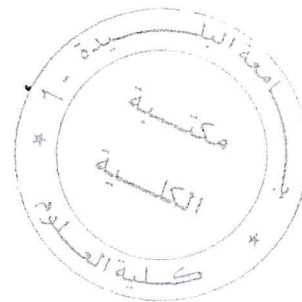
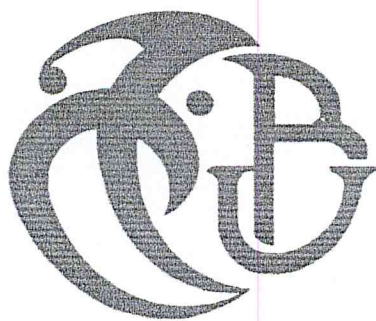


République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique

Université SAAD Dahlab de Blida



Faculté des Sciences

Département d'informatique

Projet de Fin d'Etude en vue de l'obtention du diplôme de MASTER en Informatique

**Conception et réalisation d'un système
d'information décisionnel pour
les ressources humaines**

- Evaluation des employés -

Organisme d'accueil: Sonelgaz

Réalisé par :

YOUSSEF MEHDI

SALAH RIADH

Promotrice:

Encadré par :

Mme S.Oukid

Mr Chergui Mahfoud

Président du jury:
Mme ABED

Examinateur:
M. ZAÏR

Remerciements

La première personne que nous tenons à remercier est notre encadrant Mme Ouqid , pour l'orientation, la confiance, la patience qui ont constitué un apport considérable sans lequel ce travail n'aurait pas pu être mené au bon port. Qu'elle trouve dans ce travail un hommage vivant à sa haute personnalité.

Nos remerciements s'étendent également à tout le personnel du service des ressources humaines chez la société de Sonelgaz à Blida pour leurs bonnes explications qui nous ont éclairé le chemin de la recherche et leur collaboration avec nous dans l'accomplissement de ce modeste travail.

Nous tenons à exprimer nos sincères remerciements à tous les professeurs qui nous ont enseigné et qui par leurs compétences nous ont soutenu dans la poursuite de nos études.

On n'oublie pas nos parents qui par leur contribution, leur soutien et leur patience, leurs prières et leurs encouragements, on a pu surmonter tous les obstacles.

Enfin, on remercie tous ceux qui, de près ou de loin, ont contribué à la réalisation de ce travail.

Résumé: Ce projet s'inscrit dans le cadre de conception et mise en œuvre d'un système d'informatique décisionnel pour les ressources humaines. Ce travail a été fait durant un stage pratique au niveau de l'agence de Blida de Sonelgaz. Un système d'informatique décisionnel est défini comme un ensemble d'outils basant sur un stockage des données spécifique (DataWarehouse) permettant d'intégrer les données éparpillées sur les systèmes de production d'une entreprise sous format multidimensionnelle afin de les exploiter à travers des outils d'analyse. Un SID diffère d'un système d'informatique transactionnel par ses objectifs, ainsi par son architecture: Une architecture en couche qui mis en jeu 4 éléments essentiels (sources de données, ETL, Data Warehouse «DW », application de restitution). Chaque composant nécessite le développement d'un outil particulier. Cette particularité requiert une modélisation spécifique c'est la modélisation multidimensionnelle.

Les Mots Clés: Entrepôt de données, Fait, Dimension, Magasins, Mesure, OLAP, Hierarchie

Abstract: This project is part of design and implementation of a Decision support System for *Human resources*. This work was done during an internship at the agency of Blida of Sonelgaz. A Decision support System is defined as a set of tools based on specific storage of data (DataWarehouse) to integrate scattered data on the operational systems to multidimensional format used throughout the analysis tools. A Decision support System is different from a transactional system by its goals and its architecture. Layer based architecture contain four essential elements (data sources, ETL, DataWarehouse "DW," Application of restitution). Each component requires the development of a specific tool. This special feature requires a specific modeling: is the multidimensional modeling.

Keys Word: DataWarehouse, Data Mart, Fact, dimension, hierarchie, OLAP, Measurement.

ملخص: هذا المشروع هو تصميم و تنفيذ نظام معلومات مساعد على اتخاذ القرارات موجه للموارد البشرية و قد تم هذا العمل خلال فترة تدريب في وكالة البلدية لفرع مؤسسة سونلغاز

يعرف نظام المعلومات المساعد على اتخاذ القرارات على أنه مجموعة من الأدوات القائمة على تخزين البيانات بصيغة محددة (مستودع البيانات) لدمج بيانات متناثرة على نظم المعلومات الانتاجية والمسيرة للشركة و تخزينها على شكل متعدد الأبعاد, وذلك من خلال أدوات متعددة الأبعاد للتحليل و يختلف نظام معلومات المساعد على اتخاذ القرارات عن أنظمة المعلومات التفسيرية في الأهداف و البنية ,بنية الطبقات التي تعتمد على أربعة عناصر أساسية (مصادر البيانات,محمل البيانات, مستودع البيانات و اطار العرض) كل مكون يتطلب تطوير أداة محددة و تتطلب هذه الميزة نموذج محدد هو النمذجة متعددة الأبعاد

الكلمات المفتاحية: مخزن المعطيات , الحدث , بعد , مخزن , قياس , تدرج

Liste des tableaux et des figures		
Tableau 1	Système opérationnel Vs système décisionnel	11
Tableau 2	différences entre système opérationnel et Data-Warehouse	25
Tableau 3	finalités des Data-Mart et Data-Warehouse	38
Tableau 4	Comparaison des processus OLTP et OLAP.	43
Tableau 5	La table FactEval	64
Tableau 6	La table DimPeriode	65
Tableau 7	La table DimEmploye	66
Tableau 8	La table DimPoste	67
Tableau 9	La table DimHabEnca	68
Tableau 10	La table DimHabInterPer	69
Tableau 11	La table DimHabPerso	70
Figure 1	Schéma de principe de l'informatique décisionnelle	9
Figure 2	Schéma en étoile du DataMart	61
Figure 3	Schéma du model logique	63
Figure 4	La Hiérarchie de la dimension « Période»	65
Figure 5	Hiérarchie de la dimension « Employé »	66
Figure 6	Hiérarchie de la dimension « Poste de Travail »	67
Figure 7	Hiérarchie de la dimension « Habilités d'encadrement »	68
Figure 8	Hiérarchie de la dimension « Habilités Interpersonnelles »	69
Figure 9	Hiérarchie de la dimension « Habilités Personnelles »	70

Figure 10	Schéma du modèle physique de DataMart « évaluation des employés »	71
Figure 11	Interface SQL Server R2 2008	75
Figure 12	Tableau de bord du SQL Server R2 2008	76
Figure 13	Exemple de listing de la dimension (employé)	77
Figure 14	Réalisation des jointures en étoile	78
Figure 15	Exemple du model physique	79
Figure 16	Schéma du rendement d'un employé	94
Figure 17	Schéma de comparaison entre le rendement de 5 employés	95

Table des matières		
I	Introduction à l'informatique décisionnelle	1
1	Préambule	1
2	Publics visés	2
3	Objectifs de l'informatique décisionnelle	3
4	Histoire de l'informatique décisionnelle	4
5	Le décisionnel, pourquoi faire ?	6
6	Informatique décisionnelle : système d'aide à la décision ?	7
7	Le projet décisionnel	8
II	Les composants du système décisionnel	10
1	Système décisionnel versus système	10
2	Les éléments du système décisionnel	12

3	Extraction	13
4	Stockage	17
5	Restitution	19
III	Architecture du système décisionnel	22
1	Source des données	22
2	Définition	23
3	La structure	26
4	La construction du DataWarehouse	28
5	La conception du DataWarehouse	31
6	Le Data-Mart	37
7	Le concept OLAP	40
8	Systèmes OLAP versus systèmes OLTP:	42
IV	La modélisation multidimensionnelle	44
1	Concept de Cube	44

2	Concept de Faits	45
3	Concept de Dimension	45
4	Concept d'hierarchie	46
5	Concept de Granularité	46
6	Les Schémas Multidimensionnels	47
7	Complexité Vs Redondance	49
8	Les implémentations des modèles multidimensionnels	49
9	Approche de conception multidimensionnelle	51
V	Implémentation du projet décisionnel	54
1	Problématique	54
2	Présentation de Sonelgaz	54
3	Objectifs et état des lieux	55
4	Périmètres retenus	56
5	Modélisation dimensionnelle	58
6	Le schéma en étoile	61

7	ETL (Extraction - Transformation – Chargement)	62
8	Le modèle logique	63
9	Le modèle physique	64
10	Présentation de l'information	72
11	Présentation de la solution choisie pour l'élaboration du Projet	74
12	Conclusion	96
	Bibliographie	97

I. Introduction à l'informatique décisionnelle

1. Préambule :

Le monde dans lequel nous vivons est de plus en plus complexe. Les technologies de l'information nous génèrent une multitude de données comme jamais auparavant. Le problème n'est donc plus tant d'acquérir une masse de données, mais de l'exploiter. Pour cela il faut collecter de l'information de qualité, la normaliser, la classer, l'agréger, et l'analyser, pour l'exploiter afin de prendre la bonne décision au bon moment.

Dans ce but, il est nécessaire de mettre en place un système d'information particulier, appelé système décisionnel. Ce système doit permettre de présenter de manière simple les chiffres recueillis pour mettre en lumière la conjoncture actuelle et indiquer implicitement la voie à suivre. Un système décisionnel ne remplace pas les systèmes opérationnels qui font fonctionner l'entreprise, mais il vient s'y intégrer, en y extrayant des données, afin d'en diffuser la connaissance, de la manière la plus facilement exploitable par les personnes concernées. Le système opérationnel n'est pas, à priori, modifié par la mise en place du système décisionnel, ce dernier vient le compléter par une exploitation avancée de l'information.

Il est donc nécessaire d'ajouter aux systèmes opérationnels qui permettent très bien de gérer l'entreprise au quotidien, un système offrant la capacité d'analyser le passé, le présent et de simuler l'avenir pour anticiper les changements constants de notre société.

La mise en place d'un système décisionnel permet d'apporter des réponses efficaces à tous les niveaux de l'entreprise ; cet aspect décisionnel est présent dans les organisations depuis de nombreuses années, il revêt l'apparence de rapports et de tableaux de bord. Mais, beaucoup d'entre elles s'aperçoivent que ces simples outils de Reporting ne satisfont pas entièrement leurs attentes. Elles se rendent compte que la mise en place d'un entrepôt de données global, transversal et cohérent, lié à des

outils d'analyses, est nécessaire. La Business Intelligence est devenue une priorité pour les directions informatiques, pour ne pas dire la priorité.

Une plate-forme décisionnelle intégrée est nécessaire pour transformer des données légalement utilisables en connaissance afin d'améliorer la performance à court, moyen et long terme, des organisations. Ce qu'il faut retenir : L'informatique décisionnelle a pour objectif de transformer les données déjà présentes dans et à l'extérieur de l'entreprise, en connaissance, afin de permettre la judicieuse décision au bon moment.

2. Publics visés :

Le défi majeur de l'informatique décisionnelle, bien au-delà de la technologie, est humain. En effet, ce processus de transformation de la donnée en connaissance nécessite de nombreuses compétences qui doivent s'allier pour travailler vers un objectif commun. L'objectif ici, sera de montrer aux différentes parties prenantes, l'ensemble du processus décisionnel, afin que chacun puisse mieux y participer.

Nous nous adressons donc à plusieurs profils que l'on peut répartir de la façon suivante :

- Les techniciens : Les personnes ayant des compétences informatiques doivent pouvoir prendre en main concrètement la plate-forme décisionnelle. Le processus de la chaîne du décisionnel sera vu du début à la fin, avec des approfondissements sur les choix techniques pouvant être mis en œuvre.
- Les architectes et administrateurs de système décisionnel : Les futurs chefs de projets, architectes ou administrateurs de projets décisionnels y trouveront une vision globale de la plate-forme décisionnelle afin d'en conceptualiser la structure.

- Les statisticiens : Les analystes, statisticiens ou Data Miner pourront mieux appréhender les nouvelles opportunités apportées par l'intégration de l'analyse de données dans le processus d'aide à la décision ; les fusions en amonts (ETL, administration, structure de Data Warehouse, OLAP, etc.) et en avals (Reporting, portail, procédure stockée, etc.). Il est primordial d'analyser les données nombreuses et complexes, mais surtout d'industrialiser leur intégration dans le processus décisionnel.
- Les décideurs actuels ou futurs : Il est fondamental d'aider les décideurs présents ou futurs, de mieux appréhender la face cachée de l'iceberg de l'informatique décisionnelle.
- Tout le monde dans l'entreprise, le mot décideurs donne souvent une connotation un peu prétentieuse au sujet. Le terme décideur serait plutôt à prendre au sens large, depuis les cadres ayant à prendre beaucoup de décisions, jusqu'aux employés ayant des rapports à fournir régulièrement. L'objectif est donc de donner à ces utilisateurs non techniciens, une meilleure connaissance de ces systèmes pour qu'ils les utilisent mieux et prennent une part encore plus active dans leur élaboration et leur évolution.

3. Objectifs de l'informatique décisionnelle :

Dans le monde dans le quel nous vivons, les décisions ne peuvent plus être prises uniquement sur des coups de génie. Pour bien gérer une organisation, il est nécessaire de s'appuyer sur des informations de qualité. Les managers d'aujourd'hui et surtout de demain, ont besoin d'analyses pour comprendre le présent et simuler l'avenir.

Le but de l'informatique décisionnelle est de transformer les données de l'entreprise en « Intelligence ». Il est important de préciser que le mot « intelligence » est à prendre dans son acception anglo-saxonne, c'est-à-dire connaissance de l'entreprise ; connaissance de son fonctionnement, de ses processus, de son organisation, de ses ressources humaines, de son histoire, de son présent mais aussi de son futur probable.

Les entreprises sont en perpétuel changement. Les organisations privées ou publiques ont en général comme points communs, des clients, des employés et des partenaires de plus en plus exigeants. Face à ces changements de plus en plus rapides et à cette concurrence de plus en plus forte la simple réactivité ne suffit plus : il faut anticiper. Cette anticipation ne peut être efficace qu'en s'appuyant sur des informations pertinentes, des prévisions justes, voir des simulations et la recherche d'optimum. Mais, dans leurs organisations actuelles, les données sont volatiles, surabondantes, non organisées pour la prise de décision, et souvent éparpillées dans de multiples systèmes hétérogènes. Il devient donc capital de rassembler et d'homogénéiser les données afin de permettre l'analyse des indicateurs nécessaires aux prises de décisions.

De manière communément acceptée, le terme Business Intelligence (BI) se traduit en français par informatique décisionnelle (ID) et de manière réciproque.

4. Histoire de l'informatique décisionnelle :

C'est une longue histoire, alors pour ne citer que quelques grandes dates clés, commençons en 1958, le chercheur d'IBM Hans Peter Luhn, a utilisé le terme Business Intelligence qu'il a défini comme « *la capacité de présenter les interrelations entre des faits de telle sorte que cela permet de guider les actions pour atteindre le but espéré.* »

L'informatique décisionnelle est associée au concept de Data-Warehouse, ce qui se traduit généralement en français par entrepôt de données. Ce concept a été formalisé en 1992 par Bill Inmon dans l'ouvrage "*Developing the Data Warehouse*" (Construire l'Entrepôt de Données) de la façon suivante : « *Un Data-Warehouse est*

une collection de données thématiques, intégrées, non volatiles et historisées pour la prise de décisions. ».

En 1996, Ralph Kimball, a publié le livre de référence « The Data-Warehouse Toolkit ».

Selon les ouvrages, on lit souvent qu'à la suite des infocentres des années 70 suivent les EIS (Executive Information System) des années 90, avant de parler de l'informatique décisionnelle avec le Data-Warehouse et la Business Intelligence.

Voici donc quelques définitions qui n'ont certainement pas la prétention d'être la vérité absolue mais simplement une proposition critiquable.

Voici trois définitions préliminaires :

Infocentre : collection de données pour la prise de décision. Terme utilisé d'après la littérature entre les années 70 et la fin des années 90. On dit souvent que l'infocentre est l'ancêtre du Data-Warehouse.

EIS : Executive Information System, terme apparemment utilisé dans les années 90 pour désigner des outils de restitution d'information synthétique, souvent sous forme graphique, généralement seulement pour le top management. Contrairement aux outils Business Intelligence actuels, les EIS fournissaient des rapports statiques. Il était quasiment nécessaire de faire un projet informatique dans les règles pour chaque rapport, c'est-à-dire faire un cahier des charges pour le développement et le test. Il y avait donc un décalage important entre le moment du besoin et celui de la livraison.

SIAD : Système d'Information d'Aide à la Décision ou bien aussi Système d'Information et Aide à la Décision, ou encore Systèmes d'Information et Analyse Décisionnelle.

Pour certains, ces termes : Infocentre, EIS voir SIAD sont obsolètes. Souvent, on considère que les infocentres et les EIS étaient statiques, alors que maintenant, une plateforme BI permet une interactivité. L'utilisateur peut lui-même créer la plus part de ses rapports.

Tout comme ces définitions sont très techniques, dans l'expression « informatique décisionnelle », il y a le mot « informatique » qui donne maladroitement une

connotation technique qui peut faire peur. Certes, les projets d'informatique décisionnelle requièrent de nombreuses compétences informatiques, mais il ne faut pas perdre de vue que leur objectif est de mettre en place des solutions d'aide à la décision pour des personnes non informaticiennes.

Il faut remarquer que les expressions « entrepôt de données » et « informatique décisionnelle » recouvrent tous les deux, deux notions importantes :

- Pour la première, l'entrepôt de données ou Data-Warehouse, renvoie à l'infrastructure décisionnelle dans son ensemble, ou seulement au contenant des données décisionnelles, le stockage de l'information décisionnelle.
- Pour la seconde, l'informatique décisionnelle ou Business Intelligence, renvoie à l'ensemble du processus dans sa globalité (extraction - validation - transformation - chargement - stockage - restitution - analyse de données - diffusion de la connaissance - génération de rapport - gestion de la performance - etc.) ou seulement aux applications finales de Reporting et de pilotage, partie émergée de l'iceberg, visible par l'ensemble des utilisateurs.

