

الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التعليم العالي والبحث العلمي
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE
جامعة عمّار تليجي بالأغواط
UNIVERSITE AMAR TELIDJI LAGHOUAT



كلية العلوم
FACULTE DES SCIENCES
قسم الرياضيات والإعلام الآلي

DEPARTEMENT DE MATHEMATIQUES ET INFORMATIQUE

Mémoire de MASTER

Domaine : Mathématiques et Informatique

Filière : Informatiques

Option : Système d'Information et de Décision

PAR:

GAZZAM kaltoum

Theme

L'intégration d'une base de données NoSql dans le processus de développement de système d'information géographique.

Soutenue publiquement devant le jury composé de :

Mr. MAICHA Mohammed El Habib	Encadreur	Univ.Laghouat
Mr. GUELLOUMA Younes	President	Univ.Laghouat
Mr. AMEUR Mohammed El Amine	Examineur	Univ.Laghouat

Année universitaire :2016/2017

Remerciements

Toute notre gratitude et remerciements à ALLAH qui nous a donné la force, le courage et la volonté d'élaborer ce travail.

Nous adressons notre profond remerciement à Monsieur l'encadreur MAICHA Mohammed El Habib pour avoir dirigé ce travail. Nous le remercions pour ses précieux conseils et sa disponibilité.

Nous remercions les membres du jury de nous faire l'honneur de juger notre modeste travail.

Nous remercions tous nos professeurs du département d'Informatique.

Remerciement les plus sincères à toutes les personnes qui nous ont aidé de près ou de loin à accomplir notre travail.

Dédicace

Je dédie ce mémoire

À mes chers parents ma mère et mon père

Pour leur patience, leur amour, leur soutien et leurs

Encouragements.

À mon MARI Mohamed El Amine

À mes frères Mustapha, Mohamed, Omar et Nour Eddine

À ma sœur Rouguia,

vous m'avez toujours soutenu durant toutes mes études,

je vous souhaite une vie pleine de joie de bonheur et de réussite.

À toute ma chère famille,

À mes professeurs,

À mes chers amis,

À tous ceux qui m'aiment,

À tous ceux que j'aime,

À tous ceux qui m'ont aidé de près ou de loin,

Je dédie ce travail avec hommage.

GAZZAM Kalthoum

Résumé

Le domaine des Système d'information géographique est très intéressant car il touche pratiquement la vie quotidienne des humains, néanmoins les données produites par ces humains ne cessent de grandir et cela peut provoquer des obstacles pour ce type des systèmes. Pour cela il faut réfléchir à prévenir les problèmes liés à la gestion des grandes données au sein des SIG avant qu'ils se produisent, Car les données peuvent arriver en vitesse croissantes de plusieurs sources et en des formats différents. Dans ce sens nous avons eu l'idée d'essayer de remplacer le gestionnaire de données utilisée classiquement dans les SIG par un nouveau mécanisme qui peut tolérer ou même s'adapter avec le phénomène de données non structurées qui est le NoSQL.

Mots clés : SIG, NoSQL, MongoDB, données non structurées

Abstract

The geographical information system is considered as an important domain. because it technically touches the human's daily life. Yet the resulted Data of using these systems are increasingly grew which may lead to obstacles to this type of systems, therefore we ought to think in a way to avoid the problems that are related to the management of huge Data in its volume before it happens. In illustrated way, these Data may reach a very fast speed from different sources and maybe received in various forms. These basics lead us to have the idea of changing Data processor used typically in geographical information system by new technique where we can enhance the phenomenon of nonstructural Data, this technique is called NoSQL

Key words: GIS, NoSQL, MongoDB, Unstructured Data

ملخص

يعتبر مجال نظم المعلومات الجغرافية مجالا مهما لأنه و بكل بساطة يلمس عمليا الحياة اليومية لكل البشر ، لكن المعطيات الناتجة عن إستعمالهم لهاته النظم تنمو و بشكل متزايد و هذا قد يتسبب لنا في عقبات لمثل هذا النوع من الأنظمة المعلوماتية لذلك و يجب علينا التفكير في كيفية تجنب المشاكل المتعلقة بتسيير البيانات العملاقة من حيث الحجم قبل وقوعها من الأساس ، بشكل توضيحي هاته البيانات قد تصل بسرعة فائقة و من مصادر مختلفة و قد ترد بأشكال متنوعة ، و من هذا المنطلق كانت لدينا فكرة لمحاولة إستبدال معالج البيانات و المستخدم تقليديا في نظم المعلومات الجغرافية بألية جديدة يمكن من خلالها تحسين ظاهرة البيانات غير المهيكلة . تسمى هذه التقنية **NoSQL**

كلمات مفتاحية: نظم المعلومات الجغرافية، **NoSQL**

Table des Matières

Table des Matières	vii
Liste de figures	ix
1 Généralités sur les SIG	5
1 Introduction	5
2 Information géographique	5
2.1 Les composantes de l'information géographique	5
3 Représentation de l'information géographique	6
4 Système d'Information Géographique	9
4.1 SIG	9
4.2 Les composantes d'un SIG	10
4.3 Les domaines d'application des SIG	11
4.4 Interrogation et analyse des données de SIG	11
5 Conclusion	12
2 Du relationnel au NoSQL en géomatique	13
1 Introduction	13
2 Base de données	13
3 Systèmes de Gestion de Base de Données (SGBD)	14
3.1 SGBD relationnel	14
4 NoSQL	15

5	Mais pourquoi nous avons besoin d'une nouvelle façon de sauvegarder nos données ?	15
6	Caractéristiques principales des bases de données NoSQL	16
7	Typologies de NoSQL	18
8	Principaux avantages du NoSQL	20
9	Base de données géographiques	21
10	NoSQL en géomatique	22
10.1	Présentation de quelques solutions	22
11	Constat sur le NoSQL dans le géomatique :	24
11.1	Propriétés BASE	28
12	conclusion	28
3	Résultats et discussions	30
1	Introduction	30
2	Le système d'information géographique	30
3	La base de données géographique	32
3.1	SQLite	32
4	Du coté de NoSQL	32
4.1	La BDD MongoDB	32
4.1.1	JSON	33
4.1.2	IntelliJ IDEA	33
	Bibliographie	42

Liste de figures

1.1	Image satellite.	6
1.2	Carte.	7
1.3	le mode raster	8
1.4	le mode vecteur	8
2.1	Théorème CAP	16
2.2	Partitionnement Horizontal de données	17
2.3	représentation clef/valeur	18
2.4	schéma R vs colonne	19
2.5	Les bases orientées document	20
2.6	Les bases orientées graphe	20
3.1	La Configuration et l'installation du plugin Mongo dans l'IntelliJ IDEA	35
3.2	la carte de l'Algerie dans ArcGis	35
3.3	la carte de Parking, Restaurant, Stade en ArcGis	36
3.4	la couche Algeria et ses champs	36
3.5	Résultat de Requête 3 sur ArcGIS	38
3.6	Résultat de Requête 3 sur MongoDB	38

GLOSSAIRE

Schéma données dynamique	de	modèle de données dans lequel, on peut ajouter dynamiquement des colonnes ou attributs
Système d'Information		(abr. SI) ensemble organisé de ressources (matériels, logiciels, personnel, données et procédures) qui permet de regrouper, de classier, de traiter et de diffuser de l'information sur un environnement donné
SIG		Système d'Information géographique
Métadonnées		des données qui décrivent les données, L'objectif des métadonnées est en particulier de rendre possible et de faciliter la consultation et l'échange des données et ainsi permettre de pérenniser l'utilisation des données
Scalabilité		désigne la capacité d'un produit logiciel à s'adapter à un changement d'ordre de grandeur de la demande (montée en charge). En particulier sa capacité à maintenir ses fonctionnalités et ses performances en cas de forte demande
ACID		Atomicité, Cohérence, Isolation, Durabilité
BDD		Base de données, lot d'informations stockées de manière permanente dans un dispositif informatique. Les technologies existantes permettent d'organiser et de structurer la base de données de manière à pouvoir facilement manipuler le contenu et stocker efficacement de très grandes quantités d'informations
Cloud computing		concept qui consiste à déporter sur des serveurs distants des stockages et des traitements informatiques traditionnellement localisés sur des serveurs locaux ou sur le poste de l'utilisateur
Transaction		une suite d'opérations qui font passer la base de données d'un état A - antérieur à la transaction - à un état B postérieur
MAJ		Mise à jour
API		Application Programming Interface
REST		Representational State Transfer
SERVICE REST	API	des services qui sont propres et polyvalentes pour beaucoup de types d'échange d'information entre logiciels. REST signifie "Representational State Transfer"

Introduction générale

Contexte

L'industrie est sur le point de changement technique dans comment et où les données sont stockées et traitées. Depuis plus de 30 ans (1970), le système vénérable relationnel de gestion de base de données (RDMS), en cours de fonctionnement dans les centres de données des entreprises détient la majorité des données mondiales. Cela ne peut pas continuer. La technologie (SGBDR) ne peut plus suivre le rythme de la vitesse, du volume et de variété des données en cours de création et de consommation. Pour ce nouveau monde de Big Data, les bases de données NO SQL sont nécessaire. De ce point, et que les (BDD-R) constituent désormais le support de stockage usuel des données géographique, on constate néanmoins que très peu de travaux ont parlé des fonctionnalités NOSQL qui peuvent être utilisés dans les SIG.

