

PUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'enseignement Supérieur et de la Recherche Scientifique

Université M'HAMED BOUGARA – BOUMERDES



**Faculté des sciences
Département de physique
Structure Infotronique**

**Domaine : Sciences et Techniques
Filière : Génie Electrique
Spécialité : Infotronique**

**MEMOIRE DE FIN D'ETUDES
Pour l'obtention du diplôme de
Master**

Thème : Conception et réalisation d'un Dataware house

« Métier ressources humaines »

Réalisé par :

- **Messelem Mohamed merouane**
- **Mebtouche lilia**

Encadré par : Me.Hamza hakim

Membres de jurés:

Pr.Hamadouche

D.riahla Mohamed Amine

Mme.Mechid

Introduction

1.1	Les systèmes d'aide à la décision
1.2	Le système décisionnel dans l'organisation
	a) Le système opérant
	b) Le système de pilotage
	c) Le système d'information
1.3	Qu'est-ce qu'un système d'information décisionnel
1.4	Historique de l'informatique décisionnel
1.5	Le système opérationnel vs le système décisionnel
1.6	Objectifs des systèmes d'information décisionnelle
1.7	Les composants du décisionnel
	Conclusion
2.	Data warehouse
2.1	Introduction
2.2	Définition
a-	Définition 1
b-	Définition 2
c-	Définition 3
2.2.1	Discussion de la définition
2.3	L'origine du data warehouse
2.3.1	L'infocentre
2.4	Les objectifs du data warehouse
2.5	Les éléments d'un data warehouse
a-	Les applications opérationnelles
b-	Préparation des données
c-	Présentation des données
2.5.1-	Data Mart indépendant
2.5.2-	Data Mart interconnectés
a-	Zone d'outils d'accès
b-	L'architecture du data warehouse

2.6.1	Implémentation de l'architecture	
	a- L'architecture réelle	
	b- L'architecture virtuelle	
	c- L'architecture remote	
2.7	La structure des data warehouses	
	a- Les données détaillées	
	b- Les données agrégées	
	d- Le méta données	
	e- d- Les données historisées	
3.	Principales notions qui rentrent au processus du data warhousing	
	a. Le système source	
	b. La zone de préparation	
	c. Le serveur de présentation	
	d. Data Mart	
	e. Data Warehouse	
4.	Le portail de restitution	
	a. Data Mining	
	b. Outils statistiques	
	c. E.I.S (Executive Information System)	
	d. S.I.A.D (Système Interactif D'aide à La Décision)	
	
	e. Requêteur.....	
	f. L'outil OLAP	
	Conclusion	
	Chapitre : définitions des besoins	
	1. Introduction	
	1.1. L'utilité de l'étape «définition des besoins.....	

1.1.2. Description de la démarche d'étude des besoins.....	
a. Étude préliminaire des systèmes sources et interviews avec les utilisateurs	
b. Détection des postes susceptibles d'être porteurs d'informations utiles	
c. Planification, préparation et conduite des interviews	
1.1.3. La phase de planification	
1.1.3. La phase de préparation	
2. La conduite des entretiens	
2.1. Autres moyens utilisés pour la détection des besoins.....	
2.2 Rédaction et validation du recueil récapitulatif des besoins	
3. Problèmes et obstacles rencontrés	
4. L'utilité des entretiens.....	
Conclusion	

Introduction

1.Présentation de l'organisme d'accueil	
1.1. Historique de GTP	
1.2. Métiers de GTP	
2. Secteurs d'activité de GTP:	
3. Politique commerciale	
3.1. Produits de GTP	
a. Construction	
b. Engineering	
c. Procurement	
d. Maintenance industrielle	
e. Soudage	
3.2. Concurrence	
3.3. Clients de GTP	
a. Clients nationaux	

b. Clients étrangers avec Sonatrach

c. Clients étrangers

4. Principales réalisations de GTP

4.1. Ressources et politique de l'entreprise

4.1.1. Humaines

4.1.2 Matérielles

5. Situation informatique de l'entreprise

5.1. Aspect hardware

5.1.2 Aspect software:

a. Les systèmes d'exploitation

b. Les logiciels :

c. Les applications :

d. Le réseau informatique :

6. Caractéristiques:

6.1. Autres applications réseau utilisées :

7. Les données

7.1. Interconnexion via des liaisons louées

8. Les traitements

8.1. Sous-système Gestion du Personnel et Paie:

a. La gestion du Personnel

b. La paie

c. L'interface avec le système «Comptabilité»:

d. Les tableaux de bord et statistiques:

e. Sous-système Gestion Finances et Comptabilité

f. En Comptabilité Générale

g. En Comptabilité Analytique

9. ORGANIGRAMME DE L'ENGTP

9.1. Organigramme Général

9.2. Organigramme de la direction technique

Conclusion

Chapitre 4

Introduction

4.1. Définition

4.2 Historique

- a. Engine-based
- b. Database-embedded
- c. Code-generators

4.3 Les processus d'ETL

4.3.1 Le processus d'extraction

- a. Extraction complète
- b. Extraction incrémentale

4.3.2 Le processus de transformation

4.3.3 Le processus de chargement

4.4 Politiques de l'alimentation

4.4.1 Objectif de qualité de données dans un processus ETL

4.5 Démarches de construction d'un «Data Warehouse»

- a- Etape 1 : modélisation et conception .
- b- Etape 2 : alimentation .
- c- Etape 3: mise en œuvre .
- d- Etape 4 : administration et maintenance.

4.5.1 Approche «besoins d'analyse» :

4.5.2 Approche «sources de données» :

4.5.3 Approche mixte :

4.6 Mise en œuvre du Data Warehouse

Conclusion

ABREVIATIONS

BI : Business Intelligence.

DBA : Data Base Administrator.

DCF : Direction Comptabilité et Finance.

DIM : Dimension.

DR : direction régionale.

DSS : Decision Support system.

DW : Data Warehouse (Entrepôt de données).

ED : Etude et développent.

ENGTP : Entreprise Nationale de grands travaux pétroliers

ETL : Extract, Transform and Load (ETC).

FK: Foreign Key.

FTP : File Transfer Protocol.

GTP : Grand traveaux pétrolier.

HOLAP: Hybrid On Line Analytical Process.

MOLAP: Multidimensional On Line Analytical Process.

OLAP : On Line Analytical Process.

OLTP: On Line Transactional Process.

PDG: Président Directeur General.

PK : Primary Key.

ROLAP : Relational On Line Analytical Process.

SI : Systèmes d'Information.

SID: Systèmes d'Information Décisionnels.

SIAD : Systèmes d'Information d'Aide à la Décision

SGBD : Système de Gestion de Base de Données.

SPA : Société Par Action.

SQL : Structured Query Language.

Liste des figures

Figure I.1 :

Figure I.2 :

Figure I.3 :

Figure I.4 :

Figure I.5 :

Figure I.6 :

Figure I.7 :

Figure I.8 :

Figure I.9 :

Figure I.10 :

Figure I.11 :

Figure I.12 :

Figure I.13 :

Figure I.14 :

Figure I.15 :

Figure I.16:

Figure I.17 :

Figure I.18 :

Figure I.19 :

Figure I.20 :

Figure I.21 :

Figure II.1:

Figure II.2:

Figure II.3 :

Figure II.4 :

Figure II.5 :

Figure II.6 :

Liste des figures

Figure II.7 :

Figure II.8 :

Figure II.9 :

Figure II.10 :

Figure II.11 :

Figure II.12 :

Figure II.13 :

Figure II.14:

Figure II.15:

Figure II.16:

Figure II.17:

Figure II.18:

Figure II.19 :

Figure II.20:

Figure II.21 :

Figure II.22:

Figure II.23 :

Figure II.24:

Figure II.25 :

Figure II.26 :

Figure II.27 :

Figure II.28 :

Figure II.29 :

Figure II.30 :

Figure II.31 :

Figure II.32:

Figure II.33 :

Liste des tableaux

Tableau I.1 :

Tableau I.2 :

Tableau I.3 :

Tableau I.4 :

Tableau I.5 :

Tableau II.1 :

Tableau II.2 :

Tableau II.3 :

Tableau II.4 :

Tableau II.5 :

Tableau II.6 :

Tableau II.7 :

Tableau II.8 :

Tableau II.9 :

Tableau II.10 :

Tableau II.11 :

Tableau II.12 :

Tableau II.13 :

Tableau II.14:

Tableau II.15 :

Tableau II.16 :

Tableau II.17 :

Tableau II.18 :

Tableau II.19 :

Tableau II.20 :

Tableau II.21 :

Tableau II.22 :

Liste des tableaux

Tableau II.23 :

Tableau II.24 :

Tableau II.25 :

Tableau II.26 :

Tableau II.27 :

Tableau II.28 :

Tableau II.29 :

Tableau II.30 :

Tableau II.31 :

Tableau II.32 :

Tableau II.33 :

Tableau II.34 :

I. Contexte

Avec le développement des technologies de l'information, les entreprises ou les administrations produisent de très grands volumes de données à travers leurs systèmes opérationnels. Leurs données doivent être manipulées et exploitées dans le but de visualiser l'activité de l'administration et plus particulièrement visualiser la rentabilité de ses affaires. En revanche, Il s'avère très complexe de prendre les bonnes décisions aux bons moments afin de prévoir les changements en vue de s'adapter à un contexte décisionnel performant.

Ce système décisionnel permettra aux décideurs de disposer d'informations pertinentes et d'outils d'analyses puissants pour s'orienter vers les décisions nécessaires.

L'arrivée de la notion de Business Intelligence (informatique décisionnelle) vient nous répondre au besoin d'avoir une vision globale et optimiser le patrimoine informationnel de l'entreprise, et elle est utilisée dans divers domaines comme la gestion des ressources humaines qui est la structure objet de notre étude.

Afin de mieux exploiter les outils offerts par la BI, les entrepôts de données servent non pas à faire des analyses globales dans un gisement de données mais de prendre en compte de l'utilisateur dans ses différentes navigations multidimensionnelles. En se basant sur ce constat, l'utilité des préférences utilisateurs sont primordiales dans l'analyse et l'exploration des données utiles du cube.

Dans ce contexte, un des meilleurs outils stratégiques est le data warehousing, il permet de transformer la principale richesse des entreprises, qui sont les données contenues dans les systèmes informatiques. En outre, le data warehousing permet d'identifier de nouvelles opportunités de développement, et de suivre en temps réel la progression de l'entreprise.

Selon les experts du «Data warehousing » : Bien *utilisé, le data warehouse constitue une arme stratégique de développement.*

II. Positionnement de la problématique

Tous les processus d'analyse et de prise de décision au sein du groupe GTP, se basent essentiellement sur des rapports dont les données sont extraites à partir des systèmes transactionnels d'une manière manuelle.

Le volet ressource humaine, contient des utilisateurs qui sont chargés de tous les mouvements internes et externes, ils extraient les informations depuis une application appelée «mouvement» qui contient tous les informations du personnels existant. Toute modification ou changement se fera au niveau d'une application local.

Dans cette optique, la synthèse générale se fait après l'enregistrement des données qui sont ensuite transformés en plusieurs «tableaux de bords», pour que les décideurs puissent avoir une vision claire de l'entreprise et qui les aident à prendre des décisions qui semblent utiles.

Les systèmes opérationnels paraissent peu adaptés pour servir de support à la décision car ils utilisent en générale le modèle relationnel qui ne convient pas aux applications qui gèrent l'activité quotidienne de l'entreprise.

Face à –cette ampleur, les systèmes actuels se trouvent face à :

- La difficulté de gérer et d'organiser les données pertinentes et stratégiques qui se trouve au niveau de la base de données opérationnelle.
- La difficulté de réunir les données jugées importantes dans les différentes applications des services de l'entreprise ce qui les rend inaccessible.
- Difficulté d'extraire l'historique des données des années écoulées.
- Les accès simultanés des utilisateurs et les décideurs en particulier les rendent très lourd et créent des conflits de privilèges d'accès. Pour cela il est difficile de faire une analyse quotidienne sous un système opérationnel.
- La complexité des structures conceptuelles des systèmes opérationnels (plusieurs tables et de vues) ce qui va rendre les requêtes plus complexes (trop de jointures) donc couteuses en ressource et en temps de réponse.

III. Objectifs

Pour résoudre les problèmes précédemment décrits, L'ENGTP nous a confié dans le cadre de notre projet de fin d'études, de concevoir et réaliser la première version du data warehouse, qui répond aux besoins analytiques les plus urgents, et qui simplifie le pilotage des « activités » de l'entreprise.

Les objectifs attendus de notre travail sont :

- l'analyse des besoins des décideurs en termes de données analytiques dans les activités GRH.
- Le choix de l'architecture du data warehouse.
- L'installation, la configuration et le paramétrage de l'environnement de développement du data warehouse.
- Connaître les principaux domaines d'application des data warehouses.
- Connaître le paradigme du décisionnel (et son articulation avec le paradigme transactionnel)
- Connaître les principes, les étapes et les méthodes de la modélisation dimensionnelle.

L'objectif principal est la réalisation d'un data warehouse dédié à l'analyse de l'activité :

- Performant sur les temps de réponse et l'exactitude des résultats.
- Performant d'historier tout le capital de données de L'ENGTP.
- Restituant des informations synthétiques et précises répondant aux besoins des décideurs.

L'initiation de tout projet nécessite une phase de planification, celle-ci permet de définir les tâches à réaliser.

Afin de présenter notre travail, le présent mémoire est organisé en deux grandes parties et se présente comme suit :

Après une introduction générale dans laquelle nous présentons le contexte général du projet, ainsi que la problématique et les objectifs visés. **La première partie**, présente les aspects théoriques du domaine des systèmes d'information d'aide à la décision, en évoquant leurs définitions et concepts de bases relatifs aux « Entrepôt de données » à savoir : la modélisation dimensionnelle et l'alimentation.

Dans la deuxième partie, nous présentons le travail réalisé au sein du groupe GTP à travers les chapitres suivants :

Le chapitre un, présente l'organisme d'accueil, sa structure organisationnelle, son activité et la culture de l'entreprise en matière d'utilisation des technologies de l'information.

Le chapitre deux, présente l'étude de l'existant au sein de l'entreprise, ainsi que les définitions des besoins.

Le chapitre trois, contient la conception de la partie d'entrepôt de notre solution. Il présente entre autres les modèles dimensionnels des activités recensées.

Le chapitre quatre, se présente en deux étapes :

- **Etape une** : présenter la conception de la zone d'alimentation, ainsi que les stratégies adoptées.
- **Etape deux** : elle décrit l'architecture globale de la solution, et cela en présentant les différents outils intégrés et les volets de la solution développés. Elle décrit aussi la manière dont se passe le déploiement de la solution.

Une conclusion générale est présentée afin de synthétiser le travail réalisé et de citer les perspectives du projet.

Résumé du projet

L'entreprise national de grand travaux pétroliers «**ENGTP**», assure plusieurs missions dans le domaine de l'énergie , est une filiale de l'entreprise national **SONATRACH** ,qui s'occupe de la réalisation des grands travaux pétroliers, employant de très grands moyens humains . Le groupe **ENGTP** exécute plusieurs tâches, ce qui fait de lui un acteur incontournable de l'économie nationale.

Le groupe rencontre dans le cadre de son activité de gestion quelques problèmes dans sa politique dans le métier ressources humaines. Ces difficultés sont liées notamment aux données jugées importantes qui se trouvent éparpillées dans différentes applications des services de l'entreprise ce qui les rendent inaccessibles. Ces difficultés ont rendus tout effort d'analyse vain, et c'est pourquoi les dirigeants du groupe GTP envisagent de mettre en place un système informatique qui permet de procurer les informations nécessaires et fiables qui aide ainsi à prendre dans les meilleurs délais les décisions les plus appropriées.

Dans ce contexte et afin de répondre à ces attentes grandissantes le groupe à incité sa filiale spécialisée dans les systèmes d'informations.

Le but recherché étant d'aller vers la mise en place d'un système informatique décisionnel

« **Business Intelligence** ». Ce système sera construit autour d'une base de données dédiée totalement aux décisionnel un « **Data warehouse** » et répond à tous les besoins d'analyse du groupe.

Ce présent projet a donc pour vocation première de réaliser une telle base de données.

Mots clés : Data Warehouse « Entrepôt de données », Business Intelligence « B.I » Informatique Décisionnel .

Introduction

L'objet de ce chapitre est de décrire un concept central dans la recherche en système d'information, nous aborderons tant les supports théoriques de ces systèmes en intégrant notamment l'approche de la décision en situation, ainsi que l'architecture technologique nécessaire à la mise en œuvre de ces systèmes.

1. Les systèmes d'aide à la décision [5]

Le processus décisionnel consiste à transformer le capital de données d'une entreprise en informations pertinentes à partir desquelles les décideurs peuvent tirer des connaissances afin d'aboutir à de bonnes décisions touchant tous les niveaux de l'entreprise.

La décision est un acte par lequel un décideur opère un choix entre plusieurs options, permettant d'apporter une solution satisfaisante à un problème donné où l'exécution d'une action d'un projet avec toutes les conséquences que cette décision pourrait engendrer.

2. Le système décisionnel dans l'organisation :

Toute organisation peut être décrite selon trois systèmes : [9]

- **Le système opérant** : représente l'activité productrice de l'organisation qui consiste à transformer les flux primaires pour répondre aux besoins des clients.
- **Le système de pilotage** : correspondant à l'ensemble du personnel dirigeant qui régule, pilote et adapte l'organisation par leurs décisions.
- **Le système d'information** : permettant de collecter, conserver, traiter et restituer les données produites dans l'organisation.

Les systèmes d'informations assure le lien entre les systèmes de pilotage et opérant c'est- à dire : le système opérant produit les informations stockées dans le système d'information, qui après traitement assure la transmission de ces informations au système de pilotage lui permettant ainsi de connaître l'activité opérationnelle. Les décisions prises sont répercutées vers le système opérant au travers du système d'information.

Face aux importants défis que doivent relever les organisations (concurrence, développement à l'inter national , émergence de technologie ...) , le pilotage réclame aujourd'hui des systèmes dédiés efficaces .

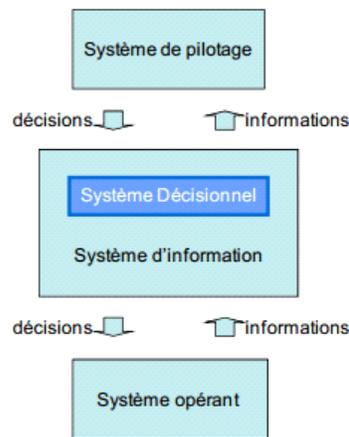


Figure I.1 : Positionnement du système décisionnel dans l'organisation.[Goglin,1998]

3. Qu'est ce qu'un système d'information décisionnel [5]

Système décisionnel ou informatique décisionnel (Business Intelligence BI), désigne les moyens, les outils et les méthodes qui concourent à rassembler, synthétiser et restituer l'ensemble des données d'une entreprise en vue de fournir à un décideur une aide à la décision.

4 .Historique de l'informatique décisionnel [5]

Dès les années 70, les données informatisées dans les organisations ont pris une importance qui n'a cessé de croître. Les systèmes informatiques gérant ces données sont utilisés essentiellement pour faciliter l'activité quotidienne des organisations et pour soutenir les prises de décision.

La démocratisation de la micro-informatique dans les années 80 a permis un important développement de ces systèmes augmentant considérablement les quantités de données informatisées disponibles. Face aux évolutions nombreuses et rapides qui s'imposent aux organisations, la prise de décision est devenue dès les années 90 une activité primordiale nécessitant la mise en place de systèmes dédiés efficaces.

A partir de ces années , les éditeurs de logiciels ont proposé des outils facilitant l'analyse des données pour soutenir les prises de décision. Les tableurs sont probablement les premiers outils qui ont été utilisés pour analyser les données à des fins décisionnelles.

Dès la fin des années 90, les acteurs importants tels que Microsoft , Oracle, IBM, SAP sont intervenus sur ce nouveau marché en faisant évoluer leurs outils et en acquérant de nombreux logiciels spécialisés ; par exemple, SAP vient d'acquérir Business Object pour 4,8 milliards d'euros. Ils disposent aujourd'hui d'offres complètes intégrant l'ensemble de la chaîne décisionnelle : E.T.L., stockage (S.G.B.D.), restitution et analyse. Cette dernière décennie connaît encore une évolution marquante avec le développement d'une offre issue du monde du logiciel libre (« open source ») qui atteint aujourd'hui une certaine maturité (Talend3, JPalo4, Jasper5).

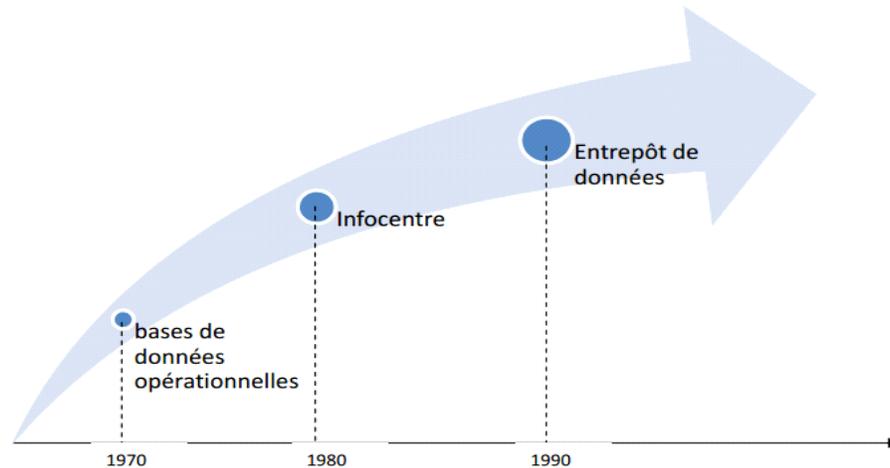


Figure I.2: évolution des systèmes décisionnels. [Goglin, 1998]

5. Le système opérationnel vs système décisionnel

Données opérationnelles	Données décisionnelles
Orientées application, détaillées, précises au moment de l'accès	Orientée 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

Tableau I.1 : Comparaison entre les données opérationnelles et décisionnelles [18]

6. Objectifs des systèmes d'information décisionnelle [6]

Les principaux objectifs d'un système décisionnel sont :

Offrir aux décideurs une vision transversale de l'entreprise en intégrant toutes ses dimensions.

- Répondre aux besoins des utilisateurs.
- Compléter les systèmes opérationnels.
- Fournir l'accès aux informations à des utilisateurs répartis de l'entreprise.
- Donner aux décideurs plus de temps de décision et moins de temps à la collecte et aux traitements des données.
- Automatiser la génération d'information.
- Faciliter la prise de décision.

7. Les composants du décisionnel

En relation étroite avec les nouvelles technologies de l'information et des télécommunications, le système décisionnel se manifeste à différents niveaux selon leurs utilités et leurs missions principales, comme illustré dans la figure suivante :

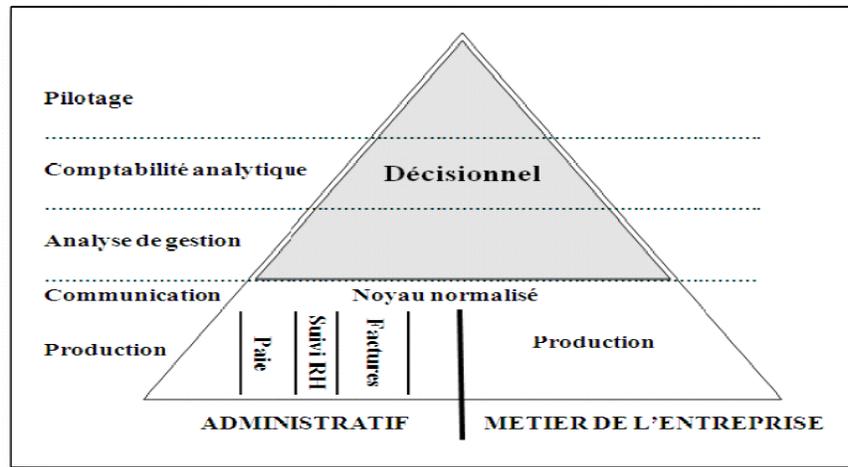


Figure I.3: le décisionnel au sein du système d'information[4]

Conclusion

Les premières applications dans les entreprises sont les outils informatiques d'aide au pilotage et à la décision, désormais appelés «*Business Intelligence*» qui consiste à transformer le capital de données d'une entreprise en informations pertinentes à partir desquelles les décideurs peuvent tirer des connaissances afin d'aboutir à de bonnes décisions touchant tous les niveaux de l'entreprise.

Introduction

L'objet de ce chapitre est de présenter les aspects théoriques des entrepôts de données «data warehouses», en évoquant leurs définitions et les concepts de bases relatifs à ces derniers. Ce chapitre introduit le contexte de nos travaux en définissant les différents concepts ayant servi de cadre à notre travail.

1. Définitions [8]

Il existe plusieurs définitions concernant le «Data warehouse», nous en présentons ci-dessous celles qui sont proposées par les théoriciens du domaine:

- **Définition 1** : selon **Bill Inmon [Inmon 2001]** considéré comme le «père du data warehousing», estime qu'un data warehouse est « un ensemble de données thématiques, cohérentes, évoluant dans le temps, fiables, sur lequel les dirigeants fondent leur processus de décision ».
- **Définition 2** : selon **Ralph Kimball [1996]** définit ce concept comme «un exemplaire de données relatives à des transactions structuré spécifiquement à des fins de consultation et d'analyse».
- **Définition 3** : selon **Barry Devlin [1997]** décrit un data warehouse comme étant « une mémoire unique, complète et cohérente de données provenant de sources diverses et mises à la disposition des utilisateurs finaux sous une forme compréhensible, et utilisable dans un contexte commercial ».

Les définitions précédemment énoncées, nous amènent naturellement à citer les caractéristiques des data warehouses :

1- Données orientées sujet : Au cœur du Data warehouse, les données sont organisées par thème. Les données propres à un thème, les ventes par exemple, seront rapatriées des différentes bases OLTP (online transaction processing) de production et regroupées.

2- Données intégrées : Les données proviennent de sources hétérogènes utilisant chacune un type de format. Elles sont intégrées avant d'être proposées à utilisation.

3- Données non volatiles : Les données ne disparaissent pas et ne changent pas au fil des traitements, au fil du temps (Read-Only).

4- Données historiques / datées : Les données non volatiles sont aussi horodatées. On peut ainsi visualiser l'évolution dans le temps d'une valeur donnée. Le degré de détail de l'archivage est bien entendu relatif à la nature des données. Toutes les données ne méritent pas être archivées.

2. Discussion de la définition

A partir des définitions et les caractéristiques fondamentales des data warehouse on peut dire de manière plus simple, que les entrepôts de données sont un prolongement des bases de données relationnelles classiques, utilisés spécifiquement dans le cadre de l'informatique décisionnelle .

Les données présentes dans cet entrepôt sont des données dérivées des sources d'informations de l'entreprise, elles sont périodiquement intégrées et réunies au sein de l'entrepôt .

3. . L'origine du data warehouse [8]

Le concept d'entrepôt de données «data warehouse»est bien plus ancien que l'on croyait ,il remonte aux années 60 durant lesquelles un intérêt croissant est porté au système décisionnel (Business intelligence BI) qui a vu le jour .

Le système décisionnel a évolué à partir de la recherche opérationnelle qui a commencé dans les années 1960, suivie par la supervision (DSS: Decision Support System) dans les années 1970, l'aide à la décision dans les années 1980 (entrepôt de données ,OLAP) et l'exploration de données (Data Mining) dans les années 1990.

Cette évolution s'est réalisée aux cours des périodes suivantes :

1960 : General Mills et l'université Dartmouth , dans un projet conjoint , créent les termes «faits « et «dimension» .

1983:Teradata introduit dans sa base de données managériale un système exclusivement destiné à la prise de décision.

1988: BarryDevlin et Paul Murphy publient l'article «une architecture pour les systèmes d'information financiers « («An architecture for a business and information system») ou ils utilisent pour la première fois le terme: «Data warhouse».

1990 :Red Brick systems crée Red Brick warehouse, un système spécifiquement dédié à la construction de l'entrepôt de données .

1991 : Bill Inmon publie Building the data warhouse(Construire l'entrepôt de données) .

1995 : Le data warehousingInstitute , une organisation à but lucratif destinée à promouvoir le data warehousing, est fondé .

1996 : Ralph Kimball publie :The data warehouseToolkit (la boîte à outils de l'entrepot de données) .

3.1 L'infocentre [4]

L'infocentre, est l'ancêtre des bases de données décisionnelles.

