Introduction to Data Exploration

Data Organization

Ira A. Fulton Schools of Engineering
Arizona State University

Objectives

Objective

Explain data warehouses and exploratory querying

What is Data?

specifics
characters
stored
quantities
Philosophy
calculation
electrical
reference
particulars
signals
assumed
computer
making
mechanical
collected
things
basis
analysis
form
optical
together
facts
statistics
magnetic
reasoning
symbols
figures
recording
operations
media
data
details
known
recorded
Computing
performed

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

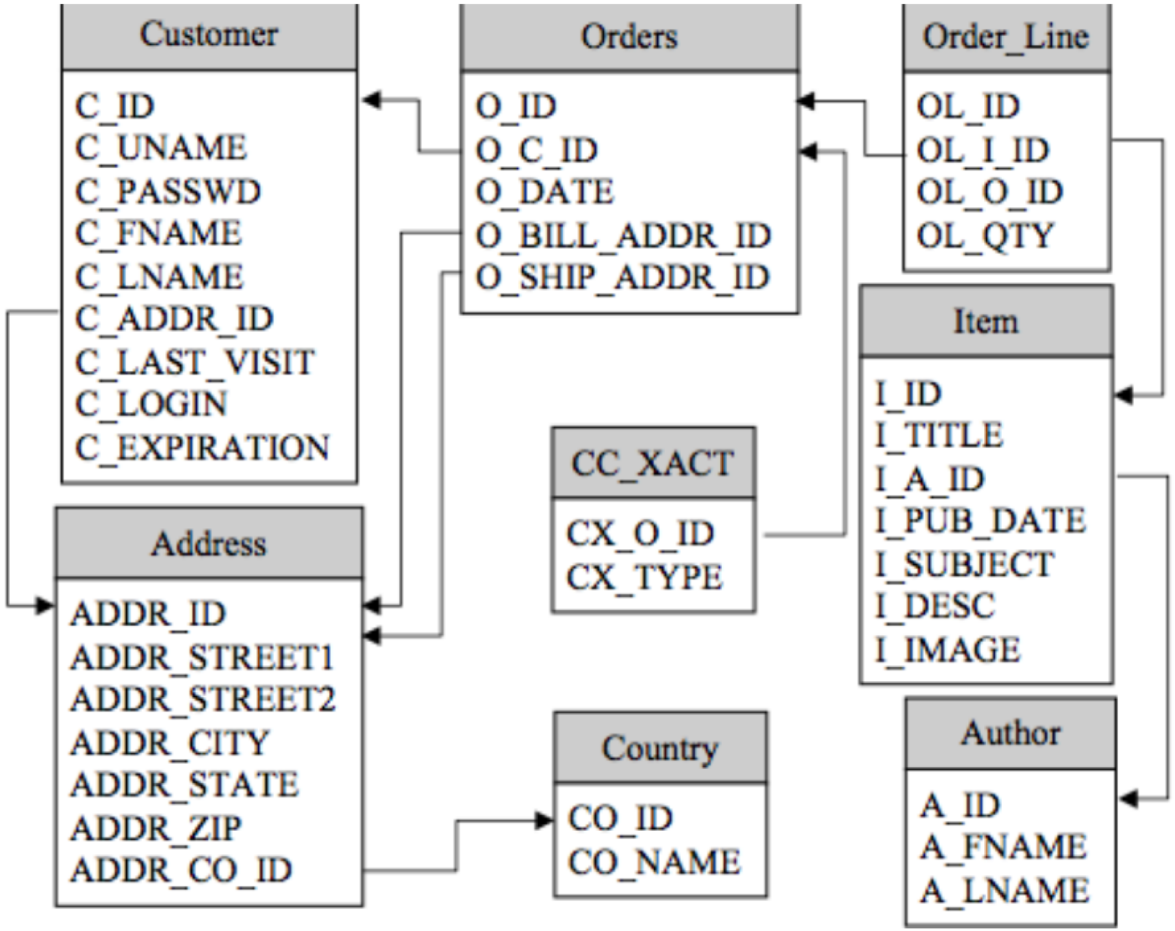
What is a “Data Schema?”

- “Schema” is described within the formalism corresponding to the underlying data model
- a formalism to describe “constraints” that describe “properties” of data
  - Hierarchical
  - Relational,
  - Object Oriented,
  - Spatial,
  - Fuzzy

Levels of Data Organization

Structured Data/Database

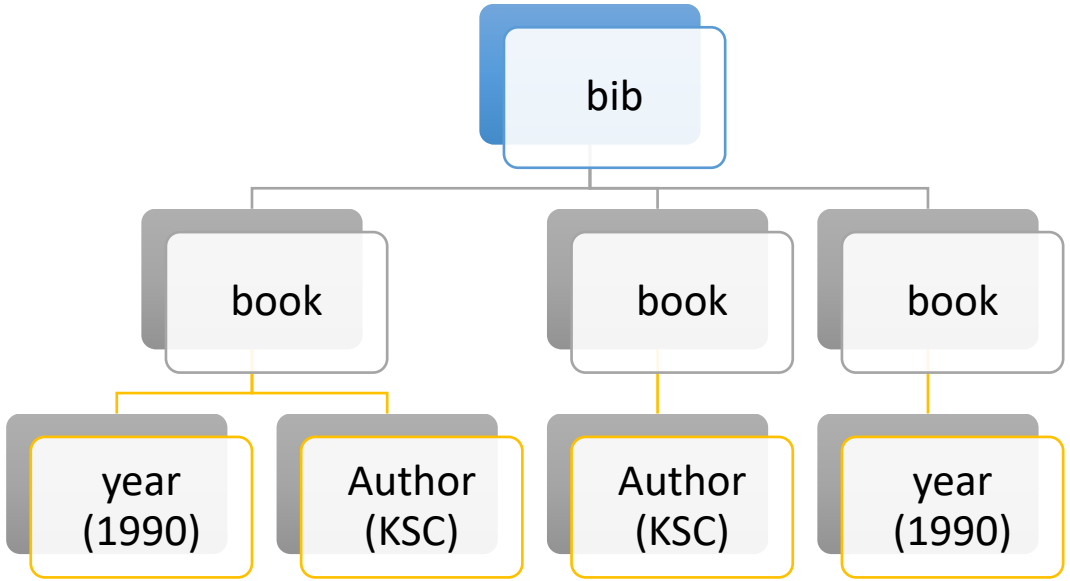
- The data is well-structured and organized
- A schema describes/enforces this structure



Semi-Structured Data

Semi-structured data/database

- The data is organized, but the organization is flexible
- The “constraints” that reflect the structure of the data are malleable
- Data is self-describing: Each item in the database describes its own schema



Levels of Data Organization

Un-structured data/data collection

- There is no pre-determined schema that describes the structure of the data
- Structure exists but is unknown

Images

Videos

Social Network

Books

Sensor Reading

Common Data Representations

- Relational Data
- Vector Space (spatial or high-dimensional) data
- Strings, sequences, and time series data
- Trees and graphs
- Fuzzy and probabilistic data