Problématique

Les SGBDR n'ont pas l'habitude de travailler sur des schémas dynamiques avec une vitesse croissante et leurs sources d'alimentations peuvent être hétérogènes, tout ça amène forcément avec le temps à un volume de données très raisonnables à savoir de Téra Octets. En d'autres termes ces problématiques majeurs peuvent être résumées en 3 points:

1. L' évolutivité faible

2. Faible adaptation aux typologies de données moderne
3. Scalabilité limitée: capacité en monter en charge

Objective principale

Notre objet est de concevoir une approche solide pour l'introduction du concept NO SQL parmi les techniques de développements de SIG. Par la suite on vise à:

1. Introduire NO SQL dans le SIG pour améliorer la performance de ce dernier;
2. Introduire NO SQL dans le SIG pour améliorer la performance de ce dernier;
3. L'application des techniques de NO SQL dans un cas d'étude.

Organisation du mémoire

Après l'introduction générale, ce mémoire se décompose en 03 Chapitres :

- Tout d'abord, le premier chapitre présente des définitions et des généralités sur les SIG.
- Le deuxième chapitre illustre un nouveau type de base de données qui sera utilisé dans le mémoire, nommé NOSQL avec ces caractéristiques, ces avantages et inconvénients.
- Dans le troisième chapitre on va détailler le processus d'installation des outils avec la configuration des plateformes de test et les résultats des requêtes spatiales.

On va clôturer ce mémoire avec une conclusion générale et quelques perspectives.

Chapitre 1

Généralités sur les SIG

1 Introduction

Ce chapitre introductif a pour objectif de définir les principaux concepts afin de comprendre le thème du travail. On va commencer par décrire qu'est-ce qu'une information géographique ensuite c'est quoi le système qui utilise et manipule cette information, et on finira par comment interroger un tel système.

2 Information géographique

L'information géographique est la représentation d'un objet ou d'un phénomène réel localisé dans l'espace à un moment donné. L'information géographique est caractérisée par une composante purement spatiale et une composante sémantique et composante topologique [1].

2.1 Les composantes de l'information géographique

- **Composante spatiale:** représente la position sur la surface terrestre et la forme d'un objet du monde réel, une position est décrite dans un système de référence explicite, par exemple un système de coordonnées.

Cette composante permet de représenter la forme de l'objet lui-même et le positionner. Autrement dit la composante spatiale d'un objet géographique représente la localisation et la forme de cet objet.

- **Composante sémantique:** représente l'information relative à la nature, et les attributs décrivant un objet ou un phénomène du monde terrestre.
- **Composante géométrique:** définit sa localisation dans l'espace et sa morphologie élémentaire (point, ligne, surface ou pixel).

3 Représentation de l'information géographique

Avant de citer les différents façons de représentation de l'information géographique, il faut d'abord illustrer les types de cette information pour ne pas confondre entre les modes de représentation et les différents types qui sert:

1. **Image satellite:** l'information géographique peut être représentée sur une image enregistrée de la surface terrestre, où l'on peut voir une multitude d'objets mais sans connaître directement leurs attributs. La figure 1.1 est une image satellite [2].

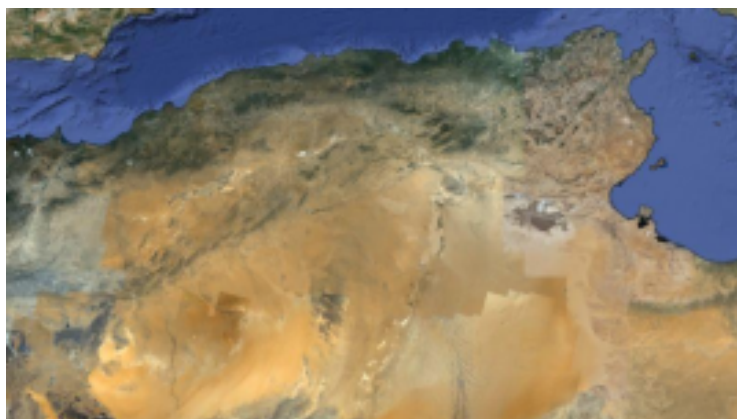


Figure 1.1: Image satellite.

2. **La carte:** L'information géographique se prête particulièrement bien à la représentation sur une carte, où l'on situe les objets et les phénomènes dans un repère général et homogène, qui donne une vue d'ensemble sur leur implantation sur le terrain.

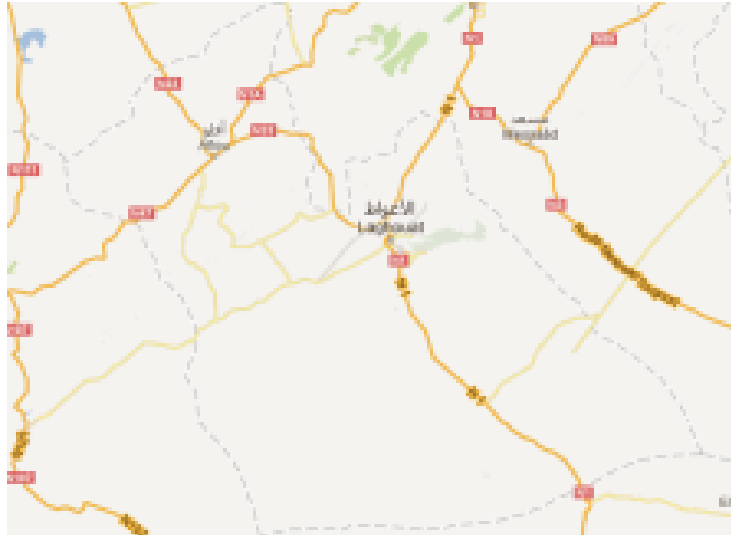


Figure 1.2: Carte.

3. **Le texte:** L'information géographique peut être représentée sous forme textuelle, un fichier de données littérales ou par des données numériques (exemple: fichier des abonnés au téléphone: nom, prénom, numéro de téléphone, adresse postale).

Il existe deux modes fondamentaux de représentation numérique des données géographiques :

- le mode matriciel **RASTERS**
- le mode vectoriel **VECTEURS**

Le mode raster: ou maillé représente un espace découpé selon une grille régulière de taille prédéterminée (la résolution). une valeur numérique unique est attribuée a chaque maille, pouvant correspondre à une mesure, à une

catégorie ou à l'identifiant d'un objet.

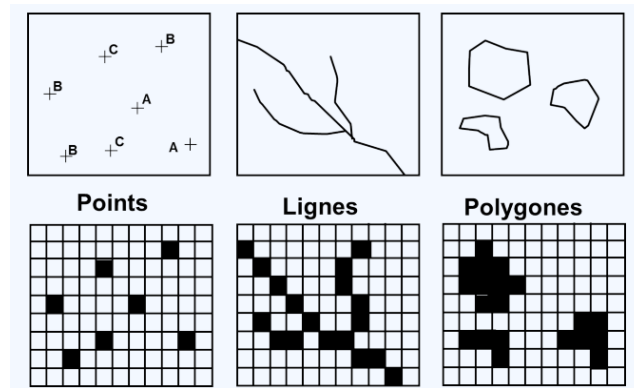


Figure 1.3: le mode raster

Le mode vecteur: permet de représenter le monde réel sous forme de points, lignes ou polygones. Ces entités spatiales sont définies par les coordonnées de leur contour dans un système de coordonnées cartésiennes en X et Y [3].

