

Bases de données

La théorie relationnelle

BD010
v233c

2020-09-04

Département d'informatique
Faculté des sciences



Christina.Khnaisser@USherbrooke.ca
<http://info.USherbrooke.ca/ckhnaisser>
Luc.Lavoie@USherbrooke.ca
<http://info.USherbrooke.ca/llavoie>

PLAN

- Préambule
- Modèles et représentations
- Types, valeurs, variables et relations
- Théorie relationnelle
- Opérateurs
- Contraintes
- De la théorie aux modèles
- Absence
- Vocabulaire
- Références
- Les colles du prof



PRÉAMBULE

- Qu'est-ce qu'une relation
- Quels que mots pour le dire
- Vérification des relations
- Au-delà du binaire
- Une intuition : la logique et les relations

PRÉAMBULE

QU'EST-CE DONC QU'UNE RELATION?

- La réponse mathématique :
 - Un sous-ensemble d'un produit cartésien d'ensembles.
- Soit les ensembles
 - Nom = {Paul, Éliane, Mohamed, Sergeï, Diego}
 - Ville = {>Δ^aσ▷^b, Blanc-Sablon, Tadoussac, Chandler, Adélaïde, San Diego}
- Un exemple de **valeur** de relation
 - **v1** = {(Paul, >Δ^aσ▷^b), (Éliane, Blanc-Sablon), (Mohamed, Blanc-Sablon), (Sergeï, Chandler), (Sergeï, San Diego)}
- Une constante de relation est aussi appelée, dans certains contextes, *constante* (de relation) ou *littéral* (de relation).

PRÉAMBULE

QU'EST-CE DONC QU'UN TYPE DE RELATION?

- Informellement, un type est un ensemble fini de valeurs.
- Un type de relation est donc un identifiant qui réfère à un ensemble de valeurs d'une même relation.

- Un exemple de **type** de relation
 - **TVN** = *Nom* × *Ville* ;

PRÉAMBULE

QU'EST-CE DONC QU'UNE VARIABLE RELATION?

- Informellement, une variable est un identifiant qui réfère (désigne, est associé...) à une valeur.
- Une variable de relation est donc un identifiant qui réfère à une valeur de relation.
- Un exemple de **variable** de relation :
 - **Visite** : TVN ;
- À sa déclaration la variable est associée à l'ensemble vide $\{\}$.
- Pour en changer l'association, on a recours à l'opérateur d'affection :
 - **Visite** := v1 ;

PRÉAMBULE

QU'EST-CE DONC QU'UN INVARIANT?

- Un invariant est une condition qui doit être vraie en tout temps.
- À la déclaration d'une variable est toujours associée un invariant qui prescrit que la valeur associée à la variable doit faire partie (être un élément) du type associé à la variable.
- Par exemple, attendu la déclaration **Visite** : **TVN** ;
 - **Visite** \subseteq **TVN**
 - **Visite** \subseteq **Nom** \times **Ville**

PRÉAMBULE

LA VÉRIFICATION DES AFFECTATIONS EST CAPITALE!

○ Soit les ensembles

- Nom = {Paul, Éliane, Mohamed, Sergeï, Diego}
- Ville = {>Δ^ςσ▷^{ςb}, Blanc-Sablon, Tadoussac, Chandler, Adélaïde, San Diego}
- v2 = {(Paul, >Δ^ςσ▷^{ςb}), (Éliane, Blanc-Sablon), (Tadoussac, Blanc-Sablon), (Sergeï, Chandler), (Sergeï, Sans Diego)}

○ La constante v2 ne fait pas partie du type TVN :

- v2 \notin Nom × Ville
- v2 \notin TVN

○ La variable Visite : TVN ne peut donc pas référer à v2.

- L'affectation « Visite := v2 » induit donc une erreur.

PRÉAMBULE
VOCABULAIRE

- L'expression « variable de relation » est souvent abrégée en *relvar*.
- Par analogie, il y a aussi *reltype* et *relcon*.
- Un identifiant est aussi appelé, dans certains contextes, *identificateur*.

PRÉAMBULE

NOTATION

- Par souci d'expressivité, on permet souvent la déclaration d'une *relvar* sous une forme compacte, par exemple
 - RELATION **Visite** {*Nom*, *Ville*}

REMARQUE

- Les ensembles sont des cas particuliers de relations en ce sens qu'il y a une bijection triviale entre une relvar définie sur un un seul ensemble et cet ensemble lui-même.
- Par exemple
 - RELATION $V \{ \text{Ville} \}$ et Ville

PRÉAMBULE

RELATION N-AIRE

- **Visite** est une variable de relation binaire, qu'advient-il des relations de degré supérieur?
- Soit les ensembles
 - Nom = {Paul, Éliane, Mohamed, Sergeï, Diego}
 - Ville = {>Δ^ϕσ▷^{ϕb}, Blanc-Sablon, Tadoussac, Chandler, Adélaïde, San Diego}
 - Matricule = {15113150, 15112354, 15113870, 15110132}
- **RELATION Etudiant (Matricule, Nom, Ville)**
 - Etudiant \subseteq Matricule \times Nom \times Ville
 - Etudiant = { (15113150, Paul, >Δ^ϕσ▷^{ϕb}),
(15112354, Éliane, Blanc-Sablon),
(15113870, Mohamed, Blanc-Sablon),
(15110132, Sergeï, Chandler),
(15110132, Sergeï, San Diego) }

PRÉAMBULE

PRÉDICAT ET PROPOSITIONS

- Une relation est décrite par un prédicat qui représente une partie de la réalité.
- Un prédicat est une fonction avec une ou plusieurs variables. En associant des valeurs aux variables du prédicat, on obtient la proposition.
- Une proposition est un énoncé vrai ou faux à propos d'une partie de la réalité.

PRÉAMBULE

LES RELATIONS ET LA LOGIQUE

- Quels sont les prédicats représentés par les relations Visite et Étudiant?
- Peut-on les déduire à partir des relations?
- Peut-on déduire les relations à partir des prédicats?
- Y a-t-il un lien entre prédicat et information?

