

N° d'ordre :  
N° de série :

**RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE**  
**Ministère de l'Enseignement Supérieur et de la Recherche Scientifique**



**UNIVERSITE ECHAHID HAMMA LAKHDAR - EL OUED**

**FACULTÉ DES SCIENCES EXACTES**

**Département D'Informatique**

**Mémoire de Fin D'étude**

**Présenté pour l'obtention du Diplôme de**

# **MASTER ACADEMIQUE**

Domaine: **Mathématique et Informatique**

Filière : **Informatique**

Spécialité : **Systèmes Distribués et Intelligence Artificielle**

Présenté par :

- **HASSI Nadjjet**
- **OTHMANI Nour elhouda**

## **Thème**

**Conception et Réalisation d'un Système  
Décisionnel Pour Les Collectivités Locales  
(Etude de cas : Des Décisions Financières)**

Soutenue le xx-xx- 2017 Devant le jury:

M.	KHALAIFA Nacer	MCA	Président
M.	NAOUI Mohamed Anwar	MAA	Rapporteur
M.	MEFTAH Mohamed Charafeddine	MCD	Encadreur

**Année Universitaire : 2016-2017**

### **Remerciements**

*Premièrement et avant toute chose, nous rendons grâce à Dieu le tout puissant pour toutes les vertus qu'il nous offre jour après jour, et pour nous avoir permis de suivre le chemin du savoir.*

*Ce modeste travail n'aurait pas pu voir le jour sans l'aide, la contribution et le soutien de plusieurs personnes. Nous voudrions leur rendre ici un hommage.*

*En premier lieu, nous remercions Dr. MEFTAH Mohamed Charafeddine de nous avoir encadrés, soutenus et si bien orientés tout au long de ce modeste projet.*

*Nous tenons à remercier également Mr Merouane Zoubeidi, Ghénabzia Ahmed, Tadjani Abdelraouf, Elanez Hafnaoui, Bougi Nissrin, Dik Ichrak et Med Daw qui n'ont épargné aucun effort pour nous aider et nous diriger durant la réalisation de ce travail*

*Nous remercions les membres du jury pour leurs attentions et intérêts portés envers notre travail. Merci de nous avoir honorés de votre présence.*

*Enfin, nous remercions l'ensemble des enseignants du département de l'informatique et toute personne qui a contribué de près ou de loin à l'élaboration de ce projet.*

## Résumé

La variation et la grande quantité des informations au sein de l'entreprise, a rendu l'opération du tri de ces informations pour choisir la meilleure et la plus exacte au moment exacte très difficile, cette chose a inquiété beaucoup les utilisateurs, ces derniers qui aiment de garantir le développement et la durabilité de l'entreprise.

L'opération de prendre la décision est une grande responsabilité, ici, les décisions prises par les responsables vont être le centre d'intérêt de tout le monde, cette chose rend ce responsable inquiet pour prendre la meilleure décision, à cause qu'il se base sur des stratégies non équilibrées qui résultent son utilisation des moyens traditionnels dans les rapports donnés par ses fonctionnaires concernant la situation financière de l'entreprise de HOUD BERKAOUI (entreprise qui dépend à SONATRACH), ces moyens sont très longs et compliqués, la chose qui les rend toujours difficiles à lire ou à comprendre, aussi bien sa préparation demande beaucoup de temps, à ce propos, la prise de décision demande aussi un plus de temps parce que si on prend une décision non étudiée ou précipitée on peut influencer négativement le processus du travail de l'entreprise.

Notre projet a pour but de construire un système d'aide de l'entreprise pour diminuer le pourcentage de l'échec des décisions prises pour l'amélioration de la situation de l'entreprise aussi bien sa compétence de développement car il est basé sur une stratégie pertinente dotée par des informations liées à la décision à travers les indices, les tableaux et les chiffres, ces derniers peuvent donner une vision très claire à la situation actuelle de l'entreprise aussi bien son développement à tout moment.

**Mots clés :** entrepôts de données, rapports, ETL, tableau de bord, le système de décision, le finance.

## **Abstract**

The diversity and large quantity of information inside the institution made the process to make the appropriate decision difficult. This made it a source of worry to those who are responsible of the decisions and who want to guarantee the institution's growth and continuity.

The process of making decisions is very important. It requires a great responsibility, where light is shed on the decisions made by the responsible. Hence, this makes the responsible confused about taking the right decisions and that through relying on inflexible strategies, which are the result of using conventional tools in order to get reports from his advisers and employers concerning the financial status of HOUDH BARKAOUI'S Institution, which is related to SONATRACH. These reports are often long and complex, what makes them difficult to be read and which, also, take a long time to be prepared. Any delay in making decisions or any incorrect conclusions taken from the reports would affect the institution negatively.

This project aims to create a Decision Support System, which helps the institution in reducing the decisions failure's rate, which are undertaken to improve the institution's condition and its ability of development. This is because the system is built on a strategy that supplies them with information, which are related to the decision, through interfaces, which consist of indexes, tables and graphs that give a clear image to the institution's current situation and its development every time.

**Key Words:** Data-warehouse, Reports, ETL, Dashboard, Decision Support System, Finance.

تنوع المعلومة وكثرتها داخل المؤسسة صعب عملية الحصول على الصحيحة منها في الوقت المناسب مما جعلها تشكل مصدر قلق رئيسي لأصحاب القرار الذين يرغبون في ضمان نمو المؤسسة و استدامتها. إن عملية اتخاذ القرار مسؤولية كبيرة و جد مهمة حيث يتم تسليط الضوء على القرارات المتخذة من المسؤول مما يجعله في حالة حيرة حول اتخاذه القرارات المناسبة عتماده على إستراتيجية غير ناتجة عن استخدامه للوسائل التقليدية في حصوله على التقارير من مستشاريه أو موظفيه في ما يخص مطولة و معقدة مما يصعب قراءتها ذ وقتنا طويلا في إجراءات إعدادها مما يؤدي أي تأخير في أخذ القرار أو أي استخلاص غير سليم من قراءة التقارير قد يؤثر سلبا على المؤسسة.

يهدف مشروع يساعد المؤسسة في التقليل من نسبة فشل القرارات المتخذة لتحسين وضع المؤسسة و قدرتها على التطوير لأنه مبني على إستراتيجية تزودهم ب ذات صلة بالقرار من خلال واجهات و التي تحتوي على مؤشرات و لبيانية رؤية واضحة حول الوضع الراهن لشركة و تطوره .

: مستودع البيانات تقارير ETL القيادة المالية.

## Sommaire :

Résumé .....	V
La liste des figures : .....	VI
La liste des tableaux : .....	VIII
Introduction générale : .....	1

### Partie I

#### Synthèse Bibliographique

Introduction synthèse bibliographique : .....	5
---	---

#### Chapitre I

##### Les systèmes décisionnels

1)Introduction : .....	6
2)Définition de système décisionnel : .....	6
3)Historique d'un système décisionnel : .....	8
4)Objectif d'un système décisionnel : .....	8
5)Les fonctionnalités de système décisionnel : .....	9
6)L'utilisation du système décisionnel : .....	12
7)Architecture des systèmes décisionnels : .....	13
8)Le déploiement et mise en œuvre d'un système décisionnel : .....	14
9)Les fouilles de données : .....	17
10)Conclusion: .....	18

#### Chapitre II

##### Les entrepôts de données

1)Introduction: .....	19
2)Définition d'un entrepôt de données (Data Warehouse) : .....	19
3)Les caractéristiques fondamentales d'entrepôt de données: .....	20
4)Historique des entrepôts de données: .....	22
5)Objectifs d'entrepôt de données: .....	23
6)La structure de l'entrepôt de données.....	24
7)Composants de base et architecture d'un entrepôt de données : .....	25
8)Méthodologie de Construction d'un entrepôt de données : .....	28
9)Alimentation d'entrepôt de données: .....	31
10)Modélisation dimensionnelle : .....	33
11)Différents modèles de la modélisation dimensionnelle : .....	35
12)Le cube de données : .....	36
13)Le concept OLAP (On Line Analytical Process): .....	37
14)Comparaison entre MOLAP, ROLAP et HOLAP : .....	39

15)La navigation dans les données : .....	39
16)Conclusion : .....	42
Partie II	
Partie Pratique	
Introduction partie pratique .....	43
Chapitre I	
Etude de cas	
« Service Finance - HAOUUD BERKAOUI »	
(Analyse de Besoin et étude de l'existant)	
1)Introduction : .....	45
2)Présentation de la HAOUUD BERKAOUI : .....	45
3)Etude de l'existant: .....	48
4)Etude des besoins : .....	49
5)Obstacles et problèmes rencontrés : .....	51
6)Conclusion : .....	52
Chapitre II	
Les conceptions	
1)Introduction : .....	53
2)La conception de la zone d'entreposage:.....	53
3)Conception de la zone « Alimentation » : .....	56
4)Conception des cubes dimensionnels: .....	60
5)Conclusion : .....	62
Chapitre III	
Réalisation	
1)Introduction : .....	63
2)L'architecture technique globale de la solution : .....	63
3)Outils utilisés : .....	63
4)Réalisation du projet : .....	65
5)Conclusion : .....	69
Conclusion	
générale	
Conclusion générale: .....	70
Bibliographie	
Bibliographie : .....	72
Liste des abréviations : .....	75

## La liste des figures :

---

### La liste des figures :

#### Partie I : synthèse bibliographique

##### Chapitre I : Les systèmes décisionnels.

<b>Figure I - 1:</b> Les 4 phases du fonction de la système décisionnel [7].....	9
<b>Figure I - 2:</b> Les utilisateurs du système décisionnel.....	12
<b>Figure I - 3:</b> Architecture des systèmes décisionnels.....	13
<b>Figure I - 4:</b> Les entrepôts de données et datamarts [10].....	14
<b>Figure I - 5:</b> Exemple d'un état de sortie en rapport[8].....	15

##### Chapitre II : Les entrepôts de données.

<b>Figure II - 1:</b> Définition de l'entrepôt de données[18].....	20
<b>Figure II - 2:</b> Orientation sujet des entrepôts de données[8].....	20
<b>Figure II - 3:</b> L'intégration dans les entrepôts de données[8].....	21
<b>Figure II - 4:</b> L'historisation des entrepôts de données[8].....	21
<b>Figure II - 5:</b> Non volatile des entrepôts de données [22].....	22
<b>Figure II - 6:</b> Evolution des bases de données décisionnelles[19].....	23
<b>Figure II - 7:</b> Structure d'un entrepôt de données[22].....	24
<b>Figure II - 8:</b> Composants de base et architecture d'un entrepôt de données[26].....	28
<b>Figure II - 9:</b> Architecture du DW selon Inmon [8].....	29
<b>Figure II - 10:</b> Architecture du DW selon Kimball [8].....	30
<b>Figure II - 11:</b> Le système d'alimentation ETL [28].....	31
<b>Figure II - 12:</b> Tableau de fait.....	34
<b>Figure II - 13:</b> Tableau dimension.....	35
<b>Figure II - 14:</b> Exemple d'un schéma conceptuel en étoile [8].....	35
<b>Figure II - 15:</b> Exemple d'un schéma de données en flocon [8].....	36
<b>Figure II - 16:</b> Exemple d'un cube de données [30].....	36
<b>Figure II - 17:</b> Schématisation de la technique M-OLAP [28].....	37
<b>Figure II - 18:</b> Schématisation de la technique R-OLAP [28].....	38
<b>Figure II - 19:</b> Schématisation de la technique H-OLAP.....	39
<b>Figure II - 20:</b> Exemple du le Tranchage [8].....	40
<b>Figure II - 21:</b> Exemple Le Découpage [8].....	40
<b>Figure II - 22:</b> Exemple Le Forage vers le haut.....	41
<b>Figure II - 23:</b> Exemple Le forage vers le bas.....	41

## La liste des figures :

---

### Partie II : Partie Pratique

<b>Figure II</b> : Le cycle de vie d'un projet décisionnel.....	44
Chapitre I : Etude de cas « Service Finance - HAOUD BERKAOUI »	
<b>Figure I - 1</b> : L'organigramme de direction régionale HAOUD BERKAOUI .....	46
<b>Figure I - 2</b> : Organigramme de la structure division finance.....	47
Chapitre II : Les conceptions	
<b>Figure II - 1</b> : Schéma en étoile du processus « Charge ».....	56
<b>Figure II - 2</b> : Cube dimensionnel « Charge ».....	61
Chapitre III : Réalisation	
<b>Figure III - 1</b> : L'architecture technique globale de la solution.....	63
<b>Figure III - 2</b> : La construction de la partie ETL. ....	66
<b>Figure III - 3</b> : La construction et chargement des cubes. ....	67
<b>Figure III - 4</b> : Exemple L'éditeur Ad Hoc. ....	68
<b>Figure III - 5</b> : Exemple d'un tableau de bord sur JasperSoft Server.....	68

## La liste des tableaux :

---

### La liste des tableaux :

Partie I : synthèse bibliographique

Chapitre II : Les entrepôts de données

<b>Tableau II - 1 :</b> Tableau comparatif entre les approches de mise en place d'un entrepôt de données selon Inmon et Kimball [4].....	30
<b>Tableau II - 2 :</b> Tableau comparatif entre R-OLAP , M-OLAP et H-OLAP [8]. .....	39

Partie II : Partie Pratique

Chapitre II : Les conceptions

<b>Tableau II - 1 :</b> Les dimensions participantes aux différents. ....	54
<b>Tableau II - 2 :</b> Tableau descriptif de la dimension « Date ». ....	54
<b>Tableau II - 3 :</b> Tableau descriptif de la dimension «Compte». ....	55
<b>Tableau II - 4 :</b> Tableau descriptif de la dimension «Région». ....	55
<b>Tableau II - 5 :</b> Liste des agrégats « Charge ». ....	55
<b>Tableau II - 6 :</b> Table de source des données.....	57
<b>Tableau II - 7 :</b> Liste des niveaux et des hiérarchies de chaque dimension. ....	60
<b>Tableau II - 8 :</b> Le cube « Charge ». ....	61

# Introduction Générale

### **Introduction générale :**

#### **Contexte du travail :**

Au cours des changements économiques et technologiques, la gestion des informations est devenue la chose la plus importante pour tous les secteurs vitaux de l'économie dans le monde entier, elle est aussi considérée comme un facteur primordial pour améliorer la performance des entreprises, avec le défi de recueillir et stocker des données qui vont être traitées avec les outils informatiques et des logiciels (Mysql, ETL,...etc) et les utiliser pour l'analyse (par l'OLAP) et prendre les décisions.

Récemment l'institution HAUD BERKAOUI a commencé à réaliser l'importance des informations, mais elle est encore loin de l'utilisation de la technologie moderne pour l'organisation des informations, l'utilisation des moyens traditionnels généralement cause généralement une désorganisation et incohérence des informations accumulées, alors ses moyens rendent des résultats non fiables à cause du manque d'objectivité et de fiabilité des données, ce qui complique la prise des décisions.

Par conséquent, l'institution a dû obtenir un système de support pour aider à prendre les décisions considéré comme un système d'informations connecté à l'unité qui provient des utilisateurs du support nécessaire pour prendre les bonnes décisions à partir de l'analyse des informations compliquées, et simuler l'opération pour avoir les conséquences des décisions avant de les réaliser, l'entrepôt des données est considéré comme l'élément le plus important des systèmes de support de la prise des décisions, alors qu'on le considère comme la base des données d'assemblage, d'organisation, d'enregistrement et de stockage d'information à partir des bases des données opérationnelles, et la fourniture de support de la prise des décisions.

Ce genre de bases de données se distingue de la similitude des composants internes avec ce que l'utilisateur a besoin comme indice et axe d'analyse, l'importance de ce système est de répondre aux demandes de l'institution et d'améliorer la performance et permettre au responsable d'avoir une idée générale des activités après avoir les données par un panneau de contrôle désignant spécifiquement pour prendre les décisions.

#### **Problématique :**

HAUD BERKAOUI est divisée en neuf sections, chacune avec sa propre base de données, un surplus majeur prend place des données qui sont généralement inorganisées et incohérentes ce qui complique le processus de recueillir et analyser les données des activités

financières, pour répondre aux questions à propos de la performance et du suivi financier de l'institution.

Les responsables de la direction financière et comptabilité déclarent qu'il se face à quelque problème durant le processus d'apprendre une décision tels que :

- Coût de la procédure de reportage : la procédure de reportage est jugée très coûteuse pour l'entreprise, et cela est principalement dû au nombre d'intervenants et des moyens mis en place pour cette dernière.
- Difficultés dans l'élaboration et la consolidation des rapports d'activités. Cette procédure se fait manuellement ce qui augmente le risque d'erreur et cause une grande perte de temps.
- rapports proposés par les différentes applications métier ne donnent plus satisfaction à cause de leurs limites (rapports figés et non paramétrés).
- Grand volume de données stockées dans plusieurs bases de données, l'exploitation de ces données est donc très difficile.
- La décentralisation et l'hétérogénéité des données : Les données sont stockées dans plusieurs bases de données et ne sont pas intégrées, alors l'information n'est pas disponible en temps voulu.
- Difficultés d'accéder aux informations en temps voulu. en vue de prendre des décisions.

### **Objectifs du projet :**

Pour résoudre les problèmes de l'organisation mentionnée ci-dessus, nous allons mettre les bases nécessaires à la mise en œuvre du système de prise de décision. Dans le cadre de ce projet, notre mission est avant tout la conception et la mise en œuvre de l'entrepôt de données qui utilise les données stockées dans différentes sources. Ensuite, nous allons concevoir et mettre en œuvre un panneau de contrôle qui contient les plus importants indicateurs de performance qui facilitent le processus de navigation dans l'entrepôt de données qui répondent le plus urgent analytique et simplifient les exigences de gestion d'entreprise. Donc les principaux objectifs de ce projet peuvent être résumés comme suit :

- Optimiser le temps d'élaboration des rapports.
- Réduire le nombre d'intermédiaire à travers une seule base de données (DW) lors de la production de rapports et améliorer aussi la qualité des rapports.
- Rendre disponible à tout moment des informations nécessaires à la prise de décision afin d'effectuer les analyses appropriées.

- L'intégration et la Centralisation des données afin de faciliter l'accès aux informations, sans négliger l'aspect sécurité.
- Faciliter l'exploitation des informations stockées (données organisées selon différentes dimensions).
- Mettre à la disposition des utilisateurs, dans des délais très courts, des états de sorties multiples et personnalisés (rapport, graphes, ...).
- Offrir à l'utilisateur la possibilité de naviguer dans les données de manière interactive à l'aide des cubes OLAP.
- Permettre d'établir un diagnostic financier spécifique aux besoins des décideurs en donnant l'accès à des indicateurs pertinents, fiables, facilement interprétables et mesurés dans le temps.
- Performant sur les temps de réponse et l'exactitude des résultats.
- La Réduction des coûts de la procédure de reportage actuelle.
- La réduction du nombre d'intervenants lors de la production de rapports.
- Offrir des informations fiables, cohérentes et pertinentes, contenant le logique business souhaité.

### **Organisation du mémoire :**

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

#### **Partie I : synthèse bibliographique**

Regroupe les chapitres suivants :

- **Chapitre I: Les systèmes décisionnels**

Ce chapitre décrit brièvement les caractéristiques d'un système décisionnel qui est le thème de ce projet de fin d'études.

- **Chapitre II: Les entrepôts de données**

Ce chapitre comprend les concepts de l'entrepôts de données à travers une présentation de ces concepts principaux, d'un ensemble de définitions, des étapes de construction d'un entrepôt de données, et enfin du concept de DataMart.

#### **Partie II : Partie pratique**

- **Chapitre I: Etude de cas « Service Finance - HAUD BERKAOUI » :**

Dans ce chapitre, nous allons faire une présentation de l'organisme d'accueil, son environnement, sa structure ainsi que ses activités. Nous passerons par la suite à l'analyse de l'existant en systèmes d'information et en systèmes décisionnels pour finir avec des critiques

de l'existant et les techniques de collecte d'information utilisées, la démarche suivie afin d'identifier les besoins pour finir avec un récapitulatif des besoins recueillis.

- **Chapitre II: Les conceptions**

Ce chapitre porte sur la conception de l'entrepôt de données à savoir : la conception de la zone d'entreposage, la conception de la zone d'alimentation, la conception des cubes dimensionnels.

- **Chapitre III: Réalisation**

Dans ce chapitre, nous allons présenter l'architecture technique de notre solution, l'ensemble des outils utilisés ainsi que les différentes étapes de sa réalisation.

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

Partie I  
Partie I

Synthèse  
Synthèse

Bibliographique  
Bibliographique

## **Introduction synthèse bibliographique :**

Les systèmes d'aide à la décision (DSS) ont été émergés, leurs applications et techniques ont parfaitement été améliorées. Ces systèmes fournissent le support approprié pour optimiser la qualité des décisions en s'appuyant sur les informations existantes dans l'institution. L'objectif de cette partie est de mettre en évidence ce que les systèmes d'aide à la décision (DSS), leur rôle, leur évolution historique et les éléments de leur construction.

# Chapitre I

## Chapitre I

Les systèmes

## Les systèmes

décisionnels

## décisionnels

## 1) Introduction :

Avec l'apparition du phénomène de la mondialisation et le développement des technologies de l'information, la mesure de la performance de l'activité d'une entreprise est devenue un élément indispensable du plan d'action des entreprises pour l'atteinte de leurs objectifs, la conservation de leurs parts de marché et l'optimisation de leurs processus métier. Un système d'information d'aide à la décision est l'un des outils permettant de répondre d'une façon parfaite à ce besoin.

A travers ce premier chapitre, nous allons voir qu'est ce que le système décisionnelle, les outils qu'elle offre en terme d'exploitation des données dans les organisations pour la production des informations pertinentes, aidant les décideurs de l'entreprise à prendre des décisions judicieuses.

## 2) Définition de système décisionnel :

Les systèmes décisionnels s'inscrivent dans la famille des systèmes d'information, ils sont souvent appelés SID (système d'information décisionnelle) en anglais : DSS pour Decision Support System ou encore BI pour Business Intelligence.

Parmi les nombreuses définitions des systèmes décisionnels que nous avons étudié, nous avons constaté que Devisy a dit : « Le SID est un système permettant aux décideurs de l'entreprise de disposer des informations pertinentes et des outils d'analyse puissants pour aider à prendre les bonnes décisions aux bons moments» [1].

