

TP Informatique 05

Instructions générales

Lancer l'application **DB Browser for SQLite**. Aller dans **Fichier**, puis **Ouvrir une base de données** et sélectionner la base de données **pays.sqlite**.

Commencer par regarder la structure de la base de données. Vous pouvez aussi parcourir la base de données pour avoir une vue d'ensemble de la table.

Aller dans **Exécuter le SQL** pour taper vos requêtes. Pour les lancer, sélectionner la requête SQL à lancer, puis cliquer sur l'icône de lancement (flèche bleue).

Sélection et projection

Dans ce TP, on utilisera uniquement la table **pays**. Pour éviter les problèmes de compatibilité, aucun accent ni tiret n'a été utilisé dans cette table.

1. Exécuter la commande suivante :

```
SELECT * FROM pays
```

Quels sont les champs ou attributs (c'est-à-dire les noms des colonnes) de la relation (aussi appelée table) **pays** ?

Pour obtenir plus d'informations sur les différents champs, en particulier pour connaître leur type, cliquer sur l'onglet **structure**. Que signifient **int**, **bigint**, et **varchar** ?

L'unité des champs **superficie**, **population** et **PIB** (c'est-à-dire produit intérieur brut) sont respectivement le km^2 , l'habitant et le dollar (américain).

Pour n'afficher que certains enregistrements ou multiplets (c'est-à-dire lignes) de la table, on utilise une **sélection** (ou **restriction**).

2. Comment obtenir toutes les informations disponibles sur la France ?

Pour n'afficher que certains champs ou attributs (c'est-à-dire colonnes) de la table, on effectue une **projection**. Tout l'intérêt réside naturellement dans la possibilité de combiner les deux opérations élémentaires que sont la sélection et la projection.

3. Afficher les noms des pays dont la population dépasse 200 millions d'habitants.
4. Afficher la liste des continents. On utilisera le mot clé **DISTINCT** pour éviter les répétitions.

On peut réaliser des opérations sur les champs d'un même enregistrement. Si deux champs sont numériques, on peut les multiplier, les additionner...

5. Afficher le nom et le PIB par habitant pour les pays dont la population est supérieure à 200 millions.
6. Afficher le produit intérieur brut de la France en milliard de dollars.

Pour réaliser des sélections plus riches, il est possible de combiner les conditions avec les mêmes opérations logiques (aussi appelées booléennes) qu'en python : **AND**, **OR**...

7. Afficher les noms des pays d'Europe dont le PIB dépasse celui de la France. On pourra copier-coller (et modifier légèrement) le résultat de la requête précédente.
8. Afficher les noms des pays américains.
9. Reprendre la question précédente en utilisant **UNION**.
10. Quel est le PIB par habitant de la Suisse ? Dans un deuxième temps, arrondir le résultat par défaut avec la fonction **round**, dont le nom est explicite.
11. Afficher le nom et la densité de population (en habitant par km²) de la Russie et des Etats-unis. Lequel de ces deux pays est le plus densément peuplé ?
12. Pour améliorer la lisibilité du résultat précédent, on peut utiliser le **renommage** qui permet de changer le nom d'une colonne. Réécrire la requête de la question précédente en renommant la colonne « densite ».
13. Afficher le nom et la population en millions d'habitants de l'Allemagne, de l'Italie et de la France. On rappelle qu'il faut répondre à la question à l'aide d'une unique requête.
14. On peut utiliser une notation plus concise pour répondre à la question précédente. La syntaxe **IN Liste** permet de tester si un élément appartient à une liste. La syntaxe des listes en SQL est (**element1, element2, ...**). Les éléments de la liste sont séparés par des virgules et encadrés par des parenthèses. Répondre à la question précédente en utilisant cette nouvelle syntaxe.
15. Afficher les pays dont la superficie est supérieure à trois millions de km². Les compter « à la main ». La suite proposera une solution plus efficace.

