

Requêtes récursives en SQL

1 Jeu de données

Nous travaillerons avec les données fournies par le Syndicat des Transports en Île-de-France (STIF), au format GTFS, que vous avez modélisées et chargées lors du TP3. Si ce n'est pas déjà fait, suivez les étapes indiquées dans la Section 2 du TP3 pour charger ces données (et relisez cette section vous vous rappeler la structure de ces données).

IMPORTANT : Dans la recherche de trajets, on considérera possible un changement de transport en commun entre deux arrêts quelconques qui se trouvent dans la même zone (c'est-à-dire deux arrêts qui ont la même station parent).

2 Requêtes sur la base

Pour chacune des questions suivantes, proposer une requête SQL qui renvoie les informations demandées ; ne pas hésiter à construire les requêtes petit à petit pour vérifier qu'elles renvoient des informations cohérentes.

Il peut être utile (voire nécessaire) d'ajouter des indexes sur les tables pour rendre l'exécution des requêtes plus efficace.

Pour vous aider, il est indiqué après chaque question le nombre de résultats que cette requête est censée renvoyer.

Pour désactiver la pagination du résultat dans le client en ligne de commande de PostgreSQL utiliser la commande `\pset pager off` au début de votre session `psql`. Vous pouvez aussi utiliser la commande `\timing` afin de comparer les temps d'exécution avec et sans index. N'hésitez pas non plus à préfixer vos requêtes avec le mot clef **EXPLAIN** pour voir le plan d'exécution de votre requête.

1. Calculez

- les zones d'arrêt qui ne sont pas de `location_type = 1`
- les arrêts avec `location_type = 1` dont le `stop_id` ne commence pas par "StopArea"
- les arrêts dont le `stop_id` commence pas "StopArea" qui ne sont pas des zones d'arrêts

(0 résultats pour les trois requêtes)

(On en déduit qu'il y a trois façons alternatives de reconnaître les zones d'arrêts :

- les arrêts qui sont parent d'un autre arrêt ;
- les arrêts qui ont `location_type = 1` ;
- les arrêts dont le `stop_id` commence par "StopArea").

2. Quelles sont les zones d'arrêts présents dans la table `stop_times` ? (0 résultats)

3. Quels sont les `stop_id` des vrais arrêts (c'est-à-dire des arrêts qui ne représentent pas des zones d'arrêts) ? (44497 résultats)

4. Quels sont les `stop_id` des zones d'arrêts ? (18944 résultats)

5. Quels sont les arrêts de transport en commun qui se trouvent dans la zone de la "GARE DE BAGNEUX" ? Pour chaque arrêt, renvoyer son identifiant et son nom. Les résultats doivent être triés par nom. (9 résultats)

6. Quels sont les lignes de transport en commun (c-à-d les routes) qui ont un arrêt dans la zone de la "GARE DE BAGNEUX" ? Pour chaque ligne, renvoyer le nom long de la ligne. Les résultats doivent être triés. (5 résultats)
7. Quels sont les arrêts accessibles en prenant un transport en commun dans la zone de la "GARE DE BAGNEUX", sans changement ? Renvoyer à la fois le nom de l'arrêt et le nom long de la ligne de transport permettant d'y accéder. Les résultats doivent être triés par nom de ligne, puis par nom d'arrêt. Ne pas oublier que pour qu'un arrêt **s2** soit accessible depuis un arrêt **s1** par une ligne de transport en commun, **s2** doit suivre **s1** dans le trajet de cette ligne (numéro de séquence plus élevé). (152 résultats)
8. Refaire la requête précédente, en sélectionnant cette fois ci uniquement les arrêts accessibles en train (exclure bus, métro, tram etc). Se rappeler que le type de transport en commun est décrit par l'attribut **route_type** de la table **routes**. (46 résultats)
9. Quelles sont les zones d'arrêt accessibles en prenant au plus deux trains successifs (c-à-dire au plus un changement) depuis la zone de la "GARE DE BAGNEUX" ? Remarquer que les lignes empruntées doivent être uniquement des trains (on ne s'intéresse pas aux autres modes de transport). Remarquer également qu'un changement peut avoir lieu entre deux arrêts quelconques qui se trouvent dans la même zone (i.e. qui ont la même station parent). Pour cette requête vous pouvez utiliser l'information que la "GARE DE BAGNEUX" en tant que zone d'arrêt porte le **stop_id** "StopArea:8775868". Renvoyer pour chaque résultat le nom de la zone d'arrêt ainsi que le nombre de trains successifs qui permettent d'y accéder depuis la zone de la "GARE DE BAGNEUX" (0, 1 ou 2). Les résultats doivent être triés par nombre de trains, puis par nom de zone. (294 résultats)
10. Refaire la requête précédente, mais renvoyer cette fois-ci, pour chaque zone d'arrêt dans le résultat, son nom ainsi que le nombre minimum de trains successifs qui permettent d'y accéder. Les résultats doivent être triés par nombre minimum de trains, puis par nom de zone. (247 résultats)
11. Refaire la requête précédente mais renvoyer cette fois-ci, pour chaque zone d'arrêt dans le résultat, son nom ainsi que les séquences possibles de lignes de train utilisables pour y accéder (utilisant un nombre minimum de trains). On pourra représenter une séquence de lignes de train par la concaténation des noms longs de la ligne, séparés par des espaces. Attention à ne pas commencer une séquence par un espace. Les résultats doivent être triés par séquence de trains, ensuite par nom de gare. (266 résultats)
12. Refaire la requête précédente mais renvoyer cette fois-ci, pour chaque zone d'arrêt dans le résultat, la concaténation de toutes les séquences possibles de lignes de train utilisables pour y accéder (utilisant un nombre minimum de trains). Utiliser la fonction d'agrégation **array_agg** pour agréger les séquences possibles pour un même arrêt. Les résultats doivent être triés par nom de gare. (247 résultats)
13. Quels sont les zones d'arrêt accessibles depuis la zone de la GARE DE BAGNEUX en effectuant au plus 3 arrêts en tout (destination incluse, origine exclue) ? On peut utiliser uniquement des transports en commun sur rail, et les changements de ligne sont autorisés (un changement de ligne compte comme un arrêt). Par exemple on peut arriver à la zone de l'"Aéroport Orly Ouest" avec 3 arrêts parce qu'on peut se rendre de la "GARE DE BAGNEUX" à la zone de la "GARE D'ANTONY" en RER B avec un seul arrêt intermédiaire ("BOURG-LA-REINE"), et donc deux arrêts en tout ; ensuite dans la zone de la "GARE D'ANTONY" on peut prendre l'"ORLYVAL" qui est direct jusqu'à la zone de l'"Aéroport Orly Ouest". Renvoyer pour chaque zone d'arrêt dans le résultat le nombre minimum d'arrêts (0,1, 2 ou 3) depuis la "GARE DE BAGNEUX". Les résultats doivent être triés par nombre d'arrêts, puis par nom d'arrêt. (15 résultats)