

110-004-218-1

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'Etat en informatique**

Option : Système d'Information

Sujet :

***Conception et réalisation d'un data warehouse  
métier pour la fonction Ressource Humaine à  
ENGTP***

**Présenté par :**  
FOURALI Mouloud  
KOULLA Walid

Promoteur : Mlle SOUAMI Feryel

**Organisme d'accueil :** ENGTP

**Soutenu le** 30/05/2008

, devant le jury composé de :

**Président :**  
**Examineur :**  
**Examineur :**

MIG-004-218-1



Je dédie ce travail

A mes parents qui m'ont toujours poussés et motivés dans mes études, à mes frères et sœur ce mémoire représente donc l'aboutissement du soutien et des encouragements qu'ils m'ont prodigués tout au long de ma scolarité. Qu'ils en soient remerciés.

A tous les membres des familles Fourali et Ouchenir spécialement à mes deux grandes mères pour leurs prières.

A tous les ami(e)s de la fac avec lesquels j'ai eu le plaisir de partager mon cursus universitaire.

A Mr.Belgacem et Mlle.Mohamedi pour la confiance qu'ils nous ont accordée en nous confiant un tel projet et à tout le staff de l'ENGTP pour leur totale coopération avec nous ainsi que leur accueil au sein de l'équipe de développement.

Je tiens spécialement à mentionner le plaisir que j'ai eu à travailler avec Mlle.Hadj Kouider dont les conseils, orientations et suggestions ont été plus que déterminants dans l'aboutissement de ce travail en la remerciant pour toute sa patience à notre égard.

A Mlle.Souami pour son encadrement technique.

A ma chère et tendre amie Metat Sihem sans qui ces années d'études n'aurais pas connu un dénouement aussi heureux et joyeux, merci Sissi.

A mes défunts grands pères Fourali Mouloud et spécialement Ouchenir Mohand qui m'a gâté depuis ma tendre jeunesse.

Je pense également à tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail.

Bien sur à mon binôme Koulla Walid pour son aide et toute la bonne humeur qu'il diffusé autour de lui ainsi qu'à toute sa famille.

Enfin et surtout au DIEU tout puissant qui m'a toujours soutenu.

Fourali Mouloud



Merci à dieu qui a voulu que ceci ai lieu

Je dédie ce travail...

A celle qui a attendu avec patience les fruits de sa bonne éducation, à la personne qui a joué le rôle principal dans mon existence ... à ma mère.

A celui qui m'a indiqué la bonne voie en me rappelant sans cesse que la volonté, le courage et la ténacité font toujours les grands hommes ... à mon père.

A mes frères et sœur qui au file de mes péripéties ont su être le soutien dont j'avais tant besoin.

A ma source de joie quotidienne, à celui qui a su rire même dans les moments les plus difficiles de notre périple...à Mouloud (Neo).

A toute ma famille et particulièrement à mes deux grands-mères, que Dieu les garde.

Je dédie en fin ce travail à tous ceux qui ont contribué de près ou de loin, physiquement ou moralement à son aboutissement et en signe de ma gratitude, je leur offre pour toujours ma reconnaissance infinie et mon profond respect pour la patience et les sacrifices qu'ils m'ont consenti : (à Fayçal, Larbi, Yacine, Reda, Mohamed, Khaled, Riadh, ...etc.)

Je remercie Mlle. Souami pour son encadrement technique qui a porté une aide décisive à l'aboutissement de ce travail.

Je remercie particulièrement Mlle. Hadj Kouider pour la patience dont elle a fait preuve à notre égare et pour sa précieuse aide qu'elle nous a généreusement offert.

Je tiens à remercier Mr. Belgacem et Mlle. Mohamedi pour la confiance qu'ils nous ont accordée en nous confiant un projet de cette envergure et à tout le staff de l'ENGTP pour leur totale coopération, ainsi que leur accueil au sein de l'équipe de développement.

KOULLA Walid

## Résumé

Afin de prendre les bonnes décisions, les entreprises ont besoin d'accéder de manière efficace à des informations précises. À l'ère du tout-technologique, la source de données consultables paraît infinie. Comment naviguer sur cet océan de données importantes, souvent contradictoires et trouver la bonne information qui aidera les analystes dans leurs prises de décisions ? Le Data warehouse ou le Data mart est le cœur de pratiquement toutes les solutions de Business Intelligence. Ils permettent de faire toutes les opérations analytiques et donner aux décideurs des moyens chiffrés pour évaluer les faits voulus.

Notre travail consiste à mettre en œuvre un data warehouse métier pour la fonction gestion de la ressource humaine au sein de l'ENGTP. L'approche retenue pour la réalisation du système est l'approche de cycle de vie dimensionnel proposée par Ralph Kimball. Notre approche est divisée en trois phases : Identification des besoins, Conception et Déploiement du système.

## Abstract

In order to make the good decisions, the companies need to reach accurate informations efficiently. At the era of all-technological, the consultable data source appears infinite. How to sail on this ocean of important, often contradictory data and to find the good information which will help the analysts in their decision makings? Data warehouse or Data mart are practically the heart of all of Business Intelligence solutions. They make it possible to make all the analytical operations and to give to the decision makers means quantified to evaluate the desired facts.

Our work consists in implementing one Data warehouse for the function of human resource management within the ENGTP. The approach adopted for the realization of the system is the dimensional life cycle suggested by Ralph Kimball. Our approach is divided into three phases: Need identification, Design and Deployment of the system.

## ملخص

من أجل أخذ قرارات جيدة ، الشركات لها حاجة لتوصل الى معلومات دقيقة و بكفاءة عالية . في عصر أين لتكنولوجيه لم تترك للبيانات مصدر له نهاية. كيف نبحر في هذا الكم الهائل من المعلومات ، أحيانا متناقضة ،و العثور على المعلومة الدقيقة التي من شأنها أن تساعد المحللين في صنع القرار؟ تخزين البيانات او بيانات السوق من الناحية العملية هو الحل الأمثل لكل هذه التساؤلات. على انها تجعل من الممكن اجراء جميع العمليات التحليليه وتقديمها الى صانعي القرار كوسيلة لتقييم كمية الحقائق المرجوة.

ويتألف عملنا هذا في وضع نظام قراري عملي لفائدة ادارة الموارد البشريه في المؤسسة الوطنية للأشغال البترولية الكبرى النهج المعتمد لتحقيق هذا النظام هو إتباع الأبعاد المقترحة من طرف رالف كيمبل. ونهجنا يقسم الى ثلاث مراحل هي : تحديد الحاجة وتصميم النظام ونشر النظام.

# SOMMAIRE

## PARTIE I :

### INTRODUCTION GENERALE

I- Avant propos.....	1
II- Problématique et cadre applicatif.....	1
III- Objectifs.....	3

## PARTIE II :

### Présentation Général du Data Warehouse

<b>1. Naissance du Data Warehouse</b> .....	4
<b>2. Historique</b> .....	5
<b>3. Informatique décisionnel</b> .....	6
<b>4. Problématique</b> .....	8
<b>5. Data warehouse</b> .....	9
5.1. Définition .....	9
5.1.1. Données orientées sujet (thématique) .....	10
5.1.2. Données intégrées .....	11
5.1.3. Données non volatiles .....	12
5.1.4. Données Historisées .....	12
5.2. Data Warehousing ? .....	13
5.3. Objectifs de data warehouse .....	13
5.4. Fonctions de data warehouse .....	15
5.5. Les concepts de base d'un data warehouse .....	17
5.5.1. Architecture générale d'un data warehouse .....	17
5.6. Fouille de données .....	20
5.6.1. Data mining .....	20
5.6.2. Outils statistiques .....	21
5.6.3. EIS .....	22
5.6.4. SIAD .....	22
5.6.5. Requêteur .....	22
<b>6. Data warehouse et base de production (SGBD)</b> .....	23
<b>7. Le concept OLAP « One Line Analytical Processing »</b> .....	24
7.1. Définition .....	24
7.2. Les règles d'OLAP .....	24
7.3. Systèmes OLTP versus systèmes OLAP .....	26
<b>8. Les composants fonctionnels d'un data warehouse</b> .....	28
8.1. Alimentation d'un data warehouse .....	28
a. Découverte des données .....	28
b. Extraction des données .....	28
c. Transformation des données .....	29

d. Chargement des données Homogénéisation .....	29
8.2. Le processus de stockage .....	30
8.3. Le processus d'accès .....	31
<b>9. Construction de data warehouse .....</b>	<b>31</b>
9.1. Modélisation des données de l'entrepôt .....	31
9.1.1. Modélisation entité relation .....	31
9.1.2. Modélisation dimensionnelle .....	32
a. Caractéristiques du modèle dimensionnel .....	33
9.1.2.1. Modélisation en Etoile .....	35
9.1.2.2. Modélisation en Flocon .....	38
9.1.2.3. Modélisation en constellation .....	39
<b>10. Structure du data warehouse .....</b>	<b>41</b>
10.1. Processus de conception pour une base de données dimensionnelle (R.Kimball) .	42
10.1.1. Choisir les processus d'activité opérationnelle à modéliser .....	42
10.1.2. Définition du grain .....	42
10.1.3. Modification du grain .....	43
10.1.4. Normalisation .....	43
<b>11. BD multidimensionnelles .....</b>	<b>43</b>
<b>12. Outils de Navigation .....</b>	<b>44</b>
12.1. OLAP Tools .....	44
12.1.1. Drill up .....	44
12.1.2. Drill down .....	44
12.1.3. Slicing .....	45
12.1.4. Dicing .....	46
12.1.5. Data Surfing .....	46
12.1.6. Pivoting (Rotate) .....	46
12.2. Data marts .....	47
<b>13. Conclusion.....</b>	<b>48</b>

## PARTIE III :

### Présentation de l'organisme d'accueil

<b>Introduction.....</b>	<b>49</b>
<b>1. Missions de « GTP » .....</b>	<b>49</b>
<b>2. Activités.....</b>	<b>50</b>
<b>3. Organisation de l'entreprise.....</b>	<b>51</b>
3.1. Organigramme général de ENGTP.....	51
3.2. Organisation.....	52
3.3. Organigramme du centre informatique.....	53
<b>4. Environnement de l'entreprise.....</b>	<b>54</b>
4.1. Politiques commerciales.....	56
4.1.1. Produit .....	56
4.1.2. Surveillance des marchés.....	56
4.1.3. La concurrence.....	57

4.1.4. Les Clients.....	57
4.1.5. Fournisseurs.....	58
4.1.6. Production.....	58
4.1.7. Partenariat.....	58
<b>5. Ressources et politique de l'entreprise.....</b>	<b>59</b>
5.1. Humaines.....	59
5.2. Matérielles.....	59
5.3. Situation informatique de l'entreprise.....	60
5.3.1. le réseau informatique.....	60
<b>6. Notoriété de GTP.....</b>	<b>63</b>
<b>7. Objectifs de GTP.....</b>	<b>63</b>
<b>8. La finalité du projet.....</b>	<b>63</b>
<b>9. Conclusion.....</b>	<b>64</b>

## **PARTIE IV : Définition des besoins**

<b>1. Introduction.....</b>	<b>65</b>
<b>2. Consistance du Projet de fin d'études.....</b>	<b>66</b>
<b>3. Choix de la méthode de définition des besoins.....</b>	<b>66</b>
<b>4. Définition des besoins.....</b>	<b>67</b>
<b>5. Objectif.....</b>	<b>68</b>
5.1 Compte rendu des entretiens.....	68
6. Les processus métiers.....	70
6.1 Identification des processus.....	70
<b>7. La dimensionnalité métier.....</b>	<b>71</b>
<b>8. Les indicateurs d'analyse de l'activité.....</b>	<b>73</b>
<b>9. Les systèmes sources.....</b>	<b>74</b>
<b>10. Conclusion.....</b>	<b>74</b>

## **PARTIE V : Conception du système**

<b>1. Modélisation dimensionnelle.....</b>	<b>75</b>
1.1. Modélisation dimensionnelle de l'activité « <b>Réalisation Personnel</b> ».....	75
1.2. Modélisation dimensionnelle de l'activité « <b>Personnel sortant</b> ».....	78
1.3. Modélisation dimensionnelle de l'activité « <b>Personnel inactif</b> ».....	80
1.4. Modélisation dimensionnelle de l'activité « <b>Mouvement personnel</b> ».....	81
1.5. Modélisation dimensionnelle de l'activité « <b>Réalisation financière</b> ».....	82
<b>2. Mise en œuvre de l'entrepôt.....</b>	<b>83</b>
2.1. Construction de la base de l'entrepôt.....	85
2.2. Construction des cubes OLAP.....	88
2.3. Construction de la zone d'alimentation.....	93

2.4. La construction du portail de restitution.....	95
<b>3. Conclusion.....</b>	<b>97</b>

**PARTIE VI :  
Déploiement du système**

<b>1. Architecture technique de notre système.....</b>	<b>98</b>
<b>2. Disposition physique de notre système.....</b>	<b>99</b>
<b>3. Les utilisateurs du système.....</b>	<b>100</b>
<b>4. Le catalogue des états.....</b>	<b>100</b>
<b>5. Logiciels utilisés.....</b>	<b>101</b>
5.1. Les composants serveurs.....	101
5.2. Les composants clients.....	101
<b>6. Matériel utilisé.....</b>	<b>101</b>
<b>7. La sécurité du système.....</b>	<b>101</b>
<b>8. Navigation.....</b>	<b>102</b>
<b>Conclusion Générale .....</b>	<b>109</b>

## Liste des figures

Fig-01	Approche du Data warehouse.....	2
Fig-02	Schéma d'un entrepot de données.....	7
Fig-03	Définition de l'entrepôt de données.....	9
Fig-04	Données orientées sujet.....	11
Fig-05	Données intégrées.....	11
Fig-06	Données non volatiles.....	12
Fig-07	Entrepot de données.....	13
Fig-08	Fonctions de data warehouse.....	16
Fig-09	L'architecture de l'entrepôt de données.....	17
Fig-10	Requetes dans les entrepots de données.....	19
Fig-11	Le data mining dans les data warehouses.....	20
Fig-12	Exploitation d'un entrepôt de données.....	21
Fig-13	Composants fonctionnels d'un data warehouse.....	28
Fig-14	Homogénéisation de données.....	29
Fig-15	Le stockage.....	30
Fig-16	Représentation des données sous forme d'un cube (multidimensionnelle).....	34
Fig-17	Modélisation en étoile.....	35
Fig-18	Modèles d'hierarchies.....	38
Fig-19	Modélisation en flocon.....	39
Fig-20	Modélisation en constellation.....	40
Fig-21	Structure de data warehouse logique d'accès aux données.....	41
Fig-22	Cube OLAP.....	44
Fig-23	Drill Up, Drill Down.....	45
Fig-24	Exemple de Slicing.....	45
Fig-25	Exemple de Dicing.....	46
Fig-26	Data Mart.....	47
Fig-27	Organigramme général de GTP.....	51
Fig-28	Organigramme du centre informatique.....	53
Fig-29	Architecture Client / Serveur à GTP.....	61

Fig-30	Exemple de cube.....	72
Fig-31	Modélisation dimensionnelle en étoile de l'activité « <b>Réalisation Personnel</b> ».....	78
Fig-32	Modélisation dimensionnelle en étoile de l'activité « <b>Personnel Sortant</b> ».....	79
Fig-33	Modélisation dimensionnelle en étoile de l'activité « <b>Personnel Inactif</b> ».....	81
Fig-34	Modélisation dimensionnelle en étoile de l'activité « <b>Mouvement personnel</b> »..	82
Fig-35	Modélisation dimensionnelle en étoile de l'activité « <b>Réalisation Financière</b> »...	83
Fig-36	Architecture détaillée de notre data warehouse.....	84
Fig-37	Mesures d'activités.....	90
Fig-38	Dimension.....	90
Fig-39	Cube MOLAP de l'activité <b>Réalisation Personnel</b> .....	91
Fig-40	Cube MOLAP de l'activité <b>Mouvement Personnel</b> .....	91
Fig-41	Cube MOLAP de l'activité <b>Personnel Inactif</b> .....	92
Fig-42	Cube MOLAP de l'activité <b>Personnel Sortant</b> .....	92
Fig-43	Cube MOLAP de l'activité <b>Réalisation Financière</b> .....	93
Fig-45	Le processus d'alimentation de notre entrepôt.....	94
Fig-46	La fréquence de chargement des systèmes sources vers l'entrepôt.....	95
Fig-47	Les outils d'accès à notre data warehouse.....	96
Fig-48	Disposition physique de notre système.....	99
Fig-49	Drill-up.....	102
Fig-50	Drill-down.....	103
Fig-51	Slice and Dice.....	103
Fig-52	Pivoting.....	104
Fig-53	Connexion au schéma et choix du classeur .....	104
Fig-54	Affichage classeur « <b>Mouvement Personnel</b> ».....	105
Fig-55	Choix du type de graphe.....	105
Fig-56	Pivot des axes.....	106
Fig-57	Affichage en mode matricielle.....	107

## Liste des tableaux

Tab-01	Qu'est ce qu'un data warehouse ? .....	10
Tab-02	Système de production vs Data Warehouse.....	23
Tab-03	OLTP vs OLAP.....	27
Tab-04	Représentation des données dans une table relationnelle.....	32
Tab-05	Représentation des données dans un tableau multidimensionnel.....	34
Tab-06	Différence entre attribut et fait.....	37
Tab-07	Evolution des effectifs par statut.....	59
Tab-08	Liste des tables de l'entrepôt de données.....	86
Tab-09	Liste des attributs de l'entrepôt de données.....	87
Tab-10	Liste des faits et des dimensions sollicitées.....	89



Partie .1



# Introduction Générale

## **I. Avant-propos**

Dans le contexte d'évolution rapide des marchés, les entreprises font face à une concurrence accrue. La multiplication des exigences des clients contribue en outre à modifier le paysage économique traditionnel.

La mondialisation conduit les utilisateurs, où qu'ils soient dans le monde, à devoir accéder aux données 24h/24 et 7j/7. De plus, les réglementations et les initiatives de mise en conformité nécessitent l'accessibilité immédiate aux informations.

Pour répondre à ces exigences complexes, les entreprises ont besoin d'une solution capable d'intégrer en temps réel, les données de leurs systèmes opérationnels disparates dans l'entreprise de façon à disposer d'une vue complète de leurs clients et de leurs activités.

Dans ce contexte, le meilleur outil stratégique est le data warehousing. En effet, il permet de transformer la principale richesse des entreprises, les données contenues dans les systèmes informatiques, en informations décisionnelles. En outre, le data warehousing permet d'identifier de nouvelles opportunités de développement, et de suivre en temps réel la progression de l'entreprise.

Bien utilisé, le data warehouse constitue une arme stratégique de développement.

## **II. Problématique et cadre applicatif**

ENGTP est l'une des plus grandes entreprises de construction et des grandes réalisations dans le secteur de l'énergie et des hydrocarbures.

Ses premières applications ont été développées en 1970 possédant comme noyau un système de comptabilité générale et analytique et des satellites permettant la prise en charge de l'ensemble des fonctions de gestion de l'entreprise. Noyau et satellites communiquaient à travers des interfaces spécialement conçues.

En 2000, elle est passée du stade d'informatique à traitement par lot vers une informatique sous base de données et en mode client/serveur. L'architecture des systèmes étant restée sensiblement la même.

Le système d'information de suivi de la ressource humaine communique avec les systèmes Comptabilité / Finance, gestion du personnel.

Cette approche est illustrée par la figure suivante :



**Fig -1-** Approche du Data warehouse

Pour l'analyse de l'activité suivi des projets de l'entreprise, les décideurs utilisent des rapports d'activité en provenance de :

- 1 SIG.
- 2 BDD (Oracle 9i, developer 2000,...).
- 3 Système de gestion du personnel (table oracle + tableaux Excel).

Les sources d'information hétérogènes diminuent les performances des analystes, vu le volume important des données, l'augmentation continue de son activité, l'analyse devient plus difficile et plus lente, ainsi l'utilisateur devra passer 70 % du temps à chercher la donnée et 30% seulement à l'analyser. Et il arrive parfois que les données disponibles restent inutilisées simplement parce que leurs utilisateurs ne connaissent pas leur existence.

Bien que l'entreprise ait des moyens pour analyser son activité comme les tableaux de bord mais les managers ont besoin d'avoir une vue consolidée des mesures d'activités et de leurs indicateurs clés de performance pour prendre les bonnes décisions en ayant le pouvoir de choisir l'axe sur lequel l'analyse sera portée ainsi que la possibilité de créer son propre état sous le format désiré.

Le besoin principal des utilisateurs est de retrouver à tout moment (temps réel) une information exacte, utile et transverse concernant un employé. Pour cela, la mise en œuvre d'un data warehouse métier pour la fonction de la ressource humaine semble nécessaire.

### **III. Objectifs**

L'ambition de GTP à travers la conception d'un système décisionnel basé sur l'architecture data warehouse est de simplifier le pilotage des activités, mieux analyser et surveiller les indicateurs économiques de l'entreprise, ceci en ayant la possibilité d'accès aux informations pertinentes en temps réel.

Notre objectif est de concevoir et déployer un data warehouse métier pour la fonction des projets.

Cette réalisation va permettre de :

1. Améliorer les performances décisionnelles de l'entreprise.
2. Retrouver une information historique et transversale de l'entreprise (non volatiles).
3. Obtenir des informations déjà agrégées selon les besoins de l'utilisateur : simplicité et rapidité d'accès.
4. Offrir à l'utilisateur liberté de navigation.
5. Homogénéiser les sources de données en ce qui concerne le suivi des projets.
6. Ajouter de la valeur aux données et les manipuler.
7. Analyser et prendre des décisions rapidement.

En effet, le principal objectif d'un système décisionnel est de donner à l'utilisateur les moyens d'être autonome, en lui permettant de naviguer dans l'information en fonction des questions qu'il se pose.

“ La mise en œuvre d'un data warehouse est une réussite lorsque l'utilisateur dit : J'ai l'information, elle est sûre, je la comprends donc je l'utilise ” [J-M. Franco, 96].



Partie .2



# Présentation générale du data warehouse

## 1. Naissance du Data Warehouse

Pour obtenir une vision complète de leur performance, les entreprises ont besoin d'intégrer l'ensemble des données relatives à leur activité, et de fournir un accès simple aux informations vitales. La valeur ajoutée du data warehouse réside dans sa capacité à opérer en des lieux géographiques différents, sur des plateformes hétérogènes et dans des environnements locaux divers.

Il peut alors fournir des informations utiles sur lesquelles des décisions efficaces peuvent être prises, et contribuer ainsi au succès de l'entreprise.

Les volumes de données à traiter dans le cadre de l'apprentissage automatique et de la fouille de données sont de plus en plus importants. L'étude des interactions possibles entre les domaines des bases de données et de l'apprentissage est donc nécessaire pour pallier les problèmes liés à la gestion de ces gros volumes de données.

Apparus pour gérer de tels volumes de données issues de sources hétérogènes, les entrepôts de données constituent l'outil essentiel de collecte et de mise à disposition des données en vue de leur analyse. De ces entrepôts de données peuvent être extraits des magasins de données, contenant des sous-ensembles de données dédiés à une analyse particulière. [Aexis.BI]

- ✓ Rapides à mettre en œuvre et simples à adapter aux changements
- ✓ Suffisamment flexible pour évoluer dans le temps et s'adapter à chaque nouveau développement de l'entreprise.
- ✓ Economiquement rentable, avec un retour sur investissement de moins d'un an.
- ✓ Conçue de manière à entreprendre aisément des projets d'intégration d'envergure.
- ✓ Complet, il restitue une vision globale des informations de l'entreprise, et cela dans le cadre d'un grand groupe international.
- ✓ Intelligent, il simplifie la manipulation d'informations provenant de sources disparates.

## 2. Historique

« Fournir à tout utilisateur reconnu et autorisé, les informations nécessaires à son travail ».

Ce slogan fait naître une nouvelle informatique, intégrante, orientée vers les utilisateurs et les centres de décision des organisations. C'est l'ère du client-serveur qui prend vraiment tout son essor à la fin des années 90 avec le développement des technologies intranet. Un environnement de concurrence plus pressant contribue à révéler l'informatique décisionnelle. Tout utilisateur de l'entreprise ayant à prendre des décisions doit pouvoir accéder en temps réel aux données de l'entreprise, doit pouvoir les traiter et en extraire les informations pertinentes pour prendre les << bonnes >> décisions. Il se pose des questions du type : << Quels sont les besoins en ressource humaine par spécialité par région pour l'année dernière ? >> ; << Quelle est l'évolution du mouvement du personnel par type de contrat et par période >> ; ou encore << Comment qualifier les employés pour le post X ? >>

Le système opérationnel ne peut satisfaire ces besoins car les bases de données opérationnelles sont trop complexes pour pouvoir être appréhendées facilement par tout utilisateur et le système opérationnel ne peut être interrompu pour répondre à des questions nécessitant des calculs importants. Il s'avère donc nécessaire de développer des systèmes d'information orientés vers la décision. Il faut garder un historique et restructurer les données de production, éventuellement récupérer des informations démographiques, géographiques et sociologiques. Les entrepôts de données ou data warehouse sont la réalisation de ces nouveaux systèmes d'information. Cette apparition est rendue possible grâce aux progrès technologiques à coûts constants (grâce à l'augmentation importante des capacités de stockage et à l'introduction des techniques du parallélisme dans l'informatique de gestion, techniques qui permettent des accès rapides à de grandes bases de données).

[ Univ3-lille ]

### **3. Informatique décisionnelle**

L'informatique décisionnelle en anglais DSS (Decision Support System ou encor BI Business Intelligence) s'est développée dans les années 70. Elle est alors essentiellement constituée d'outils d'édition de rapports, de statistiques, de simulation et d'optimisation ainsi que de nombreux algorithmes permettant d'extraire des informations à partir de données brutes. Ces algorithmes ont des origines diverses et souvent multiples. Certains sont issus des statistiques, d'autres proviennent des recherches en Intelligence Artificielle, tous ces algorithmes sont regroupés dans des logiciels de fouille de données ou Data Mining qui permettent la recherche d'informations nouvelles ou cachées à partir de données ainsi, dans le cas de systèmes à base de règles, plutôt que d'essayer d'extraire la connaissance d'experts et d'exprimer cette connaissance sous forme de règles, un logiciel va générer ces règles à partir de données. [AW.96]

Les systèmes décisionnels permettent de:

- ✓ Répondre aux besoins d'aide à la prise de décision,
- ✓ Compléter les systèmes opérationnels,
- ✓ Fournir l'accès aux informations à des utilisateurs répartis dans l'entreprise (système transversal).

La démarche adoptée par ces systèmes consiste à:

- ✓ Utiliser des données cohérentes et vérifiées,
- ✓ Appliquer les méthodes de modélisation et d'analyses,
- ✓ Accéder simplement et intuitivement aux données pertinentes,
- ✓ Etablir des règles de décisions,
- ✓ Mettre en forme des résultats.

L'analyse décisionnelle se donne pour but d'apporter une vision analytique et prospective de l'activité d'une entreprise en partant de données diverses recueillies sur les différentes parties de son système d'information.

A partir de cette moisson de contenus, elle fournit au final des tableaux de bord composés d'indicateurs de performance permettant aux décideurs de suivre l'évolution de la ressource humaine à tous les étages.

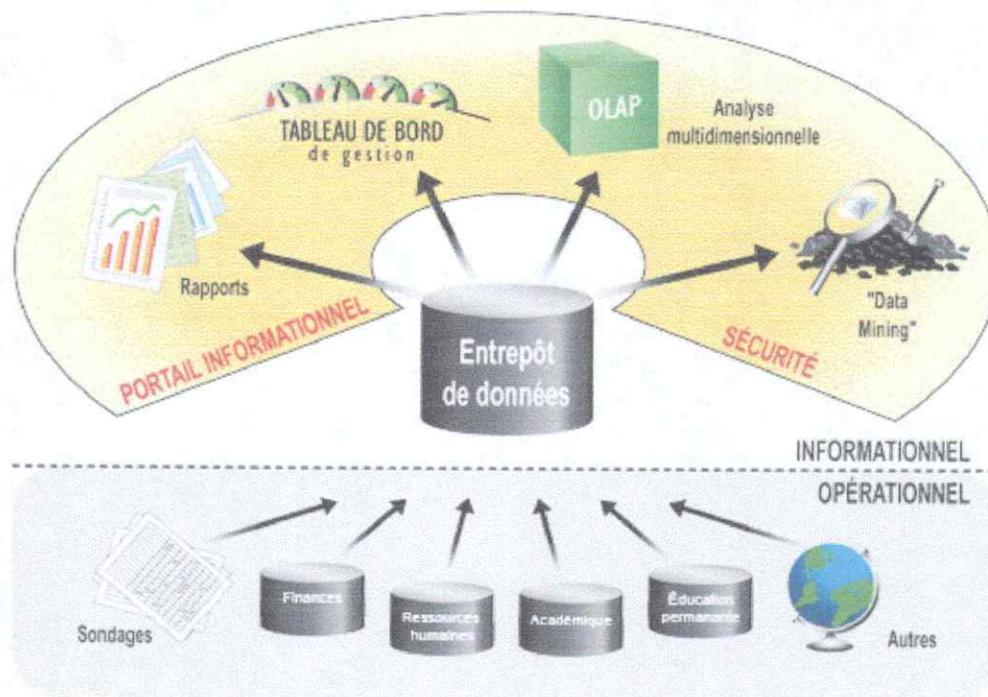


Fig -2- Schéma d'un entrepot de données

Plusieurs méthodes existent pour mettre en oeuvre la fouille de données. Le choix de l'une d'entre elles est une première difficulté pour l'utilisateur ou le concepteur. Aucune méthode n'est meilleure qu'une autre dans l'absolu. Néanmoins, l'environnement, les contraintes, les objectifs et bien sûr les propriétés des méthodes doivent guider l'utilisateur dans son choix.

Les entrepôts de données et la fouille de données sont les éléments d'un domaine de recherche et de développement très actifs actuellement : l'extraction de connaissances à partir de données ou Knowledge Discovery in Databases (KDD). Par ce terme, on désigne tout le cycle de découverte d'information. Il regroupe donc la conception et les accès à de

grandes bases de données ; tous les traitements à effectuer pour extraire de l'information de ces données ; l'un de ces traitements est l'étape de fouille de données ou Data Mining.

Les data warehouse sont construits sur une nouvelle architecture permettant d'extraire l'information, architecture bien différente de celle prévue pour l'informatique de production, basée elle sur des systèmes de gestion de bases de données relationnelles et des serveurs transactionnels. Un entrepôt de données est construit en l'alimentant via les serveurs transactionnels de façon bien choisie et réfléchie pour permettre aux procédures d'extraction de connaissances de bien fonctionner. L'organisation logique des données est particulièrement conçue pour autoriser des recherches complexes. Le matériel est évidemment adapté à cette utilisation.

## 4. Problématique

Une grande masse de données :

- Distribuées
- Hétérogènes
- Très détaillée

A traiter :

- Synthétiser / Résumer
- Visualiser
- Analyser

Pour une utilisation par :

- Des experts et des analystes d'un métier
- Non informaticiens
- Non statisticiens

## 5. Data warehouse

### 5.1. Définition

Littéralement, le data warehouse est « l'entrepôt de données ». Pratiquement, c'est une base spécialement conçue pour l'analyse des données et l'aide à la décision qui consolide toutes les données disparates de l'entreprise. Elle est construite à partir des bases de production. Elle se distingue de l'infocentre traditionnel par le fait que les données sont transformées.