PRÉAMBULE

QUE CHOISIR ET POURQUOI?

- Chaque modèle est-il « complet »?
 - Qu'entend-on par complet?
- Chaque modèle facilite certaines opérations
 - Lesquelles?

MODÈLES ET REPRÉSENTATIONS

- Avec une matrice ?
- Avec un graphe ?
- Avec un table ?
- Mieux : le modèle (info-)relationnel !

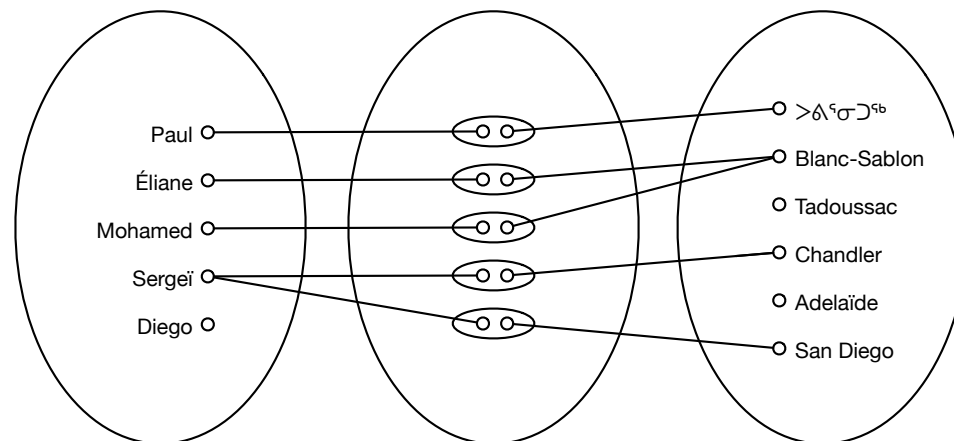
QUEL MODÈLE CHOISIR, LE MODÈLE MATRICIEL?

- Nom = {Paul, Éliane, Mohamed, Sergeï, Diego}
- Ville = {>Δ^ϕσ▷^{ϕb}, Blanc-Sablon, Tadoussac, Chandler, Adélaïde, San Diego}
- Visite = {(Paul, >Δ^ϕσ▷^{ϕb}), (Éliane, Blanc-Sablon), (Mohamed, Blanc-Sablon), (Sergeï, Chandler), (Sergeï, San Diego)}

	>Δ ^ϕ σ▷ ^{ϕb}	Blanc-Sablon	Tadoussac	Chandler	Adélaïde	San Diego
Paul	1	0	0	0	0	0
Éliane	0	1	0	0	0	0
Mohamed	0	1	0	0	0	0
Sergeï	0	0	0	1	0	1
Diego	0	0	0	0	0	0

QUEL MODÈLE CHOISIR, LE GRAPHE?

- Nom = {Paul, Éliane, Mohamed, Sergeï, Diego}
- Ville = {>Δ^ϕσ▷^{ϕb}, Blanc-Sablon, Tadoussac, Chandler, Adélaïde, San Diego}
- Visite = {(Paul, >Δ^ϕσ▷^{ϕb}), (Éliane, Blanc-Sablon), (Mohamed, Blanc-Sablon), (Sergeï, Chandler), (Sergeï, San Diego)}



QUEL MODÈLE CHOISIR, LE MODÈLE « TABULAIRE »?

- Nom = {Paul, Éliane, Mohamed, Sergeï, Diego}
- Ville = {>Δ^ϵσ▷^{ϵb}, Blanc-Sablon, Tadoussac, Chandler, Adélaïde, San Diego}
- Visite = {(Paul, >Δ^ϵσ▷^{ϵb}), (Éliane, Blanc-Sablon), (Mohamed, Blanc-Sablon), (Sergeï, Chandler), (Sergeï, San Diego)}

