#### Introduction to Data Management CSE 414

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

CSE 414 - Spring 2017

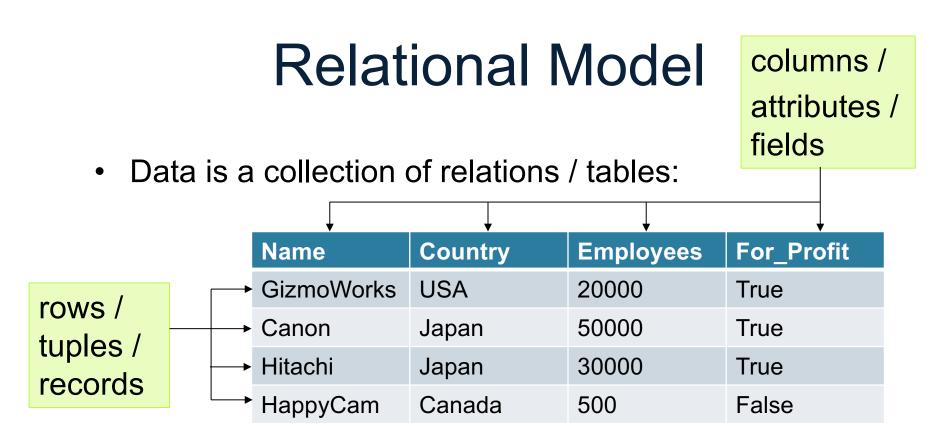
1

#### Announcements

- Office Hours are listed on the calendar – one every day, M-F
- If you have a laptop, bring to section Th
  - also look at HW1 for installing sqlite
  - can go through the examples yourself

### 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



- mathematically, relation is a set of tuples
  - each tuple appears 0 or 1 times in the table
  - order of the rows is unspecified

## **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)

- Particular data is an "instance" of that relation
  - data changes over time
  - DBMS usually just stores the current instance

# 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(<u>Name</u>: string, <u>Country</u>: string, Employees: int, 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")

- 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)
  - strings: CHAR(n), VARCHAR(n)
  - BOOLEAN
  - DATE, TIME, TIMESTAMP
- Additional types differ by vendor:
  - SQLite: http://www.sqlite.org/datatype3.html

### SQL statements

- create table ...
- drop table ...
- alter table ... add/remove ...
- insert into ... values ...
- delete from ... where ...
- update ... set ... where ...