

Support Cours & TD Datawarehouse

Pr. A. ELOUARDIGHI

Chapitre 1: Notions et Architecture d'un DW

Systèmes transactionnels vs Systèmes décisionnels

Systèmes transactionnels

- Les outils traditionnels de gestion et d'exploitation des données sont du type transactionnel ou OLTP (On-Line Transaction Processing)
- L'exploitation de données tournée vers la saisie, le stockage, la mise à jour, la sécurité et l'intégrité des données.
- Le système transactionnel est développé pour gérer les transactions quotidiennes
- Ces bases de données supportent habituellement des applications particulières telles que les inventaires de magasins, les réservations d'hôtel, etc
- Le contenu est fait de données actuelles, pas d'archives
- Les données sont très détaillées (détails de chacune des transactions)
- La mise à jour s'effectue par de nouvelles transactions
- Très souvent plusieurs de ces systèmes existent indépendamment les uns des autres

Opérations dans les systèmes transactionnels

- Ajout
- Effacement
- Mise à jour des enregistrements
- Requêtes simples
- Interrogations et modifications fréquentes des données par de nombreux utilisateurs

Systèmes décisionnels

- Le terme décisionnel « Business Intelligence » couvre l'ensemble des technologies permettant en bout de chaîne d'apporter une aide à la décision.
- SI capable d'agrèger les données internes ou externes et de les transformer en information servant à une prise de décision rapide.
- SI capable de répondre à certains types de questions:
 - Quelles sont les ventes du produit X pendant le trimestre A de l'année B dans la région C ?
 - Comment se comporte le produit X par rapport au produit Y?
 - Quel type de client peut acheter le produit X?
 - Est-ce qu'une baisse de prix de 10% par rapport à la concurrence ferait redémarrer les ventes du produit X ?
- Ces exemples mettent en évidence les faits suivants:
 - Les questions doivent pouvoir être formulées dans le langage de l'utilisateur en fonction de son secteur d'activité:
 - *Service marketing, Service économique, service relation clients...*
 - La prévision des interrogations est difficile car elles sont du ressort de l'utilisateur.
 - Les questions vont varier selon les réponses obtenues:
 - *Si le produit X s'est vendu moins bien que l'année précédente, il va être utile de comprendre les raisons: Détailler les ventes par région par type de magasin,...*
 - Des questions ouvertes vont nécessiter la mise en place de méthodes d'extraction d'informations

Des données aux décisions

- Données :
 - Points de ventes, géographiques, démographiques, ...
- Informations :
 - I vit dans R, I est âgé de A, ...
- Connaissances :
 - Dans X%, le produit Y est vendu en même temps que le produit Z, ...
- Décisions
 - Lancer la promotion de Y & Z dans R auprès des clients plus âgés que A, ...

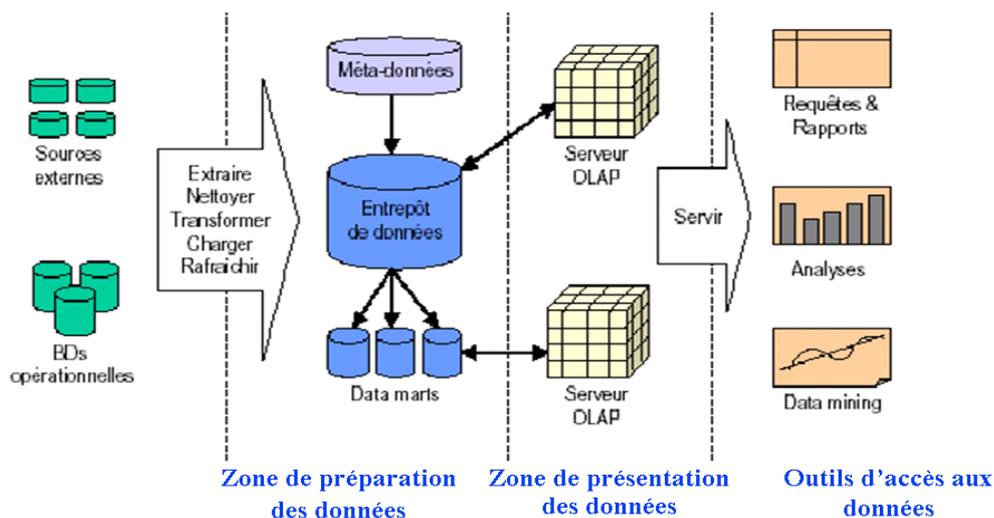
Applications transactionnelles v.s Applications décisionnelles

- Les applications transactionnelles sont constituées de traitements factuels de type OLTP (On Line Transaction Processing)
- Les applications d'aide à la décision sont constituées de traitements ensembliste de type OLAP: On Line Analytical Processing

Incompatibilités des deux activités

- Les deux activités (OLTP & OLAP) ne peuvent co-exister sur des données dans le même système d'information: leurs objectifs de performance sont exactement opposés:
 - Les requêtes complexes et lourdes dégradent les performances des systèmes transactionnels,
 - Les données temporelles sont réparties entre données actuelles et données archivées, rendant la vue historique des données très difficile ou impossible,
- Le support efficace d'une activité OLAP nécessite la constitution d'un système d'information propre: Le Datawarehouse

Naissance du Datawarehouse



Datawarehouse : Définition

- « Collection de données orientées sujets, intégrées, non volatiles et historisées, organisées pour le support du processus d'aide à la décision »
- Base de données dans laquelle sont déposées après nettoyage et homogénéisation les informations en provenance des différents systèmes de production de l'entreprise OLTP

