Big Data Overview

Neenu Ann Sunny

Assistant Teacher
Department of Computer Applications
Saintgits College of Applied Sciences Pathamuttom, Kerala, India

***Abstract*: Big data is a new driver of the world economic and societal changes. The world's data collection is reaching a tipping point for major technological changes that can bring new ways in decision making, managing our health, cities, finance and education. While the data complexities are increasing including data's volume, variety, velocity and veracity, the real impact hinges on our ability to uncover the `value' in the data through Big Data Analytics technologies. Big Data Analytics poses a grand challenge on the design of highly scalable algorithms and systems to integrate the data and uncover large hidden values from datasets that are diverse, complex, and of a massive scale. Potential breakthroughs include new algorithms, methodologies, systems and applications in Big Data Analytics that discover useful and hidden knowledge from the Big Data efficiently and effectively.**

***Keywords*: big data, big data characteristics, Type of big data, general challenges of big data.**

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.[1]. The big data era has led to the creation of a major trend towards the use of the amounts of data which is kept in the systems of big companies or on the internet and also different kinds of data produced by companies in the real world, people and other parts of the society. Thus, delay in the field of big data's research and applications not only mean the loss of industrial strategic advantage but also it is considered a weak point in the security of the virtual environment. At the moment, big data directly affects all of the aspects of modern society including business, government, healthcare, and research in any field. So, companies and organizations will improve and will be posed as a model for the analysis of services through the exploitation of potentials existing in big data [2].

2. What is Big Data?

Big data refers to the large, diverse sets of information that grow at ever-increasing rates. It encompasses the volume of information, the velocity or speed at which it is created and collected, and the variety or scope of the data points being covered. Big Data is also data but with a huge size. Big Data is a term used to describe a collection of data that is huge in volume and yet growing exponentially with time. In short such data is so large and complex that none of the traditional data management tools are able to store it or process it efficiently. In fact, the real nature and purpose of big data is the ability to use information obtained from different large amounts of the data which existed in the past heterogeneously and was always ignored [2]. But, before considering any project which deals with big data it is necessary for this big data to be organized and for the appropriate mechanisms of protection of the data to be applied from the beginning in order to avoid the drawbacks which are resulted from the data's flaw disastrous [2].

3. Characteristics of Big Data

Big data has major characteristics including big volume, variety, velocity, and veracity, value.

Figure: 1

3.1 Volume

Big data indicates a big volume of data which turn into big data through mechanisms such as human-computer interface, mobile phone internet, and brain-machine interface [2]. Volume refers to the unimaginable amounts of information generated every second from social media, cell phones, cars, credit cards, M2M sensors, images, video, and whatnot. We are currently using distributed systems**,** to store data in several locations and brought together by a software Framework like hadoop[8].

3.2 Variety

As Discussed before, Big Data is generated in multiple varieties. Compared to the traditional data like phone numbers and addresses, the latest trend of data is in the form of photos, videos, and audios and many more, making about 80% of the data to be completely unstructured [8]

3.3 Veracity

Veracity basically means the degree of reliability that the data has to offer. Since a major part of the data is unstructured and irrelevant, Big Data needs to find an alternate way to filter them or to translate them out as the data is crucial in business developments.[8]

3.4 Value

Value is the major issue that we need to concentrate on. It is not just the amount of data that we store or process. It is actually the amount of valuable, reliable and trustworthy data that needs to be stored, processed, analyzed to find insights.[8]

3.5 Velocity

Last but never least, Velocity plays a major role compared to the others, there is no point in investing so much to end up waiting for the data. So, the major aspect of Big Dat is to provide data on demand and at a faster pace.[8]

4. Types of Big Data

Three different varieties

- Structured Data
- Semi-Structured Data

- Unstructured Data

4.1 Structured Data

Structured Data is used to refer to the data which is already stored in databases, in an ordered manner. It accounts for about 20% of the total existing data and is used the most in programming and computer-related activities.There are two sources of structured data- machines and humans. All the data received from sensors, weblogs, and financial systems are classified under machine-generated data. These include medical devices, GPS data, data of usage statistics captured by servers and applications and the huge amount of data that usually move through trading platforms, to name a few.Human-generated structured data mainly includes all the data a human input into a computer, such as his name and other personal details. When a person clicks a link on the internet, or even makes a move in a game, data is created- this can be used by companies to figure out their customer behavior and make the appropriate decisions and modifications.

