

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA
RECHERCHE SCIENTIFIQUE
UNIVERSITE MOHAMED BOUDIAF - M'SILA



Mémoire présenté pour l'obtention Du diplôme de Master Académique

Par: Chenane Ichraq

Intitulé

Analyse multidimensionnelle de document

Soutenu devant le jury composé de :

Nom et prénom Enseignant

Président Dr. Brahimi Belkacem

Université de M'sila

Rapporteur Dr. Akhrouf Samir

Université de M'sila

Examineur Dr. Hellassa Madiha

Université de M'sila

Année universitaire : 2020 / 2021

Resumé

Les travaux de recherche présentés dans ce mémoire de thèse s'inscrivent dans le cadre des systèmes d'aide à la prise de décision. Ces systèmes se basent sur un processus d'analyse en ligne (OLAP) et structurent les données de manière multidimensionnelle. Nous avons proposé une nouvelle approche qui consiste à intégrer les données issues de documents XML au cœur du processus d'analyse. Jusqu'à présent, ces documents sont restés exclus des systèmes d'aide à la prise de décision, faute de méthode et d'outils adaptés. Selon [Tseng & Chou, 2006], les systèmes OLAP n'emploient que 20% des données qui transitent au sein des systèmes d'information des entreprises. Les 80% restants, des documents restent hors de portée des systèmes d'aide à la prise de décision. Dans [Sullivan, 2001], l'auteur argumente en faveur de l'intégration des données issues de documents et de l'emploi de la fouille de texte pour en permettre l'analyse. De leur côté, les auteurs de [Fankhauser & Klement, 2003] affirment que la technologie XML est assez mature pour permettre l'implantation d'outils de fouilles de texte performant. Nous avons proposé d'aller au delà de ces propos en fournissant un environnement permettant l'intégration de documents XML au sein d'un système OLAP adapté pour l'analyse de données textuelles contenues dans ces documents. Il s'agit d'une première approche en vue de l'intégration de 100% des données issues des systèmes d'information des entreprises. Pour permettre l'intégration de documents au sein de l'environnement d'aide à la prise de décision, nous avons proposé un modèle multidimensionnel associé à un ensemble d'opérations pour permettre la manipulation des concepts du modèle. Nous avons proposé une démarche pour intégrer les documents au sein de l'environnement. La validation de nos propositions a été effectuée par le développement d'un prototype permettant l'analyse multidimensionnelle de données issues de documents

abstract

The research presented in this dissertation falls within the framework of decision support systems. These systems are based on an online analysis process (OLAP) and structure the data in a multidimensional way. We have proposed a new approach that consists of integrating data from XML documents into the heart of the analysis process. Until now, these documents have remained excluded from decision support systems, due to a lack of suitable methods and tools. According to [Tseng & Chou, 2006], OLAP systems only use 20% of the data that passes through companies' information systems. The remaining 80% of documents remain out of reach of decision support systems. In [Sullivan, 2001], the author argues in favor of integrating data from documents and using text mining to enable analysis. For their part, the authors of [Fankhauser & Klement, 2003] assert that XML technology is mature enough to allow the implementation of powerful text mining tools. We proposed to go beyond these remarks by providing an environment allowing the integration of XML documents within an OLAP system suitable for the analysis of textual data contained in these documents. This is a first step towards integrating 100% of data from companies' information systems. To allow the integration of documents within the decision support environment, we proposed a multidimensional model associated with a set of operations to allow manipulation of the model concepts. We have proposed an approach to integrate documents into the environment. The validation of our proposals was carried out by the development of a prototype allowing the multidimensional analysis of data from documents

الملخص:

يندرج البحث المقدم في هذه الرسالة في إطار أنظمة دعم القرار. تستند هذه الأنظمة إلى عملية تحليل عبر الإنترنت (OLAP) وهيكل البيانات بطريقة متعددة الأبعاد. لقد اقترحنا نهجاً جديداً يتكون من دمج البيانات من مستندات XML في قلب عملية التحليل. حتى الآن ، ظلت هذه الوثائق مستبعدة من أنظمة دعم القرار ، بسبب نقص الأساليب والأدوات المناسبة. وفقاً لـ [Tseng & Chou ، 2006] ، تستخدم أنظمة OLAP فقط من البيانات التي تمر عبر أنظمة معلومات الشركات. تظل 80% المتبقية من الوثائق بعيدة عن تناول أنظمة دعم القرار. في [Sullivan ، 2001] ، يجادل المؤلف لصالح دمج البيانات من المستندات واستخدام التنقيب عن النص لتمكين التحليل. من جانبهم ، يؤكد مؤلفو [Fankhauser & Klement ، 2003] أن تقنية XML ناضجة بما يكفي للسماح بتنفيذ أدوات تنقيب عن النصوص قوية. اقترحنا تجاوز هذه الملاحظات من خلال توفير بيئة تسمح بدمج مستندات XML داخل نظام OLAP مناسب لتحليل البيانات النصية الواردة في هذه المستندات. هذه خطوة أولى نحو دمج 100% من البيانات من أنظمة معلومات الشركات. للسماح بدمج المستندات في بيئة دعم القرار ، اقترحنا نموذجاً متعدد الأبعاد مرتبطاً بمجموعة من العمليات للسماح بمعالجة مفاهيم النموذج. لقد اقترحنا نهجاً لدمج الوثائق في البيئة. تم التحقق من صحة مقترحاتنا من خلال تطوير نموذج أولي يسمح بالتحليل متعدد الأبعاد للبيانات من المستندات

Remerciements

Je tiens à remercier tout d'abord mon directeur de recherches, Professeur SAMIR Akhrouf, pour sa patience, et surtout pour sa confiance, ses remarques et ses conseils, sa disponibilité et sa bienveillance. Qu'il trouve ici le témoignage de ma profonde gratitude. Je voudrais également remercier les membres du jury pour avoir accepté d'évaluer ce travail et pour toutes leurs remarques et critiques, ainsi que le personnel et les enseignants de universités m'sila sans oublier les enseignants étrangers qui ont contribué à ma formation. Je tiens aussi à remercier monsieur de la société Algal + : Dr Abdelhak Kadri ainsi que tout le personnel et les enseignants du société pour leur soutien inestimable. A tous mes enseignants qui m'ont initié aux valeurs authentiques, en signe d'un profond respect et d'un profond amour !!!

Merci à vous tous

Dédicace

A la mémoire de mon défunt père

À cet source de tendresse, de patience et de générosité,,

À ma mère !

À mes beaux parents et à toute ma famille

À tous mes amis et collègues

À tous les étudiants de la promotion 2020/2021

Option : Faculté mathématiques et informatique

A tous ceux qui, par un mot, m'ont donné la force de continuer

Table des matières

Remerciement

Dédicace

Introduction

1. **Chapitre 1** : Présentation de la société "ALGALPLUS"

1.1 Introduction

1.2 Peu d'historique

1.3 service extrusion

2. **Chapitre 2** : System décisionnels

2.1 introduction

2.2. Déférence entre système décisionnel et opérationnel

2.3 Information décisionnelle BI et leur objectif

2.4 Entreposage de données (DW)

2.4.1 Définition DW (BILL Inmon /oracle)

2.4.2Contexte et problématique DW

2.4.3 Introduction

2.4.4 Problématique et objectif DW

2.4.5 Domaine utilisation DW par décideur

2.5 Cycle de vie d'un projet décisionnel

3. **Chapitre 3** : Analyse multidimensionnelle

3.1 Introduction

3.2 Granularité de dimension

3.3 Représentation dimensionnel

3.3.1 Dimension

3.3.2 E/R model vs dimensionnel model

3.4 OLAP

3.4.1 Définition OLAP

3.4.2 Opération OLAP

3.5 Modélisation dimensionnel

3.6 Modélisation en étoile

3.7 Modélisation en flacon

4. **chapitre 4**: réalisation de données

4.1 capture d'écran d'environnement de travail

Conclusion

Références bibliographiques

Introduction Générale

Les entrepôts de données et les systèmes d'analyse en ligne OLAP (On-Line Analytical Processing) fournissent des méthodes et des outils permettant l'analyse de données issues des systèmes d'information des entreprises. Mais, seules 20% des données d'un système d'information est constitué de données analysables par les systèmes OLAP actuels. Les 80% restant, constitués de documents, restent hors de portée de ces systèmes faute d'outils ou de méthodes adaptés. Pour répondre à cette problématique nous proposons un modèle conceptuel multidimensionnel pour représenter les concepts d'analyse. Ce modèle repose sur un unique concept, modélisant à la fois les sujets et les axes d'une analyse. Nous y associons une fonction pour agréger des données textuelles afin d'obtenir une vision synthétique des informations issues de documents. Cette fonction résume un ensemble de mots-clefs par un ensemble plus petit et plus général. Nous introduisons un noyau d'opérations élémentaires permettant la spécification d'analyses multidimensionnelles à partir des concepts du modèle ainsi que leur manipulation pour affiner une analyse. Nous proposons également une démarche pour l'intégration des données issues de documents, qui décrit les phases pour concevoir le schéma conceptuel multidimensionnel, l'analyse des sources de données ainsi que le processus d'alimentation. Enfin, pour valider notre proposition, nous présentons un prototype.

Mots-clefs

OLAP, Document XML, Galaxie, Modélisation multidimensionnelle, Démarche de modélisation, Base de données multidimensionnelle, Entrepôts de données, Entrepôts de documents.

Dans cette présentation, nous discuterons de quatre chapitres. Dans chaque chapitre, nous définirons nos buts et objectifs et aborderons le problème de l'affichage des données à travers des graphiques afin de faciliter le processus de prise de décision.

Analyse de données multidimensionnelles. Cela se fait en affichant et en visualisant les données de l'entreprise

Afin de traiter ces données, nous avons utilisé le programme Power BI d'analyse multidimensionnelle des données, qui prend en charge OLAP (Online Analytical Processing)

Le premier chapitre est représenté dans l'explication et la présentation des données et informations de la société Algal +, présentant ses services et un aperçu de son histoire.

pour le deuxième chapitre, nous avons étudié les points les plus importants de différence entre Système décisionnel et Système Opérationnel .

Nous avons expliqué comment le décideur choisit et sur quoi il dépend pour choisir la bonne décision à travers notre analyse de ces données et notre connaissance du magasin de données et informatique décisionnel (BI) et entreposage de données .

Dans le troisième Chapitre Nous avons parlé de l'analyse multidimensionnelle des données, et nous sommes arrivés à définir ce domaine et à expliquer les éléments les plus importants tels que les OLAP

Quant au quatrième chapitre, ce sont des captures d'écran que nous avons obtenues en visualisant les données de manière multidimensionnelle utilisées dans le programme Power BI

Chapitre 1

Présentation de la société «ALGALPLUS» et processus d'extrusion aluminium

1 . Présentation de la société Algal Plus algal+ Société installée à M'sila

(250 km au sud d'Alger) spécialisée dans la transformation

d'aluminium algal+ disposent :

d'un des plus importants complexes industriels de la profession en Afrique ;

d'une superficie de 127 000 m² avec trois presses d'extrusion UBE et

SCHLOEMANN de

2.500et 1.600T ; d'une capacité de filage annuelle de18.000 T. ; de deux chaînes de

thermo-laquage ; d'une chaîne complète d'anodisation et de coloration chimique

; d'un atelier de fonderie, des ateliers de parachèvement et de travail à façon

avec des

équipements à commande numérique ; d'une station ultrasophistiquée de traitement de

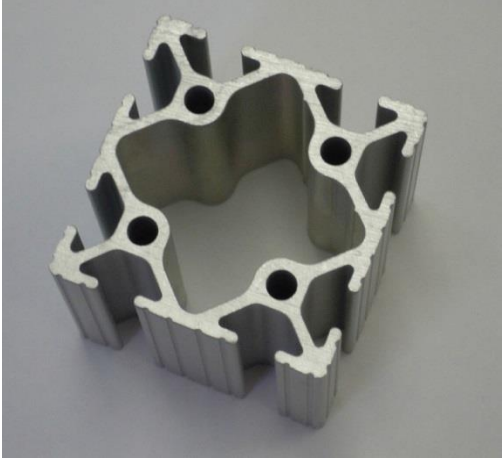
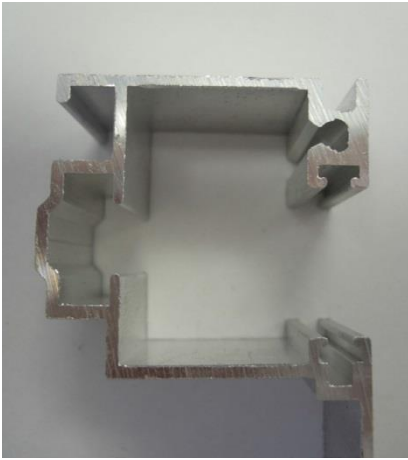

l'eau en amont et en aval ; d'une station d'électricité et de gaz permettant

d'accueillir dans un futur très proche quatre

autres presses d'extrusion.

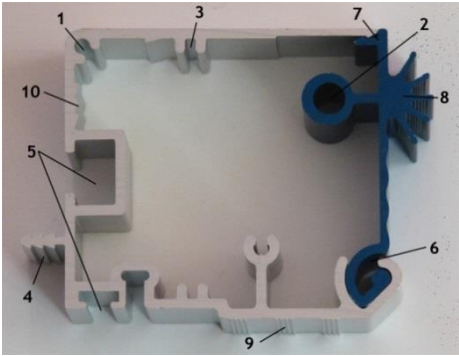
En 1979	La fondation par SNS d'une entité en vue de l'extrusion, anodisation, refonte de profilé aluminium et acquisition, à cet effet, d'un terrain de 123 000 m ² dans la zone industrielle de M'sila. A permis de créer de nombreux emplois très recherchés dans la région.
En 1982	Signature du contrat de réalisation clé en main de l'usine avec la société japonaise UBE, leader mondial en engineering dans l'aluminium.
En 1983	Individualisation du financement par le ministère des finances.
En 1985	Restructuration de SNS en groupe et prise de contrôle du complexe par METANOF.
En 1985	Entrée en production du complexe.
En 1996	Nouveaux investissement dans l'extension de l'unité d'anodisation et introduction de la coloration électrolytique.
En 1997	Nouveaux investissements dans l'extension des bâtiments existants et installation d'une ligne de thermo laquage. Désormais, l'entreprise est à même de proposer du profilé en l'état, anodisé, coloré et thermo laqué sur le même site.
En 1998	Obtention de la certification norme ISO 9001 version 2000.
En 2004	Obtention de la certification norme ISO 9001 version 2000.
En 2007	Un groupe de nouveaux actionnaires acquiert 100% des actifs industriels de l'entreprise.
En 2007	Nouveaux investissements dans l'acquisition d'une 3eme ligne d'extrusion 1600TN.
En 2008	Renouvellement de la certification ISO 9001 version 2000 Mise en place du système pour l'acquisition des labels QUALICOAT et QUALANOD.
Aperçu historique.	