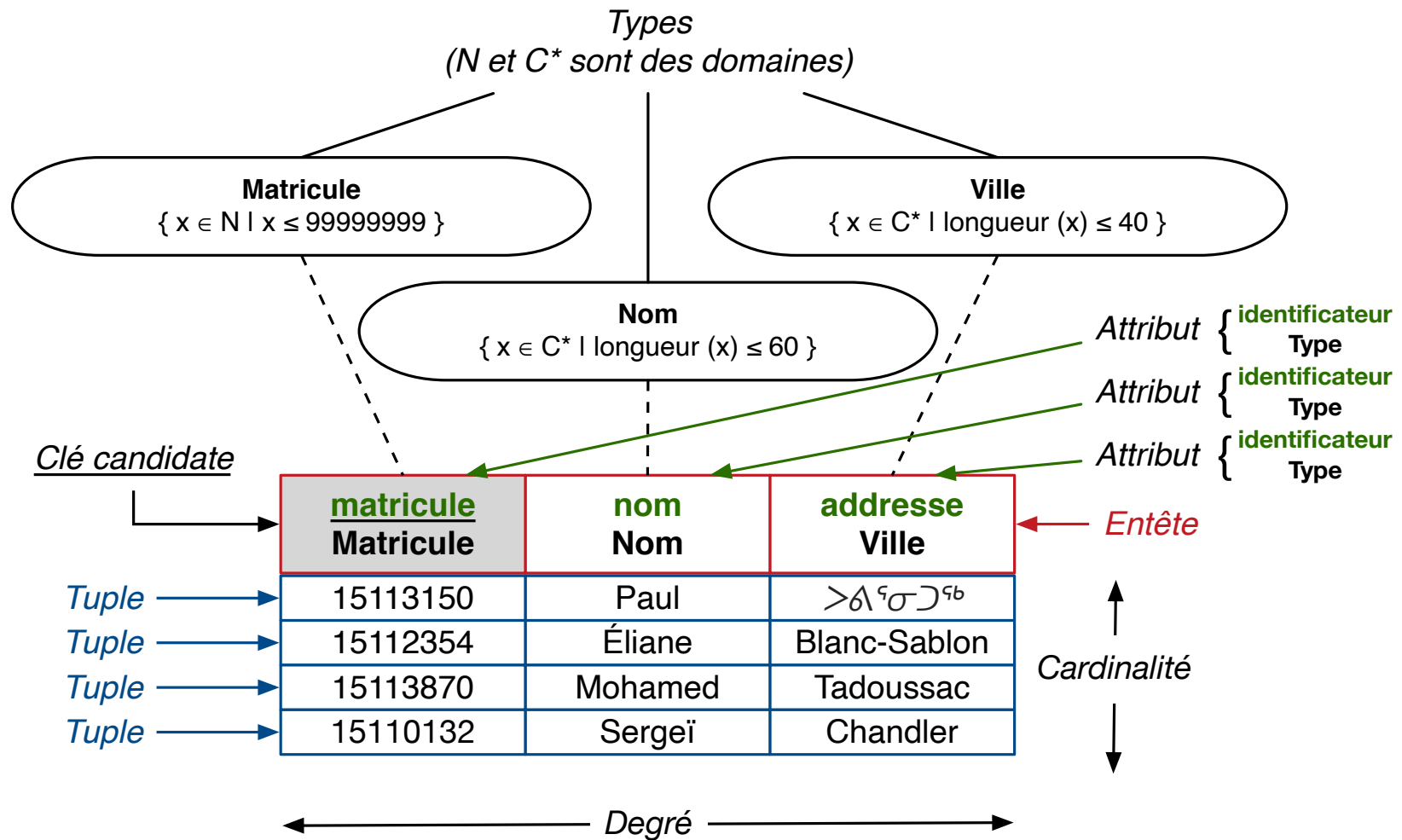
Nom	Ville
Paul	>Δ ^ϵ σ▷ ^{ϵb}
Éliane	Blanc-Sablon
Mohamed	Blanc-Sablon
Sergeï	Chandler
Sergeï	San Diego

QUE CHOISIR ET POURQUOI?

- Chaque modèle est-il « complet »?
 - Qu'entend-on par complet?
- Chaque modèle facilite certaines opérations
 - Lesquelles?
- Notre choix : le modèle « info-relationnel » !

LE MODÈLE INFO-RELATIONNEL ILLUSTRÉ PAR LA RELATION *ÉTUDIANT*

Modèle-Relationnel_PRE.graffle [R] (2020-08-29)



TYPES, VALEURS, VARIABLES ET RELATIONS

- Domaine, type et relation (bis)
- Référence, dénotation et variable
- Logique et relations (bis)

RAPPELS A

- Un **domaine** est un ensemble *fini* de valeurs *propres*.
- Un **type** est un sous-ensemble d'un **domaine** déterminé par une contrainte (qui en restreint les valeurs).
- Une **relation** de degré n est un **sous-ensemble** du produit cartésien de n domaines (de n types) :

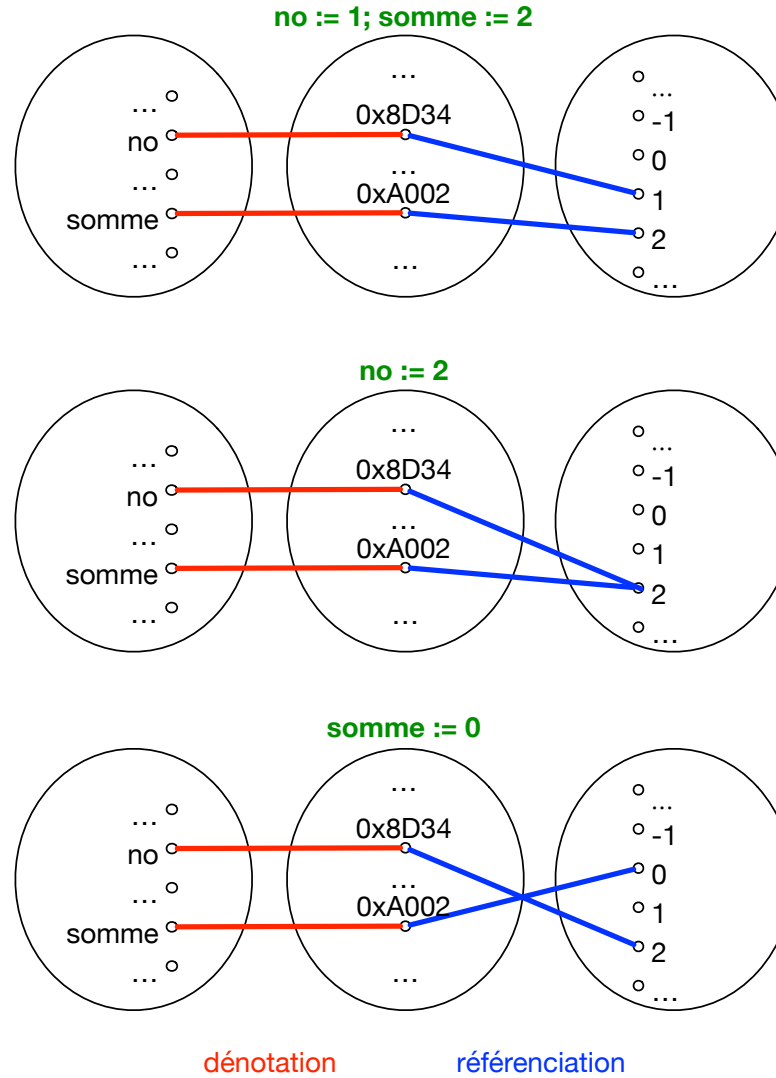
$$D_1 \times D_2 \times \dots \times D_n$$

RAPPELS B

- Une **référenciation** est une **fonction** associant une référence (adresse) à une valeur.
- Une **dénotation** est une **fonction** associant un nom (identifiant) à une référence.
- Une **variable** est une **dénotation**.
- Une variable est dite **typée** si, par construction, elle réfère toujours à une valeur d'un même type.

