

R2.06 Exploitation BDD

M. T. Pham, M. Khayata

minh-tan.pham@univ-ubs.fr

Cours 2

Fonctions de groupe et Regroupement

Plan

Fonctions de groupe

Comptage simple (COUNT)

Recherche de valeurs extrêmes (MAX/MIN)

Calcul de somme/moyenne (SUM/AVG)

Regroupement

Regroupement (GROUP BY)

Restriction de groupe (HAVING)

Comptage simple

- **COUNT (*)** → renvoie le dernier ROWNUM (le numéro du dernier t-uple) du résultat du mapping

Acteur (noSS(1), nom, prenom, salaire, pays)

Quel est le nombre d'acteurs enregistrés dans la table ?

```
SELECT COUNT ( * )  
FROM Acteur  
;
```

Remarque

- Si la table est mal construite, si elle comporte des **lignes de valeurs manquantes (NULL)**, alors ces lignes sont comptées par **COUNT (*)**
- Si la table est correcte, alors elle comporte une **clé primaire**, et **COUNT (*)** donne le nombre de valeurs de la clé.

Comptage de valeurs différents

- **COUNT (DISTINCT <colonne>)** → renvoie le nombre de valeurs différentes de la colonne

Acteur (noSS(1), nom, prenom, salaire, pays)

De combien de pays viennent les acteurs enregistrés dans la table Acteur ?

```
SELECT COUNT(DISTINCT pays)
FROM Acteur
;
```

- **Remarque** : il est possible de mettre plusieurs colonnes dans le COUNT

Exemple : **COUNT (DISTINCT nom, pays)**

Recherche de valeurs extrêmes

- MIN (<colonne>)
- MAX (<colonne>)

Acteur (noSS(1), nom, prenom, salaire, pays)

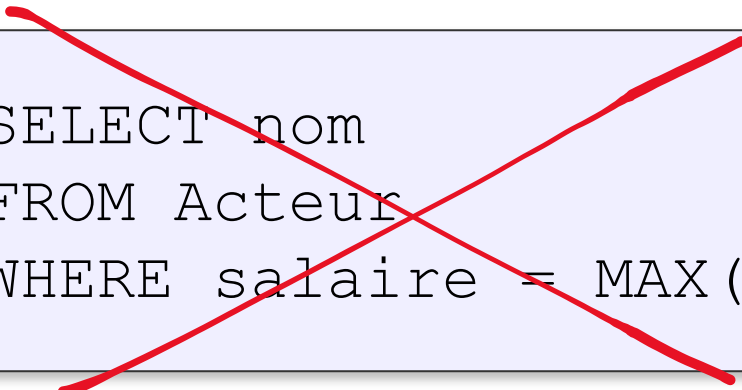
Quel est le salaire minimum des acteurs ?

```
SELECT MIN(salaire)
FROM Acteur
;
```

Recherche de valeurs extrêmes

Acteur (noSS(1), nom, prenom, salaire, pays)

Quel est l'acteur (ou les acteurs) qui gagne le plus ?



```
SELECT nom  
FROM Acteur  
WHERE salaire = MAX(salaire);
```

```
SELECT nom  
FROM Acteur  
WHERE salaire = (SELECT MAX(salaire)  
                  FROM Auteur  
                  );
```

Fonction numérique : somme

- **SUM (<colonne>)** → renvoie la somme des valeurs d'une colonne numérique

Acteur (noSS(1), nom, prenom, salaire, pays)

Quel est le total des salaires des acteurs français ?

```
SELECT SUM(salaire)
FROM Acteur
WHERE UPPER(pays) = 'France'
;
```


Fonction numérique : moyenne

- **AVG (<colonne>)** → renvoie la moyenne des valeurs d'une colonne numérique

Acteur (noSS(1), nom, prenom, salaire, pays)

Quel est le salaire moyen des acteurs français ?

```
SELECT AVG(salaire)
FROM Acteur
WHERE UPPER(pays) = 'France'
;
```

Remarque

Il ne faut pas confondre

➤ **COUNT ()** [nombre de tuples]

On peut compter des cochons, ou n'importe quoi d'autre
et

➤ **SUM ()** [somme des valeurs des tuples]

On ne peut pas faire la somme de deux **cochons**, mais seulement de
leurs **poids**