Il est utilisé comme stockage intermédiaire des données issues des applications de production et rassemble une copie des données vitales de l'entreprise, les utilisateurs finaux y puisent avec des outils de restitution et d'analyse comme les décisionnels, pour suivre l'activité ou déterminer des tendances.

« C'est l'ensemble unique, complet et consistant de données obtenues depuis une variété de sources et rendu disponible à des utilisateurs finaux à des fins de compréhension et de gestion des affaires décrites par ces données. » [B. Devlin 97]

« Le data warehouse est une collection de données thématiques, intégrées, non volatiles et historisées, organisées pour le support d'un processus d'aide à la décision » [B.Inmon.95]

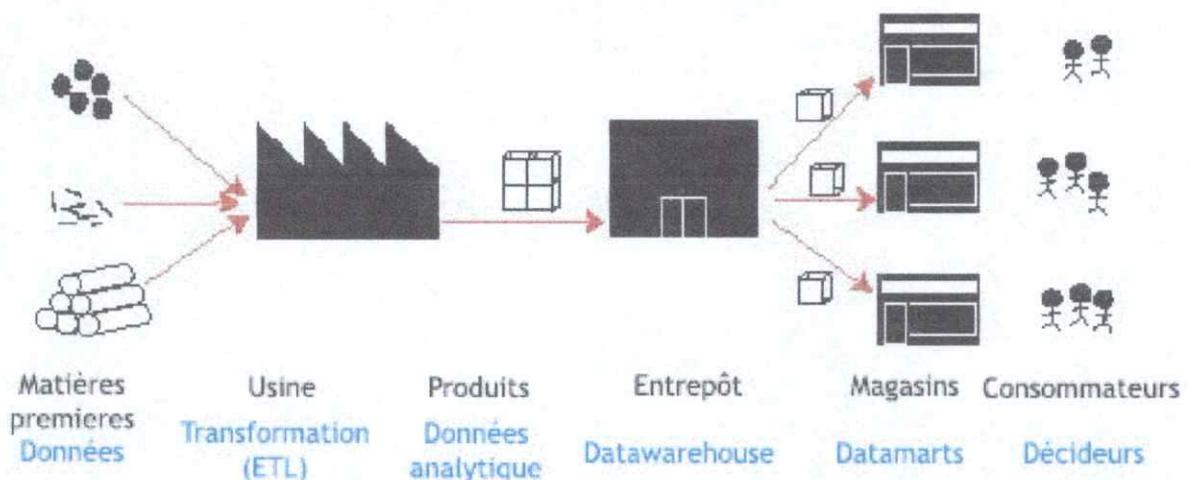


Fig -3- Définition de l'entrepôt de données

Ensemble de données	Ensemble d'outils permettant
- Destinées aux décideurs.	- De regrouper les données.
- Souvent une copie des données de production.	- De nettoyer, d'intégrer les données.
- Avec une valeur ajoutée (agrégation, historique).	- De faire requêtes, rapports d'analyse.
- Intégrées, Historisées.	- De faire du datamining.
	- Faire l'administration du data warehouse.

Tab -1- Qu'est ce qu'un data warehouse?

### 5.1.1. Données orientées sujet (thématique)

Le data warehouse est organisé autour des sujets majeurs et des métiers de l'entreprise, les données sont organisées par thème, l'intérêt de cette organisation réside dans le fait qu'il devient possible de réaliser des analyses sur des sujets transversaux aux structures fonctionnelles et organisationnelles de l'entreprise. Cette orientation permet également de faire des analyses par itération, sujet après sujet.

L'intégration dans une structure unique est indispensable pour éviter aux données concernées par plusieurs sujets d'être dupliquées, cependant, dans la pratique, il existe également des data marts. Le data warehouse est fragmenté en plusieurs bases qui supportent les orientations sujet.

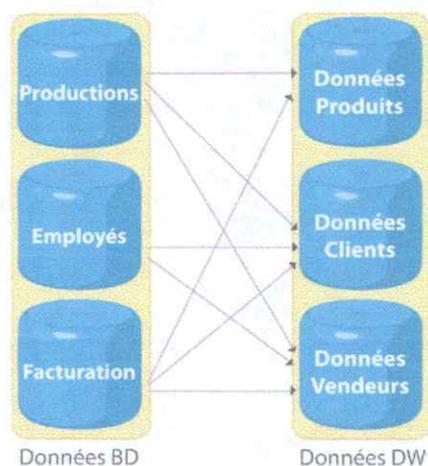


Fig -4- Données orientées sujet

### 5.1.2. Données intégrées

Les données proviennent de plusieurs sources différentes.

Un data warehouse est un projet d'entreprise, il concerne les différents services et métiers de l'entreprise. Avant d'être intégrées dans le data warehouse, les données doivent être mises en forme et unifiées afin d'avoir un état cohérent, l'intégration nécessite une forte normalisation, une bonne gestion des référentiels et de la cohérence, une parfaite maîtrise de la sémantique et des règles de gestion s'appliquent aux données manipulées. C'est ainsi que l'on pourra donner une bonne vision de l'entreprise via l'utilisation d'indicateurs.

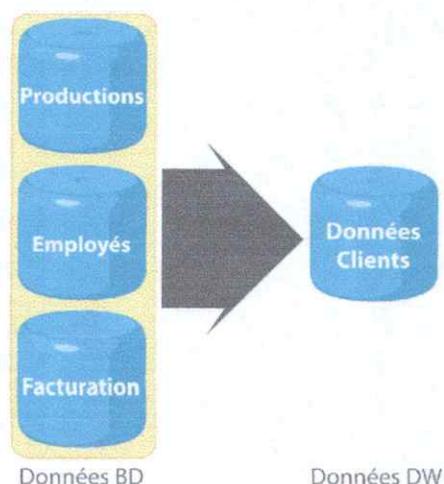


Fig -5- Données intégrées

### 5.1.3. Données non volatiles

Afin de conserver la traçabilité des informations et des décisions prises, les informations stockées au sein de l'entrepôt ne peuvent pas être supprimées, une requête lancée à différentes dates sur les mêmes données doit toujours retourner les mêmes résultats.

Une donnée introduite dans le data warehouse ne pourra donc plus être supprimée ni même modifiée. Dans le cas présent, les données ne sont pas volatiles.

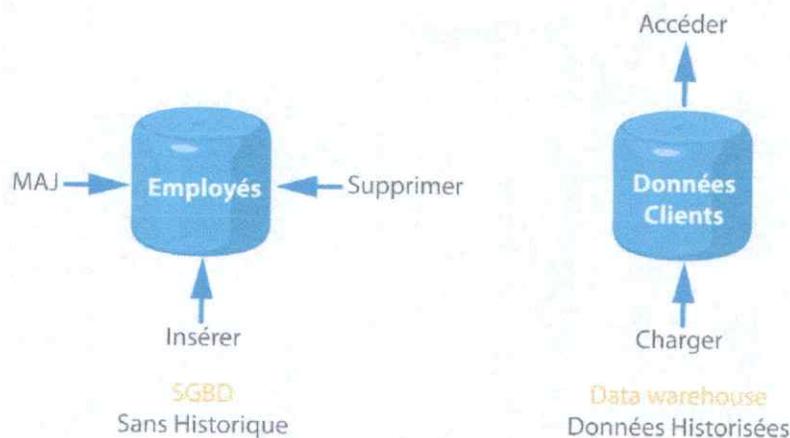


Fig -6- Données non volatiles

### 5.1.4. Données Historisées

Contrairement au système de production les données ne sont jamais mises à jour (quasiment), chaque nouvelle donnée est insérée.

L'historisation est nécessaire pour suivre dans le temps l'évolution des différentes valeurs des indicateurs à analyser. Ainsi, un référentiel temps doit être associé aux données afin de permettre l'identification dans la durée de valeurs précises. La portée temporelle des données dans un data warehouse est plus longue (fournit des informations sous une perspective historique, exemple : de 5 à 10 dernières années) que celle des bases opérationnelles (valeur courante des données et peut ne pas faire référence au temps).

## 5.2. Data Warehousing ?

« Entreposage de données » : Un processus qui permet de construire un data warehouse à partir des bases de production (toutes les sources de données (légal, juridiques, fiscales, politiques, techniques, marketing)).

Le data warehousing correspond aux techniques et méthodes nécessaires à la création d'un data warehouse ainsi qu'à son exploitation. Ce sont les processus de création et d'alimentation du data warehouse ainsi que la possibilité de faire des requêtes.

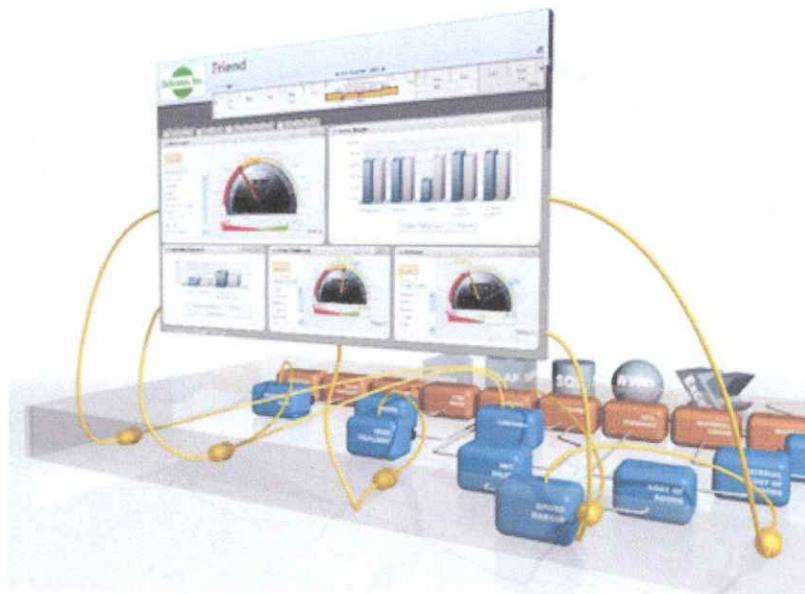


Fig -7- Entrepot de données

Un entrepôt de données définit donc un ensemble de données et un ensemble d'outils, les données sont destinées aux décideurs qui sont souvent des copies des données de production avec une valeur ajoutée (orientées objet, agrégées, historiques) et un ensemble d'outils permettant de regrouper les données des différentes sources, de les nettoyer de les intégrer ainsi que d'y accéder de différentes manières (requêtes, rapport, analyse, datamining).

### 5.3. Objectifs de data warehouse

L'essentiel est de disposer d'informations dans les meilleurs délais pour prendre des décisions rapidement. Il s'agit aussi d'un outil de statistique et de simulation. Il s'agit de construire une base indépendante de ou des bases de productions (celles qui se mettent à jour) qui consolide toutes les informations sur un sujet. Par opposition, une base de production contient des données ayant un intérêt limité dans le temps, le data warehouse doit répondre aux contraintes techniques principales suivantes : volumétrie, structure des données, performance, fiabilité, haute disponibilité. L'enjeu est de transformer les données en informations directement exploitables par les utilisateurs du métier concerné.

Les principaux objectifs du data warehouse [Kim 97] :

- ❖ Accès aux informations de l'entreprise : Le data warehouse assure l'accès aux informations de l'entreprise et de l'organisation.
- ❖ Les informations sont cohérentes : Cela signifie que les requêtes faites à des moments différents doivent fournir les mêmes résultats. La cohérence implique que les données sont chargées dans leur totalité.
- ❖ Les outils de présentation d'information font partie du data warehouse : L'entrepôt de données ne comporte pas seulement des données, mais aussi un ensemble de requêtes, d'analyses et de présentations des informations.
- ❖ Les données publiées sont stockées dans le data warehouse : Le data warehouse est le lieu où sont publiées des données qui ont déjà servi. Les données sont soigneusement rassemblées à partir de source d'informations situées à différents endroits de l'organisation. Elles sont nettoyées, leur qualité est vérifiée et elles ne sont diffusées que si elles sont prêtes à être utilisées. Si l'information est peu fiable ou incomplète, les données ne peuvent être publiées à destination de la communauté des utilisateurs.
- ❖ Qualité de l'information d'un data warehouse : La qualité de l'information d'un data warehouse est très importante. En effet comment obtenir des analyses fiables si les données brutes sont de mauvaise qualité.

## 5.4. Fonctions de data warehouse

Le data warehouse est un outil d'aide à la décision basé sur une base de données fédérant et homogénéisant les informations des différents services d'une organisation.

« Le défi du data warehouse, c'est le passage de l'informatique de gestion à la gestion de l'information ». [J.P. Minarro], Il constitue la forme la plus sophistiquée des systèmes d'aide à la décision, il peut stocker un volume très important de données. Ses fonctions sont :

- Récupérer les données existantes (bases de données de production (source), fichiers plats, fichiers externes, ...).
- Nettoyer ces données.
- Référencer les données (utilisation de métadonnées : données sur les données).
- Stocker les données historisées (constitution d'historique en vue d'analyse, comportementales en particulier).
- Mettre à disposition les données (permettre à chaque utilisateur d'accéder à l'information pertinente). Interrogation, visualisation, analyse.

## Partie 2 : Présentation Général du Data Warehouse

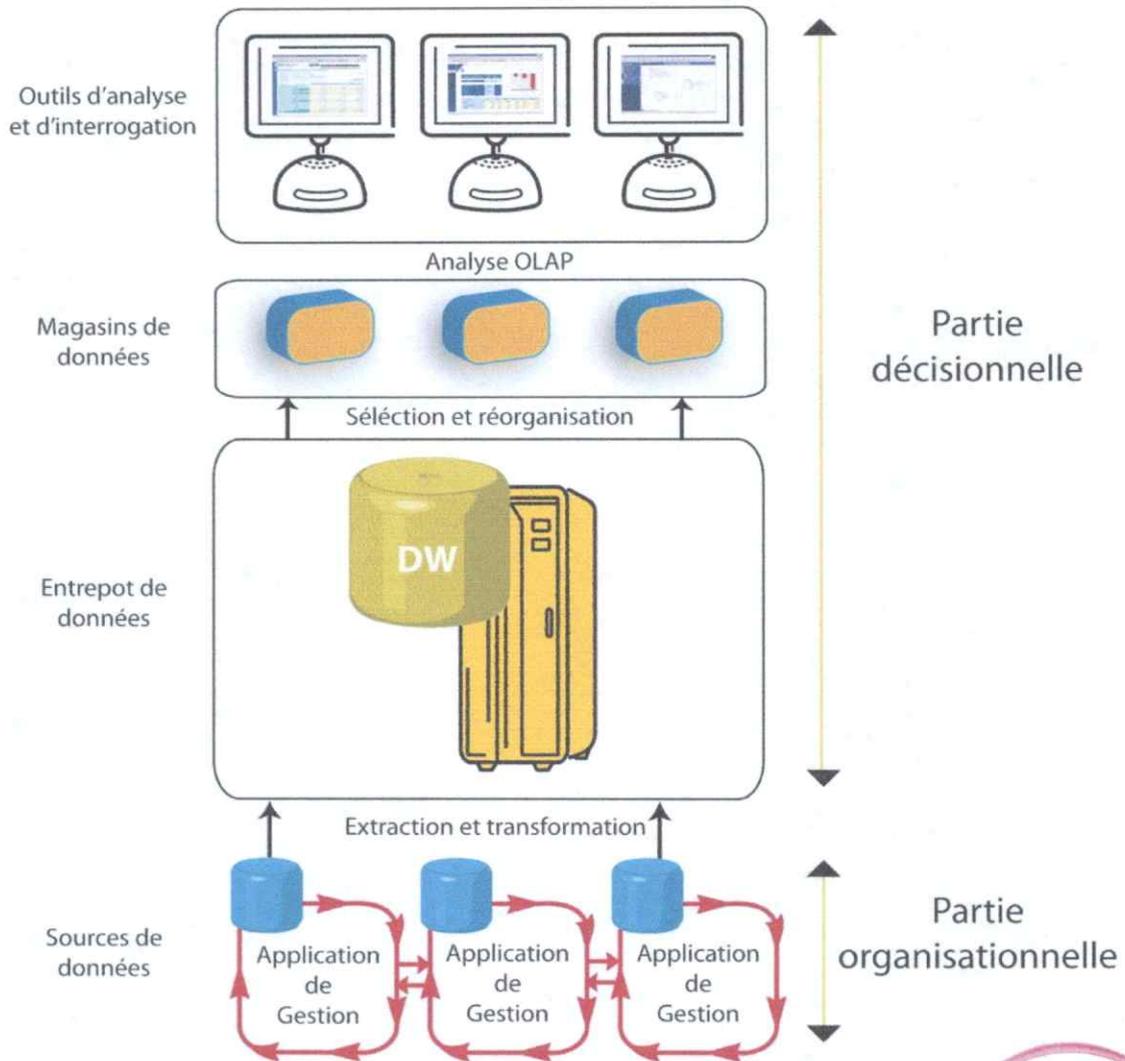


Fig -8- Fonctions de data warehouse



## 5.5. Les concepts de base d'un data warehouse

### 5.5.1. Architecture générale d'un data warehouse

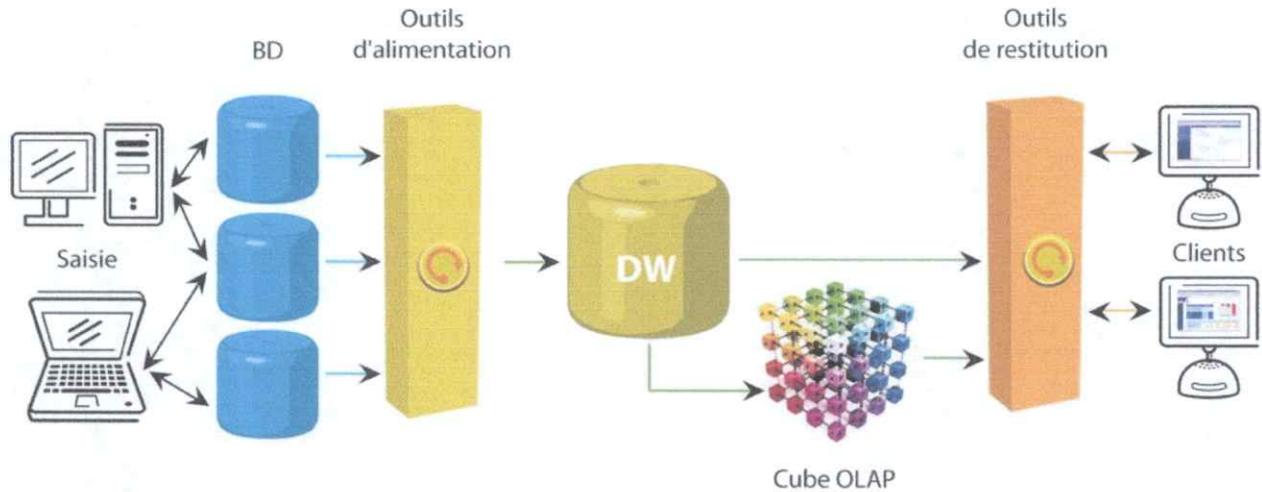


Fig -9- L'architecture de l'entrepôt de données

- **Le système source**

C'est le système opérationnel d'enregistrement, dont la fonction consiste à capturer les transactions liées à l'activité de l'entreprise [Kim 97]. Il s'agit souvent de ce que l'on appelle les applications de gestion.

- **La zone de préparation**

Elle regroupe l'ensemble des processus qui nettoient, transforment, combinent, archivent, suppriment les données. Elle prépare les données sources en vue de leur intégration et de leur exploitation dans le data warehouse. La frontière qui détermine la zone de préparation des données est que cette dernière ne doit en aucun cas être accessible à l'utilisateur final par requêtes ou par un quelconque autre service de présentation.

Les outils d'alimentation sont utilisés pour extraire les données des bases de production et des bases d'informations internes et externes, pour les convertir, les transformer et en fin les stocker dans le data warehouse.

- **Le serveur de présentation**

Machine cible sur laquelle le data warehouse est stocké et organisé pour répondre en accès direct aux requêtes provenant des utilisateurs [Kim 97]. Des générateurs d'états ou d'autres applications. Le data warehouse est supporté par une base de données relationnelle, multidimensionnelle ou objet, même si celle-ci sont assez rares ou utilisées dans des contextes assez particuliers.

- **L'entrepôt de données**

C'est la source de données interrogeable de l'entreprise [Kim 97]. Il est alimenté par la zone de préparation de données.

- **Le Data mart**

« Un sous-ensemble logique d'un entrepôt de données » [Kim 97], On considère souvent le data mart appelé aussi magasin de données comme la réduction de l'entrepôt de données à un seul processus ou un groupe de processus ciblant un groupe métier spécifique, exemple : data mart ressources humaines, comptabilité, etc.

Le data mart minimise la complexité informatique. C'est une petite structure très ciblée et pilotée par les besoins utilisateurs, il est donc plus facile de se concentrer sur les besoins utilisateurs.

Le data mart est une base de données moins coûteuse que le Data warehouse et plus légère puisque destinée à quelques utilisateurs d'un département. Il séduit plus que le Data warehouse les candidats au décisionnel.

Il a la même vocation que le Data warehouse (fournir une architecture décisionnelle), mais vise une problématique précise avec un nombre d'utilisateurs plus restreint. En général, c'est une petite base de données (SQL ou multidimensionnelle) avec quelques outils, et alimentée par un nombre assez restreint de sources de données.

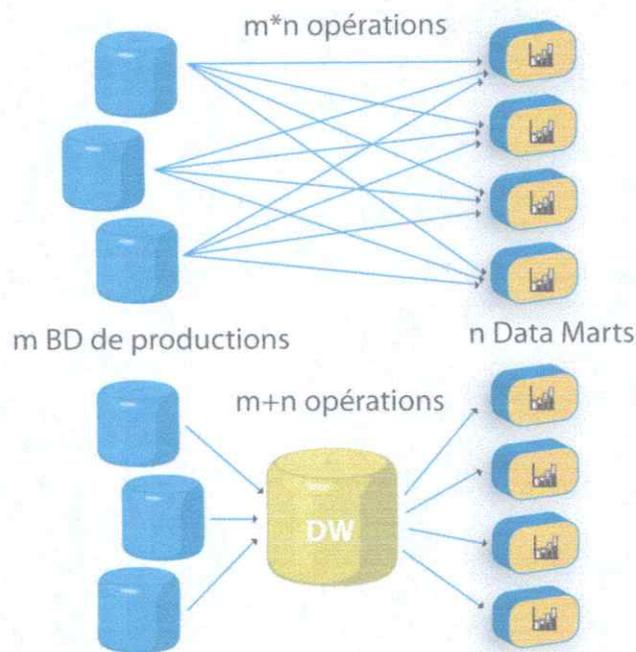


Fig -10- Requetes dans les entrepots de données

- **Le portail de restitution**

C'est la partie publique de l'entrepôt de données. Il représente ce que voient les utilisateurs, les outils avec lesquels ils travaillent. Sous ce terme sont regroupées toutes les applications qui s'appuient sur les données du data warehouse pour les restituer soit à l'utilisateur, soit à une autre application. Les services offerts par le portail de restitution sont les services d'accès aux données, les applications de modélisation et de data mining.

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

## 5.6. Fouille de données

Il existe plusieurs façons d'accéder aux données contenues dans le data warehouse ; les plus importantes sont :

### 5.6.1. Data mining

Les outils data mining, également appelés « forage des données » ou « extraction de la connaissance », s'appuient sur le constat qu'il existe au sein de chaque entreprise des informations dont le sens ou les liens sont cachés dans le gisement des données de l'entreprise. La fouille de données est l'un des outils d'interrogation de data warehouse.

Le data mining permet de faire apparaître des corrélations cachées dans des gisements de données. La figure suivante donne plus de détail sur le rapport entre les deux concepts « *entrepôt de données, fouille de données* »

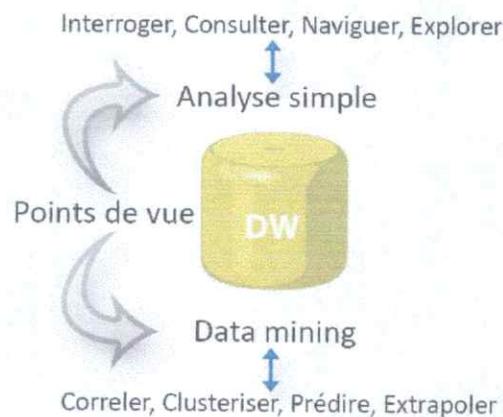


Fig -11- Le data mining dans les data warehouses

Data Mining = Processus inductif, itératif et interactif de découverte dans les BD larges de modèles de données valides, nouveaux, utiles et compréhensibles. [E-G. Talbi]

- Itératif : nécessite plusieurs passes.
- Interactif : l'utilisateur est dans la boucle du processus.
- Valides : valables dans le futur.

## Partie 2 : Présentation Général du Data Warehouse

- Nouveaux : non prévisibles.
- Utiles : permettent à l'utilisateur de prendre des décisions.
- Compréhensibles : présentation simple.

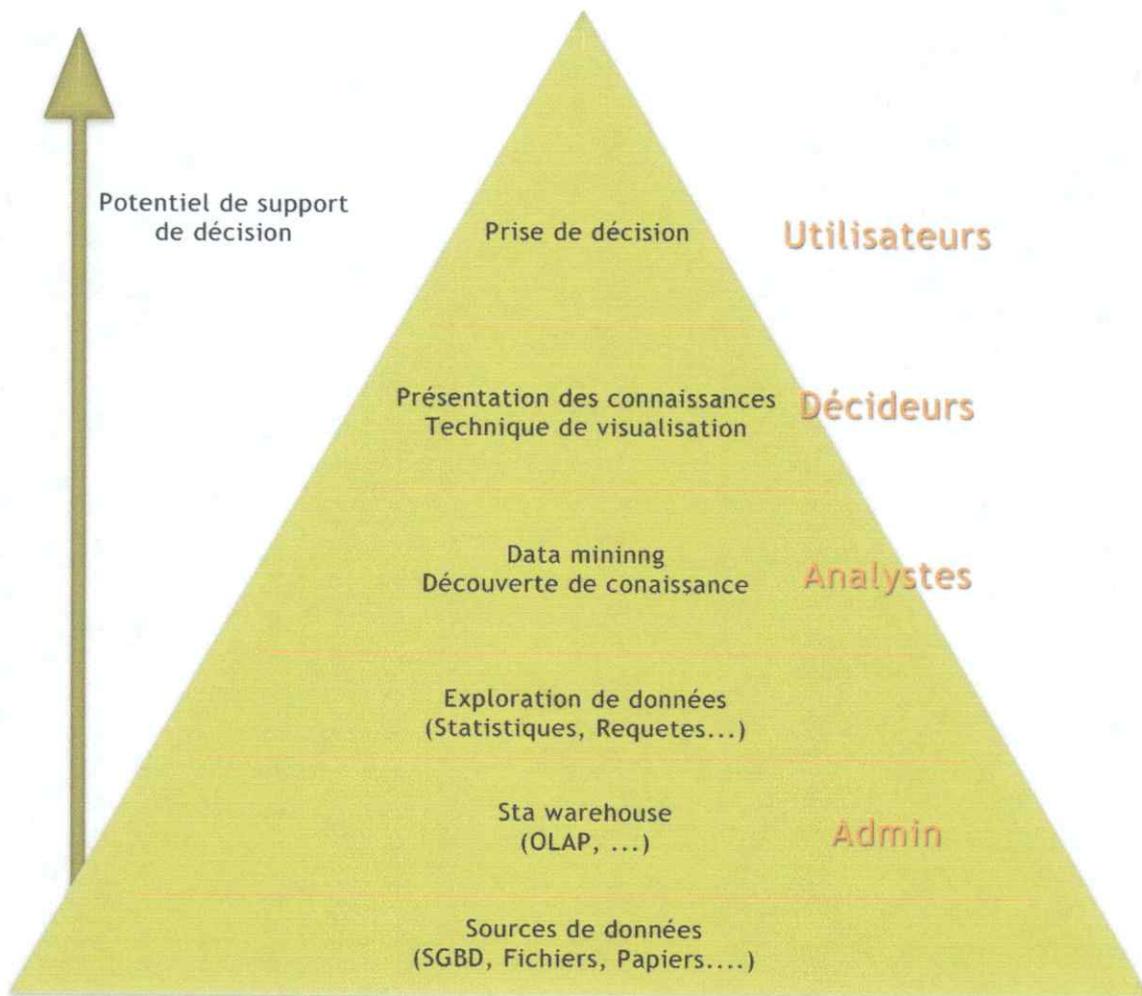


Fig -12- Exploitation d'un entrepot de données

### 5.6.2. Outils statistiques

Les applications ou les outils statistiques sont destinés à mettre en œuvre des fonctions mathématiques statistiques à des fins d'analyse.

### **5.6.3. EIS**

« Executive Information System ». Littéralement : système d'informations pour cadres. Un EIS est un ensemble d'outils qui permettent aux différents niveaux de management d'accéder aux informations essentielles de leur organisation, de les analyser et de les présenter de façon élaborée avec une interface graphique conviviale.

Le décideur a pour but :

- L'accès à l'information en temps réel.
- L'amélioration de la communication au sein de l'entreprise.
- Récolter l'information des différentes bases de données.
- Développer la compétitivité de son entreprise.
- Cerner l'information externe.

L'EIS est donc un outil adapté à la prise de décision stratégique (à long terme), c'est un outil de management de haut niveau (investissement, politiques d'expansion).

### **5.6.4. SIAD**

Un SIAD « Système Interactif d'Aide à la Décision » est un outil d'analyse et de modélisation des données de l'entreprise qui permet de créer des représentations multidimensionnelles de l'information, historiquement, il s'agit d'une terminologie et d'outils utilisés avant l'avènement et la maturité du data warehouse.

### **5.6.5. Requêteur**

Un requêteur permet à l'utilisateur final d'accéder aux données de l'entreprise de manière autonome, dans un langage propre à son métier, mais qu'il nécessite généralement la connaissance de la structure de la base accédée, et ce, en définissant lui-même les informations qu'il veut obtenir ainsi que le format des restitutions souhaitées.

## 6. Data warehouse et base de production (SGBD)

En réalité SGBD et data warehouse ont des objectifs différents. Ils stockent les données de manière différente et font l'objet de requêtes différentes. Ils sont ainsi basés sur deux systèmes différents : OLTP et OLAP

OLTP est le modèle utilisé par les SGBD, le mode de travail est transactionnel, l'objectif est de pouvoir insérer, modifier et interroger rapidement et en sécurité la base, ces actions doivent pouvoir être effectuées très rapidement par de nombreux utilisateurs simultanément sachant que chacune d'elles travaille sur de faibles quantités d'informations, et toujours sur les versions les plus récentes des données.

Les data warehouses eux reposent sur le système OLAP, ce système travaille en lecture seulement. Les programmes consultent d'importantes quantités de données pour procéder à des analyses, les objectifs principaux sont regrouper, organiser des informations provenant de sources diverses, les intégrer et les stocker pour donner à l'utilisateur une vue orientée métier, retrouver et analyser l'information facilement et rapidement cela nécessite de consulter des versions historiques de la base et peut se permettre d'ignorer temporairement les dernières mises à jour, ces bases sont souvent d'un ordre de grandeur nettement supérieur à celle des bases OLTP, du fait de la conservation de l'historique.

