



MÔ HÌNH VÀ CẤU TRÚC DỮ LIỆU KHÔNG GIAN

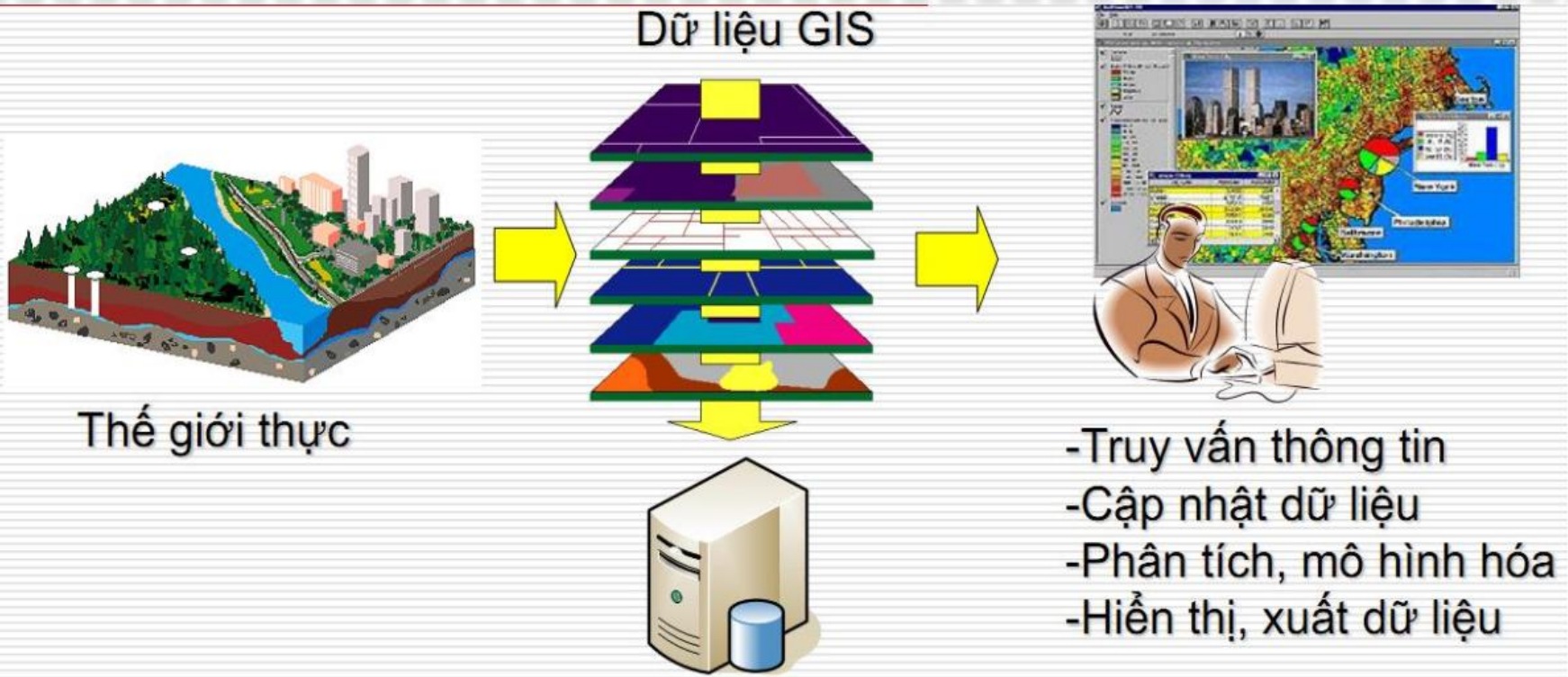
Giảng viên: Kiều Tuấn Dũng, Nguyễn Tu Trung
BM HTTT, Khoa CNTT, Trường ĐH Thủy Lợi

Hà Nội, 2019

Nội dung

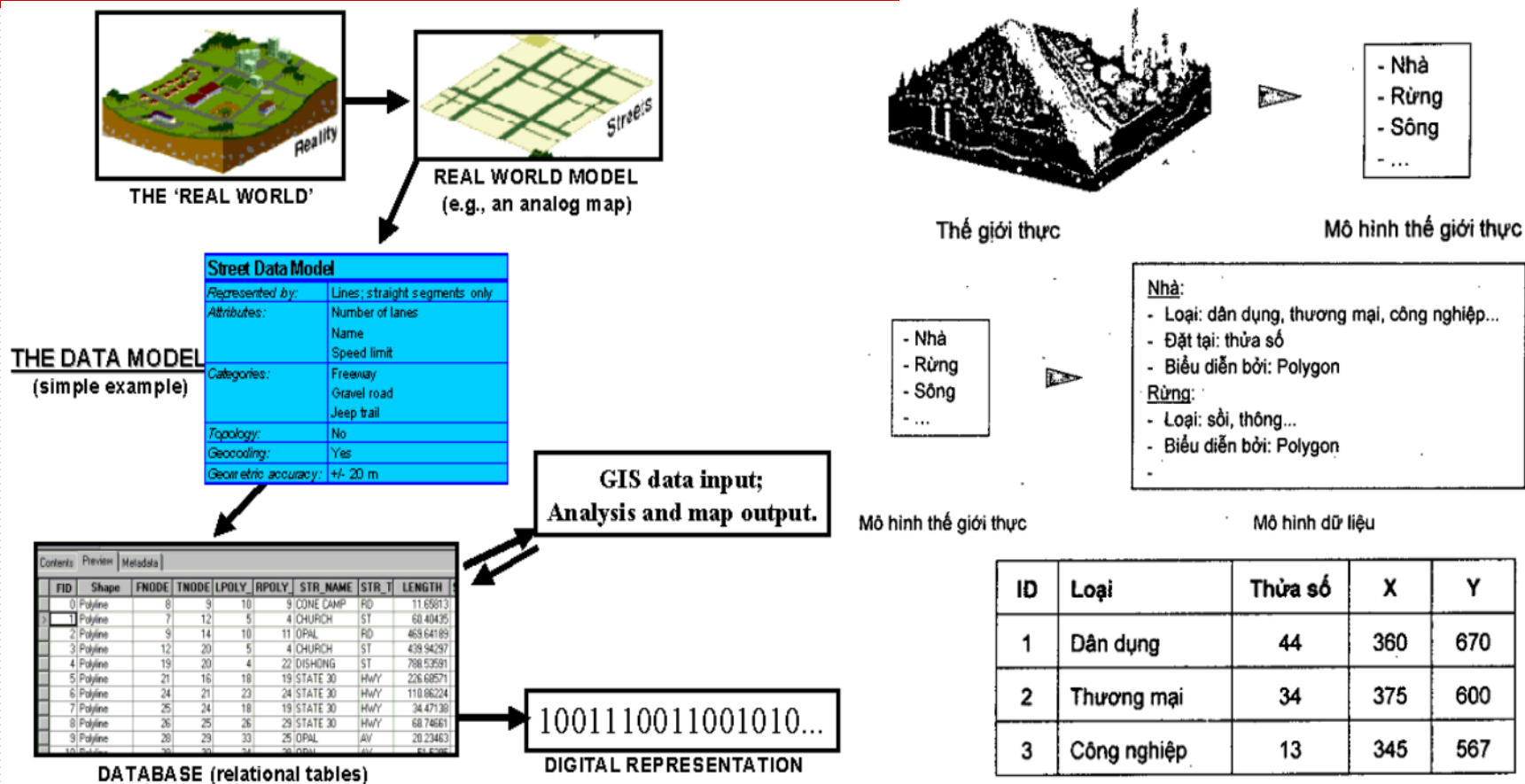
- ❖ Từ thế giới thực đến GIS
- ❖ Thực thể không gian trong GIS
- ❖ Vai trò của mô hình dữ liệu trong GIS
- ❖ Mô hình không gian
- ❖ Đối tượng không gian
- ❖ Mô hình dữ liệu không gian
- ❖ Mô hình dữ liệu Raster
- ❖ Mô hình dữ liệu Vector
- ❖ Mối quan hệ giữa các đối tượng không gian
- ❖ So sánh giữa Raster và Vector

