Modélisation de données

Un modèle relationnel… en 10 diapositives simples

TMR_02_SYN
v250a

2025-01-14

Christina.Khnaisser@USherbrooke.ca
Luc.Lavoie@USherbrooke.ca

Rappels (1)

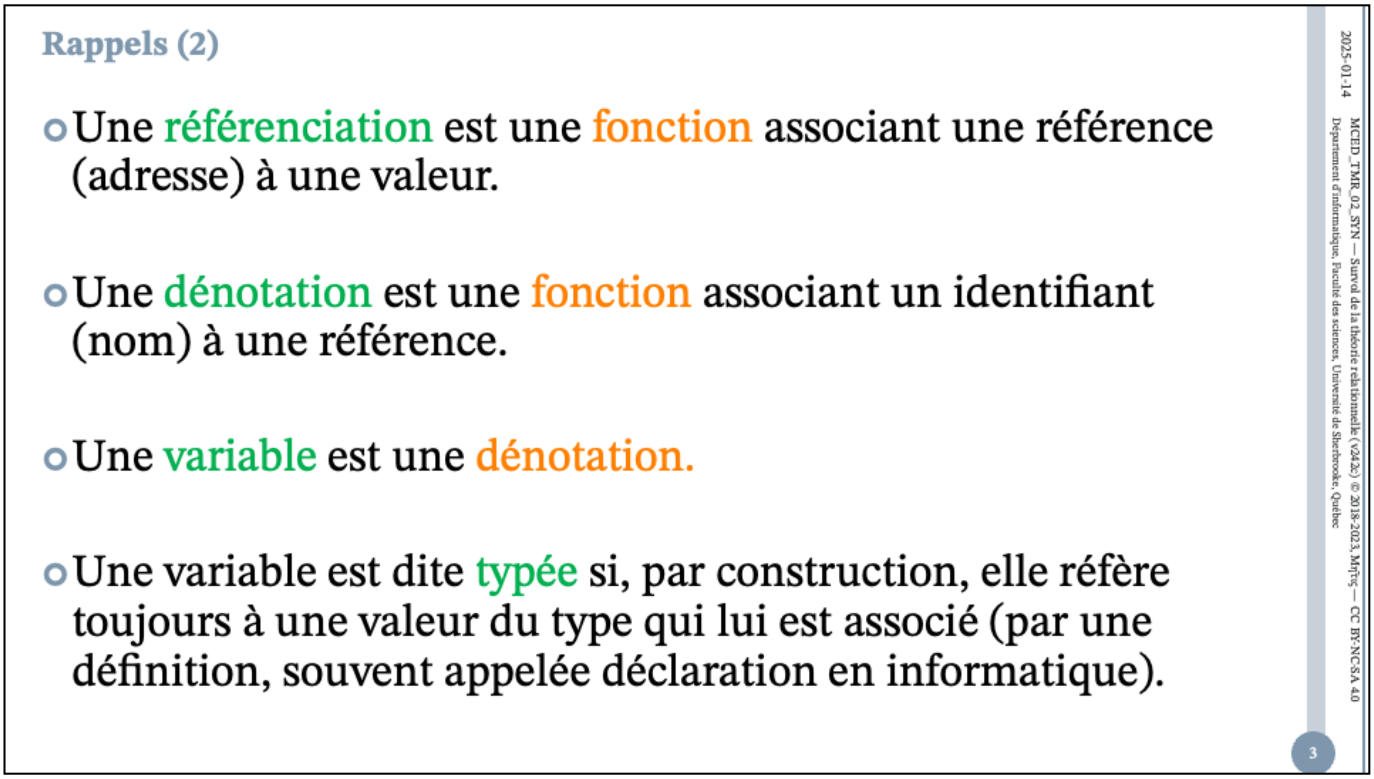
- Un type de base est un ensemble *fini* de valeurs *propres*.
- Un sous-type est un sous-ensemble d'un type de base déterminé par une contrainte (qui en restreint les valeurs).
- Une relation de degré *n* est un sous-ensemble du produit cartésien de *n* types:

$$D_1 \times D_2 \times \ldots \times D_n$$

Pourquoi fini?
Parce qu'en informatique, nous considérons uniquement les structures finies...
tant que nous n'aurons pas de dispositifs de calcul ayant des ressources infinies.

Pourquoi propre ?
Pour pouvoir déterminer avec sans ambiguïté le type de la valeur et donc les opérations qui lui sont applicables.

De la définition du sous-type, on déduit le corolaire que l'ensemble de valeurs qu'il représente est un sous-ensemble d'un type de base et conséquemment qu'il est lui-même fini.



Élaborer et motiver les définitions (voir Stoy et Strachey)

Une dénotation est une fonction associant un nom (identifiant) à une référence.

