

TP Entrepôts de données

Miage 1 FA/FC

Exercice 1 : Schéma en étoile et rapports

Soit la base de données *Comptoir*.

On veut à partir de cette base de données opérationnelle, produire un entrepôt de données pour étudier les faits suivants :

- Le fait Ventes permettant d'analyser les ventes selon le temps, les vendeurs, les produits et les clients
- Le fait Achat permettant d'analyser les achats de produits selon les fournisseurs, les produits et le temps.

Question 1 : Implémenter en SQL les tables illustrées en annexes.

Question 2 : Proposer des schémas en étoile pour étudier ces deux faits.

Question 3 : Créer une base de données *entrepôt* qui implémente les schémas en étoile de la question1.

Alimenter cet entrepôt avec des requêtes ajout et en utilisant des données provenant des tables de la base de données Comptoir

Question 4 : Réaliser sous OpenCalc des tableaux d'analyse croisée dynamiques en vous basant sur des cubes de données analysant les deux faits de l'introduction.

Question 5 : Créer une application extracteur injecteur pour alimenter l'entrepôt à partir d'un ou de plusieurs fichiers XML

Annexes

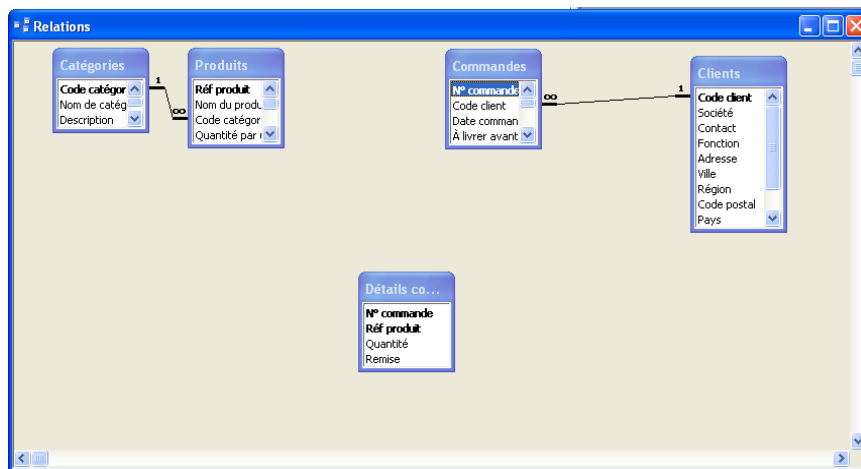


Figure 1 Relations entre les tables

Produits : Table		
Nom du champ	Type de données	Description
Réf produit	Texte	
Nom du produit	Texte	
Code catégorie	Texte	Identique au Code catégorie de la table Catégories.
Quantité par unité	Texte	(par exemple : bouteille 1 litre x 12).
Prix unitaire	Monétaire	
Unités en stock	Numérique	
Unités commandées	Numérique	
Niveau de réapprovisionnement	Numérique	Nombre minimum d'unités à garder en stock.
Indisponible	Oui/Non	'Oui' signifie que le produit n'est plus disponible.

Figure 2 Table Produit

Clients : Table		
Nom du champ	Type de données	Description
Code client	Texte	Code unique de cinq caractères basé sur le nom du client.
Société	Texte	
Contact	Texte	
Fonction	Texte	
Adresse	Texte	Rue ou boîte postale.
Ville	Texte	
Région	Texte	État ou province.
Code postal	Texte	
Pays	Texte	
Téléphone	Texte	Le numéro de téléphone inclut le code du pays ou de la zone.
Fax	Texte	Le numéro de téléphone inclut le code du pays ou de la zone.

Figure 3 Table Client

Commandes : Table		
Nom du champ	Type de données	Description
N° commande	Texte	
Code client	Texte	Identique au Code client de la table Clients.
Date commande	Date/Heure	
À livrer avant	Date/Heure	
Date envoi	Date/Heure	
Port	Monétaire	
Destinataire	Texte	Nom de la personne ou de la société responsable de la réception.
Adresse livraison	Texte	Numéro et nom de rue, ce champ ne doit pas contenir le code postal.
Ville livraison	Texte	
Région livraison	Texte	État ou province.
Code postal livraison	Texte	
Pays livraison	Texte	

Figure 4 Table Commande

Nom du champ	Type de données	Description
Code catégorie	Texte	
Nom de catégorie	Texte	Nom de la catégorie de produit.
Description	Mémo	
Illustration	Objet OLE	Image représentant la catégorie de produit.

