

**MINISTRE DES ENSEIGNEMENTS SECONDAIRE,
SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE**

**UNIVERSITE POLYTECHNIQUE
DE BOBO-DIOULASSO**

**ECOLE SUPERIEURE D'INFORMATIQUE
01 BP 1091, Tél.(226) 97 27 64
BOBO-DIOULASSO**

**BURKINA FASO
UNITE-PROGRES-JUSTICE**



**Centre National de Traitement
de l'Information
BP 1132 Ouagadougou 01
e. mail : info@cenatrin.bf**

Mise en place d'un data mart concernant la paie du personnel de l'Etat

MEMOIRE DE FIN D'ETUDE

présenté et soutenu publiquement pour l'obtention du

Diplôme d'ingénieur de conception en informatique

par

W. Stanislas KABORE

Maitre de stage
Joachim OUEDRAOGO

Directeur de memoire
Karim TRAORE

Table des matières

REMERCIEMENTS	3
INTRODUCTION	4
PARTIE I. : ETAT DES LIEUX	5
1.1. La problématique du thème.....	6
1.2. Le DPS 7000/250.....	7
1.3. Le GCOS7.....	10
1.4. Le système opérationnel actuel.....	11
PARTIE II. : METHODES ET CONCEPTS	13
2.1. Le data warehouse ou entrepôt de données.....	14
2.1.1. Présentation.....	14
2.1.2. Pourquoi un entrepôt de données.....	14
2.1.3. Définition	17
2.1.4. Concepts de base.....	19
2.1.5. Infocentre et Data Warehouse.....	21
2.1.6. Composants d'un entrepôt de données	23
2.1.7. Matrice de bus de l'entrepôt	26
2.1.8. Infrastructures	27
2.1.9. Conception d'un entrepôt de données.....	28
2.1.10. Administration des données.....	31
2.2. Le Data Mart	34
2.2.1. Définition	34
2.2.2. Mise en place	35
2.3. Le concept OLAP	37
2.3.1. Présentation.....	37
2.3.2. Les 12 règles OLAP.....	37
2.3.3. Composants d'un système OLAP	40
2.3.4. Offre du marché	43
2.4. La modélisation dimensionnelle	44
2.4.1. Définitions.....	44
2.4.2. Processus de modélisation dimensionnelle.....	45
2.4.3. Dimensions	46
2.4.4. Tables des faits.....	49
2.4.5. Clés artificielles	51
2.4.6. Mythe de la modélisation dimensionnelle	51
2.4.7. Modélisation	52
2.4.8. Gestion du projet de modélisation dimensionnelle.....	53
PARTIE III. : APPLICATION	56
3.1. Schéma global des composants du data mart.....	57
3.2. La phase Extraction Traitement Chargement (ETC)	58
3.2.1. Extraction.....	58
3.2.2. La phase de traitement et de chargement.....	59
3.3. Le modèle dimensionnel	60
3.3.1. Matrice de l'architecture en bus décisionnel.....	60
3.3.2. Diagramme de la table des faits	61
3.3.3. Détail de la table des faits	62
3.3.4. Détail de la table dimensionnelle.....	63

3.4.	Le modèle physique en étoile.....	75
3.5.	Le modèle multidimensionnel.....	76
3.6.	Les outils d'accès.....	77
3.7.	Le scénario de mise en œuvre.....	77
CONCLUSION.....		80
ANNEXE I : PRESENTATION DU CENATRIN.....		81
ANNEXE II : ABREVIATIONS.....		85
ANNEXE III : BIBLIOGRAPHIE.....		86
ANNEXE IV : RESSOURCES INTERNET.....		87
ANNEXE V : GLOSSAIRE.....		88
ANNEXE VI : LISTE DES TABLEAUX ET SCHEMAS.....		90

Remerciements

Qu'il me soit permis ici de remercier très sincèrement les dirigeants du Centre National de Traitement de l'Information (CENATRIN), notamment :

- Toute la Direction Générale du CENATRIN, en particulier le Directeur Général, pour l'opportunité qui m'a été donnée de pouvoir m'inscrire à ce cycle et de pouvoir le terminer ;
- Le Chef de la Division Etudes et Développement, pour la supervision de mes travaux, les conditions de travail agréables, son soutien et ses conseils ;
- Tous les chefs de Divisions et tout le personnel du CENATRIN ;

Je remercie également tous ceux qui, de près ou de loin, ont contribué à ma formation. Je citerai surtout :

- Tout le corps professoral de l'Ecole Supérieure d'Informatique pour la qualité des enseignements dispensés ;
- Tous les enseignants missionnaires pour leur disponibilité et leur contribution considérable à notre formation ;

Je n'oublierai pas les membres de ma famille, mes parents et mes amis pour leur soutien multiforme.

Introduction

L'Ecole Supérieure d'Informatique (ESI) a vu le jour en 1990 avec l'ouverture du Cycle des Ingénieurs de Travaux en Informatique (CITI) option : analyste programmeur. Cela au vue de la nécessité de former des cadres compétents en informatique au niveau national.

C'est dans cette même optique que l'Ecole Supérieure d'Informatique a ouvert en 1998 le Cycle de formation des Ingénieurs de Conception en Informatique (CICI). La formation théorique reçue est complétée par les études de cas, les travaux pratiques et les stages en entreprises. En dernière année du Cycle de formation des Ingénieurs en Conception, les élèves ingénieurs effectuent un stage de six (6) mois en entreprise sanctionné par la présentation d'un mémoire.

C'est dans ce cadre que nous avons effectué notre stage au Centre National de Traitement de l'Information (CENATRIN), qui est mis en liquidation administrative depuis juillet 2005. Le stage a été supervisé par Monsieur Karim TRAORE, enseignant à l'ESI et Monsieur Joachim OUEDRAOGO, chef de projet, Chef de la division Etudes et Développement du CENATRIN. Ce mémoire a pour thème « Mise en place d'un data mart concernant la paie du personnel de l'Etat », et vise à apporter une solution aux problèmes de manque d'états décisionnels auxquels sont confrontées les structures en charge du traitement des salaires des agents de l'Etat.

Ce document présente en trois (3) parties les résultats des travaux effectués durant ce stage :

- la première partie pose la problématique du sujet et fait un état des lieux ;
- la deuxième partie présente les méthodes et outils de conception et d'analyse multidimensionnelle ;
- la troisième partie propose une application des concepts et méthodes pour la mise en place d'un data mart dans le cadre de la paie du personnel de l'Etat.

Ces différentes parties sont complétées par une liste de références bibliographies et de ressources internet exploitées au cours de ce stage.

PARTIE I:
Etat des lieux

1.1. La problématique du thème

Les services en charge de la gestion de la solde des agents de l'Etat burkinabé sont régulièrement confrontés à des problèmes de non disponibilité de certains types d'états. La présence de ces états pourrait les aider dans les travaux quotidiens ainsi que dans certaines prises de décision. Le système existant est très pauvre en états de synthèse. Les responsables sont obligés de recourir aux informaticiens à chaque fois pour demander des états spécifiques non prévus à l'origine de l'application. La plupart du temps, ils sont obligés de ressaisir les données de quelques états qu'ils ont (par l'intermédiaire des informaticiens ou certains états de solde de fin du mois) sur Excel pour pouvoir les traiter et les analyser. Par exemple les états comme :

- l'évolution de la masse salariale sur les trois, cinq ou même dix dernières années ;
- la répartition annuelle de la masse salariale par province, département, localité sur une période ;
- la répartition d'une indemnité quelconque par section ou par province ;
- l'évolution d'un élément de salaire sur une période donnée ;
- la répartition d'un élément de salaire par catégorie et sur une période donnée ;

ne peuvent pas être faits avec les systèmes actuels car ils n'ont pas été conçus dans une optique d'aide à la décision.

Suite à l'arrêt du gros système le BULL DPS7000/250, il était devenu nécessaire de récupérer les informations sur les bulletins des agents de l'Etat dont la paye se faisait sur ce système jusqu'en juillet 1999 pour la majorité. Ces informations étaient capitales car elles permettaient de mettre à la disposition des agents des duplicatas de bulletin. Ces bulletins venaient en complément des pièces constitutives de dossier administratif pour les corrections de salaire au niveau de la solde, pour les dossiers de pension au niveau des caisses de retraite : la caisse

nationale de sécurité sociale (CNSS) et la caisse autonome de retraite des fonctionnaires (CARFO), et aussi des dossiers au niveau du Ministère de la Fonction Publique.

Dans un premier temps, notre travail a consisté à résoudre le problème de la récupération des informations sur les bulletins des agents de l'Etat. Ce travail consistait à réaliser sur micro-ordinateur des programmes de transfert des données se trouvant sur le DPS 7000/250 et l'application de consultation et d'édition des bulletins de paie.

L'application devrait permettre d'avoir accès aux données sur une période allant de 1994 à 2000.

Ces données en fichier texte plat représentent une taille de 2,16Go.

Par la suite, nous avons pensé qu'il était opportun de profiter des avancées technologiques pour mettre à la disposition des responsables de la solde toute la manne d'informations qu'ils avaient à travers l'ancien système de paie qui était au CENATRIN et le système actuel. Ainsi les données du DPS 7000/250, couplées à celles du même genre produites par le système opérationnel actuel SIGASPE (Système Intégré de Gestion Administrative et Salariale du Personnel de l'Etat), pourraient être orientées pour un sujet « les éléments de salaire payés », intégrées, historisées et organisées pour aider les responsables dans la prise de décision : d'où la mise en place d'un data mart. Le data mart devra permettre aux utilisateurs d'avoir des états de synthèse, des statistiques et aussi de pouvoir à partir d'échantillons assez représentatifs faire des projections, des prévisions.

Pour le data mart, en plus des 2,16 Go de données du DPS 7000/250, il y a environ 3Go de données sous format dmp (fichier export d'une base oracle) provenant du système opérationnel actuel SIGASPE.

1.2. Le DPS 7000/250

Le DPS 7000/250 se classe dans la famille des gros systèmes. Il a été installé au CENATRIN en Mai 1990 en remplacement du BULL 64 DPS 2.

Le BULL 64 DPS 2 était un système jumeau couplé : système qui partageait dynamiquement et statiquement des ressources. Ce système en exploitation depuis plus de dix (10) ans ne présentait aucune garantie, aussi il a fallu procéder à son remplacement.

Le DPS 7000/250 était un système :

- Compact : haute intégration grâce à VLSI/CMOS (Very Large Scale Integration / Complementary Metal Oxyde Semiconductor). Ceci à l'opposé du BULL 64 DPS 2 ;
- Puissant : 5,5 à 21 transactions de type débit/crédit par seconde (transaction débit/crédit définie par DATAMATION avril 1985). Il était cinq (5) fois plus puissant que le BULL64 ;
- Fonctionnant en environnement de bureau : pas de climatisation spécifique, pas de faux plancher, tension de 220 volts, consommation électrique faible ;
- Ayant des coûts d'exploitation réduits : espace au sol très réduit, fonctionnement sans opérateur, télémaintenance et télécontrôle ;
- Avec intégration des réseaux locaux : support des réseaux locaux Ethernet et Starlan.

Le DPS 7000/250 a un processeur principal ou central et des processeurs spécialisés. Le processeur central gère l'initialisation du système, les tests de diagnostics (On line/Off line), l'accès à la maintenance préventive et corrective, la télémaintenance, la gestion de la console de service, les échanges de GCOS7 avec le système et les différents processeurs spécialisés. A leur tour, les processeurs spécialisés avaient chacun un certain nombre de tâches spécifiques : exemple un processeur de disques pour les disques, un processeur de bandes pour les bandes, un processeur de réseau.

Le DPS 7000/250 possède :

- sept (7) disques de 1Go chacun avec un débit de 2,45Mo/s et un temps d'accès de 20 ms. Les disques étaient organisés comme suit : un (1)

disque système, cinq (5) disques de données et un (1) disque de secours. Contrairement au BULL64, les disques du DPS 7000/250 n'étaient pas amovibles ;

- deux (2) imprimantes à impact dont une de 750 lignes par minute et l'autre de 1200 lignes par minute ;
- Quatre (4) dérouleurs de bandes ;
- Un (1) micro-ordinateur BULL Micral 200 de 20 Mo de disque dur, 4 Mo de mémoire, un processeur 80-286. Ce micro ordinateur tournait sous DOS 3.1. Il était connecté au DPS 7000 et permettait de lire et/ou d'écrire sur les disquettes 8' et 3'5. Il assurait le transfert des fichiers dans les deux sens ;
- Des terminaux locaux au CENATRIN et des terminaux distants au Trésor, à la solde, à la direction du matériel... Les terminaux distants étaient connectés au DPS7000/250 par une liaison X25 (FASOPAC) de l'ONATEL (Office National des Télécommunications) avec un débit de 64 Ko/s. Le DPS7000/250 pouvait gérer 130 terminaux actifs.
- Une mémoire de 256 Mo étendue grâce aux techniques de mémoire virtuelle BKST (backing store). Cette mémoire était répartie entre les différentes classes de passage des travaux configurées à la personnalisation du système.

Le DPS 7000/250 effectuait essentiellement les traitements des applications de l'Etat : solde, impôts, dépense de matériel..., et celles des structures privées.

Tout cela pour dire que le DPS 7000 regorge d'une mine d'informations qu'il serait dommage de laisser disparaître.

Le DPS 7000/250 tournait sous GCOS7 (système d'exploitation pour les ordinateurs centraux BULL HONEYWELL DPS 7000) version 3.

En comparaison, le DPS 7000/250 est de la gamme de l'AS/400 B60 d'IBM.

1.3. Le GCOS7

Le GCOS7 (General Comprehensive Operating System) se compose de l'Operating system et des utilitaires auxquels se greffent les applications.

L'Operating system se compose:

- du noyau ;
- des jobs de service (reader, telecom, writers);

Ces deux (noyau et jobs de service) réalisent la gestion globale, gestion des jobs, gestion des fichiers et volumes, gestion des messages. Ceci en réagissant aux événements extérieurs, aux ordres GCL (GCOS Command Language) à la main, aux ordres JCL (Job Control Language) en batch/IOF, aux commandes et directives en IOF interactif.

- des fonctions système partageables (qui sont soit en type 0, soit en sharables modules) sont appelables par tous et réalisent des fonctions d'intérêt public.

Les utilitaires se composent :

- des utilitaires entrant dans le développement de programme (programming facilities) :
 - ✓ Gestionnaires de bibliothèques ;
 - ✓ Editeur, full screen editor ;
 - ✓ Compilateurs.
- d'autres utilitaires (dependant components) :
 - ✓ data management : au niveau fichier, volume, base de données, groupe de fichiers ;
 - ✓ développement d'applications transactionnelles : moniteur TDS, JCL de préparation et de génération d'une application, définition de grilles FORMGEN, RTP d'utilisation de grilles ;
 - ✓ gestion du système : analyse de l'exploitation, simulation de la charge, droits d'accès au site, gestion des sécurités, catalogage des

fichiers, journalisation, gestion de l'exploitation du site, installation et suivi, traces d'exploitation.

Les principales fonctions de GCOS7 sont :

- la gestion des jobs : job flow
Un travail depuis son introduction jusqu'à l'édition de tous ses états est connu du système par un numéro RON (Run Occurrence Number), indice dans la table KJOB (Knowned Job). Il transite par différents états ou phases.
- la gestion des sécurités
La sécurité au niveau des travaux et la surveillance du système.
- l'extended machine
Gestion de la mémoire (centrale, de réserve, fichiers virtuels), des évènements (gestion des entrées/sorties), du temps (gestion de l'heure et de la date).
- la gestion des données
Cela passe par la gestion des fichiers, des volumes, utilisation de fichiers, contrôle de périphériques.
- le logiciel télécom
- le distributed operation facility (DOF).

