

# The 17 V's Of Big Data

Arockia Panimalar.S<sup>1</sup>, Varnekha Shree.S<sup>2</sup>, Veneshia Kathrine.A<sup>3</sup>

<sup>1</sup> Assistant Professor, Department of BCA & M.Sc SS, Sri Krishna Arts and Science College, Tamilnadu, India

<sup>2,3</sup> III BCA, Department of BCA & M.Sc SS, Sri Krishna Arts and Science College, Tamilnadu, India

\*\*\*

**Abstract** - Big data has turned into a hot cake for many associations and can be more useful for the enterprises like banking, internet business, insurance and manufacturing and so forth to encourage their customers. Generally, at the point when the data was low in volume, it was effortlessly overseen and processed by traditional technologies. These technologies are unequipped for dealing with it as big data differs in terms of volume, velocity and value as compared to the other data. Researchers & practitioner have distinguished, characterized and explored big data in terms of its characteristics including volume, velocity, variety, value, virality, volatility, visualization, viscosity and validity. Be that as it may, these examinations have been turned out to be deficient due to the developing issues rehashed day by day. This paper has identified and defined the fourteen characteristics of big data and a new three characteristics of big data has been explored further to handle big data efficiently.

expect that volume of data will grow by 45% in the next two years, and few said it will be doubled [1]. Thus, big data is a moving target and requires more attention to capturing it, curate it, handle it and process it. Fig. 1 shows the exponential growth of big data volume with time.

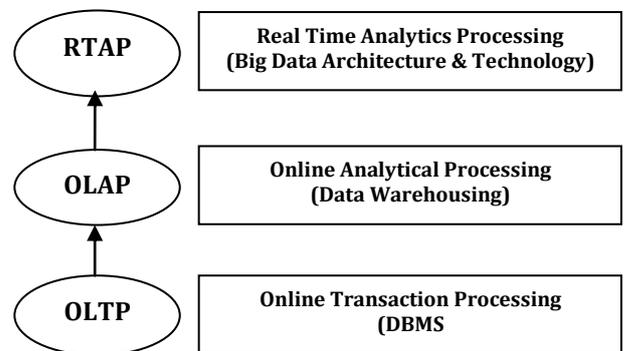


Fig 1: Growth of Data

**Key Words:** Big Data, Data, 14 V's, 1C, 17 V's, Big Data Characteristics

## 1. INTRODUCTION

Big data is a collection of data sets or a combination of data sets. The concept of big data has been endemic within digital communication and information science since the earliest days of computing. Big data is growing day by day because data is created by everyone and for everything from mobile devices, call centers, web servers, and social networking sites[1]. But the challenge is that it is too large, too fast and hard to handle for traditional database and existing technologies. Many organizations gather the massive amounts of data generated from high-volume transactions like call centers, sensors, web logs, and digital images. The success of their business depends on meeting big data challenges while continually improving operational efficiency.

Big data is continuously including more & more data sets with high volume beyond the capability of regularly used software tools to capture, curate, handle and process data set within a tolerable elapsed time. A huge amount of data sets is created every second from every part of the world i.e. the volume of data can never be reduce but increases day by day. Nearly five years ago, personal computer storage was tens to hundreds of gigabytes. Today IDC's Digital Universe Study predicts that between 2009 and 2020 digital information data will grow by 44% from 0.8 ZB to 35 ZB. Many surveys

## 2. EVOLUTION OF 14 V's AND 1C OF BIG DATA CHARACTERISTICS

### A. 3 V's of Big Data

Big data is a new idea, and it has got numerous definitions from researchers, organizations, and individuals. In 2001, industry analyst Doung Laney (currently with Gartner), articulated the mainstream of definition of big data in terms of three V's: Volume, Velocity, and Variety[2].

