

# Plan

## Bases de Données : opérations sur plusieurs relations

Stéphane Devismes

Université Grenoble Alpes

26 août 2020

- 1 Jointures
- 2 Opérations ensemblistes
- 3 Sous-requêtes
- 4 Agrégations et groupements

S. Devismes (UGA)    Opération sur plusieurs relations    26 août 2020    1 / 75

Jointures    Opérations ensemblistes    Sous-requêtes    Agrégations et groupements

### Jointure interne (ou theta-produit)

**BUT :** Former toutes les lignes  $(x_1, \dots, x_n, x'_1, \dots, x'_{n'})$

- avec  $(x_1, \dots, x_n)$  dans  $R$  et
- $(x'_1, \dots, x'_{n'})$  dans  $R'$
- telles que  $(x_1, \dots, x_n, x'_1, \dots, x'_{n'})$  vérifie la condition  $C$ .

En général,  $C$  compare les valeurs de certains attributs de  $R$  avec les valeurs de certains attributs de  $R'$ .

On préfixe les noms des attributs pour lever les éventuelles ambiguïtés : **une condition de la forme  $A = A$  sera donc notée  $R.A = R'.A$ .**

S. Devismes (UGA)    Opération sur plusieurs relations    26 août 2020    2 / 75

Jointures    Opérations ensemblistes    Sous-requêtes    Agrégations et groupements

### Jointure interne : exemples (1/3)

$R_1$		
A	B	C
a	b	c
x	y	z
u	v	w
r	s	t

$R_2$		
A	B	D
a	b	w
a	y	z
u	v	z
u	v	i

Jointure entre  $R_1$  et  $R_2$  pour la condition  $R_1.A = R_2.A$  :

$R_3$					
$R_1.A$	$R_1.B$	$R_1.C$	$R_2.A$	$R_2.B$	$R_2.D$
a	b	c	a	b	w
a	b	c	a	y	z
u	v	w	u	v	z
u	v	w	u	v	i

## Jointure interne : exemples (2/3)

R <sub>1</sub>		
A	B	C
a	b	c
x	y	z
u	v	w
r	s	t

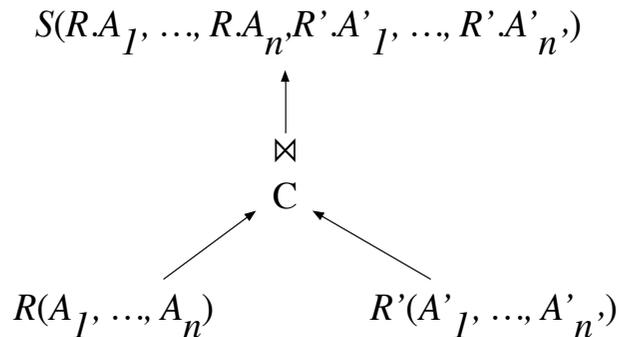
R <sub>2</sub>		
A	B	D
a	b	w
a	y	z
u	v	z
u	v	i

Jointure entre R<sub>1</sub> et R<sub>2</sub> pour la condition R<sub>1</sub>.A = R<sub>2</sub>.A et R<sub>1</sub>.B = R<sub>2</sub>.B :

R <sub>3</sub>					
R <sub>1</sub> .A	R <sub>1</sub> .B	R <sub>1</sub> .C	R <sub>2</sub> .A	R <sub>2</sub> .B	R <sub>2</sub> .D
a	b	c	a	b	w
u	v	w	u	v	z
u	v	w	u	v	i

## Jointure interne en AR

Le schéma de S est la réunion des deux schémas.



La condition C est **essentiellement de même nature** que les conditions suivant un WHERE, mais en principe elle utilise **des données venant de chacune des deux tables**.

## Jointure interne : exemples (3/3)

R <sub>1</sub>		
A	B	C
a	b	c
x	y	z
u	v	w
r	s	t

R <sub>2</sub>		
A	B	D
a	b	w
a	y	z
u	v	z
u	v	i

Jointure entre R<sub>1</sub> et R<sub>2</sub> pour la condition R<sub>1</sub>.B = R<sub>2</sub>.B et R<sub>1</sub>.C = R<sub>2</sub>.D :

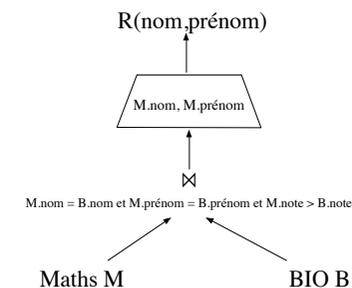
R <sub>3</sub>					
R <sub>1</sub> .A	R <sub>1</sub> .B	R <sub>1</sub> .C	R <sub>2</sub> .A	R <sub>2</sub> .B	R <sub>2</sub> .D
x	y	z	a	y	z

## Jointure interne en AR : exemples (1/4)

Quels sont les élèves qui ont une note meilleure en maths qu'en bio ?

- Schéma : R(nom : Chaîne, prénom : Chaîne)
- Spécification :  $\langle n, p \rangle \in R \iff$  l'élève de nom n et de prénom p a une note en maths m et une note en bio b telles que  $m > b$ . (« meilleur » est pris au sens strict).

