



RAPPORT DE STAGE

Conception et Développement d'une application de géolocalisation orientée santé publique

Réalisé Par

BRAHIMI ANIS

Etudiant Licence3 Informatique UPEC

Encadré par

Malika Grim-Yefsah

Enseignant-Chercheur ENSG-IGN

Table des matières

Re	emer	ciements	3
1.	IN	TRODUCTION	4
2.	Co	ntexte	5
2	2.1.	Objectifs proposés	.5
2	2.2.	L'information géographique	6
2	2.3.	Les systèmes d'Information géographique (SIG)	6
3.	Pré	ésentation de l'organisme d'accueil	7
	3.1.	Description de l'Institut national de l'information géographique et forestière (IGN)	
	3.1.	1. Définition	. 7
	3.1.	2. Quelques chiffres sur IGN	7
	3.1.	3. Les activités de L'IGN	. 7
3	3.2. Ľ	organigramme de L'IGN	8
3	3.3. Le	es produits et services IGN	9
	3.3.	1. Les services IGN	9
	3.3.	2. Les bases de données IGN1	LO
4.	La	réalisation1	1
4	1.1.	Collection de données1	.1
	Les	différentes données collectées sont :	1
2	1.2.	Modélisation par des diagrammes UML1	.2
	4.2.	1. Diagramme de classes1	13
	4.2.	2. Diagramme de cas d'utilisation	4
4	1.3.	Développement d'une base de données Neo4j1	.5
	4.3.	1. Généralité sur les bases de données1	L 5
	4.3.	2. Les bases de données NoSQL	ا5
	4.3.	3. Les différents types de base de données NoSQL	ا5
	4.3.	4. Implémentation d'une base de données orientée graphe Neo4j	٦,
4	1.4.	Transfert de données	.9
4	1.5.	Implémentation de l'interface graphique de l'application en Java2	
	4.5.		
	4.5.	2. Les autres parties	<u>2</u> 5
5.	Co	nclusion et perspectives2	6
Le	s an	nexes2	7
Le	s Ré	férences :	9

Remerciements

Je tiens à remercier Madame Malika Grim-Yefsah pour avoir proposé ce stage et de m'avoir encadré et appuyé tout au long de son avancement.

Je tiens à remercier également Monsieur Alain Chaumet pour ses précieux conseils.

1. INTRODUCTION

Etudiant en 3éme année licence en informatique à l'université Paris-Est Créteil, j'effectue un stage intitulé « Conception et Développement d'une application de géolocalisation orientée santé publique » au sein de l'institut national de l'information géographique et forestière (IGN) entre le 08/04/2019 au 28/06/2019. L'IGN est un établissement public qui assure la production et la diffusion de l'information géographique composé de plusieurs directions dont une direction d'enseignement et de recherche nommée « Ecole Nationale des sciences géographiques (ENSG-Géomatique) ». Pendant ce mois de stage, je suis localisé au centre de compétences des techniques des systèmes d'information (CCTSI) à l'ENSG. Ce centre est composé d'une dizaine d'agents : deux (02) enseignants, deux (02) chercheurs, deux (02) enseignant-chercheur, trois (03) ingénieurs et un (01) responsable du centre.

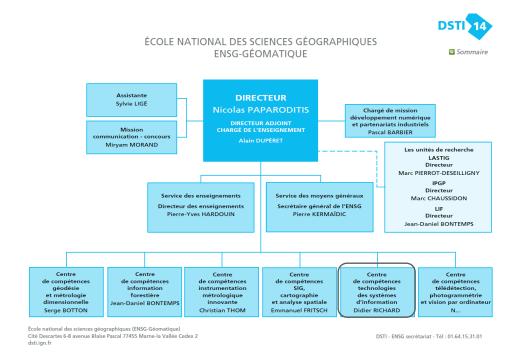


Figure 1: organigramme ENSG et ma localisation pendant le stage

Mon souhait de faire un master en informatique décisionnelle m'a poussé à choisir ce stage qui propose à la fois la modélisation d'un problème, le développement d'une base de données et la programmation d'une application.

Le but de ce stage était de concevoir et développer une application de géolocalisation orientée santé public en utilisant une base de données graphe, où nous avons opté pour la base Neo4j. Dans un premier temps nous décrirons l'entreprise et ses activités puis nous présentons les missions réalisées lors du stage et nous terminons par montrer les apports du stage.

2. Contexte

Des études ont montré que les expositions de courts et longs termes aux environnements naturels améliorent le bien-être et la santé, alors que vivre en milieu urbain (associé à une exposition réduite aux milieux naturels) a été associe à un faible bien être et une mauvaise santé. Ces études montrent aussi que ces observations semblent être liées aux facteurs socio-économiques. La collecte massive de différentes données (géographique, physiologique, socio-économique, environnementale etc.) et leurs combinaisons et leur traitement permettent de travailler sur des modèles prédictifs d'appariations et de disséminations de maladie de plus en plus performants.

L'IGN s'intéresse à cette problématique en collaboration avec l'Inserm (Institut National de Santé et de Recherche Médicale). Dans le cadre de cette collaboration que ce sujet a été proposé.

2.1. Objectifs proposés

Notre objectif est de montrer via une application l'influence de ces milieux (environnementaux, socio-économique, ..) sur la santé d'un individu en analysant les services et les infrastructures existants dans son quartier par rapport aux nombres, coût et la distance qui séparent les individus de ces infrastructures.

En raison de manque de données et leurs confidentialités, nous n'avons pas pu travailler sur la corrélation entre la présence de ces infrastructures et des phénomènes de santé publique comme, l'obésité ou la dépression.

