A Study on Data Generation, Storage and Processing of BIG DATA

Dr.S.Prem Kumar M.Tech,P.Hd[1] , **N.Archana**[2], **B.Dharani**[3]
[1, 2, 3] Dept Of Computer Science And Engineering
[1, 2, 3] G.Pullaiah College Of Engineering and Technology

Abstract- *Everyday enormous amount of data is being produced worldwide. Everyday 2.5 quintillion bytes of data is created. For gathering any kind of information relating to climatic information we make use of sensors, posts to social media sites, digital pictures and videos, transactions etc.*

Big Data analytics has brought a big opportunity for organizations. Huge amount of information are being captured by companies. Increase in the big data also has increase in breaching the privacy of individuals. For transferring the large data sets we use big data. Many organizations are trying to make use of Big Data to secure their data. In order to protect the privacy in big data different algorithms and methodologies must be implemented at different levels. Already the forward thinking players of the banking, insurance, manufacturing, retail, healthcare, transportation, communications, education are successfully using big data.

I. INTRODUCTION

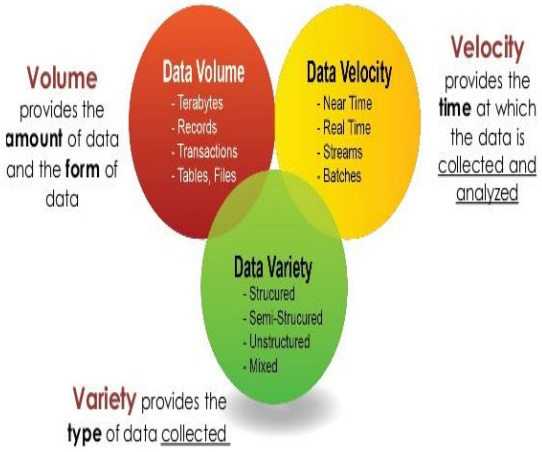
Now a day's amount of data generation rate is growing rapidly it is becoming very difficult to handle by using the traditional methods or systems [1]. Big data is very large, distributed aggregations of loosely structured data that often incomplete and inaccessible. To be more specific, Big data has the following characteristics stated below.

It works with Petabytes of data. The storage capacity of people and records in big data is above million and trillions respectively [2].Unstructured data is focused more[3].

Big data contains data sets with sizes that has the ability of commonly used software tools to manage, and process data within in a time. To understand the datasets which are very complex new technique should be required in Big data [4]. In big data volume, variety, velocity are distinguished [5]. Big data can be effectively utilize for better understanding the world. For example Amazon and Google can learn our shopping preferences and browsing habits. Facebook stores all information about personal life and social relationships. Popular websites in video sharing such as YouTube recommends us videos based on our search history. Storing and making reuse of personal information causes threat users privacy and security.

II. CHARACTERISTICS

The following are the peculiar forms of Big data:[5]



A. Volume

The amount of data generated and stored, and the form of data, range describes the value. It also describes whether the data should be considered or not.

B. Variety

Variety provides kind of data collected. Big data is also geospatial data,3D data including log files and social media.

C. Velocity

Velocity refers to the speed of data processing that provides a path of high growth and development.

III. LIFE CYCLE OF BIG DATA

Handling dimensions of big data in terms of 3v's. An efficient and effective systems is needed to design and process

huge amount of data from different sources arriving with very high speed is to be monitored. Big data goes through several stages throughout its life cycle. Data is distributed widely and being developed to process and store the data in large repositories. For example, cloud computing technologies, such as Hadoop MapReduce, are explored for big data storage and processing processing.

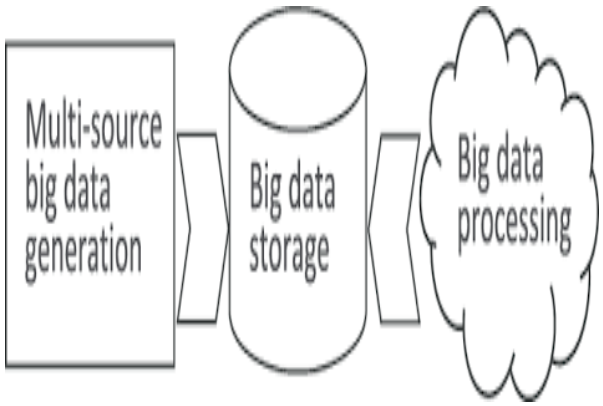


Figure: big data life cycle

A. Data generation:

The data generation is more seen in domains such as market trends, e-banking etc. 25TB of data generation is seen by social networking sites like Facebook.

