

Faculté des Sciences Exactes et de l'Informatique
Département de Mathématiques et d'Informatique
Filière : Informatique

MEMOIRE DE FIN D'ETUDES
Pour l'Obtention du Diplôme de Master en Informatique
Option : **Systemes d'Information Géographique**

THEME :

Vers une nouvelle approche SOLAP pour l'analyse des données complexes

Etudiant(e)s : « **MILOUD DAOUADJI ASMA** »

« **REZKALLAH SARA** »

✓ Encadrant(e) : « **Dr : YOUNSI FATIMA ZAHRAA** »

Année Universitaire 2016/2017

Résumé

L'étude de structure topologique dans les réseaux sociaux devient de plus en plus une question importante. La connaissance des modules de base des réseaux nous aide à bien comprendre leurs fonctionnements et comportements, et à appréhender les performances de ces systèmes.

Dans le contexte des réseaux de contact, les données peuvent être considérées comme des réseaux sociaux hétérogènes. La construction et l'analyse de tels réseaux répartis dans l'espace d'une façon hétérogène posent de nombreux problèmes. Par conséquent, l'analyse SOLAP peut apporter une réponse pour synthétiser ou agréger ce type de réseau, car l'agrégation et la visualisation sont deux points centraux pour ce type d'analyse.

L'objectif principal de ce projet de fin d'étude est de proposer un nouveau MOdèle Décisionnelle basé sur une le concept Solap et le Graph olap nommé MODS-GOLAP. Ce dernier permettra de construire, stocker et analyser plusieurs réseaux sociaux tout en combinant SOLAP et l'analyse des réseaux sociaux. Nous nous intéressons, tout d'abord, par la création d'un entrepôt de données spatiotemporelle qui permet de stocker les différents types de données (de type graphe, spatiales, thématiques et temporelles). Ensuite, ces données subissent une analyse multidimensionnelle pour des fins d'analyse et d'aide à la décision. Enfin, nous visualisons les données analysées par l'utilisation des outils de visualisation.

Mots-clés : MOdèle Décisionnelle basé sur une approche Solap et le Graph olap (MODS-GOLAP), Analyse multidimensionnelle, Spatial OLAP, Graph OLAP, Réseaux sociaux, Système d'Information Décisionnels (SID), Théories des graphes.

SOMMAIRE

Résumé

Liste des tableaux.....	i
Liste des figures.....	i

INTRODUCTION GENERALE.....	01
----------------------------	----

CHAPITRE I : RESEAUX COMPLEXES

1. Introduction.....	03
2. Définitions et notations.....	03
2.1 Qu'est ce qu'un réseau.....	03
2.2 Les graphes.....	03
2.2.1 Graphe non orienté.....	04
2.2.2 Graphe orienté.....	04
2.2.3 Sous-graphe.....	05
2.3 Les graphes statique et dynamique.....	05
2.3.1 Graphe statique.....	05
2.3.2 Graphe dynamique.....	05
3. Structures et modèles de réseaux.....	06
3.1 Les réseaux aléatoires.....	06
3.2 Les réseaux réguliers.....	07
3.3 Le modèle du Small-world (petit monde).....	07
3.3 Le modèle du Scall-free (Invariants d'échelle).....	08
4. L'analyse des réseaux sociaux.....	09
4.1 Les mesures des réseaux.....	09
5. Conclusion.....	10

CHAPITRE II : LES SYSTEMES D'INFORMATION DECISIONNELS (SID)

1. Introduction.....	11
2. Système d'information décisionnel (SID)	11
2.1 Fonction essentielle d'un SID	11
2.2 Architecture générale d'un SID.....	11
3. Les entrepôts de données.....	13
3.1 L'alimentation d'ED.....	13
4. Modélisation multidimensionnelle des données.....	14
4.1 Modélisation conceptuelle de données.....	14

5. Schéma multidimensionnel, le cube OLAP (On-Line Analytical Processing).....	16
5.1 Les architectures OLAP.....	14
5.1.1 MOLAP : Multidimensional OnLine Analytical Processing.....	17
5.1.2 ROLAP : Relational Online Analytical Processing.....	17
5.1.3 HOLAP: Hybrid Online Analytical Processing.....	17
6. Le Spatial OLAP(SOLAP)	18
6.1 Entrepôt de données Spatiales(EDS).....	18
6.2 Le concept SOLAP.....	18
6.2.1 Les dimensions spatiales.....	18
6.2.2 Les outils SOLAP.....	20
7. Les graphes et les cubes de données.....	21
7.1 Les graphes OLAP.....	21
7.1.1 Définition d'un Graph OLAP.....	21
7.1.2 L'utilisation de Graph OLAP.....	21
7.2 Les graphes SOLAP.....	23
8. Conclusion.....	23

CHAPITRE III : Conception et Modélisation

1. Introduction.....	24
2. Objectifs visés.....	24
3. Description du modèle proposé	24
3.1 La structuration.....	25
3.2 L'exécution.....	25
3.3 L'exploitation.....	26
4. Modélisation Multidimensionnelle de l'ED.....	27
4.1 Identification de fait.....	27
4.2 Définition des axes d'analyse.....	27
5. Démarche adoptée par le modèle MODS-GOLAP.....	29
6. Conclusion.....	29

CHAPITRE VI : IMPLEMENTATION ET MISE EN ŒUVRE

1. Introduction.....	30
2. Les outils de développements utilisés.....	30
2.1 Microsoft SQL server	30
2.2 Java netbeans.....	30

2.3 Microsoft visuel studio.....	30
3. Présentation du système MODS-GOLAP	31
3.1 Création de la base de données.....	32
3.2 Création de cube de données.....	34
4. Connexion au serveur OLAP et Configuration du service IIS + XMLA.....	36
5. Présentation de notre application	37
5.1 L' interface d'accueil	37
5.2 La mise à jour des données	37
5.3 Choix des dimensions et de mesures à analysés	38
6. La phase d'exploitation	39
6.1 Les requêtes MDX	39
6.2 Visualisation des résultats MDX et d'histogramme	40
7. Conclusion.....	41
CONCLUSION GENERALE ET PERSPECTIVES.....	42
Bibliographie.....	44

Liste des Tableaux

Table 1.1. Présentation des différentes structures de réseaux.....	08
Table 1.2. Les propriétés des réseaux.....	09

Liste des Figures

Figure 1.1. Exemple de graphe simple, non orienté et matrice d'adjacence associée.....	04
Figure 1.2. Exemple de graphe simple, orienté et matrice d'adjacence associée.....	04
Figure 1.3. Un sous-graphe du graphe G.....	05
Figure 1.4. Le modèle du réseau régulier.....	07
Figure 1.5. Le modèle petit monde (Small World) proposé par Watts Strogatz.....	08
Figure 2.1. Les principaux composants d'un SID.....	12
Figure 2.2. Une présentation schématique du processus ETL.....	14
Figure 2.3. Schéma en étoile	15
Figure 2.4 Schéma en flocon en neige.	16
Figure 2.5 a) Modèle logique en étoile générique b) Modèle logique en étoile pour l'analyse de la pollution de l'air.....	20
Figure 2.6. Exemple d'un graph OLAP.....	22
Figure 3.1. Le modèle MODS-GOLAP proposé	24
Figure 3.2. La présentation de la phase de structuration	25
Figure 3.3. La présentation de la phase de structuration	26
Figure 3.4. Représente une vue d'ensemble de l'étape d'exploitation.....	27

Figure 3.5. Schéma du modèle multidimensionnel	28
Figure 3.6. La démarche adoptée par MODS-GOLAP.....	29
Figure 4.1. Organigramme de travail	31
Figure 4.2. Fenêtre de SQL Server management studio.....	32
Figure 4.3. Création de la base de données.....	32
Figure 4.4. Fenêtre de création de la base de données.....	33
Figure 4.5. Création de la table.....	33
Figure 4.6. Création des colonnes pour la table commune.....	34
Figure 4.7. Exemple d'une table.....	34
Figure 4.8. Le schéma du cube de données du fait réseau social.....	35
Figure 4.9. Le déploiement du cube.....	36
Figure 4.10. Présentation la fenêtre principale.....	37
Figure 4.11. Fonction de la mise à jour de la table commune.....	38
Figure 4.12. Choix des dimensions et des mesures.....	39
Figure 4.13. Résultat de la requête.....	40
Figure 4.14. Visualisation des résultats MDX.....	40
Figure 4.15. Représentation graphique deux nombre de relations amicales par rapport aux régions.....	41

INTRODUCTION GENERALE

Contexte et problématique

Dans le contexte de l'informatique décisionnelle dont le but est de collecter, organiser, stocker et analyser l'information pour aider la prise de décision le concept d'entrepôt de données a été introduit. Un entrepôt de données est une collection de données, orientées sujet, intégrées, non volatiles et historisées, organisées comme support d'un processus d'aide à la décision. Les données sont extraites, nettoyées, transformées en un format unique qui les prépare à une analyse en ligne.

En effet, l'analyse en ligne OLAP (Online Analytical Processing) est l'une des technologies les plus importantes dans les entrepôts de données, elle permet l'analyse multidimensionnelle de données. Cela correspond à un outil d'analyse puissant, tout en étant flexible en terme d'utilisation pour naviguer dans les données, plus ou moins en profondeur. OLAP a été le sujet de différentes améliorations et extensions, avec sans cesse de nouveaux problèmes en lien avec le domaine et les données, par exemple le multimédia, les données spatiales, les données séquentielles, etc.

Contribution

Dans ce travail, nous nous intéressons à l'OLAP Spatial (Spatial OLAP ou SOLAP) qui sont des systèmes d'aide à la décision intégrant les fonctionnalités des (Systèmes d'Information Géographiques) SIG, des entrepôts de données et des systèmes OLAP. Le SOLAP étend le modèle multidimensionnel en définissant les concepts de dimension spatiale et mesure spatiale.

Dans le contexte des relations sociales, les données peuvent être considérées comme des réseaux hétérogènes de contacts. La construction et l'analyse de tels réseaux posent de nombreux problèmes. Le concept Graph OLAP a amélioré l'analyse de ces réseaux. Cependant, la construction et l'analyse de tels réseaux *répartis dans l'espace* d'une façon hétérogène posent de nombreux problèmes. A la question de comment synthétiser ou agréger un réseau, l'analyse SOLAP des réseaux de contact peut apporter une réponse car l'agrégation et la visualisation sont deux points centraux de ce type d'analyse.

L'objectif de notre travail, consiste à élaborer un MODèle Décisionnel basé sur une approche Solap et le Graph olap nommé MODS-GOLAP. Son objectif principale est d'intégrer et analyser les données thématiques, spatiales, temporelles et de type graphe.

Le modèle MODS-GOLAP est composé de deux sous modèles, à savoir :

1. *Le sous-modèle d'entrepôt de données* : permettra de créer et de stocker les données multiples d'une façon permanente dans l'entrepôt,
2. *Le sous-modèle d'analyse* : permettra d'effectuer des analyses spatiales et multidimensionnelles des données stockées.

Organisation du mémoire

Hormis le chapitre introductif et le chapitre consacré à la conclusion et perspectives, ce mémoire comporte quatre chapitres:

Chapitre 1: est dédié à un état de l'art qui énonce des rappels sur la théorie des graphes et le vocabulaire utilisés dans le travail, et dans un second temps, passe en revue les travaux connexes.

Chapitre 2: est consacré à l'étude des Système d'Information Décisionnels (SID). Nous discuterons les principes d'un SID ainsi que son architecture et nous exposerons les travaux liés à notre étude.

Chapitre 3 : dans ce chapitre, nous décrivons, en détail, notre contribution en expliquant notre démarche décisionnelle ainsi que la structure du modèle MODS-GOLAP proposé, ses composants et son fonctionnement.

Chapitre 4 : ce chapitre détaille l'aspect technique de notre contribution et présente une étude de cas qui constitue en soi une étape de validation de notre proposition.

CHAPITRE I : RESEAUX COMPLEXES

1. Introduction

Le terme de réseaux fait d'abord partie du langage courant, il évoque de nombreuses structures concrètes et familière : des réseaux technologique (voies ferrées et routières, lignes aérienne et électrique internet), des réseaux sociaux, des réseaux biologique (réactions métabolique, système nerveux, circulation sanguine, chaine alimentaires) ou la classe des réseaux en arbre (réseaux hydrographiques) dont nous verrons les propriétés particulières [Fas 03].

2. Définitions et notations

Dans cette section, nous définissons un ensemble de termes utilisés dans la littérature qui se rapportent au domaine de recherche des réseaux sociaux.

2.1. Qu'est ce qu'un réseau?

Un **réseau est un graphe** dont les nœuds et les liens présentent des caractéristiques quantitatives ou qualitatives. De manière générale, un **réseau** désigne un ensemble **d'entités** (hommes, animaux, machines, cellules, etc.), que l'on nomme les nœuds du réseau, reliées entre elles par un ensemble de connexions appelées **liaisons** ou **liens**.

Un réseau est traditionnellement défini par un graphe $G = (V, E)$ dans lequel V est l'ensemble des nœuds et E l'ensemble des liaisons du réseau. E est un ensemble de couples de nœuds tel que $E \subseteq V \times V$.

Plusieurs exemples des réseaux complexes existent: réseau neuronal, réseau sociaux, réseau du transport aérien, réseau biologique, etc.

2.2. Les graphes

Les graphes représentent de manière simple et naturelle des relations entre les objets. Les outils mathématiques et les algorithmes mises au point en théorie des graphes permettent de résoudre une multitude de problèmes, tels que les problèmes de cheminement, d'ordonnancement, d'adéquation, etc[Rob03].

2.2.1. Graphe non orienté

Pour illustrer cette définition, imaginons un groupe de sept individus dont on cherche à décrire les interactions et les possibles échanges d'information. Le graphe simple et la matrice d'adjacence qui lui est associée sont illustrés par la Figure.1.1 Le graphe pourrait représenter l'existence de discussions entre les membres du groupe sur une période donnée, ce qui se prête bien à un modèle de graphe non-orienté puisque l'échange d'information peut être réciproque.

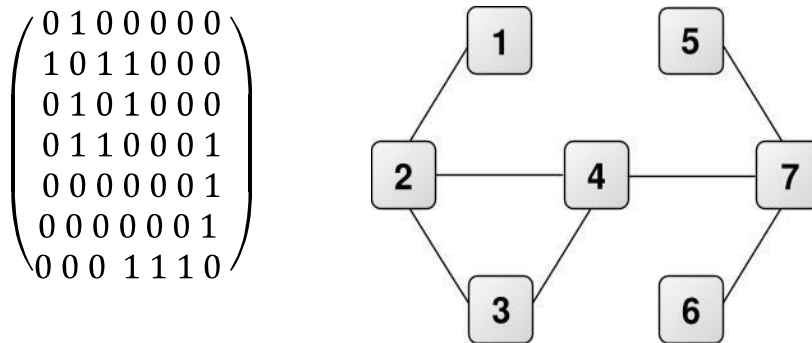


Figure 1.1. Exemple de graphe simple, non orienté et matrice d'adjacence associée.

2.2.2. Graphe orienté

Un réseau **orienté** est un réseau dont l'ensemble des liens E regroupe des couples de nœuds (donc ordonnés), appelés *liens orientés*. Dans un réseau orienté, la présence d'un lien $e_1 = (v_i, v_j)$ entre les nœuds v_i et v_j n'implique pas nécessairement l'existence d'un lien $e_2 = (v_j, v_i)$.

Dans les représentations graphiques de tels réseaux, l'orientation du lien est généralement représentée par une flèche indiquant la direction du lien (voir Figure 1.2).

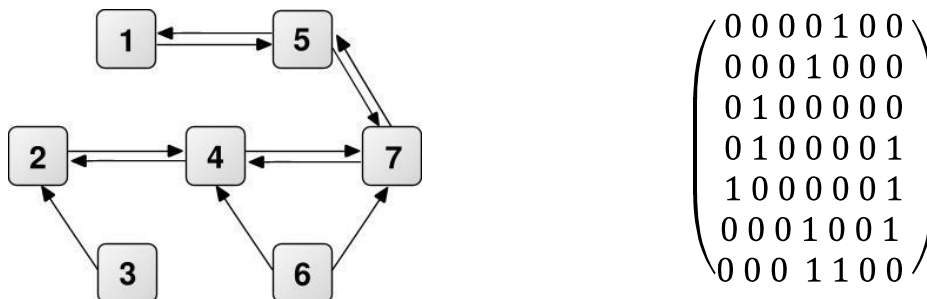


Figure 1.2. Exemple de graphe simple, orienté et matrice d'adjacence associée.

