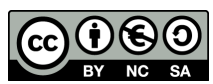
BDW - Algèbre Relationnelle

Fabien Duchateau

fabien.duchateau [at] univ-lyon1.fr

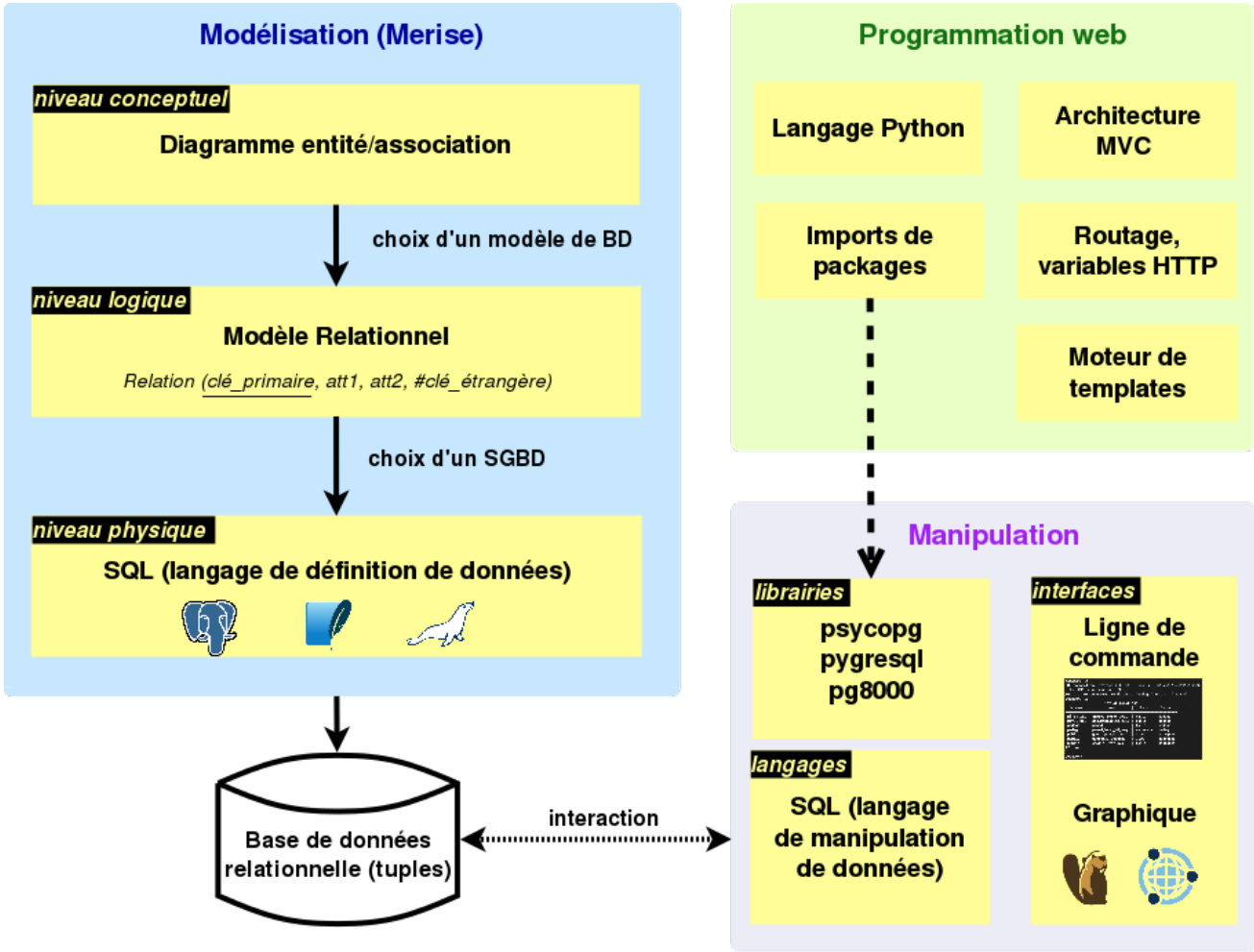
Université Claude Bernard Lyon 1

2024 - 2025



https://perso.liris.cnrs.fr/fabien.duchateau/BDW/

Positionnement dans BDW



Ces diapositives utilisent le genre féminin (e.g., chercheuse, développeuses) plutôt que l'écriture inclusive (moins accessible, moins concise, et pas totalement inclusive)

Langages de manipulation de données

Pourquoi des langages de manipulation "théoriques" ?

