

Concepteur : Didier DONSEZ
Durée : 4 heures 30
Indications :
Remarques : L'exercice 1 est à faire en binôme.

TD SUR LA MODELISATION DECISIONNELLE (PARTIE 1)

Exercice 1 : Prise en main de StarTraker

StarTraker est un logiciel fournit avec l'excellent ouvrage de Ralph Kimball (*Ralph Kimball, Entrepôts de Données, Ed. Intl Thomson Pub., 1997 et 2000, ISBN 2-84180-021-0*) qui permet de manipuler les bases de données décrites dans l'ouvrage et de générer des rapports (*report*) à partir de ces bases. L'utilisation de StarTraker est présentée en annexes D et E de l'ouvrage.

Installez StarTraker en copiant (sous Unix) le répertoire `~donsez/DW/STARTRAK` dans votre répertoire `~/DW`

Lancez StarTraker depuis Windows

Choisissez la base `Grocery.mdb`

Vous trouverez entre crochets les commandes de StarTraker utiles.

Suivez les instructions suivantes pour générer le premier rapport : « Montant des Ventes de 10/95 par rapport à 10/94 »

1. Déplacez le champ « Description » de la dimension « Product » vers le rapport
2. Déplacez le champ « Dollar_Sales » de la table de fait de ventes « Sales Fact » vers le rapport
3. Sélectionnez l'onglet « Groups » de la dimension Time et double cliquez sur le groupe « 11/95 »
4. Déplacez le champ « Dollar_Sales » de la table de fait de ventes « Sales Fact » vers le rapport
5. Déplacez le groupe « 11/94 » de la dimension Time vers la colonne « Dollar_Sales » du rapport
6. Déplacez le champ « Dollar_Sales » de la table de fait de ventes « Sales Fact » vers le rapport
7. Déplacez le groupe « 11/94 » de la dimension Time vers l'autre colonne « Dollar_Sales » du rapport
8. Sélectionnez la colonne Dollar_Sales (11/94) et sélectionnez le menu « Comparisons/Percent Difference of »
9. Cliquez « Run Report »

Générez (et sauvegardez) les rapports suivants :

1. Le rapport des montants des ventes (*dollars_sales*) par jour (*day*)
2. Le rapport des montants des ventes par mois (*month*)
3. Le rapport des marges (*margin*) et profits (*profits*) des ventes par marque (*brand*). Il s'agit d'attributs calculés
4. Le rapport des montants des ventes, du montant moyen des ventes et du montant maximal des ventes par mois (*month*)
5. Le rapport des montants des ventes et des nombres d'unités vendues par mois (*month*)
6. Le rapport des montants des ventes des 4 meilleurs jours et de 6 plus mauvais jours du mois de 10/95 [*Edit>Define Exceptions*]
7. Le rapport des montants des ventes et le rang (*rank*) par mois [*Sequences>Rank*]
8. Le rapport des montants des ventes et le cumul (*cume*) par mois [*Sequences>Cume*]
9. Le rapport des montants des ventes et le cumul mobile sur 3 mois par mois [*Sequences>Moving Sum*]
10. Le rapport des montants des ventes et des nombres d'unités vendues par mois (*month*) et par marque (*brand*)
11. Le rapport des montants des ventes par jour du mois de 10/95 et par ville (*city*) de la région

-
- Est (*eastern*).
12. Le rapport du montant des ventes de tous les produits, le ratio des boissons (*drink*), le ratio de l'alimentation (*foods*) et le ratio des fournitures (*supplies*) par mois (
 13. Le rapport précédent mais avec des lignes de rupture par année sur les mois [*Edit > Set Break Rows*]

Parcourez la dimension *Product* [*bouton Browse*]; Visualisez les descriptions [*double clic sur fields>description*], Sélectionnez les viandes, Créez un nouveau groupe « viandes » à partir des descriptions qui sont des viandes [*Group > Create Group*].