Từ thế giới thực đến GIS



- ❖ Thực thể không gian (spatial entity): là những gì tồn tại trong thế giới thực
- ❖ Đối tượng không gian (spatial object): là những thực thể không gian được biểu diễn trong máy tính số

Từ thế giới thực đến GIS

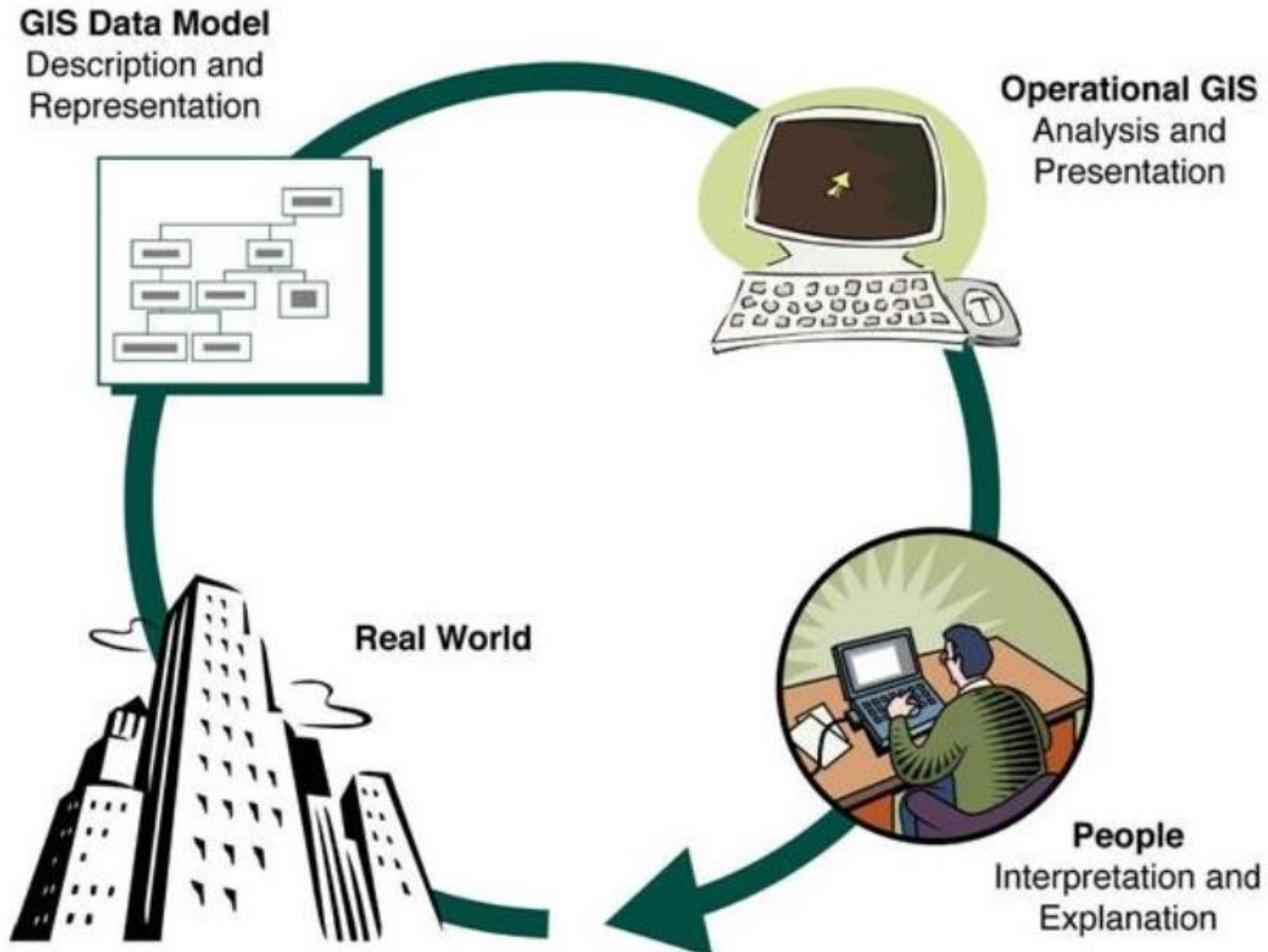


- ❖ Thế giới thực (các thực thể) được biểu diễn trong máy tính bằng những đối tượng không gian dưới dạng dữ liệu số (không gian và phi không gian)

Thực thể không gian trong GIS

- ❖ Được mô tả bằng 4 thành phần:
 - ❖ Vị trí địa lý của đối tượng được mô tả (không gian): theo hệ tọa độ địa lý
 - ❖ Mối liên hệ của đối tượng đó theo không gian (không gian): nằm trong, bên cạnh, cắt nhau, trên, dưới ...
 - ❖ Thông tin thuộc tính của đối tượng (phi không gian)
 - ❖ Thời gian

Vai trò của Mô hình dữ liệu trong GIS



Mô hình không gian

❖ Mô hình không gian:

- ❖ Là sự đơn giản hoá thế giới thực, là tập những phần tử biểu diễn các thực thể không gian trong thế giới thực

❖ Mô hình dữ liệu không gian:

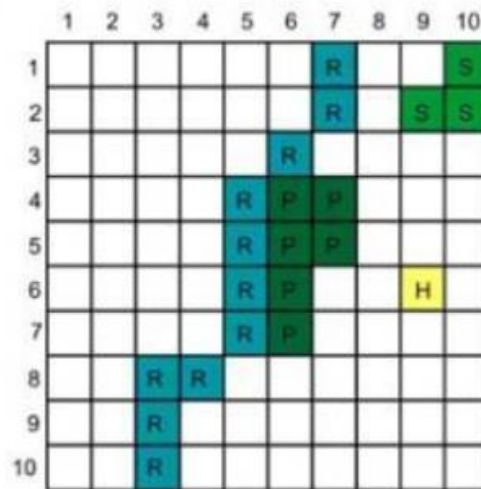
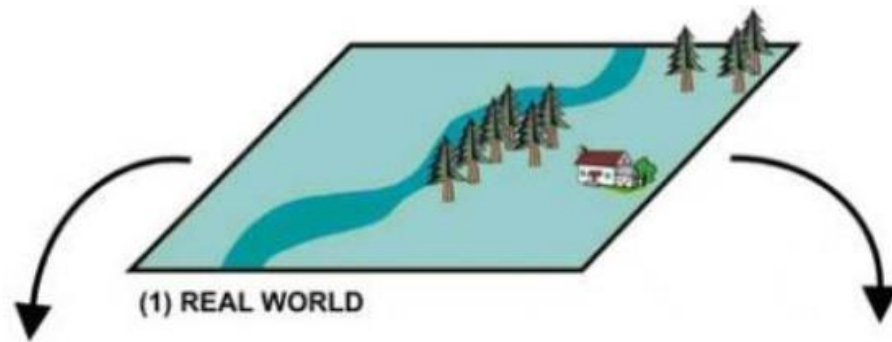
- ❖ Tập các nguyên tắc chuyển hóa thế giới thực thành các đối tượng không gian được miêu tả một cách logic

Đối tượng không gian

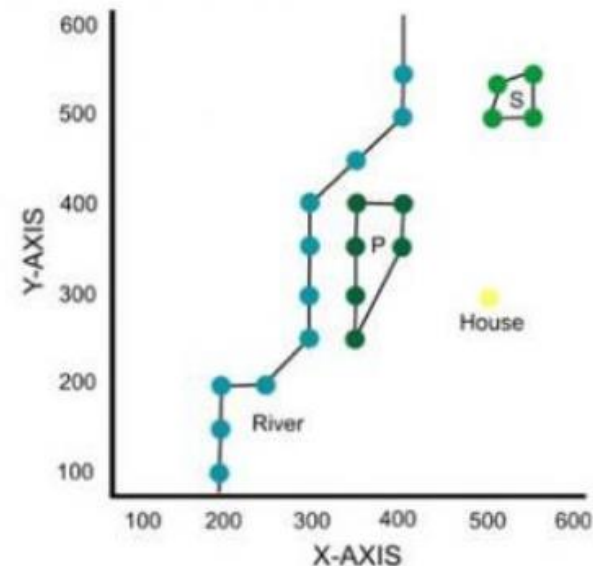
❖ Dữ liệu số về các đối tượng không gian: được biểu diễn trong máy tính dưới dạng nhị phân theo

❖ Mô hình Raster (liên tục)

❖ Mô hình Vector (rời rạc)



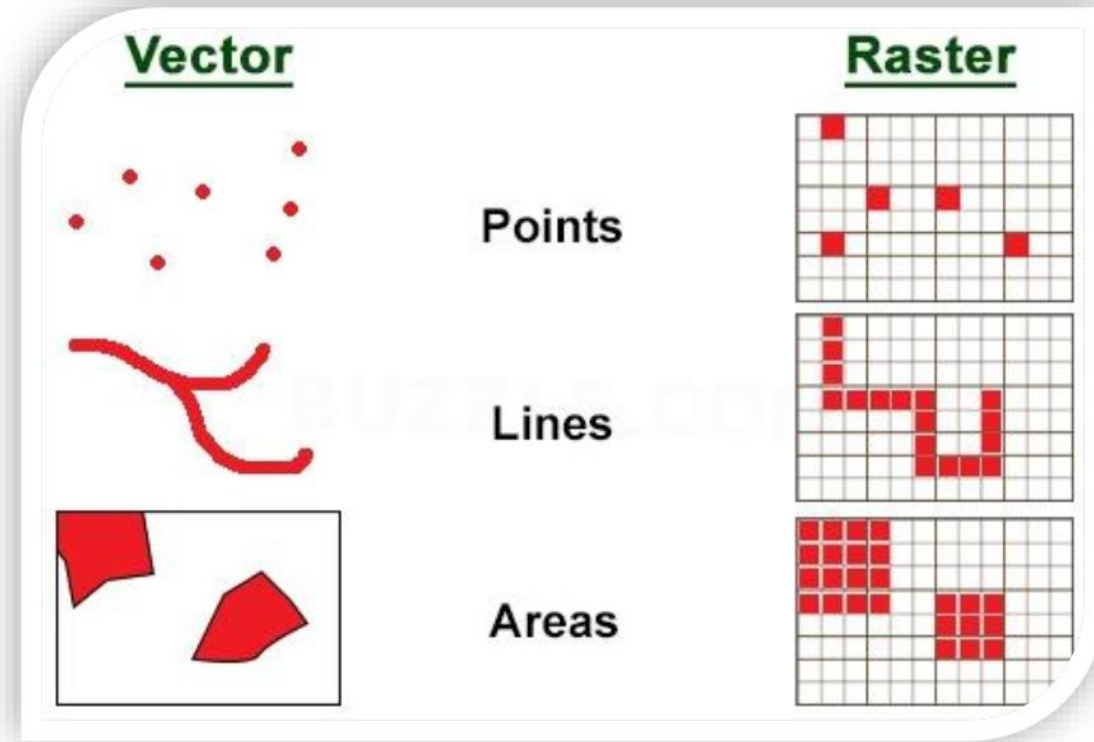
(2) RASTER REPRESENTATION



(3) VECTOR REPRESENTATION

Mô hình dữ liệu không gian

- ❖ Mô hình Raster: Các đối tượng không gian được chia thành các ô lưới bằng nhau gọi là điểm ảnh (pixel), mỗi điểm ảnh chỉ có 1 một thuộc tính
- ❖ Mô hình Vector: Các đối tượng không gian được biểu diễn như những điểm, đường, vùng

