

# Databases

## Managing Data for Retrieval, Update, & Calculation

Drilling Record			
ID	Name	Spudded	Completed
40	Exxon #1	2/4/96	6/3/96
43	Shell #5	3/14/97	6/12/96

Production (barrels/day, cfs)			
ID	Oil	Gas	Water
40	53	1200	5
55	108	2500	15

1997 Expenditures (millions of \$s)			
ID	Drilling	Production	Transport.
40	1.501	0.652	0.078
72	5.522	0.301	0.055

# GIS Data Recap

Two data types:

- ⌘ Spatial - *Where* things are, in (x, y)
  - Stored in coord. files & topology *tables*
    - Vector Model
    - Raster Model
  
- ⌘ Aspatial - *What* things are
  - stored in tables of attributes

# GIS = Lots and Lots of Tabular Data

- ⌘ How will it be managed?
  - Data Model Considerations
  - Analysis Considerations
  - Data Entry Considerations
  - Security
  - Efficiency

# HOW?

## ⌘ Goals:

- ⌘ Maximize flexibility for sorting, reordering, subsetting, searching
- ⌘ Efficient storage; *eliminate redundancy*
- ⌘ Secure entry and retrieval mechanisms
- ⌘ Rapid retrieval

## ⌘ Solution:

Database Management System (DBMS)

# Accuracy

- ⌘ Data entry can be accomplished via *forms* that require:
  - *Data definitions* - #s of attributes, the types and lengths or numerical ranges of each attribute, and how much editing will be permitted.
  - *Data Dictionary* - catalogue of attributes with their permitted values and ranges ("**Domains**").
  - *Validation Rules* - ensure data integrity.

# Database display in ArcMap

- ⌘ Displayed in tables with rows of *records* (*tuples*) and columns of *fields* (*attributes*)

SHAPE	Date_	Digitize	Instrument	Software	Unit Contact	Line_type	Cor
Polyline Z	3/14/2005	Digital Ink	Tablet	ArcMap	pC-Crh	Dashed	hick 2n
Polyline Z	3/15/2005	Digital Ink	Tablet	ArcMap	pCcc-dm	Solid	Cal south
Polyline Z	3/15/2005	Digital Ink	Tablet	ArcMap	pCcc-dm	Solid	doll South
Polyline Z	3/15/2005	Digital Ink	Tablet	ArcMap	pCcc-dm	Solid	dol Lense
Polyline Z	3/15/2005	Digital Ink	Tablet	ArcMap	pCcc-dm	Solid	Cal South
Polyline Z	3/15/2005	Digital Ink	Tablet	ArcMap	pCcc-dm	Solid	<Null>
Polyline Z	3/15/2005	Digital Ink	Tablet	ArcMap	pCcc-dm	Solid	Cal North
Polyline Z	3/15/2005	Digital Ink	Tablet	ArcMap	pCcc-dm	Solid	dol lense
Polyline Z	3/15/2005	Digital Ink	Tablet	ArcMap	pCcc-dm	Solid	cal North

Line\_type field

# Domains in a Geodatabase

The image shows a screenshot of ArcGIS Desktop with several windows and annotations. On the left, a tree view shows a geodatabase named 'Marble\_features.mdb' containing a feature class 'Contacts05'. The 'Database Properties' window is open to the 'Domains' tab, showing a table of domains for the 'Contacts05' feature class. The 'Line\_type' domain is selected, and its properties are shown in the 'Domain Properties' section. The 'Feature Class Properties' window is also open, showing the 'Fields' tab with a table of fields. The 'Line\_type' field is selected, and its properties are shown in the 'Field Properties' section. Red arrows and text annotations highlight these elements.

**Geodatabase**

**Geodatabase Domains**

**Domain Values**

**Domain applied to Contacts05**

**Contacts05 Feature Class Fields**

Domain Name	Description
Contact_type	Solid or Dashed Contact
Date_	When Acquired
Date_1	When Acquired
Date_2	When Acquired
Digitize	Digitizing technique
Dip_Plunge	Inclination of Plane or Line
Geologist	Data Collector

Field Type	Text
Domain Type	Coded Values
Split policy	Default Value
Merge policy	Default Value

