

BUT3 FI

R5.13 - Nouveaux paradigmes de base de données

Chapitre BD-Géo

Bases de données géo-spatiales

Karine ZEITOUNI

karine.zeitouni@uvsq.fr

Edition 2023-2024

Plan du chapitre

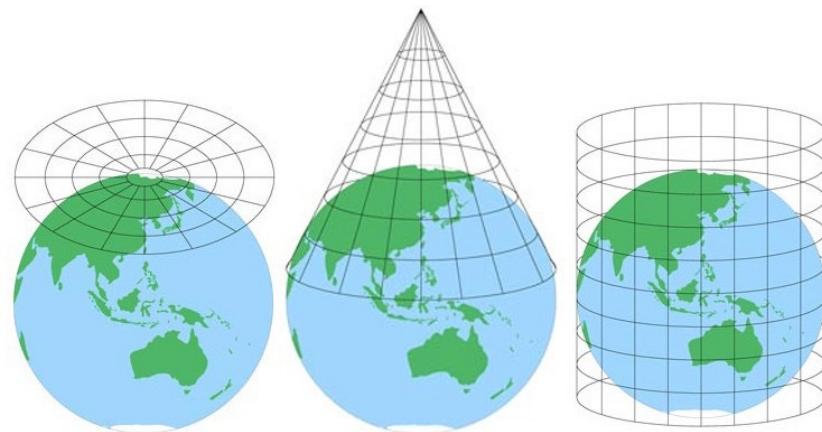
- ◆ Définitions
- ◆ Fonctions de base
- ◆ Modèle de représentation des données
- ◆ Modèle de traitement - Algèbre d'opérateurs spatiaux
- ◆ Requêtes
- ◆ Indexation spatiale
- ◆ Oracle Spatial
- ◆ Conclusion

Quelques définitions (1)

- ◆ Information géographique :
 - Information liée à une localisation sur la terre décrivant une entité physique, un phénomène ou un évènement
 - Elle est dite géolocalisée
- ◆ Caractéristiques :
 - Localisation = position et forme : coordonnées, ou position indirecte...
 - Description : attributs classiques qualitatif ou quantitatif
- ◆ Relations spatiales :
 - Tout lien topologique, métrique ou directionnel entre localisations spatiales
 - Topologique : décrit si les objets se touchent et comment ils se touchent
 - Ex : Disjoints, identiques, adjacents, l'un inclus dans l'autre, s'intersectent, ...
- ◆ Couche thématique :
 - Collection d'objets géographiques appartenant à un même thème.
- ◆ Carte :
 - Définie par une séquence (ou un empilement) de couches

Quelques définitions (2)

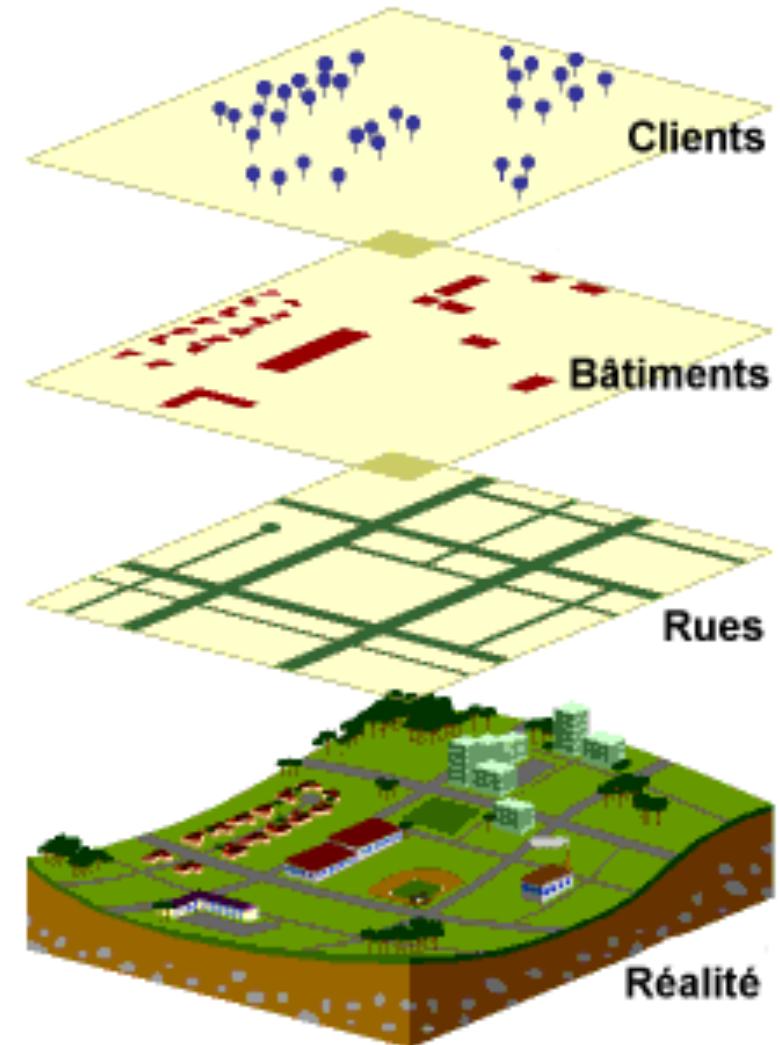
- ◆ Système de référence géographique :
 - Détermine la représentation du globe terrestre et des coordonnées géographiques
 - Se base sur un système géodésique définissant la forme ellipsoïde complexe de la terre et une méthodes de projection des coordonnées (cylindrique, conique ou azimutale) – cf. Figure
 - Ex : système géodésique WGS 84 et projection UTM (Universal Transverse Mercator) du GPS
- ◆ Méta-données géographiques :
 - Echelle, Emprise, Système de référence géographique, Qualité, Datation, ...
 - Relatifs à un ensemble (dit lot homogène) d'objets géographiques.



ICSM: 3.0

Quelques définitions (3)

- ◆ **Base de données spatiale** (ou géospatiale) :
 - Ensemble organisé d'objets géographiques où chaque objet est une association d'une description qualitative ou quantitative et d'une localisation
 - Organisé en couches thématiques
Ex : découpage administratif, Réseaux routier, Cadastre, Plan d'Occupation du Sol (POS), Topographie (courbes de niveau)...
 - Gérée au sein d'un Système d'Information Géographique (SIG)



Quelques définitions (4)

- ◆ Système de gestion de base de données spatial ou SGBD spatial :
 - ✓ C'est un SGBD qui offre au moins un **type de donnée spatial** dans son modèle de données,
 - ✓ Qui offre **des opérateurs spatiaux**, ainsi qu'**un langage de requêtes** permettant d'utiliser ce type et ces opérateurs,
 - ✓ Et ce de manière efficace.
- ◆ Il intègre des mécanismes d'optimisation tels que :
 - ✓ Un **index spatial**
 - ✓ Des algorithmes avancés pour ces opérateurs tels que la **jointure spatiale**

Exemple d'utilisation d'un type « spatial »

□ Un thème ou couche thématique :

- ◆ peut être vu comme une table avec un attribut géométrique

Table Routes

Id_route	Nom_route	Type_Surface	Nb_voies	Localisation
1	Av. Morane Saulnier	Goudronnée	4	A red circle representing a circular road route.
2	Rue Dewoitine	Goudronnée	3	A red U-shaped curve representing a road with a sharp bend.
3	Av. Europe	Goudronnée	1	A red zigzag line representing a winding road.