Donc d'un système décisionnel correspond à l'ensemble des moyens, des outils et des méthodes qui permettent de collecter, consolider, modéliser et restituer les données internes et externes, matérielles ou immatérielles, d'une entreprise en vue d'offrir une aide à la décision et de permettre aux responsables de la stratégie d'entreprise d'avoir une vue d'ensemble de l'activité traitée.

Nous pouvons élargir cette définition, en reprenant les caractéristiques suivantes [2] :

- **Petit nombre d'utilisateurs :**

L'aide à la décision stratégique est le lot de quelques personnes dans l'entreprise (décideurs, patrons).

- **Données générales et détaillées :**

On s'intéresse ici aux chiffres par mois, par année, par groupe de produit... etc. Les décideurs n'ont pas intérêt à voir la commande de tel ou tel client. Ils veulent voir l'ensemble des activités. Par contre, les analystes ont tout intérêt à pouvoir creuser dans les données pour trouver des fraudeurs par exemple.

- **Rapidité suggérée :**

Il est clair que plus c'est rapide et mieux, mais dans la prise des décisions stratégiques, on ne calcule pas à la seconde. Un décideur peut bien attendre quelques heures pour avoir une information très complexe à créer. Mais dans la plupart des cas, les temps de réponses doivent être calculés en secondes.

- **Ouverts :**

Contrairement au monde opérationnel, on laisse libre court à la curiosité des utilisateurs, les environnements de SD doivent permettre d'accéder le plus simplement possible aux données et d'en faire tout ce qu'on veut !

- **Gros volumes de données :**

Les environnements de SD doivent regrouper toutes les données de l'entreprise, de la ligne de commande au chiffre d'affaire annuel. Des années et des années d'accumulation de données génèrent des Gigas Octets qui doivent être gérés par les environnements de SD.

- **Non transactionnels :**

Ici, pas de processus rigide. L'utilisateur doit pouvoir commencer une analyse, revenir en arrière, démarrer une autre analyse en parallèle, envoyer un résultat à un collègue pour qu'il puisse creuser une autre piste.

- **Données en lecture seule :**

Dans le mode décisionnel il ya pas de perte de données. On ne supprime jamais des données mais on les archive. Si le prix d'un produit change, on veut garder trace de cela.

- **Projets très risqués :**

En 2002, 40% des projets de SD ont échoué. La principale raison est le manque d'engagements de la part de la direction et le manque des connaissances dans le domaine.

- **Centralisés :**

Toutes les données sont regroupées en un seul point. La même source pour tout le monde.

### 3) Historique d'un système décisionnel :

Il est utile de faire un bref historique qui retrace les étapes qui ont marqué la marche du concept du système décisionnel [3].

En 1962 Ken Iverson a publié le premier langage multidimensionnel (APL: A Programming language) ce langage a été traduit par un premier outil multidimensionnel visant les applications de type marketing.

Dix ans après, apparaît le moteur OLAP conçu pour les applications financières (Banques, assurances ...).

En 1993 E.Codd a dicté les règles universelles qui décrivent les moteurs OLAP (modèle relationnel). Apparaissent ensuite deux grands théoriciens Ralph Kimball et Bill Inmon qui ont bouleversé entièrement le monde de l'information en transformant les données en une information fiable et cohérente en entreprise.

Ce qui a poussé les grandes firmes de l'informatique comme MICROSOFT, IBM à créer des outils puissants d'analyse pour la mise en application des données orientées métier. En plus la concurrence a obligé ces derniers à produire plus d'outils, de logiciels pour mettre en œuvre les indicateurs stratégiques de l'entreprise et la manière de les interpréter pour définir les prévisions et détecter les cibles ou les tendances.

En 2000, Microsoft a fait un véritable lifting de l'informatique décisionnelle et qui a voulu s'imposer sur le marché des SGBD face à Oracle et IBM, leader mondial de la Business Intelligence, en créant la version SQL server 2000. Ce qui a conduit à une grande concurrence entre les firmes et notamment entre MICROSOFT et IBM qui a fait progresser d'une façon spectaculaire le SD et ses applications [4].

### 4) Objectif d'un système décisionnel :

Nos différentes recherches sur le sujet d'un système décisionnel montrent que ce dernier permet toujours aux dirigeants d'avoir une vue globale sur les activités de leurs entreprises. Le système décisionnel a pour objectif [5]:

- **Anticiper les changements :**

En étudiant les informations historiques et en établissant des prévisions, le SD permet d'expliquer la performance actuelle et prédire le futur, à l'aide des projections et des modèles vraisemblables dans l'univers du possible.

- **L'accroissement de la force de la réactivité de l'entreprise :**

Le système décisionnel exige une flexibilité de la part des entreprises afin de répondre avec agilité aux besoins d'évolution, dans un monde en perpétuel mouvement où les variables de succès sont beaucoup et ne cessent pas d'évoluer.

- **Offrir aux décideurs une vision transversale de l'entreprise, en intégrant toutes les dimensions.**

- **Assurer la protection et la sécurisation des informations :**

Le système doit permettre le contrôle d'accès à ces informations confidentielles.

- **Conversion de la masse des données en une valeur métier :**

Le système à travers les outils d'analyse permet de dégager une valeur qui aide dans l'opération de la prise de décision.

### 5) Les fonctionnalités de système décisionnel :

Un système décisionnel assure quatre fonctions fondamentales, à savoir la collecte, l'intégration, la diffusion et la présentation des données. A ces quatre fonctions s'ajoute une fonction de contrôle du système décisionnelle lui-même, l'administration [6] :

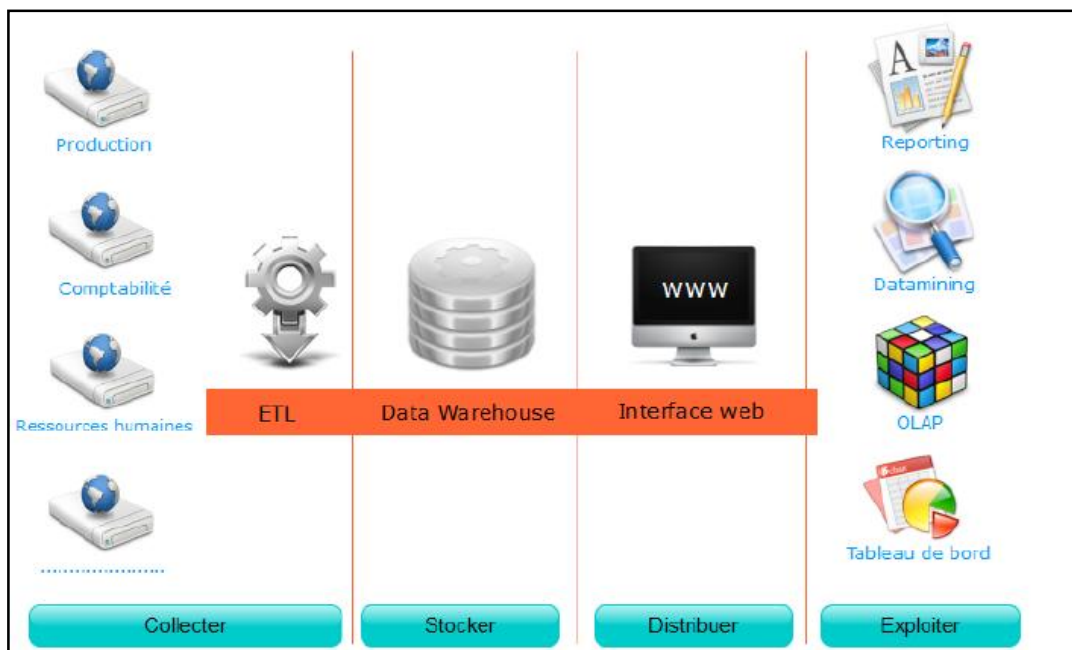


Figure I - 1: Les 4 phases du fonction de la système décisionnel [7].

- **La collecte :**

La collecte est l'ensemble des tâches qui consistent à détecter, à sélectionner, à extraire et à filtrer les données brutes issues des environnements pertinents compte tenu du périmètre du SD.

Les sources de données internes ou externes qui étaient souvent hétérogènes tant sur le plan technique que sur le plan sémantique, cette fonction est la plus délicate à mettre en place dans un système décisionnel complexe. Elle s'appuie notamment sur des outils d'ETL. Cette alimentation utilise les données sources issues des systèmes transactionnels de production, le plus souvent sous forme de :

Compte-rendu d'événement ou bien compte-rendu d'opération : il signifie le constat au fil du temps des opérations (achats, ventes, écritures comptables) le film des activités de l'entreprise.

Compte-rendu d'inventaire ou compte-rendu de stock : on peut le définir comme l'image photo prise à un instant donné (à une fin de période : mois, trimestre) de l'ensemble du stock (les clients, les contrats, les commandes, les encours). La fonction de la collecte joue également au besoin, un rôle de recodage. Une donnée représentée différemment d'une source à une autre impose le choix d'une représentation unique pour les futures analyses.

- **L'intégration :**

L'intégration consiste à concentrer les données collectées dans un espace unifié, dont le socle informatique essentiel est l'entrepôt. Élément central du dispositif, de sa part, il permet aux applications décisionnelles de bénéficier d'une source d'information commune, homogène, normalisée et fiable, susceptible de masquer la diversité de l'origine des données. Au passage les données sont épurées ou transformées par :

Un filtrage et une validation des données en vue de maintenir la cohérence de l'ensemble (les valeurs acceptées par les filtres de la fonction de la collecte sont susceptibles d'introduire des incohérences de référentiel par rapport aux autres données, donc elles doivent être soit rejetées, soit intégrées avec un statut spécial).

Une synchronisation (d'intégrer en même temps ou à la même date de valeur des événements reçus ou constatés de manière décalée ou déphasée).

Une certification (pour rapprocher les données de l'entrepôt des autres systèmes légaux de l'entreprise, comme la comptabilité ou les déclarations réglementaires). C'est également dans cette fonction qu'on effectue éventuellement les calculs et les agrégations (cumuls) communs à l'ensemble du projet. La fonction d'intégration est généralement assurée par la gestion de métadonnées, pour l'interopérabilité entre toutes les ressources informatiques, des

données structurées (bases de données accédées par des progiciels ou applications) ou des données non structurées.

- **La diffusion ( Distribuer ) :**

La diffusion met les données à la disposition des utilisateurs, selon des schémas correspondant au profil ou au métier de chacun, sachant que l'accès direct à l'entrepôt ne correspondrait généralement pas aux besoins d'un décideur ou d'un analyste. L'objectif prioritaire est de segmenter les données en contextes informationnels fortement cohérents, simples à utiliser et correspondant à une activité décisionnelle particulière. Alors qu'un entrepôt de données peut héberger des centaines ou des milliers de variables ou indicateurs, un contexte de diffusion raisonnable n'en présente que quelques dizaines au maximum. Chaque contexte peut correspondre à un DataMart, bien qu'il n'y ait pas de règles générales concernant le stockage physique. Très souvent, un contexte de diffusion est multidimensionnel, c'est-à-dire modélisable sous la forme d'un hyper cube, donc il peut être mis à disposition à l'aide d'un outil OLAP. Les différents contextes d'un même système décisionnel n'ont pas tous besoin du même niveau de détail. De nombreux agrégats ou cumuls, n'intéressant que certaines applications et n'ayant donc pas lieu d'être géré en tant qu'agrégats communs par la fonction d'intégration, relèvent donc de la diffusion. Ces agrégats peuvent être, au choix, stockés de manière persistante ou calculés dynamiquement à la demande. A ce propos, on peut distinguer trois questions à élucider pour concevoir un système de rapport :

- ✓ À qui s'adresse le rapport spécialisé ? (choix des indicateurs à présenter, choix de la mise en page).
- ✓ Par quel trajet ? (circuit de diffusion type flux de travail pour les personnes, circuits de transmission télécoms pour les moyens).
- ✓ Selon quel agenda ? (diffusion routinière ou sur événement prédéfini).

- **La présentation (Exploiter) :**

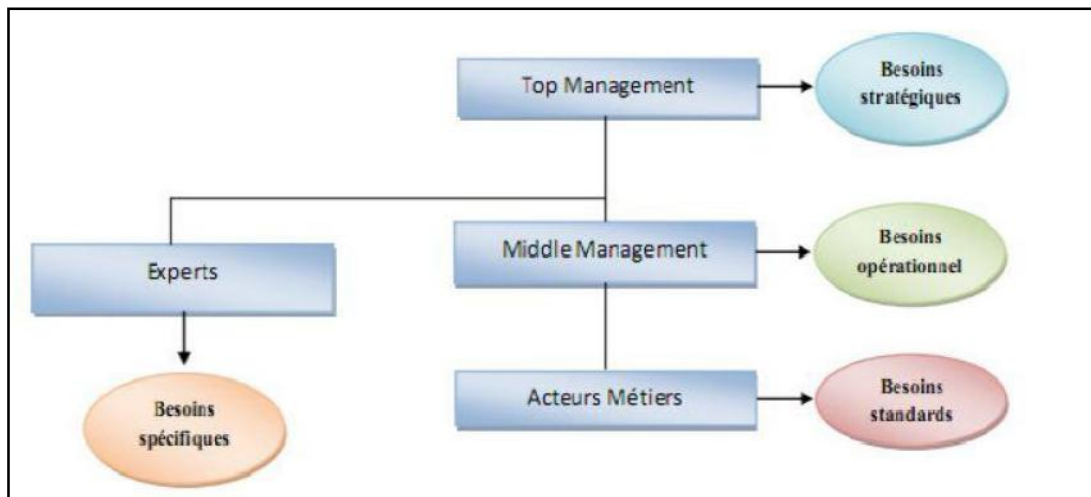
Cette quatrième fonction est la plus visible pour l'utilisateur, régit les conditions d'accès de l'utilisateur aux informations. Elle assure le fonctionnement du poste de travail, le contrôle d'accès, la prise en charge des requêtes, la visualisation des résultats sous une forme ou une autre. Elle utilise aussi toutes les techniques de communication possibles comme les outils bureautiques, raquetteurs et générateurs d'états spécialisés, infrastructure web, télécommunications mobiles... etc.

- **L'administration :**

C'est une fonction transversale qui supervise la bonne exécution de toutes les autres fonctions. Elle pilote le processus de mise à jour des données, la documentation sur les données et sur les métadonnées, la sécurité, les sauvegardes et finalement la gestion des incidents.

## 6) L'utilisation du système décisionnel :

Nous entendons par là, les personnes qui bénéficient et utilisent le système décisionnel, les premiers utilisateurs en entreprise sont les cadres supérieurs qui ont besoin d'informations synthétisées pour suivre l'activité de l'organisation. Cependant, d'autres consommateurs existent: marketing, ressources humaines, et autres directions métiers. En effet on distingue quatre classes de travailleurs, qui consomment le décisionnel selon leurs besoins propres [8] :



**Figure I - 2:** Les utilisateurs du système décisionnel.

### a) Le pilotage stratégique (Top Management) :

La première préoccupation de la direction générale concerne le pilotage de la performance par le suivi des objectifs critiques de l'organisation et la consolidation des indicateurs opérationnels.

Le système décisionnel offre à ses utilisateurs une information synthétisée sous forme d'indicateurs de pilotage présentés via des tableaux de bord.

### b) Le pilotage du système opérationnel (Middle management) :

Le pilotage opérationnel concerne le suivi quotidien des processus opérationnels, il permet l'analyse approfondie des déviations (goulots d'étranglement, événements non attendus... etc.).

L'utilisateur a ainsi la possibilité de « naviguer » et de « zoomer » dans les données afin de pouvoir analyser en profondeur la situation présente.

**c) L'utilisation standard dans les processus métier (Acteurs métiers) :**

Les acteurs métiers qui ont, au quotidien, une activité opérationnelle. Se caractérisent par des besoins de types connus à l'avance, cette spécificité réside dans le fait qu'ils sont des consommateurs d'informations uniquement, qui se contentent généralement de rapports statiques prédéfinis.

**d) Les experts :**

Désigne la catégorie d'utilisateurs finaux avec des métiers spécialisés demande une compétence avancée (finances, marketing, ressources humaines...) ils ont donc des besoins plus complexes et spécifiques à leurs métiers (consolidation budgétaire, segmentation de la clientèle...) aussi, les systèmes décisionnels doivent leur offrir des outils puissants et flexibles pour répondre efficacement à leurs attentes.

## 7) Architecture des systèmes décisionnels :

La définition de l'architecture du système d'aide à la décision est une étape cruciale de tout projet de système décisionnel. Se compose d'outils informatiques et de progiciels assurant le fonctionnement de la chaîne de traitement de l'information. Il propose d'utiliser les données transitant par le Système d'information, données de production le plus souvent, en informations susceptibles d'être exploitées à des fins décisionnelles.

L'architecture d'un système décisionnel, représenté dans la figure I - 3, comprend trois composantes principales [9]:



**Figure I - 3:** Architecture des systèmes décisionnels.

**a) Les sources de données :**

Ce sont majoritairement les données du système dit " de production", il est nécessaire de recueillir et d'intégrer les données stockées dans les diverses sources primaires et secondaires,

qui sont hétérogènes à l'origine. Les sources se composent pour la plupart des données appartenant aux systèmes opérationnels, mais peuvent aussi inclure des documents non structurés, tels que des courriels et des données reçues des fournisseurs externes. De manière générale, un effort majeur est nécessaire pour unifier et intégrer les différentes sources de données.

#### b) Les entrepôts de données et datamarts :

L'entrepôt de données - ou « DataWarehouse » en anglais est le cœur de la solution du système décisionnel, il contient les données provenant de différentes sources. L'utilisation d'outils d'extractions, de transformations et de chargements, connus sous le nom d'« ETL », sera nécessaire pour charger les données dans l'entrepôt de données.

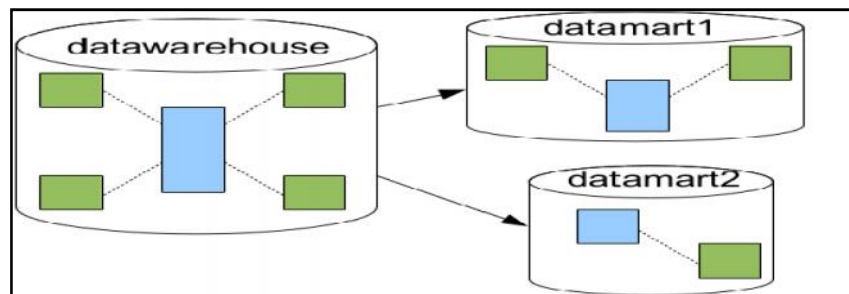


Figure I - 4: Les entrepôts de données et datamarts [10].

#### c) les outils et de restitution :

Ces outils permettent aux utilisateurs d'accéder aux données stockées et analysées par différents axes d'analyse, afin d'obtenir le correspondant qui peut soutenir les résultats de prise de décision.

## 8) Le déploiement et mise en œuvre d'un système décisionnel :

C'est la dernière étape du projet le système décisionnel l'exploitation de ce dernier se fait par le biais d'un ensemble d'outils analytiques développés autour du EDD. Cette étape nécessite l'achèvement du développement, ou de la mise en place de ces outils qui peuvent accomplir les fonctions suivantes :

### 8.1) Les rapports:

Le rapport constitue un outil décisionnel ne sont pas, à proprement parler, des instruments d'aide à la décision, mais, lorsqu'ils sont utilisés de manière appropriée, ils peuvent fournir une précieuse vue d'ensemble.

« il est la présentation périodique de rapports sur les activités et résultats d'une organisation, d'une unité de travail ou du responsable d'une fonction, destinée à en informer

ceux chargés de les superviser en interne ou en externe, ou tout simplement concernés par ces activités ou résultants »[4]. Les rapports sont alors créés par le biais d'outils de rapports qui permettent de leur donner un format prédéterminé. Les requêtes sont constituées lors de l'élaboration des rapports qui seront ensuite diffusés périodiquement en automatique ou ponctuellement à la demande des analystes.



**Figure I - 5:** Exemple d'un état de sortie en rapport[8].

### 8.1.1) Objectifs du rapports:

Le rapport achève l'élaboration d'un système décisionnel, il permet de [4] :

- Fournir une vision opérationnelle de l'activité.
- Assurer le suivi de ses campagnes et valider les business plans.
- Fournir les bilans à communiquer aux actionnaires et au public.
- Dans le cas une gestion décentralisée, il permet de vérifier que les centres de responsabilité respectent leurs engagements contractuels.
- Mettre en évidence de nouvelles opportunités ou des leviers d'amélioration de l'existant.
- Pour les managers, c'est un outil de management opérationnel (il s'agit plutôt de rapports numériques).
- Pour les équipes de production, c'est une aide quotidienne pour leurs activités.
- Pour les commerciaux c'est un outil de suivi de leur activité.

### 8.2) Tableaux de bord :

Généralement chaque tableau de bord est décrit par rapport aux fonctionnalités qu'il offre aux décideurs. Il est évident que toutes ces définitions présentent des aspects communs

tels que l'utilisation de graphes et de tableaux pour la représentation de données, ainsi que la présentation d'une vue globale des données de l'entreprise.

Nous retiendrons deux définitions qui, à notre avis, résument assez bien le concept de "Tableau de Bord".

« Le tableau de bord est un ensemble d'indicateurs peu nombreux conçus pour permettre aux gestionnaires de prendre connaissance de l'état et de l'évolution des systèmes qu'ils pilotent et d'identifier les tendances qui les influenceront sur un horizon cohérent avec la nature de leurs fonctions » [11].

Aussi, on peut dire que : « un tableau de bord est un élément de mesure de la performance nécessaire pour la prise de décision pour les acteurs de l'entreprise »[12].

C'est un outil de pilotage qui donne une vision sur l'évolution d'un processus, afin de permettre aux responsables de mettre en place des actions correctives.

### **8.2.1) Les objectifs des tableaux de bord :**

De nombreuses raisons poussent les décideurs à adopter un système de tableau de bord, on y trouve [13] :

- Disposer d'une vision globale et synthétique de l'activité de l'entreprise à tout instant.
- Automatiser la collecte et le traitement des données, et la mise à jour des indicateurs.
- Comparer la situation de l'entreprise avec les objectifs prédéfinis et mettre éventuellement en place des actions correctives.
- Améliorer la prise de décision grâce à une analyse des données à tous les niveaux.
- Disposer de graphiques clairs, conçus pour mettre en évidence les informations spécifiques.
- Responsabiliser les managers, et favoriser la communication des objectifs, stratégie, résultats et plans d'action.
- Piloter l'entreprise et aider à prendre des décisions.
- Déléguer et responsabiliser ses collaborateurs.