VARIABLE ET AFFECTATION

Domaine des noms Domaine des références Domaine des entiers



LES RELATIONS ET LA LOGIQUE

relation	prédicat modélisé par un ensemble de tuples représentée par un tableau (une table).
tuple	proposition modélisée par un ensemble d'attributs représenté par une ligne (un enregistrement).
attribut	variable typée représentée par une cellule (un champ).
contrainte	expression logique.
relvar	variable référant une (valeur de) relation.
schéma	un ensemble de définitions de relvar et un ensemble de définitions de contraintes.
base de données	un ensemble de relvars conformes à un schéma (donc aux définitions des relvars et des contraintes).

Note : dans le but de faciliter la gestion de la complexité, on permet généralement qu'une base de données puisse être définie à l'aide de plusieurs schémas; chaque identifiant défini par un schéma est alors qualifié (préfixé) par le nom du schéma.

INFORMATION ET RELVAR, RELVAR ET PRÉDICAT, TUPLE ET PROPOSITION

- À chaque relvar on associe un prédicat qui en est « l'interprétation » informationnelle.
- En substituant les valeurs des attributs d'un tuple aux variables correspondantes du prédicat de la relvar, on obtient une proposition.
- À chaque tuple est alors associée une proposition qui en est « l'interprétation » informationnelle.

REMARQUES

- L'interprétation informationnelle d'un tuple est aussi appelée, dans certains contextes, « fait ».
- L'association d'une valeur à un attribut passe par la donnée... et c'est très important!
 - C'est la donnée qui permet de rendre l'attribut « manipulable » informatiquement et, corolairement, le tuple, la relvar, le prédicat.
 - À son tour, le prédicat rendu ainsi manipulable permet la représentation et la transmission de l'information.
 - ... et bien d'autres choses encore !

PRÉAMBULE

DE LA THÉORIE DES TYPES

- La définition du type présentée ici est minimale. La plupart des auteurs s'entendent pour associer au type un ensemble d'opérateurs. La notion de type peut être développée encore plus.
- La théorie fondamentale des types a été établie dans l'article suivant :
 - Luca Cardelli and Peter Wegner. 1985.
On understanding types, data abstraction, and polymorphism.
ACM Computing Surveys 17, 4 (December 1985), 471–523.
DOI:<http://dx.doi.org/10.1145/6041.6042>
- Le cours IFT 232 approfondira cette question par la présentation de la théorie des types abstraits algébriques (qui est à l'origine de la méthode de conception par classes, aussi appelée programmation orientée-objet).

FONDEMENTS

- Attributs
- Tuples
- Relations
- Opérateurs
- Contraintes

- Un attribut est un couple formé d'un identifiant **a** et d'un type **D**, noté **a:D**.
- Par abus de langage, lorsque le contexte le permet, il est usuel de désigner l'attribut par son seul identifiant; ainsi écrit-on l'attribut **a**.
- Rappels :
 - Un domaine est un ensemble *fini* de valeurs (distinctes, dans le sens où une même valeur ne peut appartenir à deux domaines distincts).
 - Un type est un ensemble de valeurs défini par un domaine et une contrainte (l'union et l'héritage de types ne sont pas pris en compte ici).

FONDEMENTS – TUPLES

- Soit a_i des identifiants distincts et D_j des types, un tuple t est défini comme suit :
 - $t \triangleq (\{a_1:D_1, a_2:D_2, \dots, a_n:D_n\}; \{(a_1,v_1), (a_2,v_2), \dots (a_n,v_n)\})$
 - avec $\forall i : 1 \leq i \leq \text{deg}(t) \implies \text{val}(t, a_i) \in \text{def}(t, a_i)$
- Où
 - $\text{def}(t) = \{a_1:D_1, a_2:D_2, \dots, a_n:D_n\}$ entête de t
 - $\text{def}(t, a_i) = D_i$ type de a_i
 - $\text{val}(t) = \{(a_1,v_1), (a_2,v_2), \dots (a_n,v_n)\}$ valeur de t
 - $\text{val}(t, a_i) = v_i$ valeur de a_i
 - $\text{deg}(t) = n$ degré de t
 - $\text{id}(t) = \{a_1, a_2, \dots, a_n\}$ les identifiants d'attributs de t

- Soit a_i des identifiants distincts, D_j des types et t_k des tuples, une relation R est définie comme suit :
 - $R \triangleq (\{a_1:D_1, a_2:D_2, \dots, a_n:D_n\}; \{t_1, t_2, \dots, t_m\})$
 - avec $\forall i : 1 \leq i \leq \text{card}(R) \Rightarrow \text{def}(R) = \text{def}(t_i)$
- Où
 - $\text{def}(R) = \{a_1:D_1, a_2:D_2, \dots, a_n:D_n\}$ entête de R
 - $\text{def}(R, a_i) = D_i$ type de a_i
 - $\text{val}(R) = \{t_1, t_2, \dots, t_m\}$ valeur de R
 - $\text{deg}(R) = n$ degré de R
 - $\text{card}(R) = m$ cardinalité de R
 - $\text{id}(R) = \{a_1, a_2, \dots, a_n\}$ identifiants d'attributs de R

FONDEMENTS — ILLUSTRATION

Grâce aux
contraintes sur les
tuples et les
relations, la
représentation
tabulaire initiale
est donc bien
fondée.