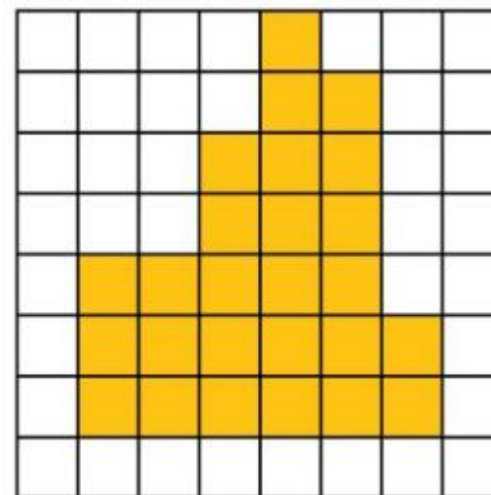


Mô hình dữ liệu Raster

- ❖ Cấu trúc dữ liệu Raster
- ❖ Biểu diễn các đối tượng:
 - ❖ Đối tượng điểm
 - ❖ Đối tượng đường
 - ❖ Đối tượng vùng
- ❖ Kỹ thuật nén dữ liệu
- ❖ Các định dạng file Raster

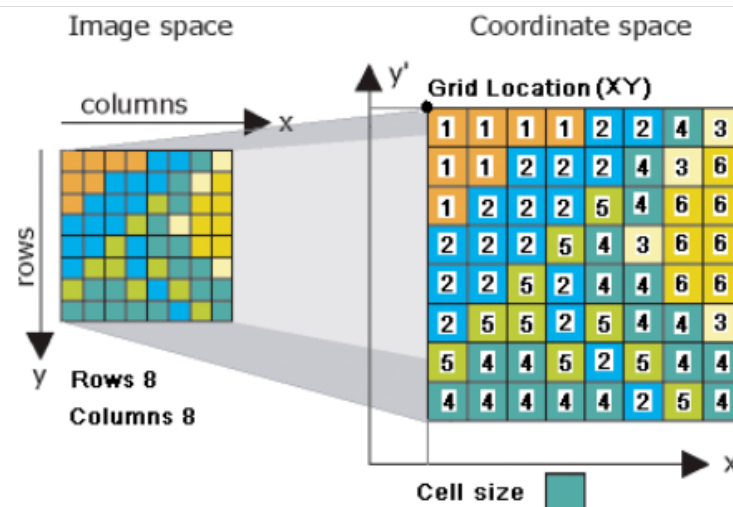
Cấu trúc dữ liệu Raster

- ❖ Mỗi điểm ảnh chỉ biểu diễn một thuộc tính xác định bởi giá trị $f(x,y)$

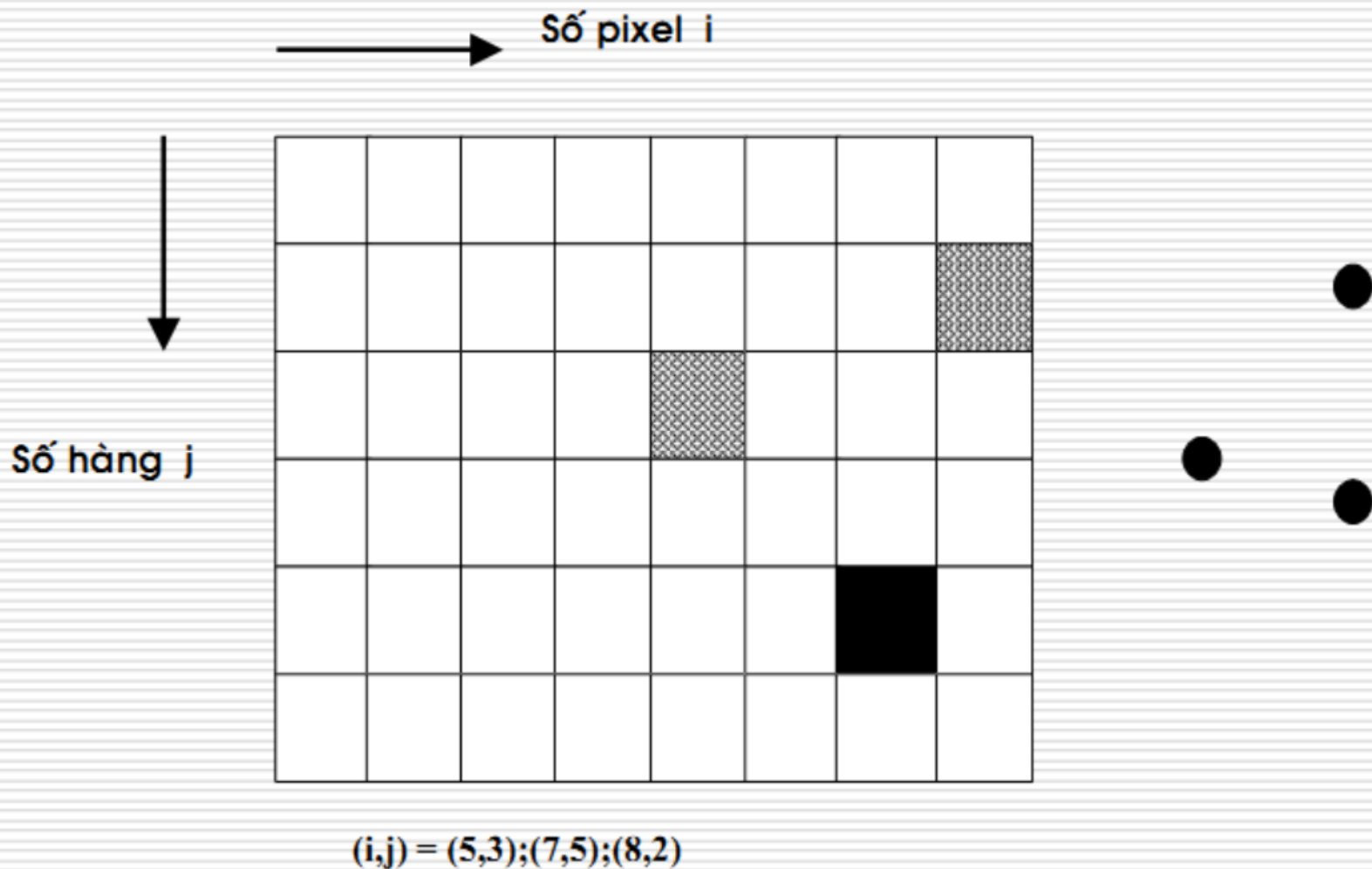


Row 1: 00001000
Row 2: 00001100
Row 3: 00011100
Row 4: 00011100
Row 5: 01111100
Row 6: 01111110
Row 7: 01111110
Row 8: 00000000