Algèbre relationnelle :



## Jointure interne en SQL : exemples (1/4)

Quels sont les élèves qui ont une note meilleure en maths qu'en bio ?

```
SELECT M.nom, M.prenom
FROM Maths M INNER JOIN Bio B
ON M.nom=B.nom AND M.prenom=B.prenom AND M.note > B.note;
```

-- ou bien

```
SELECT M.nom, M.prenom
FROM Maths M JOIN Bio B
ON M.nom=B.nom AND M.prenom=B.prenom AND M.note > B.note;
```

**Remarque :** le renommage n'est pas obligatoire. Le cas échéant, il faut préfixer par le nom de la table.

## Jointure interne en SQL : exemples (2/4)

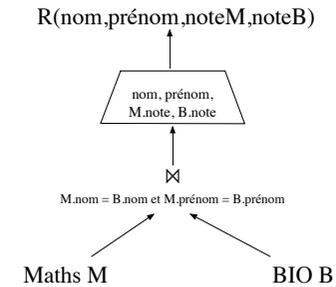
Donner les notes de maths et de bio de tous les élèves.

## Jointure interne en AR : exemples (2/4)

Donner les notes de maths et de bio de tous les élèves.

- Schéma :  $R(\underline{\text{nom}} : \text{Chaine}, \underline{\text{prénom}} : \text{Chaine}, \text{noteM} : \text{Nombre}, \text{noteB} : \text{Nombre})$ .
- Spécification :  $\langle n, p, m, b \rangle \in R \iff$  l'élève de nom  $n$  et de prénom  $p$  a pour notes  $m$  en maths et  $b$  en bio.

Algèbre relationnelle :



## Jointure avec USING : exemple

$R_1$		
A	B	C
a	b	c
x	y	z
u	v	w
r	s	t

$R_2$		
A	B	D
a	b	w
a	y	z
u	v	z
u	v	i

Jointure entre  $R_1$  et  $R_2$  avec  $USING(A,B)$  :

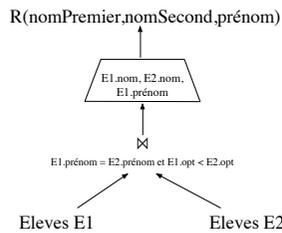
$R_3$			
A	B	$R_1.C$	$R_2.D$
a	b	c	w
u	v	w	z
u	v	w	i

## Jointure interne en AR : exemples (3/4), auto-jointure

Quels sont les élèves de même prénom dans des options différentes ?

- Schéma :  $R(\text{nomPremier} : \text{Chaine}, \text{nomSecond} : \text{Chaine}, \text{prénom} : \text{Chaine})$
- Spécification :  $\langle n1, n2, p \rangle \in R \iff$  les élèves  $n1, p$  et  $n2, p$  ont le même prénom  $p$  et suivent deux options différentes. De plus, pour éviter les « symétriques », le numéro de l'option de  $n1, p$  est strictement inférieur au numéro de l'option de  $n2, p$ .

Algèbre relationnelle :

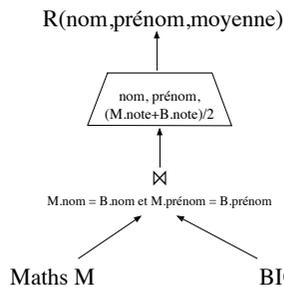


## Jointure interne en AR : exemples (4/4)

Donner la moyenne des notes de maths et de bio de chaque élève.

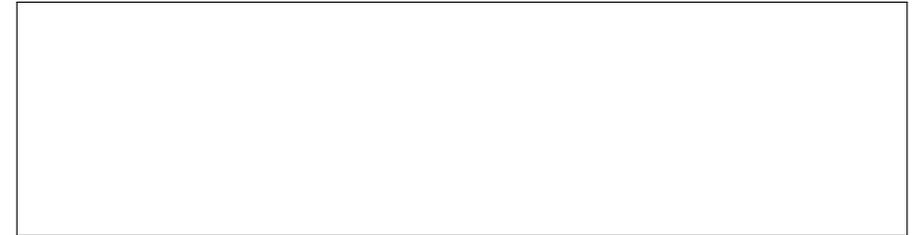
- Schéma :  $R(\text{nom} : \text{Chaine}, \text{prénom} : \text{Chaine}, \text{moyenne} : \text{Nombre})$
- Spécification :  $\langle n, p, m \rangle \in R \iff$   $m$  est la moyenne des notes de maths et de bio de l'élève  $p, n$ .

Algèbre relationnelle :



## Jointure interne en SQL : exemples (3/4), auto-jointure

Quels sont les élèves de même prénom dans des options différentes ?



## Jointure interne en SQL : exemples (4/4)

Donner la moyenne des notes de maths et de bio de chaque élève.



## Remarque : renommage

- Renommage d'un attribut en Oracle-SQL : attribut AS att ou attribut att.
- Renommage d'une relation en Oracle-SQL : Table T (mais pas Table AS T).
- Renommage d'uns sous-requête en Oracle-SQL : ...sous-requête...R.

## Remarque : abréviations

Donner, pour chaque élève, sa note de maths, sa note de bio, et la moyenne des deux.