1.4. Le système opérationnel actuel

Le système opérationnel actuel est le SIGASPE (Système Intégré de Gestion Administrative et Salariale du Personnel de l'Etat). Le système permet :

- la gestion administrative des agents publics de l'Etat burkinabé par le ministère de la fonction publique et les directions des ressources humaines des différents ministères ;
- la gestion des salaires par le Ministère des Finances et du Budget à travers la Direction Générale du Budget, la Direction Générale du Trésor

et de la Comptabilité Publique, la Direction du Contrôle Financier, et les directions financières des différents ministères.

Le système tourne actuellement sur un serveur IBM Netfinity 8500r avec :

- deux (2) processeurs Xéon P III 550 Mhz avec possibilité d'extension à quatre (4) processeurs ;
- une mémoire RAM de 1Go extensible ;
- cinq (5) disques durs de 9,10 Go chacun ;
- un streamer interne 20/40 Go ;
- le système d'exploitation est Windows 2003 server.

L'application a été développée sous Developer2000 (forms 4.5 et report 2.5) avec une base de données sous Oracle 9i.

Les utilisateurs du système sont dans les différents bâtiments de l'administration à travers toute la ville de Ouagadougou. Ces différents bâtiments sont interconnectés entre eux à travers le réseau interministériel de l'administration (RESINA).

Les postes clients sont des micro-ordinateurs avec des systèmes d'exploitation différents et des configurations variables.

PARTIE II :
Méthodes et concepts

2.1. Le data warehouse ou entrepôt de données

2.1.1. Présentation

Le concept d'entrepôt de données a été formalisé pour la première fois en 1990. L'idée de constituer une base de données orientée sujet, intégrée, contenant des informations datées, non volatiles et exclusivement destinées aux processus d'aide à la décision fut dans un premier temps accueillie avec une certaine hésitation. Beaucoup n'y voyaient qu'une autre forme du concept déjà ancien : l'infocentre.

L'entreprise doit anticiper pour faire face aux nouveaux enjeux économiques. Pour être efficace, l'anticipation peut s'appuyer sur de l'information pertinente qui est à la portée de toute entreprise qui dispose d'un capital de données gérées par ses systèmes opérationnels et qui peut en acquérir d'autres auprès de fournisseurs externes. Mais ces données ne sont pas organisées dans une perspective décisionnelle et sont éparpillées dans plusieurs systèmes hétérogènes. Il est nécessaire de rassembler et d'homogénéiser les données afin de permettre des analyses des indicateurs pertinents et de faciliter les prises de décisions.

Pour répondre à ces besoins, il a été défini et intégré une architecture qui va servir de fondation aux applications décisionnelles : l'entrepôt de données.

2.1.2. Pourquoi un entrepôt de données

2.1.2.1. La problématique des entreprises

L'entreprise construit un système décisionnel pour améliorer sa performance. Elle doit décider et anticiper en fonction de l'information disponible et capitaliser sur ses expériences.

Depuis plusieurs dizaines d'années, une importante masse d'informations est stockée sous forme informatique dans les entreprises. Les systèmes d'informations

sont destinés à garder la trace d'événements de manière fiable et intègre. Ils automatisent de plus en plus les processus opérationnels.

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

2.1.2.2. La réalité des systèmes d'information

Les données contenues dans les systèmes d'informations opérationnels sont :

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

Le tableau ci-dessous présente les différences entre les données opérationnelles et décisionnelles.

Données opérationnelles	Données décisionnelles
Orientées application, détaillées, précises au moment de l'accès	Orientées activité (thème, sujet), condensées, représentent 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 unitaire 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	Utilisées de façon aléatoire

Tableau 1: Différences entre données du système de production et données décisionnelles

S'il existe effectivement des informations importantes, il est nécessaire de construire une structure pour les héberger, les organiser et les restituer à des fins d'analyse. Cette structure est le Data Warehouse ou "entrepôt de données".

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

2.1.2.3. Les objectifs

Toutes les données, qu'elles proviennent du système de production de l'entreprise ou qu'elles soient achetées vont devoir être organisées, coordonnées, intégrées et stockées, pour donner à l'utilisateur une vue intégrée et orientée métier. L'entrepôt de données doit viser les objectifs suivants :

- L'entrepôt de données doit rendre les données de l'organisation facilement accessibles.

Le contenu de l'entrepôt de données doit être facile à comprendre. Les données doivent être parlantes et leur signification évidente pour l'utilisateur et pas seulement pour le développeur.

- L'entrepôt de données doit présenter l'information de l'organisation de manière cohérente.

Les données de l'entrepôt doivent être crédibles.

- L'entrepôt de données doit être adaptable et résistant aux changements.

Les données de l'entrepôt devront être conçues pour traiter les changements.

De ce fait, les changements ne doivent pas invalider les données existantes ou les applications.

- L'entrepôt de données doit être le socle sur lequel repose l'amélioration des prises de décision.
- L'entrepôt de données doit être accepté par les utilisateurs pour pouvoir réussir.

2.1.3. Définition

Selon Bill Inmon (1996) :

« Le data warehouse est une collection de données orientées pour un sujet, intégrées, non volatiles et historisées, organisées pour le support du processus d'aide à la décision».

De cette définition, il apparait les caractéristiques suivantes d'un data warehouse :

- Orientées sujet

L'entrepôt de données est organisé autour des sujets majeurs de l'entreprise, contrairement aux données des systèmes de production qui sont généralement organisées par processus fonctionnels. Les données sont structurées par thème. L'intérêt de cette organisation est de disposer de l'ensemble des informations utiles sur un sujet le plus souvent transversal aux structures fonctionnelles et organisationnelles de l'entreprise.

Cette orientation sujet va également permettre de développer son système décisionnel via une approche par itérations successives, sujet après sujet.

- Données intégrées

Un entrepôt de données est un projet d'entreprise. Pour y parvenir, les données doivent être intégrées. Avant d'être intégrées dans l'entrepôt de données, les données doivent être mises en forme et unifiées afin d'avoir un état cohérent. Une donnée doit avoir une description et un codage unique.

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

- Données historisées

Dans un système de production, la donnée est mise à jour à chaque nouvelle transaction. Dans un entrepôt de données, la donnée ne doit jamais être mise à jour. Un référentiel temps doit être associé à la donnée afin d'être capable d'identifier une valeur particulière dans le temps.

- Données non volatiles

La non volatilité des données est en quelque sorte une conséquence de l'historisation. Une même requête effectuée à quelques mois d'intervalle en

précisant la date de référence de l'information recherchée donnera le même résultat.

Le tableau ci-après présente les principales différences entre le système de production et l'entrepôt de données:

Critère	Système de production	Entrepôt de données
Niveau de détail des informations "utilisateurs"	Très détaillé	Synthétique, parfois détaillé
Utilisateurs	Une ou quelques fonctions de l'entreprise	Plusieurs fonctions de l'entreprise
Données figées	Non – évolution en temps réel	Oui – archivage
Historique	Non	Oui
Opérations sur les données	Ajout/mise à jour/ consultation	Consultation uniquement

Tableau 2: différences entre système de production et entrepôt de données

2.1.4. Concepts de base

2.1.4.1. La structure

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

- **Les données détaillées**

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 l'entrepôt de données n'est pas forcément le même que celui géré dans les systèmes opérationnels.

- 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.

- Les métadonnées

Elles regroupent l'ensemble des informations concernant l'entrepôt de données et les processus associés. Elles constituent une véritable aide en ligne permettant de connaître l'information contenue dans l'entrepôt de données. Elles sont 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 l'entrepôt de données (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 l'entrepôt de données).
- Aux équipes de production (procédures de changement, historique de mise à jour,...).

- Les données historisées

Un des objectifs de l'entrepôt de données 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.

2.1.4.2. L'architecture

Le stockage des données est réalisé dans un SGBD séparé du système de production. Le SGBD est alimenté par des extractions périodiques. Avant le chargement, les données subissent d'importants processus d'intégration, de nettoyage, de transformation.

L'avantage est de disposer de données préparées pour les besoins de la décision et répondant aux objectifs de l'entrepôt de données.

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

2.1.5. Infocentre et Data Warehouse

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

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

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

Dans un infocentre, chaque nouvelle valeur remplace l'ancienne valeur. Il est donc impossible de retrouver une valeur calculée dans une session préalable aux dernières alimentations. La non volatilité est une caractéristique essentielle de l'entrepôt de données.

De même, dans un infocentre, il n’y a pas de gestion d’historique des valeurs. L’infocentre sert à prendre des décisions opérationnelles basées sur des valeurs courantes. Au niveau d’un entrepôt de données, l’utilisateur travaille sur les historiques pour des prises de décisions à long terme, des positionnements stratégiques et pour analyser des tendances.

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

Le tableau ci-dessous fait une comparaison entre l’infocentre et l’entrepôt de données.

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

Tableau 3: tableau de comparaison entre infocentre et entrepôt de données

2.1.6. Composants d'un entrepôt de données

2.1.6.1. Les applications opérationnelles sources

Ce sont les applications opérationnelles qui capturent les transactions de l'organisation. Les applications sources ne conservent que très peu de données historisées et si vous avez un bon entrepôt de données, il peut libérer les applications sources d'une bonne partie de leurs responsabilités concernant la représentation du passé.

2.1.6.2. La préparation de données

La zone de préparation des données de l'entrepôt est à la fois une zone de stockage et un ensemble de processus couramment appelés ETC (Extraction/Transformation/Chargement). L'extraction est la première étape du processus d'apport de données à l'entrepôt qui se traduit par la lecture, l'interprétation et la copie des données sources dans la zone de préparation. Ensuite on passe à la transformation en vue du chargement. Il faut interdire aux utilisateurs l'accès à la zone de préparation des données.

2.1.6.3. La présentation de données

La zone de présentation des données est le lieu où les données sont organisées, stockées et offertes aux requêtes directes des utilisateurs, aux programmes de reporting et aux autres applications d'analyse. La zone de présentation des données est l'entrepôt de données tel que vu par les utilisateurs.

2.1.6.4. Les outils d'accès aux données

L'ensemble des outils d'accès aux données constitue le dernier composant majeur d'un environnement d'entrepôt de données. Les outils d'accès aux données constituent l'ensemble des moyens fournis aux utilisateurs pour exploiter la zone de

présentation en vue de prendre des décisions basées sur des analyses. Un outil d'accès aux données peut être une chose aussi simple qu'un outil de requête ou aussi complexe qu'une application de forage de données ou de modélisation.

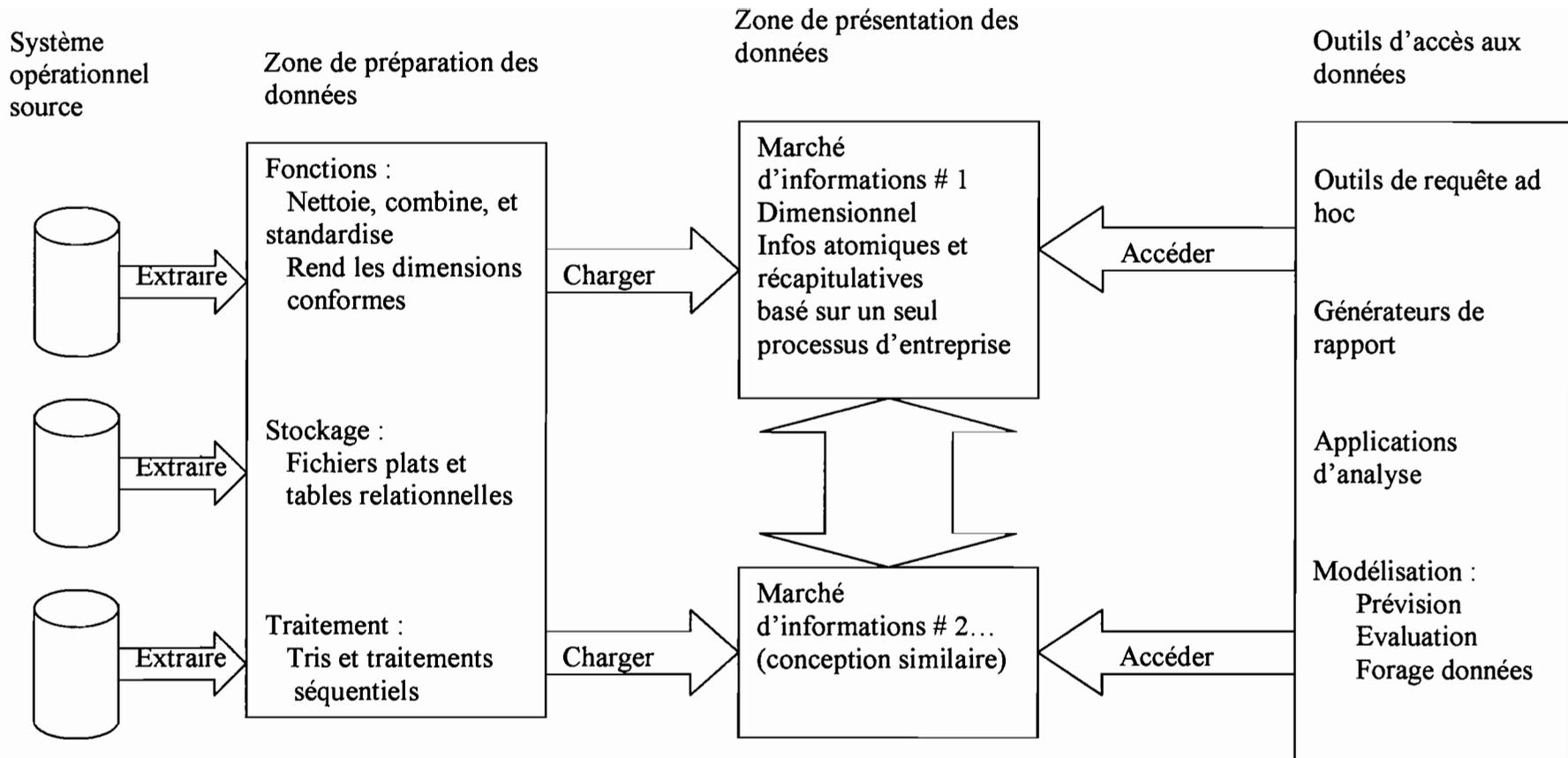


Schéma 1 : Composants de base de l'entrepôt de données

2.1.7. Matrice de bus de l'entrepôt

La matrice de bus de l'entrepôt de données permet la représentation des processus de l'entreprise sous forme de lignes et les dimensions sous forme de colonnes. Une ligne permet de voir d'un coup les dimensions du processus. La colonne décrit l'interaction entre les processus et les dimensions communes. La matrice est aussi utile comme outil de planning et de communication. Le fait de remplir les lignes et les colonnes aboutit à définir l'architecture d'ensemble de l'entrepôt de données. La matrice sert également à la communication au sein des équipes de développements. Ce document succinct présente la totalité du plan. Il est adapté aux communications avec les décideurs en matière d'informatique de gestion de l'entreprise. C'est un document très important parmi tous ceux dont la préparation doit précéder toute réalisation d'un entrepôt de données.

Exemple de matrice de bus des assurances

Processus d'entreprise	Dimensions communes								
	Date	Assuré	Risque couvert	Bien assuré	Employé	Police	Sinistre	Déclarant	Tiers
Transactions de police	x	x	x	x	x	x			
Instantané des primes	x	x	x	x	x	x			
Transactions de sinistre	x	x	x	x	x	x	x	x	x

Tableau 4: matrice de bus des assurances

2.1.8. Infrastructures

Il y a deux niveaux d'infrastructures :

- l'infrastructure technique : l'ensemble des composants matériels et logiciels (à associer aux composants fonctionnels : alimentation, stockage et accès).
- l'infrastructure opérationnelle : l'ensemble des procédures et des services pour administrer les données, gérer les utilisateurs et exploiter le système.

