

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Saad DAHLEB de Blida

Faculté des sciences

Département d'informatique



## Mémoire de fin d'études

Pour l'obtention du diplôme de Master en Informatique

Option : Génie Logiciel

Thème

Conception et Réalisation d'un système Géo Décisionnel [Tableau de bord] pour l'activité commerciale d'une entreprise.

Promoteur : Mme H.ABED

Encadreur : Mr K.BOUSSEBAT

Organisme d'accueil : Icon software.

Mémoire Présenté par : Mr Redhouane CHENTIR

Soutenu le :

devant le jury composé de :

Président

Mr...BALA

Examineur

Mlle...ELGHERSS

Examineur

Mlle...ALI.MAZIGHI

MA-004-121-1

**Résumé :**

Notre mémoire traite la conception et la réalisation d'un système d'information décisionnel à caractère géographique pour l'activité commerciale d'une entreprise de cosmétique. Ce mémoire aborde les concepts fondamentaux du cycle de réalisation du projet géo-décisionnel, en commençant par la création de l'entrepôt de données, la construction de la cartographie des profils clients, l'alimentation de l'entrepôt ainsi que l'établissement d'un tableau de bord basé sur le reporting et les analyses multidimensionnelles classiques et géographiques.

**Mots clés :** GEO, décisionnel, système décisionnel, data warehouse, OLAP, SIG, tableau de bord, intelligence économique.

**Abstract :**

Our thesis treats the conception and implementation of a business intelligence system with geographical character of a business cosmetics company. This thesis discusses basic concepts of geo business intelligence system, starting with the creation of the data warehouse, the construction of customer's map profiles, the feeding of the data warehouse and the establishment of a dashboard based on reporting, classical multidimensional analysis and geographic.

**Key words :** BI, Business intelligence, Geo, Geo BI, OLAP, Data Warehouse, GIS, dash board.

## **Problématique :**

Les professionnels du marché des cosmétiques en Algérie se disent optimistes quant à l'évolution de leurs affaires, à l'instar du groupe Venus qui veut se diversifier. La lourde compétitivité qu'impose l'ouverture du marché sur les petites et moyennes industries oblige ces dernières à adopter une nouvelle approche dans la prise de décision.

En l'absence d'un outil suffisamment fiable de planification et de prise de décision, les entreprises peuvent encourir des risques en se basant sur des intuitions et des visions globales du marché actuel. L'entreprise désire comprendre les comportements de ses clients en établissant un système décisionnel capable de gérer au mieux l'activité de pilotage au haut niveau et faciliter une prise de décision ciblée basée sur des analyses pertinentes. L'intégration de la dimension spatiale va pouvoir résoudre le problème du « QUI est mon client ? » et « OU se trouve mon client » pour mieux lui vendre des produits ciblés.

## **Objectifs :**

Avec ce thème on s'est engagée dans le développement d'un système décisionnel. Le but est de proposer un système qui assure un meilleur pilotage des activités commerciales et de distribution (PME/PMI<sup>1</sup>), ainsi, on va introduire la dimension géographique pour une meilleure analyse et compréhension du comportement du client; donc une meilleure prise de décision. L'objectif est de concevoir et de déployer un nouveau système décisionnel qui se traduit par :

- La spécification des besoins des décideurs pour l'activité commerciale.
- La construction du magasin de données.
- La constitution d'un SIG (système d'information géographique) des profils clients.
- La constitution d'un portail de restitution basé sur l'intégration des deux domaines (la BI<sup>2</sup> et le SIG<sup>3</sup>).

---

<sup>1</sup> Petites et moyennes entreprises/petite et moyennes industries.

<sup>2</sup> Business intelligence.

<sup>3</sup> Systèmes d'informations géographiques.

Notre mémoire est organisé en trois grandes parties, la première un état de l'art des principales disciplines dont nous avons eu recours pour mener à bien notre projet. Ces disciplines sont organisées dans les chapitres suivants :

Chapitre 1 : L'Intelligence Economique et les Systèmes Décisionnels.

Chapitre 2 : Les Entrepôts de Données.

Chapitre 3 : Modélisation Multidimensionnelle.

Chapitre 4 : les Systèmes d'informations Géographiques

La seconde partie est une conception de la solution proposée, elle se distingue par les chapitres suivants :

Chapitre 5 : Etude des besoins.

Chapitre 6 : Conception Multidimensionnelle.

La dernière concerne le choix des outils et la réalisation du système, cette partie comporte les chapitres suivants :

Chapitre 7 : Outils technologiques.

Chapitre 8 : Mise en Œuvre.

## Sommaire :

Titre	Page
Résumé	
Introduction générale	

## Chapitre 1 : Intelligence Economique et Systèmes Décisionnels

1. Introduction :	1
2. Gérer Le Flux d'informations :	1
3. Stratégie et information :	2
4. Des données à l'information, de l'information à la connaissance :	3
5. Concepts de base et définitions :	4
6. Cadre d'application de l'intelligence économique :	7
7. Etapes du processus d'intelligence économique :	8
8. Les systèmes décisionnels :	9
9. Définition d'un système d'information décisionnel :	9
10. Les caractéristiques des applications décisionnelles :	10
11. Conclusion :	10

## Chapitre 2 : Les Entrepôts de Données

1. Introduction :	11
2. Historique :	11
2.1- L'infocentre :	11
2.2- Les Executive Information System -EIS- :	13
2.3- Les Entrepôts de données :	13
3. Pourquoi un entrepôt de données	15
3.1- Problématique des entreprises :	15
3.2- Définition	16
4. Les composants d'un Data Warehouse	19
5. L'alimentation d'un Data Warehouse:	22
5.1- Phases d'alimentation du Data Warehouse :	22
5.1-1. Découverte des données :	22

5.2-	Exploitation des données du Data Warehouse :	23
6.	Conclusion	24

### **Chapitre 3 : Modélisation Multidimensionnelle**

1.	Introduction	25
2.	Consolidation des besoins :	25
3.	Composants du modèle multidimensionnel :	25
4.	Schéma en étoile :	28
5.	Schéma en flocon de neige :	30
6.	Passage du modèle relationnel au modèle multidimensionnel :	31
7.	Caractéristiques du modèle OLAP :	31
8.	Navigation au sein d'un cube OLAP :	32
9.	Conclusion :	33

### **Chapitre 4 : Systèmes d'informations géographiques.**

1.	Introduction :	34
2.	Définition :	34
3.	La donnée géographique :	35
4.	Caractéristiques des données géographiques :	36
5.	Classement des projections d'après les altérations	39
6.	Type de canevas de représentation des méridiens et des parallèles :	40
7.	Mode de représentation de l'information géographique dans un SIG	41
8.	Les domaines d'application :	46
9.	Conclusion	46

### **Chapitre 5 : Etude des besoins**

1.	Introduction :	48
2.	Création de questionnaire	48
3.	Questions relatives aux dirigeants :	49
4.	Questions aux responsables :	50

5.	Compte Rendu de l'entretien :	51
6.	Composants d'un entrepôt de données :	53
	a. Description des dimensions :	53
	b. Dédution des besoins :	54
7.	Conclusion	55

## **Chapitre 6 : Conception Multidimensionnelle**

1.	Introduction :	56
2.	Le Processus de conception :	56
3.	Modélisation dimensionnelle de l'activité Vente :	58
4.	Description des dimensions :	60
5.	Architecture du système :	62
6.	Construction du datawarehouse :	63
7.	L'aspect géographique :	65
8.	La restitution des informations géographiques :	66
9.	Alimentation de l'entrepôt de données :	66
10.	Conclusion :	67

## **Chapitre 7 : Outils Technologiques**

1.	Introduction :	68
2.	PostgreSQL 9.1 :	68
3.	PostGIS :	69
4.	Openlayers :	69
5.	Quantum GIS :	69
6.	Talend :	70
7.	Pentaho :	71
8.	Conclusion :	71

## **Chapitre 8 : Mise en Œuvre**

1.	Introduction :	72
2.	Construction du datawarehouse :	72

3.	Alimentation du datawarehouse :	72
4.	Création de la cartographie des profils clients :	76
5.	Création de cube dimensionnel :	80
6.	Création du tableau de bord :	81
	• Authentification et création de tableau de bord :	82
	• Génération de tableau de bord :	85
7.	Conclusion :	86

### **Conclusion Générale**

Conclusion générale :	87
-----------------------	----

### **Annexes**

Annexe A : Questionnaire	89
Annexe B : Description de la base de données	91
Bibliographie	97

## Liste des figures :

Figure	Page
Figure 1 : Importance stratégique de la gestion de l'information.	2
Figure 2 : Le cycle de l'intelligence	3
Figure 3 : Pyramide modélisant le processus de BI	6
Figure 4 : Flux informationnel lié au Business intelligence	6
Figure 5 : Etapes du processus d'intelligence économique	8
Figure 6. Approche Infocentre.	12
Figure 7 : Approche Entrepôt de données.	14
Figure 8 : Définition Entrepôt de données.	17
Figure 9 Composants de base d'un Data Warehouse	19
Figure 10 : Structure de base d'une table de faits.	27
Figure 11 : Structure de base d'une table de dimension.	28
Figure 12 : Schéma en étoile	29
Figure 13 : Schéma en flocon de neige.	30
Figure 14 : Passage de l'étoile au cube.	31
Figure 15 : Architecture OLAP Multidimensionnel	32
Figure 16 : différentes visions d'une rue	36
Figure 17 : l'objet géographique support et/ou référentiel	37
Figure 18 : le géoïde est la forme théorique de la terre	38
Figure 19 : l'ellipsoïde	38
Figure 20 : altérations différentes selon les systèmes	39
Figure 21 : canevas cylindrique	40
Figure 22 : canevas conique.	40
Figure 23 : canevas azimutal (aphylactique).	40
Figure 24 : couche raster d'une zone.	41
Figure 25 : image pixélisée.	42
Figure 26 : photo aérienne	42
Figure 27 : carte scannée	43
Figure 28 : image satellitale	43
Figure 29 : image satellitale radar.	44
Figure 30 : Représentation alphanumérique d'une donnée géographique	44
Figure 31 : points, polygones et surface Algerie nord.	45

Figure 32 : Représentation sous forme de cube	52
Figure 33 : modèle dimensionnel en flocon de neige pour l'activité vente.	61
Figure 34 : Architecture du système.	62
Figure 35 : interface pgAdmin3 (PostgreSQL 9.1)	72
Figure 36 : création de connexion à une base de donnée	73
Figure 37 : choix de la base de données et identification.	73
Figure 38 : Récupération du schéma du datawarehouse.	74
Figure 39 : choix du schéma et des tables à récupérer.	74
Figure 40 : création de job pour alimentation de tables.	75
Figure 41 : job ETL de la table tf_vente.	75
Figure 42 : jointure entre deux tables avec l'outil Tmap et alimentation.	76
Figure 43 : couche Algérie sur quantum gis.	77
Figure 44 : ajouter une nouvelle couche région.	77
Figure 45 : couche adresse, exemple adresse client Blida.	78
Figure 46 : cartographie de la répartition client.	78
Figure 47 : sauvegarde couche.	79
Figure 48 : importation du fichier .shp.	79
Figure 49 : administration de la connexion au DW	81
Figure 50 : Ajout d'un nouvel utilisateur.	82
Figure 51 : Authentification Pentaho	82
Figure 52 : interface Pentaho business analytics.	83
Figure 53 : récupération du cube publié en xml	83
Figure 54 : localisation des clients par chiffre d'affaire.	84
Figure 55 : Analyse bidimensionnelle.	84
Figure 56 : création de tableau de bord.	85
Figure 57 : exemple de tableau de bord.	85

## Liste des tableaux :

Tableau	Page
Tableau 1. Comparaison entre Infocentre et entrepôt de données	15
Tableau 2 : Questions aux dirigeants.	50
Tableau 3 : Questions aux responsables.	51
Tableau 4 : dimensions et leur description.	53
Tableau 5 : Tableau d'indicateurs des axes d'analyse.	59
Tableau 6 : Tableau descriptif des dimensions.	60
Tableau 7: Attributs de la table de fait et dimensions.	64

*ETAT DE L'ART*

*INTELLIGENCE ECONOMIQUE  
ET SYSTEME DECISIONNEL*

## 1. Introduction :

Grâce au développement rapide et continu des technologies, l'accès à l'information est aujourd'hui grandement facilité et constitue sans nul doute l'une des caractéristiques majeures de la société qui se crée chaque jour devant nous.

L'Intelligence économique vise à tirer parti de cette situation en développant des méthodes qui permettent l'identification de sources pertinentes d'information, l'analyse de l'information collectée et sa transformation en connaissance pour aider à la prise de décision. Tournée vers l'information disponible en dehors de l'entreprise, l'Intelligence économique couvre de larges domaines d'intérêts, passant de la technologie au marché ou à la réglementation. Aujourd'hui, la lente diffusion des méthodes et outils de l'Intelligence économique en Europe ne doit pas masquer le formidable potentiel d'innovation qu'il engendre, notamment pour les PME soucieuses d'affronter la concurrence et les marchés dans les meilleures conditions.[CRL, 2002]

## 2. Gérer Le Flux d'informations :

La globalisation de l'économie, la généralisation des technologies de l'information et de la communication, la construction de réseaux formels ou informels, l'accélération des échanges économiques, l'évolution des relations entre le donneur d'ordre et ses prestataires, le développement de ce qu'on nomme la gestion de la relation client (CRM: Customer Relationships Management), le raccourcissement des cycles de vie des produits... conduisent à adapter en permanence la gestion au quotidien des entreprises. Les grandes entreprises et organisations ont bien compris ces nouvelles exigences et ont développé en conséquence des démarches d'Intelligence économique répondant à leurs propres besoins. Aujourd'hui, ces défis sont presque identiques pour les petites entreprises. Le vaste champ d'investigation est plus ou moins le même et la réactivité se doit d'être la même. Pour autant, les moyens financiers, humains et techniques ne suivent pas cette logique.

Pour gérer une telle masse de données et d'informations, il est absolument indispensable d'adopter des méthodes de tri et de sélection, pragmatiques et efficaces.

### 3. Stratégie et information :

La gestion au quotidien de l'entreprise repose sur un cadre stratégique dont les racines sont fortement ancrées dans l'information.

Sans stratégie, l'entrepreneur est un peu comme un explorateur sans carte et boussole. Il aura beau obtenir autant d'informations qu'il le désire, elles ne lui seront d'aucune utilité. La stratégie est le résultat d'une dialectique entre la situation interne à l'entreprise et le monde qui l'entoure. Grâce au Benchmarking (action qui consiste à confronter son expérience à celle des autres), le chef d'entreprise élabore son propre cadre d'action, regardant vers le long terme (la stratégie) et opérant au quotidien.

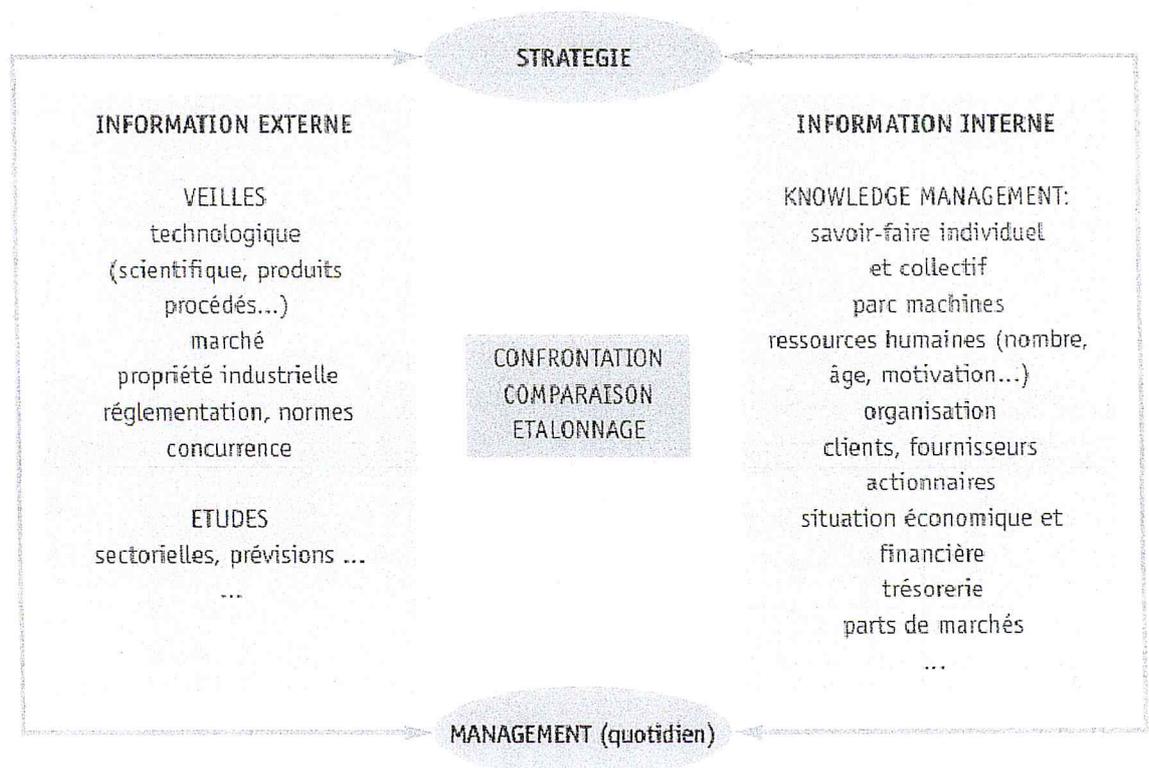


Figure 1 : Importance stratégique de la gestion de l'information. [CRL, 2002]

Aujourd'hui, l'analyse de la situation interne a pour mission de renseigner le dirigeant sur l'état réel de son entreprise, sur la base de connaissances

tangibles (procédures, capacités du parc machines, situation financière et trésorerie, organisation, carnet de commandes...) et tacites (savoir-faire, situation de la ressource humaine, relations avec les clients...). Le paysage externe apporte pour sa part de nombreuses informations, éventuellement utiles, issues d'une veille classique technologique (normes, brevets, réglementation, produits et procédés, clients, concurrents, fusions et acquisitions...) et permettant d'avoir une vision du futur (tendances, prévisions de marché, prospective, évolutions politiques et sociales...)

#### 4. Des données à l'information, de l'information à la connaissance :

La demande classique du chef d'entreprise à son système d'information est la suivante : " Je veux la bonne information au moment opportun ". Mais obtenir la bonne information au bon moment, est le résultat d'un processus permanent et d'une politique décidée au plus haut niveau.

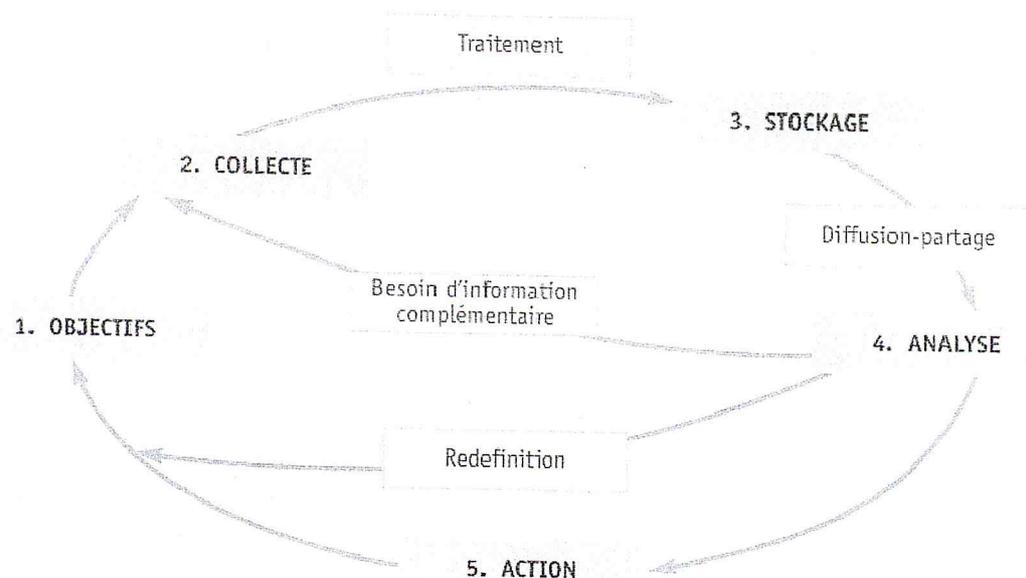


Figure 2 : Le cycle de l'intelligence. [CRL, 2002]

Une fois les objectifs globaux en matière d'information arrêtés, les missions de collecte, de stockage et d'analyse de l'information doivent être conçues de manière à aider l'utilisateur dans sa prise de décision finale. Il s'agit alors de transformer la masse de données (disponibles sous différentes formes, souvent inorganisées et collectées par différents canaux) en information, puis en connaissance.

Les méthodes et outils de l'Intelligence économique permettent de nos jours de valider les données collectées (à partir de différentes sources considérées comme fiables) en un ensemble cohérent d'information adapté au profil de l'entreprise et à ses besoins. Cette étape est aujourd'hui de la plus grande importance compte tenu du grand nombre de sources disponibles : études prospectives, littérature professionnelle, bases de données gratuites et payantes, données informelles du web, procédés, produits, règlements et normes, concurrents, fusions, partenariats, clients, situation du secteur industriel, évolutions sociétales... Il s'agit là d'un travail permanent puisque l'information doit être mise à jour continuellement. On peut donc dire qu'une fois le cycle de l'information parcouru, il faut reproduire l'opération. Et ainsi de suite.

#### 5. Concepts de base et définitions :

L'utilisation fréquente de termes dans des contextes différents, ne permet pas toujours d'avoir une vision claire. Voici quelques définitions de base : [Bellinger, 2010]

**Données** : nombres, mots, événements existants en dehors d'un cadre conceptuel de référence ; en conséquence, et en absence de contexte, les données prises individuellement n'ont pas une grande signification. Accumulation de données n'est donc pas information.

**Information** : ensemble de données, validées et confrontées, qui commencent à avoir un sens. Accumulation d'informations n'est pas connaissance.

**Connaissance** : ensemble d'informations interprétées par l'entreprise et lui permettant de prendre des décisions. Accumulation de connaissance n'est pas sagesse (intelligence).

**Intelligence** : elle apparaît lorsque les principes fondamentaux qui ont fondé la connaissance sont compris. Accumulation de sagesse (intelligence) n'est pas vérité.

A l'origine, ce concept n'est pas français. Les britanniques parlent de «business intelligence», renseignement des affaires, et les américains de «competitive intelligence», renseignement concurrentiel. En fait, il existe plusieurs définitions de l'Intelligence économique, cette notion ayant fait l'objet de divers débats conceptuels. La première définition apparaît en 1967 dans un ouvrage d'Harold Wilensky. Il y définit «l'intelligence économique comme l'activité de production de connaissances servant les buts économiques et stratégiques d'une organisation, recueillie et produite dans un contexte légal et à partir de sources ouvertes ». Cette définition permet de distinguer l'intelligence économique de l'espionnage économique car elle se développe ouvertement et utilise principalement des moyens légaux.[DUFAU, 2010]

Selon la commission européenne qui travaille sur le projet TRIPS, **l'intelligence économique [IFOACT, 2000]**est un ensemble de concepts, méthodes et outils qui unifient toutes les actions coordonnées de recherche, acquisition, traitement, stockage et diffusion d'information pertinente pour des entreprises considérées individuellement ou en réseaux, dans le cadre d'une stratégie partagée.

Ces processus cohérents, permanents, itératifs, conduisent à des modifications importantes dans les comportements individuels et collectifs, et amènent des transformations dans les mécanismes de prise de décision. Le développement de l'Intelligence économique concerne en outre tous les secteurs de l'entreprise : gestion, marketing, finance, organisation de la production, recherche, ressources humaines...

Le processus de l'intelligence économique se schématise de la façon suivante :

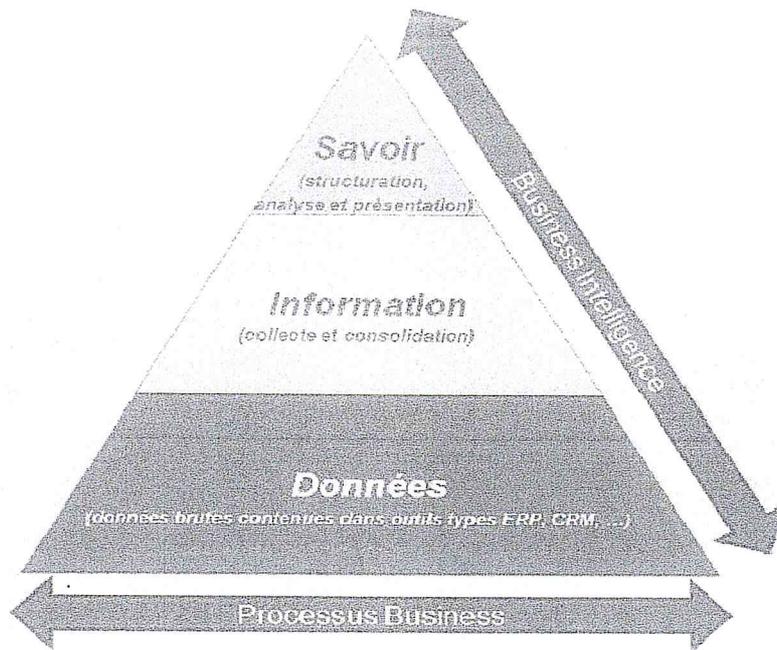


Figure 3 : Pyramide modélisant le processus de BI. [LAFARE, 2009]

Nous pouvons également modéliser le flux informationnel, c'est-à-dire l'acheminement depuis la donnée brute provenant de SI source (ERP, CRM...), à la production de reportings et aux autres tableaux de bord de la manière suivante : [CIGREF, 2009]

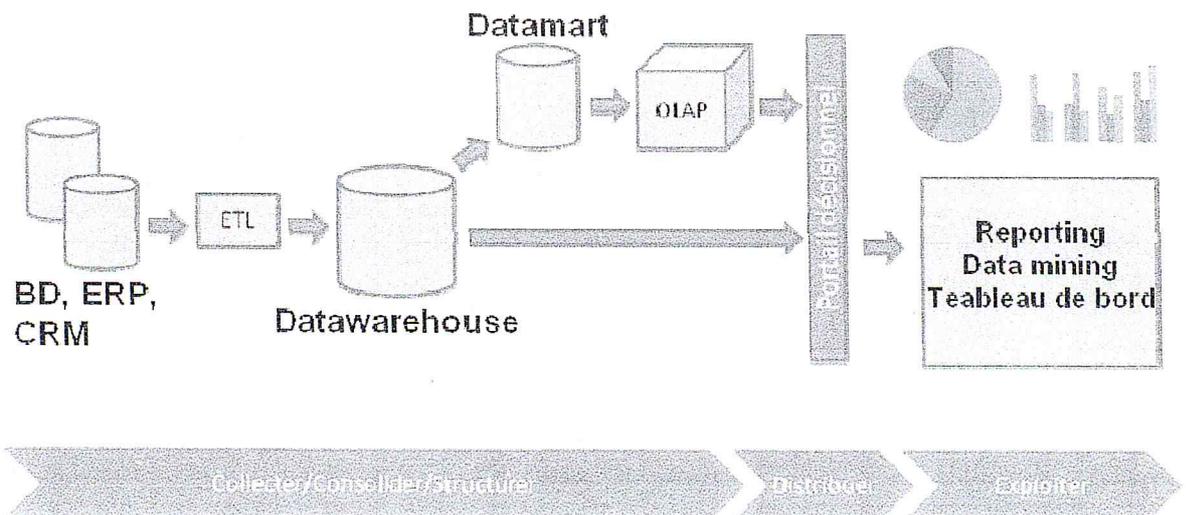


Figure 4 : Flux informationnel lié au Business intelligence. [www.piloter.org]

Il est essentiel de définir certains termes, qui seront utilisés dans la suite de ce document :<sup>1</sup>

- **Datawarehouse (entrepôt de données)** : base de données utilisées pour collecter et stocker des informations volatiles provenant d'autres bases de données.
- **Datamart** : sous ensemble logique d'un Datawarehouse. Il est généralement exploité en entreprise pour restituer des informations ciblées sur un métier spécifique.
- **Datamining** : les outils de Datamining permettent d'extraire des hypothèses à partir de grandes quantités de données, par des procédés typiquement statistiques.
- **ETL (Extract, Transform, Load)** : il s'agit d'une technologie informatique intergicielle (Middleware) permettant d'effectuer des synchronisations massives d'information d'une base de données vers une autre.
- **Cube OLAP** : représentation abstraite d'informations multidimensionnelles, les données sont rangées selon un principe de dimension correspondant étroitement aux axes de recherche des utilisateurs (par exemple les ventes de produit dans le temps par zone géographique).

