

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ ABDELHAMID IBN BADIS - MOSTAGANEM



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

**RAPPORT DE MINI-PROJET**

Option : **Ingénierie des Systèmes d'Information**

**THEME :**

**CONCEPTION ET DEVELOPPEMENT D'UN ENTREPOT DE  
DONNEES DYNAMIQUE**

Etudiante : « **AMRANE HAFIDHA** »

« **BARDAD NEZHA** »

Encadrante : « **Dr.YOUNSI FATIMA ZOHRA** »

Année Universitaire 2020-2021

## **Résumé**

Les maladies infectieuses sont chaque année la principale cause d'une morbidité et d'une mortalité importante dans le monde entier. Malgré les dispositifs sanitaires construits, les programmes de lutte appliqués sur le terrain et la mise en œuvre de mesures de prévention, il n'y a finalement que très peu de maladies infectieuses éradiquées.

L'objectif de ce mini-projet est, tout d'abord, décrire les notions de base sur la théorie des graphes et le concept des réseaux sociaux. Nous allons, également, fournir une présentation générale sur les types de réseaux les plus célèbres et expliquer l'analyse des réseaux sociaux. Ensuite, nous décrivons les Systèmes D'information Décisionnels (SID), en particulier, les entrepôts de données spatiotemporelles et dynamique permettant le stockage des différents types de données (de type graphe, spatiales, thématiques et temporelles) et son moteur OLAP. Enfin, nous présentons quelques extensions d'OLAP, à savoir : Spatial OLAP, G-OLAP.

**Mots-clés :** Analyse multidimensionnelle, Entrepôt de données spatiotemporelle et dynamique, OnLine Analytical Processing (OLAP), Réseaux sociaux, Système D'information Décisionnel(SID), Théories des graphes, Les maladies infectieuses.

## **Abstract:**

Infectious diseases are the leading cause of significant morbidity and mortality each year around the world. Despite the health systems built, the control programs applied in the field and the implementation of preventive measures, there are ultimately very few infectious diseases eradicated.

The aim of this project is first, describing the basics of graph theory and the concept of social networks. We will also provide a general overview of the most famous kinds of networks and explain the analysis of social networks. Next, we describe Decisional Information Systems; in particular, spatiotemporal and dynamic data warehouses allowing the storage of multiple data (graph, spatial, thematic and temporal) and its OLAP engine. Finally, we present some extensions of OLAP, namely: Spatial OLAP, G-OLAP.

**Keywords:** Multidimensional analysis, Spatiotemporal and dynamic data warehouse, OLAP, Social networks, Decisional Information System, Graph theories, Infectious diseases.

## Liste des figures

<b>Numéro de la figure</b>	<b>Titre de la figure</b>	<b>Page</b>
Figure 1	Exemple d'un graphe non orienté	6
Figure 2	Exemple d'un graphe orienté	6
Figure 3	Exemple d'un sous graphe	7
Figure 4	Exemple de graphe aléatoire	8
Figure 5	Exemple de graphe régulier	9
Figure 6	Exemple de graphe petit-monde(Erdős et Rényi)	9
Figure 7	Exemple de graphe sans échelle(Barabasi et Albert)	10
Figure 8	Les principaux composants d'un SID	13
Figure 9	Les principaux composants d'un ED	15
Figure 10	Schéma en étoile	15
Figure 11	Schéma en flocon de neige	16
Figure 12	Schéma en constellation	16
Figure 13	Le processus ETL	17
Figure 14	Architecture du système ROLAP	18
Figure 15	Architecture du système MOLAP	18
Figure 16	Architecture du système HOLAP	19
Figure 17	Processus ETL	23
Figure 18	la présentation de Microsoft SSIS, SSAS, SSRS	24
Figure 19	Schéma du modèle multidimensionnel	27
Figure 20	Fenêtre principale de SQL Server Management Studio	31
Figure 21	Fenêtre création de la base de données	32
Figure 22	Création d'une table	32
Figure 23	Exemple d'une table patient	33
Figure 25	La création d'un nouveau projet	33
Figure 26	Ajouter nouvelle source de données	34
Figure 27	Ajouter une vue des sources de données	34
Figure 28	l'interface d'accueil	35
Figure 29	Fenêtre de la table patient	36
Figure 30	l'interface des résultats	36

## Liste des abréviations

<b>Abréviation</b>	<b>Expression Complète</b>	<b>Page</b>
SID	Système D'information Décisionnel	12
OLAP	On-Ligne-Analytical Processing	13
SQL	Structured Query Language	13
ED	Entrepot de Données	13
BI	Business Intelligence	14
BDD	Base De Données	14
ETL	Extract Transform Load	17
SGBD	Système de gestion de base de données	17
ERP	Entreprise Resource Planning	17
ROLAP	Relational On-Ligne-Analytical Processing	17
SGBDR	Système de Gestion de Base de Données Relationnelle	17
MOLAP	Multidimensional On-Ligne-Analytical Processing	18
SGBDM	Système de Gestion de Base de Données Multidimensionnelle	18
HOLAP	Hybride On-Ligne-Analytical Processing	18
SOLAP	Spatial On-Ligne-Analytical Processing	19
SIG	Système D'information Géographique	20
GOLAP	Graph On-Ligne-Analytical Processing	20
SSAS	SQL Server Analysis Service	24
SSIS	SQL Server integration service	24
SSRS	SQL Server Reporting service	24
XML	eXtended Markup Language	31
HTML	Hyper Text Markup Language	31
VMS	Virtual Memory System	31

## Table des matières

Introduction générale.....	1
Chapitre 1 .....	3
Les réseaux complexes.....	3
1 Introduction .....	3
2 Concepts et définitions de bases.....	3
2.1 Que ce qu'un réseau ?.....	3
2.2 Que ce qu'un graphe ? .....	3
2.3 Type des graphes .....	3
2.3.1 Graphe non orienté .....	3
2.3.2 Graphe orienté .....	4
2.3.3 Sous-graphe .....	4
2.4 Les graphes statiques et les graphes dynamiques .....	5
2.4.1 Graphe statique : .....	5
2.4.2 Graphe dynamique : .....	5
3 Les différentes classes de réseaux .....	5
3.1 Les réseaux aléatoires .....	6
3.2 Les réseaux réguliers .....	6
3.3 Les réseaux petit-monde .....	7
3.4 Les réseaux sans échelle .....	8
4 L'analyse structurelle des réseaux.....	8
4.1 La mesure des réseaux .....	8
4.1.1 Diamètre : .....	8
4.1.2 Longueur de chemin:.....	8
4.1.3 Densité : .....	9
4.1.4 Degrés : .....	9
4.1.5 Distribution de degrés : .....	9
4.1.6 Coefficient de clustering: .....	9

5	L'épidémiologie par réseaux de contacts : Travaux connexes .....	9
6	Conclusion.....	10
Chapitre 2 .....		11
Les Systèmes d'Information Décisionnels (SID).....		11
1	Introduction .....	11
2	Qu'est-ce qu'un Système d'Information Décisionnel SID.....	11
3	L'architecture d'un système d'information décisionnel (SID) .....	11
3.1	Les sources de données.....	12
3.2	Entrepôts et magasins de données .....	12
3.2.1	L'entrepôt de données .....	12
3.2.2	Les magasins de données .....	12
3.3	Serveurs OLAP .....	12
3.4	Les outils d'analyse .....	12
4	Entrepôts de données (DataWarehouse) .....	13
4.1	Données Thématiques (orientées sujet).....	13
4.2	Données Intégrées.....	13
4.3	Données Historiées .....	13
4.4	Données Non volatiles.....	13
5	Les objectifs d'un ED.....	13
6	Les composants d'un ED.....	13
7	Notions de base .....	14
7.1	Table de fait .....	14
7.2	Mesure .....	14
7.3	Table de dimensions .....	14
7.4	Hiérarchie .....	14
8	Différents modèles de la modélisation dimensionnelle.....	14
8.2	Le modèle de données en flocons (schéma en flocons de neige) .....	15
8.3	Le modèle en constellation (schéma en constellation) .....	15

8.4	Le modèle mixte (schéma mixte) .....	16
9	L'intégration des données dans l'entrepôt de données.....	16
10	Le cube OLAP (On-Line Analytical Processing) .....	17
10.1	Les architectures OLAP.....	17
10.1.1	L'approche ROLAP (Relational OLAP).....	17
10.1.2	L'approche MOLAP (Multidimensionnel OLAP).....	17
10.1.3	L'approche HOLAP (Hybride OLAP).....	18
11	Spatial OLAP (SOLAP).....	18
11.1	Les types de dimensions spatiales.....	19
11.1.1	Descriptive .....	19
11.1.2	Géométrique .....	19
11.1.3	Mixte .....	19
11.2	Les outils SOLAP .....	19
11.2.1	OLAP dominant (business Objects, Cognos, Knoys) .....	19
11.2.2	SIG dominant .....	19
11.2.3	SOLAP intégré .....	19
12	Les graphes et les cubes de données .....	19
12.1	Définition d'un Graphe OLAP (G-OLAP) .....	19
12.2	L'utilisation de G-OLAP .....	20
13	Conclusion.....	20
Chapitre 3.....		21
Conception et Modélisation .....		21
1	Introduction .....	21
2	Le but visés.....	21
3	Qu'est-ce qu'un ETL ? .....	21
4	Description du modèle proposé.....	21
4.1	L'extraction .....	22
4.2	Transformation .....	22

4.3	Le chargement .....	22
5	Pourquoi intégrer un ETL dans votre architecture de données ? .....	23
6	Les analyses services (SSAS, SSIS, SSRS) .....	24
6.1	SSAS.....	24
6.2	SSIS .....	24
6.3	SSRS.....	24
7	Les taches et étapes de l'ETL.....	25
8	Avantages et Inconvénients de processus ETL .....	25
9	Modélisation multidimensionnelle de l'ED .....	26
9.1	Identification des faits : .....	26
9.2	La table de fait : .....	26
9.3	Les dimensions (axes) et les hiérarchies.....	26
10	Cube de données OLAP .....	27
11	Cube de données OLAP multidimensionnel .....	27
12	Schéma multidimensionnel OLAP.....	28
12.1	Les opérateurs d'analyse OLAP .....	28
	• Roll-up:.....	28
	• Drill-down: .....	28
13	Conclusion.....	29
	Chapitre 4 .....	30
	Implémentation et Mise en œuvre .....	30
1	Introduction .....	30
2	Les outils de développements utilisés .....	30
3	Microsoft SQL Server Management Studio 2014.....	30
4	Microsoft Visual Studio 2019 .....	30
5	Java Netbeans .....	31
6	Création de la base de données .....	31
7	Extraction du cube de données.....	33



8	Conclusion.....	37
	Conclusion générale .....	38
	Bibliographie.....	39

## Introduction générale

Dans l'entreprise le système d'information a pour objectif de faciliter l'établissement et la mise en œuvre de la stratégie, en particulier de concrètement supporter la réalisation des activités. Il est construit à partir des exigences des métiers, des processus définis par l'entreprise, et est constitué de l'ensemble des moyens (humains, logiciels, matériels) utilisés pour collecter, stocker, traiter et communiquer les informations, un entrepôt de données fait de même but comme support d'aide à la décision.

Parmi les meilleures technologies d'analyse, nous retrouvons la technologie OLAP. OLAP est un type d'application informatique orienté vers l'analyse sur le champ d'informations selon plusieurs axes, il est couramment utilisé pour l'analyse multidimensionnelle.

En effet, l'analyse en ligne (OLAP) a pour le but d'obtenir des rapports de synthèse tels que ceux utilisés en analyse financière. Les applications d'OLAP sont couramment utilisées en informatique décisionnelle. Dans ce travail de mini-projet, nous nous intéressons à la présentation des différents outils et techniques que nous allons utiliser dans notre projet de fin d'étude. Nous allons donner une vue d'ensemble des principaux éléments de chaque concept utilisé.

### Organisation du mémoire

Ce mémoire comporte deux chapitres :

**Le premier chapitre** est dédié à un état de l'art sur la théorie des graphes et les modèles de réseaux, et dans un second temps, nous passons au sein aux travaux connexes.

**Le deuxième chapitre** fournit une présentation générale des SID. Nous discutons, tout d'abord, les notions qui se rapportent aux SID et leurs différentes architectures ainsi que la modélisation multidimensionnelle. Ensuite, nous présentons le moteur OLAP et ses extensions.

**Le troisième chapitre** donne une vue générale sur l'analyse multidimensionnelle dans le monde du business intelligence, ensuite fonctionne une hiérarchie concernant notre projet avec le Cube OLAP.

