



Langage SQL

Séquence 1 : Algèbre relationnelle (Rappel)

Table des matières

1	Modèle relationnel	2
1.1	Domaine, Tuples, Attributs et Relations	2
1.2	Clé primaire d'une relation	3
1.3	Schéma d'une Base de données relationnelle	4
2	L'algèbre relationnelle	4
2.1	Les opérateurs unaires	5
2.1.1	La sélection	5
2.1.2	La projection	6
2.2	Les opérateurs binaires ensemblistes	7
2.2.1	union	7
2.2.2	L'intersection	7
2.2.3	La différence	8
2.3	Les opérateurs binaires ou n-aires	9
2.3.1	Le produit cartésien	9
2.3.2	La jointure, jointure naturelle	9

Introduction

Les bases d'un langage structuré d'interrogation de bases de données sont apparues à l'occasion de la publication d'un article de CHAMBERLIN et BOYCE en 1974. Ce premier langage s'appelait **SEQUEL**. Ensuite, **SQUARE** qui utilisait les expressions mathématiques vînt le jour en 1975. Enfin, c'est en 1980 que **SQL (Structured Query Language)** apparaît et continue à couler des jours heureux sur gros systèmes et microordinateurs. Le **SQL** est un langage indispensable en informatique pour stocker et lire des données. De nombreux fournisseurs de logiciels proposent l'utilisation de ce langage. Parmi les principaux **SGBD** (Système de Gestion de Base de Données) possédant l'interface SQL. Le **SQL** standard a été défini par **ANSI**

(American National Standard Institute), c'est un langage d'interrogation de bases de données relationnelles utilisant le concept de vues.

L'algèbre relationnelle a été inventée en 1970 par **Edgar Frank Codd**, le directeur de recherche du centre **IBM**. Cette algèbre est constituée d'un ensemble d'opérations formelles sur les relations. Les opérations relationnelles permettent de créer une nouvelle relation (table) à partir d'opérations élémentaires sur d'autres tables (par exemple l'union, l'intersection, ou encore la différence).

Les principes de l'algèbre relationnelle sont beaucoup utilisés de nos jours par les **SGBD** pour la gestion des bases de données informatiques comme le **SQL**.

1 Modèle relationnel

Le modèle relationnel propose de percevoir les données sous une forme simple de tables (relation).

- + Il s'inspire directement de la notion mathématique de relation.
- + La manipulation des données s'effectue à l'aide d'un ensemble d'opérateurs algébrique (algèbre relationnelle).

1.1 Domaine, Tuples, Attributs et Relations

Les données sont représentées sous la forme de collection de relations (fichiers). Chaque relation est présentée par une **TABLE** de valeurs.

- + Toutes les valeurs d'une même colonne sont toujours de même type (appelé domaine de l'attribut)
- + R peut être soit une entité soit une association (ex : PERSONNE, TRAVAILLER-POUR)

1.2 Clé primaire d'une relation

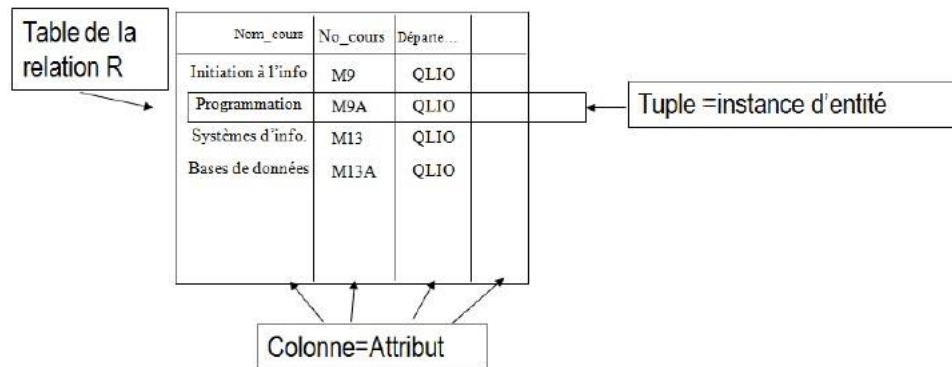


Figure 1 – Une relation avec ses différentes parties

Spécification : Nom + Type + Format + Information additionnelles (ex unités)

Exemple : Num-tél: chaîne de caractère, (xxx)xx xx xx xx avec $x \in \{0, \dots, 9\}$

- + Domaine : Ensemble des valeurs atomiques que peut prendre un attribut.