## 6. Cadre d'application de l'intelligence économique :

Pour être efficace dans son approche d'Intelligence économique, une entreprise doit définir avec précision les objectifs qu'elle entend atteindre :

- S'agit-il d'objectifs stratégiques : modernisation, innovation, expansion?
- S'agit-il d'être plus compétitif sur un marché difficile ?
- S'agit-il de maintenir sa position de leader ?

Dans un premier temps, toutes ces questions doivent être clairement identifiées, partagées et discutées par l'équipe de direction. Dans une deuxième phase, l'ensemble de ces choix sera expliqué au personnel, de

---

<sup>1</sup> Définitions des termes voir chapitre 2.

manière simple, comportant les tâches que chacun aura à accomplir. La difficulté réside dans l'élaboration d'un plan de collecte de l'information suffisamment pertinent par rapport aux enjeux de l'entreprise. Il n'y a en effet aucun intérêt à collecter des masses d'informations qui traitent de thèmes généraux si cette action n'apporte rien de déterminant et si elle nécessite un temps d'analyse important. Quand on collecte de l'information pour le compte d'une personne, il faut tout d'abord vérifier avec cette dernière que le plan et les délais collent parfaitement avec ses besoins. Il faut aussi garder à l'esprit que la collecte de certaines informations pourra prendre parfois plusieurs semaines.

### 7. Etapes du processus d'intelligence économique :

La Business intelligence a pour objectif de permettre aux personnes situées en haut de la hiérarchie de décision de disposer d'une information de valeur, à laquelle ils puissent se fier dans le cadre de leurs prises de décision. Pour cela il est nécessaire de produire une information à forte valeur à partir d'un processus disposant des phases suivantes : [GUENDOUZ, 2008]

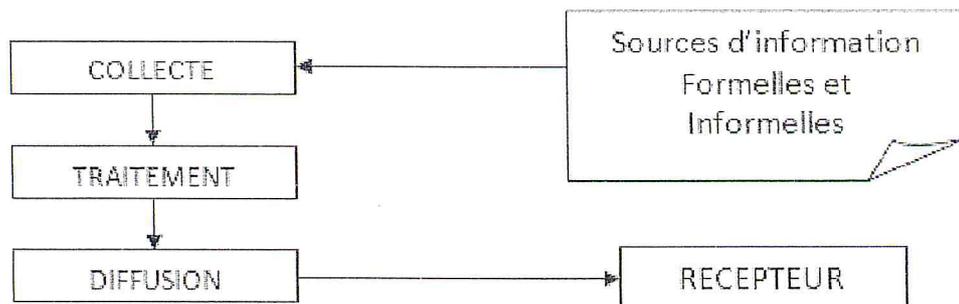


Figure 5 : Etapes du processus d'intelligence économique

- La **collecte** ou recherche, se définit comme la période de recherche où sont identifiées et exploitées les sources d'informations, ceci dans le cadre d'une planification.
- L'**analyse**, c'est-à-dire le traitement ou l'exploitation, compose l'étape au cours de laquelle les données et informations passent à l'état de

connaissance à travers un processus systématique d'évaluation, d'interprétation et de synthèse destiné à élaborer des conclusions (articulées sur des éléments significatifs) répondant aux besoins de renseignements exprimés.

- La **diffusion** est l'acheminement des renseignements sous une forme appropriée (orale, écrite ou graphique) aux organes ayant exprimé la demande. Il s'agit bien d'un cycle dans la mesure où le renseignement obtenu permet d'une part d'orienter les besoins nouveaux en renseignements et, d'autre part, de réévaluer constamment la connaissance obtenue en fonction de l'évolution de l'environnement.

### 8. Les systèmes décisionnels :

L'histoire des systèmes décisionnels commence en 1958, le chercheur d'IBM Hans Peter Luhn, a utilisé le terme Business Intelligence qu'il a défini comme «la capacité de présenter les interrelations entre des faits de telle sorte que cela permette de guider les actions pour atteindre le but espéré.» [LASSENCE, 2010]

L'informatique décisionnelle est associée au concept de Data Warehouse, ce qui se traduit en français par entrepôt de données. Ce concept a été formalisé en 1992 par Bill Inmon dans l'ouvrage "Developing the Data Warehouse" (Construire l'Entrepôt de Données) de la façon suivante : « Un Data Warehouse est une collection de données thématiques, intégrées, non volatiles et historisées pour la prise de décisions. »[KIMBALL, 1996]

### 9. Définition d'un système d'information décisionnel :

Le système d'information pour Jean-François Goglin est : « un ensemble de données organisées de façon spécifique facilement accessibles et appropriées à la prise de décision ou encore une représentation intelligente de ces données au travers d'outils spécialisés ; la finalité d'un système décisionnel et le pilotage de l'entreprise». Toujours selon Jean-François Goglin, le système décisionnel constitue une synthèse d'informations opérationnelles, internes ou externes, choisis selon leur pertinence dont le principal intérêt est d'offrir au décideur une vision transversale de l'entreprise, contrairement aux bases de production : vente clients, gestion

«qui sont incapable d'offrir aux décideurs une vue consolidée de toute les données et les transformer en information.[Abd, 2012]

#### **10. Les caractéristiques des applications décisionnelles :**

Tout d'abord, elles traitent de gros volumes de données. Les trois principales caractéristiques d'un système décisionnel sont :[GUENDOUZ, 2008]

- La capacité de traiter de gros volumes de données.
- Temps de réponse très élevés.
- Les requêtes beaucoup plus complexes du point de vue informatique, car elles contiennent de nombreuses opérations de jointure et de regroupement.

#### **11. Conclusion :**

Nous avons pu nous familiariser avec l'intelligence économique dans ses grandes lignes, un aperçu sur les systèmes d'informations décisionnels, ainsi que le processus qui permet de les réaliser, nous avons également abordé des notions telles que les stratégies assignées à cette discipline, le cycle de l'intelligence et la connaissance. La Business intelligence interprète des données d'une nature complexe pour aider les dirigeants à prendre des décisions plus au moins sans risque, ciblées, et qui auront un impact potentiel sur le marché et la vie même de l'entreprise. Si les techniques d'Intelligence économique sont de plus en plus adoptées par les grandes entreprises et les organismes de soutien au développement économique, son utilisation se fait plus lentement auprès des PME<sup>2</sup>.

---

<sup>2</sup> Petite et moyenne entreprise.

# *ENTREPOTS DE DONNEES*

## 1. Introduction :

Les entreprises passent à l'ère de l'information. Le Défi étant de transformer leur système d'information qui avait une vocation de production à un SI décisionnel dont la vocation de pilotage devient majeure. Les entreprises sont confrontées à une concurrence de plus en plus forte, des clients de plus en plus exigeants, un contexte organisationnel de plus en plus complexe et mouvant.

Pour faire face aux nouveaux enjeux économiques, l'entreprise doit anticiper. L'anticipation ne peut être efficace qu'en s'appuyant sur de l'information pertinente. Cette information est à la portée de toute entreprise disposant d'un capital de données gérées par ses systèmes opérationnels et qui peut en acquérir d'autres auprès de fournisseurs externes.

Mais actuellement, les données sont surabondantes, non organisées dans un perspectif décisionnel et éparpillées dans de multiples système hétérogènes. Pourtant, les données représentent une mine d'informations. Il devient fondamental de rassembler et d'homogénéiser les données afin de permettre d'analyser les indicateurs pertinents pour faciliter les prises de décisions.

Pour répondre à ces besoins, le nouveau rôle de l'informatique est de définir et d'intégrer une architecture qui sert de fondation aux applications décisionnelles : le Data Warehouse.

## 2. Historique :

Le Datawarehousing est le résultat d'un long processus de recherche et de développement qui est passé par les phases suivantes :

### 2.1- L'infocentre :

Il s'agit essentiellement d'un concept de mise à disposition des utilisateurs non informaticiens des entreprises, de moyens informatiques leur permettant de créer leurs propres fichiers, de manipuler les données qui les intéressent, de disposer directement d'un moyen de calcul qui offre la fonction « ordinateur personnel » dans l'entreprise.[NIEUWBOURG, 2006] L'objectif

des outils d'Infocentre est de consulter des bases de données afin d'y sélectionner des informations, de les extraire et de les restituer aux utilisateurs, sous forme de tableaux ou de graphiques. [Clainche, 2009] On a donc dupliqué les bases de données de productions, c'est-à-dire que chaque jour, chaque semaine, ou chaque mois. Une copie des informations de productions était réalisée, à l'identique, sur un autre ordinateur spécialement pour les analystes. Les équipes opérationnelles pouvaient continuer à utiliser leurs applications, sans être nullement perturbées par les requêtes analytiques, et les analystes pouvaient prendre le risque de lancer des requêtes complexes, analysant par exemple le chiffre d'affaires suivant plusieurs dimensions (clients, produits, fournisseurs, région), sur plusieurs années, mois par mois, sans prendre le risque de bloquer le système opérationnel. Ce mode de fonctionnement, en doublant totalement les données entre les deux systèmes, se révélait très onéreux. Les serveurs, les disques durs, les bases de données, devaient tous être acquis en double, uniquement pour les besoins de l'analyse. La figure suivante représente l'approche Infocentre : [BOUSSEBAT, 2003]

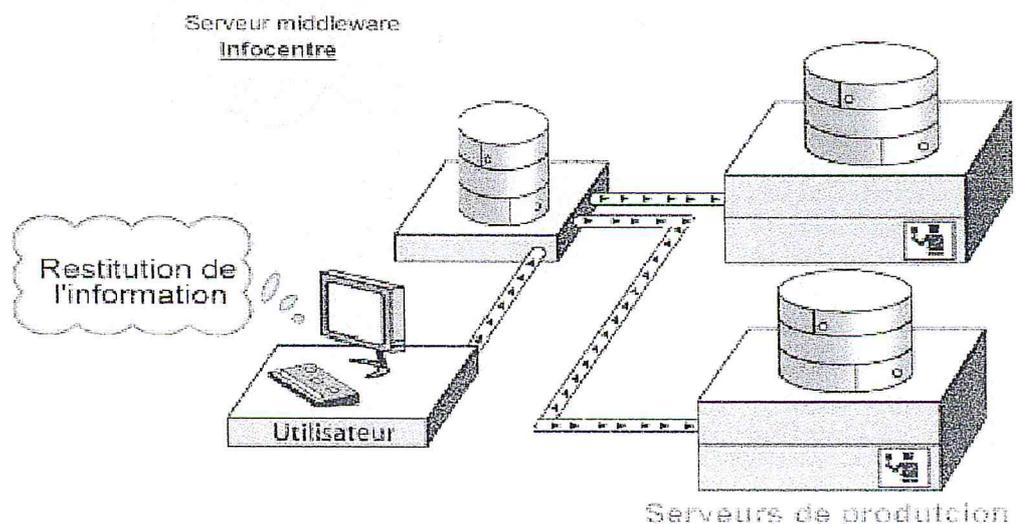


Figure 6. Approche Infocentre. [BOUSSEBAT, 2003]

## 2.2- Les Executive Information System -EIS- :

Appelé aussi Système d'information pour dirigeant, est un système d'information informatisé, spécialement conçu pour répondre aux besoins de la haute direction d'une entreprise et qui lui est réservé. Le système d'information pour dirigeants se doit de fournir une information synthétisée et à jour qui donne un aperçu général continu des activités et des opérations de l'entreprise, à partir des sources externes et internes. Il s'agit en quelque sorte du « tableau de bord » informatisé des cadres supérieurs, qui sert à la planification stratégique et à partir duquel on peut produire des rapports, des graphiques, etc., faciles à consulter rapidement. Ce système fait partie du système d'information de gestion et diffère d'un système d'aide à la décision, dans la mesure où sa fonction principale est de fournir de l'information, la plupart du temps en temps réel, plutôt que des outils d'analyse et de prise de décision. Toutefois, l'évolution technique fait que les deux types de systèmes se confondent un peu plus tous les jours. [E-BUSINESS, 2010]

## 2.3- Les Entrepôts de données :

Les managers ont besoin d'information qu'on peut obtenir à l'aide des requêtes très complexes qui exigent le traitement d'une très grande quantité d'information. L'exécution de ces requêtes occupe trop de temps et peut détériorer la performance du système comme un tout. C'est la raison pour laquelle on préfère créer des bases de données qui ne sont pas opérationnelles mais qui n'enregistrent que des faits par des intervalles de temps donnés. Les bases de données de ce type s'appellent **entrepôts de données** (*Data Warehouse*).

Après plusieurs années de duplication de données, c'est-à-dire utilisation d'infocentre, les experts ont opté pour une solution plus intelligente, ils ont en effet constaté que les informations traitées dans les applications opérationnelles étaient très différentes de celles interrogées dans les applications décisionnelles. Le nombre de tables, de fichiers interrogés dans une même requête est bien plus important dans l'aide à la décision, le nombre d'indicateurs calculés également. En revanche, les applications

décisionnelles se contentent presque toujours de lire les données. Elles n'ont jamais à écrire de nouvelles informations dans les bases de données.

Les questions posées par un décideur impliquent fréquemment des informations stockées dans plusieurs applications ou base de données. Lorsqu'on calcule la rentabilité des clients, on exporte des données de la gestion commerciale (factures, commandes), de la comptabilité (délais de règlement, impayés), mais également de la gestion de production (coût des produits fabriqués). Le fait de dupliquer ces bases de données dans un infocentre ne simplifie en rien ces extractions.

Les experts ont donc pensé à mettre en place, en sortie des bases de production, un entrepôt de données. Cet entrepôt uniquement dédié au stockage des données décisionnelles, permet de réconcilier les différentes sources initiales de données, et les applications de production.

L'entrepôt de données est construit à partir d'une base de données relationnelle, une copie des données qui nous intéressent pour la prise de décision et l'analyse sera exportée à cet entrepôt de données. Cette manœuvre va permettre d'alléger sur le long terme la base de production.

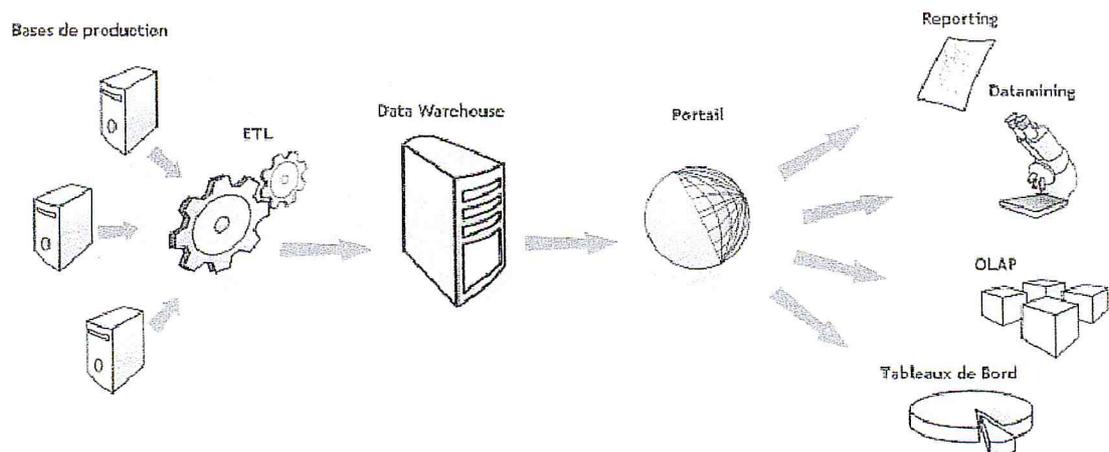


Figure 7 : Approche Entrepôt de données. [OSBI, 2010]

Certaines caractéristiques sont identiques. Mais il existe de nombreux éléments permettant de différencier les deux notions. Dans un infocentre, chaque nouvelle valeur remplace l'ancienne valeur. Il est donc impossible de

retrouver une valeur calculée dans une session préalable aux dernières alimentations. La non volatilité est une caractéristique essentielle de l'entrepôt de données.

De même, dans un infocentre, il n'y a pas de gestion d'historique des valeurs. L'infocentre sert à prendre des décisions opérationnelles basées sur des valeurs courantes. Au niveau d'un entrepôt de données, l'utilisateur travaille sur les historiques pour des prises de décisions à long terme, des positionnements stratégiques et pour analyser des tendances.

Dans un infocentre, l'intégration des données est plus ou moins poussée. Le processus d'alimentation est simple. Le tableau ci-dessous fait une comparaison entre l'infocentre et l'entrepôt de données : [KABORE, 2005]

Infocentre	Entrepôt de données
<b>Collection de données</b>	Collection de données
<b>Orientées sujet</b>	Orientées sujet
<b>intégrée</b>	Intégrée
<b>Volatiles</b>	Non volatiles
<b>Actuelles</b>	Historisées
<b>Organisées pour le support d'un processus de décision ponctuelle</b>	Organisées pour le support d'un processus de d'aide à la décision
<b>outils</b>	Architecture

Tableau 1. Comparaison entre Infocentre et entrepôt de données.[KABORE, 2005]

### 3. Pourquoi un entrepôt de données

#### 3.1- Problématique des entreprises :

La surabondance de données qui touche les organismes de toutes catégories confondues les rend difficile à organiser et à exploiter dans un contexte décisionnel. Il devient donc capital de rassembler et d'homogénéiser les données afin de permettre d'analyser les indicateurs pertinents pour faciliter les prises de décisions. L'entreprise construit un système décisionnel pour améliorer sa performance. Elle doit décider et

anticiper en fonction de l'information disponible et capitaliser sur son expérience. [HELOU, 2004]

Depuis plusieurs dizaines d'années, une importante masse d'informations est stockée sous forme informatique dans les entreprises. Les systèmes d'information sont destinés à garder la trace d'événements de manière fiable et intègre. Ils automatisent de plus en plus les processus opérationnels. L'informatique a un rôle à jouer, en permettant à l'entreprise de devenir plus entreprenante et d'avoir une meilleure connaissance de ses clients, de sa compétitivité ou de son environnement.

Il est intéressant de calculer les retours sur investissement rendus publics. Ils se calculent rarement en termes de baisse de coûts, mais en termes de gains. Par exemple, ils permettent un meilleur suivi des ventes, une meilleure compréhension des habitudes d'achats des clients, d'une adaptation des produits à une clientèle mieux ciblée.

### 3.2- Définition

La traduction littérale de Data Warehouse est "entrepôt de données". Ce mot désigne à la fois la base de données dans laquelle est stocké l'ensemble des informations mais également l'ensemble du système d'information décisionnel.

La définition classique du Data Warehouse donnée par Bill Inmon dans son ouvrage de référence *Using the Data Warehouse* est la suivante :

Le data Warehouse est une collection de données orientées sujet, intégrées, non volatiles et historisées, organisées pour le support d'un processus d'aide à la décision. [INMON, 1994]

Un entrepôt de données, ou data Warehouse, est une vision centralisée et universelle de toutes les informations de l'entreprise. C'est une structure (comme une base de données) qui a pour but, contrairement aux bases de données, de regrouper les données de l'entreprise pour des fins analytiques et pour aider à la décision stratégique. La décision stratégique étant une action entreprise par les décideurs de l'entreprise et qui vise à améliorer,

quantitativement ou qualitativement, la performance de l'entreprise. En gros, c'est un gigantesque tas d'informations épurées, organisées, historisées et provenant de plusieurs sources de données, servant aux analyses et à l'aide à la décision. L'entrepôt de données est l'élément central de l'informatique décisionnelle à l'heure où j'écris ce tutorial. En effet, l'entrepôt de données est le meilleur moyen que les professionnels ont trouvé pour modéliser de l'information pour des fins d'analyse, et il ne serait pas étonnant que d'ici quelques années un nouveau concept apparaisse pour révolutionner l'informatique décisionnelle. [GRIM, 2008]

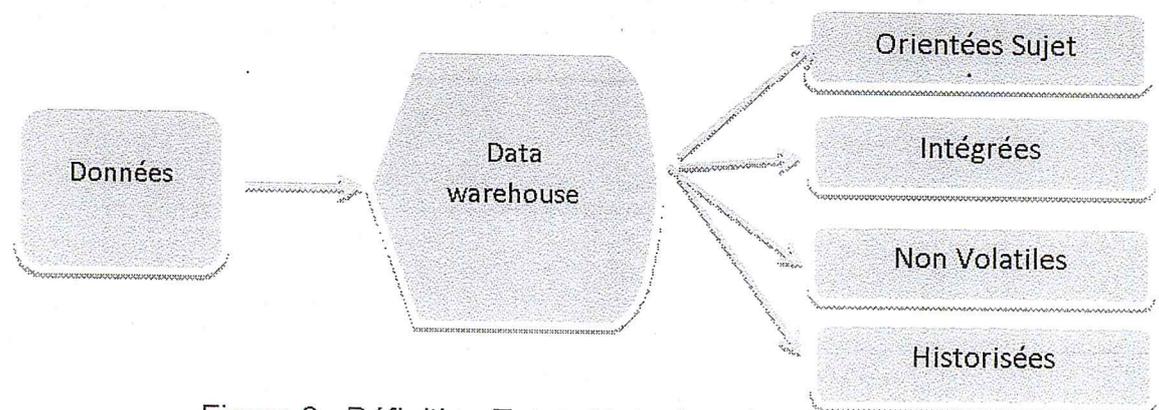


Figure 8 : Définition Entrepôt de données.

- **Données orientées sujet**

L'entrepôt de données est organisé autour des sujets majeurs et des métiers de l'entreprise. Les données sont organisées par thème, contrairement aux données des systèmes de production, organisées par processus fonctionnels. L'avantage de cette représentation demeure dans le fait qu'il devient possible de réaliser des analyses sur des sujets transversaux aux structures fonctionnelles et organisationnelles de l'entreprise. Et ainsi, de pouvoir analyser un processus dans le temps à différentes étapes de sa conception au sein du SI. Cette orientation permet également de faire des analyses par itération, sujet après sujet. L'intégration dans une structure unique est indispensable pour éviter aux données concernées par plusieurs sujets d'être dupliquées. Dans la pratique il existe également des Datamart<sup>3</sup> pouvant supporter l'orientation sujet. [DESNOS, 2007]

<sup>3</sup> Magasin de données, voir page : 29

Par exemple, une entreprise de commercialisation et de distribution de produits pétroliers peut avoir les processus fonctionnels suivants : ravitaillement, transfert, gestion des commandes client et transport produits. Alors que ses principaux sujets sont : le produit, la vente, l'approvisionnement et consommation interne.

- **Données intégrées**

Intégrer les données consiste à maîtriser leur hétérogénéité, pour donner au contenu de l'entrepôt de données une présentation homogène et pour garantir sa qualité. Avant d'être intégrées dans l'entrepôt de données, les données doivent donc être mises en forme et unifiées afin d'en avoir un état cohérent, associé à une description et un codage unique. Par exemple, la consolidation de l'ensemble des informations concernant un client est nécessaire pour donner une vue homogène de ce client. [GHOZZI, 2004]

Un Entrepôt de données est un projet d'entreprise et concerne les différents services et métiers de l'entreprise. L'intégration de données, au sein d'un entrepôt de données, est donc un processus déterminant sur la qualité et la quantité d'informations disponibles aux utilisateurs pour le processus de décision.

Cette phase, que nous verrons plus en détail avec les outils ETL, implique que les données doivent être mises en forme et unifiées afin d'avoir un état cohérent. Pour parfaire cette cohérence, l'intégration nécessite une forte normalisation de données.

- **Historisées**

L'historisation est nécessaire pour suivre dans le temps l'évolution des différentes valeurs des indicateurs à analyser. Ainsi, un référentiel temps doit être associé aux données afin de permettre l'identification dans la durée de valeurs précises.

En règle générale, dans un système de production, la donnée est mise à jour lors de chaque nouvelle transaction. Une mise à jour annule et remplace l'ancienne valeur. Dans un entrepôt de données, la donnée ne doit jamais

être mise à jour en mode « annule et remplace ». Un référentiel temps doit être associé à la donnée afin d'être capable d'identifier une valeur particulière dans le temps. [AGILTEAM, 2011]

- **Non volatiles**

La non volatilité des données est en quelque sorte une conséquence de l'historisation. Une même requête effectuée à quelque mois d'intervalle en précisant la date de référence de l'information recherchée donnera le même résultat. [SMAÏLI, 2008]

#### 4. Les composants d'un Data Warehouse

On présente les principaux composants de base d'un Data Warehouse, selon «Ralph KIMBALL» :

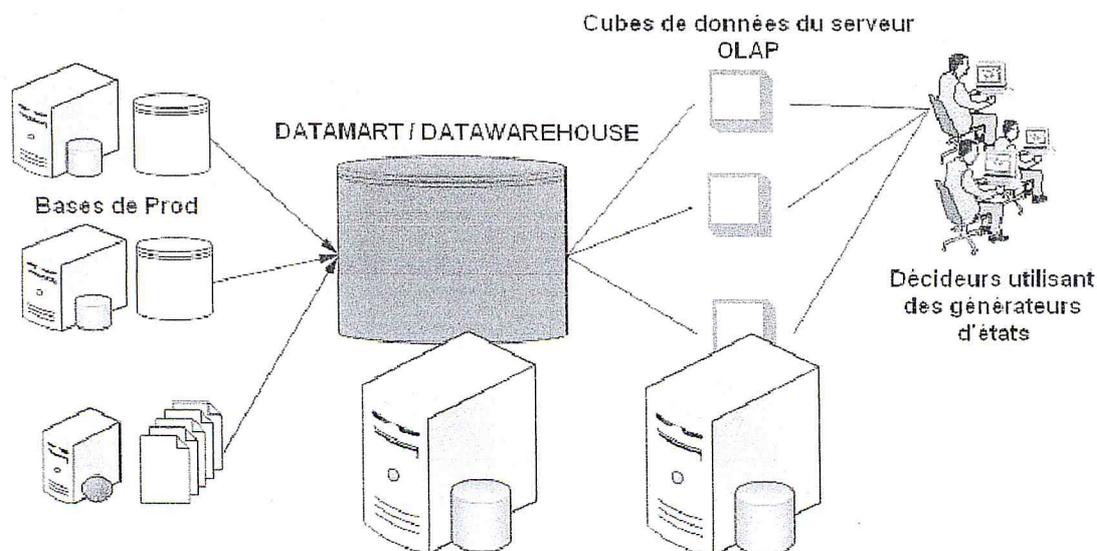


Figure 9 Composants de base d'un Data Warehouse. [TASLIMANKA, 2007]

- **Système source**

Le système source est le système opérationnel d'enregistrement dont la fonction consiste à capturer les transactions liées à l'activité. On entend généralement par système source les applications de gestion de l'entreprise, et dont la priorité est notamment le temps de disponibilité du système. Le postulat de départ consiste à considérer que ces systèmes opérationnels ne

contiennent que très peu de données historisées et qu' ne sont généralement pas orientés autour de la génération d'ils états. Ces systèmes sont souvent autonomes et sont peu homogènes entre eux. [CHALLIER, 2006]

- **Zone de préparation des données.**

La zone de préparation des données représente l'ensemble des processus qui nettoient, transforment, combinent, archivent, suppriment les doublons en vue de leur intégration et de leur exploitation dans le Data Warehouse. [KIMBALL, 2008] Quelle que soit la forme des données dans la zone de préparation (forme relationnelle, fichiers séquentiels) la frontière qui détermine cette zone de préparation est définie par l'absence de service de requête ou de service de présentation. Dès qu'un système offre de tels services, il doit être classifié dans la catégorie des systèmes de présentation.[CHALLIER, 2006]

- **Serveur de présentation**

Le serveur de présentation est la machine cible sur laquelle l'entrepôt de données est stocké et organisé pour répondre en accès direct aux requêtes émises par les utilisateurs, pour les générateurs d'états et les autres applications.[KIMBALL, 2008] C'est sur le serveur de présentation que les données sont présentées et stockées sous une est forme dimensionnelle.

- **L'entrepôt de données**

L'entrepôt de données correspond à la source de données interrogeable de l'entreprise. [KIMBALL, 2008] L'entrepôt de données est alimenté par la zone de préparation des données. Il est fréquemment mis à jour en tenant compte de la charge au fur et à mesure que les données sont corrigées, que les instantanés s'accumulent et que les libellés évoluent.

- **Le magasin de données**

Le magasin de données est défini comme un sous-ensemble logique d'un entrepôt de données. [KIMBALL, 2008] Il représente un projet réalisable. On considère souvent le magasin de données comme la réduction de l'entrepôt

de données à un seul processus ou à un groupe de processus ciblant un groupe métier spécifique.

