

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE MOULOUD MAMMARI DE TIZI-OUZOU



FACULTE DU GENIE ELECTRIQUE ET D'INFORMATIQUE  
DEPARTEMENT D'INFORMATIQUE

## **Mémoire de Fin d'Etudes de MASTER ACADEMIQUE**

Domaine : **Mathématiques et Informatique**

Filière : **Informatique**

Spécialité : **Systèmes Informatiques**

*Présenté par*

**Mohand Ameziane MAZNI**

**Adlane MEZAGUER**

Thème

# **Conception et réalisation d'un système d'aide à la décision Cas : Société Nationale des Assurances « SAA »**

*Mémoire soutenu publiquement le 30 / 06 / 2016 devant le jury composé de :*

**Président : M<sup>elle</sup>. Samia AIT ADDA**

**Encadreur : M<sup>me</sup>. Dalila TAOURI**

**Examineur : M. Hakim RADJA**

**Examineur : M<sup>elle</sup>. Yasmine YESLI**

**2015/2016**

# Remerciements

Avant tout, nous remercions Dieu, le tout puissant, de nous avoir donné le courage et nous avoir guidé pour pouvoir mener à bien ce modeste travail.

Nous tenons à exprimer en tout premier lieu notre gratitude et nos sincères remerciements, à notre promotrice, Madame Taouri Dalila. Merci de nous avoir encadrés, soutenus et si bien orientés tout au long de notre projet.

Nous remercions M. Hadji Madjid et son collègue Rezki pour les aides et les enseignements apportées durant notre stage à la SAA. Ainsi que M. Abane Mouloud et M. Benbouzid Abdelhakim pour leur accueil chaleureux.

Nous remercions les membres du jury pour leurs attentions et intérêts portés envers notre travail. Merci de nous avoir honorés de votre présence.

Enfin, nous remercions l'ensemble des enseignants de département d'informatique, et toute personne qui a contribué de près ou de loin à l'élaboration de ce projet.

*Mohand & Adlane*

# Dédicaces

*Je dédie ce modeste travail ;*

*À mes parents qui n'ont jamais cessé de m'encourager et me  
soutenir,*

*À mes soeurs ,*

*À mes grands parents,*

*À tous les membres de ma famille, paternelle et maternelle,*

*À mon binôme,*

*À tous mes amis,*

*À toutes les personnes que je connais et que je n'ai pas citées.*

***Mohand A.***

# Dédicaces

*Je dédie ce modeste travail ;*

*À mes parents qui n'ont jamais cessé de m'encourager et me  
soutenir,*

*À mes soeurs ,*

*À mes grands parents,*

*À tous les membres de ma famille, paternelle et maternelle,*

*À mon binôme,*

*À tous mes amis,*

*À toutes les personnes que je connais et que je n'ai pas citées.*

***Adlane***

# Table des matières

<b>I Synthèse bibliographique</b>	<b>20</b>
<b>1 Introduction au Business Intelligence (BI)</b>	<b>21</b>
1.1 Historique des systèmes décisionnels . . . . .	21
1.2 Définition des systèmes décisionnels . . . . .	22
1.3 Architecture du Système décisionnel . . . . .	23
1.3.1 Les sources de données (Zone de préparation) . . . . .	23
1.3.2 Les entrepôts de données et datamarts (Zone d'entreposage) . . . . .	23
1.3.3 Les outils d'analyse et de restitution (Zone de restitution) . . . . .	24
1.4 Fonctionnement d'une chaîne décisionnelle . . . . .	24
1.4.1 La phase de la collecte de données . . . . .	24
1.4.2 La phase de stockage . . . . .	25
1.4.3 Phase de d'exploitation . . . . .	25
1.5 Système décisionnel VS système transactionnel . . . . .	25
1.6 Les logiciels au cœur de la décision . . . . .	26
1.7 Pentaho . . . . .	27
1.7.1 PENTAHO DATA INTEGRATION (PDI) . . . . .	28

<b>2</b>	<b>Les Entrepôts de Données ( Data Warehouses )</b>	<b>30</b>
2.1	Quelques définitions . . . . .	30
2.2	Caracteristiques essentiels du Data Warehouse . . . . .	31
2.3	Structure de données d'un Data Warehouse . . . . .	33
2.4	Les éléments composant un Data Warehouse . . . . .	34
2.4.1	Système source . . . . .	34
2.4.2	La zone de préparation de données . . . . .	34
2.4.3	La zone de présentation de données . . . . .	35
2.4.4	Outils d'accès aux données . . . . .	35
2.5	Le reporting : . . . . .	36
2.5.1	Objectifs du reporting . . . . .	37
2.5.2	Les types de reporting . . . . .	37
<b>3</b>	<b>Modélisation multidimensionnelle</b>	<b>38</b>
3.1	Modélisation dimensionnelle et ses concepts . . . . .	38
3.1.1	La table de faits . . . . .	39
3.1.2	La table de dimension . . . . .	39
3.2	Les différents modèles de la modélisation dimensionnelle . . . . .	40
3.2.1	Modèle en étoile . . . . .	40
3.2.2	Modèle en flocon . . . . .	40
3.2.3	Modèle en constellation . . . . .	41
3.3	Le cube de données . . . . .	41
3.4	Le concept OLAP . . . . .	42
3.4.1	Qu'es ce que l'OLAP ? . . . . .	42
3.4.2	Et l'OLTP ? . . . . .	43
3.4.3	OLAP Vs OLTP . . . . .	43

3.4.4	Architecture des serveurs OLAP . . . . .	43
3.4.4.1	Architecture MOLAP « Multidimensional OLAP » . . . . .	43
3.4.4.2	Architecture ROLAP « Relational OLAP » . . . . .	44
3.4.4.3	Architecture HOLAP « Hybrid OLAP » . . . . .	44
3.4.4.4	Autres architectures OLAP . . . . .	45
3.4.4.5	Comparaison MOLAP Vs ROLAP Vs HOLAP . . . . .	45
3.5	La navigation dans les données . . . . .	45
3.5.1	Les opérations ensemblistes « Slice & Dice » . . . . .	45
3.5.2	Les opérations de granularité Roll-up & Drill-down . . . . .	46
3.6	Conduite et construction du Data Warehouse . . . . .	47
3.6.1	L'approche conductrice à la mise en place d'un Data Warehouse [Kimball ,1996] . . . . .	47
3.6.2	Alimentation du Data Warehouse . . . . .	48
3.6.3	Le déploiement et mise en oeuvre . . . . .	50
3.6.4	Maintenance et croissance . . . . .	51
<b>II</b>	<b>Analyse et conception</b>	<b>54</b>
<b>4</b>	<b>Analyse et diagnostic de la situation existante</b>	<b>55</b>
4.1	Les Assurances . . . . .	55
4.1.1	Définition de l'opération d'assurance . . . . .	55
4.1.2	L'assurance à travers l'histoire . . . . .	56
4.1.3	Les produits d'assurance . . . . .	56
4.1.4	La réassurance . . . . .	57
4.1.5	Le marché mondial des assurances . . . . .	57
4.1.6	Le marché Algérien des assurances . . . . .	58
4.1.7	Assureurs Algériens . . . . .	59

4.2	Présentation de l'organisme d'accueil « SAA » [DI]	60
4.2.1	Division et filiales	61
4.2.2	La SAA en chiffres	62
4.2.3	L'organisation générale de la SAA	63
4.2.3.1	Direction générale de la société	63
4.2.3.2	Direction régionale	64
4.2.3.3	Les agences	64
4.2.4	Organigrammes	65
<b>5</b>	<b>Conception de la solution</b>	<b>70</b>
5.1	Conception de la zone d'entreposage	70
5.1.1	Formalisme utilisé	71
5.1.2	Modélisation dimensionnelle des activités	71
5.1.2.1	Modélisation dimensionnelle de l'activité « Production »	71
5.1.2.2	Modélisation dimensionnelle de l'activité « Suivi Sinistre »	74
5.2	Conception de la zone d'alimentation	76
5.2.1	Identification des sources de donnée	77
5.2.2	Processus d'alimentation	77
5.2.2.1	Extraction des données	78
5.2.2.2	Préparation du chargement : traitements des données	78
5.2.3	Préparation du chargement des dimensions	78
5.2.3.1	Chargement des dimensions	78
5.2.3.2	Chargement des dimensions temporelles	79
5.2.3.3	Préparation du chargement des tables de faits	79
5.3	Approche adoptée	86
5.3.1	Identification des utilisateurs	87
5.3.2	Les besoins techniques	88
5.3.3	Les besoins fonctionnels	88
5.3.4	Identification des cas d'utilisation	89
5.3.4.1	Les cas d'utilisation fonctionnels	89

<b>6</b>	<b>Déploiement de la solution</b>	<b>91</b>
6.1	Architecture technique de la solution . . . . .	91
6.2	Les outils utilisés . . . . .	93
6.2.1	Serveur de base de données . . . . .	94
6.2.2	Pentaho BI suite . . . . .	94
6.2.2.1	Le serveur d'application Pentaho . . . . .	94
6.3	Présentation de l'architecture de la solution . . . . .	96
6.4	Etapes de la réalisation de la solution . . . . .	97
6.4.1	La réalisation de la zone d'entreposage . . . . .	97
6.4.2	La réalisation de la zone d'alimentation . . . . .	97
6.5	La réalisation du Reporting . . . . .	99
6.6	Présentation du prototype de l'application . . . . .	99
6.7	La sécurité du système . . . . .	102

# Table des figures

0.0.1 Architecture d'un système d'information décisionnel . . . . .	16
0.0.2 L'ancien schema entité/association . . . . .	17
1.2.1 Le décisionnel au sein du Système d'information. . . . .	23
1.4.1 Architecture du système d'information décisionnel. . . . .	24
1.5.1 Comparaison entre le système décisionnel et le système transactionnel . . . . .	26
2.2.1 Différences entre système de production et Data Warehouse . . . . .	32
2.3.1 Structure de données dans le Data Warehouse. . . . .	33
2.4.1 Finalités des Datamart et Data Warehouse . . . . .	35
2.5.1 Exemple d'un état de sortie en reporting. . . . .	36
3.1.1 Considération d'un sujet d'analyse comme un cube à plusieurs dimensions. [Franco, 1997] . . . . .	38
3.1.2 Exemple d'une table de fait. [Franco, 1997] . . . . .	39
3.1.3 Exemple d'une table de dimension. [Franco, 1997] . . . . .	39
3.2.1 Exemple d'un schéma étoile. . . . .	40
3.2.2 Exemple d'un schéma flocon. [Franco, 1997] . . . . .	40
3.2.3 Exemple d'un schéma en constillation . . . . .	41
3.3.1 Exemple d'un hyper-cube. [Web5] . . . . .	42
3.4.1 Comparaison OLTP Vs OLAP. [Web6] . . . . .	43
3.4.2 Schéma MOLAP. [Nakache 1998] . . . . .	44

3.4.3 Schéma ROLAP. [Nakache 1998] . . . . .	44
3.4.4 Comparaison entre les différentes architectures d'OLAP. [Grislin, 2000] . . . . .	45
3.5.1 Exemple de Slicing. . . . .	46
3.5.2 Exemple de Dicing. . . . .	46
3.5.3 Exemple Roll-up. . . . .	46
3.5.4 Exemple Drill-down du niveau des régions au niveau des villes . . . . .	47
3.6.1 Le système d'alimentation ETL. . . . .	49
4.1.1 Structure du marché des assurances en Algérie. . . . .	60
4.2.1 Evolution du chiffre d'affaires par branche 2010-2013 (En millions de DA) . . . . .	62
4.2.2 Evolution des indemnisations par branche (En millions de DA) . . . . .	62
4.2.3 Organigramme de la direction générale . . . . .	66
4.2.4 Organigramme des directions régionales . . . . .	67
4.2.5 Organigramme des agences : . . . . .	68
5.1.1 Formalisme adopté pour la table de faits . . . . .	71
5.1.2 Formalisme adopté pour la table de dimension . . . . .	71
5.1.3 Formalisme adopté pour la relation entre les deux tables . . . . .	71
5.1.4 Schéma de base du modèle dimensionnel de l'activité « Production » . . . . .	72
5.1.5 Dimension Vehicule . . . . .	73
5.1.6 Dimension Agence . . . . .	73
5.1.7 Dimension Garantie . . . . .	74
5.1.8 Dimension Temps . . . . .	74
5.1.9 Schéma de base du modèle dimensionnel de l'activité « Suivi sinistre » . . . . .	75
5.1.10 Dimension Accident . . . . .	76
5.1.11 Dimension Ordre de médecin . . . . .	76
5.1.12 Dimension Jugement . . . . .	77

5.2.1 Lancement de SPOON.BAT . . . . .	80
5.2.2 Creation de la transformation . . . . .	80
5.2.3 Nouvelle transformation . . . . .	81
5.2.4 Nouvelle connexion . . . . .	81
5.2.5 Configuration de la connexion . . . . .	82
5.2.6 Choix d'extraction . . . . .	82
5.2.7 Configuration de l'extraction . . . . .	82
5.2.8 Script . . . . .	83
5.2.9 Filtrage . . . . .	83
5.2.10 Flux . . . . .	84
5.2.11 Configuration du filtre . . . . .	84
5.2.12 Structure de flux . . . . .	85
5.2.13 Configuration de l'alimentation excel . . . . .	85
5.2.14 Lancement de la transformation . . . . .	85
5.2.15 Demarrer . . . . .	86
5.2.16 Resultat . . . . .	86
5.3.1 Place et importance des besoins dans un projet décisionnel. . . . .	87
5.3.2 Le futur système vu par les responsables informatiques . . . . .	88
6.1.1 Architecture du système mis en place. . . . .	92
6.2.1 Interface d'identification du serveur Pentaho. . . . .	94
6.2.2 Console d'utilisateur du serveur Pentaho . . . . .	95
6.2.3 Console d'administration du serveur Pentaho. . . . .	96
6.3.1 Architecture globale de la solution. . . . .	97
6.4.1 Création et chargement de la table temps . . . . .	99
6.6.1 Authentification . . . . .	100
6.6.2 Sélection des sources de données. . . . .	100
6.6.3 Rapport sous format d'un tableau croisé . . . . .	101
6.6.4 Rapport sous format d'un graphe. . . . .	101
6.6.5 Capture d'écran du tableau de bord. . . . .	102

# Liste des tableaux

1.1	Tableau comparatif entre l'infocentre et le Datawarehouse . . . . .	22
4.1	Les catégories d'assurance . . . . .	56
4.2	Les 10 premières entreprises mondiales d'assurance en termes de chiffres d'affaires. . . . .	58
4.3	l'évolution du chiffre d'affaires du secteur des assurances en milliards de dinars . . . . .	58
4.4	Evolution de la sinistralité (Millions DA) . . . . .	58
5.1	Liste des groupes d'utilisateurs . . . . .	87
5.2	Les exigences des responsables informatiques . . . . .	88
5.3	Liste des cas d'utilisation fonctionnels. . . . .	90
6.1	Tableau des outils . . . . .	93
6.2	Tableau des transformations utilisées . . . . .	98

# Motivations et objectifs

De nos jours, le vrai challenge auquel font face les entreprises n'est plus la collecte et le stockage de données mais plutôt l'exploitation de ce patrimoine de données à des fins d'analyse. Conscientes de l'importance des informations qu'on peut tirer de ces analyses de données, les entreprises se sont lancées sur la piste des outils d'aide à la décision sous toutes leurs formes. En 2009 par exemple, le marché du décisionnel était de 2.052 milliards d'Euro rien que pour la France et 9 milliards de dollars dans le monde. [Poletto, 2012]

Un marché qui a connu une croissance de 5.9% durant la même année, ce qui représente 3 fois la croissance du marché du logiciel. Plus encore, l'étude annuelle de Gartner : « Gartner EXP CIO Survey 2012 »<sup>1</sup> a montré que le premier centre d'intérêt des DSI<sup>2</sup> de plus de 45 pays demeure la catégorie : « Analytics and Business Intelligence », en français : « Outils d'analyse et informatique décisionnelle », juste avant les technologies mobiles et le Cloud Computing.

Comme on peut facilement le déduire, les décideurs sont les principaux utilisateurs des systèmes décisionnels. Les décideurs sont généralement des « marketeurs » ou analystes. Ces derniers établissent des plans marketing qui leur permettent de mieux cibler leurs clients, de les fidéliser etc. Et pour cela, ils ont besoin d'indicateurs et des données résumant leurs activités (ils n'ont souvent besoin du détail que pour des cas spécifiques).

Par exemple, et contrairement aux systèmes de production (ou base gestion) où les utilisateurs chercheront à connaître leurs transactions pour faire un bilan, les systèmes décisionnels quant à eux cherchent plutôt à donner un aperçu global pour connaître les tendances des clients.<sup>3</sup>

---

1. Une entreprise américaine de conseil et de recherche dans le domaine des techniques avancées. Elle tient à jour différentes statistiques et maintient un service de nouvelles spécialisées.

2. Direction des systèmes d'informations

3. <http://taslimanka.developpez.com/tutoriels/bi/>

Le marché de l'assurance s'inscrit dans un contexte en pleine mutation. Cette mutation est telle qu'elle implique une évolution du business model, aujourd'hui fondé sur un calcul de risques établi sur des données historiques. Cette évolution, voire rupture, va pouvoir s'appuyer sur de nouvelles technologies, rendant possibles l'accès et la compréhension d'un volume de données en croissance exponentielle.

Valoriser et exploiter le capital data pour personnaliser la proposition de valeur auprès de ses assurés, en concevoir de nouvelles, mieux gérer le risque grâce à des analyses prédictives, développer une image moderne : telles sont les perspectives des usages des entrepôts de données par les assureurs. La puissance des entrepôts de données peut être mise à disposition de toutes les fonctions métier sur la chaîne de valeur de l'assurance.

### **Les défis de l'informatique décisionnelle**

Les avantages que procure un système décisionnel sont nombreux, alors pourquoi ne sont-ils pas déployés massivement ? L'implémentation d'une telle technologie doit relever deux défis qui inhibent encore aujourd'hui l'adoption de cet outil par les petites et moyennes structures et qui en limitent l'usage pour les plus grandes.

Premièrement, l'informatique décisionnelle nécessite une forte expertise technique et exige le recours à des logiciels optimisés et fortement intégrés à l'infrastructure existante : systèmes d'exploitation, systèmes de gestion de bases de données avancés, outils d'intégration de données, système de reporting et d'analyse. Les gros volumes de données, les requêtes complexes et contraintes de performances induites par ces outils rendent le succès du déploiement fortement dépendant du niveau de compétences et des choix technologiques. Le second défi est sans aucun doute lié aux coûts de ces systèmes. Des technologies avancées sont exigées et les solutions d'informatique décisionnelles ont été conçues pour de grandes organisations à forts budgets. Car les solutions complexes présentant un fort coût d'acquisition et de maintenance.

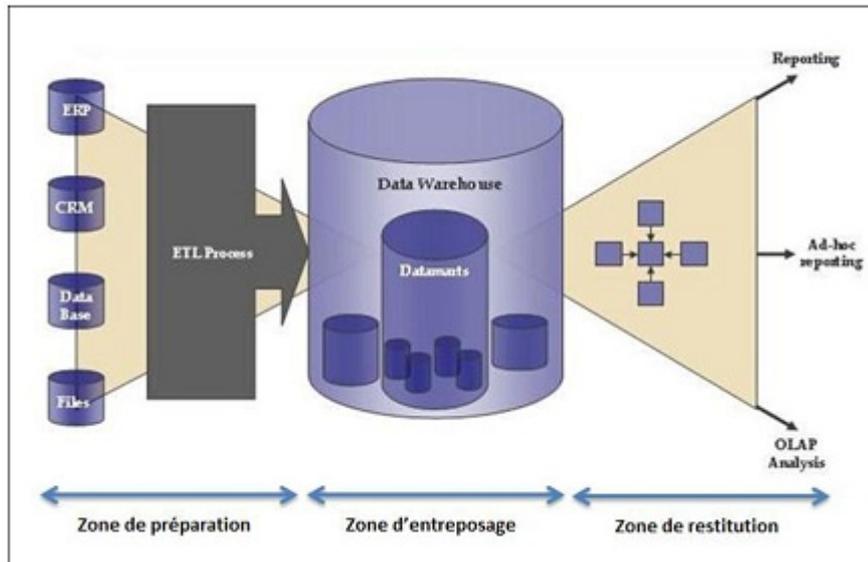


FIGURE 0.0.1 – Architecture d'un système d'information décisionnel

De nouvelles possibilités apparaissent cependant, au plus grand profit des organisations à budget limité. Des logiciels Open Source sont aujourd'hui disponibles pour toutes les fonctions d'une plateforme décisionnelle délivrant un niveau d'usage, de capacité à monter en charge et de robustesse compatibles avec des besoins critiques.

## Problématique

L'analyse de de l'environnement de la SAA permet de constater, d'une part, une explosion du volume de données manipulées (Plus d'un million d'enregistrements dans la plupart des tables), et un élargissement de la taille de la base de données opérationnelle.

D'autre part, les rapports proposés par les différentes applications métier ne donnent plus satisfaction à cause de leurs limites (rapports figés et non paramétrés) et des temps de réponses qui sont de plus en plus lents. Exemple, l'ancien schema entité/association :

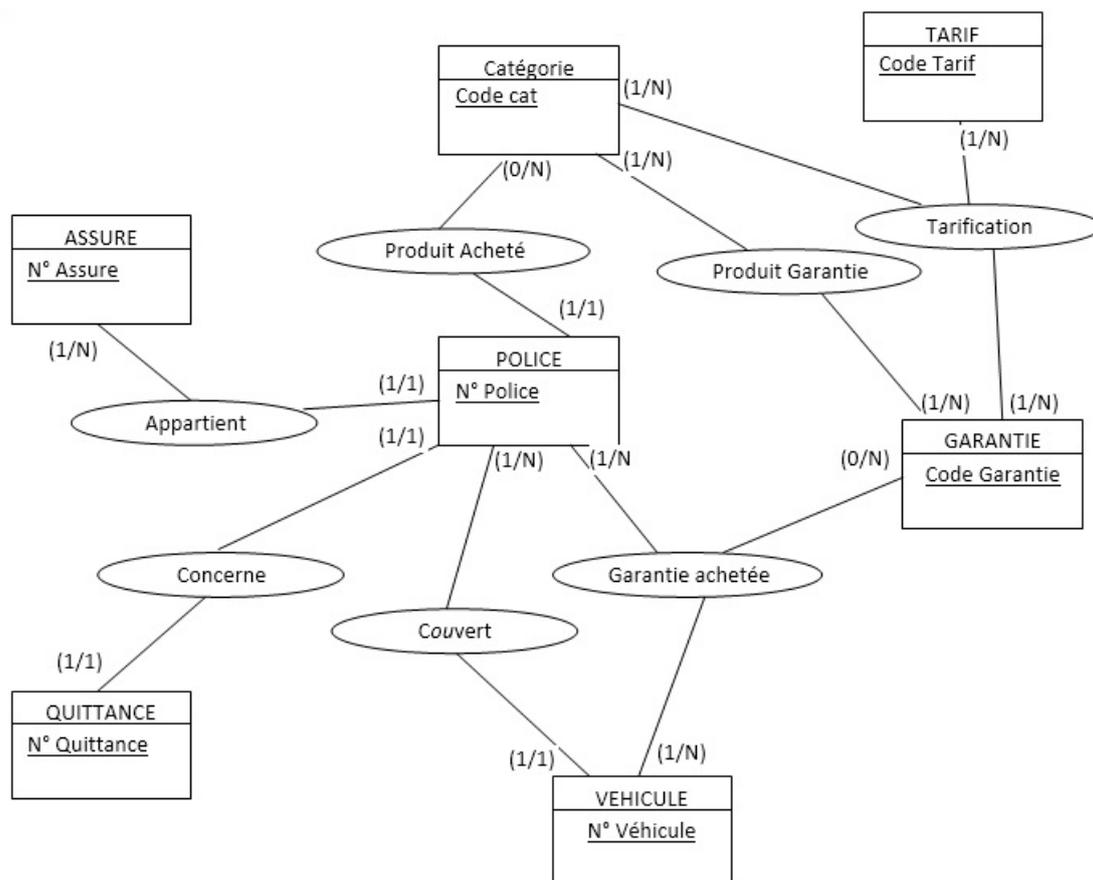


FIGURE 0.0.2 – L’ancien schéma entité/association

Encore plus, et en l’absence d’un outil destiné à supporter les décideurs dans l’accomplissement de leurs fonctions, les collaborateurs de la DSI trouvent souvent des difficultés à répondre aux requêtes des cadres dirigeants qui sont devenues fréquentes, complexes et consomment beaucoup de temps. Afin de pallier à tous ces problèmes, la SAA voudrait mettre en place un système d’information décisionnel proposant un système de reporting et un tableau de bord regroupant un ensemble d’indicateurs pour la mesure de performance de l’entreprise.

### Objectifs initiaux

Le projet de mise en place d’un système d’information décisionnel vise à :

- ▶ Réduire les temps de réponse de la base de données en réorganisant dimensionnellement les données de l’entreprise.
- ▶ Réduire la charge de travail des collaborateurs de la DSI en mettant à leur disposition un système de reporting afin de répondre aux requêtes des décideurs.

- ▶ Augmenter la réactivité des décideurs en mettant à leur disposition un ensemble d'indicateurs de performance (KPI).

## **Organisation du mémoire**

Le mémoire est organisé en deux grandes parties.

### **Partie Théorique, Regroupe les chapitres suivants :**

▶ **Chapitre I** : Introduction au Business Intelligence (BI)

Ce chapitre décrit brièvement les caractéristiques d'un système d'information décisionnel qui est le thème de ce projet de fin d'études.

▶ **Chapitre II** : Les Entrepôts de Données ( Datawarehouses )

Un chapitre dédié aux entrepôts de données qui sont considérés comme la pierre angulaire de tout système décisionnel.

▶ **Chapitre III** : Modélisation des données du Data Warehouse

Les outils de restitution de données étant la partie visible à l'utilisateur, ce chapitre reprend les principaux outils de la zone de présentation de données, notamment le reporting.

### **Partie Pratique, Regroupe les chapitres suivants :**

▶ **Chapitre IV** : Analyse et diagnostic de la situation existante

Ce chapitre reprend les principales notions du domaine des assurances qui est le secteur d'activité de l'organisme d'accueil, et présentation de ce dernier.

▶ **Chapitre V** : Conception de la solution

Conception de l'entrepôt de données et les tables de reporting.

▶ **Chapitre VI** : Déploiement de la solution

Une présentation des outils utilisés et de la politique de mise en place du système.

## Première partie

# Synthèse bibliographique

# Chapitre 1

## Introduction au Business Intelligence (BI)

Au cours des dix dernières années, la concurrence croissante dans le climat actuel des affaires a considérablement augmenté l'intérêt d'analyser des données, car l'information est le pilier central du processus de prise de décision. Aujourd'hui, la clé pour la survie dans le monde de l'entreprise est d'être capable d'analyser, de planifier et de réagir aux conditions changeantes aussi vite que possible pour prendre les bonnes décisions commerciales stratégiques et tactiques.

### 1.1 Historique des systèmes décisionnels

Les préoccupations d'ordre décisionnel ont toujours existé et l'idée de mettre en place un système spécifique pour y répondre est apparue très tôt dans l'histoire de l'informatique. Au démarrage les applications de production et d'aide à la décision cohabitaient sur un même serveur informatique et utilisaient une même base de données, ceci fut très conflictuel, car les applications d'analyse sont de fortes consommatrices de ressources et peuvent ainsi extrêmement dégrader le temps de réponse des autres applications. Ceci a motivé la séparation entre un système opérationnel et un système dédié au décisionnel.

Dans un premier temps, la réponse apportée fut la création de bases de données dites « infocentres ». C'est une base de données qui est la duplication exacte de la base de données du système opérationnel qui servira à des fins d'analyse. Mais cette dernière ne permet de répondre qu'à des besoins décisionnels simples et son utilisation (complexe) demeure réservée à un petit groupe d'informaticiens.

Afin de palier à tous les inconvénients de l'infocentre, le concept d'une base véritablement décisionnelle a commencé à voir le jour, et les termes d'entrepôt de donnée et SID sont alors apparus pour désigner cette nouvelle notion. Nous sommes ainsi passés du principe d'isoler les données en vue d'une utilisation décisionnelle (infocentre) à des bases de données structurées en fonction des problématiques d'analyse de données propres à chaque entreprise (SID).

