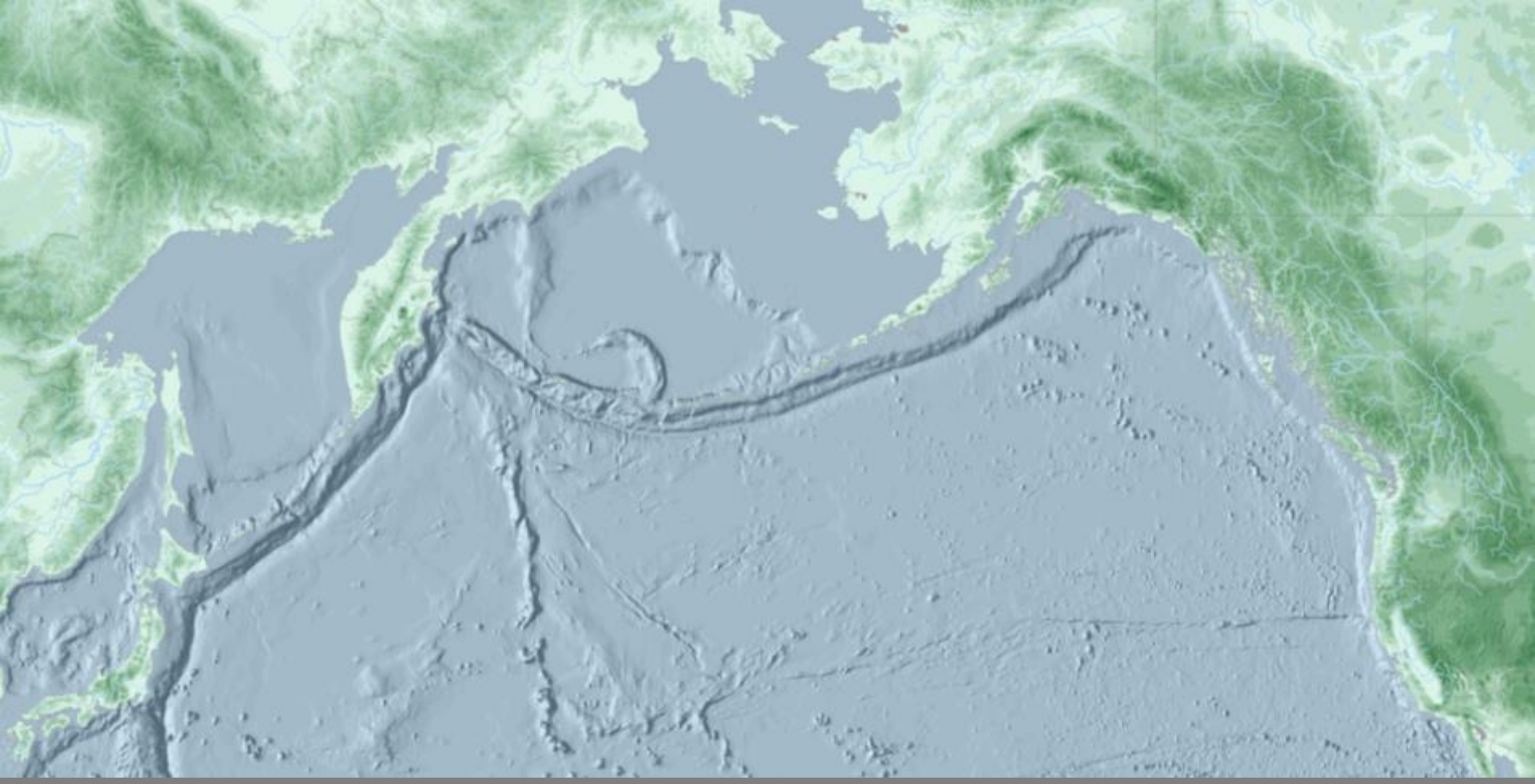


GEODATABASES

GIS Analysis| Winter 2016

Databases are special

What is a database?

- A tool for collecting and organizing information
- An organized collection of data, created within a Database Management System (DBMS)

Database Management System (DBMS)

- System or software program(s) that enables you to store, modify and extract information from a database
- Stores properties of geographic objects and the relationships among the objects
- Examples of DBMS include:
  - Microsoft Access
  - MySQL or SQL Server
  - Postgre SQL
  - Oracle

Why a database?

Can't I just use an excel spreadsheet??

- 'Flat File', 2D (rows, columns) has limitations

name	surname	address	phone #	order #	item	qty	item	qty	item	qty	item	qty
Leo	Durocher	112 Beal St	5-1307	1	CR7	1						
Rudy	Valentini	1 Hispanola Dr	4-2706	2	F15	1						
Paul	Smith	99 Upstate Ln	0-0000	3	GTO	3	F15	1	B52	1	SR71	1
Adam	Smith	1 Wall St	1-2334	4	626	1						
Atom	Ant	685 Hanbar Rd	4-1222	5	B52	2	CR7	2				
William	Smith	202 Dinkytown	9-9199	6	F111	2						
Alice	Paul	5 Free St.	4-4178	7	SR71	1						
Paul	Smith	99 Upstate Ln	0-0000	8	F15	1						

Why a database?

