

CMPSCI 105: Lecture #11 Introduction to Databases

©2014 Dr. William T. Verts

What is a Database?

- A Database, like a Spreadsheet, is a way of structuring information in order to solve problems,
- Unlike Spreadsheets, Databases are much more restrictive in where information can go,
- Those restrictions make it easier to ask certain types of questions (*Queries*)

©2014 Dr. William T. Verts

Spreadsheets vs. Databases

- Spreadsheets: good at "Figuring Things Out"
 - Networks of interrelated calculations, where a change to one cell causes a ripple effect across the rest of the sheet.
- Databases: good at "Looking Stuff Up"
 - Searching existing tables for information that matches the query,
 - Joining (synthesizing) two or more tables into one.

©2014 Dr. William T. Verts

One Skill You Will Develop

- When do you use a Spreadsheet?
 - Lots of calculations,
 - "What-If" scenarios,
 - Few cases of looking things up in existing information.
- When do you use a Database?
 - Looking for existing items that match queries,
 - Synthesizing new information from old,
 - Few calculations.
- Sometimes it's hard to tell which one to use.
- What are the penalties for guessing wrong?

©2014 Dr. William T. Verts

Types of Databases

- Hierarchical
 - File system on your computer
- Network
 - Distributed Data across widely separated nodes
 - The Internet
- Flat-File
 - All information stored in one table
- Relational
 - Multiple tables plus relationships between tables
 - Anything a Flat-File can do, a Relational can do.

©2014 Dr. William T. Verts

Tables in Flat-File & Relational

- Tables are made up of *Records* (numbered rows),
- Records: all information about a single object,
- Records are made up of *Fields*,
- Fields have:
 - Name (set by designer)
 - Data Type (set by designer)
 - Size (set by designer or determined by data type)
 - Default Value (optional)
 - Validation Rule or Validity Check (optional)

©2014 Dr. William T. Verts

Here's a Sample Table

ID	Last Name	First Name	Birthdate	Salary	Children	Married	Score
12345	Smith	Fred	1/5/1975	\$37,000.00	3	<input type="checkbox"/>	78.6
23451	Jones	Sam	7/20/1969	\$29,000.00	1	<input checked="" type="checkbox"/>	34.9
54233	Baker	Mary	8/15/1978	\$45,000.00	0	<input type="checkbox"/>	87
77464	Walker	Frieda	12/5/1987	\$52,000.00	1	<input type="checkbox"/>	72.1
27356	Smith	Joe	11/2/1974	\$87,000.00	2	<input checked="" type="checkbox"/>	66.9
56432	Welch	Tom	6/23/1988	\$42,000.00	0	<input checked="" type="checkbox"/>	56.8
10000	Sanchez	Mary	9/22/1994	\$49,000.00	0	<input type="checkbox"/>	92.1

©2014 Dr. William T. Verts

Data Types in Microsoft Access

- **Text** (1...255 characters, size set by designer)
- **Memo** (up to 64K characters of free-form text)
- **Date/Time** (8 bytes, like date/times in Excel)
- **Currency** (8 bytes, BCD numbers)
- **Yes/No** (1 bit, two-valued data)
- **Numbers** (lots of subtypes, next slide)

©2014 Dr. William T. Verts

Number Sizes (SubTypes)

- **Byte** (one byte), integer, 0..255,
- **Integer** (two bytes), integer, -32768...+32767,
- **Long Integer** (four bytes), integer, ±2 billion),
- **Single** (four bytes), floating point, ±10^{±38}, 5-6 decimal significant figures,
- **Double** (eight bytes), floating point, ±10^{±308}, 15-16 decimal significant figures.

©2014 Dr. William T. Verts

Here's the Sample Table Again

ID	Last Name	First Name	Birthdate	Salary	Children	Married	Score
12345	Smith	Fred	1/5/1975	\$37,000.00	3	<input type="checkbox"/>	78.6
23451	Jones	Sam	7/20/1969	\$29,000.00	1	<input checked="" type="checkbox"/>	34.9
54233	Baker	Mary	8/15/1978	\$45,000.00	0	<input type="checkbox"/>	87
77464	Walker	Frieda	12/5/1987	\$52,000.00	1	<input type="checkbox"/>	72.1
27356	Smith	Joe	11/2/1974	\$87,000.00	2	<input checked="" type="checkbox"/>	66.9
56432	Welch	Tom	6/23/1988	\$42,000.00	0	<input checked="" type="checkbox"/>	56.8
10000	Sanchez	Mary	9/22/1994	\$49,000.00	0	<input type="checkbox"/>	92.1