5. Service extrusion :

		
<p align="center">Fig. 1 Profilé creux.</p>	<p align="center">Fig. 2 Profilé semi-creux</p>	<p align="center">Fig. 3 Profilé plein</p>

Par le procédé d'extrusion, on obtient un produit long de section transversale constante sur toute sa longueur. La section est généralement plus complexe que celle d'une barre ou d'un tube. Selon la forme de sa section transversale, il est appelé profilé plein, semi-creux, creux.

L'extrusion est un moyen de mise en forme qui n'est pas unique à l'aluminium, mais ce métal est le seul dont la formabilité permet d'obtenir des profilés de résistance structurale aux formes très élaborées qui répondent à des besoins spécifiques. C'est ce qui rend d'ailleurs l'extrusion très attrayante pour plusieurs transformateurs.

	<ol style="list-style-type: none"> 1. Canaux de vissage 2. Tube intégré 3. Renfort mural pour vissage 4. Ergot d'assemblage 5. Canaux pour écrou ou tête devis 6. Charnière 7. Clipsage 8. Diffuseur de chaleur 9. Rainures 10. Nervures
<p>Fig. 4 conception d'un produit fabriqué à partir d'extrusions</p>	

Il est possible d'intégrer de nombreuses fonctions lors de la conception d'un produit fabriqué à partir d'extrusions. La figure suivante en illustre quelques-unes.

Ces longues barres seront tronçonnées en billettes de longueur adaptée à la presse d'extrusion.



Le client choisit l'alliage et son état métallurgique ainsi que la forme du profilé requis pour un usage spécifique. Puis qu'il s'agit d'un procédé de formage à chaud, il faut se rappeler que les alliages non traitables thermiquement conservent les propriétés de leur état recuit à moins de subir sub séquentiellement un étirage à froid (applicable à des formes simples). Les tolérances réalisables par le

procédé sont définies par les normes ANSIA35.2, Aluminium Standards and Data, EN755, DIN17615T3...Elles devront



Fig. 5 Billettes d'extrusion

correspondre à celles prévues dans la conception du profilé.

Le procédé courant d'extrusion ou filage de l'aluminium est une mise en forme à chaud (450°C à 500°C) par écoulement du métal d'une billette que l'on presse à travers l'orifice d'une filière en

acier qui reproduit la section du profil à obtenir (barres, tubes, etc.).

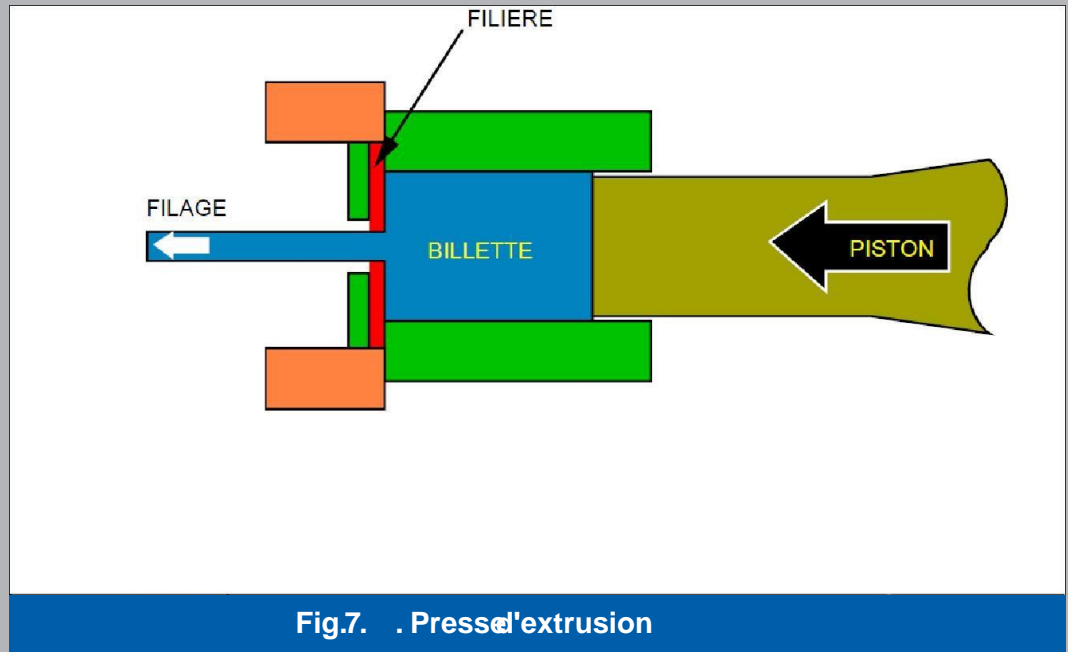
Fig. 6 Filière d'extrusion

La réussite du procédé, qui se traduit par l'atteinte des caractéristiques recherchées (finis de surface, tolérances dimensionnelles, résistance mécanique), s'appuie en grande partie sur le contrôle des températures : température de la billette préchauffée et de la filière, température à la sortie de presse et taux de refroidissement.

Le préchauffage des billettes est nécessaire pour amener le métal à la température à laquelle il sera suffisamment malléable pour s'écouler sous la pression à travers la filière. Ce préchauffage permet aussi d'atteindre la température nécessaire à la mise en solution des éléments pour les alliages traitables thermiquement.

Les billettes s'enchaînent rapidement dans la presse, donnant l'impression d'une alimentation ininterrompue de métal. Dans la presse, le métal est forcé sous la pression à passer au travers une filière (matrice ou « die »). On lui donne alors le nom de profilé. La force nécessaire pour permettre à la billette de traverser la filière est appliquée à l'aide d'un piston hydraulique.

Selon sa taille, une presse peut générer une force qui varie de quelques centaines de tonnes à plus de 20000 tonnes. Le choix de la presse (capacité requise) doit normalement être adapté aux dimensions des pièces filées. La billette utilisée est un long cylindre d'aluminium plein, dont le diamètre variant entre 6 pouces et 18 pouces, est directement relié à la capacité de la presse d'extrusion.



À sa sortie de la matrice d'extrusion, le profilé défile rapidement sous une unité de refroidissement. Selon le produit, on utilisera un jet d'eau, une brume, de l'air forcé ou non, ou une combinaison de ces derniers. Cette étape correspond à la trempe pour les alliages traitables thermiquement.

À la sortie de la presse, le profilé est coupé en longueur de plusieurs dizaines de mètres puis redressé par étirage (1 à 3% d'allongement). C'est à partir du besoin des clients que découler à la suite des opérations : coupe, vieillissement artificiel, peinture, anodisation et emballage.

Le procédé se déroule dans une usine dédiée, habituellement composée de 2 à 3 presses d'extrusion, autour des quelles sont greffés des équipements auxiliaires : ci saille ou scie à billettes, four à billettes, scie pour profilés, table de redressement, four à traitement thermique, chambre de peinture ou ligne d'anodisation.

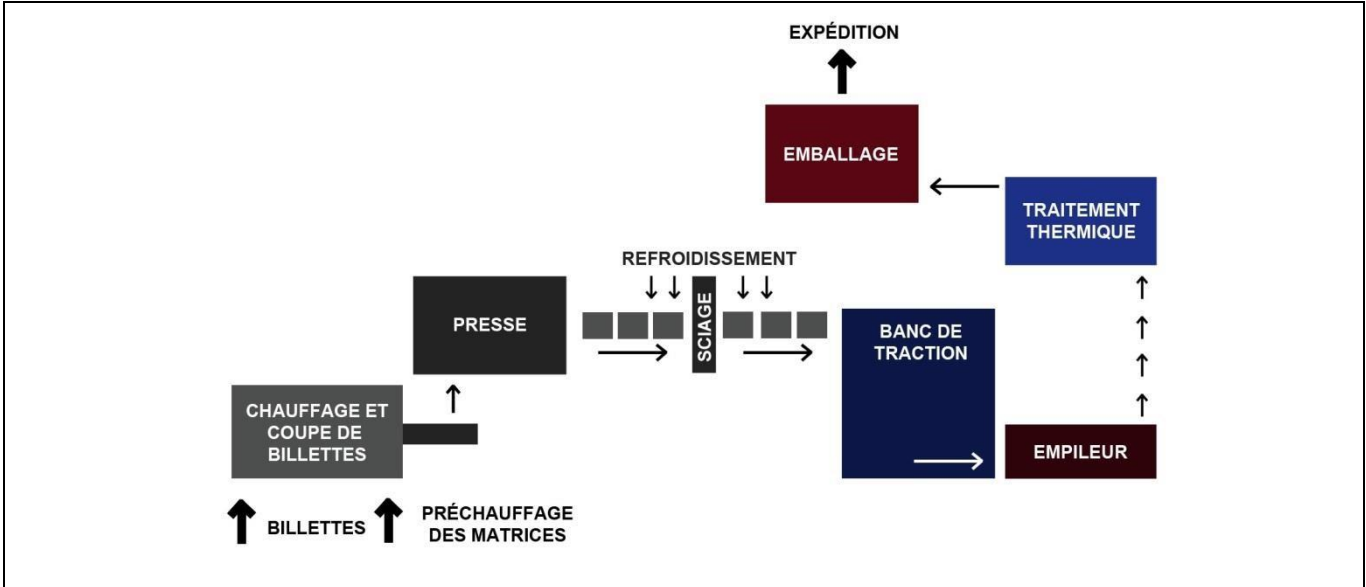


Fig. 8 Schéma général d'une ligne d'extrusion

Chapitre 2
System décisionnel

Introduction

Déférence entre system Opérationnels et System Décisionnels

Les systèmes « opérationnels » ou « de gestion », également appelés systèmes OLTP (*on-line transaction processing*), sont dédiés aux métiers de l'entreprise pour les assister dans leurs tâches de gestion quotidiennes et donc directement opérationnels. La tendance est à l'utilisation de P.G.I. (progiciels de gestion intégrée) qui regroupent tous les logiciels de gestion de l'entreprise – finances, ressources humaines, [logistique](#), ventes, etc. – en un unique progiciel paramétrable aux règles de l'entreprise, organisé autour d'une base de données, réduisant ainsi les coûts de communications entre applications.

Les systèmes « décisionnels », également appelés OLAP (*on-line analytical processing*), sont dédiés au management de l'entreprise pour l'aider au pilotage de l'activité, et donc indirectement opérationnels. Ils offrent au décideur une vision transversale de l'entreprise. La tendance pour réaliser un système décisionnel est à la mise en place d'un entrepôt de données.

Bien que les systèmes d'informations OLTP et OLAP aient le point commun de regrouper les données de l'entreprise dans un S.G.B.D. (système de gestion de bases de données) et d'en fournir l'accès aux utilisateurs, ils présentent de profondes différences, présentées ci-dessous.

Au niveau des données

Dans un système OLTP, les données ne sont conservées que sur une courte période ; elles sont détaillées, personnelles, identifiées (une facture appartient à une personne précise) et représentent généralement en volume quelques centaines de mégaoctets, voir quelques gigaoctets.

Dans un système OLAP, les données sont historisées et peuvent être agrégées. Elles peuvent aussi être anonymes ; il suffit de savoir qu'un client de tel type a effectué tel achat, l'identification n'est pas obligatoire pour en tirer des renseignements. Mais l'anonymisation des données dans l'entrepôt peut gêner la recherche d'un lien entre les différents événements concernant un même individu. La base peut atteindre des volumes con [...]

TABLE 3.5 A Comparison Between OLTP and OLAP

Criteria	OLTP	OLAP
Purpose	To carry out day-to-day business functions	To support decision making and provide answers to business and management queries
Data source	Transaction database (a normalized data repository primarily focused on efficiency and consistency)	Data warehouse or data mart (a nonnormalized data repository primarily focused on accuracy and completeness)
Reporting	Routine, periodic, narrowly focused reports	Ad hoc, multidimensional, broadly focused reports and queries
Resource requirements	Ordinary relational databases	Multiprocessor, large-capacity, specialized databases
Execution speed	Fast (recording of business transactions and routine reports)	Slow (resource intensive, complex, large-scale queries)

Figure 9 déférence entre OLAP et OLTP

L'informatique décisionnelle(BI)

Définition : L'informatique décisionnelle – ou BI (Business Intelligence) – est un processus technologique qui analyse des données pour présenter des informations exploitables par les dirigeants, les cadres commerciaux et les autres utilisateurs, afin de leur permettre de prendre des décisions plus avisées.

L'informatique décisionnelle englobe les différents outils, applications et méthodologies qui permettent à l'entreprise de collecter des données à partir de systèmes internes et de sources externes, de les préparer en vue d'une analyse, de développer des requêtes et de les appliquer à ces données.

On en tire alors diverses vues, ou modes de visualisation, tels que des rapports et des tableaux de bord, pour mettre les résultats analytiques à la disposition des décideurs et des acteurs de l'entreprise.

Objectifs de la BI

Accélération et amélioration de la prise de décisions ; optimisation des processus métier internes ; augmentation de l'efficacité opérationnelle ; génération de nouvelles recettes ; et gain d'avantages concurrentiels : autant de bénéfices potentiels qu'offrent les programmes BI.

Ces programmes contribuent également à identifier les tendances des marchés et les problèmes métier qu'il convient de résoudre.

Les données BI peuvent inclure des informations historiques, ainsi que de nouvelles données collectées auprès de systèmes source au fur et à mesure qu'elles sont générées. L'analyse BI peut ainsi soutenir des processus décisionnels tant stratégiques que tactiques.

Au départ, les outils BI étaient essentiellement utilisés par les professionnels de l'informatique, notamment les analystes des données. Ils procédaient aux analyses et généraient des rapports contenant des résultats de requêtes destinés aux utilisateurs métier. Aujourd'hui, les cadres et les opérationnels utilisent de plus en plus les logiciels BI, grâce au développement d'outils d'exploration de données et de BI en libre-service.

Variantes et déclinaisons

Interrogation et analyse ponctuelles ; rapports d'entreprise ; traitement analytique en ligne (OLAP) ; BI mobile ; BI en temps réel ; BI opérationnelle ; BI en Cloud et SaaS ; BI open source ; BI collaborative : l'informatique décisionnelle se décline dans de nombreuses applications d'analyse de données.