Le Tableau I.2 propose une comparaison entre l'infocentre et le Datawarehouse :

<b>Infocentre</b>	<b>Datawarehouse</b>
Collection de données	Collection de données
Orientés sujets	Orientés sujets
Intégrées	Intégrées
Volatiles	Non Volatiles
Actuelles	Historisées
Organisées pour le support d'un processus de prise de décision ponctuel.	Organisées pour le support d'un processus d'aide à la décision
Outils	Architecture

Table 1.1 – Tableau comparatif entre l'infocentre et le Datawarehouse

## 1.2 Définition des systèmes décisionnels

«La prise de décision implique la sélection d'un plan d'action parmi deux ou plusieurs solutions de rechange possibles en vue de parvenir à une solution à un problème donné». [Trewartha, al, 1982]

Le système décisionnel collecte les données internes et externes, les transforme en informations et les analyse pour prendre les meilleures décisions et améliorer la performance de l'entreprise.

Cela ne veut certainement pas dire qu'un tel système, aussi intelligent soit-il, peut remplacer les êtres humains pour la prise de décision. Par contre, au lieu de prendre des décisions dans le flou, les décisions seront prises en connaissance de cause.<sup>1</sup>

### ► Position du SID dans l'entreprise :

---

1. <http://informatique-decisionnelle.masteres.eisti.fr/index.php/lenseignement/ledecisionnel>

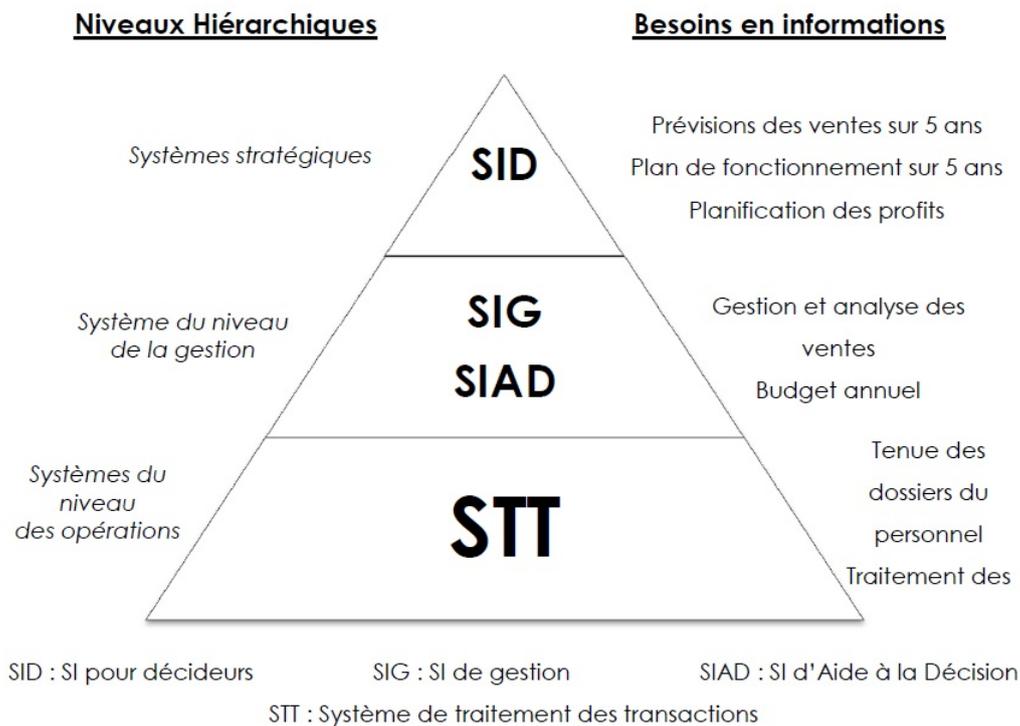


FIGURE 1.2.1 – Le décisionnel au sein du Système d'information.

La figure ci-dessus illustre parfaitement la place qui revient au décisionnel au sein d'une entreprise entre l'opérationnel et le pilotage.

### 1.3 Architecture du Système décisionnel

Tout système décisionnel est composé de trois zones importantes qui sont :

#### 1.3.1 Les sources de données (Zone de préparation)

Dans une première étape, il est nécessaire de recueillir et d'intégrer les données stockées dans les diverses sources primaires et secondaires, qui sont hétérogènes à l'origine. Les sources se composent pour la plupart des données appartenant aux systèmes opérationnels, mais peuvent aussi inclure des documents non structurés, tels que des courriels et des données reçues des fournisseurs externes. De manière générale, un effort majeur est nécessaire pour unifier et intégrer les différentes sources de données.

#### 1.3.2 Les entrepôts de données et datamarts (Zone d'entreposage)

L'entrepôt de données - ou « DataWarehouse » en anglais est le cœur du système décisionnel, il contient les données provenant de différentes sources. L'utilisation d'outils d'extractions, de

transformations et de chargements, connus sous le nom d'« ETL », sera nécessaire pour charger les données dans l'entrepôt de données.

### 1.3.3 Les outils d'analyse et de restitution (Zone de restitution)

Les données sont finalement prêtes à être analysées selon différents axes d'analyse, afin de donner un résultat qui sera utilisé pour la prise de décisions.

## 1.4 Fonctionnement d'une chaîne décisionnelle

Un projet décisionnel est composé de trois phases clés : la phase de collecte, la phase de stockage et enfin la phase d'exploitation. Voir figure 1.4.1

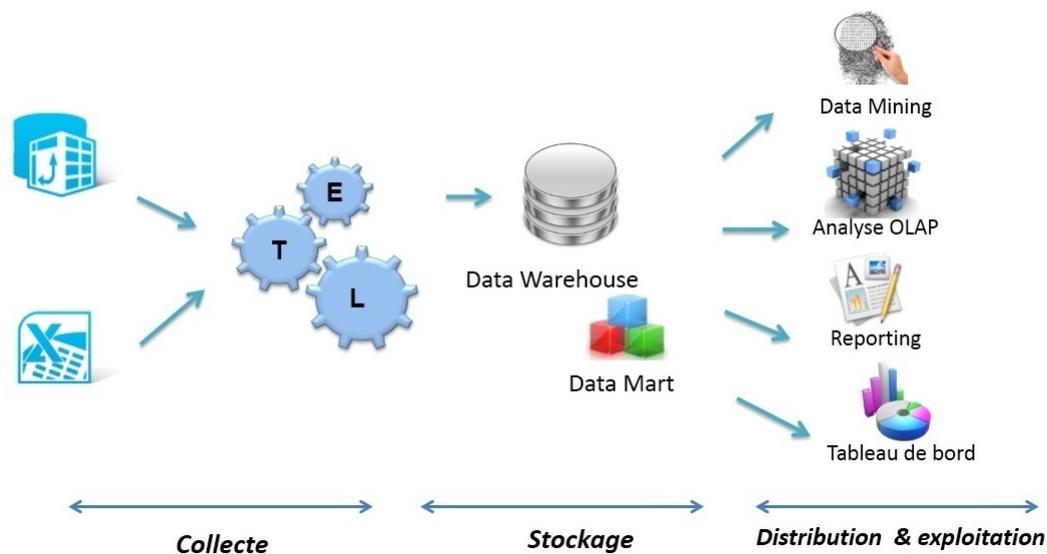


FIGURE 1.4.1 – Architecture du système d'information décisionnel.

### 1.4.1 La phase de la collecte de données

La collecte s'effectue à partir des sources de l'entreprise qui peuvent être présentées sous différents formats (fichiers plats ou bien des bases de données). Ces sources sont en général hétérogènes, c'est pourquoi il faut les manipuler avant de les stocker dans un SID (Système d'Information Décisionnel). Cette manipulation s'effectue à l'aide d'une famille d'outils dénommée ETL (Extract Transform Load) offrant des fonctions d'extraction, de transformation et de chargement des données.

### 1.4.2 La phase de stockage

Les données nettoyées et consolidées, sont stockées dans une base spécialisée (comme le Data Warehouse ou le Data Mart) afin qu'elles puissent être utilisées à des fins décisionnelles. Afin de ne pas polluer le Data warehouse, il est souhaitable d'utiliser une zone de stockage intermédiaire (ODS intermédiaire ou Staging area). C'est une structure destinée à stocker les données issues des systèmes de production opérationnels. C'est une zone de préparation des données avant leurs intégrations dans le Data warehouse.

### 1.4.3 Phase de d'exploitation

Cette dernière phase concerne la restitution des résultats. Les outils de restitution sont la partie visible offerte aux utilisateurs. On distingue à ce niveau plusieurs types d'outils :

- ▶ Les outils de reporting et de requêtes qui permettent la mise à disposition de rapports.
- ▶ Les outils d'analyse OLAP pour les analyses multidimensionnelles.
- ▶ Les outils de Datamining pour la recherche des corrélations.
- ▶ Les tableaux de bord afin de piloter la performance

## 1.5 Système décisionnel VS système transactionnel

Il est largement reconnu de nos jours les besoins spécifiques des entrepôts de données et leur divergence par rapport aux systèmes transactionnels tant du point de vue conceptuel et structurel que du point de vue utilisation et résultats. Si les uns font tourner l'organisation, les autres observent comment tourne cette dernière. Le tableau ci-après montre les différences entre le système décisionnel et le système transactionnel, en les comparant par rapport aux données qu'ils gèrent et les leurs utilisations.

Différence	Systèmes transactionnels	SID
Par les données	Orienté application	Orienté thèmes et sujets
	Situation instantanée	Situation historique
	Données détaillées et codées non redondantes	Informations agrégées cohérentes souvent avec redondance
	Données changeantes constamment	Informations stables et synchronisées dans le temps
	Pas de référentiel commun	Un référentiel unique
L'usage	Assure l'activité au quotidien	Permet l'analyse et la prise de décision
	Pour les opérationnels	Pour les décideurs
	Mises à jour et requêtes simples	Lecture unique et requêtes complexes transparentes
	Temps de réponse immédiats	Temps de réponses moins critiques
	Faibles volume à chaque transactionnel	Large volume manipulé
	Conçu pour la mise à jour	Conçue pour l'extraction
	Usage maîtrisé	Usage aléatoire

FIGURE 1.5.1 – Comparaison entre le système décisionnel et le système transactionnel

## 1.6 Les logiciels au cœur de la décision

Les entreprises ressentent de plus en plus le besoin de prendre des décisions. Ces dernières doivent être les plus fiables possibles et s'inscrivent dans des stratégies de plus en plus élaborées. Les logiciels de l'informatique décisionnelle ont accès, et utilisent, les données de la base de données de l'entreprise ainsi que de leur Progiciel de Gestion Intégrée ( "Enterprise Resource Planning" en anglais ). Voici quelques exemples de logiciels décisionnels qu'ils soient libres ou propriétaires :<sup>2</sup>

**SAS** (logiciel propriétaire) est le plus important éditeur indépendant du marché du décisionnel. Avec plus de 60 000 sites clients répartis dans 134 pays. Plusieurs applications lui sont propres : tableaux, graphiques...

**Business Objects** (logiciel propriétaire) : il permet l'accès aux applications habituelles du décisionnel (tableaux, graphiques...) ainsi qu'à internet. Birt, qui signifie Business Intelligence and Reporting Tools, est un projet de la Fondation Eclipse créé et animé depuis 2004.

**Birt** figure parmi les outils Open Source de Business Intelligence les plus largement diffusés et utilisés. Le nombre élevé de développeurs dans le monde(un million) témoigne de la facilité d'adoption de cet outil.

**Cognos** est une compagnie canadienne, filiale d'IBM depuis 2008. Fondée en 1969, la compagnie édite des solutions décisionnelles et propose des services permettant aux entreprises de piloter, superviser et comprendre leur performance. Les solutions qu'elle propose aident les entreprises à mesurer, comprendre et anticiper leur performance opérationnelle et financière et apportent un éclairage étayé sur les meilleures actions à mener. Elles

2. [https://fr.wikiversity.org/wiki/Prévision\\_décisionnelle/Application\\_de\\_la\\_prévision](https://fr.wikiversity.org/wiki/Prévision_décisionnelle/Application_de_la_prévision)

contiennent entre autres : reporting, tableaux de bord, analyse des résultats et des tendances à venir, planification et élaboration budgétaire, partage simplifiée des informations pour qu'elles soient accessibles à un grand nombre d'acteurs de l'entreprise. . .

**BI Square Software** est un éditeur de logiciels de Business Intelligence. Créé en 2010, il fournit des tableaux de bord décisionnels sur l'IT afin que les entreprises pilotent et optimisent au mieux leur infrastructure informatique (serveurs, stockage, réseau...), en s'appuyant sur les bonnes pratiques ITIL et Green IT. Soutenu par Oseo Innovation et Paris Région, il est lauréat de Scientipôle Initiative.

**SpagoBI** est une solution de Business Intelligence entièrement Open Source. SpagoBI est distribué sous licence Mozilla Public License, qui est compatible avec les usages commerciaux. SpagoBI Server est le module principal de la suite, qui offre toutes les fonctionnalités centrales et analytiques de la plateforme. Il est basé sur deux modèles théoriques (modèle analytique et modèle comportemental) et fournit un éventail de services administratifs et transversaux. Le Modèle Analytique est le noyau de SpagoBI Server. Il couvre l'ensemble des besoins analytiques, en fournissant plusieurs solutions pour chaque domaine analytique : reporting, tableaux de bord « en temps réel », modèles d'Indicateurs clé de performance pour développer et tester des modèles de monitoring des performances, analyse géo-référencée pour l'affichage des informations sur une base géographique, fichiers Office pour publier et exécuter des fichiers Office sous le contrôle du modèle comportemental, processus externes, pour exécuter les processus externes qui peuvent interagir avec des systèmes de traitement de transactions en ligne (OLTP)...

**Pentaho** est une solution, accessible en Open Source, d'informatique décisionnelle développée en Java. L'offre principale de l'Américain Pentaho repose sur sa suite qui rassemble des outils d'intégration de données (ETL), d'analyse, de reporting, de tableaux de bord et de datamining. Pentaho est de plus en plus présent en France, en Europe.

*Remarque.* Dans le cadre de notre travail, nous allons nous intéresser particulièrement à la solution Pentaho que nous allons décrire plus en détail dans ce qui suit.

## 1.7 Pentaho

*« Pentaho BI Suite est l'une des offres logiciels Open Source les plus complètes pour la réalisation de projet BI, ce logiciel est fourni par la compagnie Pentaho. Cette compagnie a été fondée en 2004 et est dirigée par des anciens responsables de sociétés ayant connu des succès importants dans le domaine de la Business Intelligence (BI. Tels que IBM, Business Objects, Hyperion, Oracle, Cognos et SAS. En plus d'avoir glané les principales récompenses dans le monde du logiciel libre, Pentaho a levé en février 2008, 12 millions de dollars de fonds, qui lui permettent d'accélérer l'évolution de sa suite » [Glauser, 2008].*

## 1.7.1 PENTAHO DATA INTEGRATION (PDI)

Pentaho Data Integration (PDI) est un ETL open source qui permet de concevoir et exécuter des opérations de manipulation et de transformation de données. Grâce à un modèle graphique à base d'étapes, il est possible de créer sans programmation des processus composés d'imports et d'exports de données, et de différentes opérations de transformation, telles que des conversions, des jointures, l'application de filtres, ou même l'exécution de fonctions Javascript. PDI, initialement connu sous le nom de Kettle, est devenu open source à partir de la version 2.2 et a rapidement intégré le projet de plateforme décisionnelle Pentaho qui l'a renommé depuis en Pentaho Data Integration. Pentaho Data Integration est disponible dans sa version 4.4.0. De nouvelles fonctionnalités sont apparues facilitant notamment le partage d'un référentiel entre plusieurs développeurs, l'utilisation de métadonnées dans les transformations (afin de personnaliser une transformation rapidement) et permettant de se connecter à toujours plus de systèmes de bases de données ou types de fichiers. Dans sa version Entreprise Edition, Pentaho, l'éditeur de PDI, propose de suivre l'exécution des transformations ETL via la console d'administration, en mode Web. Cette version propose en outre un référentiel Entreprise, qui permet le versionnement des développements.

## Les fonctionnalités de PDI

PDI permet de créer deux types de processus :

- ▶ Les transformations : traitements effectués au niveau d'une ou plusieurs bases de données comprenant des opérations de lecture, de manipulation et d'écriture. C'est à ce niveau que sont manipulées les données.
- ▶ Les tâches : traitements de plus haut niveau, combinant des actions telles que l'exécution d'une transformation PDI, l'envoi d'un mail, le téléchargement d'un fichier ou le lancement d'une application. Il est possible d'exécuter des actions différentes en fonction de la réussite ou de l'échec de chaque étape. Le rôle d'une tâche est donc d'orchestrer les différents traitements.

PDI peut se connecter sur un grand nombre de bases de données, dont Oracle, Sybase, MySQL, PostgreSQL, Informix, SQLServer et bien d'autres, et peut également utiliser des données provenant de fichiers texte, XML et Excel.

Les transformations et les tâches sont créées par glisser-déposer des différentes étapes du processus. Des assistants sont disponibles à chaque étape et permettent par exemple de générer automatiquement les requêtes de lecture et de création de tables.

## Les avantages

Dans le contexte du décisionnel, PDI va servir à constituer un datawarehouse, en réalisant l'import des données depuis différentes sources, leur consolidation et leur mise en forme. L'utilisation de PDI présente plusieurs avantages par rapport à des solutions à base de scripts :

- ▶ la création de flux de données complexes est extrêmement simple,
- ▶ la maintenance est beaucoup plus aisée car tous les flux de données sont visibles d'un simple coup d'oeil dans une interface unique et centralisée. PDI pourra également trouver sa place dans tout projet non décisionnel qui a besoin d'exécuter périodiquement des traitements sur certaines données ou des synchronisations entre différentes bases.

## Conclusion

L'informatique décisionnelle tient ses engagements : améliorer la performance de tous les services de votre entreprise en vous aidant à prendre de meilleures décisions. Dès l'instant où vous savez que vos données sont mises à jour, complètes et intègres, vous pouvez avoir confiance en chacune des décisions que vous prenez. Avec un tel niveau de maîtrise, vous pouvez améliorer la performance de votre entreprise, créer un avantage compétitif et atteindre vos objectifs stratégiques.

## Chapitre 2

# Les Entrepôts de Données ( Data Warehouses )

Toutes les entreprises disposent de données. Elles proviennent soit de leurs systèmes opérationnels, soit de l'extérieur. Le problème de ces entreprises est d'atteindre les objectifs définis par les enjeux de leur secteur en tirant parti des données accessibles.

L'entreprise actuelle croule sous les données. Cette surabondance a comme conséquence directe un rejet par saturation. Pourtant, les données représentent une mine d'informations. Elles sont un avantage dont l'entreprise doit tirer parti. Pour cela, il est fondamental de mettre en place une nouvelle informatique décisionnelle pour obtenir une meilleure compréhension de la valeur des informations disponibles, de définir des indicateurs "business" pertinents pour faciliter les prises de décisions opérationnelles et garder la mémoire de l'entreprise.

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. Cette architecture globale est le Data Warehouse.

Le concept d'entrepôt de données ou « Data Warehouse » a vu le jour en 1990. L'idée principale était de créer une base de données orientées sujet, intégrées, contenant des informations historisées, non volatiles et exclusivement destinée aux processus d'aide à la décision.

Le Data Warehouse augmente la productivité des décideurs d'entreprise en consolidant, convertissant, transformant et intégrant les données de différents systèmes opérationnels.

En peu de mots, l'objectif principal du Data Warehouse est d'intégrer les données des systèmes opérationnels dans un environnement qui permet une utilisation stratégique de ces données.[Web5]

### 2.1 Quelques définitions

Ralph Kimball publie en 1996 son ouvrage « Entrepôts de données : Guide pratique du concepteur de "Date Warehouse" », dans lequel il définit l'entrepôt de données comme étant « le conglomérat

des données d'une organisation » [kimball, 1997].

Un entrepôt de données est une aire de rassemblement et de présentation où les données opérationnelles sont spécifiquement structurées pour garantir les performances des requêtes et faciliter l'analyse et l'utilisation.

«Le Data Warehouse est un système qui extrait, nettoie, traite et rend conforme des données sources vers un espace de stockage multidimensionnel. Il permet ainsi la mise en œuvre de l'interrogation et de l'analyse à des fins d'aide à la prise de décision » [Kimball & Margy,2002]

Ralph Kimball définit également les concepts de Data Warehouse d'entreprise et celui de Data Mart ; le Data Warehouse d'entreprise étant une agglomération des Data Mart.

« Le Data Mart est un sous-ensemble, logique et physique de la présentation de l'entrepôt de données, organisé par processus métier ; c'est un ensemble flexible de données. Les Data Mart peuvent être attachés ensemble en utilisant des techniques de forage à travers leurs dimensions communes ». [Kimball, 2005]

Cette définition de l'entreposage de données n'est nullement unique, elle fut précédée par celle proposée en 1990 par William H Inmon dans son ouvrage « Building the Data Warehouse »

« Un entrepôt de données est une collection de données orientées sujet, intégrées, non volatiles, et historiées, servant de support à la prise de décision managériale. Les données de l'entrepôt de données sont aptes à être utilisées à de nombreuses fins, y compris des besoins futurs non encore connus». [H.Inmon,2005]

## 2.2 Caractéristiques essentiels du Data Warehouse

Cinq caractéristiques essentielles se dégagent de la définition donnée par Bill Inmon, le père du Data Warehouse :

### ► **Orientation sujet**

Le Data Warehouse est organisé autour des sujets majeurs de l'entreprise. Donc contrairement aux organisations traditionnelles où les données sont regroupées par processus fonctionnel, la structuration des données ne se fera pas par thèmes.

### ▷ **Données intégrées**

Les données qui seront intégrées dans le Data Warehouse proviendront de différentes sources, ces données doivent être mise en forme et unifiées afin d'avoir un état cohérent, et d'éviter toutes anomalies.

### ▷ **Données historiées**

Le Data Warehouse à l'inverse des systèmes opérationnels ne met pas à jour la donnée en supprimant l'ancienne valeur, mais sauvegarde les différentes valeurs que peut prendre cette donnée au fil du temps. Donc un référentiel temps est associé à la donnée, c'est ce qu'on appelle : l'historique.

### ▷ **Données non volatiles**

C'est une conséquence directe de l'historisation. Dans un Data Warehouse une donnée ne peut être supprimée ou mise à jour.

▷ **Données organisées pour le support d'un processus d'aide à la décision**

Ce qui signifie que les données du Data Warehouse sont organisées de manière à permettre l'exécution des processus d'aide à la décision (Reporting, Data Mining...).

- Afin de bien saisir les caractéristiques du Data Warehouse, une comparaison avec le système de production s'impose :

<b>Critère</b>	<b>Système de production</b>	<b>Data Warehouse</b>
<b>Niveau de détails des informations utilisateurs</b>	Très détaillé	Synthétique, parfois détaillé
<b>Utilisateurs</b>	Une ou quelques fonctions de l'entreprise	Plusieurs fonctions de l'entreprise
<b>Données figées</b>	Non-évolution en temps réel	Oui - Archiver
<b>Historique</b>	Non	Oui
<b>Operations sur les données</b>	Ajout, mise à jour, consultation	Consultation uniquement

FIGURE 2.2.1 – Différences entre système de production et Data Warehouse

## 2.3 Structure de données d'un Data Warehouse

Le Data Warehouse a une structure bien définie, selon différents niveaux d'agrégation et de détail des données. Cette structure est définie par Inmon [Inmon, 2005] comme suit :

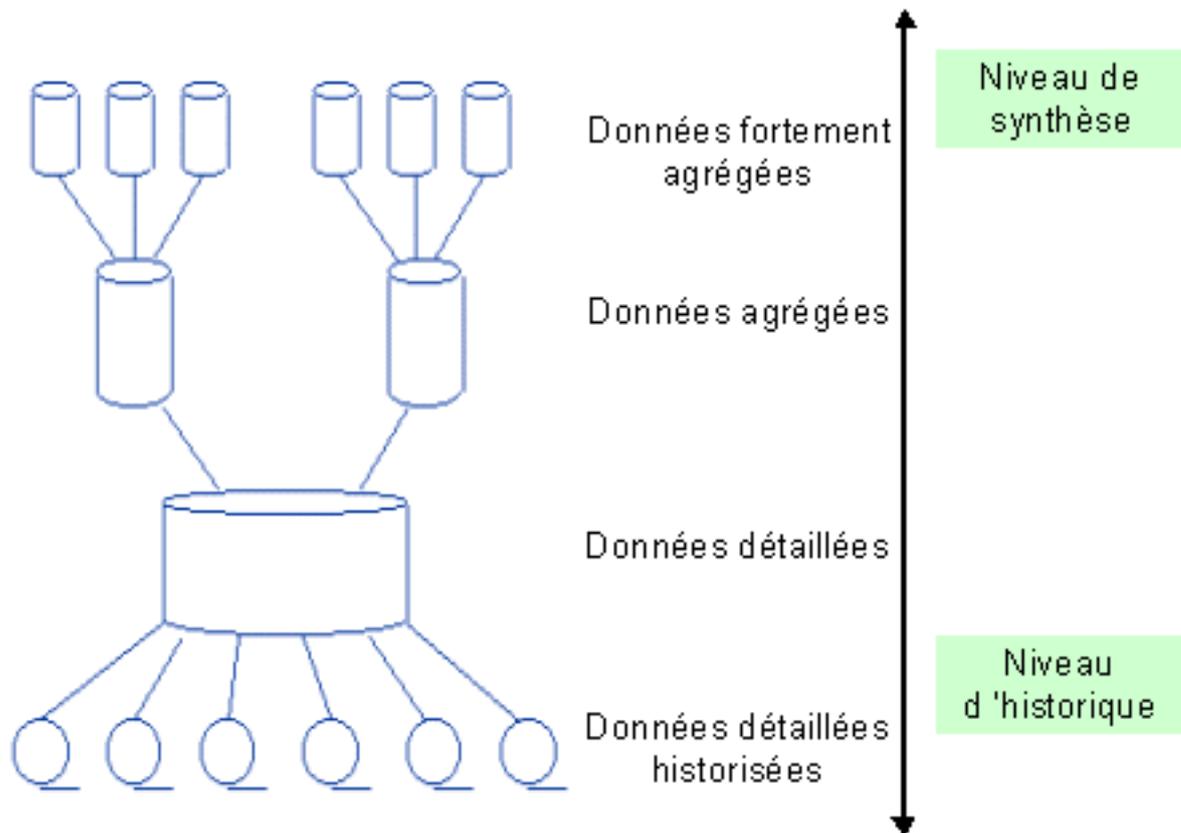


FIGURE 2.3.1 – Structure de données dans le Data Warehouse.

### A. Données détaillées

Les données détaillées reflètent les événements les plus récents. La gestion et le traitement de ces données nécessitent parfois des machines sophistiquées, car leur volume peut être très important.

Exemple : Les données provenant des systèmes de production : Police d'assurance ..

### B. Données détaillées archivées

Anciennes données rarement sollicitées, généralement stockées dans un disque de stockage de masse, peu coûteux, à un même niveau de détail que les données détaillées.

Exemple : Les dossiers d'anciens clients.

## C. Données agrégées

Données agrégées à partir des données détaillées. Ces données sont très souvent utilisées, elles constituent déjà un résultat d'analyse et de synthèse de l'information contenue dans le système décisionnel.

Exemple : Rapports d'agences ..

## D. Données fortement agrégées

Données agrégées à partir des données détaillées, à un niveau d'agrégation plus élevé que les données agrégées.

Exemple : Rapport annuel ..

## 2.4 Les éléments composant un Data Warehouse

Pour créer un Data Warehouse, il faut connaître ses composants et leurs rôles dans ce dernier. Ces composants sont au nombre de quatre : système source, zone préparation de données, zone présentation des données et outils d'accès aux données.

### 2.4.1 Système source

Evidemment il nous faut en premier lieu une ou plusieurs sources pour alimenter le Data Warehouse. Ces sources sont considérées comme externes au Data Warehouse.