5. Le décisionnel, pourquoi faire ? :

L'informatique décisionnelle doit permettre d'avoir une vision claire, nette et précise du passé, mais aussi de comprendre la situation actuelle et de prévoir ou simuler le futur.

- **Passé** : tous les utilisateurs du système décisionnel doivent pouvoir naviguer dans l'information dont ils ont besoin, depuis le détail, jusqu'au général. Ils doivent pouvoir faire des tableaux croisés, des histogrammes, etc. de manière intuitive pour faire des rapports et les partager. Un système décisionnel a notamment pour fonction d'être la mémoire de l'entreprise. La première fonction du décisionnel est de permettre à n'importe qui dans l'organisation d'accéder à l'information dont il a besoin, tout seul, de manière intuitive ; c'est-à-dire, offrir à tout le monde la possibilité, sans formation, en trois clics

de souris, d'avoir une information précise. À la lumière du passé, on découvre souvent implicitement la voie à suivre.

- **Présent** : il est important d'avoir l'information juste de la situation actuelle, consolidée avec les informations sur le marché et sur les partenaires de l'entreprise, pour savoir où on est, par rapport aux autres, dans quel environnement. L'informatique décisionnelle doit permettre de comparer des données internes avec des données externes de façon fiabilisée.

- **Futur** : un système décisionnel intégrant des outils d'analyses de données doit permettre de simuler le futur en connaissance de l'historique. Cette vision du futur reste une simulation avec une incertitude plus ou moins importante, néanmoins, elle est d'une très forte valeur ajoutée pour la prise de décision.

6. Informatique décisionnelle : système d'aide à la décision ? :

On trouve maintenant des systèmes opérationnels dans presque tous les domaines. Certaines entreprises, étant dans un milieu encore plus concurrentiel et plus complexe, ont déjà installé des outils d'informatique décisionnelle. Néanmoins très peu de grands groupes ont une plateforme décisionnelle d'entreprise intégrée leur permettant d'avoir une vision transversale, intégrant des prévisions, des simulations et de l'optimisation, pour l'analyse et la gestion de la performance. De plus, les changements sont perpétuels et les systèmes décisionnels sont donc en perpétuelle évolution.

Un système décisionnel est par définition un système d'aide à la décision. Cela ne veut certainement pas dire qu'un tel système prend des décisions à la place d'êtres humains et aussi intelligents soient ces systèmes, ils ne nous remplaceront pas. Par contre, au lieu de prendre des décisions dans le flou, les décisions seront prises en

connaissance de cause. Il ne va pas sans dire qu'une telle transformation dans le processus de prise de décision, ne se fait que très rarement sans réticences. Il est donc important d'intégrer un vrai processus de gestion du changement pour amener l'ensemble des décideurs à utiliser la connaissance apportée par le système décisionnel. Dans un monde de plus en plus complexe, il est devenu vital d'utiliser la technologie pour exploiter le déluge d'information à des fins d'analyse de la situation passée, présente, et future, pour prendre la bonne décision au bon moment.

7. Le projet décisionnel :

En résumé l'**informatique décisionnelle** permet de **transformer la donnée en connaissance**.

La chaîne du projet décisionnel est constituée de 5 étapes :

Planification ---- ETL ---- Stockage ---- Analyses ---- Restitution

- La planification : Mettre en place une plateforme décisionnelle est généralement un projet complexe, il est donc important de commencer sur des bases solides. Il faut pour cela avoir une méthodologie bien définie.
- ETL (extraction - transformation - loading) que l'on peut traduire par extraction - transformation et chargement. La chaîne du décisionnelle commence par l'intégration, la validation de la qualité des données, la transformation et le chargement de l'entrepôt de données.
- Une plateforme décisionnelle doit souvent gérer d'importants volumes de données. Les bases de données décisionnelles se mesurent parfois en dizaines voir centaines de téraoctets. La performance du stockage de cet entrepôt est cruciale.

- Analyse : le décisionnel doit permettre de transformer des données en connaissance ce qui nécessite généralement des outils de statistique, voir de d'économétrie ou de recherche opérationnelle.
- Restitution : la chaine du décisionnel se termine par la diffusion de la connaissance à travers un portail, un outil permettant à des utilisateurs non informaticiens de faire de façon intuitive en quelques cliques des rapports, des analyses et de les partager.

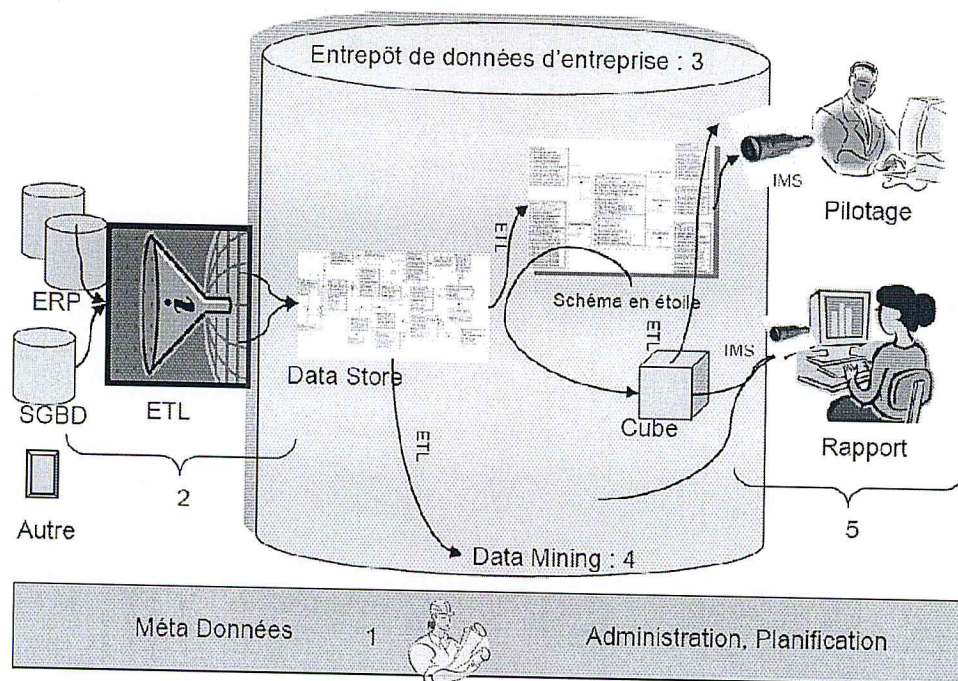


Figure 1: Schéma de principe de l'informatique décisionnelle

II. Les composants du système décisionnel

1. Système décisionnel versus système opérationnel :

Tout d'abord, avant de rentrer dans le détail des systèmes décisionnels, positionnons-les par rapport aux systèmes opérationnels.

Un système opérationnel est un système informatique permettant de gérer l'information au quotidien. Toutes les grandes entreprises ont mis en place de tel système. Parmi ces systèmes notons les ERP, Enterprise Ressources Planning, ou en français PGI : Progiciel de Gestion Intégré. De tels systèmes permettent notamment de gérer des flux d'information à travers l'entreprise, de gérer la comptabilité, de générer les factures, les feuilles de paye, la logistique, les processus, etc.

Par exemple, si un commercial vend une voiture, un flux d'information va être transmis à la production qui devra assembler cette voiture, au service comptable qui va générer la facture, aux approvisionnements qui devront prendre en compte la demande dans leur commandes, etc. De tels systèmes permettent l'industrialisation de la communication d'informations opérationnelles. Il est donc normalement possible à tout moment de savoir combien il y a de voiture dans le portefeuille de commande, où en est la commande de ce client, est-ce que les stocks sont suffisants, etc. Un système opérationnel permet l'efficacité au jour le jour.

Par contre, si l'on souhaite faire des rapports, des tableaux de bord présentant une information transversale, si l'on désire faire des prévisions et des simulations, comme par exemple savoir quel client est susceptible de commander un nouveau produit, ou quel autre client pourrait partir à la concurrence, alors ces systèmes opérationnels peuvent difficilement, et souvent pas, répondre à ces problématiques ; c'est alors qu'interviennent les systèmes décisionnels.

Un système décisionnel est un système informatique complémentaire des systèmes opérationnels auxquels il devra s'intégrer.

Techniquement, un système opérationnel est conçu pour gérer beaucoup de petite transaction par seconde, alors qu'un système décisionnel va gérer peu de transactions mais beaucoup de requêtes complexes.

Un système décisionnel doit notamment permettre de passer de la simple réactivité à l'anticipation et à la pro-activité. Etre proactif, cela veut dire par exemple qu'au lieu d'attendre que le client vous envoie la résiliation de son contrat, étant donné que vous avez déjà modélisé ce comportement, vous lui envoyez une proposition avant qu'il ne résilie, ce qui vous permet de le fidéliser. Les organisations ont pour la plus part en place leur système opérationnel pour être efficaces et réactives, elles doivent maintenant pour être proactives, déployer un système décisionnel.

➤ **Récapitulatif :**

Système opérationnel	Système décisionnel
Efficacité	Avantage compétitif
Jour le jour	Stratégie à long terme
Beaucoup de transactions par seconde	Beaucoup de requêtes complexes par seconde
Application statique	Application dynamique
Base de données changeantes	Historique évolutif
Routine	Créativité
Réactivité	Pro-activité
Piloté par les processus	Piloté par les métiers

Tableau 1 : Système opérationnel Vs système décisionnel

2. Les éléments du système décisionnel :

Les données opérationnelles sont souvent disséminées dans l'entreprise dans des systèmes hétérogènes. Pour centraliser et automatiser le traitement de ces données, le Système d'Information Décisionnel va remplir trois fonctions : extraction des données, leur stockage et la restitution des données sous une forme exploitable.

- **L'extraction**, souvent effectuée à l'aide d'un outil d'ETL (Extract, Transform, Load), consiste à aller chercher les données là où elles se situent, à les trier, et à les transformer éventuellement afin d'effectuer un pré-traitement pour faciliter l'analyse. Dans cette phase se fait également le nettoyage des données : l'homogénéisation, la suppression des doublons, la détection de données non-conformes. Ensuite, les données sont centralisées dans des bases de données du Data-Warehouse.
- **Le stockage** sert à structurer les données au sein d'un Data-Warehouse. Il s'agit de mettre en place un schéma relationnel orienté décisionnel.
- **La restitution** ou **reporting**, consiste en l'analyse des données et à la diffusion des informations. C'est la face visible du décisionnel, la partie que voient les utilisateurs. Nous distinguerons 3 sortes d'analyse de données : la simple requête vers une base (multidimensionnelle ou non), l'analyse multidimensionnelle et la constitution de tableaux de bord, et le datamining (fouille de données), consistant à mettre en relation des corrélations éventuelles entre les données afin de dégager une tendance.

Nous allons voir en détail chacune de ces fonctions.

3. Extraction :

- **Objectif :**

Cette phase a pour but l'alimentation des tables du Data-Warehouse par des données issues d'une transformation de données brutes primaires.

Garant de la qualité de la donnée présente dans le Data-Warehouse, l'étape d'extraction est critique: une analyse pratiquée sur des données erronées ou tronquées peut fausser la stratégie d'une entreprise.

Nous distinguerons 4 étapes :

- Recherche et extraction des données
- Nettoyage des données
- Transformation des données
- Chargement des données

La **phase de recherche et d'extraction** des données consiste à aller chercher seules les données pertinentes en fonction de leur source.

La **phase de nettoyage des données** consiste à les épurer : détection d'anomalies, suppression des doublons, regroupement des données identiques (par exemple, un fournisseur peut être référencé sous 3 identifiants différents), détection d'incohérences...

La **phase de transformation** regroupe les opérations de mise au format nécessaires des données, de calcul des données secondaires et de fusion ou d'éclatement des informations composites. Nous mettons les données au format désiré pour faciliter l'analyse lors de l'étape du reporting.

Enfin la **phase de chargement** a pour rôle de stocker les informations de manière correcte dans les tables de faits correspondantes du Data-Warehouse.

✓ Un outil : l'ETL

Afin d'effectuer ces différentes opérations, un outil appelé ETL (Extract, Transform & Load) est apparu, chargé d'automatiser les traitements et les rendre facilement paramétrables.

Le terme « ETL » est maintenant rentré dans le langage courant, mais pour présenter la démarche, on pourrait dire (Extraction - Transformation - Chargement)

- **Extraction** : Plus d'extraire de l'information, il faut s'intégrer aux systèmes sources, ce qui implique une communication bidirectionnelle. Le système opérationnel fournit des données brutes, détaillées, et le système décisionnel doit parfois en retour fournir de la connaissance à forte valeur ajoutée au système opérationnel.
- **Transformation** : si l'on souhaite avoir des rapports de qualité, commençons par analyser la qualité des données, il y a-t-il des doublons, des valeurs aberrantes, manquantes. La validation et la consolidation de l'information est une étape complexe cruciale.
- **Chargement** : Définition de la structure cible après création des processus d'extraction, de normalisation des données sources de transformation et leur chargement dans cette même structure cible.
- **Avantages de l'ETL** :

Les grandes forces des ETL sont :

- leur souplesse, ils sont fortement paramétrables
- leur facilité d'utilisation, utilisables par des non informaticiens après formation

- leur facilité de maintenance et pérennité

En effet, on pourrait se passer d'un ETL, les bases de données comprenant souvent des outils d'alimentation qui peuvent extraire des données, effectuer des transformations basiques et charger des données. Mais ces outils n'ont pas les avantages d'un ETL et sont souvent difficilement paramétrables, pas ergonomiques et n'offrent pas beaucoup de possibilité de transformation de données.

- **Limites de l'ETL :**

Coûts : En ressources matérielles et en temps : en effet, ils font subir différents traitements à de gros volumes de données. Un ETL nécessite donc un fort besoin en machines, débit et maintenance.

Temps réel : La durée des traitements et le temps machine consommé est tel que cela est difficilement compatible avec des traitements en temps réel. La plupart des éditeurs d'ETL pourtant essaye d'évoluer dans ce sens en développant leur propres outils, cependant, au-delà des considérations stratégiques et technologiques, on peut se demander si une application décisionnelle gagnerait en pertinence avec un Data-Warehouse actualisé en permanence.

- **Evolution de l'ETL :**

Nous recensons cinq générations d'outils d'ETL. La première génération a vu apparaître des ETL qui génèrent du code Cobol pour assurer la transition entre les données sources et celles du Data-Warehouse. Même si ces outils ont permis de faciliter la tâche, ils péchaient dans l'automatisation des environnements d'exécution. L'intervention humaine était nécessaire pour assurer l'intégration des données. La mise en place d'un ETL générateur de code nécessitait un budget prohibitif.

La seconde génération a bénéficié de l'avènement des technologies client / serveur pour les applications propriétaires pour proposer un nouveau langage : le

SQL. Cependant, l'apparition de solutions intégrées comme les ERP a créé des difficultés pour extraire les données sources.

La version suivante a permis de pallier à ce problème en proposant des adaptateurs spécifiques. La diversité des données sources est mieux supportée : les données non structurées commencent à être intégrées et des adaptateurs spécifiques permettent d'intégrer les données issues des ERP.

Dans un contexte où les délais pour prendre des décisions se raccourcissent, le quatrième type permet d'intégrer les données au fil de l'eau (intégration instantanée) et en batch (en différé). Cela permet à l'entreprise d'avoir des résultats en temps réel et aux dirigeants de prendre plus rapidement les décisions adéquates, et en cas de suractivité des ressources, on peut mettre les données à intégrer dans une file d'attente.

La dernière génération dispose d'API (Application Programming Interface) qui permet aux données échangées de pouvoir être dirigées vers des cibles multiples. Ensuite, on a réussi à standardiser l'utilisation des fonctionnalités de la précédente génération. On utilise une intégration au fil de l'eau pour les données opérationnelles car les contraintes de l'activité l'exigent et car ces données ne représentent pas un volume trop important. Les données analytiques sont quant à elles intégrées en batch car le contexte est opposé au précédent.

4. Stockage :

La centralisation et le stockage des données se font dans un entrepôt de données, ou « Data-Warehouse ». Cette base est entièrement dédiée au décisionnel.

- **Objectif :**

L'objectif d'un data-Warehouse, ou entrepôt de données, est de servir d'intermédiaire en stockant différentes données issues des applications de production en vue d'être sondé afin de recueillir les informations nécessaires à la prise de décision.

Le Data-Warehouse est ainsi le point focal, lieu unique de consolidation de l'ensemble des données de l'entreprise. Le Data-Warehouse ne produit pas d'informations, il l'extrait des systèmes de production.

- **Présentation de Data-Warehouse :**

C'est la zone de présentation des données est le lieu où les données sont organisées, stockées et offertes aux requêtes directes des utilisateurs, aux programmes de reporting et autres applications d'analyse. Elle est tout ce que les utilisateurs voient et touchent par le biais des outils d'accès. Sur le serveur de présentation, les données sont présentées et stockées sous une forme *dimensionnelle* :

- S'il repose sur une base de données relationnelle, les tables y seront organisées sous forme de schémas en étoile ou ses extensions.
- S'il est basé sur la technologie OLAP (Online Analytic Processing), les données conserveront des dimensions reconnaissables.

La majeure partie des Data-Warehouse sont implémentés sur des bases de données relationnelles. Ce qui explique pourquoi, lorsque nous évoquons les données relationnelles sous-jacentes.