### **8.2.2) Conception d'un tableau de bord :**

La conception d'un bon tableau de bord doit répondre à un certain nombre de normes et caractéristiques qui lui permettent de réaliser sa finalité, celle d'aider son utilisateur à suivre

son activité et optimiser ses actions. Ces qualités peuvent concerner aussi bien le fond que la forme [14].

La conception se fait à travers les étapes suivantes [8] :

- **Identifier les axes de développement liés à la stratégie (ou projet d'entreprise) :**

Les axes doivent être spécifique à l'entreprise, et lié à sa stratégie et son métier. Et Personnalisé en fonction des compétences du dirigeant.

- **Fixer les objectifs :**

**SMART** : o Spécifique. o Mesurable. o Ambitieux. o Réaliste. o dans le Temps.

- **Choisir les indicateurs les plus représentatifs :**

Un indicateur est une information ou un groupement d'informations contribuant à l'appréciation d'une situation par le décideur.

Un indicateur doit être clair, simple, pertinent, partagé et facilement compréhensible.

- **Construire le tableau de bord :**

Dans cette étape il faut adapter le tableau de bord à chaque niveau hiérarchique (opérationnel / stratégique) et Intégrer des graphiques pour un visuel rapide, attractif, et des commentaires.

## 9) Les fouilles de données :

Aussi connue sous les noms data mining en anglais (forage de données) est un ensemble de techniques et de méthodes du domaine des statistiques, des mathématiques et de l'information permettant l'extraction, à partir d'un important volume de données brutes, de connaissances originales auparavant inconnues[8].

« Le fouille de données est l'analyse d'un ensemble d'observations qui a pour but de trouver des relations insoupçonnées et résumer les données d'une nouvelle manière, de façon qu'elles soient plus compréhensibles et utiles pour leurs détenteurs»[15].

## 10) Conclusion :

Dans ce chapitre nous avons introduit le système décisionnel, car il améliore la performance de tous les services de l'entreprise et d'atteindre ses objectifs stratégiques en les aidant à prendre de meilleures décisions en un temps record.

Dans le prochain chapitre on abordera une phase des plus importantes dans tout projet décisionnel qui est la conception et l'implémentation d'entrepôt de données. Ce dernier est considéré comme la brique essentielle du système décisionnel et le point de départ de toutes les applications d'analyse et de système décisionnel.

# Chapitre II

## Chapitre II

Les entrepôts

de données

## 1) Introduction :

Un système décisionnel repose sur la construction d'un entrepôt de données qui héberge de manière organisée des données internes comme externes agrégées et filtrées.

Avec l'apparition de ce dernier, l'information stratégique devient plus facilement et surtout plus rapidement accessible par plusieurs utilisateurs. Ceci permet à l'entreprise d'être plus réactive.

Actuellement rapide dans ce présent chapitre on va essayer de définir ce qu'un entrepôt de données? Comment est-il construit? En quoi diffère-t-il d'une base de données relationnelle? Et bien d'autres questions relatives à ce sujet.

## 2) Définition d'un entrepôt de données (Data Warehouse) :

Les grands concepts de l'informatique décisionnelle ont été formalisés par deux théoriciens américains, Bill Inmon considéré comme le père du Datawarehousing et Ralph Kimball. Ils présentent deux visions sensiblement différentes l'entrepôt de données :

Ralph Kimball publie en 1996 son ouvrage « Entrepôts de données : Guide pratique du concepteur de "Date Warehouse" », dans lequel il définit l'entrepôt de données comme étant :

« L'entrepôt de données est un exemplaire de données relatives à des transactions structurées spécifiquement à des fins de consultation et d'analyse »[16].

L'entrepôt de données est un système qui extrait, nettoie, traite et rend conforme des données sources vers un espace de stockage multidimensionnel. Il permet ainsi la mise en œuvre de l'interrogation et de l'analyse à des fins d'aide à la prise de décision.

Dans son livre «Building the Datawarehouse » Bill Inmon définit un entrepôt de données comme suit : « Une collection de données orientées sujet, intégrées, non volatiles et historique dont le but est d'aider à la prise de décisions managériales. » [17].

Un entrepôt de données est une collection de données orientées sujet, intégrées, non volatiles, et historiées, servant de support à la prise de décision managériale. Les données de l'entrepôt de données sont aptes à être utilisées à de nombreuses fins, y compris des besoins futurs non encore connus.

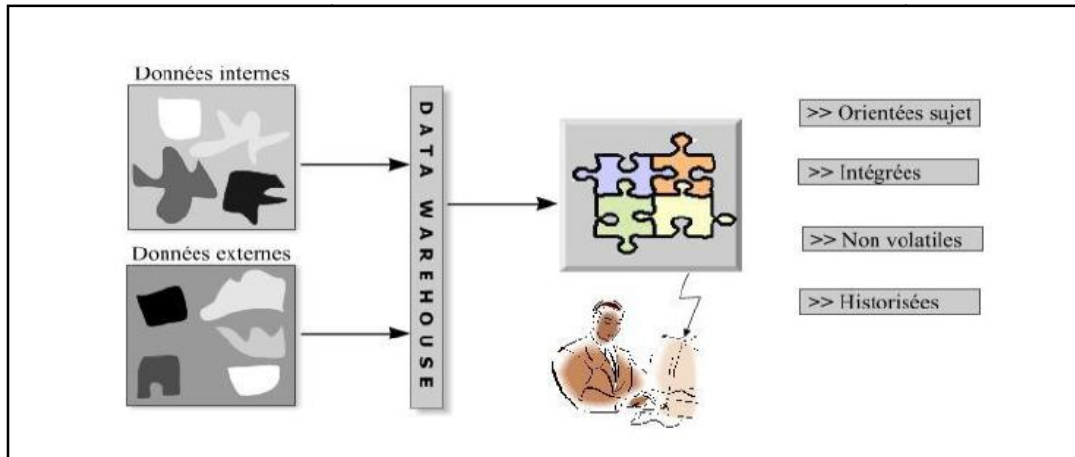


Figure II - 1: Définition de l'entrepôt de données[18].

### 3) Les caractéristiques fondamentales d'entrepôt de données :

Parmi les diverses caractéristiques d'entrepôt de données existe quatre caractéristiques essentielles [17]:

- **Orienté sujet :**

Les données sont classées et stockées par sujet et contrairement aux organisations traditionnelles où les données sont regroupées par processus fonctionnel, la structuration des données ne se fera pas par thèmes[8].

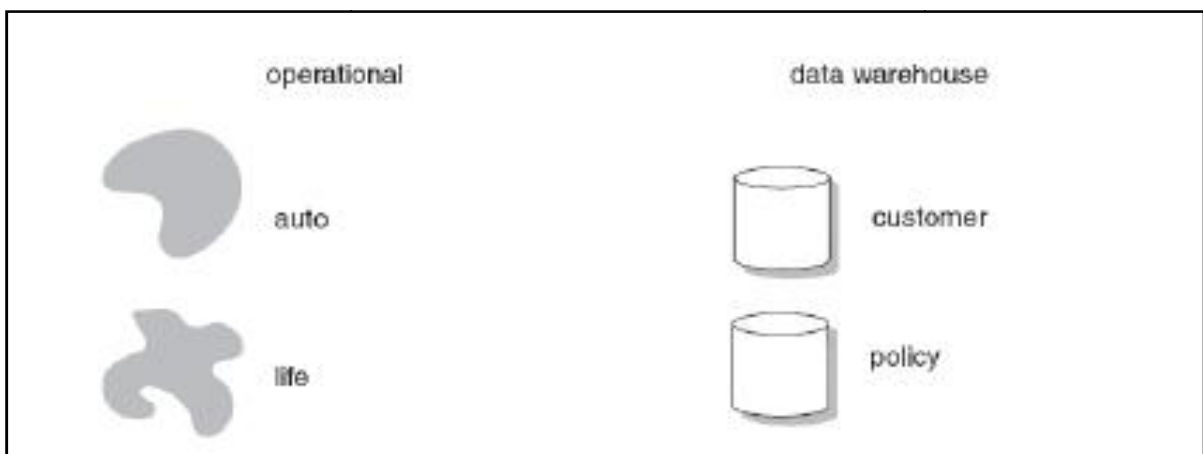
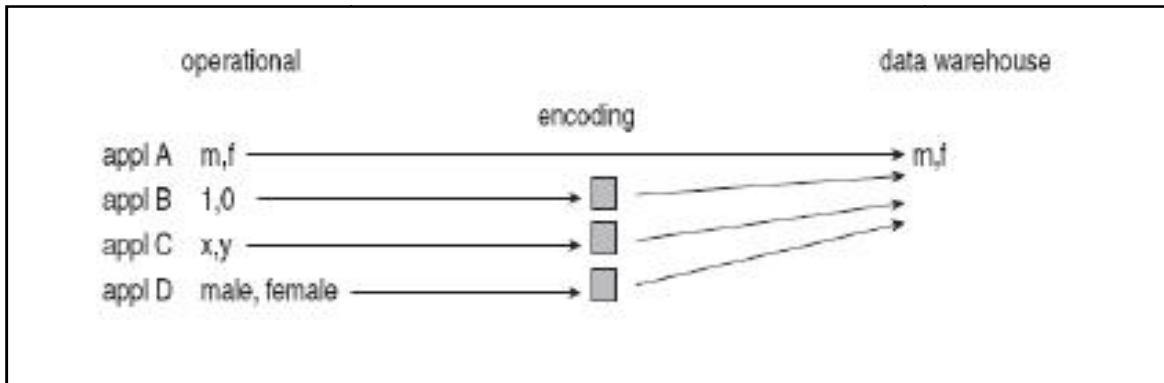


Figure II - 2: Orientation sujet des entrepôts de données[8].

• **Données intégrées :**

Les données qui seront intégrées dans l’entrepôt de données proviendront de différentes sources (Excel, Oracle... etc.) avant que les données ne soient injectées dans l’entrepôt de données, elles sont converties, reformatées, réordonnées et résumées, l’entrepôt de données permet d’intégrer les données et de les homogénéiser. C’est en offrant cette possibilité qu’il est dit intégré.

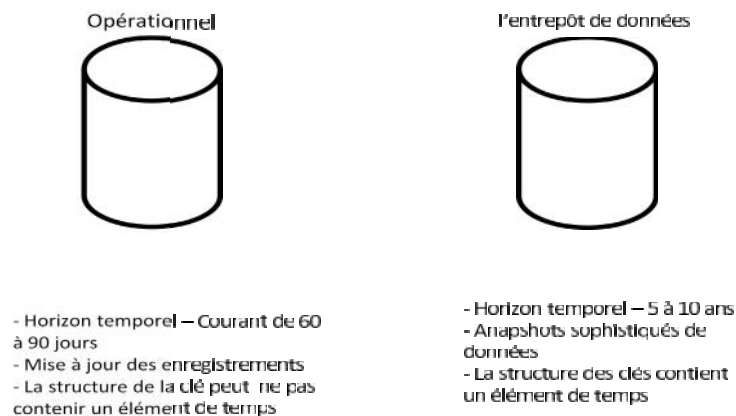


**Figure II - 3:** L’intégration dans les entrepôts de données[8].

• **Historisation :**

« L’entrepôt de données est vu comme une série temporelle où les données sont stockées sous forme de couches successives »[20].

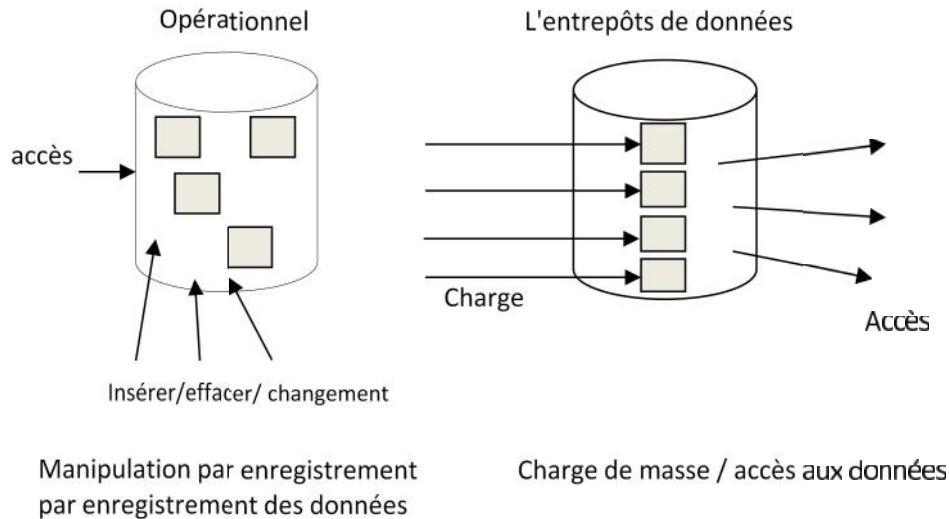
Cela signifie que les enregistrements dans l’entrepôt de données sont effectués de manière historisée, chaque élément de l’entrepôt est associé à la dimension temporelle. Cette propriété permet, entre autres, la prédiction par l’observation de la situation passée de l’organisation.



**Figure II - 4:** L’historisation des entrepôts de données[8].

- **Non volatiles :**

Non volatile pour être stable, en lecture seule, non modifiable afin de conserver la traçabilité des informations et des décisions prises [20].



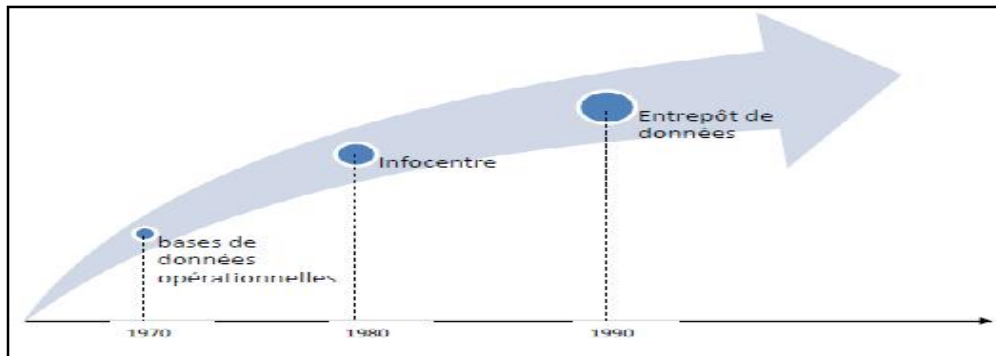
**Figure II - 5:** Non volatile des entrepôts de données [22].

#### 4) Historique des entrepôts de données :

Le concept de "entrepôt de données" a été apparu Pendant les années 80 (Data Warehouse en anglais) durant lesquelles un intérêt croissant au système décisionnel a vu le jour, dû essentiellement à l'émergence des SGBD relationnel et la simplicité du modèle relationnel et la puissance offerte par le langage SQL.

Au début, l'entrepôt de données n'était rien d'autre qu'une copie des données du système opérationnel prise de façon périodique, dédié à un environnement de support à la prise de décision. Ainsi, les données étaient extraites du système opérationnel, stockées dans une nouvelle base de données «concept d'infocentre », pour objectif de répondre aux besoins des décideurs sans altérer le bon fonctionnement du système opérationnel.

L'entrepôt de données est devenu alors une nouvelle source d'information, alimentée avec des données recueillies et consolidées des différentes sources internes et externes.



**Figure II - 6:** Evolution des bases de données décisionnelles[19].

## 5) Objectifs d'entrepôt de données :

Les informations possédées par une entreprise constituent des avantages déterminants. Ces informations se présentent généralement sous deux formes : les systèmes opérationnels qui enregistrent les données et l'entrepôt de données. Etant donné les besoins, les clients, les structures et les rythmes d'entrepôt de données sont profondément différents de ceux des systèmes opérationnels. L'entrepôt de données doit [22] :

- **Rendre l'information d'une organisation facilement accessible :**

Le contenu de l'entrepôt doit être compris d'une façon facile, l'utilisateur final doit pouvoir y naviguer facilement et avec rapidité. En effet les données sont considérées comme compréhensibles lorsqu'elles sont pertinentes et clairement définies. Ils doivent également retourner les résultats de la requête de l'utilisateur dans un temps d'attente minimal [4].

- **Rendre les informations d'une entreprise cohérentes :**

Les données de l'entrepôt doivent être crédibles, soigneusement assemblées à partir d'une variété de sources, nettoyées, et libérées afin d'être prêtes à la consommation. Les informations provenant d'une branche de l'entreprise peuvent être mises en corrélation avec celles d'une autre branche. Si deux unités de mesure portent le même nom, elles doivent alors signifier la même chose. A l'inverse, deux unités ne signifiant pas la même chose doivent être définies différemment. Une information cohérente est supposée comme une information de grande qualité [14].

- **Constituer une source d'information souple et adaptable :**

Nous avons simplement essayé de ne pas éviter le changement, ce dernier est dû à l'évolution des besoins des utilisateurs, aux changements des conditions commerciales

et l'évolution technologique (L'entrepôt de données est conçu dans la perspective de modifications perpétuelles). L'arrivée de questions nouvelles ne doit pas bouleverser les données et les Applications existantes ni les technologies [14].

- **présenter un bastion sécurisé qui protège la capitale information :**

L'entrepôt de données ne contrôle pas seulement l'accès aux données, mais il offre à ses gestionnaires une bonne visibilité des utilisations et des données même après que celles-ci ont quitté l'entrepôt. Les informations Les plus précieuses d'une organisation sont conservées dans l'entrepôt de données, il doit donc efficacement contrôler l'accès à ces données confidentielles[18].

- **Constituer la base décisionnelle de l'entreprise :**

L'entrepôt de données doit avoir les bonnes données pour appuyer la prise de décision et il n'y a qu'un seul véritable résultat concret attendu, ce sont les décisions prises grâce aux données obtenues.

## 6) La structure de L'entrepôt de données :

L'entrepôt de données est structuré à deux axes, un axe historique et un axe synthétique, chaque axe est structuré à deux classes de données. La figure suivante illustre cette structuration et positionne les classes les unes par rapport aux autres dans un cadre d'architecture de données [23] :

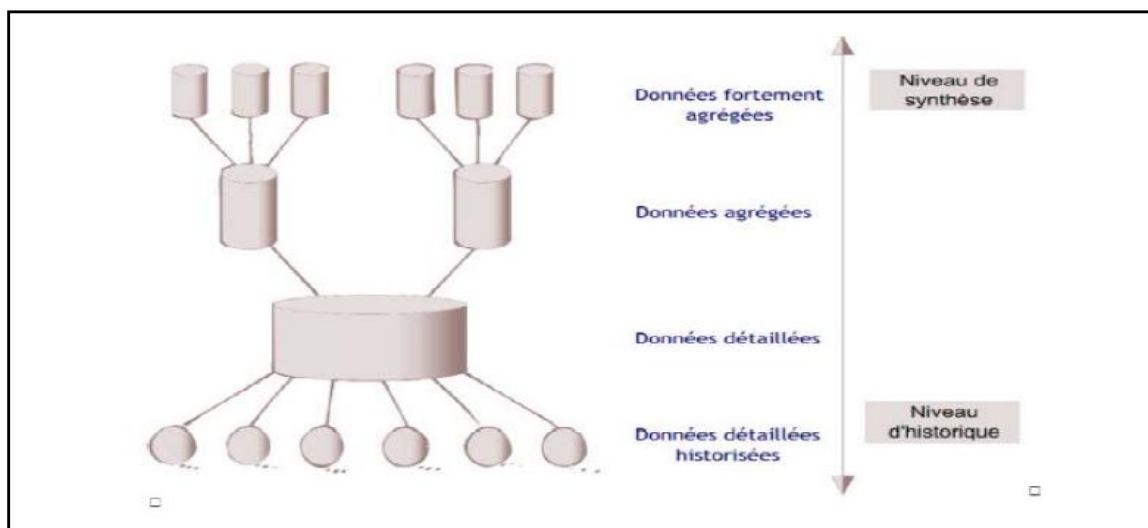


Figure II - 7: Structure d'un entrepôt de données[22].

### 6.1) Axe historique :

Les volumes à traiter et à manipuler dans cet axe sont plus importants que ceux qui sont gérés en transactionnel. Le niveau de détail géré dans l'entrepôt de données n'est pas forcément identique au niveau de détail géré dans les systèmes opérationnels. On trouve ici deux types de données : les données détaillées et les données détaillées historiées.

- **Données détaillées :**

ce sont les données qui reflètent les événements les plus récents, fréquemment consultées, La gestion et le traitement de ces données nécessitent parfois des machines sophistiquées, car leur volume peut être très important.

- **Les données détaillées historique :**

Elles reflètent les transactions archivées et détaillées de l'entreprise. L'entrepôt de données doit conserver l'historique des données et l'insertion de nouvelles valeurs ne doit jamais annuler l'ancien existant.

### 6.2) Axe synthétique :

Elles correspondent à des éléments d'analyse représentant les besoins des utilisateurs. Elles constituent déjà un résultat d'analyse et une synthèse de l'information contenue dans le système décisionnel. Elles doivent être accessibles et compréhensibles. On trouve ici deux types de données : les données agrégées et les données fortement agrégées.

- **Données agrégées :**

Données agrégées à partir des données détaillées. Ces données sont très souvent utilisées, elles constituent déjà un résultat d'analyse et de synthèse de l'information contenue dans le système décisionnel.

- **Les données fortement agrégées :**

Elles sont presque comme les données agrégées, elles ont une incidence énorme sur les performances, dans certains cas elles permettent d'exécuter les requêtes 100 fois, voir 1000 fois plus rapidement.

## 7) Composants de base et architecture d'un entrepôt de données :

Pour créer un architecte d'un entrepôt de données, il faut connaître ses composants et leurs rôles dans ce dernier. Ces composants sont au nombre de quatre :

système source, zone de préparation de données, zone de présentation des données et portail de restitution de données [24] :

### 7.1) Système source :

L'entrepôt de données regroupe généralement les données à partir de plusieurs sources. Les sources sont externes au l'entrepôt de données. L'objectif est regrouper les données nécessaires à la prise de décisions sans affecter le fonctionnement des systèmes sources. Ces sources peuvent être :

- **Sources hétérogènes :**

Document et fichier non structuré comme les documents Word, fichiers plats et feuilles de calcul[8].

- **Bases de données production :**

Appeler aussi (base de données transactionnelle) qui peuvent être nombreuses résident sur des plateformes différentes, il s'agit du portefeuille des applications « métier » de l'entreprise[25].

- **Sources externes :**

Des fichiers peuvent être acquis auprès d'autres d'entreprise spécialisées dans la constitution et la revente de données[8].

