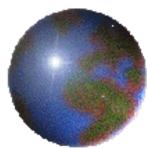




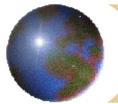
CSEG601 & CSE5601: Spatial Data Management & Application

Sungwon Jung

Bigdata Processing & DB Lab
Dept. of Computer Science and Engineering
Sogang University
Seoul, Korea
Tel: +82-2-705-8930
Email : jungsung@sogang.ac.kr



Chapter 2: Spatial Concepts and Data Models



What is a Data Model?

- What is a model? (Dictionary meaning)
 - A set of plans (blueprint drawing) for a building
 - A miniature representation of a system to analyze properties of interest
- What is Data Model?
 - Specify structure or schema of a data set
 - Document description of data
 - Facilitates early analysis of some properties, e.g. querying ability, redundancy, consistency, storage space requirements, etc.
- Examples:
 - GIS organize spatial set as a set of layers
 - Databases organize dataset as a collection of tables

3

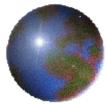


Why Data Models?

- Data models facilitate
 - Early analysis of properties, e.g. storage cost, querying ability, ...
 - Reuse of shared data among multiple applications
 - Exchange of data across organization
 - Conversion of data to new software / environment
- Example- Y2K crisis for year 2000

Many computer software systems were developed without well-defined data models in 1960s and 1970s. These systems used a variety of data models for representing time and date. Some of the representations used two digits to represent years. In late 1990s, people worried that the 2 digit representation of year may lead to erroneous behaviour. For example age of a person born in 1960 (represented as 60) in year 2000 (represented as 00) may appear negative and may be flagged as illegal data item. A large amount of effort and resources (hundreds of Billions of dollars) was spent in revising the software. Proper use of data model may have significantly reduced the costs. If time and date were modeled as abstract data types in a software, only a small portion of the software implementing the date ADT had to be reviewed and revised.

4



Types of Data Models

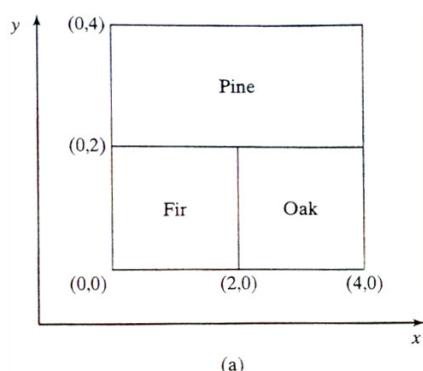
- Two Types of data models
 - Generic data models
 - Developed for business data processing
 - Support simple abstract data types (ADTs), e.g. numbers, strings, date
 - Not convenient for spatial ADTs, e.g. polygons
 - Recall a polygon becomes dozens of rows in 3 tables (Fig. 1.4, pp. 8)
 - Need to extend with spatial concepts, e.g. ADTs
 - Application Domain specific, e.g. spatial models
 - Set of concepts developed in Geographic Info. Science
 - Common spatial ADTs across different GIS applications
- Plan of Study
 - First study concepts in spatial models
 - Then study generic model
 - Finally put the two together

5



2.1 Models of Spatial Information

- Two common models
 - Field based
 - Object based
- Example: Forest stands (Fig. 2.1)
 - (a) forest stand map
 - (b) Object view has 3 polygons
 - (c) Field view has a function



(a)

Object Viewpoint of Forest Stands		
Area-ID	Dominant Tree Species	Area/Boundary
FS1	Pine	[(0,2),(4,2),(4,4),(0,4)]
FS2	Fir	[(0,0),(2,0),(2,2),(0,2)]
FS3	Oak	[(2,0),(4,0),(4,2),(2,2)]

(b)

Field Viewpoint of Forest Stands

$$f(x,y) = \begin{cases} \text{"Pine," } 2 \leq x \leq 4; 2 < y \leq 4 \\ \text{"Fir," } 0 \leq x \leq 2; 0 \leq y \leq 2 \\ \text{"Oak," } 2 < x \leq 4; 0 \leq y \leq 2 \end{cases}$$

(c)