- **Data-Wherehouse ; pour aller plus loin :**

Le rôle axial que joue le Data-Warehouse dans le système décisionnel comme trait d'union entre la phase d'extraction et celle de reporting et l'importance majeure et capitale d'une bonne construction du Data-Warehouse dans le processus du projet décisionnel comme la si bien mentionné **Bill Inmon** quand il a dit « *un Data Warehouse ne s'achète pas, il se construit* » Tous ces points et tant d'autres voudraient qu'on consacre un chapitre indépendant au concept du Data-Warehouse ou nous allons revenir ultérieurement en large et en détail sur tous les aspects techniques et architecturaux qui peuvent découler du Data-Warehouse pour mieux avancer dans la compréhension des tenants et aboutissants du système décisionnel ce qui est la quête suprême avant tout et après tout de ce modeste Mémoire .Cela dit dans un esprit logique nous avons jugés utile de parler du Data-Warehouse à ce niveau d'avancement du Mémoire sans s'étaler juste pour rester dans le fil logique d'enchaînement des idées et garder une démarche plausible et cohérente dans la succession des chapitres traitants du sujet en l'occurrence l'informatique décisionnelle.

5. Restitution :

C'est la face visible de l'iceberg décisionnel.

Cette phase permet la formulation de requêtes afin d'analyser les tendances, de recueillir des informations pertinentes, de constituer des tableaux de bords...

Les résultats doivent répondre du mieux possible aux différents impératifs de restitution: clarté et fluidité, précision mais concision.

- **Objectifs :**

L'enjeu est de tirer parti de l'information recueillie dans le Data-Warehouse.

Ce processus est appelé de manière générale la « Business Intelligence ».

La prise de décision ne pouvant pas – heureusement - être industrialisée, l'entreprise doit se munir d'outils d'aide à la prise de décision.

Ces outils doivent :

- Etre ergonomique et facile d'accès
- Etre utilisable par des non informaticiens
- Permettre d'accéder aux niveaux de données désirés
- Interroger et analyser les données de l'entreprise
- Diffuser et partager l'information
- Etre multimédias et communicants

Les utilisateurs auront à manipuler les données et profiter des services offerts par l'entrepôt de données et cela grâce aux outils et logiciels de restitution, ces derniers doivent impérativement être aussi ouverts que la philosophie et les principes du

Business Intelligence (Simplicité, Souplesse, esprit analyse, ergonomie ...). Les avancées technologiques permettent maintenant d'avoir des outils très simples et très puissants à la portée de tout le monde.

Les outils de restitution sont la face visible de l'iceberg de l'informatique décisionnelle. Mais ils ne peuvent éditer des données qui n'ont pas été correctement extraites et analysées auparavant. La valeur ajoutée d'un système décisionnel est bien plus concentrée dans la constitution de la base multidimensionnelle, que dans ces outils de présentation.

Du point de vue de l'utilisateur, ils n'en restent pas moins indispensables, car ils mettent en lumière les informations et connaissances que les étapes précédentes du processus ont contribuées à créer.

Mais le terme le plus exact à employer reste le (**Reporting**) et le terme (Restitution) n'est que la traduction française.

- **Reporting proprement dit :**

Le premier besoin, le plus demandé par les utilisateurs, celui de produire des rapports. Il est commun de distinguer trois modes de reporting, matérialisés par des rapports aux nuances suivantes :

- ✦ **Rapport statique** : il s'agit d'un rapport dont la structure est figée tant en termes de présentation que du périmètre des données présentées. Il présente l'avantage d'être généré en dehors du temps de travail de l'utilisateur et d'être consommable instantanément. Par exemple, l'acheteur logistique d'une plateforme de commerce voudra disposer de manière hebdomadaire d'un rapport présentant l'évolution de la qualité de service de ses prestataires logistiques (capacité à délivrer la quantité dans les délais et la qualité attendue) afin de peser sur les négociations futures ou établir les pénalités.

- ✦ **Rapport dynamique** : il s'agit d'un rapport ayant une structure de présentation semi statique au sein duquel le périmètre des données peut varier. Cela en offrant à l'utilisateur final la possibilité de choisir

les valeurs de paramètres dynamiques intégrés dans le rapport. Cela présente l'avantage indéniable de disposer d'une grande amplitude sur les rapports finaux puisque à partir d'une même structure de rapport on pourra par exemple aussi bien générer un rapport de comparaison annuelle qu'un rapport de comparaison trimestrielle.

✦ **Rapport Ad hoc** : il s'agit d'un rapport librement aménagé par l'utilisateur final. Il aménage les données qui l'intéressent dans une structure type tableau ou graphique en définissant le périmètre de consultation de ses axes et indicateurs. Cette approche nécessite que soit mis en place une « couche métier » d'accès aux données. Cette couche permet de présenter à l'utilisateur une approche fonctionnelle des données et de leur relation, affranchissant ce dernier de la complexité des bases de données et du langage SQL. Concrètement, les données sont regroupées au sein de thématiques métiers et reliées entre elles selon ce même logique métier. L'utilisateur n'a plus qu'à sélectionner ses données et les disposer dans son rapport. La encore, la volonté de l'utilisateur se traduit en une sollicitation de la base de données avec une performance d'affichage variable selon la complexité de la demande.

III. Architecture du système décisionnel

1. Source des données :

Les données contenues dans les systèmes opérationnels sont :

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

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

La mise en œuvre du Data-Warehouse est un processus complexe. L'objectif à atteindre est de recomposer les données disponibles pour en donner :

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

Toutes les données qu'elles proviennent du système de production de l'entreprise vont devoir être organisées, coordonnées, intégrées et stockées, pour donner à l'utilisateur une vue intégrée et orientée métier.

Ainsi, le Data-Warehouse permet la mise en place d'un outil décisionnel s'appuyant sur les informations pertinentes pour l'entreprise, centrées sur le métier utilisateur.

2. Définition

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

- **Orientées sujet :** Le Data-Warehouse est organisé autour des sujets majeurs de l'entreprise, contrairement aux données des systèmes de production. Ceux-ci sont généralement organisés par processus fonctionnels. Les données sont structurées par thèmes.

L'intérêt de cette organisation est de disposer de l'ensemble des informations utiles sur un sujet le plus souvent transversal aux structures fonctionnelles et Organisationnelles de l'entreprise.

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

L'intégration dans une structure unique est indispensable car les informations communes à plusieurs sujets ne doivent pas être dupliquées. Dans la pratique, une structure supplémentaire appelée Data-Mart (magasin de données) peut être créée pour supporter l'orientation sujet.

- **Données intégrées :** Un Data-Warehouse est un projet d'entreprise. Par exemple, le même indicateur de chiffre d'affaires intéressera autant les forces de vente que le département financier ou les acheteurs. Pour y parvenir, les données doivent être intégrées.

Avant d'être intégrées dans le Data-Warehouse, les données doivent être mises en forme et unifiées afin d'avoir un état cohérent. Par exemple, la consolidation de l'ensemble des informations concernant un client donné est nécessaire pour donner une vue homogène de ce client.

Ainsi, une donnée doit avoir une description et un codage unique.

- **Données historisées :** Dans un système de production, la donnée est mise à jour à chaque nouvelle transaction. Dans un Data-Warehouse, la donnée ne doit jamais être mise à jour. Un référentiel temps doit être associé à la donnée afin d'être capable d'identifier une valeur particulière dans le temps.
- **Données non volatiles :** La non-volatilité des données est en fait une conséquence de l'historisation. Une même requête effectuée à quelques mois d'intervalle en précisant la date de référence de l'information recherchée donnera le même résultat.

Le tableau suivant présente les principales différences entre le système relationnel et le Data-Warehouse ou le système décisionnel:

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

Tableau 2 : différences entre système opérationnel et Data-Warehouse

3. La structure

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

- **Les données détaillées :** Elles reflètent les événements les plus récents. Les intégrations régulières des données issues des systèmes de production vont habituellement être réalisées à ce niveau. Les volumes à traiter sont plus importants que ceux gérés en transactionnel.

Il est à noter que le niveau de détail géré dans le Data-Warehouse n'est pas forcément identique au niveau de détail géré dans les systèmes opérationnels. La donnée insérée dans le Data-Warehouse peut être déjà une agrégation ou une simplification d'informations tirées du système de production.

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

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

- **Les méta-données :** Elles regroupent l'ensemble des informations concernant le Data-Warehouse et les processus associés. Elles constituent une

véritable aide en ligne permettant de connaître l'information contenue dans le Data-Warehouse. Elles sont idéalement intégrées dans un référentiel.

Les principales informations sont destinées :

- à l'utilisateur (sémantique, localisation) ;
- aux équipes responsables des processus de transformation des données du système de production vers le Data-Warehouse (localisation dans les systèmes de production, description des règles, processus de transformation) ;
- aux équipes responsables des processus de création des données agrégées à partir des données détaillées ;
- aux équipes d'administration de la base de données (structure de la base implémentant le Data-Warehouse) ;
- Aux équipes de production (procédures de changement, historique de mise à jour,...).

- **Les données historisées** : Un des objectifs du Data-Warehouse est de conserver en ligne les données historisées. Chaque nouvelle insertion de données provenant du système de production ne détruit pas les anciennes valeurs, mais crée une nouvelle occurrence de la donnée. Le support de stockage dépend du volume des données, de la fréquence d'accès, du type d'accès.

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

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

4. La construction du DataWarehouse

L'objectif à atteindre est de recomposer les données disponibles pour en obtenir une vision intégrée et transversale aux différentes fonctions de l'entreprise, une vision métier au travers des différents axes d'analyses et une vision agrégée ou détaillée, adaptée à son besoin.

La transformation des données en connaissance est un processus complexe.

Le cadre général d'un Data-Warehouse comporte trois domaines principaux :

- les applications
 - les composants fonctionnels du Data-Warehouse (acquisition, stockage, accès)
 - les infrastructures (technique et opérationnelle).
- **Les applications** : Un Data-Warehouse se fait en plusieurs itérations. Chaque sujet traité est décomposé en un ensemble d'initiatives (projet décisionnel entrant en jeu dans la construction d'un Data-Warehouse dans une démarche itérative).

Les applications doivent rester maîtrisables et fournir des résultats tangibles dans un délai de moins de six mois, qui correspond au délai moyen de réalisation d'applications.

Le périmètre de chaque application doit être clairement défini (acteurs concernés, fréquences et périodicité des analyses, ...).

➤ **Les composants fonctionnels** : Trois composants caractérisent un Data-Warehouse :

✚ **L'acquisition** : Elle se compose de trois phases : l'extraction, la préparation et le chargement.

- L'extraction : collecter les données utiles dans le système de production.

- La préparation : plusieurs technologies sont utilisables :

Les passerelles, fournies par les éditeurs de base de données,
Les utilitaires de réplication, utilisables si les systèmes de production et décisionnel sont homogènes,

Les outils spécifiques d'extraction (prix élevé).

La préparation inclut la mise en correspondance des formats de données, le nettoyage, la transformation et l'agrégation.

- Le chargement : il constitue la dernière phase d'alimentation. Il est indispensable de maîtriser la structure du système de gestion de base de données - SGBD -(tables et index) pour optimiser au mieux le processus.

✚ **Le stockage** : Le composant de base est le SGBD. Il doit être spécifiquement adapté aux caractéristiques de l'accès décisionnel. Du fait de l'importance de l'historique, la structuration physique des données est également très importante.

Le SGBD apporte la transparence à l'évolution matérielle, l'indépendance, que ce soit au niveau des types et du nombre de processeurs, des disques ou des mémoires, ainsi que la transparence à l'évolution des systèmes d'exploitation.

✦ **L'accès** : Définir une architecture globale servant de support aux accès décisionnels impose des choix technologiques non structurants. Il faudra mettre en place une infrastructure commune à toutes les applications décisionnelles, tout en laissant aux utilisateurs l'opportunité d'utiliser les solutions d'accès les mieux adaptées à leur problématique.

➤ **Les infrastructures** : Il y a deux niveaux d'infrastructures :

L'infrastructure technique : l'ensemble des composants matériels et logiciels

L'infrastructure opérationnelle : l'ensemble des procédures et des services pour administrer les données, gérer les utilisateurs et exploiter le système.

Pour la construction d'un Data-Warehouse, de manière générale, il faut :

✦ Bien connaître les métiers utilisateurs :

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

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

✦ Commencer petit et voir grand

✦ Gérer l'évolutivité

✦ Alimentation et administration

5. La conception du DataWarehouse

Quatre caractéristiques ont des effets déterminants sur la démarche de conception d'un Data-Warehouse :

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

Dans tous les cas, il n'existe pas de cadre figé pour la conception d'un Data-Warehouse. Chaque entreprise doit adapter le projet à son contexte, en ne perdant pas les objectifs de vue. Cet objectif est de mettre en place un système d'informations cohérent et intégré, le système devant être décomposé en applications, chacune s'intégrant dans le Data-Warehouse.

Il est possible de proposer trois phases pour la conception :

- Définir le pourquoi du Data-Warehouse et les objectifs à atteindre (impliquer les utilisateurs)
- Définir l'infrastructure technique et organisationnelle du Data-Warehouse
- Mettre en oeuvre les applications.

✦ Découvrir et définir les initiatives :

Cette phase consiste en l'étude stratégique du Data-Warehouse et la définition du plan d'actions.

➤ L'étude stratégique :

Pendant l'étude stratégique, il faut :

- Informer et motiver les personnes concernées dans l'entreprise
- Impliquer les managers, les équipes opérationnelles, les équipes informatiques : phase d'identification et de compréhension des enjeux métier/entreprise
- Identifier les projets Data-Warehouse.

L'étude stratégique permet d'identifier la stratégie de l'entreprise, son organisation, les processus qu'elle met en œuvre, la culture de l'entreprise. Le but est de déterminer les domaines pour lesquels la mise en place d'un Data-Warehouse peut être le plus bénéfique.

A la fin de cette étude, des sous-projets ou initiatives vont être dégagés. Les acteurs consultés doivent être des spécialistes métiers, convaincus de la nécessité et de l'importance du Data-Warehouse.

➤ Le plan d'action

Pour mettre en place le plan d'action, il faut :

- Vérifier la faisabilité de chaque projet (s'assurer de l'existence et de la qualité des données, des possibilités techniques, des possibilités organisationnelles)
- Estimer les ressources pour chaque projet, les besoins
- Séquencer et planifier les projets.

Il est possible de commencer petit et voir grand. Il s'agit ainsi de limiter l'objectif à un domaine précis (Data-Mart). Ceci permet d'avoir un retour sur investissement rapide et de prouver le bien fondé du concept.

Il faut déterminer la faisabilité de chaque projet envisagé et déterminer les moyens nécessaires (techniques, humains, organisationnels) à leur réalisation.

Leur coût et leur durée de mise en œuvre seront estimés, ainsi que les bénéfices attendus et leur terme.

Il faut aussi évaluer le capital informatique, c'est-à-dire analyser les systèmes de production sur lesquels le Data-Warehouse doit s'appuyer.

✦ **L'infrastructure**

Il s'agit de déterminer l'infrastructure technologique et organisationnelle nécessaire à la mise en place du Data-Warehouse et à la conduite du changement.

➤ **L'infrastructure technique**

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

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

Enfin, il faut vérifier que les solutions choisies fonctionnent entre elles.

➤ **L'infrastructure organisationnelle**

Parallèlement aux choix technologiques, il faut :

- Déterminer la logistique et l'organisation nécessaires à la concrétisation des initiatives

- Répartir les tâches entre les équipes de développement et les équipes d'exploitation : déterminer l'alimentation du Data-Warehouse, l'administration
- Déterminer les flux d'information entre le Data-Warehouse et les utilisateurs.

✚ La formation

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

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

✚ La mise en œuvre des applications

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

- La spécification
- La conception
- La mise en œuvre et l'intégration
- Le déploiement et la mise en place des accompagnements
- Les mesures.

Pendant l'étape de spécification, les différentes étapes des initiatives sont définies et planifiées de manière plus détaillées.

Il est recommandé de faire attention aux coûts cachés que peuvent entraîner les technologies informatiques.

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

Deux visions du Data-Warehouse cohabitent dans l'approche précédente :

- Une vision entreprise : chaque projet défini dans la première phase (initiative) est construit de manière indépendante et répond à un objectif métier délimité, tout en s'intégrant dans le Data-Warehouse.
- Une vision projet : les projets identifiés deviennent des applications. Donc le processus est itératif.

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

6. Le Data-Mart

Avec un Data-Warehouse, il y a des risques d'échec. Rien n'invite l'utilisateur à se servir d'un Data-Warehouse. Le succès d'un Data-Warehouse dépend donc uniquement de son effective utilisation. Un des gros risques de la construction est de se cristalliser autour de la problématique informatique et de se détourner de l'utilisateur. Le Data-Mart minimise la complexité informatique. Il est donc plus facile de se concentrer sur les besoins utilisateurs.

➤ Définition

Le Data-Mart est une base de données moins coûteuse que le Data-Warehouse, et plus légère puisque destinée à quelques utilisateurs d'un département. Il séduit plus que le Data-Warehouse les candidats au décisionnel.

C'est une petite structure très ciblée et pilotée par les besoins utilisateurs. Il a la même vocation que le Data-Warehouse (fournir une architecture décisionnelle), mais vise une problématique précise avec un nombre d'utilisateurs plus restreint. En général, c'est une petite base de données (SQL ou multidimensionnelle) avec quelques outils, et alimentée par un nombre assez restreint de sources de données. Mais pour réussir, il y a quelques précautions à prendre, gage de son évolutivité vers le Data-Warehouse.

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

Tableau 3 : finalités des Data-Mart et Data-Warehouse

➤ **Mise en place**

Construire un ou plusieurs Data-Mart départementaux au lieu d'un Data-Warehouse central permet de valider rapidement le concept d'informatique décisionnelle.

Mais construire des Data-Mart n'est pas sans risques. En effet, dans les entreprises, des Data-Mart isolés peuvent proliférer. Ces entreprises risquent de retomber dans le piège d'une architecture composée de multiples systèmes décisionnels incohérents, contenant des informations redondantes. Cela coûte plus cher et c'est plus complexe à gérer qu'un Data-Warehouse centralisé.