**Le quatrième chapitre** traite la mise en œuvre de sujet demandé dans une espace pratique. Au sein de plusieurs fonctionnalités.



# Chapitre 1

## Les réseaux complexes

### 1 Introduction

Les réseaux sociaux permettent d'extraire des nouvelles connaissances, d'une manière synthétique les travaux sur la détection de communautés qu'elles soient globales ou locales, s'appuie relativement à la classification automatique et la théorie des graphes.

Dans ce chapitre, nous allons introduire les graphes et le langage de base avec certaines notions élémentaires sur la théorie des graphes et les réseaux.

### 2 Concepts et définitions de bases

#### 2.1 Que ce qu'un réseau ?

Un réseau est défini par un graphe qui représente un ensemble d'entités (hommes, machines, animaux, cellules ...), que l'on appelle nœuds ou sommets, reliées entre elles par un ensemble de connexions appelées arcs ou arête. Les nœuds et les liens présentent des caractéristiques quantitatives ou qualitatives.

Généralement un réseau est défini par un graphe  $G=(X, U)$  dans lequel  $X$  est l'ensemble des nœuds et  $U$  l'ensemble des liens du réseau.

#### 2.2 Que ce qu'un graphe ?

Les graphes sont des outils permettent de représenter graphiquement les connexions qui existent dans un réseau qu'ils sont évités la complexité de la représentation de tous les liens pouvant exister entre les utilisateurs ,utilisés pour le transfert d'information dans un programme, les architectures de réseau ou les systèmes routiers et de communication [1].

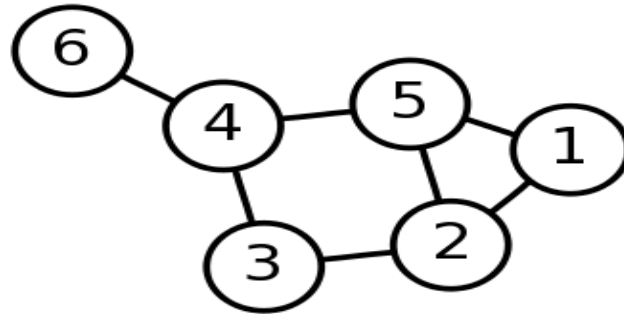
#### 2.3 Type des graphes

Dans la théorie des graphes, on distingue deux types de graphe :

##### 2.3.1 Graphe non orienté

Si on définit une relation sur un ensemble ou la notion d'ordre n'est pas importante, on représente ainsi la relation entre sommets par un lien non orienté appelée arête. On obtient

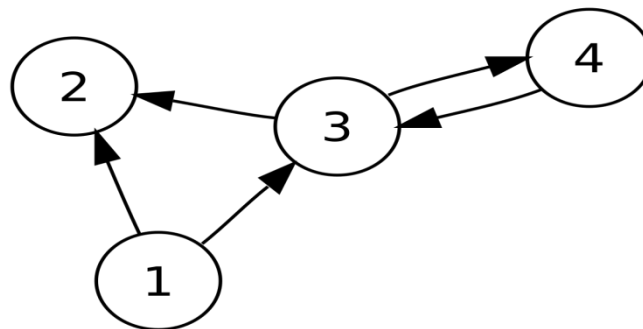
alors un graphe non orienté, noté  $G=(X, U)$ , la figure 1 représente un exemple d'un graphe non orienté.



**Figure 1** - Exemple d'un graphe non orienté

### 2.3.2 Graphe orienté

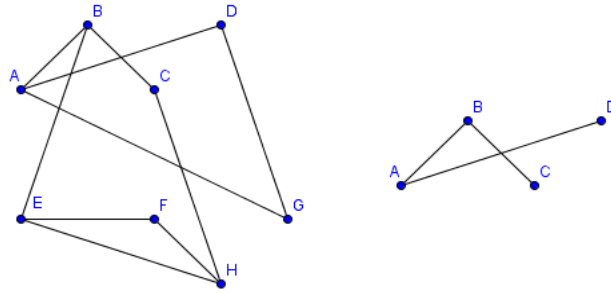
Un graphe orienté est représenté par le couple  $G=(X, U)$ , où :  
 $X = \{x_1, x_2, \dots, x_n\}$  est l'ensemble fini de  $n$  (entier positif) sommets.  
 $U = \{u_1, u_2, \dots, u_m\}$  est l'ensemble  $m$  (entier positif) arcs reliant dans un ordre bien défini ces sommets, ou un certain nombre d'entre eux, la figure 2 représente un exemple d'un graphe orienté.



**Figure 2** - Exemple d'un graphe orienté

### 2.3.3 Sous-graphe

Un sous graphe engendré par  $X_1$  est un graphe  $G_{x_1} = (X_1, U(X_1))$  tel que  $X_1$  est un sous ensemble de  $X$ , et  $U(X_1)$  sont les liens (arêtes ou arcs) induites par  $U$  sur  $X_1$ , c'est-à-dire les liens de  $U$  dont les deux extrémités sont des sommets de  $X_1$ , la figure 3 représente un exemple d'un sous graphe.



**Figure 3** - Exemple d'un sous graphe

## 2.4 Les graphes statiques et les graphes dynamiques

Nous distinguons deux grandes familles de graphes :

**2.4.1 Graphe statique :** les graphes statiques sont utilisés en particulier pour visualiser des systèmes et des situations complexes de façon statique indépendamment du temps, l'approche statique s'intéresse surtout à l'état actuel d'un réseau et aux structures communautaires qui tiennent actuellement place [2].

**2.4.2 Graphe dynamique :** est correspondant au graphe statique, ce graphe peut modéliser l'interaction entre les composants d'un ensemble et peut aussi évoluer dans le temps, suivant l'état du composant et son influence sur les autres composants. Comme exemple de graphe dynamique, on peut citer les graphes complexes, les graphes évolutifs et les graphes dédiés à la recherche opérationnelle [2]. En effet, ces graphes ou réseaux sont définis comme étant *dynamiques, temporels, variables* dans le temps ou encore *évolutifs*.

## 3 Les différentes classes de réseaux

La théorie des réseaux étudie les graphes en tant que la représentation des relations symétriques ou asymétrique entre des objets discrètes. Depuis le 17<sup>ème</sup> siècle Leonhard Euler découvrit un problème des sept ponts de Königsberg concernant l'étude des graphes, mathématiquement les sommets et les arêtes fut le fondement de la théorie des graphes. Dont les quels les terres accessibles par un nœud et les ponts par la connexion entre les nœuds .

L'utilisation de graphes comme modèles de système réels a permis de développer une nouvelle approche de l'épidémiologie par réseau de contacts, le but de cette méthode est de modéliser un réseau à l'aide d'un réseau complexe en analysant la topologie et tester des différentes stratégies de prévention et d'intervention [3].

Dans la suite, nous allons étudier les différentes classes de réseaux complexes afin de pouvoir identifier les aspects théoriques et méthodologiques de chacun. En effet, la plupart

des réseaux être généralisées, tout d'abord nous commençons par le type graphes le plus exploité : les graphes aléatoires.

### 3.1 Les réseaux aléatoires

Le modèle Erdős–Rényi, nommé ainsi en hommage à Paul Erdős et Alfréd Rényi est utilisé pour générer un graphe aléatoire dans lequel les arêtes sont définies entre les nœuds avec des probabilités égales. Il peut être employé dans une approche probabiliste afin de démontrer l'existence de graphes satisfaisant à une variété de propriétés, ou encore à procurer une définition rigoureuse de ce qu'implique une propriété pour presque tous les graphes.

Ce graphe est composé de  $N$  nœuds, le principe est de relier aléatoirement chaque couple de nœuds  $(n_i, n_j)$  par un lien avec une probabilité  $p$ . ils proposent plusieurs version de leur modèle, il existe un réseau qui s'appelle  $G_{n,p}$ , tel que, chaque nœud représente par  $(N-1)$  voisin potentiels. Les degrés moyens définis par  $K = p(N-1) \approx PN$ .

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

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

$K$  désigne le nombre des liens.

Le modèle d'Erdős et Rényi réside dans l'apparition de la phase de transitions en fonction de  $K$ , ce qui provoque de la formation d'une composante géante à partir d'une valeur critique de  $K$ . parmi de inconvénients est la distribution homogène du degré de ses nœuds qui n'est pas la distribution hétérogène qui sont trouvées dans les réseaux réels, la figure 4 représente un exemple de graphe aléatoire.

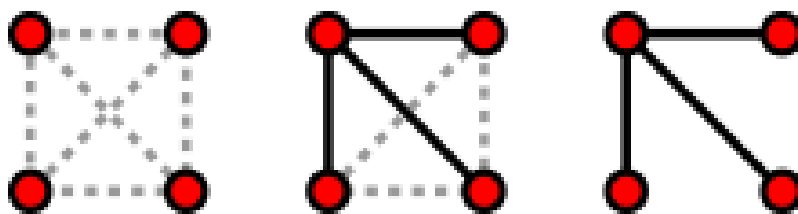
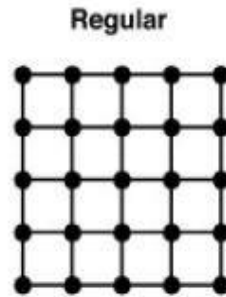


Figure 4 - Exemple de graphe aléatoire

### 3.2 Les réseaux réguliers

Sont des réseaux très ordonnées utilisée par des modèles d'automate cellulaire. Dans ce réseau chaque nœud a exactement le même nombre de lien. C'est à dire chaque nœud a 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,

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, la figure 5 représente un exemple de graphe régulier [3].



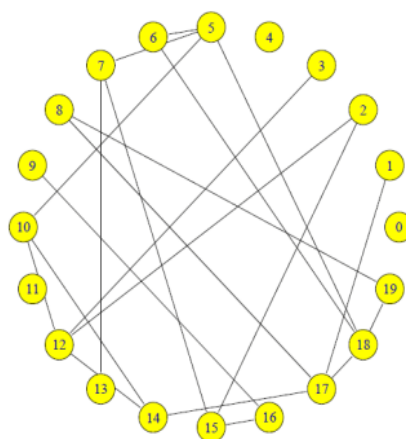
**Figure 5** - Exemple de graphe régulier

### 3.3 Les réseaux petit-monde

Le modèle de petit-monde est un modèle de génération de graphes aléatoires qui produit des graphes ayant les propriétés du petit-monde [4].

Un réseau initial est utilisé pour générer un réseau « petit-monde ». Chaque sommet d'un réseau est initialement lié à ces voisins  $K$  les plus proches. Un autre paramètre est spécifié comme probabilité de reconnexion, chaque arête a une probabilité  $p$  de se retrouver reconnectée au graphe en tant qu'arête aléatoire [3].

La figure 6 représente un exemple de graphe petit-monde [5].



**Figure 6** - Exemple de graphe petit-monde



### 3.4 Les réseaux sans échelle

Le modèle de Barabasi et Albert est un modèle de réseau aléatoire et sans échelle utilisé pour démontrer l'attachement préférentiel. Dans ce modèle, une arête a plus de probabilité de se lier à des sommets qui ont un plus haut degré qu'eux. Les deux chercheurs sont les premiers à introduire la notion des réseaux « sans échelle » ou « invariant d'échelle ». La distribution des liens des réseaux invariants d'échelle suit une loi de puissance ou la probabilité d'avoir  $K$  liens est :  $P(K) = k^{-a}$  où  $a$  est une constante, la figure 7 représente un exemple de graphe sans échelle (Barabasi et Albert) [3].

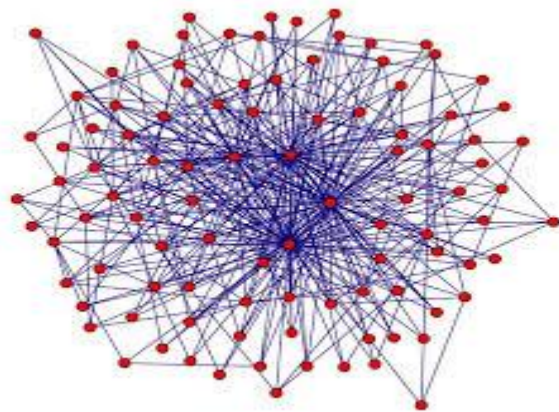


Figure 7 - Exemple de graphe sans échelle

## 4 L'analyse structurelle des réseaux

L'analyse des réseaux sociaux est une approche issue de la morphologie et la dynamique des réseaux sociaux. Elle se compose en trois phases : l'analyse du réseau, la visualisation du réseau et la récupération des données.

