1. Q Differentiate between data and information?

|  |  |
| --- | --- |
| **Data** | **Information** |
| Data is unorganised and unrefined facts | Information comprises processed, organised data presented in a meaningful context |
| Data is an individual unit that contains raw materials which do not carry any specific meaning. | Information is a group of data that collectively carries a logical meaning. |
| Data doesn’t depend on information. | Information depends on data. |
| Raw data alone is insufficient for decision making | Information is sufficient for decision making |
| An example of data is a student’s test score | The average score of a class is the information de |

2 Q How data is useful for us?

Ans-Data is important now because it provides valuable information that can be used to make informed decisions, solve problems, and drive progress in a wide range of fields, including business, healthcare, education, and government.

1. In business, data can be used to identify patterns and trends in customer behavior, sales, and marketing, which can inform strategic decisions and help companies improve their products and services.
2. In healthcare, data can be used to track the spread of diseases, predict outbreaks, and evaluate the effectiveness of treatments.
3. In education, data can be used to track student progress and inform teaching strategies.
4. In government, data can be used to track economic indicators, crime rates, and other important metrics.
5. In the field of AI, data is the backbone for training machine learning models, enabling them to make predictions and decisions.
6. In short, data is important now because it provides valuable information that can be used to make informed decisions, solve problems, drive progress and make predictions in many fields.

Q 3 What is Big data?

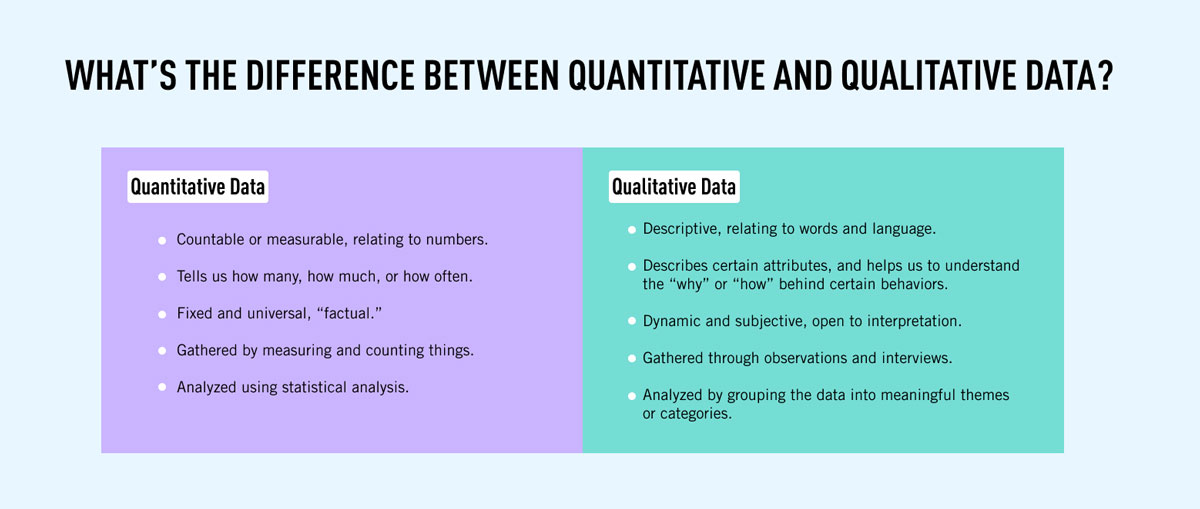
* Ans - Big Data in very simple words is a collection of large volumes of data (structured and unstructured) with has velocity and different variety of information asset. Big Data is associated with the concept of 3 V that is volume, velocity, and variety. In other words, data that is the range of 100s of TBs or PB comes into Big Data. But it doesn't mean the amount of data, the thing matters is what organization do with data. Big Data is analyzed for insights that lead to better decisions.

Q 4 **What are structured unstructured and semi-structured data?**

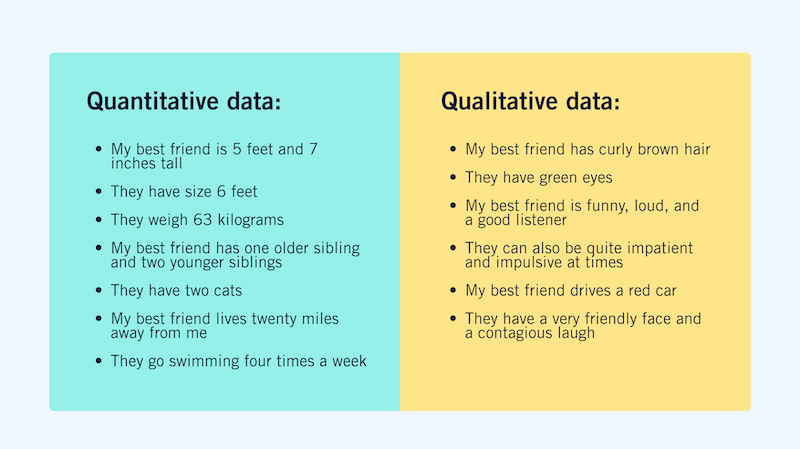
Ans - Structured data refers to data that is organized in a specific format and follows a predefined schema. Examples of structured data include spreadsheets, relational databases, and tables.