6



2.1 Object Model

❖ Object model concepts

- ❖ Objects: distinct identifiable things relevant to an application
- ❖ Objects have attributes and operations
- ❖ Attribute: a simple (e.g. numeric, string) property of an object
- ❖ Operations: function maps object attributes to other objects

❖ Example from a roadmap

- ❖ Objects: roads, landmarks, ...
- ❖ Attributes of road objects:
 - spatial: location, e.g. polygon boundary of land-parcel
 - non-spatial: name (e.g. Route 66), type (e.g. interstate, residential street), number of lanes, speed limit, ...
- ❖ Operations on road objects: determine center line, determine length, determine intersection with other roads, ...

7

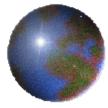


Classifying Spatial objects

- Spatial objects are spatial attributes of general objects
- Spatial objects are of many types
 - Simple
 - 0- dimensional (points), 1 dimensional (curves), 2 dimensional (surfaces)
 - Example given at the bottom of this slide
 - Collections
 - Polygon collection (e.g. boundary of Japan or Hawaii), ...
 - See more complete list in Figure 2.2

Spatial Object Types	Example Object	Dimension
Point	City	0
Curve	River	1
Surface	Country	2

8



Spatial Object Types in OGIS Data Model

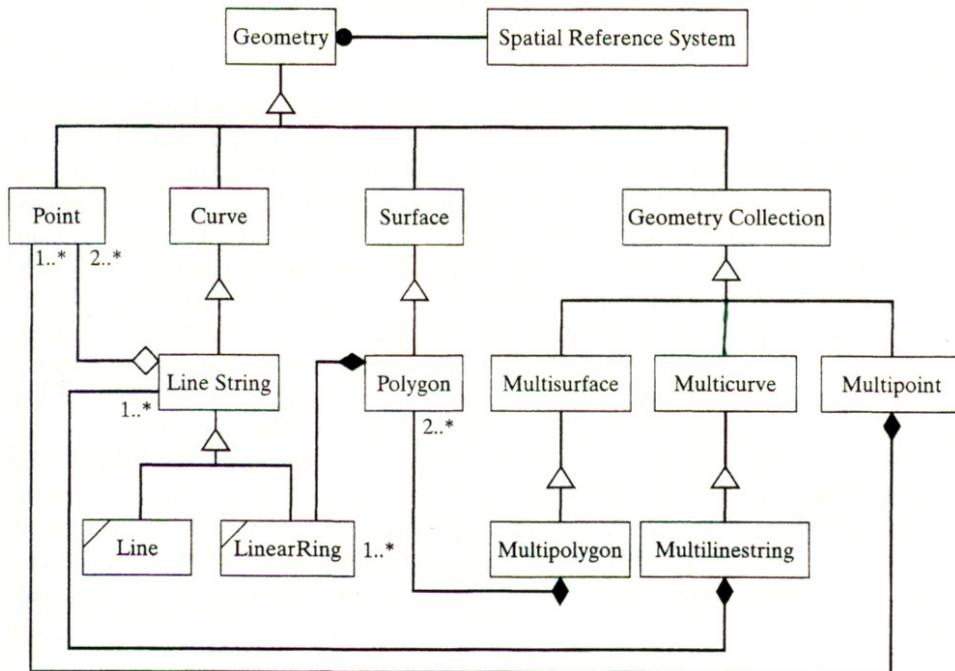


Figure 2.2. Each rectangle shows a distinct spatial object type

9

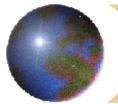


Classifying Operations on spatial objects in Object Model

- Classifying operations
 - Set based: 2-dimensional spatial objects (e.g. polygons) are sets of points
 - a set operation (e.g. intersection) of 2 polygons produce another polygon
 - Topological operations: Boundary of USA touches boundary of Canada
 - Directional: New York city is to east of Chicago
 - Metric: Chicago is about 700 miles from New York city.
 - Metric Properties: A set X is called in a metric space if the following properties are hold
 1. $d(x, y) \geq 0$ and $d(x, x) = 0$
 2. $d(x, y) = d(y, x)$
 3. $d(x, y) \leq d(x, z) + d(z, y)$ for all x, y, z in X

