Semi structured database

Semi-structured data is a type of data that is not purely structured, but also not completely unstructured. It contains some level of organization or structure, but does not conform to a rigid schema or data model, and may contain elements that are not easily categorized or classified.

1. Semi-structured data is typically characterized by the use of metadata or tags that provide additional information about the data elements. For example, an XML document might contain tags that indicate the structure of the document, but may also contain additional tags that provide metadata about the content, such as author, date, or keywords.
2. Other examples of semi-structured data include JSON, which is commonly used for exchanging data between web applications, and log files, which often contain a mix of structured and unstructured data.

Semi-structured data is becoming increasingly common as organizations collect and process more data from a variety of sources, including social media, IoT devices, and other unstructured sources. While semi-structured data can be more challenging to work with than strictly structured data, it offers greater flexibility and adaptability, making it a valuable tool for data analysis and management.

Semi-structured data is data that does not conform to a data model but has some structure. It lacks a fixed or rigid schema. It is the data that does not reside in a rational database but that have some organizational properties that make it easier to analyze. With some processes, we can store them in the relational database.

Characteristics of semi-structured Data:

- Data does not conform to a data model but has some structure.
- Data can not be stored in the form of rows and columns as in Databases
- Semi-structured data contains tags and elements (Metadata) which is used to group data and describe how the data is stored
- Similar entities are grouped together and organized in a hierarchy
- Entities in the same group may or may not have the same attributes or properties
- Does not contain sufficient metadata which makes automation and management of data difficult
- Size and type of the same attributes in a group may differ
- Due to lack of a well-defined structure, it can not used by computer programs easily

Sources of semi-structured Data:

- E-mails

- XML and other markup languages
- Binary executables
- TCP/IP packets
- Zipped files
- Integration of data from different sources
- Web pages