INTERNATIONAL RESEARCH JOURNAL OF HUMANITIES AND INTERDISCIPLINARY STUDIES
(Peer-reviewed, Refereed, Indexed & Open Access Journal)
DOI : 03.2021-11278686 ISSN : 2582-8568 IMPACT FACTOR : 5.828 (SJIF 2022)

Big Data

Smt. Reshma Dastgir Niduni
D. R. Mane Mahavidyalaya,
Kagal Dist- Kolhapur (Maharashtra, India)
E-mail: reshmaniduni@gmail.com

DOI No. 03.2021-11278686 **DOI Link ::** https://doi-ds.org/doilink/03.2022-76739781/IRJHISIC2203037

Abstract:

Big data is a term for massive data sets having large, more varied and complex structure with the difficulties of storing, analyzing and visualizing for further processes or results. The process of research into massive amounts of data to reveal hidden patterns and secret correlations named as big data analytics. This research paper explained that what is Big Data, why it is importance, benefits of Big Data. Traditional Database management system are not able to handle this vast amount of a data. Big data helps companies to generate valuable insights. Big Data is about data volume and large data set's measured in terms of terabytes or petabytes. Big Data has proved that great use since its inception, as companies started realizing its importance's for various business purpose.

In this paper presenting the characteristics of Big Data and importance of BiG Data, Types of Big Data, Sources of Big Data and Tools are used in Big Data to handle Big Data.

Keywords: *Big data, management system,*

Introduction:

The Defination of Big Data is a data thet contains greater than variety, arriving in increasing volumes and with more velocity. The three V's of Big data are Velocity, Volume, Variety. The Big Data can help you address a range of Bussiness activities. Big Data uses product development, predictive maintenance customer experience, fraud and compliance machine learning, Driveinnovation. It is an aim to solve new problems or old problems in better way. The Structure of Big Data is depends upon the Structural, Semi structured and Unstructured.

Defination:

Big Data is a collection of data that's is huge in volume yet growing expontially with time. It

is Data with so large size and complexicity.Big data is also a data but with huge size. Big Data is a term used to describe a collection of data that is huge in size and yet growing exponentially with time.

Big data best example includes stock exchange social media sites jet engines etc. the benefits of big data improved customers services better operational efficiency, better decision making. It is also data a but with huge size.

Big Data Technology can be used for creating a staging area or landing zone for new data before identifying what data should be moved to the data warehouse .Big Data deals with vast collection of heterogeneous data from different sources as compared to the data Science.

3V's of Big Data:

1. **Velocity:** The data is increasing at a very fast rate. It is estimated that the volume of data will double in every 2 years.
2. **Variety:** Now a days data are not stored in rows and column. Data is structured as well as unstructured. Log file, CCTV footage is unstructured data. Data which can be saved in tables are structured data like the transaction data of the bank.
3. **Volume:** The amount of data which we deal with is of very large size of Peta bytes.

Why impoatance of Big Data:

Big Data importance does not resolve around the amount of data company. Its importance lies in the fact that how the company utilizes the gathered data. Big Data analytics helps organization harness their data and use it to identify new opportunities. Evry company uses its collected uses its data, more rapidly its grows. The companies in the present mark ets needs to collect it and analysze it because of following points.

1. **Cost Saving :**

   Big Data technologies such as Hadoop and cloud based analytics, spark etc. bring cost saving benefits to business when they have to store large amount of data. These tools helps organization in identifying more effective ways of doing business.

2. **Time Saving :**

   Tools like Hadoop help them to analysis data immediately thus helping in making quick decision based on the learning.

3. **Understandthe market condition.**

   Big data analysis helps business to get a better understanding of market situations.

4. **Social Media Listening:**

   Companies can perform sentiment analysis using Bidg data tools to get feedback to improve their online presence.

