Bases de données (CS443)

#4, Modèle relationnel: algèbre relationnelle

Arthur Baudet

Grenoble INP - Esisar

2023–2024

Sources

- équipe BD de Lyon1
- équipe BD de Lille1

Avec leur autorisation, of course.

Le modèle relationnel

Rappel, on a une structuration en tables.

Étudiants	$NUMETUD$	$NOMETUD$	$PRENOMETUD$	AGE	$FORMATION$
	28	Codd	Edgar	20	3
	32	Armstrong	William	20	4
	53	Fagin	Ronald	19	3
	107	Bunneman	Peter	18	3

Enseignants	$NUMENS$	$NOMENS$	$PRENOMENS$	$GRADE$
	5050	Tarjan	Robert	PR
	2123	Mannila	Heikki	MCF
	3434	Papadimitriou	Spiros	PR
	1470	Bagan	Guillaume	CR

Encadre	$NUMENS$	$NUMETUD$	$DATE$
	5050	53	2005
	3434	28	2020
	5050	28	2015
	2123	32	2019

Maintenant, on veut **manipuler ces tables**, ie faire des **requêtes**.

Définition

Une requête est :

- une expression ensembliste qui calcule une relation (en algèbre relationnelle)
- une instruction (**commande**) qui retourne une table (en SQL).

Plan

Formalisme !

Attention formalisme

Mais ce n'est rien qu'une formalisation des opérations sur les tables !

Dans le modèle relationnel, les tables sont supposées construites.

Algèbre relationnelle : opérations sur les relations

Soient des relations r_i définies sur des schémas R_i :

- La projection $\pi_X(r)$ ne conserve que les attributs de X ($\subseteq R$) ;
- La sélection $\sigma_C(r)$ filtre les tuples de r (lignes) suivant la condition C.
- Le renommage $\rho[X/X']$ renomme l'attribut X' en X.
- La jointure $r_1 \bowtie r_2$ "combine" les tuples de r_1 et r_2.

On fera attention aux domaines de définition.

Projection $\pi_X(r)$

employe

NoEmp	Nom	Année	NoDep
1045	Dupond	1978	03
2067	Dupont	1965	06
0456	Martin	1981	03
0278	Martin	1987	05
0789	Blanc	1981	06

$annee\_naissance = \pi_{Nom,Annee}(employe)$ retourne (vaut) la relation :

Projection $\pi_X(r)$

employe

NoEmp	Nom	Année	NoDep
1045	Dupond	1978	03
2067	Dupont	1965	06
0456	Martin	1981	03
0278	Martin	1987	05
0789	Blanc	1981	06

$annee\_naissance = \pi_{Nom,Annee}(employe)$ retourne (vaut) la relation :

annee_naissance

Nom	Année
Dupond	1978
Dupont	1965
Martin	1981
Martin	1987
Blanc	1981

Sélection $\sigma_C(r)$

employe

NoEmp	Nom	Année	NoDep
1045	Dupond	1978	03
2067	Dupont	1965	06
0456	Martin	1981	03
0278	Martin	1987	05
0789	Blanc	1981	06

Liste des employé(e)s du département 03 né(e)s avant 1980 :

Sélection $\sigma_C(r)$

employe

NoEmp	Nom	Année	NoDep
1045	Dupond	1978	03
2067	Dupont	1965	06
0456	Martin	1981	03
0278	Martin	1987	05
0789	Blanc	1981	06

Liste des employé(e)s du département 03 né(e)s avant 1980 :

$ancien = \sigma_{annee<1980 \wedge NoDep=03}(employe)$

ANCIEN

NoEmp	Nom	Annee	NoDep
1045	Dupond	1978	03

Renommage $\rho[X/X'](r)$

Ce n'est rien d'autre qu'une substitution !

employe

NoEmp	Nom	Année	NoDep
1045	Dupond	1978	03
2067	Dupont	1965	06
0456	Martin	1981	03
0278	Martin	1987	05
0789	Blanc	1981	06

Renommons deux colonnes :

Renommage $\rho[X/X'](r)$

Ce n'est rien d'autre qu'une substitution !

employe

NoEmp	Nom	Année	NoDep
1045	Dupond	1978	03
2067	Dupont	1965	06
0456	Martin	1981	03
0278	Martin	1987	05
0789	Blanc	1981	06

Renommons deux colonnes : $new = \rho[IdEmp/NoEmp, Dpt/NoDep](employe)$

employe

IdEmp	Nom	Année	Dpt
1045	Dupond	1978	03
2067	Dupont	1965	06
0456	Martin	1981	03
0278	Martin	1987	05
0789	Blanc	1981	06

Jointure

Pour la jointure $\bowtie_C$ il nous faut d'abord la définition du produit cartésien, donc passons d'abord aux opérations ensemblistes.