Critère	Système de production	Data Warehouse
Niveau de détail des informations utilisateurs	Très détaillées	Synthétique, parfois détaillées
Utilisateurs	Une ou quelques fonctions de l'entreprise	Plusieurs fonctions de l'entreprise
Données figées	Non – évolution en temps réel	Oui archivage
Historique	Non	Oui
Opérations sur les données	Ajout/MAJ/Consultation	Consultation uniquement

Tab -2- Système de production vs Data Warehouse

## **7. Le concept OLAP « *One Line Analytical Processing* »**

Les bases de données relationnelles représentent depuis de nombreuses années la référence en matière de stockage de l'information. Cette architecture n'étant pas toujours parfaitement adaptée, un nouveau concept s'est développé depuis quelques années.

### **7.1. Définition**

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

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

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

Ralph Kimball a défini le concept OLAP comme suit: « Activité globale de requêtage et de présentation de données textuelles et numériques contenues dans l'entrepôt de données ; style d'interrogation et de présentation spécifiquement dimensionnel » [Kim 97]. La technologie OLAP est non relationnelle et presque toujours basée sur un cube de données multidimensionnel explicite. Les bases de données OLAP sont également connues sous le nom de bases de données multidimensionnelles.

La base de données multidimensionnelle OLAP est incontestablement le cœur de tout système décisionnel. La vision multidimensionnelle qu'elle apporte aux données métiers en

fait d'elle la pierre angulaire de la prise de décisions. Sans pour autant réduire l'importance des autres briques du système décisionnel, un véritable outil d'aide à la décision ne saurait être construit sans s'appuyer sur une base de données OLAP.

## **7.2. Les règles d'OLAP**

OLAP est le terme pour décrire l'approche dimensionnelle de l'aide à la décision. Tout comme OLTP, OLAP a été proposé par Edgar Frank Codd 1993 qui a prescrit douze critères de conception qui sont les suivants :

### **\* Vue multidimensionnelle**

Permet d'avoir une vision multidimensionnelle des données (ce qui n'est pas le cas avec une table unidimensionnelle).

### **\* Transparence du serveur OLAP à différents types de logiciels**

L'utilisateur ne doit pas se rendre compte de la provenance des données si celles-ci proviennent de sources hétérogènes, ces sources peuvent être un fichier Excel, une base de données de production ou même un fichier texte.

### **\* Accessibilité à de nombreuses sources de données**

OLAP est décrit comme un middleware qui se place entre les sources de données hétérogènes et un front-end (sous la forme d'un data warehouse).

### **\* Performance du système de Reporting**

Les performances ne doivent pas être diminuées lors de l'augmentation du nombre de dimensions ou de la taille de la base de données, mais proportionnel à la taille des réponses retournées.

### **\* Architecture Client/Serveur**

Il est essentiel que le produit soit Client/Serveur mais aussi que les composants serveurs d'un produit OLAP intègrent facilement ses différents clients.

**\* Dimensions Génériques**

Chaque dimension doit être équivalente par rapport à sa structure et ses capacités opérationnelles.

**\* Gestion dynamique des matrices creuses**

Le système OLAP ajuste automatiquement son schéma physique pour s'adapter au type du modèle et au volume des données (plus on dispose de places plus on peut agréger).

**\* Support multiutilisateurs**

Les outils OLAP doivent fournir des accès concurrents, l'intégrité et la sécurité.

**\* Calculs à travers les dimensions**

Les calculs doivent être possibles à travers toutes les dimensions (les agrégats doivent être faits dans toutes les dimensions).

**\* Manipulation intuitive des données**

La manipulation des données se fait directement à travers les cellules d'une feuille de calcul, sans recourir au menu ou aux actions multiples. Il doit permettre l'analyse intuitive dans plusieurs dimensions au final.

**\* Souplesse et facilité de constitution des rapports**

Lors de la création de rapports, les dimensions peuvent être présentées de n'importe quelle manière.

**\* Nombre illimité de niveaux d'agrégation et de dimension**

Dimensions et nombre d'agrégations illimitées

### **7.3. Systèmes OLTP versus systèmes OLAP**

Le terme OLTP On Line Transactional Processing regroupe les concepts mis en place par un système destiné à l'automatisation d'un processus. On utilise pour cela une base de données transactionnelle, dont la mise à jour est faite « en ligne ».

## Partie 2 : Présentation Général du Data Warehouse

Le terme OLAP On Line Analytical Processing regroupe lui les concepts mis en place par un système d'analyse en ligne de données, on utilise pour cela une base de données décisionnelle, destinée à la prise de décision. OLTP est plus ancien qu'OLAP et le traitement des transactions en mode transactionnel (OLTP) est totalement différent du domaine des systèmes dimensionnels (OLAP).

Les deux systèmes diffèrent pour :

- Les utilisateurs.
- Le contenu des données.
- La structure de la base de données.
- L'administration du système.

	Bases de production(OLTP)	Entrepôt de données (DW)
Données	<ul style="list-style-type: none"><li>• Atomique</li><li>• Orienté application</li><li>• à jour</li><li>• Dynamique</li></ul>	<ul style="list-style-type: none"><li>• Résumés</li><li>• Orienté sujet</li><li>• Historique</li><li>• Statique</li></ul>
Utilisateurs	<ul style="list-style-type: none"><li>• Employés de bureau</li><li>• Nombreux</li><li>• Concurrents</li><li>• Mises à jour</li><li>• Accès a peu de données</li></ul>	<ul style="list-style-type: none"><li>• Analystes</li><li>• Peu</li><li>• Non concurrents</li><li>• Interrogation</li><li>• Beaucoup d'information</li></ul>

**Tab -3- OLTP vs OLAP**

## 8. Les composants fonctionnels d'un data warehouse

Trois composants caractérisent un data warehouse : L'acquisition (alimentation), le stockage et l'accès.

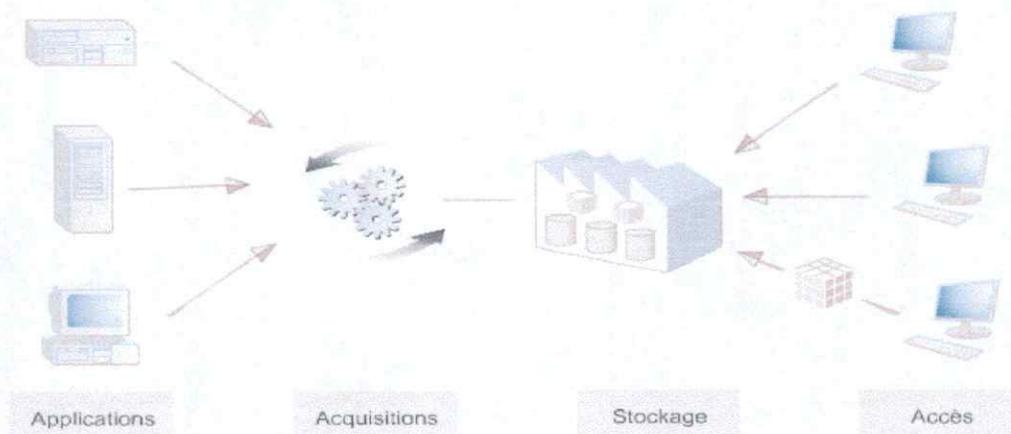


Fig-13- Composants fonctionnels d'un data warehouse

### 8.1. Alimentation d'un data warehouse

L'alimentation des données à partir des bases de production est une phase primordiale d'un data warehouse, des outils logiciels sont alors nécessaires pour intégrer ces données. On parle d'ETL (Extract, Transform, Load). Les phases de l'alimentation d'un data warehouse sont les suivantes :

#### a) Découverte des données

Il s'agit d'identifier dans les systèmes sources les données à importer dans le data warehouse, il faut prendre les données les plus judicieuses sachant qu'un mauvais choix peut considérablement compliquer les phases suivantes de l'alimentation.

#### b) Extraction des données

Il s'agit de collecter les données utiles dans les systèmes de production. Il faut identifier les données ayant été modifiées afin d'importer le minimum de données dans le data warehouse.

### c) Transformation des données

Il faut rendre les données cohérentes avec la structure du data warehouse. On applique alors des filtres sur les données. Il peut être nécessaire de convertir le format des données (EBCDIC vers ASCII par exemple) ou d'harmoniser les formats de dates (jj/mm/aaaa).

Il faut également associer les champs source avec les champs cibles. Un champ source « adresse » pourra ainsi par exemple être décomposé en « numéro », « rue », « code postal », « ville » ou l'inverse.

Enfin des données des systèmes de production doivent être agrégées ou calculées avant leur chargement.

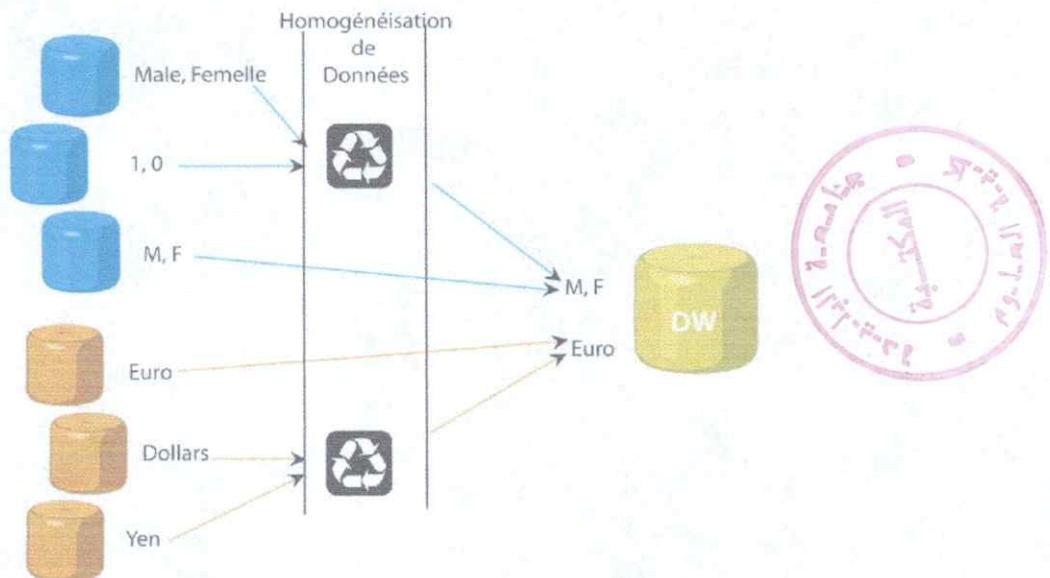


Fig -14- Homogénéisation de données

### d) Chargement des données Homogénéisation

C'est la dernière étape d'alimentation, il s'agit d'insérer les données au sein du data warehouse. C'est une phase délicate car les quantités de données sont souvent très importantes.

## 8.2. Le processus de stockage

Les bases multidimensionnelles sont conçues pour optimiser l'analyse d'indicateurs par rapport aux dimensions qui leur sont associées (temps, produit, région). Elles sont de moins en moins exigeantes en termes de volume de stockage (technique de compression). Les bases de données relationnelles sont conçues à l'origine pour le décisionnel. Elles sont plus ouvertes et s'adaptent à de nombreuses plates formes.

Le composant de base est le SGBD, il doit être spécialement adapté aux caractéristiques de l'accès décisionnel. Du fait de l'importance de l'historique, la structuration physique des données est également très importante.

Le SGBD apporte la transparence à l'évolution matérielle, l'indépendance, que ce soit au niveau des types et du nombre de processeurs, des disques ou des mémoires. Ainsi que la transparence à l'évolution des systèmes d'exploitation

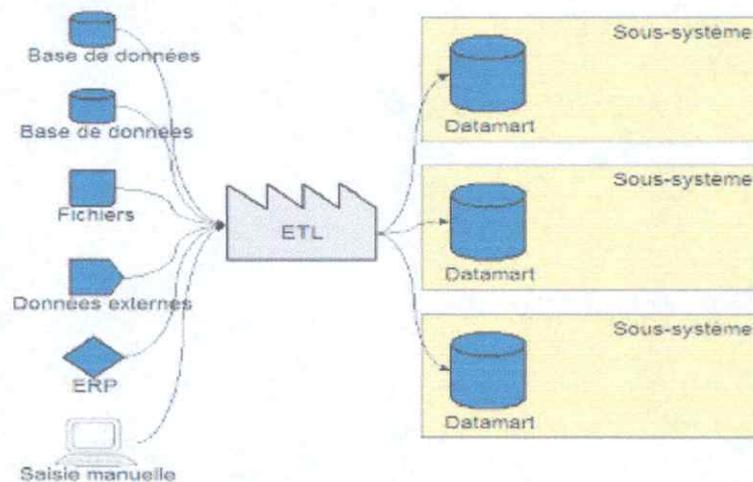


Fig -15- Le stockage

Le stockage au sein d'un data warehouse a un besoin de synthèse (agrégation des données) et un besoin de détails (conservation des données détaillées). Ce stockage peut être réalisé de trois manières différentes :

❖ **Structure directe simple**

On fait des mises à jour du data warehouse avec des laps de temps importants.

❖ **Structure de cumul simple**

On stocke les données de chaque mise à jour, les mises à jour étant fréquentes (par exemple tous les jours). On a un espace occupé important, mais on ne perd pas d'info.

❖ **Structure par résumer déroulant**

A chaque mise à jour, on stocke des données détaillées, et on synthétise les anciennes données en fonction de leur âge. Plus une donnée est vieille, moins elle est détaillée.

### **8.3. Le processus d'accès**

Définir une architecture globale servant de support aux accès décisionnels impose des choix technologiques non structurants. Il faudra mettre en place une infrastructure commune à toutes les applications décisionnelles, tout en laissant aux utilisateurs l'opportunité d'utiliser des solutions d'accès les mieux adaptées à leur problématique.

Les outils du marché sont : les requêteurs, les outils d'analyse multidimensionnelles, les EIS, le data mining, les SIG.

## **9. Construction de data warehouse**

### **9.1. Modélisation des données de l'entrepôt**

#### **9.1.1. Modélisation entité relation**

La modélisation entité relation vise à éliminer toute redondance dans les données, elle consiste à diviser les données en de nombreuses tables afin de ne retrouver les informations qu'une seule fois. Ceci apporte de nombreux avantages au niveau du traitement des transactions, qui deviennent simples et déterministes. Plus la transaction est « atomique », moins il existe d'entités concernées par cette transaction et plus l'incohérence de la base de

données est réduite. Ces diagrammes comportent souvent de nombreuses tables et jointures. La lecture d'un diagramme est souvent ardue, aucune table ne surplombe une autre.

Dans le cadre de l'entrepôt de données, la modélisation entité relation n'est pas appropriée, il est inconcevable de construire un système décisionnel avec un tel modèle.

Il faut éviter le modèle entité/relation dans les systèmes décisionnels, car l'utilisateur final ne peut ni lire ni intégrer un tel modèle (naviguer dans les données), et l'adoption d'un tel modèle pour un système décisionnel aura des conséquences catastrophiques au niveau des performances. En plus, ce modèle n'est pas interrogeable par voie logicielle.

Groupe professionnel	Tranche d'age	Effectifs
Cadres dirigeants	35-40	0
Cadres supérieurs	35-40	4
Cadres	35-40	66
Maitrise	45-50	699
Exécution	45-50	353

**Tab -4-** Représentation des données dans une table relationnelle

### **9.1.2. Modélisation dimensionnelle** [R.Kim,M .Ross]

La modélisation consiste à transformer les résultats de l'analyse des besoins en un modèle dimensionnel. Cette tâche apparaît également dans la mise en place d'un système OLTP, mais la modélisation dimensionnelle diffère fortement de la modélisation entité relation, le modèle dimensionnel est lui conçu pour être interrogé de manière particulièrement efficace.

Un data warehouse est une base dédiée au décisionnel. L'information est mise à la disposition des utilisateurs mais les mises à jour ne sont jamais faites en ligne. Les seules

mises à jour étant les phases de chargement provenant des données de production et se passant en mode batch. Une fois les nouvelles données chargées, l'intégrité de l'entrepôt ne pourra plus être remise en cause. Il devient donc envisageable d'introduire des redondances. Beaucoup de concepts liés à la modélisation de données sont alors moins cruciaux, et l'on peut donc contourner les règles d'or de la modélisation traditionnelle.

Les requêtes effectuent fréquemment dans les sélections des restrictions, des regroupements, des calculs ou autres agrégations. Pour répondre à ces besoins, il faudra très souvent manipuler des volumes importants même si le résultat final n'est constitué que de quelques lignes. Dans ce contexte, il devient très difficile d'obtenir des temps de réponse proportionnels au volume du résultat.

En fonction des éléments cités ci-dessus, il devient primordial de définir le modèle dimensionnel alliant la lisibilité de l'utilisateur à des techniques d'optimisation des requêtes.

### **a. Caractéristiques du modèle dimensionnel**

#### **\* Lisibilité pour l'utilisateur final**

L'utilisateur final (analyste, spécialiste métier, ...) doit pouvoir facilement manipuler les données. Les données doivent être simples à comprendre mais permettre plusieurs analyses différentes selon des critères différents. Plus le modèle est lisible (intuitif) pour les utilisateurs, moins sera long (coûteux) de le définir pour le rendre compréhensible.

#### **\* Performance au chargement des données**

Les entrepôts de données sont constitués de grosses quantités de données. Les données sont parfois de sources hétérogènes. Il est assez simple de les charger dans une structure simple.

Le chargement nécessite souvent une transformation / filtrage des données sources.

#### **\* Performance à l'exécution des requêtes**

L'exécution des requêtes doit être rapide pour les analystes. On estime généralement 30 secondes pour les analyses en ligne (et pas pour le reporting, ou l'imprimante est de toute

façon très lente). Il est parfois nécessaire de créer des agrégations de données et/ou des redondances. Cette création de redondance est antagoniste avec les performances de chargement. Il complique également l'administration du système.

**\* L'administration de données**

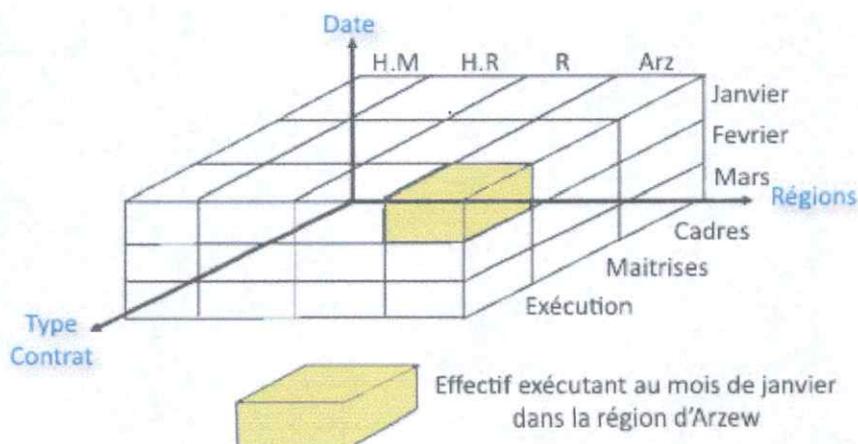
L'administration du système est fondamentale. La difficulté n'est pas tant de construire l'entrepôt mais de le faire vivre : tracer des requêtes, modification/ adaptation des objets, maîtrise et industrialisation de l'extraction des données. Les performances d'un système mal géré décroissent de façon exponentielle !

**\* L'évolutivité du modèle**

Le développement de l'entrepôt est plus incrémental qu'itératif. Chaque projet ne doit pas déboucher sur un modèle d'information isolé des autres.

	Cadre	Maîtrise	Exécution
Permanents	235	161	13
Temporaires	84	68	67

**Tab -5-** Représentation des données dans un tableau multidimensionnel



**Fig -16-** Représentation des données sous forme d'un cube (multidimensionnelle)

### 9.1.2.1. Modélisation en Etoile

Dans le datawarehouse, le modèle de données « en étoile » est typique des structures multidimensionnelles stockant des données agrégées. Le modèle en étoile est optimisé pour les requêtes d'analyse et souvent considéré (à tort) comme un modèle dénormalisé, il est implémenté sur un SGBD relationnel classique (tel Oracle ou SQL server).

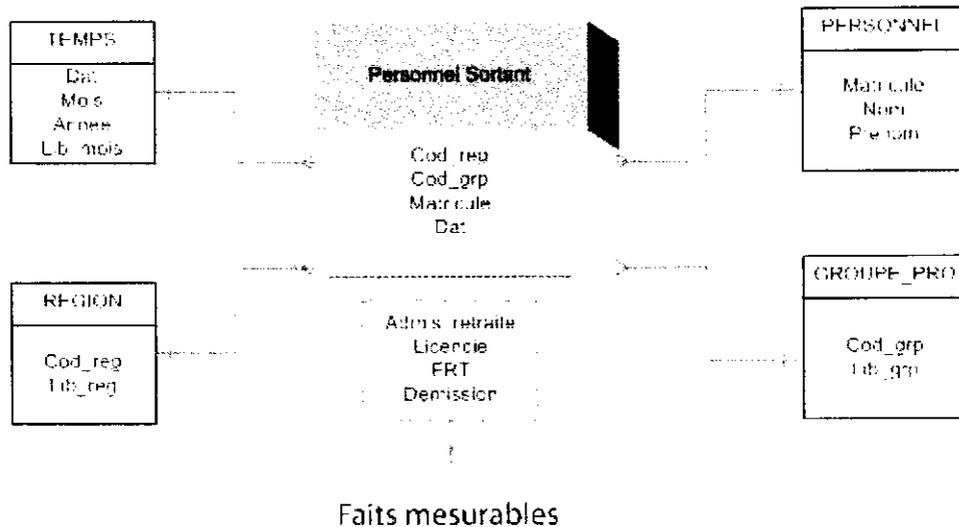


Fig -17- Modélisation en étoile

\* **Avantages** : Facilité de navigation, performance car le nombre de jointures est limité, gestion des données creuses, gestion des agrégats, fiabilité des résultats.

\* **Inconvénients** : Toutes les dimensions ne concernent pas les mesures, redondances dans les dimensions, alimentation complexe.

• **Table de faits**

Une table de faits est une table qui contient les données à analyser. Une table de faits est souvent reconnaissable par sa taille ; en effet, lorsqu'on visualise un schéma, c'est celle qui est au centre et qui est la plus grande. Ce type de table est aussi facilement reconnaissable car elle comporte un grand nombre de clés étrangères afin de lier avec des tables de dimension.

- **Grain de la table de faits**

L'unité de temps la plus petite est appelée le grain/finesse de la table de faits. Dans notre exemple, le grain est journalier. Pour connaître celui ci, il suffit de consulter la table de dimension temps et de regarder la plus petite valeur (jour\_semaine). Il s'agit donc de la plus petite unité de temps. La détermination du grain est un problème épineux! si le grain est petit, la table de faits est gigantesque et si le grain est trop gros, on ne peut pas revenir en arrière et on perd de l'information. Plus le grain est petit plus la conception est robuste.

- **Les faits**

La table de faits peut aussi contenir des champs qui ne sont pas des clés étrangères. Ce sont les faits. Les faits doivent être valorisés de façon continue et être additifs (chiffre d'affaire, quantité vendue). Si on a des valeurs discrètes, cela devrait plutôt être une dimension, donc vérifier qu'on n'a pas fait d'erreurs. Il peut exister des faits semi additifs et non additifs. Les faits semi additifs peuvent être additionnés pour certaines dimensions seulement (cas se présentant avec une agrégation : Nombre de contacts clients). Les faits non additifs ne peuvent pas être ajoutés (uniquement des comptages ou des impressions : Encours moyen de fin du mois).

- **Table de faits sans faits**

Chaque table de faits comporte plusieurs champs de clés suivis de champs de faits numériques, valorisés de façon continue et additive. Il existe dans certains cas des tables de faits dans lesquelles il n'y a pas de faits mesurés ! ces tables sont appelées tables de faits sans faits (que des comptages).

- **Table de dimensions**

Dans un schéma en étoile, les tables qui entourent la table de faits sont appelées tables de dimensions. Ces tables sont composées d'attributs qui sont souvent de type caractère et discret. Ces attributs servent à stocker la description des dimensions et sont utilisés comme source de contraintes et d'en-têtes de ligne dans le jeu de réponses de l'utilisateur.

• **Les mesures**

Une mesure est un élément de données sur lequel portent les analyses, en fonction des différentes dimensions, exemple : nombre d'accidents, ventes.

• **Attribut ou fait**

Il est important de bien comprendre la différence entre un attribut et un fait afin de placer ceux-ci dans les bonnes tables (dimensions ou faits).

	Attributs	Faits
Attribut de type numérique		X
Attribut de type caractère	X	
Valeur discrète	X	
Variation continue		X
Table de faits		X
Table de dimension	X	

**Tab -6-** Différence entre attribut et fait

• **La dimension temps**

La dimension temps figure systématiquement dans le modèle relationnel. L'entrepôt est une série temporelle. Conceptuellement, nous n'avons pas besoin d'une table de dimension temps, les hiérarchies peuvent être obtenues avec des fonctions à partir de date/heure. Pour des questions de performance, il est préférable de créer une table de dimension ; avec un grain journalier, 10 ans d'exploitation représentent environ 3653 tuples, ce qui est négligeable par rapport aux performances apportées.

• Dimensions et hiérarchies

Une dimension est un ensemble de valeurs décomposable. Les valeurs d'une dimension sont généralement organisées à l'intérieur d'une hiérarchie (jour, mois, semestre, année) :

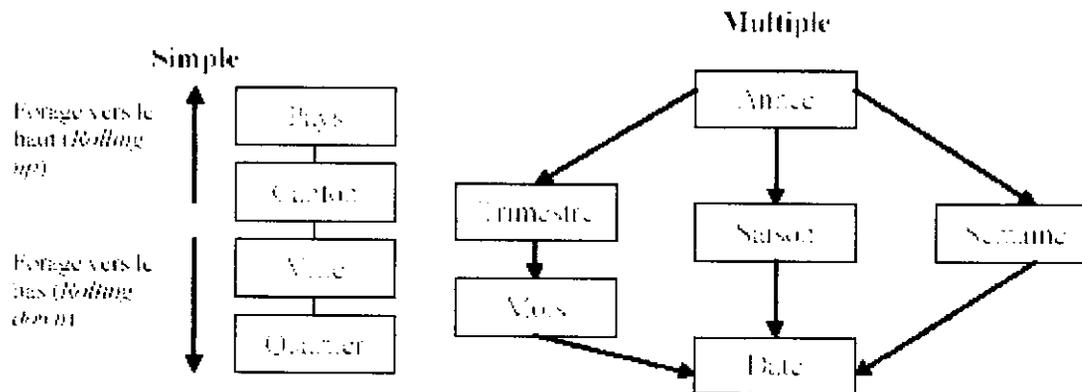


Fig -18- Modèles d'hiérarchies

Dans le cas d'une hiérarchie simple, chaque enfant n'a qu'un seul père.

Dans le cas d'une hiérarchie multiple, on a différents chemins pour créer une hiérarchie.

### 9.1.2.2. Modélisation en Flocon

Ce modèle est plus propre (3<sup>ème</sup> forme normale), la hiérarchie multiple est simplifiée et il offre un gain de place de stockage (pas vraiment un avantage, souvent négligeable sur les dimensions).

En revanche, ce modèle est plus complexe pour l'utilisateur final. Il implique aussi plus de jointures, donc il est plus lent.

Selon les spécialistes, il faut éviter le « floconnage » des dimensions, même si elles sont grandes, car les performances seront mauvaises ! On peut très bien faire de l'analyse multidimensionnelle sur un fichier plat !

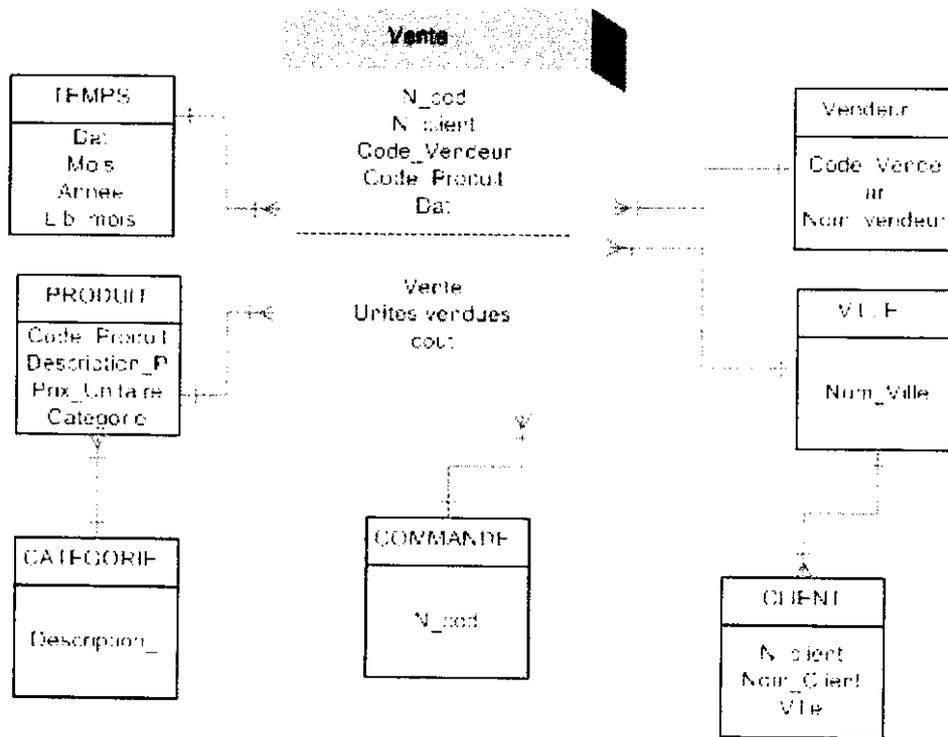


Fig -19- Modélisation en flocon

\* **Avantages** : réduction du volume, gain de place de stockage, plus propre, permet des analyses par pallier (drill down) sur la dimension hiérarchisée.

\* **Inconvénients** : navigation difficile, plus complexe pour l'utilisateur final, nombreuses jointures donc plus lent.

### 9.1.2.3. Modélisation en constellation

Une autre technique de modélisation, issue du modèle en étoile, est la modélisation 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.