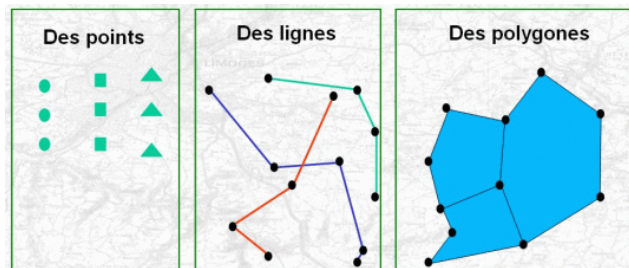


Figure 1.4: le mode vecteur

Les autres informations décrivant la donnée géographique sont portées par ses attributs (données attributaires).

Le concept couche: lors de la modélisation des systèmes d'information géographique, il faut essentiellement prévoir comment les différentes entités seront réparties en couches, par quels types d'éléments graphiques (ou

cartographiques) .

Un SIG stocke les informations du monde réel sous forme de couches thématiques pouvant être reliées les unes aux autres par la géographie. Dont chaque couche est un plan réunissant des éléments géographiques de même type.

Une couche peut aussi être vue comme un élément de la carte(une superposition des couches). Si on assemble tous les couches d'une carte on obtient la carte elles même.

4 Système d'Information Géographique

4.1 Définition

Un système d'information géographique (SIG) est un système d'information permettant de créer, d'organiser et de présenter des données alphanumériques spatialement référencées, autrement dit géoréférencées, ainsi que de produire des plans et des cartes. Ses usages couvrent les activités géomatiques de traitement, de partage et de diusion de l'information géographique. Dans les années 1960, les cartes de l'Afrique de l'Est trop nombreuses pour permettre de localiser les meilleurs endroits pour créer de nouvelles implantations forestières font naître l'idée d'utiliser l'informatique pour traiter les données géographiques, et cela fait la partition du premier SIG [4].

Objet géographique: Un objet géographique c'est le couple formé d'une description qualitative ou quantitative et d'une localisation spatiale, autrement dit un objet géographique est composé d'informations spatiales et de données attributaires. Ces données sont organisées de manière thématique.

4.2 Les composantes d'un SIG

Un SIG est constitué de cinq composants:

1. **Logiciel:** Les logiciels de SIG offrent les outils et les fonctions pour stocker, analyser et afficher toutes les informations, regroupant les aspects suivants:
 - Système de gestion de base de données. (SGBDR) qui supporte l'information géographique.
 - description Outils géographiques de requête, analyse et visualisation.
 - description Interface graphique utilisateur pour une utilisation facile.
2. **Les données géographiques:** Les données sont certainement la composante la plus importante dans un SIG. Plus les données tabulaires associées qui peuvent soit être constituées en interne, soit acquises auprès de producteurs de données.
3. **Méthodes:** La mise en œuvre et l'exploitation d'un SIG ne peut s'envisager sans le respect de certaines règles et procédures propres à chaque organisation.
4. **Matériel:** Les SIG fonctionnent aujourd'hui sur une très large gamme d'ordinateurs des serveurs de données aux ordinateurs de bureaux connectés en réseau ou utilisés de façon autonome.
5. **Utilisateurs:** Comme tous les utilisateurs des systèmes d'information géographiques ne sont pas forcément des spécialistes, un tel système propose une série de boîtes à outils que l'utilisateur assemble pour réaliser son projet. N'importe qui peut un jour ou l'autre être amené à utiliser un SIG. Ainsi que des spécialistes qui ont besoin d'analyser d'importants volumes de données géographiques selon des différents domaines.

4.3 Les domaines d'application des SIG

De nos jours l'émergence des SIG ne cesse de grandir presque dans tous les domaines, citons entre-autre:

- Tourisme (gestion des infrastructures, itinéraires touristiques)
- Marketing (localisation des clients, analyse du site)
- Planification urbaine (cadastre, POS, voirie, réseaux assainissement)
- Protection civile (gestion et prévention des catastrophes)
- Transport (planification des transports urbains, optimisation d'itinéraires)
- Hydrologie
- Forêt (cartographie pour aménagement, gestion des coupes et sylviculture)
- Géologie (prospection minière)
- Biologie (études du déplacement des populations animales)
- Télécoms (implantation d'antennes pour les téléphones mobiles)

4.4 Interrogation et analyse des données de SIG

Puisque un SIG comporte dans son ventre une BDD .Donc c'est purement claire que le langage standard d'interrogation des BDD sera utilisées .pour dialoguer avec les SIG.

Le langage SQL (Structured Query Language) est un langage de définition de données (LDD), c'est-à-dire qu'il permet de créer des tables dans une base de données relationnelle, ainsi que d'en modifier ou en supprimer. Le langage SQL peut être considéré comme le langage d'accès

normalisé aux bases de données. En fait SQL est un langage complet de gestion de bases de données relationnelles. Il a été conçu par IBM dans les années 70. Il est devenu le langage standard des systèmes de gestion de bases de données (SGBD) relationnelles (SGBDR).

Pour pousser plus loin le processus d'interrogation vers l'échelle spatiale, on fera appel au:

Requête spatiale: Est une interrogation portant sur la géométrie et la position des entités d'une ou plusieurs couches et permettant de sélectionner des entités en fonction des entités d'une autre couche. Une telle requête nécessite l'utilisation d'opérateurs de sélection géographique (intersection, inclusion, contiguïté, proximité...)[5].

5 Conclusion

Dans ce chapitre nous avons tracés le cadre général qui va supporter nos travaux ou bien nos expérimentations. Ce cadrant est délimités par les différentes notions du domaine des SIG, commençant par la définition de la géographie avec les informations géographiques, la source de données utilisées et ces différentes représentations numériques. On a aussi vu le SIG et ces composants, en précisant leur importance grâce aux nombreux domaines d'application qu'on a cités. On a fini ce chapitre par les requêtes spatiales. Car à partir de ce point on va commencer le prochain chapitre avec les composants du SIG, charger de recevoir ces requêtes spatiales et de y répondre.

Chapitre 2

Du relationnel au NoSQL en géomatique

1 Introduction

Dans ce chapitre on va rentrer dans le cœur d'un SIG là où résident les données géographiques et tabulaires, en générale il s'agit des BDD relationnelles avec leurs SGBDR accompagner du SQL pour l'interrogation. Ensuite on va entamer le nouveau paradigme du NOSQL et les techniques qu'ils offrent au SIG tout on étudiant les challenges lieu à l'interrogation de ce dernier dans les SIG

2 Base de données

Une base de données est un ensemble structuré d'informations, stocké sur un support informatique, conçu et réalisé pour que sa consultation et sa modification soient aisées, rapides et sûres, même par plusieurs utilisateurs concurrents. Une base de données est constituée de blocs élémentaires de données homogènes, regroupés sous forme d'enregistrements correspondant à des entités de même nature. Ces blocs

sont associés par des relations logiques ou géométriques. Il existe plusieurs technologies de bases de données, la plus dominante est celle fondée sur le modèle relationnel où les données sont organisées sous forme de tables dans lesquelles les lignes présentent les différentes occurrences de l'entité (les différentes parcelles de l'entité parcelle par exemple) et les colonnes les attributs de cette entité (la surface de la parcelle par exemple) [6] .

3 Systèmes de Gestion de Base de Données (SGBD)

Un système de gestion de base de données (SGBD) est une application qui sert comme son nom l'indique à stocker, manipuler et accéder aux données. La communication avec le SGBD se fait à travers un langage bien définie et structuré via des requêtes. En générale ce langage est le SQL (Standard Query Langage), mais récemment quand il s'agit des données non-structurées, on fait intervenir d'autres méthodes d'interrogation tel que le NOSQL [7].

Les SGBD sont basés sur une théorie, qu'on appelle le modèle relationnel, qui consiste à stocker toutes les données dans des tables structurées (en colonnes), avec des relations qui lient les tables entre elles. Cette modélisation est bénéfique.

D'une part en termes d'éviter la duplication des données (chaque information n'est stockée qu'à un seul endroit, il n'y a pas de redondance), et d'autre part d'optimiser au maximum les performances pour pouvoir accéder aux données et les présenter sous une forme explicite et utile.

