Ressource R5.C.06 Exploitation de la base de données (DATA5) \$\frac{\$\$545}\$

IUT La Rochelle BUT INFO3 Développeur IA

M. Sakkari

Exploitation d'un entrepôt de données OLAP

Plan

- 1. Le Processus Extract-Transform-Load (ETL)
- 2. Exploitation d'un entrepôt de données OLAP
- 3. L'Informatique Décisionnelle

I. Qu'est-ce que l'OLAP?

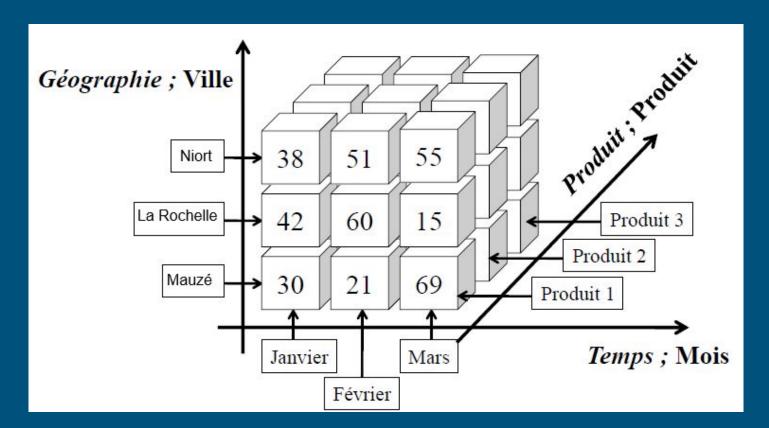
OLAP (Online Analytical Processing) est une technologie d'analyse de données:

- Permet aux utilisateurs d'analyser des données provenant de plusieurs systèmes de bases de données simultanément.
- Contrairement aux BDRs bidimensionnelles, les données OLAP sont multidimensionnelles, permettant une comparaison des informations de différentes manières.

I. Qu'est-ce que l'OLAP?

- OLAP n'est pas un logiciel ou un outil unique.
- Plutôt une technologie et une approche pour l'analyse des données multidimensionnelles.
- Fournit une analyse détaillée pour prendre des décisions stratégiques et améliorer les performances de l'entreprise.
- Basée sur des outils puissants.

- Les données multidimensionnelles sont de données organisées d'une manière permet de les analyser sous plusieurs angles ou dimensions.
- C'est-à-dire l'information peut être comparée de nombreuses façons différentes.
- Par exemple, une entreprise peut comparer ses ventes d'ordinateurs en juin avec ses ventes en juillet, puis comparer ces résultats avec les ventes d'un autre endroit, qui pourraient être stockées dans une base de données différente.



Concepts Clés des Données Multidimensionnelles :

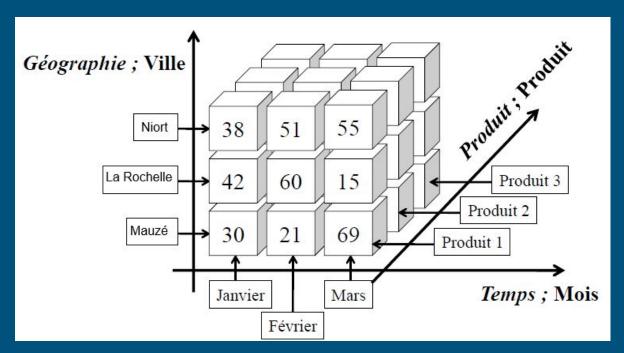
- Les dimensions sont les perspectives ou les entités selon lesquelles les données peuvent être analysées.
- Exemple : Temps (jours, mois, années), Produits (catégories, sous-catégories), Régions (pays, villes).

Concepts Clés des Données Multidimensionnelles :

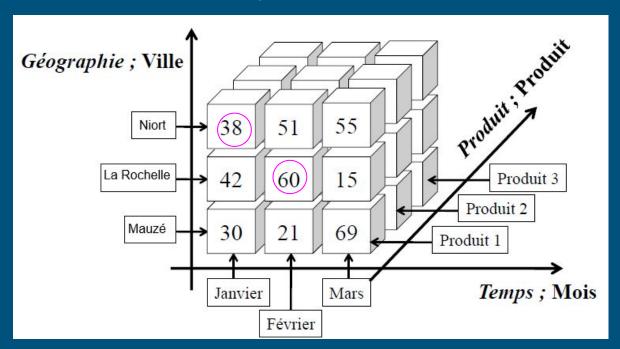
- Mesures (Faits): Les mesures sont les valeurs numériques ou quantitatives que nous souhaitons analyser.
- Exemple : Ventes, bénéfices, quantités vendues.