- Limitations of spreadsheet
  - Limit the number of items per order or have an unlimited number of columns
  - Many columns in the file will be empty

name	surname	address	phone #	order #	item	qty	item	qty	item	qty	item	qty
Leo	Durocher	112 Beal St	5-1307	1	CR7	1						
Rudy	Valentini	1 Hispanola Dr	4-2706	2	F15	1						
Paul	Smith	99 Upstate Ln	0-0000	3	GTO	3	F15	1	B52	1	SR71	1
Adam	Smith	1 Wall St	1-2334	4	626	1						
Atom	Ant	685 Hanbar Rd	4-1222	5	B52	2	CR7	2				
William	Smith	202 Dinkytown	9-9199	6	F111	2						
Alice	Paul	5 Free St.	4-4178	7	SR71	1						
Paul	Smith	99 Upstate Ln	0-0000	8	F15	1						

Why a database?

- Limitations of spreadsheet
  - Two orders from Paul Smith, which means there is redundant information for the same customer
  - Waste of space, and potential editing errors

name	surname	address	phone #	order #	item	qty	item	qty	item	qty	item	qty
Leo	Durocher	112 Beal St	5-1307	1	CR7	1						
Rudy	Valentini	1 Hispanola Dr	4-2706	2	F15	1						
Paul	Smith	99 Upstate Ln	0-0000	3	GTO	3	F15	1	B52	1	SR71	1
Adam	Smith	1 Wall St	1-2334	4	626	1						
Atom	Ant	685 Hanbar Rd	4-1222	5	B52	2	CR7	2				
William	Smith	202 Dinkytown	9-9199	6	F111	2						
Alice	Paul	5 Free St.	4-4178	7	SR71	1						
Paul	Smith	99 Upstate Ln	0-0000	8	F15	1						

Why a database?

- ***Data independence*** – allows us to make changes in the database structure in ways that are transparent and does not require a user or programmer to change their process
- ***Multiple views*** – profiles or forms can be developed that change the way data are viewed by individual user or program
- ***Centralized control & maintenance*** – one standard 'copy' of the data may be maintained and updated on a regular basis

Why a database?

- How can we improve the structure of this file by using a database?

name	surname	address	phone #	order #	item	qty	item	qty	item	qty	item	qty
Leo	Durocher	112 Beal St	5-1307	1	CR7	1						
Rudy	Valentini	1 Hispanola Dr	4-2706	2	F15	1						
Paul	Smith	99 Upstate Ln	0-0000	3	GTO	3	F15	1	B52	1	SR71	1
Adam	Smith	1 Wall St	1-2334	4	626	1						
Atom	Ant	685 Hanbar Rd	4-1222	5	B52	2	CR7	2				
William	Smith	202 Dinkytown	9-9199	6	F111	2						
Alice	Paul	5 Free St.	4-4178	7	SR71	1						
Paul	Smith	99 Upstate Ln	0-0000	8	F15	1						

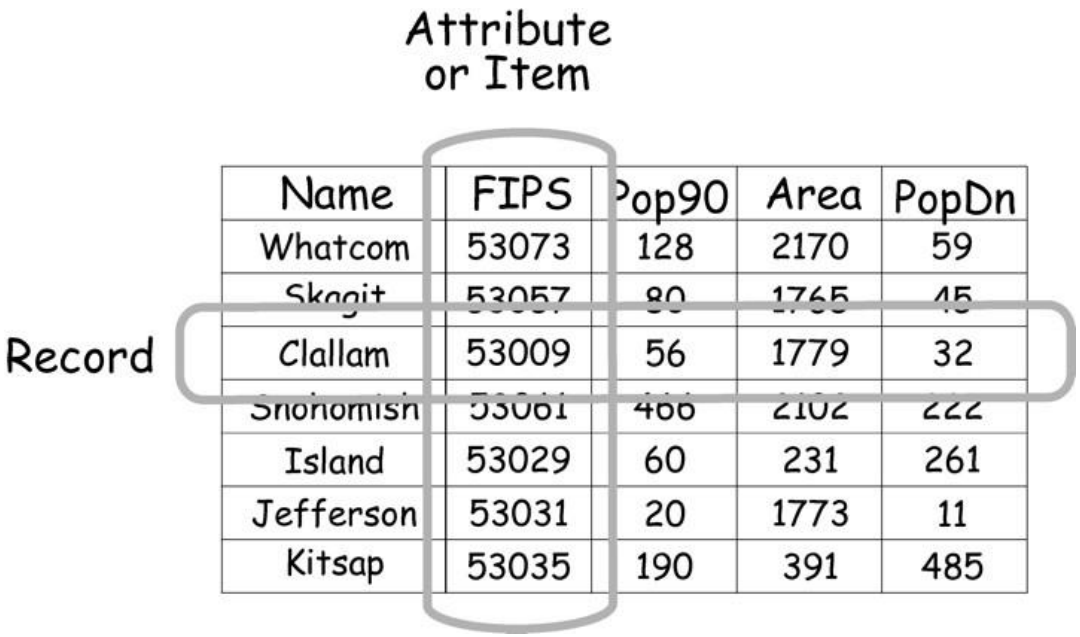
Components of a database

Components of a database

- Attributes
  - Suitable characteristics used to describe the feature
- Type
  - Real numbers, integers, text, hyperlinks, blobs
- Domain
  - Restricts acceptable values for an attribute, i.e. building type restricted to concrete, wood, stucco, brick
- Entity
  - A collection of related data items that are treated as a unit, i.e. streets, or counties, or bus stops

