

# Introduction to Databases CSE 414

## Lecture 2: Data Models

CSE 414 - Spring 2018

1

## Announcements

- HW1 and WQ1 released
  - Both due next Tuesday
- Office hours start this week
- Sections tomorrow
- Make sure you sign up on piazza
- Please ask questions!
  - Both online and offline

2

## Staff

- Instructor: Alvin Cheung
  - Office hour on Wednesdays, 1-2pm



From ACM Spring BBQ 15

CSE 414 - Spring 2018

3

## Using Electronics in Class

In the lectures:

- Opened laptops may disturb neighbors
- Please sit in the back if you take notes on laptop; pads / surfaces are OK
- Please don't check your email / youtube / fb

In the sections:

- Always bring your laptop (starting Thursday)

CSE 414 - Spring 2018

4

## Class Overview

- Unit 1: Intro
- Unit 2: Relational Data Models and Query Languages
  - Data models, SQL, Relational Algebra, Datalog
- Unit 3: Non-relational data
- Unit 4: RDMBS internals and query optimization
- Unit 5: Parallel query processing
- Unit 6: DBMS usability, conceptual design
- Unit 7: Transactions

CSE 414 - Spring 2018

5

## Review

- What is a database?
  - A collection of files storing related data
- What is a DBMS?
  - An application program that allows us to manage efficiently the collection of data files

CSE 414 - Spring 2018

6

## Data Models

- Recall our example: want to design a database of books:
  - author, title, publisher, pub date, price, etc
  - How should we describe this data?
- Data model** = mathematical formalism (or conceptual way) for describing the data

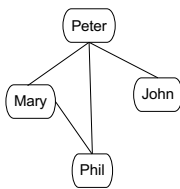
## Data Models

- Relational**
  - Data represented as relations
- Semi-structured (JSON)
  - Data represented as trees
- Key-value pairs
  - Used by NoSQL systems
- Graph
- Object-oriented

Unit 2

Unit 3

## Example: storing FB friends



Or

Person1	Person2	is_friend
Peter	John	1
John	Mary	0
Mary	Phil	1
Phil	Peter	1
...	...	...

As a graph

As a relation

We will learn the tradeoffs of different data models later this quarter

## 3 Elements of Data Models

- Instance
  - The actual data
- Schema
  - Describe what data is being stored
- Query language
  - How to retrieve and manipulate data

## Turing Awards in Data Management



Charles Bachman, 1973  
*IDS and CODASYL*



**Ted Codd, 1981**  
*Relational model*



Jim Gray, 1998  
*Transaction processing*



Michael Stonebraker, 2014  
*INGRES and Postgres*

## Relational Model

- Data is a collection of relations / tables:

	cname	country	no_employees	for_profit
rows / tuples / records	GizmoWorks	USA	20000	True
	Canon	Japan	50000	True
	Hitachi	Japan	30000	True
	HappyCam	Canada	500	False

columns / attributes / fields

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

## The Relational Data Model

- Degree (arity) of a relation = #attributes
- Each attribute has a type.
  - Examples types:
    - Strings: CHAR(20), VARCHAR(50), TEXT
    - Numbers: INT, SMALLINT, FLOAT
    - MONEY, DATETIME, ...
    - Few more that are vendor specific
  - Statically and strictly enforced

CSE 414 - Spring 2018

13

## Keys

- Key = one (or multiple) attributes that uniquely identify a record

CSE 414 - Spring 2018

14

## Keys

- Key = one (or multiple) attributes that uniquely identify a record

Key

cname	country	no_employees	for_profit
GizmoWorks	USA	20000	True
Canon	Japan	50000	True
Hitachi	Japan	30000	True
HappyCam	Canada	500	False

CSE 414 - Spring 2018

15

## Keys

- Key = one (or multiple) attributes that uniquely identify a record

Key

Not a key

cname	country	no_employees	for_profit
GizmoWorks	USA	20000	True
Canon	Japan	50000	True
Hitachi	Japan	30000	True
HappyCam	Canada	500	False

CSE 414 - Spring 2018

16

## Keys

- Key = one (or multiple) attributes that uniquely identify a record

Key

Not a key

Is this a key?

cname	country	no_employees	for_profit
GizmoWorks	USA	20000	True
Canon	Japan	50000	True
Hitachi	Japan	30000	True
HappyCam	Canada	500	False

CSE 414 - Spring 2018

17

## Keys

- Key = one (or multiple) attributes that uniquely identify a record

Key

Not a key

Is this a key?

No: future updates to the database may create duplicate no\_employees

cname	country	no_employees	for_profit
GizmoWorks	USA	20000	True
Canon	Japan	50000	True
Hitachi	Japan	30000	True
HappyCam	Canada	500	False

CSE 414 - Spring 2018

18

## Multi-attribute Key

Key = fName, lName  
(what does this mean?)

fName	lName	Income	Department
Alice	Smith	20000	Testing
Alice	Thompson	50000	Testing
Bob	Thompson	30000	SW
Carol	Smith	50000	Testing

CSE 414 - Spring 2018

19

## Multiple Keys

SSN	fName	lName	Income	Department
111-22-3333	Alice	Smith	20000	Testing
222-33-4444	Alice	Thompson	50000	Testing
333-44-5555	Bob	Thompson	30000	SW
444-55-6666	Carol	Smith	50000	Testing

We can choose one key and designate it as primary key.  
E.g.: primary key = SSN

CSE 414 - Spring 2018

20

## Foreign Key

Company(cname, country, no\_employees, for\_profit)  
Country(name, population)

Company

cname	country	no_employees	for_profit
Canon	Japan	50000	Y
Hitachi	Japan	30000	Y

Country

name	population
USA	320M
Japan	127M

CSE 414 - Spring 2018

21

## Keys: Summary

- Key = columns that uniquely identify tuple
  - Usually we underline
  - A relation can have many keys, but only one can be chosen as *primary key*
- Foreign key:
  - Attribute(s) whose value is a key of a record in some other relation
  - Foreign keys are sometimes called *semantic pointer*

CSE 414 - Spring 2018

22

## Query Language

- SQL
  - Structured Query Language
  - Developed by IBM in the 70s
  - Most widely used language to query relational data
- Other relational query languages
  - Datalog, relational algebra

CSE 414 - Spring 2018

23

## Our First DBMS

- SQL Lite
- Will switch to SQL Server later in the quarter

CSE 414 - Spring 2018

24

## Demo 1

CSE 414 - Spring 2018

25

## Discussion

- Tables are NOT ordered
  - they are sets or multisets (bags)
- Tables are FLAT
  - No nested attributes
- Tables DO NOT prescribe how they are implemented / stored on disk
  - This is called **physical data independence**

CSE 414 - Spring 2018

26