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, z, λ, ϕ)
  - Stored in coordinate & topology *tables*
    - Vector (Object) Model
    - Raster (Field) 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

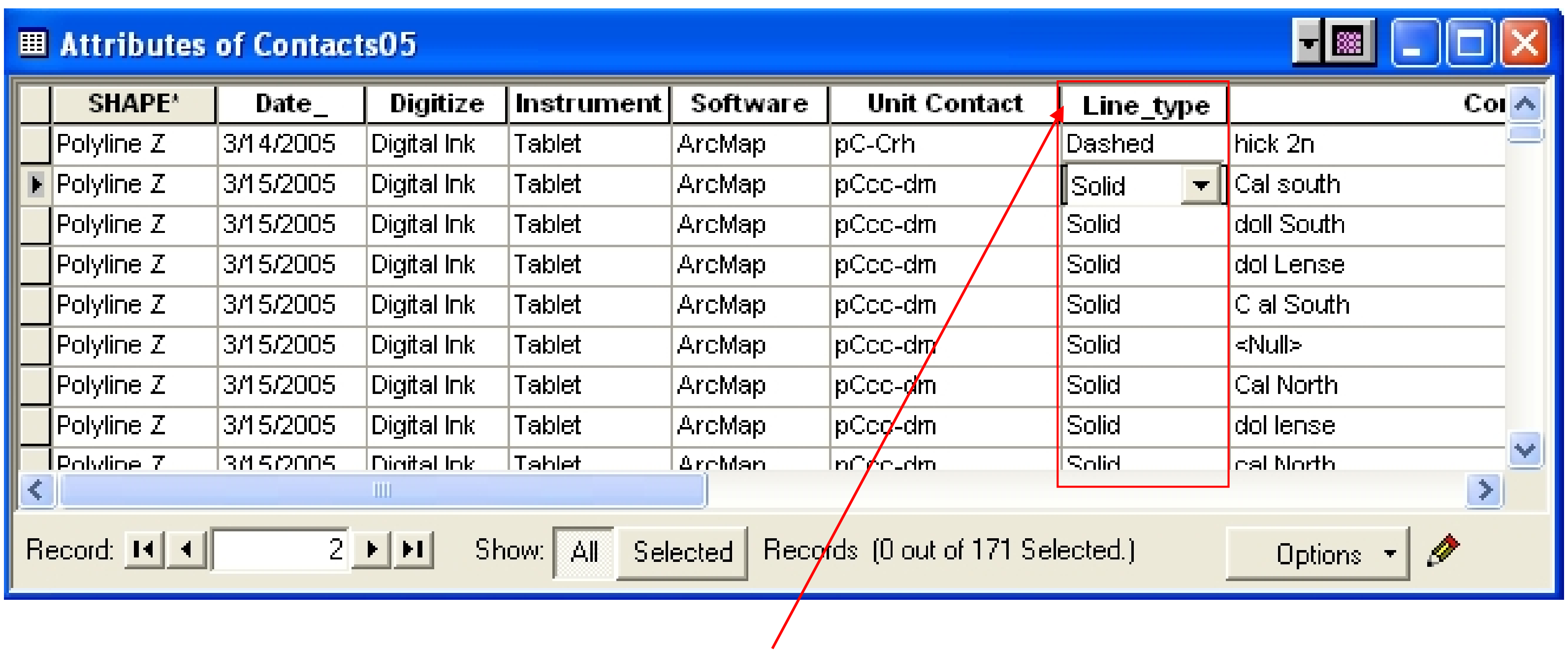
WHY? HOW?

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

Database Management System (DBMS)

Database display in ArcMap

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


Attributes of Contacts05

	SHAPE*	Date_	Digitize	Instrument	Software	Unit Contact	Line_type	Co
	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	C al 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

