

Scope of Big Data Analytics in Industrial Domain

Krishna R. Shah¹, Dr. Seema Shah²

¹Student, B. Tech Integrated Computer Science, Mukesh Patel School of Technology Management and Engineering, NMIMS University Mumbai, India

²Professor, Department of Computer Science, Mukesh Patel School of Technology Management and Engineering, NMIMS University Mumbai, India

Abstract - As the name implies, Big Data literally means large collection of data sets containing abundant information. Data is generated in huge terabytes of amount daily from modern technologies such as cloud computing and Internet of Things. Analysis of this data is crucial with the difficulty in managing such big data. Basic characteristics of big data show its potential. There are different characteristics famous as "V's" of big data. Big Data is recently becoming a trending practice that many industries and organizations are adopting. The objective of this paper is to understand the characteristics of big data which are increasingly evolving day by day and the impact of big data analytics in various sectors.

1. INTRODUCTION

Imagine a world without data storage; a place where every detail about a person or organization, every transaction performed, or every aspect which can be documented is lost directly after use. Organizations would thus lose the ability to extract valuable information and knowledge, perform detailed analyses, as well as provide new opportunities and advantages. Anything ranging from customer names and addresses, to products available, to purchases made, to employees hired, etc. has become essential for day-to-day continuity. Data is the building block upon which any organization thrives. Now think of the extent of details and the surge of data and information provided nowadays through the advancements in technologies and the internet. With the increase in storage capabilities and methods of data collection, huge amounts of data have become easily available. Every minute, more and more data are being created and needs to be stored and analyzed in order to extract value. Furthermore, data has become cheaper to store, so organizations need to get as much value as possible from the huge amounts of stored data. The size, variety, and rapid change of such data require a new type of big data analytics, as well as different storage and analysis methods. Such sheer amounts of big data need to be properly analyzed, and pertaining information should be extracted [5]. Big data term is nowadays used all over the world in every field though it is any forum or organization. Big data is nothing else but data which is in large volume that requires advance technologies to handle as existing traditional technologies cannot manage such enormous datasets, for extracting useful value information.

2. RELATED WORK

Formally, Big Data is defined from 3V's to 4V's. 4V's refers to volume, velocity, variety and veracity that is discussed in detail in [6]. However, 7V's are discussed here. The following paper [6][7] talks about 4V's i.e. characteristics of big data however this paper goes deep into discussion of characteristics and discusses 7 characteristic V's. Section III focuses on big data characteristics; Section IV shows use of big data analytics in different sectors.

3. CHARACTERISTICS OF BIG DATA

Back in 2001, Gartner analyst Doug Laney listed the 3 'V's of Big Data – Variety, Velocity, and Volume. Further the invention of V's kept increasing. Later in 2019 7 'V's were considered for characteristics of big data. 1) **Volume** – The name Big Data itself is connected to a huge volume. Volume relates to the enormous quantity of information produced daily. Data size plays a very important role in determining data value. Also, whether a particular data can actually be considered as a Big Data or not, is dependent upon the volume of data. Hence, 'Volume' is one characteristic which needs to be considered while dealing with Big Data. 2) **Velocity** – The term 'velocity' refers to the speed of generation of data. How fast the data is generated and processed to meet the demands, determines real potential in the data. Big Data Velocity deals with the speed at which data flows in from sources like business processes, application logs, networks, and social media sites, sensors, mobile devices, etc. The flow of data is massive and continuous. 3) **Variety** – The next aspect of Big Data is its variety. Variety refers to heterogeneous sources and the nature of data. It provides information about the types of data such as structured, unstructured, semi-structured etc. During earlier days, spreadsheets and databases were the only sources of data considered by most of the applications. Nowadays, data in the form of emails, photos, videos, monitoring devices, PDFs, audio, etc. are also being considered in the analysis applications. This variety of unstructured data poses certain issues for storage, mining and analyzing data. 4) **Variability** – This refers to the inconsistency which can be shown by the data at times, thus hampering the process of being able to handle and manage the data effectively. 5) **Veracity** – Big Data Veracity refers to the biases, noise and abnormality

