

Modèle relationnel

Contenu du chapitre:

- Le modèle relationnel
- Algèbre relationnelle



-  savoir écrire des requêtes dans l'algèbre relationnelle.
-  comprendre&concevoir un plan d'exécution de requête.

Table des matières

2020-2021

Modèle relationnel

- Le modèle relationnel
- Algèbre relationnelle

Le modèle relationnel: définitions

Une **BD relationnelle** est un ensemble de relations. Une relation est une table organisée en

- colonnes = attributs

(de type atomique: *int, string ok, mais pas liste, vecteur...*)

- lignes = n-uplets = tuples

Chaque ligne forme une "unité d'information" : décrit un fait réel.



Il n'y a pas d'ordre entre les lignes. En général on considère qu'il n'y a pas de doublons non plus (cela arrive!).

Donc relation = un *ensemble* de lignes (multiensemble si doublons).

Client

NSS	NOM	PRENOM	ADRESSE
1111	GROZ	Benoit	75016
3333	BIDOIT	Nicole	75014
2222	COHEN	Sarah	75008

Achat

ID_CLIENT	NOM_PRODUIT	JOUR
1111	cidre	01/01/2016
1111	pain	01/01/2016
2222	portable	05/01/2016
2222	DVD	07/01/2016

Le modèle relationnel. . . en quelques mots

Modèle Relationnel	SQL
-----------------------	-----

Relation	= Table	client
Attribut	= Colonne	adresse
Domaine	= Type	string
Clefs	= Clefs	nss : clef primaire de Client

Schéma

Schéma de relation

est constitué de:

- nom. *Client*
- liste d'attributs *nom, prenom, adresse* nom des colonnes
- Un domaine pour chaque attribut $Dom(nom) = string$ type des col.

Domaine limité aux types atomiques: int, string ... (pas ~~vector~~ ~~<int>~~)

2 caractéristiques des schémas relationnels:

1. cherchent à minimiser les redondances
2. conception surtout dirigée par les données, plus que par applications.

Représentation textuelle d'un schéma:

`Client(id, nom, prenom, adresse)`

Représentation graphique d'un schéma:

Client
<u>id</u>
nom
prenom
adresse

Schéma de la base

Schéma de base de données

est constitué du schéma de chaque relation.

Représentation graphique d'un schéma de base de données (omettant les domaines):

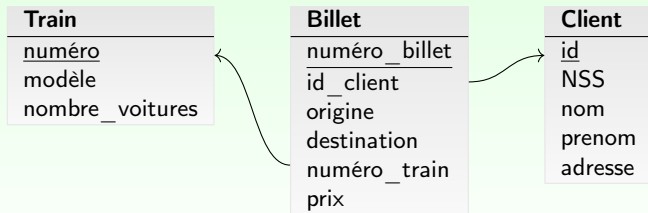


Schéma en général fixé à la création des tables (évolutions possibles mais très rares). L'instance évolue souvent (ex: ajout de nouvelles lignes).

Le schéma de la base inclut en général des clés pour garantir la cohérence des données

Clefs

Clé : groupe d'attributs minimum d'une relation qui détermine chaque tuple de façon unique.

Ex: NSS dans la table Client

Clé primaire : clé choisie par le concepteur de la base de données. Unique.

Ex: numéro_billet dans la table Billet

Clé étrangère : groupe d'attributs d'une relation R1 devant apparaître comme clé primaire dans une autre relation R2.

Ex: Billet(numéro_billet, id_client, origine,...)

référence vers Client.id

Les clés étrangères définissent des contraintes d'*intégrité référentielle*.

Clef primaire

On suppose ici que id a été définie comme clé primaire de client.

Client

id	NSS	nom	prenom	adresse
101	255081416802538	Marie	Dupond	Colombière
104	195042456002601	Jean	Legras	Plaisance
101	255081416802538	Pierre	Rolland	Paris

Deux n-uplets de la relation ne peuvent pas avoir la même clé primaire (id=101).

Parfois, il y a plusieurs attributs dans une clé. Si (nom, prénom) est clé: 2 personnes ne peuvent pas avoir à la fois le même prénom et nom.

Clef étrangère

Cours ne peut faire référence qu'à des salles (avec leur bâtiment) répertoriées dans **Salles**.

Par contre bâtiment et salle peuvent être nuls dans **Cours**.

