1.2 Big Data Characteristics

The primary

Characteristics of Big Date are:

1. Volume

Volume refers to the huge amount of data that is collected and generated every second in large organizations. This data is generated from different sources such as IoT devices, social media, victors, financial transaction and customer logs.

data was a problem earlier. But now distributed by stem such as Hadrop are used for organizing data collected from all these sources. The size of the data is crucial for understanding its value. The data is crucial for understanding its value. Also, the volume is useful in determining whether a collection of data is Big Data or not

2. Variety refers to the different sources of date of date and their nature. The sources of date have changed over the years. Earlier, it was have changed over the years. Earlier, it was only available in spreadsheet and database.

Nowadays, date is present in Shotox, audio,



PARSHWANATH CHARITABLE TRUST'S

A.P. SHAH INSTITUTE OF TECHNOLOGY Department of Computer Science and Engineering Data Science

CSE DATA SCIENCE

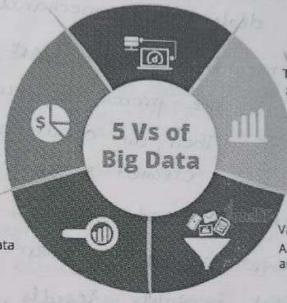
The variety of date is crucial for its storage and analysis. A variety of data can be classified unto three distinct parts:

- Structured
- Semi- structured
- Un-structured

Velocity The speed at which data is emanating and changes are occurring between diverse datasets

Value

The value that can be derived from accessing and analysing big data



The sheer volume of data being generated every second

Veracity

The discrepancies found in data

Variety

A combination of data types that are being dumped into the system

Characteristics of Big Date or 5V

Velocity refers to the speed at which the data is created or generated. This speed of data producing is also related to how fast the date is going to be processed. This is because only after analysis and processing, the date can meet

Subject Incharge; Prof. Richa Singh

Page No.

Department of CSE-Data Science | APSIT

the demands of the clients/users.

Massive amount of data are produced from sensors, social media sites and application logs and all of its is continuous. If the data flow is not continuous, there is no point in investing time or effort on it.

4. Value

Value deals with a mechanism to bring out the correct meaning of data. First of all, we need to mine data, i.e. the process to turn raw data to mine data. Then, an analysis is done on the into useful data. Then, an analysis is done on the data that we have cleaned or retrieved from the raw data. Then, we need to make sure whatever analysis we have done benefits our business, such as in finding out insights, results, etc.,

Veracity defines the degree of trustworth Veracity defines the degree of trustworth iness of the data. Veracity refers to the ansurances of quality / integrity / credibility / accuracy of the data we encounter is dete. As most of the data we encounter is unstructured, it is important to filter out the unstructured, it is important to filter out the unnecessary information and use the rest for