

الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التعليم العالي والبحث العلمي
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE
جامعة سعد دحلب البليدة 1
Université Saad Dahleb, Blida 1
Faculté des sciences
Département d'informatique



Mémoire présenté par :
Hellali Imène

Pour l'obtention du diplôme de MASTER

Spécialité : Informatique

Option : Ingénierie des logiciels

Thème :

Conception et Déploiement d'un Data Warehouse métier
Volet : Suivi des projets de l'ENGTP

Membres de jury :

Présidente : Mme N. Rezoug

Examinatrice : Mme I.Chikhi.

Promotrice : Mme F.Boumahdi

Promotion 2015/2016

Résumé :

Aujourd'hui, les nouvelles technologies de l'information permettent de concevoir des systèmes d'information particulièrement performants et novateurs. Avec l'apparition des entrepôts de données, tous les utilisateurs peuvent désormais accéder à l'information stratégique, ceci permet à l'entreprise d'être plus réactive.

Ce mémoire traite la conception et le déploiement d'un entrepôt de données métier, concernant la fonction de suivi des projets de l'entreprise GTP. Il s'agit d'un stockage intermédiaire des données issues des applications de productions, dans lesquelles les utilisateurs finaux puissent exploiter avec des outils de restitution et d'analyse.

L'approche retenue pour la réalisation du système est celui du cycle de vie dimensionnel proposée par Ralph Kimball. Le plan de notre travail est divisé en trois phases : Définitions des besoins, Conception et déploiement du Système.

SOMMAIRE

Table des matières

INTRODUCTION	
I-Contexte général	02
II-Problématique	03
III-Objectif	04
VI-Organisation du mémoire.....	05
Chapitre I : Présentation Générale du Datawarehouse	
1. Introduction	07
2. Différence entre l'informatique de production et l'informatique de décision.....	07
2.1 L'informatique de production.....	08
2.2 L'informatique décisionnelle.....	08
3. Historique :	10
4. L'infocentre	11
5. Le data warehouse.....	11
6. Infocentre Et Datawarehouse	13
7. Les Objectifs Du Datawarehouse	14
8. L'entrepôt De Données.....	15
9. Pourquoi un entrepôt de données	17
9.1 .La problématique des entreprises.....	17
9.2 La réalité des systèmes d'informations.....	18
10. Définition de l'entrepôt de données.....	20
11. L'architecture de Datawarehouse.....	23
12. Structure des Données d'un Data Warehouse.....	27
13. Conclusion	29
Chapitre II : OLAP Et Modélisation Multidimensionnelle	
1. Introduction	31
2. Présentation et définition du concept OLAP « On Line Analytical Processing »	31
3. Les douze-12- règles OLAP:.....	31
4. Les Différents Types De Serveurs OLAP.....	33
5. Les opérations sur OLAP	36
6. La différence entre OLAP l'analytique et OLTP le transactionnel	40
7. Qu'est-Ce Que La Modélisation Multidimensionnelle	42
7.1 Dimensions et indicateurs.....	42
7.2 Représentation pour l'utilisateur.....	42
8. Modélisation Conceptuelle.....	43
8.1 Définition du concept de fait.....	43
8.2 Définition du concept de dimension.....	43
8.3 Granularité des dimensions.....	44
8.4 Modèles en étoile, en flocon et en constellation.....	44
9. Conclusion	44

Chapitre III : Processus Fonctionnels Du Datawarehouse

1. Introduction	46
2. Processus d'alimentation E.T.L	46
2.1 Les phases de l'alimentation.....	47
3. Stockage des données	50
3.1. Le stockage des données externes.....	50
3.2. Optimisation physique.....	51
3.3. Les méta- données.....	51
4. L'administration des données	52
5. Conclusion	57

Chapitre IV : Conception et Déploiement Du Système

1. Introduction	59
2. 2. Approches de mise en place de DW	60
2.1 Top-Down (Bill Inmon et le CIF).....	60
2.2 Bottom-Up (Ralph Kimball et le Bus Architecture).....	60
2.3 Hybride.....	61
3. Approche de Ralph Kimball.....	61
3.1. Description du Schéma du cycle de vie dimensionnel de Ralph Kimball.....	62
4. Spécification de notre approche	64
5. Définition des besoins de l'entreprise.....	64
5.1 Les processus métiers.....	66
5.2 Les indicateurs d'analyse de l'activité.....	67
6. Modélisation dimensionnelle.....	67
6.1. Modélisation dimensionnelle de l'activité « préparation ».....	68
6.2 La modélisation dimensionnelle de l'activité « exécution ».....	70
6.3 La modélisation dimensionnelle de l'activité « Contrôle».....	72
6.4 La modélisation dimensionnelle de l'activité «Capitalisation».....	73
6.5 La modélisation dimensionnelle de l'activité «Equipement».....	74
7. Mise en œuvre de l'Architecture.....	75
7.1. Architecture détaillé du système.....	76
7.2. Construction de l'entrepôt de données.....	78
7.3 Construction des cubes OLAP	80
7.4. Construction de la zone d'alimentation.....	84
7.5. Construction du portail de restitution.....	86
8. Architecture technique de notre système.....	91
9. Disposition physique de notre système.....	92
10. Les utilisateurs du système.....	93
11. Le catalogue des états.....	94
12. Logiciels utilisés	94
13. La sécurité du système.....	94

14. Conclusion.....	95
Conclusion Générale	97
Annexe	
Références Bibliographique	

Liste Des Tableaux

Tab 1.1 : Tableau de comparaison entre infocentre et DATAWAREHOUSE.....	13
Tab 1.2 : Différence entre données du système de production et données décisionnelles....	20
Tab 1.3 : Différence entre système de production et l'entrepôt de données.....	23
Tab 2.1 : MOLAP vs ROLAP.....	36
Tab 2.2 : OLAP vs OLTP.....	41
Tab3.1 : Définition d'une métadonnée.....	53
Tab 4.1 : liste des tables de l'entrepôt de données.....	78
Tab 4.2 : liste des attributs de l'entrepôt de données.....	79
Tab 4.3- Liste des faits choisis et les dimensions sollicitées.....	80

Liste Des Figures

Fig 1.1 évolution des bases de données décisionnelles.....	10
Fig 1.2 : L'architecture de haut en bas : selon Bill Inmon.....	12
Fig 1.3 L'architecture de bas en haut : selon Ralph Kimball.....	12
Fig 1.4 : Présentation de l'approche entrepôt de données.....	17
Fig 1.5 : L'architecture de l'entrepôt de données	23
Fig 1.6 : Structure des données d'un Datawarehouse	27
Fig2.1 : Relational OLAP architecture.....	34
Fig 2.2 : MOLAP architecture.....	35
Fig 2.3 Roll-up.....	37
Fig 2.4 : Drill-down.....	38
Fig 2.5 : Slice.....	39
Fig 2.6 : Dice.....	39
Fig 2.7 : Pivot.....	40
Fig 2.8 Présentation des cubes en tableau.....	42
Fig 3.1 : Processus ETL.....	47
Fig 4.1:Schéma du cycle de vie dimensionnel.....	62
Fig4.2 : Succession des taches de notre approche.....	64

Fig 4.3 : Les processus de l'activité de la fonction « Suivi des Projet ».....	66
Fig4.4 :Modèle dimensionnel en flocon de l'activité « Analyse_Préparation ».....	70
Fig 4.5 : modèle dimensionnel en flocon de l'activité « analyse_exécution ».....	71
Fig 4.6 : modèle dimensionnel en flocon de l'activité « analyse_contrôle ».....	73
Fig 4.7 : modèle dimensionnel en flocon de l'activité « analyse_capitalisation ».....	74
Fig 4.8 : modèle dimensionnel en flocon de l'activité « analyse_équipement ».....	75
Fig 4.9 : Architecture détaillée du système.....	76
Fig 4.10 : Le Cube MOLAP de l'activité « préparation ».....	82
Fig 4.11 : Le Cube MOLAP de l'activité « exécution ».....	82
Fig 4.12 : Le Cube MOLAP de l'activité « équipement ».....	83
Fig 4.13 : Le Cube MOLAP de l'activité « contrôle ».....	83
Fig 4.14 : Le Cube MOLAP de l'activité « capitalisation ».....	84
Fig4.15: le processus d'alimentation de notre entrepôt.....	85
Fig 4.16 : La fréquence de chargement des systèmes sources vers l'entrepôt.....	86.
Fig 4.17 : Page web de connexion à l'entrepôt de données.....	87
Fig 4.18 : Page Web signalant l'erreur.....	88
Fig 4.19 : Page Web du choix des états.....	88
Fig 4.20 : Page Web de l'état « Heures dépensées par Affaire et par Région ».....	89
Fig 4.21 :Drill up.....	90
Fig 4.22 : Slice & Dice	90
Fig 4.23 : Pivoting.....	91
Fig. 4. 24: Disposition physique de notre système.....	92
Fig.4. 25:Les utilisateurs de notre système.....	94

INTRODUCTION

I-Contexte Général :

Avec la généralisation de l'informatique dans tous les secteurs d'activité, les entreprises produisent et manipulent de très importants volumes de données électroniques. Ces données sont stockées dans les systèmes opérationnels de l'entreprise au sein de bases de données (relationnelles, réseaux), et de fichiers. L'exploitation et l'exploration de ces données pour des objectifs d'analyse et d'aide à la prise de décision s'avère difficile et fastidieuse ; elle est réalisée le plus souvent de manière imparfaite par les décideurs grâce à des moyens classiques (requêtes SQL, vues, outils graphiques d'interrogation...).

Ces systèmes paraissent peu adaptés pour servir de support à la décision. Ces bases opérationnelles utilisent en général le modèle relationnel ; celui-ci convient bien aux applications gérant l'activité quotidienne de l'entreprise (commerciale, production, comptabilité ...), mais s'avère inadapté au décisionnel.

Face à cette inadéquation, les entreprises ont recours à des systèmes d'aide à la décision spécifique, basés sur l'approche des entrepôts de données. Cependant, de tels systèmes restent difficiles à élaborer et sont souvent réalisés de manière empirique, rendant l'évolution du système décisionnel délicate. Actuellement, les entreprises ont besoin d'outils et de modèles pour la mise en place de systèmes décisionnels comportant des données évolutives.

La performance de l'entreprise est souvent appréciée sous l'angle du succès commercial de ses produits et services. Toutefois, au regard du contexte actuel cette performance est aussi appréciée sous l'angle de l'aptitude de l'entreprise à être réactive globalement aux menaces et aux changements de son environnement. Les systèmes décisionnels d'aide au pilotage des activités des entreprises constituent une réponse globale aux problématiques de pilotage auxquelles elles sont confrontées aujourd'hui. Ces systèmes représentent l'ensemble des outils d'analyse et de représentation des données de l'entreprise nécessaire pour une prise de décision rapide, efficace, et sûre, car 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.

II- Problématique :

ENGTP est une filiale de l'entreprise national SONATRACH, qui s'occupe de la réalisation des grands travaux pétroliers, employant de très grands moyens humains ,matériels et logistiques .elle dispose de multitudes de systèmes opérationnels traitant ainsi de milliers de transactions qui sont effectuées chaque jour ,au niveau de chaque service de l'entreprise .ces systèmes couvrent la majeure partie de ces activités de base ,qui se trouve submerger par une montagne de données et qu'elle doit maintenant confronter les problèmes de gestions et l'organisation de celles-ci.

Face à cette ampleur, L'ENGTP se trouve dans l'incapacité de tirer profit de cette mine d'information et de se les procurer rapidement, et ce qui lui permet ainsi de prendre les bonnes décisions pour le bon déroulement de ces projets.

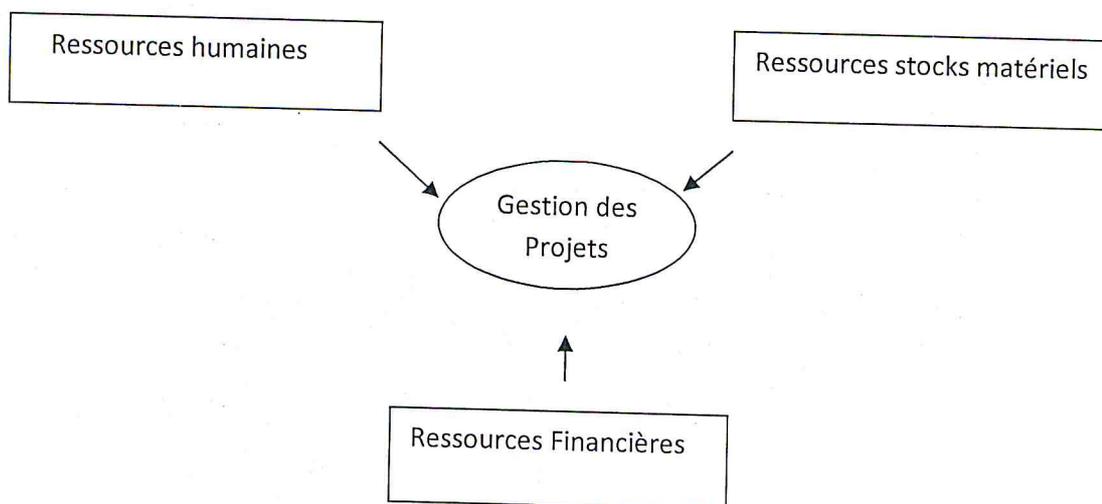
Les principales difficultés auxquelles fait face sont :

- L'incapacité de gérer et d'organiser le volume important de données pertinentes et stratégiques qui se trouve au niveau de la base opérationnelle.
- Les données jugées importantes se trouvent éparpillées dans différentes applications des services de l'entreprise ce qui les rend inaccessibles.
- L'omission totale de l'historique des données même sur une période de quelques mois.
- L'inexistence de la capitalisation des projets.
- Il est impossible de faire une analyse fiable sous un système opérationnel, car les accès simultanées des utilisateurs et les décideurs en particulier le rend très lourd et créer des conflits de privilèges d'accès.

III- Objectifs :

Notre objectif est de réaliser et de mettre en place un système d'aide à la décision capable de devenir un outil performant de pilotage et d'analyse central de l'entreprise par l'utilisation combinée de données issues du système d'information de gestion qui communique avec les systèmes de Comptabilité, du Personnel ainsi que celui de la gestion du matériel à travers des situations financières périodiques contenant les coûts du projet par corps de métier.

Cette approche est illustrée par la figure suivante :



Cette réalisation va permettre de :

Notre objectif est de concevoir et déployer un Datawarehouse métier pour la fonction des projets, cette réalisation va permettre de :

- Collecter et déterminer les informations pertinentes et stratégiques pour les décideurs à fin d'établir des analyses fiables et robustes.
- Concevoir un entrepôt de données intégré métier, il s'agit de mettre en place la base décisionnelle qui contiendra uniquement les données stratégiques pour les décideurs de l'activité.
- Retrouver une information historique et transversale de l'entreprise (non volatiles).
- Offrir à l'utilisateur liberté de navigation.
- Homogénéiser les sources de données en ce qui concerne la gestion des projets.
- Construire le portail de restitution qui permettra au management de consulter la base à travers des outils conviviaux, et dans notre cas de représenter les données sous forme multidimensionnelle grâce à l'outil OLAP.

IV-Organisation du mémoire

Notre mémoire s'articule essentiellement autour de cinq chapitres.

- Dans le premier chapitre, nous donnerons une synthèse sur l'architecture globale d'un système décisionnel basé sur l'approche « Entrepôt de données ». Nous présenterons successivement les concepts de base de l'entrepôt de données, ses objectifs et sa structure.
- Dans le deuxième chapitre, nous présentons le concept OLAP et la navigation dans les données de l'entrepôt, ainsi que les principes de modélisation multidimensionnelle.
- Dans le troisième chapitre, nous aborderons successivement les processus fonctionnels de Datawarehouse, le processus d'alimentation de l'entrepôt, le stockage des données et pour finir l'administration des données dans le Datawarehouse.
- Dans le quatrième et le dernier chapitre, nous entamerons une analyse des besoins décisionnels, qui consiste à prendre en compte la stratégie de l'entreprise, ses besoins pour la fonction des projets, ce qui conduit en suite à établir notre schéma multidimensionnel de données, pour arriver à la fin à présenter notre entrepôt et la base multidimensionnelle (cube OLAP) ainsi que les outils d'analyse utilisés et pour finir nous définissons l'architecture technique de notre système.

Le mémoire se terminera par une conclusion générale pour mettre en évidence l'intérêt d'un système d'aide à la décision basé sur l'architecture Datawarehouse.

CHAPITRE I
PRESENTATION GENERALE
DU DATAWAREHOUSE

1. Introduction

Le concept d'entrepôt de données DATAWAREHOUSE a été formalisé la première fois en 1990.

L'idée est de construire une base de données orientée sujet, intégrée, contenant des informations datées, non volatiles et exclusivement destinée au processus d'aide à la décision, fut dans un premier temps accueilli avec une certaine perplexité. Beaucoup n'y voyaient qu'un habillage d'un concept déjà ancien : l'infocentre.

Mais l'économie actuelle en a décidé autrement. Les entreprises sont confrontées à une concurrence de plus en plus forte, des clients de plus en plus exigeants, un contexte organisationnel de plus en plus complexe et mouvant.

Pour faire face aux nouveaux enjeux économiques, l'entreprise doit anticiper. L'anticipation ne peut être efficace qu'en s'appuyant sur de l'information pertinente alors qu'actuellement les données sont surabondantes, non organisées dans un contexte décisionnel et éparpillées dans de multiples systèmes hétérogènes, donc il devient fondamental de rassembler et d'homogénéiser les données afin de faciliter la prise de décision.

Pour répondre à ces besoins, les connaisseurs du domaine ont jugé opportun de définir et d'intégrer une architecture qui serve de fondation aux applications décisionnelles.

Cette architecture globale est le DATAWAREHOUSE, une notion vaste que nous allons définir par la suite dans ce chapitre en introduisant aussi les différents composants de base d'un DATAWAREHOUSE, ainsi les objectifs et la structure de ce dernier.

2. Différence entre l'informatique de production et l'informatique de décision

Les systèmes de production sont caractérisés par une activité constante composée de modifications et d'interrogations fréquentes des bases de données par de nombreux utilisateurs : ajouter une commande, modifier une adresse de livraison, rechercher les coordonnées d'un client, etc. l'intégrité des données est nécessaire pour ce genre d'applications (il faut par exemple, interdire la modification simultanée d'une même donnée par deux utilisateurs différents). La priorité est donnée en premier lieu à l'enregistrement rapide, sûr et efficace des données.

À l'inverse, nul besoin de modification ou d'enregistrement de nouvelles données dans les

systèmes d'information de décision. En effet, le but principal consiste en l'interrogation du système d'information afin de permettre l'analyse des indicateurs pertinents pour faciliter les prises de décisions. Les questions posées seront de la forme : quelles sont les ventes du produit X pendant le trimestre A de l'année B dans la région C. Une telle interrogation peut nécessiter des temps de calcul importants. Or, l'activité d'un serveur transactionnel ne peut être interrompue. Il faut donc prévoir une nouvelle organisation qui permette de mémoriser un grand volume de jeux de données et qui soit structurée de façon à avoir un moteur de recherche de connaissances efficace et appropriée.

2.1L'informatique de production

L'un des formalismes les plus utilisés pour la représentation conceptuelle des systèmes d'information est le modèle Entité Association. Dans ce genre de modèle, toute redondance est considérée comme source d'erreurs, d'incohérence, et va pénaliser les temps d'exécution et rendre complexes les procédures d'ajout, de suppression ou de modification. Les systèmes transactionnels temps réel, OLTP (On-line Transaction Processing) , garantissent l'intégrité des données. Les utilisateurs accèdent aux données de la base par de très courtes transactions atomiques et isolées. L'isolation évite la perturbation ou l'interruption de la transaction. La brièveté quant à elle garantit que les temps de réponse seront acceptables (inférieurs à la seconde) dans un environnement avec de nombreux utilisateurs. Toutefois, le modèle Entité Association et sa réalisation dans un schéma relationnel sont pourtant des obstacles importants pour l'accès de l'utilisateur final aux données. Généralement, ces modèles contiennent plusieurs dizaines d'entités. Les bases sont alors constituées de nombreuses tables, reliées entre elles par divers liens dont le sens n'est pas toujours explicite. Il devient clair que le but était de faciliter le travail des développeurs et d'améliorer des performances de tâches répétitives, prévues et planifiées tournées vers la production de documents standards (factures, commandes...), au détriment des besoins réels de l'utilisateur final. La dernière caractéristique de ces bases de données est qu'elles conservent l'état instantané du système. Dans la plupart des cas, l'évolution n'est pas sauvegardée. On garde simplement une trace des versions instantanées pour la reprise en cas de panne et pour des raisons légales.

2.2L'informatique décisionnelle

L'informatique décisionnelle désigne les moyens, les outils et les méthodes qui permettent de collecter, consolider, modéliser et restituer les données d'une entreprise en vue d'offrir une

aide à la décision et de permettre aux responsables de la stratégie d'une entreprise d'avoir une vue d'ensemble de l'activité traitée.

Ce qui caractérise d'abord les besoins, c'est la possibilité de poser une grande variété de questions au système, certaines prévisibles et planifiées comme des tableaux de bord et d'autres imprévisibles. Si des outils d'édition automatiques pré-programmés peuvent être envisagés, il est nécessaire de permettre à l'utilisateur d'effectuer les requêtes qu'il souhaite, par lui-même, sans l'intervention de programmeurs. Il sera souvent nécessaire de filtrer, d'agréger, de compter, sommer et de réaliser quelques statistiques élémentaires (moyenne, écart-type,...). La structure logique doit être prévue pour rendre aussi efficace que possible toutes ces requêtes. Pour y parvenir, il est nécessaire d'introduire de la redondance dans les informations stockées en mémorisant des calculs intermédiaires. On rompt donc avec le principe de non redondance des bases de production. La cohérence assurée par les systèmes de production est toute relative. Elle se contrôle au niveau de la transaction élémentaire mais pas au niveau global et des activités de l'organisation. A l'inverse des systèmes d'informatique décisionnelle, la cohérence requise doit être interprétable par l'utilisateur. Par exemple, si les livraisons n'ont pas été toutes saisies dans le système, comment garantir la cohérence de l'état du stock ?

Les systèmes d'informatique décisionnelle doivent donc assurer plutôt une cohérence globale des données. Pour ce faire, leur alimentation doit être une opération réfléchie et planifiée dans le temps. Les transferts de données du système opérationnel vers le système décisionnel seront réguliers avec une périodicité bien choisie dépendante de l'activité de l'entreprise. Chaque transfert sera contrôlé avant d'être diffusé. Une dernière caractéristique importante de l'informatique décisionnelle, qui est aussi une différence fondamentale avec les bases de production, est qu'aucune information n'y est jamais modifiée. En effet, on mémorise toutes les données sur une période déterminée, les données ne seront jamais remises à jour car toutes les vérifications utiles à la cohérence globale sont procédées lors de l'alimentation. L'utilisation se résume donc à un chargement périodique, puis à des interrogations non régulières, non prévisibles, parfois longues à exécuter. S'il existe effectivement des informations importantes, il n'en est pas moins nécessaire de construire une structure pour les héberger, les organiser et les restituer à des fins d'analyse. Cette structure est le Datawarehouse ou entrepôt de données. Ce n'est pas une usine à produire l'information, mais

plutôt un moyen de la mettre à disposition des utilisateurs de manière efficace et organisée. [1]

3. Historique

Avant de faire une synthèse sur les DATAWAREHOUSE, il convient de situer le datawarehousing au sein des systèmes d'information, agrégés SI, on distingue principalement les SI opérationnels et les SI dites informationnels. Le DATAWAREHOUSE ou l'entrepôt de données est l'un des composants des SI informationnels. Le but des SI informationnels est de maîtriser l'information, dans la mesure où elle est un avantage concurrentiel pour les sociétés. Le développement des premiers systèmes d'informations s'est concentré sur l'automatisation des processus opérationnels, ainsi que sur les données liées aux processus. Ces systèmes opérationnels permettent un gain de productivité non négligeable. Ils soutiennent le bon déroulement de l'activité principale de l'entreprise, l'excellence opérationnelle.

Les besoins d'analyse sont arrivés bien plus tard, ils révèlent plus de l'avantage concurrentiel que de l'excellence opérationnelle.

Les SI informationnels et les SI opérationnels ont des besoins et des structures totalement différentes. Le SI opérationnel doit être performant : les accès en écriture sont importants, même si prédéfinis. Tandis que les SI informationnels doivent être flexible, fournir des possibilités d'analyse conséquente, la lecture y est intensive et les requêtes **ad hoc** sont monnaie courante [2].

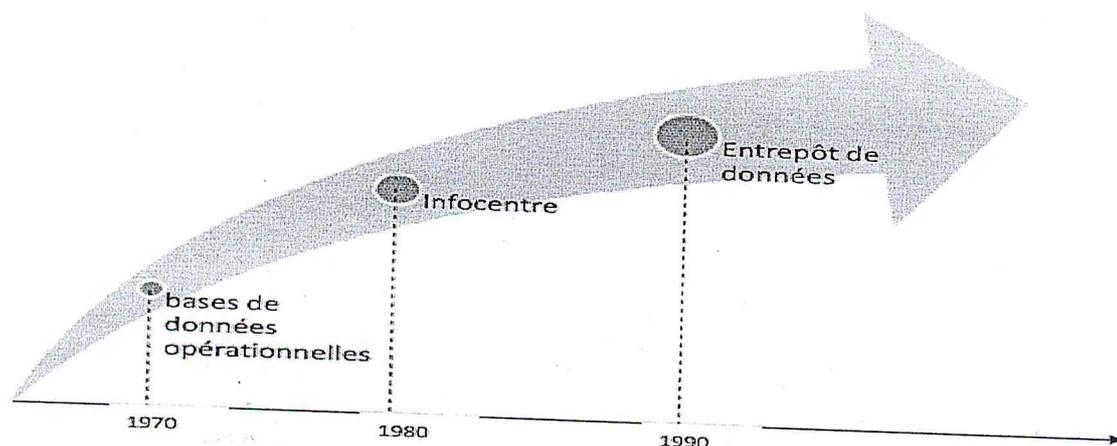


Fig 1.1 évolution des bases de données décisionnelles.

4. Infocentre

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

Chaque nouvelle valeur remplace l'ancienne valeur. Il est donc impossible de retrouver une valeur calculée dans une session préalable aux dernières alimentations, il n'y a pas de gestion d'historique des valeurs. L'infocentre sert à prendre des décisions opérationnelles basées sur des valeurs courantes. L'intégration des données est plus au moins poussée. Le processus d'alimentation est simple.

La finalité d'un infocentre est de permettre aux utilisateurs d'accéder à leurs données dans leurs propres termes. [3]

5. Le Data warehouse

Un Entrepôt de données est une base de données regroupant une partie ou l'ensemble des données fonctionnelles d'une entreprise. Il entre dans le cadre de l'informatique décisionnelle ; son but est de fournir un ensemble de données servant de référence unique, utilisée pour la prise de décisions dans l'entreprise par le biais de statistiques et de rapports réalisés via des outils de reporting. D'un point de vue technique, il sert surtout à 'délester' les bases de données opérationnelles des requêtes pouvant nuire à leurs performances.

D'un point de vue architectural, il existe deux manières de l'appréhender :

- L'architecture *de haut en bas* : selon Bill Inmon, l'entrepôt de données est une base de données au niveau détail, consistant en un référentiel global et centralisé de l'entreprise. En cela, il se distingue du Datamart, qui regroupe, agrège et cible fonctionnellement les données.