Dans notre travail nous allons modéliser via un diagramme de classe UML les facteurs qui influencent un individu repérer dans un quartier en étudiant le milieu environnemental avec la quantité d'espaces verts et bleus qui l'entourent comme les jardins, les lacs, les pistes cyclables et les lieux ou on peut pratiquer la marche-à-pied, aussi les centres sportifs et les salles de loisirs et de cultures, on s'intéressera aussi sur le milieu socio-économique d'un individu comme son niveau de revenus ou son statut d'activité professionnelle (être actif, retraité, ou chômeur).

Et pour étudier le temps et la distance qui séparent ces infrastructures à l'individu on a ajouté un autre facteur qui est le transport, celui-là influence beaucoup, par exemple un

individu qui vit dans un quartier où les moyens de transports sont près de chez lui a plus de chance d'aller à une salle de sport qu'un autre qui est loin de ces infrastructures.

2.2. L'information géographique

L'information géographique est l'information liée à la représentation spatiale d'un phénomène ou d'un objet avec toutes ses dimensions à un moment donné.

Les informations géographiques sont traditionnellement stockées sur des cartes papier avec une échelle bien spécifié. Aujourd'hui l'information géographique est traitée comme beaucoup d'autres types d'informations dans des systèmes informatisés.

On trouve deux types d'information géographique :

- Informations de base ou de référence.
- Informations thématiques qui étudient un thème spécifique pour apporter plus de précision sur un phénomène décrit par l'information de base.

L'information géographique se décompose en trois niveaux :

- Le niveau géométrique décrit la forme et la localisation de l'objet.
- Le niveau sémantique, décrit les caractéristiques de l'objet géographique (Nom, Type, superficie).
 - Le niveau topologique qui décrit les relations avec d'autres objets.

2.3. Les systèmes d'Information géographique (SIG)

Le système d'information géographique est un outil informatique qui analyse, stocke, traite, manipule et visualise des informations géographiques et spatiales sur un phénomène donné. Les SIG permettent de mieux comprendre les données, telles que les modèles et les relations spatiales en aidant les individus et les organisations à prendre des décisions plus éclairées et facilitant la création de nouveaux services de géolocalisation dans plusieurs domaines comme l'organisation du territoire, la gestion des infrastructures et réseaux, le transport. Les SIG sont utilisés par tous les secteurs (administrations, entreprise, secteur public, etc.). Plusieurs organisations privée ou publique produisent de l'information géographique. En France l'IGN est le leader dans ce domaine.

3. Présentation de l'organisme d'accueil

3.1. Description de l'Institut national de l'information géographique et forestière (IGN)

3.1.1. Définition

L'IGN est un établissement public, producteur et fournisseur de référentiels de l'information géographique depuis juin 1940 et de l'information forestière depuis janvier 2012. Ses principales missions sont de décrire la surface du territoire national, montrer les différents espaces (urbain, forestier, naturel, agricole) et leurs occupations du sol et assurer la mise à jour permanente de l'inventaire des ressources forestières, et archive et diffuse toutes les représentations produites des données rassemblées.

L'IGN est placé sous la double tutelle du ministre de la transition écologique et solidaire et du ministre de l'agriculture et de l'alimentation, il intervient en appui à l'évaluation et à la mise en œuvre des politiques publiques de prévention des risques, d'aménagement du territoire, de développement durable, de défense et de sécurité.[1]

La diffusion de ces informations géographiques en forme de données cartographiques et en informations géolocalisées ce fait par le biais d'applications, produits, référentiels et géoservices développé par l'IGN.

3.1.2. Quelques chiffres sur IGN

- IGN compte 1580 salariés dont 15 contrats d'apprentissage
- Le budget annuel de l'IGN est de 153,34 M euros en 2018 avec 40,6 % de ressources propre
 - Le salaire annuel moyen sans les apprentis à l'IGN est de 48 689 euros
 - IGN a 3 laboratoires de recherches : LaSTIG, LAREG, LIF

3.1.3. Les activités de L'IGN

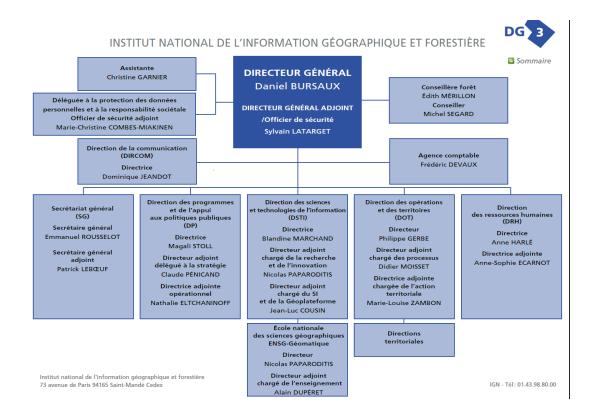
- L'IGN a pour mission de concevoir et de constituer une infrastructure géodésique cohérente avec les systèmes internationaux.

- L'IGN assure la couverture photographique aérienne régulière de l'ensemble du territoire national.
- L'IGN est chargé de l'inventaire permanent des ressources forestières nationales.
- Concevoir et maitre à jour les fonds cartographiques du territoire (généralistes ou thématiques)
- L'IGN décrit le territoire national et l'occupation de son sol de façon précise, complète et homogène.
- L'IGN conçoit des produits et services commerciaux [1]

3.2. L'organigramme de L'IGN

L'IGN est composé d'une déléguée à la protection des données personnelles et responsabilité sociétale, une conseillère forêt, une agence comptable et de 5 directions :

- Direction des programmes et l'appui aux politiques publiques ;
- Direction des sciences et technologies de l'information ;
- Direction des opérations et des territoires ;
- Direction des ressources humaines :
- Direction de la communication.



3.3. Les produits et services IGN

L'IGN développe plusieurs bases de données et services et produits afin de faciliter l'exploitation des données géographique.

