

Proceedings of the International Conference , “Computational Systems for Health & Sustainability”  
17-18, April, 2015 - by R.V.College of Engineering,  
Bangalore,Karnataka,PIN-560059

# Impact of Big Data and Emerging Research Trends

**Chandrika M**  
Research Scholar,  
Jain University,  
Bangalore.

**Parvathy J**  
Research Scholar,  
Jain University,  
Bangalore.

**Srinidhi B S**  
Research Scholar,  
Jain University,  
Bangalore.

**Vijay Bhandi**  
Research Scholar,  
Jain University,  
Bangalore.

**S Balaji**  
Professor

Centre for Emerging Technologies  
Jain Global Campus, Jain University,  
Bangalore-562112.

**Abstract:** The term big data is extensively used in many computational and decision making domains. Big data is nothing but the large data sets formed from various sources and are almost impossible to process and analyse using traditional approaches because of its complexity. Efficient analysis and processing of big data within a given time frame is essential for it to be useful. Various technologies like Hadoop, MapReduce, etc. are used to analyse the big data and hence possible to retrieve knowledge from the large datasets. This paper focuses on the impact of big data, the technologies in big data processing and its limitations and the emerging trends in big data.

**Keywords:** Big data; Analytics; Hadoop; MapReduce.

## INTRODUCTION

Large amounts of data is generated from different sources like organizations, social media and sensors. These data can be of various formats. The statistics of IBM Big Data Flood Info graphic shows that 2.7 Zettabytes of data exist in digital universe today and 100 Terabytes of data are updated daily through Facebook and other social networks resulting into 35 Zettabytes of data generated annually by 2020 [2]. As a result of these, the data generation is very fast. From this huge data, obtaining valuable information needs efficient processing. The Organisations must process the data for valuable decision making and for extracting knowledge from the decision. Understanding how big data is influencing our everyday life is the objective of the second section and what facilitates it, in terms of technology, is being discussed in the third section. Fourth and fifth sections aim at explaining the limitations and evolution & research trends respectively.

### IMPACT: HOW BIG DATA CHANGES OUR LIVES

The term Big Data was first introduced to computing world by Roger Magoulas from O'Reilly media in 2005 [2]. The relevance of big data applies when there is a huge emerge of data from different sources. From past decade, data which is being generated from various sources are of enormous amount. The sources of the data are not only Internet but from many other sources like organizations, social media and sensors. Large volumes of data need to be analysed, processed and visualized in a short time which is not possible by traditional data processing systems. The rate of data generation is very brisk that needs to be processed in a short time. The data generated are of different formats from structured formats to unstructured formats such as audio, video, image, animation and free text. It is challenging to extract the value from these raw data. The four main characteristics of Big Data are Volume, Velocity, Variety and Value [3]. NESSI White Paper [1] gives a definition for Big Data as - “Big data is a term encompassing the use of techniques to capture, process, analyse and visualize

potentially large datasets in a reasonable timeframe not accessible to standard IT technologies.” By extension, the platform, tools and software used for this purpose are collectively called “Big Data Technologies”. The impact of big data exists everywhere. Massive data from various sources will be originating daily and needs to be analysed in a short span of time. This gives insight into various decision making strategies like approach of customers towards a new product, current trends in a particular age group, market analysis for the delivery of a new product, region wise survey for the interest of customers on products. Big data technologies play an important role in the optimal decision making and for knowledge discovery for many experimental sciences [1]. It helps the stakeholders by providing a visual effect. The concept of big data lies in converting highly unstructured data into performing data. The impact of big data is seen widely in the field of tweets, online transactions, stock trading, satellite technology and also applied in various scientific and research fields [4]. The finance sector is adopting new technology changes and big data methodologies are now being widely used in transactions. Usage of big data allows fast processing of data, quick analytics and reporting and supports high fault tolerance. The finance data has the problem of growing huge in few days and light weight big data technologies help them in bringing down the cost of investment in technology [5]. Big data technology is frequently used in large data setup for monitoring its performance and efficiency. It supports large amounts of transactional data to be processed [a million sets of data in micro seconds]. This activity improves the quality, robustness and processing time consumption. The measurable strategy makes efficient utilization of man, material and money [6]. The biggest consumer of data is the digital media and communications. Here the actions have to be performed in a quick manner and data related to individuals has to be published in different manner (images, graphs). The growth of social media and sharing (tools like Facebook and Twitter) has accelerated the use of large datasets. The impact can be seen in the following areas of



social media. The data available in social media is used for creating social graphs that publishes trends about the usage. Also this data is used in understanding customers' information into a knowledge base that can serve as an input for social media targeting. The social media produces large amount of people data such as posts, likes and comments. This has to be analysed for the future scope. Understanding the user sentiment is the main process [7].