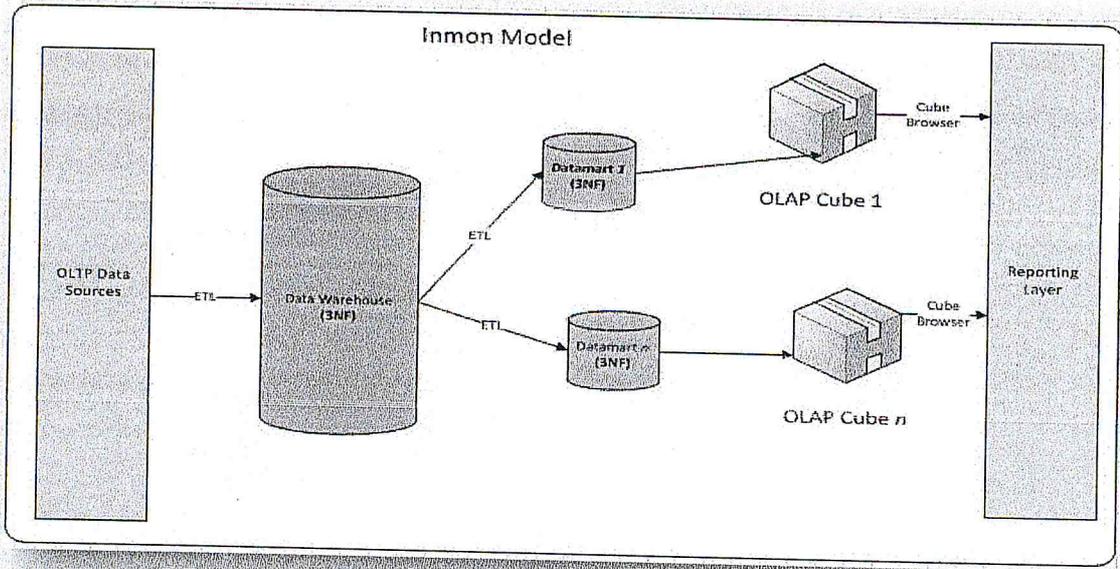


Fig 1.2 : L'architecture *de haut en bas* : selon Bill Inmon [Site1]

- L'architecture *de bas en haut* : selon Ralph Kimball, l'entrepôt de données est constitué peu à peu par les Datamarts de l'entreprise, regroupant ainsi différents niveaux d'agrégation et d'historisation de données au sein d'une même base.

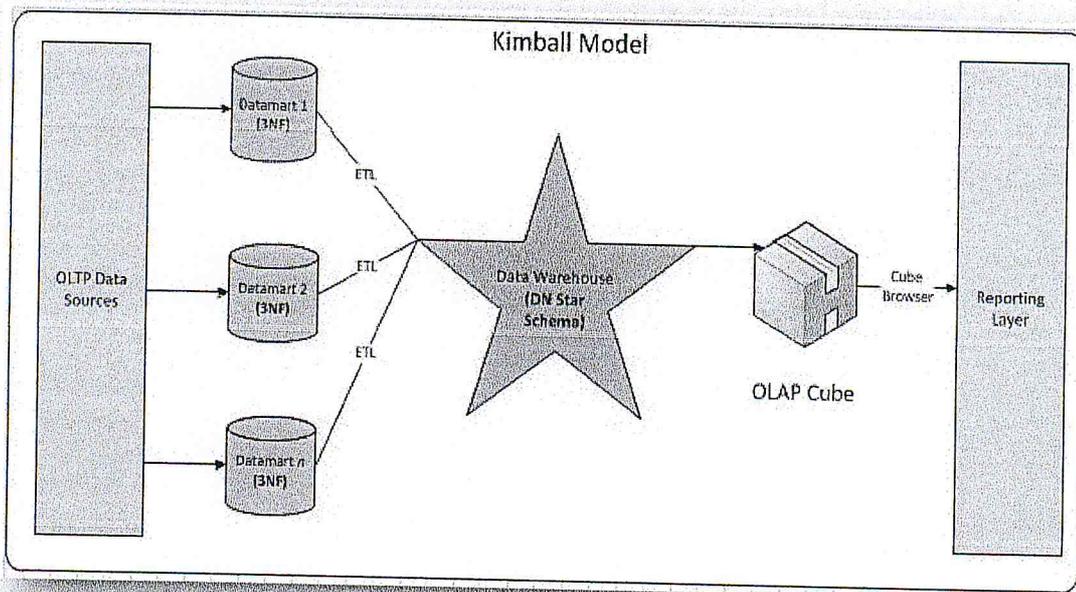


Fig 1.3 L'architecture *de bas en haut* : selon Ralph Kimball [Site1]

La définition la plus communément admise est un mélange de ces deux points de vue. Le terme *Data warehouse* englobe le contenant et le contenu : il désigne d'une part la base détaillée qui est la source de données à l'origine des Datamarts, et d'autre part l'ensemble constitué par cette base détaillée et ses Datamarts. De la même manière, les méthodes de conception actuelles prennent en compte ces deux approches, privilégiant certains aspects selon les risques et les opportunités inhérents à chaque entreprise. [Site 1].

6. Infocentre Et Datawarehouse

Certaines caractéristiques sont identiques. Mais il existe de nombreux éléments permettant de différencier les deux notions.

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

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

Dans un infocentre, chaque nouvelle valeur remplace l'ancienne valeur. Il est donc impossible de retrouver une valeur calculée dans une session préalable aux dernières alimentations. La non volatilité est une caractéristique essentielle du DATAWAREHOUSE. De même, l'historisation des données dans un infocentre, il n'y a pas de gestion d'historique des valeurs.

L'infocentre sert à prendre des décisions opérationnelles basées sur des valeurs courantes. Au niveau d'un DATAWAREHOUSE, l'utilisateur travaille sur les historiques pour des prises de décisions à long terme, des positionnements stratégiques et pour analyser des tendances.

Dans un infocentre, l'intégration des données est plus ou moins poussée. Le processus d'alimentation est simple.

La finalité d'un infocentre est de permettre aux utilisateurs d'accéder à leur données dans leurs propres termes.

Infocentre	DATAWAREHOUSE
Collection de données	Collection de données
Orientées sujet	Orientées sujet
Intégrées	Intégrées
Volatiles	Non volatiles
Actuelles	Historisées
Organisées pour le support d'un processus de décision ponctuelle	Organisées pour le support d'un processus d'aide à la décision
Outil	Architecture

Tab 1.1 : Tableau de comparaison entre infocentre et DATAWAREHOUSE[4]

7. Les Objectifs Du Datawarehouse

La problématique des entreprise aujourd'hui consiste en la gestion des informations et leurs organisation, en effet les données sont éparpillées dans différentes supports et plates formes, ce qu'il devient difficile pour l'utilisateur de rechercher de faits marquants dans un magma de données qui s'appuie sur un capital informationnel existant, lequel est constitué d'information interne à l'entreprise, géré par des systèmes d'information, qui doit être complétées par des données externes acquises auprès de partenaires ou des sociétés , ce qui donne un accès difficile et souvent prend un temps précieux aux utilisateurs.

L'utilisation du DATAWAREHOUSE résout toutes les préoccupations des entreprises, en effet il offre beaucoup d'avantage à celle-ci, ça permet d'homogénéiser les données en les collectant et stockant dans des bases de données et des fichiers pour une prise de décision meilleur et rapide.

Les principaux objectifs du DATAWAREHOUSE tels définis par Ralph Kimball, dans son livre « Entrepôt de données, Guide pratique du concepteur de DATAWAREHOUSE ».

1. Accès aux données de l'entreprise :

En effet, dans certains cas la donnée se trouve dans des systèmes archaïques dont l'accès est impossible sans l'utilisation de systèmes adéquats à leurs structures. C'est pour cela l'utilisation du DATAWAREHOUSE assure une vue transversal qui permet l'accès aux informations de l'entreprise et de l'organisation.

2. Les informations sont cohérentes :

Cela signifie que les requêtes faites à des moments différents doivent fournir les mêmes résultats. La cohérence veut aussi dire que lorsque les personnes demandent la définition de l'élément « contrat », elles obtiennent une réponse leur permettant de savoir ce qu'elles obtiendront de la base de données c'est à dire que les données en provenance du système de production doivent avoir la même signification que dans l'entrepôt de données. la cohérence implique que les données sont chargées dans leurs totalités.les données d'un DATAWAREHOUSE doivent pouvoir être séparées et combinées au moyen de toutes les mesures possibles de l'activité.

3. Les outils de présentation font partie du DATAWAREHOUSE :

Le concept de DATAWAREHOUSE n'intègre pas véritablement l'aspect « outils d'aide à la décision ». Il concerne plutôt la notion d'infrastructure pour le décisionnel. Ces outils d'aide à la décision sont cependant les éléments qui permettent d'exploiter l'architecture qu'est le DATAWAREHOUSE. L'outil d'aide à la décision est la « la partie visible de l'iceberg » pour l'utilisateur, mais aussi un ensemble de requêtes, d'analyses et de présentations d'informations.

4. les données publiées sont stockées dans le DATAWAREHOUSE :

Le DATAWAREHOUSE est un entrepôt de données. En effet les données qui le constituent sont d'origine différente, recueillies de diverses sources d'informations de l'entreprise. Il s'agit d'un stockage des données issues des applications de production, dans différents supports, sous divers formats et sur différentes plates-formes informatiques, dans lesquelles les utilisateurs finaux puisent avec des outils de restitution et d'analyse après être soigneusement agrégées, transformées et 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.

5. Qualité de l'information d'un DATAWAREHOUSE :

Le DATAWAREHOUSE offre une information de très grande qualité, il s'agit des informations sur des données stratégiques et pertinentes pour le management. En effet, dans toutes les entreprises l'information est devenue « le nerf de la guerre ». Disposer de l'information utile est savoir la rendre disponible à ceux qui sauront en tirer parti dans son entreprise et au-delà sont des éléments qui permettent de faire la différence.

En effet comment peut-on prendre des décisions d'une haute importance pour l'entreprise et obtenir des analyses fiables si les données brutes sont de mauvaise qualité ? L'entrepôt de données ne peut remédier à la mauvaise qualité des données ou à l'absence d'une donnée. [4]

8. L'entrepôt De Données

Après plusieurs années d'utilisation des infocentres, c'est-à-dire la duplication de données de production, les services informatiques ont imaginé une évolution intelligente de ce mode de stockage. Ils ont en effet constaté que les informations traitées dans les applications

opérationnelles étaient très différentes de celles interrogées dans les applications décisionnelles. Le nombre de tables, de fichiers interrogés dans une même requête est bien plus important dans l'aide à la décision, le nombre d'indicateurs calculés également. En revanche, les applications décisionnelles se contentent presque toujours de lire les données. Elles n'ont jamais à écrire de nouvelles informations dans les bases de données.

Autre constat, les questions posées par un décideur impliquent fréquemment des informations stockées dans plusieurs applications ou base de données. Lorsque vous calculez la rentabilité de vos clients, vous exportez des données de la gestion commerciale (factures, commandes), de la comptabilité (délais de règlement, impayés), mais également de la gestion de production (coût des produits fabriqués). Le fait de dupliquer ces bases de données dans un infocentre ne simplifie en rien ces extractions.

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

Fréquemment construit à partir d'une base de données relationnelle, cet entrepôt de données sert littéralement d'entrepôt. On y verse une copie des données qui serviront, un jour, à l'analyse et à la prise de décision. Cela évite également de conserver un historique trop important dans les bases de production, souvent concernées par le court terme, alors que les analystes ont besoin de recul pour détecter des tendances.

Ces périodes de référence varient d'ailleurs beaucoup en fonction des secteurs d'activité. On considère que dans la grande distribution, des données vieilles de six mois sont inexploitable tant les produits les assortiments, les marques ont évolués. La figure suivante représente l'approche entrepôt de données « DATAWAREHOUSE ».

L'informatique a un rôle à jouer, en permettant à l'entreprise de devenir plus entreprenante et d'avoir une meilleure connaissance de ses clients, de sa compétitivité ou de son environnement.

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

9.2 La réalité des systèmes d'informations

On dénombre deux catégories de systèmes d'information, les systèmes transactionnels et les systèmes décisionnels. Les premiers servent à gérer le quotidien, l'opérationnel de l'entreprise. Les seconds sont utilisés pour prendre du recul, et servir de support aux décisions de l'entreprise et de ses dirigeants.

Les systèmes transactionnels ou opérationnels seraient des mines d'or informationnelles. Ils sont les outils qu'utilisent les entreprises chaque jour. De la gestion des achats à celle des ventes. En fait, il n'en est rien.

Les données contenues dans ces systèmes sont :

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

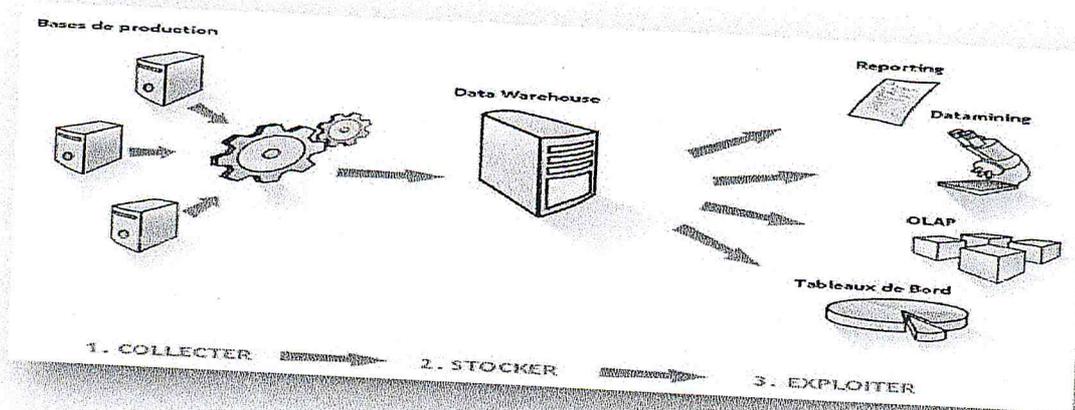


Fig 1.4 : Présentation de l'approche entrepôt de données.[1]

9. Pourquoi un entrepôt de données

9.1 .La problématique des entreprises

Aujourd'hui, les organisations, quelque soient leur secteur d'activité, se trouvent avec des données surabondantes, non organisées dans une perspective décisionnelle et éparpillée dans de multiples systèmes hétérogènes. Il devient donc capital de rassembler et d'homogénéiser les données afin de permettre d'analyser les indicateurs pertinents pour faciliter les prises de décisions. L'entreprise construit un système décisionnel pour améliorer sa performance. Elle doit décider et anticiper en fonction de l'information disponible et capitaliser sur son expérience.

Depuis plusieurs dizaines d'années, une importante masse d'informations est stockée sous forme informatique dans les entreprises. Les systèmes d'information sont destinés à garder la trace d'événements de manière fiable et intégrée. Ils automatisent de plus en plus les processus opérationnels.

Parallèlement, les entreprises réalisent la valeur du capital d'information dont elles disposent. Au-delà de ce que l'informatique leur apporte en terme fonctionnel, elles prennent conscience de ce qu'elle pourrait apporter en terme de contenu informationnel. Considérer le système d'information sous cet angle, c'est-à-dire, en tant que levier pour accroître leur compétitivité et leur réactivité n'est pas nouvelle. Par contre, étant donné l'environnement concurrentiel actuel, cela devient une question de survie.

Toutes les applications de ces systèmes répondent à la même attente : permettre la saisie d'informations, leur traitement et la production en sortie de résultats, sous forme de documents papier, de consultations à l'écran ou d'autres informations.

Elles sont par ailleurs caractérisées par des spécificités communes : la première étant la capacité à gérer de grands volumes de données, la seconde des temps de réponse très réduits et la troisième des requêtes relativement simples d'un point de vue informatique.

La décision concerne tous les départements de l'entreprise : finance, ressources humaines, ventes, marketing, production et bien entendu la direction générale. L'aide à la décision les concerne dans la même globalité.

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

Le tableau suivant présente les différences entre les données opérationnelles et décisionnelles.

Données opérationnelles	Données décisionnelles
Orientées application, détaillées, précises au moment de l'accès	Orientées activité (thème, sujet), condensées, représentes des données historiques
Mise à jour interactive possible de la part des utilisateurs	Pas de mise à jour interactive de la part des utilisateurs
Accédées de façon unitaires par une personne à la fois	Utilisées par l'ensemble des analystes, gérées par sous-ensemble
Cohérence atomique	Cohérence globale
Haute disponibilité en continu	Exigence différente, haute disponibilité ponctuelle

Uniques (pas de redondance en théorie)	Peuvent être redondantes
Structure statique, contenu variable	Structure flexible
Petite quantité de données utilisées par un traitement	Grande quantité de données utilisée par les traitements
Réalisation des opérations au jour le jour	Cycle de vie différent
Forte probabilité d'accès	Faible probabilité d'accès
Utilisées de façon répétitive	Utilisée de façon aléatoire

Tab 1.2 : Différence entre données du système de production et données décisionnelles[5]

S'il existe effectivement des informations importantes, il n'en est pas moins nécessaire de construire une structure pour les héberger, les organiser et les restituer à des fins d'analyse. Cette structure est l'Entrepôt de Données ou « *Data Warehouse* ». Ce n'est pas une usine à produire l'information, mais plutôt un moyen de la mettre à disposition des utilisateurs de manière efficace et organisée.

La mise en œuvre de l'entrepôt de données ou « *Data Warehouse* » est un processus complexe. L'objectif à atteindre est de recomposer les données disponibles pour en donner :

- ❖ Une vision intégrée et transversale aux différentes fonctions de l'entreprise.
- ❖ Une vision métier au travers de différents axes d'analyse.
- ❖ Une vision agrégée ou détaillée suivant le besoin des utilisateurs.

L'entrepôt de données permet la mise en place d'un outil décisionnel s'appuyant sur les informations pertinentes pour l'entreprise, centrées sur le métier utilisateur. [5]

10. Définition de l'entrepôt de données

De nombreuses définitions ont été proposées, soit académiques, soit par des éditeurs d'outils, de base de données ou par des constructeurs, cherchant à orienter ces définitions dans un sens mettant en valeur leur produits.

Un entrepôt de données « *Data warehouse* » est défini de plusieurs façons comme suit :

✓ « Un Data Warehouse est une collection de données concernant un sujet particulier, varie dans le temps, non volatiles et ou les données sont historisées, intégrées et organisée pour le processus de décisions » [6].

✓ « Fait référence à une collection de technologie d'aide à la décision permettant à des (managers, dirigeants et analystes) de prendre des décisions pertinentes et rapides ». [7].

✓ « Une collection de technologies décisionnelles formant un environnement permettant aux décideurs de prendre des décisions plus pertinentes et plus rapides ». [8].

✓ « Un entrepôt de données est une copie des données de transaction spécifiquement structurées pour la question et l'analyse ». [9]

Ce système correspond à une collection de données (agrégées ou atomiques) pour aider les gestionnaires dans leur prise de décision. Les données sont :

❖ **Orientées sujets**

L'entrepôt de données est organisé autour des sujets majeurs de l'entreprise, contrairement aux données des systèmes opérationnels. Ces derniers en général sont organisés par processus fonctionnels. Les données sont structurées par thèmes. L'intérêt de cette organisation est de disposer de l'ensemble des informations utiles sur un sujet le plus souvent transversal aux structures fonctionnelles et organisationnelles de l'entreprise.

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

❖ **Données intégrées**

Les données alimentant l'entrepôt de données proviennent de multiples applications sources hétérogènes. Les données des systèmes de production doivent être converties, reformatées et nettoyées, de façon à avoir une seule vision globale dans l'entrepôt de données.

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

Une donnée doit avoir une description et un codage unique. Cette phase d'intégration est très complexe et représente 60 à 90 % de la charge totale d'un projet.

❖ **Historisées**

En règle générale, dans un système de production, la donnée est mise à jour lors de chaque nouvelle transaction. Une mise à jour annule et remplace l'ancienne valeur. Dans un entrepôt de données, la donnée ne doit jamais être mise à jour en mode « annule et remplace ». Un référentiel temps doit être associé à la donnée afin d'être capable d'identifier une valeur particulière dans le temps.

❖ **Non volatiles**

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

❖ **Résumées**

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

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

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

Le tableau ci-après présente les principales différences entre le système de production et l'entrepôt de données « *Data Warehouse* ».

Critère	Système de production	Entrepôt de données
Niveau de détail des informations utilisateurs	Très détaillé	Synthétique, parfois détaillé
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/Mise à jour & Consultation	Consultation uniquement

Tab 1.3 : Différence entre système de production et l'entrepôt de données[5]

11. L'architecture de Datawarehouse

L'architecture de l'entrepôt de données doit être conçue de façon à ce que la bonne information soit aisément accessible pour la bonne personne au bon moment. Pour y parvenir, cette architecture doit d'une part permettre de transformer de l'information de façon à ce qu'elle réponde aux objectifs métiers des utilisateurs, et d'autre part d'organiser le mouvement de l'information de la source à la cible. [10]

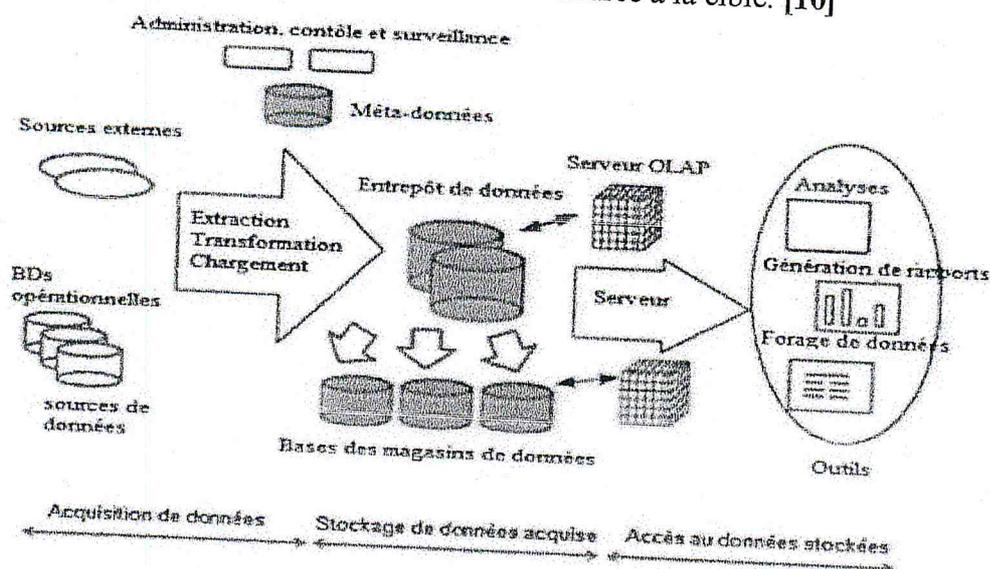


Fig 1.5 : L'architecture de l'entrepôt de données [9]

1. Le système source :

Système opérationnel d'enregistrement, dont la fonction consiste à capturer les transactions liées à l'activité de l'entreprise. Il s'agit souvent de ce que l'on appelle les applications de gestion. La principale priorité du système source est le temps de disponibilité. Les systèmes sources, de par leur nature, « vivent leur vie ». Généralement, peu d'effort doivent être consenti pour homogénéiser les dimensions de base telles que le produits, le client, le lieu ou le calendrier par rapport aux autres SI opérationnels de l'entreprise.

2. La zone de préparation des données

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

3. Le serveur de présentation

Machine cible sur laquelle l'entrepôt de données est stocké et organisé pour répondre en accès direct aux requêtes émises par des utilisateurs. Des générateurs d'états ou d'autres applications. Sur le serveur de présentation, les données seront stockées et représentées sous une forme dimensionnelle, de façon à faciliter l'accès pour l'utilisateur final. Dans la majorité des cas, le serveur est basé sur une base de données relationnelle, de sorte que les tables y soient organisées sous forme de schéma en étoile.

4. L'entrepôt de données

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

5. Le magasin de données

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

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

6. Portail de restitution

C'est la part 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 de l'entrepôt de données pour les restituer soit à l'utilisateur, soit à une autre application. Les services offerts par le portail de restitution sont les services d'accès aux données, les applications de modélisation et de Data Mining.

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

7. Meta données

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

8. Data mining

Le terme de Data Mining signifie littéralement *forage de données*, comme dans tout forage, son but est pouvoir extraire un élément : la connaissance. Ces concepts s'appuient sur le constat qu'il existe au sein de chaque entreprise des informations cachées dans le gisement de

données. Ils permettent grâce à un certain nombre de techniques spécifiques, de faire apparaître des connaissances.

L'exploration se fait sur l'initiative du système, par un utilisateur métier, et son but est de remplir l'une des tâches suivantes : classification, estimation, prédiction, regroupement par similitudes, segmentation, descriptions et dans une moindre mesure, l'optimisation.

9. Outils statistiques :

Les statistiques sont à la base de tout raisonnement sur les données. Elles permettent de synthétiser un grand nombre de valeurs pour une variable, grâce à un nombre très réduit d'information. Pour chaque variable, on va ainsi rechercher au moins deux indicateurs : un pour mesurer la tendance centrale, un pour mesurer la dispersion

10. EIS :

Un EIS (Executive Information System) est un outil de visualisation et de navigation dans les données permettant de constituer des tableaux de bord. Il est constitué d'outils qui permettent aux différents niveaux de management d'accéder aux informations essentielles de leur organisation, de les analyser et de présenter de façon élaborée. Ces outils sont dotés d'une interface graphique très conviviale et très esthétique.

L'utilisateur final ne peut visualiser que les informations initialement prévues par le concepteur du tableau de bord. Il ne permet pas de poser une question qui n'a pas été prévu dès le départ.

11. 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 DATAWAREHOUSE.

12. Requêteur :

Un requêteur permet à l'utilisateur final d'accéder aux données de l'entreprise de manière autonome, dans un langage proche de celui de son métier. Ces outils peuvent nécessiter la connaissance de la structure de la base que l'on interroge.

Différentes générations d'outils se sont succédées sur le marché. Si au départ ils étaient de simples outils d'interrogation graphique, ils tendent désormais à s'orienter vers les métiers à s'orienter vers les métiers des utilisateurs pour permettre une utilisation plus intuitive. Les premières solutions nécessitaient la connaissance de la structure de la base de données et même parfois du langage SQL.

Aujourd'hui ces éléments sont de plus en plus transparents pour l'utilisateur final. La facilité d'utilisation des outils s'est accrue ainsi que leurs coûts de mise en place et de maintenance.

13. L'outil OLAP :

OLAP (Online Analytical Processing), il utilise des vues multidimensionnelles de données agrégées afin de permettre un accès rapide à des informations stratégiques pour une analyse plus fine. OLAP permet à des analystes et gestionnaires d'accéder rapidement interactivement à des données à travers une large variété de vues possible d'informations, comme il transforme aussi des données brutes de manière à refléter la réelle dimension d'une entreprise du point de vue de l'utilisateur.

12. Structure des Données d'un Data Warehouse : [11]

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

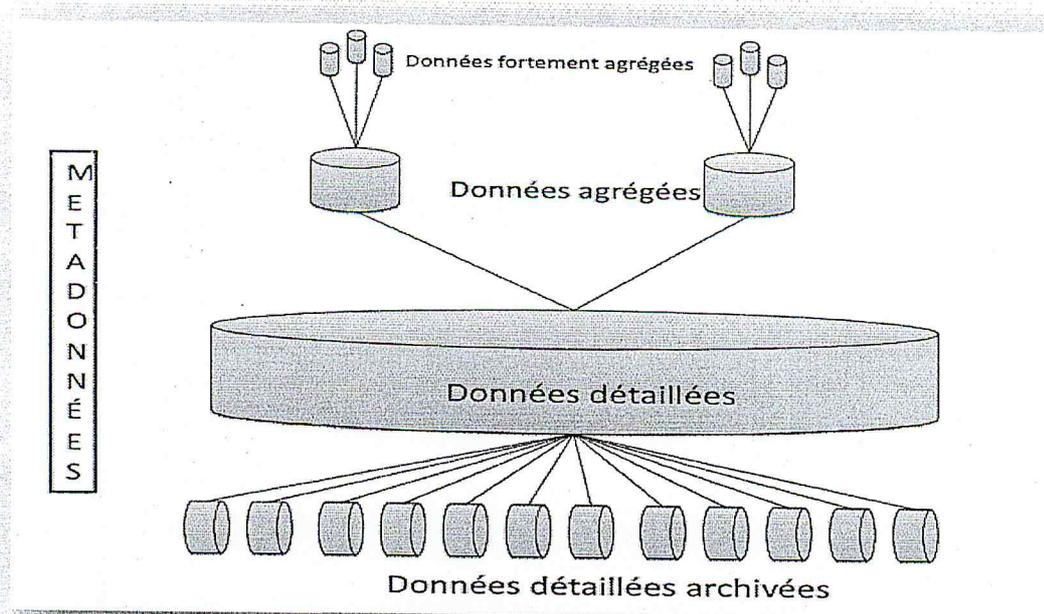


Fig 1.6 : Structure des données d'un Datawarehouse [11]

Données détaillées : ce sont les données qui reflètent les événements les plus récents, Fréquemment consultées, généralement volumineuses car elles sont d'un niveau détaillé.