- Code SQL avec erreur : noteM+noteB :

```
SELECT nom, prenom, M.note AS noteM, B.note AS noteB,
       (noteM+noteB)/2 AS moyenne
FROM Maths M JOIN Bio B USING (nom, prenom);
-- ERROR: "NOTEB": invalid identifier
```

- Code SQL sans erreur : M.note+B.note :

```
SELECT nom, prenom, M.note AS noteM, B.note AS noteB,
       (M.note+B.note)/2 AS moyenne
FROM Maths M JOIN Bio B USING (nom, prenom);
```

## Jointure naturelle (ou produit naturel)

La relation résultat  $S(B_1, \dots, B_p)$  a pour attributs l'union des attributs de  $R$  et de  $R'$  : ceux qui sont communs à  $R$  et à  $R'$  apparaissent une seule fois dans  $S$ .

**BUT :** Former toutes les lignes  $(z_1, \dots, z_p)$  obtenues en regroupant chaque ligne de  $R$  avec chaque ligne de  $R'$  qui a les mêmes valeurs aux attributs communs à  $R$  et  $R'$ .

## Jointure naturelle : observations

La jointure naturelle permet d'alléger l'écriture d'une requête et permet ainsi plus de clarté.

Cependant, la jointure naturelle peut mener à des erreurs dans le cas où plusieurs attributs « de sémantique différente » portent le même nom, ou si plusieurs attributs « de même sémantique » portent des noms différents.

Si il n'y a aucun attribut commun à  $R$  et à  $R'$ , alors  $R * R'$  correspond au produit cartésien.

## Jointure naturelle : exemple

$R_1$		
A	B	C
a	b	c
x	y	z
u	v	w
r	s	t

$R_2$		
A	B	D
a	b	w
a	y	z
u	v	z
u	v	i

Jointure naturelle entre  $R_1$  et  $R_2$  :

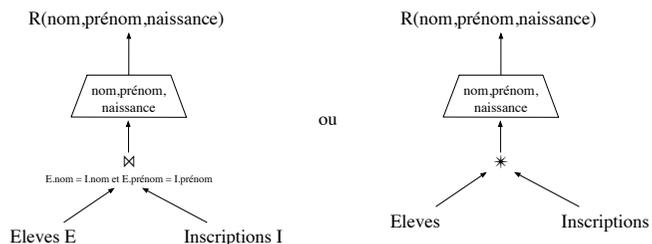
$R_3$			
A	B	C	D
a	b	c	w
u	v	w	z
u	v	w	i

## Jointure naturelle en AR : exemples (1/2)

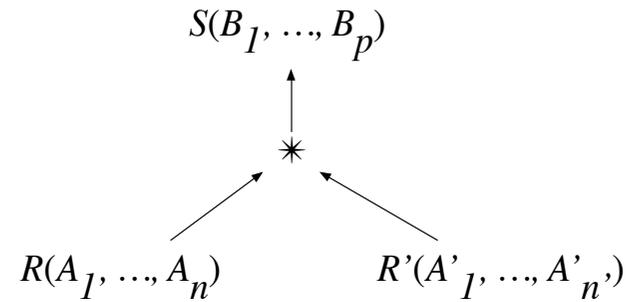
Quels sont les élèves qui font une activité, avec leur date de naissance ?

- Schéma :  $R(\text{nom} : \text{Chaîne}, \text{prénom} : \text{Chaîne}, \text{naissance} : \text{Date})$
- Spécification :  $\langle n, p, j \rangle \in R \iff$  l'élève de nom  $n$  et de prénom  $p$  pratique une activité, et sa date de naissance est  $j$ .

Algèbre relationnelle :



## Jointure naturelle en AR



## Jointure naturelle en SQL : exemples (1/2)

Quels sont les élèves qui font une activité, avec leur date de naissance ?

```
SELECT DISTINCT nom, prénom, naissance
FROM Eleves NATURAL JOIN Inscriptions;
```

Remarque : avec une jointure interne c'est plus long !

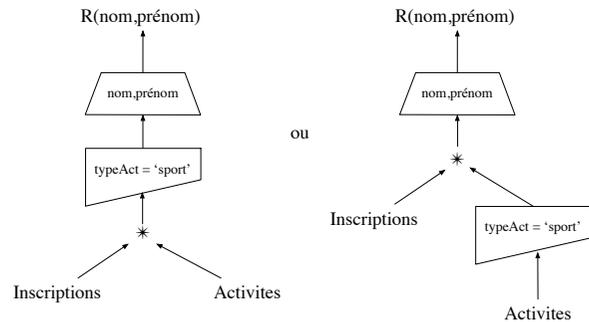
```
SELECT DISTINCT E.nom, E.prénom, naissance
FROM Eleves E JOIN Inscriptions I
ON E.nom = I.nom AND E.prénom = I.prénom;
```

## Jointure naturelle en AR : exemples (2/2)

Quels sont les élèves qui font du sport ?

- Schéma :  $R(\text{nom} : \text{Chaîne}, \text{prénom} : \text{Chaîne})$
- Spécification :  $\langle n, p \rangle \in R \iff \text{l'élève } n, p \text{ fait du sport.}$

Algèbre relationnelle :



## Jointure naturelle vs USING

Donner, pour chaque élève qui suit les 3 matières, la moyenne de ses notes en maths, bio et histoire.

