International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 04 Issue: 08 | Aug -2017 www.irjet.net p-ISSN: 2395-0072

Big Data Analysis

Narayanan.V[1], NikitaSri.G[2], Suchitra.B[3]

[1,2] *Student,Dept.of Information Technology, Sri Krishna Arts & Science College, Coimbatore, TamilNadu, India*
[3]Assistant Professor, *Dept.of Information Technology, Sri Krishna Arts & Science College, Coimbatore, TamilNadu, India*

Abstract- As technology grows , the need for data is also necessary . In today era world contains very huge amount of data that are scattered / distributed everywhere . The need for exact data is also necessary ,some of the data are hidden which cannot be used by users . The data's in the web are classified as structured data (having proper structure), unstructured data(having improper structure data of various file format) and semi-structured data (partial organized data).In order to handle unstructured data and to support large volume of data , the big data is useful. Big data supports all types of data and large volume of data . In this Paper , the need for big data and its advantages and how it is useful in IT fields are discussed.

1. Introduction

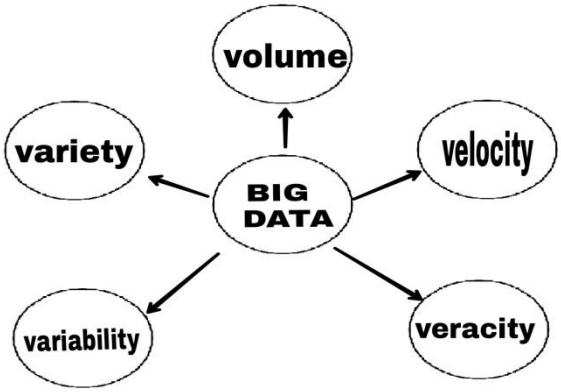
The most recent development in this type of data is in attitudes and behaviours and this is where Big Data comes in. while examing everyone's activities on the internet (i.e) their Facebook posts , Google searches, tweets,emails, and more, we have now more varities of data on every profiles . This has led to very large databases, which need to be tracked for some measures . The evolution of data is not ending anytime soon. analysis. After the birth of big data, new technologies and processes were developed at warp speed to help companies to manuplate their data into profitable way . Big data required advanced processing frameworks such as Hadoop and new databases such as NoSQL to store and manipulate it. The basic idea behind the word " BIG DATA "is that everything we do is increasingly leaving a digital trace (data) which we use and examine.

2. What actually BIG DATA is

Data has been spread everywhere whether we want it or not There are some things that are so big that they have implications for everyone, BIG DATA is one of those things an is completely transforming the way we do business and is impacting other parts of our lives. BIG DATA refers to the large collection of data sets that are so larger or complex so that traditional data processing application software is inadequate to deal with them. big data challenges include search, capturing data, data storage, data analysis, sharing , transfer, querying , updating visualization and information privacy. BIG DATA usually includes data with data philosophy encompasses unstructured, semi-structured and unstructured data ,however the main focus is on unstructured data big data size is a constantly moving target as of terabytes to many petabytes of data. Big Data is a convergence of new hardware and algorithms that allow us to discover new patterns in large data sets— patterns we can apply to making better predictions and, ultimately, better decisions. Big Data has the potential to improve lives with better services and products.

3. 5v characteristics

Big data can be described by the following 5v's characteristics



a. Volume it refers to the vast amount of data generated and stored every second . The size of the data determines the value and potential insight- and whether it can actually be considered big data or not.big data tools use distributed systems so that we can store an analyse data across databases that are dotted around anywhere in the world

b. Variety it refers to the type and nature of the data .variety of data categories into structured data (relational database (.i.e.) having proper structure) ,semi-structured data (partial organize data)and unstructured data (text , images , video , voice , etc.) This helps people who analyze it to effectively use the resulting insight.

c. Velocity it refers to the speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development. Just the data goes viral in seconds.

d. Variability Inconsistency of the data set can hamper processes to handle and manage it. Variability is different from variety. Its meaning is constantly changing it can have a huge impact on your data homogenization.

e. Veracity The quality and quantity of captured data can vary greatly, affecting the accurate analysis and that to in the inconvenient form .here Veracity refers to make sure the data is truthful, which requires processes to keep the bad data from heaping in your systems.

4. Big data as big deal

FOUR things make big data significant:

a) **The data is massive**. The data are huge so that It cant fit on a single hard drive. The volume of data far exceeds than what the human mind can think .(for example just think of a Million billion Terabytes, and then multiply that by more millions).

b) **The data is messy and unstructured**. Most important work of the big data is cleaning and converting the information so that it would be easy to search an sort . Only a few thousand experts on our planet fully know how to do this data cleanup. But in 10 years, the work for the big data will increases because data generated will also increase day by day and will become tedious one.

c) **Data has become a commodity**: now data has become a necessary commodity that can be sold and bought. companies and individuals can buy terabytes of social media and other data on Data marketplaces . data are huge that it wont be fit into any hard disk and Most of the data is cloud-based. data Buying commonly involves a subscription fee where you plug into a cloud server farm.

d) The possibilities of big data are endless. Data are very useful in our day to day life because Perhap doctors will one day predict cancer, heart attack s ,strokes and some more deadly diseases for individuals weeks before they happen it wont be helful for us therefore we should analyse the data . Airplane and automobile crashes might be reduced by predictive analyses of their mechanical data and traffic and weather patterns. Online traing might be improved by having big data experts with us .musician can find out the tune and rhythm relating to peoples taste and they can be make the tune of the current trend by analyzing data. likewise not only in this fields big data has its own scope in every field world wide these are some of examples for our understanding .In the big data only a piece of cake has been eaten .there is more and more in it. the discoveries in the big data are updating day by day

5. Problems with big data

The most challenging task in big data is

a. **Storing data** even the data which are smaller in size is difficult to store and retrieve .therefore it is a complex task to store the data and in analyzing it

b. **Processing data faster** The data store are in diferent format structured data(relational database (ie) having proper structure) ,semi-structured data (partial organize data)and unstructured data(text, images ,video, voice,etc.) so it not easy to process the data & it takes plenty of time.

6. Tools for big data

Apache Hadoop
Lumify
Apache Storm
HPCC Systems Big Data
Apache Samoa
Elasticsearch
MongoDB
Talend Open Studio for Big Data
Rapid Miner
R-ProgrammingThese are some tools for handling big data

7. CONCLUSION

Big data deals with knowledge discovery and data can be extracted such a way that it is useful by millon of users . When data are increased, the need for database is also important . The data about data is also become a important criteria . and as years go , the need & use big data is also necessary. but big data is just the starting stage of these problems. As the technology develop there is a huge chance that the data which has been collected during that period can exceed the amount of data created till humen birth. The big data plays a vital role in todays world. In this paper the advantage ,characteristics and how the database for the big data supports are seen.

8. References

1) JEFF desjards ,on The evolution of data
2) paul gil ,on What Exactly is 'BIG data'
3)Wikipedia.org/wiki/big_data
4) neelamani samal,nilamashob myshra ,on Big data process:big challenges and opportunity
5) Ashley devan on The 7V'S of big data

BIOGRAPHIES

Narayanan.V
Student
Dept .of Information technology
Sri Krishna College of Arts & Science
Coimbatore

NikitaSri.G
Student
Dept .of Information technology
Sri Krishna College of Arts & Science Coimbatore

Suchitra.B
Assistant Professor
Dept .of Information technology
Sri Krishna College of Arts & Science Coimbatore