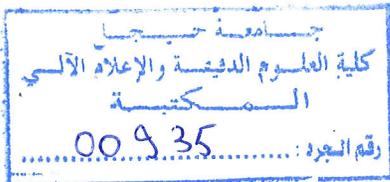


Inf.SIAD.04/18

République Algérienne Démocratique et populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Seddik Ben yahia de Jijel
Faculté des Sciences Exactes et Informatique
Département d'informatique



Mémoire de fin d'études

pour l'obtention du diplôme Master de Recherche en Informatique
Option : Système d'information et aide à la décision

Thème

Optimisation de la gestion hospitalière à
l'aide d'un
entrepôt de données actif

Élaboré et présenté par :
Doufar Sabah
Birouk Ines

Encadré par :
Mme.Bouainah Madiha

Promotion : 2018

Remerciements

*Nous tenons tout d'abord à remercier le **Dieu** tout puissant et miséricordieux, qui nous a donné la force et la patience d'accomplir ce travail.*

*Nous tenons aussi à exprimer notre sincère reconnaissance et notre profonde gratitude à notre encadreur M^{me} "**Bouainah Madiha**" pour son soutien, sa disponibilité, ses orientations, ses précieux conseils et ses encouragements qui nous ont permis d'élaborer ce travail dans des bonnes conditions.*

*Nos vifs remerciements également aux **membres du jury** pour leurs attentions et intérêts portés envers notre travail. Merci de nous avoir honorés de votre présence.*

*Sans oublier de présenter nos sincères remerciements à nos **parents** qui ont été toujours de nous encourager durant notre parcours de mes études, ainsi que pour leurs aides, leurs compréhensions et leurs soutiens.*

Notre remerciements aussi au corps professoral et administratif de la Faculté des Sciences Exactes et Informatique "Département d'informatique", pour la richesse et la qualité de leur enseignement et qui ont déployé des efforts au profit de leurs étudiants.

Nous tenons à remercier tout le personnel de l'hôpital de Jijel qui a mis à notre disposition tous les moyens nécessaires afin d'accomplir notre stage.

Nous souhaitant adresser nos remerciements les plus sincères aux personnes qui nous ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire.

Enfin, nous adressons nos plus sincères remerciements à tous nos proches et amis, qui nous ont toujours encouragées au cours de la réalisation de ce mémoire.

Merci à tous et à toutes.

Dédicaces

Tous les mots ne sauraient exprimer la gratitude, l'amour, le respect, la reconnaissance, c'est tout simplement que : Je dédie ce mémoire de fin de cycle de Master 2 à :

*Mon très cher **Père Ahmed** : Aucune dédicace ne saurait exprimer l'amour, l'estime et le respect que j'ai toujours pour vous. Rien au monde ne vaut les efforts fournis jour et nuit pour mon éducation et mon bien être. Ce travail et le fruit de vos sacrifices que vous avez consentis pour mon éducation et ma formation le long de ces années.*

*Ma tendre **Mère Naïma** : Tu représentes pour moi le symbole de la bonté par excellence, la source de tendresse et l'exemple du dévouement qui n'a pas cessé de m'encourager et de prier pour moi. Ta prière et ta bénédiction m'ont été d'un grand secours pour mener à bien mes études. Aucune dédicace ne saurait être assez éloquente pour exprimer ce que tu mérites pour tous les sacrifices que tu n'as cessé de me donner depuis ma naissance, durant mon enfance et même à l'âge adulte. Tu as fait plus qu'une mère puisse faire pour que ses enfants suivent le bon chemin dans leur vie et leurs études. Je te dédie ce travail en témoignage de mon profond amour. Puisse Dieu, le tout puissant, te préserver et t'accorder santé, longue vie et bonheur.*

*Madame **Bouainah Madiha** qui ne cesse pas de m'encourager et me conseillée. Cette humble dédicace ne saurait exprimer mon grand respect et ma profonde estime.*

À la mémoire des défunts du mes chers grands parents.

Mes très chers frères pour leurs appuis et leurs encouragements.

Mes très chères sœurs pour leurs encouragements permanents, et leurs soutiens morales.

Mes chers neveux.

Mon beau-frère.

Mes oncles et tantes et à toute ma famille.

*Mon binôme **Ines**.*

Tous mes enseignants depuis ma première année d'études.

Mes très chères amies.

Tous les membres de ma promotion.

Tous ceux qui me sentent chers et que j'ai omis de citer.

Merci d'être toujours là pour moi

Sabah

Dédicaces

Tous les mots ne sauraient exprimer la gratitude, l'amour, le respect, la reconnaissance, c'est tout simplement que : Je dédie ce mémoire de fin de cycle de Master 2 à :

*Mon très cher **Père** : Aucune dédicace ne saurait exprimer l'amour, l'estime et le respect que j'ai toujours pour vous. Rien au monde ne vaut les efforts fournis jour et nuit pour mon éducation et mon bien être. Ce travail et le fruit de vos sacrifices que vous avez consentis pour mon éducation et ma formation le long de ces années.*

*Ma tendre **Mère** : Tu représentes pour moi le symbole de la bonté par excellence, la source de tendresse et l'exemple du dévouement qui n'a pas cessé de m'encourager et de prier pour moi. Ta prière et ta bénédiction m'ont été d'un grand secours pour mener à bien mes études. Aucune dédicace ne saurait être assez éloquente pour exprimer ce que tu mérites pour tous les sacrifices que tu n'as cessé de me donner depuis ma naissance, durant mon enfance et même à l'âge adulte. Tu as fait plus qu'une mère puisse faire pour que ses enfants suivent le bon chemin dans leur vie et leurs études. Je te dédie ce travail en témoignage de mon profond amour. Puisse Dieu, le tout puissant, te préserver et t'accorder santé, longue vie et bonheur.*

*Madame **Bouainah Madiha** qui ne cesse pas de m'encourager et me conseillée. Cette humble dédicace ne saurait exprimer mon grand respect et ma profonde estime.*

Mes chères grands mères.

Mon très cher frère pour son appui et son encouragement.

Mes très chères sœurs pour leurs encouragements permanents, et leurs soutiens morales.

Mon fiancé pour son soutien et son encouragement.

Mes oncles et tantes et à toute ma famille.

*Mon binôme **Sabah**.*

Tous mes enseignants depuis ma première année d'études.

Mes très chères amies.

Tous les membres de ma promotion.

Tous ceux qui me sentent chers et que j'ai omis de citer.

Merci d'être toujours là pour moi

Ines



Table des matières

Table des matières	i
Liste des tableaux	v
Table des figures	vii
Liste des abréviations	viii
Introduction générale	1
1 Les systèmes décisionnels à base d'entrepôt de données	4
1.1 Introduction	4
1.2 Les concepts de base	5
1.2.1 Système décisionnel	5
1.2.1.1 Définition d'un système décisionnel	5
1.2.1.2 Les composants d'un système décisionnel à base d'entrepôt	5
1.3 Modélisation multidimensionnelle	7
1.3.1 Faits et mesures	8
1.3.2 Dimensions et hiérarchies	8
1.3.3 Modélisation conceptuelle	9
1.3.3.1 Le schéma en étoile	9
1.3.3.2 Le schéma en flocon de neige	10
1.3.3.3 Le schéma en constellation (mixte)	11
1.3.4 Modélisation logique	12
1.3.4.1 ROLAP (Relational OLAP)	12
1.3.4.2 OOLAP (Object OLAP)	12

1.3.4.3	MOLAP (Multidimensionnel OLAP)	13
1.4	Les systèmes d'entrepôts de données	13
1.4.1	Architecture générale d'un système d'entrepôt de données	13
1.4.2	Entrepôts et magasins de données	14
1.4.3	Les différentes Architectures d'entrepôts de données	15
1.4.3.1	Les data marts (magasin de données) indépendants	15
1.4.3.2	Architecture en bus de magasins de données	16
1.4.3.3	Architecture Hub-and-spoke	16
1.4.3.4	Entrepôt de données centralisé	17
1.4.3.5	Architecture fédérée	18
1.4.4	Réalisation d'un entrepôt de données	18
1.4.5	Comparaison entre les deux approches	19
1.5	Les opérations OLAP dans le modèle de données multidimensionnel	20
1.6	Les objectifs d'un entrepôt de données	20
1.7	Les types d'entrepôts de données	22
1.8	Conclusion	23
2	Les entrepôts de données actifs	24
2.1	Introduction	24
2.2	Les bases de données actives	24
2.2.1	Définition	24
2.2.2	Modèle de règles ECA	25
2.3	Entrepôts de données actifs	25
2.3.1	Définition	25
2.3.2	Architecture d'un EDA	26
2.3.3	Règles d'analyses	26
2.3.3.1	Approche par objet	27
2.3.3.2	Approche multidimensionnelle	27
2.3.4	Analyse multidimensionnelle	27
2.3.5	Graphes d'analyse	28
2.3.6	Taxonomie d'événements dans un entrepôt de données actif	29

2.3.6.1	Les événements OLTP	30
2.3.6.2	Les événements OLAP	30
2.3.7	Comparaison entre l'entrepôt de données actif et passif	30
2.4	Conclusion	31
3	La Conception du Système	32
3.1	Introduction	32
3.2	Domaine d'application	32
3.2.1	Présentation d'organisme d'accueil	32
3.2.2	L'organigramme de l'hôpital de Jijel	33
3.3	Architecture générale du système	36
3.4	La conception des bases de données sources(Base_Hôpital)	37
3.4.1	Diagramme de cas d'utilisation du système	37
3.4.2	Description textuelle de cas d'utilisation	38
3.4.3	Schéma conceptuel des bases de données sources(Base_Hôpital)	40
3.4.4	Le modèle relationnel	42
3.5	Conception de l'entrepôt de données (Absence_schéma)	42
3.5.1	Les dimensions et les hiérarchies	42
3.5.1.1	La dimension DIMENSION_EMPLOYE	42
3.5.1.2	La dimension DIMENSION_TEMPS	43
3.5.1.3	La dimension DIMENSION_CONGE	44
3.5.1.4	La dimension DIMENSION_SERVICE	44
3.5.1.5	La dimension DIMENSION_SPECIALITE	44
3.5.2	La table des faits	44
3.6	Conception des règles d'analyses	47
3.6.1	La syntaxe générale d'une règle d'analyse	47
3.6.2	Les règles d'analyses	47
3.6.2.1	Règle 1	47
3.6.2.2	Règle 2	48
3.6.2.3	Règle 3	48
3.6.2.4	Règle 4	48
3.7	Conclusion	49

4 L'implémentationl et réalisation	50
4.1 Introduction	50
4.2 Les outils d'implémentation	50
4.2.1 NetBeans	50
4.2.2 Oracle Database 11g Express Edition	51
4.3 Présentation de l'application	51
4.3.1 Page d'authentification	51
4.3.2 Page d'accueil	52
4.3.2.1 Bases de données sources	53
4.3.2.2 Entrepôt de données passif	54
4.3.2.3 Page de qualité de service	55
4.3.2.3.1 Page d'analyse avancée	56
4.3.2.4 Page de gestion des employés	58
4.3.2.4.1 Formulaire d' ajout d'un nouvel employé	59
4.4 Conclusion	60
Conclusion générale	61
Bibliographie	62

Liste des tableaux

1.1	Magasin de données vs Entrepôt de données.	15
1.2	comparaison entre Inmon et Kimball.	19
2.1	Comparaison entre l'entrepôt de données actif et passif	31
3.1	Les différents services de l'hôpital de Jijel	35
3.2	Description textuelle de cas d'utilisation	40
3.3	modélisation d'entrepôt de données	42
3.4	EMPLOYEE_D	43
3.5	SEXE_D	43
3.6	PERIODE_D	43
3.7	MOIS_D	43
3.8	ANNEE_D	43
3.9	CONGE_D	44
3.10	SERVICE_D	44
3.11	SPECIALITE_D	44
3.12	La table des faits (ABSENCE_F)	45

4.2	Fenêtre du menu principale de l'application.	53
4.3	Espace de base de données source.	54
4.4	Espace d'entrepôt de données.	55
4.5	Fenêtre qualité de service.	56
4.6	Fenêtre d'analyse avancée	57
4.7	l'absence des employés au niveau des services au mois de janvier 2018.	58
4.8	Fenêtre de gestion des employés.	58
4.9	Fenêtre de saisir un nouvel employé.	59
4.10	Fenêtre de cycle actif.	60

Liste des abréviations

SID	Système d'information Décisionnel
SIO	Système d'information Opérationnel
DW	Data Warehouse
DM	Data Mart
ETL	Extract Transaction Load
BDD	Base de Données
OLTP	On-Line Transaction Processing
OLAP	On-line Analytical Processing
SGBD	Système de Gestion de Base De Données
ROLAP	Relational On-Line Analytical Processing
OOLAP	Object On-Line Analytical Processing
MOLAP	Multidimensional On-Line Analytical Processing
ED	Entrepôt de Données
BI	Business Intelligence
EDA	Entrepôt de Données Actif
ECA	Événement-Condition-Action
EPH	Établissement Public Hospitalier
UML	Unified Modeling Language
DRH	Directeur des Ressources Humaines
IDE	Environnement de Développement Intégré
CDDL	Common Development and Distribution License

Introduction générale

Le système d'information décisionnel(SID)par opposition à un système d'information opérationnel(SIO), dont l'objectif est l'exécution d'un processus métier, un SID a pour but l'évaluation de la performance des processus. Il a pour vocation de faciliter la prise de décision en fournissant des réponses à des questions telles que : quelle fut l'évolution du chiffre d'affaires et de la marge brute pour chaque catégorie de produits entre le premier semestre de cette année et celui de l'année précédente ?

Un SID est donc un système d'information dédié aux décideurs d'une organisation et permettant, au moyen d'une base de données et d'une interface d'accès aux données, aux utilisateurs d'obtenir des informations utiles à la prise de décision.

Les premiers SID conçus au sein des entreprises étaient des systèmes avec très peu d'utilisateurs, élaborés parfois par les utilisateurs eux-mêmes, exploitant le plus souvent les données d'un seul SIO. Ces systèmes étaient la plupart du temps composés uniquement d'un extracteur de données et d'une petite base de données ou d'un tableur. Puis, au début des années 1990, les concepts de data warehouse(DW) et de data mart(DM) firent leur apparition, popularisés par des auteurs tels que Bill Inmon et Ralph Kimball. Aujourd'hui, le data warehouse et/ou les data marts forment le cœur de tout système d'information décisionnel digne de ce nom.

Les travaux sur les entrepôts de données ont atteint une certaine maturation à l'heure actuelle. Une des tendances actuelles des entrepôts de données est d'automatiser les scénarii d'analyse afin d'accélérer la prise de décision et afin de doter l'entrepôt d'un aspect actif.

Dans ce contexte , Les entrepôts de données actifs apportent des solutions pour pallier les insuffisances dont souffrent les entrepôts de données conventionnels. Ils permettent justement d'encapsuler des actions afin d'obtenir automatiquement par exemple des informations pour la prise de décision ou d'automatiser des tâches de rafraîchissement au niveau de l'ETL (Extract Transform Load).

le chapitre 3, est consacré aux solutions proposées pour optimiser le système de gestion hospitalière de Jijel. En effet, nous proposons dans ce chapitre les différents étapes de la conception de notre système permet d'apporter une solution à notre problème.

Enfin, nous détaillons dans le dernier chapitre, les applications du système adopté pour démontrer la validité des solutions proposées.

Nous terminons notre mémoire par une conclusion générale dans laquelle nous résumons tous les résultats obtenus.

CHAPITRE 1

Les systèmes décisionnels à base d'entrepôt de données

1.1 Introduction

Toutes les entreprises du monde disposent d'une masse de données plus ou moins considérable. Ces informations proviennent soit de sources internes (générées par leurs systèmes opérationnels au fil des activités journalières), ou bien de sources externes (web, partenaire, ... etc.).

Cette surabondance de données, et l'impossibilité des systèmes opérationnels de les exploiter à des fins d'analyse conduit, inévitablement, l'entreprise à se tourner vers une nouvelle informatique dite décisionnelle qui met l'accent sur la compréhension de l'environnement de l'entreprise et l'exploitation de ces données à bon escient.

En effet, les décideurs de l'entreprise ont besoin d'avoir une meilleure vision de leur environnement et de son évolution, ainsi, que des informations aux quelles ils peuvent se fier. Cela ne peut se faire qu'en mettant en place des indicateurs clairs et pertinents permettant la sauvegarde, l'utilisation de la mémoire de l'entreprise et offrant à ses décideurs la possibilité de se reporter à ces indicateurs pour une bonne prise de décision.

Un système décisionnel est donc avant tout un moyen qui a pour but de faciliter la définition et la mise en œuvre de stratégies gagnantes. L'entreprise a besoin d'outils qui lui permettent de déceler tous ces éléments. Parmi les outils décisionnels les systèmes d'entrepôt de données qui apparut pour la première fois en 1980, à fin des années 90 le système décisionnel est devenu basé sur l'utilisation des entrepôts de données. Sont des outils précieux dans le monde compétitif et en constante évolution d'aujourd'hui. L'idée consistait alors à réaliser une base de données destinée exclusivement au processus décisionnel. Les nouveaux besoins de l'entreprise, les quantités importantes de données produites par les systèmes opérationnels et l'apparition des technologies aptes à sa mise en œuvre ont contribué à l'apparition du concept « Data Warehouse » comme support aux systèmes décisionnels.

Au cours des dernières années, de nombreuses entreprises ont dépensé des millions de dollars

pour construire des entrepôts de données à l'échelle de l'entreprise. Nombreux sont ceux qui pensent qu'avec l'augmentation de la concurrence dans toutes les industries, l'entreposage de données est la dernière arme de marketing indispensable - un moyen de fidéliser les clients en leur permettant d'en savoir plus sur leurs besoins.