### 4.1 La mesure des réseaux

Les réseaux ont généralement des attributs qui peuvent être mesurés afin d'analyser leurs propriétés et caractéristiques. Les propriétés de réseau définissent souvent des modèles de réseau et peuvent être utilisées pour analyser le contraste de certains modèles, les principales mesures utilisées dans notre étude pour l'analyse des réseaux sont [6] :

**4.1.1 Diamètre :** est la distance la plus courte entre les deux nœuds les plus distants du réseau, une fois que la longueur de chemin la plus courte de chaque nœud à tous les autres nœuds est calculée, le diamètre est la plus grande de toutes les longueurs de chemin calculées.

**4.1.2 Longueur de chemin :** la distance moyenne du chemin reliant deux nœuds dans le réseau.

**4.1.3 Densité :** est la proportion de liens existants dans un graphe  $G$  par rapport au nombre totale de liens possible dont :  $d(G) = \frac{2m}{n(n-1)}$ , si  $m$  est de l'ordre de  $n$  on dit que le graphe  $G$  est creux.

**4.1.4 Degrés :** le degré  $K$  d'un nœud correspond au nombre d'arêtes qui lui sont connectées.

**4.1.5 Distribution de degrés :** est calculé comme suit :

$$P(K) = \frac{|L(v)|}{N}$$

$|L(v)|$  : représente le nombre du réseau  $G$  ayant le degré  $K$ .

$N$  : désigne le nombre de nœuds de  $G$ .

**4.1.6 Coefficient de clustering:** le coefficient de regroupement d'un nœud est le rapport des liens existants reliant les voisins d'un nœud au nombre maximal possible de tels liens.

## 5 L'épidémiologie par réseaux de contacts : Travaux connexes

Dans la littérature, il existe plusieurs travaux de recherche qui ont utilisé les concepts des réseaux de contact de type complexe, dans la suite nous citons les plus célèbres :

- **Les réseaux sociaux :** en ligne sont un bon exemple où Facebook peut être vu comme un graphe non orienté, puisque les "amitiés" sont bidirectionnelles, et Twitter quant à lui est un graphe orienté, puisque les "abonnements" sont à sens unique<sup>4</sup>.
- **Les réseaux d'information :** un exemple typique de ce type de réseau est le World Wide Web<sup>5</sup>, où les nœuds correspondent aux pages web contenant de l'information, et les liens sont les hyperliens permettant de naviguer d'une page à l'autre.
- **Les réseaux de technologie :** dans ce réseau, les ordinateurs et les routeurs sont les nœuds du réseau, et ces derniers sont connectés par des liens physiques comme la fibre optique modélisant les liens de ce réseau complexe.
- **Les réseaux biologiques :** Les réseaux complexes permettent également de représenter la majorité des systèmes biologiques. Parmi ces nombreux réseaux, nous pouvons citer les réseaux d'interaction protéine-protéine<sup>12</sup>, les réseaux de régulation des gènes<sup>13</sup>, les réseaux de signalisation ou encore les réseaux métaboliques<sup>14</sup>. Ils sont de ce fait très étudiés en biologie des réseaux et en bio-informatique. Dans un tel réseau, les nœuds seront les molécules biochimiques et les liens les réactions ayant permis de les obtenir [7].

## **6 Conclusion**

Dans ce chapitre, nous avons présenté des différentes notions du réseau et la théorie des graphes et en parallèle nous avons cité les différents types des réseaux complexes ainsi que leurs mesures.

Dans le chapitre suivant, nous allons mettre en place un Système D'information Décisionnel (SID) et spatiotemporel au service de l'épidémiologie.

# Chapitre 2

## Les Systèmes d'Information Décisionnels (SID)

### 1 Introduction

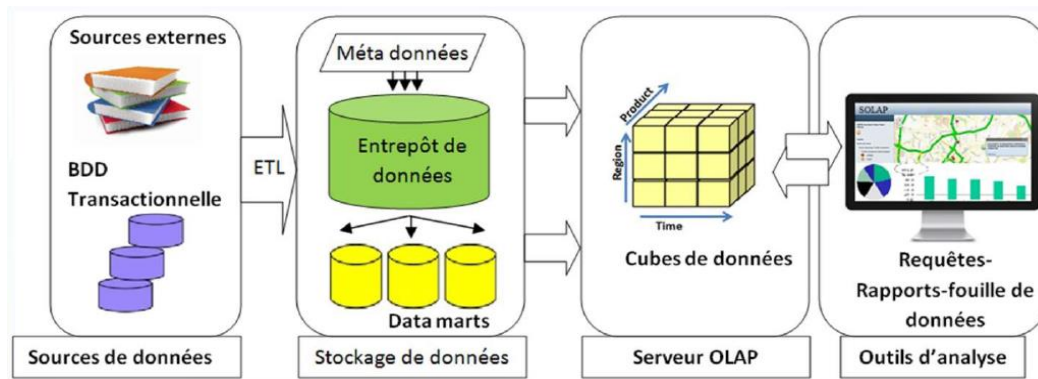
Dans ce chapitre, nous donnons un aperçu sur les notions élémentaires des Systèmes d'Information Décisionnels et sur leur moteur d'exploitation OLAP. Nous présentons l'architecture et les composantes principales d'un entrepôt de données ainsi que ses étapes de construction. Ensuite, nous basculons vers l'outil d'analyse et d'exploitation OLAP avec ses différentes architectures et l'aspect spatial dans les systèmes décisionnels.

### 2 Qu'est-ce qu'un Système d'Information Décisionnel SID

Un Système d'Information Décisionnel (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 »<sup>5</sup>. En effet, l'informatique décisionnelle ou business intelligence va permettre d'exploiter les données de l'entreprise en vue de faciliter la prise de décision par les décideurs, il désigne l'ensemble des technologies permettant aux entreprises. En ce sens il sera plus facile pour les dirigeants d'entreprise d'anticiper les actions à mettre en œuvre pour un pilotage éclairé de l'entreprise. Pour ce faire le business intelligence va tout en sécurisant l'accès à l'information, faciliter la prise de décision grâce à une meilleure qualité et cohérence des données [8].

### 3 L'architecture d'un système d'information décisionnel (SID)

Il n'existe pas de consensus sur une typologie d'architectures pour les SID. L'architecture type d'un système décisionnel basé sur l'entrepôt de données et l'OLAP, peut être représentée dans la Figure 1, elle est divisée en quatre niveaux :



**Figure 8 - Les principaux composants d'un SID**

L'architecture détaillée d'un SID, comme indiquée dans la Figure1, est composée des éléments suivants [9]:

### 3.1 Les sources de données

Les systèmes décisionnels peuvent avoir besoin de données issues de différentes sources et de différents formats de stockage. Ceci regroupe les fichiers texte, les rapports, les fichiers de base de données issus de différents systèmes de gestion de base de données, etc.

### 3.2 Entrepôts et magasins de données

#### 3.2.1 L'entrepôt de données

Est défini comme étant une collection de données orientées sujet, intégrées, non volatiles, historiées disponibles pour le support du processus de prise de décision.

#### 3.2.2 Les magasins de données

Est un sous-ensemble de l'entrepôt de données concernant un secteur particulier de l'entreprise.

### 3.3 Serveurs OLAP

Un système OLAP est une catégorie d'outils qui permet l'exploration interactive suivant une approche multidimensionnelle à plusieurs niveaux d'agrégation. Ceci peut se réaliser en exploitant un entrepôt de données. Il a pour but d'évaluer l'activité et l'aide à la décision au sein d'une organisation.

### 3.4 Les outils d'analyse

Représente l'élément le plus important pour l'utilisateur finale qui permet d'exploiter les données stockées. C'est l'élément qui correspond à la partie visible du système par rapport au décideur. On ce qui concerne les techniques de traitements des données, en distingue trois solutions : SQL, Server OLAP et Client OLAP.

## **4 Entrepôts de données (DataWarehouse)**

Inmon dans [Inmon, 1996] [10], définit l'entrepôt de données comme étant: « une collection de données thématiques, intégrées, historiées et non volatiles pour supporter le processus de prise de décision d'une organisation ».

### **4.1 Données Thématiques (orientées sujet)**

Les données sont organisées par sujet ou fait.

### **4.2 Données Intégrées**

Les données hétérogènes venues de différentes sources feront l'objet d'une intégration dans un seul espace de stockage.

### **4.3 Données Historiées**

Du fait que les données sont non volatiles, des intervalles de temps les sont associés.

### **4.4 Données Non volatiles**

Les données ne sont ni modifiables, ni supprimable.

## **5 Les objectifs d'un ED**

Parmi les objectifs d'un SID, nous citons [11] :

- Les données de l'organisation doivent être accessible facilement (données parlante, et signification évidente pas seulement pour le développeur mais surtout pour l'utilisateur).
- L'entrepôt de données doit être le socle sur lequel repose l'amélioration des prises de décisions : il doit contenir les données servant à étayer les décisions.
- La présentation des informations de manière cohérente : les données doivent être assemblées à partir des différentes sources de l'entreprise et il faut contrôler la qualité et éviter de les mettre à disposition quand elles ne sont pas nettoyées.
- L'entrepôt de données doit protéger notre richesse informationnelle : L'ED doit efficacement contrôler l'accès aux informations confidentielles de l'organisation.

## **6 Les composants d'un ED**

Dans un ED, les données sont lues soit à partir de bases de données opérationnelles soit à partir d'autres données internes ou de source de données externes par le système ETL. Ce système nettoie et prépare les données pour le traitement, cela peut être un processus complexe. Les données sont stockées dans l'ED pour les exploiter par la suite par les utilisateurs de BI qui accèdent aux données par divers outils. L'ED stocke les BDD et les

métadonnées définissant des objets, la figure 9 représente les principaux composants d'un ED. [11].

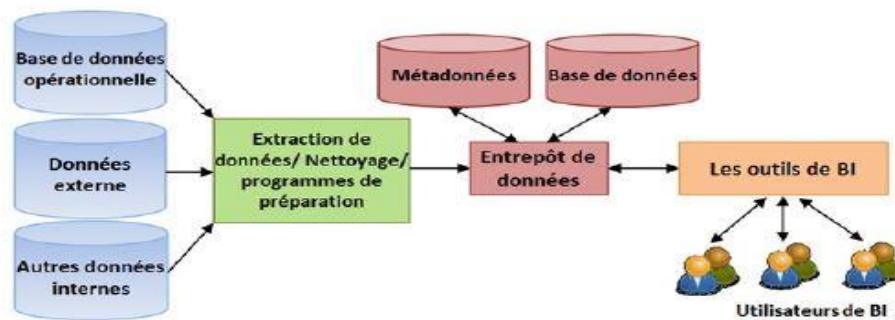


Figure 9 - Les principaux composants d'un ED

## 7 Notions de base

Dans ce qui suit, nous définissons quelques concepts de base liés à la modélisation multidimensionnelle :

### 7.1 Table de fait

C'est le concept qui modélise le sujet d'analyse. Il est formé des informations de l'activité analysée.

### 7.2 Mesure

C'est la valeur d'attribut quantitatif ou qualitatif qui évalue un fait.

### 7.3 Table de dimensions

Ensemble de paramètres qui peuvent faire varier les mesures. Elle modélise une perspective de l'analyse.

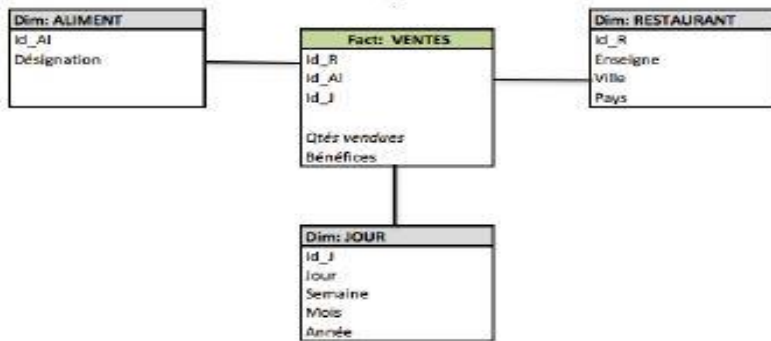
### 7.4 Hiérarchie

Permet d'organiser les membres d'une dimension selon leur niveau de détail.

