**Business** **Intelligence**

**What is structured data?**

Structured data usually resides in relational databases. Fields store length-delineated data phone numbers, Social Security numbers, or ZIP codes. Even text strings of variable length like names are contained in records, making it a simple matter to search.

Data may be human- or machine-generated as long as the data is created within an RDBMS structure. This format is eminently searchable both with human generated queries and via algorithms using type of data and field names, such as alphabetical or numeric, currency or date.

Common relational database applications with structured data include airline reservation systems, inventory control, sales transactions, and ATM activity. SQL enables queries on this type of structured data within relational databases.

Some relational databases do store or point to unstructured data such as customer relationship management applications. The integration can be awkward at best since memo fields do not loan themselves to traditional database queries.

**What is unstructured data?**

Unstructured data is essentially everything else. Unstructured data has internal structure but is not structured via pre-defined data models or schema. It may be textual or non-textual, and human- or machine-generated. It may also be stored within a non-relational database like NoSQL. Ex: text files, emails, media etc.

**What is semi-structured data?**

Semi-strucutred data is information that does not reside in a relational database, but the data has some structured organizational properties that makes it much more easier to analyze that unstructured data. With some additional cleaning and modifications, it is possible to convert semi-structured data to structured data and then store it in a relational database, but usually this does not happen. Most of the time structured data is converted to semi-structured data to reduce the amount of space required to store data. Ex: XML

**Comparison of Structured Vs Semi-Structured Vs Unstructured data**

|  |  |  |  |
| --- | --- | --- | --- |
| **Properties** | **Types of data** | | |
| *Structured* | *Semi-Structured* | *Unstructured* |
| Worth | Provides insight on the context of the data in a clear manner, so is very valuable | As this data lacks certain features to be informative, the value can be debatable as a significant effort is in due to provide real value. | Almost always has less value. |
| Habitat | Occurs mainly in relational database systems. | Occurs with meta-tags in database systems, as XML or JSON/BSON. | Can occur anywhere, as long as the habitat is not structured, occurs as log files. |
| SQL compatibility | Complex SQL commands can be performed. | Basic CRUD SQL commands can be performed. | Only SQL SELECT or similar commands can be performed. |
| Usage | Top level management | Rarely at top-level management using reporting tools but commonly in operational levels. | Only in operational levels. |
| Sources | From data warehouses, data marts etc. | Uncleaned data cubes. | Operational-level databases. |
| Accuracy | Founds a high accuracy for any given situation. | Founds considerable accuracy for suitable situations. | Accuracy cannot be considered with unstructured data. |
| Integrity | If the habitat is secure, then there are no questions about integrity. | Dependable | Dependable |
| Process performance | Easily processable as there is definite structure, so performance is high. | Streamlined processing methods must be used to define structure and process, so performance is debatable. | The larger the unstructured data, the slower it gets. |