Remarque

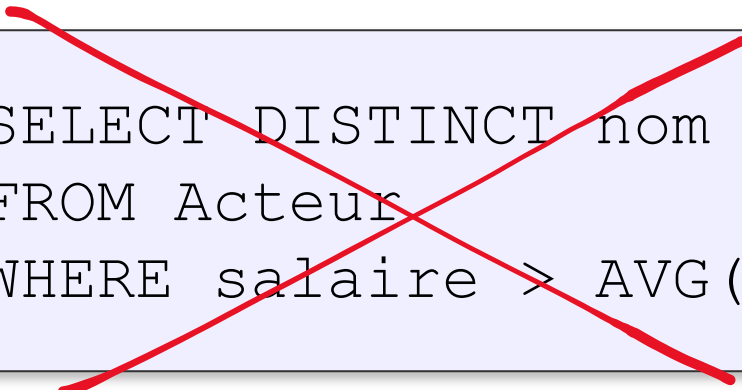
ATTENTION avec DISTINCT

- $\text{MAX/MIN}(\text{DISTINCT col}) \rightarrow$ résultat **équivalent** $\text{MAX/MIN}(\text{col})$
- $\text{SUM/AVG}(\text{DISTINCT col}) \rightarrow$ résultat **différent** $\text{SUM/AVG}(\text{col})$

Un piège !!!

Acteur (noSS(1), nom, prenom, salaire, pays)

Quels sont les noms des acteurs dont le salaire est supérieur à la moyenne ?



```
SELECT DISTINCT nom
FROM Acteur
WHERE salaire > AVG(salaire);
```

```
SELECT DISTINCT nom
FROM Acteur
WHERE salaire > (SELECT AVG(salaire)
                  FROM Auteur
                  );
```

Exemples pratiques

Compagnie (idComp (1), nomComp, pays, estLowCost)

Pilote (idPilote(1), nomPilote, nbHVol, compPil=@Compagnie(idComp))

TypeAvion (idTypeAvion(1),nbPassagers)

Qualification (unPilote=@Pilote(idPilote)(1),unTypeAvion=@TypeAvion(idTypeAvion)(1))

Avion (idAvion(1), leTypeAvion=@TypeAvion(idTypeAvion)(NN), compAv=@Compagnie(idComp)(NN))

Compagnie

idComp	nomComp	pays	estLowCost
1	Air France	France	0
2	Corsair International	France	0
3	EasyJet	Angleterre	1
4	American Airlines	Etats-Unis	0
5	Ryanair	Irlande	1

TypeAvion

idTypeAvion	nbPassagers
A320	174
A350	324
B747	279

Qualification

unPilote	unTypeAvion
1	A320
1	A350
2	A320
2	B747
3	A320
4	A320
4	A350
4	B747
5	A350
5	A320
7	A350
7	B747

Pilote

idPilote	nomPilote	nbHVol	compPil
1	Ridard	1500	1
2	Naert	450	3
3	Godin	450	5
4	Fleurquin	3000	1
5	Pham	900	4
6	Kerbellec	900	
7	Kamp	3000	4

Avion

idAvion	leTypeAvion	compAv
1	A320	1
2	A320	3
3	A350	1
4	A320	2
5	B747	1
6	A350	4
7	B747	4
8	A320	5
9	A320	5

Exemples pratiques

Afficher le nombre de pilotes.



Nombre de lignes

```
SELECT COUNT(*)  
FROM Pilote  
;
```

RESULT

COUNT(*)
7

Afficher le nombre de pilotes travaillant pour une compagnie.



Nombre de valeurs (NOT NULL)

```
SELECT COUNT(compPil)  
FROM Pilote  
;
```

RESULT

COUNT(COMPPIL)
6

Exemples pratiques

Afficher le nombre de compagnies employant (au moins) un pilote.



Nombre de valeurs distinctes

```
SELECT COUNT(DISTINCT compPil)
FROM Pilote
;
```

RESULT

COUNT(DISTINCTCOMPIL)
4

Afficher le plus grand nombre d'heures de vol.