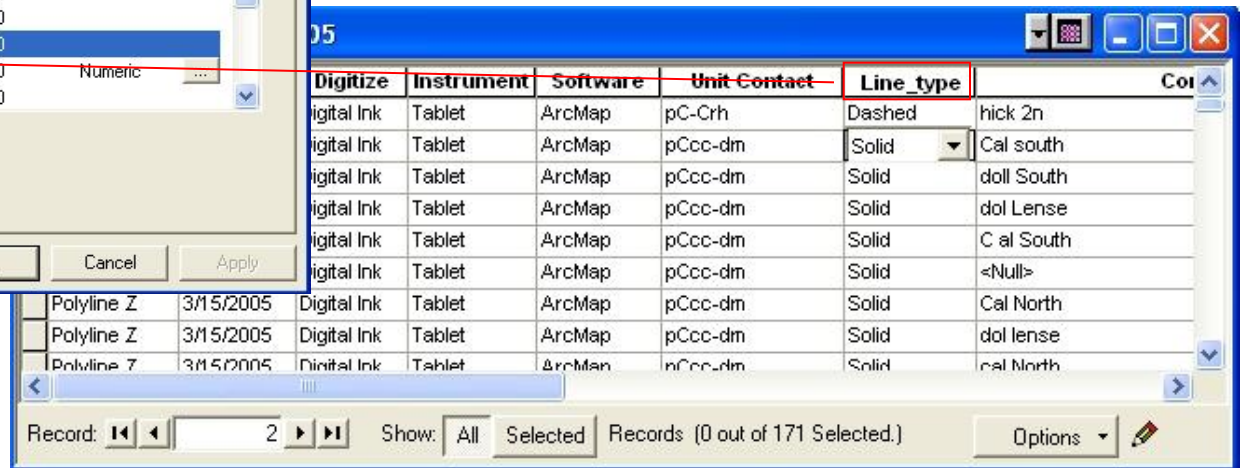
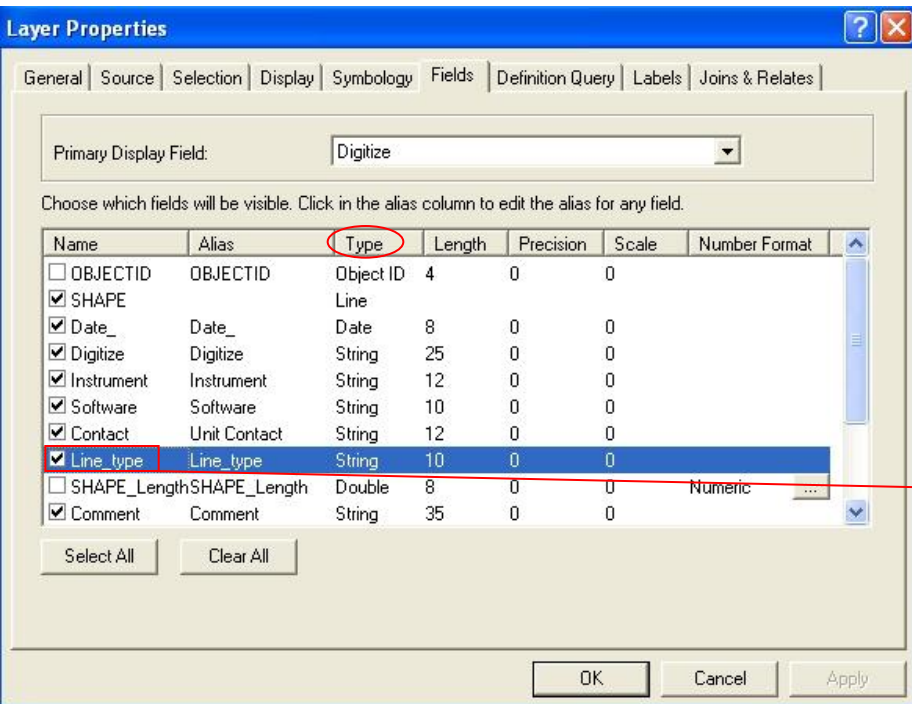
Code	Description
01	Dashed
02	Solid
03	Dotted

Field Name	Data Type
Digitize	Text
Instrument	Text
Software	Text
Contact	Text
Line_type	Text
Comment	Text
Geologist	Text
Journal	Text

Field Properties	Line_type
Alias	
Allow NULL values	Yes
Default Value	Solid
Domain	Contact_type
Length	10

# Example Field Types

⌘ Field type - how much space does database need to set aside for each field?