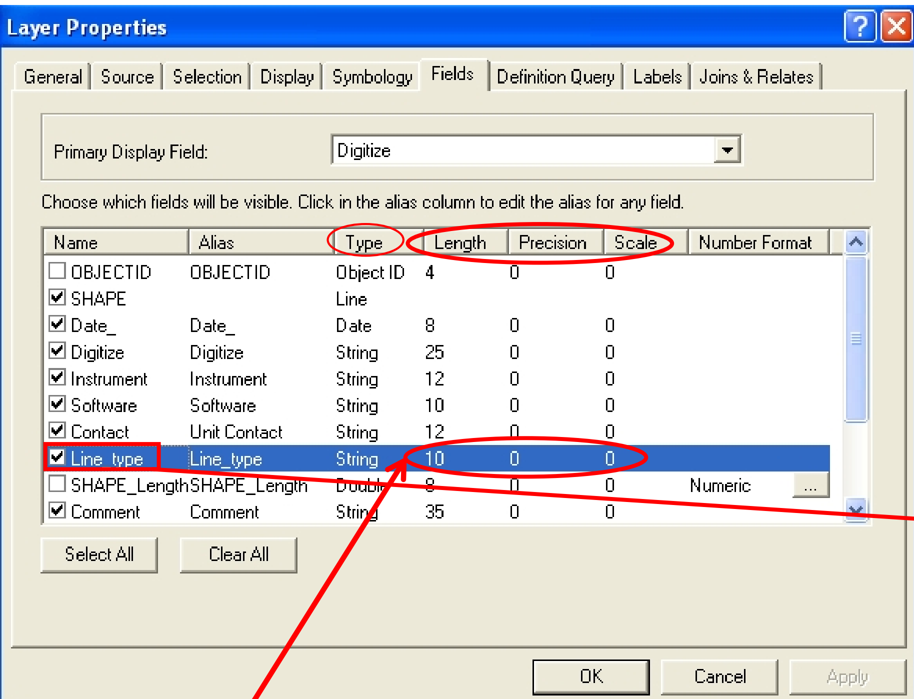
Record: 2 Show: All Selected Records (0 out of 171 Selected.) Options

Line_type **field**

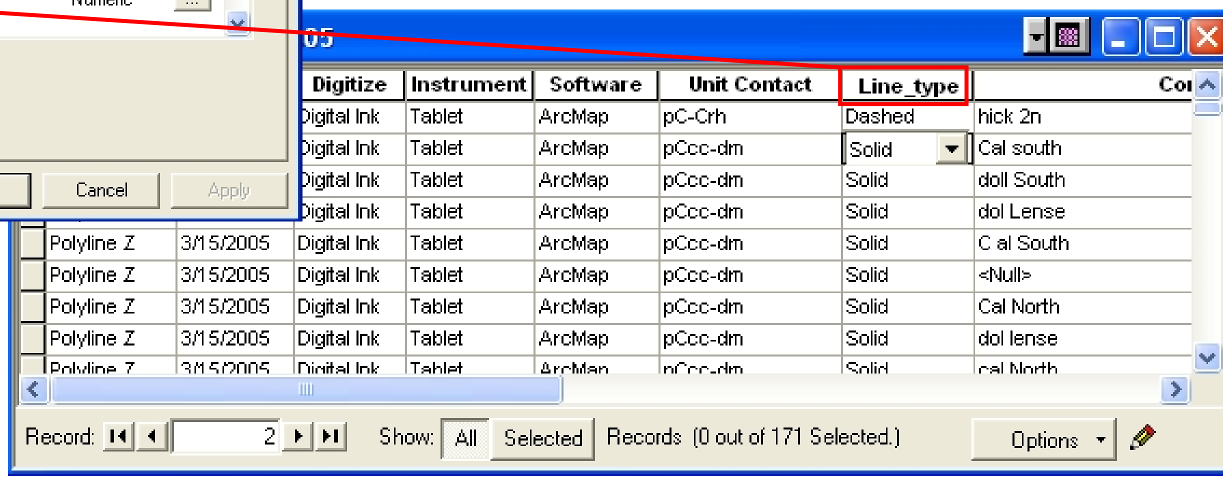
Accuracy

- 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* - catalog of attributes with their permitted values and ranges ("Domains").
  - *Validation Rules* - ensure data integrity.

Data Definition: Field "Types"



- Field type- Q: how much space does a database need to reserve for each field?



