is defined as collections of large, complex or required data which become difficult or impossible to process using current methodologies, standard database management or analytical solutions. Big Data mining is the capability of discovering knowledge and patterns from the large sets of information, streams of data due to its volume, variability, and velocity. Mobile and social Medias are major data exhaust producers. Mining this data is the new frontier for the next years. This paper represents overview of big data, its sources, challenges and opportunities to forecast the future. We have summarized various articles, research papers written by inertial scientists, researchers and scholars in the relative field.

KEY WORDS: Big Data, Big Data Mining, Volume, Velocity, Variety, Challenges and Opportunities, HADOOP.

I. INTRODUCTION

In Recent years we have seen the dramatic increase in the growth of information due to collection of data from various independent or connected applications and services. Very popular example is Internet data. Internet is growing at lightning speed. The information stored on the net is huge. Every second millions of bytes are added all over the world. The immense data and knowledge on the net is growing at a mind-blowing rate and even estimates are proven wrong every second. [6].

This massive amount of data is called as a Big Data. The 'Big Data' term is used to describe collections of large, complex or required data which become difficult or impossible to process, analyze and store using current methodologies, standard database management and analytical solutions. Big data brings together large amount of data with various data types that previously never would have been considered. Big data includes structured data, semi structured data and unstructured data. Structured data are those data formatted for use in a database management system. Semi structured and unstructured data include all types of unformatted data including multimedia and social media content. This huge amount of data is very valuable to improve quality of life and make our world a better place by extracting meaningful associations, trends and patterns, but it is unmanageable. Here comes the role of big data collection and big data analytics and algorithms to manage and extract meaningful information and patterns from the large sets of information, streams of data due to its volume, variability, and velocity.

II. RELATED WORK

Challenges and Opportunities with Big Data (A community white paper developed by leading researchers across the United States) -2012: This paper describes many technical challenges. The challenges includes some issues of scale, heterogeneity, lack of structure, error-handling, privacy, timeliness, provenance, and visualization, at all stages of the analysis pipeline from data acquisition to result interpretation. These technical challenges are common across a large variety of application domains, and therefore not cost-effective to address in the context of one domain alone. Furthermore, these challenges will require transformative solutions, and will not be addressed naturally by the next generation of industrial products. According to authors, through better analysis of the large volumes of data that are



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

becoming available, there is the potential for making faster advances in many scientific disciplines and improving the profitability and success of many enterprises. So there is a need of maximum support and encourage for fundamental research towards addressing these technical challenges of Big Data management.[1]

Big Data a New World of Opportunities: 2012:This paper highlights selected aspects of Big Data that create particular opportunities, but also challenges for the software and services industry in Europe. Therefore, NESSI address current problems and materialize opportunities associated to the use of Big Data, by applying a unifying approach, where technical aspects are aligned with business-oriented, regulatory and policy aspects. This paper states how Big Data actors can create a better momentum for Europe in terms of global competition. Thus, the overarching goal is to identify Big Data research and innovation requirements in the context of Horizon 2020. [5]

Mining Big Data: Current Status, and Forecast to the Future by *Wei Fan Huawei* (Huawei Noah's Ark Lab , Hong Kong) and *Albert Bifet* (Yahoo! Research Barcelona, Spain) -**2013.**:This paper presents some insights about big data and big data mining, Analysis tools, the main concern and the main challenges for the future that researchers and practitioners will have to deal during the next years. According to authors Big Data is going to continue growing during the next years, and each data scientist will have to manage much more amount of data every year. This data is going to be more diverse, larger, and faster. Big Data is becoming the new Final Frontier for scientific data research and for business applications. We are at the beginning of a new era where Big Data mining will help us to discover knowledge that no one has discovered before. [6]

Big Data Analytics (Ericsson White paper 288 23-3211 Uen)- **2013:**This paper explores big data analytics, the opportunities and how big data should be expanded to support analytics. It states that a common, horizontal big data analytics platform is necessary to support a variety of analytics applications. Such a platform analyzes incoming data in real time, makes correlations, produces insights and exposes those insights to various applications. This approach both enhances the performance of each application and leverages the big data investments across multiple applications. Storing and processing huge amounts of information is no longer the issue. The challenge is to know what needs to be done within the big data analytics platform in order to create specific value. Big data storage and processing techniques are necessary enablers; the goal must be the creation of the right use cases. The big data tools and technologies deployed have to support the process of finding insights that are adequate, accurate and actionable. [7]

III. SOURCES OF BIG DATA

Big data comes from various sources and in various formats. Every day we create 2.5 quintillion bytes of data. 90 % of the data in the world today has been created in the last 2 years alone. According to IBM Report (2014), from 2003 to 2010 we have created 5 billion gigabytes data. In 2011 the same amount of data was creating in two days. In 2013 the same amount of data is created in every 10 minutes. This rapid expansion is accelerated by the dramatic increase in acceptance of social networking applications, such as Face book, Twitter, mobile phones etc., providing each individual the platform to express their views with very less barriers and with social networking elements and to create contents freely and amplify the already huge Web volume. [6]

