

BI, Business Intelligence

Système interprétant des données complexes permettant aux dirigeants d'entreprise de prendre des décisions en connaissance de cause. Les données sont analysées selon plusieurs dimensions (type de produits, régions et saisons par exemple). De plus en plus, l'informatique décisionnelle se rapproche de l'intelligence d'affaires, où un système informatique permet la recherche active et l'exploitation, sur le plan décisionnel, de l'ensemble des renseignements stratégiques essentiels qu'une entreprise doit posséder, si elle veut faire face à la concurrence et occuper la première place, dans son secteur d'activités.

Editeur	Progiciel	Site web	Commentaires
Ab Initio Software Corporation	Ab Initio	Ab Initio	La Rolls des ETL
Applix Inc	Applix iTM1	Applix Inc	La nouvelle référence MOLAP temps réel avec son propre ETL ultra rapide et son plugin MS-Excel
ASG Allen Systems Group	Rochade	Rochade	gestion des métadonnées
Bayesia	BayesiaLab	Bayesia	Analyse de données, data mining et aide à la décision exploitant la puissance des réseaux Bayésiens
Business Objects	Business Objects ou B.O.	Business Objects	l'un des requêteurs les plus répandus
Business Objects	Data Integrator	Data Integrator	outil d' ETL issu du rachat d'Acta Works en 2002
Cognos	PowerPlay, ...	PowerPlay	
Cognos	Reportnet	Cognos	Solution de reporting fullweb permettant un accès unique à toutes les fonctions de pilotage nécessaires à l'Entreprise : portail, reporting, requêtes, analyses multidimensionnelles (à partir de cubes de différents éditeurs : Cognos Powerplay, Hyperion Essbase, etc.) et conceptions d'états.
Cognos	4Thought	Cognos	outil de datamining
Computer Associates	Advantage Data Transformer	CA	Issu du rachat de Platinum
Coryent Conseil	CORICO	CORICO	Analyse de données multivariées et plan d'expériences
Databeacon	Analyse & Reporting Web	Databeacon	Solution Web M-OLAP Java , Orienté PME-PMI

Data Mirror	Transformation Server	ETL	outil d' ETL
DATA	DataStudio	DATA SAS	Outil d' ETL .
Digicap	EasyOlap	Explorateur	Explorateur de données, analyse, reporting, OLAP.
ETI	ETI Extract	ETI	outil d' ETL
HarrySoftware	HarryPilot, HarryCube, ...	HARRYSuite	requêteur, moteur OLAP , ...
Hummingbird	Hummingbird ETL	Hummingbird ETL	outil d' ETL (rachat de Genio à Léonard Logics)
Hyperion Solutions	Hyperion Essbase	Hyperion Essbase	Le moteur M-OLAP le plus répandu
Hyperion Solutions	Tous les produits	Tous les produits	La suite complète des outils décisionnels de Business Performance Management
IBM	DB2	DB2	SGBD relationnel très répandu dans les grandes entreprises
IBM	Intelligent Miner	Intelligent Miner	outils de datamining
IBM	Websphere Datastage	Datastage	outil d' ETL (rachat d'Ascential en 2005)
IBM	Websphere Metastage	Metastage	gestion des métadonnées (rachat d'Ascential en 2005)
Informatica	PowerCenter ETL Manager	PowerCenter	outil d' ETL
Information Builders	ETL Manager	ETL	outil d' ETL
ISoft	Amadea	Amadea	outil d' ETL (éditeur français)
ISoft	Alice	Alice	outil de datamining
KXEN	KXEN Analytic Framework	KXEN	outil de datamining
Microsoft	SQL Server	SQL Server	l'un des SGBD relationnels les plus répandus
Microsoft	SQL Server Integration Services (SSIS)	SSIS	ETL intégré à SQL Server 2005
MicroStrategy	MicroStrategy 7i	MicroStrategy	moteur R-OLAP
NCR Corporation	Teradata	Teradata	le SGBD de référence pour les très gros volumes de données
Neolane	Neolane	Neolane	Neolane, éditeur de logiciels dédiés au marketing et à la communication client
Oracle Corporation	Oracle	Oracle	l'un des SGBD relationnels les plus répandus. L'Option Oracle Olap inclus un moteur multidimensionnel puissant.

Oracle Corporation	Warehouse builder	W. Builder	outil d' ETL
Pertinence	Pertinence Suite	Pertinence	outil d'analyse de données et d'aide à la décision pour l'optimisation de processus
Pervasive	Pervasive	Pervasive	outil d' ETL
PostgreSQL	PostgreSQL	PostgreSQL	SGBD relationnel libre et gratuit, très puissant, qui n'a rien à envier aux produits commerciaux.
Report One	MyReport	Report One	outils d'importation de données, d' ETL et de création de tableaux de bord sur mesure dans MS Excel et OpenOffice.org .
Sagent Group 1 Software	Data Flow	Data Flow	outil d' ETL
SAP	Business Information Warehouse	SAP BW	Outil décisionnel H-OLAP contenant ETL , Gestion des droits et requêteur. Possède également des process standards pour une intégration accrue avec SAP R/3.
SAS	SAS9	SAS	Plateforme décisionnelle complète : Intégration de données (ETL , qualité de données...), stockage, métadonnées uniques, portail web, reporting de masse, interactif ou non, analyse de type OLAP , analyse prédictive, datamining , textmining , applications métiers (marketing, ressources humaines, achats, grande distribution, finance, risque...) et pilotage stratégique de type balanced scorecard
SAS	Enterprise Miner	Enterprise Miner	outils de datamining
SPSS	Clementine	SPSS	outils de datamining et textmining
Sunopsis	Sunopsis	Sunopsis	outil d' ETL
SYNAXE	LATITUDES	Latitudes	Outil full web, totalement intégré : Administration et gestion des métadonnées , requêtes et reporting 'Ad-hoc', Cubes R-OLAP , tableaux de pilotage, exécution et diffusion automatisée.
Synergy SAS	Meta Analysis	Meta Analysis	gestion des métadonnées
talend	Talend Open Studio	talend	outil d' ETL en open source , basé sur une architecture distribuée de type grille de calcul
Weka	Weka	Data Mining and Machine	Logiciel libre dédié au Data Mining

		Learning	
Winsight	Reportsmith .Net	Winsight	ReportSmith.net™ permet de rendre accessible aux utilisateurs la puissance de la plateforme BI MS SQL Server 2000 ou 2005 depuis un simple navigateur Web

Le décisionnel : la réponse à la problématique managériale

La prise de décision est le coeur de la gestion d'entreprise. En effet, aujourd'hui, il ne suffit plus de produire plus et plus vite pour rester pérenne. Il faut produire mieux tout en étant rentable, tenir compte des besoins des clients, des évolutions du marché etc.

