Foundations of Data Management

Data Modeling

Bryan Shepherd, PhD
Summer 2019
Merrimack College

Key Themes

- What is Data Modeling?
- How do we do data modeling?
- What is a good data model?

Data Modeling - What is it?

Data modeling (the process) - "a method for determining what data and relationships should be stored in a database." (Watson, pg. 143)

Data model (the artifact) - 1. The database structure resulting from the data modeling process. 2. "a way of communicating database design" (ibid.)

Data Modeling - How do we do it?

Preliminary steps:

- Define the business goals
- Define the technology to be used
- Determine the facts/data required by the business goals
- Define the key relationships between the facts/data

How much you are involved in these preliminary steps is determined by your position, the size of the company, and the company's "life stage" (e.g., startup, established, etc.), but BAs and DSs are key stakeholders in the data

Data Modeling - How do we do it?

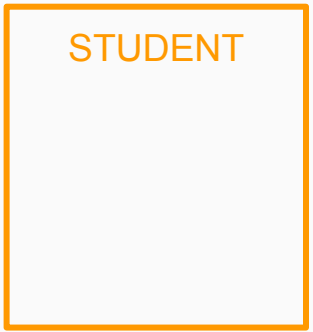
Form main components of a data model:

- Entity
- Attribute
- Relationship
- Identifier

Data Modeling - How do we do it?

Entity:

- Something we want to describe (a noun)
- Often represented by individual tables in the database



Data Modeling - How do we do it?

Attribute:

- Features or facts about entities (the things we want to describe)
- Usually represented by columns in the entity tables

Data Modeling - How do we do it?

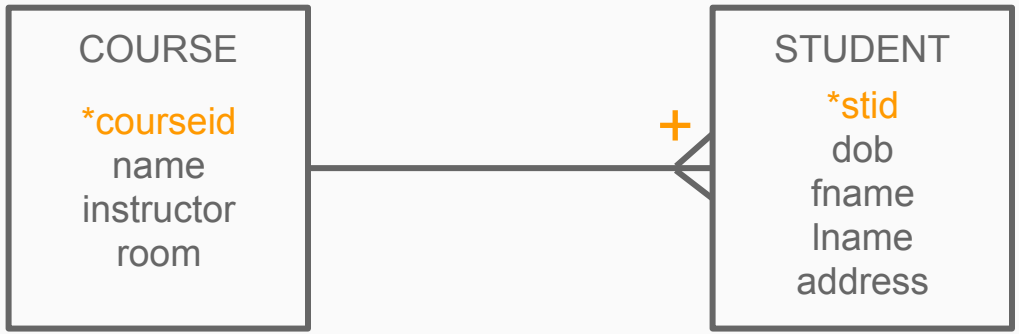
Relationship:

- Connects entities/tables
- Without them we don't need a *relational* database

Data Modeling - How do we do it?

Identifiers:

- Enable us to uniquely identify instances
- Can work within or across tables



Data Modeling - What is a good data model?

- Well-formed
  - Construction rules obeyed
  - No ambiguity
  - Entities named
  - Every entity has identifier
  - All relationships represented with correct notation
  - Relationships labeled where relevant
  - All attributes listed
  - Attribute names meaningful and unique

(Watson pg. 147)

Data Modeling - What is a good data model?

- High-fidelity
  - Construction rules obeyed
  - No ambiguity
  - Entities named
  - Every entity has identifier
  - All relationships represented with correct notation
  - Relationships labeled where relevant
  - All attributes listed
  - Attribute names meaningful and unique
- All applicable entities, attributes, and relationships are defined