

Au cœur du monde connecté, les défis de l'IOT :
Réseaux, Cloud, Plateformes, Big Data, Sécurité, Blockchain, 5G....

Objets Géomatiques

Le 8 octobre 2019

BASSET Jean-Claude

LES BASES DE DONNÉES GÉOMATIQUES

- I. Représentation de l'information géographique
- II. Sources des données géographiques et spatiales
- III. Stockage de l'information (images et vecteur)
- IV. Bases de données relationnelles SGBD-SQL
- V. Bases de données NoSQL



Le Big Data et la géographie

Les données géolocalisées sont aussi concernées par le Big Data.

Elles présentent les mêmes caractéristiques que les autres données :

- la grande quantité
- le type de données non structurées
- la demande de vitesse de réponse

OBSERVATION DE LA TERRE

Evolution du paysage - Agriculture - Océans
Evènements climatiques



FLUX DE CIRCULATION

Route - Rail - Maritime - Aéronautique



SUIVI D'OBJETS CONNECTÉS

Fabrication - Logistique – Santé – Smart Cités



SYSTEMES GÉODESIQUES

L'information géographique désigne toutes informations sur les objets localisés sur la terre. On peut représenter; un **objet** (lampe, route, cours d'eau), ou un **phénomène** (inondation, avalanche...)

3 niveaux de descriptions

- Niveau géométrique (forme et localisation)
- Niveau sémantique (données attributaires)
- Niveau topologique (relation avec les voisins)

Système géodésique

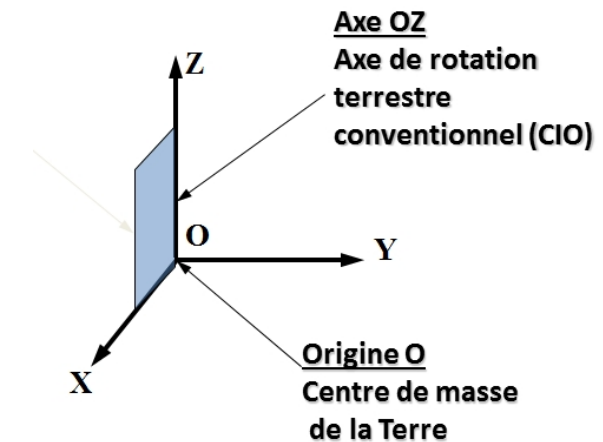
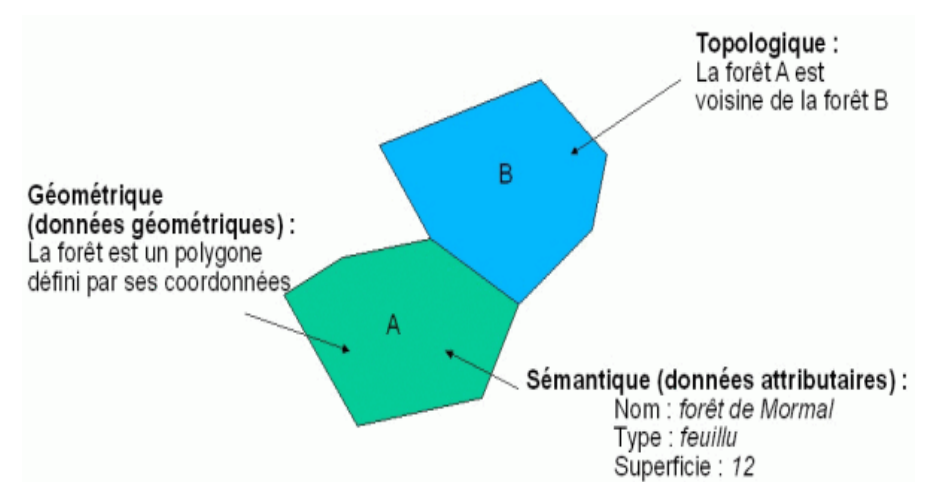
Le positionnement des points dans l'espace et dans le temps au voisinage de la terre nécessite la définition d'un référentiel

C'est un repère affine (O,i,j,k) tel que :

- O proche du centre de la terre
 - 500 mètres pour les systèmes locaux
 - <10 mètres pour les systèmes spatiaux
- OZ est proche de l'axe de rotation de la terre
- OXY est proche du plan de l'équateur
- OXZ est proche du plan méridien origine

Deux types de système

- systèmes locaux –issus de réalisation terrestre
- systèmes spatiaux qui sont mondiaux



SYSTEMES GÉODESIQUES

Différents Systèmes géodésiques

fournit des points stables et matérialisés par des bornes de coordonnées connues

ellipsoïde

déformation
mathématique



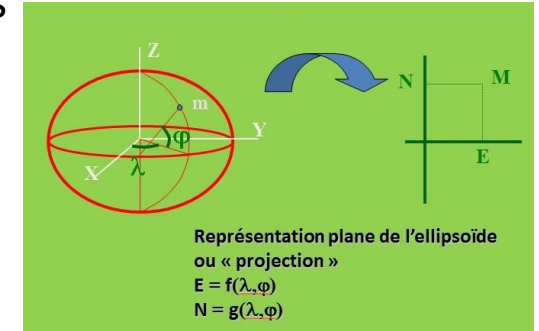
Sphère



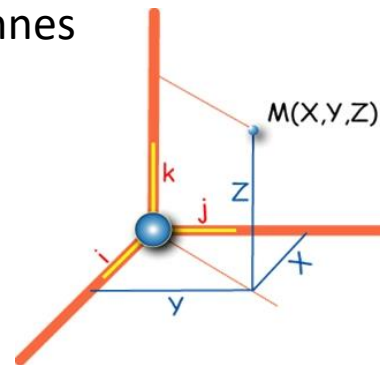
ellipsoïde



la terre réelle

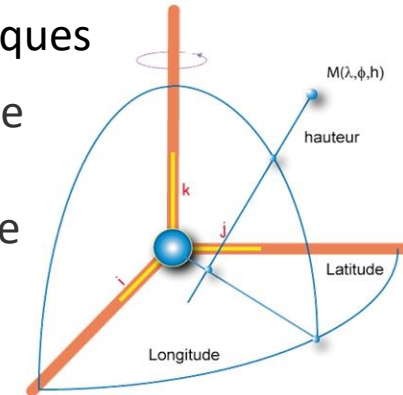


Coordonnées cartésiennes



Coordonnées géographiques

- λ : longitude géodésique
- ϕ : latitude géodésique
- h : hauteur au dessus de l'ellipsoïde

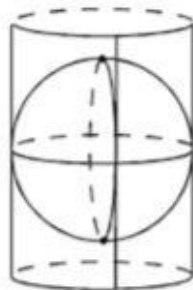


projections

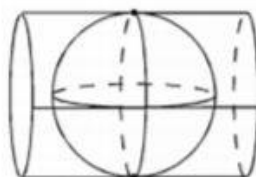
Il faut projeter l'image 3D pour avoir des coordonnées planes 2D avec perte d'information

