

Bases de Données Relationnelles

Fonctions d'agrégations/Requêtes statistiques

L2

Rafael Angarita
Maitre de Conférences
rangarit@parisnanterre.fr

Structured Query Language (SQL)

LDD

(Définition des Données)

- Création table → CREATE TABLE
- Suppression Table → DROP TABLE
- Modification table → ALTER TABLE

LMD

(Manipulation des Données)

- Insertion données INSERT INTO
- Suppression données → DELETE FROM
- Modification donnée → UPDATE ...SET

LID

(Interrogation des Données)

- Interrogation donnée →
**SELECT ... FROM... WHERE....GROUP BY
....HAVING...**

Données statistiques

Une définition minimale :
Données dénombrées et dénombrables !

Exemples :

- Le **nombre d'étudiants** de chaque formation.
- La **moyenne des notes** du module de BD.
- Le **produit le moins acheté** dans un supermarché.
- ...

Données statistiques et BD

- Dans une BD, l'information statistique n'est pas **directement** représentée ...
- On aura, sinon, de la **redondance et des incohérences!**

- Sachant que les requêtes SQL de base ne permettent qu'une extraction d'information simple...,
- la question est, donc,
comment extraire (visualiser) les données statistiques ?



Les requêtes statistiques

Exemple introductif

N° auteur	Genre	Nom auteur	Prénom auteur	A. naissance	A. décès
1	M	HUGO	Victor	1802	1885
2	F	SAND	George	1804	1876
3	M	DUMAS	Alexandre	1802	1870
4	F	YOURCENAR	Marguerite	1903	1987
5	F	De STAËL	Germaine	1766	1817
6	M	RACINE	Jean	1639	1699
7	M	VERLAINE	Paul	1844	1896

Combien y a-t-il d'auteurs dans la BD ?

Exemple introductif

N° auteur	Genre	Nom auteur	Prénom auteur	A. naissance	A. décès
1	M	HUGO	Victor	1802	1885
2	F	SAND	George	1804	1876
3	M	DUMAS	Alexandre	1802	1870
4	F	YOURCENAR	Marguerite	1903	1987
5	F	De STAËL	Germaine	1766	1817
6	M	RACINE	Jean	1639	1699
7	M	VERLAINE	Paul	1844	1896

Combien y a-t-il d'auteurs par genre dans la BD ?

Requêtes statistiques : principe

- Le calcul se fait sur les données d'un groupe de plusieurs enregistrements d'une table T_0 .
- Ces enregistrements doivent avoir **un point commun : des champs qui ont la même valeur**.
- Ils sont **agregés** pour n'en **faire plus qu'un**, qui comprend :
 - Une partie des champs communs.
 - Les résultats des calculs.
- Les enregistrements ainsi décrits forment **une nouvelle table T_1** (le résultat de la requête).

Requêtes statistiques : exemple

T0

Genre
M
F
M
F
F
M
M



T1

Genre	NB
M	4
F	3

Combien y a-t-il d'auteurs par genre dans la BD ?

- Deux groupes d'enregistrements : dans chaque groupe la valeur du champ genre est identique.
- Agréger les enregistrements de chaque groupe (pour n'en faire qu'un) **avec le résultat des calculs.**

Requêtes statistiques : SQL

- **Plusieurs types de requêtes :**

- ☛ d'interrogation.
- ☛ de manipulation : insert, update, delete...
- ☛ de définition : create, alter, drop...

Les requêtes statistiques

SELECT [DISTINCT] liste de champs , **fonctions d'agrégats**

FROM liste de tables

[WHERE prédicats]

[GROUP BY définition des groupes]

[HAVING prédicats]

Liste des champs servant à l'agrégation de plusieurs enregistrements

Autre clause de restriction qui portent sur les fonctions et champs d'agrégat.

Fonctions d'agrégations

- **Syntaxe** : **nom_de_la_fonction (expression)**.

- **Sémantique** : calcule une valeur à partir des valeurs de tous les enregistrements agrégés.

- **Quelques fonctions disponibles sont :**

- Count : compte

- Sum : somme

- Avg : moyenne

- Min : le plus petit

- Max : le plus grand

- Stddev : écart type

- Variance : variance

- First : premier

- Last : dernier

Compte le nombre d'enregistrements
(indépendamment du type)

Prennent en argument un champ numérique

Retournent le premier et le dernier élément
dans une liste (indépendamment du type)

Fonctions d'agrégations

- **AVG()** calcule la moyenne d'un ensemble de valeurs.
 - Ex : le prix du panier moyen pour de chaque client
- **COUNT()** compte le nombre de lignes concernées.
 - Ex : combien d'achats ont été effectués par chaque client
- **MAX()** récupère la plus haute valeur
 - Ex : le produit le plus cher
- **MIN()** récupère la plus petite valeur.
 - Ex : la date du premier achat d'un client
- **SUM()** calcule la somme de plusieurs ligne
 - Ex : le total de tous les achats d'un client