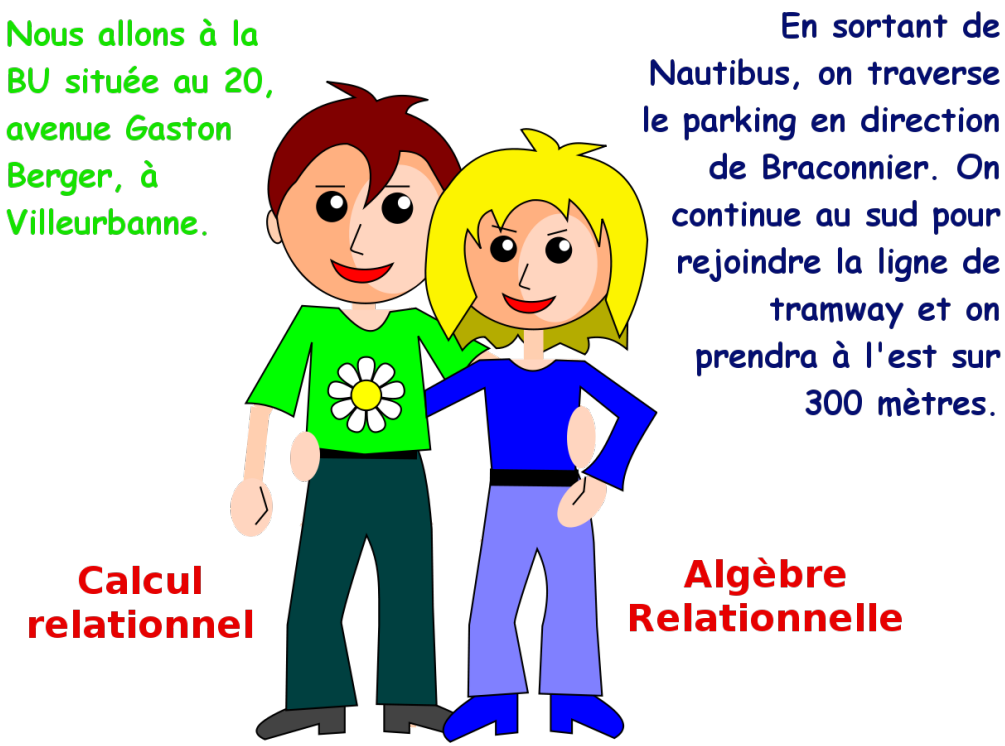
- Étudier le modèle relationnel
- Prouver des propriétés
- Obtenir une mise en œuvre efficace
- Servir de base à des langages plus conviviaux pour les utilisateurs et utilisatrices

Deux langages à la puissance d'expression équivalente :

- Algèbre relationnelle (AR)
- Calculs relationnels (CR)

Langages de manipulation de données (2)

- Calcul relationnel : décrit le résultat à obtenir (langage déclaratif)
- Algèbre relationnelle : exprime comment obtenir le résultat (langage procédural)



http://openclipart.org

Plan

L'algèbre relationnelle

Opérateurs ensemblistes

Opérateurs BD

Requêtes en AR

Intérêt de l'algèbre relationnelle (AR)

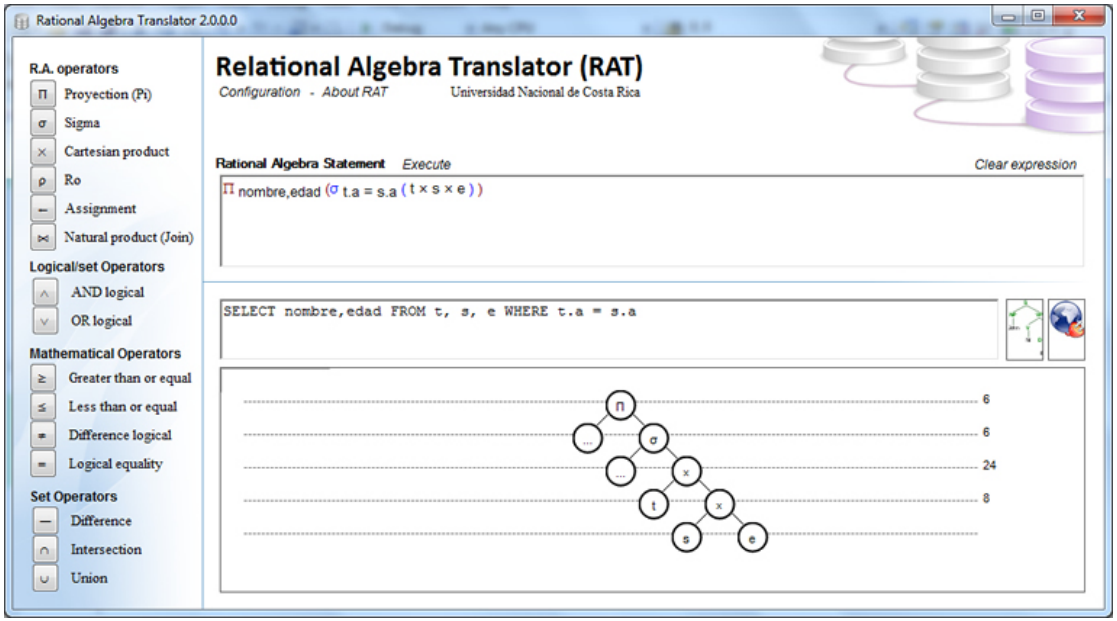
- Proposée par E. Codd, 1970
- Identification des opérateurs fondamentaux pour l'utilisation d'une base de données relationnelle
- Traduction d'une requête déclarative en opérateurs pour récupérer les données (comment le résultat est calculé par le SGBD)
- Définition des principales fonctions à optimiser dans un SGBD relationnel