Caractéristiques des données d'un DW

- **Orientées sujet :**
 - Organisées autour de sujets majeurs de l'entreprise
 - Données pour l'analyse et la modélisation en vue de l'aide à la décision, et non pas pour les opérations et transactions journalières
 - Vue synthétique des données selon les sujets intéressant les décideurs
- **Intégrées :**
 - Construit en intégrant des sources de données multiples et hétérogènes
 - BD relationnelles, fichiers, enregistrements de transactions
 - Les données doivent être mises en forme et unifiées afin d'avoir un état cohérent
 - Phase la plus complexe (60 à 90 % de la charge totale d'un projet DW)
- **Historisées :**
 - Fournies par les sources opérationnelles
 - Matière première pour l'analyse
 - Stockage de l'historique des données, pas de mise à jour
 - Un référentiel temps doit être associé aux données
- **Non volatiles :**
 - Conséquence de l'historisation
 - Une même requête effectuée à intervalle de temps, en précisant la date référence de l'information donnera le même résultat
 - Stockage indépendant des BD opérationnelles
 - Pas de mises à jour des données dans le DW

Données multidimensionnelles

- Notion de dimension : C'est une catégorie linguistique selon laquelle les données sont organisées:
 - Nom d'un attribut
 - Valeur d'un attribut

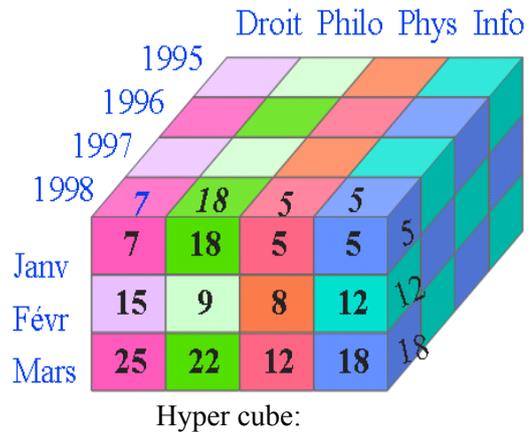
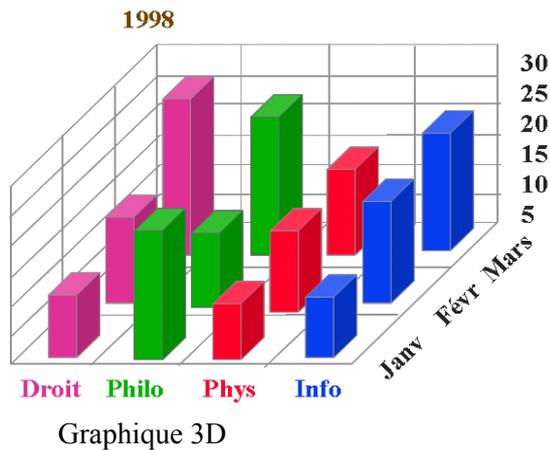
Représentation :

DuréeMoy	Départ.	Mois	Année
5	Info	Janv	1998
5	Phys	Janv	1998
18	Philo	Janv	1998
7	Droit	Janv	1998
12	Info	Févr	1998
8	Phys	Févr	1998
9	Philo	Févr	1998
15	Droit	Févr	1998
18	Info	Mars	1998
12	Phys	Mars	1998
22	Philo	Mars	1998
25	Droit	Mars	1998

Tableau simple

1998	Janv	Févr	Mars
Info	5	12	18
Phys	5	8	12
Philo	18	9	22
Droit	7	15	25

Tableau croisé



Notion de 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

Les domaines d'application du décisionnel

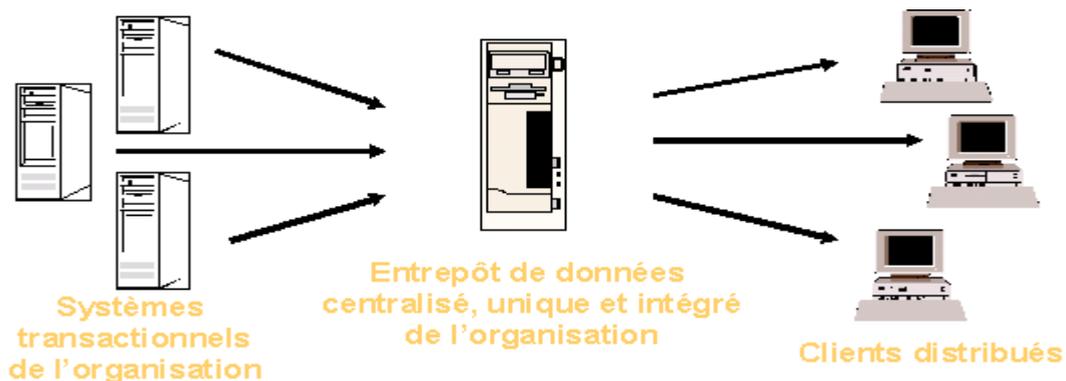
- La gestion de la relation client (CRM) est l'un des premiers champs d'application de la Business Intelligence.
- Le contrôle de gestion pour l'analyse des coûts, l'analyse de la rentabilité, l'élaboration budgétaire, les indicateurs de performance...
- La direction marketing pour le ciblage, le pilotage de gamme, les applications de géomarketing, de fidélisation clients...
- La direction commerciale pour le pilotage des réseaux, les prévisions des ventes, l'optimisation des territoires...
- Les ressources humaines pour la gestion des carrières,
- La direction de la production pour l'analyse qualité, la prévision des stocks, la gestion des flux, la fiabilité industrielle...
- La direction générale pour les tableaux de bord, indicateurs de pilotage, gestion d'alertes...

