Relational Databases

IST400/600

Jian Qin

Database

- A collection of data?
  - Everything you collected for your group project?
- A computer system?
  - File?
  - Spreadsheet?
  - Information system?

Date's criteria:
- Integration
- Sharing
- Persistence
- Entities and relationships
- Properties

Relational database

- Data are
  - Organized as tables
  - Retrievable through queries

Access example (access1.mdb)

What's a table?

- Table
  - Set of rows
- Row
  - No more than one value per column
  - Row (column, …) = tuple (element, …)
    - =record (field, …)
- Column
  - Unique (within table) name
  - Single domain: all values have same types and constraints

Access example (access1.mdb)

Basic Terminology

- Data: Raw facts that have little meaning unless they have been organized in some logical manner.
  - Letters: a, b, c, A, B, C, …
  - Numbers: 1, 2, 3, …
  - Symbols: ?, <, >, *, +, …
- Field (attribute, column): A character or group of characters that has a specific meaning.
  - Telephone number, date of birth, city, annual income, product price, …
- Record (tuple, row): A logically connected set of one or more fields that describes a person, place, or thing.
- File (relation, table) : A collection of related records.

What's a query?

- A question about a table, expressed as specific constraints on column values
  - E.g. what datasets have the abbreviation "acsoe"?
  - E.g. "Find all deployments that started before 3/25/1998"
- A query against one table (the base table) produces another table (the answer table) whose rows and columns "answer the question" (satisfy the query constraints)

Access example (access1.mdb)

Problems with single-table database

(or, why not just keep everything in Excel?)

- Redundant data
  - Attributes applied to a group of rows must be repeated in each of the rows
- Typographical errors
  - Redundant values increase probability of transcription error
- Updating data
  - If redundant values, then must change all of them
- Modifying data
  - Insertion anomaly
    - Can't insert partial row
  - Deletion anomaly
    - Must delete whole row

Multiple table databases

- Eliminate redundancy
  - By factoring single table into multiple tables
    - Each table = single kind of thing
    - Each row = single thing
- Preserve relationships by references between tables
- Replace redundant values with *reference* to unique values
  - (*key*)

Keys

- A key uniquely identifies, and can therefore be used as a reference to, a single row
  - *Primary key* (PK): column(s) whose values uniquely identify a row
    - Data values that are naturally unique
    - Arbitrary value
  - *Foreign* key (FK): reference to another row's primary key
  - **Foreign keys are how databases maintain explicit relationships between rows, within or between tables.**

Relationships

- Relationships in Conceptual Models
  - One-to-one (1:1)
    - E.g. person A has 1 spouse B
    - are A and B really different aspects of the same thing?
      - yes: merge into single table
      - no: like one-to-many
  - One-to-many (1:M)
    - E.g. person A has several children B
    - Most common relationship
      - table B has foreign keys into table A
  - Many-to-many (M:N)
    - E.g.
      - student A takes several classes B
      - class B has several students A
    - Needs a third table
      - table C has foreign keys into tables A and B

Database design

- Develop a logical data model, and then translate it into a physical data model.
- *Logical*: database-independent
  - entities and attributes
  - relationships
- *Physical*: database-dependent
  - tables and columns
  - foreign keys

Entity-Relationship (ER) Models

- Standard way to represent logical data models
- *Entities*: kinds of things
- *Attributes*: facts about things
  - "has-a"
- *Relationships*: connections between kinds of things
  - "is-a", "is-part-of", "is-a-...-of"
- Graphical representation
- Entity
  - node (e.g. box)
- Attribute
  - label (e.g. text in box)
- Relationship
  - arc (e.g. line)
  - cardinality (e.g. "crow's foot")

From an ER Model to a Relational Model

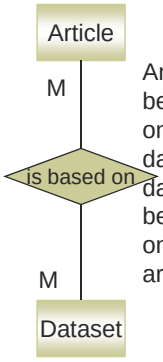
- Create 1 table per entity.
- If parent ("1:") entity then
  - create single-column primary key.
- Else if child (":n") entity then
  - Create foreign keys for primary key of each parent.

Access example (access1.mdb)

Entity-Relationship Diagram (ERD)

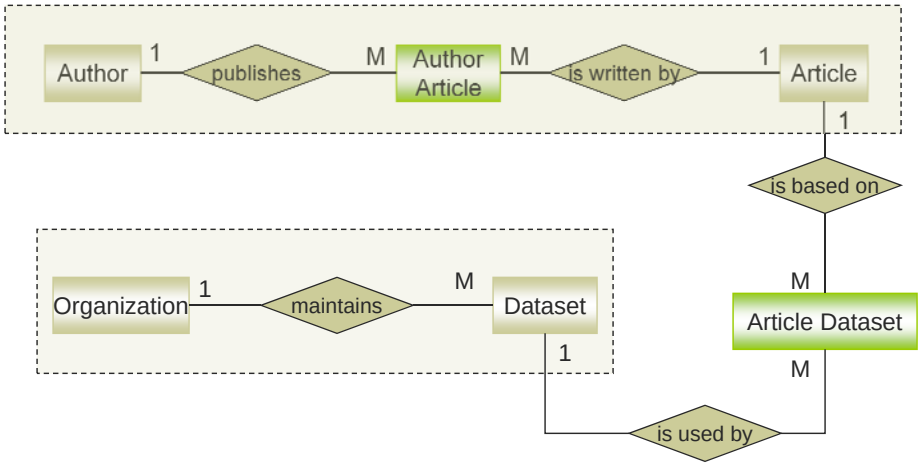
An author may publish one or more articles;
An articles may contain one or more authors.

An organization maintains one or more datasets.



An article may be based on one or more datasets; A dataset may be used in one or more articles.

Entity-Relationship Diagram (ERD)



Case analysis

- The VEMAP Data Portal http://www.cgd.ucar.edu/vemap/datasets.html
- Study the website and think about:
  - How would you describe the portal as a whole?
  - If the datasets information were to become database-driven, what entities (tables) would you create?
  - How are the tables related to one another?
  - What columns would you have for each table?