---

Codd, Edgar. A relational model of data for large shared data banks. Communications of the ACM (1970)

Quelques outils basés sur l'AR

- Relational Algebra Translator (RAT), qui convertit de l'algèbre au SQL
- ISBL (IBM)
- Relational
- Pireal
- RAdb



---

http://en.wikipedia.org/wiki/ISBL
http://ltworf.github.io/relational/
http://www.slinfo.una.ac.cr/rat/rat.html
http://centaurialpha.github.io/pireal/
http://users.cs.duke.edu/~junyang/radb/

Principe de l'AR

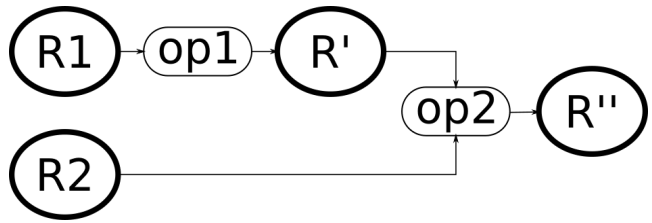
- Algèbre : ensemble d'opérateurs qui manipulent des expressions et produisent une expression
- En BD, une expression est une relation (i.e., un ensemble de n-uplets)
- L'algèbre relationnelle consiste à combiner des relations à travers des opérateurs pour formuler une requête

$$\text{op_unaire}(Exp) \rightarrow Exp'$$
$$\text{op_binaire}(Exp_1, Exp_2) \rightarrow Exp''$$

https://en.wikipedia.org/wiki/Relational_algebra

Principe de l'AR (2)

- Le résultat d'une opération est une nouvelle relation temporaire (durée de vie limitée, détruite à la fin du programme ou de la transaction qui l'a créée)
- La relation résultat a les mêmes caractéristiques qu'une relation de la base et peut être manipulée de nouveau par les opérateurs de l'algèbre $\Rightarrow$ langage fermé



Soient R_1 et R_2 deux relations et R' une relation résultat :

$$\text{op_binaire}(R_1, \text{op_unaire}(R_2)) \rightarrow R'$$

Opérateurs de l'AR

Opérateurs fondamentaux :

- **sélection**
- **projection**
- **union**
- **différence**
- **produit cartésien**

Opérateur syntaxique (modifie le schéma et pas les n-uplets) :

- **renommage**

Opérateurs déduits (en composant les précédents) :

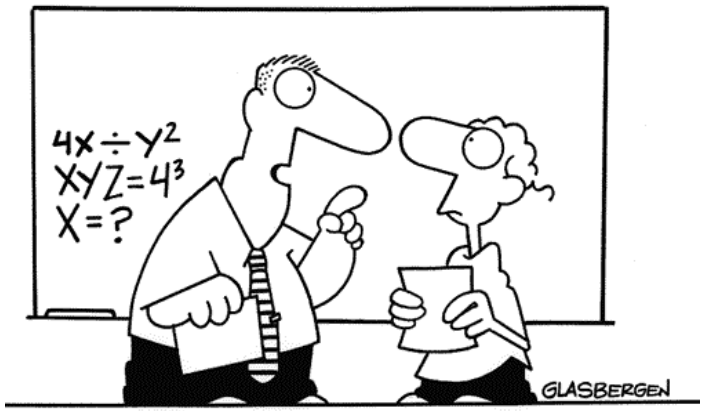
- **intersection**
- **jointures** (semi, naturelle, thêta)
- **division**

Opérateurs de l'AR (2)

On peut regrouper les opérateurs en deux catégories :

ensemblistes : union, intersection, différence, produit

spécifiques BD : sélection, projection, jointures, division, renommage



"Algebra class will be important to you later in life because there's going to be a test six weeks from now."