3.3.1. Les services IGN

Géoportail: C'est une plateforme de données et de géoservices de référence, elle facilite l'accès à l'information géographique au grand public, elle propose des services aux particuliers comme aux organismes publics et aux professionnels. Son site et le suivant www.geoportail.gouv.fr

Les géocodeurs : géocodeur est une application de géoservice conçut par IGN, elle traduit une adresse littérale en coordonnées géographiques. Ces coordonnées géographiques permettent de situer chaque adresse sur une carte numérique via un Système d'Informations Géographiques (SIG). IGN propose deux applications,

Mes adresses qui est un service en ligne de localisation d'adresse en nombre son site est : https://mesadresses.ign.fr

Mon geocodeur est une version "expert" elle propose plus de choses que la version en ligne.

Ma carte: Une application de création de carte en ligne, elle offre deux possibilités, la création d'une nouvelle carte ou l'intégration de données dans une carte son site est : https://macarte.ign.fr/ [3]

Remonter le temps : Une application qui permet de remonter le temps et de consulter les données, les cartes anciennes et les comparer avec les actuelles, ce qui nous permet de voir l'évolution du territoire national dans le temps son site est : https://remonterletemps.ign.fr

Edugéo : C'est un service en ligne développé par l'IGN en partenariat avec le ministre de l'éducation national, il est destiné à apporter aux enseignants et les élèves un accès aux données géographiques sur l'ensemble du territoire Français avec des fonctionnalités comme : outil de croquis et de légende adapté à des usages en classe, données historiques sur 30 zones pédagogiques, logiciel SIG dédié...

3.3.2. Les bases de données IGN

L'IGN fournit à ses clients publics ou privés des bases de données géographiques de différents types :

Les cartes : Des images géoréférencées et immédiatement utilisables pour connaître le territoire, localiser des informations, consulter et mettre à jour des bases de données métiers.[3]

Les ortho images: L'IGN fournit des images aériennes du territoire national avec les ortho-images, elle utilise la BD ortho pour une image aérienne de 50 cm et la BD ortho HR pour des images de haute résolution aussi elle dispose de la BD ortho Historique pour identifier l'évolution du territoire.

Les bases vectorielles : les bases vectorielles de l'IGN se différencient par la précision des points enregistrés et par le modèle conceptuel de données, la donnée vectorielle permettent d'opérer une sélection d'objet à partir de différents critères. L'IGN propose une gamme complète de base vectorielle de la base BD TOPO pour une ville jusqu'à GeoFla pour le territoire national. [3]

Parcellaire cadastrale : les données du cadastre sont composées de données graphiques et de données sémantiques, ces données permettent d'identifier, de localiser et de représenter la propriété foncière. .[3]

Modèles numérique 3D : L'IGN propose une gamme de données en 3D autre que la BD TOPO qui est une base de données 3D. L'IGN a développé des bases de données intelligentes comme le référentiel national 3D.

PS: L'IGN propose des données en open source au grand public et d'autres données professionnelles payantes.

4. La réalisation

Mon travail pendant ce stage a été divisé en 5 volets :

- 1. Etude des collections des données
- 2. Conception et Modélisation
- 3. Implémentation d'une base de données orientée graphe
- 4. Transfert des données
- 5. Développement d'interfaces graphiques de l'application

4.1. Collection de données

Pour rassembler toutes les données nécessaires pour notre base de données, on a utilisé plusieurs sources comme IGN, APUR et INSEE.

APUR: Atelier Parisien d'urbanisme (APUR), il a été créé le 3 juillet 1967 par le Conseil de Paris. Ses missions sont de documenter, analyser et développer des stratégies prospectives concernant les évolutions urbaines et sociétales à Paris et dans la Métropole du Grand Paris.[5]

INSEE: L'Institut national de la statistique et des études économiques (Insee ou INSEE) est chargé de la production, de l'analyse et de la publication des statistiques officielles en France. IL a été créé le 27 avril 1946.[6]

Les différentes données collectées sont :

Les données sur les pistes cyclable : Pour avoir les données sur les pistes cyclables nous avons pris le thème trancon_route de la base de données bdtopo_metropole, ce thème comporte des données sur les portions de voie de communication destinée aux automobiles, aux piétons, aux cycles ou aux animaux.

Ces données seront filtrées juste dans le cas des pistes cyclables, puisque la base n'était pas encore rempliée en ce qui concerne les données sur les passages piétons.

Les données sur les espaces verts, centres sportifs et équipements cultures et loisirs : Nous avons utilisé les données en open data sur APUR catégorie équipements et services pour extraire ces données.

Les données sur le revenu : nous avons pu récupérer des données Fichier Localisé Social et Fiscal des communes en France métropolitaine, Martinique et La Réunion de l'année 2014 mis en ligne le 25/04/2019 par l'INSEE. Ce fichier contient plusieurs indicateurs relatifs aux

taux de pauvreté et au niveau de vie, pour notre part on a travaillé avec un seul indicateur qui est la médiane du niveau de vie dans une commune.

Les données des individus : Les données des individus sont confidentielles, donc nous avons mis nos données à nous et nous collègues du travail après leurs accords.

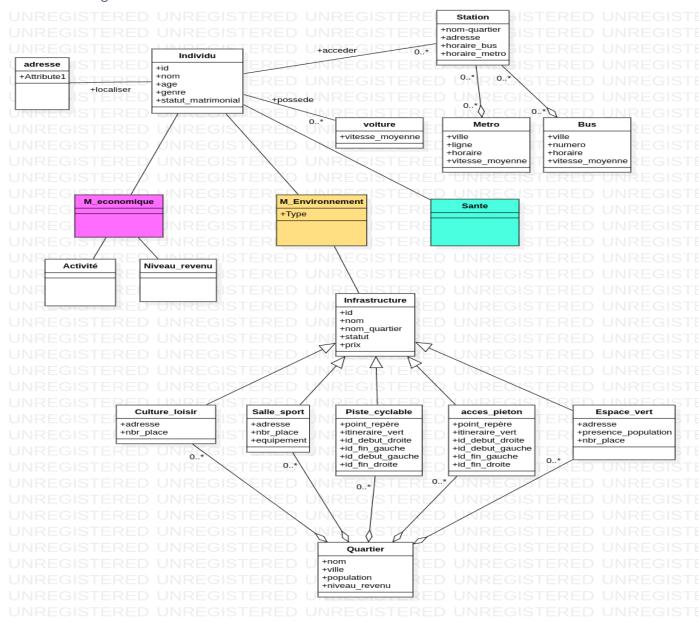
Les données sur le transport : Les données ne sont pas encore récoltées.