2.2.3. Sous-graphe

Un sous-graphe de G consiste à considérer seulement une partie des sommets de V et les liens induits par E . Par exemple si G représente les liaisons aériennes journalières entre les principales villes du monde, un sous-graphe possible est de se restreindre aux liaisons journalières entre les principales villes européennes.

La Figure 1.3 représente un sous-graphe.

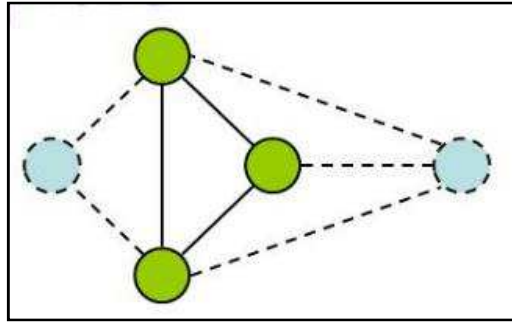


Figure 1.3. Un sous-graphe du graphe G

2.3. Les graphes statique et dynamique

Dans la littérature, il existe deux types de graphe non orienté :

2.3.1. Graphe statique

L'approche statique considère le réseau social comme un tout et ne s'intéresse pas aux processus qui ont amenés à sa création. Elle considère que l'ordre dans lequel les liens apparaissent ne couvre pas d'importance dans la compréhension du réseau. C'est cette approche qui a donné la plus grande profusion d'algorithmes. L'approche statique s'intéresse surtout à l'état actuel d'un réseau et aux structures communautaires qui tiennent actuellement place.

2.3.2. Graphe dynamique

Un graphe dynamique est une succession de graphes statiques. Nous appellerons ces graphes statiques des instantanés car ils représentent l'image du réseau dynamique à un instant donné. Le temps sera divisé en pas de temps ou instants. Un pas de temps sera l'index de l'instantané. Si les instantanés ne sont pas répartis régulièrement on peut associer au pas de temps une durée. De plus, sauf mention contraire, nous noterons l'instant suivant t comme l'instant $t + 1$ quelle que soit la durée qui les sépare.

3. Structures et modèles de réseaux:

La théorie des réseaux est présente en mathématiques depuis Euler 1736, elle a été utilisée pour résoudre le problème des 7 ponts de Königsberg. Pour résoudre ce problème, Euler schématisa la situation sous forme d'un graphe dans lequel les terres accessibles étaient représentées par un nœud, et les ponts par des connexions entre ces nœuds.

Dans le contexte d'épidémiologie, l'un des premiers modèles est basé sur un formalisme issu de la théorie des réseaux ont été proposé par Newman en 2002. Ce modèle s'est depuis imposé comme la référence dans le domaine et est le fondement de la plupart des nouveaux développements [You 16].

Nous commençons cette section par l'étude des classes de réseau complexe les plus célèbres afin de pouvoir identifier les aspects théoriques et méthodologiques de chacun. En effet, la plupart de ces réseaux peuvent être généralisés, nous commençons par le type de graphes le plus exploité : les graphes aléatoires.

3.1. Les réseaux aléatoires

La naissance de modélisation de réseaux se basant sur les réseaux aléatoires (En anglais: *Random Network*) était en 1959. Les deux chercheurs Erdős et Rényi publient un article [Erd59] dans lequel ils introduisent ce nouveau concept. Ce graphe est composé de N nœuds, et on relie d'une façon aléatoire chaque couple de nœuds (v_i, v_j) par un lien avec une probabilité p . Les auteurs ont proposé plusieurs versions de leur modèle, le plus simple de ces réseaux est appelé $G_{n,p}$. En effet, chaque nœud a $(N - 1)$ voisins potentiels, et donc le degré moyen de chaque nœud pour N est $k = p(N - 1) \approx PN$.

La distribution des liens suit une loi de Poisson de la forme suivante :

$$P(k) = e^{-\lambda k} \cdot \frac{\lambda^k}{k!}$$

D'où k désigne le nombre de liens.

Une caractéristique intéressante du modèle d'Erdős et Rényi réside dans l'apparition de la phase de transition en fonction de k , ce qui provoque la formation d'une composante géante à partir d'une valeur critique de k . L'inconvénient majeur de ce modèle réside dans la distribution homogène du degré de ses nœuds (selon une loi de Poisson) qui n'est pas la distribution hétérogène observée dans les réseaux du réel (selon une loi de Puissance).

3.2. Les réseaux réguliers

Les réseaux réguliers sont des réseaux très ordonnés et souvent utilisés dans des modèles d'automates cellulaires. Dans un réseau régulier, chaque nœud a exactement le même nombre de liens. Autrement dit, tous les nœuds ont un degré de quatre et, par conséquent, la distribution du degré présente une fonction Delta Dirac à quatre. En effet, pour affirmer une structure régulière, soit x le nombre de liens d'un nœud v_i , le nombre x doit être identique pour tous les nœuds v_i du réseau. Ainsi, $\forall v_i \in V, k_{v_i} = 2 * x$. Bien que ce type de structure s'observe en réalité très peu dans la nature, elle est en revanche souvent utilisée comme base pour la formation de réseaux plus réalistes [You 16]. La Figure 1.4 représente le modèle du réseau régulier.

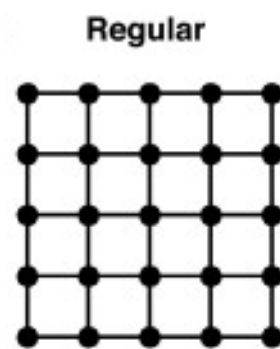


Figure 1.4. Le modèle du réseau régulier

3.3. Le modèle du Small-World (Petit monde)

Le concept (ou notion) de small-world est issu des travaux bien connus de Stanley Milgram dans les années 60. L'expérience consistait à demander à des habitants du Middle West de faire parvenir une lettre à un destinataire de la Côte Ouest en utilisant comme intermédiaire des personnes qu'ils appelaient par leur prénom. Milgram eu la surprise de constater que la moyenne des chaînes parvenues au destinataire.

Plusieurs modèles existent dans la littérature utilisant l'architecture petit monde (Small World), le modèle le plus utilisé est ce de Watts et Strogatz[Watt 98]. La Figure 1.5 représente le modèle de Watts et Strogatz[Watt 98].

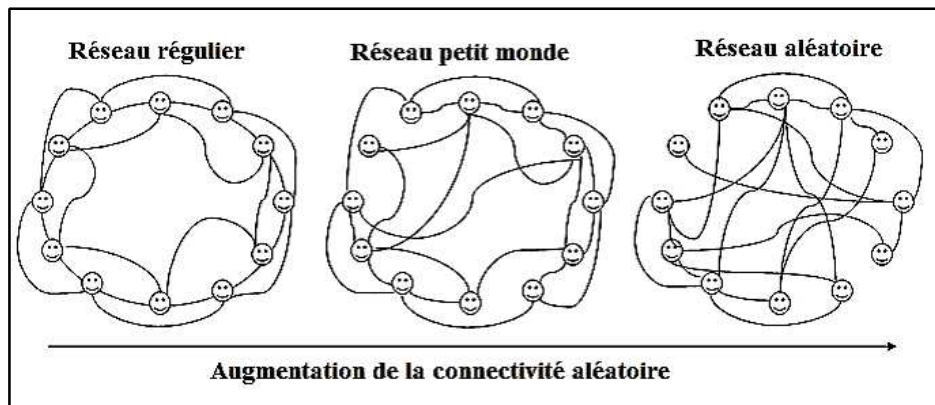


Figure 1.5. Le modèle petit monde (Small World) proposé par Watts Strogatz

3.4. Le modèle duScale-Free (Invariants d'échelle)

En 1999, Albert et Barabasi [Bar 99] ont publié un modèle dynamique permettant d'obtenir une distribution des degrés suivant une loi de puissance des nœuds entrants et sortants dans les citations d'articles scientifiques. L'approche et la méthodologie sont radicalement différentes. Les auteurs partent de l'étude de très gros graphes dont ils cherchent à modéliser la structure et les règles de construction. Contrairement à ce qui était communément admis, les graphes qu'ils étudient ne présentent pas une distribution gaussienne des degrés mais au contraire une distribution de type loi de puissance.

Le réseau est dit invariant d'échelle si, dans l'équation:

$$P(k) \sim k^{-\gamma}$$

où $P(k)$ désigne la fréquence et k le degré, γ est supérieur à 2.

Dans la Table 1.1, nous présentons un récapitulatif des différentes structures citées au-dessus présentées par [Bör07].





Structure	Régulier	Aléatoire	Petit monde	Sans échelle
Représentation				
Nœuds	42	42	42	42
Liens	84	84	84	81

Table 1.1. Présentation des différentes structures de réseaux

Une variété de travaux ont suivi ce travail, afin de comprendre le phénomène de lapropagation d'une maladie à travers ces réseaux. La Table 2.2.Illustre les propriétés de chaque réseau.

Type de réseau	Propriétés
Régulier	-Connectivité de voisinage (treillis généralement réguliers). -Fort clustering ; -Prop de courts chemins.
Petit monde	-Réseau régulier reconnecté avec des raccourcis(les plus courts circuits) ; -Fort clustering ; -Peu de courts chemins.
Aléatoire	- Nœuds connectés avec une probabilité P ; -Faible clustering ; -Trop de courts chemins.
Sans échelle	-La distribution des degrés suit une loi de puissance ; -Le nombre de nœuds de degré k est proportionnel à K^α

Table 1.2. Les propriétés des réseaux.

4. L'analyse des réseaux sociaux :

L'analyse des réseaux sociaux est une méthode d'analyse qui permet l'exploitation et lacompréhension de la morphologie et la dynamique des réseaux sociaux.L'objectif de l'analyse de réseau social est d'identifier les acteurs importants, les liens cruciaux, les sous-groupes, les rôles, les caractéristiques duréseau, et ainsi de suite, pour répondre aux questions de fond sur les structures.

4.1. Les mesures des réseaux

Les mesures caractérisant les réseaux permettent de décrire la structure de ces derniers.De nombreuses mesures ont été proposées et étudiées afin de comprendre la topologiestructurelle du réseau. Dans ce qui suit, nous présentons les principales mesures utiliséesdans notre étude, tout en montrant son intérêt pour l'analyse des réseaux [You 16] :

- **Densité** : la densité d d'un graphe G est la proportion de liens existant dans G par rapport au nombre total de liens possibles : $d(G) = 2m/n(n - 1)$. Si m estde l'ordre de n on dit que le graphe est creux (par opposition aux graphes denses).

- **Plus court chemin** : c'est la longueur du plus petit chemin reliant deux nœuds dans le réseau.
- **Diamètre** : le diamètre d'un réseau est formellement le plus long des plus courts chemins entre deux entités, ou nœuds, du réseau, via ses connexions.
- **Degré** : le degré $d(i)$ d'un nœud est le nombre d'arrêtes incidentes au nœud i , autrement dit, le nombre de nœuds voisins de i .
- **Distribution de degré** : peut-être calculer comme suit [La 2014]

$$P(k) = |\delta(v)| / N$$

$|\delta(v)|$: désigne le nombre de sommets du réseau G ayant le degré k et

N : désigne la taille de G (nombre de nœuds).

- **Coefficient de clustering (coefficient de regroupement)** : est la probabilité que deux voisins d'un nœud soient aussi voisins entre eux. Il peut être interprété comme la probabilité pour que deux plus proches voisins de i soient connectés les uns aux autres.
- **Modularité** : La modularité est une mesure pour la qualité d'un partitionnement des nœuds d'un graphe, ou réseau, en communautés. Elle a été introduite par M. E. J. Newman. C'est aussi une fonction d'optimisation pour certaines tâches de détection de communautés dans les graphes.

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i * k_j}{2m} \right] \delta(c_i, c_j)$$

Où

A_{ij} : est la valeur de la matrice d'adjacence entre les sommets i et j

K_i : est la somme des poids des arêtes adjacentes à i ,

m : est le nombre d'arêtes du graphe et

δ : est le delta de Kronecker qui vaut 1 si ses arguments sont égaux et 0 sinon.

5. Conclusion :

Dans ce chapitre, nous avons constaté que la notion de réseau et les théories de graphe qui lui sont consacrées, sont très anciennes. Cependant, elles sont à un tournant intéressant de leur évolution, puisque aujourd'hui les échanges interdisciplinaires sont nombreux. Nous avons, également, présenté les différents type de réseaux complexes ainsi que les mesures concernées par l'analyse des réseaux sociaux.

Dans le deuxième chapitre, nous allons présenter le système information décisionnelle et les concepts de base sur les entrepôts de données et son moteur OLAP, plus particulièrement, SOLAP et Graph OLAP.

CHAPITRE II : LES SYSTEMES D'INFORMATION DECISIONNELS (SID)

1. Introduction

Depuis des décennies, les systèmes d'informations décisionnels (SID) sont au cœur de la prise de décision dans l'entreprise. Ils permettent aux décideurs d'opérer un meilleur choix entre diverses alternatives possibles pour répondre à un besoin d'analyse particulier.

Dans ce chapitre, nous présentons les Systèmes d'Information Décisionnels (SID) en énumérant les principaux concepts relatifs à l'entreposage des données et à l'analyse en ligne OLAP ainsi que son extension SOLAP.

2. Système d'information décisionnel (SID)

Un nouveau concept est apparu dans les années 90, c'est l'informatique décisionnelle (Business Intelligence). Il connaît un essor considérable grâce au progrès réalisés en technologie de l'information et le partage d'informations via les réseaux à grande échelle.

Un SID est généralement défini comme étant « *un regroupement de données orientées vers certains sujets, intégrées, dépendantes du temps, non volatiles, ayant pour but d'aider les gestionnaires dans leurs prises de décision* » [Kim 05]

En effet, les SID sont des outils du système d'information de gestion orientés vers la production de tableaux de bords et d'outils de pilotage. Ces systèmes visent à faciliter la prise de décision au sein d'une organisation.

2.1. Fonction essentielle d'un SID

Un SID doit être capable d'assurer quatre fonctions fondamentales : la collecte, l'intégration, la diffusion et la présentation des données. À ces quatre fonctions s'ajoute une fonction d'administration, soit le contrôle du SID lui-même [Gou98].

2.2. Architecture générale d'un SID

Dans la littérature, il n'existe pas une architecture standard unique pour un SID, aucun travail jusqu'à présent n'a pu concevoir une architecture idéale. En effet, nous présentons dans la Figure 2.1 une architecture typique d'un SID. En effet, le processus d'entreposage de données est un processus qui est défini généralement selon trois niveaux : (1) acquisition des données, (2) stockage des données et (3) exploitation des données.

