

Big Data

Tejashree Bhong^{1*} Poonam Jadhav² Janhavi Tupe³ Vaishnavi Pawar⁴ Swati Mane⁵
Mr. Godase R C.⁶

^{1,2,3,4,5} SY Students, Department of Computer Engineering, Sahakar Maharshi Shankarrao Mohite Patil Institute of Technology and Research, Akluj, Solapur, Maharashtra, India

⁶ Lecturer, Department of Computer Engineering, Sahakar Maharshi Shankarrao Mohite Patil Institute of Technology and Research, Akluj, Solapur, Maharashtra, India

Abstract – Big data is explained as huge amounts of data that processing unit and traditional data storage can't store and process and it is a term that has become recently popular. Data in Petabytes can be considered as big data. big data, according to gartner means – “Big data” is velocity, high-volume, or variety information assets that demand low cost, innovative forms of information processing for enhanced insight and decision making.” However, this huge amount of data offers useful statistics for organizations when assessed appropriately with modern equipment, so that it can help them to make better decisions to improve their business.

Keywords – Data, Comprehend, Innovative

-----X-----

INTRODUCTION

Collection of data sets so huge or complex that it make difficult to process using hand database management system and traditional data processing applications. Cluster formation in Big Data were discovered in March 2014 at America Society of Engineering Education. Now usage of term Big Data trends to the use of user behavior analytics, use of predictive analytics, and certain other advanced data analytics methods that value from huge information or seldom to a specific length of information set. Large quantity of information produced with the aid of using human sports and system operations, the information are so complex and substantial that they cannot be comprehended with the aid of using human beings or analyzed with an identical relational database.

TYPES OF BIG DATA:

There are three types of big data – 1. Structured, 2. Unstructured, or 3. Semi-Structured. Each type has a particular purpose. Below given is the explanation of each type.

1. **Structured data** – Any data that have some pre-determined and formats and fixed organizational properties formats, and it can be analyzed, processed, or stored is called as structured data. It is easy to sort and evaluate. Because of its fixed format, each field is unique and can be retrieved individually and in combination with data

from other fields. Hence, it allows the rapid collection of data from multiple locations. Over time, geniuses in laptop technology have had wonderful achievement in growing technology and extracting price from them, running with such constant data.

2. **Unstructured data** – unlike structured data, any data with pre-defined and no particular format is considered unstructured data. Unstructured data consists of information such as dates, facts, and numerals. In addition to being massive in size, unstructured facts poses diverse demanding situations in terms of processing to derive fee from it. Pictures we post on Instagram or facebook and videos we watch on Youtube or other platforms are examples of unstructured data. Though a huge amount of data is available with organizations, they have no idea of obtaining value from it as the data is in its raw form.
3. **Semi-structured data** – semi-structured data is a combination of unstructured and structured data, which shows that it has the characteristics of saturated and unsaturated data forms. It consists of data that fails to have a specific structure and does not match the relational databases.

CHARACTERISTICS OF BIG DATA:

Now we know the types of big data, let us see the characteristics of big data. The characteristics of big data can be divided into 5 Vs – Variety, Velocity, Value, and Veracity, Volume

- **Volume** – Main characteristic of any data is the size of the data which is measured in exabytes, petabytes. The amount of data created or stored in a big data system is referred to as its Volume. These large amounts of data require the use of much more powerful advanced processing technology, and powerful, better than a standard laptop or desktop CPU. The best example for such a massive volume of data is found in Instagram or twitter, where the audience spends a lot of time watching videos, posts, commenting, etc. With this ever-exploding data, there may be amazing capacity for analysis, sample discovery, and more.
- **Variety** – Variety consists of information kinds that fluctuate in layout and the manner it's far based and made equipped for processing. Top media companies like Pinterest, Google and others generate data that can be stored and analyzed later. If databases and spreadsheets were the main sources in the initial data, pictures, videos, PDFs, emails, etc., have turn out to be extra outstanding those days.
- **Velocity** – Rate at which data is gathered affects whether the way data is categorized as large data or general data. Much of this data should be accessed in real-time, in order to enable the systems to handle the speed and quantity of data generated. The processing speed of the data refers to the availability of more data than before, at the same time, it also indicates that the data processing speed must be much more.
- **Value** – Another significant issue to be considered is Value. It is not just the amount of data we process or store that matters. It is about the reliability and value of the data and also about processing, storing and evaluating data in order to obtain statistics.
- **Veracity** – This refers back to the reliability and exceptional of the records. The price of huge records can't be puzzled if it has dependable capabilities and it meets the very best exceptional. It is almost true in the case of working with data that is updates in live time. Thus, the authenticity of data needs to be verified and balanced at all levels of the process of Big Data collection and processing.

ADVANTAGES OF BIG DATA:

1. Enhanced customer experience and satisfaction.
2. Improved functional efficiency.
3. Businesses will be able to use outside intelligence to make decisions.
4. Early risk detection associated with products and services.
5. Improved functional efficiency.

CONCLUSION:

We live in an advanced world where many technologies emerge every day. Technology is continuously attacking us in all areas of our lives. As the big data is generating on a large scale, it can become a significant asset for various companies and organizations, thus offering help to them to find new statistics and improve their businesses. In recent decades, there has been a huge growth of data with the increased use of mobile phones, social networks, streaming videos, and the Internet of Things platform.

REFERENCE

1. Adams, M.N. (2010). Perspectives on Data Mining. International Journal of Market Research 52(1), pp. 11–19.
2. Asur, S., Huberman, B. A. (2010). Predicting the Future with Social Media. In: ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 1, pp. 492–499.
3. Bakshi, K. (2012). Considerations for Big Data: Architecture and Approaches. In: Proceedings of the IEEE Aerospace Conference, pp. 1–7.
4. Cebr (2012). Data equity, unlocking the value of big data in: SAS Reports, pp. 1–44.
5. Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J.M., Welton, C. (2009). MAD Skills: New Analysis Practices for Big Data. Proceedings of the ACM VLDB Endowment 2(2), pp. 1481–1492.
6. Cuzzocrea, A., Song, I., Davis, K.C. (2011). Analytics over Large-Scale Multidimensional Data: The Big Data Revolution! In: Proceedings of the ACM International Workshop on Data Warehousing and OLAP, pp. 101–104.

7. Economist Intelligence Unit (2012): The Deciding Factor: Big Data & Decision Making. In: Capgemini Reports, pp. 1–24.

Corresponding Author

Tejashree Bhong*

SY Students, Department of Computer Engineering,
Sahakar Maharshi Shankarrao Mohite Patil Institute
of Technology and Research, Akulj, Solapur,
Maharashtra, India