Introduction to CMPSC 430 Database Design

Spring 2016 Dr. Linda Null

This Course...

- Is NOT
 - $^{\circ}$ A tutorial on using a specific DBMS
 - A tutorial on SQL
 - $^{\circ}$ A course on database implementation
- But You WILL learn
 - The foundations of database design
 - Some SQL and relational algebra
 - How to use databases
 - Practical benefits in using a DBMS

Why Study Databases?

- Databases used to be specialized applications; now they are a central component in most applications
- Shift from computation to information
- Knowledge of database concepts is essential for computer scientists
- Databases are everywhere, even when you don't see them... can you give me some examples?

Why Study Databases?

• Because data is valuable

- Examples include tax records, student records, bank account records, photos, ...
- These things must be protected, no matter what happens (disk crash, machine crash, flood, fire, etc.)
- Must also protect from people...

Why Study Databases?

• Because data is typically structured

- Tax records follow the same structure
- Bank records follow the same structure
 We can exploit this structure to help us
- Store the data efficiently
- Retrieve the data efficiently and in useful ways

Why Study Databases?

- Because the database field has made a number of contributions to the field of computer science
- DBs encompass most of CS (OS, language, theory, AI, logic)
- DB concepts can be applied to different problems in different areas
- DBMS s/w is highly successful as a commercial technology
- DB research is highly active and very interesting

Problem: Keeping An Address Book Solution 1: A blank spiral notebook • Entries recorded in pen, in time order • Advantages:

- Cheap, simple, and private
- Reliable and space efficient
- Disadvantages
 - Hard to search, update, and share
- $^{\circ}$ Hard to add info to and limited in size
- Hard to retrieve info from
- Multiple entries are repeated
- What if you lose it?

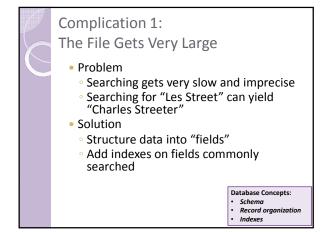
Solution 2: Loose Leaf Notebook

- Advantages
 - Easier to insert, update, and delete
- Can keep sorted
- However
 - All other disadvantages of Solution 1 still apply
 - In particular, still very difficult to search by any "key" other than the one sorted on (probably name)

Solution 3: Text File (Text Editor)

Advantages

- Free format and unlimited size
- Easily copied (can now have backup)
- Easily shared
- Substring searchable
- Cleanly updatable (easy to edit)
- However...
 - What if our requirements grow?
 - This can create some unforeseen complications



Complication 2: Data Redundancy

- Why do we have redundancy?
- Large families, frequent moves
- Might forget to update addresses of some family members
- Want a single point of update and to save space
- Notion of residence as a separate entity is an important concept (e.g., 1 Xmas card per family)
- Solution
 - Separate residences from names: 2 files, one for people and one for residence
 - · How to we associate a residence with a person?

Database Concepts: • Consistency • Normalization

Normalization
Foreign Keys

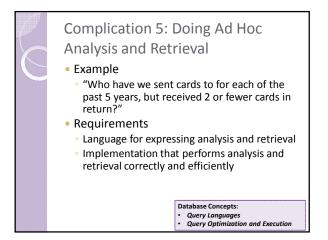
Complication 3: Multiple Associations

of Persons and Residences

- What does this mean?
- People can own, rent, manage, visit residences
- May want constraints on the number of residences per person
- Examples
 - Many-to-one (single family), many-to-many (rich people with vacation homes), one-tomany (builders)

Database Concepts: • Relationships • Cardinality Constraints

Complication 4: Need to Add Information for New Purposes • Examples • Xmas cards sent and received • Post office gives big discount for using zip+4 • Requirements • Adding fields and/or new tables



Complication 6: Want to Organize

the Data Differently for Some Users

Examples

- Other family members want to see the names and residences together
- You don't want your kids to see your business entries
- Solution
- Use stored queries as "windows" into the database
- Data not selected by query is "not there"

 Database Concep • Joins
ViewsSecurity

ts:

Complication 7: Required Existence of Associated Data

Examples

- Can't send a Xmas card to someone without an address
- $^\circ\,$ Names are not unique unless qualified by residence: the Tim Wahls living at 123 Main Street

Solutions

- Refuse to insert a name unless it is associated with an address
- · Refuse to delete an address if it is associated with a name
- Or, tolerate multiple non-unique names...

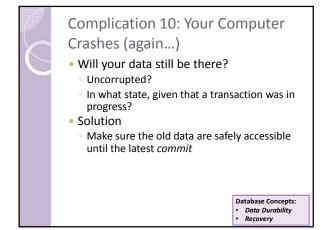
Database Concepts: Referential Integrity Consistency .

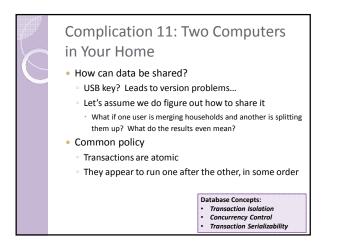
Complication 8: Want Programmed Access to Data • What does this mean? Want to write a Java or C++ program to search, display, and/or update entries Solution Use data organization to define corresponding data types Use access library to open, retrieve, update data Database Concepts: Database Schemas API : : : Embedded Querying

Complication 9: Multiple Updates on All or None Basis

- Examples
- Two households merge
- Requires changing residences of several people
- What if computer crashes halfway through? Solution
- Present illusion that all updates are done simultaneously
- Implanted by commit or rollback of entire piece of work

Da	tabase Concepts:
٠	Transactions
٠	Atomicity





Complication 12: A Home Computer and a Business Computer

Is the up and database and up 2

- Is there one database or two?
- Need speed and reliability at each site
- Logically, have one database for maintenance and one for querying
- Two database communicate
- Solutions
 - Personal data on home computer
 - Business data on business computer
 - Common logical view

Database Concepts:
• Distributed Databases
 Data Partitioning
Data Replication

Complication 13: Want to Add Family Photos, Sound, and Videos Requirements Ability to capture, store, and play new media Logical integration into existing data • Querying: all photos of Tim Wahls fishing ...

Database Concepts: Multimedia Data Query by Content

Complication 14: Your Uncle Sonny Gets the Genealogy Bug Vision

All family members pool their databases over the Internet Together, all genealogy relationships can be recorded

• But

Aunt Sara is paranoid: will not reveal birthdates

You don't want business associates in genealogy database Everyone wants complete control over safety of their own data

People use different formats for records and different name abbreviations for entries

Database Concepts: Federated Databases Data Integration

.

Complication 15: You Become President

- Of a university, a large organization, the USA
- Your address list grows to hundreds of thousands or more
- You realize it contains useful information in the large
- Examples
- Which are the top 10 zip codes on the list?
- Which zip codes have addresses that are most likely to send cards to you when you send card so them?
- Which of those zip codes are in states that had less than 5% difference in Republican/Democratic presidential votes in 2012?

Database Concepts: Data Mining Online Analytical Processing