- A: no more than 10 characters

Fields are “defined” by:

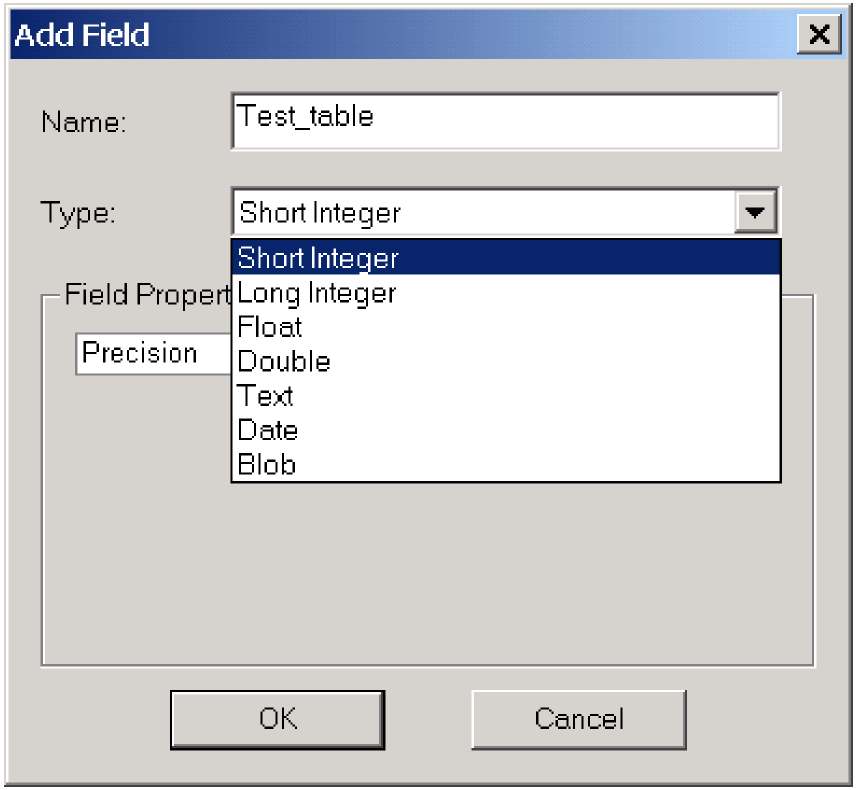
- Name – attribute (column heading)
- Field Type – number (long, short, float, double), text (“string”), or date
- Length – no. of characters in text
- Precision – no. of digits used to store numbers
- Scale – no. of digits to right of decimal point

Field Definitions

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

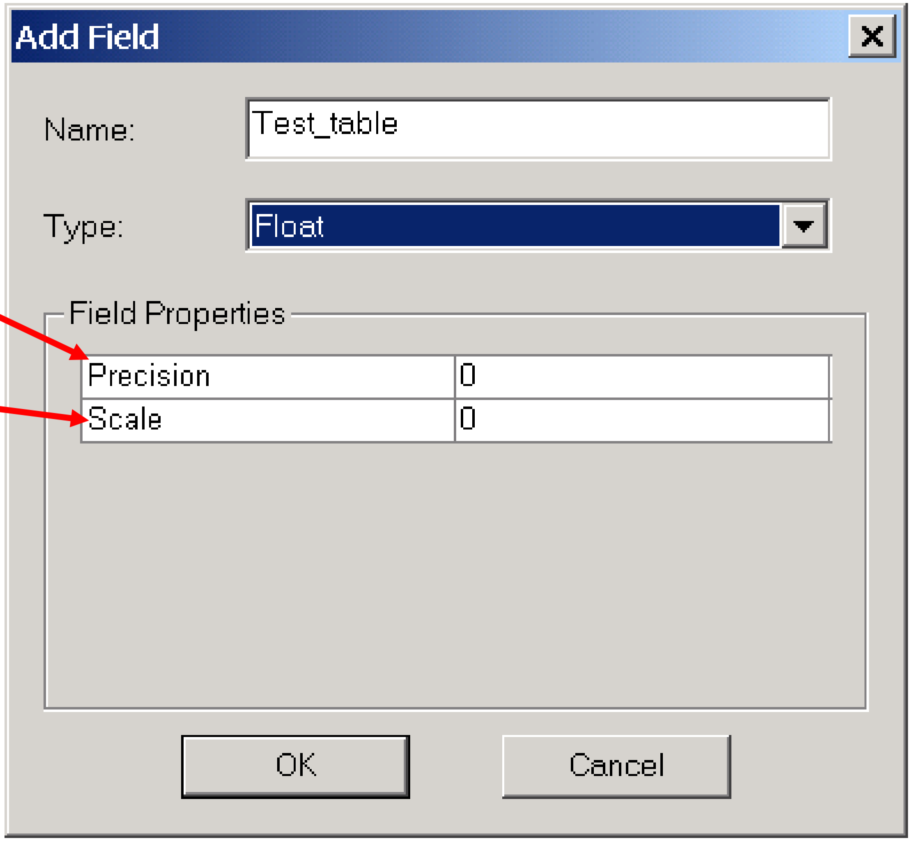


Numeric Field Types

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 E38	Single-precision, up to 6 places past the decimal. Up to 6 total numbers.
Double	8 bytes	~+/- 1.8 x E308	Double-precision; 15 places past decimal, 6-19 total numbers.

