Bharathidasan University

Tiruchirappalli – 620 023, Tamil Nadu

6 Yr. Int. M.Tech. Geological Technology and Geoinformatics

Course : MTIGT0501

GEOGRAPHIC INFORMATION SYSTEMS (GIS)

Unit-2 : GIS DATA STRUCTURES

Dr. K. Palanivel

Professor, Department of Remote Sensing

Course Objectives:

- To learn the basics and concepts of GIS
- To know the components and importance of GIS
- To study the capabilities of GIS in input, verification, analysis, modelling and outputgeneration
- To understand the importance of manipulation and their applications
- To learn the methods of spatial data analyses, simulation and modelling aspects.

MTIGT0501: GEOGRAPHIC INFORMATION SYSTEMS --- 3 credits

1. **Basics of GIS:** Definition - Usefulness of GIS - Components of GIS - Computer Hardware, Software Modules and Organizational Context of GIS. **6 Hrs.**

2. **Data Structure:** Data Structure in GIS - Types of Data (Points, Lines and Polygons)- Data Base Structures (Raster Data Structures and Vector data Structures) - Data Conversion, (Vector to Raster and Raster to Vector). **6 Hrs.**

3. **Data Input, Verification, Storage and Output:** Spatial Data Input Processes and Devices (Sources of data, - Different Types of Data Entry methods, viz., Manual input, Run length code, Digitization, Automated Scanning, etc. – Vector to Raster conversion – Raster to Vector conversion - Input devices) - Entry of non-spatial data – Linking of Spatial & Non-spatial data – Data Verification (Errors of different types) – Correction (Rubber Sheet Transformation, Bilinear interpolation, Cubic Convolution, etc.) – GIS capabilities for Data correction – Data output (Types of Output, GIS Capabilities for output, Output devices). **12 Hrs.**

4. **Methods of Spatial Interpolation and Digital Elevation Model:** Spatial Interpolation: Basic Principles of Interpolation – Methods of Interpolation (Interpolation by Joining Boundaries, viz., Simple vector maps, Theisson polygons) – Global Methods of Interpolation, Local Interpolation (Trend Surface Analysis) – Local Interpolation (Splines) - Optimal Interpolation (Kriging).

   Digital Elevation Modeling: Need for Three Dimensional Models - Methods of DEM - Products of DTM (Contour Maps, Shaded Relief Map, Maps Related To Slopes, Line of Sight Maps, Drainage Analysis, Volume Estimation, etc.) - Usefulness of DEM/DTM. **12 Hrs.**

5. **Data Analysis and Spatial Modeling:** Simple data retrieval – Data retrieval through Boolean Logic – Map Overlaying and Cartographic Modeling (Two layers, Multiple layers, Binary, Index, Regression, and Process Models) – Overlay analysis, Capabilities (Point Operations, Regional Operations, Neighbourhood Operations) – Buffering – Cartographic Modeling using Natural Language Commands – Advantages and disadvantages of Carto modeling – Net work analysis. **12 Hrs.**

6. **Current Contours: (Not for Final Exam, only for Discussion):** Recent advancements in GIS; Application of GIS in automation, decision making and query building processes in Geological Technology; Modules and capabilities of QUGIS, GRAM++, IDRISI GIS software.

1. Burrough, P.A 1986: Principles of Geographical information Systems for Land Resources Assessment, Clarandone Press, Oxford.
2. Kang - Tsung Chang, Introduction to Geographic Information System, MC Graw Hill, Boston. 2002.
3. Avery, T.V, Interpretation of Aerial Photography Burgass, Publishing Company.
4. Gautham, N.C 1970: Urban Landuse Study Through Aerial Photo interpretations Techniques, Pink Publishing House, Mathura.
5. American Society of Photogrammetry, 1983: Manual of Remote Sensing (2nd Edition), ASP Falls Church, Virginia.
6. Campbell, J 1984: introductory Cartography, Printers Hall Englewood Cliffs, N.J
7. Dent B.D 1985: Principles of Thematic Map Design, Addition - Wesley, Reading, Mass.
8. Freeman, H and GG.Pieroni 1980: Map Data Processing, Academic Press, New York.
9. Monmonier, M.A 1982: Computer Assisted Cartography - Principles and Prospects, Prentice Hall, Englewood Cliffs, NJ
10. Tomlinson, R.F Calkins, H.S and D.F.Marble 1976: Computer Handling of Geographic Data, UNESCO, Geneva.
11. Grame F. & Bonham – Carter; Geographic information Systems for Geoscientists; Modelling with GIS, Pergamon.

Course outcomes:

After the successful completion of this course, the students are able to:

1. Understand the basic concepts and virtues of this important tool providing various platforms to handle Geospatial data
2. Gain basic ideas to generate, group, store Geospatial data in effective data structures
3. Develop skills on manipulation, 3D visualization, Spatial Analysis and Spatial Modeling.
4. Handle Geologic problems Geospatially in GIS platform independently.

Unit – 2

GIS DATA STRUCTURES

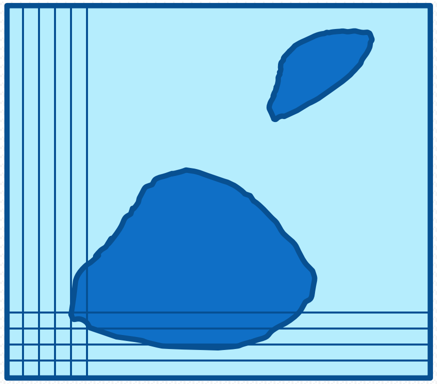
Data Structure: Data Structure in GIS - Types of Data (Points, Lines and Polygons) - Data Base Structures (Raster Data Structures and Vector data Structures) - Data Conversion, (Vector to Raster and Raster to Vector).