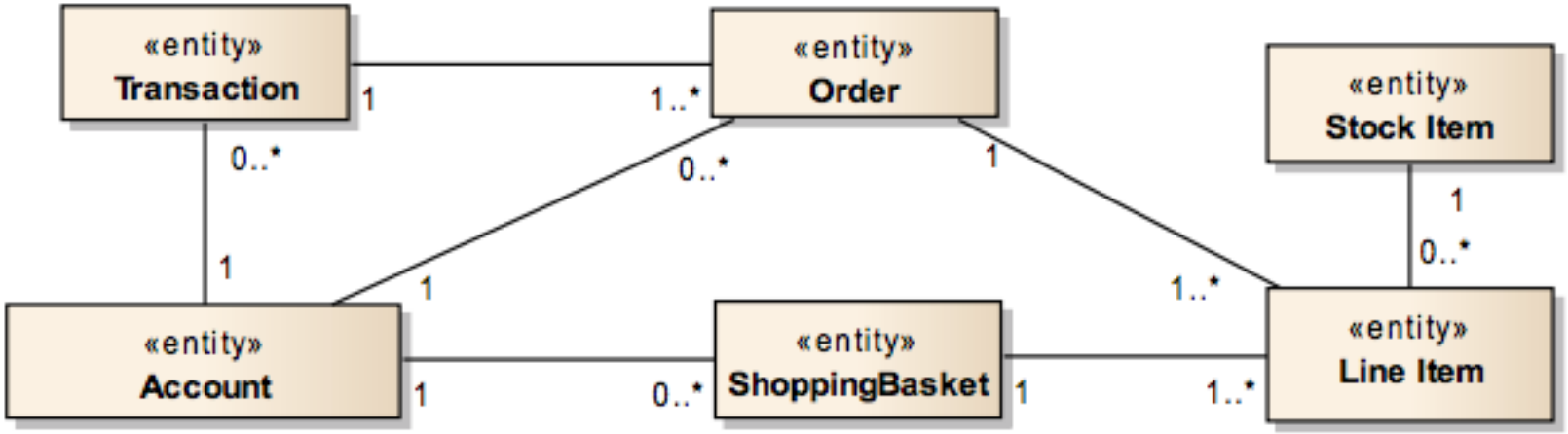
Table structure

- **Record:** a row in a database; represents one feature (a.k.a. "tuple")
- **Attribute:** a column in a database; contains attribute *values* (a.k.a. "field")



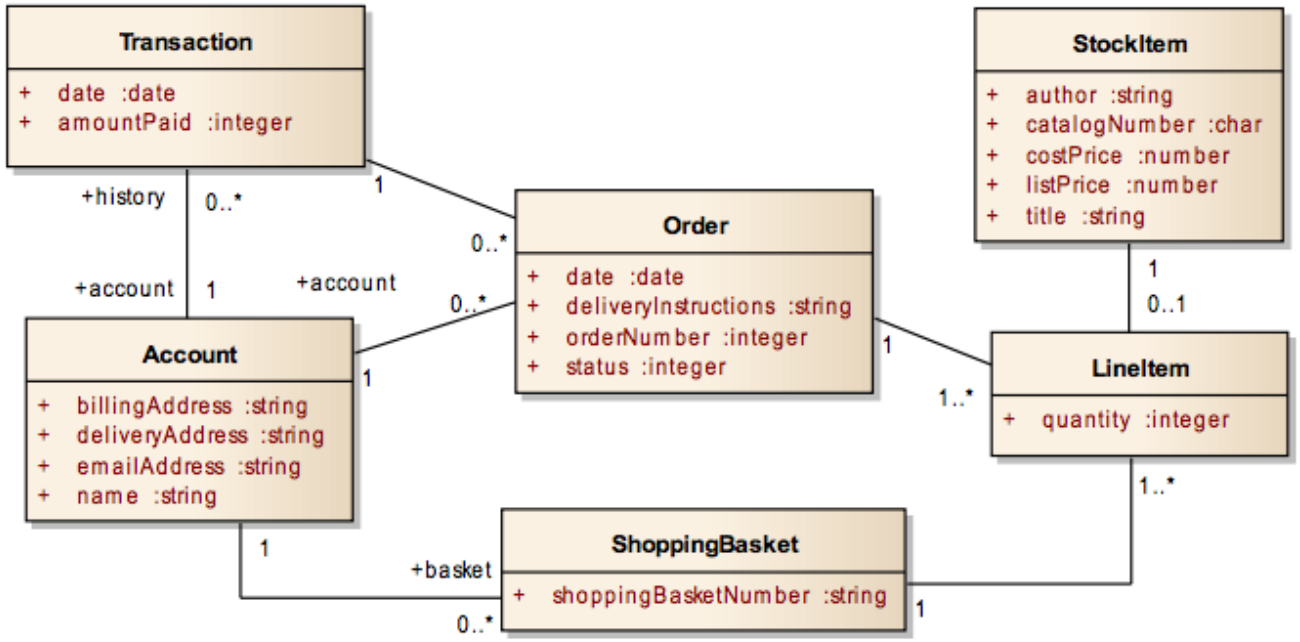
Conceptual structure

- Schema. Compact graphical representation of the conceptual model, the entities, and the relationships among them