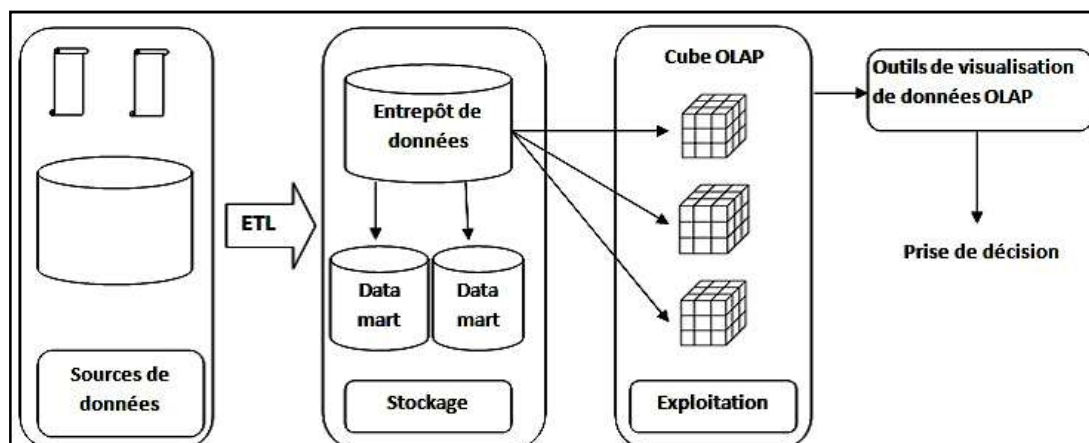


Figure 2.1. Les principaux composants d'un SID

L'architecture détaillée d'un SID, comme indiquée dans la Figure 2.1, est composée des éléments suivants [You 16] :

- **Les sources de données** sont distribuées, variées, et hétérogènes. Elles peuvent être internes ou externes à l'organisme, par exemple : bases de production, rapports, bases des partenaires, etc.
- **Le processus ETL (Extract, Transform and Load)** permet le nettoyage, l'intégration et le chargement périodiques de toutes les données au sein de l'ED nécessaires pour l'analyse.
- **L'entrepôt de données (datawarehouse)** est le lieu de stockage centralisé, il contient les données orientées métier. Une fois ces données stockées dans le Datawarehouse, on va pouvoir créer des magasins de données appelés : **Datamarts**.
- **Les magasins de données (datamarts)** : sont des extraits de l'ED. Les données sont organisées de manière adéquate pour permettre des analyses rapides à des fins de prise de décision.
- **Le cube OLAP** permet d'accéder rapidement et interactivement à des données stockées via une large variété de vues possible d'informations.
- **Les outils de visualisation de données** permettent de visualiser les données selon des axes d'analyses. L'information est visualisée via des interfaces interactives et fonctionnelles dédiées à des décideurs souvent non informaticiens (directeurs, chefs de services, etc).

3. Les entrepôts de données

L'entrepôt de données (ou Datawarehouse en anglais) est le lieu de collecter et stocker de données orientées sujet, intégrées, historisées, et non volatiles utilisées pour le support d'un processus d'aide à la décision [Kim 05].

L'entrepôt de données doit respecter de fortes contraintes de sécurité (sauvegarde, intégrité des données, protection contre l'intrusion) et être accessible par les différentes applications et utilisateurs. Et les caractéristiques des données d'EDD sont:

- **Données orientées sujet** : Les données sont organisées par sujet et structurées par thèmes. L'avantage de cette représentation est d'éviter aux données concernées par plusieurs sujets d'être dupliquées.
- **Données intégrées**: Les données proviennent de diverses sources. Elles sont donc hétérogènes, donc avant d'être intégrées il est nécessaire d'effectuer une forte normalisation pour uniformiser leur format.
- **Données historiées**: Les données ne sont jamais mises à jour. Chaque nouvelle donnée est insérée. Un référentiel temps doit être associé aux données afin de permettre consulter des historiques de changements au cours du temps.
- **Données non volatiles** : Un entrepôt de données veut conserver la traçabilité des informations et des décisions prises. Les données ne sont jamais modifiées ou supprimées.

3.2. L'alimentation d'ED

L'alimentation d'ED se déroule par le processus d'alimentation de données (ETL). En effet, l'outil ETL (Extract, Transform and Load) est un progiciel de type médiateur (middleware) qui permet de réaliser un passage en masse d'information d'une base de données vers une autre. Il est utilisé pour alimenter l'ED à partir de bases de données opérationnelles. Ainsi, il permet l'extraction, le nettoyage et l'importation des données des différentes sources et il les charge dans un entrepôt de données (data warehouse) en temps réel.

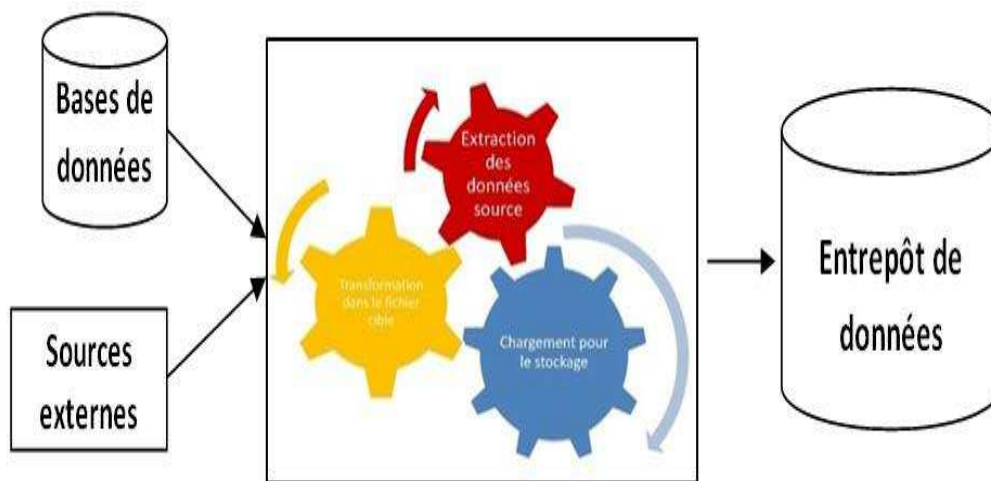


Figure 2.2. Une présentation schématique du processus ETL

4. Modélisation multidimensionnelle des données

Un entrepôt de données est une base de données dédiée à l'analyse en ligne et à l'aide à la décision. Il est conçu selon une modélisation dite **multidimensionnelle** qui se présente classiquement selon un modèle en *étoile*, en *floconde neige* ou *constellation*. Conceptuellement, cette modélisation multidimensionnelle a donné naissance aux concepts de fait et de dimension [Kim 05].

Une table dimensionnelle (ou dimension) modélise une perspective de l'analyse. Une dimension se compose de paramètres correspondant aux informations faisant varier les mesures de l'activité. Une dimension est généralement formée de paramètres (ou attributs) textuels et discrets sur lesquels portent les clauses de conditions et de groupement au sein de requêtes. Chaque dimension est définie par sa clé primaire qui assure l'intégrité référentielle avec la table(s) de faits à laquelle(s) elle est reliée. Le concepteur intervient dans le choix des dimensions de la base de données, il doit collecter les attributs fortement corrélés entre eux dans une dimension. Ces tables dimensionnelles sont équivalentes, en effet, toutes peuvent être vues comme des points d'entrée, symétriquement identiques, dans la table de faits. Chaque modèle multidimensionnel doit être constitué de la dimension temps, pour garder trace de l'historique des données.

4.1. Modélisation conceptuelle de données

Les étapes nécessaires pour la modélisation d'une base de données multidimensionnelle sont les suivantes:

- Détermination des faits représentant les sujets analysés,
- Détermination des dimensions représentant les perspectives de l'analyse,
- Définition des granularités des données de l'analyse,
- Organisation des paramètres des dimensions selon des dépendances de hiérarchie pour supporter les analyses à différents niveaux de détail.

Le modèle multidimensionnel peut être parfaitement mis en œuvre sur une plate forme relationnelle. Dans ce cas, les données peuvent être présentées sous trois schémas possibles:

- **Schéma en étoile** : qui comporte une table de fait principale contenant une clé multiple composée d'un ensemble de clés primaires. Chaque clé primaire permet de relier la table de fait à la table dimensionnelle. La figure 2.3 représente le schéma en étoile

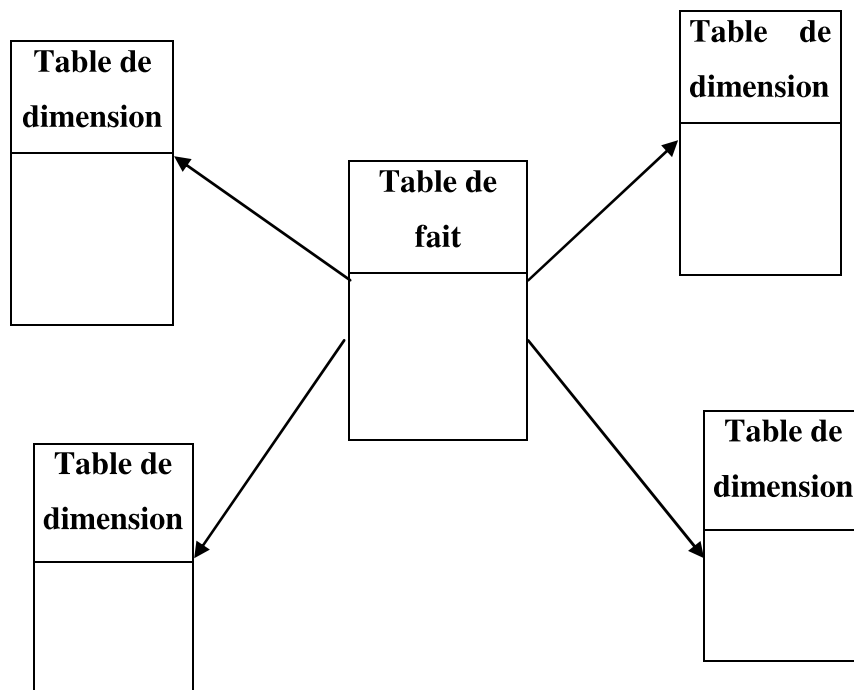


Figure 2.3. Schéma en étoile

- **Schéma en flocon** : est un schéma en étoile mais où chaque dimension peut être représentée par plusieurs niveaux en séparant les attributs qui ont un lien fonctionnel entre eux. La modélisation en flocon consiste à décomposer les dimensions du modèle en étoile en sous hiérarchie. La figure 2.4 représente le schéma en flocon en neige.

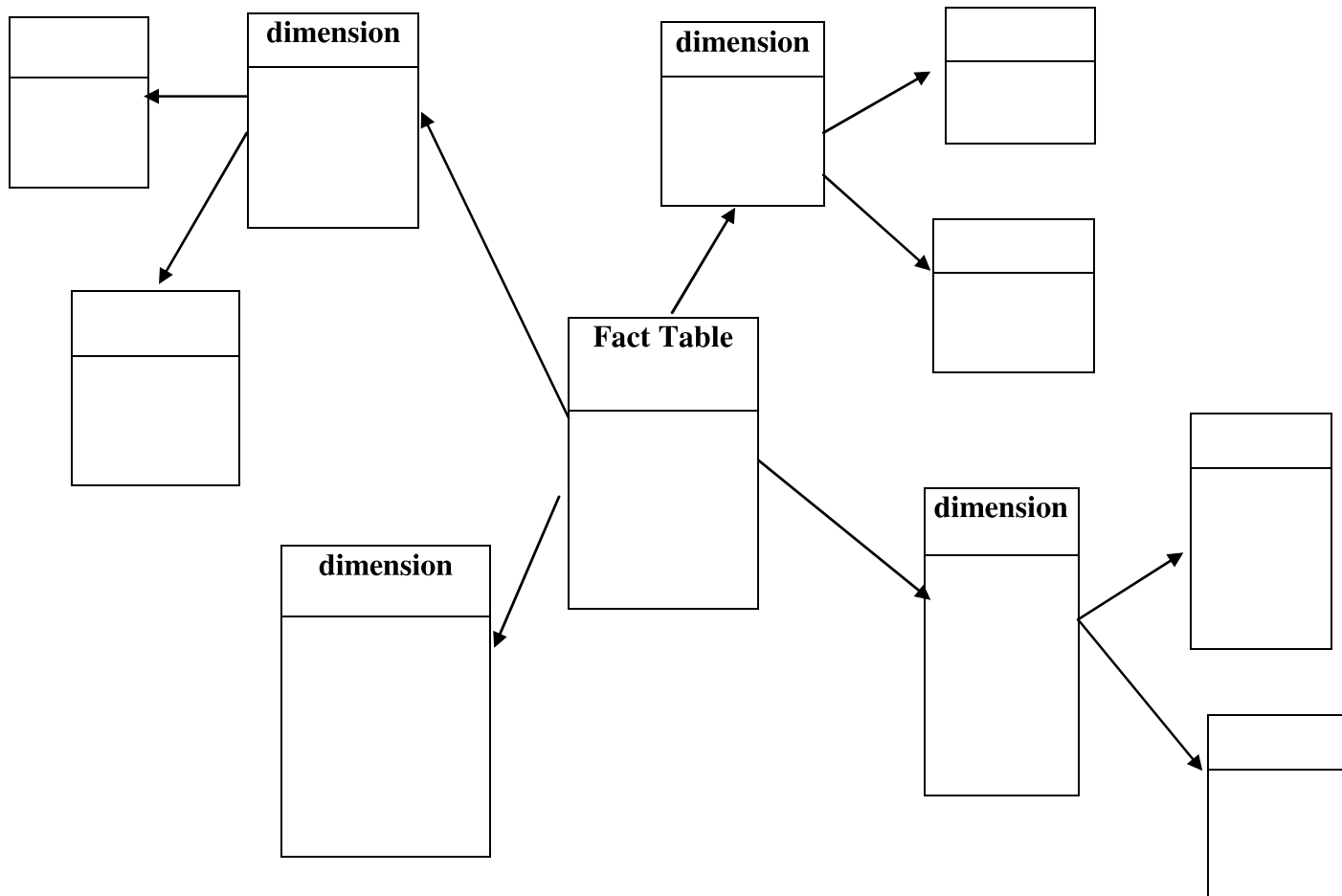


Figure 2.4. Schéma en flocon en neige

- **Schéma en constellation** : Le schéma en constellation est le fusionnement de plusieurs schémas en étoile, nous aurons plusieurs tables de faits et des dimensions communes entre plusieurs schémas en étoile.

Les applications d'aide à la décision utilisent la technologie OLAP (On Line Analytical Processing) qui aide efficacement les décideurs à prendre les meilleures décisions, en leurs fournissant les informations nécessaires[Kim 05].

5. Schéma multidimensionnel : le cube OLAP (On-Line Analytical Processing)

La modélisation multidimensionnelle permet la représentation explicite des hiérarchies et même la possibilité de manipuler à la fois le contenu et la structure des données. La

modélisation multidimensionnelle offre des opérateurs spécifiques associés à ce modèle de données (Drill-down, slice, dice, etc.).

Dans l'OLAP traditionnel, les mesures sont le plus souvent numériques et elles sont assorties d'une fonction d'agrégation (somme, moyenne, minimum, maximum, comptage, etc.) pour calculer la valeur des mesures aux différents niveaux hiérarchiques des dimensions.

5.1. Les architectures OLAP

Les types de bases de données multidimensionnelles ou OLAP permettent des architectures multidimensionnelles et des approches différentes du décisionnel. Dans ce qui suit, nous citons les trois grandes familles du moteur OLAP :

5.1.1. MOLAP: Multidimensionnel OnLine Analytical Processing

Le MOLAP, c'est de l'OLAP multidimensionnel. Dans cette architecture les données détaillées et agrégées sont stockées dans une base de données multidimensionnelle (hypercube). La modélisation se fait directement à partir du cube. Le serveur OLAP extrait les données de l'hypercube et les présente directement au module client. Ces cubes sont implémentés comme des matrices à plusieurs dimensions.

5.1.2. ROLAP : Relational Online Analytical Processing

Le ROLAP, c'est de l'OLAP relationnel. Dans une architecture ROLAP, les données détaillées et agrégées sont stockées dans une base de données relationnelle. Cette dernière est structurée particulièrement selon un : schéma en étoile ou en flocon de neige. Les requêtes OLAP sont traduites en SQL.

5.1.3. HOLAP: Hybrid Online Analytical Processing

L'HOLAP est une architecture qui combine les deux architectures citées au-dessus : MOLAP + ROLAP. Les données sont stockées dans des tables relationnelles et les données agrégées sont stockées dans des cubes. Les requêtes vont chercher les données dans les tables et les cubes.

6. Le Spatial OLAP (SOLAP)

Dans cette section, nous allons présenter des notions de bases sur l'outil SOLAP:

6.1. Entrepôt de Données Spatiales (EDS)

Un Entrepôt de Données Spatiales (EDS) est: "*Une collection de données spatiales et non spatiales : orientées-sujet, intégrées, variantes dans le temps, et non volatiles dédiées à la prise de décision spatiale*" [Ste et al. 00]. Un EDS est une reformulation d'un ED conventionnel à la seule différence d'intégrer des données de nature spatiale.