Les Data-Mart résolvent les problèmes de performance des gros Data-Warehouse. Mais ils font régresser vers le vieux problème des îlots isolés. Les entreprises vont devoir affronter des problèmes techniques complexes et coûteux pour remettre en cohérence les ensembles.

Fédérer des Data-Mart ou les faire évoluer vers une structure centralisée n'est pas facile.

On peut se poser la question s'il est préférable de bâtir un gros et unique Data-Warehouse ou bien de concevoir un réservoir plus modeste, nourri par les données d'un seul département.

Il est intéressant de commencer par un Data-Mart, à condition de respecter certaines règles :

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

Donc le Data-Mart peut préparer au Data-Warehouse. Mais il faut penser grand, avenir, et adopter des technologies capables d'évoluer.

7. Le concept OLAP

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

Par exemple, un utilisateur peut, en vue multidimensionnelle, analyser les ventes par produit mais aussi par région ou par période. Ces modèles permettent des manipulations simples : rotation, pivot ou vues par tranche, analyse de type permutations d'axes (slice and dice) ou en cascade (drill-down).

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

➤ Les 12 Règles OLAP de E.F Codd

OLAP est le terme pour décrire l'approche dimensionnelle de l'aide à la décision. Tout comme OLTP, OLAP a été proposé par E. F. Codd. Cette philosophie comprend douze critères qui représentent l'étalon de mesure servant à comparer les systèmes d'aide à la décision, à ces douze critères, 6 ont été ajoutés en 1995. Il faut noter que ces six critères supplémentaires sont rarement cités et utilisés .

✓ 1 - Vue multidimensionnelle

Permet d'avoir une vision multidimensionnelle des données.

✓ 2 – Transparence du serveur OLAP à différents types de logiciels

L'utilisateur ne doit pas se rendre compte de la provenance des données si celles-ci proviennent de sources hétérogènes; ces sources peuvent être un fichier Excel, une base de données de production ou même un fichier texte.

✓ **3 – Accessibilité à de nombreuses sources de données**

OLAP est décrit comme un middleware qui se place entre les sources de données hétérogènes et un front-end (sous la forme d'un Data-Warehouse). Il doit donner accès aux données nécessaires aux analyses demandées afin de présenter à l'utilisateur une vue simple et cohérente. Ils doivent aussi savoir de quel type de systèmes proviennent les données.

✓ **4 - Performance du système de Reporting**

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

✓ **5 - Architecture Client/Serveur**

Il est essentiel que le produit soit en Client-Serveur mais aussi que les composants serveurs d'un produit OLAP intègrent facilement ses différents clients.

✓ **6 - Dimensions Génériques**

Chaque dimension doit être équivalente par rapport à sa structure et à ses capacités opérationnelles.

✓ **7 - Gestion dynamique des matrices creuses**

Le système OLAP ajuste automatiquement son schéma physique pour s'adapter au type du modèle et au volume des données (plus on dispose de place plus on peut agréger).

✓ **8 - Support Multi-Utilisateurs**

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

✓ **9 - Calculs à travers les dimensions**

Les calculs doivent être possibles à travers toutes les dimensions (les agrégats

doivent être faits dans toutes les dimensions).

✓ **10 - Manipulation intuitive des données**

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

✓ **11 - Souplesse et facilité de constitution des rapports**

Lors de la création de rapport, les dimensions peuvent être présentées de n'importe quelle manière.

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

Dimensions et niveaux d'agrégations illimités.

8. Systèmes OLAP versus systèmes OLTP:

Les bases de données sont utilisées dans les entreprises pour organiser les importants volumes d'informations contenus dans leurs systèmes opérationnels.

Ces données sont gérées selon des processus transactionnels en ligne (OLTP : "On-Line Transactional Processing").

L'exploitation de l'information contenue dans ces systèmes opérationnels est devenue une préoccupation essentielle pour les dirigeants qui sont appelés à prendre des décisions par une meilleure connaissance de leur activité, et par conséquent les entreprises sont donc à la recherche de systèmes supportant efficacement les applications d'aide à la décision. Ces applications décisionnelles utilisent des processus d'analyse en ligne de données (OLAP : "On-Line Analytical Processing". Ces processus répondent aux besoins spécifiques des analyses d'information

	Processus OLTP	Processus OLAP
Données	<p>Exhaustives</p> <p>Courantes</p> <p>Dynamiques</p> <p>Orientées applications</p>	<p>Résumées</p> <p>Historiques</p> <p>Statiques</p> <p>Orientées sujets (d'analyse)</p>
Utilisateurs	<p>Nombreux</p> <p>Variés (employés, directeurs,...)</p> <p>Concurrents</p> <p>Mises à jour et Interrogations</p> <p>Requêtes prédéfinies</p> <p>Réponses immédiates</p> <p>Accès à peu d'information</p>	<p>Peu nombreux</p> <p>Uniquement les décideurs</p> <p>Non concurrents</p> <p>Interrogations</p> <p>Requêtes imprévisibles et complexes</p> <p>Réponses moins rapides</p> <p>Accès à de nombreuses informations</p>
Administration des données	<p>Forte disponibilité</p> <p>Sauvegardes fréquentes</p> <p>Peu de maintenance off-line</p>	<p>Disponibilité faible</p> <p>Sauvegardes peu fréquentes mais très volumineuses</p> <p>Beaucoup de maintenance mais en off-line</p>

Tableau 4: Comparaison des processus OLTP et OLAP.

IV. La modélisation multidimensionnelle

La modélisation dimensionnelle consiste à présenter les données sous la forme d'un cube, elle dérive des concepts OLAP (On Line Analytical Processing). le modèle multidimensionnel a été introduit par Ralph Kimball. Il consiste en deux nouveaux concepts fondamentaux: les faits et les dimensions. Le modèle multidimensionnel est la structure de données la plus utilisée et la plus appropriée aux requêtes et analyses des utilisateurs d'entrepôts de données. Un fait représente un sujet d'analyse, caractérisé par une ou plusieurs mesures. Ce fait est analysé selon des axes d'observation qui constituent également ses descripteurs. Ces dimensions peuvent présenter des hiérarchies qui offrent la possibilité de réaliser des analyses à différents niveaux de granularité (niveaux de détail). Un entrepôt de données présente alors une modélisation dite « Multidimensionnelle » puisqu'elle répond à l'objectif d'analyser des faits en fonction de dimensions qui constituent les différents axes d'observation des mesures.

1. Concept de Cube:

Le cube de données offre une abstraction très proche de la façon dont l'analyse voit et interroge les données. Il organise les données en une ou plusieurs dimensions qui déterminent une mesure d'intérêt. Une dimension spécifie la manière dont on regarde les données pour les analyser, alors qu'une mesure est un objet d'analyse. Chaque dimension est formée par un ensemble d'attributs et chaque attribut peut prendre différentes valeurs. Les dimensions possèdent en général des hiérarchies associées qui organisent les attributs à différents niveaux pour observer les données à différentes granularités. Une dimension peut avoir plusieurs hiérarchies associées, chacune spécifiant différentes relations d'ordre entre ses attributs

2. Concept de Faits:

Le **fait** modélise le sujet de l'analyse. Un fait est formé de **mesures** correspondant aux informations de l'activité analysée.

Les mesures d'un fait sont numériques et généralement valorisées de manière continue. Les mesures sont numériques pour permettre de résumer un grand nombre d'enregistrements en quelques enregistrements (on peut les additionner, les dénombrer ou bien calculer le minimum, le maximum ou la moyenne). Les mesures sont valorisées de façon continue car il est important de ne pas valoriser le fait avec des valeurs nulles. Elles sont aussi souvent :

- Additives : C'est-à-dire que l'opérateur d'agrégation (représenté par **SUM**) peut être appliqué pour regrouper les valeurs des mesures au cours de toutes les dimensions (exemple : Quantités vendus, Chiffre d'affaire ...).
- Semi-additives : Additives selon quelques dimensions (exemple : Nombre de transactions d'un client).
- Non-additives : C'est-à-dire que l'opérateur d'agrégation (représenté par **SUM**) n'est pas applicable sur aucune dimension (exemple : un ratio de gestion).

3. Concept de Dimension:

Le sujet analysé, c'est à dire le fait, est analysé suivant différentes perspectives. Ces perspectives correspondent à une catégorie utilisée pour caractériser les mesures d'activité analysées, on parle de dimensions.

Une dimension modélise une perspective de l'analyse. Une dimension se compose de paramètres correspondant aux informations faisant varier les mesures de l'activité. Les dimensions servent à enregistrer les valeurs pour lesquelles sont analysées les mesures de l'activité. Une dimension est généralement formée de paramètres (Ou attributs) textuels et discrets. Les paramètres textuels sont utilisés pour restreindre la portée des requêtes afin de limiter la taille des réponses. Les paramètres sont discrets,

c'est à dire que les valeurs possibles sont bien déterminées et sont des descripteurs constants.

Une dimension peut se diviser en plusieurs niveaux, pour modéliser des phénomènes à différentes échelles ou niveaux de détail. Par exemple, une dimension représentant un découpage administratif peut se diviser en niveaux «Pays», «Province», « Région économique » et « Division de recensement ».

4. Concept d'hierarchie :

Certaines dimensions peuvent comporter plus d'un ensemble de niveaux permettant de décrire un thème d'analyse. Par exemple, on peut décrire le temps selon un découpage jour-semaine ou encore jour-mois-année. On parle alors de hiérarchies multiples ou alternatives. Une dimension peut ainsi comporter plus d'une hiérarchie, offrant des chemins différents pour remonter vers les membres de niveau plus général. En règle générale, le niveau le plus détaillé est commun à toutes les hiérarchies à l'intérieur d'une dimension, puisque ce sont les membres de ce niveau qui définissent le grain de la dimension et par conséquent l'adressage des faits détaillés.

5. Concept de Granularité :

La granularité est le niveau de détail décrivant un fait. La granularité détermine le volume des données ainsi que le type des requêtes que l'utilisateur peut poser.

Prenons l'exemple d'une compagnie de téléphonie cité par (4). Il existe deux systèmes différents ; un premier système avec le détail de tous les appels, un deuxième système avec un ré-sumé mensuel de tous les appels.

Considérons la requête « Quelle est la personne qui a fait le plus d'appels? ». On peut obtenir une réponse dans le premier système mais non dans le deuxième système.

D'un autre côté la requête, « Quelle est la moyenne des appels longue distance ? », a une réponse dans les deux systèmes. Dans le premier système, il faudra calculer la moyenne de plusieurs enregistrements. Dans le deuxième système il suffit d'accéder à un seul enregistrement. Cet exemple montre qu'il y a un compromis à faire entre la

per-formance et le stockage des données. Il est ainsi intéressant de garder les données détail-lées dans le cas où il n'y aurait pas de problèmes de stockage et d'avoir des données agrégées pour assurer une meilleure performance. Quel que soit le niveau de granularité choisie, les données des différentes hiérarchies nécessitent une agrégation pour pouvoir répondre aux questions des utilisateurs. L'agrégation dans les outils OLAP assure la performance.

6. Les Schémas Multidimensionnels:

D'une manière très générale, le modèle dimensionnel possède une table de faits centrale et des tables de dimensions situées autour. Chaque enregistrement de la table de faits stocke les clés des tables de dimensions et les mesures faites à un instant précis. La taille de la table de faits peut atteindre plusieurs millions d'enregistrements, et peut nécessiter plusieurs giga, voire tera octet d'occupation sur disque.

- **Schéma en Etoile :**

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

Cette structure est constituée du fait central et des dimensions. Ce modèle représente visuellement une étoile, on parle de modèle en étoile

- La table des faits contient des mesures (par exemple le prix des produits vendus, la quantité de produits vendus) qui peut être agrégée de diverses façons.
- Les tables de dimension fournissent la base pour l'agrégation des mesures de la table de fait.
- La table de fait est liée avec toutes les tables de dimension par des relations "Un-à-Plusieurs".
- La clef primaire de la table de fait est la concaténation des clefs primaires de toutes les tables de dimension.

-L'avantage d'employer les schémas en étoile pour représenter les données est la réduction du nombre de tables, et le nombre de relations entre ces tables et donc le nombre de jointures exigé dans les requêtes des utilisateurs

- **Schéma en Flocon :**

Dans une étoile, les dimensions sont dénormalisées : les hiérarchies d'agrégation sont implicites. Un flocon de neige est une étoile où les hiérarchies sont respectées et explicites.

Un flocon peut être formé depuis une étoile en normalisant les dimensions. Le flocon diffère de l'étoile de par sa constitution : un flocon reste relativement normalisé alors qu'une étoile est dénormalisée.

Une modélisation en flocon consiste à décomposer les dimensions du modèle en étoile en sous hiérarchies. La modélisation en flocon est donc une émanation de la modélisation en étoile, le fait est conservé et les dimensions sont éclatées conformément à sa hiérarchie des paramètres.

- **Schéma en Constellation :**

Le schéma en constellation consiste à placer plusieurs schémas en étoile avec des tables de faits reliées hiérarchiquement. Les liens entre les différentes tables de faits permettent de visionner les différents niveaux de détail. Donc ce schéma consiste à fusionner plusieurs modèles en étoile qui utilisent des dimensions Communes

- **Schéma en Grappe:**

Ce schéma est apparu car il n'existe pas de schéma en étoile ou de schéma en flocon parfait. Le schéma en grappe est une dérivation de

ces deux schémas pour en former un troisième. Kimball déclare qu'un schéma en flocon n'est pas optimal, car il est trop complexe. Toujours pour les mêmes raisons de simplifications des tables, afin de pouvoir trouver facilement les informations dans l'entrepôt, le schéma en grappe apparaît alors comme un compromis entre le schéma en étoile et le schéma en flocon.

7. Complexité Vs Redondance :

Tous ces modèles sont des modèles dimensionnels, que l'on peut ou non implémenter lors de la création d'un entrepôt de données. Cependant, il faut savoir que ces modèles augmentent soit la complexité des tables, soit la redondance des données. Chaque cas est unique, et on doit bien réfléchir au schéma que l'on va adopter.

8. Les implémentations des modèles multidimensionnels :

Selon la façon dont le cube de données est stocké, il existe deux approches fondamentales pour construire des systèmes basés sur un modèle multidimensionnel et une troisième Hybride. L'approche MOLAP (Multidimensionnel OLAP) implémente le cube de données dans un tableau multidimensionnel (multidimensionnelle array). Par contre, l'approche ROLAP (Relationnel OLAP) utilise un SGBD relationnel pour gérer et stocker le cube de données

- **Les systèmes MOLAP :**

Les systèmes de type MOLAP stockent les données dans un SGBD multidimensionnel sous la forme d'un tableau multidimensionnel (Multidimensionnelle array). Chaque dimension de ce tableau est associée à une dimension du cube. Seules les valeurs de données correspondant aux données de chaque cellule sont stockées. Ces systèmes demandent un pré calcul de toutes les agrégations possibles.

En conséquence, ils sont plus performants que les systèmes traditionnels, mais difficiles à mettre à jour et à gérer.

Les systèmes MOLAP apparaissent comme une solution acceptable pour le stockage et l'analyse d'un entrepôt lorsque la quantité estimée des données d'un entrepôt ne dépasse pas quelques giga-octets et lorsque le modèle multidimensionnel évolue peu. Mais, lorsque les données sont éparses, ces systèmes sont consommateurs d'espace et des techniques de compression doivent être utilisées.

- **Les systèmes ROLAP :**

Les systèmes de type ROLAP utilisent un SGBD relationnel pour stocker les données de l'entrepôt. Ils représentent une interface multidimensionnelle pour le SGBD relationnel. Le moteur OLAP est un élément supplémentaire qui fournit une vision multidimensionnelle de l'entrepôt, des calculs de données dérivées et des agrégations à différents niveaux. Il est aussi responsable de la génération des requêtes SQL mieux adaptées au schéma relationnel. Les mesures (par exemple les quantités vendues) sont stockées dans une table qu'on appelle la table des faits. Pour chaque dimension du modèle multidimensionnel, il existe une table qu'on appelle la table de dimension (comme Produit, Temps, Client) avec tous les niveaux d'agrégation et les propriétés de chaque niveau.

Ces systèmes peuvent stocker de grands volumes de données, mais ils peuvent présenter un temps de réponse élevé. Les principaux avantages de ces systèmes sont :

- Une facilité d'intégration dans les SGBDs relationnels existants.
- Une bonne efficacité pour stocker les données multidimensionnelles.

- **Les systèmes HOLAP :**

Les données des cubes sont stockées dans une structure relationnelle et les agrégats dans une structure multidimensionnelle. L'utilisateur

formule une requête sur la base de données relationnelle et le résultat s'affiche sous une forme multidimensionnelle. La convivialité et ainsi respecté tout en permettant le stockage de grand volume de données.

9. Approche de conception multidimensionnelle :

La Conception d'un schéma conceptuel selon l'approche **RALPH KIMBALL** qui Considéré comme étant la référence dans le monde de la conception des entrepôts de données .la méthode de conception d'une base de données multidimensionnelles suivant la démarche **KIMBALL** est abordée d'une manière systématique, en envisageant quatre étapes dans un ordre bien défini :

- **Première étape : Choisir le processus d'activité à modéliser.**

Un processus d'activité est un processus opérationnel important pour l'organisation, relié par une application existante à partir de laquelle des données peuvent être collectées au profit de l'entrepôt de données. Exemples de processus d'activité : Facturation, Ventes....

- **Deuxième étape : Choisir le Grain du processus d'activité.**

Le Grain est le niveau de détail fondamental, atomique, des données figurant dans la table des faits pour ce processus. Des grains typiques sont des transactions individuelles, des récapitulatifs individuelles quotidiennes.