6 Hrs.

2. GIS DATA STRUCTURES

2.1 Definition

- Computer based data structure for the representation of real world entities
  - Two types
    - Vector data structure
    - Raster data structure



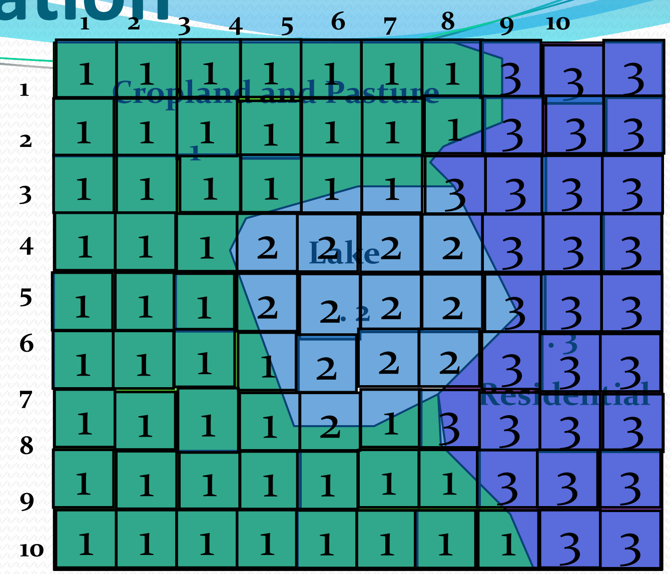
  - A data structure that uses
    - points, lines and polygons to describe world surface / spatial phenomena – VDS
    - Sets of regular shaped tessellated units to describe world surface / spatial phenomena – RDS
    - Tessellate - cover a surface by repeated use of a single shape **without gaps** or **overlapping**.

GIS Data Representation

1	1	1	1	1	1	1	3	3	3
1	1	1	1	1	1	1	3	3	3
1	1	1	1	1	1	3	3	3	3
1	1	1	2	2	2	2	3	3	3
1	1	1	2	2	2	2	3	3	3
1	1	1	1	2	2	2	3	3	3
1	1	1	1	2	1	3	3	3	3
1	1	1	1	1	1	1	3	3	3
1	1	1	1	1	1	1	1	3	3

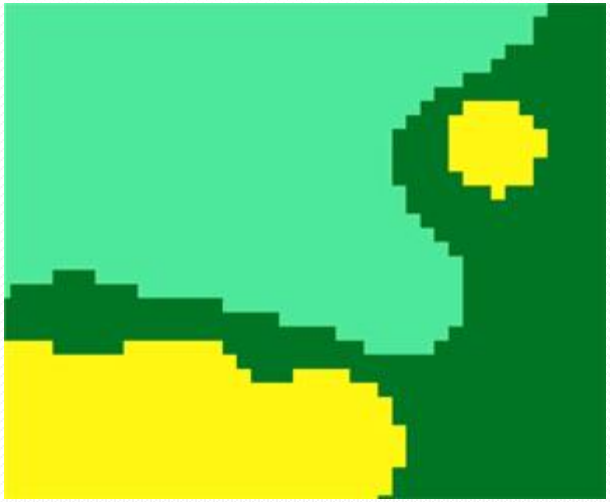
Rasterization

Vector

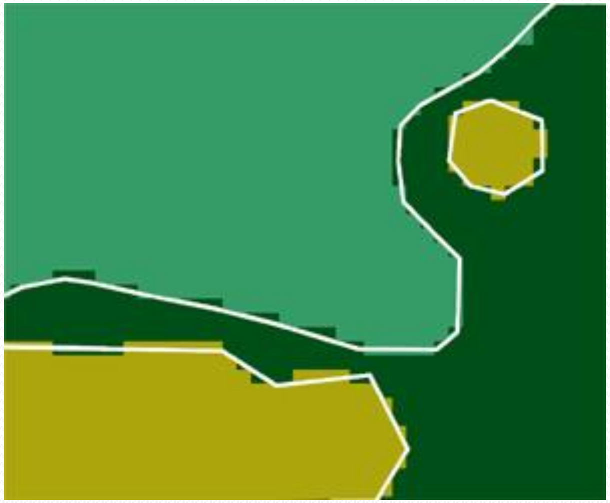


Consist arrays of cells in a grid pattern

Raster (Pixel size = 5mm x 5mm)



Vectorization



Go Back

Consist vector elements – points, lines & polygons

2.2 File and Data access

- Essential features of any **data storage system** are that they should allow data to be accessed and cross-referenced quickly.

1. Simple lists – Consider one second for checking a record or row (**n** no. of rows), the **Average Search Speed** in seconds can be calculated as, **T = (n+1)/2,** where, T = time, n=number of records in a file.

- For example, if there are **10,000 records** in a database, then it may take around a minimum of 5000.5 seconds, i.e., around **1 hour**
- **39 minutes.**
- Search event becomes a time taking one
- But easy to update the record, i.e., deletion and updation of record, and addition of new records at the end of the list in that file.

2.2 File and Data access ...contd...

2. Ordered sequential files – Nos. in Telephone directory

– Average Search Speed, T = $\mathbf{log_2(n+1)}$

- Binary Search
- For a search in a database containing 10,000 records,
- It takes 14 seconds only
- Searching becomes 50-100 times speedier
- Data updation becomes tedious
- By adding one row (i.e., a record), the entire database need to be arranged orderly – every time.