Le domaine de A_i est noté $\text{Dom}(A_i)$

Figure 2 – Domaine

- + Schéma de relation (structure) : Noté par : $R(A_1 ; A_2 ; \dots ; A_N)$ est un ensemble d'attributs $R = \{A_1 ; A_2 ; \dots ; A_N\}$ où N est le degré de la relation.
 - + Exemple : de schéma de relation Employé pour une relation de degré 6 :
 - + Employé (Nom, Nss, Tel-pers, Adresse, Tel-bur, Age)
- + Instance de relation : une instance de relation r de R est un ensemble de n-tuples $r = \{t_1 ; t_2 ; \dots ; t_m\}$ qui existent dans la base de données à un instant donné.
 - + Chaque n-tuple est constituée par une liste ordonnée de n valeurs t

$= \langle v_1; v_2; \dots; v_N \rangle$ où $\forall i \in [1, N] \ v_i \in \text{dom}(A_i)$ ou $v_i = \text{NULL}$

+ Exemples d'instance et de n-tuples correctes et non correctes

1.2 Clé primaire d'une relation

Une relation est un ensemble de tuples et les tuples sont distincts

1.3 Schéma d'une Base de données relationnelle

Instance :

ETUDIANT	Nom	Matricule	Niveau	Spécialité
	Ludovic	17	1	OGP
	Jean	8	2	OGP

Tuples : $t1 = \langle \text{Philippe}, 15, 3, \text{INFO} \rangle$
 ~~$t2 = \langle \text{Nathalie}, 16, B, \text{OGP} \rangle$~~

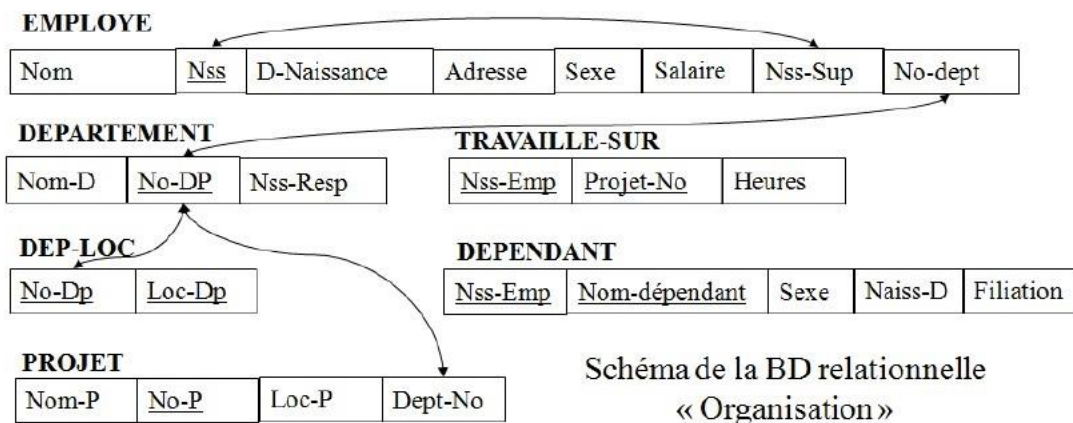
Figure 3 – Une instance

- + Chaque relation possède au moins une super-clé (ensemble de tous les attributs de cette relation).
 - + Une Clé C d'un schéma de relation R est une Super-Clé-minimale de R.
 - + Exemple : R=ETUDIANT (Nss, Nom, Age, Adresse, Tel), {Nss; Nom; Age} est une super-clé; {Nss} est une clé.
 - + Tout schéma de relation R possède obligatoirement une Clé R peut avoir plusieurs clés (Clés candidates).
 - + L'une d'elles sera choisie pour identifier les tuples (Clé Primaire) + Exemple : R=Notation (Matricule-Etudiant, Code-Module, DS, CC)
-
- + Par convention, une clé primaire est soulignée.

