

POLYTECH MONTPELLIER

IG4

TP1 - BD DECISIONNELLES SOUS ORACLE

Compte-rendu de TP à rendre avant le 2 novembre 8h

On se propose d'analyser les habitudes d'emprunts (nombre d'emprunts) des adhérents de la bibliothèque en fonction de la ville qu'ils habitent, de leur âge, de la catégorie, et de l'état des exemplaires.

N.B. Vous pourrez vous appuyer sur la doc oracle. Les tables et lignes peuvent être récupérés sur www.lirmm.fr/~laurent/POLYTECH

1. Sur quel schéma vous proposez-vous de faire reposer cette analyse très simple ? De quel type de schéma s'agit-il ?
2. Comparaison de l'utilisation de GROUP BY, GROUP BY CUBE, GROUP BY ROLLUP : créez la requête permettant d'afficher les données à analyser
 - (a) avec la clause GROUP BY
 - (b) avec la clause GROUP BY ROLLUP
 - (c) avec la clause GROUP BY CUBE
 - (d) avec la clause GROUP BY GROUPING SETS
 - (e) quelles différences font apparaître les résultats ?
Utilisez la fonction DECODE pour afficher les labels 'Toutes villes confondues' et 'Tous états confondus' lorsque les résultats sont liés à l'agrégation sur ces attributs.
1. Quelle est la densité de votre "cube" ?
2. Fonctions d'analyse :
 - (a) Utilisez la fonction RANK pour ordonner les cellules du cube
 - (b) Utilisez la fonction RATIO_TO_REPORT pour donner, pour chaque abonné, la proportion d'emprunts qu'il a réalisés par rapport à la totalité des emprunts pour sa ville.
 - (c) Affichez le nombre d'emprunts état par état en indiquant également le nombre d'emprunts cumulés à chaque ligne (sum(count(x)) pour vous aider).
 - (d) Quelle est, pour chaque catégorie, le nombre maximal d'emprunts.
3. TOP-N QUERIES :
 - (a) Utilisez la fonction RANK pour rechercher les 2 comptages les plus importants (fonction TOP)
 - (b) même question pour les deux comptages les moins importants (BOTTOM)
 - (c) même question mais pour ordonner les nombres d'emprunts catégorie par catégorie.

4. Oracle permet de créer des vues matérialisées :
 - (a) A quoi cela peut-il servir ?
 - (b) Créez une vue matérialisée pour stocker les informations nécessaires à notre analyse.
 - (c) Vous pouvez visualiser les vues créées a bien été créée en consultant la table `user_mviews`. En particulier, l'attribut `STALENESS` permet de savoir si la vue est à jour ou non.
 - (d) Comment gérer les rafraîchissements pour qu'ils s'exécutent une fois par jour ?
 - (e) Insérez un emprunt dans la table *emprunts* et réaffichez le contenu de la vue matérialisée.
 - (f) La vue est-elle à jour ? Vous pouvez le confirmer ou l'infirmier à l'aide d'une requête du type

```
select staleness from user_mviews where mview_name = 'mavue';
```
 - (g) Forcez le rafraîchissement à l'aide de `dbms_mview.refresh`
 - (h) Les modifications ont-elles été prises en compte ? Affichez le contenu de la vue et les méta-données stockées par Oracle (`staleness` dans `user_mviews`).
5. Slow Changing Dimensions
 - (a) Rappelez ce que signifie "Slow Changing Dimensions"
 - (b) Donnez un exemple dans le cadre des données bibliothèque
 - (c) Quelles sont les différentes façons de gérer ce phénomène ? Quels seraient les biais possibles sur les données bibliothèque selon le choix de gestion ?
6. ETL
 - (a) Qu'est-ce qu'un ETL ?
 - (b) Quelles transformations permettent-ils de réaliser ?
 - (c) Quels sont les outils du marché ?