Données détaillées archivées : anciennes données rarement sollicitées, généralement Stockées dans un disque de stockage de masse, peu coûteux, à un même niveau de détail que les données détaillées.

Données agrégées : données agrégées à partir des données détaillées.

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

Meta données : ce sont les informations relatives à la structure des données, les méthodes D'agrégation et le lien entre les données opérationnelles et celles du DATAWAREHOUSE. Les métadonnées doivent renseigner sur :

- Le modèle de données,
- La structure des données telle qu'elle est vue par les développeurs,
- La structure des données telle qu'elle est vue par les utilisateurs,
- Les sources des données,
- Les transformations nécessaires,
- Suivi des alimentations.

13. Conclusion :

Le système informatique orienté « Décisionnel » dans une entreprise constitue un atout considérable pour le management, il met à la disposition de ce dernier un réservoir immense de données historisées concernant l'ensemble des activités de l'entreprise et permet à travers des requêtes simples d'élaborer des rapports présents selon différents axes d'analyse en fonction des besoins du management.

Dans ce chapitre, nous avons présenté l'une des approches du système décisionnel, le DATAWAREHOUSE avec tous ces différents aspects : l'architecture, la structure et les objectifs de DATAWAREHOUSE.

Dans le chapitre suivant, nous allons présenter Le concept OLAP, ainsi que la modélisation multidimensionnelle, qui représente les techniques les mieux adaptées pour la présentation des données, ainsi que leurs conception

CHAPITRE II
OLAP ET MODELISATION
MULTIDIMENSIONNELLE

1-Introduction

Dans ce chapitre, nous allons tout d'abord présenter et définir le concept OLAP « On Line Analytical Processing », en suite, nous énumérons ces douze règles, que tout système décisionnel doit suivre et respecter. Mais aussi, Nous verrons les différences qui existent entre les deux concepts OLAP et OLTP « On Line transaction Processing ».

En deuxième lieu, nous explorons les différents concepts de la modélisation multidimensionnelle tout en présentant la modélisation conceptuelle, la phase primordiale dans le processus de réalisation de n'importe quel projet.

2-Présentation et définition du concept OLAP « On Line Analytical Processing » [12]

Le concept OLAP, repose essentiellement 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.

OLAP concerne de ce fait au moins autant le monde des serveurs, voire des structures de stockage, que celui des outils.

3. Les douze-12- règles OLAP [13]

Les 12 règles telles que définies par E.F Codd sont les suivantes :

1-Vue multidimensionnelle

L'utilisateur a l'habitude de raisonner en vue multidimensionnelle comme par exemple lorsqu'il souhaite analyser les ventes par produit mais aussi par région ou par période. Ces modèles permettent des manipulations simples : rotation, pivot ou vues par tranche, analyse de type permutations d'axes (slice and dice) ou en cascade (drill anywhere). (2.2)

2-Transparence du serveur OLAP à différents types de logiciels

Cette transparence se traduit pour l'utilisateur par un complément à ses outils habituels garantissant ainsi sa productivité et sa compétence. Elle s'appuie sur une architecture

ouverte permettant à l'utilisateur d'implanter le système OLAP sans affecter les fonctionnalités du système central.

Par ailleurs, l'utilisateur ne doit pas être concerné par l'intégration des données dans OLAP provenant d'un environnement homogène ou hétérogène.

3- Accessibilité à de nombreuses sources de données

Le système OLAP doit donner accès aux données nécessaires aux analyses demandées. Les outils OLAP doivent avoir leur propre schéma logique de stockage des données physiques hétérogènes, doivent accéder aux données et réaliser n'importe quelle conversion afin de présenter à l'utilisateur une vue simple et cohérente. Ils doivent aussi savoir de quel type de systèmes proviennent les données.

4- Performance du système de Reporting

L'augmentation du nombre de dimensions ou du volume de la base de données ne doit pas entraîner de dégradation visible par l'utilisateur.

5- Architecture Client/Serveur

La plupart des données pour OLAP sont stockées sur des gros systèmes et sont accessibles via des PC. Il est donc nécessaire que les produits OLAP soient capables de travailler dans un environnement Client/Serveur.

6- mensions Génériques

Toutes les dimensions doivent être équivalentes en structure et en calcul. Il ne doit exister qu'une seule structure logique pour toutes les dimensions. Toute fonction qui s'applique à une dimension doit être aussi capable de s'appliquer à une autre dimension.

7- Gestion dynamique des matrices creuses

Le schéma physique des outils OLAP doit s'adapter entièrement au modèle d'analyse spécifique créé pour optimiser la gestion des matrices creuses. En effet, dans une analyse à la fois sur les produits et les régions, tous les produits ne sont pas vendus dans toutes les régions.

8- Support Multi-Utilisateurs

Les outils OLAP doivent supporter les accès concurrents, garantir l'intégrité et la sécurité afin que plusieurs utilisateurs accèdent au même modèle d'analyse.

9- Calculs à travers les dimensions

Les opérations doivent pouvoir s'effectuer sur toutes les dimensions et ne doivent pas faire intervenir l'utilisateur pour définir un calcul hiérarchique.

10- Manipulation intuitive des données

Toute manipulation doit être accomplie via une action directe sur les cellules du modèle. Sans utiliser de menus ou des chemins multiples à travers l'interface utilisateur.;

11- Souplesse et facilité de constitution des rapports :

La création des rapports dans les outils OLAP doit permettre aux utilisateurs de présenter comme ils le désirent des données synthétiques ou des résultats en fonction de l'orientation du modèle.

12- Nombre illimité de niveaux d'agrégation et de dimensions

Dimensions et niveaux d'agrégations illimités.

4. Les Différents Types De Serveurs OLAP [Site 2]

(OLAP) est basé sur le modèle de données multidimensionnelles. Il permet aux gestionnaires et aux analystes d'obtenir un aperçu de l'information grâce à un accès rapide, cohérente et interactive à l'information.

Nous avons quatre types de serveurs OLAP:

- Relational OLAP (ROLAP)
- OLAP multidimensionnelles (MOLAP)
- OLAP hybride (HOLAP)
- Serveurs SQL spécialisés

1. Relational OLAP

serveurs ROLAP sont placés entre relationnel serveur back-end et les outils client frontaux. Pour stocker et gérer les données de l'entrepôt, ROLAP utilise SGBD relationnels ou étendus-relationnel.

ROLAP comprend les éléments suivants:

- La mise en œuvre de la logique de navigation d'agrégation.
- Optimisation pour chaque SGBD back-end.
- outils et services supplémentaires.

Les serveurs OLAP relationnelles sont placés entre relationnel serveur back-end et les outils client frontaux. Pour stocker et gérer les données de l'entrepôt, l'OLAP relationnelle utilise SGBD relationnels ou étendus-relationnel.

ROLAP comprend les éléments suivants:

- La mise en œuvre de la logique de navigation d'agrégation
- Optimisation pour chaque SGBD back-end
- outils et services supplémentaires

Points à retenir

- serveurs ROLAP sont hautement évolutives.
- outils ROLAP analysent de grands volumes de données à travers de multiples dimensions.
- ROLAP entrepose les outils et l'analyse des données très volatiles et changeantes.

Relational OLAP architecture :

ROLAP comprend les composants suivants:

- serveur de base de données
- serveur ROLAP
- outil frontal.

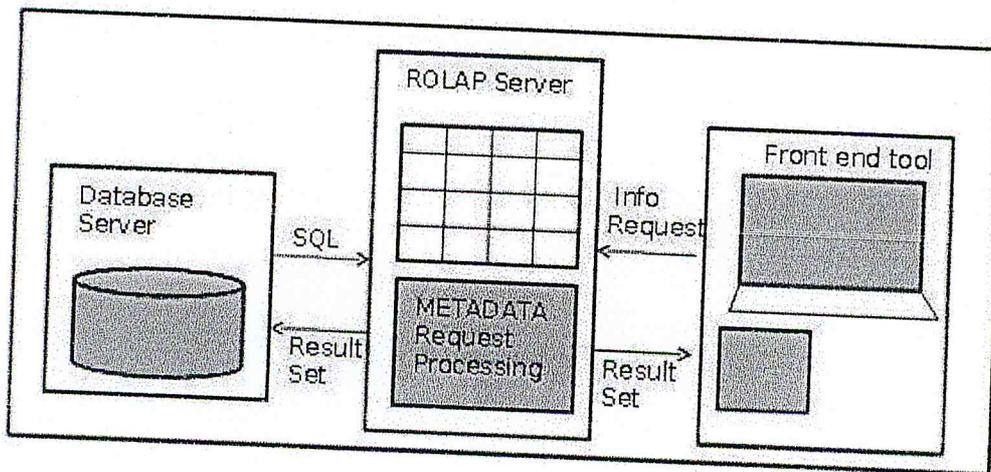


Fig2.1 : Relational OLAP architecture[site2]

Avantages

- serveurs ROLAP peuvent être facilement utilisées avec SGBDR existantes.
- Les données peuvent être stockées efficacement, étant donné qu'aucun faits à zéro peuvent être stockés.
- outils ROLAP ne pas utiliser des cubes de données pré-calculées.
- serveur DSS de micro-stratégie adopte l'approche ROLAP.

Désavantages

- les performances des requêtes pauvres.
- Certaines limitations d'évolutivité en fonction de l'architecture de la technologie qui est utilisée.

2. Multidimensionnel OLAP

MOLAP utilise des moteurs de stockage multidimensionnel tableau à base de vues multidimensionnelles des données. Avec des magasins de données multidimensionnelles, l'utilisation du stockage peut être faible si l'ensemble de données sont rares. Par conséquent, beaucoup serveur MOLAP utilisent deux niveaux de représentation de stockage de données pour gérer des ensembles denses et clairsemées données.

Multidimensional OLAP (MOLAP) utilise des moteurs de stockage multidimensionnel tableau à base de vues multidimensionnelles des données. Avec des magasins de données multidimensionnelles, l'utilisation du stockage peut être faible si l'ensemble de données sont rares. Par conséquent, de nombreux serveurs MOLAP utilisent deux niveaux de représentation de stockage de données pour gérer les ensembles de données denses et clairsemées.

MOLAP architecture :

MOLAP comprend les composants suivants:

- serveur de base de données.
- serveur MOLAP.
- outil frontal.

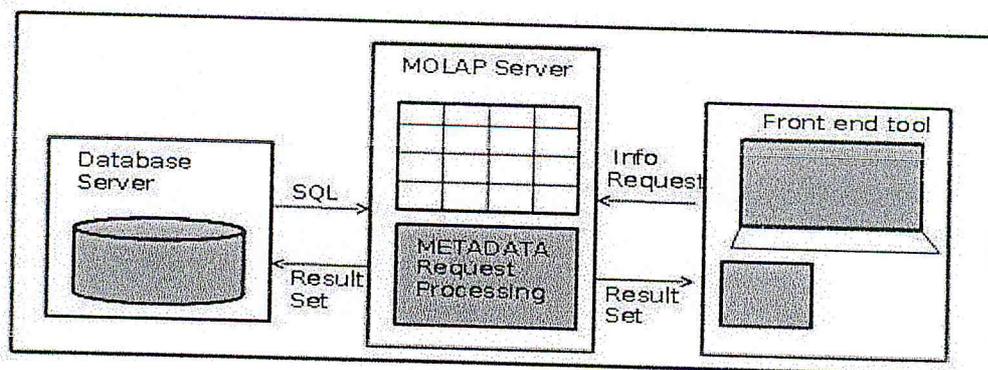


Fig 2.2 : MOLAP architecture[site2]

Avantages

- MOLAP permet la plus rapide indexation des données résumées pré-calculées.
- Aide les utilisateurs connectés à un réseau qui ont besoin d'analyser les grandes données, moins définies.
- Plus facile à utiliser, donc MOLAP est adapté pour les utilisateurs inexpérimentés.

Désavantages

- MOLAP ne sont pas capables de contenir des données détaillées.
- L'utilisation du stockage peut être faible si l'ensemble de données sont rares.

Le tableau ci-dessous présente les principales différences entre MOLAP et ROLAP

MOLAP	ROLAP
La recherche d'information est rapide.	La recherche d'information est relativement lente.
Utilise tableau creux pour stocker les ensembles de données.	Utilise table relationnelle.
MOLAP est le mieux adapté pour les utilisateurs inexpérimentés, car il est très facile à utiliser.	ROLAP est le mieux adapté pour les utilisateurs expérimentés.
Maintient une base de données distincte pour les cubes de données. installation de SGBD est faible.	Il ne peut pas exiger un espace autre que disponible dans l'entrepôt de données. installation de SGBD est forte.

Tab 2.1 : MOLAP vs ROLAP[site2]

3. OLAP hybride (HOLAP)

OLAP hybride est une combinaison des deux ROLAP et MOLAP. Il offre une évolutivité supérieure de ROLAP et de calcul plus rapide de MOLAP. serveurs HOLAP permet de stocker les gros volumes de données d'informations détaillées. Les agrégations sont stockées séparément en magasin MOLAP.

4. Serveurs SQL spécialisés

serveurs SQL spécialisés fournissent la langue de recherche avancée et le support de traitement de requête pour les requêtes SQL plus en étoile et flocon de neige schémas dans un environnement de lecture seule.

5. Les opérations sur OLAP

Comme les serveurs OLAP sont basés sur la vue multidimensionnelle des données, nous allons discuter des opérations OLAP dans des données multidimensionnelles.

Voici la liste des opérations OLAP:

- Roll-up
- Drill-down
- Émincer
- Pivot (rotation)
- Roll-up

Roll-up effectue l'agrégation sur un cube de données dans l'une des façons suivantes:

- En grimpant une hiérarchie de concept pour une dimension
- Par réduction dimensionnelle

Le schéma suivant illustre le fonctionnement de roll-up.

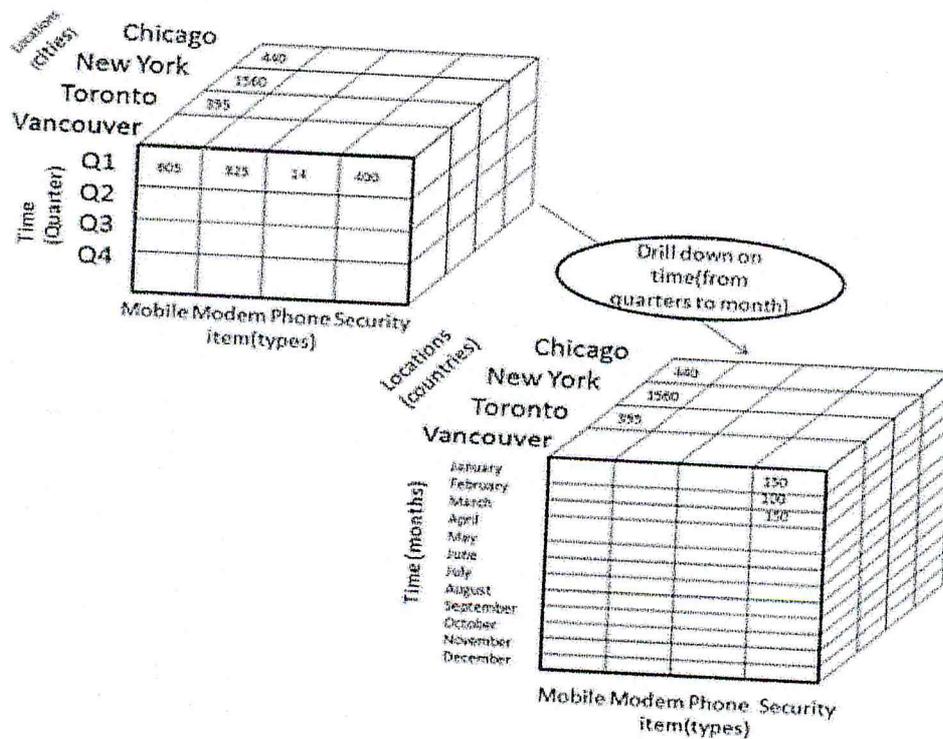


Fig 2.3 Roll-up[site2]

- Roll-up est réalisé en grimpant une hiérarchie de concept pour l'emplacement de la cote.
- Initialement, la hiérarchie de concept était «rue <ville <province <country".
- Sur enroutement, les données sont agrégées en remontant la hiérarchie de l'emplacement à partir du niveau de la ville au niveau du pays.

- Les données sont regroupées dans les villes plutôt que des pays.
- Quand enrroulement est effectué, une ou plusieurs dimensions du cube de données sont supprimés.
- **Drill-down**

Drill-down est l'opération inverse de roll-up. Elle est réalisée par l'une des manières suivantes:

- En prenant une hiérarchie de concept pour une dimension
- En introduisant une nouvelle dimension.

Le schéma suivant illustre la façon dont fonctionne drill-down:

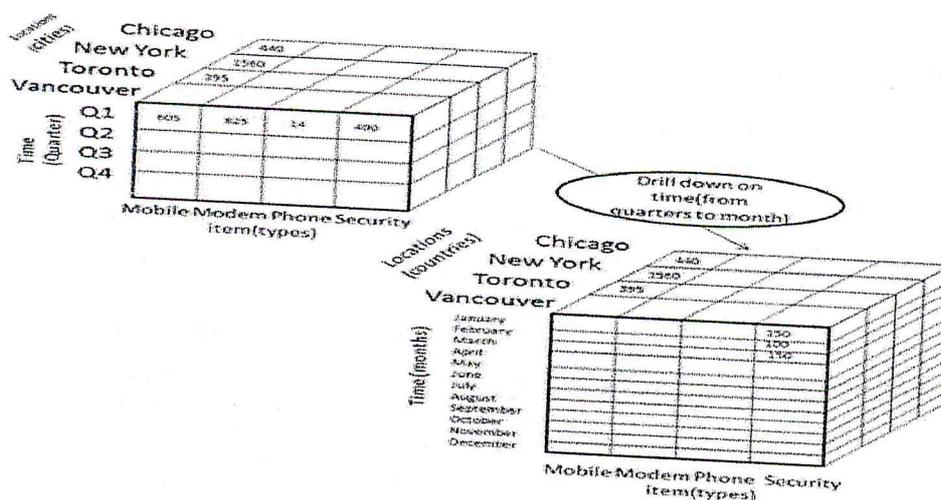


Fig 2.4 : Drill-down[site2]

- Drill-down est effectuée en renforçant une hiérarchie de concept pour le temps de dimension.
- Initialement, la hiérarchie de concept était «jour <mois <trimestre <année.»
- Le forage vers le bas, la dimension temporelle est descendu au niveau du trimestre au niveau du mois.
- Lors de forage vers le bas est réalisée, une ou plusieurs dimensions du cube de données sont ajoutées.
- Il navigue les données à partir des données moins détaillées à des données très détaillées.
- **Slice**

L'opération de tranche sélectionne une dimension particulière à partir d'un cube donné et fournit un nouveau sous-cube. Considérez le diagramme suivant qui montre comment fonctionne la tranche.

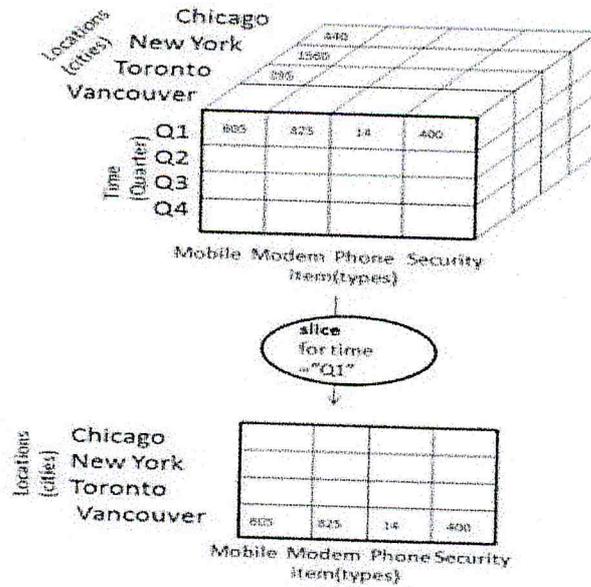


Fig 2.5 : Slice[site2]

- Ici Slice est effectuée pour la dimension «temps» en utilisant le temps de critère = "Q1".
- Il formera un nouveau sous-cube en sélectionnant une ou plusieurs dimensions.
- **Dice**

Dice sélectionne deux ou plusieurs dimensions d'un cube donné et fournit un nouveau sous-cube. Considérez le diagramme suivant, qui montre l'opération de dés.

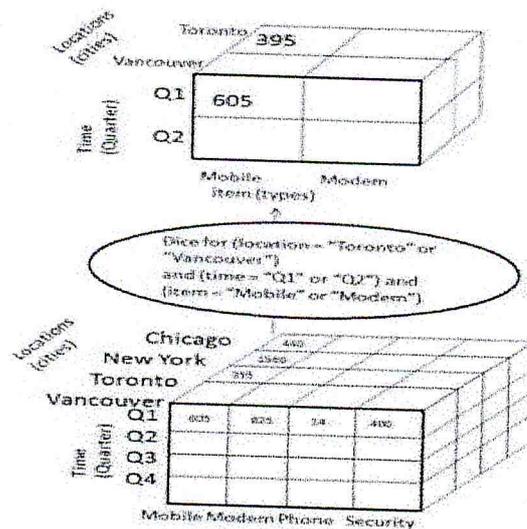


Fig 2.6 : Dice[site2]

L'opération de dés sur le cube sur la base des critères de sélection suivants comprend trois dimensions.

- (Location = "Toronto" ou "Vancouver")
- (Temps = «Q1» ou «Q2»)
- (Item = "Mobile" ou "Modem")
- **Pivot**

L'opération de pivotement est également connu comme la rotation. Il fait tourner les axes de données en mode afin de fournir une autre présentation des données. Considérez le diagramme suivant, qui montre l'opération de pivot.

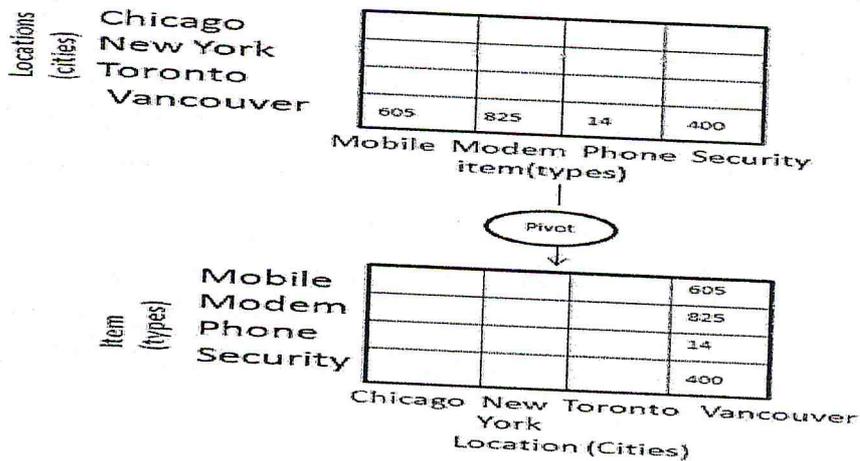


Fig 2.7 : Pivot[site2]

Dans ce l'élément et l'emplacement des axes en 2-D tranche sont mis en rotation.

6. La différence entre OLAP l'analytique et OLTP le transactionnel [14]

- Le système OLTP : Le système transactionnel OLTP (On Line Transaction Processing) est un modèle utilisé par les systèmes de gestion de base de données sources (étape 1 de la BI).

Son objectif est d'insérer, modifier et interroger rapidement et en toute sécurité les données des systèmes opérationnels individuels. Chaque requête effectuée se réfère à de faibles quantités d'informations et aux dernières données insérées pour une visualisation basique de l'information.

- Le système OLAP : Le système OLAP (On Line Analytic Processing) est un outil permettant de structurer l'information dans le Data Warehouse afin de l'analyser (étape 3 de la BI).

Des spécialistes du Datawarehouse définissent le système OLAP comme étant une 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, qui se concrétise par un certain nombre de vendeurs OLAP. Les objectifs principaux de ce systèmes sont de regrouper, organiser les informations provenant des systèmes de données

sources, les intégrer et les stocker dans l'entrepôt (Data Warehouse) pour permettre à l'utilisateur de retrouver ainsi qu'analyser l'information facilement et rapidement. Ce système peut contenir une quantité d'informations nettement plus importante que le système OLTP étant donné qu'il permet de conserver les données historiques. De plus, la mise à jour des données s'effectue en batch. C'est-à-dire que les données sont importées en arrière-plan lorsque le système est très peu sollicité; généralement la nuit. De manière plus simple, le système OLTP permet de visualiser les dernières transactions insérées dans chacun des systèmes ou bases opérationnelles alors que le système OLAP permet d'analyser multidimensionnellement toutes les informations présentes et passées provenant des systèmes opérationnels et stockées dans le Data Warehouse afin de pouvoir être utilisé par les outils de reporting, de datamining, les MIS... Tout ceci en temps réel.

On peut résumer les différences entre les systèmes OLTP et les systèmes OLAP dans le tableau suivant : [Site 2]

Data Warehouse (OLAP)	Base de données opérationnelle (OLTP)
Implique traitement historique de l'information.	Implique traitement au jour le jour.
systèmes OLAP sont utilisés par les travailleurs du savoir tels que les cadres, les gestionnaires et les analystes.	systèmes OLTP sont utilisés par des commis, CBM, ou les professionnels de la base de données.
Utile dans l'analyse de l'entreprise.	Utile dans la gestion de l'entreprise.
Il se concentre sur l'information sur.	Il se concentre sur les données dans.
Basé sur Star Schema, Flocon de neige, Schema et Fact Constellation Schema.	Sur la base de modèle entité-relation.
Contient des données historiques.	Contient des données actuelles.
Fournit des données résumées et consolidées.	Fournit des données primitives et très détaillées.
Fournit résumé et une vue multidimensionnelle des données.	Fournit vue relationnel détaillé et plat des données.
Nombre ou utilisateurs est dans des centaines.	Nombre d'utilisateurs est en milliers.
Nombre d'enregistrements accessibles est en millions.	Nombre d'enregistrements accessibles est en dizaines.
taille de la base de données est de 100 Go à 1 To	taille de la base de données est de 100 Mo à 1 Go.
Très flexible.	Fournit de hautes performances.

Tab 2.2 : OLAP vs OLTP[14]

7. Qu'est-Ce Que La Modélisation Multidimensionnelle [Site 3]

Les bases de données relationnelles modélisées selon les principes classiques de normalisation s'adaptent très mal à un contexte analytique (OLAP). En analyse, l'utilisateur doit disposer d'un modèle relativement intuitif et capable de stocker le résultat de nombreux calculs d'agrégation (ce qui, d'un strict point de vue relationnel, constitue une redondance), tandis que les problématiques classiques de contrôle de saisie et de cohérence sont inexistantes (on suppose que les données de base sont déjà saisies dans une BDD relationnelle cohérente) ou déplacées (on ne saisit plus des données mais des paramètres de calcul ou de *simulation* – voir plus bas).