According to the recent survey done by one big mobile company (Source: www.sakal.com, page No.7, 4/02/2015), there are 40 Cr online users in India in 2015. Use of internet application is increased by 20% within two years. Following table shows the use of various social media channels in every one minute for three years.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

Data/Source	Year		
	2012	2013	2014
Twits(Twitter)	98000	2.78 Lacs	3.42 Lacs
Calls(Skype)	3.7 Lac	14 Lac	14 Lac
Photo Upload(Instagram)	3480	216000	410000
Search(Google)	6.9 Lac	20 Lac	40 Lac
Post(Facebook)	79361	25 Lac	30 Lac
Mails(Email)	16.8 Cr.	20.4 cr.	20.4 cr.
post(Worldpress)	347	1106	1380
Messages(WhatsApp)	20 Million	31 Million	50 Million
Video(You-Tube)	25 Hrs.	72 Hrs.	120 Hrs.

Table1: Data generated in Every Minute

Also we are receiving large amount of data from numerous sources like, sensor networks, Government data holdings, company market lead databases, bidirectional interactions, E- Health networks, Closed-circuit television, video cameras etc. The Cameras in police car records pursuits and traffic stops as well as dash cam used for complaint handling. [1]

Following Figures Shows the Graphical Representation of above table. Figure a shows the percentage of the online data generated through various sources. According to graph data generation percentage is high from Google, then from Face book and Skype and low from Twitter.

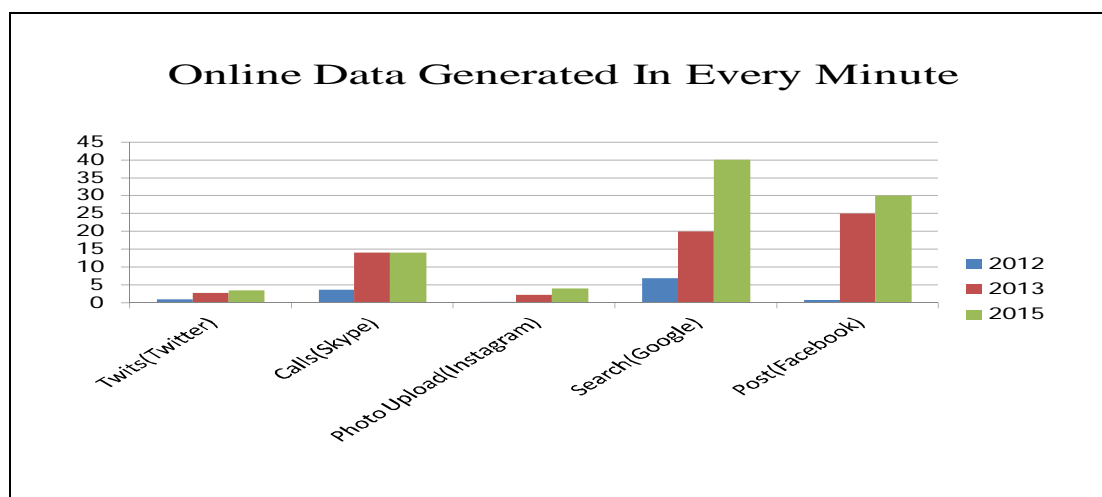


Fig. a

According to figure 2 You tube and whatsapp produce maximum data as compared to emails. The percentage of data generation is high in 2015.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

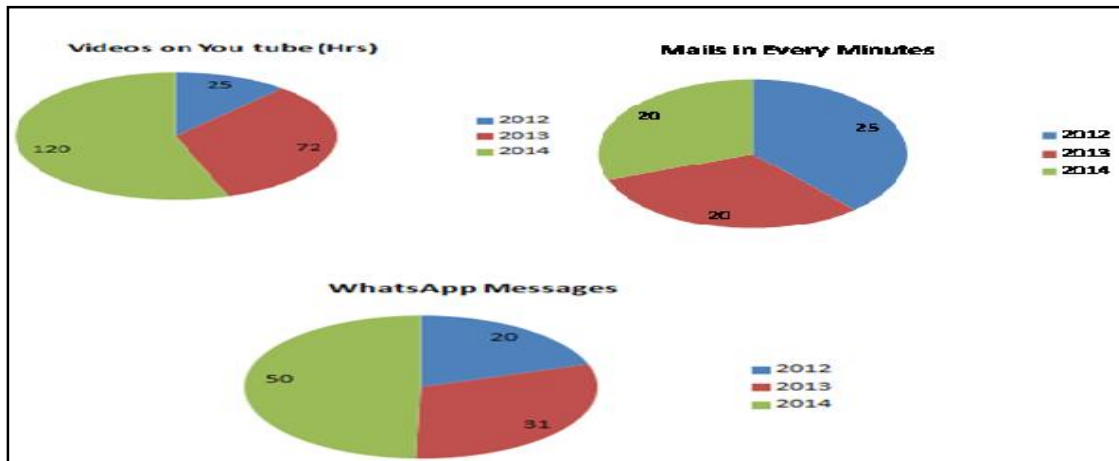


Fig. b

IV. CHALLENGES AND OPPORTUNITIES

Big data generated in growing number of ways from growing number of sources. It gives lots of opportunities in various areas as listed in the following table.