Exemple: un DW dans les télécoms

- **Sujets :**
 - Suivi du marché: lignes installées/ désinstallées, services et options choisis, répartition géographique, répartition entre public et différents secteurs d'organisations
 - Comportement de la clientèle
 - Comportement du réseau
- **Historique**
 - 5 ans pour le suivi du marché
 - 1 an pour le comportement de la clientèle
 - 1 mois pour le comportement du réseau
- **Sources**
 - Fichiers clients élaborés par les agences
 - Fichiers de facturation
- **Requêtes**
 - Comportement clientèle
 - Nombre moyen d'heures par client, par mois et par région
 - Durée moyenne d'une communication urbaine par ville
 - Durée moyenne d'une communication internationale

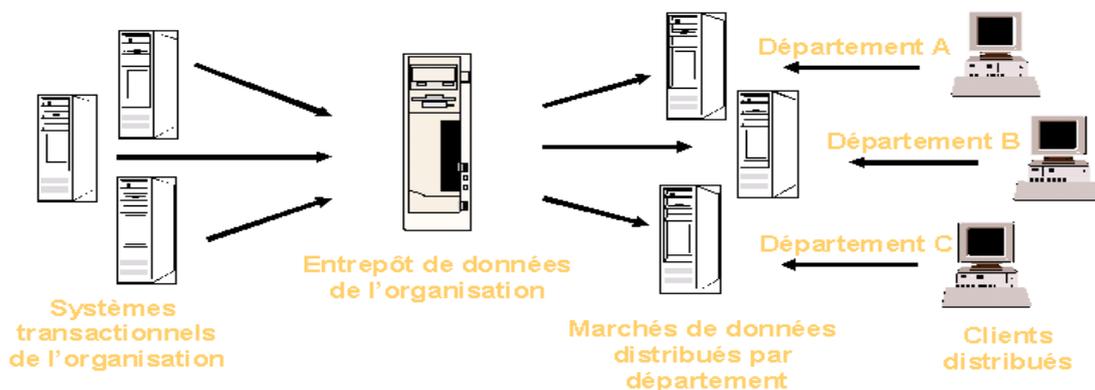
Architecture d'un Datawarehouse

Architecture centralisée



- Il s'agit de la version centralisée et intégrée d'un entrepôt regroupant l'ensemble des données de l'entreprise. Les différentes bases de données sources sont intégrées et sont distribuées à partir de la même plate-forme physique

Architecture fédérée



- Il s'agit de la version intégrée d'un entrepôt où les données sont introduites dans les marchés de données orientés selon les différentes fonctions de l'entreprise

Chapitre 2 : OLAP et Analyse multidimensionnelles

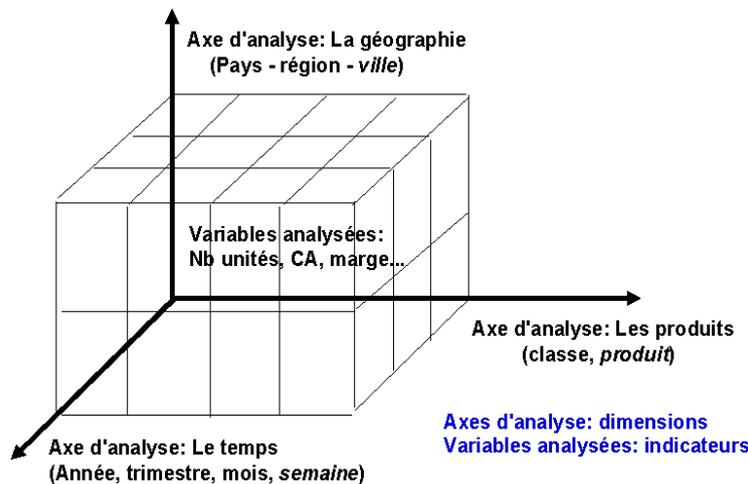
OLAP :

- « Il s'agit d'une catégorie de logiciels axés sur l'exploration et l'analyse rapide des données selon une approche multidimensionnelle à plusieurs niveaux d'agrégation ».
- OLAP vise à assister l'utilisateur dans son analyse en lui facilitant l'exploration de ses données et en lui donnant la possibilité de le faire rapidement.
 - L'utilisateur n'a pas à maîtriser des langages d'interrogation et des interfaces complexes
 - L'utilisateur interroge directement les données, en interagissant avec celles-ci

Modèle conceptuel

- Approche multidimensionnelle
 - 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

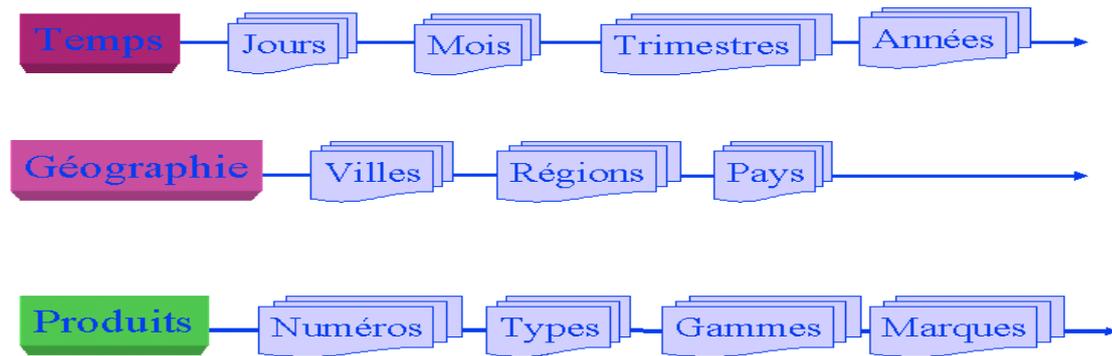
Vue multidimensionnelle



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



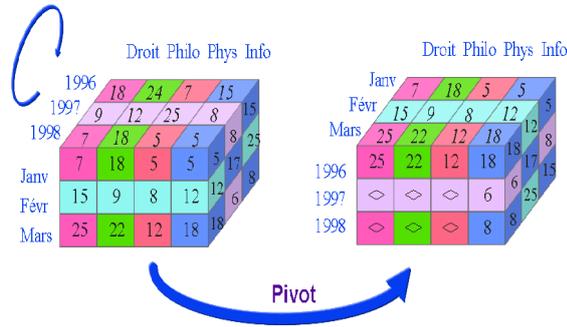
Les 12 règles OLAP (Edgar Frank Codd, 1993)