1.2 Les concepts de base

1.2.1 Système décisionnel

1.2.1.1 Définition d'un système décisionnel

Au début des travaux sur ce domaine, en 1996, le système décisionnel (SID) a été défini par l'un des pionniers dans ce domaine, le SID est : « *une hiérarchie d'espaces de stockage des données comprenant les sources de données jusqu'aux espaces appelés magasins de données contenant des données très agrégées* ». [1]

Une autre définition donnée dans [1] : « *Un système d'information décisionnel, SID, est un système qui réalise la collecte, la transformation des données brutes issues de sources de données et le stockage dans d'autres espaces ainsi que la caractérisation des données résumées en vue de faciliter le processus de prise de décision* ».

1.2.1.2 Les composants d'un système décisionnel à base d'entrepôt

Un système décisionnel est un système d'information dédié aux applications décisionnelles. Les composants d'un système décisionnel à base d'entrepôt de données mettent en jeu quatre éléments essentiels : les sources de données, l'entrepôt de données, les magasins de données et les outils d'analyse et d'interrogation. Voir la (figure 1.1) :

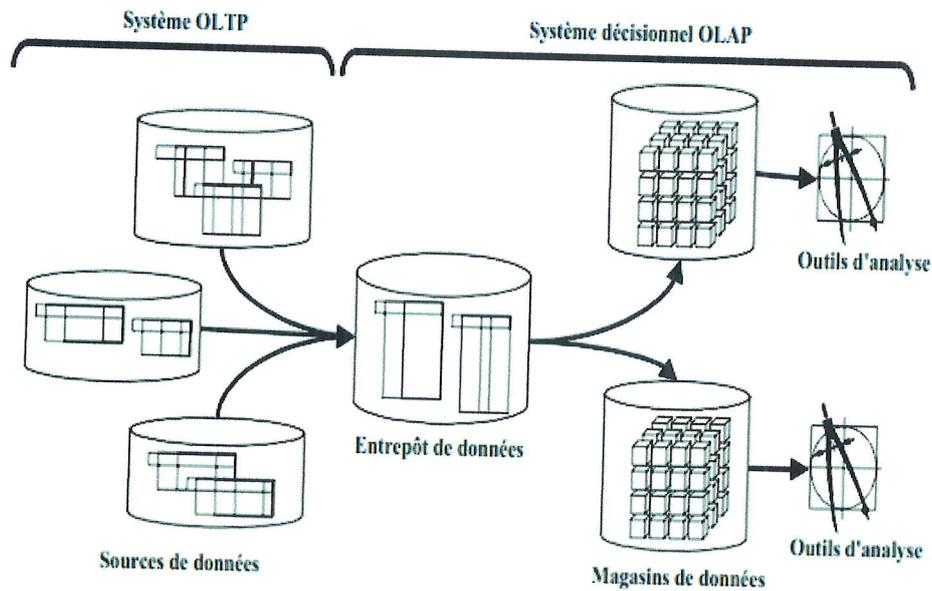


FIGURE 1.1 – Architecture des systèmes décisionnels.
[2]

- **Les sources de données** : sont nombreuses, variées, distribuées et autonomes. Elles peuvent être internes (bases de production) ou externes (Internet, bases des partenaires) à l'entreprise.[2]
- **Entrepôt de données (Data Warehouse)** : est le lieu de stockage centralisé des informations utiles pour les décideurs. Il met en commun les données provenant des différentes sources et conserve leurs évolutions.[2]

Plusieurs définitions ont été données au concept d'entrepôt de données. Nous avons retenu la définition suivante :

Selon Inmon (1994) : « *Un entrepôt de données est une collection de données orientées sujet, Intégrées, non volatiles et historisées, organisées pour le support d'un processus d'aide à la décision.* » [4] Cette définition met l'accent sur les caractéristiques suivantes :

- ◆ **orientées sujet** : la première caractéristique de l'entrepôt de données est qu'il est orienté autour des principaux sujets de l'entreprise. Les données sont structurées autour de thèmes ce qui facilite l'analyse, on regroupe les différents sujets dans une structure commune.[5]

- ◆ **intégrées** : c'est l'aspect le plus important de l'environnement d'entrepôt de données est celui des données trouvées dans l'entrepôt de données est intégré. L'intégration apparaît de différentes manières, une forte normalisation, une bonne gestion de cohérence (des variables, dans des structures de codage, attributs physiques des données ...etc.), Il faut doter ces données afin qu'elles puissent aisément s'intégrer dans le Data Warehouse.[5]
- ◆ **non volatiles** : l'entrepôt de données est qu'il est non volatile. A chaque fois lorsqu'on exécute une requête dans des différentes dates sur les mêmes données on doit avoir toujours les mêmes résultats. Donc il assure une fiabilité des résultats.[5]
- ◆ **historisées** : la prise en compte de l'évolution des données est essentielle pour la prise de décision qui, par exemple, utilise des techniques de prédiction en s'appuyant sur les évolutions passées pour prévoir les évolutions futures. [5]
- **Les magasins de données (Data marts)** : sont des extraits de l'entrepôt orientés sujet. Les données sont organisées de manière adéquate pour permettre des analyses rapides à des fins de prise de décision.[2]

Plusieurs définitions ont été proposées au concept de magasin de données parmi eux :

Le magasin de données (data mart) : *« est un extrait de l'entrepôt. Les données extraites sont adaptées à une classe de décideurs ou à un usage particulier (recherche de corrélation, logiciel de statistiques,...). L'organisation des données suit un modèle spécifique qui facilite les traitements décisionnels. »*[2]

Magasin de données : *« est un ensemble cohérent de données répondant aux besoins spécifiques d'une classe de décideurs. C'est un sous ensemble de l'entrepôt de données propre à un groupe de décideurs. »*[1]

- **Les outils d'analyse** : permettent de manipuler les données suivant des axes d'analyses. L'information est visualisée au travers d'interfaces interactives et fonctionnelles dédiées à des décideurs souvent non informaticiens (directeurs, chefs de services,...).[2]

1.3 Modélisation multidimensionnelle

Les modèles de conception des systèmes transactionnels OLTP ne sont pas adaptés aux systèmes OLAP dont les requêtes sont souvent très complexes. Pour ce type d'environnement OLAP, une nouvelle approche de modélisation a été proposée : la modélisation multidimensionnelle. Popularisée par Ralph Kimball dans les années 90, cette modélisation est aujourd'hui reconnue comme la modélisation la plus appropriée aux besoins d'analyse et

de prise de décision.[6]

Le modèle dimensionnel est basé sur la dualité des concepts fait-dimension :

1.3.1 Faits et mesures

Fait : un modèle de données multidimensionnel est organisé autour d'un thème central que l'on veut étudier. Ce thème est modélisé par une table de faits. Un fait est formé de mesures (numériques) correspondant aux informations de l'activité analysée. Ainsi que les clés de chacun des tables de dimension.[2]

Mesure : les mesures d'un fait représentant les différentes valeurs de l'activité analysée. Ils peuvent être numériques pour permettre de résumer un grand nombre d'enregistrements en quelques enregistrements (on peut les additionner, les dénombrer ou bien calculer le minimum, le maximum ou la moyenne). Les mesures sont valorisées de façon continue car il est important de ne pas valoriser le fait avec des valeurs nulles. Les mesures à des types nous trouvons des mesures additives ou semi-additives afin de pouvoir les combiner au moyen d'opérateurs arithmétiques.[2]

1.3.2 Dimensions et hiérarchies

Dimension : Les dimensions représentent les axes de l'analyse multidimensionnelle. Les dimensions sont les perspectives ou entités se compose de paramètres (ou bien d'attributs) correspondant aux informations faisant varier les mesures de l'activité. Les paramètres peuvent être textuels et discrets. Les paramètres textuels sont utilisés pour restreindre la portée des requêtes afin de limiter la taille des réponses. Les paramètres sont discrets, c'est à dire que les valeurs possibles sont bien déterminées et sont des descripteurs constants. Chaque dimension peut avoir une table ou plusieurs associée à-il. Il a appelé hiérarchie.[2]

Les dimensions représentent les axes de l'analyse multidimensionnelle. Elles sont organisées en schémas hiérarchiques. Un schéma de hiérarchie, composé par plusieurs niveaux, représente différentes granularités ou degrés de précision de l'information.[19]

Hiérarchie : représente les paramètres d'une dimension selon leur niveau (de granularité) ou de détail. Les paramètres sont ordonnés par une relation "est plus fin".[2]
C'est-à-dire composé par plusieurs niveaux, représente différentes granularités ou degrés de précision de l'information.

1.3.3 Modélisation conceptuelle

Le modèle de données entité-relation est couramment utilisé dans la conception de relations bases de données, un entrepôt de données nécessite un schéma concis et orienté sur le sujet qui facilite l'analyse des données en ligne, dans le modèle multidimensionnel au niveau conceptuel, Les modèles de conception sont totalement différents. Ils sont dé-normalisés. On en retient deux principaux : le schéma en étoile (Star Schema) et le schéma en flocon (Snowflake Schema) et le moins utilisée le Schéma en constellation aussi appelée le schéma mixte (étoile /flocon de neige).[3]

Trois schémas sont utilisés pour la modélisation multidimensionnelle :

- Le schéma en étoile.
- Le schéma en flocon de neige .
- Le schéma en constellation .

1.3.3.1 Le schéma en étoile

Le paradigme de modélisation le plus courant est le schéma en étoile, dans lequel l'entrepôt de données contient une grande table centrale (table de faits) contenant la plus grande partie des données, sans redondance, et un ensemble de tables auxiliaires plus petites (dimension tables), un pour chaque dimension. Le graphique du schéma ressemble à une étoile, avec les tables de dimension affichées dans un motif radial autour de la table de faits centrale. Comme dans la(figure 1.2) montre un schéma en étoile d'une activité de ventes. Par exemple dans la table de fait ventes l'analyse les montants et Prix Unitaires des ventes par temps, par produit, par client et par magasin.[6]



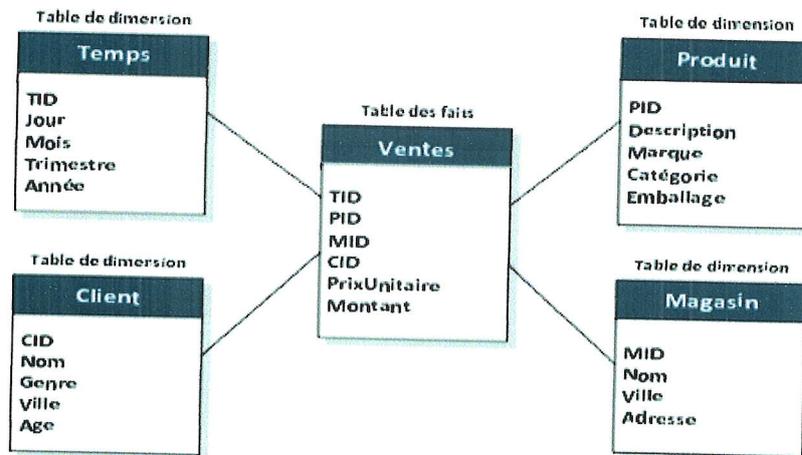


FIGURE 1.2 – Schéma en étoile.
[6]

1.3.3.2 Le schéma en flocon de neige

Le schéma de flocon de neige est une variante du modèle de schéma en étoile. La dé-normalisation des tables de dimension dans un schéma en étoile ne reflète pas les hiérarchies associées à chaque dimension. Pour mettre en évidence cette hiérarchie, le modèle en flocon de neige a été proposé. Chaque table de dimension est éclatée en un ensemble de hiérarchies. Ce schéma normalise les dimensions, réduisant la taille de chacune des relations et permettant ainsi de formaliser la notion de hiérarchie au sein d'une dimension. Les tables représentant la hiérarchie la plus fine sont directement liées à la table des faits. Les tables représentant les autres hiérarchies sont liées entre elles selon leur niveau dans cette hiérarchie. Comme dans la (figure 1.3) montre un schéma en flocon de neige. Sur ce schéma, en descendant un Hiérarchie de concepts pour le temps défini comme "jour < mois < trimestre < année." Et un Hiérarchie de concepts pour le client défini comme "ville < région ". [6]

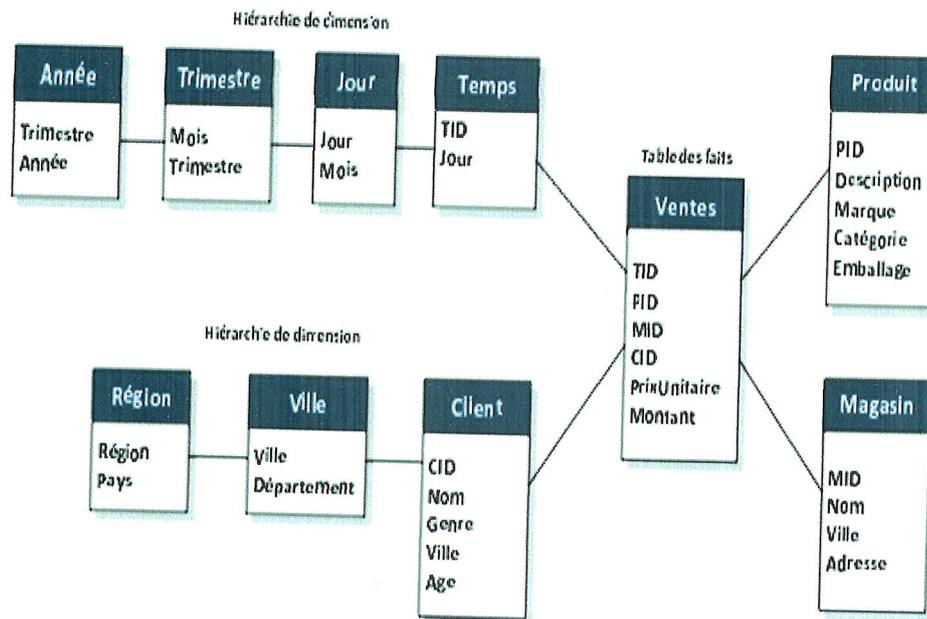


FIGURE 1.3 – Schéma en flocon de neige.
[6]

1.3.3.3 Le schéma en constellation (mixte)

Combine les deux schémas précédents (étoile & flocon de neige), Il s'agit de fusionner plusieurs modèles en étoile qui utilisent des dimensions communes. Un modèle en constellation comprend donc plusieurs faits et des dimensions communes ou non.[2]

La figure 1.4 illustre la modélisation en constellation ; nous décrivons une constellation constituée de deux schémas en étoile : l'un correspond aux ventes effectuées dans les pharmacies et l'autre analyse les prescriptions des médecins. Les dimensions Temps et Géographie sont partagées par les faits Prescriptions et Vente.

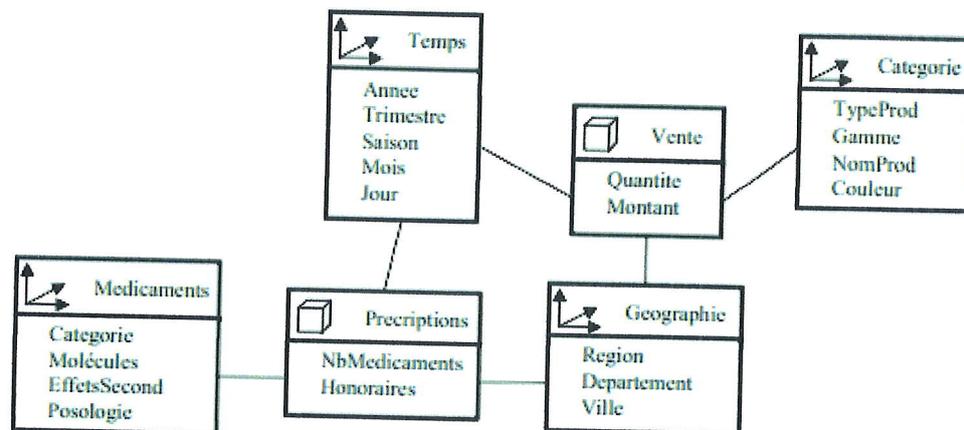


FIGURE 1.4 – modélisation en constellation.

[2]

1.3.4 Modélisation logique

Au niveau logique plusieurs possibilités sont envisageables pour la modélisation multidimensionnelle. Il est possible d'utiliser :

- un système de gestion de bases de données (SGBD) existant tel que les SGBD relationnels (ROLAP) ou bien les SGBD orientées objet (OOLAP).
- un système de gestion de bases de données multidimensionnelles (MOLAP).

1.3.4.1 ROLAP (Relational OLAP)

L'approche la plus couramment utilisée consiste à utiliser un système de gestion de bases de données relationnelles, on parle de l'approche ROLAP ("Relational On-Line Analytical Processing"). Le modèle multidimensionnel est alors traduit de la manière suivante :

- chaque fait correspond à une table, appelée table de fait.
- chaque dimension correspond à une table, appelée table de dimension. Ainsi, la table de fait est constituée d'attributs représentant les mesures d'activité et les attributs clés étrangères de chacune des tables de dimension. Les tables de dimension contiennent les paramètres et une clé primaire permettant de réaliser des jointures avec la table de fait.[2]

1.3.4.2 OOLAP (Object OLAP)

Plus récemment, une autre approche s'appuie sur le paradigme objet ; on parle de l'approche OOLAP ("Object On-Line Analytical Processing"). Le modèle multidimensionnel se traduit

ainsi :

- chaque fait correspond à une classe, appelée classe de fait.
- chaque dimension correspond à une classe, appelée classe de dimension.

