

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

Université Saad Dahlab, Blida  
USDB.

Faculté des sciences.  
Département informatique.

Mémoire pour l'obtention  
d'un diplôme d'ingénieur d'état en informatique.  
Option Système d'information

Sujet :

**REALISATION D'UN SYSTEME  
DECISIONNEL POUR L'INTERROGATION  
DES DONNEES RESSOURCES HUMAINES**

Présenté par : Seddar yagoub samia  
Benaouda rima

Promoteur : M.Bala  
Encadreur : K.Haddou

Organisme d'accueil : NAFTAL



Soutenu le: date soutenance, devant le jury composé de :

Nom. President du jury, grade, organisme

Nom examinateur 1. grade, organisme

Nom examinateur 2. grade, organisme

Président

Examinateur

Examinateur

MIG-004-165-1

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

سبحانك لا علم لنا الا ما علمتنا



## *Dédicaces*

*Je dédie ce modeste mémoire à :*

- *Mes chères parents, que dieu me les protège longtemps en vie, et qui veillent pour ma réussite.*
- *Mes très chères sœurs ASSIA, RADIA*
- *Mes très chers frères : YACINE, ABDELKRIM*
- *Toute ma famille maternelle FECIH.*
- *Toute ma famille paternelle SEDDAR YAGOUB*
- *Tous mes camarades en informatique*
- *Tous mes amis et tous ceux que je connais*
- *Tous mes enseignant du primaire à l'université,*
- *Mon binôme RIMA.*



*SAMIA*



## *Dédicace*

*Je dédie ce modeste mémoire à :*

- *Mes chères parents, que dieu me les protège longtemps en vie, et qui veillent pour ma réussite.*
- *Mes chères sœurs Amel, Farida, Zahra et Selma.*
- *Toute ma famille paternelle Benaouda.*
- *Toute ma famille maternelle Soualah.*
- *Tous mes camarades en informatique*
- *Tous mes amis et tous ceux que je connais.*
- *Tous mes enseignant du primaire à l'université.*
- *Mon binôme SAMIA.*



## *RJMA*



# Remerciements

Tout d'abord nous remercions sans cesse le bon dieu tout puissant, clément et miséricordieux qui nous a donné la volonté, la force et la patience pour suivre nos études.

Et en se soumettant au Hadith de notre prophète Mohamed صلى الله عليه وسلم "من لا يشكر الناس لا يشكر الله"

Du terme de cette étude nous tenons à exprimer notre gratitude à :

Mr BOUKACEM, Mr HOUCHATI et Mr M.YOUSFI et à tout le personnel qui nous permis d'effectuer la partie pratique ainsi que pour la documentation et les encouragements.

A notre promoteur Mr M.BALA pour les conseils qu'il n'a cessé de prodiguer, et d'avoir garde un œil attentif sur le déroulement et l'avancement du projet en ayant toujours des remarques très constructives.

Nous tenons à exprimer toute notre reconnaissance à notre encadreur Mr KHADDOU pour nous avoir proposé ce sujet, et pour la confiance qu'il nous a accordé pour la réalisation de ce projet.

Notre reconnaissance va particulièrement à, Mr B.MEKAOUI, Mr A.Zait, Mr H.SADAR, Mr B.BELAMINE, Mme Ryma BOUCHELIF Pour leur aide durant notre projet

à tous ceux qui ont contribué de près ou de loin pour la réalisation de ce travail. et à tous ceux qui nous ont aidé dans l'entreprise.

## Résumé :

Les entrepôt de données sont utilisés pour l'exploitation et l'analyse de grands volumes de données extraits des systèmes d'information en exploitation. Dans un entrepôt de données, les données sont : sélectionnées et préparées (pour répondre aux questions vitales de l'entreprise intégrée (à partir des différentes sources de renseignements) et datées (elle gardent la trace de leur origine).

Un modèle multidimensionnel organise l'entrepôt de données selon plusieurs axes d'analyse appelée <<dimensions>>.

Les systèmes OLAP (On Line analytical Processing) permettent une exploitation interactive des données contenues dans un entrepôt.

Un entrepôt de données est collection de données historiques exploitées par les applications d'aide à la décision. Il centralise des données d'intérêt pour un groupe d'utilisateurs afin de rendre leur accès rapide, peu coûteux et efficaces. Dans cette thèse, nous nous intéressons à la conception et déploiement d'un entrepôt de données métier, concernant la fonction Ressource Humaine de l'entreprise NAFTAL SPA.

L'approche retenue pour la réalisation du système est l'approche de cycle de vie dimensionnel par Ralph Kimball. Notre approche est divisée en trois phases :

Identification des besoins, Conception et Déploiement du système.

## Abstract

The data warehouses is used for the exploitation and analyzes great volumes of data, which are extracted from the information systems in exploitation. In a data warehouse, the data are : selected and prepared (to answer the vital questions of the company), integrated (starting from the various sources of information) and dated (they keep the trace of their origin).

A multidimensional model organizes the data warehouse according to several axes of analysis called "dimensions ". Systems OLAP (Online Analytical Processing) allow an interactive exploration of the data contained in a data warehouse. A data warehouse is a collection of historical data exploited by decision-support applications. It centralizes data of interest for a group of users so that access is fast, inexpensive and effective. In this work, we are interested in the conception and the spreading of a data profession warehouse, concerning the Human Resource function of the Naftal enterprise. The approach kept for the realization of the system is the approach of dimensional life cycle proposed by Ralph Kimbal. Our approach is divided into the three phases : Identification of, needs, Conception and in short the Spreading

# Sommaire

## Introduction Générale

I. Contexte général.....	-I-
II. Problématique.....	-II-
III. Objectifs.....	-III-
IV. Organisation du mémoire.....	-IV-

## Partie I : Synthèse sur les entrepôts de données

I. Préambule .....	1
II. Historique .....	1
II.1. L'Info Centre.....	2
II.2. Les entrepôts de données .....	3
II.3. Infocentre Vs Data Warehouse .....	4
III. Le décisionnel .....	5
III.1. Définition du système d'information décisionnel .....	5
III.2. Evolution des systèmes décisionnels.....	7
IV. Les concepts de base d'un Data Warehouse .....	7
IV.1. Définitions du Data Warehouse .....	7
IV.2. Objectifs du Data Warehouse.....	9
IV.3. La structure du Data Warehouse.....	10
IV.4. L'architecture du Data Warehouse.....	13
IV.5. L'implémentation du Data Warehouse.....	15
IV.5.1. Implémentation selon l'architecture réelle.....	15
IV.5.2. Implémentation selon l'architecture virtuelle.....	16
IV.5.3. Implémentation selon l'architecture remote.....	16
V. Le concept OLAP .....	17
V.1. Présentation .....	17
V.2. Les différents outils OLAP .....	17
V.2.1. Les outils MOLAP .....	17
V.2.2. Les outils ROLAP .....	18
V.2.3. Les outils HOLAP (Hybride OLAP) .....	19
V.3. Système OLTP Vs système OLAP .....	19
VI. Organisation des données d'un Data Warehouse .....	20
VI.1. Modélisation multidimensionnelle .....	20
VI.2. Eléments de modélisation multidimensionnelle.....	20
VI.3. Schémas de Modélisation dimensionnelles .....	22
VII. La construction de l'entrepôt de données.....	24
VII.1. Etude préalable et définition des besoins .....	25
VII.2. Conception du modèle de données .....	25
VII.3. La mise en œuvre de l'architecture .....	27
VII.3.1. Construction de l'entrepôt .....	27
VII.3.2. Construction des cubes OLAP .....	27
VII.3.3. Le processus d'alimentation .....	27
VII.3.3.1. Qu'est ce qu'un ETL ? .....	28
VII.3.3.2. Système d'alimentation / ETL .....	28

VIII.	Navigation dans les données .....	30
VIII.1.	Drill Down & Drill Up .....	30
VIII.2.	Slice & Dice .....	30
VIII.3.	Data Surfing .....	31
IX.	Conclusion.....	31

## Partie II : CONSTRUCTION DU DATA WAREHOUSE DE L'ENTREPRISE

### NAFTAL SPA

#### Chapitre I : Présentation générale de l'organisme d'accueil

1.	Présentation générale de NAFTAL Spa.....	32
1.1	Historique.....	32
1.2.	Missions de NAFTAL SPA .....	32
1.3.	Organisation de l'entreprise .....	33
	Conclusion .....	36

#### Chapitre II : Identification et analyse des besoins

1.	Introduction .....	37
2.	Présentation des approches .....	37
2.1.	Première approche « le questionnaire » .....	37
2.2.	Deuxième approche « l'entretien » .....	38
2.3.	La documentation .....	38
2.4.	Obstacles rencontrés lors de l'enquête .....	38
3.	Recueil récapitulatif .....	40
4.	Conclusion .....	42

#### Chapitre III : Conception du système

1.	Modélisation dimensionnelle .....	43
1.1.	Introduction .....	43
1.2.	Processus de modélisation .....	43
1.3.	Formalisme utilisé .....	44
1.4.	Modélisation dimensionnelle des activités.....	45
A.	Modélisation dimensionnel de l'aspect « effectif » .....	45
B.	Modélisation dimensionnelle de l'aspect « absentéisme » .....	51
C.	Modélisation dimensionnelle de l'aspect « conge » .....	54
D.	Modélisation dimensionnelle de l'aspect « mission » .....	55
2.	Mise en œuvre de l'architecture .....	57
2.1.	Présentation .....	57
2.2.	Architecture détaillée du Data Warehouse.....	57
2.3.	Construction de l'entrepôt de données .....	58
2.4.	Construction des cubes OLAP .....	63
2.5.	Construction de la zone d'alimentation .....	69
2.5.1.	La zone d'alimentation .....	69
2.5.2.	Système source .....	69
2.5.3.	La périodicité de l'alimentation .....	70



2.5.4. Gestion et contrôle de l'alimentation .....	70
2.6. Construction du portail de restitution .....	71
3. Conclusion .....	72
 <b>Chapitre IV: Déploiement du système</b>	
1. Disposition physique du système .....	73
2. Les utilisateurs du système .....	74
2.1. Les composants serveurs .....	74
2.2. Les composants clients .....	75
3. Sécurité du système .....	76
3.1. Sécurité physique .....	76
3.2. Sécurité logique .....	77
4. Conclusion .....	78
<b>Conclusion générale</b>	80
<b>LEXIQUE</b>	82
<b>Annexe A</b>	89
<b>Annexe B</b>	90
<b>Annexe C</b>	92
<b>Annexe D</b>	
<b>Bibliographie et webographie</b>	

## LISTE DES FIGURES

<i>N°</i>	<i>Désignation</i>	<i>Page</i>
<b>Partie I : Synthèse sur les entrepôts de données</b>		
Fig 1	L'approche Infocentre	2
Fig 2	L'approche Entrepôt de données	3
Fig. 3	Définition de l'entrepôt de données	7
Fig. 4	L'entrepôt de données. point focal de l'informatique décisionnelle	10
Fig 5	Structure d'un entrepôt de données	11
Fig 6	La structure pyramidale de l'entrepôt de données	12
Fig 7	L'architecture générale de l'entrepôt de données	13
Fig 8	Architecture d'un produit MOLAP	18
Fig 9	Architecture d'un produit ROLAP	18
Fig 10	Schéma d'un cube	20
Fig 11	Exemple de fait	21
Fig 12	Exemple de dimensions	21
Fig 13	Exemple d'un schéma en étoile	22
Fig 14	Exemple d'un schéma en flocon	23
Fig 15	Exemple d'un schéma en constellation	24
Fig 16	Système d'alimentation ETL	28
Fig 17	Exemple de Down / Up	30
Fig 18	Exemple de rotation	30
Fig 19	Exemple de slicing	30
Fig 20	Exemple de scorpig	31
<b>Partie II : Cas Naftal Spa « Direction Ressource Humaine et Direction Paie »</b>		
<b>Chapitre 1 : Présentation générale de l'organisme d'accueil</b>		
Fig 21	Organigramme de l'entreprise NAFTAL Spa	34
<b>Chapitre 3 : Conception du système</b>		
Fig 22	Formalisme utilisé pour la modélisation dimensionnelle des aspects	44
Fig 23	Modèle dimensionnel en flocon de l'aspect « Effectif »	46
Fig 24	Dimension CSP pour l'aspect Effectif	47
Fig 25	Dimension temps pour l'aspect Effectif	48

Fig 26	Dimension agent pour l'aspect Effectif	48
Fig. 27	Dimension structure pour l'aspect Effectif	48
Fig. 28	Dimension tranche_salaire pour l'Effectif	50
Fig 29	Dimension tranche_experience pour l'activité Effectif	50
Fig 30	Modèle dimensionnel en flocon de l'activité « Absentéisme »	52
Fig. 31	Dimension type_absence pour l'activité Absentéisme	53
Fig. 32	Modèle dimensionnel en flocon de l'aspect « conge »	54
Fig. 33	Modèle dimensionnel en flocon de l'activité « Mission »	56
Fig. 34	Architecture détaillée du Data Warehouse	58
Fig. 35	Cube MOLAP de l'activité « Effectif »	65
Fig. 36	Cube MOLAP de l'activité « Absentéisme »	66
Fig. 37	Cube MOLAP de l'activité « Conge »	67
Fig. 38	Cube MOLAP de l'activité « Mission »	68
Fig. 39	Illustration du processus d'alimentation	69
Fig. 40	Description du portail de restitution des données	71
<b>Chapitre 4 : Déploiement du système</b>		
Fig 41	Architecture globale du système	73
Fig. 42	Les trois niveaux de sécurité logiques	76

## LISTE DES TABLEAUX

<i>N°</i>	<i>Désignation</i>	<i>Page</i>
Partie I : Synthèse sur les entrepôts de données		
Tableau 1	Tableau comparatif entre l'Infocentre et l'entrepôt de données.	4
Tableau 2	SID VS transactionnels : la différence par les données	6

Tableau 3	SID VS transactionnels : la différence par l'usage	6
Tableau 4	Synthèse sur les architectures de stockage	16
Tableau 5	OLTP vs OLAP	19
<b>Chapitre 2 : Identification et analyse des besoins</b>		
Tableau 6	Tableau récapitulatif des besoins	41
<b>Chapitre 3 : Conception du système</b>		
Tableau 7	Liste des tables de dimensions de l'entrepôt de données	60
Tableau 8	Liste des tables de faits de l'entrepôt de données	62
Tableau 9	Tableau récapitulatif des choix de faits et des dimensions	64
<b>Chapitre 4 : Déploiement du système</b>		
Tableau 10	Tableau des logiciels par catégories d'utilisateurs	75

## LISTE DES FIGURES

<i>N°</i>	<i>Désignation</i>	<i>Page</i>
<b>Partie I : Synthèse sur les entrepôts de données</b>		
Fig 1	L'approche Infocentre	2
Fig 2	L'approche Entrepôt de données	3
Fig. 3	Définition de l'entrepôt de données	7
Fig. 4	L'entrepôt de données. point focal de l'informatique décisionnelle	10
Fig 5	Structure d'un entrepôt de données	11
Fig 6	La structure pyramidale de l'entrepôt de données	12
Fig 7	L'architecture générale de l'entrepôt de données	13
Fig 8	Architecture d'un produit MOLAP	18
Fig 9	Architecture d'un produit ROLAP	18
Fig 10	Schéma d'un cube	20
Fig 11	Exemple de fait	21
Fig 12	Exemple de dimensions	21
Fig 13	Exemple d'un schéma en étoile	22
Fig 14	Exemple d'un schéma en flocon	23
Fig 15	Exemple d'un schéma en constellation	24
Fig 16	Système d'alimentation ETL	28
Fig 17	Exemple de Down / Up	30
Fig 18	Exemple de rotation	30
Fig 19	Exemple de slicing	30
Fig 20	Exemple de scorpig	31
<b>Partie II : Cas Naftal Spa « Direction Ressource Humaine et Direction Paie »</b>		
<b>Chapitre 1 : Présentation générale de l'organisme d'accueil</b>		
Fig 21	Organigramme de l'entreprise NAFTAL Spa	34
<b>Chapitre 3 : Conception du système</b>		
Fig 22	Formalisme utilisé pour la modélisation dimensionnelle des aspects	44
Fig 23	Modèle dimensionnel en flocon de l'aspect « Effectif »	46
Fig 24	Dimension CSP pour l'aspect Effectif	47
Fig 25	Dimension temps pour l'aspect Effectif	48

Fig 26	Dimension agent pour l'aspect Effectif	48
Fig. 27	Dimension structure pour l'aspect Effectif	48
Fig. 28	Dimension tranche_salaire pour l'Effectif	50
Fig 29	Dimension tranche_experience pour l'activité Effectif	50
Fig 30	Modèle dimensionnel en flocon de l'activité « Absentéisme»	52
Fig. 31	Dimension type_absence pour l'activité Absentéisme	53
Fig. 32	Modèle dimensionnel en flocon de l'aspect « conge »	54
Fig. 33	Modèle dimensionnel en flocon de l'activité « Mission »	56
Fig. 34	Architecture détaillée du Data Warehouse	58
Fig. 35	Cube MOLAP de l'activité « Effectif »	65
Fig. 36	Cube MOLAP de l'activité « Absentéisme »	66
Fig. 37	Cube MOLAP de l'activité « Conge »	67
Fig. 38	Cube MOLAP de l'activité « Mission »	68
Fig. 39	Illustration du processus d'alimentation	69
Fig. 40	Description du portail de restitution des données	71
<b>Chapitre 4 : Déploiement du système</b>		
Fig 41	Architecture globale du système	73
Fig. 42	Les trois niveaux de sécurité logiques	76

## LISTE DES TABLEAUX

<i>N°</i>	<i>Désignation</i>	<i>Page</i>
<b>Partie I : Synthèse sur les entrepôts de données</b>		
Tableau 1	Tableau comparatif entre l'Infocentre et l'entrepôt de données.	4
Tableau 2	SID VS transactionnels : la différence par les données	6

Tableau 3	SID VS transactionnels : la différence par l'usage	6
Tableau 4	Synthèse sur les architectures de stockage	16
Tableau 5	OLTP vs OLAP	19
<b>Chapitre 2 : Identification et analyse des besoins</b>		
Tableau 6	Tableau récapitulatif des besoins	41
<b>Chapitre 3 : Conception du système</b>		
Tableau 7	Liste des tables de dimensions de l'entrepôt de données	60
Tableau 8	Liste des tables de faits de l'entrepôt de données	62
Tableau 9	Tableau récapitulatif des choix de faits et des dimensions	64
<b>Chapitre 4 : Déploiement du système</b>		
Tableau 10	Tableau des logiciels par catégories d'utilisateurs	75

---

---

INTRODUCTION GENERALE

---

---



## I. Contexte général

Avec l'apparition et le développement de phénomènes économiques comme la mondialisation, les entreprises évoluent dans un environnement difficile à appréhender. Il en résulte que la prise de décision stratégique est de plus en plus complexe (augmentation du nombre de paramètres à prendre en compte) et, en même temps, doit intervenir très rapidement pour ne pas laisser le temps aux concurrents de prendre de l'avance.

Les systèmes d'information dont disposent les entreprises sont performants et novateurs, génèrent une grande quantité de données qui sont généralement stockées dans plusieurs bases de données, fichiers...Or, l'exploitation de ces données pour permettre de réaliser des analyses ou des prises de décisions s'avère difficile et fastidieuse: elle est réalisée le plus souvent de manière imparfaite par les décideurs grâce à des moyens classiques (requêtes SQL, vues, outils graphiques d'interrogation...).

De ce constat, ces systèmes paraissent peu adaptés pour servir de support à la décision. Ceci est dû essentiellement à l'utilisation des bases opérationnelles du modèle relationnel, convenant bien aux applications transactionnelles gérant l'activité quotidienne de l'entreprise (Conge, Mission, Absentéisme, Comptabilité...), seulement ce modèle s'avère inadapté au décisionnel.

Face à cette inadéquation, les entreprises ont besoin d'outils ou de modèles pour la mise en place de systèmes décisionnels. Dans ce contexte, l'entrepôt de données représente l'élément principal dans l'élaboration de tels systèmes.

Le succès Gestion des Ressources Humaines constitue le plus souvent un puissant indicateur de performance de l'entreprise. Toutefois, cette performance saurait être plus appréciée si l'entreprise avait l'aptitude d'être réactive aux menaces et aux changements de son environnement. Les systèmes décisionnels d'aide au pilotage des activités des entreprises constituent une réponse globale aux problématiques de pilotage auxquelles elles sont confrontées. Ces systèmes représentent l'ensemble des outils d'analyse et de représentation des données de l'entreprise nécessaire pour une prise de décision rapide, efficace et sûre.

Dans ce cadre, la Direction des Ressources Humaines de l'entreprise Naftal Spa nous a sollicité pour collaborer dans l'étude, la conception et le déploiement d'un entrepôt de données métier

## II. Problématique

NAFTAL Spa est une entreprise nationale chargée de la commercialisation et de la distribution des produits pétroliers sur tout le territoire national. Avec un effectif de 30000 agents, elle se trouve confrontée à des problèmes de gestion d'une grande masse de données liées au pilotage des Ressources Humaines. Elle dispose d'un système d'information au niveau de ses structures opérationnelles qui est la gestion du personnel et d'un système centralisé qui est le système Paie. Ces systèmes couvrent la majeure partie de ses activités de base et génèrent mensuellement des milliers de transactions. A ce stade, l'entreprise se trouve dans l'incapacité d'analyser par ses systèmes actuels le volume de données régulièrement consolidé au niveau central de façon rapide, correcte et efficace.

Les principales difficultés auxquelles elle fait face sont :

- Complexité et diversité de la structure conceptuelle existante rendant l'intégration des données difficile pour une vue multiaxes.
- Grande masse de données est stockée dans plusieurs tables, plus ou moins structurées, ce qui rend difficile l'interrogation et l'analyse.
- Difficulté de garder l'historique des mouvements, même sur une période de quelques mois.

Face à cette situation et toutes ces difficultés, notre tâche sera de réaliser un système décisionnel basé sur l'architecture « Entrepôt de données » qui permettra de mieux acheminer les données et de les représenter d'une meilleure façon afin de satisfaire les besoins des décideurs de NAFTAL Spa Direction Ressource Humaine.

## III. Objectifs

La Direction Ressource Humaine de NAFTAL Spa s'est fixé l'objectif de mettre en place un système décisionnel capable de produire des rapports et analyses sur plusieurs axes et à la demande pour simplifier le pilotage des Ressources Humaines.

A cet effet il nous a été demandé :

- La conception de la base datawarehouse,
- La réalisation des opérations ETL (Extraction, Transformation, Loading) nécessaires pour la constitution d'un entrepôt de données,
- La définition des cubes d'analyse multidimensionnels,
- La création des rapports. et définition des indicateurs de performances.

Le but de la réalisation de notre travail est de permettre aux cadres dirigeants de la Direction des Ressources Humaines de disposer d'informations pertinentes relatives à leurs activités, entre autre absentéisme et congés.

#### **IV. Organisation du mémoire**

Afin de présenter le travail qui nous a été assigné, nous avons opté pour une organisation en deux principales parties. La première concerne l'aspect théorique de notre travail, la seconde la partie réalisation proprement dite.

Lors de la première partie, nous présenterons une synthèse sur les entrepôts de données. Cette partie comprendra tous les concepts relatifs au Data Warehouse, entre autres, la modélisation multidimensionnelle, le concept OLAP, le chargement et alimentation de l'entrepôt, etc.

Dans la deuxième partie, nous nous intéresserons au cas Naftal Spa Direction des Ressources Humaines. Cette partie sera divisée en quatre chapitres, chacun représente une étape dans la construction de l'entrepôt de données :

- 1. Présentation générale de l'organisme d'accueil :** Ce chapitre sera consacré à la présentation de l'entreprise, son organisation, son environnement et ses objectifs.
- 2. Identification et analyse des besoins :** la réussite du Data Warehouse repose en majorité sur la bonne compréhension des utilisateurs et de leurs besoins qui seront définis grâce aux méthode de collecte d'information.

Dans cette phase, nous devons prendre les facteurs clés qui conduisent l'entreprise a vouloir définir de manière efficace ses besoins, et les traduire pour intégrée lors de la

conception.

3. **Conception du système :** Lors de ce chapitre nous développerons un modèle dimensionnel pour notre entrepôt de données. Ensuite, nous présenterons le détail technique de chaque composant de l'architecture entrepôt de données (ETL, Entrepôt de données. Cubes OLAP et le portail de restitution).
4. **Déploiement du système :** Nous y présenterons l'architecture informatique de notre système décisionnel et son détail.

Et bien évidemment nous terminerons notre travail par une conclusion générale.

---

---

**PARTIE I :**  
**SYNTHESE SUR LES ENTREPOTS DE DONNEES**

---

---

## **I. Préambule :**

Toute entreprise possède actuellement d'importants volumes de données, stockés le plus souvent dans différents médias (bases de données, documents, papiers,...) et a besoin d'outil permettant une exploitation efficace et performante de ces données pour l'aider dans ses prises de décision.

En effet, pour faire face à de multiples enjeux économiques, l'anticipation est de rigueur. Cette anticipation ne peut être efficace qu'en ayant recours à des informations pertinentes. Or, l'entreprise dispose de beaucoup de données produites par ses systèmes d'information transactionnels, comme elle peut en acquérir d'autres à partir de son environnement externe.

Le problème qui se pose est la surabondance des données, leur non organisation et leur éparpillement dans différents systèmes hétérogènes. Ces données peuvent présenter une source inestimable d'informations décisionnelles.

Le concept d'entrepôt de données ou « Data Warehouse» a vu le jour en 1990. Le concept de base était de créer une base de données orientées sujet, intégrées, contenant des informations historisées, non volatiles et exclusivement destinées aux processus d'aide à la décision. Le Data Warehouse devait jouer le rôle d'une architecture qui servirait de fondations aux autres applications décisionnelles.

Comme pour tout nouveau concept, cette idée n'a pas été vue de bon œil voire même accueillie avec perplexité par certains. Il faut dire que le Data Warehouse représentait, pour beaucoup de personnes, simplement l'habillage de : l'Infocentre.

## **II. Historique :**

Avant l'apparition du Data Warehouse, plusieurs concepts ont vu le jour. Ces derniers avaient pour objectif de permettre à l'entreprise de disposer d'une source d'informations décisionnelles pertinentes. Parmi ces concepts citons :

## II.1. L'Infocentre :

Depuis fort longtemps, les analystes ont eu besoin des mêmes données que les opérationnels. Mais cela a vite créé quelques anomalies et problèmes, en effet, le besoin exprimé par les analystes engendre des requêtes qui généralement bloquaient celles des opérationnels et qui causaient d'importants dysfonctionnements dans les systèmes d'informations transactionnels. D'où l'idée de créer une copie des bases de données de production, et d'en attribuer une à chaque analyste. La duplication des données se faisait chaque mois, chaque semaine voire même une copie chaque jour qui est étonnant. Les informations étaient dupliquées à l'identique sur un autre ordinateur dédié aux analystes. Grâce aux copies, les opérationnels pouvaient effectuer leurs requêtes et continuer leurs opérations sans être altérées par celles des analystes, ces derniers avaient la liberté de lancer des requêtes quelque soient leurs complexités sans prendre le risque de saturer ou bloquer les autres systèmes transactionnels.

Or, ce mode de fonctionnement, qui se basait sur le dédoublement des données présentait un grand inconvénient qui n'est autre que le coût très élevé. En effet, pour que le concept soit réalisable, il fallait disposer de serveurs, de disques durs, de bases de données et d'autres éléments importants, le tout en double, et cela uniquement pour les besoins de l'analyse. La figure suivante représente l'approche Infocentre.

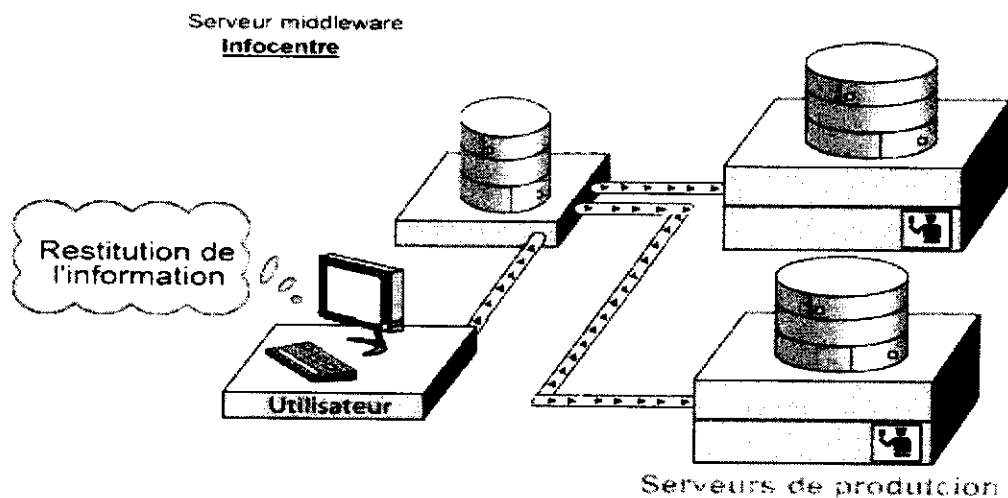


Fig. 1: L'approche Infocentre

## II.2. Les entrepôts de données :

Après maintes années d'utilisation de l'Infocentre, les services informatiques eurent une idée révolutionnaire concernant le mode de stockage des données. Ils ont effectivement constaté que les informations utilisées dans les systèmes transactionnels étaient très différentes de celles interrogées dans les applications décisionnelles. Ainsi le nombre de tables et de fichiers interrogés dans une même requête est beaucoup plus important dans le décisionnel. Mais contrairement aux systèmes transactionnels, les systèmes décisionnels n'apportent aucun changement dans les données stockées dans les bases de données, ils se contentent de les lire. En plus, les informations demandées par un décideur se trouvent généralement dans plusieurs applications et bases de données. Le fait de dupliquer ces bases de données dans un Infocentre ne simplifie en rien ces extractions.