La technologie BI comprend également les logiciels de visualisation de données (ou DataViz) – qui permettent de mettre en forme des éléments infographiques (graphes, etc.) – ainsi que différents outils d'élaboration de tableaux de bord et de cartes de scores – qui présentent, de manière facile à appréhender, des données visualisées sous forme d'éléments de mesure métier et d'indicateurs clés de performance (KPI).

Les applications BI peuvent s'acheter séparément auprès de différents éditeurs, ou auprès d'un seul dans le cadre d'une plateforme BI unifiée.

Les programmes BI peuvent également intégrer des formes d'analytique avancé, telles que l'exploration de données (data mining), l'analytique prédictif, l'exploration de texte (text mining), l'analyse statistique ou encore l'analytique Big Data.

Toutefois, dans nombre de cas, les projets d'analytique avancé sont pilotés et réalisés par des équipes distinctes de professionnels compétents du domaine, tels que les data scientists, des statisticiens, ou encore des développeurs de modèles prédictifs, tandis que les équipes BI supervisent une analyse et une interrogation des données métier à un niveau plus simple.

Sous le capot

Les données de l'informatique décisionnelle sont stockées dans des entrepôts ou, à une moindre échelle, dans des Datstores. Ces conteneurs stockent des sous-ensembles des informations d'une entreprise.

Parallèlement, les architectures BI font de plus en plus appel à des systèmes Hadoop en guise de référentiels ou de réceptacles pour données BI et analytiques, particulièrement dans le cas de données non structurées, de fichiers journaux, de données de capteurs et d'autres types de Big Data.

Avant leur utilisation dans les applications BI, les données brutes issues de différents systèmes source doivent être intégrées, consolidées et nettoyées. A cette fin, des outils d'intégration et de qualité des données garantissent que les analyses porteront sur des informations précises et homogènes.

En plus des responsables BI, les équipes d'informatique décisionnelle combinent généralement architectes et développeurs BI, analystes et professionnels de l'administration des données.

Les utilisateurs métier sont souvent présents eux aussi : ils représentent l'activité et s'assurent que le processus de développement BI répond bien à leurs besoins.

Pour contribuer à ce développement, un nombre croissant d'entreprises remplacent le développement en cascade classique par des approches Agile des entrepôts de données et de l'informatique décisionnelle. Ces approches font appel aux techniques de développement Agile pour scinder les projets BI en plusieurs parties et apporter aux utilisateurs de nouvelles fonctionnalités selon un mode incrémental et itératif. Les entreprises peuvent ainsi accélérer la mise en place de fonctions BI et optimiser ou modifier les plans de développement, à mesure que les besoins métier changent ou que de nouvelles exigences apparaissent et prennent le pas sur les précédentes.

Un peu d'Histoire

Equivalent anglais de l'intelligence décisionnelle, le terme de *Business Intelligence* faisait déjà l'objet d'un emploi sporadique dans les années 1860. Mais c'est le consultant Howard Dresner qui l'a proposé pour la première fois en 1989 pour désigner l'ensemble des techniques d'analyse des données venant à l'appui des processus de prise de décisions métier.

Les technologies qui allaient ensuite être connues sous l'appellation d'informatique décisionnelle, ou BI, sont dérivées de systèmes analytiques antérieurs, souvent exploités sur des mainframes, notamment les systèmes d'aide à la décision et les systèmes d'information destinés aux dirigeants (EIS, Exécutive Information Systèmes).

Entreposage de données(Data warehousing)

Définition:

L'entreposage de données est un ensemble de méthodes, de techniques et d'outils utilisés pour aider les travailleurs du savoir (cadres supérieurs, directeurs, gestionnaires et analystes) à effectuer des analyses de données qui aident à exécuter les processus de prise de décision et à améliorer les ressources d'information.

Entrepôt de données

Définition BILL Inmon : Un entrepôt de données est un ensemble de données qui soutient les processus de prise de décision. Il propose les fonctionnalités suivantes (Inmon, 2005) : « Collection de données orientées sujets, intégrées, non volatiles et historisées, organisée pour le support du processus d'aide à la décision »

- **Orienté sujet** : Un entrepôt de données est organisé autour des sujets clés (ou entités de haut niveau) de l'entreprise. Les principaux sujets peuvent inclure les clients, les patients, les étudiants et les produits. Les bases de données opérationnelles reposent sur de nombreuses applications spécifiques à l'entreprise.

- **Intégré** : les données hébergées dans l'entrepôt de données sont définies à l'aide de conventions de nommage, de formats, de structures de codage et de caractéristiques associées cohérents.

- **Variante temporelle** : les données de l'entrepôt de données contiennent une dimension temporelle afin qu'elles puissent être utilisées comme enregistrement historique de l'entreprise.

- **Non volatile** : les données de l'entrepôt de données sont chargées et actualisées à partir des systèmes opérationnels, mais ne peuvent pas être mises à jour par les utilisateurs finaux.

Définition de l'entrepôt de données Oracle :

Un entrepôt de données est une base de données relationnelle conçue pour les requêtes et l'analyse plutôt que pour le traitement des transactions. Il contient généralement des données historiques dérivées des données de transaction, mais peut inclure des données provenant d'autres sources. Les entrepôts de données séparent la charge de travail

d'analyse de la charge de travail de transaction et permettent à une organisation de consolider les données de plusieurs sources. CA aide:

- Tenir des registres historiques
- Analyser les données pour mieux comprendre l'entreprise et améliorer l'entreprise.

En plus d'une base de données relationnelle, un environnement d'entrepôt de données peut inclure une solution d'extraction, de transport, de transformation et de chargement (ETL), une analyse statistique, des rapports, des capacités d'exploration de données, des outils d'analyse client et d'autres applications qui gèrent le processus de collecte de données. , en les transformant en informations utiles et exploitables et en les livrant aux utilisateurs professionnels.

Contexte et problématique : entrepôts de données et de documents

Dans ce cadre, l'utilisation de bases de données multidimensionnelles (BDM) fournit un point de vue global sur les données d'une entité économique et permet aux décideurs d'avoir une vision détaillée des performances de cette entité grâce à un accès rapide et interactif aux données (Colliat, 1996). Les BDM, appelées magasins de données, agencent de manière multidimensionnelle les données issues de l'entrepôt afin de faciliter leur exploitation et analyse (cf. ?figure 1).

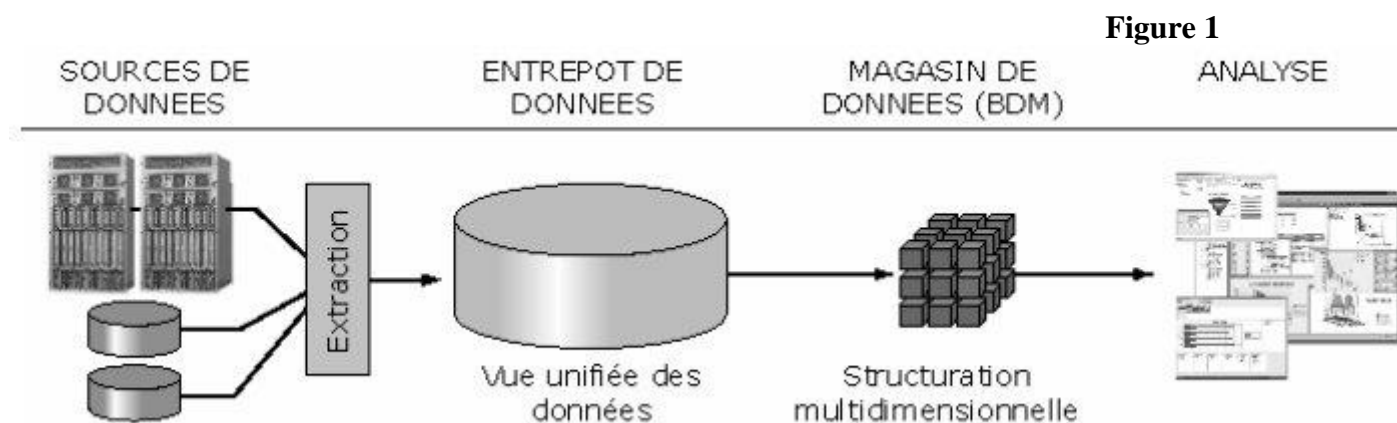


Figure 10

La modélisation multidimensionnelle (Kimball, 1996) représente les données comme des points dans un espace multidimensionnel avec la métaphore du cube (ou hypercube). Cette approche orientée utilisateur intègre structures et données multidimensionnelles dans la représentation du cube. Par exemple figure 11

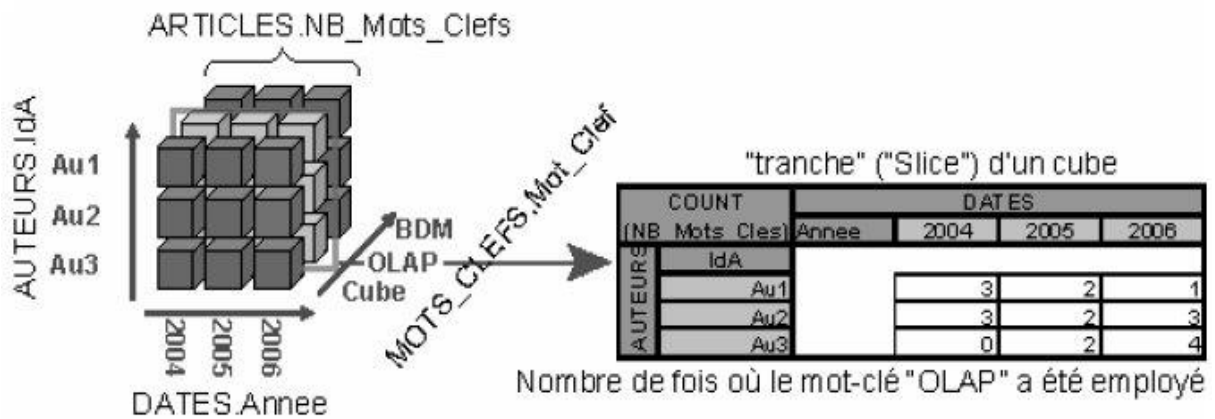


Figure 11

Objectif d'entrepôt de données (DW)

- L'objectif principal des systèmes d'entrepôt de données doit toujours être de définir correctement un processus pour transformer les données en informations.

Problématique

Comment répondre aux demandes des décideurs?

- En donnant un accès rapide et simple à l'information stratégique
- En donnant du sens aux données Mettre en place un système d'information dédié aux applications décisionnelles: un data warehouse

Les données utilisables par les décideurs :

- Données opérationnelles (de production)
- Bases de données (Oracle, SQL Server)

_ Fichiers, ...

_ Paye, gestion des RH, gestion des commandes...

■ Caractéristiques de ces données:

_ Distribuées: systèmes éparpillés

_ Hétérogènes: systèmes et structures de données différents

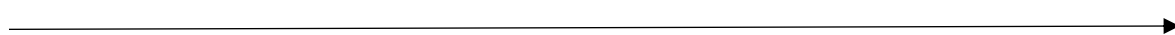
_ Détaillées: organisation des données selon les processus fonctionnels

_ Peu/pas adaptées à l'analyse : les requêtes lourdes peuvent bloquer le système **transactionnel**

_ Volatiles: pas d'historisation systématique

Champs d'application des systèmes décisionnels

Définir le problème	Rassembler les données	Analyser les données	Établir des solutions	Décider
---------------------	------------------------	----------------------	-----------------------	---------



Temps de prise d'une décision

Le processus de prise de décision :

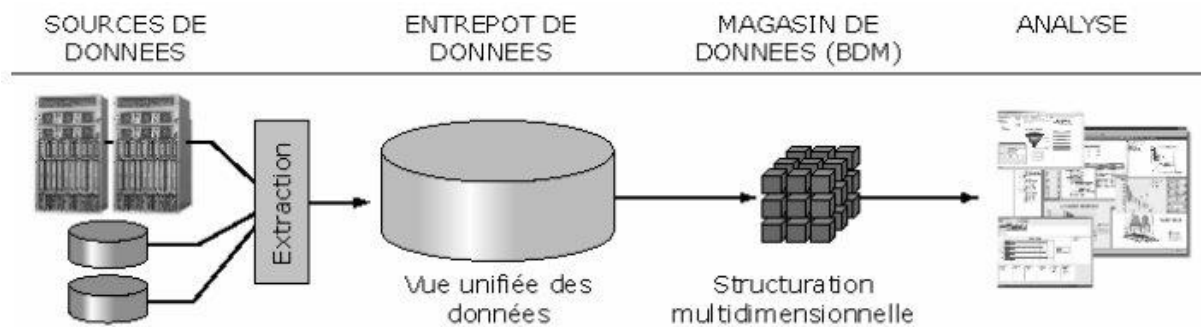


Figure12

Domaines d'utilisation des DW

- Banque
- Risques d'un prêt, prime plus précise
- Santé
- Épidémiologie
- Risque alimentaire
- Commerce
- Ciblage de clientèle
- Déterminer des promotions
- Logistique
- Adéquation demande/production
- Assurance
- Risque lié à un contrat d'assurance (voiture)
- ...

Quelques métiers du décisionnel

- **Strategic Performance Management**
Déterminer et contrôler les indicateurs clé de la performance de l'entreprise
- **Finance Intelligence**
 - Planifier, analyser et diffuser l'information financière. Mesurer et gérer les risques
- **Human Capital Management** (gestion de la relation avec les employés)
- Aligner les stratégies RH, les processus et les technologies.

- **Customer Relationship Management** (gestion de la relation client)

Améliorer la connaissance client, identifier et prévoir la rentabilité client, accroître l'efficacité du marketing client

- **Supplier Relationship Management** (gestion de la relation fournisseur)

Classifier et évaluer l'ensemble des fournisseurs. Planifier et piloter la stratégie Achat.

Magasine de donnée(Datamart)

Définition : « C'est un sous-ensemble de données dérivées du DW ciblé sur un sujet unique ».