7.1 Dimensions et indicateurs

La modélisation multidimensionnelle (formalisée notamment par Edgar F. Codd dans un article commandité par Arbor Software) propose donc d'analyser des *indicateurs* numériques (par exemple chiffre d'affaires, nombre d'individus, etc.) dans un contexte précisé par le croisement de plusieurs *dimensions* d'analyse (par exemple temps, géographie, organisation, produits, etc.), généralement présentées sous forme d'arbres hiérarchiques. Au-delà de trois dimensions, cela devient mathématiquement un hypercube (qu'il est évidemment beaucoup plus difficile de représenter graphiquement). Une base de données multidimensionnelle typique peut donc s'envisager comme un hypercube d'une petite dizaine de dimensions comprenant plusieurs millions de cellules (on parle plus couramment de *cube OLAP*).

7.2 Représentation pour l'utilisateur

Classiquement, les requêtes formulées par les utilisateurs s'expriment et se représentent sous forme de tableaux croisés :

		Janvier	Février	Mars	Trim1
France	Bières	70	70	80	220
	Vins	100	110	90	300
	Total	170	180	170	520
UK	Bières	250	220	240	710
	Vins	50	40	60	150
	Total	300	260	300	860
Total Europe		470	440	470	1380

Fig 2.8 Présentation des cubes en tableau[site3]

Cette représentation en tableau croisé (qui se traduit ensuite facilement en divers graphiques) présente de nombreux avantages pour l'utilisateur :

- elle est directement compréhensible
- elle est facilement manipulable par un analyste qui souhaite changer son point de vue sur les données (zoom, pivot, etc.)
- elle peut être directement comprise par un outil OLAP, libérant ainsi l'utilisateur non-informaticien du besoin d'apprendre un langage d'interrogation (SQL, MDX).

8. Modélisation Conceptuelle

La modélisation multidimensionnelle utilise deux concepts essentiels qui consistent en fait et dimension.

8.1 Définition du concept de fait [15]

Le sujet analysé est représenté par le concept de fait.

La table de faits contient des mesures (ex: unités_ vendues) et les clés externes faisant référence à chaque table de dimension

Une mesure est un élément de donnée sur lequel portent les analyses, en fonction des différentes dimensions Ex : coût des travaux, nombre d'accidents, ventes

Un fait représente la valeur d'une mesure, mesurée ou calculée, selon un membre de chacune des dimensions Exemple : «250 000 euros » est un fait qui exprime la valeur de la mesure « coût des travaux » pour le membre « 2002 » du niveau aimé de la dimension « temps » et le membre « Versailles » du niveau « ville » de la dimension « découpage administratif » La table de faits

Les mesures sont numériques pour permettre de résumer un grand nombre d'enregistrements en quelques enregistrements (on peut les additionner, les dénombrer ou bien calculer le minimum, le maximum ou la moyenne). Elles sont valorisées de façon continue car il est important de ne pas valoriser le fait avec des valeurs nulles, mais aussi souvent additives afin de pouvoir les combiner au moyen d'opérateurs arithmétiques.

8.2 Définition du concept de dimension

Les tables de dimension contiennent des rangées (tuples ,enregistrements). Chaque rangée est constituée de colonnes (champs). Dans une base de données standard, le premier champ de chaque enregistrement est la clé primaire, l'unique identifiant de l'enregistrement. Par contre, le modèle de données multidimensionnel est un tableau à n dimensions (hypercube). Chaque dimension possède une hiérarchie associée de niveaux de consolidation. Chaque position dans un tableau multidimensionnel, correspondant à une intersection de toutes les dimensions est appelée une cellule.

Ces dimensions peuvent être affinées, décomposées en hiérarchies, afin de permettre à l'utilisateur d'examiner ses indicateurs à différents niveaux de détail, de " descendre " dans les données, allant du niveau global au niveau le plus fin. Par exemple, dans la dimension temps, une date pourra être décomposée en <année, mois, semaine, jour>, et qui représente Un axe d'analyse essentiel. On aura alors une vision pyramidale des données, la base de la pyramide représentant le niveau le plus détaillé et le haut le niveau le plus global.

CHAPITRE III
PROCESSUS FONCTIONNELS
DU DATAWAREHOUSE

1. Introduction

Dans ce chapitre, nous aborderons successivement en détail les processus fonctionnels qui permettront le bon fonctionnement de l'entrepôt, il s'agit pour commencer : le processus d'alimentation de l'entrepôt de données « ETL », (Extraction, Transfert, Loading) permettant ainsi d'extraire les données du système source, les transformées par la suite pour en fin les chargées dans l'entrepôt pour but d'analyse par les utilisateurs de l'activité, mais aussi le stockage des données. Et pour finir, le processus d'administration des données de l'entrepôt, car Comme tout autre système informatique, un Data Warehouse s'administre. Dès la phase de conception de l'architecture, il faut penser à l'administration des données.

2. Processus d'alimentation E.T.L [5]

Les données du Data Warehouse sont, pour la plupart, issues des différentes sources de données opérationnelles de l'entreprise. Des solutions logicielles sont alors nécessaires à leur intégration et à leur homogénéisation. Celles-ci peuvent aller de l'écriture de batchs à l'utilisation de logiciels spécialisés dans l'extraction et la transformation d'informations (ETI, Prism, Carleton, ...). Ces outils ont pour objet de s'assurer de la cohérence des données du Data Warehouse et d'homogénéiser les différents foïnats trouvés dans les bases de données opérationnelles. Les solutions de réplication sont souvent citées comme pouvant répondre à ce besoin. Très liées à un éditeur et à une technologie, elles ne sont en fait adaptées qu'à un très petit nombre de cas, où les données sont issues de sources homogènes et n'ont pas besoin d'être transformées.

Alimenter le Data Warehouse est la difficulté technique majeure et la plus coûteuse.

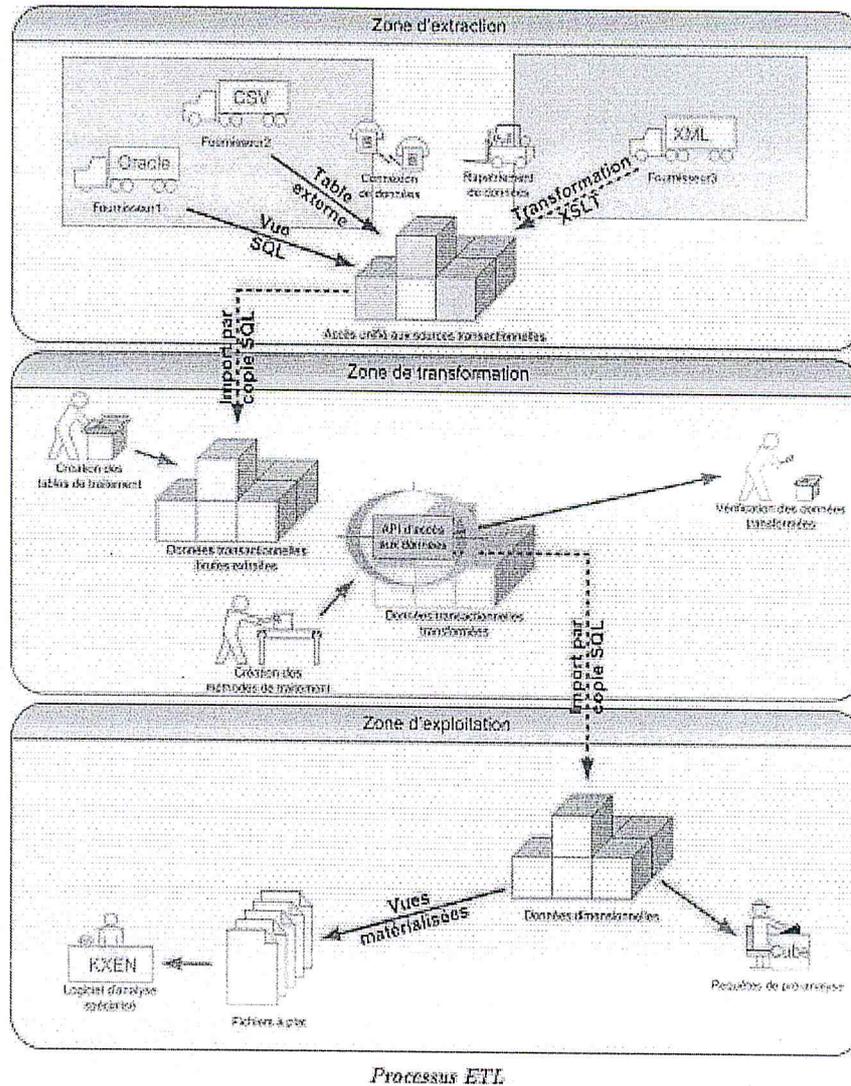


Fig 3.1 : Processus ETL [16]

2.1. Les phases de l'alimentation

Les phases de l'alimentation du Data Warehouse sont les suivantes :

Découvrir quelles sont les données à faire migrer.

L'acquisition des données se déroule en trois phases : l'extraction, la transformation et le chargement. La découverte des données consiste les localiser dans le système opérationnel et à prendre les plus judicieuses.

Un mauvais choix des données à extraire va complexifier les phases suivantes de l'alimentation

8.3 Granularité des dimensions

Lors du processus OLAP, les données sont généralement analysées en partant d'un faible niveau de détail vers des données plus détaillées pour «*forer vers le bas*». Pour définir ces différents niveaux de détail, chaque dimension est munie d'une (ou plusieurs) hiérarchie(s)

des paramètres. La hiérarchie ou la granularité sert lors des analyses pour restreindre ou accroître les niveaux de détail de l'analyse.

- **Définition de la granularité**

Une hiérarchie organise les paramètres d'une dimension selon une relation «*est_plus_fin*» conformément à leur niveau de détail.

8.4 Modèles en étoile, en flocon et en constellation

A partir du fait et des dimensions, il est possible d'établir une structure de données simple qui correspond au besoin de la modélisation multidimensionnelle. Cette structure est constituée du fait central et des dimensions.

Ce modèle représente visuellement une étoile, on parle de modèle en étoile.

9. Conclusion

Dans ce chapitre, nous avons présenté le concept OLAP, qui représente un outil fondamental pour l'analyse multidimensionnelle, parce qu'il associe des mécanismes de navigation aux données, permet d'effectuer des analyses de manière interactive dans un cadre prédéfini, limitant dès lors l'autonomie potentielle de l'utilisateur. on a vu aussi les concepts liés à la modélisation multidimensionnelle notamment la modélisation conceptuelle et logique.

Nous allons parler dans le chapitre qui suit des processus fonctionnels du Datawarehouse notamment l'ETL, qui est le processus d'extraction, de transformation et de chargement de données.

2.1.1. L'extraction des données.

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

Le processus d'alimentation doit automatiquement et périodiquement rafraîchir les données et de préférence uniquement celles qui ont été modifiées afin d'optimiser les performances. On réalisera des extractions différentielles en marquant les données (date dans un fichier LOG). Cette fonctionnalité devient importante lorsque le volume de données devient important. L'intégrité des données est indispensable et nécessite la synchronisation des différents processus d'extraction. Les problèmes liés à ce nécessaire synchronisation peuvent être complexes, soit fonctionnellement, soit techniquement dans des environnements très hétérogènes. Un autre problème est de traiter les données externes. Il faut maintenir une surveillance du système d'information pour pouvoir les identifier et s'assurer que ce sont les bonnes données qui sont recensées. De plus, la forme des données externes, qui est souvent totalement anarchique accentue la difficulté. Pour être utiles, ces données nécessitent un reformatage pour pouvoir les incorporer dans une forme exploitable pour l'entreprise. Enfin le troisième problème vient de l'apparition imprévisible de ces données qui les rend difficiles à capter. En conséquence, l'outil d'extraction doit attaquer toutes sortes de sources de données sans être perturbé et s'adapter aux futures. Pour extraire les données sources, il y a plusieurs technologies utilisables : des passerelles, fournies principalement par les éditeurs de bases de données. Ces passerelles sont généralement insuffisantes car elles sont mal adaptées

aux processus de transformation complexes ; des utilitaires de réplication, utilisables si les systèmes de production et décisionnel sont homogènes et si la transformation à appliquer aux données est légère ; des outils spécifiques d'extraction. Ces outils sont certainement la solution opérationnelle au problème de l'extraction, mais leur prix relativement élevé est un frein à leur utilisation dans les premières applications.

2.1.2. La transformation des données

Le nettoyage des données est une discipline sur laquelle de nombreux éditeurs travaillent actuellement. Outre la qualité des données qu'ils permettent d'auditer et éventuellement d'améliorer, les outils de nettoyage permettent de supprimer les doublons dans les fichiers. Il s'agit à ce stade d'appliquer des filtres prédéfinis sur les données afin d'attribuer des valeurs cohérentes aux variables mal ou non renseignées ou encore d'harmoniser les formats (date : jj/mm/aaaa). On peut également avoir à convertir les données d'un format EBCDIC vers ASCII. Dernier exemple, des données du système opérationnel doivent être agrégées ou - calculées avant leur chargement dans la base décisionnelle.

Il faut également pouvoir associer des champs sources avec des champs cibles. Il existe plusieurs niveaux de complexité pour ces associations (cardinalités 1-1, 1-N, N-1, N-N), comme par exemple : le transfert du "nom du client" vers un champ cible, la décomposition d'une "adresse" vers les champs "numéro", "rue", "ville" ou l'inverse.

Certains outils peuvent également réaliser des analyses lexicales des champs sources. Ils seront donc capables de comprendre que les champs suivants signifient la même chose "Boulvd", "Bd"», "Boulevard".

En complément, on trouve des outils d'audit et d'analyse pour assurer le suivi du processus afin notamment de contrôler les rejets.

2.1.3. Le chargement des données

Le chargement est la dernière phase de l'alimentation du Data Warehouse. C'est une phase délicate notamment lorsque les volumes sont importants. Pour obtenir de bonnes performances en chargement, il est impératif de maîtriser les structures du SGBD (tables et index synchronisation des chargements de la source vers la cible est un problème majeur. Pour réaliser ce transfert des données, on peut utiliser le transfert de fichiers ou le transfert de bases à bases. Le transfert de fichiers consiste à créer au moment de l'extraction des données, un ensemble de fichiers qui seront transférés sur le système cible afin d'alimenter la base décisionnelle. La transformation des données s'effectuera alors soit à la constitution des fichiers, soit à leur arrivée sur la cible. Le transfert de base à base est plus complexe, en effet chaque donnée est extraite de la

base source, nettoyée si nécessaire, et transférée sur la base cible. Des logiciels contrôlent souvent les phases d'extraction, de transport et de chargement. Ils créent un enchaînement automatisé et gèrent les mises à jour mais ils sont rarement capables de s'occuper également du nettoyage des données, Il faut donc souvent combiner les outils pour alimenter un Data Warehouse

.3. Stockage des données [Site 4]

Ces besoins en terme d'analyse multidimensionnelle peuvent influencer d'une part sur le modèle physique, d'autre part sur la vision que les utilisateurs ont des données. Les données peuvent être stockées soit sous forme multidimensionnelle, soit sous forme relationnelle. Les bases multidimensionnelles sont conçues pour optimiser l'analyse d'indicateurs (CA) par rapport aux dimensions qui leur sont associées (temps, produit, région). Elles sont de moins en moins exigeantes en terme de volume de stockage (technique de compression). Les bases relationnelles sont conçues à l'origine pour le décisionnel. Elles sont plus ouvertes et s'adaptent à de nombreuses plates formes.

3.1. Le stockage des données externes

Les données qui ne sont pas générées par les systèmes informatiques de l'entreprise sont appelées des données externes et sont introduites sans formatage préalable. Elles ne sont donc pas forcément structurées. Typiquement, les données externes qui sont dans l'entrepôt de données arrivent dans l'entreprise par l'intermédiaire d'un poste client. Elles sont saisies manuellement dans un tableur, ce qui laisse peu de chances d'identifier d'où vient les données ainsi que leurs contenus. D'autres types de données ont fait leur apparition, il s'agit du son et de l'image qui sont, sans compression, particulièrement volumineux. L'entrepôt de données est le lieu idéal pour stocker des données externes et non structurées. Si ces données ne sont pas stockées dans un endroit bien défini, on ne peut plus savoir d'où viennent les données et on laisse la porte ouverte aux éventuelles redondances. Les données externes peuvent être stockées au sein même de l'entrepôt de données. Toutefois, pour diminuer le coût de stockage de ces données, on préférera les stocker séparément, sur un support économique. A la détection des données externes, on choisira de rediriger le stockage de ces données vers des périphériques d'archivage dédiés. Le composant de base support du stockage

est le SGBD. Outre le stockage, le SGBD doit être spécifiquement adapté aux caractéristiques de l'accès décisionnel. Ces technologies sont principalement liées à la parallélisation des requêtes et à diverses optimisations proposées pour accélérer les sélections et les regroupements ensemblistes.

3.2. Optimisation physique

Du fait de l'importance de l'historique dans un Data Warehouse, la structuration physique des données est également importante. Parmi les optimisations proposées par les éditeurs, on notera les index bitmaps (Sybase , Oracle), ou la génération assistée des agrégations (HP, Informix, Oracle, etc.) ainsi que le partitionnement des tables.

3.2.1. Agrégation de tables

Les requêtes faisant appel à des données agrégées forment 80 % des demandes effectuées dans un système décisionnel. La construction de tables agrégées permet d'améliorer très nettement les temps de réponse de ces requêtes. Par exemple, on crée une table des ventes mensuelles représentant la somme des ventes journalières par magasin et cela sur chacun des mois.

Cette technique d'optimisation trouve ses limites lorsque les bases des données atteignent un volume important. Autre inconvénient, l'administration est complexe.

3.2.2. Partitionnement de tables

Un partitionnement physique des tables en plus petites unités selon le critère temps apporte des performances stables dans le temps, des facilités pour les reprises, les indexations, restructurations et l'archivage. Notre table des ventes sera éclatée en 12 tables, soit une par mois.

3.3. Les méta- données

Les métadonnées sont des données sur les données. Les métadonnées représentent la totalité des informations concernant le Data Warehouse et les processus associés.

Elles sont intégrées dans un référentiel. La méta- donnée va également servir de mémoire pour les données externes en recueillant les informations suivantes : l'identifiant du document, sa date d'entrée dans l'entrepôt de données, la description du

document, l'origine du document, la date d'origine du document, les mots indexés du document, la date d'épuration, la place physique du document, sa longueur...

Dans le cadre d'outils OLAP, les méta- données sont, pour l'essentiel, la structure multidimensionnelle elle-même. Pour un outil d'aide à la décision, elles ont pour vocation de présenter à l'utilisateur des éléments qui lui sont familiers et qu'il pourra donc analyser facilement, de manière intuitive.

Dans le cas d'outil ROLAP, les métadonnées ont pour but de transformer le modèle de stockage relationnel en une vision utilisateur multidimensionnelle.

Le mode de stockage privilégié pour les méta- données est le serveur relationnel. Elles sont ainsi stockées une seule fois et accessibles par l'ensemble des utilisateurs. La mise à jour de l'environnement client peut alors s'effectuer de deux manières : soit en rapatriant à chaque connexion le contenu des méta- données, soit en ramenant en local les méta- données et en les mettant à jour de manière automatique lorsqu'un changement est survenu. Cette deuxième solution est plus performante, un nouveau catalogue n'étant ramené qu'une fois sur le poste local.

4. L'administration des données [Site 4]

C'est une des fonctions les plus importantes du Data Warehouse. Cette fonction est d'autant plus importante que le Data Warehouse évolue au fur et à mesure de son utilisation. La mise en place d'un Data Warehouse doit s'accompagner de celle d'un référentiel de données, permettant de décrire, stocker et diffuser les métadonnées associées.

Cette mise en place passe par l'organisation d'une fonction d'administration des données à plusieurs niveaux, par la définition de normes et de procédure d'administration des référentiels.

1. Les métadonnées

Elles représentent toutes les informations nécessaires à l'accès, à la compréhension et à l'exploitation des données du Data Warehouse.

Type d'information I	Signification
Sémantique	Que signifie la donnée
Origine	D'où vient-elle, où, par qui est-elle créé ou mise à jour
Règle de calcul	Règle de calcul, de gestion
Règle d'agrégation	Périmètre de consolidation
Stockage, Format	Où, comment est-elle stockée, sous quel format
Utilisation	Programmes informatiques qui l'utilisent, Machines : comment et sur lesquelles, à disposition

Tab3.1 : Définition d'une métadonnée[site 4]

La donnée est forcément liée à d'autres objets du système d'information. Il est donc nécessaire de représenter, décrire et stocker les interactions avec d'autres données.

Généralement, pour représenter les métadonnées, on utilise l'informatique. Par exemple les modèles de données MERISE : et modèle logique de données.

2. Le référentiel du data Warehouse

Le référentiel du Data Warehouse est l'ensemble des outils nécessaires à la mise en œuvre de la fonction d'administration de données.

Les objectifs de l'administration de données du Data Warehouse sont :

- Assurer la cohésion du système :
 - Respecter la cohérence et la fiabilité des informations.
 - Unifier la représentation des données.
 - Respecter la cohérence des concepts.
 - Vérifier la non redondance des informations.
- Simplifier techniquement les systèmes d'information :
 - Diminuer le nombre de fichiers.
 - Unifier la saisie et le stockage des informations.
 - Organiser les mises à jour et la diffusion des informations.

Un référentiel de données pour le Data Warehouse est un référentiel de données dans lequel est décrits l'organisation et la localisation des données, ainsi que les règles de consolidation des données agrégées et historisées.

Il est conçu de manière à collecter l'ensemble des modèles de données nécessaires à la construction et à l'exploitation du Data Warehouse.

La construction du référentiel est un projet qu'il est nécessaire de prévoir en parallèle à celui du data Warehouse.

3. La mise en œuvre

Il n'existe pas une seule manière de le faire. La problématique doit être considérée sous deux aspects : organisationnel et technique.

Acheter une solution clé en main est très risquée : les fonctionnalités de l'outil conditionnent alors l'organisation du projet.

Quelque soit l'outil choisi, il faut pouvoir le faire évoluer afin de s'adapter à de nouveaux besoins.

4. Sécurité

Le Data Warehouse a pour vocation de laisser à l'utilisateur une totale autonomie en ce qui concerne la recherche et l'analyse des données. Cette liberté doit cependant être souvent restreinte, notamment pour des raisons de sécurité. L'outil doit donc permettre d'adapter l'environnement de travail à l'utilisateur qui s'y connecte selon sa fonction et donc ses droits. Afin de mener à bien cette politique de sécurité, les notions d'utilisateur et de groupe d'utilisateur sont indispensables. Cette notion doit néanmoins être dissociée de la notion d'utilisateur du serveur de données. Il est en effet préférable que l'utilisateur se connecte directement à l'outil d'aide à la décision et accède de manière transparente au serveur de données sans en connaître le nom d'utilisateur et le mot de passe. Il n'y a ainsi pas de risque que l'utilisateur se connecte au serveur par d'autres biais. Ce risque est d'autant plus important que l'utilisateur trouve le plus souvent sur son poste de travail les moyens de se connecter à un serveur de données. Il peut par exemple se connecter à partir d'Excel à un serveur de données, via le driver ODBC.

Plus les autorisations pourront être définies à un niveau fin, plus l'environnement de travail pourra être facilement adapté à chacun des profils d'utilisateur. Au delà du catalogue dans sa globalité, l'outil doit donc permettre de définir des droits sur chacun de ses éléments en restreignant le nombre de tables ou de colonnes disponibles ou, par un critère de recherche, le nombre de lignes accessibles.

Au delà des droits concernant l'accès aux données, la possibilité d'affecter des privilèges sur les requêtes et rapports prédéfinis est un facteur essentiel dans la définition d'une stratégie de diffusion.

5. Gestion des performances

La possibilité de limiter les accès aux ressources permet à l'administrateur de mieux contrôler les requêtes émises par l'utilisateur. Ce besoin est surtout essentiel pour les outils d'infocentre, qui peuvent offrir à l'utilisateur un espace de liberté et d'autonomie important. En effet contrairement aux outils basés sur un modèle multidimensionnel, les utilisateurs accèdent aux informations non selon un schéma prédéfini mais selon leurs besoins. Ils génèrent ainsi les requêtes qu'ils veulent et n'accèdent pas toujours à un serveur optimisé en conséquence.

Ce contrôle des ressources peut prendre plusieurs formes :

- Limitation du temps d'exécution ; cette fonctionnalité ne sera néanmoins pas gérée de manière optimale par l'outil qui, à moins d'avoir déjà effectué une requête identique, n'a pas moyen de connaître à priori le temps qu'elle prendra. Certains serveurs (Ingres et Oracle notamment) offrent cette fonctionnalité. L'idéal est donc d'avoir un outil capable de demander au serveur le temps que va mettre une requête et d'agir ensuite en conséquence (par exemple, refuser la demande selon l'utilisateur). Sinon, l'outil peut permettre l'abandon de requêtes trop longues.
- limitation du nombre de lignes récupérées ; l'outil peut soit permettre de ramener les lignes par paquets, soit ne rapatrier qu'un certain nombre d'enregistrements, soit abandonner une requête ramenant un volume trop important de données. L'outil, en cas de dépassement des limites, peut soit prévenir l'utilisateur et lui demander si il souhaite continuer, soit abandonner la demande.
- Le contrôle du corps des requêtes permet de limiter considérablement le nombre de demandes coûteuses. Ce contrôle peut s'exercer sur différents paramètres : le nombre de tables interrogées ; les jointures, l'administrateur peut notamment interdire tout produit cartésien. Il peut de plus être intéressant d'imposer d'utiliser que les colonnes indexées ou d'interdire un nombre trop important de théta jointures, ...

Ce besoin de contrôler les performances est dû à l'autonomie laissée aux utilisateurs par les outils d'infocentre. En effet, l'administrateur n'a que très peu de maîtrise sur ce que vont faire les utilisateurs et le modèle de données ne peut être dès le départ optimisé pour un type d'utilisation particulier. Des fonctions d'audit permettent de remédier à cela en donnant à l'utilisateur une vision des demandes les plus et les moins fréquemment utilisées. Il peut alors paramétrer et optimiser son environnement (serveur, client) en conséquence. L'outil doit donc offrir la possibilité de journaliser le texte des requêtes, leur durée d'exécution, le nom de l'utilisateur les ayant envoyées.

6. Batches

La possibilité d'effectuer des requêtes en batch est une fonctionnalité essentielle dans un contexte de Data Warehouse où une requête peut avoir besoin de manipuler un volume important de données. Un scheduler, planificateur, peut alors permettre d'envoyer une requête à un instant donné ou à intervalles réguliers. Imposer des plages horaires aux utilisateurs pour ces requêtes coûteuses est alors également une nécessité. L'administrateur peut gérer les priorités d'après les requêtes ou les utilisateurs. Le scheduler doit permettre non seulement d'envoyer une requête mais une série de requêtes. Un mécanisme d'alertes

peut alors avertir l'utilisateur si des valeurs sont exceptionnelles. On passe alors d'une logique pull (l'utilisateur part à la recherche de l'information pertinente) à une logique push (l'information vient vers l'utilisateur quand elle est jugée pertinente). Outre les requêtes définies avec l'outil d'aide à la décision, le scheduler peut également permettre de lancer des requêtes d'administration à un instant donné, pendant des plages horaires peu occupées.

5. Conclusion

Dans ce chapitre nous avons présenté les différentes étapes du processus d'alimentation de l'entrepôt, il s'agit de l'étape d'extraction des données brutes qui proviennent du système source, pour en suite les transformées pour les rendre conforme au format de l'entrepôt à fin de les chargées dans celui-ci pour but d'analyse. Mais aussi, nous avons défini le stockage des données ainsi que leur administration dans l'entrepôt.

CHAPITRE IV
CONCEPTION ET
DEPLOIEMENT DU
SYSTEME.

1. Introduction

Aucune méthode de conception n'avait pénétré le secteur industriel en 1999, mais elles faisaient l'objet de sujets de recherche. Aussi, une méthode originale de conception des datamarts a été créée pour aboutir à la modélisation physique. Cette méthode consiste à trouver des combinaisons d'agrégations (ou de niveau de détail) des axes multidimensionnels autour d'une table de faits plus ou moins agrégée, tout en répondant aux impératifs fonctionnels, de sécurité et aux contraintes de volumétrie. Pour cela, des Datamarts de premier niveau ont été identifiés, construits à partir des croisements d'axe et uniquement d'un point de vue fonctionnel. A partir de ceux ci, des datamarts de second niveau ont été construits pour couvrir les autres aspects, en particulier la volumétrie (par agrégations).