4.2. Modélisation par des diagrammes UML

La deuxième tâche consiste à réaliser des diagrammes de classes et de cas d'utilisations pour modéliser l'application. Pour la réalisation de ces diagrammes on a utilisé StarUML.

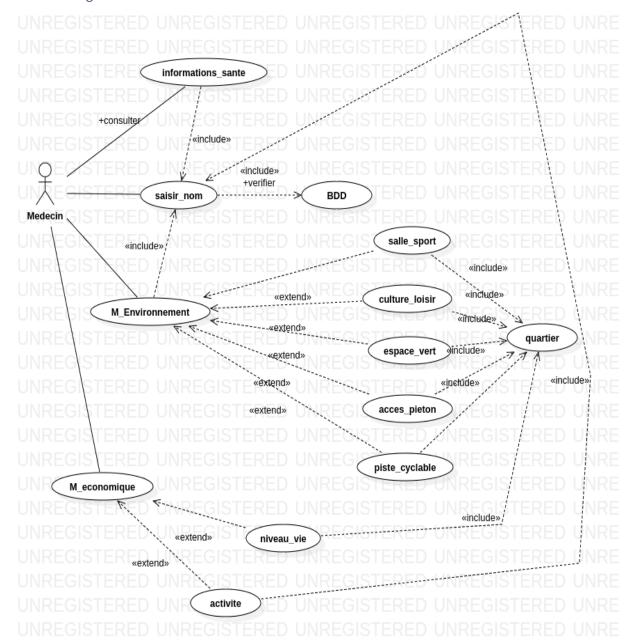
StarUML : StarUML est un logiciel de modélisation UML open source, simple d'utilisation, nécessitant peu de ressources système. Il fournit plusieurs types de diagrammes différents et accepte la notation UML 2.0. Il prend en charge l'approche MDA (Model Driven Architecture) en prenant en charge le concept de profil UML et en permettant de générer du code pour plusieurs langues. Ce logiciel constitue une excellente option pour une familiarisation à la modélisation.

4.2.1. Diagramme de classes



Le diagramme de classe est le plus important car il décrit les responsabilités, le comportement et le type de chaque objet. Il nous a permis de mieux visualiser notre problème et de ce fait mieux programmer en langage JAVA. Notre diagramme a été modifié lors de la programmation et il n'est pas définitif car nous essayons de l'améliorer. Quelques fonctions peuvent être supprimées ou déplacées. Nous avons consacré plus de temps à réaliser ce diagramme car c'est le plus complet.

4.2.2. Diagramme de cas d'utilisation



Notre diagramme de cas d'utilisation décrit les services les plus importants rendus par notre application. L'acteur principal qui va interagir avec l'application est le médecin en consultation d'un individu. Les fonctionnalités de notre application sont :

- Enregistrer un nouveau patient
- Sélectionner un patient pour avoir les informations sur les infrastructures qui l'entourent en nombre, cout, distance
 - Trouver le temps nécessaire pour se rendre à une infrastructure choisie

4.3. Développement d'une base de données Neo4j

4.3.1. Généralité sur les bases de données

Une base de données (DB), au sens le plus général, est une collection organisée de données. Plus précisément, une base de données est un système électronique qui permet d'accéder facilement aux données, de les manipuler et de les mettre à jour. Elle est gérée par un système de gestion de bases de données (SGBD) auquel ont accès simultanément tous les utilisateurs.

Le modèle relationnel des bases de données se trouve être limitant dès lors que les données sont fortement liées car une forte interconnexion des données entraine des temps de calculs assez longs. De plus, ses performances diminuent lorsque le volume de données à gérer devient trop important.

À l'heure du stockage massif de données et du big data, il devient donc fréquent de se tourner vers de nouvelles technologies de stockage, comme les bases de données Not Only SQL (NoSQL).

4.3.2. Les bases de données NoSQL

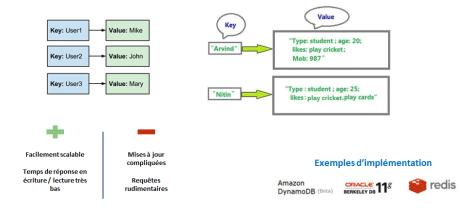
La base de données NoSQL désigne une base qui ne respecte pas toutes les règles d'un SGBD relationnel. Les systèmes basés sur NoSQL sont particulièrement utiles dans des bases de données très volumineuses. NoSQL, qui signifie "pas seulement SQL", constitue une alternative aux bases de données relationnelles traditionnelles dans lesquelles les données sont placées dans des tables et dont le schéma de données est soigneusement conçu avant la construction de la base de données.

4.3.3. Les différents types de base de données NoSQL

Il existe quatre types de systèmes de gestion de base de données NoSQL:

<u>Le modèle clé/ valeur</u>: sont les bases de données NoSQL les plus simples. Chaque élément de la base de données est stocké sous forme de nom d'attribut (ou "clé") avec une valeur qui peut être composée de plusieurs enregistrements.

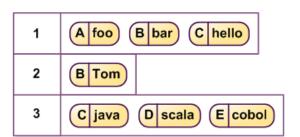
Exemple: Aerospike, Redis et Riak



<u>Le modèle colonnes</u>: Dans les bases de données NoSQL orientées colonnes, les données sont stockées par des colonnes de données plutôt que sous forme de lignes de données.