3.1 SGBD relationnel

Un SGBDR (système de gestion de base de données relationnelle) est en effet un simple SGBD, qui en plus gère les relations, c'est-à-dire qu'on

peut définir des contraintes qui garantissent l'intégrité référentielle et fonctionnelle des données.

4 NoSQL

Depuis bien des années, les Système de Gestion de Bases de données(SGBD), utilisent la méthode relationnelle pour stocker les informations. C'est le cas des grandes SGBD tels que MySQL, Oracle, Postgres ou encore SQL Server. Cette méthode qui consiste à stocker les données dans des tables liées entre elles avec des relations a fait jusque-là ses preuves. Cependant, le coté statique de cette méthode qui constitue l'une de ses principales limites a poussé les chercheurs à s'orienter vers de nouvelles solutions plus adaptées aux besoins (exigences).C'est le cas de la méthode NOSQL.

5 Mais pourquoi nous avons besoin d'une nouvelle façon de sauvegarder nos données ?

Le premier besoin fondamental auquel répond NoSQL est la performance. En effet, ces dernières années, les géants du Web comme Google et Amazon ont vu leurs besoins en termes de charge et de volumétrie de données croître de façon exponentielle. Et c'est pour répondre à ces besoins que ses solutions ont vu le jour. Les architectes de ces organisations ont procédé à des compromis sur le caractère ACID des SGBDR. Ces intelligents compromis sur la notion de relationnel ont permis de dégager les SGBDR de leurs freins à la scalabilité horizontale et à l'évolutivité. Par la suite des entreprises comme Facebook, Twitter ou encore LinkedIn ont migré une partie de leurs données sur des bases NoSQL [8].

6 Caractéristiques principales des bases de données NoSQL

- **CAP:** Les bases de données NoSQL répondent aussi au théorème du CAP d'Eric Brewer qui est plus adapté aux systèmes distribués. Ce théorème énonce que tout système distribué peut répondre aux contraintes suivantes:

- **Cohérence :** tous les noeuds du système voient exactement les mêmes données au même moment;
- **Haute disponibilité (Availability) :** en cas de panne, les données restent accessibles;
- **Tolérance au Partitionnement :** le système peut être partitionné. Mais le théorème du CAP précise aussi que seulement deux de ces trois contraintes peuvent être respectées en même temps.

D'autres caractéristiques communes aux différentes bases de données NoSQL peuvent être citées tel que le partitionnement horizontal sur plusieurs noeuds, la réplication des données.

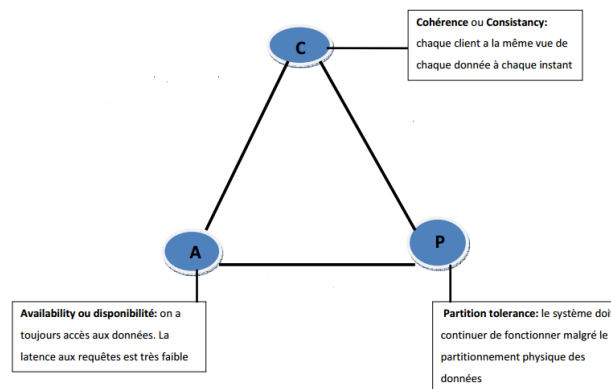


Figure 2.1: Théorème CAP

[9]

- Scalabilité horizontale:** Les bases de données de type NOSQL proposent une nouvelle représentation de l'information, en s'affranchissant des contraintes ACID du modèle SQL, elles ont le très gros avantage de fournir une architecture technique où il suffit de rajouter des serveurs pour gagner en performance sans trop se poser de questions. Cette technique consiste lorsque la charge des traitements ou des données devient très importante au niveau d'un ou de plusieurs serveurs à ajouter un ou plusieurs serveurs qui se partagent les données et les traitements. Cette approche de stockage permet d'avoir des bases de données performantes avec une disponibilité de données par rapport aux une BDD classiques qui ne peuvent égalet même en multipliant les serveurs miroirs. Un SGBDR pour répondre aux exigences de performance face aux gros volumes de données, doit se retourner vers du matériel de plus en plus rapide dotés d'une capacité élevée de mémoire tandis que, le NoSQL pour sa part, pour gérer la "montée en charge" se réfère à la répartition de la charge sur les systèmes de Cloud Computing. Il s'agit là de composant de NoSQL qui fait d'elle une solution peu coûteuse pour les grands ensembles de données.

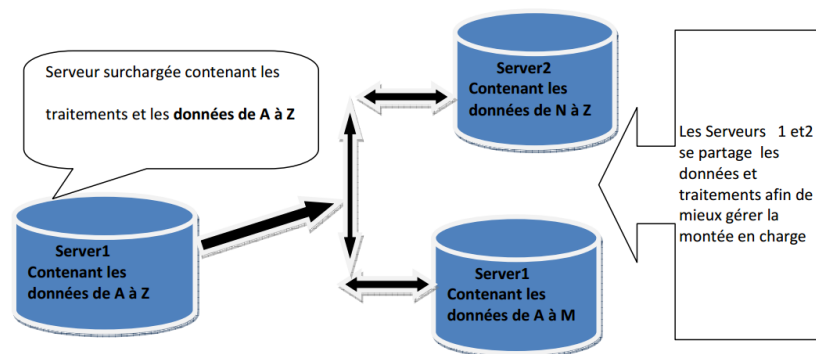


Figure 2.2: Partitionnement Horizontal de données

[10]

7 Typologies de NoSQL

Lorsque l'on parle de NoSQL, on regroupe des systèmes de base de données qui ne sont pas relationnels, mais il faut savoir qu'il existe plusieurs types de bases de données "NoSQL", et non pas un type à part entière.

Quatre grandes catégories se distinguent parmi celles-ci

- **a) Les bases clef/valeur:**

Permettent de stocker des informations sous forme d'un couple clef/valeur où la valeur peut être une chaîne de caractère, un entier ou un objet sérialisé. Une base de données Redis un (clé valeur scalable). Ce type de base de données offre de très bonnes performances par sa simplicité et peut même être utilisé pour stocker les sessions utilisateur ou le cache de votre site par exemple.

la figure 2.3 illustre un exemple d'une base clef/valeur.

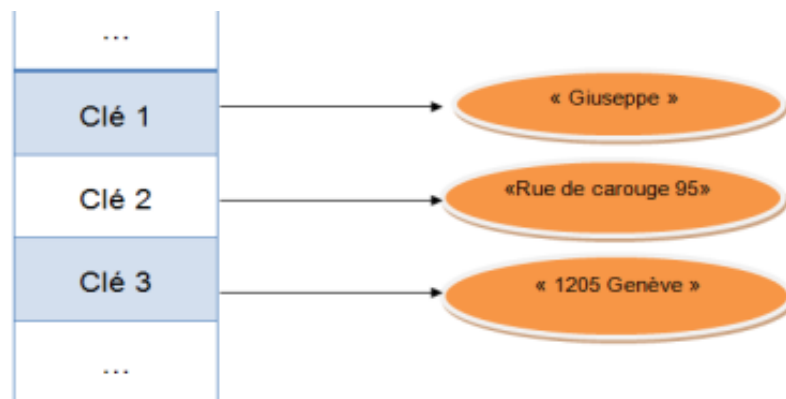


Figure 2.3: représentation clef/valeur

[11]

- **b) Les bases orientées colonnes:**

Ressemble aux bases de données relationnelles, car les données sont

sauvegardées sous forme de ligne avec des colonnes, mais se distingue par le fait que le nombre de colonnes peut varier d'une ligne à l'autre. Les solutions les plus connues sont les bases de données HBase ou Cassandra. la figure 2.4 explique la différence entre un schéma relationnel et une base orientée colonne.

	Nom	Fixe	Mobile
1	Patrick	0102030405	NULL
2	Farid	NULL	0601020304
3	Tim	NULL	NULL

Modèle d'une base de données relationnelle

1	Nom	Fixe
	Patrick	0102030405
2	Nom	Mobile
	Farid	0601020304
3	Nom	
	Tim	

Modèle d'une base de données NoSQL famille colonne

Figure 2.4: schéma R vs colonne

[12]

- **c) Les bases orientées document:**

représente les informations sous forme d'objet XML ou JSON. L'avantage est de pouvoir récupérer simplement des informations structurées de manière hiérarchique. Les solutions les plus connues sont CouchDB, RavenDB et MongoDB.