L'idée voulue était donc de mettre en place, et ceci à la sortie des bases de productions, un entrepôt de données qui servirait au stockage des données décisionnelles quelque soit leurs sources. Le terme entrepôt n'est pas exagéré, en effet le Data Warehouse sert d'entrepôt où l'on dispose une copie des données qui vont servir à l'analyse et à la prise de décision. Ceci permettra aux analystes de disposer d'un historique important qui évitera en même temps sa conservation dans les bases de production.

La figure suivante représente l'approche entrepôt de données « *Data Warehouse* ».

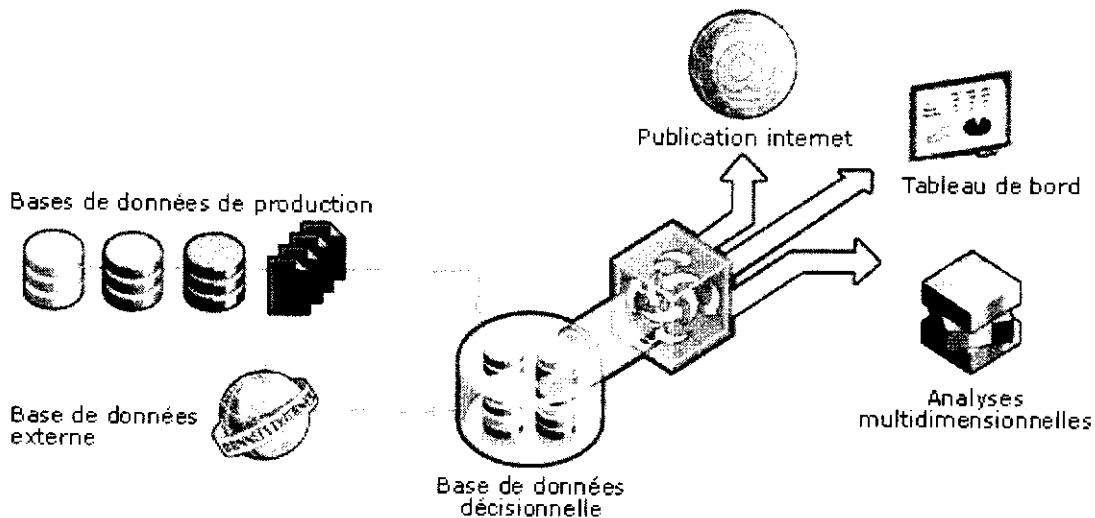


Fig 2 : L'approche entrepôt de données.



### II.3. Infocentre Vs Data Warehouse:

A priori certaines caractéristiques sont identiques. Mais il existe de nombreux éléments permettant de différencier les deux notions :

L'Infocentre est une collection de données orientées sujet, intégrées, volatiles, actuelles, organisées pour le support d'un processus de décision ponctuel.

Le Data Warehouse est une collection de données orientées sujet, intégrées, non volatiles, historisées, organisées pour le support d'un processus d'aide à la décision.

Dans un Infocentre, chaque nouvelle valeur remplace l'ancienne valeur. Il est donc impossible de retrouver une valeur calculée dans une session préalable aux dernières alimentations.

La non volatilité est une caractéristique essentielle du Data Warehouse.

L'Infocentre sert à prendre des décisions opérationnelles basées sur des valeurs courantes.

Au niveau d'un Data Warehouse, l'utilisateur travaille sur les historiques pour des prises de décisions à long terme, des positionnements stratégiques et pour analyser des tendances.

Dans un Infocentre, l'intégration des données est plus ou moins poussée. Le processus d'alimentation est simple. La finalité d'un Infocentre est de permettre aux utilisateurs d'accéder à leurs données dans leurs propres termes. [Nakache, 98].

Le tableau suivant récapitule les points de comparaison entre l'Infocentre et l'entrepôt de données.

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

**Tableau 1 : Tableau comparatif entre l'Infocentre et l'entrepôt de données.**  
[Nakache, 98].

### III. LE DECISIONNEL:

#### III.1. Définition du système d'information décisionnel :

Le système d'information décisionnel est un ensemble de données organisées de façon spécifique, facilement accessible et appropriées à la prise de décision ou encore une représentation intelligente de ces données au travers d'outils spécialisés. La finalité d'un système décisionnel est le pilotage de l'entreprise. [Web 1]

Un SID est : « un système permettant aux décideurs de l'entreprise de disposer d'informations pertinentes et d'outils d'analyse puissants pour les aider à prendre les bonnes décisions au bon moment ». [Devisy, 02].

Les SID offrent aux décideurs une vision transversale de l'entreprise (effectif global, effectif en absences, effectif en missions...). Une infrastructure décisionnelle constitue le système d'analyse des processus économiques de l'entreprise.

#### III.2. Evolution des systèmes décisionnels:

Le développement des premiers systèmes d'information s'est concentré sur l'automatisation des processus opérationnels. Il est vrai que ces systèmes d'informations offrent un gain significatif en matière de productivité et assurent le bon déroulement de l'activité principale. Mais qu'en est-il des données contenues dans ces systèmes ? Elles sont :

- ✘ **Eparpillées** : il existe souvent de multiples systèmes, conçus pour être efficaces pour les fonctions sur lesquelles ils sont spécialisés.
- ✘ **Peu structurées pour l'analyse** : la plupart des systèmes informatiques actuels ont pour objet de conserver en mémoire l'information, et sont structurés dans ce but.
- ✘ **Focalisées pour améliorer le quotidien** : toutes les améliorations technologiques se sont focalisées pour améliorer cette capacité en terme de volume, qualité, rapidité d'accès. Il manque très souvent la capacité à nous donner les moyens de tirer parti de cette mémoire pour prendre des décisions.
- ✘ **Utilisées pour des fonctions critiques** : la majorité des systèmes existants sont conçus dans le but unique de nous servir avec des temps de réponse corrects.

C'est à partir des années 90 que les entreprises ont compris que les données sont non seulement utiles dans le cadre d'une utilisation opérationnelle, mais qu'elles peuvent leur

trouver une utilisation stratégique. Ces données constituaient alors la matière première des SID.

Les applications utiles dans le processus de prise de décision sont nombreuses, et déjà fortement présentes dans le système d'information des entreprises. C'est le cas des applications d'analyse de l'effectif, de la masse salariale ... Mais elles répondent toutes au même processus : analyser des données préalablement collectées par les applications opérationnelles de l'entreprise, les mettre en forme, aider à distinguer les grandes tendances, et publier des résultats sous forme de graphiques, de tableaux ou de rapports. A l'instar des systèmes transactionnels, les systèmes décisionnels sont caractérisés par des spécificités communes. Ils travaillent sur de gros volumes de données, les requêtes qu'ils leur appliquent étant par contre beaucoup plus complexes, ils disposent de plus de temps pour les exécuter.

Les deux tableaux suivants montrent la différence entre les systèmes transactionnels et décisionnels du point de vue de leur usage et des données utilisées

<i>Systemes transactionnels</i>	<i>SID</i>
Orienté applications	Orienté thèmes et sujets
Situation instantanée	Situation historique
Données détaillées et codées non redondantes	Informations agrégées cohérentes souvent avec redondance
Données changeant constamment	Informations stables et synchronisées dans le temps
Pas de référentiel commun	Un référentiel unique

**Tableau 2 : SID VS transactionnels : la différence par les données.**

<i>Systemes transactionnels</i>	<i>SID</i>
Assure l'activité au quotidien	Permet analyse et prise de décision
Pour les opérationnels	Pour les décideurs
Mises à jour et requêtes simples	Lecture uniquement et requêtes complexes transparentes
Temps de réponse immédiats	Temps de réponse moins critiques
Faible volume à chaque transaction	Large volume manipulé
Conçue pour la mise à jour	Conçu pour l'extraction
Usage maîtrisé	Usage aléatoire

**Tableau 3 : SID VS transactionnels : la différence par l'usage.**

## IV. LES CONCEPTS DE BASE D'UN DATA WAREHOUSE :

### IV.1. Définitions du Data Warehouse :

Plusieurs définitions ont été proposées, qu'elles soient académiques ou des définitions apportées par des éditeurs d'outils, de bases de données ou même de constructeurs.

- Une collection de données thématiques, intégrées, non volatiles et historisées pour la prise de décisions. [Inmon ,02].
- Fait référence à une collection de technologies d'aide à la décision permettant à des (managers, dirigeants et analystes) de prendre des décisions pertinentes et rapides [Chaudhuri ,97 A].
- Une collection de technologies décisionnelles formant un environnement permettant aux décideurs de prendre des décisions plus pertinentes et plus rapides. [Kerkri, 01].
- Un Data Warehouse est un entrepôt central de données qui représente une vue intégrée des données d'une entreprise. Il s'agit d'un outil majeur d'aide à la décision. [Abbey, 02]

De manière générale un entrepôt de données peut être défini par : «Un entrepôt de données est un ensemble de données intégré, temporel (dépendant du temps), persistant, orienté autour d'un même sujet, et utilisé dans le processus d'aide à la décision »

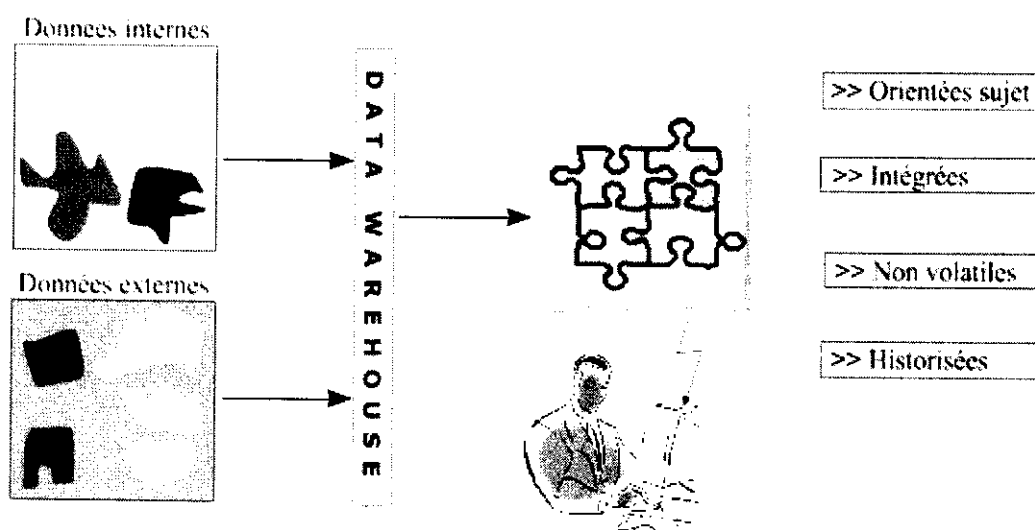


Fig 3: Définition de l'entrepôt de données

Voyons maintenant les caractéristiques des données stockées dans un entrepôt de données :

❖ **Orientées sujets**

L'entrepôt de données est organisé autour des sujets majeurs de l'entreprise, contrairement aux données des systèmes opérationnels [Nakache, 98]. Les données sont structurées par thèmes. Grâce à cette organisation, le décideur peut disposer d'un ensemble d'informations utiles sur un sujet le plus souvent transversal aux structures fonctionnelles et organisationnelles de l'entreprise.

❖ **Données intégrées**

Les données alimentant l'entrepôt de données proviennent de multiples sources hétérogènes. Avant d'être intégrées dans le Data Warehouse, elles doivent être converties, reformatées et nettoyées.

L'intégration de ces données consiste à contenir leurs hétérogénéités, afin de donner au contenu de l'entrepôt de données une présentation homogène et pour garantir sa qualité. Avant d'être incorporées dans l'entrepôt de données, les données subissent un processus de transformation et d'unification afin d'en avoir un état cohérent. Par exemple, la consolidation de l'ensemble des informations concernant un agent est nécessaire pour donner une vue homogène de cet agent.

Cette phase d'intégration est très complexe et représente 60 à 90 % de la charge totale d'un projet. [Nakache, 98].

❖ **Historisées**

Dans un système de production, la mise à jour des données se fait lors de chaque nouvelle transaction. Une mise à jour annule et remplace l'ancienne valeur. Dans un entrepôt de données, la prise en compte de l'évolution des données est primordiale pour la prise de décision et notamment les prédictions. La donnée ne doit jamais être mise à jour, une empreinte de temps doit être associée à la donnée afin d'être capable d'identifier une valeur particulière dans le temps.

**❖ Non volatiles**

Les données de l'entrepôt sont essentiellement utilisées en mode de consultation, elles sont très rarement modifiées. La non volatilité des données est en quelque sorte une conséquence de l'historisation. Une même requête effectuée à quelque mois d'intervalle en précisant la date de référence de l'information recherchée donnera le même résultat.

**❖ Résumées**

Les informations issues des sources de données doivent être agrégées et réorganisées afin de faciliter le processus de prise de décision.

**❖ Disponible pour l'interrogation et l'analyse**

Les utilisateurs doivent pouvoir consulter les données réorganisées de l'entrepôt en fonction de leur droit d'accès au travers d'outils interactifs d'aide à la manipulation et l'analyse.

**IV.2. Objectifs du Data Warehouse:**

L'objectif principal d'un entrepôt de données est de rendre l'utilisateur autonome dans l'exploitation des informations. En effet, une donnée est stockée dans divers systèmes opérationnels, sous divers formats et sur différentes plates-formes informatiques. Dans certains cas, la donnée se trouve dans des systèmes archaïques dont l'accès est impossible sans l'utilisation des programmes spéciaux adéquats à leurs structures.

Les objectifs d'un Data Warehouse se résument par les points suivants : [A.Ramdane ,04]

- ✘ Assurer l'accès immédiat, rapide et facile aux informations de l'entreprise.
- ✘ Avoir des données cohérentes, utiles et les rendre lisibles et manipulables.
- ✘ Stocker les données publiées mais aussi un ensemble de requêtes, d'analyses et de présentation de l'information.
- ✘ Ajouter de la valeur aux données en permettant de retrouver la source de la donnée.

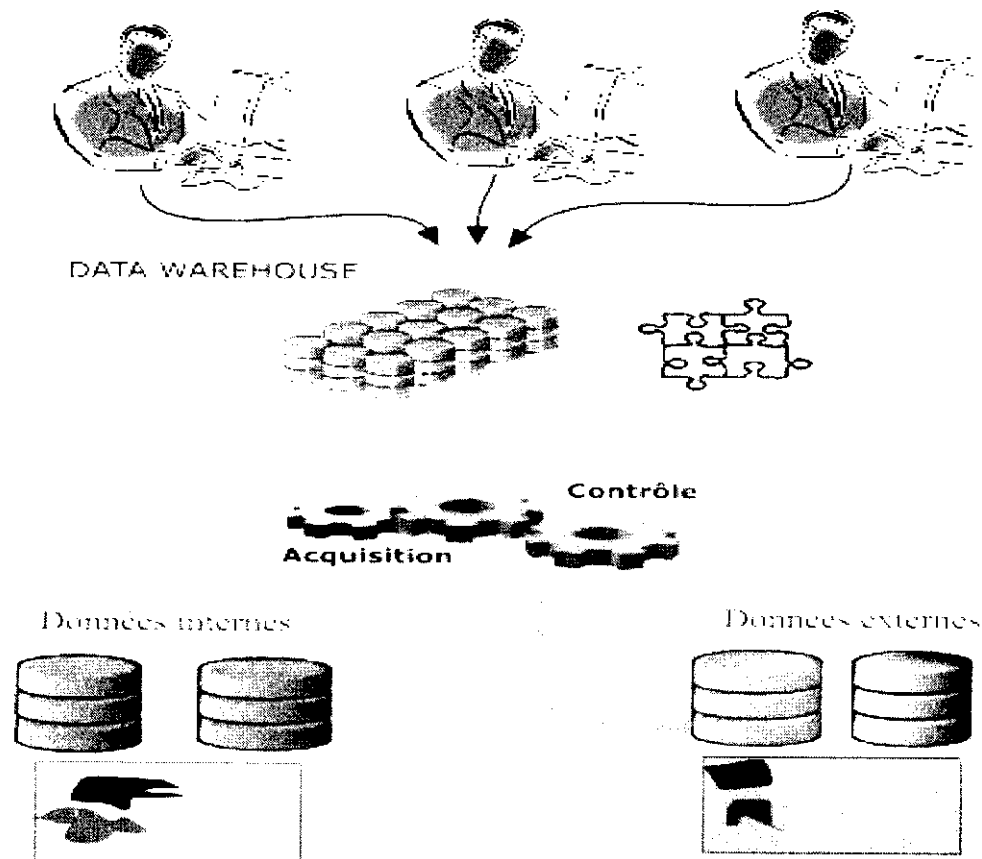


Fig 4 : L'entrepôt de données, point focal de l'informatique décisionnelle

Cette figure illustre l'objectif d'un entrepôt de données, sorte de point focal stockant en un endroit unique toute l'information utile provenant des systèmes de production et des sources externes.

#### IV.3. La structure du Data Warehouse

Un entrepôt de données se structure en quatre classes de données, organisées selon un axe historique et un axe synthétique. La figure suivante illustre cette structuration et positionne les classes les unes par rapport aux autres dans un cadre d'architecture de données.

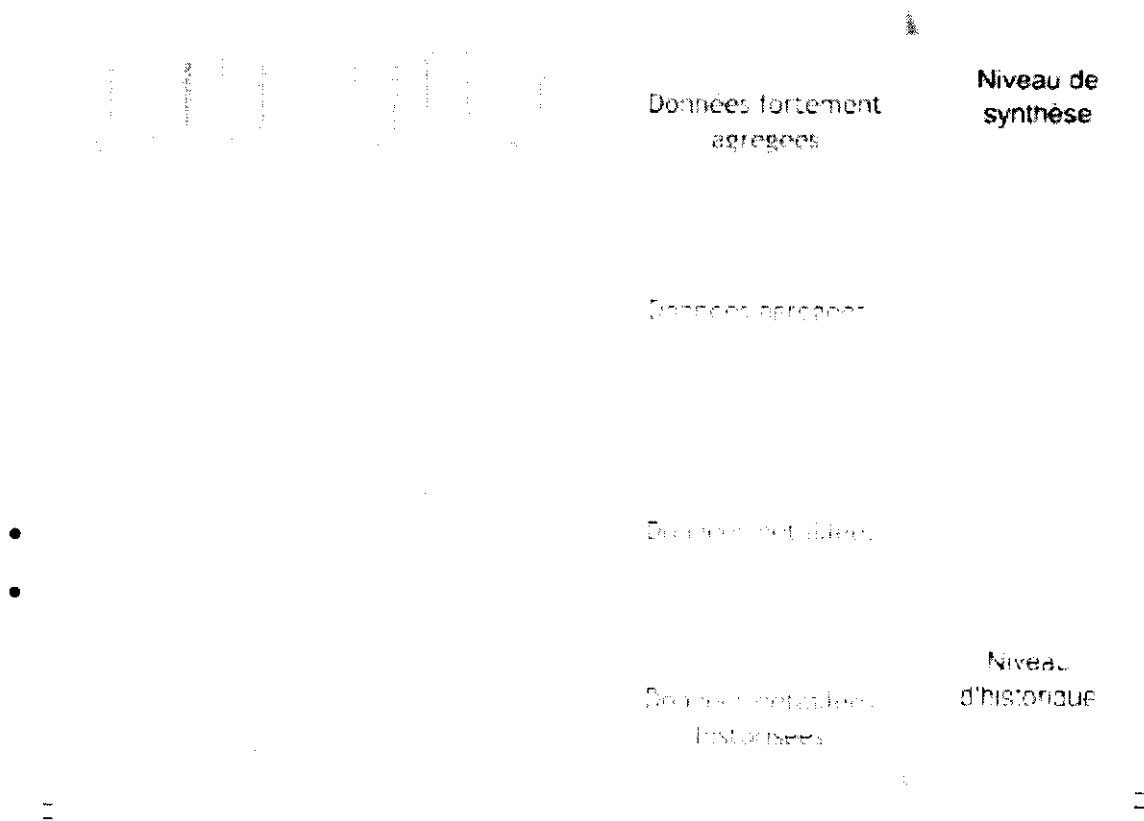


Fig 5 : Structure d'un entrepôt de données

- **Les données détaillées**

Elles reflètent les événements les plus récents. Les intégrations régulières des données issues des systèmes de productions vont habituellement être réalisées à ce niveau.

Les volumes à traiter sont plus importants que ceux gérés en transactionnel. Le niveau de détail géré dans l'entrepôt de données n'est pas forcément identique au niveau de détail géré dans les systèmes opérationnels.

- **Les données agrégées**

Elles correspondent à des éléments d'analyse représentant les besoins des utilisateurs. Elles constituent déjà un résultat d'analyse et une synthèse de l'information contenue dans le système décisionnel. Elles doivent être accessibles et compréhensibles.

Dans le cas d'un agrégat, l'information est composée du contenu présenté (nombres d'effectifs, ...) et de l'unité (par mois, par sexe, ...).



- **Les méta-données**

Les données fédérées dans l'entrepôt de données proviennent de sources très hétérogènes. Les méta-données constituent un dictionnaire et une véritable aide en ligne permettant de connaître l'information contenue dans l'entrepôt de données. Elles sont idéalement intégrées dans un référentiel.

- **Les données historisées**

Une nouvelle insertion de données provenant du système de production ne détruit pas les anciennes valeurs, mais crée une nouvelle occurrence de la donnée.

La logique d'accès aux données la plus utilisée est la suivante : les utilisateurs commencent à rechercher les données par le niveau le plus agrégé, puis approfondissent leur recherche vers les données les plus détaillées.

La figure suivante illustre la logique d'accès aux données de l'entrepôt sous la forme d'une pyramide.

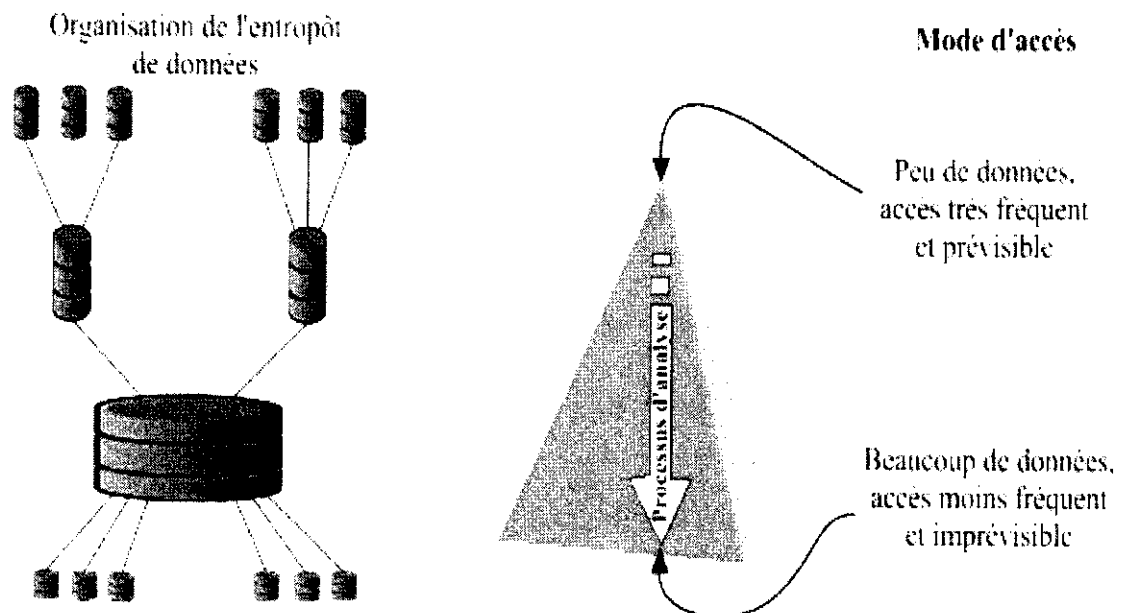


Fig 6 : La structure pyramidale de l'entrepôt de données

L'entrepôt de données réussit dans une entreprise lorsque le nombre d'utilisateurs accédant aux données de détail augmente.

#### IV.4. L'architecture du Data Warehouse

L'architecture de l'entrepôt de données doit être conçue de façon à ce que la bonne information soit aisément accessible pour la bonne personne au bon moment. Pour y parvenir, cette architecture doit d'une part permettre de transformer l'information de façon à ce qu'elle réponde aux objectifs métiers des utilisateurs, et d'autre part d'organiser le mouvement de l'information de la source à la cible.

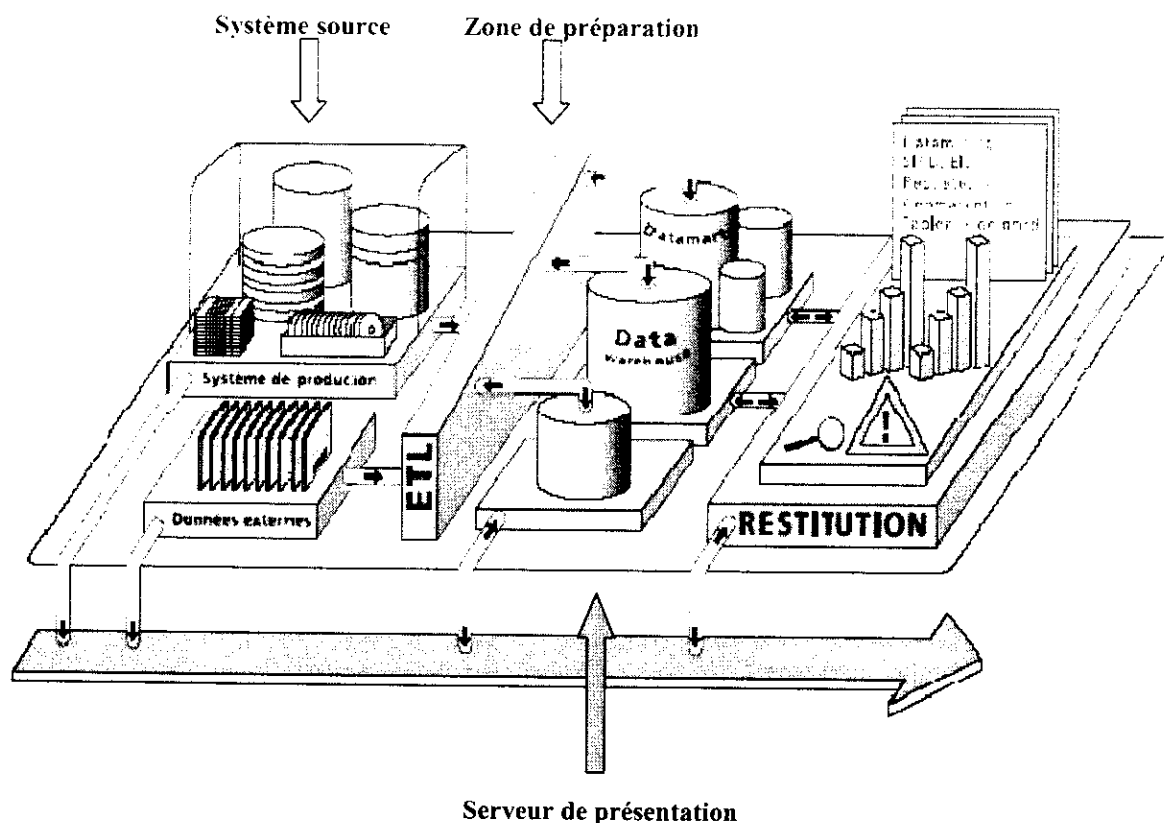


Fig 7 : L'architecture générale de l'entrepôt de données [Devisy, 02]

Cette architecture illustre le processus de construction d'un système de pilotage et d'aide à la décision. La première étape consiste à mettre en place un entrepôt de données, dont la fonction consiste à récupérer des données depuis des bases d'informations existantes et à les stocker. Il faut pour cela mettre en place une procédure d'alimentation qui permettra de rassembler toutes les données utiles, quelle que soit leur origine, comptabilité, gestion commerciale, courrier électronique, serveurs externes ou encore internet.