Après avoir évoqué tous les aspects liés aux entrepôts de données dans les chapitres précédents, et avoir présenté l'environnement de l'entreprise nous pouvons maintenant commencer l'élaboration de l'entrepôt de données.

Dans ce chapitre nous évoquerons la conception logique et physique du système, qui se manifeste par deux grandes étapes importantes :

1. l'étape de la modélisation conceptuelle de l'entrepôt de données de la fonction « Suivi des projets ».
2. l'étape de la réalisation de l'entrepôt de données métier « Suivi des projets ».

Dans la première partie de ce chapitre nous aborderons l'étape une, il s'agit de la modélisation conceptuelle de l'entrepôt de données de la fonction « Suivi des projets »,

Qui passe par deux sous étapes :

1. L'étude préalable qui définit les besoins, les objectifs des utilisateurs et précise la démarche à suivre.
- . la conception du modèle dimensionnel de données qui représente l'entrepôt conceptuellement et logiquement.

Dans la deuxième partie de ce chapitre nous entamerons l'étape de mise en œuvre de cet entrepôt de données, par quatre sous étapes :

1. construction de l'entrepôt, la base de l'entrepôt.
2. construction des cubes OLAP, la base multidimensionnelle.
3. construction de la zone d'alimentation de l'entrepôt.
4. construction du portail de restitution.

Et pour finir, nous expliquerons le déploiement de notre Système.

2. Approches de mise en place de DW [Site 5]

Il existe plusieurs approches pour mettre en place un DW. Par contre seulement trois approches sont communes. il s'agit de l'approche "Top-Down" prônée par Inmon, l'approche "Bottom-up" de Kimball et de l'approche "Hybride" qui dérivent des deux premières approches.

Si vous recherchez la meilleure approche, il n'existe pas de meilleure approche, mais plutôt recherchez l'approche la plus adaptée à votre contexte.

2.1 Top-Down (Bill Inmon et le CIF)

- L'emphase est mise sur le DW.
- Commence par concevoir un modèle de DW au niveau de l'entreprise.
- Déploies une architecture multi tiers composée de staging area, le DW, et les data mart dépendants.
- Le staging area est permanent.
- Le DW est orienté entreprise; les data marts sont orientés processus.
- Le DW contient des données atomiques; Les data marts contiennent les données agrégées.
- Le DW utilise un modèle de données normalisé de toute l'entreprise; Les data marts utilise des modèles dimensionnels orientés sujet.
- Les utilisateurs peuvent effectuer des requêtes sur le DW et les data marts.

2.2 Bottom-Up (Ralph Kimball et le Bus Architecture)

- L'emphase est mise sur les datamarts.
- Commence par concevoir un modèle dimensionnel pour le datamart.
- Utilise une architecture qui consiste en un staging area et les datamarts.
- Le staging area est en général non permanent, mais il peut devenir permanent pour implanter l'architecture en BUS (Dimensions et faits conformes)
- Les datamarts contiennent les données atomiques et les données agrégées.
- Les datamarts peuvent fournir une vue entreprise ou processus.
- Un datamart consiste en un seul star schema physique.

- Les datamarts sont implantés d'une façon incrémentale et intégrée en utilisant les dimensions conformes.
- Les utilisateur ne peuvent effectuer des requêtes sur le staging area.

2.3 Hybride

- L'emphase est sur le DW et les datamarts; utilise les deux approches "top-down" et "bottom-up".
- Commence par concevoir un modèle de données de l'entreprise en même temps que les modèles spécifiques.
- Passe 2-3 semaines à créer un modèle normalisé d'entreprise de haut niveau ;
génère les modèles des premiers datamarts.
- Charge les datamarts avec les données atomiques en utilisant un staging area temporaire.
- Les modèles des datamarts sont composés d'un ou plusieurs star schémas.
- Utilise un outil ETL pour charger les datamarts et pour échanger le métadatas avec ces derniers.
- Charge le DW à partir des datamarts lorsqu'il y'a besoin de faire des requêtes à travers plusieurs datamarts en même temps.

3. Approche de Ralph Kimball

Issue des entreprises, les entrepôts de données sont aujourd'hui un thème de recherche à part entière. L'approche multidimensionnelle, cubes OLAP, architecture de data warehousing, font l'objet de nombreux travaux. Nous allons présenter une méthodologie adaptée pour l'implémentation de notre système, Celle-ci constitue la trame de ce mémoire et définit la démarche à suivre. Nous avons opté pour Une approche basée sur la gestion de projet par prototypage, que le célèbre chercheur Ralph Kimball propose et privilégie, car elle repose essentiellement sur une modélisation multidimensionnelle des données. Le schéma de cycle de vie dimensionnel proposé par R. Kimball matérialise l'enchaînement des grandes étapes de l'implémentation d'un entrepôt de données. Il identifie le séquençement des tâches de haut niveau et souligne les contraintes entre ces différentes actions tant au niveau technologique ainsi qu'au niveau des données et des applications.

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

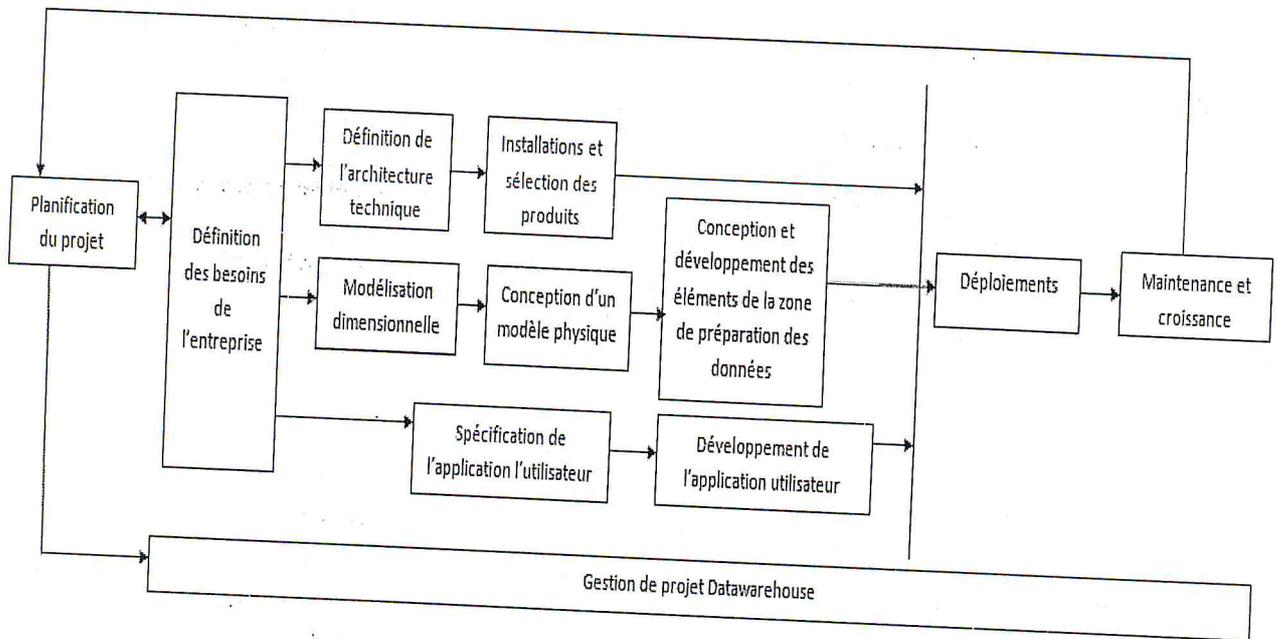


Fig 4.1:Schéma du cycle de vie dimensionnel [9]

3.1. Description du Schéma du cycle de vie dimensionnel de Ralph Kimball [Site 6]

1. **Planification du projet** : La planification aborde la définition et l'étendue du projet de l'entrepôt, y compris l'appréciation du niveau de maturité de l'organisation face à ce type d'approche. Elle se concentre sur les besoins en terme de ressources et de niveau de qualification, couplés aux affectations des tâches, à leurs durées et à leur séquençement.. La planification dépend bien évidemment des besoins, comme l'indique la flèche en double sens du schéma
2. **Définition des besoins** : Il est essentiel de bien comprendre les utilisateurs et leurs besoins, sinon l'entrepôt deviendra rapidement un exercice vain de la part de l'équipe des concepteurs.. L'approche utilisée pour identifier les besoins analytiques diffère de manière significative de la traditionnelle analyse des besoins basée sur les données. Les besoins une fois définis constituent le point de départ de trois trajectoires parallèles que sont la technologie, les données et les interfaces utilisateurs.
3. **Modélisation dimensionnelle** : C'est la définition des besoins qui détermine quelles sont les données requises pour répondre aux besoins d'analyse des utilisateurs. La conception du modèle logique de données commence par la construction d'une matrice représentant les processus métier clé et leurs dimensionnalités. A partir de cette matrice, il faut effectuer une analyse plus détaillée des données du (des) système(s) source(s) opérationnels. Le résultat de cette analyse est le modèle dimensionnel. Ce modèle identifie la granularité de la table de fait, les dimensions associées avec leurs attributs et leurs hiérarchisations. Cet ensemble d'activités s'achèvera sur le développement d'une mise en correspondance des données sources et cibles dans des méta-données.

4. **Conception du modèle physique** : La conception physique d'une base de données définit les structures nécessaires pour l'implémentation du modèle dimensionnel. Les éléments fondamentaux sont la détermination des règles de nommage des objets, la mise en place de l'environnement de la base de données. L'indexation primaire, les stratégies de partitionnement et les agrégations primaires sont également définies. La conception du modèle physique est fortement dépendante de la machine utilisée pour l'entrepôt.
5. **Conception et développement de la zone de préparation des données** : La conception de la zone de préparation des données (staging area) constitue généralement la tâche la plus sous-estimée du projet entrepôt de données. Le processus de préparation se déroule en trois phases majeures : Extraction, Transformation et le Chargement (Loading).
6. **Définition de l'architecture technique** : Cette étape définit la vision globale de l'architecture technique à mettre en oeuvre. Elle nécessite la prise en compte de trois facteurs : Les besoins; L'environnement existant et les orientations techniques stratégiques planifiées. En plus de l'architecture supportant l'entrepôt, il est nécessaire de mener des réflexions sur les outils de conception de la zone de préparation des données et des outils de restitutions.
7. **Choix technologiques et mis en oeuvre** : A partir de l'étude de l'architecture technique il faut sélectionner les composants spécifiques, telle plate-forme(s) matérielle(s) et logicielle(s), SGBD outils d'extraction et restitution à mettre en oeuvre. Une fois les produits évalués et sélectionnés, ceux-ci doivent être installés et testés méticuleusement afin de garantir une intégration adéquate d'un bout à l'autre de l'environnement de l'entrepôt.
8. **Développement de l'application utilisateur** : Il est recommandé de définir une série d'applications standard destinées aux utilisateurs finaux, car tous n'ont pas besoin d'un accès ad hoc à l'entrepôt. Les spécifications de l'application décrivent les maquettes d'états, les critères de sélection laissés à l'utilisateur et les calculs nécessaires.
9. **Déploiement** : Le déploiement est le point de convergence de la technologie, des données et des applications utilisateurs. Une planification est indispensable pour gérer le déploiement qui comprend également la formation des utilisateurs, les processus de communication, le support utilisateur, la prise en compte des demandes d'évolution et de correction.
10. **Maintenance et croissance** : Après le déploiement initial de l'entrepôt, c'est sa vie qui commence. Il faut s'assurer de fournir un service de support et de formation continue. Il faut également s'assurer que les processus mis en place pour la gestion de la zone de construction vont faire fonctionner l'entrepôt en continu et efficacement. Il est également important de mesurer périodiquement les performances de l'entrepôt et de son acceptation dans l'entreprise. L'entrepôt va donc évoluer et croître et le changement doit être perçu comme un facteur de succès et non d'échec. Des processus de hiérarchisation des priorités doivent bien sur être mis en place afin de gérer les demandes des utilisateurs en termes d'évolution et de croissance.

11. **Gestion du projet** : La gestion du projet garantit que les activités du cycle de vie restent sur la bonne voie et sont bien synchronisées. Cela consiste à contrôler l'état d'avancement du projet, la détection et la résolution des problèmes et le contrôle des changements afin de garantir l'accès aux objectifs du projet.

4. Spécification de notre approche :

La projection de notre démarche par rapport à la démarche proposée par Ralph Kimball est la décomposition de certaines phases en plusieurs et en parallèle l'élimination des autres. De plus, à cause de l'existence d'une infrastructure technique (matériel et logiciel). Nous avons abandonné les détails théoriques des composants et des services de l'architecture pour aborder les questions pratiques sur la sélection des produits.

Pour les deux étapes concernant la spécification et le développement des applications utilisateur, nous avons construit une seule étape qui regroupe ces deux dernières, il s'agit de l'étape construction du portail de restitution, comme nous avons supprimé les deux étapes : définition de l'architecture technique et l'installation et sélection des produit. L'étape maintenance et croissance se déroulera après le déploiement de l'entrepôt. Dans notre travail, cette étape est inutile car elle traite le point de continuation d'assistance et de formation des utilisateurs ainsi l'optimisation des performances de la base de données. Le schéma suivant illustre notre approche :

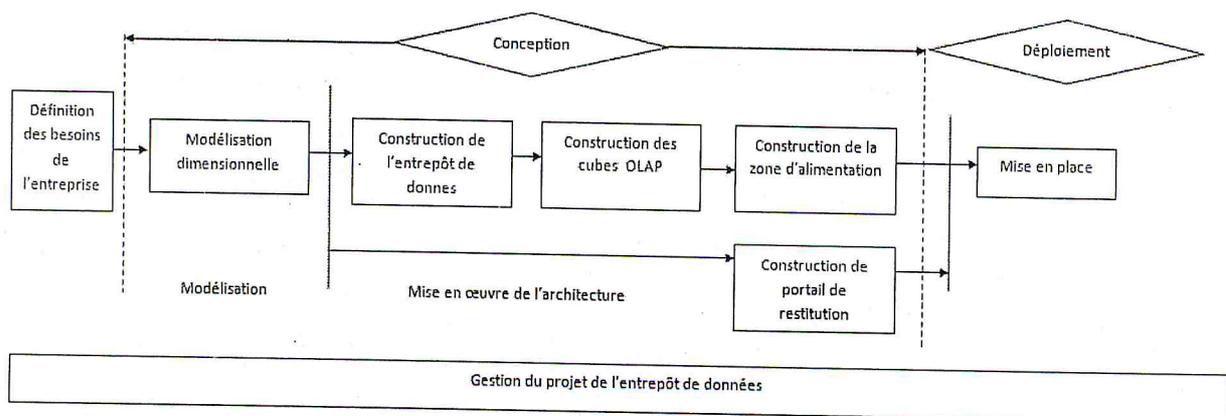


Fig4.2 : Succession des tâches de notre approche

5. Définition des besoins de l'entreprise

Comme nous l'avons évoqué dans l'introduction, la conception du modèle dimensionnel passe d'abord par une étape préliminaire qui consiste en l'étude et l'identification des besoins des utilisateurs (décideurs, analystes,...).

C'est la première phase de ce projet, elle consiste à prendre compte de la stratégie de l'entreprise, des besoins métier pour la fonction des projets.

L'étude des besoins doit déterminer le contenu de l'entrepôt et son organisation, d'après les résultats attendus par les utilisateurs, les requêtes qu'ils formuleront. Le besoin d'informatisation peut provenir du système de pilotage ou d'un service particulier de l'entreprise.

L'expression des besoins par les utilisateurs met souvent en évidence la volonté d'obtenir : des analyses sur ce qui s'est passé (par exemple comparer les performances actuelles d'un projet donné avec celles d'un autre projet de même nature réalisé l'année dernière) ou des analyses prédictives (par exemple déterminer les achats potentiels pour un type de client, déterminer les clients qui risquent d'abandonner l'entreprise, ...).

Les interviews doivent permettre de préciser les faits à suivre et l'ensemble des axes sur lesquels l'utilisateur voudrait fonder son analyse et indicateurs qui nous aiderons lors de la conception du modèle dimensionnel. Il faut recenser les données nécessaires à un bon fonctionnement de l'entrepôt, car La valeur ajoutée 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.

Il faut alors recenser les données disponibles dans les bases de production, toutes les données de production ne sont pas utiles dans l'entrepôt. C'est pour cela nous avons procédé à une série d'entretiens avec les responsables concernés de la Direction des Projet :

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.

Nous allons cités quelques besoins tels indiqués par l'utilisateur :

- Je souhaite classer mes projets selon leur nature, leur implantation et leur domaine pour savoir qu'elle est la nature des projets qui m'est rentable.

Quelle est la région qui travaille le plus à une certaine date que je détermine ?

Dans quel cas les valeurs réelles des projets coïncident avec celles qui ont été prévues ? Ainsi ces différents entretiens ont porté sur :

1. la connaissance de l'activité de l'entreprise, afin d'identifier les différents processus métiers qui la compose ;
2. l'identification des indicateurs de l'analyse de l'activité ;

Le but de cette identification est :

3. cerner les besoins et les attentes des utilisateurs ;
4. déterminer les dimensions et les faits de mesures afin d'établir la modélisation dimensionnelle ;

Ainsi à partir de tous les besoins, nous allons déterminer les différents indicateurs et axes d'analyse, mais avant cela, nous devons connaître les processus de cette activité (Suivi des projets).

5.1 Les processus métiers

5.1.1 Définition d'un processus

« Un processus est un ensemble de phénomènes conçus comme une chaîne causale progressive » **Le petit Larousse**. Cette définition apporte deux notions essentielles du processus :

-Le processus est un ensemble de phénomènes.

-Cet ensemble est conçu comme une chaîne causale et progressive. En consolidant la définition précédente avec les notions de finalité et de mesure nous aboutissons à une définition spécifique au processus organisationnel :

« Un processus est un regroupement cohérent d'activités, disposant de ressources, alimenté par des entrées, qui sont transformé en sorties en y apportant une valeur ajoutée pour le client ». [17]

5.1.2 Identification des processus

« Les processus de réalisation regroupent toutes les activités dédiées au cycle de vie de produit et peuvent englober selon les cas les activités suivantes : recherche de marketing, développement de nouveaux produits, avant vente, conception, production ». [17]

Dans notre cas et après le résultat d'une série d'entretiens, nous avons aboutis aux processus clés pour les décideurs en ce qui concerne la fonction « Suivi des projets » sont réalisés sur différentes activités : **Préparation des projets, exécution des projets, contrôle des projets, capitalisation des projets et contrôle équipements**

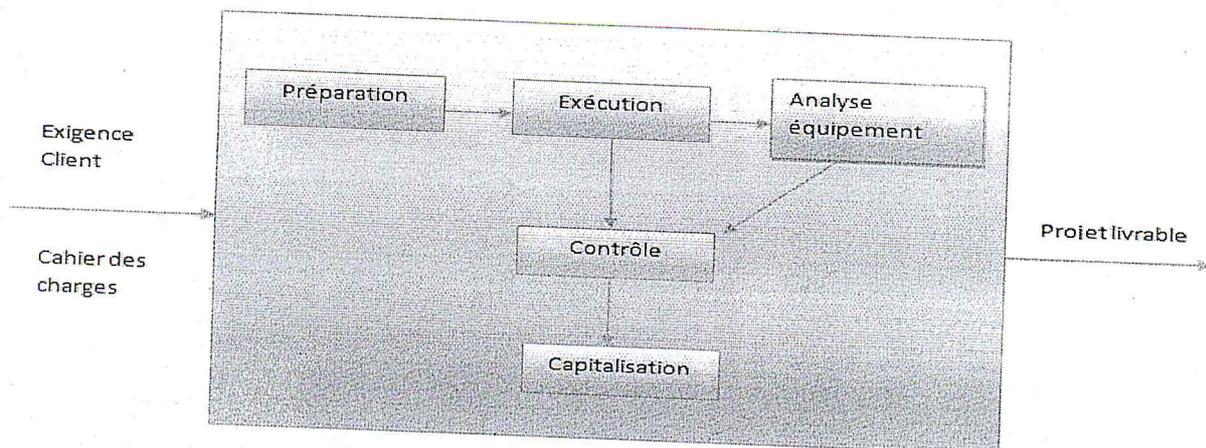


Fig 4.3 : Les processus de l'activité de la fonction « Suivi des Projets »

5.2 Les indicateurs d'analyse de l'activité

Lors des entretiens, on a compris clairement le souhait des utilisateurs d'employer les informations produites lors du déroulement des processus métiers précédentes afin de mesurer leur activité commerciale, car Le contrôle d'indicateurs clés soigneusement définis permet de fournir une vue instantanée de l'activité au sein d'un département, par exemple : le chiffre d'affaire.

Pour cela voici les principaux indicateurs que nous avons extrait :

- Le chiffre d'affaire prévu ainsi que le réel des projets par rapport aux projets, les clients, régions et le temps.
- Avancement prévu (AP) ainsi que l'Avancement réel (AR) des projets selon les axes précédents
- le Nombres d'Heures prévues (HP), le Nombre d'heures dépensées (HD) ainsi que le Nombre d'heures achevées (HA) des projets selon les axes précédents ;
- Taux de réalisation des projets (AR/AP) selon les axes précédents ;
- Taux de rendement des projets (HA/HD) selon les axes précédents ;
- Taux de mobilisation des projets (HD/HP) selon les axes précédents ;
- Nombre d'effectifs prévu et le réel pour les projets selon les axes précédents ;
- Total location matériel = taux location du matériel * nombre de jours ou il est utilisé. Les charges prévues, les charges réelles.

Les heures perdues, les heures de malfaçon.

Taux des heures perdues : $H\text{-Perdues}/HD$.

Taux de malfaçons : $H\text{-Malfaçons}/HD$.

6. Modélisation dimensionnelle

La modélisation dimensionnelle vient au secours des utilisateurs, car elle représente au mieux leurs besoins et leurs attentes qui repose sur l'analyse de la performance. Cette performance peut se matérialiser au travers d'un ensemble d'indicateurs.

La modélisation d'un data Warehouse (métier) part du principe d'un processus de conception de haut vers le bas, en effet c'est en connaissant le processus des activités qu'on pourra déterminer les dimensions et les faits qui figureront dans notre base. La modélisation dimensionnelle 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.

Le processus de la conception logique d'un data Warehouse passe par quatre première étapes [17]

Etape 1 : Sélection du processus d'entreprise à modéliser

La première étape de la conception consiste à décider quels processus d'entreprise doivent être modélisés, en confrontant notre compréhension des exigences de l'activité à la réalité des données disponibles. Le modèle dimensionnel construit doit répondre aux questions les plus pressantes des utilisateurs et ses données doivent être parmi les plus facile à extraire.

Etape 2 : Déclaration du grain

Une fois le processus identifié, l'équipe de l'entrepôt de données doit prendre une décision importante concernant la granularité. Quel niveau de détail doit-il être rendu disponible dans le modèle dimensionnel ? Le modèle dimensionnel doit être basé sur les données les plus atomiques, les plus détaillées susceptibles d'être collectées.

Etape 3 : Choix des dimensions

Une fois le grain de la table de fait choisi, nous pouvons maintenant déterminer les dimensions principales de la table de faits.

Etape 4 : Identification des Faits

La quatrième et dernière étape de la conception est la détermination des faits qui appaireront dans la table de faits. Ici, notre choix est en fonction de la déclaration du grain.

6.1. Modélisation dimensionnelle de l'activité « préparation »

6.1.1. Le processus d'activité

Après la signature du contrat avec le client et avoir accepter toutes les contraintes imposées par celui-ci, ainsi que la bonne étude du cahier de charge. Vient alors l'étape de préparation du projet afin de pouvoir l'entamé, et qui consiste à prévoir tous les besoins nécessaire pour la réalisation d'un projet comme ressources et moyens pour chaque début du mois.

Il s'agit de prévoir l'effectif qualifié dont a besoin le projet, le matériel adéquat, comme on prévoit aussi les charges.

Nous avons jugés l'importance de la présence des valeurs prévues dans l'entrepôt et leur évolution dans le temps parce que l'utilisateur pourra les consulter pour les utiliser dans un autre projet de même nature que le premier. Grâce à ces valeurs, on arrivera à les comparer avec les valeurs réelles pour tirer des conclusions sur l'avancement du projet.

. Cette phase de préparation est valable pour tout nouveau projet, que l'ENGTP entreprend.

6.1.2. Le grain du processus d'activité

En général, le grain du processus représente le niveau de détail le plus fin et atomique. Déclarer le grain du processus, revient à spécifier exactement ce que représente une ligne individuelle de la table de fait. En effet dans cette phase de préparation du projet ce fait pour chaque début du mois ; car il est clair que l'information s'obtiendra par mois et pas autrement, donc le grain est mensuel.

6.1.3. Les dimensions

D'après la granularité de la table de faits, nous avons déterminé une série de dimensions principales. Cette série comporte les dimensions suivantes : temps, projet, région, client.

Les tables de dimensions contiennent les descriptions textuelles de tous les attributs des dimensions, ils sont utilisés pour décrire des aspects d'une dimension qui peuvent être utilisé comme source de contraintes par les utilisateurs.

a. Dimension « temps »

La définition de la dimension temps selon Ralph KIMBALL « 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 (subordonné) de la base de données » [17]. Dans notre structure pour l'activité préparation du projet, nous avons opté pour un grain mensuel de données. Ainsi chacun des enregistrements de notre table de dimension temps représente un mois. Le champ année sert à stocker l'historique annuel des réalisations pour qu'on puisse comparer l'activité de cette année et l'année précédente durant le même mois.

b. Dimension « projet »

La dimension projet décrit chaque nouveau projet dans l'entreprise. Ils sont identifiés par une clé unique, Les projets sont classés par secteur d'activité (hydrocarbures, énergie, chimie et pétrochimie,...).

c. Dimension « client »

Chaque enregistrement représente un client unique de l'entreprise. Elle représente le client qui possède un contrat avec l'entreprise pour un projet donné.

d. Dimension « région »

La dimension région décrit chaque région qui représente l'endroit où le projet va être réalisé. Elle est identifiée par une clé unique, elle sert à distinguer entre les réalisations dans différentes régions durant la même période.

6.1.4. Les faits mesurés

Les faits sont les mesures de l'activité, chacune de ces mesures doit être prise à l'intersection de toutes les dimensions. Dans le cadre d'une gestion des projets ; il est important de connaître les mesures prévues, c'est pour cela notre table de fait doit contenir les mesures suivantes :

- Le chiffre d'affaire prévu.
- Les heures prévues.
- les charges prévues,
- L' avancement prévu., le nombre d'effectifs prévus.