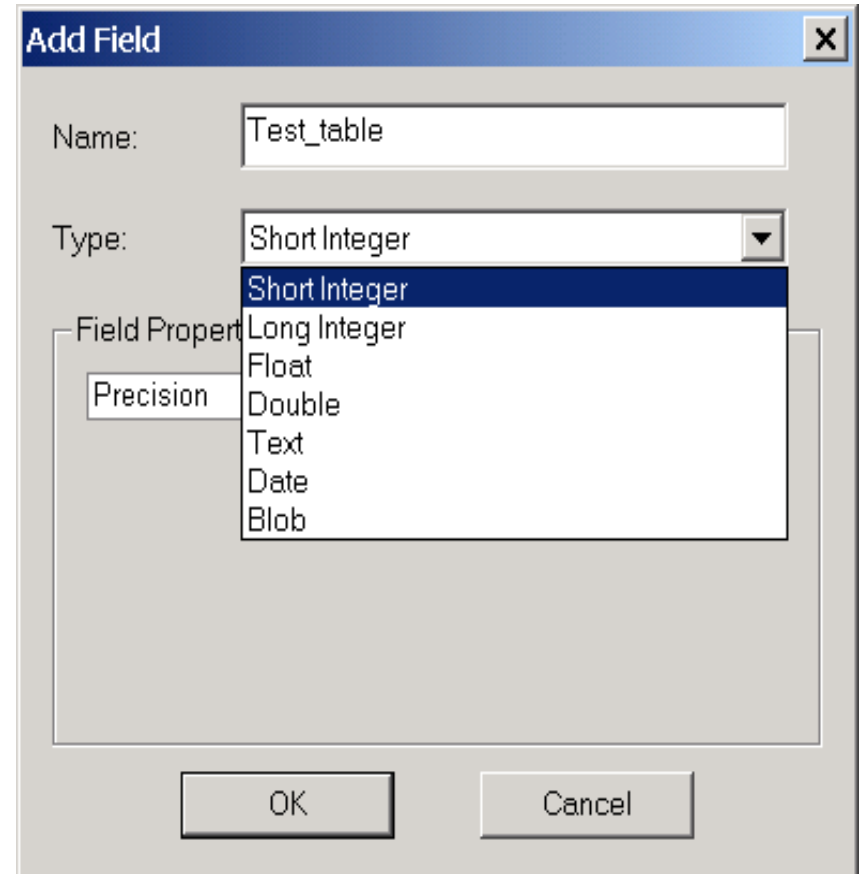
# Fields are defined:

- ⌘ Name - attribute (column heading)
- ⌘ Field Type - number (long, short, float, double), text ("string"), or date
- ⌘ Length - no. of characters in text
- ⌘ Precision - no. of digits used to store numbers
- ⌘ Scale - no. of digits to right of decimal point

Name	Type	Length	Precision	Scale	
FID	Object ID	4	0	0	
Shape	Point				
IDNUMBER	String	13	0	0	
PIT_TYPE	String	21	0	0	
STATUS	String	21	0	0	
PARISH	String	21	0	0	
CONTAINMT	String	51	0	0	
CONT_COND	String	12	0	0	
BREACHED	String	2	0	0	
RANKING	Long	6	6	0	

# Fields types in ArcGIS

- ⌘ Short Integer - 1 to 4 digits (no decimal)
- ⌘ Long Integer = 5 to 9 digits (no decimal)
- ⌘ Float = 1 - 8 digits, decimal (short real)
- ⌘ Double = 6 - 19 digits, decimal (long real)
- ⌘ Text = 1 -255 characters
- ⌘ Date = 8 character
- ⌘ Blob = binary large object



# Numeric Field Types

Data Type	Storage	Range	Description
Short Integer	2 bytes	+/- 32,768	Used for coding, e.g. lulc, veg. types, T/F
Long Integer	4 bytes	+/- 2.14 billion	Large whole numbers, e.g. populations
Float	4 bytes	+/- 3.4 x 1,038	Single-precision, up to 6 places past the decimal. Up to 8 total numbers.
Double	8 bytes	+/- 1.8 x 10,308	Double-precision; 15 places past decimal, 6-19 total numbers.

# Field Properties

- ⌘ Precision = number of digits stored in a field. Precision  $>6$  for double,  $<6$  float
- ⌘ Scale = no. of decimal places in double and float
- ⌘ E.g. 3500426.21 should be stored as float or double, precision  $\geq 9$ , scale  $\geq 2$

The screenshot shows a dialog box titled "Add Field" with a close button (X) in the top right corner. The dialog contains the following fields and controls:

- Name:** A text input field containing "Test\_table".
- Type:** A dropdown menu currently showing "Float".
- Field Properties:** A section containing a table with two rows:

Precision	0
Scale	0
- Buttons:** "OK" and "Cancel" buttons at the bottom.

# Numeric Field Properties

⌘ Short and Long integers fields:

Precision = 4

8,400

⌘ Float and Double data fields:

Precision = 9

8,400.08347

Scale = 5

# File Size Comparison, Text Fields

Text Field length	100 records	1000 records	10,000 records	100,000 records
2	0.2 Kb	1.95 Kb	19.53 Kb	195.3 Kb
50 (default)	4.88 Kb	48.83 Kb	488.3 Kb	4.88 Mb

# A DBMS provides:



- ✓ Accuracy - reduce errors during entry by use of established rules, templates
- Efficiency - rapid access & retrieval, no redundancy