Quatre tuples (ayant le même entête)

matricule : Matricule matricule : 15113150	nom : Nom nom : Paul	adresse : Ville adresse : >Δ ^σ ⊂ ^{sb}	t1
matricule : Matricule matricule : 15112354	nom : Nom nom : Éliane	adresse : Ville adresse : Blanc-Sablon	t2
matricule : Matricule matricule : 15113870	nom : Nom nom : Mohamed	adresse : Ville adresse : Tadoussac	t3
matricule : Matricule matricule : 15110132	nom : Nom nom : Sergeï	adresse : Ville adresse : Chandler	t4

Une relation comprenant quatre tuples

matricule : Matricule	nom : Nom	adresse : Ville	
matricule : Matricule matricule : 15113150	nom : Nom nom : Paul	adresse : Ville adresse : >Δ ^σ ⊂ ^{sb}	t1
matricule : Matricule matricule : 15112354	nom : Nom nom : Éliane	adresse : Ville adresse : Blanc-Sablon	t2
matricule : Matricule matricule : 15113870	nom : Nom nom : Mohamed	adresse : Ville adresse : Tadoussac	t3
matricule : Matricule matricule : 15110132	nom : Nom nom : Sergeï	adresse : Ville adresse : Chandler	t4

La représentation compacte usuelle de cette même relation

matricule : Matricule	nom : Nom	adresse : Ville
15113150	Paul	>Δ ^σ ⊂ ^{sb}
15112354	Éliane	Blanc-Sablon
15113870	Mohamed	Tadoussac
15110132	Sergeï	Chandler

FONDEMENTS – RACCOURCIS ET NOTATIONS

- Notations équivalentes à $\text{val}(t, a_i)$
 - $t.a_i$
 - $a_i(t)$
 - $t(a_i)$
 - a_i from t
- Nous utiliserons fréquemment
 - $t.a_i$

OPÉRATEURS

- Premier essai
- Renommage
- Deuxième essai
- Et encore plus !

FONDEMENTS — OPÉRATEURS NATURELS

Restriction
 $R \sigma \text{ cond}$

A	B	C
a1	b1	c1
a2	b2	c2
a3	b3	c3

Projection
 $R \pi \{A, C\}$

A	B	C
a1	b1	c1
a2	b2	c2
a3	b3	c3

Jointure naturelle
 $R \bowtie S$

A	B
a1	b1
a2	b1
a3	b3
a4	b4

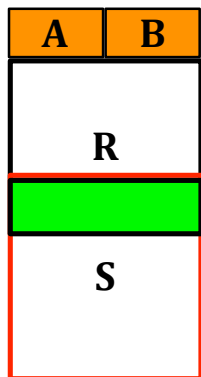
 \bowtie

B	C
b1	c1
b2	c2
b3	c3
b3	c4

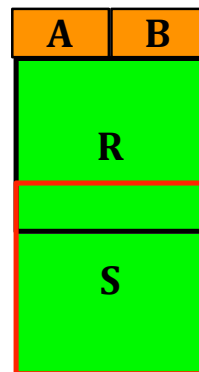
 $=$

A	B	C
a1	b1	c1
a2	b1	c1
a3	b3	c3
a3	b3	c4

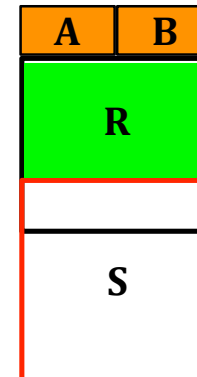
Intersection
 $R \cap S$



Union
 $R \cup S$



Différence
 $R - S$



Note : Le symbole de projection π est souvent omis, à l'instar de la multiplication dans les polynômes.

FONDEMENTS — OPÉRATEUR DE RENOMMAGE

- La présence de l'entête dans chacun des tuples et chacune des relations permet de définir un opérateur de renommage.
- L'entête d'une relation est conservé dans le catalogue du SGBDR.
- Le catalogue est la description des schémas relationnels du SGBDR sous la forme d'une BD dont le schéma est lui-même dans le catalogue, comme tous les autres schémas de toutes les autres BD du SGBDR.

Renommage
 $R \rho A:C$

A	B
a1	b1
a2	b2
a3	b3

 $\rho A:C =$

C	B
a1	b1
a2	b2
a3	b3

FONDEMENTS — OPÉRATEURS DE BASE

- On constate toutefois que $R \cap S = R - (R - S) = R \bowtie S$
- L'opérateur \cap n'est donc pas « nécessaire ».

Restriction
 $R \sigma \text{ cond}$

A	B	C
a1	b1	c1
a2	b2	c2
a3	b3	c3

Projection
 $R \pi \{A, C\}$

A	B	C
a1	b1	c1
a2	b2	c2
a3	b3	c3

Jointure (naturelle)
 $R \bowtie S$

A	B
a1	b1
a2	b1
a3	b3
a4	b4

 \bowtie

B	C
b1	c1
b2	c2
b3	c3
b3	c4

 $=$

A	B	C
a1	b1	c1
a2	b1	c1
a3	b3	c3
a3	b3	c4

Différence
 $R - S$

A	B
R	
S	

Union
 $R \cup S$

A	B
R	
S	

Renommage
 $R \rho A:C$

A	B
a1	b1
a2	b2
a3	b3

 $\rho A:C =$

C	B
a1	b1
a2	b2
a3	b3

FONDEMENTS — OPÉRATEURS NATURELS

RÈGLES D'APPLICATION

- Les trois opérateurs ensemblistes ne sont bien définis que si les entêtes des relations sont identiques.
- L'opérateur de projection n'est bien défini que si tous les identifiants d'attributs sont définis dans l'entête de la relation.
- L'opérateur de restriction n'est bien défini que si tous les identifiants d'attributs de la condition sont définis dans l'entête de la relation.
- L'opérateur de renommage n'est bien défini que si l'identifiant à changer est défini dans l'entête de la relation et que le nouvel ne l'est pas.

FONDEMENTS — PRODUIT CARTÉSIEN OU JOINTURE ?

- Constatons d'abord l'un se calcule à l'aide de l'autre :
 - $R \times S = R \bowtie S$
pourvu que R n'ait aucun attribut commun avec S
ce dont on peut s'assurer par renommage;
 - $R \bowtie S = (R \times S) \sigma (R.a_1 = S.a_1 \wedge \dots \wedge R.a_n = S.a_n)$
pourvu qu'on ait la notion pointée à disposition.