- **d) Les bases orientées graphe:**

Présentent les données sous forme de noeud et de relation. Cette structure permet de récupérer simplement des relations complexes. Un exemple de base graphe est Neo4J.

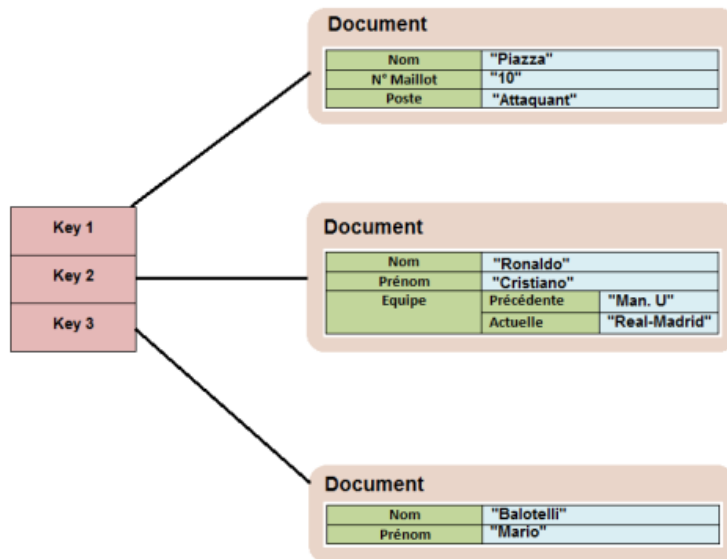


Figure 2.5: Les bases orientées document

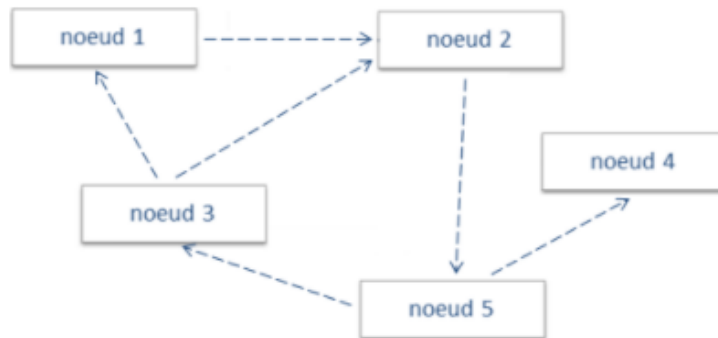


Figure 2.6: Les bases orientées graphe

8 Principaux avantages du NoSQL

Les bases de données No-SQL ont les avantages suivants :

- Leurs performances ne s'écroulent jamais quel que soit le volume traité. Leur temps de réponse est proportionnel au volume (on observe une dérive quadratique dans les SGBDR classiques).

- Elles se migrent facilement, contrairement aux SGBDR classiques, il n'est pas nécessaire de procéder à une interruption de service pour effectuer le déploiement d'une fonctionnalité impactant les modèles de données.
- Elles sont facilement scalables. A titre d'exemple, le plus gros cluster de NoSQL fait 400 To tandis qu'Oracle sait traiter jusqu'à une vingtaine de Téraoctet (pour des temps de réponse raisonnable).
- Elles sont consistantes de manière pratique (pour l'utilisateur une requête aura toujours la même réponse quel que soit le nœud du cluster).
- Elles s'intègrent facilement aux SI déployés dans les Clouds du marché.
- Elles possèdent un modèle extensible (le nombre de colonne d'une table n'est pas défini).

9 Base de données géographiques

C'est un ensemble des données spatiales et non spatiales structurées et organisées de manière à être interrogeables et analysables de façon interactive ou automatique. Une base de données géographique concerne habituellement une zone définie et elle est gérée par un logiciel SIG , aussi elle intègre les données avec leurs métadonnées. Ces données sont représentées sous forme numérique peuvent se distinguer en trois catégories :

- **Les données géométriques** liées au positionnement et à la forme des objets ;

- **Les données attributaires** qui matérialisent les caractéristiques des objets ;
- **Les métas données** qui décrivent les données.

10 NoSQL en géomatique

Plusieurs implémentations des bases de données NoSQL existe aujourd'hui comme cela a été souligné dans le début de ce chapitre. Certaines incluent nativement la gestion des données géo spatiale et d'autre par le biais des extensions. Une autre catégorie de solutions n'a pas été conçus pour les applications géo spatiales, mais ont été mis en œuvre pour gérer les données spatiales. La section suivante présente quelques solutions qui sont couramment utilisées pour gérer les données spatiales [13].

10.1 Présentation de quelques solutions

(a) MongoDB

MongoDB est une implémentation du NoSQL dans le domaine géo-spatial. Elle se base sur le modèle orienté document et est disponible sur la plus part des plateformes. Elle permet la manipulation des objets structurés sous le format BSON et permet d'indexer n'importe quel attribut, sans schéma prédéterminé. Elle a été conçue pour optimiser les requêtes avec JavaScript et pour améliorer les services de l'API REST.

MongoDB permet l'indexation géo spatiale nativement (sans extension). Plusieurs applications de géolocalisation utilisent MongoDB ,par exemple :

- **Foursquare** (<https://foursquare.com>): Un outil web qui donne permet à l'utilisateur de trouver facilement sa position. Il permet également de facilement trouver les endroits publics tels que les restaurants, les salles de cafés ou même les magasins.
- **Scrabbly** : pour gérer et géo localiser les joueurs de scrabble dans le monde.

(b) **Cassandra**

est un système NoSQL puissant en géolocalisation, originellement créé par FACEBOOK sorti en 2008. Il est orienté colonne, et open source.

SimpleGeo : est l'une des implémentations qui utilise Cassandra pour fournir des services et des produits basés sur la géolocalisation.

(c) **Google Big Table**

BigTable est un système de gestion de base de données développé par Google. Il est optimisé pour la compression des données et offre une haute performance en termes de gestion des données. Cette implémentation du NoSQL dans le domaine géo-spatiale est cependant disponible au travers de la plateforme Google App Engine. Plusieurs applications ont tirés leurs racines à partir de ce système de bases de données. C'est le cas par exemple de Google Earth.

(d) CouchDB

Tout comme MongoDB, CouchDB est une base de données orientée documents qui permet de manipuler des objets structurés en JSON, sans schéma, ni table. Elle possède plusieurs extensions de gestion de données, parmi lesquelles GeoCouch, qui est l'extension spatiale intégré à CouchBase simple Server.

11 Constat sur le NoSQL dans le géomatique :

La base de données relationnel est le support de stockage usuel des données géographique, et le NOSQL est un nouveau type de base de données qui a de la souplesse, il support un très grand volume de données (BIG DATA), ces principaux caractéristiques mentionnés dans la section 6.

Dans cette section on a voulez regroupé tous les idées constatées et les connaissances rencontrées dans différents articles et mémoires [14],[15] , [8] , [16] , [17] tout au long de notre recherche bibliographique sur ce domaine, donc on va passer à ce que le NoSql fait ,et ce qu'il ne fait pas et qu'est ce qu'il offre ou propose :

• Ce qu'il fait :

- (a) L'absence de modèle unique
- (b) Le théorème de CAP : un système distribué qui ne garantit pas
 - La cohérence ;
 - Haut disponibilité ;
 - Tolérance au partitionnement.
- (c) L'anti abstraction : auparavant les éditeurs sont forcés de séparer les couches physiques aux couches logiques, mais les

implémentations NOSQL fait l'inverse, par des algorithmes afin de coupler l'infrastructure, les couches middleware et les logiciels.

- (d) Les besoins : (régler par NOSQL)
- Extensibilité : l'ajout d'un enregistrement ne doit pas dégrader le fonctionnement ;
 - Résilience : (Théorème CAP) en cas de fail;
 - Performance : le coût d'une jointure (en temps) des tables volumineux d'un SGBDR n'existe pas en NOSQL;
 - Souplesse de structuration : le mode de mémorisation de donnée (colonne ou document) permet l'ajout dynamique des données.