1.3 Schéma d'une Base de données relationnelle

Définition. Un schéma de BD relationnelle est un ensemble de schémas de relations $S = \{R_1; R_2; \dots; R_M\}$ et un ensemble de contraintes d'intégrité CI.

- + Une instance de BD relationnelle est un ensemble d'instances de relations $\{r_1; r_2; \dots; r_M\}$ tel que chaque r_i est une instance de R_i et que tous les R_i satisfont les contraintes d'intégrité spécifiées dans CI.



2 L'algèbre relationnelle

Nous pouvons distinguer trois familles d'opérateurs relationnels :

- + Les opérateurs unaires (la **sélection** et la **projection**), qui sont les plus simples, permettent de produire une nouvelle table à partir d'une autre table.
- + Les opérateurs binaires ensemblistes (l'**union**, l'**intersection** et la **différence**) permettent de produire une nouvelle relation à partir de deux relations de même degré et de même domaine.

2.1 Les opérateurs unaires

Figure 4 – Schéma d'une base de données

- + Les opérateurs binaires ou n-aires (le **produit cartésien**, la **jointure** et la **division**) permettent de produire une nouvelle table à partir de deux ou plusieurs autres tables.

Dans ce cours, nous utilisons notre propre notations car, elles ne sont pas standardisées (normées)

2.1 Les opérateurs unaires

2.1.1 La sélection

Définition :

La sélection génère une relation regroupant exclusivement toutes les occurrences de la relation R qui satisfont l'expression logique E. Notation :

$\sigma(E)R$.

Il s'agit d'une opération unaire essentielle dont la signature est :

relation \times expression logique \rightarrow relation

La sélection permet ainsi de choisir (i.e. sélectionner) certaines lignes dans une table. Le résultat de la sélection est donc une nouvelle relation qui a les mêmes attributs que R. Si R est vide (c'est-à-dire sans aucune occurrence), la relation qui résulte de la sélection est vide.

Soit la relation PERSONNE suivante :

Ci-dessous un exemple de sélection $\sigma(idPersonne \geq 5)$ PERSONNE sur la relation PERSONNE :

2.1 Les opérateurs unaires

PERSONNE		
idPersonne	nom	prénom
5	Durand	Caroline
1	Germain	Stan
12	Dupont	Lisa
3	Germain	Rose-Marie

Figure 5 – La relation Personne

$\sigma(idPersonne \geq 5)$ PERSONNE		
idPersonne	nom	prénom
5	Durand	Caroline
12	Dupont	Lisa

Figure 6 – Résultat de la sélection des personne dont l'**idPersonne** est plus grand que 5

2.1.2 La projection

Définition :

La projection consiste à supprimer les attributs autres que A_1, A_2, \dots, A_n d'une relation et à éliminer les n-uplets en double apparaissant dans la nouvelle version.

Notation : $\Pi_{(A_1, A_2, \dots, A_n)} R$

Il s'agit d'une opération unaire essentielle dont la signature est : relation \times liste d'attributs \rightarrow relation

En d'autres termes, la projection permet de choisir des colonnes dans une table. Si R est vide, la relation qui résulte de la projection est vide, mais pas forcément équivalente étant donné qu'elle contient généralement moins d'attributs.

Exemple de projection sur la table PERSONNE :

$\Pi_{(\text{nom})}$ PERSONNE
nom
Durand
Germain
Dupont

Figure 7 – Résultat de la projection sur la colonne Nom de la relation Personne

2.2 ensemblistes

2.2 Les opérateurs binaires ensemblistes

2.2.1 union

Définition :

l'union est une opération portant sur deux relations R1 et R2 ayant le même schéma et construisant une troisième relation constituée des n-uplets appartenant à l'une ou l'autre des deux relations R1 et R2 sans doublon.