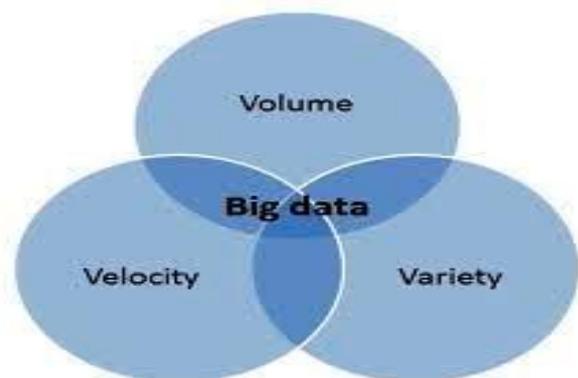


Fig 2: 3 V's of Big Data

### B. 4 V's of Big Data

SAS (Statistical Analysis System) has added two additional dimensions i.e. Variability and Complexity. Further, Oracle has defined big data in terms of four V's i.e. Volume, Velocity, Variety and Veracity[3].

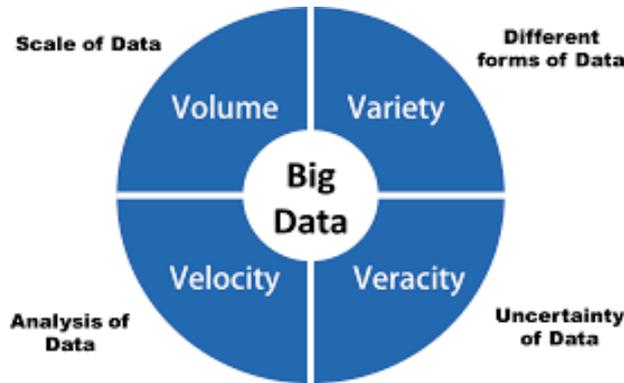


Fig 3: 4 V's of Big Data

### C. 5 V's of Big Data

Oguntimilehin. A, presented big data in terms of five V's as Volume, Velocity, Variety, Variability, Value and a Complexity[1].



Fig 4: 5 V's of Big Data

### D. 10 V's of Big Data

In 2014, Data Science Central, Kirk Born has defined big data in 10 V's i.e. Volume, Variety, Velocity, Veracity, Validity, Value, Variability, Venue, Vocabulary, and Vagueness [6].

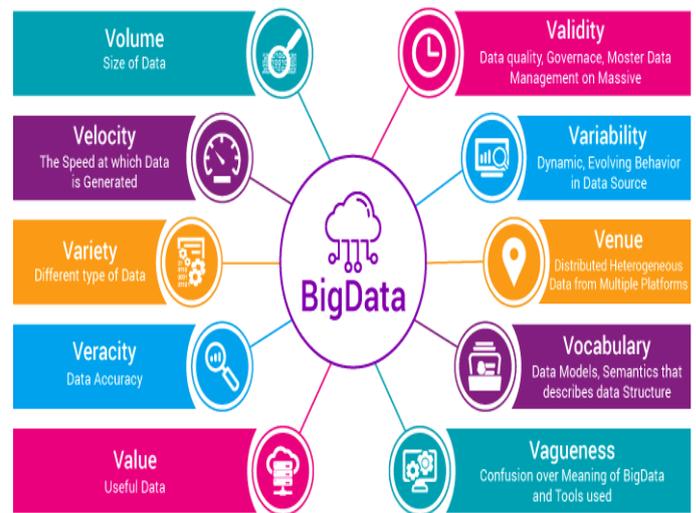


Fig 5: 10 V's of Big Data

### E. 14 V's of Big Data

All the big data characteristics has been listed and defined in table 1. These characteristics provide research horizon to the researcher and practitioners in order to effectively manage big data. The whole research in big data revolves around these characteristics (Fourteen V's and 1C defined in Table 1) in order to effectively manage and use big data efficiently & effectively. But some gap still exists which need to be addressed in order to get better insight in the area.

