Geographic Information System (GIS)

Dr. Eman Omar

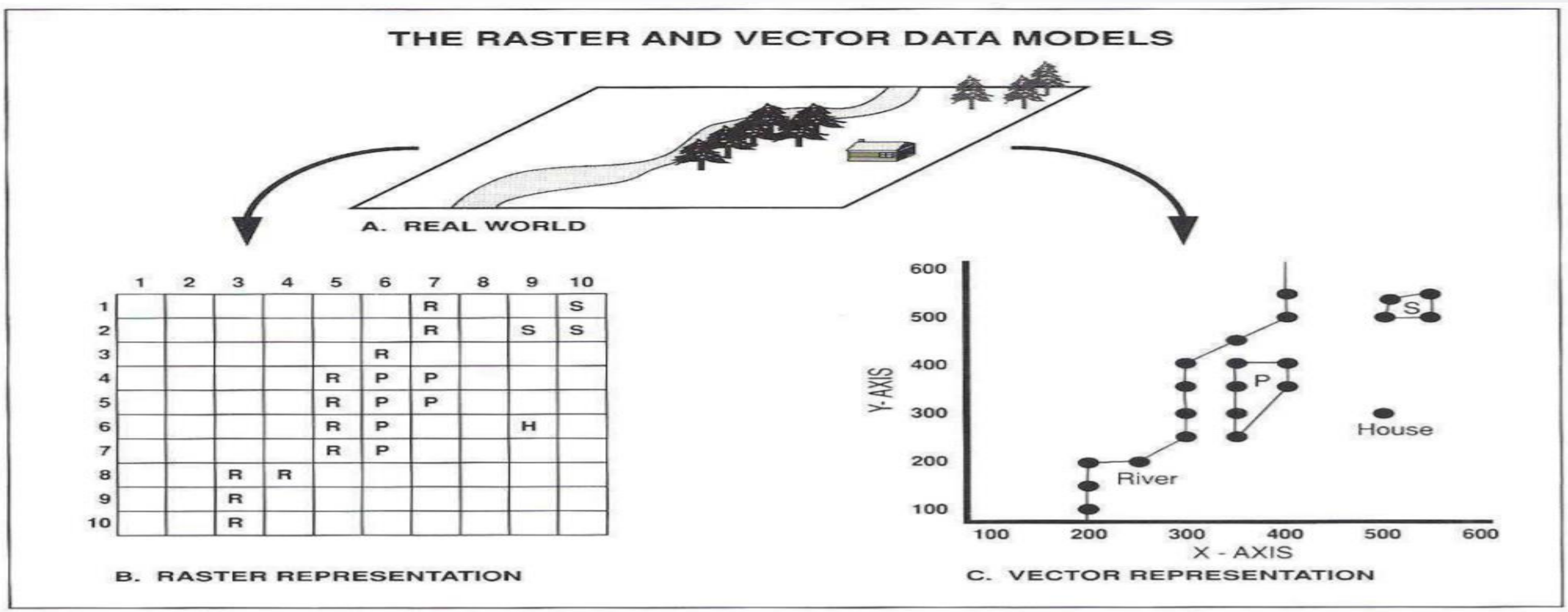
Spatial Data Models.(Vector and Raster Data Models)

The nature of Geographic Data

- The map is perhaps the most familiar form in which geographic data are represented.
- The information for a geographic feature has four major components:
  1. Locations.
  2. Non-spatial attributes.
  3. Time.
  4. Spatial relationships among geographic features.

Spatial Data Models

- There are two fundamental approaches to the representation of the spatial component of geographic information: the vector model and the raster model.



Raster Vs Vector Models

- In both models , the spatial information is represented using homogeneous units.
- In the raster approach, the homogeneous units are the cells. (The area within a cell is not subdivided and the cell attribute applies to every location within the cell.)
- In the vector approach, the homogeneous units are the points, lines, and polygons. Relative to the raster approach , these units are relatively few in number.

Comparison of Raster & Vector Data Models

RASTER MODEL

Advantages:

1. It is a simple data structure.
2. Overlay operations are easily and efficiently implemented.
3. High spatial variability is efficiently represented in a raster format.
4. The raster format is more or less required for efficient manipulation and enhancement of digital images.