Chaque Data Mart doit être représenté par un modèle dimensionnel, et au sein d'entrepôt de données, un tous ces Data Marts doivent être construits à partir de dimensions conformes et de faits conformes.

Les magasins de données sont basés sur des données détaillées et peuvent contenir des agrégats de données destinées à optimiser les performances, contrairement à la définition de Bill Inmon, qui conçoit le magasin de données comme un ensemble de données agrégées.

- **Portail de restitution**

C'est la part publique de l'entrepôt de données.[KIMBALL, 2008] Il représente ce que voient les utilisateurs, les outils avec lesquels ils travaillent. Sous ce terme sont regroupées toutes les applications qui s'appuient sur les données de l'entrepôt de données pour les restituer soit à l'utilisateur, soit à une autre application. Les services offerts par le portail de restitution sont les services d'accès aux données, les applications de modélisation et de datamining.<sup>4</sup>

Les services d'accès aux données comprennent : la navigation dans l'entrepôt, dans les méta-données, la surveillance de l'activité, la gestion des requêtes et la génération d'états standards. Les applications de modélisation offrent différents types d'analyses basées sur des modèles tels que modèles financiers, systèmes d'évaluation des clients, optimisation des processus et prévision, ainsi que les activités centrales du datamining telles que la catégorisation, la classification, l'estimation et la prédiction et finalement le regroupement par affinité.

- **Meta données**

Les métas données « données sur les données » présentes dans l'entrepôt, décrivent les caractéristiques des données stockées ; origine, date de dernière mise à jour, mode de calcul, procédure de transformation... Elle sont

---

<sup>4</sup> Fouille ou exploration sémantique de données.

utiles aussi bien aux utilisateurs qu'aux administrateurs. Elles permettent aux premiers de comprendre les données qui leurs sont présentées et aux seconds d'avoir les moyens d'exploitation et de maintenance de l'entrepôt de données. [HAOUIET, 2008]

## **5. L'alimentation d'un Data Warehouse:**

Comme pour n'importe quelle structure de données, il est nécessaire de pouvoir extraire les données, les stocker et d'y accéder. Pour cela il est indispensable d'identifier les données intéressantes de celles qui ne le sont pas.

Les données exploitées sont connues pour leurs hétérogénéités, car souvent on a affaire à plusieurs systèmes de gestion de bases de données. Elles sont connues aussi par leurs diffusions, c'est-à-dire leurs présences dans plusieurs environnements différents au même moment. En dernier ; la complexité de la donnée qui est due à la nature logique et physique.

Il est nécessaire de faire appel à des logiciels capable d'intégrer et d'homogénéiser, leurs rôles est de s'assurer en premier lieu de la cohérence des données du Data Warehouse.

### **5.1- Phases d'alimentation du Data Warehouse :**

Les plus importantes phases de l'alimentation d'un entrepôt de données sont les suivantes :

- Extraction.
- Préparation.
- Chargement.

#### **5.1-1. Découverte des données :**

Une phase cruciale dans l'alimentation de l'entrepôt de données, car elle détermine les données susceptibles d'être prises à partir du système opérationnel. Un mauvais choix des données va avoir des répercussions négatives sur les phases suivantes.

- **L'extraction :**

Elle consiste à collecter les données utiles dans le système de production à l'aide d'un outil spécialisé comme ETL (Extract, transform and load).

Pour être utiles, ces données nécessitent un reformatage pour pouvoir les incorporer dans une forme exploitable pour l'entreprise.

- **La Préparation :**

La préparation inclut la mise en correspondance des formats de données, le nettoyage, la transformation et l'agrégation. Les outils de nettoyage permettent de supprimer les doublons dans les fichiers. Il s'agit à ce stade d'appliquer des filtres prédéfinis sur les données afin d'attribuer des valeurs cohérentes aux variables mal ou non renseignées ou encore d'harmoniser les formats.

- **Le chargement des données :**

Le chargement constitue la dernière phase d'alimentation. Il est indispensable de maîtriser la structure du SGBD (tables et index) pour optimiser au mieux le processus. C'est une phase délicate notamment lorsque les volumes sont importants.

## 5.2- Exploitation des données du Data Warehouse :

Après alimentation de l'entrepôt de données il est intéressant d'exploiter ces données. L'utilisateur final, le décideur ou le manager doit pouvoir interroger les données à l'aide d'outils simples et conviviaux, qui leurs permettent de répondre à leurs nouveaux besoins.

On distingue un moyen très répandu pour l'exploitation des données d'un entrepôt :

- **Tableau de bord (Dash board) :**

Un tableau de bord est un instrument de mesure de la performance. Le tableau de bord contribue à réduire l'incertitude et facilite la prise de risque inhérente à toute décision. Un des objectifs du tableau de bord est de détecter une activité anormale. Pour ce faire, il utilise

généralement des figures et des couleurs permettant d'identifier rapidement le problème.[SMAÏLI, 2008]

## 6. Conclusion

Dans ce chapitre, nous avons vu comment un entrepôt de données permet au décideur de travailler dans un environnement informationnel, référencé, homogène et historisé. Cette technique l'épargne des problèmes liés à l'hétérogénéité des systèmes informatiques et à l'hétérogénéité des différentes définitions de données issues de l'historique de l'organisation.

Le Data warehouse (entrepôt de données) est un ensemble de données historisées variant dans le temps, organisé par sujets, consolidé dans une base de données unique, géré dans un environnement de stockage particulier, aidant à la prise de décision dans l'entreprise.

Les applications décisionnelles permettent ensuite d'extraire à partir d'entrepôt de données, une connaissance partielle de l'activité de l'entreprise selon les axes qui préoccupent le décideur à un instant donné.

*MODELISATION  
MULTIDIMENSIONNELLE*

Un modèle dimensionnel contient les mêmes informations qu'un modèle Entité/Relation, mais présente les données dans un format symétrique plus approprié pour faire l'analyse de données. La modélisation dimensionnelle est une approche dédiée aux systèmes décisionnels. Elle part du principe que l'objectif majeur de ce type de système est l'analyse de la ventilation de données quantitatives (les faits) par rapport à des données qualifiantes (les dimensions). [MIDOUNI, 2005]

Chaque modèle dimensionnel se compose d'une table contenant une clé multiple, la table des faits, et d'un ensemble de tables plus petites nommées tables dimensionnelles :

- **Tables de faits** : Une table de fait est une table qui contient les données observables (les faits) que l'on possède sur un sujet et que l'on veut étudier, selon divers axes d'analyse (les dimensions).

Les « faits », dans un entrepôt de données, sont numériques, puisque d'ordre quantitatif. Il peut s'agir du montant en argent des ventes, du nombre d'unités vendues d'un produit. Les tables de faits représentent des associations dont l'existence d'une occurrence dépend de l'existence des occurrences correspondantes dans les tables dimensionnelles, c'est-à-dire la table de fait contient l'ensemble des mesures correspondant aux informations de l'activité à analyser. Mais rappelons que certaines tables de faits peuvent contenir aucun attribut et représentent que des liaisons entre tables dimensionnelles.

Tous les éléments qui pointent sur la table de faits sont liés à une sémantique exprimable par une phrase. Par conséquent, la table de faits est la matérialisation d'une association entre  $n$  entités. [EURISE, 2008]

- Structure de base d'une table de fait

Une table de fait devrait avoir la structure suivante :

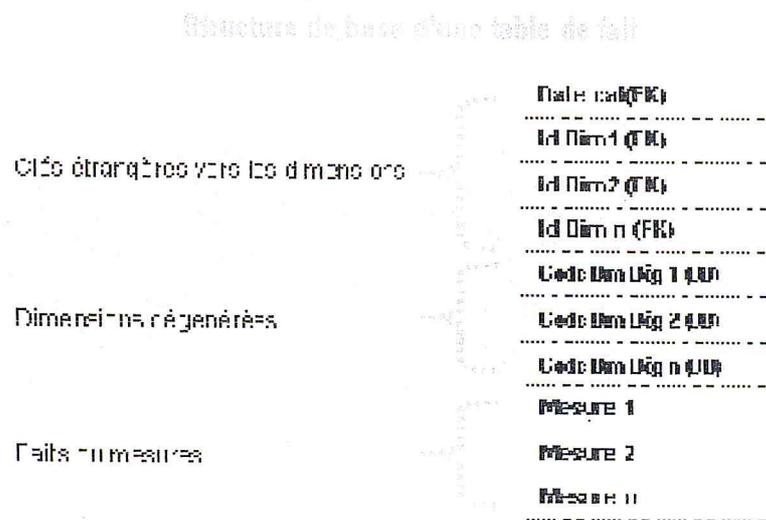


Figure 10 : Structure de base d'une table de faits. [DECIZIA, 2009]

- **Tables dimensionnelles** : est une table qui contient les axes d'analyse (les dimensions) selon lesquels on veut étudier des données observables (les faits) qui, soumises à une analyse multidimensionnelle, donnent aux utilisateurs des renseignements nécessaires à la prise de décision.

On appelle donc « dimension » un axe d'analyse. Il peut s'agir des clients ou des produits d'une entreprise, d'une période de temps comme un exercice financier, des activités menées au sein d'une société. [www.dwfacile.com]

- Structure de base d'une dimension

Une dimension devrait avoir la structure suivante :

### Structure de base d'une table de dimension

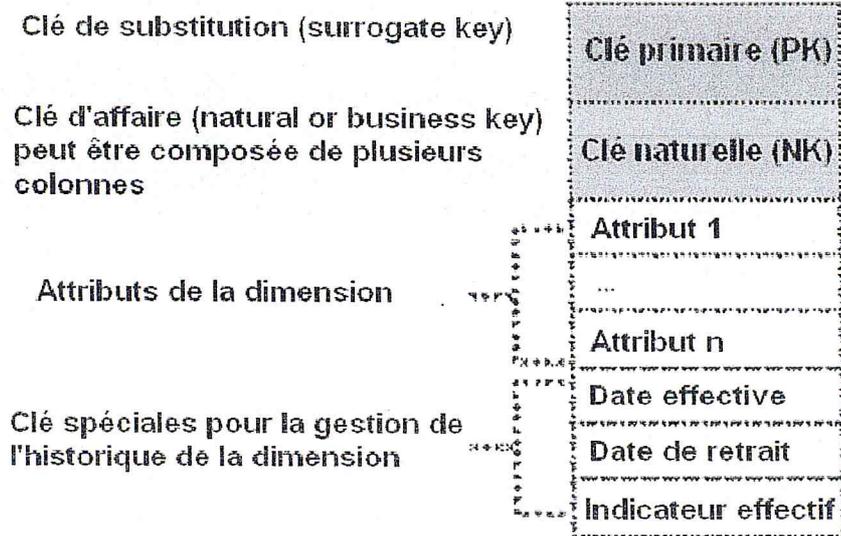


Figure 11 : Structure de base d'une table de dimension. [www.dwfacile.com]

#### 4. Schéma en étoile :

Un schéma en étoile consiste en une table de faits centrale et plusieurs tables dimensions. Les mesures d'intérêt pour l'OLAP<sup>5</sup> sont stockées dans la table des faits (eg., ventes, stock). Pour chaque dimension du modèle multidimensionnel il existe une table (eg., région, produit, temps). Cette table stocke les informations relatives aux dimensions. [SILHADI, 2000]

<sup>5</sup> OLAP : Online Analytical Processing, type d'application informatique orienté vers l'analyse sur-le-champ d'informations selon plusieurs axes, source : wiki.org

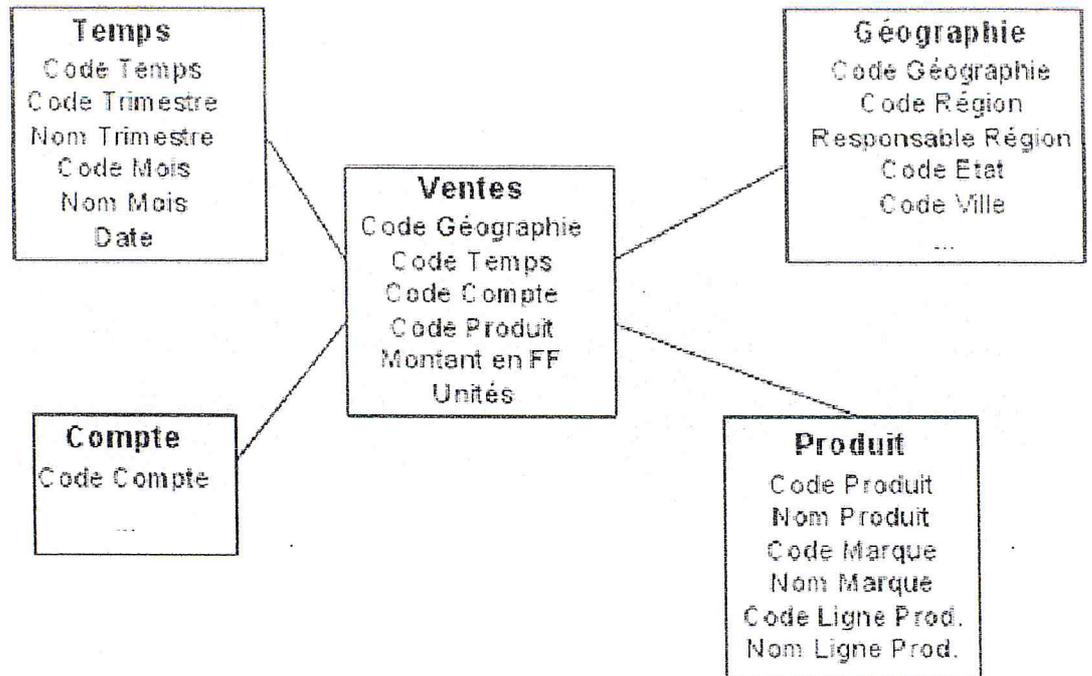


Figure 12 : Schéma en étoile.

- **Avantage :**

On dit souvent que le modèle en étoile est plus performant. Cela est dû au fait qu'il y a moins de jointures à faire que sur un modèle en flocons. C'était vrai il y a quelques années, quand les SGBD étaient moins performants. Actuellement la différence est minime, les SGBD gérant mieux les jointures multiples. De plus les jointures supplémentaires mettent en œuvre généralement des tables à faible volumétrie. [MALKOVICH, 2010]

- **Contre-performance :**

Une table de dimension en jointure réflexive dans un modèle en flocons est à proscrire. Par exemple si on a une table de hiérarchie décrivant la structure de l'entreprise, une colonne "père" permettant de faire le lien sur la même table. Le SGBD va lire plusieurs fois la même table, les accès concurrents se faisant sur le même disque les performances seront très dégradées. Il vaut mieux créer plusieurs tables différentes, chacune représentant un niveau de la hiérarchie.

Un autre schéma a été développé afin de surpasser ce problème, nommé : « schéma en flocon de neige ». [MALKOVICH, 2010]

### 5. Schéma en flocon de neige :

Le schéma flocon de neige consiste à avoir une table centrale et une table pour chaque niveau dans chaque dimension. Les attributs sont ainsi « normalisés » dans des « sous tables ». ce type de schéma est bien adapté aux Data warehouse volumineux, il convient quand on a beaucoup de hiérarchies et de grandes tables. [CANIVET, 2002]

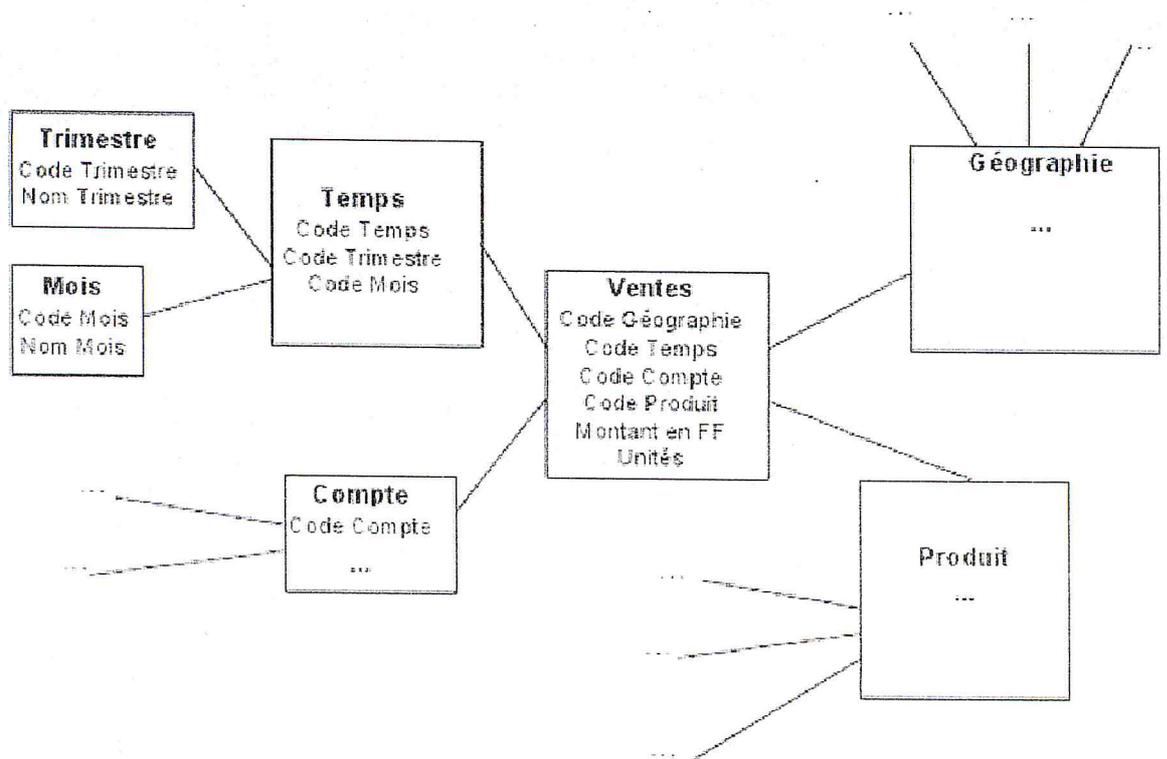


Figure 13 : Schéma en flocon de neige.

- **Avantage :**

Pas de redondance des données, gain d'espace mémoire sur le long terme.

- **Contre-performance :**

Le schéma perd en simplicité, et le parcourt du système devient plus complexe.

## 6. Passage du modèle relationnel au modèle multidimensionnel :

Les tables correspondant aux branches de l'étoile pourront alors être dépilées pour libérer la troisième dimension nécessaire à la construction de cubes et ainsi passer d'un monde relationnel à un monde multidimensionnel.

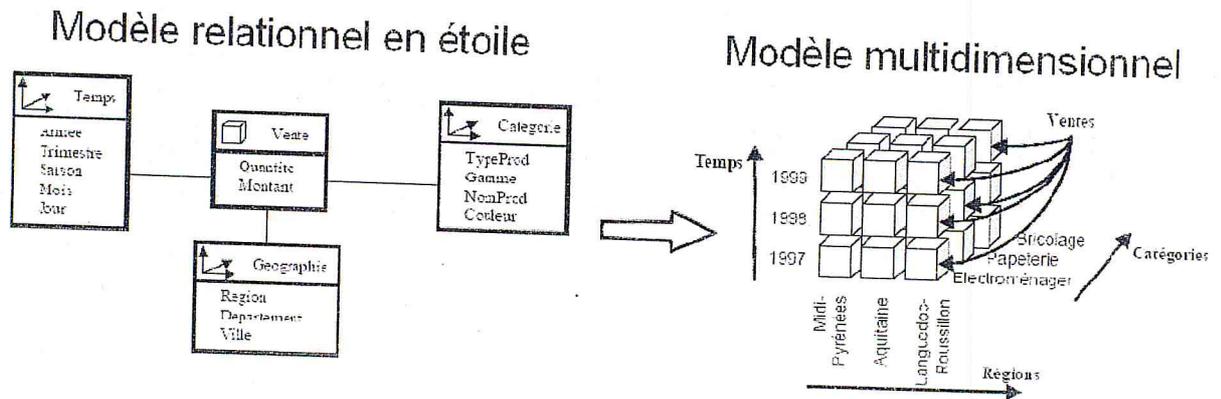


Figure 14 : Passage de l'étoile au cube.

## 7. Caractéristiques du modèle OLAP :

Les outils **OLAP** (*On Line Analytical Process*) reposent sur une base de données multidimensionnelle, destinée à exploiter rapidement les dimensions d'une population de données.

La plupart des solutions OLAP reposent sur un même principe : restructurer et stocker dans un format multidimensionnel les données issues de fichiers plats ou de bases relationnelles. Ce format multidimensionnel, connu également sous le nom d'hypercube, organise les données le long de dimensions. Ainsi, les utilisateurs analysent les données suivant les axes propres à leur métier.

Ce type d'analyse multidimensionnelle nécessite à la fois l'accès à un grand volume de données et des moyens adaptés pour les analyser selon différents points de vue. Ceci inclut la capacité à discerner des relations nouvelles ou non prévues entre les variables, la capacité à identifier les paramètres nécessaires à manier un volume important de données pour créer un nombre illimité de dimensions et pour spécifier des expressions et conditions inter dimensions. Ces dimensions représentent les chemins de consolidation. [HELOU, 2004]

Ceci a évolué pour aboutir à une méthode d'analyse, qui a permis aux décideurs de disposer d'un accès rapide et pertinent visualisé sous forme de Cube.

OLAP est conçue exclusivement pour l'analyse multidimensionnelle, avec un mode de stockage optimisé par rapport aux chemins d'accès prédéfinis. Ainsi, toute valeur d'indicateur associée à l'axe temps sera pré-calculée au chargement pour toutes ses valeurs hebdomadaires, mensuelles, etc. [HELOU, 2004]

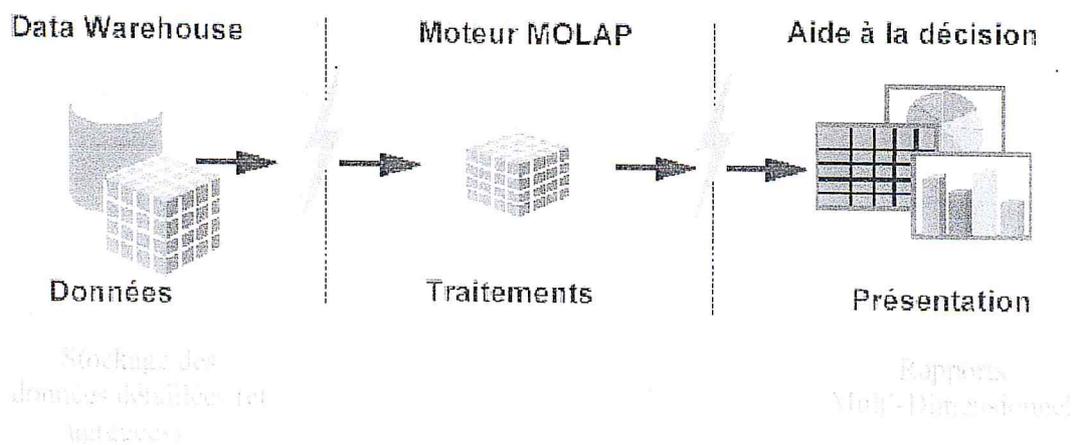


Figure 15 : Architecture OLAP Multidimensionnel. [HELOU, 2004]

### 8. Navigation au sein d'un cube OLAP :

Un des principaux points de la Business Intelligence est l'analyse multidimensionnelle. Il devient primordial de pouvoir étudier une même information suivant différents axes. Par exemple, il est très intéressant pour une entreprise de pouvoir interpréter son chiffre d'affaires par zone géographique, par période temporelle, par produit, par client. [ROUSSET, 2009]

Les outils OLAP utilisent des opérateurs particuliers pour la navigation dans les hypercubes.

**Roll-up** Passage de mesures détaillées à résumées en remontant dans la hiérarchie de la dimension.

**Drill-down** Descendre dans la hiérarchie de la dimension.

**Rotate** Rotation des axes du cube pour fournir une vue alternative des données

**Slicing** Extraction d'une tranche d'informations : Sélection d'une dimension pour passer à un sous-cube.

**Dice** Extraction d'un bloc de données : Sélection de deux ou plusieurs dimensions.

**Drill-across** Exécution de requêtes impliquant plus d'un cube ayant une dimension commune. [DIEUNGANG, 2009]

### 9. Conclusion :

Dans ce présent chapitre nous avons introduit les notions de modélisation multidimensionnelles et nous avons élucidé sa relation avec les systèmes décisionnels ainsi que les besoins pour l'établissement de notre modèle.

Cette introduction dans les approches dimensionnelles va nous mener à étudier dans les prochains chapitres les démarches de modélisation orientées objets ainsi que l'établissement des besoins des décideurs. Notre projet porte sur la conception d'un système décisionnel (tableau de bord se focalisant spécialement sur l'information géographique), s'inscrivant dans un environnement décisionnel axé sur une modélisation multidimensionnelle.

*SYSTEME D'INFORMATION  
GEOGRAPHIQUE*

## 1. Introduction :

Les outils de prise de décision commerciale (CRM, décisionnels, data mining etc), mettent en évidence les corrélations existantes entre le comportement d'achat des clients et certaines de leurs caractéristiques individuelles. La localisation géographique est pratiquement toujours la caractéristique la mieux renseignée. Il en découle essentiellement :

Les **SIG** (Système d'Information Géographique), qui permettent de visualiser sur des cartes géographiques des données statistiques, comptables, voire factuelles.

Il y a quelques années la mise en place de SIG paraissait réservée aux grandes structures. Aujourd'hui les progrès informatiques et les possibilités offertes en matière de gestion et d'analyse, conduisent à la généralisation de l'outil à tous les échelons du territoire. Cependant la mise en place d'un SIG, quelle que soit la taille du territoire, reste un investissement humain et financier important.

## 2. Définition :

- Un SIG (Système d'Information géographique) est un Système informatique permettant, à partir de diverses sources, de rassembler et d'organiser, de gérer, d'analyser et de combiner, d'élaborer et de présenter des informations localisées géographiquement, contribuant notamment à la gestion de l'espace. [SFPT, 1989]
- Un système d'information géographique est un "système informatique de matériels, de logiciels, et de processus conçus pour permettre la collecte, la gestion, la manipulation, l'analyse, la modélisation et l'affichage de données à référence spatiale afin de résoudre des problèmes complexes d'aménagement et de gestion". [FICCDC, 1988]
- Un système d'information géographique est un "ensemble de données repérées dans l'espace, structuré de façon à pouvoir en extraire commodément des synthèses utiles à la décision". [MICHEL, 1990]

Les Systèmes d'Information Géographique (SIG) ont donné naissance à une discipline récente, la géomatique, combinant la géographie et l'informatique.

De manière générale un logiciel cartographique permet de collecter, combiner et utiliser des informations géographiquement localisées, qu'elles soient sous forme d'images (raster<sup>6</sup>) ou de données vectorielles et de les représenter sur des cartes.

### 3. La donnée géographique :

Avant d'aller de l'avant, il est crucial d'éclaircir certains points sur la donnée géographique elle-même. De très nombreuses données ont un caractère spatial, comme par exemple, des adresses, des codes postaux, des pays, des coordonnées latitudinales et longitudinales.

De manière générale, il s'agit donc d'associer aux objets une thématique et leur localisation sur la surface terrestre. Ce second critère est essentiel, il permet de définir des relations topologiques entre les objets.

En ajoutant la dimension temporelle, généralement déjà bien gérée dans les systèmes décisionnels : « L'information géographique est la représentation d'un objet ou d'un phénomène réel, localisé dans l'espace à un moment donné » [QUODVERTE, 2003]

Toute information géographique doit préciser son système de localisation. Il peut s'agir du type direct (mathématique) : tout point terrestre peut, en effet, être repéré par ses coordonnées latitude et longitude. Pour le type indirect, la description est sémantique, indiquant le nom de l'endroit (exemple : adresse postale) ou l'itinéraire pour s'y rendre. Afin de permettre le passage d'une référence indirecte à une référence directe, les descriptions textuelles sont classées dans des répertoires, diffusés par des organismes spécialisés.

Pour représenter le monde réel, trois primitives géographiques peuvent être utilisées : le point, la ligne et la surface. Grâce à ces entités géométriques, on peut les assembler.[TRANCHANT, 2011]

---

<sup>6</sup> Définition en page §

#### 4. Caractéristiques des données géographiques :

Les données géographiques présentent un certain nombre de caractéristiques spécifiques lors de leur modélisation et de leur traitement informatique comme la richesse de leur sémantique, l'importance de la notion de précision liée notamment aux procédures de saisie, et l'existence d'une multiplicité de représentations géométriques (figure 1).