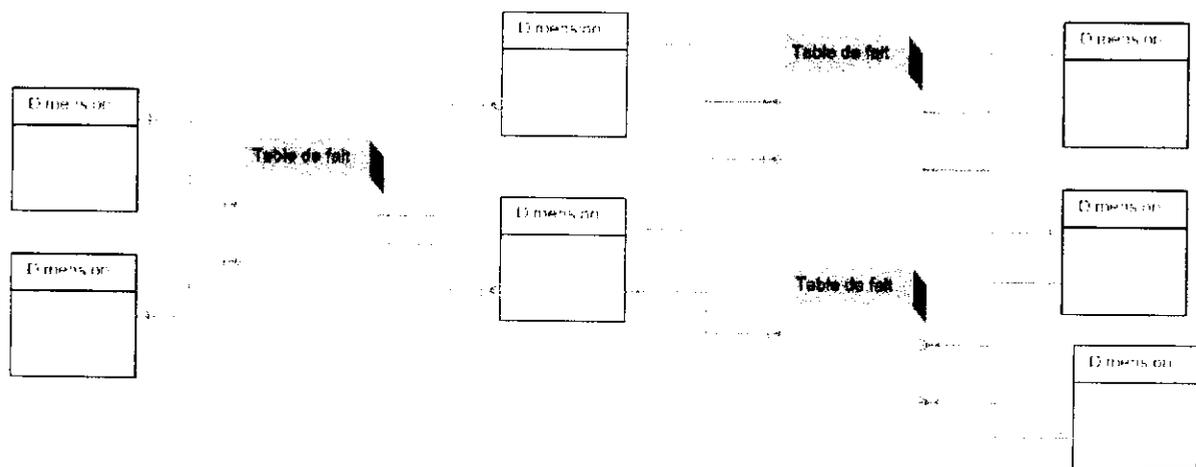


Fig -20- Modélisation en constellation

\* **Avantages** : réduction du volume, gain de place de stockage, couplage de plusieurs tables de faits (analyses) dans un seul schéma.

\* **Inconvénients** : conception et navigation difficiles, pareil que pour le modèle en flocon les nombreuses jointures le rendent plus lent et gourmand en ressources, risques d'incohérence.

## 10. Structure du data warehouse

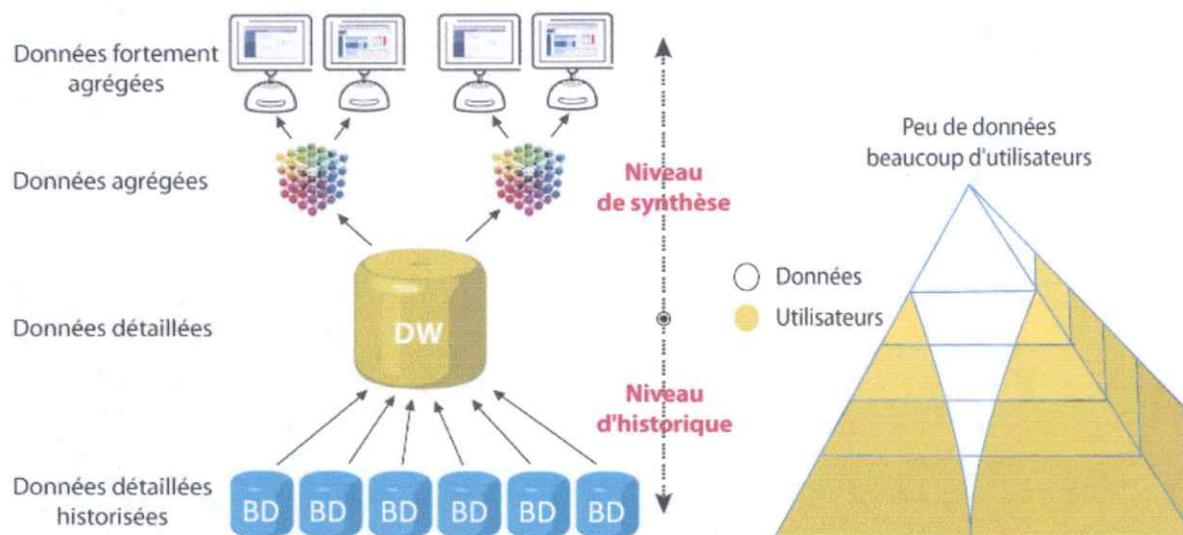


Fig -21- Structure de data warehouse logique d'accès aux données

Un DW se structure en quatre classes de données, organisées selon un axe historique et un axe synthétique.

➤ Les données détaillées

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

➤ 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.

➤ Les métadonnées

Très souvent, les données à fédérer dans le data warehouse proviennent des sources très hétérogènes. Cela rend indispensable la présence d'un dictionnaire unique qui sait gérer l'ensemble des fonctions de l'entrepôt. Cette cohérence du dictionnaire est décrite au sein des métadonnées du dictionnaire du data warehouse. Elles constituent l'ensemble des

données qui décrivent des règles ou des processus attachés à d'autres données : répertoire, règles d'extraction et de conversion, règles d'agrégation.

➤ Les données historisées

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

Les supports les plus couramment utilisés sont les disques, les disques optiques numériques, les cassettes.

## **10.1. Processus de conception pour une base de données dimensionnelle [Kim 97]**

### **10.1.1. Choisir les processus d'activité opérationnelle à modéliser**

La première étape de la conception consiste à choisir les processus d'activité à modéliser, sur la base d'une connaissance de l'activité et des sources de données disponibles. Exemple : les commandes, la facturation, les stocks, les ventes, etc.

La modélisation doit transformer l'expression des besoins en un schéma dimensionnel.

### **10.1.2. Définition du grain**

La seconde étape de la conception est le choix du grain de la table des faits pour chaque processus d'activité. Un entrepôt de données exige presque toujours des informations exprimées dans le grain le plus bas possible pour chaque dimension.

Non parce que les requêtes portent sur les enregistrements individuels mais parce que les requêtes doivent faire des coupes dans la base de données selon des critères bien précis.

### **10.1.3. Modification du grain**

L'énoncé précis du grain détermine les dimensions principales de la table des faits. Il est possible d'ajouter ensuite des dimensions supplémentaires, à condition que celles ci ne puissent prendre qu'une valeur pour chaque combinaison des dimensions élémentaires.

S'il on veut une dimension supplémentaire qui détruit le grain choisi initialement en entraînant la génération d'enregistrements supplémentaires, il faut alors revoir la définition du grain pour reprendre en compte la dimension supplémentaire.

### **10.1.4. Normalisation**

La table de faits atteint naturellement un haut degré de normalisation, tout effort de normalisation des tables d'une base de données dimensionnelle afin d'économiser l'espace disque et une perte de temps. Les tables de dimension ne doivent pas être normalisées mais rester des tables plates : des tables de dimension normalisées ne sont plus navigables, les gains d'espace disques sont de l'ordre de 1% par rapport à l'ensemble du schéma.

## **11. BD multidimensionnelles**

L'analyse multidimensionnelle est la capacité à analyser des données qui ont été agrégées suivant plusieurs dimensions. On veut donc accéder à des données déjà agrégées selon les besoins de l'utilisateur, de façon simple est rapide, on utilise pour cela des hypercubes OLAP :

- Les données sont représentées dans des hypercubes à n dimensions.
- Les données sont structurées suivant plusieurs axes d'analyses (dimensions) comme le temps, la localisation ...
- Une cellule est l'intersection des différentes dimensions.
- Le calcul de chaque cellule est réalisé au chargement. Le temps de réponse est ainsi stable quelque soit la requête.

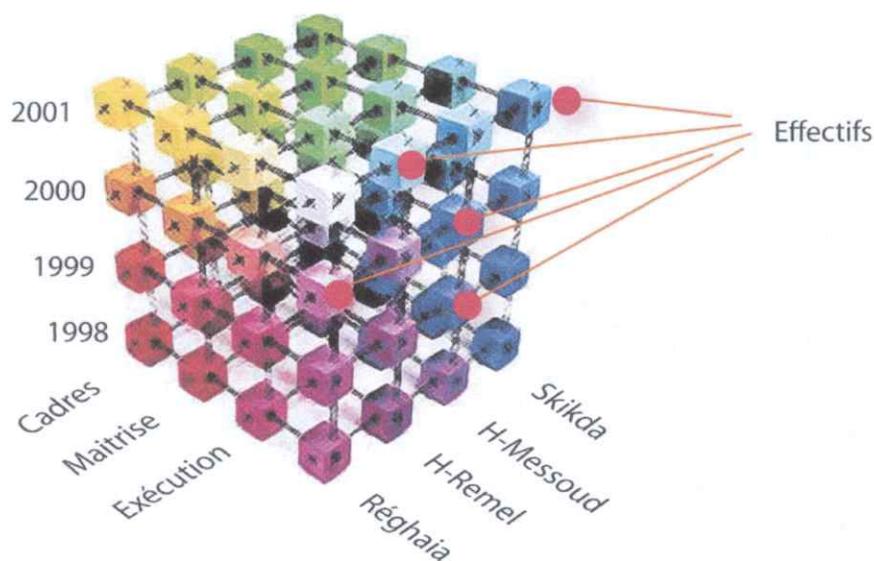


Fig -22- Cube OLAP

## 12. Outils de Navigation

### 12.1. OLAP Tools

#### 12.1.1. Drill up : Agréger selon une dimension

Ou forage vers le haut : consiste à représenter les données du cube à un niveau de granularité supérieur conformément à la hiérarchie définie sur la dimension. Utilisation de la fonction d'agrégation (somme, moyenne, etc.) spécifiée pour la mesure et la dimension.

#### 12.1.2. Drill down : Détailler selon une dimension

Ou forage vers le bas : consiste à représenter les données du cube à un niveau de granularité de niveau inférieur, donc sous une forme plus détaillée.

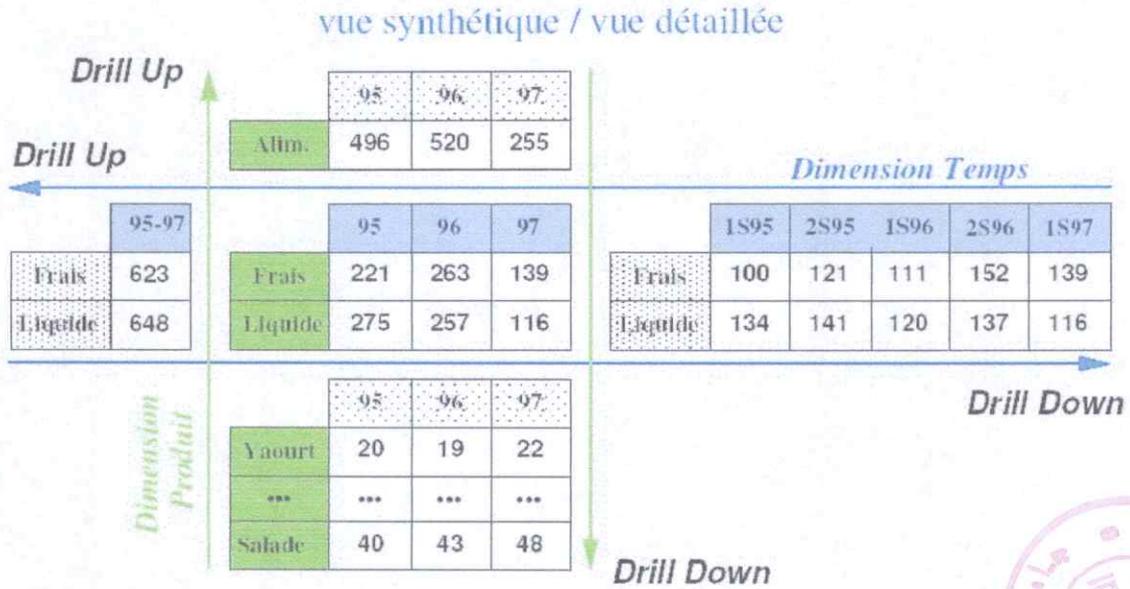


Fig -23- Drill Up, Drill Down



### 12.1.3. Slicing : Sélection selon un axe

Sélection de tranches du cube par des prédicats selon une dimension. Filtrer une dimension selon une valeur, exemple : Slice (2004) ; on ne retient que la partie du cube qui correspond à cette date.

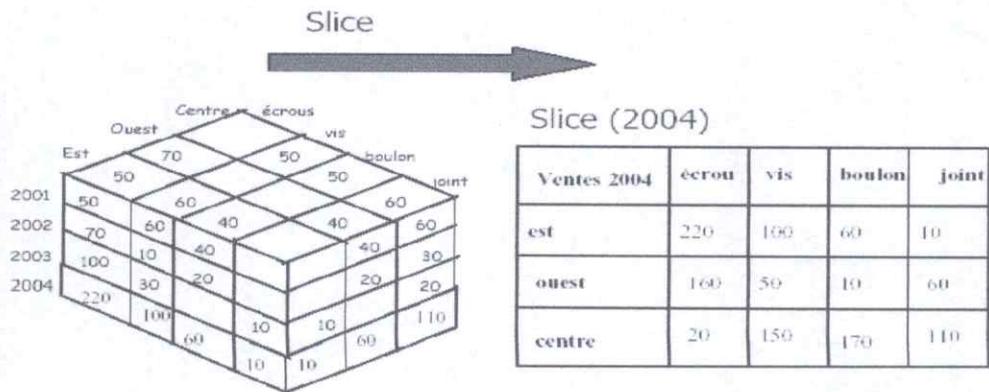


Fig -24- Exemple de Slicing

#### 12.1.4. Dicing : Projection selon un axe

Extraction d'un sous-cube.

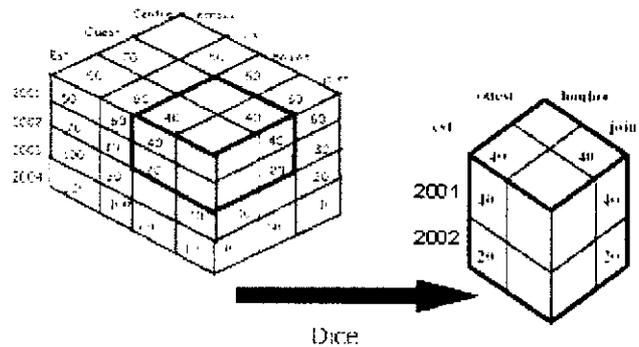


Fig -25- Exemple de Dicing

#### 12.1.5. 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. Ces mécanismes s'appliquent sur le modèle défini soit par l'administrateur, soit par l'utilisateur. Ce modèle doit être assez complexe pour adresser l'ensemble des demandes de l'utilisateur 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.

#### 12.1.6. Pivoting (Rotate) : Tourner le cube pour visualiser une face.

Possibilité de modifier l'aspect d'un rapport en déplaçant un champ (ou un groupe de champs) de ligne en colonne ou inversement. On peut également ajouter des champs en les sélectionnant dans une liste de choix. (Région, Produit) → (Région, Mois).

## 12.2. Data marts

On y stockera des données portant sur une des activités de l'entreprise. Ces mini data warehouses peuvent alors être considérés comme des espaces d'analyse, du fait que les données sont bien moins nombreuses et surtout qu'elles sont thématiques. Ils peuvent également servir de bases de construction à des cubes de données.

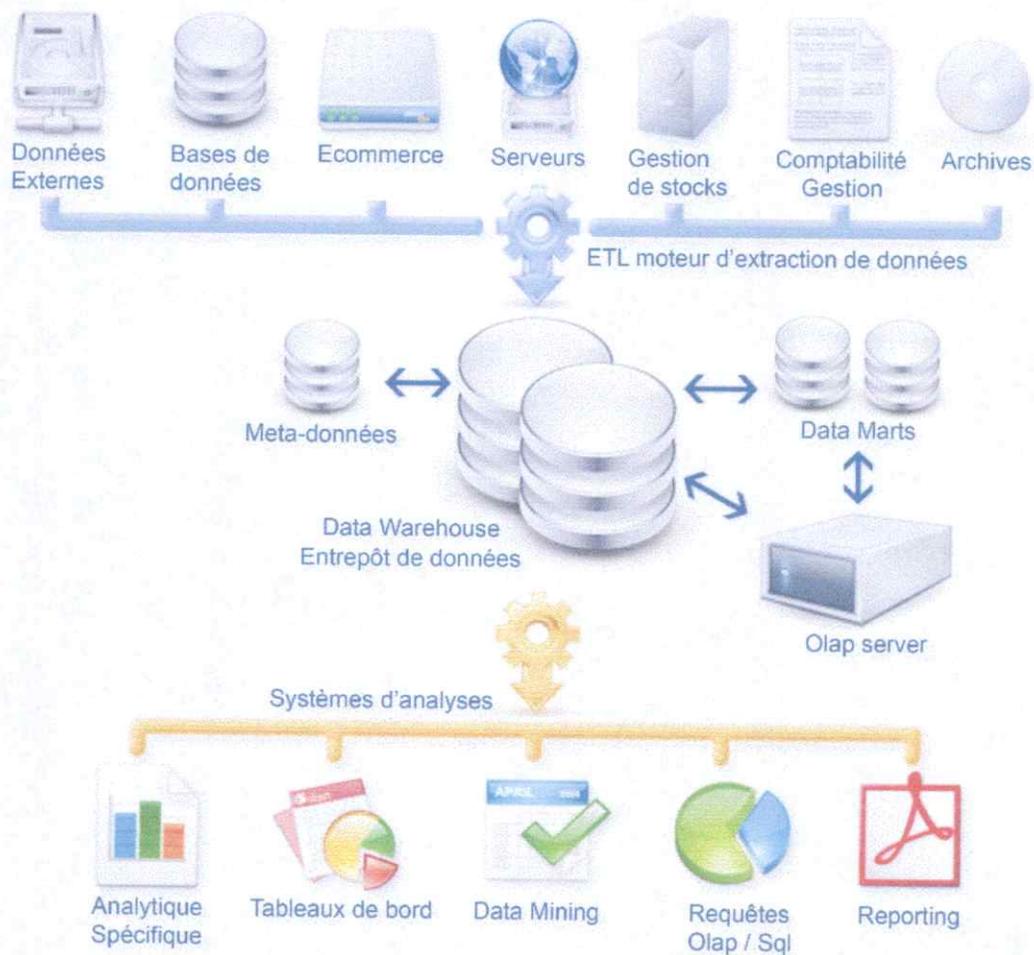


Fig -26- Data Mart

## 13. Conclusion

La gestion des gros volumes de données dans un contexte décisionnel nécessite des solutions appropriées depuis le recueil des données jusqu'à l'exploitation de certaines données par des algorithmes d'extraction automatique de connaissance.

Cet état de l'art introduit l'ensemble des entités impliquées dans ce type de solutions. Ces entités sont les entrepôts de données et les magasins de données, reposant sur le modèle multidimensionnel, et les outils d'analyse de type OLAP et OLAP mining.

L'utilisation des bases de données multidimensionnelles pour la fouille de données (OLAP Mining) suppose une collaboration entre le système de gestion de bases de données et les systèmes de découverte de connaissance, ce qui constitue un point très intéressant de l'approche décrite ici.

Dans le chapitre suivant, nous allons présenter notre démarche pour faire la conception de l'entrepôt de données. Cette démarche est une collection d'étapes successives commençant par l'extraction des besoins et finissant par la mise en œuvre et le déploiement. Mais avant cela, nous allons faire une petite présentation de l'entreprise GTP.



Partie .3



# Présentation de GTP

## **Introduction**

L'Entreprise Nationale de Grands Travaux Pétroliers « ENGTP » est une Société par Actions créée depuis le 19 février 1989, ses actions étaient détenues d'une part par la société mère Sonatrach à 51% et d'autre part par la Société de Gestion de participation à 49% jusqu'au 13 Décembre 2005 où toutes ses actions sont détenues à 100 % par Sonatrach.

Elle est dotée d'un Conseil d'Administration de 07 membres et d'une équipe de Cadres Dirigeants menée par un Président Directeur Général.

Son capital social est actuellement de 2.790.000.000 DA. Son Objet social est l'étude et la réalisation des projets d'installation matérielle notamment dans les domaines des hydrocarbures, de l'hydraulique, de l'énergie, de l'agroalimentaire, des matériaux de construction et des industries s'y rapportant, à l'intérieur du territoire et à l'étranger. La maintenance d'installations industrielles en exploitation. La formation dans ce domaine de soudage, contrôle et activités annexes.

Son siège social est fixé à : BP 38, zone industrielle de Réghaia, Alger. Elle est implantée sur le territoire national d'est en ouest et du nord au sud, notamment à Skikda, Arzew, Hassi R'mel, Hassi Messaoud et In Amenas où elle est représentée par des directions régionales.

Afin de bien comprendre le problème de notre étude, nous allons présenter l'entreprise pour mieux la connaître.

### **1. Missions de « GTP »**

- ✓ Etudes et la réalisation des projets d'installations matérielles dans le domaine des hydrocarbures et des industries s'y rapportant.
- ✓ Etudes générales industrielles, résolution des problèmes technico-économiques et expertise.
- ✓ Contrôle et réception de tous matériaux, matériels et installations industrielles.
- ✓ La réalisation des grands ensembles industriels dans le domaine des Hydrocarbures et des industries se rapportant à son objet.
- ✓ La formation des hommes.
- ✓ Réalisation de réseaux de collectes et désertes d'Hydrocarbures liquides et gazeux et des installations de surfaces y afférents.

Ses différents domaines d'intervention sont :

### Partie 3 : Présentation de GTP

- ✓ Secteur Hydrocarbures.
- ✓ Secteur Chimie et Pétrochimie.
- ✓ Secteur Energie Electrique.
- ✓ Secteur Agroalimentaire.
- ✓ Secteur Matériaux de Construction.

## **2. Activités**

- Etudes, procurement, supervision, essais et mise en service d'installations industrielles.
- Génie-civil et bâtiments industriels (travaux et terrassements généraux de fondations des équipements, de VRD...).
- Génie-mécanique (préfabrication et montage de structure métallique de tuyauteries d'équipements statiques ou dynamiques...).
- Génie-électrique (installation d'équipements électriques et d'instrumentations, étalonnage et montage des instruments ...).
- Canalisation (topographique, bardage de tubes, soudage et contrôle, testes hydrostatiques et protection cathodiques).
- Soudage (procèdes SMA W, GTA W, GMA W, FCA W, SA W).
- Contrôle (non destructif par rayons x, rayons Gamma et destructif par compression, traction, pliage, résilience et macrographie).
- Traitement thermique.
- Maintenance industrielle (intervention préventive par la mise à disposition d'équipes d'entretien ou curative par la remise en état d'équipements ou d'organes d'une usine en exploitation).

### 3. Organisation de l'entreprise

#### 3.1. Organigramme général de l'ENGTP

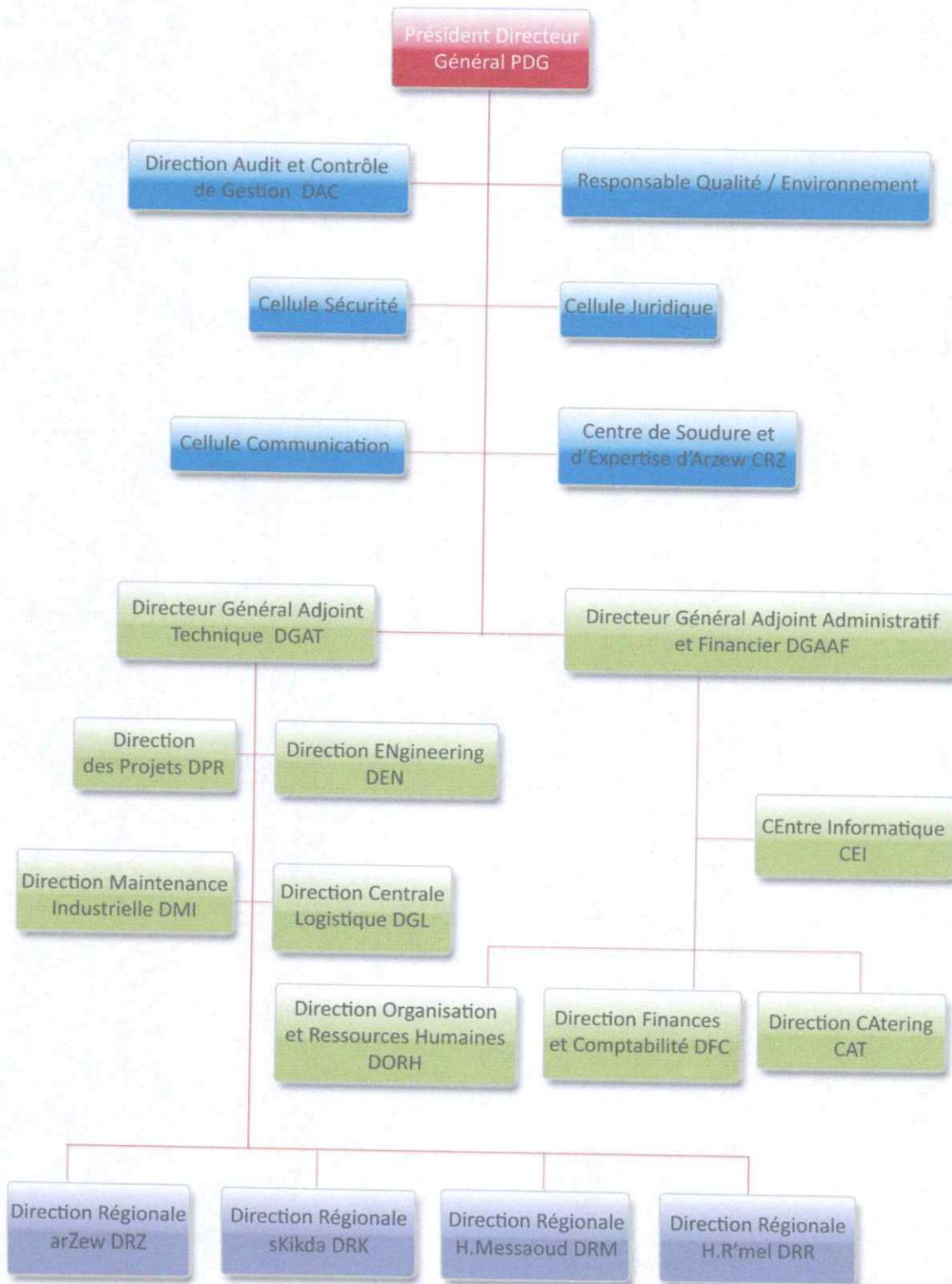


Fig -27- Organigramme général de GTP

### Partie 3 : Présentation de GTP

Cette structure organisationnelle adoptée depuis juillet 2003 est de type « staff and line » ou hiérarchico-fonctionnelle. Théoriquement, elle vise à renforcer la compétence des chefs hiérarchiques sur le plan de la stratégie en leur adjoignant des conseillers spécialisés. Sur le plan pratique, comme il a pu être constaté, il naît des relations tendues entre staff et line évacuant les avantages qu'une telle structure pouvait donner.

## **3.2. Organisation**

La structure d'organisation de l'ENGTP/Spa est fixée, par décision du Président Directeur Général (PDG/A.001.R12 du 12 mars 2003), comme suit :

Le staff du PDG est chargé de l'élaboration des politiques et stratégies de développement de l'entreprise et l'élaboration de leur mise en œuvre.

Le Président Directeur Général.

Le staff du PDG est composé des organes suivants :

- La Direction Audit et Contrôle de gestion (DAC)
- Le Responsable Qualité.
- La cellule sécurité (HSE et SIE).
- La cellule juridique.
- La cellule communication.
- Le Centre de Soudure et d'expertise d'Arzew (CSZ).

Les structures techniques ou opérationnelles classées sous l'autorité d'un directeur général adjoint technique, chargées des activités d'engineering, de logistique, de maintenance industrielle et de la réalisation des projets de l'entreprise.

Le Directeur Général Adjoint Technique (DGAT) et les organes dépendants, à savoir :

- La Directions des Projets (DPF).
- La Direction Engineering (DEN).
- La Direction Maintenance Industrielle (DMI).
- La Direction Centrale logistique (DCL).
- Les Directions Régionales de H.Messaoud (DRM), de H.R'mel (DRR), d'Arzew (DRZ) et de Skikda (DRK).

### Partie 3 : Présentation de GTP

Les structures fonctionnelles placées sous l'autorité d'un directeur général adjoint administratif et financier, chargé de l'administration et de la gestion des ressources humaines et financière de l'entreprise.

-Le Directeur Général Adjoint Administratif et Financier (DGAAF) et les organes dépendants, à savoir :

- La Direction Organisation et Ressources Humaines (DORH)
- La Direction Finances et Comptabilité (DFC)
- La Direction Catering (CAT)
- Le Centre Informatique (CEI).

### 3.3. Organigramme du centre informatique

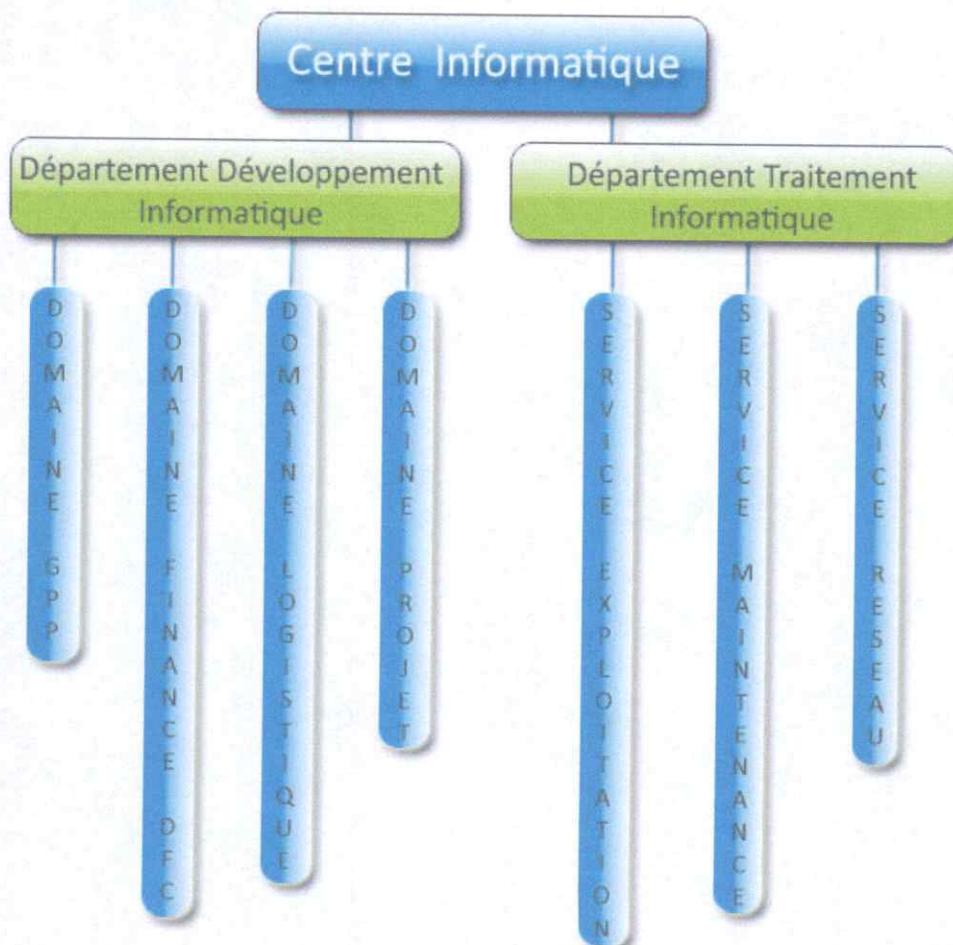


Fig -28- Organigramme du centre informatique

### Partie 3 : Présentation de GTP

DGP : Domaine Gestion Personnel Paie communique avec DORH (Direction organisation des ressources humaines).

DFC : Domaine de Finances et Comptabilité communique avec DFC (Direction finances et comptabilité).

DL : Domaine (matériel, outillage, maintenance, stock) communique avec DCL (Direction centrale logistique).

DOP : DOmaine Projet communique avec DPR (Direction des projets).