Préférez-vous le chocolat ou le cacao ?

FONDEMENTS — OPÉRATEURS : ENSEMBLE MINIMAL ?

- En pratique, deux ensembles de six opérateurs de base sont proposés, selon les auteurs :
 - Codd : $\pi \sigma \times \cup - \rho$
 - Date : $\pi \sigma \bowtie \cup - \rho$
- Un ensemble minimal comprend aussi peu que deux opérateurs, voir l'algèbre \mathcal{A} proposée par Date (2007).
- Nous utiliserons toutefois tantôt l'un, tantôt l'autre, car ils proposent un bon équilibre entre minimalisme et expressivité.

FONDEMENTS — OPÉRATEURS CONSTRUITS

Extension (augmentation)

$$R \xi C:f = (R \bowtie F) \rho f':C$$

A	B
a1	b1
a2	b1
a3	b3
a4	b4

 $\xi C:f =$

A	B	C
a1	b1	f(a1,b1)
a2	b1	f(a2,b1)
a3	b3	f(a3,b3)
a4	b4	f(a4,b4)

Note : Le symbole de l'extension varie beaucoup d'un auteur à l'autre.

Semi-jointure (matching)

$$R \times S = (R \bowtie S) \pi R$$

A	B
a1	b1
a2	b1
a3	b3
a4	b4

 \times

B	C
b1	c1
b2	c2
b3	c3
b3	c4

 $=$

A	B
a1	b1
a2	b1
a3	b3

Produit

$$R \times S = R \bowtie S$$

(jointure sans attributs partagés)

A
x
y

 \times

B
a
b
c

 $=$

A	B
x	a
x	b
x	c
y	a
y	b
y	c

Semi-différence (not matching, anti-join...)

$$R \times S = R - (R \bowtie S)$$

A	B
a1	b1
a2	b1
a3	b3
a4	b4

 $-$

B	C
b1	c1
b2	c2
b3	c3
b3	c4

 $=$

A	B
a4	b4

FONDEMENTS — OPÉRATEURS CONSTRUITS

- Se reporter au module BD012 qui comprend une suite d'exemples de requêtes relationnelles utilisant les opérateurs de base et quelques opérateurs construits.

CONTRAINTES

- Contraintes générales
- Contraintes spécifiques
- Vocabulaire

CONTRAINTES

○ Contrainte (générale)

- Une contrainte est une expression logique applicable aux objets d'une base de données; l'expression qui doit être maintenue vraie tout au long de l'existence de la base de données.

○ Contrainte (spécifique)

- Toute autre contrainte spécifique est un cas particulier de contrainte générale.
- On désigne souvent une contrainte spécifique selon
 - sa portée par rapport à un objet particulier de la base de données : contrainte d'attribut, contrainte de relation, contrainte de schéma ;
 - son rôle : clé candidate, clé référentielle

CONTRAINTES SPÉCIFIQUES (PORTÉE)

- Contrainte de domaine
 - restreindre l'ensemble des valeurs d'un domaine (pour en faire un type, par exemple).
- Contrainte d'attribut
 - limiter la valeur des attributs d'un tuple entre eux.
- Contrainte de relation
 - limiter globalement les valeurs de la relation (donc celles des tuples de la relation entre eux);
 - exemple : clés candidates.
- Contrainte d'intégrité référentielle
 - limiter globalement les valeurs d'attributs d'une relation relativement à une clé candidate d'une autre;
 - exemple : clés référentielles (étrangères).

CONTRAINTES SPÉCIFIQUES (RÔLE)

○ Clé (d'une relation R)

- sous-ensemble d'attributs déterminant un tuple unique au sein de la relation.
 - Soit X une clé de R :
 $\#R = \#(R \pi \{X\})$

○ Clé référentielle (d'une relation R sur une relation S)

- sous-ensembles d'attributs dont les valeurs sont restreintes par la clé d'une (autre) relation
 - Soit X une clé référentielle de R sur S :
 $(R \pi \{X\}) \subseteq (S \pi \{X\})$ et (X est une clé de S)

VOCABULAIRE

- Clé irréductible
 - clé dont on ne peut retirer aucun attribut sans qu'elle cesse d'être une clé.
- Clé candidate
 - synonyme : clé irréductible.
- Surclé
 - tout ensemble d'attributs contenant une clé.

CONTRAINTES, UNE SUITE ?

- Nous reprendrons l'étude des clés (et des dépendances fonctionnelles) quand nous aurons mieux maîtrisé la théorie relationnelle.

DE LA THÉORIE AUX MODÈLES

Plusieurs modèles ont été proposés sur la base d'une même théorie

- Modèle de Codd I
- Modèle de Codd II
- Modèle de Date
- Modèle d'Ullman
- Modèles SQL
- ...
- Et la liste n'est pas complète!
- ...
- Que choisir ?

DE LA THÉORIE AUX MODÈLES THÉORIE, MODÈLES ET LANGAGES

- Il y a **une** théorie relationnelle.
- Il y a **plusieurs** modèles relationnels, par exemple
 - Modèle de Codd I
 - Modèle de Codd II
 - Modèle de Date
 - Modèle d'Ullman
 - Modèles SQL (ANSI:1992, ISO:1999, ISO:2006... ISO:2016...)
- Pour **chacun** de ces modèles, il y a **plusieurs** langages (et même, plusieurs dialectes pour certains de ces langages)

DE LA THÉORIE AUX MODÈLES QUE CHOISIR ?

- Pour l'exposé des principes relationnels, nous utiliserons toujours le modèle de Date.
- Pour la programmation SQL, nous présenterons des techniques permettant d'être aussi proche que possible du modèle de Date, en indiquant les écarts possibles en fonction du modèle SQL ISO 9075:2016.
- Certains des autres modèles seront couverts par les activités IFT 287, IGE 487 et IFT 723.
- Le calcul des propositions et des prédicats est couvert par l'activité MAT 115.

