INTRODUCTION TO DATA

1.Differentiate between Data & Information

DATA	INFORMATION
● Data refers to raw facts that have no specific meaning.	● Information refers to processed data that has a purpose and meaning.
● Data is independent of the information.	● Information is dependent on data.
● Data or raw data is not enough to make a decision.	● The information is sufficient to help make a decision in the respective context.
● Understanding is difficult.	● Understanding is easy.
● Eg:Test result of a student	● Average score of a class.

2.How data is useful for us?

Data allows organizations to measure the effectiveness of a given strategy.When strategies are put into place to overcome a challenge, collecting data will allow you to determine how well your solution is performing, and whether or not your approach needs to be tweaked or changed over the long-term.

3.What is big data?

The term Big Data is used in the data definition to describe the data that is in the petabyte range or higher and also it is said to be a huge amount of data.And it is a combination of structured and unstructured data. So big data is just what it sounds like a whole lot of data.

4.Differentiate between structured,semi-structured and unstructured data?

STRUCTURED DATA	SEMI STRUCTURED DATA	UNSTRUCTURED DATA
Clearly defined data models;easy to search.	Loosely coupled data models.	No defined data models;difficult to search.
It is possible to version over tables,rows and tuples.	It is possible to version over graphs or tuples.	It is possible to version the data as a whole.
Types ● CHAR ● TEXT ● LONGTEXT ● INT	Types ● JSON files ● HTML web pages ● CSV files	Types ● Images ● Video,audio files ● Text ● Emails
Highly organized data. Storage - Relational Database	Partially organized data. Storage - Organized by metatags	Unorganized data. Storage - Data lake

5.What are Quantitative Data and Qualitative Data?

Quantitative Data

- Countable or measurable , relating to numbers.
- Tell us how many , how much or how often.
- Gathered by measuring and counting things.
- Eg:Length,size,weight and many more.

Qualitative Data

- Descriptive , relating to words and language.
- Describes certain attributes , and helps us to understand the “why” or “how” behind certain behaviors.
- Gathered through observations and interviews.
- Eg:Color,Gender,Textures and many more.

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

It has 5 V’s

- Volume
- Variety
- Veracity
- Value
- Velocity

7.Name some popular tools used in big data.

- Apache Spark
- Apache Hadoop
- Apache Flink
- Google Cloud Platform
- MongoDB
- Sisense
- RapidMiner

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

- The different types of data are Quantitative(Numerical) and Qualitative(Categorical) .

- Quantitative(Numerical) has two types of data .
  - Discrete Data - A variable which can take only certain values.
    - eg: The number of students in a class.
  - Continuous Data - A variable which can take any value in particular limit.
    - eg: The speed of cars.

- Qualitative(Categorical) have two types of data .
  - Ordinal Data - Can be arranged in order.
    - eg: Ranking.
  - Nominal Data - Cannot be arranged in any particular data.
    - Eg: Eye color.