



BIG DATA - A PROBLEM OR A SOLUTION

Shaikh Aamer

Student at Computer Science Department
P.E.S. College Of Engineering, Aurangabad,
Maharashtra, India.

Razvi Zeba

Student at Computer Science Department
P.E.S. College Of Engineering, Aurangabad,
Maharashtra, India.

Abstract - Data is a most powerful and efficient unit of information. Data is considered as an important raw material. Its impact on the government performances and research endeavors. Big Data is a term which defines the huge amount of complex data sets which are difficult to process using the conventional processing tools. Big Data is characterized by high volume, velocity, variety and veracity. This type of data becomes very essential source for the valuable insights. It helps to make more powerful and precise decisions. This paper provides the overview of the Big Data and discuss whether it is a problem or a solution. If Big Data is a problem then we will also discuss about it's solutions.

KEYWORDS - Big Data, Data Analytics, Applications of Big Data, OLAP, OLTP, SAP HANA, IBM, Oracle, Apache Hadoop, Big Data Problems, Big Data Solutions.

I. INTRODUCTION

As we are living in a full of digital world today. The Data is generated in a very huge amount. The term Big Data means that we are dealing with the huge amount of data which cannot be processed by using the conventional processing tools. The main sources of generating this data is social media, mobile phones, sensors, medical diagnosis, photos, documents, web logs, IOT devices, Online transactions like banks and Other and many more.

Also now-a- days whatever data is getting generate in the world most of it comes under the Big Data. So, many of us thinking that we are generating huge amount of data that's the good thing that we are getting more data. But, the real scenario behind the scene is that in actual, the Big Data is the biggest problem that we are facing. Because the data has different varieties and it's tedious to manage get removed by veracity of data.

II. DEFINATIONS OF BIG DATA

According to several technologies leaders like Microsoft, IBM, SAP, Cloudera, Oracle there are many definitions of Big Data. They are as follow –
According to SAP the definition of Big Data is: OLTP and OLAP

OLTP: OLTP is Online Transaction Processing that supports the transaction oriented applications. It is the short queries are get run like transactions from one different types of the data and it is not possible to manage from the traditional database by every problem has a solution.

According to several technologies there are many different definitions of Big Data that are as follows. The business media firm has given 3 V's of Big Data they are as follows –

Volume - The amount of data is very at huge amount. It collect the data from various sectors like social media, interactions, etc. It ranges from terabyte to the petabytes.

Variety – The variety includes the data from a wide range of sources and formats. It comes in a format like E- mail, videos, articles, photos, stickers, PPT, PDF, etc. This data helps to shape the future of a company while tracking programs.
This type of data is in very huge amount.

Velocity – The velocity is a speed of the data that is as we know that is as we know that increasingly, businesses have stringent requirements from the time data is generated, to the time actionable insights are delivered to the users.
Therefore, data needs to be collected, stored, processed, and analyzed within relatively given short windows of time its ranging from daily to real-time also. If in an organization is not able to handle these 3 V's then we can say that there is big data problem in that organization.

Then , IBM added 4th V that is Veracity.



Veracity – It means that truthfulness of the data means some time data may contain wrong information or corrupted data or unexpected data. It will account to another.

OLAP: OLAP is Online Analytical Processing, in it the long-time queries are get processed. It characterized by the large volume of data.

Oracle: Oracle says that the data that cannot be processed in the traditional data management system are known as Big Data.

IBM: IBM is a term applied to data sets whose size and type is beyond the ability of traditional, relational Database capture, manage and process the data with low latency.

III. MAIN PURPOSE OF THIS WORK

This paper explores the problem of Big Data in the field of Big Data and also discuss the solutions to tackle the problem of big data. We investigate the influence of big data in various domains and various sectors.

SO WHY DO WE NEED BIG DATA?

As we are seeing that many more organizations even don't realize that they are facing to the big data problem or we can say that they are don't want to think in terms of big data. In general scenario an organization is likely to be benefit from big data technologies when existing databases and applications can no longer scale to support sudden increases the 3 V's in volume, variety, and velocity of data.

IV. TYPES OF DATA

Mainly there are three types of data, they are as follows –

1. Structured Data
2. Unstructured Data
3. Semi-Structured Data

1. Structured Data: Any type of data which can be stored, accessed and processed in the form of fixed format is called as a structured data. Now a days the size of such data grows to a huge amount, typical sizes are being in the range of multiple Zettabytes.

2. Unstructured Data: This refers to the data that lacks any specific format or structure. It makes very difficult and time consuming to process and analyze the unstructured data.

3. Semi-Structured Data: It pertains the data containing both the formats structured data and unstructured data. It refers to the Data that has not been classified under a particular database.

V. APPLICATIONS OF BIG DATA

Big Data is a fast growing and important part now-a-days.

The importance of big data leads to intense competition and increased demand for big data professionals. Big Data is used in many sectors by various sources in various domains. So, here we will discuss some of the applications as -

1. Health care

Big data has already started to create a huge difference in this sector. Health care like pharmaceutical industries, hospitals has already adopted these facilities. It provides personalized healthcare services to the individual patients, fitness, earables, telemedicine, remote monitoring and so on. It is also used in sending reports, giving appointments, analyzing the condition of the patients and keeping their records.