5. **Solve advertise problem and offer marketing Insights:**

Big data analytics problem helps in changing the company prodyuct line. Big Data analytics shape all business operations. It enables campanies to fulfill customer expectations.

6. **The Driver of Innovation and Product Development**

Big Data makes companies capable to innovate and revdevelop their products.

Types of Big Data:

Big Data includes huge volume, high velocity, and extensible variety of data. These Big data is classified in three ways:

- Structured Data
- Unstructured Data
- Semi-Structured Data

STRUCTURED DATA:

Structured data are stored in structured form, In row and column structure. We can easily retrieve and analyze necessary data form structured data. For data mining structured data are very useful. So we can do mining and retrieve useful knowledge from them. ex. Excel, Database, Table, etc..

SEMI STRUCTURED DATA:

Semi-structured data is a form of structured data that does not obey the formal structure of data models associated with relational databases or other forms of data tables, but nonetheless contains tags or other markers to separate semantic elements and enforce hierarchies of records and fields within the data. Therefore, it is also known as self- describing structure.

In semi structured data, the entities belonging to the same class may have different attributes even though they are grouped together, and the attributes order is not important. Semi structured data are increasingly occurring since the advent of the Internet where full text documents and databases are not the only forms of data anymore, and different applications need a medium for exchanging information. In object-oriented databases, one often finds semi-structured data [8].

UNSTRUCTURED DATA:

Unstructured data are data which are not in structured format. All unformat data are known as unstructured data. Worlds 90% data are unstructured data. All PDF, Word, Audio, Video, Image file contain unstructured data.

For any analyses we have to convert over unstructured data into structured data. These things are done by different tools.

The Sources of Big Data:

Big Data is the data that is characterized by such informational feature as the log of events

nature and statically correctness and that imposes data processing and easy scalability of the solution. There are two types of Big data sources are internal and external once. Data is internal if a company generates owns and control it. External data is public data or the generates outside the company correspondingly the company neither owns nor control it.

Some example of internal data sources are corporate ERP module, international documents, sensor, controller, in house call centers website logs.

- **Black box data :**

  It Includes flight crew voices, microphones recording and aircrafts performance information.

- **Social media :**

  It is derived from social media platform through tweet, retweets, likes ,video, upload and comments shared on facebook , instagram, twitter, you tube ,linkedIn etc.

- **Stock exchang data :**

  This is data from stock exchanges about the share selling and buying decision made by customers.

- **Transactional data:**

  It is information gatheres via online and offline transaction during different points of sale. It is a key source of business intelligence. The unique characteristics of transactional data are its time print. The sources of transactional data include payments order invoices, storage records and E-receipts.

- **Search Engine data:**

  This is one of the most significant sources of Bid data. Search engines have vast databases where they get their data.

Benefits of Big Data:

There are many companies like Amazon, Netflix, Spotufy, LinkedIn, Swiggy etc which Big Data analytics, Banking sector makes the maximum use of Big Data Analytics. Education Sector is also using data analytics to enhance students performance as well as making teaching easier for instructors.

Big Data analytics has expanded its roots I all the field. This results in the use of Big Data in a wide range of Indusrties including finance and banking, Healthcare, eduction, governments, Retails, manufacturing and many more.

Big Data is used for organizational to improve the efficiency, understand the untapped market and It enhance competitivenessascmapared to Data Science.Data Science supposedly uses theoretically as well as practical approaches to big information from the Big Data which plays an important.

The Basic concept Big data is analysis, data filtering. Big Data relates more with technology computer tools and software. Big Data is highly velocity or high variety information assests that demand cost effective innovation forms of information processing that enables enhanced insight, decision making and process automation.

The amount of data that can be collected by companies are huge and they pertain to Big Data but utilization of the data to extract valuable information , so that Data Science is needed.

Tools are used in Big Data:

Data is meaningless until it turns into useful information and knowledge which can aid the management in decision making. For this purpose, we have several top big data software available in the market. This software help in storing, analyzing, reporting and doing a lot more with data.