4.2 Semi-Structured Data

The line between unstructured data and semi-structured data has always been unclear since most of the semi-structured data appear to be unstructured at a glance. Information that is not in the traditional database format as structured data, but contains some organizational properties which make it easier to process, are included in semi-structured data. For example, NoSQL documents are considered to be semi-structured, since they contain keywords that can be used to process the document easily.Big Data analysis has been found to have definite business value, as its analysis and processing can help a company achieve cost reductions and dramatic growth. So it is imperative that you do not wait too long to exploit the potential of this excellent business opportunity.

4.3 Unstructured Data

While structured data resides in the traditional row-column databases, unstructured data is the opposite- they have no clear format in storage. The rest of the data created, about 80% of the total account for unstructured big data. Most of the data a person encounters belong to this category- and until recently, there was not much to do to it except storing it or analyzing it manually.Unstructured data is also classified based on its source, into machine-generated or human-generated. Machine-generated data accounts for all the satellite images, the scientific data from various experiments and radar data captured by various facets of technology.Human-generated unstructured data is found in abundance across the internet since it includes social media data, mobile data, and website content. This means that the pictures we upload to Facebook or Instagram handle, the videos we watch on YouTube and even the text messages we send all contribute to the gigantic heap that is unstructured data.Examples of unstructured data include text, video, audio, mobile activity, social media activity, satellite imagery, surveillance imagery – the list goes on and on.

The following image will clearly help you to understand what exactly unstructured data is

The Unstructured data is further divided into –

- Captured
- User-Generated data

a. *Captured data*

It is the data based on the user's behavior. The best example to understand it is GPS via smartphones which help the user each and every moment and provides a real-time output.

b. User-generated data:

It is the kind of unstructured data where the user itself will put data on the internet every movement. For example, Tweets and Re-tweets, Likes, Shares, Comments, on Youtube, Facebook, etc.

5. General Challenges of Big Data

Considering the mentioned key features, it is seen that big data is always facing with several challenges and among them some are created through inherent characteristics of big data in the stages of data achievement,storage, processing, transfer,and management Challenges regarding size, implementation, continuous access, veracity of the work load, security, and cost are created by current models and approaches for the analysis of big data and limitations of the current processing systems . Thus, a lot of claims regarding big data and its analysis and views point to the proverb that studying challenges of big data is like finding a needle in a hay stack [2].

CONCLUSTION

Big Data is BackBone of Devolopment of today's Cyber World.Ever Enterprise and Gigantic Compaines are Totally dependent on this technology for storing data and their analysis.

REFERENCES

[1]. Big Data Analytics:A Literature Review Paper. Nada Elgendy and Ahmed Elragal.
[2] Majlesi journal of multimedia processing volu.4 No.1 march 215.Bigdata and its challenges.
[3] Yasutomo Yamamoto, "It Platforms for utilization of big data", *Hitachi Review* Vol. 63 (2014), No. 1.
[4] Xiaolong Jin, Benjamin W. Wah, Xueqi Cheng, Yuanzhuo Wang, "Significance and challenges of big data research" , *Science Direct* , article in press
[5] H.V. Jagadish, "Big data and science: Myths and reality", *Science Direct*, Article in press
[6] 4 Big Data Challenges that universities face– Tanya Roscorla / April 9, 2014, available at www.centerdigitaled.com
[7] Fully – Day Tutorial – Big Data: Finding the needle in the haystack enter preners, available at www.hicss.hawaii.edu/hicss46_ /swt46/BigData.pdf
[8] Big Data Characteristics:know the 5'vs of big dataTech Enthusiast working as a Research Analyst at Edureka. Curious about learning more about Data Science and Big-Data Hadoop.Ravi kiran