Il est possible de trier les résultats d'une requête en utilisant la commande **ORDER BY**. On peut ainsi préciser un champ sur lequel doit porter le tri. Par défaut le tri est croissant. Pour changer ce comportement, il faut rajouter le mot clé **DESC**.

16. Afficher les pays par PIB décroissant. Quel est le pays dont le PIB est le plus élevé ?
17. Utiliser **ORDER BY** pour trouver le pays dont le PIB par habitant est le plus élevé.

Fonctions d'agrégation

Les *fonctions d'agrégation* réalisent un traitement sur toutes les valeurs d'une même colonne. On se limitera aux fonctions d'agrégation suivantes : **SUM**, **AVG**, **COUNT**, **MIN** et **MAX**. Ces fonctions permettent par exemple de déterminer la valeur minimale, la valeur maximale ou bien encore la valeur moyenne d'un champ d'une table. La syntaxe générale la suivante :

```
SELECT fonction(champ) FROM table
```

Le cas de la fonction **COUNT** est particulier. Pour obtenir le nombre total de lignes d'une table, on peut utiliser la syntaxe

```
SELECT COUNT(*) FROM table
```

18. En se basant sur leur nom, prédire l'utilité de chacune des fonctions d'agrégation ci-dessus. Exécuter ensuite les commandes suivantes :

```
SELECT population FROM pays WHERE superficie > 3000000
```

puis

```
SELECT MIN(population) FROM pays WHERE superficie > 3000000
```

Le comportement de la fonction MIN est-il en accord avec la réponse à la question précédente ?

19. Afficher la population mondiale. Pour faciliter la lecture, afficher ensuite la population mondiale en milliards d'habitants. Noter que la base de données date un peu.
20. Combien de pays sont recensés dans cette base de données ?
21. Quelle est la population de l'Amérique du Sud exprimée en millions d'habitants ?
22. Afficher le produit intérieur brut de l'Afrique en milliards de dollars.
23. Quelle est la population du Bénélux ? Pour améliorer la lisibilité, on affichera le résultat en millions d'habitants en arrondissant par défaut.

La commande GROUP BY, dont le nom est explicite, permet de regrouper des enregistrements suivant certains critères. À titre d'exemple, la requête de gauche produit le résultat de droite.

```
SELECT continent,  
       round(SUM(population)/1E6)  
       AS 'pop. totale en millions'  
FROM pays GROUP BY continent
```

continent	pop. totale en millions
Afrique	1042
Amerique du Nord	545
Amerique du Sud	392
Asie	4133
Europe	685
Océanie	37

La commande GROUP BY est très souvent utilisée avec les fonctions d'agrégation.

24. Pour chaque continent, afficher le nom du continent et le nombre de pays du continent.
25. Pour chaque continent, afficher le nom du continent et le nombre de pays ayant une population supérieure à 10 millions d'habitants.
26. Sans utiliser ORDER BY, obtenir le PIB par habitant le plus élevé. Vérifier la compatibilité avec la réponse à la question 17.

BONUS

27. Combien de pays ont un nom commençant par la lettre 'Z' ? On pourra utiliser la relation d'ordre sur les chaînes de caractères (c'est-à-dire l'ordre lexicographique).
28. Combien de pays ont un nom commençant par la lettre 'H' ?
29. Exécuter la commande suivante

```
SELECT nom FROM pays WHERE nom LIKE 'H%'
```

Vérifier que le résultat est compatible avec la réponse précédente. Quel est le rôle du mot clé LIKE ? Proposer une traduction en français. Le caractère % permet de remplacer n'importe quelle chaîne de caractères. Le mécanisme utilisé ici est un exemple d'expression régulière.

30. Afficher tous les pays dont le nom contient "unis".
31. Un pays protège particulièrement bien ses données bancaires et financières. Trouver lequel.
32. Pour chaque continent, afficher :
 - le nom du continent
 - le PIB par habitant du continent