Big data Tools: The tools that are used to store and analyze a large number of data sets and processing these complex data are known as big data tools. A large amount of data is very difficult to process in traditional databases. So that's why we can use big data tools and manage our huge size of data very easily.

Apache Hadoop:

This is the topmost big data tool. In fact, over half of the Fortune 50 companies use Hadoop. Some of the Big names include Amazon Web services, Hortonworks, IBM, Intel, Microsoft, Facebook, etc.

It is a Highly useful for R & D purposes, Provides quick access to data. Highly scalable, Highly-available service resting on a cluster of computers

Xplenty:

Xplenty will help you make the most out of your data without investing in hardware, software, or related personnel. Xplenty provides support through email, chats, phone, and an online meeting. Xplenty is an elastic and scalable cloud platform.

You will get immediate connectivity to a variety of data stores and a rich set of out-of-the-box data transformation components. Xplenty's powerful on-platform transformation tools allow you to clean, normalize, and transform data while also adhering to compliance best practices.

Adverity:

It is a flexible end-to-end marketing analytics platform that enables marketers to track marketing performance in a single view and effortlessly uncover new insights in real-time.It isExcellent customer support, High security and governance, Strong built-in predictive analytics, Easily analyze cross-channel performance with ROI Advisor.

Skytree:

It is one of the best big data analytics tools that empowers data scientists to build more

accurate models faster. It offers accurate predictive machine learning models that are easy to use.Skytree via the easy-to-adopt GUI or programmatically in Java, Model Interpretability

Spark:

It is an open source framework for data analytics, machine learning algorithms, and fast cluster computing. This is written in Scala, Java, Python, and R.

It helps to run an application in Hadoop cluster, up to 100 times faster in memory, and ten times faster on disk.It is one of the open source data analytics tools that offers lighting Fast Processing. Support for Sophisticated Analytics

Lumify:

It is Scalable, Secure, Supported by a dedicated full-time development team. Supports the cloud-based environment. Works well with Amazon's AWS.

It is a big data fusion, analysis, and visualization platform. It is one of the best big data analysis tools that helps users to discover connections and explore relationships in their data via a suite of analytic options.

Talend:

Talend is one of the most powerful data integration ETL tools, cloud computing, and big data integration tools available in the market.It is specialized in Big Data because it has all the plugins to integrate with big data efficiently.Talend is used to unify the repository for storing and reusing the metadata.

R-Programming:

R is one of the most comprehensive statistical analysis packages. It is open-source, free, multi-paradigm and dynamic software environment. It is written in C, Fortran and R programming languages.

It provides a suite of operators for calculations on arrays, in particular, matrices, It provides coherent, integrated collection of big data tools for data analysis. It provides graphical facilities for data analysis which display either on-screen or on hardcopy.

Conclusion:

In this paper we learn that abstract of concept of Big Data helps us to understand inefficiency and opportunities in out company. There are three types of data. which are structured, unstructured and semi structured. They have different characteristic and applications. World 10 % data are structured and 90% data are unstructured. We need structured data for analysis so we get easily useful knowledge. We have to convert unstructured data to structured data by different tools for analysis. By using these tools, we get interesting pattern and knowledge from data.You need to choose the right Big Data tool wisely as per your project needs.

The Big Data technologies discussed here will help any company to increase its profits, understand its customers better and develop quality solutions. And the best part is, you can start learning these technologies from the tutorials and resources available on the Internet.

It plays a major role in shaping the organization growth. The Big data is backbone of Bussiness of business in the modern industry. It analysis helps companies to make growth strategies for both the presents and future. It is pivotal for studying the graph and customer needs.

References:

1. Big data - Wikipedia
2. What is Big Data? Introduction, Types, Characteristics, Examples (guru99.com)
3. Introduction to Big Data (w3schools.in)
4. What is Big Data - javatpoint