Disadvantages:

1. The raster data structure is less compact. Data compression techniques can often overcome this problem.
2. Topological relationships are more difficult to represent.
3. The output of graphics is less aesthetically pleasing because boundaries tend to have a blocky appearance rather than the smooth lines of hand-drawn maps. This can be overcome by using a very large number of cells, but may result in unacceptably large files.

VECTOR MODEL

Advantages:

1. It provides a more compact data structure than the raster model.
2. It provides efficient encoding of topology, and, as a result, more efficient implementation of operations that require topological information, such as network analysis.
3. The vector model is better suited to supporting graphics that closely approximate hand-drawn maps.

Disadvantages:

1. It is a more complex data structure than a simple raster.
2. Overlay operations are more difficult to implement.
3. The representation of high spatial variability is inefficient.
4. Manipulation and enhancement of digital images cannot be effectively done in the vector domain.

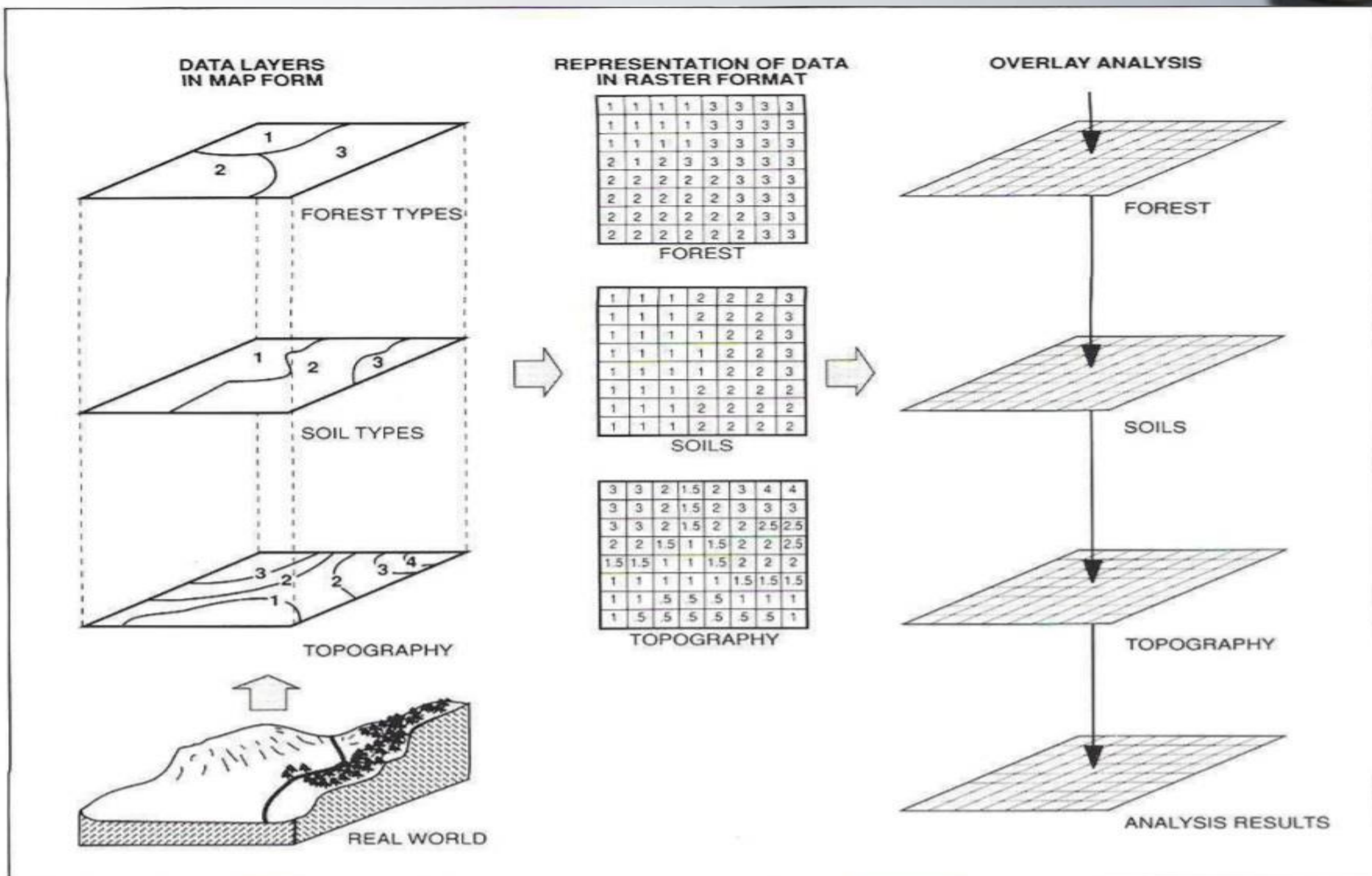
The Raster Data Model

- The raster data model consists of a regular grid of square or rectangular cells .
- The location of each cell or pixel (for picture element t) is defined by its row and column numbers.
- The value assigned to the cell indicates the value of the attribute it represents.