Algèbre Relationnelle (suite)

Puisque les relations sont des ensembles de tuples, on bénéficie en plus de tous les opérateurs ensemblistes.

- A condition d'avoir $R = S$:
  - Différence ($r_1 \setminus r_2$).
  - Intersection ($r_1 \cap r_2$).
  - Union ($r_1 \cup r_2$).
- A condition d'avoir $R \cap S = \emptyset$
-
  - Produit cartésien ($r_1 \times r_2$). La relation obtenue est sur le schéma $R_1 \cup R_2$.

Pour les conditions sur le schéma, on peut les "forcer" par le renommage préalable

Union, Intersection, Différence

EMPLOYE 1

NoEmp	Nom	Annee	NoDep
1045	Dupond	1978	03
2067	Dupont	1965	06
0456	Martin	1981	03

EMPLOYE 2

NoEmp	Nom	Annee	NoDep
1045	Dupond	1978	03
0278	Martin	1987	05
0789	Blanc	1981	06

Que valent $EMPLOYE1 \oplus EMPLOYE2$ avec $\oplus \in \{\cup, \cap, \setminus\}$?

Domaines

On fera attention aux domaines de définition des opérations.

Produit cartésien

Définition

Le produit cartésien de deux relations $r_1 \times r_2$ (de cardinal n_i) est une relation sur le schéma $R_1 \cup R_2$ (les schémas sont supposés disjoints). Les tuples de la relation sont la concaténation d'un élément/tuple de r_1 et d'un élément/tuple de r_2.

On obtient alors une table de taille de $n_1 * n_2$ tuples

Source : équipe BD LIP6

Produit cartésien, exemple

Employe3

NoEmp	Nom	Annee	NoDep
2067	Dupont	1965	06
0456	Martin	1981	03

Departement

NoDep2	Intitule	Taille
03	Comptabilité	6
06	Informatique	10

$RES = EMPLOYE3 \times DEPARTEMENT$ fournit la table :

Produit cartésien, exemple

Employe3

NoEmp	Nom	Annee	NoDep
2067	Dupont	1965	06
0456	Martin	1981	03

Departement

NoDep2	Intitule	Taille
03	Comptabilité	6
06	Informatique	10

$RES = EMPLOYE3 \times DEPARTEMENT$ fournit la table :

NoEmp	Nom	Annee	NoDep	NoDep2	Intitule	Taille
2067	Dupont	1965	06	03	Comptabilité	6
2067	Dupont	1965	06	06	Informatique	10
0456	Martin	1981	03	03	Comptabilité	6
0456	Martin	1981	03	06	Informatique	10

Jointure (def)

Soient :

- deux tables R et S avec $R \cap S = \emptyset$,
- F une formule logique comportant au moins un atome $A_i \oplus B_j$ ($\oplus$ opérateur de comparaison) avec A_i (resp B_j) attribut de R (resp. S)

Alors la jointure est le sous-ensemble des tuples du produit cartésien $R \times S$ qui satisfont F, cad :

$$R \bowtie_F S =_{def} \sigma(R \times S)$$

Types de jointures

- équijointure : la formule F n'utilise que l'égalité
- θ-jointure : la formule F utilise (aussi) des comparaisons
- la jointure dite naturelle.