Exemple: Accumulo, Cassandra et HBase

	Α	В	С	D	E
1	foo	bar	hello		
2		Tom			
3			java	scala	cobol

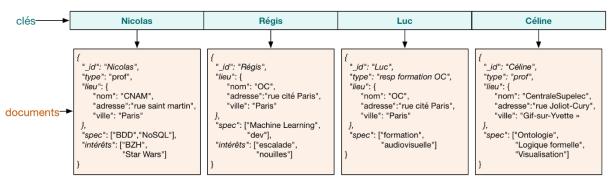


Organisation d'une table dans une BDD relationnelle

Organisation d'une table dans une BDD orientée colonnes

<u>Le modèle document</u>: Le model clés/ valeurs est similaire au model document sauf que dans ce dernier la valeur contient des données complexes appelées documents.

Exemple : MongoDB, CouchDB Server, MarkLogic sont des bases orientées document.



Dans notre projet nous allons utiliser le quatrième modèle qui est la base de données orientée graphe.

<u>Les bases de données orientées graphe</u>: Comme leur nom l'indique, les bases de données orientées graphe reposent sur la théorie des graphes et sur les représentations associées. Cela permet notamment d'effectuer des recherches de plus court chemin, de proximité, d'éloignement ou de sous-graphes (patterns) et ce à partir de n'importe quel point d'entrée dans le graphe.

Les différents composants d'une base de données graphe sont :

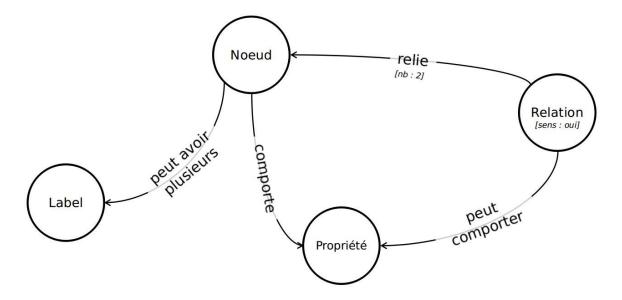
Les propriétés : c'est l'enregistrement d'une donnée sous la forme propriété : valeur>.
Bien que n'étant pas représentées physiquement dans le graphe, elles sont à l'origine du stockage de l'information.

Les nœuds : ce sont les principaux composants physiques du graphe. Dans la base de données, ce sont eux qui stockent majoritairement les données grâce aux propriétés qui leurs sont attachées.

Les relations : composants indispensables du graphe, elles relient les nœuds uni directionnellement selon des règles prédéfinies et contiennent ainsi intrinsèquement de

l'information. Elles permettent aussi de stocker des données grâce aux propriétés qui peuvent leur être attaché.

Les labels : chaque nœud du graphe peut se voir attribuer un ou plusieurs labels, qui permettent de regrouper sémantiquement des nœuds correspondant à un même type de réalité. Par exemple, un nœud peut comporter les labels Personne et/ou Etudiant. [7]



En comparaison aux autres modèles de bases de données, le stockage sous-forme de graphe est particulièrement intéressant lorsque les données sont fortement interconnectées car les outils mis à disposition par la théorie des graphes sont alors très efficaces.[7]

Les bases de données sont par exemple utilisées dans les réseaux sociaux, l'épidémiologie, la détection de fraude, les réseaux de neurones (IA) ou le e-commerce, entre autres. [8]

4.3.4. Implémentation d'une base de données orientée graphe Neo4j

Neo4j est une base de données orientée graphe développé en java par la société Neo Technology en 2000, la version 1.0 est sortie en 2010. Neo4j est leader dans le stockage de données dans le format graphe. [9]

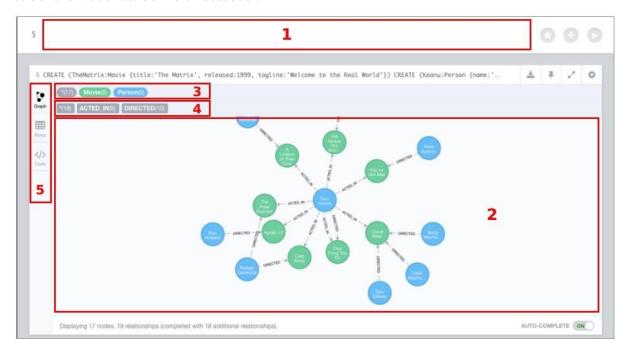
Neo4j respectent les principes ACID (atomicité, cohérence, isolation et durabilité) malgré qu'elle soit une base de données NoSQL.

Ses avantages:

- La possibilité d'exécuter plusieurs millions de requêtes par seconde sur des relations dans un nœud
 - Chaque transaction est fiable et unique au sein de cette base

Neo4j propose deux édition, édition l'édition Community, sous licence GPLv3, et l'édition commerciale, qui offre des options en plus comme la possibilité de faire des sauvegardes à

chaud, en plus de ça il y'a la version Desktop avec laquelle nous avons travaillé tout au long du projet de pair avec le Neo4J Browser qu'elle contient et qui permet la gestion des données et dont le visuel est donné ci-dessous :



PS : Nous pouvons accéder à l'interface d'administration de notre base de données par l'adresse :

http://localhost:7474

- 1- La zone de saisie des requêtes avec le langage Cypher
- 2- La zone d'affichage des résultats des requêtes
- 3- La liste des labels de nœuds : à chaque label corresponds une couleur et le nombre total des nœuds
 - 4- La liste des labels des relations et leur nombre total
 - 5- Le mode d'affichage : Graphe, Table, Texte, Code

A chaque lancement d'une requête, une nouvelle zone d'affichage s'ouvre ce qui nous permet de voir les actions antérieures avec leurs requêtes au-dessus.

Neo4j intègre son propre langage qui est Cypher.