6.2. Le concept SOLAP

Le concept SOLAP s'appuie sur l'ajout de la dimension spatial au concept de l'OLAP avec adaptation des opérations d'analyses, enrichissement de l'interface et affichage cartographique. Il est défini par Yvan Bédard comme : "*Une plate-forme visuelle supportant l'exploration et l'analyse spatio-temporelle faciles et rapides des données selon une approche multidimensionnelle à plusieurs niveaux d'agrégation via un affichage cartographique, tabulaire ou en diagramme statistique*" [Bédet al.97].

Les systèmes d'EDS et SOLAP se basent sur le modèle spatio- multidimensionnel. Ce modèle étend le modèle multidimensionnel des systèmes d'ED et de l'OLAP avec de nouveaux concepts spatio-multidimensionnels. Ces nouveaux concepts concernent la reformulation des concepts classiques de dimensions et de mesures pour prendre en compte la composante spatiale.

6.2.1. Les dimensions spatiales

Est une dimension qui contient au moins un niveau d'agrégation spatial et permet la représentation de l'information spatiale relative aux localisations géographiques et non géographiques des faits en axes d'analyse. Il existe trois types de dimensions spatiales :

- **Descriptive** : les données n'ont que la valeur nominale de la référence spatiale (nom du pays, nom de l'état, nom de la ville).
- **Géométrique** : chaque élément de la dimension est associé à un objet géométrique.
- **Mixte** : qui contient en même temps des données géométriques et non géométriques.

6.3. La modélisation multidimensionnelle des données spatiales

Plusieurs travaux se sont intéressés à la modélisation multidimensionnelle des données spatiales dans les entrepôts de données [Bédet al.97].

Nous présentons dans la figure 2.5-a un modèle logique générique pour l'analyse multidimensionnelle des données spatiales. Ce modèle est un schéma en étoile classique [Kim 96] composé de N dimensions thématiques et de N dimensions spatiales. Chaque dimension spatiale est composée de niveaux spatiaux et chaque niveau de la dimension spatiale possède une géométrie qui le représente dans l'analyse cartographique et qui est considérée comme un attribut géométrique, en addition aux attributs descriptifs qu'il possède.

Un niveau spatial qui inclut une géométrie peut être représenté dans le modèle logique de deux manières différentes (Malinowski et Zimányi 2007):

- 1) Inclure la géométrie des membres à haut niveau de granularité (ex. les villes) dans la géométrie des membres du niveau parent (ex. pays). Ceci permet d'assurer la contrainte d'inclusion, qui stipule que chaque ville, par exemple, se trouve dans la géométrie qui représente le pays auquel elle appartient.
- 2) Attribuer aux membres de chaque niveau de granularité spatiale une géométrie, puis inclure les contraintes topologiques explicitement. Par exemple, veiller à ce que la géométrie d'un membre du niveau ville se trouve dans la géométrie du membre du pays qui lui correspond.

La figure 2.5-b montre le modèle logique en étoile correspondant au modèle spatio multidimensionnel. Ce modèle logique est constitué des tables de dimensions : Location, Polluant et Temps et de la table de faits. Les dimensions sont dé-normalisées et la table de fait est reliée au niveau le plus détaillé de chaque dimension (Station, Jour et Polluant). Les niveaux spatiaux de la dimension Location disposent d'attributs géométriques (Région Geom, Département Geom et Station Geom) qui représentent les géométries des membres de chaque niveau.

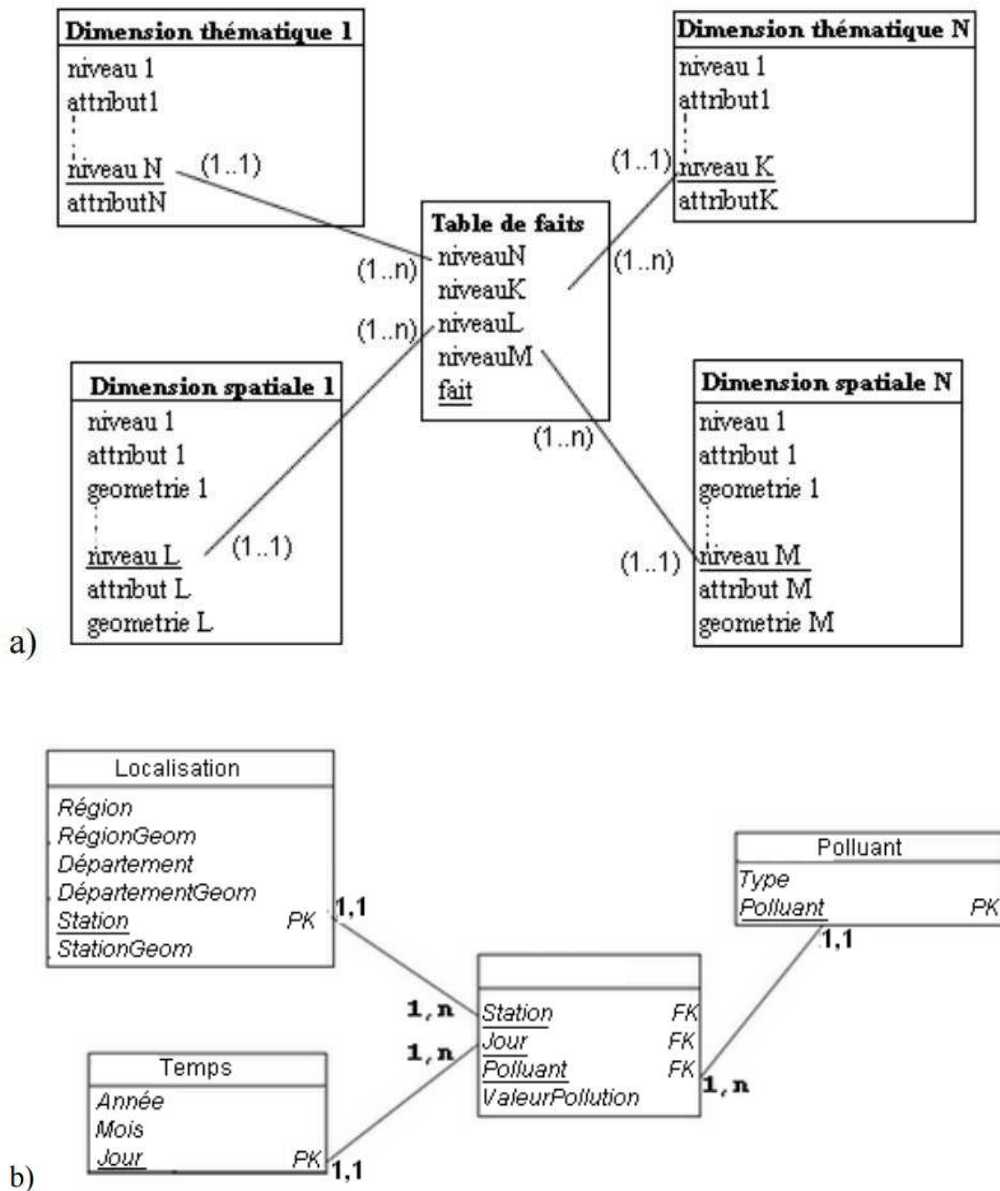


Figure 2.5 a) Modèle logique en étoile générique b) Modèle logique en étoile pour l'analyse de la pollution de l'air

6.4. Les outils SOLAP

Dans cette section, nous allons résumer les outils actuellement disponibles pour SOLAP. Ces outils peuvent être divisés en trois catégories différentes :

- **OLAP dominant** (Business Objects, Cognos, Knosys) : la composante spatiale est accessoire ;
- **SIG dominant** : la composante décisionnelle est accessoire ;
- **SOLAP intégré** : les deux composantes décisionnelle et spatiale sont fortement couplées, elles font appel autant aux fonctions OLAP que SIG (Geo cube, SOVAT).

7. Les graphes et les cubes de données

Dans le contexte des relations sociales, les données peuvent être considérées comme des réseaux hétérogènes de contacts. La construction et l'analyse de tels réseaux posent de nombreux problèmes. A la question de comment synthétiser ou agréger un réseau, l'analyse SOLAP peut apporter une réponse car l'agrégation et la visualisation sont deux points centraux de ce type d'analyse [Che et al.08].

7.1. Les graphes OLAP

Chen et al. [Che et al.08] présentent les fondements et proposent un cadre général appelé *Graph OLAP*.

7.1.1. Définition d'un Graph OLAP:

Nous modélisons les données examinées par le graphe OLAP en tant que collection de clichés de réseau (snapshots) $G = \{G_1, G_2, \dots, G_N\}$, d'où chaque cliché $G_i = (I_{1,i}, I_{2,i}, \dots, I_{k,i}; G_i)$ tel que $I_{1,i}, I_{2,i}, \dots, I_{k,i}$ sont les attributs informationnels décrivant le snapshot dans son ensemble et $G_i = (V_i, E_i)$ est le graphe. Il existe également des attributs de nœud associés à $v \in V_i$ et les attributs de liens attachés à tout $e \in E_i$. Notons que, depuis $\{G_1, G_2, \dots, G_N\}$ on représente des différentes observations, $\{V_1, V_2, \dots, V_N\}$ correspondent réellement au même ensemble d'objets dans les applications réelles.

7.1.2. L'utilisation de Graph OLAP

Graph OLAP est utilisé comme cadre formel par d'autres équipes de recherche. Dans l'OLAP utilisant les graphes, en reprenant la terminologie de [Che et al.08], il peut y avoir plusieurs types de dimensions : les dimensions informationnelles (comme dans l'OLAP traditionnel) et les dimensions topologiques. Ainsi en résulte un nouveau graphe montrant les interactions entre les institutions. Nous pensons que les dimensions topologiques constituent une vraie valeur ajoutée dans la modélisation car elles permettent de modéliser les relations qui existent entre les objets.

La Figure 2.6 illustre un exemple de Graph OLAP de co-auteurs

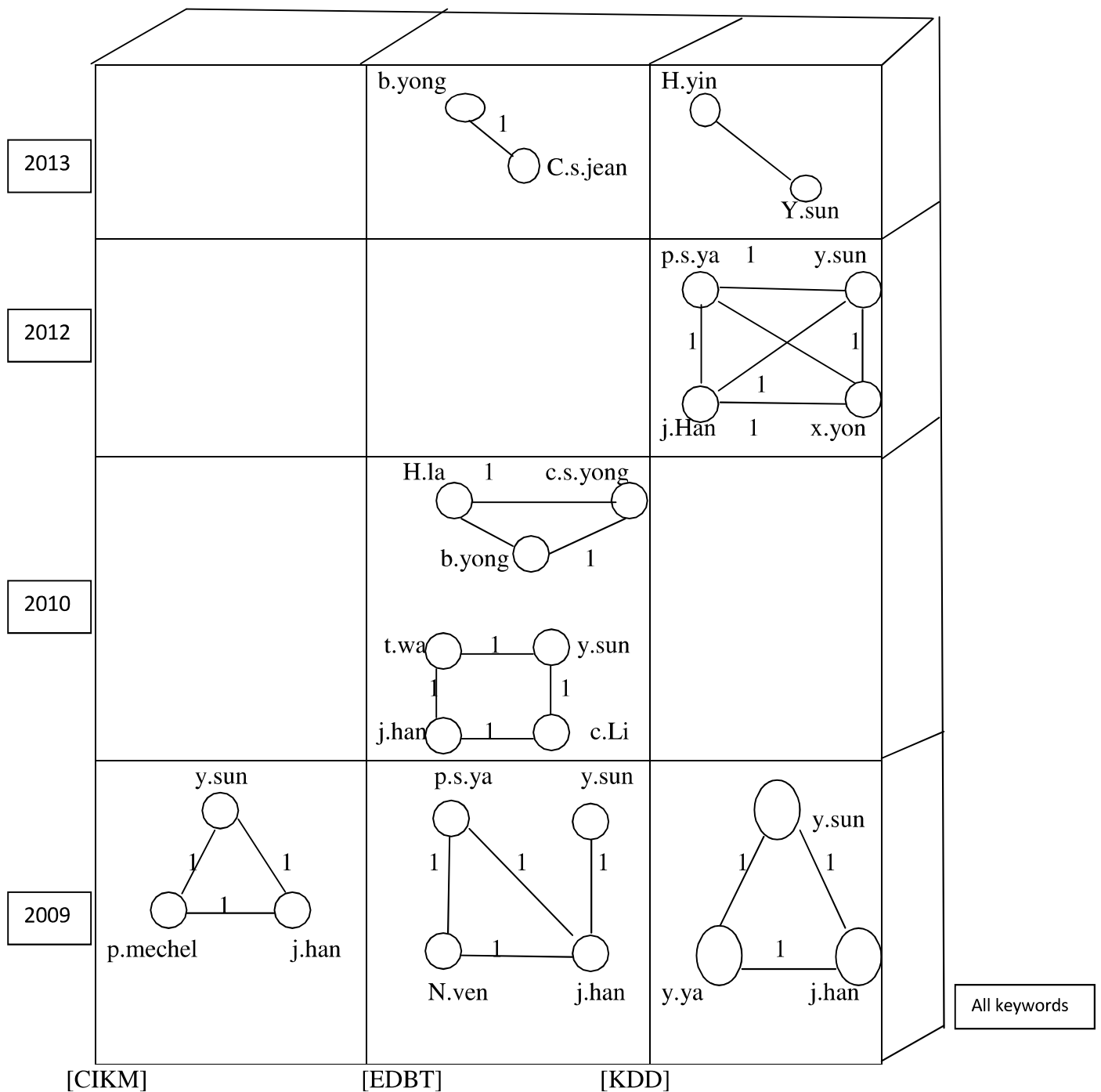


Figure 2.6 Exemple d'un graph OLAP

Les réseaux de co-auteurs sont hétérogènes et représentés sous la forme de graphes dans lesquels on trouve des entités et des liens de différents types :

- Le réseau des co-auteurs avec des auteurs comme nœuds et des arêtes représentant les liens de co-rédaction d'un papier ;
- Le réseau des citations où les nœuds représentent les papiers où les liens entre un papier et les papiers qu'il cite sont représentés par des arcs orientés ;

- Le réseau des thématiques peut relier des papiers ou des auteurs abordant les mêmes thématiques ;
- Le réseau des conférences liant deux conférences quand les mêmes auteurs ont publié dans ces deux conférences.

7.2. Les graphes SOLAP

A notre connaissance, il n'existe pas, dans la littérature, des études consacrées au couplage de graph OLAP et la dimension spatiale. En effet, nous proposons, dans notre projet de fin d'étude, un nouveau cadre pour construire, stocker et analyser plusieurs réseaux de contact et leur dynamique, tout en combinant le Graph OLAP et la donnée spatiale. Les deux idées principales sont d'une part d'être capables de construire et d'analyser plusieurs réseaux issus des mêmes données ; d'autre part de combiner l'analyse Graph OLAP à la donnée spatiale pour enrichir les possibilités d'analyse.

8. Conclusion

Dans ce chapitre, nous avons présenté les concepts de base sur les SID, l'OLAP et son extension SOLAP ainsi que l'analyse multidimensionnelle. Nous avons, également, abordé le nouveau concept les graphes OLAP et les graphes SOLAP.

Dans le chapitre suivant, nous allons présenter en détail notre modèle décisionnel proposé basé sur les deux approches SOLAP et Graph OLAP.

Chapitre III : Conception et Modélisation

1. Introduction

Dans les chapitres précédents, nous avons abordés les principaux concepts sur les réseaux complexes ainsi que le SID.

Dans ce chapitre, nous allons décrire d'une façon détaillée Modèle Décisionnel basé sur une approche Solap et le Graph olap (MODS-GOLAP). Ce dernier intègre plusieurs variantes qui contribuent aux mieux pour analyser les données multiples.

2. Objectifs visés

L'objectif principal de notre étude est de :

- Modéliser et élaborer un MOdèle Décisionnel basé sur une approche Solap et le Graph olap (MODS-GOLAP) pour stocker et analyser les données multiples.

Les objectifs spécifiques étant de :

- Concevoir un entrepôt de données spatiotemporelles et de type graphe;
- Stocker les données multiples dans l'entrepôt de données;
- Effectuer des analyses multidimensionnelles des données.

