

Big Data Analytics: Applications, Challenges and Security Issues

¹A.Subashini, ²A.Jenita Mary

¹Assistant Professor, ²Assistant Professor

Department of Computer Applications

Faculty of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur, Tamilnadu.

ABSTRACT

We are living in the modern computer age and using very hi-technology. This is making us to deal the huge amount of data in in-built server storages or through online in the web. Each data is connected with some other data source and it is very tough to estimate the volume of data through digitally. This huge data is called as big data that indicates the amount of data in the range of some terabytes and beyond. Mostly, this type of data is in the form of images, videos and different documents with different structure format. To process and analyse this large amount of data using the technique is called “Hadoop platform”. It is very high speed process and cost effective too.

Keywords: Hadoop, Bigdata,.

1. INTRODUCTION

The terminology “big data” refers to the huge volume of unstructured data created by the lots of applications, twitter and other social network sites. Big data means describing the huge amount of data which are available in different type of format: structured, semi structured or unstructured. The data when it is unable to process through traditional databases and software technologies then such type of data is called as big data. The term “Big data” is originated from one of the web companies who were handled the both structured and unstructured data. On an average, we create 2.5 quintillion bytes of data, in that almost 90% of data has been created in the last two years only. Everyday lots of data is being processed, collected and warehoused. [1]

- Web data, e-commerce
- Bank/Credit card transactions
- Social Network

This modern society is very much sophisticated with more online transactions: Banking, ticket booking and on-line payments on daily basis. As a result, organisations are creating, processing, and storing huge amount of data and these data is becoming more worthy and valuable for a computer industry. [2]

2. PROPERTIES OF BIG DATA

There are so many properties are associated with big data concept, namely, volume, velocity, veracity and variety.

Volume: Big data is huge volume of data and the data is created by the networks, on-line transaction devices and from the social media sites. The size of the data is beyond terabytes or petabytes.

Velocity: It means the rate of data transactions happens and data flow in on-line transactions, devices and social media networks. These real time data helps the companies, research and business people to make the correct decisions and transactions etc. The data flow includes data steaming as well.

Veracity: It refers the biases, noise and irregularity in data. The data is being processed and stored during data analysing. It talks about the uncertainty of data, before taking data analysing it should be clear.

Variety: The variety refers to the types of data and sources that are using the big data concept. It can be either structured or unstructured but the data come in the form of any images, videos, emails, audio or pdf files. This type o data is mostly unstructured and creates trouble in storage while analysing the data.

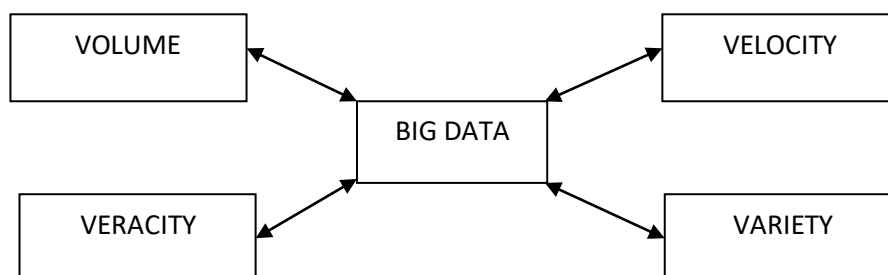
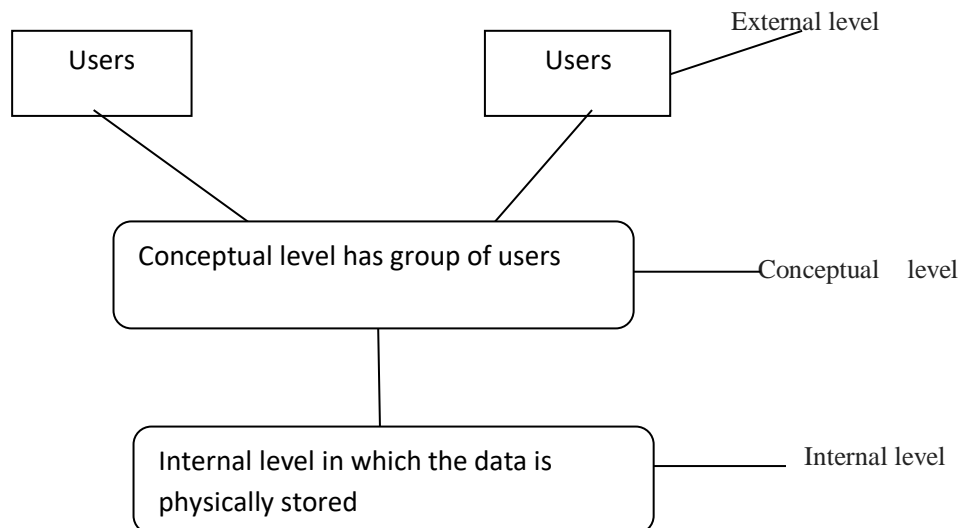


Figure-1: Four V's of Big Data

3. DATABASE MANAGEMENT SYSTEMS VS BIG DATA ANALYTICS

Relational database systems are based on the structured **data** i.e. **Data** is stored in a fixed format or fields in a file.