1. Vue multidimensionnelle:
 - Comme par exemple lorsqu'on souhaite analyser les ventes selon plusieurs dimension: par produit par région ou par période.
2. Transparence du serveur OLAP à différents types de logiciels
 - Elle s'appuie sur une architecture ouverte permettant à l'utilisateur d'implanter le système OLAP sans affecter les fonctionnalités du système central.
3. Accessibilité à de nombreuses sources de données
 - Le système OLAP doit donner accès aux données nécessaires aux analyses demandées.
 - Les outils OLAP doivent avoir leur propre schéma logique de stockage des données physiques
4. Performance du système de Reporting
 - L'augmentation du nombre de dimensions ou du volume de la base de données ne doit pas entraîner de dégradation visible par l'utilisateur.
5. Architecture Client/Serveur
 - La plus part des données pour OLAP sont stockées sur des gros systèmes et sont accessibles via des terminaux . Il est donc nécessaire que les outils OLAP soient capables de travailler dans un environnement Client/Serveur.
6. Dimensions Génériques
 - Toutes les dimensions doivent être équivalentes en structure et en calcul.
 - Toute fonction qui s'applique à une dimension doit être aussi applicable à une autre dimension.
7. Gestion dynamique des matrices creuses
 - Le schéma physique des outils OLAP doit s'adapter entièrement au modèle d'analyse spécifique créé pour optimiser la gestion des matrices creuses
8. Support Multi-Utilisateurs
 - Les outils OLAP doivent supporter les accès concurrents,
 - Garantir l'intégrité et la sécurité afin que plusieurs utilisateurs accèdent au même modèle d'analyse.
9. Opération sur les dimensions
 - Les opérations doivent pouvoir s'effectuer sur toutes les dimensions.
10. Manipulation intuitive des données
 - Toute manipulation doit être accomplie via une action directe sur les cellules du modèle sans utiliser de menus ou des chemins multiples à travers l'interface utilisateur.
11. Souplesse et facilité de constitution des rapports
 - La création des rapports dans les outils OLAP doit permettre aux utilisateurs de présenter comme ils le désirent des données synthétiques ou des résultats en fonction de l'orientation du modèle.
12. Nombre illimité de niveaux d'agrégation et de dimensions
 - Tout outil OLAP doit gérer au moins 15 à 20 dimensions.

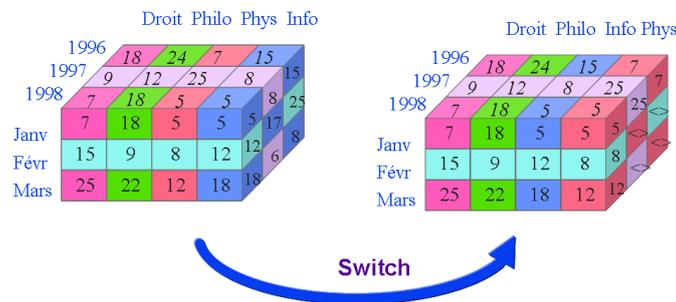
Opérations OLAP

Opérations sur la structure des cubes

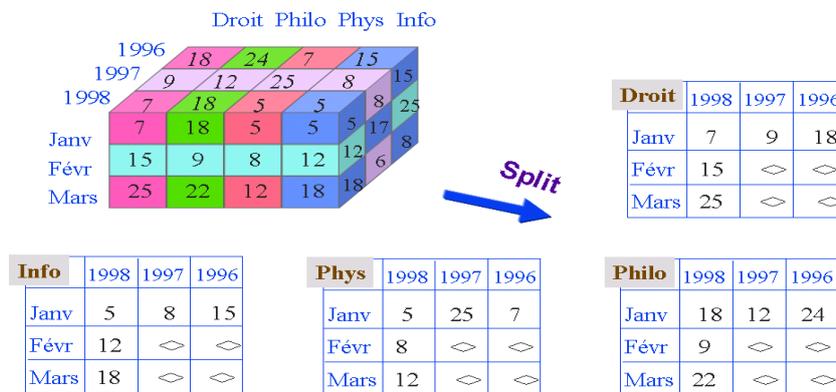
- Pivot (Rotation)



- Switch (Permutation)

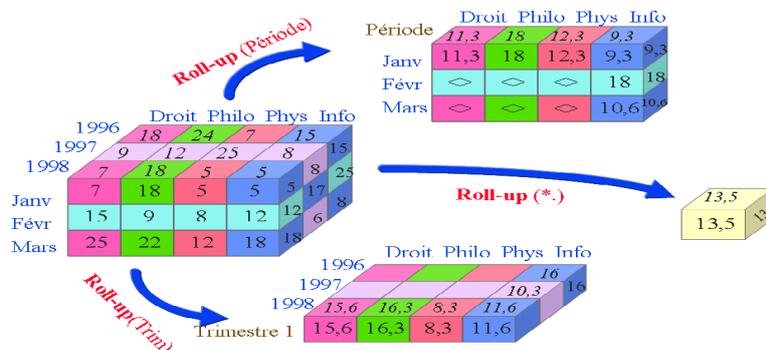


- Split (Décomposition)