©2014 Dr. William T. Verts

Here's the Structure of the Table

Field Name	Data Type	Description
ID	Number	
Last Name	Text	
First Name	Text	
Birthdate	Date/Time	
Salary	Currency	
Children	Number	
Married	Yes/No	
Score	Number	

©2014 Dr. William T. Verts

A Note About "Fields"

- Fields have two definitions, a precise definition and a sloppy definition:
- The precise definition is that a field is a component of a record, as in "The Children field of record 12" (one datum),
- The sloppy definition is that a field is a column of a table, as in "The Children field of the table" (the whole column).

©2014 Dr. William T. Verts

Let's Build a Table about People

- We have to decide how many fields we need.
- For each field we need:
 - Name (should describe the field contents)
 - Type (Text, Number, Currency, Yes/No, etc.)
 - Size (where we are allowed to do so)
 - Default Value (common value for new records)
 - Validation Rule (constraints on field values)

©2014 Dr. William T. Verts

A Rule to Follow

- When deciding Data Type or Size, pick the smallest one that entirely covers the expected range of values.
- Too small and not all values can be covered,
- Too large and you are wasting space.

©2014 Dr. William T. Verts

Field: Last Name

- Type: **Text**
- Size: ??? (Make a guess: 15? 20?)
 - Too few characters cuts off names
 - Too many characters wastes space
 - You can always widen/narrow the field later
 - Caution: characters lost when narrowing field are gone forever
- No Default Value or Validation Rule

©2014 Dr. William T. Verts

Field: First Name

- Type: **Text**
- Size: ??? (longest first name that I see on a regular basis is "Christopher" – 11 characters, I usually even it out at 12 characters)
- No Default Value
- No Validation Rule

©2014 Dr. William T. Verts

Field: Middle Initial

- Type: **Text**
- Size: **1**
- No Default Value
- No Validation Rule

©2014 Dr. William T. Verts

Field: Salary

- Type: **Currency**
- Size: We don't get a choice (8 bytes)
- No Default Value
- Validation Rule:
 - >=20000 AND <=100000**
 (prevents stupid data entry errors such as 5000 or 500000)

©2014 Dr. William T. Verts

Field: **Birthdate**

- Type: **Date/Time** (Not Text!)
- Size: We don't get a choice (8 bytes)
- No Default Value
- Validation Rule:
 $\geq 1/1/1949$ AND $\leq 12/31/1996$

©2014 Dr. William T. Verts

Field: **Children**

- Type: **Number**
- Size: **Byte**
- Default Value: **0** (how many kids assumed?)
- Validation Rule:
 ≤ 20
 (no lower limit needed because Byte enforces a lower limit of 0, but we do need to pick an upper limit)

©2014 Dr. William T. Verts

Field: **Married** (option 1)

- Type: **Yes/No**
- Size: We don't get a choice (1 bit)
- Default Value: **No**

©2014 Dr. William T. Verts

Field: **Married** (option 2)

- Type: **Text**
- Size: **1**
- Default Value: **S**
- Validation Rule:
 $= "S" \text{ OR } "M" \text{ OR } "D" \text{ OR } "W"$

©2014 Dr. William T. Verts

Field: **Sex**

- Type: **Text**
- Size: **1**
- No Default Value
- Validation Rule:
 $= "M" \text{ OR } "F"$

©2014 Dr. William T. Verts

Field: **Score**

- Type: **Number**
- Size: **Single**
- No Default Value
- Validation Rule:
 ≥ 0.0 AND ≤ 100.0

©2014 Dr. William T. Verts

Structure vs. Data

- A Table can:
 - Exist
 - Have structure
 - Take up space on the disk
 - Yet contain no data (be empty)
- Only when the structure has been defined can data be added
- The structure can be modified after data have been added.

©2014 Dr. William T. Verts

Running Queries Depends on the DB

- dBase used text-based pseudo-English sentences:
 - LIST LastName, FirstName FOR Salary < 40000 .AND. Children >= 2
 - Requires only very modest computing power
- More modern databases (MS Access) use QBE (Query By Example) graphical screen-based forms to give examples of what is being searched for.
 - Easier to use than typing sentences
 - Requires more powerful computers to do gracefully.

©2014 Dr. William T. Verts