On distingue sept parties constituant l'architecture de l'entrepôt de données :

⊕ **Le système source**

*Système opérationnel d'enregistrement, dont la fonction consiste à capturer les transactions liées à l'activité de l'entreprise [Kim, 97]. Il peut être représenté par les systèmes de production ou bien de données externes. La principale priorité du système source est le temps de disponibilité.*

⊕ **La zone de préparation des données**

*Ensemble des processus qui nettoient, transforment, combinent, archivent, suppriment les doublons, c'est-à-dire prépare les données sources en vue de leur intégration puis de leur exploitation au sein de l'entrepôt de données [Kim, 97]. Cette zone comprend tout ce qui se trouve entre les systèmes sources et le serveur de présentation. La frontière qui détermine la zone de préparation des données est que la zone de préparation ne doit en aucun cas être accessible à l'utilisateur final par requête ou par un quelconque autre service de présentation.*

⊕ **Le serveur de présentation**

*Machine cible sur laquelle l'entrepôt de données est stocké et organisé pour répondre en accès direct aux requêtes émises par des utilisateurs [Kim, 97]. Des générateurs d'états ou d'autres applications. Sur le serveur de présentation, les données seront stockées et représentées sous une forme dimensionnelle.*

⊕ **L'entrepôt de données**

*L'entrepôt de données correspond à la source de données interrogeable de l'entreprise [Kim, 97]. Il est alimenté par la zone de préparation des données.*

⊕ **Le magasin de données**

*Le magasin de données ou Datamart est défini comme un sous-ensemble logique d'un entrepôt de données [Kim, 97]. Il représente un projet réalisable. On considère souvent le magasin de données comme la réduction de l'entrepôt de données à un seul processus ou à un groupe de processus ciblant un groupe métier spécifique.*

### ⊕ **Portail de restitution**

*C'est la part publique de l'entrepôt de données [Kim, 97]. Il représente ce que voient les utilisateurs, les outils avec lesquels ils travaillent. Les services offerts par le portail de restitution sont les services d'accès aux données, les applications de modélisation et de data mining.*

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

### ⊕ **Meta données**

Ce sont toutes les informations de l'environnement de l'entrepôt de données qui ne constituent pas les données proprement dites. Ce sont les « données sur les données ».

## **IV.5. L'implémentation du Data Warehouse**

Pour implémenter le stockage de l'entrepôt de données, trois types d'architectures sont possibles :

### ***IV.5.1. Implémentation selon l'architecture réelle***

Elle est généralement retenue pour les systèmes décisionnels. Le stockage des données est réalisé dans un SGBD séparé du système de production. Le SGBD est alimenté par des extractions périodiques.

Avant le chargement, les données subissent d'importants processus d'intégration, de nettoyage, de transformation. L'avantage est de disposer de données préparées pour les besoins de la décision et répondant aux objectifs de l'entrepôt de données.

Les inconvénients sont le coût de stockage supplémentaire et le manque d'accès en temps réel.

#### IV.5.2. Implémentation selon l'architecture virtuelle

Cette architecture n'est pratiquement pas utilisée pour l'entrepôt de données.

Les données résident dans le système de production. Elles sont rendues visibles par des produits middleware ou par des passerelles.

Il en résulte deux avantages : pas de coût de stockage supplémentaire et l'accès se fait en temps réel. L'inconvénient est que les données ne sont pas préparées.

#### IV.5.3. Implémentation selon l'architecture remote

C'est une combinaison de l'architecture réelle et de l'architecture virtuelle. Elle est rarement utilisée.

L'objectif est d'implémenter physiquement les niveaux agrégés afin d'en faciliter l'accès et de garder le niveau de détail dans le système de production en y donnant l'accès par le biais de middleware ou de passerelle.

Les différents éléments d'appréciation sont repris dans le tableau récapitulatif ci-dessous :

	<i>Architecture Réelle</i>	<i>Architecture Virtuelle</i>	<i>Architecture remote</i>
<i>Utilisation</i>	Retenue pour les systèmes décisionnels	Rarement utilisée	Rarement utilisée
<i>Stockage</i>	SGBD séparé du système de production, alimenté par des extractions périodiques	Données résident dans le système de production	Combinaison des architectures réelle et virtuelle
<i>Avantage</i>	Données préparées pour les besoins de la décision	Pas de coût de stockage supplémentaire, accès en temps réel	
<i>Inconvénient</i>	Coût de stockage supplémentaire, manque d'accès en temps réel	Données non préparées	

Tableau 4 : Synthèse sur les architectures de stockage

## **V. LE CONCEPT OLAP :**

### **V.1. Présentation :**

Le concept OLAP est apparu pour la première fois en 1993 dans un livre intitulé « Providing OLAP to User Analysts » : An IT Mandate, réalisé par le docteur Edgar Codd, l'un des concepteurs des bases de données relationnelles.

Littéralement On-Line Analytical Processing, OLAP désigne une catégorie d'applications et de technologies permettant de collecter, stocker, traiter et restituer des données multidimensionnelles, à des fins d'analyses.

Ralph KIMBALL a défini le concept OLAP comme : « Activité globale de requête et de présentation de données textuelles et numériques contenues dans l'entrepôt de données : style d'intégration et de présentation spécifiquement dimensionnel ». La technologie OLAP est non relationnelle et presque toujours basée sur un cube de données multidimensionnel explicite.

La base de données multidimensionnelle OLAP est incontestablement le cœur de tout système décisionnel. La vision multidimensionnelle qu'elle apporte des données métiers a fait d'elle la pierre angulaire de la prise de décisions. Sans pour autant réduire l'importance des autres briques d'un système décisionnel, un véritable outil d'aide à la décision ne saurait être construit sans s'appuyer sur une base de données OLAP.

### **V.2. Les différents outils OLAP :**

Les moteurs du décisionnel, les outils OLAP, complètent les Data Warehouse. Le stockage des informations peut être réalisé de plusieurs façons : MOLAP (Multidimensionnel OLAP), ROLAP (Relationnel OLAP) et HOLAP (Hybride OLAP) :

#### **V.2.1. Les outils MOLAP :**

La base MOLAP est l'application physique du concept OLAP elle adopte réellement une structure multidimensionnelle. L'intérêt est que les temps d'accès sont optimisés ; cette approche nécessite de redéfinir des opérations pour manipuler ces structures multidimensionnelles. Les jointures sont déjà faites. MOLAP agrège tous par défaut. Cette base est plus rapide et performante mais limitée pour l'instant au Giga Octet.

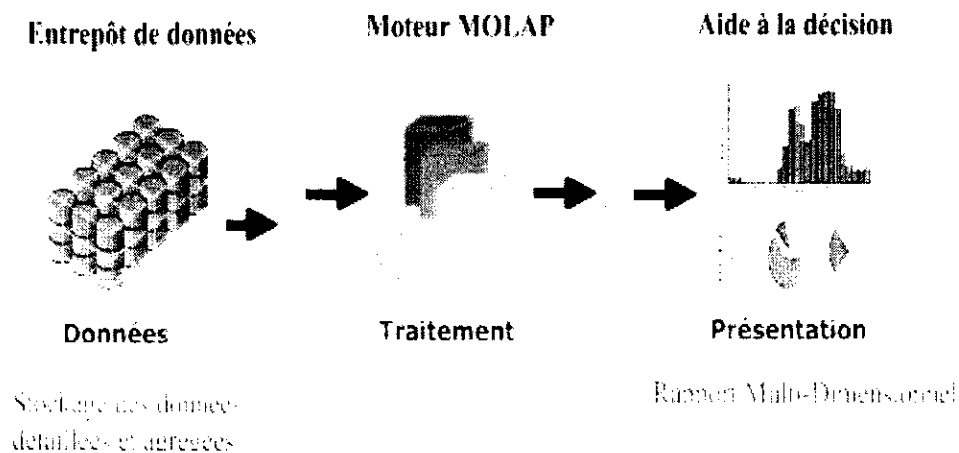


Fig 8 : Architecture d'un produit MOLAP

### V.2.2. Les outils ROLAP :

Ce système est capable de simuler le comportement d'un SGBD multidimensionnel en exploitant un SGBD classique. On entend par cela qu'il conserve la table de la donnée de base sur le serveur OLAP d'origine (ou dans un entrepôt de données) mais utilise un jeu indépendant de tables relationnelles pour stocker les références dimensionnelles et les données agrégées, ce qui dispense de l'acquisition d'un SGBD multidimensionnel, il utilise aussi un middleware pour implémenter les opérations spécifiques de l'OLAP. De ce fait, lorsque l'utilisateur interroge la base de données relationnelle, un cube multidimensionnel est fourni en réponse.

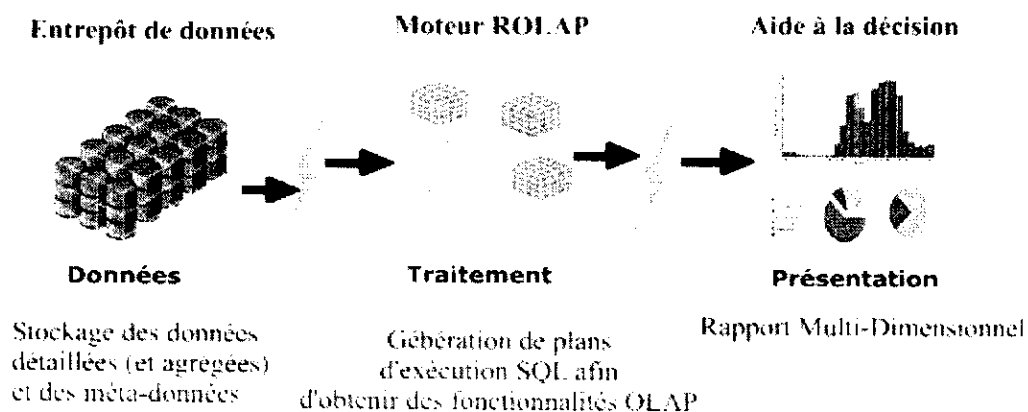


Fig 9 : Architecture d'un produit ROLAP

### V.2.3. Les outils HOLAP (Hybride OLAP) :

La base HOLAP est un compromis entre les deux concepts précédents : une base MOLAP pour les données souvent consultées, une base ROLAP pour les autres données. La solution HOLAP (aussi appelé DOLAP, Database OLAP) combine les avantages des deux solutions MOLAP et ROLAP.

### V.3. Système OLTP Vs système OLAP :

On peut résumer les différences entre les systèmes OLTP (On-Line Transactional Processing) et les systèmes OLAP dans le tableau suivant :

	<i>Processus OLTP</i>	<i>Processus OLAP</i>
<b>Conception</b>	Orientées applications Structure statique	Orientées sujets (d'analyse) Structure évolutive
<b>Données</b>	Dynamiques Courantes Complètes	Statiques Historiques Résumées
<b>Utilisateurs</b>	Nombreux Variés (employés, directeurs, ...) Concurrents	Peu nombreux Uniquement les décideurs Non concurrents
<b>Utilisation</b>	Mises à jours et interrogations Requêtes prédéfinies Réponses immédiates Accès à peu d'information	Interrogations Requêtes imprévisibles et complexes Réponses moins rapides Accès à de nombreuses informations

Tableau 5 : OLTP vs OLAP



## VI. ORGANISATION DES DONNEES D'UN DATA WAREHOUSE :

Pour pouvoir exploiter facilement les données au niveau du Data Warehouse, le concepteur doit les organiser d'une manière bien particulière. Dans le cadre du Data Warehouse, la modélisation entité relation n'est pas appropriée, car elle est souvent utilisée pour créer un modèle complexe regroupant tous les processus de l'entreprise. C'est pourquoi la modélisation dimensionnelle a fait son apparition. Elle permet de créer des modèles individuels pour traiter des sujets de gestion discontinue.

### VI.1. Modélisation multidimensionnelle :

« Les données à analyser au niveau d'un magasin doivent refléter la vision d'une classe d'analystes ». Cette vision correspond à une structuration des données selon plusieurs dimensions pouvant représenter des notions variées telles que le temps et la localisation géographique. On parle alors de modélisation et de traitement multidimensionnels des données.

Les analyses décisionnelles sont basées sur des traitements OLAP directement reliés à la modélisation de l'information sous une forme conceptuelle proche de la perception qu'en a l'analyste. Cette perception de l'information est basée sur une vision multidimensionnelle des données.

### VI.2. Eléments de modélisation multidimensionnelle :

- **Cube :**

C'est la représentation des données de la base multidimensionnelle au niveau conceptuel. Chaque point du cube contient les mesures relatives à la combinaison des différentes dimensions d'analyse qui le comportent.

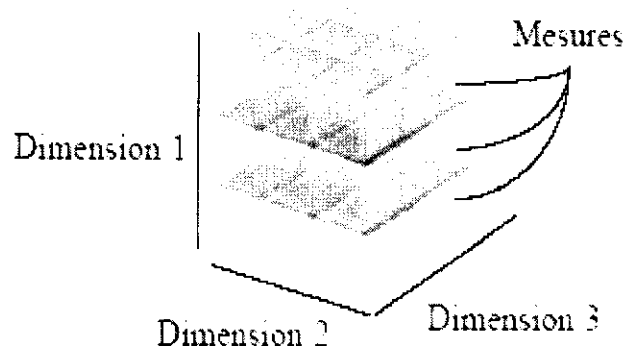


Fig 10 : Schéma d'un cube

- **Table de fait :**

Elle capture les données qui résument les activités de l'entreprise (effectifs en absences, ...). Elle comporte généralement une multitude de lignes, représentant des centaines de millions d'enregistrements lorsqu'elle couvre une ou plusieurs années d'historique.

Sa caractéristique principale est de ne contenir que des données numériques « faits » (les faits ne doivent pas être des données autres que celles des champs de mesure numérique) qui peuvent être synthétisées afin de fournir des informations sur l'historique des activités de l'entreprise. Chaque table de fait a un index qui contient, comme clés étrangères, les clés primaires des tables de dimension associées.

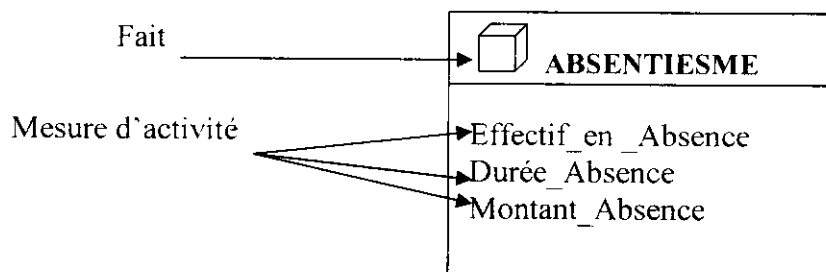


Fig 11 : Exemple de fait

- **Table de dimension :**

Elle contient des attributs qui décrivent les enregistrements de faits de la table de fait. Les dimensions servent à enregistrer les valeurs pour lesquelles sont analysées les mesures de l'activité. Une dimension est généralement formée de paramètres (ou attributs) textuels et discrets. La modélisation dimensionnelle génère des tables de dimension contenant chacune des attributs de faits indépendants de ceux figurant dans d'autres dimensions.

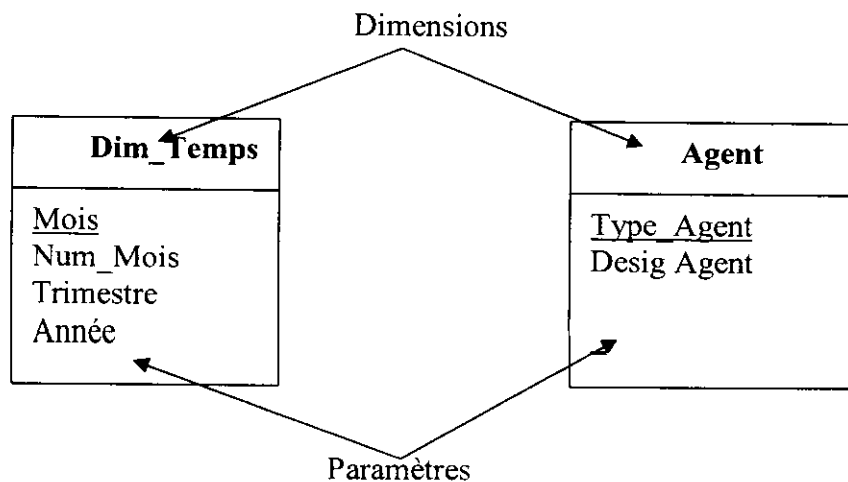


Fig. 12 : Exemple de dimensions

### VI.3. Schémas de Modélisation dimensionnelles :

#### ⊕ Schéma en Etoile :

Dans ce modèle, au centre de l'étoile se trouve la table de fait. L'identifiant de cette table est une clé multiple composée de la concaténation des clés de chacune des dimensions d'analyse. Autour de la table de faits figurent tous les éléments caractérisant les dimensions d'analyse. Ces caractéristiques sont regroupées dans des tables de dimension. Les plus grands avantages de ce type de modèle sont la lisibilité et la performance.

- ▲ *La lisibilité* : ce modèle est très parlant pour l'utilisateur et sa finalité est évidente. Il est naturellement orienté sujet et définit clairement les indicateurs d'analyse.
- ▲ *La performance* : les chemins d'accès à la base de données sont prévisibles.

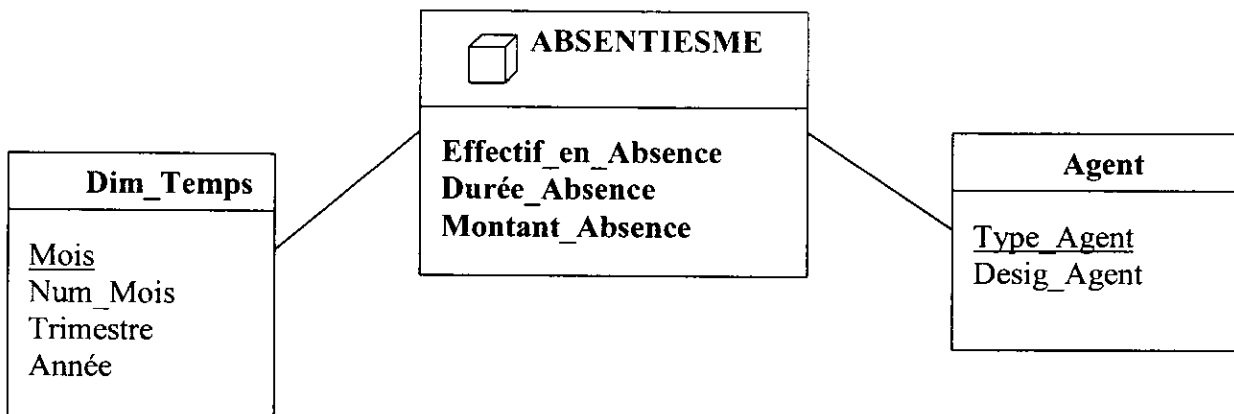


Fig 13 : Exemple d'un schéma en étoile

#### ⊕ Schéma en Flocon :

La modélisation en flocon est une modélisation en étoile pour laquelle on éclate les tables de dimensions en sous tables selon la hiérarchie de cette dimension. L'avantage de cette modélisation est de formaliser une hiérarchie au sein d'une dimension. Par contre, la modélisation en Flocon induit une dé normalisation des dimensions générant une plus grande complexité en terme de lisibilité et de gestion.

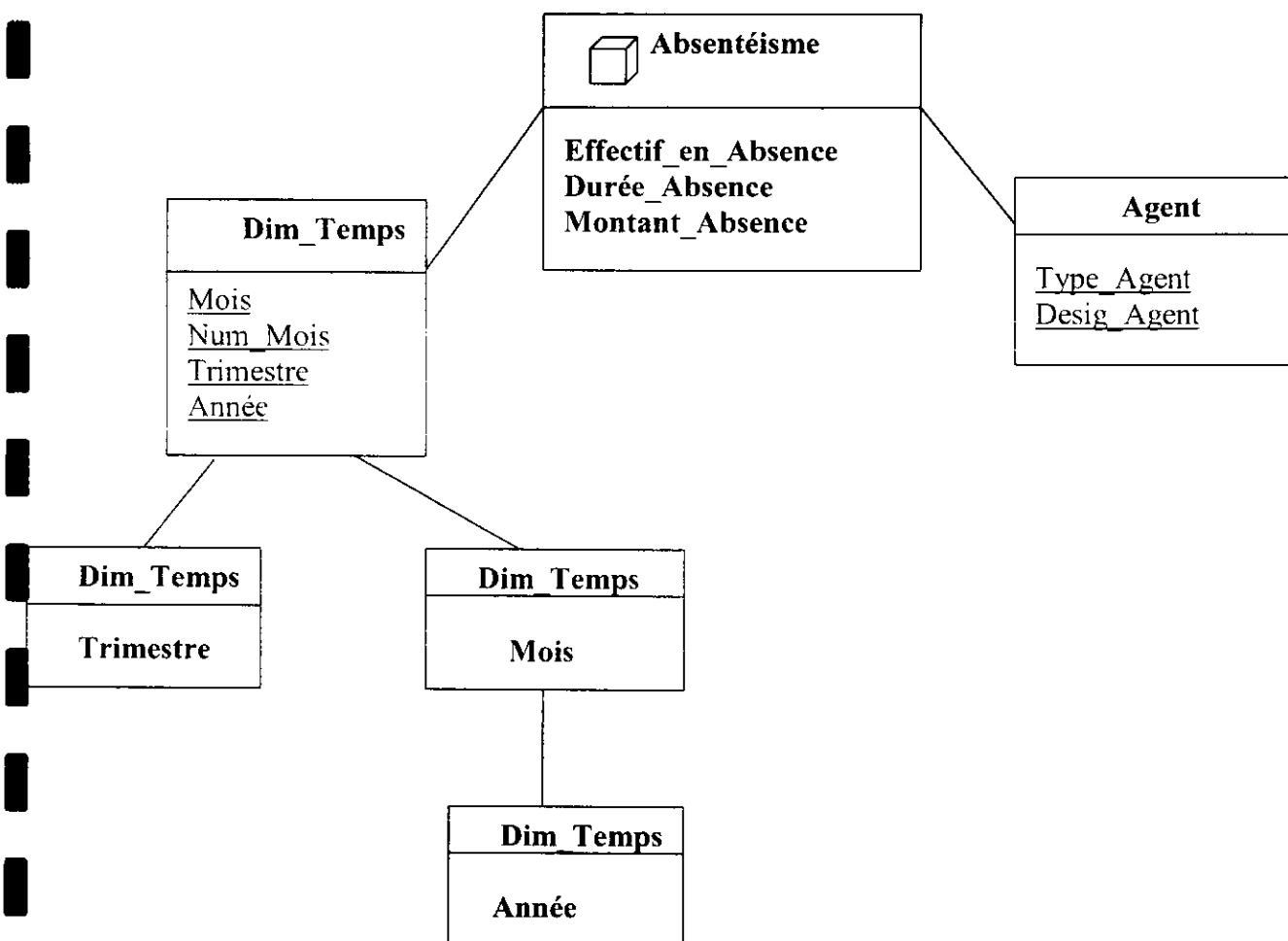


Fig. 14 : Exemple d'un schéma en flocon

### ⊕ Schéma en Constellation :

Il s'agit de fusionner plusieurs modèles en étoile qui utilisent des dimensions communes. Un modèle en constellation comprend donc plusieurs faits et des dimensions communes ou non.

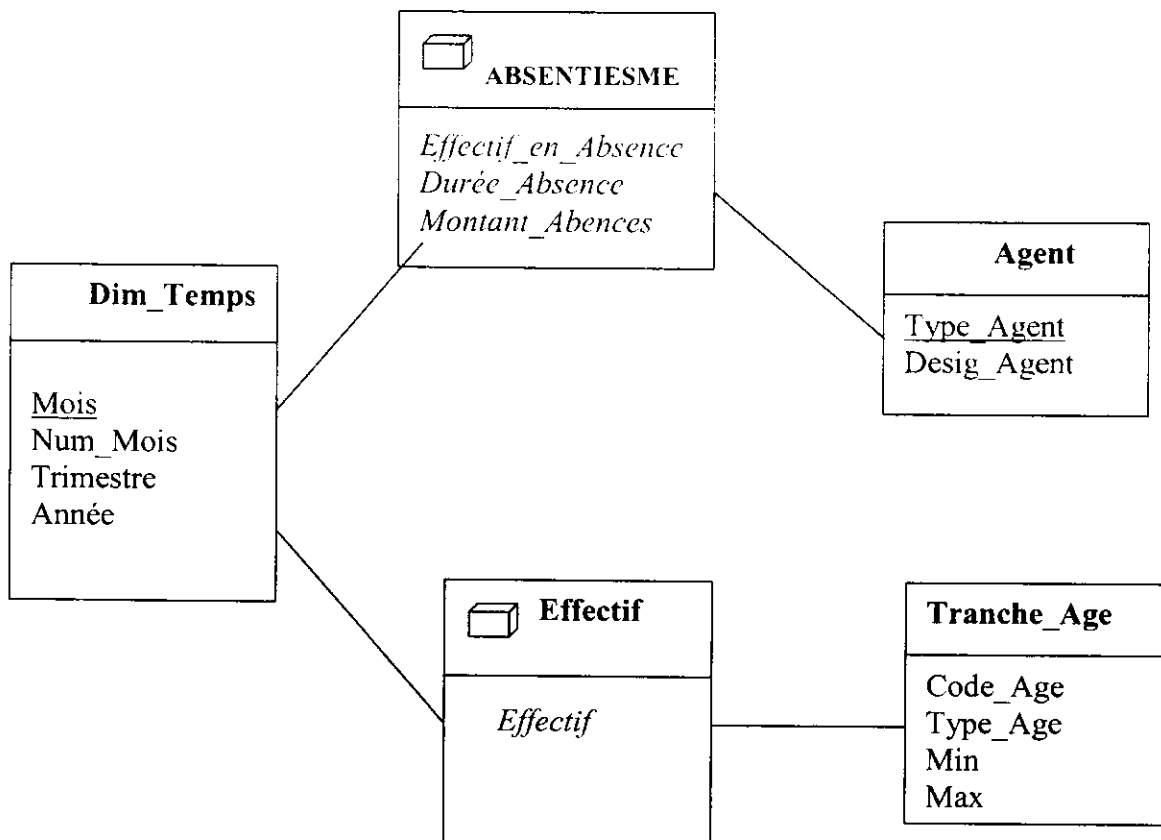


Fig. 15: Exemple d'un schéma en constellation

## VII. La construction de l'entrepôt de données

Construire un entrepôt de données dimensionnel revient à faire correspondre les besoins de la communauté des utilisateurs avec la réalité des informations disponibles [Kim, 97].

L'entrepôt n'est pas un produit ou un logiciel mais un environnement. Il est bâti et ne s'achète pas. Les données sont puisées dans les bases de production, nettoyées, normalisées, puis intégrées.

La construction d'un entrepôt de données se fait en trois étapes interdépendantes :

- 1) Etude préalable et définition des besoins, objectifs et la démarche suivie.
- 2) Conception du modèle dimensionnel de données qui représente l'entrepôt conceptuellement et logiquement.
- 3) Mise en œuvre de l'architecture selon l'ordre des phases suivantes :
  - a) Construction de la base de l'entrepôt.
  - b) Construction des cubes OLAP, la base multidimensionnelle.
  - c) Construction de la zone d'alimentation (l'examen des données, le choix des méthodes et des dates auxquelles les données entreront dans l'entrepôt).

⊕ Par la suite, nous détaillons chaque étape :

### **VII.1. Etude préalable et définition des besoins :**

Cette partie de l'étude ressemble à toute étape préliminaire à l'implémentation d'un nouveau système d'information. Une bonne conception de l'entrepôt de données repose sur une bonne compréhension des besoins des utilisateurs finaux. L'étude des besoins doit déterminer le contenu de l'entrepôt et son organisation, d'après les résultats attendus par les utilisateurs, les requêtes qu'ils formuleront, les projets qui ont été définis et leur priorité.

L'expression des besoins par les utilisateurs met souvent en évidence la volonté d'obtenir : des analyses sur ce qui s'est passé (par exemple comparer les effectif actuelles d'une structure avec celles de l'année dernière) ou des analyses prédictives (par exemple déterminer le nombre d'absences potentiels pour un type d'absence). Il faut alors recenser les données disponibles dans les bases de production, toutes les données de production ne sont pas utiles dans l'entrepôt, donc il faut identifier les besoins des utilisateurs en terme de données utiles.

### **VII.2. Conception du modèle de données :**

Pour présenter un modèle idéal, il faut d'abord avoir une description complète de l'environnement de l'organisation, permettant la conception adéquate de l'entrepôt de données dimensionnel.

Une fois lancer dans la conception nous devons toujours nous demander si nous répondons avec la meilleure efficacité possible aux besoins les plus importants de l'organisation.

Ralph KIMBALL propose une démarche qui se résume à neuf décisions, dans un ordre bien défini, ces décisions portent sur les points suivants :

- 1) *Choisir les processus d'activité à modéliser* : on va identifier les processus majeurs de l'entreprise dans lesquelles les informations sont collectées au profit de l'entrepôt à partir des applications existantes (par exemple : nombre effectifs en absence, nombre effectifs en mission, la masse salariale ...). une fois les processus identifiés, une ou plusieurs tables de faits sont construites.
- 2) *Choisir le grain de chaque table de faits* : le grain est la signification précise d'un enregistrement du plus bas niveau dans la table de faits.
- 3) *Choisir les dimensions de chaque table de faits* : le choix des dimensions s'accompagne de la définition de tous les attributs textuels (les champs) qui garniront la table de dimension. Exemple de dimension (temps, sexe, type agent...).
- 4) *Choisir les faits mesurés que contiendra chaque enregistrement de table de faits* : une table de faits ne peut contenir que des attributs de type quantités numériques additives telles que : durée d'absence, montant d'absence.
- 5) *Les attributs des dimensions* : avec des descriptions complètes.