Etudier les données de production telles quelles ne suffit pas non plus. Le décideur a besoin de collecter les informations opérationnelles internes et externes, les croiser et les recouper, de disposer en temps réel, des informations pertinentes pour prendre rapidement les bonnes décisions. Autant de préoccupations auxquelles répond l'informatique décisionnelle.

Le décisionnel ou informatique décisionnelle englobe les solutions informatiques d'aide à la décision avec, en bout de chaîne, des rapports, des tableaux de bords, des statistiques prévisionnelles etc. Des outils pour le suivi, le pilotage de l'activité pour une prise de décision précise et rapide.

Le portail : la porte d'entrée vers vos ressources

Le portail est une plateforme donnant accès aux données, ressources, services de l'entreprise regroupées au sein d'une interface unique. C'est la «porte d'entrée» vers les données de l'entreprise autant pour l'ensemble de son personnel que pour ses partenaires, conformément à la politique de sécurité de l'entreprise.

Le portail intègre aussi une dimension de personnalisation visuelle et fonctionnelle. L'utilisateur construit son propre espace de travail, il sélectionne les ressources qui l'intéressent dans celles qui lui sont proposées, il peut même les agencer à sa guise.

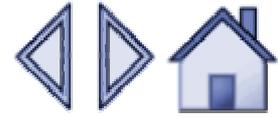
C'est dans cet esprit qu'est conçu le portail décisionnel: il allie les capacités de fédération et de personnalisation du portail avec les capacités de synthèse et de restitution de l'information pertinente du décisionnel pour offrir un outil utilisable à tous les niveaux hiérarchiques de l'entreprise.

A qui s'adresse le portail décisionnel ?

Il est évident que le portail décisionnel s'adresse aussi bien à l'équipe de direction, aux décideurs, aux chefs opérationnels, à la direction financière qu'à la direction du marketing.

En effet, il propose un panel d'outils qui répondent à leurs différents besoins :

- des outils qui permettent de définir les orientations stratégiques, de suivre les objectifs critiques, d'analyser et de mesurer les écarts de performance, comme les tableaux de bords, les regroupements analytiques d'indicateurs, la navigation multidimensionnelle etc., pour les besoins de pilotage stratégique et de pilotage de la performance.
- des outils pour le suivi quotidien des processus opérationnels tels que les rapports statiques, et dynamiques.
- des outils d'analyse multidimensionnelle pour les besoins particuliers d'élaboration et de consolidation budgétaires comme de gestion de la relation client par la segmentation du marché.



Introduction à la gestion de la relation client

Le client est généralement la principale source de revenus pour les entreprises. Or, avec le changement de l'économie dû notamment à l'intégration des nouvelles technologies dans les relations client-entreprise, la concurrence devient de plus en plus serrée et les clients peuvent ainsi désormais se permettre de choisir leur fournisseur ou d'en changer par un simple clic. Les critères de choix des clients sont notamment des critères financiers, de réactivité de l'entreprise mais également des critères purement affectifs (besoin de reconnaissance, besoin d'être écoutés, ...). Ainsi dans un monde de plus en plus concurrentiel, les entreprises souhaitant augmenter leurs bénéfices ont plusieurs alternatives :

- Augmenter la marge sur chaque client,
- Augmenter le nombre de clients,
- Augmenter le cycle de vie du client, c'est-à-dire le fidéliser.

Les nouvelles technologies permettent aux entreprises de mieux connaître leur clientèle et de gagner leur fidélité en utilisant les informations les concernant de telle manière à mieux cerner leurs besoins et donc de mieux y répondre.

Ainsi il s'est avéré que fidéliser un client coûtait 5 fois moins cher que d'en prospector des nouveaux. C'est la raison pour laquelle un grand nombre d'entreprises orientent leur stratégie autour des services proposés à leurs clients.

Qu'est-ce que le CRM ?

Le CRM (*Customer Relationship Management*, ou en français **GRC**, *gestion de la relation client*) vise à proposer des solutions technologiques permettant de renforcer la communication entre l'entreprise et ses clients afin d'améliorer la relation avec la clientèle en automatisant les différentes composantes de la relation client :

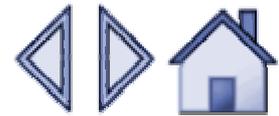
- L'avant-vente : il s'agit du marketing, consistant à étudier le marché, c'est-à-dire les besoins des clients et à démarcher les prospects. L'analyse des informations collectées sur le client permet à l'entreprise de revoir sa gamme de produits afin de répondre plus précisément à ses attentes. L'*Enterprise Marketing Automation (EMA)* consiste ainsi à automatiser les campagnes marketing.
- Les ventes : L'Automatisation des forces de ventes (*Sales Forces Automation, SFA*), consiste à fournir des outils de pilotage aux commerciaux afin de les assister dans leurs démarches de prospection (gestion des prises de contact, des rendez-vous, des relances, mais aussi aide à l'élaboration de propositions commerciales, ...).
- La gestion du service clientèle : le client aime se sentir connu et reconnu de l'entreprise et ne supporte pas devoir récapituler, à chaque prise de contact, l'historique de sa relation à l'entreprise.
- L'après-vente, consistant à fournir une assistance au client notamment via la mise en place de [centres d'appel](#) (appelés généralement *Call centers*, *Help Desk* ou *Hot-Line*) et via la mise en ligne d'informations de support technique.