Unstructured data refers to data that does not have a predefined structure or schema. Examples of unstructured data include emails, social media posts, videos, images, and text documents. This type of data can be difficult to process and analyze using traditional data processing methods.

Semi-structured data is a combination of structured and unstructured data. It contains some organization or structure, but also includes unstructured elements. Examples of semi-structured data include XML and JSON files, which have a structure but also allow for variability in the data that is contained within them. Semi-structured data can be easier to work with than unstructured data, but may still require specialized tools and techniques for processing and analysis.

[Q 5 what are the quantitative and qualitative data](https://www.quora.com/profile/Robin-Thomas-16" \t "_blank)? A

Example –



Q 6 Name some popular tools used in big data analytics?

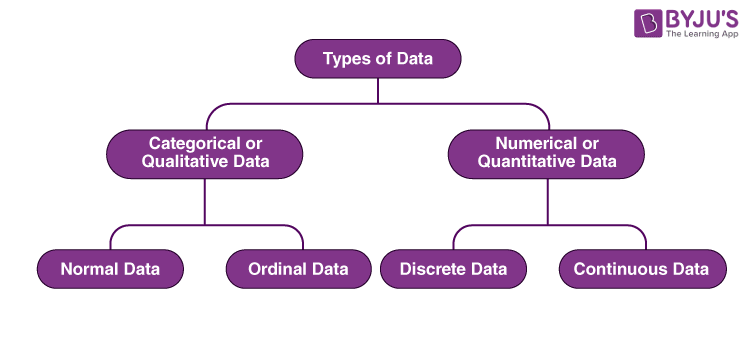
* Ans-Hadoop - helps in storing and analyzing data.
* MongoDB - used on datasets that change frequently.
* Talend - used for data integration and management

Cassandra - a distributed database used to handle chunks of data.

* Spark - used for real-time processing and analyzing large amounts of data.

Q 7 What are the diffrent type of data? Explain

Ans -The data is classified into majorly four categories:

* Nominal data
* Ordinal data
* Discrete data
* Continuous data
* Further, we can classify these data as follows:
* 
* Let us discuss the different types of data in Statistics herewith examples.
* **Qualitative or Categorical Data**
* Qualitative data, also known as the [categorical data](https://byjus.com/maths/categorical-data/), describes the data that fits into the categories. Qualitative data are not numerical. The categorical information involves categorical variables that describe the features such as a person’s gender, home town etc. Categorical measures are defined in terms of natural language specifications, but not in terms of numbers.
* Sometimes categorical data can hold numerical values (quantitative value), but those values do not have a mathematical sense. Examples of the categorical data are birthdate, favourite sport, school postcode. Here, the birthdate and school postcode hold the quantitative value, but it does not give numerical meaning.
* **Nominal Data**
* Nominal data is one of the types of qualitative information which helps to label the variables without providing the numerical value. Nominal data is also called the nominal scale. It cannot be ordered and measured. But sometimes, the data can be qualitative and quantitative. Examples of nominal data are letters, symbols, words, gender etc.
* The nominal data are examined using the grouping method. In this method, the data are grouped into categories, and then the frequency or the percentage of the data can be calculated. These data are visually represented using the pie charts.
* **Ordinal Data**
* Ordinal data/variable is a type of data that follows a natural order. The significant feature of the nominal data is that the difference between the data values is not determined. This variable is mostly found in surveys, finance, economics, questionnaires, and so on.
* The ordinal data is commonly represented using a bar chart. These data are investigated and interpreted through many visualisation tools. The information may be expressed using tables in which each row in the table shows the distinct category.
* **Quantitative or Numerical Data**
* Quantitative data is also known as numerical data which represents the numerical value (i.e., how much, how often, how many). Numerical data gives information about the quantities of a specific thing. Some examples of numerical data are height, length, size, weight, and so on. The quantitative data can be classified into two different types based on the [data sets](https://byjus.com/maths/data-sets/). The two different classifications of numerical data are discrete data and continuous data.
* **Discrete Data**
* Discrete data can take only discrete values. Discrete information contains only a finite number of possible values. Those values cannot be subdivided meaningfully. Here, things can be counted in whole numbers.
* **Example:** Number of students in the class
* **Continuous Data**
* Continuous data is data that can be calculated. It has an infinite number of probable values that can be selected within a given specific range.
* **Example:** Temperature range