C'est un système de gestion de base de données relationnelle présentant une copie de travail d'une partie de la base de production, mise à jour périodiquement.

L'infocentre était une première solution pour soulager le système de production des requêtes complexes du décideur.

Il permettait en effet de transférer les données de base de type non relationnel dans un univers plus propice à l'interrogation inattendue.

Avec l'accroissement des besoins en matière de décision, que ce soit en termes de quantité de données collectées ou en nombre d'utilisateurs potentiels, l'infocentre se révéla bien insuffisant. Le Data warehouse l'a rapidement remplacé.

4. Les objectifs du data warehouse

Le problème des entreprises d'aujourd'hui consiste en la gestion des informations et de leurs organisations, en effet les données sont dans les différents supports et plates formes, ce qu'il devient difficile pour l'utilisateur de rechercher des faits marquants dans une montagne de données qui s'appuie sur un capital informationnel existant, lequel est constitué d'informations internes à l'entreprise, géré par ses systèmes d'information ,qui doit être complète par des données externes, acquises auprès de partenaire ou des société, ce qui donne un accès difficile et souvent prend un temps précieux aux utilisateurs ainsi que des traitements supplémentaires.

L'utilisation du Data warehouse résout toutes les préoccupations des entreprises, en effet il offre beaucoup d'avantage à celle-ci, et 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 data warehouse tels définis par Ralph Kimball , dans son livre «entrepôt de données, guide pratique du concepteur de data warehouse».

- **Accès aux informations de l'entreprise :** en effet, dans certains cas la donnée se trouve dans des systèmes anarchiques dont l'accès est impossible sans l'utilisation des programmes spéciaux adéquats à leurs structures. C'est pour ce la l'utilisation du data warehouse assure une vue transversal qui permet l'accès aux informations de l'entreprise.
- **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 les données en provenance du système de production doivent avoir la même signification que dans «l'entrepôt de données» où les données doivent être chargées dans leurs totalités.

En sorte que les données du «data warehouse» peuvent séparées et combinées au moyen de toutes les mesures possibles de l'activité.

- **Les outils de présentation d'information font partie du data warehouse :**

Le concept de data warehouse 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 data warehouse. L'outil d'aide à la décision est «la partie visible de l'iceberg» pour l'utilisateur, mais aussi un ensemble de requêtes, d'analyse et de présentations des informations.

- **Les données publiées sont stockées dans le data warehouse :**

Le data warehouse est un entrepôt de données dont les données qui le constituent sont d'origine différentes, recueillies de diverses sources d'informations de l'entreprise. Ce concept décrit un stockage des données issus des bases opérationnelles, dans différents supports, sous divers formats et sur différentes plates-formes informatiques.

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.

- **Qualité de l'information d'un data warehouse[3]**

Le data warehouse offre une information d'une 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 et 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.

La question posée est : 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.

5. Les éléments d'un data warehouse[3]

L'environnement du Data Warehouse est constitué essentiellement de quatre composantes : les applications opérationnelles, la zone de préparation des données, la présentation des données et les outils d'accès aux données.

- **Les applications opérationnelles** : ce sont les applications du système opérationnel de l'entreprise et dont la priorité est d'assurer le fonctionnement de ce dernier et sa performance. Ces applications sont extérieures au Data Warehouse.
- **Préparation des données** : la préparation englobe tout ce qu'il y a entre les applications opérationnelles et la présentation des données. Elle est constituée d'un ensemble de processus appelé ETL, « Extract, transform and Load », les données sont extraites et stockées pour subir les transformations nécessaires avant leur chargement.
« Un point très important, dans l'aménagement d'un entrepôt de données, est d'interdire aux utilisateurs l'accès à la zone de préparation des données, qui ne fournit aucun service de requête ou de présentation ». [5]
- **Présentation des données** : c'est l'entrepôt où les données sont organisées et stockées.

Si les données de la zone de préparation sont interdites aux utilisateurs, la zone de présentation est tout ce que l'utilisateur voit et touche par le biais des outils d'accès.

L'entrepôt de données est constitué d'un ensemble de Data Mart. Ce dernier est défini comme étant une miniaturisation d'un Data Warehouse, construit autour d'un sujet précis d'analyse ou consacré à un niveau départemental.

Cette différence de construction, autour d'un sujet ou au niveau départemental, définit la façon d'implémentation du Data Mart au niveau de l'entrepôt. On distingue, en effet, deux architectures internes du Data Warehouse :

5.1- Data Mart indépendant

Les Data Mart sont des versions miniaturisées du Data Warehouse au niveau départemental, alimentées par le Data Warehouse et basées sur les besoins départementaux en informations.

5.2- Data Mart interconnectés

Les Data Mart sont construits autour de sujets , interconnectés grâce aux tables des faits contenues dans le Data Warehouse , ce dernier se compose alors des Data Mart et ces tables des faits , appelées bus .

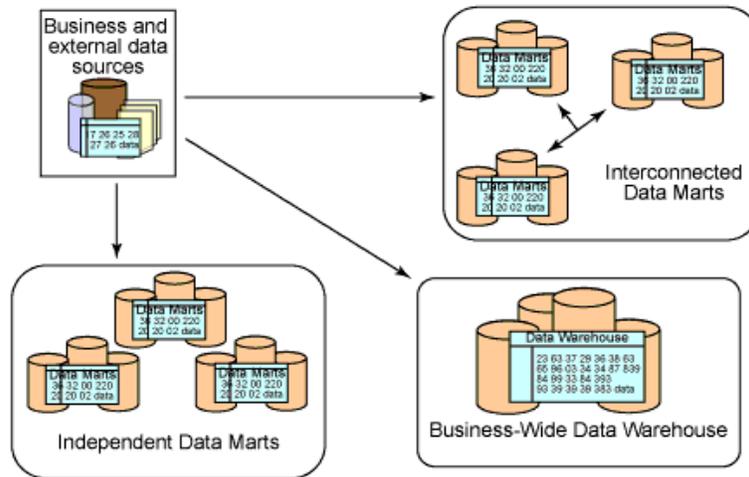


Figure I.4 :Les Data marts indépendant & interconnectés.[Inmon, 2002].

Zone d'outils d'accès [7]

c'est l'ensemble des moyens fournis aux utilisateurs du Data Warehouse pour exploiter la zone de présentation des données en vue de la prise de décision. Ces outils varient des simples requêtes ad hoc aux outils permettant l'application de forage de données plus complexes . Environ 80 à 90% des utilisateurs sont desservis par des applications d'analyses préfabriquées , consistant essentiellement en des requêtespréétablies.

6. L'architecture du data warehouse [7]

Chapitre 2 : Les concepts du data warehousing.

Après avoir exposé et défini chacun des éléments constituant l'environnement d'un data warehouse, nous présenterons les composants de base d'un data warehouse .

Pour cela, il serait intéressant de connaître le positionnement de ces éléments dans une architecture globale comme le montre la figure ci-dessous :

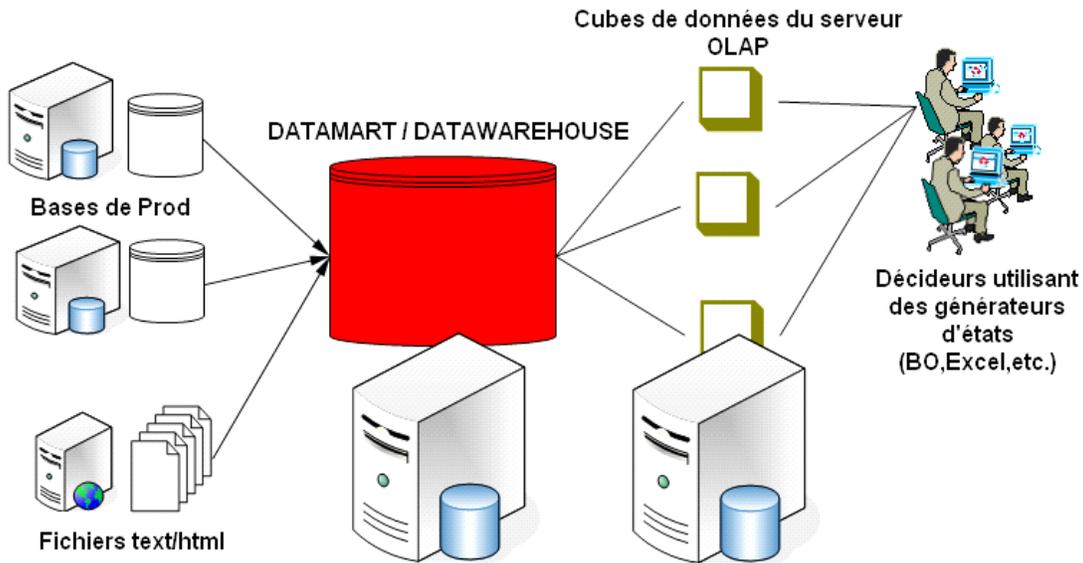


Figure I.5 : Architecture globale d'un entrepôt de données DW. [Kimball, 2002].

6.1 Implémentation de l'architecture[8]

Pour implémenter un Data Warehouse , trois types d'architectures sont possibles :

- L'architecture réelle.
 - L'architecture virtuelle.
 - L'architecture remote.
-
- **L'architecture réelle**

Elle est généralement retenue pour les systèmes décisionnels .

Le stockage des données est réalisé dans un SGBD séparé du système de production . Le SGBD est alimenté par des extractions périodiques.

Avant le changement , les données subissent d'importants processus d'intégration , de nettoyage ,de transformation.

L'avantage est de disposer de données préparées pour les besoins de la décision et répondant aux objectifs du data warehouse .

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

- **L'architecture virtuelle**

Cette architecture n'est pratiquement pas utilisée pour le data warehouse.

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

Il en résulte deux avantages : pas de coût de stockage supplémentaire et l'accès se fait en temps réel.

L'inconvénient est que les données ne sont pas préparées.

- **L'architecture remote**

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

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

7. La structure des data warehouses[3]

Un data warehouse peut se structurer en quatre classes de données, organisées selon un axe historique et un axe de synthèse. Cette structure est définie comme suite :

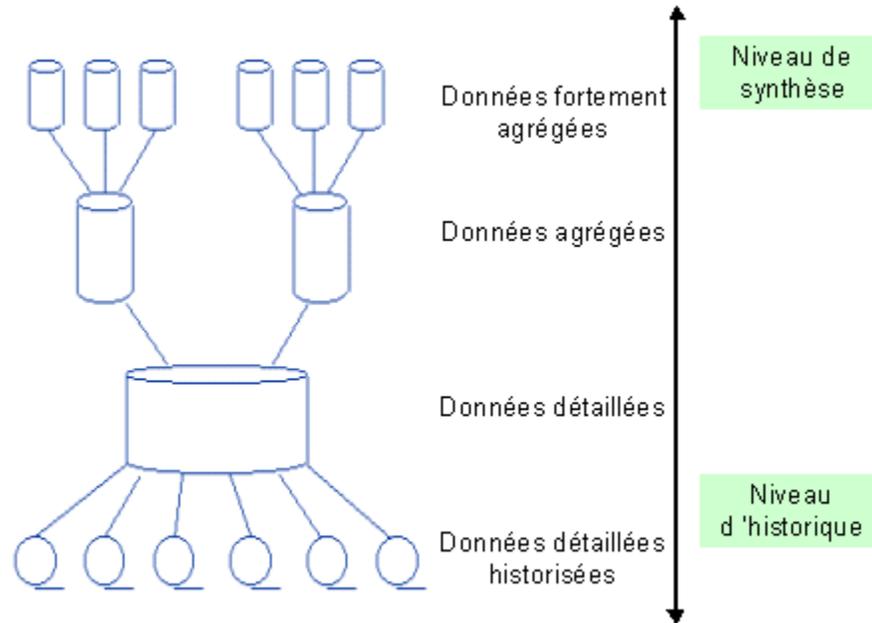


Figure I.6 : Structure des données d'un data warehouse.[20]

- **Les données détaillées**

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

Les volumes à traiter sont plus importants que ceux gérés en transactionnel. le niveau de détail géré dans le data warehouse n'est pas forcément identique au niveau de détail géré dans les systèmes opérationnels. La donnée insérée dans le data warehouse peut être déjà une agrégation ou une simplification d'informations tirées du système de production.

- **Les données agrégées**

Elles correspondent à des éléments d'analyse représentatifs des besoins utilisateurs. Elles constituent déjà un résultat d'analyse et une synthèse de l'information contenue dans le système décisionnel , et doivent être facilement accessibles et compréhensibles. La facilité d'accès est apportée par des structures multidimensionnelles qui permettent aux utilisateurs de naviguer dans les données suivant une logique intuitive, avec des performances optimales. (Certains SGBD du

marché sont conçus pour faciliter la mise en place des agrégations et la navigation au sein de celle-ci) .

La définition complète de l'information doit être mise à la disposition de l'utilisateur pour une bonne compréhension. Dans le cas d'un agrégat, l'information est composée du contenu présenté (moyenne des ventes...) et de l'unité (par mois , par produit ...) .

- **Les méta données**

Elles regroupent l'ensemble des informations concernant le data warehouse et les processus associés. Elles constituent une véritable aide en ligne permettant de connaître l'information contenue dans le data warehouse. Elles sont idéalement intégrées dans un référentiel.

Les principales informations sont destinées :

A l'utilisateur (sémantique, localisation) .

Aux équipes responsables des processus de transformation des données du système de production vers le data warehouse(localisation dans les systèmes de production, description des règles , processus de transformation) .

Aux équipes responsables des processus de création des données agrégées à partir des données détaillées.

Aux équipes d'administration de la base de données (structure de la base implémentant le data warehouse) .

Aux équipes de production (procédures de changement, historique de mise à jour) .

- **Les données historisées**

Un des objectifs d'un data warehouse est de conserver en ligne les données historisées. Chaque nouvelle insertion de données provenant du système de production ne détruit pas les anciennes valeurs,mais crée une nouvelle occurrence de la donnée. Le support de stockage dépend du

volume des données, de la fréquence d'accès , du type d'accès. Les support les plus couramment utilisées sont les disques , les disques optiques numériques ,les cassettes .

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

L'accès des données se fait également directement par les données détaillées et historisées , ce qui conduit à des brassages de données lourds, demandant ces machines très puissants.

8. Principales notions qui rentrent au processus du data warehousing [4]

1- Le système source

Cela correspond à des données disparates dans toutes l'entreprise , des informations de toutes sortes , en provenance d'une multitude de sources , de niveaux de détail différents , et s'échelonnant dans le temps.Elle peuvent être internes à l'entreprise ou externes .

Dans un système de production , la donnée est mise à jour à chaque nouvelle transaction. Ils sont en effet construits dans le but de traiter des opérations individuelles qui peuvent impliquer différents métiers de l'entreprise et surtout, ne se préoccupent pas de leur compilation ou historisation dans le temps .

2- La zone de préparation

Cette zone est la frontière entre la zone de présentation des données et les système source . En effet, les données doivent obligatoirement transitées dans cette zone qui comprend des processus qui nettoient, transforment, combinent, et archivent les données, afin de les intégrées dans l'entrepôt de données pour l'interrogation, l'exploitation et l'analyses de ces dernières. Elle ne doit en aucun cas être accessible à l'utilisateur final par requêtes ou par un quelconque autre service de présentation.

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

3- Le serveur de présentation

Ce composant représente la machine cible ,la ou tout a été soigneusement stocker et organiser. Après avoir préparer les données sources et alimenter l'entrepôt de données, elles sont désormais totalement interrogeables et exploitables pour répondre en accès direct aux requêtes provenant des utilisateurs, de générateurs d'états ou d'autres applications.

Généralement le data warehouse est conçu sur des bases de données relationnelles , multidimensionnelles ou objet .

Dans la majorité des cas , le serveur est basé sur une base de données relationnelle, de sorte que les tables y sont organisées sous forme de schémas en étoile.

4 -Data Mart

Le terme «data mart « est réservé à des bases plus petites, dédiées à une activité stratégique (data Martcommercial , data Mart financier,...) ou un département.

D'après Jean-MichelFranco , le data Mart ne constituera qu'un ensemble du capital informationnel de l'entreprise.

5- Data Warehouse

Le data warehouse , ou entrepôt de données, est utilisé comme stockage intermédiaire des données issues des applications de production et rassemble une copie des données vitales de l'entreprise. Il peut ainsi contenir les commandes, la facturation , la production , les prospects...

Les utilisateurs finaux y puisent avec des outils de restitution et d'analyse , comme les décisionnels , pour suivre l'activité.

Le data warehouse est la forme la plus sophistiquée des systèmes d'aide à la décision.

6 -Le portail de restitution

C'est la partie visible de l'entrepôt de données , elle représente l'interface homme machine ainsi que l'ensemble d'outils avec lesquels ils travaillent. Sous ce terme sont regroupées toutes les

applications qui s'appuient sur les données du data warehouse pour les restituer soit à l'utilisateur, soit à une autre application. Les services offerts par le portail de restitution sont les services d'accès aux données, les applications de modélisation et data mining.

7- Data Mining[6]

Le terme de data mining signifie littéralement forage de données.

Comme dans tout forage, son but est de 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, description et, dans une moindre mesure, l'optimisation.

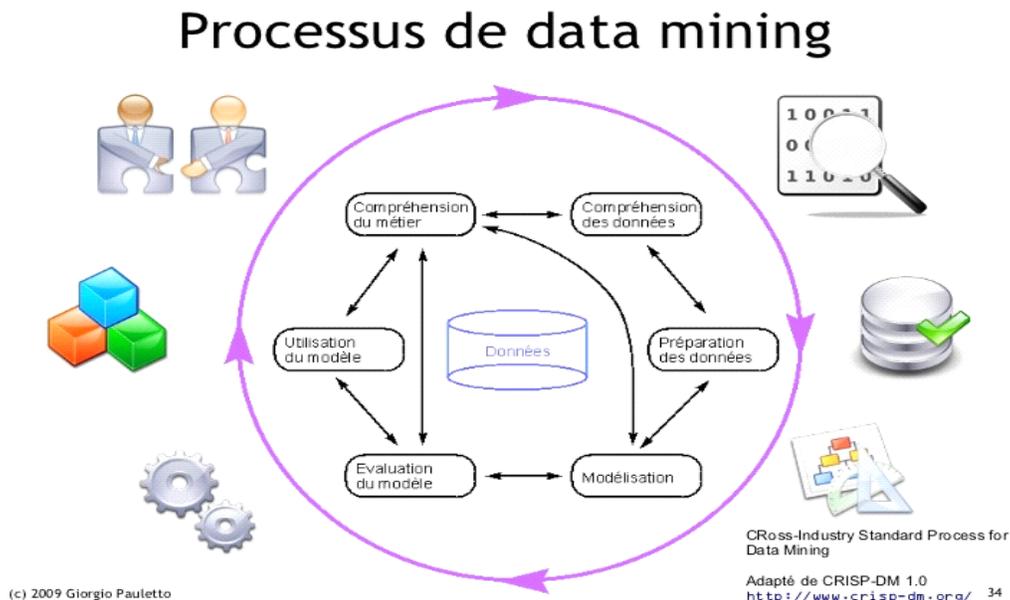


Figure I.7: Le data mining dans les entrepôt de données. [Goglin, 1998].

8- 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'informations. Pour chaque variable , on va ainsi rechercher au moins deux indicateurs : un pour mesurer la tendance centrale , un pour mesurer la dispersion .

9- E.I.S (Executive Information System)

Un E.I.S 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 les 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.

10- S.I.A.D (Système Interactif D'aide à La Décision)

Un S.I.A.D est un outil d'analyse et de modélisation des données de l'entreprise qui permet de créer des représentations multidimensionnelles de l'information, historiquement , il s'agit d'une terminologie et d'outils utilisés avant l'avènement et la maturité du data warehouse.

11- 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 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 couts de mise en place et de maintenance.

12- L'outils OLAP

On Line 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 vu de l'utilisateur.

Conclusion

Le « Data Warehouse » est une collection de données thématiques , intégrés , non volatiles et historisées pour la prise de décision . Il est apparu comme une réponse à des besoins dans le domaine décisionnel. Son adaptabilité et sa capacité de fournir les données nécessaires à une bonne analyse, ont fait de lui un atout majeur et incontournable pour toute entreprise soucieuse du suivi de ces performances.

Introduction

La modélisation dimensionnelle souvent appelée modélisation OLAP se présente comme une alternative au modèle relationnel. C'est une méthode de conception logique qui vise à présenter les données sous une forme standardisée intuitive et qui permet des accès hautement performants. Elle aboutit à présenter les données non plus sous forme de tables mais de cube centré sur une activité.

1. La modélisation dimensionnelle et ses concepts [21]

Le modèle dimensionnel est désormais très largement employé pour la conception d'entrepôts de données.

Il existe plusieurs techniques de modélisation dimensionnelle, des plus simples aux plus avancées. Après avoir commencé par exposer les recommandations de conception fondamentales. On passe aux directives de conception de modèles dimensionnels qui seront illustrés par des études de cas tirées du monde réel.

La maîtrise de l'ensemble des techniques permettant de créer, gérer et consulter des bases de données dimensionnelles d'entreprise.

1.1 Concept de fait ou indicateur

Les faits représentent les informations quantifiées de l'entreprise. Nous pouvons les nommer faits, indicateurs ou encore mesures. Ce sont les données à analyser qui correspondent à l'activité de l'entreprise. Les indicateurs ont la particularité d'être additifs. Ils sont contenus dans une table physique de la base de données décisionnelle. Nous nommerons « portefeuille d'indicateurs », « table des faits » ou « table des mesures » le regroupement de plusieurs indicateurs.

Les indicateurs n'ont d'intérêt que s'ils sont mis en valeur par des informations. Une ligne de faits correspond aux valeurs de l'intersection des tables des dimensions. Grâce aux dimensions, nous déterminons le grain (la finesse) des résultats contenus dans la table des faits.

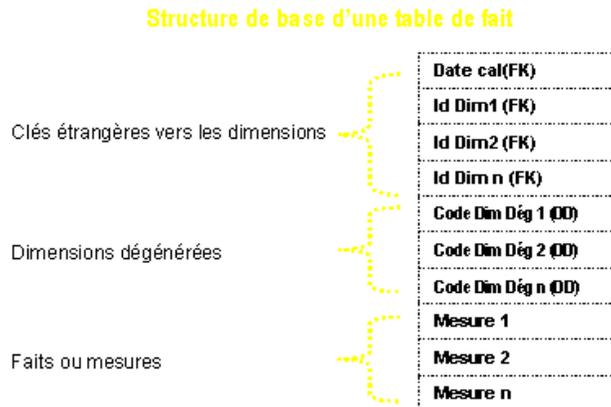


Figure I.8: Structure d'une table de « fait ». [Nakache,1998]

1.2 Concept d'attribut et de dimension [22]

- **Concept d'attribut :** il s'agit généralement des champs textuels qui décrivent une caractéristique d'un élément que chacun peut constater, l'attribut est souvent connu d'avance et il n'est pas mesurable .
- **Concept de dimension :**

Tout comme les faits, les dimensions sont contenues dans des tables physiques de la base de données.

Ce sont des informations qui vont mettre en évidence les données contenues dans les tables des faits. Lorsque nous parlons de dimension nous parlons également « d'axe d'analyse ». Une dimension regroupe les valeurs de même type. Par exemple dans la dimension géographique nous pourrions avoir : le continent, pays, région, ville, quartier, rue, bâtiment, étage, porte. Grâce à cet exemple nous comprenons mieux le sens d'« axe d'analyse », puisque nous distinguons tout de suite une hiérarchie au sein de la dimension.

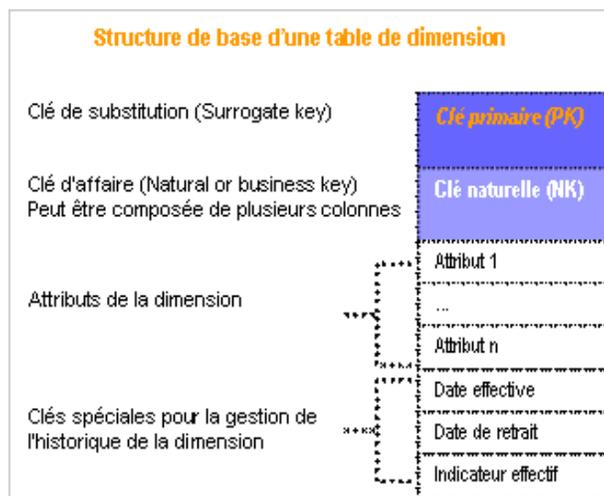


Figure I.9: structure de base d'une table de dimension[Kimball,2002]

Nous nous apercevons immédiatement de la corrélation entre les tables des faits et les tables des dimensions. Et de là on tire deux principaux rôles pour la «Dimension» :

- **Afficher les données** : Ce seront les entêtes des lignes ou des colonnes pour regrouper les faits, nous caractérisons ainsi la donnée brute contenue dans la table des faits.
- **Filtrer les données** : Nous allons choisir telle ou telle valeur de dimension afin d'obtenir un tableau correspondant à nos attentes.

2. Comparaison entre les tables « fait » et « dimension »

Le tableau suivant récapitule la différence entre ces deux tables :

	Tables de faits	Tables de dimensions
Structure	Peu de colonnes beaucoup de lignes	Peu de lignes beaucoup de colonnes
Données	Mesurable, généralement numérique	Descriptives généralement textuelles
Référentiel	Plusieurs clés étrangères	Une clé primaire
Valeur	Prend de nombreuses valeurs	Plus ou moins constantes
Manipulation	Participe à des calculs	Participe à des contraintes
Signification	Valeurs de mesure	Descriptive
Rôle	Assure les relations entre les dimensions	Assure l'interface homme / entrepôt de données

Tableau I.2 : Tableau qui différencie les deux tables. [Nakache,1998]

3. Les modèles de modélisation dimensionnelle [23]

Nous venons de voir les concepts de table des dimensions et table des faits. Celles-ci peuvent se structurer de différentes façons. Nous allons étudier trois modélisations multidimensionnelles :

- La modélisation en étoile.
- La modélisation en flocon de neige.
- La modélisation en Constellation.

3.1 La modélisation en étoile

La Modélisation en étoile représente une table de faits connectée à un ensemble de tables de dimensions, Cette modélisation ne tient pas compte des formes normales, car elle a uniquement une préoccupation, celle de l'analyse (lecture des données). La table des faits est la seule table à contenir des jointures avec les dimensions.

Cette déclinaison nous permet de dessiner la granularité des dimensions : temps, acteur (étudiant), géographie.

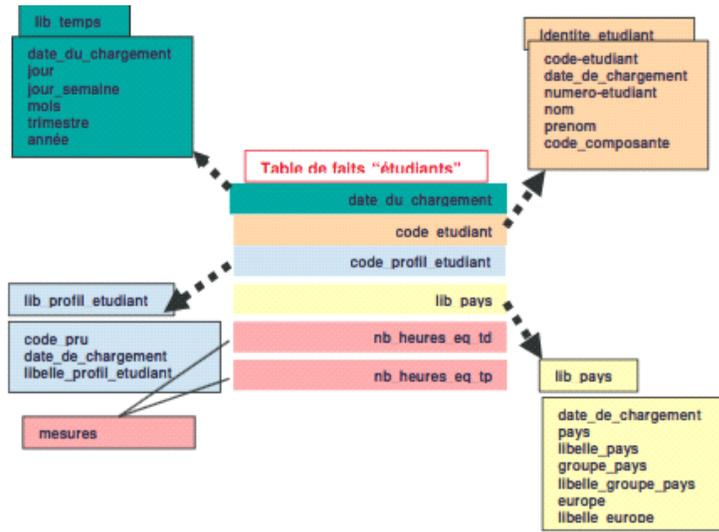


Figure I.10 : modélisation en étoile.[27]

3.2 La modélisation en flocon de neige

Le modèle en flocon de neige est constitué d'une table des faits au centre et des tables des dimensions autour, comme pour le modèle en étoile. La différence se situe au niveau des tables des dimensions, qui peuvent également se diviser en plusieurs branches différentes. Ces branches sont souvent utilisées pour modéliser des hiérarchies.

Cependant d'après **Ralph Kimball**, elles peuvent engendrer un certain nombre de points négatifs:

- Difficulté de compréhension par des non informaticiens.
- Requêtes alourdies par un nombre grandissant de jointure.