Pour la construction d'un entrepôt de données, de manière générale, il faut :

1. Bien connaître les métiers utilisateurs

- Impliquer les utilisateurs dans les projets,
- Faire participer l'utilisateur à la définition et à l'évolution des métadonnées,
- Former l'utilisateur pour l'aider à comprendre la logique du Data Warehouse,
- Avoir un chef de projet orienté utilisateur.

2. Assurer une véritable conduite de projet

Le chef de projet fédère l'ensemble des besoins. Il organise et planifie les différentes phases de déploiement en fonction de la culture de l'entreprise.

3. Commencer petit et voir grand

Il faut commencer petit, qu'il s'agisse d'un prototype ou d'un Data Mart.

4. Gérer l'évolutivité

Alimentation et administration sont les points critiques du processus de construction de l'entrepôt de données.

2.1.9. Conception d'un entrepôt de données

Quatre caractéristiques ont des effets déterminants sur la démarche de conception d'un entrepôt de données:

- Les évolutions technologiques : un système d'information peut se construire par intégration d'un certain nombre de composants, chacun pouvant être choisi par rapport à son contexte d'utilisation. L'entreprise définit son architecture en fonction de ses besoins.
- La stratégie de l'entreprise : l'entrepôt de données est très proche de la stratégie de l'entreprise. L'objectif de l'entrepôt de données se définit en terme métier. Il faut donc impliquer les utilisateurs ayant le plus de connaissances dans leur entreprise ou dans leur métier.
- L'amélioration continue : un entrepôt de données doit évoluer en fonction des demandes utilisateurs ou des nouveaux objectifs de l'entreprise.
- La maturité de l'entreprise : certaines entreprises ont déjà un système décisionnel, d'autres n'ont aucun acquis.

Dans tous les cas, il n'existe pas de cadre figé pour la conception d'un entrepôt de données. Chaque entreprise doit adapter le projet à son contexte, en ne perdant pas les objectifs de vue.

Il est possible de proposer trois phases pour la conception :

- Définir le pourquoi de l'entrepôt de données et les objectifs à atteindre (impliquer les utilisateurs).

- Définir l'infrastructure technique et organisationnelle de l'entrepôt de données.
- Mettre en œuvre les applications.

2.1.9.1. L'infrastructure

Il s'agit de déterminer l'infrastructure technique et organisationnelle nécessaire à la mise en place de l'entrepôt de données et à la conduite du changement.

L'infrastructure technique

Des choix technologiques en phase avec la politique de l'entreprise doivent être faits à plusieurs niveaux :

- Les fournisseurs : faut-il prendre un seul fournisseur (ce qui facilite la politique d'intégration et en réduit les coûts de mise en œuvre) ou assembler les meilleures offres du marché (ce qui apporte une flexibilité, une adaptation à chaque projet, mais coûte beaucoup en intégration).
- Les outils : faut-il construire, acheter ou faire avec l'existant.
- Comment sera utilisé l'entrepôt de données, par qui, comment sera structurée l'organisation qui l'exploitera.
- Faut-il une architecture centralisée (entrepôt de données), distribuée (plusieurs Data Mart), ou une architecture répliquée (un entrepôt de données et plusieurs Data Mart).
- La structure de stockage, sera-t-elle relationnelle, multidimensionnelle, hybride (Entrepôt de données en relationnel, Data Mart en multidimensionnel).
- Choisir le matériel : selon les volumes envisagés, les utilisateurs concernés, l'architecture visée, la flexibilité attendue.
- Organiser l'administration des systèmes et la gestion de la sécurité.

Il faut vérifier que les solutions choisies fonctionnent entre elles. (Ne pas se fier aux plaquettes commerciales en ce qui concerne la portabilité).

L'infrastructure organisationnelle

Parallèlement aux choix technologiques, il faut :

- Déterminer la logistique et l'organisation nécessaires à la concrétisation des initiatives.
- Répartir les tâches entre les équipes de développement et les équipes d'exploitation : déterminer l'alimentation de l'entrepôt de données, l'administration.
- Déterminer les flux d'information entre l'entrepôt de données et les utilisateurs.

2.1.9.2. La formation

Selon l'expérience de l'entreprise en matière de décisionnel et des technologies utilisées, un plan de formation sera nécessaire.

Il sera impératif que les membres participants au projet soient favorables au changement.

2.1.9.3. La mise en œuvre des applications

La mise en œuvre est réalisée pour chaque initiative. La démarche proposée est une démarche en cinq étapes :

- la spécification,
- la conception,
- la mise en œuvre et l'intégration,
- le déploiement et la mise en place des accompagnements,
- les mesures.

Ces étapes correspondent à celles de mise en place d'un projet informatique.

Pendant l'étape de spécification, les différentes étapes des initiatives sont définies et planifiées de manière plus détaillées. Il est recommandé de faire attention aux coûts cachés que peuvent entraîner les technologies informatiques.

L'étape de mesure permet de faire le bilan de la réalisation et de capitaliser les réussites et échecs rencontrés pendant le développement de l'application.

Il n'existe pas de démarche complète et universelle pour la mise en œuvre d'un entrepôt de données. Toute approche doit être adaptée à l'entreprise.

2.1.10. Administration des données

Comme tout système informatique, un entrepôt de données s'administre. Dès la phase de conception de l'architecture, il faut penser à l'administration des données qui est une des fonctions les plus importantes de l'entrepôt de données. Cette fonction est d'autant plus importante que l'entrepôt de données évolue au fur et à mesure de son utilisation.

La mise en place d'un entrepôt de données doit s'accompagner de celle d'un référentiel de données, permettant de décrire, stocker et diffuser les métadonnées associées.

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

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

2.1.10.1. Le référentiel l'entrepôt de données

Le référentiel de l'entrepôt de données est l'ensemble des outils nécessaires à la mise en œuvre de la fonction d'administration de données. Un référentiel de données pour l'entrepôt de données est un référentiel de données dans lequel sont décrits l'organisation et la localisation des données, ainsi que les règles de

consolidation des données agrégées et historisées. Il est conçu de manière à collecter l'ensemble des modèles de données nécessaires à la construction et à l'exploitation de l'entrepôt de données.

La construction du référentiel est un projet qu'il est nécessaire de prévoir en parallèle à celui de l'entrepôt de données.

2.1.10.2. La sécurité

L'entrepôt de données a pour vocation de laisser à l'utilisateur une totale autonomie en ce qui concerne la recherche et l'analyse des données. Cette liberté doit cependant être souvent restreinte, notamment pour des raisons de sécurité. L'outil doit donc permettre d'adapter l'environnement de travail à l'utilisateur qui s'y connecte selon sa fonction et donc ses droits. Il peut par exemple se connecter à partir d'Excel à un serveur de données, via le driver ODBC.

2.1.10.3. La gestion des performances

La possibilité de limiter les accès aux ressources permet à l'administrateur de mieux contrôler les requêtes émises par l'utilisateur. Ce contrôle des ressources peut prendre plusieurs formes :

- limitation du temps d'exécution ;
- limitation du nombre de lignes récupérées ;

Le contrôle du corps des requêtes permet de limiter considérablement le nombre de demandes coûteuses. Ce contrôle peut s'exercer sur différents paramètres :

- le nombre de tables interrogées ;
- les jointures, l'administrateur peut notamment interdire l'utilisation des produits cartésiens et aussi imposer d'utiliser uniquement les colonnes indexées ou d'interdire un nombre trop important de thêta-jointures, ...

Ce besoin de contrôler les performances est dû à l'autonomie laissée aux utilisateurs par les outils.

2.1.10.4. Les batchs

La possibilité d'effectuer des requêtes en batch est une fonctionnalité essentielle dans un contexte d'entrepôt de données où une requête peut avoir besoin de manipuler un volume important de données. Un planificateur, peut alors permettre d'envoyer une requête à un instant donné ou à intervalles réguliers. Imposer des plages horaires aux utilisateurs pour ces requêtes coûteuses est alors également une nécessité.

2.2. Le Data Mart

2.2.1. Définition

Le Data Mart est une base de données moins coûteuse que l'entrepôt de données, et plus légère puisque destinée à quelques utilisateurs d'un département. C'est une petite structure très ciblée et pilotée par les besoins utilisateurs. Il a la même vocation que l'entrepôt de données (fournir une architecture décisionnelle), mais vise une problématique précise avec un nombre d'utilisateurs plus restreint.

Le tableau ci-après donne des éléments de comparaison entre un data warehouse et un data mart.

	Entrepôt de données	Data Mart
Cible utilisateur	Toute l'entreprise	Département
Implication du service informatique	Elevée	Faible ou moyen
Base de données d'entreprise	SQL type serveur	SQL milieu de gamme, bases multidimensionnelles
Modèles de données	A l'échelle de l'entreprise	Département
Champ applicatif	Multi sujets, neutre	Quelques sujets, spécifique
Sources de données	Multiples	Quelques unes
Stockage	Base de données	Plusieurs bases distribuées
Taille	Centaine de GO et plus	Une à 2 dizaines de GO
Temps de mise en place	9 à 18 mois pour les 3 étapes	6 à 12 mois (installation en plusieurs étapes)

Tableau 5: finalités du data mart et de l'entrepôt de données

2.2.2. Mise en place

Construire un ou plusieurs Data Marts départementaux au lieu d'un entrepôt de données central permet de valider rapidement le concept d'informatique décisionnelle. Mais construire des Data Marts n'est pas sans risques :

- En effet, dans les entreprises, des Data Marts isolés peuvent proliférer. Ces entreprises risquent de retomber dans le piège d'une architecture composée de multiples systèmes décisionnels incohérents, contenant des informations redondantes. Cela coûte plus cher et c'est plus complexe à gérer qu'un entrepôt de données centralisé.
- Les Data Marts résolvent les problèmes de performance des gros entrepôts de données. Mais ils font régresser vers le vieux problème des îlots isolés. Les entreprises vont devoir affronter des problèmes techniques complexes et coûteux pour remettre en cohérence les ensembles.
- Fédérer des Data Marts ou les faire évoluer vers une structure centralisée n'est pas facile.

On peut se poser la question s'il est préférable de bâtir un gros et unique entrepôt de données ou bien de concevoir un réservoir plus modeste, nourri par les données d'un seul département. Il est intéressant de commencer par un Data Mart, à condition de respecter certaines règles :

- Impliquer les utilisateurs.
- Ne pas construire de multiples Data Marts isolés.
- Bannir les redondances.

Les data marts sont petits, moins complexes et plus facile à déployer que les entrepôts de données. Les data marts peuvent évoluer facilement vers un entrepôt de données. Les différents data marts indépendants peuvent être dynamiquement couplés pour se métamorphoser en entrepôt de données.

Les data marts ne se résument qu'à une seule information métier (exemple : les éléments de bulletins de paye).

Donc le Data Mart peut préparer à l'entrepôt de données. Mais il faut penser grand, avenir, et adopter des technologies capables d'évoluer.

2.3. Le concept OLAP

2.3.1. Présentation

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

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

2.3.2. Les 12 règles OLAP

Afin de formaliser le concept OLAP, fin 1993, à la demande de Arbor Software, Edgar F. Codd publie un article intitulé "*Providing OLAP to User Analysts*" aux Etats-Unis. Dans cet article, il définit 12 règles que tout système de pilotage multidimensionnel devrait respecter.

Ces règles s'énoncent ainsi :

Règle 1 : Vue Conceptuelle Multidimensionnelle des données

Ces modèles permettent des manipulations simples : rotation, pivot ou vues par tranche, analyse de type permutations d'axes (*slice and dice*) ou en cascade (*drill anywhere*).

Règle 2 : Le système est transparent pour l'utilisateur

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

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

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

Règle 3 : Accessibilité à toutes les données utiles à la décision

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

Règle 4 : La performance demeure stable quel que soit le volume de données

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

Règle 5 : L'architecture est Client / Serveur

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

Règle 6 : Toutes les dimensions sont équivalentes en structure et en calcul

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

Règle 7 : Le système gère dynamiquement les Matrices Creuses

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

Règle 8 : L'accès possible à plusieurs utilisateurs simultanément

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

Règle 9 : Il n'y a pas d'opérations restrictives sur les dimensions

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

Règle 10 : Manipulation intuitive des données

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

Règle 11 : Souplesse de Création de Rapports

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

Règle 12 : Nombre illimité de dimensions et de niveaux d'agrégation

Tout outil OLAP doit gérer au moins 15 à 20 dimensions.

2.3.3. Composants d'un système OLAP

Un système OLAP est composé de différents éléments : les sources de données, un serveur OLAP et des clients.

2.3.3.1. Les sources OLAP

La source d'un système OLAP dépend de l'utilisation du produit OLAP qui en est faite, il peut s'agir d'un data warehouse, d'une base de données quelconque d'une entreprise, d'une collection de feuilles de calculs contenant des données.

La capacité d'un produit OLAP à exploiter des données provenant d'un grand nombre de sources est importante.

2.3.3.2. Le serveur OLAP

C'est le noyau d'un système OLAP. Il effectue tout le travail, c'est le lieu où les données sont stockées. Le type de stockage des données est fonction de l'architecture.

MOLAP

Le système MOLAP (Multidimensionnal OLAP) est le plus courant. On peut stocker des données de nature multidimensionnelle et fournir des temps de réponses de requêtes très rapides.

Avantages :

- Des modèles qui reflètent la réalité ;
- Des accès très rapides sans SQL ;
- Des données récapitulatives préétablies.

Inconvénients :

- Gère mal les très gros volumes de données ;
- Technologie non encore optimisée ;
- Risques d'explosion de la base de données.

ROLAP

Le ROLAP (Relational OLAP) s'appuie sur une base de données relationnelle. Les données sont enregistrées dans une base de données relationnelle dont le schéma est en général en étoile ou en flocon.

Avantages :

- Idéal pour les grands volumes ;
- Technologie optimisée et reconnue ;
- Evolution facile et souple.

Inconvénients :

- Peu efficace pour les requêtes complexes ;

HOLAP

Le HOLAP (Hybrid OLAP) est une architecture hétérogène composé de tout ou partie des architectures précitées. Les données sont stockées dans des tables relationnelles et les données agrégées sont stockées dans des cubes. Les requêtes vont chercher les données dans les tables et les cubes.

DOLAP

Le DOLAP (Desktop OLAP) décrit une catégorie de produits qui ne sont pas nécessairement connectés à un serveur. Ils peuvent s'exécuter sur un client avec la possibilité d'utiliser une source de données sous la forme d'un « Data Cube » construit et stocké localement sur une machine utilisateur.

2.3.3.3. Le client OLAP

Le client est le composant utilisé pour manipuler et visualiser les données d'une base OLAP. Un client peut être, une simple feuille de calculs avec des fonctionnalités OLAP, un outil de reporting spécialisé ou une application OLAP. Le web est la forme la plus récente de client OLAP.

Les outils de reporting

Un outil de requête ou un outil de reporting est une forme simple d'accès à des données OLAP. Les outils supportant la technologie OLAP produisent des rapports interactifs contrairement aux autres qui font des rapports statiques.

Des fonctionnalités OLAP sur les feuilles de calculs

Les feuilles de calculs sont beaucoup utilisées pour faire différentes formes d'analyses. Mais une fois le rapport généré, il est difficile de le visualiser sous différents aspects. L'ajout de fonctionnalités OLAP permet, que lorsque le rapport est généré, de pouvoir faire les manipulations voulues sur ce rapport en vue d'obtenir l'information désirée.