A ce stade, la conception de la structure logique est terminée, et les autres étapes concernent la structure physique :

- 6) *Comment suivre les dimensions à évolution lente* : la dimension agent change constamment, car les humains changent de nom, se marient et divorcent et change d'adresse, et dans une compagnie d'assurances il est essentiel de savoir le statut de la personne assurée à l'époque d'un sinistre passé et non le statut actuel. Nous appelons ce type de dimensions des dimensions à évolution lente. Pour le traitement de ces changements, on a trois solutions :
  - a) Recouvrir et perdre les valeurs anciennes dans l'enregistrement de la dimension, renonçant ainsi à la possibilité de suivre les événements ou situations passées.
  - b) Créer un enregistrement de dimension supplémentaire lors du changement, comportant les nouvelles valeurs de l'attribut, ce qui revient à segmenter l'historique très exactement selon l'ancienne et la nouvelle description.

- c) Créer de nouveaux champs « actuel » à l'intérieur de l'enregistrement d'origine de la dimension, tout en conservant en même temps les valeurs enregistrées en premier lieu. Ceci permet de décrire l'historique en amont du changement en utilisant les valeurs d'origine, et de décrire l'historique aval selon les nouvelles valeurs des attributs.
- 7) *Les agrégats* : se sont les dimensions hétérogènes, les minidimensions, les modes de requêtes et autres décisions sur le stockage physique.
- 8) *L'étendue historique de la base de données* : on peut permettre à l'historique de s'accumuler au-delà des trois ans envisagés pour n'importe quels entrepôt de données.
- 9) *L'urgence avec laquelle les données doivent être extraites et chargées dans l'entrepôt de données* : généralement le chargement se fait chaque jour les premières heures du matin, ou attendre jusqu'à la fin de la semaine.

### **VII.3. La mise en œuvre de l'architecture :**

Cette phase se passe en trois étapes :

#### **VII.3.1. Construction de l'entrepôt :**

C'est une base de données dont l'objectif est de centraliser l'information décisionnelle, elle représente la base de données physique de notre entrepôt.

Cette base contient toutes les tables de faits et les tables de dimensions définies dans la conception de la structure logique.

#### **VII.3.2. Construction des cubes OLAP :**

Les cubes représentent la modélisation multidimensionnelle des données. Ils constituent une façon de faciliter l'analyse.

#### **VII.3.3. Processus d'alimentation:**

Les projets de l'entrepôt de données sont pilotés par les besoins de gestion. Au sein de l'entreprise, un entrepôt de données met à la disposition des décideurs une information précise extraite des systèmes opérationnels. Les entreprises qui ont mis en œuvre, avec succès, des



projets de ce type ont vu leurs efforts récompensés par un avantage compétitif significatif et un profit accru.

Un tel environnement est par nature complexe : il est nécessaire d'intégrer les "méta-données" qui décrivent ses outils et ses bases de données. Il est également indispensable de permettre l'adjonction de nouveaux outils.

Mais l'évolution de l'environnement de l'entrepôt de données ne concerne pas que la mise en œuvre de nouveaux outils. Des événements ponctuels, telles que la modification de structure des données opérationnelles, doit aussi être pris en compte.

Dans l'environnement entrepôt de données, le point clé est l'administration des "méta-données". Les méta-données sont habituellement gérées au sein d'un dictionnaire ou d'un référentiel. Souvent, les logiciels de l'entrepôt de données utilisent un moyen "propriétaire" de gestion des méta-données, créant ainsi des flots d'informations isolées. Les technologies informatiques devenant de plus en plus complexes et multiples, il devient indispensable d'assurer une gestion centralisée des méta-données, en particulier pour la pérennité des environnements de l'entrepôt de données.

### VII.3.3.1. Qu'est ce qu'un ETL ?

Pour en extraire les données utiles, l'outil d'ETL (Extraction Transfer Loading) doit pouvoir se connecter aux sources, qu'il s'agisse des applications ou des bases en production. En ce sens, il joue un rôle d'intégration au niveau des données.

### VII.3.3.2 Système d'alimentation / ETL :

Le système d'alimentation a pour but de transformer les données primaires (par exemple issues des systèmes de production) en informations stockées dans le Data Warehouse. Pour cela il est nécessaire de réaliser les fonctions suivantes :

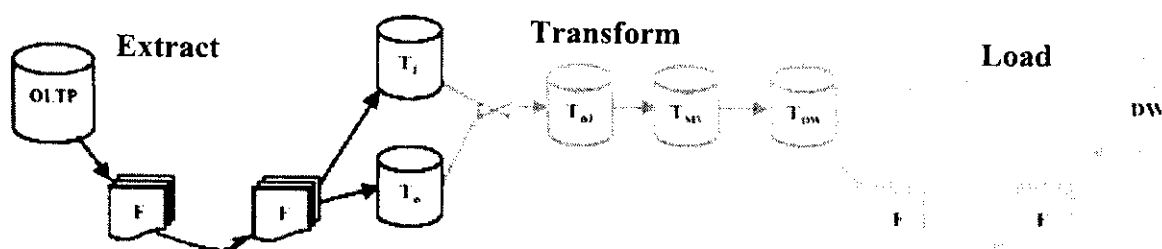


Fig 16 : Système d'alimentation ETL

**▲ L'Extraction des données :**

Elle consiste à collecter les données utiles dans le système de production. Pour rafraîchir la base décisionnelle, il faut identifier les données ayant évolué afin d'extraire le minimum de données, puis planifier ces extractions afin d'éviter les saturations du système de production.

Il existe plusieurs technologies utilisables pour extraire les données :

- Des passerelles, fournies principalement par les éditeurs de bases de données.
- Les utilitaires de réplication, utilisable si les systèmes de production et décisionnels sont homogènes et si la transaction à appliquer aux données est légère.
- Des outils spécifiques d'extraction. Ces outils sont certainement la solution opérationnelle au problème de l'extraction.

**▲ La transformation des données :**

Elle inclut la mise en correspondance des formats de données, le nettoyage, la transformation et l'agrégation (données mensuelles). Ainsi, le calcul des données secondaires et de fusion ou d'éclatement des informations composites.

**▲ Le chargement des données :**

Il consiste la dernière phase d'alimentation. Il est indispensable de maîtriser la structure du Système de Gestion de la Base de Données (SGBD) pour optimiser au mieux le processus.

**VIII. Navigation dans les données :**

**VIII.1. Drill Down & Drill Up :**

Le Drill est une action de zoom avant ou arrière opérée pour obtenir un niveau de détail plus fin (Drill down) et plus élevé (Drill up) sur une dimension d'un système multidimensionnel.

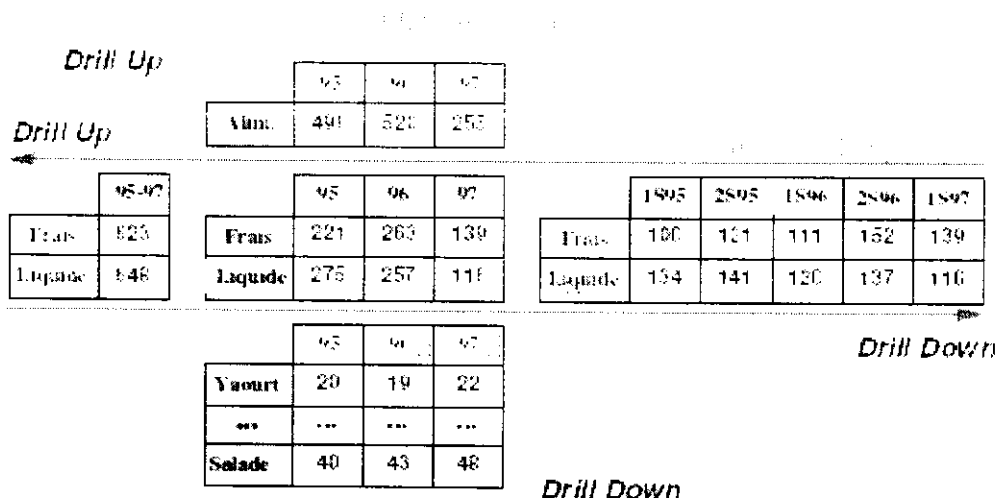


Fig 17 : Exemple de Down / Up

**VIII.2. Slice & Dice :**

Le « Slice and Dice » (Tranche et dé en Français) désigne la possibilité de faire pivoter dynamiquement les axes du tableau d'analyse croisée. Il est possible ainsi de passer d'un tableau représentant les ventes par magasin en lignes et jour en colonnes par un tableau similaire représentant les ventes par magasin en colonnes et jour en ligne.

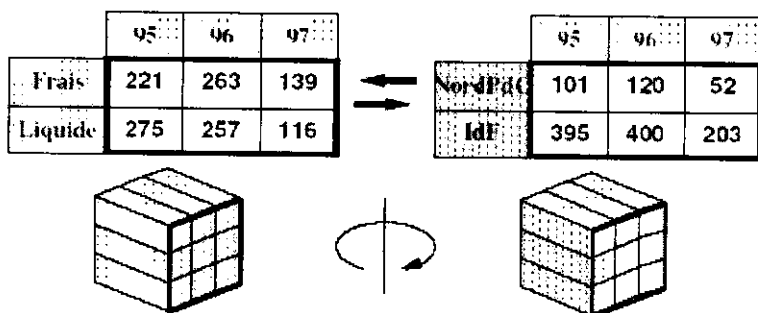


Fig 18 : Exemple de rotation

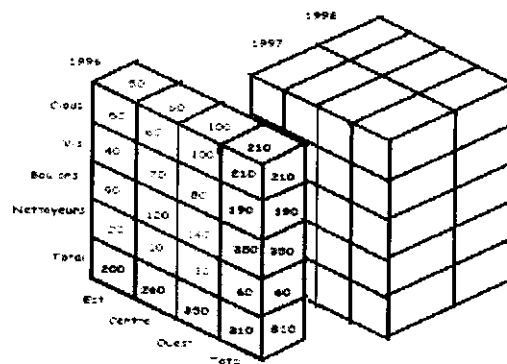


Fig 19: Exemple de slicing

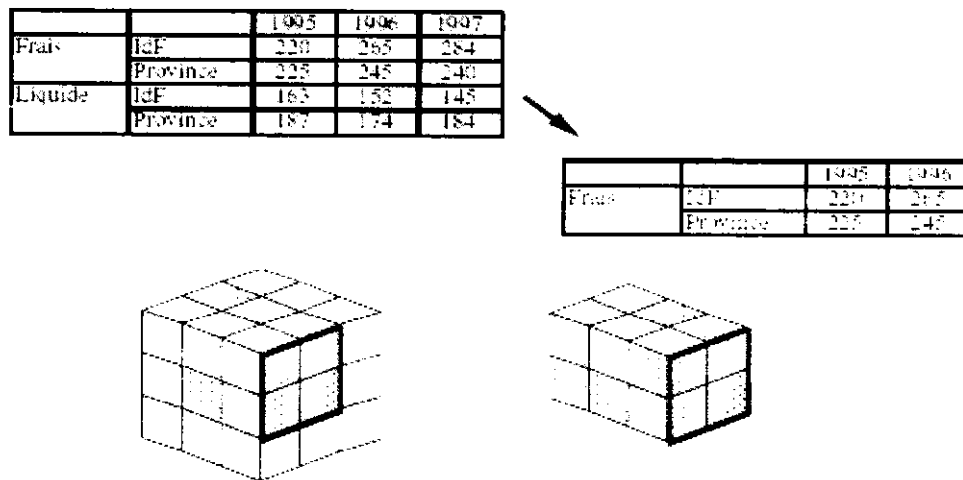


Fig 20 : Exemple de scoping

### VIII.3. Data Surfing :

Le data surfing est la possibilité laissée à l'utilisateur de circuler librement, de manière intuitive et ergonomique dans un modèle dimensionnel, au delà d'un « simple » Drill Down ou Slice and Dice. L'utilisateur peut alors modifier dynamiquement ses axes d'analyse ou appliquer un nouveau filtre à ses données.

Ce modèle doit être assez complexe pour adresser l'ensemble des demandes des utilisateurs et assez souple pour que l'utilisateur puisse le personnaliser à son goût, en créant de nouveaux attributs ou de nouveaux axes d'analyse ou en définissant de nouvelles métriques calculées.

## IX. CONCLUSION :

Dans cette partie, nous avons vu comment un entrepôt de données permet aux décideurs de travailler dans un environnement informationnel, référencé, homogène et historisé. Cette technique l'affranchit des problèmes liés à l'hétérogénéité des systèmes informatiques et à l'hétérogénéité des différentes définitions de données issues de l'historique de l'organisation.

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

---

---

**PARTIE II :**  
**CONSTRUCTION DU DATA WAREHOUSE DE**  
**L'ENTREPRISE NAFTAL SPA**

---

---

---

---

**PARTIE II :**

---

**CHAPITRE 1 :**

**PRESENTATION GENERALE DE L'ORGANISME D'ACCUEIL**

---

---

## **I. Présentation générale de NAFTAL Spa :**

### **I.1. Historique :**

C'est en 1880 vers la fin du 19<sup>ème</sup> siècle que remonte l'exploration pétrolière dans le bassin de CHLEF. L'exploration pétrolière en ALGERIE fut couronnée de succès par la découverte, dès 1954, de gaz à Djebel BERGA (AIN SALAH) puis, durant l'année 1956 en juin par la découverte du champ de (HASSI R'mel). Donc, il fallait tirer le produit du fond de la terre, le raffiner, et le distribuer. C'est dans ce cadre que l'état Algérien a créé SONATRACH le 31/12/1963, et qui a pour tâches de suivre toutes les opérations depuis l'extraction du brut jusqu'à sa commercialisation. Dès lors, SONATRACH détient le monopole dans le domaine des hydrocarbures jusqu'au jour où l'état Algérien a pris l'initiative de restructurer plusieurs entreprises parmi lesquelles figurait SONATRACH.

Ce qui a donné lieu à plusieurs entreprises et en particulier E.R.D.P. (Entreprise National de Raffinage de Distribution des Produits Pétroliers) qui a été créée le 6 Avril 1980 et mise en place le 1<sup>er</sup> janvier 1982. Elle est chargée de l'industrie, du raffinage et de la distribution des produits pétroliers sous le sigle de NAFTAL Spa (l'appellation de NAFTAL Spa provient de « NAFT » qui signifie 'Pétrole' en Arabe et « AL » en référence à 'AL\_DJAZAIRE').

En 1987 NAFTAL a fait l'objet d'une restructuration (par le décret 87\_189) qui l'a divisée en:

- ⊕ *NAFTEC* : pour ce qui concerne le raffinage (industrie du raffinage).
- ⊕ *NAFTAL* : pour ce qui concerne la distribution et la commercialisation.

La raison sociale de la société change suite à cette séparation des activités, et NAFTAL Spa est désormais chargée de la commercialisation et distribution des produits pétroliers.

En 1998, NAFTAL devient une FILIALE à 100% de SONATRACH.

### **I.2. Missions de NAFTAL SPA :**

Les missions principales de Naftal Spa sont :

- ⊕ Stockage, Distribution et commercialisation des Carburants, GPL (gaz de pétrole liquéfié), Lubrifiants, Bitumes, Pneumatiques, GPL/carburant, Produits spéciaux.
- ⊕ Transport des produits pétroliers.
- ⊕ Veiller à l'application et au respect des mesures relatives à la sécurité industrielle, la sauvegarde et la protection de l'environnement.
- ⊕ Procéder à toute étude de marché de consommation des produits pétroliers.

- ⊕ Définir et développer une politique en matière d'audit, concevoir et mettre en œuvre des systèmes intègres d'information.
- ⊕ Développer et mettre en œuvre les actions visant à une utilisation optimale et relationnelle des infrastructures et moyens.
- ⊕ Développer une image de marque de qualité.

### **I.3. Organisation de l'entreprise :**

Le schéma d'organisation de la macrostructure de la société NAFTAL Spa s'articule autour des principales structures ci-après énumérées:

- ⊕ La direction générale.
- ⊕ Les structures fonctionnelles (directions exécutives, direction centrales, directions de soutien).
- ⊕ Les structures opérationnelles (Branches).

Le schéma ci-dessus représente l'organisation de NAFTAL Spa, et dans ce qui suit, nous présentons chaque structure séparément :



## ORGANIGRAMME GENERAL DE NAFTAL

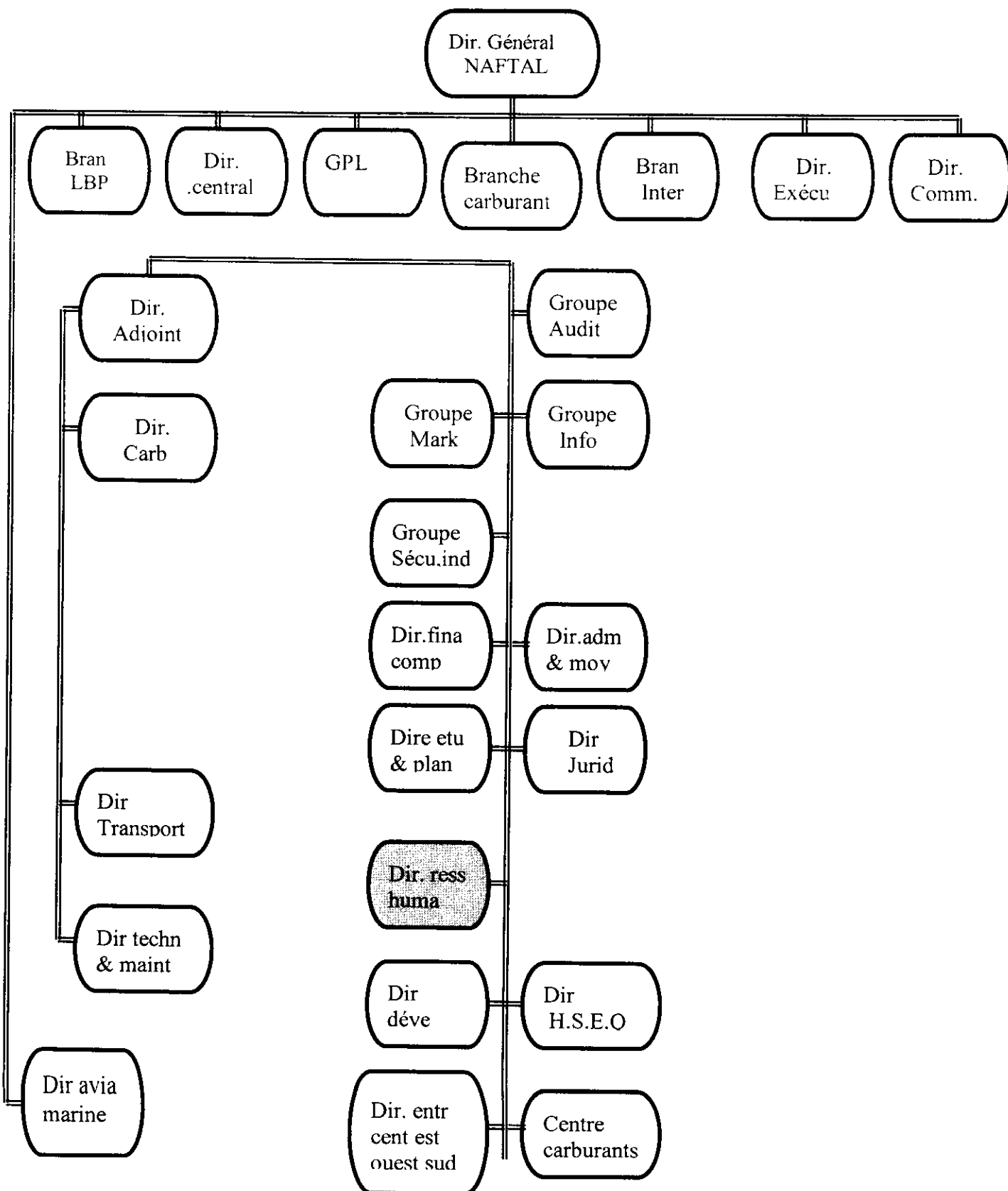


Fig.22 : Organigramme de l'entreprise NAFTAL

**I.3.1 La direction générale** : composée d'un président directeur général qui est secondé par :

- ⊕ *Conseillers et assistants* : chargés d'assister le PDG dans la prise de décision stratégique.
- ⊕ *Le comité exécutif* : ce comité se charge de définir et d'arrêter les grands axes stratégiques de développement de la société.
- ⊕ *Le comité directeur* : assiste le PDG dans la gestion des activités de la société.

**I.3.2 Les structures fonctionnelles** : sont organisées en quatre directions exécutives, cinq centrales et deux de soutien :**a) Les directions exécutives :**

- ⊕ *Dir.Stratégie / Planification & Economie* : parmi ses activités la planification, l'étude économique, le suivi des marchés et l'évaluation des projets.
- ⊕ *Dir.Finances* : chargée du financement, le contrôle budgétaire et trésorerie.
- ⊕ *Dir.Systèmes d'Information & Procédures* : se charge des fonctions informatiques, des systèmes d'informations et les procédures de gestion.
- ⊕ *Dir.Ressources Humaines* : détermine la réglementation et les relations de travail, le suivi des emplois, des formations et la gestion des carrières. (Elle fera objet de notre étude).

**b) Les directions centrales :**

- ⊕ *Dir.Marketing* : est chargée de la stratégie d'image de la société.
- ⊕ *Dir.Recherche & Développement* : assure les activités de recherche sur le plan technologique en rapport avec les activités de la société.
- ⊕ *Dir.Audit* : assure la conduite des missions d'audit dans la société.
- ⊕ *Dir.Qualité, Hygiène, Sécurité industrielle & Environnement* : établit les règles et consignes de la sécurité industrielle, elle est chargée également d'adapter les normes relatives à la protection de l'environnement.
- ⊕ *Dir.Affaires Sociales et Culturelles* : assure les activités de la santé, du social, du sport et la culture au profit des salariés et des ayants droits.

c) *Les directions de soutien* : assurent la gestion administrative du siège social de la société, ces directions sont :

⊕ *Dir. Administration générale.*

⊕ *Dir. Sécurité interne de l'Etablissement.*

**1.3.3 Les structures opérationnelles** : sont organisées en ligne de produit (cinq branches) :

a) *Branche Carburants* : a pour mission principale la commercialisation et la distribution des produits carburants.

b) *Branche Commercialisation* : assure la commercialisation des produits pétroliers, pour satisfaire les besoins de la clientèle.

c) *Branche Activités Internationales* : s'occupe de toutes les activités et relations internationales de l'entreprise.

d) *Branche lubrifiants et pneumatique (LBP)* : a pour mission et activité principales la commercialisation des produits lubrifiants terres, produits spéciaux et pneumatiques.

e) *Branche « gaz de pétrole liquéfiés » (GPL)* : a pour mission et activité principales la commercialisation et le conditionnement des gaz de pétrole liquéfiés.

## **2. CONCLUSION :**

Lors de ce chapitre, nous avons pu donner une présentation générale de notre organisme d'accueil « NAFTAL SPA ». Nous avons décrit les différentes structures constituant cette organisation. Notre entrepôt de données se limitera à la Direction Ressource Humaine.

Le chapitre suivant sera consacré à l'identification de l'ensemble des besoins des décideurs et cadres dirigeants, une étape cruciale avant la conception de tout entrepôt de données.

---

---

**PARTIE II :**

---

**CHAPITRE II :**  
**IDENTIFICATION ET ANALYSE DES BESOINS**

---

---

## I. INTRODUCTION :

La construction de tout entrepôt de données doit passer par la phase de collecte et d'identification des besoins, car ce dernier doit répondre aux attentes des décideurs et des analystes. Cette identification représente une étape primordiale avant d'entamer la modélisation même de l'entrepôt.

Dans ce chapitre, nous aborderons les différentes approches utilisées qui nous permettront de comprendre l'activité de l'entreprise et ainsi mieux cerner les processus clés de ces activités.

## II. PRESENTATION DES APPROCHES :

### II.1. Première approche « le questionnaire » :

Le but attendu du questionnaire est de permettre l'extraction des indicateurs par rapport à différents axes nécessaires à la conception du modèle dimensionnel.

#### Elaboration du questionnaire :

Le questionnaire que nous avons élaboré comprend des questions qui concernent essentiellement l'activité des gestionnaires. Nous distinguons trois catégories de personnes à questionner « les membres de la haute direction, les responsables et analystes, les auditeurs de données ». Dans notre questionnaire nous sommes intéressés qu'aux deux premières catégories.

➤ *Partie réservée aux dirigeants* : les questions de cette partie portent sur :

- Des informations relatives au cadre dirigeant interrogé.
- Des informations permettant de comprendre de façon globale les activités de l'entreprise et ses orientations.
- Des informations concernant les problèmes liés aux prises de décisions.
- Des informations sur les métriques d'évaluation.

Les informations attendues de cette partie du questionnaire nous permettront de faire le lien entre les données collectées au sein de l'organisation.

- *Partie réservée aux responsables et analystes* : dans cette partie nous essayerons d'entrer d'avantage dans le détail. Elle comprend des questions portant sur :
- Des informations relatives au responsable ou analyste interrogé.
  - Les principaux aspects des activités des personnes interrogées.
  - Analyse de l'activité : elle permet de déterminer les différentes dimensions et faits à mesurer.
  - Analyse des besoins : elle permet de connaître les types d'analyses effectuées actuellement et celle souhaitées par les décideurs.

En ce qui concerne les audits de données, il a été préférable de réaliser des entretiens avec les collaborateurs de la Direction GRH.

### **II.2. Deuxième approche « l'entretien » :**

L'entretien se révèle comme le moyen le plus efficace pour collecter les informations nécessaires à notre travail. L'entretien permet, entre autres de connaître les impressions de notre interlocuteur. La personne interviewée se trouve généralement plus à l'aise à répondre à nos questions verbalement que de se restreindre à des réponses limitées écrites. L'interviewé peut en revanche proposer des suggestions et nous éclaircir sur des points pas assez précis dans le questionnaire.

Lors de notre enquête, nous avons pu nous entretenir avec les chefs de départements de la Direction GRH. Le but des entretiens avec ces derniers est d'établir l'existence des données requises par les besoins évoqués par les utilisateurs.

### **II.3. La documentation :**

Elle consiste à étudier les rapports utilisés par les départements qui constituent la Direction GRH. Les principaux rapports étudiés étant les rapports d'activités mensuels pour l'année 2006 et 2007 concernant essentiellement la paie et les effectifs.

### **II.4. Obstacles rencontrés lors de l'enquête :**

Comme lors de toute enquête menée, nous avons rencontré des difficultés et des obstacles pour parvenir à la collecte d'informations. Chaque approche utilisée présente des inconvénients :

**Le questionnaire :**

Cette méthode est le plus souvent utilisée lorsqu'on a affaire à un grand nombre de personnes à interroger. Parmi les difficultés que nous avons rencontrées lors de notre enquête avec cette méthode, nous citons :

- ⊕ La surcharge de travail qui pèse sur les personnes interrogées. En effet, la distribution du questionnaire a été réalisée au milieu de l'année, période où le travail est très dense (bilans annuels...).
- ⊕ Refus de communiquer des informations que la personne interrogée juge confidentielles.
- ⊕ La personne interrogée ne peut pas formuler son avis car elle se limite au contexte du questionnaire.

**L'entretien :**

Cette approche nous a posé les problèmes suivants :

- ⊕ La personne entretenue peut être source de problèmes (ne répond pas aux questions, annule ou rate un rendez-vous).
- ⊕ Difficulté de planifier l'entretien vu la surcharge de travail.

**La documentation :**

Parmi les inconvénients observés citons :

- ⊕ Abondance des documents à étudier...etc.

**Remarque :** afin de mener notre enquête, nous avons utilisé les trois approches citées auparavant de façon complémentaire.

### III. RECUEIL RECAPILATIF:

Après avoir récupéré la majorité des questionnaires distribués, nous avons réalisé une synthèse avec les informations recueillies lors des entretiens. Cette synthèse servira de recueil récapitulatif qui contient les différents volets d'activités, les dimensions et les indicateurs clés de notre travail.

#### Recueil récapitulatif des besoins

##### Contexte du recueil :

Les cadres dirigeants et décideurs de la Direction des Ressources Humaines au sein de NAFTAL SPA ont émis un vif désir d'exploiter les informations contenues dans le système opérationnel (Gestion des Ressources Humaines et Gestion de la Paie) ainsi que des multiples fichiers de différents types (en particulier de type Excel) afin de pouvoir les utiliser dans leur prise de décisions. Nous avons pu identifier quatre principaux processus clés dans notre travail.

##### Les processus clés :

Les quatre principaux processus clés qui intéressent les décideurs de la Direction Ressource Humaine sont les suivants :

- ◇ Effectif : il est important dans une entreprise, de pouvoir connaître à tout moment le nombre d'employés global.
- ◇ Absentéisme : consiste un aspect important dans le comportement des salariés mesurable à travers le bilan social.
- ◇ Conge : nombre d'effectif en conge.
- ◇ Mission : nombre d'effectif en mission.



**Les métriques d'évaluation de performances :**

Les décideurs veulent disposer d'informations sur les quatre processus cités auparavant selon les axes : Branche, sexe, age, tranche expérience, type agent, tranche salaire, niveau académique, type absence, mouvement, CSP, dans une période de temps.

