

COURS 2

Éléments d 'algèbre relationnelle

Motivation : Interroger des bases de données

Qu'est-ce qu'une requête ?

une expression dans une algèbre dite algèbre relationnelle.

i.e., des opérations sur des tables

- **Filtrer :**

- Sélection (filtre / critères)

- Projection (choix de colonnes)

- Différence (suppression de lignes)

- **Combiner :**

- Produit cartésien (toutes les possibilités)

- Jointure (complément d'information)

- Union (union d'ensembles de lignes)

Quelques définitions

- Ligne

Une ligne L est une séquence de valeurs V_i , $L = \langle V_1, \dots, V_i, \dots, V_n \rangle$

- Extraction

La ligne $\langle V_i \rangle$ construite à partir de L est notée $L(i)$

- Concaténation

Soient deux lignes $L_1 = \langle V_1, \dots, V_n \rangle$ et $L_2 = \langle W_1, \dots, W_m \rangle$,
la concaténation de L_1 et de L_2 , notée $L_1 \bullet L_2$ est :

$$L_1 \bullet L_2 = \langle V_1, \dots, V_n, W_1, \dots, W_m \rangle$$

- Table

Une table est un ensemble de lignes

Sélection (restriction)

T =

c1 (nom)	c2 (âge)	c3 (adresse)	c4 (néA)
Bob	13	Lyon	Nice
Sam	7	Nice	Nice
Cathy	13	Brest	Brest
Julie	20	Lyon	Brest

- Sélection / constante : Personnes habitant Lyon

$\sigma_{c3="Lyon"}(T)$

✓
✓

c1 (nom)	c2 (âge)	c3 (adr)	c4 (néA)
Bob	13	Lyon	Nice
Julie	20	Lyon	Brest

- Sélection / critère inter-colonnes : Personnes nées dans la ville d'habitation

$\sigma_{c3=c4}(T)$

✓
✓

c1 (nom)	c2 (âge)	c3 (adr)	c4 (néA)
Sam	7	Nice	Nice
Cathy	13	Brest	Brest

Sélection (restriction)

- Autre comparateur : Personnes n'ayant pas plus de 14 ans

$$\sigma_{c2 \leq 14}(T)$$

✓
✓
✓

c1 (nom)	c2 (âge)	c3 (adr)	c4 (néA)
Bob	13	Lyon	Nice
Sam	7	Nice	Nice
Cathy	13	Brest	Brest

- Composition de sélections : Habitants de Lyon n'ayant pas plus de 14 ans

$$\sigma_{c3="Lyon"}(\sigma_{c2 \leq 14}(T))$$

✓

c1 (nom)	c2 (âge)	c3 (adr)	c4 (néA)
Bob	13	Lyon	Nice

- Conjonction de critères :

$$\sigma_{c3="Lyon"}(\sigma_{c2 \leq 14}(T)) \Leftrightarrow \sigma_{c3="Lyon" \text{ and } c2 \leq 14}(T)$$

Une formalisation de la sélection (restriction)

- Sélection / constante :

L'ensemble des lignes L de la table T telles que $L(i)=a$ est noté $\sigma_{i=a}(T)$

$$\sigma_{i=a}(T) = \{ L \mid L \text{ dans } T \text{ et } L(i) = a \}$$

- Sélection / inter-colonnes :

L'ensemble des lignes L de la table T telles que $L(i)=L(j)$ est noté $\sigma_{i=j}(T)$

$$\sigma_{i=j}(T) = \{ L \mid L \text{ dans } T \text{ et } L(i) = L(j) \}$$

- Autres comparateurs : $\neq < > \leq \geq$ peuvent être employés à la place de $=$
- Conjonction de comparaisons : Pour alléger les notations des conjonctions de critères de sélection peuvent être employées

Projection

T =

c1 (nom)	c2 (âge)	c3 (adr)	c4 (néA)
Bob	13	Lyon	Nice
Sam	7	Nice	Nice
Cathy	13	Brest	Brest
Julie	20	Lyon	Brest

- Projection sur les colonnes 1 et 2 :