in data. Is the data that is being stored, and mined meaningful to the problem being analyzed. Veracity in data analysis is the biggest challenge when compared to things like volume and velocity. In scoping out your big data strategy you need to have your team and partners work to help keep your data clean. 6) **Value** – That is, if you’re going to invest in the infrastructure required to collect and interpret data on a system-wide scale, it’s important to ensure that the insights that are generated are based on accurate data and lead to measurable improvements at the end of the day. 7) **Volatility** – Big data volatility refers to how long is data valid and how long should it be stored. In this world of real time data, you need to determine at what point is data no longer relevant to the current analysis. Big data clearly deals with issues beyond volume, variety and velocity to other concerns like variability, veracity, value and volatility. The prime objective of big data analysis is to process data of high volume, velocity, variety, veracity, value, variability and volatility using various traditional and computational intelligent techniques. The following Figure 1 refers to the definition of big data.

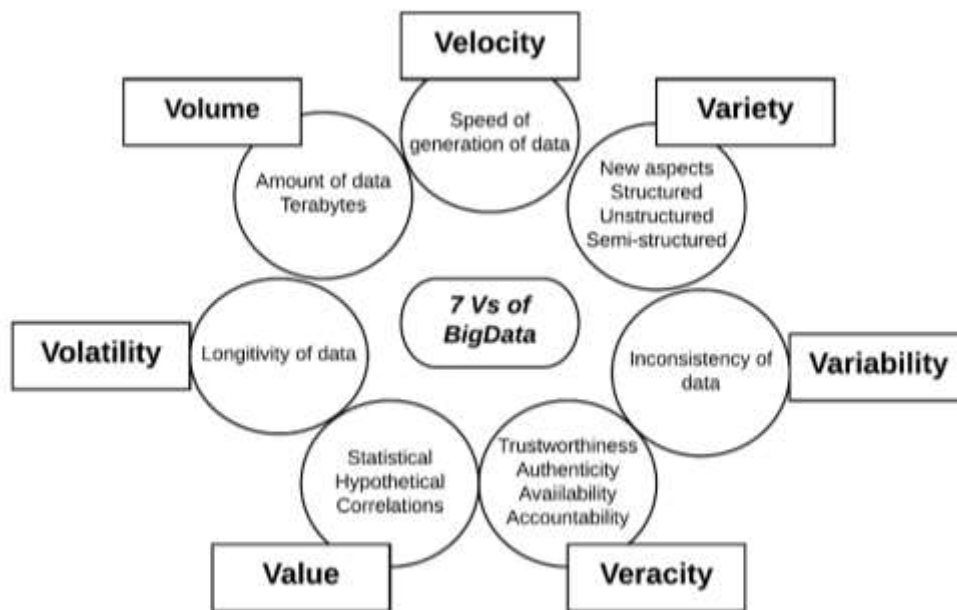


Fig 1: Characteristics of Big Data

4. FIELDS THAT GENERATE DATA

All most all fields generate big data. Some major fields where big data plays a major role is

- Social networking sites: social media that carry information, posts, links etc of different peoples from all over world like Facebook twitter etc.
- Search engines: there are lots of data from different databases that retrieve from search engines.
- Medical history: medical history of patients for various health issues from hospitals
- Online shopping: shopping online help to know the preferences of customers on different products.
- Stock exchange: shares of different companies hold by stock

5. BIG DATA IN INDUSTRIAL SECTORS

- Big Data Analytics in IoT

Internet has restructured global interrelations, the art of businesses, cultural revolutions and an unbelievable number of personal characteristics. Currently, machines are getting in on the act to control innumerable autonomous gadgets via internet and create Internet of Things (IoT). Thus, appliances are becoming the user of the internet, just like humans with the web browsers. Internet of Things is attracting the attention of recent researchers for its most promising opportunities and challenges. It has an imperative economic and societal impact for the future construction of information, network and