# Efficiency & Flexibility

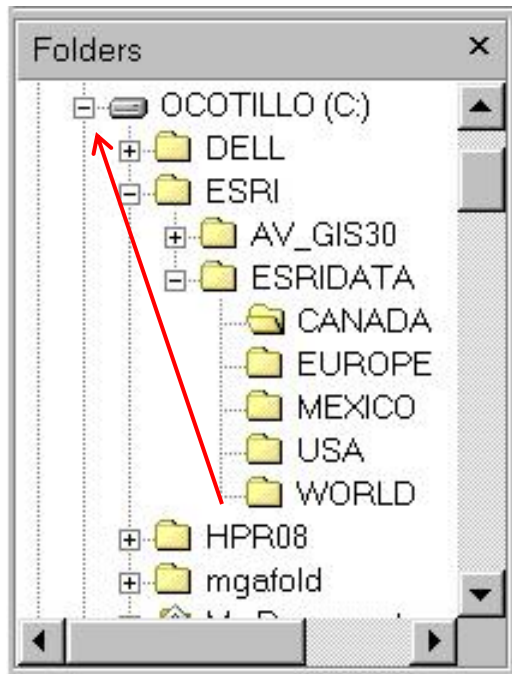
- ⌘ Relies on database structure (data model):
  - Hierarchical
  - Network
  - Relational
  - Object-oriented

# GIS attribute data models

- ⌘ *Hierarchical* - pre-1980
- ⌘ *Relational* - 1980's, 1990's; still dominant today
- ⌘ *Object-oriented* - late '90's; newest, implemented by some GISs - still undergoing R&D

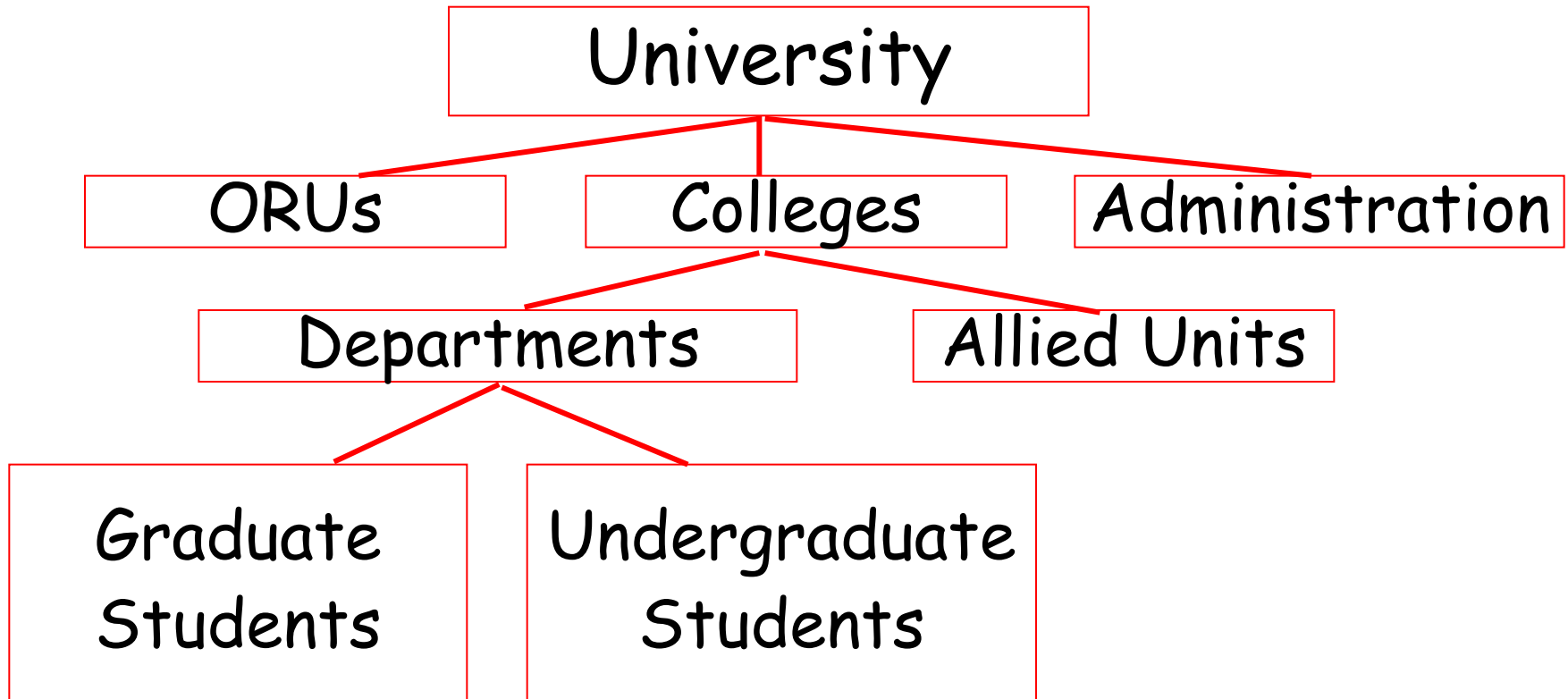
# Hierarchical Structure

E.g. Filing cabinet or folders on a hard drive



⌘ File address for storage and retrieval is a **linear path**, e.g.  
`C:\ESRI\ESRIDATA\CANADA\cities.shp`

# Hierarchical Structure



# Hierarchical - Limitations

⌘ 1) Linear structure can't deal with multiple "memberships"

⌘ E.g. a single well might be stored under different databases for taxes, production, drilling history, water quality, etc.