Voilà un schéma représentatif :

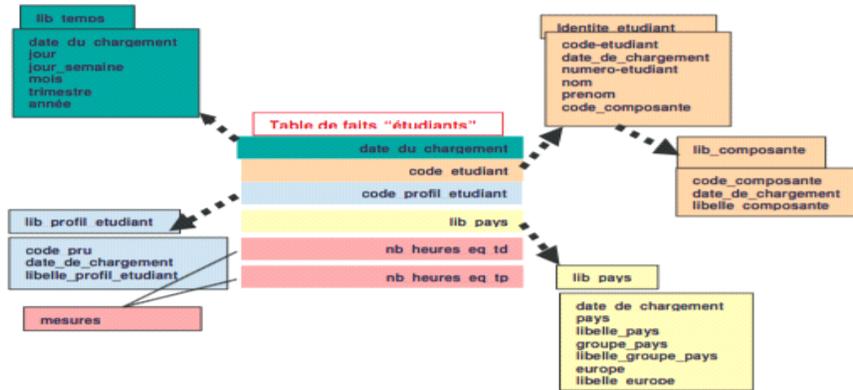


Figure I.11: modèle en flocon de neige. [27]

3.3 La modélisation en constellation

Les modèles en étoile ou en flocon ne gèrent qu'une seule table des faits. Par contre, il est très fréquent pour décrire une activité d'entreprise que nous ayons plusieurs tables des faits, donc plusieurs étoiles. Ces différentes étoiles auront peut-être des dimensions communes. Si nous relient ces dimensions ensemble nous obtenons une constellation, voilà un schéma en constellation :

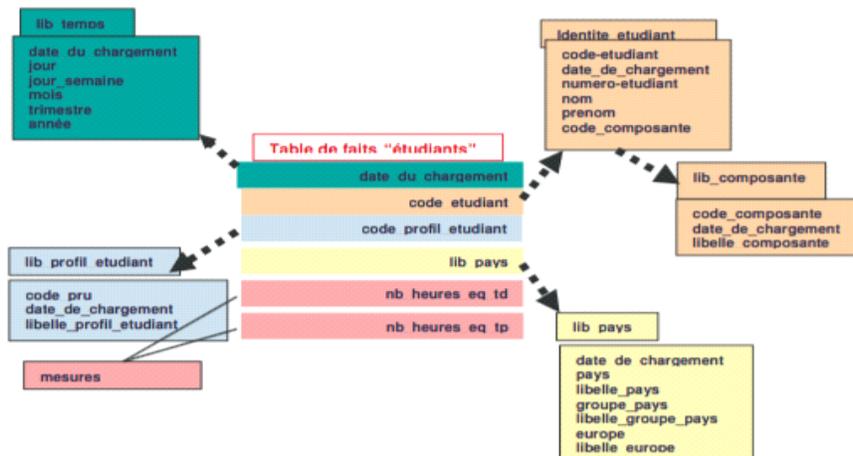


Figure I.12 : modèle en constellation . [27]

4. Concept OLAP [24]

La finalité d'un entrepôt de données est de stocker et historiser des volumes importants de données. Et pour ce faire, il existe plusieurs phases pour arriver à concevoir cet entrepôt de données, on présente « OLAP » :

4.1 Définition

On Line Analytical Processing est un système d'accès aux données en lecture uniquement, les programmes accédant aux informations travaillent sur de très grandes quantités de données, ce qui permet de réaliser des analyses complexes.

4.2 Les règles de OLAP

Le système OLAP regroupe l'information provenant de diverses sources. Il les regroupe, les intègre, les stocke, tout ceci afin de donner une vue métier à l'utilisateur. Cette vue métier va l'aider à retrouver l'information rapidement.

Une notion importante est l'historisation des données au sein des bases OLAP. Ceci entraîne avec une architecture différente, une grandeur de base de données supérieure aux bases de données classiques, En 1993 E.F. Codd a défini dans « Providing OLAP to user-analyst » le concept OLAP, il a mis en évidence 12 règles que doivent respecter les bases pour être OLAP :

1. Multidimensional Conceptual View (Vue conceptuelle multidimensionnelle)

Permet d'avoir une vision multidimensionnelle des données. L'inverse se nomme les tables unidimensionnelles.

2. Transparency (Transparence)

L'utilisateur ne doit pas se rendre compte de la provenance des données si celles-ci proviennent de sources hétérogènes. Ces sources peuvent provenir des bases de données de production, de fichiers à plats,

3. Accessibility (Accessibilité)

L'utilisateur doit disposer d'un accès aux données provenant de sources multiples en faisant abstraction des conversions et extractions de celles-ci.

4. **Consistence Reporting Performance (Performance continue dans les rapports)**

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

5. **Client-Server Architecture (Architecture client-serveur)**

Il est essentiel que le produit soit client-serveur. Le serveur stocke les données et le client les restitue.

6. **Generic Dimensionality (Dimensionnement générique)**

Chaque dimension doit être équivalente par rapport à sa structure et à ses capacités opérationnelles pour ne pas fausser les analyses.

7. **Dynamic Sparse Matrix Handling (Gestion dynamique des matrices creuses)**

Certaines cellules de l'hypercube peuvent être vides. Elles doivent être stockées de manière à ne pas détériorer les temps d'accès.

8. **Multi-User support (Support multi-utilisateurs)**

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

9. **Unrestricted Cross-dimensional Operations (Opération non restrictive entre les dimensions)**

Les calculs doivent être possibles à travers toutes les dimensions qui sont régies par les règles de gestion. Toutes les tranches de cube doivent être visualisées.

10. **Intuitive Data Manipulation (Manipulation intuitive des données)**

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

11. Flexible Reporting (Flexibilité dans la création des rapports)

La création des rapports ou des graphiques se doit d'être simple et efficace pour les utilisateurs.

12. Unlimited Dimensions & Aggregation Levels (Nombre illimité de niveaux d'agrégation et de dimensions)

Dimensions et niveaux d'agrégation illimités, afin d'autoriser les analyses les plus pointues.

4.3 Architectures des serveurs OLAP [24]

Ces 12 règles ont pour but de normaliser une base de données décisionnelle. Cette base de données peut être un système de gestion de base de données relationnelle ou multidimensionnelle, respectivement SGBDR et SGBDM, dans les SGBDM, le stockage des données se base sur le principe des hyper cubes, un hyper cube est une matrice décisionnelle avec au minimum quatre dimensions d'analyse, nous parlons également du cube, qui est une matrice décisionnelle avec trois dimensions.

Le concept d'OLAP est décliné en plusieurs « sous concepts » qui orientent la structure physique des données ou les techniques de traitements.

- **M-OLAP:** (Multidimensional on line analytical processing) est la forme la plus classique. Elle utilise les tables multidimensionnelles pour sauvegarder les informations et réaliser les opérations. Les données sont stockées dans une base de données multidimensionnelle.

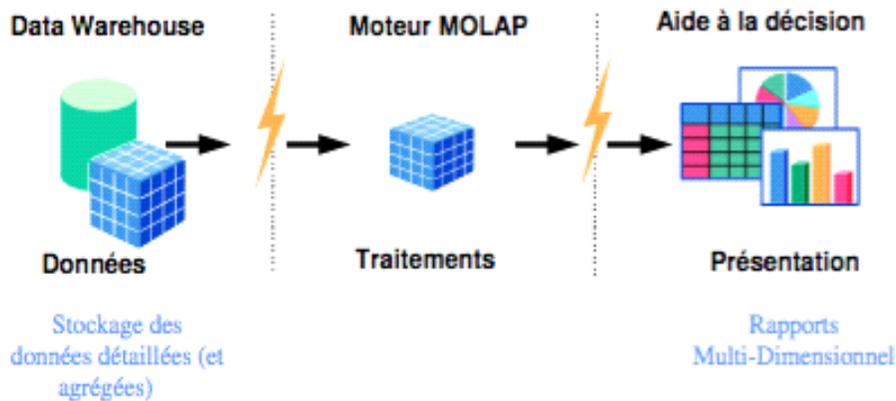


Figure I.13: Principe de l'architecture M-OLAP.[Nakache, 1998]

- **R-OLAP:** (Relationnal on line analytical processing) Il utilise une structure de base de données relationnelle, son avantage réside en la simplicité de mise en place puisqu'elle ne nécessite aucun investissement dans une base multidimensionnelle.

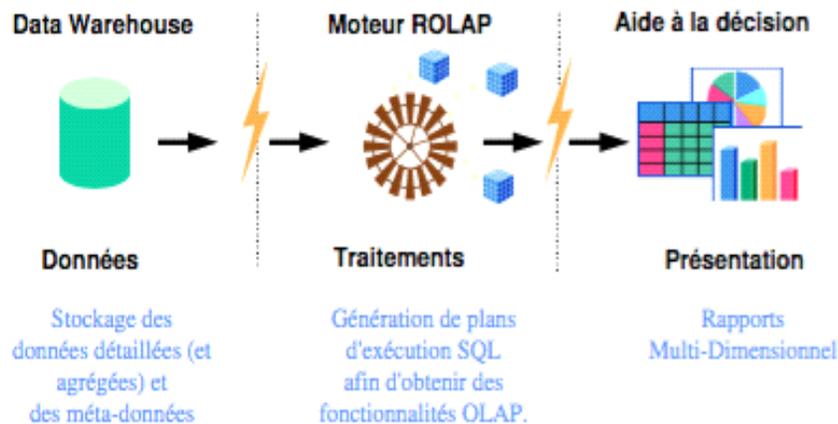


Figure I.14: Principe de l'architecture R-OLAP.[Nakache, 1998]

- **H-OLAP:** (Hybrid on line analytical processing)
Il utilise R-OLAP et M-OLAP en fonction des données qu'il traite. Sur les données agrégées il utilise M-OLAP, par contre sur les données plus détaillées, il utilise R-OLAP.

4.4 Définition OLTP[25]

On Line Transaction Processing est le modèle utilisé dans les bases de données de production. Il utilise un mode de travail transactionnel. Son rôle principal est l'interaction sur les données avec les actions suivantes : ajout, suppression et mise à jour. Il permet également l'interrogation des données avec des requêtes simples. OLTP permet l'accès à ces données et ces traitements à un grand nombre d'utilisateurs simultanés. Les transactions ainsi générées travaillent sur de petits ensembles de données.

5. Comparaison entre « OLAP » et « OLTP »

Critère	Online Transaction Processing (OLTP)	Online Analytical Processing (OLAP)
But	Contrôler et exécuter les tâches quotidiennes et fondamentales de l'entreprise	Assister dans la planification, la résolution de problème et la prise de décision
Types de données	Données opérationnelles (transactions)	Données historiques consolidées
Sources de données	BD transactionnelles	Entrepôts de données ou magasins de données
Ce que montre les données	Portrait instantané des processus d'affaires de l'entreprise	Vue multidimensionnelle de plusieurs activités d'affaires de l'entreprise
Insertions et mises-à-jour	Courtes requêtes d'insertion et de mise-à-jour lancées par les usagers finaux	Longs traitements en lot servant à rafraichir les données
Requêtes	Simple requêtes retournant quelques enregistrements (lignes) de la BD	Requêtes complexes impliquant souvent plusieurs tables et faisant l'agrégation de valeurs
Temps de réponses	Instantané	Quelques secondes à 1 minute max.

Tableau I.3 : Tableau comparatif entre « OLTP » et « OLAP »[Kimball,1996]

6. La navigation dans les données [24]

La vocation de l'OLAP est de réaliser une analyse interactive et multidimensionnelle des données de l'entrepôt de données. On parle de navigation dans les données et d'analyse exploratoire, Cette analyse en ligne agrège les données pour pouvoir les explorer et les visualiser, L'OLAP dispose d'un ensemble d'opérateurs de navigation qui vont lui permettre de visualiser les informations contenues dans le cube, de sélectionner un sous cube, de modifier l'ensemble des dimensions à prendre en compte ou de changer leur granularité, Comme mentionné précédemment, ces opérateurs de navigation sont généralement décomposés en deux catégories : opérateurs de de sélection (Slice, Dice) et d'agrégation (Roll-up, Drill-down).

Les opérations les plus utilisées sont :

6.1 Opérateurs de sélection : « Slice » et « Dice »

L'opérateur Slice (littéralement, trancher), permet de sélectionner un sous ensemble du cube, selon une ou plusieurs valeurs d'une dimension particulière. Slice est un opérateur de sélection portant sur les valeurs d'une seule dimension.

L'application de Dice équivaut à l'application de l'opérateur Slice à plusieurs dimensions.

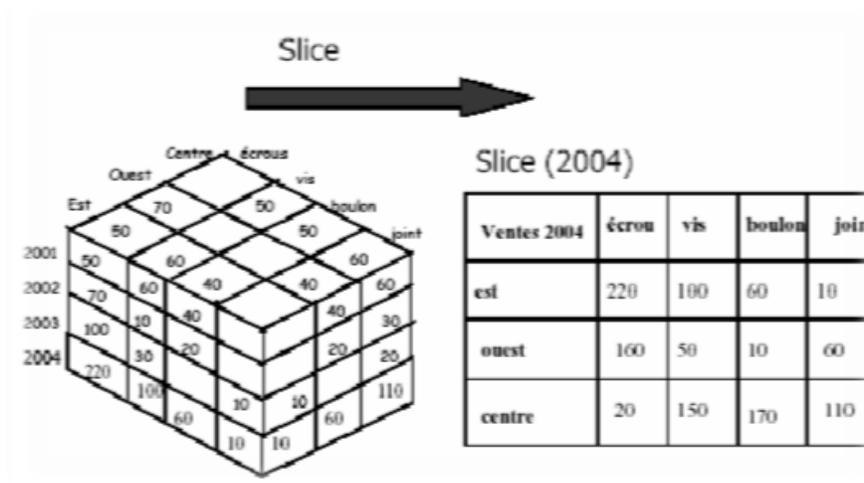


Figure I.15: Exemple de « slice » [Kimball,1996]

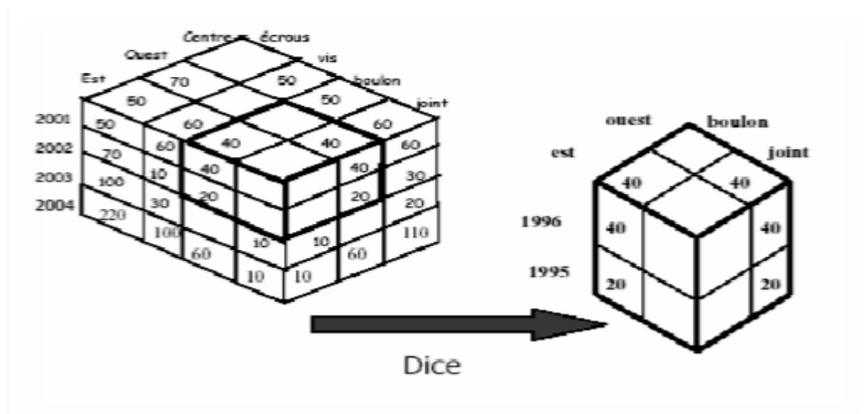


Figure I.16 : Exemple de « dice » [Kimball, 1996]

6.2 Opérateurs d'agrégation « Roll-up » et « Drill-down » [26]

Ce sont des opérateurs servant à naviguer entre différentes granularités d'une ou de plusieurs dimensions. Le Drill-down permet d'afficher les données avec une granularité plus fine.

Le Roll-up est l'inverse de Drill-down et consiste à remonter d'un niveau dans une hiérarchie de dimension vers un niveau plus agrégé.

Dans le deux cas, le nombre de dimensions du cube reste inchangé, en d'autres termes, l'application d'un Roll-up/Drill-down sur un cube à n dimensions résulte en un cube à « n » dimensions.

Les technologies OLAP, par leur aspect dynamique et synthétique complètent les outils de reporting.

Les outils de reporting sont généralement utilisés afin de fournir des vues statiques au travers de rapports instantanés à partir des données de l'entrepôt.

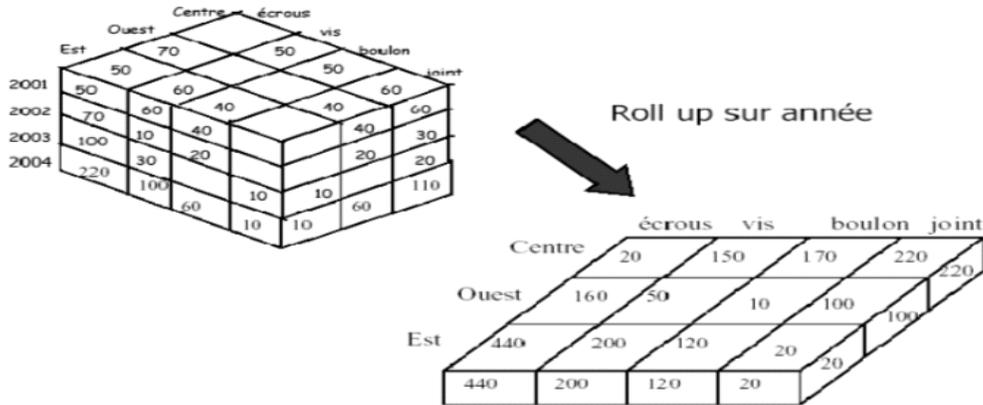


Figure I.17 : Exemple de « Roll-up » moins de détails sur les années. [Kimball,1996]

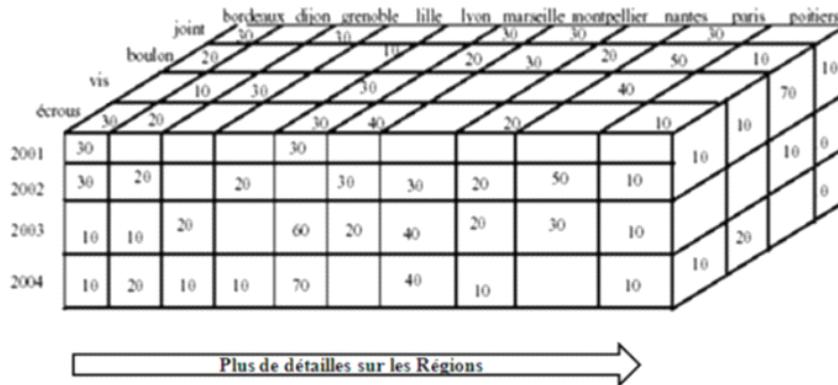


Figure I.18: Exemple de « Roll-up » plus de détails sur les régions. [Kimball,1996]

Conclusion

La modélisation dimensionnelle «OLAP», désigne une classe de technologie conçue pour l'accès aux données et le pouvoir d'analyse instantané.

La modélisation dimensionnelle est l'aspect le plus intéressant qui permet la navigation dans les données de l'entrepôt à la demande.

Introduction

La base cible est une base standard ou un entrepôt de données, les bases sources sont toujours utilisées par des applications. Les outils les plus communément utilisés dans ce domaine sont appelés des ETL (Extract, Transform, Load ou Extraction, Transformation, Chargement). Ce type d'outil est destiné à extraire les données de diverses sources (bases de données, fichiers, etc.), à les transformer et à les charger dans une base cible, permettant ainsi d'effectuer des synchronisations massives d'information d'une base de données vers une autre.

1. Définition

L'ETL,(Extract, Transform, Load ou Extraction, Transformation, Chargement) ou l'alimentation du Data Warehouse , est une étape des plus importantes dans un projet Data Warehouse , elle représente 80% de la charge de travail . Cette étape a pour objectif d'assurer l'acheminement des données des systèmes sources jusqu'à l'entrepôt de données , en passant par les différentes phases de nettoyage et de transformations nécessaires. [14]

2. Historique

Avant les années 90 chaque processus ETL est codé manuellement. Le développeur doit faire appel à plusieurs langages et différentes technologies pour mettre en place ces processus. Ceux-ci sont flexibles car le développeur code le processus de A à Z. Les limitations rencontrées sont celles liées aux technologies utilisées. Les inconvénients sont multiples : réalisation coûteuse et peu réutilisable, maintenance complexe, compatibilité limitée, etc. Au milieu des années 90, la première génération d'outils ETL voit le jour. Beaucoup de ces outils génèrent des codes pour extraire des données stockées dans des mainframe, la génération automatique du code permet d'éviter la lourde tâche de codage manuel des processus ETL. Les ETL sont performants avec les systèmes patrimoniaux mais nécessitent des connaissances approfondies en C ou en COBOL et ne sont pas adaptés pour les bases de données relationnelles. A la fin des années 90 apparaît la seconde génération d'outils ETL. Ceux-ci ne génèrent plus de code mais s'appuient sur des moteurs internes capables de gérer les processus ETL définis dans l'application.

Les ETL se regroupent en trois familles, en fonction de leur mode de fonctionnement et plus particulièrement en fonction des traitements effectués :

1 . Engine-based

Les transformations sont effectuées par le moteur de l'ETL en fonction d'un référentiel. Il offre l'avantage de pouvoir effectuer des transformations multi-base.

2. Database-embedded

Les transformations sont réalisées par la base de données sources. Il offre l'avantage d'avoir un accès complet au traitement.

3.Code-generators

Un code est généré en fonction des transformations à apporter, Il offre l'avantage d'être complètement indépendant de la source de données.

3. Les processus d'ETL

3.1 Le processus d'extraction [15]

Il y a plusieurs façons d'extraire et de charger les données pour un ETL. Les critères tels que l'architecture physique, la taille des données ou la disponibilité des serveurs vont permettre de choisir le mode de chargement.

Le chargement des données peut se faire de deux façons distinctes au sein d'un ETL : complète ou incrémentale.

1. Extraction complète

Le chargement complet consiste à vider la table de destination avant de réintégrer les données de la table source. Elle est intéressante pour les structures de données simples et de taille modeste.

Avantages

- Simplicité de mise en œuvre.
- Aucune différence de traitement entre les anciennes et les nouvelles données.

Inconvénients

-Besoin de beaucoup de ressources surtout si la source de données est importante ou que les traitements sont lourds,

- Gestion de l'historique impossible, car nous perdons la trace de l'existant.

2. *Extraction incrémentale*

La mise à jour incrémentale consiste à comparer la précédente remontée d'informations vis-à-vis de la nouvelle remontée, toutes les modifications seront intégrées, il existe plusieurs solutions afin de trouver les différences entre les remontées :

-Comparaison des remontées.

-Marquer les modifications au niveau de la source de données.

-Analyser les fichiers de log des moteurs de base de données.

L'extraction incrémentale apporte la possibilité d'historiser les données, par contre son implémentation s'avère plus compliquée que l'extraction complète.

Avantages

-Rapidité sur de gros ensembles.

-Historisation des données possibles.

Inconvénients

-Difficulté de mise en œuvre.

-Historique difficile à rechercher.

3.2 Le processus de transformation [16]

La transformation est une tâche complexe qui nécessite beaucoup de réflexion et de ressource. Après avoir extrait les données, il faut effectuer plusieurs traitements en vue de les :

-**Dénormaliser**: Dans un système décisionnel, la duplication délibérée de certaines données peut s'avérer utile, particulièrement en termes de performance.

-**Nettoyer** : Certaines données du système de production peuvent être fausses, par exemple une erreur de frappe sur le nom d'une ville. Les erreurs doivent pouvoir être détectées et corrigées afin de ne pas entraîner de répercussion sur les analyses.

-**Réconcilier** : Un vendeur est un vendeur, il ne doit y avoir qu'une seule dimension vendeur avec toutes les informations le concernant même si elles proviennent de source de données différentes.

-**Préparer** : Les données doivent respecter le format requis par les systèmes cible, par exemple, la modélisation en étoile ; ou encore être préparées de façon à pouvoir suivre ultérieurement, l'évolution d'un changement d'état dans le temps. Voici des exemples de transformations qui peuvent également avoir lieu :

L'affectation des clés de substitution qui consiste à remplacer chaque clé primaire par une clé numérique sans signification, améliorant les performances et garantissant l'indépendance du système décisionnel.

Autre exemple, la recherche d'équivalent textuel, qui consiste à faire correspondre des noms ou des adresses à certaines valeurs déjà enregistrées et qui peuvent parfois ne pas concorder à cause de différences orthographiques. Dans ce cas, il faut détecter ces différences et les corriger afin de faire correspondre les valeurs.

Enfin, on peut également citer les opérations de calcul, la gestion des tables de dimension et la gestion des tables de faits.

3.3 Le processus de chargement [17]

La dernière étape consiste à charger les données préalablement extraites, transformées et préparées dans des cibles hétérogènes. Il s'agit le plus souvent d'un entrepôt de données. Le chargement consiste à insérer ou mettre à jour les données cibles tout en conservant les données modifiées devant l'être, afin de conserver une traçabilité des informations (desquelles découlent les décisions prises). Les informations stockées dans un entrepôt de données ne doivent jamais disparaître dans la mesure du possible. Cette étape peut être complexe : il faut ajouter les nouvelles lignes, détecter si des lignes ont été modifiées ou supprimées ; il faut notifier ces modifications et veiller à ne pas charger des données en double.

On peut en déduire l'existence de deux types de chargement de données:

-Le chargement complet ou initial.

-Le chargement incrémental correspondant aux mises à jour périodiques.

-Le chargement initial est obligatoire pour le 1er peuplement puis les chargements incrémentaux peuvent ensuite être mis en place. Dans le cas où l'intégralité des données est extraite à chaque rafraichissement on parle de rechargement complet.

L'étape de chargement est relativement simple à exécuter dans les cas où les données ont été correctement préparées. Pour les fortes volumétries, il est recommandé d'utiliser l'utilitaire de chargement de masse (appelé aussi mode bulk) de la base de données, la plupart des S.G.B.D. proposent ce mode qui offre de bien meilleures performances.

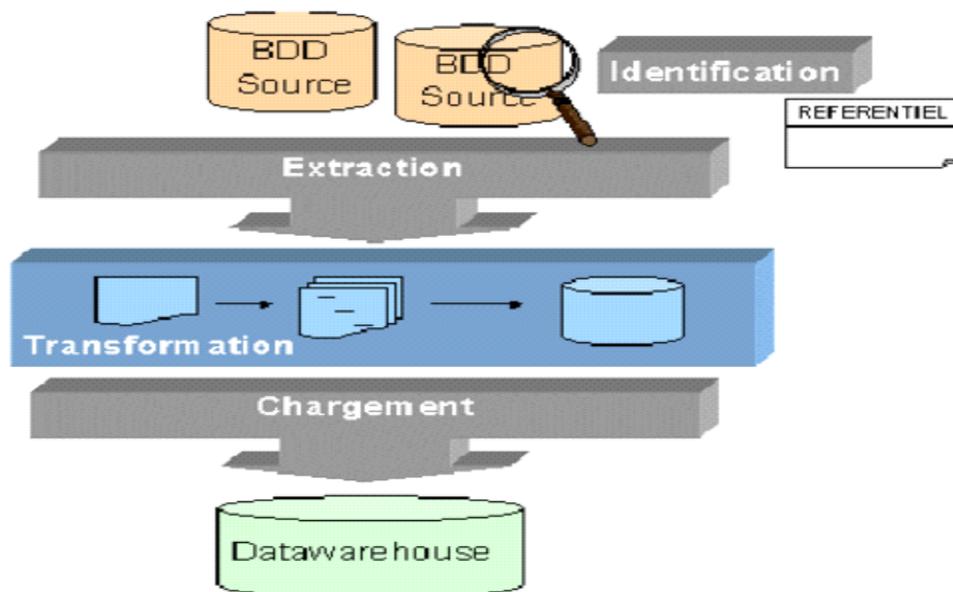


Figure I.19: représentation d'un ETL.[Inmon,2002]

4. Politiques de l'alimentation [18]

Le processus d'alimentation peut se faire de différentes manières, le choix de la politique de chargement dépend des sources : disponibilité et accessibilité, ces politiques sont :

Push: dans cette méthode, la politique de chargement est dans le système de production, il pousse les données vers la zone de préparation quand il en a l'occasion, l'inconvénient est que si le système est occupé, il ne poussera jamais les données.

Pull: contrairement de la méthode précédente, le pull tire les données de la source vers la zone de préparation, l'inconvénient de cette méthode est qu'elle peut surcharger le système s'il est en cours d'utilisation.

Push-pull : c'est la combinaison de deux méthodes, la source prépare les données à envoyer et indique à la zone de préparation qu'elle est prête, la zone de préparation va alors récupérer les données.

5. Objectif de qualité de données dans un processus ETL [19]

Le processus d'alimentation doit répondre à certaines exigences :

Sure: assure l'acheminement des données et leurs livraison.

Rapide: le processus d'alimentation doit palier au problème de lenteur et assure le chargement du Data Warehouse dans des délais acceptables.

Correctif: le processus d'alimentation doit apporter les correctifs nécessaires pour améliorer la qualité des données.

Transparent: le processus de l'ETL doit être transparent afin d'améliorer la qualité de données.

