

Concepteurs : Didier DONSEZ
Durée : 2 heures
Remarques : Calculatrice et tout document autorisé
Conseil : Lire le sujet jusqu'au bout.
Annexe : Résultats du benchmark TPC/H

Rendre les réponses aux 2 problèmes sur des copies séparées

Modélisation décisionnelle des actions de téléspectateurs interactifs

L'opérateur BreeFox¹ propose à ses abonnés un boîtier routeur ADSL qui groupe un accès à IP, le téléphone et un bouquet de chaînes de télévision interactives. Nous ne nous intéresserons qu'à cette dernière fonction dans notre entrepôt de données.

Les chaînes de télévision proposées par BreeFox dans son bouquet sont des chaînes nationales et des chaînes à péage. Seulement, les interruptions publicitaires (des chaînes nationales et à péage) peuvent être personnalisées en fonction du profil du foyer (ou des adultes du foyer quand le contrôle parental² est déverrouillé). Le téléspectateur peut réagir (interactivement) de plusieurs manières aux émissions et aux publicités qu'il reçoit au moyen de sa télécommande (ie *zapette*):

- 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 (la durée de visualisation est importante).
- Pendant une publicité, il peut demander plus d'information sur le produit présenté (la durée de visualisation est importante) puis revenir à l'émission en cours .

¹ Toute ressemblance avec une société existante est purement fortuite !

² Pour éviter que les enfants du foyer apprennent trop jeunes la Biologie !

L'entrepôt de données est centré sur les actions du téléspectateur. L'objectif de cet entrepôt 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 l'entrepôt est constitué des tables suivantes (les clés primaires sont soulignées)

Date(CléDate, Année, Mois, JourDeMois, JourDeSemaine, TrancheHoraire, Heure, Minute, DrapeauVacances, DrapeauManifestation)

Foyer(CléFoyer, NomAbonné, AnnéeNaissanceAbonné, Région, Département, District, Ville, Quartier, SituationFamille, RevenuFoyer, CatégorieSocioProfessionnel, SousCatégorieSocioProfessionnel, DomaineActivité, NombreAdulte, NombreEnfant)

Emission(CléEmission, Chaîne, DateDébut, DuréeSeconde, TypeEmission, Catégorie, Annonceur)
TypeEmission= "Programme", "AnnoncePublicitaire"
Catégorie= "Météo", "Journal", "Variété", "Jeu", "JeuAvecPognion", "Film", "Foot", ...
Annonceur=seulement pour les annonces publicitaires

Action(CléDate, CléFoyer, CléEmission, TypeAction, DuréeAction, DuréeRestante, DrapeauCtrlParentalDéverrouillé)

- TypeAction= « Zap Autre Chaîne », « Eteindre », « Allumer », « Zap Pub Suivante », « Demande Information », « Passif » (je pas d'action sur la zappette)
- **DuréeAction** = Durée écoulée entre le début de l'émission et le début de l'action. **Quand** TypeAction est différent de « Demande Information »
- **DuréeAction** = Durée écoulée entre le début de l'émission et la fin du temps supplémentaire d'information avant que le spectateur regarde avant de retourner à son émission, **Quand** TypeAction est égal de « Demande Information »
- DuréeRestante = Durée restante entre l'action et la fin de l'émission (=0 si TypeAction= « Passif »)
- DrapeauCtrlParentalDéverrouillé indique le contrôle parental a été déverrouillé

Rétro-Conception

Q1: Quelle est la table de fait dans cet entrepôt ? Justifiez en 2 lignes !

Q2: Que pensez vous de l'attribut TypeAction de Action ?

Q3: A votre avis, il y a t'il des dimensions douteuses dans cet entrepôt ? Rappelez la définition et justifiez en 3 lignes

Q4: Donnez les nouvelles tables si on décide de diminuer la taille de la table Foyer par une mini-dimension démographique

Dimensionnement

Q5: Donnez le nombre de faits présents dans la table de fait.

Nombre de foyers abonnés	3 Millions
Nombre de actions par foyer et par heure	10
Un foyer regarde la télévision 320 jours par an, 5 heures par jour	
Nombre de tranche horaire	24
Nombre d'années	3
Taille des clés	4 octets
Taille des attributs numériques	4 octets
Taille des attributs discrets (comme les types !)	1 octet
Taille des attributs booléens (comme les drapeaux !)	1 octet

Donnez la taille d'un enregistrement de la table de fait ?

Donnez la taille (en Octets) de stockage de la table de fait.

Configuration Matérielle

Q6: A partir des résultats du benchmark TPC/H (http://www.tpc.org/tpch/results/tpch_results.xls) donné en annexe, choisissez la configuration matérielle et logicielle (complète) qui est la plus adaptée à votre infocentre pour une performance minimale de 25000 QphH ? Quels sont vos critères de choix ?

Remarque : vous négligerez la taille des tables de dimensions.

Rapports

Q7: Donnez la requête SQL qui donne le temps cumulé pour chaque type d'émission et pour chaque tranche horaire

Q8: Donnez la requête SQL qui donne le top 10 des type d'émission les plus regardées en temps cumulé

Q9: Donnez le rapport mensuel de 'progression' du nombre de demande d'information (ie TypeAction =« Demande Information ») et de la durée associée regardée

Q10: Donnez le rapport précédent mais avec une moyenne glissante avec les 2 mois précédents.

Conception physique

Q11 : Rappelez le principe d'index B-Tree et celui d'un index BitMap ? Lequel est il généralement mieux adapté aux entrepôts de données ?(en 10 lignes)

Q12 : Pourquoi est il intéressant d'avoir des enregistrements de taille fixe pour la table de fait (justifiez en 5 lignes)

Q13 : Est il judicieux d'utiliser les index bitmap sur les fichiers (stockage physique des tables) dont les enregistrements ont un taille variable (justifiez en 5 lignes)

Q14 : Pourquoi est il intéressant de partitionner la table de fait sur des disques différents ? (2 raisons à justifier en 8 lignes)