Jointure naturelle impossible !

```
SELECT nom, prénom, (M.note+B.note+H.note)/3 AS moyenne
FROM Maths M JOIN Bio B USING (nom, prénom)
JOIN Histoire H USING (nom, prénom);
```

## Jointure naturelle en SQL : exemples (2/2)

Quels sont les élèves qui font du sport ?

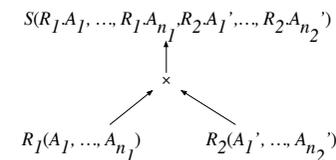


## Jointure croisée (ou produit cartésien)

**BUT :** Former toutes les lignes  $(x_1, \dots, x_{n_1}, x'_1, \dots, x'_{n_2})$  avec  $(x_1, \dots, x_{n_1})$  dans  $R_1$  et  $(x'_1, \dots, x'_{n_2})$  dans  $R_2$

Le produit cartésien est rarement utilisé car rarement intéressant.

De plus, il est très gros :  $S$  comporte  $n_1 + n_2$  colonnes et  $p_1 \times p_2$  lignes, où  $p_1$  est le nombre de lignes de  $R_1$  et  $p_2$  le nombre de lignes de  $R_2$ .



## Jointure croisée : exemple

R <sub>1</sub>		
A	B	C
a	b	c
x	y	z
u	v	w
r	s	t

R <sub>2</sub>		
A	B	D
a	b	w
a	y	z
u	v	z
u	v	i

Jointure croisée entre R<sub>1</sub> et R<sub>2</sub>

R <sub>3</sub>					
R <sub>1</sub> .A	R <sub>1</sub> .B	R <sub>1</sub> .C	R <sub>2</sub> .A	R <sub>2</sub> .B	R <sub>2</sub> .D
a	b	c	a	b	w
a	b	c	a	y	z
a	b	c	u	v	z
a	b	c	u	v	i
x	y	z	a	b	w
x	y	z	a	y	z
x	y	z	u	v	z
x	y	z	u	v	i
u	v	w	a	b	w
u	v	w	a	y	z
u	v	w	u	v	z
u	v	w	u	v	i
r	s	t	a	b	w
r	s	t	a	y	z
r	s	t	u	v	z
r	s	t	u	v	i

## Jointure croisée en SQL : exemple

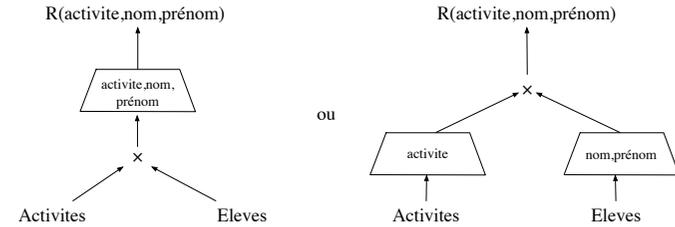
Quelles sont toutes les combinaisons (activité, élève) possibles ?

## Jointure croisée en AR : exemple

Quelles sont toutes les combinaisons (activité, élève) possibles ?

- Schéma :  $R(\text{activité} : \text{Chaîne}, \text{nom} : \text{Chaîne}, \text{prénom} : \text{Chaîne})$
- Spécification :  $\langle a, n, p \rangle \in R \iff$  l'activité  $a$  est proposée par l'école et  $n, p$  est un élève de l'école.

Algèbre relationnelle :



## Remarque sur les jointures

Pour une jointure avec conditions, certaines conditions portent sur la jointure et d'autres portent sur une seule des tables de la jointure.

A priori il y a 3 approches possibles, une est conseillée et les deux autres déconseillées, bien qu'elles retournent le même résultat :

- **Conseillé** : avec JOIN ON (ou JOIN USING) et WHERE, en mettant les conditions de jointure dans le JOIN et les autres dans le WHERE.
- **DÉConseillé** : avec JOIN ON (ou JOIN USING), en mettant toutes les conditions dans le JOIN.
- **DÉConseillé** : avec CROSS JOIN (ou une simple virgule) et WHERE, en mettant toutes les conditions dans le WHERE (c'est l'utilisation « à l'ancienne » de la jointure croisée).

# Opérations ensemblistes et connecteurs logiques

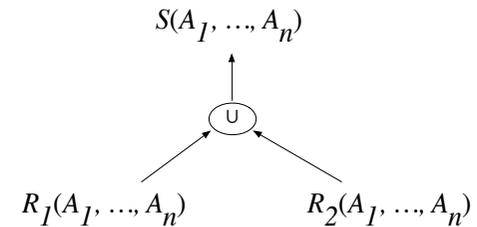
union	$\cup$	UNION	ou	$\vee, +$	OR
intersection	$\cap$	INTERSECT	et	$\wedge, \cdot$	AND
complémentaire	$-, \setminus$	MINUS	non	$\neg$	NOT

# Union

$R_1 \cup R_2$ . UNION (résultat sans répétition), UNION ALL (résultat avec répétition, n'existe pas en AR).

Sur deux relations  $R_1$  et  $R_2$  de même schéma.

Conserver les lignes qui sont dans  $R_1$  ou dans  $R_2$  (ou dans les deux : le "ou" n'est pas exclusif).