Maximum

```
SELECT MAX(nbHVol)
FROM Pilote
;
```

RESULT

MAX(NBHVOL)
3000

Exemples pratiques

Afficher le nombre d'heures de vol moyen.



```
SELECT AVG(nbHVol)
FROM Pilote
;
```

RESULT

AVG(NBHVOL)
1457,14286

Afficher l'écart-type du nombre d'heures de vol.



```
SELECT STDDEV(nbHVol)
FROM Pilote
;
```

RESULT

STDDEV(NBHVOL)
1111,46621

Exemples pratiques

Afficher les pilotes ayant un nombre d'heures de vol supérieur à la moyenne.



```
SELECT nomPilote
FROM Pilote
WHERE nbHVol >
      (
        SELECT AVG(nbHVol)
        FROM Pilote
      )
;
```

RESULT

NOMPILOTE
Ridard
Fleurquin
Kamp

Pour afficher aussi le nombre d'heures de vol.



```
SELECT nomPilote , nbHVol
FROM Pilote
WHERE nbHVol >
      (
        SELECT AVG(nbHVol)
        FROM Pilote
      )
;
```

RESULT

NOMPILOTE	NBHVOL
Ridard	1500
Fleurquin	3000
Kamp	3000

Plan

Fonctions de groupe

Comptage simple (COUNT)

Recherche de valeurs extrêmes (MAX/MIN)

Calcul de somme/moyenne (SUM/AVG)

Regroupement

Regroupement (GROUP BY)

Restriction de groupe (HAVING)

Regroupement

Problème typique :

- Une table d'acteurs contient les acteurs de plusieurs nationalités ; comment donner des résultats **par nationalité** ?

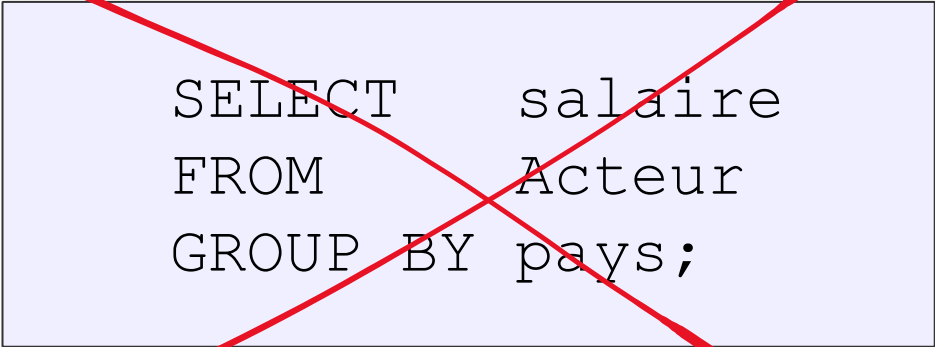
La solution :

- Partage de table en **sous-tables** suivant une colonne : $R \setminus A$
- La clause **GROUP BY** permet de **séparer (virtuellement)** la table étudiée **en sous-tables** suivant les différentes valeurs d'une colonne

Regroupement

- La colonne de projection est **obligatoirement** la colonne de regroupement

Acteur (noSS(1), nom, prenom, salaire, pays)



```
SELECT    salaire
FROM      Acteur
GROUP BY  pays;
```

Regroupement

- La colonne de projection est **obligatoirement** la colonne de regroupement
- **L'intérêt du regroupement** est qu'on peut ajouter à la colonne de projection une fonction comme COUNT, SUM, AVG, etc.

Acteur (noSS(1), nom, prenom, salaire, pays)

Quel est le nombre des auteurs par pays ?

```
SELECT COUNT (*)  
FROM   Acteur  
GROUP BY pays;
```

Quels sont les salaires moyens par nationalité ?

```
SELECT pays, AVG(salaire)  
FROM   Acteur  
GROUP BY pays;
```

GROUP BY and ORDER BY

Acteur (noSS(1), nom, prenom, salaire, pays)

Quels sont les salaires moyens par nationalité ordonnés par ordre décroissant ?

```
SELECT    pays, AVG(salaire)  salairePays
FROM      Acteur
GROUP BY  pays
ORDER BY  salairePays DESC;
```

Restriction de groupe

- On peut restreindre les tuples à l'intérieur d'un groupe, c'est-à-dire d'une sous-table par la clause **HAVING**.

Acteur (noSS(1), nom, prenom, salaire, pays)

Quels sont les pays représentés par au moins 4 acteurs ?

```
SELECT pays, COUNT (*)  
FROM   Acteur  
GROUP BY pays  
HAVING COUNT (*) >= 4;
```

- **Remarque** : à ne pas confondre avec WHERE !!!