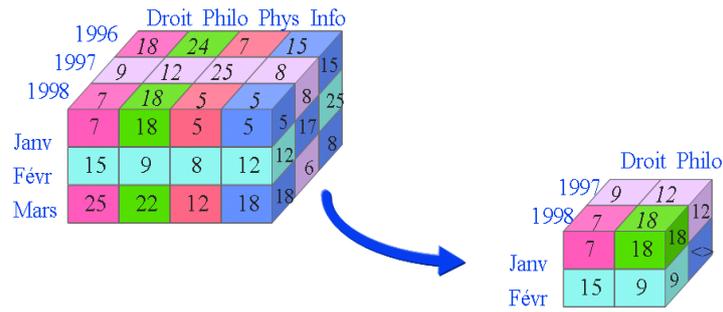


Opérations sur le contenu des cubes

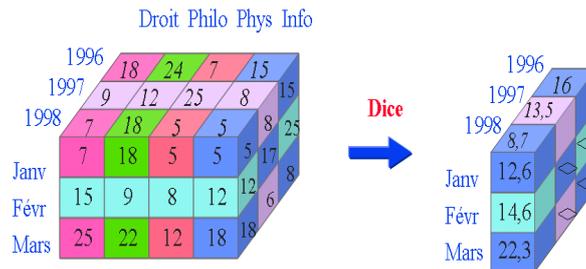
- Roll-up (passage au grain supérieur) / Drill-down (passage au grain inférieur)



- Slice (Restriction)

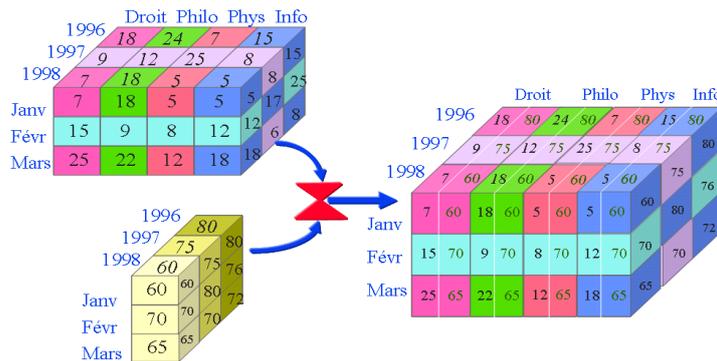


- Dice (Projection)

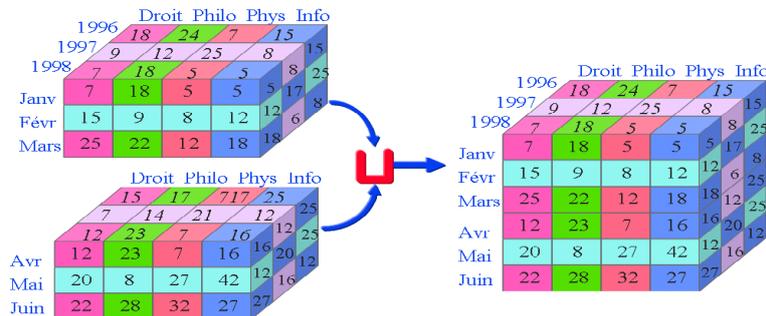


Opérations entre cubes

- Jointure



- Union



Exemple d'outils OLAP

- BusinessObjects, BusinessObjects : <http://www.businessobjects.fr/>
- SAS: <http://www.sas.com/offices/europe/france/software/technologies/olap/index.html>
- HarrySoftware , HarryPilot, HarryCube : <http://www.harrysoftware.com/jahia/Jahia/pid/61>

Chapitre 3: Modélisation et Conception d'un DW

Construction d'un Datawarehouse

Caractéristiques:

- Le Datawarehouse est différent des bases de données de production:
 - Les besoins pour lesquels on veut le construire sont différents
 - Il contient des informations historisées, organisées selon les métiers de l'entreprise pour le processus d'aide à décision
- Le Datawarehouse n'est pas un produit ou un logiciel mais un environnement, qui se bâtit et ne s'achète pas.

Phases de construction d'un DW:

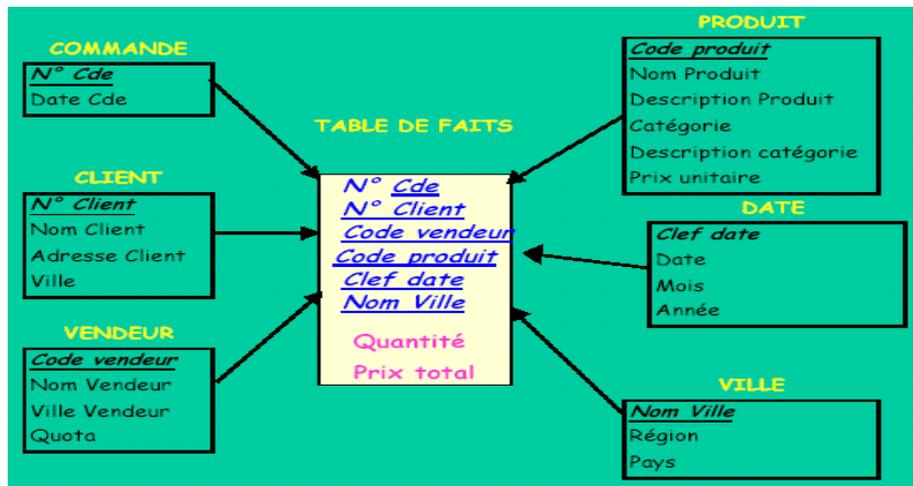
- Il y'a trois parties interdépendante qui relève la construction d'un Datawarehouse:
 - L'étude préalable qui va définir les objectifs, la démarche à suivre, le retour sur investissement,...
 - L'étude du modèle de données qui représente le DW conceptuellement et logiquement
 - L'étude de l'alimentation du Datawarehouse