Plusieurs de ces critères peuvent être simplement modélisés avec l'approche OOLAP, tels que le besoin d'explicitier les hiérarchies des dimensions, de supporter correctement l'agrégation et les relations multivaluées entre le fait et les dimensions, de gérer et manipuler simplement le temps. L'avantage de l'approche OOLAP par rapport à l'approche ROLAP est donc lié au niveau d'abstraction plus fort de l'objet qui prend en compte des concepts plus riches, permettant ainsi de modéliser facilement certaines caractéristiques des données multidimensionnelles.[2]

1.3.4.3 MOLAP (Multidimensionnel OLAP)

Une alternative à ces deux approches consiste à utiliser un système multidimensionnel "pur" qui gère des structures multidimensionnelles natives; on parle de l'approche MOLAP ("Multidimensional On-Line Analytical Processing"). Les structures multidimensionnelles natives utilisées sont des tableaux à n dimensions. Dans la littérature, les termes de cube, hyper cube et table multidimensionnelle sont utilisés de manière interchangeable. Nous utiliserons le terme d'hypercube pour désigner des structures à deux, trois ou à plus de trois dimensions. Cette approche permet de stocker les données de manière multidimensionnelle. L'intérêt est que les temps d'accès sont optimisés, mais cette approche nécessite de redéfinir des opérations pour manipuler ces structures multidimensionnelles.[2]

1.4 Les systèmes d'entrepôts de données

1.4.1 Architecture générale d'un système d'entrepôt de données

L'entrepôt de données joue un rôle stratégique dans la vie d'une entreprise. Il stocke des données pertinentes aux besoins de prise de décision en provenance des systèmes opérationnels de l'entreprise et d'autres sources externes. A la différence d'une base de données classique supportant des requêtes transactionnelles de type OLTP (On-Line Transaction Processing), un entrepôt de données est conçu pour supporter des requêtes de type OLAP (On-Line Analytical Processing). L'interrogation est l'opération la plus utilisée dans le contexte d'entrepôt de données où la mise à jour consiste seulement à alimenter l'entrepôt. Le processus de construction d'un entrepôt de données est composé de trois principales phases [6] :

- (1) extraction des données à partir des différentes sources.

- (2) organisation et intégration des données dans l'entrepôt.
- (3) accès aux données intégrées et analyse de ces dernières dans une forme efficace et flexible.

Dans la première et la deuxième phase, les données issues de différentes sources de données sont extraites, nettoyées et intégrées dans l'entrepôt de données. Les métadonnées contiennent des informations utiles sur la création, l'utilisation et la gestion de l'entrepôt. Durant la troisième phase, un serveur OLAP se charge de présenter les informations demandées par les utilisateurs sous plusieurs formes : tableaux, rapports, statistiques, etc.[6]

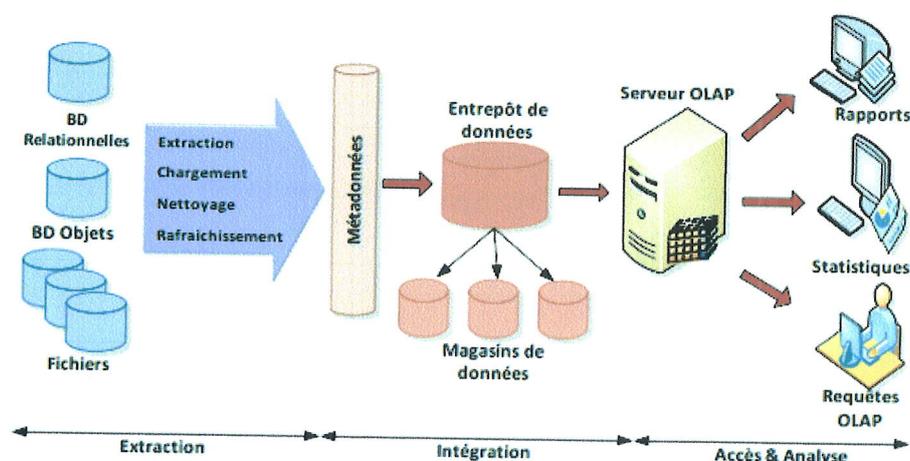


FIGURE 1.5 – Architecture d'un entrepôt de données.
[6]

1.4.2 Entrepôts et magasins de données

Le tableau suivant présente les différences entre l'entrepôt de données et magasin de données on quelques points essentiels :

Table des figures

1.1	Architecture des systèmes décisionnels.	6
1.2	Schéma en étoile.	10
1.3	Schéma en flocon de neige.	11
1.4	modélisation en constellation.	12
1.5	Architecture d'un entrepôt de données.	14
1.6	Les data marts indépendants architecture.	16
1.7	architecture Hub and spoke.	17
1.8	entrepôt de données centralisé.	18
2.1	Architecture conceptuelle d'un entrepôt de données actif.	26
2.2	Graphe d'analyse.	29
2.3	Classification des événements.	29
3.1	L'organigramme de l'E.P.H de Jijel	34
3.2	L'organigramme des différents services de l'hôpital de Jijel	36
3.3	Architecture générale du système	37
3.4	Diagramme de cas d'utilisation du système	38
3.5	Schéma conceptuel de bases de données sources	41
3.6	schéma en flocon de neige	46
3.7	La syntaxe générale d'une règle d'analyse	47
3.8	la règle d'analyse N°1	47
3.9	la règle d'analyse N°2	48
3.10	la règle d'analyse N°3	48
3.11	la règle d'analyse N°4	48
4.1	Fenêtre de l'authentification.	52

Caractéristique	Magasin de données	Entrepôt de données (ED)
Portée	Un domaine d'analyse	Plusieurs domaines d'analyse
Temps de développement	Mois	Années
Coûts de développement	\$10,000 à \$ 100,000 +	\$ 1, 000,000 +
Complexité de développement	Faible à moyenne	Grande
Taille des données	Mb à plusieurs Gb	Gb jusqu'à plusieurs Pb
Horizon des données	Courantes et historiques	La plupart du temps historiques
Transformation des données	Faible à moyenne	importante
Fréquence des mises à jour	Horaire, journaliser ou hebdomadaire	Peut aller jusqu'à mensuel
Nombre d'utilisateurs simultanés	Dizaines	Centaines à milliers
Types d'utilisateur	Analystes dans le domaine spécifique et gestionnaires	Analystes d'entreprise et cadres seniors
Objectifs d'affaires	Optimisation des activités dans le domaine spécifique	Optimisation inter-fonctionnelle et support à la décision

TABLE 1.1 – Magasin de données vs Entrepôt de données.
[22]

1.4.3 Les différentes Architectures d'entrepôts de données

Dans le domaine d'entrepôt de données il existe principalement cinq types d'architectures ce sont : les data marts indépendants, architecture en bus de magasins de données, architecture Hub-and-spoke, entrepôt de données centralisé, architecture fédérée

1.4.3.1 Les data marts (magasin de données) indépendants

C'est l'architecture la plus simple, développe pour les départements, des domaines fonctionnels ou des objectifs spécialisés qui opèrent de manière indépendante ou isolement. Les organisations disposant d'un certain nombre des inconvénients à cette architecture on peut résumer comme suit[9] :

- ◆ Incohérences des données dans les data marts et manquent de conformité.
- ◆ Les données et les traitements sont redondants.
- ◆ Non évolutif.

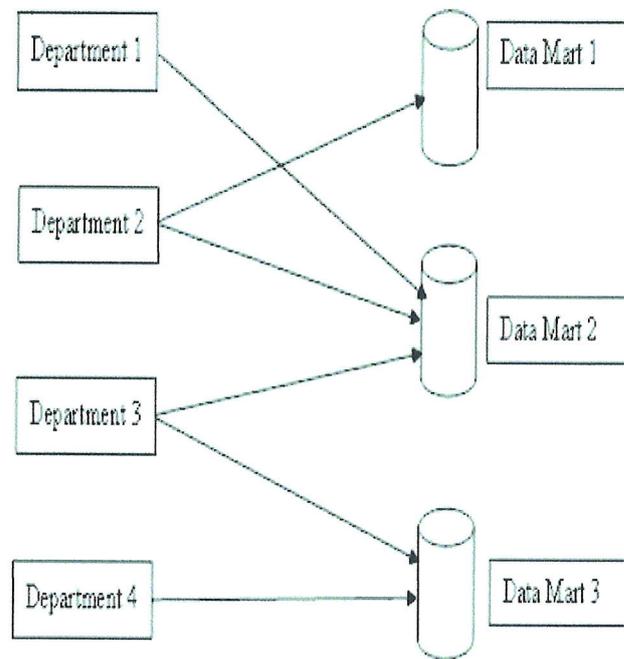


FIGURE 1.6 – Les data marts indépendants architecture.
[9]

1.4.3.2 Architecture en bus de magasins de données

Cette architecture est développée par des processus métiers (d'affaires) spécifiques, en se basant sur des dimensions et les faits conformes permettent l'intégration de données supplémentaires pour former une vue d'ensemble de l'organisation à l'échelle de l'organisation. Les données sont modélisées dimensionnellement dans un schéma en étoile. Elle caractérise par [9] :

- ◆ Intégration des données assurée par les dimensions conformes.
- ◆ Approche bottom-up.
- ◆ Donne des résultats rapidement.
- ◆ Modélisation dimensionnelle (schéma en étoile).
- ◆ Les data marts utilisent des "dimensions normalisées et conformes".
- ◆ L'entrepôt est "conceptuel" créé par le "bus" de dimensions conformes.

1.4.3.3 Architecture Hub-and-spoke

Est développée de manière itérative, dans cette architecture les données de niveau atomique sont conservées dans l'entrepôt sous la 3^{ème} forme normale. Cette architecture est caractérisée par [9] :

- ◆ Une table de fait et un ensemble des tables (tables de dimension) relie à cette table.
- ◆ Les points (table de dimension) reçoivent des données de la source centrale (table de fait).
- ◆ Basé sur les données nécessaires à la prise de décision.
- ◆ Évolutif
- ◆ Intégration et consolidation complète et des données de l'entreprise.
- ◆ Les données des data marts suivent le modèle dimensionnel.
- ◆ Les data marts reçoivent les données de l'entrepôt de données.

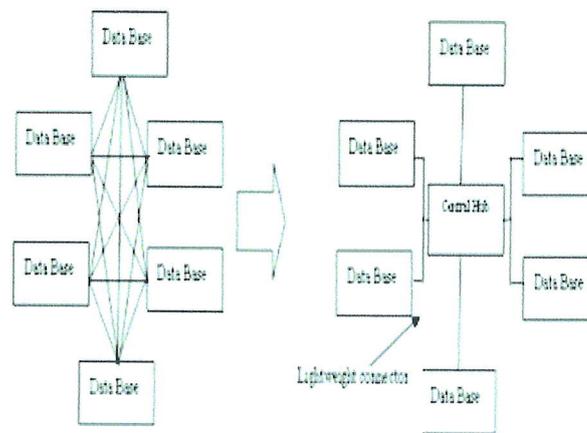


FIGURE 1.7 – architecture Hub and spoke.
[9]

1.4.3.4 Entrepôt de données centralisé

On peut dire que cette architecture est similaire à l'architecture en étoile, mais n'a pas de dépendance data marts. Les caractéristiques de cette approche est comme suit [9] :

- ◆ Pas de dépôts de données dépendants.
- ◆ Consolide les data marts dans l'entrepôt de données.
- ◆ L'entrepôt contient à la fois des données atomiques (de détail) et résumé.

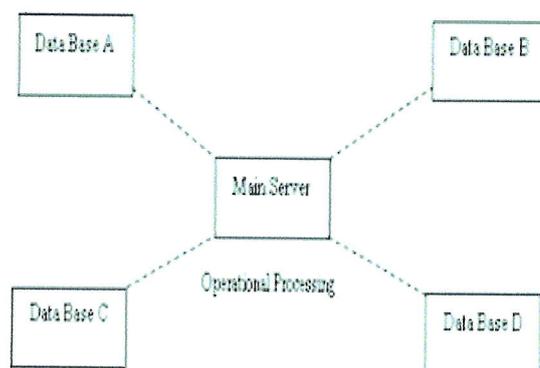


FIGURE 1.8 – entrepôt de données centralisé.
[9]

1.4.3.5 Architecture fédérée

L'architecture fédérée s'appuie sur la collecte de données à partir de plusieurs structures telles que les data marts, les entrepôts de données et les systèmes transactionnels, elle est opérée de manière transparente. Cette approche est caractérisée par [9] :

- ◆ Ne pas réorganiser les structures de données existantes (telles que data marts, entrepôts ou systèmes transactionnels).
- ◆ Intégrer logiquement ou physiquement les données.
- ◆ Les requêtes distribuées et les métadonnées associent les données.
- ◆ Accéder aux données simultanément sur plusieurs systèmes.
- ◆ Utile dans les fusions et acquisitions.

1.4.4 Réalisation d'un entrepôt de données

D'une manière générale, dans le domaine d'entrepôt de données il existe principalement deux approches d'architectures introduites par Bill Inmon et Ralph Kimball ce sont successivement : le traitement descendant (top-down processing) et le traitement ascendant (bottom-up processing), la combinaison de ces deux approches donne une autre approche appelée Middle-Out (Middle-Out processing).

Top-down : c'est l'approche descendante commence par la conception et la planification globales. C'est la méthode la plus lourde, la plus contraignante et la plus complète en même temps. Elle consiste en la conception de tout l'entrepôt (ie : toutes les étoiles), puis en la réalisation de ce dernier. Imaginez le travail qu'une telle méthode implique : savoir à l'avance

toutes les dimensions et tous les faits de l'entreprise, puis réaliser tout ça. Le seul avantage que cette méthode comporte est qu'elle offre une vision très claire et très conceptuelle des données de l'entreprise ainsi que du travail à faire.[7]

Bottom-up : c'est l'approche inverse, elle consiste à créer les étoiles une par une, puis les regrouper par des niveaux intermédiaires jusqu'à obtention d'un véritable entrepôt pyramidal avec une vision d'entreprise. L'avantage de cette méthode est qu'elle est simple à réaliser (une étoile à la fois), l'inconvénient est le volume de travail d'intégration pour obtenir un entrepôt de données ainsi que la possibilité de redondances entre les étoiles (car elles sont faites indépendamment les unes des autres).[7]

Middle-Out : c'est l'approche hybride, et conseillée par les professionnels du business intelligence (BI). Elle consiste en la conception totale de l'entrepôt de données (ie : concevoir toutes dimensions, tous les faits, toutes les relations), puis créer des divisions plus petites et plus gérables et les mettre en œuvre. Cela équivaut à découper notre conception par éléments en commun et réaliser les découpages un par un. Cette méthode tire le meilleur des deux précédentes sans avoir les contraintes. Il faut juste noter que cette méthode implique, parfois, des compromis de découpage (dupliquer des dimensions identiques pour des besoins pratiques).[7]

1.4.5 Comparaison entre les deux approches

Le tableau 1.2 présente des différents points de la comparaison entre les deux approches :

	Inmon	Kimball
source requise	✓	✓
mise en scène	✓	✓
ETL	✓	✓
Data marts	✓	✓
exigences de l'entreprise	✓	✓
attribut de temps des données	✓	✓
entreprise dw	✓	×
outils dimensionnels	×	✓
outils relationnel	✓	×
orienté processus	×	✓
modèle de données normalisé	✓	×
complexe à concevoir	✓	×
cadre temporel continu et discret	✓	×
changer lentement le cadre temporel	×	✓

TABLE 1.2 – comparaison entre Inmon et Kimball.

1.5 Les opérations OLAP dans le modèle de données multidimensionnel

Dans le modèle multidimensionnel, les données sont organisées en plusieurs dimensions et chaque dimension contient plusieurs niveaux d'abstraction définis par des hiérarchies de concepts. Cette organisation offre aux utilisateurs la possibilité de visualiser les données sous différents angles. Un certain nombre d'opérations de cube de données OLAP existent pour matérialiser ces différentes vues, permettant une interrogation interactive et une analyse des données disponibles. On trouve les opérations suivantes [3] :

- **Roll-up** : l'opération de roll-up (également appelée opération de drill-up par certains fournisseurs) effectue l'agrégation sur un cube de données, soit en gravissant une hiérarchie de concepts pour une dimension ou par réduction de dimension. Lorsque l'enroulement est effectué par réduction de dimension, une ou plusieurs dimensions sont retiré du cube donné.[3]
- **Drill-down** : est l'inverse de roll-up. Il navigue à partir de données moins détaillées des données plus détaillées. L'exploration peut être réalisée soit en supprimant une hiérarchie de concepts pour une dimension, soit en introduisant des dimensions supplémentaires.[3]
- **Slice and dice** : l'opération de découpe effectue une sélection sur une dimension du cube donné, résultant en un sous-cube.[3]
- **Pivot (rotate)** : est une opération de visualisation qui fait pivoter les données axes en vue de fournir une présentation alternative des données.[3]

1.6 Les objectifs d'un entrepôt de données

Selon Kimball avant de plonger dans les détails de la modélisation et de son implémentation, il convient de s'interroger sur les objectifs fondamentaux de l'entrepôt de données on peut les comprendre en se promenant dans les couloirs de toute organisation et en écoutant les gestionnaires d'entreprise. On retrouve inévitablement des thèmes récurrents [10] :

- « *Nous avons des montagnes de données dans cette société mais nous ne pouvons pas y accéder.* »
- « *Nous avons besoin de faire des coupes en tranches et en dés dans les données de toutes sortes de façons.* »

- « *Il faut faire en sorte que les gestionnaires puissent accéder aux données facilement et directement.* »
- « *Montrez-moi seulement ce qui est important.* »
- « *Je suis fou de rage quand deux personnes me présentent les mêmes mesures de performance dans une réunion, mais avec des chiffres différents.* »
- « *Nous voulons que les gens puissent utiliser les informations pour baser leurs prises de décision sur des faits.* »

Nous avons constaté que ces préoccupations sont tellement universelles qu'elles représentent le fondement du cahier des charges d'un entrepôt de données. Transformons ces commentaires de gestionnaire d'entreprise en articles du cahier des charges d'un entrepôt de données.