L'objet du CRM est d'être plus à l'écoute du client afin de répondre à ses besoins et de le fidéliser. Un projet de CRM consiste donc à permettre à chaque secteur de l'entreprise d'accéder au [système d'information](#) pour être en mesure d'améliorer la connaissance du client et lui fournir des produits ou services répondant au mieux à ses attentes.

Intégration du CRM dans l'entreprise

La mise en place de solutions de CRM dans une entreprise ne consiste pas uniquement à installer un logiciel ad-hoc, mais à modifier l'organisation de l'entreprise tout entière, ce qui implique une nécessaire prise en compte d'un projet de conduite de changement. En effet la mise en place d'une stratégie de CRM impose des modifications structurelles, de compétences et de comportements.

Informatique
décisionnelle
(Business intelligence)



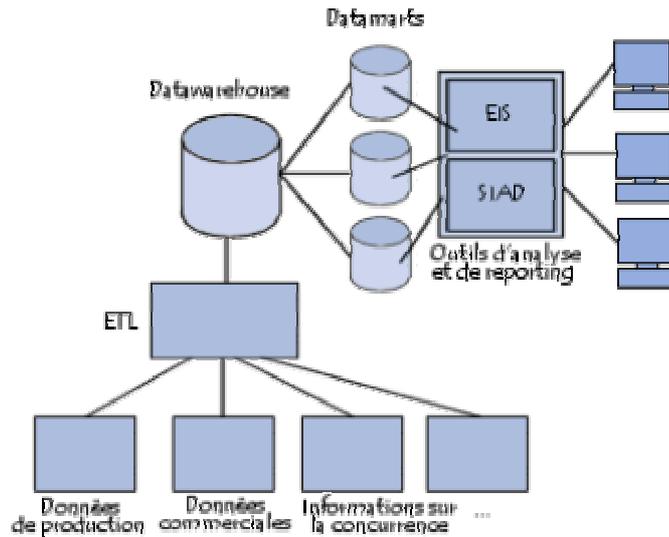
Introduction à l'informatique décisionnelle

On qualifie d'**informatique décisionnelle** (en anglais « **Business intelligence** », parfois appelé tout simplement « *le décisionnel* ») l'exploitation des données de l'entreprise dans le but de faciliter la prise de décision par les décideurs, c'est-à-dire la compréhension du fonctionnement actuel et l'anticipation des actions pour un pilotage éclairé de l'entreprise.

Les outils décisionnels sont basés sur l'exploitation d'un [système d'information](#) décisionnel alimenté grâce à l'extraction de données diverses à partir des données de production, d'informations concernant l'entreprise ou son entourage et de données économiques.

Un outil appelé **ETL** (*Extract, Transform and Load*) est ainsi chargé d'extraire les données dans différentes sources, de les nettoyer et de les charger dans un *entrepôt de données*.

Enfin des outils d'analyse décisionnelle permettent de modéliser des représentations à base de requêtes afin de constituer des tableaux de bord, on parle ainsi de *reporting*.



Enterprise
Application
Integration



Introduction à la notion d'EAI

L'objet de l'**EAI** (*Enterprise Application Integration*, traduisez *intégration des applications de l'entreprise*) est l'interopérabilité et l'organisation de la circulation de l'information entre des applications hétérogènes, c'est-à-dire faire communiquer les différentes applications constituant le [système d'information](#) de l'entreprise, voire même celles des clients, des partenaires ou des fournisseurs.

Un projet d'EAI consiste donc dans un premier temps à mettre en place une architecture dans laquelle les différentes applications communiquent entre elles. Il s'agit donc de développer des **connecteurs** (*middleware*) permettant d'interfacer des applications utilisant des [protocoles de communications](#) différents (généralement propriétaires).

Toutefois le projet d'EAI va au-delà de l'interopérabilité entre les applications : il permet de définir un [workflow](#) entre les applications et constitue ainsi une alternative aux [ERP](#) avec une approche plus modulaire.

Néanmoins, l'EAI conserve des limites liées à la rigidité de l'existant (appelé *legacy*, traduisez *héritage*), si bien qu'il est nécessaire de modifier le connecteur lors de modifications importantes des applications. Les [web services](#)

Entrepôt de données

Structure informatique dans laquelle est centralisé un volume important de données consolidées à partir des différentes sources de renseignements d'une entreprise (notamment les bases de données internes)/ L'organisation des données est conçue pour que les personnes intéressées aient accès rapidement et sous forme synthétique à l'information stratégique dont elles ont besoin pour la prise de décision.

Entrepot de Donnees

Introduction

Problematique

Les décideurs d'une entreprise doivent pouvoir répondre à un certain nombre de question pour diriger leur entreprise :

Qui sont mes clients ?

Pourquoi sont ils mes clients ?

Comment cibler ma clientèle ?

Quel est l'évolution de tel produit ?

Qui sont mes employés ?

...

L'objectif est donc d'apporter aux décideurs d'une entreprise les moyens de répondre à ces questions.

Utilite d'un datawarehouse

Les sources de données d'une entreprise proviennent essentiellement des bases de production. Ces données sont éparpillées dans des systèmes multiples, pas nécessairement compatibles entre eux. Ces bases sont conçues pour être efficaces pour les fonctions sur lesquelles elles sont spécialistes. Elles sont donc peu structurées pour l'analyse, avec souvent comme objectif principal de conserver l'information. Comme bases de production elles sont focalisées sur les fonctions critiques de l'entreprise, et doivent être en mesure de servir l'utilisateur avec un temps de réponse rapide et structurées dans ce but.