- **Troisième étape : Choisir les dimensions applicables a chaque enregistrement de la table de faits.**

Des dimensions typiques sont le temps, le produit, le magasin, le client. Le choix d'une dimension s'accompagne de la définition de tous les attributs textuels qui garniront la table de dimension

- **Quatrième étape : Choisir les faits mesures que contiendra chaque enregistrement de la table de faits.**

Des faits mesures typiques sont des quantités numériques additives.

La construction d'un entrepôt de donnée multidimensionnel selon **KIMBALL** revient à faire correspondre les besoins de la communauté utilisateurs avec la réalité des informations disponibles. Les neuf décisions majeures qui jalonnent la conception de la base de données sont toutes subordonnées aux besoins des utilisateurs et aux réalités que représentent les données.

- **Les neuf décisions :**

Les neuf décisions à prendre pour la conception complète d'un entrepôt de données dimensionnel portent sur les points suivants:

1. Les processus, et partant, l'identité des tables de faits.
2. Le grain de chaque table de fait.
3. Les dimensions de chaque table de faits.
4. Les faits, y compris les faits précalculés.
5. Les attributs des dimensions, avec des descriptions complètes et la terminologie adéquate.
6. Comment suivre les dimensions à évolution lente.
7. Les agrégats, les dimensions hétérogènes, les mini-dimensions, les modes de requêtes et autres décisions sur le stockage physique.
8. L'étendue historique de la base de données.
9. L'urgence avec laquelle les données doivent être extraites et chargées dans l'entrepôt de données.

Cette méthodologie est une méthodologie descendante, qui commence par identifier les processus majeurs de l'entreprise dans la quelle les informations sont collectées. Car les concepteurs d'entrepôt de donnée d'après KIMBALL doivent commencer avec des sources des données existantes et réellement utilisées, pour éviter de perdre du temps en rêvant à des sources d'information qui n'existent pas.

Une fois les processus identifiés, une ou plusieurs tables de faits sont construites à partir de chacun des processus choisis. Avant de pouvoir concevoir en détail une table de faits, une décision doit être prise sur la signification précise d'un enregistrement de plus bas niveau dans la table de faits. Cette signification est le grain de la table de faits. Lorsque le grain de la table de faits est connu, les dimensions et leurs grains respectifs peuvent être identifiés. Il y aura des dimensions supplémentaires qui ne seront pas strictement requises pour décider du grain de la table de faits.

Le choix des dimensions est le point clé de la conception, une fois les dimensions choisies, il faut définir toutes les mesures de la table de faits, on peut ensuite définir complètement le contenu des enregistrements de dimension.

A ce stade la conception de la structure logique principale est terminée, et nous pouvons porter notre attention sur les aspects les plus généraux de la structure physique.

V. Implémentation du projet décisionnel

1. Problématique

Le deuxième axe de ce Mémoire consiste à établir un socle méthodologique architectural et technique pour le développement d'une application décisionnelle. Cette partie à tout d'abord nécessité un travail de documentation important dont le résultat est synthétisé dans les chapitres précédents.

Le but de ce chapitre est non seulement de clarifier les concepts manipulés dans la suite de l'étude de cas mais aussi d'explicitier certains choix architecturaux.

Pour commencer nous allons tout d'abord faire une petite présentation de la société Sonelgaz, société qui est le sujet de notre travail, ensuite nous allons décrire l'enchaînement logique de notre travail au sein des services des ressources humaines de la filiale de Blida de la compagnie « Sonelgaz ».

2. Présentation de Sonelgaz

Sonelgaz (Société Nationale de L'Electricité et du Gaz) est une compagnie chargée de la production, du transport et de la distribution de l'électricité et du gaz en Algérie. Elle a été créée en 1969 en remplacement de l'unité précédente : Electricité et Gaz d'Algérie (EGA) et on lui a donné un monopole de la distribution et de la vente de gaz naturel dans le pays, de même la production, la distribution, l'importation et l'exportation de l'électricité.

En 2002 le décret présidentiel N02-195 la convertit en une société par action SPA entièrement détenue par l'état, on parle maintenant de groupe Sonelgaz .

En 2003, elle produisait 29 milliards de Kwh par an, vendait 4.6 milliards de mètres cube de gaz par an .En 2006 elle employait environ 28000 personnes.

3. Objectifs et état des lieux

Dans le cadre de l'achèvement de notre Mémoire nous avons été retenus pour faire un stage pratique au sein du service de ressources humaines de l'agence de Blida de la compagnie « Sonelgaz » .

Présents sur les lieux du stage ou nous étions bien accueillis et encadrés, il était question tout au début de fixer un but commun avec le service des ressources humaines, but qui devrait bien sur corroborer notre objectif principal qui est d'élaborer une application informatique décisionnelle pour couronner le volet théorique que nous avons acquis et développer au fur et à mesure de notre cursus universitaire et grâce à la documentation appropriée et la recherche.

Après beaucoup de réunions et de concertations avec le personnel du service concerné, il était aussi question d'échanges productifs ou nous avons expliqués les notions de base du décisionnel et tout le bien sur le niveau logistique qu'il peut leur apporter pour notamment plus de faciliter et de bon sens dans leur travail qui consistait à gérer les informations des employés. En contre partie nous étions à l'écoute des contraintes soulevées par le personnel pour mieux aviser notre étude de cas, et ce fut le cas jusqu'au moment ou avec ce personnel nous avons enfin fixer un objectif principal dont le titre était « L'évaluation des employés » selon des critères précis que nous avons fixer ensemble mais surtout avec eux et selon aussi des axes fixes qui regroupent les informations professionnelles et personnelles des employés. L'objectif étant fixer nous pouvions enfin passer à l'évaluation de l'état des lieux et quelle fut notre surprise de ne pas trouver une base de donnée au vrai sens du terme pour gérer les informations des employés mais juste des fichiers pats de type Excel et l'argument avancé par le personnel des ressources humaines que le logiciel Excel permettait amplement de subvenir au besoin logistique de ces derniers du fait du nombre raisonnable des employés et restreint des informations qui ne nécessitaient pas l'utilisation d'une base de donnée .

Mais c'était sans compter sur notre abnégation et déterminisme que nous avons décidé coûte que coûte de relever les défis de déployer une base de données de type décisionnel pour le profit du service des ressources humaines de cette agence en l'occurrence et qui pourrait leur être d'un secours précieux.

4. Périmètres retenus

Une fois l'objectif principal fixé il nous fallait retenir les périmètres des axes sur lesquels nous allions évaluer les employés et c'était bien sur avec l'étroite collaboration du personnel du service des ressources humaines en étant toujours à leur écoute ,ce qui nous a permis de dégager trois grands axes d'évaluation comportant en leur sein des point bien précis convergeant tous vers une seule finalité qui coulait dans l'objectif principal de la base des donnée.

Ces trois axes et leurs points détaillés sont :

➤ **L'habilité d'encadrement**

1-planification et organisation : Propose, détermine des objectifs en terme de résultats à atteindre reliés aux activités à réaliser. Elabore des plans d'actions en établissant des priorités et des échéanciers de réalisation. Prévoit les modes et les périodicités de contrôle. Emploie des outils pour l'utilisation rationnelle du temps (tableau de bord, planification). Prévoit les solutions de rechange en anticipant les sources de problèmes entravant la performance. Définir et répartit les rôles et taches de ses collaborateurs, selon leur responsabilités er compétences.

2-Décision : Fait face aux situations et ne les évites pas. Prend connaissance de toute information relative aux situations. S'assure de l'exactitude des faits (recoupements, déplacement...) Procède, si nécessaire aux consultations de parties concernées (collaborateurs, hiérarchies...) quant aux mesures à prendre. Prend des mesures concrètes et donne les instructions nécessaires. Accepte et assume les conséquences de ses décisions.

3-Contrôle : Etablit un échéancier de contrôle. Fait des suivis périodiques pour s'assurer de l'état d'avancement des travaux. Identifie et analyse des écarts au plan des réalisations par rapport aux objectifs fixés. Evalue le rendement de ses collaborateurs. Introduit les ajustements nécessaires. Produit et diffuse les bilans, comptes rendus et différentes pièces attestant des actes de contrôle.

4-Délégation : Se libère des tâches relevant des collaborateurs. Délègue les responsabilités aux collaborateurs, en fonction de leurs compétences. Spécifie les produits ou résultats attendus ainsi que échéances. Transmet les informations pertinentes à ses collaborateurs. Assiste des collaborateurs. Confie à ses collaborateurs des missions de représentation. Fait de sa satisfaction suite à la prise d'initiatives positives.

5-Gestion du développement des équipes et personnes : Offre son soutien et son assistance. Donne l'occasion à ses collaborateurs d'assumer des responsabilités. Met à la disposition de ses collaborateurs toute information utile à l'accomplissement de leurs activités. Contribue à l'intégration des nouvelles recrues. Favorise la participation des collaborateurs à toute action de développement des compétences.

➤ **Habiletés interpersonnelles**

1-Sens de la communication : Transmet des messages clairs et précis. Soigne la présentation matérielle de ses documents. S'assure de la compréhension de son message par son interlocuteur et reformule si nécessaire. Suscite de l'intérêt et l'attention à ses propos. Prend le temps d'écouter ses interlocuteurs.

2-Animation des équipes de travail : Aide l'équipe à définir ses objectifs. Propose et met en place des modalités de travail, favorisant les points d'information et de décision. Incite chacun des membres de l'équipe à prendre ses décisions. Amène l'équipe à se fixer des échéances en fonction de la contribution de chacun. Favorise la concertation et assure la coordination. Veille à la réduction des tensions et l'instauration d'un climat serein.

3-Relation avec l'environnement intérieur/extérieur : S'informe des moyens, besoins, présentations, développés par les structures de l'unité de l'entreprise ou les structures externes. Informe les structures environnantes de toutes ses activités ou besoins. Participe aux réunions de coordination inter structures. Intègre les orientations des structures hiérarchiques ou fonctionnelles. Rend compte à la hiérarchie de l'évolution de ses activités. Encourage le

développement d'une atmosphère de travail favorisant les échanges avec les autres structures de l'entreprise ou groupes de travail.

➤ **Habiletés personnelles**

1-Autonomie et sens de l'initiative : Etablit lui-même son propre programme d'activités. Recherche par lui-même les moyens d'atteindre les résultats. N'a pas recours à la hiérarchie systématiquement. Initie des actions en réponses à des situations urgentes, puis en informe les niveaux hiérarchiques. Introduit des ajustements au programme de travail en tenant compte de nouveaux facteurs. Fait part à sa hiérarchie des progrès obtenus et des contraintes à lever pour d'autres améliorations.

2-Dynamisme : S'engage pleinement dans son travail en termes de temps et d'implication. Démonstre beaucoup d'effort. Affiche sa disponibilité. Insuffle de l'énergie à ses collaborateurs (encouragement, aide à la relance...) Affiche de l'engouement et de la motivation. Favorise l'efficacité.

Et vers la fin il était aussi question de donner une note d'évaluation du rendement trimestriel qui devait poursuivre l'employé dans son dossier ou dans notre base de données.

Une fois arrivé à ce stade nous disposons de l'ensemble des informations des employés sous forme d'un fichier Excel et que nous avons déjà dégagés l'objectif principal et retenus les périmètres d'étude, nous pouvions ainsi passer au vif du sujet qui consistait à la modélisation dimensionnelle de notre entrepôt de données et que nous allions élaborer par la même occasion.

5. Modélisation dimensionnelle

La modélisation dimensionnelle consiste à présenter les données sous forme de cube constitué :

- De N dimension : matérialisant les axes d'analyses possibles.
- De données numériques : appelées faits.

Ainsi il suffit pour l'utilisateur de naviguer dans le cube selon les différents axes pour effectuer son analyse. Cette présentation de l'information est particulièrement intuitive puisque l'utilisateur manipule au travers des dimensions et des faits des notions bien connues de son domaine métier.

Un cube est implémenté à l'aide d'un modèle en étoile. Cette structure de données est constituée de tables de dimensions qui forment les branches de l'étoile et au centre d'une table de faits qui enregistre les événements qui se produisent à l'intersection de chaque dimension. La table de faits contient les clés des différentes dimensions ainsi que les valeurs qui correspondent aux indicateurs que l'on souhaite suivre. Les données de la table de faits sont numériques et dans les dimensions on retrouve tous les attributs le plus souvent textuels.

Dans notre cas d'étude nous avons opté pour une modélisation dimensionnelle en quatre étapes préconisée par l'approche de RALPH KIMBALL comme nous l'avons déjà explicité dans le volet théorique de ce Mémoire.

Cette approche se résume dans les quatre points suivant :

- Choisir les DataMarts
- Déclarer la granularité de la table de faits
- Définir les dimensions
- Définir les faits à mesurer

✚ Choisir les DataMarts :

Dans notre cas nous avons identifié un seul DataMart qui pourrait être agrandi dans le futur en un DataWarehouse mais nous avons opté pour la simplicité et la solidité du Le DataMart que nous avons identifié est « L'évaluation des employés » dont la source de données de ce dernier est essentiellement un fichier plat de type Excel.

✚ Déclarer la granularité de la table de faits :

Cette étape est essentielle dans la conception de l'entrepôt de données, sans avoir clairement défini le niveau de granularité de la table de faits, il est inutile d'aller plus loin dans la conception au risque d'avoir à tout refaire lorsque les utilisateurs découvriront le produit final.

Le but est de spécifier ce que représentera un enregistrement de la table de faits, dans notre cas d'étude notre table de faits permet d'évaluer chaque employé en un temps T déterminé à l'avance et selon son poste de travail et sa catégorie professionnelle et par rapport à ses habilités d'encadrement, interpersonnelles et personnelles.

✦ **Choisir des dimensions :**

Les dimensions sont élaborées en étroite collaboration avec le personnel du service des ressources humaines et elles reflètent à juste titre leurs attentes en matière d'analyse et entre autre elles répondent à leurs contraintes.

Et comme nous l'avons déjà expliqué dans les périmètres retenus quand nous avons identifié trois axes d'analyse, sur lesquels nous allons ajouter cette fois ci l'axe « Temps » qui est primordial et incontournable dans le décisionnel par soucis d'historisation, également nous allons ajouté sans aucune surprise l'axe « Employé » qui renferme les informations pertinentes sur les employés, ainsi pour finir le tableau nous avons ajoutés l'axes « Poste de travail », comme ça nous avons une vue exhaustive sur les principaux axes qui vont permettre d'analyser le cube voir même faire des requêtes qui vont aider à la prise de décision.

En résumé nous avons retenu ces sept dimensions :

- Dimension **Employé**
- Dimension **Temps**
- Dimension **Poste de travail**
- Dimension **habilités d'encadrement**
- Dimension **habilités interpersonnelles**
- Dimension **habilités personnelles**

✦ **Définition des faits à mesurer :**

Toujours avec le consentement du service des ressources humaines nous avons retenu un seul fait pertinent qui est l'évaluation du rendement de l'employé au sein de l'entreprise, ce fait sera une note numérique octroyée par le service des ressources humaines à l'égard des employés traduisant leur rendement dans le cadre professionnel, cette note sera graduée de zéro à dix.

6. Le schéma en étoile :

Ainsi notre DataMart bien défini avec sa table de faits et ces différentes dimensions, nous pouvons à présent illustrer les connexions entre ces différentes tables sous la forme d'un schéma en étoile caractéristique de l'approche simple et percutante préconisée par RALPH KIMBALL qui est la référence de notre démarche dans ce cas d'étude, le schéma est illustrée si dessous.

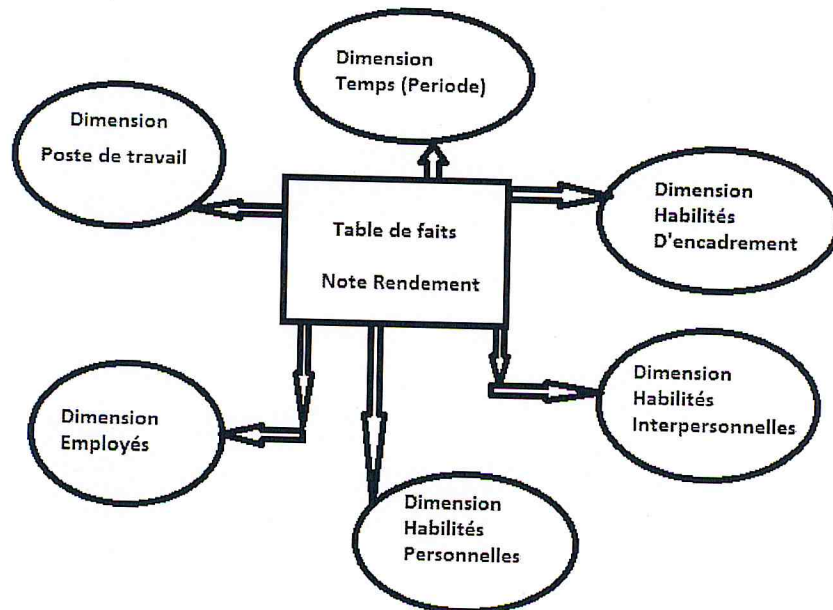


Figure 2: Schéma en étoile du DataMart

7. ETL (Extraction - Transformation – Chargement)

Source des données : dans notre cas d'étude, nous devions faire avec l'existant aussi maigre soit il, ainsi en partant d'un simple fichier plat de type Excel, en lui ajoutant d'autres fichiers que nous avons créés nous même sous la houlette du service des ressources humaines pour combler les lacunes dans l'optique de l'obtention d'un modèle dimensionnel solide et cela en suivant les étapes cités ci-dessous :

- **Extraction :** dans un premier temps les données sont extraites des fichiers plats de types Excel qui se trouvent être l'unique source de données en notre possession ce qui va bon gré mal gré restreindre notre marge de manœuvre car nous aurions souhaité obtenir d'autres sources de données pour enrichir notre DataMart, mais nous devions faire avec ce que nous avons sous les mains mais tant que faire alors bien faire sans regarder les obstacles voir même les surmonter.
- **Transformation :** une fois que les données sont extraites, elles allaient subir une série de traitements de manière à les rendre présentables pour les utilisateurs et dignes d'intérêt pour l'entreprise et les opérateurs, dans notre cas d'étude ça consistait à faire des vérifications du contenu des données pour éventuellement améliorer leurs qualités ainsi qu'à la gestion des redondances des informations, leurs cohérences avec la logique, leurs pertinences et enfin leurs bien fondés.
- **Chargement :** une fois les étapes précédentes effectuées avec succès, cette étape n'allait devenir qu'une formalité, il fallait juste assurer la concordance des informations à charger avec les diverses cibles qui allaient les accueillir dans le DataMart.