Système d'Information Géographique SIG

◆ Fonction d'un SIG (les 5 A) :

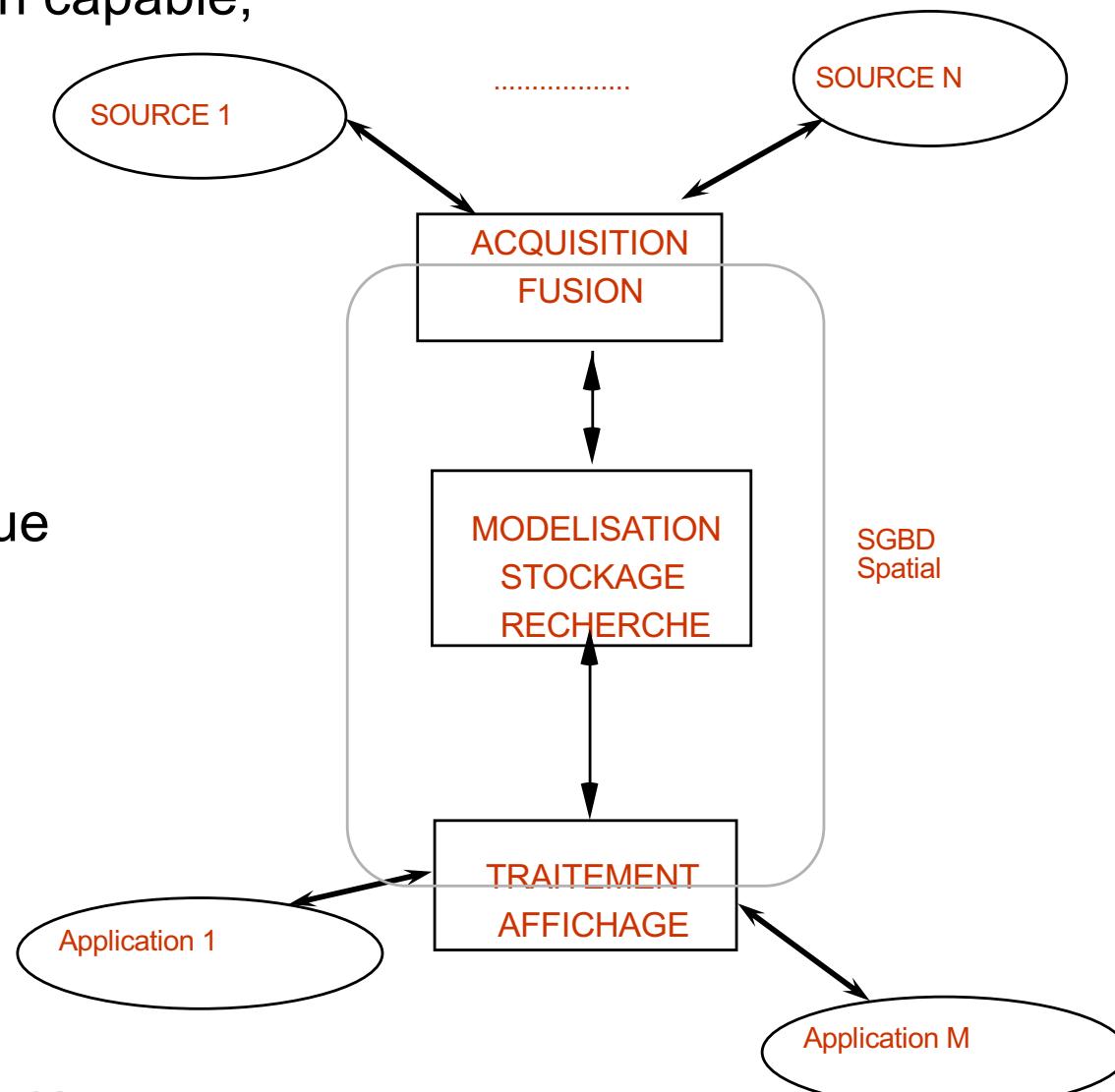
C'est un Système d'Information capable,

- d'Acquérir,
- d'Assembler,
- d'Archiver,
- d'Analyser
- et d'Afficher

de l'information géographique

◆ La suite

- ❖ Passe en revue
- ❖ Les 5 A
- ❖ En détaillant le SGBD



Plan du chapitre

- ◆ Définitions
- ◆ Fonctions de base
- ◆ Modèle de représentation des données
- ◆ Modèle de traitement - Algèbre d'opérateurs spatiaux
- ◆ Requêtes
- ◆ Indexation spatiale
- ◆ Oracle Spatial
- ◆ Conclusion

Acquisition des données

- ◆ Sources :

- cartes papiers, photos aériennes, images satellites (SPOT, LANSAT),
 - relevés sur le terrain, GPS, ...
 - ou simples adresses

- ◆ Edition / correction de cartes :

- table à digitaliser, scanner, photogrammétrie, traitement d'images
 - **Géocodage** : Processus automatique permettant de transformer des adresses en coordonnées absolues sur une carte. (ex. Google Map)

- ◆ Difficultés :

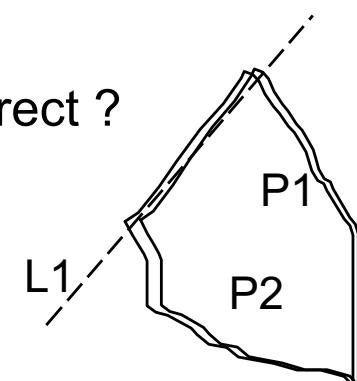
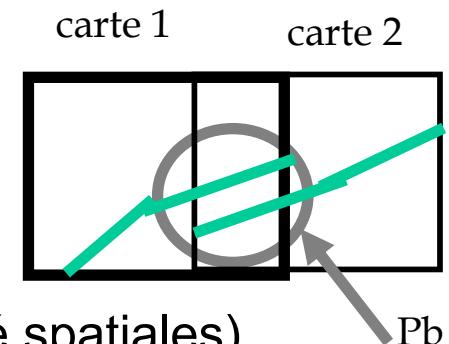
- processus long (ex: la BD Topo 87->2007) et coûteux
 - qualité difficile à évaluer et la précision dépend de l'échelles de mesures (ex: 10 à 50 m pour la BD Carto, 5 à 10m pour GéoRoute)
 - difficultés augmentant si mises à jour fréquentes
 - Ambiguïté des adresses lors du géocodage e dépendance aux adresse pré-localisées

Acquisition / Echange des données

- ◆ Echange = Conversion d'un format source de représentation des données au format cible (concerne les données et les métadonnées)
- ◆ Utilise des standards d'échange des données
 - Principaux formats d'échanges :
 - **Shapefile** (des SIG ArcInfo-ArcGIS d'ESRI) -> STANDARD DE FAIT
 - DIGEST-VPF, DLMS (norme de l'OTAN)
 - DLG & SDTS (USGS -US Geological Survey-)
 - NTF (GB)
 - GDF (dérivé de NTF - Centre Européen de Rech. en Carto. Officielle)
 - SAIF (développé au Canada)
 - **GML** (format XML de l'OGC – Open Geospatial Consortium)
 - **KML** (format XML de l'OGC, orienté visualisation, adopté par Google map)
 - GeoJSON (encodage standard avec JSON)
 - Outils
 - Les SIG offrent souvent des outils d'import/export dans les formats (exemple shp2oracle ou shp2pgsql)
 - Outil de conversion universel : FME (commercial), GDAL-OGR/ ogr2ogr (gratuit)

Assemblage / Fusion des données

- ◆ Fusion horizontale :
 - Juxtaposition de cartes
 - Problème sur les bords (pour respecter la continuité spatiales)
- ◆ Fusion verticale :
 - **Appariement d'objets**
 - Problème d'imperfection géométrique – lequel est correct ?
 - Problèmes amplifiés si données non uniformes
 - Cartes de formats numériques différents
 - Cartes à des échelles différentes (appariement)
 - Cartes imprécises (qualité)
=> pré-traitements importants



