

IS2A3 – BDA – T.P. 3a

© Polytech Lille

Résumé

Ce TP aborde les index sous PostgreSQL.

N'hésitez pas à consulter la documentation en ligne sur le site de PostgreSQL : <https://www.postgresql.org/docs/13/index.html>

Mise en place

- Créer une base de données postgres qui aura pour nom `bda_tp3a_votreNomLogin`.

```
$ createdb -U votreNomLogin 1 bda_tp3a_votreNomLogin
```

Rappel, pour exploiter le serveur postgres de l'école, positionnez la variable `PGHOST` :

```
export PGHOST=serveur-etu.polytech-lille.fr
```

- Se connecter à votre nouvelle base de données :

```
$ psql -U votreNomLogin -d bda_tp3a_votreNomLogin
```

Rappel, le mot de passe de base est "postgres".

- Vous écrirez toutes vos commandes SQL dans un fichier et vous importez votre fichier SQL dans la base de données :

```
# \i votreFichierSQL
```

I Base de données issue de la MEL

- I.1. Lors de ce TP, nous allons utiliser comme base de données, des données qui ont été générées par les différentes stations de Vlille durant le mois de janvier 2022. La création des différentes tables ainsi que le peuplement se trouve dans le fichier `init_bd_vlille_220101.sql` que vous trouverez sur Moodle.
- I.2. Importez ce fichier dans la base de données que vous avez créée pour le TP.
Attention : l'importation peut être un peu longue. N'importez qu'une seule fois la base de données.
- I.3. Combien de stations de Vlille existait-il en janvier 2022 ?
- I.4. Combien de données avez-vous pour chaque station ?

II Comprendre avec `explain analyse`

Pour chacune des requêtes suivantes, préfixez la requête avec la commande `explain analyse`

II.1. `select * from compteur;`

II.2. `select count(*) from compteur;`

II.3. `select id, count(*) from compteur group by id;`

Est-ce que les requêtes sont toutes effectuées de la même façon ? Expliquez

1. Cette option n'est pas obligatoire quand votre nom de login UNIX est le même que votre nom de login PostgreSQL.

III Construction d'index

- III.1. Faites une requête pour connaître les différentes valeurs que peuvent prendre la colonne `service` de la table `compteur` et leurs nombre d'occurrences.
- III.2. Pour chacune des différentes valeurs possibles de `service`, regardez le temps effectué pour les requêtes où on affiche les `id` des stations en fonction de ces différentes valeurs possibles de `service` (utilisez `explain analyse`). Refaîtes plusieurs fois la même requête pour voir si le temps est toujours identique.
- III.3. Construisez un index `index_service` sur la colonne `service` de la table `compteur`.
- III.4. Refaîtes les commandes de III.2 et comparez (méthodes, temps, ...).
- III.5. Supprimez l'index `index_service`.

IV Index composite

- IV.1. Montrez les `id` des stations et le `timestamp` heure pour chaque compteur qui a exactement 3 places de vélo et au moins 10 vélos disponibles.
- IV.2. Calculez le temps de cette requête :
 - (a) sans index
 - (b) avec seulement un index sur le nombre de vélos
 - (c) avec seulement un index sur le nombre de places
 - (d) avec seulement deux index, un sur le nombre de vélos et un sur le nombre de places
 - (e) avec un index composite sur le nombre de vélos et le nombre de places.Qu'en concluez-vous ?

V Index partiel

- V.1. Montrez les `id` des stations et le `timestamp` heure pour chaque compteur qui a au moins 25 places de vélo disponibles et qui est *HORS SERVICE*.
- V.2. Calculez le temps de cette requête :
 - (a) sans index
 - (b) avec seulement un index sur le nombre de places
 - (c) avec seulement un index partiel sur le nombre de places où les stations sont *HORS SERVICE*
 - (d) avec le même index mais en échangeant l'ordre des conditions dans le `where` de la requêteQu'en concluez-vous ?

VI Index fonctionnel

- VI.1. On va s'intéresser ici à la capacité d'une station, c'est à dire à la somme du nombre de places disponibles et du nombre de vélos disponibles. Montrez les `id` des stations et le `timestamp` heure pour chaque compteur qui a au moins une capacité de 27.
- VI.2. Calculez le temps de cette requête :
 - (a) sans index
 - (b) avec seulement un index sur le nombre de vélos
 - (c) avec seulement un index sur le nombre de places
 - (d) avec seulement un index fonctionnel sur la capacité
 - (e) avec le même index mais en regardant une capacité plus grande que 10, que 20, que 30 et que 40Qu'en concluez-vous ?

VII Index couvrant

VII.1.

VII.2. Montrez les `id` des stations et le `timestamp` heure pour chaque compteur qui a au moins 30 places de vélo disponibles.

VII.3. Calculez le temps de cette requête :

- (a) sans index
 - (b) avec seulement un index sur le nombre de places
 - (c) avec seulement un index couvrant
- Qu'en concluez-vous ?

VIII Optimisation

Proposez pour chacune des questions suivantes, la requête appropriée et un ou plusieurs index qui vous permettront d'optimiser cette requête :

VIII.1. Montrez l'ensemble des lignes des compteurs entre 08h00 et 10h00.

VIII.2. Montrez l'ensemble des lignes des compteurs entre 08h00 et 10h00 des stations sans TPE.

VIII.3. Montrez le numéro des stations où le minimum entre le nombre de places de vélos disponibles et le nombre de vélos disponibles est supérieur à 20 au moins une fois sur toute la période.

Aide : Utiliser la fonction `least` pour calculer le minimum.

IX Pour ceux qui veulent un vrai défi ...

Vous travaillez actuellement à la métropole Lilloise à la gestion des stations Vville.

Un utilisatrice vous envoie le message suivant :

Je fais souvent le trajet en Vville de l'arrêt Canon d'or à l'arrêt Université catholique le matin vers 8h et le retour le soir vers 17h. Il arrive souvent qu'il n'y ait plus de vélo disponible au départ ou il n'y a plus de places à l'arrivée.

Suite à ce message, votre manager vient vous voir et propose 3 solutions :

- Proposez aux usagers une station de repli où prendre un vélo proche ou en laisser un.
- Proposez aux usagers un horaire moyen où le dernier vélo peut être pris et où le dernier vélo peut être posé avant l'heure indiquée.
- Proposez une rerépartition des vélos dans les stations : prendre des vélos dans une station pleine ou presque pleine et en mettre dans une station vide ou presque vide.

Proposez une procédure stockée pour chacune des solutions proposées.