

Atelier 03 : Entrepôt de données

Correction de l'exercice de cours :

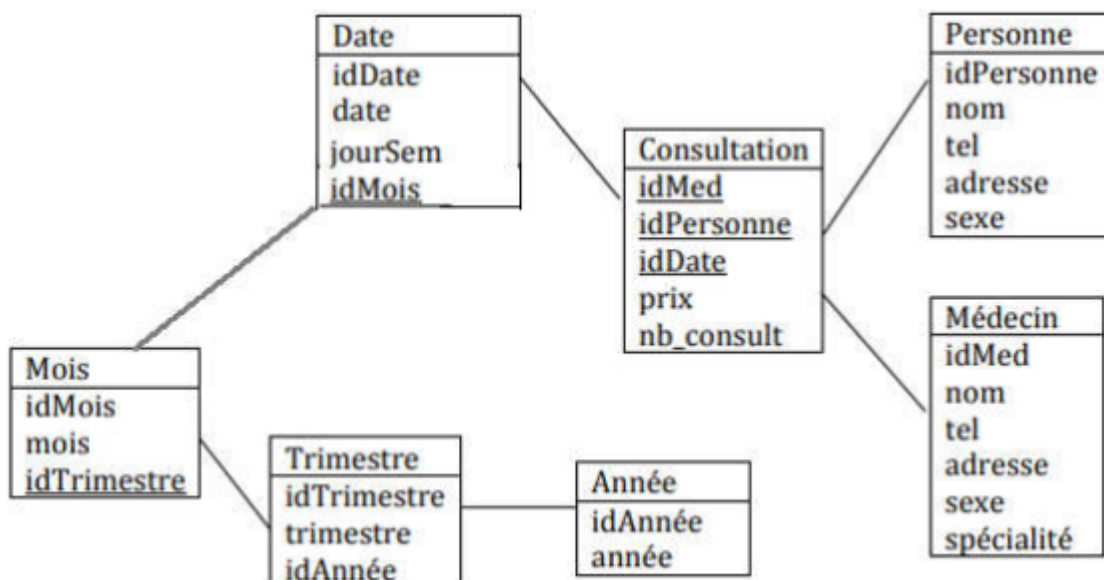
On veut construire un entrepôt de données afin de stocker les informations sur les consultations d'un pays. On veut notamment connaître le nombre de consultations, par rapport à différents critères (personnes, médecins, spécialités, etc.) Ces informations sont stockées dans les relations suivantes :

Personne (id, nom, tel, adresse, sexe)

Médecin (id, tel, adresse, sexe, spécialité)

Consultations (id_med, id_pers, date, prix)

Proposer un schéma conceptuel qui tient en compte de la date, du jour, de la semaine, du mois, du trimestre et de l'année.



a) Quelle est la table des faits ?

Consultation

b) Quelles sont les mesures ?

Le prix et le nombre de consultations (nb_consult)

c) Combien de dimensions ont été retenue ? Quelles sont-elles?

Trois dimensions : Médecin, Personne, Temps

d) Quelles sont les hiérarchies des dimensions ?

Date : Année <-- Trimesre <-- Mois <-- Date

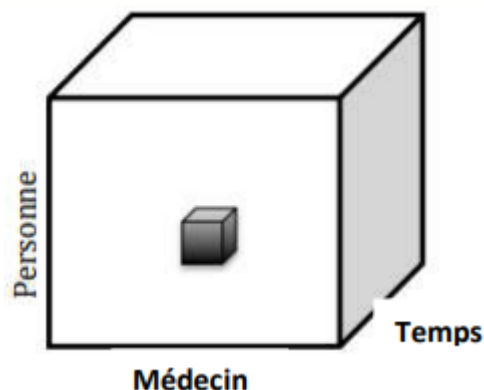
Date : JourSemaine <-- Date

Médecin : Spécialité <-- Médecin

Médecin : Sexe <-- Médecin

Personne : Sexe <-- Personne

e) Faites une représentation du cube OLAP sans tenir compte des hiérarchies



f) A partir de ce cube, indiquer quelles opérations OLAP (roll up, drill down, slice, dice) faut-il appliquer pour obtenir les informations suivantes :

- Le coût total des consultations par médecin en 2012 :

-Roll up sur Temps de Date à Année

-Roll up sur Personne de idPersonne (un seul individu) à Tous

-Slice Année = 2012

g) Le nombre de consultations par jour de la semaine par spécialité et par sexe du patient