B. Data Storage:

It is a Storage infrastructure that is designed specifically to store, manage data. Big data storage enables the storage and sorting of big data i such a way that it can easily be accessed used and processed by applications [6]. Data management is to ensure high level of data quality and accessbility for business intelligence.

C. Data processing:

Data may be coming from different sources like in form of text, audios, videos so data collection is needed. Data transmission and data collection refers to data processing. By removing the meaningless and redundant data memory will be saved.

IV. APPLICATIONS OF BIG DATA

The drastic increase in demand of Big data information management spent on software firms is more than $15 billion. 4.6 billion mobile-phone worldwide [7]. One third of worldwide stores information in formats of image data and in alpha numeric text [8].The applications of big data are Healthcare, Media, Education, Education etc.

A. Healthcare

Big data is useful at healthcare, in improving report of patients which is automated [9]. Not only managing the report but also prescribed medicine, clinical treatments and so on. To process the information, it includes records of the patient and generate data by specified doctor, many sensors are used to detect the information. There should be a greater need of paying attention towards this type of environments.

B. Media

To understand how the media utilizes big data, With the smart phones and associated digital media becoming the major source of entertainment. Media creators and distributers must embrace big data analytics to create a connection with their customers.

C. Government

Government organizations lack in efficiency of information, cost complexity, where big data provides these features to the government, without flaws as proposed in [10].These analytics requires different part of government sectors like local or central that work with combination in creating new processes to deliver the desired outputs.

D. Manufacturing

Big data also supports the developers to develop the product with less cost, and at low time rate. Past years, Six sigma program was implemented, where the manufactures were able to produce variety of product with high quality and abundant of output per units of inputs was drastically improved. Chemicals, Pharmaceutical environments cannot produce optimal result even after applying the techniques. To process and analyse the complex production of industrial information without duplication activities different techniques are needed in manufacturing [11].

V. CONCLUSION

One of the biggest computations over previous decades is big data technology. The lifecycle stages perform arranging processing and collecting or storing the information of records respectively. Federated government also involve in the development of big data all over the world. The biggest advantage and feature of big data that, it performs low-cost commodity hardware and more on. The key feature that big data provides is efficiently in data, productivity, revenue, and profitability in terms of business and IT sector.

REFERENCES

[1] J.Manyika et al., Big data: The Next Frontier for Innovation, Competition,
and Productivity. Zürich, Switzerland: McKinsey Global Inst., Jun. 2011.

[2] Snijders, C.; Matzat, U.; Reips, U.-D. (2012). "'Big Data': Big gaps of knowledge in the field of Internet International Journal of Internet Science.

[3] Dedić, N.; Stanier, C.(2017)."Towards Differentiating Business Intelligence, Big Data,,Data Analytics and Knowledge Discovery".

[4] Ibrahim; Targio Hashem, Abaker; Yaqoob, Ibrar; Badrul Anuar, Nor; Mokhtar, Salimah; Gani, Abdullah; Ullah Khan, Samee (2015). "big data" on cloud computing: Review and open research issues". Information Systems.

[5] DT&SC 7-3: What is Big Data?. YouTube 12 August 2015.

[6] The Information Society.

[7] "Impending Challenges for the Use of Big Data". International Journal of Radiation Oncology*Biology*Physics.

[8] O'Donoghue, John; Herbert, John (1 October 2012). Journal of Data and Information Quality. Retrieved 16 June 2016 – via ACM Digital Library.

[9] Couldry, Nick; Turow, Joseph (2014). "Advertising, Big Data, and the Clearance of the Public Realm: Marketers' New Approaches to the Content Subsidy". International Journal of Communication.

[10] Computerworld. 21 March 2012. Retrieved 12 September *2016*.[60]

[11] Lee, Jay; Wu, F.; Zhao, W.; Ghaffari, M.; Liao, L (January 2013). "Prognostics and health management design for rotary machinery systems—Reviews,methodology and applications". Mechanical Systems and Signal Processing.