CHANDIGARH UNIVERSITY

Discover. Learn. Empower.

Apex Institute of Technology

Department of Computer Science & Engineering

Introduction to Data Science

(CSF-284)

Dr. Jitender Kaushal
Associate Professor
CSE(AIT), CU

DISCOVER . LEARN . EMPOWER

Big Data Characteristics

INTRODUCTION

- 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.
- 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**.
- But the **challenge** is that it is **too large, too fast, and hard to handle** for **traditional databases and existing technologies**.
- Many organizations gather massive amounts of data generated from **high-volume transactions like call centers, sensors, weblogs, and digital images**. The **success** of their business depends on **meeting big data challenges** while continually **improving operational efficiency**.

- Big data is continuous including **more & more data sets with high volume** beyond the capability of regularly used software tools **to capture, curate, handle and process data sets 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 reduced but increases day by day**.
- Nearly five years ago, personal computer storage was tens to hundreds of gigabytes. International Data Corporation: **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 expect that volume of data will grow by **45%** in the next two years, and few said it will be doubled. Thus, big data is a moving target and requires more attention to capture it, curate it, handle it and process it. Fig. 1 shows the exponential growth of big data volume with time.

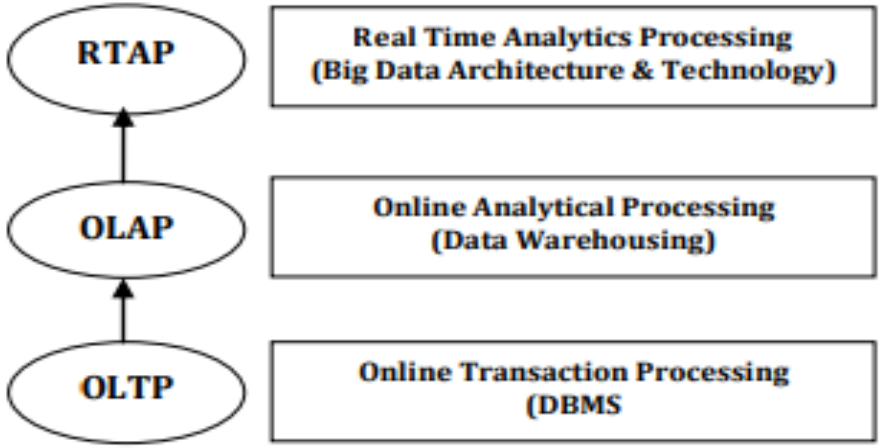
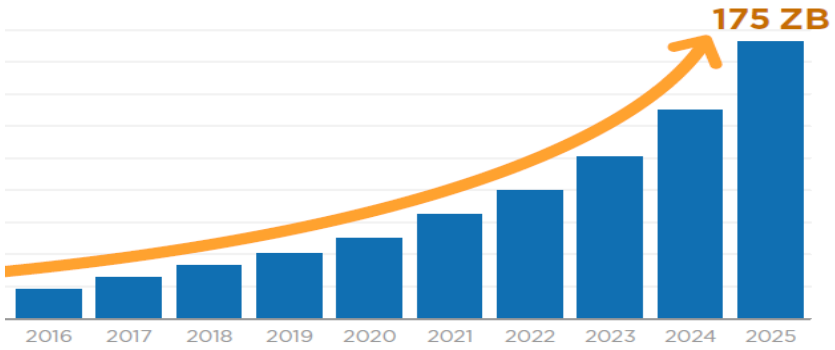