-Roll up sur Temps de idDate à idJourSemaine

-Roll up sur Medecin de idMed à spécialité

-Roll up sur Personne de idPersonne à sexe

h) Le coût des consultations pour toutes les patientes pour le mois d'octobre

-Roll up sur Temps de idDate à idMois

-Roll up sur Personne de idPersonne à sexe

--Roll up sur Medecin de idMed à tous

-Dice (Mois = Octobre, sexe = Féminin)

i) Le coût des consultations par patiente pour le mois d'octobre

-Roll up sur Temps de idDate à idMois

-Roll up sur Medecin de idMed à tous

-Dice (Mois = Octobre, sexe = Féminin)

Etude de cas:

La société Orion :

Cette société fictive, présente au niveau mondial, est spécialisée dans la commercialisation d'articles de sport et d'extérieur. Les données disponibles regroupent des informations sur :

- les employés
- les produits
- les clients
- les commandes

- les fournisseurs

Le siège social aux États-Unis, gère des filiales en Belgique (depuis 1999), Pays Bas, Allemagne, Royaume-Uni, Danemark, France, Italie, Espagne et Australie. Les produits sont vendus en magasin, par catalogue et par internet. Une carte de fidélité : 'Orion Star Club', propose beaucoup d'avantages. L'historique d'information va du 1er janvier 1998 au 31 décembre 2002.

Structure de l'organisation

Le siège social héberge la majeure partie des fonctions administratives, soit un nombre important d'employés, entre 600 et 800. Le siège social centralise aussi la gestion des stocks, la vente par catalogue, la vente par internet et l'import - export. Néanmoins, certains employés gèrent aussi ces fonctions depuis les différentes filiales.

Les employés sont enregistrés dans la base de données selon cinq niveaux :

- Pays
- Compagnie
- Département
- Section
- Groupe

Les informations complémentaires sur les employés sont notamment :

- Date d'entrée et de départ de l'employé
- Date de début et de fin de contrat (pour certain contrat)
- Adresse
- Sexe
- Salaire
- Responsable hiérarchique

L'offre

La société propose environ 5500 références. Certaines ne sont pas vendues dans tous les pays, d'autres, de part les volumes commercialisés, reflètent certaines particularités régionales, certains sports nationaux. Tous les noms sont fictifs.

Les produits sont organisés selon 4 niveaux :

- Ligne de produit
- Catégorie de produit
- Groupe de produit
- Produit

Chaque produit a un coût et un prix de vente. Le système informatique gère tous les prix en dollars. En utilisant les dates de début et de fin, ces prix varient en fonction du temps. Cet historique est sauvegardé. Le système gère aussi les remises pour certains produits, à certaines périodes. Les prix sont généralement uniques de part le monde.

Les clients

Les clients sont repartis à travers le monde, notamment dans les pays où se trouvent des filiales, mais pas uniquement. Les noms et adresses sont fictifs, même si les villes, régions/comtés et pays, sont réels. La base de données enregistre environ 90 000 clients, pas tous actifs.

L'adresse des clients comprend tout ou partie des informations suivantes :

- Rue
- Code postal
- Ville
- Région / département / comté
- Etat
- Pays
- Continent

Les clients sont classés dans des groupes en fonction de leur activité d'achat.

Les commandes

Chaque commande pointe vers le commercial qui a enregistré la vente. Environ 980 000 commandes sont enregistrées, commandes qui reflètent notamment les saisonnalités. Chaque commande comprend une ou plusieurs lignes, une ligne par produit.

Les fournisseurs

Chaque produit provient d'un fournisseur qui est basé dans un pays, mais toutes les commandes sont passées par le siège social. Il y a 64 fournisseurs, mais un seul fournisseur par produit.

La société Orion souhaite améliorer sa performance à l'aide d'un système décisionnel. Voici quelques questions qui ont été recensées et auxquelles devrait répondre le système mis en place :

- Quels sont les produits qui se vendent le mieux ?
- Quels sont les produits en perte de vitesse ?
- Quels sont les produits qui contribuent très peu au chiffre d'affaire pour un pays et une année donnés ? Est-ce que ces produits peuvent être remisés ?
- Quelle est la marge générée par ce groupe de produit ?
- Est-ce que la marge dépend de la quantité vendue ?
- Est-ce que les remises font augmenter les ventes ?
- Est-ce que les remises font augmenter la marge ?
- Quels sont les commerciaux qui font le plus de ventes ?
- Quels sont les commerciaux qui performant le mieux par pays, sexe, âge, salaire ?
- Quels groupes de clients sont identifiés ?
- Quels sont les clients les plus rentables ?
- Quels fournisseurs proposent des produits rentables ?
- Quelle est la moyenne et l'écart-type du chiffre d'affaire ?