*n*n	$\log_2 n$
1	0
2	1
4	2
8	3
16	4
32	5
64	6
128	7
256	8
512	9
1024	10
2048	11
4096	12
8192	13
16384	14

2.2 File and Data access ...contd...

3. Indexed files – Word search in a dictionary

- Ordered and indexed for quick access
  - **Direct files** **(n1+1)/2 + (n2+1)/2**
  - **Indirect files** Where, n1 = No. of steps in the index
  - **Index inverted files** n2 = No. of items in the data block

  - Quick searching
  - Data updation is tedious – the entire database need to be ordered and indexed every time on deletion, editing or updating events.

Inverted files

Soil Profile No.	Series	pH	Depth	Drainage	Texture	Erosion
1	A	4	Deep	Good	Sandy	---
2	B	5	Shallow	Good	Clay	Yes
3	C	6	Shallow	Poor	Sandy	No
4	D	7	Deep	Good	Clay	Yes
5	E	4	Deep	Poor	Clay	No
6	F	5	Shallow	Poor	Clay	No

Index (Inverted file)

Topic	Soil Profiles (sequential numbers in original file)					
	1	2	3	4	5	6
Deep	1			4	5	
Shallow		2	3			6
Good Dr.	1	2		4		
Poor Dr.			3		5	6
Sandy	1		3			
Clay		2		4	5	6
Eroded		2		4		

2.3 Database structures and management

1. **Hierarchical** database structure
2. **Network** database structure
3. **Relational** database structure
4. **Object-oriented** database structure
5. **Full topological polygon network** db structure
6. **TIGER** (**T**opologically **I**ntegrated **G**eographic **E**ncoding and **R**eferencing)

- Other developments in database structures

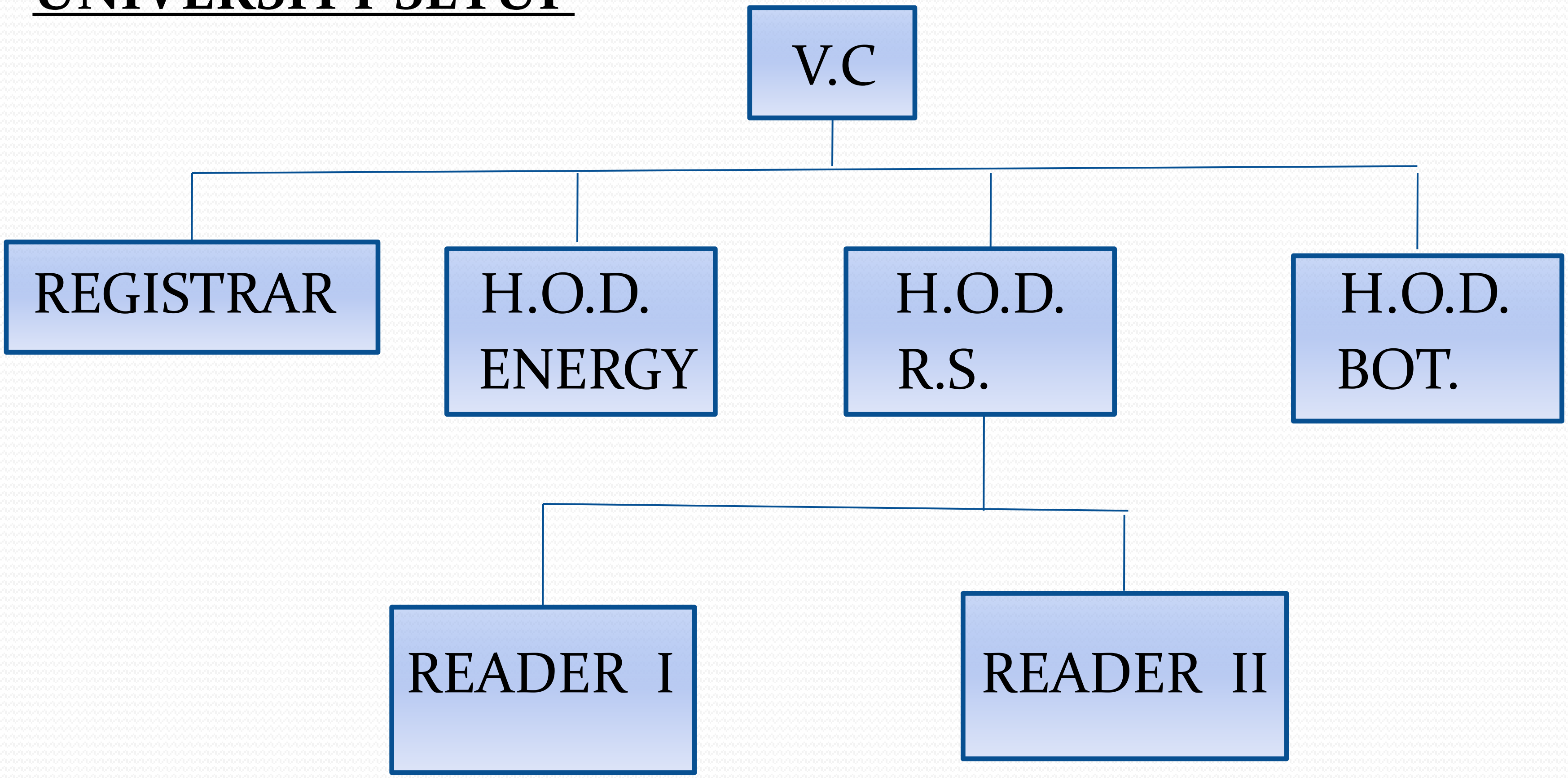
2.3 Database structures and management ...contd...