L'entrepôt de données doit rendre les données de l'organisation facilement accessibles. Le contenu d'entrepôt doit être facile à comprendre. Les données doivent être parlantes et leur signification évidente pour l'utilisateur et pas seulement pour le développeur. Le contenu de l'entrepôt de données doit être étiqueté de manière significative. Les utilisateurs veulent séparer et combiner les données de toutes sortes de façons, un processus que l'on nomme découpage en tranches et en dés. Les outils d'accès doivent être simples et faciles à utiliser. Le renvoyer des résultats de requêtes à l'utilisateur avec des temps d'attente minimales.[10]

L'entrepôt de données doit présenter l'information de l'organisation de manière cohérente. Les données doivent être assemblées à partir de différentes sources de l'organisation et nettoyées. Il faut contrôler leur qualité et ne les publier que lorsqu'elles sont propres à la consommation par les utilisateurs. Les informations d'un processus d'entreprise doivent correspondre à celles d'un autre processus d'entreprise. Si deux mesures de performance portent le même nom, elles doivent vouloir dire la même chose. Inversement, si deux mesures ne veulent pas dire la même chose, elles doivent être appelées différemment. La cohérence implique une qualité élevée. La cohérence exige, en outre, que les définitions communes de l'entrepôt de données soient disponibles pour les utilisateurs.[10]

L'entrepôt de données doit être adaptable et résistant aux changements. Nous ne pouvons tout simplement pas éviter les changements. Les besoins des utilisateurs, les conditions de l'activité, les données et la technologie sont les uns comme les autres exposés au temps qui passe. Les données de l'entrepôt doivent être conçues pour traiter ces changements inévitables. Les modifications de l'entrepôt de données doivent se faire en douceur, ce qui veut dire qu'elles ne doivent pas invalider les données existantes ou les applications. Les données existantes et les applications ne doivent pas être modifiées ou bouleversées lorsque la communauté des utilisateurs pose de nouvelles questions ou que de nouvelles données sont ajoutées à l'entrepôt de données. Si les données descriptives de l'entrepôt de données sont modifiées, nous devons rendre compte convenablement de ces modifications.[10]

L'entrepôt de données doit être un bastion sur protégeant notre richesse informationnelle. Les plus précieuses parmi les informations d'une organisation sont conservées dans l'entrepôt de données. Au minimum, l'entrepôt de données contient le plus souvent des informations sur ce que nous vendons, à qui, à quel prix ? des précisions dangereuses si elles tombent en de mauvaises mains. L'entrepôt de données doit efficacement contrôler l'accès aux informations confidentielles de l'organisation. [10]

L'entrepôt de données doit être le socle sur lequel repose l'amélioration des prises de décision. Il doit contenir les données servant à étayer les décisions. L'entrepôt de données ne produit qu'une seule sortie : les décisions prises sur la base des réalités qu'il révèle. Ces décisions sont la valeur ajoutée de l'entrepôt de données. L'appellation antérieure à celle de l'entrepôt de données est toujours celle qui décrit le mieux ce que nous mettons en place : un système d'aide à la décision. [10]

L'acceptation de l'entrepôt de données par la communauté des utilisateurs est l'une des conditions de réussite. Cela ne sert à rien de concevoir une solution élégante utilisant les produits et les plates-formes les plus prisés. Si la communauté des utilisateurs n'a pas accueilli l'entrepôt de données à bras ouverts et continué de s'en servir régulièrement six mois après la formation, alors nous avons échoué au test de réception. Contrairement à la nouvelle version d'une application opérationnelle, dont les utilisateurs sont toujours obligés de se servir, l'entrepôt de données peut n'être qu'une option pour ses utilisateurs potentiels. L'acceptation des utilisateurs est avant tout liée à la simplicité d'utilisation. Comme le montre cette liste, le succès d'un entrepôt de données demande bien plus qu'une excellente maîtrise de la technique et des bases de données. Lors d'un projet d'entrepôt de données, nous avons un pied solidement posé dans les technologies de l'information et d'autre sur le terrain moins familier des gestionnaires. Nous devons être à cheval sur les deux domaines et faire évoluer certaines de nos talents confirmés pour les adapter aux exigences particulières des entrepôts de données. Il est clair que nous avons besoin de compétences variées requises par un comportement hybride administrateur de données/ gestionnaire. [10]

1.7 Les types d'entrepôts de données

Il existe trois types d'entrepôt de données on peut citer comme suit :

- ◆ Entrepôt de données passif (Classique).
- ◆ Entrepôt de données temps réel (Real Time).
- ◆ Entrepôt de données actif.

On peut les classer selon domaine d'application :

Domaine géographie : un système d'informations géographique permet d'acquérir, de structurer, de mémoriser, d'analyser et de visualiser les données géographiques. L'intégration



des données géographiques dans l'analyse en ligne est un enjeu majeur. La modélisation des entrepôts de données géographiques tout comme l'adaptation des fonctionnalités des systèmes d'entrepôt de données classiques pour les données géographiques. Les entrepôts de données ne permettent pas une analyse automatique des données ce qui ne permet pas de générer automatiquement des rapports d'analyse.[11, 12]

Domaine industriel : Les entrepôts de données actifs permettent d'une analyse automatique utilisée bien dans le domaine industriel.[12]

1.8 Conclusion

Dans ce chapitre, nous avons introduit les principes à la base de l'entreposage de données et de l'analyse OLAP avec les problématiques associées à la modélisation multidimensionnelle de données :

- Dans un premier temps, nous avons expliqué les différents composants qu'il intègre, comme les diverses sources, les types de données et les différents outils pour arriver à la construction d'un système d'aide à la décision à base d'entrepôt de données.
- Dans une deuxième partie, nous avons décrit les différents modèles multidimensionnels pour la construction d'un entrepôt de données, ainsi que les différentes opérations pour la manipulation des données multidimensionnelles.

Les entrepôts de données actifs

2.1 Introduction

L'évolution des bases de données aux entrepôts de données a permis le passage de la gestion à l'analyse des données. Les entrepôts de données offrent un support limité pour automatiser les tâches de prise de décision qui se produisent fréquemment. Ces entrepôts de données conventionnels sont considérés comme des entrepôts passifs. Toutes les tâches reliées à l'analyse des données et à la prise de décision doivent être effectuées manuellement par les analystes.

Aujourd'hui, une nouvelle génération de systèmes décisionnels, est en train de se propager dans le domaine du décisionnel, appelée entrepôt de données actif (EDA). Les entrepôts de données actifs apportent des solutions pour pallier les insuffisances dont souffrent les entrepôts de données conventionnels. L'idée des entrepôts de données actifs a été inspirée par des techniques utilisées dans les bases de données actives, appelons les règles actives qui utilisent le paradigme événement-condition-action (ECA).

Dans ce chapitre, nous proposons dans un premier temps de se familiariser avec les concepts de bases de données actives. Dans un deuxième temps, nous présentons le nouveau concept d'entrepôt de données actif.

2.2 Les bases de données actives

2.2.1 Définition

Un système de base de données actif est un système de base de données qui surveille les situations d'intérêt, et quand ils se produisent, déclenche une réponse appropriée en temps opportun. Le comportement souhaité est exprimé dans des règles actives (également appelées règles d'événement-condition-action), qui sont définies et stockées dans la base de données.

Cela a l'avantage que les règles peuvent être partagées par de nombreux programmes d'application, et que le système de base de données peut optimiser leur implémentation.[15]

2.2.2 Modèle de règles ECA

Le Modèle ECA (Événement, Condition, Action) est le modèle généralisé appliqué pour spécifier les règles des bases de données actives, dans ce modèle, la forme générale d'une règle est constitué de trois composants : lorsque survient un événement E, si la condition C est satisfaite alors exécuter l'action A.[16]

On Event

If Condition

Then Action

2.2.2.1 Événement

« *Un événement est quelque chose qui se produit à un instant donné et qui a une importance toute particulière : le pouvoir de déclencher une règle (ou plusieurs)* ». [16]

L'événement à le pouvoir de provoquer l'évaluation de la condition de la règle ECA et l'exécution de sa partie Action si la Condition est satisfaite.

2.2.2.2 Condition

Une condition est une formule qui doit être réalisée pour que l'action soit exécutée.[16]

2.2.2.3 Action

La partie action dans le modèle ECA spécifie les opérations à effectuer lorsque la règle est déclenchée et sa condition est satisfaite. Cette partie peut être un ordre arbitraire d'insertion, de mise à jour, de suppression, etc. [15]

2.3 Entrepôts de données actifs

2.3.1 Définition

L'entrepôt de données actif(EDA), est un système qui offre la possibilité d'automatiser la prise de décision des tâches de routine, ce système réalise une approche de décision en boucle fermée, nommée cycle actif d'entrepôt de données.[14]

Un entrepôt de données actif peut exporter des décisions aux systèmes OLTP directement ou après confirmation de l'utilisateur. Pour automatiser l'analyse et la prise de décision, on utilise l'idée de règles ECA à partir de systèmes de bases de données actifs. Les règles ECA qui imitent le travail d'un analyste sont appelées règles d'analyse, elles peuvent être exécutées automatiquement sans interaction explicite de l'utilisateur ou demande d'application.[14]

2.3.2 Architecture d'un EDA

Pour automatiser les tâches de décision qui se produisent fréquemment, on peut prolonger l'architecture conventionnelle d'entrepôt de données en ajoutant des règles d'analyse. Ces règles seront spécifiées une fois par l'analyste et seront ensuite exécutées automatiquement par l'entrepôt de données. Ainsi, les règles existantes peuvent être facilement adaptées aux exigences changeantes et peut être utilisé comme modèle pour résoudre de nouveaux problèmes.[14] L'idée fondamentale derrière les règles d'analyse est de prolonger les règles conventionnelles ECA des systèmes actifs de base de données avec des dispositifs multidimensionnels (cubes, opérateurs OLAP) des entrepôts de données et d'une approche pour la prise de décision.[12]

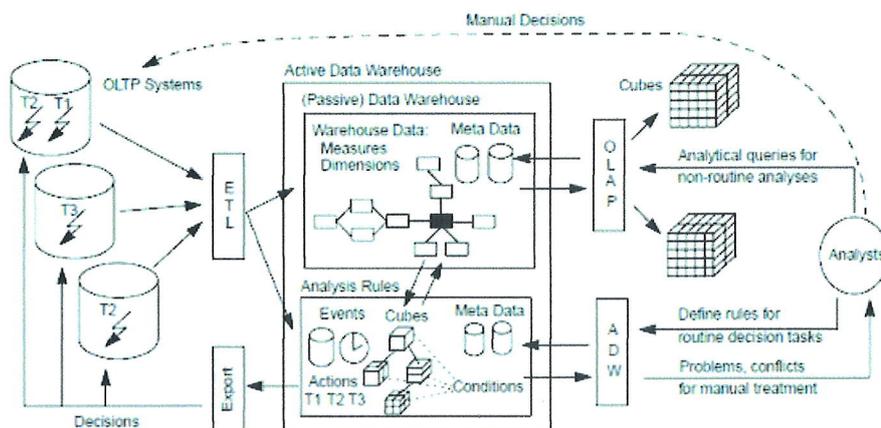


FIGURE 2.1 – Architecture conceptuelle d'un entrepôt de données actif.
[14]

2.3.3 Règles d'analyses

Les règles d'analyse ont la même structure que les règles ECA, mais effectuent des analyses beaucoup plus complexes à l'aide d'opérateurs OLAP sur des données d'entrepôt au lieu d'évaluer des conditions simples comme le cas des règles ECA.[12]

Les composants des règles d'analyse font partie de la méta-connaissance d'un entrepôt de données actif, qui comprend toutes les informations supplémentaires qui doivent être disponibles dans l'entrepôt de données actif afin que les analystes puissent spécifier leurs règles

actives pour automatiser les tâches de décision courantes.[14]

La définition d'une règle d'analyse selon Thalhammer [14], comprend les éléments suivants :

- **Dimension primaire** : représente les instances d'un niveau de dimension particulier pour lequel la règle est définie. La dimension primaire détermine aussi la partie action de la règle d'analyse.
- **Événement** : spécifié par des points temporels pour le déclenchement de la règle d'analyse.
- **Condition primaire** : permet d'analyser uniquement le sous-ensemble d'instances de niveau qui appartiennent au niveau de la dimension primaire.
- **Graphe d'analyse** : représente tous les cubes qui sont utiles pour l'analyse, reliés entre eux par des opérateurs OLAP (Roll-Up, Drill-Down,... etc).
- **Étapes de décision** : représentent les conditions dans lesquelles une décision peut être prise.
- **Action** : représente la tâche de décision de la règle.

La spécification d'une règle d'analyse nécessite que les données de l'entrepôt de données soient affichées d'une perspective orientée objet et d'une perspective multidimensionnelle.[14]

2.3.3.1 Approche par objet

Le principal raison d'utiliser une approche orientée objet est que les objets qui doivent être analysés l'utilisation de règles actives doit être associée à certaines méthodes qui seront utilisées comme actions des règles d'analyse.[14]

2.3.3.2 Approche multidimensionnelle

En utilisant l'approche multidimensionnelle, les analystes définissent quels cubes doivent être inspecté et quelles conditions sont nécessaires à l'analyse de ces cubes. Le résultat des analyses multidimensionnelles est la décision d'exécuter la méthode dans un OLTP système ou d'effectuer d'autres analyses manuelles.[14]

2.3.4 Analyse multidimensionnelle

Lorsqu'un analyste interroge l'entrepôt de données pour prendre la décision d'exécuter une transaction dans un système OLTP, il suit une approche descendante incrémentielle dans la création et l'analyse des cubes. Tout d'abord, l'analyste crée un cube très grossier qui décrit les objets pour lesquels une décision doit être donnée .En second lieu, dans la réalisation des analyses multidimensionnelles l'analyste compare, transforme,... etc., les cellules de ce cube. En conclusion, l'analyste peut déterminer pour chaque niveau d'instance du niveau de

la dimension primaire si une décision peut être prise ou si une analyse plus poussée utilisant un cube plus détaillé est nécessaire.[14]

Les règles d'analyse suivent la même approche. Pour automatiser la prise de décision, une règle d'analyse doit connaître les cubes nécessaires aux analyses multidimensionnelles qui constituent le graphe d'analyse spécifié une seule fois par l'analyste. Les n dimensions de chaque cube du graphe d'analyse sont classées en une dimension primaire, qui représente le niveau d'instances du niveau de dimension primaire, et $n-1$ dimensions d'analyse, qui représentent l'espace multidimensionnel pour l'analyse, puisqu'un niveau d'instance du niveau de dimension primaire est décrite par une ou plusieurs cellules d'un cube, l'analyse multidimensionnelle signifie comparer, transformer, etc. les valeurs de mesure de ces cellules.[14]

Les analyses multidimensionnelles qui sont effectuées sur les cubes du graphe d'analyse sont appelées étapes de décision. Chaque étape de décision analyse les données d'exactitude d'un cube du graphe d'analyse. Ainsi, le graphe d'analyse et les étapes de décision représentent les connaissances pour l'analyse multidimensionnelle et la prise de décision d'une règle d'analyse.[14]

2.3.5 Graphes d'analyse

Les graphes d'analyse représentent tous les cubes qui sont nécessaires pour une analyse multidimensionnelle. Un graphe a un seul cube racine qui est le cube le plus grossier, ce cube est créé en appliquant l'opérateur Roll-Up de OLAP sur un cube de base. Chaque cube se trouvant à l'intérieur du graphe d'analyse est dérivé directement ou indirectement du cube racine en appliquant les opérateurs : Drill-Down, Slice ou Intersection. Ces opérateurs sont utilisés au niveau du cube grossier pour le raffiner et par conséquent avoir un cube plus spécifique.[12]

- **L'opérateur Slice** : permet de sélectionner un sous-ensemble de cellules dans un cube.
- **L'opérateur Drill-Down** : permet de passer d'un niveau grossier de dimension à un niveau de dimension plus fin en présentant les mesures avec un niveau de granularité plus fin.
- **L'opérateur Intersection** : crée un cube plus spécifique qui représente un sous-ensemble de cellules, pour lesquelles la mesure possède un niveau de granularité plus fin.
- **L'opérateur Roll-Up** : permet de passer d'un niveau de dimension fine à un niveau de dimension plus gros.

Pour automatiser la prise de décision, le graphe d'analyse constitue une partie indispensable aux règles d'analyse. Ces règles doivent connaître les cubes qui sont nécessaires à l'analyse multidimensionnelle.[12]

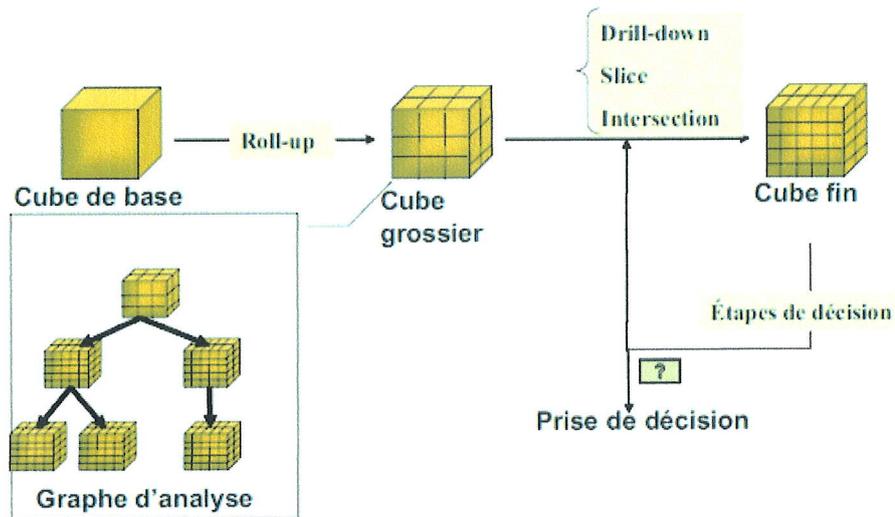


FIGURE 2.2 – Graphe d’analyse.
[12]

2.3.6 Taxonomie d’événements dans un entrepôt de données actif

Nous distinguons deux grandes catégories d’événements selon la classification de Bouattour [13] : les événements OLTP et les événements OLAP.