- Quelles sont les variables qui expliquent le mieux l'importance du chiffre d'affaire ?
- Y-a-t'il une différence significative entre la moyenne de la somme du chiffre d'affaire géré par les commerciaux de sexe féminin et celle des commerciaux de sexe masculin ?

Il faut donc construire un entrepôt de données capable de répondre aux besoins de requête, de reporting, et d'analyses avancées

1. Schéma de l'entrepôt

Il existe plusieurs types de schémas utilisés dans la conception d'entrepôts de données, chacun ayant ses propres avantages et inconvénients. Le **schéma en étoile** (Star Schema) est le plus simple et le plus utilisé, où une table de faits centrale est reliée à plusieurs tables de dimensions. Ce modèle est particulièrement efficace pour les requêtes OLAP (Online Analytical Processing) car il simplifie la structure des données et optimise les performances des requêtes. Les dimensions dans un schéma en étoile sont dénormalisées, ce qui signifie que toutes les informations pertinentes à une dimension sont stockées dans une seule table de dimension, facilitant l'accès et l'analyse.

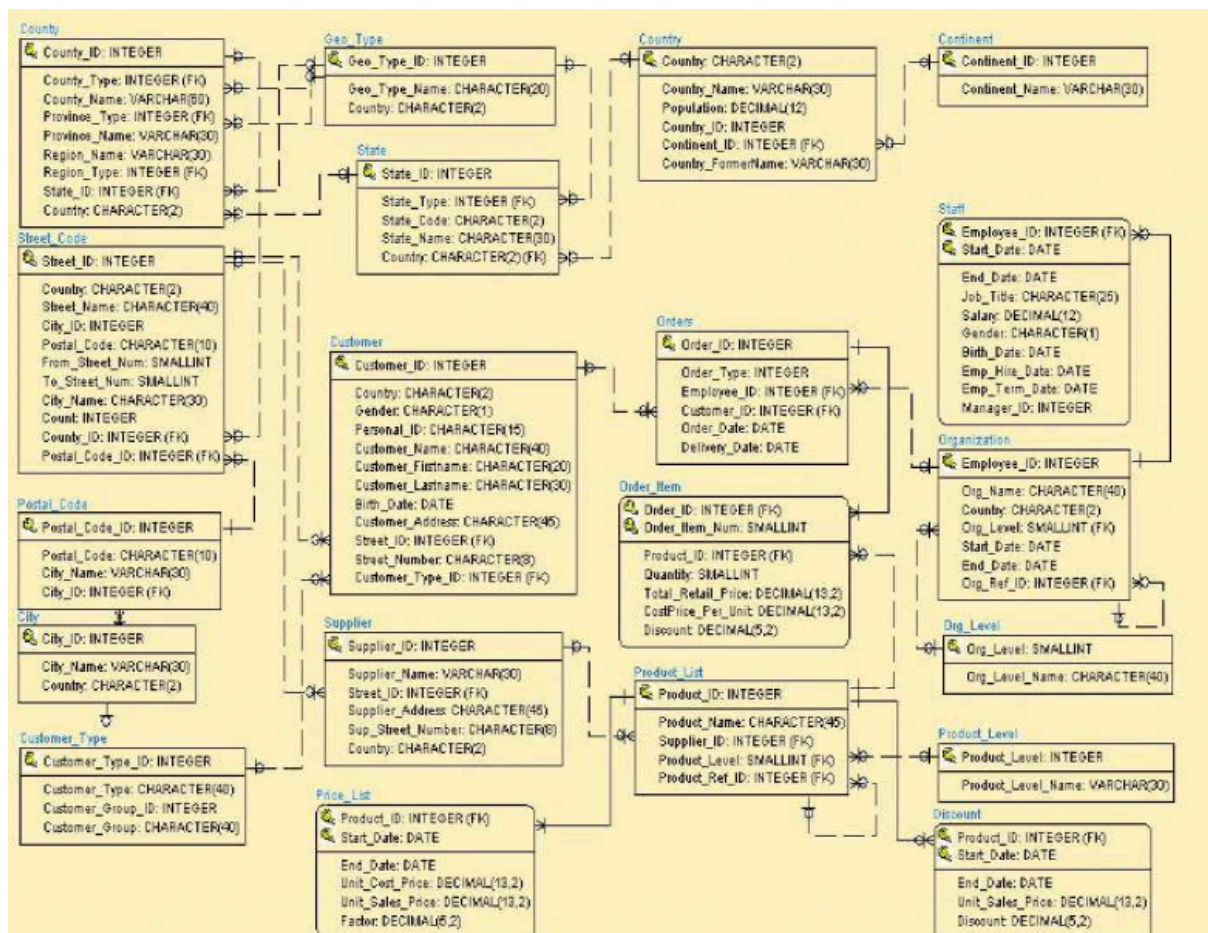
- Proposez un schéma en étoile pour l'entrepôt de données de la société Orion.
- À partir du schéma en étoile proposé, écrivez les commandes SQL nécessaires pour créer les tables de l'entrepôt de données de la société Orion. Assurez-vous d'inclure les clés primaires et les clés étrangères appropriées pour chaque table.
- Utilisez des commandes SQL pour répondre aux questions suivantes à partir des données de l'entrepôt :