Plan

L'algèbre relationnelle

Opérateurs ensemblistes

Opérateurs BD

Requêtes en AR

Union $\cup$

- Soient R et S deux relations
- $R \cup S$ crée une relation comprenant tous les n-uplets existants dans l'une ou l'autre des relations R et S
- Les 2 relations doivent avoir le même nombre d'attributs, et les mêmes types (*i.e.* même domaine)
- Par convention, la relation résultat possède le schéma du premier opérande (relation R)
- Élimination des doublons

Exemple d'union

nomU	ville	effectif
INSA	Lyon	36000
UCB	Lyon	15000
UJF	Grenoble	10000
UJM	Saint-Étienne	21000

Table UNIVERSITÉ

nomU2	ville2	effectif2
UPMC	Paris	32000
UCBN	Caen	24000
UCB	Lyon	15000

Table UNIV2

Résultat de UNIVERSITÉ ∪ UNIV2 ?

nomU	ville	effectif
INSA	Lyon	36000
UCB	Lyon	15000
UJF	Grenoble	10000
UJM	Saint-Etienne	21000
UPMC	Paris	32000
UCBN	Caen	24000

Différence – ou \

- Soient R et S deux relations
- R \ S crée une relation de même schéma et de population égale à l'ensemble des n-uplets de R moins ceux de S, i.e., les n-uplets qui se trouvent dans R mais pas dans S
- Les 2 relations doivent avoir le même nombre d'attributs, et les mêmes types (*i.e.* même domaine)
- Par convention, la relation résultat possède le schéma du premier opérande (relation R)

Exemple de différence

nomU	ville	effectif
INSA	Lyon	36000
UCB	Lyon	15000
UJF	Grenoble	10000
UJM	Saint-Étienne	21000

Table UNIVERSITÉ

nomU2	ville2	effectif2
UPMC	Paris	32000
UCBN	Caen	24000
UCB	Lyon	15000

Table UNIV2

Résultat de UNIVERSITÉ \ UNIV2 ?

nomU	ville	effectif
INSA	Lyon	36000
UJF	Grenoble	10000
UJM	Saint-Etienne	21000

Intersection ∩

- Soient R et S deux relations
- R ∩ S crée une nouvelle relation de même schéma et de population égale à l'ensemble des n-uplets de R tels qu'il existe un n-uplet de même valeur dans S
- Les 2 relations doivent avoir le même nombre d'attributs, et les mêmes types (*i.e.* même domaine)
- Par convention, la relation résultat possède le schéma du premier opérande (relation R)

Exemple d'intersection

nomU	ville	effectif
INSA	Lyon	36000
UCB	Lyon	15000
UJF	Grenoble	10000
UJM	Saint-Étienne	21000

Table UNIVERSITÉ

nomU2	ville2	effectif2
UPMC	Paris	32000
UCBN	Caen	24000
UCB	Lyon	15000

Table UNIV2

Résultat de UNIVERSITÉ ∩ UNIV2 ?

nomU	ville	effectif
UCB	Lyon	15000

Produit cartésien $\times$

- Soient R et S deux relations
- R $\times$ S crée une nouvelle relation où chaque n-uplet de R est associé à chaque n-uplet de S
- Le nombre de n-uplets est égal à $\mid R \mid \times \mid S \mid$ (où $\mid R \mid$ est le nombre de n-uplets dans la relation R)
- Intérêt principalement par rapport à la jointure

Exemple de produit cartésien

nomU	ville	effectif
INSA	Lyon	36000
UCB	Lyon	15000
UJF	Grenoble	10000

Table UNIVERSITÉ (réduite)