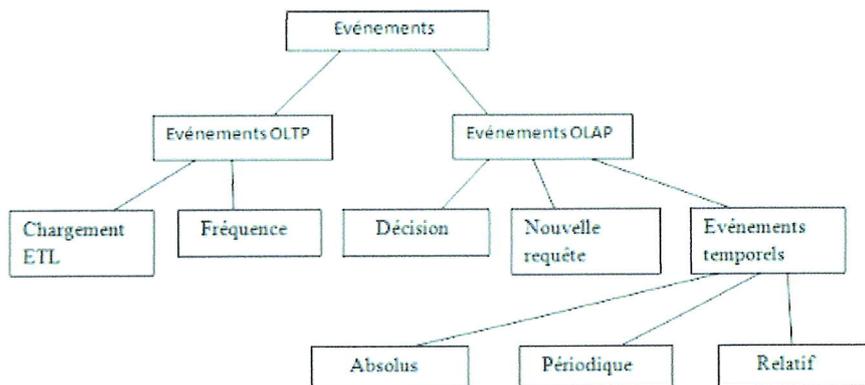


FIGURE 2.3 – Classification des événements.
[13]

2.3.6.1 Les événements OLTP

Les événements OLTP sont généralement les événements classiques rencontrés dans les bases de données transactionnelles, ces événements peuvent concerner l'insertion, la modification et la suppression des données. Ils ont composé de :

- **Chargement ETL** : Le chargement constitue la phase de l'alimentation de l'entrepôt qui peut se déclencher suite à une insertion, une modification ou une suppression dans les sources OLTP.
- **Fréquence** : périodique pour calculer le nombre d'occurrence d'une opération OLTP.

2.3.6.2 Les événements OLAP

Nous distinguons 3 catégories d'événements OLAP :

- **Décision** : correspond à une décision d'un décideur.
- **Nouvelle requête** : c'est l'appelle d'une méthode OLTP.
- **Temporel** : Il existe trois types d'événements temporels :
 - ◆ **événements temporels absolus** : est une date du calendrier, par exemple : 20 avril 2018.
 - ◆ **événement temporel périodique** : peut être la fin de chaque trimestre, chaque mois,... etc.
 - ◆ **événements temporels relatifs** : représentent des points temporels positionnés par rapport à un événement donné, par exemple : 20 jours après le lancement d'un produit.

2.3.7 Comparaison entre l'entrepôt de données actif et passif

La table 2.1 résume les principales différences entre un entrepôt de données actif et passif selon Thalhammer [14] et Bouattour [13] :

Caractéristique	l'entrepôt de données actif	l'entrepôt de données passif
Le but	La prise de décision	Aide à la décision
Le contenu	Mémoriser les données et les connaissances	Mémoriser les données
Analyse de données	Automatique pour les tâches de décision courantes et manuelle pour les tâches de décision non courantes	Toutes les analyses de données sont faites explicitement par les utilisateurs (analystes)
Le cycle	Avec cycle	Sans cycle

TABLE 2.1 – Comparaison entre l'entrepôt de données actif et passif

2.4 Conclusion

Dans ce chapitre, nous avons présenté une nouvelle approche pour automatiser la prise de décision dans les entrepôts de données, on l'appelle entrepôt de données actif. L'idée de base des entrepôts de données actifs est d'étendre les règles ECA conventionnelles avec des fonctionnalités multidimensionnelles pour analyser les données et prendre des décisions. La technologie des entrepôts de données actifs se développe de plus en plus dans le monde du décisionnel. Elle répond à la croissance constante des volumes de données et aux besoins des utilisateurs.

CHAPITRE 3

La Conception du Système

3.1 Introduction

La conception consiste à enrichir la description du logiciel de détails d'implémentation. Elle consiste aussi à traduire les besoins en spécifiant comment l'application pourra être satisfaisante avant de procéder à sa réalisation avec des outils de développement appropriés.

La conception d'un entrepôt de données actif est une tâche complexe. Elle se compose de plusieurs étapes. Dans un premier temps, nous devons analyser l'ensemble des sources de données. Cette analyse sert à la sélection de l'ensemble de données à stocker dans l'entrepôt de données. La deuxième étape consiste à organiser ces données à l'intérieur de l'entrepôt. Pour ce faire, nous devons concevoir le modèle multidimensionnel à utiliser. Finalement, la troisième étape consiste à déterminer les règles d'analyses nécessaires pour l'automatisation de la décision.

La première partie de ce chapitre montre la présentation du domaine d'application et la spécification d'architecture globale de notre système. La deuxième partie concerne la conception de base de données source et de l'entrepôt de données ainsi que les règles d'analyses.

3.2 Domaine d'application

3.2.1 Présentation d'organisme d'accueil

L'établissement public hospitalier Mohammed Seddik Ben Yahia (E.P.H) est considéré comme l'un des établissements les plus importants de la wilaya de Jijel : le premier hôpital de la wilaya a été créé en 1933 sous le nom de "France Fanant", Sur une superficie de 9100 m^2 , dont 3900 m^2 construits ou 42,56%, et 5200 m^2 non construits, avec une capacité de 160 lits, et suivant la division administrative de l'année 1974, qui a séparé Jijel de la wilaya de

Constantine, un nouvel hôpital nommé Mohammed Seddik Ben Yahia a été créé et l'hôpital "France Fanant" ne répond plus aux demandes croissantes de la population dans le domaine de la santé, l'hôpital a été lancé le 8 novembre 1983, en vertu de la résolution ministérielle n° 81/242.

En 2008, le nom de l'hôpital Mohammed seddik Ben Yahia est passé du secteur de la santé à l'hôpital public conformément au décret exécutif n° 07/140 du 19 mai 2007. L'E.P.H de Jijel est un établissement public à caractère administratif dotée d'une personnalité morale, et d'une indépendance financière, et placé sous la tutelle du Wali. IL est composé de la structure de diagnostic, de traitement, d'hospitalisation et de réadaptation médicale, et couvre la population d'une municipalité ou d'un groupe de municipalités. L'E.P.H de Jijel est situé sur le côté est de la ville de Jijel, accueillant actuellement 431 lits et couvrant la population de la municipalité de Jijel.

3.2.2 L'organigramme de l'hôpital de Jijel

L'organisation de la direction de l' E.P.H de Jijel s'articule autour de quatre sous-directions :

- La Sous-direction des finances et des moyens.
- La Sous-direction des ressources humaines.
- La Sous-direction de la maintenance du matériel médical et des équipements associés.
- La Sous-direction des services de santé.

La figure (3.1) illustre l'organigramme générale de l' E.P.H de Jijel.

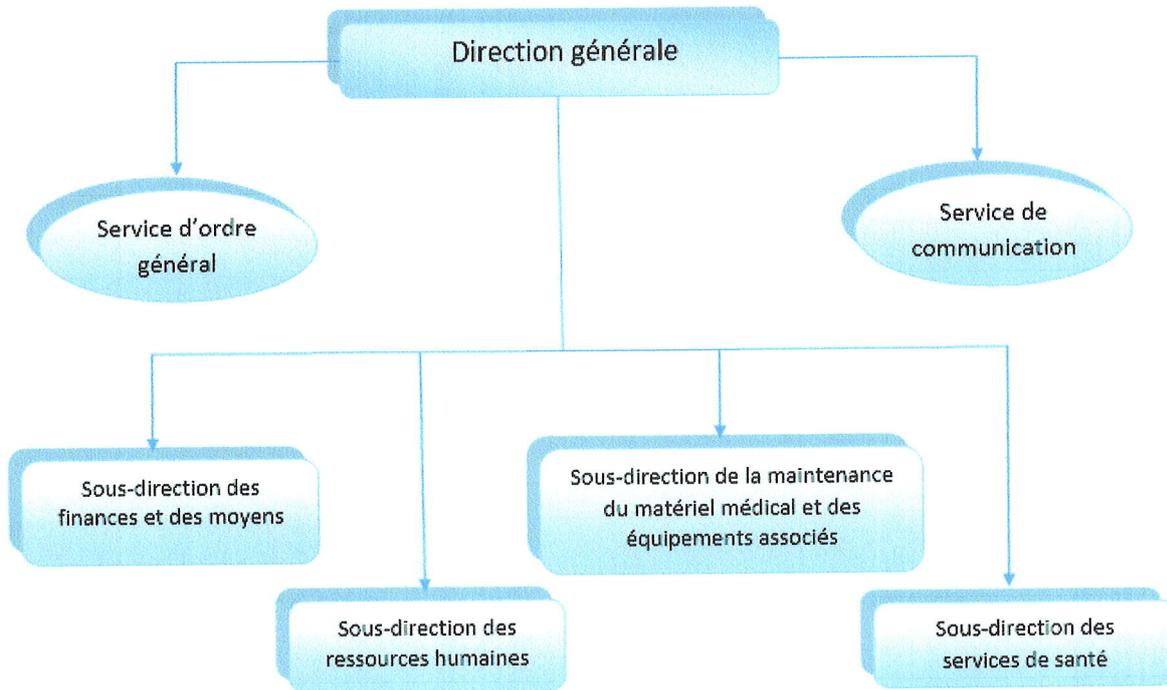


FIGURE 3.1 – L’organigramme de l’E.P.H de Jijel

D’après notre stage à l’hôpital de Jijel on a pu connaître ses différents services et le nombre des employés du côté médicale (médecins spécialistes, médecins généralistes, sage femmes, infirmiers, aide soins, pharmaciens, psychologues, laborantins) qui figurent dans le tableau (3.1) :



Service	Nombre d'employés
Service de médecine interne	32
Service de pédiatrie	41
Service de chirurgie générale	29
Service d'anesthésie et réanimation	34
Service de prévention	04
Service de néphrologie et hémodialyse	30
Service de chirurgie orthopédique	16
Service de maladie thoracique	06
Service de radiologie	02
Service d'obstétrique et gynécologie	78
Service des maladies cancéreuses	21
Service de maladies infectieuses	10
Service d'ophtalmologie	04
Service d'hospitalisation à domicile	02
Service des urgences	65
Service d'anatomie pathologique	07
Service de maladie psychiatrique	05
Service de médecine de travail	06
Service d'urologie	09
Service de médecine légale	03
Service de pharmacie	09
Laboratoire central	15
Centre de transfusion sanguine	09
Salle des opérations	31

TABLE 3.1 – Les différents services de l'hôpital de Jijel

Notre travail se base sur ses services qui sont groupés en trois classes :

- Services hospitaliers.
- Service technique.
- Autres services.

La figure (3.2) ci-dessous montre l'organigramme de différents services de l'hôpital de Jijel.

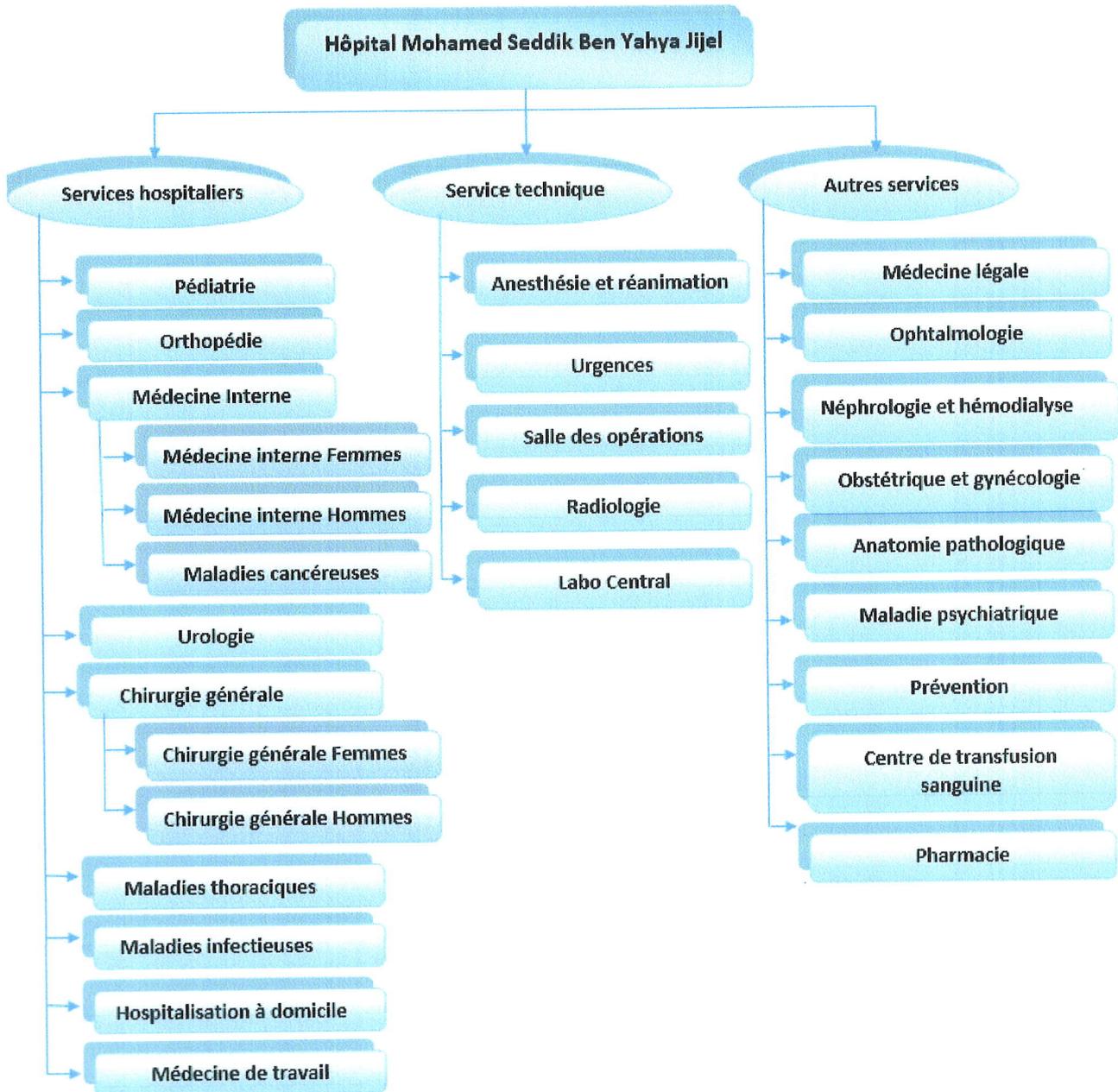


FIGURE 3.2 – L’organigramme des différents services de l’hôpital de Jijel

3.3 Architecture générale du système

On présente l’architecture générale de notre système selon le schéma suivant :

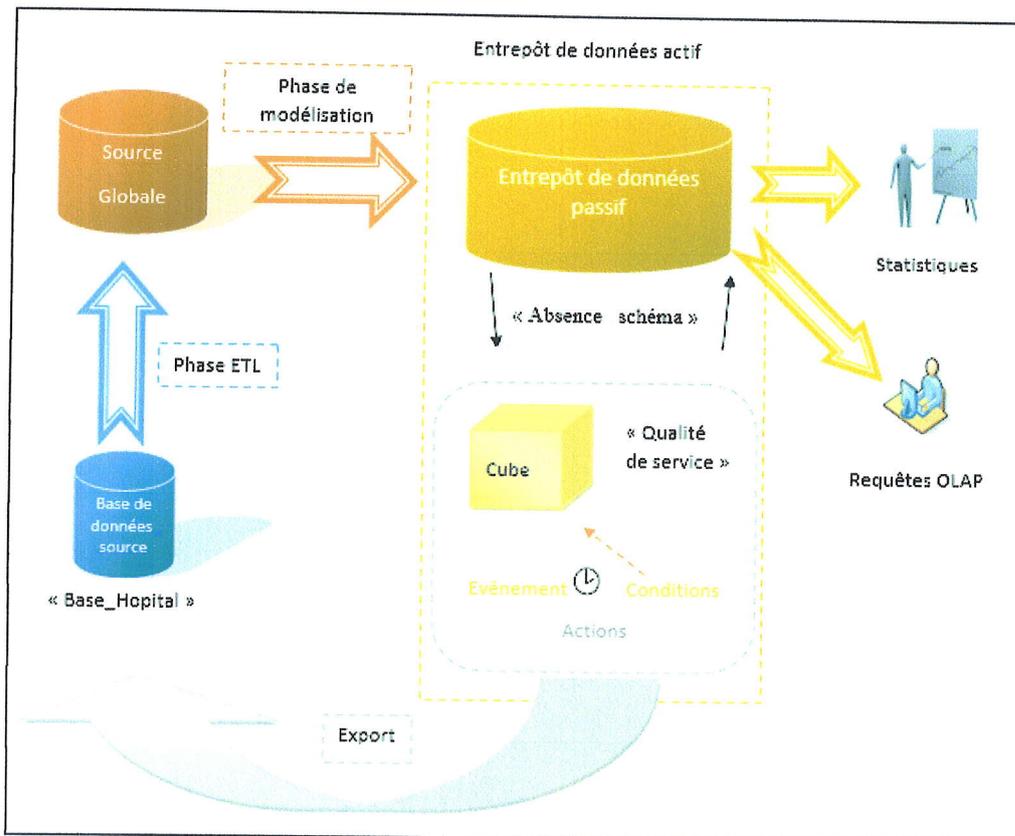


FIGURE 3.3 – Architecture générale du système

- **Base de données source** : la base de données « base-hôpital » contient l'ensemble des données nécessaires pour notre travail.
- **Source globale** : l'intégration des sources de données dans une source globale virtuelle, à partir de laquelle l'entrepôt de données est construit.
- **Entrepôt de données passif** : après la phase de modélisation on obtient un entrepôt de données où les données manipulées sont représentées sous forme multidimensionnelle.
- **Cube** : un cube organise les données en une ou plusieurs dimensions qui déterminent une mesure d'intérêt, il sera analysé par les règles d'analyse.