## 8 Différents modèles de la modélisation dimensionnelle

### 8.1 Le modèle de données en étoile (schéma en étoile)

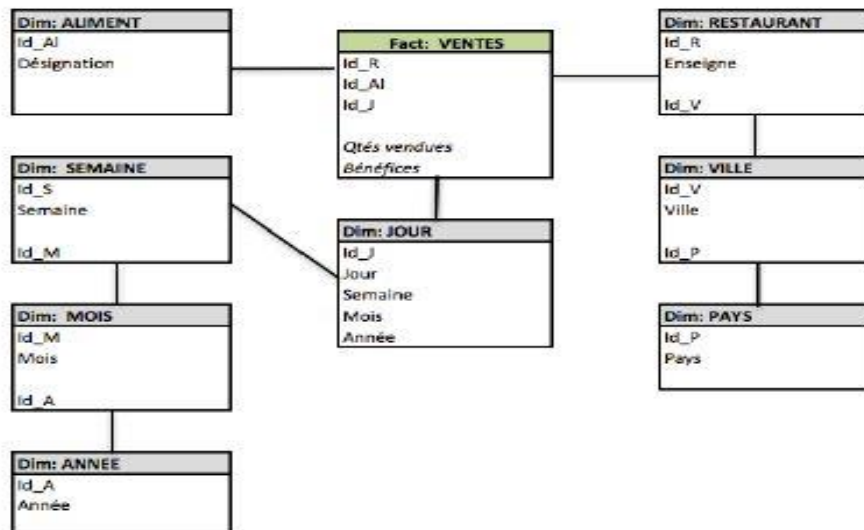
Un schéma conceptuel constitué du fait central et des dimensions représentées de manière dénormalisée. Les données dans un modèle en étoile sont dénormalisées pour permettre une meilleure performance des requêtes, la figure 10 représente le schéma en étoile.



**Figure 10 - Schéma en étoile**

### 8.2 Le modèle de données en flocons (schéma en flocons de neige)

Dans ce schéma la table de faits est conservée et les dimensions sont éclatées conformément à sa hiérarchie. Ce modèle permet d'éviter le problème de redondance qu'on peut trouver dans le modèle en étoile, la figure 11 représente le schéma en flocon de neige.

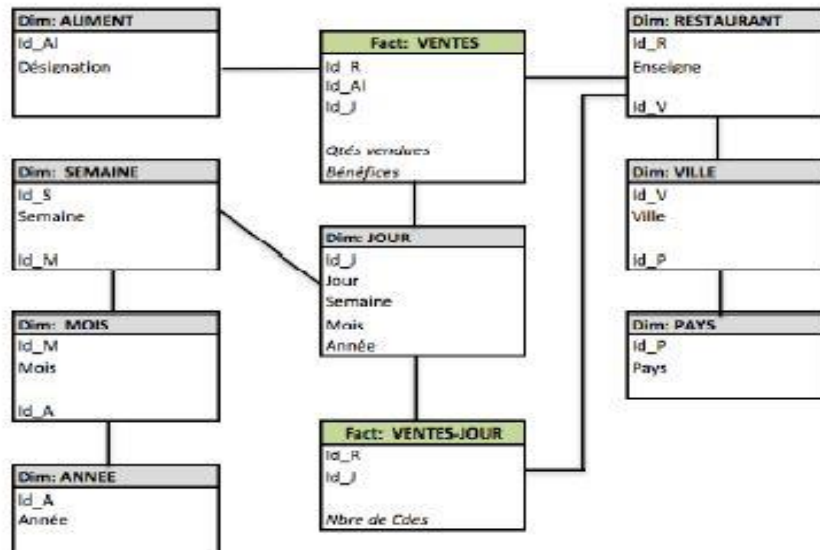


**Figure 11 - Schéma en flocon de neige**

### 8.3 Le modèle en constellation (schéma en constellation)

Cette technique consiste à fusionner plusieurs modèles en étoile utilisant des dimensions communes. Dans ce modèle, le partage des dimensions par plusieurs faits permet d'éviter de les définir plusieurs fois, la figure 12 représente le schéma en constellation.





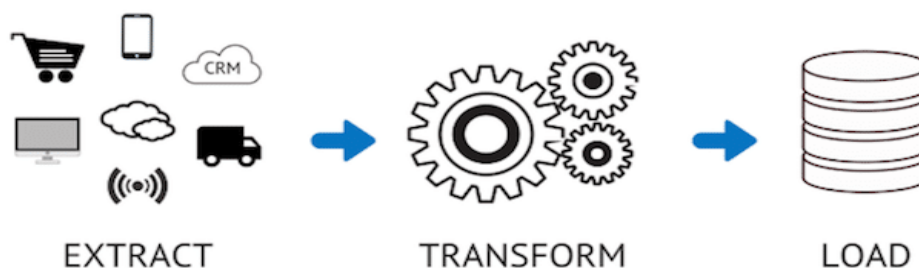
**Figure 12 - Schéma en constellation**

#### 8.4 Le modèle mixte (schéma mixte)

Cette technique consiste à fusionner plusieurs modèles en étoile et en flocon de neige dans un seul schéma.

### 9 L'intégration des données dans l'entrepôt de données

Les outils ETL gèrent toutes les étapes de la collecte des données au sein des systèmes d'information hétérogènes : SGBD, ERP, applications spécifiques, fichiers plats, bases hiérarchiques... depuis le nettoyage des données collectées, la consolidation et la mise en concordance des données éparées jusqu'à leur distribution auprès des applications cibles ou des systèmes décisionnels. Le processus ETL est une opération de migration de données qui consiste aussi à la rendre aisément consommable [Leblanc, 2012] [12], la figure 13 représente le processus ETL.



**Figure 13 - Le processus ETL**

## 10 Le cube OLAP (On-Line Analytical Processing)

Le concept « OLAP » est défini comme : «*Activité globale de requête et de présentation de données textuelles et numériques contenues dans l'entrepôt de données; Style d'interrogation spécifiquement dimensionnel* » [13].

OLAP propose une approche multidimensionnelle ce qui nous amène à la notion d'(hyper)cube. Un cube représente un ensemble de mesures organisées selon un ensemble de dimensions. Une dimension est un axe d'analyse c'est-à-dire une base sur laquelle seront analysées les données. Les applications OLAP sont des applications d'aide à la décision.

### 10.1 Les architectures OLAP

Dans la littérature des systèmes OLAP, il existe différentes implémentations, nous citons dans les sections suivantes, les trois implémentations les plus connues à savoir [11]:

#### 10.1.1 L'approche ROLAP (Relational OLAP)

Dans ce type d'implémentation un système de gestion de bases de données relationnelles (SGBDR) est utilisé pour le stockage des données. Le serveur ROLAP permet de simuler la vue multidimensionnelle et de communiquer entre le SGBDR et les outils client OLAP. L'architecture ROLAP permet une grande souplesse dans la gestion de gros volumes de données et une meilleure administration des données, la figure 14 représente l'architecture du système ROLAP.



**Figure 14 - Architecture du système ROLAP**

#### 10.1.2 L'approche MOLAP (Multidimensionnel OLAP)

dans ce type d'implémentation un système de gestion des bases de données multidimensionnelles (SGBDM) natifs est utilisé. Dans ce cas les données décisionnelles sont maintenues dans des structures multidimensionnelles. Les cellules des différents hyper cubes sont pré-calculées à partir de l'ED ou des bases sources et stockées dans ces structures. L'architecture MOLAP offre moins de performances que la précédente

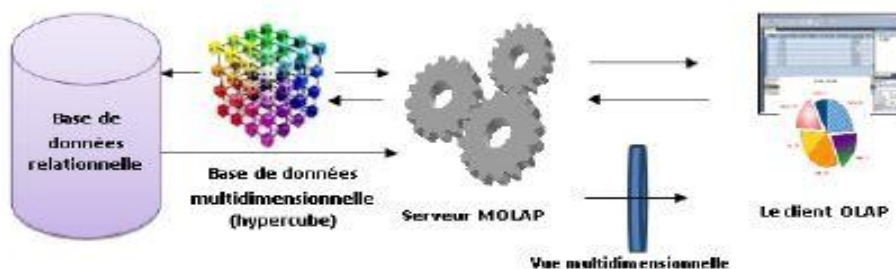
concernant l'administration de données à cause du pré-calcul des cellules, la figure 15 représente l'architecture du système MOLAP.



**Figure 15 - Architecture du système MOLAP**

### 10.1.3 L'approche HOLAP (Hybride OLAP)

ce type d'implémentation combine les deux technologies ROLAP et MOLAP. En effet, une partie des données dont les utilisateurs accèdent le plus fréquemment est stocké dans des structures multidimensionnelles gérées par un SGBDM. Une autre partie est stockée dans des structures relationnelles gérées par un SGBDR. L'architecture HOLAP combine les deux architectures ROLAP et MOLAP, la figure 16 représente l'architecture du système HOLAP.



**Figure 16 - Architecture du système HOLAP**

## 11 Spatial OLAP (SOLAP)

Le SOLAP est un système d'aide à la décision qui intègre les concepts et méthodologies issues à la fois de Système d'Information Géographique (SIG) et d'OLAP.

Le SOLAP enrichit les capacités d'analyse des systèmes OLAP classiques car la visualisation des mesures sur une carte permet de comprendre la distribution géographique d'un phénomène et de mettre en relation les différents phénomènes spatiaux par rapport aux axes d'analyse alpha numériques et de comparer ces phénomènes à diverses granularités

géographiques. De plus, la composante cartographique dans l'OLAP représente une interface vers l'entrepôt de données spatiales.

### **11.1 Les types de dimensions spatiales**

Est une dimension qui contient au moins un niveau d'agrégation spatial et permet à 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 :

#### **11.1.1 Descriptive**

Les membres de la dimension ont une référence spatiale textuelle (nom de lieu par exemple).

#### **11.1.2 Géométrique**

à tous les niveaux des hiérarchies d'une dimension, les membres ont une composante géométrique référencée dans l'espace.

#### **11.1.3 Mixte**

Combinaison des deux premières, certains membres comportent une composante géométrique, d'autres une références spatiale nominale seulement.

### **11.2 Les outils SOLAP**

Nous résumons les outils actuellement disponibles pour SOLAP. Ces outils peuvent être divisés en trois catégories différentes [14] :

#### **11.2.1 OLAP dominant (business Objects, Cognos, Knoys)**

La composante spatiale est accessoire ;

#### **11.2.2 SIG dominant**

La composante décisionnelle est accessoire ;

#### **11.2.3 SOLAP intégré**

Les deux composantes décisionnelles et spatiale sont fortement couplées, elles font appel autant aux fonctions OLAP que SIG.

## **12 Les graphes et les cubes de données**

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 types d'analyse.

### **12.1 Définition d'un Graphe OLAP (G-OLAP)**

L'analyse OLAP a été développée et utilisée dans un contexte de données assez classiques, souvent structurées dans des bases des données. Dans le paysage des données

complexes, les réseaux d'information constituent un type de données particulièrement riche. La modélisation sous forme de graphe avec des nœuds et les arêtes peut prendre différentes formes selon les besoins de représentation : graphe value ou non pour pondération des arcs, graphe homogène (un seul type de nœud) ou hétérogène, pour illustrer les réseaux d'information, les données bibliographiques se prêtent bien à la représentation sous forme de graphes.

## **12.2 L'utilisation de G-OLAP**

Dans les approches initiales de Graphe OLAP, au niveau de la modélisation, deux types de dimensions ont été définis : les dimensions informationnelles (OLAP traditionnel) qui propose différentes opérations de navigation dans les données, et les dimensions topologique quant à, elles, se rapportent à la modélisation des réseaux eux-mêmes dans les cellules vers un but de prendre un résultat qu'il s'agit d'un nouveau graphe. Initialement, la combinaison de l'OLAP et des graphes s'est donc faite au travers de plusieurs approches basées sur des cubes de graphes avec dans leurs cellules [15].

## **13 Conclusion**

Dans ce chapitre, nous avons présenté une approche théorique des systèmes d'information décisionnels. Initialement, nous avons introduit les concepts de base. Puis, nous avons décrit les principaux composants d'un SID ainsi que la description des différents termes dans le monde du SID. Finalement, nous avons abordé les concepts SOLAP et G-OLAP.

# Chapitre 3

## Conception et Modélisation

### 1 Introduction

Dans les deux derniers chapitres, nous avons couvert quelques concepts fondamentaux liés aux réseaux sociaux et au SID.

Par conséquent, ce chapitre présentera en détail les approches Solap et Golap, ainsi qu'un modèle de prise de décision pour aider à l'analyse de plusieurs ensembles de données.