La jointure naturelle se fait sur des schémas comportant des attributs en commun :

$$R(\mathbf{X}, Y) \bowtie_{JN} S(\mathbf{X}, Y') =_{def} \pi_{S.X, Y, Y'}\Big(\sigma_{R.X=S.X}\big(\underline{\rho[X/R.X](R) \times \rho[X/S.X](S)}\big)\Big)$$

- Le renommage souligné sert à effectuer un produit cartésien correct.
- La partie en bleue filtre celui-ci pour ne garder que les lignes dans lesquelles $S.X = R.X$
- La projection sert à éliminer une des deux colonnes $S.X$ (ici) ou $R.X$
- (il resterait à renommer l'autre colonne)

Exemple d'équi-jointure

Employe

n°empl	nom_e	ville_e	age_e	n°chef_e
141	dupond	paris	40	500
36	durand	tours	40	500
251	parent	agen	25	60

Chef

n°chef	nom_c	age_c
500	albert	50
60	jacquet	40

Quelle jointure pour :

n°empl	nom_e	ville_e	age_e	n°chef_e	n°chef	nom_c	age_c
141	dupond	paris	40	500	500	albert	50
36	durand	tours	40	500	500	albert	50
251	parent	agen	25	60	60	jacquet	40

Exemple d'équi-jointure

Employe

n°empl	nom_e	ville_e	age_e	n°chef_e
141	dupond	paris	40	500
36	durand	tours	40	500
251	parent	agen	25	60

Chef

n°chef	nom_c	age_c
500	albert	50
60	jacquet	40

Quelle jointure pour : $Employe \bowtie_{nchef_e = nchef} Chef$

n°empl	nom_e	ville_e	age_e	n°chef_e	n°chef	nom_c	age_c
141	dupond	paris	40	500	500	albert	50
36	durand	tours	40	500	500	albert	50
251	parent	agen	25	60	60	jacquet	40

Exemple de θ-jointure

n°empl	nom_e	ville_e	age_e	n°chef_e)
141	dupond	paris	40	500
36	durand	tours	40	500
251	parent	agen	25	600
27	barbier	paris	53	500
125	lefevre	paris	30	523
208	legrand	evry	39	523

n°chef	nom_c	age_c
500	albert	50
60	jacquet	40
523	durieux	35

Quelle jointure pour ?

n°empl	nom_e	ville_e	age_e	n°chef_e	n°chef	nom_c	age_c
27	barbier	paris	53	500	500	albert	40
208	legrand	evry	39	523	523	durieux	35

Exemple de θ-jointure

n°empl	nom_e	ville_e	age_e	n°chef_e)
141	dupond	paris	40	500
36	durand	tours	40	500
251	parent	agen	25	600
27	barbier	paris	53	500
125	lefevre	paris	30	523
208	legrand	evry	39	523

n°chef	nom_c	age_c
500	albert	50
60	jacquet	40
523	durieux	35

Quelle jointure pour ? $Employe \bowtie_{nchef_e = nchef \wedge age_e > age_c} Chef$

n°empl	nom_e	ville_e	age_e	n°chef_e	n°chef	nom_c	age_c
27	barbier	paris	53	500	500	albert	40
208	legrand	evry	39	523	523	durieux	35

Exemple de jointure naturelle

EMPLOYE

NoEmp	Nom	Année	NoDep
1045	Dupond	1978	03
2067	Dupont	1965	06
0456	Martin	1981	03
0278	Martin	1987	05
0789	Blanc	1981	06

DEPARTEMENT

NoDep	Intitulé	Taille	NoResp
03	Compta	6	0456
06	Info	10	1249
05	Achats	3	0278

INFO-EMP = EMPLOYE $\bowtie_{JN}$ DEPARTEMENT

INFO-EMP

NoEmp	Nom	Année	NoDep	Intitulé	Taille	NoResp
1045	Dupond	1978	03	Compta	6	0456
2067	Dupont	1965	06	Info	10	1249
0456	Martin	1981	03	Compta	6	0456
0278	Martin	1987	05	Achats	3	0278
0789	Blanc	1981	06	Info	10	1249

Schéma de T :
on ne garde
qu'une seule fois
les attributs communs

Pause : Practise time !