3.4 La conception des bases de données sources(Base_Hôpital)

3.4.1 Diagramme de cas d'utilisation du système

Ce diagramme est l'un des diagrammes d'UML. « *Un cas d'utilisation est une manière spécifique d'exprimer l'utilisation d'un système.* » [17] et capture son comportement.

Dans ce qui suit, on décrit le diagramme de cas d'utilisation associées au système à réaliser.

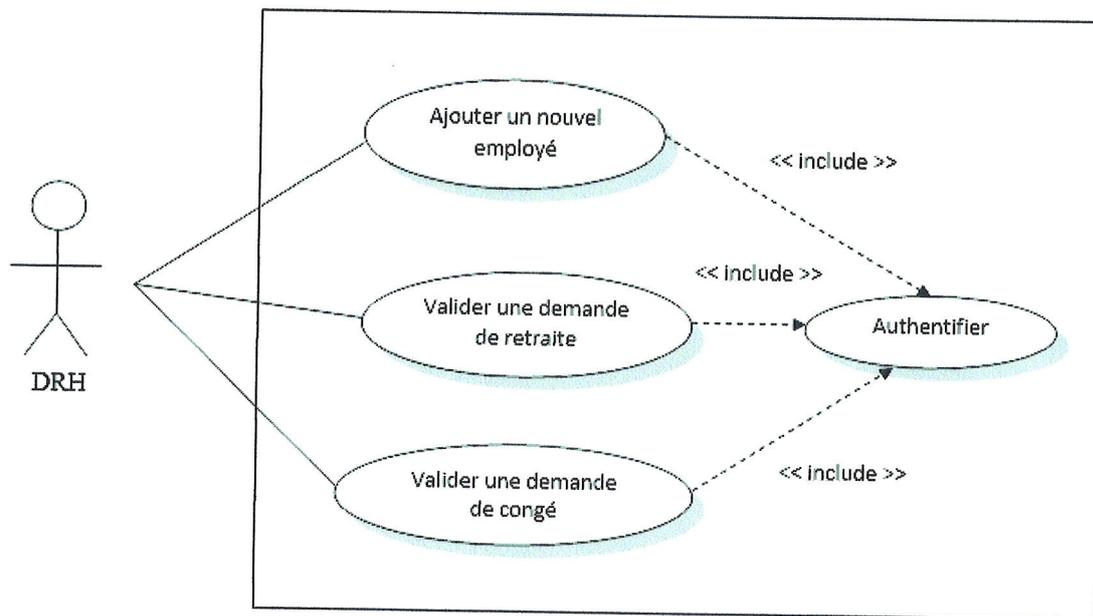


FIGURE 3.4 – Diagramme de cas d'utilisation du système

3.4.2 Description textuelle de cas d'utilisation

Nous avons utilisé la structure proposée par « Pascal Roques » [18] pour la description textuelle de cas d'utilisation qui est organisée de la façon suivante :

- **Sommaire d'identification.**
- **Description des enchainements.**

Cas d'utilisation : Authentifier	
Sommaire d'identification	
Titre	Connexion de l'utilisateur au système
But	Connecter au système
Acteur	L'utilisateur (DRH (directeur des ressources humaines))
Description des enchainements	
Pré condition	Aucune
Démarrage	l'utilisateur demande la connexion
Enchainements	<p>e1 :L'utilisateur demande une connexion au système.</p> <p>e2 :Le système demande le nom d'utilisateur et le mot de passe puis valide</p> <p>e3 :Le système vérifie l'existence de l'utilisateur et la validité de son mot de passe</p> <p>e4 :Le système ouvre une session et affiche l'espace d'utilisateur</p>
Arrêt	lorsque la confirmation d'authentification
Exception	Dans le cas où les informations fournies sont incorrectes, le système renvoi un message d'erreur, ré-affiche le formulaire d'authentification et attend que l'utilisateur ressaisisse ses informations correctes
Post condition	Ouverture de l'espace d'utilisateur
Cas d'utilisation : Ajouter un nouvel employé	
Sommaire d'identification	
Titre	Ajouter un nouvel employé
But	Pour sauvegarder les informations de l'employé
Acteur	L'utilisateur (DRH)
Description des enchainements	
Pré condition	L'utilisateur est authentifié
Démarrage	L'utilisateur demande le formulaire du nouvel employé
Enchainements	<p>e1 :L'utilisateur demande le formulaire d'ajout</p> <p>e2 :Le système afficher le formulaire d'ajout</p> <p>e3 :L'utilisateur saisit les informations de l'employé</p> <p>e4 :Le système enregistre les informations de l'employé</p>
Arrêt	Lorsque le système confirme l'enregistrement d'employé
Exception	Aucune
Post condition	Saisir un nouvel employé

Cas d'utilisation : Valider une demande de retraite	
Sommaire d'identification	
Titre	Valider une demande de retraite
But	Ce cas d'utilisation permet de supprimer un employé âgé
Acteur	L'utilisateur (DRH)
Description des enchainements	
Pré condition	l'existence d'employé
Démarrage	la saisie des informations de l'employé qui veut la retraite
Enchainements	e1 :L'utilisateur demande le formulaire de demande de retraite e2 :Le système afficher le formulaire de demande de retraite e3 :L'utilisateur saisit les informations de l'employé e4 :Le système supprime les informations de l'employé
Arrêt	Lorsque le système supprime l'employé
Exception	Suppression impossible si l'employé ne dépasse pas la période de travail
Post condition	un employé à supprimer
Cas d'utilisation : Valider une demande de congé	
Sommaire d'identification	
Titre	Valider une demande de congé
But	Ce cas d'utilisation permet de saisir un nouveau congé
Acteur	L'utilisateur (DRH)
Description des enchainements	
Pré condition	l'existence d'employé
Démarrage	L'utilisateur demande le formulaire de demande de congé
Enchainements	e1 :L'utilisateur demande le formulaire de demande de congé e2 :Le système afficher le formulaire e3 :L'utilisateur saisit les informations de congé e4 :Le système enregistre le congé
Arrêt	Lorsque le congé est enregistré
Exception	Si l'employé n'a pas le droit à un congé
Post condition	Ajouter un nouveau congé

TABLE 3.2 – Description textuelle de cas d'utilisation

3.4.3 Schéma conceptuel des bases de données sources(Base_Hôpital)

Les bases de données (BDD) constituent le cœur du système d'information. La conception

de ces bases est une tâche très importante du processus de développement du système d'information. Le schéma suivant représente notre schéma conceptuel :

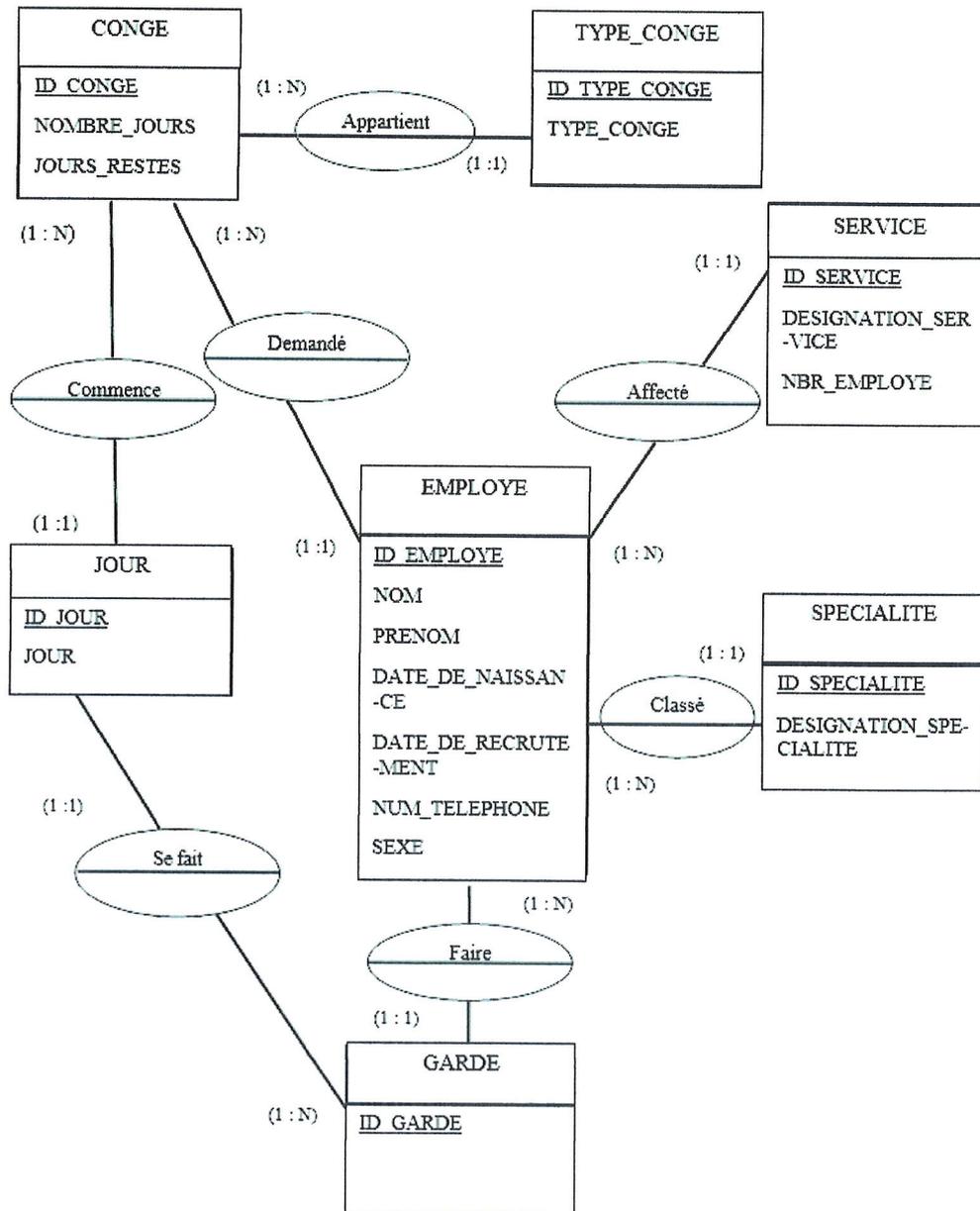


FIGURE 3.5 – Schéma conceptuel de bases de données sources

3.4.4 Le modèle relationnel

Notre modèle de base de données relationnel est définie comme suit :

SERVICE (ID_SERVICE,DESIGNATION_SERVICE, NBR_EMPLOYE)
 SPECIALITE (ID_SPECIALITE,DESIGNATION_SPECIALITE)
 EMPLOYE (ID_EMPLOYE,NOM,PRENOM, DATE.DE.NAISSANCE,DATE.DE.RECRUTEMENT,
 NUM.TELEPHONE,SEXE,# ID_SERVICE,# ID_SPECIALISTE)
 JOUR (ID_JOUR, JOUR)
 GARDE (ID_GARDE, # ID_EMPLOYE, # ID_JOUR)
 TYPE_CONGE (ID_TYPE_CONGE, TYPE_CONGE)
 CONGE (ID_CONGE, NOMBRE_JOURS, JOURS_RESTES, #ID_TYPE_CONGE,#
 ID_EMPLOYE,# ID_JOUR)

3.5 Conception de l'entrepôt de données (Absence schéma)

Concernant la modélisation conceptuelle, nous proposons un schéma en Flocon de Neige qui nous permet d'assurer une grande flexibilité dans les données et mieux exprimer nos besoins pour avoir des résultats clairs et rapides.

3.5.1 Les dimensions et les hiérarchies

Nous allons présenter les données nécessaires pour définir chacune des dimensions et hiérarchies.

Mesures	Dimensions	Hiérarchie	
NBR_ABSENCE	EMPLOYE_D →	SEXE_D	
TAUX_ABSENCE	PERIODE_D →	MOIS_D →	ANNEE_D
	SERVICE_D		
	SPECIALITE_D		
	CONGE_D		

TABLE 3.3 – modélisation d'entrepôt de données

3.5.1.1 La dimension DIMENSION_EMPLOYE

Cette dimension est définie par la hiérarchie qui contient les niveaux suivant :

— EMPLOYE_D :

Nom d'attribut	Type d'attribut	Description
NUM_EMPLOYE	Number	La clé primaire
NOM	Varchar2	Le nom d'employé
PRENOM	Varchar2	Le prénom d'employé
DATE_NAISSANCE	Date	Date de naissance
DATE_RECRUTEMENT	Number	Le recrutement d'employé
NUM_TELEPHONE	Varchar2	Numéro de téléphone
ID_SEXE	Varchar2	La clé étrangère pour la table SEXE_D

TABLE 3.4 – EMPLOYE_D

— SEXE_D :

Nom d'attribut	Type d'attribut	Description
ID_SEXE	Varchar2	La clé primaire
SEXE	Varchar2	Le sexe d'employé

TABLE 3.5 – SEXE_D

3.5.1.2 La dimension DIMENSION_TEMPS

Cette dimension est définie par la hiérarchie qui contient les niveaux suivant :

— PERIODE_D :

Nom d'attribut	Type d'attribut	Description
NUM_PERIODE	Varchar2	La clé primaire
DESIGNATION_PERIODE	Varchar2	La désignation du jour
NUM_MOIS	Number	La clé étrangère pour la table MOIS_D

TABLE 3.6 – PERIODE_D

— MOIS_D :

Nom d'attribut	Type d'attribut	Description
NUM_MOIS	Number	La clé primaire
DESIGNATION_MOIS	Varchar2	La désignation du mois
NUM_ANNEE	Number	La clé étrangère pour la table ANNEE_D

TABLE 3.7 – MOIS_D

— ANNEE_D :

Nom d'attribut	Type d'attribut	Description
NUM_ANNEE	Number	La clé primaire
DESIGNATION_ANNEE	Varchar2	La désignation d'année

TABLE 3.8 – ANNEE_D

3.5.1.3 La dimension DIMENSION_CONGE

Cette dimension contient :

— CONGE_D :

Nom d'attribut	Type d'attribut	Description
ID_CONGE	Varchar2	La clé primaire
NOMBRE_JOURS	Number	La durée du congé
JOURS_RESTES	Number	Les jours restes du congé
TYPE_CONGE	Varchar2	Type du congé

TABLE 3.9 – CONGE_D

3.5.1.4 La dimension DIMENSION_SERVICE

Cette dimension contient :

— SERVICE_D :

Nom d'attribut	Type d'attribut	Description
NUM_SERVICE	Number	La clé primaire
DESIGNATION_SERVICE	Varchar2	Désignation du service
NBR_EMPLOYE	Number	Nombre d'employé dans chaque un des services

TABLE 3.10 – SERVICE_D

3.5.1.5 La dimension DIMENSION_SPECIALITE

Cette dimension contient :

— SPECIALITE_D :

Nom d'attribut	Type d'attribut	Description
ID_SPECIALITE	Varchar2	La clé primaire
DESIGNATION_SPECIALITE	Varchar2	Désignation de la spécialité

TABLE 3.11 – SPECIALITE_D

3.5.2 La table des faits

La table « ABSENCE_F » contient les mesures d'analyse et les clés étrangères qui référencent les dimensions d'analyse mentionnées précédemment, elle représente le nombre d'absence et le taux d'absentéisme pour chacun des employés.