The terminology big data refers a collection of new database technologies that is designed to handle the specific requirements better than relational databases. Big data help and used in the semi structured and unstructured data to improve the variety of data grouped from various sources: Different subscribers, customers and industry audiences. The implementation of data analytics on huge data sets is known as big data analytics. Analytics is the process of raw data into identifying and verifying the patterns of the data. This is one of the delayed processing and time consuming tasks of analytics. [3], that is, to prepare the data for analysis sample, Analytics are executed on big amount of data that requires efficient procedure to store, filter, convert and retrieve the data. Cloud analytics solutions need to consider the lot of cloud deployment models adopted by various enterprises like: Private, Public and Hybrid methods.

Private: It mainly works on the private network, handled by the specific companies or organisation itself or sometimes by the third party vendors. A private Cloud is more appropriate for businesses that require the highest level of security controls and data privacy.

Public: It works with off-site over the web and exists to the general public. Public Cloud offers with low cost but high efficiency and shared resources. The quality of services such as Security, privacy, and availability is listed in a contract.

Hybrid: It combines both the Clouds where additional resources from a public Cloud can be provided as needed to a private Cloud or vice versa. Customers can be developed and deployed the analytic applications using a private environment.

NECESSITY OF SECURITY IN BIG DATA

The big data is primarily used by various businesses but they might not have any assets from security point of view. When any security incidents happens the big data then it may come out with even more troubles and serious problems. In general, most of the companies use this technology to store the data of petabytes range regarding to the respective industry business and customers. This creates the severe criticality in the classification of information to secure the data we either need encryption or decryption log or we should use honey pot techniques. The challenge of ducting risks, threats and malicious introduces must be solved when using the big data style analysis. [3]