INEFFICIENT

⌘ Can't assemble all this data for query in a hierarchical database

# Hierarchical - Limitations

2) Can't deal with exceptions to linear scheme - entities may not belong to next higher class but could instead contain it.

E.g. Structure Oil Well Database by:

State

County

Oil Field

Well

Pay zone

What of field that spans several counties with wells that produces from more than one pay zone?

⌘ i.e. No "*one-to-many*" relationships

# Relational Database advantages

- ⌘ Data stored in separate files
  - Easy update, editing, searching without affecting or using all data
- ⌘ Flexibility
  - Using key(s), can extract and assemble records and attributes to form new tables
  - Subsets of database can be queried by standard means - SQL

# Relational Database Structure

- ⌘ Consists of "relations" (tables) with multiple attributes (columns) per record
- ⌘ Every record (row) has a unique identifier (marker or *key* attribute)
  - Key is the glue between files that can be used to extract and/or assemble records and attributes

# Parts of a Relation

Primary Key

Production (barrels/day, cfs)			
ID	Oil	Gas	Water
40	53	1200	5
55	108	2500	15

Record or tuple

Attribute or field

# Properties of Relations

- ⌘ Each row has to be unique; no row-to-row dependency
- ⌘ Row order irrelevant
- ⌘ Column order irrelevant
- ⌘ All attribute values must be stored in separate rows ("first normal form")

# Relational Database Structure

Drilling Record			
ID	Name	Spudded	Completed
40	Exxon #1	2/4/96	6/3/96
43	Shell #5	3/14/97	6/12/96

File

Production (barrels/day, cfs)			
ID	Oil	Gas	Water
40	53	1200	5
43	108	2500	15

File

1997 Expenditures (millions of \$s)			
ID	Drilling	Production	Transport.
40	1.501	0.652	0.078
72	5.522	0.301	0.055

File

⌘ Key Field = ID

# One-to-One Table Join

Drilling Record			
ID	Name	Spudded	Completed
40	Exxon #1	2/4/96	6/3/96
43	Shell #5	3/14/97	6/12/96

Production (barrels/day, cfs)			
ID	Oil	Gas	Water
40	53	1200	5
43	108	2500	15

⌘ One record from source table (production) is joined to one record of destination (drilling record) table to create a "View" - virtual combination

# Result of One-to-One Table "Join"

- ⌘ Joined Production table and Drilling Record:

Drilling Record						
ID	Name	Spudded	Completed	Oil	Gas	Water
40	Exxon #1	2/4/96	6/3/96	53	1200	5
43	Shell #5	3/14/97	6/12/96	108	2500	15

- ⌘ View can't be edited - destination table can be

# One-to-Many Join

Drilling Record				
ID	Well Name	Spudded	Completed	Field_ID
40	Exxon #1	2/4/96	6/3/96	2
43	Shell #5	3/14/95	6/12/96	2
72	Amoco #3	4/8/88	4/8/89	2
55	BP #2	6/8/90	8/8/91	Wildcat

⌘ One record from source table joined to many records of destination table

Oil/Gas Fields			
Field_ID	Name	Discovered	Total_Oil
1	Longview	1/20/56	13000564
2	Katy	2/3/48	85640
3	Anhuac	4/11/73	3587889

# One-to-Many Join Result

Drilling Record							
ID	Name	Spudded	Completed	Field_ID	Name	Discovered	Total_Oil
40	Exxon #1	2/4/96	6/3/96	2	Katy	2/3/48	85640
43	Shell #5	3/14/95	6/12/96	2	Katy	2/3/48	85640
72	Hess #3	4/8/88	4/8/89	2	Katy	2/3/48	85640
55	BP #2	6/8/90	8/8/91	Wildcat			

# Many-to-Many Join example - USGS DLGs

- ⌘ Join "lookup table" with feature codes tables to obtain feature descriptions
  - Feature descriptions stored once, used many times
  - Primary key is feature code