Cours

cours	horaire	bât	salle	nb_étudiant
Prog Web	...	640	C101	24
Algo	...	640	C101	38

Salles

bât	salle	type
640	C101	TP
620	A101	Amphi
120	D101	Amphi

2020-2021


Modèle relationnel

- Le modèle relationnel
- Algèbre relationnelle
 - Algèbre relationnelle: généralités
 - Selection, Projection, Jointure, Renommage
 - Opérateurs ensemblistes

Le modèle relationnel

Modèle de données

Le modèle relationnel définit (indépendamment du stockage physique)

- une représentation des données et contraintes sur les données
(*Langage de définition des données* (DDL))
- un langage pour manipuler les données:
 - langage de mise à jour
 - langage d'interrogation

Algèbre relationnelle

Langage d'interrogation de données

Décrite par Codd.

- Fournit une base théorique aux langages de requêtes relationnels (SQL).
- Langage *procédural*: décrit des opérations à effectuer (par opposition à langage *déclaratif* comme SQL: ne décrit que le résultat souhaité).
- Un des deux langages de requêtes *formels* associés au modèle relationnel (le second est le *calcul relationnel*)
- L'algèbre=ensemble d'opérations sur les relations. Les opérations sont des fonctions retournant une relation et donc peuvent être composées.
- Considère les relations comme des *ensembles* (pas de doublons).

Algèbre relationnelle

- 2 *opérateurs unaires* : projection, sélection
- 3 *opérateurs binaires* : jointure, union, différence
- 1 *opérateur de renommage*
- *opérateurs dérivés* : produit cartésien, intersection, division, ...
- Typage : chaque *opérateur* définit une *application* de l'ensemble des instances d'un schéma de base de données (source) dans l'ensemble des instances d'un schéma de relation (cible).

Requête suppose le schéma connu, pas l'instance!
- Typage : *schéma cible* déterminé par le schéma source et l'opérateur.

Les transparents de ce cours sont largement repris du cours de N.Bidoit

Sélection σ

permet d'éliminer des nuplets (lignes):

$\sigma_C(\mathbf{r})$

Soit $R(V)$ schéma d'une relation \mathbf{r} , et C condition sur V .

$\sigma_C(\mathbf{r})$ retourne les lignes de \mathbf{r} qui satisfont la condition C .

Une condition sur V est une combinaison booléenne de termes de la forme $A \text{ op } a$ ou $A \text{ op } B$. Formellement:

$C ::= A \text{ op } a$

| $A \text{ op } B$

| $C_1 \wedge C_2$

| $C_1 \vee C_2$

| $\neg C_1$

($A, B \in V$, a constante, $\text{op} \in \{=, \neq, <, >, \leq, \geq\}$)

Sémantique:

$$\sigma_{A=B}(\mathbf{r}) = \{u \in \mathbf{r} \mid u|_A = u|_B\}$$

$$\sigma_{A=a}(\mathbf{r}) = \{u \in \mathbf{r} \mid u|_A = a\}$$

Sélection

Extraire les informations du film "Speed2": $\sigma_{Titre='Speed2'}(\text{film})$

film

Titre	M-en-S	Acteur
Speed 2	Jan de Bont	S. Bullock
Speed 2	Jan de Bont	J. Patric
Speed 2	Jan de Bont	W. Dafoe
Marion	M. Poirier	C. Tetard
Marion	M. Poirier	M-F Pisier
Marion	M. Poirier	M. Poirier

$\sigma_{Titre='Speed2'}(\text{film})$

Titre	M-en-S	Acteur
Speed 2	Jan de Bont	S. Bullock
Speed 2	Jan de Bont	J. Patric
Speed 2	Jan de Bont	W. Dafoe

Projection π

permet de supprimer des colonnes.

$$\pi_{A_1, \dots, A_k}(\mathbf{r})$$

Retourne les colonnes A_1, \dots, A_k de \mathbf{r} . Plus précisément:

$$\pi_W(\mathbf{r}) = \{u|_W \mid u \in \mathbf{r}\}$$

Donc schéma (A_1, \dots, A_k) du résultat diffère en général de celui de \mathbf{r} .



Les doublons sont éliminés (sémantique *ensembliste* de l'algèbre rel.).