2. Academia

Big Data also helps in enhancing the education. Education is no more limited to the physical bounds of the classroom. Now-a-days students are getting jobs with their online degrees by many colleges and the universities. The data generated from the schools, colleges, universities is in very huge amount.

The type of data generated from this is personal information, attendance, academic progress of every students, staffs and colleagues. Administrative records such as course plan, financial status, Details, activities, etc.

3. Agriculture

India ranks second in the field of agriculture department in world. Agriculture plays an important role in the development of the country.

It is based on the geographical and climatic factors. The major factors on which crop production are climate, temperature, rainfall, cultivation, fertilizers, pesticides, etc. The green houses are equipped



With latest sensors, which help the farmers to know the atmosphere inside the green houses. This is very Essential for crop production.

4. I.T.

One of the largest users of big data is I.T. companies around the world

Are using big data. Data to optimize

Their functioning, enhance employee productivity and minimize risks in business operations. Big data technologies with machine learning and artificial intelligence, the IT sector is powering innovation to find solution.

5. Transportation

Big Data holds immense value for the transportation industry. Big data technologies to optimize route planning, control traffic, manage road congestion and improve services. Big Data to revenue management, drive technological innovation, enhance logistics and of course to gain the upper hand in the market.

VI. SO HOW WE WILL SOLVE THE PROBLEM

By using the various new tools that addresses the entire data management cycle, the big data technologies make it technically and economically feasible, not only to collect and store larger datasets, but also to analyze them in order to cover new precious and valuable insights. In most of the cases, big data processing involves a common data flow that is from collection of raw data to the consumption of actionable information.

So, there are mainly four parts which help to solve this problem, they are as follows-

1. Data Collection
2. Data Storing
3. Data Processing and Data Analyzing
4. Data Consuming and Data Visualizing

1. Data Collection

The first thing is that we will collect the data like as we mentioned above. It is first challenge that many organizations are face when dealing with the big data. So, we required a good platform that will perform the task of collecting the data in real time.

2. Data Storing

It is very important in the terms of big data. Every big data platform has need a very scalable, secure And the durable repository to store prior or even after processing tasks. It's totally depends on your specific

requirements. We may also need temporary Stores for data in- transit. The huge amount of data will be stored in small parts for heir future uses.

3. Data Processing and Data Analyzing

In this step, data is transformed from its raw state into a usable applicable. Usually by means of sorting, joining, aggregation and even we can perform more advanced algorithms and functions. The throughput data sets are then stored for further processing or made available for consumption through the data visualization tools and business intelligence.

4. Data Consumption and Data Visualization

Big Data is nothing but the getting high amount of values, Actionable insights your data assets. In general, data is given to stakeholders through self-service business intelligence and The Agile data visualization tools that is used for the fast and easy exploration of datasets.

So, these are the methods from which we can solve the big data problem.

VII. FUTURE SCOPE

In this today's world, the amount of data collected from various sources from all over the world in many forms. Big Data is applicable to process and store the data in light form. There are many new technologies and tools for the record, measure and process the data. Data is collected by the big data not only structured but it also contains unstructured. The main domains Of challenges in big data is data storage, data processing, data quality and relevance, data privacy and security and scalability. In future, big data will gets many benefits, specialized frameworks, limitations of framework and the Reviews.

VIII. CONCLUSION

To handle the Big Data and to work with it and it obtain the Benefits from a branch of science called as Data Science.

It deals with the knowledge from the huge sets of data, by virtue of data influence and exploration. It is useful in various platforms like healthcare, retailers, finance, sports hubs and communication. In The digital companies like Google and Yahoo. In Social networking sites like Facebook, Twitter, Instagram are requiring many data scientists. This paper covers the Big Data topic and conclude that Big Data is a problem and for this it shows the many Solutions.



IX. ACKNOWLEDGEMENT

We are thankful to Prof. S. N. Kakarwal Madam, P. E. S. College Of Engineering, Aurangabad, Maharashtra for their valuable support and guidance in writing this paper.

X. REFERENCES

- [1] Website - <https://aws.amazon.com/big-data>
- [2] Website-<https://www.gartner.com/it-glossary/big-data/>
- [3] Website - <https://www.sap.com/india/trends/big-data.html>
- [4] Website - <https://www.oracle.com/big-data/guide/what-is-big-data.html>
- [5] Website - <https://www.coursera.org/big-data/>
- [6] Website - <https://en.m.wikipedia.org/big-data/>
- [7] Website - <https://www.makeinindia.com/>
- [8] Sabia and Shetal Katra, *Applications of Big Data : Current Status and Future Scope Paper*.
- [9] Bhawna Dhupla and M.Usha Rani (IADS), Sri Padmavati Mahila Visvavidyalayam, *Research Challenges in Big Data Solutions – chapter in Springs Briefs in Applied Sciences and Technology Jan 2019*.
- [10] Manish Kumar Kakhani, Sweeti Kakhani and S. R. Biradar, *International Journal of Application or Innovation in Engineering of Management (IJAIEEM), Vol2, Issue 8, August 2018*.
- [11] Khushboo Wadhvani, Bradley University , *Big Data Challenges and Solutions, February 2017*.
- [12] Ranjit Kumar, *Book for Research Methodology_A Step-by-Step_Guide for Beginners*, by
- [13] Author C. R. Kothari. *Book Research Methodology, Methods and Techniques*