6. Démarches de construction d'un «Data Warehouse» [20]

Plusieurs théoriciens du domaine ont essayés de proposer des démarches pour la réalisation d'un projet «Data Warehouse» , ces démarches sont définies dans plusieurs étapes chevauchées et qui se complètent entre elles :

- **Etape 1** :modélisation et conception .
- **Etape 2** :alimentation .
- **Etape 3**: mise en œuvre .
- **Etape 4** : administration et maintenance.

La première étape : «**modélisation et conception**» est l'étape clé du travail , car les chercheurs ont proposés des démarches pour la réalisation d'un projet Data Warehouse , ces démarches se croisent essentiellement dans l'étape «1» qui décrit les approches existantes

Les approches les plus connues dans la conception des Data Warehouse sont :

- L'approche basée sur les besoins d'analyse.
- L'approche basée sur les sources de données.
- L'approche Mixte.

Aucune des approches citées n'est ni parfaite, ni applicable à tous les cas.

Elles doivent être étudiées pour choisir celle qui s'adapte le mieux à notre cas.

Quel que soit l'approche adoptée pour la conception d'un Data Warehouse , la définition de celui-là reste la même . En étant un support d'aide à la décision , le Data Warehouse se base sur une architecture dimensionnelle.

6.1 Approche «besoins d'analyse» :

Le contenu du Data Warehouse sera déterminé selon les besoins de l'utilisateur final . Cette approche est aussi appelée « approche descendante » (Top-Down Approach) et est illustrée par R. Kimball grâce à son cycle de vie dimensionnel comme suit :

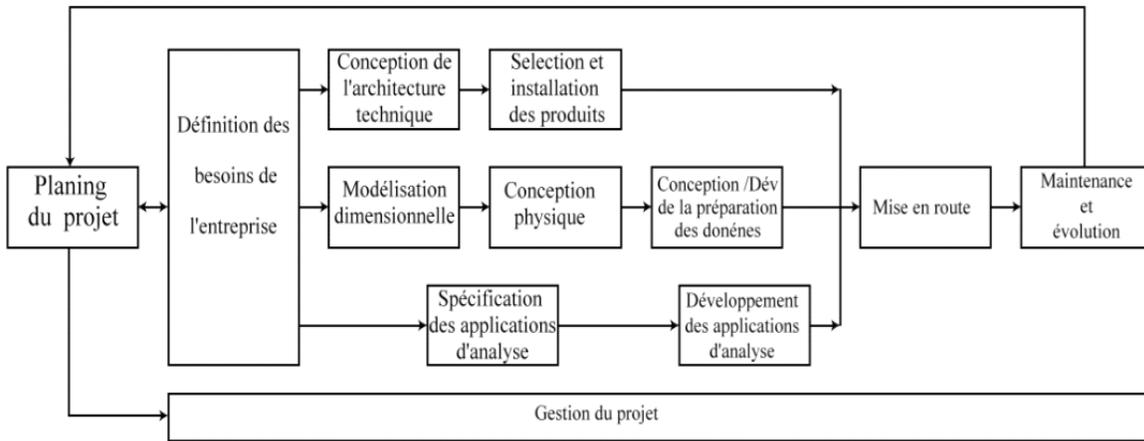


Figure I.20: illustration de l'approche besoin d'analyse[Kimball,2002]

6.2 Approche «sources de données» :

Le contenu du Data Warehouse est déterminé selon les sources de données . Cette approche est appelée : Approche ascendante (Bottom-up Approach).

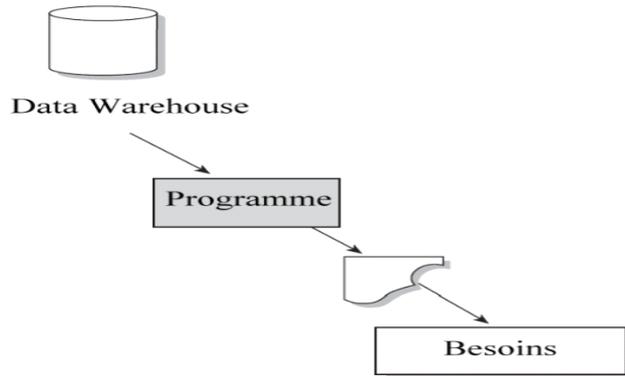


Figure I.21: illustration de l'approche sources de données [Inmon,2002]

Inmon considère que l'utilisateur ne peut jamais déterminer ses besoins dès le départ , « **Donnez-moi ce que je vous demande, et je vous dirai ce dont j'ai vraiment besoin** »[19]

Il considère que les besoins sont en constante évolution.

6.3 Approche mixte :

Une combinaison des deux approches appelée hybride ou mixte peut s'avérer efficace. Elle prend en considération les sources de données et les besoins des utilisateurs.

Cette approche consiste à construire des schémas dimensionnels à partir des structures des données du système opérationnel, et les valider par rapport aux besoins analytiques. Cette approche cumule les avantages et quelques inconvénients des deux approches déjà citées, telles que la complexité des sources de données et la difficulté quant à la détermination des besoins analytiques.

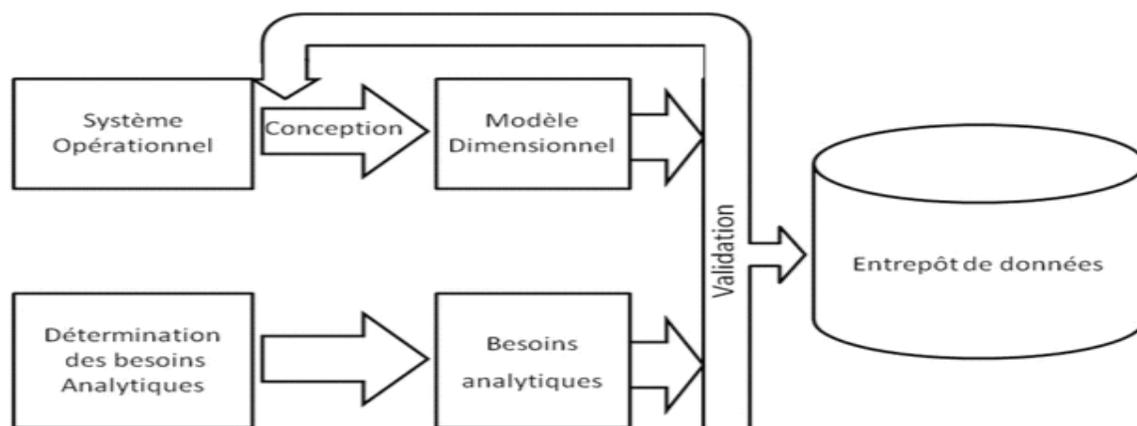


Figure I.22: illustration de l'approche mixte. [Kimball,2002]

Cette étape aboutit à l'établissement du modèle dimensionnel validé du Data Warehouse. Ce modèle dimensionnel sera transformé en modèle physique, qui différera du modèle dimensionnel

7. Mise en œuvre du Data Warehouse

L'exploitation du Data Warehouse se fait par le biais d'un ensemble d'outils analytiques développés autour du Data Warehouse, donc cette étape nécessite l'achèvement du développement, ou de la mise en place de ces outils qui peuvent accomplir les fonctions suivantes :

- **Requêtage ad-hoc:**

Le requêtage ad-hoc reste très fréquent dans ce type de projet, en effet, les analystes seront amenés à interagir avec le DW via des requêtes ad-hoc dans le but de faire les analyses requises par leurs métiers et d'élaborer aussi des rapports et des tableaux de bords spécifiques.

L'accès a ce genre de services peut se faire via différentes méthodes et outils, cependant, les spécialistes ont la matière préconisent de laisser la possibilité à l'utilisateur de choisir les outils qui lui paraissent les plus adéquats.

- **Reporting:**

Le terme reporting désigne une famille d'outils de business intelligence destiné à assurer la réalisation, la publication et la diffusion de rapports d'activité selon un format chiffrés ou d'un suivi d'avancement.

- **Analyse dimensionnelle des données:**

L'analyse dimensionnelle est sans doute celle qui exploite et fait ressortir au mieux la capacité de l'entrepôt de données, le but par l'analyse dimensionnelle est d'offrir aux utilisateurs la possibilité d'analyser les données selon différents critères.

Cette analyse se fait selon le principe OLAP, offrant de ce fait aux utilisateurs les possibilités de recourir à différentes opérations facilitant la navigation dans les données.

- **Tableaux de bord:**

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

Cette forme de restitution a la particularité de se limiter à l'essentiel, c'est-à-dire, la mise en évidence de l'état d'un indicateur par rapport à un objectif, tout en adoptant une représentation graphique de l'information.

- **Data Mining:**

Le Data Mining permet grâce à un certain nombre de techniques de découvrir ces connaissances en faisant apparaitre de corrélation en ces données.

Le Data Warehouse constituera alors la première source de données sur laquelle s'exécutera le processus de découverte de connaissance.

Le recours à ce genre de méthode est plus en plus utilisé dans les entreprises modernes, les applications et outils implémentant ces solutions sont rarement développés en interne, en effet les entreprises préfèrent se reposer sur des valeurs sûres du marché afin d'exploiter au plus vite les données en leurs possession.

Conclusion

L'alimentation des données est l'étape à laquelle il faut accorder le plus de temps et d'attention. En effet, elle est le garant de contenance de l'entrepôt en données fiables et correctes. Une fois l'alimentation terminée, l'exploitation des données peut alors se faire par différentes méthodes.

Introduction

Dans ce chapitre on verra en détail les spécificités du système à étudier en commençant par la représentation générale de l'organisme d'accueil.

Une bonne étude de l'organisme d'accueil est très importante dans tout projet décisionnel.

1. Présentation de l'organisme d'accueil

L'entreprise Nationale de grand travaux pétroliers «ENGTP» est une société par actions créée depuis le 19 février 1989, ses actions étaient détenues d'une part par la société mère sonatrach à 51 % et d'autre part par la société de gestion de participation à 49 % jusqu'au 13 décembre 2005 où toutes ses actions sont détenues à 100 % par sonatrach . Elle est dotée d'un conseil d'administration de 7 membres et d'une équipe de cadres dirigeants menée par un président directeur général.

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

Son siège social est fixé à : BP 38, zone industrielle de réghaia, Alger.elle est implantée sur le territoire national d'est en ouest et du nord au sud, notamment à skikda,Arzew, Hassi R'mel , Hassi Messouad et In amenas où elle est représentée par des directions régionales. Afin de bien comprendre le problème de notre étude, nous allons présenter l'entreprise pour mieux la connaître.

2. Historique de GTP

L'entreprise nationale de grands travaux pétroliers (GTP), a été créée par décret N° 80/103 du 04 Avril 1980 en tant qu'entreprise publique sous l'égide du ministère de l'énergie

- Elle est passée au statut d'EPE en date du 19 février 1989
- Elle est passée filiale à 100% SONATRACH en date du 17 décembre 2005

Son capital social est de 6.390.000.000 de dinars .

De par son historique, GTP a bénéficié de la capitalisation du savoir faire de

- **UIE:** (Union nationale français des industries et Entreprise de l'eau et l'environnement)
- **ALTRA:**(entreprise Algérienne des Travaux) société d'économie mixte dans laquelle SONATRACH détenait 51% des actions.
- **SONATRACH:** la compagnie algérienne de recherche, d'exploitation, de transport par canalisation, de transformation et de commercialisation des hydrocarbures et de leur dérivé.

1968 : Sonatrach et le Groupe français UIE créent une société de canalisation et de montage de montage industriel dénommée ALTRA

1972 : ALTRA devient filiale de Sonatrach à 100 %

1980 : Création de GTP qui hérite du patrimoine d'ALTRA

1989 : Statut EPE/SPA pour GTP

2004 : Sonatrach détient 51 % du capital de GTP et Holding Public TRAVEN détient 49%

2005 : GTP devient filiale du Groupe Sonatrach Holding SPP à 100 %

3. Politique Qualité et Environnement :

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

4. Métiers de GTP

GTP est une entreprise spécialisée dans la réalisation des projets.

Ses principaux métiers sont:

- Construction et/ou montage de grands ensembles industriels
- Engineering et contrôle de réalisation
- Préfabrication de tuyauterie et de charpente métallique
- Maintenance industrielle
- Expertise et contrôle de soudage
- Canalisation et collectes

5. Secteurs d'activité de GTP:

Le secteur des hydrocarbures a longtemps constitué le champ d'action principal de l'activité de GTP. Elle intervient dans d'autres secteurs, tel que :

- Chimie et pétrochimie
- Energie électrique
- Secteur industriel
- Les matériaux de constructions (cimenteries)
- L'agroalimentaire (semouleries)

6. Politique commerciale

6.1. Produits de GTP

Les produits de GTP sont:

- **Construction:** génie civil, préfabrication et montage de tuyauterie, réservoirs et sphère de stockage, électricité, instrumentation, canalisation, réseaux de collectes et de dessertes.
- **Engineering:** études de faisabilité, petit engineering de base, engineering détaillé, contrôle de réalisation.
- **Procurement:** approvisionnement, achats, transport, dédouanement.
- **Maintenance industrielle:** Prestations de maintenance, arrêts programmés et interventions.
- **Soudage:** contrôle de soudage, expertise, formation

6.2. Concurrence

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

Des concurrents sérieux se positionnent sur les marchés du montage industriel à l'instar des entreprises issues de Sonelgaz : Etterkib pour le montage, Kanagaz pour la canalisation, Inerga pour le génie civil, du secteur public avec des sociétés étrangères) à l'exemple de Sarpi, Entral, Safir, Benco, etc.

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

6.3 Clients de GTP

Les principaux clients de GTP sont:

Clients nationaux

- SONATRACH, NAFTAL, SONELGAZ, NAFTEC, ENIP, ASMIDAL, ERCE

Clients étrangers avec Sonatrach

- SH-CEPSA, SH-ANADARKO (groupement Berkine), SH-OO (organisation OURHOUD), SH-REPSOL (groupement TFT), SONAHNESS

Clients étrangers

- JGC/KBR (Japon-USA), GE ENERGY (Italie), BECHTEL (canada), FMC (USA), DENYS (Belgique), SAIPEM (Italie), ABB (Italie), JGC(Japon), KELLOG (USA)

7. . Principales réalisations de GTP

Les principaux projets réalisés par GTP sont:

- Montage unité d'extraction Rhourd_Nouss (Hassi-Messaoud)
- Récupération assistée du pétrole Rhou-El-Baguel (Hassi-Messaoud)
- Station compression SC9/SC12 (Hassi-Messaoud)
- Centre de traitement gaz (Hassi R'Mel)
- Centre de traitement d'huile BRN (Hassi-Messaoud)
- Unité de traitement de gaz TFT (Hassi-Messaoud)
- Unité de traitement (5eme train) d'Alger (Hassi-Messaoud)
- Pipe GPL Nord & Sud TFT-Haoud el Hamra 5430km (Hassi-Messaoud)
- Boosting (Hassi R'Mel)
- Unité de traitement d'huile OURHOUD (Hassi-Messaoud)
- Unité de traitement gaz (In Saleh)
- Unité de traitement de gaz_ Tiguentourine (In Amenas)
- Centrales électriques : Arzew(|&|), jijel, Tilghmet, Adrar ; Illizi, Tan-Tan(Maroc), Skikda.
- Cimenteries : Meftah, Ain El Kebira.
- CSD (centre de stockage et distribution) : Remchi, El-Eulma, Skikda, Berrahal.
- PEHD pour polymed (Skikda)
- Centres enfûteurs (Sphères)

7.1 Ressources et politique de l'entreprise

7.1.1. Humaines

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

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

- Evolution des effectifs par statut

Unité: agent

Année	Permanent		Temporaire		Total effectif	
	Nbre	%	Nbre	%	Nbre	%
2002	4593	57.9	3343	42.1	7936	100.0
2003	4230	50.1	4209	49.9	8439	100.0
2004	3920	48.0	4243	52.0	8163	100.0
2005	3666	46.3	4248	53.7	7914	100.0
2006	3487	59.4	2382	40.6	5869	100.0
2007	3385	49.3	3511	50.7	6896	100.0

Tableau II.1 : Evolution des effectifs par statut

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

7.1.2 Matérielles

GTP dispose de:

- Un atelier de soudage.
- Des ateliers de Tuyauterie.
- Des Laboratoires de Contrôle Destructif et Non Destructif.
- D'un parc matériel roulant de 3000 équipements, dans les principaux domaines ci-après :
- Le Terrassement et Génie Civil.
- Le soudage et contrôle.
- Le Levage (Grue de 18 T à plus de 300T).

8. Situation informatique de l'entreprise

8.1. Aspect hardware

Désignation	Région						total par type
	Siège	DRZ	DRK	DRM	DRR	CSZ	
Micro-ordinateurs de Bureau	579	68	79	267	141	23	1157
SERVEURS	6	2	3	3	2	0	16
Imprimantes de Bureau	313	35	58	157	92	20	675
Imprimantes LIGNE	8	3	3	6	3	0	23
Onduleurs de bureau	118	32	57	143	73	10	433
Onduleurs de site	4	1	1	2	1	0	9
s/total	1028	141	201	578	312	53	
%	44,44	6,10	8,69	24,99	13,49	2,29	
total général	2313						

Tableau II.2: Répartition des équipements Informatiques par Direction régionale et par Type

8.1.2 Aspect software:

- **Les systèmes d'exploitation**

- Windows NT4.
- Windows 98.
- Windows 2000 professionnel.
- Windows 2003 server.
- Windows 2008 server.
- Windows XP professionnel.

- **Les logiciels :**

- SGBD oracle server 8.
 - SGBD oracle server 9i.
 - SGBD oracle 10g.
 - Logiciel de développement: developper 2000.
 - Logiciel de bureautique: Microsoft office XP, office 2000 (Word, Excel,...).
 - Logiciel technique de dessin : Autocad
- Logiciel antivirus KASPERSKY : installé en réseau.

- **Les applications :**

- Gestion de paie et personnel.
- Gestion de stock.
- Gestion comptable.

- **Le réseau informatique :**

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

Des réseaux LAN (Local Area Network) sont installés au niveau siège et dans toutes les régions de l'entreprise.

9. Caractéristiques:

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

Autres applications réseau utilisées :

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

10. Les données

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

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

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

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

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

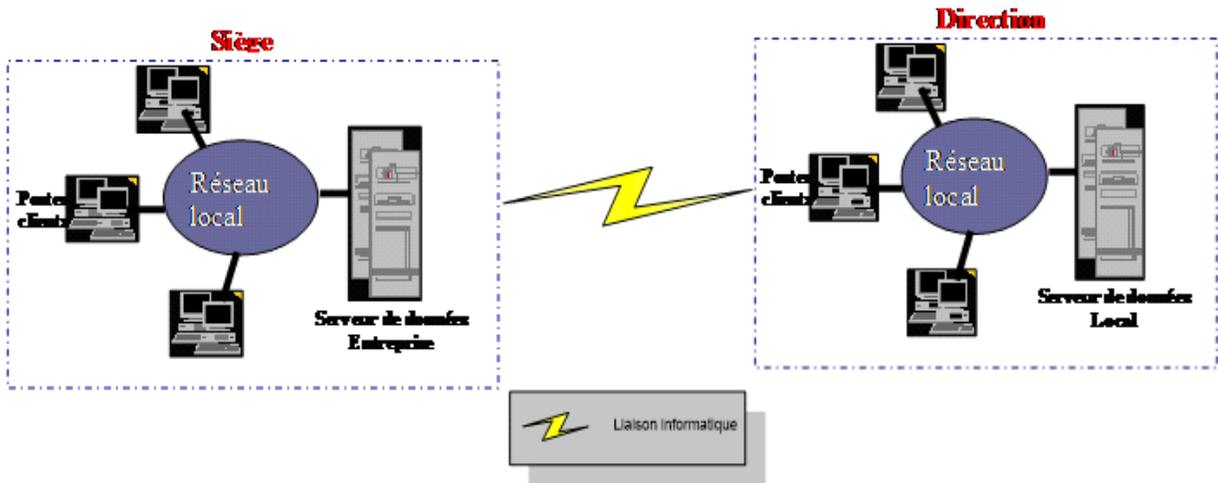


Figure II.2: Architecture Client / Serveur à GTP

Interconnexion via des liaisons louées

Le siège et les directions régionales sont interconnectés à travers des liaisons spécialisées louées auprès du fournisseur d'accès "Algérie Télécom". Ces liaisons sont disponibles par le biais de câbles concédés installés entre chaque direction et le centre amplificateur (CA) le plus proche.

11. . Les traitements

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

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

Après un transfert des données des régions vers le siège, une consolidation s'effectue à ce niveau. La mise à jour des bases de données permet de disposer de la situation des produits et des charges de l'ensemble des centres de frais de l'entreprise à travers la production d'états et de rapports synthétiques mis à la disposition des managers.

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

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

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

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

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

Sous-système Gestion Finances et Comptabilité

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

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

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

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

Pour fournir les résultats suivants:

En Comptabilité Générale

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

En Comptabilité Analytique

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

12. . ORGANIGRAMME DE L'ENGT

12.1 Organigramme Général

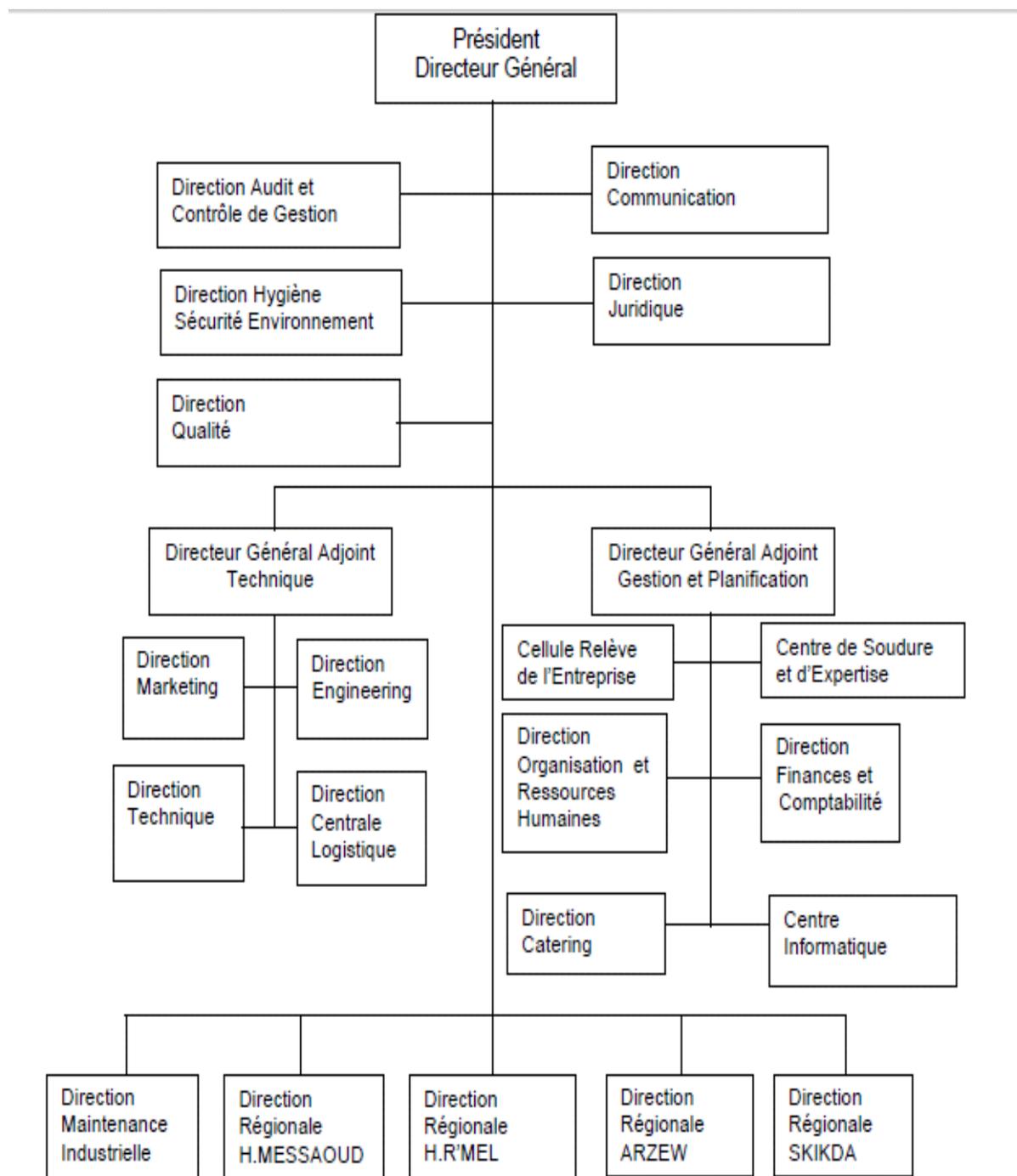


Figure II.3 : Organigramme général de L'ENGTP.

12.2. Organigramme de la direction technique

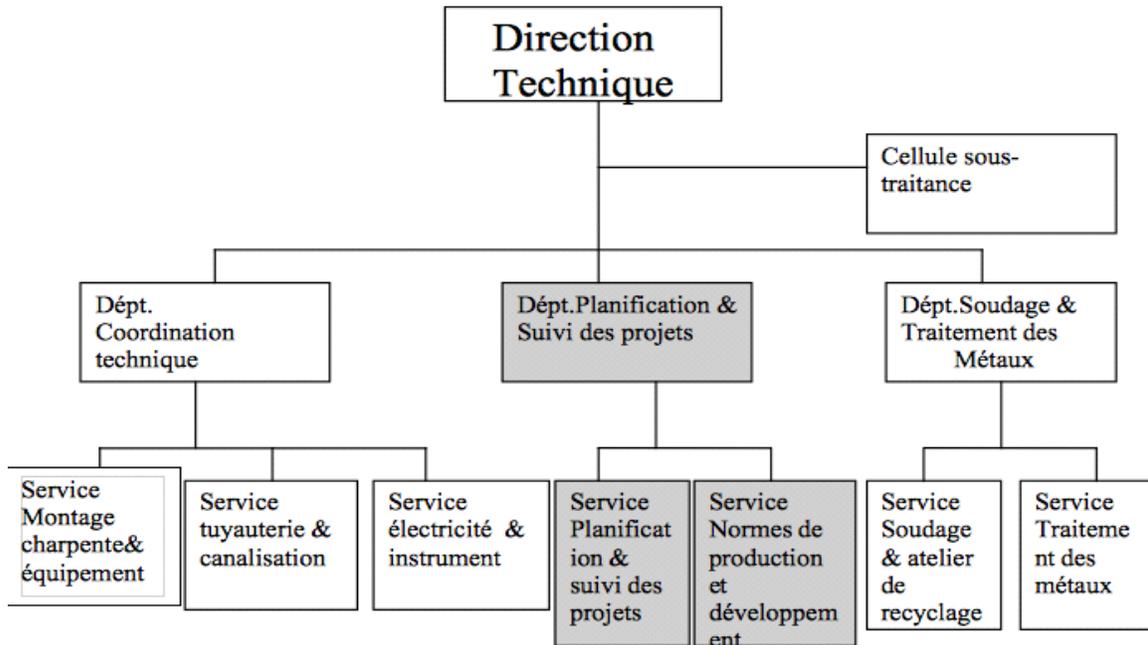


Figure II.4 : Organigramme de la direction

Conclusion

Dans cette étape, nous avons pu cerner l'organisme d'accueil, en présentant l'environnement dans lequel il évolue. Après avoir étudié globalement ses ressources, ses politiques, ses concurrents et sa place dans le marché. Notre entrepôt de données se limitera à la fonction commerciale que nous avons jugée comme fonction principale de GTP. Dans la prochaine étape, nous allons présenter les différents besoins des décideurs cette étape est le point important placé au centre de l'univers de l'entrepôt de données.

Introduction

L'entreprise accueillante possède des informations, qui se présentent dans des systèmes opérationnels dont le rôle est l'enregistrement des données et l'analyse.

A travers ce chapitre nous allons essayer, de faire une analyse complète de l'existant pour avoir une vision générale des procédures d'élaboration des rapports et de consolidation des données, ensuite on définit le point de départ du projet et des besoins en particulier en connaissant les lacunes existantes.

Ce chapitre a pour principale vocation d'exposer et de décrire une étape plus que nécessaire dans un projet Data Warehousing. C'est à partir de cette étude que se décidera la manière de construction de l'entrepôt de données et de son architecture.

1. Etat du décisionnel au sein du groupe « GTP »

Il est important de savoir que l'entreprise, dispose d'un système d'aide à la décision. Tous les processus d'analyse et de prise de décision au sein du groupe GTP, se basent essentiellement sur des rapports dont les données sont extraites à partir des systèmes transactionnels d'une manière manuelle.