GROUP BY

- Regroupe les lignes qui ont les mêmes valeurs dans un résumé
 - Ex : Trouver le nombre de client dans chaque pays

Table: Customers

customer_id	first_name	last_name	age	country
1	John	Doe	31	USA
2	Robert	Luna	22	USA
3	David	Robinson	22	UK
4	John	Reinhardt	25	UK
5	Betty	Doe	28	UAE

`SELECT country, COUNT(*) AS number
FROM Customers
GROUP BY country;`

country	number
UAE	1
UK	2
USA	2

HAVING

- Similaire au WHERE sauf qu'il permet de filtrer en utilisant des fonctions d'agrégation
- HAVING est très souvent utilisé en même temps que GROUP BY mais ce n'est pas obligatoire

Fonctions d'agrégation : Sum

Sum : fournit la somme d'un champ

Stock

N° Produit	désignation	quantité
1	Marteau	40
2	Pelle	10
3	Truelle	5
4	Tournevis	100

La quantité de tout le matériel
en stock ?



```
SELECT Sum(quantité)  
FROM Stock
```

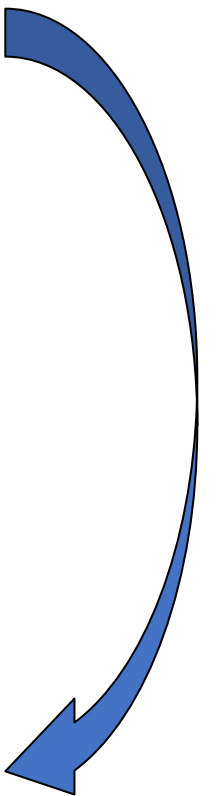


Sum(quantité)
155



QG
155

```
SELECT Sum(quantité) as QG  
FROM Stock
```



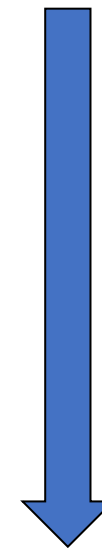
Fonctions d'agrégation : Max

Max : fournit la valeur maximale d'un champ

Avion

Code Avion	Type	Capacite
A0001	Boeing 747	420
A0002	AirBus 320	300
A0003	AirBus 300	280
A0004	Boeing 737	250
A0005	DC 10	120
A0006	Boeing 747	420
A0007	AirBus A380	800
A0008	Boeing 777	550
A0009	AirBus 300	280
A0010	Boeing 737	250

Quelle est la capacité maximale des avions?



```
SELECT Max(capacite) as capacité_max  
FROM Avion
```

Capacité_max

800



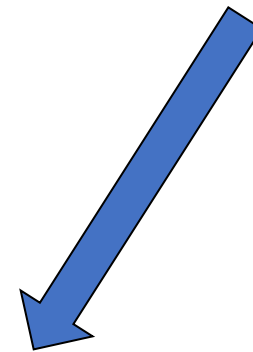
Fonctions d'agrégation : Count (1/2)

Count : permet de compter le nombre d'enregistrements

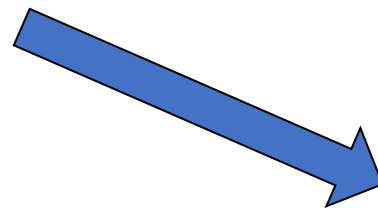
étudiants

N° étudiant	Nom	Prénom	formation
105230	MARTIN	Marie	MMIA
105234	BERNARD	Louis	Gestion
105237	THOMAS	Alice	Anglais
105239	MARTIN	Bernard	Anglais
105250	MARIE	Philippe	Gestion
105211	Mounier	Alice	Gestion

Quel est le nombre total d'étudiants ?



```
SELECT count ([N° étudiant]) as Nombre_étudiants  
FROM étudiant
```



Nombre_étudiants
6

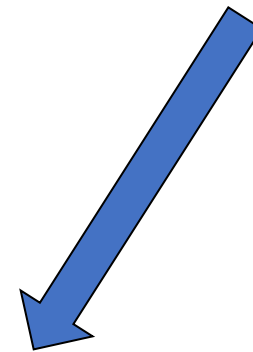
Fonctions d'agrégation : Count (1/2)

Count : permet de compter le nombre d'enregistrements

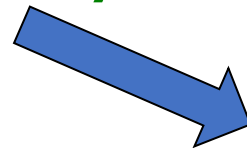
étudiants

N° étudiant	Nom	Prénom	formation
105230	MARTIN	Marie	MMIA
105234	BERNARD	Louis	Gestion
105237	THOMAS	Alice	Anglais
105239	MARTIN	Bernard	Anglais
105250	MARIE	Philippe	Gestion
105211	Mounier	Alice	Gestion

Quel est le nombre d'étudiants
par formation ?