nomU2	ville2	effectif2
UPMC	Paris	32000
UCB	Lyon	15000

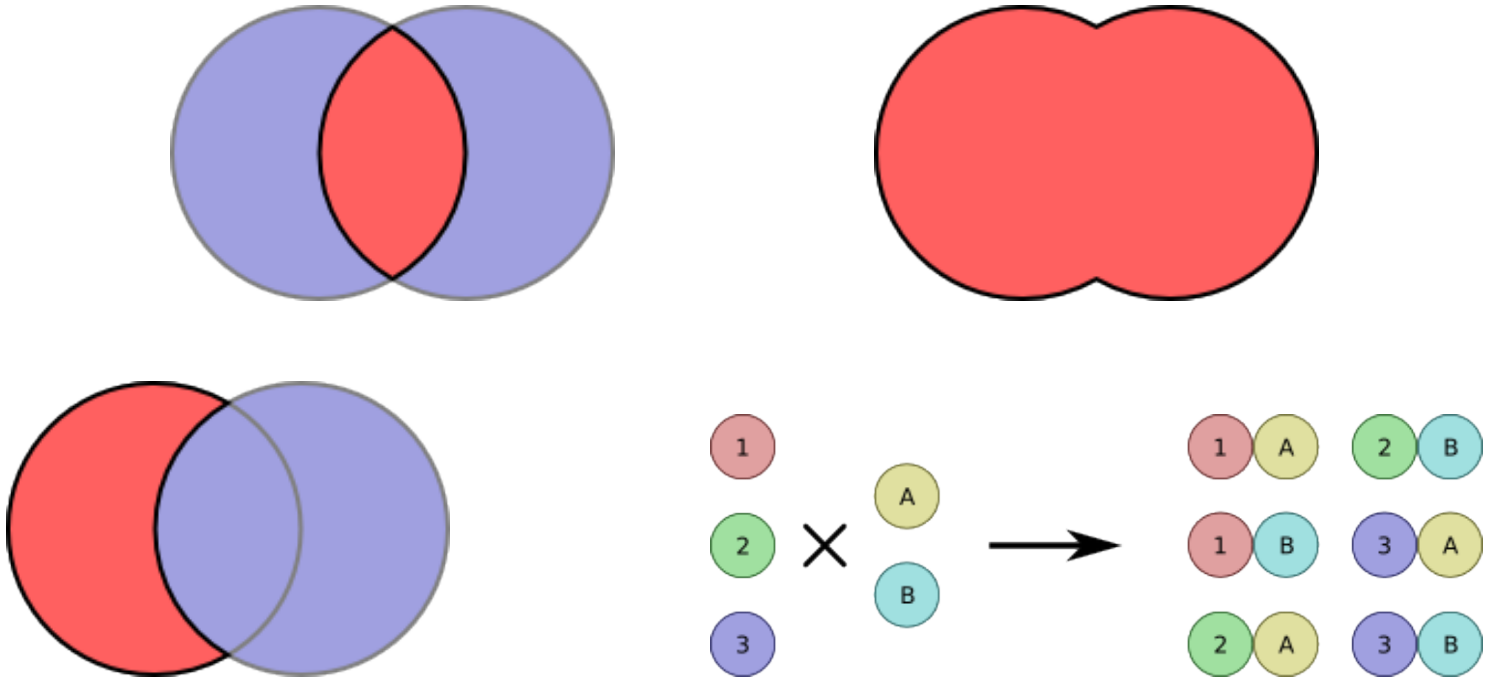
Table UNIV2 (réduite)

Résultat de UNIVERSITÉ × UNIV2 ?

nomU	ville	effectif	nomU2	ville2	effectif2
INSA	Lyon	36000	UPMC	Paris	32000
INSA	Lyon	36000	UCB	Lyon	15000
UCB	Lyon	15000	UPMC	Paris	32000
UCB	Lyon	15000	UCB	Lyon	15000
UJF	Grenoble	10000	UPMC	Paris	32000
UJF	Grenoble	10000	UCB	Lyon	15000

En résumé...

Quatre opérateurs ensemblistes binaires



Que représente $R \setminus (R \setminus S)$?

---

https://en.wikibooks.org/wiki/Probability/Mathematical_Review

Plan

L'algèbre relationnelle

Opérateurs ensemblistes

Opérateurs BD

Requêtes en AR

Renommage ρ

- Soit R une relation
- Renommage d'un ou plusieurs attributs d'une relation R, avec A_1 qui devient A'_1, ..., A_k qui devient A'_k :

$$\rho_{A'_1/A_1,\ldots,A'_k/A_k}(R)$$

- Utilisé en cas d'homonymie ou avant certaines opérations ensemblistes
- Il existe une autre notation qui permet en plus de renommer la relation R en R' :

$$\rho_{R'(A_1\to A'_1,\ldots,A_k\to A'_k)}(R)$$

Exemple de renommage

nomU	ville	effectif
INSA	Lyon	36000
UCB	Lyon	15000
UJF	Grenoble	10000
UJM	Saint-Étienne	21000

Table UNIVERSITÉ

Résultat de $\rho_{\text{nomUniversité/nomU,villeUniversité/ville}}$ (UNIVERSITÉ) ?

nomUniversité	villeUniversité	effectif
INSA	Lyon	36000
UCB	Lyon	15000
UJF	Grenoble	10000
UJM	Saint-Etienne	21000

Projection Π

- Soit R une relation
- $\Pi_{A_1,\ldots,A_k}(R)$ ne garde que les attributs $A_1, \ldots, A_k$ de la relation R
- Projection $\Rightarrow$ filtre sur les colonnes
- Élimination des n-uplets doublons (il ne peut y avoir deux fois le même élément dans un ensemble)

Exemple de projection

nomU	ville	effectif
INSA	Lyon	36000
UCB	Lyon	15000
UJF	Grenoble	10000
UJM	Saint-Étienne	21000

Table UNIVERSITÉ

Résultat de $\Pi_{\text{nomU, effectif}}$ (UNIVERSITÉ) ?