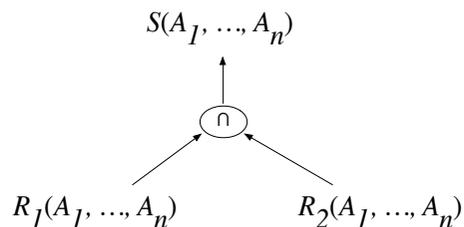


# Intersection

$R_1 \cap R_2$ . INTERSECT (résultat sans répétition).

Sur deux relations  $R_1$  et  $R_2$  de même schéma.

Conserver les lignes qui sont à la fois dans  $R_1$  et dans  $R_2$ .

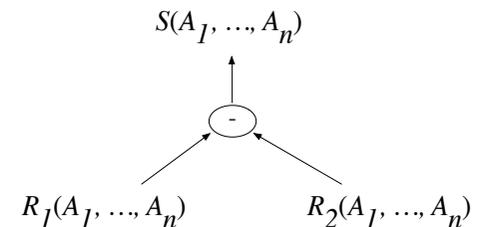


# Différence

$R_1 - R_2$ . MINUS (résultat sans répétition).

Sur deux relations  $R_1$  et  $R_2$  de même schéma.

Conserver les lignes qui sont dans  $R_1$  mais pas dans  $R_2$ .



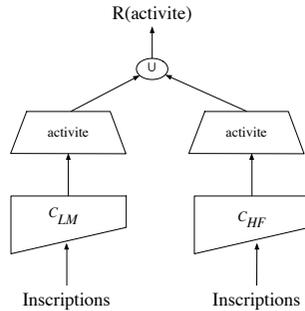
**Attention :**  $R_1 - R_2$  est différent de  $R_2 - R_1$ .

## Union en AR : exemple

Quels sont les activités choisies par Michel Leblanc ou par Franck Hassan ?

- Schéma :  $R(\text{activité} : \text{Chaîne})$
- Spécification :  $\langle a \rangle \in R \iff$  l'activité  $a$  est choisie par Michel Leblanc ou par Franck Hassan.

Algèbre relationnelle :  $C_{LM}$  pour nom = 'Leblanc' et prénom = 'Michel'  
 $C_{HF}$  pour nom = 'Hassan' et prénom = 'Franck'

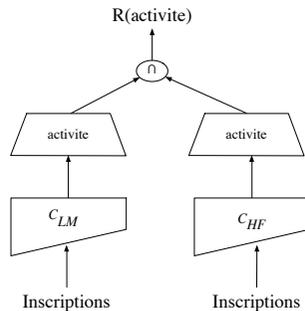


## Intersection en AR : exemple

Quels sont les activités choisies par Michel Leblanc et par Franck Hassan ?

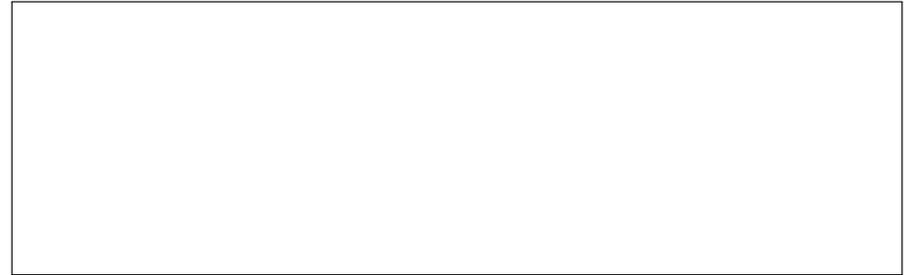
- Schéma :  $R(\text{activité} : \text{Chaîne})$
- Spécification :  $\langle a \rangle \in R \iff$  l'activité  $a$  est choisie par Michel Leblanc et par Franck Hassan.

Algèbre relationnelle :  $C_{LM}$  pour nom = 'Leblanc' et prénom = 'Michel'  
 $C_{HF}$  pour nom = 'Hassan' et prénom = 'Franck'



## Union en SQL : exemple

Quels sont les activités choisies par Michel Leblanc ou par Franck Hassan ?



## Intersection en SQL : exemple

Quels sont les activités choisies par Michel Leblanc et par Franck Hassan ?



## Erreur classique

Quels sont les activités choisies par Michel Leblanc et par Franck Hassan ?

```
SELECT activite
FROM Inscriptions
WHERE (nom = 'Hassan' AND prenom = 'Franck')
AND (nom = 'Leblanc' AND prenom = 'Michel') ;
-- no row selected
```

## Sous-requêtes (requêtes imbriquées) (2/2)

- **tous (ALL).**  
Comparer une expression avec **toutes** les valeurs dans  $R'$  :  
 $A_1 > \text{tous } R'$ ,  
Cette condition a pour valeur « vrai » si et seulement si pour **tout** élément de  $R'$  on obtient « vrai ».  
Si la sous-requête  $R'$  ne retourne aucune ligne alors la condition a pour valeur « vrai ».
- **au moins un (ANY).** Comparer une expression avec **au moins une** des valeurs dans  $R'$  :  
 $A_1 > \text{au moins un } R'$ ,  
Cette condition a pour valeur « vrai » si et seulement si pour **au moins un** élément de  $R'$  on obtient « vrai ».  
Si la sous-requête  $R'$  ne retourne aucune ligne alors la condition a pour valeur « faux ».