Areas	Use of Big Data
Web & e-tailing	Recommendation Engines, Ad Targeting, Search quality, Abuse and Click Fraud Detection
Telecommunication	Customer Churn Prevention, Calling Data Record (CDR) Analysis, Network Performance Optimization, Analyzing Network to Predict Failure
Healthcare	Health Information Exchange, Gene Sequencing Serialization, Drug Safety
Government	Welfare Schemes, Justice, Fraud Detection and Cyber Security
Banks and Financial services	Modeling True Risk, Threat Analysis, Fraud Detection, Trade Surveillance, Credit Scoring and Analysis, Personalized Banking Services
Retail	Point of Sales Transaction Analysis, Customer Churn Analysis, Sentiment Analysis

Table2: Opportunities of Big Data

But it becomes difficult to process using on-hand database management tools or traditional data processing applications due to its large volume and variety. The challenges include capture, curation, storage, search, sharing, transfer, analysis, and visualization. So, special techniques are needed to analyze big data.

Top level management need to become familiar with the big data methodologies, adopt the new technologies for their business and ensure the employees to develop skill with big data. Many IT companies attempt to manage big data challenges using a NOSQL database, such as Cassandra or HBase and distributed computing system such as HADOOP. NOSQL databases are typically key-value stores that are non-relational, distributed, horizontally scalable and schema-free. Traditional way focuses on resolving the complexity of relationship among schema-enabled data. These



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

considerations do not apply to NO-SQL. As a result old ways are no longer apply. So, we need new methodology, efficient algorithms and good analytics to overcome big data challenges.

V. THE IMPORTANCE OF BIG DATA IN THE BUSINESS WORLD

Big data is very important in business world. The significance of big data is derived from the data collected through various data sources. Previously business world purely relied on structured data collected and stored in a traditional database. Data collected from social media and Internet provides 90% unstructured data in various formats like xml files, word files, PDF, text, email, videos, images etc. which is very important to take business decisions. Analysis of these data can give new insights for executives. It will provide more accurate, relevant and detailed data allowing top level management to analyze performance variability, useful trends, hidden behavioral patterns and plan action plan to retain existing customers and attract new one by minimizing risks. New product, services, strategies and business models can develop from analysis of big data which will help to improve the quality of life.

VI. CONCLUSION

Big data poses opportunities and challenges for business. Previously data are able to be stored and processed using traditional databases. Unstructured data is unable to handle by RDBMS. Big data gives opportunities to business executives to make effective business decisions and improve quality of products, services and business models. So there is a need to follow trend in big data carefully to make the proper decisions. Big data NOSQL and Hadoop distributed system helps to handle big data. Creating data models and developing efficient algorithms will help managing big data.

REFERENCES

1. Challenges and Opportunities with Big Data(Journal proceedings of the VLDB Endowment Volume 5 Issue 12, August 2012, pages 2032-2033 ,A community white paper developed by leading researchers across the United States, 2012)
2. Total-number-of-websites-size-of.html (<http://www.factshunt.com/2014/01/>)
3. Data Modeling for Big Data (<http://www.ca.com/us/~media/files/articles/ca-technology-exchange/data-modeling-for-big-data-zhu-wang.aspx> by Jinbao Zhu and Allen Wang,2012)
4. Research Trends Special Issue on Big Data Issue 30 September 2012
5. Big Data A New World of Opportunities:NESSI White Paper, December 2012
6. Mining Big Data: Current Status, and Forecast to the Future (SIGKDD Explorations, Volume 14, Issue 2 by Wei Fan Huawei Noah's Ark Laband Albert Bifet Yahoo! Research Barcelona ,2013]
7. Big Data Analytics (Ericsson White paper 288 23-3211 Uen | August 2013)

BIOGRAPHY

Poonam Sawant is an Asst. Professor in Sinhgad Institute of Management and Computer Application, Pune, Maharashtra, India. She has completed her M.Phil from Shivaji University, Kolhapur, Maharashtra, India in Oct. 2012 with "O" grade. She obtained her Master's in Computer Application from SIBER, Shivaji University, Kolhapur, Maharashtra, India in 2005. Now she is pursuing her Ph.D in Computer Management from Savitribai Phule Pune University, Pune, Maharashtra, India. Her research interests are Web Mining, Big Data and Algorithms.

Dr. B.L.Desai is a Manager, BIM Learning & Development and Big Data Academy, Capgemini, Pune, Maharashtra, India.He Obtained his Ph.D from Shivaji University, Kolhapur, Maharashtra, India. He is a recognized Ph.D guide of Savitribai Phule Pune University, Pune, Maharashtra, India. His research interests are Knowledge Management, Data Mining, Expert System and Big Data.