• **Ce qu'il ne fait pas :**

- (a) Outils de manipulation de donnée limitée
- Pas de fonction relationnelle ;
 - Pas de langage de requête performant;
 - Open Source est très actif, et il se progresse mais encore loin des SGBDR.
- (b) Qualité de transaction
- Difficile d'assurer ACID (Atomicité, Cohérence, Isolation, Durabilité)
- (c) Classification classique
- Base de données Clé-Valeur : point d'entrée unique de donnée;
 - Base de données orienté document Permet l'ajout, la suppression des champs et l'indexation (JSON);

- Base de données orienté colonnes (ex : Facebook);
 - * Peut atteindre des millions de colonnes;
 - * Peut s’adapté au M.A.J fréquent ;
 - * Interrogation est asses limité.
- Il offre :

Type de Base	Bases	Implémentations célèbres
Clé – Valeur	Redis, Riak, Voldemort	Linkedin
Colonnes	BigTable, HBase, Cassandra, DynamoDB	Amazon, Facebook, Google
Document	MongoDB, CouchDB, CouchBase, RavenDB	Expedia, SAP, Salesforce
Graphe	Neo4j, Oracle Graph	WalMart, Accenture, Boomberg, eBay

[18]

- Le NoSql est proposé par :
 - Oracle NOSQL, Oracle graphe ;
 - PostgreSQL : Fournit une solution d’indexation des fonctions pour manipuler les données JSON ;
 - MongoDB : un acteur de type document, multi instance et support le sharding, création d’index et il a un interpréteur de

- requête. en général en ligne de commande mais il existe des GUI (Ghengis, Umongo ...);
- Hadoop : applique Map/Reduce permet d'agréger des traitements sur des millions (milliard) d'enregistrement, une représentation cartographique rapide (des heures en SQL);
 - GeoCouch : orienté Document interface graphique FUTON ;
 - Neo4j : orienté Graphe Open Source.
- Quelques réflexions : Ici on regroupe les éléments importants rencontrés durant notre lecture des différents documents sous forme de points:
 - (a) Pas de substitution en mode relationnel, difficile de remplacer SGBDR par NOSQL ;
 - (b) Une offre devenir ; ex : Google (Big Table) et google earth permet de confirmer son intérêt.
 - (c) Le maintien de la modélisation des données géographique NOSQL ne signifie pas que la modélisation devient inutile ;
 - (d) o Choix d'une solution par rapport au besoin non à l'offre Une étude préalable est nécessaire pour la réussite (colonne, document et graphe) ex : oublier NoSQL pour gérer 500 Mo de données;
 - (e) Une évolution technique SIG en corrélation avec la technique SI le SIG suit les technologies récentes du SI, utilisation de JavaScript, JSON ;
 - (f) o Le NEWSQL : l'approche « Best of Breed » doit apporter le meilleur des deux modes [19].

11.1 Propriétés BASE

(a) SGBD issus de la nuagique (cloud computing) et des systèmes distribués :

- privilégiant la haute disponibilité des données (distribuées), la rapidité, la simplicité
- au détriment de la cohérence, de l'exactitude de la réponse

(b) Propriétés BASE :

- Basically Available : le système doit toujours être accessible (ou indisponible sur de courtes périodes)
- Soft state : l'état de la BD n'est pas garanti à un instant donné (les mises à jour ne sont pas immédiates : cf. cohérence à terme)
- Eventual consistency : la cohérence des données à un instant donné n'est pas primordiale (mais assurée à terme : verrouillage optimiste en reportant à plus tard la vérification de l'intégrité)

12 conclusion

Ce chapitre nous a faciliter la compréhension de SIG en général, et on a cité les avantages de NOSQL qui est nécessaire pour la manipulation de SIG, Aujourd'hui le NoSQL apporte une nouvelle façon d'appréhender la modélisation des données. Cette nouvelle technologie n'est pas là pour remplacer les bases de données relationnelles, elle répond à des besoins différents mais les deux approches peuvent cohabiter. Le choix de l'une ou de l'autre sera donc fortement dépend du contexte et du besoin. Toute l'acheminement de nos idées et nos réflexions dans ce sens , on a dessinés avec les cartes du Mind Mapping et ils sont disponible dans l'Annexe 1.

Le prochain chapitre sera consacré à l'étude expérimentale où nous allons citer tous les processus d'installation et de l'apprentissage des outils SIG et NoSQL, plus l'interrogation des deux types de base de données.

Chapitre 3

Résultats et discussions

1 Introduction

Dans ce chapitre nous allons présenter les différents outils que nous avons utilisés pour la réalisation d'une base de données géographique et l'autre base NOsql , ainsi que les démarches que nous avons poursuivies à la fois pour l'installation délicate des outils surtout du côté NoSQL, et pour l'apprentissage des outils de SIG.

Ensuite on va détailler avec des captures d'écrans tout le processus d'interrogation des deux bases de données que nous avons conçues pour faire les tests. Du côté ArcGis ça sera le SQL, et pour le MongoDB ça sera le NoSQL (Orienté document).

2 Le système d'information géographique

ArcGIS est un système complet qui permet de collecter, organiser, gérer, analyser, communiquer et diffuser des informations géographiques. En tant que principale plateforme de développement et d'utilisation des systèmes d'informations géographiques (SIG) au monde, ArcGIS est

utilisé par des personnes du monde entier pour mettre les connaissances géographiques au service du gouvernement, des entreprises, de la science, de l'éducation et des médias. ArcGIS permet la publication des informations géographiques afin qu'elles puissent être accessibles et utilisables par quiconque. Le système est disponible partout au moyen de navigateurs Web, d'appareils mobiles tels que des smartphones et d'ordinateurs de bureau.

Les professionnels dans de nombreux secteurs utilisent ArcGIS dans un large éventail d'applications, incluant la planification et l'analyse, la gestion d'actifs, la sensibilisation opérationnelle, les opérations sur le terrain telles que l'inspection mobile et le déploiement d'urgence, les études de marché, la gestion des ressources, la logistique, l'éducation et la sensibilisation. En général, ArcGIS vous permet de:

- Créer, partager et utiliser des cartes intelligentes;
- Rassembler les informations géographiques;
- Créer et gérer des bases de données géographiques;
- Résoudre des problèmes avec l'analyse spatiale;
- Créer des applications basées sur la carte;
- Communiquer et partager des informations grâce à la puissance de la géographie et de la visualisation.

L'installation de ArcGIS nécessite elle-même une formation, pour cela on a suivi la formation au mode vidéo qui nous à pris avec l'apprentissage de l'environnement .

3 La base de données géographique

On a laissé le choix de la bdd par défaut dans ArcGis qui est le SQLite.

3.1 SQLite

SQLite est une bibliothèque écrite en C qui propose un moteur de base de données relationnelle accessible par le langage SQL. SQLite c'est le moteur de base de données le plus utilisé au monde, grâce à son utilisation dans de nombreux logiciels grand public comme Firefox, Skype.

Nous avons travaillé sur la carte de l'Algérie avec ses 48 Wilayas dont chaque wilaya contient des communes et pour chacune d'elles on a les informations (superficie, population, nbstade, nbrestaurant et nbparking), on est arrivé à une BDD de taille 15,6 Mo.

La distribution des communes dans chaque wilaya avec leurs noms et superficie, sont des données réelles, tandis que pour le restes des attributs nous n'avons pas trouver des données réelles, donc on utilisés un script random pour générer ces données et continuer nos tests.

4 Du côté de NoSQL

4.1 La BDD MongoDB

Nous avons déjà décrit MongoDB dans la section [5a](#).

Le processus d'installation de MangoDB dans Windows 7 est une etape fastidieuse par ce que a l'origine MongoDB est sensé roulé.

4.1.1 JSON

JSON est le seul langage d'interprétation de données reconnue par le mongoDB.

Définition

JSON (JavaScript Object Notation – Notation Objet issue de JavaScript) est un format léger d'échange de données. Il est facile à lire ou à écrire pour des humains. Il est aisément analysable ou général par des machines. Il est basé sur un sous-ensemble du langage de programmation.

JSON se base sur deux structures:

- Une collection de couples nom/valeur. Divers langages la réifient par un objet, un enregistrement, une structure, un dictionnaire, une table de hachage, une liste typée ou un tableau associatif.
- Une liste de valeurs ordonnées. La plupart des langages la réifient par un tableau, un vecteur, une liste ou une suite [20].

4.1.2 IntelliJ IDEA

Définition

IntelliJ IDEA est un IDE Java commercial développé par JetBrains, il est un environnement de programmation complet qui se présente sous la forme d'une application. L'outil aide les développeurs à concevoir et à documenter leurs codes comme un traitement de texte aide à produire des documents écrits.