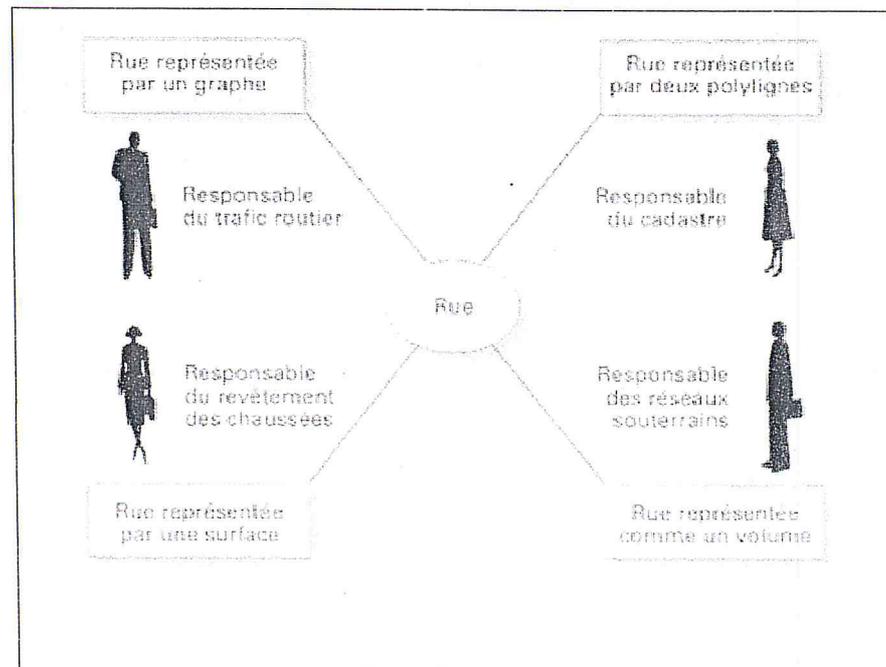


Figure 16 : différentes visions d'une rue. [LAURINI, 1993]

Dans la réalité, tous les objets occupent un certain volume, mais dans un SIG, les objets sont symbolisés par des objets ponctuels, linéaires, surfaciques ou volumiques. Ainsi, un bâtiment est considéré comme un polygone complexe dans un SIG à grande échelle (1/500) et un point dans un SIG à petite échelle (1/50 000). Une rivière est représentée comme un objet surfacique à grande échelle et comme un objet linéaire à petite échelle. D'une manière générale, le choix de l'échelle et du type de représentation influe sur la qualité des données et des résultats. [LAURINI, 1993]

La figure 1 donne un exemple de rue, dont la représentation géométrique varie selon les utilisateurs (multi-représentation) :

- le responsable du trafic routier peut voir le réseau routier comme un graphe de manière à établir les plans de circulation ou de jalonnement ;
- le responsable du Service du Cadastre modélise la rue comme deux polygones délimitant ses deux côtés ;
- la personne chargée du revêtement des chaussées, de manière à déterminer les quantités de goudron, a besoin de connaître la surface de la rue, donc la voit comme une aire ;
- enfin, le chef de service des réseaux souterrains (eau, assainissement, gaz, etc.), en vue de gérer le sous-sol, doit voir la rue comme un volume.

Un des aspects particuliers est la notion de localisant. Par localisant, on entend des informations permettant de localiser les objets dans l'espace (adresses postales, coordonnées-Lambert, etc.).

- **Une Information Localisée :**

L'objet géographique est à la base du SIG, et est le support d'autres données et / ou le référentiel permettant de positionner d'autres objets. La donnée est géographique dès lors qu'elle est localisable directement par des coordonnées, ou indirectement par des données littérales de type adresse, numéro de commune, numéro de borne kilométrique, code postal, numéro de parcelle cadastrale, ... [IAAT, 2003]

Support pour de la donnée littérale.

parcelle n° : 920  
section : ZK  
Propriétaire : Mme Dubois  
Adresse : 22, rue des lilas  
Commune : Bord de Mer  
N° INSEE : 99395  
Code postal : 99140



Référentiel permettant de positionner des objets de façon relative

Borne à incendie N° 211 se trouvant à la limite du 20 et 22, rue des lilas, côté droit.

Figure 17 : l'objet géographique support et/ou référentiel. [IAAT, 2003]

Une Information localisée sur quoi ?

- Les coordonnées terrestres :

La terre est une sphère, mais une sphère imparfaite :

- Le Géoïde est la forme théorique qui se rapproche le plus de la surface réelle de la terre c'est-à-dire le niveau moyen des mers. Il sert de référence pour déterminer les altitudes.

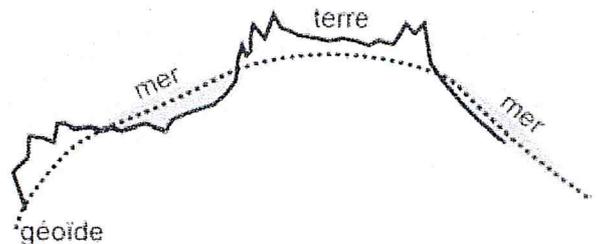


Figure 18 : le géoïde est la forme théorique de la terre. [IAAT, 2003]

- L'ellipsoïde est la surface mathématique qui se rapproche le plus de la forme du géoïde ; grâce à ce dernier, on peut calculer les coordonnées géographiques en LONGITUDE et en LATITUDE.

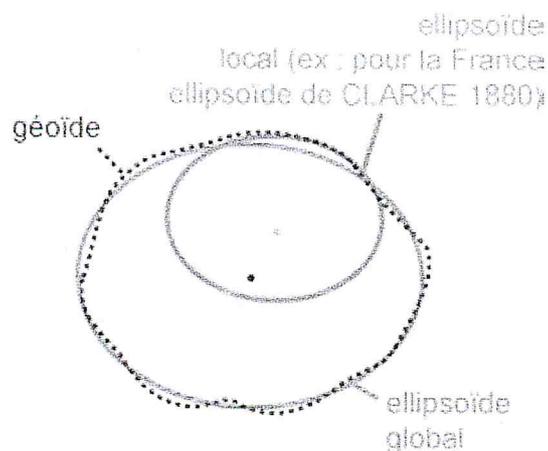


Figure 19 : l'ellipsoïde. [IAAT, 2003]

- Le procédé mathématique qui permet le passage de l'ellipsoïde au plan se nomme : **système de représentation plane** ou **système de projection** ou **projection**.

Cette transformation ne va pas sans déformation (linéaire, surfacique, angulaire) [IAAT, 2003]

### 5. Classement des projections d'après les altérations

- La déformation linéaire : aucune projection ne conserve sur la carte toutes les longueurs.
- La déformation angulaire : les projections conformes conservent les angles au détriment des surfaces. Elles sont utilisées pour des données à grande échelle de type topographiques.
- La déformation surfacique : les projections équivalentes conservent les surfaces mais pas les angles. Elles sont utilisées pour le cadastre et pour des données à petite échelle. [IAAT, 2003]

Les projections dites aphylectiques<sup>7</sup> ne conservent ni les angles ni les surfaces mais sont un compromis compensant au mieux les altérations, utilisées pour les représentations de type planisphère.

Il existe de nombreuses projections et les algorithmes mathématiques des représentations planes permettent de créer de nouvelles quasiment à l'infini. [Botton, 2004]



Figure 20 : altérations différentes selon les systèmes. [Botton, 2004]

<sup>7</sup> Voir figure n°8

## 6. Type de canevas de représentation des méridiens et des parallèles : [Botton, 2004]

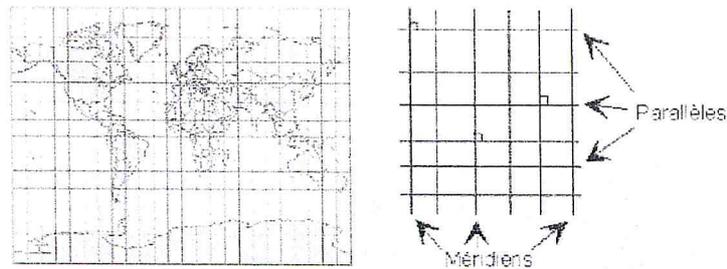


Figure 21 : canevas cylindrique.

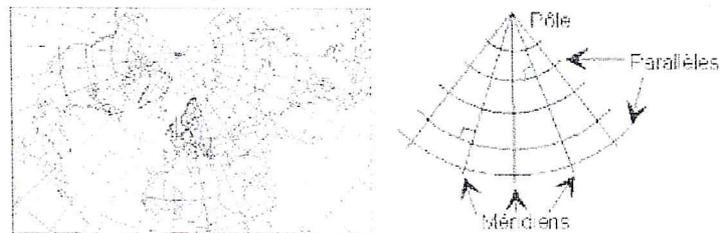


Figure 22 : canevas conique.

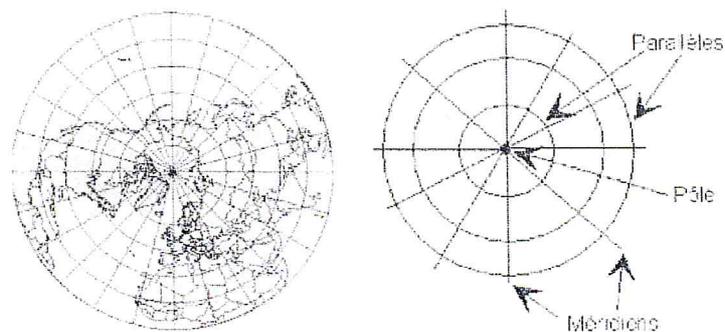


Figure 23 : canevas azimutal (aphylactique).

- **La notion d'échelle**

L'échelle est le rapport de taille entre deux réalités géographiques par une approche multi-scalaire. Mais elle est surtout utilisée en cartographie où elle renvoie au rapport de réduction entre l'objet référent du terrain et la carte (objet référé). Ce rapport mesure une transformation réalisée par

l'intermédiaire de la sémiologie graphique propre à la discipline cartographique. [LEVY, 2005]

La donnée numérique et les outils de zoom des logiciels permettent une grande liberté dans les échelles de visualisation de la donnée.

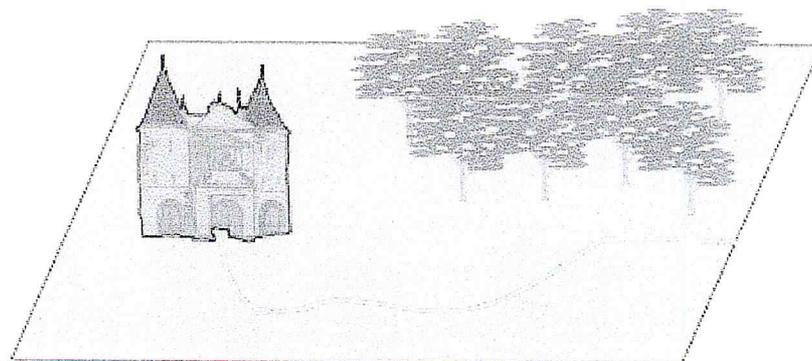
Avec les SIG, on parle plutôt d'échelle d'utilisation, c'est à dire le ratio entre l'échelle à laquelle la donnée a été numérisée et les limites de son exploitation.

L'échelle d'utilisation dépend de l'usage de la base de données, de la précision géométrique ainsi que de l'exhaustivité des informations géographiques voulues. [IAAT, 2003]

## 7. Mode de représentation de l'information géographique dans un SIG

- **Données raster :**

La réalité est décomposée en une grille régulière et rectangulaire, organisée



Ce que vous voyez sur le terrain

Figure 24 : couche raster d'une zone.

en lignes et en colonnes, chaque maille de cette grille ayant une intensité de gris ou une couleur. La juxtaposition des points recrée l'apparence visuelle du plan et de chaque information. Une forêt sera "représentée" par un ensemble de points d'intensité identique. [IRD, 2001]

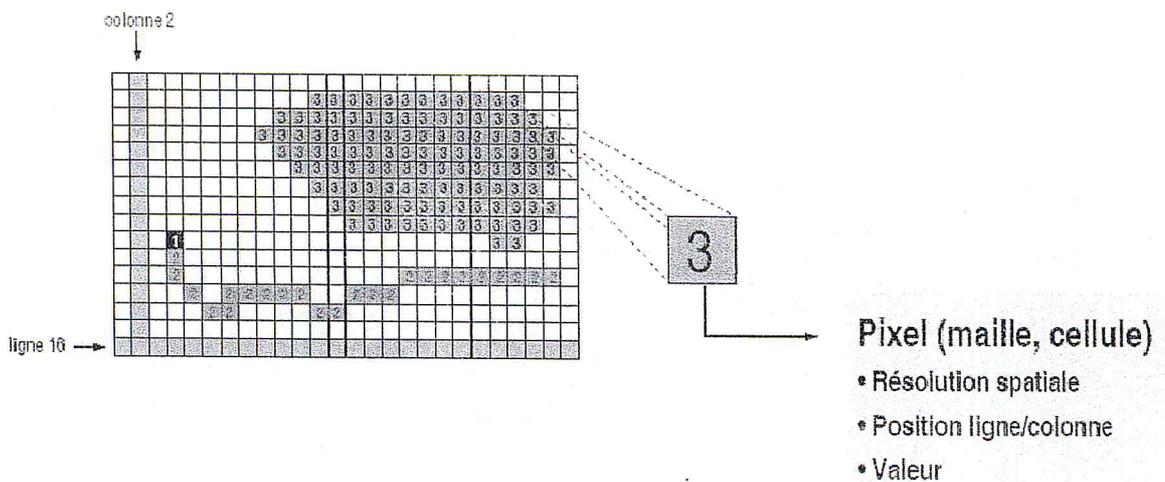


Figure 25 : image pixélisée. [IRD, 2001]

• **Photo aérienne :**

La photo est la base de nombreuses données géographiques. A partir d'appareil photo ou de caméra aéroportée (avion, ballon, ...) il est possible d'avoir de nombreux détails de la surface de la terre. Elle peut être :

- scannée,
- numérique (directement intégrable sur un disque dur),
- orthorectifiée (corrigée des déformations d'échelle dues aux différentes altitudes, à l'assiette de l'avion, ... on obtient une orthophotographie). [IAAT, 2003]

La précision de la photo aérienne dépend de la dimension du plus petit détail visible (notion de résolution).

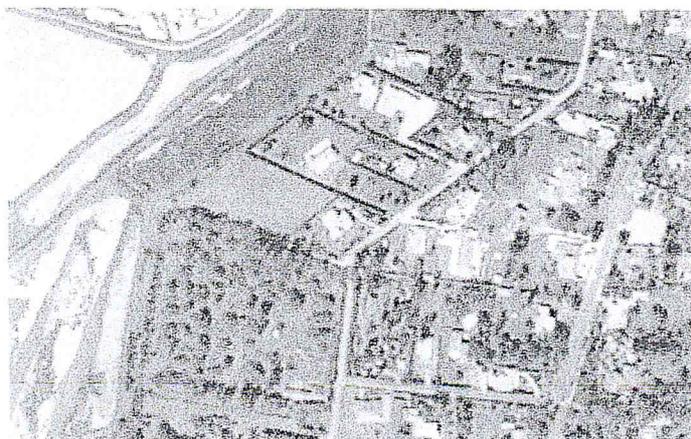


Figure 26 : photo aérienne. [IAAT, 2003]

- **Plan scanné ou carte scannée :**

C'est la représentation d'une information déjà interprétée. Ceci montre ses limites. Par contre, la carte scannée est un bon référentiel visuel car elle est souvent issue de carte papier destinée au grand public.



Figure 27 : carte scannée.[IAAT, 2003]

- **Image satellitale :**

Image issue de capteurs embarqués dans des satellites d'observation placés sur des orbites de 500 à 36000 km d'altitude. L'image représente le rayonnement solaire réfléchi par les objets au sol dans le domaine visible ou proche infra-rouge. Elle doit subir plusieurs traitements radio-métriques et géométriques avant d'être utilisable dans un SIG.

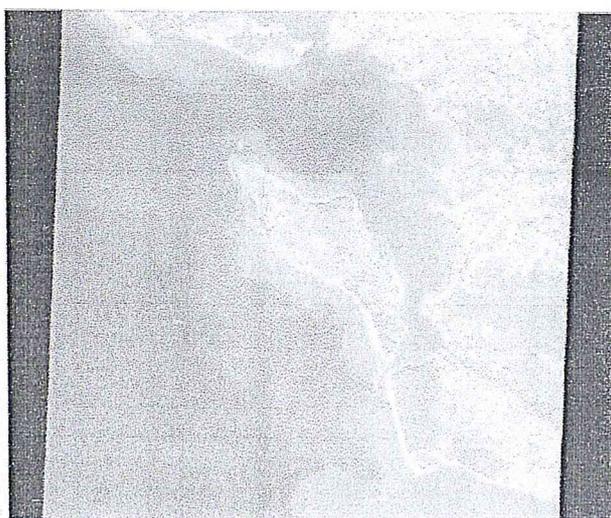


Figure 28 : image satellitale.[IAAT, 2003]

- **Image satellitale radar :**

Image enregistrée par des capteurs embarqués dans des satellites d'observation Elle représente la réponse du sol à l'onde envoyée par le capteur (principe du flash ou du sonar).



Figure 29 : image satellitale radar. [IAAT, 2003]

- **Données vectorielles :**

Les limites des objets spatiaux sont décrites à travers leurs constituants élémentaires, à savoir les points, les arcs, et les arcs des polygones. Chaque objet spatial est doté d'un identifiant qui permet de le relier à une table attributaire.[IRD,2001]

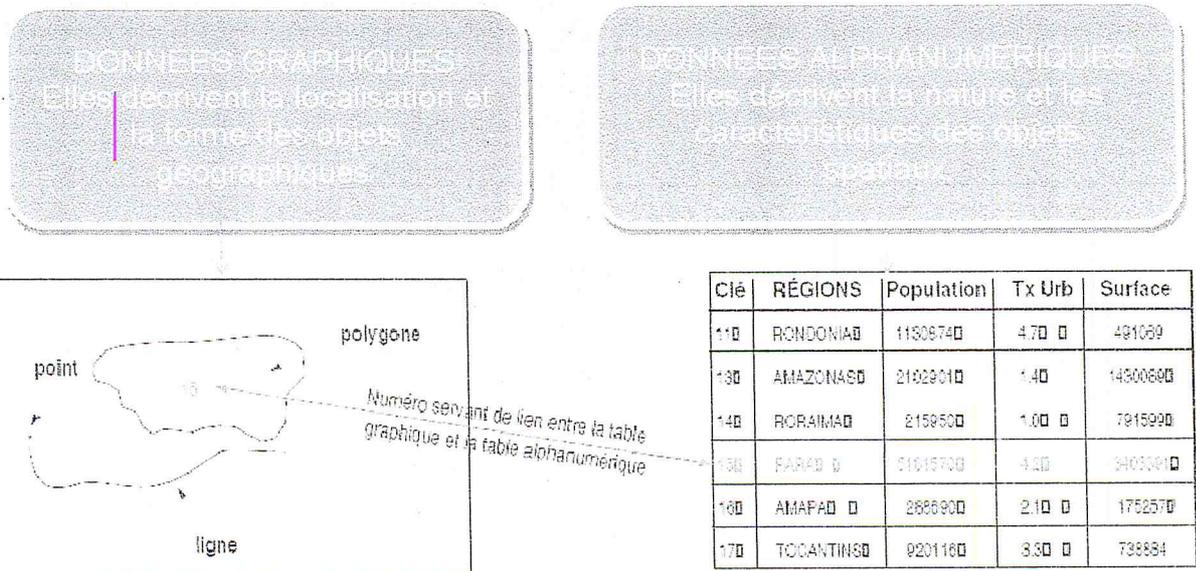


Figure 30 : Représentation alphanumérique d'une donnée géographique

Pour représenter les objets à la surface du globe, les SIG utilisent trois objets géométriques qui sont le point, la ligne et la surface.

#### **Le point :**

L'objet le plus simple, il peut représenter à grande échelle des arbres, des bornes d'incendie, des collecteurs d'ordures, .... Mais à des échelles plus petites de type carte routière au 1/1 000 000ème, il représente une capitale régionale. [IRD,2001]

#### **La ligne :**

La ligne représente les réseaux de communication, d'énergie, hydrographiques, d'assainissement, etc.. Elle peut être fictive, en représentant l'axe d'une route, ou virtuelle en modélisant des flux d'information, d'argent, ... [IRD,2001]

#### **La surface :**

Elle peut matérialiser une entité abstraite comme la surface d'une commune ou des entités ayant une existence géographique comme une forêt, un lac, une zone bâtie, ... [IRD,2001]

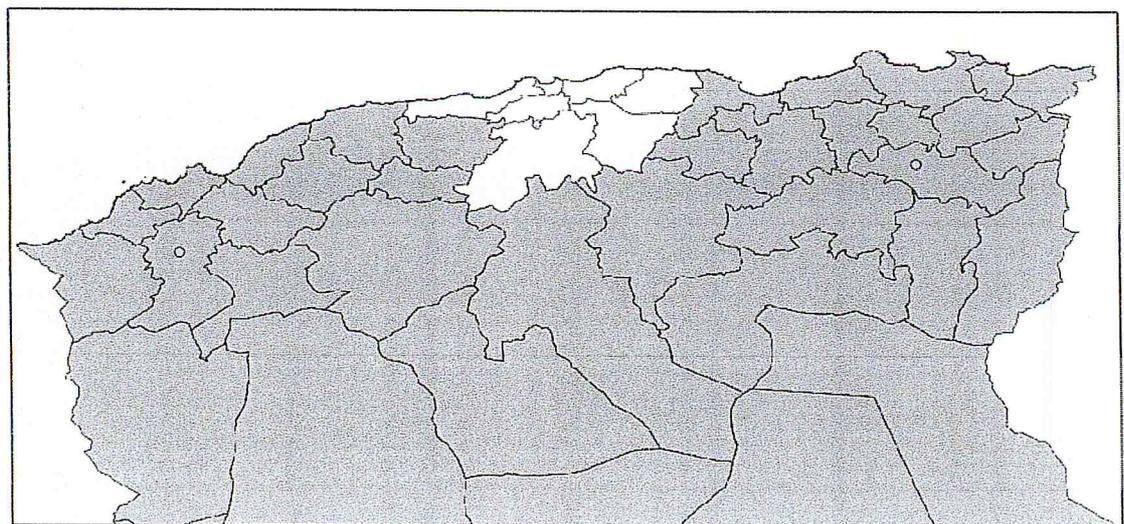


Figure 31 : points, polygones et surface Algérie nord.

## 8. Les domaines d'application :

Les domaines d'application des SIG sont aussi nombreux que variés.

[IRD,2001]

Citons cependant :

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

## 9. Conclusion

Ce chapitre a permis de mettre en évidence le fait que les SIG constituent un domaine à part entière, aux contours délimités, qui se démarque de la cartographie ou des simples SGBD<sup>8</sup>. L'étude de ce domaine était enrichissante et avantageuse car elle a permis une meilleure compréhension de la donnée géographique, cette étude s'est faite au sens large et superficiel de ce domaine très profond et dont les experts ne cessent d'agréments depuis 1969. Les cartes créées avec un SIG peuvent désormais facilement intégrer des rapports, des vues 3D, des images photographiques et toutes sortes d'éléments multimédias. Notre projet ne comportera pas tous les aspects vu dans ce chapitre, ces notions ont été abordées à titre profitable et instructif.

---

<sup>8</sup> Système de gestion de base de données.

Après avoir défini les concepts de notre étude dans la première partie de ce mémoire qui est Etat de l'art, nous nous étalerons dans la prochaine partie sur l'étude des besoins et la conception.

*CONCEPTION  
DE LA SOLUTION*

# *ETUDE DES BESOINS*

## 1. Introduction :

Les besoins des décideurs sont au cœur de la préoccupation du développeur pour la conception de l'entrepôt de données, il est donc nécessaire de se pencher sur ce sujet bien en profondeur en se rapprochant des décideurs et de comprendre le mécanisme actuel de leurs entreprises, leurs façon d'analyser les activités, et tenter d'apporter un plus, qui est souvent très méconnu des décideurs.

### 1. Création de questionnaire

- Définition des besoins

- Conception du questionnaire :

Nous avons essayé à l'aide de ce questionnaire d'extraire l'ensemble des axes et indicateurs pouvant nous aider dans la conception de notre modèle dimensionnel. Ce dernier a été exprimé aux décideurs sous forme d'interview. Pour atteindre les objectifs tracés nous effectuons une enquête aux près de plusieurs personnes de différentes classes hiérarchique.

Notre projet de développement d'entrepôt de donnée ainsi que son système décisionnel géographique a été initié dans le but de répondre aux exigences des décideurs de SARL Venus©. L'un des objectifs de ce projet est d'aider ces décideurs à avoir une vision plus précise de leurs propres besoins. L'entreprise a fait appel à nous dans le but de l'aider dans la mise en œuvre d'un système décisionnel. Ceci lui permettra d'une part de devenir leader dans le marché du cosmétique et d'une autre part l'aider à comprendre ses propres besoins. Notre enquête a révélé que la plupart des décideurs ne savent pas ce qu'ils veulent ou ce qu'ils attendent de leurs propres système décisionnel, cette constatation nous a mené à nous approcher de plusieurs entreprises pour pouvoir aider la nôtre à cerner ses propres besoins.

Notre enquête a ciblé trois entreprises de secteurs différents, Venus© pour la cosmétique, Sicam© pour l'agroalimentaire et Peugeot© pour les

voitures. Le point en commun de ces trois PME/PMI<sup>9</sup> est la vente de produits finis à une gamme de clientèle très large sur un territoire géographique vaste.

La première partie de cette enquête a été alloué au recueil d'informations, l'approche des décideurs et le rapport des discussions. Bien que nous portions un grand intérêt à la fiabilité de cette enquête, il est impossible de vérifier les affirmations des décideurs à l'échelle de l'entreprise.

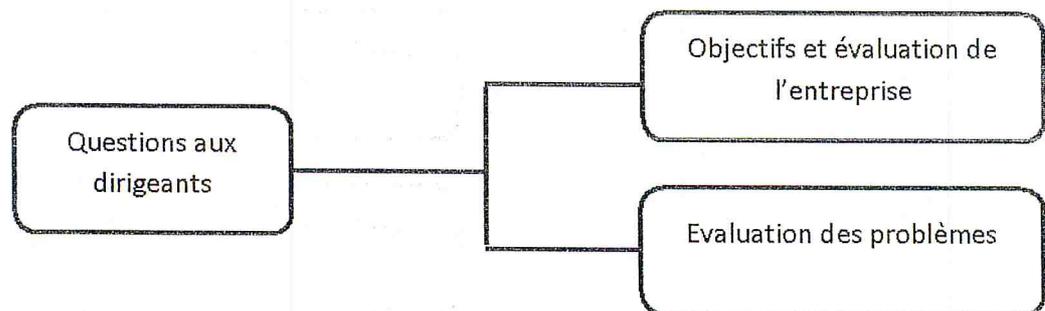
Nous avons divisé le questionnaire<sup>10</sup> de l'enquête en 2 parties :

1. Questions aux dirigeants.
2. Questions aux responsables.

Le choix de cette répartition est pour alléger le volume de questions pour chaque corps de l'entreprise, et cibler les secteurs d'activité.

### 3. Questions relatives aux dirigeants :

Nous avons rencontré les cadres des entreprises citées auparavant dans le but de récolter les informations qu'on juge nécessaire. Les questions relatives aux dirigeants se trouvent dans l'Annexe A et se divisent en deux parties :



Le tableau suivant résume les objectifs et les attentes des questions relatives aux dirigeants :

<sup>9</sup> PME/PMI : Petites et moyennes entreprises /petites et moyennes industries.