Nom et age de toutes les personnes ✓ $\pi_{c1,c2}(T)$

c1 (nom)	c2 (âge)
Bob	13
Sam	7
Cathy	13
Julie	20

- Projection sur les colonnes 1 et 3

Nom et adresse de toutes les personnes

✓ $\pi_{c1,c3}(T)$

(renumérotation des colonnes)

c1 (nom)	c2 (adr)
Bob	Lyon
Sam	Nice
Cathy	Brest
Julie	Lyon

Composition (Combinaison)

T =

c1 (nom)	c2 (âge)	c3 (adr)	c4 (néA)
Bob	13	Lyon	Nice
Sam	7	Nice	Nice
Cathy	13	Brest	Brest
Julie	20	Lyon	Brest

- Quels sont les noms des personnes habitant Lyon ?

1) Sélection :

$\sigma_{c3="Lyon"}(T)$

c1 (nom)	c2 (âge)	c3 (adr)	c4 (néA)
Bob	13	Lyon	Nice
Julie	20	Lyon	Brest

2) ... suivie d'une projection :

$\pi_{c1}(\sigma_{c3="Lyon"}(T))$

✓
✓

c1 (nom)
Bob
Julie

Une formalisation de la projection

- Projection sur colonnes :

L'ensemble des lignes de la table T obtenues en ne conservant que les colonnes de numéros i_1, i_2, \dots, i_k est noté $\pi_{i_1, i_2, \dots, i_k}(T)$

$$\pi_{i_1, i_2, \dots, i_k}(T) = \{ L(i_1)L(i_2) \dots L(i_k) \mid L \text{ dans } T \}$$

- Retour sur la composition : le *principe de clôture*

✓ **Le résultat d'une opération portant sur des tables est aussi une table**
⇒ composition possible de différentes opérations

e.g., des sélections et des projections

NB. Propriété fondamentale pour expliquer le confort des interrogations

Produit cartésien

homme=

c1 (nom)	c2 (adr)
Bob	Lyon
Sam	Nice

femme=

c1 (nom)	c2 (adr)
Cathy	Brest
Julie	Lyon
Linda	Lyon

- Quels sont tous les couples homme-femme ?

homme \otimes femme

à partir de 'Bob' {

à partir de 'Sam' {

c1 (nom)	c2 (adr)	c3 (nom)	c4 (adr)
Bob	Lyon	Cathy	Brest
Bob	Lyon	Julie	Lyon
Bob	Lyon	Linda	Lyon
Sam	Nice	Cathy	Brest
Sam	Nice	Julie	Lyon
Sam	Nice	Linda	Lyon

- Quels sont les couples homme-femme d'une même ville ?

$\sigma_{c2 = c4}$ (homme \otimes femme)

c1 (nom)	c2 (adr)	c3 (nom)	c4 (adr)
Bob	Lyon	Julie	Lyon
Bob	Lyon	Linda	Lyon

Produit cartésien

homme=	c1 (nom)	c2 (adr)	ville=	c1 (nomV)	c2 (nbHab)	c3 (départ)
	Bob	Lyon		Nice	340.000	Alp-Mar
	Sam	Nice		Brest	160.000	Finistère
				Lyon	420.000	Rhône

- Dans quels départements habitent les personnes de la table homme ?

1) Combinaison par produit cartésien

(homme \otimes ville)		c1 (nom)	c2 (adr)	c3 (nomV)	c4 (nbHab)	c5 (départ)
à partir de 'Bob' {	{	Bob	Lyon	Nice	340.000	Alp-Mar
		Bob	Lyon	Brest	160.000	Finistère
		Bob	Lyon	Lyon	420.000	Rhône
à partir de 'Sam' {	{	Sam	Nice	Nice	340.000	Alp-Mar
		Sam	Nice	Brest	160.000	Finistère
		Sam	Nice	Lyon	420.000	Rhône

Produit cartésien

2) ... suivi d'une sélection :

$\sigma_{c2 = c3}$ (homme \otimes ville)

c1 (nom)	c2 (adr)	c3 (nomV)	c4 (nbHab)	c5 (départ)
Bob	Lyon	Lyon	420.000	Rhône
Sam	Nice	Nice	340.000	Alp-Mar