- **Données archivées :**

Les systèmes opérationnels sont destinés à la gestion actuelle de l'organisation, ils sont régulièrement purgés des données les plus anciennement stockées vers des fichiers et des documents d'archivage [25].

- **Données internes :**

Représentent les divers documents et des fichiers que les employés d'une organisation gardent en leur possession. Feuilles de calcul, profils clients et bases de données départementales. Ces données peuvent être extraites des systèmes de production mais une grande partie d'entre-elles est contenue dans des fichiers privés au sein de l'organisation[25].

### 7.2) La zone de préparation de données :

C'est à ce niveau que se fait le traitement des données brutes et chargées à partir du système source. Elle est constituée d'un ensemble des processus qui nettoient, transforment, combinent, archivent, suppriment les doublons, c'est-à-dire préparent les données sources en vue de leur intégration puis de leur exploitation au sein d'entrepôt de données.

Aussi appelé « staging area », il s'agit de la zone comprise entre les systèmes sources et les serveurs de présentation de données. Cette zone n'offre aucun service de présentation de données ou de requête.

Le traitement de la qualité des données implique plusieurs étapes distinctes : vérification des valeurs valides, cohérences entre les valeurs, vérification des doublons. L'objectif final de la zone de préparation des données est l'obtention de données prêtes à être chargées sur un serveur de prestation (moteur OLAP ou SGBDR).

### **7.3) La zone de présentation de données :**

C'est tout ce que l'utilisateur peut manipuler par le biais des outils d'accès. Elle est constituée l'entrepôt de données ou d'une série de Datamarts.

### **7.4) Le portail de restitution de données :**

Ce sont les clients de l'entrepôt de données. Ils permettent de manipuler les données suivant des axes d'analyses. Il existe plusieurs outils d'accès aux données d'un entrepôt de données tels que :

- Des outils d'interrogation ad hoc. (pour faciliter l'accès aux données en fournissant une interface conviviale au langage de requêtes).
- Des outils de rapport.
- Des outils de fouille de données (pour découvrir des liens sémantiques entre les données).
- Des applications.
- Des outils d'analyse en ligne OLAP (pour la synthèse et l'analyse des données multidimensionnelles).

Les outils cités en ci-dessus sont exploités par les utilisateurs pour accéder à la zone de présentation d'un entrepôt de données. Ces outils doivent être faciles d'utilisation pour les non informaticiens. Il est impératif que les données d'un entrepôt de données soient présentées à l'utilisateur de manière significative et propre à son activité, afin qu'elles puissent être interprétées facilement.

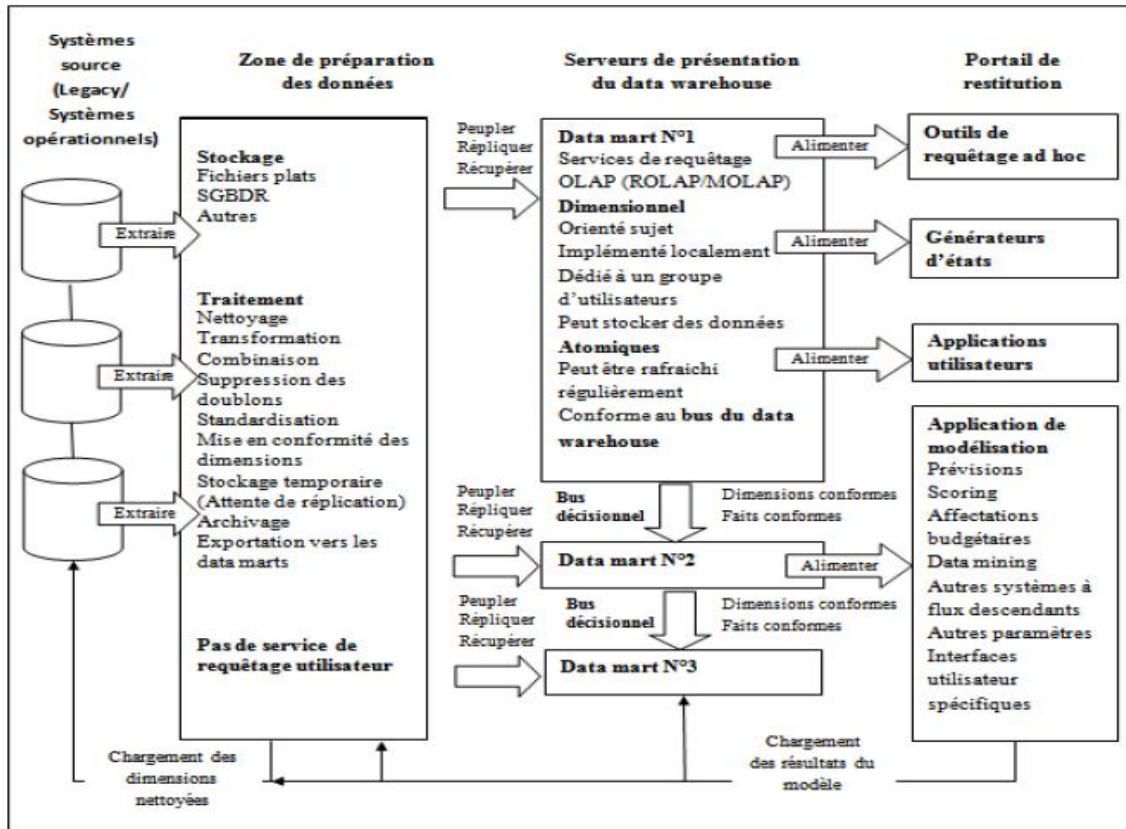


Figure II - 8: Composants de base et architecture d'un entrepôt de données[26].

## 8) Méthodologie de Construction d'un entrepôt de données :

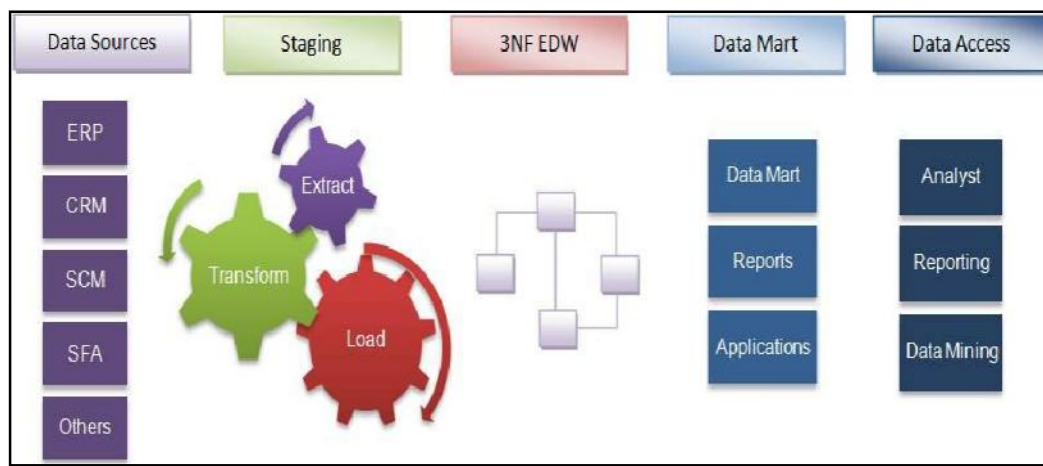
Il y a trois méthodes communément utilisées pour mettre en place un entrepôt de données (Data Warehouse), haut en bas (Top-Down) et bas en haut (Bottom-up) et hybride (Middle-out) relative aux deux théoriciens de l'informatique décisionnel Inmon et Kimball.

### 8.1) Approche de haut en bas:

Plus connue sous le nom «Top-down», C'est une approche descendante qui consiste à créer un entrepôt de donnée centrale, normalisé (3FN) qui regroupe des données atomiques (détaillées) provenant des différents systèmes de production de l'entreprise. Une fois que l'entrepôt de donnée est conçue, on commence à créer les DataMarts, chaque Datamart contient les informations nécessaires pour répondre aux besoins analytiques d'un métier donné.

« Vous pouvez attraper tous les vairons de l'océan, les regrouper ensemble et ils ne pourront tout de même pas faire une baleine » [17].

Cette approche consiste à construire le schéma de l'entrepôt de données qui a été proposé par Inmon [27], qui démarre du fait que les besoins des utilisateurs sont en perpétuelle évolution ce qui rend inutile de donner une solution basé sur les besoins, il préfère donc se baser sur les sources de données et d'essayer de centralisé toutes les données venant des différentes sources dans un entrepôt de donnée ou ils seront historiées et intégrées. Une fois que l'entrepôt est conçu, on passe vers une autre étape qui est la compréhension des besoins d'utilisateurs puis leurs livrer les données qu'ils ont besoins sous forme des DataMarts, la figure suivante illustre cette démarche.



**Figure II - 9:** Architecture du DW selon Inmon[8].

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.

## 8.2) Approche de bas en haut :

Plus connue sous le nom «Bottom-up». C'est une approche ascendante qui consiste à créer des DataMarts avec des faits et des dimensions conformes qui vont faire l'objet de l'entrepôt de données.

« [...] L'entrepôt de données n'est rien d'autre que l'union de tous les Datamarts.» [26].

Cette approche consiste à construire le schéma de l'entrepôt de données qui a été proposé par R. Kimball [16], qui démarre du fait que l'objectif principale d'un entrepôt de données est de répondre aux besoins analytiques des utilisateurs, pour cela il faut passer par une phase de collecte de besoins afin de bien les définir, Kimball suppose que le fait d'avoir des DataMarts, avec des faits et des dimensions conformes,

permet de faire introduire de nouvelles données dont le but de répondre à d'autres besoins non prévus au départ.

L'avantage de cette méthode est qu'elle est simple à réaliser (une étoile à la fois) par contre 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).

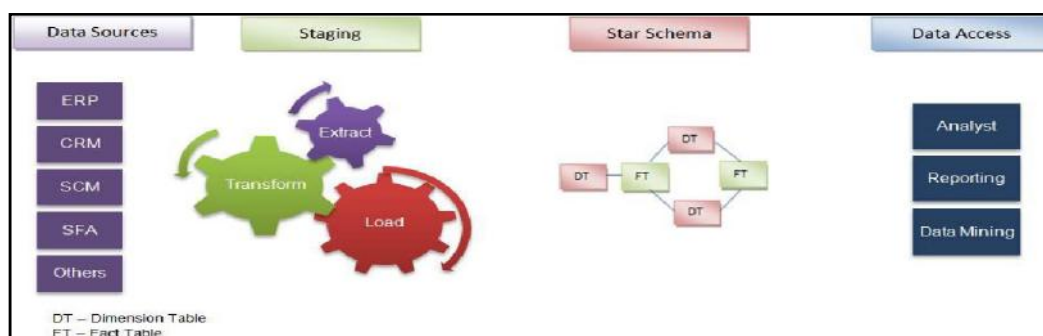


Figure II - 10: Architecture du DW selon Kimball[8].

### 8.3) Approche hybride (Middle-Out) :

C'est une combinaison entre les deux autres approches afin de profiter des avantages de chacune. L'idée dans cette approche réside dans le fait de construire des schémas candidats à partir des données. Ainsi, le schéma construit constitue une réponse aux besoins réels d'analyse et il est également possible de le mettre en œuvre avec les sources de données [28].

### 8.4) Approche de haut en bas vs Approche de bas en haut :

Haut en bas Bill Inmon	Bas en haut Ralph Kimball
Commence par concevoir un modèle de DW au niveau de l'entreprise.	Commence par concevoir un modèle dimensionnel pour le data mart.
Déploies une architecture multi tiers composée de staging area, le DW, et les data mart dépendants.	Utilise une architecture qui consiste en un staging area et les data marts.
Le DW est orienté entreprise; les data marts sont orientés processus.	Les data marts peuvent fournir une vue entreprise ou processus
Les utilisateurs peuvent effectuer des requêtes sur le DW et les data marts.	Les utilisateurs ne peuvent effectuer des requêtes sur le staging area.

Tableau II - 1: Tableau comparatif entre les approches de mise en place d'un entrepôt de données selon Inmon et Kimball [4].

## 9) Alimentation d'entrepôt de données :

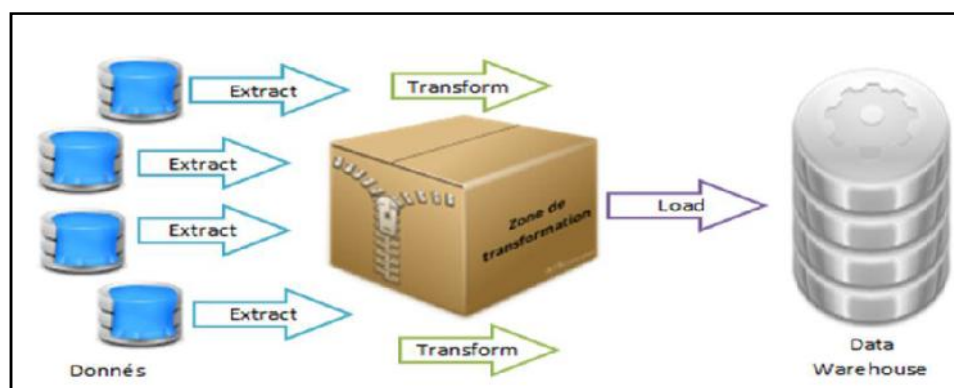
Pour alimenter l'entrepôt de données d'une façon régulière, les données d'un ou de plusieurs systèmes opérationnels doivent être extraites et copiées dans l'entrepôt de données, d'où la nécessité d'établir des flux de données entre l'entrepôt de données et les bases de données spécialisées où elles ont été créées (les bases de production). Cette alimentation le plus souvent appelée processus ETL « **Extract-Transform-Load** » se déroule en 3 phases qui sont :

- Extraction des données (issues par exemple des systèmes de production).
- Transformation des données.
- Chargement des données.

Ces étapes décrivent une mécanique cyclique qui a pour but de garantir l'alimentation d'entrepôt de données en données homogènes, propres et fiables [28] :

### 9.1) Processus Extraction – Transformation - Chargement:

« **Extract-Transform-Load** » est connu sous l'acronyme ETL (ou parfois : Datapumping). Il s'agit d'une technologie informatique middleware permettant d'effectuer des synchronisations massives d'information d'une banque de données vers une autre. Les données sont extraites de différentes sources de données, après elles vont être transformées en une structure plus appropriée pour l'analyse et les rapports enfin, elles sont chargées dans la base de données.



**Figure II - 11:** Le système d'alimentation ETL [28].

#### 9.1.1) L'Extraction des données :

« L'extraction est la première étape du processus d'apport de données à l'entrepôt de données, extraire, cela veut dire lire et interpréter les données sources et les copier dans la zone de préparation en vue de manipulations antérieures » [26].

L'extraction des données est la première des processus ETL. Le but de cette étape, est comme son nom l'indique : la lecture et l'extraction des données de différentes sources internes et externes, structurées et / ou non structurées. Il existe trois types d'extraction de données qui sont [28] :

• **L'extraction complète :**

C'est la capture de l'ensemble des données à un instant donné (snapshot de l'état opérationnel) ce type d'extraction est appliqué lors de l'alimentation initiale de l'entrepôt ou bien pour un rafraichissement complet des données. Ce type d'extraction prend beaucoup de temps surtout dans le cas de gros volumes de données cela peut prendre une journée jusqu'à une semaine.

• **L'extraction incrémentale :**

C'est le type d'extraction le plus appliqué, il est basé sur une métadonnée date du dernier chargement, qui est fournis par les systèmes sources. Ce type d'extraction évite d'extraire les données déjà chargées dans l'entrepôt, il est nécessaire d'appliquer ce type d'extraction dans le cas d'une alimentation journalière de l'entrepôt de données, afin de traiter que les données nouvelles ou celles qui ont été modifiées.

• **L'extraction sous évènement transactionnel :**

Elle consiste à détecter en temps réel les enregistrements modifiés au niveau des systèmes source, cette détection est faite par consultation du fichier journal des transactions (fichier log) ce type d'extraction est le moins utilisé par rapport au deux types d'extraction précédentes.

### **9.1.2) Transformation des données :**

La transformation est la seconde phase du processus ETL. Il s'agit de filtrer et trier les données disponibles dans le staging en utilisant des attributs spécifiques et faire des calculs. C'est la tâche la plus complexe et qui demande beaucoup de réflexion.

Pour garantir la fiabilité des données et leurs qualités. Voici les grandes fonctionnalités de transformation [12] :

- Nettoyage des données.
- Standardisation des données.
- Conformité des données.
- Gestion des tables de faits.
- Gestion des dimensions.

- Affectation des clés de substitution (surrogate key).
- Gestion de l'évolution lente (Slowly changing dimension).
- Gestion des faits arrivants en retard (Late arriving fact).
- Gestion des lookups.

Cette opération se solde par la production d'informations dignes d'intérêt pour l'entreprise et de et sont donc prêtes à être entreposées.

### **9.1.3) Le chargement des données :**

C'est la dernière phase de l'alimentation d'un entrepôt de données. Le processus de chargement permet de transférer les données extraites et transformées vers leur destination finale et spécifiée (ODS, Data Warehouse ou DataMart). Donc il est indispensable de maîtriser la structure du Système de Gestion de la Base de Données (SGBD) pour optimiser au mieux le processus [4]. Il existe 3 cas de figures se présentent, selon l'architecture mise en place [12] :

#### **a) Chargement des données dans l'entrepôt de données :**

Si la politique retenue a été de construire un entrepôt de données avec une base de données, les données seront chargées dans l'entrepôt. Cette approche est proche à celle de Bill Inmon. Il sera dès lors possible d'utiliser des fonctionnalités analytiques comme Oracle le permet.

#### **b) Chargement des données dans les cubes de données :**

Les données seront directement chargées dans des cubes de données sans les stocker dans l'entrepôt de données. Cette approche est certainement la plus proche à celle de Ralph Kimball.

#### **c) Le mode hybride :**

Ce cas de figure est celui qui offre le plus d'avantages mais demande par contre plus d'effort. Le chargement des données s'effectue à la fois sur l'entrepôt de données et les Datamarts :

- Un premier chargement des données dans un entrepôt de données.
- Un deuxième chargement dans des cubes de données.

## **10) Modélisation dimensionnelle :**

La modélisation dimensionnelle est une méthode de conception logique qui a comme but de présenter les données sous une forme standardisée et intuitive, qui permet des accès hautement performants. Elle repose sur deux concepts fondamentaux

de faits et de dimensions. Ce mode de stockage est prévu pour l'analyse statistique des données [28].

### 10.1) Table de fait :

Le terme « fait » représente d'un sujet d'analyse et regroupe un ensemble d'indicateurs appelés mesures.

La table de fait sert à stocker les mesures de l'activité. Une mesure est une valeur numérique, additives qui associe un fait à un axe d'analyse appelé dimension. Une ligne d'une table fait correspond à une mesure.

Une table de faits assure les liens plusieurs à plusieurs entre les dimensions. Elle sont composées d'un moins de deux clés étrangères, elles se connectent aux clés primaires des tables de dimensions [28].

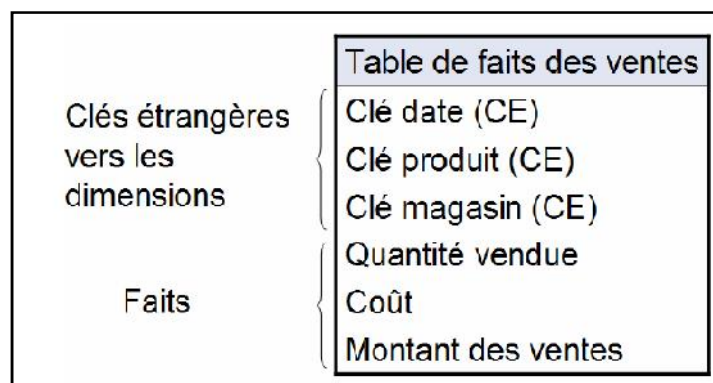


Figure II - 12: Tableau de fait.

### 10.2) Table des dimensions :

Les tables de dimensions sont des tables qui sont accompagnées obligatoirement d'une table de faits, elles contiennent les descriptions textuelles des dimensions de l'activité. Une table de dimension est constituée de nombreuses colonnes qui décrivent une ligne. C'est grâce à cette table que l'entrepôt de données est compréhensible et utilisable, elles permettent des analyses en tranches et en dés. Une dimension est généralement constituée d'une clé artificielle, d'une clé naturelle et des attributs [28].

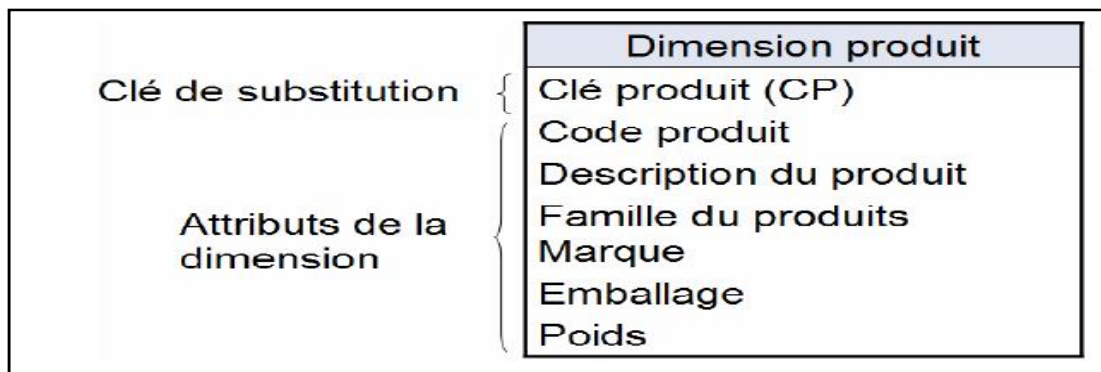


Figure II - 13: Tableau dimension.

## 11) Différents modèles de la modélisation dimensionnelle :

### 11.1) Schéma en étoile :

Appelé schéma en étoile car le diagramme qui le représente ressemble à une étoile, Constitué d'une table de faits centrale reliée à des dimensions. Ces dernières n'ont pas de liaisons entre elles.

Le schéma en étoile présente un nombre de jointures limité qui facilite énormément la navigation [13].

