Introduction to Data Management
CSE 414

Lecture 2: Data Models & SQL
(Ch. 2.1-2.3)

Announcements
• Office Hours are listed on the calendar
  – one every Monday, Tuesday, Wednesday, and Friday

Data Models
• language / notation for talking about data
• models we will use:
  – relational: data is a collection of tables
  – semi-structured: data is a tree
• other models:
  – key-value pairs: used by NoSQL systems
  – graph data model: used by RDF (semi-structured can also do)
  – object oriented: often layered on relational, J2EE

Relational Model
• Data is a collection of relations / tables:
  rows / tuples / records
  columns / attributes / fields
  Name Country Employees For_Profit
  GizmoWorks USA 20000 True
  Canon Japan 50000 True
  Hitachi Japan 30000 True
  HappyCam Canada 500 False

Relational Schema
• Each column has a “domain” (or type)
  – SQL has Java-like types for numbers, strings, etc.
  – domain is a constraint on the data allowed in the table
• Names and types part of the “schema” of the table:
  Company(Name: string, Country: string,
           Employees: int, For_Profit: boolean)

Keys
• Key = subset of columns that uniquely identifies tuple
• Another constraint on the table
  – no two tuples can have the same values for those columns
• Examples:
  – Movie(title, year, length, genre): key is (title, year)
  – what is a good key for Company?
• Part of the schema (book notation is underline):
  Company(\texttt{Name}, string, \texttt{Country}, string, \texttt{Employees}: int, \texttt{For_Profit}: boolean)
Keys (cont.)

• Can have multiple keys for a table

• Only one of those keys may be “primary”
  – DBMS often makes searches by primary key fastest
  – other keys are called “secondary”

• “Foreign key” is a column (or columns) whose value
  is a key of another table
  – i.e., a reference to another row in another table

SQL (“sequel”) – Cont.

• Provides standard types. Examples:
  – BOOL: BOOLEAN
  – DATE, TIME, TIMESTAMP
    • DATE: Stores year, month, and day values
    • TIME: Stores hour, minute, and second values
    • TIMESTAMP: Stores year, month, day, hour, minute, and second values
• Additional types differ by vendor:
  – SQLite: http://www.sqlite.org/datatype3.html

CREATE TABLE Company(  
  name VARCHAR(20) PRIMARY KEY,
  country VARCHAR(20),
  employees INT,
  for_profit CHAR(1));

SQL (“sequel”)

• Standard query language for relational data
  – used for databases in many different contexts
  – inspires query languages for non-relational (e.g. SQL++)
• Everything not in quotes (‘…’) is case insensitive
• Provides standard types. Examples:
  – numbers: INT, FLOAT, DECIMAL(p,s)
    • DECIMAL(p,s): Exact numerical, precision p, scale s. Example: decimal(5,2) is a number that has 3 digits before the decimal and 2 digits after the decimal
  – strings: CHAR(n), VARCHAR(n)
    • CHAR(n): Fixed-length n
    • VARCHAR(n): Variable length. Maximum length n

SQL statements

• create table …
• drop table …
• alter table … add/remove …
• insert into … values …
• delete from … where …
• update … set … where …

DROP TABLE Company;
ALTER TABLE Company
ADD CEO VARCHAR(20);

INSERT INTO Company VALUES
('GizmoWorks', 'USA', 20000, 'y');

DELETE FROM Company
where name = 'GizmoWorks';

UPDATE Company
SET employees = employees + 120
where name = 'GizmoWorks';

Demo on Sqlite
• E.g., type sqlite3 in Cygwin
• .exit - exit from sqlite3