Les différentes missions accomplies par le centre informatique sont :

- Participer à l'élaboration du plan de développement informatique notamment en ce qui concerne les prévisions des ressources matérielles nécessaires à l'atteinte des objectifs.
- Gérer et veiller à la maintenance des équipements informatiques de l'entreprise, on peut citer quelques investissements : unité centrale, les écrans, les imprimantes, les scanners, les onduleurs.

## **4. Environnement de l'entreprise**

L'entreprise ENGTP, comme toutes les entreprises algériennes, est contrainte de s'adapter à un environnement en pleine mutation qui se caractérise :

✓ Au plan interne, par :

La fin du monopole qui liait jusque la l'entreprise ENGTP à la société pétrolière Sonatrach sur le plan des projets de construction d'ouvrages industriels : l'approche du marché doit se substituer à l'approche monopolistique.

La négociation en cours et l'adhésion prochaine à l'OMC (l'Organisation Mondiale du Commerce) qui à terme ouvrira quasi totalement le marché algérien aux produits et services en provenance des pays membres de cette organisation; dans les faits les étapes d'engineering et de procurement sont déjà l'œuvre de grandes compagnies internationales ; le domaine de la construction d'ensembles industriels dont lesquels, excelle l'ENGTP, commence à être soumis à rude concurrence par des entreprises de pays émergents (chine, l'inde...) pour les projets d'envergure et par des sociétés mixtes et des entreprises publiques diversifiant leurs activités notamment dans la construction de petits projets.

### Partie 3 : Présentation de GTP

La donne sociale est à considérer dans le nouveau contexte fortement concurrentiel ; l'essaimage d'activités subsidiaires devient un impératif pour recentrer les efforts de gestion sur la maîtrise du métier avec à terme l'essor de l'activité recherche et développement seule à même de pérenniser l'entreprise.

Sur le plan de l'environnement réglementaire, dispositifs mis en place à l'exemple du Bausem, engagent de plein fouet l'entreprise dans la compétition.

✓ Au plan externe l'entreprise doit tenir compte des effets :

#### • **La mondialisation**

Qui se répercute pour l'entreprise par :

- l'homogénéisation des marchés : les clients ont de plus en plus d'exigences.
- l'accélération des innovations technologiques tous secteurs confondus.
- la nécessité pour l'entreprise d'investir de façon continue dans leur outil industriel afin de demeurer économiquement concurrentielle.

#### • **Les nouveaux facteurs de compétitivité**

- Les compétences managériales.
- La maîtrise des technologies.
- Le partenariat.
- L'innovation.

La grande taille de l'entreprise, si elle lui est favorable pour les grands et moyens projets industriels, elle l'est moins pour des petits projets où la flexibilité et la maîtrise des coûts sont des impératifs majeurs ; ajouté à cela, la tendance drastique du marché international à réduire substantiellement les délais de réalisation de projets avec comme conséquence une tendance à une segmentation de projets en lots distincts, souvent par corps de métier.

L'activité de l'entreprise relève déjà de la typologie de production unitaire pour son activité principale liée à la réalisation de projets industriels ; or, il est communément admis que cette typologie de production se caractérise par la complexité et l'incertitude des activités à gérer dans le domaine, d'où l'exigence de centraliser la décision autour de la synergie d'une équipe compétente et pluridisciplinaire.

Comme toutes les entreprises algériennes, l'ENGTP est confrontée au passage d'un environnement simple qu'elle maîtrise, caractérisé par la certitude et la stabilité d'un marché restreint et fermé où les concurrents utilisent les mêmes armes, à un environnement complexe qui lui échappe caractérisé par l'incertitude, le changement

### Partie 3 : Présentation de GTP

rapide, un marché plus vaste et ouvert et des concurrents plus aguerris et qui utilise des armes plus sophistiquées et changeantes.

Pour faire face au triptyque des 3C (Clients, Concurrence, Changement) et imposer son positionnement, l'entreprise ne peut y demeurer sans un engagement sans équivoque sur des domaines tels que ceux du marketing et de la communication.

Enfin, les nouvelles technologies de l'information sont des atouts actuels de l'environnement qui permettent à l'entreprise de transcender les grands espaces de son aire géographique d'intervention et dont les anciens dirigeants de l'entreprise n'avaient pas eu accès et alors de façon centralisée la réalisation des projets par le passé.

#### **4.1. Politiques commerciales**

Au regard des prévisions budgétaires pessimistes, l'entreprise n'affiche pas une dynamique commerciale favorable ; son management subit l'évolution de l'environnement et du marché.

##### **4.1.1. Produits**

Sur le plan de la gamme de produits que peut proposer GTP, celle-ci est relativement dense.

On pourrait citer comme activités principales :

- L'engineering
- Le procurement
- La construction
- Les prestations de services diverses (formation, contrôles, assistances).

##### **4.1.2. Surveillance des marchés**

L'entreprise n'a pas d'ambitions pour l'exportation de ses services et alors qu'elle a des références à faire valoir dans les pays de la région et dans le secteur des hydrocarbures. Un client étranger a admis que « l'entreprise GTP doit cesser sa mentalité à faire concurrence à des entreprises locales, elle doit être plus agressive afin de s'élever et se comparer aux sociétés internationales !! »

L'entreprise ne dispose pas d'informations régulières sur l'évolution des marchés nationaux ; ce qui constitue un manquement à rattraper sur les aspects essentiels comme l'évolution quantitative de la demande et des besoins, l'évolution des produits substitués et de l'offre des concurrents ainsi que leurs stratégies. De façon générale, veille des concurrents, de

### Partie 3 : Présentation de GTP

veille des fournisseurs, de veille du secteur des hydrocarbures et de veille d'informations économiques générales.

#### **4.1.3. La concurrence**

Il importe de mentionner qu'une concurrence vive, jusque là cantonnée aux aspects engineering et procurement, s'élargit de plus en plus au secteur de la construction ou l'entreprise GTP était privilégiée par son principal client Sonatrach.

Des concurrents sérieux se positionnent sur les marchés du montage industriel à l'instar des entreprises issues de Sonelgaz : Etterkib pour le montage, Kanagaz pour la canalisation, Inerga pour le génie civil, ainsi que d'autres, telles : Cosider, Tarsi, Enac, GCB, Baticic et enfin des sociétés mixtes (entreprises du secteur public avec des sociétés étrangères) à l'exemple de Sarpi, Entral, Safir, Benco, etc.

L'entreprise connaît une forte concurrence et pour y faire face, elle doit améliorer ses performances décisionnelles.

#### **4.1.4. Les clients**

Il faut signaler l'ouverture du secteur des hydrocarbures vers les multinationales particulièrement durant les dernières années ; cette évolution s'est traduite par la mise en place du Baosem.

L'entreprise est soumise à la pression de clients aux travers de leur exigence sur :

- La réduction des prix, ce qui a mené Sonatrach à renégocier les dernières conventions maintenance industrielle en sa faveur en mettant en compétition GTP, et alors que sur le plan qualité ce client en est satisfait.
- La qualité avec le niveau exigé du zéro défaut.
- L'environnement car les entreprises étrangères ont une longueur d'avance à ce sujet par l'application du référentiel ISO 14001.

Malgré cela, il faut signaler tout de même le satisfecit de clients envers les prestations fournies par GTP.

Nous pouvons citer quelques clients algériens: Sonatrach, Naftal, Naftec, Sonelgaz, Saidal, Eriad Sétif, Eriad Alger, Asmidal. Ainsi que des sociétés étrangères telles: Entrepouse, JGC, ARCO, KBR Pétrofac, Agip, Brown and Root.

#### **4.1.5. Fournisseurs**

La contrainte à ce niveau est réelle, et elle n'est pas à sous estimer car une conjoncture pourrait traduire par une alliance avec un concurrent avec notamment le développement de sociétés mixtes ; la réalisation de la sphère d'éthylène à Arzew a subi plusieurs mois de retard pour des considérations d'approvisionnement et donc de fournisseur.

#### **4.1.6. Production**

La réalisation de projets industriels est articulée autour des 3 niveaux EPC (Engineering ; Procurement ; Construction).

Nous pouvons constater que la production est structurée autour du chapitre construction qui domine l'activité de GTP ; ensuite vient le second chapitre qui est la fonction Engineering / Procurement dont l'activité est réduite au sein de GTP ;

Un autre volet enfin permet d'examiner la fonction de service par la mise à disposition de moyens humains et matériels au client Sonatrach.

#### **4.1.7. Partenariat**

(Coopération : GTP/NUOVO PIGNONE pour la fabrication des AEROREFRIGERANTS)

Dans le souci de satisfaire les exigences accrues liées au plan de développement de l'énergie et des hydrocarbures en Algérie, GTP et NUOVO PIGNONE ont convenu d'engager une collaboration ayant pour objet la réalisation de nouvelles installations industrielles et la modernisation d'installations déjà existantes.

Cette collaboration a été entreprise par Sonatrach dans le cadre de sa politique de relance des ressources nationales, entreprise qu'elle a engagée avec succès et continue de mener à bien comme en témoignent les réalisations accomplies. S'appuyant sur une expérience de trente ans qui a pu développer, en Algérie, grâce à la confiance qui lui a été accordée par Sonatrach et par d'autres sociétés algériennes. NUOVO PIGNONE, grâce à sa technologie de pointe, est en mesure de fournir les meilleures installations de surfaces.

De son côté, GTP occupe une position de premier plan dans le domaine de réalisation d'installations destinées à l'industrie du pétrole et du gaz. Ceci explique pourquoi les deux sociétés ont passé, en mai 1991, un accord de coopération portant sur la fabrication conjointe d'aéroréfrigérants conformes aux standards et aux normes internationales.

22 aéroréfrigérants ont déjà été fabriqués dans le cadre du projet HAMRA.

### Partie 3 : Présentation de GTP

La fabrication d'autres aéroréfrigérants est également réalisée dans le carnet de commandes GTP-NUOVO PIGNONE au titre de projets Sonatrach :

- Centre de Hassi R'mel.
- Centre de production de pétrole de BIR REBAA NORD.
- Centre de récupération des gaz associés de OUED NOUMER.
- GPL de Hassi Messaoud.
- Projets de la « Chaîne GPL ».

## **5. Ressources et politique de l'entreprise**

### **5.1. Humaines**

L'entreprise GTP emploie en moyenne 7800 agents, dont les principaux domaines d'activité sont les suivant :

La Réalisation : 54%, la Maintenance Industrielle : 15%, la logistique : 13%, gestion : 7%, catering : 9%, sécurité : 2%

Année	Permanent		Temporaire		Total effectif
	Nombre	%	Nombre	%	Nombre
2002	4593	57.9	3343	42.1	7936
2003	4230	50.1	4209	49.9	8439
2004	3920	48.0	4243	52.0	8163

**Tab -7- Evolution des effectifs par statut**

Les emplois de l'entreprise se situent autour de 8000 agents ; la proportion de temporaires a évolué de 10% durant la période d'observation au détriment de celle des permanents qui régresse du même taux.

### **5.2. Matérielles**

GTP dispose de :

- Un atelier de soudage.
- Des ateliers de Tuyauterie.
- Des Laboratoires de Contrôle Destructif et Non Destructif.

### Partie 3 : Présentation de GTP

- D'un parc matériel roulant de 3000 équipements, dans les principaux domaines ci-après :
- Le Terrassement et Génie Civil.
- Le soudage et contrôle.
- Le Levage (Grue de 18 T à plus de 300T).

#### ❖ Politique Qualité et Environnement :

De renommée internationale prouvée, GTP est certifiée ISO 9001/2000 depuis 2004.

## **5.3. Situation informatique de l'entreprise**

### **5.3.1. Le réseau informatique**

Un réseau est une configuration de plusieurs micros ordinateurs reliés à un serveur ou une application stockée au niveau du serveur peut être utilisée par un ou plusieurs utilisateurs en même temps.

Des réseaux LAN (local area network) sont installés au niveau siège et dans toutes les régions de l'entreprise.

#### **Caractéristiques : (réseau base de données)**

- Réseau local de type Ethernet.
- Architecture étoile.
- La plate forme de serveur : Intel biprocesseur avec un contrôleur RAID pourvu d'un lecteur des bandes pour la sauvegarde des données et contient 6 disques durs de capacité 18 Giga Octet pour chacun.
- Système d'exploitation : - Serveur (Windows NT4).  
- Client (Windows 98, XP professionnel).
- Le protocole utilisé: TCP/IP.
- SGBD oracle 9i (en mode client/serveur).

#### **Autres applications réseau utilisées**

- Système de messagerie: MS exchange2000 serveur de Microsoft.
- Système de gestion du parc informatique: Logiciel administratif (SMS System Management Server).

### 5.3.2. Les données

L'entreprise dispose d'une base de données centrale contenant l'ensemble des données et des bases de données régionales sont identiques à celle centrale mais ne sont accessibles en mode mise à jour que pour les données de la région concernée. Les bases de données sont installées sur des serveurs : Un serveur central entreprise et des serveurs régionaux.

Les postes de travail qui ont accès aux bases de données sont reliés aux serveurs à travers des réseaux locaux. Ces postes de travail sont appelés Clients par rapport aux Serveurs. C'est le mode Client / Serveur.

L'accès aux bases de données se fait à travers des applications développées par le centre informatique et installées sur les postes Clients conformément au schéma suivant :

Des postes clients et un serveur de données entreprise au niveau siège interconnectés via un réseau local.

Des postes clients et un serveur de données régional implantés au niveau de chacune des quatre directions régionales interconnectés via des réseaux locaux.

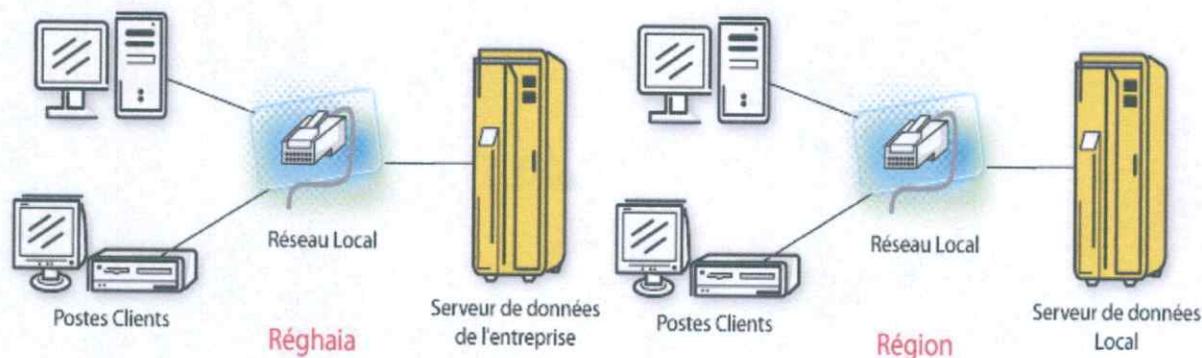


Fig -29- Architecture Client / Serveur à GTP

### 5.3.4. Les traitements

Les bases de données régionales sont mises à jour de deux manières différentes :

En local pour les différents mouvements effectués dans la région, par les services informatiques de l'entreprise, après les traitements de consolidation.

La mise à jour des bases de données permet de visualiser la situation des produits et des charges de l'ensemble des centres de frais de l'entreprise à travers la production d'états et de rapports synthétiques mis à la disposition des managers.

❖ **Sous-système Gestion du Personnel et Paie** Il intègre trois fonctions :

- **La gestion du Personnel** : Elle prend en charge la gestion de l'agent de son recrutement jusqu'à la fin de la relation de travail.

- **La paie** : La paie est un système indemnitaire et ses éléments constitutifs sont regroupés autour d'un ensemble de paramètres.

- **L'interface avec le système « Comptabilité »** : Cette fonction permet l'intégration des résultats issus du calcul paie au niveau du de la comptabilité générale et analytique.

✓ **Les tableaux de bord et statistiques** : Ce sont l'ensemble des éléments constituant les informations décisionnelles issues des traitements après paie.

❖ **Sous-système Gestion Finances et Comptabilité**

Le sous-système traite l'ensemble des mouvements comptables de l'entreprise. Ces mouvements proviennent des sources différentes :

- Des différents services de comptabilité (générale et analytique) du siège et des directions régionales.

- Journaux financiers, trésorerie, avis d'encaissement, demandes et règlements, opérations diverses, comptabilité analytique.

- Des autres sous-systèmes : gestion du personnel et paie, gestion du matériel, gestion des stocks, gestion de la maintenance.

Pour fournir les résultats suivants :

• **En Comptabilité Générale**

Journal auxiliaire, grand livre, balance générale consolidée, balance générale par région.

• **En Comptabilité Analytique**

Grand livre, balance analytique, résultats analytiques détaillés, résultats analytiques récapitulés.

## **6. Notoriété de GTP**

La liste des réalisations de GTP est longue à énumérer tant l'entreprise a été leader dans le domaine de la construction de grands ensembles industriels dans le secteur des hydrocarbures.

Sa notoriété est donc bel et bien établie au sein du secteur.

L'entreprise a toujours honoré ses engagements et a tiré des satisfactions de reconnaissances de ses partenaires en sus de résultat financiers.

Dans le domaine des reconnaissances pourraient être citées :

- Attestation de satisfaction de l'entreprise Naftal pour les études, la réalisation et la mise en route de centres enfûteurs effectués par GTP.
- Attestation de satisfaction faite par le groupement TFT « Sonatrach-Total-Repsol » pour la construction du projet : développement des champs de gaz de TFT, réalisé par GTP.
- Attestation de reconnaissance faite par Sonarco / Kellogg GTP sur la satisfaction de la productivité déployée et le succès du projet.

## **7. Objectifs de GTP**

Les objectifs que cherche à atteindre GTP sont :

- ❖ Le développement et l'amélioration de la rentabilité de ces activités.
- ❖ La fidélisation de sa clientèle.
- ❖ Une exploitation de toutes les opportunités offertes par la mondialisation tant au niveau national qu'international.
- ❖ La pénétration des marchés extérieurs.
- ❖ L'augmentation de son chiffre d'affaires et de sa part de marché.

## **8. La finalité du projet**

Après avoir présenté l'entreprise, nous pouvons constater qu'elle est instable et connaît une concurrence accrue dans ses domaines d'intervention surtout avec l'apparition de nouvelles entreprises privées de plus petites tailles.

Pour prendre de meilleures décisions, les managers ont besoin d'avoir une vue consolidée des mesures d'activités et de leurs indicateurs clés de performance. Certes, les dirigeants

### Partie 3 : Présentation de GTP

disposent de tableaux de bord mais ils sont statiques et ne permettent pas de changer les indicateurs.

Donc un data warehouse métier est nécessaire aux décideurs de l'ENGTP pour analyser leurs activités et retrouver l'information recherchée (historisée) en temps et en heure et surtout capitaliser sur les expériences.

L'ambition de GTP à travers ce projet est de simplifier le pilotage de la ressource humaine afin de mieux gérer et surveiller les indicateurs socio-économiques, dans la mesure où l'ensemble des ressources, qui concourent à la réalisation des projets de l'entreprise sont gérées à travers des systèmes informatisés intégrés, utilisant différentes applications comme réservoir de stockage des données, la recherche de l'information à l'intérieur de ses systèmes devient difficile surtout avec l'augmentation du volume de ces bases de production.

Notre travail consiste à concevoir une base qui contiendra une copie des données vitales de l'entreprise, cette base sera un entrepôt de données qui va collecter les données disparates dans l'entreprise, concernant la ressource humaine, les alimenter et les charger afin de pouvoir les exploiter à des fins d'analyse. De plus les nouvelles technologies permettent de traiter des données provenant de diverses sources d'information et de formats différents (tables oracle, fichiers textes, tableaux Excel etc.).

## **9. Conclusion**

Dans cette étape, nous avons pu cerner l'organisme d'accueil, en présentant l'environnement dans lequel il évolue. Après avoir étudié globalement ses ressources, ses politiques, ses concurrents et sa place dans le marché. Notre entrepôt de données se limitera à la fonction suivi des ressources humaines que nous avons jugée comme l'une des fonctions principales de GTP.

Dans la prochaine étape, nous allons présenter les différents besoins des décideurs cette étape est le point important placé au centre de l'univers de l'entrepôt de données.



Partie .4



# Définition des besoins



## **1. Introduction**

Maîtriser l'information est devenu un enjeu stratégique pour l'ENGTP, prendre les décisions de façon pertinente et rapide est aujourd'hui indispensable à tous les niveaux. L'efficacité dépend de la maîtrise de l'information et de la capacité de réaction. L'utilisation d'un data warehouse contribue fortement à cette efficacité ; L'ENGTP utilise actuellement divers systèmes sources pour gérer ses activités et leurs processus. Les systèmes décisionnels permettent d'avoir la meilleure vision possible des activités d'une entreprise, ils permettent aux dirigeants d'avoir les données adéquates pour prendre les bonnes décisions. Conscient de cet atout majeur les décideurs au sein de l'entreprise ont donné le feu vert pour l'élaboration d'un système d'information décisionnel basé sur un Data warehouse (entrepôt de données). La conception d'un tel system se fait en plusieurs phases dont la première consiste en une déclinaison de la stratégie de l'entreprise en objectifs métiers pour la fonction suivi de la ressource humaine. Cette phase permet de dresser la liste des indicateurs nécessaires au pilotage des différents processus de réalisation ainsi que les données sources (matières premières) pour les obtenir.

La valeur d'un entrepôt de données tient à la qualité des données qu'il contient. Il convient donc de ne l'alimenter qu'avec des données indispensables, fiables et cohérentes.

Cette phase passe par un travail de recensement et de compréhension des besoins des utilisateurs (analystes et décideurs de l'entreprise). Ce travail, par ailleurs indispensable, permet de :

- Traduire de la façon la plus précise possible les besoins exprimés en données concrètes et manipulables.
- Déterminer les liaisons entre ces données (axes d'analyse).
- Nettoyer et mettre en cohérence ces données.

Ce travail trouve sa concrétisation dans la création d'un dictionnaire de données, véritable référentiel d'entreprise.

## **2. Consistance du Projet de fin d'études**

La réalisation des projets au niveau de l'entreprise nécessite la combinaison des trois ressources de base :

- Financières ;
- Matériels ;
- Humaines ;

Alliées à la maîtrise du planning et des avancements des travaux de réalisation.

Le projet de mise en place d'un entrepôt de données et d'un système d'information décisionnel a pour ambition de prendre en charge la gestion de l'ensemble de ces ressources à travers la maîtrise de processus métiers basés sur des sous entrepôts ou datamarts. Compte tenu de la dimension de celui ci et du temps qui nous est alloué, nous avons pris option pour le processus métier « Ressources humaines ».

## **3. Choix de la méthode de définition des besoins**

Cette première phase peut être effectuée de plusieurs manières différentes :

- A travers des entretiens avec les responsables des structures impliquées dans le processus de réalisation.
- Par l'envoi aux structures de questionnaires préalablement établis.
- A travers la lecture des documents nécessaires circulant dans l'entreprise.
- L'organisation de réunions de travail.

Notre recherche était menée selon deux méthodes que nous avons jugées appropriées et complémentaires, la première était la documentation, nous avons pu lire tous les documents mis à notre disposition et concernant les activités de l'entreprise, cela nous a permis de recenser toutes les informations qui ont un lien direct avec la conception du data warehouse.

La deuxième méthode était une série d'entretiens avec les responsables concernés de la direction des ressources humaine :

- 1- Service gestion et ressources humaines,
- 2- Service formation et recrutement,
- 3- Département finances et comptabilité,

## **4. Définition des besoins**

Nous citons ci-dessous les différents documents que nous avons lus ainsi que les entretiens établis dans les différentes directions.

### **✓ Les documents**

Les documents que nous avons pu voir sont :

- Le tableau de bord mensuel.
- Le bilan annuel.

### **✓ Les entretiens**

Nous avons pu discuter avec de nombreuses personnes connaissant et maîtrisant leur travail. Grâce à ces entretiens, nous avons pu recueillir les données nécessaires à la réalisation de cette étude. Ces entretiens ont porté sur :

- La connaissance de l'activité principale de l'entreprise afin d'identifier les processus métiers ;
- Les indicateurs d'analyse de l'activité, le but de cette identification est de :
  - Cerner les besoins et les attentes d'utilisateurs.
  - Déterminer les dimensions et les faits de mesure afin d'établir la modélisation dimensionnelle.

Avant d'élaborer les entretiens, nous avons préparé quelques questions (annexe) que nous avons pu tirer après lecture des documents et compréhension de l'activité de l'entreprise.

Nous avons décidé d'effectuer deux interviews, le premier était mené avec un membre de la haute direction (Directeur des ressources humaines).

Le second était avec des responsables et analystes (Chef de service).

La date des entretiens était fixée par les interviewés, le premier s'est déroulé le début du mois mais le second était fixé pour la fin du mois (les prévisions du mois en cours et le contrôle des valeurs du mois précédent se font au début de chaque mois).

## **5. Objectif**

L'objectif à travers ces entretiens et la lecture de ces documents est d'extraire l'ensemble des axes sur lesquels l'utilisateur voudrait fonder son analyse et les indicateurs essentiels qui nous aideront lors de la conception du modèle dimensionnel.

- Les premières questions sont au nombre de trois et avaient pour but de savoir plus sur les processus métiers et les relations entre eux que les dirigeants pilotent et sur lesquels ils essayent d'influer.
- Les autres questions avaient pour but de tenir compte de la manière qui permet de décrire leurs activités, les données utiles à la prise de décision. Ainsi que le type d'analyse qu'ils effectuent en général (sur quoi portent-ils leurs analyses) et surtout s'il leur est facile d'avoir les données recherchées dans les systèmes de production pour faire leurs rapports d'analyse.

Ces interviews nous ont aidés à :

- Recueillir les besoins utilisateurs, analyser les processus métiers et l'organisation interne afin de définir les informations & indicateurs pertinents
- Identifier les données sources des systèmes opérationnels internes et externes (existence & qualité des données)
- Définir les données et le modèle de données cible associé
- Choisir la solution technique la mieux adaptée au métier à traiter parmi les outils du marché (connaissance des solutions du marché, critères de choix)

✓ Ce que nous avons constaté d'après ces entretiens que les besoins des utilisateurs diffèrent concernant la manière qu'ils souhaitent voir les informations désirées, mais les indicateurs clés sont toujours les mêmes et se répètent dans chaque entrevue.

### **5.1 Compte rendu des entretiens**

Après chaque entretien, nous avons rédigé un compte rendu qui regroupe le contenu de ces entretiens.

#### Partie 4 : Définition des besoins

Nous avons déchiffré les entretiens et traduit en modèle de données, nous allons citer quelques besoins tels indiqués par l'utilisateur :

- Donner aux utilisateurs une vision quasi instantanée de l'activité
- Permettre un pilotage efficace et rapide ainsi qu'une grande réactivité automatiser les tableaux / états / reporting, Publier à une cadence prédéfinie, Restituer l'information plus vite aux clients / utilisateurs.
- Aider au contrôle qualité des systèmes d'informations
- Disposer d'un niveau de granularité poussé (rendre possible le suivi « un par un » par exemple employé par employé)
- Génériquement, collecter, historiser et tracer l'information

A partir de tous les besoins, nous avons déterminé les différents indicateurs et axes d'analyse, mais avant cela, nous devons connaître les processus de cette activité (Suivi de la ressource humaine). Pour cela nous allons décrire la démarche suivie pour chaque employé depuis son recrutement jusqu'à sa fin de relation de travail.

Après avoir rencontré les différents acteurs (utilisateurs) du tableau de bord, nous leurs avons finalement demandé de rédiger une liste qui est un échantillon de questions (Business Questions) qu'ils peuvent et veulent poser à notre système.

Au département RH :

1. Effectif de sexe féminin, permanent, par groupe professionnel par région
2. Effectif de sous-traitance par région
3. Répartition du personnel admis à la retraite en activité par structure et par groupe professionnel, par direction et par région
4. Répartition du personnel par ancienneté

Au département Finance :

1. Masse salariale par mois par région par type de contrat.

## Partie 4 : Définition des besoins

Ainsi par exemple on peut découler depuis la 3ieme question du département RH le nombre et la date exacte de postes qui seront vacants ce qui a pour effet de déclencher soit un processus de recrutement ou bien des promotions de postes.

## **6. Les processus métiers**

« Un processus est un ensemble d'activités organisées dans le temps produisant un résultat précis et mesurable »

Nous pouvons définir un processus d'une entreprise en tant qu'un enchaînement d'activités corrélées ou interactives.

Un processus reçoit des objets en entrée et leur ajoute de la valeur, par le moyen de ressources, tout en fournissant des objets de sortie remplissant les besoins et les exigences (atteindre les objectifs) dans le temps.

### **6.1 Identification des processus**

Lors de la série d'entretiens que nous avons eus, nous avons pu constater que les processus clés pour les décideurs en ce qui concerne la fonction suivi de la ressource humaine sont :  
Mouvement personnel, Personnel sortant, Personnel inactif, Réalisation financière et le processus le plus important qui est la Réalisation personnel, pour analyser ses performances, en général, l'entreprise dispose d'un document textuel mensuel « le Tableau de bord », ce document contient toutes les informations et les remarques qui ont été faites sur la ressource humaine. Donc, nous avons jugé qu'il serait important d'importer les différents indicateurs de ce document en tant qu'axes d'analyses sur notre entrepôt.

#### **1. Réalisation personnel**

Ce processus est le plus important vu qu'il englobe la majorité des dimensions nécessaires à l'analyse tel que l'activité, la région, la qualification...etc. ce qui permet selon le choix des critères d'avoir une vue globale ou bien plus en détail.

#### **2. Mouvement personnel**

Ce dernier permet de voir tous les nouveaux admis en tant qu'employés au sein de la société ainsi que les personnes qui ont atteint leur date de FRT (fin de relation de travail), pour mieux cerner les besoins de l'entreprise en ressource humain.

## Partie 4 : Définition des besoins

### **3. Personnel sortant**

Comme son nom l'indique ce processus prend en charge l'effectif sortant, c'est un sous-processus du Mouvement personnel mais qui gère uniquement les deux aspects suivants :

Le personnel admis à la retraite.

Le personnel démissionnaire.

Le personnel licencié.

Il permet donc de prévoir les recrutements ou promotions de postes avenir.

### **4. Personnel inactif**

Il existe une multitude de raisons pour lesquelles un employé devient non actif tel que une maladie, une maternité, indiscipline etc..

Ce dernier permet de lister ces cas afin que les décideurs puissent combler les vides (postes vacants).

### **5. Réalisations financières**

Englobe toutes les dépenses et investissements qui touchent à la ressource humains tel que les salaires, Primes, Rappels, transport etc....

## **7. La dimensionnalité métier**

Les dimensions sont les axes sur lesquels l'utilisateur portera son analyse, pour chaque processus nous avons déterminé l'ensemble des dimensions qui lui sont associées.

Nous rappelons que c'est grâce aux entretiens que nous avons pu savoir sur quoi le décideur et l'analyste voudrait baser son analyse.

- La première dimension qui est présente dans tout entrepôt est le temps. Bien évidemment, comment on pourrait suivre l'évolution des employés sans avoir la date ? ainsi l'utilisateur pourra naviguer dans le data warehouse en déterminant l'année, le mois ou le trimestre auquel il souhaite faire son analyse.

- La deuxième dimension est l'employé lui même, cette dimension est la plus importante de notre analyse car c'est sur elle que vont s'exécutés tout les processus d'estimation.