Caractéristiques :

- Orienté vers un sujet unique Ex : comportement de la clientèle
- Données fortement agrégées (Le DW joue le rôle de source et d'historique pour le Datamart)
- Organisation multidimensionnelle (cubique), Dont l'une des dimensions indique souvent le temps
- Lien dynamique avec le DW (Association entre valeur agrégée et valeur détaillée)
- Interfaces simples et conviviales Approche multidimensionnelle : Modèle conceptuel
- Souvent représentés par une structure à plusieurs dimensions
- Une dimension est un attribut ou un ensemble d'attributs : (Temps, Géographie, Produits. Clients)
- Les cellules contiennent des données agrégées appelées Faits ou Indicateurs : (Nombre d'unités vendues, Chiffre d'Affaire, Coût).
 - Représentations : (Relations, Cube de données, hyper cube de données)
- Agrégation des données : Plusieurs niveaux d'agrégation
 - Les données peuvent être groupées à différents niveaux de granularité
 - Les regroupements sont pré-calculés, (par exemple, le total des ventes pour le mois dernier calculé à partir de la somme de toutes les ventes du mois.)

Granularité : niveau de détail des données emmagasinées dans un Datawarehouse.

Dimension

- Dimension = liste d'éléments
- Dimension contient des membres organisés en hiérarchie :
- Chacun des membres appartient à un niveau hiérarchique (ou niveau de granularité) particulier¹¹
- Granularité d'une dimension : nombre de niveaux hiérarchiques
- Temps : année – semestre – trimestre – mois

Architecture d'un système décisionnel : Three-Tier Architecture

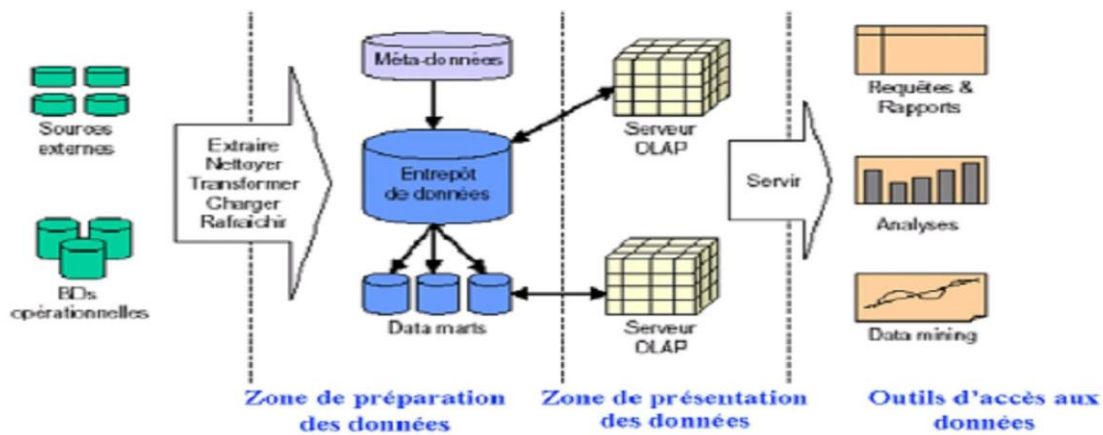


Figure 13 : Three-Tier Architecture

Le cycle de vie d'un projet décisionnel

Généralement, le développement d'un système décisionnel se fait en suivant l'une de ces deux méthodes : celle de Kimball ou celle d'Inmon.

La méthode d'Inmon est Top-Down approche, cependant Kimball méthode est Bottom-Up Approche.

6.1 Le cycle de vie d'un projet décisionnel/ Méthode KIMBALL

Un projet de système décisionnel se structure selon quatre grands axes :

1. Étude des besoins et de l'existant
 - Étude des besoins utilisateurs
 - Étude des données existantes
2. Modélisation et conception
 - Modélisation dimensionnelle
 - Architecture technique
 - Spécification des outils d'exploitation
3. Implémentation du data warehouse
 - Implémentation du DW et des DM
 - Mise en place de l'ETL
4. Implémentation des outils d'exploitation
 - Implémentation des outils de reporting
 - Implémentation des outils d'exploration
 - Implémentation des outils de prédiction

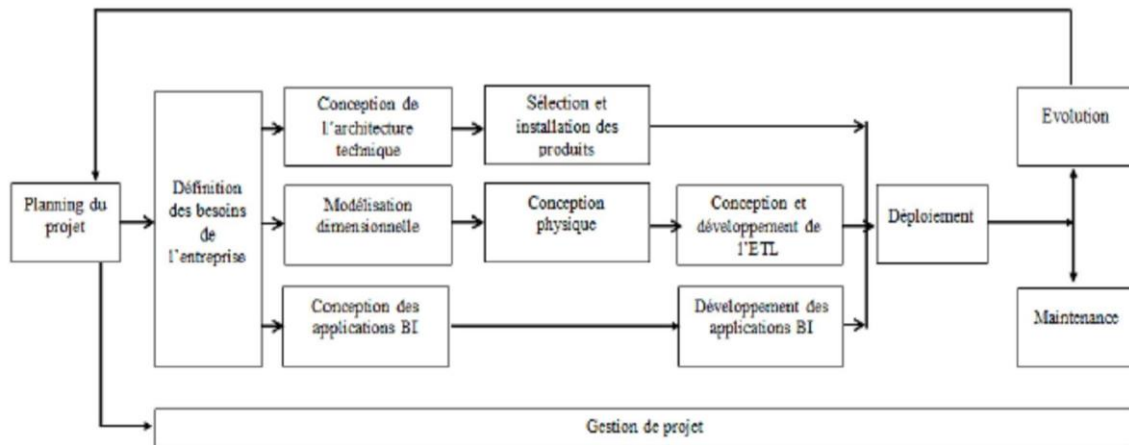


Figure 8 Le cycle de vie d'un projet décisionnel (KIMBALL & ROSS, THE Data Warehouse Toolkit : The Definitive Guide, Figure 27)

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

- La conception de l'architecture technique : on choisit l'architecture technique et les outils matériels et logiciels nécessaires pour la mise en place de l'entrepôt de données.
- La modélisation dimensionnelle : on définit la modélisation dimensionnelle de l'entrepôt et des magasins de données, et on définit les outils ETL.
- Conception des applications BI: développement des applications décisionnelles tels que : les rapports, les tableaux de bord...

Ces trois chemins convergent en fin du projet au moment du déploiement.

Le processus complet est répété pour chaque nouveau magasin de données demandé par les utilisateurs finaux tout en assurant l'évolution et la maintenabilité du système.

6.2 Le cycle de vie d'un projet décisionnel/ Méthode INMON

La méthode d'Inmon est basée sur les points suivants (INMON W. , 2002)

- La création de l'entrepôt de données, ce dernier permettra d'alimenter les autres magasins de données des différents départements ce qui garantit la cohérence de données.
- Une connaissance approfondie du domaine métier de l'entreprise. □ La modélisation de données repose sur le modèle Entité/Association et la normalisation en 3ème forme normale.

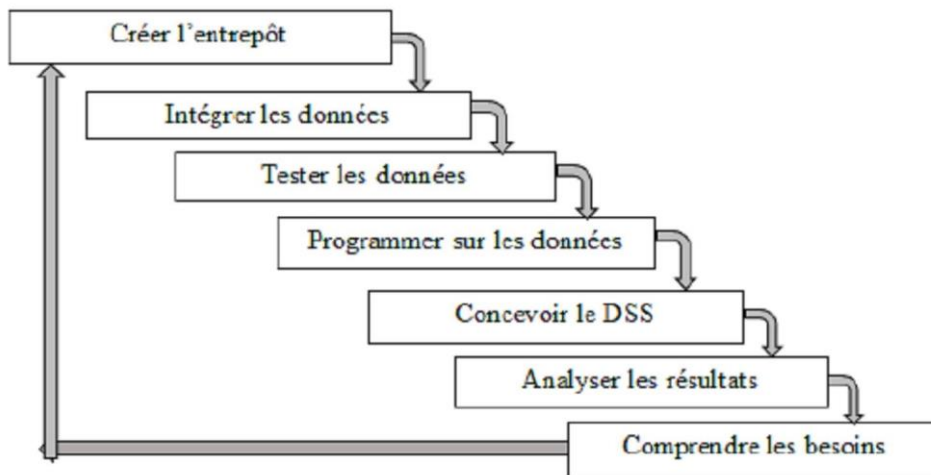


Figure 9 Le cycle de vie d'un projet décisionnel selon la méthode Inmon (INMON W. , 2002)

Figure 28

Selon le schéma ci-dessus, la création de l'entrepôt est la première étape du projet, et selon les résultats de la conception retenus à la fin du cycle, les besoins définis par le client seront bien compris par l'équipe projet ce qui permet de reprendre le cycle du projet pour permettre l'évolution de l'entrepôt de données et non pas sa révolution.

Le cycle de vie de développement d'entrepôt de données

PROCESSUS

Que chacun construise ce qu'il veut, on intégrera ce qu'il faudra quand il faudra!

On ne fait rien tant que tout n'est pas désigné, le datawarehouse doit être exhaustif!



Ralph Kimball
Kimball Group
www.kimballgroup.com



Bill Inmon
Corporate Information Factory
www.inmoncif.com



(SDLC) de KIMBALL vs Immon Un cycle de vie de développement d'entrepôt de données est une séquence de tâches de haut niveau requises pour une conception, un développement et un déploiement

efficaces d'entrepôt de données. Différents auteurs ont des points de vue radicalement différents sur l'ordre de ces tâches dans le cycle de développement des entrepôts de données.

Ces différences sont présentées dans les paragraphes suivants.

- Inmon préconise l'utilisation d'une méthode basée sur les données. Cela signifie que le traitement DSS commence par les données et se termine par les exigences.
- Inmon appelle cette méthode le CLDS (l'inverse de SDLC) .
- Selon Inmon, un entrepôt de données commence par la création d'un magasin de données central pour un domaine, qui est alimenté à partir de systèmes opérationnels. En tant que capacité analytique du pour un entrepôt de données et pour définir les besoins en personnel pour l'équipe d'entreposage de données.
- Au fur et à mesure que la capacité d'analyse du nouvel entrepôt de données est découverte, la demande pour un magasin de données intégré pour un autre domaine augmentera et ce processus se répétera jusqu'à ce qu'un entrepôt de données complet ait été développé.
- Contrairement à l'approche d'Inmon, Kimball préconise l'utilisation d'une méthode axée sur les exigences. L'entrepôt de données commence par la planification du projet pour déterminer l'état de préparation de l'organisation
- Une compréhension claire des exigences de l'entreprise est le facteur de succès le plus important, et Kimball déclare que ce processus de collecte des exigences diffère considérablement de l'analyse des exigences basée sur les données.
- Les exigences opérationnelles établissent la base des trois volets parallèles axés sur la technologie, les données et les applications d'utilisateur final.

Entreposage de données SDLC par rapport au traitement des transactions en ligne

Inmon présente une approche totalement différente du développement d'un système d'entrepôt de données. Il soutient que bien que l'OLTP soit développé à partir des exigences comme point de départ,

l'entreposage de données commence par la mise en œuvre de l'entrepôt de données et se termine par une compréhension claire des exigences. Le cycle de vie de développement de l'entrepôt de données est axé sur les données et OLTP est axé sur les exigences.

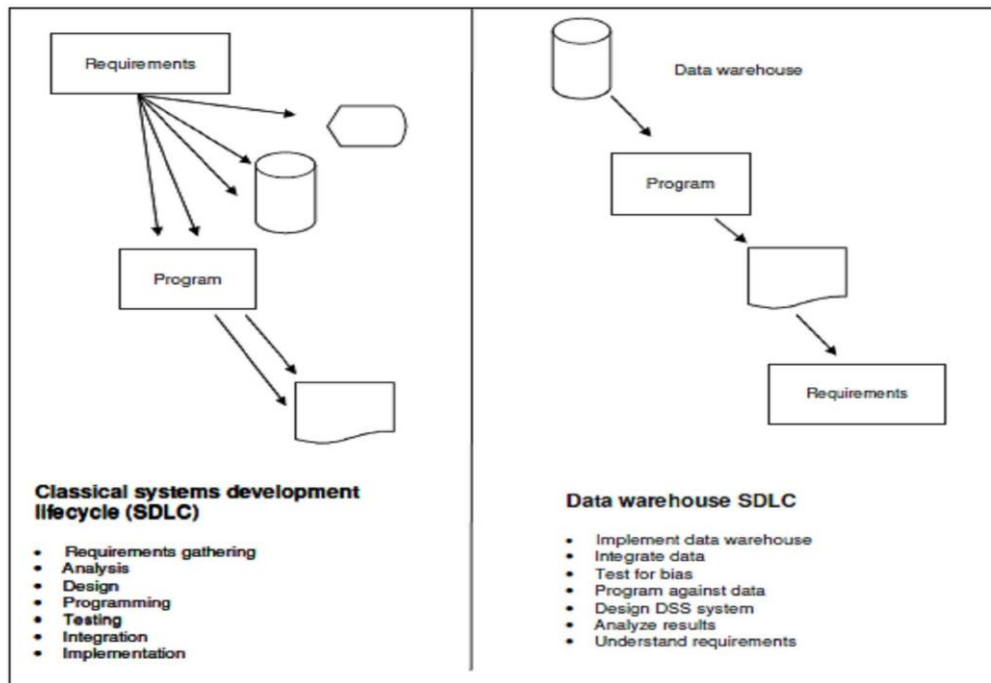


Table 4-2 The SDLC for OLTP vs. OLAP systems (Inmon, 1996:24)

Figure 29

Chapitre 3

Analyse multidimensionnelle via OLAP

Introduction

L'essor des technologies de l'information a considérablement accru la quantité de données et de documents numériques disponibles. Le volume des informations a été démultiplié et de par sa taille, ce volume rend la masse d'informations difficilement appréhendable. Toutefois, les systèmes de traitement d'analyse en ligne OLAP (On-Line Analytical Processing) (Codd et al., 1993) permettent l'analyse de grands volumes d'informations en générant une vision synthétique des données.

La modélisation multidimensionnelle (Kimball, 1996) représente les données comme des points dans un espace multidimensionnel avec la métaphore du cube (ou hypercube). Cette approche orientée utilisateur intègre structures et données multidimensionnelles dans la représentation du cube. Par exemple, en figure 5 .