3. Description du modèle proposé: MODS-GOLAP

Dans cette section, nous présentons, en détails le modèle MODS-GOLAP proposé. Ce dernier est composé de trois principales étapes : la structuration du modèle, l'exécution du modèle et l'exploitation (la concrétisation des résultats).

La Figure 1 représente une vue d'ensemble du modèle MODS-GOLAP proposé :

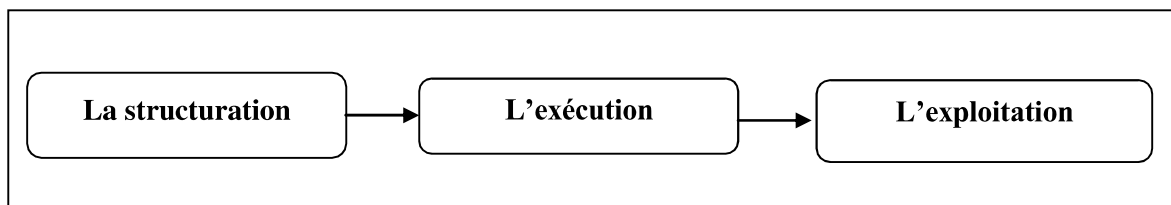


Figure 3.1.Le modèle MODS-GOLAP proposé

3.1. La structuration

L'objectif de cette étape est de constituer un ED pertinent pour l'application visée. C'est une étape nécessaire dans MODS-GOLAP. Cette étape permettra d'identifier les différentes sources de données et de déterminer le mode d'extraction des données désirées comme nous devons aussi tenir compte de la nature particulière des données. Les données utilisées sont de type : spatiale, graphe, temporel et thématique.

La figure 2 représente une vue d'ensemble de cette étape.

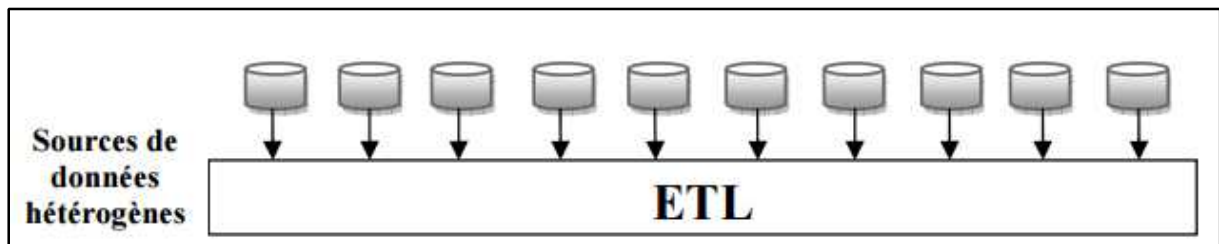


Figure 3.2. La présentation de la phase de structuration

3.2. L'exécution

Cette étape passera le stockage de l'ensemble de ces données dans l'ED pour l'analyse et la visualisation des données historiques.

Le processus ETL joue un rôle primordial dans le système d'entreposage, il permettra l'extraction, la transformation et le chargement des données multiples dans l'ED.

Durant cette étape, le processus ETL doit prendre en considération les points suivants :

- 1) l'utilisation de pilotes ODBC/JDBC pour connecter aux bases de données sources,
- 2) la compréhension de la structure de données de sources, et
- 3) la gestion des données sources de différente nature. Le processus ETL rafraîchit l'ED avec les données modifiées et ajoutées dans les données source. Ce processus est périodique selon le cycle de rafraîchissement et les besoins de l'utilisateur.

La figure 3 représente une vue d'ensemble de l'étape d'exécution.

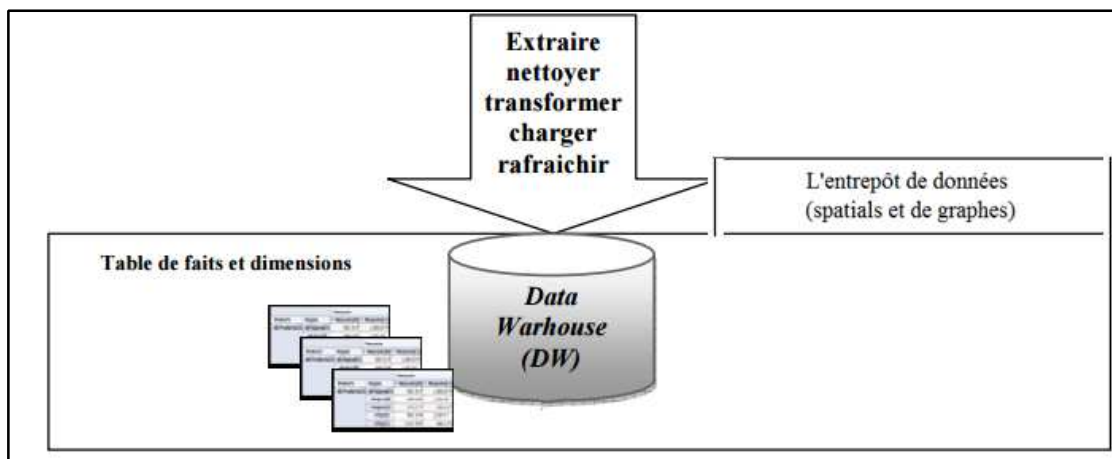
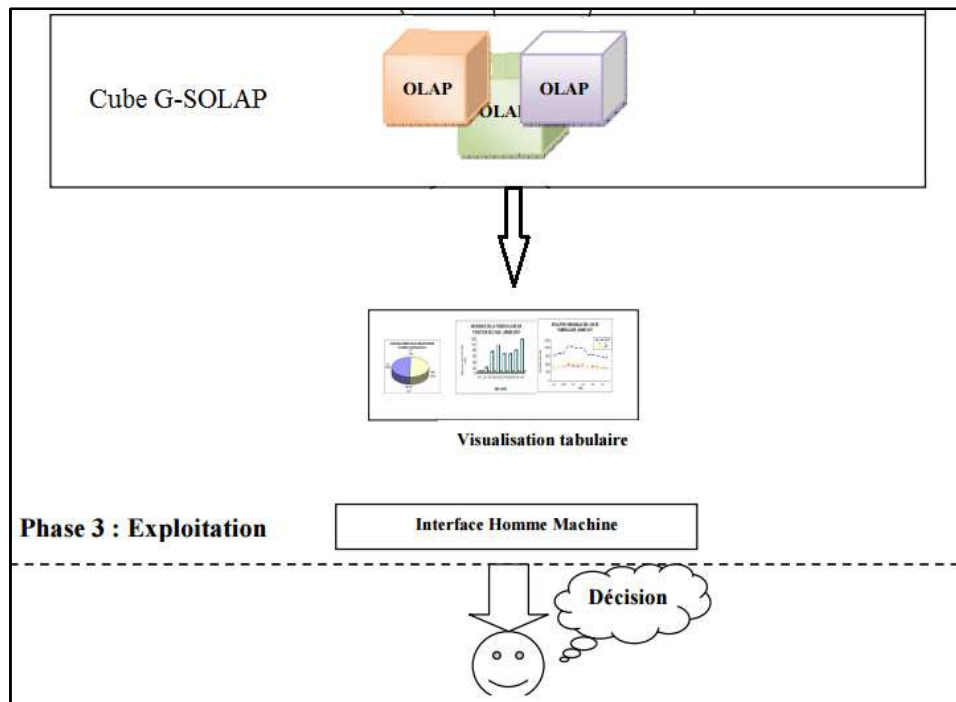


Figure 3.3. La présentation de la phase de structuration

3.3. L'exploitation

Face aux problèmes socio-économiques, les décideurs manquent d'outils simples et puissants pour explorer, synthétiser et analyser des données spatio-temporelles. En effet, la visualisation des foyers épidémiques sur des cartes géographiques est gérée par un SIG. Toutefois, cette technologie est incapable de supporter les problèmes multidimensionnels suite aux progrès des solutions issues du domaine de l'informatique décisionnelle. La solution la plus diffusée pour surmonter les insuffisances des SIG, est de les coupler avec le moteur OLAP. Cette intégration permet la maîtrise de la complexité des problèmes décisionnels à référence spatiale qui nécessite l'utilisation de méthodes, de techniques et d'outils d'analyse puissants, afin de pouvoir non seulement gérer mais aussi analyser des données multiples. Des solutions, telles que SOLAP permettant de supporter les données spatiales dans le G-SOLAP.

C'est une proposition qui se situe au cœur des systèmes d'information géo-décisionnelle. Le couplage (SOLAP–Graph OLAP), permet une représentation des résultats issus des cubes G-SOLAP.



La figure 3.4.Représente une vue d'ensemble de l'étape d'exploitation.

4. Modélisation multidimensionnelle de l'ED

Dans cette section, nous nous intéressons à la conception d'ED dans le but de bien gérer les données multiples.

La modélisation multidimensionnelle est caractérisée par le fait qu'à chaque fait (sujet d'analyse) sont associées plusieurs valeurs observées (mesures), chose qui n'existe pas dans le modèle relationnel.

4.1. Identification des faits : Une table de faits est identifiée formant un modèle en étoile:

La table de fait social network : liée à quatre dimensions, et comporte comme mesures :

- *Nb_frendly_relationship* : le nombre de relations amicales dans un réseau social ;
- *Nb_professional_relationship* : le nombre de relations professionnelles dans un réseau social ;
- *Nb_family_relationship* : le nombre de relations professionnelles dans un réseau social ;

4.2. Définition des axes d'analyse :

Dans ce qui suit, nous décrivons les différentes dimensions du modèle multidimensionnel. Nous distinguons des dimensions qui portent une seule hiérarchie et d'autres plusieurs.

Axes d'analyse thématiques : elle contient les dimensions suivantes :

age_group : 0-5 ans, 6-17 ans, 18-59 ans, 60+ ans ; Gender : homme (H) ou femme (F) ; Age : de 1 an jusqu'à 95 ans.

Axes d'analyse temporelle : c'est une dimension conforme. Elle est liée aux six tables de faits. Nous avons défini cinq dimensions temporelles (Date (jour), Week, Month, Quarter, Year)

Axes d'analyse spatiale géométrique : elle comprend les dimensions suivantes : Commune, décrite par nom de la commune, pollution : la quantité des déchets produite est mesurée par : Kg/Habitat/jour ; périmètre : périmètre du polygone commune, _Name : longitude, _Name_2 : altitude, ID : référence spatial générée par le SQL Spatial, _ID : limite de la zone géographique ; Région : dans le découpage administratif de la ville d'Oran, on trouve trois régions, chaque région regroupe un ensemble de communes (Est-Ouest-Sud).

Axes d'analyse graphe : elle contient les dimensions suivantes : degré, modularité, coefficient de clustering, APL

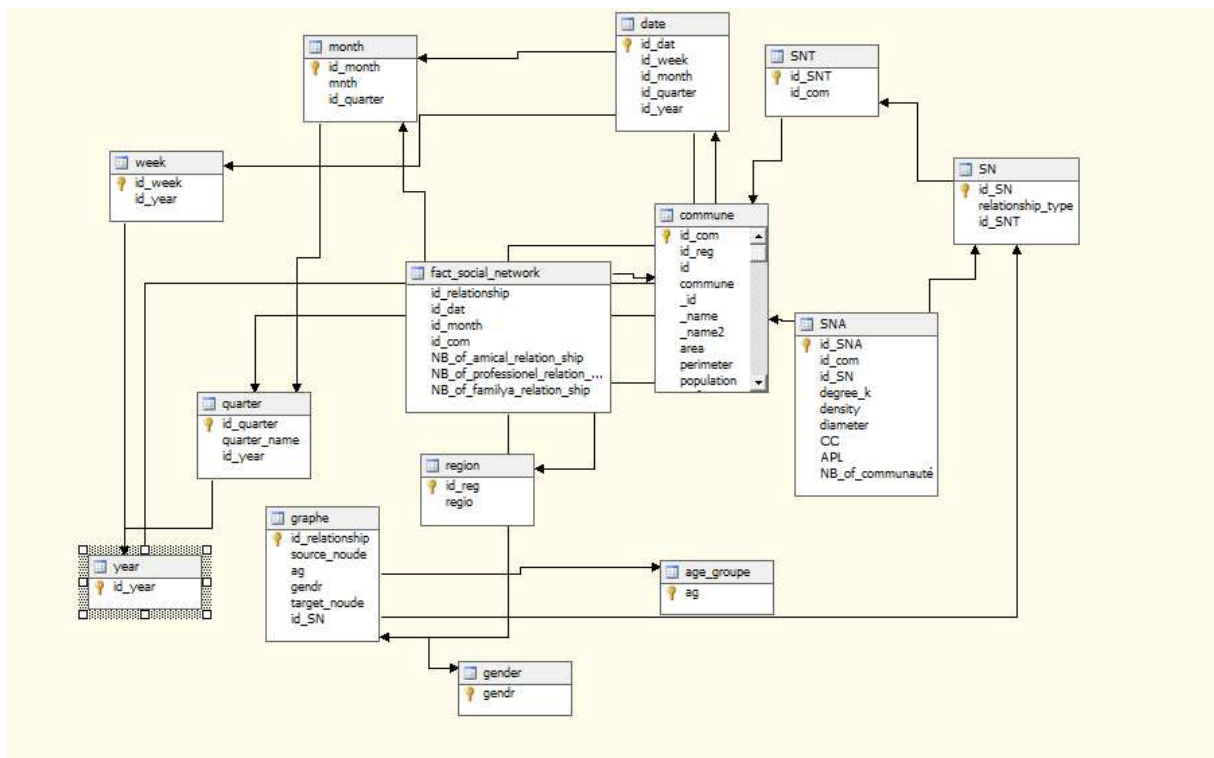


Figure 3.5 – Schéma du modèle multidimensionnel

5. Démarche adoptée par le modèle MODS-GOLAP

La démarche adoptée est donnée par l'organigramme suivant :

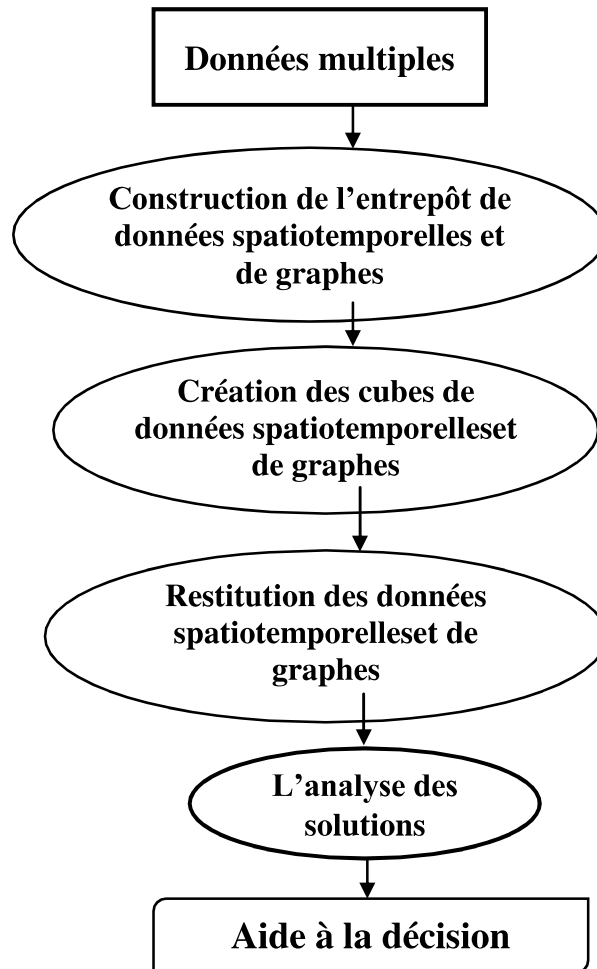


Figure 3.6 : La démarche adoptée par MODS-GOLAP

6. Conclusion

Dans ce chapitre, nous avons décrit le système MODS-GOLAP d'une façon détaillée. Nous avons, également, présenté ces différents composants, à savoir: le sous modèle de structuration, le sous modèle de d'exécution et le sous modèle d'exploitation.

L'objectif du chapitre suivant est de mettre en ouvre le système MODS-GOLAP.