- Conservation des directions : projection conforme
- Conservation des surfaces : projection équivalente
- Conservation des distances : projection équidistante (aphylactique)

Projection cylindrique directe
ex : Mercator

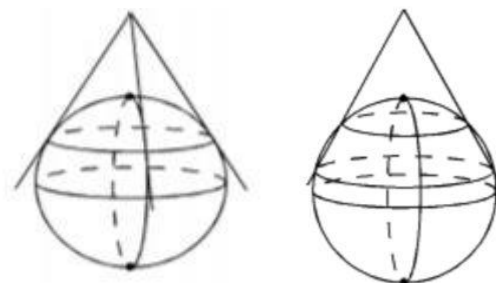


Projection cylindrique transverse
UTM = Universal Transverse Mercator



Projection conique
ex : Lambert

A surface de projection est
Un cone tangent ou sécant

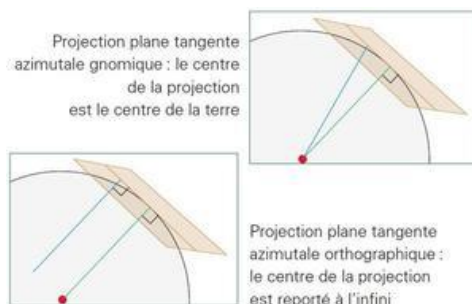


tangent

sécant

Projection azimutale

Le plan est lui-même tangent
au modèle de la terre



Projection plane tangente
azimutale gnomonique : le centre
de la projection
est le centre de la terre

Projection plane tangente
azimutale orthographique :
le centre de la projection
est reporté à l'infini

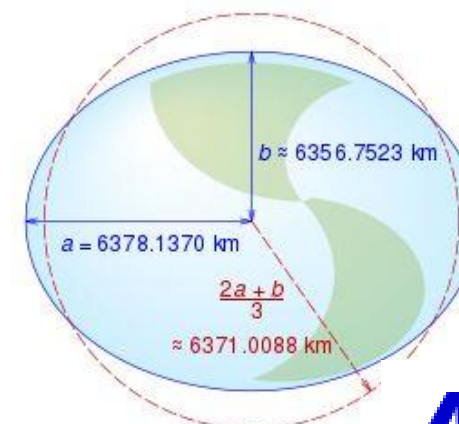
Quelques systèmes de cartographie

Lambert : ancien système de triangulation
français (NTF) - conique secant en 4 zones

Lambert93 : système légal en France
système géophysique RGF93
conique avec 23 centres de référence
ellipsoïde IAG GRS80 en 9 zones

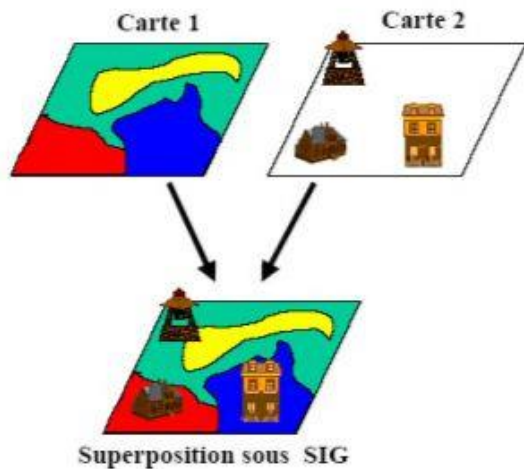
ETRS89 ; système européen – GRS80
remplace ED50

WGS84 : World Geodetic Systems 1984
Ellipsoïde IAG GRS80
centrage sur le centre de la terre
positionnement par GPS



Modélisation de l'information géographique

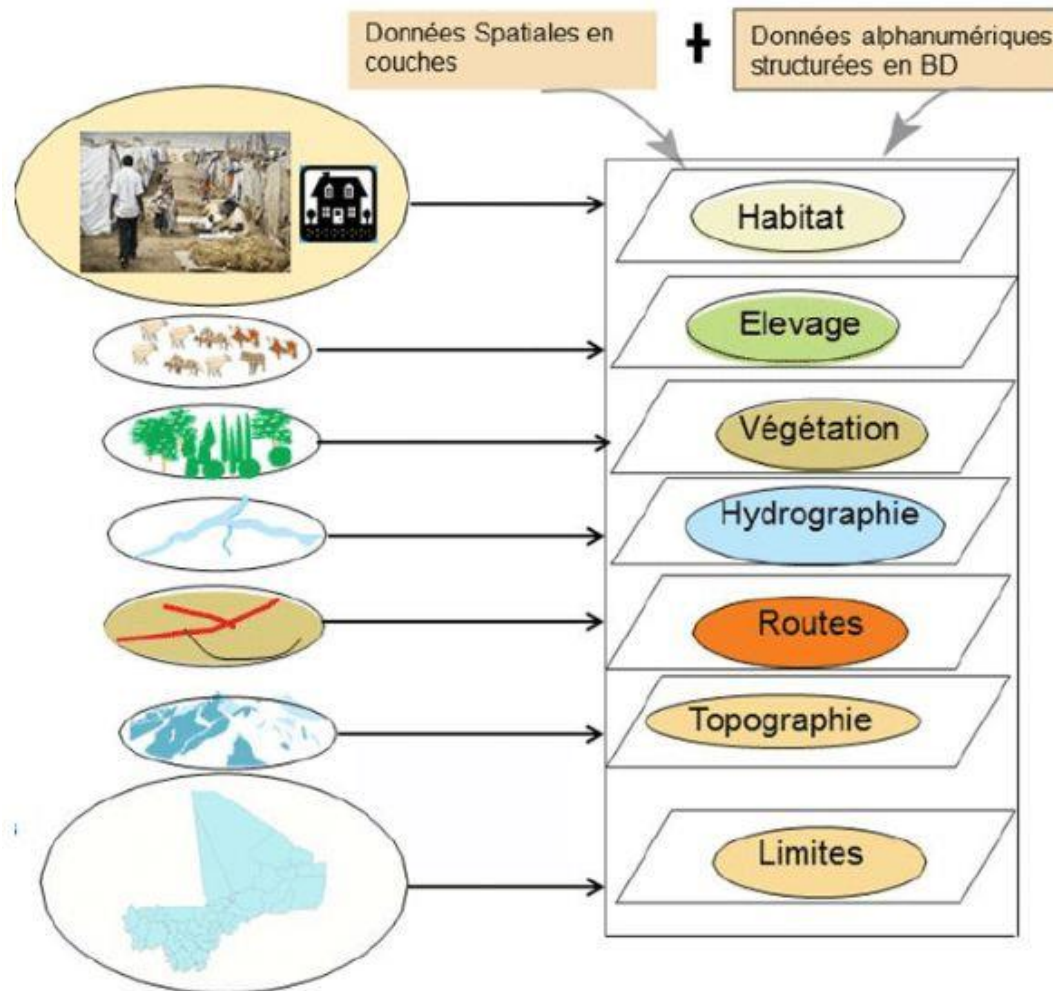
Pour modéliser la réalité sous Système d'Informations Géographiques (SIG), on répartie l'ensemble des informations géographiques de même type (l'ensemble des tronçons de cours d'eau) sous forme de plan 2D, dit « couche ».