Set theory based	Union, Intersection, Containment,
Toplogical	Touches, Disjoint, Overlap, etc.
Directional	East,North-West, etc.
Metric	Distance

10



Example of Topological and Nontopological Operations

Topological

endpoint(point, arc)	A point at the end of an arc.
simple-nonself-intersection (arc)	A nonself-intersecting arc does not cross itself.
on-boundary (point, region)	Vancouver is on the boundary of Canada and the United States.
inside (point, region)	Minneapolis is inside of Minnesota state.
outside (point, region)	Madison is outside of Minnesota state.
open (region)	Interior of Canada is an open region (excludes all its boundaries).
close (region)	Carleton County is a closed region (includes all its boundaries).
connected (region)	Switzerland is a connected region, whereas Japan is not (given any two points in the area, it is possible to follow a path from one point to the other such that the path is entirely within the area).
inside (point, loop)	A point is within the loop.
crosses (arc, region)	A road (arc) passes through a forest (region).
touches (region, region)	Minnesota (region) is a neighboring state of Wisconsin (region).
touches (arc, region)	Interstate freeway 90 (arc) passes by Lake Michigan (region).
overlap (region, region)	Land cover (region) overlaps with land use (region).

Nontopological

Euclidean-distance (point, point)	Distance between two points.
direction (point, point)	Madison city is east of Minneapolis city.
length (arc)	Length of a unit vector is one unit.
perimeter (area)	Perimeter of a unit square is 4 units.
area (region)	Area of unit square is one square unit.

11



Topological Relationships

❖ Topological Relationships

- invariant under elastic deformation (without tear, merge).
- Two countries which touch each other in a planar paper map will continue to do so in spherical globe maps.

❖ Topology is the study of topological relationships

❖ Example queries with topological operations

- What is the topological relationship between two objects A and B ?
- Find all objects which have a given topological relationship to object A ?

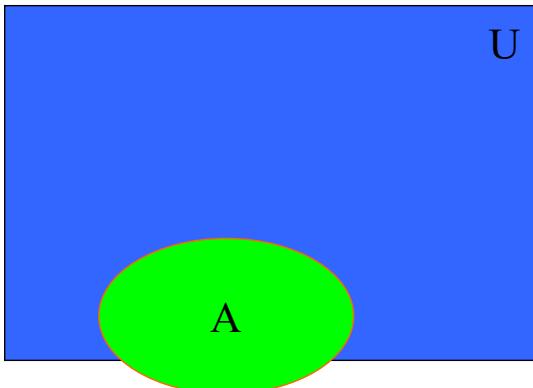
12



Topological Concepts

- ⊕ Interior, boundary, exterior

■ Let A be an object in a "Universe" U.



Green is A interior (A°)

Red is boundary of A (∂A)

Blue –(Green + Red) is
A exterior (A^-)

- ⊕ Question: Define Interior, boundary, exterior on curves and points.

13

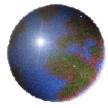


Nine-Intersection Model of Topological Relationships

- Many topological Relationship between A and B can be
 - specified using 9 intersection model
 - Examples on next slide
- Nine intersections
 - intersections between interior, boundary, exterior of A, B
 - A and B are spatial objects in a two dimensional plane.
 - Can be arranged as a 3 by 3 matrix
 - Matrix element take a value of 0 (false) or 1 (true).
- Q? Determine the number of many distinct 3 by 3 boolean matrices .

$$\Gamma_9(A, B) = \begin{pmatrix} A^\circ \cap B^\circ & A^\circ \cap \partial B & A^\circ \cap B^- \\ \partial A \cap B^\circ & \partial A \cap \partial B & \partial A \cap B^- \\ A^- \cap B^\circ & A^- \cap \partial B & A^- \cap B^- \end{pmatrix}$$

