Organization Of Data Bases

CSE 100

Data is everywhere, but it doesn't become useful until it is organized and structured so that it can be manipulated conveniently. Relational data bases -- one of the great successes of computer science research in the 1970's -- provide the structure and organization

CSE 100 Data Representation

- Though the bits and bytes of digital data represent numbers, letters, strings, etc., the information content is embodied in how the data is grouped or associated
- The basic unit of data base information is a *record*, a sequence (of fixed length) of primitive data values, called *fields*
  - Example, an address record:

William J. Clinton,	1600 Pennsylvania Ave.,	Washington DC,	20020
Name Field	Address Field	State Field	ZIP Field

CSE 100 Aspects Of Record ...

- Each field has a name, called an *attribute* or *field name*
- Each field also has a type such as integer or string
- Notice that fields are generally primitive types, so if the constituents of a field like name are needed, e.g. first name and last name, then they should be defined to be separate fields

William J. Clinton,	1600 Pennsylvania Ave.,	Washington DC,	20020
Name Field string	Address Field string	State Field string	ZIP Field integer

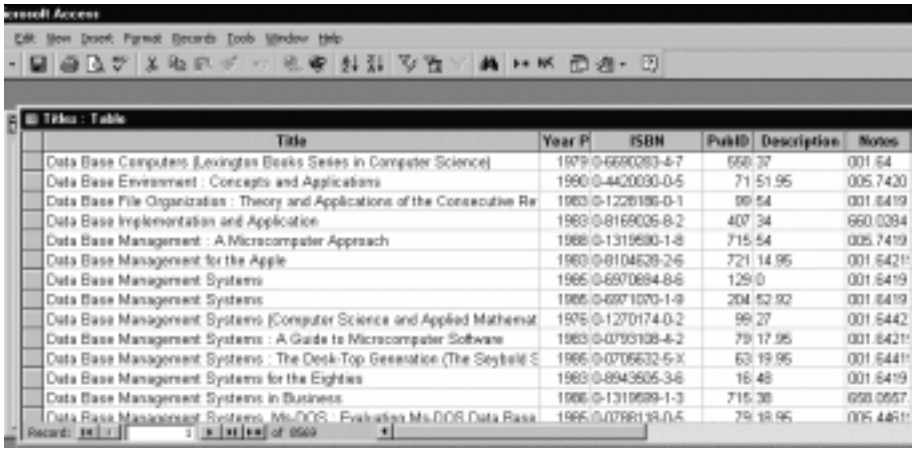
CSE 100 Tables Of Records

- A set of records naturally organize into a table
  - Rows of the table are records
  - Columns of the table are fields
  - Column headings are the attributes
- Tables are also known as *relations* in DB theory

Name	Adm_1	Adm_2	State	Party
William J. Clinton	1992	1996	Arkansas	Dem
George H. W. Bush	1988		Texas	R
Ronald W. Reagan	1880	1984	California	Rep
Jimmy Carter	1976		Georgia	Demo

- Notice variations and errors are possible

CSE 100 The Bibliography Data Base



CSE 100 Titles Table

Titles : Table

Field Name	Data Type	
Title	Text	Book title
Year Published	Number	Year Published
ISBN	Text	International Standard Book Numbe
PubID	Number	Publisher's ID Number
Description	Text	Table of Contents, Pages, Index
Notes	Text	
Subject	Text	
Comments	Memo	

General	Lookup
Field Size	Integer
Format	
Decimal Places	Auto
Input Mask	
Caption	
Default Value	0
Validation Rule	
Validation Text	
Required	No
Indexed	No

F6 = Switch panes. F1 = Help

CSE 100 Keys

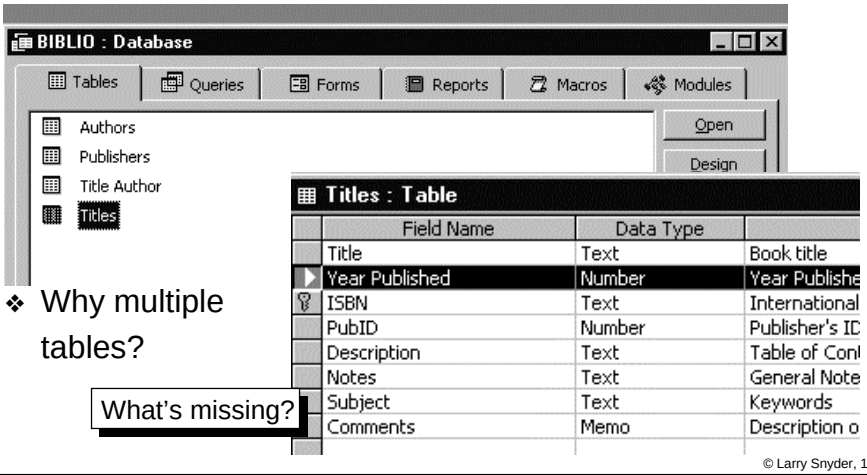
- In a table there must be some field (or collection of fields) that has a unique value for each record … this is a *primary key*
- A key can be used to identify a record, basis for search, etc.

Name	Adm_1	Adm_2	State	Party
William J. Clinton	1992	1996	Arkansas	D
George H. W. Bush	1988		Texas	R
Ronald W. Reagan	1980	1984	California	R
Jimmy Carter	1976		Georgia	D

- Keys for presidents?

CSE 100 A Data Base ...

- Data bases are a collection of tables
  - For Example, Biblio has four tables …

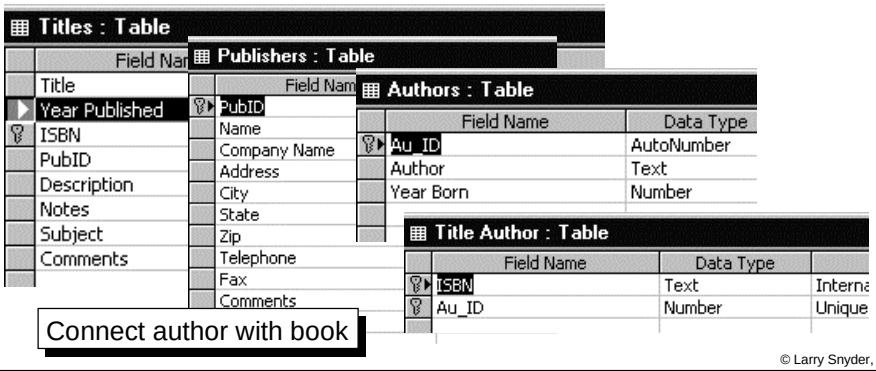


- Why multiple tables?

What's missing?

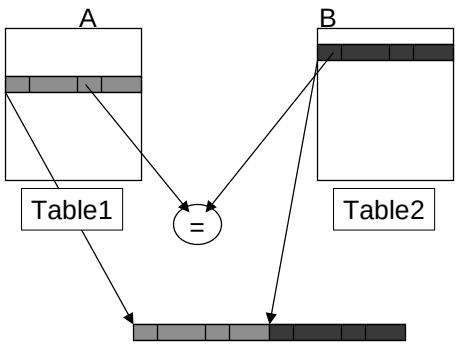
CSE 100 Redundancy Is Bad

- Ideally, the information in a database should appear in exactly one place … if it is in two or more places, it can become inconsistent
- Separate tables can specialize information



CSE 100 Combining Tables

- To be useful, it is necessary to combine information from multiple tables
- This process is called *join*



`Table1 JOIN Table2 ON FieldA = FieldB` means … for every record in Table1 find all records in Table2 that match, and make a new record for the resulting table