1. Hierarchical database structure

- Maintains a hierarchy as per their importance and level
- Vertical downward or upward connectivity amongst levels are possible
- Lateral connectivity amongst the database is not possible
- Data redundancy occurs
- Data updation is tedious

HIERARCHICAL DATA STRUCTURE

UNIVERSITY SETUP



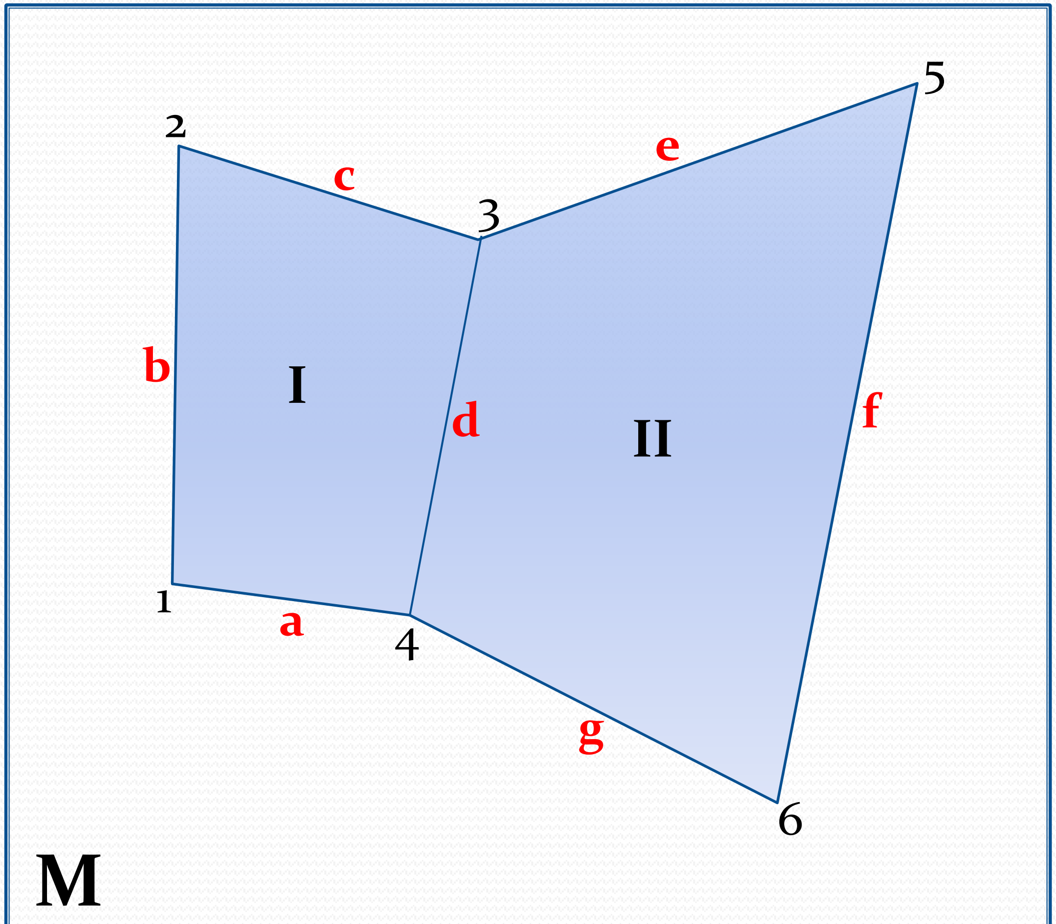
Go Back

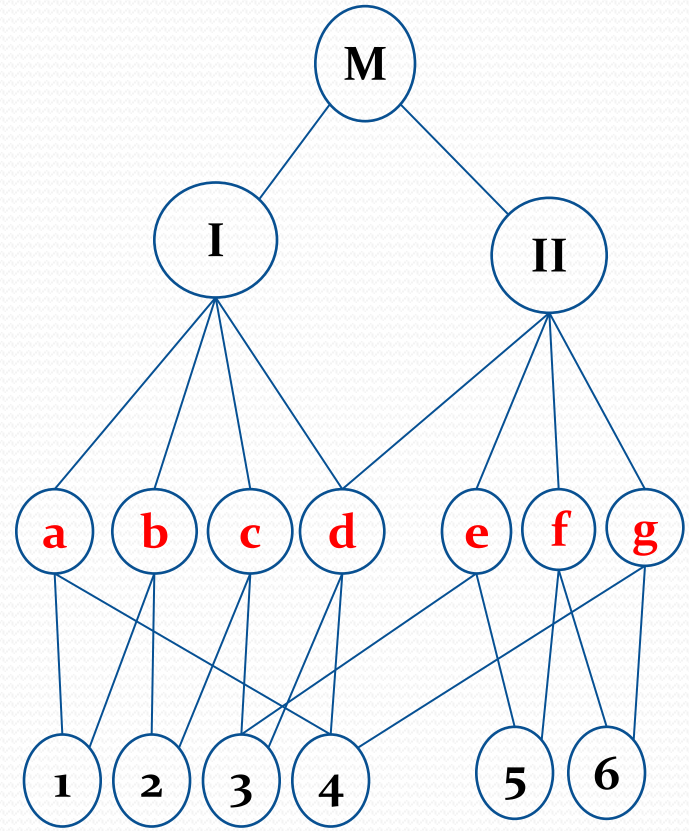
2.3 Database structures and management ...contd...

2. Network database structure

- Eliminates lateral connectivity amongst database – lateral linkage can be easily provided using networks and pointers
- Data repetivity – data redundancy is totally avoided
- Data updation is tedious.