Générez les rapports suivants :

14. Le rapport des montants des ventes de viandes par mois
15. Générez le rapport des montants des ventes en 10/94 par marque. Connaissant la structure hiérarchique liant les attributs « marque », « conditionnement » et « sous-catégorie », utiliser le forage vers le bas pour détailler ce rapport par conditionnement, puis le forage vers le haut pour un détail par sous-catégorie.
16. Réalisez une étude donnant les 5 premiers produits les plus vendus en 10/94. [*Edit>create study*]
17. Utilisez l'étude précédente pour faire le suivi des ventes concernant ces mêmes produits en 10/95.

A chaque fois, visualisez la requête SQL qui permet de générer le rapport [*File > Show SQL*]. Exportez le premier rapport [*File > Copy Report*] vers MS Excel et produisez en des « Charts ».

Pensez à donner un titre à chaque rapport [*File>Report Title*] et aux colonnes [*Edit>Edit Heading*]

Vous pourrez tester des entrepôts supplémentaires en copiant les bases Access situées dans le répertoire `~donsez/DW/STARTRAK_EXTRADB` dans votre répertoire `~/DW/STARTRAK`

Exercice 2: Modélisation décisionnelle d'un Entrepôt de Données pour une Société de Gestion d'un Réseau Autoroutier

Une société de gestion d'un réseau autoroutier, la SFGRA, souhaite mettre en place un entrepôt de données dans le but de mieux suivre le comportement de sa clientèle (principalement les professionnels de la route : routier, VRP, ...) et offrir des services mieux ciblés à celle-ci ou à des catégories de celle-ci . Ces services sont par exemple des cartes de fidélisation, des cartes d'abonnés, des avantages chez les partenaires (stations essence, restaurants, ...) qui exploitent les aires sur le réseau.

Un client se comporte de la manière suivante :

"Il aborde l'autoroute depuis une entrée de l'autoroute, passe une ou plusieurs barrières de péage auxquelles il s'acquitte en payant par chèque, carte bancaire ou en monnaie. Le dernier paiement est considéré comme une sortie de l'autoroute. Tout au long du réseau autoroutier, le client peut s'arrêter pour prendre du carburant, faire des achats ou se restaurer. Les paiements peuvent être effectués par chèque, carte bancaire ou en monnaie."

Pour construire son entrepôt, la SFGRA dispose de plusieurs bases de données opérationnelles qui sont :

- la base PAIEMENT des paiements aux péages
- la base ABONNE des abonnés (possédant une carte de passage rapide)
- les bases des sociétés partenaires PARTENAIRE (stations essence, restaurants, ...) dans lesquelles les clients font des achats durant leur voyage. (Ces sociétés reçoivent en échange des données fournis à la SFGRA les résultats des études

menées par la SFGRA). Chaque partenaire utilise un système d'information historiquement différent de celui des autres partenaires donc une base de données opérationnelle modélisée différemment.

- la base CONDITION des conditions de circulation (météo, trafic, ...)

Vous supposerez que les schémas de ces trois bases suivantes sont les suivants :

Base PAIEMENT

Passage(Date, Heure, NumBarriere, NumTypePaiement, Montant, NumeroCompteBancaire*)

* peut être NULL si le paiement est en monnaie

NumBarriere référence Barriere

NumTypePaiement référence TypePaiement

TypePaiement(NumTypePaiement, DescrTypePaiement)

Les valeurs de NumTypePaiement sont 0 pour du liquide ou 1 pour un cheque ou un carte bancaire

Barriere(NumBarriere, LieuBarriere)

Tronçon(NumBarriereDepart, NumBarriereArrivee, Distance)

NumBarriereDepart référence Barriere

NumBarriereArrivee référence Barriere

Aire(NumBarriereDepart, NumBarriereArrivee, DistanceDepuisDepart, LieuAire)

(NumBarriereDepart, NumBarriereArrivee) référence Tronçon

Base ABONNE

ClientAbonne(Nom, Prenom, Adresse, NumImmatriculation, NumeroCompteBancaire)