Étudiants	*NUMETUD*	*NOMETUD*	*PRENOMETUD*	*AGE*	*FORMATION*
	28	Codd	Edgar	20	3
	32	Armstrong	William	20	4
	53	Fagin	Ronald	19	3
	107	Bunneman	Peter	18	3

Enseignants	*NUMENS*	*NOMENS*	*PRENOMENS*	*GRADE*
	5050	Tarjan	Robert	PR
	2123	Mannila	Heikki	MCF
	3434	Papadimitriou	Spiros	PR
	1470	Bagan	Guillaume	CR

Encadre	*NUMENS*	*NUMETUD*	*DATE*
	5050	53	2005
	3434	28	2020
	5050	28	2015
	2123	32	2019

- Prénom et nom de tou.te.s les étudiant.e.s
- Prénom et nom des enseignant.e.s qui ont le grade de PR
- Nom(s) des enseignant.e(s) qui encadrent l'étudiant 107
- Num des étudiant.e.s qui n'ont pas d'encadrant.e
- Prénom et nom de tou.te.s les étudiant.e.s et enseignant.e.s

Requêtes algébriques

- Prénom et nom de tou.te.s les étudiant.e.s :

$$\pi_{PRENOMETUD,NOMETUD}(Etudiants)$$

- Prénom et nom des enseignant.e.s qui sont PR :

$$\pi_{PRENOMEND,NOMENS}(\sigma_{GRADE='PR'}(Enseignants))$$

- Nom(s) des enseignant.e(s) qui encadrent l'étudiant 107 :

$$\pi_{NOMENS}(Enseignants \bowtie_{\sigma_{NUMETUD=107}} (Encadre))$$

- Num des étudiant.e.s qui n'ont pas d'encadrant.e :

$$\pi_{NUMETUD}(Etudiants) \setminus \pi_{NUMETUD}(Encadre)$$

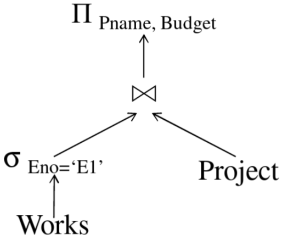
- Nom de tou.te.s les étudiant.e.s et enseignant.e.s :

$$\rho_{NOMETUD/NOM}(\pi_{NOMENS}(Etudiants)) \bigcup \rho_{NOMENS/NOM}(\pi_{NOMENS}(Enseignants))$$

Requête sous forme d'arbre !

Chaque requête peut être décrite sous forme d'arbre :

$$\Pi_{\text{Pname, Budget}}(\text{Project} \bowtie \sigma_{\text{Eno='E1'}}(\text{Works})))$$



Pour quoi faire ?

Calcul Relationnel à Variable Tuples

- Syntaxe :

$$\{x^{(n)}|F(x)\}$$

  où $x^{(n)}$ est un n-uplet (c'est à dire un tuple à n champs) et F est une formule logique du premier ordre ; $F(x)$ exprime donc de façon déclarative les conditions que chaque tuple x doit vérifier pour appartenir au résultat.
- x est une variable libre de $F(x)$.
- On introduit si besoin des variables liées par des quantificateurs $\exists$ ou $\forall$. Ces variables permettent par exemple de parcourir les relations, pour être comparées à x.

Exemples Calcul Relationnel

- Quel est le prenom et le nom de tous les étudiants
  - $\{x = (x_1, x_2) \mid \exists x' \in Etudiants((x_1, x_2) = x'[PRENOMETUD, NOMETUD])\}$
- Quel est le prenom et le nom des enseignants qui sont PR
  - $\{x = (x_1, x_2) \mid \exists x' \in Enseignants(x'[GRADE] =' PR' \wedge (x_1, x_2) = x'[PRENOMENS, NOMENS])\}$
- Quel est le nom des enseignants qui encadrent l'étudiant 53 ?
  - $\{x = (x_1) \mid \exists x' \in Encadre(x'[NUMETUD] = 107 \wedge \exists y' \in Enseignants(x'[NUMENS] = y'[NUMENS] \wedge y'[NOMENS] = x_1))\}$
- Quels est le num des étudiants qui n'ont pas d'encadrant.
  - $\{x = (x_1) \mid \exists x' \in Etudiants(x'[NUMETUD] = x_1 \wedge \forall y' \in Encadre(x'[NUMETUD] \neq y'[NUMETUD]))\}$
- Lister le nom de tous les étudiants et enseignants.
  - $\{x = (x_1) \mid \exists x' \in Etudiants(x'[NOMETUD] = x_1)\} \cup \{x = (x_1) \mid \exists x' \in Enseignants(x'[NOMENS] = x_1)\}$

Next

Un langage pour ces requêtes ?

Bilan

1. Algèbre Relationnelle