nomU	effectif
INSA	36000
UCB	15000
UJF	10000
UJM	21000

Sélection σ

- Soit R une relation
- $\sigma_C(R)$ sélectionne les n-uplets de R qui vérifient la condition C
- Sélection $\Rightarrow$ filtre sur les lignes
- Condition $=$ combinaison (via les connecteurs $\vee, \wedge, \neg$) de comparaisons (via les opérations $=, <, >, \leq, \geq$)
  - entre deux attributs
  - ou entre un attribut et une constante

Exemple de condition : effectif $\leq 20000 \wedge$ ville $=$ 'Lyon'

Exemple de sélection

nomU	ville	effectif
INSA	Lyon	36000
UCB	Lyon	15000
UJF	Grenoble	10000
UJM	Saint-Étienne	21000

Table UNIVERSITÉ

Résultat de $\sigma_{\text{effectif} > 20000}$ (UNIVERSITÉ) ?

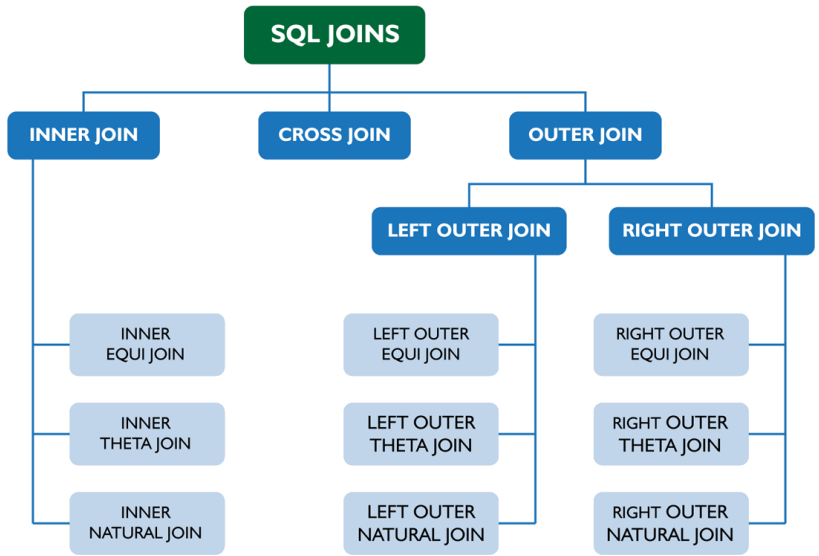
nomU	ville	effectif
INSA	Lyon	36000
UJM	Saint-Etienne	21000

Jointures

Une jointure permet de lier deux relations selon une condition (de jointure) entre leurs attributs

Classification des jointures selon :

- La condition $=$ jointure naturelle, équi-jointure, θ-jointure
- Les n-uplets conservés dans le résultat $=$ jointure interne, anti-jointure, jointures externes
- Les attributs conservés dans le résultat $=$ semi-jointure



Jointure naturelle $\bowtie$

- Soient R et S des relations qui ont les attributs $A_1, \ldots, A_k$ en commun
- La jointure naturelle $R \bowtie S$ produit l'ensemble des n-uplets constitués à partir de n-uplets n_1 de R et de n-uplets n_2 de S ayant les mêmes valeurs pour les attributs $A_1, \ldots, A_k$
- Les n-uplets obtenus sont construits comme suit :
  - ajout au n-uplet n_1 la valeur des attributs du n-uplet n_2 (uniquement pour les attributs qui ne sont pas dans R)

Exemple de jointure naturelle

nomU	ville	effectif
INSA	Lyon	36000
UCB	Lyon	15000
UJF	Grenoble	10000
UJM	Saint-Étienne	21000

Table Université

ville	département
Paris	75
Lyon	69
Grenoble	38

Table VilleDept

Résultat de Université ⋈ VilleDept ?

nomU	ville	effectif	département
INSA	Lyon	36000	69
UCB	Lyon	15000	69
UJF	Grenoble	10000	38

θ-jointure $\bowtie_C$

- Soient R et S deux relations
- La θ-jointure $R \bowtie_C S$ lie les relations R et S selon la condition de jointure C
- Utile lorsqu'il n'y a pas d'attributs en commun entre R et S ou pour des comparaisons autres que l'égalité
- L'équi-jointure est simplement une θ-jointure avec un comparateur d'égalité

Exemple de θ-jointure

nomU	ville	effectif
INSA	Lyon	36000
UCB	Lyon	15000
UJF	Grenoble	10000
UJM	Saint-Étienne	21000