Field Properties For Numbers

- Precision = number of digits stored in a field. Precision up to for 19 double, 8 for float
- Scale = no. of decimal places in double and float
- E.g. 3500426.21 (a typical easting) should be stored as “Double”, precision 9, scale 2



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

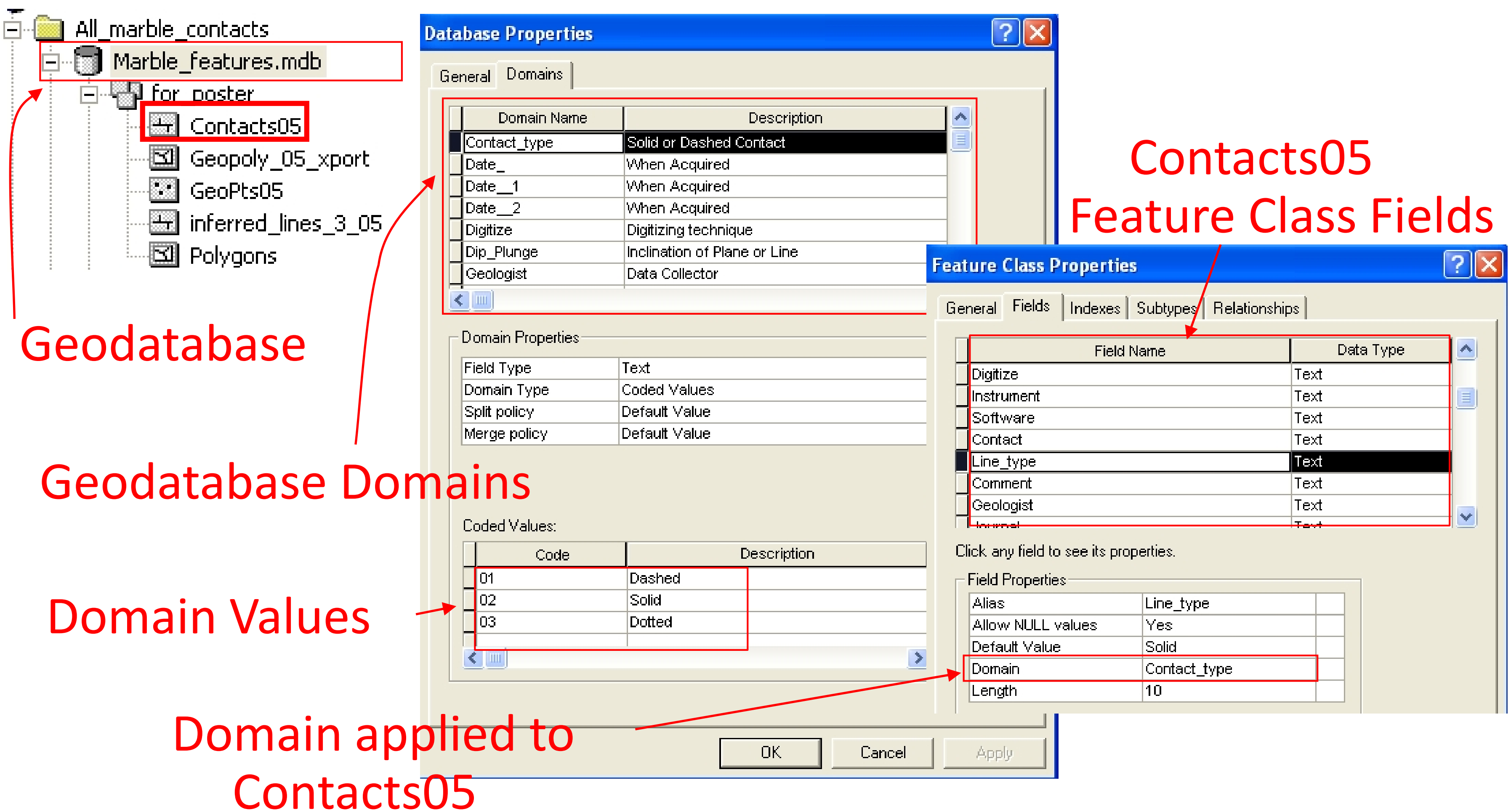
Accuracy

- 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* - catalog of attributes with their permitted values and ranges ("Domains").
  - *Validation Rules* - ensure data integrity.