### 2 Le but visés

Dans des autres objectifs nous décrivons quelques notions comme suit :

- Le premier objectifs de cette recherche est d'étudier les approches Golap et Solap avec la construction de cube OLAP ;
- Construire un entrepôt de données ;
- Développer un OLAP adapté pour visualiser et analyser les réseaux, tout en prenant leurs dynamiques ;
- Appliquer l'analyse multidimensionnelle des données [16].

### 3 Qu'est-ce qu'un ETL ?

Même si l'on connaît les différences entre ETL et EAI, il convient de rappeler ce qu'apporte un ETL à une architecture de Système d'Information. Ses tâches incluent la collecte de données provenant de diverses sources, leur traitement et leur chargement dans un entrepôt de données. Un EAI, en revanche, automatise le flux de données entre les applications [17].

### 4 Description du modèle proposé

Cette partie va poser une masse d'information sur notre modèle en détails, en prenons en considération les trois étapes qui on a les passer dont l'extraction de modèle comme premier, ensuite transformation des données du modèle, en dernière le chargement [18] :

## **4.1 L'extraction**

Un ETL doit pouvoir collecter tous type de données brutes, structurées, en temps réelle au par lots, à partir de diverses sources au sein du système d'information de l'entreprise, tant sur site que sur le Cloud.

## **4.2 Transformation**

Cette étape consiste à extraire toute les bases de données de l'entreprise (Oracle, SAP, MySQL, etc.). Cette opération nécessite l'utilisation de technologies variées selon le type de source de données :

- Vérifier les données ;
- Éliminer les doublons ;
- Nettoyer les données
- Standardiser selon les politiques de l'entreprise.

L'objectif est de fournir des données propres qui peuvent être placées dans l'entrepôt de données et utilisées comme source de solutions analytiques.

## **4.3 Le chargement**

Après cette étape, les informations stockées sont facilement accessibles et utilisables :

- Les outils OLAP sont utilisés pour analyser les données ;
- L'exploration de données est utilisée pour rechercher des corrélations ;
- Participer à l'ensemble d'outils de Reporting ;
- Des outils de Business Intelligence sont utilisés pour exploiter les données ;

Ces trois étapes, extraction de la charge de transformation, sont fréquemment effectuées en parallèle pour des avantages en termes de performances. La figure 17 représente le processus ETL.

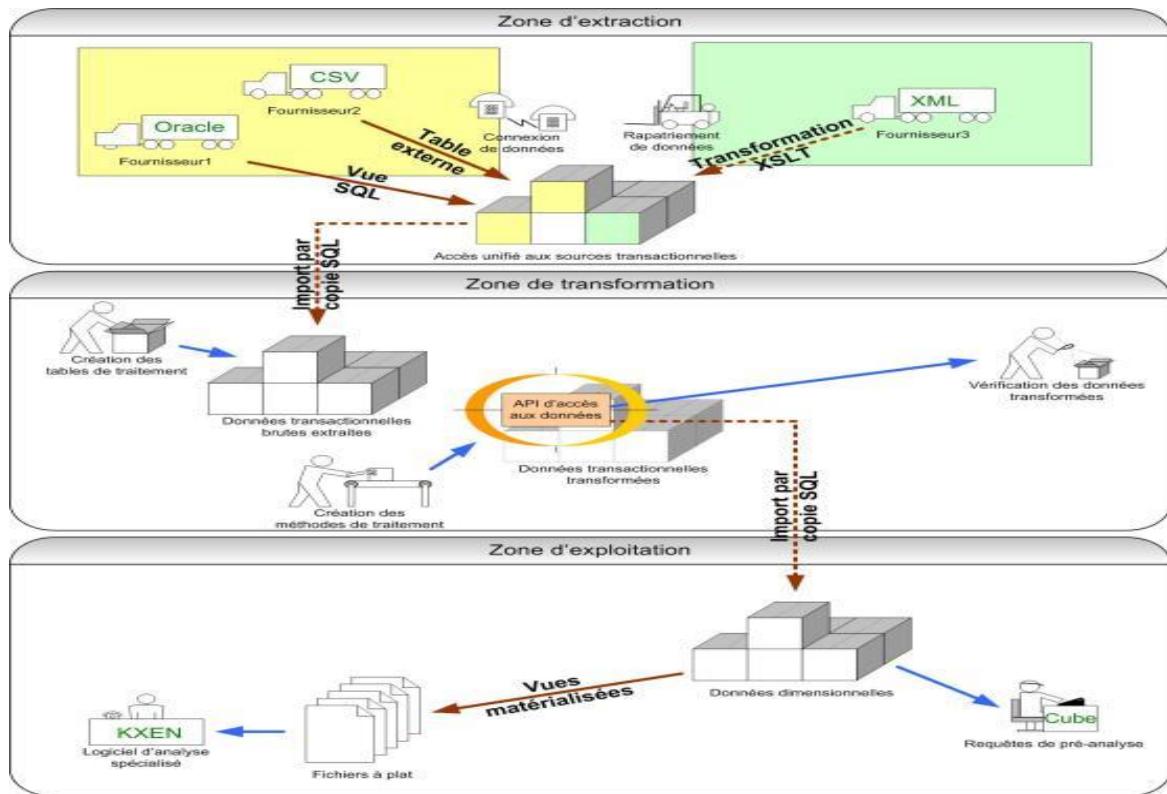


Figure 17– Processus ETL

## 5 Pourquoi intégrer un ETL dans votre architecture de données ?

Les systèmes d'information d'entreprise sont de nature hétérogène, avec plusieurs sources de données, différents niveaux de génération, ERP, CRM, gestion de production, et autres programmes. Toutes ces bases de données ont été optimisées pour des transactions aux vocations et applications variées.

La solution pour L'architecture ETL est la collectivité de toutes les données pertinentes de l'entreprise dans un DataWarehouse ou entrepôt de données. L'outil idéal pour améliorer la qualité des données [19].



## 6 Les analyses services (SSAS, SSIS, SSRS)

Microsoft SQL Server offre une variété d'outils pour gérer et analyser les données avec ces outils [20] :

### 6.1 SSAS

Est le service d'analyse de Microsoft SQL Server, qui est un outil OLAP (traitement analytique en ligne), d'exploration de données et de création de rapports utilisé dans Business Intelligence pour faire fonctionner les données .

### 6.2 SSIS

Lors de la construction d'un entrepôt de données ou d'un magasin de données, les données doivent être extraites de divers systèmes transactionnels et fichiers plats, transformées et chargées dans une base de données où elles peuvent être analysées et rapportées.

### 6.3 SSRS

Est un acronyme pour SQL Server Reporting Services. Une fois que les données sont dans leur état final, que ce soit dans un système transactionnel local, un Data-Mart ou un entrepôt de données, SSRS fournit les outils dont vous avez besoin pour créer des rapports afin de mieux comprendre les données.

Ces trois outils, qui font partie de Microsoft SQL Server et du package Microsoft Business Intelligence, sont fréquemment utilisés ensemble pour gérer les besoins d'analyse de données. La figure représente la présentation de Microsoft SSIS, SSAS et SSRS. La figure 18 représente les fonctionnalités de Microsoft SSIS, SSAS, SSRS.

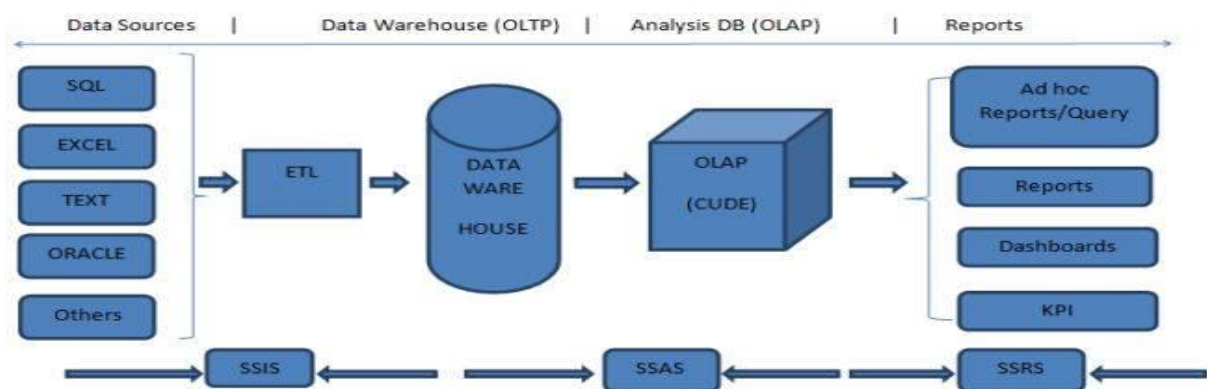
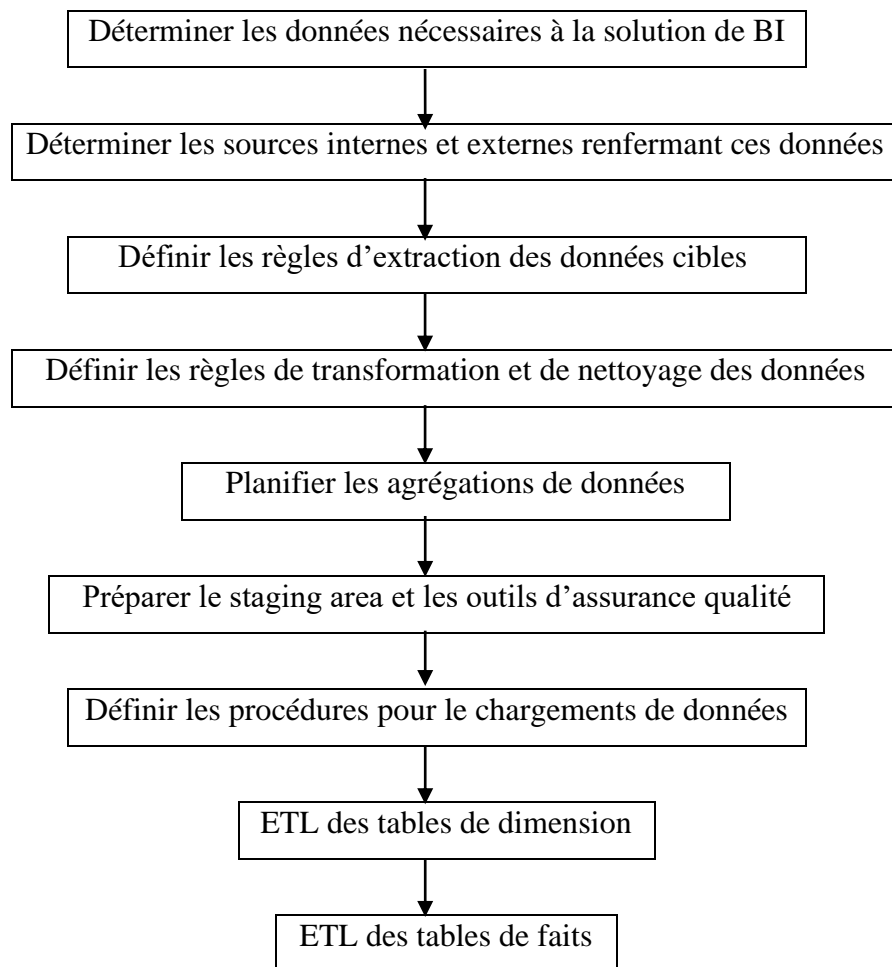


Figure 18 - la présentation de Microsoft SSIS, SSAS, SSRS

## 7 Les tâches et étapes de l'ETL



## 8 Avantages et Inconvénients de processus ETL

- Processus de développement long et couteux ;
- Permet la réutilisation des processus et transformations (ex : packages dans SSIS) ;
- Gestion des chargements nécessaire ;
- Peut traiter de grandes quantités de données dans une même exécution (traitement en lot) ;
- Optimiser la structure de l'entrepôt de données ;
- Latence des données entre la source et le centre de données ;
- Exécuté indépendamment du besoin réel [21].

## 9 Modélisation multidimensionnelle de l'ED

Le tout en présentant les vocabulaires communs utilisés dans les systèmes d'aide à la décision, ce qui permet l'analyse de données dans de gros volumes de données.

Le traitement analytique avec un modèle multidimensionnel est soutenu par un portefeuille de données ED, qui est structuré comme suit :

- **Un fait** : La clé de voute de modélisation est identifiée par les mesures, qui sont aussi des métriques ; une mesure donnée peut être calculée à partir d'une autre mesure ou à partir des propriétés des membres d'une dimension.