Overlay Analysis Using Raster Representation

- Each cell in a raster file is assigned only one value. So, different attributes are stored in separate files.
- The soil types and forest cover for an area would be stored as separate soil and forest data files.
- Operations on multiple raster files involve the retrieval and processing of the data from corresponding cell positions in the different data files. Conceptually, the process is like stacking the files as shown in the following Figure and using the vertical stack of cell values to analyze each cell location.

Overlay Analysis Using Raster Representation



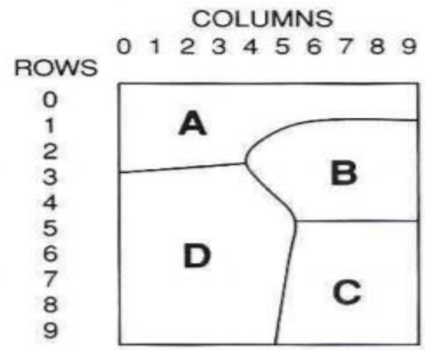
Run-Length Encoding

- Raster file could be very large.
- However , many of the cells may contain the same value as neighbouring cells. Where there is considerable redundancy in the size o f the raster file can b e a achieved by using various methods of data compression , such as run - length encoding and quad trees .
- Data compression is (The representation of data in a more compact form)

Run-Length Encoding

- In run-length encoding, adjacent cells along a row that have the same value are treated as a group termed a run.
- Instead of repeatedly storing the same value for each cell, the value is stored once, together with information about the size and location of the run.
- Several run-length encoding strategies have been developed , two of which are illustrated in the following Figure:
  1. **standard run length encoding:** in it the value of the attribute, the number of cells in the run, and the row number are recorded.
  2. **value point encoding**: the cells are assigned position numbers starting in the upper left corner, proceeding from left to right and from the top to bottom. The position number for the end of each run is stored in the point column. The value for each cell in the run is in the value column.

Run-Length Encoding



MAP

A. Full Raster Encoding (100 Values)

COLUMNS

ROWS	0	1	2	3	4	5	6	7	8	9
0	A	A	A	A	A	A	A	A	A	A
1	A	A	A	A	A	A	A	A	A	A
2	A	A	A	A	B	B	B	B	B	B
3	A	A	A	B	B	B	B	B	B	B
4	D	D	D	D	B	B	B	B	B	B
5	D	D	D	D	D	B	B	B	B	B
6	D	D	D	D	D	C	C	C	C	C
7	D	D	D	D	D	C	C	C	C	C
8	D	D	D	D	D	C	C	C	C	C
9	D	D	D	D	D	C	C	C	C	C

B. Run-Length Encoding (54 Values)

VALUE	LENGTH	ROW
A	10	0
A	10	1
A	4	2
B	6	2
A	3	3
B	7	3
D	4	4
B	6	4
D	5	5
B	5	5
D	5	6
C	5	6
D	5	7
C	5	7
D	5	8
C	5	8
D	5	9
C	5	9

