
TP5 : Bases de données

Pour ce TP, il y a trois méthodes différentes pour accéder à la base de données :

1. se connecter à l'adresse

`http://www.semwebtech.org/sqlfrontend/`

mais le nombre de connexions simultanées est limité ;

2. utiliser le fichier sur le site `mondial.db` et utiliser un explorateur de base de données (comme SQLite-Browser par exemple) ;
3. télécharger le fichier `mondial.db` et simuler un curseur avec Python en utilisant l'entête présent sur le site des classes de PT/PT*.

C'est cette dernière méthode qui sera présentée dans le corrigé et en TP. Les résultats pourront être légèrement différents lorsqu'on utilise la base de données en ligne compte-tenu de sa mise à jour régulière. Vous téléchargerez les fichiers `mondial.db` `TP5_nom.py` qui donnent le cadre du TP. On prendra garde au chemin d'accès au fichier.

1 Présentation de la base de données

Nous allons travailler avec deux tables : `country` qui décrit les pays du monde et `city` qui décrit les villes. Les attributs de ces deux tables sont les suivants.

Pour la table `country` :

- `name` (chaîne de caractères) le nom du pays ;
- `code` (chaîne de caractères) le code du pays ;
- `capital` (chaîne de caractères) la capitale du pays ;
- `province` (chaîne de caractères) la province à laquelle appartient la capitale ;
- `area` (entier) la superficie totale du pays ;
- `population` (entier) la population du pays.

Pour la table `city` :

- `name` (chaîne de caractères) le nom de la ville ;
 - `country` (chaîne de caractères) le code du pays auquel la ville appartient ;
 - `province` (chaîne de caractères) la province à laquelle la ville appartient ;
 - `population` (entier) la population de la ville ;
 - `elevation` (flottant) l'altitude de la ville (par rapport au niveau de la mer) ;
 - `latitude` (flottant) la latitude géographique de la ville ;
 - `longitude` (flottant) la longitude géographique de la ville.
1. Afficher toutes les entrées de la table `country` puis celle de `city` (une requête pour chaque, l'une après l'autre).
 2. Afficher toutes les entrées de la table `city` qui concerne le pays France dont le code est `F`.
 3. Afficher toutes les entrées de la table `city` qui concerne le pays France et dont la population est d'au moins 150 milliers d'habitants. Le faire également avec une intersection de table (syntaxe : `SELECT * FROM INTERSECT SELECT ...`).
 4. Afficher toutes les entrées de la table `city` qui concerne le pays France et dont la population est d'au moins 150 milliers d'habitants, classées par ordre alphabétique du nom de la ville.
 5. Afficher toutes les entrées de la table `city` qui concerne le pays France et dont la population est d'au moins 150 milliers d'habitants, classées par nombre d'habitants dans l'ordre décroissant. (Rappel : `ASC` signifie ordre ascendant, et `DESC` ordre descendant).

2 Agrégation

On rappelle les principales fonctions d'agrégation : `COUNT`, `MAX`, `MIN`, `SUM`, `AVG` qui donnent respectivement le nombre, le maximum, le minimum, la somme et la moyenne.

- Donner le nombre d'entrées de la table `city`, de même que pour la table `country` (une requête à chaque fois).
- Quelle est la moyenne du nombre d'habitants par ville dans le pays France (code `F`) ?
- Donner le nombre d'habitants du pays France vivant dans les villes enregistrées dans la base de données `city`. On affichera le résultat en milliers d'habitants.
- Faire une requête donnant simultanément l'altitude maximale et l'altitude minimale des villes mondiales.
- La requête précédente a montré (normalement) que l'altitude d'une ville est comprise entre -28 et 4330 mètres. Quelle est le nombre d'altitudes atteintes dans l'intervalle $[-28, 4330]$? (Réponse : 929)
- Donner pour chaque pays le nombre d'habitants (c'est la somme du nombre d'habitants de chaque ville : on se contentera d'interroger la table `city` et on donnera simplement le code du pays et le nombre total d'habitants).
- Quelle est la liste des pays (on se contentera là aussi d'interroger la table `city` et de donner le code du pays) ayant au moins 500000 habitants ? Afficher le résultat par nombre d'habitants décroissant (vous devez trouver 144 résultats).

3 Jointures

On va maintenant s'intéresser à des résultats concernant les deux tables `city` et `country`. On rappelle que pour faire une jointure, on utilise `JOIN nom_autre_table ON attribut`. On précise le nom des attributs à garder dans une sélection par `nom_table.nom_attribut` (nécessaire si le nom de l'attribut figure dans les deux tables, facultatif sinon). Il est courant et pratique de procéder à un renommage de la table (avec `AS` ou simplement en juxtaposant le nouveau nom à l'ancien) pour écourter les requêtes.

- Écrire une requête permettant d'obtenir la capitale la moins peuplée du monde (réponse : Plymouth, Montserrat) et une autre pour la plus peuplée (réponse : Lilongwe, Malawi).
- Quelle est la population moyenne des villes nord-américaines (États-Unis et Canada réunis) ? (réponse : environ 340 174)
- En considérant que la population d'un pays est égale à la somme des populations des villes qu'il abrite, donner la liste des pays par population décroissante. Quel est le pays le plus peuplé ? (Réponse : la Chine)

4 Sous-requêtes et requêtes complexes

On rappelle que l'on peut utiliser le résultat d'une requête comme une table, et imbriquer ainsi des requêtes. Voici quelques cas particuliers avec des requêtes complexes :

- si le résultat d'une requête (`R`) ne comporte qu'une seule ligne et une colonne, on peut utiliser le résultat dans une autre requête : `SELECT ... FROM ... WHERE attribut=(R)` ;
 - si le résultat d'une requête (`R`) ne comporte qu'une seule colonne, on peut tester l'appartenance dans une autre requête : `SELECT ... FROM ... WHERE attribut IN (R)` ;
 - on peut tester si le résultat d'une requête fournit au moins un résultat avec `SELECT ... FROM ... WHERE EXISTS (R)`.
- Quel est le nom de la (ou les) ville(s) ayant l'altitude maximale ? (Réponse : 'Cerro de Pasco', 4330)
 - Quelle est le nombre de villes situées au dessus de l'altitude moyenne des villes mondiales ? On n'écrira qu'une seule requête. (Réponse : 786).
 - La fonction `ROUND()` permet d'arrondir à l'entier le plus proche. Quelle est le nombre de villes situées à l'altitude moyenne des villes mondiales, arrondie à l'entier le plus proche ?
 - En considérant que la population totale d'un pays est égale à la somme des populations des villes qu'il abrite, calculer la population moyenne d'un pays. (Réponse : environ 9025482)
 - Quel est la capitale du pays le moins peuplé ?
(Réponse : Adamstown, capitale de Pitcairn.)