Data Dictionary: Domains

- Permitted attribute values or range of values for a field:
  - E.g. dip of bedding: permissible range from 0-90^{o}
  - E.g. type of geologic contact: permissibly covered, inferred, exposed
  - E.g. rock type: permissibly sandstone, shale, limestone

Domains in a Geodatabase



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?
- Security – access and use can't corrupt data
- Easy updating

Efficiency (± Flexibility)

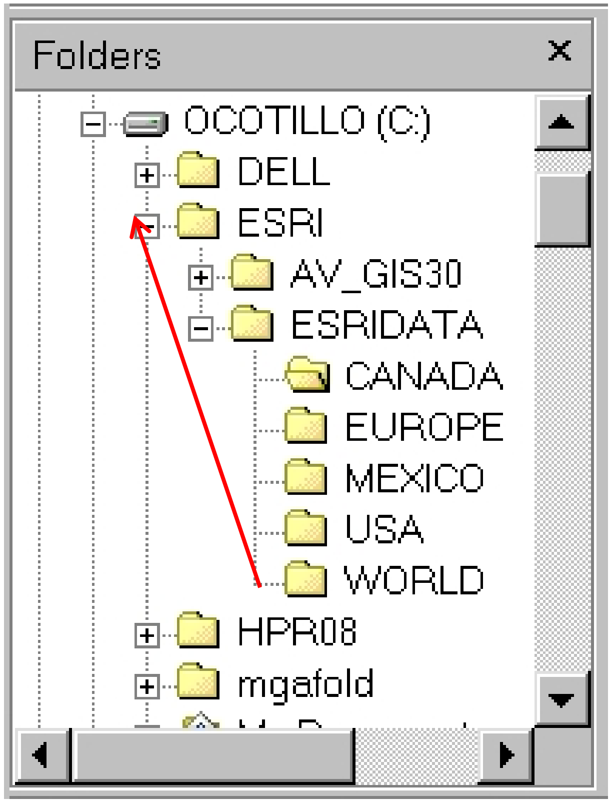
- Relies on database structure (data model):
  - Hierarchical
  - Network
  - Relational
  - Object-oriented