Base de données spatialisées



- Bases thématiques
- Bases métiers



Modes de représentation de l'information géomatiques (2 modes) /1

MODE 1 : Mode VECTEUR

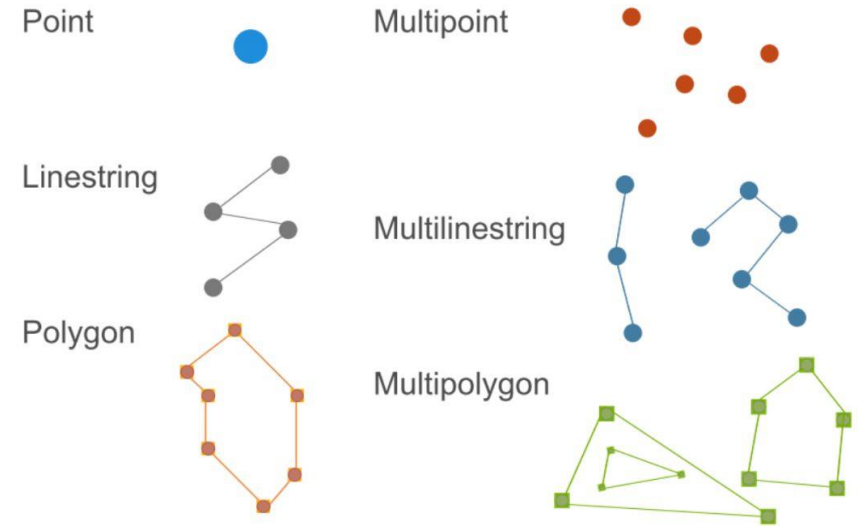
Un vecteur a **deux composantes** : une composante **attributaire** (les informations qui lui sont associées) et une composante **graphique** (sa forme, sa géométrie).

Dans sa composante graphique, un vecteur est composé de nœuds (des points dans l'espace) et des formules mathématiques pour calculer les arcs (des lignes) qui relient ces nœuds entre eux et qui forment ainsi une géométrie. Ces objets géométriques sont définis par leurs coordonnées (X,Y) **MODE 1 : Mode VECTEUR** dans un système de projection

Format. Un format de fichier bien connu pour l'information géographique est le « fichier de formes », mieux connu sous le nom de **shapefile**. Son extension est **.shp**, il est accompagné de deux autres fichiers de meme nom et d'extension :

- **DBF**, qui contient les données attributaires relatives aux objets du shapefile
- **SHX**, qui stocke l'index de la géométrie

Type des attributs de l'objet

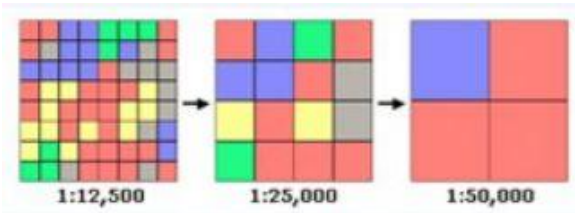


Modes de représentation de l'information géomatiques (2 modes) /2

MODE 2 : Mode RASTER - IMAGE (ou MAILLÉ)

L'information géographique représentée sous forme raster (image) est constituée d'un ensemble de xellules appliquées Pixels (picture élément) organisés sous forme d'une grille en ligne /colonne.

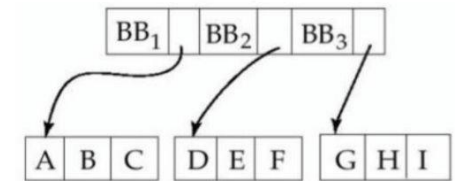
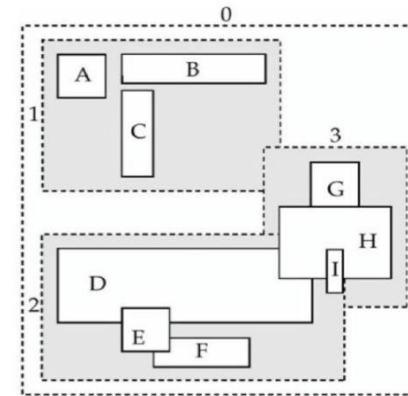
Résolution spatiale : Le pixel est l'unité élémentaire de base de l'image dont la surface correspond à sa résolution spatiale :



Profondeur de l'image : A chaque pixel est associée une ou plusieurs valeurs numériques décrivant les caractéristiques de l'espace. La profondeur définit la quantité de comptes numériques disponibles pour chaque pixel dans une image.

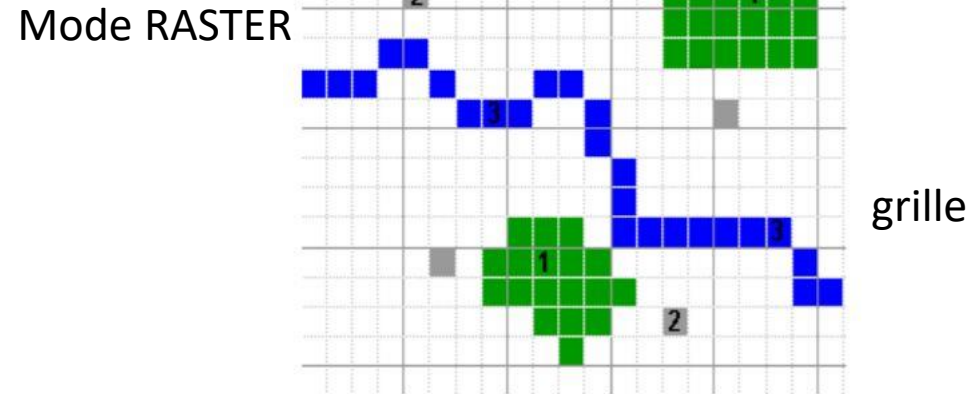
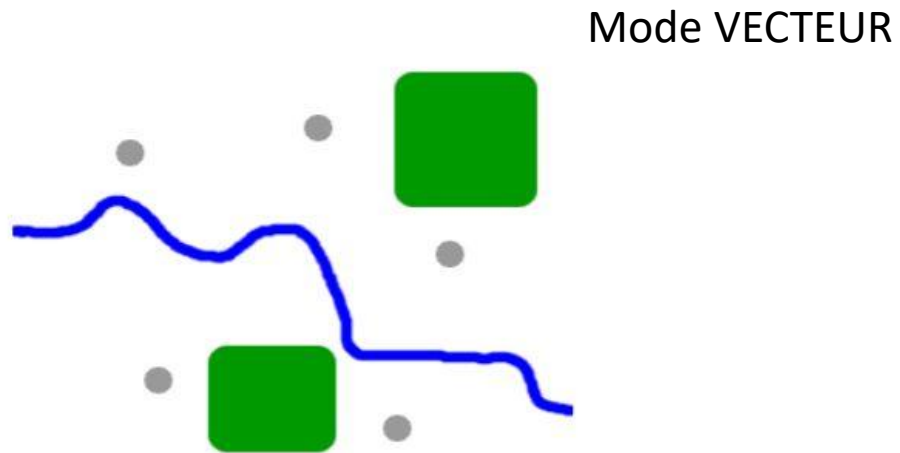
Les fichiers raster sont souvent lourds et sont enregistrés dans des formats compressés connus dans secteur de la géomatique : **basé sur les ondelette** : **.ecw, .sid**.