- **Dimensions** : organisées d'une façon hiérarchique.

**9.1 Identification des faits** : L'utilisation de modèle en étoile demande une table de fait.

**9.2 La table de fait** : Medical\_Socioeconomics\_Statistics

Comprenant trois dimensions, fait des analyses sur les mesures suivant:

- nb\_Patient\_Health\_History: le nombre des personnes (patients) avec des tares.
- nb\_Contact\_Health\_History: le nombre des personnes (contacts) avec des tares.
- nb Incidence: le totale des infectés.
- nb\_positif\_contact: le nombre des non infectés.
- nb\_Personnal\_relationship: le nombre des relation personnel dans un réseau social.
- nb\_Professionnal\_relationship: le nombre des relations professionnelles dans un réseau social.

**9.3 Les dimensions (axes) et les hiérarchies**

La structure de base de notre application multidimensionnelle est les dimensions ;

Les dimensions sont organisées d'une façon hiérarchique toute entière et toutes les hiérarchies s'il y a plusieurs, comme une seule et même dimension.

Axes d'analyse thématiques: caractérisée par les dimensions suivantes:

Age\_group: 0-5ans, 6-18ans, 19-59ans, 60+.

Gender: F (féminin), M(masculin).

Age: 0-98ans.

Axes d'analyse temporelle: comporte cinq dimensions de temps (Date, Week, Month, Quarter, Year).

Axes d'analyse spatiale géométrique: caractérisé par la présence d'attribut géométrique (Neighbourhond, Wilaya, Commune), on trouve un ensemble des quartiers (49) au plusieurs communes de Est-Ouest-Nord dans la wilaya de Oran.

Axes d'analyse graphe: contient les dimensions suivantes (Graph, Patient, Contact, Social-Network, Relationship). La figure 19 représente le schéma du modèle multidimensionnel.

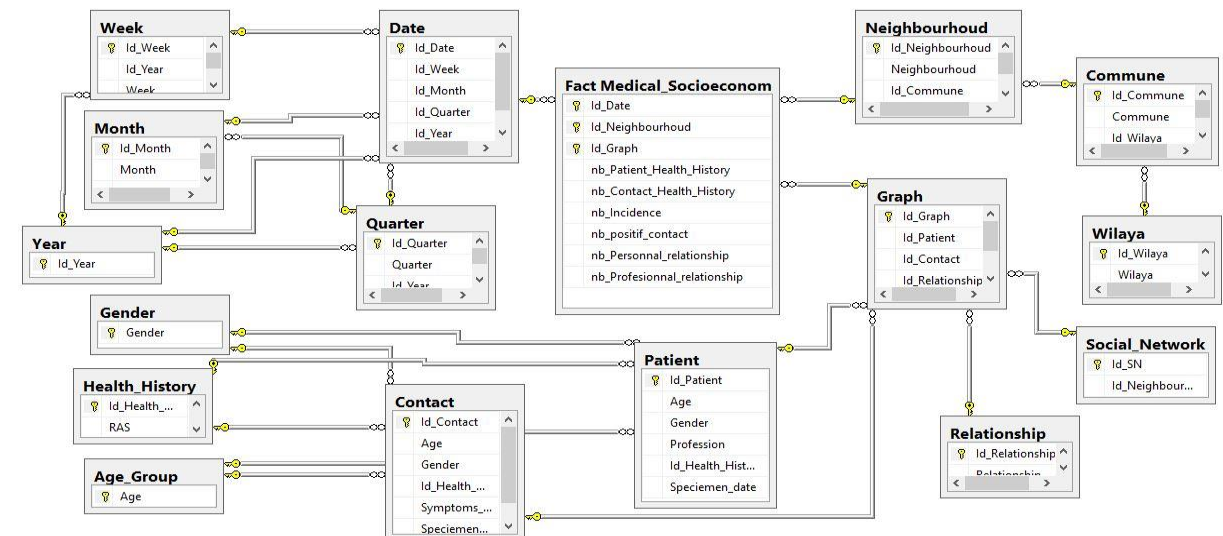


Figure 19– Schéma du modèle multidimensionnel

## 10 Cube de données OLAP

La disposition résultante de ces tables de données crée une structure d'information communément appelée « cube de données » ou « cube OLAP ». Ce cube de données est dit multidimensionnel car vous disposez de plusieurs axes d'analyse grâce à vos tables de dimensions : graph, date, etc [22].

## 11 Cube de données OLAP multidimensionnel

Une fois vos données uniformisées et stockées dans un entrepôt de données ou un cube OLAP multidimensionnel, elles sont prêtes pour l'analyse, l'évaluation et la présentation finale aux utilisateurs. Les cubes de données OLAP, par exemple, sont conçus pour stocker un grand nombre de lignes de données pouvant être lues en quelques secondes. Le cube est configuré en forme d'étoile, avec une table de faits liée aux tables de dimension par des relations de clé étrangère. En conséquence, il est possible d'interroger le cube pour les mesures et indicateurs clés. Ce projet concerne la modélisation de cubes. Il existe de nombreuses options logicielles de stockage de données, telles que l'analyse du serveur Microsoft SSAS SQL ou la base de données Oracle [23].

## 12 Schéma multidimensionnel OLAP

OLAP est défini comme une activité complète de récupération et de présentation de données textuelles et numériques stockées dans un référentiel de données. [Kimball et al. 2005] R.KimBall En fait, il permet la création de cubes de données pour exécuter l'analyse des données de manière plus efficace. Le traitement basé sur OLAP est fréquemment utilisé pour l'exploration de données, qui vise à découvrir de nouvelles relations entre divers ensembles de données. Tous les outils associés à l'exploration de données sont optimisés pour les requêtes et les rapports. Cela contraste avec les applications OLTP, qui sont principalement concernées par les tâches basées sur les transactions. Les outils OLAP permettent aux utilisateurs d'analyser des données multidimensionnelles de manière interactive en utilisant de nombreuses vues [24].

### 12.1 Les opérateurs d'analyse OLAP

Le but des systèmes OLAP est de représenter les intersections entre ces plusieurs dimensions en représentant des données sous forme de cube ou d'hypercube, les informations contenues dans le cube, et en passant d'un niveau d'agrégation à un autre afin de comprendre les détails des données.

Les analyses peuvent ensuite effectuer cinq types d'opérations de manipulation d'un cube à partir de ces bases de données [25] :

- **Roll-up:** effectue l'agrégation sur un cube de données soit en montant dans la hiérarchie, soit en réduisant les dimensions.

- **Drill-down:** cette opération permet d'obtenir plus de détails soit en descendant dans la hiérarchie, soit en ajoutant une dimension.

- **Slice:** l'opération de découpe sélectionne une dimension particulière à partir d'un cube donné et fournit un nouveau sous-cube.

- **Dice:** cette opération est similaire à une Slice.

- **Pivot:** les opérations de Pivot permettent une réorientation d'une analyse pour fournir une vue alternative des données en faisant tourner les axes de données du cube.

Le logiciel OLAP localise alors l'intersection des dimensions et les affiche. Le résultat est la "mesure", chaque cube OLAP possède au moins une à peut-être des centaines de mesures, qui sont dérivées des informations stockées dans les tables de fait et de l'entrepôt de données.

## **13 Conclusion**

Ce chapitre sert de préambule aux deux chapitres précédents, dans lesquels nous avons discuté en détail du système de prise de décision, en particulier de la phase ETL et de son rôle dans l'analyse approfondie, suivi d'un large aperçu de notre modèle multidimensionnel.

La suite de ce chapitre se trouve dans le chapitre suivant, qui a pour objectif de mettre notre modèle en action

# Chapitre 4

## Implémentation et Mise en œuvre

### 1 Introduction

Dans le chapitre précédent nous allons détaillés théoriquement concernant la hiérarchie du modèle OLAP. Ensuite dans ce chapitre nous aurons entrain de profonds au sein du l'implémentation et la mise en œuvre cette analyse.

### 2 Les outils de développements utilisés

Notre choix était basé sur l'outil SQL server pour créer notre base de données, le Visual Studio pour accomplir le cube de données et le Java Netbeans pour développer les interface de notre projet :

### 3 Microsoft SQL Server Management Studio 2014

Est un système de gestion de base de données SGBD gratuite de Microsoft SQL Server, fiable et perfomant qui offre un stockage de données riches, pour des sites web légers et des applications de bureau. Développé et proposé au téléchargement par la société Microsoft. cette version contient une liste complète des fonctinnalités sous les systèmes d'exploitation Microsoft Windows ainsi que sur les Linux depuis la version 2017 [26].

### 4 Microsoft Visual Studio 2019

Microsoft Visual Studio est un ensemble d'outils de développement pour Windows et Mac OS créé par Microsoft. Visual Studio est un ensemble d'outils de développement qui vous permet de créer des applications Web ASP.NET, des services Web XML, des applications métier et des applications mobiles. De plus, ces langages permettent de profiter plus facilement des fonctionnalités du framework. NET, qui, grâce à Visual Web Developer, vous donne accès aux technologies clés qui facilitent le développement d'applications Web ASP et de services Web XML [27].

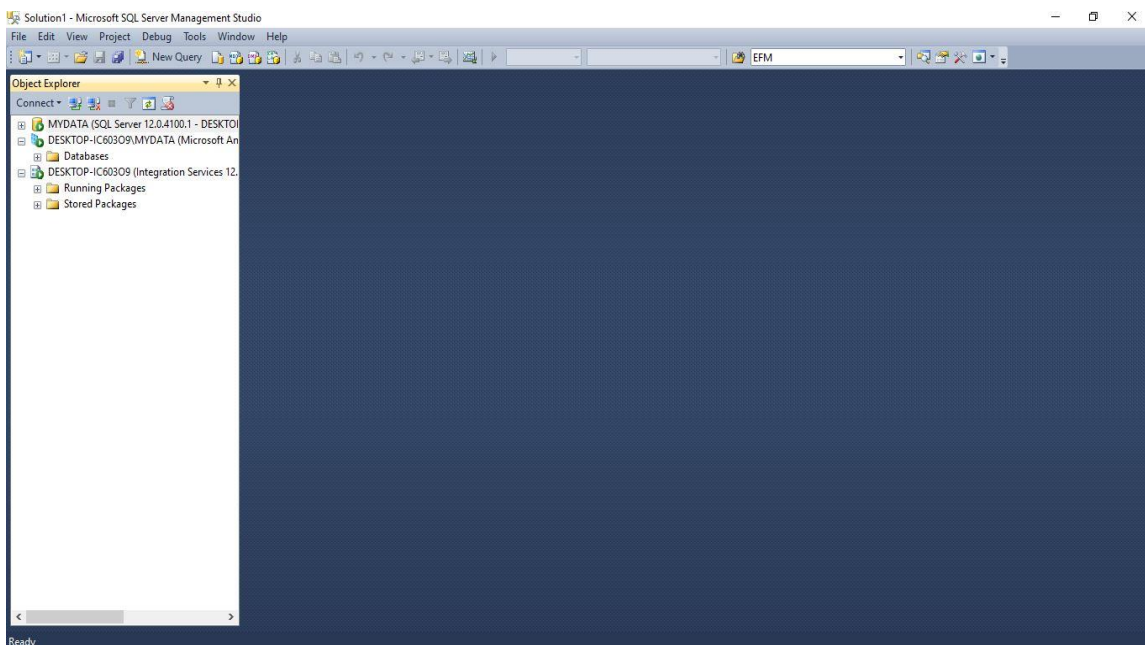
## 5 Java Netbeans

NetBeans est un environnement de développement intégré Java qui a été publié en open source. En plus de Java, Netbeans permet également de supporter différents autre langage comme python, C , C++ , XML et HTML. Il comprend toute les caractéristiques d'un IDE moderne , il disponible sous Linux, Windows, Solaris et Open VMS [28].

## 6 Création de la base de données

Dans cette étape, nous allons créer notre base de données en utilisant SQL Server Management Studio 2014.