Table UNIVERSITÉ

nomVille	département
Paris	75
Lyon	69
Grenoble	38

Table VILLEDEPT2

Résultat de UNIVERSITÉ $\bowtie_{ville=nomVille}$ VILLEDEPT2 ?

nomU	ville	effectif	nomVille	département
INSA	Lyon	36000	Lyon	69
UCB	Lyon	15000	Lyon	69
UJF	Grenoble	10000	Grenoble	38

Exemple de θ-jointure (2)

nomU	ville	effectif
INSA	Lyon	36000
UCB	Lyon	15000
UJF	Grenoble	10000
UJM	Saint-Étienne	21000

Table UNIVERSITÉ

nomVille	département
Paris	75
Lyon	69
Grenoble	38

Table VILLEDEPT2

Résultat de UNIVERSITÉ $\bowtie_{effectif < departement}$ VILLEDEPT2 ?

nomU	ville	effectif	nomVille	département

Remarques :

- On obtient un résultat vide
- La requête en soi n'a pas vraiment de sens...

Semi-jointure $\ltimes_C$ ou $\rtimes_C$

- Soient R et S deux relations

- La semi-jointure est similaire à la θ-jointure, mais seuls les attributs de la relation gauche ($\ltimes$) ou de la relation droite ($\rtimes$) sont conservés dans la relation résultat

Exemple de semi-jointure

nomU	ville	effectif
INSA	Lyon	36000
UCB	Lyon	15000
UJF	Grenoble	10000
UJM	Saint-Étienne	21000

Table UNIVERSITÉ

nomVille	département
Paris	75
Lyon	69
Grenoble	38

Table VILLEDEPT2

Résultat de UNIVERSITÉ $\ltimes_{ville=nomVille}$ VILLEDEPT2 ?

nomU	ville	effectif
INSA	Lyon	36000
UCB	Lyon	15000
UJF	Grenoble	10000

Division $\div$

- Soit la relation R dont les attributs sont $A_1, \ldots, A_k, A_{k+1}, \ldots, A_{k+n}$
- Soit la relation S dont les attributs sont $A_{k+1}, \ldots, A_{k+n}$ (le schéma de S contient un sous-ensemble des attributs de R)
- Alors la division $R \div S$ retourne une relation $T = A_1, \ldots, A_k$
- La division revient à donner "tous les n-uplets r de R tels que pour tout n-uplet s de S, (r, s) est un tuple de R"
  - $\forall s \in S, (r, s) \in R$

Exemple de division

nomU	ville	dept
UCB	Lyon	69
ENS	Lyon	69
ENS	Cachan	94
UJM	Saint-Etienne	63

Table Univ3

ville	dept
Lyon	69

Table VilleDept3

Résultat de Univ3 ÷ VilleDept3 ?

nomU
UCB
ENS

Exemple de division (2)

nomU	ville	dept
UCB	Lyon	69
ENS	Lyon	69
ENS	Cachan	94
UJM	Saint-Etienne	63

Table UNIV3

ville	dept
Lyon	69
Cachan	94

Table VILLEDEPT4

Résultat de UNIV3 ÷ VILLEDEPT4 ?

nomU
ENS

En résumé...

Cinq opérateurs spécifiques aux bases de données : renommage, projection, sélection, jointures, division

Comment écrire une θ-jointure avec les opérateurs fondamentaux ?

Comment écrire une semi-jointure avec les opérateurs fondamentaux ?

Plan

L'algèbre relationnelle

Opérateurs ensemblistes

Opérateurs BD

Requêtes en AR

Requête algébrique

Une requête en AR est une combinaison d'opérateurs

Soient deux relations `R(a, b)` et `S(a, c)`, quelques requêtes :

- La liste des *a* de la relation R

$$\Pi_a(R)$$

- Les *b* de la relation R pour les n-uplets dont le *a* vaut *'abcde'*

$$\Pi_b(\sigma_{a='abcde'}(R))$$

- Les *b* et *c* des n-uplets de R et S qui ont la même valeur pour *a*

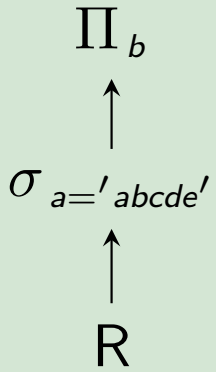
$$\Pi_{b,c}(R \bowtie_a S)$$

Arbre algébrique

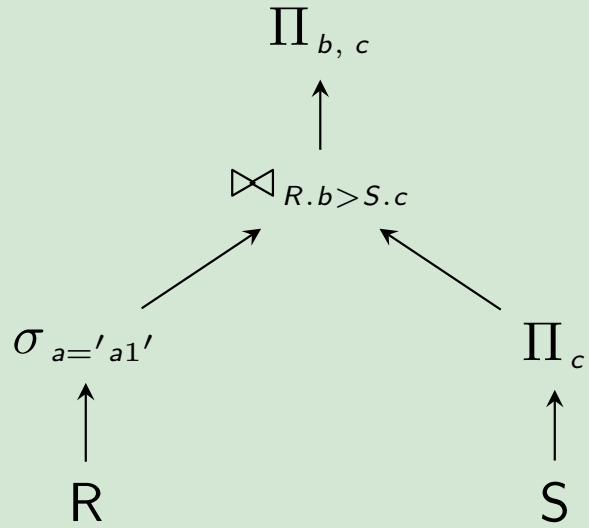
Possibilité de représenter une requête AR sous forme d'arbre :