Le web

Le web est considéré comme le dernier venu dans la famille des clients OLAP. On trouve actuellement des produits Internet-OLAP très différents, d'où la nécessité de définir clairement les besoins avant de choisir un déploiement OLAP sur le web.

Les applications

Les applications sont similaires à des outils de requêtes ou à des outils de reporting mais elles comportent des fonctionnalités beaucoup plus approfondies et elles sont plus robustes.

Le développement

Il existe des outils de développement permettant la création d'applications spécifiques OLAP. Il s'agit souvent d'un environnement de développement constitué d'une interface graphique qui supporte le développement orienté objet. Il existe un guide API (Application Program Interface) OLAP fourni par le conseil OLAP : MDAPI. Ce document qui modélise en UML (Unified Modeling Language) toutes les métas données, requêtes et éléments d'infrastructure, a pour but de fournir aux clients tous les éléments pour développer eux-mêmes des outils d'interrogation des bases de données OLAP.

2.3.4. Offre du marché

Il existe plusieurs produits sur le marché :

Oracle : avec son offre *Express* sous deux formes qui sont *Express Server* et *Personal Express*.

Arbor : principal concurrent d'Oracle, propose *Essbase*.

Applix : a comme produit *TM-1* et a introduit le concept de RAP suivant lequel toutes opérations OLAP se fait en temps réel.

MicroStrategy : leader dans l'industrie ROLAP avec son produit *DSS*.

Cognos : son offre *Impromptu* et *PowerPlay* est produit DOLAP par excellence.

Microsoft : avec son offre *Plato* intégrant des services OLAP dans SQL Server 7.

SAS Institue : propose son produit *MDDB*.

Informix : propose *Metacube*.

2.4. La modélisation dimensionnelle

La modélisation dimensionnelle dérive des concepts OLAP (On Line Analytical Processing). Les termes dimension et fait sont apparus dans un projet de recherche commun de General Mills et de l'université de Dartmouth au cours des années 1960. Dans les années 1970, AC Nielsen et IRI ont utilisé ces termes pour décrire leurs offres d'informations considérées aujourd'hui comme des marchés d'informations dimensionnels destinés à des données de vente.

L'approche dimensionnelle n'est vraisemblablement pas l'invention d'une seule personne. Il s'agit d'une force irrésistible dans la conception de bases de données qui s'impose dès lors que le concepteur fait de l'intelligibilité et de la performance ses objectifs prioritaires.

2.4.1. Définitions

Table de faits

Une table de faits est la table principale d'un modèle dimensionnel où les mesures de performance sont stockées. Une mesure est une ligne dans une table de fait. Les faits les plus utiles d'une table de faits sont numériques et additifs.

Table de dimension

Les tables de dimension sont les points d'entrée dans la table de faits. Des attributs de dimension nombreux permettent de varier les possibilités d'analyse en tranches et en dés. Les dimensions établissent l'interface homme/entrepôt de données.

2.4.2. Processus de modélisation dimensionnelle

Le processus de modélisation dimensionnelle peut être décomposé en quatre (4) étapes.

Etape 1. Sélection du processus d'entreprise à modéliser

Un processus d'entreprise est une activité normale de votre organisation, généralement assisté par un système source collectant des informations.

Etape 2. Déclaration du grain

Le grain définit exactement ce que représente une ligne individuelle de table de faits. Il répond à la question « comment décrivez-vous une ligne unique de la table de faits? ».

Etape 3. Choix de dimension

Cela revient à choisir les dimensions qui s'appliquent à chaque ligne de la table de faits. Les dimensions résultent de la question : « Comment les gestionnaires décrivent-ils des données qui résultent du processus concerné ? ».

Etape 4. Identification des faits

C'est la détermination des faits qui apparaîtront dans la table de faits. A ce niveau, le choix est fonction du choix du grain et comme pour l'étape précédant, la détermination des faits peut conduire à revoir la déclaration du grain ou le choix des dimensions.

2.4.3. Dimensions

2.4.3.1. Les dimensions additionnelles

Ces dimensions additionnelles peuvent être ajoutées à la conception sans rien perturber, même après la mise en production de l'entrepôt de données. Cela est possible parce qu'elles ne modifient pas le grain de la table de faits.

- Dimensions causales

Ces dimensions peuvent donner des indications intéressantes sur la cause d'un événement.

- Dimensions de dates multiples ou d'horodatage

Elles sont utilisées spécialement lorsque la granularité de la table de faits est la ligne d'un document de suivi.

- Dimensions dégénérées

Elles sont généralement réservées à des identificateurs de transactions opérationnelles. Elles ne doivent pas servir de prétexte pour placer dans la table de faits un code inintelligible sans en donner le décodage dans une table de dimension.

- Dimensions à jeu de rôle

Le jeu de rôle intervient dans le cas d'une dimension physique unique apparaissant plusieurs fois dans une table de faits, chaque apparition pointant sur une table de dimension logique distincte créée par le biais de vues ayant ses propres noms de colonne.

- Dimensions d'état

Ces dimensions identifient l'état actuel d'une transaction ou d'un instantané mensuel, par exemple l'état d'un compte.

- Dimension audit

Cette dimension sert à suivre l'origine et la qualité des données.

- Dimensions fourre-tout

Une dimension fourre-tout sert à regrouper commodément des drapeaux et des indicateurs de faibles cardinalités. Elle permet d'éliminer les drapeaux de la table de faits tout en les plaçant dans un cadre dimensionnel utile.

2.4.3.2. Les dimensions changeantes

- Dimensions à valeurs multiples

Un attribut ayant un nombre de variables multiples peut être associé à une ligne de dimension par l'intermédiaire d'une table passerelle reliant les diverses valeurs d'attributs à la dimension.

- Dimensions à évolution lente

Cela concerne le changement d'attribut des dimensions. Des techniques et des approches hybrides existent pour le traitement des changements d'attributs.

Type 1 : écrasement de la valeur précédente

Dans ce cas, l'ancienne valeur de l'attribut est remplacée par la nouvelle valeur dans la ligne de dimension concernée.

La solution de type 1 est facile à mettre en œuvre, mais elle ne conserve aucune trace de valeurs antérieures des attributs.

Type 2 : ajout d'une ligne de dimension

Dans ce cas, une nouvelle ligne est ajoutée indiquant la nouvelle valeur de l'attribut. La solution de type 2 est la principale technique

permettant de suivre les attributs des dimensions à évolution lente correctement.

Type 3 : ajout d'une colonne de dimension

Dans ce cas, on ajoute une nouvelle colonne pour refléter le changement d'attribut. La solution de type 3 permet de voir les données antérieures à la fois selon les nouvelles et les anciennes valeurs d'attribut.

Technique hybride 1 : changements prévisibles et application aux données de multiples versions des attributs modifiés.

Cette technique est fréquemment utilisée pour traiter des modifications de la structure commerciale.

Technique hybride 2 : changements prévisibles avec application aux données antérieures de la version actuelle de l'attribut modifié.

Cette technique convient s'il faut préserver la vision exacte du passé dans le contexte de modifications d'attributs imprévisibles tout en ayant le moyen de présenter les données antérieures selon les valeurs actuelles de l'attribut modifié.

- **Dimensions à évolution rapide**

C'est le cas où un attribut change fréquemment. Une solution consiste à isoler ces attributs qui changent rapidement et à les placer dans une ou plusieurs dimensions distinctes.

2.4.3.3. Les minidimensions

Elles constituent la meilleure approche pour parcourir les principaux attributs et en suivre les changements dans les dimensions de très grande

taille. L'approche consiste à extraire de la table un ou plusieurs petits groupes d'attributs n'ayant qu'un nombre limité de valeurs et à les placer chacun dans une minidimension.

2.4.4. Tables des faits

2.4.4.1. Les familles de tables des faits

Chaînes et cercles

Plusieurs tables de faits sont nécessaires pour prendre en compte les activités comprenant plusieurs processus. Chaque processus engendre une ou plusieurs tables. Lorsque les processus sont naturellement ordonnancés, on peut qualifier l'ensemble de chaîne de valorisation.

Schéma de produits hétérogènes

Plusieurs tables des faits sont nécessaires lorsqu'une activité engendre des produits hétérogènes dont les faits sont naturellement différents mais qui possèdent la même base client.

Exemple : une table des faits personnalisée dédiée à l'activité comptes courants d'une banque.

Schéma des transactions

Dans la table des faits, l'enregistrement d'une transaction contient en général un fait unique : la valeur de la transaction. Le schéma des faits transactionnels n'est pas aussi souple que celui des instantanés.

Les entreprises munies d'un historique transactionnel riche mais dont les transactions ne sont pas constitutives d'une recette ont généralement besoin de deux tables de faits pour représenter l'ensemble de l'activité. Une des tables des faits sert à la capture des transactions et l'autre à celle des instantanés périodiques.

Schéma des instantanés

La table des instantanés est la table compagne de la table des transactions. La flexibilité des tables d'instantanés permet d'ajouter progressivement des faits à mesure que l'on comprend mieux les besoins des utilisateurs.

Les instantanés permettent de mesurer rapidement l'état de l'activité tandis que les transactions donnent une vue aussi complète que possible d'un comportement détaillé.

Agrégats

Les agrégats sont des récapitulatifs que l'on construit en vue d'optimiser les performances des requêtes. Les agrégats sont stockés dans des tables qui leur sont dédiées et non dans les tables des faits d'origine, qui contiennent les données non agrégées.

2.4.4.2. Les tables des faits sans faits

Les tables de faits sans faits sont très intéressantes pour décrire les événements et les couvertures de manière à permettre des requêtes sur des faits qui ne se sont pas produits.

Le premier type de table des faits sans faits est celui qui enregistre un événement. Par exemple la table des faits sans faits représentant les événements présence au cours au sein d'un institut universitaire.

Le second type de table des faits sans faits est celui de la table de couverture. Par exemple la table des faits sans faits représentant les ventes promotionnelles par point de vente et par période permet de répondre à la question concernant des événements qui ne sont pas produits comme la liste des produits en promotion qui ne sont pas vendus.

2.4.5. Clés artificielles

Les clés artificielles sont des entiers affectés séquentiellement lors du garnissage d'une dimension. Toute jointure entre des tables de dimension et des tables de faits doit se baser sur des clés entières artificielles sans signification. Il faut éviter d'utiliser les codes opérationnels naturels des systèmes de productions. Aucune clé de l'entrepôt de données ne doit être intelligible c'est-à-dire suffire à elle seule à donner une indication sur la ligne concernée.

2.4.6. Mythe de la modélisation dimensionnelle

Malgré son acceptation, certains malentendus existent toujours dans la profession à son sujet : ce sont les mythes de la modélisation dimensionnelle.

Mythe 1. Les modèles dimensionnels et les marchés d'informations ne servent que pour les données cumulées.

Rien dans un modèle dimensionnel n'interdit le stockage d'informations historiques. Le volume de données historiques disponibles dans un marché d'informations doit dépendre des besoins de l'activité observée.

Mythe 2. Les modèles dimensionnels et les marchés d'informations sont des solutions par service et non au niveau de l'entreprise.

Les marchés d'informations doivent être organisés autour de processus d'entreprise.

Mythe 3. Les modèles dimensionnels et les marchés d'informations ne sont pas extensibles.

Les tables de faits modernes abritent des milliards de lignes. La modélisation dimensionnelle a été appliquée à presque toutes les activités comme la banque, les assurances, la téléphonie, les administrations, les jeux, la fabrication, la distribution, l'éducation et biens d'autres.

Mythe 4. Les modèles dimensionnels et les marchés d'informations ne conviennent que s'il existe des schémas d'utilisations prévisibles.

Les structures dimensionnelles sont extrêmement souples et adaptables au changement. Le secret de la souplesse en matière de requête est la construction de tables de faits au grain le plus fin. Le point de départ correct d'un modèle dimensionnel est d'exprimer les données au plus petit détail possible en vue d'une souplesse et d'une extensibilité maximales.

Mythe 5. Les modèles dimensionnels et les marchés d'informations ne peuvent pas être intégrés et conduisent par suite à des solutions verticales.

Les modèles dimensionnelles et marchés d'informations peuvent être intégrés s'ils sont conformes à l'architecture de bus des entrepôts de données. On ne peut rendre la modélisation dimensionnelle responsable de l'incapacité de certaines organisations à respecter l'un de ses principes fondamentaux.

2.4.7. Modélisation

2.4.7.1. La modélisation en étoile

Dans un modèle en étoile, tous les faits sont définis dans une simple table relationnelle. Cette table des faits va être reliée par clés primaires à d'autres tables correspondant aux dimensions. En fait, le modèle en étoile essaie de superposer une structure multidimensionnelle au dessus d'un modèle relationnel normalisé à deux dimensions.

Le modèle en étoile simplifie le modèle logique normalisé en organisant les données de manière optimale pour les traitements d'analyse.

2.4.7.2. La modélisation en flocon

La modélisation en flocon est une modélisation en étoile pour laquelle on éclate les tables de dimensions en sous-tables selon la hiérarchie de cette dimension. C'est un modèle en étoile normalisé.

2.4.7.3. La modélisation en constellation

La modélisation en constellation consiste à fusionner plusieurs modèles en étoile qui utilisent des dimensions communes. Un modèle en constellation comprend donc plusieurs tables de faits et des tables de dimensions communes ou non à ces tables de faits.

2.4.8. Gestion du projet de modélisation dimensionnelle

La gestion d'un projet de modélisation dimensionnelle consiste surtout à faire circuler le résultat de la conception entre les personnes. Une bonne vision du projet est indispensable à une communication efficace. Cela à travers les quatre outils graphiques suivants:

- la matrice de l'architecture en bus décisionnel ;
- le diagramme d'une table des faits ;
- le détail d'une table des faits ;
- le détail d'une table dimensionnelle.

2.4.8.1. La matrice de l'architecture en bus décisionnel

La matrice développée par l'équipe de conception pourra être révisée en vue de servir de support de présentation à l'intention d'autres concepteurs, administrateurs et utilisateurs. La matrice peut être utilisée comme une introduction générale à la conception ; elle procure alors à chaque interlocuteur une vue de ce que sera l'entrepôt de données une fois terminé.

2.4.8.2. Le diagramme de la table des faits

Le diagramme de la table des faits ne se contente pas d'illustrer les spécifiés de la table des faits ; il la situe aussi dans le contexte du data mart. Il nomme la table des faits, énonce clairement sa granularité et présente toutes les dimensions auxquelles elle est reliée. Il donne également un aperçu de toutes les dimensions qui ont été identifiées pour l'activité concernée. Le diagramme représente les dimensions non concernées sans liens et cela permet d'éveiller la créativité (leur connexion ultérieure peut devenir un objectif d'extension) et l'attention (le fait qu'elles ne soient pas connectées montre clairement qu'elles ne seront pas disponibles pour cette table des faits).

Un diagramme de table des faits est également un vecteur de communication multidirectionnel qui facilite :

- Les liaisons entre le service informatique et les utilisateurs. Le diagramme facilite le partenariat entre le service informatique et les utilisateurs. Il suscite l'appropriation du projet par les utilisateurs ; il peut aussi servir lors de la phase de validation de la conception par les utilisateurs.
- La sélection des outils d'accès aux données. Une version moins détaillée de ce diagramme pourra être soumise aux éditeurs d'outils.
- La formation des utilisateurs. Le diagramme de la table des faits a montré son efficacité en tant qu'outil de formation dans la mesure où il présente les données aux utilisateurs sous forme compréhensible.

