TP 13 : Requêtes dans des bases de données

1 Ouverture de la base de données `communes_francaises.sqlite`

1.1 Présentation de la base

Commencez par télécharger la base de données sur mon site web. Nous allons travailler avec deux tables : `depts` qui décrit les départements de France métropolitaine et `communes` qui décrit les communes. Les attributs de ces deux tables sont les suivants.

Pour la table depts :

— id (un entier) la clé primaire de l'enregistrement ;
— code (une chaîne de caractères) le code administratif du département ;
— nom (une chaîne de caractères) le nom du département ;
— chef_lieu (un entier) l'identifiant du chef-lieu du département.

Pour la table communes :

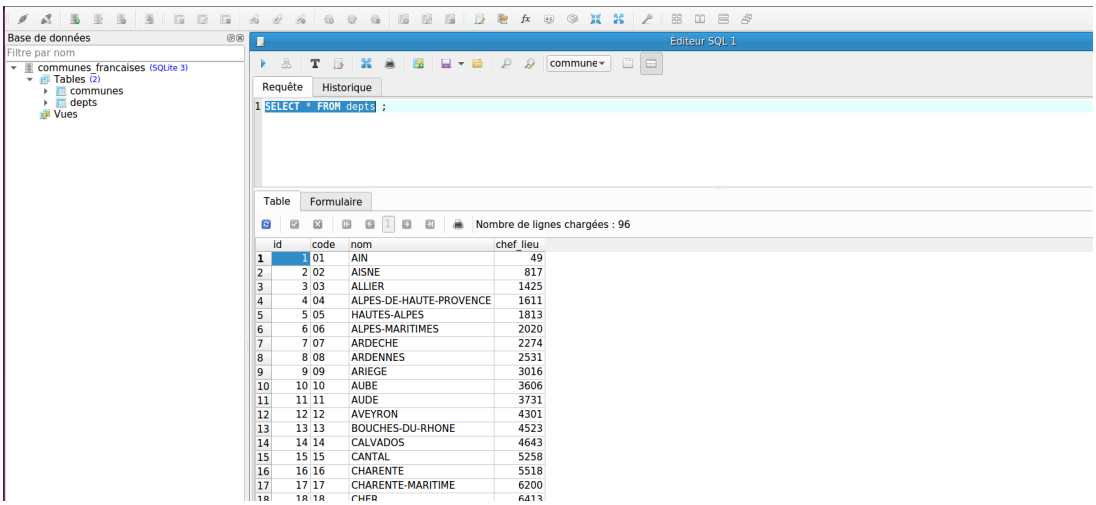
— id (un entier) la clé primaire de l'enregistrement ;
— code (une chaîne de caractères) le code administratif de la commune ;
— postal (une chaîne de caractères) le code postal de la commune ;
— nom (une chaîne de caractères) le nom de la commune ;
— superficie (un flottant) la superficie de la commune, en hectares ;
— altitude (un entier) l'altitude de la commune, en mètres ;
— population (un flottant) la population d'une commune, en millier d'habitants ;
— dep_id (un entier) l'identifiant du département où se trouve la commune.

Question 1. Avec cette description, quelles sont les clés primaires ? les clés étrangères ?

1.2 Ouverture de la base

Je vous donne deux choix pour travailler, que je détaille ci-dessous.