L'affichage des fichiers raster est souvent accéléré en procédant à une indexation de type **quadtree** nommé dans certains logiciels "pyramide"



Modes de représentation de l'information géomatiques (2 modes) /1

Comparaison des deux modes



- **donne** une représentation très conforme à la réalité
- **la localisation et les dimensions** des objets sont calculés avec précision
- **on peut individualiser les objets**, donc leur attacher des attributs
- **le poids du fichier est réduit.**

- facilité d'utilisation : données sont sous forme de tableau . Par rapport au mode vecteur, la dimension thématique est donnée par des valeurs numériques de la grille et la dimension spatiale est déduite par la position relative du pixel dans la grille.
- **le croisement des données est facile à réaliser** : toutes les grandeurs sont ramenées à la même unité de base (le pixel).
- **il se prête bien à certains types de traitements numériques** car chaque pixel contient une valeur numérique (ex : classification supervisée). Cette valeur est stockée dans un canal (raster monocanal) ou plusieurs canaux (raster multi-canal), chacun d'eux représentant une info distincte.
Le canal est au raster ce que l'attribut est au vecteur.

Les METADONNÉES

Les **métadonnées** sont définies comme "**données sur les données**" ou "**information sur les données**". Elles sont indispensables pour garder la trace de l'acquisition des données et de permettre le transcodage des données.

Les métadonnées sont une liste structurée d'information qui décrivent les données ou les services (incluant les données numériques ou non) stockés dans les systèmes d'information.

- Des informations d'identification
 - Intitulé, description, dates, version, résumé, intervenants, éditeur, contact
 - Etendue des données, aperçu des sources, contraintes légales de la sécurité
- La description du contenu
- Le système de coordonnées
- Les informations de géolocalisations et organisation des données
- des informations de qualité
 - Précision géométrique temporelle, sémantique, échelle de référence
 - Des informations de généalogie (description des sources et des processus liés aux sources)
- Modalités de diffusion : droit, tarifs, restrictions
- Aide pour un catalogue et l'interopérabilité

Normalisation

- **ISO/TC211** – ISO 19100
- **UE** - standard INSPIRE – inspire.ec.europa.eu
- **US** – **FGDC** - www.fgdc.gov (fédéral Geographic Data Committee)
- **IGN**- portail: www.geoportail.gouv.fr

SOURCES des DONNÉES

1 – Données primaires : données brutes non traitées

- **Positionnement** : par mesures, calculs ou interprétation par expert
- **imagerie aérienne** satellite , ballon , drone

2 – sources de géodonnées

- **Données disponibles globales sur géo serveurs (serveurs WEB)**

Raster : géo-image en tuile (zoom enchainé des pixels)

Requêtes http : **WMS** = Web Map Service (image) et **WFS** = Web Feature Service (vecteur)

Principales sources

- OpenStreetMap
- GoogleMap
- BingMap
- MapQuest
- YandexMap (russe)
- ArcGis (open source)
- Global Mapper

- **Imagerie satellitaire**

Open source (10/30 mètres)

- ESA – Sentinel
- USGS – Earth explorer
- USGS – Global Data explorer
- NASA - Reverb

Sources commerciales (haute résolution)

- Satellite Imaging corporation

- **Services spécialisées**

- Natural earth (générales)
- WorlClim (climatiques)
- Soilsgrids (sols)
- Unep-pnue (environnements)
- Wordpop (population)

STOCKAGE de l'information géographique mode VECTEUR /1

Deux types de données ;

- données statiques(massif,routes,batiments)
- données dynamiques (flux trafic routier, débit cours d'eau, inondations ..)

1 – données à stocker

- Données relatives à une couche
- Au niveau de l'objet
 - Géométrie (type, coordonnées des sommets, système de projection)
 - Les données attributaires (propriétaire, type culture ..)
- Au niveau de la collection d'objet
 - Les systèmes de projection
 - La symbiotique (couleur, transparenc, epaisseur, opacité)
 - La topologie (assemblage, connectivité des objets –riviere/sens)

2 – formats de stockage

- Persistance
 - Stockage en **fichiers simples**
- Structuration
 - Stockage en **fichiers semi-structure** : base de données autonomes
- Centralisation des accès
 - Fichiers ou bases en mode **client-serveur**
- Fonctionnalités de gestion
 - Système de gestion cde bases de données : **SGBD – NoSQL ,autres**

STOCKAGE de l'information géographique mode VECTEUR /2

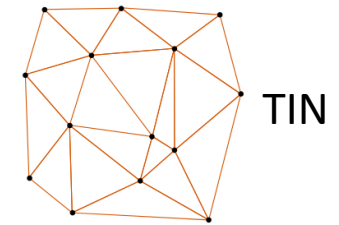
3 – Stockage en FICHIERS SIMPLES

A) Well Known text - WKT

Le format Well-known text, abrégé en **WKT**, peut se traduire par « texte bien lisible ». C'est un format Standard en mode texte ASCII utilisé pour représenter des objets **WKT** en binaire .

Les objets géométriques pouvant être représentés par ce langage sont:

- le point,
- la ligne,
- le polygone,
- le TIN ([Triangulated irregular network](#)), une surface au relief irrégulier traité par une surface couplée à des points et des lignes,
- le [polyèdre](#).



Ces objets peuvent être des géométries multiples, autrement dit une collection de géométries de même type. Les coordonnées peuvent être en deux dimensions (x, y) ou en trois dimensions (x, y, z) auquel cas la désignation de la géométrie se termine par la lettre Z.