2.4.8.3. Le détail de la table des faits

Le détail de la table des faits représente la liste complète des faits de la table. On y trouve les faits présents physiquement dans la table, les faits dérivés présentés au travers des vues des systèmes de gestion de bases de données (SGBD) et des faits calculés à partir de ceux des deux premiers groupes. Chaque fait doit posséder ses propres règles d'agrégation afin que la personne qui les consulte soit avertie que tel ou tel fait est semi-additif ou non additif. La liste des faits d'une table donnée

peut parfaitement refléter des données indisponibles mais demandées par les utilisateurs, à condition d'indiquer clairement cette indisponibilité momentanée. Il se peut que ces éléments prévus n'aient aucune source potentielle aujourd'hui, ou bien que cette source n'entre pas dans les objectifs du projet. Ils représentent des objectifs futurs.

2.4.8.4. Le Détail de la table dimensionnelle

Le détail de la table dimensionnelle présente les attributs d'une dimension. Chaque dimension a son propre diagramme qui explicite sa granularité. Le diagramme affiche les cardinalités approximatives de chaque attribut dimensionnel et donne à l'utilisateur une vue d'ensemble des nombreuses hiérarchies et relations entre les attributs. Les relations définissent les chemins de forage (drill) par défaut. Ce diagramme permet aussi d'inclure des attributs demandés par les utilisateurs mais non encore disponibles ou non prévus dans le projet initial.

PARTIE III:
Application

3.1. Schéma global des composants du data mart

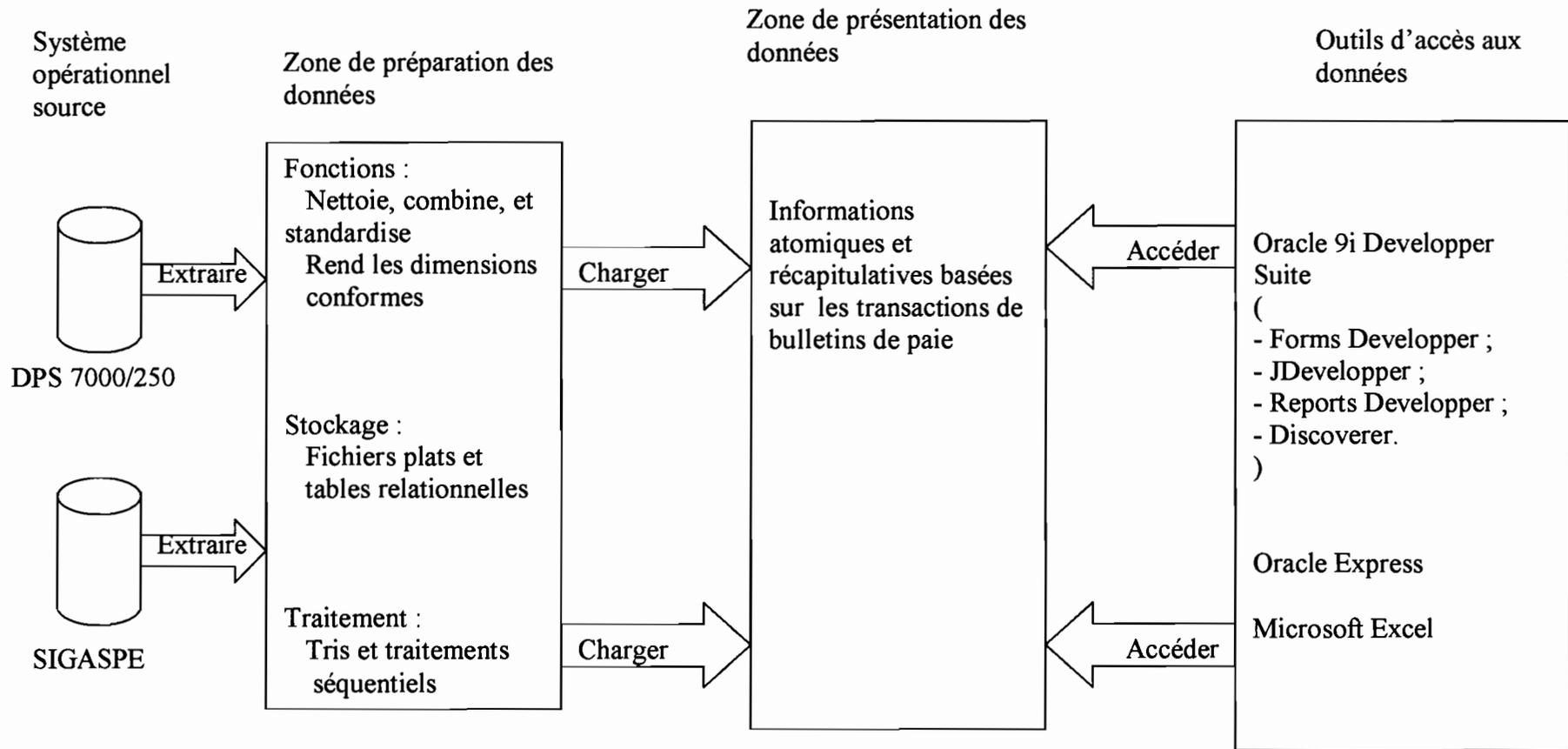


Schéma 2: composant d'un data mart dans le cadre de la paie du personnel de l'Etat.

3.2. La phase Extraction Traitement Chargement (ETC)

Au regard de la différence des deux systèmes opérationnels sources (DPS 7000/250 et SIGASPE), la phase d'extraction traitement chargement nécessitera des procédures différentes selon les systèmes.

3.2.1. Extraction

3.2.1.1. Sur le système DPS 7000/250

Les données ont été extraites dans des fichiers texte avec le point virgule (;) comme séparateur de colonnes. L'extraction est faite par un programme COBOL qui écrit les données dans des fichiers texte de taille comprise entre 8Mo et 1Ko. Les fichiers sont transférés du DPS 7000/250 vers le Bull Micral 200. A partir du Bull Micral 200 les fichiers sont facilement récupérables sur les micro-ordinateurs. Pour un mois donné, on peut avoir 2 à 6 fichiers. Les fichiers sont constitués par paire : un fichier concernant les bulletins et un autre fichier concernant les éléments de bulletins. Les noms des fichiers ont la nomenclature suivante :

- BULxxyyz ou Bxxyy ou BExxyy ou B40xxyy pour les bulletins ;
- ELMxxyyz ou Exxyy ou EExxyy ou E40xxyy pour les fichiers des éléments de bulletins.

Avec xx représentant l'année sur deux(2) positions, yy le mois sur deux (2) positions et z un numéro séquentiel.

A partir de la structure des fichiers texte, un modèle physique est conçu pour permettre la récupération des données.

La procédure de chargement proposée est la suivante :

- Regroupement de tous les fichiers dans un seul répertoire « charger »;
- Listage des fichiers dans un autre fichier texte par la commande : dir /B /C /L > nomfichier.txt;

- Création d'une table dans la base de données pour charger la liste contenue dans nomfichier.txt ;
- Chargement de la table en utilisant l'outil d'Oracle SQLLOADER ;
- Création de deux (2) fichiers de script (un pour les bulletins « fchargebu.sql » et l'autre pour les éléments de bulletins « fchargelt.sql ») à partir d'un programme PL/SQL ;
- Lancement du script de création des deux tables (bulletin_7000 et r_bulletin_7000) ;
- Lancement des scripts fchargebu.sql et fchargelt.sql pour charger l'ensemble des données en une seule fois.

3.2.1.2. Sur le système SIGASPE

L'extraction à ce niveau est moins complexe car l'application tourne sous une base de données Oracle. On utilise un fichier de sauvegarde logique fait à l'aide de l'outil EXPORT d'Oracle. Ce fichier a pour extension « dmp ». A partir de ce fichier on fait un chargement en utilisant l'outil IMPORT d'Oracle.

Afin de résoudre le problème de la disponibilité des anciens bulletins de paie nécessaires aux agents de l'Etat pour compléter leurs différents dossiers, une application a été développée. Cette application permet de consulter et de faire l'édition des bulletins de paie au format de papier A4 et papier pré imprimé. Cette application a été développée à l'aide des outils de développement Oracle Developer.

3.2.2. La phase de traitement et de chargement

Le traitement et le chargement se résument en un ensemble de programmes PL/SQL que l'on peut exécuter sous forme de script.

Avant de procéder au chargement des données il est nécessaire de mettre en place le modèle dimensionnel car il définit la structure qui doit recevoir les données extraites et traitées provenant des systèmes opérationnels.

3.3. Le modèle dimensionnel

3.3.1. Matrice de l'architecture en bus décisionnel

Processus d'entreprise	Dimensions									
	Période	Agent	Section /service	classification	Lieu	Elément de salaire	Billeteur	Fonction	Banque/mode de paiement	Qualité/situation matrimoniale
Transaction de bulletins de paie	x	x	x	x	x	x	x	x	x	x

Tableau 6: matrice de bus d'un data mart de la paie du personnel de l'Etat

3.3.2. Diagramme de la table des faits

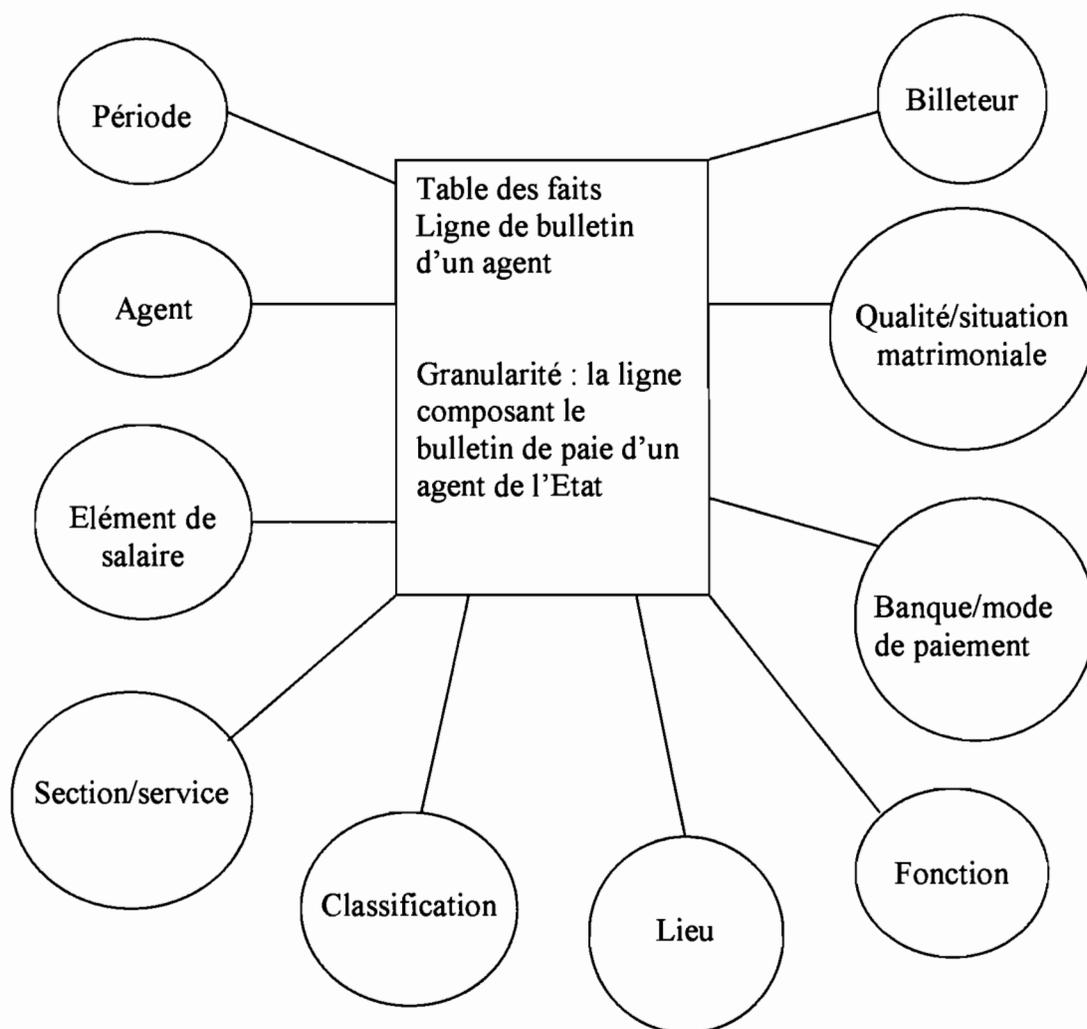


Schéma 3: diagramme de la table des faits

3.3.3. Détail de la table des faits

Table des faits Ligne de bulletin d'un agent

clé_période

clé_agent

clé_classification

clé_section/service

clé_fonction

clé_lieu

clé_billeteur

clé_banque_modepaiement

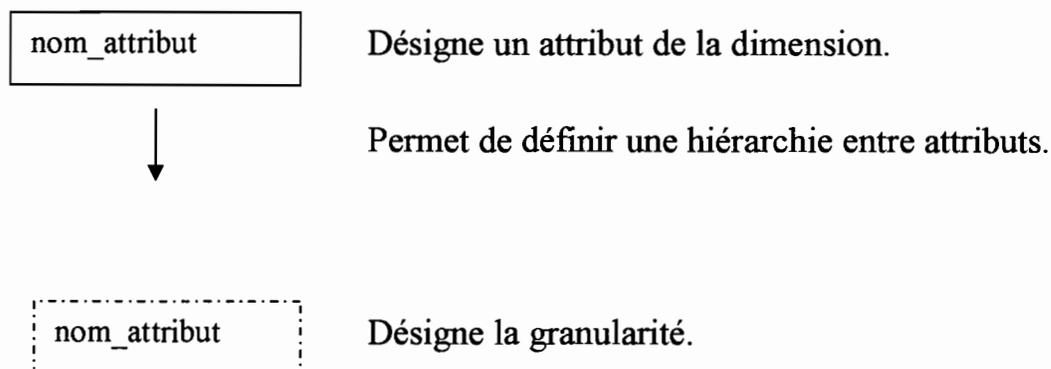
clé_qualite_sit_matrimoniale

clé_element_salaire

montant

3.3.4. Détail de la table dimensionnelle

Pour la représentation du diagramme de détail de la table dimensionnelle le formalisme suivant a été utilisé :



Les autres flèches matérialisent une relation entre les autres attributs et l'attribut clé de la dimension.