Archivage

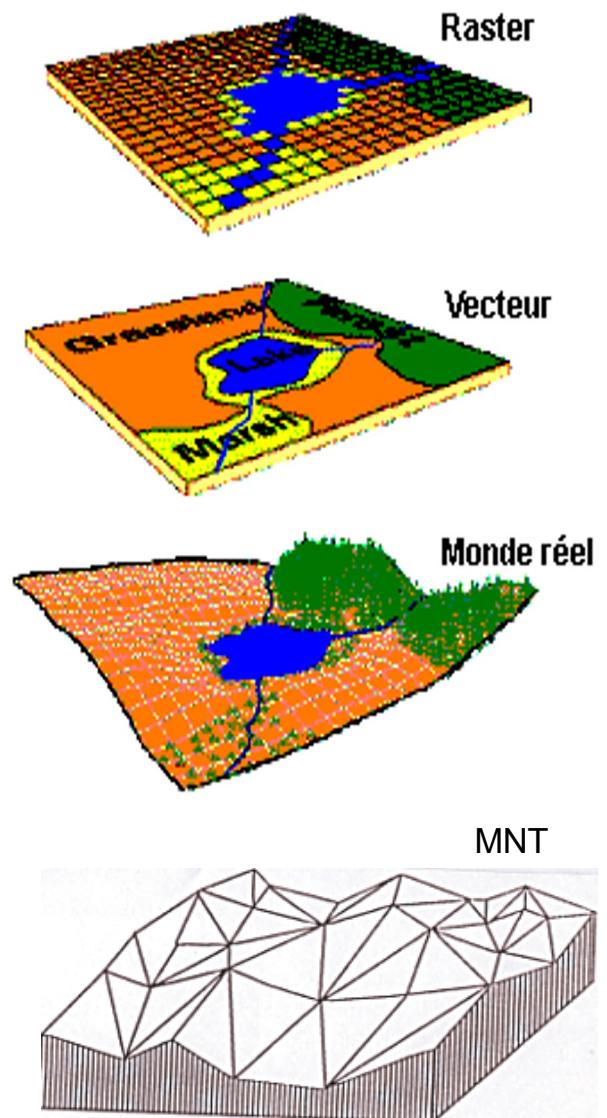
- ◆ C'est principalement le rôle du SGBD
- ◆ Apports d'un SGBD spatial:
 - Différents niveaux d'abstraction
 - Modélisation de données
 - Outils d'interrogation et de développements intégrés
 - Langage de requêtes
 - Facilité et puissance d'expression
 - Gestion de grandes bases de données
 - Persistance
 - Fonctionnalités / Efficacité
 - Opérateurs / optimisation
 - Sécurité, partage, fiabilité
 - Transactions, concurrence
 - Interopérabilité

Plan du chapitre

- ◆ Définitions
- ◆ Fonctions de base
- ◆ Modèle de représentation des données
- ◆ Modèle de traitement - Algèbre d'opérateurs spatiaux
- ◆ Requêtes
- ◆ Indexation spatiale
- ◆ Oracle Spatial
- ◆ Conclusion

Archivage / Modèles de représentation

- Modèle vecteur (prédominant)
Objets représentés par un point, une poly-ligne (routes, fleuves...) ou un polygone (lac, parcelle...)
- Modèle raster (continu)
Espace représenté par une matrice de cellules (pixel associée à des mesures (températures, couleur...))
- Triangulation (continu)
Ensemble de points de mesure reliés par des faces triangulaires (ex. modèle numérique de terrain (MNT))



Archivage / Modèles de représentation

➤ Modèle par référence

Il s'agit de référence géographique indirecte

- Soit par adresse + Géocodage
- Soit par position relative à un objet existant

➤ Exemple: Linear Referencing

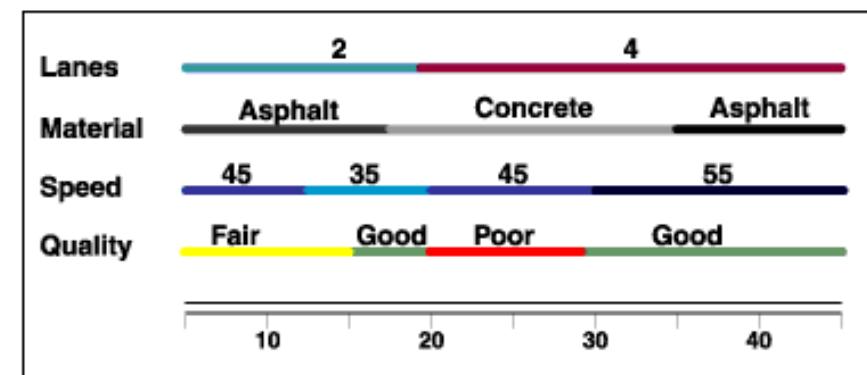
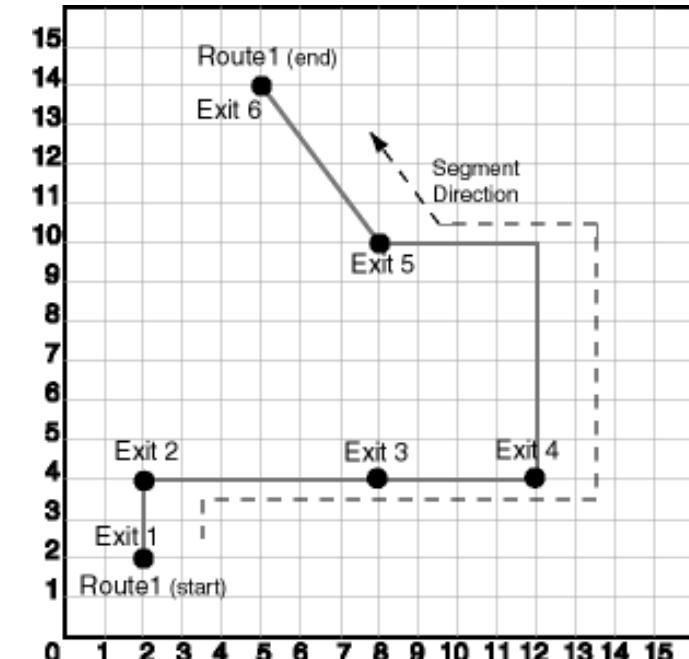
Mode de représentation par position relative à un objet linéaire existant. Des conversions existent avec les coordonnées absolues.

ex. sortie d'autoroute, position d'un bâtiment dans une rue.

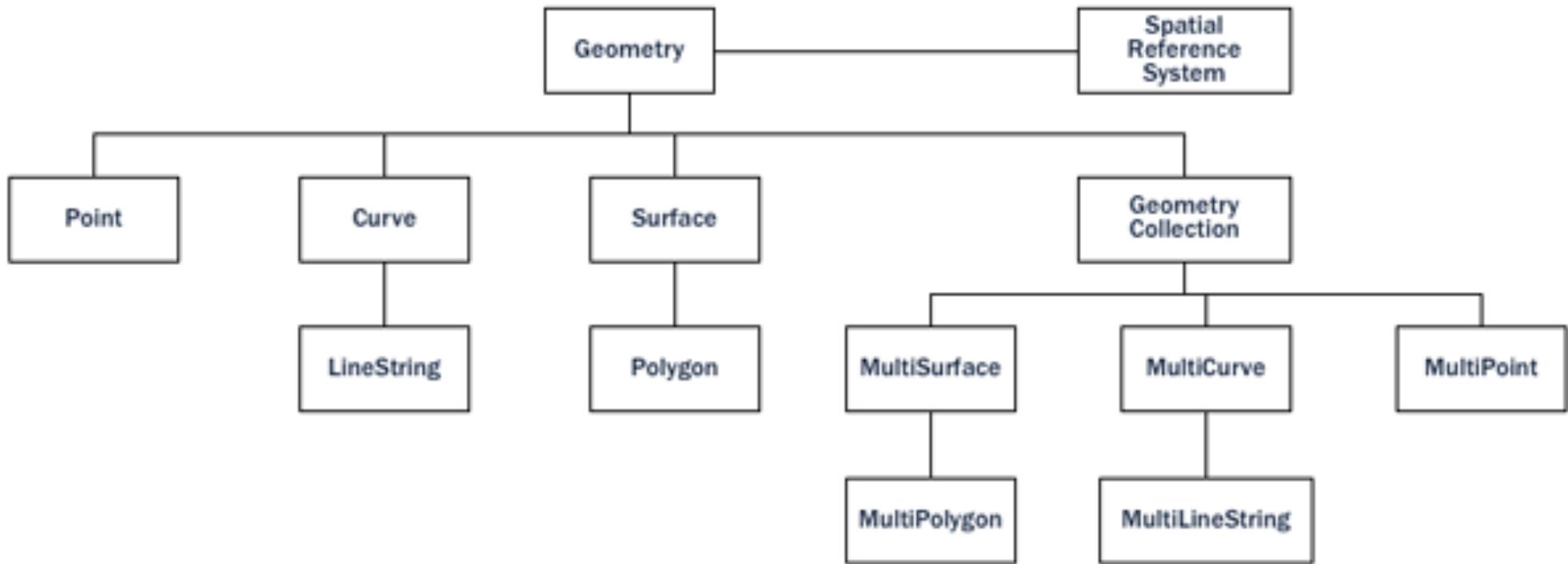
Implémenté dans la plupart de SIG :

Oracle LRS / ArcGIS / PostGIS ...