Une formalisation du produit cartésien

L'ensemble des lignes pouvant être obtenues par concaténation d'une ligne d'une table T1 avec une ligne d'une table T2 est noté $T1 \otimes T2$

$$T1 \otimes T2 = \{ L1L2 \mid L1 \text{ dans } T1 \text{ et } L2 \text{ dans } T2 \}$$

Le produit $T1 \otimes T2$ est souvent suivi d'une sélection dans laquelle intervient une colonne issue de T1 et une colonne issue de T2

⇒ définition d'un nouvel opérateur : **la jointure**

Jointure

homme=

c1 (nom)	c2 (adr)
Bob	Lyon
Sam	Nice

femme=

c1 (nom)	c2 (adr)
Cathy	Brest
Julie	Lyon
Linda	Lyon

- Couples homme-femme d'une même ville ?

$$\text{homme} \bowtie_{c2=c2} \text{femme} \iff \sigma_{c2=c4} (\text{homme} \otimes \text{femme})$$

✓

✓

✓

c1 (nom)	c2 (adr)	c3 (nom)	c4 (adr)
Bob	Lyon	Julie	Lyon
Bob	Lyon	Linda	Lyon

Jointure

homme=

c1 (nom)	c2 (adr)
Bob	Lyon
Sam	Nice

ville=

c1 (nomV)	c2 (nbHab)	c3 (départ)
Nice	340.000	Alp-Mar
Brest	160.000	Finistère
Lyon	420.000	Rhône

- Dans quels départements habitent les hommes ?

$$\text{homme} \bowtie_{c2=c1} \text{ville} \quad \Leftrightarrow \quad \sigma_{c2=c3}(\text{homme} \otimes \text{ville})$$

c1 (nom)	c2 (adr)	c3 (nomV)	c4 (nbHab)	c5 (départ)
Bob	Lyon	Lyon	420.000	Rhône
Sam	Nice	Nice	340.000	Alp-Mar

Une formalisation de la jointure

{combinaison par produit cartésien et sélection}

L'ensemble des lignes L pouvant être obtenues par concaténation d'une ligne $L1$ d'une table $T1$ avec une ligne $L2$ d'une table $T2$ telle que $L1(i) = L2(j)$ est noté :

$$T1 \bowtie_{i=j} T2$$

$$T1 \bowtie_{i=j} T2 = \{ L1 \bullet L2 \mid L1 \text{ dans } T1 \text{ et } L2 \text{ dans } T2 \\ \text{et } L1(i)=L2(j) \}$$

Cette opération est appelée **jointure de $T1$ avec $T2$**

Une formalisation de la jointure

- **Lien avec le produit cartésien et la sélection**

Soit m le nombre de colonnes de $T1$,

$T1 \bowtie_{i=j} T2$ est identique à $\sigma_{i=(j+m)} (T1 \otimes T2)$

Remarque : comme pour la sélection, les comparateurs $\neq < > \leq \geq$ ainsi que les conjonctions de critères peuvent être employés

Union

homme=

c1 (nom)	c2 (adr)
Bob	Lyon
Sam	Nice

femme=

c1 (nom)	c2 (adr)
Cathy	Brest
Julie	Lyon
Linda	Lyon

- Quel est l'ensemble des personnes ?

homme \cup femme

c1 (nom)	c2 (adr)
Bob	Lyon
Sam	Nice
Cathy	Brest
Julie	Lyon
Linda	Lyon

- Quel est l'ensemble des personnes habitant Lyon ?