# Digital Line Graph Example

Lookup Table →

Key

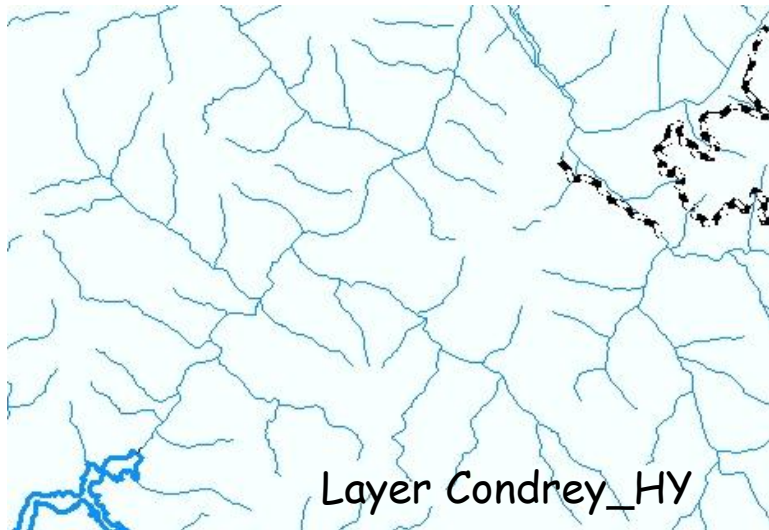
OID	LAYER	COLO	LTYPE	MAJ	MIN	NA	FEAT_DESC
390	50_300	5	CONTINUOUS	0	Spring_50_300		Spring
391	50_301	5	CONTINUOUS	0	Nonflowing_well_50_301		Nonflowing well
392	50_302	5	CONTINUOUS	0	Flowing_well_50_302		Flowing well
393	50_303	5	CONTINUOUS	0	Riser_50_303		Riser
394	50_304	5	CONTINUOUS	0	Geyser_50_304		Geyser
395	50_305	5	CONTINUOUS	0	Windmill_50_305		Windmill
396	50_306	5	CONTINUOUS	0	Cistern_50_306		Cistern
397	50_4	5	CONTINUOUS	0	Stream_50_4		Stream entering water body
398	50_400	5	CONTINUOUS	0	Rapids_50_400		Rapids
399	50_401	5	CONTINUOUS	0	Falls_50_401		Falls
400	50_402	5	CONTINUOUS	0	Gravel_50_402		Gravel pit or quarry filled with water
401	50_403	5	CONTINUOUS	0	Gaging_50_403		Gaging station
402	50_404	5	CONTINUOUS	0	Pumping_50_404		Pumping station
403	50_405	5	CONTINUOUS	0	Water_50_405		Water intake
404	50_406	5	CONTINUOUS	0	Dam_50_406		Dam or weir
405	50_407	5	CONTINUOUS	0	Canal_lock_50_407		Canal lock or sluice gate
406	50_408	5	CONTINUOUS	0	Spillway_50_408		Spillway
407	50_409	5	CONTINUOUS	0	Gate_flood_50_409		Gate (flood, tidal, head, or check)
408	50_410	5	CONTINUOUS	0	Rock_50_410		Rock
409	50_411	5	CONTINUOUS	0	Crevasse_50_411		Crevasse
410	50_412	5	CONTINUOUS	0	Stream_50_412		Stream

TPMINOR_1	TPMAJOR_2	TPMINOR_2	TPMAJOR_3	TPMINOR_3	Code
300	-999	-999	-999	-999	50_300
300	-999	-999	-999	-999	50_300
301	-999	-999	-999	-999	50_301
300	-999	-999	-999	-999	50_300
412	-999	-999	-999	-999	50_412
412	-999	-999	-999	-999	50_412
412	-999	-999	-999	-999	50_412
412	-999	-999	-999	-999	50_412
412	-999	-999	-999	-999	50_412

Hydrography feature attributes

# Result of Many-to-Many Join

⌘ Symbolize on joined field



Attributes of Condrey\_HY

C	C	Condrey_HY	m	m	m	m	maplayer.MAJ	MIN	NA	maplayer.FEAT	DESC
-	-9	50_301	3	50	5	C	0			Nonflowing_well_50_301	Nonflowing well
-	-9	50_301	3	50	5	C	0			Nonflowing_well_50_301	Nonflowing well
-	-9	50_301	3	50	5	C	0			Nonflowing_well_50_301	Nonflowing well
-	-9	50_305	3	50	5	C	0			Windmill_50_305	Windmill
-	-9	50_412	4	50	5	C	0			Stream_50_412	Stream
-	-9	50_412	4	50	5	C	0			Stream_50_412	Stream
-	-9	50_412	4	50	5	C	0			Stream_50_412	Stream
-	-9	50_412	4	50	5	C	0			Stream_50_412	Stream

Record: 0 Show: All Selected Records (0 out of 3035 Selected.)

- Ditch or canal
- Left bank
- Nonflowing well
- Right bank
- Shoreline
- Siphon
- Spring
- Stream
- Windmill

# A DBMS provides:

- ✓ Accuracy - reduce errors during entry by use of established rules, templates
- ✓ Efficiency - rapid access & retrieval, no redundancy
- **Flexibility** - robust structure for query - e.g. What is where?