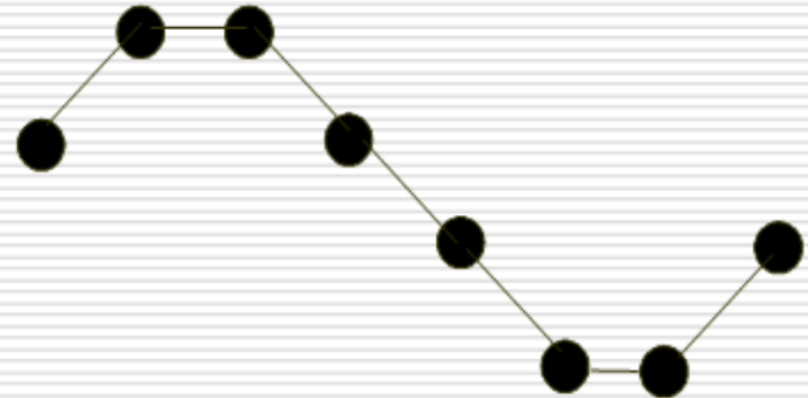
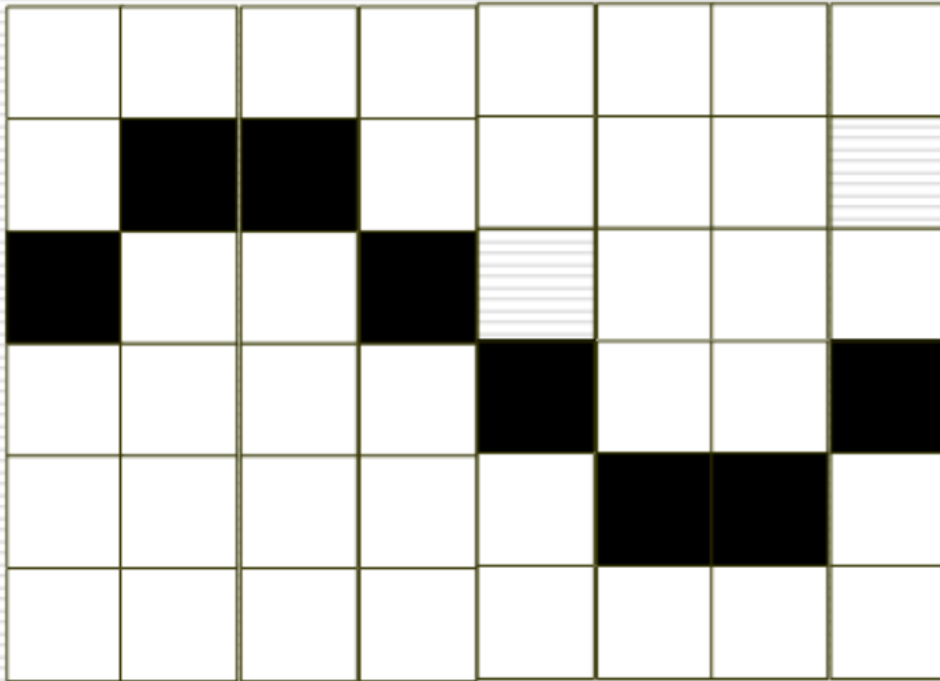
- ❖ Khi thay đổi độ phân giải, kích thước điểm ảnh thay đổi, dung lượng dữ liệu thay đổi theo



CTDL Raster – Đối tượng điểm (Point)



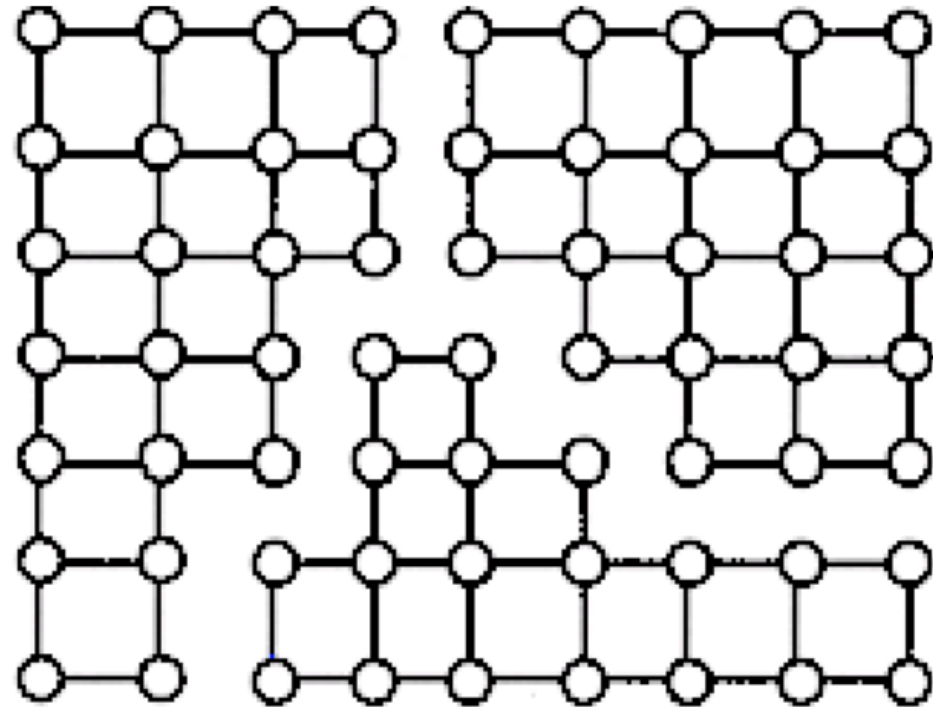
CTDL Raster – Đối tượng đường (Line)



(1,3);(2,2);(3,2) ;(4,3); (5,4)
;(6,5) ;(7,5) ;(8,4)

CTDL Raster – Đối tượng vùng (Polygon)

A	A	A	A	B	B	B	B
A	A	A	A	B	B	B	B
A	A	A	C	C	B	B	B
A	A	A	C	C	C	B	B
A	A	C	C	C	C	C	C
A	A	C	C	C	C	C	C



(4A,4B), (4A,4B),(3A,2C,3B)
(3A,3C,2B), (2A,6C), (2A,6C)

CTDL Raster – Kỹ thuật nén dữ liệu

- ❖ Khó khăn: Không gian cần dùng để lưu trữ dữ liệu raster lớn
=> Cần nén dữ liệu
- ❖ Một ảnh Raster thường bao gồm hàng triệu Pixel
- ❖ Các kỹ thuật nén dữ liệu được sử dụng để giảm kích thước file ảnh
 - ❖ PP Run-Length Encoding: các Run là nhóm pixel cùng giá trị nằm liên tục nhau trên cùng dòng gồm (giá trị/chiều dài/dòng)
 - ❖ PP Value Point Encoding: dữ liệu được lưu trữ là (dòng/cột) của điểm cuối cùng trong chuỗi pixel liên tục cùng giá trị
 - ❖ PP Quadrees: chia ảnh thành các tiểu vùng, mỗi tiểu vùng có cùng giá trị (tiểu vùng nào có nhiều giá trị khác nhau thì lại tiếp tục chia)