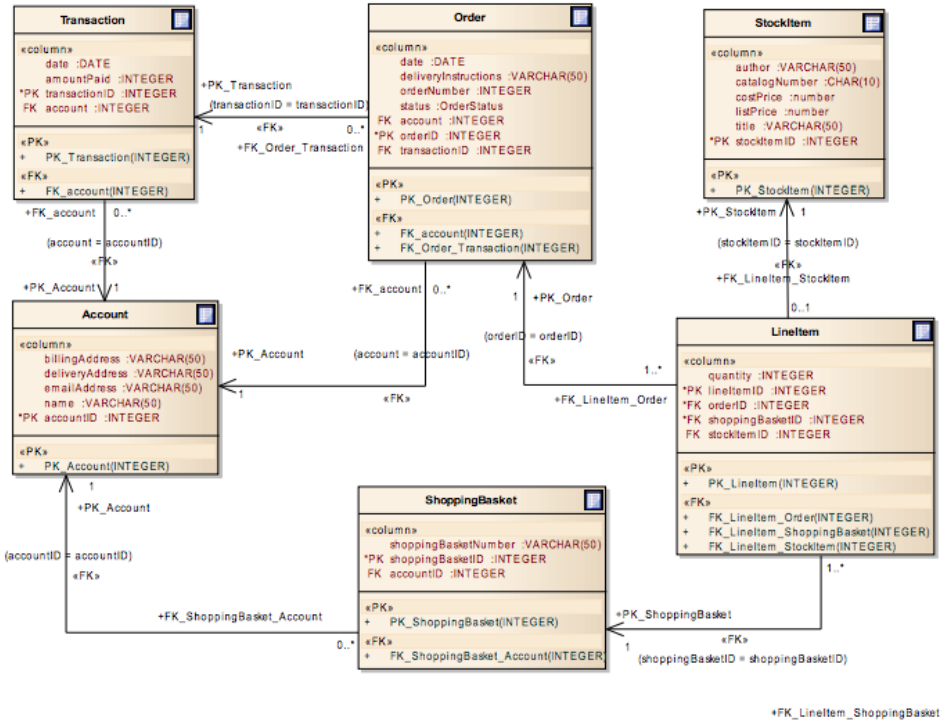
Logical structure

- Constrains the way a conceptual model may be implemented to define the structure and interaction of database components – represents the user's view of interrelationships between data



Physical structure

- Physically cluster or link data used together in processes so that they can be performed more quickly and efficiently.
- Actual physical location of different parts of the database within the file system of the computer



Relational databases

Relational Database (RDBMS)

- Most common database design. Why?
  - Flexible, does not restrict processing or queries
  - Easy to understand, learn, and implement (relative)
  - Wide range of data types
  - Not necessary to know what queries, sorting, that will be performed ahead of time
- Can ***relate tables*** through ***primary keys***

Relational database

Forests

Forest Name	Forest-ID	Location	Size
Nantahala	1	N. Carolina	184,447
Cherokee	2	N. Carolina	92,271

Trails

Trail Name	Forest-ID
Bryson's Knob	1
Slickrock Falls	2
North Fork	1
Cade's Cove	1
Cade's Cove	2
Appalachian	1
Appalachian	2

Characteristics

Trail Name	Feature	Difficulty
Bryson's Knob	Vista	E,M
Bryson's Knob	Ogrth	E,M
Slickrock Falls	Ogrth	M
Slickrock Falls	Wfall	M
North Fork	-	M
Cade's Cove	Ogrth	E
Cade's Cove	Wlife	E
Appalachian	Wfall	M,D
Appalachian	Ogrth	M,D
Appalachian	Vista	M,D
Appalachian	Wlife	M,D
Appalachian	Cmp	M,D

Recreational features

Feature	Description	Activity1	Activity2
Wfall	Waterfall	Photography	Swimming
Ogrth	Old-Growth Forest	Photography	Hiking
Vista	Scenic Overlook	Photography	Viewing
Wlife	Wildlife Viewing	Photography	Birding
Cmp	Camping	Camping	-

Relational database

Forests

Forest Name	Forest-ID	Location	Size
Nantahala	1	N. Carolina	184,447
Cherokee	2	N. Carolina	92,271

Trails

Trail Name	Forest-ID
Bryson's Knob	1
Slickrock Falls	2
North Fork	1
Cade's Cove	1
Cade's Cove	2
Appalachian	1
Appalachian	2

