

Cours de bases de données,
<http://sql.bdpedia.fr>

SQL: agrégation

SQL, agrégation

Cette session présente les agrégats en SQL. Elle consiste à regrouper des lignes et à appliquer à chaque groupe une fonction d'agrégation.

Contenu :

- La clause `group by`
- Fonctions d'agrégation
- La clause `having`

Ces diapositives correspondent au support en ligne disponible sur le site <http://sql.bdpedia.fr/>

Principe général

On définit des **groupes** de lignes ayant en commun une ou plusieurs valeurs.

On ramène chaque groupe à une seule valeur en appliquant une **fonction d'agrégation**.

Cas le plus simple : un seul groupe, obtenu par un bloc `select from where`.

```
select count(*) as nbPersonnes, count(prénom) as nbPrénoms,  
       count(nom) as nbNoms  
from Personne
```

nbPersonnes	nbPrénoms	nbNoms
7	6	7

Le group by

La clause `group by att1, ..., attn` **partitionne** le résultat d'un bloc `select from where` en fonction des `att1, ..., attn`

Chaque **groupe** contient les lignes qui partagent les mêmes valeurs pour `att1, ..., attn`.

```
select idAppart, sum(quotePart) as totalQP
from Propriétaire
group by idAppart
```

Procède en deux étapes : d'abord on groupe, puis on agrège.

Décomposons : l'étape de regroupement

On obtient une structure intermédiaire, avec autant de lignes que de valeurs distinctes pour les attributs de regroupement (ici, idAppart).

idAppart	Groupe (idPersonne, idAppart, quotePart)
100	[(1, 100, 33), (5, 100, 67)]
101	[(1, 101, 100)]
102	[(5, 102, 100)]
103	[(2, 103, 100)]
104	[(2, 104, 100)]
201	[(5, 201, 100)]
202	[(1, 202, 100)]

Ce n'est pas une table en première forme normale.

L'étape d'agrégation

La fonction d'agrégation ramène un groupe à une valeur

idAppart	SUM(quotePart)
100	SUM (33, 67) = 100
101	SUM (100) = 100
102	SUM (100) = 100
103	SUM (100) = 100
104	SUM (100) = 100
201	SUM (100) = 100
202	SUM (100) = 100

Cette fois c'est une table en première forme normale.

La clause having

Exprime un critère de sélection sur le résultats de la fonction d'agrégation.

Bien distinguer de la clause where qui s'applique aux nuplets

```
select idAppart, count(*) as nbProprios
from Propriétaire
group by idAppart
having count(*) >= 2
```

idAppart	nbProprios
100	2

À retenir

Agrégats = extension de SQL.

- S'applique au **résultat** d'une requête standard
- Partitionne en groupes de nuplets partageant les mêmes valeurs de regroupement
- Réduit chaque groupe à une valeur grâce à une fonction d'agrégation
- On peut filtrer les groupes obtenus avec **having**

Attention aux valeurs à `null` et aux doublons.