#### Partie 4 : Définition des besoins

- La troisième dimension est la région, qui sont les suivantes: Réghaia, Hassi-Messaoud, Hassi R'mel, Skikda et Arzew.

Donc l'utilisateur a besoin de comparer entre les effectifs selon un critère qu'il choisit (age, sexe, tranche d'âge, ...etc) dans deux régions différentes.

- La quatrième dimension est le type d'activité, chaque employé appartient à un secteur donné, donc l'utilisateur peut avoir les différents effectifs par secteur.

- La cinquième dimension est la direction, l'utilisateur peut effectuer toutes les analyses selon la direction souhaitée.

- La sixième dimension est le type de contrat qu'il soit permanent ou bien temporaire.

- La septième dimension est la qualification tel que (chef de service, cadre technique, ...etc).

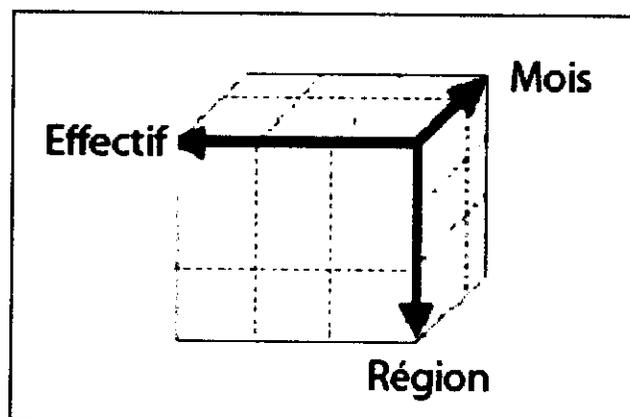
- La huitième dimension est le groupe professionnel qui se compose de trois principaux groupes, à savoir Cadre, Maîtrise et Exécution.

Donc l'utilisateur peut avoir sous forme textuelle toutes les remarques positives ou négatives ainsi que les étapes effectuées pour l'aboutissement du projet. Le principe de dimensions sera illustré dans le schéma qui suit.

Par exemple, un décideur se pose la question :

« Quel est l'effectif admis à la retraite dans telle région le mois dernier ? »

La présentation en dessous exprime ce besoin en termes de cube.



**Fig -30- Exemple de cube**

## **8. Les indicateurs d'analyse de l'activité**

Les données utilisées dans des applications d'aide à la décision ne sont pas identiques à celles utilisées dans des applications de gestion classiques. Elles se distinguent par le fait qu'elles sont "dédiées", c'est à dire proches du métier de chacun des utilisateurs.

Il ne faut pas, pour réussir la mise en place d'un Data warehouse, partir d'un "modèle générique de données", mais suivre le chemin inverse, et à partir des données existantes dans les systèmes d'information de l'entreprise, les synthétiser et aboutir au modèle de données propre aux applications d'aide à la décision.

Le contrôle d'indicateurs clés soigneusement définis permet de fournir une vue instantanée de l'activité au sein d'un département, par exemple : la masse salariale.

Les mesures doivent être également communiquées à l'ensemble des personnes qui tiennent un rôle important dans l'atteinte des objectifs.

Les utilisateurs (analystes et décideurs) souhaitent utiliser les informations produites par les processus métiers précédents afin de mesurer leur activité commerciale « Ce qui ne se mesure pas, ne s'améliore pas » [Kim 97] et pour mieux cerner leur environnement (employés, activités, etc.) .

A partir documents lus et les entretiens effectués, nous avons pu avoir les besoins suivants :  
Pour cela, voici les principaux indicateurs que nous avons pu extraire :

- ✓ Effectif actif / opérationnel / réel / écart
- ✓ Effectif en formation
- ✓ Effectif prévu
- ✓ Effectif payé
- ✓ Effectif arrivé / départ / écart
- ✓ Effectif admis à la retraite
- ✓ Masse salariale
- ✓ Effectif en congé spécial rémunéré (stand by)
- ✓ Détachement
- ✓ Maternité

## Partie 4 : Définition des besoins

- ✓ Recyclage
- ✓ Maladie longue / courte durée

## **9. Les systèmes sources**

Notre entrepôt sera alimenté à travers trois applications différentes qui sont:

- **L'application Personnel (Pers)**

C'est une application qui gère le suivi de l'employé depuis son recrutement jusqu'à sa sortie de l'entreprise. Elle regroupe la paie, fournit le nombre d'effectifs pour chaque projet ainsi que les heures dépensées qui sont les deux indicateurs qui nous intéressent et qui vont être chargés dans l'entrepôt.

- **L'application Formation**

Cette dernière permet de gérer tous les employés en formation qu'ils soient détachés vers des instituts de formations externes ou bien localement sur un des locaux de la société.

- **L'application Recrutement**

Nous permet de connaître avec précision les besoins en ressource humaine de chaque département et chaque spécialité ce qui permet une traçabilité des employés. Elle recense tous les contrats et gère le nouveau personnel.

## **10. Conclusion**

Au cours de ce chapitre, nous avons pu établir une liste des différents objectifs que cherchent à atteindre les décideurs de l'ENGTP, et ce à partir d'une analyse de leurs besoins en ce qui concerne l'aspect ressources humaines à l'aide d'entretiens et consultation des documents.

Nous avons aussi défini les processus clés qui contribuent à l'analyse de la fonction, c'est à partir de ces processus et indicateurs que sera faite notre modélisation dimensionnelle.



Partie .5



Conception

Ce chapitre traite la conception logique et physique de notre système décisionnel. Nous avons divisé ce chapitre en deux sections. La première consacrée à la modélisation dimensionnelle des sujets précédemment sélectionnés et la deuxième pour la mise en œuvre de l'architecture du système.

## **1. Modélisation dimensionnelle**

La modélisation multidimensionnelle consiste à considérer un sujet analysé comme un point dans un espace à plusieurs dimensions. Les données sont organisées de manière à mettre en évidence le sujet analysé et les différentes perspectives de l'analyse.

« Les données à analyser au niveau d'un magasin doivent refléter la vision d'une classe d'analystes » [Kim, 96].

La conception logique d'un data warehouse passe par quatre premières étapes :

- Choix du processus d'activité à modéliser.
- Choix du grain du processus d'activité.
- Choix des dimensions applicables à chaque table de faits.
- Choix des mesures que contiendra chaque enregistrement de la table de faits.

### **1.1. Modélisation dimensionnelle de l'activité**

#### **« Réalisation Personnel »**

##### **1.1.1. Le processus d'activité**

Sur la base de notre étude, et grâce aux entretiens qui nous ont aidés à mieux connaître l'activité et les ressources de données disponibles. Le premier processus important choisi est l'activité Réalisation Personnel, les analystes en ont besoin pour voir l'effectif des employés selon leurs état (actif, opérationnel, formation) ainsi que ceux qui ont été payé.

##### **1.1.2. Le grain du processus d'activité**

Le grain est important parce qu'il détermine les différentes dimensions de la base de données. La granularité de la table des faits choisie est la plus fine possible, plus le niveau de détail est fin plus la conception est robuste. Ainsi chaque enregistrement sera crée dans la table de faits qui sera consultable chaque mois.

Ce chapitre traite la conception logique et physique de notre système décisionnel. Nous avons divisé ce chapitre en deux sections. La première consacrée à la modélisation dimensionnelle des sujets précédemment sélectionnés et la deuxième pour la mise en œuvre de l'architecture du système.

## **1. Modélisation dimensionnelle**

La modélisation multidimensionnelle consiste à considérer un sujet analysé comme un point dans un espace à plusieurs dimensions. Les données sont organisées de manière à mettre en évidence le sujet analysé et les différentes perspectives de l'analyse.

« Les données à analyser au niveau d'un magasin doivent refléter la vision d'une classe d'analystes » [Kim, 96].

La conception logique d'un data warehouse passe par quatre premières étapes :

- Choix du processus d'activité à modéliser.
- Choix du grain du processus d'activité.
- Choix des dimensions applicables à chaque table de faits.
- Choix des mesures que contiendra chaque enregistrement de la table de faits.

### **1.1. Modélisation dimensionnelle de l'activité**

#### **« Réalisation Personnel »**

##### **1.1.1. Le processus d'activité**

Sur la base de notre étude, et grâce aux entretiens qui nous ont aidés à mieux connaître l'activité et les ressources de données disponibles. Le premier processus important choisi est l'activité Réalisation Personnel, les analystes en ont besoin pour voir l'effectif des employés selon leurs état (actif, opérationnel, formation) ainsi que ceux qui ont été payé.

##### **1.1.2. Le grain du processus d'activité**

Le grain est important parce qu'il détermine les différentes dimensions de la base de données. La granularité de la table des faits choisie est la plus fine possible, plus le niveau de détail est fin plus la conception est robuste. Ainsi chaque enregistrement sera créé dans la table de faits qui sera consultable chaque mois.

### **1.1.3. Les dimensions**

D'après les granularités de la table de faits, nous avons déterminé une série de dimensions principales suivantes : Activité, Direction, Temps, Région, Type contrat, Qualification, Tranche d'âge, Sexe, Groupe professionnel.

Les tables de dimensions contiennent les descriptions textuelles de tous les attributs des dimensions.

#### **a. Dimension « Temps »**

La seule dimension qui figure systématiquement dans tous les entrepôts de données, car en pratique tout entrepôt est une série temporelle.

« Le temps est le plus souvent la première dimension dans le classement sous jacent de la base de données ». [Kim, 97]. Cette dimension est à un grain mensuel, donc chaque enregistrement de la table de dimension représente un mois.

#### **b. Dimension « Direction »**

Désigne la direction à laquelle est affecté l'employé, on en dénombre 24 tel que « Cellule Juridique », « Direction qualité » « Direction engineering » ...etc.

#### **c. Dimension « Activité »**

Décrit les différentes branches principales de métiers au sein de l'entreprise qui sont au nombre de 15 comme « Maintenance », « Sécurité », « Réalisation »...etc.

#### **d. Dimension « Région »**

ENGTP est présente sur différents sites (régions) « Réghaia », « Arzew », « Skikda », « Hassi R'mel », « Hassi messaoud » ainsi que l'école de soudure et d'expertise d'Arzew et les services DMI, Catering qui sont considérés comme des régions à part entières.

#### **e. Dimension « Type contrat »**

Au niveau de l'ENGTP il existe cinq différents types de contrats qui relient l'employé à l'entreprise : Temporaire, Permanent, Stagiaire, Cadre dirigeant, Syndicat.

#### **f. Dimension « Qualification »**

Cette dernière est plus précise que le groupe de travail car elle désigne le poste occupé par l'employé ciblé, vu la taille de l'entreprise il en existe une multitude tel que chef d'atelier, chef de centre médico-social, magasinier ...etc.

## Partie 5 : Conception et mise en œuvre du système

### **g. Dimension « Tranche d'âge »**

Les employés de l'entreprise subissent une décomposition selon leurs âges qui débute des moins de 20 ans jusqu'aux plus de 60 ans sur des tranches de 5 ans.

Exemple : 20-25, 25-30.

### **h. Dimension « Sexe »**

Utile pour voir les disparités et la distribution des deux sexes (Masculin, Féminin) sur les différents postes de travail.

### **i. Dimension « Groupe professionnel »**

Désigne la hiérarchie des postes de travail (grades), elle comporte neuf éléments comme les cadres dirigeants, les ingénieurs, les techniciens...etc.

## **1.1.4. Les faits mesurés**

Le fait modélise le sujet de l'analyse. Un fait est formé de mesures correspondant aux informations de l'activité analysée.

« Les mesures d'un fait sont numériques et généralement valorisées de manière continue »  
[Kim, 96].

Pour cette activité, les faits que nous avons enregistrés sont :

- Effectif actif.
- Effectif payé.
- Effectif en formation.
- Effectif opérationnel.

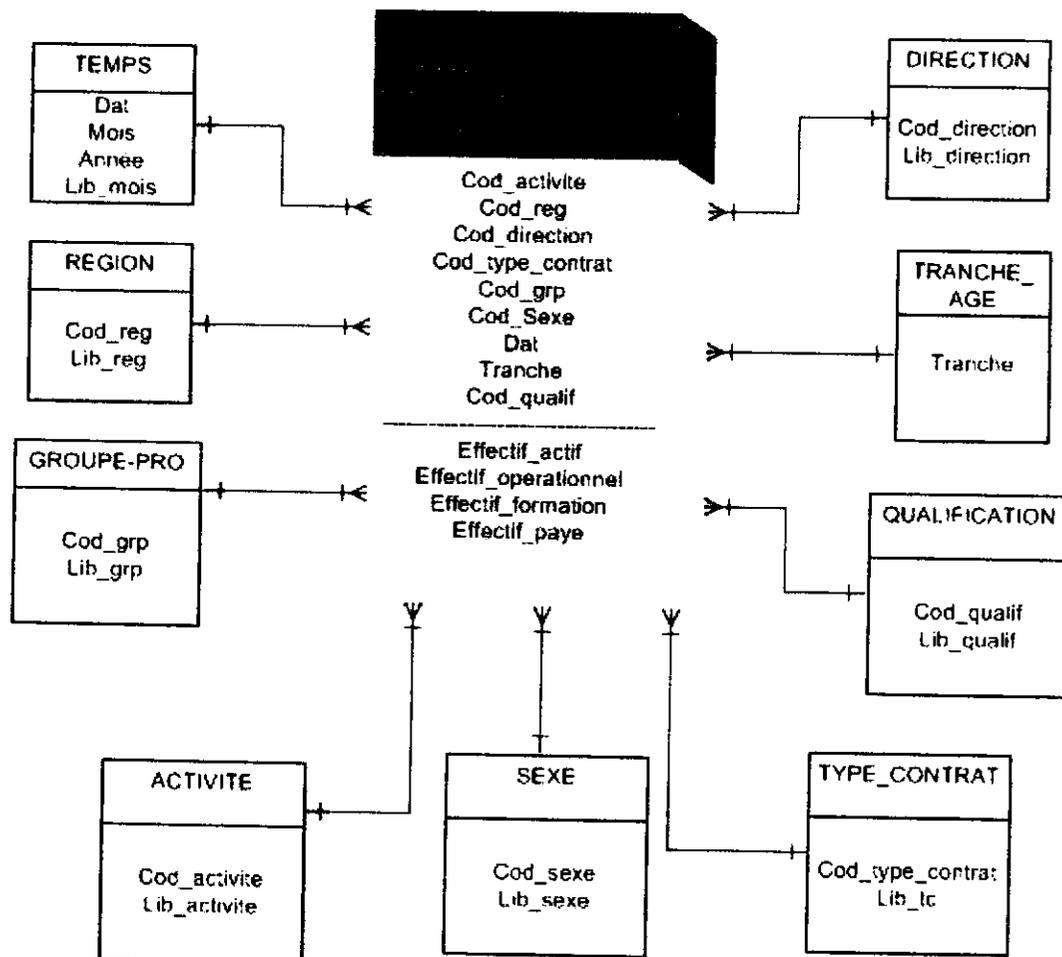


Fig -31- Table de fait Réalisation Personnel

## 1.2. Modélisation dimensionnelle de l'activité

### « Personnel sortant »

#### 1.2.1. Le processus d'activité

Ce processus permet de voir les différents effectifs sortant de l'entreprise dont les causes sont en nombre de quatre : Admis à la retraite, Fin de relation de travail (FRT), Licencié, Démissionnaire.

Il permet de voir les écarts entre l'effectif actif c'est-à-dire directement productif et l'effectif inactif, ainsi que d'analyser entre autres les différentes causes qui poussent un employé X à quitter son poste se qui donnera suite aux démarches nécessaires pour remédier à ses lacunes.

### 1.2.2. Le grain du processus d'activité

Dans ce processus, il est important de connaître mensuellement le personnel sortant pour prévoir les recrutements/promotions de postes avenir.

### 1.2.3. Les dimensions

Ce processus partage trois dimensions avec le précédent à savoir la région, le groupe professionnel et le temps ainsi qu'une autre dimension « Personnel ».

#### d. Dimension « Personnel »

Permet de désigner avec exactitude la personne concerné par se processus en listant plusieurs informations qui le lient à l'entreprise le tout à travers son matricule unique qui est affecté à chaque employé lors de son recrutement.

### 1.2.4. Les faits mesurés

Pour ce processus voila les faits qu'on a enregistré

- Effectif admis à la retraite.
- Effectif en fin de relation de travail (FRT).
- Effectif licencié.
- Effectif démissionnaire.

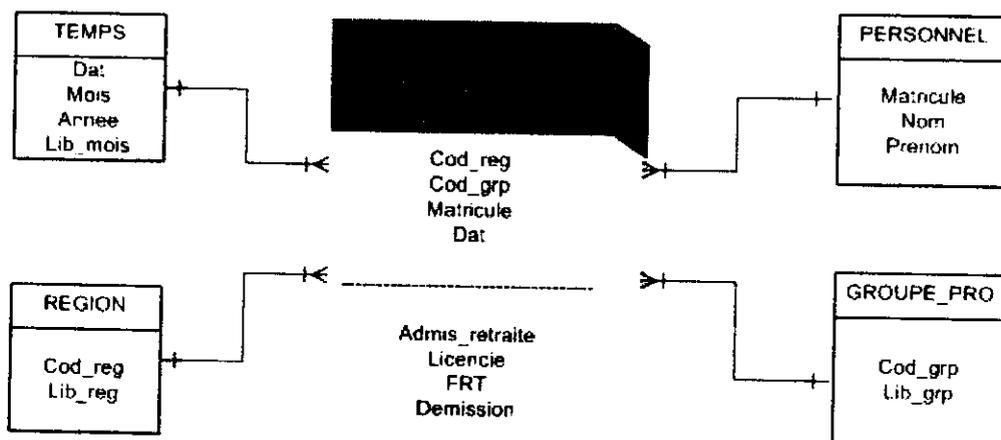


Fig -32- Table de fait Personnel Sortant

### **1.3. Modélisation dimensionnelle de l'activité**

#### **« Personnel inactif »**

##### **1.3.1. Le processus d'activité**

Personnel inactif recense les différents effectifs qui ne sont pas directement productifs au sein de l'entreprise, c'est en fait le contraire du fait mesurable « effectif actif » de l'activité Réalisation Personnel, il se distingue par des précisions quant à la cause de leurs inactivité qui seront cités dans les faits mesurables.

Ce processus est d'autant plus utile qu'il permet par exemple aux décideurs de donner l'acceptation ou le refus d'accorder des congés selon le nombre de personnes actives et la fréquence des arrêts de travail ou bien encor d'envoyer un employé en recyclage.

##### **1.3.2. Le grain du processus d'activité**

Du a l'importance de ce processus et a l'improbabilité des causes d'inactivités les décideurs on juger qu'grain d'un mois est le plus raisonnable afin de leurs permettre de faire leurs choix.

##### **1.3.3. Les dimensions**

Cette activité est liée par cinq dimensions déjà décrite en haut qui sont : la région, le groupe professionnel, la direction, le type de contrat et le temps.

##### **1.3.4. Les faits mesurés**

- Effectif en Stand by.
- Effectif en maternité.
- Effectif mis en disponibilité.
- Effectif détaché.
- Effectif en recyclage.

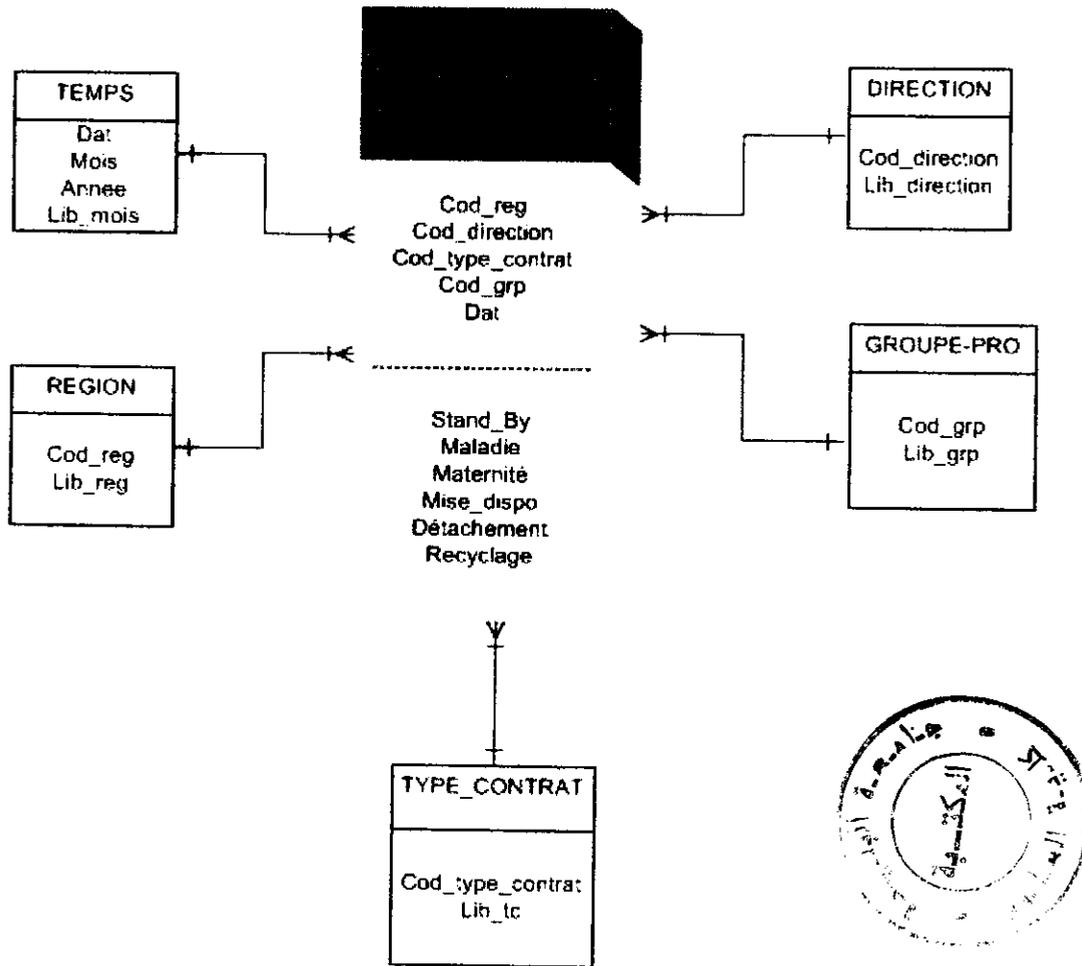


Fig -33- Table de fait Personnel Inactif

## 1.4. Modélisation dimensionnelle de l'activité

### « Mouvement personnel »

#### 1.4.1. Le processus d'activité

Il répertorie toutes les nouvelles recrues sur les différentes directions et régions, ce processus est très utile aux décideurs car il leur permet de cerner les besoins de l'ENGTP en matière de ressources humaines ainsi que son évolution à travers le temps.

#### 1.4.2. Le grain du processus d'activité

Au niveau de l'ENGTP les signatures de contrats se font aux débuts et fins de mois donc une granularité mensuel est la plus adaptée à cette activité.

### 1.4.3. Les dimensions

Cette activité est concernée par trois dimensions : La direction, La région et le temps.

### 1.4.4. Les faits mesurés

- Effectif arrivé.

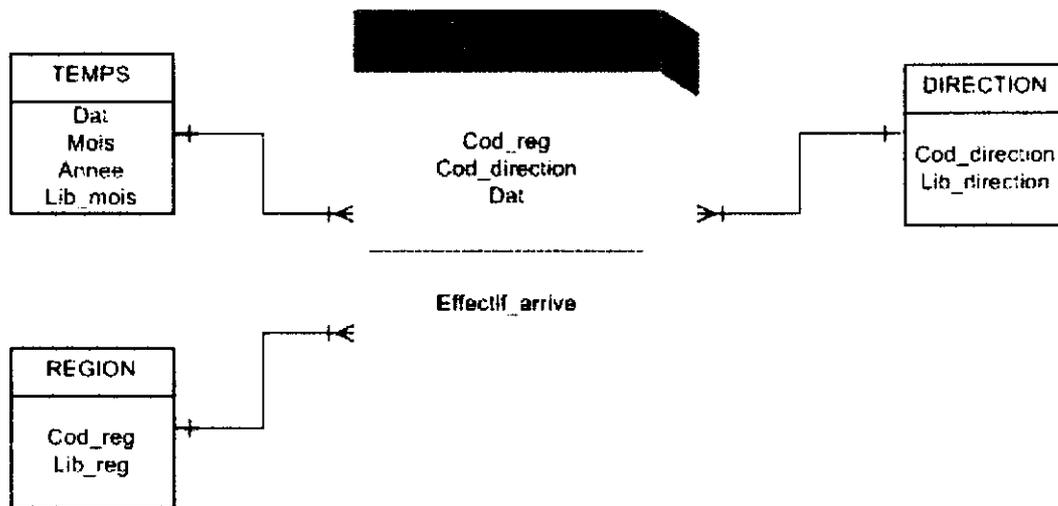


Fig -34- Table de fait Mouvement Personnel

## 1.5. Modélisation dimensionnelle de l'activité

### « Réalisation financière »

#### 1.5.1. Le processus d'activité

Sert au calcul de la masse salarial à partir des bulletins de pais mensuels et comme cette dernière représente la majorité du budget de l'entreprise on peut à partir de la décider des attributions de budget aux autres activités tels que la formation ...etc.

#### 1.5.2. Le grain du processus d'activité

Un grain mensuel est le plus adapté à cette activité vu les fluctuations que subit la masse salarial du au nombre de paramètres nécessaires à son calcul tels que les indemnités (Kilométrique, Panier), les primes (Bilan, Inventaires), les retenus, etc.

### 1.5.3. Les dimensions

Cette activité est concernée par quatre dimensions : La région, Le type contrat et le temps.

### 1.5.4. Les faits mesurés

- Masse salariale.

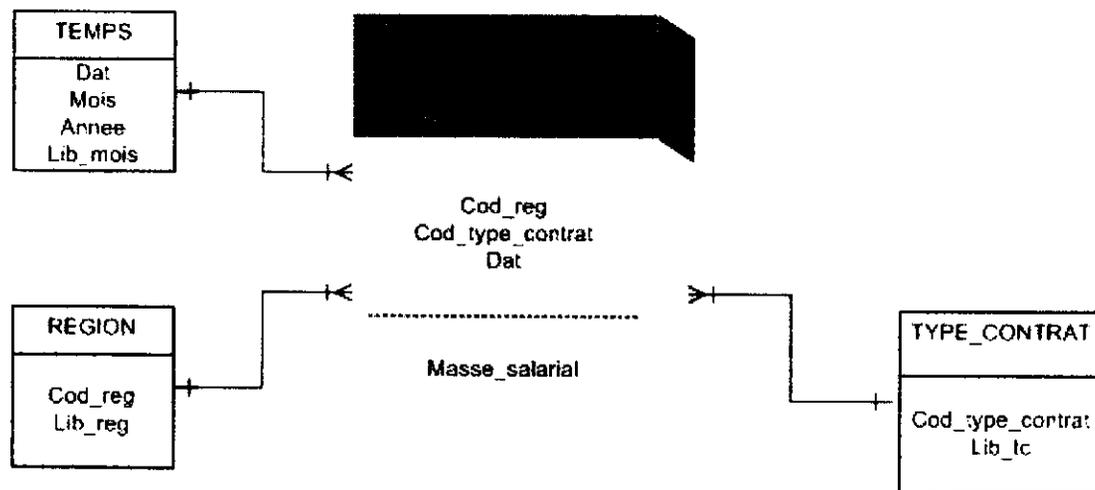


Fig -35- Table de fait Réalisation Financiere

## 2. Mise en œuvre de l'entrepôt

Nous allons passer dans ce chapitre à la mise en œuvre de notre entrepôt de données.

Cette phase comporte trois grandes étapes, chacune d'elle représente un composant de l'architecture de data warehouse qui sont:

- Une base de données et de méta-données constituant l'interface entre l'équipe informatique et la communauté des utilisateurs.
- Un ensemble de transformations simples et expressives appliquées aux données sources.
- La définition d'outils logiciels permettant de réaliser les traductions des données stockées au niveau de l'entrepôt en une information utile.

L'architecture de notre data warehouse est représentée par le schéma suivant, avec exposition des correspondances entre chaque composant et l'étape qui lui correspond dans le processus de mise en œuvre.

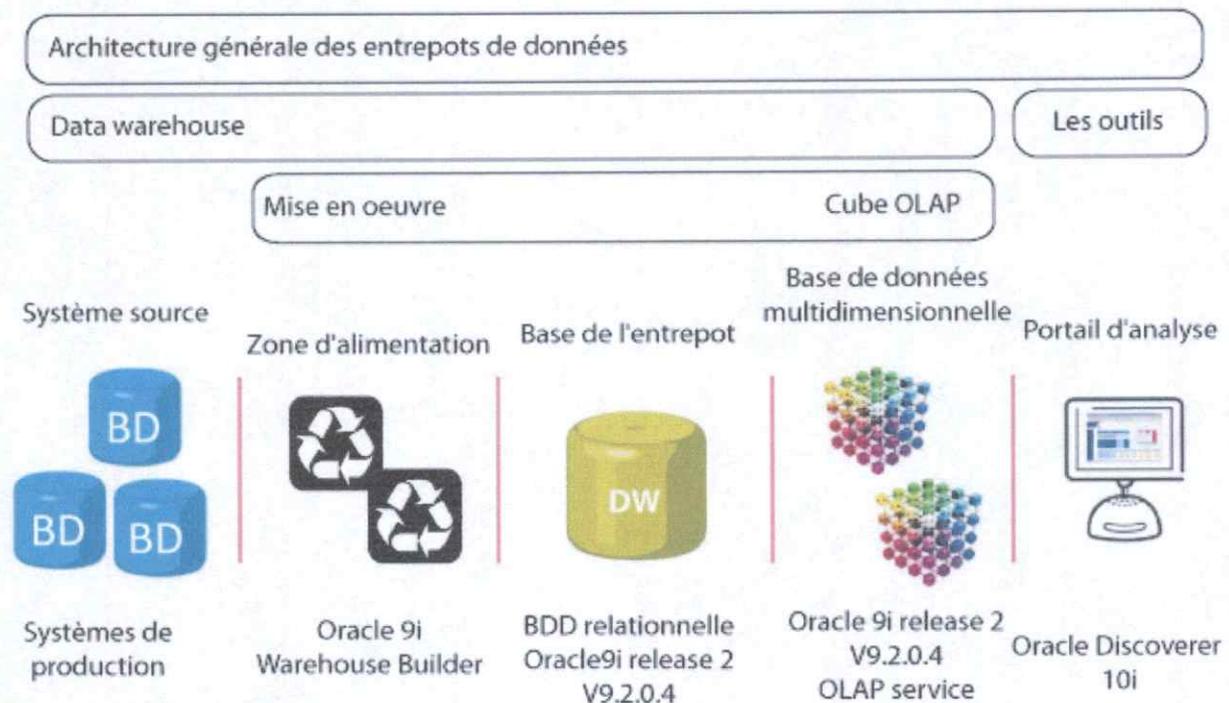


Fig -36- Architecture détaillée de notre data warehouse

- **ETL « Couche d'alimentation »** : Les données stockées dans l'entrepôt de données sont différentes de celles du système source en termes d'intégration, d'agrégation et de consolidation, pour cela le rôle de cette couche est de préparer et nettoyer les données du système source pour qu'elles puissent être stockées dans l'entrepôt de données.