- Les feuilles (en bas) sont les relations
- La racine (en haut) est le résultat
- Les noeuds sont des opérations

$\Pi_b(\sigma_{a='abcde'}(R))$



$\Pi_{b,c}((\sigma_{a='a1'}(R)) \bowtie_{R.b>S.c} (\Pi_c(S)))$



Exercice

idE	nomE	moyenneLycée	effectifLycée
123	Ana	19.5	1000
234	Bob	18	1500
345	Chloé	17.5	500
456	Damien	19.5	1000
543	Chloé	17	2000
567	Éléonore	14.5	2000
654	Ana	19.5	1000
678	Farid	19	200
765	Joana	14.5	1500
789	Gisèle	17	800
876	Irène	19.5	400
898	Hector	18.5	800

Table Élève

nomU	ville	effectif
INSA	Lyon	36000
UCB	Lyon	15000
UJF	Grenoble	10000
UJM	Saint-Étienne	21000

Table Université

idE	nomU	département	décision
123	INSA	informatique	O
123	UCB	électronique	N
123	UCB	informatique	O
123	UJM	électronique	O
234	INSA	biologie	N
345	UJF	bioinformatique	O
345	UJM	bioinformatique	N
345	UJM	électronique	N
345	UJM	informatique	O
543	UJF	informatique	N
678	UCB	histoire	O
765	UCB	histoire	O
765	UJM	histoire	N
765	UJM	psychologie	O
876	UCB	informatique	N
876	UJF	biologie	O
876	UJF	biologie marine	N
898	INSA	informatique	O
898	UCB	informatique	O

Table Candidature

Le nom et l'identifiant des élèves qui ont candidaté dans une université de Grenoble

$$\Pi_{idE,nomE}((\sigma_{ville='Grenoble'}(\text{Université})) \bowtie \text{Candidature} \bowtie \text{Élève})$$

idE	nomE
345	Chloé
543	Chloé
876	Irène

Exercice (2)

idE	nomE	moyenneLycée	effectifLycée
123	Ana	19.5	1000
234	Bob	18	1500
345	Chloé	17.5	500
456	Damien	19.5	1000
543	Chloé	17	2000
567	Éléonore	14.5	2000
654	Ana	19.5	1000
678	Farid	19	200
765	Joana	14.5	1500
789	Gisèle	17	800
876	Irène	19.5	400
898	Hector	18.5	800

Table ÉLÈVE

nomU	ville	effectif
INSA	Lyon	36000
UCB	Lyon	15000
UJF	Grenoble	10000
UJM	Saint-Étienne	21000

Table UNIVERSITÉ

idE	nomU	département	décision
123	INSA	informatique	O
123	UCB	électronique	N
123	UCB	informatique	O
123	UJM	électronique	O
234	INSA	biologie	N
345	UJF	bioinformatique	O
345	UJM	bioinformatique	N
345	UJM	électronique	N
345	UJM	informatique	O
543	UJF	informatique	N
678	UCB	histoire	O
765	UCB	histoire	O
765	UJM	histoire	N
765	UJM	psychologie	O
876	UCB	informatique	N
876	UJF	biologie	O
876	UJF	biologie marine	N
898	INSA	informatique	O
898	UCB	informatique	O

Table CANDIDATURE

Le nom et l'identifiant des élèves qui ont une moyenne inférieure à celle de Gisèle (idE = 789)

$$\Pi_{idE,nomE}(\text{ÉLÈVE} \bowtie_{moyenneLycee<moy789} (\rho_{moy789/moyenneLycee}(\Pi_{moyenneLycee}(\sigma_{idE=789}(\text{ÉLÈVE})))))$$

idE	nomE
567	Éléonore
765	Joana

En résumé...

Algèbre relationnelle :

- Une dizaine d'opérateurs (fréquents)
- Langage fermé (relation(s) en entrée, relation en sortie), ce qui permet de combiner les opérateurs
- Représentation sous forme d'arbre (optimisation)

Inconvénient de l'AR : l'écriture d'une requête nécessite de connaître les opérateurs et leur signification