

Serie de TD N°1

DATA WAREHOUSE

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_dat, 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

Une entreprise de fabrication de vaisselle jetable souhaite mettre en place un système d'information Décisionnel sous la forme d'un data mart (un mini entrepôt de données) pour observer son activité de ventes au niveaux des différents lieux de distributions de ses articles et cela dans plusieurs villes. Ces lieux de distributions sont renseignés par leur enseigne, leur type (en fonction de leur surface), leur adresse (code postal et ville), leur département, leur région. Les ventes sont renseignées selon une période qui se décline en mois, en trimestre et année. Les ventes sont observées par le nombre d'articles selon le type, et le chiffre d'affaire.

- Quel est le fait à observer ?

- Quels sont les axes d'analyse, et les mesures ?

- Construire le modèle en étoile de ce data mart.

Exercice 3

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 de ce DW?
4. Donner le schéma en étoile de ce DW.
5. Formulez trois requêtes différentes qu'on peut poser au DW.
6. Donnez le nouveau schéma du DW si on décide de diminuer la taille de la table Foyer par une mini-dimension démographique.

Exercice4

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 5

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.