Quels sont les produits qui se vendent le mieux ?

Quels sont les commerciaux qui font le plus de ventes ?

Quels sont les clients les plus rentables ?

Voici le schéma relationnel de la base de données opérationnelle de l'entreprise, qui fournira les données à l'entrepôt

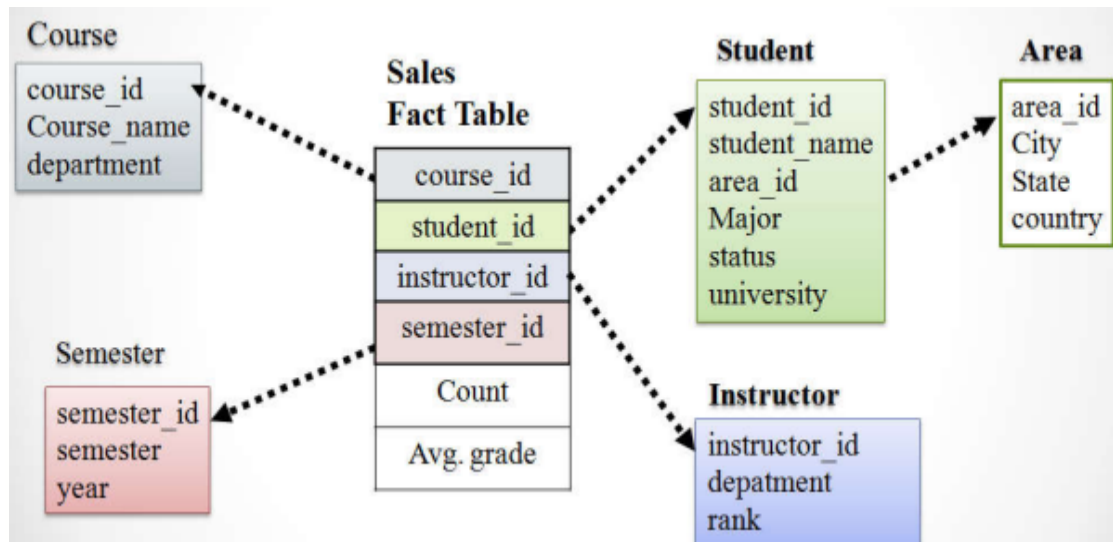


Bon travail

Atelier 03 : Entrepôt de données

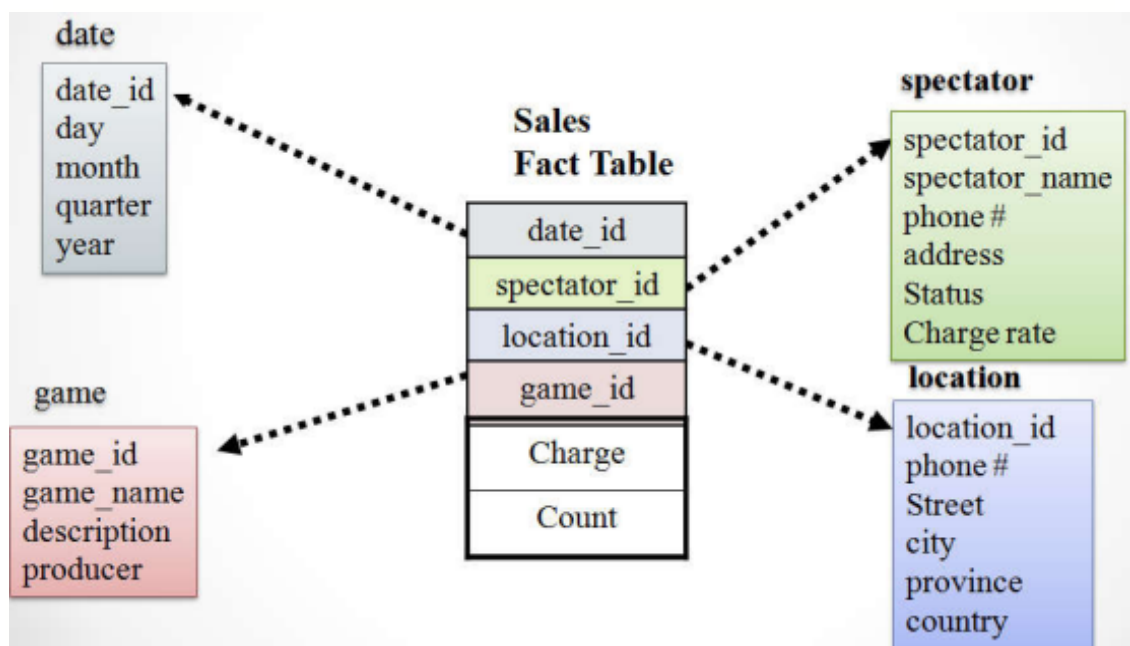
Exercice 1 :

1. On considère le schéma conceptuel suivant :



Indiquer quelles opérations OLAP (roll up, drill down, slice, dice) faut-il appliquer pour obtenir la note moyenne des cours en Informatique pour chaque étudiant.

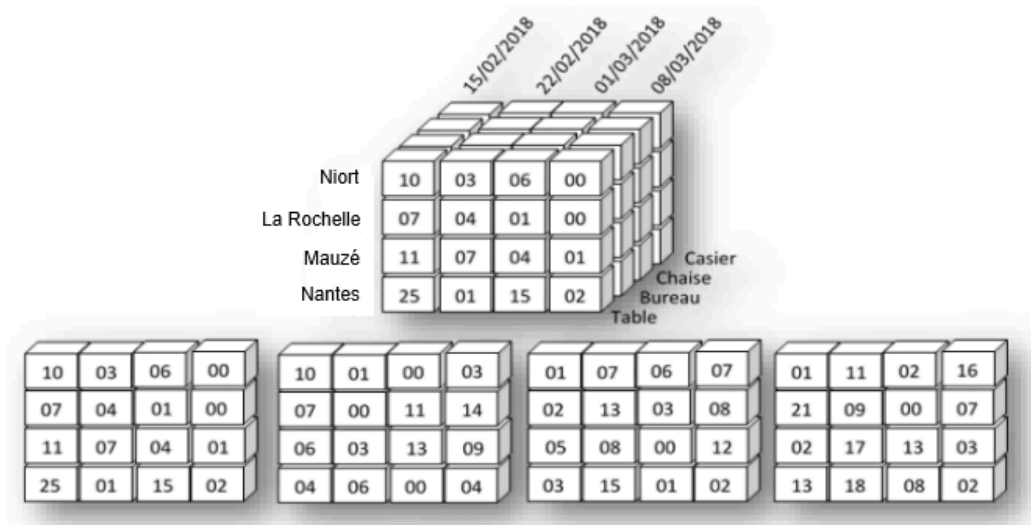
2. Indiquer quelles opérations OLAP (roll up, drill down, slice, dice) faut-il appliquer pour obtenir la charge totale payée par un étudiant spectateur à "GM Place" en 2010



3. Donner l'équivalent "Opération OLAP" de chacune des requêtes suivantes :

- SELECT TIME, LOCATION, PRODUCT ,SUM(REVENUE) AS PROFIT FROM SALES
GROUP BY ROLLUP(TIME, LOCATION, PRODUCT);
- SELECT TIME, LOCATION, PRODUCT ,SUM(REVENUE) AS PROFIT FROM SALES
GROUP BY ROLLDOWN (TIME, LOCATION, PRODUCT);
- SELECT PRODUCTS, SUM(REVENUE) FROM SALES
WHERE PRODUCTS= 'OPV'
GROUP BY PRODUCTS ;
- SELECT PRODUCTS, SUM(REVENUE) FROM SALES
WHERE PRODUCTS= 'EL' AND LOCATION='EUROPE'
GROUP BY PRODUCTS;

4. Soient les données suivantes : Dimension : Location, Produit, Temps



Déterminer les dimensions et les hiérarchies des dimensions.

5. Donner le détail des opérations qui ont permis d'avoir ces cubes