Notation : $R1 \cup R2$

Il s'agit d'une opération binaire ensembliste commutative essentielle dont la signature est :

relation \times relation \rightarrow relation

Comme nous l'avons déjà dit, R1 et R2 doivent avoir les mêmes attributs et si une même occurrence existe dans R1 et R2, elle n'apparaît qu'une seule fois dans le résultat de l'union. Le résultat de l'union est une nouvelle relation qui a les mêmes attributs que R1 et R2. Si R1 et R2 sont vides, la relation qui résulte de l'union est vide. Si R1 (respectivement R2) est vide, la relation qui résulte de l'union est identique à R2 (respectivement R1). Ci-dessous un exemple d'union :

R ₁		R ₂		R ₁ \cup R ₂	
nom	prénom	nom	prénom	nom	prénom
Durand	Caroline	Dupont	Lisa	Durand	Caroline
Germain	Stan	Juny	Carole	Germain	Stan
Dupont	Lisa	Fourt	Lisa	Dupont	Lisa
Germain	Rose-Marie			Germain	Rose-Marie
				Juny	Carole
				Fourt	Lisa

Figure 8 – La réunion

2.2.2 L'intersection

Définition :

L'intersection est une opération portant sur deux relations R_1 et R_2 ayant le même schéma et construisant une troisième relation dont les n-uplets sont constitués de ceux appartenant aux deux relations. Notation : $R_1 \cap R_2$

Il s'agit d'une opération binaire ensembliste commutative dont la signature est :

relation \times relation \rightarrow relation

Comme nous l'avons déjà dit, R_1 et R_2 doivent avoir les mêmes attributs.

Le résultat de l'intersection est une nouvelle relation qui a les mêmes attributs

2.2 ensemblistes

que R_1 et R_2 . Si R_1 ou R_2 ou les deux sont vides, la relation qui résulte de l'intersection est vide.

Ci-dessous un exemple d'intersection :

R ₁		R ₂		R ₁ \cap R ₂	
nom	prénom	nom	prénom	nom	prénom
Durand	Caroline	Dupont	Lisa	Durand	Caroline
Germain	Stan	Juny	Carole	Dupont	Lisa
Dupont	Lisa	Fourt	Lisa	Juny	Carole
Germain	Rose-Marie	Durand	Caroline		
Juny	Carole				

Figure 9 – L'intersection

2.2.3 La différence

Définition :

La différence est une opération portant sur deux relations R_1 et R_2 ayant le même schéma et construisant une troisième relation dont les n-uplets sont constitués de ceux ne se trouvant que dans la relation R_1 . Notation : $R_1 - R_2$.

Il s'agit d'une opération binaire ensembliste non commutative essentielle dont la signature est :

relation \times relation \rightarrow relation

Comme nous l'avons déjà dit, R_1 et R_2 doivent avoir les mêmes attributs. Le résultat de la différence est une nouvelle relation qui a les mêmes attributs que R_1 et R_2 . Si R_1 est vide, la relation qui résulte de la différence est vide aussi. Si R_2

est vide, la relation qui résulte de la différence est identique à R1. Exemple de différence entre deux relations :

R ₁		R ₂		R ₁ - R ₂	
nom	prénom	nom	prénom	nom	prénom
Durand	Caroline	Dupont	Lisa	Germain	Stan
Germain	Stan	Juny	Carole	Germain	Rose-Marie
Dupont	Lisa	Fourt	Lisa		
Germain	Rose-Marie	Durand	Caroline		
Juny	Carole				

Figure 10 – La différence

2.3 Les opérateurs binaires ou n-aires

2.3.1 Le produit cartésien

Définition :

Le produit cartésien est une opération portant sur deux relations R1 et R2 et qui construit une troisième relation regroupant exclusivement toutes les possibilités de combinaison des occurrences des relations R1 et R2. Notation :

$R1 \times R2$.