S. No	Big Data Character-istics	Elucidation	Description
1	Volume	Size of Data	Quantity of collected and stored data. Data size is in TB
2	Velocity	Speed of Data	The transfer rate of data between source and destination
3	Value	Importance of Data	It represents the business value to be derived from big data
4	Variety	Type of Data	Different type of data like pictures, videos and audio arrives at the receiving end
5	Veracity	Data Quality	Accurate analysis of captured data is virtually worthless if it's not accurate
6	Validity	Data Authenticity	Correctness or accuracy of data used to extract result in the form of information

7	Volatility	Duration of Usefulness	Big data volatility means the stored data and how long is useful to the user
8	Visualization	Data Act/ Data Process	It is a process of representing abstract
9	Virality	Spreading Speed	It is defined as the rate at which the data is broadcast /spread by a user and received by different users for their use
10	Viscosity	Lag of Event	It is a time difference the event occurred and the event being described
11	Variability	Data Differentiation	Data arrives constantly from different sources and how efficiently it differentiates between noisy data or important data
12	Venue	Different Platform	Various types of data arrived from different sources via different platforms like personnel system and private & public cloud
13	Vocabulary	Data Terminology	Data terminology likes data model and data structures
14	Vagueness	Indistinctness of existence in a Data	Vagueness concern the reality in information that suggested little or no thought about what each might convey
15	Complexity	Correlation of Data	Data comes from different sources and it is necessary to figure out the changes whether small or large in data with respect to the previously arrived data so that information can get quickly

**Table 1: 14 V's of Big Data Characteristics**

### 3. NEED FOR MORE EXPLORATION OF BIG DATA

S.Vikram Phaneendra and Madhusdhan Reddy discussed that the data was less and can be easily handled by RDBMS but

now-a-days it is not possible through RDMS tools, to manage big data. Because big data is different from other data in terms of five characteristics likes Volume, Variety, Velocity, Variability and Value[4].

Kiran Kumar Reddy and DnvsI Indira, have explained that big data is in the form of Structured, un-structured, massive homogenous and heterogeneous. They have also advised to use a better and modified model to handle and transfer of big data over the network [15].

Wei fan and Albert Bifet explored big data mining, as the capability of taking out the useful information from large data sets due to its characteristics likes volume, variability and velocity. It was not possible to do it before. So, researchers and practitioners have explored the big data in terms of volume, velocity, variety, variability, velocity, variety, value, virality, volatility, visualization, viscosity and validity [10].

In 2016, Experian discussed that 97% of US businesses are seeking to achieve a complete view of their customer, but the biggest problem organisations are facing is big data management [7]. US government accountability office has shown that five of the six most damaging data thefts of all time have happened in the last two years [8].

In November 2015, New York Times has published an article over DNA, which discusses about a genomics company in China. The company has been generating so many data that it could not be transmitted electronically. Hence instead of going up online they have been using disc data transfer method.

On the basis of the above discussion, it can be inferred that the research in big data exploring only these characteristics (Fourteen V's and a C) is not sufficient. In addition, because of the huge volume of big data, traditional methods for managing extracting and analyzing the same are not very useful as these may not provide accurate result for decision making. For using big data in a managed way, some more characteristics of the same must be explored and defined. As it is already discussed that the available research on big data is not sufficient to manage and use the same, it must be explored further to identify its more characteristics. The research on newly identified characteristics of big data may provide simple and effective management and use.

### 4. NEW 3 V's OF BIG DATA CHARACTERISTICS

Most of the organisations in today's world are dealing with huge amount of data. The big nature raises several issues. Though there are many solutions offered to solve the issues but as discussed in above section, problems still exist. This inspires to develop in depth understandings of big data. This will help to solve issues related to big data effectively. While going through the literature about big data, the study

present in the paper has dug out three more characteristics of it. These characteristics are defined as:

### A. Verbosity

Big data is a massive data that comes from different sources which may be structured or unstructured data, good/bad data. Bad data refers to the information which is wrong, out of date or incomplete. The consequences of storing these types of information may be dangerous sometimes. So, it is recommended to check that the stored data is secured, relevant, complete, and trustworthy. If a suitable technique at the initial stage is applied to decide whether the information is useful or not, then storage space, as well as processing time can be saved. Keeping in mind the verbose nature of the big data, we have identified 'verbosity' as one of the characteristic of big data which is defined as "The redundancy of the information available at different sources."