Le volet ressource humaine, contient un ensemble d'utilisateurs qui sont chargés de tous les mouvements d'arrivé et de départ, ils extraient les informations depuis une application appelée «mouvement» qui contient tous les informations du personnels existant. Toute modification ou bien changement se fera au niveau d'une application local.

Dans cette optique, la synthèse générale se fait après l'enregistrement des données qui sont ensuite transformés en plusieurs «tableaux de bords», pour que les décideurs puissent avoir une vision claire de l'entreprise qui les aident à prendre des décisions qui semblent utiles.

Le diagramme suivant présente l'état du décisionnel au sein du groupe « GTP»

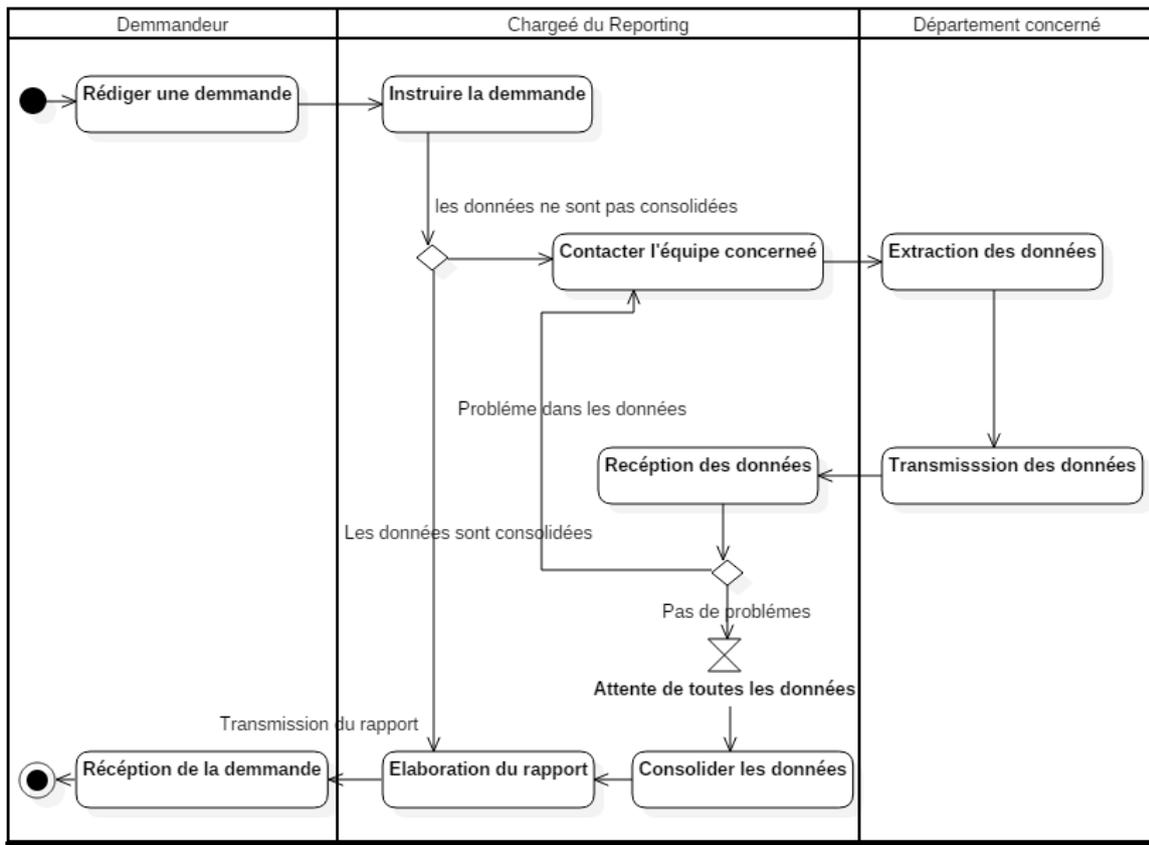


Figure II.5: Diagramme d'activité, processus de construction d'état.

Les principales difficultés auxquelles fait face le système existant sont :

- La difficulté de gérer et d'organiser les données pertinentes et stratégiques qui se trouve au niveau de la base de données opérationnelle.
- Les données jugées importantes se trouvent éparpillées dans les différentes applications des services de l'entreprise ce qui les rend inaccessible.
- Difficulté d'extraire l'historique des données sur une période de quelques années.
- Les accès simultanées des utilisateurs et les décideurs en particulier le rend très lourd et créer des conflits de privilèges d'accès. Pour cela il est difficile de faire une analyse quotidienne sous un système opérationnel.

L'étude des besoins est une étape plus que nécessaire dans un projet Data warehouse . C'est, en effet, à partir de cette étude que se décidera la manière de construction de l'entrepôt de données et de son architecture.

2. L'utilité de l'étape «définition des besoins»

Tout Data Warehouse doit être en mesure de répondre aux attentes des utilisateurs . Cela ne peut, évidemment, se faire sans une étude approfondie de leurs besoins.

Il existe deux approches qui décrivent la démarche : l'approche « Bottom Up » et l'approche « Top Down ». Ces approches permettent donc de recueillir, corriger et de modérer les attentes des utilisateurs dès le départ, tout en détectant d'éventuels besoins non exprimés.

Durant l'étude des besoins on ne peut se limiter aux interviews avec les utilisateurs , néanmoins, il faudrait absolument prendre en compte l'avis des Administrateurs des bases de données des systèmes sources, d'après **Kimball** « Les DBA sont les principaux experts sur les applications existantes susceptibles d'alimenter l'entrepôt de données . Leurs interviews servent à confronter aux réalités certains des thèmes qui surgissent lors des rencontres avec les utilisateurs finaux. »

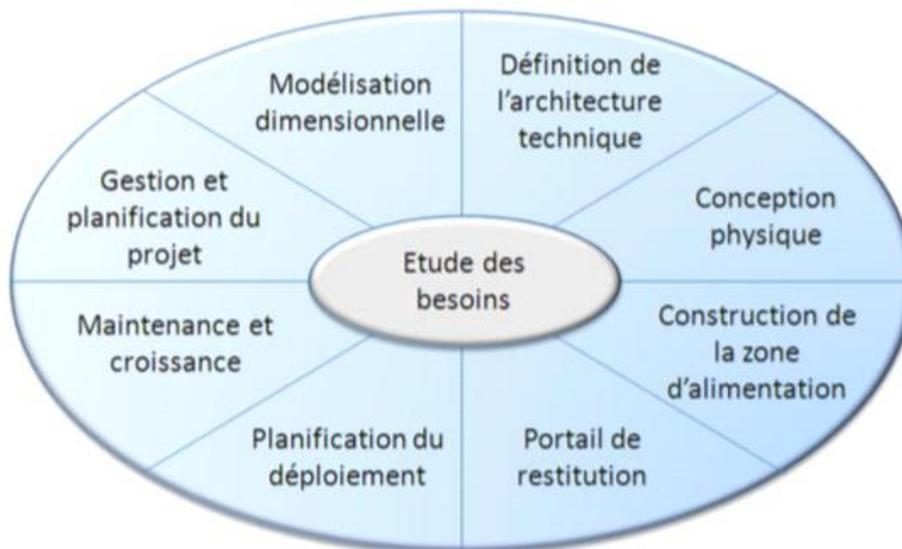


Figure II.6: La place de l'étape d'étude des besoins dans un projet Data Warehouse.[20]

3 . Description de la démarche d'étude des besoins

Afin de faire une étude aussi complète que possible , nous avons choisi d 'adopter une démarche qui nous a permis de combiner , d'une manière assez satisfaisante , entre l'approche « Bottom Up » et l'approche « Top Down ».

Ainsi, notre démarche peut se résumer à travers des étapes suivantes:

- Étude préliminaire des systèmes sources et interviews avec les utilisateurs : chef de section mouvement du personnels , chef de service des statistiques ressources humaine.
- Détection des postes susceptibles d'être porteurs d'informations utiles.
- Utilisation d'autres moyens pour la détection des besoins.
- Rédaction et validation du recueil récapitulatif des besoins.

a. Étude préliminaire des systèmes sources et interviews avec les utilisateurs

Cette étude représente une étape de compréhension des systèmes opérationnels et de leur environnement. Elle a pour mérite de consolider les connaissances acquises durant l'étude de l'existant, ainsi que le jargon et le fonctionnement de l'entreprise. En outre, cette étape permet de détecter, de manière succincte, les postes susceptibles d'interagir avec le Data Warehouse. Elle est de ce fait indispensable pour un bon recensement des besoins.

Les utilisateurs (ingénieurs de statistiques, chef de département ressources humaines les ingénieurs en informatique), qui sont pour la plupart des chefs et gestionnaires , dont la fonction principale est de définir les règles de gestion des applications et de s'assurer du respect des procédures propres à la distribution , ont été une source d'information assez bénéfique, eu égard à leur connaissance des applications .

b. Détection des postes susceptibles d'être porteurs d'informations utiles

Vu le grand nombre d'employés activant au sein du groupe GTP , ainsi que la dispersion géographique des filiales , il nous a été , bien sûr , impossible de voir et d'interviewer toute cette population. Notre étude s'est axée sur les utilisateurs des applications .

Cette étape nous a permis , donc, d'identifier, en premier lieu , les services qui peuvent être porteurs d'informations .

c. Planification, préparation et conduite des interviews

Avant de détailler cette étape, il est nécessaire de justifier notre choix de l'entretien « interviews » comme méthode de collecte des besoins. En effet, cette méthode a l'avantage d'être, plus ou moins, facilement planifié et d'ouvrir le dialogue avec les interviewés, qui sont pour la plus part des analystes.

Dans le souci de conduire à bien cette étape, qui du reste est très critique et délicate, il est nécessaire de passer par différentes phases, à savoir :

1) La phase de planification

La planification se fait généralement plusieurs jours à l'avance. Elle nous permet de prendre les rendez-vous nécessaires et de prévenir les futurs interviewés de notre arrivée et du motif de cet entretien.

Cette phase, qui est un préalable indispensable, s'est avérée être une tâche très ardue. En effet, la nature organisationnelle et la dispersion des structures liées à la distribution ont posé des problèmes que nous évoquerons plus loin. Cependant ces mêmes facteurs nous ont poussés à entreprendre ce genre de démarches dans un souci de bonne conduite du projet et afin de ne pas perdre de temps dans des allers et retours improductifs.

Aussi, nous avons essayé de planifier nos entretiens de manière à avoir une certaine alternance entre les interviews des potentiels utilisateurs et les entretiens avec les DBA et les gestionnaires de manière à combiner entre les besoins d'analyse et les potentialités des systèmes sources et de leurs données.

2) La phase de préparation

Une fois la planification de l'entretien terminée, sa préparation doit se faire de telle sorte à ce qu'il se déroule dans les meilleures conditions possibles.

La préparation se résume essentiellement en l'identification des questions à poser, des points à soulever et des thèmes à éviter afin de ne pas trop s'éparpiller. Les questions à poser sont classées en deux catégories, selon le poste de la personne interviewée. Ainsi certaines questions sont destinées aux décideurs alors que d'autres sont destinées aux analystes.

3) La conduite des entretiens

Dernière phase de l'étape, la conduite des interviews représente la réalisation sur le terrain des phases précédentes . Le but escompté étant d'amener les interviewés , au travers de leurs réponse à nos questions, de présenter leur travail et la manière dont ils procèdent pour le faire. L'identification des besoins se fera alors en détectant les métriques qu'ils utilisent et les informations sur lesquelles ils s'appuient pour la prise de décision.

d. Autres moyens utilisés pour la détection des besoins

Bien que les entretiens représentent une source importante d'informations et aident grandement à l'identification des besoins des utilisateurs , leur utilisation exclusive n'est pas conseillée dans la construction d'un entrepôt de données . Cela tient principalement au fait que les utilisateurs ne peuvent , même avec la meilleure volonté du monde , exprimer tous leurs besoins.

De ce fait , il est fait appel à l'étude des rapports déjà demandés et des données disponibles a même de fournir des informations exploitables . L'étude des rapports offre un certain apport à notre démarche d'études des besoins , dans la mesure où les utilisateurs peuvent ne pas mentionner des besoins qui leur paraissent évidents ou qui ne leur viennent pas à l'esprit durant nos interviews, ces derniers peuvent être, en effet, influencés par nos questions.

L'étude des données , quant à elle , sert à détecter des besoins non déclarés et qui peuvent se faire sentir ultérieurement, le but de cette démarche étant de construire un entrepôt de données capable de répondre à ces éventuels nouveaux besoins.

e. Rédaction et validation du recueil récapitulatif des besoins

L'étude des besoins nous a permis de recenser les besoins que nous avons classés par volets spécifiques , le tableau suivant synthétise les besoins recensés :

Volet détecté	Les critères d'analyse
Volet personnel	<p><u>Utilisateur</u> : ce volet a été évoqué et sollicité presque à tous les niveaux et par différents postes. Cependant, les utilisateurs des services recrutement et ressource humaine, ont exprimé un vif intérêt pour ce volet.</p> <p><u>Besoins</u> : les utilisateurs ont besoin de connaître l'évolution de l'effectif de l'entreprise dans le temps et selon différents critères à savoir : (par direction, catégorie socioprofessionnelle, sexe, tranche d'âge ...etc.).</p>
Volet recrutement	<p><u>Utilisateur</u> : Services recrutement</p> <p><u>Besoins</u> : les utilisateurs ont besoin de connaître les recrutements de l'entreprise dans le temps : (nombre de contrats signés, nombre de postes, nombre de candidats retenus...etc.).</p>
Volet Formation	<p><u>Utilisateur</u> : Services formations.</p> <p><u>Besoins</u> : les utilisateurs ont besoin de connaître : (effectif formé, budget prévu, coût par action ...etc.).</p>
Volet Mouvement	<p><u>Utilisateur</u> : Services recrutement, ressource humaine.</p> <p><u>Besoins</u> : ce volet contient les informations relatives au mouvement interne et externe de l'entreprise (Fin relation travail « FRT » ; démission) l'utilisateur devrait être en mesure de suivre l'évolution de ces statistiques selon différents critères d'analyse : (par nombre de droit au congé, effectifs en FRT, nombre de démissionnaire).</p>
Volet Œuvre Social	<p><u>Utilisateur</u> : Services œuvre sociale.</p> <p><u>Besoins</u> : les utilisateurs ont besoin de connaître : (prime, prêts, veuve ...etc.).</p>

Tableau II.3: synthèse des besoins recensés.

4. Problèmes et obstacles rencontrés

Bien que notre étude des besoins ait donné lieu aux résultats escomptés , à savoir leur identification et leur prise en charge , il n'empêche que le travail ne s'est pas effectué sans obstacles, dont les plus importants sont :

Difficulté de planifier les entretiens à cause de :

- L'emploi du temps chargé des interviewés.
- Les interviewés ne possèdent pas le même jargon.

5. L'utilité des entretiens

Les différents entretiens ont porté sur :

- La connaissance de l'activité de l'entreprise, afin d'identifier les différents processus métiers qui la compose.
- L'identification des indicateurs de l'analyse.
- Cerner les besoins et les attentes des utilisateurs.
- Déterminer les dimensions et les faits de mesure afin d'établir la modélisation dimensionnelle.

Conclusion

L'étude de l'existant et définition des besoins sont deux étapes plus que nécessaire d'un projet «data warhouse».

L'étude de l'existant permet de définir le point de départ du périmètre du projet, ainsi que les insuffisances du système actuel en soulignant les points faibles.

L'étude des besoins permet de connaître d'avantage des détails et d'identifier les besoins analytiques de l'entreprise.

Cette étape s'est déroulée sur 4 semaines d'entretiens et à concerné quatre utilisateurs occupant différents postes. L'ensemble des entretiens a duré des heures et nous a permis tout d'abord d'appliquer les méthodes d'analyse et de collecte d'information..

Les besoins étant recensés, la construction du «Data Warehouse» peut alors commencer. Cette construction fera l'objet du chapitre suivant.

Introduction

Une fois les besoins des utilisateurs sont connus, et les volets sont bien définis, nous avons eu recours à la modélisation dimensionnelle qui est souvent associée aux entrepôts de donnée compte tenu de ses avantages.

Avant de se lancer dans la modélisation, il est intéressant de classer les sujets recensés selon l'intérêt réalisation.

1. Processus de la modélisation dimensionnelle :

La conception d'un modèle dimensionnel passe par quatre étapes essentielles :

- Définition de l'activité.
- Définition des dimensions qui décrivent une ligne de la table de fait.
- Définir les indicateurs du fait.
- Construire le modèle de l'activité.

2. Métier étudié GRH « Gestion des ressources humaines » :

La notion d'administration du personnel, au sein de l'entreprise, a évolué à travers le temps pour devenir la gestion des ressources humaines (GRH).

Ce dernier concept est beaucoup plus large puisqu'il dépasse la simple administration pour englober le management du capital humain de l'entreprise, avec toutes les nouvelles fonctions qui en découlent.

A ce titre, La gestion des ressources humaines est la fonction organisationnelle qui s'occupe du recrutement, de la gestion, du perfectionnement et de la motivation du personnel, y compris de fournir du soutien et des systèmes fonctionnels et spécialisés pour favoriser la participation des employés ainsi que des systèmes de gestion pour favoriser le respect réglementaire des normes liées à l'emploi et aux droits de la personne. La GRH se rapporte à la réglementation, à la paie, aux avantages sociaux et aux autres aspects transactionnels qui font partie du mandat des ressources humaines.

L'ENGTP, comme toute entreprise comporte un nombre important de travailleurs de statut de rôle différents, avec une courbe d'ancienneté et d'expérience variante d'année à une autre et

plusieurs éléments dont il faut en tenir compte, ainsi la disponibilité de ces informations s'avère essentielles pour les décideurs de l'entreprise .

Cette partie conçoit les volets de notre « Data Warehouse » :

- Volet « **personnel** ».
- Volet « **recrutement** ».
- Volet « **formation** ».
- Volet « **mouvement** ».
- Volet « **œuvre social** ».

1) Volet « **personnel** » :

a. Définition:

« Le personnel est l'ensemble des individus qui apportent leur travail dans le cadre d'un contrat de travail en contrepartie d'une rémunération versée par l'employeur. Traditionnellement, le personnel a une fonction productive selon sa fonction, il participe à la création de richesse dans l'entreprise. Parmi le personnel, on trouve différents statuts : cadres, employés, ouvriers. » .

b. Les dimensions participantes du modèle :

Les dimensions ont pour objectif de décrire le fait, donc on essaye de recenser toutes les informations qui décrivent un effectif et qui peuvent intéresser les décideurs.

1. **Dimension période (PERIODE_DIM)** : la dimension période est la seule dimension qui figure systématiquement dans tout entrepôt de données.

PERIODE_DIM
PERIODE
EXERCICE

Figure II.7 : dimension période .

Chapitre 3 : La zone d'entreposage

Dans cette dimension, il est utilisé une clé PERIODE comme une clé primaire, le tableau suivant donne plus de détails sur cette dimension :

Désignation	Détails
Periode	La clé Primaire de la dimension Periode_dim
Exercice	L'exercice de la dimension Periode_dim

Tableau II.4 : table descriptive de la dimension « période ».

2. Dimension fonction (FONCTION_DIM) : est la dimension qui décrit les fonctions du personnel.

FONCTION_DIM
CD_FONCTION
LIB_FONCTION

Figure II.8 : dimension fonction .

Dans cette dimension, il est utilisé une clé CD_FONCTION comme une clé primaire, le tableau suivant donne plus de détails sur cette dimension :

Désignation	Détails
CD_FONCTION	La clé Primaire de la dimension fonction_dim
LIB_FONCTION	Libelle de la dimension fonction_dim

Tableau II.5 : table descriptive de la dimension «fonction ».

3. Dimension chantier (CHANTIER_DIM) : est la dimension qui décrit le chantier.

CHANTIER_DIM
N_CHANTIER
CD_DIRECTION

Figure II.9 : dimension chantier.

Dans cette dimension, il est utilisé une clé N_CHANTIER comme une clé primaire, le tableau suivant donne plus de détails sur cette dimension :

Désignation	Détails
N_CHANTIER	La clé Primaire de la dimension chantier_dim
CD_DIRECTION	La clé Primaire de la dimension direction_dim

Tableau II.6 : table descriptive de la dimension « chantier ».

4. Dimension groupe professionnel (GRPE_PROF_DIM) : est la dimension qui décrit le groupe professionnel.

GRPE_PROF_DIM
CD_GRPE
LIB_GRPE

Figure II.10 : dimension groupe professionnel.

Dans cette dimension, il est utilisé une clé CD_GRPE comme une clé primaire, le tableau suivant donne plus de détails sur cette dimension :

Désignation	Détails
CD_GRPE	La clé Primaire de la dimension GRPE_PROF_DIM
LIB_GRPE	Libelle du groupe professionnel de la dimension GRPE_PROF_DIM

Tableau II.7: table descriptive de la dimension « Groupe professionnel ».

5. Dimension personnel (PERSONNEL_DIM) : est la dimension qui décrit le personnel.

PERSONNEL_DIM
MATRICULE
CD_MATRICULE
NOM
PRENOM
DATE_NAISS
ADRESSE
SEXE
SIT_FAMIL
CD_AFFECT
CODE_POSTE
DT_ENTREE
DT_FRT

Figure II.11: dimension personnel.

Dans cette dimension, il est utilisé une clé MATRICULE comme une clé primaire, le tableau suivant donne plus de détails sur cette dimension :

Désignation	Détails
MATRICULE	La clé Primaire de la dimension PERSONNEL_DIM
CD_MATRICULE	Le code du matricule associé aux personnels
NOM	Le nom de l'employé
PRENOM	Le Prénom de l'employé
DATE_NAISS	La date de naissance de l'employé
ADRESSE	L'adresse de l'employé
SEX	Le sexe de l'employé
SIT_FAMIL	La situation familiale de l'employé
CD_AFFECT	Le code d'affectation de l'employé (numéro de l'affectation)
CODE_POSTE	Le code du poste de l'employé
DT_ENTREE	La date d'entrée de l'employé
DT_FRT	La date de sortie « FIN DE RELATION DE TRAVAIL » de l'employé.

Tableau II.8 : table descriptive de la dimension « personnel ».

C. Les indicateurs : appelés aussi mesurables sont des champs numériques sur lesquels on va faire nos analyses.

Les indicateurs qui correspondent à ce volet et qui permettent de mesurer les performances de cette activité sont : effectif payé, effectif permanent, effectif contractuel, affectif actif, effectif prévu, effectif opérationnel, effectif productif.

Table de fait «FAIT_PERSONNEL » :

Pour remplir la table « FAIT_PERSONNEL », il faut créer les contraintes d'intégrité avec les tables de dimensions.

- Les clés primaires des tables de dimensions sont des clés étrangères de la table de fait.
- La clé primaire de la table de fait est la concaténation des clés étrangères.

Désignation	Détails
MATRICULE	La clé Primaire de la dimension PERSONNEL_DIM
CD_GRPE	La clé primaire de la dimension GRPE_PROF_DIM
N_CHANTIER	La clé primaire de la dimension CHANTIER_DIM
CD_FONCTION	La clé primaire de la dimension FONCTION_DIM
PERIODE	La clé primaire de la dimension PERIODE_DIM
Effectif payé	L'indicateur qui indique personnel payé
Effectif permanant	L'indicateur qui indique personnel permanant
Effectif contractuel	L'indicateur qui indique personnel contractuel
Effectif actif	L'indicateur qui indique personnel actif
Effectif prévu	L'indicateur qui indique personnel prévu
Effectif opérationnel	L'indicateur qui indique personnel opérationnel
Effectif productif	L'indicateur qui indique personne productif

Tableau II.9 : La table descriptive « FAIT_PERSONNEL »

D .Le modèle en étoile du volet « PERSONNEL » :

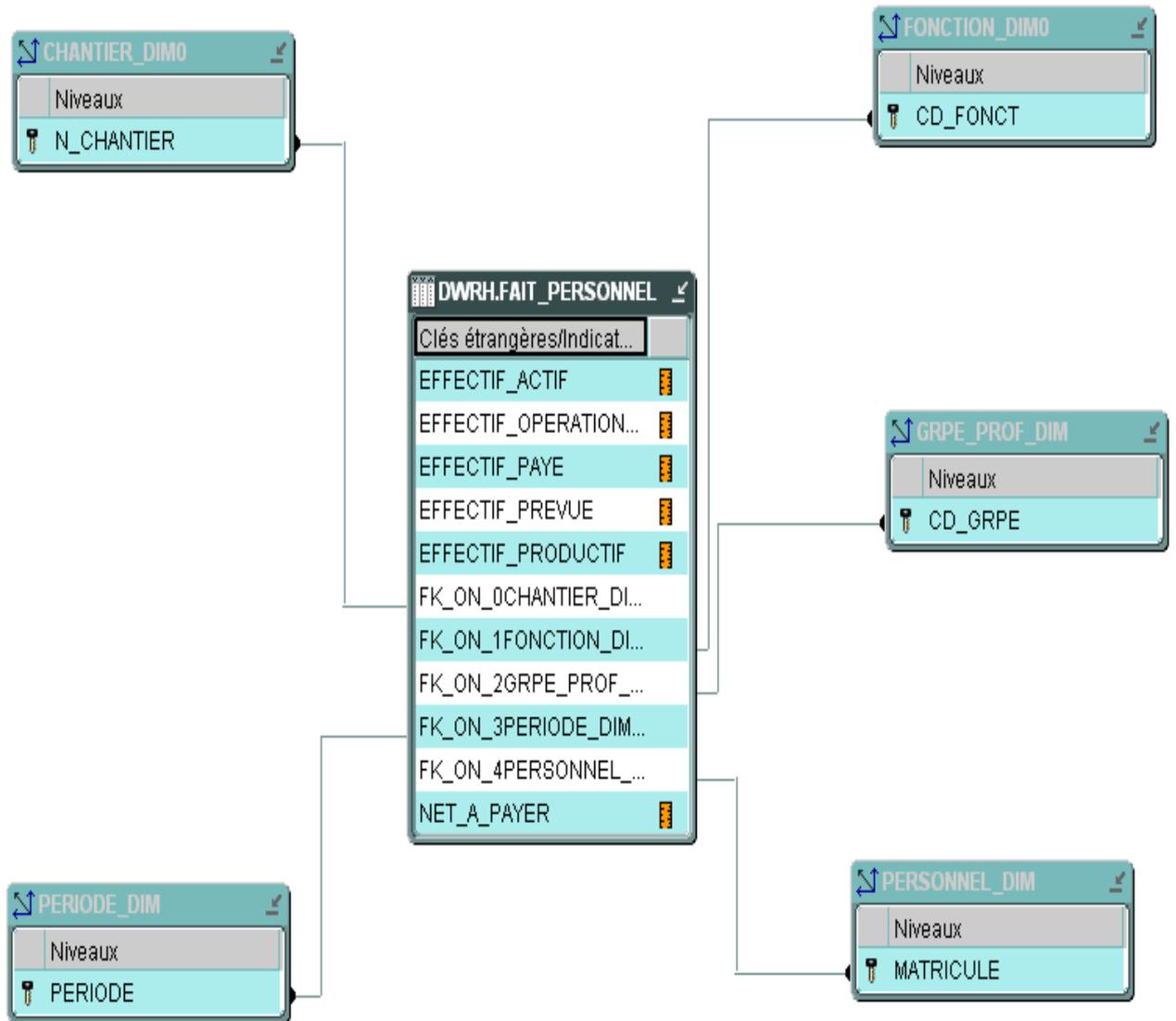


Figure II.12: capture du modèle en étoile du volet « PERSONNEL » .

2) Volet «**Recrutement**» :

a. Définition :

« Le recrutement est une étape fondamentale et incontournable pour chaque organisme professionnel, il s'agit d'un ensemble d'actions mises en œuvre pour trouver un candidat correspondant aux besoins d'une organisation pour un poste donné. » .

b. Les dimensions participantes du modèle :

1. Dimension accord dérogatoire (ACCORD_DEROG_DIM) : Un accord dérogatoire est un accord collectif dérogeant à une ou plusieurs prescriptions réglementaires.

ACCORD_DEROG_DIM
ID_DEROG
DATE_DEROG
CODE_POSTE
CD_DIRECTION

Figure II.13 : dimension Accord dérogation.

Dans cette dimension, il est utilisé une clé ID_DEROG comme une clé primaire, le tableau suivant donne plus de détails sur cette dimension :

Désignation	Détails
ID_DEROG	La clé Primaire de la dimension ACCORD_DEROG_DIM
DATE_DEROG	La date de dérogation de la dimension ACCORD_DEROG_DIM
CODE_POSTE	La clé primaire de la dimension POSTE_TRAVAIL_DIM
CD_DIRECTION	La clé Primaire de la dimension DIRECTION_DIM

Tableau II.10: table descriptive de la dimension « accord_derog ».