L'objectif reste de réunir les données nécessaires à la prise de décisions sans affecter le fonctionnement des systèmes sources. Ces sources peuvent être :

- ▶ Sources hétérogènes  
Document non structuré comme les documents Word, fichiers plats et feuilles de calcul.
- ▶ Bases de données production  
Ou base de données transactionnelle.
- ▶ Fichiers externes :  
Qui sont acquis auprès d'autres entreprises.

### 2.4.2 La zone de préparation de données

C'est à ce niveau que se fait le traitement des données brutes chargées à partir du système source. Elle est constituée d'un ensemble de processus appelé ETL, « Extract, transform and Load ». Le traitement de la qualité des données implique plusieurs étapes distinctes : vérification des valeurs valides, cohérences entre les valeurs, vérification des doublons... L'objectif final de la zone de préparation des données est l'obtention de données prêtes à être chargées sur un serveur de prestation (moteur OLAP ou SGBDR).

### 2.4.3 La zone de présentation de données

La zone de présentation est tout ce que l'utilisateur peut voir et toucher par le biais des outils d'accès. C'est l'entrepôt où les données sont organisées et stockées. L'entrepôt de données est constitué d'un ensemble de Datamart. Le Datamart est une miniaturisation du Data Warehouse, construit autour d'un sujet d'analyse précis ou consacré à un niveau départemental.

#### *Datamart*

« Datamart est défini comme un sous-ensemble logique d'un entrepôt de données ». [Kimballe & Margy,2002]

Le Datamart est une base de données plus réduite que le Data Warehouse, puisque destinée à quelques utilisateurs d'un département. Il séduit plus que le Data Warehouse les candidats au décisionnel. C'est une petite structure très ciblée et pilotée par les besoins utilisateurs. Il a la même vocation que le Data Warehouse (fournir une architecture décisionnelle), mais ciblant un groupe de métier ou sujet spécifique.

	<b>Data Warehouse</b>	<b>Data Mart</b>
<b>Cible utilisateur</b>	Toute l'entreprise	Département
<b>Implication de service informatique</b>	Elevée	Faible ou moyenne
<b>Base de données d'entreprise</b>	SQL type serveur	SQL milieu de gamme, bases multidimensionnelles
<b>Modèles de données</b>	A l'échelle de l'entreprise	Département
<b>Sources de données</b>	Multiples	Quelques-unes
<b>Champ applicatif</b>	Multi sujets, neutre	Quelques sujets spécifiques
<b>Stockage</b>	Base de données	Plusieurs bases distribuées
<b>Taille</b>	De centaines de GO à dizaines de TO	Centaines de GO

FIGURE 2.4.1 – Finalités des Datamart et Data Warehouse

### 2.4.4 Outils d'accès aux données

Les outils cités en ci-dessous sont exploités par les utilisateurs pour accéder à la zone de présentation du Data Warehouse. Ces outils doivent être faciles d'utilisation pour les non informaticiens. Il est impératif que les données du Data Warehouse soient présentées à l'utilisateur de manière significative et propre à son activité, afin qu'elles puissent être interprétées facilement.

On trouve :

- ▶ Des outils d'interrogation ad hoc.
- ▶ Des outils de Data Mining.
- ▶ Des outils de reporting.

*Remarque.* Dans notre cas, nous nous sommes intéressés à l'outil de restitution : Reporting.

## 2.5 Le reporting :

Le reporting est un ensemble d'indicateurs de résultat, construit à posteriori, de façon périodique, afin d'informer la hiérarchie des performances de l'unité. Le reporting constitue un outil décisionnel aussi bien pour la Direction Générale que pour les Directions Opérationnelles. Il centralise les informations clés susceptibles d'aider à faire le point sur une partie de l'activité de l'entreprise (marketing, commerciale, financière, ..).[web6]

Qu'il soit destiné aux managers, marqueteurs ou financiers le reporting est généralement l'activité qui consiste à rendre compte périodiquement de ses performances ou actions à l'égard de sa direction.

Année	C.A.	CA Ventes	CA Production	CA Services	Résultat Net
2001	259 715 000	0	0	0	259 715 000
2003	331 150 000	88 252 000	239 012 000	3 886 000	331 150 000
2004	352 731 000	75 062 000	274 705 000	2 963 000	352 731 000
2005	388 202 000	68 617 000	316 584 000	3 000 000	388 202 000
2006	388 202 000	68 617 000	316 584 000	3 000 000	388 202 000

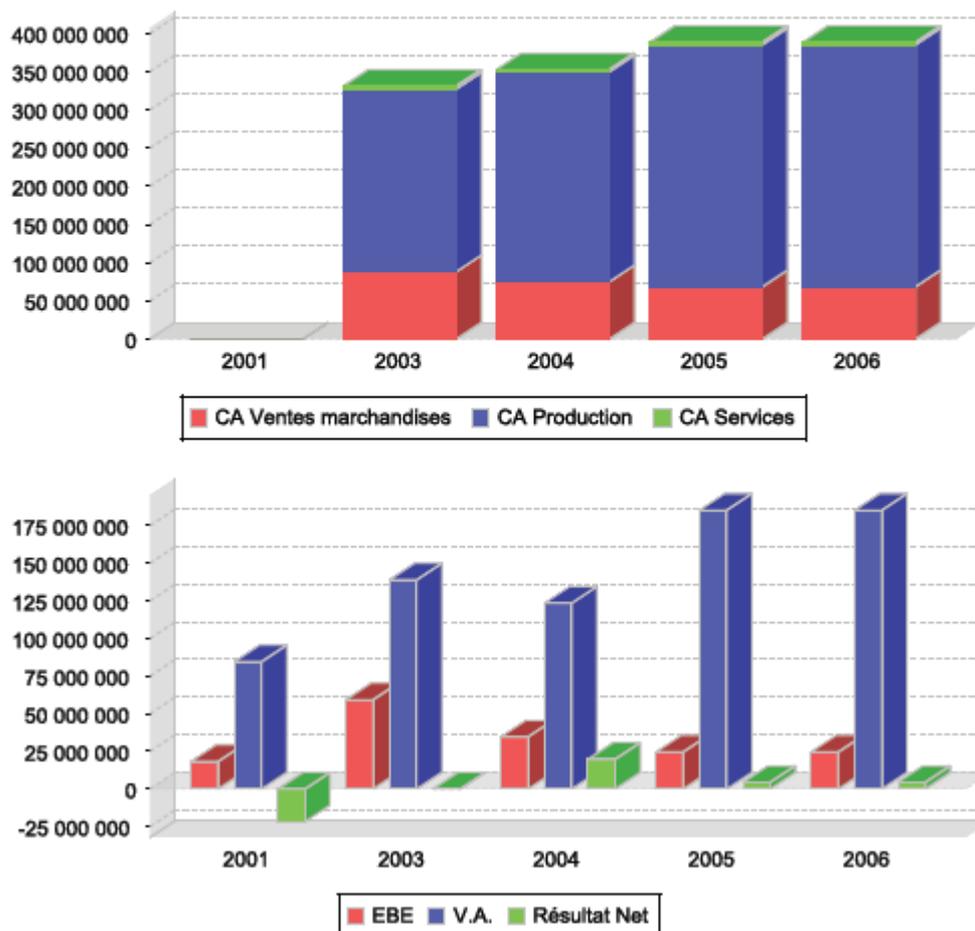


FIGURE 2.5.1 – Exemple d'un état de sortie en reporting.

### 2.5.1 Objectifs du reporting

- ▶ Le reporting achève l'élaboration d'un système décisionnel. Il permet de :
- ▶ Fournir une vision opérationnelle de l'activité.
- ▶ Assurer le suivi de ses campagnes, et valider les business plans.
- ▶ Fournir les bilans à communiquer aux actionnaires et au public.
- ▶ Dans le cas d'une gestion décentralisée, il permet de vérifier que les centres de responsabilité respectent leurs engagements contractuels.
- ▶ Mettre en évidence de nouvelles opportunités ou des leviers d'amélioration de l'existant.
- ▶ Pour les managers, c'est un outil de management opérationnel (il s'agit plutôt de rapports numériques).
- ▶ Pour les équipes de production, c'est une aide quotidienne pour leurs activités.
- ▶ Pour les commerciaux c'est un outil de suivi de leur activité.

### 2.5.2 Les types de reporting

On distingue deux types de reporting :

▶ **Le reporting de masse :**

Ce sont des solutions de publication de documents aux formats élaborés pour un grand nombre d'utilisateurs. La mise en forme des documents diffusés est soigneusement travaillée et les modalités de publication sont variées (web, mail, papier, ...). L'utilisateur (ou destinataire) est un consommateur.

▶ **Le reporting BI :**

Les rapports sont généralement générés à la demande à des fins d'analyse. La donnée y est plus importante que la forme. Le nombre d'utilisateurs est restreint ce sont généralement des auteurs.

## Conclusion

Tout système d'information décisionnel s'appuie sur un Data Warehouse. Le Data Warehouse est une base de données regroupant l'ensemble des données fonctionnelles d'une entreprise, son but est de fournir l'information nécessaire à la prise de décision. Cependant la collecte de données n'est que la première étape dans le processus de la mise en place d'un système décisionnel, comment transformer et représenter ces données afin qu'elles soient exploitables et visibles pour le décideur ?

Dans le chapitre suivant, nous tentons de répondre à cette question en introduisant le concept de la modélisation dimensionnelle.

## Chapitre 3

# Modélisation multidimensionnelle

La modélisation multidimensionnelle a été introduite par Ralph Kimball afin de répondre au besoin décisionnel et aux exigences des SGBD relationnels. Le modèle dimensionnel est la structure de données la plus utilisée et la plus appropriée aux requêtes et analyses des utilisateurs.

### 3.1 Modélisation dimensionnelle et ses concepts

Il ne faut pas perdre de vue que le Data Warehouse n'est qu'une étape dans la mise en place d'un système décisionnel. D'où le besoin d'une structure simple qui permet de représenter les données de manière visuelle pour les utilisateurs finaux. Cette structure est le modèle dimensionnel. Il consiste à représenter les données en tant que points dans un espace multidimensionnel (cube). Les données sont vues comme des sujets d'analyse (faits) étudiés selon plusieurs axes (dimensions).

« La modélisation dimensionnelle permet non seulement de présenter les données sous une forme standardisée intuitive mais permet aussi des accès hautement performants. » [Inmon,2008]

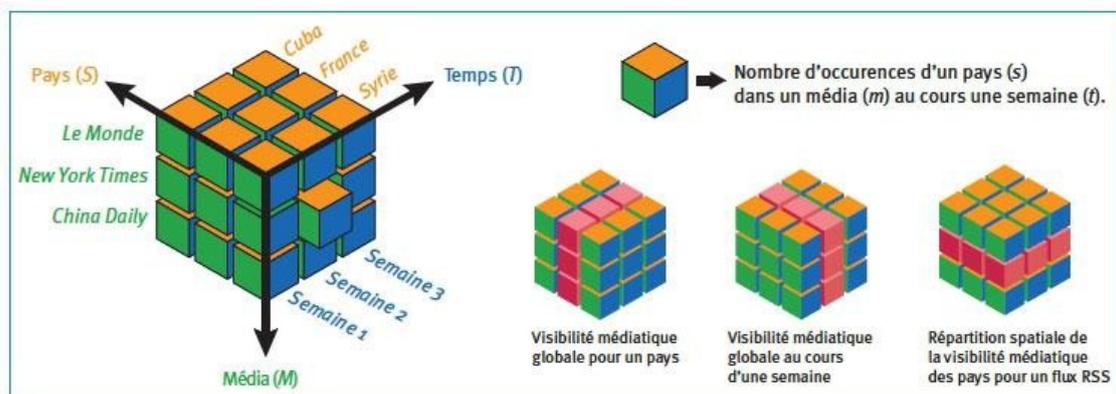


FIGURE 3.1.1 – Considération d'un sujet d'analyse comme un cube à plusieurs dimensions. [Franco, 1997]

### 3.1.1 La table de faits

Le terme « fait » représente une mesure économique sur le sujet analysé. La table de fait sert à stocker les mesures de l'activité. Chacune de ces mesures est prise à l'intersection de toutes les dimensions. Une ligne d'une table fait correspond à une mesure. Ces mesures sont généralement des valeurs numériques, additives. Une table de faits assure les liens plusieurs à plusieurs entre les dimensions. Elles comportent des clés étrangères, qui ne sont autres que les clés primaires des tables de dimension.

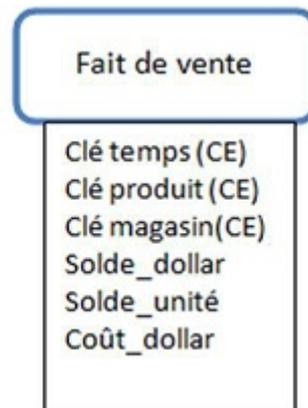


FIGURE 3.1.2 – Exemple d'une table de fait. [Franco, 1997]

### 3.1.2 La table de dimension

Les tables de dimensions accompagnent une table fait, servent à enregistrer les descriptions textuelles des dimensions de l'activité. Elle est composée de plusieurs colonnes qui décrivent une ligne. C'est la table de faits qui rend le Data Warehouse compréhensible et utilisable. « Une table de dimension établit l'interface homme / entrepôt, elle comporte une clé primaire », [Kimball, 2002].



FIGURE 3.1.3 – Exemple d'une table de dimension. [Franco, 1997]

## 3.2 Les différents modèles de la modélisation dimensionnelle

### 3.2.1 Modèle en étoile

Appelé modèle en étoile car le diagramme qui le représente ressemble à une étoile. Il y a une grande table (table fait), dominant les autres, au centre du schéma. C'est la seule table du schéma comportant de multiples jointures la reliant aux autres tables (tables de dimensions). Les autres tables n'ont qu'une seule jointure les reliant à la table de faits. C'est le modèle le plus utilisé pour sa lisibilité et performance.

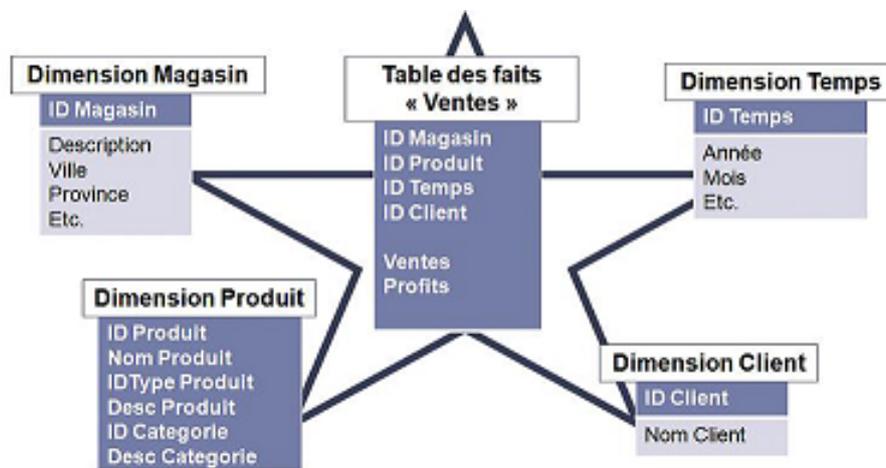


FIGURE 3.2.1 – Exemple d'un schéma étoile.

### 3.2.2 Modèle en flocon

Ce schéma n'est qu'une variante du modèle en étoile. La différence est que chaque attribut de niveau hiérarchique est mis dans une table dimension à part. Prenons comme exemple la dimension temps :

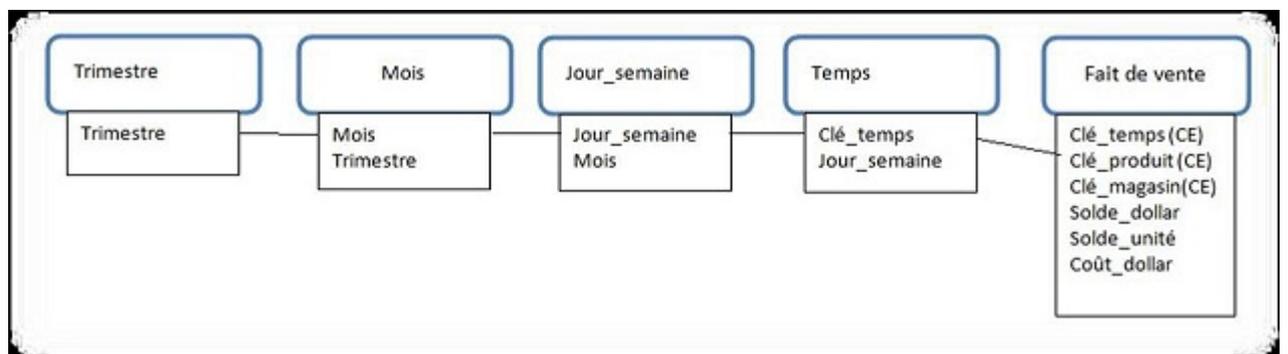


FIGURE 3.2.2 – Exemple d'un schéma flocon. [Franco, 1997]

### 3.2.3 Modèle en constellation

Le modèle en constellation est un ensemble de modèles en étoiles reliés entre eux par des dimensions communes.

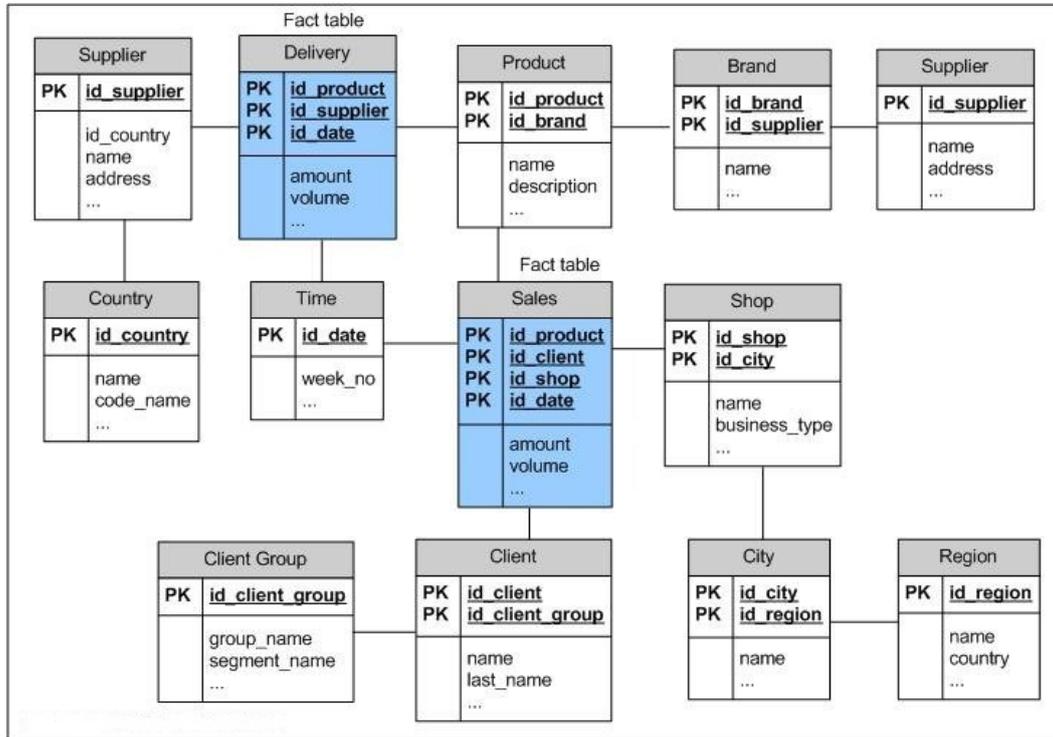


FIGURE 3.2.3 – Exemple d’un schéma en constellation

### 3.3 Le cube de données

La base de données multidimensionnelles s’appuie sur un hyper cube (cube à n dimensions). L’administrateur définit les dimensions, qui représentent une façon de trier l’information. Cette modélisation permet à l’utilisateur sans l’aide de l’informaticien d’avoir l’information dont il a besoin. L’utilisateur choisit deux ou trois critères à visualiser sous forme de tableau ou de cube. Il peut également faire pivoter les axes d’analyse pour projeter les informations sous un angle différent. Ainsi, après avoir examiné les ventes par région, il peut permuter les axes pour une visualisation par mois. Ce critère est certainement le critère-clé du concept OLAP car il reflète la dimensionnalité de l’entreprise telle que la perçoivent ses membres qui ne sont autres que les utilisateurs du système.

Exemple : Sur le schéma ci-dessous on distingue les dimensions : produit, temps et région. Les faits sont la vente. Pour chaque combinaison des dimensions on peut accéder à la valeur numérique associée au fait vente.

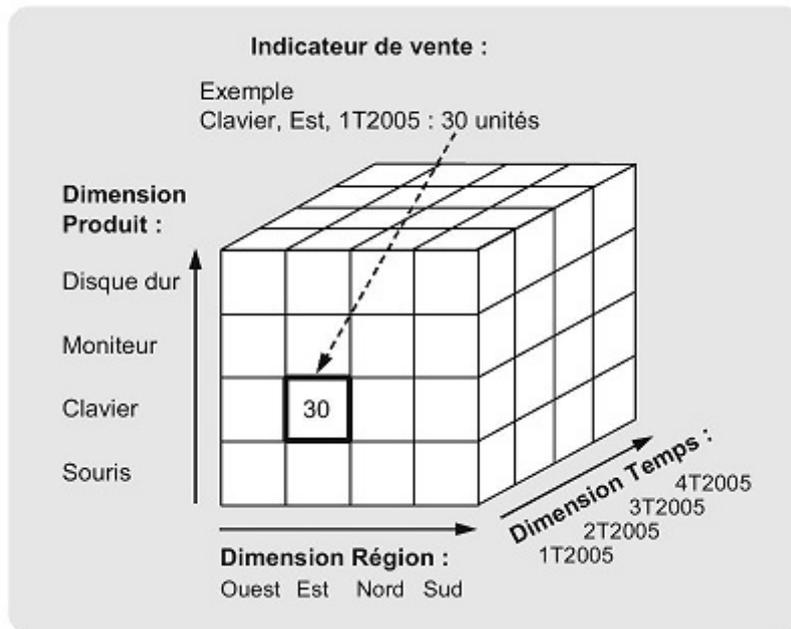


FIGURE 3.3.1 – Exemple d’un hyper-cube. [Web5]

## 3.4 Le concept OLAP

La différence d’objectifs entre les SGBD et les Data Warehouses a suscité l’utilisation de deux systèmes différents :

- ▶ OLTP (On Line Transaction Processing) pour les SGBD.
- ▶ OLAP (On Line Analytical Processing) pour les Data Warehouses.

### 3.4.1 Qu’es ce que l’OLAP ?

Le concept OLAP fut introduit et défini en 1993 par E.F Codd, le père des bases de données relationnelles. Dans ce document Codd a défini les 12 règles du traitement analytique en ligne OLAP. [Codd, 1993]

R. Kimball définit le concept « OLAP » comme « *Activité globale de requêtage et de présentation de données textuelles et numériques contenues dans l’entrepôt de données ; Style d’interrogation spécifiquement dimensionnel* » [Kimball,2005].

Donc les Data Warehouses reposent sur le système OLAP, il est caractérisé par un volume très important de données. Il fonctionne seulement en lecture et a pour objectifs principaux de rassembler, gérer, traiter et de présenter des données multidimensionnelles et venant de sources diverses à des fins d’analyse et de décision, offrant ainsi à l’utilisateur une vue orientée métier.

### 3.4.2 Et l'OLTP ?

Est un type d'application informatique qui sert à effectuer des modifications d'informations en temps réel. Ce type d'application est utilisé dans des activités opérationnelles, typiquement des transactions commerciales (opérations bancaires, achats de biens, billets, réservations).

Ce type d'application se connecte à des bases de données en lecture et écriture.

Ce type d'application est typiquement opposé au traitement OLAP (pour online analytical processing ou traitement analytique en ligne) qui fonctionne en principe en lecture.

### 3.4.3 OLAP Vs OLTP

Avant de pouvoir comparer OLAP -qu'on a défini auparavant- et OLTP, nous nous devons de parler ne serait-ce que brièvement de ce dernier. OLTP est le système utilisé par les SGBD. Il a un mode de travail transactionnel. OLTP a pour objectif de pouvoir insérer, modifier, intégrer rapidement et par plusieurs utilisateurs simultanément, les données de la base de données. Les versions de données traitées sont toujours les plus récentes.

Caractéristiques	OLTP	OLAP
Utilisation	SGBD (production)	Data Warehouse
Operation typique	Mise à jour	Analyse
Type d'accès	Lecture/Ecriture	Lecture
Niveau d'analyse	Elémentaire	Globale
Quantité d'information échangées	Faible	Importante
Orientation	Ligne	Multidimensionnelle
Taille DB	Faible (max quelque GB)	Importante (>TB)
Ancienneté des données	Récente	Historique

FIGURE 3.4.1 – Comparaison OLTP Vs OLAP. [Web6]

### 3.4.4 Architecture des serveurs OLAP

Lors de la création d'un cube OLAP, il faut décider de la manière de stocker les données automatiques ainsi que les agrégats (regroupements). Il y a plusieurs versions d'OLAP, mais les trois les plus répandues sont :

- ▶ L'architecture MOLAP « Multidimensional OLAP ».
- ▶ ROLAP « Relational OLAP ».
- ▶ HOLAP « Hybrid OLAP ».

#### 3.4.4.1 Architecture MOLAP « Multidimensional OLAP »

MOLAP est la forme la plus classique d'OLAP. Elle s'appuie sur une base de données multidimensionnelle. Le stockage MOLAP est le plus approprié pour les cubes qui sont utilisés fréquemment

et nécessitent un temps de réponse minimal. Son inconvénient est qu'il ne supporte pas de très grands volumes de données.

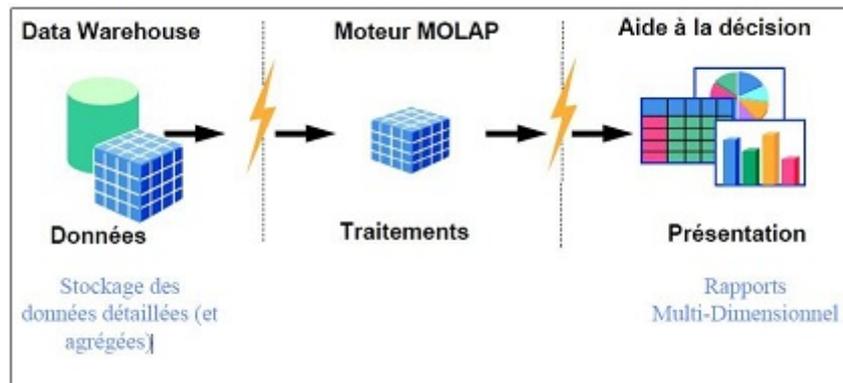


FIGURE 3.4.2 – Schéma MOLAP. [Nakache 1998]

#### 3.4.4.2 Architecture ROLAP « Relational OLAP »

« Ensemble d'interfaces utilisateurs et d'applications qui donnent une vision dimensionnelle à des bases de données relationnelles » [Kimball, 2005]. Les données du cube et les agrégations sont stockées dans l'entrepôt existant. (Bases de données relationnelles). Les systèmes ROLAP sont en mesure de simuler le comportement d'un SGBD multidimensionnel en exploitant un SGBD relationnel. L'utilisateur aura ainsi l'impression d'interroger un cube multidimensionnel alors qu'en réalité il ne fait qu'adresser des requêtes sur une base de données relationnelles. En général il est utilisé pour les grands jeux de données qui ne sont pas sollicités fréquemment. Il permet aussi de stocker de gros volumes de données.

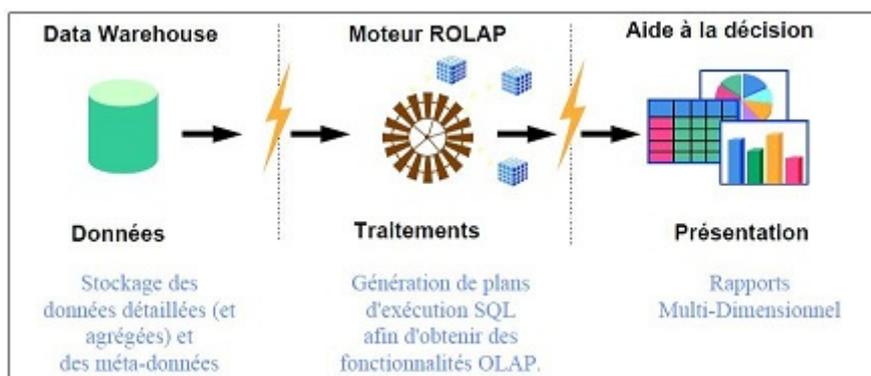


FIGURE 3.4.3 – Schéma ROLAP. [Nakache 1998]

#### 3.4.4.3 Architecture HOLAP « Hybrid OLAP »

Cette architecture est une sorte de compromis entre les deux architectures précédentes, en effet les données du cube sont stockées dans l'entrepôt relationnel existant, et les agrégations dans une structure multidimensionnelle. Cette combinaison donne à ce type de système les avantages du ROLAP et du MOLAP en utilisant tour à tour l'un ou l'autre selon le type de données.