Các định dạng file Raster

❖ Tham khảo:

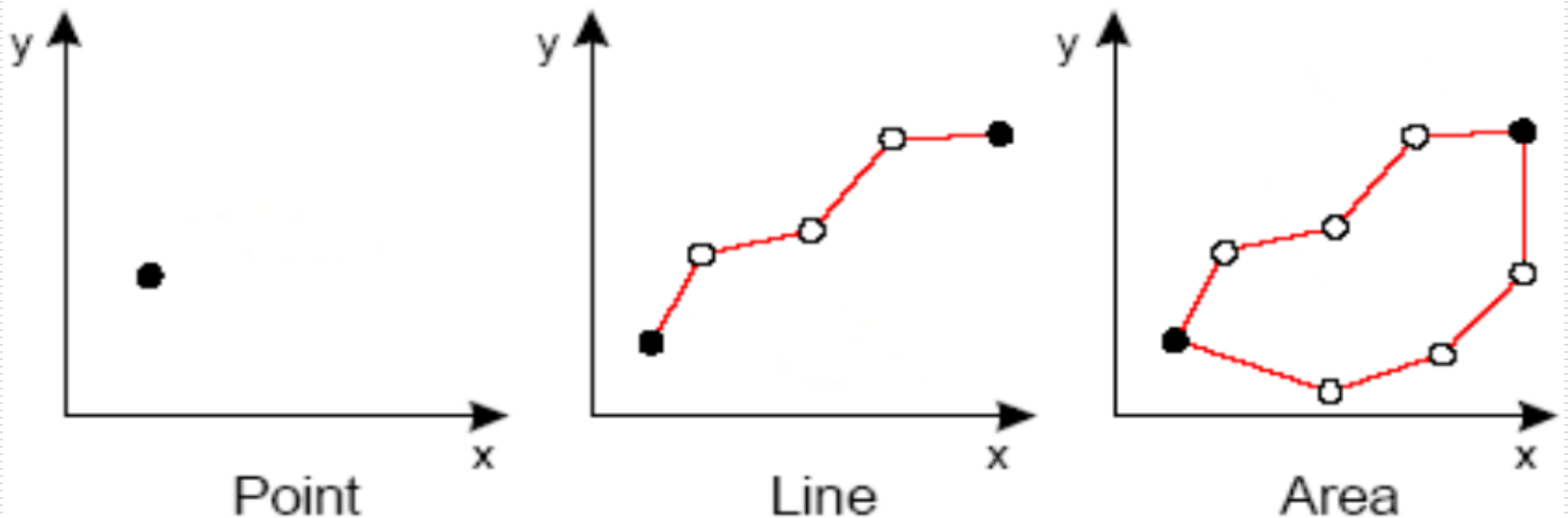
https://en.wikipedia.org/wiki/GIS_file_formats

Mô hình dữ liệu Vector

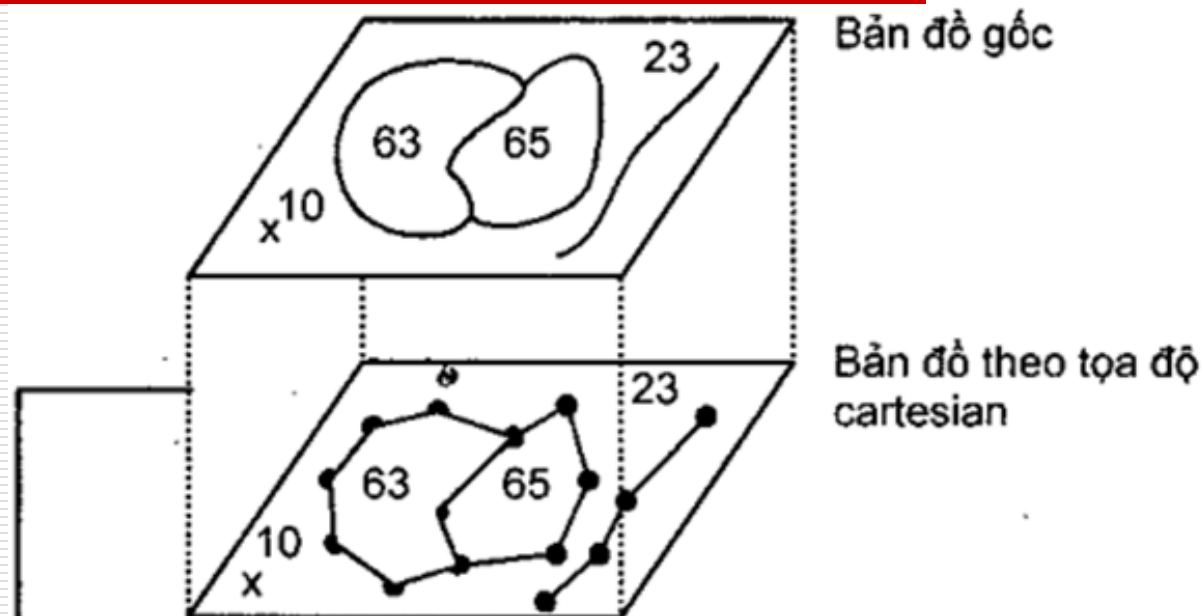
- ❖ Khái quát về mô hình Vector
- ❖ Mô hình cấu trúc dữ liệu Vector
- ❖ Các định dạng file Vector phổ biến

Khái quát về mô hình Vector

- ❖ Sử dụng các đối tượng điểm, đường, vùng để biểu diễn các thực thể không gian theo một hệ tọa độ xác định
- ❖ Mỗi điểm được xác định bởi một cặp tọa độ (x,y)
- ❖ Mỗi đường được xác định bởi một chuỗi liên tiếp các điểm $\{(x_1,y_1), (x_2,y_2) \dots\}$
- ❖ Mỗi vùng được xác định bởi những đường khép kín