## Sous-requêtes (requêtes imbriquées) (1/2)

Certaines conditions utilisent un ensemble de valeurs, qui est donné par une **sous-requête**  $R'$  renvoyant une à plusieurs colonnes. Ci-dessous nous considérons uniquement le cas à une colonne<sup>1</sup>.

- $\in$  (IN),  $\notin$  (NOT IN).  
Appartenance d'un attribut à un ensemble  $R'$  de valeurs :  
 $A_1 \in R'$
- **il existe (EXISTS), il n'existe pas (NOT EXISTS).**  
Vérifier si une sous-requête fournit au moins une valeur :  
**il existe**  $R'$ ,  
Cette condition a pour valeur « vrai » si et seulement si la valeur de la sous-requête comporte au moins une ligne (il existe au moins un élément dans la relation correspondant à la sous-requête).

1. Dans le cas où il y a  $x > 1$  colonnes, il faut remplacer dans la suite l'attribut par un x-uplet d'attributs de la forme  $(A_1, \dots, A_x)$

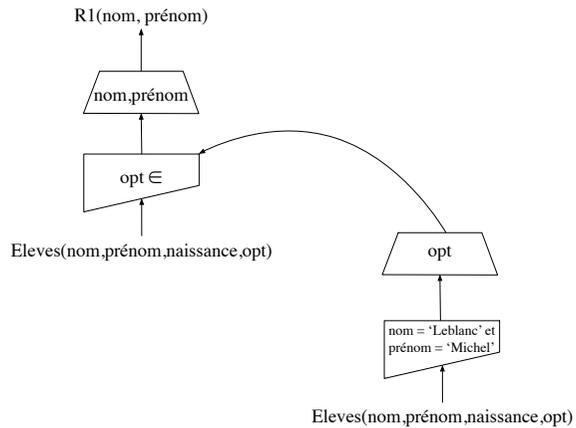
## Sous-requêtes : exemples en AR (1/2)

Quels sont les élèves qui suivent la même option que Michel Leblanc ?

- Schéma :  $R(\text{nom} : \text{Chaine}, \text{prenom} : \text{Chaine})$
- Spécification (2 interprétations de l'énoncé) :
  - $\langle n, p \rangle \in R_1 \iff$  l'élève  $p, n$  suit la même option que Michel Leblanc (Michel Leblanc inclus).
  - $\langle n, p \rangle \in R_2 \iff$  l'élève  $p, n$  suit la même option que Michel Leblanc. (Michel Leblanc exclus).

## Sous-requêtes : exemples en AR (2/2)

Quels sont les élèves qui suivent la même option que Michel Leblanc ?



## Sous-requêtes : exemples en SQL

Quels sont les élèves qui suivent la même option que Michel Leblanc ?



### Définition

Une **agrégation** est une opération qui part d'une relation  $R(A_1, \dots, A_n)$  et calcule une nouvelle relation  $S(B_1, \dots, B_n)$  en effectuant une opération sur les lignes (« verticales »).

Par exemple : « Calculer la moyenne des notes à l'examen de maths ».

Souvent, on commence par **classer (ou grouper)** les lignes de la table selon la valeur de certains attributs, et on effectue l'opération « verticale » pour chaque groupe.

Par exemple : « Calculer, pour chaque option, la moyenne des notes à l'examen de maths ».

**Remarque** : quand on groupe les lignes de la table selon la valeur de certains attributs, on forme une **partition** de la relation (vue comme un ensemble de  $n$ -uplets).

### Opérateurs d'agrégation

Les opérations d'**agrégation** sont des opérations « verticales » : MAX, MIN, AVG, SUM, COUNT, COUNT DISTINCT.

Elles calculent, pour un attribut  $A$ , une quantité qui dépend de la valeur de  $A$  dans **toutes les lignes** de la table : minimum MIN( $A$ ), maximum MAX( $A$ ), moyenne AVG( $A$ ), somme SUM( $A$ ).

On peut aussi compter le nombre de valeurs pour  $A$  par COUNT( $A$ ),

De plus, le nombre de lignes de la table est donné par COUNT(\*).

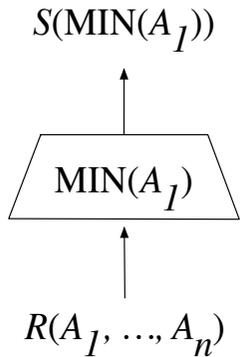
Le nombre de valeurs **distinctes** pour  $A$  est donné par COUNT(DISTINCT  $A$ ).

Une opération d'agrégation concerne **un seul attribut**, c'est valable même pour COUNT (sauf bien sûr COUNT(\*)).

Dans tous les cas, le résultat est une table avec **une seule ligne** ; en particulier COUNT( $A$ ) ou COUNT(DISTINCT  $A$ ) ou COUNT(\*) renvoie 0 si la table est vide.