Il s'agit d'une opération binaire commutative essentielle dont la signature est :

relation \times relation \rightarrow relation

Le résultat du produit cartésien est une nouvelle relation qui a tous les attributs de R1 et tous ceux de R2. Si R1 ou R2 ou les deux sont vides, la relation qui résulte du produit cartésien est vide. Le nombre d'occurrences de la relation qui résulte du produit cartésien est le nombre d'occurrences de R1 multiplié par le nombre d'occurrences de R2. Exemple :

PERSONNE		CADEAU		PERSONNE \times CADEAU			
nom	prénom	article	prix	nom	prénom	article	prix
Fourt	Lisa	livre	45	Fourt	Lisa	livre	45
Juny	Carole	poupée	25	Fourt	Lisa	poupée	25
		montre	87	Fourt	Lisa	montre	87
				Juny	Carole	livre	45
				Juny	Carole	poupée	25
				Juny	Carole	montre	87

Figure 11 – Le produit cartésien

2.3.2 La jointure, jointure naturelle

+ La jointure

Définition :

La jointure est une opération portant sur deux relations R1 et R2 qui construit une troisième relation regroupant exclusivement toutes les possibilités de combinaison des occurrences des relations R1 et R2 qui satisfont l'expression logique E. La jointure est notée : $R1 \bowtie_E R2$.

Il s'agit d'une opération binaire commutative dont la signature est :
 $\text{relation} \times \text{relation} \times \text{expression logique} \rightarrow \text{relation}$

Si R1 ou R2 ou les deux sont vides, alors la relation qui résulte de la jointure est vide. En fait, la jointure n'est rien d'autre qu'un produit cartésien suivi d'une sélection :

$R1 \text{ on}_E R2 = \sigma_{(E)}(R1 \times R2)$ Exemple de jointure :

PERSONNE			CADEAU		
nom	prénom	Age	âgeC	article	prix
Fourt	Lisa	6	99	livre	30
Juny	Carole	42	6	poupée	60
Fidus	Laure	16	20	baladeur	45
			10	déguisement	15

		PERSONNE $\bowtie_{(\text{âge} \leq \text{âgeC}) \wedge (\text{prix} \leq 50)}$ CADEAU			
nom	prénom	âge	âgeC	article	prix
Fourt	Lisa	6	99	livre	30
Fourt	Lisa	6	20	baladeur	45
Fourt	Lisa	6	10	déguisement	15
Juny	Carole	42	99	Livre	30
Juny	Carole	42	99	Livre	30
Juny	Carole	42	20	baladeur	45

Figure 12 – Une jointure

+ La jointure naturelle

Définition :

Une jointure naturelle est une jointure dans laquelle l'expression logique E est un test d'égalité entre les attributs qui portent le même nom dans les relations R1 et R2. Dans la relation construite, ces attributs ne sont pas dupliqués, mais fusionnés en une seule colonne par couple d'attributs. La jointure naturelle est notée $R1 \text{ on } R2$. Si la jointure ne doit porter que sur un sous-ensemble des attributs communs à R1 et R2 il faut préciser explicitement ces attributs de la manière suivante : $R1 \text{ on}_{A1, \dots, An} R2$

Généralement, R1 et R2 n'ont qu'un attribut en commun. Dans ce cas, une jointure naturelle est équivalente à une équi-jointure dans laquelle l'attribut de R1 et celui de R2 sont justement les deux attributs qui portent le même nom. Pour effectuer une jointure naturelle entre R1 et R2 sur un attribut A1 commun à R1 et R2, il vaut mieux écrire $R1 \text{ on}_{A1} R2$ que $R1 \text{ on } R2$. En effet, si R1 et R2 possèdent

deux attributs portant un nom commun, R^1 $\text{on}_{A_1} R^2$ est bien une jointure naturelle sur l'attribut A_1 , R^1 $\text{on}_{A_1, A_2} R^2$ est une jointure naturelle sur le couple d'attributs A_1 et A_2 , ce qui produit un résultat très différent.

PERSONNE			CADEAU		
nom	prénom	âge	âge	article	prix
Fourt	Lisa	6	40	livre	45
Juny	Carole	40	6	poupée	25
Fidus	Laure	20	20	montre	87
Choupy	Emma	6			

Figure 13 – Une jointure naturelle

Ci-dessous la jointure naturelle PERSONNE on_{age} CADEAU qui peut également s'écrire PERSONNE on_{age} CADEAU.

PERSONNE			CADEAU		
nom	prénom	âge	âge	article	prix
Fourt	Lisa	6	40	livre	45
Juny	Carole	40	6	poupée	25
Fidus	Laure	20	20	montre	87
Choupy	Emma	6			

Figure 14 – Une jointure naturelle