Projection

Extraire les titres de tous les films:

film

Titre	M-en-S	Acteur
Speed 2	Jan de Bont	S. Bullock
Speed 2	Jan de Bont	J. Patric
Speed 2	Jan de Bont	W. Dafoe
Marion	M. Poirier	C. Tetard
Marion	M. Poirier	M-F Pisier
Marion	M. Poirier	M. Poirier

$\pi_{Titre}(\mathbf{film})$

Titre
Speed 2
Marion

Les opérateurs renvoient une relation; on peut donc les composer pourvu que les schémas soient compatibles:

$\sigma_{Titre \neq 'Marion'}(\pi_{Titre}(\mathbf{film}))$

Titre
Speed 2

Par contre:

$\sigma_{Acteur \neq 'J.Patrick'}(\pi_{Titre}(\mathbf{film}))$
est indéfini.

Projection

Requête retournant le nom du cinéma et l'horaire des séances de 'Marion' entre 17h00 et 20h00?

programme

Nom-Ciné	Titre	Horaire
Français	Speed 2	18h00
Français	Speed 2	20h00
UGC	Speed 2	22h00
Français	Marion	16h00
Trianon	Marion	18h00
Trianon	Marion	22h00

?

Nom-Ciné	Horaire
Trianon	18h00

Produit cartésien \times

$\mathbf{r} \times \mathbf{s}$

Soient \mathbf{r} et \mathbf{s} deux relations de schéma respectifs $R(A_1, \dots, A_n)$ et $S(B_1, \dots, B_m)$.

$$\mathbf{r} \times \mathbf{s} = \{uv \mid u \in \mathbf{r} \text{ et } v \in \mathbf{s}\}$$

Schéma: concaténation des schémas de R et S (attr. de R puis S , dans l'ordre). Il faut que les schémas soient disjoints. Sinon, on renomme les attributs (ex: $A \rightarrow \mathbf{r}.A$) avant d'effectuer l'opération (d'autres conventions pour traiter les schémas non-disjoints existent).

La cardinalité du résultat est $|\mathbf{r}| \times |\mathbf{s}|$.

Example:

\mathbf{r}		
a	b	c
1	2	10
4	4	10

\mathbf{s}	
d	e
10	'aa'
15	'cc'

$\mathbf{r} \times \mathbf{s}$				
a	b	c	d	e
1	2	10	10	'aa'
1	2	10	15	'cc'
4	4	10	10	'aa'
4	4	10	15	'cc'

Opération coûteuse!

Jointure \bowtie

Opération binaire qui combine deux relations:
produit cartésien des nuplets avec même valeur sur attributs communs.

Jointure naturelle $\mathbf{r} \bowtie \mathbf{s}$, et θ -jointure $\mathbf{r} \bowtie_c \mathbf{s}$

Soient \mathbf{r} et \mathbf{s} deux relations de schéma respectifs $R(V)$ et $S(W)$, avec $V \cap W = A_1, \dots, A_k$.

$$\mathbf{r} \bowtie_c \mathbf{s} = \sigma_c(\mathbf{r} \times \mathbf{s})$$

$$\mathbf{r} \bowtie \mathbf{s} = \{u \mid u|_V \in \mathbf{r} \text{ et } u|_W \in \mathbf{s}\}$$

Remark:

$$\mathbf{r} \bowtie \mathbf{s} = \mathbf{r} \bowtie_{\mathbf{r}.A_1=\mathbf{s}.A_1 \wedge \dots \wedge \mathbf{r}.A_k=\mathbf{s}.A_k} \mathbf{s}$$

Example:

\mathbf{r}		
a	b	c
1	2	10
4	4	10
5	6	20
7	7	30

\mathbf{s}	
c	e
10	'aa'
10	'bb'
15	'cc'
20	'dd'

$\mathbf{r} \bowtie \mathbf{s} = \mathbf{r} \bowtie_{\mathbf{r}.c=\mathbf{s}.c} \mathbf{s}$			
a	b	c	e
1	2	10	'aa'
1	2	10	'bb'
4	4	10	'aa'
4	4	10	'bb'
5	6	20	'dd'

θ -Jointure

Example:

r

c	d	e
1	2	'b'
1	2	'dd'
5	6	'dd'

s

a	b
'x'	'aa'
'y'	'dd'

r $\bowtie_{r.e=s.b}$ **s**

c	d	e	a	b
1	2	'dd'	'y'	'dd'
5	6	'dd'	'y'	'dd'

Jointures: exercices

Jointure $r \bowtie_{r.B=s.B} s?$ $r \bowtie s?$

r

A	B	C
a ₁	b ₁	c ₁
a ₂	b ₁	c ₁
a ₂	b ₂	c ₂
a ₃	b ₂	c ₃

s

B	C	D
b ₁	c ₁	d ₁
b ₂	c ₃	d ₃
b ₄	c ₂	d ₁

Jointure

film(Titre,M-en-S, Acteur) **programme**(Nom-Ciné, Titre, Horaire)

Cinémas projetant un film dans lequel joue *Pisier* (nom-ciné, et horaire du film)?

les films avec leur MeS et acteurs dans lesquels joue M-F. Pisier?

