

## Techniques de data warehousing

### - Mini projet -

Le but de ce projet est d'étudier la conception et la réalisation d'un cube et ses applications dans un entrepôt de données. Le projet est divisé en plusieurs étapes. Avant de commencer le projet, il vous faut créer une nouvelle table de faits *sales\_fact\_projetDW* en réunissant la table *sales\_fact\_1997*, *sales\_fact\_1998* et la table *sales\_fact\_dec\_1998* dans votre base de données MySQL.

La première étape vise la *conception et réalisation d'un cube de données* en utilisant l'éditeur de Schéma (Schema editor) dans le logiciel Mondrian Schema Workbench pour produire un cube de données à analyser. Dans l'étape suivante on utilise le cube des données produit pour *les exploiter et étudier des différents phénomènes* dans la base de données *foodmart*.

Il est possible de travailler en binôme pour ce projet.

#### Etape 1

Vous êtes demandé de créer d'abord un cube de données, nommé *projetDW*, avec les descriptions suivantes :

- La table *sales\_fact\_projetDW* est la table de faits de ce cube.
- Les tables de dimension *Store*, *Time* et *Product* présentent les trois dimensions partagées (Dimension Usage) pour ce cube.
- En inspirant le cube *Sales*, vous ajoutez les dimensions *Customers* et *Yearly\_Income* dans le cube *projetDW*. Il n'est pas nécessaire de mettre les propriétés de membre pour la dimension *Customers*.
- Vous créez et intégrez les dimensions *Education*, *Occupation* et *Number\_children\_at\_home* à partir de la table *customer* dans la base de données *foodmart*.
- On utilise les mesures « Store sales », « Store cost » et « Units sales ».

Toutes les définitions concernant le cube *projetDW* seront sauvegardées dans le fichier nommé ***ProjetDW\_votreNom1\_votreNom2.xml***. Ce fichier et tous les scripts SQL pour créer votre table de faits seront retournés ensemble avec le projet.

## Etape 2

Vous utilisez l'outil éditeur de requête MDX (MDX Query editor) dans le Schema Workbench pour analyser des tendances de clients.

Vous êtes demandé d'analyser, d'observer et, s'il est possible, de trouver des tendances particulières de certains groupes de clients.

Les clients dans ce cube pourraient être catégorisés en fonction de leur revenu annuel (Yearly\_Income), leur profession (Occupation), leur niveau d'éducation (Education) et le nombre d'enfants à la maison (Num\_children\_at\_home) avec des informations dans la dimension Customers.

Votre analyse doit comprendre au moins des aspects suivants :

1. Le pourcentage de dépenses pour chaque famille de produits (Product family) par rapport aux différents niveaux de revenu, aux différentes professions et aux différents nombres d'enfants à la maison. Ensuite, les mêmes analyses pour toutes les catégories (Product category).
2. Pour 1997 et 1998, montrer les moyennes mobiles centrées de « Store sales », « Store cost » et « Units sales » pour toutes les catégories (Product category) dans la famille de boissons (drink), dans la famille de nourritures (food), dans la famille de non consommable (non-consumable) pour les différents niveaux de revenu, pour les différentes professions, pour les différents niveau d'éducation et pour les différents nombres d'enfants à la maison.
3. Trouvez les premières cinq marques (Brand name) ayant les plus grands chiffres d'affaires (Store sales) au niveau de catégories (Product category) dans la famille de nourritures (food) pour les différents niveaux de revenu, pour les différentes professions, pour les différents niveaux d'éducation et pour les différents nombres d'enfants à la maison.
4. Similaire à l'analyse précédente, trouvez les premières cinq marques (Brand name) ayant les plus grands bénéfices (Store sales – Store cost) au niveau de catégories (Product category) dans la famille de nourritures (food) pour les différents niveaux de revenu, pour les différentes professions, pour les différents niveaux d'éducation et pour les différents nombres d'enfants à la maison. Est-ce qu'il y a des remarques particulières ?
5. Est-ce que l'hypothèse « les familles nombreuses ont la tendance d'acheter des nourritures (food) moins chers » est vraie dans ces données? Donner votre opinion avec des justifications.
6. Réfléchissez au moins une autre analyse pour observer des tendances de clients.

## Etape 3

Effectuer les mêmes analyses en utilisant le SQL avec la base de données *foodmart* dans le *MySQL*.

Rédigez un résumé pour votre analyse et votre découvertes, avec des justifications et des schémas, dans un rapport simple nommé ***rapportProjetDW\_votreNom1\_votreNom2.pdf***. Les requêtes MDX et les requête SQL que vous avez utilisées doivent aussi être retournés ensemble avec votre projet.