14



Specifying topological operation in 9-Intersection Model

Figure 2.3. 9 intersection matrices for a few topological operations

$\begin{pmatrix} o & o & 1 \\ o & o & 1 \\ 1 & 1 & 1 \end{pmatrix}$ disjoint	$\begin{pmatrix} 1 & 1 & 1 \\ o & o & 1 \\ o & o & 1 \end{pmatrix}$ contains	$\begin{pmatrix} 1 & o & o \\ 1 & o & o \\ 1 & 1 & 1 \end{pmatrix}$ inside	$\begin{pmatrix} 1 & o & o \\ o & 1 & o \\ o & o & 1 \end{pmatrix}$ equal
$\begin{pmatrix} o & o & 1 \\ o & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ meet	$\begin{pmatrix} 1 & 1 & 1 \\ o & 1 & 1 \\ o & o & 1 \end{pmatrix}$ covers	$\begin{pmatrix} 1 & o & o \\ 1 & 1 & o \\ 1 & 1 & 1 \end{pmatrix}$ coveredBy	$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ overlap

15



Using Object Model of Spatial Data

❖ Object model of spatial data

- ❖ OGIS standard set of spatial data types and operations
- ❖ Similar to the object model in computer software
- ❖ Easily used with many computer software systems
 - Programming languages like Java, C++, Visual basic
 - Example use in a Java program is in section 2.1.6
 - Post-relational databases, e.g. OODBMS, ORDBMS
 - Example usage in chapter 3 through 6

16



2.2 Three-Step Database Design

- Database applications are modeled using a three-step design process
 - Conceptual-datatypes,relationships and constraints(ER model)
 - Logical-mapping to a Relational model and associated query language(Relational Algebra)
 - Physical-file structures, indexing,
- Scope
 - We discuss conceptual and logical data models in section 2.3
 - Physical model is discussed in chapter 4

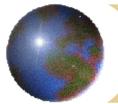
17



Example Application Domain

- Database design is for a specific application domain
 - Often a requirements document is available
 - Designers discuss requirements with end-users as needed
 - We will use a simple spatial application domain
 - to illustrate concepts in conceptual and logical data models
 - to illustrate translation of conceptual DM to logical DM
- Spatial application domain
 - A *state-park* consists of *forests*,
 - A forest is a collection of *forest-stands* of different species
 - State-Park is accessed by *roads* and has a *manager*
 - State-Park has *facilities*
 - *River* runs through state-park and supplies water to the facilities

18



2.2.1 Conceptual DM: The ER Model

➊ 3 basic concepts

- Entities have an independent conceptual or physical existence.
 - Examples: Forest, Road, Manager, ...
- Entities are characterized by Attributes
 - Example: Forest has attributes of name, elevation, etc.
- An Entity interacts with another Entity through relationships.
 - Road allow access to Forest interiors.
 - This relationship may be name "Accesses"

➋ Comparison with Object model of spatial information

- Entities are collections of attributes are like objects
- However ER model does not permit general user defined operations
- Relationships are not directly supported in Object model
 - but may be simulated via operations

19



Relationship Types

➊ Relationships can be categorized by

- cardinality constraints
- other properties, e.g. number of participating entities
 - Binary relationship: two entities participate

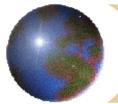
➋ Types of Cardinality constraints for binary relationships

- One-One: An instance of an entity relates to a unique instance of other entity.
- Many-One: Many instances of an entity relate to an instance of an other.
- Many-Many: Many instances of one entity relate to multiple instances of another.

➌ Exercise: Identify type of cardinality constraint for following:

- Many facilities belong to a forest. Each facility belong to one forest.
- A manager manages 1 forest. Each forest has 1 manager.
- A river supplies water to many facilities. A facility gets water from many rivers.

20



ER Diagrams Graphical Notation

- ER Diagrams are graphic representation of ER models
 - Several different graphic notation are used
 - We use a simple notation summarized below
 - Example ER Diagram for Forest example in next slide
- Q? Compare and contrast “Attributes” and “Multi-valued attributes”.

