Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 5, May -2017

# Vital Requirements of Big Data V's Virtues

Dr. Gaytri Devi[1], Rajesh Dawar[2]

[1]GVM Institute of Technology and Management, DCRUST, Sonipat
[2]GVM Institute of Technology and Management, DCRUST , Sonipat

**ABSTRACT :-** *In the present era, the world is moving around the Internet. People are interacting via Internet only. With the increasing number of Internet users and Internet applications, enormous amount of data is transferred on the Internet. This huge amount of data is termed as big data. The Big Data not only includes traditional files but also audio, video and graphics. Big Data is now rapidly expanding in all fields with the growth of networking. Big data plays very important role in the development of organizations to take strategic and competitive decisions. The decisions that were previously made on guesswork, can now be made on data itself. To get valuable insights, big data needs to be analyzed and processed and to get its maximum advantage, the requirements of big data should be fulfilled. Thus, this paper addresses the requirements of big data by relating with its attributes.*

***Keywords-*** *Big Data, Big Data V's , Big data analytics.*

## 1. INTRODUCTION

In this digital universe, The voluminous data are generated from a variety of users and devices on the Internet. The big data generated from Internet, social media, mobile devices, business etc is not limited to gigabytes but is measured in Exabytes($10^{18}$ or million of gigabytes) or Zettabytes ($10^{21}$) or even larger amount of information.
The amount of data that is travelling across the internet today, not only large, but complex as well. The transferred big data may contain semi-structured or unstructured data, such as transaction based data , graphics, audios, videos. The Big Data is the combination of structured, semi-structured, unstructured, homogeneous and heterogeneous data.
Thus, Big Data is a comprehensive term for collection of large-volume and complex data from the variety of sources.

To extract big value, big data needs analyzing and processing of data. Big Data analytics here comes into role. Big data analytics involves examining and understanding the big data. It is the process of collecting, organizing and analyzing large sets of data to discover useful information. The big data analysis pipeline includes multiple stages ranging from Data acquisition, information extraction and cleaning, data integration, analysis**,** interpretation [4]. The overall life cycle of big data consists of multiple stages ranging from generation, collection, aggregation, analyzing, processing, transmission and delivery of data.

Big data analytics can help organizations to better understand the information contained within the data and also help to identify the data that is most important to the business like market trends, customer preferences and future business decisions. The analytical findings can lead to more effective marketing, new revenue opportunities, better customer service, improved operational efficiency and competitive advantages over rival business parties. By using big data carefully and intelligently, companies can take their development to a new level.

Big Data applications includes smart decision making, planning strategies cost reduction ,time reduction, new product development, risk strategies . Big data now guides every aspect of our modern society including mobile service ,retail ,manufacturing ,financial services ,life science and physical sciences, banking, education ,manufacturing, retail, government, health sector, economic, energy disaster management etc.

Big data is not limited to the current running data on Internet but it also refers to everything that is being stored digitally. Within the past decade, everything from banking transactions to medical history has migrated from physical documents to digital storage. This has revolutionized the big data analysis. Scientific research has been transformed by big data sharing and analysis . Thus, Big data is a research frontier that can accelerate progress.[5]

The huge demands on big data processing imposes a heavy load on computation, storage, and communication so it requires exponential increase in computer power, memory storage .This also requires creation of large data sets and consequently the storage of large amounts of data. There should be an efficient transfer technique to move large amounts of data quickly and easily without impacting other users or applications. The challenge is to build an efficient network infrastructure.

 

To analyze and process the big data, it requires a clear understanding of all the above requirements and challenges. Thus this paper discusses the issues of its attributes and requirements. Section 2 describes the big data V's characteristics along with its challenges and requirements. Section 3 concludes the paper by highlighting future research directions.

## 2. V's Virtues of Big Data

"Big data technologies describe a new generation of technologies and architectures designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis.[2]

All Big data attributes or characteristics can be expressed on V's concept [1][3][5]. Here we have explored and extended Big data characteristics and challenges by correlating with V's concept-

**Volume**- Volume refers to the size of data that is being stored, analyzed and processed to get the objective. It doesn't refer to any specific quantity, yet it is often expressed in Petabytes, Terabytes, Exabytes and Zettabytes.
This is the fundamental virtue of Big data as it is the quantity of data which determines the value and prospective of concern data. This is the huge volume which actually makes big data.

The challenge to store such a vast amount of data on the devices. The traditional devices are not suitable for big data. Here is a demand for new hardware infrastructure with new software technologies.