***Volet Effectif :***

- Quel est le nombre d'effectif global par Branche, sexe, age, tranche expérience, tranche salaire, niveau académique, mouvement, CSP, type agent, temps ?

***Volet Absentéisme :***

- Quel est l'effectif en absence par Branche, sexe, age, CSP, type absence, type agent, temps ?
- Quelles sont le montant d'absence et la durée d'absence par sexe, type absence, type agent, temps ?

***Volet Mission :***

- Quel est l'effectif en mission par Branche, sexe, age, CSP, temps?
- Quels sont les frais et le coût de mission par Branche, temps?

***Volet Conge :***

- Quel est l'effectif en conge par Branche, sexe, age, CSP, temps ?

**Tableau 6 : Tableau récapitulatif des besoins**

**IV. CONCLUSION :**

Au cours de ce chapitre, nous avons pu identifier une liste de besoins des décideurs grâce à différentes approches d'enquête.

C'est à partir de ces besoins que seront construits les indicateurs et les axes d'analyse de notre modélisation dimensionnelle (chapitre suivant).

Cette phase du projet de construction de l'entrepôt de données constitue une étape très importante car elle met en évidence les bases de notre modèle dimensionnel, qui fera objet du prochain chapitre.

---

---

**PARTIE II :**

---

**CHAPITRE 3 :**  
**CONCEPTION DU SYSTEME**

---

---

Au cours de ce chapitre nous traiterons de la conception logique et physique de notre entrepôt de données. Nous aborderons principalement la modélisation dimensionnelle qui est la plus souvent associée aux entrepôts de données.

## I. MODELISATION DIMENSIONNELLE:

### I.1. Introduction :

Tout au long de son activité, la direction Natal Spa a implémenté plusieurs SI opérationnels. Ces SI s'appuient principalement sur le modèle relationnel mis en œuvre par plusieurs systèmes de gestion de bases de données (SGBD) comme Oracle. Dans notre cas d'étude, la Direction des Ressources Humaines possède son propre système transactionnel à l'échelle nationale fonctionnant sous SGBD Oracle. Aussi, la gestion de la paie possède un système transactionnel à l'échelle national fonctionnant sous paradox. Notre objectif sera donc de proposer une modélisation multidimensionnelle des données et des schémas de bases de données afin de permettre aux utilisateurs de bien comprendre et exploiter au mieux les données stockées au niveau du système opérationnel paie et système gestion du personnel.

Dans la modélisation dimensionnelle on distingue les faits et les attributs. Dans notre cas, les faits que nous traiterons sont de type numérique. Les attributs sont généralement un champ textuel. L'attribut n'est donc pas mesurable à l'inverse du fait. Ces attributs sont organisés en dimensions. Les principales dimensions identifiées dans la base de données de PAIE et R.H sont : Agent, Position, Temps, Structure, Catégorie socioprofessionnelle, Tranche d'âge, Tranche salaire, Tranche expérience, Tranche congé, Branche, Sexe.

Dans ce chapitre, nous présenterons la modélisation multidimensionnelle des processus clés cités dans le chapitre précédent.

### I.2. Processus de modélisation :

La modélisation conceptuelle de chaque schéma multidimensionnel pour les activités citées ci-dessus s'effectue en quatre étapes successives. Ces étapes ont déjà été présentées dans la première partie, voici un bref rappel :

- **Choix de processus d'activité à modéliser** : représente un processus opérationnel important pour l'organisation. Pour notre cas, les processus d'activités sont effectif, absentéisme, congé, mission.

- *Choix de grain du processus d'activité* : est le niveau de détail atomique figurant dans la table de faits d'un processus donné. 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 la quantité de masse salariale, le netpay et nombre effectif.

### I.3. Formalisme utilisé :

Avant d'entamer la modélisation dimensionnelle des activités, il est nécessaire de décrire le formalisme à utiliser. La légende suivante illustre ce formalisme.

<b>Nom de la table de fait</b>
<u>Concaténation des clés</u>
Liste des faits

Formalise adopté pour la table de faits

Nom de la table de dimension
<u>Clés primaires</u>
Liste des attributs

Formalise adopté pour la table de dimension

\_\_\_\_\_

Formalise adopté pour la relation entre les deux tables

#### **I.4. Modélisation dimensionnelle des aspects:**

##### **A. Modélisation dimensionnelle de l'aspect « effectif » :**

###### **A.1. Présentation de l'aspect « Effectif » :**

L'effectif est le premier aspect à modéliser dans la conception de notre entrepôt de données.

L'aspect effectif concerne les agents de Natal sur le territoire national.

Nous sommes en mesure de savoir en détail quel est l'effectif en mission, en congé, en absence, dans chaque branche et structure et durant quels mois.

###### **A.2. Grain d'aspect Effectif:**

Dans le cadre de la gestion Ressource Humaine, il est important de connaître l'effectif de l'entreprise. C'est pourquoi nous avons commencé par l'analyse de cet aspect .Ainsi, chaque enregistrement de la table des faits est une mesure réalisée par CSP, Sexe, Tranche d'age et par Date.

De là, nous obtenons notre premier modèle dimensionnel en flocon concernant l'aspect « effectif ».

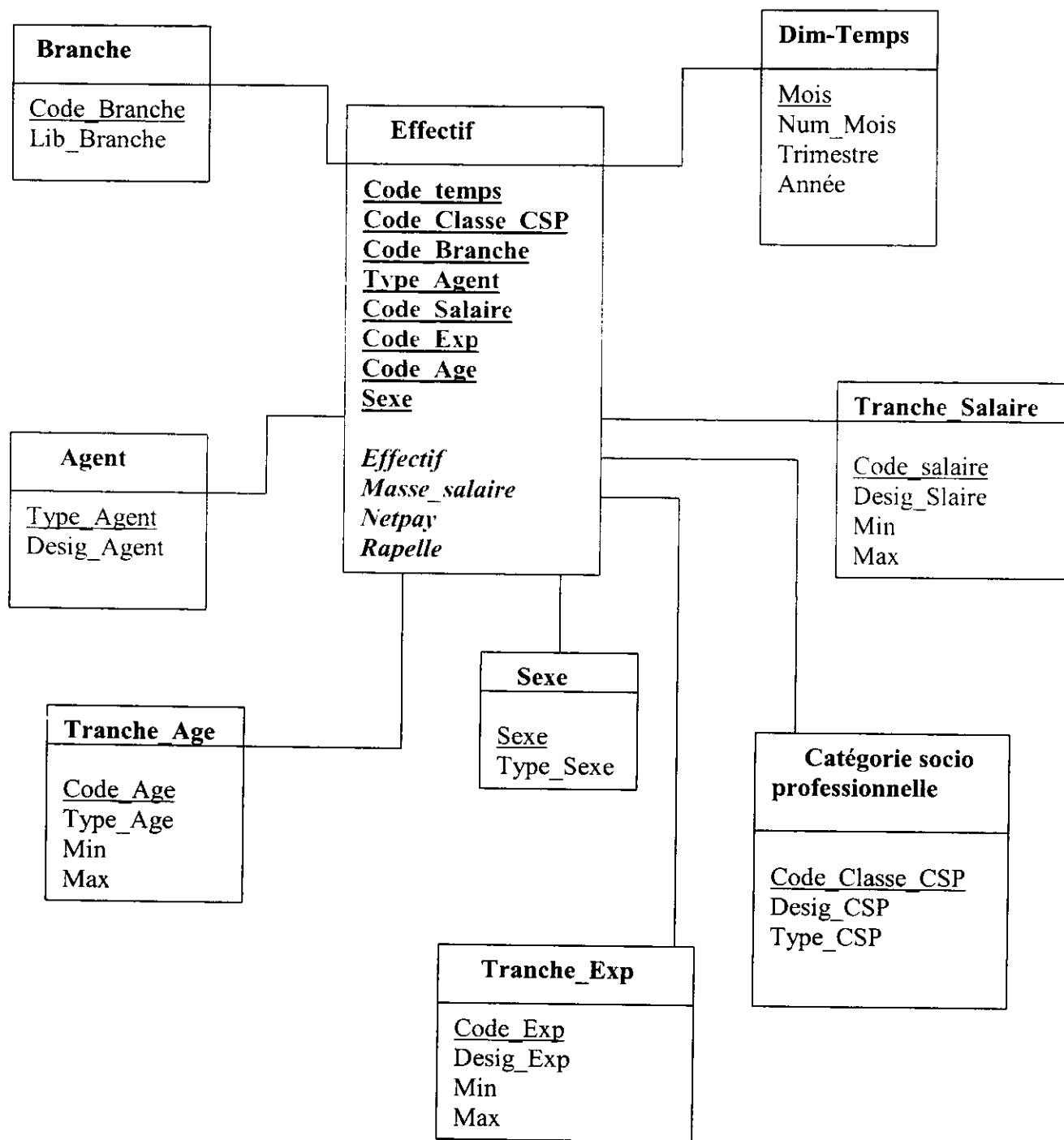


Fig23. Modèle dimensionnel en étoile de l'aspect « Effectif »

### A.3. Les dimensions de l'aspect «Effectif » :

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

#### A.3.1. Dimension « CSP » :

La dimension CSP «catégorie socioprofessionnelle» décrit l'ensemble de ce que contient l'entreprise comme catégorie socioprofessionnelle. Dans notre cas les catégories socioprofessionnelles sont essentiellement les cadres supérieurs, cadres, maîtrises, exécution et apprentis.

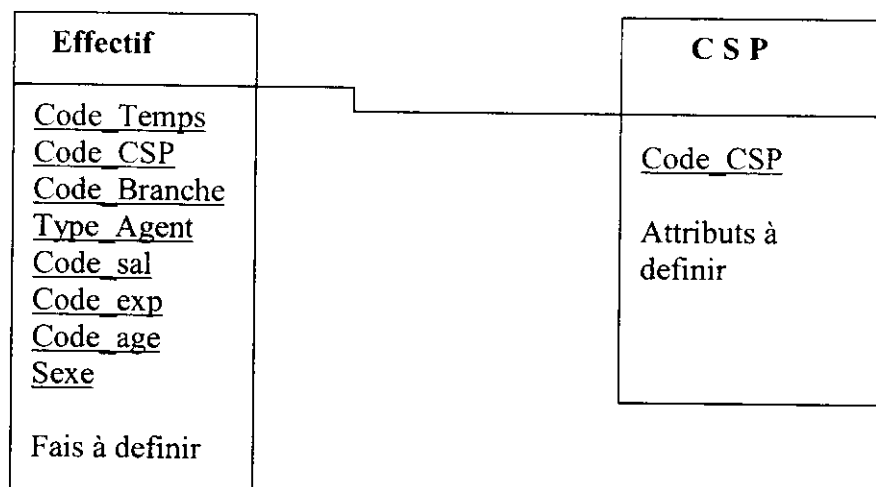


Fig. 24 : Dimension CSP pour l'aspect Effectif



### A.3.2. Dimension « Temps » :

Selon Ralph Kimball, la dimension temps est : « *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* » [Kim, 00].

Pour l'aspect effectif de NAFTAL SPA, l'enregistrement mensuel des effectifs permet de mieux les examiner. Les analystes s'intéressent les plus souvent aux effectifs mensuelles, par trimestre et bien évidemment annuelles. Mais pour mieux observer ces effectifs nous avons opté pour l'enregistrement mensuel des mouvements. Donc chaque enregistrement de la table de dimension Temps représente un mois.

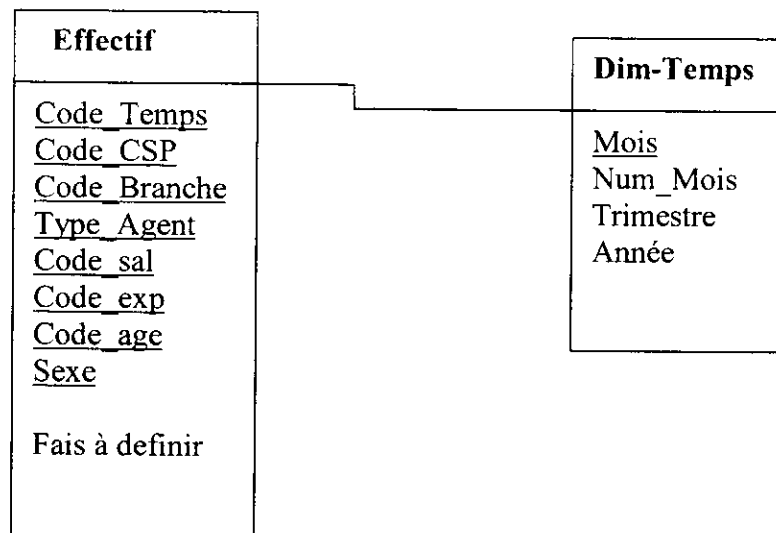


Fig. 25 : Dimension temps pour l'aspect Effectif

Dans cette dimension, le champ « Mois » c'est une date, par exemple '01/02/2006'. Les autres champs serviront à indiquer les effectifs pendant une période donnée.

### A.3.3. Dimension « type agent » :

La dimension type agent représente pour NAFTAL SPA les types des agents car on destine deux types (permanents ou temporaire).

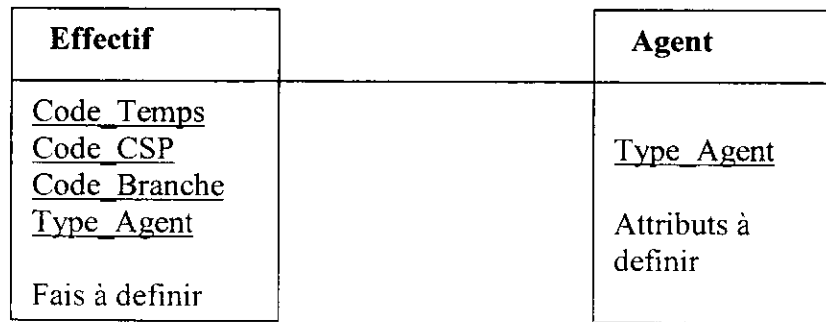


Fig. 26 : Dimension Agent pour l'aspect Effectif

La dimension «type agent » a la particularité d'être une dimension à évolution lente. Les données concernant un type agent ont la particularité d'avoir une basse fréquence de mise à jour.

#### A.3.4. Dimension « Structure » :

Les structures sont classées selon la branche à laquelle ils se réfèrent. Pour notre cas les branches sont Direction Centrale, Bitumes, Branche Carburant, Branche GPL et Branche COM.

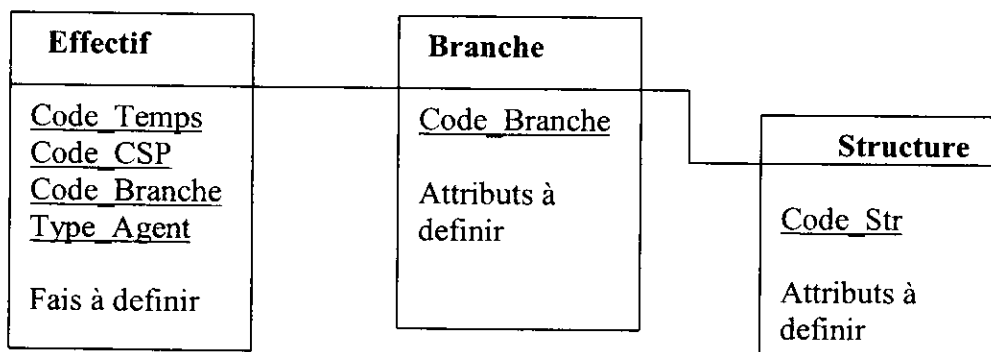


Fig. 27 : Dimension Structure pour l'activité Effectif

#### A.3.5. Dimension « Tranche salaire » :

Pour la société NAFTAL SPA , il existe plusieurs tranches de salaire pour les différents agents de l'entreprise, Ces tranches se répartissent en : <10000DA, 10000-20000DA, 20000-30000DA,30000-40000DA,40000-50000DA,50000-60000DA,60000-70000DA,70000-80000DA,80000-90000DA,90000-100000DA,100000-200000DA,>200000DA.Cette

dimension est importante dans le cas où on veut analyser les créances de l'entreprise par agent et par l'activité de l'agent.

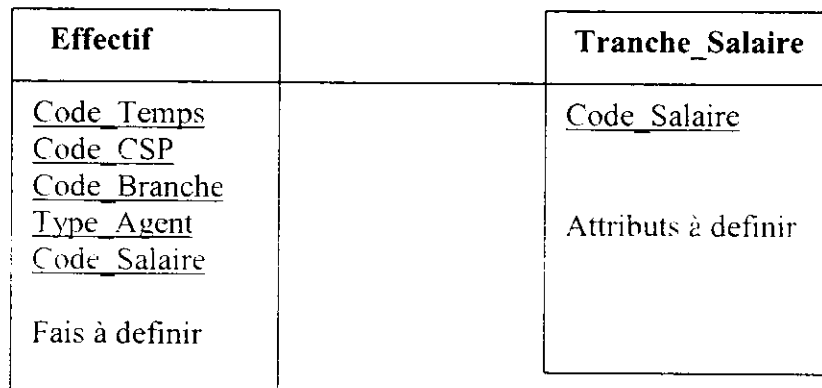


Fig. 28 : Dimension Tranche\_Salaire pour l'activité Effectif

#### A.3.6. Dimension « Tranche Expérience » :

La dimension « Tranche Expérience » nous renseigne sur l'expérience de l'agent dans et hors de l'entreprise lors de l'opération de calcul de l'effectif.

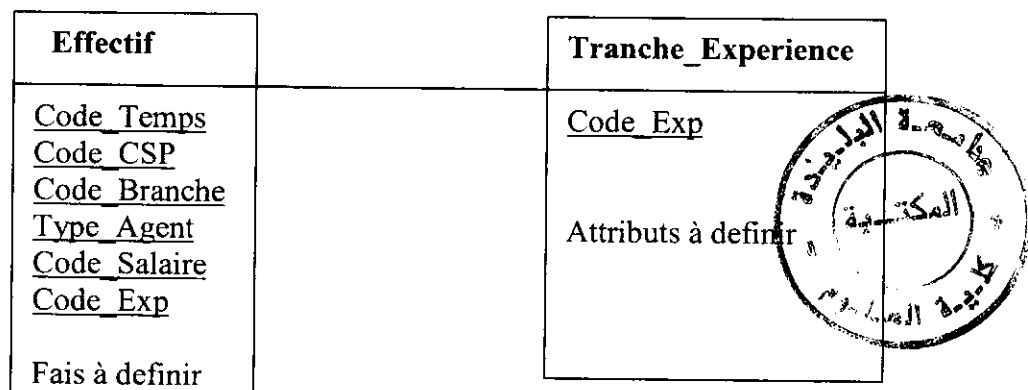


Fig. 29 : Dimension Tranche\_Experience pour l'activité Effectif

La dimension « Tranche Expérience » est une dimension à évolution lente. Les données concernant l'expérience d'un agent ont la particularité d'avoir une basse fréquence de mise à jour.

### **A.3.5. Dimension « Tranche age » :**

Pour la société NAFTAL SPA, il existe plusieurs tranches d'âges pour les différents agents de l'entreprise, Ces tranches se répartissent en : moins de 20 ans, 20-25ans, 25-30ans, 30-35ans, 35-40ans, 40-45ans, 45-50ans, 50-55ans, 55-60ans, plus de 60ans. Cette dimension est importante dans le cas où on veut analyser l'effectif et la masse salariale pour chaque tranche d'âge et par chaque type agent.

### **A.4. Les faits mesurés de l'aspect « Effectif » :**

Les analystes et les cadres dirigeants de la Direction des Ressources Humaines s'intéressent tout particulièrement à connaître l'effectif, la masse salariale et le netpay au niveau des branches et des structures. Ainsi, la table de faits doit contenir des indicateurs relatifs aux effectifs et au montant de la masse salariale et de netpay. Ces faits mesurables seront notés dans la table de faits par «*Effectif*» pour l'effectif de l'entreprise, «*Masse\_Salariale*» pour le montant de masse salariale et «*Netpay*» pour le montant de netpay.

Tous ces faits sont additifs sur toutes les dimensions. Ce qui facilite les coupes en tranches et en dés de la table.

## **B. Modélisation dimensionnelle de l'aspect « Absentéisme » :**

### **B.1. Présentation de l'aspect « Absentéisme » :**

L'absentéisme sera notre deuxième volet à modéliser lors de la conception de notre entrepôt de données. Ce dernier constitue un aspect important dans le comportement des salaires mesurable à travers le bilan social. Nous sommes en mesure de savoir en détail quels est l'effectif en absence, dans quel branche et structure et durant quel mois.

### **B.2. Grain d'aspect Absentéisme:**

Dans le cadre de l'estimation de nombres des agents en absence de NAFTAL SPA, il est impératif de connaître les différents agents de l'entreprise. Pour chaque absence réalisée, nous avons la possibilité de déterminer le type d'absence auquel il a été absent, en quelle durée et à quel montant. Le choix de la granularité consiste à indiquer qu'un nouvel enregistrement sera créé dans la table des faits pour chaque absence. Ainsi, chaque enregistrement de la table des faits est une ligne d'absence réalisée par type d'agent agent par CSP par branche et par date. De là, le diagramme de notre premier schéma dimensionnel commence à prendre forme.

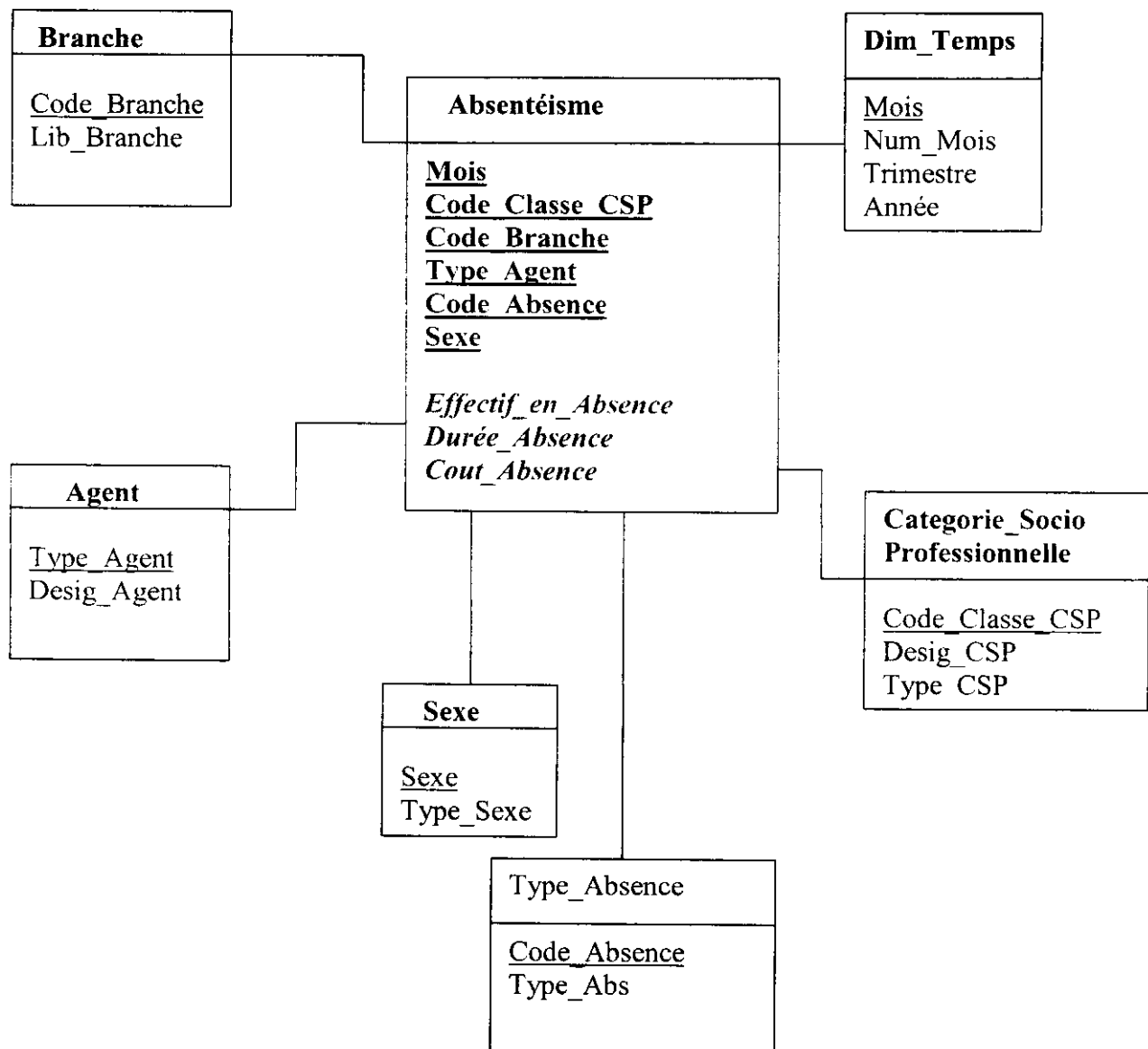


Fig. 30 : Modèle dimensionnel en étoile de l'aspect « Absentéisme »

### B.3. Les dimensions de l'aspect « Absentéisme » :

La granularité de la table des faits indique explicitement les dimensions de base. Parmi ces dernières, on trouve quelques unes entrant en commun avec l'aspect effectif.

#### B.3.1. Dimensions conformes :

Les dimensions «*Branche* », «*Temps* », «*CSP* », «*Type Agent*», «*Sexe*» et «*tranche age*» ont déjà été étudiées au point A.3. Ces dimensions sont dites « conformes » dans la mesure où elles sont identiques pour les deux tables des faits étudiées. Cette conformité rend

possible le rapprochement des deux tables de faits « Effectif » et « Absentisme » au travers des dimensions qu'elles ont en commun.

### B.3.2. Dimension « Type Absence » :

Cette dimension représente le type d'une absence réalisée par un agent. On distingue plusieurs types d'absences au niveau de Naftal Spa. Ces types se répartissent en Absence non autorisées. Absences autorisées. Abs légales. Mesures disciplinaires. Retenues mois incomplet et Cumulées Abs/sanct.

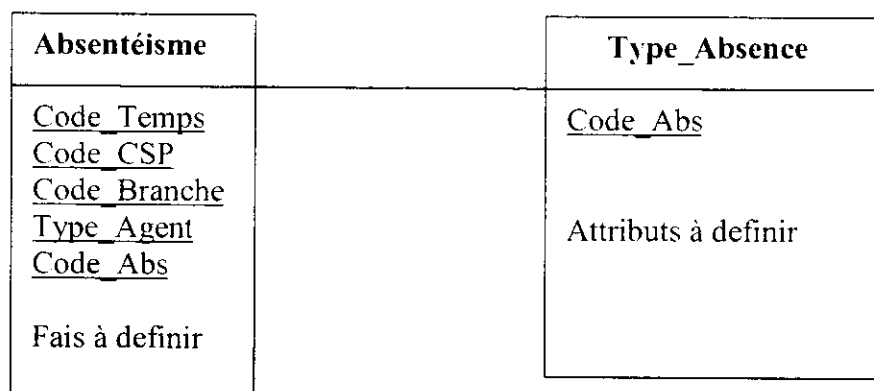


Fig. 31 : Dimension Type\_Absence pour l'aspect Absentisme

### B.4. Les faits mesurés de l'aspect « Absentisme » :

Dans l'absentisme, il est impératif de connaître les détails des absences des agents, entre autre l'effectif en absence, la durée d'absence et le montant des absences. Pour cela la table des faits doit contenir des faits reflétant les données essentielles à l'analyse.

Les faits mesurables pour ce volet sont l'effectif en absence «Effectif\_en\_Absence», la durée d'absence « Duree\_Absence » et le montant d'absence « Mont\_Absence ».

Ces trois faits appliqués aux axes définis précédemment permettent de réaliser des coupes de la table des faits en tranches et en dés.

## C. Modélisation dimensionnelle de l'aspect « Conge » :

### C.1. Présentation de l'aspect « Conge » :

Le troisième volet dans la conception de notre entrepôt de données correspond à l'aspect conge. Ce dernier constitue un aspect important dans la mesure de déterminer en détail l'effectif en conge et dans quelle période donnée.

### C.2. Grain d'aspect :

Dans le cadre de l'estimation de nombres des agents en conge de NAFTAL SPA, il est impératif de connaître les différentes structures auquel a eu conge. Pour chaque conge réalise, nous avons la possibilité de déterminer le type d'agent auquel il a eu conge. Le choix de la granularité consiste à indiquer qu'un nouvel enregistrement sera créé dans la table des faits pour chaque conge. Ainsi, chaque enregistrement de la table des faits est une ligne de conge réalisée par type d'agent agent par CSP par branche et par date. De là, le diagramme de notre premier schéma dimensionnel commence à prendre forme.

Nous présentons le modèle dimensionnel en flocon de l'activité « conge ».

