Data Modeling Concepts

Data Models

✔ Model: “set of concepts to describe the structure of a DB”
✔ High-level: close to the way users perceive the data
✔ Low-level: how the data is stored in the computer
✔ Analysis and design typically start at the top of work down

Many intermediate levels are possible

High-level modeling concepts

✔ DB Schema
  – vs. DB instance
✔ 3-schema architecture
  – External (view) level: what users are allowed to see
  – Conceptual level: abstract level, whole DB
    • our focus in this course
  – Internal level: storage structure

Independence between the levels is good

Implementing the schema

✔ Developed by system analysts
✔ Maintained by DBA
✔ DDL vs DML
  ✔ Data dictionary/system catalog
    – Active vs passive

Entity-Relationship (ER) Model

✔ Conceptual level
✔ Graphical conventions
✔ Easily turned into relational model
✔ Has limitations

Overview: identify and name the pieces of data needed in the application and identify key relationships and constraints

Entities and Attributes

✔ Entities: represent things in the real world
  – e.g. student, book, building, course, product
  – usually described by nouns in English
  – often corresponds to a struct in a programming lang.
✔ Attributes: represent qualities of entities
  – e.g. price, color, status, age, room number
  – attribute values are often adjectives in English
  – ALWAYS attached to entities
  – often corresponds to a simple type in a programming lang.
    • but compound and composite attributes exist too.
Types and instances

✔ Entity types vs entity instances
  – Person / FDR, Chelsea, Elizabeth R
  – Book / GWTW, Fire Your Boss, DB for Dummies
✔ Attributes vs values
  – Color / red, chartreuse, {red, white & blue}
  – Income / $4,000,000, $13,200
  – Address / 1600 Pennsylvania Avenue, 415 4th St. SE

Relationships

✔ Connections between entities
  – never between attributes
✔ “John owns a copy of GWTW”
  – John and GWTW are entities
  – The two entities are related (“owns”)
✔ “John owns a first edition of GWTW”
  – ‘first edition’ is an attribute of GWTW
✔ “John was born in 1932”
  – 1932 is an attribute of John (no relationship)

Keys

✔ Attributes with unique values
  – Students: each has a unique SSN
  – Buildings: each has a unique address
  – May be more than one possible key
✔ Sometimes there is no natural key
  – List of plants encountered on a nature walk
  – List of first names of employee’s dependents
  – Called “weak” entities if no proper key
  – Needs a relationship to a “strong” entity

Relationship Cardinalities

✔ 1 to 1 (1:1)
  – husband to (current) wife
  – company to its president
✔ Many to 1 (N:1) or 1 to many (1:N)
  – person to her children
  – company to its employees
✔ Many to many (N:M)
  – club member to clubs
  – student to classes

Relationship Attributes

✔ Sometimes attributes seem to attach to relationships rather than the entities involved
✔ “Faust sold his soul for $10”
  – “sold” relationship
  – $10 is a property of the relationship rather than of either Faust or his soul

E/R Notation

✔ Once the entities, attributes, relations, keys, cardinalities, etc. have been identified, the E/R drawing is straightforward!
✔ rectangles for entities
✔ diamonds for relationships
✔ ovals for attributes
✔ appropriate connecting lines and labels
The Real Problem

✔ Drawing the diagram isn’t so hard.
   – Hint: use a big sheet of paper and write in pencil
✔ What’s hard is understanding the problem well enough to know what entities, attributes, and relationships are needed to support the desired application.
✔ This may require an extended period of systems analysis and several iterations

E/R Model

Strengths and weakness

✔ Easy to understand
✔ Maps well to relational model
✔ “Semantically impoverished”
   – so is the relational model!
✔ E/R has been extended (Chapter 21)
✔ Other models exist (Chapter 21)