1. Different between data information?

DATA

* Word 'DATA' is Originated from the word 'Datum' that means single piece of information.
* Collection of a distinct small unit of information. It can be used in variety of forms like text, numbers, media, bytes etc...

INFORMATION

* Processed data is information which is accurate, systematize, understandable, relevant and timely.
* Information = data + Meaning.
* Information is organized.

Example:

₹33,000, ₹25.000, ₹28,500

Here have an example of some numbers the shown set of numbers is data but it describes nothing. We do not whether it is the salary of person or rent of a house.

Furniture Whole Sale & Retail:

First months sale = ₹33,000

Second month sale = ₹25,000

Third month sale = ₹28,800

Does it make any sense now? Yes, if it is the sale of some products month wise you can see here this is information.

2. How data is useful for us?

Today, in the internet-focused world we live in, data is widely considered to be among the world's most valuable resources because of how much potential revenue and business value it can provide. Moreover, it's available for sale and purchase.

3. What is big data?

Big data analytics is the process of examining, cleaning, transforming, and modeling large sets of data to discover useful information, inform business decisions, and support decision-making. The goal of big data analytics is to uncover patterns, trends, and insights that would be difficult or impossible to detect using traditional methods.

4. Difference between structured, semi structured, and unstructured data?

* Structured data

Structured data is organized in tabular format (ie. Rows & columns). A data structure is a way to store and organize data in a system so that we can use it efficiently.

* Semi – structured data

Semi-structured data refers to data that is not captured or formatted in conventional ways. Semi-structured data does not follow the format of a tabular data model or relational databases because it does not have a fixed schema

* Unstructured data

One of the most common types of unstructured data is **text**. Unstructured text is generated and collected in a wide range of forms, including Word documents, email messages, PowerPoint presentations, and survey responses.

5. What are quantitative data and qualitative data?

Quantitative data is numerical data that can be measured and expressed in numbers or quantities. Examples of quantitative data include things like age, weight, height, and temperature. This type of data can be analyzed using statistical methods and is often used in scientific research to support or disprove hypotheses.

Qualitative data is non-numerical data that describes or characterizes a phenomenon. It is often used to understand people's attitudes, beliefs, or experiences. Examples of qualitative data include things like interview transcripts, open-ended survey responses, and unstructured observations.

6. What are the different V’s in big data?

* Volume  
  Huge amount of data.
* Variety  
  Different formats of data from various sources.
* Velocity  
  High speed of accumulation of data.
* Value  
  Extract useful data.
* Validity/Veracity  
  Inconsistencies and uncertainty in dada.

7. Name some popular tools used in big data?

* Apache Spark.
* Apache Hadoop.
* Apache Flink.
* Google Cloud Platform.
* MongoDB.
* Sisense.

8. What are the different types of data? Explain.

* **Structured data**: This refers to data that is organized in a specific format, such as rows and columns in a spreadsheet or database table. Structured data is easy to search, sort, and filter, and is often used in traditional databases and business intelligence systems.
* **Unstructured data**: This refers to data that does not have a specific format, such as text documents, images, and videos. Unstructured data is more difficult to analyze and process, but can provide valuable insights when properly analyzed.
* **Semi-structured data**: This refers to data that has some structure, but not as much as fully structured data. Semi-structured data includes formats such as JSON and XML, which contain both structured and unstructured elements.
* **Transactional data**: This refers to data that is generated as a result of business transactions, such as sales data, customer information, and inventory levels. Transactional data is often used in operational systems and is used to track the day-to-day operations of a business.
* **Operational data**: This refers to data that is used to support the ongoing operations and decision-making processes of an organization. Operational data can include data from transactional systems, as well as data from other sources such as sensor data, social media data, and log files.
* **Analytical data**: This refers to data that is used for analytical and decision-making purposes, such as data from data warehouses, data marts, and other data stores that are optimized for reporting and analysis.
* **Master data**: This refers to data that is considered the "single source of truth" for an organization, such as customer, product, and vendor information. Master data is typically used as a reference point for other data in the organization.
* **Reference data**: This refers to data that is used to classify, categorize, or provide context for other data, such as codes, lists, and definitions.