# Agrégation en AR

**Exemple :** Retourner le minimum des valeurs de  $A_1$ .

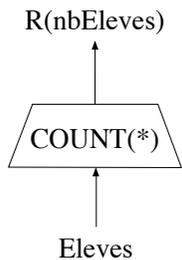


# Agrégation en AR : exemples (1/3)

**Quel est le nombre d'élèves dans l'école ?**

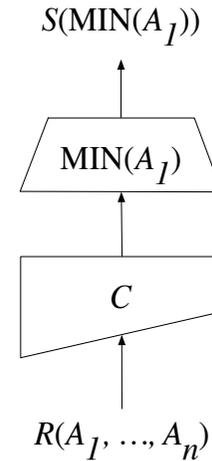
- Schéma :  $R(\text{nbEleves} : \text{Nombre})$
- Spécification :  $\langle nb \rangle \in R \iff$  Il y a  $nb$  élèves dans l'école. En particulier  $nb = 0$  s'il n'y a aucun élève. La relation a exactement une ligne.

Algèbre relationnelle :



# Agrégation après restriction en AR

**Exemple :** Conserver les lignes vérifiant une condition  $C$  et retourner le minimum des valeurs de  $A_1$  sur ces lignes.



# Aggrégation en SQL : exemples (1/3)

**Quel est le nombre d'élèves dans l'école ?**

```
SELECT COUNT(*) AS nbEleves
FROM Eleves;
```

Attention la requête suivante provoque une erreur.

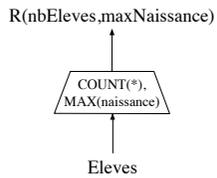
```
SELECT COUNT(nom, prenom) AS nbEleves
FROM Eleves;
-- invalid number of arguments
```

## Agrégation en AR : exemples (2/3)

Quel est le nombre d'élèves dans l'école et la date de naissance du plus jeune ?

- Schéma :  $R(\text{nbEleves} : \text{Nombre}, \text{maxNaissance} : \text{Date})$
- Spécification :  $\langle nb, j \rangle \in R \iff$  Il y a  $nb$  élèves dans l'école et les élèves les plus jeunes sont nés le jour  $j$ . La relation a exactement une ligne. S'il n'y a aucun élève cette ligne est  $(0, \text{NULL})$ .

Algèbre relationnelle :

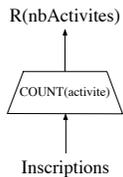


## Agrégation en AR : exemples (3/3)

Quel est le nombre d'activités pratiquées par les élèves ?

- Schéma :  $R(\text{nbActivites} : \text{Nombre})$
- Spécification :  $\langle nb \rangle \in R \iff$  Il y a  $nb$  activités pratiquées par les élèves de l'école. En particulier  $nb = 0$  si aucune activité n'est pratiquée. La relation a exactement une ligne.

Algèbre relationnelle :



## Aggrégation en SQL : exemples (2/3)

Quel est le nombre d'élèves dans l'école et la date de naissance du plus jeune ?

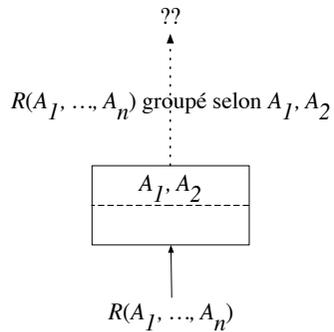
## Aggrégation en SQL : exemples (3/3)

Quel est le nombre d'activités pratiquées par les élèves ?

## Groupements en AR (1/2)

On peut classer (ou **grouper**) les  $n$ -uplets d'une relation selon les valeurs de un ou plusieurs attributs (GROUP BY).

*Grouper les lignes selon la valeur de  $A_1$  et  $A_2$ .*



## Groupements, remarque

Dans la projection il est interdit de projeter selon un attribut qui ne sert pas à grouper (ici un attribut autre que  $A_1$ ) car cela n'a pas de sens : on doit avoir exactement **une ligne pour chaque groupe**, or un attribut qui ne sert pas à grouper pourrait prendre plusieurs valeurs différentes à l'intérieur d'un groupe.

**Tous les champs « non-agrégés » qui apparaissent dans la projection doivent apparaître dans le groupement ! (l'inverse n'est pas vrai)**

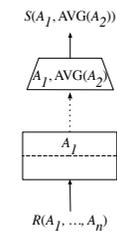
## Groupements en AR (2/2)

Le groupement est une opération « virtuelle » : c'est seulement une étape intermédiaire pour calculer une nouvelle table.

Après avoir formé des groupes, et éventuellement conservé seulement les groupes satisfaisant une certaine condition, on doit utiliser une agrégation (opération « verticale ») **par groupe**.

On obtient une table avec **exactement une ligne pour chaque groupe**.

**Grouper les lignes en fonction de la valeur de  $A_1$  puis dans chaque groupe retourner la valeur commune de  $A_1$  et la moyenne des valeurs de  $A_2$ .**

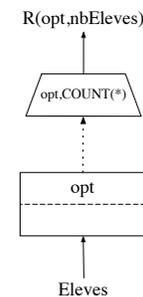


## Groupement en AR : exemples (1/2)

**Quel est le nombre d'élèves dans chaque option ?**

- Schéma :  $R(\underline{\text{opt}} : \text{Nombre}, \text{nbElevés} : \text{Nombre})$
- Spécification :  $\langle o, nb \rangle \in R \iff \text{Il y a } nb \text{ élèves dans l'option } o.$

Algèbre relationnelle :