8. Le modèle logique

L'architecture proposée est une architecture décisionnelle classique avec un système source et un DataMart alimenté via un ETL à partir de ce système source.

Le DataMart est hébergé sur un serveur de type machine locale et à partir de cette machine nous pouvons requêter afin de permettre l'analyse de l'information et éventuellement la prise de décision.

Le schéma ci-dessous illustre le modèle logique de notre cas d'étude.

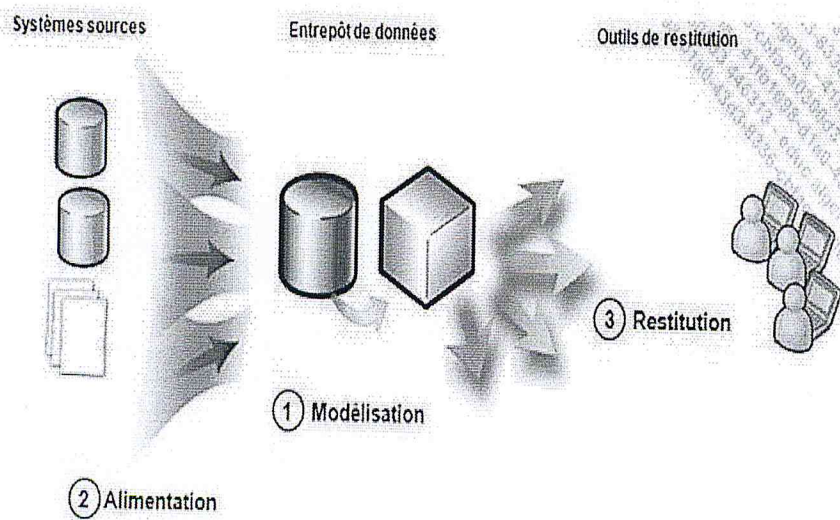


Figure 3: Schéma du model logique

9. Le modèle physique

Le déroulement des étapes de la conception (logique) selon l'approche de RALPH KIMBALL permet de passer à la conception physique et comme tout travail de conception, il est itératif et incrémental, de ce fait le modèle présenté dans la fin de cette partie est déjà considérablement enrichi par rapport au schéma du modèle en étoile, tout en gardant bien sur la même granularité qui est censée ne pas bouger car elle constitue le socle du modèle en étoile.

Mais avons de passer au schéma du model physique, nous devons expliciter la nature physique des données des tables de notre DataMart (faits et dimensions) ainsi que les différentes éventuelles hiérarchies dans les tables de dimensions comme ça notre schéma sera plus plausible et compréhensible.

➤ Table de faits « Evaluation Employé »

Attribut	Description	Type information
Période_FK	Clé étrangère vers la dimension période	Int
Employe_FK	Clé étrangère vers la dimension Employé	Int
Poste_FK	Clé étrangère vers la dimension Poste de travail	Int
Hab_Enca_FK	Clé étrangère vers la dimension Habilités d'encadrement	Int
Hab_InterPer_FK	Clé étrangère vers la dimension Habilités Interpersonnelles	Int
Hab_Perso_FK	Clé étrangère vers la dimension Habilités Personnelles	Int
NoteRendement	La note du Rendement de l'employé	Float

Tableau 5: La table FactEval

➤ **Table de la dimension « Période »**

Attribut	Description	Type information
Période_PK	Clé primaire	Int
Année	L'année de l'évaluation	Int
Trimestre	Trimestre de l'évaluation	Varchar(50)
Jour	Jour d'évaluation	Varchar(50)

Tableau 6: La table DimPeriode

• **Hiérarchie de la dimension « Période »**

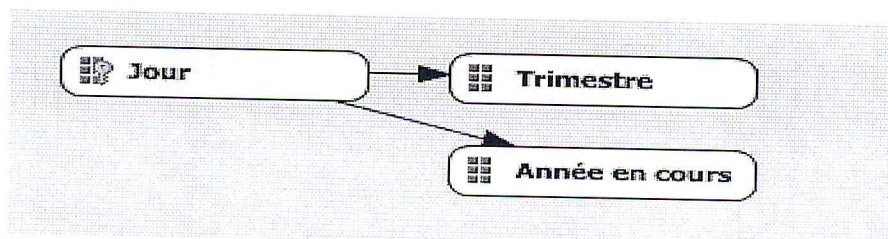


Figure 4: La Hiérarchie de la dimension « Période »

➤ **Table de la dimension « Employé »**

Attribut	Description	Type information
Employe_PK	Clé primaire	Int
NomEmploye	Le nom de l'employé	Varchar(50)
PrenomEmploye	Le prénom de l'employé	Varchar(50)
MatriculeEmploye	Le prénom de l'employé	Varchar(50)
SexeEmploye	Le sexe de l'employé	Char(10)
SituFamEmploye	La situation familiale de l'employé	Char(10)
DateNaisEmploye	La date de naissance de l'employé	Date
DateRecruEmploye	La date de recrutement de l'employé	Date

Tableau 7 : La table DimEmploye

• **Hiérarchie de la dimension « Employé »**

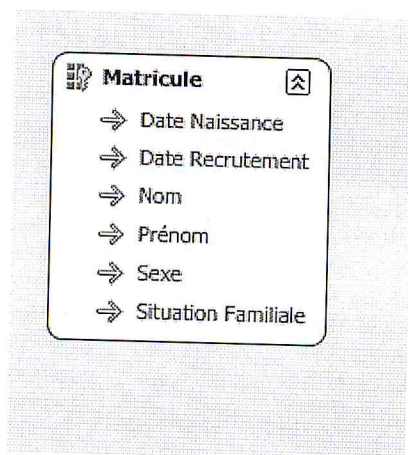


Figure 5 : Hiérarchie de la dimension « Employé »

➤ **Table de la dimension « Poste de Travail »**

Attribut	Description	Type information
Poste_PK	Clé primaire	Int
CodePoste	Le code du poste	Varchar(50)
PosteTravail	Le poste de travail	Varchar(50)
CodeLocalité	Le code de la localité	Varchar(50)
Localité	La localité	Varchar(50)

Le tableau 8 : La table DimPoste

• **Hierarchie de la dimension « Poste de Travail »**

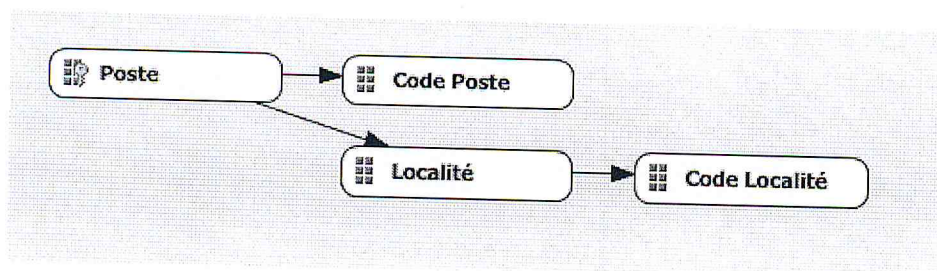


Figure 6 : Hiérarchie de la dimension « Poste de Travail »

➤ **Table de la dimension « Habilités d'encadrement »**

Attribut	Description	Type information
Hab_Enca_PK	Clé primaire	Int
CotePlanification	La cote octroyée à cette habilité	Varchar(50)
CoteDécision	La cote octroyée à cette habilité	Varchar(50)
CoteControle	La cote octroyée à cette habilité	Varchar(50)
CoteDéléation	La cote octroyée à cette habilité	Varchar(50)
CoteGestion	La cote octroyée à cette habilité	Varchar(50)

Tableau 9 : La table DimHabEnca

• **Hiérarchie de la dimension « Habilités d'encadrement »**

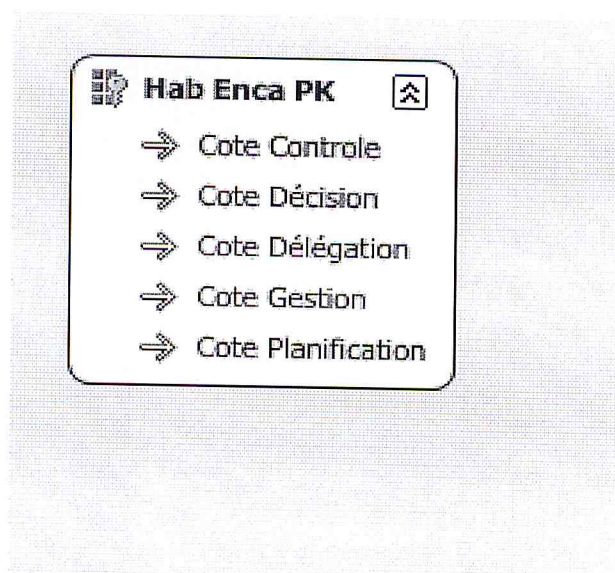


Figure 7 : Hiérarchie de la dimension « Habilités d'encadrement »

➤ **Table de la dimension « habilités Interpersonnelles »**

Attribut	Description	Type information
Hab_InterPer_PK	Clé primaire	Int
CoteCommunication	La cote octroyée à cette habilité	Varchar(50)
CoteAnimation	La cote octroyée à cette habilité	Varchar(50)
CoteRelation	La cote octroyée à cette habilité	Varchar(50)

Tableau 10 : La table DimHabInterPer

• **Hiérarchie de la dimension « Habilités Interpersonnelles »**

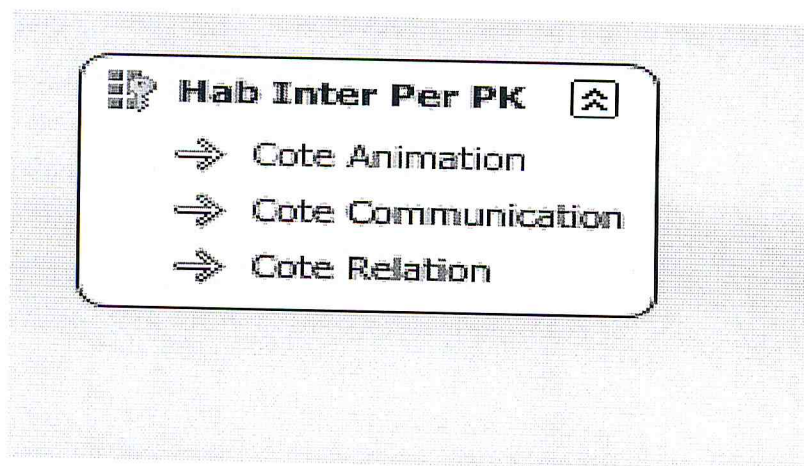


Figure 8 : Hiérarchie de la dimension « Habilités Interpersonnelles »

➤ **Table de la dimension « Habilités Personnelles »**

Attribut	Description	Type information
Hab_Perso_PK	Clé primaire	Int
CoteAutonomie	La cote octroyée à cette habilité	Varchar(50)
CoteDynamisme	La cote octroyée à cette habilité	Varchar(50)

Tableau 11 : La table DimHabPerso

• **Hiérarchie de la dimension « Habilités Personnelles »**

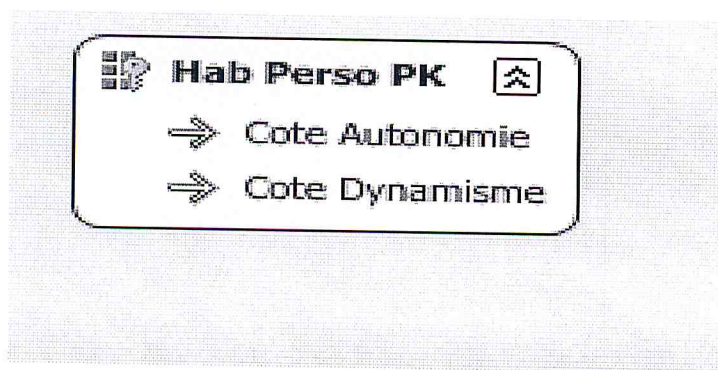


Figure 9 : Hiérarchie de la dimension « Habilités Personnelles »

✚ Schéma du model physique

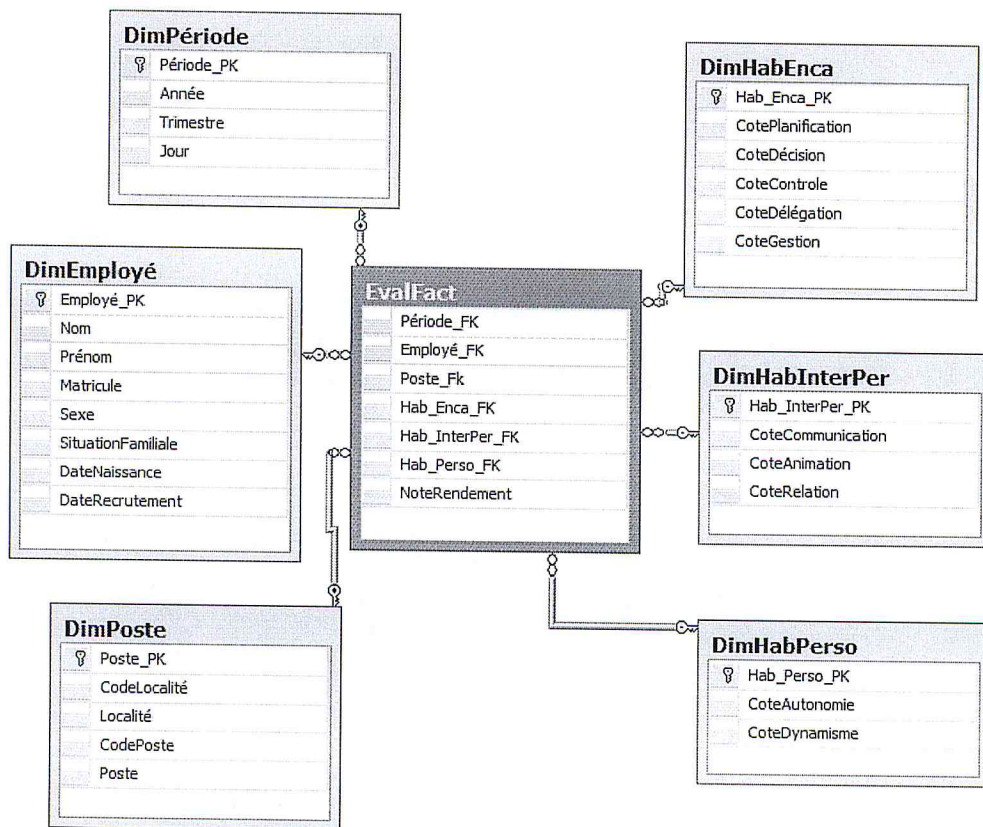


Figure 10 : Schéma du modèle physique de DataMart « évaluation des employés »

10. Présentation de l'information :

Partie émergée de l'iceberg, la présentation de l'information n'est pas celle qui occupe la plus grande part du temps de travail, C'est pourtant bien par là que l'on aborde le plus souvent l'informatique décisionnelle négligeant de prendre en compte les autres niveaux de l'architecture. Le parti pris dans cette étude a été, comme l'ont démontré les précédents chapitres, de ne pas commettre cette erreur et de ne pas brûler les étapes dans la conception du DataMart. Ce n'est pas pour cela que la présentation doit être négligée. Bien au contraire, Il s'agit du point d'entrée du DataMart qui va permettre de valoriser le travail effectué en amont. Si belle est la mécanique sous-jacente, si l'application qui permet de visualiser les données ne répond pas au besoin ou n'est pas ergonomique, elle sera rejetée par les utilisateurs et le projet sera enterré.

Les utilisateurs souhaitent souvent créer leurs propres requêtes sur le DataMart et il y a aussi Les navigateurs multidimensionnels permettent de manipuler des tableaux dynamiques. Ces tableaux ont une forme et un niveau de détail défini initialement mais contrairement au reporting, l'utilisateur peut analyser plus finement les données en naviguant dans le cube grâce à la technologie OLAP. Par opposition aux structures OLTP des applications à vocation transactionnelle, ces structures analytiques sont appelées **OLAP** (*OnLine Analytical Processing*). Ce type de structure peut-être implémentée de deux manières :

- Dans une base de données relationnelle classique, il faudra alors utiliser un moteur **ROLAP** (Relationnel) pour créer les cubes,
- Directement dans une base de données multidimensionnelle OLAP, c'est le concept **MOLAP** (Multidimensionnel).

La première solution permet de conserver une structure relationnelle ce qui facilite le chargement de gros volume de données. En revanche, les performances lors de

l'interrogation sont moins bonnes puisque le moteur ROLAP doit créer les agrégats en mémoire.

A l'inverse, les moteurs MOLAP stockent directement les données sous forme dimensionnelle et calculent les agrégats au cours de l'alimentation ce qui alourdit le processus.

Dans notre cas d'étude nous avons recensé les requêtes les plus pertinentes qui vont en adéquation avec l'objectif principal du DataMart qui est l'évaluation des employés de l'entreprise offrant ainsi aux personnels du service des ressources humaines un panel de choix outre la capacité de pouvoir par eux même naviguer dans le cube élaboré grâce à la modélisation dimensionnelle de notre DataMart et ceci pour répondre à leur besoin en matière de vue logistique générale qui doit surplomber le travail des employés pour, en déterminer les plus compétents parmi eux afin de prendre les décisions adéquate en matière cette fois ci de valorisation et promotion des travailleurs méritants pour créer un climat de concurrence bénéfique à l'émancipation de la compagnie dont le DataMart joue un rôle axiale dans cette perspective d'où l'importance avérée de l'intronisation du concept de l'informatique décisionnelle ou la business intelligence car sans cela rien n'aurait été permis et comme ça notre contrat si on ose dire est bel et bien rempli.