- *Dimension période*

Description détaillée des attributs

Nom attribut	Description	Gestion dimension changeante
Cle_période	Clé artificielle	Pas de mise à jour
Mois	Représente le mois de paie	Pas de mise à jour
Trimestre	Ensemble de trois mois de paie	Pas de mise à jour
Année	Année de paie	Pas de mise à jour

Tableau 7 : description détaillée des attributs de la dimension période

Diagramme de détail de la table dimensionnelle

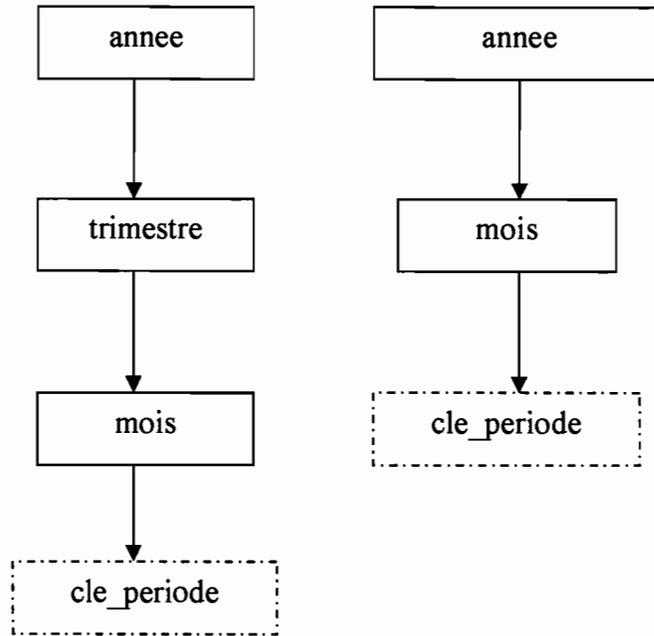


Schéma 4: diagramme de détail de la table dimensionnelle période

- *Dimension fonction*

Description détaillée des attributs

Nom attribut	Description	Gestion dimension changeante
Cle_fonction	Clé artificielle	Pas de mise à jour
Code_fonction	Clé du système opérationnel	Pas de mise à jour
Lib_fonction	Libellé de la fonction	Ecrasement
type_fonction	Les différents types de fonction	Ecrasement

Tableau 8: description détaillée des attributs de la dimension fonction

Diagramme de détail de la table dimensionnelle

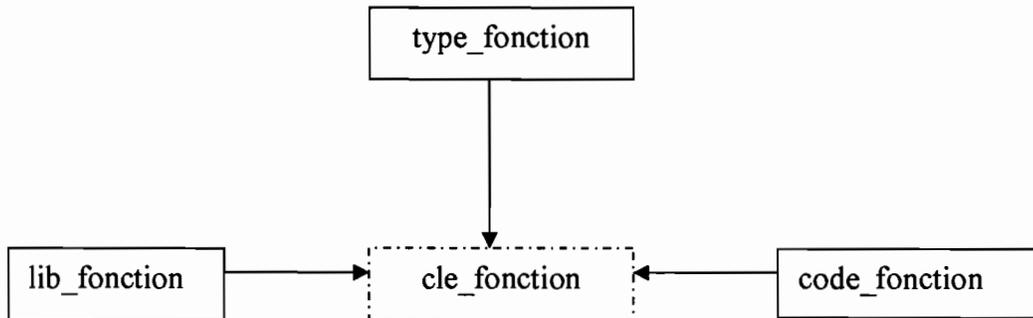


Schéma 5: diagramme de détail de la table dimensionnelle fonction

- *Dimension lieu*

Description détaillée des attributs

Nom attribut	Description	Gestion dimension changeante
Cle_lieu	Clé artificielle	Pas de mise à jour
Code_lieu	Clé du système opérationnel	Pas de mise à jour
Lib_lieu	Libellé du lieu	Ecrasement
Cd_province	Code de la province. Il regroupe un ensemble de départements.	Pas de mise à jour
Cd_depart	Code du département. C'est un ensemble de localités.	Pas de mise à jour
Cd_localité	Code de la localité. Le numéro de la localité dans le département.	Pas de mise à jour

Tableau 9: description détaillée des attributs de la dimension lieu

Diagramme de détail de la table dimensionnelle

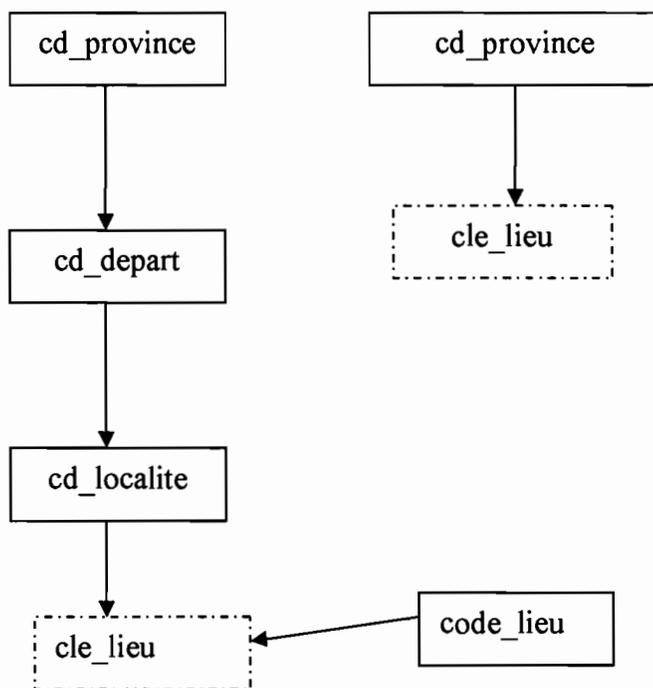


Schéma 6: diagramme de détail de la table dimensionnelle lieu

- *Dimension banque/mode de paiement*

Description détaillée des attributs

Nom attribut	Description	Gestion dimension changeante
Cle_banque_mode	Clé artificielle	Pas de mise à jour
Code_banque	Code de la banque dans le système opérationnel	Ecrasement
Code_modepaie	Code du mode de paiement	Ecrasement
Lib_banque_mode	Libellé de la banque ou mode de paiement	Ecrasement

Tableau 10: description détaillée des attributs de la dimension banque/mode paiement

Diagramme de détail de la table dimensionnelle

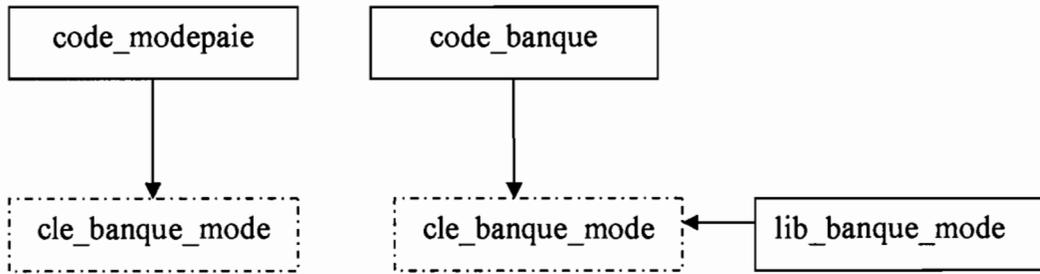


Schéma 7: diagramme de détail de la table dimensionnelle banque/mode paiement

- *Dimension élément de salaire*

Description détaillée des attributs

Nom attribut	Description	Gestion dimension changeante
Cle_eltbase	Clé artificielle	Pas de mise à jour
Code_elt	Code de l'élément de salaire utilisé dans le système opérationnel	Pas de mise à jour
Lib_eltbase	Libellé de l'élément de salaire	Ecrasement
Cd_signe	Indique le signe de l'élément : positif ou négatif	Ecrasement
Type_elt	Indique le type de l'élément de salaire : gain ou retenue	Ecrasement
Cd_imposable	Indique si l'élément de salaire est imposable	Ecrasement
Cd_saisissable	Indique si l'élément de salaire est saisissable	Ecrasement

Cd_qcessible	Permet de dire si l'élément de salaire participe à la quotité cessible	Ecrasement
Cd_retraite	Définit si l'élément de salaire est pris en compte ou pas dans la base de retraite	Ecrasement
Cd_recap	Critère de regroupement récapitulatif des éléments de salaire.	Ecrasement
Cd_parag	Définit le paragraphe de l'élément de salaire	Ecrasement

Tableau 11: description détaillée des attributs de la dimension élément de salaire

Diagramme de détail de la table dimensionnelle

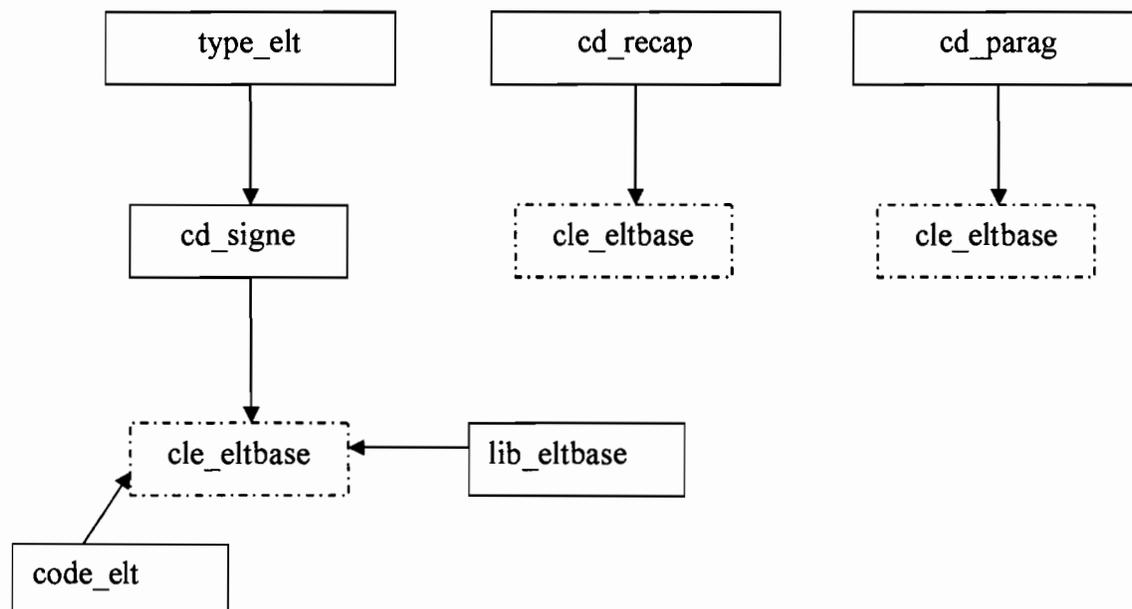


Schéma 8: diagramme de détail de la table dimensionnelle élément de salaire

- *Dimension agent*

Description détaillée des attributs

Nom attribut	Description	Gestion dimension changeante
No_matricule	Numéro matricule de l'agent	Pas de mise à jour
Lettre_cle	Lettre clé matricule	Pas de mise à jour
Nom	Nom de l'agent	Ecrasement
prenoms	Prénoms de l'agent	Ecrasement
Nom_jf	Nom de jeune fille	Ecrasement
Dt_naissance	Date de naissance	Ecrasement
Sexe	Sexe de l'agent	Ecrasement
Dt_recrutement	Date de recrutement	Ecrasement
Cd_nationalite	La nationalité de l'agent	Ecrasement
Ty_integration	Type d'intégration	Ecrasement
Cd_recrutement	Mode de recrutement	Ecrasement

Tableau 12: description détaillée des attributs de la dimension agent

Diagramme de détail de la table dimensionnelle

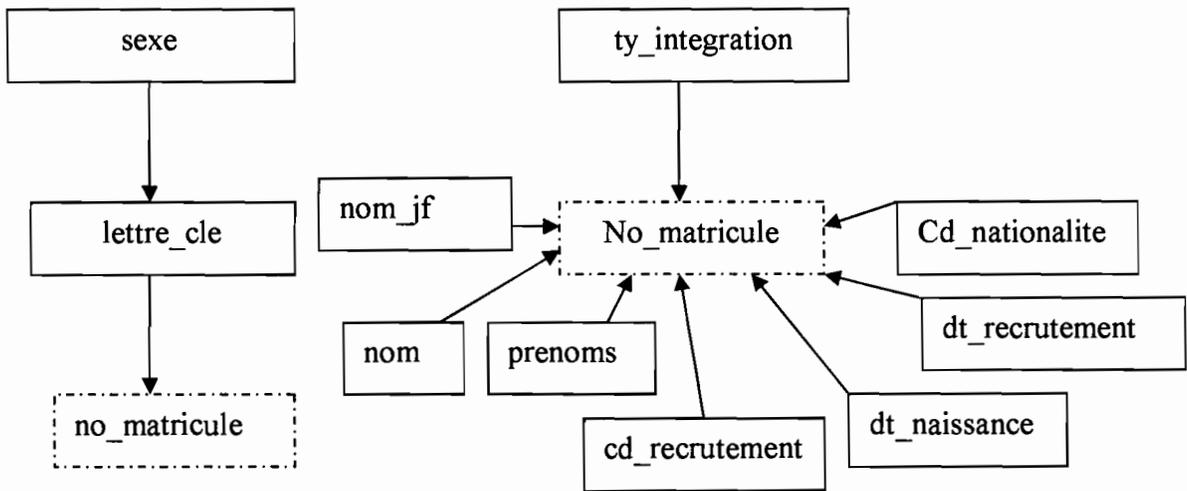


Schéma 9: diagramme de détail de la table dimensionnelle agent

- Dimension qualité/situation matrimoniale

Description détaillée des attributs

Nom attribut	Description	Gestion dimension changeante
Cle_qual_sitmatri	Clé artificielle	Pas de mise à jour
Code_qualite	Donne la qualité de l'agent	Pas de mise à jour
Code_sit_mat	Donne la situation matrimoniale	Pas de mise à jour
Lib_qual_sitmatri	Libellé qualité et situation matrimoniale	Ecrasement

Tableau 13: description détaillée des attributs de la dimension qualité/sit. matrimoniale

Diagramme de détail de la table dimensionnelle

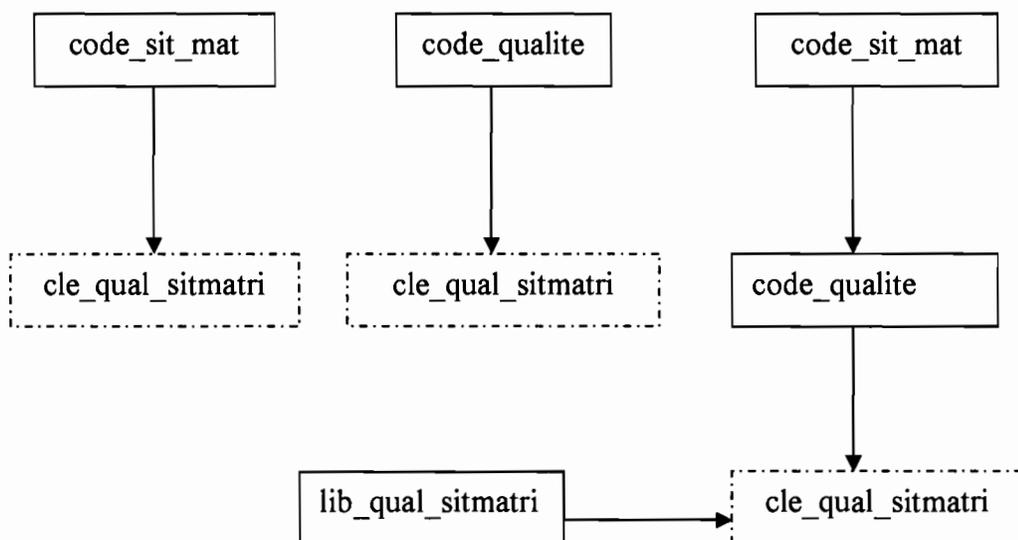


Schéma 10: diagramme de détail de la table dimensionnelle qualite/situation matrimoniale

- Dimension billeteur

Description détaillée des attributs

Nom attribut	Description	Gestion dimension changeante
Cle_billeteur	Le numéro du billeteur	Pas de mise à jour
No_matricule	Numéro matricule du billeteur	Pas de mise à jour
Dt_debut	Date de début d'activité du billeteur	Ecrasement
Dt_fin	Date de fin d'activité	Ecrasement

