

# Introduction to Data Exploration

## Data Organization

# Objectives

---




Objective

Explain data, data  
models, and data  
organization

# What is Data?

da·ta

/ˈdɑdə, ˈdādə/ 

*noun*

noun: data

facts and statistics collected together for reference or analysis.

*synonyms:* facts, figures, [statistics](#), details, particulars, specifics; [More](#)

- **COMPUTING**

the quantities, characters, or symbols on which operations are performed by a computer, being stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media.

- **PHILOSOPHY**

things known or assumed as facts, making the basis of reasoning or calculation.

# How is Data Organized?

---

## | What is a database?

- Collection of data, organized in some fashion

## | What is a data model?

- a formalism to describe “constraints” that describe “properties” of data
  - Hierarchical
  - Relational,
  - Object Oriented,
  - Spatial,
  - Fuzzy

# What is a “Data Schema?”

- **a set of constraints that**

- describe the “properties” of data
- describe the structure of the data.
- enable validation and efficient storage of the data
- enable querying and retrieval of data
  - comparison,
  - indexing,
  - query optimization
  - Query processing

| “Schema” is described within the formalism corresponding to the underlying data model

# Levels of Data Organization

---

- | Structured Data/Databases

- | Semi-Structures Data/Databases

- | Unstructured Data/Databases



# Structured Data/Databases

---

## | The data are well-structured and organized

- A “schema” describes this structure
- A Database Management System (DBMS) enforces this structure

## – Advantages

- Data organization is predictable
  - easier to query
  - easier to optimize
  - easier to explore

# Example: Relational Data Models

- Informally, data is organized in tabular form
  - Example: Data about an employee
- Schema for each table consists of attributes
  - Each attribute has a domain
- Functional dependencies
  - Describe the relationships among the attributes in the schema
  - Example: A key uniquely identifies a given tuple in the table

The diagram shows a table with four columns: NAME, SSN, OFFICE, and DESC. The table is annotated with two labels: 'Schema' and 'Attribute'. The 'Schema' label is in a box above the table, with a line pointing to the header row. The 'Attribute' label is in a box to the left of the table, with a line pointing to the first column. The table contains data for two employees: J. Doe and J. Smith, and two placeholder rows with '..'.

NAME	SSN	OFFICE	DESC
..	..	..	..
J. Doe	555-5555	GWC 999	Asst. Prof
J. Smith	333-3333	GWC 989	Prof
..	..	..	..



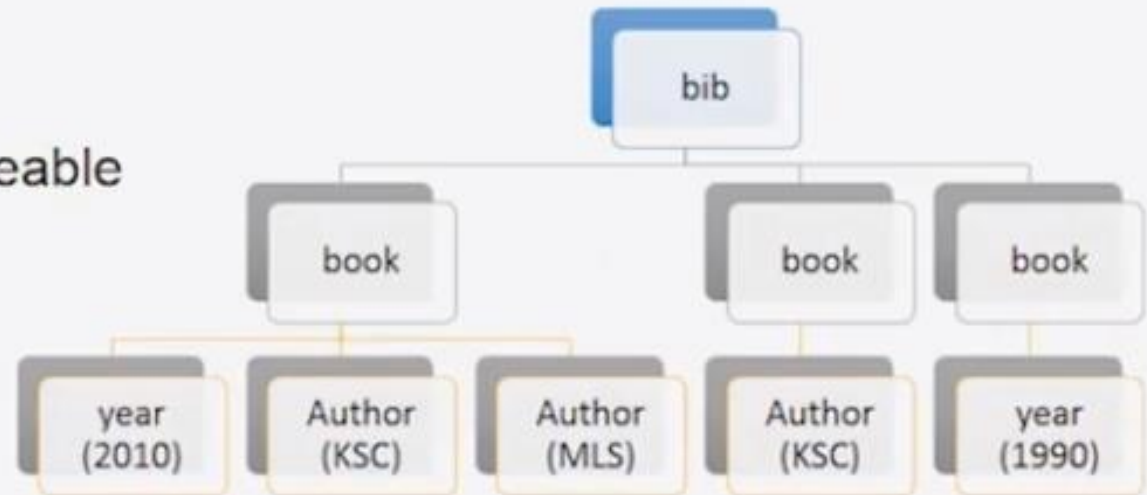
# Semi-Structured Data

| The “constraints” that reflect the structure of the data are flexible

- ability to say “or” in the schema
- missing attributes (null values) or attributes which repeat itself (multivalued attributes)
- Data is self-describing: Each item in the database describes its own schema

## –Advantages

- Data organization is flexible/malleable
  - easier to integrate
  - easier to exchange



# Semi-Structured Data

| The “constraints” that reflect the structure of the data are flexible

- ability to say “or” in the schema
- missing attributes (null values) or attributes which repeat itself (multivalued attributes)
- Data is self-describing: Each item in the database describes its own schema

## –Advantages

- Data organization is flexible/malleable
  - easier to integrate
  - easier to exchange