Le langage Cypher:

Cypher est un langage informatique de requêtes déclaratif spécifique à Neo4j qui permet une interrogation et une mise à jour d'un graphe de propriétés. Cyphe est un langage simple mais puissant, il facilite l'expression des requêtes de base de données très compliquées, cependant il reste très proche du SQL dans son utilisation, notamment parce qu'il est lui aussi basé sur les opérations CRUD (Create, Read, Update, Delete). [10]

OPÉRATION	SQL	CYPHER	
Création	CREATE	CREATE	
Recherche	SELECT	MATCH RETURN	
Nature des données	FROM	(: Label)	
Jointure	JOIN [table] ON	()-[]-()	
Jointure	JOIN (SELECT) ON	UNION	
Condition	WHERE	WHERE	
Ajout	INSERT	CREATE / MERGE	
Mise à jour	UPDATE [table] SET	SET / REMOVE	
Retrait	DELETE	DELETE	

Tableau de correspondance des principales commandes entre SQL et Cypher

Nous remarquons que les commandes sont très similaires entre les deux langages SQL et Cypher. Ce qui est différent c'est la commande de la jointure, dans le langage Cypher nous représentons le nœud entre parenthèse et la relation entre crochet (...)-[...]-(...), voici la structure complète entre deux nœuds n et m avec la relation r :

```
(n: Label<sub>n</sub> {prop<sub>n,1</sub>: ... , prop<sub>n,2</sub>: ...})-[r: Nature<sub>r</sub> {prop<sub>r,1</sub>: ...}]-(m: Label<sub>m</sub> {prop<sub>m,1</sub>: ... , prop<sub>m,2</sub>: ...})
```

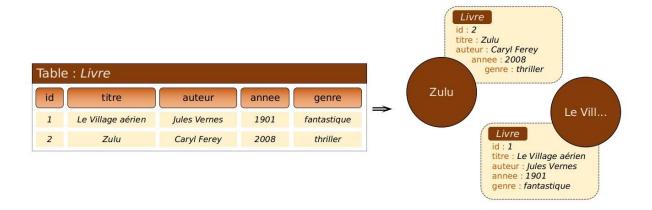
- Chaque nœud a un label et peut avoir plusieurs propriétés.
- Chaque relation a sa nature et peut avoir plusieurs propriétés

4.4. Transfert de données

Tous les fichiers de la base de données que nous avons collectée nous les avons mis en format CSV, puisque le langage Cypher contient une fonction dédiée au chargement de ce genre de fichier.

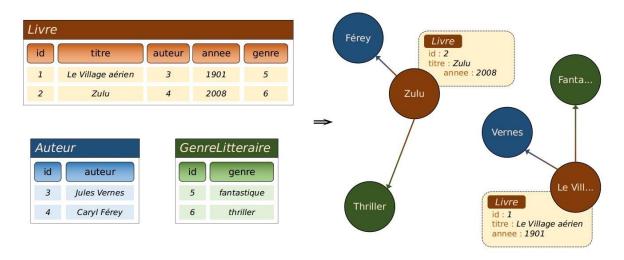
Donc le passage de la base de données relationnelle (tableaux) vers la base de données graphe neo4j se fait selon certaines règles :

- 1- Une ligne de table devienne un nœud du graphe.
- 2- Le nom d'une table devient le label attribué aux nœuds issus de cette table
- 3- Les attributs d'une table deviennent les propriétés associées aux nœuds issus de cette table



Exemple de génération de noeuds à partir d'une table relationnelle

- 4- Une clé étrangère devient une relation entre les nœuds issus de chaque table
- 5- Une table d'association devient une relation entre les nœuds issus de chaque table



Exemple de génération de graphe à partir d'une table relationnelle

Les étapes d'importation des fichiers CSV vers la base Neo4j :

- 1- Crée une base de données vide
- 2- Transformer tous les fichiers en CSV
- 3- Mettre ces fichiers dans le répertoire **importer** de l'arborescence de Neo4J, créé à l'installation
- 4- Rédiger les requêtes qui permettent la récupération et le chargement de ces fichiers dans la base neo4j

Revenons à la dernière étape celle qui permet le chargement des fichiers CSV et leurs mises en forme dans la base de données Neo4j:

LOAD CSV [WITH HEADER] FROM 'file:/nom_du_fichier' [FIELDTERMINATOR ';'] Chargement du fichier d'une ligne d'entête (si besoin) Définition explicite du délimiteur (optionnel)

Structure de la requête Cypher de chargement d'un fichier CSV

Ce que nous pouvons remarquer :

- 1- Il y'a deux façons de charger le fichier CSV soit avec une ligne d'entête là nous ajoutons « **WITH HEADER** » ou sans entête donc nous n'ajoutons pas ce bout de code.
- 2- Nous pouvons aussi définir le délimiteur souhaité en ajoutant « FIELDTERMINATOR ' ;']

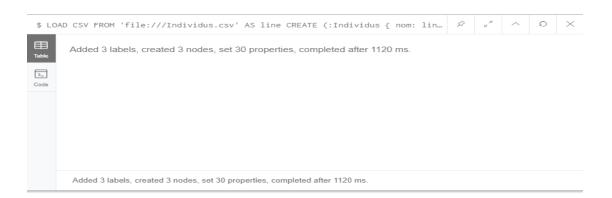
Voilà un exemple d'importation d'un fichier sur notre base de données :

```
LOAD CSV FROM 'file:///Individus.csv' AS line

CREATE (:Individus { nom: line[1], prenom: line[2],age: toInteger(line[3]),genre: line[4],numero_Rue :toInteger(line[5]),nom_Rue : line[6],code_Postal: toInteger(line[7]), statut : line[8], statut_matrimonial: line[9],code_insee : toInteger(line[10]) })
```