Voici quelques exemples:

```
POINT(6 10)
LINESTRING(3 4,10 50,20 25)
POLYGON((1,5 1,5 5,1 5,1 1))
MULTIPOINT((3.5 5.6), (4.8 10.5))
MULTILINESTRING((3 4,10 50,20 25), (-5 -8,-10 -8,-15 -4))
MULTIPOLYGON(((1 1,5 1,5 5,1 5,1 1)), (2 2,2 3,3 3,3 2,2 2)), ((6 3,9 2,9 4,6 3)))
```

B) SHAPEFILE

Le **shapefile**, ou « fichier de formes » est un format de fichier pour les données géospatiales(SIG). Initialement développé par ESRI pour ses logiciels commerciaux, ce format est désormais devenu un standard de facto, et est utilisé par un grand nombre de logiciels libres ([MapServer](#), [Grass](#), [UDig](#), [MapGuide Open Source \(en\)](#),

[QGIS](#), [GvSIG](#), etc.) comme propriétaires ([VectorWorks](#), AutoCAD Map 3D, etc.).

Il contient toute l'information liée à la géométrie des objets décrits, **shapefile** est en fait un regroupement de plusieurs fichiers formatés pour représenter différents aspects des géodonnées est constitué au minimum de 3 fichiers

- **.shp** : entité qui contient la géométrie des entités elle meme
- **.dbf** :données attributaire
- **.shx** : index des entité géographiques
- **.prj** : parametres de projection (WKT)
- **.sbn** : index relatifs aux requetés

Exemple : cartes des départements

Les données proviennent de la base de données cartographique communautaire et libre [OpenStreetMap](#).

places.shp : noms des villes ou des quartiers

roads.shp : toutes les voies de passage de l'autoroute au chemin piéton

buildings.shp : l'espace bâti

railways.shp : les voies ferrées

waterways.shp :le réseau hydraulique

points.shp : une liste de point d'intérêt

natural.shp : zones vertes

landuse.shp : occupation des sol

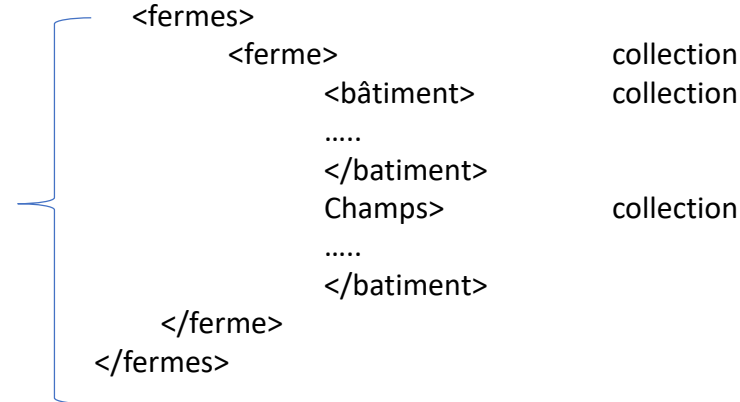
admin-departement.shp : le département

STOCKAGE de l'information géographique mode VECTEUR /4

4 – Stockage en FICHIERS STRUCTURÉS

A) Modèle XML ; balise hiérarchisée

Contenu de la couche
Pas de hiérarchie



Le **Format GPX**, dérivé du XML est un format d'interopérabilité du **GPS** [liste de **waypoint-wpt** (cheminement)]

B) Modèle GML : Géography Markup Language : dérivé du XML

Le **Geography Markup Language (GML)** est un langage dérivé du **XML** pour encoder, manipuler et échanger des données géographiques. C'est un standard développé par l'Open Geography Consortium pour garantir l'interopérabilité des données dans le domaine de la géomatique.

Le GML consiste en un ensemble de schémas XML qui définissent un format ouvert pour l'échange de données géographiques et permettent de construire des modèles de données spécifiques pour des domaines spécialisés, comme l'urbanisme, l'hydrologie ou la géologie.

Le langage GML permet de décrire :

- les objets géographiques
 - Les systèmes de projection
 - La géométrie
 - La topologie
 - le temps,
 - les unités de mesures,
- et les attributs des objets géographiques.

STOCKAGE de l'information géographique mode VECTEUR /5

4 – Stockage en FICHIERS STRUCTURÉS

C) Format GeoJSON

- **GeoJSON** (de l'anglais Geographic JSON, signifiant littéralement JSON géographique) est un format ouvert d'encodage d'ensemble de données géospatiales simples utilisant la norme JSON (JavaScript Object Notation).
- Il permet de décrire des données de type point , ligne, chaîne de caractère, polygone, ainsi que des ensembles et sous-ensembles de ces types de données et d'y ajouter des attributs d'information qui ne sont pas spatiales.
- Développé par un groupe de travail internet (IETF – RFC7945)en langage python
- Exemple

TopoJSON

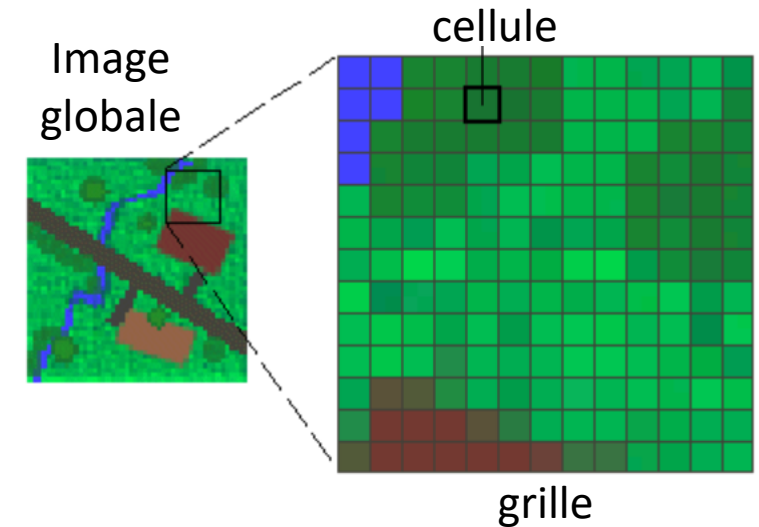
C'est une extension du format GeoJSON. Le format TopoJSON est plus léger que le format GeoJSON et évite les redondances. Ce format est très utile pour faire des représentations cartographiques de données sur le web

```
{  
  "type": "Feature",  
  "geometry": {  
    "type": "Point",  
    "coordinates": [125.6, 10.1]  
  },  
  "properties": {  
    "name": "Dinagat Islands"  
  }  
}
```

STOCKAGE de l'information géographique mode RASTER /1

Dans sa forme la plus simple, un **raster** se compose d'une matrice de cellules (ou pixels) organisées en lignes et en colonnes (grille) dans laquelle chaque cellule contient une valeur représentant des informations ; la température, par exemple.