2. **Dimension candidat (CANDIDAT_DIM) :** le candidat est une personne qui postule à un emploi dans une entreprise afin de décrocher un poste.

CANDIDAT_DIM
NUM_DOSSIER
NOM
PRENOM
SEXE
SIT_FAMIL
DATE_NAISS
ADRESSE
DT_ENTREE
DT_FRT
CD_AFFECT
CODE_POSTE

Figure II.14: dimension candidat.

Dans cette dimension, il est utilisé une clé NUM_DOSSIER comme une clé primaire, le tableau suivant donne plus de détails sur cette dimension :

Désignation	Détails
NUM_DOSSIER	La clé Primaire de la dimension CANDIDAT_DIM
NOM	Le nom du candidat
PRENOM	Le Prénom du candidat
SEXE	Le sexe du candidat
SIT_FAMIL	La date de naissance du candidat
DATE_NAISS	La date de naissance du candidat
ADRESSE	L'adresse du candidat
DT_ENTREE	La date d'entrée du candidat
DT_FRT	La date de fin de relation de travail du candidat
CD_AFFECT	Le code d'affectation.
CODE_POSTE	La clé primaire de la dimension POSTE_TRAVAIL_DIM

Tableau II.11 : table descriptive de la dimension « candidat ».

3. Dimension contrat (CONTRAT_DIM) : Le contrat est un acte de type privé qui lie juridiquement l'employé avec l'entreprise dans le cadre du contrat de travail.

CONTRAT_DIM
CODE_ CONTRAT
NUM_DOSSIER
MATRICULE
DATE_DEBUT
DUREE
FONCTION

Figure II.15: dimension contrat.

Dans cette dimension, il est utilisé une clé CODE_ CONTRAT comme une clé primaire, le tableau suivant donne plus de détails sur cette dimension :

Désignation	Détails
CODE_CONTRAT	La clé Primaire de la dimension CONTRAT_DIM
NUM_DOSSIER	Le numéro de dossier du contrat
MATRICULE	La clé primaire de la dimension PERSONNEL_DIM
DATE_DEBUT	La date qui débute le contrat
DUREE	La durée du contrat
FONCTION	Le nom de la fonction

Tableau II.12 : table descriptive de la dimension « contrat ».

4. Dimension poste de travail (POST_TRAVAIL_DIM) : Le poste de travail est la fonction occupée par le salarié au sein de l'entreprise.

POST_TRAVAIL_DIM
CODE_ POSTE
CATEGORIE
CD_FONCT
LIB_POSTE
CD_GRPE

Figure II.16: dimension poste de travail.

Dans cette dimension, il est utilisé une clé CODE_ POSTE comme une clé primaire, le tableau suivant donne plus de détails sur cette dimension :

Désignation	Détails
CODE_POSTE	La clé Primaire de la dimension POST_TRAVAIL_DIM
CATEGORIE	La catégorie du poste de travail
CD_FONCT	La clé primaire de la dimension FONCTION_DIM
LIB_POSTE	Libelle poste de travail
CD_GRPE	La clé primaire de la dimension GRPE_PROF_DIM

Tableau II.13 : table descriptive de la dimension « poste de travail ».

5. Dimension direction (DIRECTION_DIM) :c'est l'ensemble des cadres qui est dirigés par l'entreprise

DIRECTION_DIM
CD_DIRECTION
ABRG
LIBLLE

Figure II.17 : dimension direction.

Dans cette dimension, il est utilisé une clé CD_DIRECTION comme une clé primaire, le tableau suivant donne plus de détails sur cette dimension :

Désignation	Détails
CD_DIRECTION	La clé primaire de la dimension DIRECTION_DIM
ABRG	L'abréviation de la direction
LIBLLE	Libelle direction

Tableau II.14 : table descriptive de la dimension « direction ».

- c. **Les indicateurs** : Les indicateurs qui correspondent à ce volet et qui permettent de mesurer les performances de cette activité sont : nombre de demande du personnel, nombre des postes, nombre des contrats signés, nombre des candidats restés, nombre des candidats retenus.

Table de fait «FAIT_RECRUTEMENT » :

Pour remplir la table « FAIT_RECRUTEMENT », il faut créer les contraintes d'intégrité avec les tables de dimensions. Le tableau suivant donne plus de détails sur cette table :

Désignation	Détails
NUM_DOSSIER	La clé Primaire de la dimension CANDIDAT_DIM
CD_DIRECTION	La clé primaire de la dimension DIRECTION_DIM
CODE_POSTE	La clé primaire de la dimension
CODE_CONTRAT	POSTE_TRAVAIL_DIM
ID_DEROG	La clé primaire de la dimension CONTRAT_DIM
Nombre de demande personnel	La clé primaire de la dimension
Nombre des postes	ACCOR_DEROG_DIM
Nombre des contrats signés	L'indicateur qui indique demande personnel
Nombre des candidats restés	L'indicateur qui indique les nombres de poste
Nombre des candidats retenus	L'indicateur qui indique les contrats signés
	L'indicateur qui indique candidats restés
	L'indicateur qui indique candidats retenus

Tableau II.15 : La table descriptive « FAIT_RECRUTEMENT »

D .Le modèle en étoile du volet « RECRUTEMENT » :

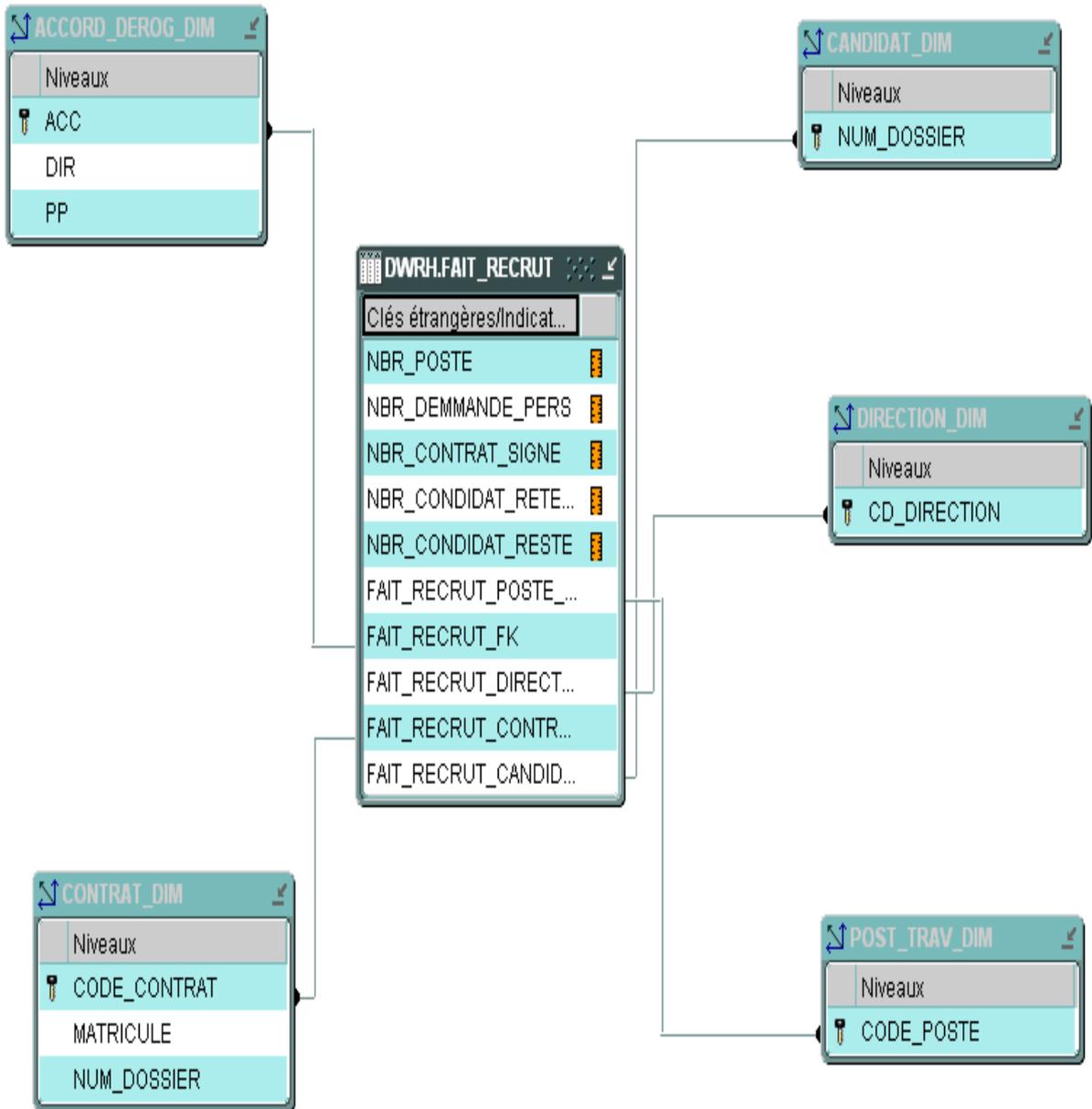


Figure II.18: capture du modèle en étoile du volet « RECRUTEMENT » .

3) Volet « **FORMATION** » :

a. Définition : « La formation professionnelle est un apprentissage nécessaire à l'exercice d'une activité professionnelle. L'objectif est de fournir au marché du travail des individus possédant un savoir-faire sans cesse mis à jour. Le financement est assuré par les entreprises, soit par cotisation, soit par dépenses en formation interne. ».

b. Les dimensions participantes du modèle :

1. **Dimension action (ACTION_DIM) :** Une action est une fraction du capital social, elle représente les droits dans une entreprise .

ACTION_DIM
CODE_ACTION
CODE_FORM
LIB_ACTION
DUREE
DIPLOME
CODE_ORG

Figure II.19: dimension action.

Dans cette dimension, il est utilisé une clé `CODE_ACTION` comme une clé primaire, le tableau suivant donne plus de détails sur cette dimension :

Désignation	Détails
CODE_ACTION	La clé primaire de la dimension ACTION_DIM
CODE_FORM	La clé primaire de la dimension FORMATION_DIM
LIB_ACTION	Libelle de l'action
DUREE	La durée de l'action
DIPLOME	Le diplôme de la dimension action
CODE_ORG	La clé primaire de la dimension ORGANISME_DIM

Tableau II.16: table descriptive de la dimension « action ».

2. **Dimension formation (FORMATION_DIM) :** La formation est un droit, indépendamment du statut professionnel, grâce à une formation, le personnel peut améliorer ses compétences dans de nouvelles techniques. La formation est financée par l'entreprise.

FORMATION_DIM
CODE_FORM
LIB_FORM

Figure II.20 : dimension formation.

Dans cette dimension, il est utilisé une clé CODE_FORMATION comme une clé primaire, le tableau suivant donne plus de détails sur cette dimension :

Désignation	Détails
CODE_FORM	La clé primaire de la dimension FORMATION_DIM
LIB_FORM	Libelle de la formation

Tableau II.17 : table descriptive de la dimension « Formation ».

3. **Dimension organisme (ORGANISME_DIM) :** Un organisme est l'ensemble des éléments qui compose une structure fonctionnelle.

ORGANISME_DIM
CODE_ORG
LIB_ORG
ADR_ORG

Figure II.21 : dimension organisme.

Dans cette dimension, il est utilisé une clé CODE_ORG comme une clé primaire, le tableau suivant donne plus de détails sur cette dimension :

Désignation	Détails
CODE_ORG	La clé primaire de la dimension ORGANISME_DIM
LIB_ORG	Libelle de l'organisme
ADR_ORG	Adresse de l'organisme

Tableau II .18 : table descriptive de la dimension « organisme ».

C. Les indicateurs :

Les indicateurs qui correspondent à ce volet et qui permettent de mesurer les performances de cette activité sont : effectif formé, budget prévu, coût par agent, cout par action, homme /jour.

Table de fait «FAIT_FORMATION » :

Chapitre 3 : La zone d'entreposage

Pour remplir la table « FAIT_FORMATION », il faut créer les contraintes d'intégrité avec les tables de dimensions. Le tableau suivant donne plus de détails sur cette table :

Désignation	Détails
CODE_ACTION	La clé Primaire de la dimension ACTION_DIM
CD_DIRECTION	La clé primaire de la dimension DIRECTION_DIM
CODE_FORM	La clé primaire de la dimension FORMATION_DIM
CODE_ORG	La clé primaire de la dimension ORGANISME_DIM
MATRICULE	La clé primaire de la dimension PERSONNEL_DIM
Effectif formé	L'indicateur qui indique l'effectif formé dans la formation
Budget prévu	L'indicateur qui indique le budget prévu pour la formation
Coût par agent	L'indicateur qui indique le coût par agent
Coût par action	L'indicateur qui indique le coût par action
Homme /jour	L'indicateur du coût d'un individu par jour durant la formation

Tableau II .19: La table descriptive « FAIT_FORMATION»

D .Le modèle en étoile du volet « RECRUTEMENT» :

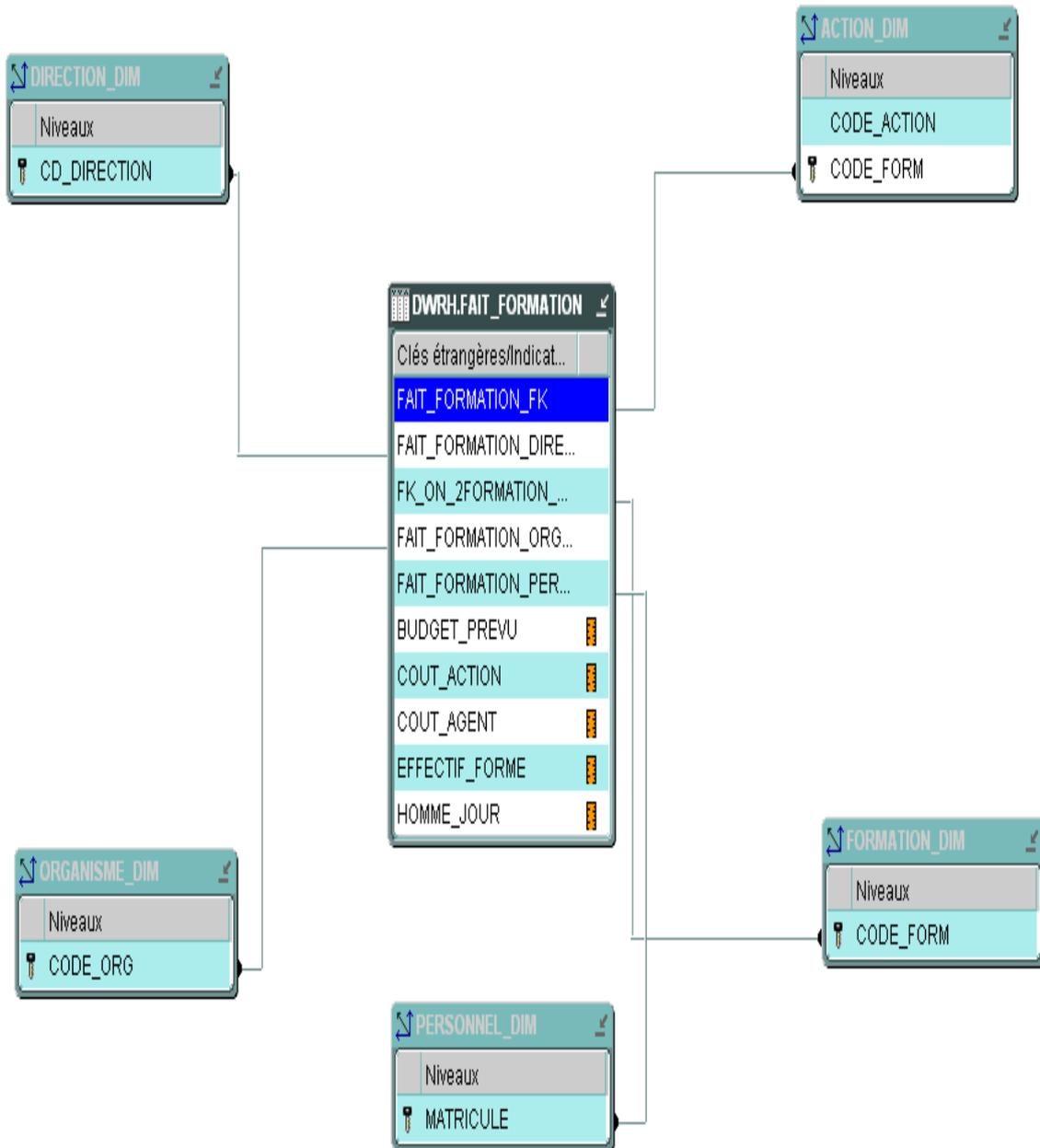


Figure II.22: capture du modèle en étoile du volet «FORMATION» .

4) Volet « MOUVEMENT » :

a. Définition :

ce volet contient les informations relatives au mouvement interne et externe de l'entreprise .

b. Les dimensions participantes du modèle :

1. Dimension période (PERIODE_DIM) :

PERIODE_DIM
PERIODE
EXERCICE

Figure II.23 : dimension période

Dans cette dimension, il est utilisé une clé PERIODE comme une clé primaire.

c. Les indicateurs :

Les indicateurs qui correspondent à ce volet et qui permettent de mesurer les performances de cette activité sont : nombre de droits au congé, nombre de consommation, moyenne mensuelle, effectif en FRT, nombre de démissionnaires.

Table de fait «FAIT_MOUVEMENT » :

Pour remplir la table « FAIT_MOUVEMENT », il faut créer les contraintes d'intégrité avec les tables de dimensions. Le tableau suivant donne plus de détails sur cette table :

Désignation	Détails
N_CHANTIER	La clé Primaire de la dimension CHANTIER_DIM
PERIODE	La clé primaire de la dimension PERIODE_DIM
MATRICULE	La clé primaire de la dimension PERSONNEL_DIM
Nombre de droits congé	L'indicateur qui indique les droits de congé
Nombre de consommation	L'indicateur qui indique le nombre de consommation
Moyenne mensuelle	L'indicateur qui indique la moyenne mensuelle
Effectif en FRT	L'indicateur qui indique l'effectif en fin de relation de travail
Nombre démissionnaires	L'indicateur qui indique les démissionnaires

Tableau II.20: La table descriptive « FAIT_MOUVEMENT »

D .Le modèle en étoile du volet « MOUVEMENT » :

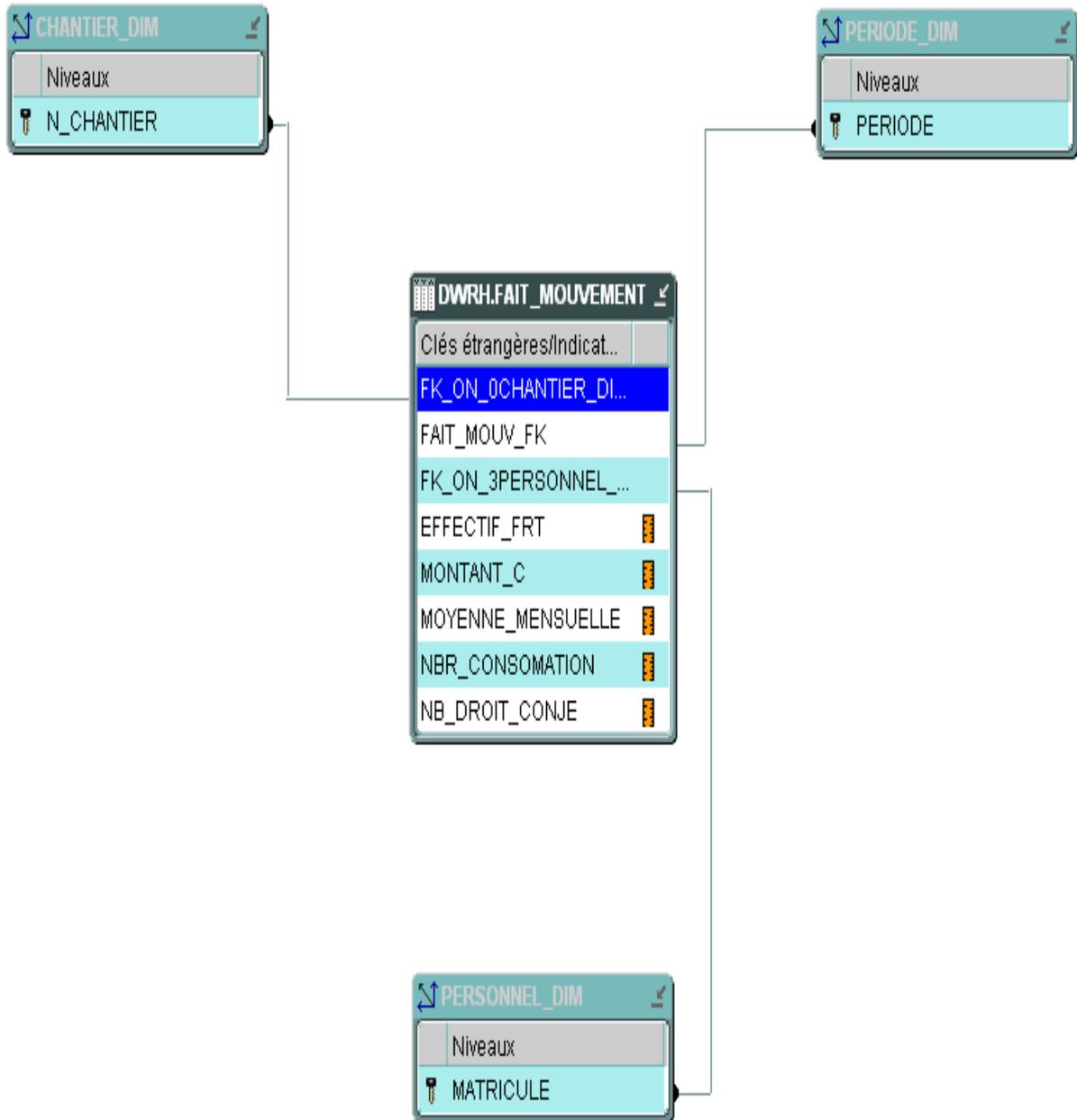


Figure II.24 : capture du modèle en étoile du volet «MOUVEMENT» .

5) Volet « Œuvre social » :

- a. **Définition** : Les œuvres sociales sont les organismes prenant en charge les activités sociales et culturelles établies dans l'entreprise au profit des salariés ou de leurs familles.

b. Les dimensions participantes du modèle :

1. **Dimension veuve (VEUVE_DIM)** : est la dame dont le conjoint est mort, elle reçoit des aides de l'entreprise.

VEUVE_DIM
NOM_V
PRENOM_V
ADRESSE
TEL
SIT_FAMIL
MATRICULE_MARI

Figure II.25 : dimension veuve.

Dans cette dimension, il est utilisé une clé NOM_V et PRENOM_V comme des clés primaires, le tableau suivant donne plus de détails sur cette dimension :

Désignation	Détails
NOM_V	La clé primaire de la dimension VEUVE_DIM
PRENOM_V	La clé primaire de la dimension VEUVE_DIM
ADRESSE	L'adresse de la veuve
TEL	Le téléphone de la veuve
SIT_FAMIL	La Situation familiale de la veuve
MATRICULE_MARI	Le matricule du mari de la veuve

Tableau II .21 : table descriptive de la dimension « veuve ».

a. Les indicateurs :

Les indicateurs qui correspondent à ce volet et qui permettent de mesurer les performances de cette activité sont : allocation familiales, nombre de jour maladie, montant CNAS, montant prêt demandé, nombre de veuve.

Table de fait «FAIT_OEUVRE SOCIAL » :

Pour remplir la table « FAIT_OEUVRE SOCIAL », il faut créer les contraintes d'intégrité avec les tables de dimensions. Le tableau suivant donne plus de détails sur cette table :

Désignation	Détails
CD_DIRECTION	La clé Primaire de la dimension DIRECTION_DIM
PERIODE	La clé primaire de la dimension PERIODE_DIM
MATRICULE	La clé primaire de la dimension PERSONNEL_DIM
NOM_V	La clé Primaire de la dimension VEUVE_DIM
PRENOM_V	La clé Primaire de la dimension VEUVE_DIM
Allocations familiales	L'indicateur qui indique le nombre d'enfant
Nombre de jour maladie	L'indicateur qui indique les jours de maladie
Montant CNAS	L'indicateur qui indique montant CNAS
Montant prêt demandé	L'indicateur qui indique les prêts demandés
Nombre de veuve	L'indicateur qui indique les veuves

Tableau II .22 : table descriptive de « FAIT_OEUVRE SOCIAL»

D .Le modèle en étoile du volet « ŒUVRE SOCIAL » :

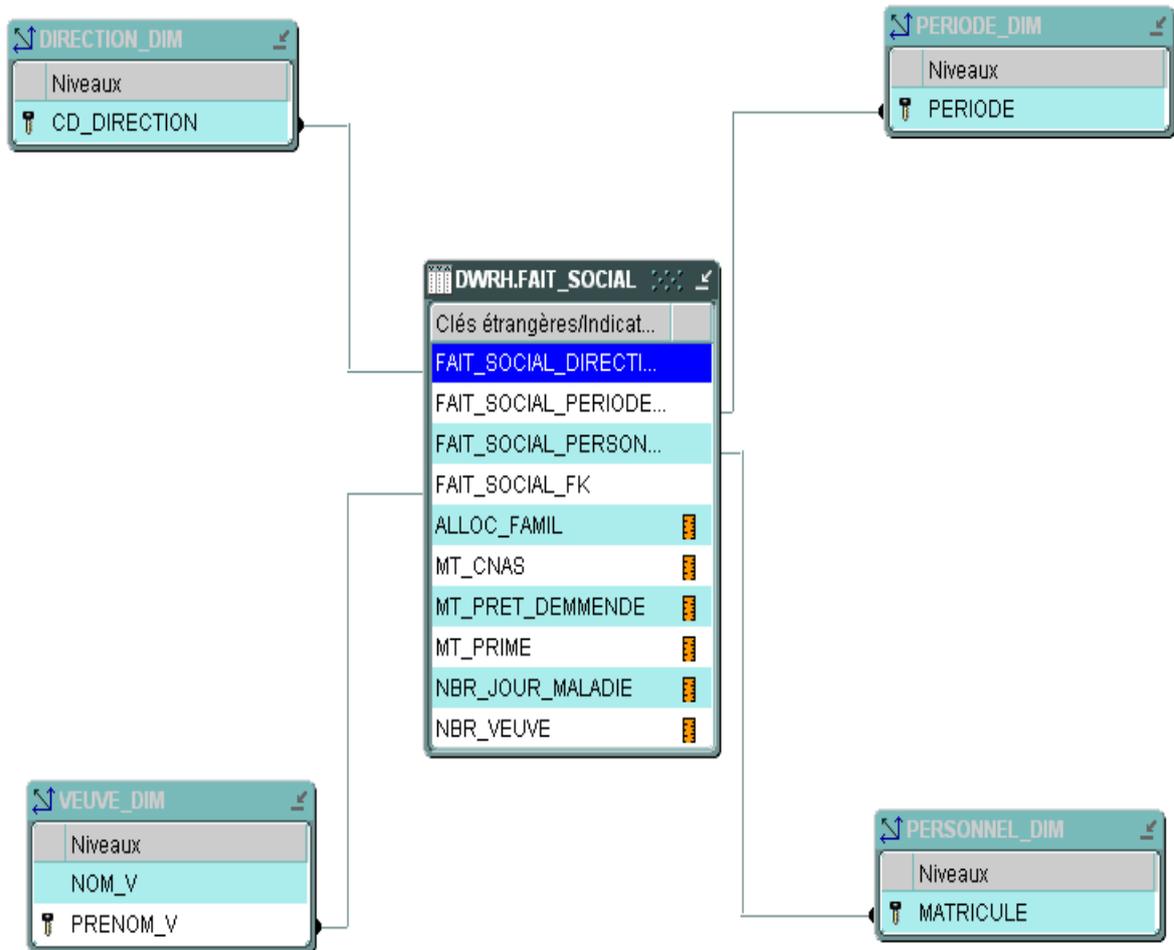


Figure II.26: capture du modèle en étoile du volet «ŒUVRE SOCIAL»

Conclusion

La zone d'entreposage constitue la zone exploitable par les utilisateurs. La modélisation de cette zone se fait grâce à la modélisation dimensionnelle. Cette manière de représenter les données offre aux utilisateurs des modèles intuitifs et compréhensibles permettant de naviguer et de manipuler les données, détaillées sans difficulté afin de satisfaire leurs besoins en analyse.