SELECT formation, **count** ([N° étudiant]) **as** Nombre
FROM étudiant
Group by formation



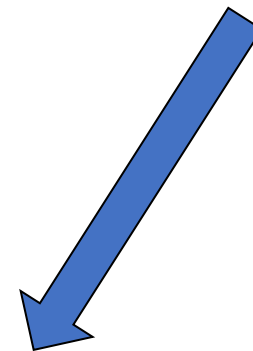
formation	Nombre
MMIA	1
Gestion	3
Anglais	2

Fonctions d'agrégation : Count (2/2)

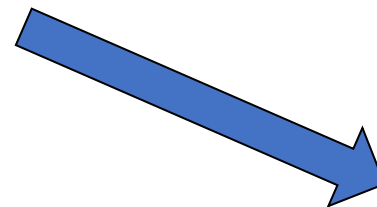
étudiants

N° étudiant	Nom	Prénom	formation
105230	MARTIN	Marie	MMIA
105234	BERNARD	Louis	Gestion
105237	THOMAS	Alice	Anglais
105239	MARTIN	Bernard	Anglais
105250	MARIE	Philippe	Gestion
105211	Mounier	Alice	Gestion

Quelles sont les formations de plus de 2 étudiants?



```
SELECT formation, count([N° étudiant]) as Nombre
FROM   étudiant
Group by formation
Having count([N° étudiant]) > 2
```



formation	Nombre
Gestion	3

Fonctions d'agrégation :un autre exemple

SELECT

```
COUNT(*) AS Flights,  
COUNT(DISTINCT Company) AS Companies,  
SUM(Passengers) AS PSum,  
AVG(Passengers) AS PAvg,  
MIN(Passengers) AS PMin,  
MAX(Passengers) AS PMax
```

FROM Flights

Flight	Company	Destination	Passengers
OK251	CSA	New York	276
LH438	Lufthansa	Stuttgart	68
OK012	CSA	Milano	37
OK321	CSA	London	156
AC906	Air Canada	Toronto	116
KL7621	KLM	Rotterdam	75
KL1245	KLM	Amsterdam	130

Flights	Companies	PSum	PAvg	PMin	PMax
7	4	858	123	37	276

SELECT : HAVING (exemple)

- Combien de vols chaque compagnie a planifié ?
 - Cependant, nous ne sommes pas intéressés par les vol vers Stuttgart et Munich
 - Et nous ne voulons pas les compagnies qui ont effectué un vol ou moins

SELECT : HAVING (exemple)

Flight	Company	Destination	Passengers
OK251	CSA	New York	276
LH438	Lufthansa	Stuttgart	68
OK012	CSA	Milano	37
OK321	CSA	London	156
AC906	Air Canada	Toronto	116
KL7621	KLM	Rotterdam	75
KL1245	KLM	Amsterdam	130

Flight	Company	Destination	Passengers
OK251	CSA	New York	276
OK012		Milano	37
OK321		London	156
AC906	Air Canada	Toronto	116
KL7621	KLM	Rotterdam	75
KL1245		Amsterdam	130

Company	Flights
CSA	3
Air Canada	1
KLM	2

Company	Flights
CSA	3
KLM	2

```
SELECT Company, COUNT(*) AS Flights
FROM Flights
WHERE (Destination NOT IN ('Stuttgart', 'Munich'))
GROUP BY Company HAVING (Flights > 1)
```

Important !

- ➡ Tous les champs sur lesquels on fait la projection (**clause Select**) doivent servir dans la **clause d'agrégation (clause Group By)**.
- ➡ Mais, un champ peut figurer dans la clause d'agrégat sans servir à la projection.
- ➡ De même, les champs utilisés dans la restriction **Where** peuvent ne pas servir à la projection.

Examples

SELECT **formation**, count([N° étudiant])
FROM étudiant
Group by **formation**
Having count([N° étudiant]) >2



SELECT **formation**, **prénom**, count([N° étudiant])
FROM étudiant
Group by **formation**
Having count([N° étudiant]) >2



SELECT **formation**, **prénom**, count([N° étudiant])
FROM étudiant
Group by **formation**, **prénom**
Having count([N° étudiant]) >2



SELECT **formation**, count([N° étudiant])
FROM étudiant
Group by **formation**, **prénom**
Having count([N° étudiant]) >2