DONNÉES ABSENTES

- Pourquoi une donnée serait-elle absente?
- Un modèle simple
- Solutions AVEC annulabilité
- Solutions SANS annulabilité
- Impact sur la logique devant être utilisée dans le modèle
- Et, non, le problème n'est pas résolu!

DONNÉES ABSENTES

POURQUOI UNE DONNÉE SERAIT-ELLE ABSENTE?

- Discussion en classe.

DONNÉES ABSENTES

UN MODÈLE SIMPLE (PROPOSÉ PAR CODD)

○ NA

- L'information n'est **pas applicable**.
- Dans ce cas, l'utilisation de l'annulabilité est à remettre en question; une bonne modélisation permet généralement d'éviter d'y avoir recours.

○ IN

- L'information est **inconnue**.
- Dans ce cas, l'annulabilité pourrait être légitime; la question est de savoir comment la représenter pour que cela pose le moins de problèmes possible.

○ X

- L'information n'est **pas accessible**.
 - **À court terme** : dans ce cas, le gestionnaire transactionnel permet d'éviter l'utilisation de l'annulabilité. L'information peut être partagée entre plusieurs utilisateurs en même temps avec un contrôle des accès concurrents. L'exécution d'une transaction doit préserver la cohérence de la BD.
 - **À long terme** : équivalent à IN.

DONNÉES ABSENTES

IMPACT SUR LA LOGIQUE DES MODÈLES RELATIONNELS

- On en conclut que
 - le cas **X**, qui se réduit en **IN** ou **NA**, peut être ignoré ;
 - le cas **NA**, pourrait être ignoré, si la modélisation est (toujours) adéquate.
- Que faire du cas **IN** ?
 - Plusieurs théoriciens, dont Date, ont décidé de ne pas intégrer ce cas au modèle relationnel et de le traiter aussi par la modélisation, préservant ainsi la logique **bi-valuée** (faux, vrai) et donc booléenne.
 - Les membres du Comité de normalisation du langage SQL ont décidé d'intégrer ce cas à leur modèle et, conséquemment, d'utiliser une logique **tri-valuée** (faux, vrai, inconnu).

DONNÉES ABSENTES

SOLUTIONS SANS ANNULABILITÉ (DATE ET CIE)

○ Principes

- Séparer les propositions complètes des incomplètes.
- Conserver les causes d'absence séparément.

○ Pour un inventaire des techniques de modélisation

- <http://www.dcs.warwick.ac.uk/~hugh/TTM/Missing-info-without-nulls.pdf>

DONNÉES ABSENTES

SOLUTIONS AVEC ANNULABILITÉ (SQL ET CIE)

- Trois étapes à faire dans l'ordre :
 - si possible, corriger cette lacune à la source (dans la réalité);
 - si possible, modifier le modèle conceptuel pour en tenir compte;
 - sinon, introduire le concept d'annulabilité dans la théorie relationnelle, ce qui induit le recours
 - à l'un des deux artifices suivants :
 - un **marqueur** NUL (une propriété des attributs) ou
 - une **valeur** NULLE ajoutée à tous les domaines.
 - et à une logique **à 3 valeurs**
 - (afin de pouvoir définir l'égalité, essentielle aux opérations d'affectation, de restriction, de jointure, d'union...)

DONNÉES ABSENTES

EXEMPLE DE LOGIQUE À 3 VALEURS (SQL)

V : Vrai
F : Faux
I : Inconnu

A	B	$A \vee B$	$A \wedge B$	$\neg A$
V	V	V	V	F
V	I	V	I	F
V	F	V	F	F
I	V	V	I	I
I	I	I	I	I
I	F	I	F	I
F	V	V	F	V
F	I	I	F	V
F	F	F	F	V

REMARQUES

- Codd, dans son deuxième modèle, a proposé d'introduire deux marqueurs (IN et NA), ce qui induit une logique à 4 valeurs.
- D'autres chercheurs ont proposé d'autres systèmes logiques à 4 et même 5 valeurs.
- Humblement, nous nous en tiendrons, le plus souvent, à la logique booléenne dans le cadre du présent cours.

DONNÉES ABSENTES

Et non,
le problème
n'est pas
résolu!

Nous y
reviendrons
donc dans le
module BD028

De: oracle-acct_ww@oracle.com
Objet: Nom d'utilisateur de votre compte Oracle
Date: 2 octobre 2014 19:44
À: luc.lavoie@usherbrooke.ca

ORACLE

Cher/Chère NULL !,

Vous avez demandé à recevoir par email le nom d'utilisateur de votre compte Oracle.

Votre nom d'utilisateur est : **luc.lavoie@usherbrooke.ca**

Merci !

L'équipe de gestion des comptes Oracle

Mettez votre compte à jour :

> [Abonnez-vous aux communications](#) dédiées aux thèmes qui vous intéressent.

> [Devenez membre des communautés Oracle.](#)

> [Pour modifier votre adresse email, votre mot de passe](#) ou toute autre information de votre compte, cliquez sur le lien [Compte](#) en haut des pages Oracle.com.

Obtenir de l'aide

> Des questions ? [Aide \(page Account Help\)](#)

> Se connecter

- [Envoyer une demande d'aide](#)
- [profilehelp_ww@oracle.com](#)

Hardware and Software
ORACLE
Engineered to Work Together



Copyright © 2014, Oracle et/ou ses filiales.
Tous droits réservés.