La figure 20 représente la fenêtre principale SQL Server Management Studio :



**Figure 20** – Fenêtre principale de SQL Server Management Studio



Une fenêtre s'ouvre où nous allons préciser le nom de la base de données : BaseBD

La figure 21 représente la fenêtre création de la base de données :

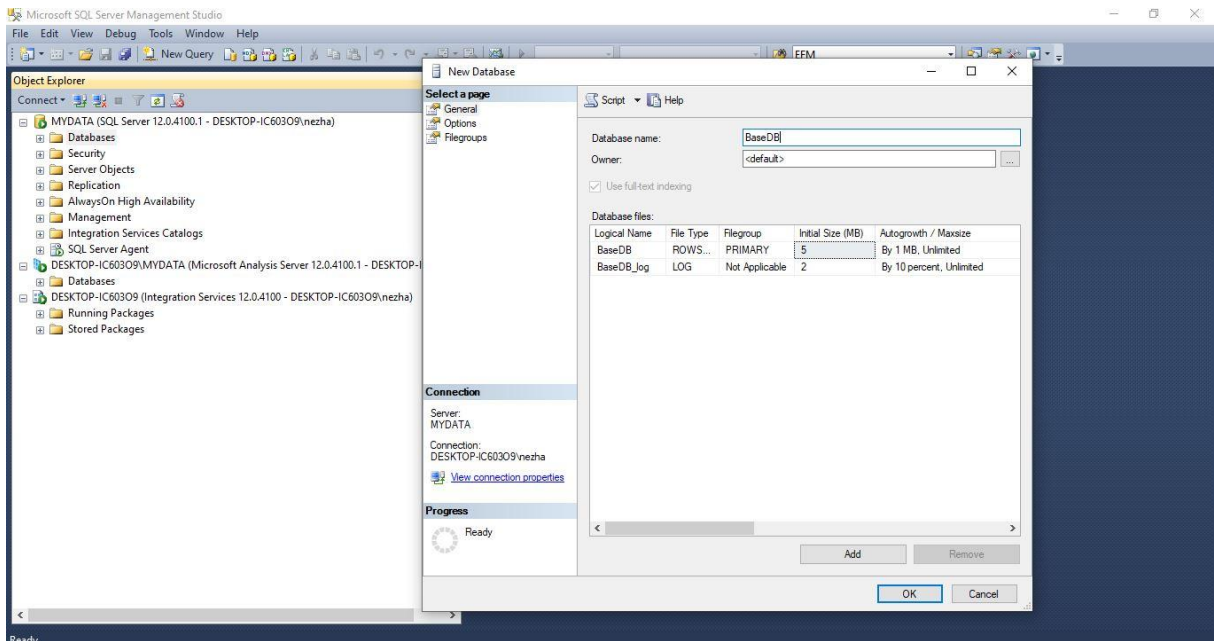


Figure 21 – Fenêtre création de la base de données

Après la création de la base de données, nous pouvons créer les tables que nous avons besoin dans notre base. La figure 19 représente un exemple d'une création de la table :

Depuis cette fenêtre, nous allons créer des colonnes pour la table Patient, nous spécifiant les types de données et la clé primaire de la table Patient, la figure 22 représente la création des colonnes pour la table Patient :

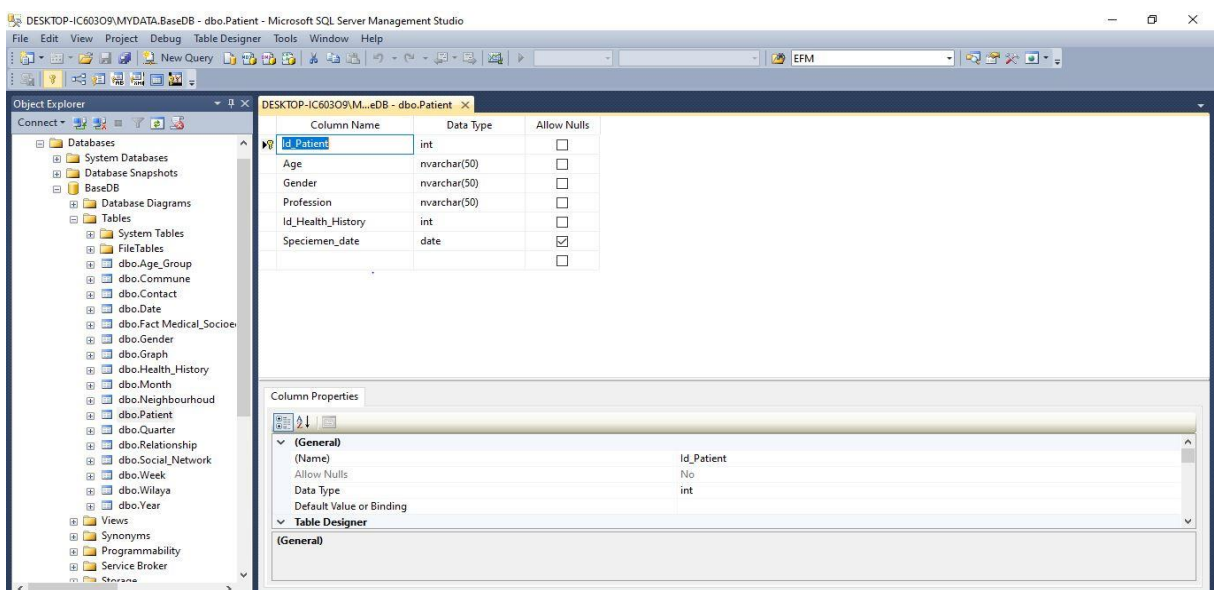


Figure 22 – Création d'une table

Après la création des tables, nous allons remplir les différents champs par les données, La figure 23 représente le processus de remplissage :

Id_Patient	Age	Gender	Profession	Id_Health_Hist...	Specimen_date
1	19-59	M	Fonctionnaire	1	2021-01-20
2	19-59	M	Clandestin	2	2021-01-31
3	19-59	F	Sans	1	2021-01-19
4	60+	F	sans	1	2021-01-25
5	19-59	F	infectiologue	3	2021-01-28
6	19-59	M	Chauffeur de tr...	4	2021-01-28
7	19-59	M	sans	5	2021-01-05
8	60+	M	retraité	6	2021-01-04
9	19-59	F	enseignante ec...	3	2021-01-04
10	60+	M	sans	1	2021-01-03
11	19-59	F	Secrétaire au C...	3	2021-01-19
12	19-59	F	medecin	1	2020-12-29
13	60+	F	sans	7	2021-01-10
14	19-59	F	commerçant	1	2021-01-07
15	19-59	M	hygieniste	1	2021-01-10
16	19-59	F	étudiante méde...	1	2021-01-04
17	19-59	M	dentiste	1	2021-01-04
18	60+	F	sans	8	2021-01-11
19	19-59	F	femme de men...	9	2021-01-11
20	19-59	M	commerçant	9	2021-01-07
21	19-59	F	etudiant	9	2021-01-07
22	19-59	M	sans	9	2021-01-13
23	19-59	M	sans	9	2021-01-27
24	19-59	F	administrateur	9	2021-01-21

Figure 23 – Exemple d’une table patient

## 7 Extraction du cube de données

Tout d’abord choisir un des projet multidimensionnel pour analyse service d’après l’ajoute du extensions pour SSAS , SSIS et SSDS pour SQL server 2014 BI par extension sous logiciel Visual Studio 2019. La figure 24 représente la création d’un nouveau projet.

