**Introduction to Data**

**Q1.** Differentiate between Data and Information?

Ans:

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Data** | **Information** |
| 1 | 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. |
| 2 | Data doesn't depend on information. | Information depends on data. |
| 3 | Data is unorganised and unrefined facts. | Information comprises processed, organised data presented in a meaningful context. |
| 4 | Raw data alone is insufficient for decision making. | Information is sufficient for decision making. |
| 5 | Bits and Bytes are the measuring unit of data. | Information is measured in meaningful units like time, quantity, etc. |
| 6 | Data are text and numerical values. | Information is refined form of actual data. |
| 7 | Data can be easily structured as the following: 1.Tabular data 2.Graph 3.Data tree. | Information can also be structured as the following: 1. Language 2. Ideas 3. Thoughts. |
| 8 | It is low-level knowledge. | It is the second level of knowledge. |
| 9 | An example of data is a student’s test score | The average score of a class is the information derived from the given data. |

|  |
| --- |
|  |
| **Q2.** How Data is useful for us?  Ans: Data is essentially the plain facts and statistics collected during the operations of a business. They can be used to measure/record a wide range of business activities - both internal and external. While the data itself may not be very informative, it is the basis for all reporting and as such is crucial in business. |

**Q3.** What is Big Data?

Ans: The term ‘big data’ is self-explanatory − a collection of huge data sets that normal computing techniques cannot process. The term not only refers to the data, but also to the various frameworks, tools, and techniques involved.

Data which are very large in size is called Big Data. Normally we work on data of size MB (WordDoc ,Excel) or maximum GB(Movies, Codes) but data in Peta bytes i.e. 10^15 byte size is called Big Data.

It can be defined as data sets whose size or type is beyond the ability of traditional [relational databases](https://www.ibm.com/in-en/analytics/relational-database) to capture, manage and process the data with low latency.

**Q4.** Differentiate between Structured, Semi- Structured and Un-Structured Data?

Ans:

| **Properties** | **Structured data** | **Semi-structured data** | **Unstructured data** |
| --- | --- | --- | --- |
| Technology | It is based on Relational database table. | It is based on XML/RDF (Resource Description Framework). | It is based on character and binary data. |
| Transaction management | Matured transaction and various concurrency techniques | Transaction is adapted from DBMS not matured | No transaction management and no concurrency |
| Version management | Versioning over tuples,row,tables | Versioning over tuples or graph is possible | Versioned as a whole |
| Flexibility | It is schema dependent and less flexible | It is more flexible than structured data but less flexible than unstructured data | It is more flexible and there is absence of schema |
| Scalability | It is very difficult to scale DB schema | It’s scaling is simpler than structured data | It is more scalable. |
| Robustness | Very robust | New technology, not very spread | — |
| Query performance | Structured query allow complex joining | Queries over anonymous nodes are possible | Only textual queries are possible |
| Data Structure | The information and data have a predefined organization. | The contained data and information have organizational properties- but are different from predefined structured data. | There is no predefined organization for the available data and information in the system or database. |

**Q5.** What are Quantitative data and Qualitative data?

Ans:

**Qualitative data:** It is Categorical Variable. 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 info graphics.

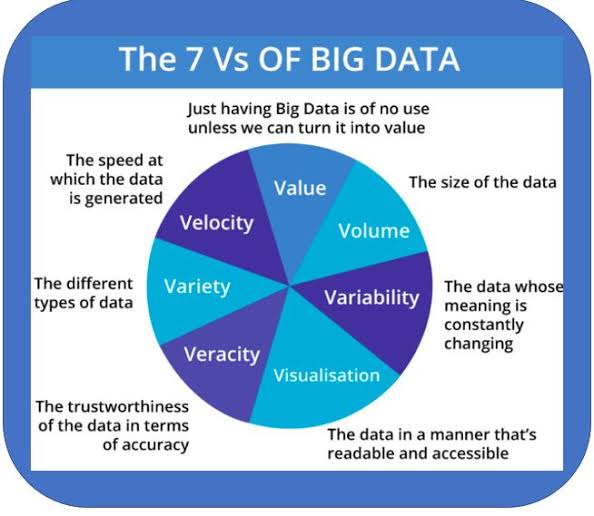
For Eg. : - The texture of the skin, the colour of the eyes, etc.

**Quantitative data:** Quantitative data is Numerical Variable. Quantitative data refers to any information that can be quantified. If it can be counted or measured, and given a numerical value, it’s quantitative data. Quantitative data can tell you “how many,” “how much,” or “how often”.

For Eg. : - If data are collected on annual income.

**Q6.** What are the different V’s in Big Data?

Ans: The **7 V’s** sum it up pretty well- **V**olume, **V**elocity, **V**ariety, **V**ariability, **V**eracity, **V**isualization, and **V**alue.



**Q7.** Name some popular Tools used in Big Data?

Ans:

1. APACHE Hadoop

2. Cassandra

3. Qubole

4. Xplenty

5. Spark

6. Mongo DB

7. Apache Storm

8. SAS

9. Data Pine

10. Rapid Miner

11. Tableau

**Q8.** What are the different types of Data? Explain?

Ans: 