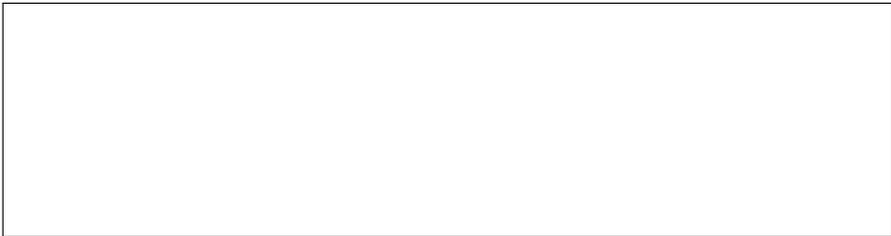
## Groupement en SQL : exemples (1/2)

Quel est le nombre d'élèves dans chaque option ?

```
SELECT opt, COUNT(*) AS nbEleves
FROM Eleves
GROUP BY opt;
```

## Groupement en SQL : exemples (2/2)

Quel est le nombre d'élèves nés en 1989 dans chaque option ?

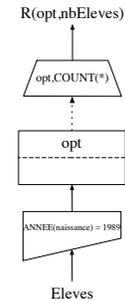


## Groupement en AR : exemples (2/2)

Quel est le nombre d'élèves nés en 1989 dans chaque option ?

- Schéma :  $R(\underline{opt} : \text{Nombre}, nbEleves : \text{Nombre})$
- Spécification :  $\langle o, nb \rangle \in R \iff$  Il y a  $nb$  élèves nés en 1989 dans l'option  $o$ . La relation a une ligne pour chaque option ayant au moins un élève né en 1989, les autres options n'apparaissent pas.

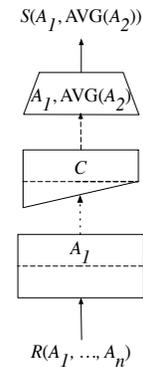
Algèbre relationnelle :



## Conditions sur les groupes

Groupier les  $n$ -uplets d'une relation selon les valeurs de un ou plusieurs attributs (GROUP BY), puis conserver les groupes satisfaisant une certaine condition (HAVING).

Groupier les lignes en fonction de la valeur de  $A_1$  puis conserver les groupes vérifiant la condition  $C$  puis dans chaque groupe retourner la valeur commune de  $A_1$  et la moyenne des valeurs de  $A_2$ .

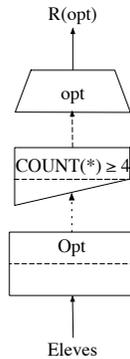


## Conditions sur les groupes en AR : exemples (1/2)

Quels sont les options suivies par au moins 4 élèves ?

- Schéma :  $R(\underline{opt} : \text{Nombre})$
- Spécification :  $\langle o \rangle \in R \iff$  Il y a au moins 4 élèves qui suivent l'option  $o$ .

Algèbre relationnelle :

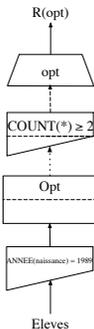


## Conditions sur les groupes en AR : exemples (2/2)

Quelles sont les options suivies par au moins 2 élèves nés en 1989 ?

- Schéma :  $R(\underline{opt} : \text{Nombre})$
- Spécification :  $\langle o \rangle \in R \iff$  Il y a au moins 2 élèves nés en 1989 qui suivent l'option  $o$ .

Algèbre relationnelle :



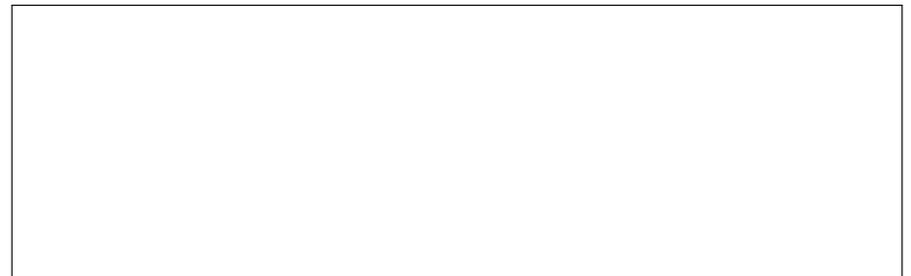
## Conditions sur les groupes en SQL : exemples (1/2)

Quels sont les options suivies par au moins 4 élèves ?

```
SELECT opt FROM Eleves
GROUP BY opt
HAVING COUNT(*) >= 4;
```

## Conditions sur les groupes en SQL : exemples (2/2)

Quelles sont les options suivies par au moins 2 élèves nés en 1989 ?



## Double agrégation

On peut avoir deux agrégations successives, la première portant sur un groupement et la seconde sur l'ensemble de la table obtenue.

Dans ce cas on peut (et c'est conseillé) regrouper les deux agrégations.

## Double agrégation en SQL : exemple

Quel est le nombre moyen d'élèves par option ?



## Double agrégation en AR : exemple

Quel est le nombre moyen d'élèves par option ?

- Schéma :  $R(\text{nbMoyen} : \text{Nombre})$   
( $R$  comporte une seule ligne).
- Spécification :  $\langle nbm \rangle \in R \iff nbm$  est le nombre moyen d'élèves par option.

Algèbre relationnelle :