Electronic commerce (e-commerce) has become the Point of Sale and process huge data. The data type variants are huge in this field. The main application of big data in retail commerce revolves around customer personalization. It is an approach of improving the sales experience with different classes of consumers like running offers for first thousand customers, special discounts for frequent visits are few among them. The dynamic pricing is one aspect of the retailers in customer offer management. The big data process allows to collect data from varied sources and derive a competitive pricing model in comparison with others. Another important factor in retail is the predictive analysis which allow the online stores by offering products and services before a customer actually likes and finding a potential customer on purchasing a specific item [8].

The research in the field of health sciences vastly improved the life expectancy of the people around the world. The healthcare sector generates large data in pathology, patient management and medical inventions. The use of big data techniques allows a healthcare professional to identify common symptoms of a particular disease. The pharmacy companies hugely benefit the geographic analysis of clinical trial data for checking the efficiency of medicine. It supports the governments and NGOs to understand the particular need in varied geographies. For example, a data set which represents a group of persons with rare blood groups can immensely help a patient in a distant geography [1].

#### **TECHNOLOGY AND ANALYSIS: EQUIPMENT FOR WORKING WITH BIG DATA**

A bunch of tools are available for big data analysis. Hadoop is one of the best tools available in big data area. Also, there are other data tools such as Cassandra, MongoDB are highly scalable and available over the cloud environment. To capture value from big data, organisations will have to deploy new technologies (e.g. storage, computing, analytical software) and techniques (new types of analysis). The range of technology challenges and the priorities set for tackling them will differ depending on the data maturity of the institution. Legacy systems and incompatible standards and formats too often prevent the integration of data and the more sophisticated analytics that create value from big data. New problems and growing computer power will spur the development of new analytical techniques. There is also a need for ongoing innovation in technologies and techniques that will help individuals and organisations to integrate, analyse, visualise, and consume the growing torrent of big data. Specific development targets include technology for stable operation of large-scale data management systems that compress, transfer, and store big data, technology for efficiently retrieving truly necessary knowledge by means of

search, comparison, and visualization across diverse information, and the mathematical methods and algorithms enabling such services. In pursuing these studies, with a view to overall system design up to the creation of value for society from big data, the creation, advancement, and systematization of next-generation common core technology highly acceptable to the public need to be undertaken, through active efforts at fusion with fields outside of information and communication technology. There are 6 key trends that are already happening to a greater or lesser degree that are shaping further development of technologies for big data [13]

**Big Data on Cloud:** Addressing big data is a challenging and time-demanding task that requires a large computational infrastructure to ensure successful data processing and analysis, which can be effectively addressed by utilising cloud service. Big data is driving public cloud adoption for fairly obvious reasons:

- The cloud cost is a fraction of that to purchase big data resources on demand.
- Cloud-to-cloud and cloud-to-enterprise data integration got much better in the last few years, so it is easy to set up massive databases in the clouds and sync them with any number of operational databases, cloud-based or on-premise.
- In most cases, public clouds can provide better performance and scalability for most big data systems because they can provide auto-scaling and auto-provisioning.

**Unstructured Data Processing:** Unstructured data is heterogeneous and variable in nature and comes in many formats, including text, document, image, video, and more. Unstructured data is growing faster than structured data. The development of technology [new software tools and database systems] for large, unstructured datasets processing and analysis is the latest trend in big data. An important development in the commercial market is that it is now possible to mine structured data sources alongside unstructured data sources to merge datasets that were previously incompatible.

**Real Time Analytics:** Big data analytics is a technology-enabled strategy for gaining richer, deeper, and more accurate insights into customers, partners, and the businesses—and ultimately gaining competitive advantage. By processing a steady stream of real-time data, organizations can make time-sensitive decisions faster than ever before, monitor emerging trends, course-correct rapidly, and jump on new business opportunities. The interconnected world of web and mobile apps, distributed sensor networks and cloud computing clusters require a new breed of data capture and analytics infrastructure that can handle the increasing volume and velocity of data.

**Monetization of Big Data:** It helps to transform the business to exploit big data and capture valuable insights that drive new revenue streams and enhance the customer experience. Big data technology adds transparency to revenue assurance functions on the transactional level and

can therefore serve as a gateway to effective data monetization as well. Hence, many organizations will spend effort and money in developing monetization strategies for big data.