Ces systèmes sont donc peu adaptés à la vision à long terme et donc à la prise de décision. Le datawarehouse va avoir pour objectif d'agréger et de valoriser ces données provenant de différentes sources. Il va permettre à l'utilisateur d'y accéder de manière simple et ergonomique.

Definition

Définition de Bill Inmon (1996):

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

Orientés sujet : Les bases de production sont le plus souvent organisées par processus fonctionnels. Le datawarehouse est lui organisé autour des sujets majeurs de l'entreprise. Les données sont donc structurés par thèmes, ces thèmes étant souvent transverses par rapport aux structures fonctionnelles et organisationnelles de l'entreprise (et donc transverses par rapport

aux systèmes de production).

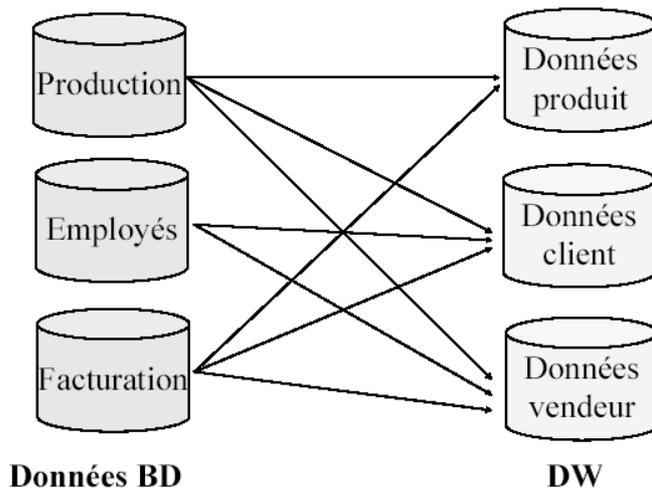
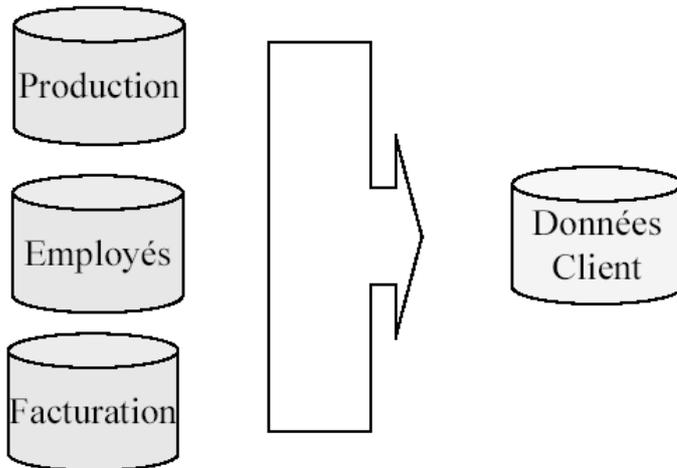


Illustration 1: Orienté sujet (source : C Vangenot, Laboratoire de Bases de Données)

Données intégrées : Les données proviennent de plusieurs sources différentes. Avant d'être intégrées au sein du datawarehouse elles doivent être mise en forme et unifiées afin d'en assurer la cohérence. Cela nécessite une forte normalisation, de bénéficier d'un référentiel unique et cohérent ainsi que de bonnes règles de gestion. Cette phase est très complexe et représente une charge importante dans la mise en place d'un datawarehouse.



: C Vangenot, Laboratoire de Bases de Données)

Illustration 2: Données intégrées (source

Données historisées : Contrairement au système de production les données ne sont jamais mises à jour. Chaque nouvelle données est insérées. Un référentiel de temps doit être mis en place afin de pouvoir identifier chaque donnée dans le temps.

Données non volatiles : Un datawarehouse veut conserver la traçabilité des informations et des décisions prises. Les données ne sont ni modifiées ni supprimées. Une requête émise sur les mêmes données à plusieurs mois d'intervalles doit donner le même résultat.

Un datawarehouse définis donc à la fois un ensemble de données et un ensemble d'outils. Il s'agit de données destinés aux décideurs, qui sont souvent une copie des données de production avec une valeur ajoutées (orientés objet, agrégés, historisées). Et c'est un ensemble d'outils permettant de regrouper les données des différentes sources, de les nettoyer et de les intégrer, ainsi que d'y accéder de différentes manières (requêtes, rapport, analyse, datamining).

SGBD et Datawarehouse

On pourrait se demander pourquoi ne pas utiliser un SGBD pour réaliser cette structure d'informatique décisionnelle. En réalité SGBD et datawarehouse ont des objectifs différents. Ils stockent les données de manière différentes et font l'objet de requêtes différentes. Ils sont ainsi basés sur deux systèmes différents : OLTP et OLAP

OLTP

OLTP (On Line Transaction Processing) est le modèle utilisé par les SGBD. Le mode de travail est transactionnel. L'objectif est de pouvoir insérer, modifier et interroger rapidement et en sécurité la base. Ces actions doivent pouvoir être effectuées très rapidement par de nombreux utilisateurs simultanément.

Chaque transaction travail sur de faibles quantités d'informations, et toujours sur les versions les plus récentes des données.

OLAP

Les datawarehouses eux reposent sur le système OLAP (On Line Analytical Processing). Ce système travail en lecture seulement. Les programmes consultent d'importantes quantités de données pour procéder à des analyses. Les objectifs principaux sont regrouper, organiser des informations provenant de sources diverses, les intégrer et les stocker pour donner à l'utilisateur une vue orientée métier, retrouver et analyser l'information facilement et rapidement. Cela nécessite de consulter des versions historiques de la base et peut se permettre d'ignorer temporairement les dernières mises à jour.

Ces bases sont souvent d'un ordre de grandeur nettement supérieur à celle des bases OLTP, du fait de la conservation de l'historique.