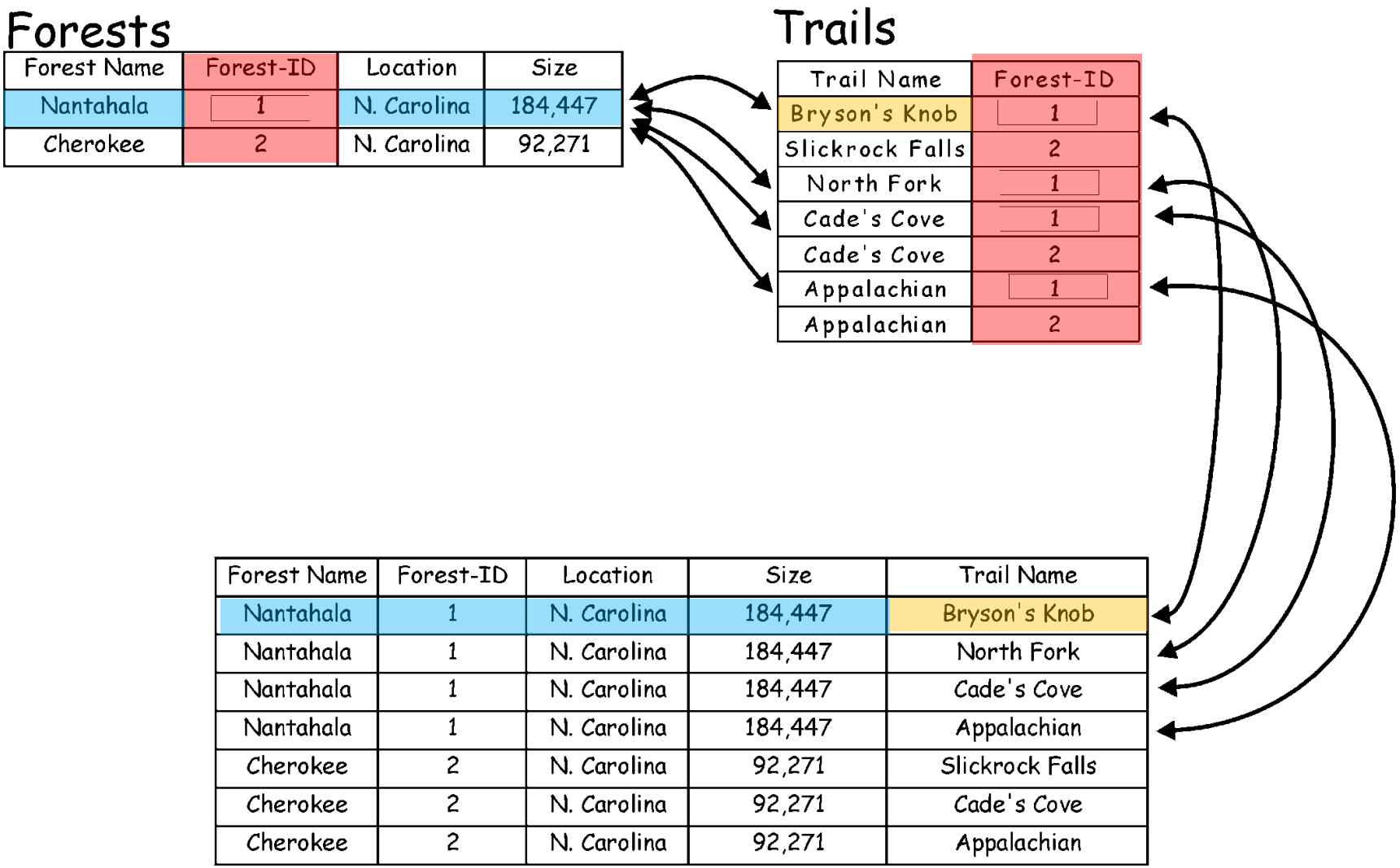
Characteristics

Trail Name	Feature	Difficulty
Bryson's Knob	Vista	E,M
Bryson's Knob	Ogrth	E,M
Slickrock Falls	Ogrth	M
Slickrock Falls	Wfall	M
North Fork	-	M
Cade's Cove	Ogrth	E
Cade's Cove	Wlife	E
Appalachian	Wfall	M,D
Appalachian	Ogrth	M,D
Appalachian	Vista	M,D
Appalachian	Wlife	M,D
Appalachian	Cmp	M,D

Recreational features

Feature	Description	Activity1	Activity2
Wfall	Waterfall	Photography	Swimming
Ogrth	Old-Growth Forest	Photography	Hiking
Vista	Scenic Overlook	Photography	Viewing
Wlife	Wildlife Viewing	Photography	Birding
Cmp	Camping	Camping	-

Relational databases



Object-relational databases

- Object models aim to address data objects in a more natural way
  - Include information and operations ('methods') into discrete objects
- Approach is essentially the same as Relational databases
  - Data resides in the database and is manipulated collectively with queries
- Bridge the gap between relational databases & object-oriented modeling techniques