#### 3.4.4.4 Autres architectures OLAP

Il existent d'autres architectures OLAP moins utilisées comme l'architecture OOLAP « Object OLAP », ou alors DOLAP « Desktop OLAP » qui décrivent une catégorie de produits qui ne sont pas nécessairement connectés à un serveur et qui s'appuient sur une source de données (un cube) construites, stockées et exploitées en local sur la machine de l'utilisateur.

#### 3.4.4.5 Comparaison MOLAP Vs ROLAP Vs HOLAP

	<b>MOLAP</b>	<b>ROLAP</b>	<b>HOLAP</b>
<b>Stockage de données de base (détaillées)</b>	BD multidimensionnelle	BD relationnelle	BD relationnelle
<b>Stockage des agrégations</b>	BD multidimensionnelle	BD relationnelle	BD multidimensionnelle
<b>Performance requêtes</b>	Elevée	Peu performante	Moyenne

FIGURE 3.4.4 – Comparaison entre les différentes architectures d'OLAP. [Grislin, 2000]

### 3.5 La navigation dans les données

Une fois que le serveur OLAP a construit ou simulé le cube multidimensionnel, plusieurs opérations sont possibles sur ce dernier, offrant ainsi la possibilité de naviguer dans les données qui le constituent « Data Surfing ». Ces opérations adressent l'ensemble des données et permettent à l'utilisateur de naviguer de manière libre et intuitive dans le modèle dimensionnel.

**Les opérations sont :**

- ▶ Opérations ensemblistes : Slicing & Dicing (extraction).
- ▶ Opérations de granularité : Roll-up, Drill-down.

#### 3.5.1 Les opérations ensemblistes « Slice & Dice »

- ▶ Le Slicing consiste à faire une sélection de tranches du cube selon des prédicats et selon une dimension « *filtrer une dimension selon une valeur* ». [Chouder, 2007]

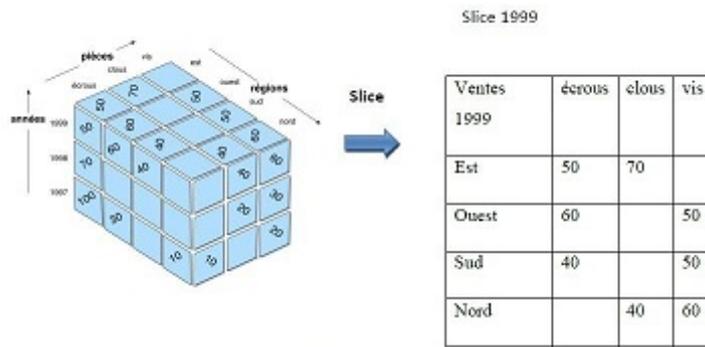


FIGURE 3.5.1 – Exemple de Slicing.

► Le Dicing, quant à lui peut être vu comme l'extraction ou sélection d'un sous cube.

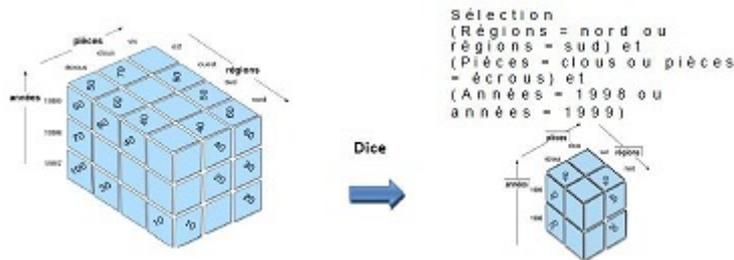


FIGURE 3.5.2 – Exemple de Dicing.

### 3.5.2 Les opérations de granularité Roll-up & Drill-down

► Le Roll-up consiste à représenter les données du cube à un niveau de granularité supérieur conformément à la hiérarchie définie sur la dimension.

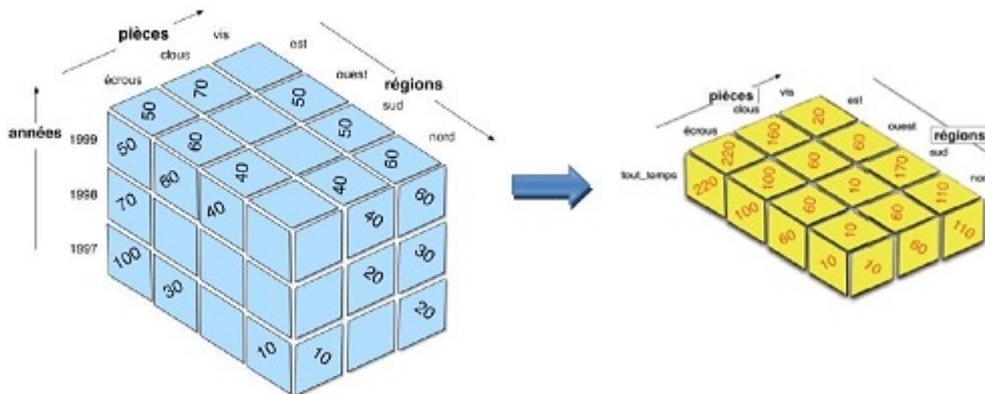


FIGURE 3.5.3 – Exemple Roll-up.

- Le Drill-down consiste à représenter les données du cube à un niveau de granularité de niveau inférieur, donc sous une forme plus détaillée (selon la hiérarchie définie de la dimension).

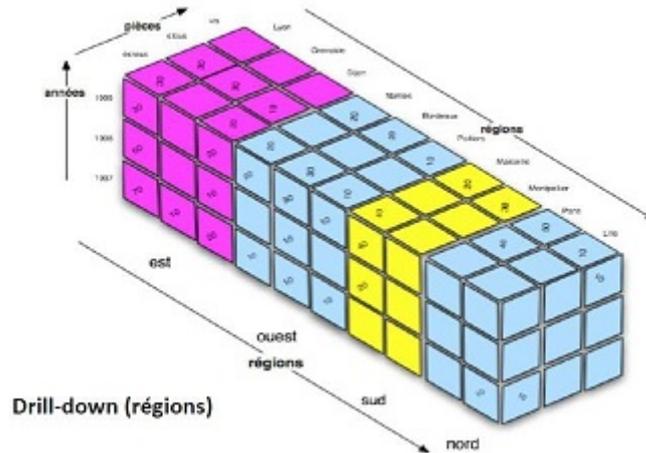


FIGURE 3.5.4 – Exemple Drill-down du niveau des régions au niveau des villes

## 3.6 Conduite et construction du Data Warehouse

Pour la construction du Data Warehouse on suivra les démarches suivantes :

- Méthodologies conductrices à la mise en place d'un Data Warehouse.
- L'alimentation du data warehouse.
- La mise en œuvre du data warehouse.
- L'administration et la maintenance du data warehouse.

### 3.6.1 L'approche conductrice à la mise en place d'un Data Warehouse

[Kimball ,1996]

#### A. Top-Down

C'est la méthode la plus lourde, la plus contraignante et la plus complète en même temps. Elle consiste à la conception de tout l'entrepôt (i.e. : toutes les étoiles), puis à la réalisation de ce dernier.

Imaginez le travail qu'une telle méthode implique : savoir à l'avance toutes les dimensions et tous les faits de l'entreprise, puis réaliser tout ça... Le seul avantage que cette méthode comporte est qu'elle offre une vision très claire et très conceptuelle des données de l'entreprise ainsi que du travail à faire.

## B. Bottom-Up

C'est l'approche inverse, elle consiste à créer les étoiles une par une, puis à les regrouper par des niveaux intermédiaires jusqu'à obtention d'un véritable entrepôt pyramidal avec une vision d'entreprise.

L'avantage de cette méthode est qu'elle est simple à réaliser (une étoile à la fois), l'inconvénient est le volume de travail d'intégration pour obtenir un entrepôt de données ainsi que la possibilité de redondances entre les étoiles (car elles sont faites indépendamment les unes des autres).

## C. Middle-Out

C'est l'approche hybride, et conseillée par les professionnels du BI. Elle consiste en la conception totale de l'entrepôt de données (i.e. : concevoir toutes les dimensions, tous les faits, toutes les relations), puis créer des divisions plus petites et plus gérables et les mettre en œuvre. Cela équivaut à découper notre conception par éléments en commun et réaliser les découpages un par un. Cette méthode tire le meilleur des deux précédentes sans avoir les contraintes. Il faut juste noter que cette méthode implique, parfois, des compromis de découpage (dupliquer des dimensions identiques pour des besoins pratiques).

### 3.6.2 Alimentation du Data Warehouse

Pour alimenter le data warehouse de façon régulière, les données d'un ou de plusieurs systèmes opérationnels doivent être extraites et copiées dans le data warehouse, d'où la nécessité d'établir des flux de données entre le data warehouse et les bases de données spécialisées où elles ont été créées (les bases de production). Pour ce faire, on met en place des interfaces d'extraction, de transformation et d'alimentation. C'est ce qu'on appelle un outil Extract, Transform, Load.

#### Le processus **Extract-Transform-Load** < **ETL** > [Kimball ,1996]

L'outil d'alimentation permet de paramétrer des règles de gestion, propres à l'entreprise et à son secteur d'activité. Ces règles visent elles aussi à assurer la cohérence entre les données et à ne stocker dans l'entrepôt de données que des informations préalablement mises en relation les unes avec les autres.

**Définition** « Extract-Transform-Load » est connu sous l'acronyme ETL (ou parfois : Data-pumping). Il s'agit d'une technologie informatique middleware permettant d'effectuer des synchronisations massives d'information d'une banque de données vers une autre. Selon le contexte, on traduira par « alimentation », « extraction », « transformation », « constitution » ou « conversion », souvent combinées.

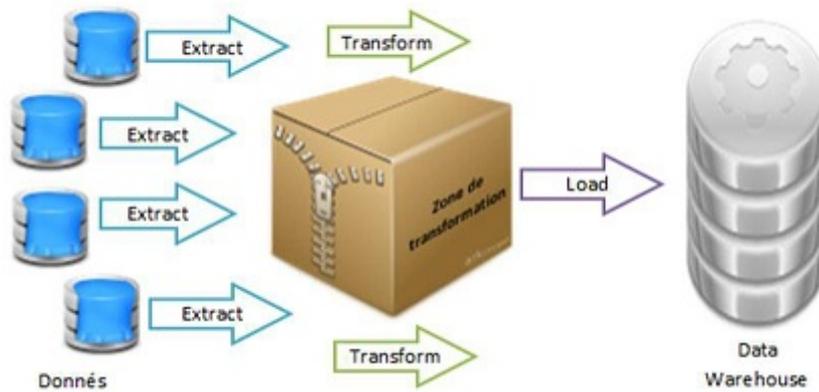


FIGURE 3.6.1 – Le système d'alimentation ETL.

## Les phases constructives d'un ETL

### I. Extract

L'extraction des données est la première des étapes des systèmes ETL. Le but de cette étape, est comme son nom l'indique : la lecture et l'extraction des données du système source. On imagine facilement que cette étape s'avère être critique. En effet, dans le cas où le système source doit fonctionner en permanence (24h/24 et 7jours sur 7), il faut que l'extraction, opération coûteuse en ressources, doit être faite le plus rapidement possible et souvent durant un laps de temps précis (souvent nommé « extract window »), décidé en collaboration des équipes chargés de la gestion et/ou de l'exploitation de ce système source. La complexité de l'extraction n'est pas dans le processus de lecture, mais surtout dans le respect de l'« extract window ». Cette contrainte est la principale raison de la séparation extraction / transformation. D'autre part, on essaye au maximum d'extraire seulement les données utiles (Mise à jour ou ajouter après la dernière extraction) et pour se faire on pourrait s'entendre avec le responsable du système source pour ajouter soit un flag ou encore des dates dans chacune des tables extraites, au moins deux dates : Date de création de l'enregistrement dans la table et la date de mise à jour (en général la plupart des systèmes sources disposent de ces deux dates). Par ailleurs pour ne pas perdre les données suite aux problèmes d'extraction, il est important de s'assurer que le système source ne purge pas les données avant que l'entrepôt ne les ait extraites.

### II. Transform

La transformation est la tâche la plus complexe et qui demande beaucoup de réflexion pour garantir la fiabilité des données et leurs qualités. Voici les grandes fonctionnalités de transformation :

- ▶ Nettoyage des données.
- ▶ Standardisation des données.
- ▶ Conformité des données.

- ▶ Gestion des tables de faits.
- ▶ Gestion des dimensions.
- ▶ Affectation des clés de substitution (surrogate key).
- ▶ Gestion de l'évolution lente (Slowly changing dimension).
- ▶ Gestion des faits arrivants en retard (Late arriving fact).
- ▶ Gestion des lookups.

### III. Load

Le chargement permet de transférer les données vers leur destination finale. 3 cas de figures se présentent, selon l'architecture mise en place.

- ▶ Chargement des données dans l'entrepôt de données :

Si la politique retenue a été de construire un entrepôt de données avec une base de données, alors les données seront chargées dans l'entrepôt. Cette approche est proche à celle de Bill Inmon. Il sera dès lors possible d'utiliser des fonctionnalités analytiques comme Oracle le permet.

- ▶ Chargement des données dans les cubes de données :

Les données seront directement chargées dans des cubes de données sans les stocker dans un Data Warehouse.

- ▶ Le mode hybride :

Ce cas de figure est celui qui offre le plus d'avantages mais demande par contre plus d'effort. Le chargement des données s'effectue à la fois sur le Data Warehouse et les Datamarts :

- ▶ Un premier chargement des données dans un entrepôt de données.
- ▶ Un deuxième chargement dans des cubes de données.

### 3.6.3 Le déploiement et mise en oeuvre

#### A. Reporting

Le terme de « Reporting » désigne une famille d'outils de Business intelligence destinés à assurer la réalisation, la publication et la diffusion de rapports selon un format prédéterminé. Ils sont essentiellement destinés à faciliter la communication de résultats chiffrés ou d'un suivi d'avancement.

#### B. Requête ad hoc

Aujourd'hui largement utilisé dans les entrepôts de données actifs, les requêtes ad hoc permettent d'accélérer la récupération des informations vitales pour répondre aux requêtes interactives dans une application à mission critique. L'accès à ce genre de service peut se faire via différentes méthodes et outils. Cependant, les spécialistes préconisent de laisser la possibilité à l'utilisateur de choisir les outils qui lui paraissent les plus adéquats.

## C. Analyse dimensionnelle des données

L'analyse dimensionnelle est sans doute celle qui exploite et fait ressortir au mieux les capacités de l'entrepôt de données. Le but par l'analyse dimensionnelle est d'offrir aux utilisateurs la possibilité d'analyser les données selon différents critères afin de confirmer une tendance ou suivre les performances de l'entreprise. Cette analyse se fait selon le principe OLAP, offrant de ce fait aux utilisateurs les possibilités de recourir à différentes opérations facilitant la navigation dans les données. La mise en place de ces outils est une option très intéressante dans la mesure où les données seront accessibles en analyses instantanées. Plusieurs fournisseurs de solution OLAP existent sur le marché et offrent des solutions construites sur des méthodes et technologies différentes. C'est d'ailleurs pour cela que le choix de la solution doit se faire au préalable, selon les besoins en utilisation, la taille de l'entrepôt et les moyens techniques disponibles.

## D. Tableau de bord

Le tableau de bord est un ensemble d'indicateurs peu nombreux, conçus pour permettre aux gestionnaires de prendre connaissance de l'état et de l'évolution des systèmes qu'ils pilotent et d'identifier les tendances qui les influenceront sur un horizon cohérent avec la nature de leurs fonctions. [Bouquin, 2003]

C'est un outil de pilotage qui souligne l'état d'avancement dans lequel se trouve le processus afin de permettre au responsable de mettre en place des actions correctives.

## E. Data mining

Ensemble de techniques et de méthodes du domaine des statistiques, des mathématiques et de l'informatique permettant l'extraction, à partir d'un important volume de données brutes, de connaissances originales auparavant inconnues par des méthodes automatiques ou semi-automatiques.

L'utilisation industrielle ou opérationnelle de ce savoir dans le monde professionnel permet de résoudre des problématiques très diverses, allant de la gestion de relation client à la maintenance préventive, en passant par la détection de fraudes ou encore l'optimisation de sites web.

### 3.6.4 Maintenance et croissance

L'évolution du schéma du Data Warehouse nécessite un suivi constant compte tenu des besoins d'optimisation de performance ou d'expansion. Il est donc nécessaire d'investir dans les domaines suivants :

#### **a. Support aux utilisateurs**

Il apporte une assistance de premier niveau aux utilisateurs dans l'utilisation du Data Warehouse pour la résolution et la correction des problèmes rencontrés.

#### **b. Formation**

Il est indispensable d'offrir un programme de formation aux utilisateurs du Data Warehouse.

#### **c. Support technique**

Le support technique surveille avec la plus grande vigilance les performances et les tendances concernant la charge du système.

#### **d. Management de l'évolution**

Il faut toujours s'assurer que l'implémentation répond aux besoins de l'entreprise. Les revues systématiques à certains points de contrôle sont un outil clé pour détecter et définir les possibilités d'amélioration. En plus du suivi et de la maintenance du Data Warehouse, des demandes d'expansion sont envisageables pour de nouveaux besoins, de nouvelles données ou pour des améliorations. [Kimball, 2002]

### **Conclusion**

Le modèle dimensionnel est la structure de données la plus utilisée et la plus appropriée aux requêtes et analyses des utilisateurs d'entrepôts de données. Elles sont simples à créer, stables et intuitivement compréhensibles par les utilisateurs finaux. Le modèle dimensionnel est la fondation même pour la construction des cubes OLAP.

## Conclusion de la partie synthèse bibliographique

Dans cette première partie, nous avons introduit les concepts de base des systèmes décisionnels et des Data Warehouse nécessaires pour le reste du mémoire.

Nous avons commencé par introduire brièvement les systèmes décisionnels et leur rôle crucial au sein d'une entreprise, et nous avons continué en décrivant les Data Warehouse, en soulignant leur différence par rapport aux systèmes de bases de données traditionnelles. Nous avons ensuite abordé la modélisation dimensionnelle, en recouvrant les différents modèles dimensionnels existants, qui sont avec le cube de données, la base même du Data Warehouse, à savoir le modèle étoile, flocon, et constellation.

Par la suite nous avons discuté le concept OLAP sur lequel repose tout le Data Warehouse, ainsi que ses différentes architectures (ROLAP, MOLAP, HOLAP...). La mise en œuvre du Data Warehouse a également été détaillée dans le dernier chapitre ainsi que sa maintenance et sa croissance.

Nous appliquerons les concepts vus précédemment dans la phase conception et déploiement de notre solution au cours de la deuxième partie de notre mémoire.

Deuxième partie

**Analyse et conception**

## Chapitre 4

# Analyse et diagnostic de la situation existante

Nous avons décomposé ce chapitre sur deux grandes parties. La première traite des assurances en générale, et la deuxième, restriction sur notre organisme d'accueil (SAA).

### 4.1 Les Assurances

Les assurances ont pris une place importante dans la vie économique contemporaine. Leur liaison est désormais bien établie avec l'ensemble des activités qui s'appuient très souvent sur elles. Elles sont réellement devenues un rouage d'une machine qui tournerait plus difficilement sans leur intervention. Outre les garanties qu'elle offre, l'assurance fournit à l'économie une épargne importante favorable à son développement. Malheureusement le domaine des assurances en Algérie reste peu exploité et principalement dominé par les entreprises étatiques.

Dans ce présent chapitre nous présentons brièvement les différentes notions relatives au domaine des assurances, et exposons quelques statistiques de ce marché en Algérie et à travers le monde.

#### 4.1.1 Définition de l'opération d'assurance

Selon la formulation proposée par le professeur Hémard : L'assurance est une opération par laquelle une partie, l'assuré, se fait promettre, moyennant une rémunération (la prime ou cotisation), pour lui ou pour un tiers en cas de réalisation d'un risque, une prestation par une autre partie, l'assureur, qui prenant en charge un ensemble de risques, les compense conformément aux lois de la statistique. [Yeatman, 1998]

Dans la loi Algérienne : L'assurance est, au sens de l'article 619 du code civil, un contrat par lequel l'assureur s'oblige, moyennant des primes ou autres versements pécuniaires, à fournir à l'assuré ou au tiers bénéficiaire au profit duquel l'assurance est souscrite, une somme d'argent, une rente ou une autre prestation pécuniaire, en cas de réalisation du risque prévu. [CNA, 2012]

### 4.1.2 L'assurance à travers l'histoire

L'assurance est une institution relativement récente, en réalité elle n'a fait son apparition qu'à la fin du moyen âge sous la forme de l'assurance maritime, comme conséquence du développement du commerce de mer dans tous les pays du bassin méditerranéen. Les banquiers, dans un but spéculatif, proposait des formules aux armateurs en guise de garantie de leurs cargaisons. Ces formules n'étaient pas des opérations d'assurances mais juste un déplacement du risque qui, au lieu d'être supporté par les propriétaires de la marchandise, était pris en charge par ces banquiers spéculateurs.

Enfin, et après plusieurs tractations entre banquiers et législateurs, le premier contrat d'assurance a vu le jour en 1347 et a été signé à GENES. Il couvrait la cargaison du « Santa Clara » pour un voyage de Gênes à Majorque.

Après l'assurance maritime, l'assurance terrestre fut son apparition sous forme d'assurance contre l'incendie. Ceci a été en Angleterre après le grand incendie de Londres de 1666. Quant à l'assurance vie, elle doit ses prémices au banquier Italien Napolitain Lorenzo Tonti, qui a proposé au 17ème siècle une formule par laquelle les survivants profitèrent de la part des prémourants. [Lezoul, 2011]

### 4.1.3 Les produits d'assurance

Les produits d'assurances à travers le monde sont structurés en deux grandes familles comme présenté dans la figure ci-dessous.

Assurance non vie			Assurance vie
Assurance de Biens	Assurances de responsabilités	Assurance santé	Assurances vie (vie, décès, épargne, retraite)
Assurances de dommages		Assurances de personnes	

TABLE 4.1 – Les catégories d'assurance

Une distinction fondamentale des produits d'assurances est celle qui sépare les assurances de dommages des assurances de personnes.

#### ► Les assurances de dommages :

Elles ont pour but d'indemniser l'assuré contre les conséquences d'un événement accidentel affectant son patrimoine. L'assureur de dommages garantit, sous les conditions du contrat, qu'après survenance d'un sinistre, le patrimoine de l'assuré sera reconstitué en valeur comme si ce sinistre n'avait pas eu lieu. A leur tour les assurances de dommages se subdivisent en deux grandes catégories :

1. Les assurances de choses : Qui garantissent les biens appartenant directement à l'assuré.
2. Les assurances de responsabilité : Qui garantissent les dommages que l'assuré pourrait causer à d'autres personnes. Il s'agit d'une garantie indirecte du patrimoine de l'assuré puisque l'assureur s'engage à payer à sa place les sommes nécessaires à la réparation des dommages causés.

► **Les assurances de personnes :**

Elles ont pour objet le versement de prestations forfaitaires en cas d'événements affectant la personne même de l'assuré. Seules les personnes physiques sont concernées par ce type d'assurance. A leur tour les assurances de personnes regroupent deux grandes familles :

- Les assurances individuelles accidents et maladie.
- Les assurances sur la vie.

#### 4.1.4 La réassurance

La réassurance est l'opération par laquelle une compagnie d'assurance (appelée dans ce cas : la cédante) se fait garantir par une autre compagnie d'assurance (appelée dans ce cas le réassureur) pour se couvrir d'une partie des risques. En d'autres termes c'est l'assurance des assurances.

En réassurance il y a deux catégories de couverture : les traités et les facultatives.

- **Les traités :** Dans le cadre du traité, la cédante (qui est la compagnie d'assurance) a pour obligation contractuelle de céder une partie spécifique de certains types ou catégories de risques. De l'autre côté le réassureur a pour obligation d'assumer ces derniers.
- **Les facultatives :** Dans le cadre d'une facultative, la cédante cède au (x) réassureur (s) tout ou une partie du risque couvert par la police d'assurance spécifique unique. La facultative est négociée séparément pour chaque police réassurée. Généralement les facultatives sont achetées par les assureurs pour des risques individuels qui ne sont pas couverts par leurs traités de réassurance.

Exemple : Je dois souscrire un contrat pour lequel le capital assuré est très important, je ne peux pas prendre le risque de l'assurer intégralement. Donc je prends en charge 20 % du risque (par exemple) et je négocie avec un ou plusieurs de mes réassureurs pour les 80 % restants.

#### 4.1.5 Le marché mondial des assurances

En 2010 le volume des cotisations (primes) dans le domaine des assurances au niveau mondial a atteint 4339 Milliards de Dollars, soit une hausse de plus de 2,7% par rapport à 2009. Le chiffre d'affaires de l'assurance non vie en hausse de 2,1% a enregistré un niveau de 1819 Milliards de Dollars. En contrepartie le chiffre d'affaires de l'assurance vie était estimé à 2520 Milliards de Dollars, soit une hausse de 3,2% en 2010.

Le tableau suivant liste les plus grosses entreprises mondiales d'assurances en termes de chiffres d'affaires. Il est extrait du Classement Global 500 du magazine Fortune de 2015.

Rang	Entreprise	Pays	Chiffres d'affaires (Milliards \$)
1	Allianz	Allemagne	124.9
2	AXA	France	111.0
3	Generali	Italie	79.3
4	China Life Insurance	Chine	66.2
5	American International Group	USA	58.3
6	Nippon Life Insurance Company	Japon	56.5
7	Prudential plc	Royaume-Uni	51.4
8	Zurich Fianancial Services	Suisse	48.5
9	Aviva	Royaume-Uni	31.7
10	Aegon	Pays-Bas	22.9

TABLE 4.2 – Les 10 premières entreprises mondiales d'assurance en termes de chiffres d'affaires.

#### 4.1.6 Le marché Algérien des assurances

L'Algérie se caractérise avant tout par la faiblesse relative de son marché. L'assurance algérienne occupe le 68ème rang mondial avec une part de marché de 0,016% du marché mondial. L'Algérie participe seulement à hauteur de 1,3% du marché de l'assurance du continent africain qui lui-même ne pèse guère plus de 1,1% du marché mondial de l'assurance, avec pourtant une population de plus de 900 millions d'habitants. En faisant impasse sur l'Afrique du Sud qui totalise à elle seule 82 % du marché continental, l'Algérie ne représente que 7 % du marché africain.

Depuis la libéralisation du secteur des assurances en Algérie, ce dernier ne cesse d'évoluer. Par exemple en 2010, les compagnies d'assurances en Algérie ont enregistré un chiffre d'affaires de 81.3 milliards de dinars<sup>50</sup>. Le taux de pénétration, qui représente la part du secteur dans le produit intérieur brut (PIB), est de 0.7% en 2010. La participation du secteur dans la production nationale demeure très faible. En effet, l'Algérie reste en retard, même par rapport à ses voisins (en Tunisie ce taux est de 2% et en Maroc il est de 3%).