La finalisation de la conception d'une étoile d'entrepôt nous permet de passer à la construction de la zone d'alimentation. Cette zone constitue l'objet du prochain chapitre.

Introduction

L'ETL, ou l'alimentation du Data Warehouse, est une étape des plus importantes dans un projet Data Warehouse. Cette étape a pour objectif d'assurer l'acheminement des données des systèmes sources jusqu'à l'entrepôt de données, en passant par les différentes phases de nettoyage et de transformations nécessaires.

Pour la réalisation et la mise en place de la solution, il a été nécessaire de recourir à un certain nombre d'outils et mettre en place des environnements d'exécutions.

Ce chapitre contient deux parties :

Partie 1 : définit le concept de l'alimentation de notre entrepôt.

Partie 2 : décrit l'environnement de mise en place et les outils utilisés, ainsi que l'environnement existant (matériels et logiciels), et dans lequel évoluera notre système.

Partie 1 :

La conception du processus d'alimentation nécessite les étapes suivantes :

- Etude préliminaire.
- Choix de l'architecture.
- Conception des processus de chargement :
 - Processus de chargement des tables de dimension / Processus de chargement des dimensions changeantes (mise à jour).
 - Processus de chargement des tables des faits.

1. L'étude préliminaire :

1.1 Les sources de données et leurs emplacements :

Afin de définir l'emplacement des informations dans les différents systèmes sources et d'en choisir les emplacements les plus fiables, nous avons travaillé de manière étroite avec les DBA et les gestionnaires du groupe ENGTP.

Les sources de données de notre entrepôt sont l'ensemble des bases de données opérationnelles des ressources humaines à savoir : gestion du personnel, gestion des

recrutements, gestion des mouvements, gestion des formations et mises à niveaux, gestion des œuvres sociales.

Le nombre important des tables, la redondance des données et l'intervention de différents systèmes, rendent cette tâche très ardue. Afin de mener à bien cette détection, nous avons :

- Lister les données nécessaires à partir des étoiles conçues.
- Lister les emplacements de chaque donnée.
- Choisir la source la plus fiable et la valider comme source de chargement.
- Dresser un tableau qui établit le lien entre données sources et données cibles avec les transformations nécessaires.

Cette étape s'achève par l'élaboration entre les données sources et les données cibles.

1.2 Définition de la périodicité de chargement :

La périodicité de chargement est étudiée pour chaque étoile séparément. , ce qui n'empêche pas une synchronisation dans les chargements des dimensions communes.

Avant de décider de la périodicité du chargement, les contraintes suivantes doivent être prises en considération :

- La quantité de données à charger.
- Le temps de non activité des systèmes sources.

L'étoile qui engendrera les chargements les plus importants, en termes de volume, n'est autre que l'étoile du « mouvement » et celle du « personnel », et ce volume ne dépasse pas le seuil pour un chargement moins que mensuel.

2. Choix de l'architecture :

L'élaboration d'une architecture du processus de l'alimentation « ETL » dès le début du projet est très importante. En effet, le choix d'une architecture affecte pratiquement toutes les composantes du projet. Tout changement de celle-ci engendrera nécessairement d'importants changements dans l'ensemble du projet.

Partant de ce constat, il est nécessaire de mettre en place une architecture consistante à même de prendre en charge toutes les contraintes auxquelles on doit faire face.

Le processus de l'ETL peut se faire de différentes manières. Nous avons choisi la méthode

« Push-Pull ». Puisque on a la possibilité de se connectée d'une façon directe aux sources de données via réseaux, les systèmes sources poussent « Push » les données vers la zone de préparation, puis le « Pull » tire ces données vers notre entrepôt.

Au niveau du serveur « ELIT » il est procédé à l'affectation de clés artificielles et à quelques transformations nécessaires avant le chargement final dans la zone d'entrepasage.



Figure II.27: architecture du processus «ETL » [20]

3. Conception des processus de chargement :

Le diagramme d'activités suivant décrit le processus général de l'alimentation de l'entrepôt de données dès sa mise en service :

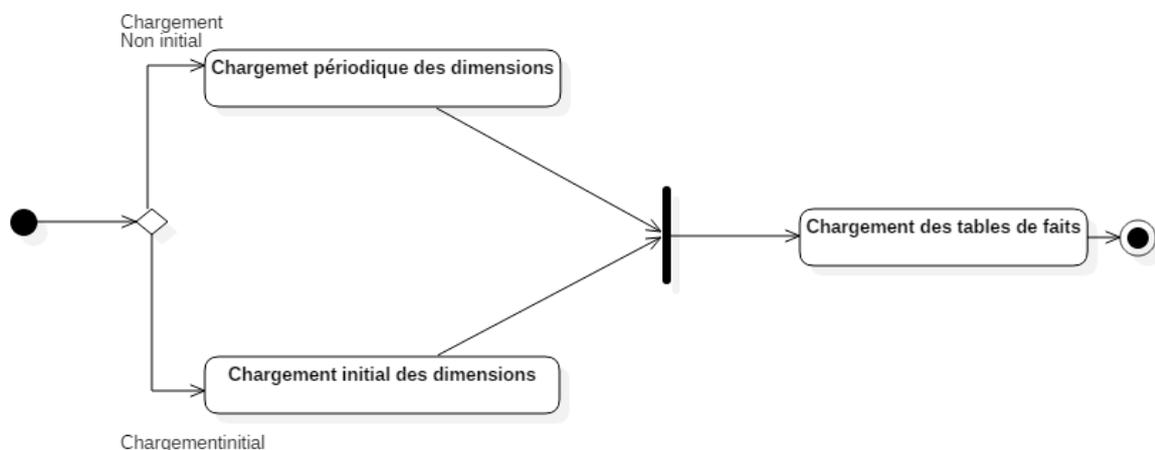


Figure II.28 : Diagramme d'activité du processus d'alimentation.

Deux types de tables dans l'entrepôt de données « faits, dimensions » doivent être distingués. Chaque type de table diffère dans les informations qu'elle contient, et d'où donc l'adoption de deux processus de chargement.

La stratégie d'extraction adoptée consiste à comparer des chargements successifs afin de détecter les changements. Cette stratégie, étant la plus fiable, est incontournable pour la capture des changements pour des raisons de non fiabilité des champs date, généralement non renseignés. Cette détection se fait au niveau de la zone de préparation améliorant sensiblement les performances.

3.1 Processus de chargement des tables de dimension et la mise à jour des dimensions changeante :

Les tables de dimension constituent le contexte de la table de faits. Elles représentent le point d'entrée au « Data Warehouse ». Une dimension est généralement constituée : d'une clé artificielle, d'une clé naturelle et des attributs.

Le processus de chargement de dimension doit, outre le chargement des données, assurer :

- **La gestion des clés artificielles:** affectation des clés et mise en correspondance avec les clés naturelles.

- **La gestion de l'évolution de dimension :** gérer les changements que subissent les dimensions. Il existe trois types de traitement par rapport à l'évolution d'une dimension :
 - **Type 1 « écrasement » :** consiste à mettre à jour l'attribut subissant un changement.
 - **Type 2 « création d'un nouvel enregistrement » :** consiste à créer un nouvel enregistrement afin de sauvegarder tout le cycle d'évolution de la dimension.
 - **Type 3 « déplacement de la valeur a changée dans un attribut ancien » :** consiste à prévoir des attributs pour enregistrer les changements éventuels. Il permet de sauvegarder un nombre défini de changements.

Le diagramme suivant illustre le processus de chargement de dimension:

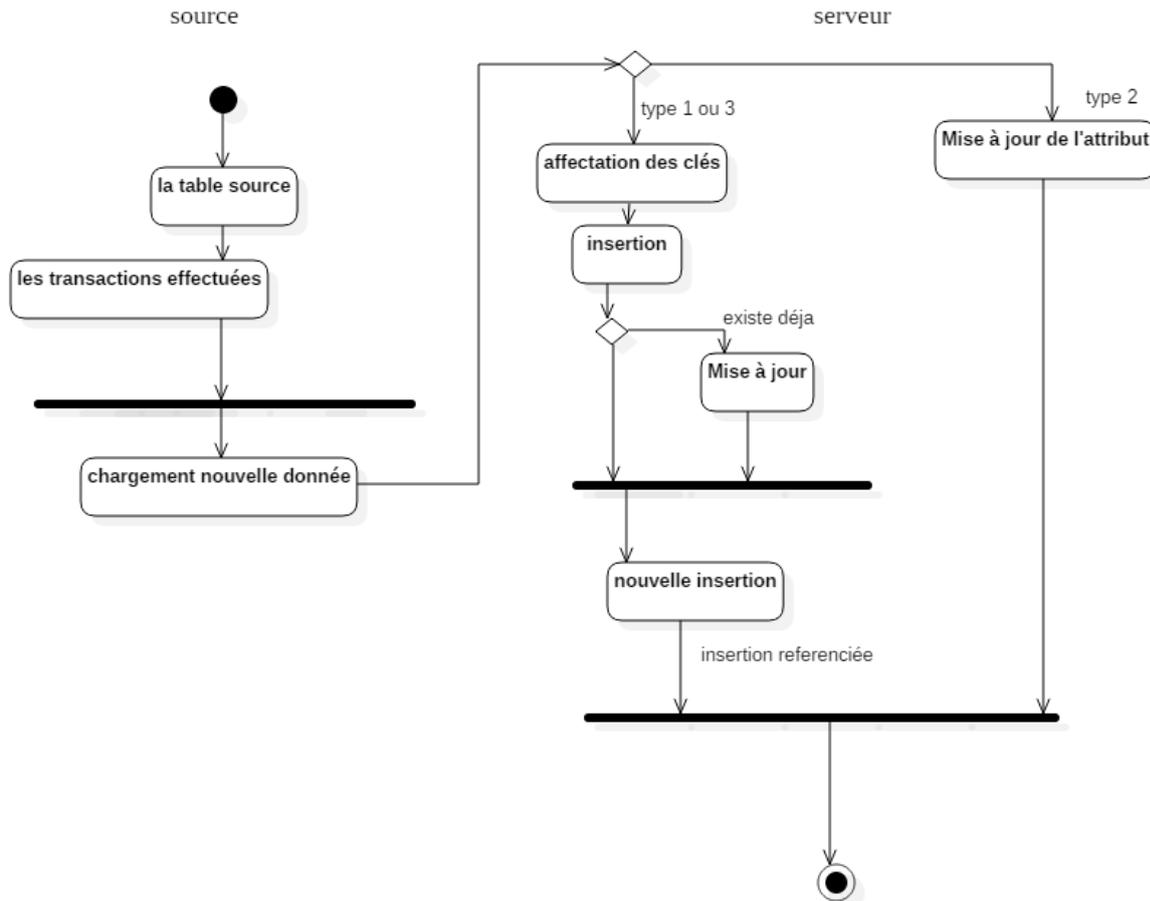


Figure II.29: diagramme d'activité du processus de chargement de dimension

La stratégie pour la détection des changements se fait comme suit :

Tout d'abord le script de mise à jour va procéder à la détection de changement à partir de la table source. Après un nouvel enregistrement est nécessaire avec incrémentation de la clé artificielle.

Après on refait le chemin inverse pour détecter la ligne changée, en comparant les lignes de la dimension avec celle de la source, si cette ligne n'est plus existante il ajoute une date d'expiration juste histoire de garder la trace historisée du changement.

3.2 Processus de chargement des tables de faits :

L'extraction des faits se fait avec les clés naturelles utilisées dans les systèmes sources. L'étape qui précède le chargement de la table des faits consiste à

remplacer les clés naturelles par les clés artificielles. La substitution peut se faire directement par le biais des tables de dimension, ce qui est correct mais très lent. Pour cela on utilise des tables de référencement.

Le processus de chargement des tables des faits doit garantir l'intégrité référentielle vis-à-vis des tables de dimensions.

Le diagramme d'activité suivant illustre le processus de chargement adopté pour une table de faits :

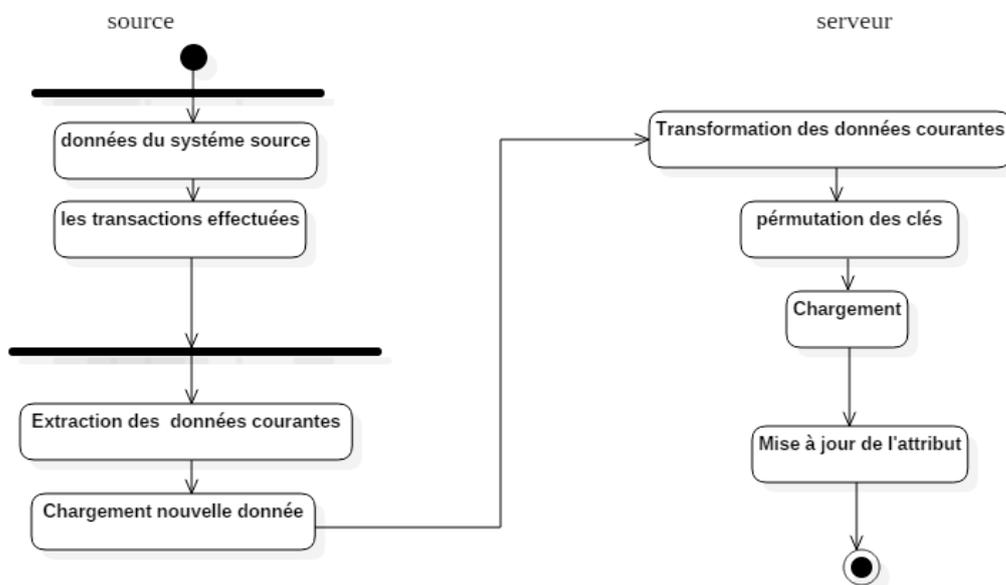


Figure II.30 : diagramme d'activité du processus de chargement de faits.

Partie 2 :

1. Implémentation

1.1 Périmètre technique et fonctionnel

Cette partie décrit les infrastructures déjà en place. En effet, cette dernière est une étape à ne pas négliger, car la diversité des sources et leurs plateformes techniques pourront engendrer des problèmes de compatibilité.

1.1.1 Matériel

Les systèmes sources sont installés sur différentes plateformes :

- ❖ Machine INTEL : DELL, ou HP.

1.1.2 Systèmes d'exploitation

Lors de notre étude il a été recensé les systèmes suivants :

- ❖ « **Windows XP pack 3** ».
- ❖ « **Windows 7 Professional** ».

L'étude du matériel existant nous a permis de prévoir des solutions à certains problèmes, tels que le non compatibilité des machines Windows 7 avec la version de «**Oracle 10g** » nécessaire pour l'exécution des programmes d'extractions.

1.2 Architecture technique de la solution

La figure suivante illustre la structure et l'architecture technique de la solution proposée :

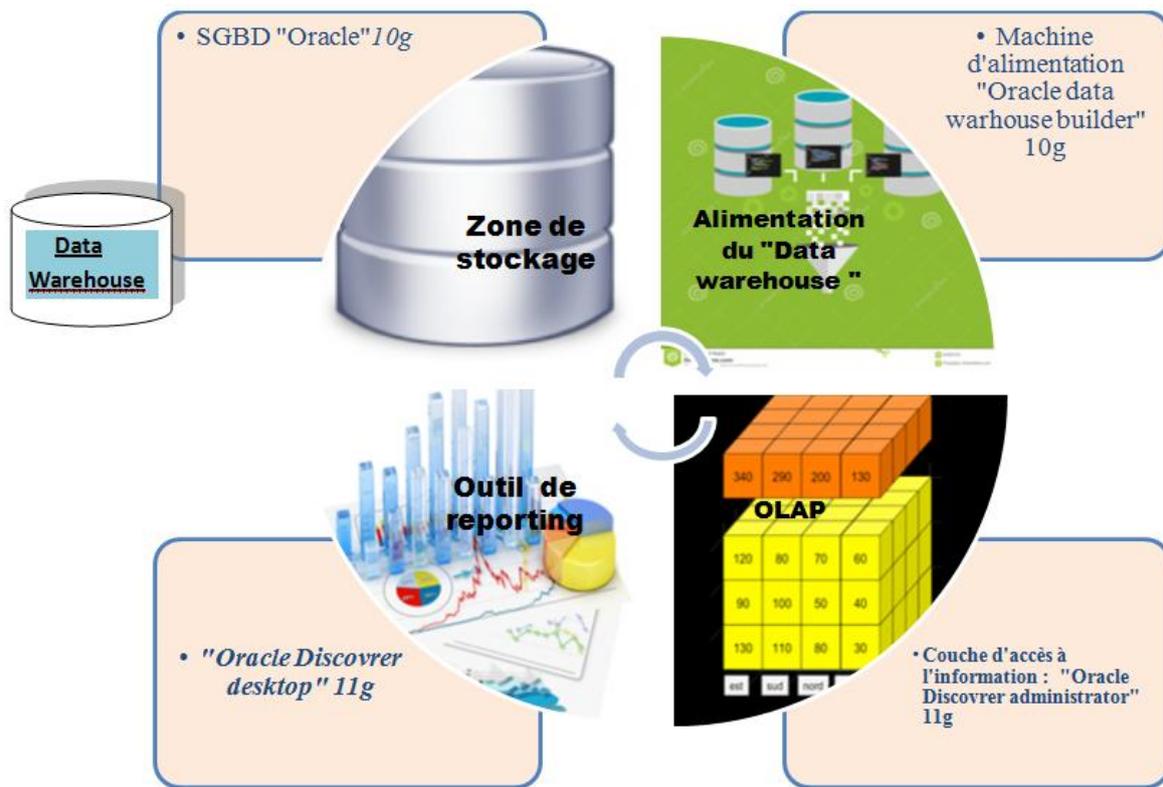


Figure II.31 : Architecture technique de la solution

1.3 Zone de stockage

Lors de la conception, nous avons conçu et mise en place une seul base de données : la base de données de la zone d'entrepasage.

Le choix du système de gestion de base de données s'est fait en fonction de la finalité de la base de données, de son utilisation et des données qu'elle contiendra.

- **Notre base de données « DWRH » :**

La base de données « DWRH », a été implémentée sous le SGBD « **Oracle 10g** », ce système intègre un ensemble d'outils d'administration et de configuration. Aussi ce SGBD est pré configuré pour la mise en place d'un Data Warehouse.

1.4 Zone d'alimentation de l'entrepôt :

Notre implémentation du processus de chargement a eu recours à l'utilisation du langage « SQL » langage de requête structurée, pour l'alimentation de l'entrepôt.

Pour cela nous avons utilisé « **oracle datawarehouse builder** » dans sa version 10g connu comme un client oracle performant.

1.5 Zone de restitution :

Cette zone représente l'interface entre l'utilisateur et le Data Warehouse. Elle est constituée d'un ensemble d'outils qui doivent permettre aux utilisateurs d'exploiter le système mis en place dans les meilleures conditions possibles. Ainsi plusieurs outils et serveurs ont été mis en place. Tous les outils que nous avons utilisés ont été exigés par l'entreprise et appartiennent à la gamme « **oracle** » :

- ❖ **Oracle discoverer administrator « 11g »** : nous avons utilisé cet outil afin de créer les cubes et la couche EUL (End User Layer).
- ❖ **Oracle discoverer desktop « 11g »** : cette composante est une application windows (Client/serveur) qui permet à l'utilisateur de construire des rapports. Il est installé sur le serveur d'application de l'entreprise, accessible depuis le site intranet.

➤ Présentation de l'outil

« **Oracle Business Intelligence Discoverer Desktop** » est un outil d'accès aux données. Nous l'avons utilisé pour voir les informations dans les bases de données dans le but d'aider les décideurs de l'entreprise à voir les données qu'ils désirent, et les analyser pour prendre en charge les décisions et créer des rapports à conserver pour le suivi des choses.

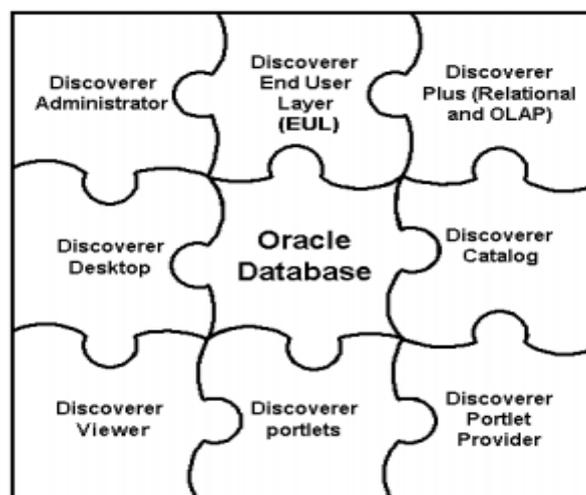


Figure II.32 : Composants Oracle Discoverer [9]

2. Déploiement :

Pour mieux décrire le déploiement de la solution, on utilise le modèle de déploiement U.M.L qui permet de présenter l'architecture de déploiement d'une manière simple et compréhensible.

Comme on peut le voir, notre solution comporte trois zones : zone d'alimentation, zone de stockage et zone de restitution. Afin d'illustrer cela, on propose deux diagrammes de déploiements : Un diagramme pour la zone d'alimentation et un autre diagramme pour la zone de restitution. La zone de stockage étant liée directement aux deux autres zones sera décrite dans les deux modèles.

2.1 Déploiement de la zone d'alimentation :

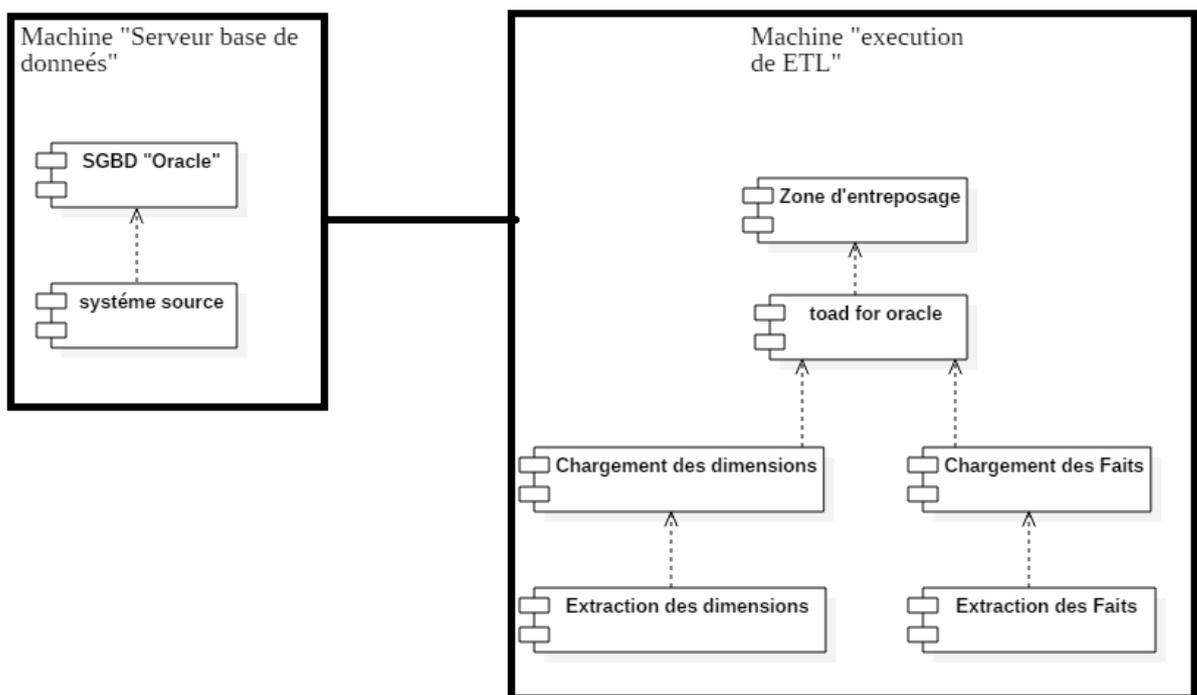


Figure II.33 : Diagramme de déploiement de la zone d'alimentation.

2.2 Déploiement de la zone restitution :

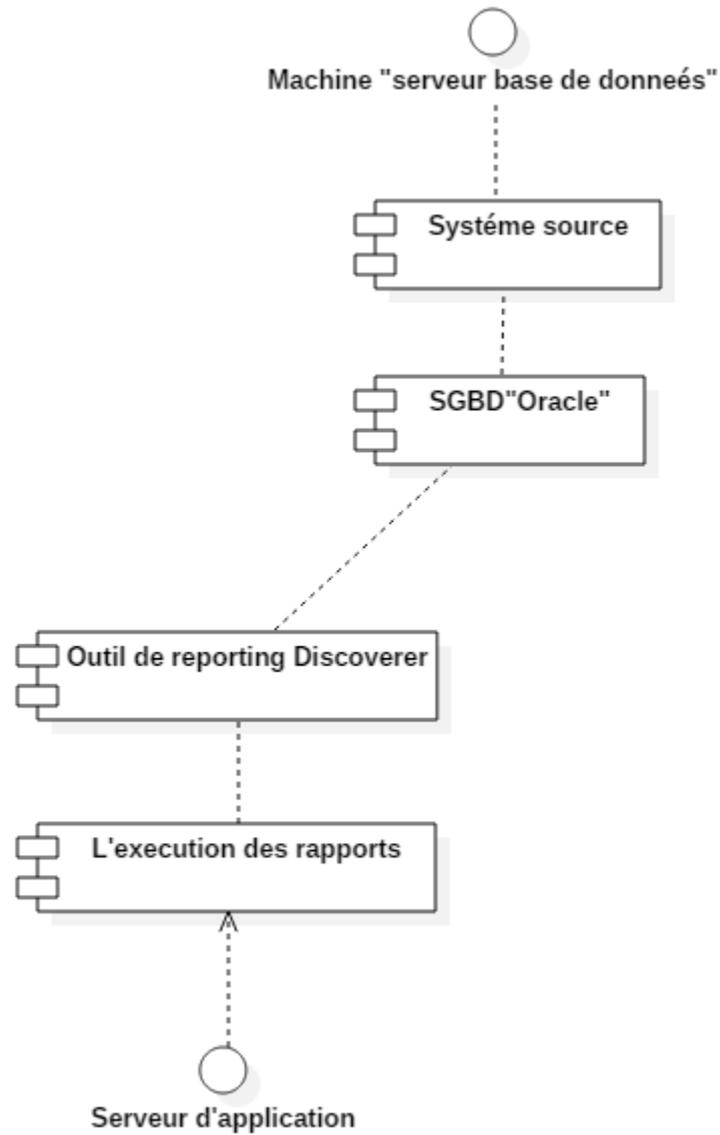


Figure II.34 : Diagramme de déploiement de la zone de restitution.

Conclusion

Le processus ETL assure la livraison des données conformes, cohérentes et correctes. Lors de la conception de la zone d'alimentation, nous avons travaillé pour assurer le chargement des données et leurs qualités, aussi pour garantir la transparence des processus de consolidation et de chargements qui assurent la qualité des données.

La partie de restitution est la partie sur laquelle l'utilisateur final aura à interagir.

La zone de stockage a été implémentée sous le SGBD « Oracle », dans sa distribution «Oracle 10g ».

Le déploiement de la solution se divise en deux parties :

- Déploiement de la zone d'alimentation.
- Déploiement de la zone de restitution.

Le déploiement se fait pour le moment sur l'intranet, et sera étendu à l'ensemble des postes décideurs dès que les résultats des tests fonctionnels auront été approuvés.

Exploiter les données à disposition de l'entreprise afin de leur donner de la valeur ajoutée, tel est le défi des entreprises modernes.

Dans ce cadre, et afin de palier à des problèmes récurrents dans le processus de prise de décision, ENGTP a initié le projet de réalisation d'un Data Warehouse afin de remettre en place un système décisionnel fiable et efficace.

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

Bien que la méthode des entretiens soit l'outil principal pour la collecte des besoins dans ce travail, l'étude des besoins déjà exprimés sous forme de rapports et de processus de prise de décision nous ont été d'une grande utilité pour définir de manière claire le périmètre et les besoins réels des utilisateurs. Cette étude a fait ressortir cinq volets d'analyses dignes d'intérêt pour la gestion des ressources humaines de l'entreprise qui sont : Personnel, recrutement, formation, mouvement, œuvres sociales.

Dans un deuxième temps, la modélisation de la zone de stockage des données s'est faite grâce aux principes de la modélisation dimensionnelle. Cette conception offre une vision claire et une compréhension intuitive des modèles proposés. Nous avons de ce fait proposé des modèles en étoiles des cinq activités recensés.