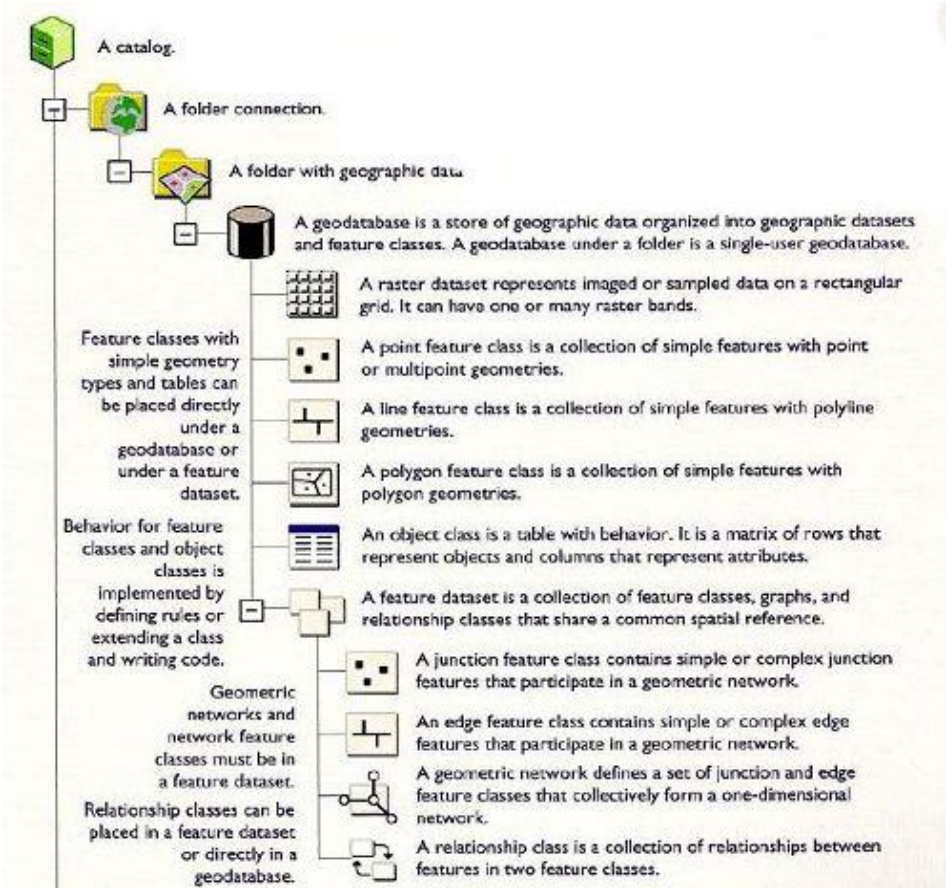
Geodatabases are *really* special

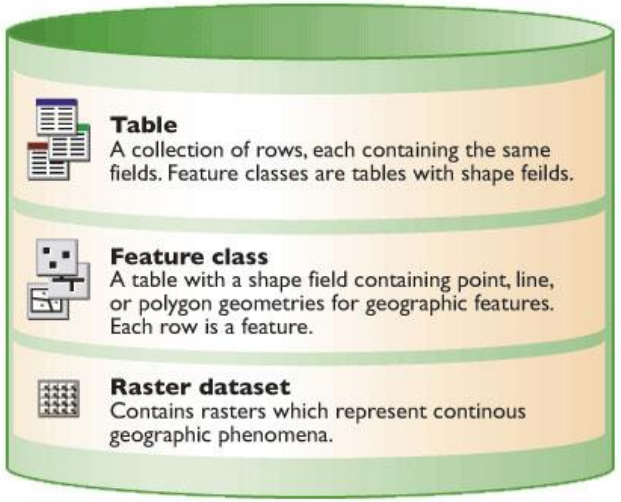
Geodatabase as multiple meanings in GIS

1. Native data structure for GIS & primary data format used
2. Physical store of data, using a DMBS or file system
3. Comprehensive information model for representing & managing geographic data
4. Software logic provides the common application logic used throughout ArcGIS for accessing & working with a variety of file formats
5. Transaction model for managing GIS data workflows

ESRI Geodatabase

Object-relational database management system





Geodatabase Components

- Features datasets
- Subtypes
- Topology
- Networks
- Raster Datasets
- Relationships
- Survey Data
- Label Annotation
- Attribute Domains

Geodatabase Components

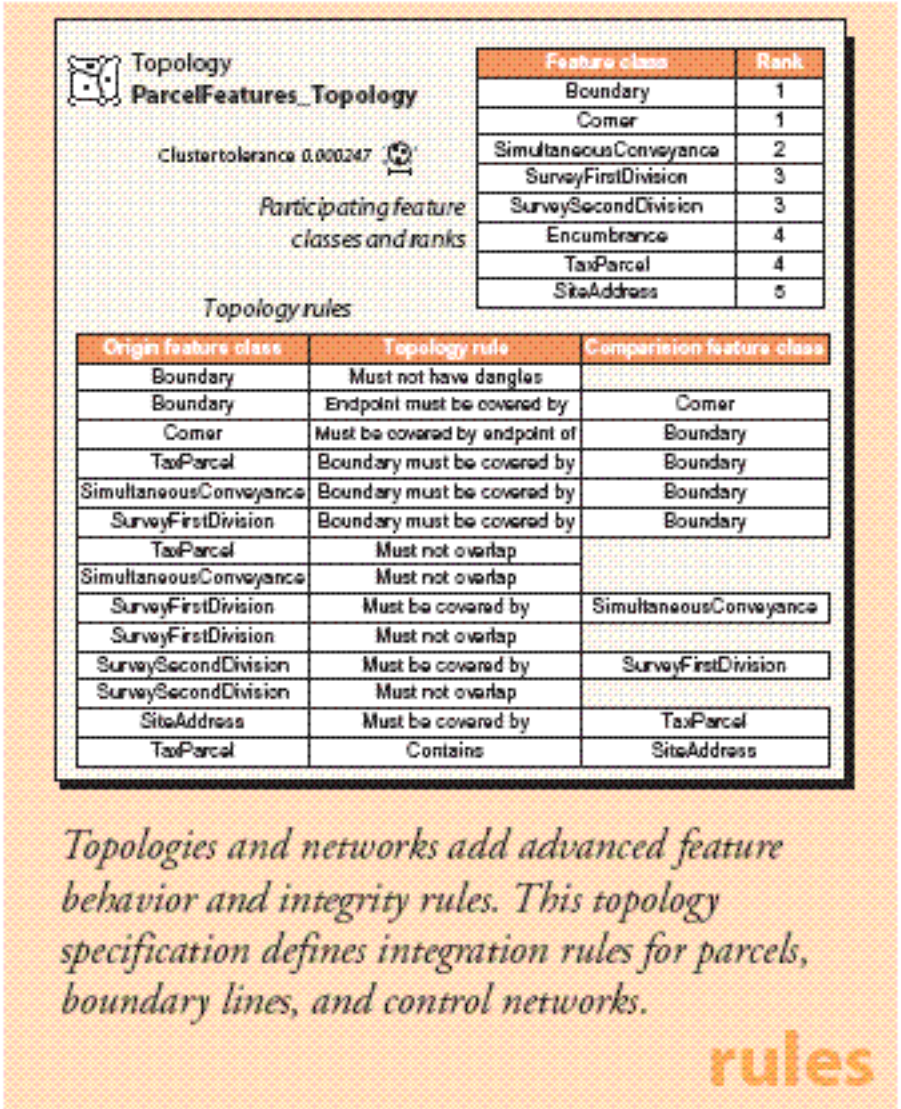
Feature datasets. A collection of related feature classes that share a common coordinate system. Allows for spatial or thematic integration of related feature classes. Common purpose is to organize related feature classes to build topology, network dataset, or terrain dataset.