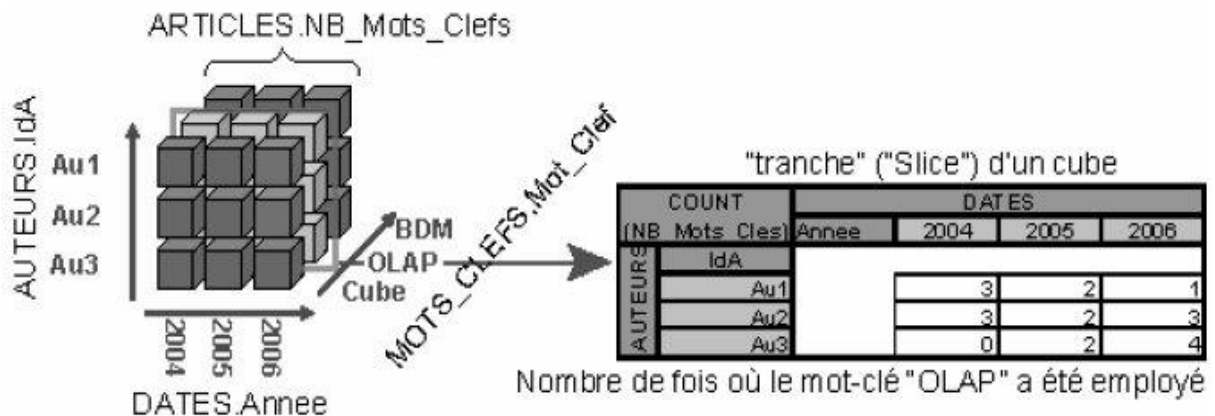


Figure 14

Approche multidimensionnelle : Modèle conceptuel

- Souvent représentés par une structure à plusieurs dimensions
- Une dimension est un attribut ou un ensemble d'attributs: (Temps, Géographie, Produits, Clients)
- Les cellules contiennent des données agrégées appelées Faits ou Indicateurs: (Nombre d'unités vendues, Chiffre d'Affaire, Coût).
- Représentations: (Relations, Cube de données, hyper cube de données)
- Agrégation des données :Plusieurs niveaux d'agrégation
 - Les données peuvent être groupées à différents niveaux de granularité
 - Les regroupements sont pré-calculés, (par exemple, le total des ventes pour le mois dernier calculé à partir de la somme de toutes les ventes du mois.)

Granularité : niveau de détail des données emmagasinées dans un Datawarehouse.

Granularité des dimensions

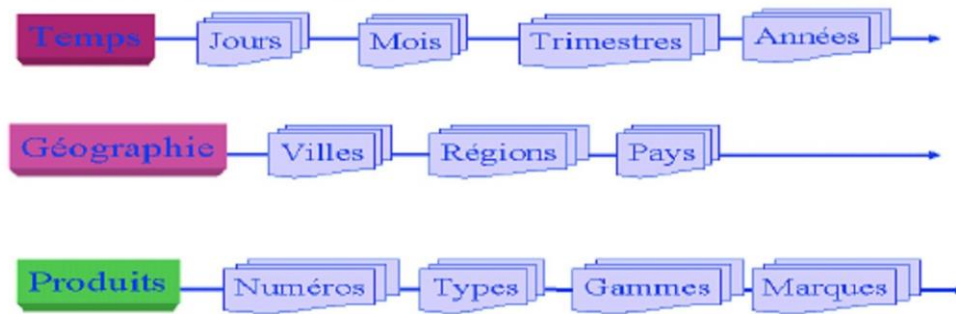


Figure 15 Granularité des dimensions

Dimension

- Dimension = liste d'éléments
- Dimension contient des membres organisés en hiérarchie :
- Chacun des membres appartient à un niveau hiérarchique (ou niveau de granularité) particulier
- Granularité d'une dimension : nombre de niveaux hiérarchiques
- Temps : année – semestre – trimestre – mois
- Les axes de dimension doivent fournir des règles de calcul d'agrégat pour chaque mesure:

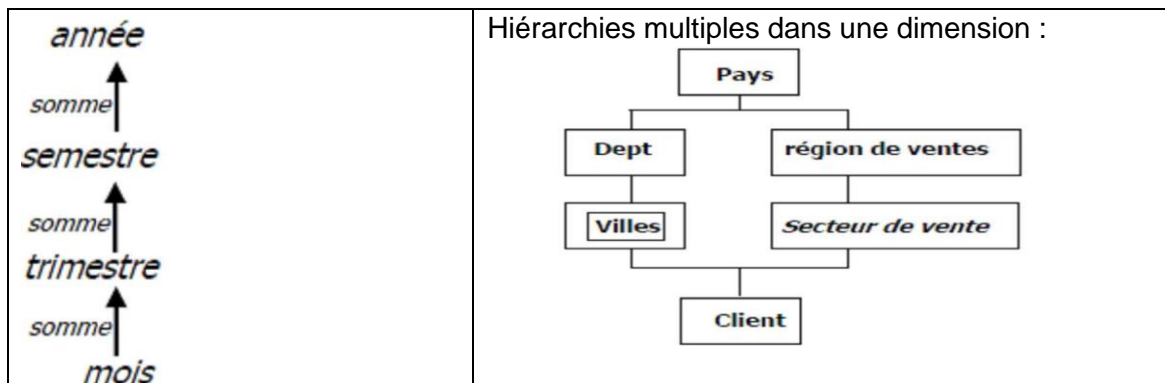


Figure 16

E/R Model vs dimensionnel Model

- La modélisation E/R est une technique de conception dans laquelle nous stockons les données sous une forme hautement normalisée dans une base de données relationnelle.

- Une table est en troisième forme normale (3NF) si chaque colonne non-clé est indépendante des autres colonnes non-clé et ne dépend que de la clé.
- Le modèle E/R se concentre essentiellement sur trois choses, les entités, les attributs et les relations.
- L'objectif de la normalisation est de minimiser la redondance en n'ayant pas les mêmes données stockées dans plusieurs tables. Par conséquent, la normalisation peut minimiser les problèmes d'intégrité, car les mises à jour SQL ne doivent alors être appliquées qu'à une seule table. Cependant, les requêtes, en particulier celles impliquant de très grandes tables, qui incluent une jointure des données stockées dans plusieurs tables normalisées peuvent nécessiter • des efforts et une programmation supplémentaires pour obtenir des performances acceptables.
- E/R et modélisation dimensionnelle, bien que liées, sont extrêmement différentes. Bien entendu, tous les modèles dimensionnels sont également de véritables modèles E/R. Cependant, lorsque nous parlons de modèles E/R, nous entendons des modèles E/R normalisés. Les modèles dimensionnels sont dénormalisés.
- Lorsque vous travaillez avec des applications d'entreposage de données, la modélisation E/R peut être utile pour les rapports et les requêtes fixes, mais la modélisation dimensionnelle est généralement meilleure pour les requêtes et analyses ad hoc (ad hoc signifie quelque chose qui n'est pas prévu mais qui peut l'être).
- personnalisé sur le moment pour atteindre n'importe quel objectif
- Les applications d'entreposage de données sont pour la plupart en lecture seule et peuvent donc généralement bénéficier de la dénormalisation. La dénormalisation est une technique qui consiste à dupliquer des données dans une ou plusieurs tables pour minimiser ou éliminer les jointures chronophages.
- Un modèle dimensionnel est aussi communément appelé schéma en étoile.

- Le modèle dimensionnel se compose de deux types de tableaux ayant des caractéristiques différentes. Ce sont : la table des faits et la table des dimensions.

- Types de modèles dimensionnels

- Il existe trois types de base de modèles dimensionnels, et ils sont : modèle en étoile, modèle en flocon de neige modèle multi-étoiles.

Modèles de représentation dimensionnelle:

- Cubes . Hyper Cube
- Étoile & flocon

Cube Hyper Cube

Hyper cube est BD multidimensionnelle

- Axes: dimensions (date, type de produits, région),
- Chaque cellule de l'hyper cube contient une mesure calculée (vente de produit).

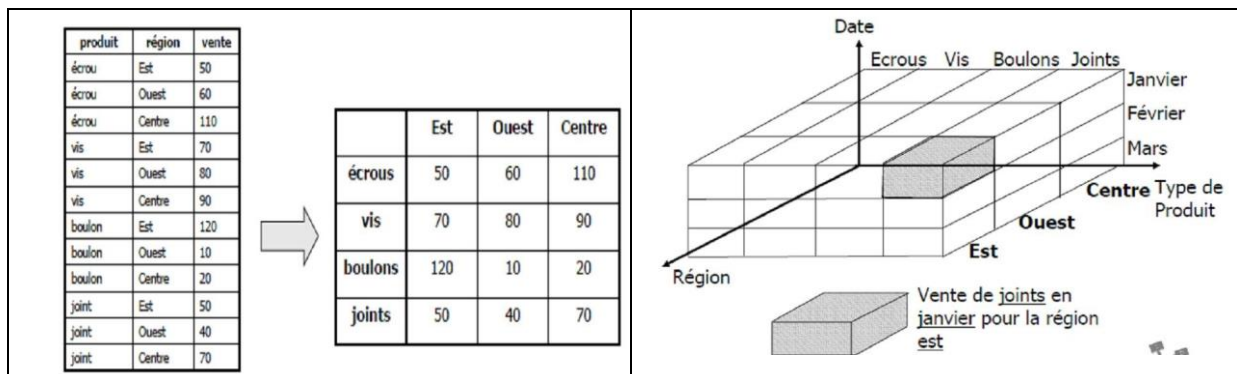


Figure17

Qu'est-ce qu'OLAP ?

OLAP « Processus analytique en ligne »

C'est la technologie derrière de nombreuses applications de Business Intelligence (BI). OLAP est une technologie puissante pour la découverte de données, y compris des capacités de visualisation de

rappports illimitées, des calculs analytiques complexes et un scénario prédictif « et si » (budget, prévisions)

Planification

Opérations OLAP (Manipulation d'un cube)

Opérateurs appliqués sur le cube sont algébriques (le résultat est un autre cube) et peuvent être combinés.

les Opérateurs sont:

- Slicing & Dicing (extraction) *Opérateur sur le cube*

- o Trancher et dés est l'un des termes les plus abusés dans la littérature sur les entrepôts de

données.

- o Nous définissons le slicing comme une opération qui réduit le nombre de dimensions du cube après avoir défini l'une des dimensions à une valeur spécifique.

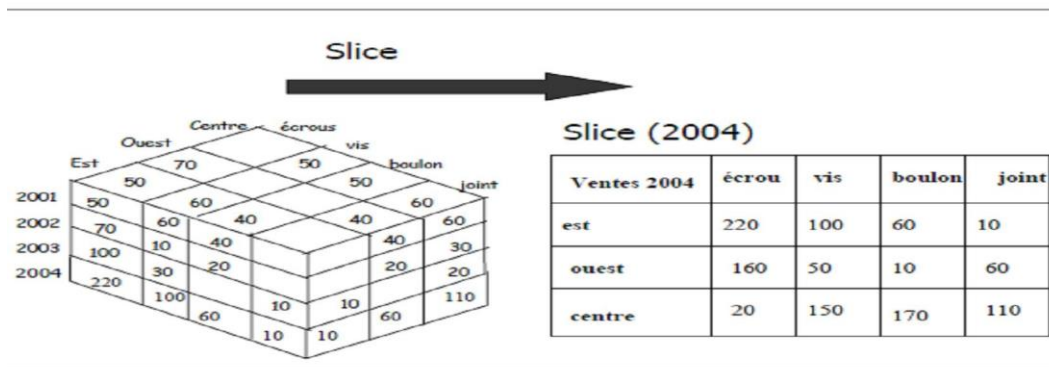
- o Le découpage est une opération qui réduit l'ensemble des données analysées

- Changement de la granularité d'une dimension Opérateurs sur les dimensions

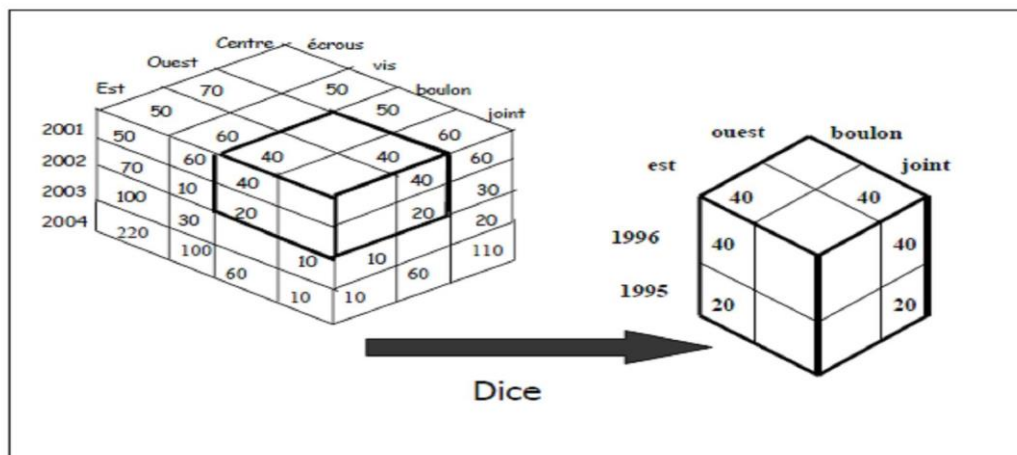
- o Roll up (agrégation d'une dimension => CV)

- o Explorer (plus détaillé)

Exemple: slicing



Exemple: Dicing



Opérations liées à la granularité

Les opérations agissant sur la granularité d'observation des données caractérisent la hiérarchie de navigation entre les différents niveaux.

- Roll-up : consiste à représenter les données du cube à un niveau de granularité supérieur conformément à la hiérarchie définie sur la dimension. (Utilisation de la fonction d'agrégation (somme, moyenne, etc) spécifiée pour la mesure et la dimension)
- Drill-down : consiste à représenter les données du cube à un niveau de granularité de niveau inférieur, donc sous une forme plus détaillée.

Drill-Down exemple

Drill-down ~ opération inverse de Roll-up
Drill-down du niveau des régions au niveau villes

	joiat	bordeaux	dijon	grenoble	lille	lyon	marseille	montpellier	nantes	paris	poitiers
vis	20	30	30	10	20	30	30	20	50	10	10
boulon	10	30	30	40	20	30	40	10	10	70	10
écrous	30	20	30	40	20	20	10	10	10	10	10
2001	30	20	30	30	20	50	10	10	10	10	10
2002	30	20	20	30	30	20	50	10	10	10	10
2003	10	10	20	60	20	40	20	30	10	20	10
2004	10	20	10	10	70	40	10	10	10	10	10

Modélisation en étoile, en flacons (snowflake), constellation (multi star)

- Modélisation conceptuelle pour BD : entité et relation
- Modélisation pour DW : dimension et mesure
- Les mesures sont les valeurs numériques que l'on compare (ex : montant ventes, qte_vendue). Ces valeurs sont le résultat d'une opération d'agrégation des données
- Les dimensions sont les points de vues depuis lesquels les mesures peuvent être observées : Ex : date, localisation, produit, etc.