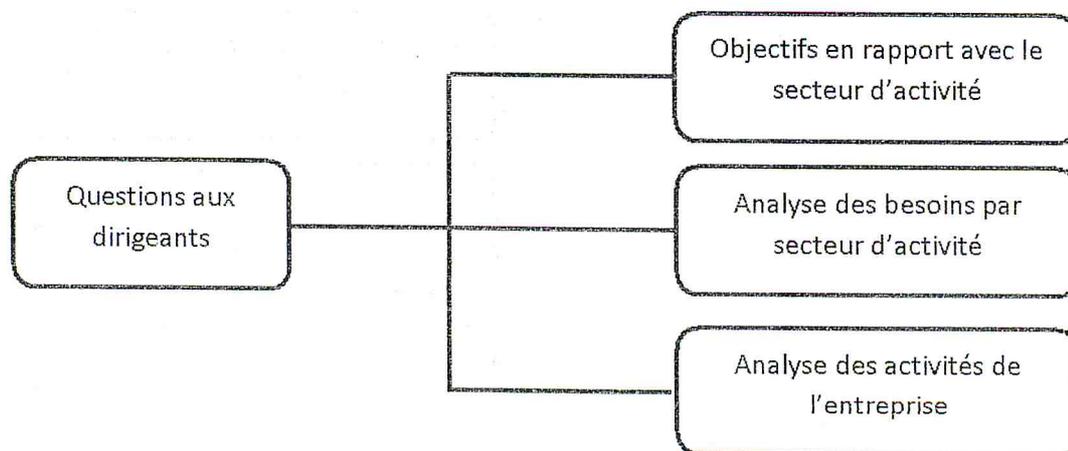
<sup>10</sup> Une copie du questionnaire se trouve en Annexe 1.

Thème des questions	BUT
<ul style="list-style-type: none"> <li>• <b>Objectifs et évaluation de l'entreprise</b></li> </ul>	<p>Nous cherchons à savoir quels sont les objectifs de l'entreprise, et qu'est-ce que cette dernière essaye d'accomplir. Il est nécessaire de comprendre et connaître les méthodes et fréquences d'évaluation des décideurs et leurs interprétations de la réussite.</p>
<ul style="list-style-type: none"> <li>• <b>Evaluation des problèmes</b></li> </ul>	<p>Pour comprendre en profondeur les attentes des dirigeants il est important de connaître la nature des problèmes dont les dirigeants font face.</p>

Tableau 2 : Questions aux dirigeants.

**4. Questions aux responsables :**

Questions relatives aux objectifs d'un service donné (Analyse, prédiction...). On note que les questions ne diffèrent pas beaucoup des questions adressées aux dirigeants de l'entreprise, elles portent essentiellement sur l'aspect technique de la gestion de l'information et la façon avec laquelle les responsables analysent leurs activités et préviennent les risques. Ces questions se divisent en trois parties :



Thème des questions	But
<b>Objectifs en rapport avec le secteur d'activité</b>	Les réponses révèlent la manière avec laquelle les responsables décrivent leurs activités, la façon avec laquelle ils procèdent à évaluations ainsi que les prestations qu'ils offrent. Toutes ces informations constituent des caractéristiques primordiales pour l'établissement de l'entrepôt de données.
<b>Analyse des besoins par secteur d'activité</b>	Les réponses nous donnent une idée sur le type d'analyse effectuées au sein de l'entreprise, également les analyses souhaitées dans le futur.
<b>Analyse des activités de l'entreprise</b>	Les réponses apportées par les responsables nous ont permis de cerner notre travail, ainsi que de déterminer à partir de ces dernières le fruit même de la modélisation multidimensionnelle, comme les faits et mesures.

Tableau 3 : Questions aux responsables.

### 5. Compte Rendu de l'entretien :

Après chaque entretien avec les décideurs, nous avons dressé un tableau récapitulatif qui regroupe le contenu des interviews par thème.

Compte rendu d'entretien
<p><b>Objectif :</b></p> <p>Les informations fournies par l'étape <i>étude des besoins</i> proviennent des interviews réalisées avec les décideurs des entreprises Venus, Sicam et Peugeot. Elles sont venues compléter le manque d'anticipation décisionnelle remarqué auprès de l'entreprise qui nous a octroyé le projet, Venus.</p> <p>Nous avons déchiffré les réponses pour déduire des axes d'analyse et les</p>

activités majeurs de l'entreprise.

### Les objectifs commerciaux

Les activités clés des décideurs de Venus concernant la fonction commerciale sont la vente, la gestion du stock, et la gestion de la commande.

### Les métriques d'évaluation :

Les décideurs de l'entreprise souhaitent employer les informations récoltées par les activités précédentes pour mieux anticiper les tendances du marché et comprendre les clients, le mode de paiement, le produit.

On retrouve si dessous quelques exemples de questions que se pose un décideur :

- Je voudrais connaître le chiffre d'affaire par rapport à l'article, au client, la période, et la zone géographique ou se vend ce produit.
- Je voudrais connaître le nombre d'articles que j'ai vendu à une période donnée par rapport au client, la zone géographique, le mode de paiement et le Mois.

### La dimension métier :

Un décideur se posera par exemple une question sur le produit en lui-même :

- Quels sont les résultats de ventes par famille de produit par zone géographique, par client, durant un mois donné.

Nous présentons un besoin dans la figure suivante sous forme de cube.

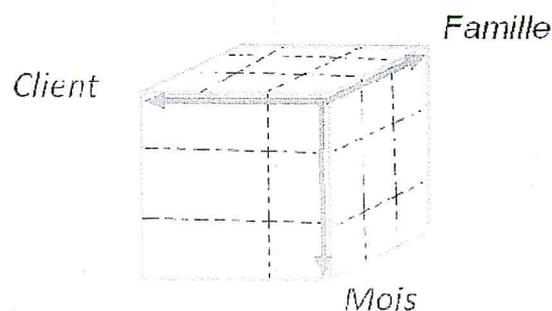


Figure 32 : Représentation sous forme de cube

A partir des comptes rendus d'entretiens avec les décideurs et responsables des entreprises citées auparavant, nous avons déduit les dimensions

suivantes en concordance avec les informations mises à notre disposition par l'entreprise VENUS, car en absence de données concrètes dans la base de données source, l'analyse multidimensionnelle est impossible.

## 6. Composants d'un entrepôt de données :

L'entrepôt de données se compose essentiellement de **dimensions** qui sont les points de vue sur lesquelles la mesure est observable. De **faits** qui présentent la valeur d'une mesure selon les dimensions citées, l'entrepôt se compose aussi d'indicateurs qui sont les valeurs comparables. Ces valeurs concrétisent l'analyse en fonction des dimensions.<sup>11</sup>

### a. Description des dimensions :

Après l'étude des réponses auxquelles les décideurs nous ont fait part. Nous nous sommes penchés sur la représentation des dimensions de notre Data Warehouse. Le tableau suivant nous présente ces dernières :

Dimension	Description
<b>Temporelle</b>	Dimension présente dans tous les Data Warehouse, car elle présente un aspect majeur de ces derniers qui est l'historisation des données.
<b>Géographique</b>	La donnée géographique représente les adresses des clients, dans notre cas ce sont des coordonnées spatiales visibles à l'aide d'une carte géographique.
<b>client</b>	Les informations relatives aux clients, comme la nature juridique, et la mention légale.
<b>Article</b>	Représente tous les articles proposés par Venus.
<b>Famille</b>	La famille de chaque article.
<b>Mode paiement</b>	Regroupe tous les modes de paiement effectués par les clients.
<b>Type de vente</b>	Les types de transactions entre les clients et l'entreprise.

Tableau 4 : dimensions et leur description.

### Description des activités :

Notre Data warehouse s'intéresse à l'activité commerciale de Vague de Fraicheur Venus©,

<sup>11</sup> Voir chapitre « Modélisation multidimensionnelle »

Cette dernière est axée sur une activité/processus majeur (la Vente).

### b. Déduction des besoins :

Analyse des ventes des produits :

- Chiffre d'affaire par **client**
  - Catégorie client (grossiste, commerçant).
  - Nature juridique (SARL, EURL).
  - Par région (wilaya).
  - Par état de facture.
- Chiffre d'affaire par période.
- Chiffre d'affaire par type de vente (normale, délégué, promotion).
- Chiffre d'affaire par mode de paiement (à terme, cheque, espèce, versement, virement).
- Chiffre d'affaire par **Article**
  - Famille (cosmétique, parfumerie).
  - Par désignation (savon, déodorant, ....
  - Par référence.
  - Par référence et par période.
- Chiffre d'affaire par :
  - Reference article et catégorie client.
  - Famille article et catégorie client.
- Le nombre d'article vendu :
  - Par période.
  - Par type de vente.
  - Par mode de paiement.

- Par client.
- Par catégorie client.
- Par nature juridique.
- Par région.
- Par client et par période.
- Par famille article.
- Par désignation
- Par référence.
- Par degré d'alcool.
- Par catégorie client et famille article.
- Par référence article et période.

#### **7. Conclusion :**

Dans ce chapitre, nous avons défini les besoins des décideurs résultants d'une longue phase d'interview et de questionnaires auxquels nous avons pu extraire les indicateurs d'analyse ainsi que les dimensions adéquates pour une analyse précise. Dans le prochain chapitre nous aborderons les grandes lignes de la modélisation multidimensionnelle qui est une étape fondamentale pour la réalisation d'un entrepôt de données.

*CONCEPTION  
MULTIDIMENSIONNELLE*

## 1. Introduction :

Nous traitons dans ce chapitre la conception logique et physique de notre système décisionnel. Ce dernier se compose de deux parties la première étant consacrée à la modélisation dimensionnelle et la seconde à l'architecture du système.

L'entreprise Vague de Fraicheur Venus© a implémenté un SI à caractère commercial pour pouvoir soutenir ses activités. Ce SI est basé sur le modèle relationnel mis en œuvre par un système de gestion de bases de données relationnel (SGBDR) PostgreSQL 9.1 réalisé par la boîte informatique de développement logiciel IconSoft©. L'objectif de notre modélisation dimensionnelle est de permettre aux utilisateurs ainsi que les décideurs de disposer d'une vision compréhensible et exploitable des données stockées au niveau du système source (Application gestion commerciale Vague de fraicheur).

La modélisation dimensionnelle met le point entre les concepts Faits et Attributs. Généralement les faits sont méconnus à l'avance. Le Fait est une constatation des tendances du marché par rapport à plusieurs dimensions. Dans le cas de notre entreprise, la majeure partie des faits sont à caractère numérique, c'est ce qui les différencie des attributs qui sont généralement de type textuel. Un attribut décrit une caractéristique d'un élément réel. Les attributs de notre base de production sont par exemple la désignation des articles, l'adresse et la nature juridique du client. Nous retrouvons dans la base de données de notre entreprise Vague de fraicheur Venus© au moins 3 dimensions : Temps, Client et Article.

Nous allons présenter dans ce chapitre dédié à la modélisation dimensionnelle et la conception les sujets cités dans le chapitre précédent. La démarche utilisée pour la modélisation de chaque sujet est un processus de plusieurs étapes.

## 2. Le Processus de conception :

La conception logique du schéma dimensionnel de chaque sujet mentionné dans le chapitre précédent s'inspire de la méthode instaurée par Ralph

Kimball dans son livre *The Data Warehouse Toolkit*, elle se déroule en quatre étapes dans cet ordre :

*The first dimensional model built should be the one with the most impact—it should answer the most pressing business questions and be readily accessible for data extraction.*

*Le premier modèle dimensionnel construit devrait être celui avec le plus d'impact, il doit répondre à des questions sur les affaires les plus pressantes et être facilement accessible pour l'extraction des données.*

- **Choix de processus d'activité à modéliser** : La première étape de la conception est de décider quels sont les processus d'affaires en combinant une compréhension des besoins de l'entreprise avec une compréhension de base des données disponibles. [KIMBALL, 2002]
- **Choix de grain du processus d'activité** : Une fois que le processus d'affaires a été identifié, l'équipe de l'entrepôt de données est confronté à une décision sérieuse au sujet de la granularité. Quel est le niveau de détail des données qui doivent être disponibles dans le modèle dimensionnel ? Cela nous amène à un bout de conception important.

Un entrepôt de données exige presque toujours des informations exprimées dans le grain le plus bas possible pour chaque dimension. Non parce que les requêtes portent sur les enregistrements individuels, mais parce que les requêtes doivent faire des coupes dans la base de données selon des critères bien précis. [MEYLAN, 2003]

- **Choix de la dimension** : Une fois le grain de la table de faits a été choisi, la date le produit et le client tombent comme étant des dimensions potentielles en plus des autres dimensions déductibles à partir des activités à modéliser. Le choix d'une dimension s'accompagne par la définition de tous les attributs textuels (les champs) qui garniront la table de dimension. [KIMBALL, 2002]

- **Choix des faits** : La quatrième et dernière étape dans la conception est de faire une détermination soigneuse des faits. Encore une fois, la déclaration du grain aide à cerner les faits. Autrement dit, les faits doivent être conformes aux grains. Les faits mesurés typiques sont des quantités numériques additives telles que le chiffre d'affaire et la quantité vendue. [KIMBALL, 2002]

### 3. Modélisation dimensionnelle de l'activité Vente :

- **Le processus d'activité** :

Pour notre projet, l'activité principale choisie dans le cadre de la création d'un entrepôt de données est l'activité *vente*. Nous pourrions ainsi voir en détail quels articles sont vendus, à quel prix, dans quelle région, durant quelle période. Le grain de cette base de données est « l'unité de vente par région par mois par client par article ».

- **Le grain du processus d'activité** :

Dans une gestion commerciale, il est important de connaître les chiffres de ventes réalisées. Notre analyse s'intéresse à ce sujet. Ainsi, chaque enregistrement de la table de faits est une ligne de vente réalisée par un client (région) par date par article. Le choix de la granularité consiste à indiquer qu'un nouvel enregistrement sera créé dans la table des faits pour chaque transaction de vente. A ce niveau, les dimensions client, article, région, mois s'imposent.

- **Les dimensions** :

La granularité de la table de faits nous a déterminés à priori une série de dimensions. Ces dimensions sont les suivantes : Article, date, client, mode de paiement, type de vente, zone géographique (adresse, wilaya, région).

• **Les faits mesurés :**

La table de faits doit contenir des indicateurs qui concernent la quantité des articles vendus et le chiffre d'affaire. Les deux faits sont : « Qte » et « CA ».

Activité	Mesure	Axe d'analyse
Activité vente.	CA: Chiffre d'affaire.	<ul style="list-style-type: none"> <li>• Catégorie client (grossiste, commerçant).</li> <li>• Nature juridique (sarl, eurl).</li> <li>• Par région (wilaya).</li> <li>• Par état de facture.</li> <li>• Période.</li> <li>• Type de vente (normale, délégué, promotion).</li> <li>• Mode de paiement (à terme, cheque, espèce, versement, virement).</li> <li>• Famille (cosmétique, parfumerie).</li> <li>• Par désignation (savon, déodorant, ....)</li> <li>• Par référence.</li> <li>• Par degré d'alcool.</li> <li>• Par référence et par période.</li> <li>• Reference article et catégorie client.</li> <li>• Famille article et catégorie client.</li> </ul>
	Qte : quantité vendue	<ul style="list-style-type: none"> <li>• Par période.</li> <li>• Par type de vente.</li> <li>• Par mode de paiement.</li> <li>• Par client.</li> <li>• Par catégorie client.</li> <li>• Par nature juridique.</li> <li>• Par région.</li> </ul>

- Par client et par période.
- Par famille article.
- Par désignation
- Par référence.
- Par degré d'alcool.
- Par catégorie client et famille article.
- Par référence article et période.

Tableau 5 : Tableau d'indicateurs des axes d'analyse.

#### 4. Description des dimensions :

Dimension	Description
Article	Contient toutes les informations concernant l'article
Client	Contient toutes les informations concernant le client
Modepaiement	Contient le mode de paiement que peut effectuer le client
Typevente	Représente les types de vente de la transaction
Adresse	Contient les adresses des clients
Wilaya	Contient les wilaya des clients
Région	Représente les régions selon les wilaya (nord, sud, Est,...)
Famille	Famille des articles
Catégorie	Catégorie du client
Temps	Contient la date des transactions.

Tableau 6 : Tableau descriptif des dimensions.

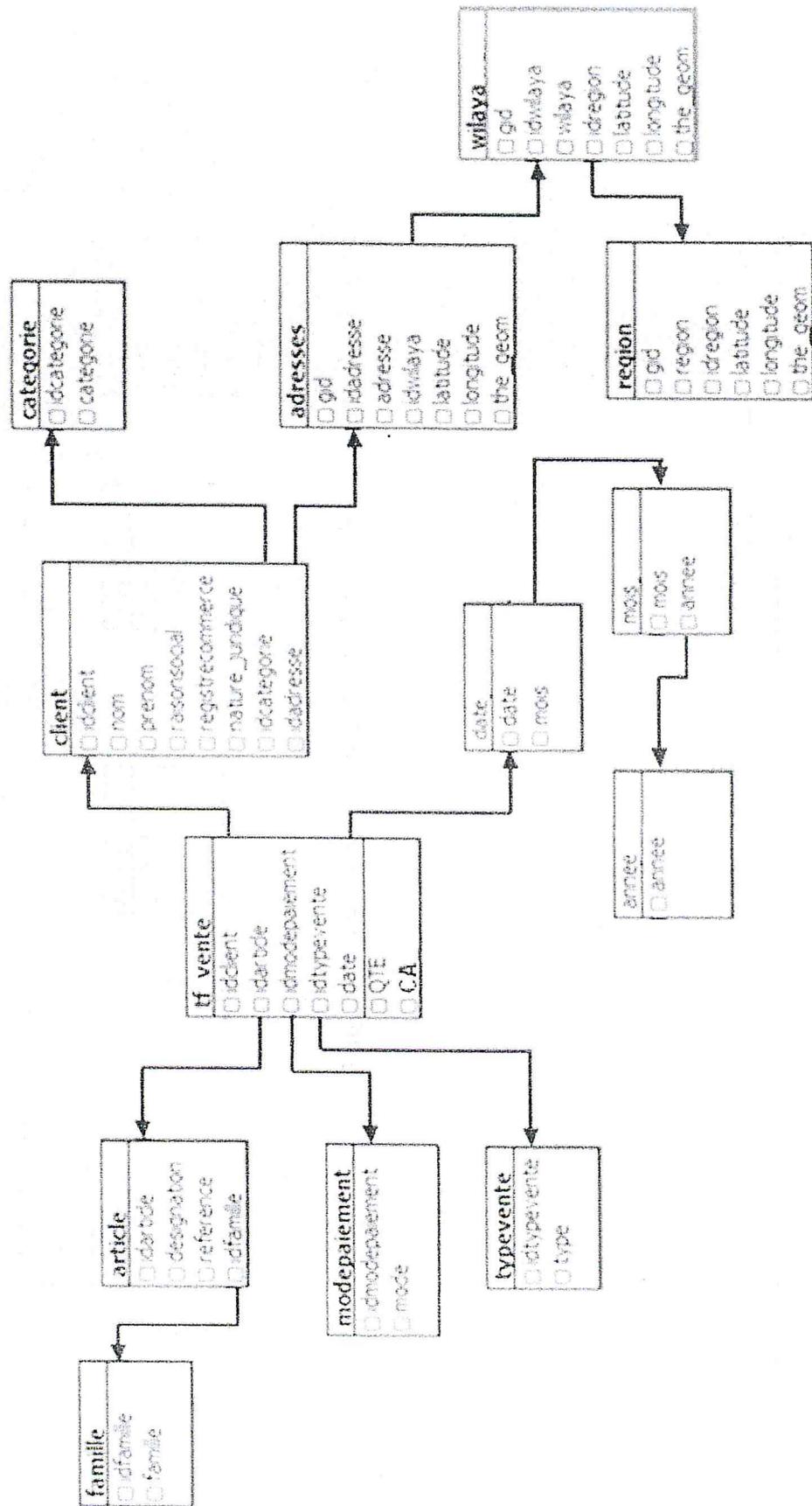


Figure 33 : modèle dimensionnel en flocon de neige pour l'activité vente.

## 5. Architecture du système :

La figure suivante illustre l'architecture globale de notre système décisionnel, cette architecture comprend les différents outils et composants du système décisionnel, ces derniers seront présentés dans le chapitre suivant (outils technologiques) :

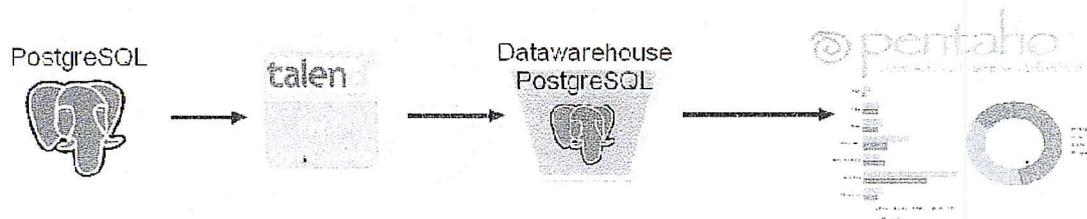


Figure 34 : Architecture du système.

Cette architecture est développée selon le concept de l'architecture trois-Tier. Cette architecture est une extension du modèle Client/serveur<sup>12</sup>, l'architecture 3-tier ou architecture à trois niveaux est l'application du modèle plus général qu'est le multi-tiers. L'architecture logique du système est divisée en trois niveaux ou couches : [LECOCCQ, 2009]

- **La couche présentation :**

La couche présentation est complètement normalisée par les deux organismes de normalisation internationale (ITU-T et ISO), La couche présentation se charge de la représentation des informations échangées entre systèmes ouverts, elle offre plusieurs fonctionnalités, parmi lesquelles les plus caractéristiques sont : [GARDIE, 2004]

- La préservation de la sémantique des données échangées.
- La négociation des syntaxes de transfert.
- L'accès des applications aux services de la couche session.

<sup>12</sup> Un mode de communication à travers un réseau entre plusieurs programmes ou logiciels : l'un, qualifié de client, envoie des requêtes ; l'autre ou les autres, qualifiés de serveurs, attendent les requêtes des clients et y répondent

- **La couche métier ou traitement :**

Ce sont des traitements qui vont s'opérer sur les données en fonction des requêtes des usagers. Différentes règles de gestion et de contrôle peuvent également être mises en œuvre à ce niveau. [MASTER, 2009] Appelée aussi couche Business, elle contient la logique d'affaires et /ou les règles d'ouverture de l'application. Une couche séparée strictement logique métier dans une architecture Trois-Tier est un atout majeur dans la mesure où les modifications qui doivent être apportées aux règles d'entreprise peuvent être faites ici sans avoir aucun effet sur d'autres applications. [EXFORSYS, 2007]

- **La couche accès aux données :**

Les données d'entreprise sont habituellement réparties, selon leur nature, dans différents systèmes de stockage tels qu'un annuaire, une base de données ou un système de fichiers. De manière standard, la plateforme J2EE propose un certain nombre d'API facilitant l'accès aux données. De plus, la communauté Open Source met à disposition un ensemble de frameworks<sup>13</sup> éprouvés, permettant de compléter ou de faciliter l'utilisation de ces API. [CLINKAST, 2008]

## **6. Construction du datawarehouse :**

Comme nous l'avons décrit précédemment le datawarehouse est basé sur une approche multidimensionnelle basée sur deux concepts essentiels : le fait et la dimension.

Notre table de fait « *vente* » comprend les clés étrangère des dimensions ainsi que les indicateurs/mesures, les dimensions contiennent une clé primaire et des attribues faisant référence aux tables de notre base de donnée source.

---

<sup>13</sup> Framework : est un kit de composants logiciels structurels, qui sert à créer les fondations ainsi que les grandes lignes de tout ou d'une partie d'un logiciel (architecture).

## Tableau descriptif du datawarehouse :

Table	attributs
Tf_vente	<ul style="list-style-type: none"> <li>• Idclient</li> <li>• Idarticle</li> <li>• Idmodepaiement</li> <li>• Idtypevente</li> <li>• Date</li> <li>• Qte</li> <li>• CA</li> </ul>
Client	<ul style="list-style-type: none"> <li>• Idclient</li> <li>• Nom</li> <li>• Prenom</li> <li>• Raisonsocial</li> <li>• registrecommerce</li> <li>• Naturejuridique</li> <li>• Idcategorie</li> <li>• idadresse</li> </ul>
Article	<ul style="list-style-type: none"> <li>• Idarticle</li> <li>• Designation</li> <li>• Reference</li> <li>• Idfamille</li> </ul>
Modepaiement	<ul style="list-style-type: none"> <li>• Idmodepaiement</li> <li>• Mode</li> </ul>
Typevente	<ul style="list-style-type: none"> <li>• Idtypevente</li> <li>• type</li> </ul>
Temps	<ul style="list-style-type: none"> <li>• Date</li> </ul>
Catégorie	<ul style="list-style-type: none"> <li>• Idcategorie</li> <li>• Categorie</li> </ul>
Famille	<ul style="list-style-type: none"> <li>• Idfamille</li> <li>• Famille</li> </ul>

<b>Adresse</b>	<ul style="list-style-type: none"> <li>• gid</li> <li>• ldadresse</li> <li>• Adresse</li> <li>• Idwilaya</li> <li>• Latitude</li> <li>• Longitude</li> <li>• The_geom</li> </ul>
<b>Wilaya</b>	<ul style="list-style-type: none"> <li>• gid</li> <li>• Idwilaya</li> <li>• Wilaya</li> <li>• Idregion</li> <li>• Latitude</li> <li>• Longitude</li> <li>• The_geom</li> </ul>
<b>Région</b>	<ul style="list-style-type: none"> <li>• gid</li> <li>• Idregion</li> <li>• region</li> <li>• latitude</li> <li>• longitude</li> <li>• the_geom</li> </ul>

Tableau 7: Attributs de la table de fait et dimensions.

### 7. L'aspect géographique :

L'entrepôt de données que nous avons réalisé bénéficie d'une caractéristique majeure lui permettant d'introduire la donnée géographique dans son aspect physique, c'est-à-dire que le modèle sur lequel on s'est basé pour créer le datawarehouse offre toutes les fonctions géométriques pour calculer les coordonnées des attributs qu'on veut étudier, dessiner des polygones et faire d'autres opérations à caractère spatial. Le modèle s'appelle `template_postgis` qui est issu du plugin PostGis que nous présenterons dans le chapitre suivant.

## 8. La restitution des informations géographiques :

Notre table client bénéficie de sous dimensions adresses wilaya et région conformément au modèle multidimensionnelle en flocon de neige, ces dernières représentent dans notre cas les données géographiques relatives aux clients. Chaque table (adresse, wilaya, région) contient des attributs classiques ([idadresse,adresse], [idwilaya,wilaya], [idregion,région]) respectivement, et qui permettent de communiquer avec la table client en suivant la hiérarchie clé\_principale/clé\_étrangère comme dans la figure 8. Les attributs gid, latitude, longitude, the\_geom sont générés à l'aide d'un SIG qui s'appelle Quantum GIS que nous présenterons dans le chapitre suivant.

La restitution est possible en 2 méthodes, la première est l'établissement de la cartographie des clients par couches adresse, wilaya et région. La création de ces couches génère les informations présentes dans les tables citées précédemment. Puis exporter chaque couche en format *.shapefile*<sup>14</sup> qui est un format standard et approuvé dans le domaine des SIG et enfin récupérer ce fichier à l'aide du plugin PostGis dans la plateforme Postgresl.

La deuxième méthode consiste à connecter l'entrepôt de données avec le SIG et générer les données géographiques d'une manière automatique à chaque insertion d'un nouvel attribut. Cette méthode est à titre expérimentale dans notre cas.

## 9. Alimentation de l'entrepôt de données :

Le chargement de données et l'alimentation s'est faite d'une façon classique en utilisant un ETL open source *Talend data intégration*, cet ETL s'occupe de la migration des données de la base de données source présente sous PostgreSQL 9.1 (vaguedefraicheur.sql) vers l'entrepôt de données présent également sous PostgreSQL (geo\_vf). L'interface graphique de *Talend* offre une solution facile pour la création de jointures complexes entre les tables.

---

<sup>14</sup> Fichier forme

### 10. Conclusion :

Nous avons présenté dans ce chapitre la modélisation multidimensionnelle de notre datawarehouse, l'architecture du système et ses composants. Dans le prochain chapitre nous présenterons les outils pour la mise en œuvre de notre système décisionnel ainsi que son déploiement.

*MISE EN OEUVRE*

*OUTILS TECHNOLOGIQUES*

## 1. Introduction :

Dans ce chapitre nous allons présenter les outils technologiques qui nous ont permis de mener à bien notre projet de système décisionnel, ces outils sont pour la plus part open source et bénéficient d'une documentation large et structurée alimentée en permanence par une communauté d'initiés internationaux.

## 2. PostgreSQL 9.1 :

PostgreSQL™ est un système de gestion de bases de données relationnelles objet (ORDBMS) fondé sur *Postgres, version 4.2™*. Ce dernier a été développé à l'université de Californie au département des sciences informatiques de Berkeley. POSTGRES

est à l'origine de nombreux concepts qui ne seront rendus disponibles au sein de systèmes de gestion de bases de données commerciales que bien plus tard.