- Un cube OLAP est une structure de données supérieure aux bases de données relationnelles.
- Les cubes peuvent afficher et additionner de grandes quantités de données, tout en permettant de parcourir le contenu des points de données.
- Les données sont regroupées et/ou segmentées pour gérer la plus grande variété de questions de l'utilisateur.

Exemple Cube OLAP de 3 dimensions.

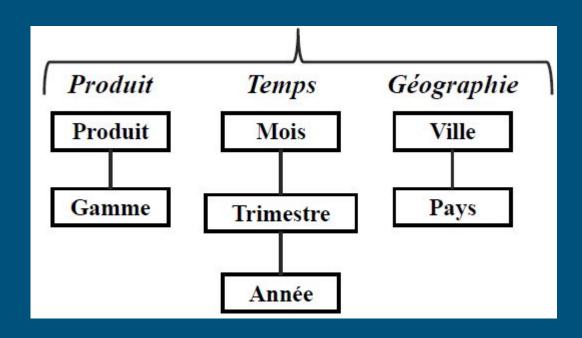


Il montre le chiffre d'affaires par Ville, Mois et Produit



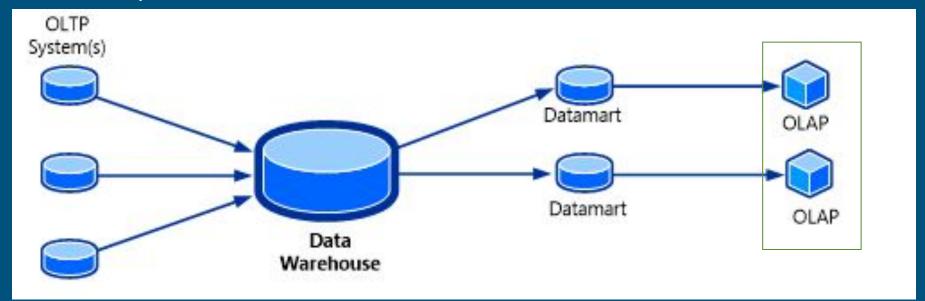
La hiérarchie des trois dimensions (Produit , Temps et Géographie) de ce

cube est comme suit:



 Le cube OLAP, également appelé cube multidimensionnel ou hypercube, est une structure de données permet une analyse instantanée des données. La topologie du système est indiquée dans l'illustration suivante..

• Les cubes OLAP sont en quelque sorte l'apport final d'une solution d'entrepôt de données.



- OLTP (Online Transaction Processing) est une catégorie de systèmes d'information qui gèrent les transactions quotidiennes de manière rapide, fiable et efficace.
- Les systèmes OLTP sont essentiels pour les opérations commerciales de base, telles que les achats en ligne, les opérations bancaires, les réservations de billets et bien d'autres activités transactionnelles

- Un Data Warehouse est une base de données relationnelle. Il recueille des données de sources variées et hétérogènes dans le but principal de soutenir l'analyse et faciliter le processus de prise de décision.
- En matière d'intégration dans le système de données existant, le fonctionnement du Data Warehouse est basé sur le processus ETL permettant de charger les données issues des différentes applications.

Un Data Warehouse est défini comme un ensemble de données orientées sujet, intégrées, variables dans le temps et non volatiles.

- a. Orienté sur le sujet : Organisé par thème, il est possible d'utiliser le Data Warehouse pour analyser n'importe quel secteur particulier de l'entreprise.
- b. Intégré : Avant toute utilisation, les données récupérées de sources hétérogènes internes ou externes sont intégrées au Data Warehouse.

- c. Variante temporelle : Les données passées sont également conservées dans le Data Warehouse, contrairement à certains des systèmes transactionnels traditionnels où seules les données les plus récentes sont stockées. Cela permet de visualiser l'évolution dans le temps des différentes données.
- d. Non volatile : Les données une fois stockées dans l'entrepôt de données ne peuvent jamais être modifiées.

- Un Data Mart est le sous-ensemble d'un Data Warehouse. Tandis que le Data Warehouse couvre plusieurs sujets, un Data Mart est spécialisé sur un seul thème.
- Le Data Warehouse doit stocker une grande quantité de données historiques et permettre des requêtes rapides et complexes à travers les données en utilisant des technologies comme OLAP et ses dérivés voire In-memory.

