

Travaux Dirigés Datawarehouse

Pr. A. ELOUARDIGHI

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_date, 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 boîtier 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 où 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 et les indicateurs de ce DW?
4. Donner le schéma en étoile de ce DW.
5. 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 2010?
- Quel est le nombre de réussites aux examens d'un cours obligatoire, pour l'année 2010?
- Quel est le nombre de réussites aux examens par sexe (féminin, masculin), pour l'année 2010?
- 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 2009?

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.