ISO 19148 – 2012 Standard.



Archivage / Modèle vectoriel du standard OGC



**Hiérarchie de types spatiaux
Standard Simple Feature Types de l'OGC(*)**

(*) <http://www.opengeospatial.org/standards/sfs>

Plan du chapitre

- ◆ Définitions
- ◆ Fonctions de base
- ◆ Modèle de représentation des données
- ◆ Modèle de traitement - Algèbre d'opérateurs spatiaux
- ◆ Requêtes
- ◆ Indexation spatiale
- ◆ Oracle Spatial
- ◆ Conclusion

Modèle de traitements

- ◆ 2 niveaux de manipulation :
 - Niveau des objets spatiaux : opérations élémentaires unaires ou binaires :
 - Prédicats : relations topologiques ou métriques (ex. distance < 100)
 - ou fonctions (calcul d'intersection, fusion)
 - Niveau des collections d'objets : opérateurs ensemblistes ou navigationnels
 - Algèbre d'opérateurs spatiaux
 - utilise les opérations élémentaires

Niveau objets - Relations topologiques

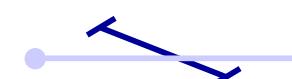
disjoints



adjacents



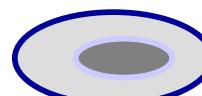
intersection



inclusion avec adjacence



inclusion sans adjacence

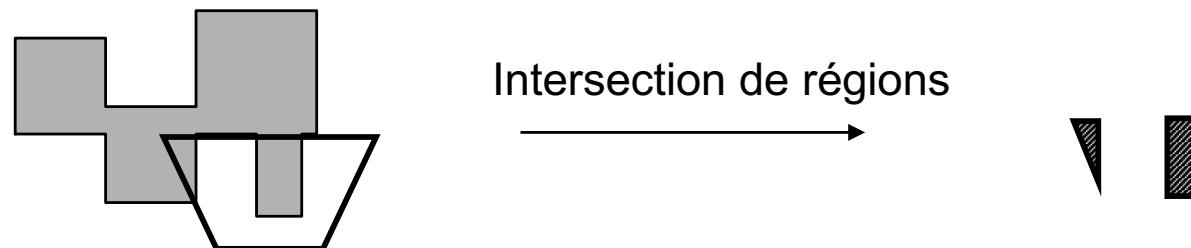


égaux

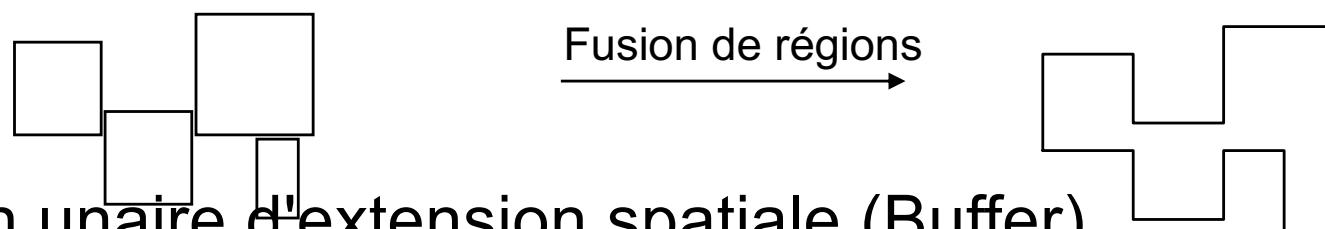


Niveau objets - Fonctions spatiales

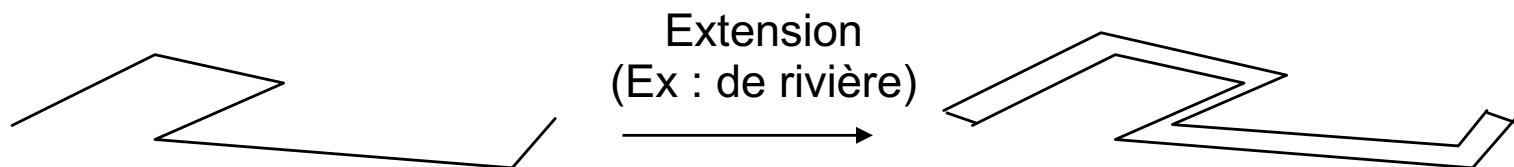
- ◆ Fonction binaire d'intersection spatiale



- ◆ Fonction agrégat fusion spatiale

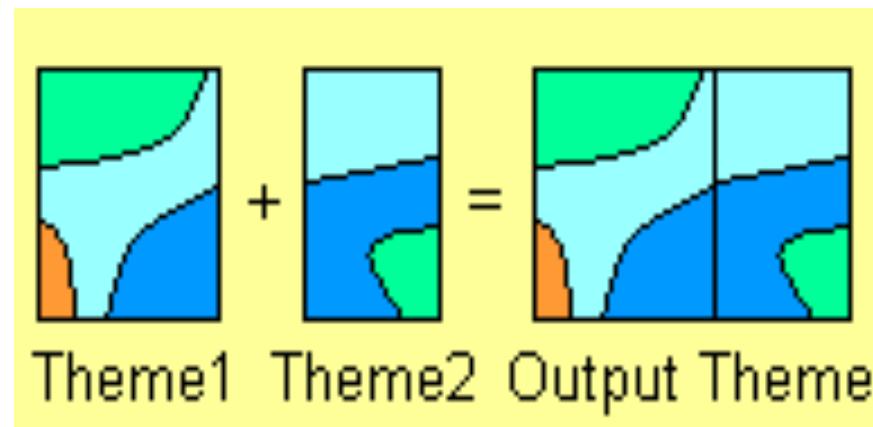


- ◆ Fonction unaire d'extension spatiale (Buffer)



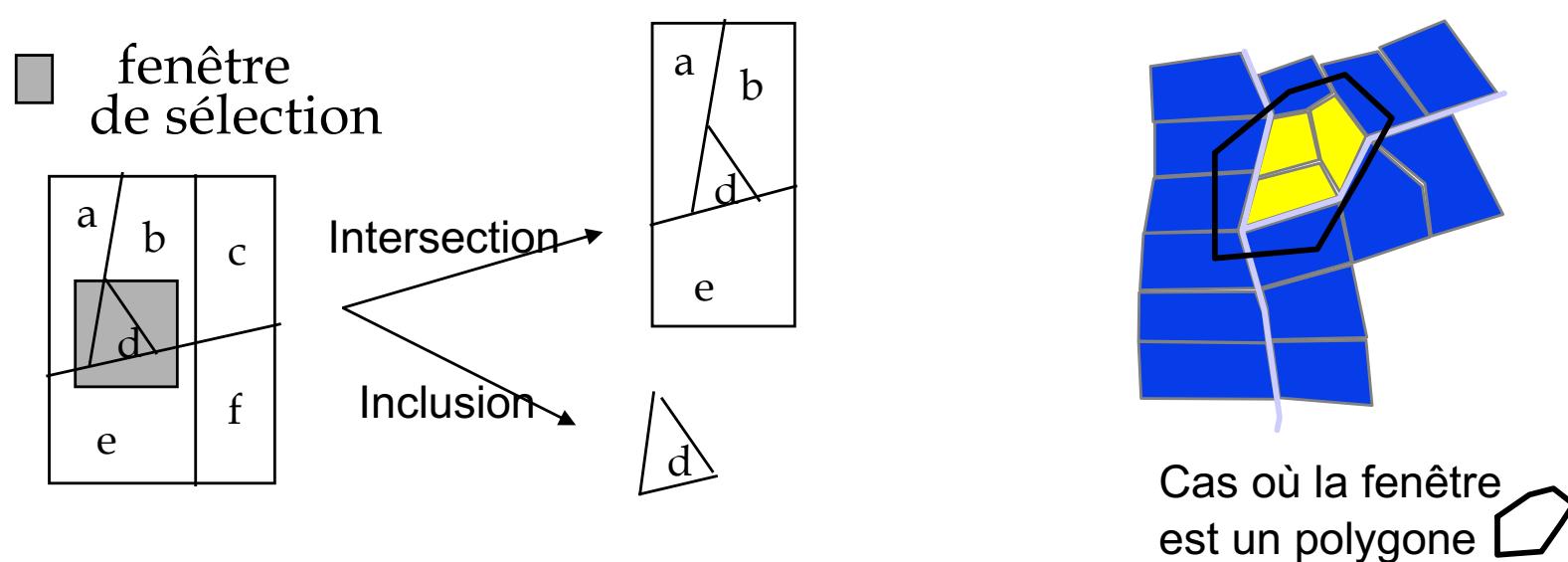
Niveau collection - Algèbre spatiale (1)