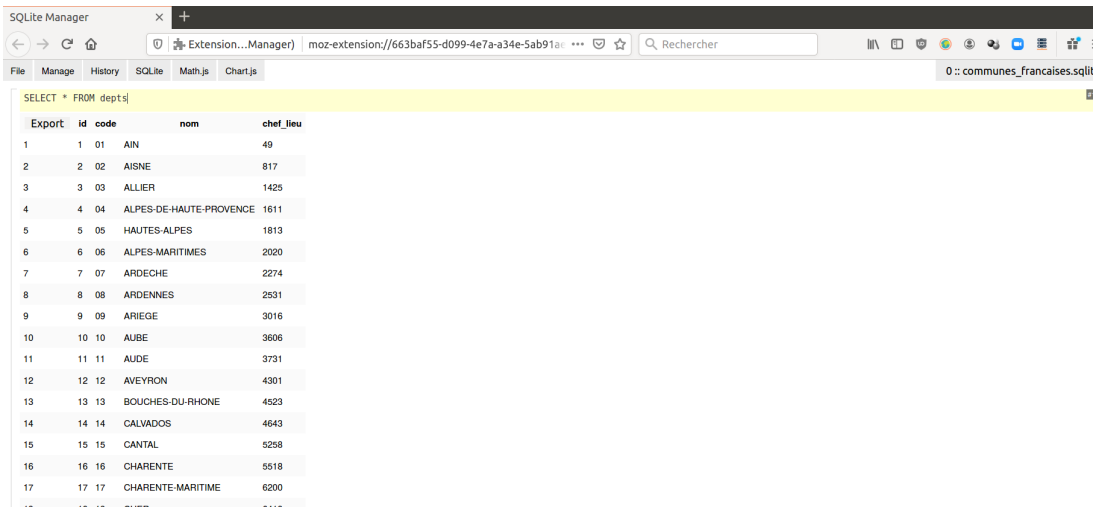
Méthode 1 : le logiciel `SQLite Studio`



Pour travailler avec `SQLite Studio`, le télécharger à l'adresse en bas de page [1]. Il s'agit d'une simple archive à dézipper, qui contient notamment un exécutable (SQLite Studio), cliquez pour lancer ! Ouvrez la base avec SQLite Studio (Database → Add a database, ou bien CTRL + O). Rajouter « Databases » dans l'onglet « View » pour avoir un aperçu de la base.

Pour écrire des requêtes SQL, ouvrir l'éditeur SQL (Tools → Open SQL Editor, ou ALT + E). Tapez la requête `SELECT * FROM depts` et exécutez avec la petite flèche bleue ou F9. Vous êtes prets pour la section suivante si tout fonctionne !

Méthode 2 : le plugin `SQLite Manager` pour le navigateur `Firefox` (qui nécessite naturellement d'avoir installé `Firefox`) .



Export	id	code	nom	chef_lieu
1	1	01	AIN	49
2	2	02	AISNE	817
3	3	03	ALLIER	1425
4	4	04	ALPES-DE-HAUTE-PROVENCE	1611
5	5	05	HAUTES-ALPES	1813
6	6	06	ALPES-MARITIMES	2020
7	7	07	ARDECHE	2274
8	8	08	ARDENNES	2531
9	9	09	ARIEGE	3016
10	10	10	AUBE	3606
11	11	11	AUDE	3731
12	12	12	AVEYRON	4301
13	13	13	BOUCHES-DU-RHONE	4523
14	14	14	CALVADOS	4643
15	15	15	CANTAL	5258
16	16	16	CHARENTE	5518
17	17	17	CHARENTE-MARITIME	6200

Rendez-vous sur le lien en bas de page [2] pour télécharger le plug-in Firefox SQLite Manager. Faites simplement File : Open a database pour ouvrir la base `communes_francaises.sqlite`. Entrez simplement la requête `SELECT * FROM depts` dans la barre de recherche en jaune claire et faites Entrée. Vous êtes prets pour la section suivante si cela fonctionne !

2 Où on commence vraiment le TP

2.1 Requêtes dans une seule table

On rappelle la syntaxe générale d'une requête de recherche :

```
SELECT attribut(s)
FROM table1
JOIN table2 ON ... JOIN table3 ON ... (* jointures *)
WHERE condition (* sélection *)
GROUP BY attribut(s) (* agrégation *)
HAVING condition (*séléction post-agrégation *)
ORDER BY attribut(s) (ASC | DESC) (* ordonnancement *)
LIMIT n (* limitation du nombre de résultats affichés à n *)
OFFSET p (* ignorer les p premiers résultats *)
```

Question 2. Afficher toutes les entrées de la table `communes` puis celle de `depts` (une requête pour chaque, l'une après l'autre).