NETWORK DATA STRUCTURE





Go Back

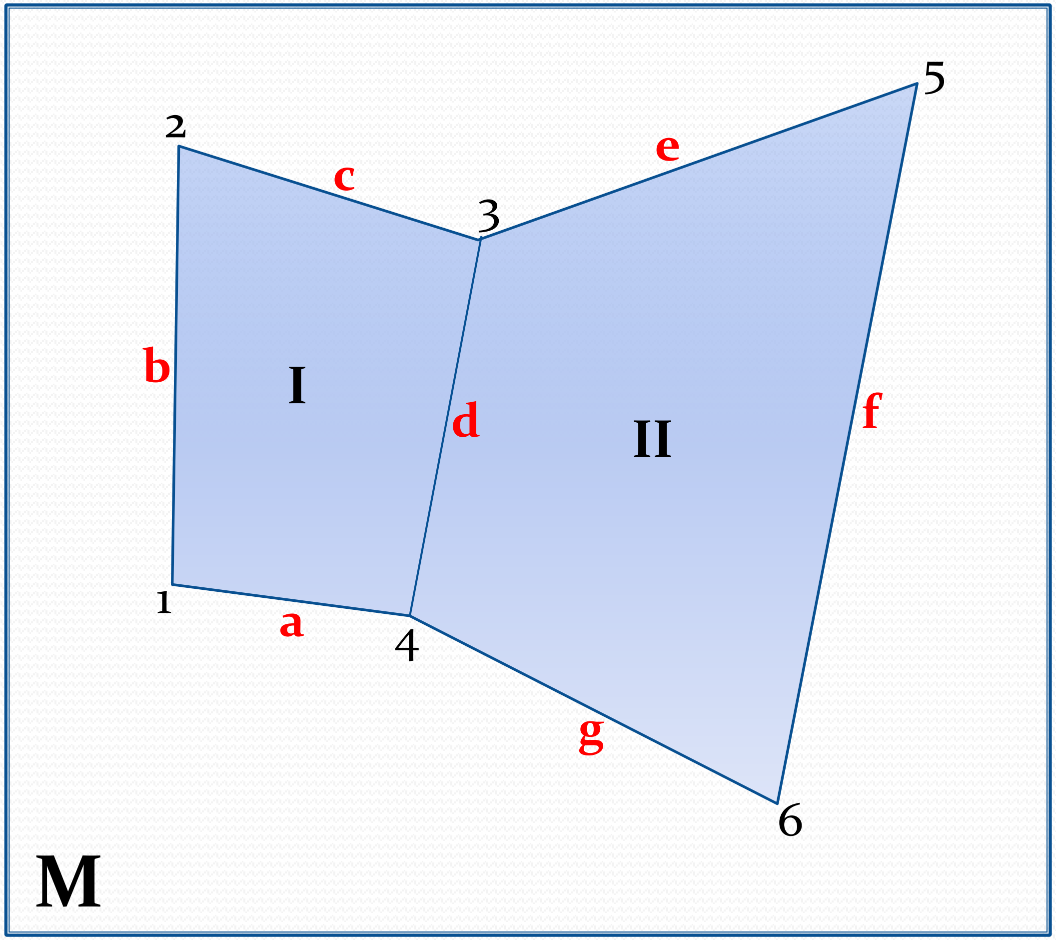
Map M consists of 2 polygons, 7 lines (6+1 shared) and 6 nodes,

2.3 Database structures and management ...contd...

3. Relational database structure

- Integrated database structure – a combination of Index, Hierarchy and Network structures
- Limited data redundancy is allowed
- Clarity in the database
- Every table is networked with Label Identifiers for the features
- Data updation is also easy
- Uses moderate storage space but effective searching is done

RELATIONAL DATA STRUCTURE



Map:

M	I	II

Polygons:

I	a	b	c	d
II	d	e	f	g

Lines:

I	a	4	1
I	b	1	2
I	c	2	3
I	d	3	4
II	e	3	5
II	f	5	6
II	g	6	4
II	d	3	4

Co-ordinates:

Go Back

4. OBJECT-ORIENTED GIS

Object-oriented data model is one application of object-oriented technology.

- It uses objects to organize spatial data.
- Unlike a geometric object of a point, line or area, an ***OBJECT*** is defined here as something that has a set of properties and can perform operations upon requests.
- That means, almost everything one uses in a GIS is an object.
- For e.g., a land use map is an object, which has properties, such as its coordinate system and feature type and can respond to requests such as zoom in, zoom out and a query.

There are aspects in Object oriented GIS and classified into two : 1) Structural aspects and 2) Behavioural aspects.

1) Structural aspects

Principles: To group objects : association, aggregation, generalization, instantiation and specialization aspects are required.

Association: Describes the relationships between objects of two types.

- If owner and land parcel represent two types of objects, the relationships between them can follow the rules that
  - ✓an owner can own one or more parcels and
  - ✓a parcel can be owned by one or more owners.

Aggregation: Asymmetric association in a whole-part relationship.

- For e.g., block groups are connected to form a census tract and census tracts are connected to form a county.

Generalization: Identifies the commonality among objects, and groups objects of similar types into a higher-order type.

- For e.g., parcel, zoning and census tract maps may be grouped into a higher-order class called boundary.
- Grouping of objects forms a hierarchical structure, which organizes objects into classes and classes into superclasses and subclasses.

Instantiation: An object of a class can be created from an object of another class.

- For e.g., a high-density residential area object may be created from a residential area object.

Specialization: Differentiates objects of a given class by a set of rules.

- For e.g., roads may be separated by average daily traffic volume.