L'outil que nous avons choisi est oracle9i warehouse builder.

- **Entrepôt de données** : C'est une base de données qui contient les données déjà alimentées à partir du système source par l'intermédiaire de la couche ETL. Ces données sont utilisées par la communauté des utilisateurs pour l'analyse. Cette composante regroupe une deuxième base de données qui est la base de métadonnées, cette base contient les données sur les données de l'entrepôt, comme le détail de chargement (La relation entre les objets sources et les objets d'entrepôt), la fréquence d'alimentation, etc.

Nous avons utilisé oracle 9i qui est basé sur le langage PL-SQL pour la conception de cette base.

## Partie 5 : Conception et mise en œuvre du système

- **Base OLAP** : Ou la base de données multi-dimensionnelle. L'objectif de cette couche est de stocker les agrégations des données de l'entrepôt dans des cubes multidimensionnels, ces cubes seront interrogés par les utilisateurs finaux. Notre serveur OLAP utilisé est Oracle 9i OLAP Service.

- **Portail de restitution** : C'est la couche visible aux utilisateurs. Elle contient tous les outils d'analyse et de reporting de données. Notre portail de restitution comporte deux applications, la première application de bureau sur la base d'oracle discoverer, et le deuxième via Intranet.

A présent, nous allons présenter en détail le processus de mise en œuvre de ces couches qui constituent l'architecture de notre système décisionnel.

### **2.1. Construction de la base de l'entrepôt**

Cette base répond à notre modèle de données, en d'autres termes c'est la transformation de la conception logique en une base de données physique.

Chaque entité dans notre modélisation, table de faits ou dimension, est transformée en une table de base de données relationnelle.

✓ Liste des tables qui constituent notre base décisionnelle

Table	Identifiant
MOUVEMENT_PERS	COD_DIRECTION, COD_REG, DAT
PERSONNEL_INACTIF	COD_DIRECTION, COD_GRP, COD_REG, COD_TYPE_CONTRAT, DAT
PERSONNEL_SORTANT	COD_GRP, COD_REG, MATRICULE, DAT
REALISATION_PERS	COD_ACTIVITE, COD_REG, COD_DIRECTION, COD_TYPE_CONTRAT, COD_GRP, TRANCHE, COD_SEXE, COD_QUALIFICATION, DAT
REALISATION_FINANCIERE	COD_REG, COD_TYPE_CONTRAT, DAT
ACTIVITE	COD_ACTIVITE, LIB_ACTIVITE
DIRECTION	COD_DIRECTION, LIB_DIRECTION
GROUPE_PRO	COD_GRP, LIB_GRP
PERSONNEL	MATRICULE, NOM, PRENOM
QUALIFICATION	COD_QUALIFICATION, LIB_QUAL
REGION	COD_REG, LIB_REG
SEXE	COD_SEXE, LIB_SEXE
TEMPS	DAT, MOIS, ANNEE, LIB_MOIS
TRANCHE_AGE	TRANCHE
TYPE_CONTRAT	COD_TYPE_CONTRAT, LIB_TC

Tab -7- Liste des tables de l'entrepôt de données

✓ Description des attributs :

La liste des attributs, leurs types de données et les tables à qui ils appartiennent sont résumés dans le tableau suivant :

Désignation	Code	Table	Type de données
Code Activité	COD_ACTIVITE	ACTIVITE	VARCHAR2(2)
Libellé Activité	LIB_ACTIVITE	ACTIVITE	VARCHAR2(50)
Code Direction	COD_DIRECTION	DIRECTION	VARCHAR2(6)
Libellé Direction	LIB_DIRECTION	DIRECTION	VARCHAR2(50)
Code Groupe	COD_GRP	GROUPE_PRO	VARCHAR2(2)
Libellé Groupe	LIB_GRP	GROUPE_PRO	VARCHAR2(50)
Matricule	MATRICULE	PERSONNEL	NUMBER(5)

Partie 5 : Conception et mise en œuvre du système

Nom	NOM	PERSONNEL	VARCHAR2(20)
Prénom	PRENOM	PERSONNEL	VARCHAR2(20)
Code Qualification	COD_QUALIFICATION	QUALIFICATION	NUMBER(10)
Libellé Qualification	LIB_QUAL	QUALIFICATION	VARCHAR2(50)
Code Région	COD_REG	REGION	VARCHAR2(2)
Libellé Région	LIB_REG	REGION	CHAR(50)
Code Sexe	COD_SEXE	SEXE	VARCHAR2(2)
Libellé Sexe	LIB_SEXE	SEXE	VARCHAR2(20)
Date	DAT	TEMPS	VARCHAR2(7)
Mois	MOIS	TEMPS	VARCHAR2(10)
Année	ANNEE	TEMPS	VARCHAR2(4)
Libellé Mois	LIB_MOIS	TEMPS	VARCHAR2(10)
Tranche d'âge	TRANCHE	TRANCHE_AGE	VARCHAR2(10)
Code Type Contrat	COD_TYPE_CONTRAT	TYPE_CONTRAT	NUMBER(10)
Libellé Type Contrat	LIB_TC	TYPE_CONTRAT	VARCHAR2(50)
L'effectif arrivé	EFFECTIF_ARRIVE	MOUVEMENT_PERS	NUMBER(10)
Stand By	STAND_BY	PERSONNEL_INACTIF	NUMBER(10)
Maternité	MATERNITE	PERSONNEL_INACTIF	NUMBER(10)
Détachement	DETACHEMENT	PERSONNEL_INACTIF	NUMBER(10)
Recyclage	RECYCLAGE	PERSONNEL_INACTIF	NUMBER(10)
Mise en Dispo	DISPO	PERSONNEL_INACTIF	NUMBER(10)
Admis à la retraite	ADMIS_RETRAITE	PERSONNEL_SORTANT	NUMBER(10)
Licencié	LICENCIE	PERSONNEL_SORTANT	NUMBER(10)
Démission	DEMISSION	PERSONNEL_SORTANT	NUMBER(10)
Fin de relation de travail	FRT	PERSONNEL_SORTANT	NUMBER(10)
La masse salariale	MASSE_SALARIAL	REALISATION_FINANCIERE	NUMBER(10)
Effectif actif	EFFECTIF_ACTIF	REALISATION_PERS	NUMBER(10)
Effectif opérationnel	EFFECTIF_OPERATIONNEL	REALISATION_PERS	NUMBER(10)
Effectif en formation	EFFECTIF_FORMATION	REALISATION_PERS	NUMBER(10)
Effectif payé	EFFECTIF_PAYE	REALISATION_PERS	NUMBER(10)

Tab -8- Liste des attributs de l'entrepôt de données

## 2.2. Construction des cubes OLAP

La modélisation multidimensionnelle des données constitue une façon de faciliter l'analyse commerciale en ligne ainsi que les performances des requêtes. Le Gestionnaire OLAP nous permet de transformer les données stockées dans des bases de données relationnelles en informations commerciales pertinentes et faciles à exploiter, grâce à la création d'un cube de données.

Dans cette étape, nous allons concevoir notre base de données multidimensionnelle (cubes OLAP).

La technique de modélisation utilisée pour la mise en œuvre des cubes OLAP est la modélisation MOLAP « Multidimensional OLAP », puisqu'elle est facile à mettre en place en plus de faciliter l'analyse multi-axes sur les données agrégées. Le serveur OLAP utilisé pour notre travail est Oracle 9i -OLAP Service-.

- **Concept de fait** : Le fait modélise le sujet de l'analyse. Un fait est formé de mesures correspondantes aux informations de l'activité analysée.

- **Concept de dimension** : Une dimension modélise une perspective de l'analyse. Une dimension se compose de paramètres correspondant aux informations faisant varier les mesures de l'activité.

Le tableau suivant résume les tables des faits choisies avec leurs mesures et leurs dimensions d'analyse sollicitées.

Partie 5 : Conception et mise en œuvre du système

Table de fait	Mesures	Dimensions sollicité
Réalisation personnel	Effectif actif Effectif payé Effectif en formation Effectif opérationnel	Activité Région Direction Type Contrat Groupe professionnel Tranche d'âge Sexe Qualification
Personnel Sortant	Effectif admis à la retraite FRT Effectif licencié Effectif démissionnaire	Groupe professionnel Personnel Région Temps
Personnel Inactif	Effectif en Stand by Effectif en maternité Effectif mis en disponibilité Effectif détaché Effectif en recyclage Effectif en maladie	Direction Groupe Région Type Contrat Temps
Mouvement personnel	Effectif arrivé	Direction Région Temps
Réalisation financière	Masse salarial	Région Type Contrat Groupe professionnel Temps

**Tab -9-** Liste des faits et des dimensions sollicitées

➤ **Formalisme utilisé**

- **Concept de fait** : Nous représentons le fait par un cube englobant les différentes mesures d'activité qu'il contient. En outre le symbole d'un cube estampille.

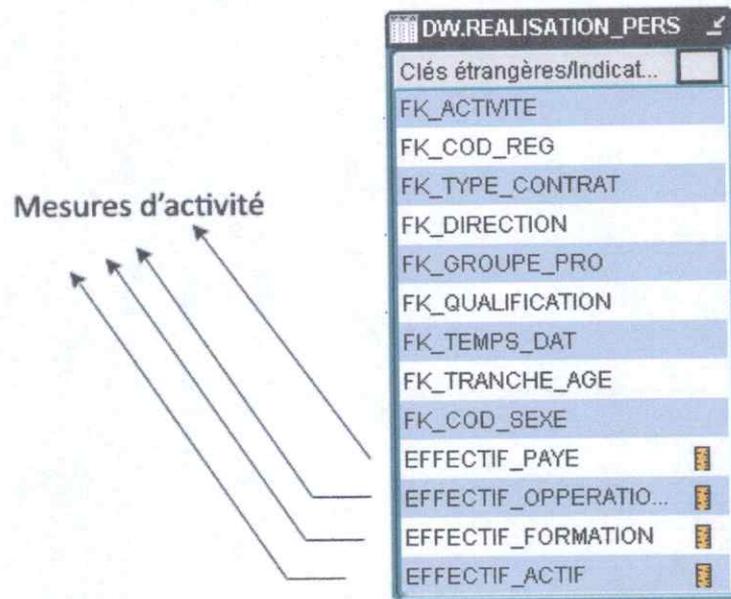


Fig -37- Mesures d'activités

- **Concept de dimension** : Nous représentons la dimension par un rectangle englobant les différents paramètres d'analyse qu'il contient. En outre le symbole représentant trois axes estampille les dimensions pour les distinguer du fait.

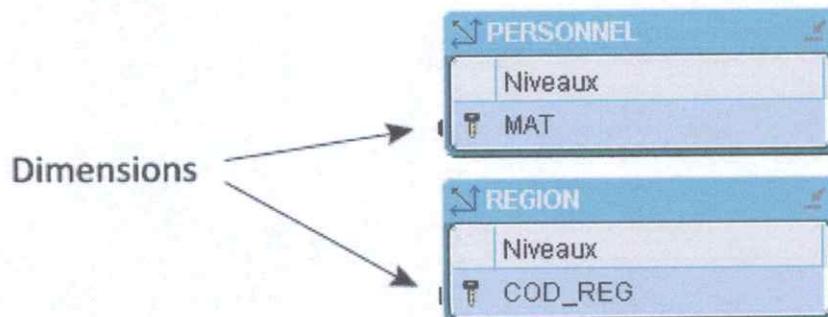


Fig -38- Dimension

➤ Schéma de modèle MOLAP des cubes

Les figures suivantes illustrent le modèle physique de chaque cube MOLAP selon la plateforme Oracle 9i OLAP Service. Chaque cube correspond à une activité précise.

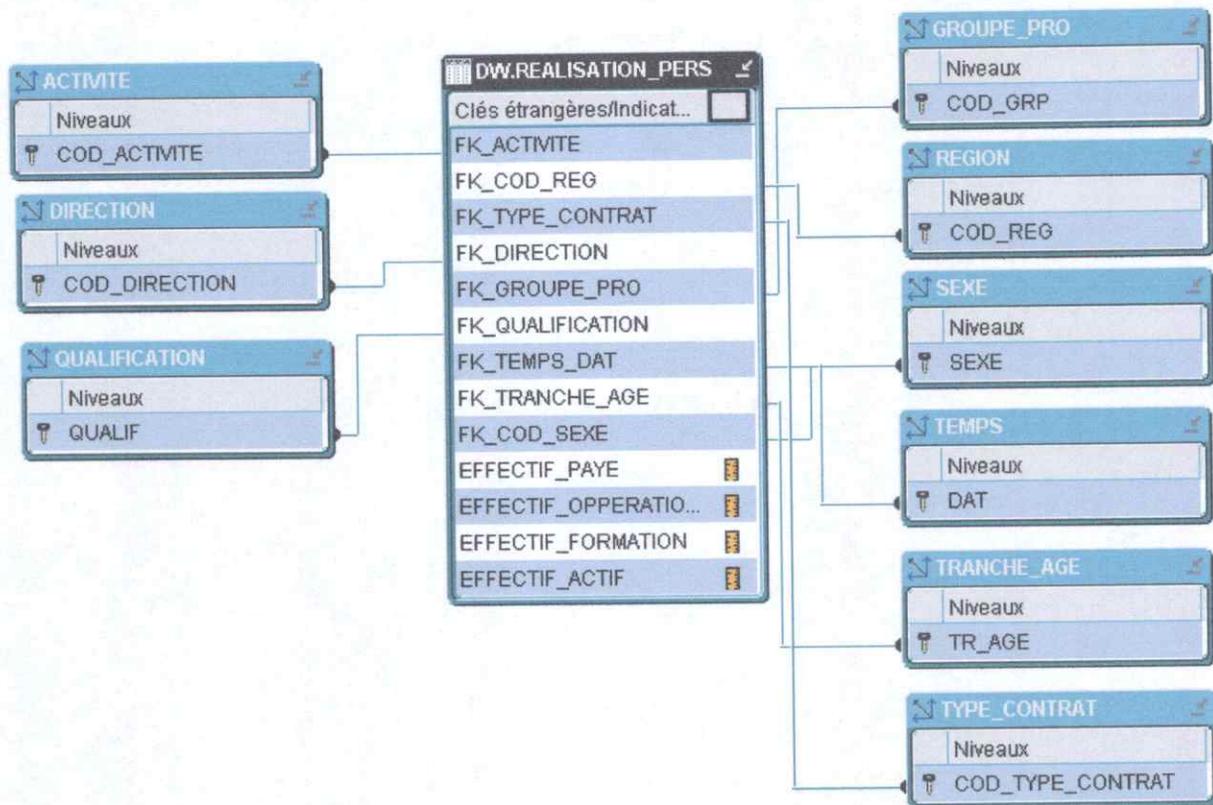


Fig -39- Cube MOLAP de l'activité Réalisation Personnel

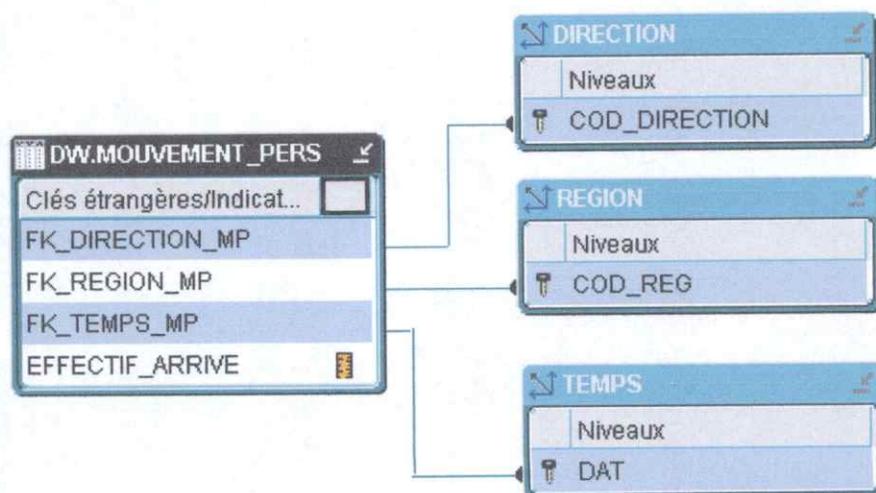


Fig -40- Cube MOLAP de l'activité Mouvement Personnel

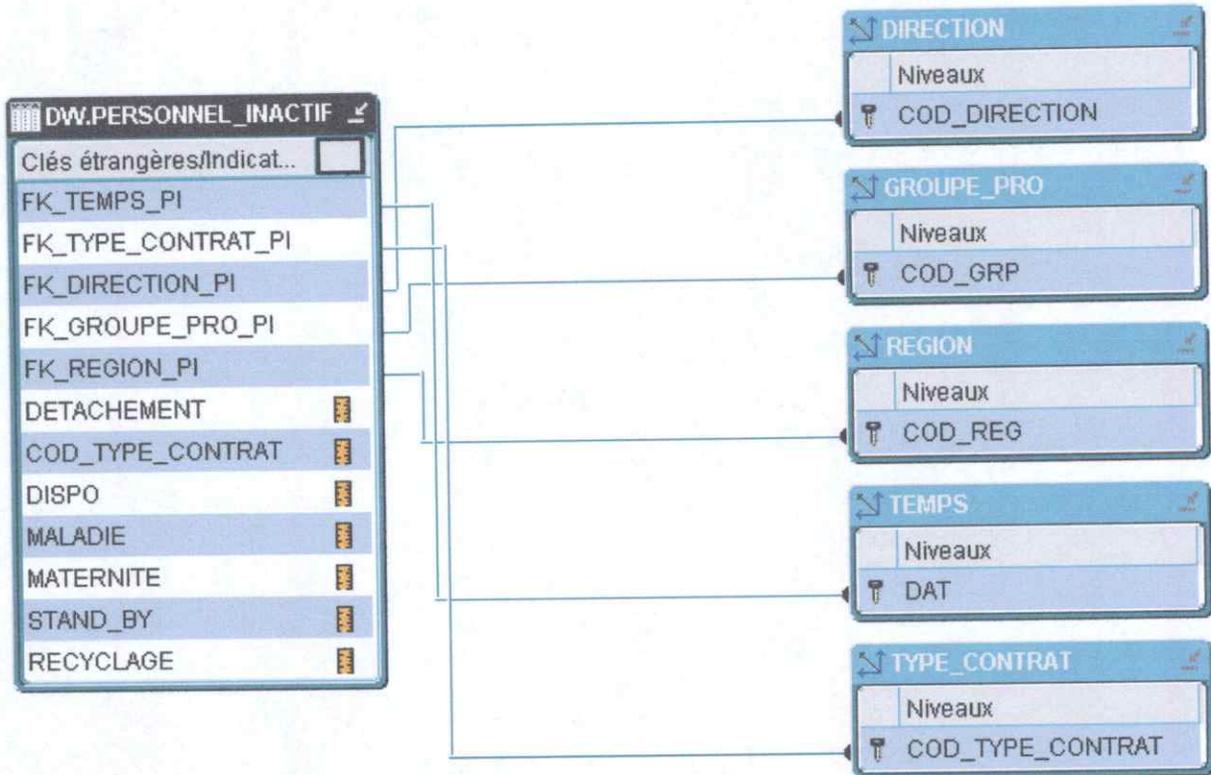


Fig -41- Cube MOLAP de l'activité Personnel Inactif

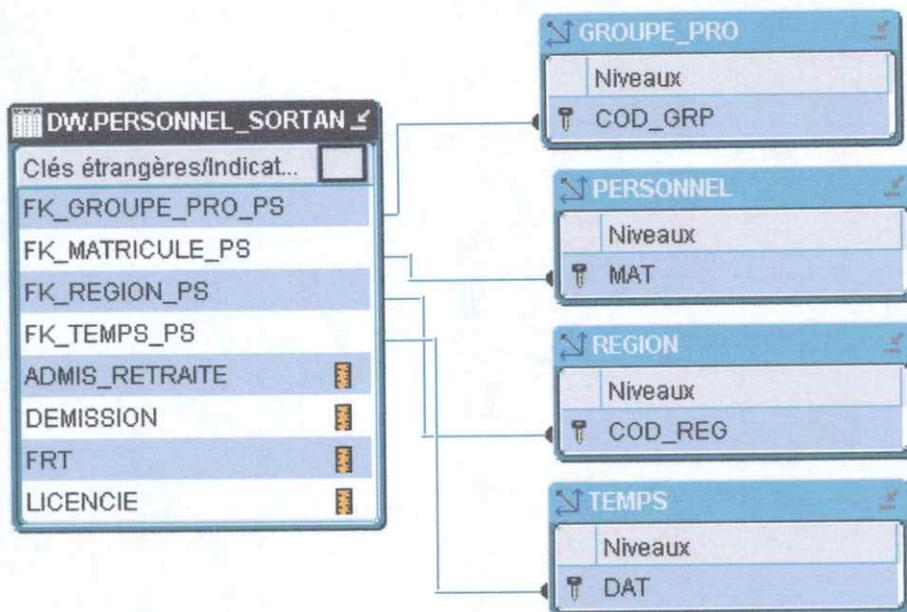


Fig -42- Cube MOLAP de l'activité Personnel Sortant

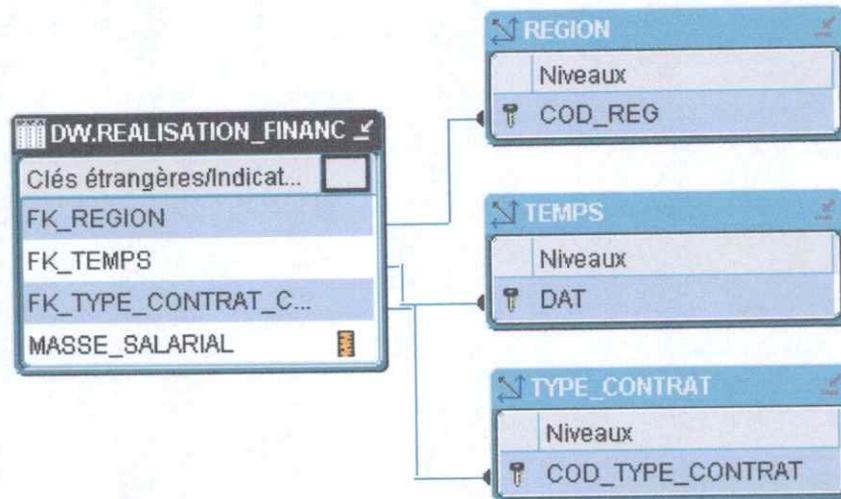


Fig -43- Cube MOLAP de l'activité Réalisation Financière

### 2.3. Construction de la zone d'alimentation

Cette opération permet de transférer les données du système opérationnel vers l'entrepôt de données, elle nécessite la construction de méta-données qui expliquent les flux et les transformations de données.

Notre processus d'alimentation se déroule en trois étapes nécessaires. L'extraction depuis le système source, le regroupement des données en un seul ensemble et en fin le chargement dans l'entrepôt de données. Pour réaliser ces étapes, nous avons besoin d'une base de données qui stocke les informations sur les données. Il s'agit d'une métabase qui nous décrit les données que doit contenir l'entrepôt de données, en donnant où et comment extraire les données nécessaires. Ainsi, un changement de type de données simples (à gauche dans les BDD) ne nécessite en fait qu'un léger changement dans l'ETL. (Ces changements sont : des calculs d'agrégation, des changements de format ou de présentation, des jointures ...).

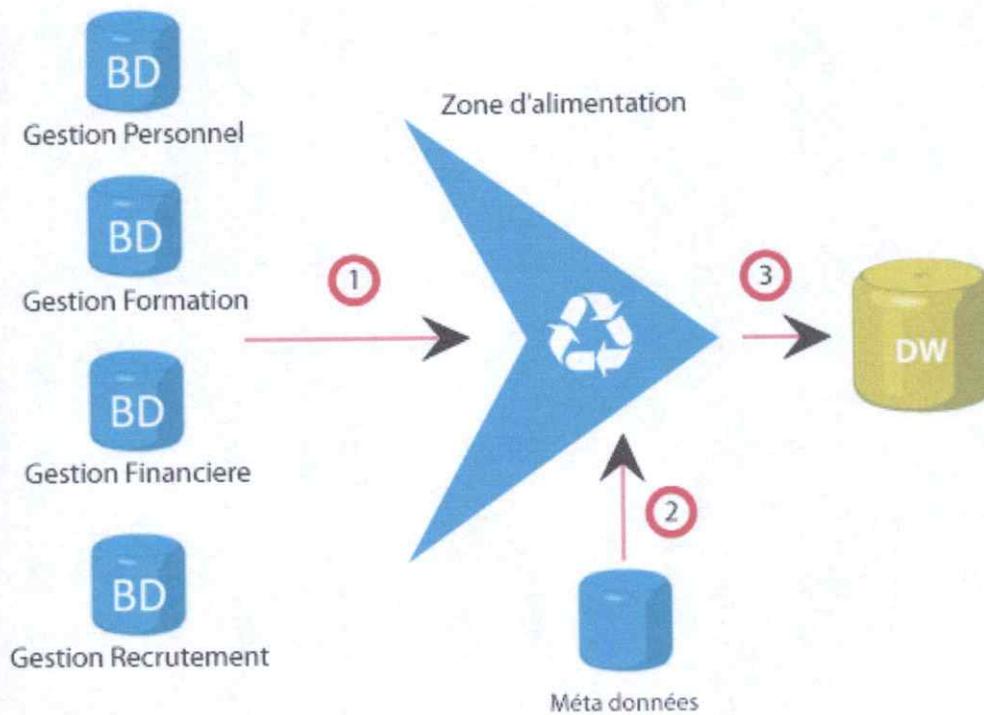


Fig -45- Le processus d'alimentation de notre entrepôt

1 : Extraction des données.

2 : Transformation des données

3 : Chargement des données.

L'opération de transformation se fait à l'aide d'un outil appelé "Oracle 9i Warehouse Builder".

Cet outil permet de modéliser graphiquement le processus de transformation de données, quelque soit la source de données.

Il possède aussi des fonctions qui permettent de transformer les données issues des systèmes sources avant chargement dans l'entrepôt de données (to\_char, to\_number, concat, substr, etc...) vu que les données sources sont hétérogènes.

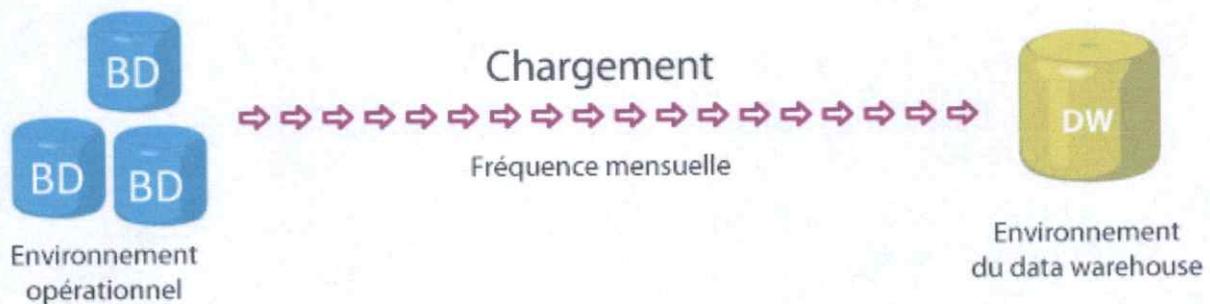
La première étape consiste à générer un mapping (tracer) qui va contenir les tables sources des différents systèmes et tables cibles de l'entrepôt avec les données après extraction.

La deuxième étape est la transformation, ce sont en réalité des transformateurs qui manipulent les données (agrégations, filtres, conversions...), par exemple : Nous avons la date (temps) du schéma personnel qui est sous la forme « Année/Mois » et qui par souci de lisibilité nous avons opté pour le format « Mois/Année ».

Et enfin, la dernière étape qui est le chargement des données dans l'entrepôt et l'automatisation d'un job qui va permettre l'alimentation des dimensions et des cubes dans un modèle en étoile sur une période donnée, vu que les données sources sont mises à jours chaque mois. Notre job a été programmé mensuellement.

Notre analyse sera suivie par l'hierarchie suivante: Année → Mois.

Donc l'opération de transformation de la source vers l'entrepôt sera exécutée chaque début du mois. La fréquence d'alimentation est illustrée par la figure suivante.



**Fig -46-** La fréquence de chargement des systèmes sources vers l'entrepôt

## 2.4. La construction du portail de restitution

Après avoir présenté les aspects relatifs aux outils d'arrière plan de notre système décisionnel, ainsi que la conception et le développement de la base de données. Nous allons maintenant passer à la mise en œuvre d'une application qui permettra d'exploiter les données de notre entrepôt ou des cubes OLAP et de présenter l'information dans des formats utiles pour que les gens puissent accéder, comprendre et utiles.

Cette partie est l'unique point d'entrée visible pour le consommateur de l'information. Pour cela, notre portail de restitution va lui permettre de manipuler les données à l'intérieur du cube.

Il faut signaler que les utilisateurs ont des besoins différents tels que :

- Des besoins d'interrogation classique aux fins de retrouver des rapports préalablement établis sur un thème défini.

## Partie 5 : Conception et mise en œuvre du système

- Des besoins de consultation évolués (Analyse OLAP multidimensionnelle dans le cadre d'application d'aide à la décision). En mettant en réalisation les indicateurs de performances avec les dimensions, permettant de voir l'impact d'une décision.

Notre portail de restitution va dépendre forcément des besoins des consommateurs, donc l'utilisation des informations de notre data warehouse va se baser sur deux modes différents :

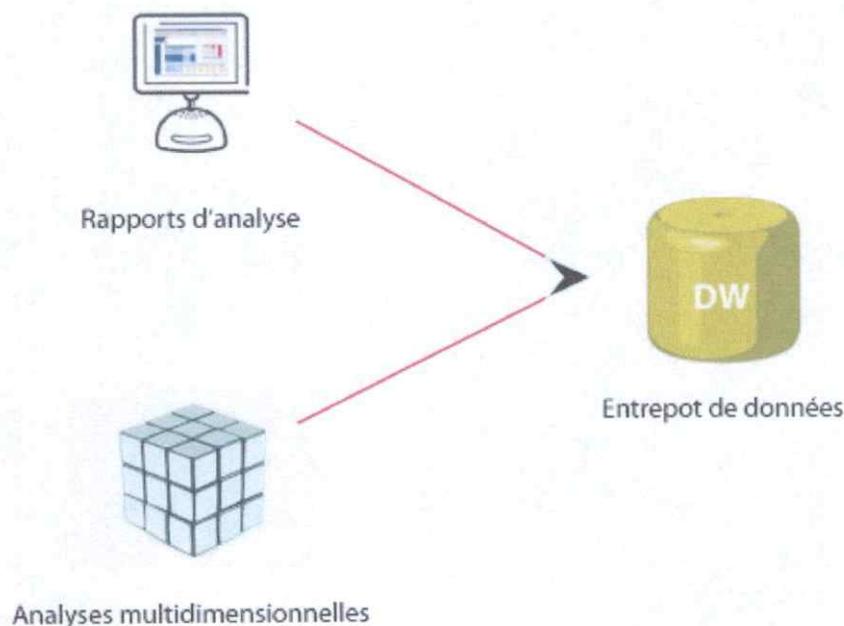
### ➤ Le reporting :

Il s'agit de mettre en œuvre une interface utilisateur efficace, et les informations y sont préstructurées sous la forme de tableaux de bord ou d'états prédéfinis, ces rapports sont diffusés sur l'Intranet de l'entreprise.