Remarque

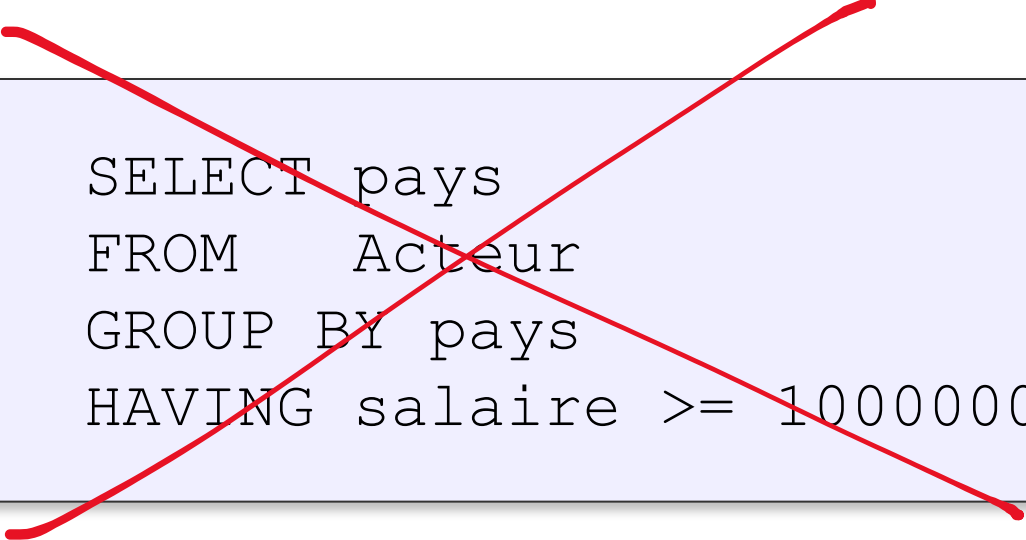
Attention : **NE PAS CONFONDRE**

- **WHERE** → qui s'applique à chaque tuple de la table,
avec
- **HAVING** → qui s'applique au groupe de tuples (dont il faut toujours avoir un GROUP BY)

Attention !!!

Acteur (noSS(1), nom, prenom, salaire, pays)

Quels sont les nationalités dont la moyenne des salaires des acteurs est ≥ 1000000 ?



```
SELECT pays  
FROM Acteur  
GROUP BY pays  
HAVING salaire  $\geq$  1000000;
```

Solution incorrecte !!!

Attention !!!

Acteur (noSS(1), nom, prenom, salaire, pays)

Quels sont les pays dont la moyenne des salaires des acteurs est ≥ 1000000 ?

```
SELECT pays
FROM   Acteur
GROUP BY pays
HAVING AVG(salaire) >= 1000000;
```

Solution correcte !!!

Composition de fonctions

Acteur (noSS(1), nom, prenom, salaire, pays)

Quels est le nombre maximum d'acteurs regroupés par nationalité ?

```
SELECT MAX (COUNT (*) )  
FROM   Acteur  
GROUP BY pays;
```