**Privacy in Big Data:** Big data creates tremendous opportunity for the world economy not only in the field of national security, but also in areas ranging from marketing and credit risk analysis to medical research and urban planning. At the same time, the extraordinary benefits of big data are tempered by concerns over privacy and data protection. Privacy advocates are concerned that the advances of the data ecosystem will upend the power relationships between government, business, and individuals, and lead to racial or other profiling, discrimination, over-criminalization, and other restricted freedoms. Finding the right balance between privacy risks and big data rewards may very well be the biggest challenge going forward.

**Rise of Smart Cities:** With rise of smart cities we encounter big problems of managing scarce resources and providing critical public services such as security, transportation, energy and water. The ability to harness real-time, highly granular data across a wide range of city operations and services is changing the way we manage and experience the urban environment. For this reason, the benefits offered by big data are a key element of many smart city strategies. Important challenges and opportunities for big data provided by rise of smart cities as below:

- Predictive analysis of traffic and transport patterns can reduce congestion and improve the efficiency of public transport services
- City resources for public safety, social care, and other key services can be targeted more effectively using up-to-date analysis
- Energy efficiency programs can be directed at the most vulnerable households and at suitable buildings for retrofit programs
- Opening up city data can increase citizen engagement and encourage new forms of creativity and innovation amongst developers and other service providers.

#### LIMITATIONS AND DELUSIONS: CRITICAL QUESTIONS FOR BIG DATA

##### Can big data deliver everything?

This is a basic question that arises when it comes for the big data analytics. Certainly not .We cannot expect miracle to happen on day by analytics .It is an engineering process which has to be studied regularly to predict a pattern for a probable forecasting related to their study. Also “Can a company acquire profits and read customer behaviour using the historical data?” To certain extent companies can identify the customer patterns, only by this factor alone may not be suitable .There are always an environment factor that gains prominence.

##### What is further investigation in big data?

Can machine learning algorithms be adapted so that, it gives automatic predictions when once system is setup? Automating the big data is an excellent thought as to make it to be a learning system on its own experiences. But can it manage the dynamic variables, and situation is what has to be analysed, also the amount of risk involved in automating the process also needs to be judged before we take this that extent.

##### Is it feasible to only infer and act according to analytics?

This decision has to be made before modelling the system .If a system is modelled in a proper way, yes we can infer .But it has the limitation, when the system is not modelled in a proper way there are threat always to have wrong inferences.

##### Should it be infrastructure wise advanced or analytically?

This is the basic dilemma a new investigator always get into .But according to us it is the good combination of both we certainly work. At a large scope big data profits but has its own limitations like scoping the big data in terms of years to be considered. For example, For an E-commerce application should it consider that past data like 80’s because it has got the recent exponential growth. Conflict of legal issues while considering the social network data in reading consumer behaviours. Sentiment analysis an important factor which all big data algorithms have to be adapted to. Availability of apt data might be an issue for some key projects with the rare domain. Though there are limitation in big data and some basic questions , this can be answered by including the basic modelling right with what all resources is required and it has to be planned and modelled well so that we reduce the limitations of big data .Also there are excellent hardware support we get, algorithms implemented plays a key role.

#### SCOPE FOR RESEARCH

The deluge of data has spread into every area of world economy. Various universities, research centres and organizations across the world are taking initiatives in conducting research programmes on big data. In recent years, major research topics in various fields include big data as a part of it. A big data project deals with correlating different data structures, framing a testable hypothesis designing accordingly and evaluating the obtained result with the expected result. According to Google Trends, big data is one of the revolutionary topics gaining high interest among Internet users.

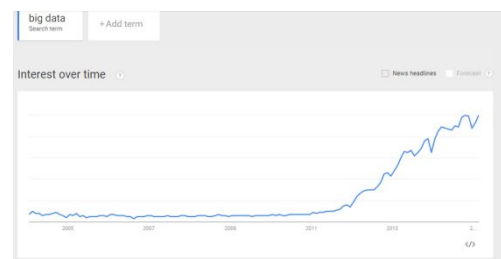


Fig 1: Big Data Interest among Internet Users over Time [15].



According to a study carried out by Gali Halevi and Dr.Henk Moed [8], the growth of research articles about big data has been increasing enormously since 1970’s, i.e., when the term big data was first coined on atmospheric and oceanic surroundings (see Fig 2).