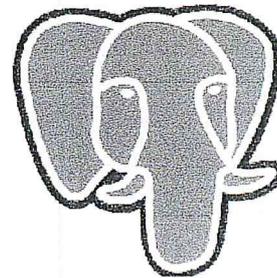
PostgreSQL™ est un descendant OpenSource du code original de Berkeley. Il supporte une grande partie du standard SQL tout en offrant de nombreuses fonctionnalités modernes : [[postgresql.com](http://postgresql.com)]

- Requêtes complexes
- Clés étrangères
- Triggers
- Vues
- Intégrité des transactions ;
- Contrôle des accès concurrents (mvcc ou multiversion concurrency control).

De plus, PostgreSQL™ est extensible par l'utilisateur de plusieurs façons. En ajoutant, par exemple :

- de nouveaux types de données
- de nouvelles fonctions

# PostgreSQL

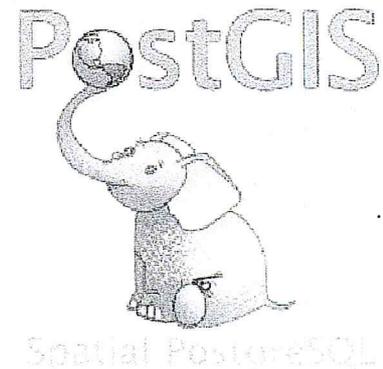


- de nouveaux opérateurs
- de nouvelles fonctions d'agrégat
- de nouvelles méthodes d'indexage ;
- de nouveaux langages de procédure.

### 3. PostGIS :

PostGIS est le "PostgreSQL Spatial" et ajoute les fonctionnalités "spatiales" à PostgreSQL, un type "geometry" en plus des types usuels tels que "varchar", "char", "integer", "date", etc. De nouvelles fonctions dédiées au type "geometry" telles que ST\_Distance(geometry, geometry), ST\_Area(geometry), ST\_Length(geometry),

ST\_Intersects(geometry, geometry). Il propose un mécanisme d'indexation qui permet aux requêtes spatiales de fonctionner rapidement dans de grandes bases de données. [[postgis.fr](http://postgis.fr)]



### 4. Openlayers :

OpenLayers est un logiciel libre. Il constitue une bibliothèque de fonctions JavaScript



assurant un noyau de fonctionnalités orienté vers la mise en place d'applications clientes Web cartographiques fluides. OpenLayers permet d'afficher des fonds cartographiques tuilés ainsi que des marqueurs provenant d'une grande variété de sources de données. Une partie de cette bibliothèque permet aussi de gérer l'ergonomie proposée à l'utilisateur, mais ce n'est pas directement son rôle. [[openlayers.com](http://openlayers.com)]

### 5. Quantum GIS :

Quantum GIS (QGIS) est un système d'information géographique (SIG) aisé



à prendre en mains et s'utilisant sous Linux, Unix, Mac OS X, et Windows.

QGIS supporte les formats vectoriels, rasters et de bases de données. QGIS est publié sous la GNU<sup>15</sup> Public License.

Le nom n'a pas d'autre signification que de commencer par la lettre Q, QGIS utilisant l'excellente plateforme Qt de Nokia.

Fonctionnalités de QGIS :

- Visualisation de données vecteurs et rasters dans différents formats et projections sans nécessité de convertir dans un autre format. Les formats supportés sont :
  - Les tables avec un support spatial activé tel que PostgreSQL avec PostGIS ou Sqlite avec SpatiaLite,
  - La plupart des formats vecteurs supportés par l'utilitaire\* OGR, ce qui inclut les shapefiles d'ESRI, MapInfo, SDTS et GML.
- Création de carte et exploration interactive des données spatiales avec une interface conviviale. Les nombreux outils disponibles comportent :
  - La projection à la volée,
  - Un composeur de carte,
  - Un panneau d'aperçu,
  - Des signets spatiaux,
  - Etc...
- Création, édition et export de données spatiales.
- Réalisation d'analyses spatiales en utilisant l'extension fTools ou l'extension GRASS intégrée.[[qgis.org](http://qgis.org)]

## 6. Talend :

Talend développe des solutions middleware open source qui permettent aux organisations de tirer un meilleur parti de leurs applications, systèmes et bases de données. Faisant voler en éclats le modèle propriétaire traditionnel, Talend démocratise le marché de l'intégration en fournissant des

The Talend logo is displayed in a large, light grey, lowercase sans-serif font. It is positioned to the right of the main text block, partially overlapping the right edge of the page.

<sup>15</sup> Licence publique générale gratuite GNU (GPL)

technologies open source d'entreprise qui adressent à la fois les besoins d'intégration de données et d'intégration d'applications des organisations de toute taille.

La plateforme d'intégration unifiée de Talend couvre des projets tels que l'intégration de données, l'ETL, la qualité de données, le master data management et l'intégration d'applications. Avec des performances démontrées, une vraie facilité d'utilisation, une extensibilité et une robustesse totales, les solutions de Talend sont les solutions d'intégration les plus utilisées et déployées dans le monde. Nous avons utilisé Talend Open Studio For Data Intégration comme ETL pour alimenter notre entrepôt de données. [fr.talend.com]

### 7. Pentaho :

Pentaho a été fondée en 2004, Il offre une suite de veille stratégique open source (BI) appelés *Business Analytics*. Pentaho permet



l'intégration des données, les services OLAP, le Reporting, les tableaux de bord, le datamining et un ETL performant sous le nom de *Pentaho data integration*. Pentaho est basé à Orlando, Floride, aux Etats-Unis. [wiki.org]

Pentaho lie étroitement l'intégration de données à l'analyse d'entreprise dans une plate-forme moderne qui permet aux services informatiques et aux utilisateurs d'accéder, de visualiser et d'explorer facilement toutes les données ayant un impact sur les résultats de l'entreprise. [pentaho.fr]

### 8. Conclusion :

Nous avons présenté dans ce chapitre l'ensemble des outils technologiques dont nous avons eu recours pour réaliser notre projet. Le prochain chapitre nous présenterons les étapes de réalisation de notre système décisionnel et sa mise en œuvre.

*MISE EN OEUVRE*

## 1. Introduction :

Nous allons présenter dans ce chapitre la mise en œuvre de notre système décisionnel, en s'appuyant sur des illustrations nous montrerons les étapes de création du datawarehouse, de son alimentation à l'aide de Talend, l'étape de l'analyse multidimensionnelle ainsi que la création du tableau de bord sous pentaho.

## 2. Construction du datawarehouse :

La création du datawarehouse se fait en conformité avec le schéma multidimensionnel que nous avons présenté dans le chapitre 4, en utilisant le système de gestion de bases de données PostgreSQL 9.1. La figure suivante nous illustre ceci :

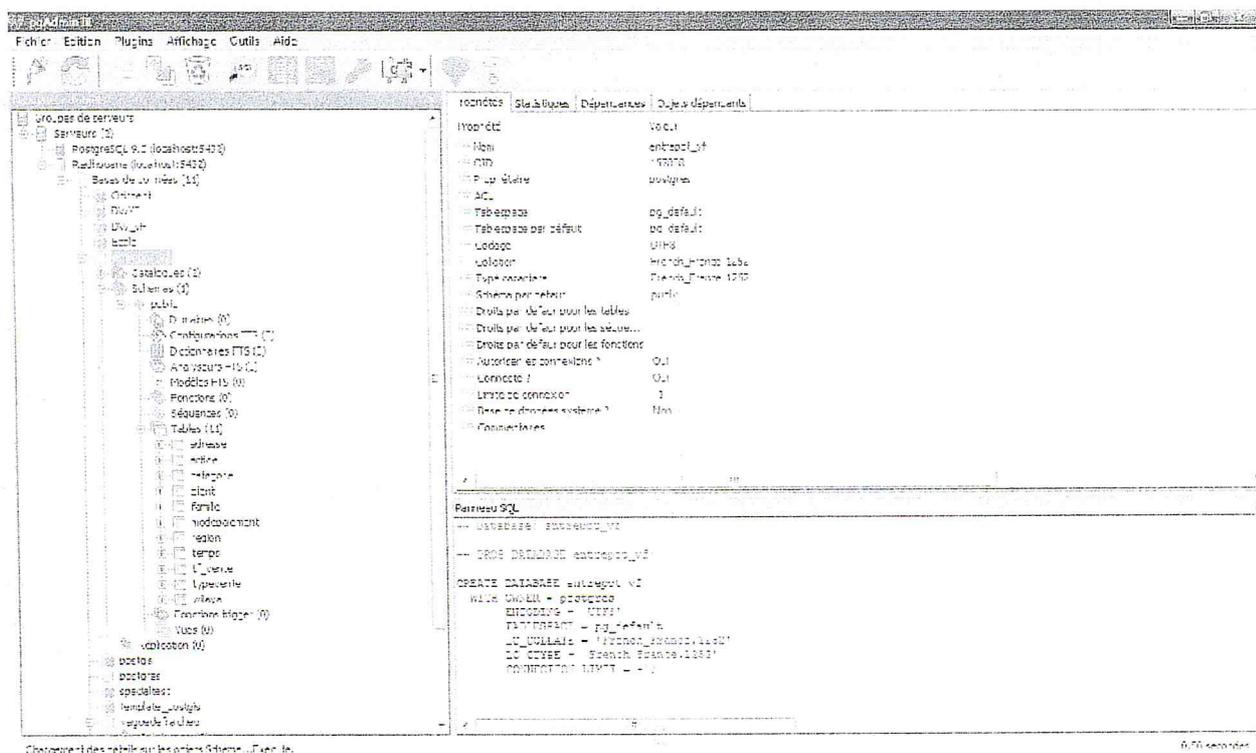


Figure 35 : interface pgAdmin3 (PostgreSQL 9.1)

## 3. Alimentation du datawarehouse :

L'alimentation du datawarehouse se fait à l'aide de l'ETL *Talend open studio for data integration*, une solution open source que nous avons choisi et dont les justificatifs se trouvent dans le chapitre précédent (outils technologiques).

La figure suivante nous montre comment créer une connexion server à partir Talend avec Postgresql.

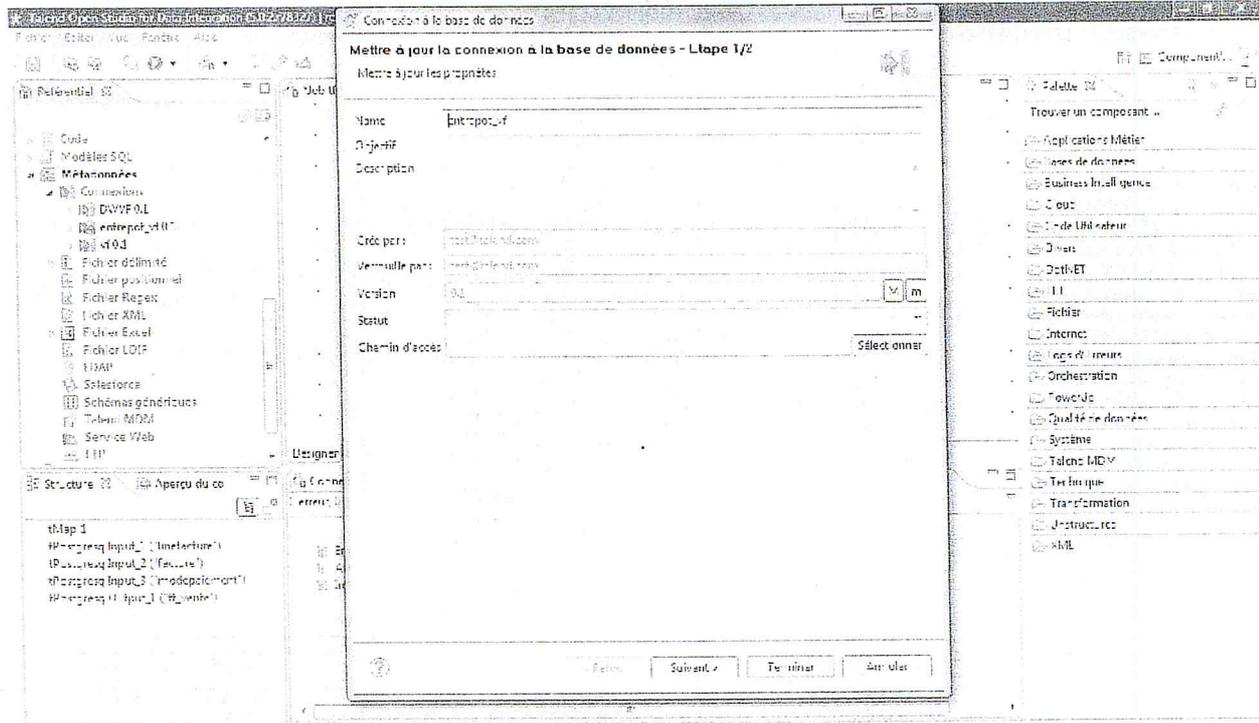


Figure 36 : création de connexion à une base de données.

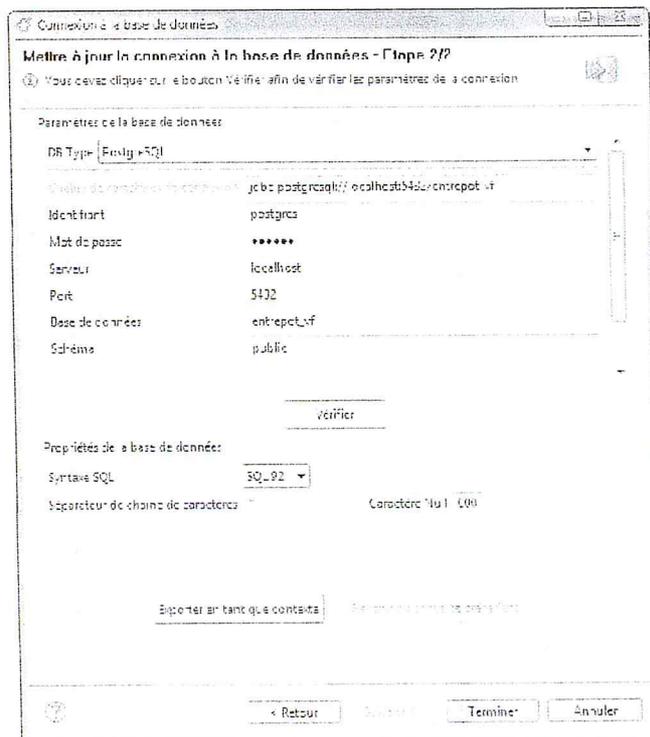


Figure 37 : choix de la base de données et identification.

La connexion se fait pour la base de données source et le datawarehouse de la même façon ainsi que la récupération des schémas des bases comme nous le montre la figure suivante :

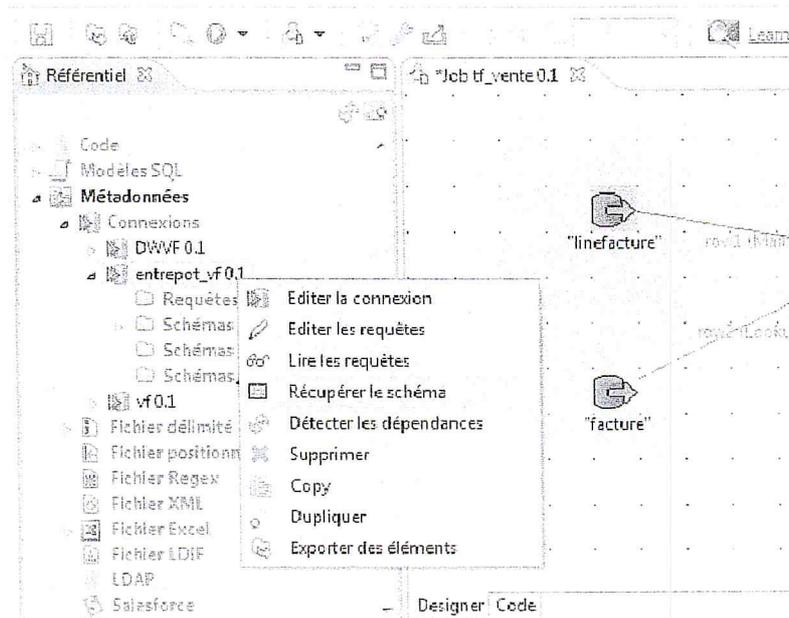


Figure 38 : Récupération du schéma du datawarehouse.

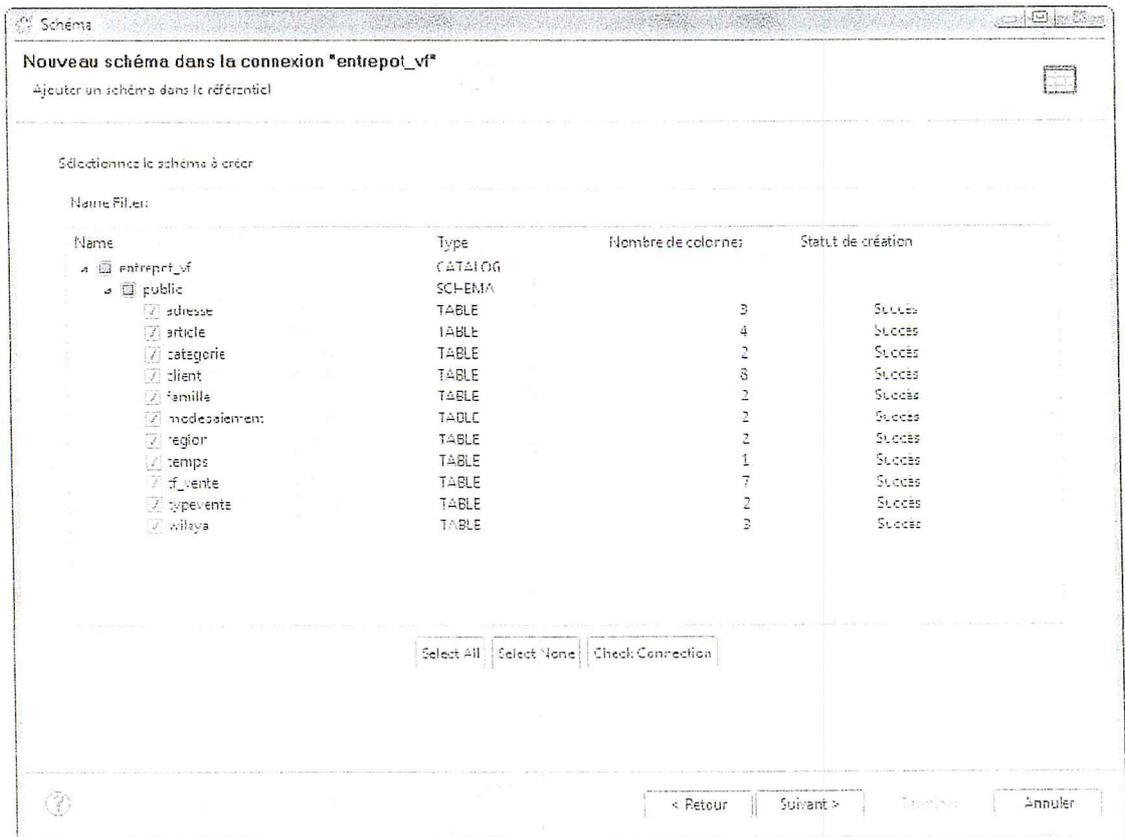


Figure 39 : choix du schéma et des tables à récupérer.

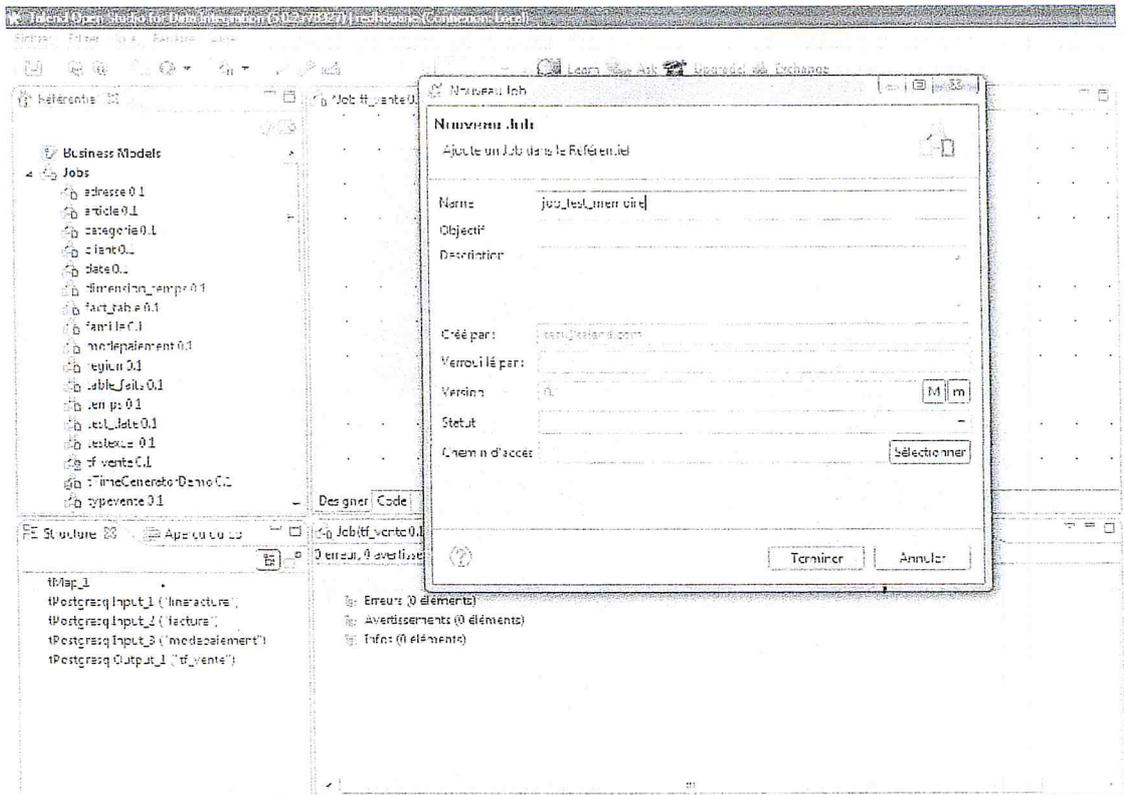


Figure 40 : création de job pour alimentation de tables.

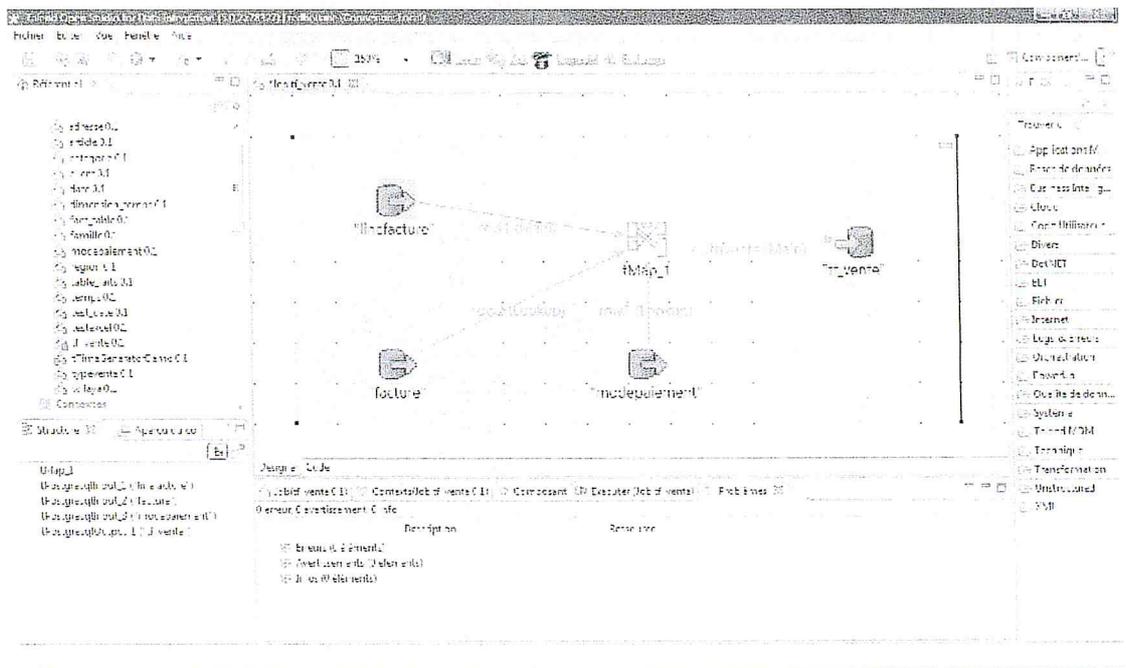


Figure 41 : job ETL de la table tf\_vente.



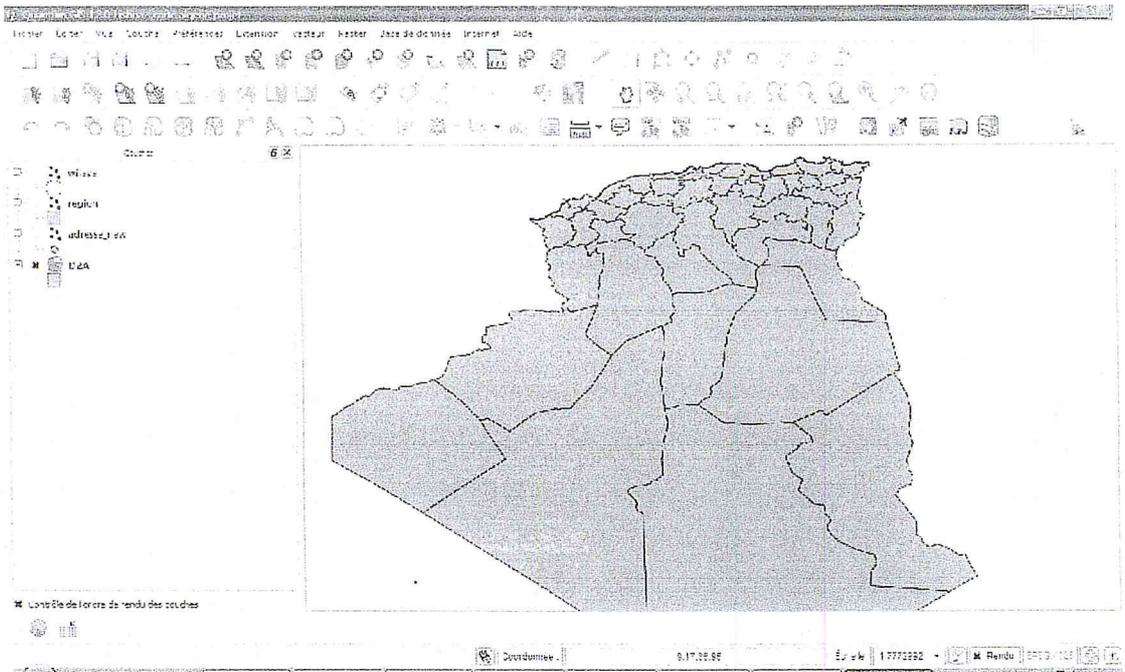


Figure 43 : couche Algérie sur quantum gis.

Chaque couche représente une table sauf la couche DZA qui est la carte d'Algérie et dont les attributs sont des polygones, qu'on a ajoutés juste pour nous aider à situer nos adresses, wilaya et régions.