A ce stade nous n'avons pas trouvé une meilleure façon de décrire tous le processus dès l'installation des outils de teste, en passant par le placage des données sur la carte ArcGIS et la migration vers MongoDB, jusqu'au

la phase des requêtes, mieux qu'un scénario doté de captures d'écran pas à pas

- Etape1 :
Conversion du fichier ArcMap en fichier Geojson dans QGis pour pouvoir ensuite le charger vers le site web ¹ qui nous donne un fichier en format standard json reconue par MongoDB.
- Etape2 :
 - a) Importation des données de la collection Algeria depuis le fichier Algeria.json par la commande **MongoImport** .
 - b) Importation des données des collections STADE, PARKING, RESTAURANT, depuis le fichier Algeria.json par la commande MongoImport.
- Etape3 :
 - (a) Pour pouvoir manipuler nos données stockées dans MongoDB sans passer par la ligne de commande, on a fait appel à IntelliJ IDEA.
 - (b) La Configuration et l'installation du plugin Mongo dans l'IntelliJ IDEA .

¹web = <http://www.convertcsv.com/csv-to-json.htm>

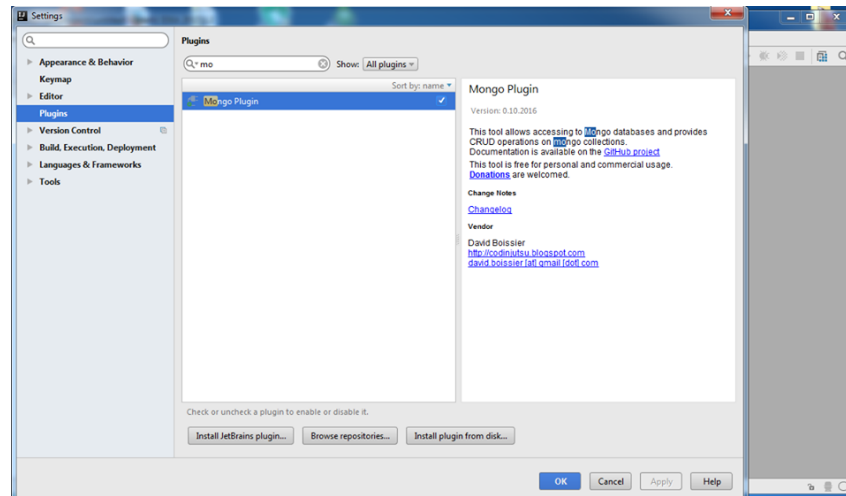


Figure 3.1: La Configuration et l'installation du plugin Mongo dans l'IntelliJ IDEA

- Etape 4:

Affichage de la carte de l'Algerie dans ArcGis

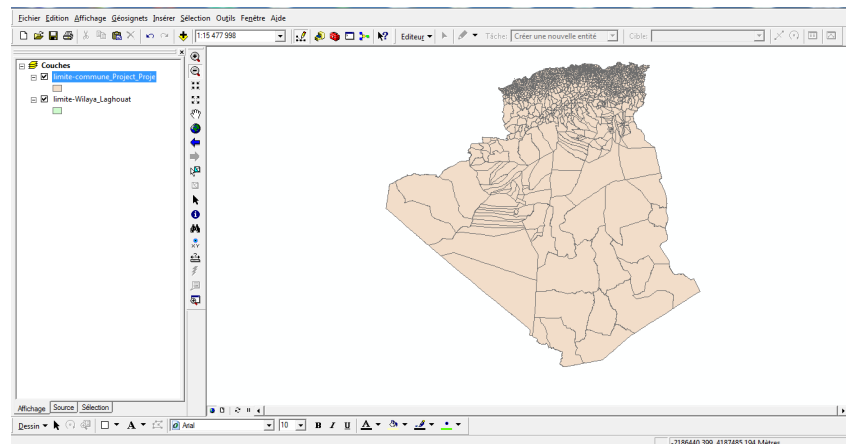


Figure 3.2: la carte de l'Algerie dans ArcGis

Affichage de la carte de l'Algerie avec les points de : Parking, Restaurant, Stade en ArcGis

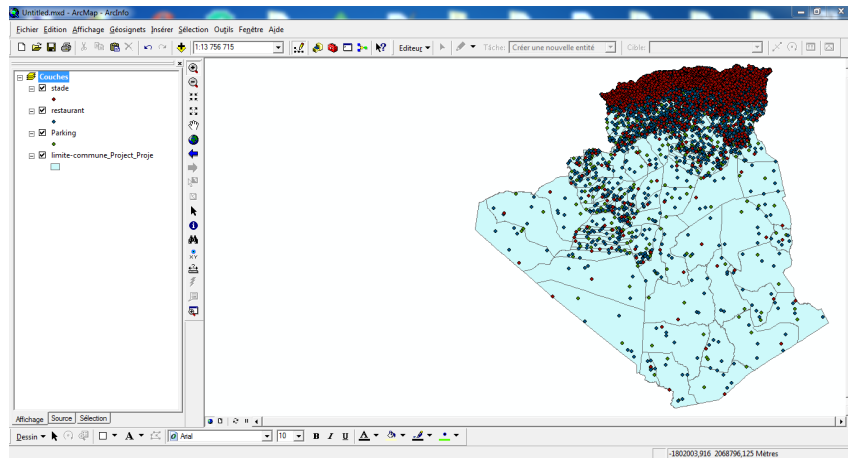


Figure 3.3: la carte de Parking, Restaurant, Stade en ArcGis

Propriétés de la Tables Attributaire de la couche Algeria et ses champs

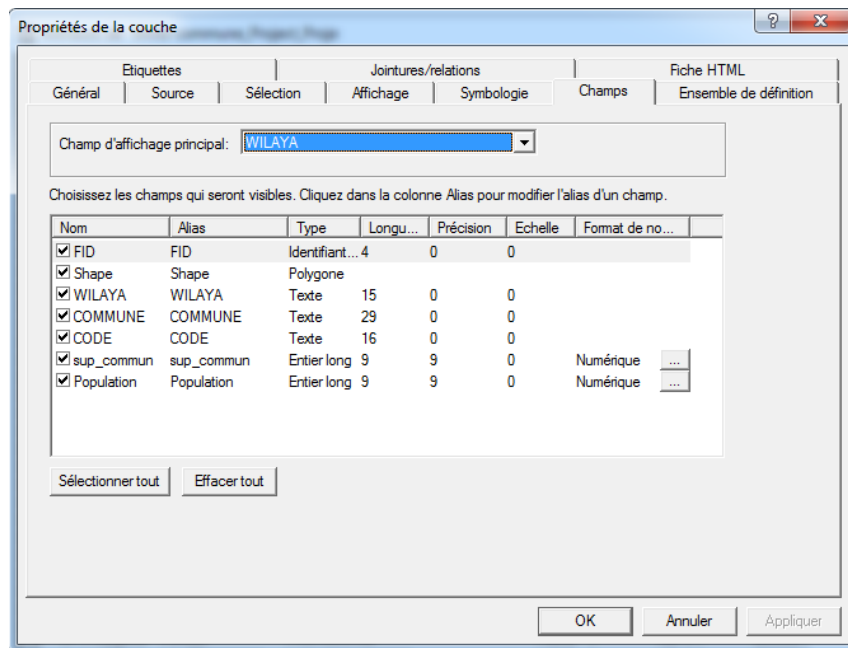


Figure 3.4: la couche Algeria et ses champs

- Etape 5: Les requêtes spatiales

On a choisi 2 requêtes spatiales complexes et une requête spatiale simple pour tester les deux environnements qui sont totalement différentes, ArcGis et MongoDB à travers IntelliJ IDEA ;

Requête 1 : Sélection de la commune Hassi Messaoud et rechercher les Stade les plus proche d'une restaurant données.

Résultat Le Stade le plus proche est de numéro 5552 d'une distance 70.315 mètres

Requête 2 :

Trouvez deux parkings proches situer dans de deux communes différents. Les étapes :

- (a) Sélection de la première commune
- (b) Sélection de la deuxième commune
- (c) sélection Les parkings de la première commune

Résultat : Le parking le plus proche des parking de commune 1 est le parking numéro 2512 avec une distance de 171.038 mètres.