Question 3. *Sélection*. Afficher toutes les entrées de la table `communes` qui concerne le département des Alpes maritimes, dont l'identifiant est 6.

Question 4. *Sélection multiples*. Afficher toutes les entrées de la table `communes` qui concerne le département des Alpes maritimes et dont la population est d'au moins 10 milliers d'habitants (rappel : la population est en milliers). Le faire également avec une intersection de tables (syntaxe : `SELECT * FROM ... INTERSECT SELECT ...`).

1. `https://sqlitestudio.pl/`
2. `https://addons.mozilla.org/en-US/firefox/addon/sqlite-manager-webext/`

Question 5. *Ordonnancement*. Afficher toutes les entrées de la table `communes` qui concerne le département des Alpes maritimes et dont la population est d'au moins 10 milliers d'habitants, classées par ordre alphabétique du nom de la commune.

Question 6. *Ordonnancement bis*. Afficher toutes les entrées de la table `communes` qui concerne le département des Alpes maritimes et dont la population est d'au moins 10 milliers d'habitants, classées par nombre d'habitants dans l'ordre décroissant. (Rappel : ASC signifie ordre ascendant, et DESC ordre descendant).

Question 7. *Limitation de l'affichage*. LIMIT permet de limiter l'affichage à un certain nombre de résultats. Afficher les données de la commune la plus peuplée du département en reprenant la requête précédente.

Question 8. *Ignorer les premiers résultats*. OFFSET n permet d'ignorer les n premiers résultats (et ne peut s'utiliser sans LIMIT). Donner la liste des communes des Alpes maritimes classées par superficie décroissante, de la 101ème à la 120ème seulement.

Question 9. *Utilisation de LIKE, hors programme*. On rappelle que l'on peut filtrer des motifs de chaînes de caractères avec `LIKE ...`, sachant que le joker `%` filtre toute chaîne de caractères. Donner la liste des communes dont le nom commence par "AB" (il y en a 49).

2.2 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.

Question 10. *Nombre*. Donner le nombre d'entrées de la table communes, de même que pour la table depts (une requête à chaque fois).

Question 11. *Moyenne*. Quelle est la moyenne du nombre d'habitants par commune dans le département des Alpes maritimes ?

Question 12. *Somme*. Donner le nombre d'habitants des Alpes-maritimes, on affichera le résultat en nombre d'habitants et pas en milliers.

Question 13. *Min et Max*. Faire une requête donnant simultanément l'altitude maximale et l'altitude minimale des communes françaises.

Question 14. *Utilisation de DISTINCT*. La requête précédente a montré (normalement) que l'altitude d'une commune française est comprise entre 0 et 2713 mètres. Quelle est le nombre d'altitudes atteintes dans l'intervalle $[\![0, 2713]\!]$? (Réponse : 1650)

Question 15. *Regroupement*. Pour indiquer quels groupes de résultats sont concernés par une agrégation, on utilise GROUP BY. Donner pour chaque département le nombre d'habitants (c'est la somme du nombre d'habitants de chaque commune. On se contentera d'interroger la table communes et on donnera simplement le `dep_id` et le nombre total d'habitants).

Question 16. *Sélection après une agrégation*. Pour sélectionner après une agrégation, on utilise HAVING. Notez aussi qu'on peut renommer un attribut avec AS. Quelle est la liste des départements (on se contentera là aussi d'interroger la table communes et de donner le `dep_id`) ayant au moins 500000 habitants ? Afficher le résultat par nombre d'habitants décroissant (vous devez trouver 50 résultats).