Reste à faire une présentation du choix de l'application utilisé et donner des exemples concrets pour mieux permettre aux lecteurs de ce Mémoire de bien cerner notre travail.

11. Présentation de la solution choisie pour l'élaboration du projet

➤ Présentation de la B.I. et SQL Server 2008 R2

Produit par Microsoft, SQL Server 2008 R2 est un système de gestion de bases de données relationnelles. Le stockage, la manipulation et l'analyse de ces données se font au sein de son moteur de bases de données. Ce service permet la réalisation de nombreuses applications, requêtes, et transactions. En plus de son moteur, SQL Server 2008 contient trois plateformes, indispensables pour réaliser un projet BI :

- Integration Services, qui permet d'intégrer des données provenant de différentes sources pour les ranger dans un entrepôt central ;
- Analysis Services, qui permet d'analyser les données grâce à des fonctions d'analyse multidimensionnelle ;
- Reporting Services, qui permet de créer, gérer et publier des rapports résultant des analyses réalisées antérieurement.

Et voici des exemples schématiques de différentes interfaces graphiques du SQL Server R2 2008 avec aussi des exemples non exhaustifs de rapports qu'on a pu réaliser grâce à ce dernier :

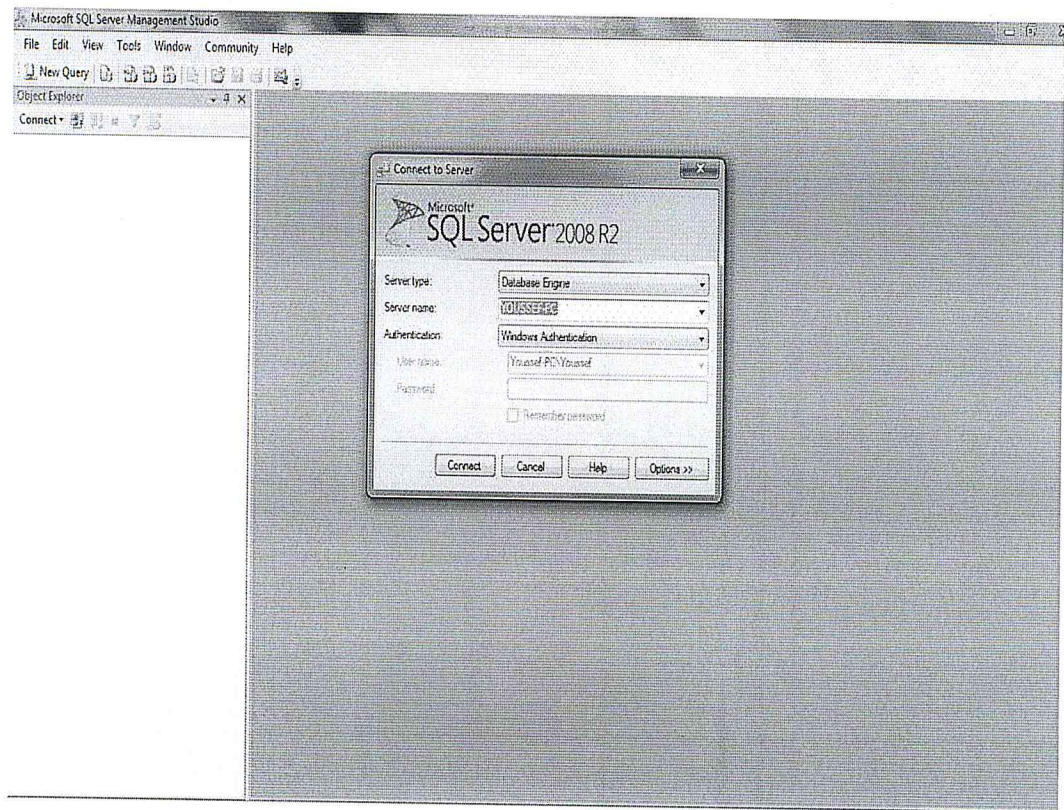


Figure 11 : Interface SQL Server R2 2008

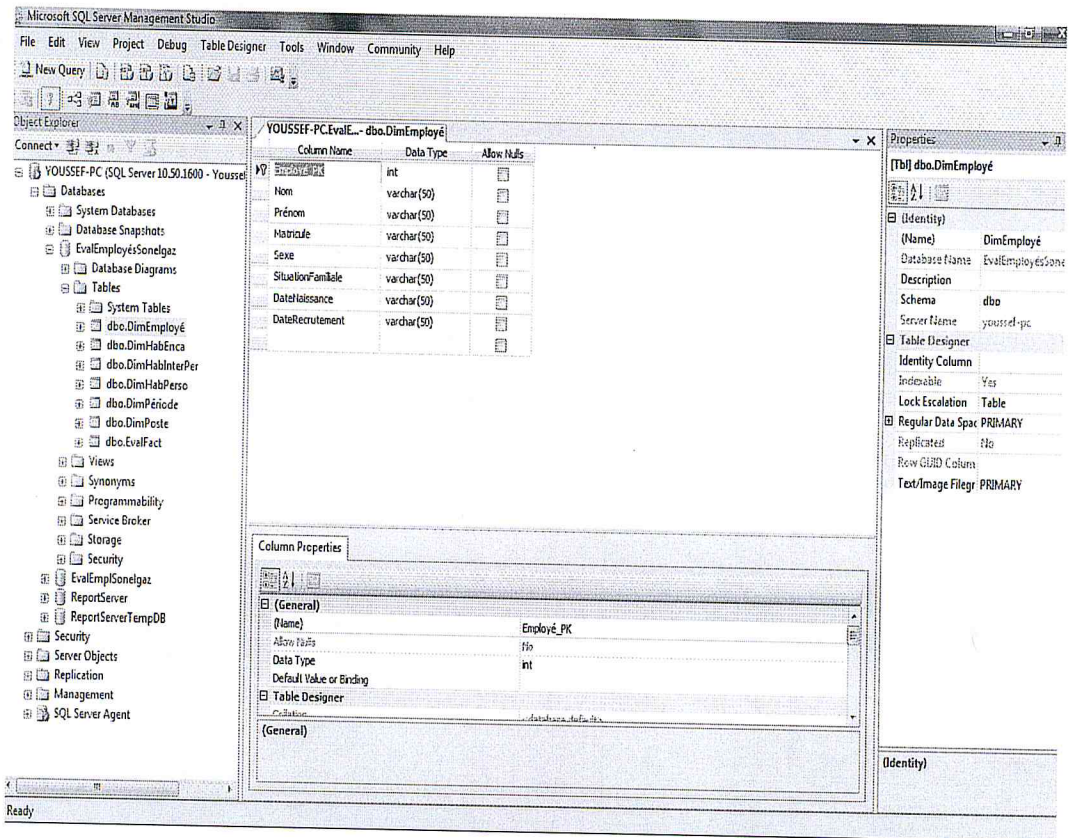


Figure 12 : Tableau de bord du SQL Server R2 2008

Microsoft SQL Server Management Studio

File Edit View Query Project Debug Tools Window Community Help

New Query

master

Object Explorer

Connect

YOUSSEF-PC (SQL Server 10.50.1600 - Youssef)

Databases

- System Databases
- Database Snapshots
- EvalEmployesSonegaz
- Database Diagrams
- Tables
 - System Tables
 - dbo.DimEmploye
 - dbo.DimHabEnca
 - dbo.DimHabInterPer
 - dbo.DimHabPerso
 - dbo.DimPériode
 - dbo.DimPoste
 - dbo.EvalFact
- Views
- Synonyms
- Programmability
- Service Broker
- Storage
- Security
- EvalEmpSonegaz
- ReportServer
- ReportServerTempDB
- Server Objects
- Replication
- Management
- SQL Server Agent