Quant à la densité d'assurance, qui représente le montant des primes d'assurance par habitant, elle s'est établie à 2322 dinars en 2010 contre 2259 dinars en 2009. Même sur ce plan l'Algérie demeure en retard par rapport à ses voisins Tunisiens, dont la densité a atteint 5000 dinars par habitant, et Marocains dont la densité était de 4500 dinars en 2010. [CNA, 2012]

Année	2003	2004	2005	2006	2007	2008	2009	2010
Chiffre d'affaires	31.2	36.7	41.4	46.5	53.8	76.6	77.3	81.3
Taux de croissance (%)	7.22	17.63	12.81	12.32	15.7	25.65	14	5

TABLE 4.3 – l'évolution du chiffre d'affaires du secteur des assurances en milliards de dinars

Année	2004	2005	2006	2007	2008	2009	2010
Sinistres réglés	20 150	53 273	24 658	25 650	34 809	35 927	35 927
Sinistres à payer	65 616	32 515	38 956	44 045	48 652	52 455	55 372

TABLE 4.4 – Evolution de la sinistralité (Millions DA)

#### 4.1.7 Assureurs Algériens

### Historique

Avant 1962, il y'avait 270 société d'assurances opérées en Algérie. La plupart de ces sociétés étaient privées et avaient leur siège soit en France, en Tunisie, ou au Maroc, dont 04 seulement parmi elles, avaient leur siège en Algérie.

Le contrôle de ces sociétés s'effectuait par l'intermédiaire des services compétent du ministère de finance.

Ils se limitaient à la transmission des documents statistiques, se chargeaient de régler des compagnies et veillaient au respect stricte de la réglementation en vigueur.

Au début de 1962, les assurances ont fait face à une période difficile dans laquelle on remarque le manque des cadres nationaux.

La loi du 08 juin 1963 a modifié et complété le cadre juridique des assurances, ainsi la CAAR (caisse Algérienne d'assurance et de réassurance) est le premier établissement crée après l'indépendance.

Après, on a vu la naissance de la première compagnie d'assurance en Algérie qui est la SAA ; société mixte Algéro-égyptienne.

Cette dernière ainsi que les autres compagnies à savoir MAATEC, CNMA, CAAT, CCR sont normalisées par l'institution de monopole de l'état sur les opérations d'assurances.

C'est à partir de 1976, que la SAA s'est vu spécialisée dans l'exploitation de la branche automobile, la CAAR dans les risques industriels et simples, la CAAT dans le transport.

Grâce à la spécificité de cette branche orientée principalement vers les particuliers et les PME/PMI (automobile), la SAA a procédé à l'expansion de son réseau, jusqu'à 24 unités de production, 307 agences et 35 centres d'expertise repartis sur l'ensemble de territoire national.

Cette expansion reposait sur la nécessité, la mise en place d'une prestation de service assez bonne, se rapproche et être disponible pour la clientèle.

L'avènement des réformes économiques a permis de transformer la SAA en EPE le 27 février 1989. La déspecialisation causée par ces réformes a élargi son domaine d'activité pratiquée actuellement à toutes les opérations des compagnies.

Ses potentiels et sa dimension à l'échelle nationale ont collaboré à l'augmentation de son capital social jusqu'à 9,8 milliard de DA.

Ainsi en janvier 1995, l'ordonnance N° 07/95 sur les assurances consacre la libération marché aux capitaux privés nationaux, internationaux, et étrangers.

#### ► **Compagnies d'assurance Public**

SAA, CAAR, CAAT, CASH, CNMA, CARAMA, SAPS, TALA, MAATEC, Gulf Insurance Group (GIG), AXA Vie Mixte (Public et Privé), AXA Dommage Mixte (Public et Privé), Algérien Gulf life Insurance compagnie (ALGIC)

► **Compagnies d'assurance Privé**

CIAR, ALLIANCE, 2A, GAM, SALAMA, TRUST, Cardif, MacirVie, La Mutualiste NHBB, Assurances Diligence, Assurance Courtag, Algérien Gulf life Insurance compagnie (ALGIC).

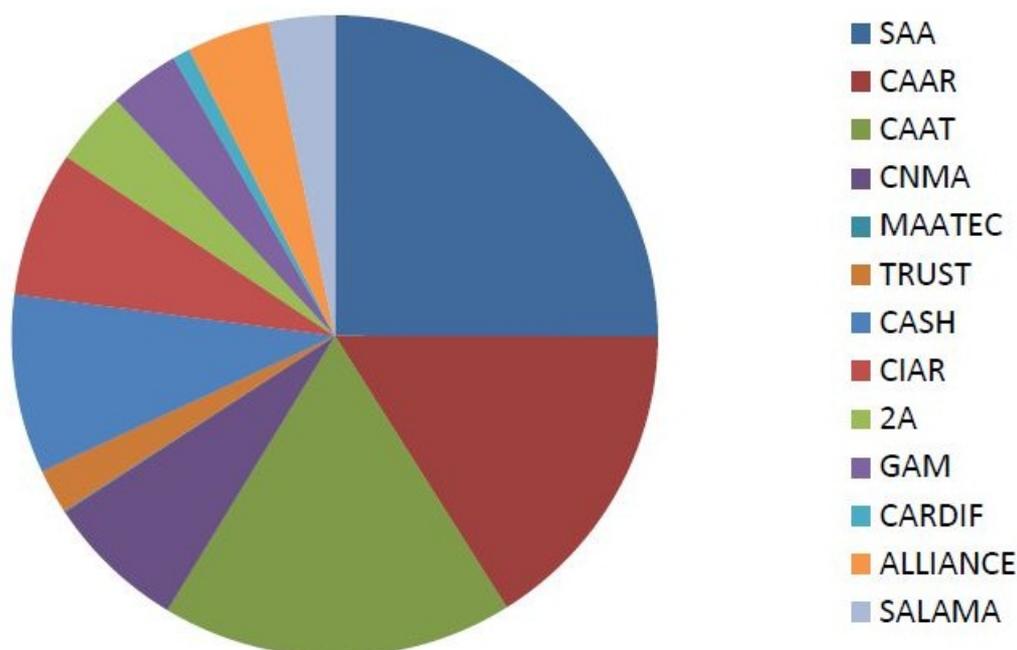


FIGURE 4.1.1 – Structure du marché des assurances en Algérie.

## 4.2 Présentation de l'organisme d'accueil « SAA » [DI]

La SAA « société nationale des assurances » est une société publique, de prestation de service spécialisée dans le domaine des assurances et réassurances.

### Siège social et capital de la SAA

Le siège social de la société se situe à Alger centre ; 05, Boulevard Ernesto Che Guevara, sous la tutelle du ministère des finances. Son capital social s'élève à 20.000.000.000 DA.

### Moyens et domaine de travail

La SAA dispose actuellement de :

- 14 Directions régionales ;

- ▶ 500 Agences ;

Son domaine d'intervention couvre pratiquement toutes les opérations d'assurances des diverses branches :

- ▶ Assurances d'automobiles ;
- ▶ Assurances risques divers : risque simple, industriel, et engineering et construction ;
- ▶ Assurance des risques transport ;
- ▶ Assurances des risques agricoles ;

La société assure aussi des prestations de service après-vente aux assurés à savoir :

- ▶ Visites de risque ;
- ▶ Expertise automobile et risques divers ;
- ▶ Indemnisation ;

Elle pratique également le conseil en assurance :

- ▶ communication et publicité ;
- ▶ Formation.

#### **4.2.1 Division et filiales**

Pour la SAA, la filialisation revêt un caractère stratégique visant à mieux organiser ses activités techniques et à se départir des services de soutien dans le but de concentrer ses forces potentielles sur le métier d'assurance.

C'est ainsi qu'elle a créé des filiales spécialisées dans les domaines suivants :

#### **SAE**

La société Algérienne des Expertises, par abréviation SAE, dont la mission principale est de réaliser des travaux d'expertise , en automobile et en risques divers, aussi bien pour le compte des structures de la SAA que pour les autres compagnies du secteur des assurances.

#### **AMANE**

Société d'Assurances de Prévoyance et de Santé, en partenariat avec la MACIF, la BDL et la BADR, dont la mission consiste à développer les produits d'assurances vie et de capitalisation.

#### **IPA**

La filiale assistance au véhicule par abréviation IPA « Inter Partner Assistance », dont la mission porte sur l'assistance aux véhicules et leurs occupants en cas de panne.

## IMPRIMERIE DES ASSURANCES

La filiale imprimerie des assurances, en partenariat avec la CCR, dont la mission porte sur la conception et la réalisation de travaux d'impression tant pour le compte de la SAA, de la CCR que pour le compte des tiers.

### 4.2.2 La SAA en chiffres

Branches / Année	2010	2011	2012	2013	Evolution 10-13
<b>Risque Obligatoire</b>	2 911	2 908	3 060	3 090	6%
<b>Risques Non Obligatoires</b>	11 663	12 470	15 396	17 402	49%
<b>Total Automobile</b>	14 574	15 378	18 456	20 492	41%
<b>Risques des Particuliers et Professionnels</b>	1 514	1 712	1 721	1 915	26%
<b>Assurances Risques Industriels</b>	1 895	2 047	2 231	2 438	29%
<b>Assurance Transport</b>	287	366	292	312	9%
<b>Assurances Agricoles</b>	184	218	256	445	142%
<b>Total Risques d'Entreprises</b>	2 366	2 631	2 779	3 195	35%
<b>Assurance de Personnes</b>	1 619	1 426	208	155	-90%
<b>TOTAL</b>	20 073	21 147	23 165	25 757	28%

FIGURE 4.2.1 – Evolution du chiffre d'affaires par branche 2010-2013 (En millions de DA)

Branches / Années	2010	2011	2012	2013	Evolution 10-13
<b>Auto. Matériels</b>	5 911	7 214	9 274	9 939	68%
<b>Auto. Corporels</b>	3 077	3 145	3 294	3 664	19%
<b>TOTAL AUTOMOBILE</b>	8 988	10 359	12 568	13 603	51%
<b>RISQUES DES PARTICULIERS &amp; PROFESSIONNELS</b>	160	180	203	226	41%
<b>Assurances Risques Industriels</b>	167	317	311	398	138%
<b>Assurances Transport</b>	325	78	326	21	-94%
<b>Assurances Agricoles</b>	62	95	94	82	32%
<b>TOTAL RISQUES D'ENTREPRISES</b>	554	490	731	501	-10%
<b>ASSURANCES PERSONNELS</b>	426	443	653	212	-50%
<b>TOTAL</b>	10 128	11 472	14 155	14 541	44%

FIGURE 4.2.2 – Evolution des indemnités par branche (En millions de DA)

### 4.2.3 L'organisation générale de la SAA

L'organisation générale de la société est divisée en trois niveaux de structure :

#### 4.2.3.1 Direction générale de la société

La direction générale est organisée en « staff and line » évolué où l'on a une double organisation : l'une « en line » ou hiérarchique et l'autre « en staff » ou fonctionnelle. L'organisation « en line » est constituée de trois (03) pôles importants d'activité :

1. Un pôle technico-commercial, chargé :
  - Directement ou indirectement de la réalisation du chiffre d'affaire dans les différents marchés ciblés par la stratégie de développement de l'entreprise ;
  - De mettre en œuvre des politiques de souscription et de sélection adéquates conduisant vers la réalisation de résultats techniques positifs ;
  - D'élaborer et d'actualiser la politique marketing à travers la connaissance permanente des marchés, le développement de nouveaux produits, etc. ;
2. Un pôle ressources, chargé de la gestion des ressources humaines, matérielles et financières de l'entreprise, en application de la politique de développement retenue par l'entreprise ;
3. Un pôle système d'informations, audit et contrôle général chargé de l'organisation, de la rationalisation et de la modernisation de la gestion au sein de l'entreprise. Une telle organisation permet au Directeur Général de disposer de huit (08) cadres dirigeants au niveau central, constitués outre des deux (02) adjoints du PDG, de six (06) chefs de divisions du fait de la décomposition :
  - ▶ Du pôle technico-commercial en cinq (05) structures : la première chargée des risques d'entreprises, la deuxième des assurances des particuliers, des professionnels et des collectivités, la troisième de l'assurance automobile, la quatrième des assurances de personnes et la cinquième du marketing ;
  - ▶ Du pôle, audit, contrôle général et système d'information en une (01) structure chargée à la fois de l'informatique et des systèmes d'informations, de l'audit et de la maintenance des procédures ainsi que du contrôle général. Le pôle ressources (humaines, matérielles et financières) est placé sous la responsabilité directe du Directeur Général Adjoint Administratif. La même organisation « en staff » permet au Directeur Adjoint Technique de superviser l'activité des cinq divisions pré citées ainsi que celle des deux directions dont il est directement responsable :
  - Direction Réassurance chargée de la politique et de l'activité de réassurance (activité transversale importante pour sécuriser les engagements de l'entreprise, en relation avec les quatre (04) Divisions techniques ;
  - Direction du contentieux et de la réglementation chargé du suivi de l'évolution de la législation et du contrôle de sa mise en œuvre au sein de l'entreprise, ainsi que généralement du suivi des dossiers contentieux relevant de l'activité générale de l'entreprise (dans le domaine civil, commercial, administratif, social, pénal, immobiliers, fiscal, etc.) et particulièrement du suivi des sinistres lorsqu'ils sont très importants. Elle est aussi compétente

pour rédiger et se prononcer sur tout acte administratif ou juridique entre l'entreprise et tous tiers (fournisseur ou autre)

#### **4.2.3.2 Direction régionale**

Les directions régionales sont au nombre de quatorze(14) implantées ainsi :

- Centre (4) : Alger\_Bir Mourad Rais ; Alger\_château neuf ; Mouzaia ; Tizi-ouzou.
- Ouest(4) : Oran ; Ghelizane ; Sidi bel Abbes ; Tlemcen.
- Est(4) : Constantine ; Sétif ; Batna ; Annaba.
- Sud(2) : Ouargla ; Béchar.

Chaque direction régionale est constituée de cinq départements, leur mission est d'encadrer, d'assister, d'animer et de contrôler les agences qui leurs sont rattachées.

Elles sont organisées comme suit :

- ▶ 01 directeur d'unité ;
- ▶ Département administration ;
- ▶ Département contentieux et sinistre ;
- ▶ Département production ;
- ▶ Département commercial (marketing) ;
- ▶ Département comptabilité et finances.

#### **4.2.3.3 Les agences**

Les agences sont au nombre de 500, et sont classées en trois catégories :

- ▶ Agence directe ;
- ▶ Agence à revenu proportionnel ;
- ▶ Agence générale.

Chaque agence est composée de trois services :

- ▶ Service production ;
- ▶ Service sinistre ;
- ▶ Service comptabilité et finance.

L'agence est la cellule vivante chargée d'accomplir les objectifs qui lui sont attribués par l'unité et œuvre à répondre aux besoins de la clientèle :

- ▶ Vente et gestion des contrats d'assurance ;
- ▶ Prospection de nouveau gisement du marché des assurances ;
- ▶ Règlement des sinistres et assurer un contact permanent avec la clientèle.

Chaque agence est constituée de trois services :

1. Service production : Au niveau du service production, on distingue trois types de section à savoir la section production automobile, la section production risques divers et la section production assurance personnel.

2. Service sinistre : Le service sinistre est composé de la section sinistre corporel d'une part et section sinistre matériel d'autre part.
3. Service comptabilité et finance : Il est chargé d'établir la situation comptable et finance de l'agence. Notamment, les versements de la liquidité de l'agence à la banque en fin de journée, de porter les écritures comptables dans les bordereaux appropriés et d'envoyer les situations de la production et du sinistre à la structure de rattachement au niveau de la direction régionale.

#### **4.2.4 Organigrammes**

On distingue trois types de diagrammes :



Organigramme des directions régionales :

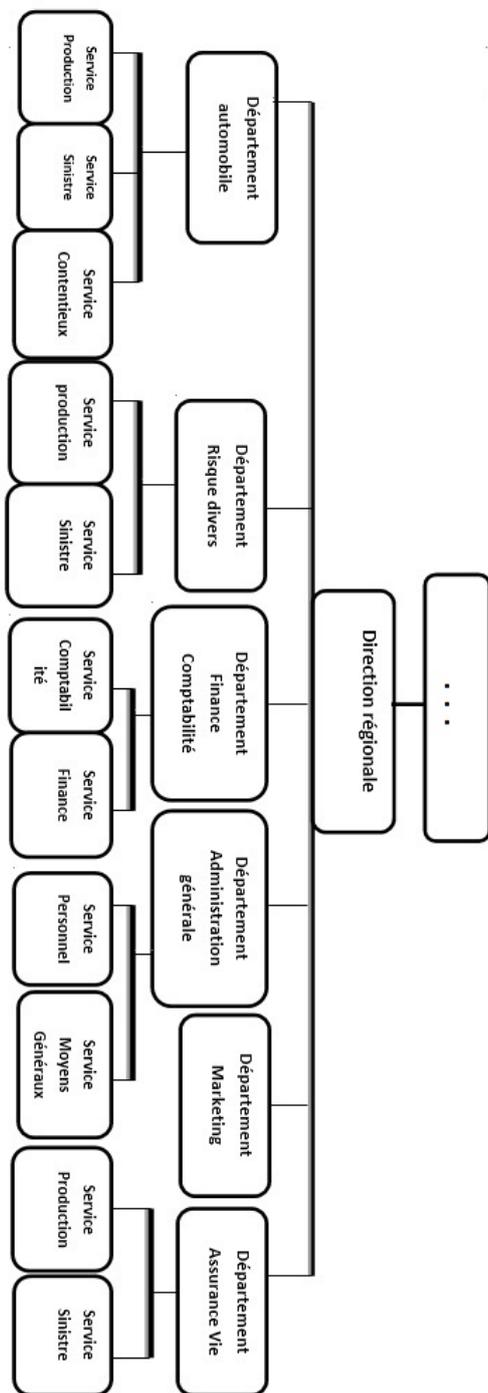


FIGURE 4.2.4 – Organigramme des directions régionales

Organigramme des agences :

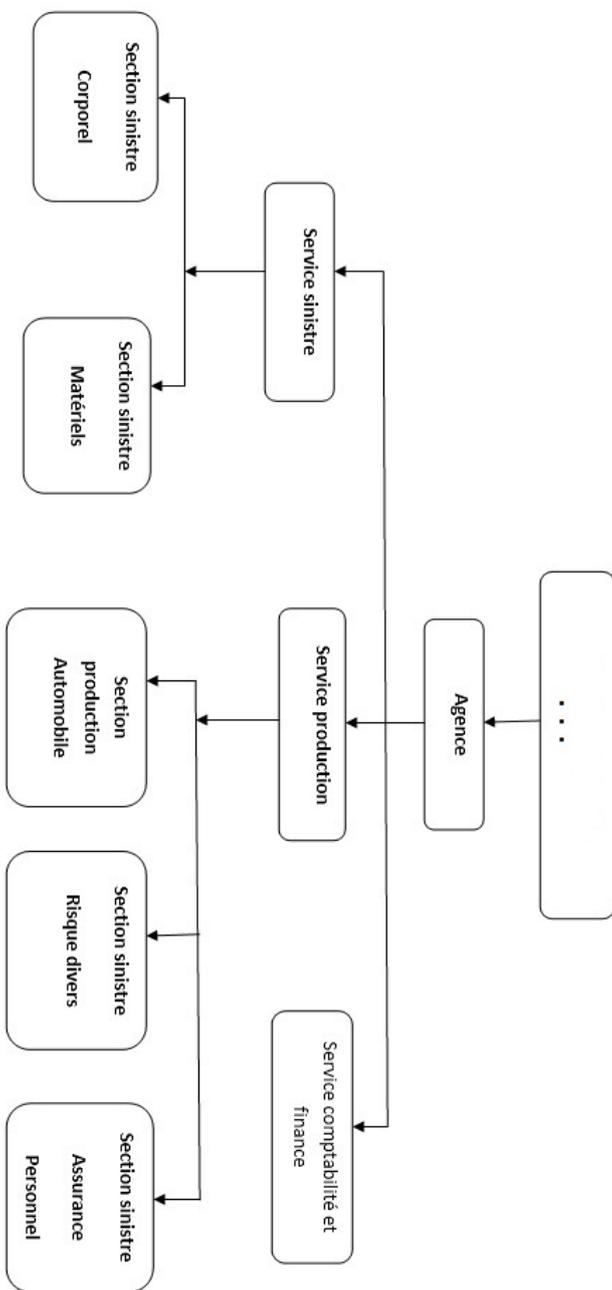


FIGURE 4.2.5 – Organigramme des agences :

## Conclusion

Le marché des assurances connaît de profondes mutations compte tenu de son rôle économique et social. D'une manière générale, l'Etat a déployé des efforts considérables afin de lever les restrictions et obstacles qui se dressent devant les investisseurs dans le domaine des assurances. Les opportunités d'investissements sont encore plus grandes dans le secteur des assurances de personnes, compte tenu du faible taux de pénétration et des nouveaux besoins de sécurité et de prévoyance.

Malgré la réforme du cadre réglementaire du secteur, il demeure largement en retard par rapport aux besoins de l'économie Algérienne, et aurait besoin d'être largement boosté. Le secteur tout entier est soutenu par l'assurance automobile obligatoire et l'assurance des personnes à cause essentiellement des assurances de groupes et de l'assurance exigée pour l'obtention d'un crédit bancaire. La dynamisation du marché des assurances par le développement des assurances de personnes et de capitalisation.

De plus denses et plus pointus efforts demeurent encore à fournir par l'ensemble des acteurs du secteur, à commencer par les assureurs, en vue, notamment, d'améliorer l'aspect prospectif de leurs politiques respectives de marketing et commercialisation.

En seconde partie du chapitre, nous avons introduit l'organisme d'accueil, la SAA. Cette présentation nous a permis de mieux comprendre le fonctionnement de ses différents départements, et présenter les différentes filiales.

# Chapitre 5

## Conception de la solution

Après avoir étudié et analysé les besoins des utilisateurs pour lesquels l'entrepôt de données doit répondre, la conception de ce dernier est nécessaire avant sa réalisation. Cette partie est consacrée pour la conception de l'entrepôt de données, que nous avons effectué en quatre étapes :

- ▶ La conception de la zone d'entrepôt : consiste à définir le schéma conceptuel de l'entrepôt (les tables de faits, les tables de dimensions,...).
- ▶ La conception de la zone d'alimentation : consiste à définir le déroulement du processus ETL, qui sert à intégrer les données dans le Data warehouse.
- ▶ La conception du reporting et du tableau de bord.
- ▶ La conception de l'architecture de la solution.

### 5.1 Conception de la zone d'entrepôt

Pour concevoir le modèle multidimensionnel de notre entrepôt de données, nous avons utilisé la méthode Faits/Dimensions qui permettent au mieux de répondre aux demandes des utilisateurs, en leur facilitant l'exploitation des données de l'entrepôt. Le processus de modélisation qu'on a suivi est celui de Kimball, qui est composé comme suit :

- ▶ **Choix de processus d'activité à modéliser :**

Un processus d'activité est un processus opérationnel important pour l'organisation. Pour notre cas, le processus clé est : le suivi des sinistres

- ▶ **Choix de grain du processus d'activités :**

C'est le niveau atomique des données figurant dans la table des faits pour ce processus. Un entrepôt de données exige presque toujours des informations exprimées dans le grain le plus bas possible pour chaque dimension.

- ▶ **Choix des dimensions applicables à chaque enregistrement de la table des faits :**

Le choix d'une dimension s'accompagne par la définition de tous les attributs textuels qui garniront la table de dimension.

- ▶ **Choix des faits mesurés que contiendra chaque enregistrement de table de faits :**

Qui sont généralement des quantités numériques additives telles que le montant des remboursements, le nombre d'accidents par région ...

### 5.1.1 Formalisme utilisé

Avant de faire la conception dimensionnelle des trois activités, on va décrire le formalisme à utiliser. Les schémas suivants nous illustrent ce formalisme.

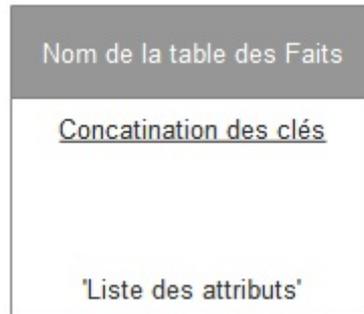


FIGURE 5.1.1 – Formalisme adopté pour la table de faits



FIGURE 5.1.2 – Formalisme adopté pour la table de dimension



FIGURE 5.1.3 – Formalisme adopté pour la relation entre les deux tables

## 5.1.2 Modélisation dimensionnelle des activités

### 5.1.2.1 Modélisation dimensionnelle de l'activité « Production »

#### Présentation de l'activité

Le choix du processus d'activité à modéliser représente la première étape de la conception. L'activité 'Production' est le premier processus à modéliser dans la conception de notre Data Mart. Cette activité concerne la souscription de nouveaux clients.

## Grain du processus d'activité

La granularité de la table des faits choisie est la plus fine possible, le choix de la granularité consiste à indiquer qu'un enregistrement sera créé dans la table des faits pour chaque souscription d'un nouveau client (transaction unitaire) par : Vehicule, garantie, agence, temps. A partir de ça notre schéma dimensionnel de la première activité commence à prendre forme. La figure suivante illustre le modèle dimensionnel de base de ce processus.

