

Concept, Need and Application Areas of Real Time Big Data– At a Glance

Anju Bala¹ and Priti²

¹Research Scholar, DCSA, M.D.U. Rohtak, Haryana (India)
anjusangwan1988@gmail.com

²Assistant Professor, DCSA, M.D.U. Rohtak, Haryana (India)
prish@yahoo.co.in

Publishing Date: April 24, 2018

Abstract

Big data is gaining momentum due to advancement in technology such as Internet, Internet of things, satellite communication and other emerging IT technology. Ninety percent of data is originated in last few couple of years. Due to its rapid growth it has attracted attention from academia, industry, and governments, business around the world. In this paper, we introduce the concept of big data, comprehending its definition, characteristic and challenges faced due to its characteristics and solutions for them. We then define the significance of big data inculcating many areas all over the world. The paper also concludes various applications of big data along with their corresponding challenges. This paper allows the readers to expand application wise work on big data.

Keywords: *Big data, significance, Applications Areas.*

1. Introduction

Quick Google search bring to light millions of pages each with their own spin on the definition of Big Data. Big data is not a new term rather than a new keyword big is added before data. Big data is increasing enormously now a days. The amount of data is exploding at an extraordinary rate as a result of developments in different Web technologies, social media, mobile and sensing devices and satellite communications. As the user requirements/demands are increasing, more and more data is contributed towards data flood. In 2011, McKinsey Global Institute defined big data as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse" Knowledge discovery and better decision making is possible if information is fetched efficiently from this

large collection of data. Big Data is a collection of large data sets that can be processed using traditional tools and techniques. Due to increase in data at colossal rate a number of challenges are faced by traditional system and techniques such as Heterogeneity, Incompleteness, Scale, Timeliness, privacy, security, provenance, visualization, out of context, Lack of talent pool etc.[1]. Compared to traditional data, the features of big data can be characterized by 4V, namely, huge Volume, high Velocity, high Variety and high value[2].

Dimensions of data such as volume (Large amount of data crossing the scale of Peta bytes offers a lot of challenges including Processing Performance, Curse of Modularity, Class Imbalance, Curse of Dimensionality etc.), Velocity is nothing but producing data at an alarming rate for example real time data coming in form of streams. Variety means data is available in different form such audio, video, text, images etc. Value states trust worthiness of data[3][4]

2. Tools of Big data used across different areas

Big data analytics' performance is better than the traditionally existing tools and techniques due to existence of a no. of tools supporting this framework. Some of most commonly used tools are as follows.

1. **Apache Hadoop:** It is open source, distributed, reliable and scalable tool used for processing of data. It is used for distributed processing of large data by

using HDFS (hadoop distributed file system).

2. **Apache Hive:** This tool is used for providing SQL like interface to user. It translates the queries in to hadoop understandable format.
3. **Pig:** It is data flow language. It supports multi query approach by using Pin Latin language which is much like SQL.
4. **Apache Storm:** It is used distributed real time processing of big data. It processes vast amount of data in fault tolerant and scalable method by using master slave architecture.
5. **Map reduce:** It is a programming model for distributed computing, an essential core component of Hadoop, scalable and distributes and clubs the task by using two functions map and reduce.
6. **Spark:** This tool is used for speeding up the computational process by using in memory cluster approach, provide the support for multiple languages.
7. **Sqoop:** SQL to Hadoop and Hadoop to SQL. It is used to transfer the data, to make big data analytics cooperative with existing relational data base system.
8. **Hbase:** It is a column oriented tool used for providing quick access to huge amount of structured data built on the top of hadoop.

Real time applications requires fast responses, to extract the information from this huge stream of data requires effective tools and techniques. Big data underlying framework and tools addresses the challenges (event transfer, analyses, decision making, and response execution) faced during fast and instantaneous response delivery and providing its significance in a number of application areas.

3. Application Areas of Big Data

Response time for real time applications is not equivalent. Different applications require time in

various scales like Military decisions are to taken in seconds, large scale emergency responses should be delivered within minutes, Surveillance and financial markets requires the response within milliseconds [4]. Due to these reasons big data analytics comes into play for real time processing of streamed data to provide the response within time limit.

