Module M4
Base de données

Chapitre 1
Algèbre des relations

Auteur : Laura Monceaux / Véronique Laime

L'algèbre relationnelle a été inventée en 1970 par Edgar Frank Cood et correspond à un ensemble d'opérations formelles sur des relations. Les principes de l'algèbre relationnelle ont été utilisés par les systèmes de gestion de bases de données (SGBD) pour gérer les bases de données comme en SQL que nous verrons dans le chapitre suivant.

Une bonne maîtrise de l'algèbre relationnelle permet de concevoir des requêtes aussi complexes soit elle avant de la mettre en œuvre à l'aide du langage SQL.

Avant de présenter les opérations de bases disponibles de l'algèbre relationnelle et les opérations dérivées, nous allons définir ce qu'est une relation : opérande des opérations en algèbre relationnelle.

1.1 Quelques définitions

Domaine : ensemble fini ou infini de valeurs distinctes que peut prendre une donnée

Exemple : {a,b,c,d} ; Entier ...

T-uple : élément <$v_1, v_2, \ldots, v_n$> tel que quelque soit i $v_i \in D_i$, où D_i est le domaine de la donnée v_i

Exemple : Soient les domaines D1 = {a,b} et D2 = {c,d}

t-uples possibles : <a,c> <a,d> <b,c> <b,d>

Table / Relation : ensemble de t-uples, identifié par un nom

- représentation sous forme de tableau
- une ligne = un t-uple
- une colonne identifiée par un nom : attribut unique de la relation

Degré d'une relation : nombre d'attributs de la relation (δ(R))

Cardinalité d'une relation : nombre de t-uples de la relation (card(R) ou |R|)

Attention : tous les t-uples d'une relation doivent être distincts.

Exemple : relation PERSONNE

Nom	Prénom	Age	VilleNaissance
Dupont	Marc	45	Nantes
Durand	Claude	38	Paris
Martin	Chloé	21	Saint Herblain
Dupont	Pierre	15	Nantes
Martin	Isabelle	35	Rezé

où δ(PERSONNE) = 4 et card(PERSONNE) = 5

Clé d'une relation : ensemble minimum d'attributs permettant de distinguer chaque t-uple de la Relation par rapport à tous les autres

Exemple : clé de la relation PERSONNE = {Nom, Prénom}

Schéma d'une relation : nom de la table, liste des attributs et des contraintes (dont les domaines des valeurs que peut prendre les attributs)

Exemple : relation de nom PERSONNE possédant 4 attributs Nom (chaîne de caractères - cdc), Prénom (chaîne de caractères), Age (entier compris entre 0 et 110), VilleNaissance (chaîne de caractères) dont la clé est {Nom,Prénom}

PERSONNE (Nom cdc, Prénom cdc, Age [0..110],VilleNaissance cdc)

remarque : souvent on souligne les attributs clés.

Schéma d'une base de données : ensemble des schémas des relations de la base de données

Exemple : extrait d'une base de données des films sorties depuis 2009

FILM

idfilm	titre	réalisateur	annee
f1	Alice aux pays des merveilles	T. Burton	2010
f2	Camping 2	F. Onteniente	2010
f3	Incognito	E. Lavaine	2009
f4	Public Enemies	M. Mann	2009

ACTEUR

idacteur	nom	prénom
a1	Depp	Johnny
a2	Wasikowska	Mia
a3	Dubosc	Franck
a4	Anconina	Richard
a5	Seigner	Mathilde

JOUER

idfilm	idacteur
f1	a1
f1	a2
f2	a3
f2	a4
f2	a5
f3	a3
f4	a1

1.2 *Opérateurs de base de l'algèbre relationnelle*

A. Opérateurs de base unaires

Les opérations de bases unaires sont des opérations opérant sur une seule table permettant d'éliminer des attributs ou des t-uples d'une relation. Il existe deux opérations unaires de bases :

- La projection

Objectif : réduire le nombre d'attributs pour ne conserver que ceux considérés « utiles »

Notation : soit R(A1,...,An) une relation

projection de R sur les attributs a1,...,ak : $\pi_{a1,..,ak}$ (R)

Résultat obtenu : T(A1,...,Ak) correspondant aux t-uples de R dont on n'a gardé que les attributs a1, ...,ak

Attention : les doublons sont supprimés.