- **Au niveau de la cellule** (pixel)
 - Coordonnées de la cellule (X,Y)
 - La valeur de la variable
- **Au niveau de la grille**
 - Système de projection
- **Formats grilles**
 - **Formats matriciels**
 - Grille ESRI(.asc)
 - **Formats listes –XYZ (.txt; .cvs; .dat)**
 - **Formats images**
 - **Formats matriciels (images en tif, jpg, png, gif)**
 - **Palette calorimétrique**
- **Information d'accompagnement (accompagnement)**
 - **géotiff**



L'objectif des spécifications du GeoTIFF consiste à permettre de décrire toute information « cartographique » associée à une image TIFF provenant d'un système d'imagerie satellite, de photographie aérienne scannée, de cartes scannées, de modèle d'élévation digital, ou du résultat d'analyse géographique

INFORMATIONS SPATIALES ET CARTOGRAPHIQUES

BASES de DONNÉES

- ✓ SPATIALES (SQL)
- ✓ NoSQL
- ✓ Autres

BASES de DONNÉES (SIG)

A – BASES RELATIONNELLES SPATIALES

Une base de données est une collection de données persistantes les transactions doivent respecter les critères, **ACID (atomicité, coherence, isolation, durabilité)**
Les bases spatiales sont basées sur les bases relationnelles du type **SGBD** , avec une évolution du standard SQL compatible **SQL**
Le Spatial SQL qui est élaboré par les opérateurs SQL du standard MapInfo Professionnel et des opérateurs du **OGC** (Open Geospatial Consortium) Spécification des fonctionnalités simple d'**OpenGIS** pour la révision de SQL- document de projet OpenGIS 99-049 du 5 mai 1999

1 – différences

➤ **Types des données** – deux classes : **géométriques et géographiques**

Types de données **géométriques** sont planes , comme les lignes et les surfaces bidimensionnelles . Les types de données **géographiques** sont ellipsoïdes et tiennent compte de la courbure de la Terre , telles que les coordonnées GPS et les chemins .

Implémentations spécifiques de types de données spatiales dans les systèmes de gestion de bases de données diffèrent entre les vendeurs , mais les concepts de base sont les mêmes

Exemples de types simples :

[Point](#)

[LineString](#)

[CircularString](#)

[CompoundCurve](#)

[Polygone](#)

[CurvePolygon](#)

Exemples de types de collection :

[MultiPoint](#)

[MultiLineString](#)

[MultiPolygon](#)

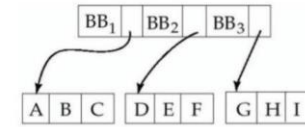
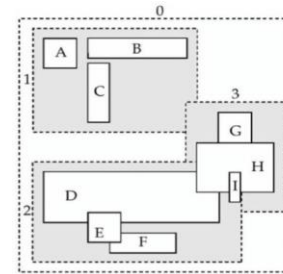
[GeometryCollection](#)

A – BASES RELATIONNELLES SPATIALES /2

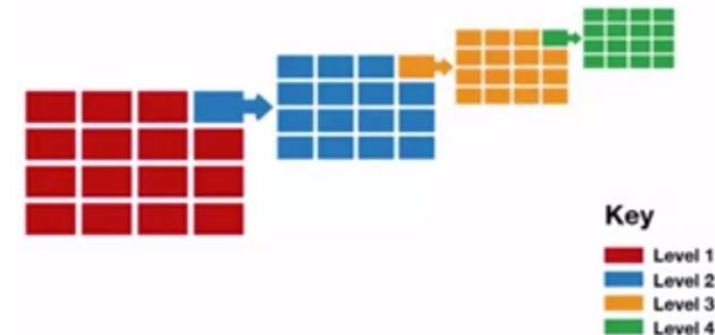
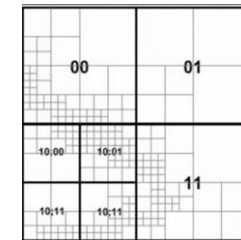
2 – indexation

- BD non spatiale
 - B-Rree- arborescence hiérarchique
 - Hachage
- BD Spatiale
 - -Tree :classent les objets géographiques sous forme de rectangles, de sous-rectangles etc. Cette structure d'index gère automatiquement la densité et la taille des objets. Voisinage de même nature
 - QuadTree : chaque nœud à quatre fils- l'espace d'un nœud est divisé en quatre espaces, puis ceux-ci en quatre – "quadtree -regions" (tuilage de google)- Map service Google
Un **nœud** représente une image complète
 - Grid : tessaliation régulière divisant une surface 2D en une série de cellules contiguës, auxquelles peuvent être associés des identifiants uniques, pour les besoins de l'indexation spatiale