CHAPITRE VI : IMPLEMENTATION ET MISE EN ŒUVRE

1. Introduction

Dans le chapitre précédent, nous avons présenté l'aspect théorique de notre travail. Nous avons présenté notre modèle MODS-GOLAP ainsi que nous avons détaillé tout le processus de notre approche.

Dans ce chapitre, nous allons détailler la mise en œuvre du modèle MODS-GOLAP proposé.

2. Les outils de développements utilisés :

Notre choix était porté sur l'outil Microsoft SQL server 2012 pour implémenter notre base de données, le Visual Studio pour réaliser le cube de données et le Java netbeans pour le développement de notre interface :

2.1. Microsoft SQL Server 2012

Microsoft SQL Server est une application utilisée pour créer des bases de Données informatiques pour la famille des systèmes d'exploitation de Microsoft Windows. Il fournit un environnement utilisé pour produire des bases de données Accessibles à partir des postes de travail, du web ou d'autres média tels qu'un Assistant numérique personnel. SQL Server 2012, la plate-forme de gestion et D'analyse des données la plus complète[1].

2.2. Java Netbeans

NetBeans est un environnement de développement intégré (EDI), placé en open source par Sun en juin 2000 sous licence CDDL (Common Development and Distribution License) et GPLv2. En plus de Java, NetBeans permet la prise en charge native de divers langages tels le C, le C++, le JavaScript, le XML, le Groovy, le PHP et le HTML, ou d'autres (dont Python et Ruby) par l'ajout de greffons. Il offre toutes les facilités d'un IDE moderne (éditeur en couleurs, projets multi-langage, refactoring, éditeur graphique d'interfaces et de pages Web)[2].

2.3. Microsoft Visual Studio 2010

Visual Studio est un ensemble complet d'outils de développement permettant de générer des applications Web ASP.NET, des Services Web XML, des applications bureautiques et des applications mobiles. Ces outils permettent de créer notre application [3].

3. Présentation du système MODS-GOLAP:

Le processus de mise en place de l'application SOLAP peut être schématisé selon l'organigramme suivant :

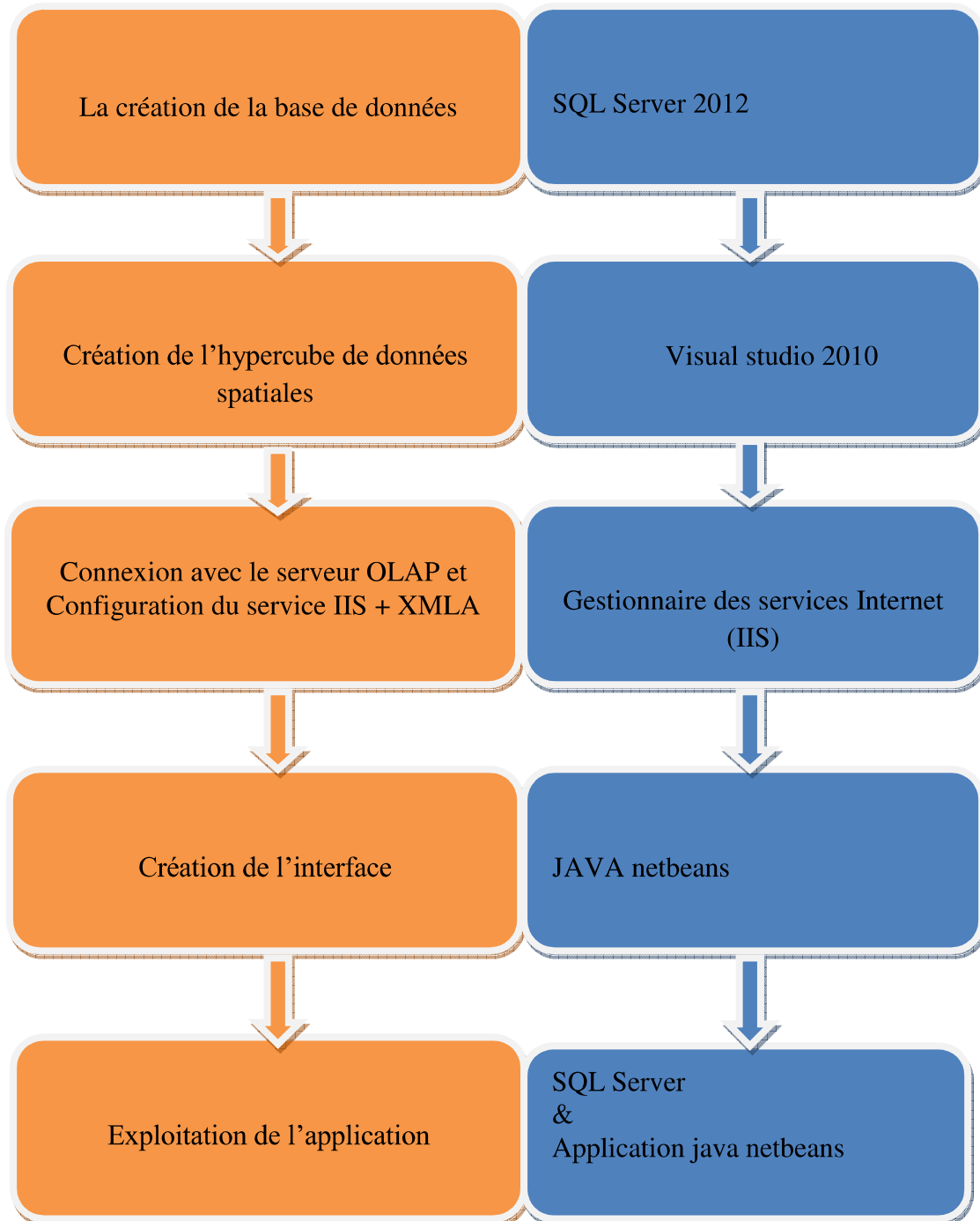


Figure 4. 1. Organigramme de travail

3.1. Création de la base de données

Dans cette étape, nous allons créer notre base de données en utilisant SQL Server2012 plus précisément SQL Server Management Studio.

La figure ci-dessus représente la Fenêtre principale de SQL Server management studio :

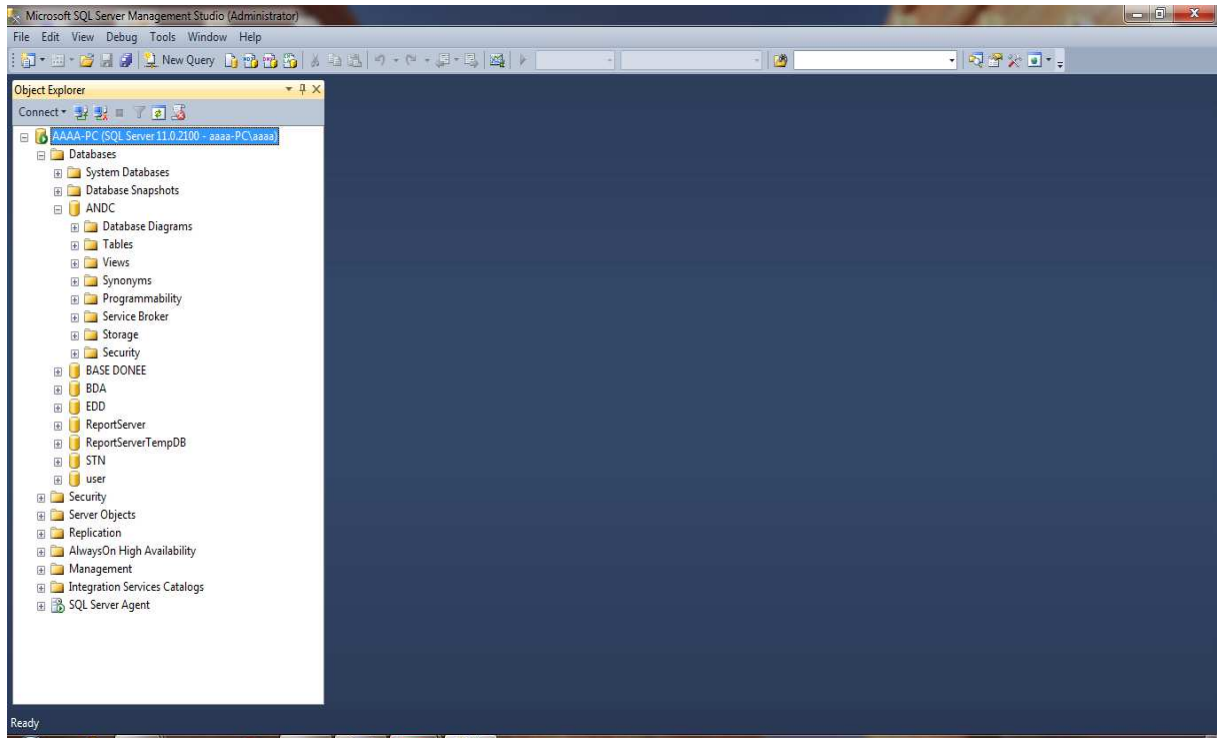


Figure 4.2 : Fenêtre de SQL Server management studio.

La création de la base de données se fait à travers la fenêtre illustrée par la figure 4.2 dans le SQL Server Management Studio(Voir figure 4.3) :

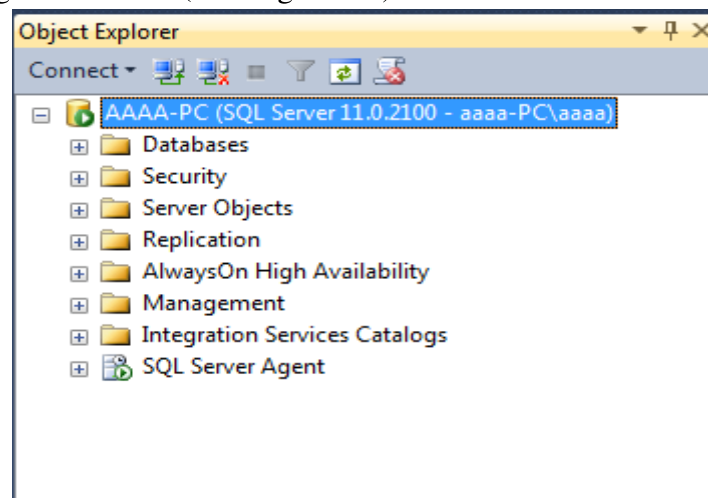


Figure 4.3 : Création de la base de données.

Une fenêtre s'ouvre où nous allons préciser le nom de la base de données : ANDC et le nom serveur : aaaa-PC\aaaa. (Voir figure 4.4)

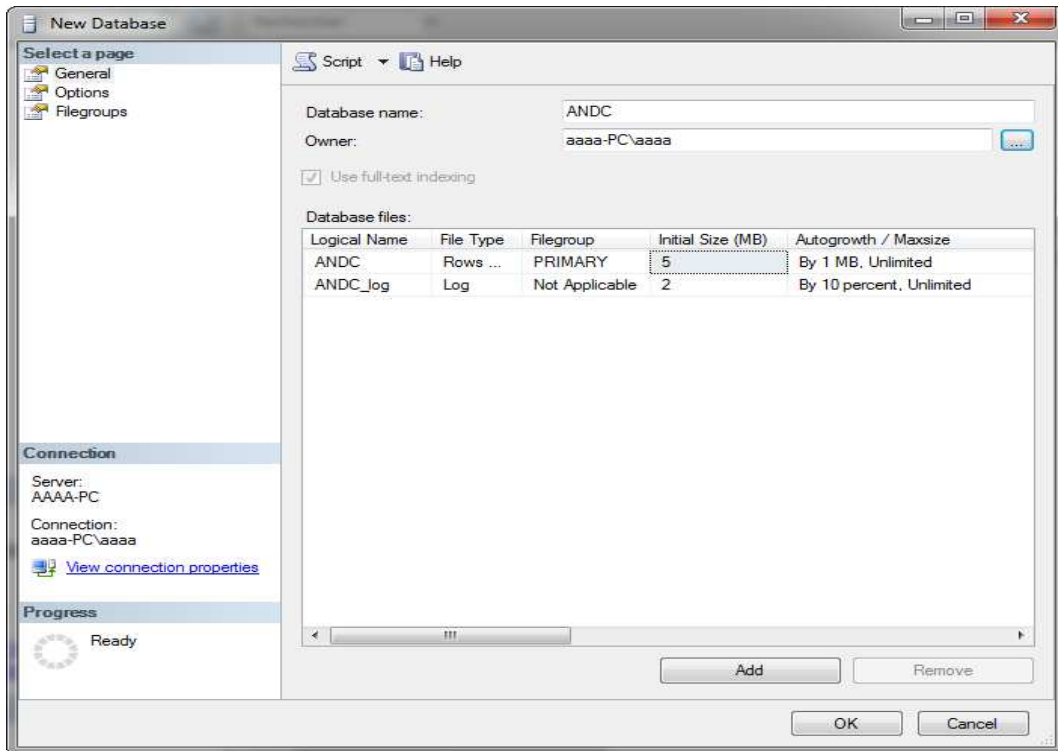


Figure 4.4 : Fenêtre de création de la base de données.

Après la validation de la base de données, on accède à une autre section où nous pouvons créer les tables que nous avons besoin dans notre base. (Voir figure 4.5)

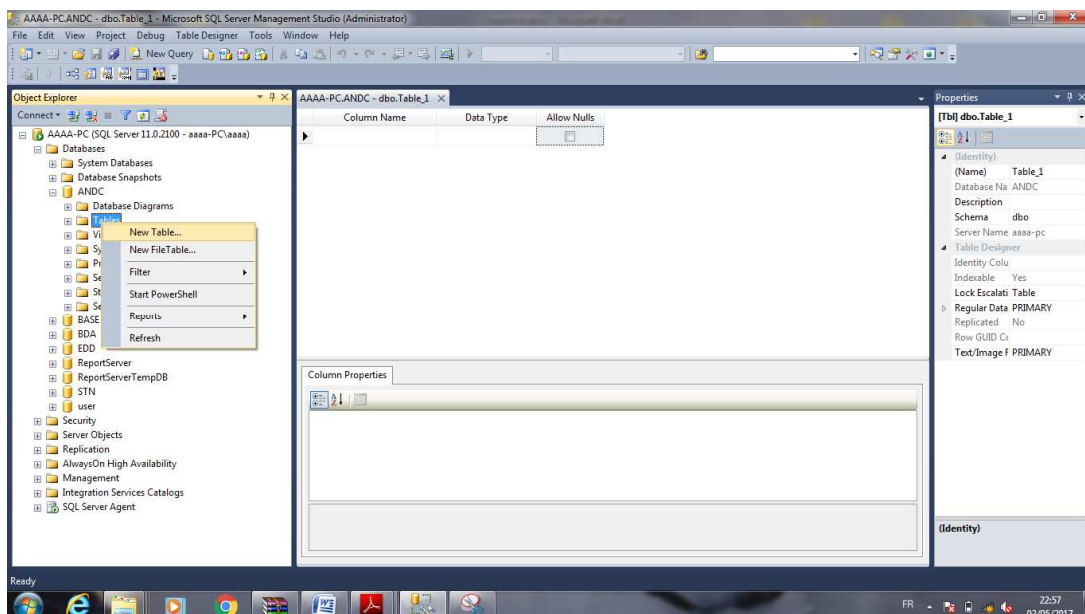


Figure 4.5 : Création de la table.

Depuis cette fenêtre, nous pouvons créer des colonnes pour la table choisie, en spécifiant le type de donnée et la clé primaire de chaque table. (Voir figure 4.6)

Column Name	Data Type	Allow Nulls
id_com	int	<input type="checkbox"/>
id_reg	int	<input checked="" type="checkbox"/>
id	int	<input checked="" type="checkbox"/>
commune	nvarchar(50)	<input checked="" type="checkbox"/>
_id	nvarchar(50)	<input checked="" type="checkbox"/>
_name	bigint	<input checked="" type="checkbox"/>
_name2	bigint	<input checked="" type="checkbox"/>
area	float	<input checked="" type="checkbox"/>
perimeter	float	<input checked="" type="checkbox"/>
population	bigint	<input checked="" type="checkbox"/>
surface	float	<input checked="" type="checkbox"/>
densite	bigint	<input checked="" type="checkbox"/>

Figure 4.6 : Création des colonnes pour la table commune.