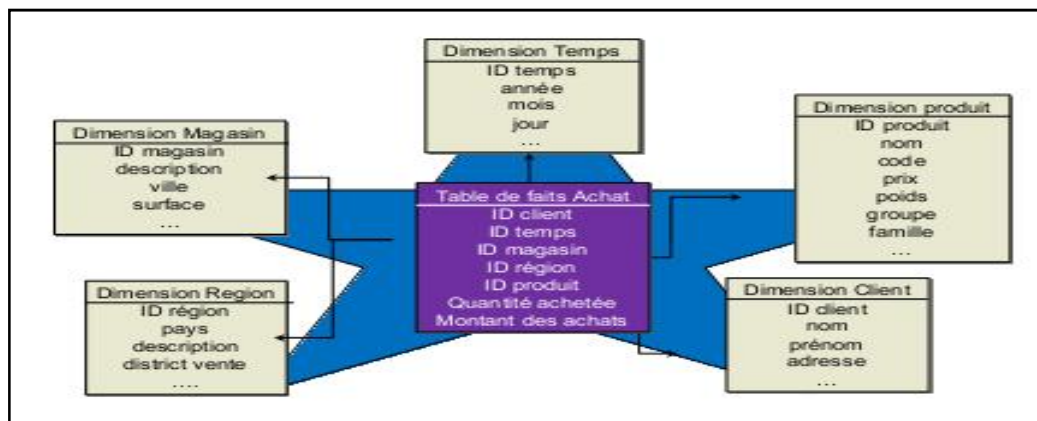


Figure II - 14: Exemple d'un schéma conceptuel en étoile[8].

### 11.2) Schéma en flocon :

Le schéma en flocons de neige est une variante du schéma en étoile avec une normalisation des dimensions. Cette normalisation permet des économies en espace disque mais rend le modèle plus complexe et les requêtes moins performants [13].

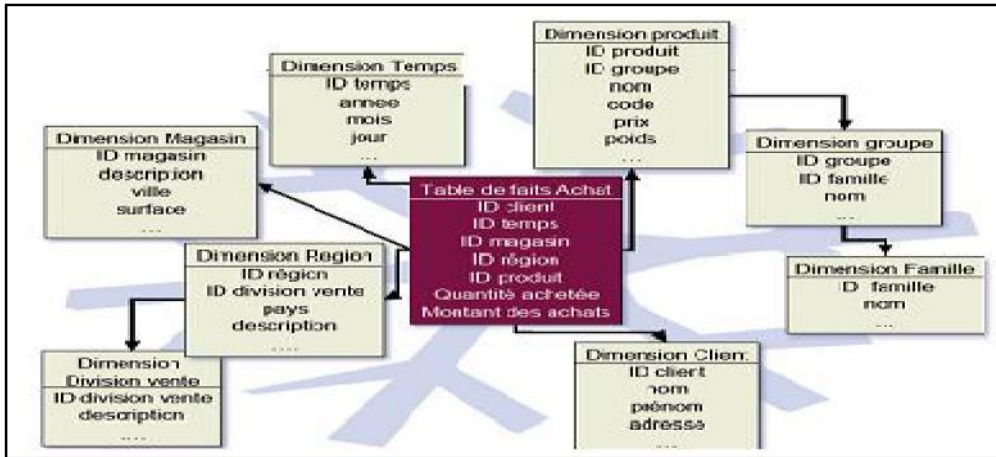


Figure II - 15: Exemple d'un schéma de données en flocon[8].

### 11.3) Schéma en constellation :

Le schéma en constellation est un ensemble de modèles en étoiles reliés entre eux par des dimensions communes.

### 12) Le cube de données :

Le cube est une modélisation d'une base de données multidimensionnelle. Il permet aux utilisateurs de l'entrepôt de données de pouvoir trouver, extraire et évaluer les données dont ils ont besoin sans l'aide des informaticiens. Cette méthode permet aux analystes et décideurs d'avoir les informations dont ils ont besoin. L'utilisateur choisit deux ou trois critères à visualiser sous forme de tableau ou de cube. Il peut également faire pivoter les axes d'analyse pour projeter les informations sous un angle différent [29].

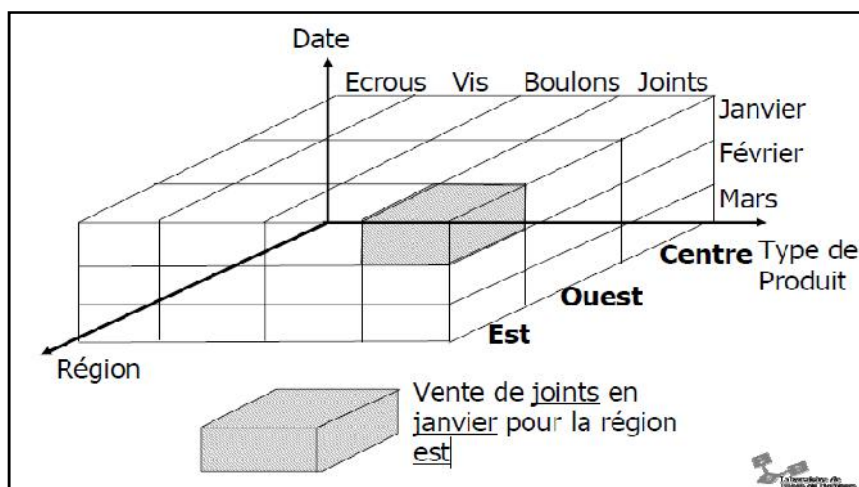


Figure II - 16: Exemple d'un cube de données [30].

### 13) Le concept OLAP (On Line Analytical Process) :

Ralph Kimball a défini OLAP comme suit : « [ ...] présentation de données textuelles et numériques contenues dans l'entrepôt de données, style d'interrogation spécifiquement dimensionnel, qui se concrétise par un certain nombre de (vendeurs OLAP) » [26].

La technologie OLAP est non relationnelle, elle est presque toujours basée sur un cube de données multidimensionnel tel qu'il permet de restructurer et stocker les données dans un format Multidimensionnel explicites. Les bases de données OLAP sont également connues sous le terme de base de données multidimensionnelle. Il existe plusieurs versions d'OLAP, Nous en citons les plus utilisés :

- MOLAP « Multidimensionnel OLAP ».
- ROLAP « Relational OLAP ».
- HOLAP «Hybrid OLAP ».

#### 13.1) M-OLAP (OLAP Multidimensionnelle) :

« Est un ensemble d'interfaces d'applications et de technologies de bases de données propriétaires dont l'aspect dimensionnel est prépondérant » [24].

Le Multidimensionnel OLAP désigne l'une des premières technologies multidimensionnelles.

Et elle s'appuie sur une base de données multidimensionnelle. L'intérêt est que les temps d'accès sont optimisés et ne supporte pas de très gros volumes de données.

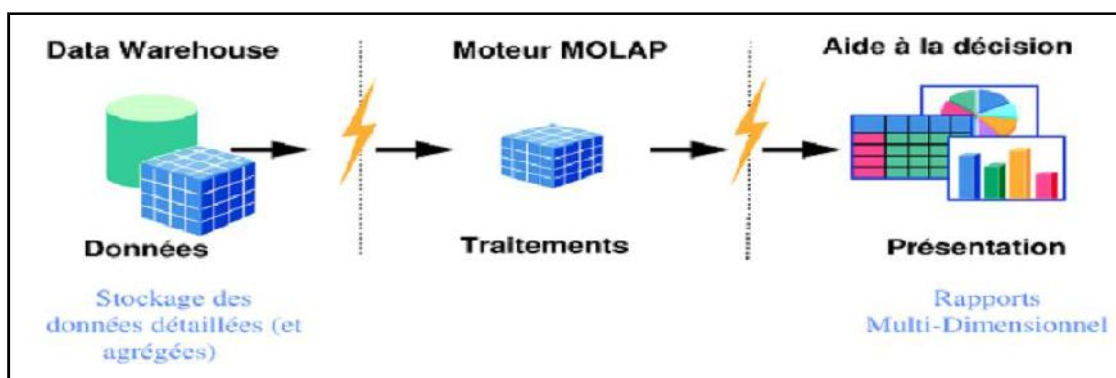


Figure II - 17: Schématisation de la technique M-OLAP [28].

### 13.2) R-OLAP (OLAP Relationnel) :

« le ROLAP est un ensemble d'interfaces utilisateurs et d'applications qui donnent une vision dimensionnelle à des bases de données relationnelles » [26].

Les données détaillées de base ainsi que les données agrégées d'entrepôt de données sont stockées sous forme de tables dans une base de données relationnelle, celle-ci doit être structurée selon un modèle particulier (étoile, flocon, ...).

Contrairement à la technique précédente, la technique R-OLAP est facile à mettre en œuvre, et elle permet de stocker de gros volumes de données. Mais aussi, elle est moins performante lors des phases de calcul.

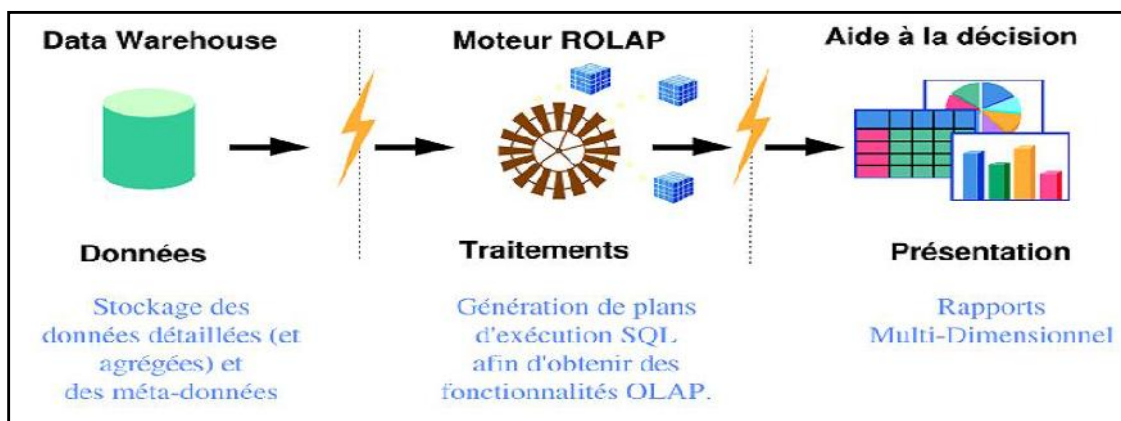


Figure II - 18: Schématisation de la technique R-OLAP [28].

### 13.3) H-OLAP (OLAP Hybride) :

Cette approche est une sorte de compromis entre les deux approches précédentes, Elle représente un bon compromis entre coûts et performance. Son principe est de [13]:

- Stocker les données de base (détaillées) dans une base de données relationnelle.
- Stocker les données agrégées dans un cube (BDD multidimensionnelle).

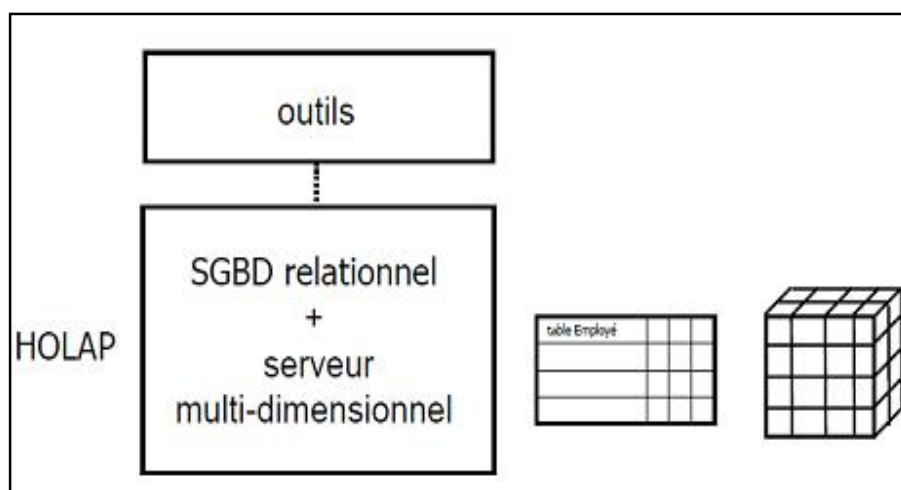


Figure II - 19: Schématisation de la technique H-OLAP.

### 14) Comparaison entre MOLAP, ROLAP et HOLAP :

	MOLAP	ROLAP	HOLAP
Stockage de données de base (détaillées)	BD multidimensionnelle	BD relationnelle	BD relationnelle
Stockage des aggregations	BD multidimensionnelle	BD relationnelle	BD multidimensionnelle
Performance requêtes	Elevée	Peu performante	Moyenne

Tableau II - 2 : Tableau comparatif entre R-OLAP, M-OLAP et H-OLAP[8].

### 15) La navigation dans les données :

Les opérations de navigation dans les données sont [8] :

- **Opérations ensemblistes** : le Tranchage (Slicing) & le Découpage (Dicing) (extraction).
- **Opérations de granularité** : le forage vers le haut (Roll-up) & le forage vers le bas (Drill-down).

### 15.1) Les opérations ensemblistes «Le Tranchage (Slice) & Le Découpage (Dice) » :

#### 15.1.1) Le Tranchage (Slicing) :

Il consiste à faire une sélection de tranches du cube selon des prédicats et selon une dimension « filtrer une dimension selon une valeur »[31].

Cette opération consiste à filtrer une dimension selon une valeur ou une plage de valeurs afin de se concentrer sur une partie de données.

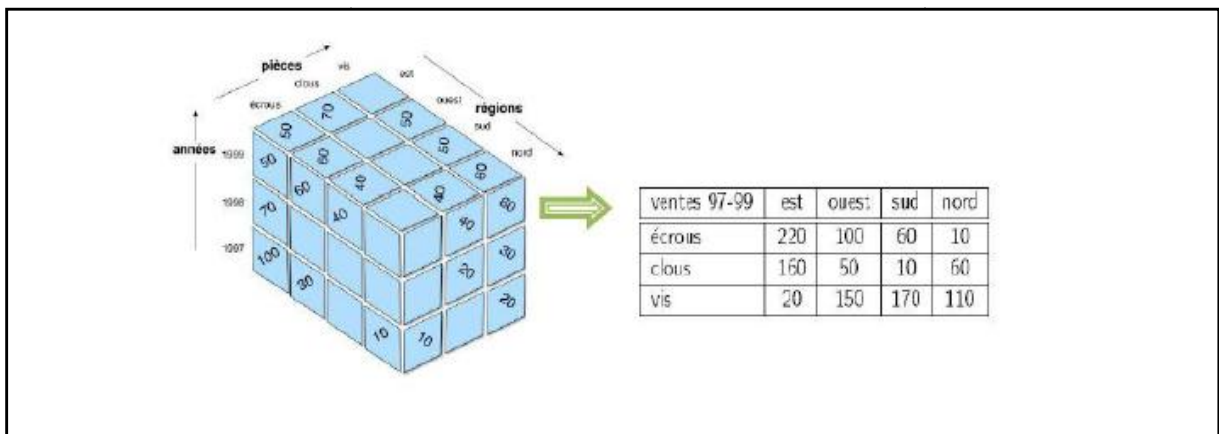


Figure II - 20: Exemple du le Tranchage [8].

#### 15.1.2) Le Découpage ( Dicing) :

Quant à lui, peut être vu comme étant une extraction d'un sous cube.

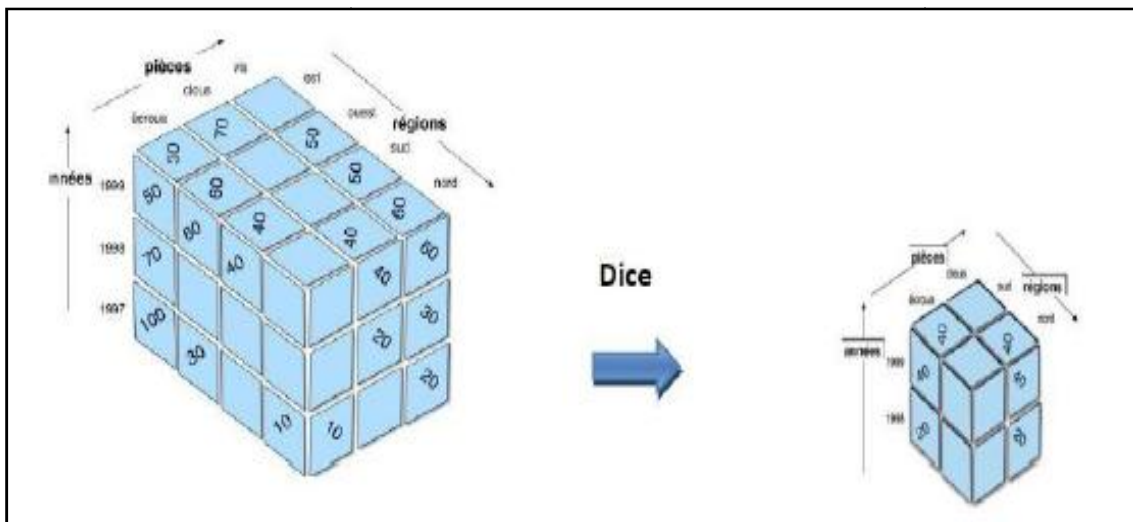
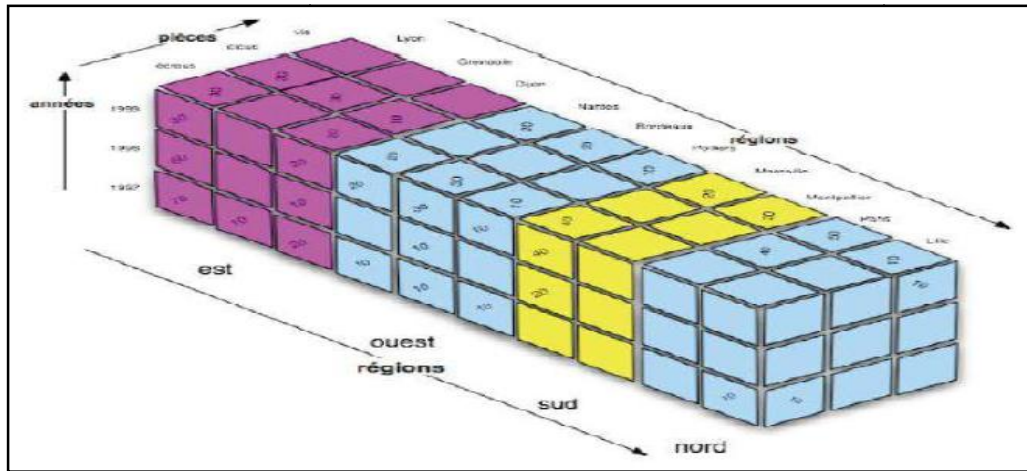


Figure II - 21: Exemple Le Découpage[8].

**15.2.1) Le forage vers le haut :**

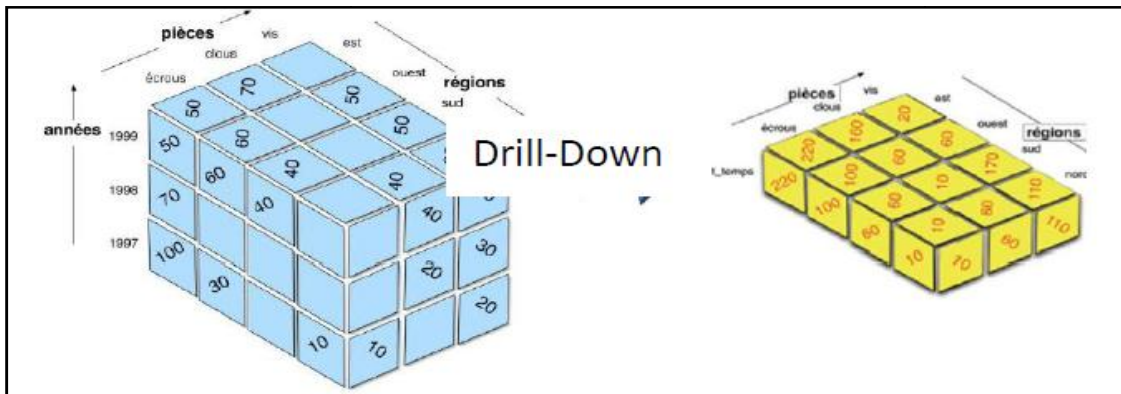
Cette méthode appelée aussi (Roll-up) elle consiste à représenter les données du cube au niveau de granularité supérieur conformément à la hiérarchie définie sur la dimension.



**Figure II - 22:** Exemple Le Forage vers le haut.

**15.2.2) Le forage vers le bas :**

Cette méthode appelée aussi (Drill-down) elle consiste à représenter les données du cube au niveau de granularité de niveau inférieur, donc sous une forme plus détaillée (selon la hiérarchie définie de la dimension).



**Figure II - 23:** Exemple Le forage vers le bas.

## **16) Conclusion :**

Dans ce chapitre, nous avons parlé de quelques généralités sur des entrepôts de données, leurs objectifs, leurs composantes, nous avons aussi présenté quelques architectures d'entrepôt de données. Le choix entre ces différentes architectures dépend des besoins et contraintes posés par l'entreprise et nous avons défini tous les différentes étapes de la mise en place d'un entrepôt de données et l'importance de chaque étape. Au cours de la deuxième partie, nous allons présenter l'organisme d'accueil ainsi les différents étapes de construction de notre système.

Partie II  
Partie II  
Partie  
pratique

## Introduction partie pratique

En effet, les tableaux de bords sont considérés comme des plus importantes applications de business intelligence, ils présentent des informations à partir de la machine pour en profiter vu que ce système est moins coûteux dans certains travaux que l'institution effectue. Dans cette partie nous allons aborder la définition de l'institution et recueillir les besoins nécessaires des utilisateurs finals à travers de l'étude de l'existant. A partir de ces besoins nécessaires nous allons établir le système en assurant son développement, son maintien au futur et citer les étapes de la création de l'entrepôt de données.