Les différentes opérations OLAP:

1. Slice : cette opération consiste à extraire un sous-ensemble de données en sélectionnant une dimension spécifique.

Par exemple, si vous avez un cube de données sur les ventes, vous pourriez choisir de ne voir que les ventes d'un certain produit pour une année spécifique.

Les différentes opérations OLAP:

2. Dice : La fonction "dice" permet de créer un sous-cube en sélectionnant plusieurs dimensions et en appliquant des conditions spécifiques sur ces dimensions.

Par exemple, vous pourriez vouloir voir les ventes de plusieurs produits dans plusieurs régions sur une période donnée.

Les différentes opérations OLAP:

3. **Drill Down**: le "drill down" permet d'affiner les données en passant à un niveau de détail plus granulaire.

Par exemple, au lieu de voir les ventes par année, vous pouvez choisir de les afficher par trimestre, puis par mois.

Les différentes opérations OLAP:

4. Drill Up (ou Roll-Up): L'opération inverse du "drill down", le "drill up" regroupe les données pour les afficher à un niveau plus agrégé.

Par exemple, vous pourriez regrouper les ventes par année au lieu de par trimestre.

Les différentes opérations OLAP:

5. Pivot (ou Rotate) : Cette opération permet de changer la perspective d'analyse des données en faisant pivoter les dimensions.

Par exemple, vous pouvez afficher les ventes par produit en lignes et par région en colonnes, puis inverser cette configuration.

Les différentes opérations OLAP:

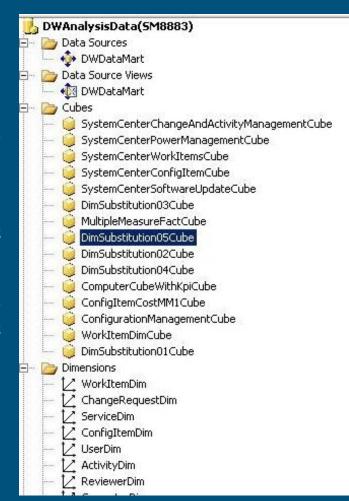
6. Agrégation: L'agrégation consiste à résumer les données sur une ou plusieurs dimensions, souvent en utilisant des fonctions comme SUM, AVG, MIN, et MAX. Cela permet d'obtenir des résumés rapides des données.

Les différentes opérations OLAP:

6. Agrégation: L'agrégation consiste à résumer les données sur une ou plusieurs dimensions, souvent en utilisant des fonctions comme SUM, AVG, MIN, et MAX. Cela permet d'obtenir des résumés rapides des données.

IV. Cubes OLAP de BIDS

- L'illustration suivante est une capture d'écran de SQL Server Business Intelligence Development Studio (BIDS) montrant les principaux éléments requis pour les cubes OLAP.
- Ces éléments sont la source de données, la vue de source de données, les cubes et les dimensions
- La vue de source de données contient toutes les relations établies entre les tables, telles que les clés primaires et étrangères



IV. Cubes OLAP de BIDS

- **S**QL Server Business Intelligence Development Studio (BIDS) est un environnement de développement intégré fourni par Microsoft pour la création de solutions de business intelligence.
- BIDS est utilisé pour créer des projets de rapports, des projets d'intégration de données (ETL) et des projets d'analyse multidimensionnelle.

Construction d'un Entrepôt de Données pour les Consultations Médicales

Nous voulons construire un entrepôt de données pour stocker et analyser les informations sur les consultations médicales dans un pays. L'objectif est de connaître le nombre de consultations selon différents critères comme les personnes, les médecins, et les spécialités.

Les informations sont stockées dans les relations suivantes :

PERSONNE: Contient les informations sur les patients.

Schéma : PERSONNE(id, nom, tel, adresse, sexe)

MÉDECIN: Contient les informations sur les médecins.

Schéma: MEDECIN(id, tel, adresse, spécialité)

CONSULTATION: Contient les informations sur les consultations.

Schéma: CONSULTATION(id_med, id_pers, date, prix)

- Proposer un schéma relationnel qui tient compte de la date, du jour de la semaine, du mois, du trimestre et de l'année.
- Quelle est La Table Des faits ?
- Quels sont les faits?
- 4. Combien de dimensions ontété retenues? Quelles sont-elles?
- 5. Quelles sont les hiérarchies des dimensions?
- Donner une représentation du cube OLAP.

- 7. Quelles opérations OLAP il faut appliquer pour obtenir les informations suivantes :
- Coût total des consultations par médecin en 2012 et 2013.
- Nombre de consultations par de la semaine, par spécialité et par sexe du patient.
- Le coût des consultations par patiente pour les mois d'octobre.

Fin

Merci Pour Votre Attention