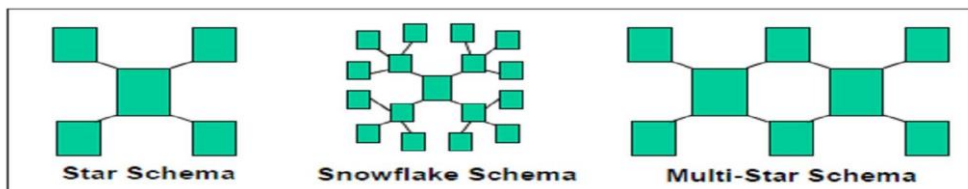


Figure 18
Les dimensions

- Une dimension peut être définie comme : un thème, ou un axe (attributs), selon lequel les données seront analysées. Ex : Temps, Découpage administratif, Produits
- Une dimension contient des membres organisés en hiérarchie : Chacun des membres appartient à un niveau hiérarchique (ou niveau de granularité) particulier. Ex : pour la dimension Temps: année – semestre – mois – jour

Les mesures

□ Une mesure est un élément de donnée sur lequel portent les analyses, en fonction des différentes dimensions: Ex : coût des travaux, nombre d'accidents, ventes

Les faits

Un fait représente la valeur d'une mesure, mesurée ou calculée, selon un membre de chacune des dimensions.

Exemple : «250 000 euros » est un fait qui exprime la valeur de la mesure « coût des travaux » pour le membre « 2002 » du niveau année de la dimension « temps » et le membre « Versailles » du niveau « ville » de la dimension « découpage administratif ».

La table de faits

- Les mesures sont stockées dans les tables de faits
- Table de fait contient les valeurs des mesures et les clés vers les tables de dimensions

Modèle logique de données en étoile

- Objectifs du modèle dimensionnel La modélisation par schéma en étoile, par opposition aux schémas normalisés en 3NF.
- permet de répondre à deux besoins caractéristiques des systèmes décisionnels : la performance et la simplicité des requêtes.

Le modèle en étoile

- Une (ou plusieurs) table(s) de faits comprenant une ou plusieurs mesures.
- Plusieurs tables de dimension dénormalisées: descripteurs des dimensions.
- Les tables de dimension n'ont pas de lien entre elles *Avantages* :
- Facilité de navigation
- Performances : nombre de jointures limité ; gestion des données creuses.
- Gestion des agrégats

Inconvénients :

- Toutes les dimensions ne concernant pas les mesures □ Redondances dans les dimensions □ Alimentation complexe.

Exemple

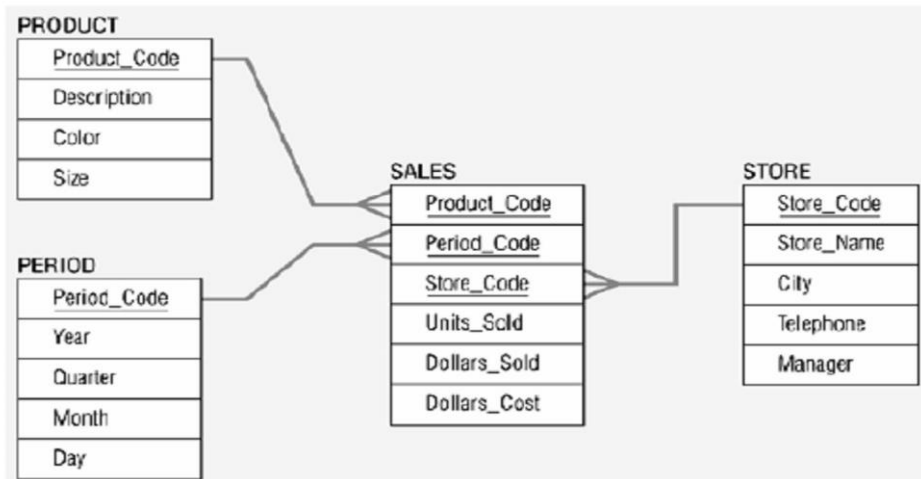


Figure 19

Exemple de modèle en étoile

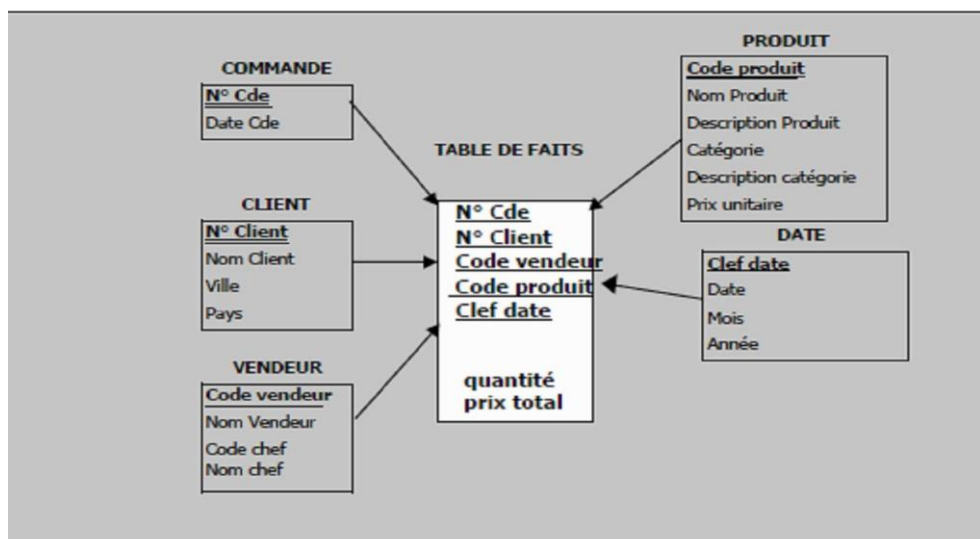


Figure 20

Le modèle en flocons (snowflake)

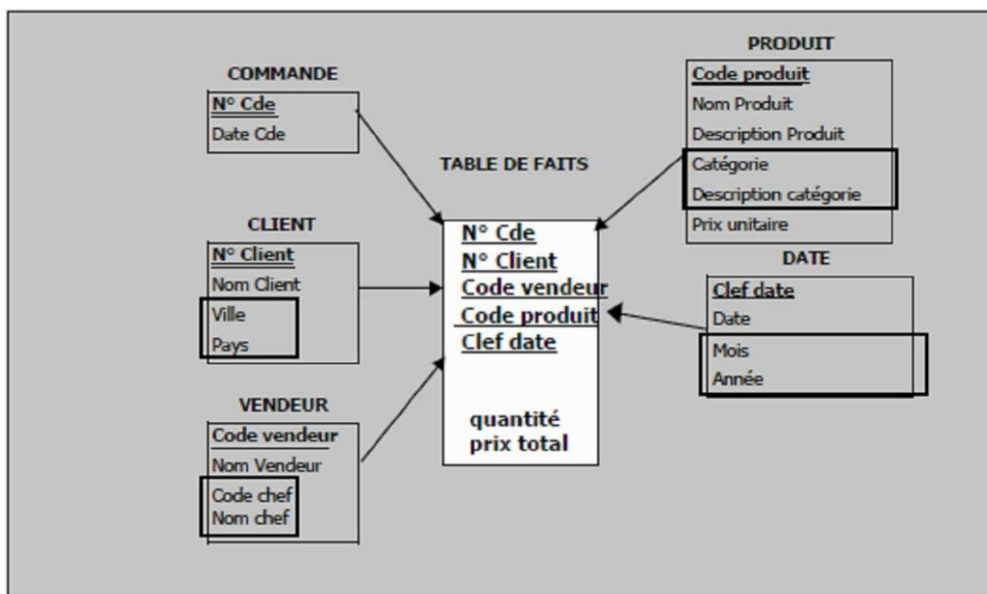
- Le schéma en flocon est dérivé du schéma en étoile où les tables de dimensions sont normalisées (la table des faits reste inchangée).
- Avec ce schéma, chacune des dimensions est décomposée selon sa (ou ses) hiérarchie(s). Exemple : Commune, Département, Région, Pays, Continent

- Modèle en étoile + normalisation des dimensions Lorsque les tables sont trop volumineuses *Avantages* :
- réduction du volume, □ Inconvénients :
- navigation difficile,
- nombreuses jointures.

Figure 21

Client	Continent	Pays	Region	Département	Commune
Pepone	Europe	France	RhôneAlpes	Rhône	Lyon1
Testut	Europe	France	RhôneAlpes	Rhône	Lyon2
Soinin	Europe	France	RhôneAlpes	Rhône	Lyon3
Vepont	Europe	France	Ile de France	Paris	Paris1
Martin	Europe	France	Ile de France	Paris	Paris2
Elvert	Europe	France	Ile de France	Yvelines	Versailles

Modèle en flocon



Le modèle en flocons

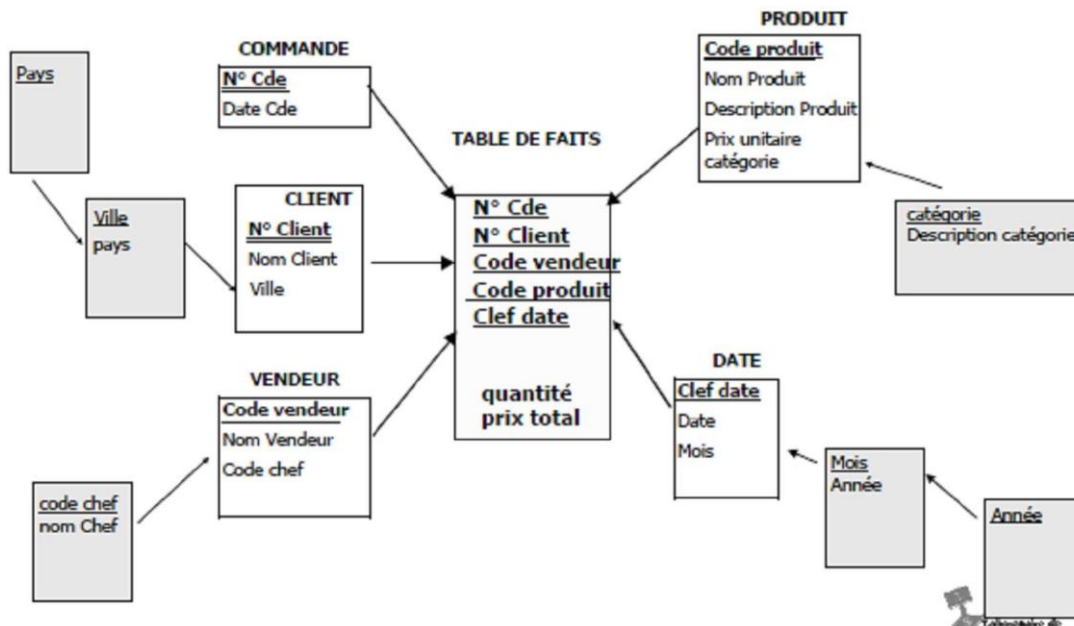


Figure 22

Data warehouse Implémentation :

Approches pour créer un DW 3 possibilités:

(1) Relational OLAP (ROLAP)

- Données sont stockées dans un SGBD relationnel

- adopter

ROLAP nécessite un middleware spécialisé, également appelé moteur multidimensionnel, entre les serveurs back-end relationnels et les composants frontaux.

- Un moteur OLAP permet de simuler le comportement d'un SGBD multidimensionnel
- Le middleware reçoit les requêtes OLAP formulées par les utilisateurs dans un outil front-end et les transforme en instructions SQL pour une application back-end relationnelle avec le support de méta-données par un critère de sélection.
- Des vues (matérialisées) sont utilisées pour la représentation multidimensionnelle

- Utilisation d'index spéciaux : bitmap

Avantages/inconvénients

- Souplesse, évolution facile, permet de stocker de gros volumes.
- Mais peu efficace pour les calculs complexes

(2) Multidimensional OLAP (MOLAP)

□ Structure de stockage en cube

- Accès direct aux données dans le cube
- Le cube est indexé sur ses dimensions *Avantages/inconvénients:*
- rapide
- formats propriétaires
- ne supporte pas de très gros volumes de données

(3) Hybrid OLAP (HOLAP)

- Données stockées dans des tables relationnelles
 - Données agrégées stockées dans des cubes.
- Les requêtes vont chercher les données dans les tables et les cubes
 - Les requêtes vont chercher les données dans les tables et les cubes

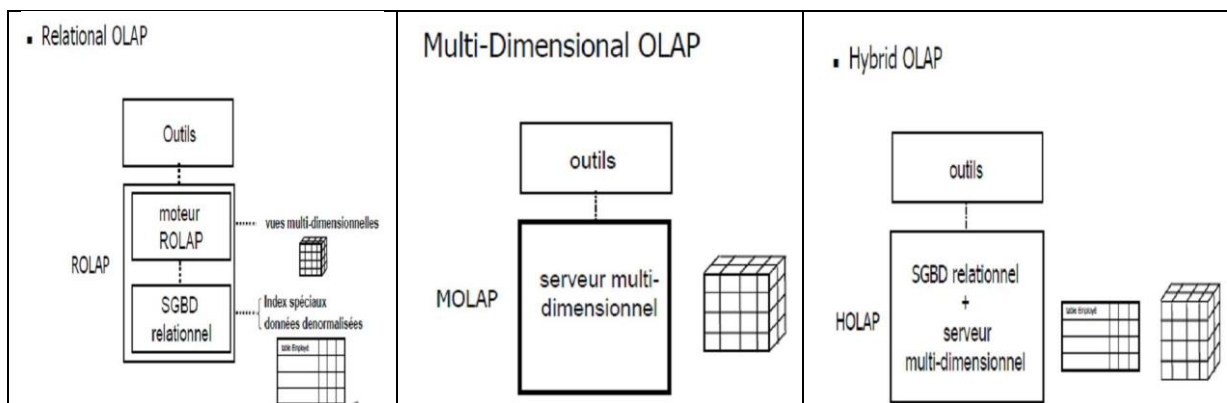


Figure 23

Architectures d'entrepôt de données

La littérature scientifique distingue souvent cinq types d'architecture pour les systèmes d'entrepôt de données, dans lesquels les mêmes couches de base mentionnées dans les paragraphes précédents sont combinées de différentes manières

1. Architecture de data marts indépendante

- Dans une architecture de data marts indépendante, différents data marts sont conçus et construits séparément de manière non intégrée (Figure 1-5).
- Cette architecture peut être initialement adoptée en l'absence d'un parrainage fort envers un projet d'entreposage à l'échelle de l'entreprise, ou lorsque les divisions organisationnelles qui composent l'entreprise sont faiblement couplées.
- Cependant, elle tend à être rapidement remplacée par d'autres architectures qui réalisent mieux l'intégration des données et le reportant croisé.