### Méthodes classiques de développement d'un projet décisionnel

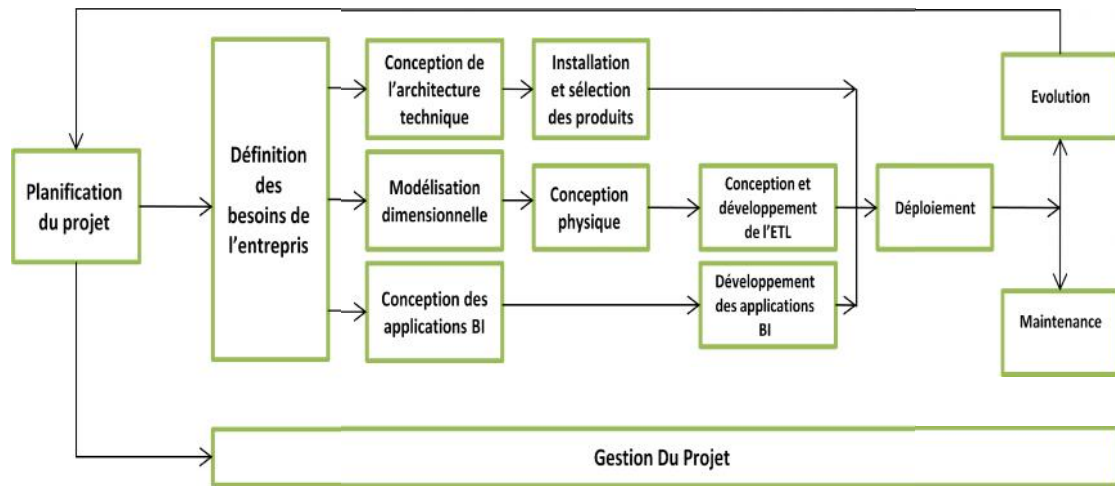
Si nous ne considérons pas les besoins des futurs utilisateurs, l'entrepôt de données deviendrait un exercice vain pour les concepteurs, il doit s'adapter à leurs besoins et non l'inverse. L'approche utilisée pour identifier les besoins analytiques des employés diffère de l'analyse traditionnelle des besoins pilotés par les données.

Généralement, le développement d'un système décisionnel se fait suivant l'une de cas une méthode Kimball est caractérisé par :

- La modélisation de l'entrepôt de données est basée sur le modèle en étoile et ses variantes ( flocon de neige et constellation).
- L'architecture en bus, qui consiste à créer un ensemble de magasins de données liés à chaque département de l'entreprise, ces magasins, lorsque assemblés, créent l'entrepôt de données. Cette approche est possible lorsque les dimensions sont conformes, c'est-à-dire qu'elles représentent exactement la même chose d'un magasin à l'autre.
- La méthodologie de Kimball est mené par les besoins des clients et les exigences métiers.

### Présentation de la méthode

Le cycle de vie d'un projet décisionnel suivant la méthode Kimball est donné par le schéma suivant :



**Figure II :** Le cycle de vie d'un projet décisionnel.

La méthode de Kimball encourage l'approche itérative en impliquant le client, le plus possible. Cette méthode débute par les besoins du client. Par la suite, la création de l'entrepôt de données, tels que trois chemins sont empruntés en parallèles, car ils ne visent pas les mêmes éléments de l'environnement :

- La conception de l'architecture technique : dans cette étape, on doit choisir l'architecture technique et les outils matériels et logiciels nécessaires pour la mise en place de l'entrepôt de données.
- La modélisation dimensionnelle : dans ce chemin, on définit la modélisation dimensionnelle de l'entrepôt et des magasins de données et on définit les outils ETL que nous allons utiliser.
- Conception des applications SD : c'est un chemin dédié pour le développement des applications décisionnelles tels que : les rapports, les tableaux de bord...

Ces trois chemins sont intégrés en fin du projet au moment du déploiement. Le processus complet est répété pour chaque nouveau de données demandé par les utilisateurs finaux. Tout en assurant l'évolution et la maintenabilité du système.

Chapitre I  
Chapitre I  
Etude de cas

« Service Finance - HAOUID BERKAOUI »

(Analyse de Besoin et étude de l'existant)

(Analyse de Besoin et étude de l'existant)

## **1) Introduction :**

Dans ce chapitre on va définir l'institution du HAOUD BERKAOUI et le service où notre étude a été effectuée, on va analyser le système existant dans cette institution. L'utilisation des différentes techniques pour recueillir les informations afin de bien connaître les utilisateurs intéressés et leurs besoins primordiaux pour la réussite de la établir entrepôt de données.

## **2) Présentation de la HAOUD BERKAOUI :**

### **2.1) Historique :**

Le champ de HAOUD BERKAOUI découvert en mars 1965 par le sondage OK-101, celui-ci a été foré au sommet de la structure et a mis en évidence un gisement d'huile dans la série inférieure du trias argilo-gréseux. Actuellement on note l'implantation de plus de 400 puits dans cette région (bloc438) dont environ 25 puits ont été foré comme injecteur, participant au maintien de pression du réservoir.

Un autre système de récupération (gaz lift) a été mis en place à l'instar du système d'injection d'eau afin de maximiser la récupération, en effet durant le premier trimestre de l'année 1992, la production moyenne est passée à 6000 T/J.

### **2.2) Représentation de la région :**

La région de HAOUD BERKAOUI représente une des 09 (neuf) principales productrices d'hydrocarbures de la DP du Sahara algérien.

Sur la route nationale n°49, dite des pétroliers, reliant Ghardaïa à HASSI MESSAOUD et à 30 Km au sud-ouest de Ouargla, un carrefour indique la présence d'un champ pétrolier, il s'agit de BERKAOUI.

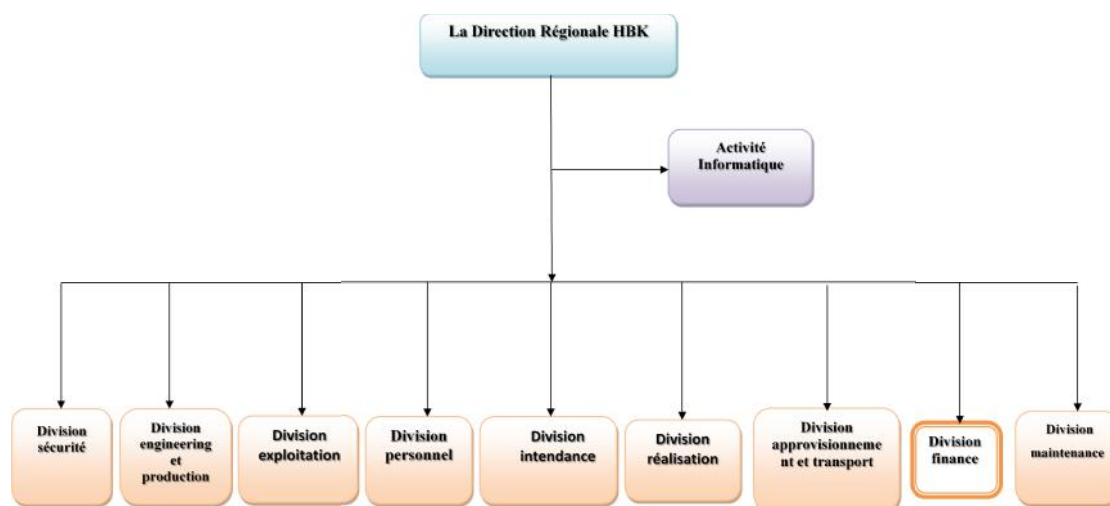
Cette région se situe à 100 Km Nord-ouest de HASSI MESSAOUD et à 750 Km au sud-est de la capitale (Alger) elle est importante en vue de sa part de production des hydrocarbures du pays (elle participe de 6% de la production journalière du pays).

Elle s'étend du sud-est de Ghardaïa jusqu'au champ extrême BOUKHZANA près de la route de Touggourt. La région de HBK se compose de trois principaux centres de production :

- HAOUD BEERKAOUI (HBK).

- GUELLALA (GLA).
- BENKAHLA (BKH).

L'organigramme de direction régionale HAOUD BERKAOUI est identique et illustré dans le schéma suivant :



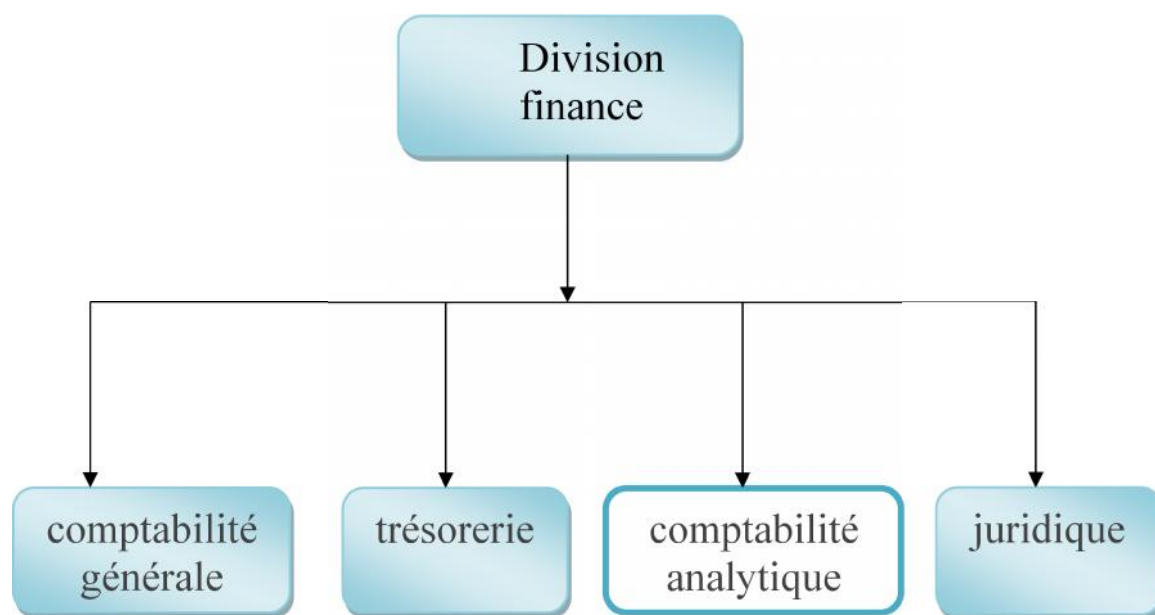
**Figure I - 1 :** L'organigramme de direction régionale HAOUD BERKAOUI.

Après la définition de l'institution, nous parlerons du service où l'objet d'étude de notre sujet a été effectué et qui a été axé sur le service financier, plus exactement, au niveau du service de la comptabilité analytique :

### **2.3) Division finance :**

La division finance assure toutes les procédures financières de la région envers les tiers et structurée comme suit :

- Service comptabilité general.
- Service trésorerie.
- Service juridique.
- Service la comptabilité analytique.



**Figure I - 2 :** Organigramme de la structure division finance.

### 2.3.1) la service comptabilité analytique :

La comptabilité analytique est un mode de traitement de l'information comptable et économique. Elle fait partie de la comptabilité de gestion et constitue par la même un outil de la comptabilité générale, elle permet de montrer au gestionnaires les performances dans le processus d'exploitation et de la réalisation des investissements.

Elle fournit les renseignements qui permettent à la division de contrôler les opérations en cours, de planifier et de prendre des décisions rationnellement.

- Elle peut être un élément de preuve en cas de litige.
- D'évaluer la rentabilité des prix de vente et services en les comparant aux coûts correspondants.
- D'éclairer les choix budgétaires des domaines de l'investissement et de l'exploitation.
- De développer le contrôle interne de l'activité dont l'élément essentiel demeure le coût, particulièrement dans sa forme standard.

La mise en place d'un service de comptabilité analytique au sein de la région « HAOUD BERKAOUI », nous expose à suivre les démarches qui déterminent les activités des structures engagées pour le contrôle budgétaire.

L'objectif essentiel de la comptabilité analytique au sein de la région « HAOUD BERKAOUI », se détermine sur la base de rendement ayant un but économique, et qui est

justifié par les calculs des coûts et le prix de revient de tonne de brute de pétrole par périmètre, ainsi le suivi du budget par nature de charge.

### **3) Etude de l'existant:**

Dans cette étape, nous travaillerons sur l'étude des programmes et systèmes existants utilisés dans les services de la comptabilité analytique.

#### **3.1) Situation informatique de Comptabilité Analytique :**

Nous allons présenter en quelques lignes la situation informatique de HOUD BERKAOUI exactement dans la service comptabilité analytique, elle existe en division finance.

##### **3.1.1) Applications informatiques :**

- **GFAO** : Application métier pour la comptabilité générale, analytique et budget
- ORACLE.
- VMware.
- Système d'exploitation Windows et Unix.
- **Inventaire** : application métier pour la gestion des inventaire.
- Autres applications développées en interne.

##### **3.1.2) Standards de développement :**

L'équipe information technologies de HOUD BERKAOUI privilège la méthode MERISE pour la conception des déferents logiciels en interne, en coté base de données ils utilisent le SGBD "ORACL".

#### **3.2) Etude détaillée des applications métiers :**

C'est une solution logiciel pour la gestion d'une activité donnée au sein d'un entreprise et application métiers de HBK sont :

##### **3.2.1) Système de gestion des bases de données :**

SONATRACH utilise principalement Oracle comme système de gestion de base de données.

##### **3.2.2) Système d'exploitation :**

Un système d'exploitation est une couche logicielle qui facilite la communication homme-machine et offre une plateforme pour d'autres applications. La grande majorité de nos Serveurs utilise Microsoft Windows 2003 Server et les postes clients utilisent Microsoft XP ou Vista. Par ailleurs, nous utilisons aussi UNIX pour les serveurs de base de données.

##### **3.2.3) Serveur d'application :**

Est un logiciel un contexte d'exécution central aux application.

## 4) Etude des besoins :

Après la finition de la phase d'étude actuelle, on va examiner les différentes techniques utilisées pour recueillir les informations nécessaires afin de nous aider à établir un entrepôt de données réussi. Comme il est indiqué ci-dessous :

La phase de collecte et d'identification des besoins est primordiale pour la construction de notre système. A fin de répondre aux attentes des décideurs.

« Les chances de succès d'entrepôt de données se trouvent considérablement accrues par la bonne compréhension des utilisateurs et de leurs besoins » [24].

### 4.1) Techniques de collecte d'information :

#### 4.1.1) Le questionnaire :

C'est un outil qui permet de quantifier et comparer l'information. Cette façon la plus rapide et facile d'obtenir beaucoup d'informations et cette information est collectée auprès d'un échantillon représentatif de la population visée par l'évaluation. C'est un ensemble de questions construit afin d'obtenir l'information correspondant aux questions de l'évaluation[8].

❖ **Avantage :**

- Permet de recueillir une quantité importante d'informations.
- Permet de consulter un grand nombre de bénéficiaires finaux.

❖ **Désavantage :**

- Nécessite des ressources importantes.
- La formulation peut influencer les réponses.
- Les réponses ne sont ni complètes ni détaillées.

#### 4.1.2) L'entretien :

C'est une technique destinée à recueillir des informations à l'aide de questions posées verbalement par l'analyste. Les entretiens sont menés soit individuellement, ou par petits groupes. C'est sans doute la technique la plus utilisée et orientée vers un but. Afin de pouvoir alimenter la recherche d'informations sur un thème fixé [8].

❖ **Avantage :**

- Il ne nécessite pas beaucoup de ressources ni de personnel.
- Il permet d'établir un contact plus personnel avec les populations ciblées.
- Il permet d'obtenir des informations très fines et très détaillées.

**❖ Désavantage :**

- C'est une technique difficile à préparer et à mettre en œuvre.
- Il demande souvent du temps.
- Coûteux en temps et argent.

**4.1.3) La documentation :**

pour connaître le fonctionnement de l'entreprise et tirer des informations sans interrompre son fonctionnement par des questionnaires ou entretiens, utilisés par les différents services de l'entreprise [28].

❖ **Avantages :** informations complètes, n'interrompt pas le fonctionnement de l'entreprise, l'information existe déjà.

❖ **Inconvénients :** coûteuse en temps, il se peut parfois que l'information soit incomplète par rapport à ce qu'on cherche.

**4.1.4) L'observation :**

C'est un mode de collecte des données précises par lequel le chercheur observe de lui-même des processus ou des comportements se déroulant dans une organisation, pendant une période de temps délimité [4].

**❖ Avantage :**

- Elle produit des informations sur les pratiques réelles.
- Elle permet de constater des décalages entre le dire et le faire.
- Elle ne nécessite pas beaucoup de ressources.

**❖ Désavantage :**

- Il y a un risque de passer à côté des données importantes ou d'orienter les résultats.
- Elle demande souvent du temps.
- Difficultés à interpréter certains comportements.

**4.1.5) Les réunions :**

permet d'explorer un sujet en profondeur à travers la discussion de groupe et ces réunions peuvent favoriser une libre expression créative avec un nombre limité de participants.

**❖ Avantages :**

rapide et fiable pour obtenir des impressions communes, obtentions d'informations clés, efficace pour obtenir des informations complètes et détaillées en peu de temps.

**❖ Inconvénients :**

difficulté d'analyse des réponses, difficile à planifier à cause de l'emploi du temps des intervenants.

## 4.2) Justification de L'approche Utilisée :

Nous avons choisi d'adopter une démarche qui nous a permis de combiner, d'une manière assez satisfaisante, entre l'approche bas en haut « Bottom Up » de Ralph Kimball, basée sur l'étude des besoins des utilisateurs et l'approche haut en bas « Top Down » de Bill Inmon qui étudie les besoins relatifs aux données réelles.

Après l'étude des différentes approches de collecte d'informations, notre choix s'est porté sur l'approche d'entretien, compte tenu de ses multiples avantages, mais cette approche présente certains inconvénients dont notamment la difficulté d'entretien un grand nombre de personnes à cause de leur contrainte d'emploi du temps. Complétée par documentation qui permet de recueillir des informations complètes. Pour résumer les démarches que nous avons adoptées :

- Etude des systèmes sources, entretiens avec l'administrateur de la base de données.
- Entretiens avec les Utilisateurs de la base de données, et les utilisateurs potentiels du système décisionnel.
- Récapitulatif des besoins détectés.
- Collecte des documents requis.

### 4.2.1) Première approche « Documentation » :

Nous avons utilisé cette approche lors de l'étude du système opérationnel de l'entreprise, la documentation était disponible sur le site officiel.

### 4.2.2) Deuxième approche « l'entretien » :

A cause de ses avantages précédemment cités. Et en rencontrant les représentants des utilisateurs, nous avons pu collecter les informations détaillées. Lors de notre enquête, nous avons pu nous entretenir avec :

- Administrateur de la base de données.
- La directrice système d'information.

## 5) Obstacles et problèmes rencontrés :

Nous avons pu collecter et identifier les besoins de l'entreprise ainsi que les processus qui la pilotent malgré les difficultés rencontrés à tous les niveaux qui ont ralenti très sensiblement la progression de notre travail. Nous citons notamment :

- L'emploi du temps chargé des interviewés.
- Les imprévus, tels que les déplacements, missions, congés.
- L'indisponibilité de personnes concernées par les entretiens et les annulations.

- Les résistances au changement.
- La rétention d'informations sous couvert de confidentialité.

La réticence de certaines personnes, que ce soit par rapport au projet ou par manque de motivation.

## **6) Conclusion :**

A travers de ce chapitre nous avons pu recueillir toutes les informations à travers des réunions et des documentations qui nous ont permet d'identifier les utilisateurs de notre futur système et leurs besoins. Cette étape définira la manière de construire l'entrepôt de données et ses données.

# Chapitre II

## Chapitre II

# Les conceptions

## 1) Introduction :

Dans ce chapitre nous allons mettre en place la conception de la zone d'entreposage, la conception de l'alimentation et la conception des cubes dimensionnels, afin de mieux naviguer dans les données et exploiter de manière efficace les informations contenues dans l'entrepôt de données.

Cette étape d'alimentation représente en général 80% de la charge de travail du projet d'où la nécessité de faire bien attention quant au choix de l'outil à utiliser. Au cours de notre travail nous avons accordé une attention particulière au traitement de la qualité des données, en utilisant des différentes méthodes dont le Master Data Management (MDM). La conception de ces cubes est basée principalement, sur les besoins recensés auprès des utilisateurs et décideurs de l'entreprise.

## 2) La conception de la zone d'entreposage :

### 2.1) Processus de la Modélisation dimensionnelle :

#### a) Sélection du processus à modéliser :

On peut définir un processus comme l'activité naturelle effectuée dans l'entreprise qui est généralement pris en charge par une source de gestion et collectant des informations.

#### b) Déclarer le grain du processus :

Cette phase veut dire, spécifier exactement ce que représente une ligne individuelle de table de faits. Le grain représente le niveau de détail des mesures de la table des faits.

#### c) Choisir les dimensions participantes à ce processus :

Les dimensions participantes résultent de la question suivante "comment peut-on décrire les données issues du processus ?" Donc nous voulons enrichir nos tables de faits d'un ensemble fourni de dimensions représentant toutes les descriptions possibles susceptibles de prendre des valeurs particulières dans le contexte de chaque mesure.

#### d) Identifier les faits numériques :

Les faits numériques qui vont renseigner chaque ligne de la table de faits. Les faits sont déterminés en répondant à la question suivante : "qu'est-ce qu'on mesure ?". Les décideurs sont particulièrement intéressés par l'analyse des mesures de performance de l'activité.

### 2.2) Les dimensions participantes :

Avant de se lancer dans la conception des processus, nous avons élaboré un tableau ci-dessous qui présente la dimension participant aux différents processus la comptabilité analytique (charge). Le but étant d'éviter une éventuelle répétition des dimensions :

Dimensions	Charge
<b>Date</b>	+
<b>Compte</b>	+
<b>Région</b>	+

**Tableau II - 1:** Les dimensions participantes aux différents.

## 2.3) Modélisation du processus « charge » :

### 2.3.1) Définition du processus charge :

Coût des moyens mis en œuvre par une entreprise pour arriver à réaliser : achats, frais de fonctionnement, personnel, impôts et taxes, frais financiers, amortissements, provisions. Les charges figurent au débit dans le compte de résultat [32].

### 2.3.2) Les dimensions participantes :

#### ❖ Dimension « Date » :

C'est « la seule dimension qui figure systématiquement dans tout l'entrepôt de données, car en pratique tout d'entrepôt de données est une série temporelle. Le temps est le plus souvent la première dimension dans le classement sous jacent de la base de données » [33].

« Date »	
Désignation	Détail
<b>Id_Dat</b>	Identifiant de la dimension « Date »
<b>Date</b>	Format complet de la date
<b>Num_Day</b>	Position du jour dans le mois
<b>Day</b>	Nom du jour dans la semaine
<b>Num_Week</b>	Numéro de semaine de l'année
<b>Week_Year</b>	Numéro de la semaine dans l'année
<b>Num_Month</b>	Numéro du mois
<b>Month_Year</b>	Numéro du mois dans l'année
<b>Month</b>	Nom du mois
<b>Num_Quarter</b>	Numéro trimestre de l'année
<b>Num_Semester</b>	Numéro semestre
<b>Year</b>	Année

**Tableau II - 2 :**Tableau descriptif de la dimension « Date ».

❖ **Dimension « Compte » :**

Cette dimension contient les informations de tous les comptes analytique, les comptes sont regroupés par nature de compte.