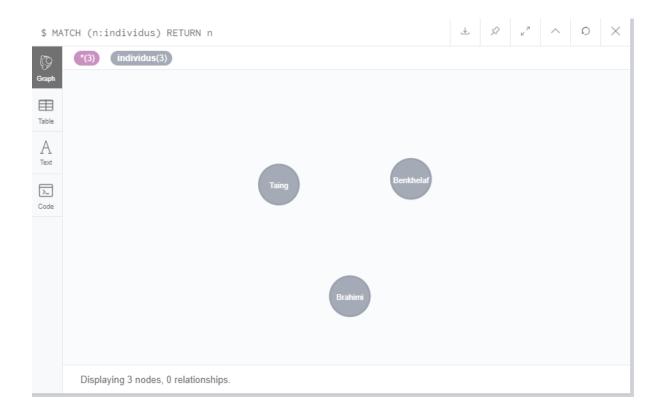
- 1. Cette commande permet d'importer le fichier Individus sur notre base Neo4j et de créé des nœuds pour chaque individu avec ses propres propriétés.
- 2. Si on veut importer un fichier à partir d'une adresse internet nous remplaçons 'file' par 'https'.
- 3. La fonction toInteger permet de récupérer les informations en nombre entier.

Le résultat de cette commande est le suivant :



Le résultat de l'exécution de cette commande est l'ajout de 3 nœuds et 30 propriétés, le temps d'exécution est de 1120 ms.

Pour afficher les nœuds ajoutés nous exécutons la commande : Match (n:Individus) return n



Les résultats peuvent être affichés avec plusieurs formes, en graphe comme la capture d'écran au-dessus ou dans un tableau ou format texte ou en format code.

Nous récupérons tous les fichiers avec la même méthode et nous les mitons dans notre base de données.

Pour créer les relations entre les nœuds on exécute cette commande :

```
A 1 MATCH (a:Equipement_sport),(b:individus)

2 WHERE a.code_postal = b.code_Postal

3 CREATE (b)-[r:Nos_equipements_sport]→(a)

4 RETURN r
```

Cette commande permet de créer une relation entre les individus et les équipements de sports qui se trouvent au sein du quartier de l'individu. Son nom est 'Nos équipements sport'.

Le résultat de cette commande est le suivant :



On a ajouté 49 relations de type Nos_equipements_sport entre les individus et les équipements sportifs. Ces relations sont beaucoup plus visibles sur la zone d'affichage.

Apres l'ajout de tous les nœuds et les relations nécessaires pour notre application nous sommes passés à la dernière étape qui est la programmation d'une interface graphique en java.

4.5. Implémentation de l'interface graphique de l'application en Java

Cette dernière étape consiste à réaliser une interface graphique en java en utilisant les données récupérées sur notre base données Neo4j. Nous avons pour but de développer plusieurs fonctionnalités pour notre application :

- 1- L'ajout d'une personne à notre base de données
- 2- La récupération des informations concernant le patient et les afficher.
- 3- Montrer le temps et la distance qui sépare un individu de ces infrastructures

Pour la réalisation de cette application en java nous avon utilisé Eclipse avec Maven

Eclipse : C'est un environnement de développement (IDE) utilisé par le langage java, mais avec le système de plugins d'autres langages peuvent l'utiliser comme C/C++ et PHP. Eclipse nécessite une machine virtuelle Java (JRE) pour fonctionner. Mais pour compiler du code Java, un kit de développement (JDK) est indispensable [12]

Maven : Maven est un outil de gestion et de construction d'un projet, il est utilisé pour automatiser et faciliter certaines tâches. Il peut aussi générer la documentation et gérer les dépendances.

Il existe deux manières d'utiliser Neo4j à partir de java :

- 1- Utiliser un serveur autonome, celui-là peut être installer dans n'importe quel ordinateur, puis accessible via son API http.
 - 2- Utiliser un Neo4j intégré dans le processus JVM

Pour notre part nous avons choisi de travailler avec un serveur autonome, comme nous avions travaillé déjà avec neo4j desktop pour l'importation des données et la réalisation des relations entre les nœuds donc nous avons préféré cette première option puisqu'elle nous permet aussi de voir le résultat sur le serveur directement.

Apres l'installation de Maven sur java, nous avons créé un nouveau projet 'Maven project'. Dans la racine de ce projet il y'a un fichier **pom.xml** c'est le fichier de description et configuration du projet Maven, nous ajouton dans ce fichier les dépendances.

```
<modelVersion>4.0.0</modelVersion>
   <groupId>ign
   <artifactId>neo4j0</artifactId>
5
   <version>0.0.1-SNAPSHOT</version>
7⊖ <dependencies>
8⊖ <dependency>
9
   <groupId>org.neo4j.driver</groupId>
10
   <artifactId>neo4j-java-driver</artifactId>
   <version>1.7.2
11
12
   </dependency>
   </dependencies>
13
14 </project>
```

Une fois c'est fait, on peut commencer la programmation

Du point de vue graphique pour que nos interfaces s'affichent nous avons utilisés les méthodes implémentées dans les différentes classes selon ce que doit afficher la fenêtre mais aussi en redirigeant les boutons choisis vers de nouvelles fenêtres.

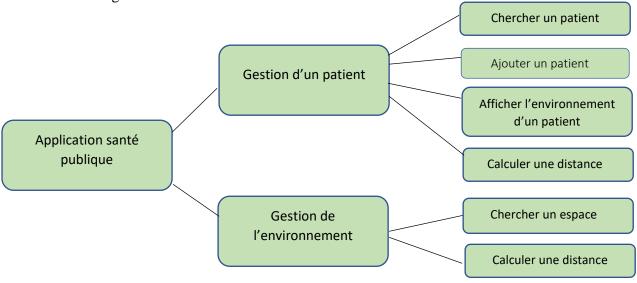


Figure 2: Schéma de l'application santé publique

Ce schéma montre les différentes parties qu'nous allons programmer pour notre application. Nous avons déjà réalisé les deux parties : chercher un patient et ajouter un patient et il nous reste les autres parties à finir dans les 3 semaines qu'ils nous restent

```
4.5.1. Partie 1:
```

Cette partie consiste à programmer une interface qui va demander à une personne de s'identifier en entrant son nom, prénom, et sa date de naissance puis elle vérifiera s'il existe dans notre base si non elle l'ajoutera avec les informations demandées.