Comparaison

Voici un tableau récapitulatif des différences entre OLTP et OLAP :

Caractéristiques	OLTP	OLAP
Utilisation	SGBD (base de production)	Datawarehouse
Opération typique	Mise à jour	Analyse
Type d'accès	Lecture écriture	Lecture
Niveau d'analyse	Elémentaire	Global
Quantité d'information échangées	Faible	Importante
Orientation	Ligne	Multidimension
Taille BD	Faible (max qq GB)	Importante (pouvant aller à plusieurs TB).
Ancienneté des données	Récente	Historique

L'objectif des bases OLTP est de pouvoir répondre rapidement à des réponses simples, exemple : les ventes du produit X.

Les bases OLAP permettent des requêtes plus complexes : les ventes du produit X par vendeur, région et par mois.

Separation physique

Il est important de séparer les bases de production (SGBD) du datawarehouse. Pour des raisons de performances premièrement car les systèmes de production ne sont pas prévus pour répondre efficacement aux requêtes des systèmes d'aide à la décision. De plus les systèmes de production ne conservent pas leurs données, alors qu'un datawarehouse repose sur des données historisées. De plus un datawarehouse se repose sur des systèmes de production différents dont les données ne sont pas nécessairement uniformisées

SGBD et Datawarehouse

On pourrait se demander pourquoi ne pas utiliser un SGBD pour réaliser cette structure d'informatique décisionnelle. En réalité SGBD et datawarehouse ont des objectifs différents. Ils stockent les données de manière différentes et font l'objet de requêtes différentes. Ils sont ainsi basés sur deux systèmes différents : OLTP et OLAP

OLTP

OLTP (On Line Transaction Processing) est le modèle utilisé par les SGBD. Le mode de travail est transactionnel. L'objectif est de pouvoir insérer, modifier et interroger rapidement et en sécurité la base. Ces actions doivent pouvoir être effectuées très rapidement par de nombreux utilisateurs simultanément.

Chaque transaction travail sur de faibles quantités d'informations, et toujours sur les versions les plus récentes des données.

OLAP

Les datawarehouses eux reposent sur le système OLAP (On Line Analytical Processing). Ce système travail en lecture seulement. Les programmes consultent d'importantes quantités de données pour procéder à des analyses. Les objectifs principaux sont regrouper, organiser des informations provenant de sources diverses, les intégrer et les stocker pour donner à l'utilisateur une vue orientée métier, retrouver et analyser l'information facilement et rapidement. Cela nécessite de consulter des versions historiques de la base et peut se permettre d'ignorer temporairement les dernières mises à jour.

Ces bases sont souvent d'un ordre de grandeur nettement supérieur à celle des bases OLTP, du fait de la conservation de l'historique.

Comparaison

Voici un tableau récapitulatif des différences entre OLTP et OLAP :

Caractéristiques	OLTP	OLAP
-------------------------	-------------	-------------

Caractéristiques	OLTP	OLAP
Utilisation	SGBD (base de production)	Datawarehouse
Opération typique	Mise à jour	Analyse
Type d'accès	Lecture écriture	Lecture
Niveau d'analyse	Elémentaire	Global
Quantité d'information échangées	Faible	Importante
Orientation	Ligne	Multidimension
Taille BD	Faible (max qq GB)	Importante (pouvant aller à plusieurs TB).
Ancienneté des données	Récente	Historique

L'objectif des bases OLTP est de pouvoir répondre rapidement à des réponses simples, exemple : les ventes du produit X.

Les bases OLAP permettent des requêtes plus complexes : les ventes du produit X par vendeur, région et par mois.

Separation physique

Il est important de séparer les bases de production (SGBD) du datawarehouse. Pour des raisons de performances premièrement car les systèmes de production ne sont pas prévus pour répondre efficacement aux requêtes des systèmes d'aide à la décision. De plus les systèmes de production ne conservent pas leurs données, alors qu'un datawarehouse repose sur des données historisées. De plus un datawarehouse se repose sur des systèmes de production différents dont les données ne sont pas nécessairement uniformisées

Architecture Technique

Schema general

Ce schéma représente l'architecture générale du datawarehouse :

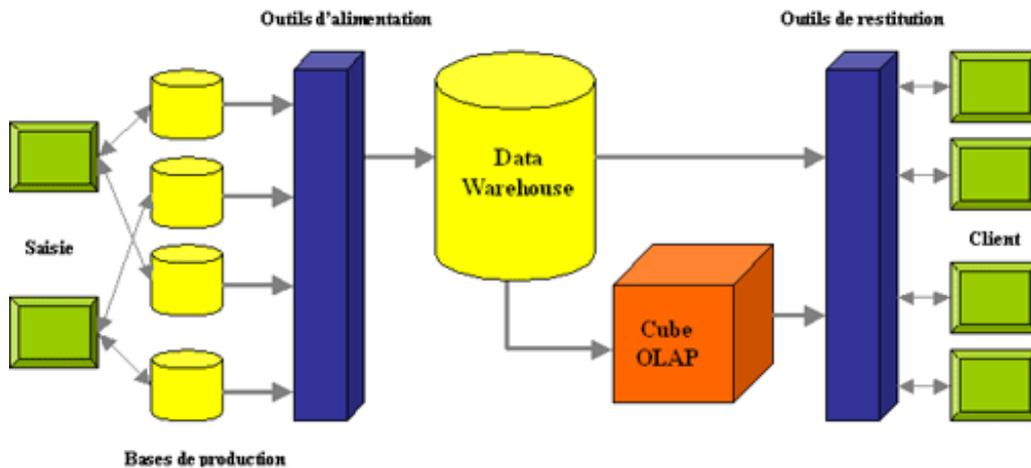


Illustration 3: Architecture générale (source : J Detroyes, supinfo)

Alimentation

L'alimentation des données à partir des bases de production est une phase primordiale d'un datawarehouse. Des outils logiciels sont alors nécessaires pour intégrer les données dans le datawarehouse. On parle d'ELT (Extract, Transform, Load). Les phases de l'alimentation d'un datawarehouse sont les suivantes :

Découverte des données : Il s'agit d'identifier dans les systèmes sources les données à importer dans le datawarehouse. Il faut prendre les données les plus judicieuses. Un mauvais choix peut considérablement compliquer les phases suivantes de l'alimentation.

