

TD BI (Conception Conception d'un DW)

Exercice 1

Une agence de voyage aimerait pouvoir analyser ses données afin de planifier de meilleures campagnes de promotion auprès de ses clients. Plus particulièrement, elle aimerait analyser le nombre et le montant des ventes en fonction:

- De la destination: hôtel, ville, pays, région, catégorie de région (ex: bord de mer, alpine, etc.), catégorie de destination (ex: familial ou non), catégorie hôtel (ex: 1-4 étoiles) ;
 - De la date d'achat: jour de l'année, jour de la semaine, mois, année, saison touristique (ex: basse ou haute saison);
 - De la date de départ: jour de l'année, jour de la semaine, mois, année, saison touristique (ex: basse ou haute saison);
 - Du forfait: nombre de personnes, nombre de nuits, type de forfait (ex: tout inclus, repas inclus, etc.), type de chambre (ex: standard, suite, penthouse, etc.) ;
 - Du client: groupe d'âge, sexe, adresse, type d'acheteur (ex: nouveau, récurrent, etc.) ;
 - Du canal de vente: catégorie (ex: magasin, internet, etc.) ;
 - De la promotion: catégorie (ex: 2 pour 1, rabais 10%, rabais 25%, etc.), début et fin de validité ;
 - Du mode de paiement: catégorie (ex: crédit, comptant, etc.) ;
- 1) Proposez un schéma en étoile permettant de faire ces analyses. Identifiez clairement les clés primaires et étrangères des tables de faits et de dimension;
 - 2) Identifiez, pour chaque table de dimension, une hiérarchie de niveaux de granularité n
(e.g., attribut1 \leftarrow attribut2 \leftarrow ...) ;
 - 3) Proposez une stratégie d'agrégation ajoutant une nouvelle table de faits agrégés.

Donnez le code SQL permettant de créer cette nouvelle table.

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.

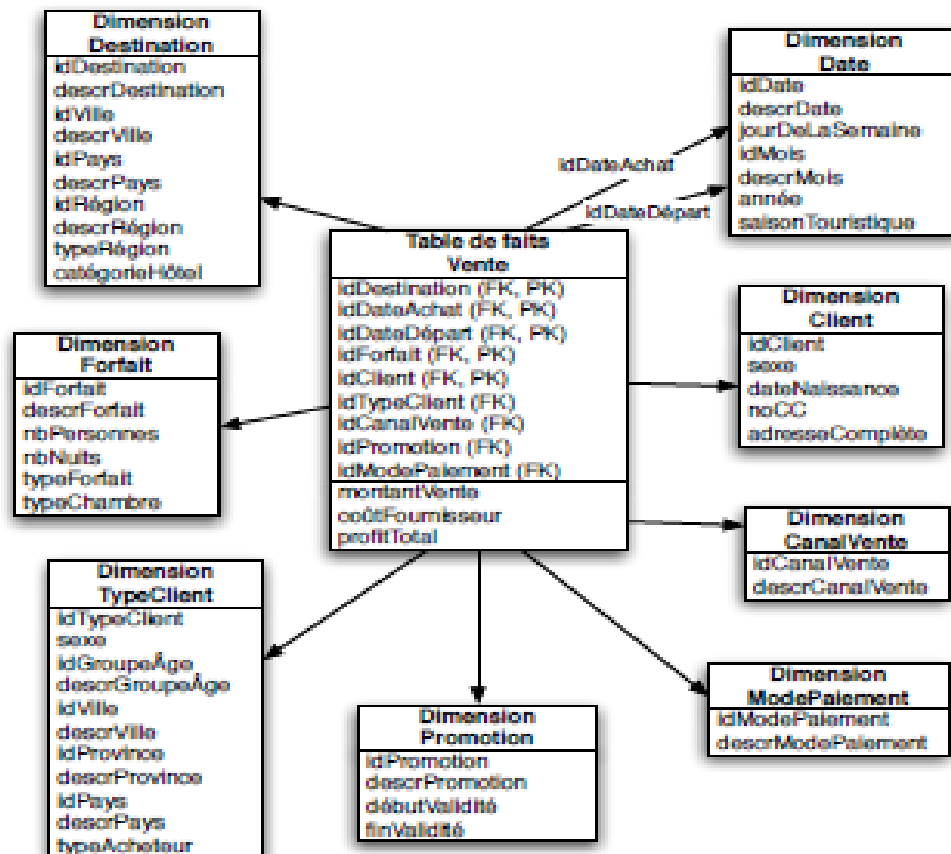
Les analyses se font suivant plusieurs axes, et à plusieurs niveaux, en analysant les mouvements des produits par jour et par mois, par ville et par région, par fournisseur et par catégorie de produit.

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.

Correction

a)



Notes:

- La plupart des attributs dimensionnels ont un ID ainsi qu'un champ descriptif. Par exemple, dans la table Date, le mot 'Novembre' n'est pas suffisant pour identifier avec précision ce mois, car on le retrouve dans chacune des années. Il faut donc un attribut idMois (ex: '11/2010') ainsi qu'un attribut descriptif descrMois (ex: 'Novembre'). C'est la même chose pour l'attribut ville: le même nom de ville peut se trouver dans plusieurs pays ou même plusieurs fois dans un même pays;

- Nous avons créé une table TypeClient selon la stratégie de mini-dimension. L'avantage est que la table TypeClient peut être pré-générée (toutes les combinaisons possibles de sexe, ville, groupe d'âge, etc.). De même, les tables Destination, Date, Forfait, Promotion et CanalVentes peuvent également être pré-générées et ne sont (presque) jamais modifiées. Seule la table de dimension Client est modifiée à chaque fois qu'un client s'ajoute au système;
- La clé primaire de la table de faits Vente est une clé composée car il est très rare que l'on accède individuellement les lignes de cette table. En revanche, les clés primaires des tables de dimension sont toujours des clés artificielles simples (ex: NUMBER).

b)

Les niveaux d'une hiérarchie doivent avoir une relation 1 à plusieurs: un parent peut avoir plusieurs enfants (ex: une année a plusieurs mois) mais chaque enfant n'a qu'un seul parent (ex: le mois '11/2010' appartient uniquement à l'année 2010).

Table de dimension	Hiérarchies
Destination	idDestination ← idVille ← idPays ← idRégion ← tous
Date	idDate ← idMois ← année ← tous
Forfait	idForfait ← tous
Client	idClient ← tous
TypeClient	idTypeClient ← idVille ← idProvince ← idPays ← tous
CanalVente	idCanal ← tous
Promotion	idPromotion ← tous
ModePaiement	idModePaiement ← tous

c)

Pour définir la stratégie d'agrégation, il faut choisir, pour chaque dimension, un niveau hiérarchique permettant de faciliter l'analyse. L'objectif est d'accélérer les calculs en pré-calculant les agrégations faites dans les requêtes analytiques fréquentes.

Ainsi, on prévoit que les analyses se feront aux niveaux suivants:

Dimension	Niveau hiérarchique retenu
Destination	idPays
Date (achat)	tous
Date (départ)	idMois
Forfait	idForfait
Client	tous
TypeClient	idProvince
CanalVente	idCanal
Promotion	idPromotion
ModePaiement	tous

Avec ces niveaux d'agrégation, on pourrait analyser les ventes par pays, mois de départ, forfait, province de client, canal de vente et promotion utilisée.