- ◆ Extension de l'algèbre relationnelle aux données spatiales :
 - Opérateurs d'union, intersection et différence
 - Sélection, projection, jointure et agrégats
 - Opérateurs de fermeture transitive (navigation : le plus court chemin)
- ◆ Exemple d'union :
 - de deux thèmes ayant les mêmes attributs



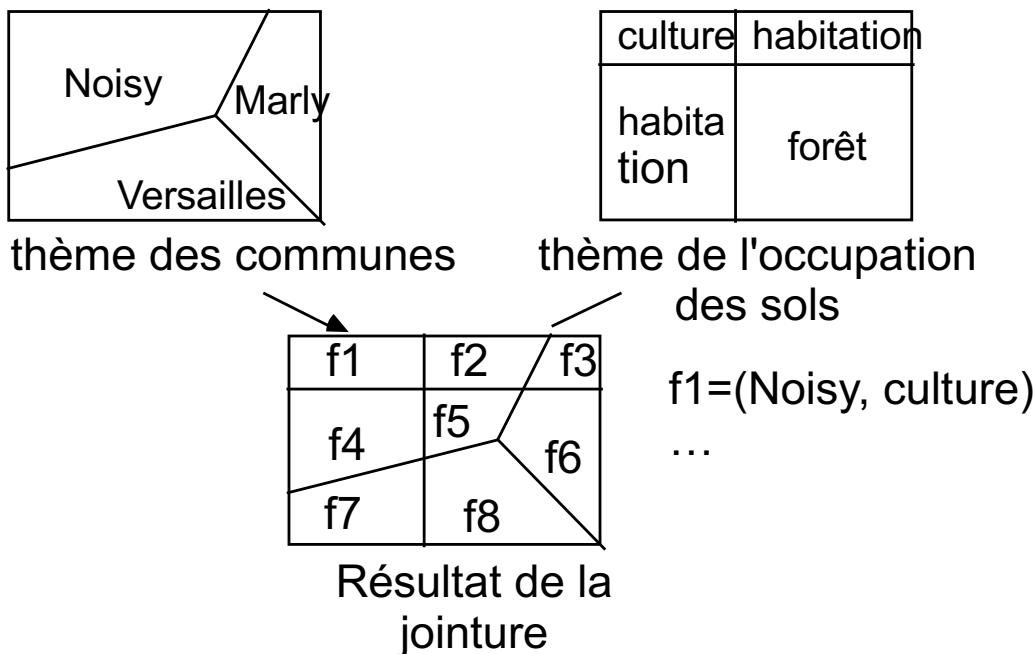
Niveau collection - Algèbre spatiale (2)

- ◆ Opérateur de sélection spatiale :
 - Sélection du sous-ensemble d'objets spatiaux par comparaison avec une constante
 - ⇒ application d'un **prédicat spatial** (inclusion, intersection,...) **entre les objets et une valeur spatiale donnée** (point, ligne ou région). On parle de fenêtre de sélection.



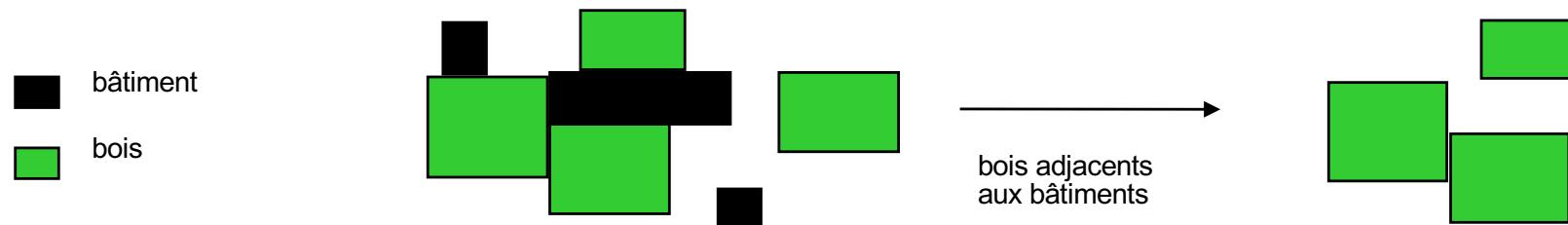
Niveau collection - Algèbre spatiale (3)

- ♦ Jointure spatiale :
 - Mise en relation de plusieurs collections sur critère spatial
- ♦ Exemple de jointure par intersection (map overlay)



Niveau collection - Algèbre spatiale (4)

- ♦ Jointure spatiale par adjacence (ici une semi-jointure)



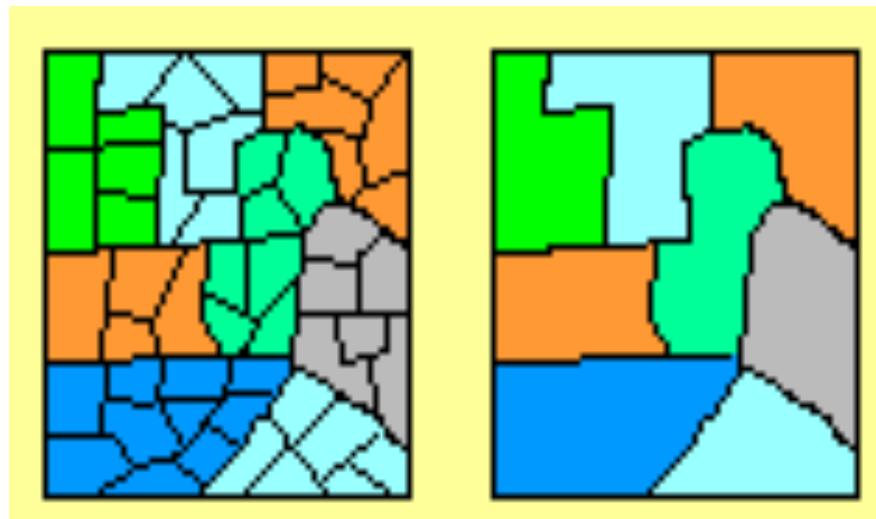
- ♦ Jointure sur critère de distance (ici une semi-jointure)



Niveau collection - Algèbre spatiale (5)

- ◆ Fonction d'agrégat

- Partitionnement selon un sous-ensemble d'attributs et pour chaque groupe (partie), application d'une fonction agrégat.
- Ex : Fusion de zones adjacentes



Plan du chapitre

- ◆ Définitions
- ◆ Fonctions de base
- ◆ Modèle de représentation des données
- ◆ Modèle de traitement - Algèbre d'opérateurs spatiaux
- ◆ Requêtes
- ◆ Indexation spatiale
- ◆ Oracle Spatial
- ◆ Conclusion