Ces faits sont parfaitement additifs sur toutes les dimensions. Après avoir défini les faits et les attributs des dimensions, à présent voici notre schéma conceptuel de l'activité préparation du projet qui se présente comme suit

NB : pour un souci de lisibilité, nous avons représenté dans les schemas, que les clés dans des faits et dimension. Le reste des attributs sont décrits plus loin

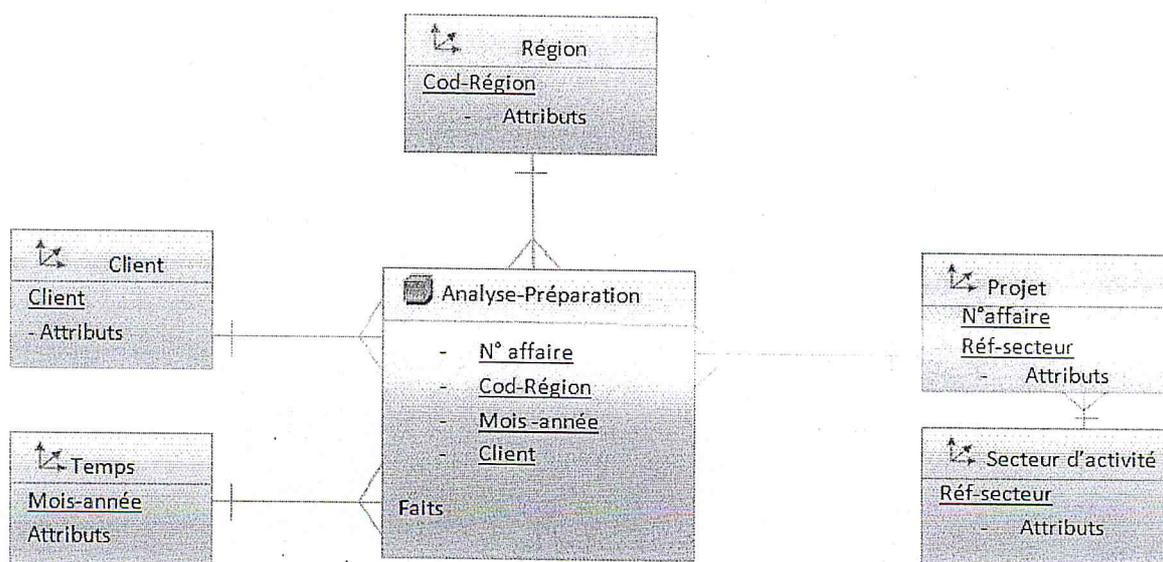


Fig4.4 :Modèle dimensionnel en flocon de l'activité « Analyse_Préparation »

6.2 La modélisation dimensionnelle de l'activité « exécution »

6.2.1. Le processus de l'activité

Ce processus est aussi jugé important dans notre cas vu qu'il nous retourne les valeurs exactes concernant l'avancement du projet ainsi que les ressources nécessaires (humaines et matérielles). L'utilisateur pourra consulter notre système pour voir la différence entre les réalisations de cette année avec celles des années précédentes concernant des projets similaires

par rapport à différents paramètres (nature du projet, le chantier où se déroule le projet, la région, etc.)

6.2.2. Le grain du processus de l'activité

le niveau de détail des informations obtenu pour la région d'exécution du projet est à la fin de chaque mois qu'on pourra établir une situation sur l'état d'avancement du projet, c'est un bilan qui se fait à la fin de chaque mois. Dans ce processus, il est important de connaître mensuellement les réalisations

6.2.3. Les dimensions conformes

Les perspectives avec lesquelles les utilisateurs de cette activité font leurs analyse sont en réalité les entités manipulées dans cette phase, sachant que c'est les même que la précédente (5.1 l'activité « Préparation ») : projet, temps, client, région.

6.2.4. Les faits mesurés

C'est des données numériques toujours, qui changent périodiquement, ce sont des indicateurs permettant aux utilisateurs d'analyser ce processus d'activité d'exécution de projet , il s'agit de :

- chiffre d'affaire réel et l'avancement réel.
- les heures dépensées et achevées.
- le nombre d'effectifs réel.
- les charges réelles.

Notre schéma conceptuel de l'activité « exécution » se présente comme suit :

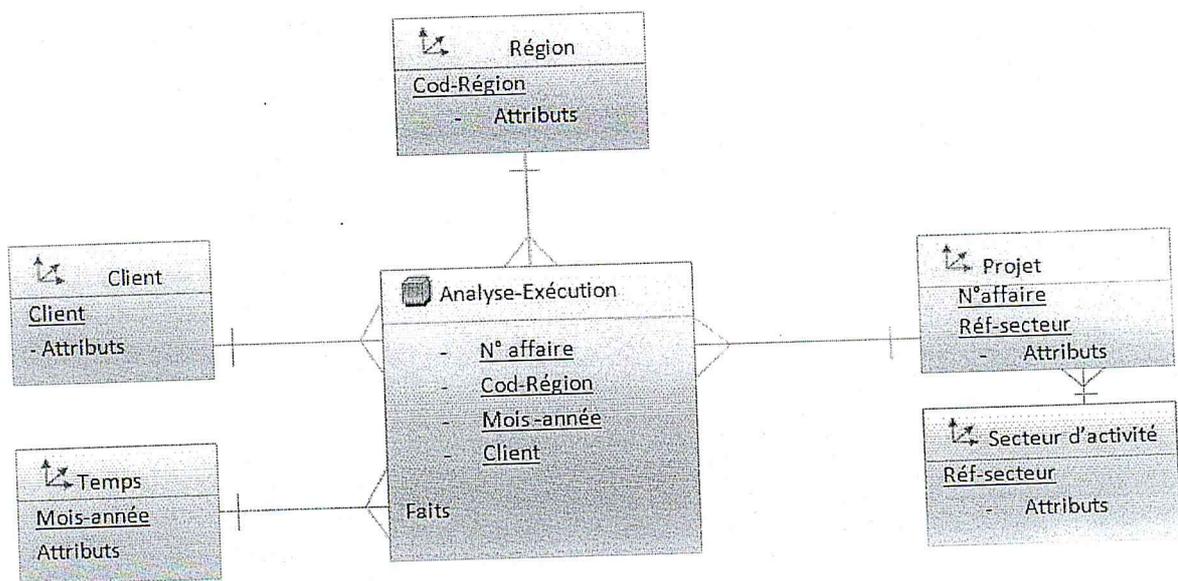


Fig 4.5 : modèle dimensionnel en flocon de l'activité « analyse_exécution »

6.3 La modélisation dimensionnelle de l'activité « Contrôle »

6.3.1 Processus de l'activité

Il s'agit de contrôler le bon avancement du projet, en comparant les données prévisionnelles à celles réalisées et voir si les données qui ont été prévues correspondent bien à celles réalisées ou bien le contraire il y a une différence, soit le prévu dépasse le réel et dans ce cas tout est bon, c'est à dire qu'on a pas besoin de rajouter des quantités de ressources, sinon dans le cas contraire, on est appelé à rajouter le manque qu'a besoin le projet. Cette activité se déroule tous les fins de mois depuis le commencement du projet jusqu'à sa fin.

6.3.2. Le grain de l'activité

Sachant que les données réelles concernant l'état d'avancement parviennent tout les chaque mois, ainsi que les prévisions pour la préparation du projet ce fait mensuellement, donc il est normale que le contrôle ce fait à la fin tout les chaque mois.

6.3.3. Les dimensions conformes

Les perspectives avec lesquelles les utilisateurs de cette activité font leurs analyse sont en réalité les entités manipulées dans cette phase, sachant que c'est les même que l'activité précédente : projet, temps, client, région.

6.3.4. Les faits mesurés

Il s'agit d'indicateurs numériques et additifs toujours, à fin de permettre aux utilisateurs de mieux analyser l'activité et qui sont les suivants :

- le taux de Réalisation (**AR/AP**).
- le taux de Rendement (**HA/HD**).
- le taux de Mobilisation (**HD/HP**).

Notre schéma conceptuel de l'activité contrôle du projet se présente comme suit :

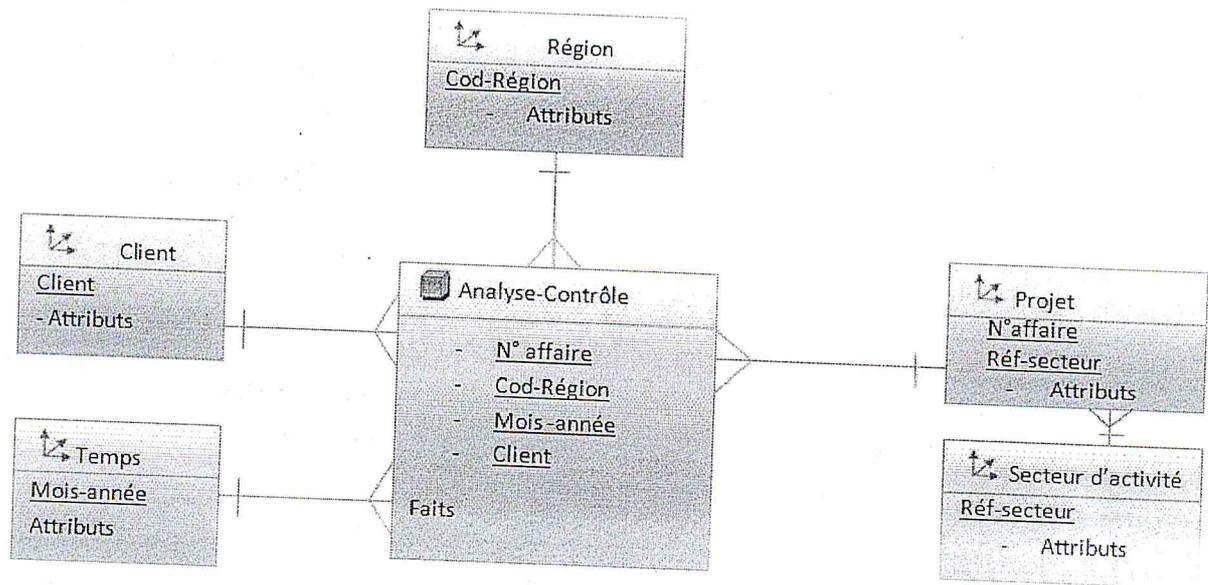


Fig 4.6 : modèle dimensionnel en flocon de l'activité « analyse_contrôle »

6.4 La modélisation dimensionnelle de l'activité « capitalisation »

6.4.1 Le processus d'activité

Il s'agit d'élaborer un bilan complet qui synthétise le cycle de vie du projet pour chaque mois.

On mentionnera toutes les remarques qui correspondent à toutes les étapes du déroulement du projet ainsi que les anomalies qui peuvent y avoir, à fin de permettre aux utilisateurs d'analyser et de prendre la bonne décision. Durant chaque projet, les décideurs auront besoin d'avoir une synthèse sur son déroulement.

6.4.2. Le grain du processus

Cette capitalisation du projet se fait à la fin de chaque mois, car c'est un bilan mensuel qui vient après une durée de réalisation. A fin que les décideurs soient au courant de toutes les étapes du projet ainsi que toutes les anomalies qui se sont produites durant la réalisation du projet.

6.4.3. Les dimensions de l'activité

6.4.3.1. Les dimensions conformes

Il s'agit les mêmes dimensions manipulées dans les précédentes activités. Sa ne change en rien puisque les entités concerne toujours : projet, temps, client, région.

6.4.3.2. Les dimensions supplémentaires

Pour ce processus d'activité de capitalisation du projet, il nous faudra une entité qui décrira bien le projet pour chaque mois en spécifiant toutes les remarques, il s'agit de la dimension :

Document textuel.

6.4.4. Les faits mesurés

C'est des faits qui serviront pour l'analyse du processus de l'activité. Qui correspond à :

Les heures perdues (**HPerdues**).

Les heures de malfaçons (**HM**).

Le taux des heures perdues (**HPerdues/HD**).

Le taux de malfaçon (**HM/HD**)

Voici notre schéma conceptuel qui représente l'activité capitalisation du projet :

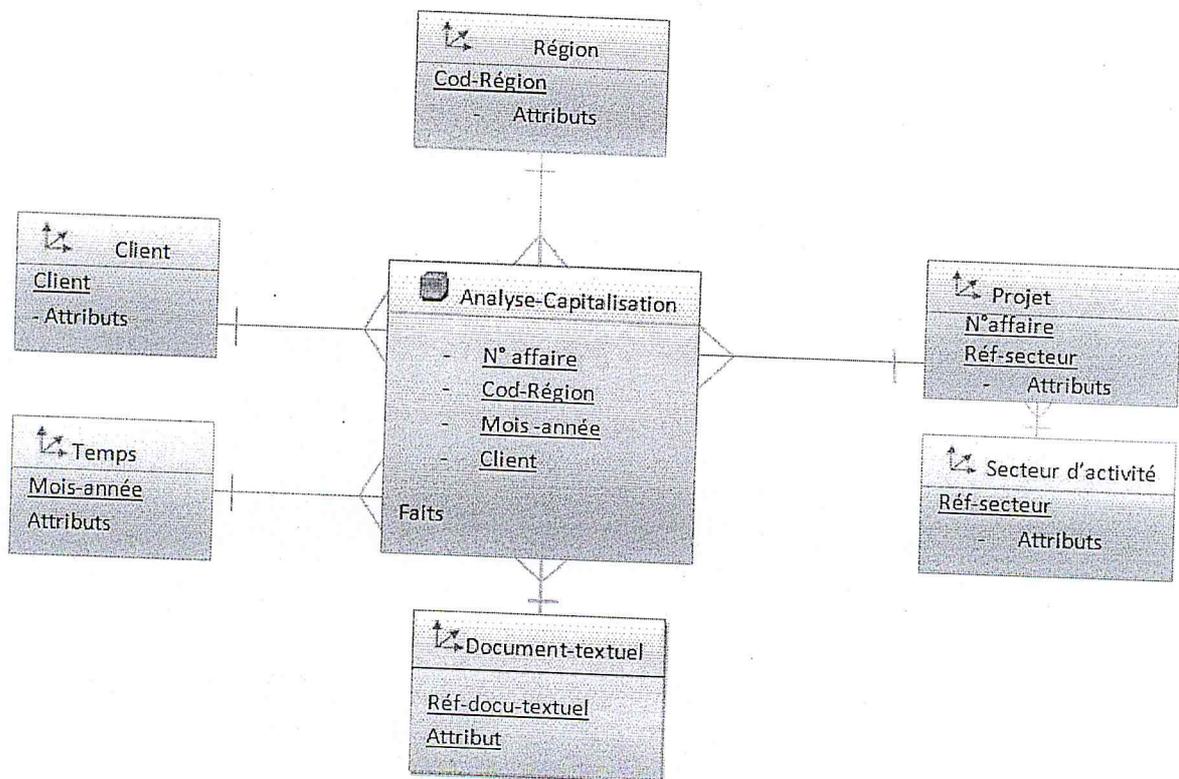


Fig 4.7 : modèle dimensionnel en flocon de l'activité « analyse_capitalisation »

6.5. La modélisation dimensionnelle de l'activité « équipement »

6.5.1. Le processus de l'activité

Cette activité s'exécute comme une sorte de contrôle du mouvement du matériel. Que nous avons jugé importante pour l'analyse de la quantité totale du matériel utilisé.

6.5.2. Le grain du processus

Le niveau de détail concernant ce processus d'activité correspond à une information qui s'établira pour chaque mois, ainsi on connaîtra le nombre de matériels utilisés.

6.5.3. Les dimensions de l'activité

6.5.3.1. Les dimensions conformes

C'est des dimensions de perspective d'analyse pour les utilisateurs identiques à celles de la table de fait précédente et qui représente les dimensions suivantes : projet, temps, client, région.

6.5.3.2. Les dimensions supplémentaires

C'est une dimension qui décrit le matériel, il s'agit de la dimension : matériel.

6.5.4. Les faits mesurés

C'est des faits numériques et additifs, a fin de permettre aux utilisateurs une meilleure analyse de l'activité, une analyse rapide, pertinente et facile. Il s'agit des mesures suivantes :

-le nombre de jours. -le Total.

Voila le schéma conceptuel qui représente l'activité équipement :

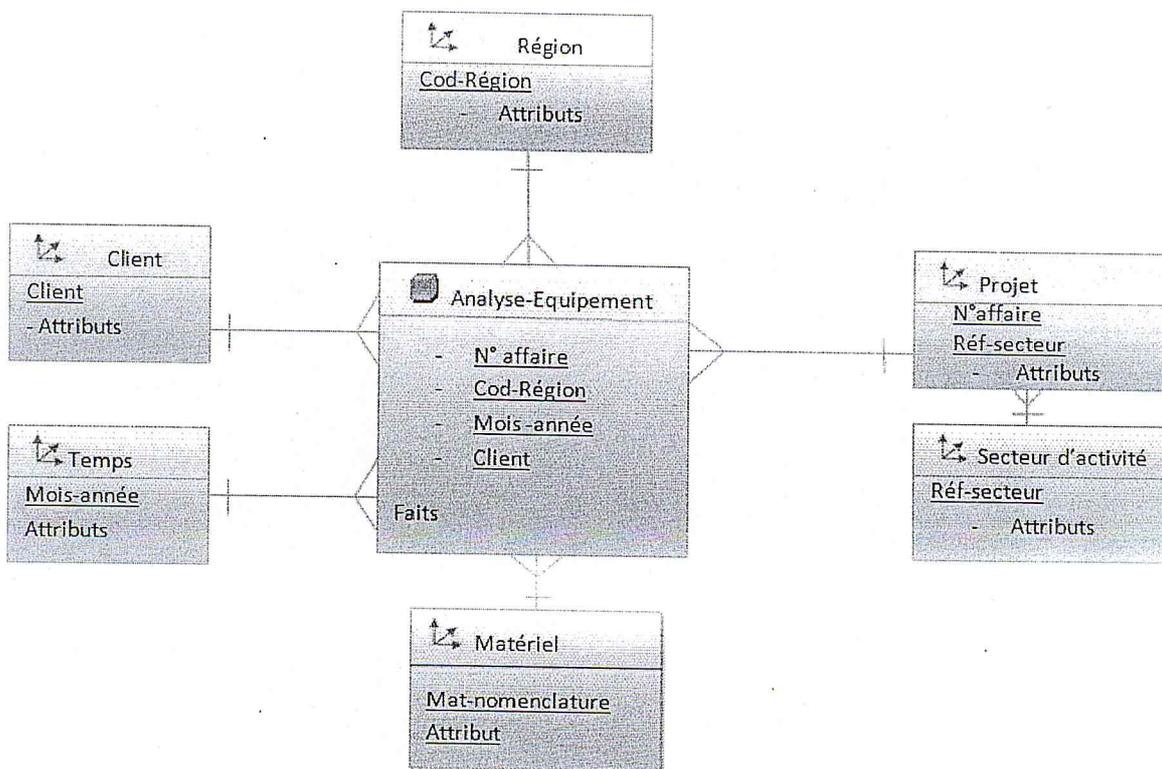


Fig 4.8 : modèle dimensionnel en flocon de l'activité « analyse_équipement »

7. Mise en œuvre de l'Architecture

Dans cette partie, nous allons présenter le processus de mise en œuvre de notre système décisionnel, ce processus est constitué de plusieurs composantes visibles dans l'architecture des systèmes décisionnels basés sur le concept « Entrepôt de données ».

Notre méthodologie de mise en œuvre du système décisionnel est basée sur trois éléments essentiels : une base de données et de métadonnées constituant l'interface entre l'équipe informatique et la communauté des utilisateurs, un ensemble de transformation simples et expressives appliquées aux données sources et 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 dataWarehouse est représentée par le schéma suivant :

7.1. Architecture détaillé du système

La figure suivante illustre bien les différents composants qui constituent notre système, c'est le processus de la méthodologie suivie, à fin de mieux mettre en œuvre notre système.

Chaque objet de l'architecture, fait objet d'une étape du processus de mise en œuvre du système.

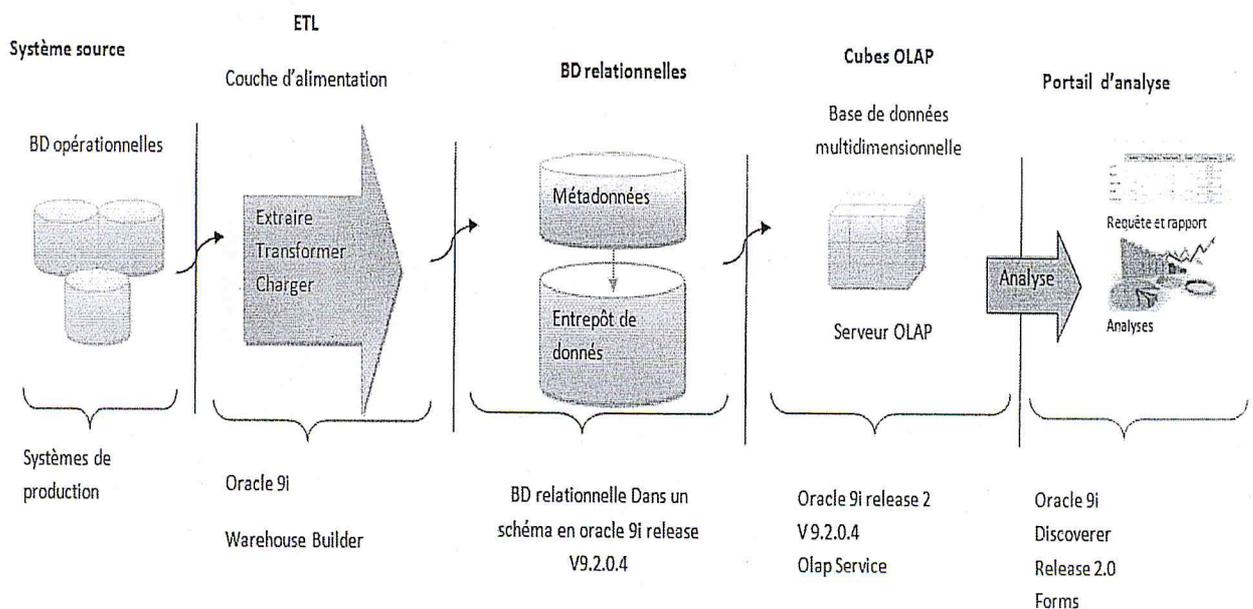


Fig 4.9 : Architecture détaillée du système

1. ETL

Les données stockées dans l'entrepôt de données sont différentes de celles du système source en terme 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. Notre entrepôt sera alimenté à travers quatre applications différentes qui sont:

- **L'application personnel (pers)**

C'est une application qui gère le suivi des employés depuis le recrutement jusqu'à sa sortie de l'entreprise. Elle regroupe aussi la paie, et puisque les projets nécessitent des ressources humaines alors cette application 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 comptabilité (compta)**

C'est une application qui gère les ressources financières de l'entreprise, pour établir à la fin de chaque exercice des rapports d'analyses comme les bilans et le tableau des résultats. A partir de ce système, nous pouvons extraire le chiffre d'affaire, les charges et les résultats pour chaque projet.

- **L'application matériel (gmat)**

Les projets nécessitent les ressources matérielles, donc cette application fournit tous les équipements (chaque équipement possède son taux de location) utilisés avec le nombre de jours ou ce matériel a été utilisé. La valeur qui va être chargée dans notre entrepôt sera le produit entre le taux de location et le nombre de jours, le résultat représentera le total location matériel qui va être utile dans l'analyse du matériel et la situation du stock.

- **L'Application suivi des Projets**

A partir de cette application, nous pouvons extraire tous les renseignements concernant les projets, tel que la région de sa réalisation, le délai de réalisation...etc., qui sont nécessaires pour le chargement de l'entrepôt.

Donc notre travail va consister à collecter toutes ces différentes données (indicateurs) à partir des systèmes sources, les modéliser dans un entrepôt de données qui sera exploité par l'utilisateur.

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

3. Base OLAP :

C'est la base de données multidimensionnelle, qui stocke les données dans des cubes OLAP qui seront interrogés par l'utilisateur final.

Notre serveur OLAP utilisé est Oracle 9i Server OLAP.

4. Le 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 la 2^{ème} 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.

7.2. Construction de l'entrepôt de données

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.

7.2.1. La liste des tables :

Le tableau suivant illustre les différentes tables qui constituent notre entrepôt de données. D'après la modélisation logique, les entités sont transformées en des tables relationnelles. En voici la liste des tables de l'entrepôt de données

Nom de la table	Identifiant.
Analyse-préparation	N_affaire , Mois année, Client, cod-Région
Analyse_ exécution	N affaire , Mois aimée, Client, cod-Région
Analyse_ contrôle	N_affaire , Mois_année, Client, cod-Région
Analyse_ capitalisation	N_affaire, Mois_année, Client, cod-Région Réf Document_Textuel
Analyse_ équipement	N_affaire, Mois_année, Client, cod-Région, Mat nomenc -
Projet	N_affaire
Temps	Mois_année
Client	Client
Région	Cod-Région
Secteur-activité	Réf secteur
Document textuel	Réf-document-textuel

Tab 4.1 : liste des tables de l'entrepôt de données

Désignation	Code	Table	Type de données
Numéro d'affaire du projet	N_affaire	Projet	Varchar2(6)
Date début du projet	Date_debut	Projet	Date
Date fin du projet	Date_fin	Projet	Date
Libellé du projet	Libellé_affaire	Projet	Varchar2(100)
Nature du projet	Nature	Projet	Varchar2(100)
Délai de réalisation du projet	Délai_réalisation	Projet	Varchar2(100)
Montant du projet	Montant	Projet	Varchar2(100)
Volume horaire du projet	Manhours	Projet	Varchar2(100)
Implantation du projet	Implantation	Projet	Varchar2(100)
Référence de la région du projet	Cod-Région	Région	Varchar2(40)
Volume horaire	V_horaire	Projet	Number(5)
Référence secteur d'activité	Réf_secteur	Secteur d'activité	Number(2)
Secteur d'activité	Secteur_activité	Secteur d'activité	Varchar2(7)
Le mois et l'année du projet	Mois_année	Temps	Varchar2(7)
Le mois	Mois	Temps	Char(30)
L'année	Année	Temps	Char(30)
Nomenclature du matériel	Mat_nomenc	Matériel	Char(30)
Désignation du matériel	Mat_desig	Matériel	Char(30)
Taux de location	Taux_location	Matériel	Char(30)
Référence du document textuel	Réf_doc_textuel	Document textuel	Varchar2(40)
Description de la qualité	Description_qualite	Document textuel	Varchar2(40)
Etapas	Etapas	Document textuel	Varchar2(40)
Remarques	Remarques	Document textuel	Varchar2(40)
Identifiant du client	Client	Client	Varchar2(40)
Nom du client	Nom_client	Client	Varchar2(40)
L'adresse du client	Adresse	Client	Varchar2(40)
Nombre d'effectifs prévu	Nb_eff_prévu	Analyse préparation	Number(30)
Chiffre d'affaire prévu	CA_prevu	Analyse préparation	Number(30)
Les heures prévues	Heures_prevues	Analyse préparation	Number(30)
Les charges prévues	Charge_prevues	Analyse préparation	Number(30)
L'avancement prévu	Avancement_prevu	Analyse préparation	Number(30)
Nombre d'effectif réel	Nb_eff_réel	Analyse exécution	Number(30)
Les heures achevées	Heures_achevé	Analyse exécution	Number(30)
Les heures dépensées	Heures_depensée	Analyse exécution	Number(30)
Les charges réelles	Charges_réel	Analyse exécution	Number(30)
Chiffres d'affaire réel	CA_réel	Analyse exécution	Number(30)
L'avancement réel	Avancement_réel	Analyse exécution	Number(30)
Heures perdues	H_Perdues	Analyse capitalisation	Number(30)
Heures Malfaçons	H_malfaçon	Analyse capitalisation	Number(30)
Taux des heures perdues	Taux-h-perdues	Analyse capitalisation	Number(30)
Taux des heures de malfaçon	Taux-h-malfaçon	Analyse capitalisation	Number(30)
Taux de réalisation	Taux_réalisation	Analyse contrôle	Number(30)
Taux de rendement	Taux_rendement	Analyse contrôle	Number(30)
Taux de mobilisation	Taux_mobilisation	Analyse contrôle	Number(30)
Nombres de jour de location du matériel	NB_jour	Analyse équipement	Number(30)
Le total du taux de location	Total	Analyse équipement	Number(30)

Tab 4.2 : liste des attributs de l'entrepôt de données

7.3 Construction des cubes OLAP

La modélisation multidimensionnelle facilite énormément l'interrogation des données, car elle permet une analyse en ligne rapide et efficace. Le gestionnaire OLAP nous permet de transformer les données de la base de données relationnelles en des données faciles à visualiser, interrogées et en accédant rapidement grâce à la création des cubes.

Nous allons procéder dans cette étape à la construction de la base multidimensionnelle (cubes OLAP).

La technique de modélisation utilisée pour mettre en œuvre nos cubes OLAP est la modélisation MOLAP « Multidimensionnel OLAP », vu qu'elle facilite l'analyse multi-axes sur les données pertinentes et agrégées. Le serveur OLAP utilisé pour construire les cubes est l'Oracle 9i Release - OLAP service.