Après la création des tables nous allons remplir les différents champs par les données. La fenêtre ci-dessus montre le processus de remplissage (Voir figure 4.7)

	id_com	id_req	id	commune	_id	_name	_name2	area	perimeter
1	1	2	1	TAFRAOUI	limiteR1007	1007	1006	187303000	86377,414
2	2	3	2	SIDI CHAMI	limiteR935	935	934	68928952	45138,527
3	3	3	3	EL KERMA	limiteR953	953	952	64145668	56631,242
4	4	2	4	MISSERGHIN	limiteR924	924	923	433650000	165357,02
5	5	2	5	BOUTLELIS	limiteR942	942	941	139710000	66587,602
6	6	2	6	AIN EL KERMA	limiteR929	929	928	106692000	50955,113
7	7	3	7	ES-SENIA	limiteR943	943	942	50388748	36493,441
8	8	2	8	OUED TLELAT	limiteR986	986	985	84957416	44243,086
9	9	2	9	EL BRAYA	limiteR964	964	963	56506924	37961,559
10	10	1	10	BOUFATIS	limiteR931	931	930	98461880	64143,129
11	11	1	11	HASSI BOUNIF	limiteR922	922	921	31934642	28796,828
12	12	2	12	AIN EL TURCK	limiteR884	884	883	30710626	36770,828
13	13	2	13	BOUSFER	limiteR888	888	887	45371072	37447,449
14	14	2	14	EL ANCOR	limiteR912	912	911	65358084	45717,438
15	15	3	15	ORAN	limiteR898	898	897	62475328	64458,617
16	16	1	16	MERS EL KEBIR	limiteR909	909	908	13690840	35786,922
17	17	3	17	BIR EL DJIR	limiteR889	889	888	39154936	38333,25
18	18	1	18	HASSI MEFSOU...	limiteR861	861	860	25752668	30035,736
19	19	1	19	AIN EL BYA	limiteR852	852	851	47216516	40030,094
20	20	1	20	BENFREHA	limiteR902	902	901	70944072	47853,543
21	21	1	21	BETHIOUA	limiteR863	863	862	94235952	75493,898
22	22	1	22	HASSI BEN OKBA	limiteR885	885	884	36648908	29805,217
23	23	1	23	GDYEL	limiteR826	826	825	96998568	53481,43

Figure 4.7 : Exemple d'une table.

3.2. Création du cube de données

Pour la création du cube de données nous allons suivre les étapes suivantes:

1. Création de la source de données.
2. Création de la vue de source de données. (Voir figure 4.8)

Pour définir les groupes de mesures et les dimensions d'un cube, nous devons utiliser l'Assistant Cube. La table de faits apparaît en jaune et les tables de dimension apparaissent en bleu.

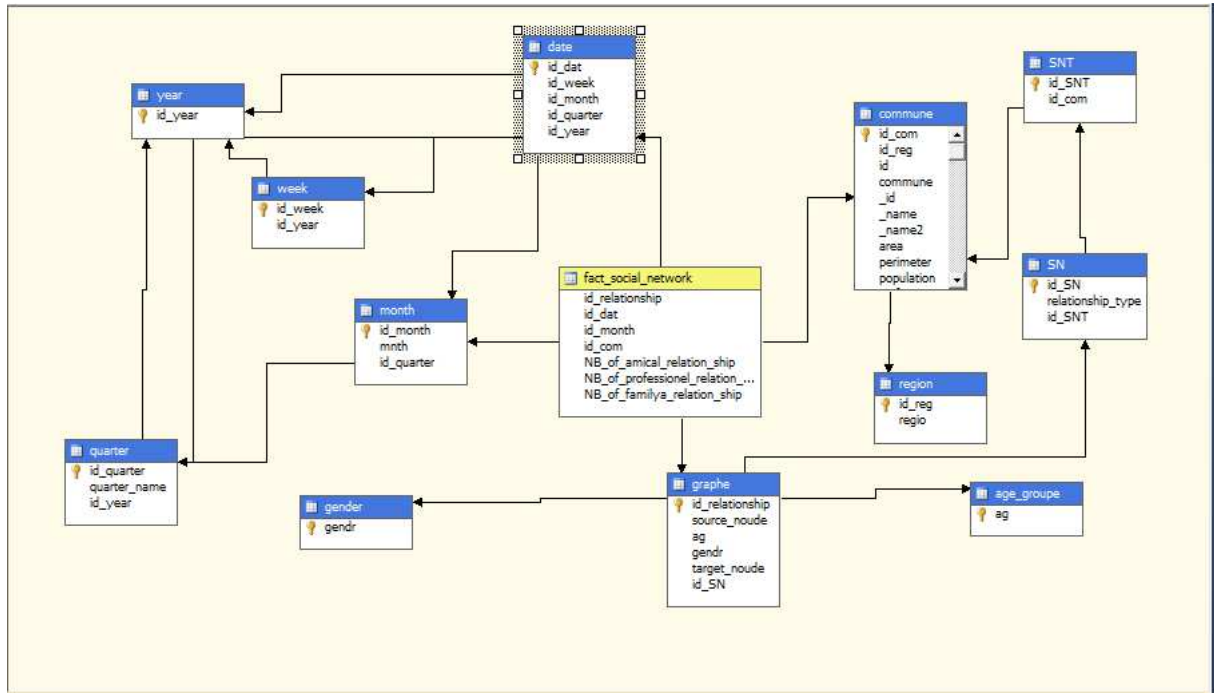


Figure 4.8 : Le schéma du cube de données du fait réseau social.

Pour afficher les données de cube et les données de dimension du cube, nous devons déployer le projet puis traiter le cube et ses dimensions. Le déploiement d'un projet entraîne la création des objets dans une instance d'Analysis Services. Le traitement entraîne la copie des données à partir des sources de données sous-jacentes dans le cube.

Vérifions maintenant les propriétés de déploiement du projet. La nouvelle boîte de dialogue présente les propriétés de la configuration Active(Development).

La réalisation de ces étapes est suivie par le déploiement pour vérifier que le cube est correctement créé. (Voir figure 4.9)

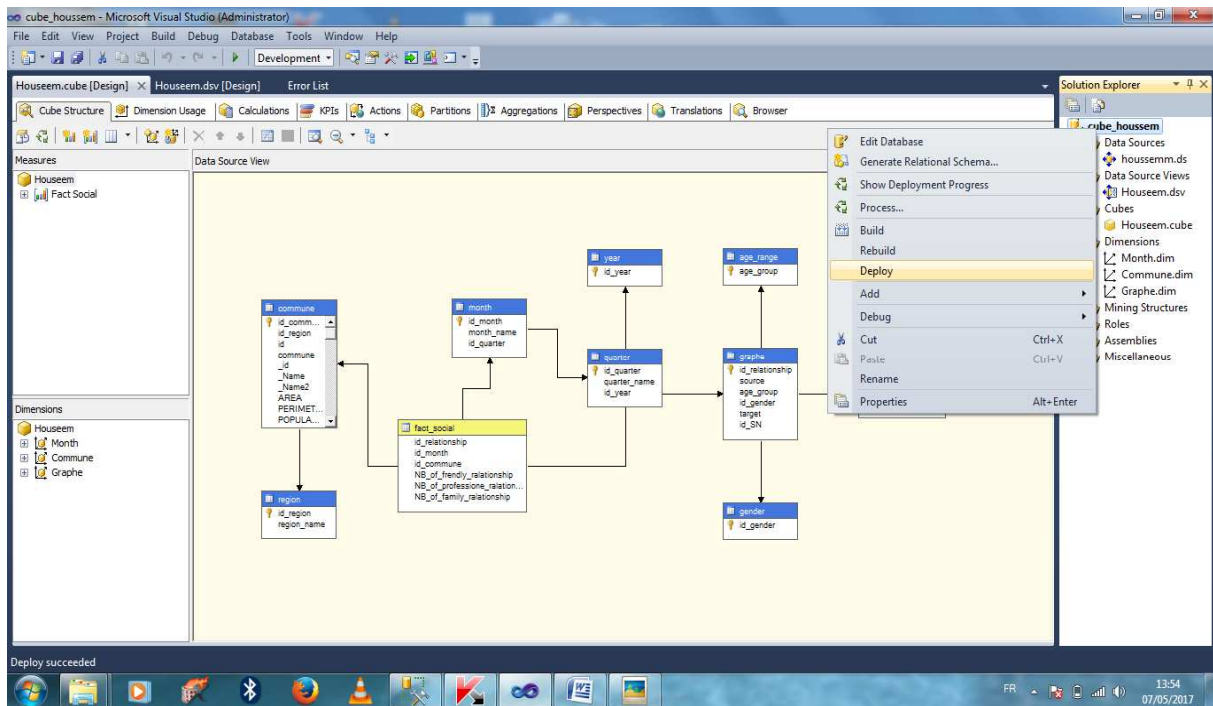


Figure 4.9 : le déploiement du cube.

4. Connexion au serveur OLAP et Configuration du service IIS + XMLA

Pour accéder à une instance Analysis Services, nous devons configurer un point de terminaison HTTP. On peut activer l'accès HTTP en configurant MSMDPUMP.dll, une extension ISAPI qui s'exécute dans Internet Information Services (IIS) et qui pompe des données entre des applications clientes et un serveur Analysis Services.

Pour configurer IIS, nous suivrons les étapes suivantes :

1. Copier les fichiers MSMDPUMP dans un dossier du serveur web.
2. Créer un pool d'applications et un répertoire virtuel dans IIS.
 - Ouvrez le dossier serveur, cliquez avec le bouton droit sur **Pools d'applications** →Ajouter un pool d'applications → nommé le pool **OLAP** →OK
 - Dans le Gestionnaire IIS, ouvrez **Sites**, puis **Site Web par défaut**. Un dossier **Olap** doit s'afficher. Il s'agit d'une référence au dossier OLAP que vous avez créé sous \inetpub\wwwroot.
3. Configurer l'authentification IIS et ajouter l'extension.

5. Présentation de notre application

Nous avons utilisé le langage java NetBeans pour la programmation de notre application.

5.1. L' interface d'accueil

La réalisation technique commence par la présentation de la fenêtre principale, il s'agit de ce qu'il voit l'utilisateur (un Menu) qui comporte plusieurs tables de dimensions liées à notre problématique et une partie dédiée aux analyses. Dans cette fenêtre, nous avons quatre types de dimensions.

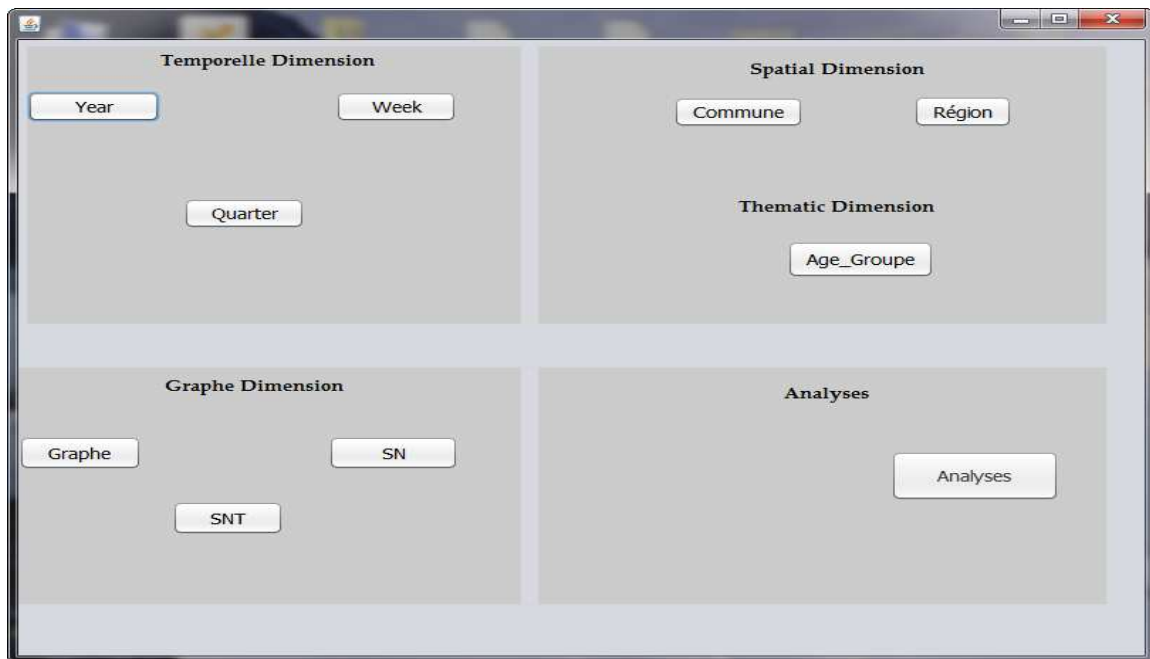


Figure 4. 10 : Présentation la fenêtre principale.

5.2. La mise à jour des données

La fenêtre illustrée par la figure 4.11, permet de insérer, de supprimer ou de modifier dans les tables de la base. Nous présentons le cas de la table commune.

The image shows a software window titled "tab_commune". Inside the window, there is a list of 12 fields on the left, each with a corresponding empty input box to its right. The fields are: id_com, id_région, id, commune, _id, _name, _name2, area, perimeter, population, surface, and densité. To the right of the input boxes, there are three buttons: "Insert", "Delete", and "Update". The "Insert" and "Delete" buttons are positioned higher than the "Update" button.

Figure 4. 11. Fonction de la mise à jour de la table commune.

5.3. Choix des dimensions et de mesures à analysés

Cette fenêtre a été créée pour être facile à utiliser et manipuler des différentes opérations, elle est composée par :

- des différents composants qui permettent de choisir les paramètres des requêtes MDX, tels que :
 - Choix de mesure (Indicateur).
 - Choix de dimension.

Elle offre aussi un contrôle de tous ces modules :

- Supprimer les requêtes et faire des nouvelles.
- Requêtes personnalisées.
- Différents modes de visualisation des résultats.

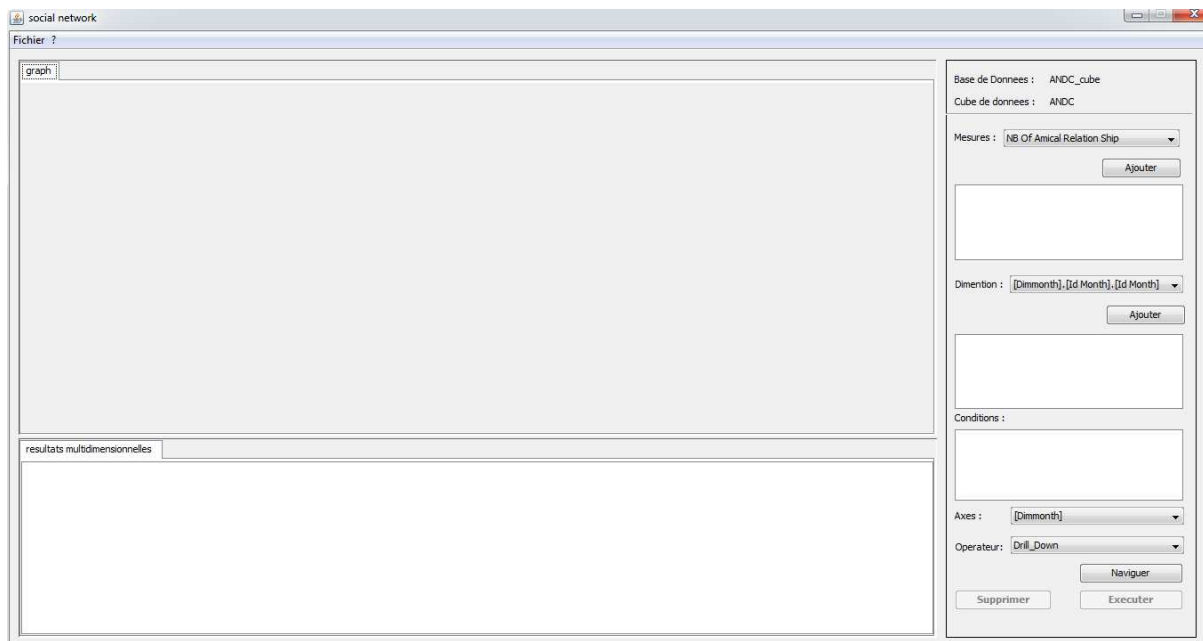


Figure 4. 12 : Choix des dimensions et des mesures

6. La phase d'exploitation

L'exploitation se fait en deux manières :

- A travers les requêtes MDX sur SQL Server
- A travers les requêtes MDX sur l'application java

6.1. Les requêtes MDX :

Le MDX pour « **Multidimensionnel Expressions** » est un langage de requête dédié aux données structurées de manière multidimensionnelles.