Fig 1: Growth of Data



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

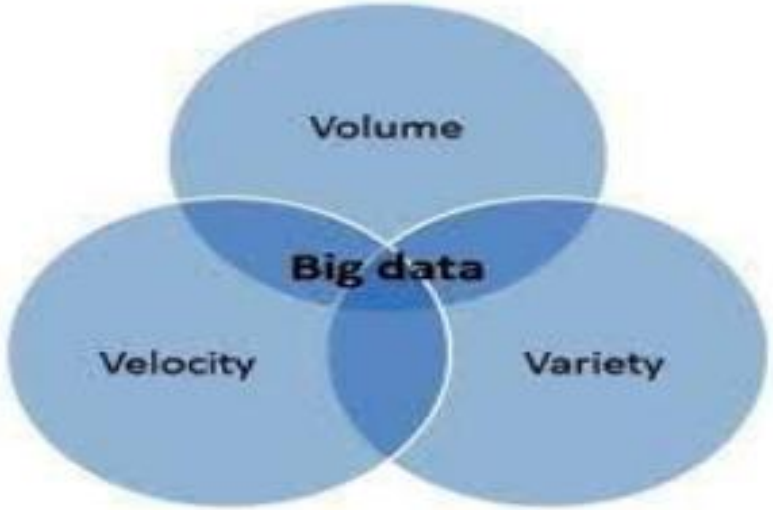


Fig 2: 3 V's of Big Data

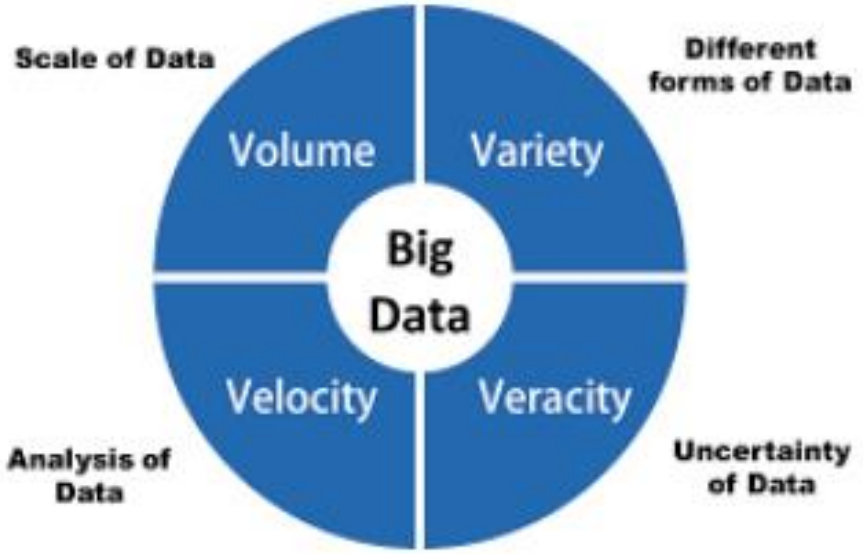
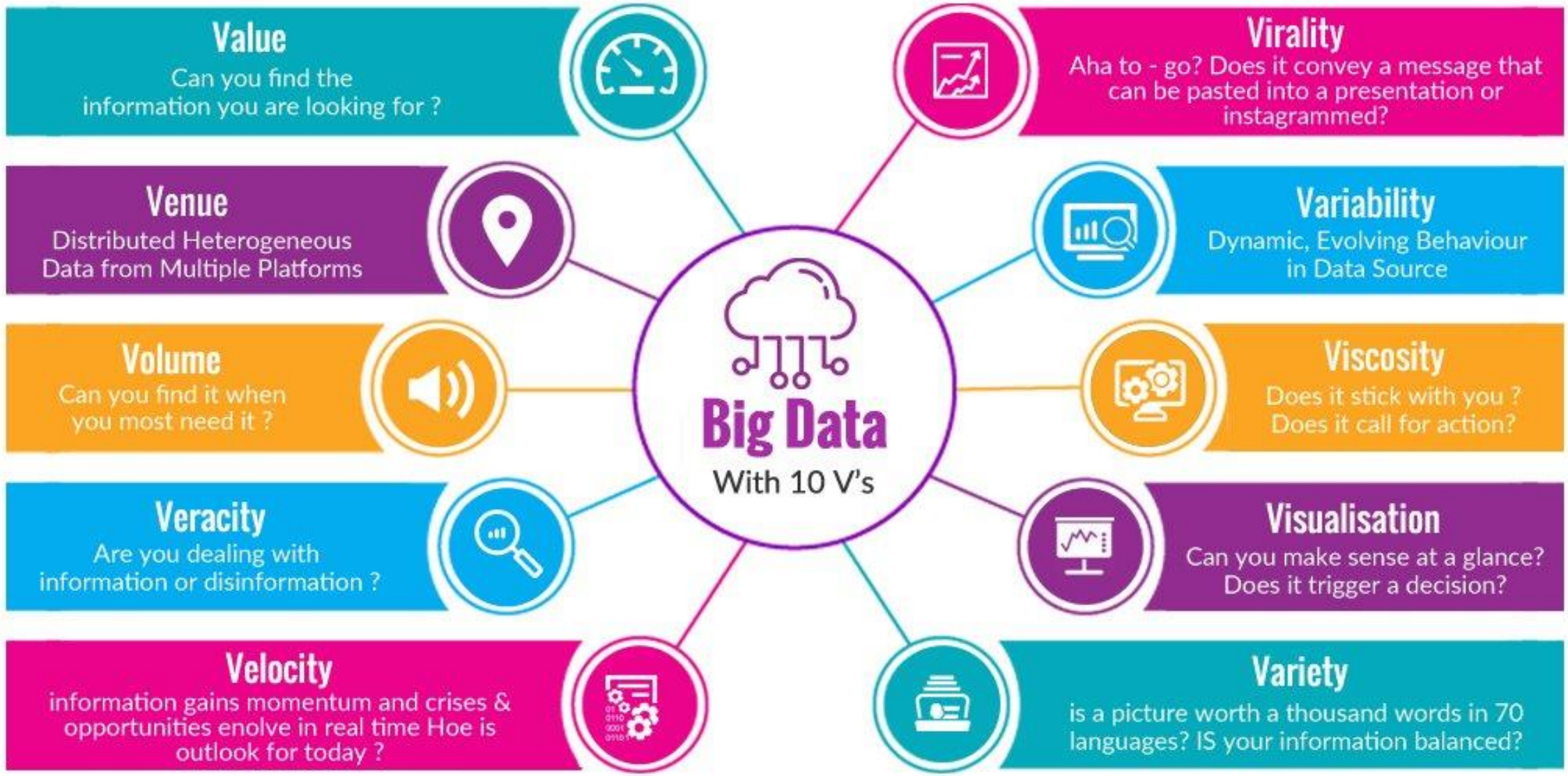