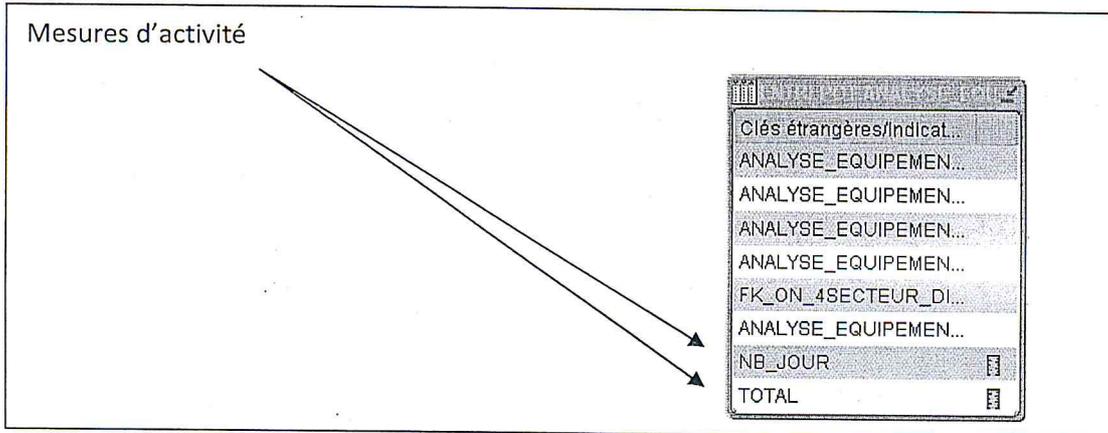
Mesures	Dimension Sollicitées	
Préparation	<ul style="list-style-type: none"> • CA-prévu • Avancement prévu • Heures-prévues • Charges prévues • NBR-EFF-Prévu 	<ul style="list-style-type: none"> ▪ Projet -> Secteur d'activité ▪ Région ▪ Temps ▪ Client
Exécution	<ul style="list-style-type: none"> • CA-réel • Avancement réel • Heures-Achevées • Charges Réelles • Heures dépensées • NBR-EFF-Prévu 	<ul style="list-style-type: none"> ▪ Projet -> Secteur d'activité ▪ Région ▪ Temps ▪ Client
Contrôle	<ul style="list-style-type: none"> • Taux- Mobilisation • Taux -rendement • Taux-réalisation 	<ul style="list-style-type: none"> ▪ Projet -> Secteur d'activité ▪ Région ▪ Temps ▪ Client
Capitalisation	<ul style="list-style-type: none"> • Heures perdues • Heures malfaçon • Taux-heures-Perdues • Taux-Heures-Malfaçon 	<ul style="list-style-type: none"> ▪ Projet -> Secteur d'activité ▪ Région ▪ Temps ▪ Client
Equipement	<ul style="list-style-type: none"> • Nb-Jour • TOTAL 	<ul style="list-style-type: none"> ▪ Projet -> Secteur d'activité ▪ Région ▪ Temps ▪ Client

Tab 4.3- Liste des faits choisis et les dimensions sollicitées

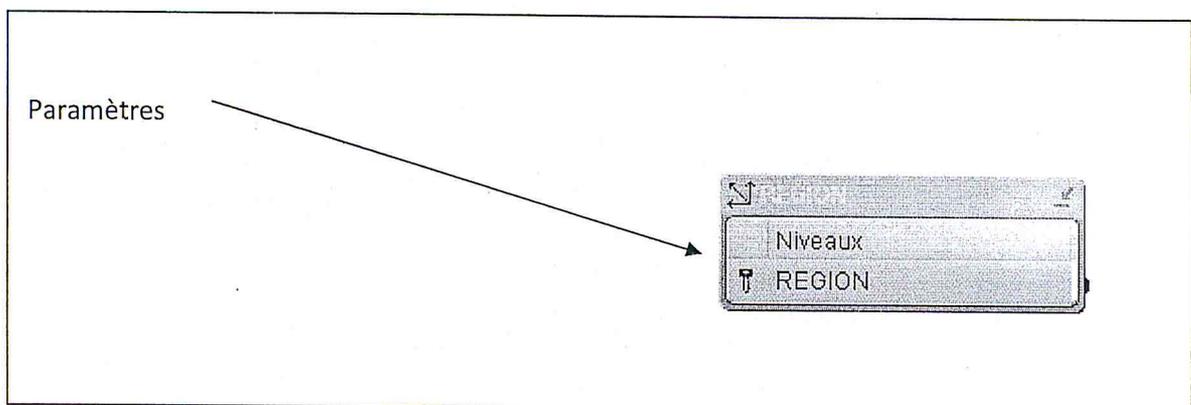
7.3.1 Formalismes utilisés :

➤ **Concept de Fait :**

Nous représentons le concept de fait par une table englobant les différentes mesures d'analyse de l'activité. Le fait est estampillé par une table.



Nous représentons la dimension par un rectangle contenant les différents paramètres d'analyse. Les trois flèches estampillent les dimensions.



7.3.2 Schémas de modèle MOLAP des cubes :

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

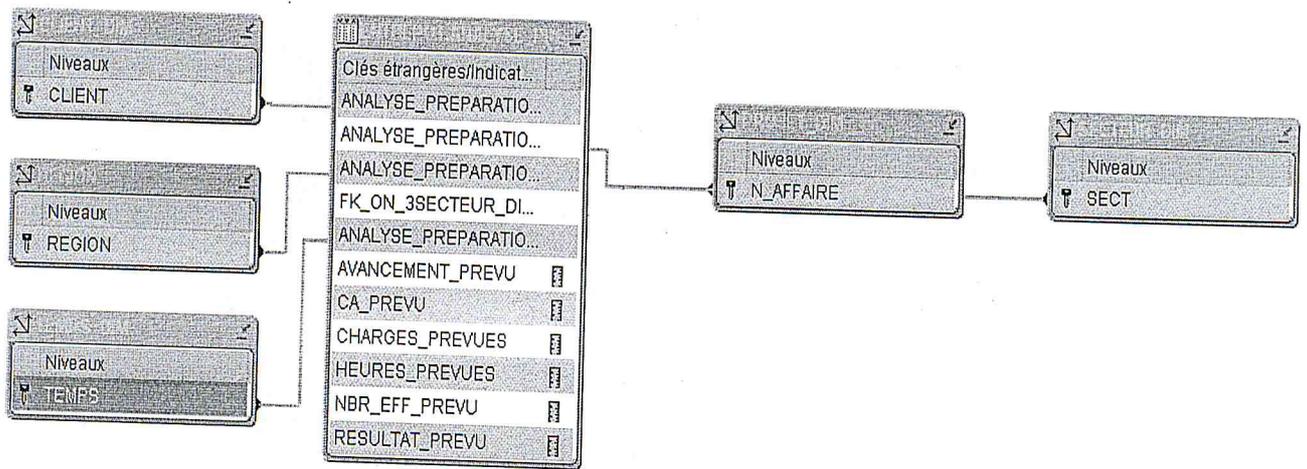


Fig 4.10 : Le Cube MOLAP de l'activité « préparation »

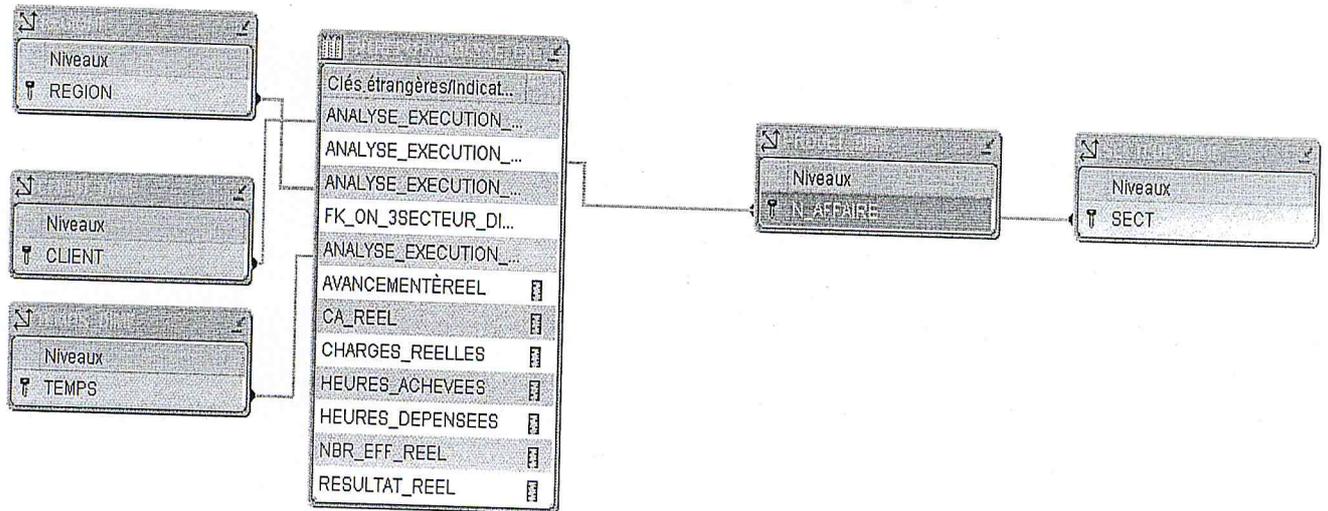


Fig 4.11 : Le Cube MOLAP de l'activité « exécution »

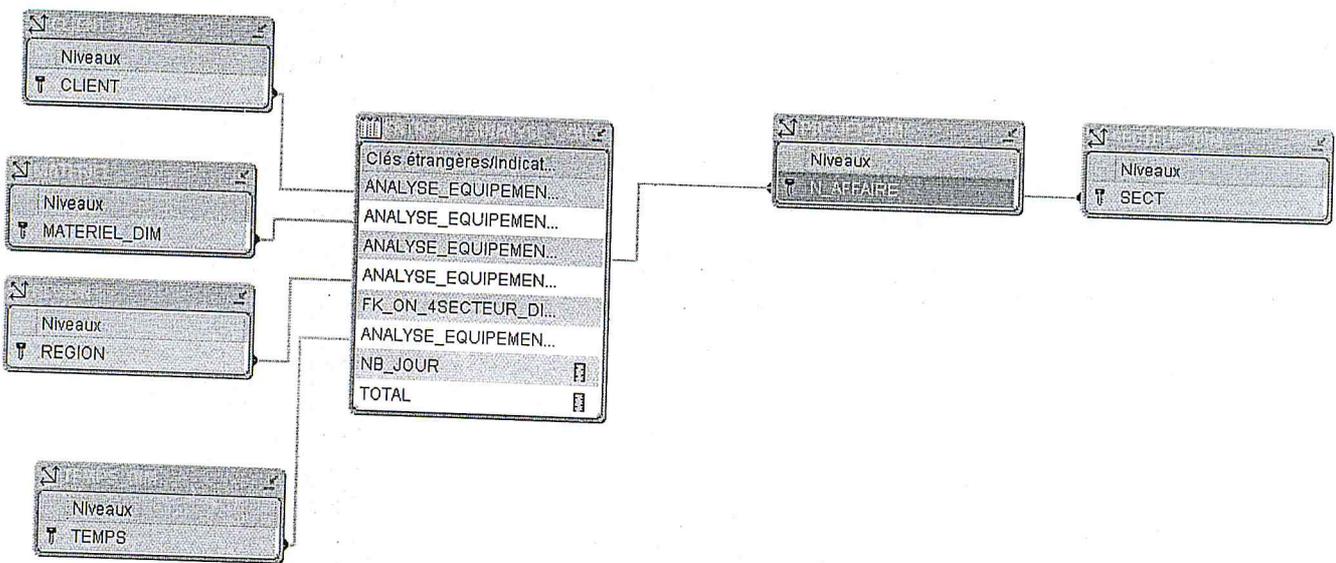


Fig 4.12 : Le Cube MOLAP de l'activité « équipement »

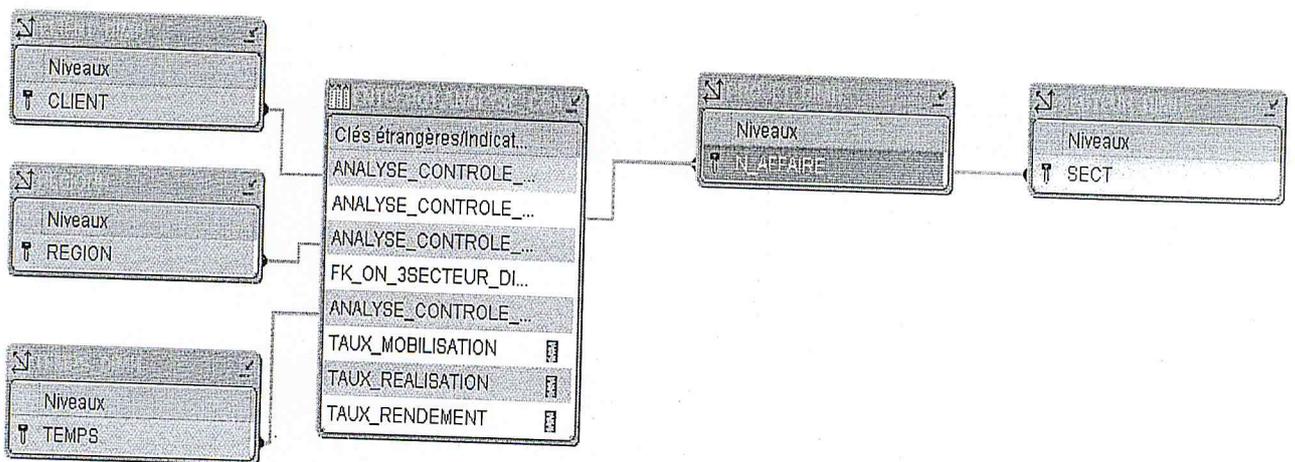


Fig 4.13 : Le Cube MOLAP de l'activité « contrôle »

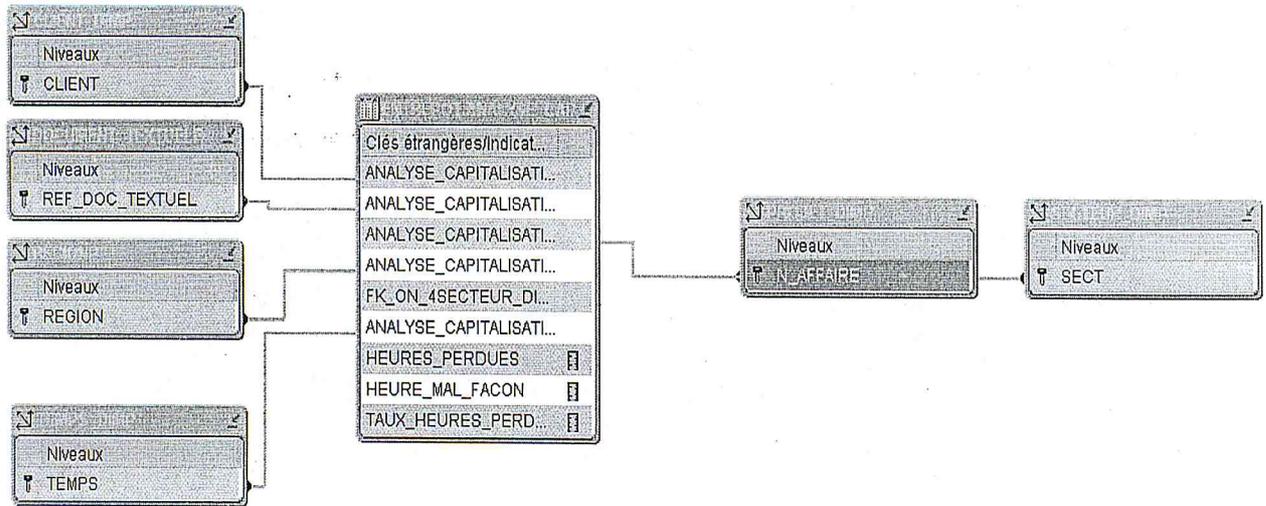


Fig 4.14 : Le Cube MOLAP de l'activité « capitalisation »

7.4. Construction de la zone d'alimentation

La construction de la zone d'alimentation de l'entrepôt, permettrait le transfert des données de la base opérationnelle vers l'entrepôt de données, qui se déroule en trois étapes élémentaires, il s'agit d'extraction des données du système source que doit contenir la base de l'entrepôt, ensuite regrouper l'ensemble de ces données, pour à la fin les chargées dans l'entrepôt de données. Pour réaliser ce processus d'alimentation de l'entrepôt, il est nécessaire de construire une base qui stock tous les informations sur les données, il s'agit de la méta- base, qui décrit les données que doit contenir l'entrepôt, en donnant où et comment extraire les données nécessaire du système source . Ainsi, un changement de type de données simple ne nécessite en fait qu'un léger changement dans l'ETL (Ces changement sont :des calculs d'agrégation, des changements de format ou de présentation,desjointures...).

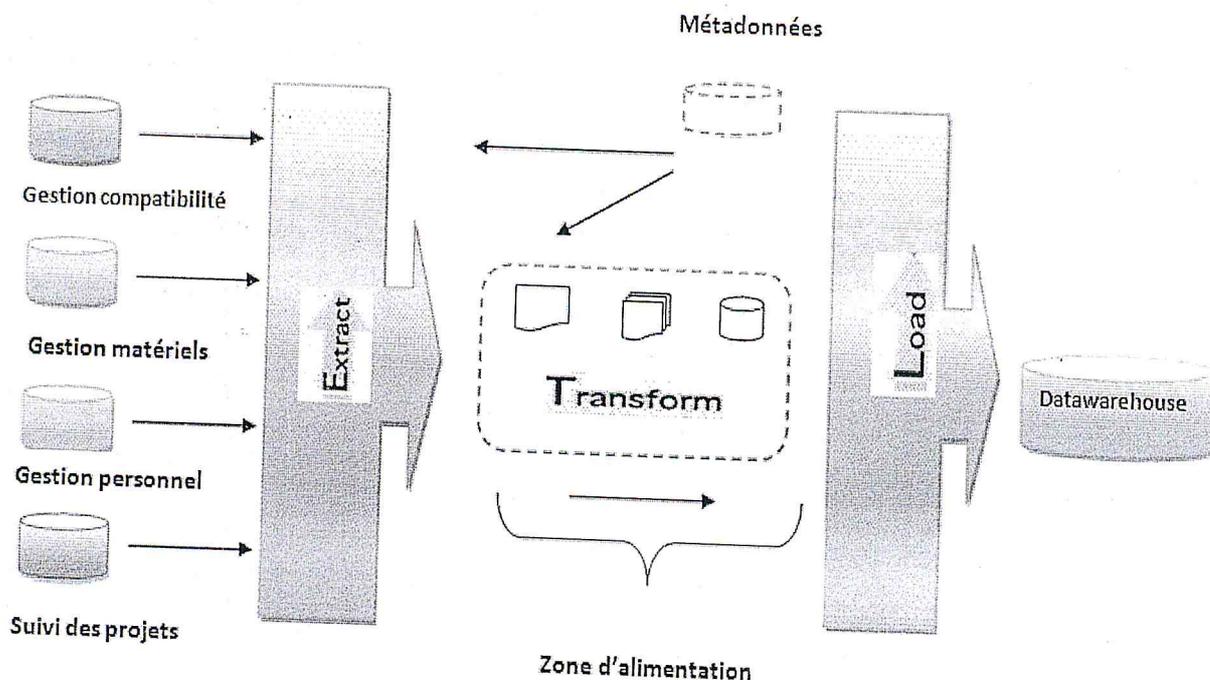


Fig4.15: le processus d'alimentation de notre entrepôt

Pour l'opération de transformation des données, On a utilisé un logiciel de la gamme oracle que l'entreprise utilise, c'est un produit qui automatise le processus de préparation des données, il s'agit en fait d'extraire ces dernières du système source, en suite les transformées pour être conforme au format de l'entrepôt et à la fin les chargées dans la zone de stockages qui correspond à l'entrepôt de données. Le produit oracle dont dispose l'entreprise est Oracle 9i Warehouse Builder release2.

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 - 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 donnée

réf_projet, dans l'application comptabilité est sous la forme n_affaire est codée sur 6 positions, dans l'application matériel : cod_chantier sur 4 position et dans l'application personnel : n_projet sur 4 positions et utiliser une fonction substr pour n_affaire de l'application comptabilité. Donc, elle va être transformée en n_affaire sur 4 positions avant d'être chargée dans l'entrepôt de données.

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 mensuellement, vu que les données sources sont mises à jours chaque mois.

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 suivant :

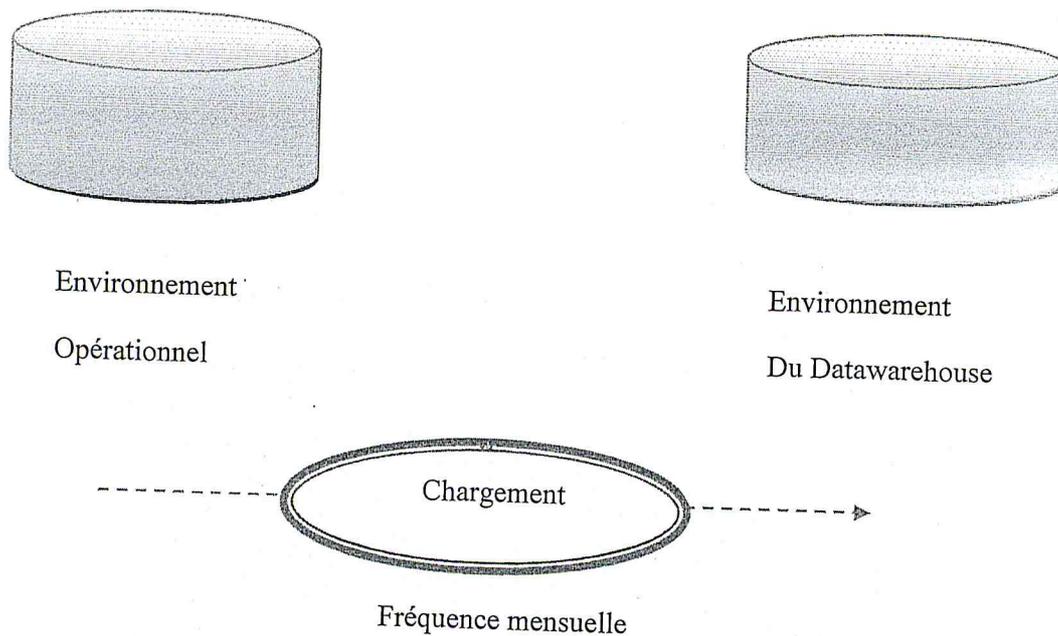


Fig 4.16 : La fréquence de chargement des systèmes sources vers l'entrepôt

7.5. 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 utilisateurs puissent accéder et comprendre.

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.

- 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 aux besoins des consommateurs, donc .

l'utilisation des informations de notre data Warehouse va se baser sur deux modes différents :

- **Le reporting** : il s'agit d'établir un rapport de données, qui peut être diffusé sur n'importe quel canal.
- **L'analyse OLAP** : il s'agit d'un processus d'analyse en ligne. Les données sont représentées sous forme multidimensionnelle.

7.5.1 . Le reporting

Il s'agit de développer des rapports prédéfinis, dont la diffusion peut être soit planifiée ou elle est laissée à l'initiative de l'utilisateur. Les informations y sont préstructurées sous la forme de tableaux de bord ou d'états prédéfinis.

7.5.2 Connexion à l'entrepôt de données

La figure 4.17 présentée ci-dessous illustre la fenêtre de connexion à l'entrepôt de données qui permettra aux utilisateurs autorisés d'y accéder après avoir introduire le nom d'utilisateur et le mot de passe.

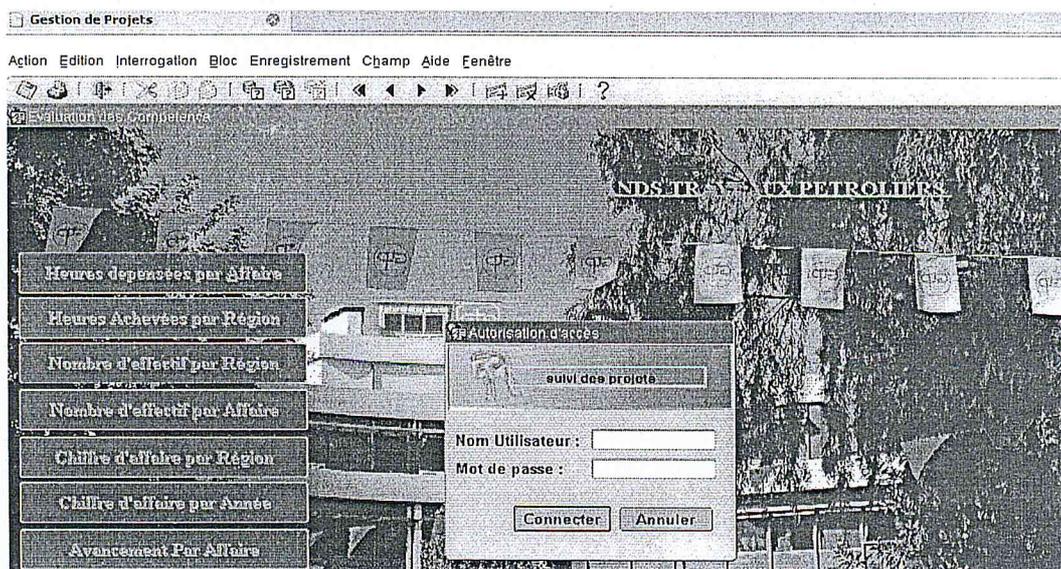


Fig 4.17 : Page web de connexion à l'entrepôt de données

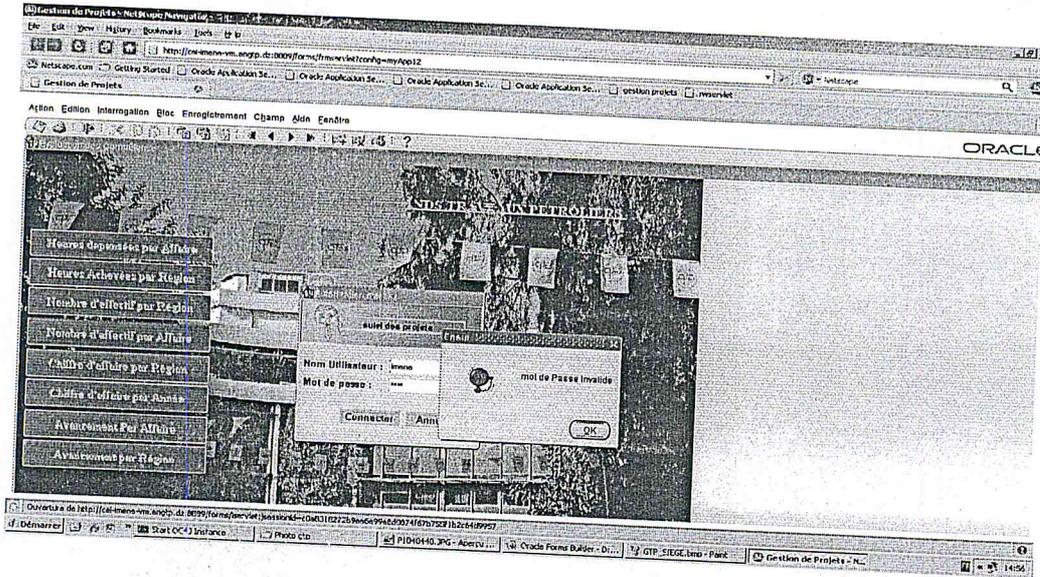


Fig 4.18 : Page Web signalant l'erreur

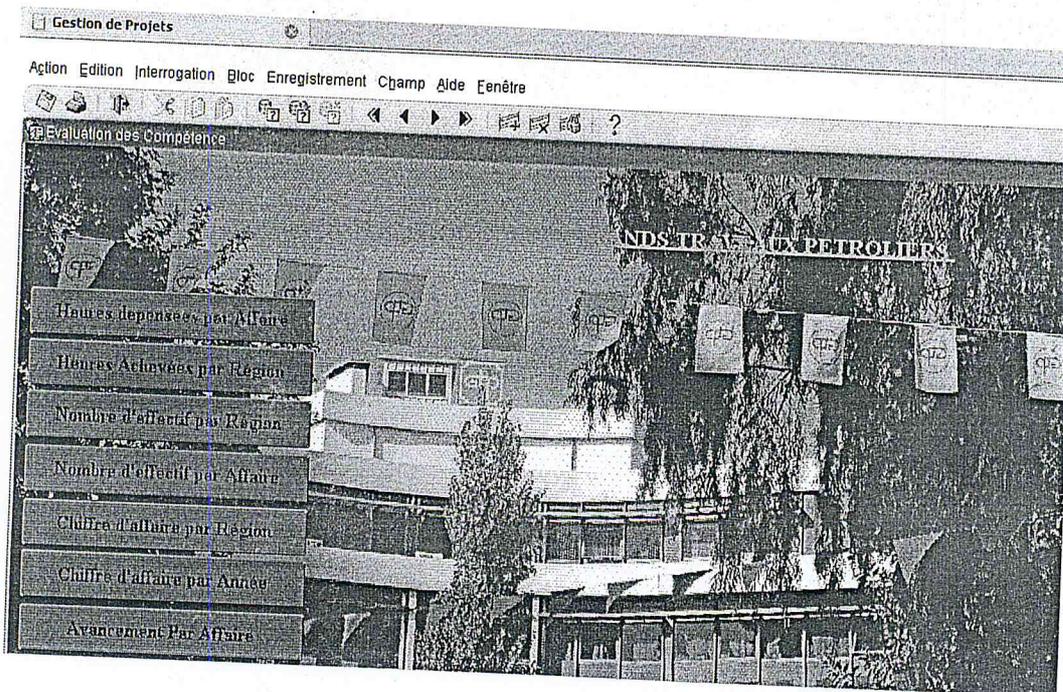


Fig 4.19 : Page Web du choix des états

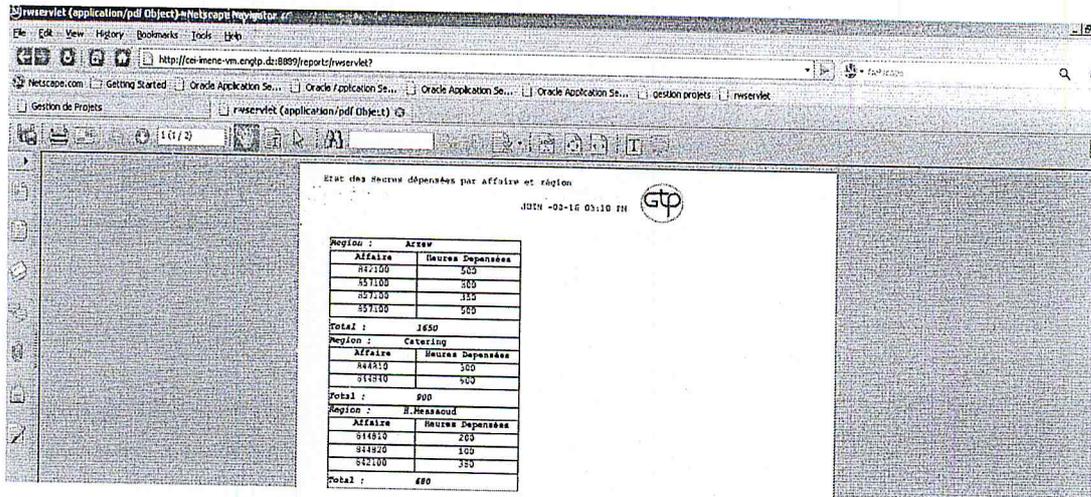


Fig 4.20 : Page Web de l'état « Heures dépenées par Affaire et par Région »

7.5.3 L'analyse Multidimensionnelle

L'analyse multidimensionnelle se fait grâce à la représentation multidimensionnelle des données et cela en utilisant les outils et les moteurs OLAP, elle nous permet de construire dynamiquement nos analyses à partir des données qui sont à sa disposition.