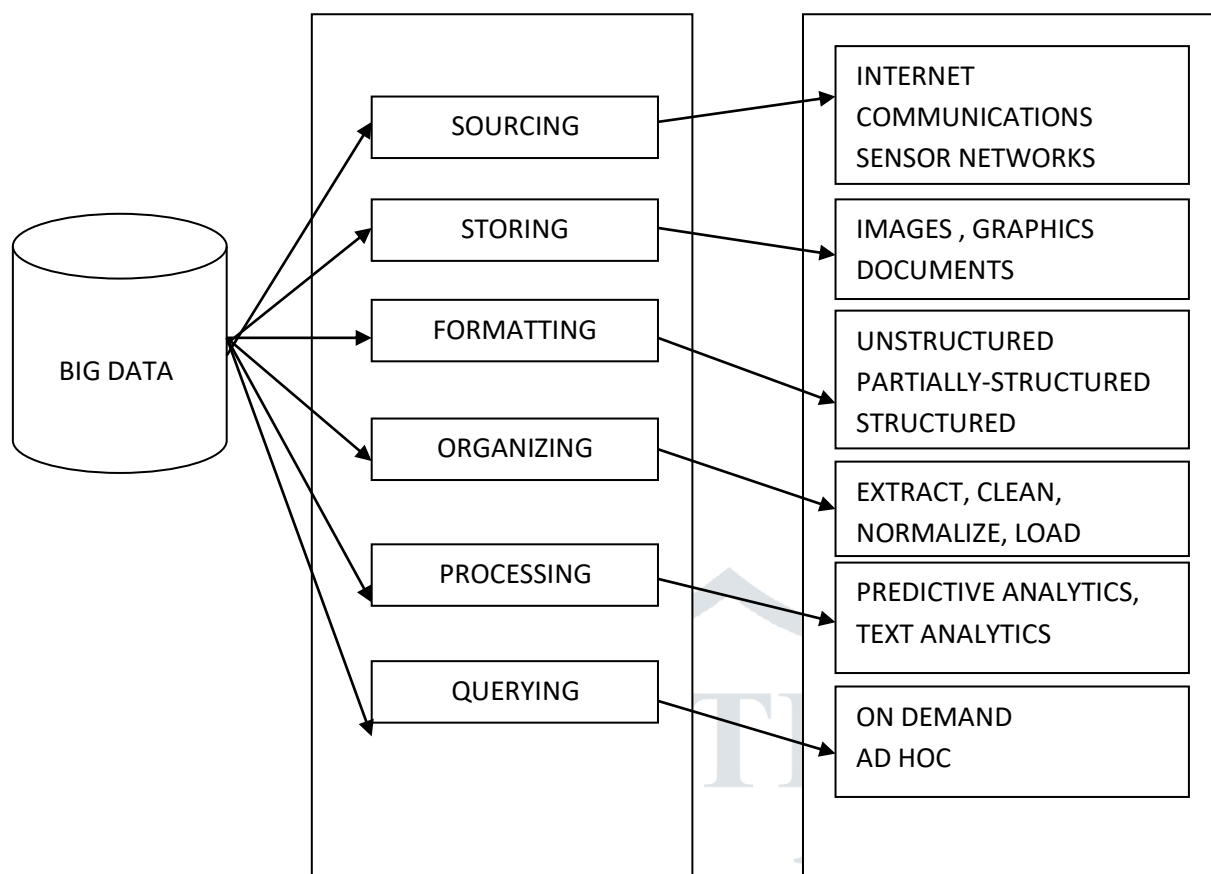


Figure-3: View of Big Data

4. ADVANTAGES OF BIG DATA

The big data allows each and every individual to verify and analyse the threats that he or she faces internally by knowing onto the complete data background over the company using the quality set of tools that the software supporting the big data provides. This is unique and advantage of big data since it permits the users to make the data secure and safe. Mostly the companies and organizations will aware and use the advantages from the cloud storage's speed, capacity and scalability. Big data also permits the end users to visualize the data and most of the companies and industries will find the new job and business opportunities. Also when the individual is allowed to personalize the content or to see and feel the real time websites are one more notable advantage in the big data analytics.

4. BIGDATA APPLICATIONS

In the existing period of data explosion, parallel processing is very much important for executing a huge amount of data in a timely manner. The following techniques "Parallelisation" and algorithms are used to reach the better scalability and ensure the performance for processing the big data. In general, most of the industries and academics are used the tool or model popularly called as "Map reduce". The two primary advantages of map reduce are encapsulations of data storage and distribution of replication details. It is a very simplest way of use by the programmers to compute for the map reduce task. Since the map reduce is schema and index free format, it needs parsing of each records at reading point map reduce has received a many of attentiveness in the fields of data mining, information and image retrieval, etc. The coding and computation becomes very difficult to be managed by traditional data processing which creates the development of big data apps. The big data provides and infrastructure supports for handling transparency in manufacturing industry having the ability to ambiguous and uncertainties that exists in the component performance and availability. The field of bioinformatics is another application of big data which requires large scale data analysis. .

TECHNICAL CHALLENGES IN BIG DATA

In the recent past new technologies evolved much and people meet with more challenges in various aspects. Once the functional constraints are in place, the next one is the technical issues. The big data is facing many technical issues which are on the road way of the research people.

Failure handling:

Developing foolproof reliable systems is not the easy task, system can be created in such a way that the probability of failures must be controlled within the permitted threshold fault tolerance is the technology challenge in the big data. When the process is initiated it might involve with numerous network nodes and the entire computation process becomes very unwieldy to retaining the checkpoints and fixing the threshold level for the process restart in case of any loss or failure are bigger concerns.

Data heterogeneity:

Big data deals with unstructured, semi structured and structured data. The process of linking unstructured data with either semi structure or with structured data and converting its data from one format to another specified format needs a lot more research.

5. CONCLUSION

There is no doubt Big Data is the hot edges of today's data development headway. The measure of data at present delivered by the diverse activities of the overall population has never been so tremendous, and is being made at a consistently extending speed. Through better investigation of the significant volumes of data that are getting the chance to be available, there is the potential for making speedier advances in a few trains and upgrading the productivity and achievement of various ventures. At long last, to totally benefit from Big Data, the above communicated troubles ought to be dealt with.

REFERENCES

- [1]. Jui-Chien Hsieh , Ai-Hsien Li and Chung-Chi Yang, "Cloud, and Big Data Computing", Int.J.Environ.Res.Public Health 2013,10,6131-6153
- [2]. Marcos D. Assunção, Rodrigo N. Calheiros, Silvia Bianchi, Marco A.S. Netto, RajkumarBuyya, "Big Data computing and clouds: Trends and future directions", journal of parallel and distributed Computing. Aug 25 2014.
- [3]. Mythreyee S, Poornima Purohit, Apoorva D.R, "A Study on Use of Big Data in Cloud Computing Environment ", International journal of Advance Research, Ideas and Innovations in Technology. 2017.

