

# Overview of Big Data

Shivam Kashyap<sup>1</sup>, Amit Choudhary<sup>1</sup>, Dr. Deepak Chahal<sup>2</sup>

MCA Student, Department of IT, Jagan Institute of Management Studies, Sector-05, Rohini, New Delhi, India<sup>1</sup>

Professor, Department of IT, Jagan Institute of Management Studies, Sector-05, Rohini, New Delhi, India<sup>2</sup>

**ABSTRACT:** In today's generation, as the internet has been revolutionised by the way IT industries are also growing day by day. The size of data on which the industries are working is petabyte and exabyte. Data size has been growing exponentially by various means such as mobile phones, remote sensing, aerial sensory technologies e.t.c. The study shows that big data has revolutionised the era of IT industries by analysing big data in account of various technologies for better decision making in developing a particular industry. This article also tells how big data is beneficial and how data can be managed or analysed by a particular software.

**KEYWORDS:** Complexity, structured, volume, velocity, variety, variability, analytics, predictive, statistical.

## I. INTRODUCTION

In digital era, data is growing at a rapid scale from digital sources which leads to the growth of big data, which is a large collection of data. Big Data is a term which is used to refer to the enormous amount of data that is complex in its own way that the traditional database management systems and software techniques are unable to process it. Big data are also the data sets that are so big and complex that traditional data process or software can't deal or process it. It basically includes those data sets with an extremely large sizes which can be beyond the ability with commonly used software tools can process it.



Fig.1 Use of Big Data

It includes the semi-structured or unstructured data, i.e., the information that doesn't have a predefined data model and is not organised in a structured manner. It requires a set of techniques and technologies with new forms that could execute diverse data sets, complexity & at a massive scale. Security is looked in terms of how data is stored, coded, transmitted, encrypted and deleted. Various statistics has shown that companies take security of the data of an individual with very high priority[1].

## II. CHARACTERISTICS OF BIG DATA

Big Data is characterised having the triplets of V's, i.e., 3V's. These 3 V's were Volume, Velocity & Variety. Later on as technology grows rapidly at a pace, the fourth V was also added which is Variability.

1. **Volume** :- In big data, Volume is the size of data that decides size limit, that is data is big or not. The data is either in Tb's or in Pb's. Massive data needs to be analysed to extract useful information.

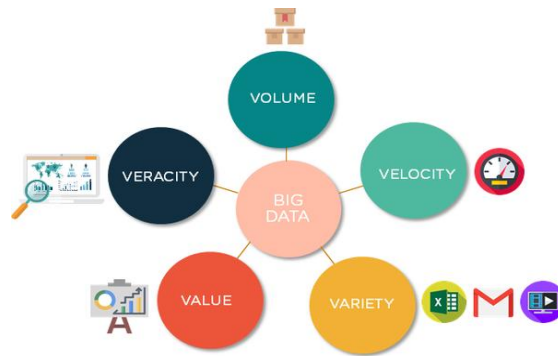


Fig.2 : Types of Big Data

2. Velocity :-In big data , velocity is the speed at which the data is generated or processed to fulfil the requirements of the industry . Data is growing equivalent to the growth of the digital world .The rate of growth of the data is directly proportional to the growth of digitalization .
3. Variety :-In this characteristic , variety specifies the type & nature of a data , that is text , image ,audio & video . It also determines the data is structured , semi – structured or unstructured .
4. Variability :- Variability , which is also known as Veracity defines the quality and accuracy of the data which is captured . It includes the availability as well as the accountability of data .

#### **A.BIG DATA ADVANTAGES**

These days in business as well as ITsector , every industry uses big data and become too much beneficial . The benefits of big data are :-

- Big data provides better decision making as it is highly used in data analytics and provide better clarity for attainment of goal in business and provide proper growth of companies.
- Increased productivity is another benefit in Big data as it helps users improving business productivity which use big data tools like Hadoop.
- Big data is also helpful for identifying the root causes & failures that occurs in real time .
- Greater Innovation is the another benefit of big data which helps employees investing in data analytics and bring out new products and services to the market .
- Big data is also useful in improving customer engagement and increasing customer trust as well as loyalty.
- Big data is also beneficial for increasing the revenue of various organisations.
- Big Data is most beneficial in financial tradings , sports , weather forecast , predictions ,security, polling e.t.c

#### **B.BIG DATA DISADVANTAGES**

There are many drawbacks that occurs while implementing the big data initiatives . These are :-

- There is a lack of a talent which is good in data analysis and other data skills . It is difficult to find a skilled data scientist as it takes time to train or acquire.
- Sometimes data can be founded unstructured which is difficult to manage or make it structured .
- Big data results tends to be mislead if data is not analysed properly .
- Big Data is not useful for analysing data for smaller duration as it needs a large amount of time to be analysed or being processed .
- Big data can take higher cost or memory to be stored as it doesn't store in a disk .
- It also tends to violate the privacy principles as third party can access the data .