La partie d'alimentation de la zone de stockage « implémentation physique des modèles dimensionnels sur un SGBD relationnel » a été sans nul doute la partie du projet la plus fastidieuse et consommatrice en temps ; nous permettant de vérifier le postula disant qu'il est nécessaire d'y consacrer plus de 80% du temps de réalisation d'un Data Warehouse.

Cette étape nous a permis de concevoir et de réaliser, grâce à des outils de la gamme « oracle », les routines d'extraction, transformation et chargement des données.

Le « BI », pour la diffusion et la gestion des documents, et d'outils de reporting et de navigation dans les données, offrant à l'utilisateur la possibilité d'exploiter les données de l'entrepôt via n'importe quel poste dédié aux décideurs. La partie restitution a aussi nécessité le développement des volets de gestion des utilisateurs, d'administration de l'entrepôt et de supervision de ce dernier.

Pour la mise en route de la solution, nous avons entamé le travail de déploiement en choisissant des sites pilotes. Ce déploiement obéit à une démarche clairement illustrée grâce à des digrammes de déploiement présentés dans le dernier chapitre du rapport.

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

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

- Suivre le déploiement actuel et recueillir les correctifs et remarques des utilisateurs.
- Etendre le déploiement de manière à couvrir, à terme, la totalité des structures de GTP.

- Utilisation des méthodes et algorithmes de « Data Mining » pour une meilleure exploitation des données.

- Utiliser des techniques d'optimisation des requêtes pour le data warehouse tels que :
 - Les vues matérialisées.
 - Les indexe bitmap.
 - La fragmentation horizontale.

Ouvrages

[1] : Le Moigne J.L., « *La théorie du système général, théorie de la modélisation* », P.U.F., 1977.

[2] : Yazid Grim (Business Intelligence) , article sur l'introduction au monde de l'informatique décisionnelle ; Fleur-Anne .Blain ; 07 Mai 2008 .

[3] : Jean-Michel Franco; « *Le Data Warehouse, le Data Mining* » ; Eyrolles 1997.

[4]: W. H. Inmon ;« *Building the Data Warehouse Third Edition*» ; Wiley Computer Publishing 2002.

[5]: [**Le Moigne, 1977**] : Le Moigne J.L., « La théorie du système général, théorie de la modélisation », P.U.F., 1977

[6] : [Nakache, 1998] : Didier Nakache; « Data Warehouse et Data Mining »; Conservatoire National des Arts et Métiers de Lille; Version 1.1; 15 juin 1998

[7] : R. Kimball ; « Entrepôts de données : Guide pratique du concepteur de Data Warehouse » ;**Wiley Computer Publishing 1996.**

[8] : Matthieu Grimaud ; «comment assurer l'intégration des données structurées dans l'entrepot de données » ;mémoire de fin d'étude ;Technologie de l'information ; Paris ;années 2007 .

[9] : Building Data Warehouse by Inmon chapter 2: The Data Warehouse EnvironmentIT-Slideshares <http://it-slideshares.blogspot.com/> .

[10] : Entrepôts de données et bases multidimensionnelles, Paris, 2002

[11] : <http://www.intranetjournal.com>

[12] Franco J.-M., Le data warehouse : le data mining, Paris, 1997

[13] : <http://www.systemeetl.com>

[14] : <http://www.redbooks.ibm.com>

[15] : J.F. Goglin; « La Construction du Datawarehouse : du Datamart au Dataweb »

[16]: Lamri Chouder ; « *Entrepôt Distribué de Données* » ; Thèse de Magistère Option : SI; Institut National de Formation en Informatique (I.N.I) 2007.

Bibliographie

[17] : R. Kimball et J. Caserta ; « The Data warehouse ETL Toolkit » ;WileyPublissing, INC 2004.

[18]: <http://www.information-management.com>

[19] : R. Kimball et M. Ross ; « *Entrepôts de Données : Guide Pratique de Modélisation Dimensionnelle 2ème édition* » ; Vuibert 2002.

[20] : Conception et réalisation d'un Data Warehouse pour la mise en place d'un système décisionnel

Articles web et Thèses

[21] : Abdenour Bouzghoub ; « *Modélisation des Entrepôts de données XML : Application au domaine de la sécurité sociale* » ;

Bibliographie

[22] : Chuck Ballard, Dirk Herreman, Don Schau, Rhonda Bell, Eunsang Kim, Ann Valencic; Data Modeling Techniques for Data Warehousing; International Technical Support Organization février 1998.

[23] : E. F. Codd ; « *Providing OLAP (On-Line Analytical Processing) to User- Analysts : an IT mandate.* » ; Technical report ; E.F. Codd& Associates; 1993.

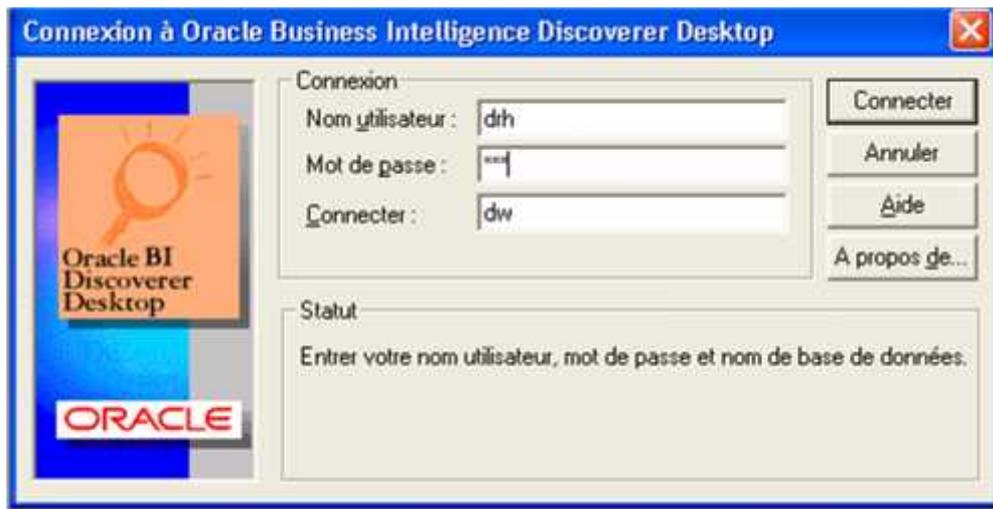
[24] :Daan Van Beck, Norman Manley, The ETL product survey 2007, A passionned International research paper, 2007.

[25] : Cécile Favre; «*Évolution de schémas dans les entrepôts de données*»; Thèse de doctorat ; Université Lumière Lyon 2 «École Doctorale Informatique et Information pour la Société» ; Décembre 2007.

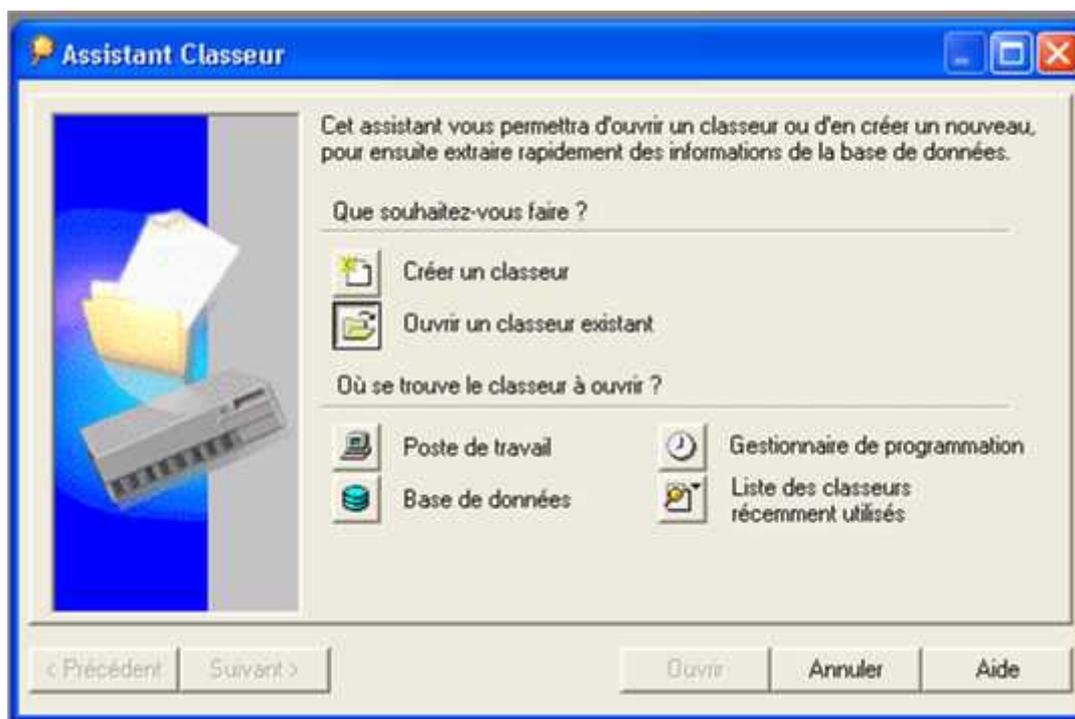
[26] : Ahmed Haciane ; « *Conception d'un datawarehouse Orienté CRM* »; Thèse de magistère Option : SI ; Institut National de Formation en Informatique (I.N.I); 2006.]

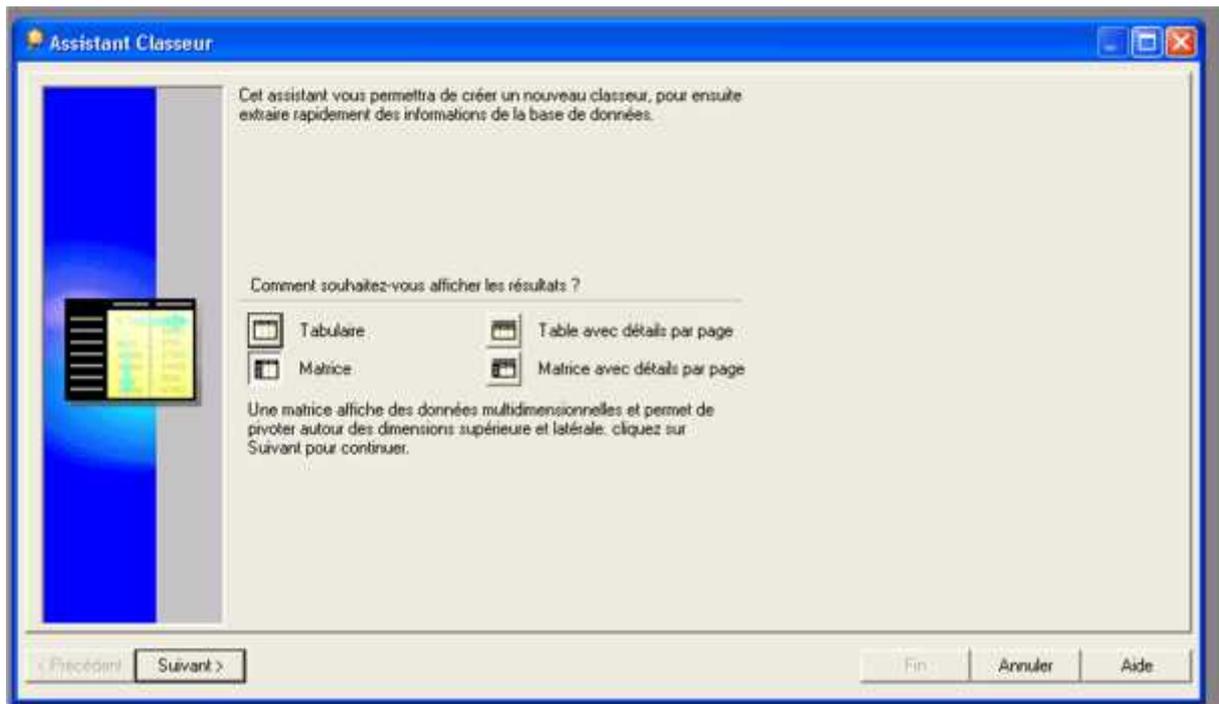
[27] :article : Veille stratégique : Mise en œuvre et valorisation de la veille stratégique en entreprise, Alger, Algérie, 2005

L'Outil Discoverer



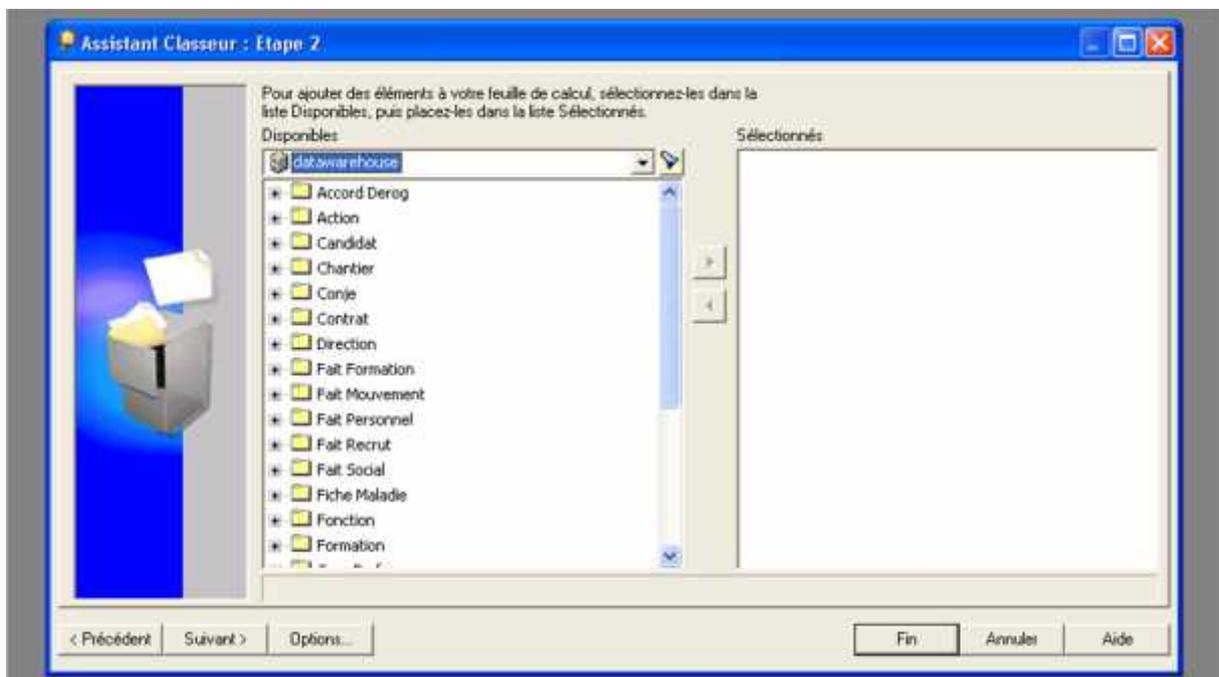
Cet outil nous permet de créer des classeurs pour afficher notre résultats, Il faut d'abord Créer un classeur puis choisir sous quelle forme on veut afficher notre classeur.



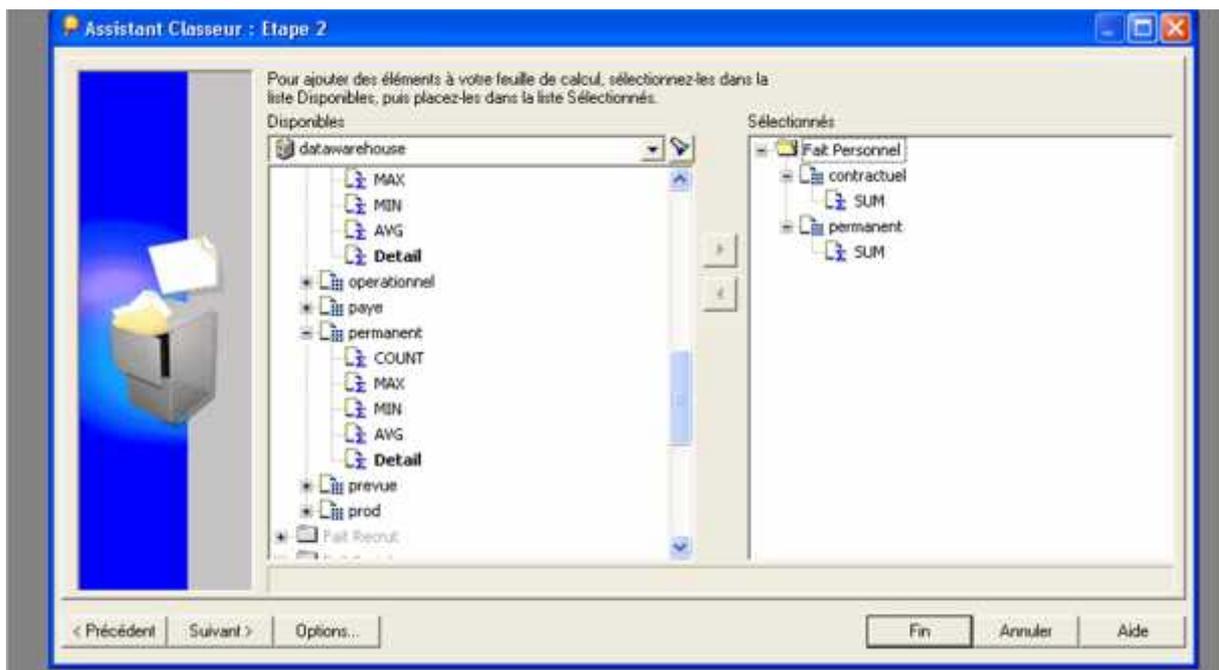
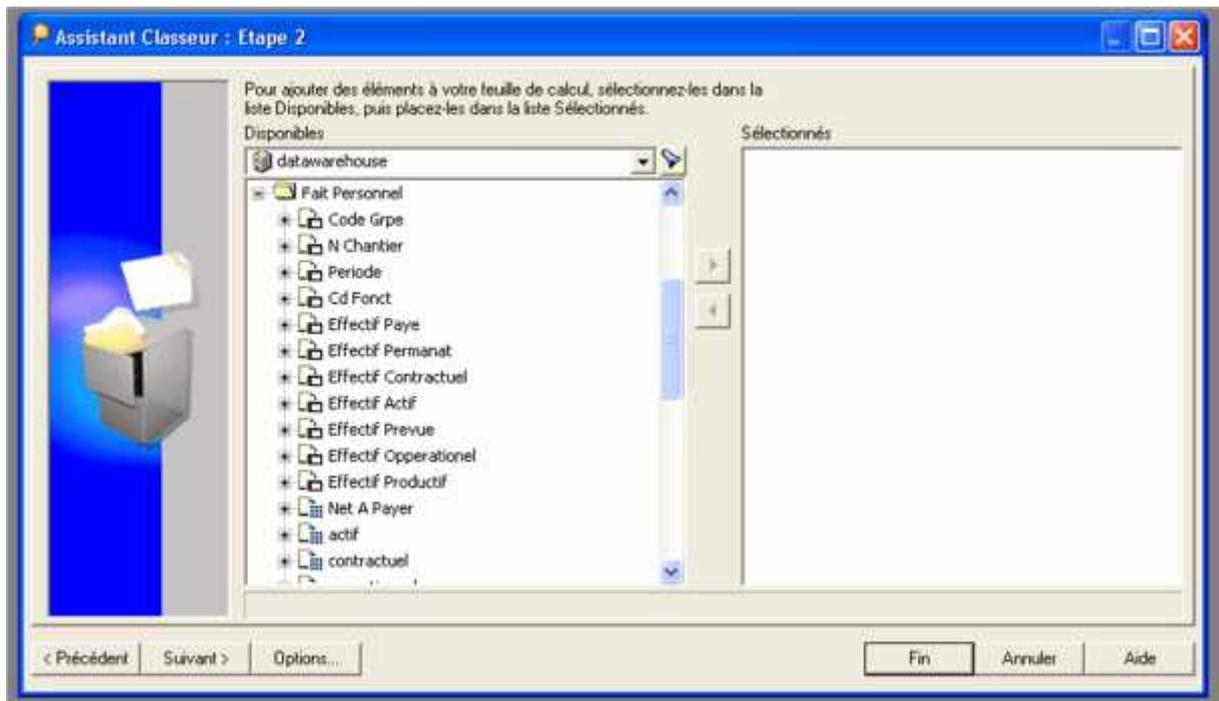


Après avoir choisi le type de classeur, on va ensuite sélectionner les éléments de ma feuille de rapport qu'on veut analyser.

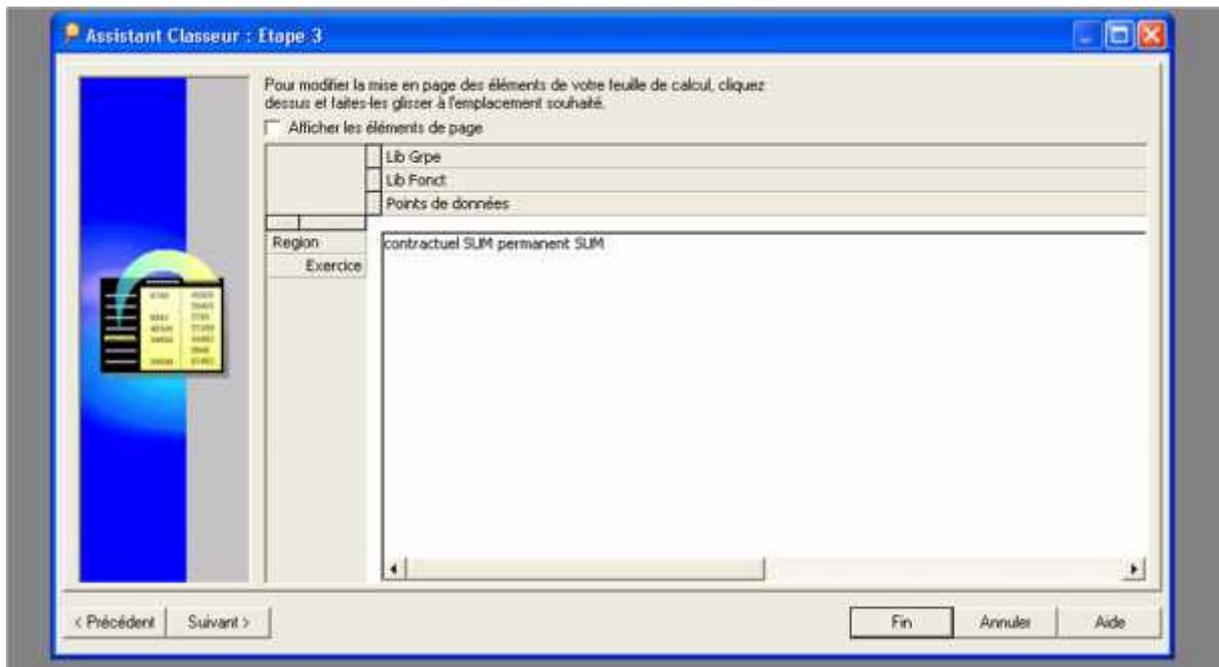
Choisir d'abord le fait sur lequel on va faire notre analyse.



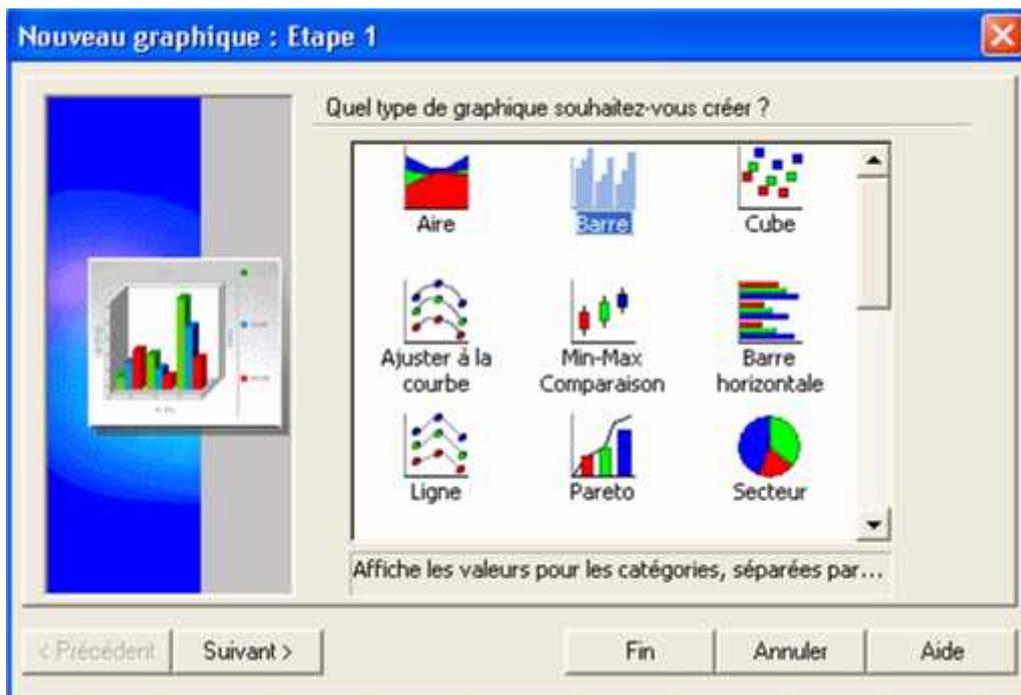
Puis choisir l'ensemble des indicateurs.



Ensuite pour choisir l'axe d'analyse, on peut choisir sur quel axe on base notre analyse.



Ensuite, il faut choisir le type de Graphique.



En fin l'exécution va donner un tableau claire (feuille d'analyse), on peut rajouter des calculs, pourcentage, des cumuls, rajouter d'autre axes d'analyse, ou bien modifier l'ordre... etc

On peut manipuler et changer l'axe d'analyse dans le tableau et faire des Drill-down et Roll-up et des pivoting

Oracle Business Intelligence Discoverer Desktop - [Clicieurt]

Entrepriee Nationale de Grands Travaux Pétroliers

DATA WAREHOUSE

REPARTITION DU PERSONNEL ENTREPRISE PAR AFFECTATION (REGION)

Eléments Page : Periode: 2017/03

	ARZEW		ECOLE DE SONDRY		HASSI MESSAOUD		HASSI RMEL		SIEGE		SKIKDA	
	contractuel	permanent	contractuel	permanent	contractuel	permanent	contractuel	permanent	contractuel	permanent	contractuel	permanent
CADRES	100	190			1730	3270	1030	1890	100	190	150	300
EXECUTIONS	105	210			830	1340	830	1290	235	390	240	360
INGENIEURS	35	80	25	40	540	1580	585	1240	300	690	130	310
MAITRISES	450	720	25	40	4925	7233	1590	2703	460	1330	300	530
TECH. SUPERIEURS	1064	1512	310	532	4301	6854	2305	3814	4222	6040	1060	1659
	1754	2712	360	612	12326	20277	6340	10967	5317	8640	1080	3179

Pivoting :

Eléments Page : Periode: 2017/02 Grpe prof: CADRES

	contractuel	permanent
SKIKDA	135	270
SIEGE	1465	2920
HASSI RMEL	1205	2170
HASSI MESSAOUD	1670	3160
ARZEW	100	190
	4575	8710
	4575	8710

Eléments Page : Grpe prof: CADRES

	2017/02		2017/03		2017/04	
	contractuel	permanent	contractuel	permanent	contractuel	pern
SKIKDA	135	270	150	300	135	2
SIEGE	1465	2920	100	190	100	1
HASSI RMEL	1205	2170	1030	1890	975	11
HASSI MESSAOUD	1670	3160	1730	3270	1600	31
ARZEW	100	190	100	190	85	1
	4575	8710	3110	5840	2895	5

Roll-up :

Eléments Page : Grpe prof: CADRES		
	2017	
	contractuel	permanent
SKIKDA	420	840
SIEGE	1665	3300
HASSI RMEL	3210	5860
HASSI MESSAOUD	5000	9460
ARZEW	285	540
	10580	20000

Drill-down (de région vers chantier) :

REPARTITION DU PERSONNEL					
Eléments Page : Grpe prof: CADRES					
	2017/02		2017/03		contr
	contractuel	permanent	contractuel	permanent	
SKIKDA	135	270	150	300	1
SIEGE	1465	2920	100	190	1
8006	25	40	25	40	2
8007	25	40	25	40	2
8011			25	40	2
9005	5	20			
9016			10	40	1
9081	15	30	15	30	1
9086	0	0	0	0	1
9096	1395	2790			
	1465	2920	100	190	1
HASSI RMEL	1205	2170	1030	1890	9
HASSI MESSAOUD	1670	3160	1730	3270	16

Voilà une représentation par graphe (plusieurs type de graphe, selon le désire de l'utilisateur) :