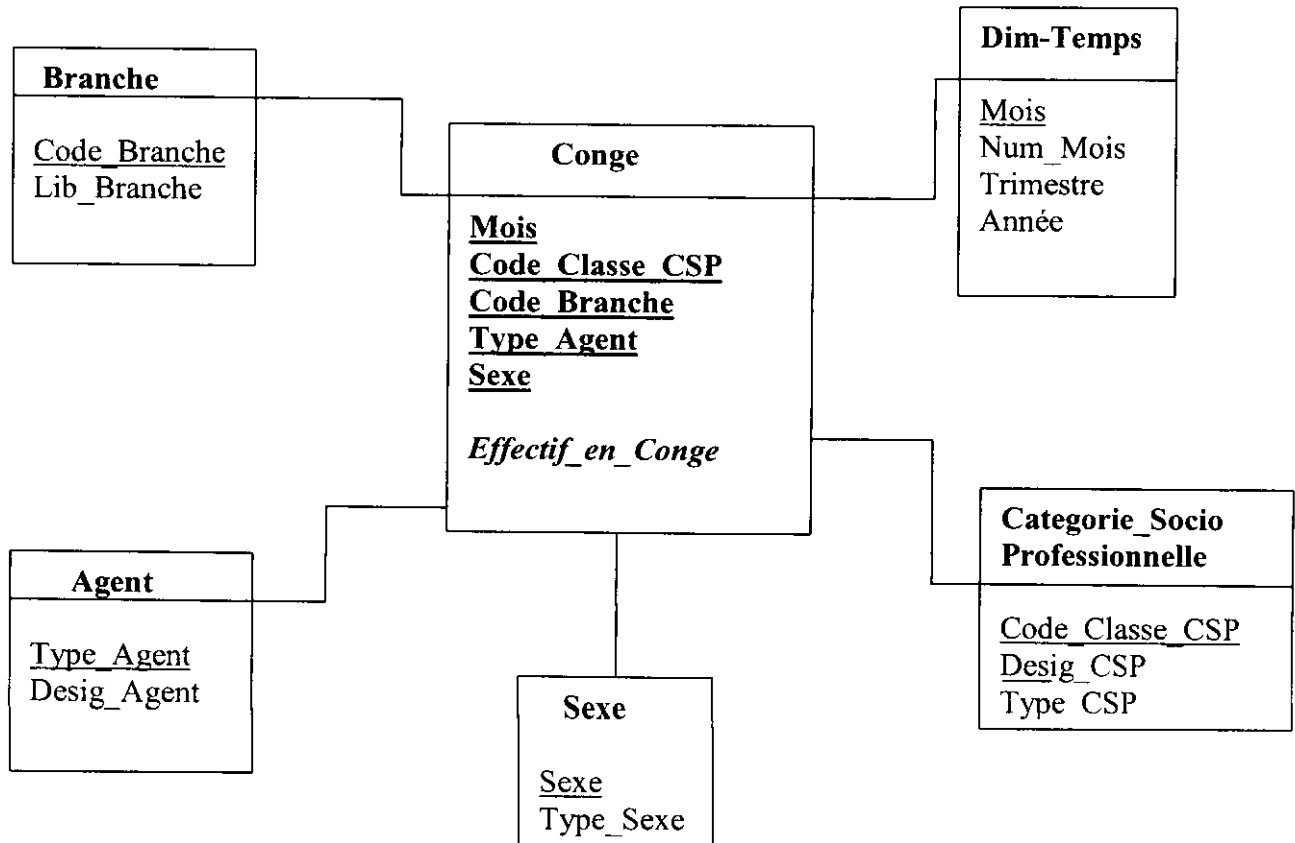


Fig. 32: Modèle dimensionnel en étoile de l'aspect « conge »

### **D.3. Les dimensions de l'aspect «Conge » :**

#### ***D.3.1. Dimensions conformes :***

Les dimensions « *Type agent* », « *Temps* », « *CSP* », « *Sexe* », « *Branche* » ont déjà été étudiées au point 4.3.

### **D.4. Les faits mesurés de l'aspect « conge » :**

Dans l'aspect conge il est impératif de connaître les détails des congés effectués par les agents. entre autre l'effectif en conge. Pour cela la table des faits doit contenir des faits reflétant les données essentielles à l'analyse.

Le seul fait mesurable pour cette activité est l'effectif en conge «Effectif\_en\_Conge ».

## **D. Modélisation dimensionnelle de l'aspect « Mission » :**

### **D.1. Présentation de l'activité « Mission » :**

Le quatrième et dernier aspect dans la conception de notre entrepôt de données correspond à l'aspect mission. Ce dernier constitue un aspect important dans la mesure de déterminer en détail l'effectif en mission et dans quelle période donnée.

### **D.2. Grain d'aspect Mission :**

Le grain choisi est important car il a son incidence sur la taille et le volume de la base de données.

Dans le cadre de l'estimation de nombres des agents en mission de NAFTAL SPA, Pour chaque mission réalise, nous avons la possibilité de déterminer la structure auquel a eu la mission, à quel frais et à quel coût. Le choix de la granularité consiste à indiquer qu'un nouvel enregistrement sera créé dans la table des faits pour chaque mission. Ainsi, chaque enregistrement de la table des faits est une ligne de mission réalisée par un type agent par CSP par structure et par date. De là, le diagramme de notre premier schéma dimensionnel commence à prendre forme, la figure suivante illustre le modèle dimensionnel de base.

Ainsi nous présentons le modèle dimensionnel en flocon de l'activité « mission ».



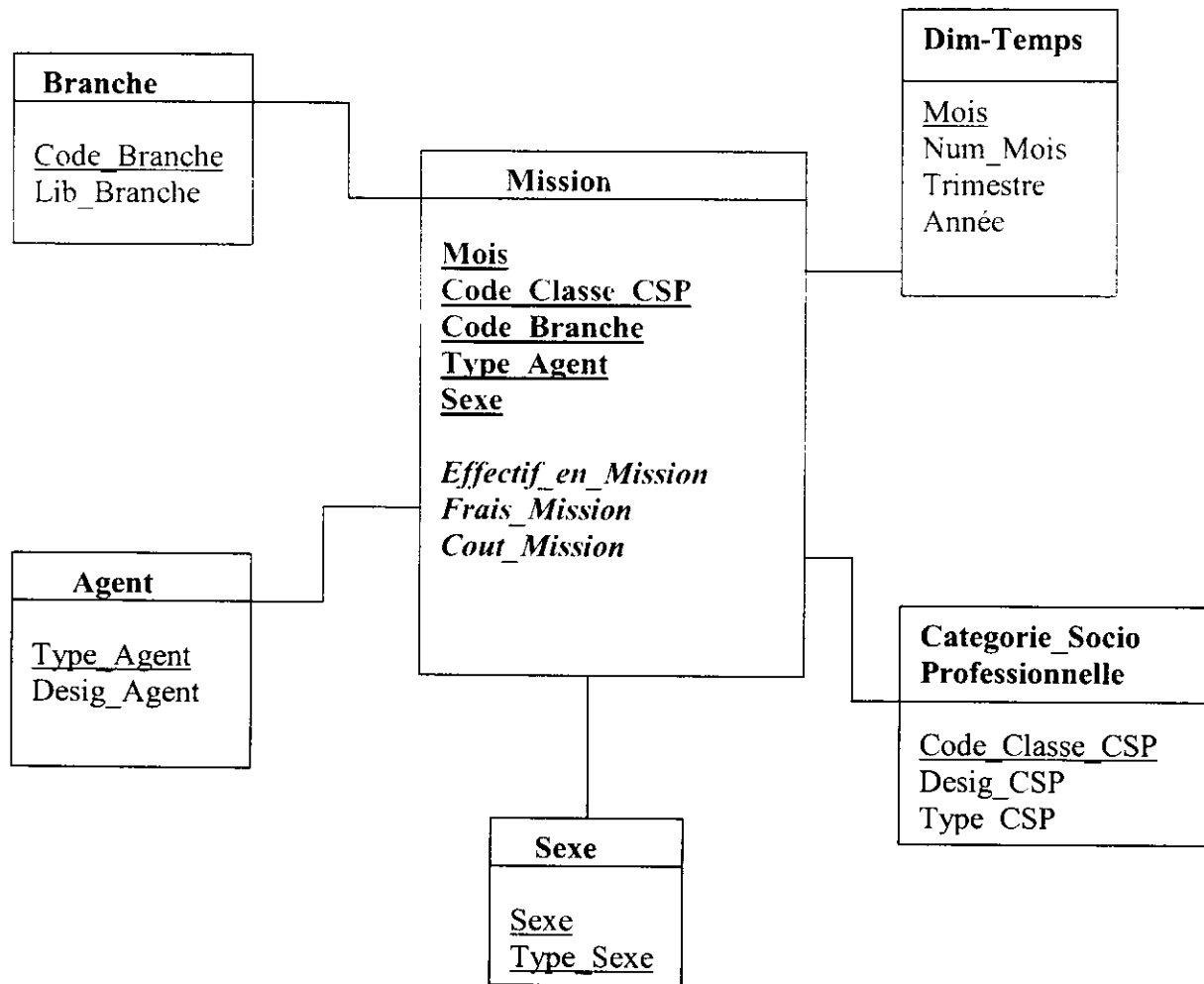


Fig. 33 : Modèle dimensionnel en étoile de l'aspect « Mission »

Notre choix de la granularité correspondra à l'ajout d'un nouvel enregistrement dans la table des faits pour chaque nouvelle mission.

### D.3. Les dimensions de l'aspect «Mission » :

#### D.3.1. Dimensions conformes :

Les dimensions « *Tranche age* », « *Temps* », « *CSP* », « *Sexe* », « *Branche* » ont déjà été étudiées au point 4.3.

**D.4. Les faits mesurés de l'aspect « Mission » :**

Dans l'aspect mission il est impératif de connaître les détails des missions effectuées par les agents, entre autre l'effectif en mission, frais de mission et coût de mission. Pour cela la table des faits doit contenir des faits reflétant les données essentielles à l'analyse.

Les faits mesurables pour cette activité sont l'effectif en mission «Effectif\_en\_Mission», les frais de mission « Frais\_Mission» et le coût de mission « Cout\_Mission».

Ces trois faits appliqués aux axes définis précédemment permettent de réaliser des coupes de la table des faits en tranches et en dés.

**II. MISE EN OEUVRE DE L'ARCHITECTURE :****II.1. Présentation :**

Dans cette partie, nous présenterons le mécanisme de mise en œuvre du système décisionnel. Ce mécanisme, que nous étudierons dans les points suivants, est organisé en plusieurs composants visibles dans l'architecture des systèmes décisionnels basés sur le concept « Entrepôt de données ».

Pour cette opération, nous suivrons une méthodologie composée de trois éléments essentiels : la base de données et méta-données, un ensemble de transformations appliquées aux données sources afin d'alimenter notre entrepôt de données et des outils logiciels permettant de présenter les données stockées dans le Data Warehouse sous forme d'informations utiles.

**II.2. Architecture détaillée du Data Warehouse :**

La figure suivante illustre l'architecture détaillée de notre entrepôt de données. Cette architecture comprend les différents composants du système décisionnel.

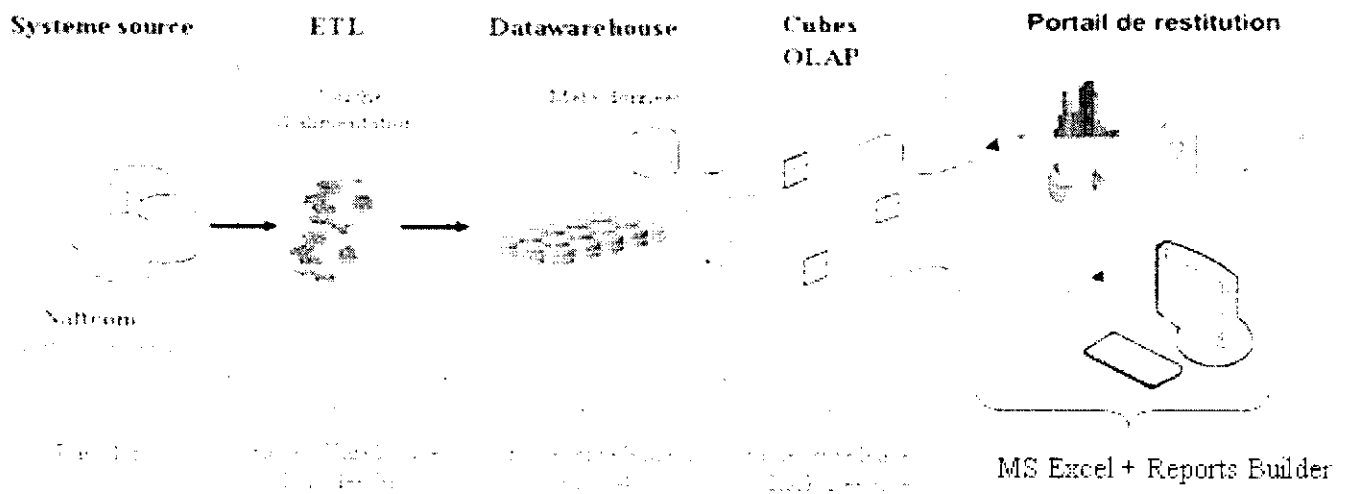


Fig 34 : Architecture détaillée du Data Warehouse

### II.3. Construction de l'entrepôt de données :

La construction de l'entrepôt de données consiste à transformer la conception logique de ce dernier en une base de données physique. Cette transformation impliquera certains changements dans la structure du modèle afin de l'adapter aux particularités du SGBD utilisé.

La différence essentielle entre le modèle physique et logique est plutôt due à la spécification détaillée des caractéristiques de la base de données physique, à commencer par le type des données, les flux, la segmentation des tables, les paramètres de stockage des tables et la répartition multi-disques.

### II.3.1. Liste des tables et les attributs de l'entrepôt de données:

#### Les tables de dimensions :

Table	Attributs	Description	Type
Dim_Temps	Mois	Mois de l'année	Date
	Mo_Month_Name	Nom du mois	Varchar2 (9)
	Num_Mois	Numéro de Mois de l'année	Varchar2 (2)
	Trimestre	Numéro du trimestre de l'année	Varchar2(1)
	Année	Année	Varchar2 (4)
Tranche_Conge	Code_Conge	Identifiant conge	Varchar2(3)
	Type_Conge	Désignation CDS	Varchar2(25)
	Min	NB de jour minimum de conge	Varchar2(4)
	Max	NB de jour maximum de conge	Varchar2 (4)
Agent	Type_Agent	Identifiant de type de l'agent	Varchar2 (2)
	Desig_Agent	Désignation agent	Varchar2 (10)
Branche	Code_Branche	Identifiant de branche	Varchar2 (4000)
	Lib_Branche	Désignation de branche	Varchar2 (30)
NC	Code_NC	Identifiant de niveau classification	Varchar2 (10)
	Salaire	montant	Varchar2 (8)
Tranche_Exp	Code_Exp	Identifiant de l'Experience	Varchar2 (2)
	Type_Exp	Type d'expérience	Varchar2 (30)
	Min	Valeur minimum de la tranche expérience	Varchar2 (3)
	Max	Valeur maximumde la tranche Experience	Varchar (3)
Sexe	Sexe	Identifiant sexe	Varchar2 (2)
	Type_Sexe	Désignation du type sexe	Varchar2 (10)

Categori_Socioprofessionnelle	Code_Classe_csp	Identifiant de classe Catégorie socioprofessionnelle	Varchar2 (2)
	Desig_csp	Désignation de classe Catégorie socioprofessionnelle	Varchar2 (50)
	Type_csp	Type de Catégorie socioprofessionnelle	Varchar2 (1)
Mouvement	Code_MV	Identifiant mouvement	Varchar2 (2)
	Type_MV	Le type de mouvement	Varchar2 (1)
	Desig_MV	Désignation mouvement	Varchar2 (30)
Structure	Code_Str	Identifiant structure	Varchar2 (3)
	Lib	Libellé structure	Varchar2 (60)
	Libdivision	Libellé division	Varchar2 (10)
Tranche_Age	Code_Age	Identifiant d'âge	Varchar2 (2)
	Type_Age	Désignation du code age	Varchar2 (25)
	Min	Valeur minimale de la tranche age	Varchar2 (10)
	Max	Valeur maximale de la tranche age	Varchar2 (10)
Type_Absence	Code_Absence	Identifiant d'absence	Varchar2 (4)
	Type_ABS	La valeur d'absence	Varchar2 (30)
Niveau_Academique	Code_NA	Identifiant du niveau académique	Varchar2 (5)
	Type_NA	Le type du niveau académique	Varchar2 (5)
	Desig_NA	Désignation du niveau académique	Varchar2 (30)
Position	Code_Position	Identifiant de la position	Varchar2 (10)
	Type_Position	Type de la position	Varchar2 (15)
	Lib_Position	Désignation de la position	Varchar2 (100)

Tableau 7 : Liste des tables de dimensions de l'entrepôt de données

## Les tables de faits :

Table	Attributs	Description	Type
Mission	Code_Str	Identifiant du structure	Varchar2(25)
	Code_Age	Identifiant du l'age	Varchar2 (2)
	Code_Csp	Identifiant du catégorie socioprofessionnelle	Varchar2 (2)
	Code_Temps	Identifiant du temps	date
	Sexe	Identifiant du sexe	Varchar2 (2)
	<i>Effectif_en_Mission</i>	Nombre d'agent en mission	Number
	<i>frais_Mission</i>	Frais de mission	Number
	<i>Coût_Mission</i>	Coût de mission	Number
	<i>Distination</i>	Distination de la mission	Varchar2 (100)
Absentéisme	Code_Str	Identifiant du structure	Varchar2 (25)
	Code_Temps	Identifiant du temps	Date
	Code_Csp	Identifiant du catégorie socioprofessionnelle	Varchar2 (2)
	Code_Age	Identifiant du l'age	Varchar2 (2)
	Sexe	Identifiant du sexe	Varchar2(2)
	Type_Agent	Identifiant de l'agent	Varchar2(2)
	Code_ABS	Identifiant de l'absence	Varchar2 (25)
	Code_NC	Identifiant du niveau de classification	Varchar2 (10)
	<i>Effectif_en_Absence</i>	Nombre d'agent en absence	Number
	<i>MONT_ABS</i>	Montant d'absence des agents	Number
	<i>DUREE_ABS</i>	Coût d'absence des agents	Number

Conge	Code_Str	Identifiant structure	Varchar2 (25)
	Code_Temps	Identifiant du temps	Date
	Sexe	Identifiant du sexe	Varchar2(2)
	Type_Agent	Identifiant de l'agent	Varchar2(2)
	Code_Csp	Identifiant du catégorie socioprofessionnelle	Varchar2 (2)
	Code_Conge	Identifiant du conge	Varchar2 (3)
	<i>EFFECTIF_EN_CONGE</i>	Nombre d'effectifs en conge	Number
Effectif	Code_Str	Identifiant du structure	Varchar2(25)
	Code_Position	Identifiant de la position	Varchar2(10)
	Code_Temps	Identifiant du temps	date
	Code_Csp	Identifiant du catégorie socioprofessionnelle	Varchar2 (2)
	Code_MV	Identifiant de mouvement	Varchar2 (2)
	Sexe	Identifiant du sexe	Varchar2 (2)
	Code_NC	Identifiant du niveau de classification	Varchar2 (10)
	Type_Agent	Identifiant de l'agent	Varchar2(2)
	Code_SF	Identifiant de situation familiale	Varchar2(2)
	Code_salaire	Identifiant de Montant de salaire	Varchar2(2)
	Code_NA	Identifiant de niveau académique	Varchar2(2)
	Code_EXP	Identifiant de Experience	Varchar2(2)
	Code_Age	Identifiant de age	Varchar2(2)
	Code_Branche	Identifiant de la branche	Varchar2(2)
	<i>Effectif</i>	Nombre effectif global	Number
	<i>Masse_Salariale</i>	Montant de la masse salariale	Number
	<i>Rappel</i>	Montant de rappel	Varchar2(28)
<i>Netpay</i>	Montant de netpay	Number	

Tableau 8 : Liste des tables des faits de l'entrepôt de données

## **II.4. Construction des cubes OLAP :**

Afin de faciliter l'analyse de la gestion Ressource Humaine, il est indispensable de recourir à la modélisation multidimensionnelle des données. Grâce au concept OLAP, nous pouvons transformer les données stockées dans des bases de données relationnelles en informations ressources humaines pertinentes et faciles à exploiter, grâce à la création d'un cube de données.

Cette partie sera consacrée à la conception de notre base de données multidimensionnelle caractérisée par les cubes OLAP. Nous utiliserons pour cela la modélisation MOLAP « *Multidimensionnel OLAP* ». Notre choix de cette modélisation est justifié par la facilité de mise en place et de l'analyse multiaxes sur les données agrégées. Le serveur OLAP utilisé est « Oracle 9i release 2 OLAP Service ».

### **II.4.1. Choix des faits et dimensions :**

Lors de la modélisation dimensionnelle des activités « Effectif, absentéisme, conge, mission » nous avons pu distinguer les différentes tables de faits et tables de dimensions. Dans le tableau suivant nous illustrons les tables de faits choisies, leurs mesures et les dimensions requises pour l'analyse.



<i>Mesures</i>	<i>Dimensions requises</i>
<b>Effectif</b>	
<ul style="list-style-type: none"> <li>• Effectif global</li> <li>• Masse salariale</li> <li>• Netpay</li> </ul>	<ul style="list-style-type: none"> <li>- Type_agent → CSP → Position → Sexe → Tranche_age</li> <li>- Structure → Sexe → CSP → Type_agent → mouvement</li> <li>- Tranche_age → Sexe → CSP → NA → Type_agent</li> <li>- Tranche_exp → CSP → Sexe → Type_agent</li> <li>- Tranche_salaire → Sexe → CSP → Type_agent</li> </ul>
<b>Mission</b>	
<ul style="list-style-type: none"> <li>• Effectif en mission</li> <li>• Frais mission</li> <li>• Coût mission</li> </ul>	<ul style="list-style-type: none"> <li>- Tranche_age → CSP → Sexe</li> <li>- Structure → Tranche_age</li> <li>- CSP → Sexe → Structure</li> </ul>
<b>Absentéisme</b>	
<ul style="list-style-type: none"> <li>• Effectif en absence.</li> <li>• Montant absence.</li> <li>• Durée absence.</li> </ul>	<ul style="list-style-type: none"> <li>- Sexe → Structure → Type_agent → Type_abs → CSP</li> <li>- CSP → Sexe → Tranche_age → Type_abs → Type_agent</li> </ul>
<b>Conge</b>	
<ul style="list-style-type: none"> <li>• effectif en conge</li> </ul>	<ul style="list-style-type: none"> <li>- Sexe → Structure → Type_agent → CSP</li> <li>- CSP → Sexe → Tranche_age → Type_agent</li> </ul>

Tableau 9 : Tableau récapitulatif des choix de faits et des dimensions

**II.4.3. Schémas du modèle MOLAP des cubes :**

Les figures suivantes illustrent le modèle physique de chaque cube MOLAP selon la plate forme Oracle 9i OLAP Service. Chaque activité étudiée est représentée par un cube.

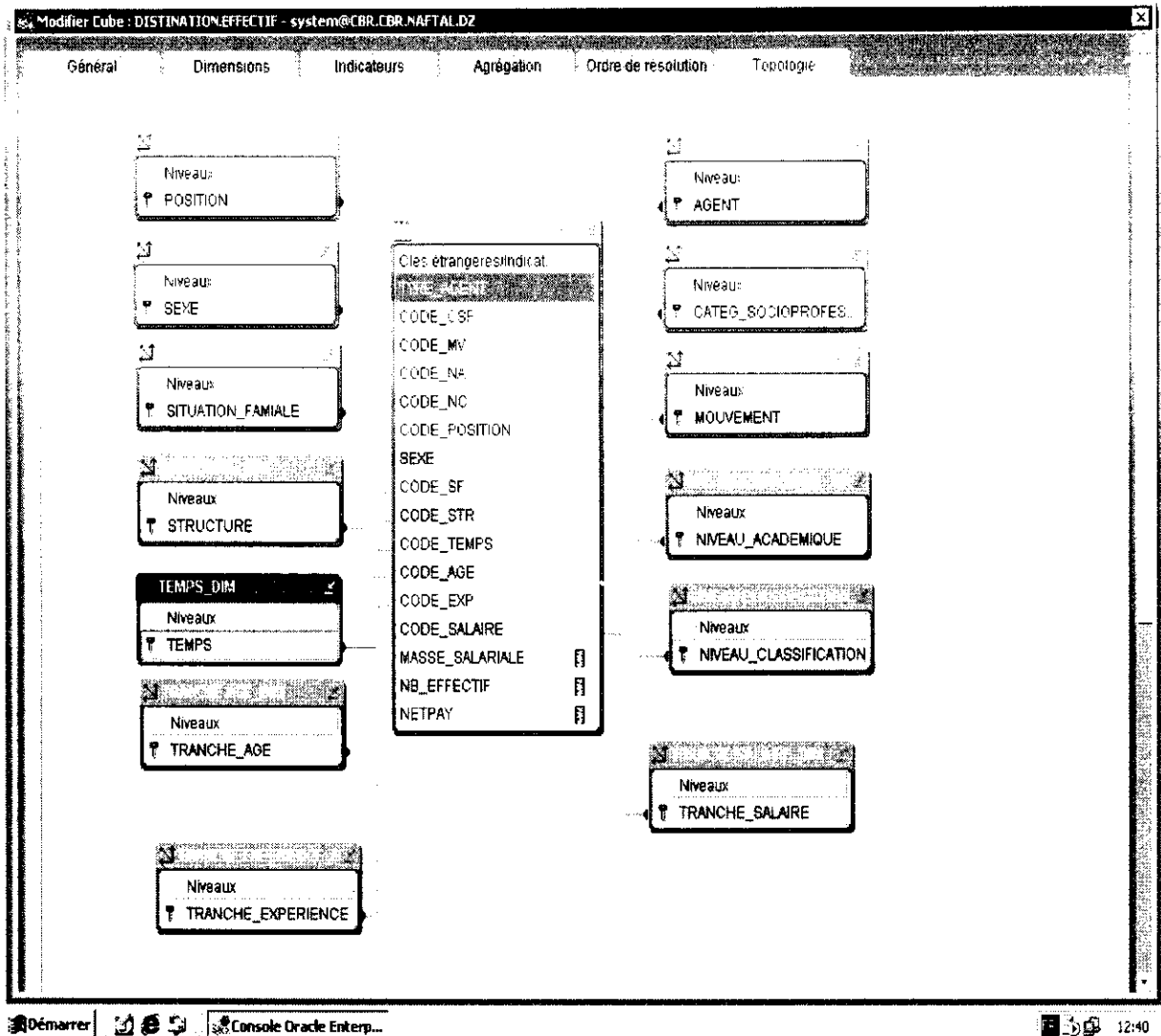


Fig 35 : Cube MOLAP de l'aspect « Effectif »

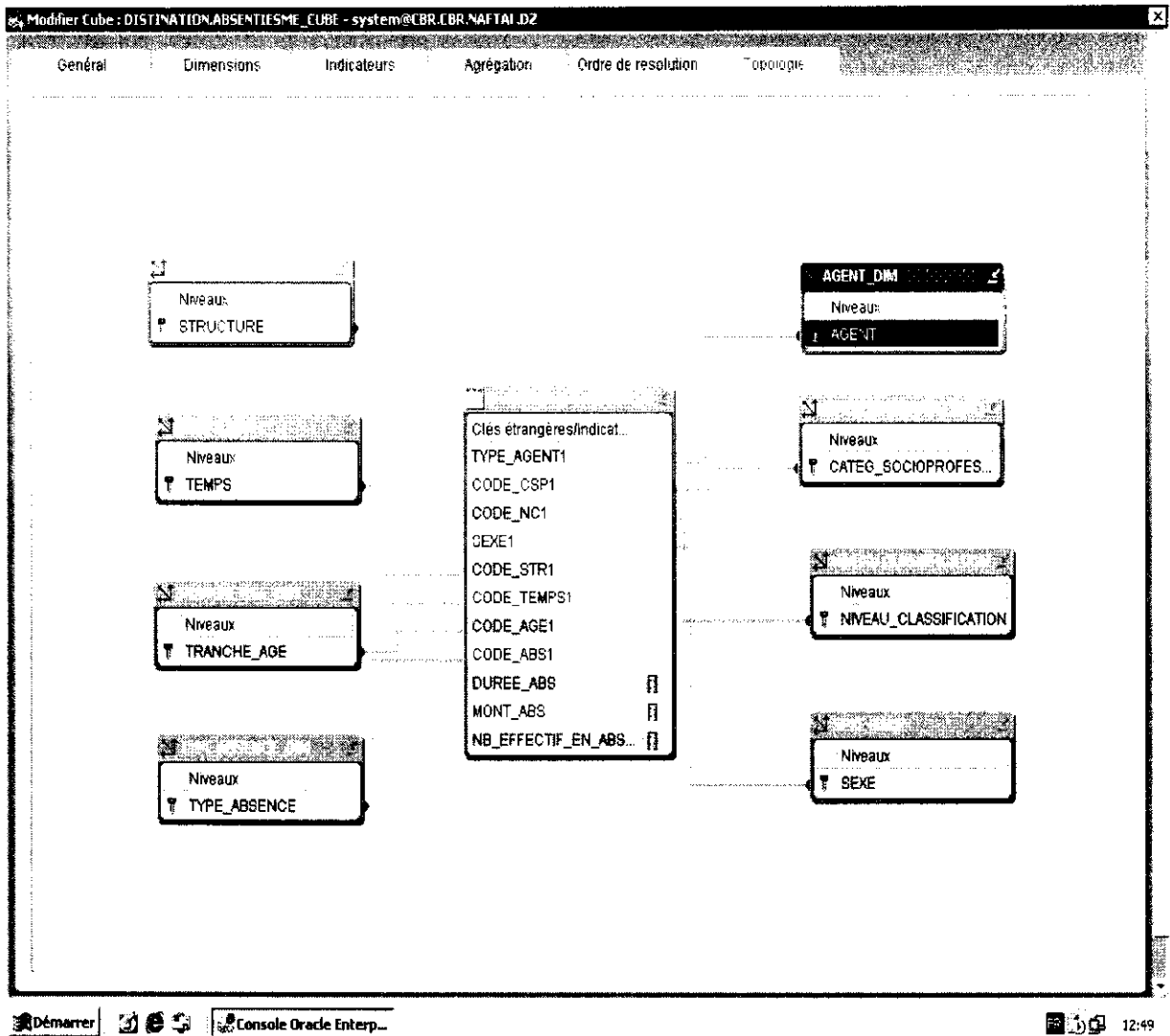


Fig. 36 : Cube MOLAP de l'aspect « Absentéisme »