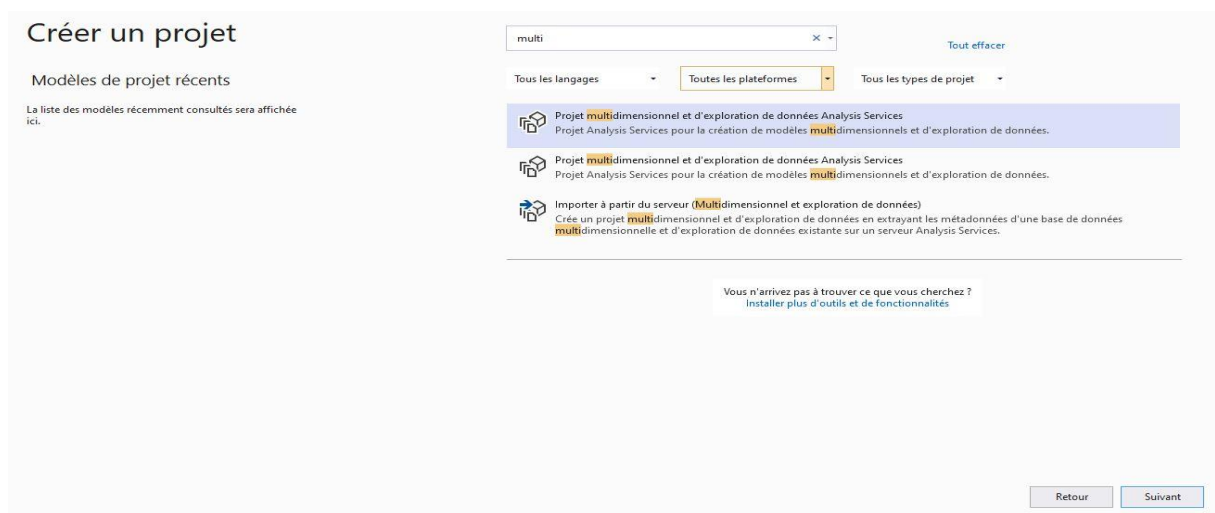


Figure 24 – la création d’un nouveau projet

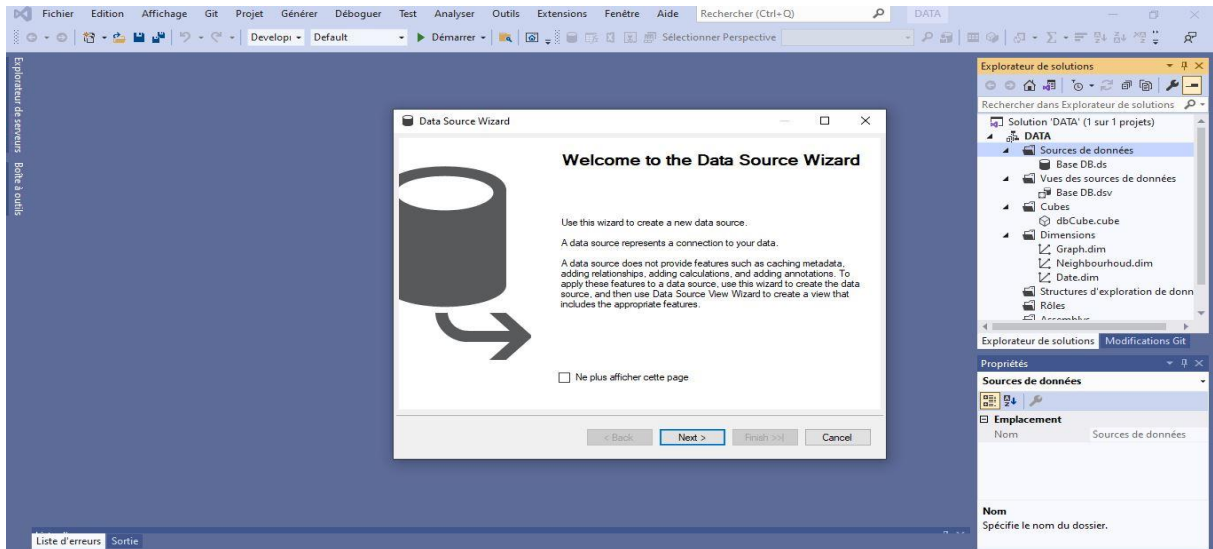


Figure 25 – Ajouter nouvelle source de données

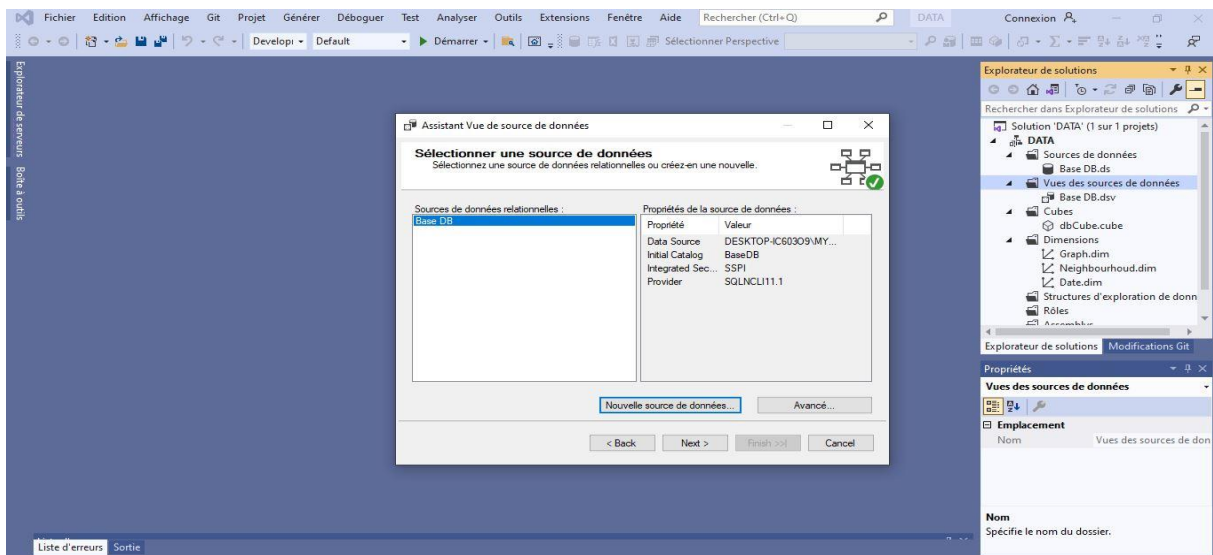
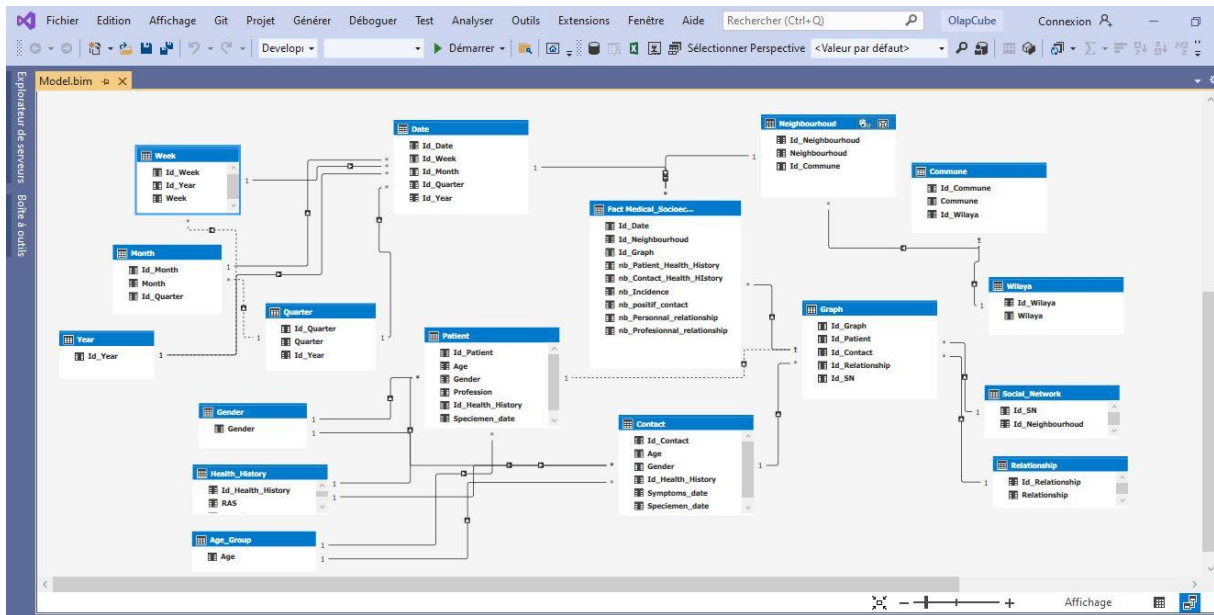


Figure 26 – Ajouter une vue des sources de données

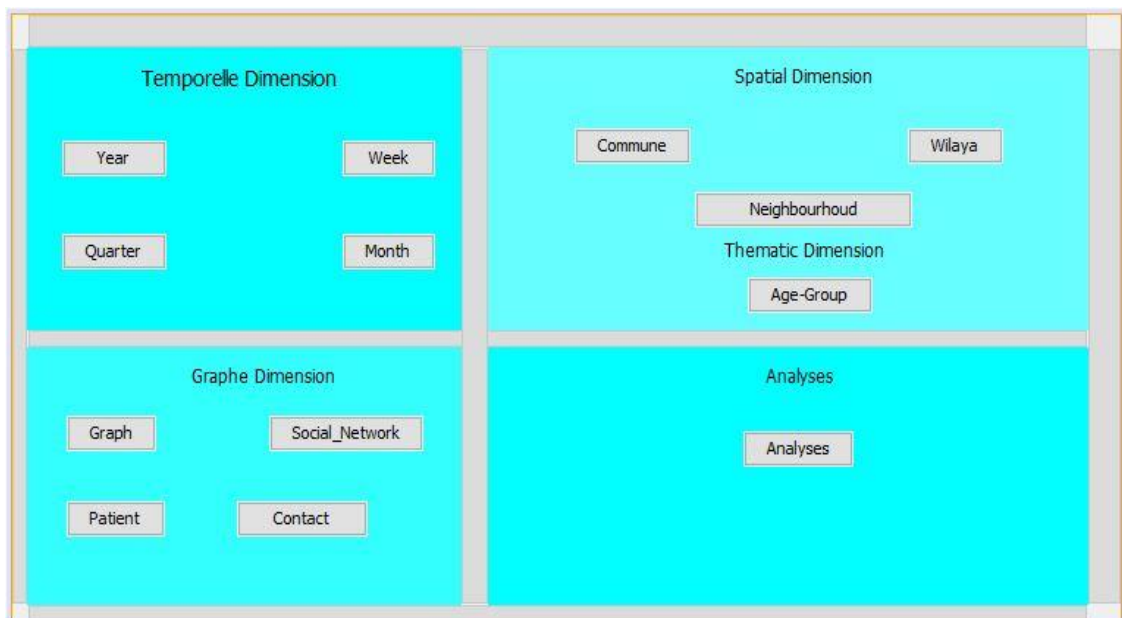


**Figure 27** – le schéma de cube de données

Pendant la création du cube en voir que la table de fait en jaune et les hiérarchies du avec les mesures.

Les interface de notre application :

Premier interface comme fenêtre apparu comme suit, la figure 28 représente l’interface d’accueil.



**Figure 28** – l’interface d’accueil

La figure 29 représente la fenêtre de la table Patient avec ces attributs.

The screenshot shows a window titled "table patient". It contains several input fields and buttons. On the left, there are labels for "id-Patient", "Age", "Gender", "Profession", "Health history", and "Specimen date", each followed by a text input field. The "Health history" field is a dropdown menu currently showing "RAS". In the center, there are three buttons: "INSERT", "DELETE", and "UPDATE". Below the "Health history" field, there are four more dropdown menus, each showing "RAS", followed by a text input field.

**Figure 29** – Fenêtre de la table patient

Dans cette fenêtre ici affiche un graphe qui représente des résultats des mesures et dimension, la figure 30 représente l'interface des résultats.

The screenshot shows a window titled "File ?" with a "GRAPH" section on the left and a "RESULTAT MULTIDIMENSIONNELLES" section on the right. The "RESULTAT MULTIDIMENSIONNELLES" section contains a large empty area. On the right side, there are several controls: "Base de Donnees : DATA\_cube", "Cube de donees: DATA", "Mesures : NB\_Patient\_Health\_History" with a dropdown and an "Ajouter" button, a text input field, "Dimension : Graphe" with a dropdown and an "Ajouter" button, a "jTextField2" text input field, "axes : Item 1" with a dropdown, "Operateur : Item 1" with a dropdown, an "Nviquer" button, a "jButton5" button, and an "Executer" button.

**Figure 30** – l'interface des résultats

## **8 Conclusion**

Ce chapitre utilise plusieurs outils pour implémenter des différents composants pour la conception et la réalisation d'un entrepôt de données

## Conclusion générale

L'entrepasage des données est une technique de collection, transmettre, et de stockage des données pour faciliter le business intelligence des entreprises.

Dans un environnement décisionnel, un entrepôt de données est le dépôt central contenant des informations, il désigne une base de données utilisée pour l'entrepasage provenant de base de données opérationnelles ou de base de données relationnelles et fournir ainsi un socle à l'aide à la décision en l'entreprise.

Parmi les critères qui nous conduisent à recommander l'implémentation d'un entrepôt de données, nous pouvons citer :

- La croissance rapide et la taille volumineuse des données a traitées ;
- L'hétérogénéité des formats des données de l'entreprise qui sont stockées dans plusieurs bases;
- L'extraction des données prend du temps et nécessitent de nombreuses manipulations.

Ce travail a souligné les différents concepts concernant un entrepôt de données afin de mettre en évidence l'intérêt de cette approche. Nous avons présenté les différentes notions de bases sur les SID et les réseaux sociaux à utiliser dans notre prochain projet de fin d'étude.

En effet, notre futur projet consiste à analyser les données à l'aide des outils d'analyse existants. Tout d'abord, nous allons modéliser notre entrepôt de données par l'un des modèles multidimensionnels qui existe dans la littérature. Nous allons traiter les données avec les systèmes OLAP qui permettre à l'analyse multidimensionnelle plus précisément des graphes OLAP (G-OLAP) qui née afin de résoudre les nombreux problèmes de réseaux, ainsi que l'analyse spatiotemporelle SOLAP qui ouvre plusieurs manipulations en partant de la modélisation et la visualisation.

## Bibliographie

- [1]-[www.futura-sciences.com/sciences/definitions/mathematiques-theorie-graphes-4712](http://www.futura-sciences.com/sciences/definitions/mathematiques-theorie-graphes-4712)
- [2]-vers une nouvelle approche SOLAP pour l'analyse des données complexes intitulée par Miloud Daouadji Asma et Rezkallah Sarah 2016.
- [3]- [https://fr.wikipedia.org/wiki/Th%C3%A9orie\\_des\\_r%C3%A9seaux](https://fr.wikipedia.org/wiki/Th%C3%A9orie_des_r%C3%A9seaux)
- [4]- <https://journals.openedition.org/ei/593>
- [5]-[www.labri.fr/perso/maylis/ACA/supports/Cours](http://www.labri.fr/perso/maylis/ACA/supports/Cours)
- [6]-[fr.wikipedia.org/wiki/Analyse\\_des\\_r%C3%A9seaux\\_sociaux](http://fr.wikipedia.org/wiki/Analyse_des_r%C3%A9seaux_sociaux)
- [7]-[fr.wikipedia.org/wiki/R%C3%A9seau\\_complexe#Les\\_r%C3%A9seaux\\_complexe\\_dans\\_la\\_vie\\_r%C3%A9elle](http://fr.wikipedia.org/wiki/R%C3%A9seau_complexe#Les_r%C3%A9seaux_complexe_dans_la_vie_r%C3%A9elle)
- [8]-[www.lebigdata.fr/-business-intelligence--definition](http://www.lebigdata.fr/-business-intelligence--definition)
- [9]- Cours de Monsieur bensalloua Charef
- [10]-[www.lebigdata.fr/data-warehouse-entrepot-donnees-definition](http://www.lebigdata.fr/data-warehouse-entrepot-donnees-definition)
- [11]-mise en place d'un SID pour le suivi et la prévention des épidémies intitulée par Mme Younsi Fatima Zohra 2015.
- [12]- <https://dumas.ccsd.cnrs.fr/dumas-01076648/document>
- [13]- <https://www.lebigdata.fr/olap-online-analytical-processing>
- [14]- mise en place d'un SID pour le suivi et la prévention des épidémies intitulée par Mme Younsi Fatima Zohra 2015.
- [15]- vers une nouvelle approche SOLAP pour l'analyse des données complexes intitulée par Miloud Daouadji Asma et Rezkallah Sarah 2016.
- [16]- vers une nouvelle approche SOLAP pour l'analyse des données complexes intitulée par Miloud Daouadji Asma et Rezkallah Sarah 2016.
- [17]- [www.oracle.com/fr/database/processus-etl-definition.html](http://www.oracle.com/fr/database/processus-etl-definition.html)
- [18]- <https://www.talend.com/fr/resources/guide-etl/>
- [19]- <https://www.axysweb.com/etl-vs-eai/>
- [20]- <https://reussir-son-crm.fr/2017/07/25/definition-ssis-ssas-ssrs/>
- [21]- <https://www.astera.com/fr/type/blog/etl-vs-elt-whats-the-difference/>
- [22]- <https://www.theses.fr/195008421>
- [23]- <https://www.theses.fr/195008421>
- [24]- <https://www.theses.fr/195008421>
- [25]- <https://www.theses.fr/195008421>
- [26]-<https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1203605-sql-server-structured-query-language-server-definition-traduction/>



[27]- [https://en.wikipedia.org/wiki/Microsoft\\_Visual\\_Studio](https://en.wikipedia.org/wiki/Microsoft_Visual_Studio)

[28]- <https://www.techno-science.net/definition/5346.html>