Langage de requêtes SQL

Description de données

- Approche SQL étendu : continuité du standard d'accès
 - Extension de SQL aux domaines et opérateurs spatiaux
 - Intégré dans le standard SQL3
- Comprend :
 - Langage de description de données
 - Langage de manipulation de données
 - Spécification des méta-données (systèmes de référence)
 - Déclaration de méthodes d'accès

```
create table commune
(   code      integer,
    nom       character (20),
    departement integer,
    loc        geometry )
```

Langage de requêtes SQL

Manipulation de données

◆ Typologie des requêtes

■ Sélection

- Ville de plus de 10 000

```
select c.nom, c.loc  
from commune c  
where c.population > 10000
```

■ Sélection spatiale

- Accès aux objets situés dans une fenêtre donnée

```
select c.nom, c.loc  
from commune c  
where INTERSECT (c.loc, &tel_polygone)
```

■ Accès par relation à un objet

- Communes au bord de le N10

```
select c.*  
from commune c, route r  
where r.nom = 'N10'  
and TOUCH (c.loc, r.loc)
```

■ Association d'objets de 2 thèmes

- Zones d'intersection des communes et des forêts (map overlay)

```
select c.nom, f.nom, Intersection (c.loc, f.loc)  
from commune c, foret f  
where INTERSECT (c.loc, f.loc)
```

■ Agrégation

- Fusion des communes par département pour calculer les frontières des départements

```
select departement, UNION (loc)  
from commune  
group by departement
```

Opérateurs du standard OGC

OGC spatial operators defined on the class geometry

<i>Classes</i>	<i>Operators</i>	<i>Operator Functions</i>
Basic Operators	Spatial Reference	Returns the reference system of the geometry
	Envelope	Returns the minimum bounding rectangle of the geometry
	Export	Converts the geometry into a different representation
	IsEmpty	Tests if the geometry is the empty set or not
	IsSimple	Returns TRUE if the geometry is simple
	Boundary	Returns the boundary of the geometry
Topological Operators	Equal	Tests if the geometries are spatially equal
	Disjoint	Tests if the geometries are disjoint
	Intersect	Tests if the geometries intersect
	Touch	Tests if the geometries touch each other
	Cross	Tests if the geometries cross each other
	Within	Tests if a geometry is within another geometry
	Contain	Tests if a given geometry contains another geometry
	Overlap	Tests if a given geometry overlaps another given geometry
	Relate	Returns TRUE if the spatial relationship specified by the 9-Intersection matrix holds
Spatial Analysis Operators	Distance	Returns the shortest distance between any two points of two given geometries
	Buffer	Returns a geometry that represents all points whose distance from the given geometry is less than or equal to a specified distance
	ConvexHull	Returns the convex hull of a given geometry
	Intersection	Returns the intersection of two geometries
	Union	Returns the union of two geometries
	Difference	Returns the difference of two geometries
	SymDifference	Returns the symmetric difference (i.e. the logical XOR) of two geometries

Plan du chapitre

- ◆ Définitions
- ◆ Fonctions de base
- ◆ Modèle de représentation des données
- ◆ Modèle de traitement - Algèbre d'opérateurs spatiaux
- ◆ Requêtes
- ◆ Indexation spatiale
- ◆ Oracle Spatial
- ◆ Conclusion

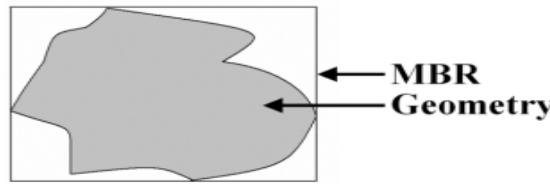
Indexation spatiale (1)

- ◆ Problème :
 - Comment obtenir un débit transactionnel acceptable ?
- ◆ Spécificités de la sélection et de la jointure spatiale :
 - Comparateur basé sur des tests géométriques coûteux
 - Accès aux objets dans une fenêtre => objets spatialement proches
=> en 2D à l'accès par intervalle, mais pas d'ordre total.
 - Données plus volumineuses
 - Nécessitent des algorithmes et structures de données adaptées
- ◆ Rôle de l'index spatial :
 - Limitation des comparaisons géométriques (temps CPU)
 - Organisation (placement) des données par proximité géométrique, de manière à minimiser les E/S

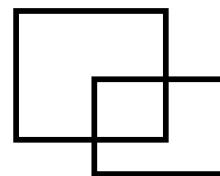
Indexation spatiale (2)

- ♦ L'objet spatial est **complexe** => on utilise souvent son Rectangle Englobant minimum (*Minimum Boundig Rectangle - MBR*) pour dégrossir.

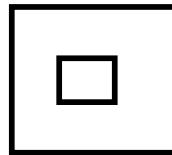
- Exemples



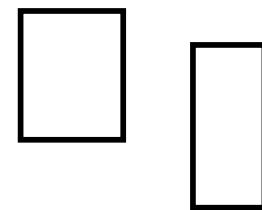
- ♦ Intérêt :
 - L'intersection se ramène de simples tests sur xmin, xMax, ymin, yMax des MBR



ou



les objets pourraient s'intersecter



les objets ne s'intersectent pas

- ♦ Conséquence
 - L'index retourne des objets candidats et non le résultat définitif
=> joue un rôle de 1er filtre

Indexation spatiale (3)

- ◆ \exists 3 approches :
 - Index avec recouvrement (*overlapping spatial index*)
 - Regroupement des MBR en MBR plus grands, avec des intersections possibles
 - Exemples : R-tree (extension du B-tree)
 - Index sans recouvrement (*non overlapping spatial index*)
 - Partitionnement de l'espace en rectangles disjoints
 - Exemple : R⁺-tree (inspiré du R-tree), kd-tree
 - Index indirect en associant un code numérique aux objets :
 - On calque une grille dans l'espace 2D dont les cellules sont numérotées et ordonnées. Cas du codage linéaire de Peano (Z-order) ;
 - L'objet est associé aux cellules de la grille et recherché en passant par les numéros des cellules (plus faciles à retrouver pour une zone de recherche).

Approche 1 : R-tree

- ◆ R-tree :
 - inspiré du B-tree, où les MBR des nœuds parents incluent ceux des fils. La descente dans l'arbre est l'inclusion des MBR !
- ◆ Un R-tree est un arbre tel que :
 - (1) Les données initiales de type (MBR de OBJ, @OBJ) où OBJ est un objet spatial quelconque, sont regroupées dans des pages disque constituant les feuilles de l'arbre. (@ désigne un pointeur)
 - (2) Un nœud non feuille regroupe des informations de type (MBR, @NŒUD) où MBR inclut les MBR du nœud fils.
 - (3) Le nombre de rectangles regroupés dans un nœud autre que la racine est compris entre une borne inférieure et une borne supérieure.