Type d'attribut	Nom d'attribut	Type du champ	Description
Clés étrangères	NUM_EMPLOYE	Number	Représente l'identifiant d'un employé
	NUM_SERVICE	Number	l'appartenance de chaque employé à un service
	NUM_PERIODE	Varchar2	Le jour d'absence
	ID_SPECIALITE	Varchar2	La spécialité d'employé
	ID_SCONGE	Varchar2	La spécialité d'employé
Mesures	NBR_ABSENCE	Number	Nombre d'absence associé à chaque employé
	TAUX_ABSENCE	Number	Le taux d'absentéisme

TABLE 3.12 – La table des faits (ABSENCE_F)

Le schéma en flocon de neige de l'entrepôt de données (Absence_schéma) :

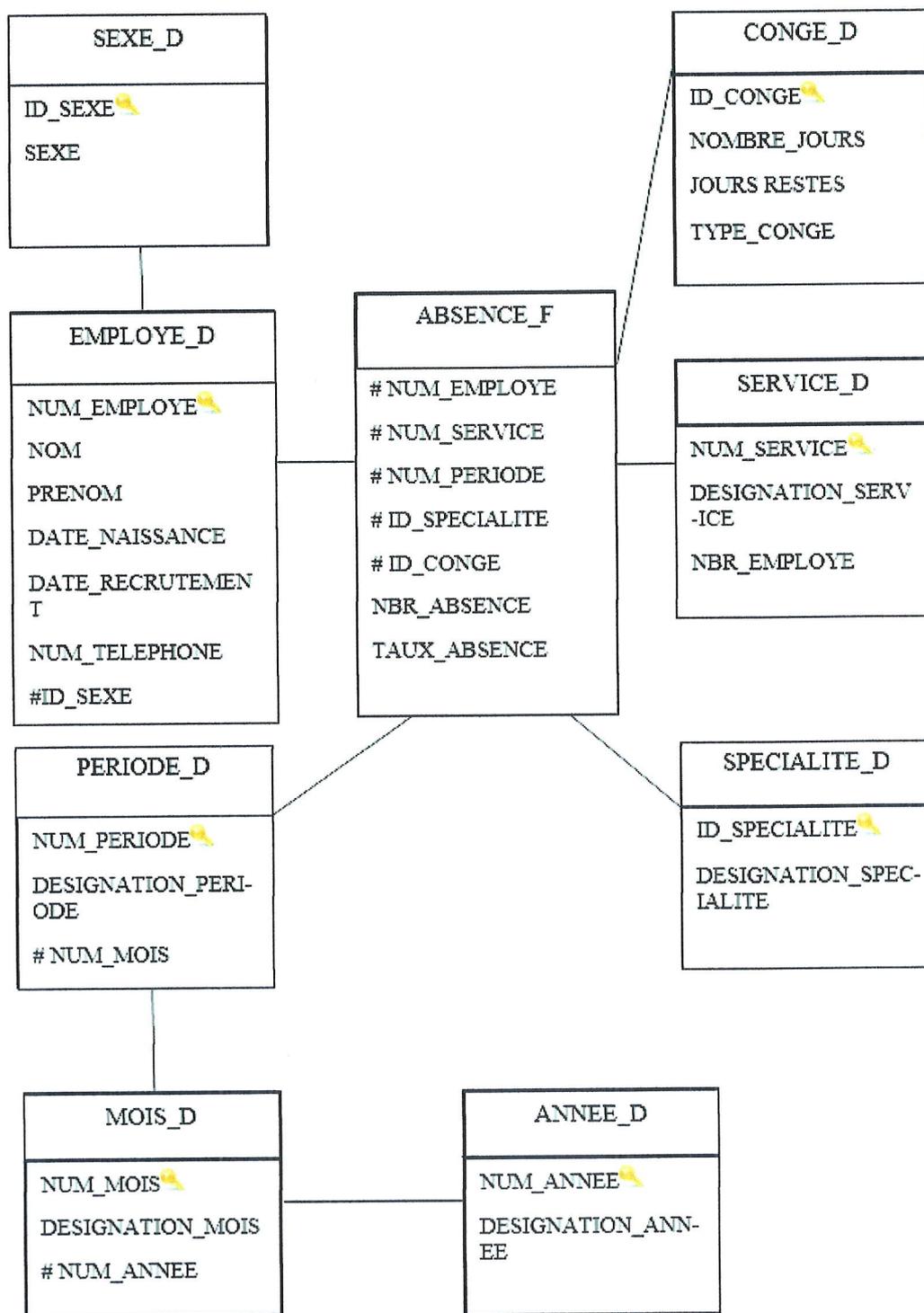


FIGURE 3.6 – schéma en flocon de neige

3.6 Conception des règles d'analyses

3.6.1 La syntaxe générale d'une règle d'analyse

Nous avons présentés une syntaxe générale d'une règle d'analyse qu'on a inspirée de « Thalhammer ». [14]

La règle : Nom de la règle.
Evenement : After/Before {Nom d'événement, Type d'événement}
Partie analyse :
Using cube {Nom de cube} (Les dimensions concernées)
If (Condition) {Les opérations OLAP}
From Table de fait
Partie trigger :
INSERT/DELETE/UPDATE { Nom de la table }

FIGURE 3.7 – La syntaxe générale d'une règle d'analyse

3.6.2 Les règles d'analyses

Dans cette partie on va décrire les règles d'analyses de notre travail :

3.6.2.1 Règle 1

La règle : nouveau recruté.
Evenement : Before { insert on employé, Fréquence}
Partie analyse :
Using cube Qualité de service (Dimension_Employe p)
ROLLUP p TO Employe_d
From Absence_F
Partie trigger :
INSERT INTO {employe}

FIGURE 3.8 – la règle d'analyse N°1

3.6.2.2 Règle 2

La règle : fin de carrière.
Evenement : Before { delete on employé, Fréquence}
Partie analyse :
Using cube Qualité de service (Dimension_Employe p)
If (période de travail >=30) ROLLUP p TO Employe.d
SLICE p.num_employe = "..."
From Absence.F
Partie trigger :
DELETE from {employe}

FIGURE 3.9 – la règle d'analyse N°2

3.6.2.3 Règle 3

La règle : Validation de congé.
Evenement : Before { insert on congé, Fréquence}
Partie analyse :
Using cube Qualité de service (Dimension_Conge c)
If (nombre_jours_restés <= Durée de congé) ROLLUP c TO conge.d
From Absence.F
Partie trigger :
INSERT INTO {conge}

FIGURE 3.10 – la règle d'analyse N°3

3.6.2.4 Règle 4

La règle : Calcule le reste de congé.
Evenement : After { fin_du_mois, Périodique}
Partie analyse :
Using cube Qualité de service (Dimension_Conge c)
ROLLUP c TO conge.d
From Absence.F
Partie trigger :
UPDATE {conge}

FIGURE 3.11 – la règle d'analyse N°4

3.7 Conclusion

Dans ce chapitre, nous avons présenté une brève présentation sur l'hôpital de Jijel et son organisation. Nous avons présenté aussi l'organigramme de l' E.P.H et ses différents services. Ainsi que l'architecture globale de notre système avec ses différents composants et la conception de base de données source, conception de l'entrepôt de données et la conception des règles d'analyses.

Dans le chapitre suivant, nous allons implémenter et mettre en œuvre ce que nous avons proposés dans ce chapitre, en d'autres termes, la réalisation de notre système.

L'implémentationl et réalisation

4.1 Introduction

Dans ce chapitre, consacré à la présentation des différents outils qui ont permis la réalisation et la mise en œuvre de notre application ainsi qu'une aperçu général sur la phase pratique qui consiste à décrire les différentes fonctionnalités disposées lors de l'utilisation de notre logiciel, ce qui permet d'éclaircir chaque étape de la réalisation.

4.2 Les outils d'implémentation

Pour réaliser notre projet nous avons utilisé, les outils suivants :

- Oracle Database 11g Express Edition.
- Langage de programmation : Java.
- Netbeans comme environnement de développement intégré (IDE) version 8.1.

4.2.1 NetBeans

NetBeans est un projet open source (libre) ayant un succès et une base d'utilisateur très large, une communauté en croissance constante, et près 100 partenaires mondiaux et des centaines de milliers d'utilisateur à travers le monde. Sun Microsystems a fondé le projet open source NetBeans en Juin 2000 et continue d'être le sponsor principal du projet.[20]

L'EDI NetBeans est un environnement de développement intégré, un outil pour les programmeurs pour écrire, compiler, déboguer et déployer des programmes. Il est écrit en Java, mais peut supporter n'importe quel langage de programmation. Il y a également un grand nombre de modules pour étendre l'EDI NetBeans. L'EDI NetBeans est un produit gratuit, sans aucune restriction quant à son usage.[20]

Également disponible, La Plateforme NetBeans; une fondation modulaire et extensible utilisée comme brique logicielle pour la création d'applications bureautiques. L'IDE offre également un support pour le développement d'applications PHP et C / C ++. Les partenaires privilégiés fournissent des modules à valeurs rajoutées qui s'intègrent facilement à la Plateforme et peuvent être utilisés pour développer ses propres outils et solutions. Les deux produits sont open source et gratuits pour un usage commercial et non-commercial. Le code source est disponible pour réutilisation sous la Common Development and Distribution License (CDDL).[20]

4.2.2 Oracle Database 11g Express Edition

Oracle Database 11g Express Edition (Oracle Database XE) est une base de données d'entrée de gamme de petite taille basée sur la base de code Oracle Database 11g Release 2. C'est gratuit à développer, déployer et distribuer, rapide à télécharger, et simple à administrer. Oracle Database XE est une excellente base de données de démarrage pour [21] :

- Développeurs travaillant sur les applications Node.js, Python, PHP, Java, .NET, XML et Open Source.
- DBA qui ont besoin d'une base de données gratuite pour la formation et le déploiement.
- Les éditeurs de logiciels indépendants (ISV) et les fournisseurs de matériel qui souhaitent distribuer gratuitement une base de données de démarrage.
- Les établissements d'enseignement et les étudiants qui ont besoin d'une base de données gratuite pour leur programme d'études.

Avec Oracle Database XE, vous pouvez désormais développer et déployer des applications avec une infrastructure performante et éprouvée, puis la mettre à niveau si nécessaire sans migrations complexes et coûteuses. Il peut être installé sur n'importe quelle machine hôte avec un nombre quelconque de processeurs (une base de données par machine), mais XE peut stocker jusqu'à 11 Go de données utilisateur, utiliser jusqu'à 1 Go de mémoire et utiliser un processeur sur la machine hôte.[21]

4.3 Présentation de l'application

4.3.1 Page d'authentification

Quand vous lancerez le logiciel, l'interface d'authentification s'apparaîtra sur l'écran, cette étape est pour sécuriser les données et protéger l'application, l'utilisateur doit s'identifier par son nom d'utilisation et son mot de passe pour pouvoir accéder à page d'accueil.



FIGURE 4.1 – Fenêtre de l’authentification.

4.3.2 Page d’accueil

Si le système valide le nom d’utilisateur et le mot de passe, l’interface principale sera affichée :

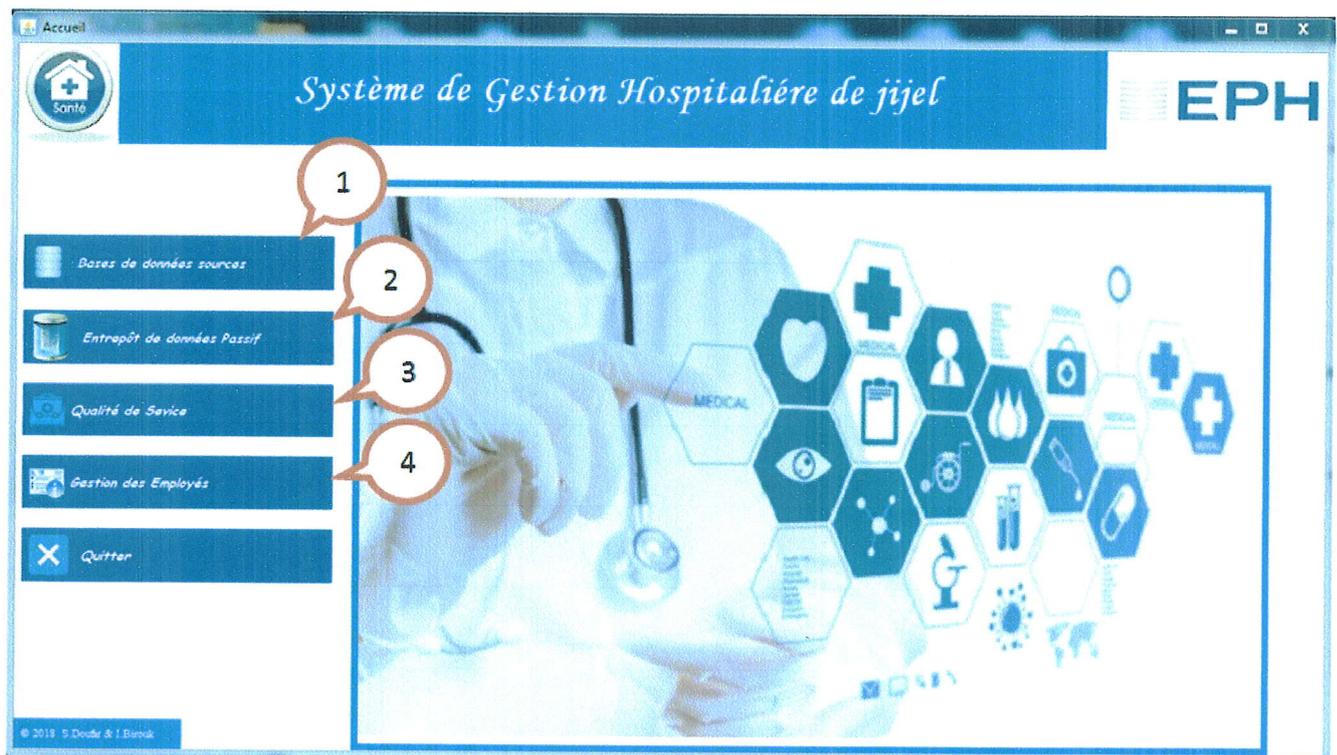


FIGURE 4.2 – Fenêtre du menu principale de l'application.

L'interface est composée de quatre boutons :

1. pour accéder à la page de bases de données sources.
2. pour accéder à la page d'entrepôt de données passif.
3. pour accéder à la page de la qualité de service.
4. Pour accéder à la page de gestion des employés.

4.3.2.1 Bases de données sources

Si l'utilisateur choisi « bases de données sources », l'interface suivante s'affiche :



FIGURE 4.3 – Espace de base de données source.

1. pour permettre à l'utilisateur de choisir la table de base de données source qu'il veut consulter.
2. Pour visualiser le choix de l'utilisateur à travers le bouton « Valider ».

4.3.2.2 Entrepôt de données passif

Si l'utilisateur choisi « Entrepôt de données passif », l'interface suivante s'affiche :

Dimension

Dimension Employé:

Dimension Spécialité:

Dimension Service:

Dimension Congé:

service	spécialité	type conge	nombre absence	taux absence
Service de pédiat...	medecin generali...	conge_paye	2	6.48
Service de medec...	medecin generali...	conge_paye	10	33.33
Service des urge...	medecin generali...	conge_paye	4	14.28
Service des urge...	medecin generali...	conge_paye	9	29.03
Service de pédiat...	medecin generali...	conge_paye	5	16.12
Service des urge...	medecin generali...	conge_paye	10	32.25
Service de méde...	medecin générali...	congé payé	14	45.16

Schéma Multidimensionnel (Cube)

specialite_D: id_specialite, designation

Employe_D: Num_empl..., Nom, Prenom, Date_Nais..., Date_Reeru..., Num_Telep..., Id_Sexe

Absence_F: Num_employe, Num_service, Num_periode, Id_specialite, Id_conge, Nbr_absence, Taux_absence

Service_D: Num_service, Designation_s..., Nbr_employe

Conge_D: Id_conge, Nombre_Jo..., Jours_Restes, Type_conge

Période_D: Num_periode, Designation_pe..., Num_mois

Sexe_D: Id_Sexe, Sexe

Mois_D: Num_mois, Designation_m..., Num_annee

Année_D: Num_annee, Designation_a...

FIGURE 4.4 – Espace d’entrepôt de données.

1. pour permettre à l'utilisateur de choisir la dimension qu'il veut consulter.
2. Pour visualiser le choix de l'utilisateur à travers le bouton « Analyser ».
3. Pour visualiser le schéma multidimensionnel.

4.3.2.3 Page de qualité de service

Si l'utilisateur choisi « qualité de service », l'interface suivante s'affiche :



FIGURE 4.5 – Fenêtre qualité de service.

1. pour permettre à l'utilisateur de choisir le service, la spécialité, le mois et le sexe dont il veut consulter les absences.
2. Pour visualiser le choix de l'utilisateur à travers le bouton « Analyser ».
3. Pour afficher le taux d'absence total selon les dimensions sélectionnées.
4. Pour accéder à la page « Analyse avancée ».