Etude préalable

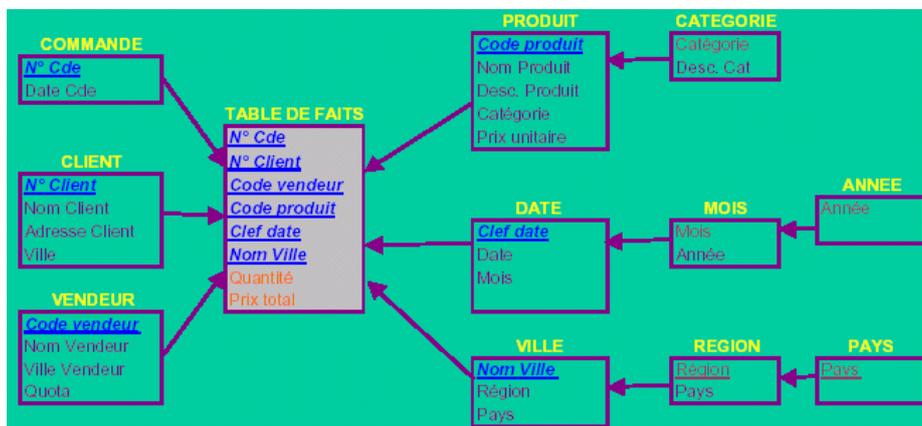
- Etude des besoins:
 - Définir les objectifs du DW
 - Déterminer le contenu du DW et son organisation, d'après:
 - Les résultats attendus par les utilisateurs,
 - Les requêtes qu'ils formuleront,
 - Les projets qui ont été définie
 - Recenser les données nécessaires à un bon fonctionnement du DW:
 - Recenser les données disponibles dans les bases de production
 - Identifier les données supplémentaires requises
 - Choisir les dimensions
 - Typiquement: le temps, le client, le produit, le magasin...
 - Choisir les mesures de fait
 - De préférences de quantités numériques additives
 - Choisir la granularité des faits
 - Niveau de détails des dimensions
- Coûts de déploiement:
 - Nécessite des machines puissantes, souvent une machine parallèle
 - Capacité de stockage très importante (historisation des données)
 - Evaluer la capacité de stockage
 - Equipes de maintenance et d'administration
 - Les coûts des logiciels
 - Les logiciels d'administration du DW
 - Les outils ETL (Extract-Transform- Loading)
 - Les outils d'interrogation et de visualisation
 - Les outils de Datamining

Modélisation

- Niveau conceptuel:
 - Un DW est basé sur une modélisation multidimensionnelle qui représente les données dans un cube
 - Un cube permet de voir les données suivant plusieurs dimensions:
 - Tables de dimensions
 - La table des faits contient les mesures et les clés des dimensions
- Niveau Logique:
 - Plusieurs schémas types sont proposés pour représenter un DW:
 - Schéma en étoile;
 - Schéma en flocon;
- Schéma en étoile
 - Une (ou plusieurs) table(s) de faits : identifiants des tables de dimension ; une ou plusieurs mesures.
 - Plusieurs tables de dimension : descripteurs des dimensions.



- Schéma en flocons
 - Raffinement du schéma étoile avec des tables normalisées par dimensions.

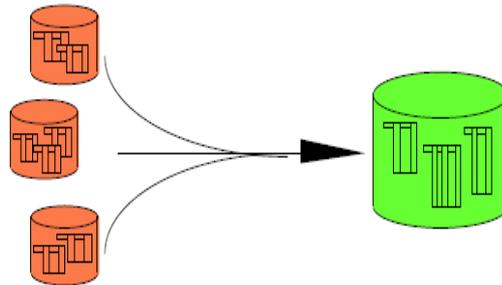


Alimentation

- L'alimentation est la procédure qui permet de transférer des données du système opérationnel vers le DW
- La conception de cette opération est une tâche complexe
 - 60 à 90 % de la charge totale d'un projet DW
- Elle doit être faite en collaboration avec l'administrateur des bases de productions
- Il est nécessaire de déterminer:
 - Quelles données seront chargées
 - Les transformations et les vérifications nécessaires
 - La périodicité et le moment de transferts des données

Alimentation d'un DW (ETL)

- Extraction (Extract)
- Transformation (Transform)
 - Filtrer
 - Homogénéiser
 - Nettoyer
 - Etc ...
- Chargement (Loading)



Extraction:

- Depuis les bases sources
- Périodique et Répétée
- Dater ou marquer les données envoyées
- Difficulté:
 - Ne pas perturber les applications OLTP
- Différentes techniques d'extraction:
 - Méthode Push: Le système opérationnel qui au fil des transactions alimente le DW
 - Méthode Pull : Le système décisionnel cherche périodiquement les données dans les bases de production

Transformation

- C'est une suite d'opérations qui a pour but de rendre les données cibles homogènes et puissent être traitées de façon cohérente.
 - Unification des modèles
 - Convertir / uniformiser les noms des attributs
 - Uniformiser les valeurs d'attributs
 - Nettoyer (Valeurs manquantes, aberrantes...)