Figure 5 Table Catégorie

Nom du champ	Type de données	Description
N° commande	Texte	Identique au N° commande de la table Commandes.
Réf produit	Texte	Identique à la Réf produit de la table Produits.
Quantité	Numérique	
Remise	Numérique	

Figure 6 Table Détails Commandes

Exercice 2 : Rappels SQL et requêtes de classement sous Oracle

Sur les tables Emp et Dept créée à partir du script « Exercice_1.sql »

- Quel est le classement des salaires des employés par département pour les départements 10 et 30 ?
- Idem en ôtant les trous dans le classement.
- Quelle est la moyenne arrondie des salaires par département ?
- Attribuer un numéro en fonction de l'ordre du salaire décroissant par département, pour les départements 10 et 20.
- Quel est le salaire total versé par profession (répondre avec deux méthodes possibles avec et sans group by) ?
- Quel est la différence entre un group by et un partition by dans une requête SQL ?
- Quel est le montant total des salaires versés, tout département et job confondus, par département et par département et job ?
- Idem en ôtant toute confusion avec les valeurs nulles (répondre avec deux méthodes possibles).

Exercice 3 : OLAP

Dans le monde de l'informatique décisionnelle, le ROLAP (Relational On-Line Analytical Processing) est une solution OLAP qui voit un entrepôt de données comme un SGBD relationnel et qui permet de l'interroger en utilisant des requêtes SQL étendu (en général très complexes et très exigeantes en terme de ressources et de temps d'exécution).

Une requête ROLAP est en général exprimée comme suit.

- On calcule le joint de la table des faits et des relations dimensionnelles.
- On sélectionne des tuples en fonctions des données dimensionnelles.
- On groupe ces données suivant certaines dimensions.

– On calcule une valeur agrégée (le plus souvent une somme).

Ce type de requêtes est décrit d'avantage dans le tutorial joint à ce TP.

Une entreprise souhaite analyser ses données de vente. Chaque vente est définie par:

- – Un client décrit par un identifiant, un nom, une ville, une "région" (divers suivant les pays), et un pays)
- – Un produit, décrite par un identifiant, un nom, une catégorie, une sous-catégorie
- – Une date, décrite par un identifiant, une année, un trimestre, un mois, un jour. Tous sont des numéros.

Question 1 : Représenter, sur un diagramme conceptuel, les données sous la forme d'un schéma en étoile.

Question 2 : Les tables correspondantes sont implantées sous forme basique par le script "Exercice_2.sql". Observer ce script, modifiez le si nécessaire et l'exécuter pour générer la base.

Question 3 : Répondre aux questions suivantes par des requêtes :

- i. Donner la moyenne des ventes par année, région, catégorie, par année, région et par année, pour les années 2009 et 2010.
- ii. Donner la moyenne des ventes par année, selon les dimensions région et catégorie pour les années 2009 et 2010.
- iii. Quel est le produit le plus vendu par année et par catégorie (cf. RANK) ?
- iv. Pour chaque année, donner le total des ventes ainsi que le total des ventes par catégorie. On ne veut pas le résultat pour le total des années (utiliser la fonction GROUPING_ID).
- v. Donner tous les totaux de ventes par année selon la dimension catégorie, ainsi que selon la dimension cl_name (nom du client). Utilisez la commande GROUPING_SETS.
- vi. Quel est le meilleur mois de vente du produit "Sirop d'érable" pour chacune des années ?
- vii. Quelle est la quantité de produits vendus pour chaque catégorie, les 5 premiers jours de chaque mois de 2010 ?
- viii. Quelle est la répartition par tiers des catégories selon leurs quantités totales vendues en 2010 ?