Lecture 2: First Steps in DB Design - ER Model

Mohamed Zahran (aka Z)
mzahran@cs.nyu.edu
http://www.mzahran.com
Semantic Data Model

ER Model
Entity-Relationship (ER) Model

- Part of the conceptual design
- Semantic data model
- Can be Pictorially represented.
- A high-level description of the data to be stored in the DB
- Facilitates discussion among all the people involved in the design process.
- ER diagrams then relational model
• Is an object in the real world
• Similar entities are grouped into **entity set**
• Entity sets need not be disjoint

**Set of Attributes**

- **Entity**: An object in the real world
- **Domain of Possible Values**
- **Key**: Minimal set of attributes whose value uniquely identify an entity in the set

**ER Diagram**

**Relationships**: An association among two or more entities
Employees

Attributes

- ssn
- name
- lot

Key

Entity Set

Employees
• Relationship set
• Can have its own **descriptive attributes**
• Several relationship sets might involved the same sets
• A relationship must be uniquely identified by its participating entities (without the need of its description entity).
• Relationship set
• Can have its own descriptive attributes
• Several relationship sets might involved the same sets
• A relationship must be uniquely identified by its participating entities (without the need of its description entity).
Relationship Types

1-to-1
1-to Many
Many-to-1
Many-to-Many
Although an employee can manage several departments, each department is managed by at most one employee.

This arrow means that given a "Departments" entity, we can **uniquely** determine the "Manages" relationship in which it appears.
ER Feature: Participation
Constraints

Total participation
As opposed to partial participation
Every employee works in at least one department.

Every department has at least one employee.

Every department must have a manager.
ER Feature: Weak-Entity Set

- A weak entity can be identified uniquely only by considering the primary key of another (owner) entity.
  - Owner entity set and weak entity set must participate in a one-to-many relationship set (one owner, many weak entities).
  - Weak entity set must have total participation in this identifying relationship set.
ER Feature: Weak-Entity Set

- **Employees**
  - ssn
  - name
  - lot

- **Policy**
  - cost

- **Dependents**
  - pname
  - age

**Identifying owner set**

**Identifying Relationship**

**Weak entity set**

The thick line means total participation

Partial key
ER Feature: Class Hierarchies

Specialization

Generalization
ER Feature: Class Hierarchies

• Attributes are inherited
• If we declare A ISA B, every A entity is also considered to be a B entity
• Reasons for using ISA:
  – To add descriptive attributes specific to a subclass.
  – To identify entities that participate in a relationship.
ER Feature: Class Hierarchies

• **Overlap constraints**: Can Joe be an Hourly_Emps as well as a Contract_Emps entity? *(Allowed/disallowed)* (default is no overlap)

• **Covering constraints**: Does every Employees entity also have to be an Hourly_Emps or a Contract_Emps entity? *(Yes/no)* (default is no covering)
ER Feature: Aggregation

• Used when we have to model a relationship involving (entity sets and) a relationship set.
  – i.e. allows us to treat a relationship set as an entity set for purposes of participation in (other) relationships

• Used when we need to express a relationship among relationships
How to measure the quality of an ER model?
Some Rules of Thumb

1. Avoid redundancy whenever possible.
2. Limit the use of weak entity sets.
3. Don’t use an entity set when an attribute will do.
Good or Bad?

name

cOMPANY

Company Addr

Software
Good or Bad?

Car

name

ManfBy

name

addr

Manfs

manf
Good or Bad?

- Car
- ManfBy
- Manfs

name
name
addr
Design Choices

• Should a concept be modeled as an entity or an attribute?
• Should a concept be modeled as an entity or a relationship?
• Identifying relationships: Binary or ternary? Aggregation?
Entity Vs Attribute

• Should *address* be an attribute of Employees or an entity (connected to Employees by a relationship)?

• Depends upon the use we want to make of address information, and the semantics of the data:
  • If we have several addresses per employee, *address* must be an entity.
  • If the structure (city, street, etc.) is important, e.g., we want to retrieve employees in a given city, *address* must be modeled as an entity (since attribute values are atomic).
Entity Vs Attribute: A Rule of Thumb

An entity set should satisfy at least one of the following conditions:

- It is more than the name of something; it has at least one nonkey attribute.
- It is the “many” in a many-one or many-many relationship.

Is this ER good or bad?
• First ER diagram OK if a manager gets a separate discretionary budget for each dept.

• What if a manager gets a discretionary budget that covers all managed depts?
  - **Redundancy**: `dbudget` stored for each dept managed by manager.
  - **Misleading**: Suggests `dbudget` associated with department-mgr combination.
Binary vs. Ternary Relationships

- If each policy is owned by just 1 employee, and each dependent is tied to the covering policy, first diagram is inaccurate.

- What are the additional constraints in the 2nd diagram?
What Will You do?

"I would like my customers to be able to browse my catalog of books and place orders over the Internet. Currently, I take orders over the phone. I have mostly corporate customers who call me and give me the ISBN number of a book and a quantity; they often pay by credit card. I then prepare a shipment that contains the books they ordered. If I don't have enough copies in stock, I order additional copies and delay the shipment until the new copies arrive; I want to ship a customer's entire order together. My catalog includes all the books I sell. For each book, the catalog contains its ISBN number, title, author, purchase price, sales price, and the year the book was published. Most of my customers are regulars, and I have records with their names and addresses. New customers have to call me first and establish an account before they can use my website. On my new website, customers should first identify themselves by their unique customer identification number. Then they should be able to browse my catalog and to place orders online."
Conclusions

- ER model is a widely used semantic data model. It follows requirements analysis.
- ER model is a good starting point for conceptual schema.
  - Constructs are expressive, close to the way people think about their applications.
  - Note: There are many variations on ER model.
- Reading: Chp2 till and including 2.5