Structure du R-Tree

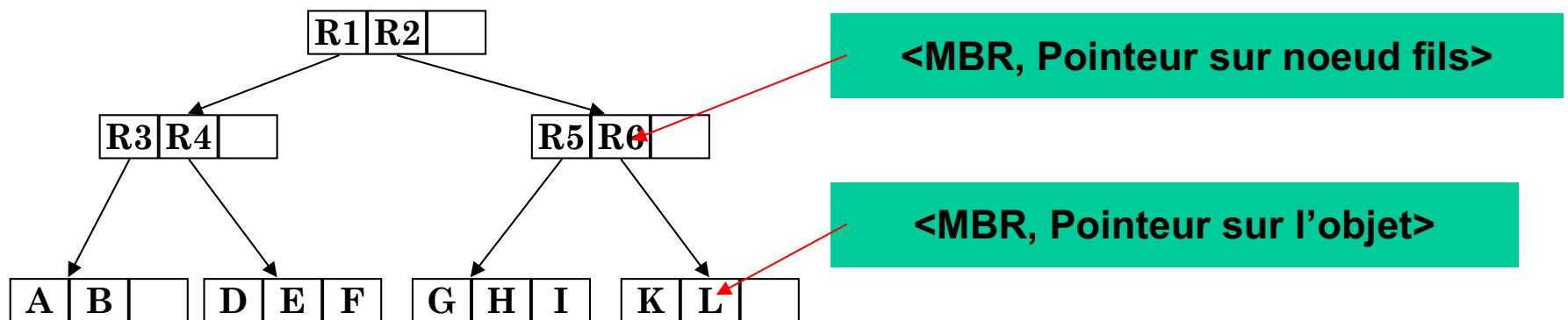
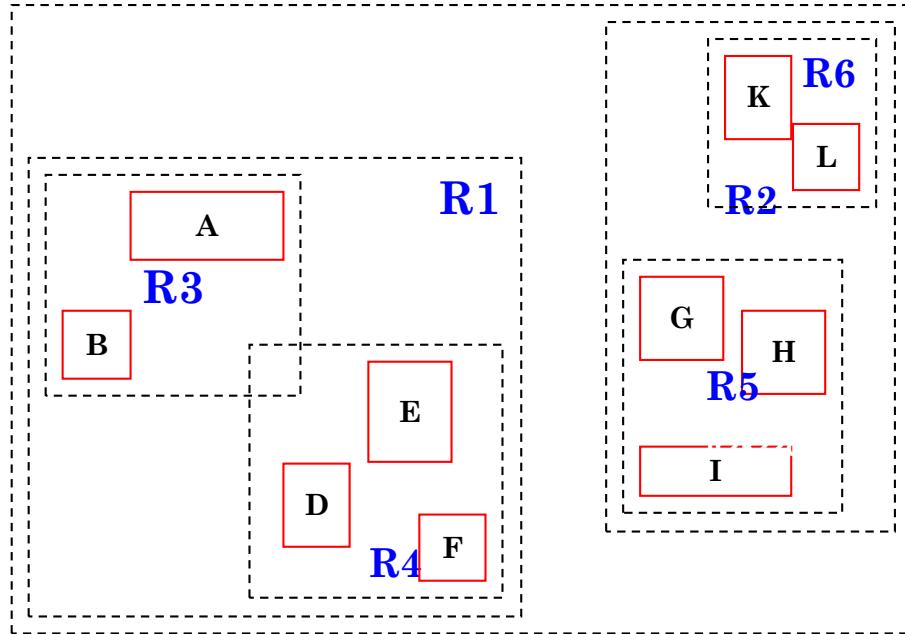
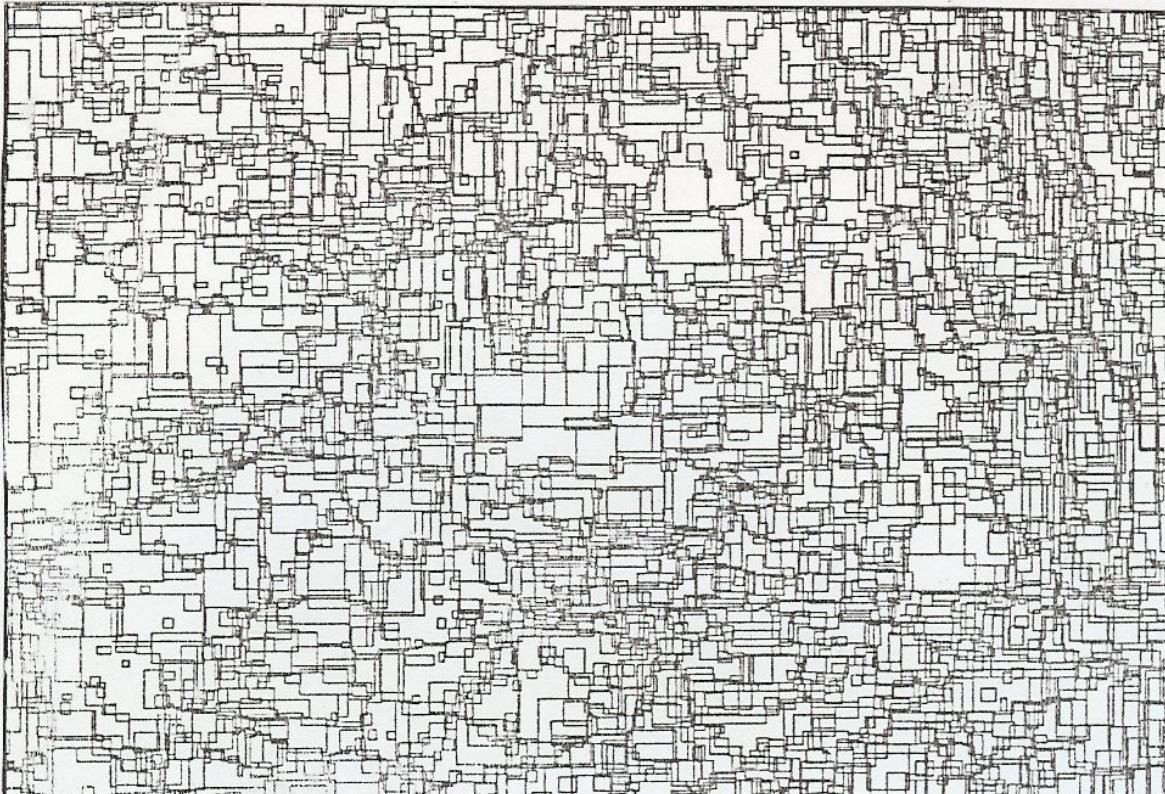
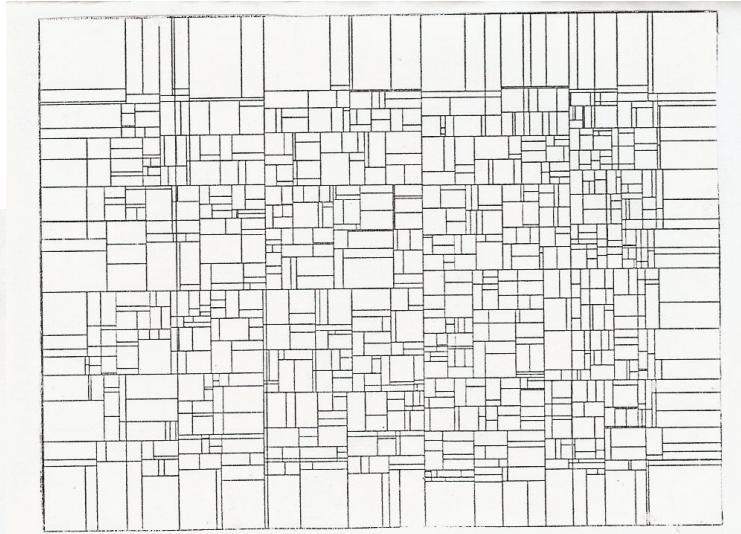