Behavioural aspects of objects:

Inheritance: It is the basic principle in explaining the behaviours of objects:

- subclasses inherit properties and operations from a superclass, and
- objects inherit properties and operations from a subclass.
- For e.g., the residential area is a superclass and low-density area and high-density area are the subclasses.

- All properties of the class residential area are inherited by its own subclasses.
- Through inheritance, properties need only be defined once in the class hierarchy.

Encaptulation : Encaptulation refers to the mechanism to hide the properties and operations of an object so that the object can perform an operation by responding to a predefined message or request.

- For e.g., a polygon object can respond to a request called ReturnCenter (return the center of) by returning the physical center of the polygon.

Polymorphism: It allows the same operation to be implemented in different ways in different objects.

- For e.g., the same request called GetDimension (get the dimension of) can be sent to a point, a line, or a polygon but the result differs depending on the feature type.
- The number 0 is returned if the object is point,
  1 if the object is a line, and
  2 if the object is a polygon.

2.4 Data structures for Raster data

I. Simple raster database structure
II. Hierarchical database structure &
III. Compact data structures
   1. Chain encoding
   2. Run length encoding
   3. Block encoding
   4. Quadtree and binary tree encoding…

I. Simple Raster Data Structure

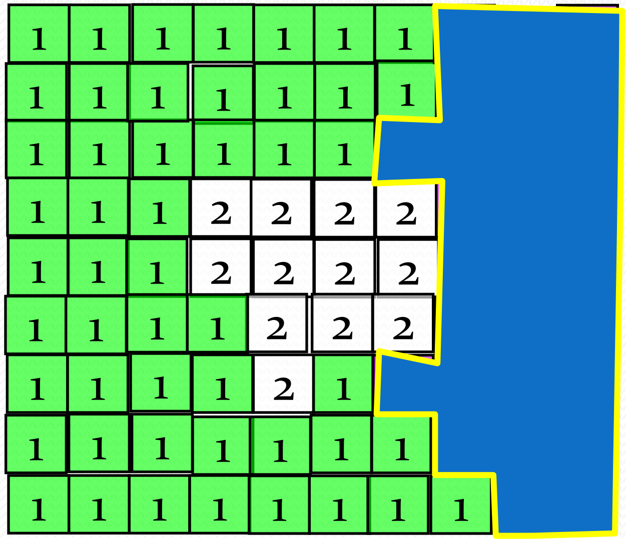
Consist arrays of cells in a grid pattern – a grid cell, i.e., pixel = picture element,

A Pixel is referenced by a row and column (x, y or Longitude, Latitude) and an attribute value (a number, alphabet or both to represent the real world object)

A point is represented by a single pixel / grid cell

A line by an array of pixels as a string along a direction continuously and

An area, i.e., polygon by an agglomeration of neighbouring cells.



This simple list raster database structure is easy to handle in computer.

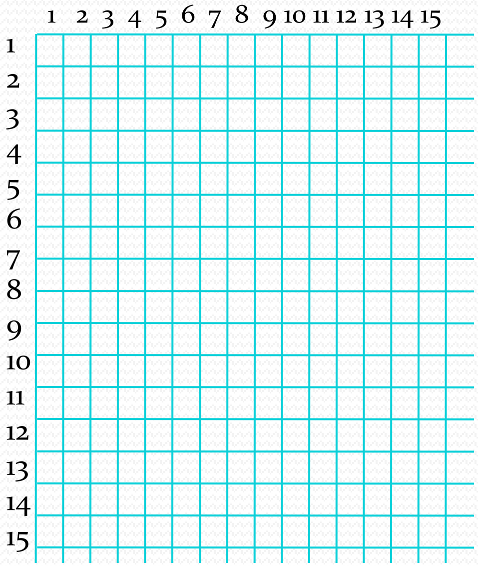
But, raster coding can affect accuracy of geometrical properties of objects – length / distance, area, perimeter, etc., due to cartesian effect of cells.

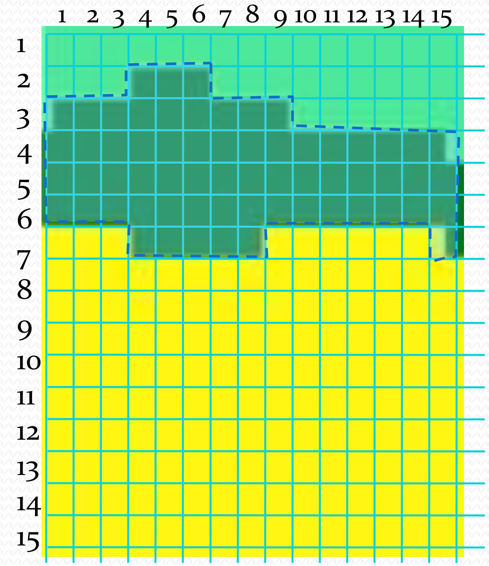
II. Hierarchical Raster Data Structure

- Based on the ascending or descending orderly arrangement, the spatial features / objects, i.e., each mapping units are referenced directly in this structure.
- Recoding a map requires one number per mapping unit which can consists of set of spatial features handled at a time – thus it has many advantages using many-to-one relationship.
- But, both are difficult when the database consists of too many objects and information covering a very large area.

