

Database Management System – 3 (Structured, Semi-structured and Unstructured data)

Ajay James
Asst. Prof in CSE
Government Engineering College Thrissur

Data

- Known facts that can be recorded and have an implicit meaning
 - descriptions, observations, and numbers used in decision making
- Big Data - deals with large amount of data
 - Structured
 - Semi Structured
 - Unstructured Data

Structured data

- Data whose elements are **addressable** for effective analysis
- Organized into a **formatted** repository that is typically a database
- Concerns all data which can be stored in database SQL in a table with rows and columns
- Have relational keys and can easily be mapped into pre-designed fields
- *Example:* Relational data

Structured data

id	name	age
1	Jim	28
2	Pam	26
3	Michael	42

id	subject	Teacher
1	Languages	John Jones
2	Track	Wally West
3	Swimming	Arthur Curry
4	Computers	Victor Stone

student_id	subject_id	grade
2	1	98
1	2	100
1	4	75
3	3	60
2	4	76
3	2	88

Semi-structured data

- Information that does not reside in a relational database
- Have some organizational properties that make it easier to analyze
- With some process, you can store them in the relation database
- *Example: XML data, JavaScript Object Notation (JSON), HTML files*

Semi-structured data

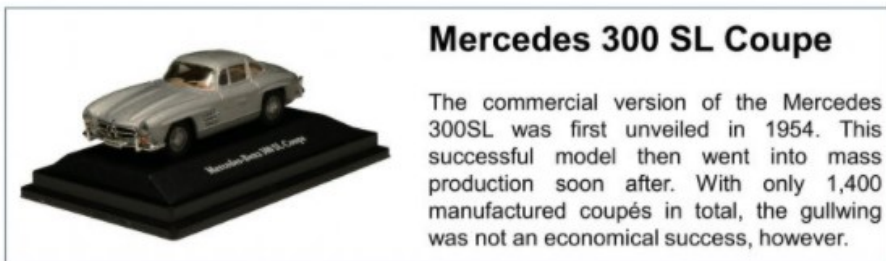
```
## Document 1 ##
{
  "customerID": "103248",
  "name":
  {
    "first": "AAA",
    "last": "BBB"
  },
  "address":
  {
    "street": "Main Street",
    "number": "101",
    "city": "Acity",
    "state": "NY"
  },
  "ccOnFile": "yes",
  "firstOrder": "02/28/2003"
}
```

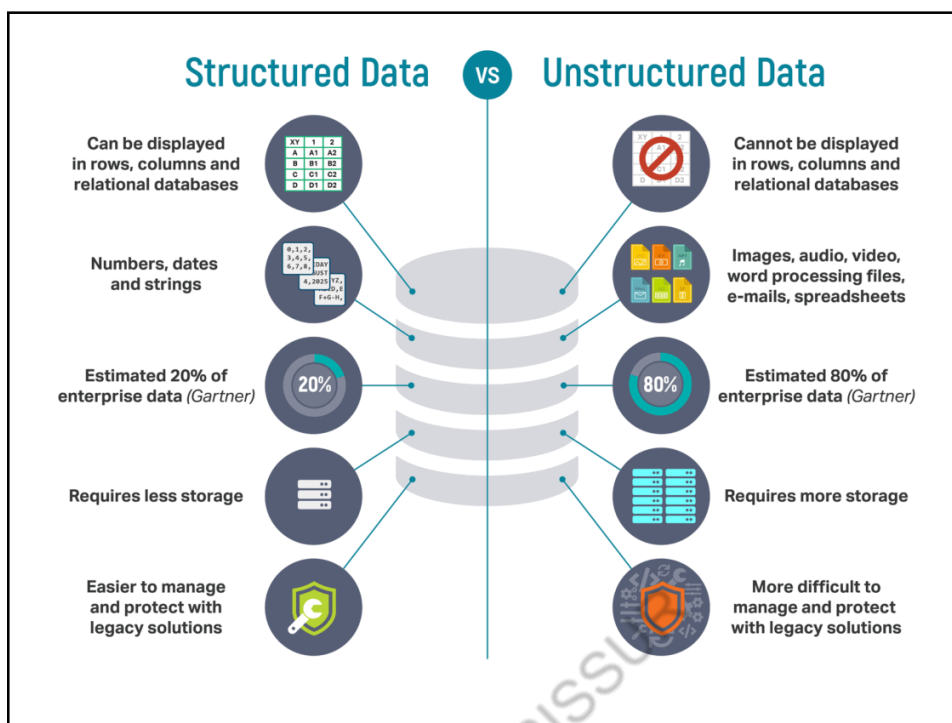
```
<?xml version="1.0"?>
<job>
  - <production>
    <ApprovalType>WebCenter</ApprovalType>
    <Substrate>carton 150 gr</Substrate>
    <SheetSize>220-140</SheetSize>
    <press>SuperFlat2</press>
    <finishing>standard</finishing>
    <urgency>normal</urgency>
  </production>
  - <customer>
    <name>FruitCo</name>
    <number>2712</number>
    <currency>USD</currency>
  </customer>
</job>
```

Unstructured data

- Data which is not organized in a predefined manner
- Does not have a predefined data model
- Not a good fit for a mainstream relational database
- Used by organizations in a variety of business intelligence and analytics applications
- *Example:* Word, PDF, Text, Media logs.

Unstructured data





Properties	Structured data	Semi-structured data	Unstructured data
Technology	Based on Relational database table	Based on XML/RDF(Resource Description Framework).	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 rows, tables	Versioning over tuples or graph is possible	Versioned as a whole
Flexibility	Schema dependent and less flexible	More flexible than structured data but less flexible than unstructured data	More flexible and there is absence of schema
Scalability	Very difficult to scale DB schema	Scaling is simpler than structured data	More scalable
Query performance	Structured query	Queries over anonymous nodes are	Only textual queries

Reference

- <https://www.geeksforgeeks.org/difference-between-structured-semi-structured-and-unstructured-data/>
- <https://k21academy.com/microsoft-azure/dp-900/structured-data-vs-unstructured-data-vs-semi-structured-data/>

Thank you