Illustration sur un cas réel

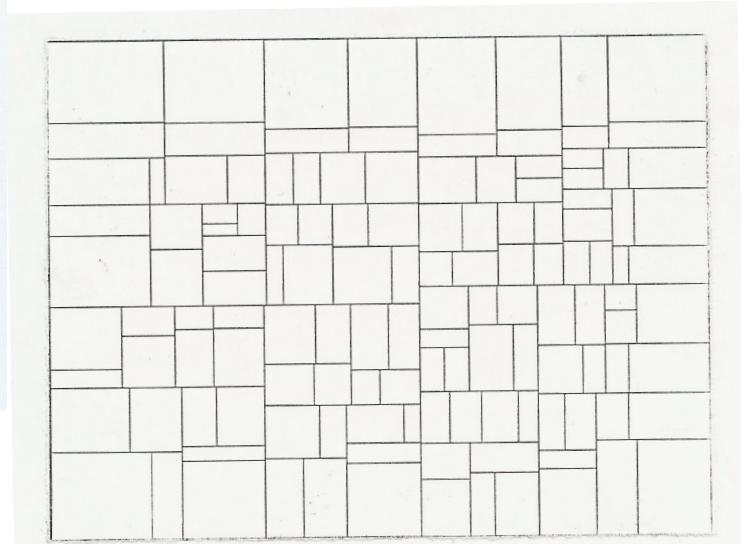
ici niveau d'un arbre R+ tree



MBR des Objets



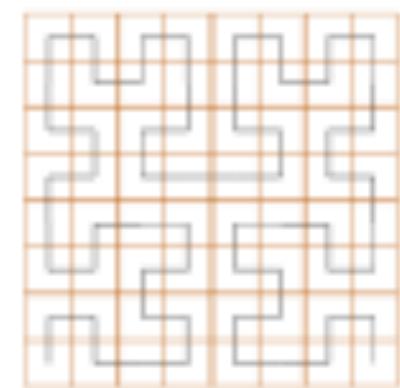
MBR des feuilles



MBR des nœuds parents

Approche 3 : Codage linéaire selon une grille

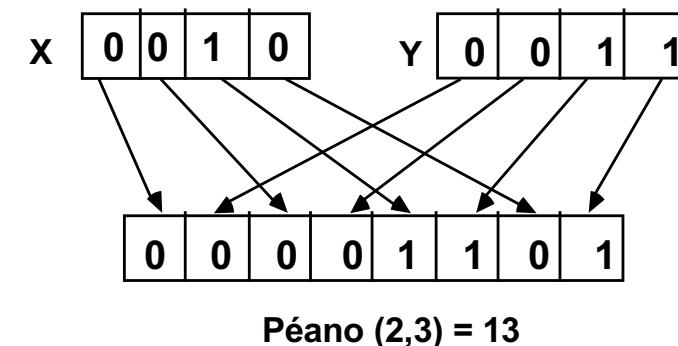
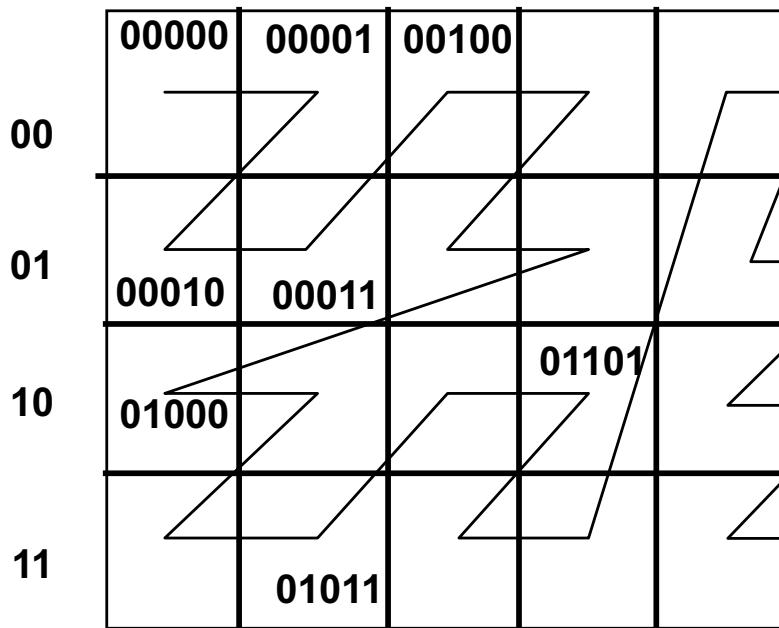
- ◆ Problème du domaine spatial :
 - Pas d'ordre total évident sur les objets dans **un espace 2D**
- ◆ Solution :
 - Définir un ordre total artificiel en discrétisant l'espace 2D selon une grille, puis numérotation des cellules de cette grille
 - Ceci forme une **courbe de remplissage** du plan (*filling curve*)
 - Ensuite affecter les numéros de cellules aux objets localisés dedans
 - Il existe différentes façons d'ordonner les cellules
 - En Z, en N, Hilbert
 - Ont une propriété fractale, permettant plusieurs niveaux de granularité



Courbe de Hilbert

Approche 3 : Codage linéaire selon une grille

- ◆ Idée : Identification par les cellules d'une grille
 - L'ordre en Z (courbes en Z) s'obtient facilement :
 - par une fonction d'entrelacement des codes binaires de X et Y
 - On pré-calcule l'intersection des objets avec les cellules de la grille
 - on associe dans la base de données l'objet et le numéro de cellule (ajout d'un attribut ou d'une table de correspondance)
 - Il suffit d'indexer la table selon cet attribut avec un simple arbre-B.



On remarque que le préfixe tel que 000 est commun aux 4 cellules adjacentes et le préfixe 001, 010, puis 011 forment un Z.
=> Propriété fractale

Aperçu de quelques produits SIG

	Commerciaux	Libres
Produits desktop	ArcGIS (ESRI), MapInfo, Géoconcept GIS (Fr)	QGIS, GDAL, Deegree
Serveurs SGBD spatiaux	Oracle Spatial, MS SQL Server	Postgres/PostGis, MySql
Orientés analyse & rasters	ArcGIS Analyst	GRASS, IDRISI, GeoNetwork
API Web mapping	ArcGIS	GeoServer, MapServer, GeoTool, Deegree (côté serveur) OpenLayer (côté client) Leaflet (côté client et mobile)
Cloud GIS	Map2Net (Fr), ArcGIS Online, GIS Cloud, CartoDB, MapInfo Stratus	MapBox, Mango Map,... OpenStreetMap, Google Map

Plan du chapitre

- ◆ Définitions
- ◆ Fonctions de base
- ◆ Modèle de représentation des données
- ◆ Modèle de traitement - Algèbre d'opérateurs spatiaux
- ◆ Requêtes
- ◆ Indexation spatiale
- ◆ Oracle Spatial
- ◆ Conclusion

Oracle Spatial (cf. Document dédié)

- ❑ C'est une extension prédéfinie d'Oracle qui offre les types Geometry ainsi que les fonctions spatiales associées.
- ❑ Implémente le standard *Simple Features Specification* (*) disponibles dans SQL (ST_xx), standard de l'OGC et ISO SQL-MM Spatial
- ❑ ajoute des fonctions d'analyse spatiales telles que SDO-NN.
- ❑ implémente l'indexation par R-Tree.
- ❑ Seul inconvénient comme dans les autres SGBD spatiaux
 - ◆ Ne fournit pas nativement les fonctions de rendu et d'intéraction cartographique.
 - ◆ Des outils comme QGIS ou Geoserver permettent cette visualisation depuis un SGBD, mais également depuis des fichiers comme le format shapefile.

(*) <http://www.opengeospatial.org/standards/sfs>

Conclusion

- ◆ De nombreux produits aujourd’hui
- ◆ Basés sur des standards de l’OGC :
 - pour l’échange de données (GML, WKT, KML)
 - Pour les services web (WFS, WMS)
- ◆ Tous étendent les SGBD relationnels :
 - Par l’ajout de types et de librairie de fonctions prédéfinies
 - Par l’ajout de méthodes d’indexation et l’extension de leur optimiseur
 - Illustration du modèle Objet-Relationnel pour représenter un type abstrait géométrie (natif)

Parallèle BD relationnelles – Spatiales

- ◆ Cette comparaison montre que :
 - les SGBD Géographiques sont spécifiques,
 - mais ils peuvent être vus comme une extension des SGBD relationnels

	RELATIONNEL	SPATIAL
Données	Entier, Réel, Texte, ...	Plus complexes: Point, Ligne, Région ...
Prédicats et calculs	Tests : $=, >, \dots$ Calculs : $+, /, \dots$ et fonctions simples	Prédicats et calculs géom. et topologiques: Tests : intersecte, adjacent à, ... Fonctions géom. : intersection, surface...
Manipulation	Opérateurs de l'algèbre : Sélection, Projection, Jointure... Agrégats : Count, Sum, Avg...	Manipulation par thème ou inter-thèmes Sélection et jointure sur critère spatial Agrégats : fusion d'objets adjacents
Liens entre objets	Par clés de jointures	Liens spatiaux (souvent) implicites
Méthodes d'accès	Index B-tree, hachage	Index R-tree, quad-tree, etc.