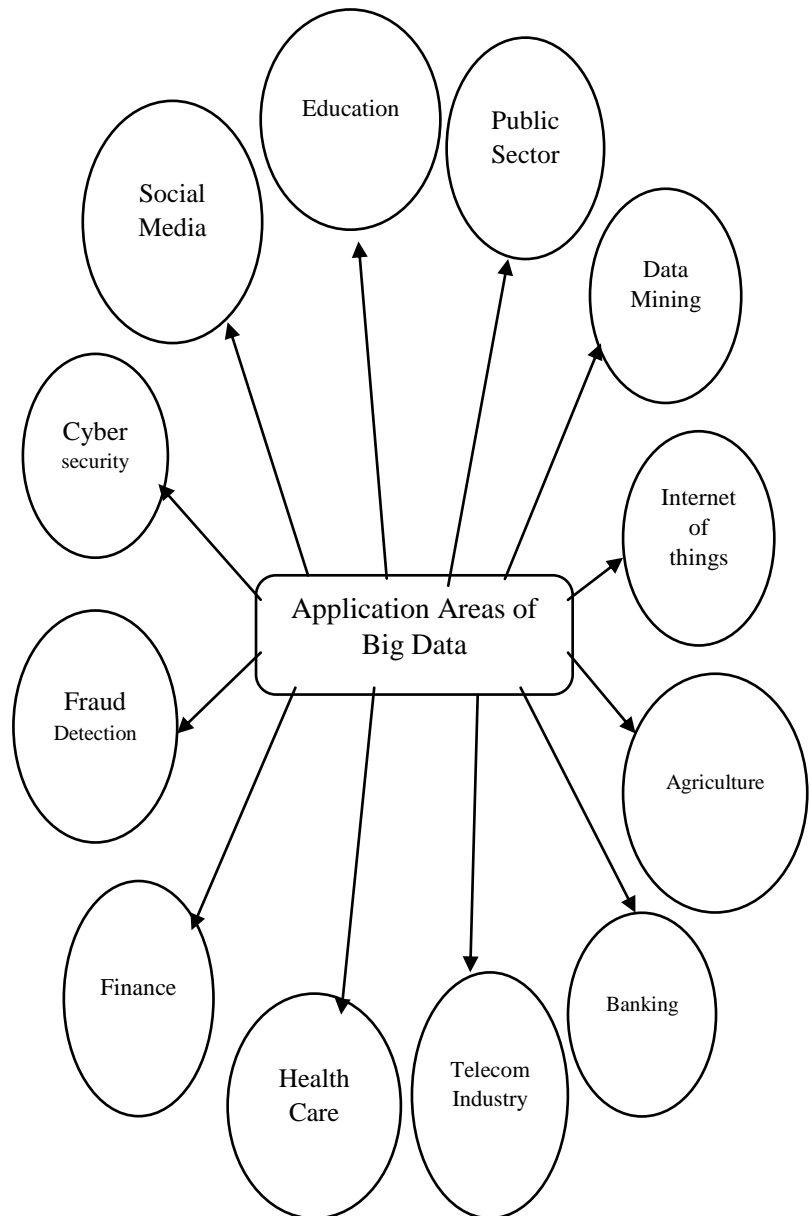


Figure 1: Applications of big data

The field of education is not untouched by the benefits of big data. It has influenced learning practices, method, and way of material collection techniques due to advancement in technology. As, the data involved in education is getting larger and larger day by day due to no of researchers, the applications of Big Data techniques become more and more necessary in learning environments. The way of delivering education is changing constantly with the growth of technology Massive Open Online Courses (MOOCs) are available for students. At this time more than 4,70,000 student in UK are registered for MOOCs there is the ability for academia to gather insights from data in the process of learning itself. Big data has affected education in all of its aspects with its new capability and new tools such as academic analytics, learning analytics, institutional analytics and information technology analytics[5]Big Data techniques can be used in a variety of ways in learning analytics as listed 9below.

With the ability to deal with large volumes, heterogeneity (in form of text, audio, video, image, hypertext) of both structured and unstructured data from different sources, big data analytical tools hold the promise to store and retention of this vast amount of data for analysing purpose and design the predictive model for taking appropriate decision related to health records[6][7].

Big data provides its significance in health care due its tools for addressing the challenges faced in health. A number of solution are available in big data analytics like On-site options such as Teradata, Vertica (HP) and Netezza (IBM) as well as Cloud – hosted Software such as SaaS(SaaS is an important technology for democratizing the results of big data. SaaS-based solutions allow healthcare entities that control subsets of data to expose access through services that eliminate some of the aggregation and integration challenges), Tableau, which supports visualization[8].

With rapid development in E-commerce and ICT online payment system and easy use of credit cards is gaining popularity. Criminal activities are also increasing at the same pace. Fraud can be defined in the following way by(legal dictionary)

1. Wrongful deception with the intent to gain personally or financially.
 2. Intentional deceit in order to persuade another person to part with something of value.
 3. A person who pretends to be something or someone he is not.
- Frauds can be of many types depending upon the situation in which they arise:

Online fraud detection is very tedious task as the no of transaction are increasing day by day generating a big amount of data where the traditional storage and processing system required a lot of time and lost their efficiency. Big Data analytics can help in identification and detection of fraud, it also facilitate digital forensic analysis for better decision making. Big data analytic framework using technologies like Apache Hadoop, Map reduce, Apache Spark, Apache flink, Strom provides valuable support for detection of frauds [9]. Big Data analytics tools have their own advantages by using any of the tools/techniques involve basically the following steps for fetching information/making insights for better decision making:

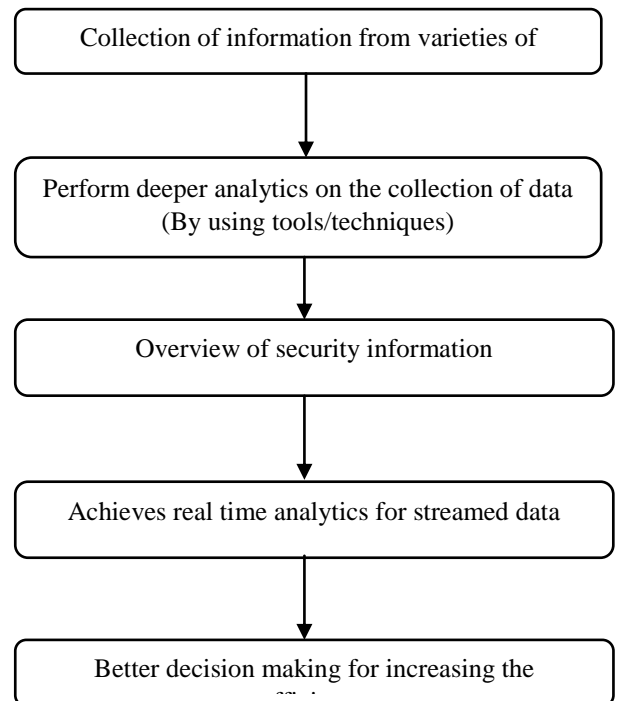


Figure 2: Real Time Analytics [10]

In this era of modern technology every industry and company want to make revolutionary increase in their profits and return on investment. Telecommunication industry is collecting the data from massive data sources including call data , user online data, user location data, wireless basic data, configuration data, report data, GPRS data , warning data, network performance data etc. [11]. Analysis of this huge amount of data is very challenging task for the telecom operators, faced a no of issues due to their traditional tools and techniques. Big data analytic framework provide a unique place for storage , collection and analysis of this data. Big data with the help of big data analytics tools, telecom industry is gaining popularity among the world wide citizens. Big data analytics provide a lot of fruitful outcomes to telecom operators some of which are discussed here[12].

- I. Creates transparency.
- II. Supply chain management
- III. Targeted Marketing
- IV. Predictive analysis
- V. Better understand the customers.
- VI. Improve performance
- VII. Insight generation
- VIII. Data driven decision making

Big Data provides useful output all of above.

4. Conclusion

This paper describes the concept of big data along with different tools used to handle the big data. Then the paper discusses different applications of the big data. It is clear that different tools can be used for distinct applications using big data. Moreover big data can be expanded in different application areas. In future the analysis of big data using different tools can also be done.

References

- [1] Challenges and Opportunities with Big Data A community white paper developed by leading researchers across the United States.
- [2] Xiaolong Jin, Benjamin W. Wah, Xueqi Cheng, Yuanzhuo Wang, Significance and Challenges of Big Data Research, Big Data Research, Volume 2, Issue 2, June 2015, Pages 59-64, ISSN 2214-5796,
- [3] Kuchipudi Sravanthi, Tatireddy Subba Reddy, Applications of Big data in Various Fields, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (5) , 2015, 4629-4632.
- [4] Mohamed N, Al-Jaroodi J. Real-time big data analytics: applications and challenges. In: High performance computing & simulation (HPCS), 2014 international conference; 2014. p. 305–10.
- [5] Daniel, B. (2014). Big data and analytics in higher education: opportunities and challenges. *British Journal of Educational Technology*. doi: [10.1111/bjet.12230](https://doi.org/10.1111/bjet.12230).
- [6] Archana, J., and Anita, M., A survey of big data analytics in healthcare and government. *Procedia Comput. Sci.* 50: 408–413, 2015.
- [7] Ms. K. Gayathri, 2 Dr. N. Uma Maheswari, 3 Dr. R. Venkatesh, 4Mr. B. Ganesh Prabu, "An Overview on Healthcare in Big Data", advances in natural and applied sciences, 2017 May 11(7): pages 536-542.
- [8] Big Data is the Future of Healthcare, white paper.
- [9] Y. Dai, J. Yan, X. Tang, H. Zhao, M. Guo, "Online credit card fraud detection: A hybrid framework with big data technologies", *IEEE Trustcom/BigDataSE/ISPA*, pp. 1644-1651, 2016.
- [10] Sathyapriya, Thiagarasu, "Big Data Analytics Techniques for Credit Card Fraud Detection: A Review", International Journal of Science and Research (IJSR), Volume(2015)
- [11] "Comparing Hadoop, MapReduce, Spark, Flink, and Storm" [online], Available: <http://www.metistream.com/comparing-hadoopmapreduce-spark-flink-storm/>.
- [12] R. Jony et al., "Big Data Use Case Domains for Telecom Operators", IEEE Int'l. Conf. on Smart City/SocialCom/SustainCom (SmartCity), 2015.