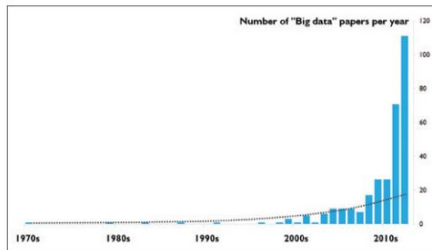


Fig 2: Timeline of Big Data as a Topic of Research [8].

A very interesting finding in his studies is that the conference papers are higher in figures than any other type of documents (see Fig 3 and Fig 4).

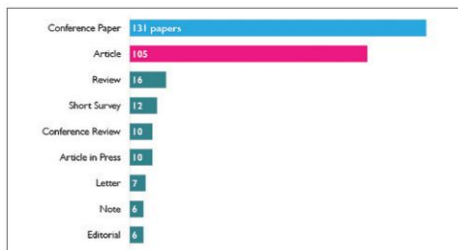


Fig 3: Document Types of Big Data Papers [8].

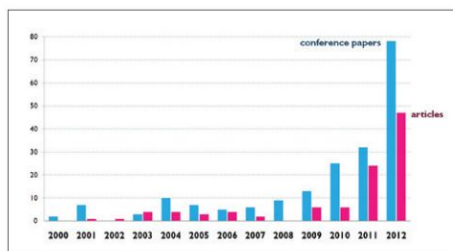


Fig 4: Conference Papers and Article Growth over Time [8].

Even though the top subject area for big data research is computer science, the graph (see Fig 5) also shows the research interests of other disciplines in the development of society by making use of big data. Having a look at the geographical distribution of big data papers, USA is leading followed by China, Germany, Japan and other countries (see Fig 6).

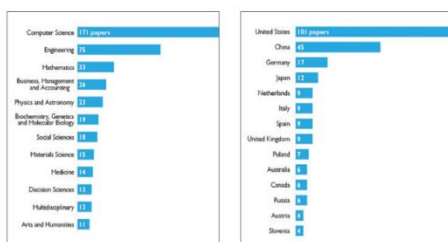


Fig 5: Subject Areas Researching Big Data [8]. Fig 6: Geographical Distribution of Big Data Papers [8].

## CONCLUSION

The big data revolution is impacting almost every aspect of our lives. In this review paper, we have presented our initial study on the impact of big data on finance, commerce and social media. We further explored about how big data handling is taking an interesting turn with various research happening on the big data technology. This paper attempted to touch upon the available technologies for big data and a view on the important research trends in the field of big data analysis like emergence of cloud computing, unstructured data handling etc. The paper also shed light on the limitations of big data and tried to find answers for them. Finally we presented important areas of big data where future research can take direction, like big data modelling, big data as a service etc.

## REFERENCES

- [1]. “Big Data A New World of Opportunities,” NESSI White Paper, December 2012.
- [2]. F. C. P. A. A. M. V. Elena Geanina ULARU, “Perspectives on Big Data and Big Data Analytics,” Database Systems Journal, vol. III, no. 4, 2012.
- [3]. L. J. Yue, “How Big Data can Shape Next-Generation GIS,” Wuhan University, Wuhan, China.
- [4]. “Big Data Development, Challenges and opportunities,” Lean Big Data, 24 February 2014.
- [5]. “Social Network Analytics,” LeanBigData, 5 April 2014.
- [6]. “Data Modelling For Big Data,” CA TechnologieData.
- [7]. “Big Data Model or Not Model?,” SQL Server.
- [8]. D. H. M. Gali Halevi, “Evolution of Big Data as a Research and Scientific Topic,” Research Trends, no. 30, September 2012.
- [9]. C. K. Boyd D, “Critical Questions for Big data,” Information, Communication & Society, no. 15 (5), p. 662, 10 May 2012.
- [10]. “Big data: The Next Frontier for Innovation, Competition, and Productivity,” McKinsey Global Institute.
- [11]. B. D. & C. K, “Six Provocations for Big Data,” in A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, 2012.
- [12]. G. M. & G. M. De Mauro A, “What is Bi Data? A Consensual Definition and a Review of Key Research: Topics,,” in 4th International Conference on Integrated Information, Madrid, 2014.
- [13]. “Emerging Technologies: Big Data - A Horizon Scanning Research Paper,” The emerging technologies big data community of interest.
- [14]. Rossano, “The Pragmatic Side of Using Big Data,” 2013.
- [15]. Google Trends, 2015.