

Data Warehouse
Série N° 2
Année universitaire 2016-2017
Mr. Imade BENELALLAM

Objectif de la série : Cube OLAP, Conception dimensionnelle et dimensionnement des entrepôts de données.

Exercice 1 : Extensions OLAP pour SQL

Soit une table de fait R (A, B, C, M) où A, B et C sont les dimensions et M est la mesure. Supposons que les attributs A, B, C prennent chacun deux valeurs différentes, et que toutes les combinaisons de valeurs sont présentes dans la table de faits.

- 1) Combien de n-uplets ont les résultats des requêtes suivantes (justifiez vos réponses) :
Q1. Select A, B, C, sum(M) From R Group by rollup (A, B, C)
Q2. Select A, B, C, sum(M) From R Group by cube (A, B, C)
- 2) Si l'on ne dispose pas des opérateurs cube et rollup, combien faudrait-il de requêtes group by pour chacune de ces requêtes ?

Exercice 2 : Cube OLAP

On considère un entrepôt de données permettant d'observer les ventes de produits d'une entreprise. Le schéma des tables est le suivant :

CUSTOMER (*cust_id, city, state, region, country*)

PRODUCT (*prod_id, prod_name, category, cost_price, sell_price, supplier*)

TIME (*time_key, month, month_name, quarter, year*)

PURCHASES (*prod_id, time_key, cust_id, ship_date, purchase_price, shipping_charge*)

- 1) Indiquez quelles sont les tables de faits et les tables de dimension du schéma en étoile de cet entrepôt.
- 2) On veut transformer ce schéma en schéma en flocon. Donnez la nouvelle représentation de la table TIME.
- 3) Ecrire la requête SQL permettant d'afficher la moyenne des ventes (*purchase_price*) par catégorie (*category*) et par année (*year*).
- 4) Calculer le cube permettant d'obtenir la somme des ventes par catégorie, par année et par région en utilisant la clause GROUP BY CUBE.
- 5) On considère la même requête, mais en remplaçant la clause GROUP BY CUBE par la clause GROUPE BY ROLLUP. Quelles différences ce changement implique-t-il?

Exercice 3 : Modélisation décisionnelle d'un entrepôt de données pour assurance maladie

Le Ministère de la Santé et du Bien-Etre de Grolang vous sous-traite la réalisation d'un entrepôt de données pour réaliser des études sur les dépenses de santé dans son beau-pays. Les bases de production de cet entrepôt sont les systèmes d'information des centres de sécurité sociale et des assurances santé complémentaire de Groland qui gèrent les dossiers (électroniques) des assurés.

Le schéma de l'entrepôt est constitué des tables suivantes (les clés primaires sont soulignées)

***Date**(CléDate, Année, Mois, JourDeMois, JourDeSemaine, TrancheHoraire, DrapeauVacances)*

***Assuré**(CléAssuré, MoisNaissance, AnnéeNaissance, MoisDécès, AnnéeDécès, Région, Département, District, Ville, Quartier, RevenuAssuré, RevenuFoyer, CatégorieSocioProfessionnel, SousCatégorieSocioProfessionnel, DomaineActivité, CaissePrimaire, CaisseComplémentaire, DrapeauAssuréPrincipal)*

***Praticien**(CléPraticien, Spécialité, SousSpécialité, Région, Département, District, Ville, Quartier, MoisNaissance, AnnéeNaissance, DrapeauConventionné)*

***Acte**(CléDate, CléAssuré, CléPraticien, CléPathologie, MontantActes, MontantPriseEnChargeCaissePrimaire, MontantPriseEnChargeCaisseComplémentaire, NombreMedicamentsPrescrits, MontantPharmacologieGenerique, MontantPharmacologieNonGenerique, MontantDesActesComplémentaires, DrapeauActesComplémentairesBiologie, DrapeauActesComplémentairesChirurgie, DrapeauActesComplémentairesKinésithérapie, DrapeauActesComplémentairesRadiologie, NombreDeJoursDArrêtDeTravail, CoutJoursDArret).*

***Pathologie**(CléPathologie, DesignationNormalisé, Spécialité, SousSpécialité, TauxDIncapacité, DuréeTraitement, Chronicité, DrapeauMaladieProfessionnelle)*

Rétro-Conception

- 1) Quelle est la table de fait dans cet entrepôt ? Justifiez !
- 2) A votre avis, il y a t'il des dimensions douteuses dans cet entrepôt ? Rappelez la définition et justifiez.
- 3) Donnez les nouvelles tables si on décide de diminuer la taille de la table Assurée par une mini-dimension démographique

Dimensionnement

- 4) Donnez le nombre de faits présents dans la table de fait.
 - Nombre d'assurés 60 Millions
 - Nombre d'actes par praticien et par jour 20
(Un praticien travaille 300 jours par an)
 - Nombre de praticiens 300 000
 - Montant total moyen d'un acte 100 Euros
 - Nombre d'actes supplémentaires prescrit par acte 0,1
 - Nombre d'années 6
 - Coûts annuel des actes 180 Milliard d'Euro

Taille des clés 4 octets

Taille des attributs numériques 4 octets

Taille des attributs booléens (comme les drapeaux !) 1 octet

- 5) Donnez la taille d'un enregistrement de la table de fait.
- 6) Donnez la taille (en Octets) de stockage de la table de fait.

Rapports

- 7) Donnez la requête SQL qui donne le top 10 des sous-spécialités des pathologies qui ont entraîné le plus de dépenses (montant des actes + montant pharmacologies)?
- 8) Donnez le rapport mensuel de progression du ratio des montants des médicaments génériques par rapport aux médicaments non génériques.