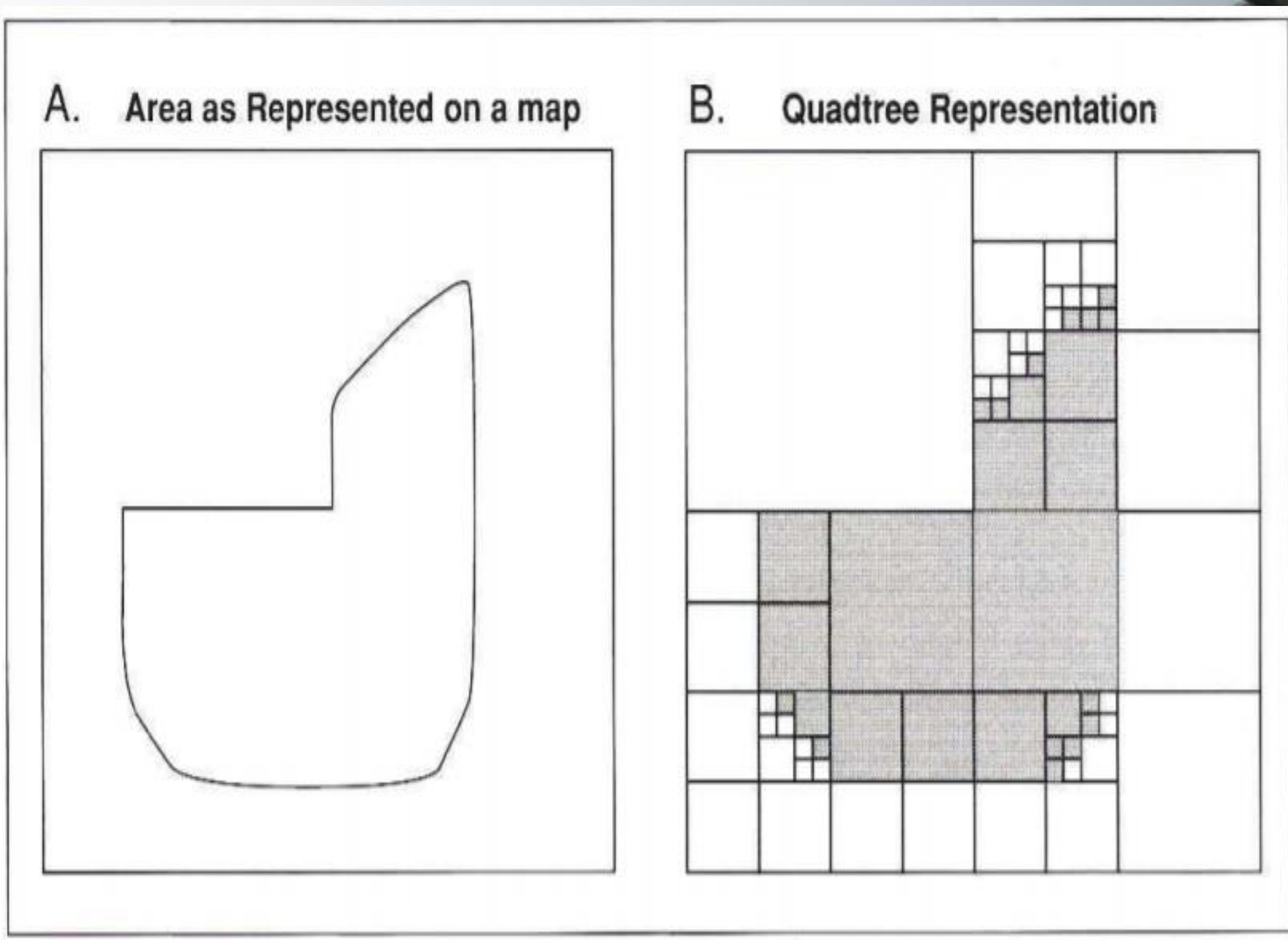
C. Value Point Encoding (32 Values)

VALUE	POINT
A	23
B	29
A	32
B	39
D	43
B	49
D	54
B	59
D	64
C	69
D	74
C	79
D	84
C	89
D	94
C	99

The Quadtree data model

- The quadtree data model provides a more compact raster representation by using a variable-sized grid cell.
- In stead o f dividing an area into cells of one size, finer subdivisions are used in those areas with finer details.
- It recursively subdivides a raster image into quarters. The subdivision process continues until each cell is classed.

The Quadtree data model cont.

