1) Difference between data and information.

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

2) How data is useful to us?

1. To analyze, detect patterns, trends and relationships in data sets
2. To predict consumer behavior or to identify business and operational risks
3. Better Productivity
4. Faster results
5. Less time
6. Great Profit

3) What is Big data?

Big data is a combination of structured, semi structured and unstructured data collected by organizations that can be mined for information and used in machine learning projects, predictive modeling and other advanced analytics applications.

Systems that process and store big data have become a common component of data management architectures in organizations, combined with tools that support big data analytics uses.

4) Difference between structured, unstructured and semi structure data.

1. **Structured Data**
    - Well organised data
    - It is less flexible and difficult to scale. It is schema dependent
    - It is based on relational database.
    - Versioning over tuples,row,tables
    - Easy analysis
    - Financial data, bar codes are some of the examples of structured data
2. **Unstructured Data**
    - Not organised at all
    - It is flexible and scalable. It is schema independent
    - It is based on character and binary data
    - Versioning is like as a whole data
    - Difficult analysis
    - Media logs, videos, audios are some of the examples of unstructured data
3. **Semi-structured Data**
    - Partially organised
    - It is more flexible and simpler to scale than structured data but lesser than unstructured data
    - It is based on XML/ RDF
    - Versioning over tuples is possible
    - Difficult analysis compared to structured data but easier when compared to unstructured data
    - Tweets organised by hashtags, folder organised by topics are some of the examples of unstructured data

5) Define qualitative and quantitative data.

Quantitative data are measures of values or counts and are expressed as numbers. Quantitative data are data about numeric variables (e.g. how many; how much; or how often).

Qualitative data are measures of 'types' and may be represented by a name, symbol, or a number code. Qualitative data are data about categorical variables (e.g. what type).

Data collected about a numeric variable will always be quantitative and data collected about a categorical variable will always be qualitative. Therefore, you can identify the type of data, prior to collection, based on whether the variable is numeric or categorical.

6) What are the different types of v's in big data?

Big data is a collection of data from many different sources and is often describe by five characteristics: volume, value, variety, velocity, and veracity.

1. **Volume:**the size and amounts of big data that companies manage and analyze
2. **Value:**the most important "V" from the perspective of the business, the value of big data usually comes from insight discovery and pattern recognition that lead to more effective operations, stronger customer relationships and other clear and quantifiable business benefits
3. **Variety:**the diversity and range of different data types, including unstructured data, semi-structured data and raw data
4. **Velocity:**the speed at which companies receive, store and manage data – e.g., the specific number of social media posts or search queries received within a day, hour or other unit of time
5. **Veracity:**the "truth" or accuracy of data and information assets, which often determines executive-level confidence
6. The additional characteristic of variability can also be considered
7. **Variability:**the changing nature of the data companies seek to capture, manage and analyze – e.g., in sentiment or text analytics, changes in the meaning of key words or phrases

7) What are the popular tools used in big data?

1. Apache Spark
2. Apache Hadoop
3. Apache Flink
4. Google Cloud Platform
5. MongoDB
6. Sisense
7. RapidMiner

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

1. **Qualitative or Categorical Data**

    Qualitative data, also known as the 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. Eg: birthdate, favourite sport, school postcode.

    - **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.

2. **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. 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

        - **Interval data**

            It categorizes and ranks data, and introduces precise and continuous intervals, e.g. temperature measurements in Fahrenheit and Celsius, or the pH scale. Interval data always lack what's known as a 'true zero.' In short, this means that interval data can contain negative values and that a measurement of 'zero' can represent a quantifiable measure of something.

        - **Ratio data**

            It categorizes and ranks data, and uses continuous intervals (like interval data). However, it also has a true zero, which interval data does not. Essentially, this means that when a variable is equal to zero, there is none of this variable. An example of ratio data would be temperature measured on the Kelvin scale, for which there is no measurement below absolute zero (which represents a total absence of heat).