FIGURE 1-5
Independent data
marts architecture

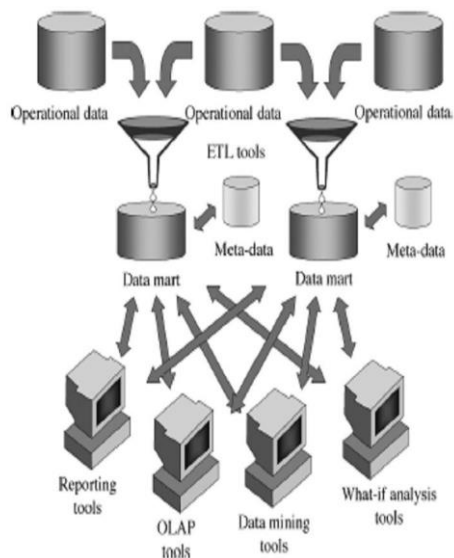


Figure 24

2. L'architecture de bus

, recommandée par Ralph Kimball, est apparemment similaire à l'architecture précédente, avec une différence importante. Un ensemble de base de dimensions conformes (c'est-à-dire des dimensions d'analyse qui conservent la même signification dans tous les faits auxquels elles appartiennent) est

adopté et partagé en tant que directive de conception commune. Cela garantit l'intégration logique des magasins de données et une vue des informations à l'échelle de l'entreprise.

3. L'architecture en étoile

- Dans l'architecture hub-and-spoke, l'une des plus utilisées dans des contextes de taille moyenne à grande, une grande attention est accordée à l'évolutivité et à l'extensibilité, ainsi qu'à l'obtention d'une vue des informations à l'échelle de l'entreprise.
- Les données atomiques normalisées sont stockées dans une couche réconciliée qui alimente un ensemble de magasins de données contenant des données résumées sous forme multidimensionnelle .
- Les utilisateurs accèdent principalement aux magasins de données, mais ils peuvent occasionnellement interroger la couche réconciliée.

FIGURE 1-6
Hub-and-spoke
architecture

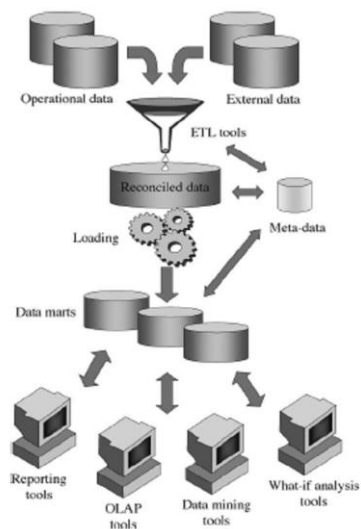


Figure 25

4. L'architecture centralisée, recommandée par Bill Inmon, peut être considérée comme une implémentation particulière de l'architecture en étoile, où la couche réconciliée et les magasins de données sont regroupés dans un seul référentiel physique.

5. L'architecture fédérée

- L'architecture fédérée est parfois adoptée dans des contextes dynamiques où les entrepôts de données/data marts préexistants doivent être intégrés de manière non invasive pour fournir un environnement d'aide à la décision unique et interorganisationnel (par exemple, dans le cas de fusions et acquisitions).
- Chaque entrepôt de données/entrepôt de données est virtuellement ou physiquement intégré aux autres, s'appuyant sur une variété de techniques avancées telles que les requêtes distribuées, les ontologies et l'interopérabilité des métadonnées.

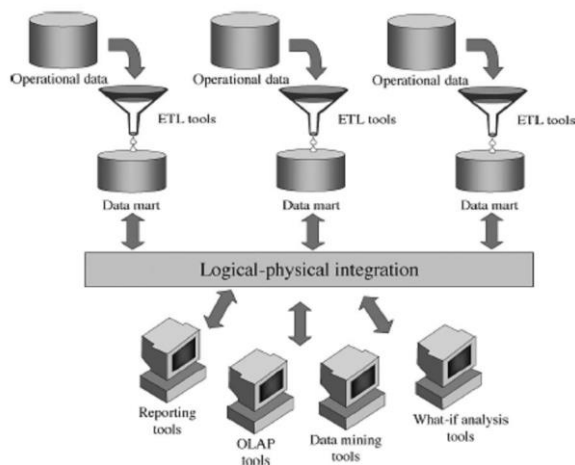


FIGURE 1-7 Federated architecture

Figure 26

Choisir l'une de ces architectures

La liste suivante comprend les facteurs qui sont particulièrement influents lorsqu'il s'agit de choisir l'une de ces architectures :

- La quantité d'informations interdépendantes échangées entre les unités organisationnelles d'une entreprise et le rôle organisationnel joué par le promoteur du projet d'entrepôt de données peuvent conduire à la mise en œuvre d'architectures à l'échelle de l'entreprise, telles que des architectures de bus, ou des architectures spécifiques au département, telles que des données indépendantes. marchés.
- Un besoin urgent d'un projet d'entrepôt de données, des restrictions sur les ressources économiques et humaines, ainsi que de faibles compétences du personnel informatique peuvent suggérer qu'un type d'architecture « rapide », comme des magasins de données indépendants, devrait être mis en œuvre.

- Le rôle mineur joué par un projet d'entrepôt de données dans les stratégies d'entreprise peut vous faire préférer un type d'architecture basé sur des magasins de données indépendants plutôt qu'un type d'architecture en étoile.

Types de faits

Il existe trois types de faits :

1 Faits additifs

Les faits additifs peuvent être utilisés avec n'importe quelle fonction d'agrégation comme Sum(), Avg() etc. L'exemple est la quantité, le montant des ventes, etc.

2. Faits semi-additifs

Les faits semi-additifs sont ceux où seules quelques fonctions d'agrégation peuvent être appliquées.

Par exemple, considérez les détails du compte bancaire. Vous ne pouvez pas appliquer la Sum() sur le solde bancaire qui ne donne pas de résultats utiles mais les fonctions min() et max() peuvent renvoyer des informations utiles.

3. Faits non additifs

Vous ne pouvez pas utiliser de fonctions d'agrégation numériques telles que Sum(), Avg(), etc. sur des faits non additifs.

Par exemple, un fait non additif est tout type de rapport ou de pourcentage. Les faits non numériques peuvent également être des faits non additifs.

Types de tables de faits

1. Tableaux de faits sans faits

Une table de faits qui ne contient aucune mesure est une table de faits sans faits. Cette table ne contiendra que les clés de différentes tables de dimension. Ceci est souvent utilisé pour résoudre un problème de cardinalité plusieurs-à-plusieurs.

Par exemple, une table de faits qui n'a que productID et date key est une table de faits sans faits. Voici des exemples courants de tables de faits sans faits :

- Identification des événements de promotion des produits (pour déterminer les produits promus qui ne se sont pas vendus)
- Suivi de la présence des étudiants ou des événements d'inscription
- Suivi des accidents liés à l'assurance
- Identifier les horaires des bâtiments, des installations et des équipements pour un hôpital ou une université

Les tableaux de faits sans faits sont utilisés pour suivre un processus ou collecter des statistiques. Elles sont appelées ainsi car la table de faits n'a pas de valeurs numériques ni d'informations abrégées. Il existe deux types de tables de faits sans faits : celles qui décrivent des événements et celles qui décrivent des conditions. Les deux peuvent jouer un rôle important dans vos modèles dimensionnels.

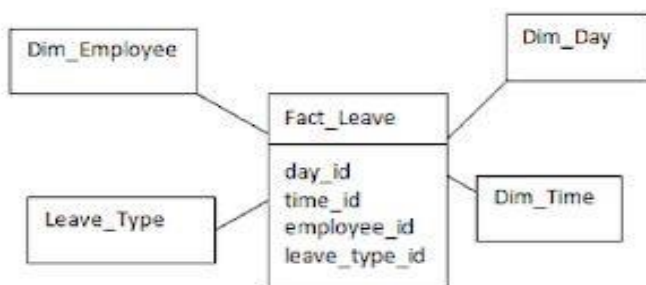


Fig 29 exemple modèle Dimension

Le fait ci-dessus est utilisé pour capturer le congé pris par un employé. Chaque fois qu'un employé prend un congé, un enregistrement est créé avec les dimensions. En utilisant le fait

FACT_LEAVE nous pouvons répondre à de nombreuses questions telles que

Nombre de congés pris par un salarié

Le type de congé que prend un employé

Détails de l'employé qui a pris un congé

Types des Dimensions

1. Dimensions qui changent lentement
2. Dimensions en évolution rapide
3. Dimensions indésirables
4. Dimensions inférées
5. Dimensions conformes
6. Dimensions dégénérées
7. Dimensions du jeu de rôle
8. Dimensions réduites
9. Dimension Statique

Chapitre 4

Réalisation power BI

Introduction

Nous utilisons l'environnement de travail Power BI pour réaliser donnée « Arrêt extrusion » en forme Excel de la société algal +

Nous utilisons un tableau de 5 éléments cette information ce forme numérique données intitulé par

1. presse
2. time début
- 3 .time fin
4. type arrêt
5. Temp arrêt

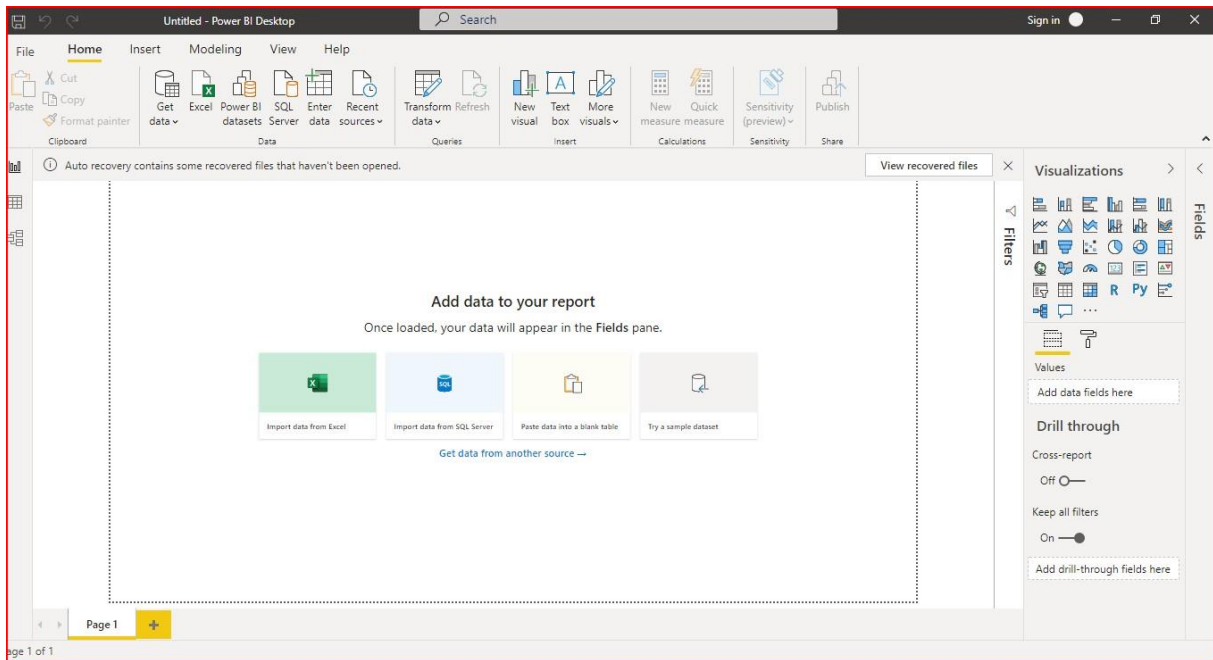


Figure 30

The screenshot shows the ALGALPLUS software interface. The title bar reads 'ALGALPLUS'. The interface includes a menu bar with options like 'Refont', 'Extrusion', 'PSPF', 'Laguage', 'Commercial', 'Anodisation', 'Faux Bois', 'Labo', 'T.Eaux', and 'Correction.F'. The main area displays a list of items and a data table. The table has columns for 'Type de Billette', 'Taux Declasse', 'Week', 'Heurs de Marche', 'Cadence Horaire Brute', 'Cadence Horaire Nette', 'OBJECTIF', 'ConsBilletes (moy)', 'Production Nette (moy)', and 'Cons Billetes'. The table displays data for years 2013 through 2020, with a 'Grand Total' row at the bottom.