**OK avec Oracle !
Ne fonctionne pas avec
MySQL !!!**

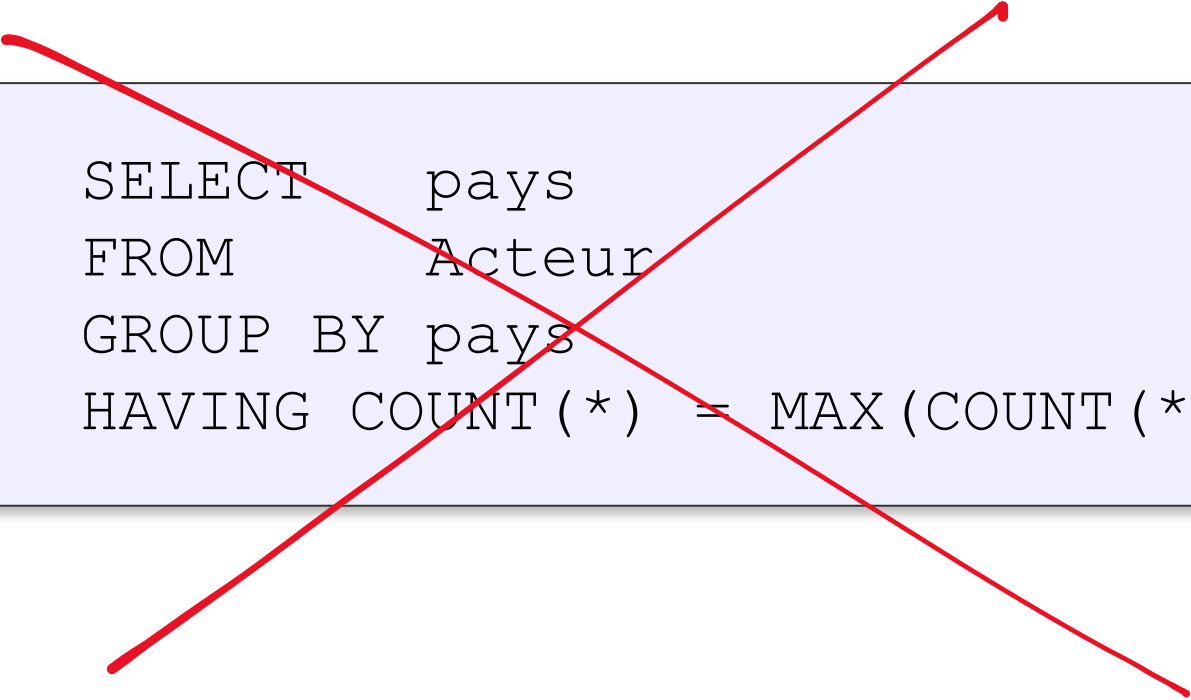
```
SELECT MAX(nombre)  
FROM ( SELECT COUNT(*) AS nombre  
        FROM   Acteur  
        GROUP BY pays  
      ) ;
```

OK avec Oracle/MySQL

Attention !!!

Acteur (noSS(1), nom, prenom, salaire, pays)

Quelle(s) est (sont) la (les) nationalités ayant le plus d'acteurs ?



```
SELECT    pays
FROM      Acteur
GROUP BY  pays
HAVING    COUNT (*) = MAX (COUNT (*)) ;
```

Attention !!!

Acteur (noSS(1), nom, prenom, salaire, pays)

Quelle(s) est (sont) la (les) nationalités ayant le plus d'acteurs ?

```
SELECT    pays
FROM      Acteur
GROUP BY  pays
HAVING    COUNT (*) = (SELECT MAX (COUNT (*) )
                       FROM Acteur
                       GROUP By pays
                       ) ;
```

Solution correcte !!!

Exemples pratiques

Compagnie (idComp (1), nomComp, pays, estLowCost)

Pilote (idPilote(1), nomPilote, nbHVol, compPil=@Compagnie(idComp))

TypeAvion (idTypeAvion(1),nbPassagers)

Qualification (unPilote=@Pilote(idPilote)(1),unTypeAvion=@TypeAvion(idTypeAvion)(1))

Avion (idAvion(1), leTypeAvion=@TypeAvion(idTypeAvion)(NN), compAv=@Compagnie(idComp)(NN))

Compagnie

idComp	nomComp	pays	estLowCost
1	Air France	France	0
2	Corsair International	France	0
3	EasyJet	Angleterre	1
4	American Airlines	Etats-Unis	0
5	Ryanair	Irlande	1

TypeAvion

idTypeAvion	nbPassagers
A320	174
A350	324
B747	279

Qualification

unPilote	unTypeAvion
1	A320
1	A350
2	A320
2	B747
3	A320
4	A320
4	A350
4	B747
5	A350
5	A320
7	A350
7	B747

Pilote

idPilote	nomPilote	nbHVol	compPil
1	Ridard	1500	1
2	Naert	450	3
3	Godin	450	5
4	Fleurquin	3000	1
5	Pham	900	4
6	Kerbellec	900	
7	Kamp	3000	4

Avion

idAvion	leTypeAvion	compAv
1	A320	1
2	A320	3
3	A350	1
4	A320	2
5	B747	1
6	A350	4
7	B747	4
8	A320	5
9	A320	5

Exemples pratiques

Pour chaque compagnie, afficher le nombre d'avions.



```
SELECT compAv, COUNT(*)  
FROM Avion  
GROUP BY compAv  
;
```