2.3 Jointures

On va maintenant s'intéresser à des résultats concernant les deux tables communes et depts. On rappelle que pour faire une jointure, on utilise `JOIN nom_autre_table ON condition`. On précise le nom des attributs à garder dans une selection 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.

Question 17. Donner la liste des dix communes les plus vastes de France, classées par ordre décroissant de taille, en précisant pour chacune d'elles le nom du département où elles se situent. (La plus vaste est Arles).

Question 18. Écrire une requête permettant d'obtenir le chef-lieu le moins peuplé de France (réponse : Privas) et une autre pour le plus peuplé (réponse : Toulouse [3]). Réfléchissez à la jointure que vous allez faire pour cette question, et faites une requête qui ne produit qu'un résultat [4].

Question 19. Quelle est la superficie moyenne (en hectares) des communes corses ? Rappelons que la Corse est constituée de deux départements de codes administratifs (ce sont des chaînes de caractères) 2A et 2B (Réponse : environ 2433).

Question 20. En considérant que la superficie d'un département est égal à la somme des superficies des communes qu'il abrite, donner la liste des départements par superficie décroissante. Quel est le département le plus vaste ? (Réponse : la Gironde)

2.4 Sous-requêtes, 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 (hors-programme normalement, mais ce n'est pas très clair...) :

- 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)`.

Question 21. Quel est le nom de la (ou les) commune(s) ayant l'altitude maximale ? Remarque : votre requête ne produira qu'un seul résultat.

Question 22. Quelle est le nombre de communes situées au dessus de l'altitude moyenne des communes françaises ? On n'écrira qu'une seule requête. (Réponse : 11825).

Question 23. ROUND permet d'arrondir à l'entier le plus proche. Quelle est le nombre de communes situées à l'altitude moyenne des communes françaises, arrondie à l'entier le plus proche ? (Réponse : 58).

Question 24. En considérant que la superficie d'un département est égal à la somme des superficies des communes qu'il abrite, calculer la superficie [5] moyenne (en km^2) d'un département français (Réponse : 5712 km^2 environ)

Question 25. Quel est le chef-lieu du (du, pas de !) département le moins peuplé de France ? (Réponse : Mende, chef-lieu de la Lozère.)

Il est possible de faire une jointure sur une table et une copie de cette table, mais dans ce cas le renommage est obligatoire. Lorsque cette table est en plus le résultat d'une requête, il est possible d'utiliser la syntaxe suivante : `WITH nom_table AS (R) SELECT ...`. La requête qui suit le `SELECT` peut utiliser le résultat de la requête (R) renommé en `nom_table`.

Question 26. En utilisant la syntaxe précédente, trouver toutes les communes des Alpes-maritimes (dep_id = 6) qui sont à la même altitude. La requête renverra des couples de communes. Il est nécessaire de faire une jointure d'une table avec elle même, renommage obligatoire ! Pensez à éliminer les doublons. (Réponse : il y a deux communes aux altitudes 64, 130, 164, 204, 242, 367, 955 et 1062 mètres, et trois à l'altitude 850 mètres).

Question 27. En utilisant EXISTS décrit plus haut, donner le nombre de communes des Alpes-maritimes ayant la même altitude qu'une autre. Remarque : en procédant par renommage, vous pouvez utiliser un attribut externe dans une requête interne, comme ceci : `SELECT ... nom AS n FROM ... (SELECT ... WHERE nom <> n ...) ...` Vous devez trouver 19, ce qui est cohérent avec la requête précédente.

3. Paris, Marseille et Lyon sont stockées par arrondissement, ce qui explique pourquoi Toulouse apparaît en premier...
4. Utiliser LIMIT. Attention, `SELECT c.nom, MAX(population)...` n'a aucun sens en aglèbre relationnelle et est à proscrire. Bien qu'elle fonctionne ici, cette requête n'est pas rigoureuse en SQL et pourra ne pas fonctionner avec certains logiciels de bases de données.
5. Rappel : un hectare correspond à 0.01 km^2