Concept	Symbol
Entities	rectangle
Attributes	oval
Multi-valued Attributes	double-lined oval
Relationships	diamond
Cardinality of Relationship	1:1, M:1, M:N

21



ER Diagram for “State-Park”

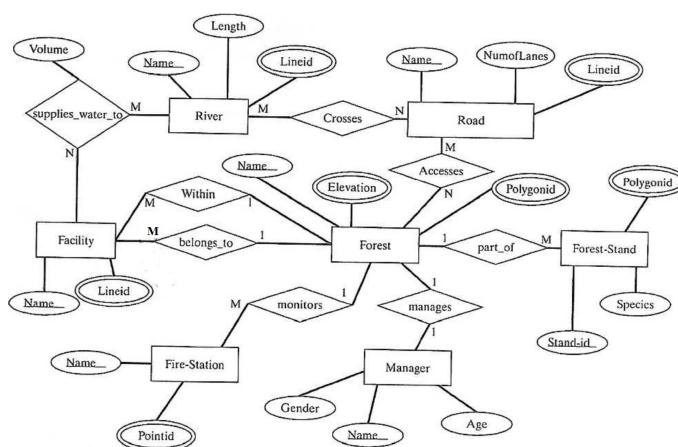
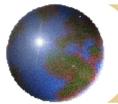


FIGURE 2.4. An ER diagram for the *State-Park* example.

- Exercise:
 - List the entities, attributes, relationships in this ER diagram
 - Identify cardinality constraint for each relationship.
 - How many roads “Accesses” a “Forest_stand”? (one or many)

22



2.2.2 Logical Data Model: The Relational Model

- Relational model is based on set theory
- Main concepts
 - Domain: a set of values for a simple attribute
 - Relation: cross-product of a set of domains
 - Represents a table, i.e. homogeneous collection of rows (tuples)
 - The set of columns (i.e. attributes) are same for each row
- Comparison to concepts in conceptual data model
 - Relations are similar to but not identical to entities
 - Domains are similar to attributes
 - Translation rules establishing exact correspondence are discussed in 2.2.3

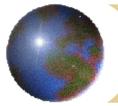
23



Relational Schema

- Schema of a Relation
 - Enumerates columns, identifies primary key and foreign keys.
 - Primary Key :
 - one or more attributes uniquely identify each row within a table
 - Foreign keys
 - R's attributes which form primary key of another relation S
 - Value of a foreign key in any tuple of R match values in some row of S
- Relational schema of a database
 - collection of schemas of all relations in the database
 - Example: Figure 2.5
 - A blue print summary drawing of the database table structures
 - Allows analysis of storage costs, data redundancy, querying capabilities
 - Some databases were designed as relational schema in 1980s
 - Nowadays, databases are designed as E R models and relational schema is generated via CASE tools

24



More on Relational Model

⊕ Integrity Constraints

- Key: Every relation has a primary key.
- Entity Integrity: Value of primary key in a row is never undefined
- Referential Integrity: Value of an attribute of a Foreign Key must appear as a value in the primary key of another relationship or must be null.

⊕ Normal Forms (NF) for Relational schema

- Reduce data redundancy and facilitate querying
- 1st NF: Each column in a relation contains an atomic value.
- 2nd and 3rd NF: Values of non-key attributes are fully determined by the values of the primary key, only the primary key, and nothing but the primary key.
- Other normal forms exists but are seldom used
- Translating a well-designed ER model yields a relational schema in 3rd NF
 - satisfying definition of 1st, 2nd and 3rd normal forms

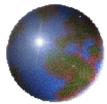
25



2.2.3 Mapping ER to Relational

- Highlights of transaltion rules (section 2.2.3)
 - Entity becomes Relation
 - Attributes become columns in the relation
 - Multi-valued attributes become a new relation
 - includes foreign key to link to relation for the entity
 - Relationships (1:1, 1:N) become foreign keys
 - M:N Relationships become a relation
 - containing foreign keys or relations from participating entities
- Example and Exercise
 - Compare Fig. 2.4 and Fig. 2.5
 - Identify the relational schema components for
 - entity Facility, its attributes and its relationships
 - Note an empty relation box in Fig. 2.5. Fill in its schema.

