What is Big Data?

- Data which are very large in size is called Big Data.
- **Big Data** is a collection of data that is huge in volume, yet growing exponentially with time.
- Normally we work on data of size MB or maximum GB but data in Peta bytes i.e. 10^15 byte size is called Big Data.
- Big data is also a data but with huge size.

Types of Big Data

Following are the types of Big Data

1. Structured
2. Unstructured
3. Semi-Structured

1. Structured

Any data that can be stored, accessed and processed in the form of fixed format is termed as a 'Structured 'data.

Data that is stored in a relational database management system is an example of structured data.

for example employee table ,which is present in database , the data is an structure format

Emp ID	Emp Name	Gender	Department	Salary
232	Palvi	Female	Finance	50,000

2. Unstructured

Any data with unknown form or the structure is classified as unstructured data. This makes it very difficult and time-consuming to process and analyze unstructured data.

A typical example of unstructured data is a heterogeneous data source containing a combination of simple text files, images, videos etc.

Email is an example of unstructured data.

3. Semi-Structured

Semi structured is the third type of big data.

Semi-Structured data can contain both the forms of data. Example of semi-structured data is a data represented in an XML file.

The web application data that is unstructured contains transaction history files, log files, etc.

Sources of BIG DATA

These data come from many sources like

Social Networking Sites: Facebook, Google, LinkedIn all these sites generates huge amount of data on a day to day basis as they have billions of users worldwide.

E-Commerce Sites: Sites like Amazon, flipkart generates huge amount of logs which users buying trends can be traced.

Weather Station: All the weather station and satellite gives very huge data which are stored and manipulated to forecast weather.

Telecom Company: Telecom giants like Airtel, Vodafone study the user trends and accordingly publish their plans and for this they store the data of its million users.

Share Market: Stock exchange across the world generates huge amount of data through its daily transaction.

Characteristics of Big Data

In recent years, Big Data was defined by the "3Vs' but now there is ”5Vs” which are also termed as the characteristics of Big Data as follows:

1. Volume

- Hence while dealing with Big Data it is necessary to consider a characteristic Volume
- Volume is one of the characteristics of big data. We already know that Big Data indicates huge 'volumes' of data that is being generated on a daily basis from various sources like social media platforms, machines, networks, human interactions, etc.

2) Velocity

- Velocity essentially refers to the speed at which data is being created in real-time.
- In Big Data velocity data flows in from sources like machines, networks social media mobile phones etc.
- **Example** there are more than 3.5 billion searches per day are made on Google. Also FaceBook users are increasing by 22%(Approx.) year by year.

3) Variety:

- It refers to nature of data that is structured, semi-structured and unstructured data.
- It also refers to heterogeneous sources.
- It refers to structured, unstructured, and semi structured data that is gathered from multiple sources.

I. **Structured data:** This data is basically an organized data generally refers to data that has defined the length and format of data.
II. **Semi Structured**: this data is basically a semi Organized data. It is generally a form of data that do not conform to the formal structure of data
III. **Unstructured data:** This data basically refers to unorganized data.

4) Veracity:

- It refers to inconsistencies and uncertainty in data. That is available can sometimes get messy and quality and accuracy are difficult to control.
- Veracity is the process of being able to handle and manage data efficiently..
- For example, Facebook posts with hash tags.

5) Value:

Value is an essential characteristic of big data. It is not the data that we process or store. It is **valuable** and **reliable** data that we **store, process**, and also **analyze**.

Benefits of Big Data

Below are the top advantages of using big data in business

- Better decision making
- Greater innovations
- Improvement in education sector
- Recommendation engines
- life saving application in healthcare industry