R-Tree



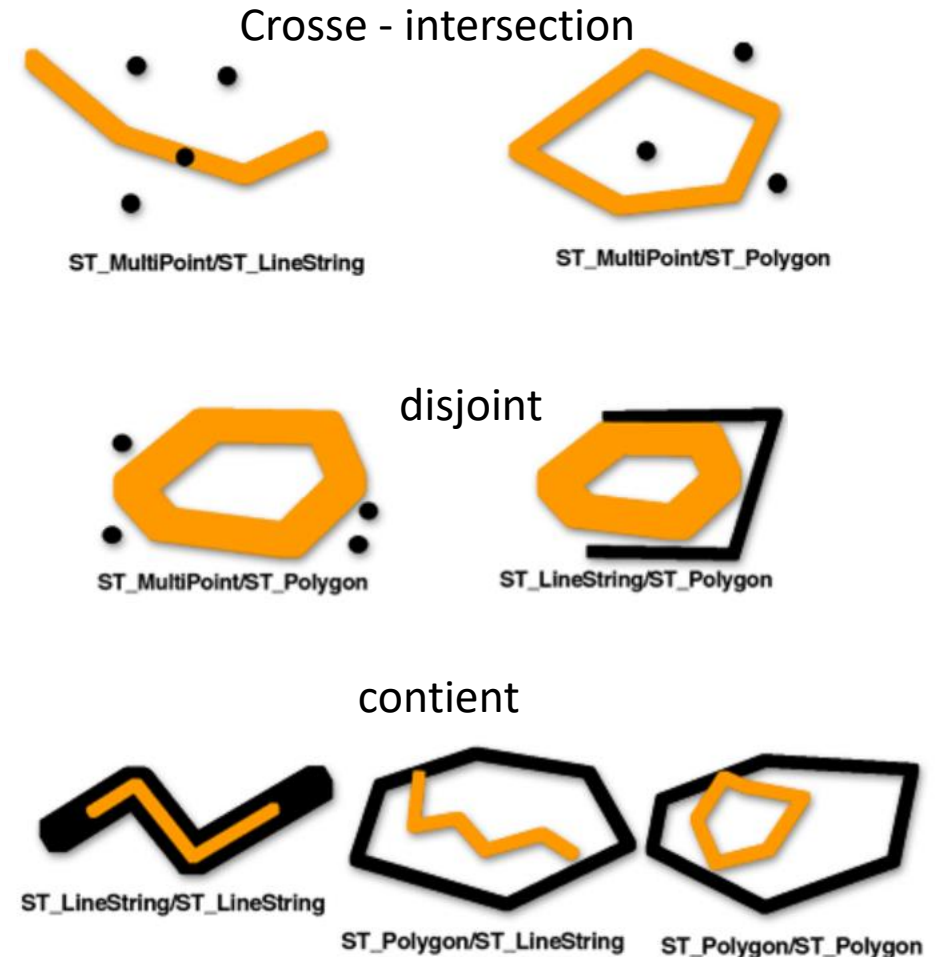
QuadTree



A – BASES RELATIONNELLES SPATIALES /3

3 – fonctions

- **Mesures spatiales** : longueur, surface
 - Distance entre deux objets
- **Opérateurs topologiques** : (opérateur et test)
 - Equals ,
 - Disjoint
 - Touches,
 - Contient; within (Est Contenu)
 - crosses (intersection)
- **Fonctions spéciales** créant de nouvelles entités
 - Union
 - inclusion
 - Intersection
 - Difference : obtenue par soustraction
 - Simplify : géométrie simplifiée
 - Buffer : surface élargie



A – BASES RELATIONNELLES SPATIALES /4

3 – OGC – Open Geospatial consortium

L'**Open Geospatial Consortium**, ou **OGC**, est un [consortium](#) international pour développer et promouvoir des standards ouverts, les *spécifications OpenGIS*, afin de garantir l'intéropérabilité des contenus, des services et des échanges dans les domaines de la géomatique et de l'information géographique. En [2009](#), l'OGC compte 371 partenaires du monde entier : structures commerciales et industrielles, agences gouvernementales et internationales, universités...

4 – fournisseurs de BASES RELATIONNELLES

De nombreux fournisseurs de SGBD ont adoptés l'extension spatiale et mis des logiciels de gestion cartographique à leurs catalogues.

- IBM Db2 avec ST_Geometry
- SQL Server ,Microsoft Azure
- Oracle/PLSQL:SIG
- PostgréSQL avec PostGIS (open)
- AmazonRDS avec Oracle Spatial
- SAP HANA associé a ESRI (ArcGIS)
- Sqlite avec Spatiallite (open)

Ces bases sont accessibles directement ou par l'intermediaire de logiciels clients

PS – il existe de nombreux logiciels poi u les systemes d'information géometrique , consultez

Wikipédia : liste des logiciels SIG

A – BASES NoSQL

Une **base de données SIG** est une collection de données, conceptuellement structurée, décrivant les caractéristiques des données et les relations existant entre leurs entités correspondantes, destinée à aider dans divers champs d'application

- Fichiers simple ; shapefile , WKT
- Fichiers structurés : XML, GML, JSON
- Fichiers spatiaux : GeoJSON

	CouchDB	MongoDB	Neo4j	BigTable
Basic Concepts	Document-oriented	Document-oriented	Network-oriented	Column-oriented
Indexing	R-Tree Only2D	2D 2Dsphere	R-Tree 2D and partially 3D	B-Tree Only2D
Vector Data Types	Fully	Fully	Fully Basic Types and Limited MultiGeometry Types	Fully Basic Types
Topological functions	Only Within() Contains()	Only Within(Point) Contains(Point)	Almost fully	Not supported
Analysis and metric functions	Only Distance()	Only Distance (Point)	Fully	Only Distance()
Set functions	Not supported	Only Intersection (Point)	Fully	Not supported
Input/output Format	Input: .SHP Output: .KML .CSV .GeoJSON	Input: GeoJSON Output: GeoJSON	Input: .SHP .OSM Output: SLD styled PNG	Input .JSON .KML Output .JSON .KML

Trois types de base NoSQL

- **Type colonne :**
 - Orienté : **bigTable** (data base); **HBASE (Hadoop)**
- **type document**
 - orienté CP (coherence/Partition) : **MongoDB**
 - Orienté AP (Available/Partition) : **CouchDB**
- **Type Graphe : Neo4**

B – BASES NoSQL

Théorème de CAP

Dans toute base de données, vous ne pouvez respecter au plus que 2 propriétés parmi la *cohérence*, la *disponibilité* et la *distribution*.

Consistency (Cohérence) : Une donnée n'a qu'un seul état visible quel que soit le nombre de réplicas

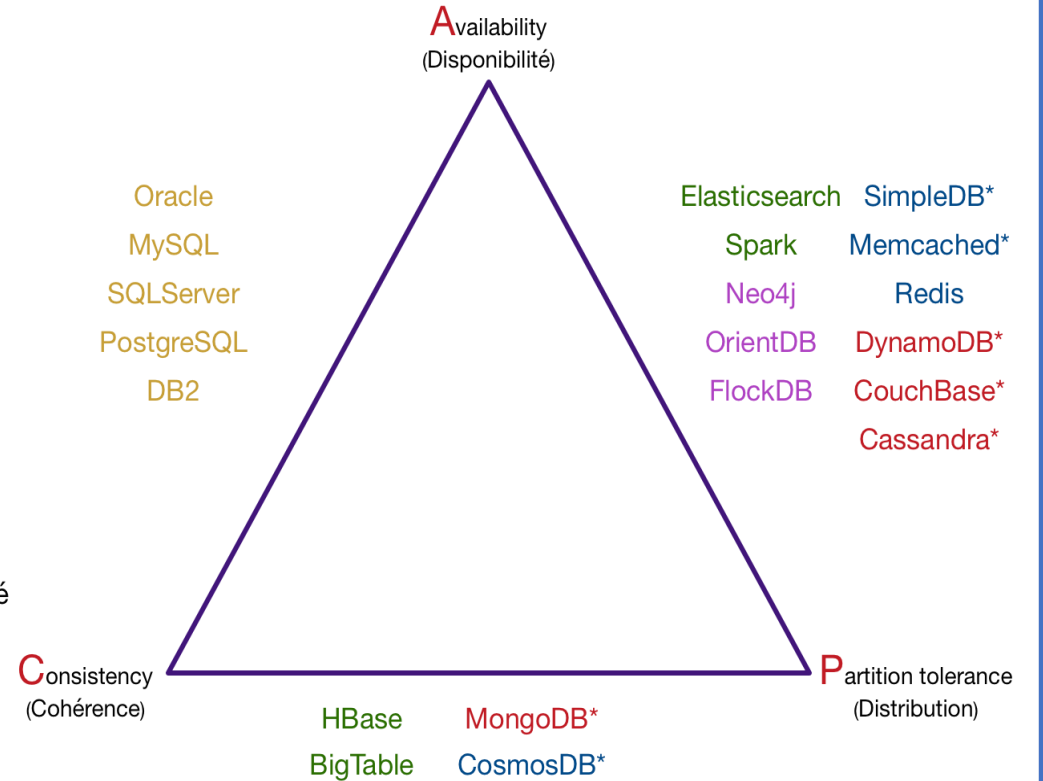
Availability (Disponibilité) : Tant que le système tourne (distribué ou non), la donnée doit être disponible