communication technology. The new regulation of future will be eventually, everything will be connected and intelligently controlled. The concept of IoT is becoming more pertinent to the realistic world due to the development of mobile devices, embedded and ubiquitous communication technologies, cloud computing, and data analytics. Moreover, IoT presents challenges in combinations of volume, velocity and variety. In a broader sense, just like the internet, Internet of Things enables the devices to exist in a myriad of places and facilitates applications ranging from trivial to the crucial. Several diversified technologies such as computational intelligence, and big-data can be incorporated together to improve the data management and knowledge discovery of largescale automation applications. Much research in this direction has been carried out by Mishra, Lin and Chang [2]. Knowledge acquisition from IoT data is the biggest challenge that big data professionals are facing. Therefore, it is essential to develop infrastructure to analyses the IoT data. An IoT device generates continuous streams of data and the researchers can develop tools to extract meaningful information from these data using machine learning techniques. Understanding these streams of data generated from IoT devices and analyzing them to get meaningful information is a challenging issue and it leads to big data analytics. Machine learning algorithms and computational intelligence techniques is the only solution to handle big data from IoT prospective.

- Big Data Analytics in Health Sector

By definition, big data in healthcare refer to electronic health data sets so large and complex which are difficult to manage by traditional software or hardware neither by any traditional tools and methods. Big data analytics plays a vital role in health sector. Benefits of health with related to big data are demonstrated in 3 areas namely to prevent disease, identify risk factor for disease, define intervention for health behavior change. The health care from age has generated voluminous amount of data in the form of records, regulatory requirements, patient care etc. This data is stored in hard copy form most commonly but now everything is rapidly turning to digitization. This reduces the quality of healthcare meanwhile reducing the cost. Big data supports wide range of medical and health care functions to find any previously untapped intelligence. By understanding patterns and trends within the data, big data scientists by the help of big data analytics could improve care, save lives as well as reduce cost.

- Big Data Analytics in Banking

Over a period of time banking systems have undergone some intense process of invention and innovation by which it has allowed bank to diversify their activities, to create new as well as complex products. The banking industry is generating huge amount of data day by day where previously this data was failed to utilize by banks. But nowadays, banks are using this data to reach the main objective of marketing. This data is unlocking secrets of money movements, helping to prevent major disaster and frauds as well as it helps to understand customer's behavior. This banking and financial industry is one of the biggest adopters of big data technologies. Banks internationally have started to harness the power of data to derive utility across various parts of their functioning. Big data in financial industry is defined as tool that allows an organization to create, manipulate, and manage large sets of data in a given time and the storage that supports such voluminous data.

6. CONCLUSION

In recent years data are generated at a dynamic pace. Analyzing these data is a challenge to mankind. The paper talks about all kinds of industries that use big data analytics. The flowchart diagram helps to understand the characteristics at a glance and a detailed information of the sectors that use and manage big data. We believe that in future, researchers will pay more attention to these new evolving characteristics to solve problems of big data analysis effectively and efficiently.

REFERENCES

- [1] M. K.Kakhani, S. Kakhani and S. R.Biradar, Research issues in big data analytics, International Journal of Application or Innovation in Engineering & Management, 2(8) (2015), pp.228-232.
- [2] N. Mishra, C. Lin and H. Chang, A cognitive adopted framework for iot big data management and knowledge discovery prospective, International Journal of Distributed Sensor Networks, 2015, (2015), pp. 1-13
- [3] M. D. Assuno, R. N. Calheiros, S. Bianchi, M. a. S. Netto and R.Buyya, Big data computing and clouds: Trends and future directions, Journal of Parallel and Distributed Computing, 79 (2015), pp.3-15.
- [4] I. A. T. Hashem, I. Yaqoob, N. Badrul Anuar, S. Mokhtar, A. Gani and S. Ullah Khan, The rise of big data on cloud computing: Review and open research issues, Information Systems, 47 (2014), pp. 98-115.
- [5] Nada Elgendy and Ahmed Elragal, Big Data Analytics: A Literature Review Paper, Conference Paper, August 2014

[6] D. P. Acharjya, Kauser Ahmed P, A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 2, 2016

[7] Pritee Chunarkar-Patil, Akshanda Bhosale, Big data analytics, Review Article, Open Access Journal of Science, September 24, 2018

[8] Althaf Rahaman.Sk, Sai Rajesh.K, Girija Rani K, Challenging tools on Research Issues in Big Data Analytics, 2018 IJEDR