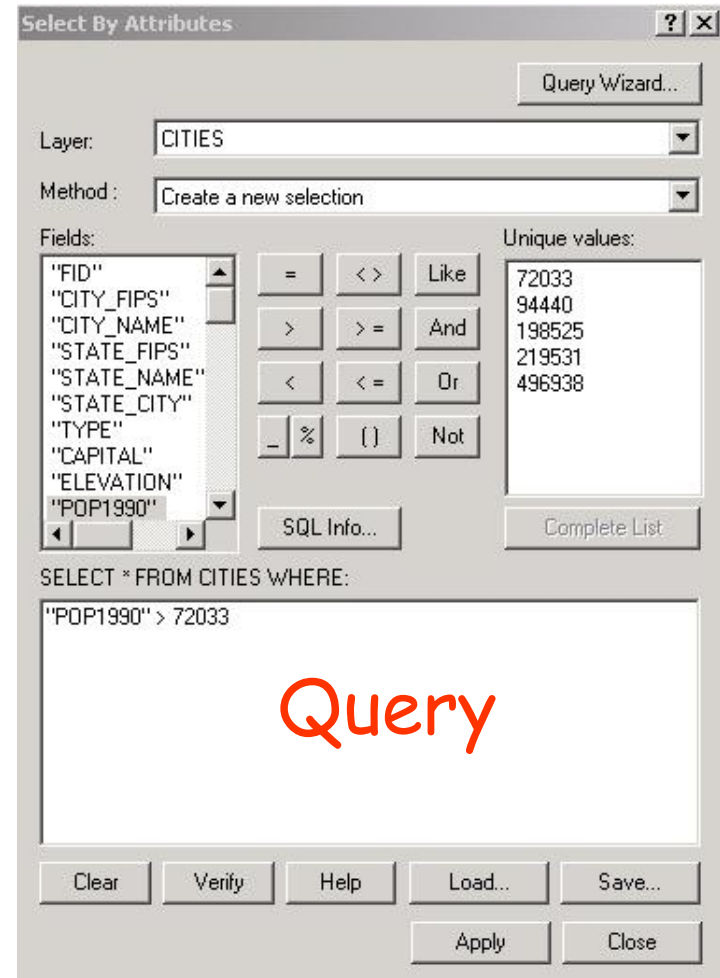
# Flexibility



- ⌘ Using primary key(s), can extract and assemble records and attributes to form new tables
- ⌘ Subsets of database can be queried by standard means - SQL

# ArcMap Query Builder

⌘ E.g. Find all cities in Louisiana where 1990 population exceeded 72,033





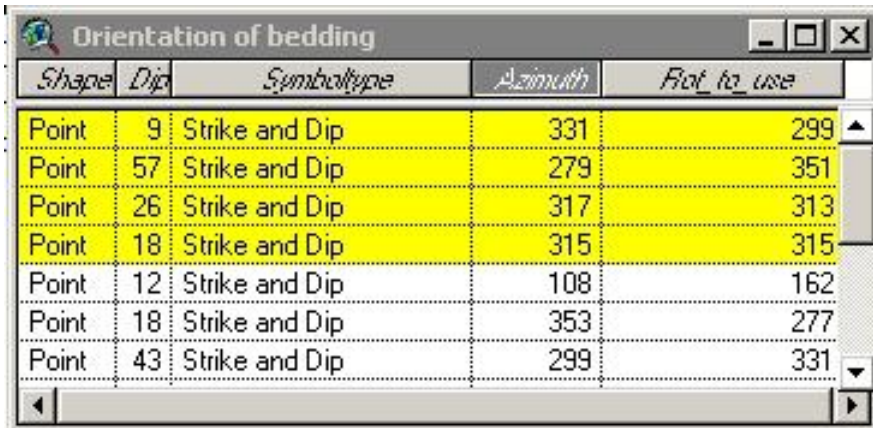
# Relational DBMSs Permit:



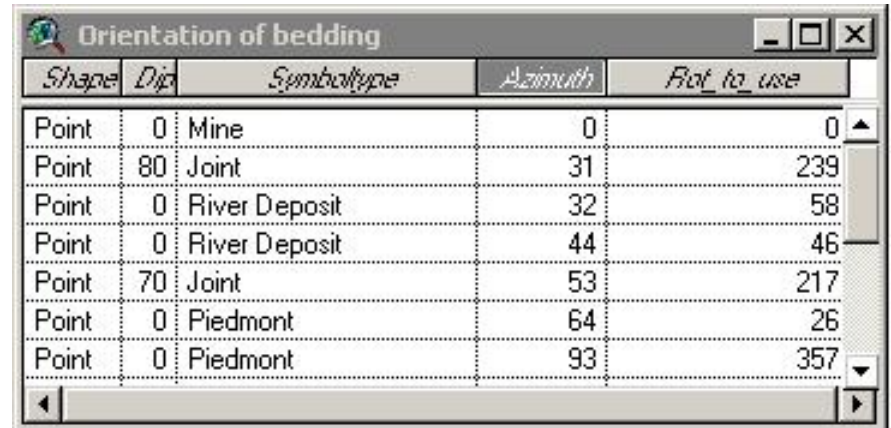
- ⌘ File updating
- ⌘ Data retrieval via query using a standard language (SQL)
- ⌘ Sorting (reordering) by field values
- ⌘ Calculations and field statistics
- ⌘ Report generation
- ⌘ Multi-user access