Partition Tolerance (Distribution) : Quel que soit le nombre de serveurs, toute requête doit fournir un résultat correct

Modèle

- Relationnel
- Orienté Clé-Valeur
- Orienté Colonne
- Orienté Document
- Orienté Graphe

* Possibilité de changer le mode de cohérence
Cohérence <-> Disponibilité



B – 1 – base clé-valeur (kvstore)



- Amazon dynamo DB
- Oracle NoSQL database
- Infinity
- Redis
- MemcacheDB
- Aerospike
- Oracle Berkeley
- Riak Kv
- Voldemort
- Apache Ignite

A – BASES NoSQL

Base type colonne

1 – BigTable (google)

- Application : google Earth et google sky
- Développement en **KML**(Keynote Markup Language)
- Saisie : Landsat8, navette)
- Outils de developpement de couches

2 - Hadoop

- SpatialHadoop : un cadre map-reduce pour données spatiales
- RESQUE : moteur de recherche spatiales
- Fichiers HFDS : système distribué

Base type document

3 – MongoDB

- Intègre nativement les fonctionnalités spatiales (objets et requêtes)
- Données en mode GeoJSON
- Indices géospatiaux
 - **2dsphere** : Les index [2dsphere](#) prennent en charge les requêtes qui calculent les géométries sur une sphère terrestre (WGS 84).
 - **2d** : Les index [2d](#) prennent en charge les requêtes qui calculent les géométries sur un plan directionnel
- Création de couches avec shapefile (PyMongo) et MongoLayers avec QGIS
- Créateur du sharding (éclatement des données)

A – BASES NoSQL

Base type document

3 – CouchDB

- Projet officiel de la fondation Apache
- Extension spatiale : GeoCouch conforme OGC
- Utilise mapReduce
- Permet la réplication
- Importe des shapefile contrairement à MongoDB
- CouchDB constitue un monde en soi. Tout est codé à l'intérieur des bases, leur permettant de fonctionner comme des applications autonomes et portables (réplication). Elle est très différente de MongoDB ou SimpleGeo (Cassandra) où seules les données sont stockées et où les traitements sont faits de l'extérieur.

4 – GeoMesa

- une suite d'outils **open source** qui permet des requêtes géospatiales et des analyses à grande échelle sur les systèmes informatiques distribués.
- GeoMesa fournit une indexation spatio-temporelle au-dessus des bases de données **Accumulo, HBase, Google Bigtable et Cassandra** pour le stockage massif des données ponctuelles, ligne et polygones

A – BASES NoSQL

Base type Graphe

4 – Neo4J

- est une base de données orientée graphe développée en JAVA. Elle est développée en Java et est bien documentée. Elle offre d'excellentes performances et permet de traiter rapidement de grandes quantités de relations .
- Elle stocke des données structurées sous forme de graphe.
- dispose d'un module spatial appelé *Neo4J Spatial* , intégrant les principales fonctions de recherche topologique, de routage et de création d'index spatiaux.
- **Neo4J Spatial** fournit une couche de traitement géographique eux SIG si ces derniers ne disposent pas de ces fonctions .Supporte les objets conforme à **OpenGIS** (du point au multi-polygone)
- Plusieurs langages tels que Gremlin, SPARQL et Crypter permettent de parcourir des graphes dans Neo4J.

GREMLIN est le **langage graphique** traversant [d'Apache TinkerPop](#). Gremlin est un langage [fonctionnel](#) de flux de [données](#) qui permet aux utilisateurs d'exprimer succinctement des traversées complexes sur (ou des requêtes) du graphique de propriété de leur application. Chaque traversée Gremlin est composée d'une séquence d'étapes (potentiellement nichées)

C2 – Data Lake

Les 3V du big data et de l'IOT font référence à :

- **Volume** car les masses de données à traiter sont sans cesse croissantes.
- **Vitesse** car le traitement des données doit se faire souvent en temps réel ou le plus vite possible.
- **Variété** car les données sont de formats très variés et pas toujours structurées

le **lac de données** (ou **data lake** en anglais) désigne un espace de stockage global des informations présentes au sein d'une organisation.

On trouve donc dans un lac de données des données de natures différentes : des données structurées issues notamment de base de données relationnelles (lignes et colonnes), des données semi-structurées (CVS, journaux, XML, JSON...), et des données non structurées (emails, tweet, documents, [PDF](#)) et fichiers de type [blob](#) (images, audio, vidéo notamment).

Le positionnement est de plus en plus une donnée fondamentale de ces flux d'information

Plate forme d'accueil

Cloud → **Amazon AWS-Redshift**, **Microsoft Azure Cosmos**, **oracle** big data Cloud, **IBM Db2 Big SQL**

Google Cloud

Outils → Snowflake sur AWS et Azure

Indexa sur AWS

Talend

Cloudera Altus sur IBM

Ataccama ONE

Outils APACHE -GÉOSPATIAL

De nombreux projets Open-Source sont développés et accessibles pour enrichir le mode MapReduce

1 – librairie SIS

Apache Spatial Information System (SIS) est une bibliothèque de langage Java pour développer des applications géospatiales. SIS fournit des structures de données pour les entités géographiques et les métadonnées associées ainsi que des méthodes pour manipuler ces structures de données. Implémentation de [GeoAPI 3.0.](#) (OGC)

2 – SpatialHadoop (université Minnesota)

SpatialHadoop est une extension MapReduce d'Apache Hadoop conçue spécialement pour fonctionner avec des données spatiales.

3 – Hadoop-GIS

Hadoop-GIS est un système de stockage de données spatiales évolutif et hautes performances fonctionnant sur Hadoop. Il prend en charge le partitionnement spatial, l'interrogation spatiale personnalisée à l'aide de RESQUE qui s'exécute sur le framework Map-Reduce.

4 – Outils ESRI –SIG pour Hadoop

Ce sont des outils open source qui s'exécuteraient sur la plateforme ArcGIS. Ceux-ci permettent l'intégration de Hadoop avec le logiciel d'analyse de données spatiales, c'est-à-dire ArcGIS Desktop

5 – GeoSpark

GeoSpark est un système informatique en grappe pour le traitement de données spatiales à grande échelle. GeoSpark étend Apache Spark / SparkSQL avec un ensemble de jeux de données distribués résilients spatiaux prêts à l'emploi (SRDD) /

Un prototype SIG pour analyser la résilience urbaine : application à la ville de Dublin

Vertigo – la revue électronique en sciences de l'environnement – volume 13 , numéro3
Serge Lhomme, Richard Laganier, Youssef Diab et Damien Serre

