# **Data-** Data is collection of facts ,(number of word, measurement and observation) that has been translated in a form that computers can process. And also, data is everything which is present around you is called data.

**Uses of Data**

It helps in understanding more about the data by identifying relationship that may exist between two variables.

Data is instead the raw materials which when processed when processed and given meaning by people it becomes information.

**How its useful for us:-**

>Data is useful for us because

1-It saves our time

2-Data can correct problem in real time.

3-Data can be used market better.

4-Data helps us to make decision in every aspect of your business.

Que-What is Big data.. ?  
Big Data :- Big data refers to large ,diverse sets of information that grows at ever increasing rates .

It encompasses the volume of information the velocity or speed at which it is created and collected the variety scope of the data points being covered.  
Its collection of data from many different sources.  
It’s a data with so large size and complexity .  
  
Que-Differentiate between structured ,semi -structured and unstructured data .  
:-Structured Data - Structured data exists in a format created to be captured, stored, organized and analysed.   
 It is neatly organized for easy access. If structured data was an office it would contain many file cabinets that are efficiently set up, clearly labelled and easy to access.

For that reason, structured data brings inherent benefits when dealing high volumes of information.   
  
:-Unstructured Data- Unstructured data is information that is not arranged according to a present data model or schema, and therefore cannot be stored in a traditional relational database or RDBMS. Text and multimedia are two common types of unstructured content. Many business documents are unstructured, as are email messages, videos, photos, webpages, and audio files.  
  
:-Semi-Structured Data- Semi-structured data is data that does not conform to a data model but has some structure. It lacks a fixed or rigid schema. It is the data that does not reside in a rational database but that have some organizational properties that make it easier to analyse. With some processes, we can store them in the relational database.   
Que-Difference between Data and Information:-  
1.Data is unorganized while Information is structured or organized.  
2.Data does not depends on information while information depends on data.  
3.Data is an individual unit that contains raw material which do not carry any specific meaning while Information is a group of data that collectively carries a logical meaning.  
4.Example of data is a :-student’s test score . Example of information: -the average score of a class is the information derived from the given data.

* Que-What is Big data..?  
  Big Data :- Big data refers to large ,diverse sets of information that grows at ever increasing rates .It encompasses the volume of information the velocity or speed at which it is created and collected the variety scope of the data points being covered.  
  Its collection of data from many different sources .  
  It’s a data with so large size and complexity .  
    
  What are the Quantitative data and Qualitative data.?  
  Quantitative Data :- Quantitative data refers to any information that can be quantified. If it can be counted or measured, and given a numerical value, it’s quantitative data. Quantitative data can tell you “how many,” “how much,” or “how often”—for example, how many people attended last week’s webinar? How much revenue did the company make in 2019? How often does a certain customer group use online banking?  
    
  Qualitative Data: - Unlike quantitative data, qualitative data cannot be measured or counted. It’s descriptive, expressed in terms of language rather than numerical values.  
    
  For example-Researchers will often turn to qualitative data to answer “Why?” or “How?” questions. For example, if your quantitative data tells you that a certain website visitor abandoned their shopping cart three times in one week, you’d probably want to investigate why—and this might involve collecting some form of qualitative data from the user.

Que-What are the different V’s in Big Data?  
1. Volume:  
>The name ‘Big Data’ itself is related to a size which is enormous.  
>Volume is a huge amount of data.  
>To determine the value of data, size of data plays a very crucial role. If the volume of data is very large then it is actually considered as a ‘Big Data’. This means whether a particular data can actually be considered as a Big Data or not, is dependent upon the volume of data.  
  
2.Velocity :  
>refers to the high speed of accumulation of data.  
>In Big Data velocity data flows in from sources like machines, networks, social media, mobile phones etc.  
  
3. Variety:  
>It refers to nature of data that is structured, semi-structured and unstructured data.  
>It also refers to heterogeneous sources.  
>Variety is basically the arrival of data from new sources that are both inside and outside of an enterprise. It can be structured, semi-structured and unstructured.  
  