# Reordering In ArcMap

1. Order selected records by sorting    
⌘ ascending or descending field values
2. Sort records by selected attributes



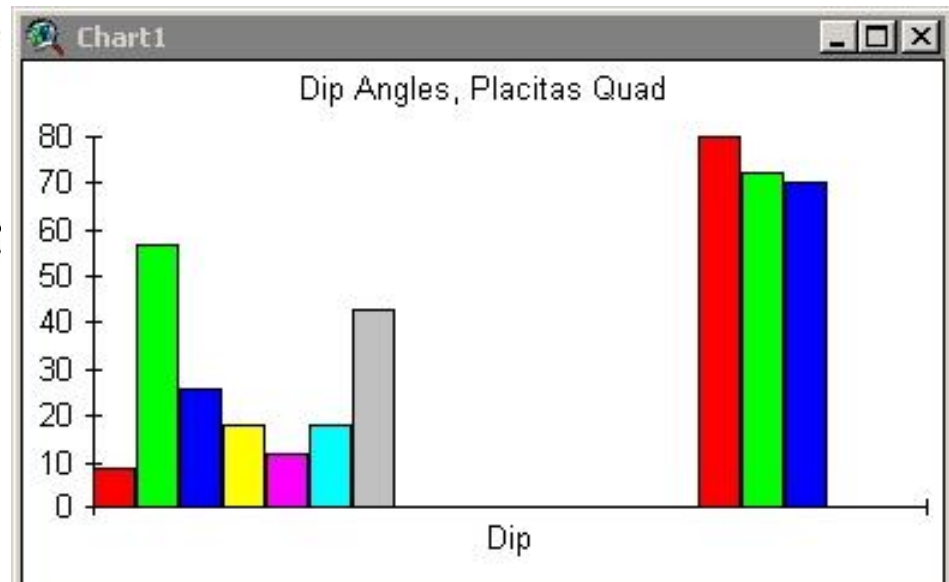
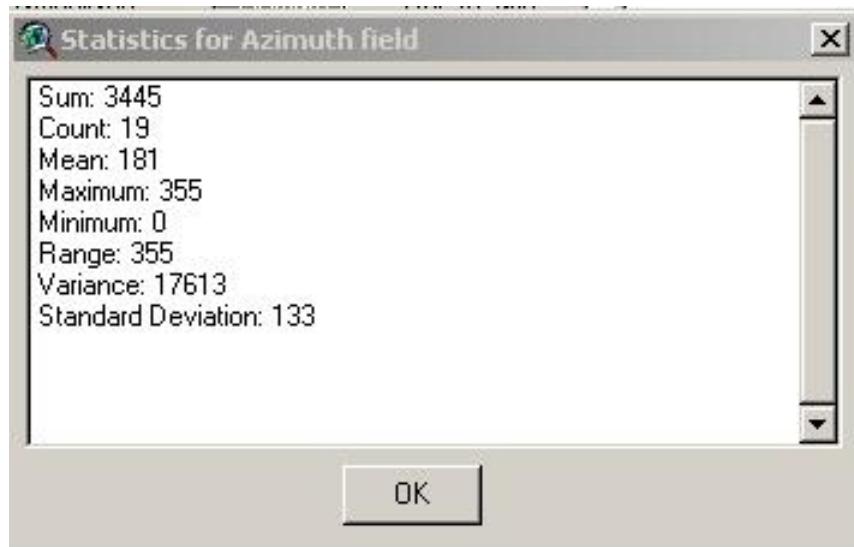
Shape	Dip	Symboltype	Azimuth	Rot. to use
Point	9	Strike and Dip	331	299
Point	57	Strike and Dip	279	351
Point	26	Strike and Dip	317	313
Point	18	Strike and Dip	315	315
Point	12	Strike and Dip	108	162
Point	18	Strike and Dip	353	277
Point	43	Strike and Dip	299	331



Shape	Dip	Symboltype	Azimuth	Rot. to use
Point	0	Mine	0	0
Point	80	Joint	31	239
Point	0	River Deposit	32	58
Point	0	River Deposit	44	46
Point	70	Joint	53	217
Point	0	Piedmont	64	26
Point	0	Piedmont	93	357

# Field Statistics In ArcMap

1. Get stats. & graphs on selected attributes



# GIS' are Spatial Databases

## ⌘ Coverage and Shapefile models

- Spatial information stored in spatial attribute files, attributes in relational database table
  - Feature ID is key
  - Spatial information can't participate in relational database advantages

## ⌘ Geodatabase model

- All information, spatial and aspatial, are stored together in a relational database