[Aide \(page Account Help\)](#) | [Ne pas envoyer d'email](#) | [Mentions légales](#) | [Conditions d'utilisation](#) | [Confidentialité](#)

VOCABULAIRE

LANGAGES ET NOTATIONS

- Un script est la description textuelle d'un modèle de données conforme à une théorie (entité-association, relationnelle, etc.).
- Un diagramme est une représentation graphique (le plus souvent incomplète) d'un script.
- Nous utiliserons souvent en cours un langage textuel inspiré de l'algèbre relationnelle (sous le nom « langage relationnel »).
- SQL comprend un langage de description textuelle de schémas relationnels... c'est l'un des langages de programmation les plus utilisés au monde.
- Nous utilisons en cours deux notations graphiques
 - la notation graphique classique pour les diagrammes relationnels (des rectangles pour les relations, des flèches pour les clés référentielles);
 - la notation graphique classique pour les diagrammes entité-association (des rectangles, des ovales, des losanges et divers types de flèches).
- UML est une autre notation graphique fréquemment utilisée pour les diagrammes de classes or, plusieurs l'utilisent aussi pour représenter des diagrammes entité-association.

VOCABULAIRE

... ET ARTICLES ENCYCLOPÉDIQUES D'INTRODUCTION

- Une théorie (mathématique) est un ensemble d'affirmations dont certaines sont des axiomes et les autres des théorèmes démontrables à partir de ces axiomes et au moyen de règles d'inférence (exprimée à l'aide de la) logique.
 - <http://fr.wikipedia.org/wiki/Théorie>
- Un modèle est une représentation conforme à une théorie.
 - <http://fr.wikipedia.org/wiki/Modèle>
- Un langage (formel) est un formalisme permettant de décrire des propositions sémantiquement interprétables en termes d'un modèle.
 - http://fr.wikipedia.org/wiki/Langage_formel
- Une théorie peut être à l'origine de plusieurs modèles, un modèle de plusieurs langages, un langage de plusieurs dialectes.
- Pour en savoir plus sur le calcul des prédicats :
 - http://fr.wikipedia.org/wiki/Calcul_des_prédicats

THÉORÈME, COROLAIRES ET CONSTATS.

- Théorème d'incomplétude de Gödel :
 - *Toute théorie cohérente ayant un nombre fini d'axiomes exprimés dans un langage qui permet de décrire l'arithmétique comprend des propositions indécidables.*
 - Exemples :
 - l'arithmétique de Peano,
 - la théorie des ensembles.
- Par ailleurs, un modèle n'est pas forcément complet relativement à une théorie.
- Par ailleurs, un langage n'est pas forcément complet relativement à un modèle.

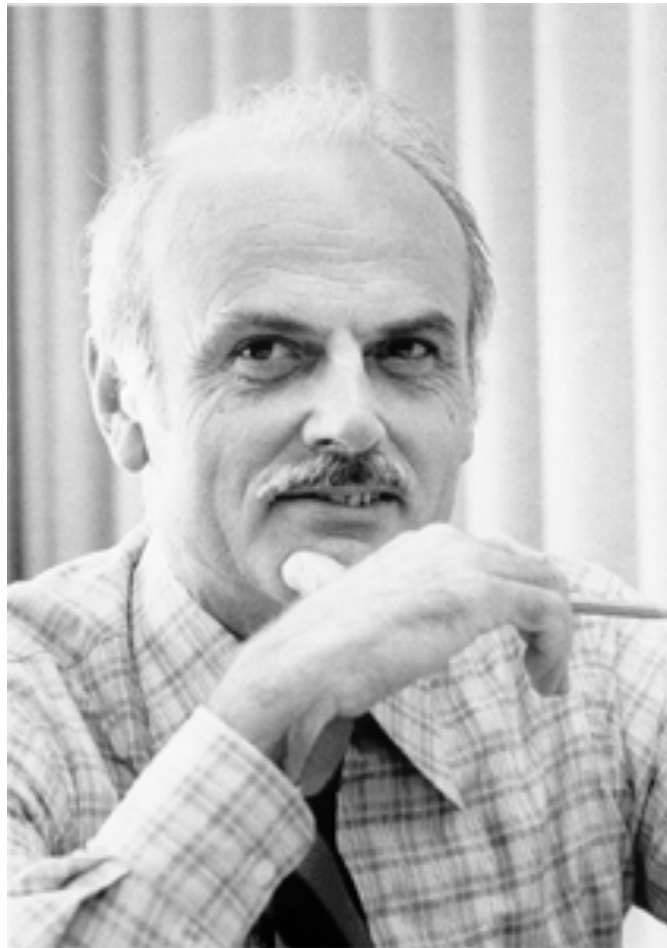
RÉFÉRENCES

- Théorie relationnelle
 - E.F. Codd. 1990.
The Relational Model for Database Management: Version 2.
Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.
 - C.J. Date, H. Darwen. 2007.
Databases, types and the relational model: the third manifesto.
Reading, Mass.: Addison-Wesley.
 - F. de Sainte Marie. 2013.
Bases de données relationnelles et normalisation : de la première à la sixième forme normale.
<https://fsmrel.developpez.com/basesrelationnelles/normalisation/>
 - H. Darwen. 2006.
How To Handle Missing Information Without Using NULL.
<http://www.dcs.warwick.ac.uk/~hugh/TTM/Missing-info-without-nulls.pdf>
- Manuels classiques
 - [C. J. Date 2004], chapitre 3.
 - [Elmasri and Navathe 2004], chapitre 4.
 - [Elmasri and Navathe 2011], chapitre 3.
 - [Elmasri and Navathe 2016], chapitre 8.
 - [Ullman and Widom 2008], chapitre 3.

LES COLLES DU PROF

- Quelles différences existe-t-il entre
 - un domaine et type?
 - un tuple et une relation?
 - une relation et une variable de relation?
 - un schéma et une base de données?
 - une théorie et un modèle?
 - un modèle et un langage?
 - une clé candidate et une superclé?
- Quels sont les opérateurs de base proprement relationnels?
- En quoi se distinguent-ils des opérateurs ensemblistes?

EDGAR FRANK CODD ET CHRISTOPHER J. DATE



https://en.wikipedia.org/wiki/Edgar_F._Codd



Photo of Chris Date by Douglas Robertson, Edinburgh

https://en.wikipedia.org/wiki/Christopher_J._Date