III. Compressed Data structures

1. Chain encoding





$0^3, 1^1, 0^3, 3^1, 0^3, 3^1, 0^5, 3^1, 0^1, 3^3, 2^1, 1^1, 2^6, 3^1, 2^5, 1^1, 2^3, 1^4$

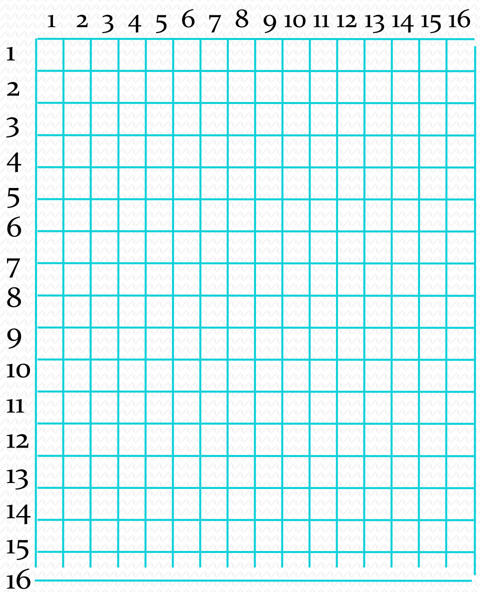
Conditions:

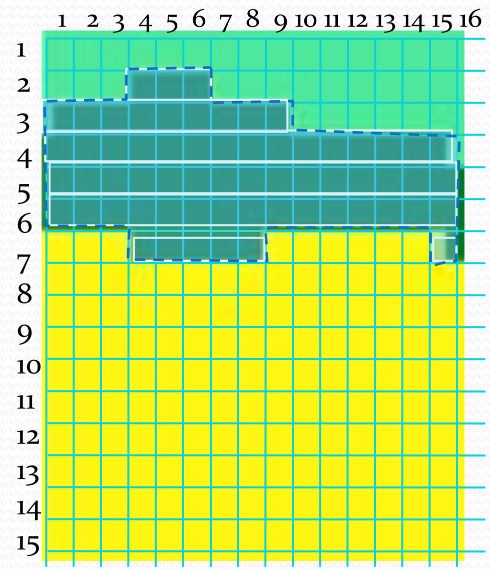
1. Define Origin: X & Y of starting cell
2. Sequence of unit vectors / cells along the outer boundary of the region in cardinal directions
3. Directions can be numbered (East = 0
   North = 1
   West = 2
   South = 3

4. Boundary of the region is coded clockwise.

Go Back

2.Run length encoding





Row wise from left to right – begin cell, end cell of a particular class

Row 1 ……

Row 2 4,6

Row 3 1,9

Row 4 1,15

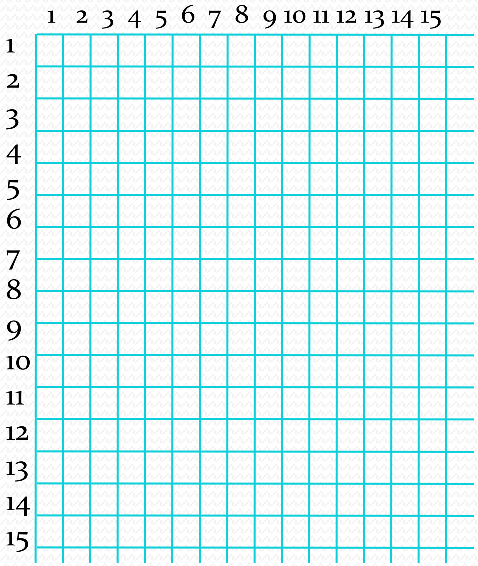
Row 5 1,15

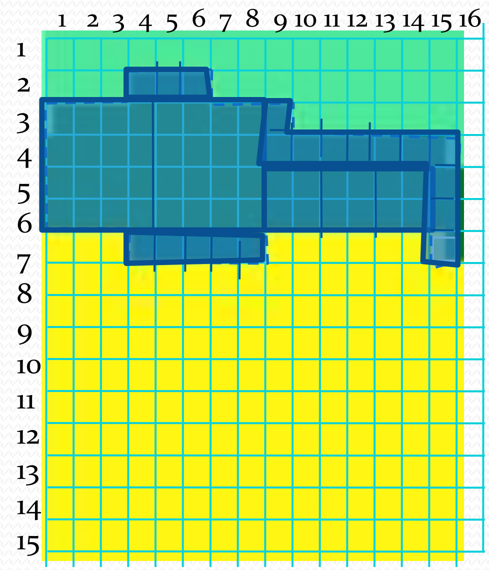
Row 6 1,15

Row 7 4,8 15,15

Go Back

3. Block encoding





lllr to RLE, but 2 dimensionally, using square blocks

Data structure consists of 3 nos.

1. The origin – X, Y(centre or bottom left)
2. Radius of each square – single unit squares, 4-square blocks, 16-square blocks, etc.

For Green Region 'B',

2 16-squares

3 4-squares

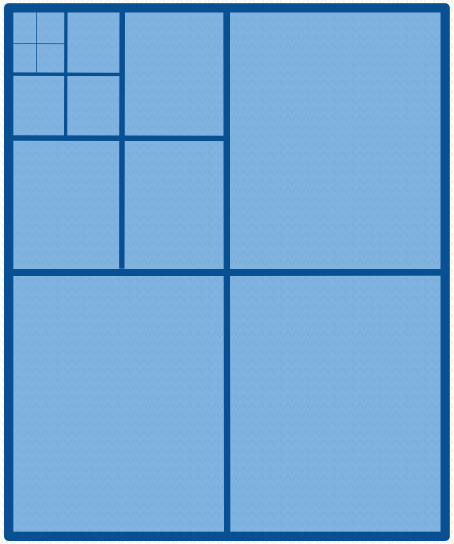
16 unit squares

2 coordinates for each square (42)
3 for cell sizes,
Total=45 nos.