4.3.2.3.1 Page d'analyse avancée Si l'utilisateur choisi « Analyse avancée », l'interface suivante s'affiche :

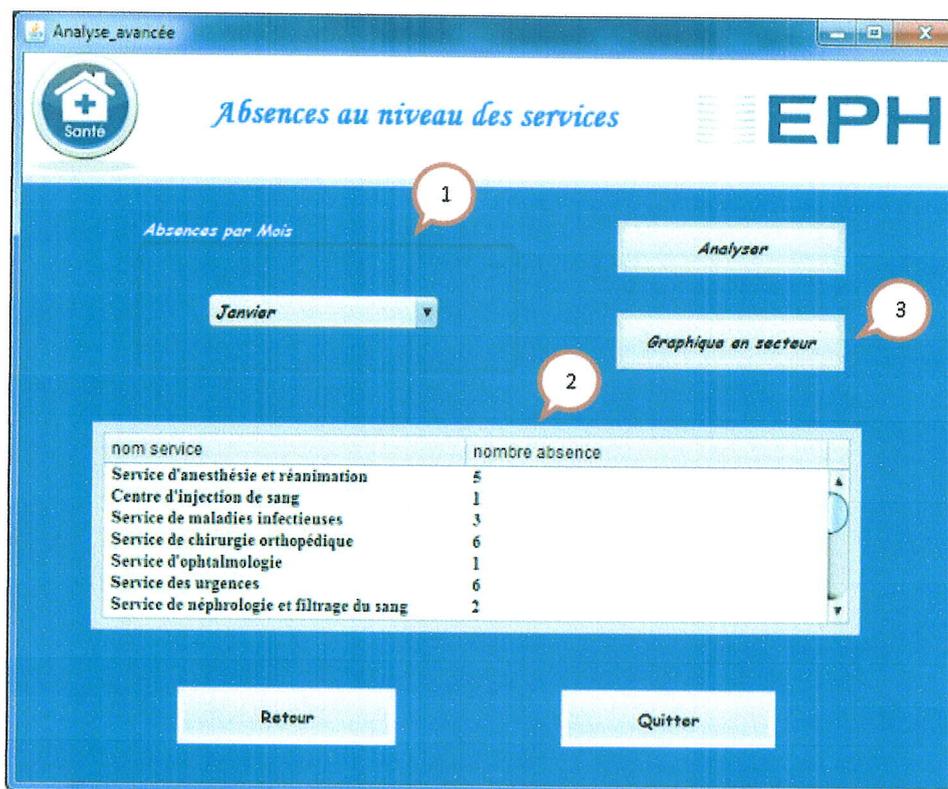


FIGURE 4.6 – Fenêtre d'analyse avancée

1. pour permettre à l'utilisateur de choisir le mois dont il veut consulter les absences au niveau des services.
2. Pour visualiser le choix de l'utilisateur à travers le bouton « Analyser ».
3. Pour afficher le résultat obtenu sous la forme graphique suivante :

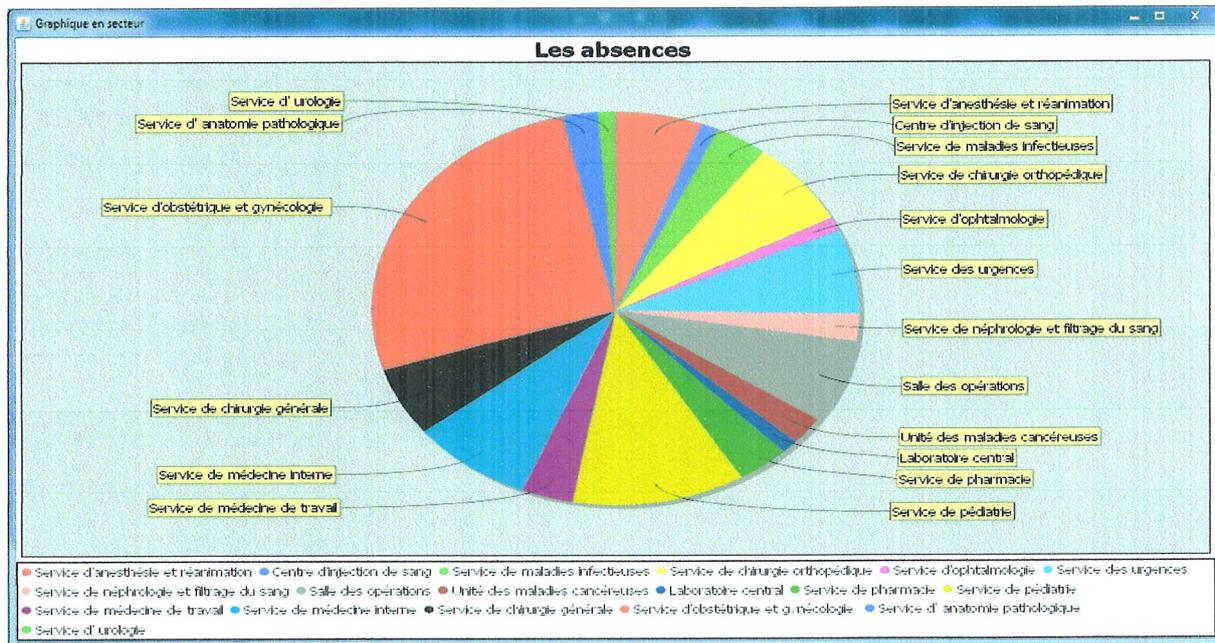


FIGURE 4.7 – l’absence des employés au niveau des services au mois de janvier 2018.

4.3.2.4 Page de gestion des employés

Si l'utilisateur choisi « gestion des employés », l'interface suivante s'affiche :

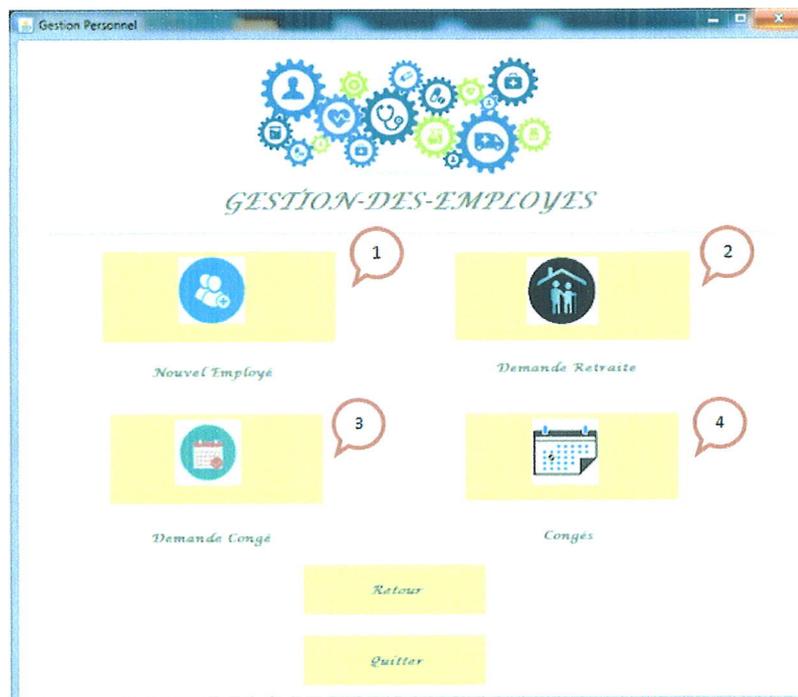
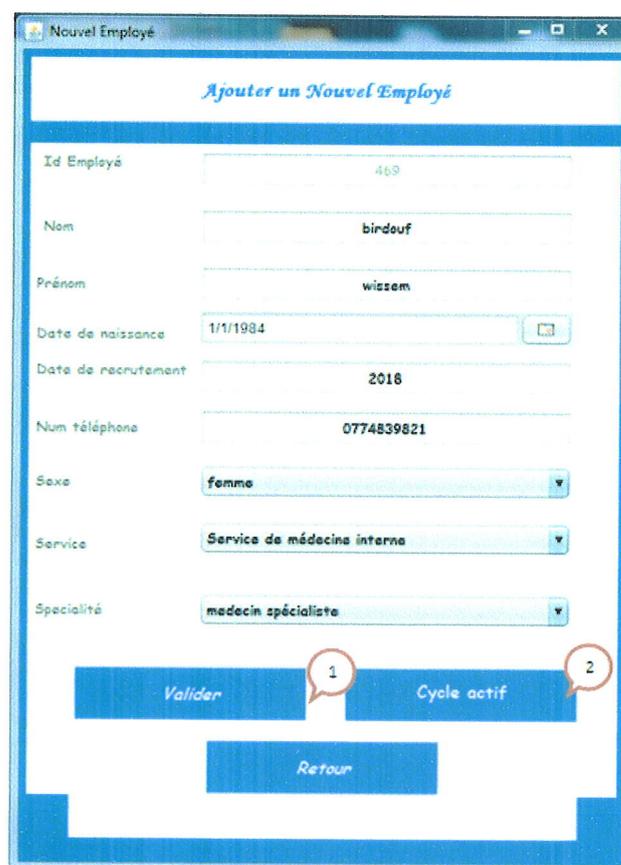


FIGURE 4.8 – Fenêtre de gestion des employés.

L'interface est composée de quatre boutons :

1. pour accéder à la page de nouvel employé.
2. pour accéder à la page de demande retraite.
3. pour accéder à la page de demande congé.
4. Pour accéder à la page des congés.

4.3.2.4.1 Formulaire d'ajout d'un nouvel employé Si l'utilisateur veut ajouter un nouvel employé, l'interface suivante s'affiche :



The screenshot shows a web application window titled "Nouvel Employé". The main heading is "Ajouter un Nouvel Employé". The form contains the following fields and values:

Id Employé	469
Nom	birdouf
Prénom	wizsem
Date de naissance	1/1/1984
Date de recrutement	2018
Num téléphone	0774839821
Sexe	femme
Service	Service de médecins interne
Spécialité	medecin spécialiste

At the bottom of the form, there are three buttons: "Valider" (with a callout bubble containing the number 1), "Cycle actif" (with a callout bubble containing the number 2), and "Retour".

FIGURE 4.9 – Fenêtre de saisir un nouvel employé.

1. Pour valider l'ajout d'un nouvel employé.
2. Pour afficher le cycle actif qui se réalise lors de l'ajout, ce ci est illustré dans l'interface suivante :

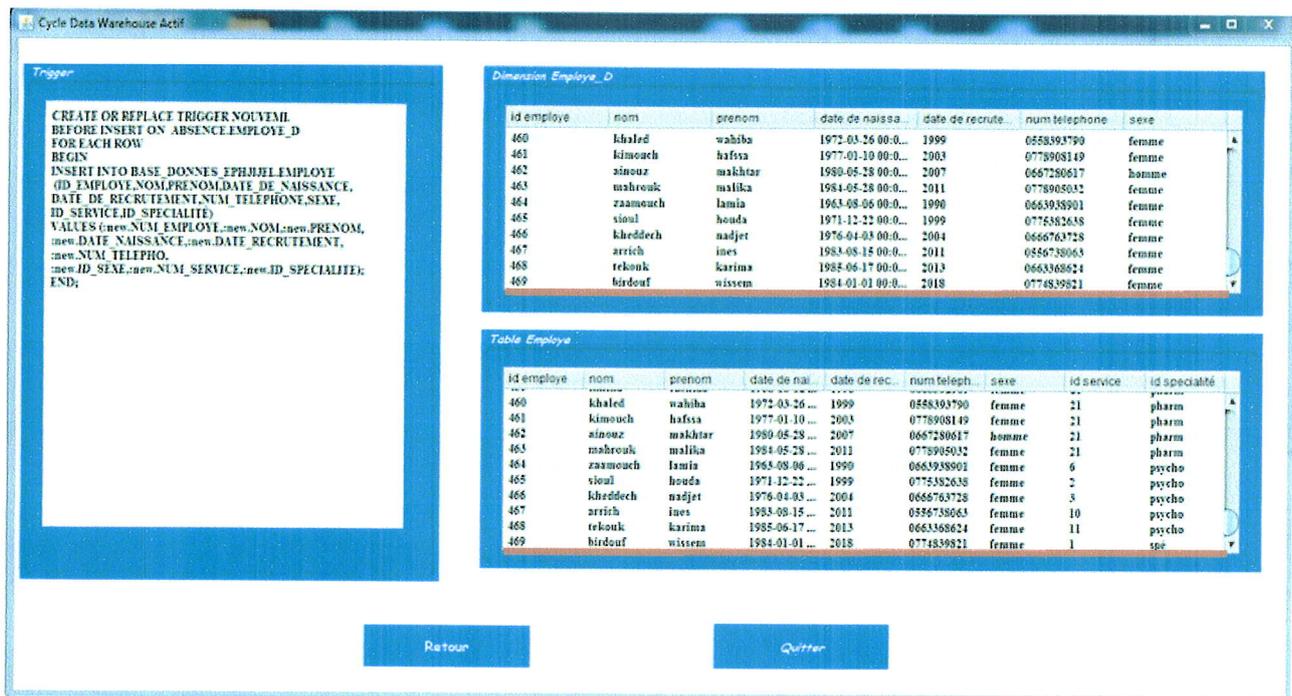


FIGURE 4.10 – Fenêtre de cycle actif.

Remarques : Les interfaces de demande de retraite, demande de congé et congés ont le même principe que l'interface de l'ajout d'un nouvel employé.

4.4 Conclusion

Dans ce chapitre nous avons présenté l'étape de la réalisation qui contient la description du logiciel ainsi que sa mise en œuvre, les interfaces principales de notre application et ses fonctionnalités.

Conclusion générale

Dans le cadre de ce mémoire, nous avons réalisé un outil d'aide à la décision pour améliorer la gestion hospitalière de l'hôpital de Jijel.

D'après le stage qu'on a fait dans établissement publique hospitalière de Jijel, nous avons vu qu'il y a une mauvaise exploitation des ressources. L'organisation hospitalière est peu développée. L'enregistrement des informations manuellement sur des supports de papier les faisant circuler d'un service à un autre engendrait beaucoup de problèmes tels que la perte de temps dans la recherche de ces informations... etc. En conséquence nous avons rencontré des problèmes durant la phase de collecte de données.

Tout au long de notre travail, nous avons abordé les points suivants :

- Nous avons tout d'abord proposé un système d'aide à la décision à base d'entrepôt de données pour améliorer la gestion administrative du personnel de l' Hôpital Mohamed Seddik Ben Yahia de Jijel.
- Dans un deuxième temps, nous avons défini une syntaxe pour la représentation des règles d'analyse. Notre proposition relative à la règle d'analyse consiste à décrire une formalisation simplifiée et généralisant l'utilisation des règles d'analyse dans le cadre des entrepôts de données actifs, qu'on a inspirée de « *Thalhammer* ». Cette syntaxe regroupe des opérateurs OLAP et OLTP.
- Enfin, nous avons fourni une implémentation de l'entrepôt de données actif pour automatiser le cycle de décision.

Bibliographie

- [1] Annoni Estella. Eléments méthodologiques pour le développement des systèmes décisionnels dans un contexte de réutilisation, 2007.
- [2] Olivier TESTE. Modélisation et manipulation d'entrepôts de données complexes et historisées, 18 décembre 2000.
- [3] Jiawei Han, Micheline Kamber, and Jian Pei Morgan Kaufmann. *Data Mining : Concepts and Techniques*. 2011.
- [4] W. Inmon and R.D. Hackathorn. *Using the Data Warehouse*. 1994.
- [5] María Trinidad Serna Encinas. Entrepôts de données pour l'aide à la décision médicale : conception et expérimentation, 27 Juin 2005.
- [6] Kamel Boukhalfa. De la conception physique aux outils d'administration et de tuning des entrepôts de données, 2009.
- [7] <https://grim.developpez.com/cours/businessintelligence/concepts/conception-datawarehouse/>.
- [8] <https://www.computerweekly.com/tip/Inmon-or-Kimball-Which-approach-is-the-suitable-for-your-data-warehouse>.
- [9] Sandeep Singh and Sona Malhotra. Data warehouse and its methods. *Journal of Global Research in Computer Science*, May 2011.
- [10] Traduction de Claude Raimond, *ENTREPOT DE DONNEES Guide pratique de modélisation dimensionnelle*. 2002.
- [11] Sandro Bimonte. L'information géographique et les entrepôts de données. *INSA de Lyon*, Décembre 2008.
- [12] Sonia Bouattour. Les entrepôts de données actifs : automatisation des scénarii d'analyse. Master's thesis, Université Lumière Lyon 2, 22 Juin 2007.
- [13] Sonia Bouattour, Jamel Feki, Hanene Ben Abdallah, Riadh Ben Messaoud, and Omar Boussaid. Proposition d'une taxonomie d'événement dans les entrepôts de données actifs. *Laboratoire ERIC*.

- [14] Thalhammer, T.M. Schrefl, and M. K.Mohania. Active data warehouses : Complementing olap with active rules, 2001.
- [15] Umeshwar Dayal, Eric N. Hanson, and Jennifer Widom. Active database systems, 1994.
- [16] Thierry Coupaye. Un modèle d'exécution paramétrique pour systèmes de bases de données actifs, 1992.
- [17] Hung Ledang. Des cas d'utilisation à une spécification b, (2001).
- [18] Pascal Roques and Franck Vallée. *UML2 en action de l'analyse des besoins à la conception*. 4^{ème} édition.
- [19] Faiza Ghozzi Jedidi. conception et manipulation de base de données dimensionnelles à contraintes. 18 novembre 2004.
- [20] https://netbeans.org/index_fr.html.
- [21] <http://www.oracle.com/technetwork/database/database-technologies/express-edition/overview/index.html>.
- [22] S.Chafki and C.Desrosiers. Mti820-entrepôts de données et intelligence d'affaires architecture des entrepôts de données, Hiver 2011.



RÉSUMÉ

L'entrepôt de données est une phase du processus décisionnel qui supporte efficacement le processus OLAP, il est né dans les entreprises pour l'aide à la prise de décision. Ainsi, les principaux utilisateurs de ces technologies font partie intégrante de l'entreprise, Nous pouvons citer les secteurs qui peuvent tirer profit des outils d'aide à la décision comme les institutions médicales, les banques et les assurances. Dans ce travail nous proposons une solution à base d'entrepôt de données actif permettant d'intégrer dans une même base cible des données et des traitements. Le mécanisme qui rend une telle approche active se base sur les règles d'analyse. Ces derniers permettent désormais de pallier à certains aspects statiques des entrepôts conventionnels en les dotant de mécanismes automatiques, qui sont capables de gérer d'une manière dynamique le cycle de vie d'un entrepôt.

L'objectif de notre travail consiste à réaliser un système d'aide à la décision pour automatiser certaines tâches de la gestion hospitalière au niveau de l'hôpital de Jijel.

Mots clés : Entrepôt de données, Entrepôt de données Actif, Entrepôt de données Médicales, Règle d'analyse.

ABSTRACT

The data warehouse is a phase of the decision-making process that effectively supports the OLAP process, it was born in companies for decision support. The main users of these technologies are an integral part of the company. With it we can name the sectors that can benefit from the decision support tools such as medical institutions, banks and insurance. In this work we propose a solution for an active data warehouse allowing data and processing to be integrated into the same target database. The mechanism that makes such an approach active is based on the rules of analysis. The latter now make it possible to compensate for certain static aspects of conventional warehouses by equipping them with automatic mechanisms, which are capable of dynamically managing the life cycle of a warehouse.

The objective of this work is to implement a decision support system to automate certain hospital management tasks at the Jijel hospital.

Key words : Data Warehouse, Active data warehouse, Medical Data Warehouse, Rule of analysis.