Tableau 14: description détaillée des attributs de la dimension billeteur

Diagramme de détail de la table dimensionnelle

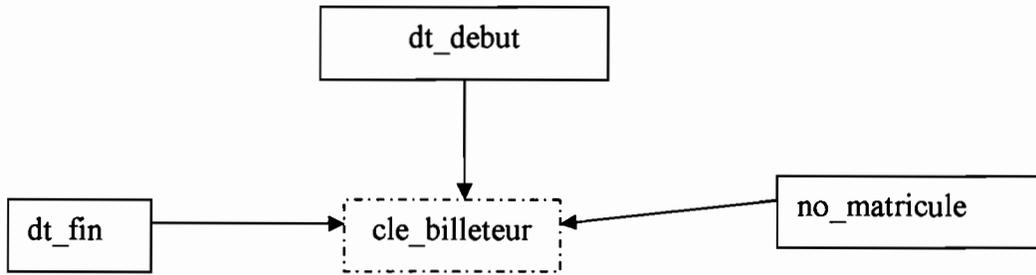


Schéma 11: diagramme de détail de la table dimensionnelle billeteur

- *Dimension section/service*

Description détaillée des attributs

Nom attribut	Description	Gestion dimension changeante
Cle_sect_service	Clé artificielle	Pas de mise à jour
Code_section	Numéro de la section dans le système opérationnel	Pas de mise à jour
Code_service	Code du service dans le système opérationnel	Pas de mise à jour
No_struct_1	Premier niveau	Pas de mise à jour
No_struct_2	Deuxième niveau	Pas de mise à jour
No_struct_3	Troisième niveau	Pas de mise à jour
No_struct_4	Quatrième niveau	Pas de mise à jour
Lib_long_sce	Libellé long du service	Ecrasement
Lib_court_sce	Libellé court du service	Ecrasement
No_ex	L'exercice	Ecrasement
Titre	Le titre du service	Ecrasement
Chap	Le chapitre du service	Ecrasement

Tableau 15: description détaillée des attributs de la dimension section/service

Diagramme de détail de la table dimensionnelle

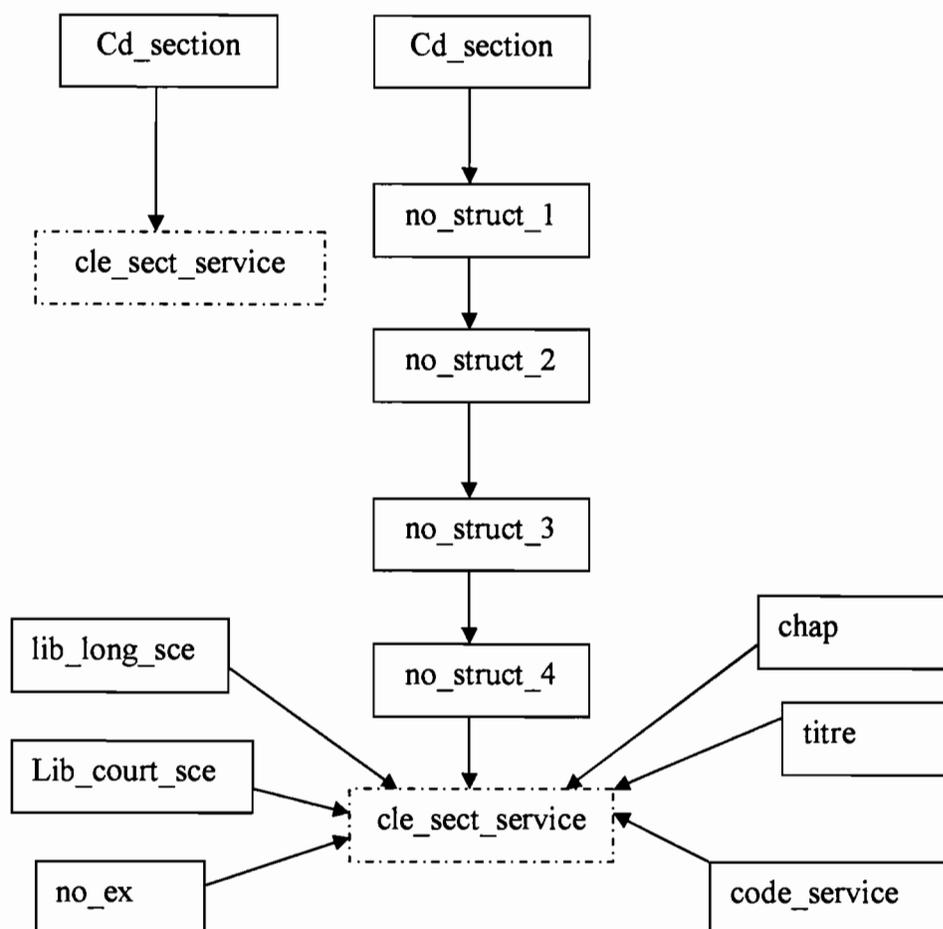


Schéma 12: diagramme de détail de la table dimensionnelle section/service

- *Dimension classification*

Description détaillée des attributs

Nom attribut	Description	Gestion dimension changeante
Cle_classif	Clé artificielle	Pas de mise à jour
Code_categorie	La categorie de l'agent	Pas de mise à jour
Code_echelle	L'échelle de l'agent	Pas de mise à jour
Code_grade	Le grade de l'agent	Pas de mise à jour
Code_echelon	L'echelon de l'agent	Pas de mise à jour
Lib_classif	Libellé classification	Ecrasement

Tableau 16 : description détaillée des attributs de la dimension classification

Diagramme de détail de la table dimensionnelle

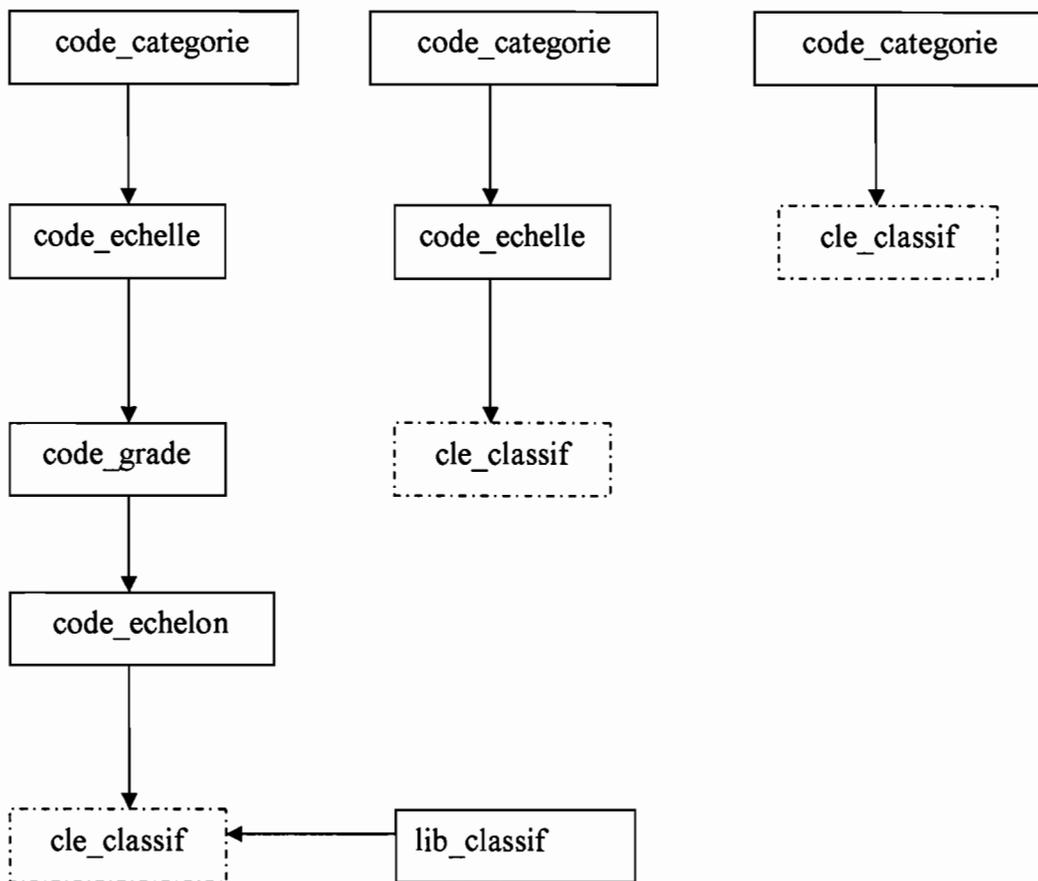


Schéma 13: diagramme de détail de la table dimensionnelle classification

3.4. Le modèle physique en étoile

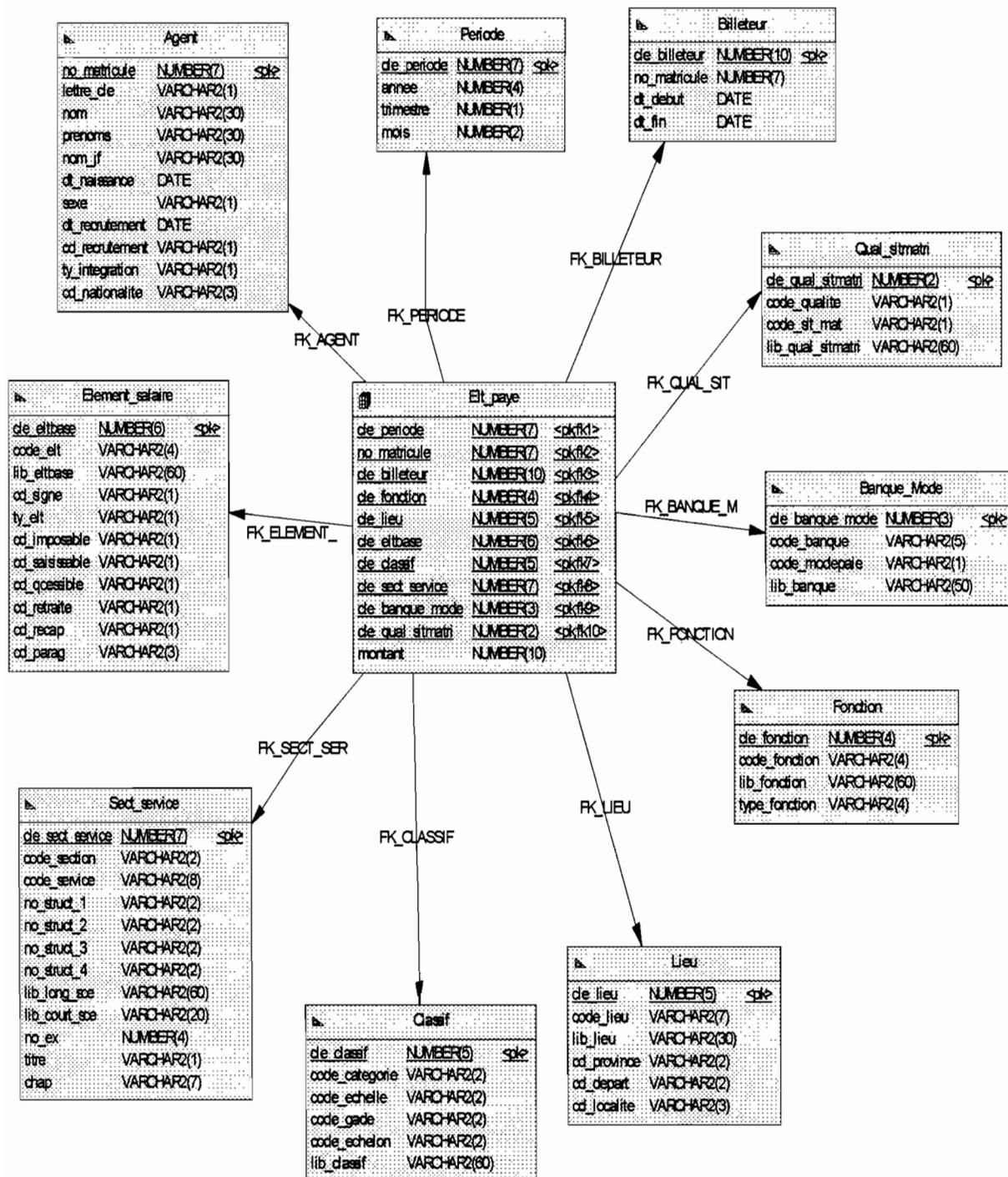


Schéma 14: modèle physique en étoile

3.5. Le modèle multidimensionnel

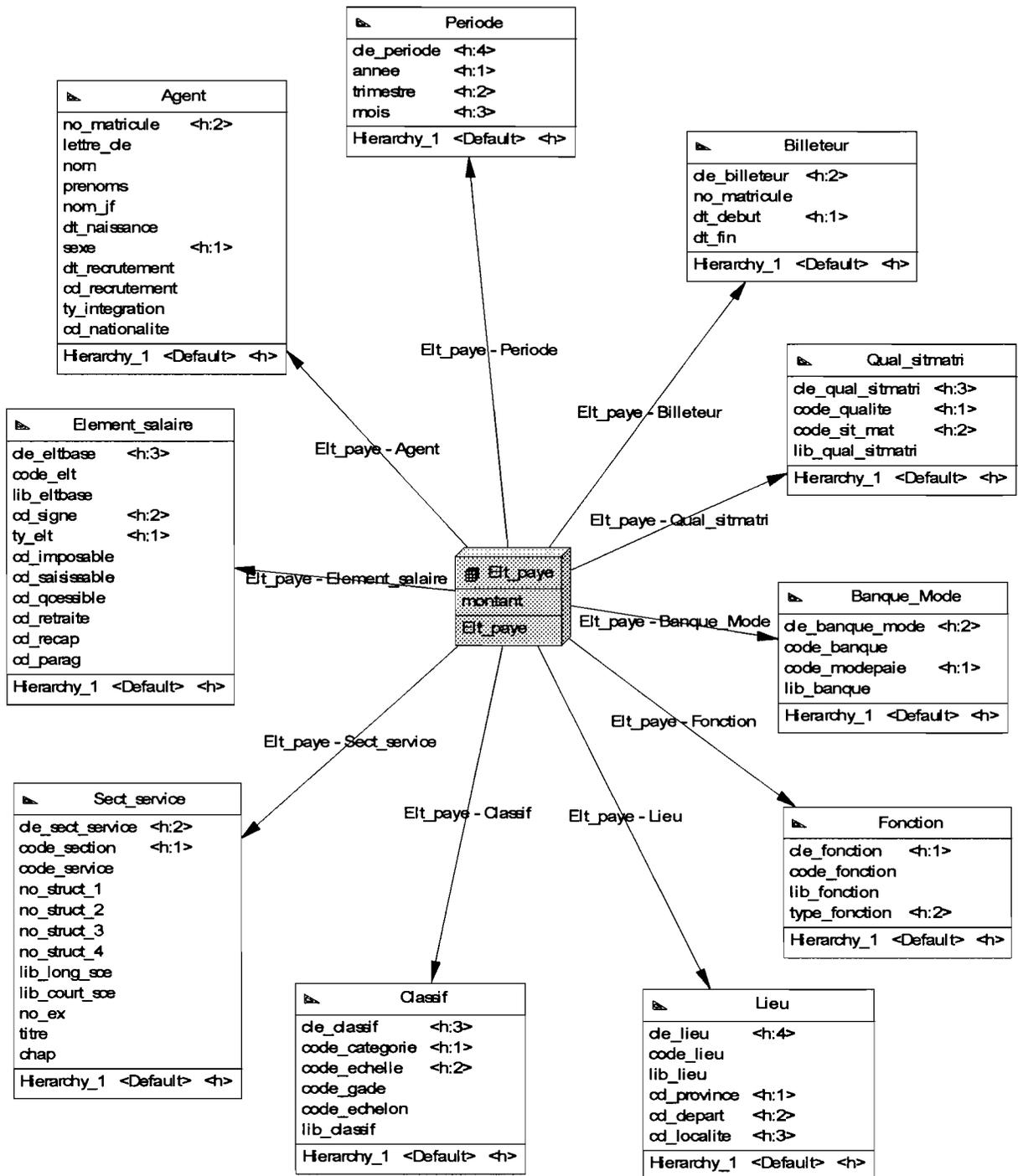


Schéma 15 : modèle multidimensionnel

3.6. Les outils d'accès

Le choix des produits est guidé par l'existant. En effet, la structure possède la gamme des outils Oracle nécessaires, et au regard de l'offre du marché, nous nous sommes rendus compte que les produits Oracle étaient adaptés.