Type de Billette	Taux Declasse	Week	Heurs de Marche	Cadence Horaire Brute	Cadence Horaire Nette	OBJECTIF	ConsBilletes (moy)	Production Nette (moy)	Cons Billetes
HEURSP	Production Declasse								
	Production Nette								
annee	Mois	Date	1600	2500	Grand Total				
2013			1 965 026,00	3 590 650,00	5 555 676,00				
2014			2 229 518,00	2 633 239,00	4 862 757,00				
2015			2 829 538,00	3 407 112,00	6 236 650,00				
2016			2 808 831,00	3 874 700,00	6 683 531,00				
2017			2 552 911,00	3 858 798,00	6 411 709,00				
2018			3 584 720,00	4 521 619,00	8 106 339,00				
2019			3 199 515,00	4 429 286,00	7 628 801,00				
2020			21 305,00	20 555,00	41 860,00				
Grand Total			19 191 364,00	26 335 959,00	45 527 323,00				

Figure 31

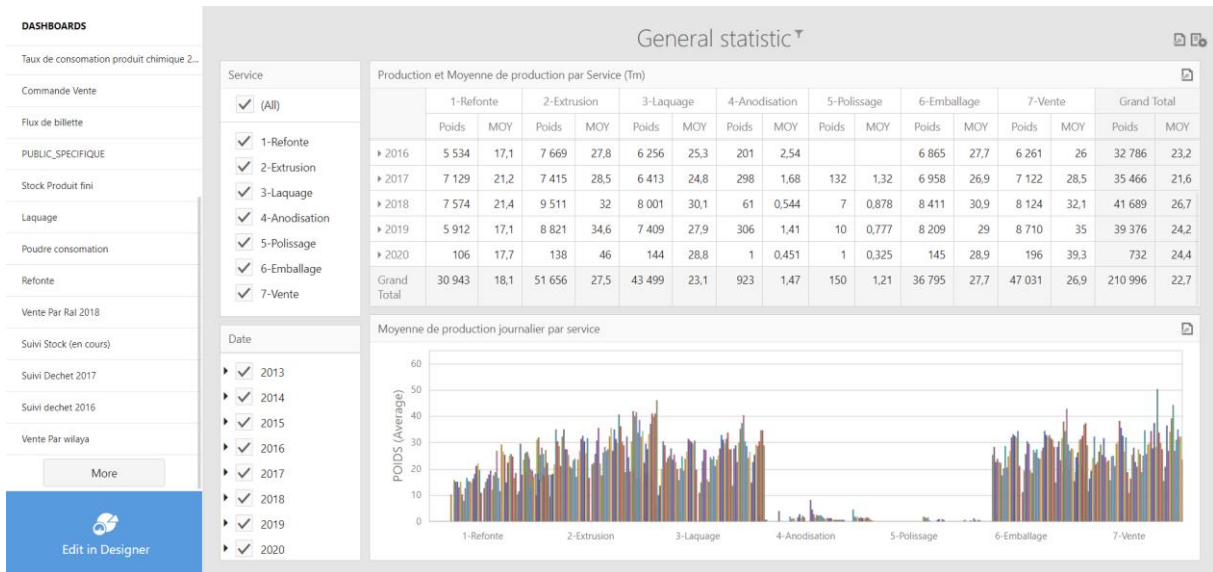


Figure 32 Dashboard

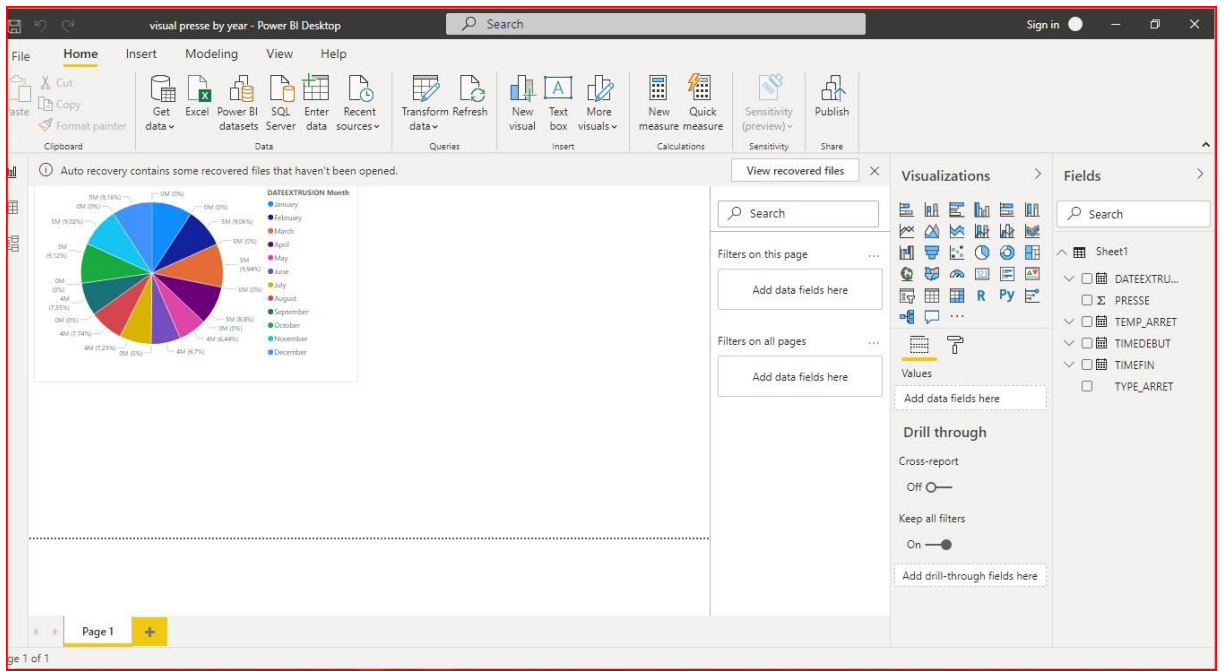


Figure 33 Visual presse by Year

Conclusion

Les travaux de recherche présentés dans ce mémoire de thèse s'inscrivent dans le cadre des systèmes d'aide à la prise de décision. Ces systèmes se basent sur un processus d'analyse en ligne (OLAP) et structurent les données de manière multidimensionnelle. Nous avons proposé une nouvelle approche qui consiste à intégrer les données issues de documents XML au cœur du processus d'analyse. Jusqu'à présent, ces documents sont restés exclus des systèmes d'aide à la prise de décision, faute de méthode et d'outils adaptés. Selon [Tseng & Chou, 2006], les systèmes OLAP n'emploient que 20% des données qui transitent au sein des systèmes d'information des entreprises. Les 80% restants, des documents restent hors de portée des systèmes d'aide à la prise de décision. Dans [Sullivan, 2001], l'auteur argumente en faveur de l'intégration des données issues de documents et de l'emploi de la fouille de texte pour en permettre l'analyse. De leur côté, les auteurs de [Fankhauser & Klement, 2003] affirment que la technologie XML est assez mature pour permettre l'implantation d'outils de fouilles de texte performant. Nous avons proposé d'aller au delà de ces propos en fournissant un environnement permettant l'intégration de documents XML au sein d'un système OLAP adapté pour l'analyse de données textuelles contenues dans ces documents. Il s'agit d'une première approche en vue de l'intégration de 100% des données issues des systèmes d'information des entreprises. Pour permettre l'intégration de documents au sein de l'environnement d'aide à la prise de décision, nous avons proposé un modèle multidimensionnel associé à un ensemble d'opérations pour permettre la manipulation des concepts du modèle. Nous avons proposé une démarche pour intégrer les documents au sein de l'environnement. La validation de nos propositions a été effectuée par le développement d'un prototype permettant l'analyse multidimensionnelle de données issues de documents.

Références bibliographiques

- [En ligne](#) Abelló A., Samos J., Saltor F., “YAM²: A Multidimensional conceptual model extending UML”, *Information Systems (IS)*, vol. 31, (6), Elsevier, 2006, p. 541-567.
- [En ligne](#) Beyer K.S., Chamberlin D.D., Colby L.S., Ozcan F., Pirahesh H., Xu Y., “Extending XQuery for Analytics”, *ACM SIGMOD Int. Conf. on Management of Data (SIGMOD)*, ACM Press, 2005, p. 503-514.
- [En ligne](#) Bordawekar R., Lang C. A., “Analytical processing of XML documents: opportunities and challenges”, *SIGMOD Record*, vol. 34, (2), ACM Press, 2005, p. 27-32.
- [En ligne](#) Boussaid O., Messaoud R.B., Choquet R., Anthoard S., “X-Warehousing: An XML-Based Approach for Warehousing Complex Data”, *10th East European Conf. on Advances in Databases and Information Systems (ADBIS 2006)*, LNCS 4152, Springer, 2006, p. 39-54.
- Codd E.F., Codd S.B., Salley C.T., Providing OLAP (On Line Analytical Processing) to user analyst: an IT mandate, rapport technique, E.F. Codd and associates, (white paper de Hyperion Solutions Corporation), 1993.
- [En ligne](#) Colliat G., “OLAP, Relational, and Multidimensional Database Systems”, *SIGMOD Record*, vol. 25, (3), ACM Press, 1996, p. 64-69.
- Fankhauser P., Klement T., “XML for Data Warehousing Chances and Challenges”, (Extended Abstract), *5th Int. Conf. on Data Warehousing and Knowledge Discovery (DaWaK 2003)*, LNCS 2737, Springer, 2003, p. 1-3.
- [En ligne](#) Golfarelli M., Rizzi S., Vrdoljak B., “Data Warehouse Design from XML Sources”, *4th ACM Int. Workshop on Data Warehousing and OLAP (DOLAP 2001)*, ACM Press, 2001, p. 40-47.
- Gray J., Bosworth A., Layman A., Pirahesh H., “Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Total”, *12th Int. Conf. on Data Engineering (ICDE)*, IEEE Computer Society, 1996, p. 152-159.

- Gyssens M., Lakshmanan L.V.S., “A Foundation for Multi-dimensional Databases”, *23rd Int. Conf. on Very Large Data Bases (VLDB)*, Morgan Kaufmann, 1997, p. 106-115.
- Horner J., Song I-Y., Chen P.P., “An analysis of additivity in OLAP systems”, *7th ACM Int. Workshop on Data Warehousing and OLAP (DOLAP 2004)*, ACM Press, p. 83-91, 2004.
- Jensen M.R., Møller T.H., Pedersen T.B., “Specifying OLAP Cubes On XML Data”, *13th Int. Conf. on Scientific and Statistical Database Management (SSDBM)*, IEEE Computer Society, 2001, p. 101-112.
- Keith S., Kaser O., Lemire D., “Analyzing Large Collections of Electronic Text Using OLAP”, *APICS 29th Conf. in Mathematics, Statistics and Computer Science*, Acadia University, 2005, p. 17-26.
- Khrouf K., Soulé-Dupuy C., “A Textual Warehouse Approach: A Web Data Repository”, in *Intelligent Agents for Data Mining and Information Retrieval*, Masoud Mohammadian (Eds.), Idea Publishing Group (IGP), ISBN : 1-59140-277-8, 2004, p. 101-124.
- Kimball R., *The data warehouse toolkit*, Ed. John Wiley and Sons, 1996, 2nd ed. 2003.
- En ligne Malinowski E., Zimányi E., “Hierarchies in a multidimensional model: From conceptual modeling to logical representation”, *Journal of Data & Knowledge Engineering (DKE)*, vol. 59, (2), 2006, p. 348-377.
- McCabe C., Lee J., Chowdhury A., Grossman D. A., Frieder O., “On the design and evaluation of a multi-dimensional approach to information retrieval”, *23rd Int. ACM Conf. on Research and Development in Information Retrieval (SIGIR)*, ACM Press, 2000, p. 363-365.
- En ligne Mothe J., Chrisment C., Dousset B., Alau J., “DocCube: Multi-dimensional visualisation and exploration of large document sets”, *Journal of the American Society for Information Science and Technology (JASIST)*, vol. 54, (7), Wiley Periodicals, 2003, p. 650-659.
- Nassis V., Rajugan R., Dillon T.S., Rahayu J.W., “Conceptual Design of XML Document Warehouses”, *6th int. Conf. on Data*

Warehousing and Knowledge Discovery (DaWaK 2004), LNCS 3181, Springer, 2004, p. 1-14.

- En ligne Niemi T., Niinimäki M., Nummenmaa J., Thanisch P., “Constructing an OLAP cube from distributed XML data”, *5th ACM Int. Workshop on Data Warehousing and OLAP (DOLAP)*, ACM Press, 2002, p. 22-27.
- En ligne Park B-K., Han H., Song I-Y., “XML-OLAP: A Multidimensional Analysis Framework for XML Warehouses”, *6th Int. Conf. on Data Warehousing and Knowledge Discovery (DaWaK)*, LNCS 3589, Springer, 2005, p. 32-42.
- En ligne Pedersen D., Riis K., Pedersen T.B., “XML-Extended OLAP Querying”, *14th Int. Conf. on Scientific and Statistical Database Management (SSDBM)*, IEEE Computer Society, 2002, p. 195-206.
- Pérez-Martinez J.M., Berlanga-Llavori R.B., Aramburu-Cabo M.J., Pedersen T.B., “Contextualizing data warehouses with documents”, *Decision Support Systems (DSS)*, Elsevier, (In Press) disponible en ligne : <http://dx.doi.org/10.1016/j.dss.2006.12.005>
- En ligne Pokorný J., “Modelling Stars Using XML”, *4th ACM Int. Workshop on Data Warehousing and OLAP (DOLAP)*, ACM Press, 2001, p. 24-31.
- Ravat F., Teste O., Tournier R., “OLAP Aggregation Function for Textual Data Warehouse”, *9th International Conference on Enterprise Information Systems (ICEIS 2007)*, vol. DISI, INSTICC Press, 2007, p. 151-156.
- En ligne Ravat F., Teste O., Tournier R., Zurfluh G., “Querying Multidimensional Databases”, *11th East-European Conference on Advances in Databases and Information Systems (ADBIS 2007)*, LNCS 4690, Springer, 2007, p. 298-313.
- Sullivan D., *Document Warehousing and Text Mining*, Wiley John & Sons, 2001.
- Torlone R., “Conceptual Multidimensional Models”, Capitre 3 de *Multidimensional Databases: Problems and Solutions*, M. Rafanelli (ed.), Idea Publishing Group (IGP), 2003, p. 69-90.
- En ligne Tseng F.S.C., Chou A.Y.H., “The concept of document warehousing for multi-dimensional modeling of textual-based

- business intelligence”, *Decision Support Systems (DSS)*, vol. 42, (2), Elsevier, 2006, p. 727-744.
- Vrdoljak B., Banek M., Rizzi S., “Designing Web Warehouses from XML Schemas”, *5th Int. Conf. on Data Warehousing and Knowledge Discovery (DaWaK)*, LNCS 2737, Springer, 2003, p. 89-98.
 - En ligne Vrdoljak B., Banek M., Skocir Z., “Integrating XML Sources into a Data Warehouse”, *2nd Int. Workshop on Data Engineering Issues in E-Commerce and Services (DEECS 2006)*, LNCS 4055, Springer, 2006, p. 133-142.
 - Wang H., Li J., He Z., Gao H., “OLAP for XML Data”, *5th Int. Conf. on Computer and Information Technology (CIT)*, IEEE Computer Society, 2005, p. 233-237.
 - En ligne Wiwatwattana N., Jagadish H.V., Lakshmanan L.V.S., Srivastava D., “X³: A Cube Operator for XML OLAP”, *23rd Int. Conf. on Data Engineering (ICDE)*, IEEE Computer Society, 2007, p. 916-925.
 - En ligne Yin X., Pedersen T.B., “Evaluating XML-extended OLAP queries based on a physical algebra”, *7th ACM Int. Workshop on Data Warehousing and OLAP (DOLAP)*, ACM Press, 2004, p. 73-82.
 - Zhang J., Ling T.W., Bruckner R.M., Tjoa A.M., “Building XML Data Warehouse Based on Frequent Patterns in User Queries”, *5th Int. Conf. on Data Warehousing and Knowledge Discovery (DaWaK)*, LNCS 2737, Springer, 2003, p. 99-108.