Chargement

- C'est l'opération qui consiste à charger les données nettoyées et préparées dans le DW.
- C'est une opération qui peut être longue :
 - Mettre en place des stratégies pour assurer de bonnes conditions à sa réalisation
 - Définir la politique de rafraîchissement.
- C'est une phase plutôt mécanique et la moins complexe

Outils ETL

- BusinessObjects, Data Integrator,
<http://www.businessobjects.fr/>
- Oracle Corporation, Warehouse builder,
<http://www.oracle.com/technology/products/warehouse/index.html/>
- IBM, Websphere Datastage , <http://www.ascential.com/products/datastage.html>

Travaux dirigés

Exercice 1

Il s'agit de modéliser Le Datawarehouse des ventes d'une entreprise commerciale. Cette entreprise vend des produits regroupés par familles de produits. Une vente correspond à un produit et un seul; la vente est effectuée par l'un des vendeurs du service de vente spécialisé dans le produit. Le Datawarehouse doit pouvoir fournir le chiffre d'affaires des ventes d'un produit, par date, client, et vendeur, ainsi que toutes les sommations possibles de chiffre d'affaires.

Les objets Du Datawarehouse sont les suivants:

- **produit**, caractérisé par : code_produit, code_famille, etc...
- **client**, caractérisé par : code_client, nom, CSP (catégorie socio-professionnelle), etc ...
- **vente**, caractérisée par : code_date, code_produit, code_client, code_vendeur, Chiffre d'affaires
- **vendeur**, caractérisé par : code_vendeur, nom, code_service, etc...
- **date**, caractérisée par : code_dat, semaine, mois, année, etc...

1. Donner les définitions des termes suivants : table de faits, table de dimension, indicateur, hiérarchie.
2. Tracer le schéma en étoile dimensionnel du Datawarehouse, en précisant pour chaque table sa nature dimensionnelle (table de faits ou table de dimension), ses clés, ainsi que la nature des champs.

Exercice 2

Un opérateur Télécom propose à ses abonnés un boiter routeur ADSL qui groupe un accès à Internet, le téléphone et un bouquet de chaînes de télévision interactives. Nous ne nous intéresserons qu'à cette dernière fonction dans notre Datawarehouse (DW).

Les chaînes de télévision proposées par cet opérateur dans son bouquet sont des chaînes nationales et des chaînes à péage. Le téléspectateur peut réagir (interactivement) de plusieurs manières aux émissions et aux publicités qu'il reçoit au moyen de sa télécommande:

- A tout moment, il peut zapper vers une autre chaîne
- A tout moment, il peut éteindre son poste
- Pendant une publicité, il peut zapper à la pub suivante sans attendre la fin de la publicité en cours.
- Pendant une publicité, il peut demander plus d'information sur le produit présenté.

Le DW est centré sur les actions du téléspectateur. L'objectif de ce DW est de profiler au plus précis le foyer (ou plutôt le téléspectateur qui détient la télécommande) afin de maximiser la demande d'information sur les produits présentés par les publicités. En effet, l'opérateur perçoit plus d'argent de la part de l'annonceur quand le spectateur demande plus d'information au moment au l'annonce est passée. Le schéma de ce DW est constitué des tables suivantes (les clés primaires sont soulignées)

Date (CléDate, Année, Mois, JourDeMois, JourDeSemaine, TrancheHoraire, Heure, Minute)

Foyer (CléFoyer, NomAbonné, AnnéeNaissanceAbonné, Région, Département, District, Ville, Quartier, SituationFamille, RevenuFoyer, CS, NombreAdulte, NombreEnfant)

Emission (CléEmission, Chaîne, DateDébut, DuréeSeconde, TypeEmission, Catégorie, Annonceur)

Action (CléDate, CléFoyer, CléEmission, TypeAction, DuréeAction, DuréeRestante)

1. Quelle est la table de fait dans ce DW ? Justifiez votre réponse.
2. La granularité de la table de faits.
3. Quelles sont les dimensions de ce DW?
4. Donner le schéma en étoile de ce DW.
5. Formulez trois requêtes différentes qu'on peut poser au DW.
6. Donnez le nouveau schéma du DW si on décide de diminuer la taille de la table Foyer par une mini-dimension démographique.

Exercice 3

Une université cherche à étudier les facteurs influant sur la réussite de ses étudiants aux examens.

Pour cela elle décide de construire un Datawarehouse.

Elle souhaite pouvoir répondre aux questions suivantes:

- Quel est le nombre de réussites aux examens par cours, pour l'année 2007?
- Quel est le nombre de réussites aux examens d'un cours obligatoire, pour l'année 2007?
- Quel est le nombre de réussites aux examens par sexe (féminin, masculin), pour l'année 2007?
- Combien d'étudiants ayant un âge de 22 ans ont réussi leurs examens de base de données relationnelle?
- Quel est le nombre de réussites aux examens pendant le semestre d'hiver 2006?

Pour cela elle dispose des données suivantes: Pour chaque examen passé, on connaît l'âge et le sexe de l'étudiant, le nom du cours (les cours peuvent être regroupés en cours obligatoire et cours à option), la date de l'examen, la note obtenue et si l'examen est réussi ou non.

Proposez un modèle en étoile pour cette application. Recherchez tout d'abord les différentes dimensions et proposez une hiérarchie pour ces dimensions.

Exercice 4

Un distributeur (grossiste) approvisionne plusieurs magasins en produits, en effectuant au plus une livraison par jour et par magasin. Les informations qui figurent sur chaque bon de livraison sont les suivantes : le numéro du bon de livraison, la date de livraison, la référence du magasin, et pour chaque type de produit livré sa référence et la quantité livrée (le nombre d'articles). Ces informations sont stockées chez le distributeur, et accumulées pendant des longues périodes afin de les analyser pour améliorer le service de distribution.