**Velocity** -Velocity measures the speed of data formation, flow and collection. It represents the rate at which data arrive, the time in which it must be processed and the speed at which data moves around. Thus it refers to both transmission speed and processing speed. In depth view, the velocity is the pace with which the data is created, stored, analyzed and visualized. In the past, analyzing the data using batch processing was common practice. But today, data is created in real-time or near real-time.. it should be passed at the moment when it is created. Big data need to be analyzed sometimes while it is being generated, without ever putting it into databases.
Big data can be analyzed or processed with artificial intelligence techniques, data mining(extracting information), predictive analytics, the algorithm structure and implementation., data visualization(meaningful presentation), statistical analysis..
For meeting the transmission speed of big data, the communication network architecture should be designed in such a way to produce network digital highway.
Thus in context with speed, the challenges here are to provide big data analytics methods and parallel processing techniques to provide timely and required information.

**Variety** –Variety represents the type of big data. Big data can be structured data i.e. numeric data or records that can fit into tables in traditional databases and unstructured i.e. text documents, email, video, audio, sensors, stock ticker data. It involves homogeneity and heterogeneity of data types. These all types of data need to be sorted and accessed at one platform. Dealing with a variety of structured and unstructured data greatly increases the complexity of both storing and analyzing big data. The wide variety of data requires a different approach as well as different techniques to store all raw data. Variety of data can slow down the processing.
So, the challenge is to discover new databases to sort and handle the heterogeneity nature of data types.

**Value:** It refers to the quality of data**.** Data value measures worth and usefulness of data in making decisions. Thus, it points to usability of the data. 'Value' is the most important V of Big Data.
Some data is more valuable than other data – temporally, spatially, contextually, etc.[4] . The huge quantity of big data that is being stored is not always completely useful. So, it require data filtering to select worthy data to draw accurate conclusions.
Here the challenge is sorting and cleaning up data.

**Veracity:** Veracity refers to the reliability and accuracy of data. The quality of captured data can vary greatly, affecting accurate analysis as the data has not been validated and verified by any formal methodology. This unverified data compromises the loyalty of the information and affects the accuracy of conclusions drawn. Data is virtually worthless if it's not trustworthiness.
The challenge is to discover verification methods to keep the correct data and leaving the dirty data.

**Variability**: Variability means the quality of being changeable .Here it refers to change in data with time or inconsistency of data sets.
Most of the data items are time varying- the same data can be collected over and over with the different values based on time stamp.
Variability can slow the processes to handle and manage it. Daily, seasonal and event-triggered peak data loads can be difficult to manage. It requires estimation and prediction based analysis.
The challenge is to design methods for the time lag to impact data consistency.

 

**Vulnerable**: means the data at risk. The data that is being transferred on Internet is vulnerable specially on social media, banks and health sectors. . Big data carries big risks as most of the times the data sets are of personal and private information like credit card data, personal ID information and other sensitive assets. This data represents a severe security concern .It must be secured with respect to privacy and security laws - regulations. There should be security mechanism to safeguard big data. Defining security levels can also be the done [2].
Thus the challenge is to have security measures to maintain privacy and confidentiality.

**Variegated:** Variegated means divergent. Here we have related this to divergent sources of data. Data is loosely connected or disassociated as data comes from multiple sources. It difficult to link, match, cleanse and transform data across systems. This is also termed as complexity. Complexity measures the degree of interconnectedness and interdependence in big data structures These data need to be linked, connected and correlated in order to be able to grasp the information that is supposed to be conveyed by these data.
To get data in right shape, all systems that send data should be standardized. Here the challenge is to make synchronization across the data sources.
So Big Data program must have technologies to overcome the challenges that "V's" raises to get again '**V'** i.e. **Victory** to fulfil the objectives of big data.

## 2. Conclusion

Big data is an emerging trend. Big Data represent large amounts of data. In this era of Internet-mobile technology, flood of data is generated by machines ,networks, smartphones , human interaction on social web sites. So the size of available data has been growing at an increasing rate. For companies/organizations, this data can help researchers and businesses make valuable decisions that provide strategic competitive advantages. To get advantage over others, companies needs careful big data analysis. This paper has discussed the challenges in the life of big data .The challenges include capturing analysis storage, searching, sharing ,visualization transferring and privacy violations. To meet these challenges, scalable computing infrastructures, data storage frameworks, network infrastructure, parallel programming frameworks, data mining frameworks should be explored. The hope is to develop better and better techniques and technologies towards finding solutions for big data problems in future research trends.

## REFERENCES

1. A. A. TOLE "Big Data Challenges" *Database Systems Journal* vol. IV, no. 3/2013€
2. Gantz, J. and E. Reinsel. 2011. "Extracting Value from Chaos", IDC's Digital Universe Study, sponsored by EMC
3. S. Kaisler , J. Alberto ,F.Armour,W.Money, "Big Data: Issues and Challenges Moving Forward " 46th Hawaii International Conference on System Sciences , 1530-1605/12 IEEE ,2012
4. "Challenges and Opportunities with Big Data "A community white paper developed by leading researchers across the United States
5. Ishrapa, J.Anuradha ,"A brief introduction on big data 5vs characteristics and Hadoop Tehnology " International conference on Intelligent computing , communication and convergence (ICCC-2015) ,procedia Computer science 48(2015) 319-325,www.sciencedirect.com

 