Comme outils d'accès au data mart nous proposons :

- La mise en place d'une application développée à l'aide des outils de développement Oracle comme Forms Developer, Reports Developer, ainsi que Oracle Express qui permet de développer des applications OLAP. Ceci en vue de permettre la mise à disposition des utilisateurs d'un ensemble d'états répertoriés et validés par les utilisateurs.
- La mise à la disposition des utilisateurs d'un outil de requête et de représentation graphique comme Discoverer. Cet outil permet de faire des requêtes et stocke les résultats dans des tableaux croisés que l'utilisateur a la possibilité de modifier. Les données des tableaux peuvent être sélectionnées pour faire des graphiques. Discoverer permet le renvoi des données vers Microsoft Excel. Cet outil permettra aux utilisateurs d'avoir une certaine autonomie dans la conception et la réalisation de leurs états.

3.7. Le scenario de mise en œuvre

Pour la mise en œuvre, nous proposons au choix les solutions suivantes basées sur deux architectures :

- Architecture client/serveur

Pour le serveur nous proposons les caractéristiques suivantes :

- processeur Xéon (bi processeur) ;
- mémoire RAM 1 Go ;
- unité de sauvegarde ;
- carte réseau 10/100 Mb/s ;
- cinq (5) disques SCSI hot Plug de 36 Go chacun en RAID ;

- système d'exploitation : Unix, Linux, Microsoft Windows NT.

Pour les postes clients :

- processeur pentium III au moins
- mémoire RAM 256 Mo
- disque dur de 40 Go
- carte réseau 10/100 Mb/s
- système d'exploitation : Microsoft Windows 98/NT/2000/XP

- Architecture 3Tiers

Pour le serveur de données :

- processeur Xéon (bi processeur) ;
- mémoire RAM 1Go ;
- unité de sauvegarde ;
- carte réseau 10/100 Mb/s ;
- cinq (5) disques SCSI hot Plug de 36 Go chacun en RAID ;
- système d'exploitation : Unix, Linux, Microsoft Windows NT.

Pour le serveur d'application :

- processeur Xéon (quadri processeur) ;
- mémoire RAM 2Go ;
- carte réseau 10/100 Mb/s ;
- cinq (3) disques SCSI hot Plug de 36 Go chacun en RAID ;
- système d'exploitation : Unix, Linux, Microsoft Windows NT.

Pour les postes clients :

- processeur pentium III au moins ;
- mémoire RAM 128Mo ;
- disque dur de 40 Go ;
- carte réseau 10/100 Mb/s ;

- système d'exploitation : Microsoft Windows 98/NT/2000/XP.

Les utilisateurs concernés par le data mart sont en premier les responsables des services en charge des salaires du Ministère des Finances et du Budget. L'accès au data mart peut être élargi à d'autres structures comme les Directions Administratives et Financières des différents ministères.

Tous les utilisateurs devront accéder au data mart par identification. Cette identification pourra se faire selon le cas : soit par le serveur de données dans l'environnement client/serveur, soit par le serveur d'application dans l'environnement 3Tiers.

Le data mart sera géré par la Direction des Services Informatiques du Ministère des Finances et du Budget qui a en charge la gestion des applications du ministère.

Une session de formation des utilisateurs à l'utilisation de l'application et à l'utilisation des outils de requête sera organisée avant la mise en exploitation du système adopté.

Conclusion

La mise en place d'un data mart peut être perçue comme un élément d'amélioration des prises de décision au niveau de l'entreprise. Nos travaux nous ont permis de faire une étude théorique sur les entrepôts de données ou data warehouse, les data marts, la modélisation dimensionnelle et le concept OLAP. Cette étude était indispensable pour la mise place d'un data mart concernant la paie des agents de l'Etat.

La mise en œuvre d'un data mart nécessite la connaissance et la maîtrise des concepts et méthodes étudiés. Aussi, comme tout projet informatique, pour que le projet réussisse, il faut :

- une maîtrise du sujet et une bonne organisation de l'équipe de projet ;
- une forte implication des futurs utilisateurs ;
- un engagement des responsables de l'entreprise.

Au regard des processus multiples qui peuvent exister dans l'entreprise, l'approche par les data marts pour la mise en place d'un système global d'aide à la décision de l'entreprise peut s'avérer une solution efficace. Cela revient à l'application de la théorie du « Diviser pour régner ».

ANNEXE I : Présentation du CENATRIN

Le Centre National de Traitement de l'Information (CENATRIN) a été créé en décembre 1970 et a pris son autonomie financière en 1975. Il est érigé en Etablissement Public à caractère Industriel et Commercial (EPIC) par le décret n°77/308/PRES/MF du 17 août 1977.

Le CENATRIN a été transformé en Société d'Etat au capital de 400 000 000 F CFA par le décret n°94/421/PRES/MICM/MEFP du 23 novembre 1994. Prestataire de services informatiques, il est sous la tutelle technique de la Délégation Générale à l'Informatique (DELGI). Le CENATRIN est en liquidation administrative depuis le 22/07/2005 par le décret n° 2005-406/PRES/PM/MCPEA/MFB.

1.1 Objectifs et missions

Le CENATRIN a pour objectifs :

- la conception et/ou la commercialisation de solutions adaptées aux besoins locaux basés sur une utilisation optimale des Nouvelles Technologies de l'Information et de la Communication (NTIC) ;
- la fabrication, le montage et/ou la commercialisation d'appareillages informatiques adaptés aux conditions locales d'utilisation ;
- la réalisation de tous travaux nécessaires aux traitements de l'information (développement de logiciels spécifiques, installation de réseaux locaux, installation de systèmes, gestion de centres informatiques, maintenance et dépannage d'appareillages informatiques), commandés par les services de l'Etat, les collectivités locales, les entreprises et organismes publics, parapublics et privés ;
- la formation des utilisateurs ;
- la formation des professionnels de l'informatique ;
- le recyclage et le perfectionnement des informaticiens.

1.2 Organisation de la société

Le CENATRIN comprend une Direction Générale, six (6) divisions et un centre à Bobo Dioulasso. Les principales divisions sont :

- la division Administrative et Financière ;
- la division Commerciale ;
- la division Etudes et Développement ;
- la division Exploitation ;
- la division Formation ;
- la division Système, Maintenance et Nouvelles Technologies de l'Information et de la Communication.

1.3 Activités

Entreprise de prestation de services, le CENATRIN a pour activités :

- la réalisation de schémas directeurs informatiques (SDI) ;
- les études préalables d'informatisation ;
- l'audit informatique ;
- les études et développement de logiciels standards et spécifiques ;
- la vente de logiciels ;
- l'assistance et le conseil en informatique ;
- la formation aux logiciels standards ;
- la saisie de données et le traitement de données ;
- le conditionnement (déliassage, massicotage, reliure) ;
- la vente d'équipements informatiques (matériel, consommables) ;
- la formation à la micro-informatique (Bureautique, Internet, etc.) ;
- la vente de pièces de rechange ;
- la maintenance à la carte et sur contrat ;
- l'installation et configuration d'équipements informatiques et réseautiques ;
- le câblage réseaux ;
- la fourniture d'accès à Internet ;

- la réalisation et hébergement de sites WEB ;
- la gestion des cybercafés.

1.4 Matériel et logiciel

1.4.1 Matériel

Le CENATRIN, dans la recherche d'une spécialisation dans la vulgarisation des nouvelles technologies, s'est doté d'équipements informatiques de divers constructeurs et de plate-forme variée, notamment des micro-ordinateurs Pentium III et IV, des imprimantes à laser, des serveurs, des routeurs, des onduleurs, des modems, etc.

Egalement on y trouve un mini-ordinateur BULL (DPS7000), des imprimantes à aiguille, des dérouleurs de bandes magnétiques BULL,...

1.4.2 Logiciel

Sont utilisés au CENATRIN les logiciels suivants :

- systèmes d'exploitation : Windows 95, Windows 98, Windows 2000 et NT, Unix, Netware de Novell, Linux ;
- développement : Access, Visual Basic, Delphi, Oracle, Hyper Text Markup Language (HTML), AMC*Designor ;
- bureautique: Word, PowerPoint, Excel, Internet Explorer.

Les logiciels suivants ont été développés :

- SGPA (Système Intégré de Gestion de Paie et Administrative) ;
- KIRA_WIN (Gestion de courrier) ;
- LAFIA (Gestion de stock de pharmacie) ;
- GESTBUD (Gestion Budgétaire des administrations Publiques) ;
- GIMAT (Gestion Intégrée de Matériel) ;
- FASO COMPTA (Comptabilité Générale);
- SGCH (Système de Gestion des Centres Hospitaliers) ;
- GestCom (Gestion Commerciale) ;

- GestImmo (Gestion des Immobilisations) ;
- Système d'Information et de Gestion de l'Office National des Télécommunications (S.I.G ONATEL sous ACCESS);
- Système Intégré de Gestion Administrative et Salariale du Personnel de l'Etat (SIGASPE volet salarial sous ORACLE);
- Système d'information de la Caisse Autonome de Retraite des Fonctionnaires (sous ORACLE);
- Logiciel de gestion des impôts pour la Direction Générale des Impôts (DBASE) ;
- Comptabilité Intégrée de l'Etat (CIE sous ORACLE) ;
- Système Intégré de Gestion Administrative et Salariale des Armées (SIGASA sous *ORACLE*).

ANNEXE II : Abréviations

API	Application Program Interface
CENATRIN	Centre National de Traitement de l'Information
DOLAP	Desktop OLAP
DOS	Disk Operating System
GCOS	General Comprehensive Operating System
GCL	GCOS Command Language
HOLAP	Hybrid OLAP
JCL	Job Control Language
KJOB	Knowned Job
MOLAP	Multidimensionnal OLAP
ODBC	Open DataBase Connectivity
OLAP	On Line Analytical Processing
OLTP	On Line Transaction Processing
PL/SQL	Procedural Language/Structured Query Language
ROLAP	Relational OLAP
RON	Run Occurrence Number
SGBD	Système de Gestion de Base de Données
SIGASPE	Système Intégré de Gestion Administrative et Salariale du Personnel de l'Etat
SQL	Structured Query Language
UML	Unified Modeling Language
VLSI/CMOS	Very Large Scale Integration / Complementary Metal Oxyde Semiconductor

ANNEXE III : Bibliographie

Ralph KIMBALL, Laura REEVES, Margy ROSS, Warren THORNTHWAITE – Le data warehouse : guide de conduite de projet, Eyrolles, deuxième édition, février 2005.

Ralph KIMBALL, Margy ROSS – Entrepôts de données : guide pratique de modélisation dimensionnelle, Vuibert, deuxième édition 2003.

Georges GARDARIN – Internet/intranet et base de données, Eyrolles, première édition 1999.

BULL - Bull DPS7 : gestion du système GCOS7, 1988.

BULL – Migration pupitrage GCOS7_v2 vers GCOS7_v3.

Cours de base de données – Geneviève JOMIER Avril2005.

ANNEXE IV : Ressources internet

<http://www-adele.imag.fr/~donsez/cours/bdcpdw.pdf>

Cours sur la modélisation dimensionnelle.

<http://perso.wanadoo.fr/bernard.lupin/>

Site web où l'on peut trouver des informations sur le concept OLAP.

<http://www.olapreport.com/>

Site web entièrement consacré à OLAP.

<http://www.xrobion.com/produits.htm>

Page web présentant le produit Oracle Express.

<http://www.quod.ch/webquod/dwh01.htm>

Page web sur le data warehouse.

<http://www.epfl.ch/f/teaching/courses/SlidesIBD/DW.pdf> –

Document PDF sur le data warehouse.

ANNEXE V : Glossaire

Attribut : un fait décrivant chaque position d'une dimension.

Base de données : est un ensemble structuré et organisé permettant le stockage de grandes quantités d'informations afin d'en faciliter l'exploitation.

Data mart : l'ensemble des données se rapportant à un des métiers de l'entreprise.

Data warehouse : entrepôt de données. Ce terme anglais est utilisé pour désigner l'ensemble des informations d'une entreprise, enregistrées sous un format informatique.

DOLAP : Desktop OLAP. Désigne un petit produit OLAP faisant de l'analyse multidimensionnelle en local.

Dimension : un ensemble de données du même type, permettant de structurer la base multidimensionnelle.

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

HOLAP : Hybrid OLAP. Désigne les outils d'analyse multidimensionnelle qui récupèrent les données dans des bases relationnelles ou multidimensionnelles, de manière transparente pour l'utilisateur.

MOLAP : Multidimensionnal OLAP. Ce terme désigne plus spécifiquement une technologie de stockage plus cartésien. Les jointures entre tables de dimension et de faits sont déjà faites, ce qui explique les performances.

Multidimensionnel : structure de données ayant au moins trois dimensions indépendantes.

OLAP : On Line Analytical Processing. Désigne une catégorie d'applications et de technologies permettant de collecter, stocker, traiter et restituer des données multidimensionnelles, à des fins d'analyse.

OLTP : On Line Transaction Processing. Terme associé à l'origine aux activités et aux applications servant à entrer de manière fiable des données dans une base de données relationnelles.

SGBDR : Système de Gestion de Base de Données Relationnelle. Permet le stockage, le traitement et la restitution de données stockées dans des tables relationnelles.

ROLAP : Relationnal OLAP. Cette technique permet de faire de l'analyse multidimensionnelle à partir de données stockées dans des tables relationnelles.

ANNEXE VI : Liste des tableaux et schémas

Tableau 1: Différences entre données du système de production et données décisionnelles...	16
Tableau 2: différences entre système de production et entrepôt de données	19
Tableau 3: tableau de comparaison entre infocentre et entrepôt de données	22
Tableau 4: matrice de bus des assurances	26
Tableau 5: finalités du data mart et de l'entrepôt de données	34
Tableau 6: matrice de bus d'un data mart de la paie du personnel de l'Etat.....	60
Tableau 7 : description détaillée des attributs de la dimension periode.....	63
Tableau 8: description détaillée des attributs de la dimension fonction	64
Tableau 9: description détaillée des attributs de la dimension lieu	65
Tableau 10: description détaillée des attributs de la dimension banque/mode paiement	66
Tableau 11: description détaillée des attributs de la dimension élément de salaire	68
Tableau 12: description détaillée des attributs de la dimension agent.....	69
Tableau 13: description détaillée des attributs de la dimension qualité/sit. matrimoniale.....	70
Tableau 14: description détaillée des attributs de la dimension billeteur	71
Tableau 15: description détaillée des attributs de la dimension section/service	72
Tableau 16 : description détaillée des attributs de la dimension classification	74
Schéma 1 : Composants de base de l'entrepôt de données	25
Schéma 2: composant d'un data mart dans le cadre de la paie du personnel de l'Etat.....	57
Schéma 3: diagramme de la table des faits.....	61
Schéma 4: diagramme de détail de la table dimensionnelle periode	64
Schéma 5: diagramme de détail de la table dimensionnelle fonction	65
Schéma 6: diagramme de détail de la table dimensionnelle lieu	66
Schéma 7: diagramme de détail de la table dimensionnelle banque/mode paiement	67
Schéma 8: diagramme de détail de la table dimensionnelle élément de salaire	68
Schéma 9: diagramme de détail de la table dimensionnelle agent	70
Schéma 10: diagramme de détail de la table dimensionnelle qualite/situation matrimoniale..	71
Schéma 11: diagramme de détail de la table dimensionnelle billeteur	72
Schéma 12: diagramme de détail de la table dimensionnelle section/service	73
Schéma 13: diagramme de détail de la table dimensionnelle classification.....	74
Schéma 14: modèle physique en étoile	75
Schéma 15 : modèle multidimensionnel	76