### C. ANALYTICS IN BIG DATA

Data can be defined as a collection of characters that is organized to serve a purpose. It can be anything ranging from text, number to images, videos [2]. Analytics in big data is the method or a technique to test large amount of data to identify hidden patterns, correlations & other insights. The process for analysing big data is complex as it examines large and varied portion of data to uncover the information.

Big data analysis is very helpful for organisations to provide better business based decisions. As big data is complex, it involves complex norms and applications with elements like predictive models & statistical algorithms, that are helpful for high performance analytics systems.

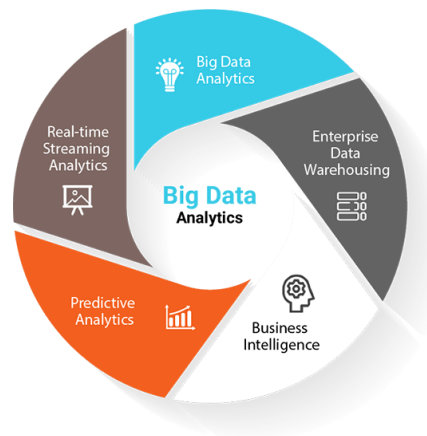


Fig 3 :Working of Big Data Analytics

In Big data analytics, data can be analysed directly in Hadoop cluster (as big data analysts use Hadoop data lake to form a data repository to store raw data) & then sound data management technique is applied. and then data can be stored in HDFS which can further be organised, partitioned and refined. Then performance should be checked and it can be loaded as well as extracted. Once the data is ready, it can be analysed with software for further processing which can be done by :-

- Data mining
- Predictive Analysis
- Machine Learning
- Deep Learning

### D. BIG DATA TOOLS

The amount of data collected from different applications all over the world across a wide variety of fields. Since big data is too large and complex, it can't be dealt with traditional data processing methods, it requires a set of tools to get analysis & gain insights from them.

In market, there are many number of tools and technologies available to handle big data and its analysis and these tools are :-

1. Hadoop
2. Spark
3. Storm
4. Apache Cassandra
5. MongoDB
6. Statwing
7. CouchDB
8. Flink
9. Cloudera



10. Drill
11. Plotly
12. Splunk

### ***E.APPLICATIONS OF BIG DATA***

A focus is made on machines as machines cannot be understood by verbal communication it forms abstractions and concepts [3]. Today , big data is very useful in every field . The whole world is dependent of enormous data as it makes them aware and known about a particular thing . So let us discuss applications of big data one by one .

1. In government sector , as it is helpful to predict the leader or winning party by surveys or responds by people .
2. In social media analytics by providing real time insights in promotions , marketing advertisements e.t.c .And by these , various companies can decide pricing and availability of product.
3. Technology:-various online business can use big data technologies to handle the back –end operations ,queries user ends.
4. In smart phones ,by face recognition and fingerprints ,also provides backup or reminder of previous stored data .

### **III.CONCLUSION**

So , the article concludes that big data creates major impact in IT and business industry . It creates a new revolution of data and the whole world is dependent on data . Big data can be managed by various tools and provides useful information in like better decision making and improved marketing. Analysis of big data is also important to process it and better understanding . The opportunities offered by big data are also unlimited.

### **REFERENCES**

1. Bhutani S. et al, Data privacy and security issues in India: An empirical study, International Journal of Research in Engineering, Volume 1; Issue 4; October 2019; Page No. 15-17.
2. Chahal D. et al. A Survey on Cryptography, Encryption and Compression Techniques, International Research Journal of Engineering and Technology, Volume: 06 Issue: 11 | Nov 2019.
3. L. Kharb et al (2019) “Brain Emulation Machine Model for Communication” in International Journal of Scientific & Technology Research (IJSTR). pp 1410-1418.