Exemple :

Soit la relation R(A,B,C) suivante :

A	B	C
1	2	3
3	3	4
2	5	5
3	5	4

$\pi_{A,C}(R)$

A	B	C
1	2	3
3	3	4
2	5	5
3	5	4

La projection de R sur les attributs A,C nous retourne une table contenant 3 t-uples, car après avoir supprimé la colonne B, il ne faut pas oublier de supprimer les doublons (ici deux fois le tuple (3,4)).

- La sélection

Objectif : réduire le nombre de t-uples pour ne conserver que ceux répondant à une condition Q

Notation : soit R(A1,...,An) une relation

sélection de R satisfaisant la condition Q : σ_{Q} (R)

Résultat obtenu : T(A1,...,An) correspondant aux t-uples de R satisfaisant la condition Q

Attention : les doublons sont supprimés.

Exemple :

Soit la relation R(A,B,C) suivante :

A	B	C
1	2	3
3	3	4
2	5	5
3	5	4

$\sigma_{B<4}(R)$ →

A	B	C
1	2	3
3	3	4
~~2~~	~~5~~	~~5~~
~~3~~	~~5~~	~~4~~

B. Opérations de base ensemblistes

Ces opérations permettent de combiner deux relations pour en former une troisième : il s'agit de l'union, l'intersection, le produit cartésien et la différence.

- L'union

Objectif : regrouper des données similaires

Notation : soit R(A1,...,An) et S(A1,...An)

union de R et de S : R U S

Résultat obtenu : T(A1,...,An) correspondant aux t-uples de R et de S

Attention : les doublons sont supprimés.

Attention : l'union se fait entre 2 relations possédant les mêmes attributs (même nom, même domaine)

Exemple :

Soit les relations R(A,B,C) et S(A,B,C) :

R

A	B	C
1	2	3
3	3	4
2	5	5

S

A	B	C
1	2	3
4	5	5

R U S →

T

A	B	C
1	2	3
3	3	4
2	5	5
4	5	5

- ## L'intersection

<u>Objectif</u> : recupérer les t-uples communs à deux relations

<u>Notation</u> : soit R(A1,...,An) et S(A1,...An)

intersection de R et de S : R ∩ S

<u>Résultat obtenu</u> : T(A1,...,An) correspondant aux t-uples de R et de S communs

Attention : les doublons sont supprimés.

Attention : l'intersection se fait entre 2 relations possédant les mêmes attributs (même nom, même domaine)

<u>Exemple</u> :

Soit les relations R(A,B,C) et S(A,B,C) :

R

A	B	C
1	2	3
3	3	4
2	5	5

S

A	B	C
1	2	3
4	5	5

R ∩ S

T

A	B	C
1	2	3

- ## Le produit cartésien

<u>Objectif</u> : combiner les t-uples d'une relation avec tous les t-uples d'une autre relation

<u>Notation</u> : soit R(A1,...,An) et S(B1,..,Bm)

produit cartésien de R et de S : R X S

<u>Résultat obtenu</u> : T(A1,...,An) correspondant à l'ensemble des t-uples ayant n+m attributs dont les n premiers forment les t-uples de R et les m derniers forment les t-uples de S

<u>Exemple</u> :

Soit les relations R(A,B,C) et S(A,D) :

R

A	B	C
1	2	3
3	3	4
2	5	5

S

A	D
1	2
4	5

R X S →

T	R.A	B	C	S.A	D
	1	2	3	1	2
	1	2	3	4	5
	3	3	4	1	2
	3	3	4	4	5
	2	5	5	1	2
	2	5	5	4	5

Attention : les relations R et S possédent un attribut commun, ainsi dans la table résultat, les attributs communs sont préfixés par le nom de leur table d'origine.

- La différence

Objectif : réduire l'ensemble des t-uples d'une relation par l'ensemble des t-uples d'une autre relation

Notation : soit R(A1,...,An) et S(A1,...,An)

différence entre R et S : R - S

Résultat obtenu : T(A1,...,An) correspondant à l'ensemble des t-uples de R n'appartenant pas à S (différence ensembliste)

Attention : la différence se fait entre 2 relations possédant les mêmes attributs (même nom, même domaine)

Exemple :

Soit les relations R(A,B,C) et S(A,D) :

R	A	B	C
	1	2	3
	3	3	4
	2	5	5

S	A	B	C
	1	2	3
	4	5	5

R - S →

T	A	B	C
	3	3	4
	2	5	5

1.3 Opérateurs dérivés

Il existe des opérations dérivées de l'algèbre relationnelle, qui peuvent être construites à partir des opérations de base : les jointures. Celles-ci sont les opérations majeures utilisées en base de données. Ces opérations permettent d'enrichir des t-uples et de mettre en relation des attributs de différentes relations. Nous présentons ainsi dans la suite : la jointure interne, la jointure naturelle, la semi-jointure et la jointure externe.