Khái quát về mô hình Vector



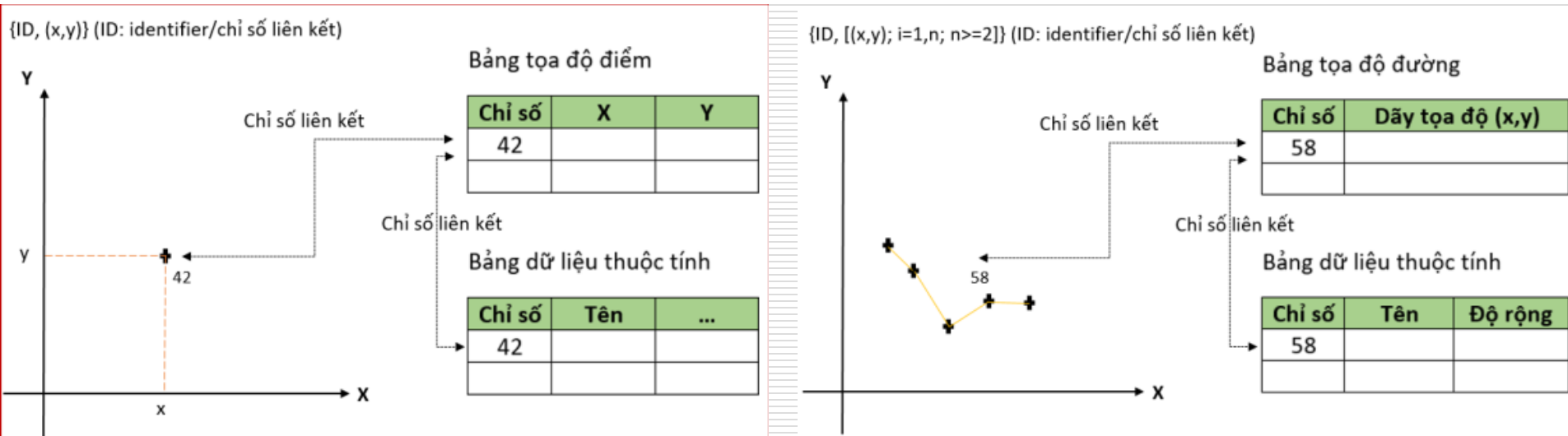
Cấu trúc dữ liệu

Đối tượng	Số	Vị trí
Point	10	$x_1 y_1$
Line	23	$x_1 y_1 x_2 y_2 \dots x_n y_n$
Polygon	63	$x_1 y_1 x_2 y_2 \dots x_n y_n x_1 y_1$
	65	$x_1 y_1 x_2 y_2 \dots x_n y_n x_1 y_1$

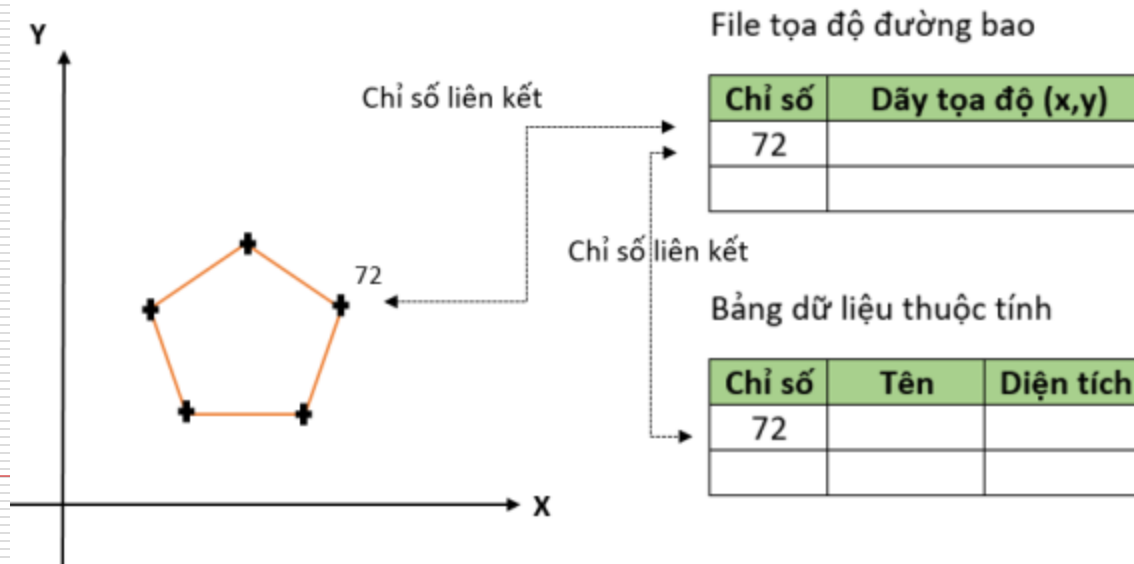
Mô hình cấu trúc dữ liệu Vector

- ❖ Tùy thuộc cách thức lưu trữ, người ta chia ra thành các mô hình:
 - ❖ Spaghetti
 - ❖ Topological
 - ❖ Trianglular Irregular Network (TIN)

Mô hình Spaghetti



$\{ID, [(x,y); i=1,n; n \geq 3; (x_1,y_1) \sim (x_n,y_n)]\}$ (ID: identifier/chỉ số liên kết)



Mô hình Spaghetti

❖ Ưu điểm:

- ❖ Đơn giản, dễ trình bày, biên tập, in ấn

❖ Nhược điểm:

- ❖ Không mô tả được mối quan hệ không gian giữa các đối tượng, vì thế rất khó khăn khi thực hiện các phép phân tích không gian
- ❖ Đối với dữ liệu vùng, vùng ranh giới giữa hai vùng được ghi nhận 2 lần, mỗi lần cho 1 vùng

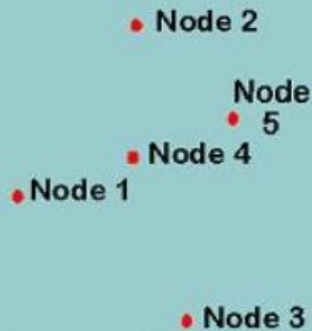
Mô hình Topological

- ❖ Thông tin về vị trí không gian
- ❖ Thông tin về quan hệ không gian:
 - ❖ Liên thông với nhau: đường – điểm nối (arc-node)
 - ❖ Kề nhau: đường bao (Polygon-Arc)
 - ❖ Nằm trong nhau, phủ nhau

<u>Quan hệ không gian</u>	<u>Đặc tính không gian</u>
<u>Mỗi đường (arc) bắt đầu bằng 1 node và kết thúc bằng node khác</u>	<u>Độ dài của đường</u> <u>Hướng đường</u>
<u>Các đường nối với nhau tại các điểm</u>	<u>Tính liên tục</u>
<u>Các đường nối với nhau tại tạo thành đường bao của vùng (polygon)</u>	<u>Tính tạo vùng (có diện tích, chu vi)</u>
<u>Các đường tham gia định nghĩa vùng cả bên trái và bên phải</u>	<u>Tính kề nhau</u>