### B. Voluntariness

Big data is a set of huge amount of data which can be used as a volunteer by different organisations without any interference. Big data voluntarily help numerous enterprises. It assist retailers by giving them knowledge of customer preferences, urban planning by visualization of environment modelling and traffic patterns, manufacturers by predicting product issues to optimize their productivity and to improve the equipment and customers performance, energy companies to meet out energy demands during peak time and consequently increase production and improving efficiency by reducing the losses, healthcares professionals to prevent diseases and improving patient health [1], research organisations to obtain quality of research and revolutionize life science, physical science, medical science and scientific research[11][12], financial service organizations to identify and prevent fraud, government agencies to improve services in their respective fields. Keeping in mind the voluntary behaviour of the big data, 'voluntary' has been defined as one of the characteristic of big data which is defined as "The will full availability of big data to be used according to the context."

### C. Versatility

Big data is evolving to satisfy the needs of many organisations, researchers and Government. It facilitate the urban planning, environment modelling, visualization, analysis, quality classification, securing environment, computational analysis, biological understanding, designing and manufacturing process required by organisations and cost-effective models as well as elegant exploration of the result. Keeping in mind the resourceful/adaptable nature of the big data, we have identified 'versatility' as one of the characteristic of big data which is defined as "The ability of

big data to be flexible enough to be used differently for different context."

The three characteristics obtained in the research may prove to be a milestone for the purpose of research, if explored properly. Further research in these characteristics may resolve many issues related to big data. It may also help to differentiate the big data nature.

## 5. CONCLUSION

Big data is a collection of data sets which is growing day by day because data is created by everyone and for everything from mobile device and call centre. This paper revolves around the big data and its characteristics in terms of V's like volume, velocity, value, variety, veracity, validity, visualization, virality, viscosity, variability, volatility, venue, vocabulary, vagueness, and complexity. The day by day reported issues reflects that available research is not sufficient to manage and process big data. Hence the research presented in the paper has explored big data further to identify the three new 'V' characteristics i.e. Verbosity, Voluntariness, and Versatility. It is expected that the research on 17 V's and 1C (volume, velocity, value, variety, veracity, validity, visualization, virality, viscosity, variability, volatility, venue, vocabulary, vagueness, verbosity, voluntariness, and versatility and complexity) identified characteristics will provide simple and effective management of big data which can be used in value added applications and research environment.

## 6. REFERENCES

- [1]Oguntimilehin A., Ademola E.O. "A Review of Big Data Management, Benefits and Challenges," Journal of Emerging Trends in Computing and Information Sciences, vol-5, pp-433-437, June 2014.
- [2] Stephen Kaisler, Frank Armour, J.Alberto Espinosa and Wolliam Money "Big Data: Issues and Challenges Moving Forward," Hawaii International Conference on System Sciences 46th, pp-995-1003, 2013
- [3] Oracle (2013), --Information Management and Big Data: A Reference Architecture||, www.oracle.com/.../ infomgmt-big-data-retrieved 20/03/14.
- [4]S.Vikram Phaneendra & E.Madhusudhan Reddy "Big Data-solutions for RDBMS problems- A survey," In 12th IEEE/IFIP Network Operations & Management Symposium (NOMS 2010) (Osaka, Japan, Apr 19{23 2013).
- [5]<https://www.youtube.com/watch?v=S89o3INzIjc>
- [6]<http://www.datasciencecentral.com/profiles/blogs/top-10-list-the-v-s-of-big-data>

[7]<http://www.inc.com/bill-carmody/biggest-problem-with-big-data-management-in-2016.html>

[8]<http://www.kdnuggets.com/2015/08/iegbigdatainnovation-boston-4-problems.html>

[9]<http://data-magnum.com/how-many-vs-in-big-data-the-characteristics-that-define-big-data>

[10]<http://data-magnum.com/how-many-vs-in-big-data-the-characteristics-that-define-big-data>

[11]<http://www.ibmbigdatahub.com/presentation/current-challenges-and-opportunities-bigdata-and-analytics-emergency-management>

[12]<http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>

[13]<https://software.intel.com/sites/default/files/article/402274/etl-big-data-with-hadoop.pdf>

[14]<http://www.nytimes.com/2013/03/12/technology/new-us-research-will-aim-at-flood-of-digital-data.html>

[15]Kiran kumara Reddi & DnvsI Indira "Different Technique to Transfer Big Data: survey," IEEE Transactions on 52(8) (Aug.2013)