Base du partenaire X PARTENAIREX

Achat(Date, Heure, NumAire, NTypePaiement, NtypeAchat, Montant, NumeroCompteBancaire, NumCarteFidelite*)

* peut être NULL si le client ne donne pas sa carte de fidelité au moment du paiement

TPaiement(NTypePaiement, DescrTypePaiement)

Les valeurs de NTypePaiement sont "MON" pour du liquide, "CHE" pour un cheque ou "CAR" pour une carte bancaire

TAchat(NTypeAchat, DescrTypeAchat)

Les valeurs de NTypeAchat sont "ESS" pour du carburant, "RES" pour de la restauration ou "XTR" pour autre chose.

CarteFidelite(NumCarteFidelite, Nom, Prenom, Adresse)

Base CONDITION

Meteo(Date, Heure, Region, ConditionEffective, ConditionPrévueMoins12H, ConditionPrévueMoins24H, ConditionPrévueMoins48H)

Trafic(Date, Heure, NumBarriereDepart, NumBarriereArrivee, EtatTrafic, ComptageVehicule, TypeIncident)

(NumBarriereDepart, NumBarriereArrivee) référence Tronçon

Les bases de données enregistrent des faits de paiement au niveau de la transaction.

Le premier entrepôt de données est centré sur les dépenses au niveau des parcours (entrée sur l'autoroute par un premier paiement, sortie considérée au moment du dernier paiement). On cherche à étudier les dépenses, les moyens de paiement, le temps de parcours et le

kilométrage parcouru.

Le second entrepôt de données est centré sur les dépenses au niveau des tronçons.

On ne chargera dans l'entrepôt que les voyages identifiables (c'est à dire pour lesquels les paiements ont été réalisés par chèque ou par carte bancaire : on utilise le numéro de compte bancaire présent sur le chèque et sur la carte bancaire)

Conception

Ex1: Rappelez le principe de chargement de l'entrepôt de données en commentant de 10 lignes un schéma.

Ex2: Donnez le schéma (en étoile) de l'entrepôt de données centré dépense au niveau des parcours (Table de Fait / Tables de Dimension, Précisez l'additivité des faits, s'il y a des dimensions dégénérés, des dimensions douteuses, ...)

Ex3: Donnez le schéma (en étoile) de l'entrepôt de données centré dépense au niveau des tronçons en réutilisant les dimensions de l'entrepôt précédent (Table de Fait / Tables de Dimension, Précisez l'additivité des faits)

Ex4: Donnez la taille (en Octets) de stockage des différentes tables des 2 entrepôts.

Nombre de Clients	10 000 000
Nombre de Voyages par jour	200 000
Nombre de Paiement par jour	800 000
Nombre moyen de barrières franchies par voyage	4
Nombre de jours	2400

Rapports

Ex5: Donnez la requête SQL qui donne le résumé mensuel de l'activité du client.

Ex6: Donnez la requête SQL qui donne la vitesse moyenne des clients.

Ex7: Donnez la requête SQL (étendu à l'OLAP) qui donne la moyenne mouvante sur 3 mois par tronçon

Chargement

Ex7: Donnez les requêtes d'initialisation de l'entrepôt

Ex8: Donnez les tables supplémentaires nécessaires aux requêtes de chargement.

Ex9: Donnez les requêtes de chargement de l'entrepôt à partir des bases de données opérationnelles.

Ex10 (optionnel) : Ajoutez cet entrepôt à StarTraker en suivant les indications de l'annexe E de l'ouvrage de Kimball. Testez les rapports précédents

Configuration Matérielle

Ex11: A partir des résultats du benchmark TPC/H (http://www.tpc.org/tpch/results/tpch_results.xls), choisissez la configuration matérielle et

logicielle qui est la plus adaptée à votre infocentre (deux entrepôts) pour des performances minimum égales à 1300 QphH ? idem à 2600 QphH ? Quels sont vos critères de choix ?