La syntaxe MDX, finalement assez proche du SQL, se base sur une structure de type « SELECT ... FROM ... WHERE ». Le « SELECT » contient les mesures et les membres de dimensions à afficher en rangées et en colonnes (équivalent au « GROUP BY » en SQL), le « FROM » contient le cube de données définissant la structure multidimensionnelle des données et le WHERE permet d'appliquer des coupes sur plusieurs membres de dimensions.

Example:

```
SELECT [Measures]. [NB Of profession Relation Ship] ON COLUMNS, [Dimcommune]. [Id Com].members ON ROWS FROM [ANDC]; (Voir figure 4.13)
```

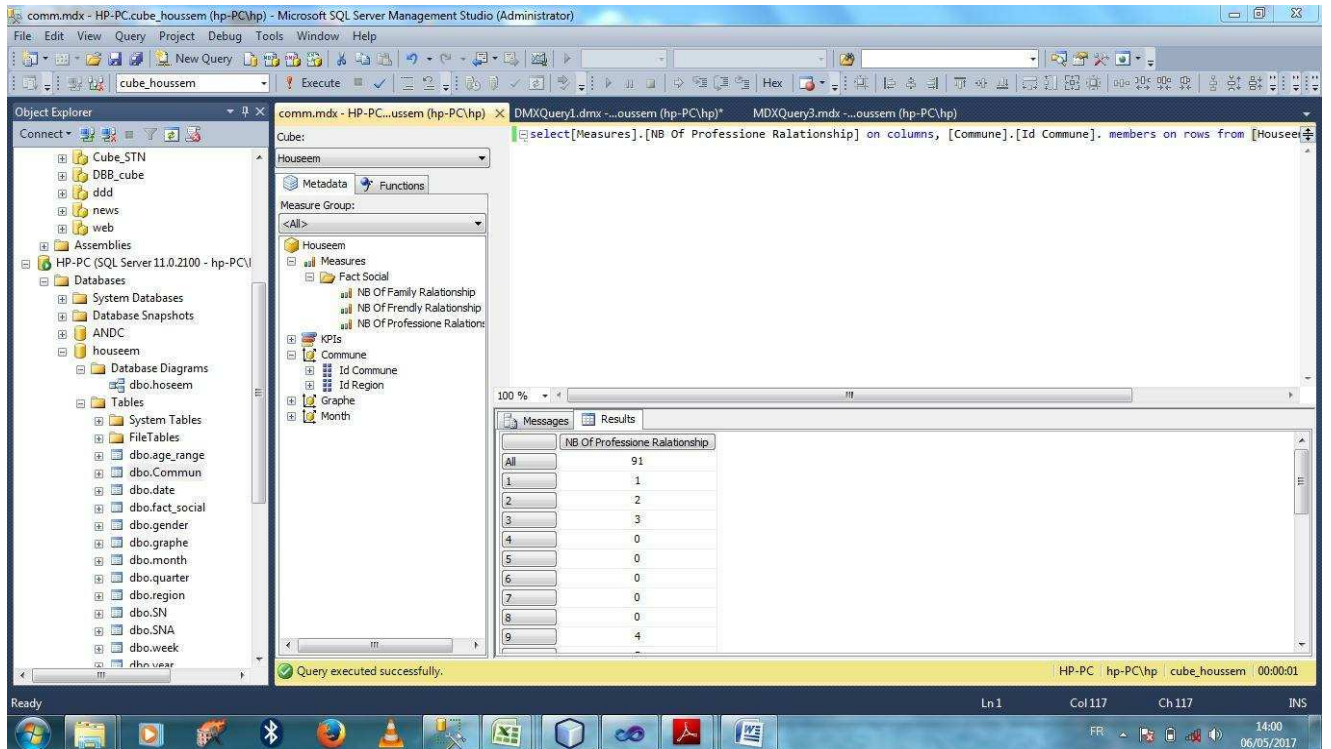


Figure 4. 13. Résultat de la requête

6.2. Visualisation des résultats MDX et d'histogramme :

La représentation des résultats d'histogramme sont affiché en mode tabulaire sous la forme classique d'un tableau à des lignes et des colonnes (montré dans la figure ci-dessous), qui donne les mesures de l'analyse effectuée sous forme d'une liste (voir figure 4.14).

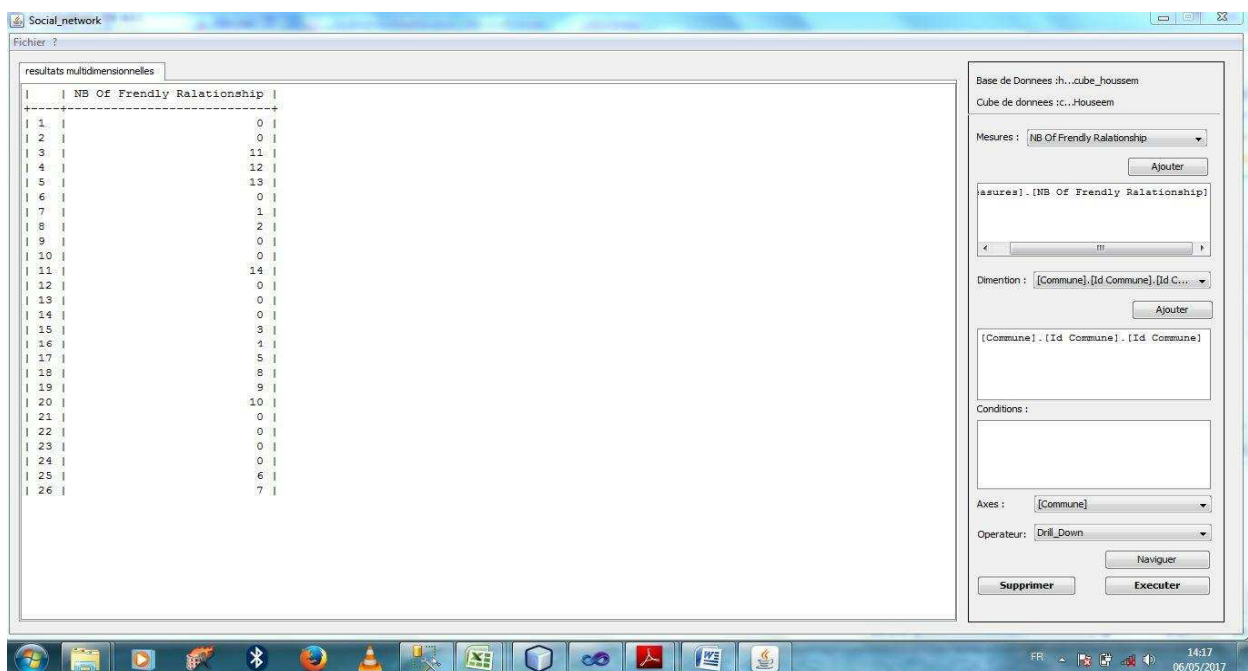


Figure 4. 14. Visualisation des résultats MDX

La représentation des résultats d'histogramme sont affichés en mode tabulaire sous la forme classique d'un tableau à des colonnes (montré dans la figure ci-dessous : on choisie le résultat des relations familière par rapports à la table région), qui donne les mesures de l'analyse effectuée (voir figure 4.15).

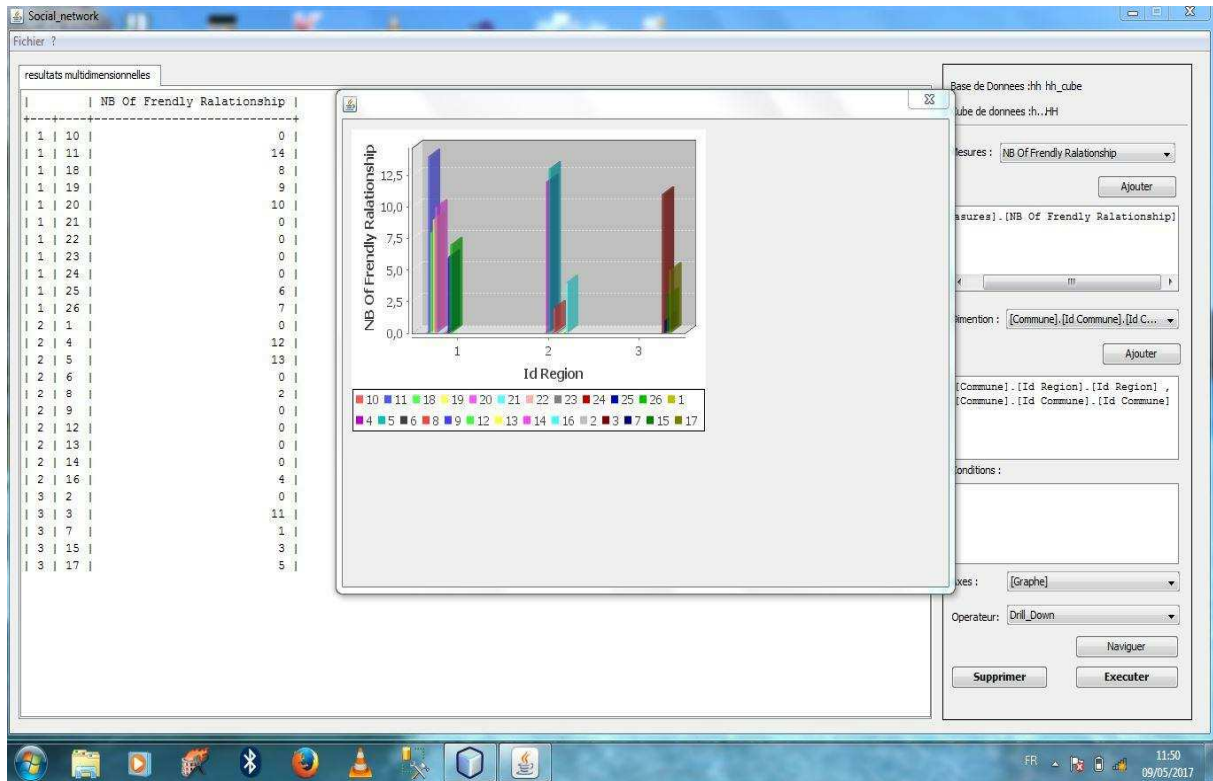


Figure 4. 15 : Représentation graphique deux nombre de relations amicales par rapport aux régions

7. Conclusion

Dans ce chapitre, nous avons présenté les outils utilisés pour l'implémentation de notre application en exposant les différents composants et fonctionnalités permettant le développement d'une application décisionnelle.

CONCLUSION GENERALE ET PERSPECTIVES

Le concept de réseau offre un modèle de représentation pour une grande variété d'objets et de systèmes, aussi bien naturels que sociaux, dans lesquels un ensemble d'entités homogène ou hétérogènes interagissent entre elles. La construction et l'analyse de tels réseaux posent de nombreux problèmes donnant naissance à l'outil Graph OLAP.

Les systèmes SOLAP peuvent améliorer les connaissances des intervenants, faciliter la consultation des données, permettre une gestion rapide des risques à plusieurs niveaux de détails. Cependant, à la question de comment synthétiser ou agréger des réseaux de contact répartis dans l'espace, l'analyse SOLAP peut apporter une réponse car l'agrégation et la visualisation sont deux points centraux de ce type d'analyse. Cela nécessite une évolution de l'outil d'analyse SOLAP pour analyser des données issues de réseaux de contact et modélisées sous forme de graphes.

En effet, l'objectif principal de notre projet de fin d'étude est de concevoir, modéliser et élaborer un MOdèle Décisionnelle basé sur une approche Solap et le Graph olap nommé MODS-GOLAP qui permet d'enrichir le cube SOLAP.

Le modèle traite un problème décisionnel spatiotemporel de type graphe. Il permet, en premier lieu, de concevoir et de réaliser un entrepôt de données multiples. Ensuite, MODS-GOLAP applique le processus d'analyse multidimensionnelle afin de comprendre la structure d'un réseau social par rapport au temps, à l'espace et au type de graphe. Enfin, il fournit des résultats et permet des visualisations des résultats.

Nous avons abordé cette problématique à travers quatre chapitres :

Le chapitre 1 est dédié à la présentation des réseaux complexes et les définitions relatives aux modèles de réseaux.

Dans le chapitre 2, nous avons effectué quelques rappels sur les SID et son moteur OLAP ainsi que son extension SOLAP. Nous avons, également, présenté l'architecture générique d'un SID et discuté ses modèles logiques. Ces notions sont nécessaires à une bonne compréhension des contributions

Dans le troisième chapitre, nous avons décrit notre proposition MODS-GOLAP.

Dans le quatrième et dernier chapitre, nous avons détaillé la conception et la mise en œuvre du modèle MODS-GOLAP.

Perspectives

A l'issue de cette étude, nous avons pu atteindre notre objectif qui portait sur l'élaboration du modèle décisionnel basé sur une approche SOLAP et Graph OLAP. Nous proposons comme perspective à notre travail :

1. D'enrichir le modèle d'aide à la décision proposé par l'utilisation d'un modèle d'épidémie,
2. De simuler et de visualiser la mobilité des individus entre les différentes régions.
3. D'améliorer l'affichage cartographique des résultats par l'utilisation d'un véritable SIG afin d'exploiter ses différentes fonctionnalités.

Bibliographie

- [Bar 99]: Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *science*, 286(5439), 509-512.
- [Béd et al. 97]: Bédard, Y., Bernier, E., Larrivé, S., Nadeau, M., Proulx, M. J., & Rivest, S. (1997, November). Spatial olap. In *Forum annuel sur la RD, Géomatique VI: Un monde accessible* (pp. 13-14).
- [Bull 09]: Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3), 186-198.
- [Che et al. 08]: Chen, C., Yan, X., Zhu, F., Han, J., & Philip, S. Y. (2008, December). Graph OLAP: Towards online analytical processing on graphs. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on* (pp. 103-112). IEEE.
- [Erdős 59]: Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *science*, 286(5439), 509-512.
- [Fas 03]: Fassin, D. (2003). Le capital social, de la sociologie à l'épidémiologie. *Rev D'épidémiologie Santé Publique*, 51, 403-13.
- [Gou 98]: Gouarné, J. M. (1998). *Le projet décisionnel: enjeux, modèles et architectures du Data Warehouse*. Eyrolles.
- [Kim 05]: Kimball, R. (2005). *Le data warehouse: Guide de conduite de projet*. Eyrolles.
- [kim96] Kimball, R. (1996). The data warehouse toolkit: practical techniques for building dimensional data warehouse. *John Wiley&sons, New York*.
- [La 2014] *Réseaux sociaux et structures relationnelles: Que sais-je?» n° 3399*. Presses universitaires de France.
- [Rob 03]: Robert, C., & Cogis, O. (2003). Théorie des graphes. *Au-delà des ponts de Königsberg: problèmes, théorèmes, algorithmes*. Paris: Vuibert.
- [Ste et al. 00]: Stefanovic, N., Han, J., & Koperski, K. (2000). Object-based selective materialization for efficient implementation of spatial data cubes. IEEE Transactions on Knowledge and Data Engineering 12(6) 938-958

Webographie

- [1] http://fablab-robert-houdin.org/wiki/doku.php?id=tuto_installation_de_eclipse.
- [2] <https://netbeans.org>
- [3] [https://msdn.microsoft.com/fr-fr/library/6b1f4\(v=vs.90\).aspx](https://msdn.microsoft.com/fr-fr/library/6b1f4(v=vs.90).aspx).