26



Relational Schema Example

- Identify relations with primary key and foreign keys

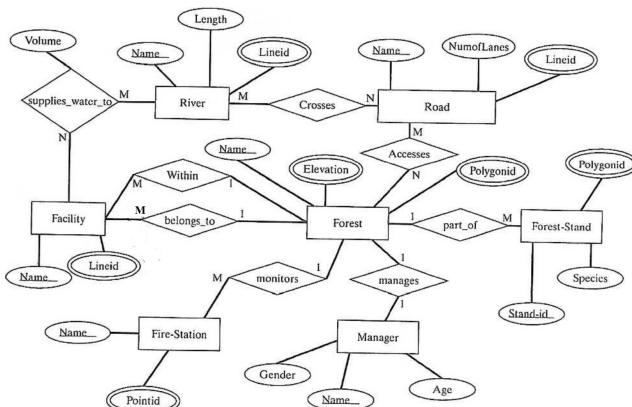


FIGURE 2.4. An ER diagram for the *State-Park* example.

Forest-Stand	Fstand-Geom
Stand-id (Integer)	Polygonid (Integer)
Species (varchar)	
Forest-name (varchar)	
River	River-Geom
Name (varchar)	Name (Integer)
Length (Real)	Lineid (Integer)
Road	Road-Geom
Name (varchar)	Rname (varchar)
NumofLanes (Integer)	Lineid (Integer)
Facility	Facility-Geom
Name (varchar)	Name (varchar)
Forest-name (varchar)	Pointid (Integer)
Forest-name-2 (varchar)	
Forest	Forest-Geom
Name (varchar)	Name (varchar)
Fire-Station	Firestation-Geom
Name (varchar)	Name (varchar)
ForName (varchar)	Poindid (Integer)
Supplies_Water_To	
FacName (varchar)	
RivName (varchar)	
Volume (Real)	
Road-Access-Forest	
RoadName (varchar)	
ForName (varchar)	
Manager	
Name (varchar)	
Age (Integer)	
Gender (varchar)	
ForName (varchar)	

Figure 2.5. Relational schema for the State-Park example.

27



Relational Schema for “Point”, “Line”, “Polygon” and “Elevation”

- Relational model restricts attribute domains
 - simple atomic values, e.g. a number
 - Disallows complex values (e.g. polygons) for columns
 - Complex values need to be decomposed into simpler domains
 - A polygon may be decomposed into edges and vertices (Fig. 2.5)

Polygon	Line	
Polygonid (Integer)	Seq-no (Integer)	Pointid (Integer)
Point	Elevation	
Pointid (Integer)	Forest-name (varchar)	Pointid (FK.) (Integer)
		Elevation (Real)

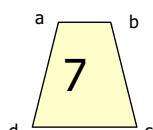
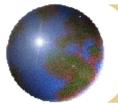


Figure 2.6. Schema for point, line, polygon, and elevation.

28



2.3 Extending ER with Spatial Concepts

- Motivation

- ER Model is based on discrete sets with no implicit relationships
- Spatial data comes from a continuous set with implicit relationships
- Any pair of spatial entities has relationships like distance, direction, ...

- Explicitly drawing all spatial relationship

- clutters ER diagram
- generates additional tables in relational schema
- Misses implicit constraints in spatial relationships (e.g. partition)

- Pictograms

- Label spatial entities along with their spatial data types
- Allows inference of spatial relationships and constraints
- Reduces clutter in ER diagram and relational schema
- Example: Fig. 2.7 (next slide) is simpler than Fig. 2.4