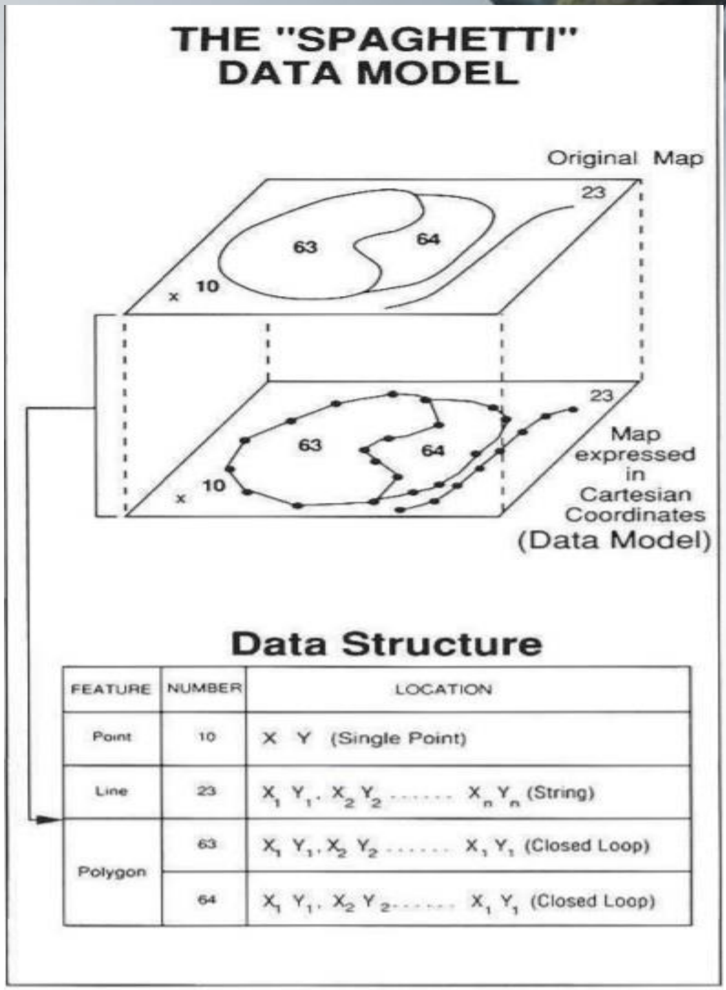


The Vector Data Model

- The approach used in the vector model is to precisely specify the position of the points , lines, and polygons used to represent features of interest.
- The location of features on the earth's surface are referenced to map positions using an XY coordinate system (termed a Cartesian coordinate system).
- Geographic features are commonly record ed on two dimensional maps as points, lines, and areas.
- A point feature is recorded as a single XY coordinate pair, a line as a series of XY coordinates, and an area as a closed loop of XY coordinate pairs that define the boundary of the area. (An area bounded by a closed loop of straight line segments is termed, a polygon.)

Vector Data Models (The Spaghetti Data Model)

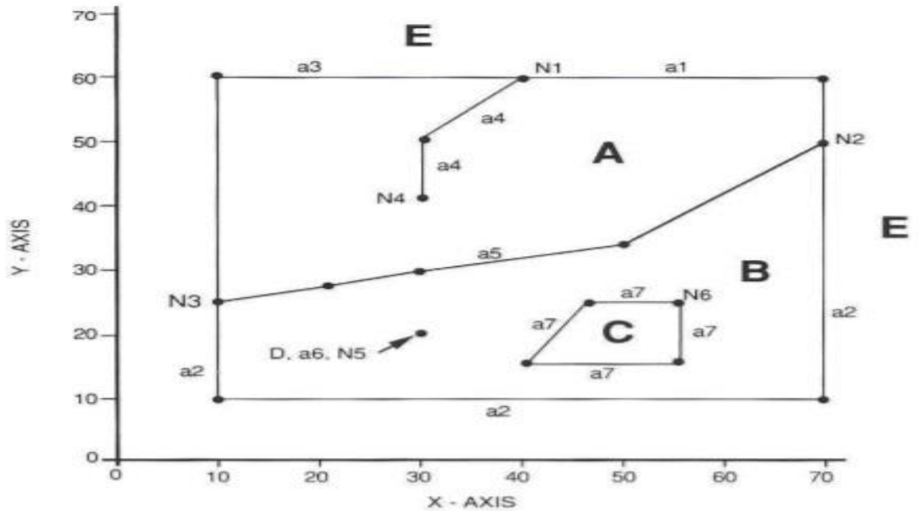
- The paper map is translated line- for-line into a list of XY coordinates.
- A point is encoded as a single XY coordinate pair and a line as a string of XY coordinate pairs. An area is represented by a polygon and is recorded as a closed loop of XY coordinates that define its boundary.
- The common boundary between adjacent polygons must be recorded twice, once for each polygon. A file of spatial data constructed in this manner is essentially a collection of coordinate strings with no inherent structure — hence the term **spaghetti model.**



FEATURE	NUMBER	LOCATION
Point	10	X Y (Single Point)
Line	23	X_1 Y_1, X_2 Y_2 .... X_n Y_n (String)
Polygon	63	X_1 Y_1, X_2 Y_2 .... X_1 Y_1 (Closed Loop)
	64	X_1 Y_1, X_2 Y_2 .... X_1 Y_1 (Closed Loop)

Vector Data Models (The Topological Data Model)

- The topological model is the most widely used method of encoding spatial relationships in a GIS.
- Topology is the mathematical method used to define spatial relationships.