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

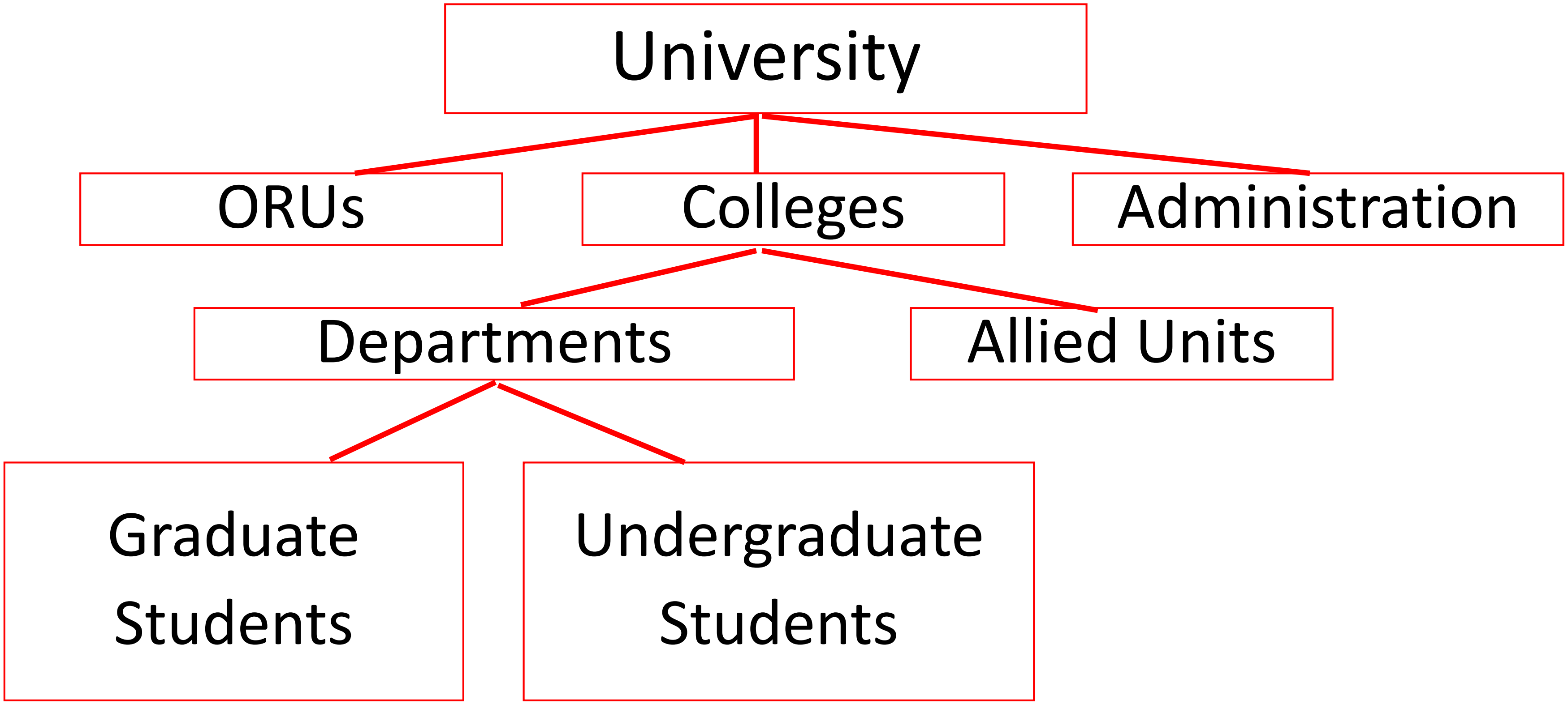
E.g. Filing cabinet or folders on a hard drive



- 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 many times in 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 do we do with a oil 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, relatively small, tables
  - Easy update, editing, searching without affecting or using all data
- Flexibility
  - Using “key” fields, 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, fields) 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

Record or tuple

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

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 ("relation")

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

File ("relation")

- "Key" Field = ID

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

File ("relation")

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 table* (drilling record) to create a “View” – virtual combination

Result of One-to-One Table "Join"

- Joined Production and Drilling Record tables:

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

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 record from source table joined to many records of destination table

One-to-Many Join Result

Table of Wells in Katy Field

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			

- Note that "Discovered" and "Total_Oil" fields in joined table pertain to Katy Field, not to individual wells! This could be a problem…

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

Attributes of maplayer

OID	LAYER	COLO	LTYPE	W	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

Show: All Selected Records (0 out of 503 Selected.) Options

Attributes of Condrey_HY

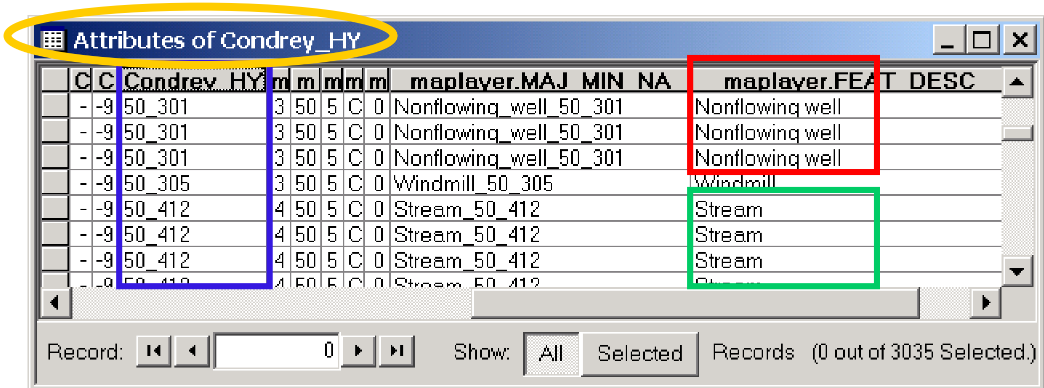
TPMINOR_1	TPMAJOR_2	TPMINOR_2	TPMAJOR_3	TPMINOR_3	Code
300	-999	-999	-999	-999	50_300
300	-999	-999	-999	-999	50_300
301	-999	-999	-999	-999	50_301
300	-999	-999	-999	-999	50_300
412	-999	-999	-999	-999	50_412
412	-999	-999	-999	-999	50_412
412	-999	-999	-999	-999	50_412
412	-999	-999	-999	-999	50_412

Record: 1 Show: All Selected Records (0 out of 3035 Selected.)