Extraction des données : Il s'agit de collecter les données utiles dans les systèmes de production. Il faut identifier les données ayant été modifiées afin d'importer le minimum de données dans le datawarehouse.

Transformation des données : Il faut rendre les données cohérentes avec la structure du datawarehouse. On d'applique alors des filtres sur les données. Il peut être nécessaire de convertir le format des données (EBCDIC vers ASCII par exemple) ou d'harmoniser les formats de dates (jj/mm/aaaa). Il faut également associer les champs source avec les champs cibles. Un champ source « adresse » pourra ainsi par exemple être décomposé en « numéro », « rue », « code postal », « ville » ou l'inverse. Enfin des données des systèmes de production doivent être agrégées ou calculées avant leur chargement.

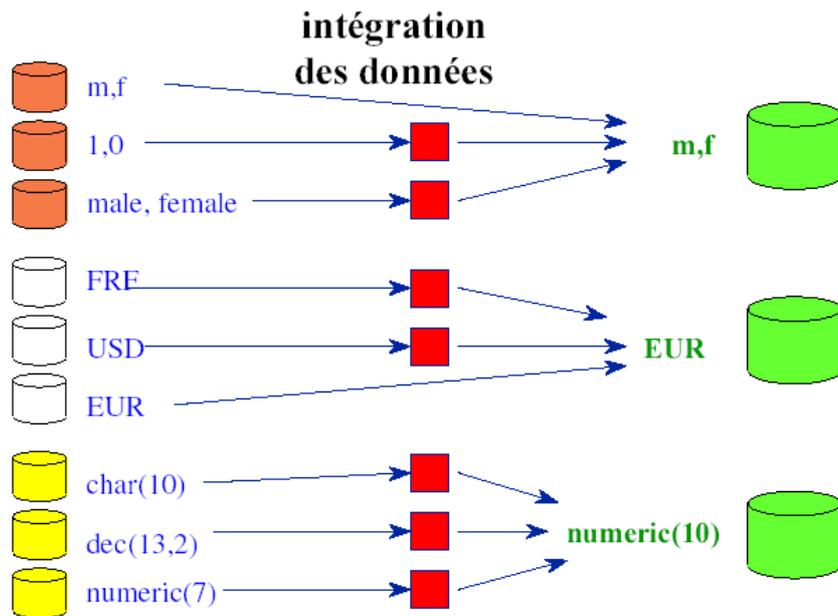


Illustration 4: Intégration des données (source : D Donsez, Université Joseph Fournier)

Chargement des données : C'est la dernière phase de l'alimentation d'un datawarehouse. Il s'agit d'insérer les données au sein du datawarehouse. C'est une phase délicate car les quantités de données sont souvent très importantes.

Datamart

Un datamart (ou magasin de données) est une vue partielle du datawarehouse mais orientée métier. C'est un sous-ensemble du datawarehouse contenant des informations se rapportant à un secteur d'activité particulier de l'entreprise ou à un métier qui y est exercé. Il se situe en aval du datawarehouse et est alimenté par celui-ci. On peut donc créer plusieurs datamart correspondant au différent besoin des utilisateurs.

Cela permet de réduire le nombre d'opération sur les bases de production. De plus cela permet d'offrir aux utilisateurs un outil spécifiquement adapté à leurs besoins. Cet outil sera plus petit et permettra donc un accès plus rapide à l'information.

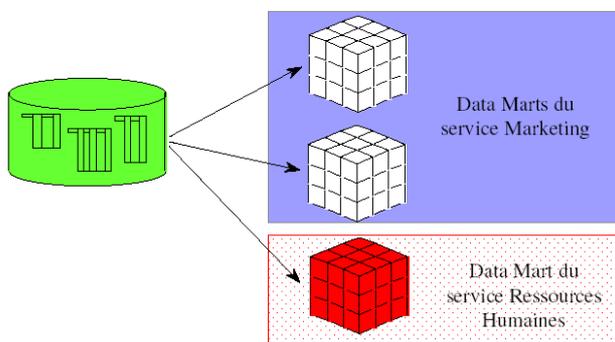


Illustration 15: Un datawarehouse, des datamarts (source : D Donsez, Université Joseph Fournier)

Bases de Données Multidimensionnelles

L'analyse multidimensionnelle est la capacité à analyser des données qui ont été agrégées suivant plusieurs dimensions. On veut donc accéder à des données déjà agrégés selon les besoins de l'utilisateur, de façon simple et rapide.

On utilise pour cela des hypercubes OLAP. Les données sont représentées dans des hypercubes à n dimensions. Les données sont structurées suivant plusieurs axes d'analyses (dimensions) comme le temps, la localisation ...

Une cellule est l'intersection des différentes dimensions. Le calcul de chaque cellule est réalisé au chargement. Le temps de réponse est ainsi stable quelque soit la requête.