Fig 3: 4 V's of Big Data

Fig 4: 5 V's of Big Data


Value
Can you find the information you are looking for ?
Venue
Distributed Heterogeneous Data from Multiple Platforms
Volume
Can you find it when you most need it ?
Veracity
Are you dealing with information or disinformation ?
Velocity
information gains momentum and crises & opportunities enolve in real time Hoe is outlook for today ?
Big Data
With 10 V's
Virality
Aha to - go? Does it convey a message that can be pasted into a presentation or instagrammed?
Variability
Dynamic, Evolving Behaviour in Data Source
Viscosity
Does it stick with you ? Does it call for action?
Visualisation
Can you make sense at a glance? Does it trigger a decision?
Variety
is a picture worth a thousand words in 70 languages? IS your information balanced?

14 V's of Big Data

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

NEED FOR MORE EXPLORATION OF BIG DATA

1. If data was less and can be easily handled by RDBMS but nowadays it is not possible through RDMS tools, to manage big data. Because **big data is different from other data in terms of five characteristics like Volume, Variety, Velocity, Variability, and Value**.

2. Big data is in the form of Structured, un-structured, massive homogenous and heterogeneous. Therefore it is required to use a **better and modified model to handle and transfer of big data over the network.**

- The capability of taking out useful information from large data sets due to its characteristics like volume, variability, and velocity. It was not possible to do it before. So, exploration of big data in terms of volume, velocity, variety, variability, velocity, variety, value, virality, volatility, visualization, viscosity, and validity.

NEW 3 V's OF BIG DATA CHARACTERISTICS

The 3 V's of Big Data characteristics are defined as:

A. **Verbosity:** Big data is massive data that **comes from different sources** which may be **structured or unstructured data, good/bad data**.

- Bad data refers to information that 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 big data, **'verbosity' is one of the characteristics 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 amounts of data that can be used as a **volunteer by different organizations without any interference**. Big data voluntarily help numerous enterprises.
- It assists **retailers** by giving them knowledge of customer preferences, **urban planning** by visualization of the environment modeling and traffic patterns, **manufacturers** by predicting product issues to optimize their productivity and to improve the equipment and customers' performance, **energy companies** to meet energy demands during peak time and consequently, increase productivity and improve efficiency by reducing losses, **healthcare** professionals to prevent diseases and improve patient health, research organizations to obtain the quality of research and revolutionize life science, physical science, medical science, and scientific research **financial service organizations** to identify and prevent fraud, and government agencies to improve services in their respective fields.
- Keeping in mind the voluntary behavior of the big data, **'voluntary' has been defined as one of the characteristics of big data** which is defined as **"full availability of big data to be used according to the context."**

C. Versatility

- Big data is evolving to satisfy the needs of **many organizations, researchers, and Governments**. It facilitates urban planning, environment modeling, visualization, analysis, quality classification, securing environment, computational analysis, biological understanding, designing and manufacturing process required by organizations, and cost-effective models as well as elegant exploration of the result.

- Keeping in mind the resourceful/adaptable nature of big data, we have identified **'versatility' as one of the characteristics of big data** which is defined as **"The ability of big data to be flexible enough to be used differently for a different context."**

Thank You