A. Jointure interne

<u>Notation</u> : soit R(A1,...,An) et S(B1,...,Bm)

jointure interne entre R et S selon la condition Q : $R \bowtie_Q S$

<u>Résultat obtenu</u> : T(A1,...,An,B1,...,Bm) : t-uples de R x S vérifiant la condition Q

$$T = R \bowtie_Q S = \sigma_Q (R \times S)$$

Attention : si un attribut est commun aux deux relations R et S, dans la relation résultante, celui ci sera préfixé par sa table origine et présent deux fois.

<u>Exemple</u> :

Soit les relations R(A,B,C) et S(A,D) :

R	A	B	C
	1	2	3
	3	3	4
	2	5	5

S	A	D
	1	4
	4	3

$R \bowtie_{B<D} S$

1ère étape : faire le produit cartésien de R et S

2ème étape : ne conserver que les tuples vérifiant B<D

T	R.A	B	C	S.A	D
	1	2	3	1	4
	1	2	3	4	3
	3	3	4	1	4
	~~3~~	~~3~~	~~4~~	~~4~~	~~3~~
	~~2~~	~~5~~	~~5~~	~~1~~	~~4~~
	~~2~~	~~5~~	~~5~~	~~4~~	~~3~~

B. Jointure naturelle

<u>Notation</u> : soit R(X1,...,Xk,A1,...,An) et S(X1,...,Xk,B1,...,Bm)

jointure naturelle entre R et S : $R \bowtie S$

Résultat obtenu : T(X1,...,Xk,A1,...,An,B1,...,Bm) : ensemble des t-uples ayant n+m+k attributs dont les n+k premiers forment un t-uple de R et les k premiers et les m derniers forment un t-uple de S.

$$T = R \bowtie S = \pi_{X1,\ldots,Xk,A1,\ldots,An,B1,\ldots,Bm}\ (\sigma_{\forall i \in [1..k]\ R.Xi=S.Xi}\ (R \times S))$$

Attention : pour faire une jointure naturelle, il faut au moins un attribut commun entre les 2 relations.

Attention : on supprime les doublons de la relation obtenue s''il y en a

Exemple :

Soit les relations R(A,B,C) et S(A,D) :

R

A	B	C
1	2	3
3	3	4
2	5	5

S

A	D
1	4
2	3

$R \bowtie S$

T

R.A	B	C	~~S.A~~	D
1	2	3	~~1~~	4
~~1~~	~~2~~	~~3~~	~~2~~	~~3~~
~~3~~	~~3~~	~~4~~	~~1~~	~~4~~
~~3~~	~~3~~	~~4~~	~~2~~	~~3~~
~~2~~	~~5~~	~~5~~	~~1~~	~~4~~
2	5	5	~~2~~	3

1ère étape : faire le produit cartésien de R et S

2ème étape : ne conserver que les tuples vérifiant où R.A = S.A

3ème étape : on ne conserve qu'une seule fois l'attribut commun A (suppression du prefixe

Résultat obtenu :

A	B	C	D
1	2	3	4
2	5	5	3

C. Semi jointure

Notation : soit R(X1,...,Xk,A1,...,An) et S(X1,...,Xk,B1,...,Bm)

semi jointure entre R et S avec R relation directrice : $R \ltimes S$

Résultat obtenu : T(X1,...,Xk,A1,...,An) : projection sur les attributs de la relation directrice de la jointure naturelle de R avec S.

$$T = R \ltimes S = \pi_{X1,...,Xk,A1,...,An}(R \bowtie S)$$

Attention : pour faire une semi jointure, il faut au moins un attribut commun entre les 2 relations.

Attention : on supprime les doublons de la relation obtenue s''il y en a.