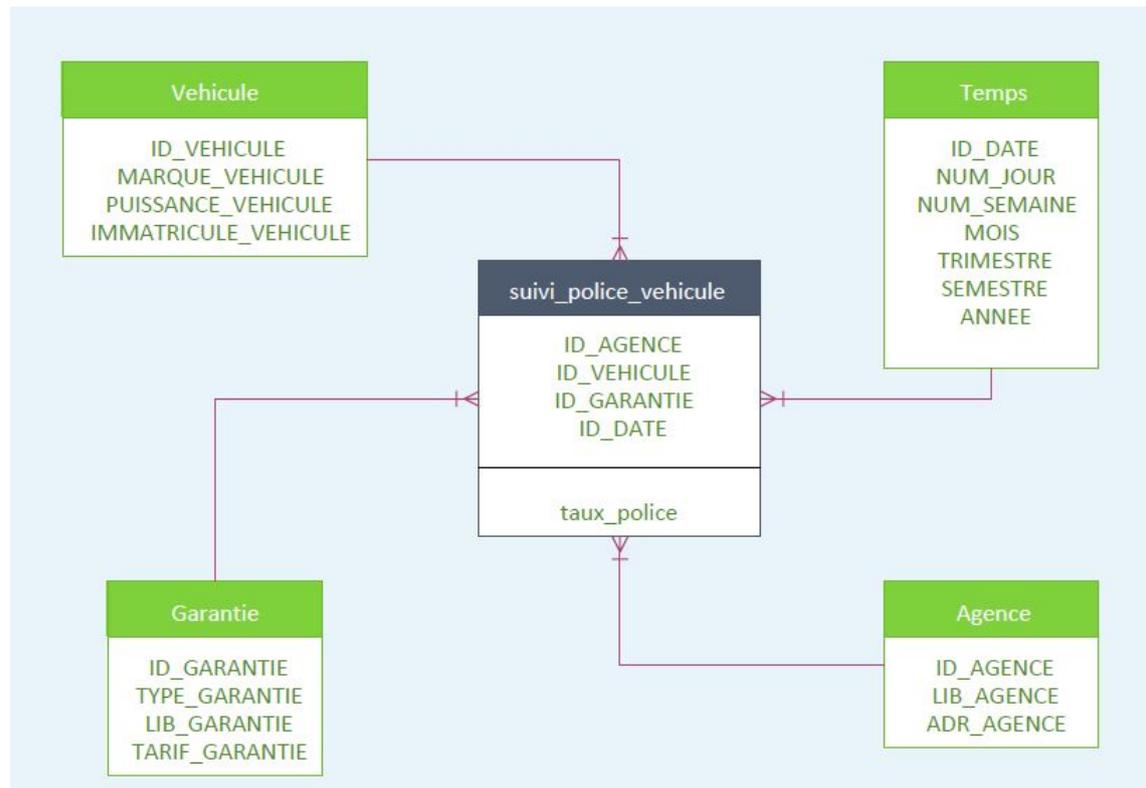


FIGURE 5.1.4 – Schéma de base du modèle dimensionnel de l'activité « Production »

## Les dimensions de l'activité « Production »

A partir du schéma initial illustré ci-dessus, nous avons déterminé une série de dimension qui sont les suivantes :

- **Dimension Vehicule :**

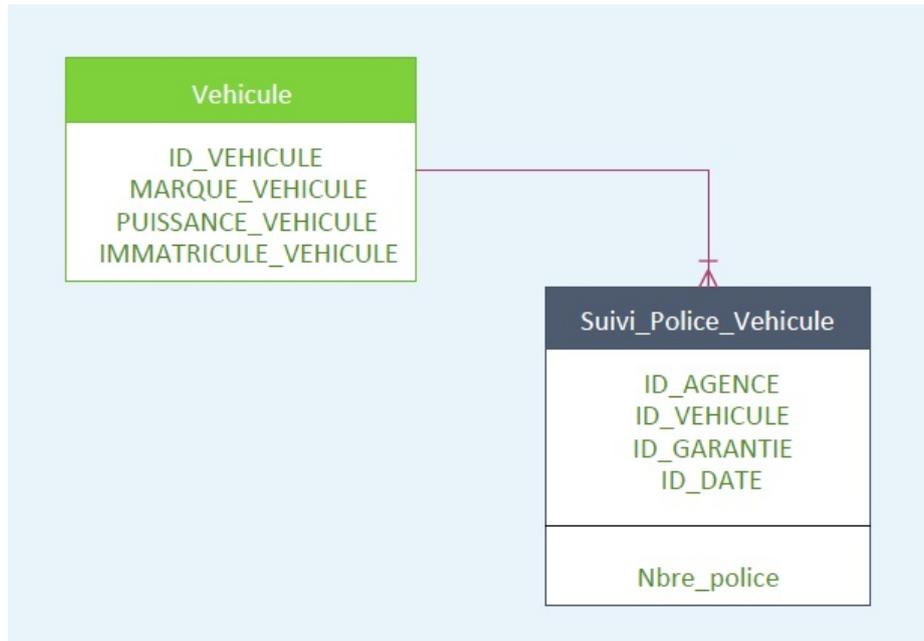


FIGURE 5.1.5 – Dimension Vehicle

► Dimension Agence :

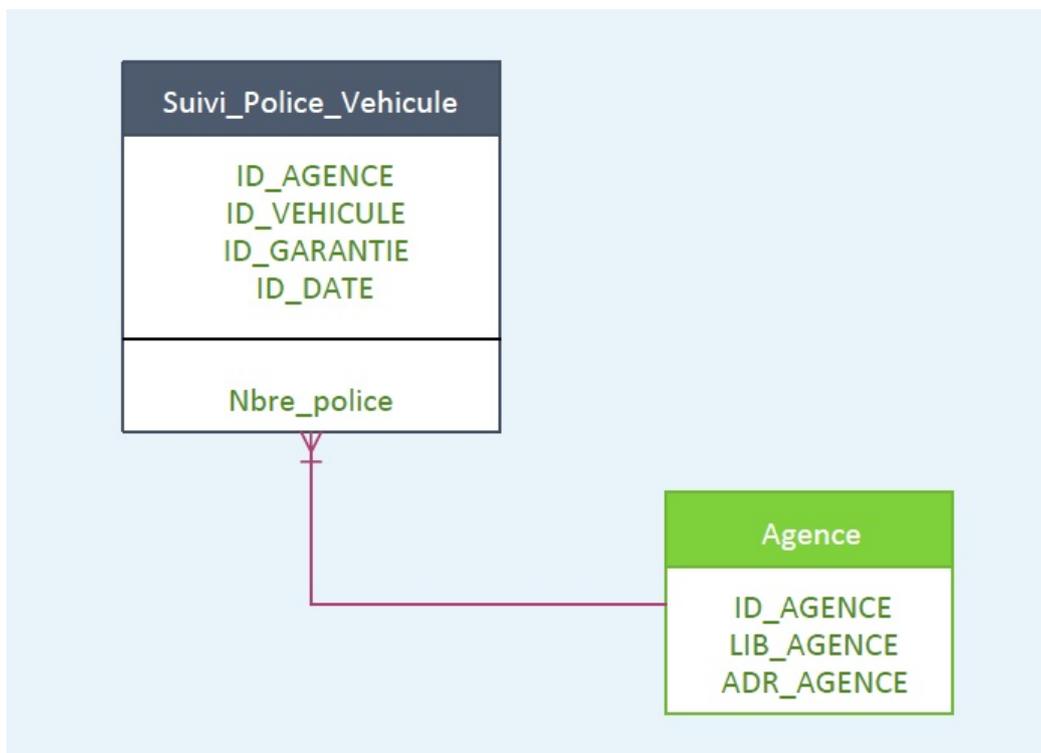


FIGURE 5.1.6 – Dimension Agence

► Dimension Garantie :

► Dimension Temps :

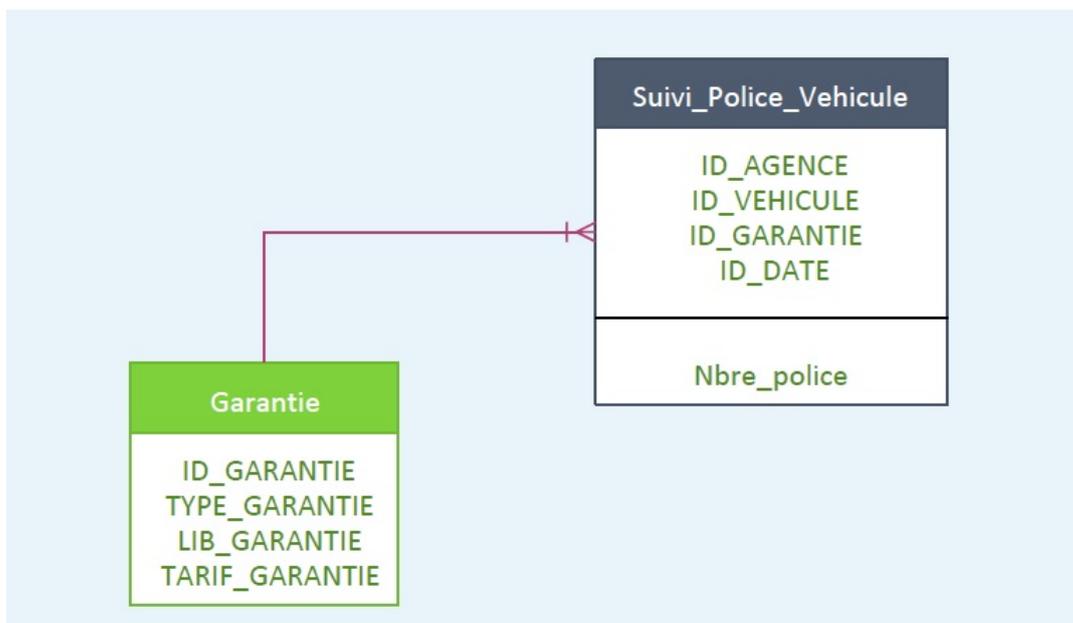


FIGURE 5.1.7 – Dimension Garantie

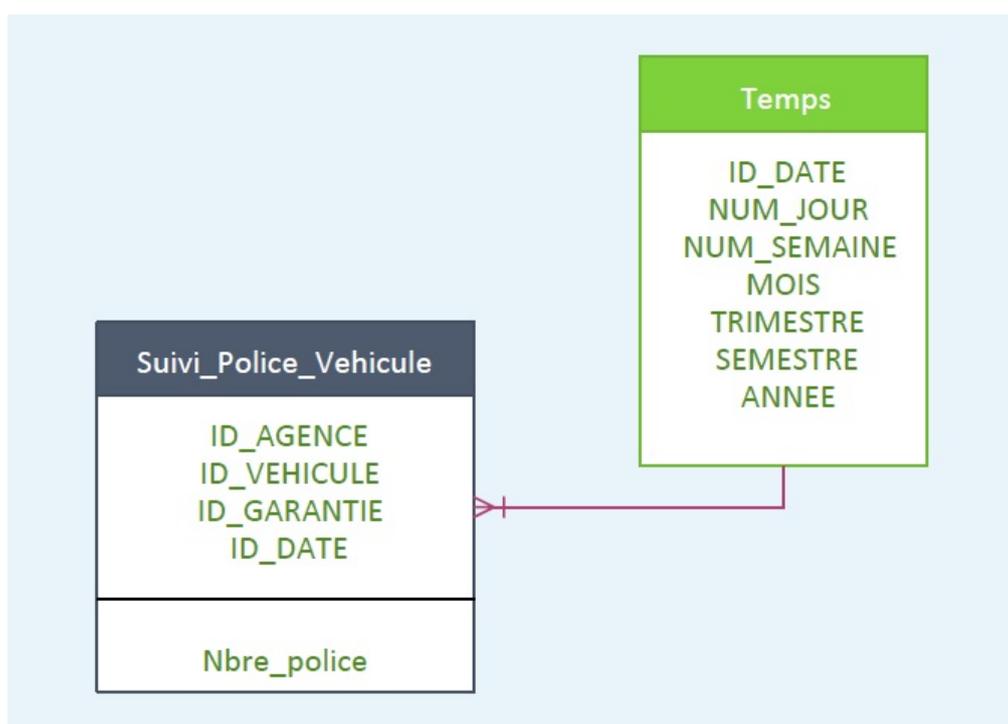


FIGURE 5.1.8 – Dimension Temps

### 5.1.2.2 Modélisation dimensionnelle de l'activité « Suivi Sinistre »

#### Présentation de l'activité

Le deuxième processus clé choisi dans la création de notre magasin de données est le « Suivi des sinistres ». Ce processus d'activité concerne les assurés victimes d'un accident.

Les clients de la SAA, ont une semaine d'intervalle pour déclarer leurs sinistres.

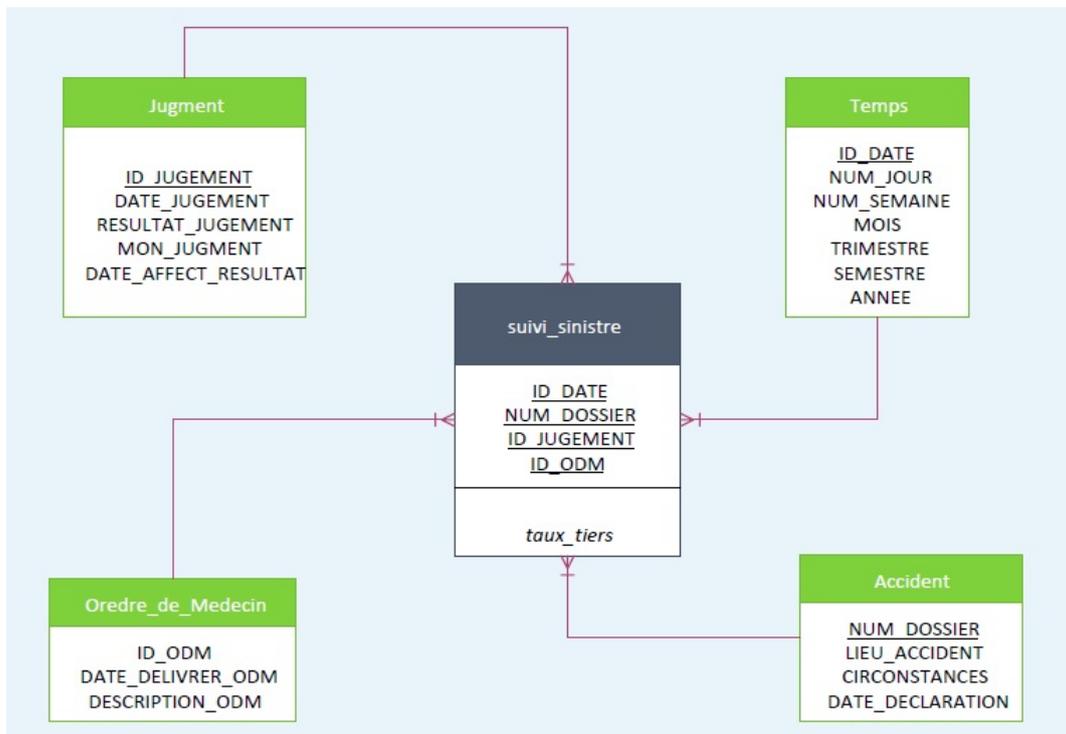


FIGURE 5.1.9 – Schéma de base du modèle dimensionnel de l'activité « Suivi sinistre »

### Les dimensions de l'activité « Suivi sinistre »

La granularité de l'activité « Suivi sinistre » détermine elle-même une série de dimensions principales, parmi ces dernières, on trouve une entrant en commun avec l'activité « Production » qui est « Temps ».

La dimension temps est selon Ralph Kimball : « la seule dimension qui figure systématiquement dans tout entrepôt de données, car en pratique tout entrepôt de données est une série temporelle. Le temps est le plus souvent la première dimension dans le classement sous jacent de la base de données ».

Cette dimension est dite « conforme » dans la mesure où elle est indiquée pour les deux tables des faits étudiées. Cette conformité rend possible le rapprochement des deux tables de faits « Production » et « Suivi sinistre » à travers la dimension qu'elles ont en commun.

- ▶ Dimension Accident
- ▶ Dimension Ordre de médecin
- ▶ Dimension Jugement

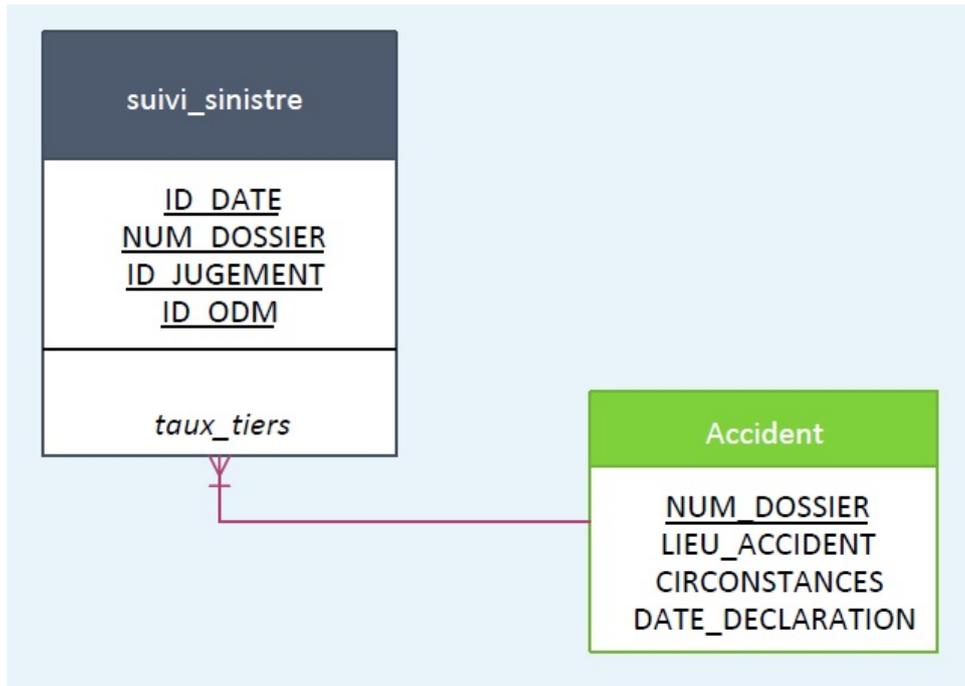


FIGURE 5.1.10 – Dimension Accident

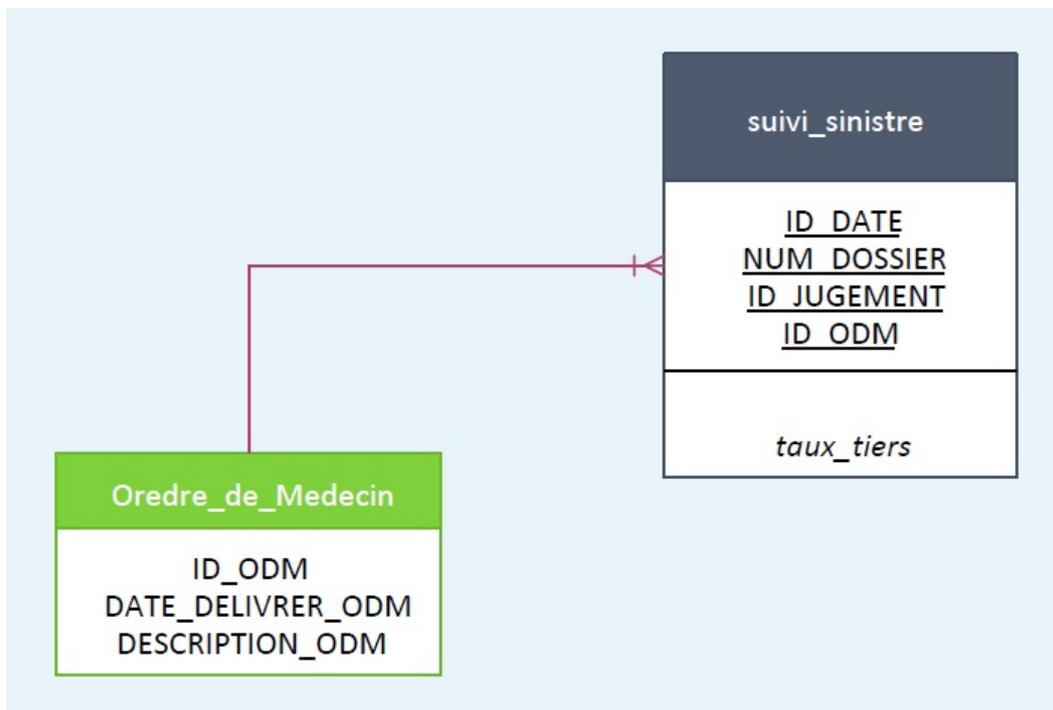


FIGURE 5.1.11 – Dimension Ordre de médecin

## 5.2 Conception de la zone d'alimentation

L'alimentation est la procédure qui permet de transférer des données du système opérationnel vers l'entrepôt de données en les adaptant. La conception de cette opération est une tâche assez complexe (elle constitue 70% d'un projet décisionnel en moyenne). Il est nécessaire de déterminer

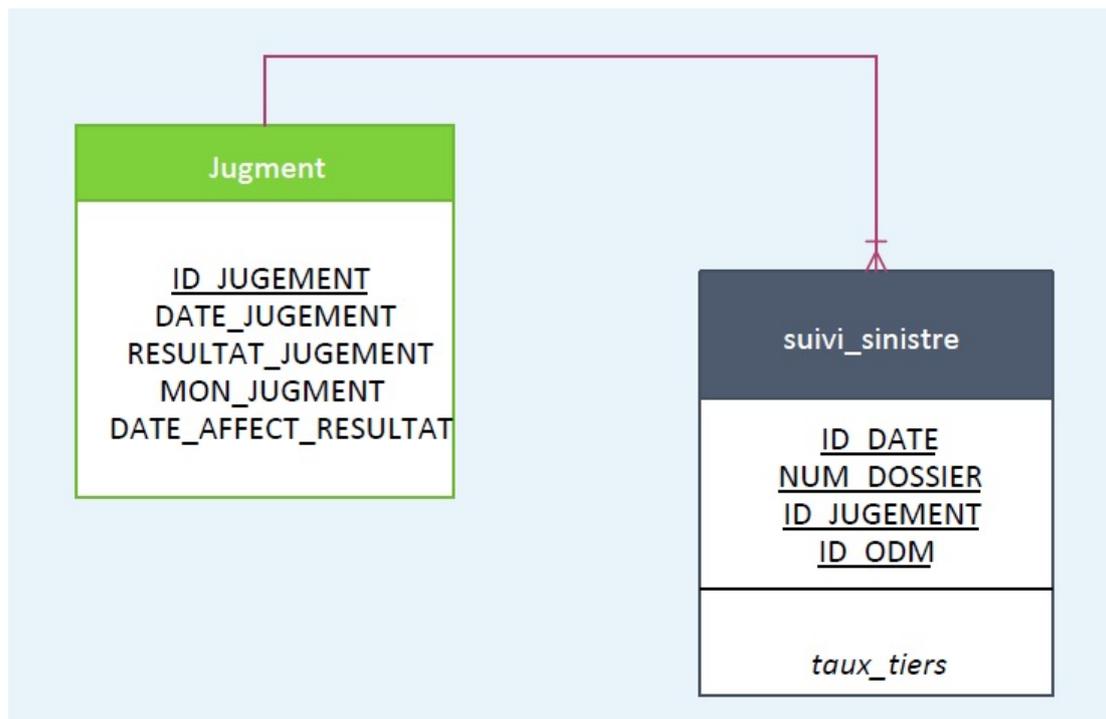


FIGURE 5.1.12 – Dimension Jugement

quelles données seront chargées, quelles transformations et vérifications seront nécessaires, la périodicité et le moment auxquels les transferts auront lieu.

### 5.2.1 Identification des sources de donnée

Avant de commencer la conception de l'ETL, il faut bien répondre aux questions : «comment sont mes sources? Et quelles données de production faut-il sélectionner pour alimenter le DW? ».

Les sources de données de notre entrepôt et la BDD de SAA. Afin de déterminer l'emplacement des données à charger dans l'entrepôt, nous avons :

1. Listé l'ensemble des informations dont nous avons besoin à partir des modèles en étoile .
2. Déterminé leurs emplacements sur la BDD .

Une fois la source définie, le processus d'alimentation peut commencer

### 5.2.2 Processus d'alimentation

Ce processus passe par trois étapes principales : Extraction, transformation et chargement.

1. L'extraction des données, en accédant aux différentes bases de données dont nous avons besoin.
2. La transformation.
3. Le chargement, pour alimenter les Data Marts et le Data Warehouse.

### 5.2.2.1 Extraction des données

L'extraction des données à partir du système source est la première étape du processus ETL. Cette opération peut commencer une fois le plan global de préparation de données établi et les sources de données identifiées précisément.

Dans un premier temps, toutes les données source du système, identifiées comme étant pertinentes, sont extraites et injectées dans zone de préparation de données (staging area) ; c'est là que seront opérés les différents traitements de ces données avant leur chargement dans les cubes multidimensionnels. Il s'agit de la première extraction sur le système source ; elle concerne par conséquent les données les plus récentes disponibles sur ce système. Cette technique d'extraction correspond à l'étape de mise en route du Data Warehouse.

### 5.2.2.2 Préparation du chargement : traitements des données

Les données extraites à partir du système source sont stockées dans la zone préparation de données en vue d'être nettoyées puis chargées dans les dimensions et les tables de faits. L'objectif de la staging area est l'obtention de données prêtes à être chargées dans une structure de Data Warehouse. Ainsi, toutes les étapes du traitement des données se font au sein de l'environnement relationnel de la staging area. Une fois la provenance des données précisément établie, il est nécessaire de définir les opérations de transformation des données.

## Operateurs de type transformation

Ces operateurs permettent de transformer les données des objets sources avant leur chargement dans les objets cibles. Les opérateurs de transformation les plus utilisés sont :

- ▶ Filtre : Permet de définir une condition « Where » sur les données
- ▶ Convertisseurs : Permet de convertir les types des données
- ▶ Déuplicateur : Supprime tous les doublons trouvés
- ▶ Expression personnalisée : Permet à l'utilisateur de définir un code SQL personnalisé pour la transformation
- ▶ Agrégat : Pour le calcul des fonctions d'agrégat

### 5.2.3 Préparation du chargement des dimensions

#### 5.2.3.1 Chargement des dimensions

Le chargement des dimensions est une tâche relativement simple ; considérant que tous les traitements sur les données ont été faits, il ne reste qu'à les insérer dans les structures représentant les dimensions.

Le chargement des données se fait avec un l'outil de chargement offert par PDI.

### 5.2.3.2 Chargement des dimensions temporelles

La particularité de la dimension temps est qu'elle n'existe pas dans le système source. Son chargement ne s'effectue qu'une seule fois dans le cycle de vie d'un Data Warehouse. Ceci est entièrement géré par l'outil de chargement fourni par PDI; cette dimension contiendra toutes les dates correspondant aux faits actuellement stockés dans le Data Warehouse ou qui le seront dans le futur.

### 5.2.3.3 Préparation du chargement des tables de faits

Chaque enregistrement (ligne) de la table de fait devrait, en principe, pouvoir être associé à chacune des dimensions du modèle. Pour garantir cela, un nombre de traitements intermédiaires sont nécessaires; cela va de la constitution de jointures entre les tables du système source à l'ajout de champs calculés et d'agrégats. Afin de permettre une optimisation des traitements en termes de ressources et de temps, des tables intermédiaires au chargement des cubes multidimensionnels ont été construites. La succession des chargements de données dans ces tables constitue le processus de traitement et de préparation du chargement des tables de faits (cubes). L'une des conditions du Data Warehouse étant de ne garder que les données fiables et pertinentes; un nombre d'autres traitements s'impose :

- ▶ La conversion de données textuelles en données numériques pour les calculs,
- ▶ La détection des anomalies et des données incohérentes,
- ▶ La suppression des lignes nulles,
- ▶ Le traitement des enregistrements avec certains champs nuls,

Le processus de préparation des données représente la partie la plus délicate du projet, car en plus d'extraire les données du système source, il doit assurer le chargement de données correctes, cohérentes, conformes et avec des performances raisonnables.

*Notre cas pratique :*

## La première Transformation avec Pentaho Data Integration

Notre première transformation, et surtout, ça va être l'occasion de découvrir l'interface de l'outil Pentaho Data Intégration. De nombreuses captures d'écrans sont prévues pour présenter les mécanismes principaux et faciliter la prise en main.

## Fenêtre de lancement de PDI

Le démarrage de PDI sous Windows se passe par un double clic sur le fichier exécutable Spoon.bat

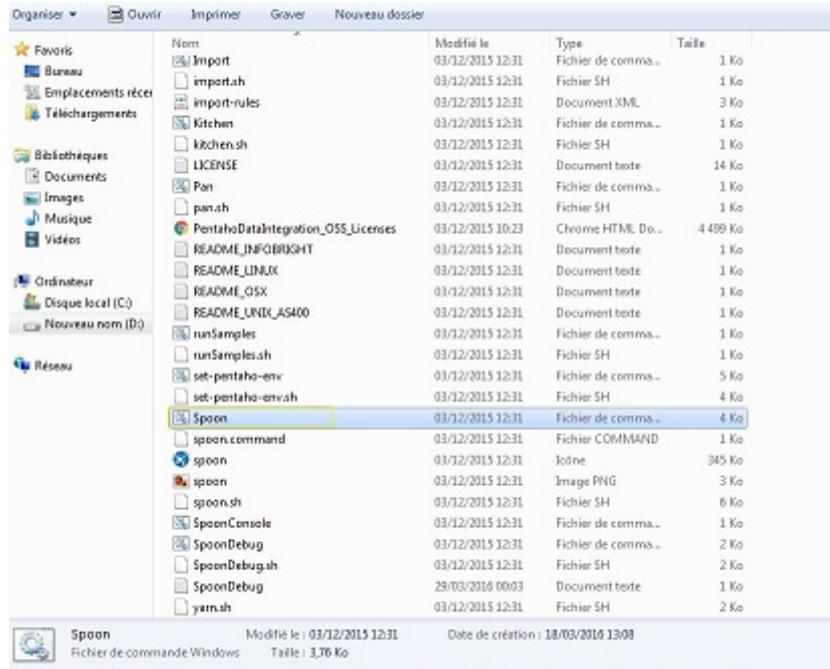


FIGURE 5.2.1 – Lancement de SPOON.BAT

## Création d'une nouvelle transformation

La création d'une nouvelle transformation s'effectue par la commande de menu Fichier -> Nouveau -> Transformation comme suit :

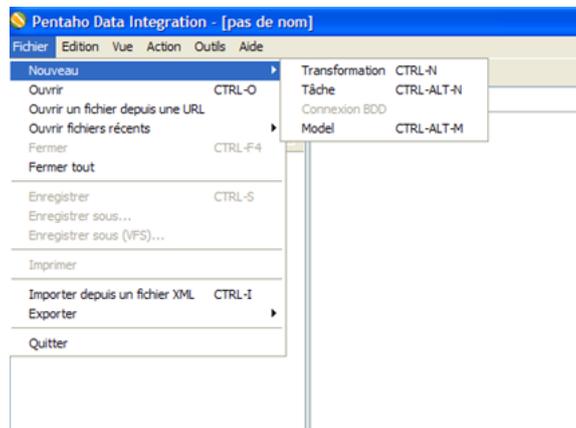


FIGURE 5.2.2 – Creation de la transformation

L'affichage suivant est obtenu :

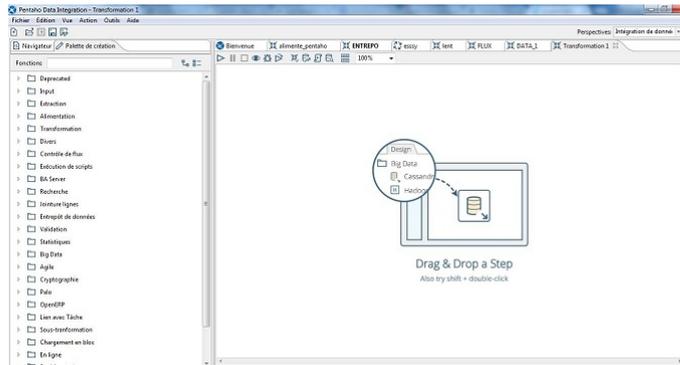


FIGURE 5.2.3 – Nouvelle transformation

Le nom par défaut de la transformation est « Transformation 1 » s’affiche en intitulé de l’onglet correspondant à la transformation en préparation (partie droite de l’écran), Sur le coté gauche de l’écran, la palette de création de PDI (lise des étapes de transformation) se présente sur un onglet. Les composants sont rangés en catégories pour faciliter la recherche.