SQLQuery1.sql - Y...f-PCYoussef (SS) | YOUSSEF-PC.EvalE...-dbo.DimEmploye

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP 1000 [Poste_PK]
, [CodeLocalité]
, [Localité]
, [CodePoste]
, [Poste]
FROM [EvalEmployesSonegaz].[dbo].[DimPoste]

```

Properties

Current connection parameters

Aggregate Status

Connection failed

Elapsed time: 00:00:00.7910453

Finish time: 20/11/2012 09:10:59

Name: YOUSSEF-PC

Rows returned: 873

Start time: 20/11/2012 09:10:59

State: Open

Connection

Connection name: YOUSSEF-PC (Youssef)

Connection Details

Connection elapsed: 00:00:00.7910453

Connection finish: 20/11/2012 09:10:59

Connection rows: 873

Connection start: 20/11/2012 09:10:59

Connection state: Open

Display name: YOUSSEF-PC

Login name: Youssef-PC\Youssef

Server name: YOUSSEF-PC

Server version: 10.50.1600

Session: Training 36

SPID: 55

Name

The name of the connection.

Results

Poste_PK	CodeLocalité	Localité	CodePoste	Poste
151	18001400	SDC DFC	C842	AGT GEST REGLEMENT FINANCIER NIV2
152	18001500	SDC DPT RH	X251	CHEF SCE ASSIST CONTR DP NIV1
153	18001500	SDC DPT RH	W752	COP RESSOURCES HUMAINES NIV2
154	18001500	SDC DPT RH	B71	AGT GEST PAL PERSONNEL HC NIV1
155	18001500	SDC DPT RH	A251	CHARGE DES ETUDES NIV1
156	18001500	SDC DPT RH	A251	CHARGE DES ETUDES NIV1
157	18001500	SDC DPT RH	A251	CHARGE DES ETUDES NIV1
158	18001500	SDC DPT RH	D341	AGT GEST PAL PERSONNEL NIV1
159	18001500	SDC DPT RH	D341	AGT GEST PAL PERSONNEL NIV1
160	18001500	SDC DPT RH	W782	CHEF SCE DEVELOPPEMENT RH NIV1
161	18001500	SDC DPT RH	W802	CHEF SCE GESTION RH NIV2
162	18001500	SDC DPT RH	AD81	SECRETARE PAL NIV1

Query executed successfully. YOUSSEF-PC (10.50 RTM) | Youssef-PC\Youssef (SS) | master | 00:00:00 | 873 rows

Ready

Ln1 Col1 Ch1 BNS

Figure 13 : Exemple de listing de la dimension (employé)

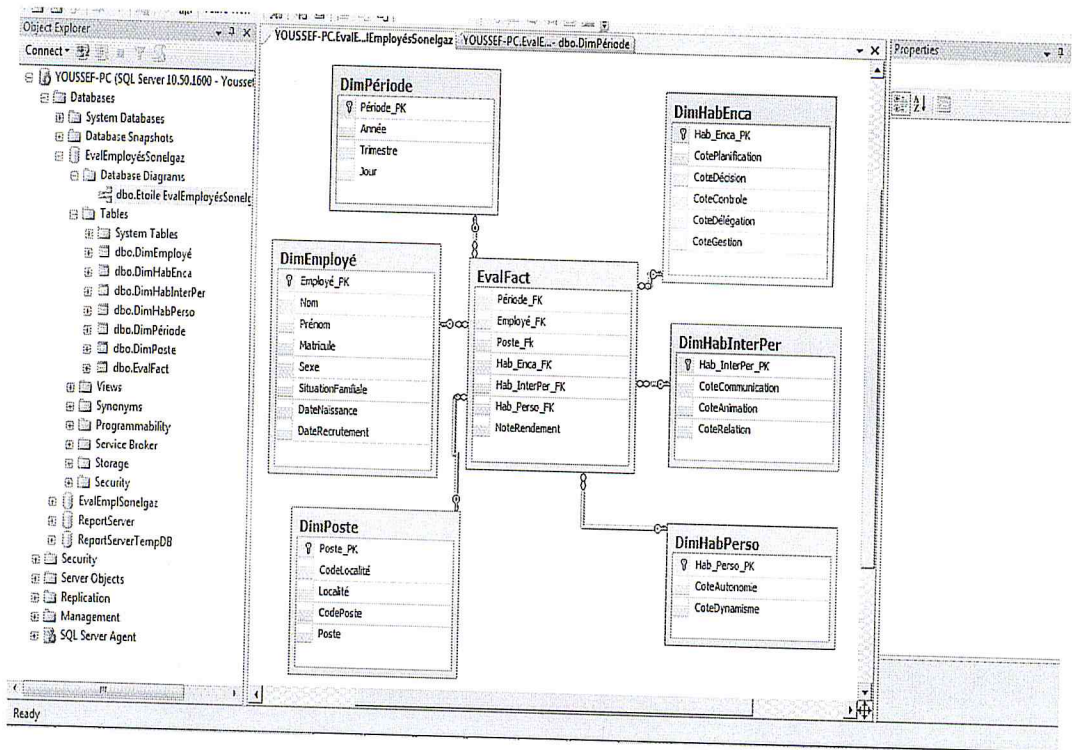


Figure 14 : Réalisation des jointures en étoile

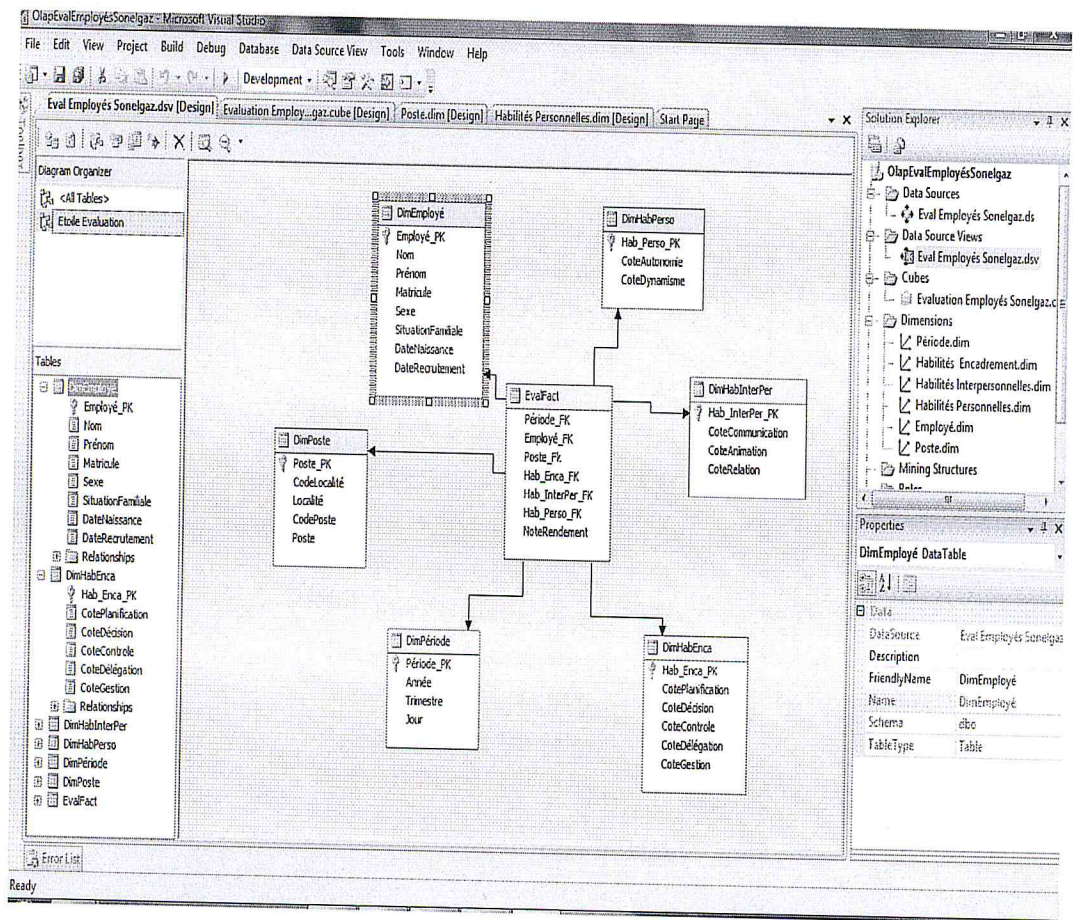


Figure 15 : Exemple du model physique

➤ Exemple de rapport des notes de rendements des employés sur deux années d'une localité donnée :

	(2010-2011)	(2011-2012)	Total
1593	3,25	5,25	4,25
SLIMANI	3,25	5,25	4,25
DJILALI	3,25	5,25	4,25
M	3,25	5,25	4,25
AGT GESTIO N CLIENTE LE NIV1	3,25	5,25	4,25
32841	4,5	5,75	5,125
CHACHI	4,5	5,75	5,125
RABAH	4,5	5,75	5,125
M	4,5	5,75	5,125
MAITRE OUVR OPER POLY PI NIV1	4,5	5,75	5,125
32876	4,5	4,5	4,5
CHEMLAL	4,5	4,5	4,5
YOUCEF	4,5	4,5	4,5
M	4,5	4,5	4,5
OPERAT EUR POLYVA LENT PFTT	4,5	4,5	4,5

	NIV2			
37035		5	4,5	4,75
SAIDI		5	4,5	4,75
NADIR		5	4,5	4,75
M		5	4,5	4,75
	CHEF SCE COMME RC (AGENC E) NIV1	5	4,5	4,75
37044		5,5	5,5	5,5
GUELLATI		5,5	5,5	5,5
MOHAMED		5,5	5,5	5,5
M		5,5	5,5	5,5
	ATTACH E COMME RCIAL PAL NIV2	5,5	5,5	5,5
39325		6,25	3,75	5
LAHSEN		6,25	3,75	5
MOURAD		6,25	3,75	5
M		6,25	3,75	5
	AGT GESTIO N CLIENTE LE NIV1	6,25	3,75	5
39758		5,25	4	4,625
LAGUEL		5,25	4	4,625
RADOUANE		5,25	4	4,625

	M		5,25	4	4,625
		EMPL POLYV TECH- COM PAL NIV1	5,25	4	4,625
46202			5,25	5,75	5,5
		OTSMANE	5,25	5,75	5,5
		MOHAMED	5,25	5,75	5,5
	M		5,25	5,75	5,5
		MAITRE OUVR OPER POLY PI NIV1	5,25	5,75	5,5
46204			5	4,25	4,625
		CHERFOUH	5	4,25	4,625
		DJAMEL	5	4,25	4,625
	M		5	4,25	4,625
		EMPL POLYVA LENT TECHNI CO- NIV3	5	4,25	4,625
47370			4,75	4,75	4,75
		TOUATI	4,75	4,75	4,75
		TOUFIK	4,75	4,75	4,75
	M		4,75	4,75	4,75
		OPERAT EUR POLYVA LENT PETI NIV1	4,75	4,75	4,75
47374			3,75	5,25	4,5

FERRADJI		3,75	5,25	4,5
KARIMA		3,75	5,25	4,5
F		3,75	5,25	4,5
	CHARGE DES ETUDES NIV1	3,75	5,25	4,5
71307		5,25	3,75	4,5
LAROUI		5,25	3,75	4,5
MOHAMED		5,25	3,75	4,5
M		5,25	3,75	4,5
	EMPL POLYV TECH- COM PAL NIV1	5,25	3,75	4,5
75228		4,5	5,25	4,875
AIT MEBAREK		4,5	5,25	4,875
MOHAMED		4,5	5,25	4,875
M		4,5	5,25	4,875
	ATTACH E COMME RCIAL NIV1	4,5	5,25	4,875
87599		5,5	5,5	5,5
DRARENI		5,5	5,5	5,5
MOURAD		5,5	5,5	5,5
M		5,5	5,5	5,5
	ATTACH E COMME	5,5	5,5	5,5

	RCIAL NIV1			
87600		6,5	4,75	5,625
	ALLIG	6,5	4,75	5,625
	SID ALI	6,5	4,75	5,625
	M	6,5	4,75	5,625
	ATTACH E COMME RCIAL NIV1	6,5	4,75	5,625
87602		5	5,5	5,25
	ALLAL	5	5,5	5,25
	MENAOUER	5	5,5	5,25
	M	5	5,5	5,25
	ATTACH E COMME RCIAL NIV1	5	5,5	5,25
87603		5	3,75	4,375
	REMILI	5	3,75	4,375
	MOHAMED	5	3,75	4,375
	M	5	3,75	4,375
	ATTACH E COMME RCIAL NIV1	5	3,75	4,375
87604		4,5	4	4,25
	BOUZIANE	4,5	4	4,25
	FOUZIA	4,5	4	4,25

	F		4,5	4	4,25
		ATTACH E COMME RCIAL NIV1	4,5	4	4,25
87605			5	5,5	5,25
		KARTALI	5	5,5	5,25
		MAHMOUD	5	5,5	5,25
		M	5	5,5	5,25
		ATTACH E COMME RCIAL NIV1	5	5,5	5,25
87611			7	5,25	6,125
		DJANATI	7	5,25	6,125
		TOUFIK	7	5,25	6,125
		M	7	5,25	6,125
		OPERAT EUR POLYVA LENT PETI NIV1	7	5,25	6,125
88667			6	5	5,5
		MOUDER	6	5	5,5
		MOHAMED	6	5	5,5
		M	6	5	5,5
		ATTACH E COMME RCIAL NIV1	6	5	5,5

88676		6	5,5	5,75
	KARAR	6	5,5	5,75
	DJAFFAR	6	5,5	5,75
	M	6	5,5	5,75
		6	5,5	5,75
				CHARGE DES ETUDES NIV1
93423		4,25	4	4,125
	CHOULI	4,25	4	4,125
	ABDELAZIZ	4,25	4	4,125
	M	4,25	4	4,125
		4,25	4	4,125
				ATTACH E COMME RCIAL NIV1
93426		5,5	4,75	5,125
	RAHIM	5,5	4,75	5,125
	MOURAD	5,5	4,75	5,125
	M	5,5	4,75	5,125
		5,5	4,75	5,125
				ATTACH E COMME RCIAL NIV1
93427		5	6,75	5,875
	TIBERKAK	5	6,75	5,875
	RACHID	5	6,75	5,875
	M	5	6,75	5,875
		5	6,75	5,875
				ATTACH

	E COMME RCIAL NIV1			
96717		4,75	5,5	5,125
TABTI		4,75	5,5	5,125
MEROUANE		4,75	5,5	5,125
M		4,75	5,5	5,125
	OPERAT EUR POLYVA LENT PETI NIV1	4,75	5,5	5,125
99935		5,5	4,75	5,125
YOUNSI		5,5	4,75	5,125
ABDELMOUMENE		5,5	4,75	5,125
M		5,5	4,75	5,125
	ATTACH E COMME RCIAL NIV1	5,5	4,75	5,125
99936		5,25	4,5	4,875
BEGHDAD		5,25	4,5	4,875
MOHAMED		5,25	4,5	4,875
M		5,25	4,5	4,875
	ATTACH E COMME RCIAL NIV1	5,25	4,5	4,875
99937		5,5	5	5,25
ADEL		5,5	5	5,25

	MOHAMED	5,5	5	5,25
	M	5,5	5	5,25
	ATTACH E COMME RCIAL NIV1	5,5	5	5,25
99939		5,75	4,25	5
	HABIS	5,75	4,25	5
	ABDELAZIZ	5,75	4,25	5
	M	5,75	4,25	5
	ATTACH E COMME RCIAL NIV1	5,75	4,25	5
99940		5,75	5,5	5,625
	BENHADJA	5,75	5,5	5,625
	FAYCAL	5,75	5,5	5,625
	M	5,75	5,5	5,625
	ATTACH E COMME RCIAL NIV1	5,75	5,5	5,625
99941		5	3,5	4,25
	BENELKAID	5	3,5	4,25
	YASSINE	5	3,5	4,25
	M	5	3,5	4,25
	ATTACH E COMME RCIAL NIV1	5	3,5	4,25

99948		4	5,25	4,625
CHEKAIMI		4	5,25	4,625
AHMED		4	5,25	4,625
M		4	5,25	4,625
	ATTACH E COMME RCIAL NIV1	4	5,25	4,625
Total		5,136363636 36364	4,871212121 21212	5,003787878 78788

- Exemple de rapport détaillé sur des employés et leurs notes de rendement d'une localité donnée au cours d'une année donnée

						(2011-2012)
						Quatrième Trimestre
Nom	Prénom	Poste	Cote Décision	Cote Gestion	Cote Relation	2012-09-30
ADEL	MOHAMED	ATTACHE COMMERCIAL NIV1	à parfaire	acquise	à développer	6
AIT MEBAREK	MOHAMED	ATTACHE COMMERCIAL NIV1	à parfaire	à développer	à acquérir	3
ALLAL	MENAOUER	ATTACHE COMMERCIAL NIV1	à acquérir	à développer	acquise	7
ALLIG	SID ALI	ATTACHE COMMERCIAL NIV1	acquise	à parfaire	acquise	6
BEGHDAD	MOHAMED	ATTACHE COMMERCIAL NIV1	à acquérir	acquise	à développer	5

BENELKAI D	YASSINE	ATTACHE COMMERCIAL NIV1	acquise	à parfaire	à parfaire	4
BENHADJ A	FAYCAL	ATTACHE COMMERCIAL NIV1	à parfaire	à développer	à parfaire	4
BOUZIAN E	FOUZIA	ATTACHE COMMERCIAL NIV1	à acquérir	acquise	à parfaire	3
CHACHI	RABAH	MAITRE OUVR OPER POLY PI NIV1	à parfaire	à acquérir	à parfaire	7
CHEKAIM I	AHMED	ATTACHE COMMERCIAL NIV1	à parfaire	à parfaire	à développer	5
CHEMLAL	YOUCEF	OPERATEUR POLYVALENT PETI NIV2	à développer	acquise	acquise	6
CHERFOU H	DJAMEL	EMPL POLYVALENT TECHNICO- NIV3	acquise	à acquérir	à acquérir	4
CHOULI	ABDELAZIZ	ATTACHE COMMERCIAL NIV1	à développer	à développer	à acquérir	3
DJANATI	TOUFIK	OPERATEUR POLYVALENT	acquise	à développer	à développer	4

		T PETI NIV1				
DRARENI	MOURAD	ATTACHE COMMERCIAL NIV1	à acquérir	à acquérir	acquise	7
FERRADJI	KARIMA	CHARGE DES ETUDES NIV1	à parfaire	acquise	acquise	3
GUELLATI	MOHAMED	ATTACHE COMMERCIAL PAL NIV2	acquise	à parfaire	acquise	6
HABIS	ABDELAZIZ	ATTACHE COMMERCIAL NIV1	à parfaire	à acquérir	à acquérir	3
KARAR	DJAFFAR	CHARGE DES ETUDES NIV1	à parfaire	à acquérir	à développer	5
KARTALI	MAHMOUD	ATTACHE COMMERCIAL NIV1	à acquérir	à acquérir	à parfaire	3
LAGUEL	RADOUANE	EMPL POLYV TECH-COM PAL NIV1	à parfaire	à acquérir	à développer	4
LAHSEN	MOURAD	AGT GESTION CLIENTELE NIV1	à développer	à parfaire	à développer	3
LAROUÏ	MOHAMED	EMPL POLYV TECH-COM	à acquérir	acquise	acquise	4

		PAL NIV1				
MOUDER	MOHAMED	ATTACHE COMMERCIAL NIV1	à développer	à parfaire	à acquérir	4
OTSMANE	MOHAMED	MAITRE OUVR OPER POLY PI NIV1	à développer	à acquérir	à développer	6
RAHIM	MOURAD	ATTACHE COMMERCIAL NIV1	acquise	à acquérir	acquise	6
REMILI	MOHAMED	ATTACHE COMMERCIAL NIV1	à développer	à parfaire	à acquérir	5
SAIDI	NADIR	CHEF SCE COMMERC (AGENCE) NIV1	à développer	acquise	à développer	5
SLIMANI	DJILALI	AGT GESTION CLIENTELE NIV1	à acquérir	acquise	acquise	6
TABTI	MEROUANE	OPERATEUR POLYVALENT PETI NIV1	à parfaire	à acquérir	à acquérir	7
TIBERKAK	RACHID	ATTACHE COMMERCIAL NIV1	à acquérir	à développer	à acquérir	7
TOUATI	TOUFIK	OPERATEUR	acquise	à acquérir	à acquérir	4

		POLYVALENT PETI NIV1				
YOUNSI	ABDELMOUMENE	ATTACHE COMMERCIAL NIV1	acquise	à développer	à acquérir	7

➤ Exemple d'évaluation du rendement d'un employé choisi sous forme de schéma

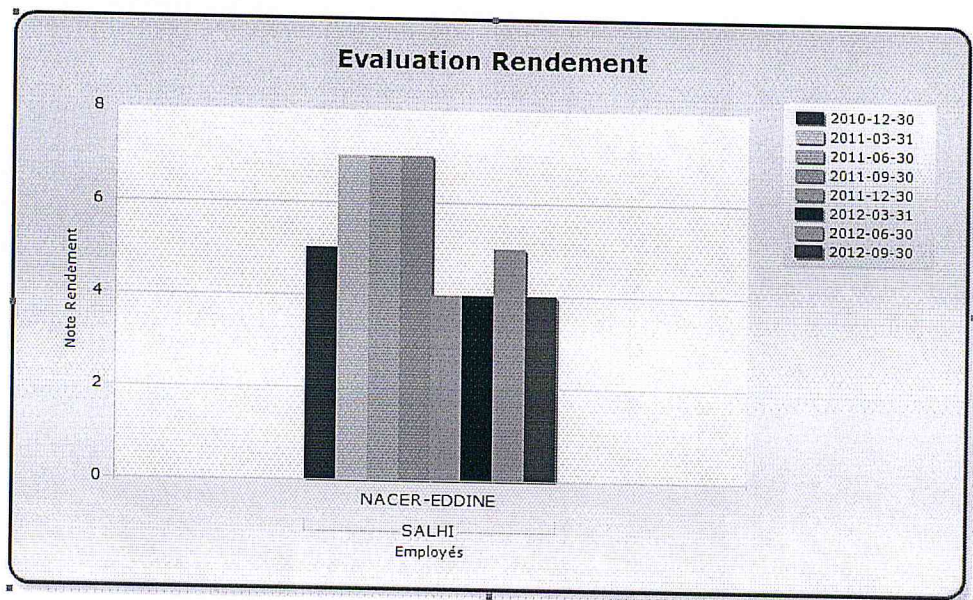


Figure 16 : Schéma du rendement d'un employé

- Exemple de comparaison entre le rendement de cinq employés sous forme d'un schéma

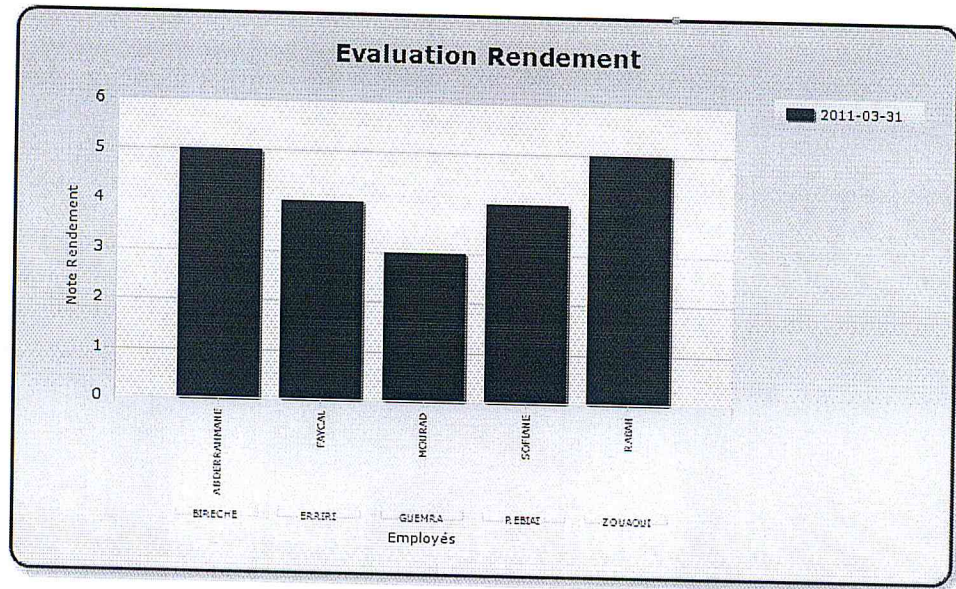


Figure 17 : Schéma de comparaison entre le rendement de 5 employés

12. Conclusion

Les nouvelles technologies de l'information permettent de concevoir des systèmes d'informations particulièrement performants. Ces derniers fournissent d'importantes informations mais ne sont pas conçus pour permettent leurs utilisations dans un processus d'aide à la décision.

Aussi, le Datawarehouse permet au décideur de travailler dans un environnement informationnel, référencé, homogène et historisé. Cette technique l'affranchit des problèmes liés à l'hétérogénéité des systèmes informatiques, ainsi que celle des différentes définitions de données issues de l'historique de l'organisation.

Par ailleurs, l'analyse multidimensionnelle permet d'extraire du DataWarehouse, une connaissance quantitative de l'activité de l'organisation selon les dimensions ou les aspects qui préoccupent le décideur à un instant donné.

Ainsi, Sonelgaz qui est confrontée à un environnement de plus en plus rude se dotera d'outils performants destinés à assister les managers dans leurs prises de décisions, aussi bien sur le plan ressources humaines et pourquoi pas même sur le plan financier, que sur les plans commercial.

Désormais, les managers pourront accéder à l'information stratégique. Ceci permet à l'entreprise d'être plus réactive mais pose aussi de nouveaux problèmes (problèmes de confidentialité, de compétences pour l'analyse,...).

Dans ce contexte, et en utilisant ces nouveaux outils mis à sa disposition, le contrôle de gestion a un rôle important à jouer en terme de coordination et de challenge de la performance auprès des managers et de la Direction Générale des ressources humaines.

Bibliographie

- “Piloter l’entreprise grâce au datawarehouse” Jean-Michel Franco et Sandrine de Lignerolles Eyrolles - 2001
- "La construction du datawarehouse" - Jean-François Goglin – 2001
- Entrepôts de données Guide pratique de modélisation dimensionnelle de *Ralph Kimball et Margy Ross* – 2002
- Le Datawarehouse - Guide de conduite de projet de *Ralph Kimball - Laura Reeves - Margy Ross - Warren Thornthwaite* – 2005
- Business Intelligence avec SQL Server 2008 R2 de *Sébastien Fantini* - 2010
- « Découverte de connaissances à partir de données » Marc Tommasi et Rémi Gilleron (2000)
<http://www.grappa.univ-lille3.fr/polys/fouille/index.html>
- Comment ça marche – la Business Intelligence – Jean-François Pillou (2004)
<http://www.commentcamarche.net/entreprise/business-intelligence.php3>
- Conception d'un entrepôt de données (Data Warehouse) par Yazid Grim 2010
<http://grim.developpez.com/cours/businessintelligence/concepts/conception-datawarehouse/>
- Initiation au décisionnel (Business Intelligence, DataWarehouse, OLAP) par Mohamed Taslimanka Sylla 2007
<http://taslimanka.developpez.com/>

- Tutoriel Microsoft SQL-Server 2008 : Base de données de films et de réalisateurs par Alex Vimory 2009
<http://a-vimory.developpez.com/tutoriels/sql-server-2008-films/>

- Mettre en place un entrepôt de données multidimensionnel par Carlos Da Costa 2011
<http://business-intelligence.developpez.com/>