Go Back

4. Quadtree encoding

Successive, hierarchical division of 2^n x 2^n array

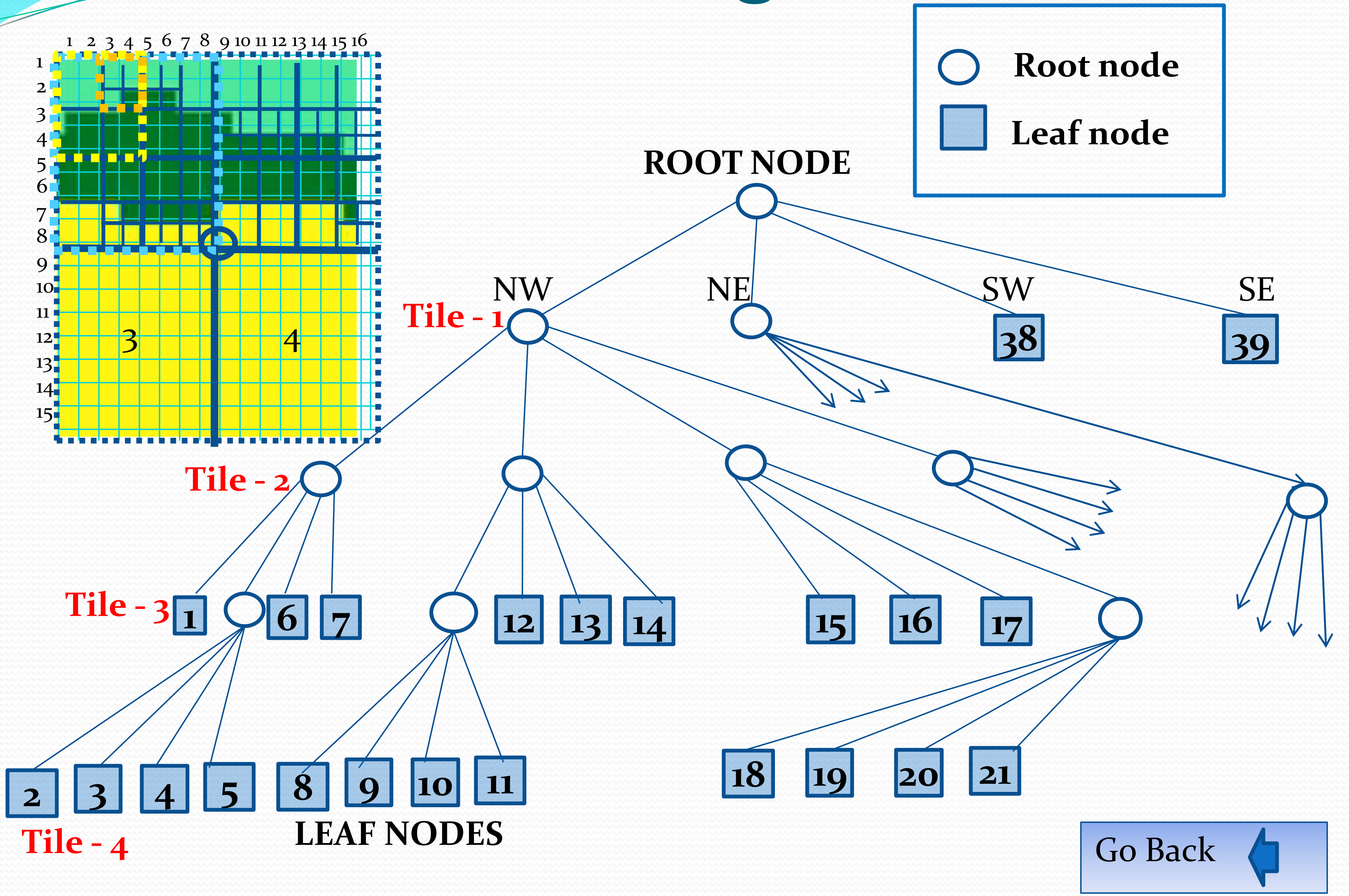


If the region is divided into two halves each time – **binarytree**

If the region is tiled into four quadrants each time - **quadtree**

Lowest limit of the division is - single cell - pixel.

Quadtree encoding...contd...



2.5 Vector – Raster comparison

ADVANTAGES AND DISADVANTAGES OF RASTER DATABASE

ADVANTAGES	DISADVANTAGES
1. DATA STRUCTURE IS SIMPLE	1. IF THE PIXEL SIZE IS LARGER TO REDUCE THE VOLUME THEN DETAILS ARE LOST
2. OVERLAYING OF THEMATIC DATA WITH REMOTELY SENSED DATA IS EASY	2. NETWORK LINKAGE IS DIFFICULT
3. SPATIAL ANALYSIS IS EASY	3. PROJECTION & TRANSFORMATION IS TIME CONSUMING
4. SIMULATION IS EASY	4. OUTPUT LOOKS CLUMSY

2.5 Vector – Raster comparison ...CONTD...

ADVANTAGES AND DISADVANTAGES OF VECTOR DATA BASE

ADVANTAGES

1. COMPACT DATA STRUCTURE
2. TOPOLOGY WITH CLEAR NETWORK
3. ACCURATE GRAPHICS
4. RETRIVAL AND UPDATING IS POSSIBLE

DISADVANTAGES

1. COMPLEX DATA STRUCTURE
2. COMBINATION OF SEVERAL THEMES CREATE PROBLEMS
3. SIMULATION AND PROGNESTIGATION IS DIFFICULT
4. SPATIAL ANALYSIS IMPOSSIBLE.