SPATIAL DATA ENCODING

POLYGON TOPOLOGY	
POLYGON	ARCS
A	a1, a5, a3
B	a2, a5, 0, a6, 0, a7
C	a7
D	a6
E	area outside map coverage

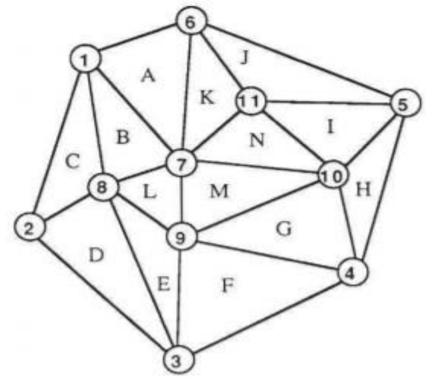
NODE TOPOLOGY	
NODE	ARCS
N1	a1, a3, a4
N2	a1, a2, a5
N3	a2, a3, a5
N4	a4
N5	a6
N6	a7

ARC TOPOLOGY				
ARC	START NODE	END NODE	LEFT POLYGON	RIGHT POLYGON
a1	N1	N2	E	A
a2	N2	N3	E	B
a3	N3	N1	E	A
a4	N4	N1	A	A
a5	N3	N2	A	B
a6	N5	N5	B	B
a7	N6	N6	B	C

ARC COORDINATE DATA			
ARC	START X, Y	INTERMEDIATE X, Y	END X, Y
a1	40, 60	70, 60	70, 50
a2	70, 50	70,10; 10,10	10, 25
a3	10, 25	10,60	40, 60
a4	40, 60	30,50	30, 40
a5	10, 25	20,27; 30,30; 50,32	70, 50
a6	30, 20		30, 20
a7	55, 27	55,15; 40,15; 45,27	55, 27

Vector Data Models (The Triangulated Irregular Network-TIN)

- The Triangular Irregular Network or TIN is a vector-based topological data model that is used to represent terrain data.
- A TIN represents the terrain surface as a set of interconnected triangular facets. For each of the three vertices, the XY coordinate (geographic location) and the Z coordinate (elevation) values are encoded.



X–Y COORDINATES

node	coordinates
1	x1, y1
2	x2, y2
3	x3, y3
:	
11	x11, y11

Z COORDINATES

node	coordinate
1	z1
2	z2
3	z3
:	
11	z11

EDGES

Δ	adjacent Δ
A	B,K
B	A,C,L
C	B,D
D	C,E
E	D,F,L
F	E,G
G	F,H,M
H	G,I
I	H,J,N
J	I,K
K	A,J,N
L	B,E,M
M	G,L,N
N	I,K,M

NODES

Δ	node
A	1,6,7
B	1,7,8
C	1,2,8
D	2,3,8
E	3,8,9
F	3,4,9
G	4,9,10
H	4,5,10
I	5,10,11
J	5,6,11
K	6,7,11
L	7,8,9
M	7,9,10
N	7,10,11

