Journal of Advances and Scholarly Researches in Allied Education
Vol. 16, Issue No. 2, February-2019, ISSN 2230-7540

Feature, Planning and Uses of Big Data

Ankit Garg*

Assistant Professor in Computer Science, R.K.S.D. (P.G.) College, Kaithal

Abstract – In the era of Computing Technology, data plays a role of lifeblood in various fields to make decisions. Without consistent and reliable data, the organizations cannot make decisions at right time. To design the data in an effective way, traditional methods were not proved to be effective because data is increasing from Kilo Bytes to Giga Bytes and then to Petta Bytes. Because of large diversity of data that is changing very rapidly and massively, big data is an emerging technology that takes an initiative to overcome these problems. In addition to the large volume of data, big data also handles varieties, velocity, veracity and value of data. Various tools being used to analyze the big data like Hadoop, MapReduce are also discussed in this paper. Architecture of Hadoop which is used to convert diverse and unstructured data into uniform and structured form is also described in this paper. Big data plays a significant role in various areas like banking, industries, E-Governance etc.

Keywords — Big Data, 5 V's, Architecture, Hadoop, MapReduce, Applications of Big Data.

----------X----------

I. INTRODUCTION

Here, in the era of emerging technology each field whether it is banking, hospitals, school, government organization, social media and lot more produces a lake of data in a structured, semi-structured and unstructured format called **BIG DATA.** Big data describes collection of data sets so large and complex that it becomes difficult to process, manage using relational database system and traditional database system. Its tool like data mining, text analytics, predictive analytics, statistics, cloud computing are so large that are impractical to manage with traditional software tools.

Big data can be collected through two sources: internal and external sources. Its efforts are currently focus on analyzing internal 0 data to extract result and whereas some organization look outside to get such as social media. Internal data are enterprise data transaction details (which occur in e-shopping, banking, etc.) log files (consider example of railway ticket booking at IRCTC and we make payment in reserving our seats. The amount of respective seats is deducted from account but the seat is not reserved which leads to inconsistency then this log file generate unique Id. With the help of this Id within 24 hours our deducted amount will be cash back in an account. External data sources are social media data which is increasing day by day and platform promoted this is Facebook, Instagram Google Duo, Pinterest, Twitter and a lot more. In short, big data is a combination of social data and enterprise data. The 5V's of Big Data: Volume, Velocity, Value, Variety and Veracity as shown in Figure 1.

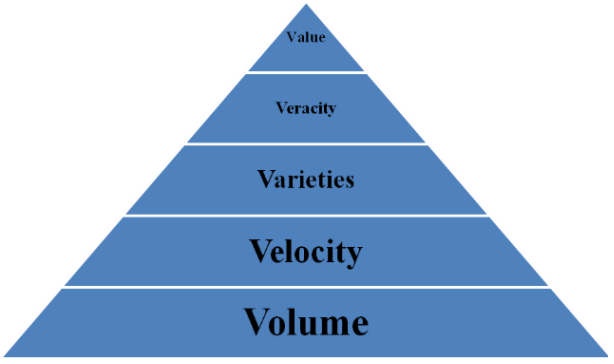

Figure 1 Representation of 5 V's of Big Data

VOLUME:-

Big Data is like a pyramid where volume is the base. Today, Facebook gives 600 TB of data everyday consisting of 20 billion messages, 5 billion times, that the 'like' is pressed and photos uploaded and a lot more. Data is produced at an astronomical rates from gigabytes to Petabytes and now Zettabytes because sources of data has increased from enterprises, social media sites, cell phones , cars , M2M sensors etc. in each and every second.

VELOCITY: -

As we know data is being generated at an alarming rate. Velocity refers to the increasing speed 0 at which data is created so the increasing speed at which data can be analysed, processed and stored in a relational database. After every 60seconds there are more than 1 Lakh tweets, 695000 status