

Fonctions d'agrégation

1 Count

Les fonctions d'agrégation permettent d'effectuer des opérations avancées sur les solutions d'une requête comme : compter, sélectionner le maximum etc. Une des opérations les plus courantes est de compter. `COUNT(col)` permet de compter les résultats d'une requête. Pour compter les pays en Europe, on écrira :

```
SELECT COUNT(countrycode)
FROM world.country
WHERE continent='Europe';
```

Cette requête renvoie une table ayant une ligne et une colonne contenant le nombre de réponse à la requête. Admettons qu'on veuille compter les pays par continent. On serait tenté d'écrire :

```
SELECT continent ,COUNT(countrycode)
FROM world.country;
```

Cependant, cette requête ne fonctionnera pas en SQL. SQL est incapable de deviner comment regrouper les différentes lignes pour compter. On doit lui spécifier clairement cela avec la clause `GROUP BY` :

```
SELECT continent ,COUNT(countrycode)
FROM world.country
GROUP BY continent;
```

Cette requête regroupe les lignes de la table `country` par continent et pour chaque groupe compte le nombre de `countrycode` y apparaissant. Lorsque plusieurs lignes sont susceptibles d'avoir la même valeur, on peut compter seulement le nombre d'occurrence distinctes avec `COUNT(DISTINCT col)`.

1. Écrire une requête qui compte le nombre de langues parlées dans chaque pays.
2. Écrire une requête qui compte le nombre de langues parlées dans le monde.
3. Écrire une requête qui compte le nombre de langues officielles par pays.

2 Sum, Max, Min, Avg

Une autre fonction importante est la fonction `SUM(col)` qui effectue la somme des valeurs (numériques) d'une colonne :

```
SELECT SUM(population_country)
FROM world.country;
```

renvoie la population mondiale. On peut de même utiliser `GROUP BY` pour faire des paquets :

```
SELECT continent ,SUM(population_country)
FROM world.country
GROUP BY continent;
```

renvoie la population de chaque continent. On peut même faire des opérations sur la colonne à l'intérieur de `SUM`. Par exemple : `SUM(percentage/100)`.

1. Écrire une requête qui renvoie la surface de chaque région.
2. Écrire une requête qui compte le nombre de francophones dans le monde.

On peut utiliser de la même façon la fonction `MIN` (resp. `MAX`) qui renvoie la plus petite (resp. grande) valeur ou `AVG` qui renvoie la moyenne.

1. Combien de personnes vivent dans une capitale européenne ?
2. Quel est la capitale européenne la moins peuplée ?
3. Quelle est la langue la plus parlée dans le monde ?

3 Having

Parfois, on veut filtrer les requêtes en fonction du résultat d'une fonction d'agrégation. Par exemple, pour connaître les langues officielles dans plus de 10 pays, on serait tenté d'écrire :

```
SELECT language FROM world.countrylanguage  
WHERE COUNT(countrycode) > 10 and isofficial GROUP BY language;
```

Cependant, cela ne fonctionne pas. **WHERE** applique une condition sur chaque ligne de la table pour les filtrer, par exemple, garder seulement les langues officielles. Ici, on veut *ensuite* sélectionner les lignes après avoir regroupé par langue et compté. On utilisera alors **HAVING**, après le **GROUP BY** :

```
SELECT language FROM world.countrylanguage WHERE isofficial  
GROUP BY language HAVING COUNT(countrycode) > 10;
```

4 TP

1. Écrire une requête qui renvoie le nombre de pays par régime.
2. Écrire une requête calculant le nombre de personnes vivant dans des villes de plus d'un million d'habitant.
3. Écrire une requête qui calcule le nombre de personnes vivant dans une ville qui n'est pas listées dans la table `city`. (Indice : comparer la population du pays avec la somme sur les villes).
4. Écrire une requête qui compte le nombre moyen de langue parlée par pays dans chaque région.
5. Écrire une requête qui donne la liste des pays ayant deux langues officielles parlées par plus de la moitié de la population.

Devoir maison

Écrire une fonction `biggest_city(countryname character varying(49))` qui étant donné le nom d'un pays, renvoie le nom de la ville la plus peuplée de ce pays.