« Compte »	
Désignation	Détail
Cpt_Cod	Code de Compte Comptabilité Analytique
Cpt_Lib	Libellee de Compte
Nat_Cod	Nature de Compte

**Tableau II - 3 :** Tableau descriptif de la dimension «Compte».

❖ **Dimension « Région » :**

Décrit la zone géographique ou le fait a eu lieu, après l'étude des besoins au sein du group, il parait intéressant de faire des comparaisons par rapport aux zones géographique.

« Région »	
Désignation	Détail
Reg_Cod	Code de Région
Reg_Lib	Libellee de Région
Grp_Reg	Group de Region

**Tableau II - 4 :** Tableau descriptif de la dimension «Région».

**2.4) Les mesurables :**

D'après notre étude des besoins, nous avons détecté les mesurables suivants :

Montant : toutes les charges qui ne se dépense que pour la production, soit des produits, des salaires ou d'autre charges.

**2.5) Les agrégats :**

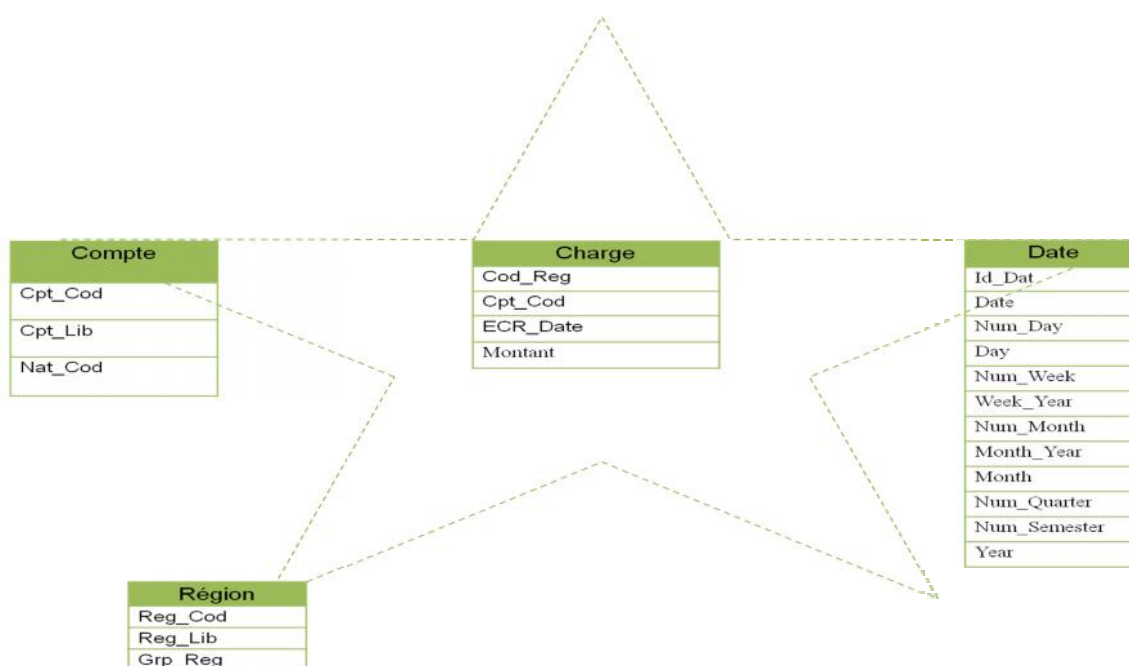
Dimension	Agrégat
Date	Semaine, Mois,Année
Compte	Cpt_Cod, Nat_Cod
Région	Reg_Cod, Grp_Reg

**Tableau II - 5 :** Liste des agrégats « Charge ».

La modélisation par le schéma en étoile, par opposition aux schémas normalisés en 3NF, permet de répondre à trois besoins caractéristiques des systèmes décisionnels :

- Facilité de navigation.
- Nombre de jointures limité.
- Les outils Microsoft ont une sensibilité Kimball.

Le modèle en étoile de l'activité « Charge » :



**Figure II - 1:** Schéma en étoile du processus « Charge ».

### 3) Conception de la zone « Alimentation » :

#### 3.1) Etudes des sources de données :

C'est l'étape qui prend plus de temps, car il nous a fallu étudier et analyser les bases de données du système, connaître l'utilité et le rôle de chaque table, afin de pouvoir déceler les tables porteuses d'informations utiles à la prise de décision.

Nos sources de données sont les bases de données de GFAO :

Table des source	Destination(Zone d'entreposage)	Les attributs de destination)
Feaa_S61	Charge	Cod_Reg Cpt_Cod ECR_Date Montant
Fefa_cpta Fefa_cpt Feaa_nat	Compte	Cpt_Cod Cpt_Lib Nat_Cod
Fefa_ggggrp Fefa_region	Région	Reg_Cod Reg_Lib Grp_Reg

**Tableau II - 6 :** Table de source des données.

### 3.2) Conception de l'architecture ETL :

L'élaboration de l'architecture ETL dès le début du projet est primordiale, car le choix de l'architecture affecte la suite du projet. Tout changement d'architecture engendrera des changements lors de la partie réalisation.

Donc il est important de mettre en œuvre une architecture ETL consistante, qui répond aux exigences et contraintes de l'entreprise, qui dans notre cas sont :

L'architecture du système opérationnel en place.

#### 3.2.1) Quantité importante de données :

La quantité de données importante que nous devons traiter influence fortement la conception de l'architecture ETL et la manière de procéder. En effet le système de production contient une grande masse de données.

### 3.3) Processus d'alimentation d'entrepôt de données :

Nous devons extraire, transformer et charger les données à partir des sources précédemment citées, dans l'entrepôt de données tout en passant par la zone de préparation de données (staging area) :

#### 3.3.1) Extraction de données :

Cette étape consiste à extraire les données de leurs sources, auxquelles nous avons un accès direct via notre poste de travail, et à les charger dans la zone de préparation de données, en évitant de faire beaucoup de transformations sur elles, pour éviter un long chargement de données et par conséquent une longue indisponibilité du système de production. Cette étape

est primordiale vu la quantité de données importante que nous allons traiter, mais aussi, dans le cas où une erreur se produit pendant le processus, comme les données sont dans la zone de préparation, nous n'aurons pas à revenir au système de production et de le réquisitionner à nouveau.

### 3.3.2) Transformation de données :

Maintenant que les données sont dans la zone de préparation, nous devons effectuer une deuxième opération d'ETL : contrôler, nettoyer et transformer les données que nous allons charger dans l'entrepôt de données à partir de la zone de préparation de données.

### 3.3.3) Chargement de données :

C'est la dernière étape du processus ETL, les données sont chargées dans un entrepôt central. Nous pouvons maintenant combinées, agrégés et chargés dans des cubes ou Datamart si cela est jugé nécessaire.

Dans cette phase nous effectuons les types de chargement de données [28] :

- Chargement de la dimension temps.
- Chargement des dimensions.
- Chargement des tables de faits.

#### ❖ Chargement de la dimension temps :

« La dimension temps la plus commune et utile est celle de la dimension temps calendrier avec une granularité journalière » [34].

Un processus ETL a pour mission d'assurer la transmission de données conformes, cohérentes et correctes tout en assurant des performances acceptables.

Lors de la conception du processus d'alimentation de l'entrepôt de données, il est important de respecter certaines règles et objectifs de conception tel que l'assurance du chargement des données de qualités ou la transparence des processus de chargement et de consolidation qui assurent la qualité de données.

Nous avons aussi accordé beaucoup d'importance aux sources de données et à s'assurer de leur authenticité afin de ne pas nuire aux performances de notre système.

Pour avoir un bon système, il faut pouvoir garder l'historique des chargements des données, pour permettre au décideur un suivi de l'avancement des chargements et une correction d'une mauvaise mise à jour.

**❖ Chargement des dimensions :**

Le chargement des dimensions se fait à partir de la zone de préparation de données, où nous avons précédemment chargé les données du système de production, cela en effectuant les transformations et contrôles nécessaires sur les données.

Nous avons relevé dans notre travail deux catégories de dimensions, leur chargement est un peu différent, nous allons donc détailler les étapes par catégories :

**a) Dimensions non changeantes :**

Sont toutes les tables référentielles de la base de données source, ce sont les tables qui resteront inchangées sur une longue période de notre système, elles nécessitent un seul chargement :

- Base de données source.
- Zone préparation de données.
- L'entrepôt de données.

**b) Dimensions changeantes :**

Ce sont les tables de transactions. Ces dimensions nécessitent un chargement initial et des chargements périodiques, afin d'insérer les nouvelles données. Par soucis d'efficacité et rapidité nous n'allons insérer que les nouvelles données à chaque chargement périodique.

**❖ Chargement des tables de faits :**

« Les tables de fait soutiennent les mesures d'une entreprise. La relation entre les tables de fait et les mesures sont extrêmement simples. Si une mesure existe, elle peut être modulé comme une ligne de la table de fait. Si une ligne table de fait existe, donc c'est une mesure ».

Après avoir chargé toutes les dimensions nous passons aux tables de faits. La source est la même que pour les dimensions, c'est-à-dire la zone de préparation de données. Pour chaque enregistrement de table des faits dans l'entrepôt de données, l'identificateur issu du système source est remplacé par une clé de substitution. Cependant l'intégrité référentielle doit impérativement être maintenue, C'est-à-dire : pour chaque clé étrangère d'une table des faits, il doit exister systématiquement une clé correspondante dans la table dimensionnelle.

Au final, l'identificateur de chaque table des faits sera composé des clés des dimensions participantes au fait.

**3.4) La périodicité de chargement de données :**

La périodicité de l'alimentation d'entrepôt de données, dépend essentiellement de la fréquence de la mise à jour ou de l'insertion des tables « Opérationnelles » de la base de données. De ce fait, il faut prendre en considération trois facteurs importants :

- La quantité de données à charger.
- Le temps nécessaire pour le chargement.
- Le moment idéal pour lancer le chargement.

Le chargement périodique des données est assuré par l'ETL, c'est pourquoi ce paramètre joue un rôle décisif quant au choix de l'outil ETL. Pour notre projet, nous avons opté pour une fréquence d'alimentation quotidienne.

#### 4) Conception des cubes dimensionnels :

##### 4.1) Définitions des niveaux et des hiérarchies :

Une dimension est une étape très importante dans la conception des cubes. Les valeurs d'une dimension sont généralement organisées à l'intérieur d'une hiérarchie. L'accès au niveau supérieur dans une hiérarchie est appelé « rolling up », et au niveau inférieur « drill down ».

Dans la première partie de notre démarche, nous allons identifier les niveaux des attributs de chaque dimension avec sa hiérarchie correspondante. Au final nous obtiendrons le tableau récapitulatif suivant :

Dimension	Colonne	Niveaux	Hiérarchies
Date	Id_Dat	Niveau 1 = N1	Hiérarchie 1 H1= N1   N2   N3 N4   N5   N6   All
	Date		
	Num_Day		
	Day	Niveau 2 = N2	
	Num_Week		
	Week_Year	Niveau 3 = N3	
	Num_Month		
	Month_Year		
	Month	Niveau 4 = N4	
	Num_Quarter		
Num_Semester	Niveau 5 = N5		
Year	Niveau 6 = N6		
Compte	Cpt_Cod	Niveau 1 = N1	Hiérarchie 2 H1= N1   N2   All
	Cpt_Lib		
	Nat_Cod	Niveau 2 = N2	
Région	Reg_Cod	Niveau 1 = N1	Hiérarchie 3 H1= N1   N2   All
	Reg_Lib		
	Grp_Reg	Niveau 2 = N2	

**Tableau II - 7 :** Liste des niveaux et des hiérarchies de chaque dimension.

### 4.2) Le niveau « All » :

Ce niveau représente le niveau le plus haut d'une hiérarchie. De ce fait, il est le niveau le plus agrégé. Le niveau « ALL » n'apparaît pas systématiquement dans toutes les hiérarchies[35].

### 4.3) Le cube « Charge » :

Dans le tableau ci-dessous, nous allons dresser le cube à réaliser, Pour ce cube la, on donnera les dimensions participantes ainsi que le mesurable présents dans ce cube :

Nom du cube	Les mesurables	Les dimensions
Charge	Montant	Date
		Compte
		Région

Tableau II - 8 : Le cube « Charge ».

### 4.4) Presentation des cubes dimensionnels :

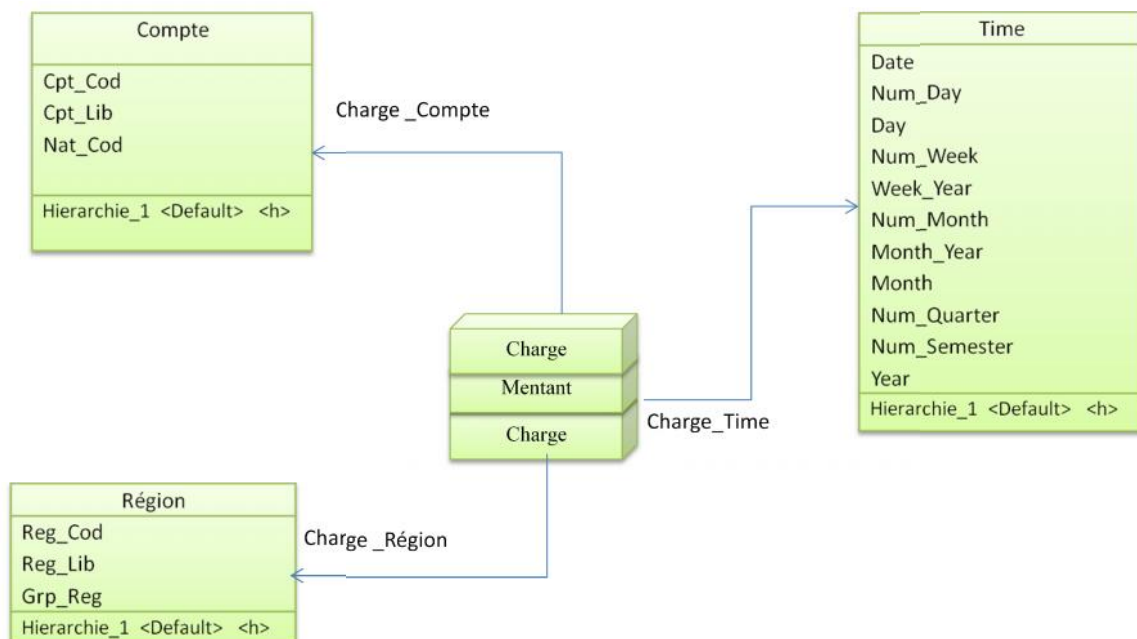


Figure II - 2: Cube dimensionnel « Charge ».

## 5) Conclusion :

Les cubes dimensionnels est une étape très importante dans tout projet l'entrepôt de données. Dans ce chapitre nous avons mis en place la conception des cubes, claire et intuitive afin de faciliter aux utilisateurs une meilleure exploitation des données contenues dans l'entrepôt de données.

Notre approche a consisté à déterminer le grain de l'activité, les dimensions, les mesures et enfin les agrégats.

Dans ce chapitre nous avons défini les différents cubes dimensionnels en précisant les dimensions participantes, les niveaux de leurs attributs et les hiérarchies utilisées. Le chapitre suivant portera sur la partie réalisation et déploiement du projet.

Chapitre III

# Chapitre III

Réalisation

## 1) Introduction :

Dans ce chapitre, on va faire la description du milieu de travail (Logiciel et Matériel) dont on va utiliser pour le système à étudier, aussi qu'on va présenter en détail la solution suggéré. La réalisation est réaliser en quatre étapes :

- Construction de la zone d'entreposage.
- La construction de la partie ETL.
- La construction et chargement des cubes.
- Construction de l'outil de restitution.

## 2) L'architecture technique globale de la solution :

L'architecture technique globale de la solution est illustrée dans la figure suivante :

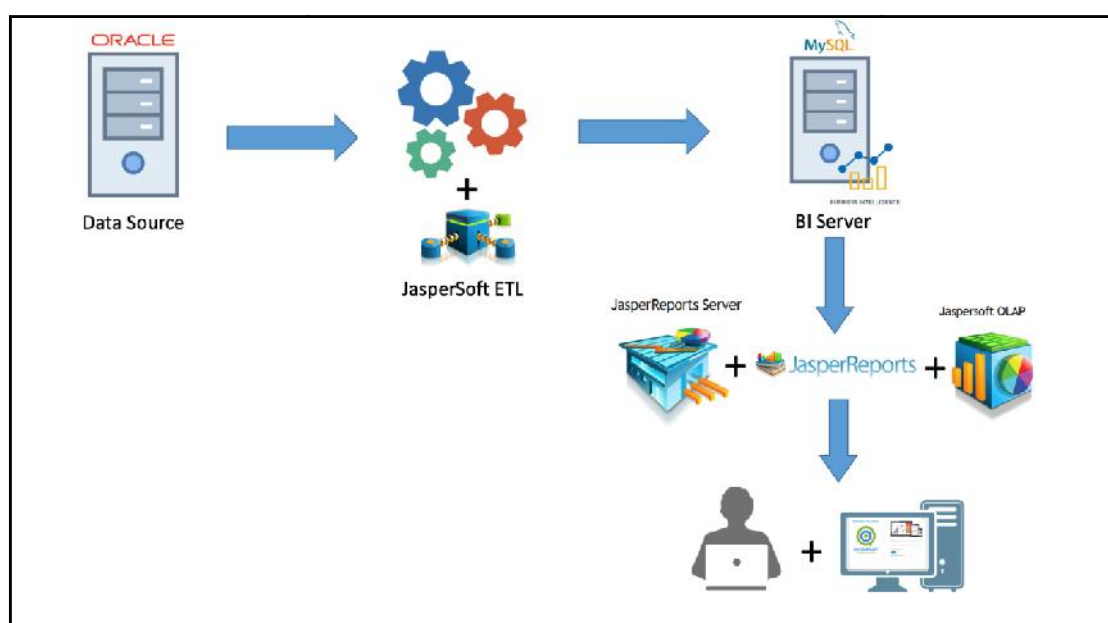


Figure III - 1: L'architecture technique globale de la solution.

## 3) Outils utilisés :

### 3.1) Stockage de la base de données «MySql» :

MySQL est un SGBDR rapide et facile à utiliser utilisé pour de nombreuses petites et grandes entreprises. MySQL est développé, commercialisé et soutenu par MySQL AB, une entreprise suédoise. MySQL devient tellement populaire en raison de nombreuses bonnes raisons[36]:

- MySQL est publié sous licence open source. Donc, vous n'avez rien à payer pour utilise le.

- MySQL est un programme très puissant à part entière. Il gère un grand sous-ensemble de la fonctionnalité des packages de base de données les plus coûteux et les plus puissants.
- MySQL utilise une forme standard de la langue de données SQL bien connue.

### **3.2) Extraction, transformation et chargement « JasperSoft ETL » :**

Jaspersoft ETL est un concepteur de travail ETL complet et prêt à exécuter avec un vaste ensemble de capacités d'intégration de données. Il extrait et transforme de manière cohérente et précise les données de plusieurs systèmes et les charge dans des magasins de données optimisés pour les rapports et les analyses "tels que les marges et les entrepôts de données de schéma étoiles ou flocons de neige. Il permet de suivre facilement les performances des autres outils ETL principaux [37].

Les composants clés et les fonctionnalités principales de JasperETL comprennent notamment [38].

- Job Designer – éditeur graphique qui fournit une vue fonctionnelle des processus d'intégration de données.
- Transformation Mapper – éditeur graphique qui fournit une vue des mappings et des transformations les plus complexes.
- Business Modeler – fournit une vue graphique non-technique d'un besoin métier de gestion de flux de données.

### **3.3) JasperSoft Server :**

Le serveur Jasper Reports s'appuie sur la bibliothèque Jasper Reports pour fournir un serveur de rapports entièrement fonctionnel. Jasper Reports Server fournit des fonctionnalités d'application importantes telles que la sécurité, un référentiel et une planification. Pour permettre l'intégration avec d'autres applications et la possibilité de personnaliser facilement les fonctionnalités, Jasper Reports Server expose des interfaces publiques complètes[39].

Le cœur de Jasper soft BI Suite est Jasper Reports Server, qui permet de :

- Créez facilement de nouveaux rapports à l'aide d'une interface de génération de rapports Ad Hoc intuitive basée sur le Web. (Capacité Edition Professionnelle)
- Gérer de manière efficace et sécurisée plusieurs rapports.
- Interagissez avec les rapports, y compris la saisie des paramètres et le forage des données.

- Organisez des rapports et du contenu Web pour créer des tableaux de bord attrayants et riches en données qui transmettent rapidement les tendances commerciales. (Capacité Edition Professionnelle)
- Accueillir des organisations distinctes de manière sécurisée et transparente par multi-tenancy. (Capacité Enterprise Edition)

### **3.4) Construction du cube « Pentaho Olap Schema Workbench » :**

Le Schema workbench, du projet Pentaho d'analyse de données OLAP « Mondrian » est une interface créatrice qui vous permet de créer et de tester virtuellement des schémas cubes Mondrian OLAP. Le moteur Mondrian traite les requêtes MDX avec les schémas ROLAP (OLAP relationnel). Ces fichiers de schéma sont des modèles de métadonnées XML créés dans une structure spécifique utilisée par le moteur Mondrian. Ces modèles XML peuvent être considérés comme des structures en forme de cube qui utilisent les tableaux FACT et DIMENSION existants dans votre RDBMS. Il n'exige pas qu'un cube physique réel soit construit ou maintenu; Seulement que le modèle de métadonnées est créé. Il fournit les fonctionnalités suivantes[40] :

- Éditeur de schéma intégré à la source de données sous-jacente pour la validation. (Voir au dessus).
- Test des requêtes MDX contre le schéma et la base de données Capture d'écran.
- Parcourir les bases de données sous-jacentes.

## **4) Réalisation du projet :**

### **4.1) Construction de la zone d'entreposage :**

Consiste à créer un dépôt de données relationnelle conforme aux l'architecture en étoile que nous avons proposées dans le chapitre : « Conception de la zone d'entreposage ». Cela s'est fait via My SQL .

Dans cette dépôt de données, les dimensions sont formées de tables avec clés primaires , générées à l'aide de séquences, alors que la tables faits sont des tables avec des clés étrangères de toutes tables dimensions en relation avec le fait, il y a un autre type de table, qui est la staging table, ou zone de préparation de données, qui est une table intermédiaire qui contiendra toutes les données extraites du système source.