### ➤ L'analyse OLAP :

Il s'agit de réaliser un outil d'analyse multidimensionnelle qui offre aux utilisateurs le pouvoir de naviguer dynamiquement dans le catalogue des méta-données afin de réaliser des états d'analyse sous différents formats (graphes, tables, matrices,...) et d'exporter les états vers une autre cible (HTML, Fichier texte, Excel,...).

Notre outil multidimensionnel permettra de naviguer à l'intérieur d'un cube par l'intermédiaire de Drill, Pivoting ou Slice and Dice



**Fig -47-** Les outils d'accès à notre data warehouse

### **3. Conclusion**

Dans cette partie, nous avons pu modéliser notre data warehouse en commençant par une conception logique qui a intégré le concept de faits et de dimensions, en utilisant la modélisation dimensionnelle qui représente l'information d'une manière adéquate aux processus d'analyse et les technologies du système OLAP.

Le modèle de données de notre entrepôt est basé sur un schéma en étoile comportant plusieurs faits et une hiérarchie ainsi détaillée des dimensions, ces dimensions peuvent être les mêmes dans plusieurs tables de faits.

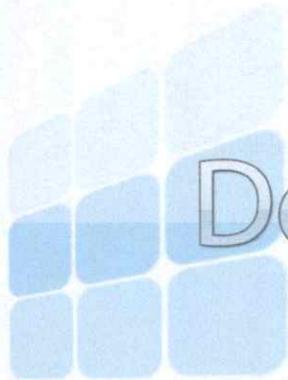
Dans ce chapitre nous avons aussi présenté l'architecture type d'un système décisionnel basé sur l'approche entrepôt de données, cette architecture est divisée en trois composantes essentielles (ETL, Données, Portail de restitution).

Nous avons construit une base multidimensionnelle puissante qui supporte toutes les requêtes d'analyse.

Dans ce qui suit, nous allons présenter le déploiement de notre système. Le déploiement est la spécification de la disposition physique de notre système décisionnel, en spécifiant l'architecture logicielle utilisée, le matériel de l'infrastructure réseau, la sécurité des utilisateurs du système.



Partie .6



Déploiement

## Partie 6 : Déploiement du système

Après avoir conçu notre data warehouse, nous allons passer au déploiement du prototype que nous avons réalisé en intégrant les fonctionnalités des cubes OLAP et les avantages de l'architecture multidimensionnelle.

La mise en place de notre système dépend donc de la plate-forme existante, l'architecture technique est présentée ci-dessous.

### **1. Architecture technique de notre système**

L'entreprise GTP implémente au niveau de tous ses départements un réseau Intranet basé sur la plate-forme Microsoft (Windows NT4 server, Oracle9i comme serveur de base de données et Exchange 2000 server comme serveur de messagerie électronique).

Ce réseau Intranet sera étendu vers un réseau Extranet pour objectif de regrouper les autres structures en un seul réseau ouvert.

Notre système décisionnel est destiné à deux types d'utilisateurs selon leurs besoins :

Le premier type englobe les analystes et les décideurs qui ont un accès ad hoc au data warehouse pour naviguer librement à l'intérieur du cube et réaliser des états et des présentations souhaités, l'architecture informatique que nous avons adopté pour cette application est une architecture deux tiers.

Le deuxième type d'utilisateurs est l'ensemble des statisticiens qui auront un accès via l'Intranet de l'entreprise à des rapports d'analyse préalablement définis.

## 2. Disposition physique de notre système

Notre système décisionnel est basé sur l'architecture illustrée dans la figure suivante.

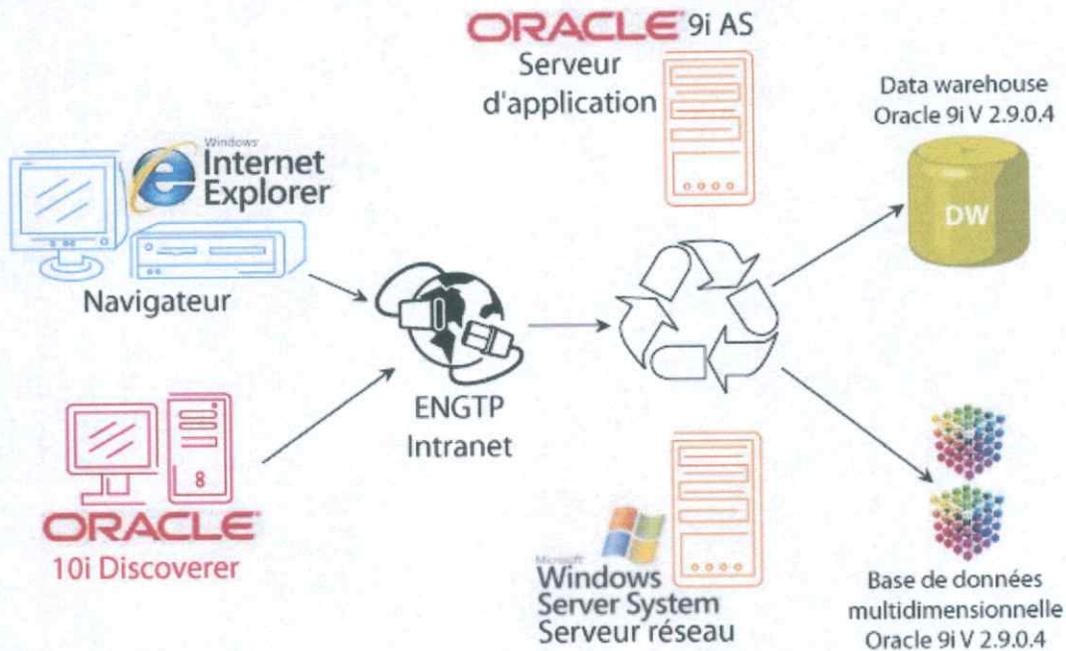


Fig -48- Disposition physique de notre système

Cette architecture est divisée en trois couches essentielles : présentation, métiers (traitements) et stockage (données), le détail de ces couches est présenté dans ce qui suit.

- ❖ **La présentation** : Deux types de clients implémentés dans cette couche, le premier est un simple navigateur, le rôle de ce client est d'accéder à l'information à travers le portail décisionnel développé, Le deuxième client est une application développée à travers un outil Oracle discoverer. cette application sera installée et configurée sur le poste client.

L'objectif final de cette couche est de présenter les données stockées aux serveurs de données en informations utiles et simples à analyser.

- ❖ **Le traitement** : Cette couche regroupe en général tous les processus (Objets) métiers des applications de l'entreprise (La logique métier). La connexion aux serveurs de données, aux serveurs de sécurité et gestion des requêtes utilisateurs sont des processus de base dans un système informatique de l'entreprise.

## Partie 6 : Déploiement du système

Elle contient dans notre cas le serveur de sécurité, le serveur réseau et le serveur d'application.

Comme la première, 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 Oracle9i Server pour l'entrepôt de données et le serveur de base de données multi-dimensionnelle Oracle9i OLAP Service.

### **3. Les utilisateurs du système**

Notre système est utilisé par deux catégories d'utilisateurs. La première catégorie est l'ensemble des analystes y compris les décideurs, c'est l'ensemble des utilisateurs le plus intéressant par le déploiement d'un système d'aide à la décision basé sur les dimensions, leur objectif est l'analyse multidimensionnelle sur les activités de l'entreprise. La deuxième catégorie des utilisateurs est l'ensemble des statisticiens, leurs besoins informationnels ont pour objectif d'avoir un aperçu sur les états d'analyse des activités de base de l'entreprise.

La figure suivante présente les deux catégories des utilisateurs de notre système, avec leurs caractéristiques.

### **4. Le catalogue des états**

Ce catalogue contient l'ensemble des états et des présentations réalisés et enregistrés via l'application de bureau (Discoverer). Ces états analytiques sont accessibles par deux manières. La première est avec l'application de bureau en mode lecture écriture et la deuxième par Intranet en mode lecture seule.

La technologie qu'on a adoptée pour implémenter ce catalogue est XML DB de Oracle9i release2, qui permet d'enregistrer des fichiers légers de type XML dans la base de données. Les fichiers XML décrivent bien ces états en sauvegardant leurs requêtes, leurs types et leurs formats.



## **5. Logiciels utilisés**

### **5.1. Les composants serveurs**

- Microsoft Windows NT4 server comme serveur de sécurité et serveur réseau.
- Oracle9i Entreprise Edition version 9.2.0.4 comme serveur de données qui implémente le data warehouse.
- Oracle9i Entreprise Edition version 9.2.0.4 OLAP Service comme serveur de base de données multidimensionnelle qui implémente les cubes OLAP.

### **5.2. Les composants clients**

Les deux types d'application que nous avons réalisée, orientés vers l'analyse de données ou vers le reporting, sont utilisés comme clients de notre système décisionnel, il suffit de les configurer sur le réseau Intranet. Ces deux applications sont : l'analyseur OLAP Oracle discoverer 10i installée sur le poste de l'utilisateur, et le reporting accessible via l'Intranet selon le protocole HTTP.

## **6. Matériel utilisé**

Le matériel utilisé pour le déploiement de notre système est basé sur la plate-forme Intel biprocesseur avec un contrôleur RAID pourvu d'un lecteur des bandes pour la sauvegarde des données et contient 6 disques durs de capacité 18 Giga Octets pour chacun.

Ce serveur sert à héberger la base de données du data warehouse, la base de données multidimensionnelle et le catalogue des états d'analyse.

## **7. La sécurité du système**

L'information tend à devenir l'un des biens les plus précieux de l'entreprise, sa destruction ou bien sa perte d'intégrité entraîne des conséquences graves au fonctionnement du système d'information.

La qualité de fonctionnement d'un système d'information suppose au préalable une prévention et une protection suffisante et cohérente vis-à-vis à des risques encourus, il est impératif que ce système projeté doit être en mesure de se protéger contre les indiscretions

## Partie 6 : Déploiement du système

et les destructions. Ceci n'est possible que par la mise en place d'un plan de sécurité rigoureux afin de minimiser les risques.

En général, deux aspects de sécurité possible : la sécurité physique et la sécurité logique. La première contre les dégâts matériels (Machine tombe en panne) et la deuxième contre les dégâts logiciels (Virus, Perte de mot de passe, etc.).

Notre approche de sécurité pour les deux types de pannes est basée sur les normes standards.

**Sécurité physique** : les données de l'entrepôt et de la base multidimensionnelle sont sauvegardées périodiquement dans des bandes magnétiques de grande capacité.

**Sécurité logique** : consiste à protéger les données logiquement contre les dégâts provoqués par des logiciels tiers (par exemple les virus, accès non autorisé). Notre solution de sécurité est basée sur la sécurité du système d'exploitation et la sécurité du SGBD pour les accès à notre système décisionnel.

Notre outil multidimensionnel permettra de naviguer à l'intérieur d'un cube par l'intermédiaire de Drill, Pivoting ou Slice and Dice.

## 8. Navigation

- Un **Drill-up** Permet de réaliser des cumuls ou des consolidations d'indicateurs en agrégeant les données suivant un axe d'analyse, ceci en naviguant dans les dimensions pour mieux affiner les résultats.

Eléments Page :			
	Annee	Date	Effectif Arrive
1	06	01/06	3
2	07	01/07	58
3	08	01/08	655
4	72	01/72	1

Eléments Page : Année: 08 ▼			
	Cod Reg	Date	Effectif Arrive
1	0	01/08	34
2	2		76
3	3		128
4	4		280
5	5		137

Fig -49- Drill-up

## Partie 6 : Déploiement du système

- Un **Drill-down** Permet de naviguer à l'intérieur de la dimension et détailler l'analyse en affinant la granularité de l'axe d'analyse. par exemple en passant de l'effectif arrivé par an vers l'effectif arrivé par mois.

Eléments Page : Année: 08 ▼		
	<b>Date</b>	<b>Effectif Arrive</b>
1	01/08	655
2	02/08	474
3	03/08	15

Fig -50- Drill-down

- Un **Slice and Dice** Permet d'intervenir deux axes d'analyse.  
(De année par région vers région par année)

Eléments Page : Region: ARZEW				✓ ARZEW
	<b>Annee</b>	<b>Date</b>	<b>Effectif Arrive</b>	ECOLE DE SOUDURE ET D'EXPERTISE D'ARZEW
1	08	01/08	76	HASSI MESSAOUD
2	75	01/75	1	HASSI R'MEL
3	76	01/76	1	REGHAIA
4	77	01/77	2	SKIKDA
5	78	01/78	1	

Fig -51- Slice and Dice

- Un **Pivoting** Permet de transposer deux colonnes par un simple glissé-déplacer.  
Exemple : On peut faire une duplication de notre table sous forme de matrice avant de faire un Pivoting.

## Partie 6 : Déploiement du système

Eléments Page : **Region: ECOLE DE SOUDURE ET D'EXPERTISE D'ARZEW**

	05	06	07	08	77	78	80	81	82	83	85	86	87
<b>01/07</b>			1										
<b>01/78</b>						1							
<b>02/08</b>				4									
<b>02/78</b>						4							
<b>02/81</b>								1					

↓

Eléments Page : **Region: ECOLE DE SOUDURE ET D'EXPERTISE D'ARZEW**

	01/07	01/78	02/08	02/78	02/81	04/05	04/07	05/07	05/80	05/83	06/78	08/80	09/06	09/07	09/81	09/87	10/07	10/82
05						1												
06													1					
07	1						2	2						3			1	
08			4															
77																		

Fig -52- Pivoting

Les illustrations suivantes décrivent l'interface destinée à l'utilisateur final ainsi qu'une panoplie d'options qui sont utile à l'analyse.

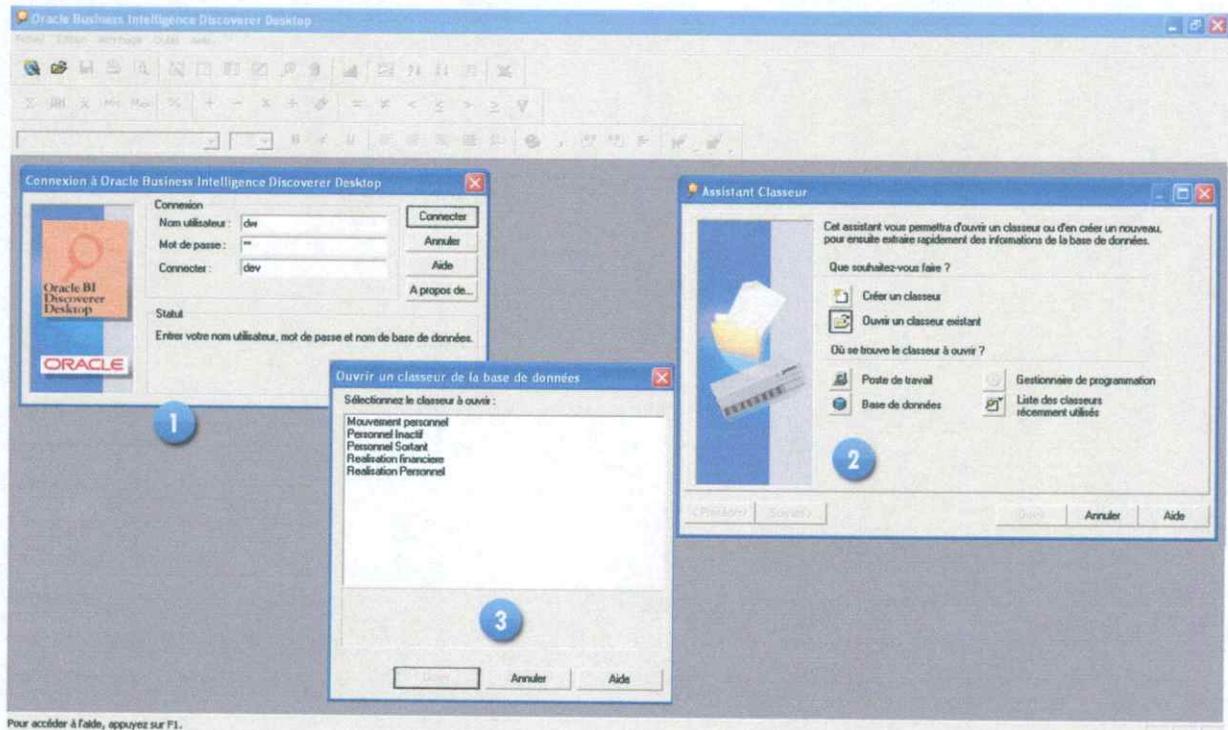


Fig -53 - Connexion au schéma et choix du classeur

1. Insertion des informations de connexion au schéma (Nom Utilisateur, Mot de passe, choix du schéma).

## Partie 6 : Déploiement du système

2. Choix de la source.
3. Choix du classeur.

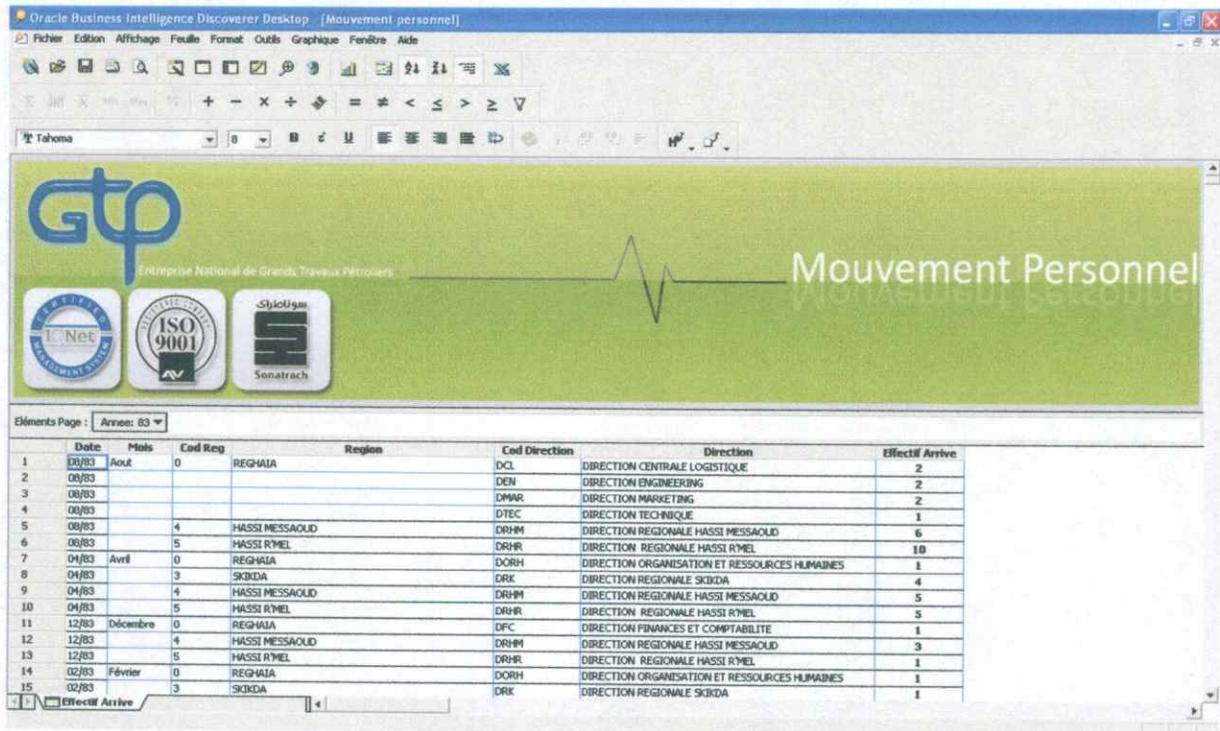


Fig -54- Affichage classeur « Mouvement Personnel »

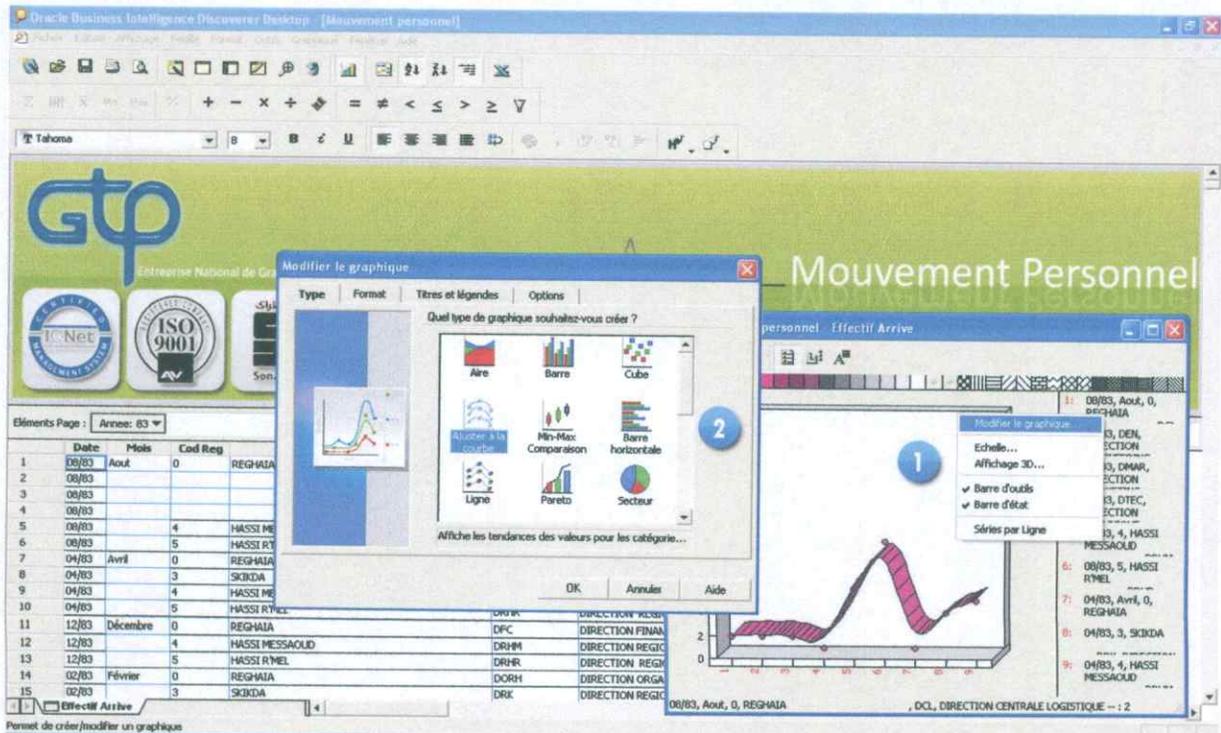


Fig -55- Choix du type de graphe

## Partie 6 : Déploiement du système

1. Cliquez droit sur le graphe.
2. Choix du type de graphe

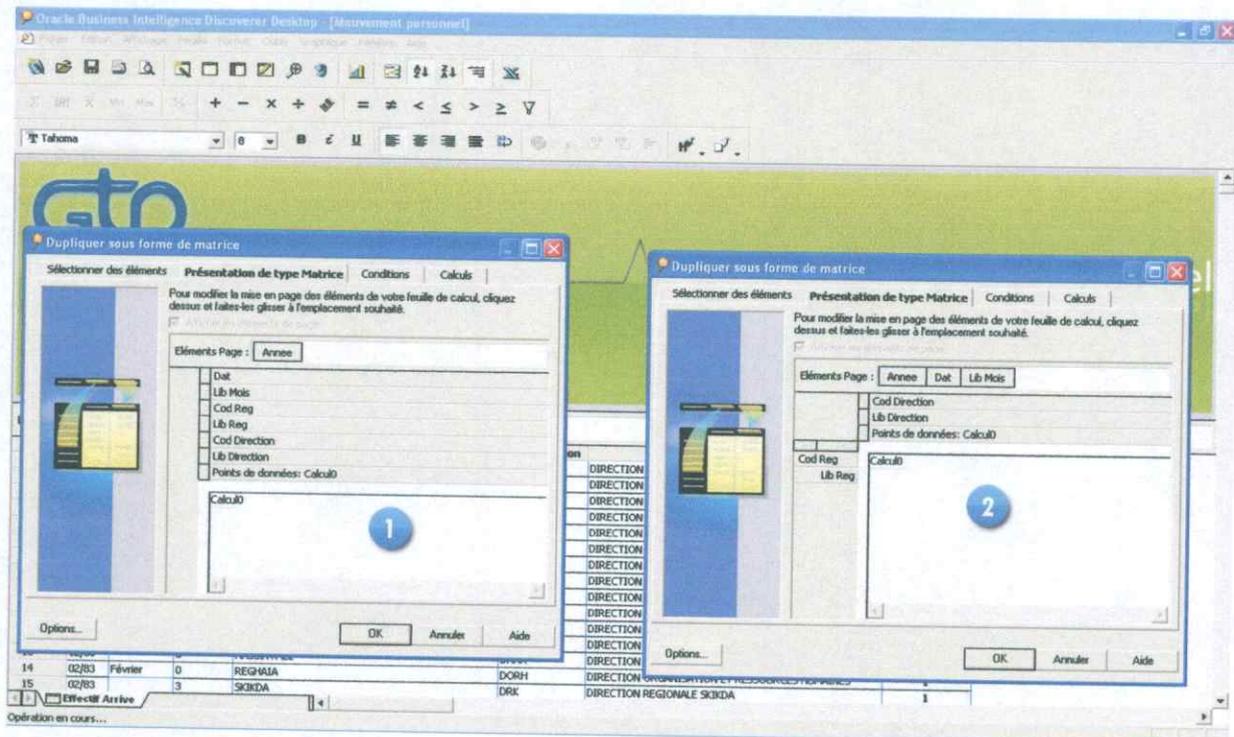


Fig -56- Pivot des axes

1. Après avoir choisi « dupliquer au format matrice ».
2. Pivot des dimensions (axes d'analyse).

## Partie 6 : Déploiement du système

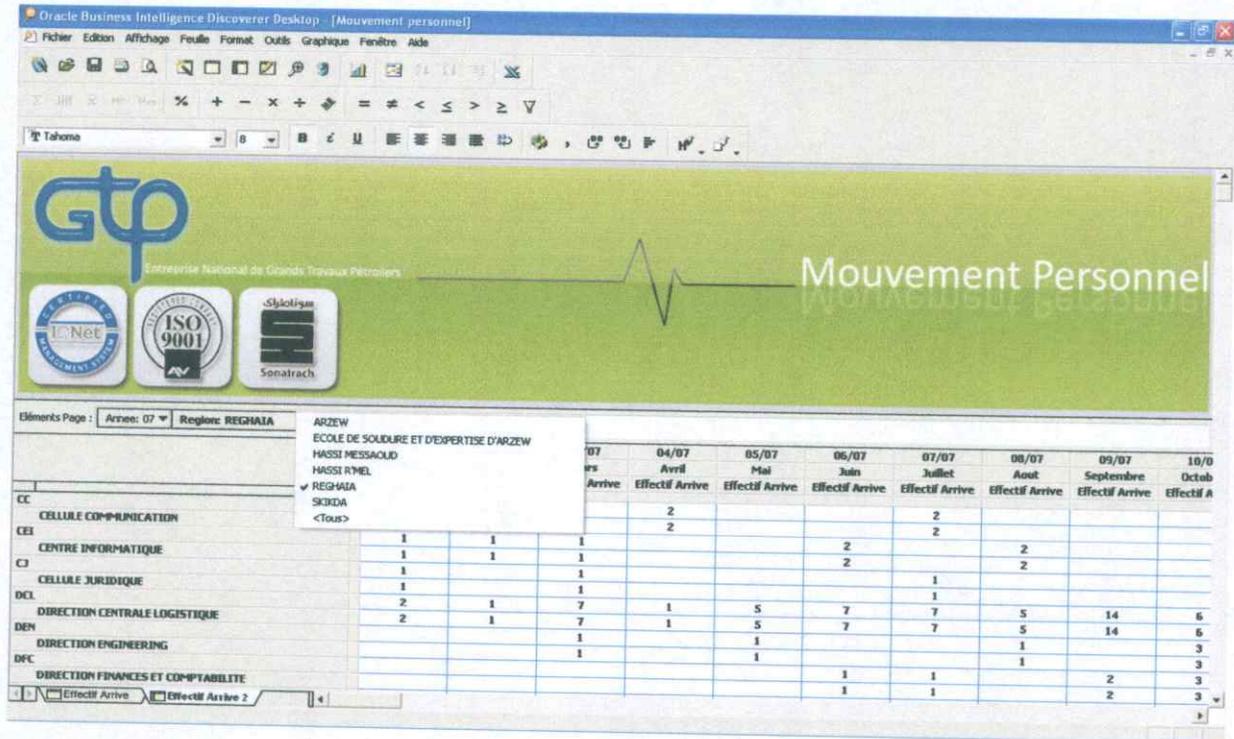
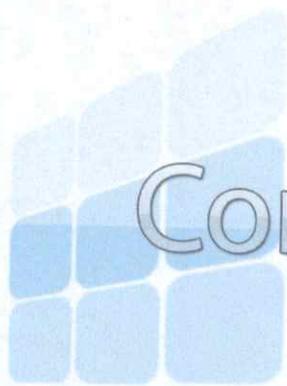


Fig -57- Affichage en mode matricielle.



Partie .7



# Conclusion générale

## Conclusion Général

Les avantages d'un système d'aide à la décision sont de donner aux collaborateurs, partenaires et fournisseurs les informations dont ils ont besoin pour gagner en efficacité, anticiper les décisions stratégiques, optimiser le pilotage de l'entreprise, séduire et fidéliser les clients en leur permettant d'accéder aux informations pertinentes. Conscient de tous ces avantages l'ENGTP nous a offert la chance de réaliser un projet de cette envergure que nous avons entamé de la manière suivante :

- ✓ Identification des besoins, ceci grâce aux entretiens établis, cette étape englobe tous les besoins en métier de pilotage et d'analyse de l'activité suivi des projets.
- ✓ Conception du système, qui elle même se décompose en trois processus qui sont :
  - La conception du modèle multidimensionnel qui regroupe les dimensions et les différents indicateurs associés pour chaque activité, sa modélisation dans des cubes en utilisant oracle 9i OLAP service.
  - La mise en place de la zone d'alimentation qui est à déclenchement automatique sur une durée prédéterminée (un mois dans notre cas), ce processus récupère les données de la source vers le data warehouse.
  - Le portail de restitution qui sert à reconnaître et analyser les données stockées dans l'entrepôt.
- ✓ Déploiement du système, qui permet de valider notre travail dans l'entreprise. L'entrepôt est crée et reliée avec le système opérationnel par l'intermédiaire de la zone d'alimentation que nous avons programmé, Nous avons aussi créé la bd multidimensionnelle tournant sous 'Oracle9i OLAP service' et installé l'application de l'interface utilisateur 'Oracle Discoverer 10i'. Notre système décisionnel est implémenté sur Oracle9i comme serveur de données de l'entrepôt, Intranet comme infrastructure réseau.

Ce projet nous a permis de mieux connaître les systèmes d'information décisionnels en particulier le data warehouse, il représente 9 mois de travail. Il a été aussi une expérience professionnelle très enrichissante et moyen de mettre en pratique cinq ans d'études.