<u>Exemple</u> :

Soit les relations R(A,B,C) et S(A,D) :

R

A	B	C
1	2	3
3	3	4
2	5	5

S

A	D
1	4
2	3

$R \ltimes S$

T

A	B	C	~~D~~
1	2	3	~~4~~
2	5	5	~~3~~

on rajoute une étape à la jointure naturelle

4ème étape : ne conserver que les attributs de la relation directrice

D. Jointure Externe

<u>Notation</u> : soit R(A1,...,An) et S(B1,...,Bm)

jointure externe entre R et S

avec R relation directrice : $R \bullet\!\bowtie_Q S$

avec S relation directrice : $R \bowtie_Q \bullet\, S$

avec R et S relations directrices : $R \overset{\bullet}{\bowtie}_Q S$

<u>Résultat obtenu</u> : T(A1,...,An,B1,...,Bm) : t-uples de R X S vérifiant la condition Q (jointure interne) + ajout des t-uples de la relation directrice ne vérifiant pas la condition en complétant les attributs de l'autre relation par des *null.*

Exemple :

R	A	B	C
	1	2	3
	3	3	4
	2	5	5

S	A	D
	1	4
	4	3
	3	1

1ère étape : jointure interne entre R et S avec la condition B < D

R.A	B	C	S.A	D
1	2	3	1	4
1	2	3	4	3
3	3	4	1	4

2ème étape : ajout des t-uples de la relation directrice ne vérifiant pas B<D avec complétion des attributs de l'autre relation par *null*

Si R est directrice : $R \bullet\!\bowtie_{B<D} S$

R.A	B	C	S.A	D
1	2	3	1	4
1	2	3	4	3
3	3	4	1	4
2	5	5	null	null

Si S est directrice : $R \bowtie_{B<D}\!\bullet S$

R.A	B	C	S.A	D
1	2	3	1	4
1	2	3	4	3
3	3	4	1	4
null	null	null	3	1

Si R et S directrices : $R \overset{\bullet}{\bowtie}_{B<D} S$

R.A	B	C	S.A	D
1	2	3	1	4
1	2	3	4	3
3	3	4	1	4
null	null	null	3	1
2	5	5	null	null

1.4 Expressions relationnelles

Une ***expression relationnelle*** est une expression composée d'opérateurs de l'algèbre relationnelle sur des relations.

Le résultat d'une expression relationnelle est la relation obtenue après avoir appliqué toutes les opérations contenues dans l'expression relationnelle. Celles ci sont évaluées de la plus interne à la plus externe.

Exemple : Soit les relations R(A,B,C) et S (A,D,E)

R

A	B	C
1	2	3
2	6	7
3	5	6

S

A	D	E
3	2	4
3	5	6
2	4	6
4	5	5

Expression Relationnelle (exp1) : $\Pi_{A,B}$ ($\sigma_{C>5}$ (R))

1ère étape : sélection C>5 (en rouge)

2ème étape : projection sur A,B (en bleu)

A	B	C
1	2	3
2	6	7
3	5	6

Expression Relationnelle (exp2) : $\Pi_{A,B,E}$ ($\sigma_{D>3}$ (R $\bowtie$ S))

1ère étape : jointure naturelle (table ci dessous)

2ème étape : sélection D>3 (en rouge)

3ème étape : projection sur A,B,E (en bleu)

A	B	C	D	E
2	6	7	4	6
3	5	6	2	4
3	5	6	5	6

Attention : Pour réduire le coût d'une expression relationnelle, on cherche à appliquer les opérations de sélection et de projection sur les tables le plus tôt possible; afin de réduire considérablement les relations avant de réaliser des opérations binaires (qui sont beaucoup plus coûteuse si les relations sont importantes).

Ainsi l'expression relationnelle $\Pi_{A,B,E}$ ($\sigma_{D>3}$ (R $\bowtie$ S)) peut s'optimiser en appliquant les opérations de sélection et de projection dès le départ :

(exp3) ($\Pi_{A,B}$ (R)) $\bowtie$ ($\Pi_{A,E}$ ($\sigma_{D>3}$ S))

1ère étape : projection de A,B de la relation R (en bleu)

2ème étape : sélection de D>3 sur S (en rouge)

3ème étape : projection sur A,E de la relation obtenue à la deuxième étape (en bleu)

4ème étape : jointure naturelle

R

A	B	C
1	2	3
2	6	7
3	5	6

S

A	D	E
3	2	4
3	5	6
2	4	6
4	5	5

jointure naturelle :

A	B	E
2	6	6
3	5	6

Dans l'expression exp2, la jointure naturelle se fait entre la relation R (degré 3 et cardinalité 3) et la relation S (degré 3 et cardinalité 4). Dans l'expression exp3, la jointure naturelle se fait entre deux relations de degrè et de cardinalité moindres. Le coût est donc moins important.

Evidemment pour cet exemple, la différence de coût est minimale mais pour des expressions relationnelles utilisant des relations d'une base de données réelle, cela peut avoir toute son importance. Cette notion d'optimisation des expressions relationnelles pourra être étudiée dans des cours de niveau supérieure.