

Exercice 1 : Modélisation d'un cas basique

1. Donner le schéma multidimensionnel en étoile d'une entreprise de ventes de produits. L'entreprise veut analyser le chiffre d'affaire, les quantités et les notes par client, par produit et par date. Le chiffre d'affaire correspond au montant de vente ; la quantité correspond à la quantité vendue et la note correspond à une évaluation faite par le client sur la qualité du produit acheté. Les notes sont attribuées sur une échelle de 0 à 10. Chaque produit est caractérisé par sa référence, son nom et son prix et chaque client est caractérisé par un numéro, son nom, son numéro de téléphone et son adresse.

1. Supposons que dans le schéma en étoile de l'exercice précédent, chaque produit appartient une seule catégorie (alimentation, produits de nettoyage...) et que l'entreprise veut analyser l'activité par catégorie (regrouper les données). Aussi, l'entreprise veut analyser son activité par jour, par mois, par année, et par ville du client. Modifier le schéma précédent en incluant les nouvelles informations d'analyse. Identifier les hiérarchies.

2. Transformer le schéma précédent vers un schéma en flocon de neige.

3. Supposons que l'entreprise veut également analyser son activité par trimestre (groupe de 3 mois), et par semestre (groupe de six mois) et par quadrimestre (groupe de quatre mois). Modifier le schéma en flocon de neige pour inclure ces informations. Identifier les différentes hiérarchies de cette dimension.

4. Supposons que chaque produit peut également être catégorisé par rapport à sa taille (petit, moyen et grand) et sa fréquence de demande (rare, normale, fréquente). Ajouter ces informations au schéma en flocon de neige.

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 suivant le temps de visionnage du spectateur au moment où l'annonce est passée.

Q 1.1 Donnez le schéma des tables du data-warehouse. Quelle est la table de fait ? Quelles sont les dimensions de ce DW ? Justifiez votre réponse

Q 1.2 Donner le schéma en étoile de ce DW.

Exercice 3 :

TU Hôtels est une petite chaîne d'hôtels ayant des propriétés dans plusieurs états américains. L'entreprise possède une base de données centralisée pour stocker et faire le suivi des réservations de ses clients. En 2008, ils ont installé des restaurants appelés Café in the Hotel dans plusieurs de leurs hôtels. Un système est employé pour faire le suivi des commandes et les relayer aux employés dans les cuisines.

TU Hôtels aimerait utiliser les données qu'ils ont emmagasinées pour mieux comprendre la performance de leurs hôtels et restaurants. Ils ont également accès à une base de données de critiques de clients provenant du site web HotelComplainer.com. La tâche est de faire la conception de deux magasins de données (data marts) utilisant les données provenant des trois sources mentionnées ci-haut.

Vous devrez faire un schéma en étoile pour chaque magasin de données en choisissant les dimensions, les faits, et les attributs à partir des sources, dont le schéma relationnel est fourni à la page suivante.

La table suivante présente les questions analytiques auxquelles devra répondre vos magasins de données :

Data mart 1: Performance des hôtels	Data mart 2: Performance des restaurants
<ul style="list-style-type: none">• Durant quel mois y a-t-il le plus grand nombre de réservations de chambre?• Quelle est la saison morte pour les hôtels situés dans une région particulière?• Quels hôtels génèrent le plus de revenus (non attribuables aux restaurants)?• Quel est la durée moyenne des séjours dans les hôtels de 4.5 étoiles ou plus?• Les fumeurs restent-ils plus longtemps que les non-fumeurs?• Pour un hôtel donné, combien y a-t-il de clients provenant d'un autre état?	<ul style="list-style-type: none">• Quels restaurants génèrent le plus de revenus?• Les restaurants les mieux cotés génèrent-ils plus de revenus?• Quel est l'item commandé le plus souvent dans une région particulière?

Pour compléter l'exercice, vous devrez suivre les étapes suivantes :

- 1) Identifiez le principal événement d'affaires pour chaque magasin de données; Posez-vous la question suivante : "Quel est l'évènement d'affaires qui génère la (les) métrique(s) de performance?"
- 2) Identifiez les attributs associés aux faits. Posez-vous la question suivante : "Comment l'évènement d'affaires est-il mesuré?"
- 3) Identifiez les dimensions et leurs attributs. Posez-vous la question suivante : "Quelles données caractérisent les différents aspects de l'évènement d'affaires?"
- 4) Élaborez le schéma en étoile selon les principes vus en classe.