$\sigma_{c2="Lyon"}(\text{homme} \cup \text{femme})$

c1 (nom)	c2 (adr)
Bob	Lyon
Julie	Lyon
Linda	Lyon

Différence

homme=

c1 (nom)	c2 (adr)
Bob	Lyon
Sam	Nice

femme=

c1 (nom)	c2 (adr)
Cathy	Brest
Julie	Lyon
Linda	Lyon

- Quelles sont les villes pour lesquelles on connaît au moins un homme et aucune femme ?

$$\pi_{c_2}(\text{homme}) - \pi_{c_2}(\text{femme})$$

c1 (adr)
Nice

Formalisation des deux opérations

◆ Une formalisation de l'union (ensembliste)

Soient T1 et T2 deux tables ayant même nombre de colonnes, l'ensemble contenant les lignes de T1 ainsi que celles de T2 est noté $T1 \cup T2$

$$T1 \cup T2 = \{ L \mid L \text{ dans } T1 \text{ ou } L \text{ dans } T2 \}$$

◆ Une formalisation de la différence (ensembliste)

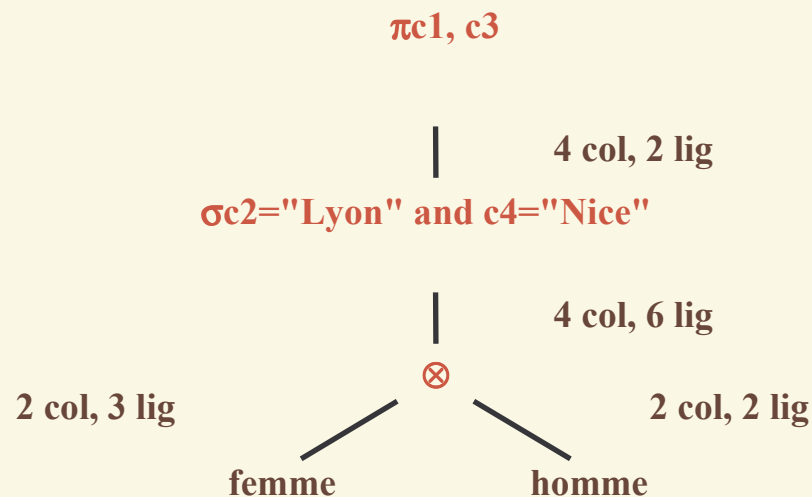
Soient T1 et T2 deux tables ayant même nombre de colonnes, l'ensemble contenant les lignes de T1 ne se trouvant pas dans T2 est noté $T1 - T2$

$$T1 - T2 = \{ L \mid L \text{ dans } T1 \text{ et } L \text{ n'est pas dans } T2 \}$$

Arbre et transformation de requêtes

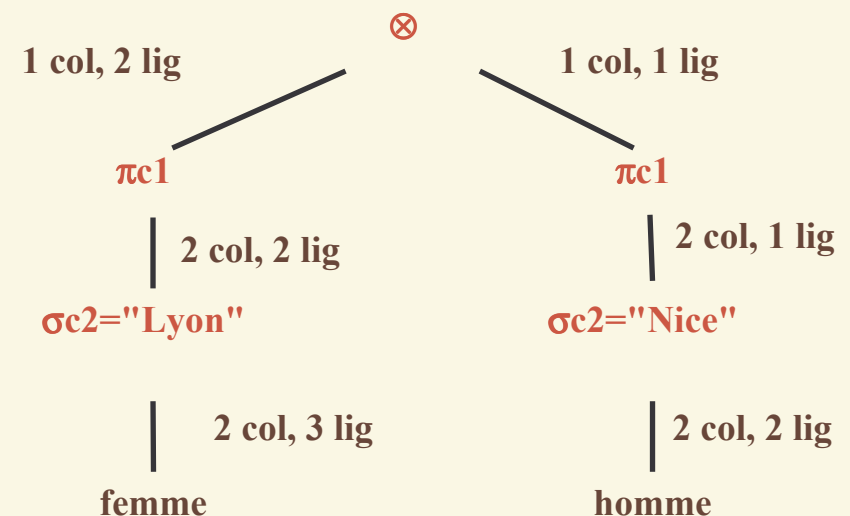
- Quels sont les couples possibles avec une femme de Lyon et un homme de Nice ?

c1	c2
Julie	Sam
Linda	Sam



$$2*3+2*2+4*6+4*2 =$$

42 cases manipulées



$$2*3+2*2+2*2+2*1+1*2+1*1 =$$

19 cases manipulées