Les analyses se font suivant plusieurs axes, et à plusieurs niveaux, en analysant les mouvements des produits par jour et par mois, par ville et par région, par fournisseur et par catégorie de produit.

On supposera qu'un fournisseur peut fournir au distributeur des produits dans plusieurs catégories et qu'une catégorie de produit peut être fournie par plusieurs fournisseurs.

Définir le schéma dimensionnel du Datawarehouse permettant d'analyser la quantité livrée par ville et catégorie de produits, en faisant apparaître clairement les dimensions et les indicateurs.

Etude de cas

Cas d'une compagnie d'assurance

Une compagnie d'assurance de biens (automobile, immobilier, responsabilité civile) possède une application transactionnelle de production permet de gérer les polices (contrats) de ses clients ainsi que les sinistres (accidents) déclarés par ces clients.

Gestion des polices

Pour gérer les polices, les agents d'assurance peuvent effectuer les transactions suivantes :

- Créer, mettre à jour ou supprimer une police d'assurance
- Créer, mettre à jour ou supprimer un risque (pour une police donnée)
- Créer, mettre à jour ou supprimer des biens assurés (voiture, maison) sur un risque
- Chiffrer ou refuser le risque
- Valider ou refuser la police

On enregistre dans ces transactions un grand nombre d'informations, et notamment : date d'écriture (date de la transaction), date d'effet (date de début d'assurance), client (personne(s) privée(s), personne morale), opérateur (employé, agent: chiffrage, vérificateur : validation), risque (produit vendu par la compagnie d'assurance), couverture (description des biens assurés), police (numéro de police, « note » de la police ou du risque,...) , transaction (code transaction).

Gestion des sinistres

Pour gérer les sinistres déclarés par les clients, les agents d'assurance ont à leur disposition les transactions suivantes :

- Créer, mettre à jour ou supprimer une déclaration de sinistre
- Créer, mettre à jour ou supprimer une expertise
- Créer, mettre à jour ou supprimer des paiements
- Clore le sinistre

Ces transactions comportent notamment : date d'écriture (date de la transaction), date d'effet (date de déclaration), client, opérateur, risque, biens sinistrés, police, les tiers impliqués dans le sinistre, les montants financiers (limites, déjà payé, reste à payer, ...), code transaction.

Conception

A partir de cette application transactionnelle, on veut créer un Datawarehouse permettant de répondre aux questions suivantes :

- Pour chaque bien assuré, on veut connaître le montant de la prime (somme annuelle payée par le client pour assurer le bien) associée au bien assuré, et le nombre de transactions du mois pour ce bien.
 - De même on veut pouvoir sortir des tableaux de bord par sinistre avec le total payé dans le mois et le total reçu dans le mois pour ce sinistre.
1. Faire le schéma en étoile d'un Datamart « **Police** » ne prenant pas en compte les sinistres.
 2. De même, faire le schéma en étoile d'un Datamart « **Sinistre** ».
 3. Faire un seul Datawarehouse de ces deux Datamarts.

Cas d'une banque

Une banque distribue une carte de paiement « carte de crédit » à ses clients. Elle décide de réaliser un Datawarehouse (DW) afin de faire le suivi des paiements suivants effectués avec la carte :

- a. Voyages en avion,
- b. Locations de voiture,
- c. Hôtellerie.

Elle veut faire un suivi indépendant de chacun des paiements a, b ou c, mais aussi avoir la possibilité d'un suivi global.

A chaque déplacement en avion, la compagnie aérienne lui envoie un fichier contenant les éléments suivants: identification de la carte de paiement, coordonnées du client et de la compagnie aérienne; ville de départ, ville d'arrivée, n° du vol, date du vol, n° du billet, classe du siège, distance parcourue, date d'achat et prix payé.

Les loueurs de véhicule transmettent après chaque location: identification de la carte de paiement, coordonnées du client et de la société de location de véhicules, catégorie du véhicule, date de début de location, date de fin de location, nombre de jours, distance parcourue, date de réservation et prix payé.

L'hôtel transmet à chaque séjour: identification de la carte de paiement, coordonnées du client et de l'hôtel, catégorie de chambre, date de début de séjour, date de fin de séjour, nombre de nuitées, date de réservation, prix de l'hébergement et prix de la restauration.

1. Un premier DW ne concerne que les déplacements en avion.

Etablir le modèle dimensionnel. Faire clairement apparaître les dimensions et les indicateurs. Ce DW doit permettre de répondre aux questions suivantes : quel est le chiffre d'affaires (CA) par client, par date de voyage (et par mois, trimestre et année), par compagnie aérienne, par ville de destination?

2. De même, établir deux autres modèles dimensionnels, l'un pour les locations de voiture, l'autre pour l'hôtellerie.

Dans le cas de la location de voiture, on souhaite éditer le CA, le nombre de jours de location, et le kilométrage pour chaque: client, date de réservation, ville, loueur, et catégorie de véhicule.

Dans le cas de l'hôtellerie, on veut des tableaux de bord par client, hôtel, ville, date de début de séjour, catégorie de chambre, faisant apparaître le nombre de nuitées, le prix total payé.

3. On veut maintenant regrouper ces trois DW en un seul, afin de répondre aux questions supplémentaires suivantes :

Quel est le CA total induit par un déplacement en avion ? Quelle est la durée du séjour ? Quel est le CA en location de voiture ? En hôtellerie ? On désire ici pouvoir éditer les détails de CA par période de temps et par client, ville de destination, ville de location (si différente), ville d'hébergement (si différente), compagnie aérienne, loueur et hôtelier, et faire tous les regroupements utiles.

Figurer le modèle dimensionnel d'un tel DW, en faisant clairement apparaître les dimensions et les indicateurs.