Tout à fait à gauche de l’écran, un troisième onglet « Navigateur » présente la structure de la transformation.

Et bien sur, toutes les commandes de l’outil PDI sont aussi accessibles depuis un système de menus déroulants en haut de l’écran.

Et maintenant on va commencer le développement de notre première Transformation PDI .

## Objectif de notre transformation

Notre transformation consiste à importer des données à partir d’une Table d’une base de données MySQL, à manipuler l’information s’y trouvant et à créer une nouvelle structure en bout de ligne. Ce type de processus est fréquent lorsque nous devons importer de l’information dans un fichier (Excel, texte, CSV, etc.) à partir d’une base de données (ou n’importe quelle source de données).

### ► Création de la connexion à la base SampleData

Dans l’onglet Navigation on fait un clic droit sur le dossier « connexion » et la fenêtre suivante s’affiche :

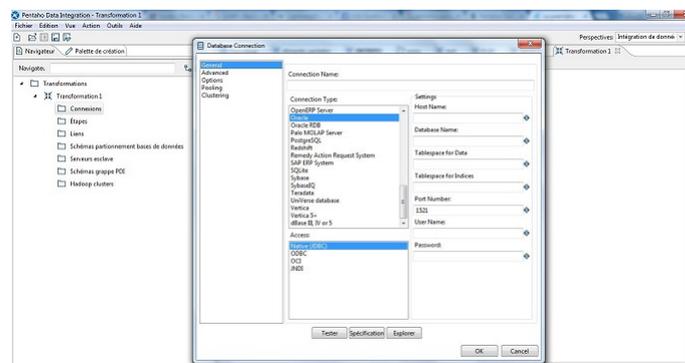


FIGURE 5.2.4 – Nouvelle connexion

Puis on saisi les paramètres de notre connexion :

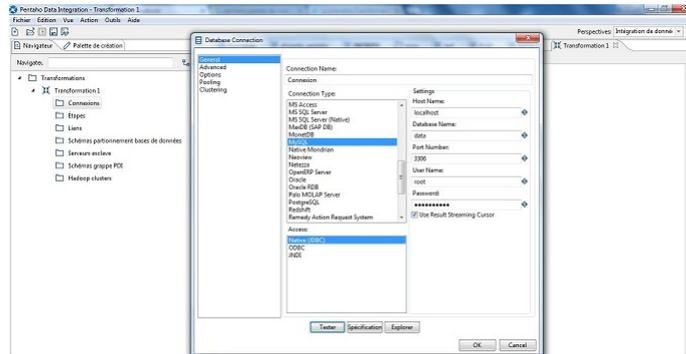


FIGURE 5.2.5 – Configuration de la connexion

### ► Création de la première step (étape de la transformation)

Dans l'onglet « Palette de création » on choisi le dossier Extraction (domaine Extraction)

Et on fait un drag-drop de l'étape « Extraction depuis table » dans l'espace design au milieu de la page :

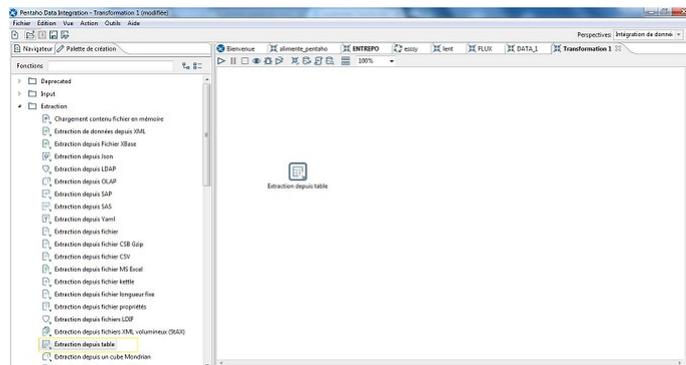


FIGURE 5.2.6 – Choix d'extraction

Puis on nomme l'étape par « Ma table d'extraction » en double cliquant sur l'étape et on choisi la base et on clic sur « Obtenir script SQL » pour choisir la table :

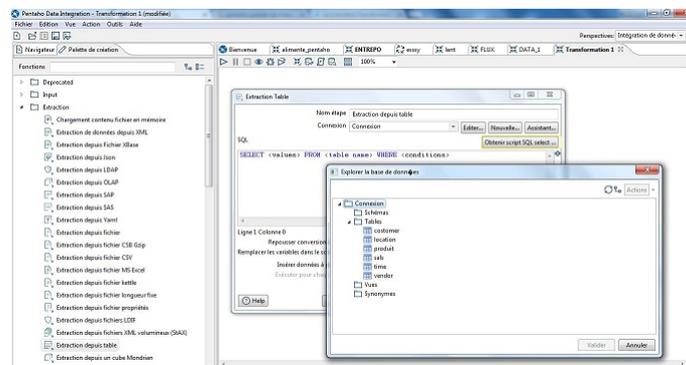


FIGURE 5.2.7 – Configuration de l'extraction

on inclus les noms des champs dans le script, et on obtient comme résultat :

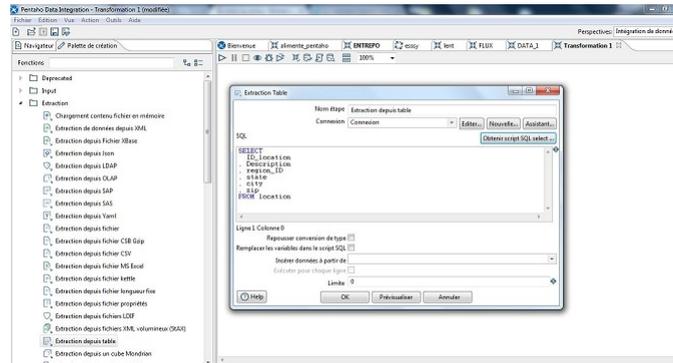


FIGURE 5.2.8 – Script

### ► Création de la Deuxième ”Step”

Dans cette étape on va réduire le flux pour ne laisser que les clients canadiens avec le step « Filtrage Lignes » :

Dans l’onglet « Palette de création » à la rubrique « Contrôle de Flux » on va glisser-déposer le step « Filtrage Lignes » vers l’espace ”design” :

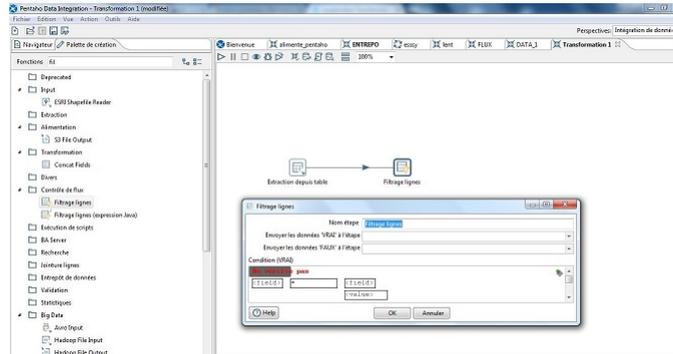


FIGURE 5.2.9 – Filtrage

Et pour l’étape qui va contenir le résultat « Vrai » :

Dans l’onglet ”Palette de création” à la rubrique Transformation on va glisser déposer la step « Altération Structure Flux» vers l’espace Design :

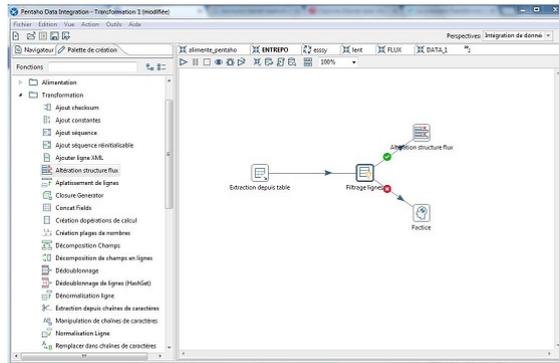


FIGURE 5.2.10 – Flux

Retournons à notre étape de filtrage, en double cliquant sur « Filtrage Ligne » on aperçoit que les champs d’envoi sont renseignés :

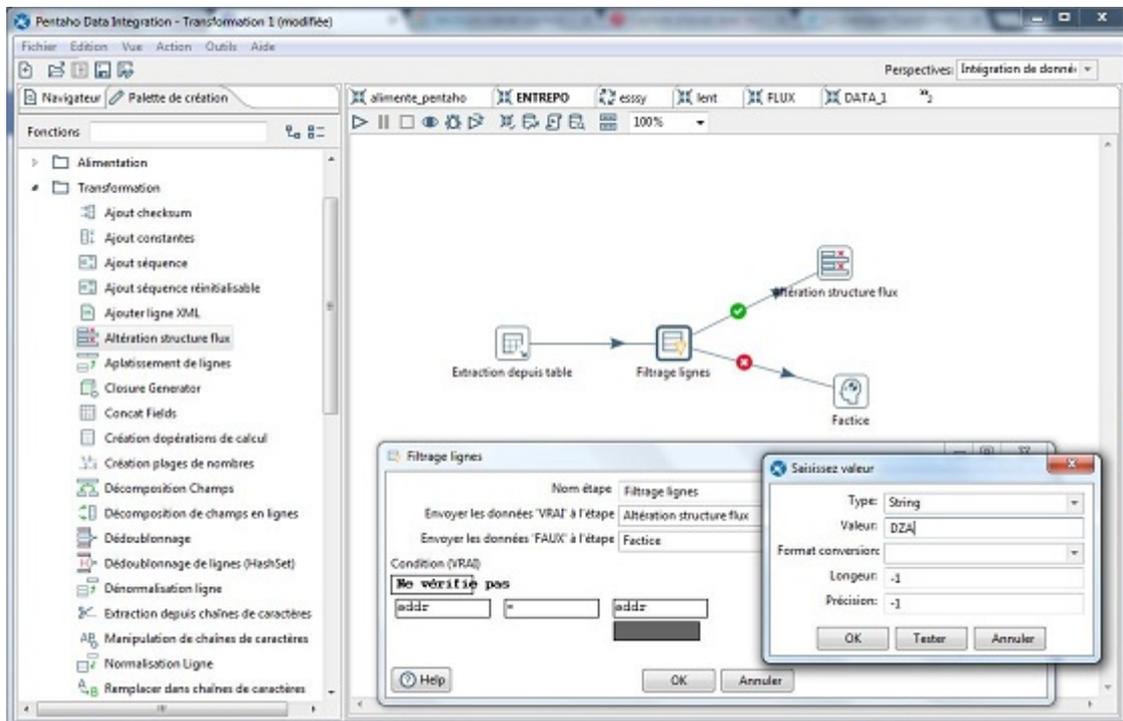


FIGURE 5.2.11 – Configuration du filtre

Puis on click sur l’étape « Alteration structure flux » pour supprimer les champs dont on n’aura pas besoin :

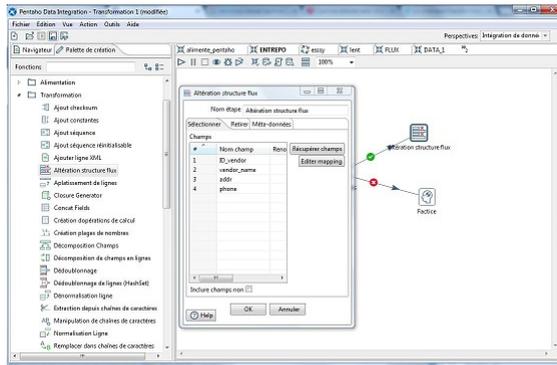


FIGURE 5.2.12 – Structure de flux

Et en ne laissant que l'information pertinente pour un référentiel des clients.

- Alimentation du fichier Excel par le référentiel des clients

Ici on va utiliser une étape appelée « Alimentation fichier Excel » dans la rubrique "Alimentation" :

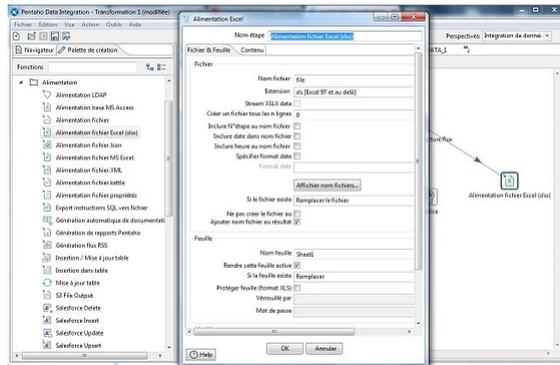


FIGURE 5.2.13 – Configuration de l'alimentation excel

Puis on sauvegarde notre travail avec un Contrôle+S, et on lance la transformation en cliquant sur le triangle bleu en haut de l'espace « design » :

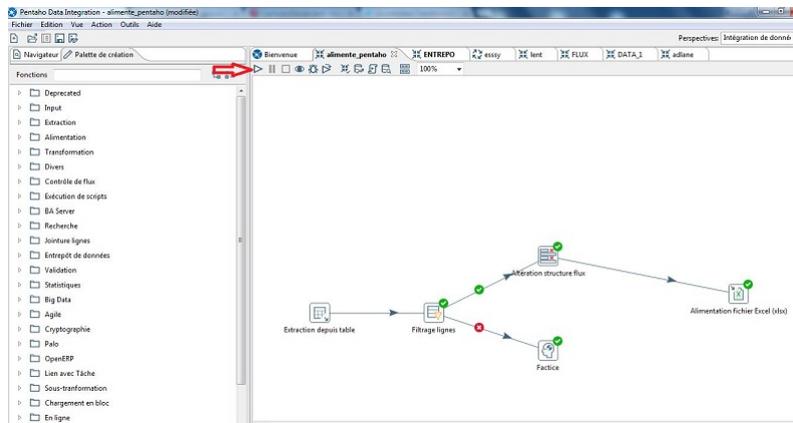


FIGURE 5.2.14 – Lancement de la transformation

On obtient la fenêtre suivante ou on peut spécifier le niveau de débogage et le type d'exécution et on clique sur le bouton "Démarrer" :

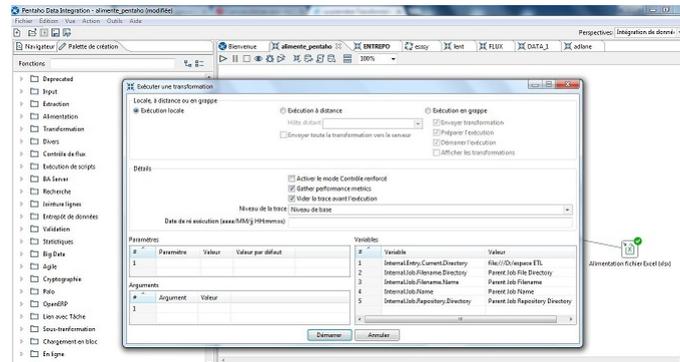


FIGURE 5.2.15 – Démarrer

Le résultat de l'exécution s'affiche en bas de la page :

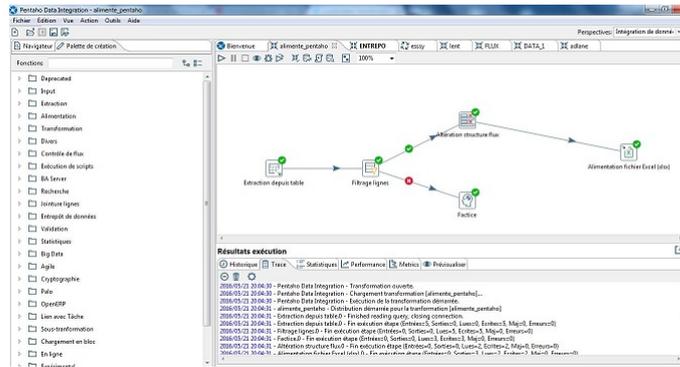


FIGURE 5.2.16 – Resultat

## 5.3 Approche adoptée

Ralph Kimball place les besoins de l'utilisateur au centre de l'univers du Datawarehouse.

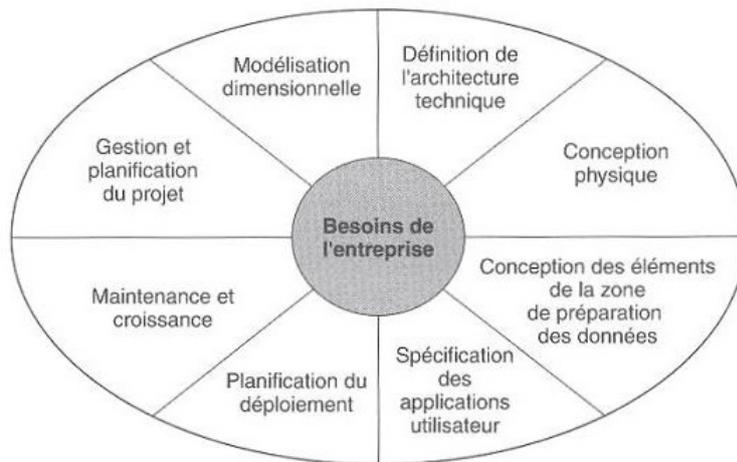


FIGURE 5.3.1 – Place et importance des besoins dans un projet décisionnel.

Afin de s’assurer de la réussite du nouveau système, nous avons adopté une approche centrée sur les utilisateurs finaux de ce système. Une approche qui s’inscrit pertinemment dans la logique de la démarche de Kimball. Elle consiste à :

► **Identifier les utilisateurs du futur système .**

► **Recenser les exigences des responsables informatiques :**

Cette étape consistait à identifier les caractéristiques techniques du nouveau système.

► **Recenser les besoins fonctionnels des futurs utilisateurs :**

Puisque la réussite de tout nouveau système se mesure par la satisfaction de ses utilisateurs, cette étape intervient pour recueillir les attentes de ces derniers à travers des interviews et des réunions.

► **Identifier les cas d’utilisation :**

Une étape qui vise à répondre à la question suivante : comment se serviront les utilisateurs, préalablement identifiés, du futur système ?

### 5.3.1 Identification des utilisateurs

La liste des utilisateurs potentiels du nouveau système. Ces utilisateurs sont exclusivement des employés de la direction générale et peuvent être rassemblés en groupes (rôles).

N°	Groupe	Direction	Position
1	Administrateur	DSI	Le DSI et les ingénieurs de support et d’organisation
2	Utilisateur métier	Directions métier et support	Certains agents de production et de sinistre
3	Cadre dirigeant	PDG, Directions métier et support	PDG, DG et les responsables des directions métier et support

TABLE 5.1 – Liste des groupes d’utilisateurs

### 5.3.2 Les besoins techniques

Le projet d'élaboration d'un système décisionnel constitue un besoin pour la **SAA**. De ce fait, la direction des systèmes d'information avait tracé certains critères que le futur système doit remplir. Ces critères sont résumés dans le tableau ci-dessous :

N°	Elément	Remarque
1	Interface web pour le client	Obligatoire
2	Interface web pour l'administration	Obligatoire
3	Serveur de base de données Oracle	Obligatoire
4	Système d'exploitation Windows	Souhaitable
5	Edition automatique des rapports	Souhaitable

TABLE 5.2 – Les exigences des responsables informatiques

Ces exigences peuvent être schématisées par la figure ci-dessous :

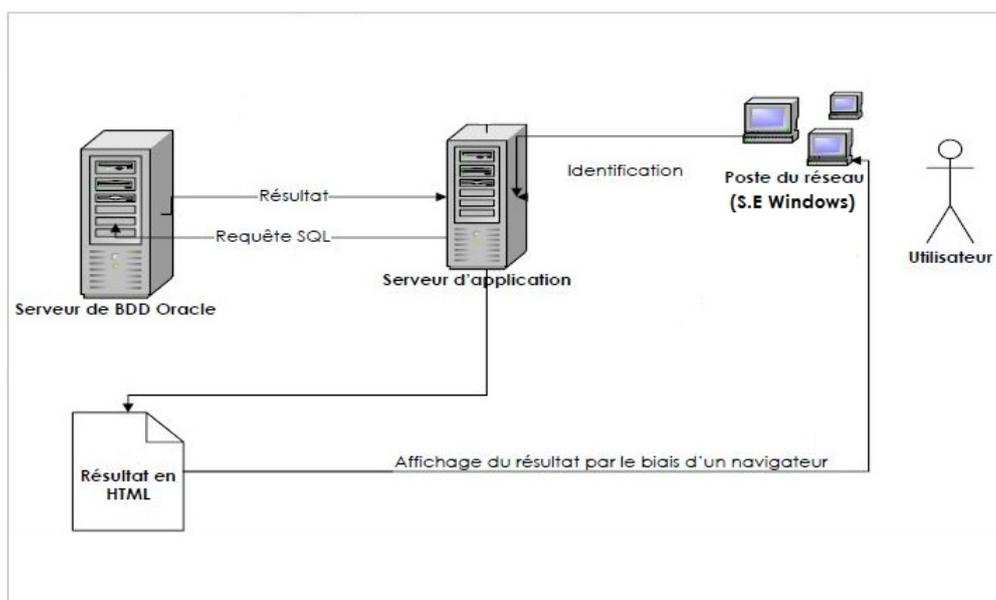


FIGURE 5.3.2 – Le futur système vu par les responsables informatiques

### 5.3.3 Les besoins fonctionnels

#### En termes de rapports

Les rapports souvent exigés par les cadres dirigeants ou les entités de tutelle se résument à des synthèses d'activité en production et en sinistre.

#### Production :

Les rapports demandés en production se résument à des synthèses des contrats souscrits selon des axes d'analyse. Ces axes d'analyse peuvent être :

- ▶ Les périodes.
- ▶ Les catégories d'assurés.
- ▶ Les agences.
- ▶ Les directions régionales.
- ▶ Les classes.
- ▶ Les branches.
- ▶ Les formules
- ▶ Les conventions.
- ▶ Les wilayas.
- ▶ Les villes...etc.

### **Sinistre :**

Les rapports demandés en sinistre se résument aussi à des synthèses des sinistres enregistrés, provisionnés, ou réglés selon des axes d'analyses. Au-delà des axes déjà mentionnés en production, d'autres axes peuvent être demandés en sinistre. On en trouve :

- ▶ Taux de sinistres réglés.
- ▶ L'âge du tiers.
- ▶ La Compagnie du tiers.
- ▶ Le type de sinistre...etc.

## **5.3.4 Identification des cas d'utilisation**

### **5.3.4.1 Les cas d'utilisation fonctionnels**

Le tableau ci-dessous présente la liste des cas d'utilisation fonctionnels.

N°	Intitulé du cas	Groupes concernés	Description
1	Se connecter	Tous les groupes	Chaque utilisateur peut se connecter au serveur d'application en saisissant un identifiant et un mot de passe valides.
2	Consulter rapport	Tous les groupes	Selon ses privilèges un utilisateur peut visualiser un rapport déjà créé.
3	Télécharger rapport	Tous les groupes	Un utilisateur peut télécharger un rapport déjà créé.
4	Ajouter rapport	Administrateur	Un utilisateur peut, s'il a le droit, interroger la base et mettre le résultat sous forme de rapport
5	Modifier rapport	Administrateur	Un administrateur peut changer certaines informations relatives à un rapport déjà existant.
6	Supprimer rapport	Administrateur	A la demande d'un responsable ou en détectant une erreur dans une requête, l'administrateur peut supprimer un rapport existant.
7	Consulter tableau de bord	Administrateur et cadre dirigeant	Un cadre dirigeant ou un administrateur peut consulter certains indicateurs clés regroupés dans un tableau de bord.
8	Ajouter indicateur	Administrateur	Un administrateur peut ajouter certains indicateurs au tableau de bord à la demande d'un responsable.
9	Supprimer indicateur	Administrateur	Un administrateur peut éventuellement et à la demande d'un responsable supprimer un indicateur du tableau de bord.

TABLE 5.3 – Liste des cas d'utilisation fonctionnels.

## Conclusion

Nous avons présenté avec détails les étapes de la conception de notre entrepôt de données, et montrer, étape par étape le processus ETL. Nous avons aussi expliqué l'approche adopté dans notre conception.

Il nous reste la dernière étape de notre chaine décisionnel, qui est la restitution des données. Cette dernière fera l'objet d'une partie de notre suivant chapitre.

## Chapitre 6

# Déploiement de la solution

Dans ce dernier chapitre, nous allons décrire la mise en place de notre solution, en présentant en détails sa réalisation et son déploiement. Pour cela, il a été nécessaire de recourir à un certain nombre d'outils et mettre en place l'environnement d'exécution. Ce chapitre contient l'environnement et l'architecture technique de la solution, les étapes de la réalisation des différentes zones de données conçues dans le chapitre de « Conception (zone d'entreposage, zone d'alimentation, les cubes OLAP, le Reporting et le tableau de bord) », pour terminer avec l'aspect sécurité de la solution. Nous avons utilisé des prises d'écran afin de figurer le travail fait et d'illustrer les différentes fonctionnalités offertes par le système.

### 6.1 Architecture technique de la solution

Le système mis place peut être schématisé comme le montre la figure ci-dessous :

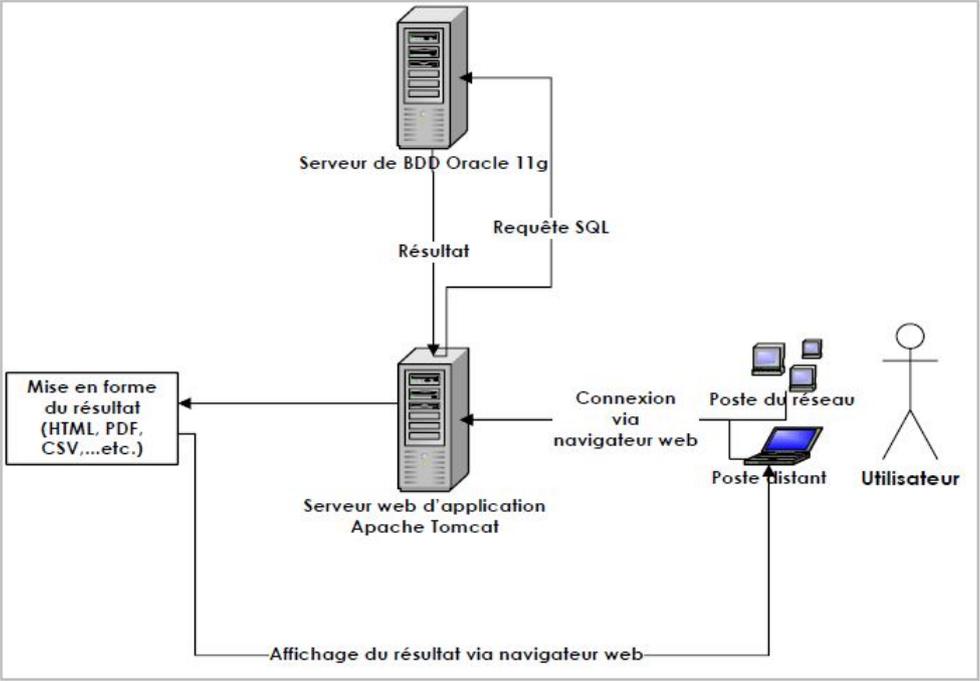


FIGURE 6.1.1 – Architecture du système mis en place.

## 6.2 Les outils utilisés

Oracle Data base  
Express Edition

Comme la base de données centralisée de la société est sous Oracle 11g et ainsi la disposition de la License Oracle. Donc après avoir leur accord, nous avons utilisé les outils présentés dans le tableau ci-dessous.