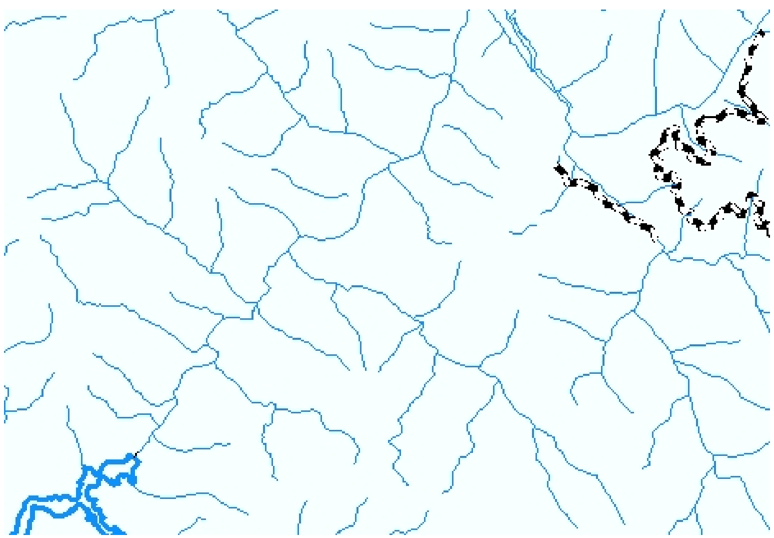
Hydrography feature attributes

Result of Many-to-Many Join

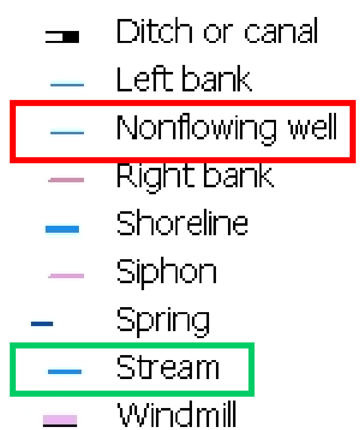
- Symbolize on joined field (FEAT_DESC)



C	C	Condrey_HY	m	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



Layer = Condrey_HY



A DBMS provides:

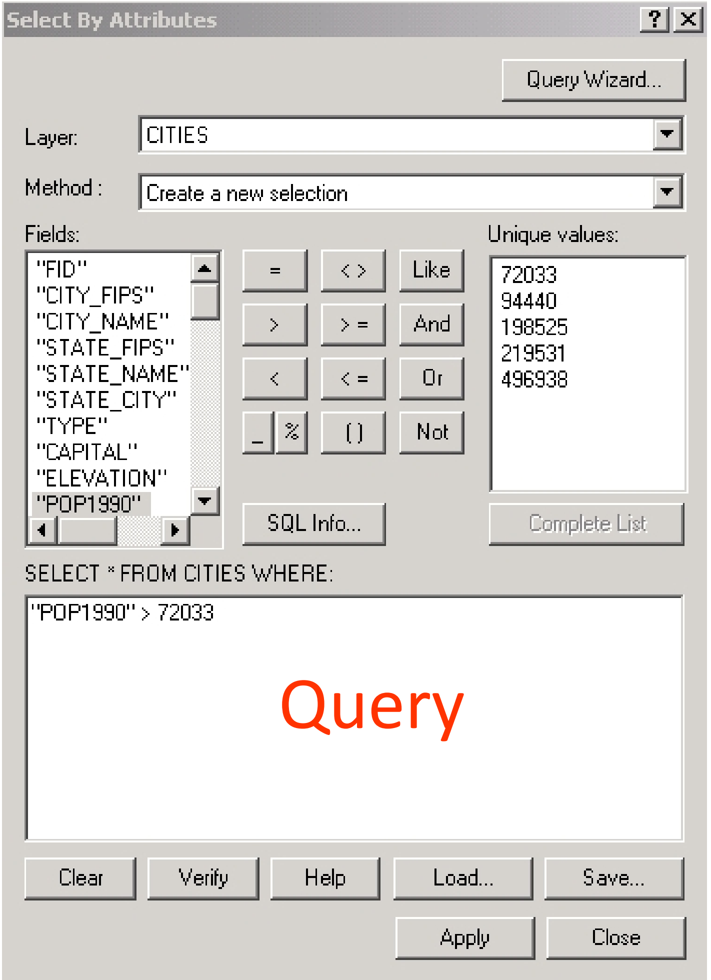
- ✓ Accuracy - reduce errors during entry by use of established rules, templates
- ✓ Efficiency - rapid access & retrieval, no redundancy
- Flexibility - robust structure for query – e.g. What is where?
- Security – access and use can't corrupt data
- Easy updating