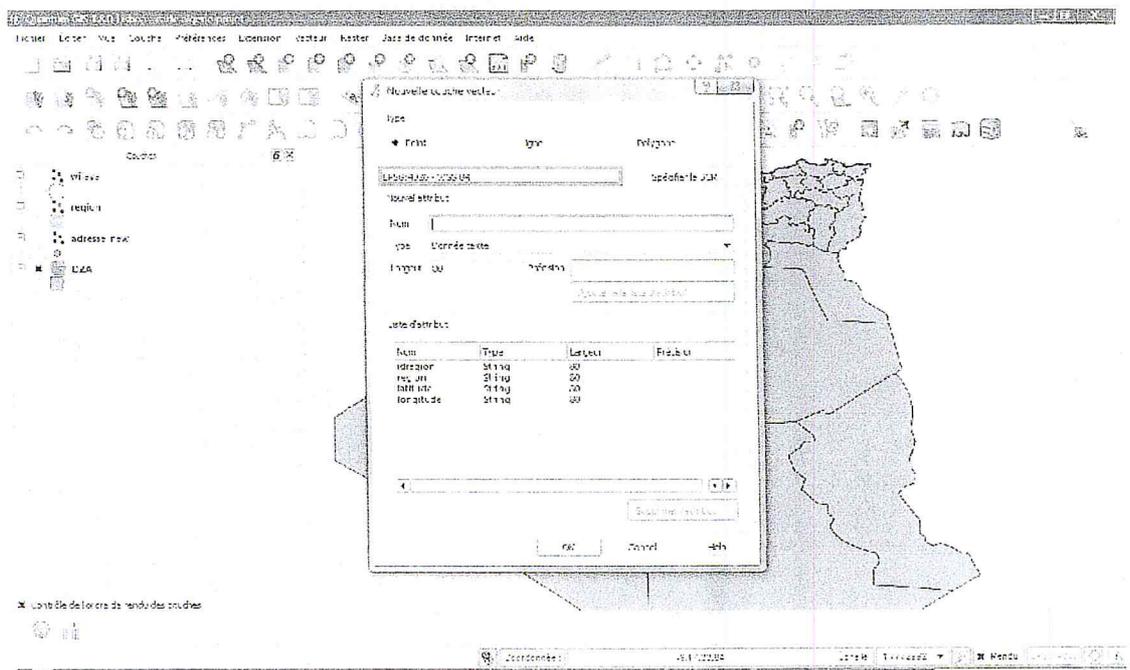


Figure 44 : ajouter une nouvelle couche région.

Le champ encadré en rouge dans la figure 44 EPSG :4326 – WGS 84 est un système de coordonnées géographiques ce système s'appelle World Geodetic System[wiki.org] et qui est un système mondial très utilisé notamment avec le GPS<sup>16</sup>

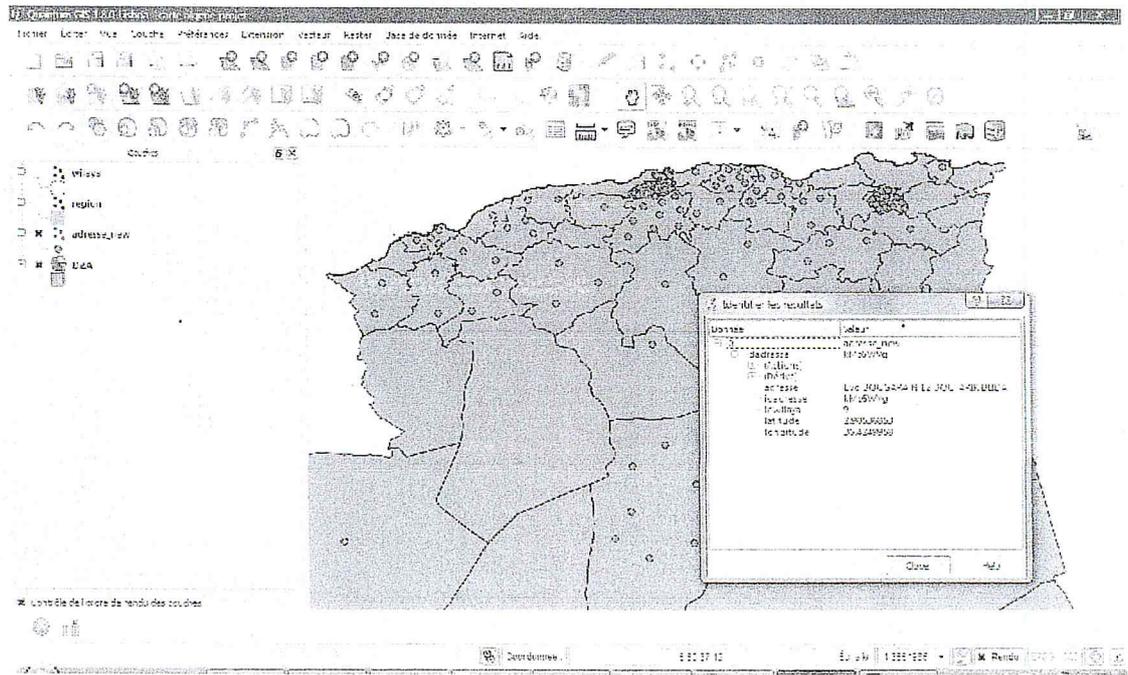


Figure 45 : couche adresse, exemple adresse client Blida.

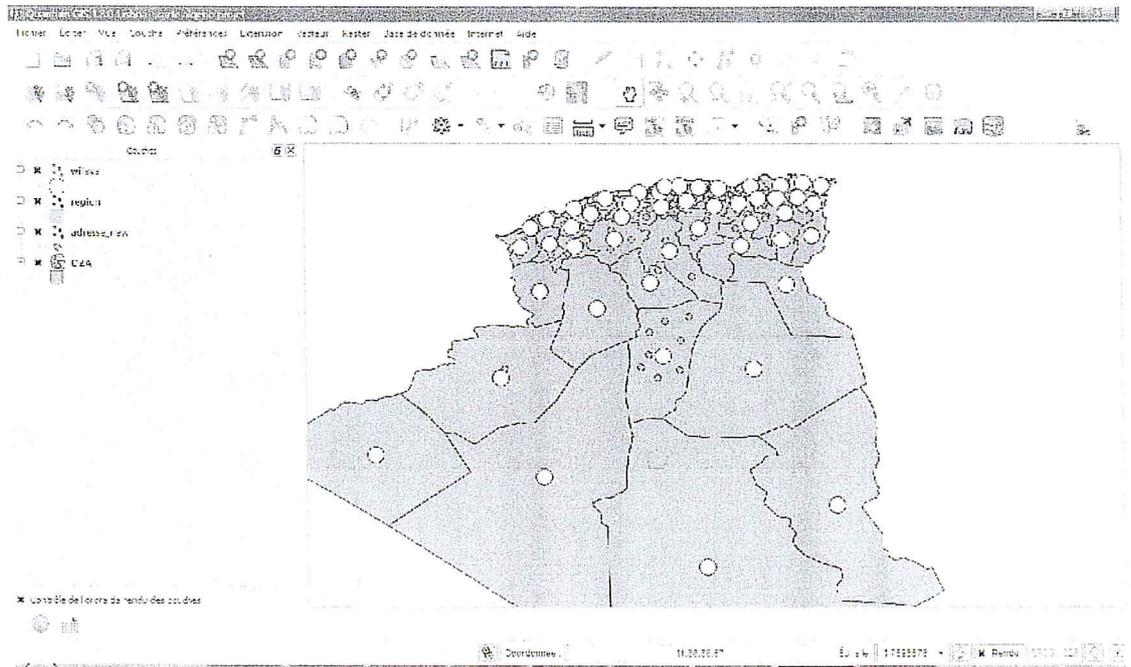


Figure 46 : cartographie de la répartition client.

<sup>16</sup> Global Positioning System/ système de localisation mondial.

Sauvegarder la couche une par une sur Quantum GIS ;

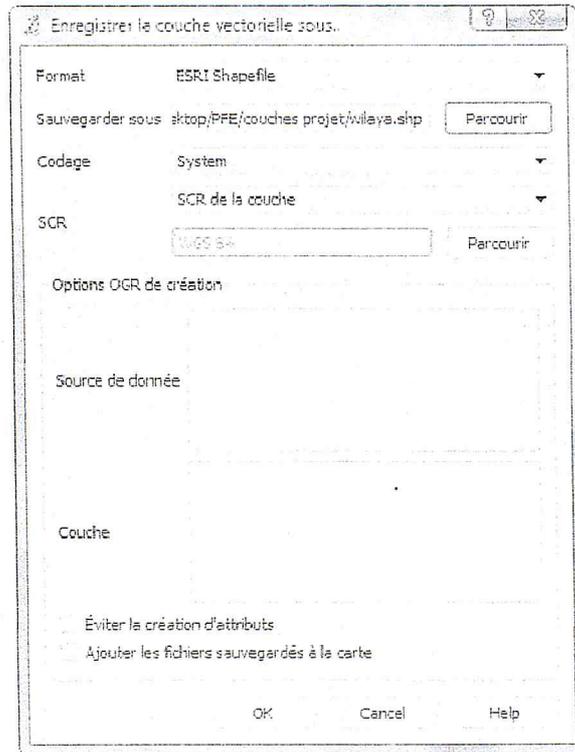


Figure 47 : sauvegarde couche.

Puis importation de la couche en format .shp qui veut dire ShapeFile/ fichier forme avec le plugin PostGis de PostgreSQL.

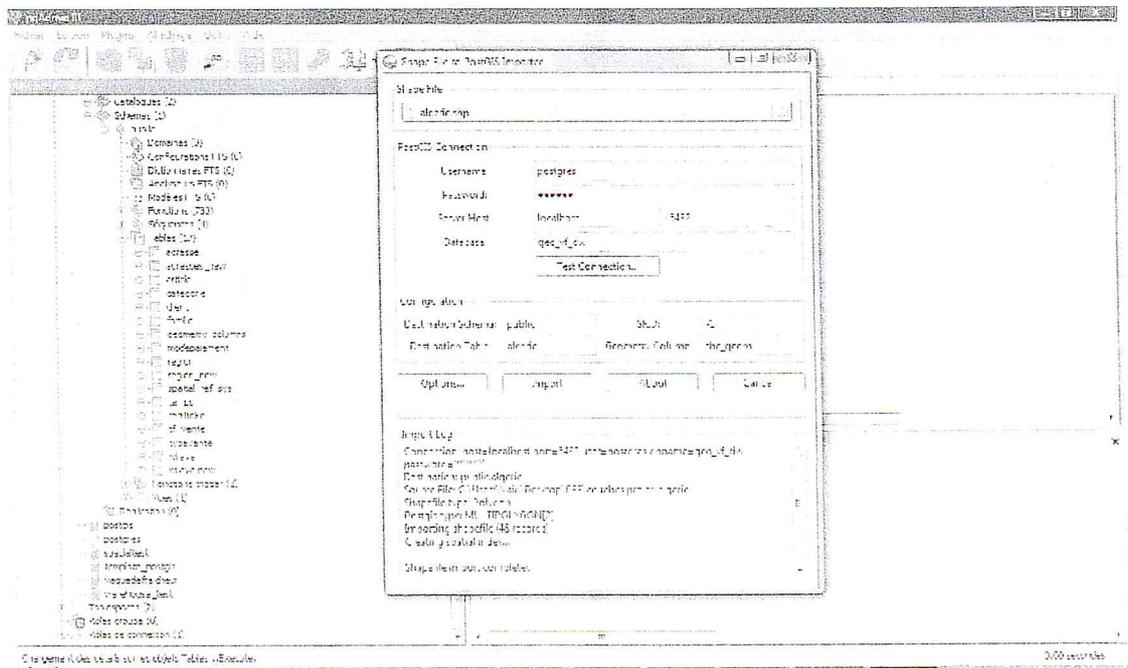


Figure 48 : importation du fichier .shp.

## 5. Création de cube dimensionnel :

Après l'alimentation du data warehouse avec Talend et l'importation des couches géographiques de nos tables adresses wilaya et région à l'aide de l'outil PostGis l'entrepôt de données spatial est prêt pour l'analyse et le Reporting. Afin de pouvoir analyser en profondeur les dimensions de notre activité de vente il est nécessaire de générer un cube Olap à l'aide de l'outil Schema Workbench que propose la suite Pentaho business analytics.

Voici un exemple de cube que nous avons réalisé en xml : un exemple

```
<Schema name="schema parfait">
  <Cube name="cube parfait" visible="true" cache="true" enabled="true">
    <Table name="sf_vente" schema="public">
    </Table>
    <Dimension type="StandardDimension" visible="true" foreignKey="idclient" highCardinality="false" name="client">
      <Hierarchy name="number 1" visible="true" isAll="true" primaryKey="idclient" primaryKeyTable="client">
        <Join leftKey="idcategorie" rightKey="idcategorie">
          <Table name="client" schema="public">
          </Table>
          <Table name="categorie" schema="public">
          </Table>
        </Join>
      </Hierarchy>
    </Dimension>
  </Cube>
</Schema>
```

Le code xml nous montre une déclaration de schema et de cube, ainsi que la table de fait et d'une dimension client. La balise <Join></Join> permet de faire un drill down et de descendre en profondeur dans la hiérarchie de notre entrepôt de données. Dans l'exemple ci-dessus du client à la catégorie client.

```

    <Level name="adresse" visible="true" table="adresses_new" column="ladresse" type="string" uniqueNumbers="false"
    hideNumberId="Never" description="adresses des clients">
      <Annotations>
        <Annotation name="Data.Role">
          <Text>Data.Role</Text>
        </Annotation>
        <Annotation name="Geo.Role">
          <Text>Geo.Role</Text>
        </Annotation>
      </Annotations>
      <Property name="Latitude" column="longitude" type="numeric">
      </Property>
      <Property name="Longitude" column="latitude" type="numeric">
      </Property>
    </Level>
  </Hierarchy>
</Dimension>
```

Dans ce code xml la balise <Level></Level> définit les attributs de notre dimension, dans notre cas c'est l'adresse qui est en jointure avec le client dans une balise <Join></Join>, les balises <Annotations> permettent d'ajouter des spécificités d'affichage aux attributs des dimensions de notre cube. Dans notre cas c'est un Data.Role *Geography* c'est-à-dire l'attribut adresse à une coordonnées géographiques, *Location* par contre va effectuer

un mapping<sup>17</sup> entre les propriétés de l'attribut représentées par les balises <Property></Property>, ces dernières font référence aux colonnes longitude et latitude de la table adresses.

Après création du schéma il est possible de le publier directement à l'aide de l'outil schema workbench de Pentaho et de le retrouver dans la plateforme d'analyse business analytics de pentaho. Ou bien de le copier manuellement dans le répertoire suivant : chemin/pentaho/server/biserver-ee/pentaho-solution/geo\_vf\_dw

## 6. Création du tableau de bord :

La création du tableau de bord geodécisionnel se base sur la plateforme *Pentaho business analytics*, cette dernière bénéficie du business analytics server auquel se connecte le datawarehouse. Pentaho accède directement aux sources des données et exerce sur elles toutes sortes d'analyse et de reporting. La figure suivante nous montre le panneau de configuration du serveur Pentaho accessible à cette adresse <http://localhost:8088> après démarrage de l'*Entreprise Console*.

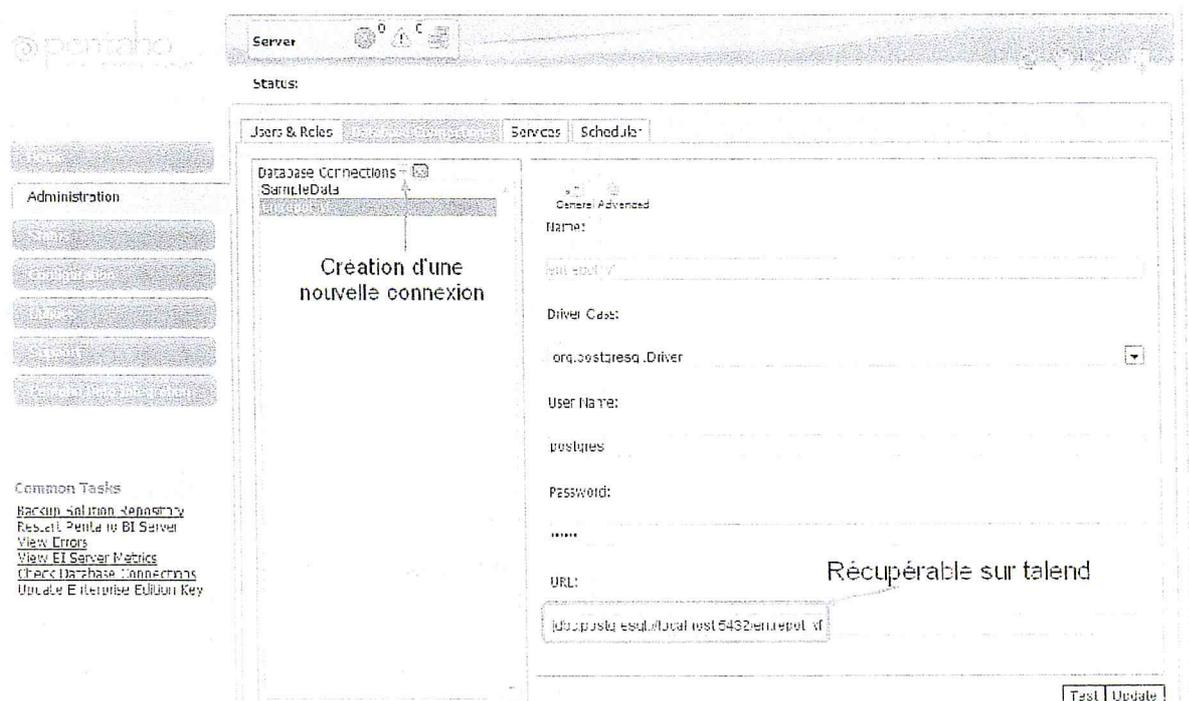


Figure 49 : administration de la connexion au DW.

<sup>17</sup> Mise en correspondance.

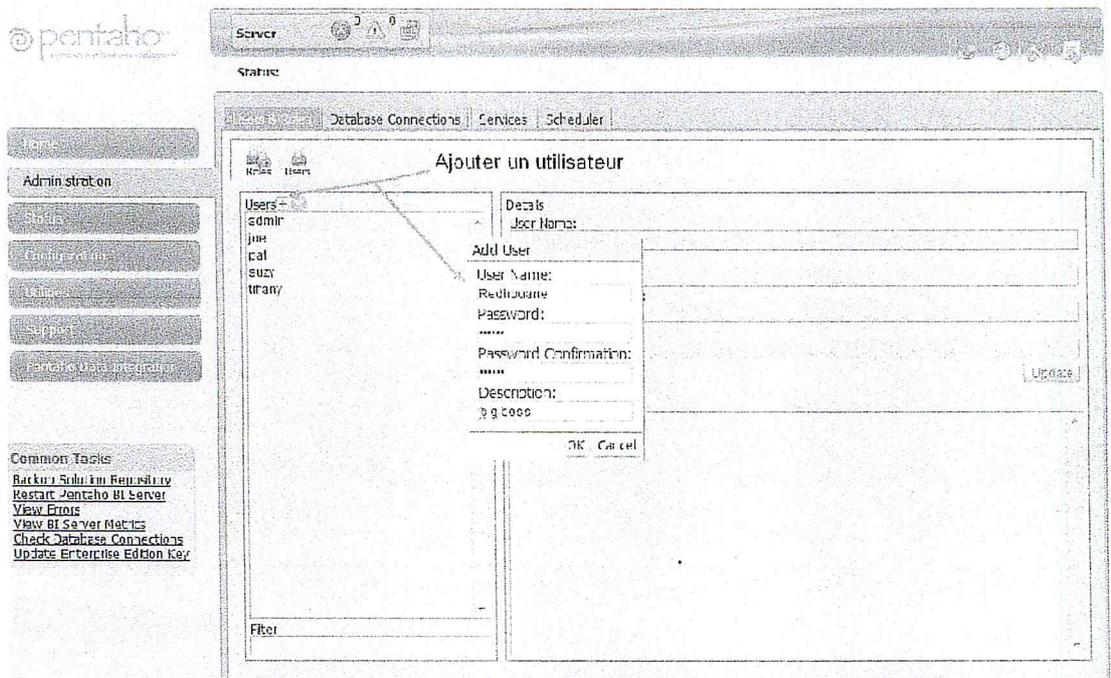


Figure 50 : Ajout d'un nouvel utilisateur.

- **Authentification et création de tableau de bord :**

La figure suivante illustre l'interface graphique web de pentaho business analytics :

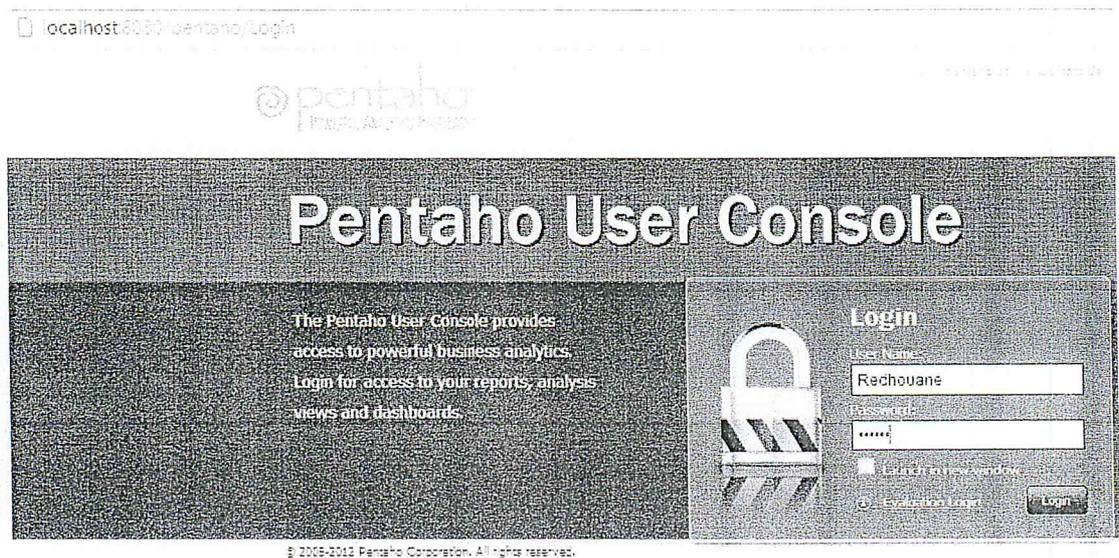


Figure 51 : Authentification Pentaho.

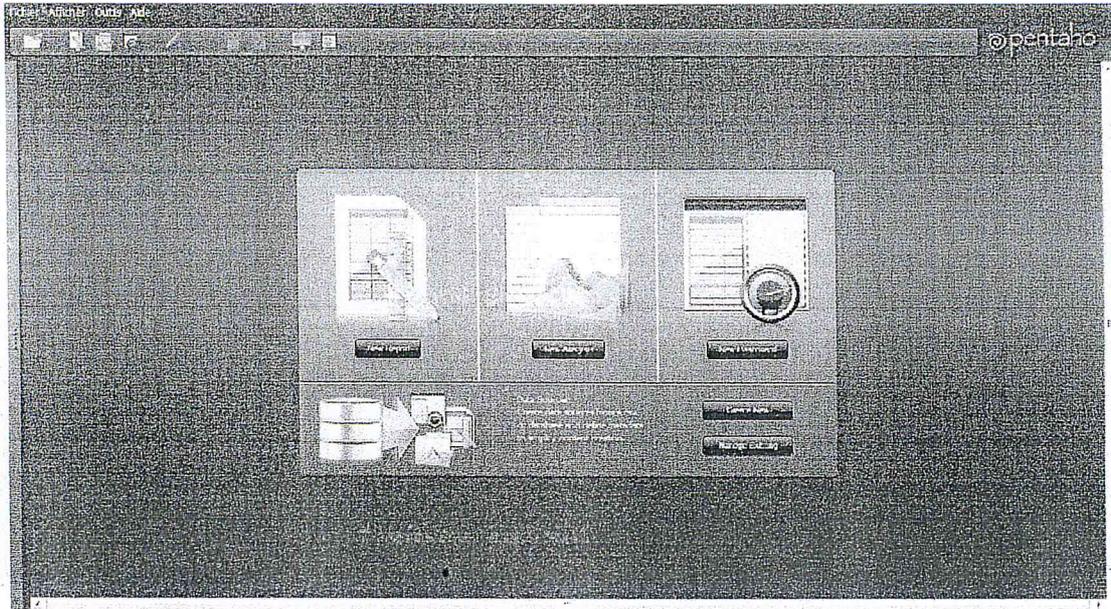


Figure 52: interface Pentaho business analytics.

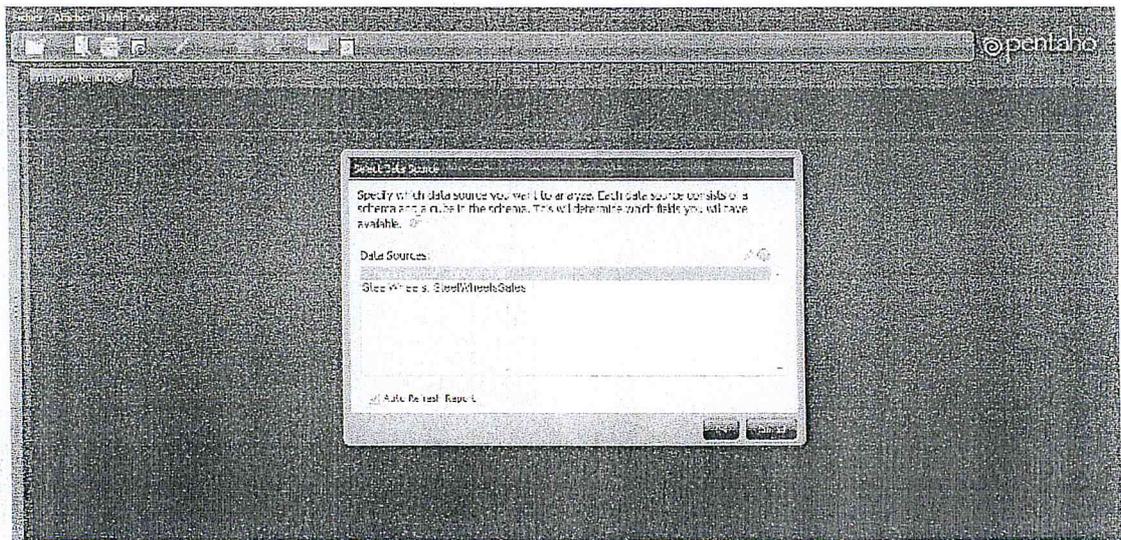


Figure 53 : récupération du cube publié en xml

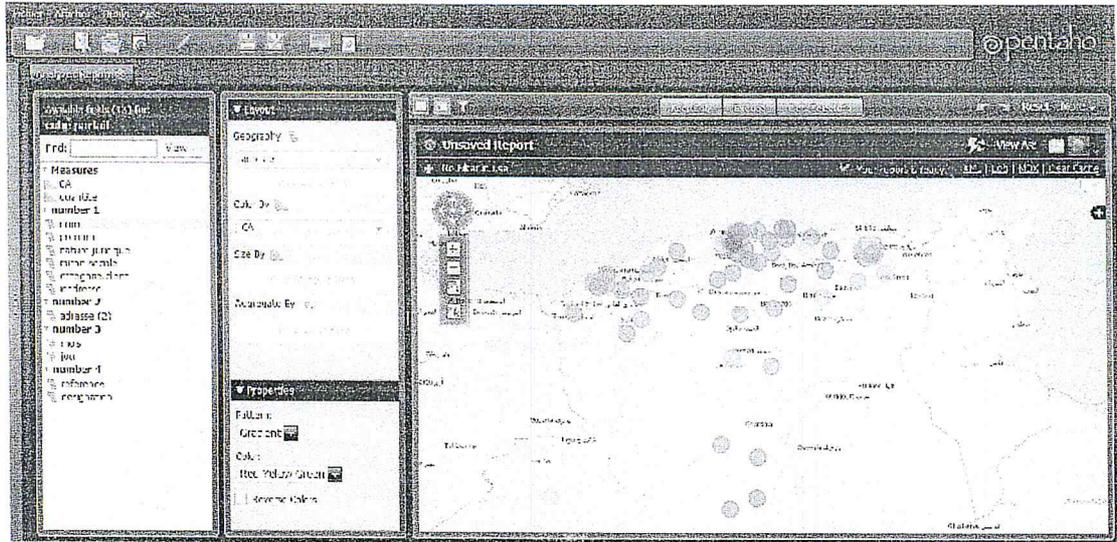


Figure 54 : localisation des clients par chiffre d'affaire.

Le passage de la souris sur les points rouges permet d'afficher le chiffre d'affaire réalisé avec le client ou la quantité de produits vendu pour un client donné, ou le chiffre d'affaire par wilaya ou région.

