**Data Analytics**

**Q1 – Differentiate between data and information.**

Ans:

|  |  |
| --- | --- |
| Data | Information |
| * Data is a collection of facts, statistics or items of information * It is unprocessed and unstructured * Might be meaningless on its own. * May be difficult to understand * Comes in forms like numbers, figures and statistics. | * Information is data that is processed, organized   and structured.   * It is processed and structured * Always meaningful * Easy to understand * Comes as word, thoughts and ideas. |

**Q2 – How data is useful for us?**

Ans: Data is useful to us in various ways:

* Improve People's lives
* Make Informed decisions
* Stop Molehills from turning into mountains
* Find Solutions to problems.
* Fraud and Risk Detection.
* Healthcare.
* Internet Search.
* Targeted Advertising.

**Q3 – What is big data?**

Ans: **Big Data** is a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. Big data is also a data but with huge size.

Example of big data includes social media on which ***500+terabytes*** of new data get ingested into the databases of site **Facebook**, every day. This data is mainly generated in terms of photo and video uploads, message exchanges, putting comments etc.

**Q4 – Differentiate between structured, semi-structured and unstructured data.**

|  |  |  |
| --- | --- | --- |
| Structured | Semi-structured | Unstructured |
| * Structured data is well organised data * It is based on relational database management system. * It is easy to secure the data * Difficult to scale * Data mining is easier | * Semi-structured data is organised only up to certain limit * It is based on Xml/Rdf * Data is moderately secure * More scalable * Data mining is less-easier than structured | * Unstructured data is non-organised. * It is based on character and binary data * Data is unsecured. * Highly scalable * Data mining is difficult |

**Q5 – What are quantitative and qualitative data?**

Ans: Quantitative data is data that can be counted or measured in numerical values. The two main types of quantitative data are discrete data and continuous data. Height in feet, age in years, and weight in pounds are examples of quantitative data.

Qualitative data is information that cannot be counted, measured or easily expressed using numbers. It is collected from text, audio and images and shared through data visualization tools, such as word clouds, concept maps, graph databases, timelines and infographics.

**Q6 – What are different V’s in big data?**

Ans: Big data is used to describe the data that is in petabyte range. The five Vs in big data are

1. Variety
2. Volume
3. Value
4. Veracity
5. Velocity

**Q7 – Name some popular tools used in big data?**

Ans: Some of the popular tools used in big data are as follows:

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

Q8 – What are the different types of data?

Ans: There are mainly two types of data

1. Quantitative data
2. Qualitative data

Quantitative data is further divided into two sub types

1. Discrete – Data that cannot be broken
2. Continuous – Data that can be broken

Qualitative data is also further divided into two sub types

1. Nominal – Data that is categorized only
2. Ordinal – Data that is categorized and ranked also