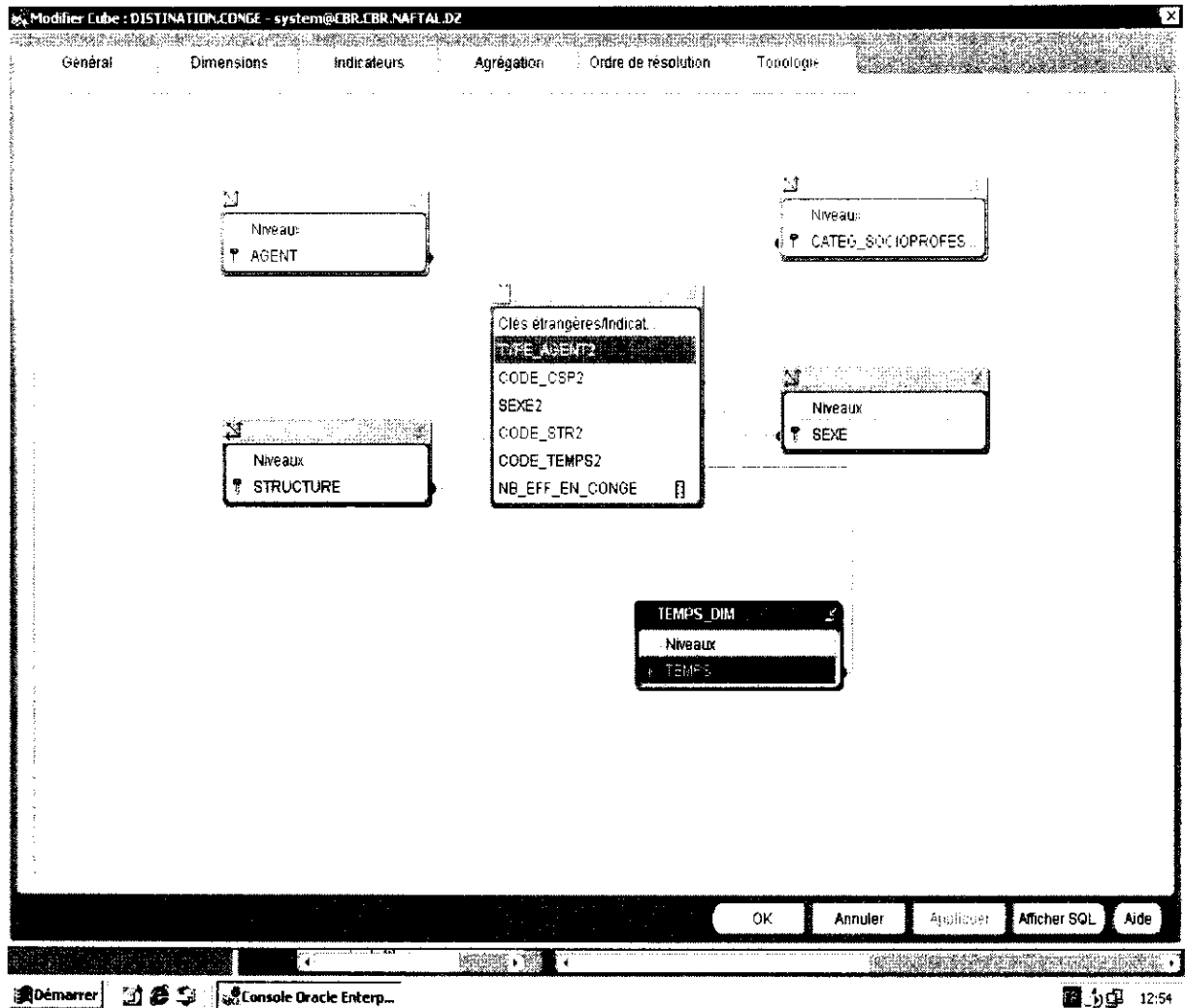


Fig. 37 : Cube MOLAP de l'aspect « Conge »

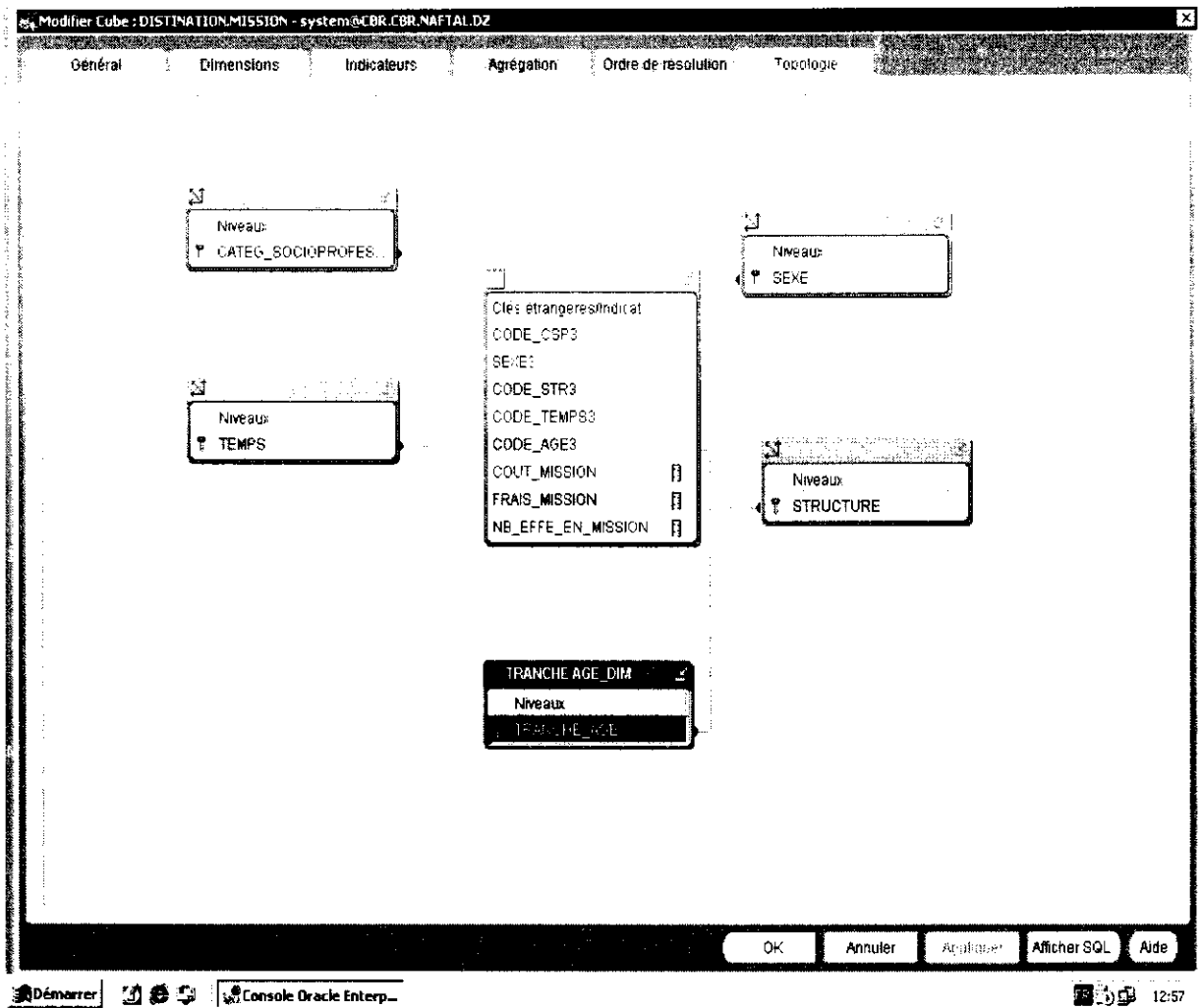


Fig. 38: Cube MOLAP de l'aspect « Mission »

## II.5. Construction de la zone d'alimentation :

### II.5.1. La zone d'alimentation :

Le processus d'alimentation consiste à transférer les données du système transactionnel (Paie et Grh) vers l'entrepôt de données et ceci après y avoir effectué quelques transformations dans leur structure. Cette phase de la conception de l'entrepôt de données est très importante et complexe. Elle représente environ 70% de l'opération de réalisation de notre entrepôt de données. Il est indispensable de déterminer quelles données seront chargées, quelles transformations y seront faites, la périodicité à laquelle les transferts s'effectueront. La figure suivante illustre le processus d'alimentation.

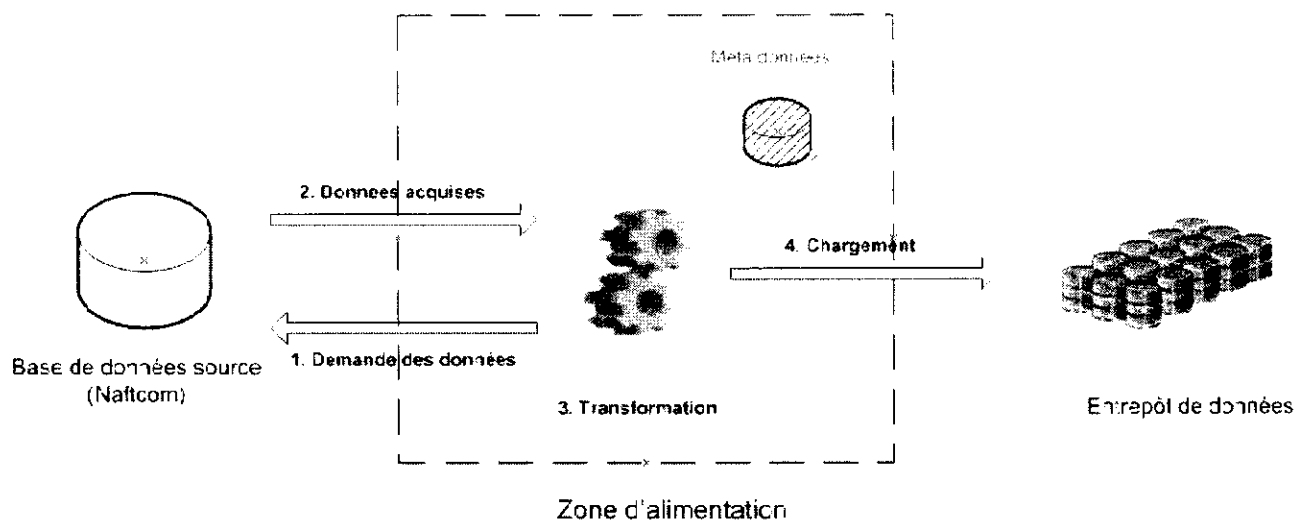


Fig 39 : Illustration du processus d'alimentation

Le processus de transformation est réalisé grâce à OWB ou « Oracle Warehouse Builder ». Cet outil permet de modéliser graphiquement l'opération de transformation des données à partir de leur source et de les charger vers l'entrepôt de données.

### II.5.2. Système source :

Comme cité auparavant dans la problématique, il existe plusieurs sources de données susceptibles d'être chargées dans le Data warehouse, tels des fichiers de type Excel et la base de données paie et Grh. Après analyse et étude des nombreux fichiers de type paradox manipulés par les cadres de la direction Paie, nous avons constaté que ces fichiers contiennent des informations déjà existantes dans Gestion Paie et GRH (redondance d'informations). Ces

fichiers sont réédités manuellement avec des données provenant du système paie et GRH. Ceci démontre l'absence d'outils adéquats permettant de tirer profit de ces données.

### **II.5.3. La périodicité de l'alimentation :**

Il est indispensable de bien choisir la fréquence pour laquelle l'entrepôt de données sera alimenté. Le chargement périodique peut être assuré grâce à des Jobs. Un Job est un programme qui s'exécute à des intervalles de temps bien définis.

Dans notre cas, nous distinguons deux types de Jobs : Job pour le chargement des données alimentant les tables de dimensions et Job pour le chargement des tables de faits.

La fréquence d'alimentation des tables de dimensions sera mensuelle car elles sont dites des dimensions à évolution lente. La fréquence d'alimentation des tables de faits sera par mois. Cette fréquence permettra de réaliser des rapports d'activité périodique selon les besoins des dirigeants. Ce choix de fréquence peut être justifié car le délai d'un mois permet de collecter un maximum d'informations. Bien évidemment, ces fréquences peuvent être changées ultérieurement si le besoin se fait sentir.

Un autre mode de chargement peut être réalisé grâce à l'exécution d'un script SQL. Ce Script (chargement.sql) permet de forcer le chargement des données sans tenir compte de la fréquence d'alimentation et sans en altérer le fonctionnement.

### **II.5.4. Gestion et contrôle de l'alimentation :**

Afin de permettre le contrôle du processus d'alimentation de notre entrepôt de données : date de chargement, nombre de champs chargés, nombre de champs erronés... nous avons créé quatre vues qui permettront la gestion de l'alimentation. Ces vues sont :

*all\_iv\_map\_run* : permet d'avoir des informations sur les chargements effectués ultérieurement : date de chargement, package exécuté, état, nombre de champs chargés...

*all\_iv\_process* : permet de connaître les dates de création des Jobs.

*all\_iv\_run\_error* : permet de connaître les erreurs commises lors des chargements.

*all\_iv\_run\_target* : permet de connaître les tables alimentées lors des chargements.

## II.6. Construction du portail de restitution :

Lors de tous les points étudiés auparavant, nous avons présenté les aspects relatifs aux outils d'arrière-plan de notre système décisionnel, ainsi que la conception de la base de données. Dans cette partie, nous présenterons brièvement le portail de restitution qui nous permettra de présenter les données stockées dans l'entrepôt de données ou dans les cubes OLAP.

Le portail de restitution est la partie publique du système. Il représente ce que voient les utilisateurs. Le portail permet aussi de réaliser des requêtes sur l'entrepôt de données, ceci grâce à une interface efficace minimisant, pour l'utilisateur, le nombre d'actions afin d'obtenir le résultat voulu.

L'accès aux données du Data warehouse est réalisé selon deux types :

- ❖ **Reporting** : consiste en des rapports prédéfinis diffusés sur un canal. L'utilisateur y accède selon ses droits et privilèges. Ces rapports ont été édités grâce à l'outil Oracle Reports Builder.
- ❖ **Analyse OLAP** : Cet outil permet de naviguer dynamiquement dans les données afin de réaliser des rapports sous différents formats. Pour cette analyse nous utiliserons Microsoft Excel et Oracle Discover.

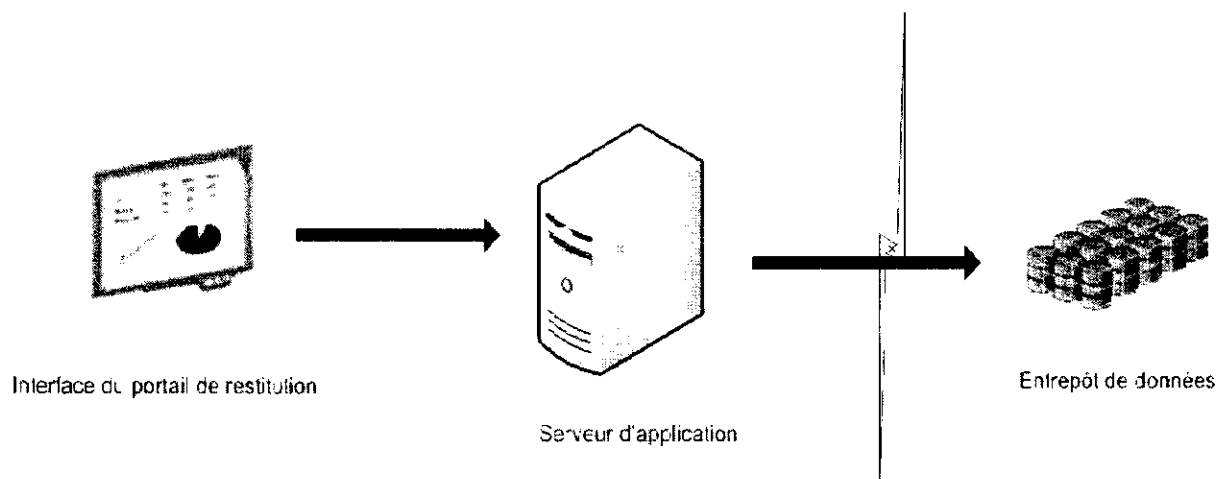


Fig 40: Description du portail de restitution des données.



### **III. Conclusion :**

Dans ce chapitre, nous avons présenté la modélisation dimensionnelle de notre entrepôt de données. Cette modélisation repose sur un schéma multidimensionnel qui intègre les concepts de faits et de dimensions. Le modèle de données de notre entrepôt est basé sur un schéma en flocon. Les dimensions peuvent être partagées par plusieurs tables de faits formant un schéma en constellation. Le partage des dimensions entre plusieurs faits limite les redondances et la complexité.

Nous avons aussi présenté l'architecture type d'un système décisionnel basé sur l'approche entrepôt de données, cette architecture est divisée en trois composantes essentielles (ETL, Données, Portail de restitution), chaque couche peut être implémentée à part.

Le déploiement du système consiste à spécifier l'architecture physique de notre Data warehouse à savoir les logiciels utilisés, la sécurité et les utilisateurs du système.

---

---

**PARTIE II :**

---

**CHAPITRE 4 :**  
**DEPLOIEMENT DU SYSTEME**

---

---

Après avoir décrit le modèle conceptuel de notre entrepôt de données, nous allons présenter notre solution informatique qui permettra de diffuser les informations décisionnelles aux cadres dirigeants et responsables. Lors de ce chapitre nous présenterons l'architecture générale de cette solution, ainsi que les pré requis matériels et logiciels.

### I. Disposition physique du système :

Pour présenter l'architecture globale de notre système décisionnel, nous l'illustrons grâce à la figure suivante.

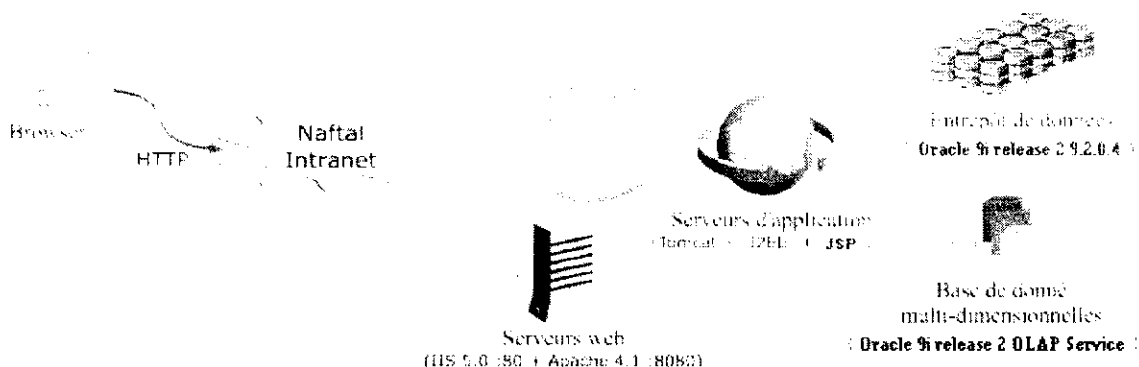


Fig 41 : Architecture globale du système

On constate que cette architecture est divisée en trois couches essentielles : présentation, traitements et stockage. En ce qui suit sera présenter ces couches en détail.

- **La présentation :** elle consiste en un navigateur web qui permet d'accéder aux états préparés et d'en modifier la mise en page. Cette application est développée avec la technologie J2EE (JSP 2.0). les rapports sont diffusés sur le site web grâce à Oracle Reports Builder (Oracle discoverer).
- **Le traitement :** cette couche contient dans notre cas le serveur de sécurité, le serveur web et le serveur d'application.

Le rôle de cette couche est de séparer les traitements sur les données afin de minimiser la charge sur les serveurs de stockage, et aussi de simplifier l'administration des systèmes informatiques.

- **Les données :** ou stockage. Cette couche englobe tous type des serveurs de données, soit un SGBD, un serveur de fichiers ou un serveur de messagerie, etc. Dans notre

cas, cette couche contient le serveur de données Oracle 9i version 9.0.2 pour l'entrepôt de données et le serveur de base de données multi-dimensionnelle Oracle 9i version 9.0.2 OLAP Service.

## **II. Les utilisateurs du système :**

Deux catégories d'utilisateurs sont répertoriées pour l'usage de notre système décisionnel. La première catégorie est l'ensemble des analystes et décideurs de la Direction des Ressources Humaines. c'est l'ensemble des utilisateurs le plus intéressé par le déploiement d'un système d'aide à la décision basé sur les dimensions, leurs objectifs est l'analyse dimensionnelle sur les activités de l'entreprise. La deuxième catégorie des utilisateurs est l'ensemble du personnel de la direction planification, leurs besoins informationnels à pour objectif de construire un modèle pré visionnaire stable et dynamique sur les activités de base de l'entreprise.

## **III. Les logiciels utilisés :**

### **III.1. Les composants serveurs :**

Nous citons les logiciels utilisés coté serveurs :

- Oracle 9i Entreprise Edition version 9.0.2 comme serveur de données implémentant l'entrepôt de données.
- Oracle 9i Entreprise Edition version 9.0.2 OLAP Service comme serveur de bases de données multidimensionnelles qui implémente les cubes OLAP.
- Tomcat 5 comme serveur web sur le port 8080 et comme serveur d'application de la technologie J2EE.

### **III.2. Les composants clients :**

En plus de notre site web qui permet de diffuser des rapports destinés à l'analyse des activités modélisées, on peut utilisé tout autre type d'application bureau orientée vers l'analyse de données comme client de notre système décisionnel, il suffit de la configurer sur le réseau intranet.

Le tableau ci-dessous donne un petit aperçu des logiciels pouvant être utilisés pour la manipulation des données contenues dans notre système décisionnel. Ces logiciels sont classés selon la catégorie des utilisateurs.

<i>Utilisateur novice outil simple</i>	<i>Utilisateur intermédiaire</i>	<i>Utilisateur averti ou statisticien</i>	<i>Expert</i>
Alice (Isoft) Statbox (GrimmerSoft) Smartminer (Grimmersoft) Analyst (Objectmine) Scenario (Cognos) Kate (Acknosoft) See-K (Trivium) Business miner (BO)	D-map (Complex Systems) Storm (Elseware) 4 Thought (Cognos)	StatLab (SLP infoware) SPAD (CISIA) SPSS (SPSS) Clementine (SPSS) Saxon (PMSI) K-wiz (Gentia) Darwin (Oracle)	SAS enterprise miner Intelligent miner (IBM)

Tableau 10: Tableau des logiciels par catégories d'utilisateurs

#### **IV. Sécurité du système :**

Lors de la réalisation de tout système, le facteur risque existe toujours. Ce facteur doit être pris en considération lors de chaque étape de conception et de déploiement du système.

Nous avons vu que l'information est un bien précieux, et que le processus de transformation des données en informations stratégiques n'est des plus faciles. Alors, voir la perte ou la destruction de ces informations constitue une perte énorme d'une part pour les concepteurs du système et d'autre part pour les utilisateurs finaux. D'où la nécessité de sécuriser et de protéger notre système et d'adopter une politique de sécurité.

Une politique de sécurité est un ensemble de règles qui fixent les actions autorisées et interdites dans le domaine de la sécurité. On entend par là que la politique de sécurité est un ensemble de normes portant sur les moyens (humains, techniques et financiers) à mettre en œuvre afin de se protéger contre toute éventualité de menace. [Laouis, 05]

Nous distinguons deux aspects de sécurité : sécurité physique et sécurité logique. La première contre les dégâts matériels. La deuxième consiste à protéger les données logiquement contre les dégâts par des logiciels tiers (Virus...) ou contre des intrusions de personnes malintentionnées.

### IV.1. Sécurité physique :

Le risque principal qui pourrait atteindre le matériel est la panne qui pourrait surgir sur les serveurs (d'applications et de données). La première solution à ces pannes peut être de restreindre l'accès à la salle contenant les serveurs uniquement aux informaticiens chargés de la configuration. La deuxième solution serait de réaliser des sauvegardes sur un serveur de secours. Pour cela on peut utiliser l'utilitaire d'Import/Export d'Oracle. Ces sauvegardes permettraient de protéger les données contre toute perte accidentelle en cas de défaillance du matériel. Ces sauvegardes devraient être réalisées périodiquement. Grâce à cet utilitaire, on peut facilement restaurer les données effacées ou modifiées malencontreusement, à partir de la dernière copie réalisée.

### IV.2. Sécurité logique :

La sécurité logique concerne l'accès aux données. Nous distinguons trois niveaux de protection des données. Ces trois niveaux seront illustrés par la figure suivante :

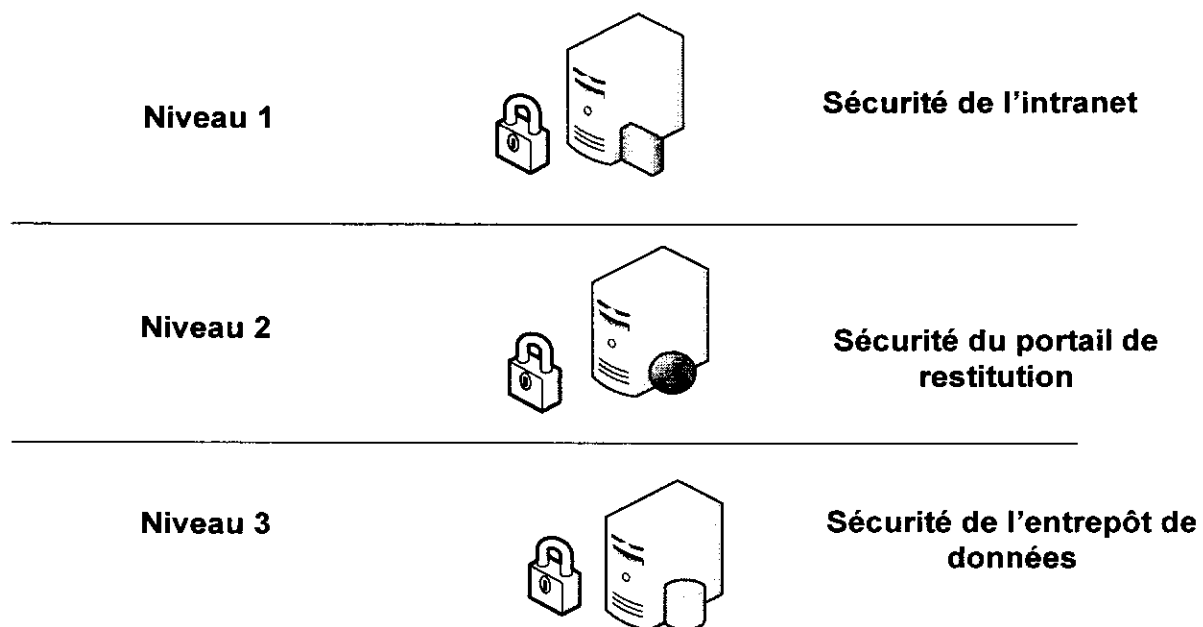


Fig 42 : Les trois niveaux de sécurité logiques

**❖ 1<sup>er</sup> niveau :**

La solution adoptée pour ce niveau est d'installer un pare-feux permettant d'empêcher les intrusions ou toute attaque provenant de l'extérieur de l'entreprise. En interne, il serait judicieux d'installer des antivirus pour pallier aux attaques de virus ou de tout autre programme qui nuirait à l'intégrité des données.

**❖ 2<sup>ème</sup> niveau :**

Ce niveau correspond à la sécurité de notre portail de restitution de données implémenté sur le serveur Web.

- Notre portail de restitution ne permet que des accès en consultation des cubes. Donc aucune modification sur les cubes ne peut être effectuée à partir de ce portail.
- Les utilisateurs du portail accèdent aux données d'après les privilèges qui leur sont accordés.

**❖ 3<sup>ème</sup> niveau :**

C'est la sécurité liée au serveur de bases de données Oracle :

- Authentification Windows et Oracle pour les accès des utilisateurs.

**V. CONCLUSION:**

Cette partie de notre travail, qui en constitue la dernière, traite du déploiement de notre système décisionnel. Nous y avons abordé l'aspect physique de ce dernier, en citant son dispositif physique, ses utilisateurs, les logiciels utilisés ainsi que le volet sécurité.

Cette étape constitue la plaque tournante de notre travail, car elle servira à décrire l'environnement technique de notre système décisionnel. Le bon fonctionnement de notre système est conditionné par sa mise en œuvre efficace.