Requête 3 :

- Sur ArcGis : Il existe 57 communes ou la population est mois de 40.000 habitants.
- Sur MongoDB : Exécution de la commande find() sur l'attribut communes, avec le critère de population moins de 40.000 habitant

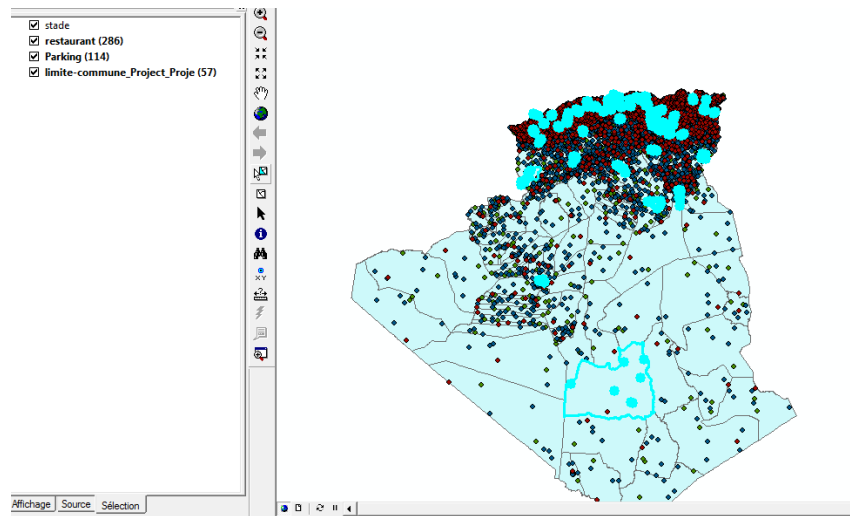


Figure 3.5: Résultat de Requête 3 sur ArcGIS

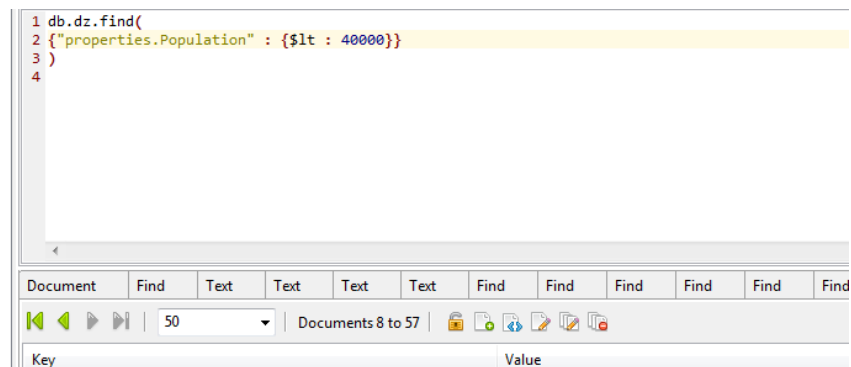


Figure 3.6: Résultat de Requête 3 sur MongoDB

Conclusion

Dans ce chapitre nous avons montré les différents outils que nous avons utilisés pour la réalisation d'une base de données géographique, ensuite on a illustré comment on a fait la migration vers une base de données complètement différente celui de NOsql. Ensuite on a détaillé avec des captures d'écrans tout le processus d'interrogation des deux bases de données que nous avons conçus pour faire les tests. Du côté

ArcGis les trois (3) requêtes donnent des résultats, tandis que, pour le MongoDB seul la dernière requête est exécutable sans erreurs et cela dû au fait que le MongoDB nécessite encore d'autre mécanisme très poussé pour pouvoir manipuler des données géospatiales.

Conclusion générale

Dès le début de ce travail, on a fixé un objectif bien précis, il était pour nous d'étudier la possibilité d'intégrer une base de données de type NOSQL au lieu d'une Bases de Données Relationnelle dans le cœur dans un système d'information géographique. Pour ce faire, nous avons commencé par définir l'information géographique avec ces composantes ainsi que leur mode de représentation, tout en montrant leur gestionnaire de base de données géo-spatiale.

Par la suite, on a présenté une nouvelle approche très répondue de nos jours pour la gestion de toute type de données spécialement les données volumineuses que le paradigme NOSQL, et qui ne se base plus sur l'approche classique du relationnelle, son objectif étant l'atteindre des meilleures performances quel que soit la nature et la structuration des informations. Nous avons souligné que le NOSQL regroupe quatre grandes famille de Base de Données chacune peut être utilisé dans le domaine géométrique.

La dernière partie de notre mémoire est consacré à l'utilisation du NOSQL dans le processus de développement d'un SIG, après voir acquis les connaissances de base de domaine géo-spatiale, plus l'apprentissage des outils de manipulation de l'information géographique. Et aussi après un grand parcours des solutions qui sont actuellement utilisés dans le NOSQL, nous nous sommes

intéressé à MongoDB qui possède plusieurs applications de gestion de données comme IntelliJ IDEA et STUDIO 3T qui permet de faire l'interrogation des données spatiale.

Finalement on peut dire qu'on a atteint notre objectif même que nous n'avons pas arrivé à exécutés les requêtes spatiales complexes sur les données de type orienté document, mais au moins on a ouvert la voix pour d'autre chercheurs qui viennent après nous. On sait très bien qu'ils vont trouver les choses claires et ils vont arriver à continuer nos travaux avec succès.

PERSPECTIVES

On voulait pousser plus loin notre travail avec la visualisation des résultats obtenus dans MongoDB sur une carte comme dans le cas de ArcMap, mais malheureusement ça nécessite encore un effort d'apprentissage de nouvelle technique et du temps.

Dans le futur INCHALLAH, nous voulons étudier la possibilité d'utiliser le reste de la bench MEAN STACK ¹ pour produire des systèmes d'information géographique très rapides, puissants, efficaces et dans des délais brefs avec des couts réduits.

¹The MEAN stack is MongoDB, Express.js, AngularJS (or Angular), and Node.js

Bibliographie

- [1] information géographique, . URL <https://www.aquaportail.com/definition-6409-information-geographique.html>. 5
- [2] définition, typologie, exemples. URL <http://seig.ensg.ign.fr/fiche.php?NOFICHE=FP1&NOCONT=&NOCHEM=&NOLISTE=&NORETOUR=FP15>,. 6
- [3] Glossaire des sig. URL http://www.ente-aix.fr/documents/118-demoGeo/demo/4_BasesIG/co/20_gr_PubliRasterVecteur.html,. 8
- [4] définition de sig. URL <http://www.sig-geomatique.fr/sig-sig.html>,. 9
- [5] définition de rs. URL http://www.emse.fr/tice/uved/SIG/Glossaire/co/Requete_spatiale.html,. 12
- [6] Base de données. URL http://www.emse.fr/tice/uved/SIG/Glossaire/co/SGBD_1.html,. 14
- [7] définition de sgbd. URL http://www.emse.fr/tice/uved/SIG/Glossaire/co/SGBD_1.html,. 14
- [8] KOUEDI Emmanuel. Approche de migration d'une base de données relationnelle vers une base de données nosql orientée colonne., . 15, 24

- [9] KOUEDEI Emmanuel. Approche de migration d'une base de données relationnelle vers une base de données nosql orientée colonne., .
- [10] KOUEDEI Emmanuel. Approche de migration d'une base de données relationnelle vers une base de données nosql orientée colonne., . [17](#)
- [11] KOUEDEI Emmanuel. Approche de migration d'une base de données relationnelle vers une base de données nosql orientée colonne., . [18](#)
- [12] Etude comparative bdd relationnelle versus nosq, . URL <https://blog.talanlabs.com/etude-comparative-bdd-relationnelle-versus-nosql/>. [19](#)
- [13] CHRISTIAN CAROLIN. Bases de données no sql et sig : d'un existant restreint à un avenir prometteur. . [22](#)
- [14] Meyer Léonard. *L'AVENIR DU NoSQL*. 2014. [24](#)
- [15] Rémy Frenoy. Design and implementation of a nosql solution on a software as a service platform. [24](#)
- [16] Olivier Guibert. *Cours de Bases de Données : NoSQL*. 2016. [24](#)
- [17] Adriano Girolamo PIAZZA. *NoSQL Etat de l'art et benchmark*. 2013. [24](#)
- [18] KOUEDEI Emmanuel. Approche de migration d'une base de données relationnelle vers une base de données nosql orientée colonne., . [26](#)

- [19] CHRISTIAN CAROLIN. Bases de données no sql et sig : d'un existant restreint à un avenir prometteur. . 27
- [20] Json. URL <http://slidewiki.org/slide/29108>,. 33

Annexe 1