RESULT Pour chaque groupe, on compte le nombre de lignes

COMPAV	COUNT (*)
1	3
2	1
4	2
5	2
3	1

Exemples pratiques

Pour chaque compagnie, afficher le nombre de types d'avion.



```
SELECT compAv, COUNT(DISTINCT leTypeAvion)
FROM Avion
GROUP BY compAv
;
```

RESULT Pour chaque groupe, on compte le nombre de valeurs distinctes dans une colonne

COMPAV	COUNT(DISTINCTLETYPEAVION)
1	3
2	1
5	1
4	2
3	1

Exemples pratiques

Pour chaque compagnie désignée par son nom (et triée dans l'ordre alphabétique), afficher le nombre de types d'avion.



```
SELECT nomComp, COUNT(DISTINCT leTypeAvion)
FROM Compagnie, Avion
WHERE idComp = compAv
GROUP BY nomComp
ORDER BY nomComp
;
```

RESULT

NOMCOMP	COUNT(DISTINCTLETYPEAVION)
Air France	3
American Airlines	2
Corsair International	1
EasyJet	1
Ryanair	1

Exemples pratiques

Pour chaque pilote, afficher le nombre de qualifications (éventuellement 0)



```
SELECT unPilote , COUNT(unTypeAvion)
FROM Qualification
GROUP BY unPilote
;
```



Il manque un pilote

UNPILOTE	COUNT(UNTYPEAVION)
1	2
2	2
3	1
4	3
5	2
7	2



```
SELECT idPilote , COUNT(unTypeAvion)
FROM Pilote
LEFT JOIN Qualification ON idPilote = unPilote
GROUP BY idPilote
;
```

RESULT

IDPILOTE	COUNT(UNTYPEAVION)
1	2
2	2
3	1
4	3
5	2
6	0
7	2

Exemples pratiques

Pour chaque compagnie, afficher le nombre de pilotes (éventuellement 0)



```
SELECT compPil, COUNT(idPilote)
FROM Pilote
GROUP BY compPil
;
```



Il manque une "étiquette"

COMPIL	COUNT(IDPILOTE)
1	2
	1
5	1
4	2
3	1



```
SELECT idComp, COUNT(idPilote)
FROM Compagnie
LEFT JOIN Pilote ON idComp = compPil
GROUP BY idComp
;
```

RESULT

IDCOMP	COUNT(IDPILOTE)
1	2
2	0
3	1
4	2
5	1

Exemples pratiques

Afficher le plus grand nombre de pilotes dans une compagnie.



```
SELECT MAX( nbPilotes )  
FROM  
(  
  SELECT compPil, COUNT(idPilote) nbPilotes  
  FROM Pilote  
  GROUP BY compPil  
)  
;
```

RESULT

MAX(NBPILOTES)
2

Afficher les compagnies ayant deux pilotes.



```
SELECT compPil, COUNT(idPilote) nbPilotes  
FROM Pilote  
GROUP BY compPil  
HAVING COUNT(compPil) = 2  
;
```

RESULT

COMPIL	NBPILOTES
1	2
4	2

Exemples pratiques

Afficher les compagnies ayant le plus grand nombre de pilotes.



```
SELECT compPil
FROM Pilote
GROUP BY compPil
HAVING COUNT(idPilote) =
    (
        SELECT MAX(COUNT(idPilote))
        FROM Pilote
        GROUP BY compPil
    )
;
```

RESULT

COMPPIL
1
4

Références

- <https://fr.wikipedia.org/>
- Cours BDD Anthony Ridard, <https://math-ridard.fr/>
- Christian Soutou, *Modélisation des bases de données : UML et les modèles entité-association*, 3^{ème} édition, Groupe Eyrolles.
- Christian Soutou, *SQL pour Oracle : Optimisation des requêtes et schémas*, 5^{ème} édition, Groupe Eyrolles.
- Elisabetta De Mria, *Cours L2 Informatique*, UFR Sciences, Université Côte d'Azur, <https://www.i3s.unice.fr/%18edemaria/>
- Oracle SQL Tutorials, <https://www.w3schools.com/sql/default.asp>
- MySQL Tutorials, <https://www.w3schools.com/MySQL/default.asp>