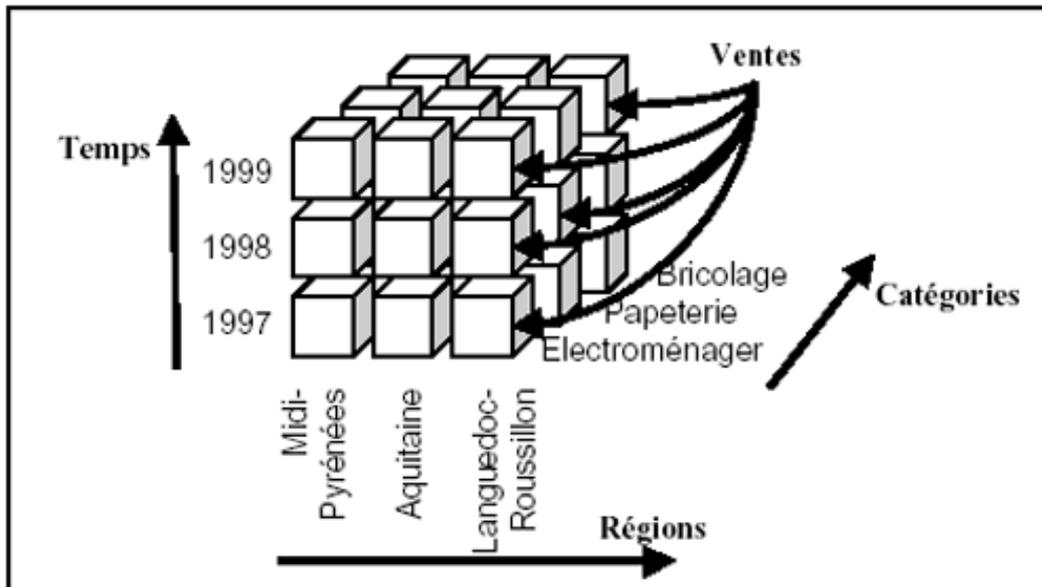


Illustration 16: Représentation d'un hypercube (source : J Detroyes, supinfo)

Cela permet, dans l'exemple du schéma ci-dessus, d'analyser la répartition de l'indicateur « vente » suivant le temps, les catégories de produit et les régions. En outre, des hiérarchies seront définies pour chaque axe d'analyse (par exemple, l'année, puis la saison, le mois et la semaine, pour l'axe temps). Une fois cette structure multidimensionnelle établie, l'outil OLAP propose des méthodes de navigation dans les données : drill-down/drill-up, rotate, slicing, scoping.

Drill-up / down

Cela permet d'aller vers les informations détaillées dans une hiérarchie ou au contraire de remonter d'un niveau de granularité. Il s'agit donc de « zoomer ou de dézoomer » sur une dimension.

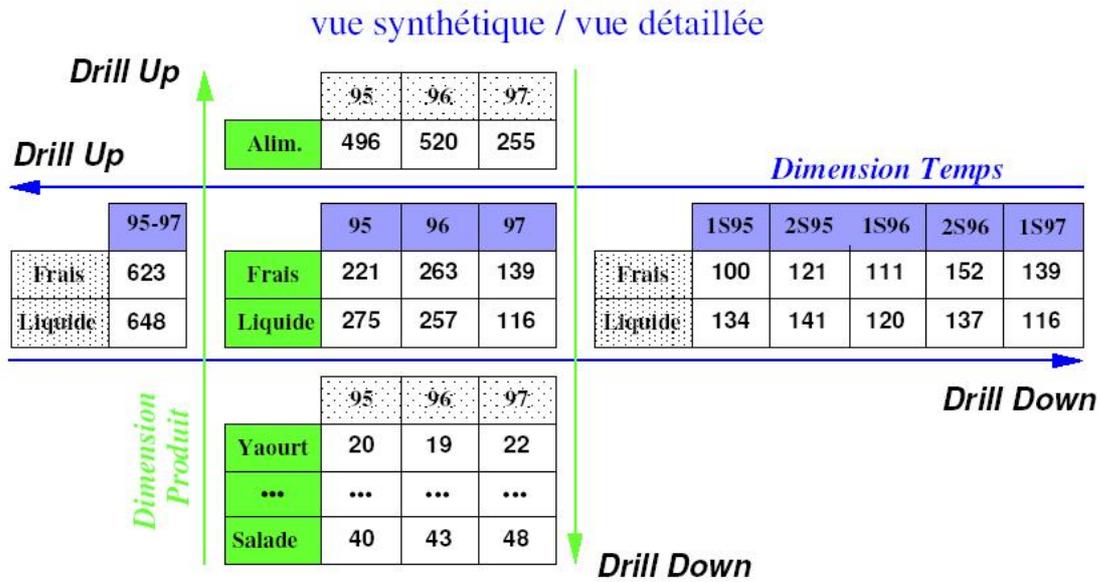


Illustration 17: Drill Up/ Drill Down (source : D Donsez, Université Joseph Fournier)

Rotate

Consiste à effectuer une rotation de l'hypercube afin de présenter une face différente. Il s'agit donc de modifier une dimension de lecture.

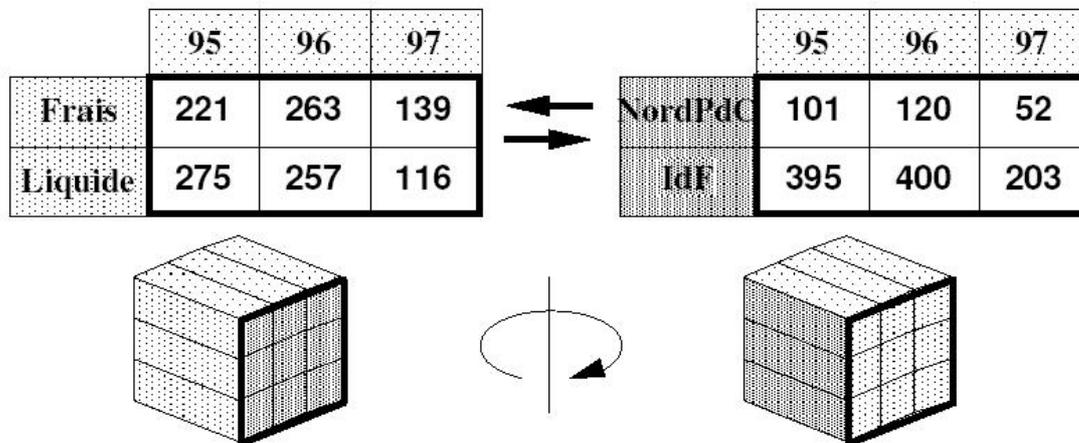


Illustration 18: Rotate (source : D Donsez, Université Joseph Fournier)

Slicing

Consiste à ne travailler que sur une tranche de l'hypercube. Une des dimensions est alors réduite à une seule valeur .