Pour l'approche de notre solution DS (Système Décisionnel) nous avons opté pour une analyse multidimensionnelle s'appuyant sur des cubes OLAP prédéfinis.

L'outil OLAP nous permet grâce à ces moteurs "Drill " de naviguer, visualiser et de manipuler les données dynamiquement, car il est à la possibilité de l'utilisateur de choisir

Comment représenter les données et d'en faire des coupes. Les drills permettent de raffiner la recherche des données, d'intervertir des axes d'analyses de ces indicateurs par simple clic.

- Un **Drill-up** permet de faire le cumul ou la consolidation des indicateurs en agrégeant les données suivant un axe d'analyse. Exemple

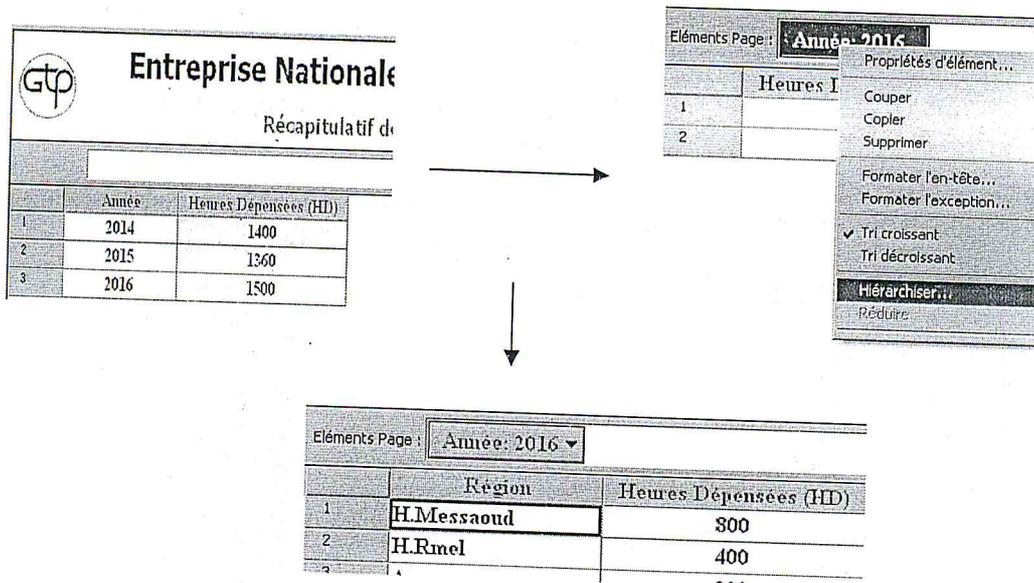


Fig 4.20 : Drill up

- Un **Drill-Down** permet de détailler une analyse en raffinant la granularité d'un axe d'analyse. Exemple .

Eléments Page : Année: 2016 ▼		
	Mois Année	Heures Dépensées (HD)
1	01/2016	340
2	02/2016	300
3	03/2016	400
4	04/2016	250
5	05/2016	210

Fig 4.21 : Drill-down

- Un **Slice & Dice** (sélection) permet d'intervenir deux axes d'analyse, par exemple pour analyser les heures dépensées du projet par rapport à l'axe temps et l'axe région. Exemple :

Eléments Page : Region: H.Rmel ▼		
	Année	Heures Dépensées (HD)
1	2014	450
2	2015	380
3	2016	400

Fig 4.22 : Slice & Dice

➤ Un **Pivoting** permet de transposer deux colonnes par un simple glisser-déplacer.

Exemple : On peut faire une duplication de notre table sous forme de matrice avant de faire un pivoting.

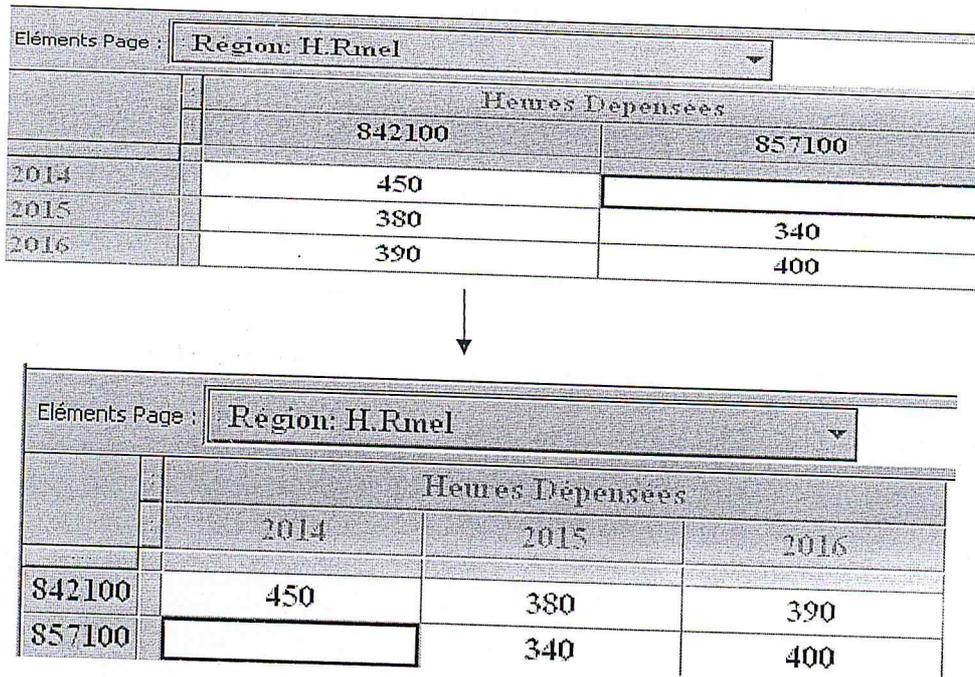


Fig 4.23 : Pivoting

8. Architecture technique de notre 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 (Oracle 9i comme serveur de base de données et Exchange 2000 server comme serveur de messagerie électronique).

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.

9. Disposition physique de notre système

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

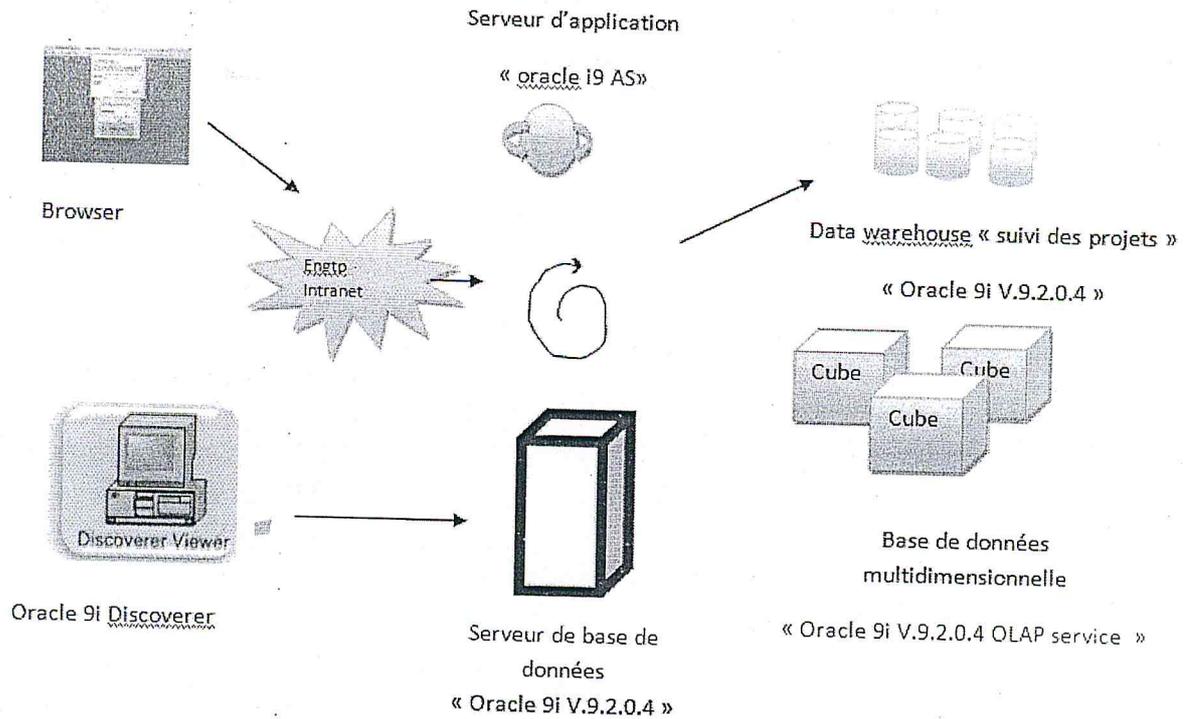


Fig.4. 24: Disposition physique de notre système

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.

Elle contient dans notre cas le serveur de Base de données 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 multidimensionnelle Oracle9i OLAP Service.

10. 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.

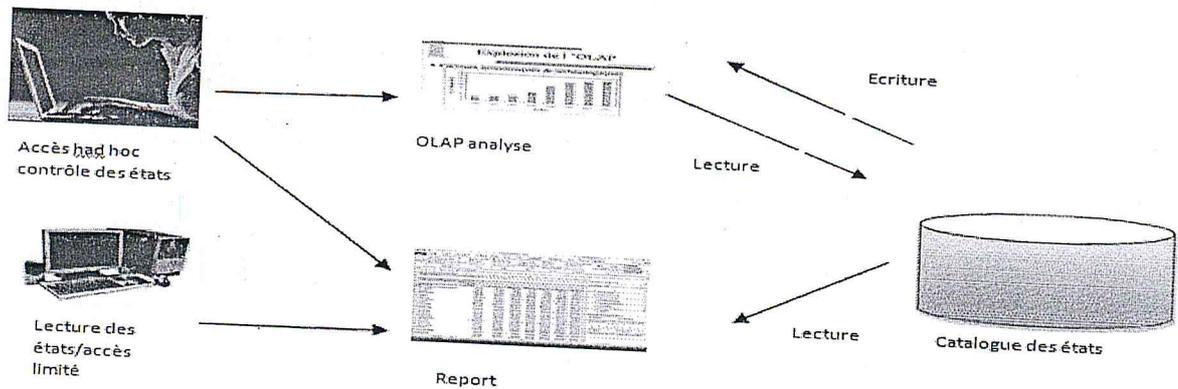


Fig.4. 25: Les utilisateurs de notre système

11. 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.

12. Logiciels utilisés

12.1. Les composants serveurs

- 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.
- Oracle 9i DB Developer (Forms Builder & Reports Builder).

12.2. Les composants clients

Les deux types d'application que nous avons réalisés, 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'analyse OLAP installée sur le poste de l'utilisateur, et le reporting accessible via l'Intranet selon le protocole HTTP

13. 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

Le système projeté doit être en mesure de se protéger contre les indiscretions 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 sont possibles : la sécurité physique et la sécurité logique.

La première est contre les dégâts matériels (Machine tombe en panne) et la deuxième est contre les dégâts logiciels (Virus, Perte de mots de passe, etc.).

Notre approche de sécurité pour les deux types de pannes est basée sur la norme standard de la Sécurité logique qui 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.

14. Conclusion

Dans ce chapitre nous avons présenté en premier lieu, la conception de l'entrepôt, en suivant le processus de Ralph Kimball, qui se base sur une représentation dimensionnelle des données, sous forme d'un diagramme en flocon qui intègre les concepts de faits et de dimensions. En effet, cette modélisation multi-dimensionnelle de la représentation des données est la mieux adaptée pour les analyses OLAP.

Dans ce chapitre nous avons aussi présenté l'architecture type d'un système décisionnel basé sur l'approche de l'entrepôt de données. Cette architecture est divisée en trois composantes essentielles (ETL, données, portail de restitution). Chaque couche peut être implémentée indépendamment des autres. Afin de valider notre modèle conceptuel dédié à l'entrepôt de données, nous avons réalisé un prototype d'un système d'aide à la décision dédié à la fonction « suivi des projets » dont le déploiement est sus-détaillé.

**CONCLUSION
GENERALE**

Conclusion Générale

Le but final de notre mémoire est d'apporter une solution aux problèmes sus notés,

Notre travail est d'être à l'écoute de l'entreprise. Une écoute systématique des décideurs nous ramène constamment à ce que nous sommes censés faire, identifier des enjeux, spécifier des modèles d'accès aux données, choisir une architecture sont les actes fondateurs les plus décisifs pour le développement d'un système décisionnel. Ils sont d'ailleurs, bien plus que les outils, les véritables éléments distinctifs d'un projet décisionnel.

Construire un système décisionnel, c'est d'abord se donner les moyens de dépasser ou de contourner des limites propres aux systèmes d'information existant. Or ces limites sont plus organisationnelles et méthodologiques que technologiques, et c'est là que réside la difficulté majeure.

La finalité de ce projet réside dans la réalisation d'un système décisionnel métier basé sur l'approche data Warehouse, dont l'intérêt majeur est d'offrir aux décideurs une vision plus expressive, à travers l'analyse multidimensionnelle, ou avec des états préalablement établis.

La méthodologie utilisée lors de la conception et le déploiement du système projeté suit l'approche du cycle de vie dimensionnel proposée par R. Kimball. Notre démarche est structurée en trois phases principales : Définition des besoins, conception du système et enfin le déploiement.

Dans la première phase, nous avons mené une série d'entretiens, à fin de mieux cerner les besoins des utilisateurs, et comprendre leurs attentes de ce système, c'est ce qui nous a permis en premier lieu d'identifier les sujets d'activités clés de la fonction « suivi des projet », pour pouvoir établir une liste des indicateurs exprimés par les responsables à fin de mieux piloter l'activité, de prendre de meilleures décisions et de rentabiliser ainsi leurs investissement.

Ce qui nous conduit, directement à la deuxième phase, il s'agit de la conception de notre système. Elle propose une modélisation dimensionnelle. Plusieurs sujets intéressants ont été identifiés, leurs tables de faits et leurs dimensions sont illustrées. On a fait correspondre les besoins de la communauté des utilisateurs avec la réalité des informations disponibles.

Dans la phase de mise en œuvre, nous avons construit le modèle physique de notre entrepôt. Ce dernier est basé sur le modèle relationnel. Les cubes OLAP construits sauvegarde l'agrégation des données stockées dans l'entrepôt.

La dernière phase concerne le déploiement du système. La base de données de l'entrepôt est créée et reliée avec le système opérationnel par l'intermédiaire de la zone d'alimentation. Nous avons aussi créé la base de données multidimensionnelle et nous avons également installé l'application de l'interface utilisateur. Notre système décisionnel est implémenté sur : Oracle9i Server comme serveur de données pour l'entrepôt et la base multidimensionnelle et l'architecture deux-tiers et trois tiers comme technologie de développement.

Ce mémoire est une base de réflexion concrète qui peut toujours être améliorée, corrigée ou complétée.

En perspectives, nous proposons ce qui suit :

- Généraliser le système décisionnel pour qu'il englobe tous les autres services de l'entreprise sous intranet.
- Prise en charge réelle des satisfactions clients.
- La qualité des données stockées au sein de l'entrepôt. Pour qu'une base de données soit correcte, il faut que toutes les données ne soient oubliées pour quelque raison que ce soit.

Ceci dit nous espérons que ce modeste travail pourrait aider toute personne intéressée par le monde du décisionnel. Et qu'il attire l'intérêt des responsables de l'entreprise pour l'exploitation des techniques décisionnelles dans l'amélioration de la qualité de management.

BIBLIOGRAPHIE

[1] « Développement d'un outil Extraction Transformation Load (ETL) », MARHOUMI Fatima Ezzahra, Mémoire de fin d'études en vue de l'obtention du grade d'Ingénieur Civil Informaticien en Sciences Appliquées/ Université Libre de Bruxelles/Faculté des Sciences Appliquées, 2005 – 2006.

[2] « Le Data Warehouse, le Data Mining », Jean-Michel Franco, EDS- institut Prométhus, Eyrolles, 1996.

[3] « L'informatique Décisionnelle », Victor Sandoval, Hermes, 1997.

[4] « Entrepôt de données, Guide pratique du concepteur de Datawarehouse », Ralph Kimball, Vuibert, 2001.

[5] « Conception et déploiement d'un entrepôt de données métier », Khaled BOUSSEBAT, Mohamed ABDELLI, Mémoire de fin d'étude, Promotion : 2002-2003

[6] « Building the Datawarehouse », Inmon W.H, John Wiley & Sons.

[7] « An Overview of Data Warehousing and OLAP Technology » Chaudhuri S, ACM SIGMOD Record. 1997

[8] El Mostapha Kerki, « Processus de mise en œuvre de l'entrepôt de données approche sémantique » Université de Bourgogne, 1998

[9] « Concevoir et déployer un Datawarehouse guide de conduite de projet », Ralph Kimball, Eyrolles, 2000.

[10] Ralph Kimball, Margy Ross « Entrepôt de Données: Guide Pratique de Modélisation Dimensionnelle » Vuibert, Paris, 2003

[11] « What is a Data Warehouse », B. Inmon, Article 2000.

[12] « Le Datawarehouse, le Datamining », Jean Michel Franco, EDS : institut Prométhus- Eyrolles-1996.

[13] « Providing OLAP(Online analytical processing) to user analysts : an IT mandate », Codd E.F Technical Report, 1993

[14] « Management Information Systems pour un service public ». David Pereira, MIS pour un service public.

[15] « Datawarehouse », C.Vangenot laboratoire de base de données, Eyrolles,2000

[16] Introduction à l'ETL et application avec Oracle/ Stéphane Crozat ;12 février 2015

[17] « Entrepôt de données : guide pratique de modélisation dimensionnelle », Ralph Kimball, Margy Ross ; Vuibert, paris, 2003.

Les Sites :

Site 1 : **<http://www.Wikipédia.org>**

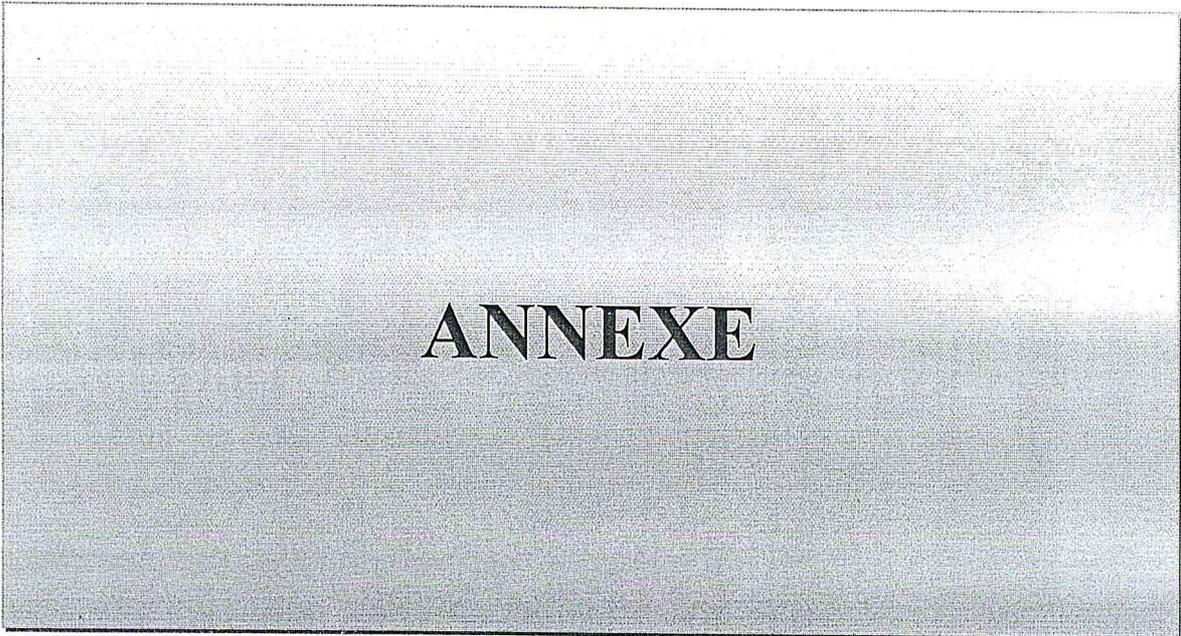
Site 2 : **http://www.tutorialspoint.com/dwh/dwh_olap.htm**

Site 3 : **<http://blog.developpez.com/bi/p7422/informations>**

Site 4 : **<http://www.tutorialspoint.com>**

Site 5 : **<http://www.dwfacile.com/>**

Site 6 : **<http://decizia.com/>**



ANNEXE

1. Présentation générale de l'entreprise

GTP est une entreprise de grande envergure spécialisée dans la construction en tous corps de métiers, de Grands Ensembles Industriels et de Canalisations dans, principalement, les domaines Oil & Gas et Energy.

Sa présence sur le marché depuis maintenant plus de 40 années, lui a permis de développer un large portefeuille d'activités et d'accumuler un savoir-faire, une expertise et des capacités qui l'on hissé au statut d'Entreprise Leader.

GTP a collaboré avec les grands constructeurs internationaux (General Contractors) dans la réalisation de la majorité des installations dont dispose Sonatrach dans le secteur des Hydrocarbures.

2. Organisation

2.1 Missions :

1. Construction de Grands Ensembles Industriels dans le domaine des hydrocarbures (Oil & Gas) et de l'énergie.
2. Pose de Canalisations (Pipelines).
3. Maintenance des Equipements et des Installations Industrielles
4. Engineering, Procurement et Contrôle de réalisation
5. Expertise et Contrôle dans le domaine du soudage

2.2 Domaines d'intervention

- 1- Production et traitements des hydrocarbures
- 2- Transport des hydrocarbures
- 3- Chimie et Pétrochimie
- 4- Production d'énergie électrique
- 5- Secteur matériaux de construction
- 6- Secteur agro-alimentaire

2.3 Activités

- 1- L'Engineering et le Procurement
- 2- Les activités de Réalisation
- ✓ Génie civil
- ✓ Montage industriel (MEI)
- ✓ Pose de Canalisations
- 3- La Préfabrication de Tuyauteries et de Charpentes Métalliques
- 4- La Maintenance Industrielle
- 5- Les activités de logistique et de soutien à la construction

3. Ressources de l'entreprise

3.1- Ressources Humaines

L'entreprise emploie actuellement près de 10.000 agents, répartis comme suit:

6. Productif : 74%
7. Support : 17%
8. Management : 09%
9. Permanents: 25%
10. Temporaires: 75%

3.2- Ressources Matérielles :

L'entreprise dispose d'un parc matériel composé de plus de 5.600 équipements de construction se répartissant comme suit:

11. Matériel roulant : 1087 U 20 %
12. Terrassement et génie civil : 345 U 06 %
13. Soudage: 1900 U 34 %
14. Canalisation : 402 U 07 %
15. Le levage (grues de 18T à 400T): 298 U 05 %
16. Matériel fixe ateliers de fabrication : 1114 U 20 %
17. Auxiliaires : 456 U 11 %

3.3- Capacités d'accueil :

Pour l'hébergement et la restauration de son personnel en région sud, GTP dispose de bases de vie dont la capacité totale est de 9000 places réparties comme suit:

18. 60 % mobiles
19. 40 % fixes

3.4-Corporate school

Pour la formation et la mise à niveau de son personnel dans les principaux métiers de l'entreprise dont, notamment, le Soudage, la Tuyauterie et le Contrôle Non Destructif, GTP dispose depuis 1978 d'une école située à Arzew, appelée

« Centre de Soudure et d'Expertise ».

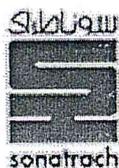
4.Principaux projets réalisés

Projets importants réalisés de 2001 à 2014 :

Intitulé du projet	Client	Volume Horaire (H)	Durée (Mois)	Période
Préfabrication Piping & Supports Arzew GNL 3	SAIPEM	1 022 380	26	2009 / 2012
EL Merk CPF MEI & Telecom work	PETROFAC	2 359 450	18	2010 / 2012
Gassi Touil CPF – MEI	JGC	2 520 300	24	2010 / 2012

5.Principaux clients de l'entreprise

1- Nationaux



2- Internationaux