Subtypes. Manage a set of attribute subclasses in a single table. This is often used to manage different behaviors on subsets of the same features.

Geodatabase Components

Topology.

Defining topologies to enforce valid connectivity between features. The collection of rules established enable the geodatabase to more accurately model geometric relationships.



Feature class	Rank
Boundary	1
Corner	1
SimultaneousConveyance	2
SurveyFirstDivision	3
SurveySecondDivision	3
Encumbrance	4
TaxParcel	4
SiteAddress	5

Origin feature class	Topology rule	Comparision feature class
Boundary	Must not have dangles	
Boundary	Endpoint must be covered by	Corner
Corner	Must be covered by endpoint of	Boundary
TaxParcel	Boundary must be covered by	Boundary
SimultaneousConveyance	Boundary must be covered by	Boundary
SurveyFirstDivision	Boundary must be covered by	Boundary
TaxParcel	Must not overlap	
SimultaneousConveyance	Must not overlap	
SurveyFirstDivision	Must be covered by	SimultaneousConveyance
SurveyFirstDivision	Must not overlap	
SurveySecondDivision	Must be covered by	SurveyFirstDivision
SurveySecondDivision	Must not overlap	
SiteAddress	Must be covered by	TaxParcel
TaxParcel	Contains	SiteAddress

Geodatabase Components

Networks. Making geometric networks for tracing and network analysis, commonly created for modeling transportation networks.

Survey data. Integrating survey projects with vertices and points in survey-aware features classes

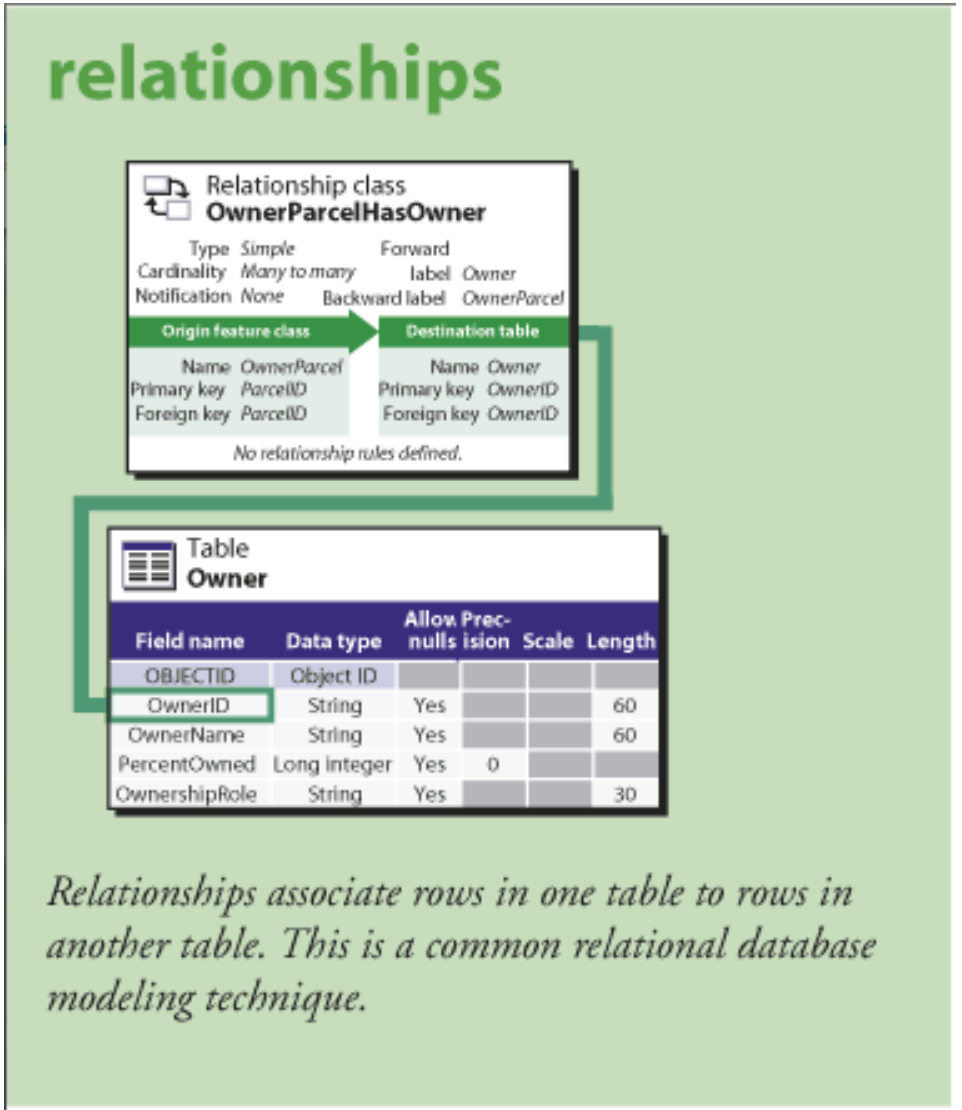
Raster data. Using catalogs for time-related data and mosaics for base maps