Editeur	Editeur	Version	Fonction
		11g Version 2	Hébergement des bases de données (SGBD)
	Pentaho data intégration (PDI)	6.0.1	Extraction, Transformation et Chargement
	BI server - Pentaho suite	5.0.1	Construction et chargement des cubes
	Pentaho User Console	6.0.1	Outil d'analyse et de Reporting

TABLE 6.1 – Tableau des outils

## 6.2.1 Serveur de base de données

Le serveur de base de données utilisé est un serveur Oracle 11g R2.

## 6.2.2 Pentaho BI suite

Pentaho BI suite est une suite décisionnelle open source contenant un ETL, des outils d'analyse et de reporting. La suite est éditée sous une licence GNU General Public License Version 2, avec d'autres parties sous LGPL V2, ou The Common Public License et The Mozilla Public License.

La version utilisée dans ce projet est la version communautaire (Pentaho Community Edition) dans sa version 4.8.

### 6.2.2.1 Le serveur d'application Pentaho

La suite Pentaho intègre un serveur web Apache Tomcat 6.0 Servlet. Ce dernier interagit avec l'ensemble des clients Pentaho pour afficher les différents rapports, tableaux de bord, cubes, . . . etc. à l'aide d'un navigateur web récent.

Le serveur Pentaho offre deux consoles. La première console est pour les utilisateurs et la deuxième est réservée aux administrateurs.

Ci-dessous la fenêtre de connexion du serveur Pentaho. Chaque utilisateur est identifié par un nom unique et un mot de passe. En saisissant un login et un password valides, l'utilisateur pourra se connecter au serveur d'application Pentaho.



FIGURE 6.2.1 – Interface d'identification du serveur Pentaho.

Après s'être connecté, l'utilisateur est redirigé vers la console d'utilisateur du serveur Pentaho. L'utilisateur pourra à présent visualiser les rapports auxquels il a accès, consulter certains indicateurs du tableau de bord ou bien gérer des cubes de donnée.

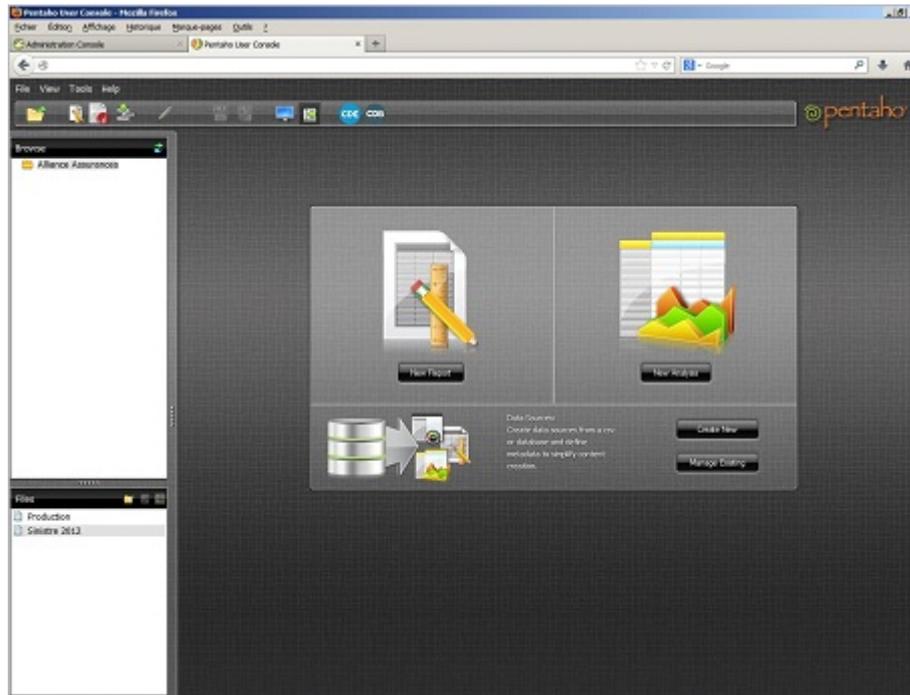


FIGURE 6.2.2 – Console d'utilisateur du serveur Pentaho

La suite Pentaho BI offre aux administrateurs une interface web interactive afin de gérer le serveur BI. Cette interface permet de :

- ▶ Ajouter/Désactiver/Supprimer des utilisateurs.
- ▶ Ajouter/Supprimer des privilèges à un utilisateur.
- ▶ Ajouter/Modifier/Désactiver/Supprimer des rôles (groupes d'utilisateurs).
- ▶ Ajouter/Supprimer des privilèges à un groupe d'utilisateurs.
- ▶ Vérifier l'état du serveur (connecté/déconnecté).
- ▶ Détecter les erreurs en cas de dysfonctionnement du serveur.

Ci-dessous une capture d'écran de l'interface web d'administration du serveur Pentaho :

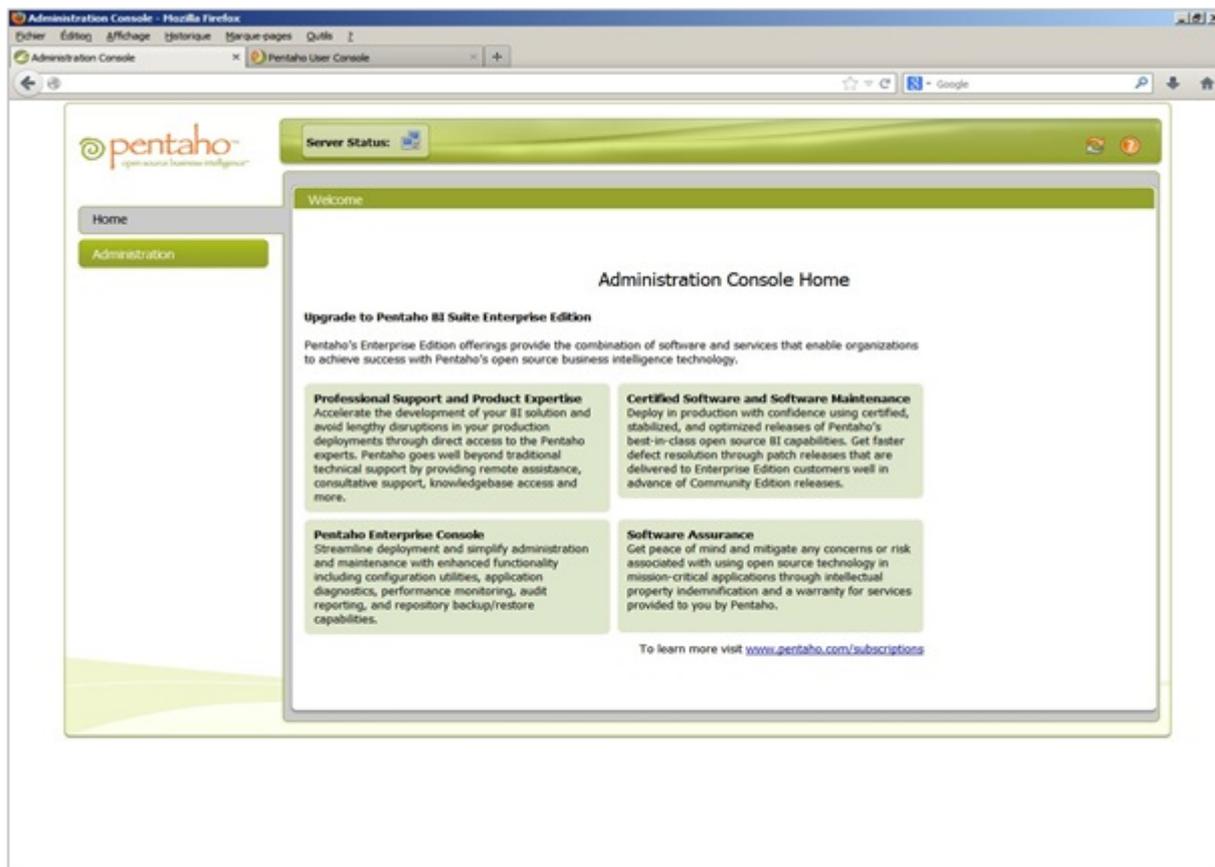


FIGURE 6.2.3 – Console d’administration du serveur Pentaho.

## Les prérequis de déploiement :

### ► Machine client :

Pour qu’un utilisateur puisse accéder aux tableaux de bord et visualiser les rapports, il doit disposer d’un navigateur web.

### ► Machine serveur :

Pour déployer la solution, l’environnement suivant doit être installé : Oracle 11g et Pentaho Enterprise Edition.

## 6.3 Présentation de l’architecture de la solution

La figure suivante illustre l’architecture détaillée de notre Data Warehouse. Cette architecture comprend les différents composants du système décisionnel à savoir l’entrepôt de données, l’ETL et le portail de restitution.

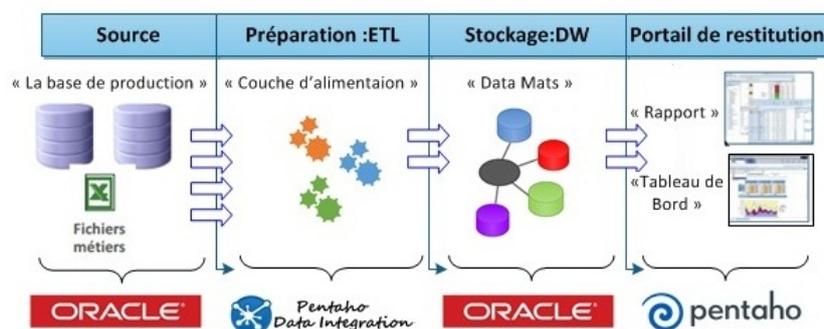


FIGURE 6.3.1 – Architecture globale de la solution.

## 6.4 Etapes de la réalisation de la solution

### 6.4.1 La réalisation de la zone d’entreposage

Il s’agit de l’implémentation des DataMarts constituant le DW. Pour les implémenter physiquement, nous avons utilisé l’outil ROLAP qui permet de gérer la structure multidimensionnelle dans un SGDB relationnel. Nous avons créé une base de données relationnelle conforme aux schémas conçus dans la partie « conception de la zone d’entreposage » à l’aide de Oracle 11g(le SGBD utilisé par la SAA). Dans cette BDD, chaque fait du schéma conceptuel correspond à une table appelée table de fait et chaque dimension correspond à une table appelée table de dimension. La table de dimension contient une clé primaire (clé de substitution) et un ensemble d’attributs. Tandis que la table de fait possède comme attributs les mesures d’activités et les clés étrangères vers les tables de dimension.

### 6.4.2 La réalisation de la zone d’alimentation

Une fois la base de données cible prête, nous pouvons procéder au chargement des données. Le chargement de la dimension période se fait à l’aide d’une fonction qui à chaque nouvelle année insère les mois, les trimestres et l’année.

Comme on a déjà mentionné, nous avons utilisé l’outil Pentaho Data Intégration qui est aussi un composant Pentaho BI server, pour la réalisation des différents programmes ETL. Cet outil présente de grandes capacités de traitement des gros volumes de données, il offre aussi plusieurs fonctionnalités telles que la reprise sur incident, la planification et la synchronisation des différents programmes ETL ainsi que la signalisation des erreurs.

Le tableau ci-dessous présente les fonctionnalités de Pentaho Data Intégration que nous avons utilisé dans le processus ETL.

TABLE 6.2 – Tableau des transformations utilisées

Le tableau ci-dessous présente les fonctionnalités de Pentaho Data Intégration que nous avons utilisé dans le processus ETL.

Composant	Description
	Lire des données à partir d'une table dans la bdd
	Insertion ou mettre a jour les lignes dans une table
	Filterer les lignes grâce à de simples équations
	Sélectionner ou retirer des champs dans une ligne, Possibilité de modifier le type, longueur et précision du champ.
	Ce type d'etape ne fait rien.Elle est cependant pratique lorsque vous desirez tester des transformation
	Planification des taches
	Verification d'une connexions
	Execution d'une tansformation
	Remplacer la valeur du champ par une valeur par défaut si elle est nulle.
	Générer un nombre de linges
	Créer de nouveaux champs a partir calculs simples
	Plugiciel permettant decrire des sripte java
	Ajouter une ou plusieurs constantes aux flux d'entree
	Recherche valeurs provenants d'un flux dans une une transformation
	Tri de lignes suivant certains chapms (ascendant ou descandant )

Table 1 Composant

Maintenant nous pouvons procéder au chargement des données. On commence par le chargement de la dimension Temps qui se fait à l'aide d'une fonction qui charge cette table avec une période de temps.

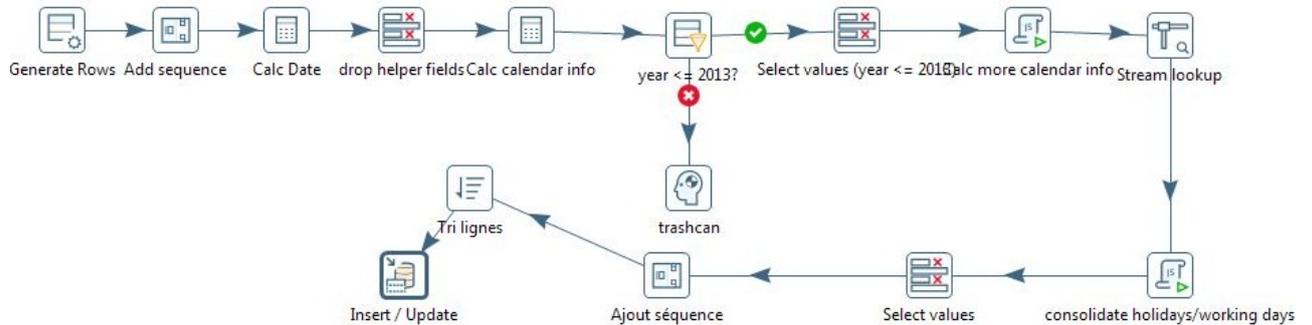


FIGURE 6.4.1 – Création et chargement de la table temps

Les autres données sont extraites à partir des tables de la base de données , et ceci après avoir eu recours aux transformations nécessaires.

## 6.5 La réalisation du Reporting

Pour terminer avec ce projet, la dernière étape consistait à créer des tableaux de bord récapitulatifs des différents indicateurs présentés dans la partie « conception du tableau de bord ». Pour cela, nous avons utilisé Pentaho, qui offre la possibilité de créer, modifier et enregistrer des tableaux de bord.

## 6.6 Présentation du prototype de l'application

Dans ce qui suit nous allons montrer des captures d'écran un scénario d'utilisation simple qui décrit ce que peut faire l'utilisateur actuellement avec notre solution.

### Authentification :

Premièrement l'utilisateur entre son nom d'utilisateur et mot de passe.



FIGURE 6.6.1 – Authentication

- **Accès à l'espace personnel**

Cet utilisateur (admin) possède des droits d'accès aux cubes créés, il trouve des cubes de données correspond à la définition d'un schéma de faits.

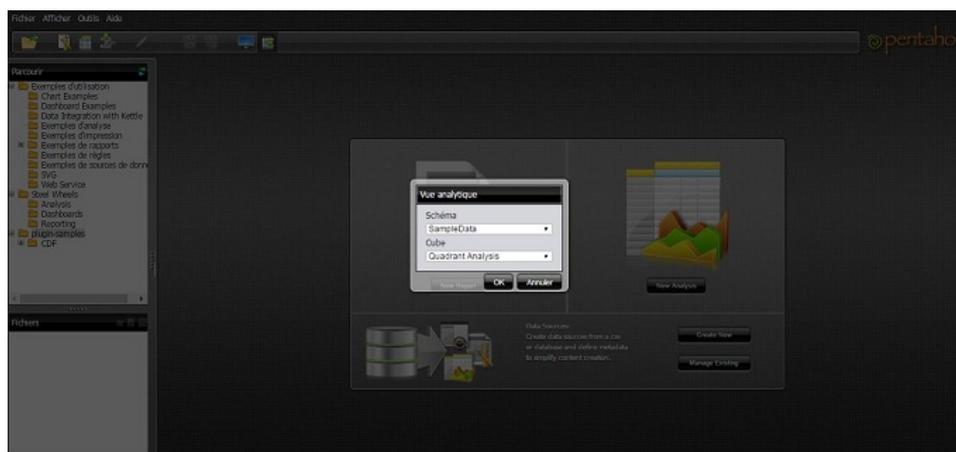


FIGURE 6.6.2 – Sélection des sources de données.

- **Création d'un nouveau rapport**

Après la sélection de Data Mart dont on veut naviguer , un tableau croisé dynamique s'affichera avec les mesures déjà définies, et les axes d'analyse, Puisque l'utilisateur interroge un cube OLAP il peut choisir l'axe qu'il veut.

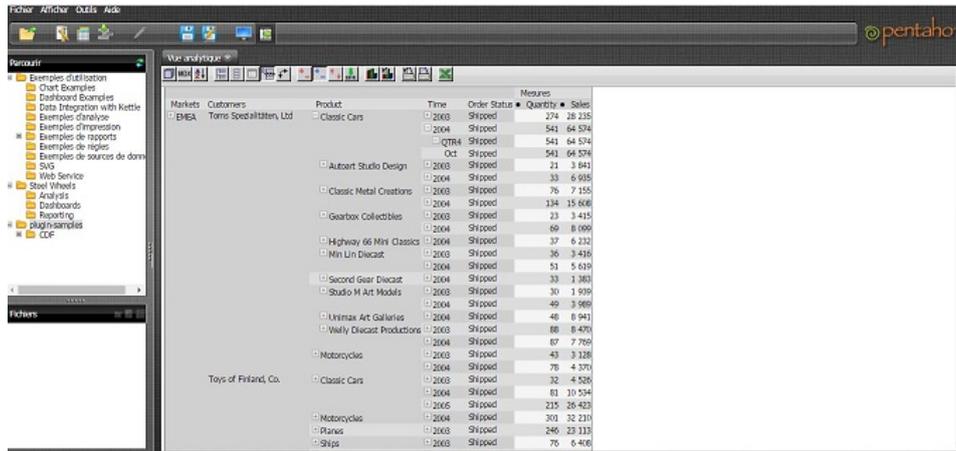


FIGURE 6.6.3 – Rapport sous format d'un tableau croisé

- **Résultat de du rapport en graphique**

L'utilisateur veut voir l'état en utilisant une représentation graphique, il n'a qu'à choisir le type de présentation (courbe, barres, etc.), il l'obtiendra facilement en spécifiant les axes (X\_axe :agence\_code et Série nombre de clients) sur lesquels il veut afficher les résultats.

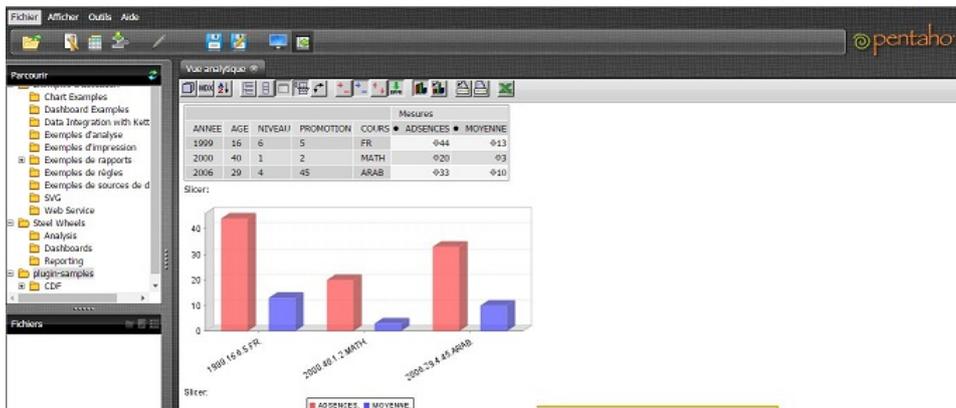


FIGURE 6.6.4 – Rapport sous format d'un graphe.

- **Exporter le rapport vers une autre Format**

L'utilisateur peut exporter le rapport créé vers un format choisi comme PDF or EXCEL pour l'imprimer.

- **Enregistrement de l'état dans un tableau de bord personnel**

L'utilisateur qui a les privilèges, a la possibilité de sauvegarder le résultat de sa requête dans un tableau de bord personnel, il peut créer plusieurs tableaux de bord, et il peut choisir de partager certains avec d'autres utilisateurs.

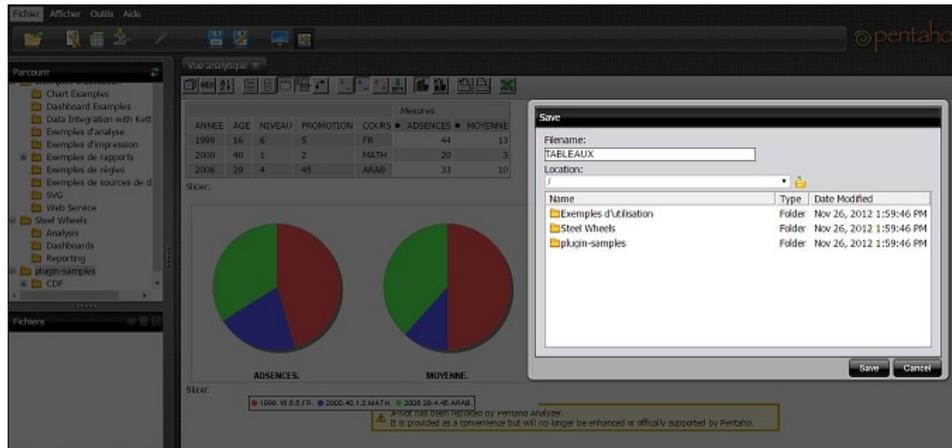


FIGURE 6.6.5 – Capture d'écran du tableau de bord.

## 6.7 La sécurité du système

Au niveau de l'application utilisateur, nous avons utilisé la stratégie de sécurité disponible avec Pentaho BI Server qui est basée sur le mécanisme d'autorisation et d'authentification permettant d'affecter des droits d'accès pour chaque utilisateur. Au niveau de la base de données qui est notre data warehouse, nous utilisons la gestion les mécanismes offert par le SGBD Oracle 11 g, ce dernier est paramétré pour autoriser ou interdire les accès et la manipulation des données du DW.

## Conclusion

Dans cette partie nous avons décrit l'architecture technique qui va recevoir notre solution ainsi que l'environnement de développment qui ont été pris en compte lors de la phase de réalisation du DW. Nous avons aussi présenté tous les outils que nous avons utilisés et qui sont Oracle 11g et Pentaho BI open source qui nous a permis la collecte, l'intégration et l'organisation des données issues de différentes sources et la restitution des informations résultantes de l'analyse sous différentes formes : rapports et tableaux de bord.

# Conclusion et perspectives

De nos jours, l'information devient la principale matière première de l'économie et constitue un instrument de compétition, c'est pourquoi les entreprises sont conscientes que la maîtrise, le contrôle et la diffusion de l'information doivent être utilisées. C'est pour ce la que leurs services doivent se doter d'outils analytiques sophistiqués qui leurs permettent de prendre des décisions éclairés en disposant d'informations nécessaires.

Afin de préserver sa place dans un marché extrêmement concurrentiel, la SAA a initié le projet de mise en place d'un système d'information décisionnel qui a fait l'objet de notre projet de fin d'études.

Tout au long de notre travail de conception et de réalisation, nous avons essayé de suivre une démarche mixte, alliant de ce fait entre deux approches connues dans le domaine du Data Warehousing, à savoir l'approche « Besoins d'analyse » et l'approche « Sources de données ». Cette démarche a permis de répondre aux attentes et besoins des utilisateurs tout en exploitant au mieux les données générées par les systèmes opérationnels de manière à anticiper sur des besoins non exprimés.

Afin de collecter les besoins auprès des utilisateurs finaux, et ce grâce aux questionnaires ainsi que la documentation existante, et d'anticiper ceux non exprimés par les utilisateurs en faisant des réunions de travail avec notre promotrice afin comprendre et d'exploiter les données des systèmes de production. Nous avons pu ressortir les différents sujets d'analyse .

Nous avons ensuite conçu la zone d'entreposage, la zone de préparation des données. Pour représenter l'ensemble des schémas constituant la zone d'entreposage, nous avons suivi la modélisation dimensionnelle qui offre une facilité de compréhension et une simplicité d'alimentation.

Une fois les modèles logiques définis, nous avons passé à l'alimentation de la zone de stockage, ce qui nous a consommé beaucoup plus de temps (extraction de l'ensemble des données, vérification de leur qualité, apport de transformations nécessaires et chargement dans la zone de stockage) et ce par le biais d'outil ETL. Une fois l'alimentation terminée, nous sommes passés au chargement de nos DataMarts.

Afin de permettre à l'utilisateur final d'exploiter les données stockées dans le DW via n'importe quel client léger en utilisant un navigateur web, nous avons mis en place une plateforme BI qui permet de :

- ▶ Diffuser l'ensemble des rapports prédéfinis établis à l'aide d'outils de Reporting, ce qui permet de réduire le temps de leur élaboration de plusieurs jours à quelques minutes.
- ▶ Faciliter le pilotage des activités marketing et de disposer d'un ensemble d'informations fiables en temps voulu pour prendre de bonnes décisions.
- ▶ Gérer les rôles et privilèges pour chaque utilisateur.

Enfin nous avons présenté les différentes étapes de la réalisation, l'architecture de déploiement de notre solution ainsi que les différentes technologies utilisées.

Pour finir, et avant de citer les perspectives du projet, nous pouvons dire que ce stage au niveau de la SAA, nous a permis d'acquérir une très bonne expérience professionnelle et d'évoluer dans un domaine qui nous était totalement inconnu à savoir le domaine des systèmes décisionnels, et au sein d'un environnement regroupant des équipes de professionnels.

Comme un projet Data Warehouse n'est jamais complètement terminé, nous pouvons citer les perspectives et les développements suivants :

- ▶ Suivre le déploiement actuel et recueillir les correctifs et remarques des utilisateurs.
- ▶ Utiliser des méthodes et algorithmes de Data Mining pour une meilleure exploitation des données.
- ▶ Optimiser le temps de réponse des requêtes en mettant en place des indexes. Actuellement, le temps de réponse est bon, mais la taille des tables de fait augmente, pour cela, il faut mettre en place une stratégie d'indexation.
- ▶ Continuer le développement de portail de restitution (ajouter d'autres rapports et d'autres tableaux de bord).

# Bibliographie

- [Poletto, 2012] Poletto Maxime, L'informatique décisionnelle, Thèse professionnelle à l'Ecole Supérieure d'Informatique CESI EXIA, 2012.
- [Trewartha, al, 1982] Trewartha, Newport, M.G., « Management », Dallas and Business Publication, 3rd Edition, 1982
- [Glauser ,2008] Glauser Claude-Alain, Senior Consultant Business Intelligence
- [Kimball, 1997] : R. Kimball, «Entrepôts de Données», Intl Thomson Pub., 1997.
- [Kimball, 2002] : R. Kimball et M. Ross; « Entrepôts de Données : Guide Pratique de Modélisation Dimensionnelle 2ème édition »; Vuibert 2002.
- [Kimball, 2005] Ralph Kimball, « Le Data Warehouse, Guide de conduite de projet », 2nd Edition, Eyrolles, 2005.
- [Inmon, 2005] William H. Inmon, « building the data warehouse », Wiley publishing, 2005.
- [Inmon, 2008] William H. Inmon « DW 2.0 The architecture for the next generation of data warehousing », Elsevier Press, 2008.
- [Franco, 1997] J.Franco, « Le Data Warehouse, Le Data Mining », 1st Edition, Eyrolles, 1997.
- [Kimball ,1996] [Kimball, 1996] : R. Kimball; « Entrepôts de données : Guide pratique du concepteur de Data Warehouse »; Wiley Computer Publishing 1996.
- [Bouquin, 2003] [Bouquin, 2003] : Bouquin Henry ; « Le contrôle de gestion »; P.U.F; 2003.
- [Yeatman, 1998] Yeatman Jérôme, Manuel International de l'assurance, Economica, Paris, 1998.
- [CNA, 2012] Conseil National des assurances, Bulletin des assurances N°21 4ème trimestre 2012, Alger, 2012.
- [Lezoul, 2011] Lezoul Mohamed, La situation actuelle du secteur des assurances en Algérie : Quelles sont les alternatives?, Sétif, 2011.
- [Fernandez, 2003] Fernandez Alain, Les nouveaux tableaux de bord des managers le projet décisionnel dans sa totalité, Editions d'Organisation, Paris, 2003.
- [DI] Documents Internes.