---

---

CONCLUSION GENERALE

---

---



## Conclusion générale :

L'environnement dans lequel évoluent les entreprises est un monde en perpétuel changement marqué par d'incessants bouleversements économiques. Afin de pouvoir progresser dans cet environnement l'entreprise doit posséder les informations nécessaires pour prendre les bonnes décisions aux bons moments. Ceci peut être facilité par la mise en œuvre de systèmes décisionnels et d'outils d'aide à la décision. Le but espéré par ces systèmes est d'améliorer le processus décisionnel et de diminuer les erreurs dues au manque d'informations stratégiques impliquant de mauvaises décisions.

Dans ce contexte, Naftal Spa avec sa Direction Ressource Humaine nous a confié la tâche de concevoir et de mettre en œuvre un système décisionnel basé sur l'architecture entrepôt de données.

Afin de mener notre travail de façon efficace, nous avons utilisé une méthodologie suivant l'approche de cycle de vie dimensionnel<sup>3</sup> de Ralph Kimball, pionnier dans le domaine des entrepôts de données. Cette approche est constituée de trois phases : définition et étude des besoins, conception du système et puis son déploiement.

Dans une première phase, nous avons identifiés les différents besoins à partir d'une enquête menée auprès des utilisateurs et décideurs de l'entreprise. Nous avons utilisé trois approches complémentaires afin de cerner tous besoins en matières de pilotage et d'analyse de l'activité Ressource Humaine.

La deuxième phase a été consacrée à la conception de notre système décisionnel. Elle propose une modélisation multidimensionnelle. Dans cette phase nous avons proposé des schémas dimensionnels en flocon des aspects auxquelles les besoins ont été identifiés. Lors de cette modélisation nous avons fait correspondre les besoins de l'entreprise avec la réalité des informations disponibles. Notre modèle offre comme avantage majeur d'être simple et facilement compréhensible par toute personne, même non informaticienne. L'architecture adoptée pour notre système, révèle les différentes parties qui composent un système décisionnel fondé sur l'approche entrepôt de données. Nous avons construit le modèle physique de notre entrepôt. Ce dernier est basé sur le modèle relationnel. Les cubes OLAP construits sauvegardent l'agrégation des données stockées dans l'entrepôt. En matière d'alimentation, nous avons développé un script SQL exécutant des packages permettant ainsi

---

<sup>3</sup> Voir annexe

d'alimenter l'entrepôt à partir du système source à des fréquences prédéfinies. Le prototype d'un portail de restitution des données est proposé. Ce dernier est capable de faire l'analyse croisée et le reporting. Il est réalisé à l'aide d'une application Web qui s'exécute sur l'intranet de l'entreprise.

La dernière phase concernait le déploiement du système. La base de données de l'entrepôt est créée et reliée avec les systèmes opérationnels (logiciel Gestion Ressource Humaine et Gestion Paie) par l'intermédiaire de notre script SQL. Nous avons aussi créé la base de données multidimensionnelle et nous avons également installé l'application de l'interface utilisateur. Notre système décisionnel est implémenté sur : Oracle 9i release 2 comme serveur de données pour l'entrepôt et la base multidimensionnelle, intranet comme infrastructure réseau et l'architecture trois-tiers comme technologie de développement web.

Comme tout projet, notre travail se devait de respecter les délais de réalisation. Il nous a permis d'acquérir une connaissance approfondie des systèmes décisionnels en général, et les entrepôts de données en particulier. Ce travail représente le fruit de quatre années d'études et d'une année d'expérience dans le monde professionnel. Nous avons pu affronter les difficultés qu'on rencontre lors de la réalisation de tout nouveau projet. D'autre part, ce stage nous a permis d'aborder des sujets d'actualité tel que l'implémentation et l'administration des Data Warehouse.

Un travail ne pouvant être terminé sans proposer des perspectives, les notre sont les suivantes :

- Permettre l'extension du portail vers Internet de manière à permettre aux décideurs de consulter les données quelque soit leur localisation.
- Etendre le système décisionnel vers d'autres Structure et Directions Ceci peut être réalisé car l'architecture de notre entrepôt de données a été réalisée de manière à faciliter son extension.

---

---

# LEXIOUE

---

---

## LEXIQUE

EIS	Exécutive Information System
ETL	Extract, Transform and Load
HOLAP	Hybrid On-Line Analytical Processing
MOLAP	Multidimensional On-Line Analytical Processing
ROLAP	Relational On-Line Analytical Processing.
Cube	Un nom pour une base de données, utilisé habituellement lorsqu'on se réfère uniquement au cas simple du temps ou du l'effectif
Data Warehouse	Entrepôt de données.
OLAP	On-Line Analytical Processing
OLTP	On-Line Transactional Processing
Attribut	Un fait décrivant chaque position d'une dimension
SGBD	Système de Gestion de Bases de Données
SI	Systèmes d'Information
SID	Systèmes d'Information Décisionnel
Dimension	Une unité indépendante du modèle d'une organisation, servant de point d'entrée ou de mécanisme pour faire des coupes dans les mesures additives des activités de l'organisation
Drill down	Forage vers le bas
Fait	Donnée numérique servant de base à la définition des indicateurs dans un modèle multidimensionnelle.
Indicateur	Chiffre significatif de la situation économique de l'entreprise
Modèle dimensionnel	Technique de modélisation, consistant à créer physiquement ce modèle sur une base de données relationnelle à partir de l'identification des faits à analyser et des dimensions d'analyses qui leur sont associées
Modèle en étoile	Technique de modélisation dimensionnelle, consistant à créer physiquement ce modèle sur une base de données relationnelle. Ce modèle distingue physiquement les tables de faits des tables de dimensions
Modèle en flocon	Technique de modélisation dimensionnelle, dérivant de la modélisation en étoile. Dans ce modèle, les tables de dimensions y sont dénormalisées

---

Modèle relationnel	Technique de modélisation, consistant à modéliser une base de données en la décomposant en entité et en relations corrélant ces entités
Table de faits	La table centrale du schéma de jointure en étoile
Modélisation dimensionnelle	Multidimensional On-Line Analytical Processing
Data mining	Fouille de données
Méta données	Données sur les données

---

---

**ANNEXE A**

---

---

Dans ce qui suit nous donnons une présentation du système source Gestion Ressource Humaine et du système source Gestion Paie.

## 1- Description de l'application Gestion Ressource Humaine

Cette application est utilisée au niveau de la Direction Ressource Humaine ainsi que dans tous les CDF (Centre De Frais). On trouve la base de données consolidée de toutes les bases de données de l'application Gestion Ressource Humaine au niveau du groupe informatique de Naftal Spa.

	Domaine fonctionnel	Système d'exploitation	Accès aux données	Langage de développement
Données générales	Gestion Ressource Humaine			
Environnement Technique		Windows XP Professional	Paradoxe Oracle9i	Delphi 5 Entreprise

Le SGBD de la base de données consolidée est en Oracle9i.

### 1.1- Liste des tables :

Il existe plusieurs tables dans la base de données Gestion Ressource Humaine (70 tables), mais nous ne citerons que celle concernées par notre étude.

Nom de Table	Désignation
<i>AGENT</i>	Agent
EXCO NGE	Les congés
FTRUB	Les rubriques
ITEM	Les mouvements
<i>LISTMV</i>	Signification des mouvements
MISSION	Les missions
<i>MOUV</i>	Mouvement de la personne
TPOSITION	Historique de Position
CSP	Catégorie Socioprofessionnelle
NC	Niveau Classification
NIVEAU_ACADEMIQUE	Niveau Académique
POSITION	Position
SEXE	Sexe
SITUATION_FAMILIALE	Situation Familiale

## 1.2- Dictionnaire des données

Nom	Désignation
<i>Agent</i>	<i>Agent</i>
MATRIC	Matricule
NOM	Nom de l'agent
PRENOM	Prénom de l'agent
CODA1	Code agence
CODE_P	Code personne
SEXE	Sexe
POSI	Position
DATE_N	Date de naissance
DATE_REC	Date de recrutement
SITU	Situation
ADRESSE	Adresse
PRENOM_PER	Prénom de père
NO_PRENOM_MER	Nom et prénom de mère
N_SS	Numéro de sécurité sociale
SIT_FAM	Situation Familiale
NB_ENFANT	Nombre d'enfants
LIEU_NAISS	Lieu de naissance



CODSTR	Code structure
ECH	Echelon
DATE_POSIT	Date position
SERVICE_NAT	Service national
DATE_CONF	Date conférence
CODE_DIP	Code diplôme
CODE_AN	Code année
CODE_BANQUE	Code banque
CODE_AGENCE	Code agence
NUM_BANQ	Numéro banque
NUM_ADH	Numéro adhérons
LIEU	Lieu
SITU_ACT	Situation actuelle
PRESUME	Présumé
MATADM	Matricule ADM
GROUPE	Groupe
NIV_AC	Niveau académique
NOM_JFILLE	Nom de jeune fille
DIR	Direction
DIV	Division
BR	Branche
EXERCICE	Exercice
QTE	Quantité
POSITION	Position
RELIQUAT	Reliquat
ETAT_ORG	Etat organisation
AGE	Age
EXP	Expérience
ADD_WILAYA	Adresse wilaya
LIEU_NAIS_WILAYA	Lieu de la wilaya de naissance
XMATRIC	Matricule ancien
FLG	Flag
CODADM	Code ADM
<b><u>Exconge</u></b>	<b><u>Conge</u></b>
MATRIC	Matricule
EXERCICE	Exercice
QTE	Quantité
POSITION	Position
RELIQUAT	Reliquat
DROIT	Droit de conge
RE_QTE	Quantité ancienne
RE_RELIQUAT	Reliquat ancien
RE_DROIT	Droit ancien
<b><u>ITEM</u></b>	
CODE_IT	Lettre Majuscule
TYPE_IT	Numéro du Bon de Régularisation Ventes
DESIG_IT	Numéro d'opération
<b><u>LISTMV</u></b>	<b><u>Liste mouvement</u></b>

CODMV TYPE_MV DESIG_MV	Code mouvement Type mouvement Désignation mouvement
<b>MISSION</b>	<b>Mission</b>
NUMDOC CODEP MATRIC DEST RAIS MTRAN DATAL DATR FAITA LE	Numéro document  Matricule Destination Raison Matricule Date aller Date retour Faite aller Date mission
RESPONSABLE VISA	Responsable Visa
<b>MOUVI</b>	<b>Mouvement de personne</b>
NUM_DOC MATRIC CODE_P CODEA2 TYPE_MVI DATE_MVI MOTIF DATE_ETAB ETAT	Numéro document Matricule Code personne Code aller Type mouvement Date mouvement Motif Date établissement
XCODE_P XCODSTR VALIDE NDRHM DATE_RETOUR SITU_ANC ANC_ETAT XDATE_POSI XMATRIC DIRECTION XDIRECTION POSI XNUM_DOC	Date de retour Situation ancienneté
<b>TPOSITION</b>	<b>Position</b>
Type Code Libelle	Type position Code position Libelle position
<b>Categorie Socioprofessionnelle</b>	<b>Catégorie socioprofessionnelle</b>
CODE_CLASSE_CSP DESIG_CSP TYPE_CSP	Code classe catégorie socioprofessionnelle Désignation catégorie socioprofessionnelle Type catégorie socioprofessionnelle
<b>NC</b>	<b>Niveau Classification</b>
CODE_NC	Code Niveau classification

SALAIRE	Salaire
<b>NIVEAU ACADEMIQUE</b>	<i>Niveau Académique</i>
CODE_NA	Code Niveau Académique
TYPE_NA	Type Niveau Académique
DESIG_NA	Désignation Niveau Académique
<b>POSITION</b>	
CODE_POSITION	Code Position
TYPE_POSITION	Type position
LIB_POSITION	Libellé position
<b>SEXE</b>	<i>Type sexe</i>
SEXE	Code sexe
TYPE_SEXE	Type sexe
<b>SITUATION FAMILIALE</b>	<i>Situation familiale</i>
CODE_SF	Code situation familiale
TYPE_SF	Type situation familiale
DESIG_SF	Désignation situation familiale

## 2- Description de l'application Gestion Paie

Cette application est utilisée au niveau de la Direction Paie . On trouve la base de données consolidée de toutes les bases de données de l'application Gestion Paie au niveau du groupe informatique de Naftal Spa.

	Domaine fonctionnel	Système d'exploitation	Accès aux données	Langage de développement
Données générales	Gestion Paie			
Environnement Technique		Windows XP Professional	Paradoxe Oracle9i	Delphi 5 Entreprise

Le SGBD de la base de données consolidée est en Paradox. Or, les ingénieurs du groupe informatique sont en train de migrer les données sous Oracle 9i. Nous avons participé à ce

projet de migration des données vers Oracle, en transférant les données des tables dont nous avons besoin.

## 2.1- Liste des tables :

Il existe plusieurs tables dans la base de données Gestion Paie , mais nous ne citerons que celle concernées par notre étude.

Nom de Table	Désignation
<i>PAIE</i>	Paie
PAIERUB	Paie rubrique
<i>STRUCTURE</i>	Structure
FTRUB	
PERS	Personne

## 2.3- Dictionnaire des données

Nom	Désignation
<i>PAIE</i>	<i>Paie</i>
PAI_MATR	Matricule
PAI_CDG	
PAI_CUNIT	Code unitaire
PAI_CDF	Centre de frais
PAI_SERV	Service
PAI_CANAL	Code analytique
PAI_GIMP	Gain imposable
PAI_RGIMP	Retenu gain imposable
PAI_GNIMP	Gain non imposable
PAI_RGNIMP	Retenu gain non imposable
PAI_RETIMP	Retenu imposable
PAI_RETNIP	Retenu non imposable
PAI_RRET	Tracteur
PAI_CHPAT	Matricule Camion
PAI_RCHPAT	Matricule Tracteur
PAI_NETIMP	Distance
PAI_NETPAY	Numéro de la convention

PAI_DATEFF	Type du Bon de livraison facture
<b>PAIERUB</b>	<b>Paie rubrique</b>
PL_MATR	Lettre Majuscule
PL_CDG	Numéro de l'avoir05
PL_CUNIT	Numéro de l'opération
PL_CDF	Taux de taxe
PL_SERV	Code produit
PL_DATEFF	Prix
PL_RUB	CDS
PL_CNAT	Libellé produit
PL_SENS	Quantité
PL_MONT	Montant
PL_TAUX	Client
PL_EL ST	Mode de règlement
PL_DATD	Lettre Majuscule
PL_DATF	Numéro du Bon de Livraison
FLG	Type Bon de Livraison
<b>STRUCTURE</b>	<b>Structure</b>
CODE	Lettre Majuscule
NFUS	Numéro du Bon de Livraison
LIB	Type Bon de Livraison
LIBDIVISION	Numéro d'opération
NOMBRE	Taux de taxe
<b>Ftrub</b>	
TR_CRUB	Lettre Majuscule
TR_LRUB	Numéro du Bon de Régularisation Ventes
TR_CPTGEND	Date de l'opération
TR_CPTGENC	CDS
FLG	Numéro Bon de Livraison facture
<b>Pers</b>	<b>Personne</b>
P_MAT	Lettre Majuscule
P_DIVDIR	Numéro du Bon de Régularisation Ventes
P_CDF	Numéro d'opération
P_SERV	Taux de taxe
P_LIEUT	Code produit
P_CODPOS	Emballage
P_NOMP	Prix
P_DATN	Date operation
P_SEXE	CDS
P_DATR	Libellé produit
P_DATC	Quantité
P_SF	Montant
P_SC	Client
P_QF	Mode de règlement
P_ENFPP10	Lettre Majuscule
P_ENFPM10	Numéro du Bon de réception
P_ENFCP10	Numéro d'opération
P_ENFCM10	CDS
TYPEACHAT	Type d'achat

P_NC P_ECH P_TIEP P_ADRES P_CPTANAL P_CATEG P_CORP P_GROUP P_CDIP P_CSAL P_SALD P_CLOG P_CTRSP P_CPAIE P_CFONCT P_LIBF P_CBANQ P_CAG P_CPTBNQ P_AFF P_MIP P_CAIS	CDS approvisionnement Code Raffinerie Code Fournisseur Nom du Fournisseur Référence du document Date référence Moyen de transport Matricule camion citerne Tracteur Matricule tracteur Distance Navire Date de chargement Date de réception Coût d'approvisionnement Matricule du chauffeur Nom du chauffeur Numéro de la convention Numéro du jour comptable Montant HT Montant TTC Etat
P_CAF P_CMAF P_CMDJ P_NBAMDJ P_NADH P_NSS P_CVOIT P_MTVEH P_MENSV P_DATPV	Lettre Majuscule Numéro du Bon de réception Code Produit Emballage Date d'opération CDS Quantité Prix Montant Coût d'approvisionnement
P_CPRET P_MTSOC P_MENSS P_DATPS P_RUB1 P_SENS1 P_MONT1 P_RUB2 P_SENS2 P_MONT2 P_RUB3 P_SENS3 P_MONT3 P_RUB4 P_SENS4 P_MONT4 P_RUB5	Lettre Majuscule Numéro du Bon de Ravitaillement Sortie Date d'opération CDS CDS destinataire Référence du document Moyen de transport Coût de transport Date de chargement Camion citerne Matricule camion citerne Tracteur Matricule tracteur Matricule du chauffeur Nom du chauffeur Numéro de convention Distance

P_SENS5 P_MONT5 P_RUB6 P_SENS6	Numéro du jour comptable Montant HT Montant TTC Etat
P_MONT6 P_RUB7 P_SENS7 P_MONT7 P_RUB8 P_SENS8 P_MONT8 P_RUB9 P_SENS9	Lettre Majuscule Numéro du Bon de Ravitaillement Sortie Code produit Emballage Date d'opération CDS Quantité Prix Montant
P_MONT9 P_RUB10 P_SENS10 P_MONT10 P_RUB11 P_SENS11 P_MONT11 P_RUB12 P_SENS12 P_MONT12 TYP	Lettre Majuscule Numéro du Bon de Ravitaillement Entrée Date d'opération CDS CDS Emetteur Moyen de transport Coût de transport Date de chargement Date de déchargement Camion citerne Matricule camion citerne

---

---

**ANNEXE B**

---

---



## Manuel d'installation du système

Afin d'installer notre système décisionnel, il est indispensable de suivre les étapes suivantes :

- Installation d'Oracle Data Base sur le serveur de base de données.
- Création de la base de données globale « Distination » qui contiendra le schéma de l'entrepôt de données.
- Importation du fichier « Distination.dmp » qui contient les tables, les vues, les packages, les dimensions et les cubes de l'entrepôt de données.
- Exécution du fichier « CréerJob » pour créer les Jobs qui permettront l'alimentation automatique et fréquente de l'entrepôt de données.
- Exécution du fichier « Chargement.sql » pour alimenter l'entrepôt de données pour la première fois.
- Installation du serveur d'application « Tomcat 5 » sur le poste du serveur d'application.
- Installation de l'application « NaftalWeb » dans le dossier « Webapps » du répertoire « Tomcat ».
- Installation du package Java « Jdk 1.5 ».
- Déclaration des variables d'environnement (JAVA\_HOME) pour le JDK et (PATH) pour le Tomcat.

---

---

ANNEXE C

---

---

## Méthodologie de Ralph Kimball :

Le schéma de cycle de vie dimensionnel proposé par R. Kimball [Kim, 00] matérialise l'enchaînement des grandes étapes de l'implémentation d'un entrepôt de données.

Il identifie l'enchaînement des tâches de haut niveau et souligne les contraintes entre ces différentes actions tant au niveau technologique qu'au niveau des données et des applications.

Cette démarche est illustrée par le schéma ci-dessous. Ce schéma représente la succession des tâches de haut niveau nécessaires à la conception, au déploiement d'entrepôt de données efficaces.

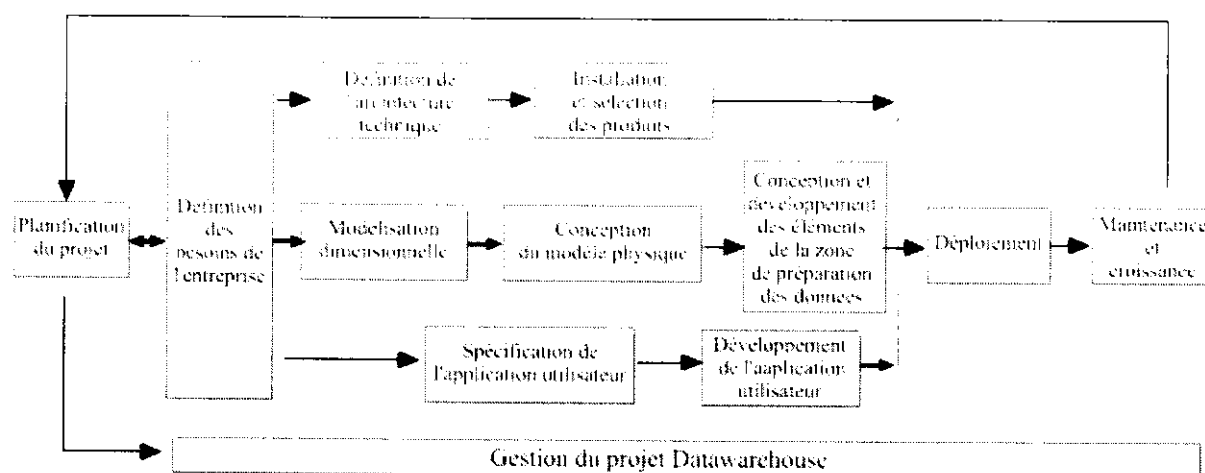


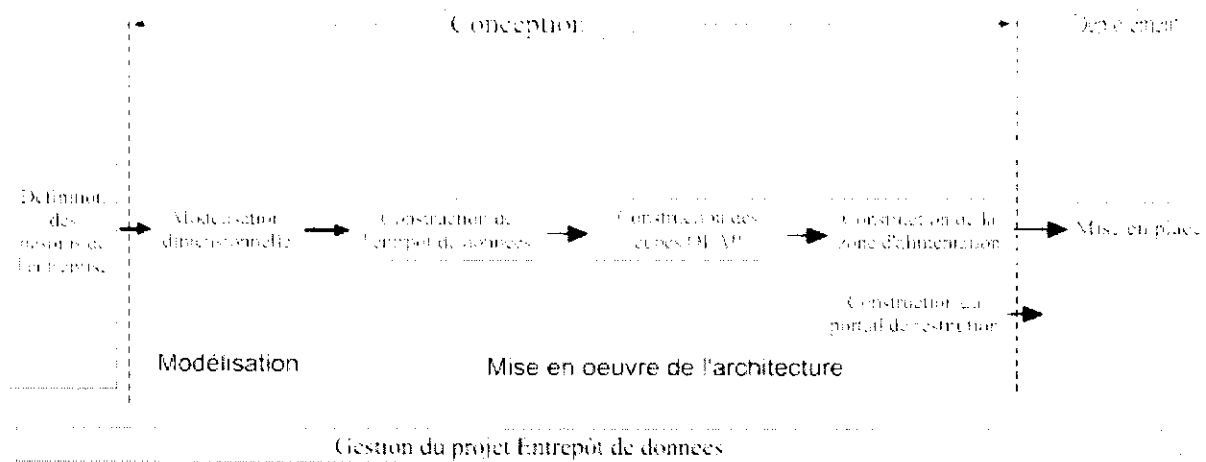
Schéma du cycle de vie dimensionnel [Kim, 00]

## Personnalisation de la méthodologie :

Afin d'adapter la méthodologie proposée par Ralph Kimball, nous avons projeté notre démarche à la sienne. Cela a été réalisé par la décomposition de certaines phases en plusieurs et en parallèle l'élimination des autres. Pour les deux étapes concernant la spécification et le développement des applications utilisateur, nous avons construit une seule étape qui regroupe ces deux dernières.

L'étape maintenance et croissance se déroulera après le déploiement de l'entrepôt. Dans notre travail, cette étape est inutile car elle traite le point de continuation d'assistance et de formation des utilisateurs.

Le schéma suivant illustre notre approche :



### Personnalisation de la méthodologie de Ralph Kimball

---

---

**ANNEXE D**

---

---

## **Conception et Réalisation d'un Entrepôt de Données :**

### **POUR NAFTAL SPA**

#### **Objectif : Questionnaire**

Nous avons l'honneur de vous solliciter afin de répondre à ce questionnaire. En effet, nous sommes en vigueur de préparer notre projet de fin d'étude d'ingénieur en informatique. Pour cela, il nous a été proposé l'implémentation d'un système décisionnel relatif au suivi du personnel et sa rémunération, cette activité 'Ressource Humaine' constitue un chapitre très important dans toute entreprise.

Le système décisionnel à implémenter permet au décideur de lui donner une vue multidimensionnelle de la Direction Ressource Humaine, afin qu'il puisse prendre les bonnes décisions ce qui veut dire que le décideur peut visualiser à l'aide de graphes ou de rapports (automatiquement générés par le logiciel) les indicateurs de performances selon les axes ou dimensions qu'il souhaite voir. Nous vous prions de bien vouloir prendre la peine d'y répondre. Merci.

## Questionnaire

Date :

Direction :

Division :

❖ Questions adressées aux dirigeants :

N°	Question
1.	Quels sont les objectifs de votre secteur d'activités ?
2.	Quels sont les processus décisionnels sur lesquels vous avez une influence dans ces activités ?
3.	Comment se fait votre intervention dans ces processus ?
4.	Quels sont les indicateurs clés qui vous aident à détecter la réussite ou non dans votre activité ?
5.	Quels sont les problèmes rencontrés lors de la prise de décisions ?

❖ **Questions adressées aux responsables et dirigeants :**

<b>Analyse des activités</b>	
<b>N°</b>	<b>Question</b>
1.	Quelles sont les activités qui vous concernent ?
2.	Quels sont les indicateurs ou métriques qui vous permettent d'évaluer votre activité ?
3.	Quelle est la fréquence d'évaluation de votre activité ?
4.	Quels sont les axes d'analyse importants (type agent, structure...) dans votre analyse? Quels sont les critères ?
5.	Quels sont les obstacles ou problèmes rencontrés lors de la prise de décisions ?

<b>Analyse des besoins</b>	
<b>N°</b>	<b>Question</b>
1.	Existent-ils des analyses dans votre activité ? de quels types ?
2.	Quels sont les analyses que vous souhaitez avoir et qui n'existent pas actuellement ?
3.	Quelles sont les informations que vous jugez pertinentes ? quelles sont leurs sources ? quel usage faites vous de ces informations
4.	De quelles données historisées avez-vous besoin ?
5.	Quels sont les processus clés qui vous concernent ?



---

---

# **BIBLIOGRAPHIE**

---

---

---

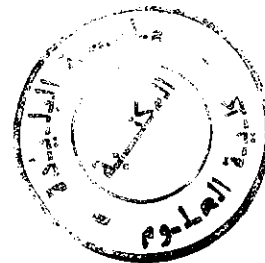
---

## Bibliographie et Webographie

- [Abbey, 02]: Michael Abbey, Mike Corey, Ian Abramson: « Oracle 9i, Notions fondamentales », Campus Press ,2002,USA.
- [A.Ramdane ,04]: A.Ramdane-Cherif, L.Sid. Mémoire de fin d'étude « conception et réalisation d'un outil Web d'analyse et d'un Data warehouse ». INI, 2003/2004
- [Chaudhuri .97 A] : S.Chaudhuri: « Data warehouse and OLAP for decision support », Microsoft research. Redmond: SIGMOD Record. 1997. USA.
- [Devisy, 02] : A.Devisy : « Livre blanc e-Business Intelligence. l'effet Internet sur le décisionnel » www.aubay.com
- [Inmon, 02]: W.Bill Inmon: « Building the Data warehouse ». Edition Wiley Computer Publishing, USA 2002.
- [Kerkri, 01] : E.Kerkri : « Processus de mise en œuvre d'entrepôt de données 'Approche sémantique' », Rapport de recherche entre l'hôpital de Bocage et l'université Bourgogne, Dijon, France 2001.
- [Kim, 97] : Ralph Kimball, « *Entreôpt de données* », International Thomson Publishing Fance, 1997.
- [Kim, 00]: Ralph Kimball « Concevoir et déployer un Datawarehouse, Guide de conduite de projet » édition Eyrolles, France, 2000.
- [Nakache, 98] : Didier Nakache, « *Data warehouse et data mining* », Conservatoire National des Arts et Métiers de Lille, Version 1.1, 15 Juin 1998
- [Web1] : « *Système d'Information Décisionnel et Data warehouse* », [www.decisionnel.net/datawarehouse/dwh.htm](http://www.decisionnel.net/datawarehouse/dwh.htm)

### Quelques Sites intéressants :

- [www.decisionnel.net/datawarehouse/dwh.htm](http://www.decisionnel.net/datawarehouse/dwh.htm)
- [www.dwinfocenter.org](http://www.dwinfocenter.org)
- [www.datawarehousingonline.com](http://www.datawarehousingonline.com)
- [www.rkimball.com](http://www.rkimball.com)
- [www.billimmon.com](http://www.billimmon.com)



- 
- [www.nordnet.fr/dnakache/valeur](http://www.nordnet.fr/dnakache/valeur)
  - [www.datawarehousing.com](http://www.datawarehousing.com)
  - [www.aubey.com](http://www.aubey.com)
-