Geodatabase Components

Relationships.

Modeling how features and objects are related to each other. Build relationships between two tables using a common key.

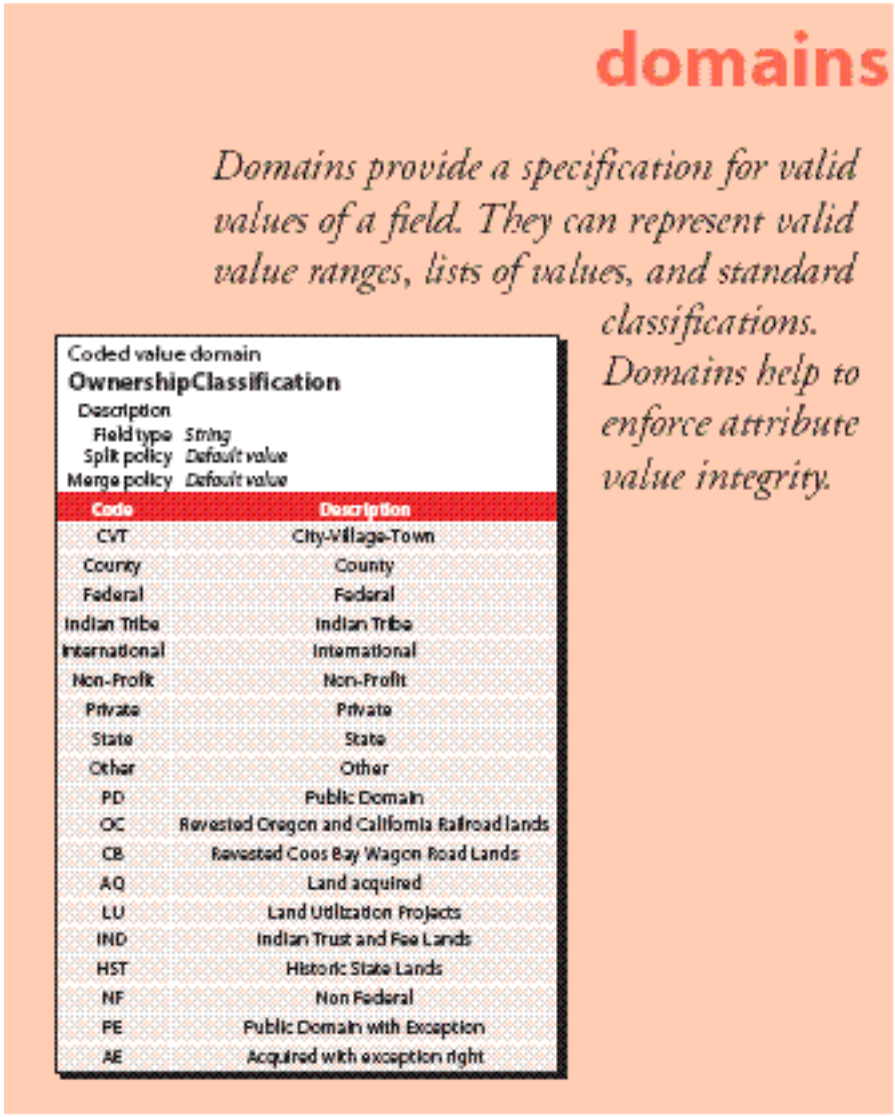
Example: a building can be associated with a parcel



Geodatabase Components

Attribute Domains.

Rules that describe the valid values for a field type, providing a method for enforcing data integrity. It is a declaration of acceptable attribute values.



Geodatabase Components

Label Annotation. Labels that are stored as a feature class. All features (labels) have a geographic location and attributes. Annotations can be text or a graphic shape.

Type of Geodatabases

- Personal
  - Single User
  - Microsoft Access
  - Up to 2GB storage
- File
  - Single user
  - File folder structure
  - Up to 1TB storage
- ArcSDE
  - Multi-user
  - Support versioning
  - Oracle, Microsoft SQL Server, IBM DB2, IBM Informix, PostgreSQL
  - Storage limit based on DBMS type