4. Veracity:  
>It refers to inconsistencies and uncertainty in data, that is data which is available can sometimes get messy and quality and accuracy are difficult to control.  
>Big Data is also variable because of the multitude of data dimensions resulting from multiple disparate data types and sources.  
  
5.Value:   
>After having the 4 V’s into account there comes one more V which stands for Value!. The bulk of >Data having no Value is of no good to the company, unless you turn it into something useful.  
>Data in itself is of no use or importance but it needs to be converted into something valuable to extract Information. Hence, you can state that Value! is the most important V of all the 6V’s.  
  
6. Variability:  
>How fast or ,available data that extent is the structure of your data is changing ?.  
>How often does the meaning or shape of your data to change?.  
>example : if you are eating same ice-cream daily and the taste just keep changing.  
  
Que-Name some popular tools used in Big Data?  
1-APACHE Hadoop:-  
Hadoop also offers cross-platform support for its users. Today, it is the best big data analytic tool and is popularly used by many tech giants such as Amazon, Microsoft, IBM, etc.  
>Free to use and offers an efficient storage solution for businesses.  
>Offers quick access via HDFS (Hadoop Distributed File System).  
 2-Cassandra:-  
>APACHE Cassandra is an open-source NoSQL distributed database that is used to fetch large amounts of data.   
>It’s one of the most popular tools for data analytics and has been praised by many tech companies due to its high scalability and availability without compromising speed and performance.   
3-Qubole:-  
>It’s an open-source big data tool that helps in fetching data in a value of chain using ad-hoc analysis in machine learning. Qubole is a data lake platform that offers end-to-end service with reduced time and effort which are required in moving data pipelines.  
>Supports ETL process: It allows companies to migrate data from multiple sources in one place.  
4-Xplenty:-  
>It is a data analytic tool for building a data pipeline by using minimal codes in it. It offers a wide range of solutions for sales, marketing, and support. With the help of its interactive graphical interface, it provides solutions for ETL, ELT, etc. >The best part of using Xplenty is its low investment in hardware & software and its offers support via email, chat, telephonic and virtual meetings.   
5-Spark:-  
>APACHE Spark is another framework that is used to process data and perform numerous tasks on a large scale.   
>It is also used to process data via multiple computers with the help of distributing tools.  
6-Mongo DB:-  
Came in limelight in 2010, is a free, open-source platform and a document-oriented (NoSQL) database that is used to store a high volume of data. It uses collections and documents for storage and its document consists of key-value pairs which are considered a basic unit of Mongo DB.

Q- What are the different types of data? Explain.

There are different types of data-

1. Qualitative Data Type- Qualitative or Categorical Data describes the object under consideration using a finite set of discrete classes. It means that this type of data can’t be counted or measured easily using numbers and therefore divided into categories. The gender of a person (male, female, or others) is a good example of this data type.
2. Quantitative Data Type- This data type tries to quantify things and it does by considering numerical values that make it countable in nature. The price of a smartphone, discount offered, number of ratings on a product, the frequency of processor of a smartphone, or ram of that particular phone, all these things fall under the category of Quantitative data types.
3. Nominal- These are the set of values that don’t possess a natural ordering. Let’s understand this with some examples. The color of a smartphone can be considered as a nominal data type as we can’t compare one color with others.
4. Ordinal- These types of values have a natural ordering while maintaining their class of values. If we consider the size of a clothing brand then we can easily sort them according to their name tag in the order of small < medium < large. The grading system while marking candidates in a test can also be considered as an ordinal data type where A+ is definitely better than B grade.
5. Discrete - The numerical values which fall under are integers or whole numbers are placed under this category. The number of speakers in the phone, cameras, cores in the processor, the number of sims supported all these are some of the examples of the discrete data type.
6. Continuous - The fractional numbers are considered as continuous values. These can take the form of the operating frequency of the processors, the android version of the phone, wifi frequency, temperature of the cores, and so on.