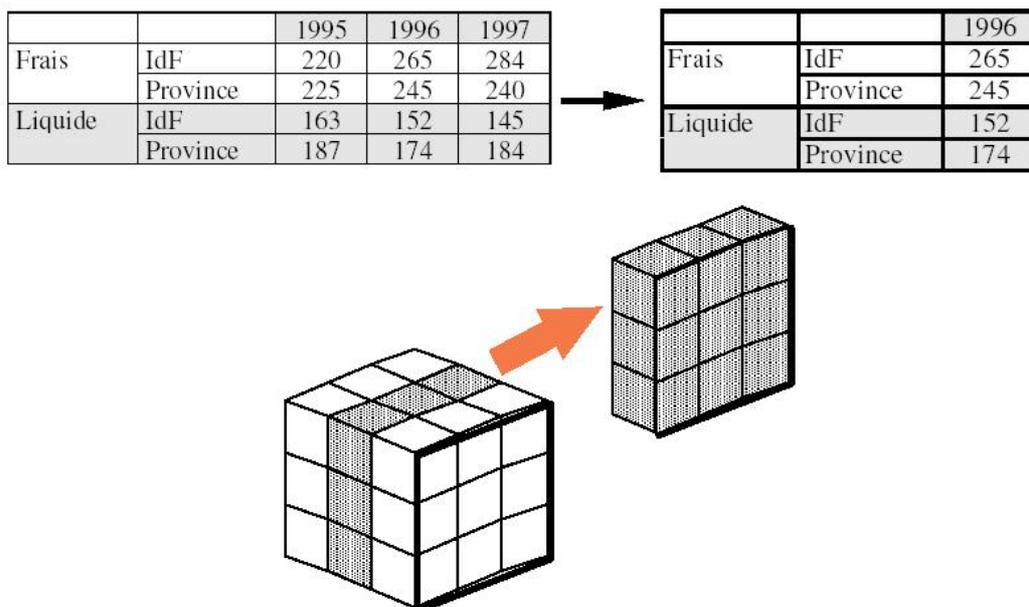


Illustration 19: Slicing (source : D Donsez, Université Joseph Fournier)

Scoping

Consiste à ne travailler que sur un sous-cube. On s'intéressera alors seulement à une partie des données.

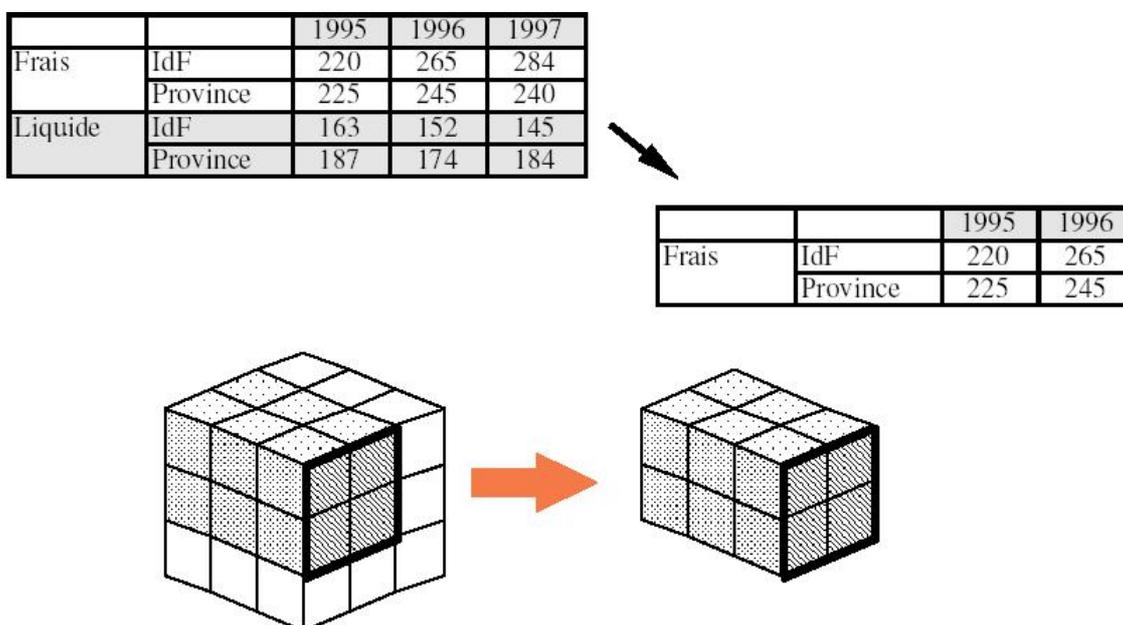


Illustration 20: Scoping (source : D Donsez, Université Joseph Fournier)

Differents OLAP

Au niveau logique il existe plusieurs possibilités pour la modélisation multidimensionnel : ROLAP, MOLAP, HOLAP et OOLAP

ROLAP : Dans le Relational OLAP les données sont stockées dans une base de données

relationnelle. Un moteur OLAP permet de simuler le fonctionnement d'un hypercube. Cela permet une facilité dans la mise à jour des données.

MOLAP : Le Multidimensional OLAP consiste à utiliser un système multidimensionnelle pur, qui gère des structures multidimensionnelles natives. Elles utilisent des tableaux à n dimensions. L'accès aux données se fait directement dans le cube. Cela permet une rapidité d'accès à l'information mais augmente le temps de mise à jour.

HOLAP : HOLAP (Hybrid OLAP) est un hybride entre ROLAP ET MOLAP. Les parties tables de faits et tables de dimensions sont stockées dans une base relationnelle standard tandis que le reste des données (les calculs) sont stockées dans une base multidimensionnelle.

OOLAP : C'est la technologie la plus récente, Object OLAP, qui s'appuie sur le paradigme objet. Le modèle multidimensionnel se traduit ainsi :

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

Le principal avantage est d'augmenter le niveau d'abstraction.