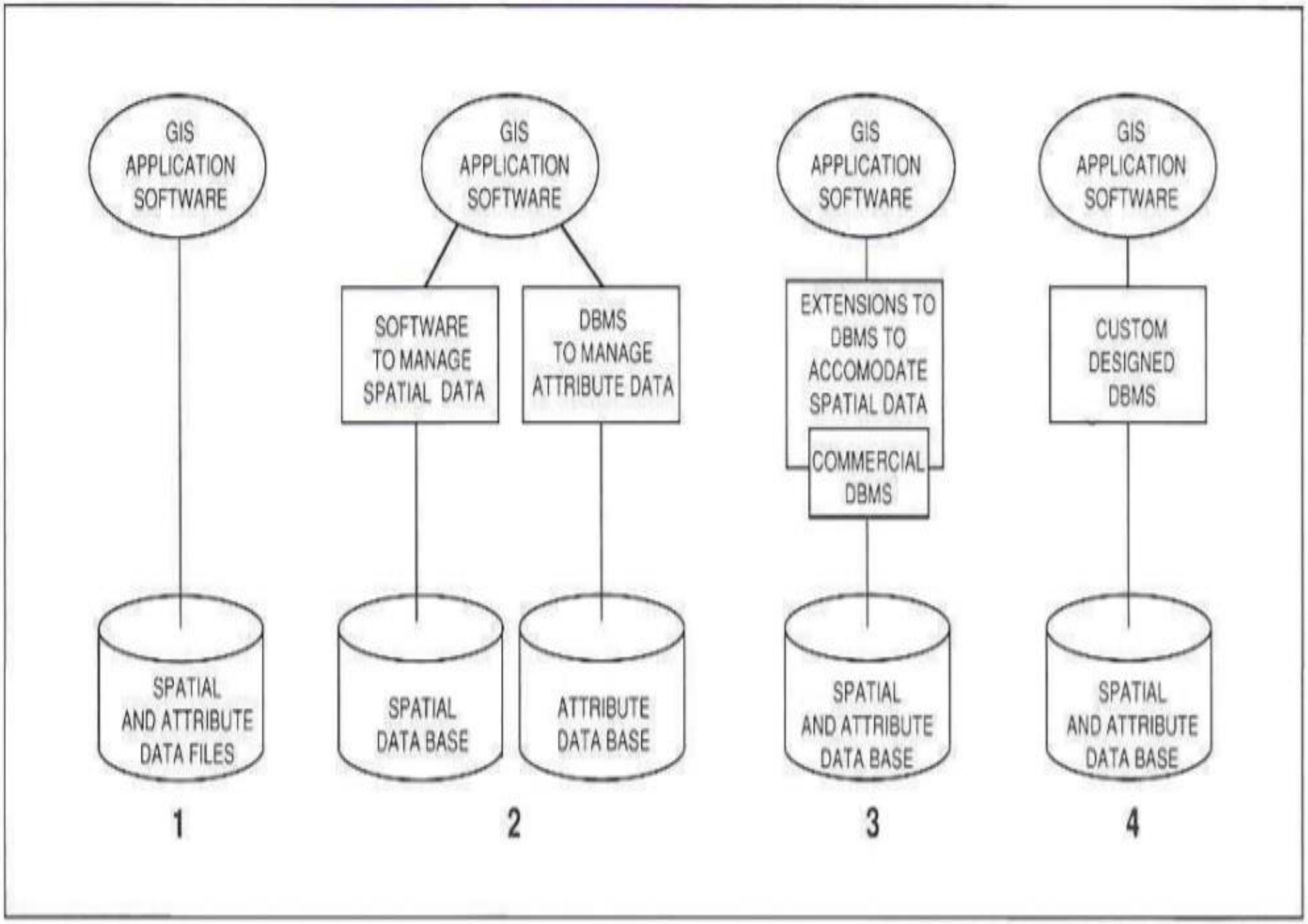
Limitation of General Purpose DBMS for GIS Applications

1. The spatial data records used in a GIS are variable length records which are needed to store variable numbers of coordinate points, whereas general purpose database systems are designed to handle fixed length records.
2. Manipulation of geographic data involves spatial concepts, such as proximity, connectedness, containment, and overlay, that are not easily accommodated by general purpose database query languages.
3. A G1S requires sophisticated graphics capabilities that are not normally supported b y a general purpose DBMS.
4. Geographic information is complex. The representation of a single geographic feature requires multiple records in multiple files.
5. The highly interrelated nature of G1S data records requires a more sophisticated security system than the record locking approach used by general purpose DBMS.

Approaches Used to Implement a GIS

- Various practical approaches were taken to provide data management services for a GIS. They may be broadly grouped into the following four, somewhat overlapping, strategies:
  1. Develop a proprietary system providing the individual data management services required by the different application modules. This is the file processing approach.
  2. Develop a hybrid system using a commercially available DBMS (usually a relational one) for storage of the nonspatial attributes. Develop separate software to manage the storage and analysis of the spatial data, using the services of the relational DBMS to access the attribute data.
  3. Use an existing DBMS, usually a relational one, as the core of the GIS. Then develop extensions to the system where needed. Although the spatial and attribute data may be managed by the DBMS, a significant amount of software is generally added to the DBMS to provide the spatial functions and graphics display used in geographic analysis.
  4. Start from scratch and develop a spatial database capable of handling the spatial and non-spatial data in an integrated fashion.

Approaches to GIS Systems Design



Thank You

Questions ??