29



ER Diagram with Pictograms: An Example

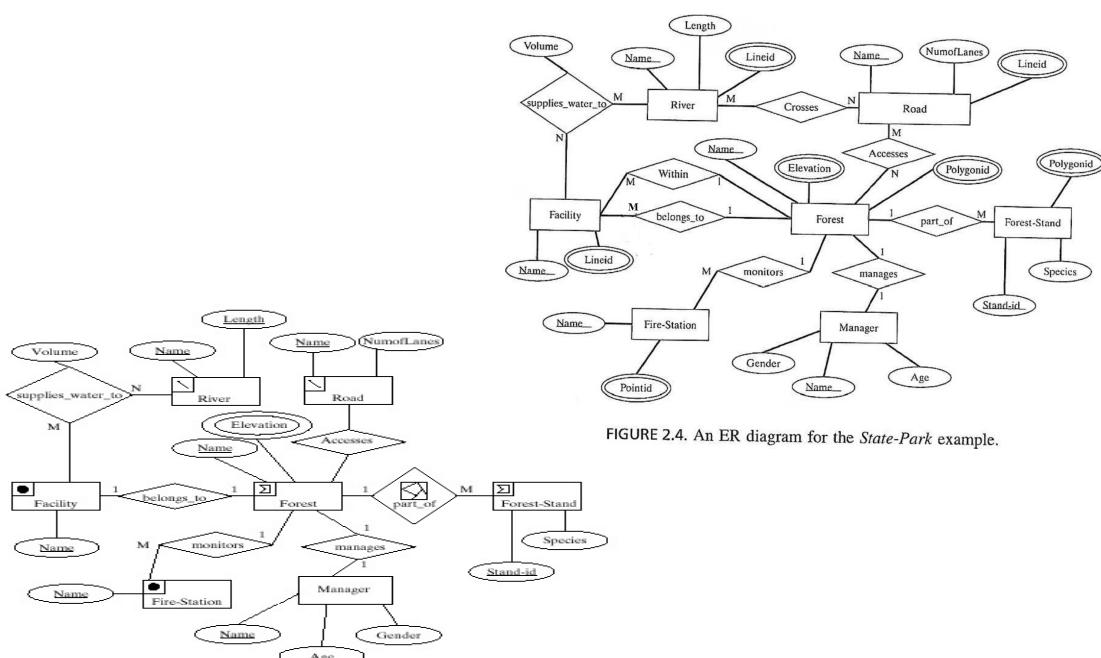


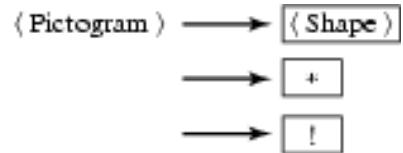
Figure 2.7. ER diagram for the State-Park example, with pictograms.

30

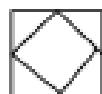
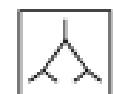


Specifying Pictograms

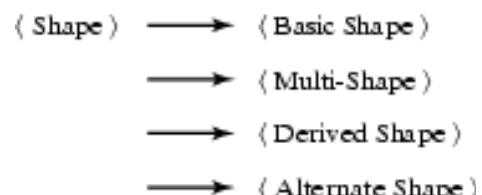
- Grammar based approach
 - Rewrite rule
 - like English syntax diagrams
- Classes of pictograms
 - Entity pictograms
 - basic: point, line, polygon
 - collection of basic
 - ...
 - Relationship pictograms
 - partition, network



Grammar (for Pictogram)



Part_of(Network) Part_of(Partition)



Grammar (for Shape)

Pictograms for Relationships

31



Entity Pictograms: Basic shapes, Collections



Grammar (for Basic Shape)

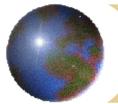
Pictograms for Basic Shapes



Grammar (for Cardinality)

Pictograms Multishapes
(using cardinality)

32



Entity Pictograms: Derived and Alternate Shapes

- Derived shape example is city center point from boundary polygon
 - ✓ Derive the shape of the US from the shapes of its state boundaries
- Alternate shape example: A road is represented as a polygon for construction or as a line for navigation

(Derived Shape) → (Basic Shape)



Grammar (for Derived Shape)

Pictograms for Derived Shapes

(Alternate Shape) → (Basic Shape) (Derived Shape)
→ (Basic Shape) (Basic Shape)

Grammar (for Alternate Shape)



Pictograms for Alternate Shapes

33



Summary

- ⊕ A data model is a high level description of the data
 - ▢ it can help in early analysis of storage cost, data quality
- ⊕ There are two popular models of spatial information
 - ▢ Field based and Object based
- ⊕ Database are designed in 3-steps
 - ▢ Conceptual, Logical and Physical
- ⊕ Pictograms can simplify Conceptual data models

34