Flexibility

- Using primary key(s), can extract and assemble records and attributes to form new tables, as discussed
- 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

Unsorted

Orientation of bedding

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

Sorted by "Azimuth"

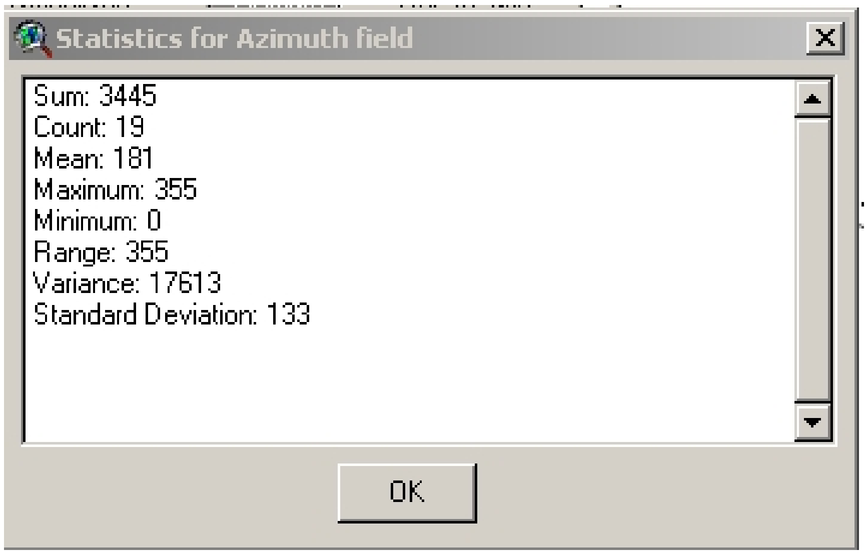
Orientation of bedding

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

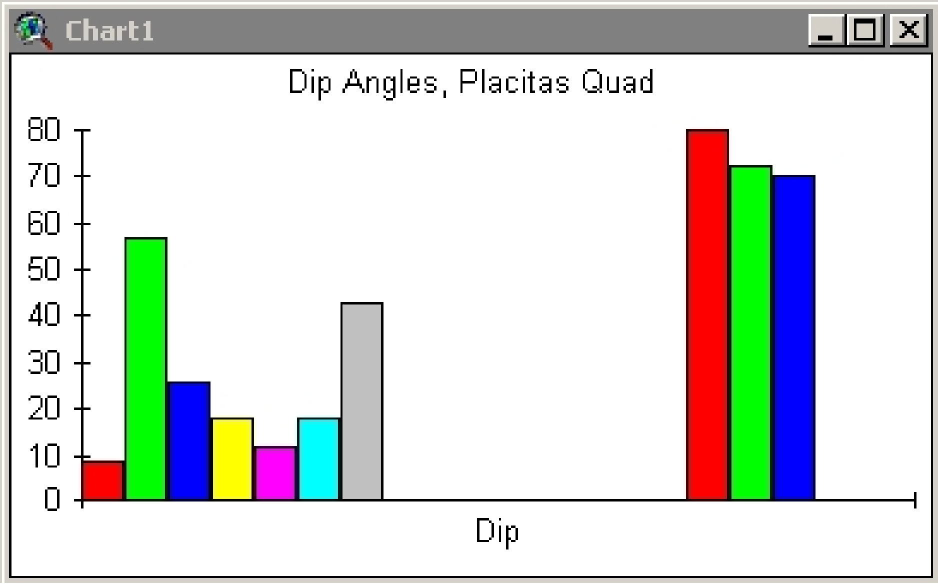
Field Statistics In ArcMap

- Get stats. & graphs on selected attributes

Statistics for "Azimuth"



Histogram for "Dip"



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