### **4.2) La construction de la partie ETL:**

La dépôt de données cible étant prête, nous pouvons commencer la phase ETL, grâce à Jaspersoft ETL qui est aussi un composant Jaspersoft BI.

Jaspersoft ETL, est un outil décisionnel, qui sert à créer des procédés d'extraction / transformation destinés à récupérer des données provenant de bases de données opérationnelles et les envoyer vers l'entrepôt de données.

La première étape consiste à extraire les données du système source vers la staging table sans transformations, afin d'effectuer cette tâche dans les plus brefs délais et libérer le système source, En second étape, une fois la staging table chargée avec les données qu'on a au préalable sélectionnées, nous pouvons commencer à transformer et charger ces données dans les tables dimensions et tables faits.

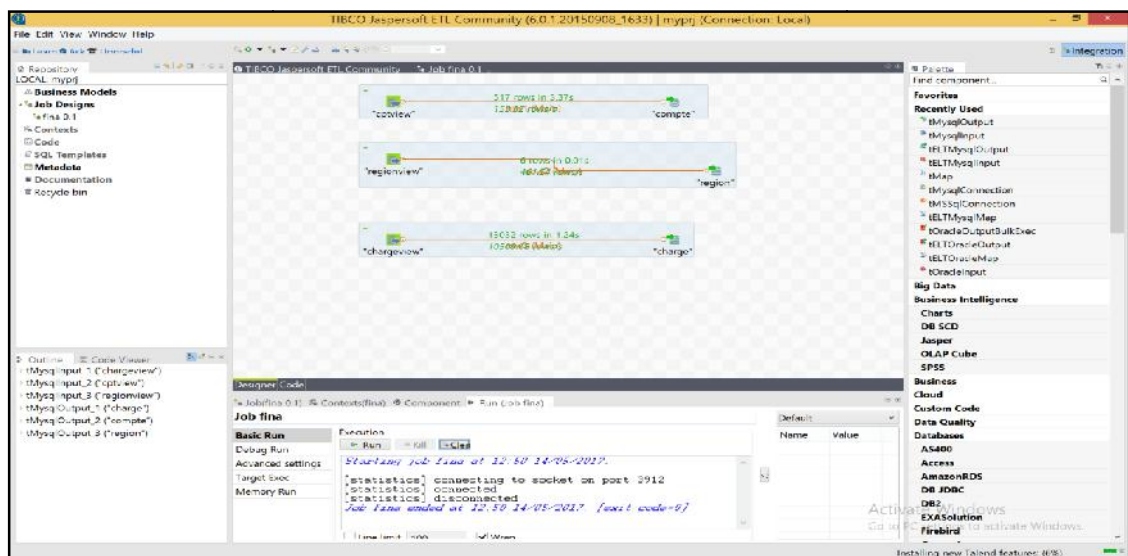


Figure III - 2 : La construction de la partie ETL.

#### 4.3) La construction et chargement des cubes:

Maintenant nous disposons d'une dépôt de données relationnelles qui peut alimenter nos cubes, pour la construction de ces derniers nous utilisons Pentaho Olap Schema Workbench.

Pentaho Olap Schema Workbench, il offre aux utilisateurs la capacité de concevoir et gérer des collections de types de données multidimensionnelles dans My SQL. Cette phase s'est déroulée en cinq étapes :

- Création des dimensions, leurs niveaux et hiérarchies.
- Construction des mappings, c'est-à-dire la correspondance avec les dimensions du Datamart.
- Création des cubes et leurs mesures.
- Construction du mapping des cubes : la correspondance avec les tables de faits du Datamart.
- Chargement des dimensions et cubes.

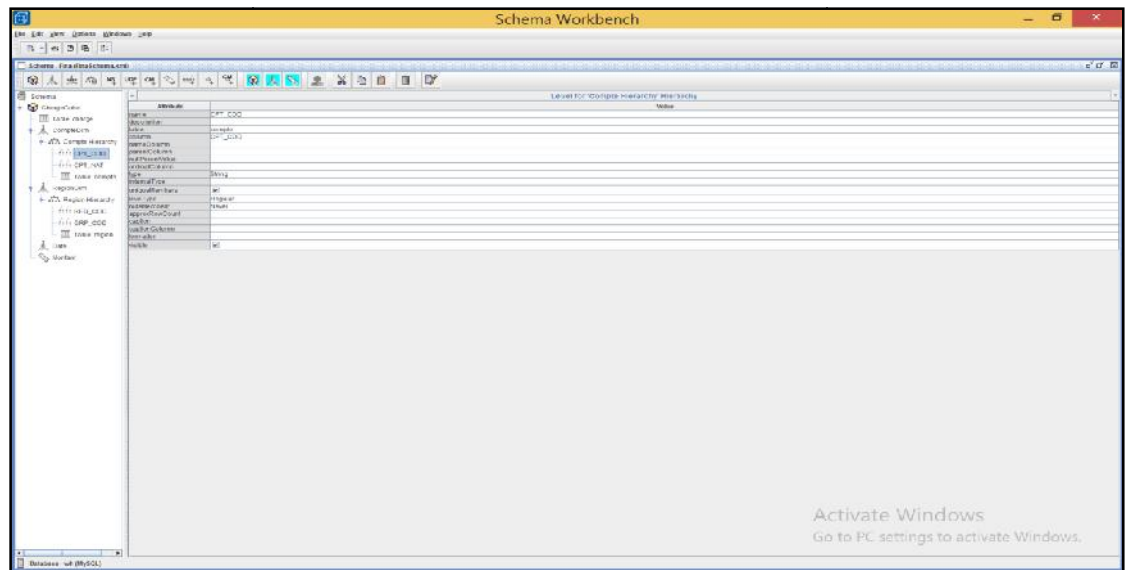


Figure III - 3: La construction et chargement des cubes.

## 4.4) Construction de l'outil de restitution:

### 4.4.1) L'éditeur Ad Hoc :

L'éditeur Ad Hoc prend en charge la création de vues pour différents types de rapports: tableaux, tableaux croisés et graphiques. Vous interagissez intuitivement avec l'éditeur pour créer ces vues en faisant simplement glisser et déposer des éléments. Vous pouvez ajouter et résumer des champs, définir des groupes, étiqueter et signaler le rapport et formater les données pour chaque champ. Vous pouvez également utiliser l'éditeur pour explorer et analyser les données de manière interactive[41].

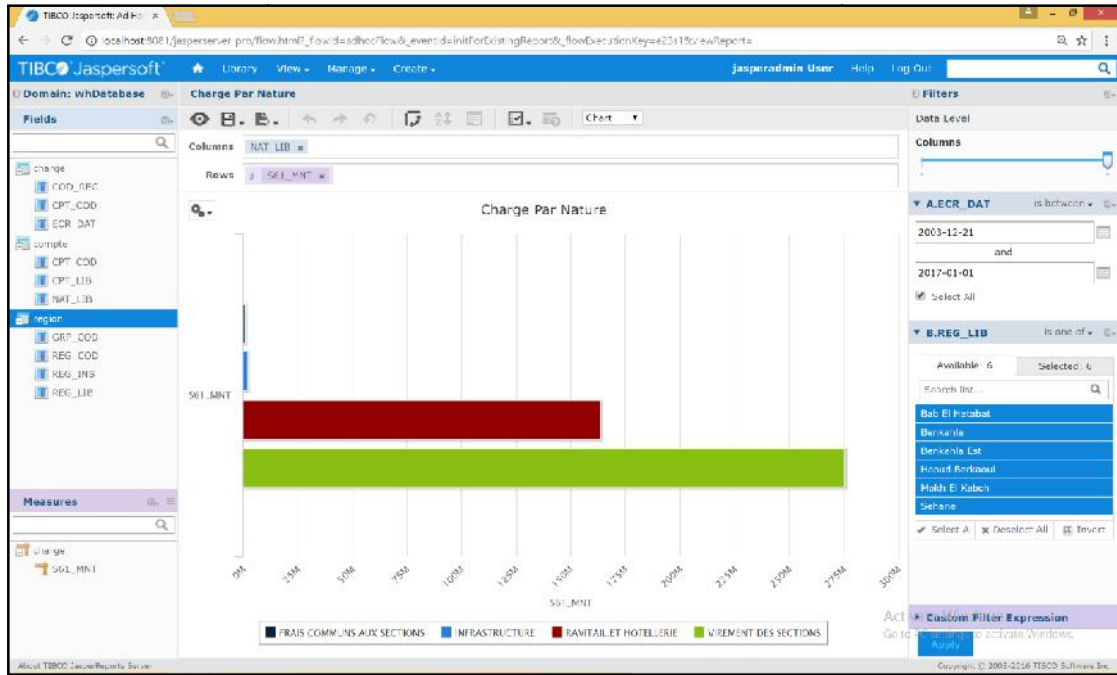


Figure III - 4: Exemple L'éditeur Ad Hoc.

4.4.2) Jasper Server :

Jasper Server est la plateforme décisionnelle de Jasper Soft, société qui développe également le générateur d'états Jasper Reports disponible depuis 2001. non seulement il offre une gamme de produits et fonctionnalités BI complète : tableaux de bord, rapports, prédiction... Il est basé sur une architecture web moderne, très facile d'utilisation, qui offre réellement une nouvelle génération de fonctionnalités BI.

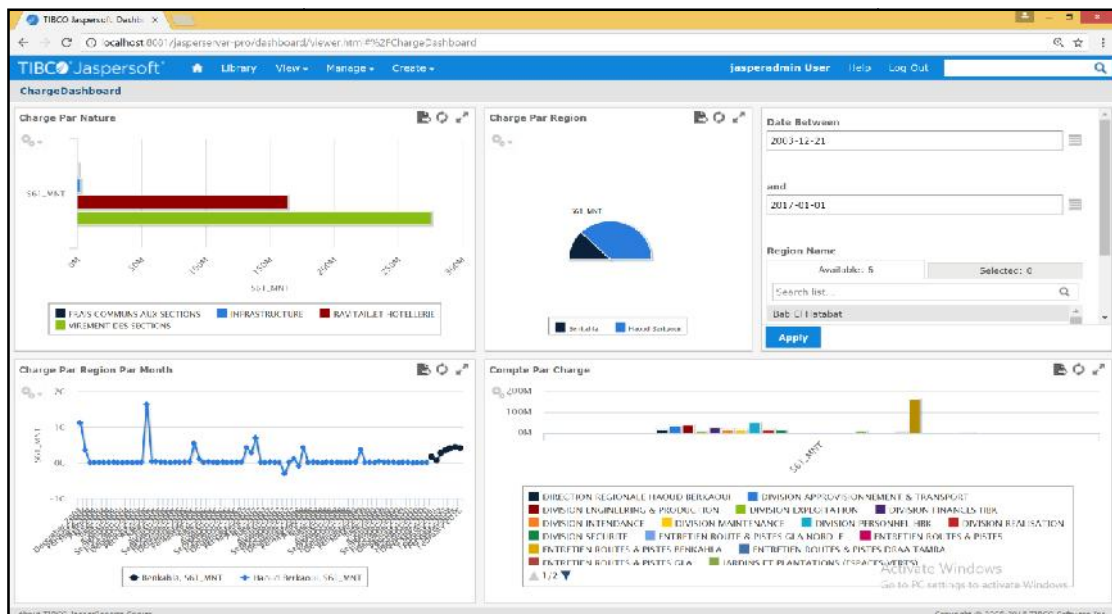


Figure III - 5: Exemple d'un tableau de bord sur JasperSoft Server.

## 5) Conclusion :

Dans ce chapitre, nous avons présenté la réalisation et le déploiement de notre système décisionnel. En premier lieu, nous avons défini l'architecture technique ainsi que les outils utilisés. Ensuite nous avons montré les différentes étapes de la réalisation de notre projet

Conclusion

Générale

## **Conclusion générale :**

Les systèmes décisionnel, occupent une place de premier plan dans la liste des priorités pour l'entreprise et cela dépend de la nécessité du responsable de cette entreprise à une vision globale de ses travaux et être attentifs à leur performance et déterminer la faiblesse, donc cet entreprise doit avoir les informations nécessaires pour prendre de bonnes décisions. L'objectif principal de ces systèmes de circonstances est d'améliorer le processus de prise de décision.

La méthodologie adopté pour la conception et la mise en œuvre de notre thèse de fin d'étude et la méthode " analyse des besoin" créé par Kimball qui est très connue dans le domaine de stockage des données, cette méthode permet de répondre aux nécessités des utilisateurs en profitant el maximum des données créé par le system d'exploitation pour anticiper les besoins non déclarés.

A fin de connaitre les besoins des utilisateurs, nous avons fait une étude complète et détaillé sur l'état d'apprendre les décision au sein de l'entreprise, des interviews sont réaliser avec les utilisateurs et les responsables pour savoir leurs besoins et opinions et pour assurer cette étude on a étudier le contenu des rapports et fichiers demandés par les responsables avant prendre n'importe quelle décision concernant le budget et les couts, ses rapport nous ont bien aider afin de savoir les besoins réels de l'administrateur.

Nous avons commencé ce projet en réalisant le processus de recherche pour les systèmes décisionnels existants car c'est une étape très importante .la chose qui nous a permis d'avoir une idée des systèmes de prise de décision et d'identifier les éléments nécessaires pour les mettre en œuvre. Sur la base de ces circonstances d'étude, nous avons conçu un ensemble orienté objet. Un coût de Datamart.

Phase de conception, nous avons commencé dans la conception de la zone de d'entreposage, au niveau de cette partie, nous avons suivi les principes de dimensions de modélisation qui offrent une vision et une compréhension claire, simple et intuitive. Une fois un modèle en trois dimensions est conçus, où on effectue la création du schéma étoile en trois dimensions pour le sujet à analyser, nous passons par le pouvoir de la zone de l'alimentation, l'étape la plus longue du projet, Dont le but est d'extraire, transformer et charger des données grâce à des outils professionnels certifiés.les cubes et dimensions chargés de notre datamars réaliser par des outils professionnels.

Grâce à la mise en œuvre de ce programme, nous avons pu analyser et recueillir toutes les informations nécessaires pour prendre des décisions correctes à moindre coût et dans les meilleurs délais et que ces circonstances sont des décisions fondées sur le SD et de la stratégie de centralisation des données.

A la fin de ce projet, nous avons pu atteindre les objectifs que nous avons définis préalablement.

Comme pour tout projet informatique, on peut toujours apporter des améliorations après implémentation. Pour cela nous proposons les perspectives suivantes :

- ✓ Création d'autres rapports et cubes de données afin de répondre au mieux aux attentes des futurs décideurs.
- ✓ Recueillir les remarques des utilisateurs et apporter les corrections nécessaires.
- ✓ Implémenter le système sur des plateformes mobiles.

Nous notons que le stage que nous avons fait aux sein de l'entreprise HAUD BERKAOUI nous a permis d'apprendre beaucoup sur le milieu du travail et apprendre sur le coté pratique par des professionnels dans la direction de l'entreprise ce qui nous a permis de travailler avec eux pour acquérir une bonne expérience dans le domaine du système de prise de décision. Ce travail présente l'essence de cinq années d'études.

# Bibliographie

### **Bibliographie :**

#### **Ouvrages :**

- [1] : A. Devisy, « Livre blanc e-Business intelligence, l'effet internet sur le décisionnel », Aubay, 2002.
- [3] : Bertrand Burquier, « Business Intelligence avec SQL Server 2008- Mise en œuvre d'un projet décisionnel », Editions DUNOD ,2009.
- [4] : HOUARI S , KARIM A Conception et réalisation d'un DATA WAREHOUSE pour la mise en place d'un système décisionnel au sein de la SNTR-groupe, Ecole Nationale Supérieure d'Informatique, Alger, 2011
- [5] : ZERGUINI G , Conception et implémentation d'un système b.i pour l'analyse des ventes territoriales des produits pharmaceutiques et leur impact sur les performances d'une force de vente , Ecole Nationale Supérieure d'Informatique, Alger, 2013.
- [7] : LAGRID A Conception et réalisation d'un tableau de bord sur l'entrepôt des bacheliers et aide à la décision pour l'orientation, Ecole Nationale Supérieure d'Informatique, Alger, 2012.
- [8] : GHENABZIA A , OUAMER D , Conception et réalisation d'un système décisionnel pour l'activité stock, approvisionnement et vente au sein de l'actspecialist , Ecole Nationale Supérieure d'Informatique, Alger, 2013.
- [9] : Carlo Vercellis, « Business Intelligence: Data Mining and Optimization for Decision Making », Wiley publishing, 2009.
- [10] : Stéphane Crozat, Introduction au domaine du décisionnel et aux data warehouses, 2016.
- [11] : Bouquin Henry, « Le contrôle de gestion », P.U.F, 2003.
- [13] : GUEROUMI M , LAZIB M , Elaboration d'un système d'information décisionnel au profit de la SPA Alliance Assurances , Ecole Nationale Supérieure d'Informatique, Alger,2013.
- [14] : Mr . KARCH A Mr . ZARROUK N , Conception et réalisation d'un système d'information D'aide à la décision pour la SGA , Ecole Nationale Supérieure d'Informatique, Alger, 2013
- [15] : David Hand « Le CRM analytique les outils d'analyse OLAP et Data Mining » 2001.
- [16] : R. Kimball ; « Entrepôts de données : Guide pratique du concepteur de Data Warehouse », Wiley Computer Publishing 1996.
- [17] : Inmon W.H., Building The Datawarehouse 3ème édition, John Wiley & Sons Inc, New York, 2002.

- [18] : Mr . AFRIT Med , Mr . MESSAOUDI M Contribution à la conception et au déploiement du datawarehouse ,Volet : suivi d'activation des puces et de rechargement des cartes des clients pré-payés , 2007
- [19] : FILALI A , KEDJNANE S , Conception et réalisation d'un Data Warehouse pour la mise en place d'un système décisionnel , Ecole Nationale Supérieure d'Informatique, Alger, 2010.
- [20] : Kimball Ralph et Ross Margy. 2002. The Data Warehouse Toolkit. 2. s.l. : Wiley Computer Publishing, 2002.
- [21] : Louisa Demmou , Exploration de problèmes de performance d'un entrepôt de données, Faculté des Sciences Université de Sherbrooke, 2010.
- [22] : R. Kimball et M. Ross ; « Entrepôts de Données : Guide Pratique de Modélisation Dimensionnelle 2ème édition » ; Vuibert 2002.
- [23] : William H. Inmon, « building the data warehouse », Wiley publishing, 2005.
- [24] : R. Kimball, M. Ross et W. Thornthwaite, « Le Data Warehouse – Guide de conduite de projet », Editions Eyrolles, 2007-06-11.
- [25] : BELHOUET Y , HAMID S, Mise En Place D'une Solution Data Warehouse Pour Le Controle De Gestion Au Sein De Natixis Algerie , Ecole Nationale Supérieure d'Informatique, Alger, 2013.
- [26] : Ralph Kimball, « Le Data Warehouse, Guide de conduite de projet », 2nd Editioin, Eyrolles, 2005.
- [27] : W.H. Inmon, «Building the Data Warehouse», New York: John Wiley & Sons, 1996.
- [28] : MERAGHNI O , YAHI I, Conception et réalisation d'un DATA WAREHOUSE pour la mise en place d'un système décisionnel au sein du groupe TOYOTA ALGERIE , Ecole Nationale Supérieure d'Informatique, Alger, 2012.
- [29] : TESTE, Olivier, 2009. Modélisation et manipulation des systèmes OLAP : de l'intégration des documents à l'utilisateur. HDR. S.l. : Université Paul Sabatier - Toulouse III.
- [30] : Vangenot . C, Datawarehouse,
- [33] : R. Kimball, « Concevoir et déployer un Data Warehouse, Guide de conduite de projet », Editions Eyrolles 2000, France.
- [34] : Ralph Kimball, « The Data Warehouse ETL Toolkit », Wiley publishing, 3rd Edition, 2004.
- [35] : Fon Silvers, « Building and maintaining a data warehouse », CRC Press, 2008.

### Articles et thèses :

[12] : Fernandez, Alain. 2004. Indicateur des performances des tableaux de bord.

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

### Webographie :

[2] : yazid GRIM, « passez en mode BI », <http://grim.developpez.com/articles/concepts/bi-intro/>

[6] : [http://www.memoireonline.com/04/14/8861/m\\_Extraction-des-connaissance--partir-d-un-datamart--l-aide-de-l-arbre-de-decision-application-au8.html#toc30](http://www.memoireonline.com/04/14/8861/m_Extraction-des-connaissance--partir-d-un-datamart--l-aide-de-l-arbre-de-decision-application-au8.html#toc30)

[32] : <http://www.deficreation.com/modules.php?name=Encyclopedia&op=content&tid=448>

[36] : [https://www.tutorialspoint.com/mysql/mysql\\_tutorial.pdf](https://www.tutorialspoint.com/mysql/mysql_tutorial.pdf)

[37] : <http://community.jaspersoft.com/wiki/getting-started-jaspersoft-etl>

[38] : <http://community.jaspersoft.com/wiki/getting-started-jaspersoft-etl>

[39] : <http://information-systems.smile.eu/Tout-savoir-sur/Decisionnel-les-solutions-open-source/Les-suites-decisionnelles/Jaspersoft-bi-suite>

[40] : <http://mondrian.pentaho.com/documentation/workbench.php>

[41] : <http://community.jaspersoft.com/documentation/jasperreports-server-user-guide/overview-ad-hoc-editor>

**Liste des abréviations :**

**B :**

**BDD** : Base de Données.

**BI** : Business Intelligence.

**D :**

**DSS** : Décision Support System.

**DW** : Data Warehouse (Entrepôt de données).

**E**

**EDD** : Entrepôt De Données

**ETL** : Extract, Transform and Load.

**G :**

**GLA** : Guellala.

**H :**

**HBK** : Haoud Beerkaoui.

**HOLAP** : Hybrid on Line Analytical Process.

**M :**

**MDM** : Master Data Management

**MOLAP** : Multidimensional On Line Analytical Process.

**O :**

**ODS** : Opérationnel Data Store

**OLAP** : On Line Analytical Process.

**R :**

**ROLAP** : Relational On Line Analytical Process.

**S :**

**SD** : Système Décisionnel.

**SID** : Systèmes d'Information Décisionnels.

**SGBD** : Système de Gestion de Base de Données.

**SGBDR** : Système de Gestion de Bases de Données Relationnel.

**SQL** : Structured Query Language.

**3FN** : Troisième Forme Normale.