1.2 Big Data Characteristics

The primary Characteristics of Big Data 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, videos, financial transaction and customer logs.

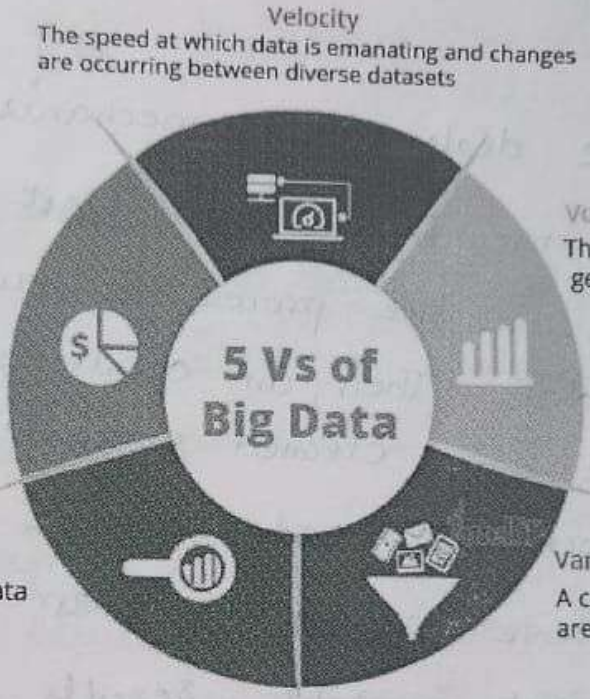
Storing and processing this huge amount of data was a problem earlier. But now distributed system such as Hadoop are used for organizing data collected from all these sources. The size of the data is crucial for understanding its value. Also, the volume is useful in determining whether a collection of data is Big Data or not

2. Variety :

Variety refers to the different sources of data and their nature. The sources of data have changed over the years. Earlier, it was only available in spreadsheet and database. Nowadays, data is present in photos, audio,

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

- Structured
- Semi-structured
- Un-structured



Characteristics of Big Data or 5V

3. Velocity :- 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 data is going to be processed. This is because only after analysis and procesing, the data can meet

the demands of the clients / users.
Massive amount of data are produced from sensors, social media sites and application logs and all of it 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 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.,

5. Veracity :-

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