Les titres des films dans lesquels joue M-F. Pisier et qui sont à l'affiche?

Renommage ρ

permet de renommer un ou plusieurs attributs:

$\rho_{A_1 \rightarrow B_1, \dots, A_k \rightarrow B_k}(\mathbf{r})$

$R(A_1, \dots, A_k)$. On suppose la fonction de renommage ρ injective de A_1, \dots, A_k dans B_1, \dots, B_k . Retourne une relation similaire à \mathbf{r} dans laquelle chaque colonne A_i est renommée en B_i :

$$\rho_{A_1 \rightarrow B_1, \dots, A_k \rightarrow B_k}(\mathbf{r}) = \{v \mid \exists u \in R, \forall i : u|_{A_i} = v|_{B_i}\}$$

Note: on omettra en indice les $A_i \rightarrow B_i$ tels que $A_i = B_i$.

Intérêt: donner des noms distincts à des attributs homonyme (jointure...), ou inversement permettre certaines opérations ensemblistes (union...) en donnant le même nom à des attributs distincts.

Renommage

film

Titre	M-en-S	Acteur
Speed 2	Jan de Bont	S. Bullock
Speed 2	Jan de Bont	J. Patric
Speed 2	Jan de Bont	W. Dafoe
Marion	M. Poirier	C. Tetard
Marion	M. Poirier	M-F Pisier
Marion	M. Poirier	M. Poirier

$\rho_{\text{Acteur} \rightarrow \text{Acteur} \text{ Actrice}}(\text{film})$

Titre	M-en-S	Acteur _ Actrice
Speed 2	Jan de Bont	S. Bullock
Speed 2	Jan de Bont	J. Patric
Speed 2	Jan de Bont	W. Dafoe
Marion	M. Poirier	C. Tetard
Marion	M. Poirier	M-F Pisier
Marion	M. Poirier	M. Poirier

Union \cup

Opérations ensemblistes supportées par l'algèbre: $\cup, \cap, -, \times$

$r \cup s$

Soient R et S deux relations *de même schéma*.

$$r \cup s = \{u \mid u \in r \cup s\}$$

 Les doublons sont éliminés!

r

Titre	M-en-S	Acteur
Speed 2	Jan de Bont	S. Bullock
Speed 2	Jan de Bont	J. Patric
Speed 2	Jan de Bont	W. Dafoe

s

Titre	M-en-S	Acteur
Marion	M. Poirier	C. Tetard
Marion	M. Poirier	M-F Pisier
Marion	M. Poirier	M. Poirier

$r \cup s$

Titre	M-en-S	Acteur
Speed 2	Jan de Bont	S. Bullock
Speed 2	Jan de Bont	J. Patric
Speed 2	Jan de Bont	W. Dafoe
Marion	M. Poirier	C. Tetard
Marion	M. Poirier	M-F Pisier
Marion	M. Poirier	M. Poirier

Différence —

Opérations ensemblistes supportées par l'algèbre: $\cup, \cap, -, \times$

r - s

Soient R et S deux relations *de même schéma*.

$$r - s = \{u \in r \mid u \notin s\}$$

r

Titre	M-en-S	Acteur
Speed 2	Jan de Bont	S. Bullock
Speed 2	Jan de Bont	J. Patric
Marion	M. Poirier	C. Tetard
Marion	M. Poirier	M-F Pisier
Marion	M. Poirier	M. Poirier

s

Titre	M-en-S	Acteur
Speed 2	Jan de Bont	W. Dafoe
Marion	M. Poirier	C. Tetard
Marion	M. Poirier	M-F Pisier
Marion	M. Poirier	M. Poirier

r - s

Titre	M-en-S	Acteur
Speed 2	Jan de Bont	S. Bullock
Speed 2	Jan de Bont	J. Patric

Exemples

film(Titre, M-en-S, Acteur)

Proposez des requêtes de l'algèbre relationnelle pour calculer les relations suivantes:

Les films dirigés par au moins deux metteurs en scènes

Les personnes ayant travaillé sur le tournage du film 'Marion'

Les acteurs qui ne sont pas metteurs en scène