Mô hình Topological

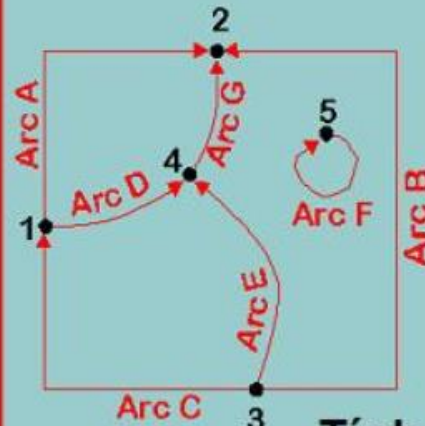
NODES



Node	Lat	Long
1	25	82
2	30	78
3	21	76
4	26	79
5	27	75

Đầu, cuối, giao của đường

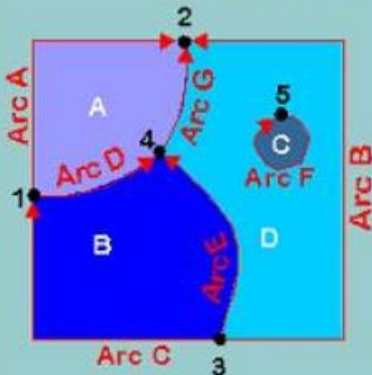
ARCS



ARC	From Node	To Node
A	1	2
B	3	2
C	3	1
D	1	4
E	3	4
F	5	5
G	4	2

Tính liên tục

CONTIGUITY

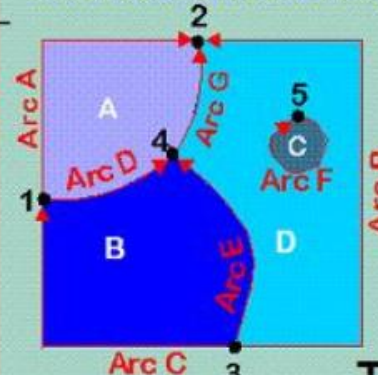


Arc	Left Poly	Right Poly
A	Universe	A
B	D	Universe
C	Universe	B
D	A	B
E	B	D
F	D	C
G	A	D

Tính kề nhau

POLYGONS

Use arc and arc direction to combine polygons



Poly	No. of Arcs	Arc List
A	3	A,D,G
B	3	C,D,E
C	1	F
D	4	B,E,G,-F

Tính tạo vùng

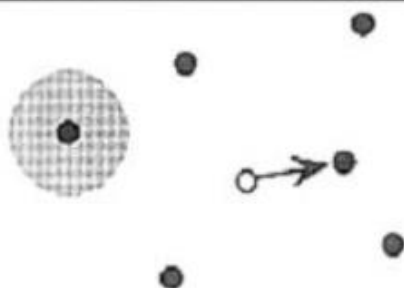
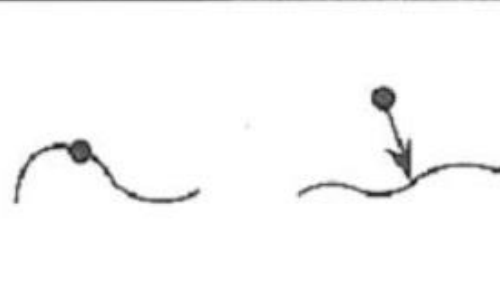
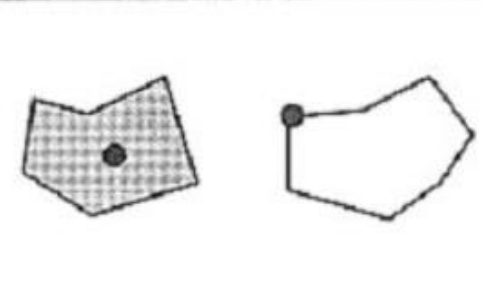
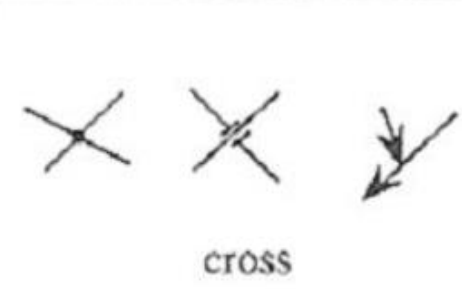
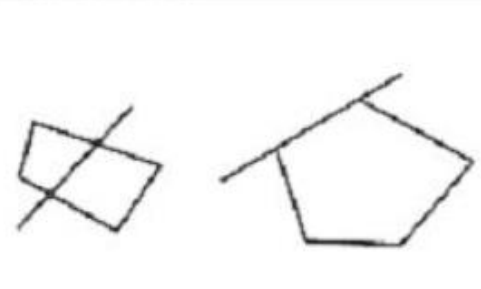
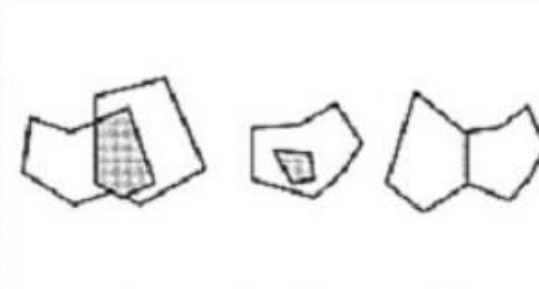
Các định dạng file Vector phổ biến

❖ https://en.wikipedia.org/wiki/GIS_file_formats

Định dạng shapefile

- ❖ Shapefile là định dạng dữ liệu không gian địa lý vector phổ biến cho các phần mềm GIS
- ❖ Được tổ chức thành các tập tin riêng rẽ, tối thiểu cần có 3 tập tin với phần mở rộng là “.shp”, “.shx”, “.dbf”
 - ❖ Tập tin có phần mở rộng là “.shp” chứa các thông tin về đặc điểm, hình dạng hình học của đối tượng
 - ❖ Tập tin có phần mở rộng là “.shx” chứa các thông tin về thứ tự của các đối tượng
 - ❖ Tập tin có phần mở rộng là “.dbf” chứa các thông tin về hàng dữ liệu thuộc tính của đối tượng
- ❖ Shapefile là một định dạng lưu trữ vector số để lưu trữ vị trí hình học và thông tin thuộc tính liên quan
- ❖ Một shapefile khi hiển thị trong phần mềm GIS được gọi là lớp dữ liệu

Mối quan hệ giữa các đối tượng không gian

Point - Point	Point - Line	Point - Area
 <p>is within nearest to</p>	 <p>on line nearest to</p>	 <p>in area on area</p>
Line - Line	Line - Area	Area - Area
 <p>intersect cross flow into</p>	 <p>intersect border</p>	 <p>overlap inside adjacent to</p>

So sánh giữa Raster và Vector

	Raster	Vector
Cấu trúc dữ liệu	Đơn giản	Phức tạp
Dung lượng	Lớn	Bé
Chất lượng đồ họa	Trung bình	Tốt
Chồng lớp	Đơn giản	Phức tạp
Phân tích mạng	Không	Dễ dàng
Độ chính xác hình học	Thấp	Cao
Topology	Khó khăn	Dễ dàng
Phân tích không gian	Đơn giản	Phức tạp