Etablir la connexion:

Pour établir la connexion entre notre programme java et le système de gestion de base de données Neo4j on a utilisé Neo4j java driver, après nous établions la connexion avec ce bout de code :

driver = GraphDatabase.driver(uri, AuthTokens.basic(user, password));

Uri: "bolt://localhost:7687"

Le user et le mot de passe sont les même identifiants que le user et mot de passe de notre base de données

4.5.2. Les autres parties

Les autres parties sont en cours de réalisation pour les 3 semaines qui nous reste.

5. Conclusion et perspectives

Notre objectif pour ce stage est de réaliser une application qui pourra aider le personnel médical à positionner un patient dans son environnement. Cet environnement consiste à une localisation géographique vis-à-vis des espaces verts, espaces bleu, installations sportives et bases de loisirs. Toutes les informations de cet environnement se retrouvent dans les données de l'IGN, APUR et l'INSEE.

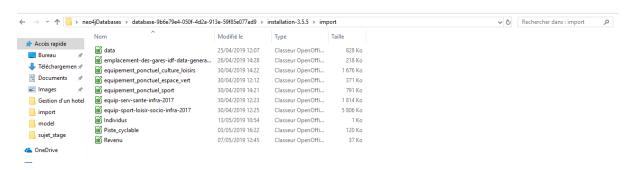
Au cours de ce stage j'ai pu acquérir de nouvelles connaissances : premièrement, j'ai découvert le monde de l'information géographique qui était totalement inconnu pour moi grâce à la documentation sur les activités de l'IGN et aussi à l'exploration de ses bases de données. Deuxièmement j'ai pu aussi apprendre à manipuler un logiciel de modélisation à savoir StarUML. Troisièmement j'ai acquis des connaissances en base de données Not Only SQL et j'ai pratiqué la base de données orienté graphe Neo4j et un langage spécifique Cypher. Quatrièmement j'ai mis en pratique mes connaissances en programmation java acquise en L3.

Nous avons rencontré quelques difficultés notamment la manipulation de Neo4j et son intégration sur java.

Pour la suite de notre travail, l'accent devra être mis sur la programmation de notre interface graphique durant les 3 semaines qui nous reste.

Les annexes

Les données récoltées :



Deux exemples de base de données récoltées :



	А	В	С	D
1	75056		26194.7	<i>D</i>
2		Paris 1er Arrondissement	31574.5	
3		Paris 2e Arrondissement	29296	
4		Paris 3e Arrondissement	30560.7	
5		Paris 4e Arrondissement	30385.5	
6		Paris 5e Arrondissement	32717.5	
7		Paris 6e Arrondissement	38213.3	
8		Paris 7e Arrondissement	41501.3	
9		Paris 8e Arrondissement	39906	
10		Paris 9e Arrondissement	32319	
11		Paris 10e Arrondissement	24684.2	
12		Paris 11e Arrondissement	25972.7	
13		Paris 12e Arrondissement	26700	
14		Paris 13e Arrondissement	23397	
15		Paris 14e Arrondissement	27047.3	
16		Paris 15e Arrondissement	30246.7	
17		Paris 16e Arrondissement	38506.8	
18		Paris 17e Arrondissement	29715.6	
19		Paris 18e Arrondissement	20526	
20		Paris 19e Arrondissement	18968.3	
21		Paris 20e Arrondissement	20345.3	
22		Yvetot	19026	
23		Yville-sur-Seine	20956.7	
24		Achères-la-Forêt	27680.7	
25		Amillis	22215.4	
26		Ampony ille	24094.4	
27		Andrezel	27976	
28		Annet-sur-Marne	26751.4	
29		Arbonne-la-Forêt	26825.4	
30		Argentières	25116.7	
31		Armentières-en-Brie	19321	
32		Arville	26644.3	
33		Aubepierre-Ozouer-le-Repos	24144	
34		Aufferville	22264	
35		Augers-en-Brie	21514.4	
36		Aulnoy	24129.6	
37	77014		21545	
38	77015			
39		Bagneaux-sur-Loing	18258	
40		Bailly-Romainvilliers	24063.5	
41		Balloy	23314	
42		Bannost-Villegagnon	24548.5	
43		Barbey	21319	
44		Barbizon	30270.5	
45		Rarou	22027	

Les Références :

http://www.ign.fr/institut/institut/statut-et-missions [1]

http://www.ign.fr/publications-de-l-ign/Institut/Publications/RA/RA 2018/ign ra 2018.pdf [2]

http://www.ign.fr/publications-de-l-ign/Institut/Publications/Autres_publications/CataloguePro-2016/CatalogueOffre2016.pdf [3]

http://www.ign.fr/institut/kiosque/presentation-lign [4]

https://www.apur.org/fr/atelier [5]

https://fr.wikipedia.org/wiki/Institut national de la statistique et des %C3%A9tudes %C3%A9co nomiques [6]

https://neo4j.com/product/ [7]

Neo Technology. GraphDay in Paris, November 2018 [8]

https://neo4j.com/company [9]

https://fr.wikipedia.org/wiki/Cypher_(langage) [10]

https://neo4j.com/developer/desktop-csv-import/ [11]

https://www.commentcamarche.net/download/telecharger-34075627-eclipse-ide-pour-developpeurs-java [12]

https://neo4j.com/developer/java/ [13]

https://fr.wikipedia.org/wiki/StarUML [14]

https://fr.calameo.com/read/0011885828344255b2b15 [15]

https://fr.wikipedia.org/wiki/Information g%C3%A9ographique [16]