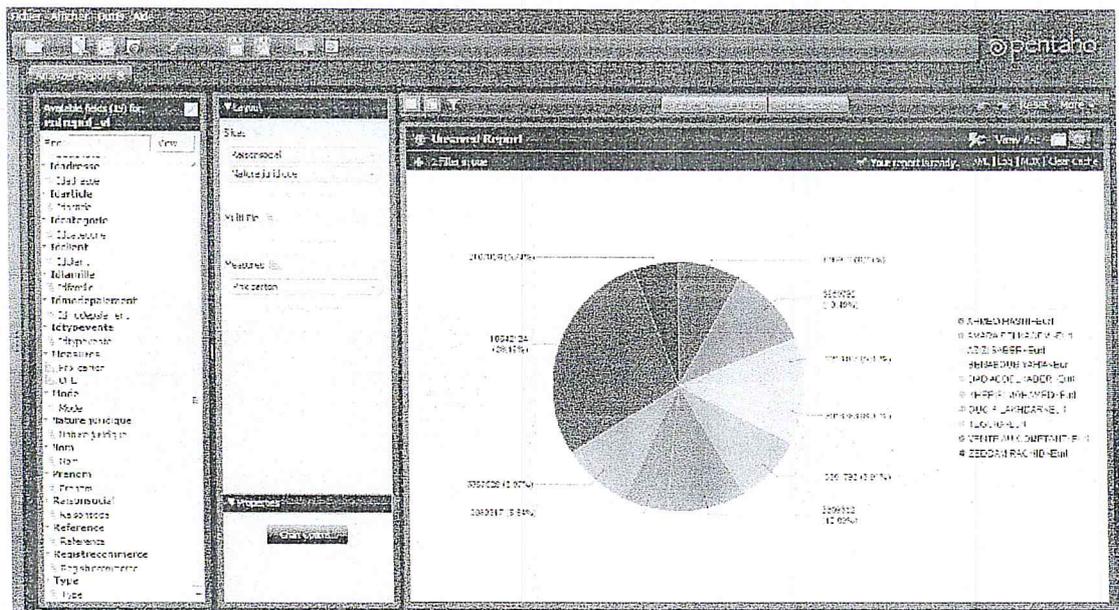


Figure 55 : Analyse bidimensionnelle.

La figure 13 est une analyse des 10 meilleurs clients selon la nature juridique et le chiffre d'affaire. Le résultat de l'analyse nous montre que 100% des clients sont de catégorie Eurl.



## 7. Conclusion :

Nous avons présenté dans ce chapitre la démarche avec laquelle nous avons réalisé notre système Géo décisionnel argumenté de quelques exemples de tableaux de bord.

## CONCLUSION GENERALE

## Conclusion générale :

Notre projet permet d'avoir une vue plus nette du monde du business intelligence, il met en clair les concepts de base du décisionnel, du datawarehouse et de l'introduction de la donnée géographique dans ce cycle décisionnel. Les systèmes d'informations décisionnels sont dédiés à améliorer la gestion de l'entreprise, c'est dans ce but que l'entreprise a fait appel à *Icon Software* afin de l'aider à cerner ses besoins en matière de gestion et d'analyse des activités commerciales. La particularité d'un projet décisionnel est qu'il se compose de trois phases majeures, l'étude des besoins de l'entreprise, la modélisation et conception de la solution proposée et enfin la mise en œuvre du système. Vient s'ajouter à toutes ces phases quelques customisations propres à la dimension géographique et qui se concentrent plus particulièrement dans l'aspect technique du projet.

Afin de bien mener notre projet nous nous sommes étalés sur les disciplines fondamentales dont nous avons jugé nécessaires à la réalisation du projet. Et c'est dans la première partie qui est un état de l'art que nous avons apporté ce soutien théorique.

La deuxième partie de notre projet a été consacré à la modélisation et la conception de notre système en se basant sur une étude de besoins réalisée en collaboration avec trois entreprises, Peugeot Blida, Sicam, Venus vague de fraîcheur. Ce choix d'approcher plusieurs entreprises vient combler le manque d'expérience en matière de business analyse que caractérise l'entreprise qui nous a octroyé le projet (venus).

La conception du système géo-décisionnel s'est divisée en deux étapes, la première étant la création de la cartographie des profils clients et la deuxième la création du datawarehouse en se basant sur les données géographiques des clients.

La dernière partie de notre projet a été consacrée à la mise en œuvre du système géo-décisionnel en créant un portail de restitution basé sur un tableau de bord ergonomique comportant des analyses multidimensionnelles, du Reporting et une visualisation géographique des

profils clients en fonctions des axes d'analyses vues précédemment. L'utilisation de l'outil Pentaho Business analytics entreprise édition nous a permis de nous familiariser avec l'analyse multidimensionnelle en nous proposant un outil open source très puissant et fiable capable de s'adapter aux exigences du client et aux besoins de l'informaticien en matière de flexibilité.

Durant notre projet, nous avons sans cesse pensé à améliorer notre système en le rendant plus riche et plus sur-mesure aux besoins du décideur, certaines des perspectives suivantes sont actuellement en cours d'expérimentation et à un stade très avancé de concrétisation :

- L'hébergement de l'entrepôt spatial geo\_vf\_dw sur Geoserver qui est une plateforme d'hébergement de données géographique. Et intégrer uniquement la carte géographique souhaitée comme API dans la plateforme Pentaho BA.
- La création d'un tableau de bord de restitution basé sur la plateforme Android.
- Créer un fichier de mapping standard pour les communes villes et wilaya d'Algérie afin de standardiser l'utilisation des données spatiales et éviter les customisations

# ANNEXES

## Questionnaire

Date :

Direction :

Questions aux dirigeants :

Questions relatives aux objectifs de l'entreprise et de la façon dont cette dernière procède à l'évaluation de la réussite.

Objectifs et évaluation de l'entreprise	
Questions	
1.	<ul style="list-style-type: none"><li>• Quels sont les objectifs de votre secteur d'activités ? qu'essayez-vous d'accomplir ?</li></ul>
2.	<ul style="list-style-type: none"><li>• Quelles sont vos méthodes pour évaluer la réussite ? Comment savez-vous que vous êtes sur la bonne voie ? Quelle est la fréquence des évaluations ?</li></ul>
3.	<ul style="list-style-type: none"><li>• Qu'est-ce qui vous empêche d'atteindre vos objectifs ? Quelles en sont les conséquences sur l'entreprise ?</li></ul>

Evaluation des problèmes	
Questions	
1.	<ul style="list-style-type: none"><li>• Quel genre de problèmes affrontez-vous d'une façon générale ?</li></ul>
2.	<ul style="list-style-type: none"><li>• Comment arrivez-vous à voir que vous êtes en face d'un problème, ou vous allez l'être prochainement ?</li></ul>
3.	<ul style="list-style-type: none"><li>• Croyez-vous que le recours à l'utilisation d'un système décisionnel va vous aider à éviter de tomber dans les mêmes problèmes ?</li></ul>

### Informations sur le Répondeur :

Quel poste occupez-vous ?

- Directeur général
- Directeur exécutif
- Directeur central
- Responsable informatique

Autres (précisez svp) :

.....

Vous occupez ce poste depuis combien d'année ? :.....année(s)

Questions aux responsables :

Questions relatives aux objectifs d'un service donné (Analyse, prédiction...)

Objectifs en rapport avec le secteur d'activité	
Questions	
1.	• Quels sont les objectifs de votre service ?
2.	• Comment vous procédez pour atteindre vos objectifs ?
3.	• Comment vous évaluez votre réussite ? comment vous savez si vous êtes sur la bonne voie ?
4.	• Quelle est la fréquence de votre auto-évaluation ?
5.	• Quels sont les contraintes que vous rencontrez durant votre travail ?
6.	• Parlez-nous de vos produits ? comment classifiez-vous vos produits ? comment les recherchez alors qu'ils sont sans doute nombreux ?
7.	• Combien de fournisseurs avez-vous ?

Analyse des besoins par secteur d'activité	
Questions	
1.	• Effectuez-vous des analyses sur les données que vous avez ? quelle est leurs fréquences ? comment obtenez-vous les données ?
2.	• Quel genre d'analyse aimerez-vous réaliser dans le futur ?
3.	• Quelles sont les données archivées dont vous avez besoin ?
4.	• Comment utilisez-vous l'information tirée à partir de vos analyses ?

Analyse des activités de l'entreprise	
Questions	
1.	• A quelle période examinez-vous l'activité de votre entreprise ?
2.	• Votre chiffre d'affaire se calcule d'une façon classique ? ou il dépend de plusieurs facteurs ?
3.	• Votre analyse d'activité s'intéresse aux clients ?
4.	• pensez-vous que la donnée géographique peut apporter un plus à votre analyse ?
5.	• Voulez-vous suivre les ventes par région géographique ?
6.	• Quels sont les activités que vous voulez analyser ?
7.	• Décrivez votre système opérationnel. (SI et BD)
8.	• Quels sont les outils d'analyse que vous utilisez ? quel processus utilisez-vous pour diffuser l'information collectée ?
9.	• Envisagez-vous de concevoir un entrepôt de donnée ?

Informations sur le Répondeur :

Quel poste occupez-vous ?

- Directeur général
- Directeur exécutif
- Directeur central
- Responsable informatique

Autres (précisez svp) : .....

Vous occupez ce poste depuis combien d'année ? :.....année(s)

## Introduction :

Dans cette partie on présente la base de données de Vague de fraîcheur Venus©, cette partie comprend :

- La description de l'application.
- La liste des tables de la base de données.
- Le modèle physique des données.
- Le dictionnaire de données contenant le nom, le code et le type des données :

## Description de l'application :

L'application de Vague de fraîcheur Venus© est utilisée au niveau du service commercial de l'entreprise.

	Domaine fonctionnel	Système d'exploitation	Accès aux données	Langage de développement	Chargé de maintenance	Nombre d'incidents constatés
Données générales	Gestion commerciale					
Environnement Technique		Windows XP	Postgre-sql	Java EE		
Maintenance					Ingénieur Icon-software	Non communiqué
Observation Générale	Type de solution : Application Desktop réalisée en JEE. Fonction de sauvegarde et mise à jour existante. Authentification sécurisée avec droit d'accès pour administrateurs.					

Tab 1 : Description de l'application.

### Liste des tables :

Deux types de tables caractérisent la base de données, des tables statiques et tables dynamiques :

### Liste des tables statiques :

Nom de la table	Désignation
<b>Famille</b>	Famille de l'article
<b>Article</b>	Article
<b>Modepaiement</b>	Mode de paiement
<b>Client</b>	Liste des clients
<b>Catégorie</b>	Catégorie des clients
<b>Prixvente</b>	Prix de vente de chaque article
<b>Typevente</b>	Type de vente qu'entreprend le client

Tab 2 : Liste tables statiques.

### Liste des tables Dynamiques

Nom de la table	Désignation
<b>Facture</b>	Facture de chaque client
<b>Linefacture</b>	Liste chaque transaction pour chaque produit qu'achète un client
<b>Boncommande</b>	Bon de commande du client
<b>Reglement</b>	Etat de la facture
<b>lineboncommande</b>	Table Intermédiaire entre article et boncommande
<b>facturesf</b>	Table intermédiaire entre boncommande et état de la facture « Reglement »

Tab 3 : Liste tables dynamiques.

Dictionnaire des données :

	Article	Désignation
Famille		Famille de l'article
Famille		Famille de l'article
Description		Description de la famille
Idfamille		Identifiant numérique de la famille
Article		Article
designation		Nom de l'article
Reference		Reference de l'article
prixunitaireht_old		Prix unitaire hors taxe
Tva		TVA
Unitemesure		Unité de mesure de l'article
quantite_en_carton		Quantité de carton d'article en stock
Volume		Volume de l'article par unité de mesure
Degrealcool		Degré d'alcool du produit
Droitfixe		Droit fixe
quantite_en_unite		Quantité de chaque article
Codearticle		Code article
Idarticle		Identifiant article
Idfamille		Identifiant famille article
unite_par_carton		Quantité de l'article par carton
Modepaiement		Mode de paiement
Mode		Mode de paiement
Designation		Désignation de paiement
idmodepaiement		Identifiant du monde de paiement
Client		Désignation
Nom		Nom client
Prenom		Prénom client
Adresse		Adresse client
Telephone		Telephone
Ajouterdate		Date de l'ajout du client
Codeclient		Code du client
Raisonsocial		Nom de la société
Registrecommerce		Code du registre de commerce
Immatriculfiscal		Immatriculation fiscale
Articleimposition		Article imposition
Codeactivite		Code de l'activité
Solde		Solde
Idclient		Identifiant du client
Idcategorie		Identifiant de la catégorie du client
nature_juridique		Nature juridique de l'entreprise du client
	Catégorie	Catégorie des clients
Categorie		Catégorie client
Description		Description

<b>Idcategorie</b>	Identifiant catégorie
<b>Prixvente</b>	Prix de vente de chaque article
<b>Idprixvente</b>	Identifiant du prix de vente
<b>Prix</b>	Prix de l'article
<b>Date</b>	Date de la création du prix
<b>Actif</b>	Etat de l'article
<b>Idarticle</b>	Identifiant article
<b>Idtypevente</b>	Identifiant type de vente
<b>Facture</b>	Facture de chaque client
<b>Totalht</b>	Total hors taxes
<b>Totaltva</b>	Total TVA
<b>Totalgeneral</b>	Total général
<b>Volumetotalalcool</b>	Volume total d'alcool
<b>Droitfixe</b>	Droit fixe
<b>Date</b>	Date
<b>Numeroefacture</b>	Numéro de facture
<b>Etat</b>	Etat
<b>Idfacture</b>	Identifiant facture
<b>Idboncommande</b>	Identifiant bon de commande
<b>Modedepaiement</b>	Mode de paiement
<b>Totallettre</b>	Total en lettre
<b>Totalhtalcool</b>	Total hors taxes alcool
<b>Totalhtcosmetique</b>	Total hors taxes cosmétiques
<b>Totaltvacosmetique</b>	Total TVA cosmétique
<b>Totalhtparfumerie</b>	Total hors taxes parfumerie
<b>Totaltvaparfumerie</b>	Total TVA parfumerie
<b>Doit</b>	Droit
<b>Linefacture</b>	Liste chaque transaction
<b>Quantité</b>	Quantité en carton achetés
<b>Montantht</b>	Montant de la transaction en hors taxe
<b>Idlinefacture</b>	Identifiant de la table linefacture
<b>Idarticle</b>	Identifiant de l'article acheté
<b>Montanttva</b>	Montant de la TVA sur la quantité achetée
<b>Prix</b>	Prix de l'article par unité
<b>Idfacture</b>	Identifiant de la facture du client
<b>Boncommande</b>	Bon de commande du client
<b>Numeroebc</b>	Numéro bon de commande
<b>Date</b>	Date
<b>Etat</b>	Etat
<b>Idboncommande</b>	Identifiant bon de commande
<b>Idclient</b>	Identifiant client
<b>Idtypevente</b>	Identifiant type de vente
<b>Reglement</b>	Etat de la facture
<b>Date</b>	Date
<b>Montant</b>	Montant

<b>Modereglement</b>	Mode règlement
<b>Idreglement</b>	Identifiant règlement
<b>Idfacture</b>	Identifiant facture
<b>Typereglement</b>	Type de règlement
<b>mode_payement</b>	Mode de paiement
<b>lineboncommande</b>	Table Intermédiaire entre article et boncommande
<b>Quantite</b>	Quantité
<b>Montantht</b>	Montant hors taxe
<b>Idlineboncommande</b>	Identifiant ligne bon de commande
<b>Idboncommande</b>	Identifiant bon de commande
<b>Idarticle</b>	Identifiant article
<b>Numeroelbc</b>	Numéro ligne de bon de commande
<b>Quantitetotal</b>	Quantité totale
<b>Montanttva</b>	Montant TVA
<b>Volumefixe</b>	Volume fixe
<b>Prix</b>	Prix
<b>facturesf</b>	Table intermédiaire entre boncommande et état de la facture « Reglement »
<b>Totalht</b>	Total hors taxes
<b>Totaltva</b>	Total TVA
<b>totalgeneral</b>	Total général
<b>Volumetotalalcool</b>	Volume total alcool
<b>Droitfixe</b>	Droit fixe
<b>Date</b>	Date
<b>Numeroefacture</b>	Numéro facture
<b>Etat</b>	Etat
<b>Idfacturesf</b>	Identifiant facture intermédiaire
<b>Idboncommande</b>	Identifiant bon de commande
<b>Modepaiement</b>	Mode de paiement
<b>Totallettre</b>	Total en lettre
<b>Doit</b>	droit

Le modèle physique de l'application de vague de fraîcheur Venus® :

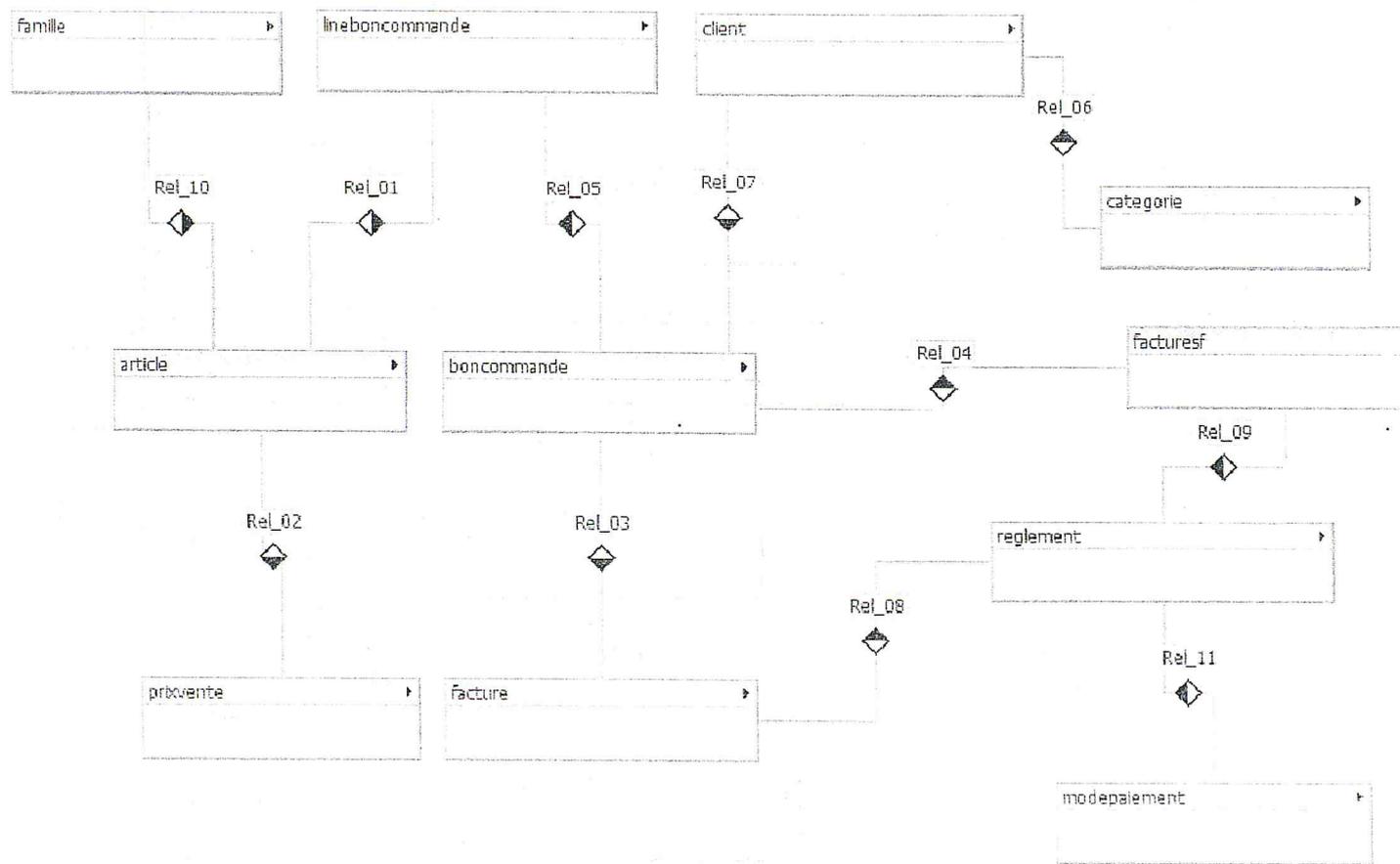


Fig 1 : Modèle physique de l'application.

*BIBLIOGRAPHIE*

## **Chapitre : Bi et systèmes décisionnels :**

[Crl, 2002] : Conseil Régional De Lorraine, Intelligence Economique : Un Guide Pour Débutants Et Praticiens, 2002, Isbn: 84-451-2389-0

[Bellinger, 2010] : Gene Bellinger : Knowledge Management – Emerging Perspectives ([www.Outsights.Com/Systems/Kmgmt/Kmgmt.Htm](http://www.Outsights.Com/Systems/Kmgmt/Kmgmt.Htm))

[Dufau, 2010] : l'intelligence Economique, M. Jean-Pierre Dufau, Juillet 2010

[Ifoact, 2000] : Commission Européenne, Programme Innovation: Projet Trips: Ifoact: 1998-2000.

[Lafare, 2009] : Matthieu Lafare, Thèse Professionnelle Hec : « Business Intelligence ». 2009

[Cigref, 2009 ] : Place De La Bi Et Pilotage Des Projets Décisionnels Dans Les Grandes Organisations Françaises, Cigref, Octobre 2009.

[Guendouz, 2008] : Moussa Guendouz, Mustapha Labdani Abdessalam, Conception, Implémentation Et Déploiement d'un Système Décisionnel De l'activité Commerciale d'algérie Télécom, 2007-2008.

[Lassence, 2010] : Grégoire De Lassence, Introduction A l'informatique Décisionnelle, 2010

[Kimball, 1996] : Ralph Kimball, A Publié Le Livre De Référence "The Data Warehouse Toolkit », 1996

[Abd, 2012] : Zara Abd, Composants d'un Système d'information Décisionnel, 2012

## **Chapitre Datawarehouse :**

[Nieuwbourg, 2006] : Philippe Nieuwbourg, Analyste Business Intelligence, Qu'est-Ce Qu'un Infocentre ? Décembre 2006.

[Clainche, 2009] : Alexandre Le Clainche, Gestion De Projet Dans Les Ntic, Infocentre, Septembre 2009.

[Boussebat, 2002] : Khaled Boussebat, Mohamed Abdelli, Conception Et Déploiement d'un Entrepôt De Données Métier Cas : La Fonction Commerciale De Naftal Spa, 2002

[E-Business, 2010]: l'encyclopédie E-Business ;  
[http://www.Journaldunet.Com/Encyclopedie/Definition/437/51/20/Executive Information System.Shtml](http://www.Journaldunet.Com/Encyclopedie/Definition/437/51/20/Executive%20Information%20System.Shtml)

[Osbi, 2010] : Osbi.Fr, l'informatique Décisionnelle Open Source Pour Tous.

[Kabore, 2005] : W. Stanislas Kabore, Mise En Place d'un Data Mart Concernant La Paie Du Personnel De l'etat, 2004 – 2005

[Helou, 2004] : Georges El Helou Et Charbel Abou Khalil, Data Mining Techniques d'extraction Des Connaissances, Février 2004.

[Desnos, 2007] : J.-F. Desnos, Entrepôt De Données, 2007

[Ghozzi, 2004] : Faiza Ghozzi Jedidi, Conception Et Manipulation De Bases De Données Dimensionnelles A Contraintes, Novembre 2004.

[Agilteam, 2011] : Agilteam, Les Facteurs De Succès d'un Projet Datawarehouse, 2011.

[Smaïli, 2008] : K. Smaïli, Systèmes d'information Décisionnels Et Datamining. 2008

[Taslimanka, 2007] : Mohamed Taslimanka Sylla, Initiation Au Décisionnel, 2007, Developpez.Com

[Challier, 2006] : Philippe Challier, l'entrepot De Donnees « Sanctuaire » Des Historiques Fiscaux, 2006

[Kimball,2008] : Ralph Kimball, Laura Reeves, Margy Ross, Warren Thornthwaite ; Le Data Warehouse : Guide De Conduite De Projet, 2008

[Haouet, 2008] : Chaker Haouet, informatique décisionnelle et management de la performance de l'entreprise.

### **Chapitre SIG :**

[Sfpt, 1989] : Société Française De Photogrammétrie Et De Télédétection, 1989.

[Ficcdc, 1988] : Comité Fédéral De Coordination Inter-Agences Pour La Cartographie Numérique (Ficcdc, 1988)

[Michel, 1990] : Michel Didier, Conseil National De l'information Géographique, (1990)

[Quodverte, 2003] : Quodverte Philippe, « Système d'information Géographique ». In Jacques Lévy Et Michel Lussault, Avril 2003.

[Tranchant, 2011] : Michael Tranchant, Informatique Décisionnelle, Recherche En Géospatial, Développement, Géomatique, 2011 : [Http://Tranchant.Name/2011/11/Information-Geographique/](http://Tranchant.Name/2011/11/Information-Geographique/)

[Laurini, 1993] : Robert Laurini Et Milleret-Raffort (F.). – Les Bases De Données En Géomatique. Hermès (1993)

[laat, 2003] : Institut Atlantique d'aménagement Du Territoire Poitou-Charentes, Cahier Méthodologique Sur La Mise En Œuvre d'un Sig, 2003

[Botton, 2004] :Serge Botton, Classification Des Représentations Planes Ou Projections, 2004.

[Lévy, 2005] : Lévy J., In Dictionnaire De La Géographie Et De l'espace Des Sociétés, 2005.

[Ird,2001] : Institut De Recherche Pour Le Développement, Qu'est-Ce Qu'un Système d'information Géographique, 2001.

[Esri, 2010] : Environmental Systems Research Institute, Qu'est-Ce Qu'un Sig ?, 2010 : <http://www.esrifrance.fr/Sig4.aspx>

### **Chapitre modélisation multidimensionnelle :**

[ACE, 2010] : ACE , The Electoral Knowledge Network; Projet d'administration et coût, <http://aceproject.org/main/francais/et/ete02.htm>

[Base Plus, 2011] : Base Plus, traitement informatique de la Donnée Marketing, <http://www.base-plus.fr/Une-donnee-pertinente/une-donnee-pertinente-presentation.html>

[Stibo, 2010] : Stibo Systems, The Strategic Information Management Company,  
<http://www.stibosystems.fr/French/societe/Glossaire/A/Accessibilit%C3%A9-des-donn%C3%A9es.aspx>

[Midouni, 2005] : Midouni Sid Ahmed Djallal, Modélisation multidimensionnelle des données complexes : application aux données médicales, Version -9- 2005

[Eurise, 2008] : [http://eurise.univ-st-etienne.fr/bibliographie/fichiers/cf\\_udltddfplcldedd2002.pdf](http://eurise.univ-st-etienne.fr/bibliographie/fichiers/cf_udltddfplcldedd2002.pdf)

[Decizia, 2009] : Decizia, Articles sur la Business Intelligence et le Data Warehousing, <http://decizia.com/blog/category/mdd>

[dwfacile.com] : Data warehouse facile, Modélisation dimensionnelle et data warehousing, [http://www.dwfacile.com/def\\_dim.htm](http://www.dwfacile.com/def_dim.htm)

[Silhadi, 2000] : Samira Silhadi-Hacid, Malika Tarafi ; Du DataWarehouse au WebHouse : le croisement des réseaux et des bases de données.

[Malkovich, 2010]: John Malkovich, starschema vs snowflake, 2010

[Canivet, 2002] : M. Canivet, A.Devisy, B.Dupuy, R.Salmeron, M.Simon, Y.Singer, Livre blanc e-Business intelligence, l'effet internet sur le décisionnel, 2002.

[Helou, 2004] : Georges El Helou et Charbel Abou Khalil, Data Mining Techniques d'extraction des connaissances, 2004

[Rousset, 2009] : Alexis Rousset, Business intelligence, Mai 2009, <http://blog.netapsys.fr/index.php/post/2009/05/28/Naviguation-au-sein-dun-Cube-OLAP>

[Dieungang, 2009] : Marlyse Dieungang – Khaoula Ghilani, Datawarehouse: Cubes OLAP 2009

### **Chapitre conception multidimensionnelle :**

[Kimball, 2002]: Ralph Kimball Margy Ross ; The Data Warehouse Toolkit - The Complete Guide to Dimensional Modeling Second Edition 2002 P33.

[Meylan, 2003] : Eddy Meylan ; Informatique décisionnelle, Ressources humaines & Modélisation dimensionnelle. University of Applied Sciences of Western Switzerland 2003.

[Le Cocq, 2009] : M. Le Cocq, Application Client/Serveur et Web, 2009

[Gardie, 2004] : Michel Gardie, La couche présentation : La syntaxe ASN.1, 2004

[Master, 2009] : Technologies du Web : Accès à une source de données via une interface Web (architecture LAPP), 2009

[Exforsys, 2007]: Exforsys Inc, N-Tier Architecture: The Business Rules Layer, 2007

[ClinKast, 2008] : ClinKast, Architecture J2EE, 2008 : [http://www.clinkast.fr/conseil\\_j2ee02.php](http://www.clinkast.fr/conseil_j2ee02.php)