Une variable est une dénotation référant à une valeur d'un type,

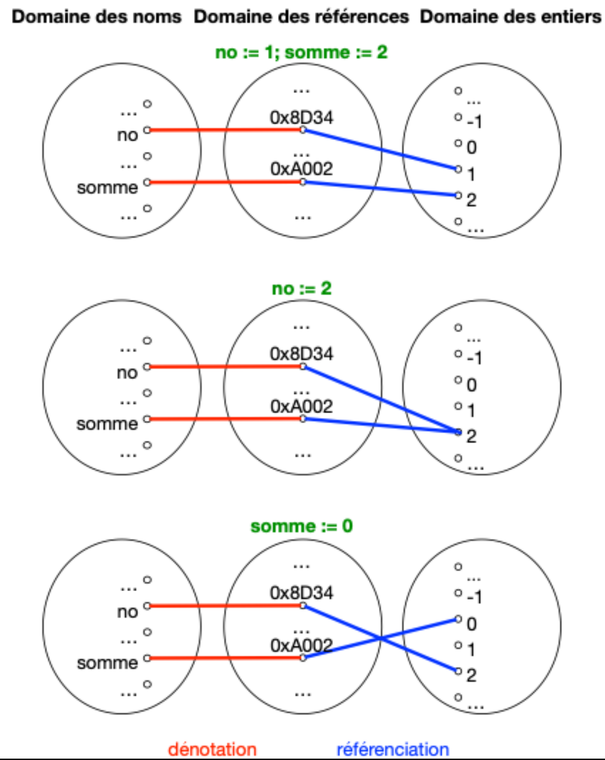
la référence associée peut être modifiée grâce à une opération appelée affectation.

L'admission d'un type de base dont les valeurs seraient des références départagent deux classes de modèles: les modèles «à pointeurs» des autres.

Entre les modèles d'une même classe, la gestion des références (les opérations disponibles) est souvent un facteur de différenciation.

Les modèles relationnel de Codd et tout comme celui de Date sont «sans pointeurs». Le modèle SQL est «avec pointeur».

Variable et affectation



Fundements — Attribut

- 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.

Un attribut correspond à une dimension de l’observation, à un «aspect» d’un fait.

Un tuple correspond à une observation, un «fait».

La relation est un ensemble d’observations de même type, de même «nature».

Deux tuples sont de même type si leurs entêtes sont les mêmes.

Transposé dans le domaine de la logique

- un tuple est une proposition (un énoncé vrai sur le monde);
- une relation est un prédicat (...).

Un prédicat peut être défini

- par énumération (l’ensemble de tous les énoncés vrais et eux seuls);
- par compréhension (la caractérisation nécessaire et suffisante des relations entre les variables).

Une base de données est un ensemble de variable de relation définies par leurs valeurs (donc par énumération) et leurs contraintes (donc par compréhension).

En conclusion, une base de données est la représentation d’un système logique.

Dans la pratique, il est souvent difficile d’établir un ensemble de contraintes nécessaires et suffisantes.

Mais on tente de s’en approcher le plus possible.

Pour une théorie des types plus complète, voirIFT 232, IFT 339 et IGE 487 (entre autres).

Fondements — Tuple

- Soit a_i des identifiants distincts et D_j des types, un tuple t est défini comme suit:
  - $t \triangleq (\{a_1{:}D_1, a_2{:}D_2, \ldots, a_n{:}D_n\}; \{(a_1,v_1), (a_2,v_2), \ldots, (a_n,v_n)\})$
  - avec $\forall\ i: 1 \leq i \leq deg(t) \Longrightarrow val(t, a_i) \in def(t, a_i)$
- où

$def(t) = \{a_1{:}D_1, a_2{:}D_2, \ldots, a_n{:}D_n\}$	entête de t
$def(t, a_i) = D_i$	type de l'attribut a_i de t
$val(t) = \{(a_1,v_1), (a_2,v_2), \ldots, (a_n,v_n)\}$	valeur de t
$val(t, a_i) = v_i$	valeur de de l'attribut a_i de t
$deg(t) = n$	degré de t
$id(t) = \{a_1, a_2, \ldots, a_n\}$	les identifiants d'attributs de t

Un tuple est la représentation d'une proposition logique (un énoncé vrai sur le monde).

Un tuple correspond à une observation, un «fait».

Un attribut correspond à une dimension de l'observation, à un «aspect» d'un fait.

La relation est un ensemble d'observations de même type, de même «nature».

Deux tuples sont de même type si leurs entêtes sont les mêmes.

Transposé dans le domaine de la logique

- un tuple est une proposition (un énoncé vrai sur le monde);
- une relation est un prédicat (…).

Un prédicat peut être défini

- par énumération (l'ensemble de tous les énoncés vrais et eux seuls);
- par compréhension (la caractérisation nécessaire et suffisante des relations entre les variables).

Une base de données est un ensemble de relations définies par leurs valeurs (donc par énumération) et leurs contraintes (donc par compréhension). En conclusion, une base de données est la représentation d'un système logique.

Le cas deg(t) = 0, est important.

Il existe un seul tuple possible (pourquoi?):

- t0 = ({}, {})

Notation simplifiée fondée sur l'ordre d'énumération des attributs
(les identifiants d'attributs et leurs types étant déterminés par ailleurs):

- t = <v_1, v_2, ..., v_n>

Finalement, la notation $t.a_i$ désigne l'attribut a_i:D_i dans def(t).

Fondements — Relation

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

Une relation est la représentation d'un prédicat logique.

Un prédicat peut être défini

- par énumération (l'ensemble de tous les énoncés vrais et eux seuls);
- par compréhension (la caractérisation nécessaire et suffisante des relations entre les variables).

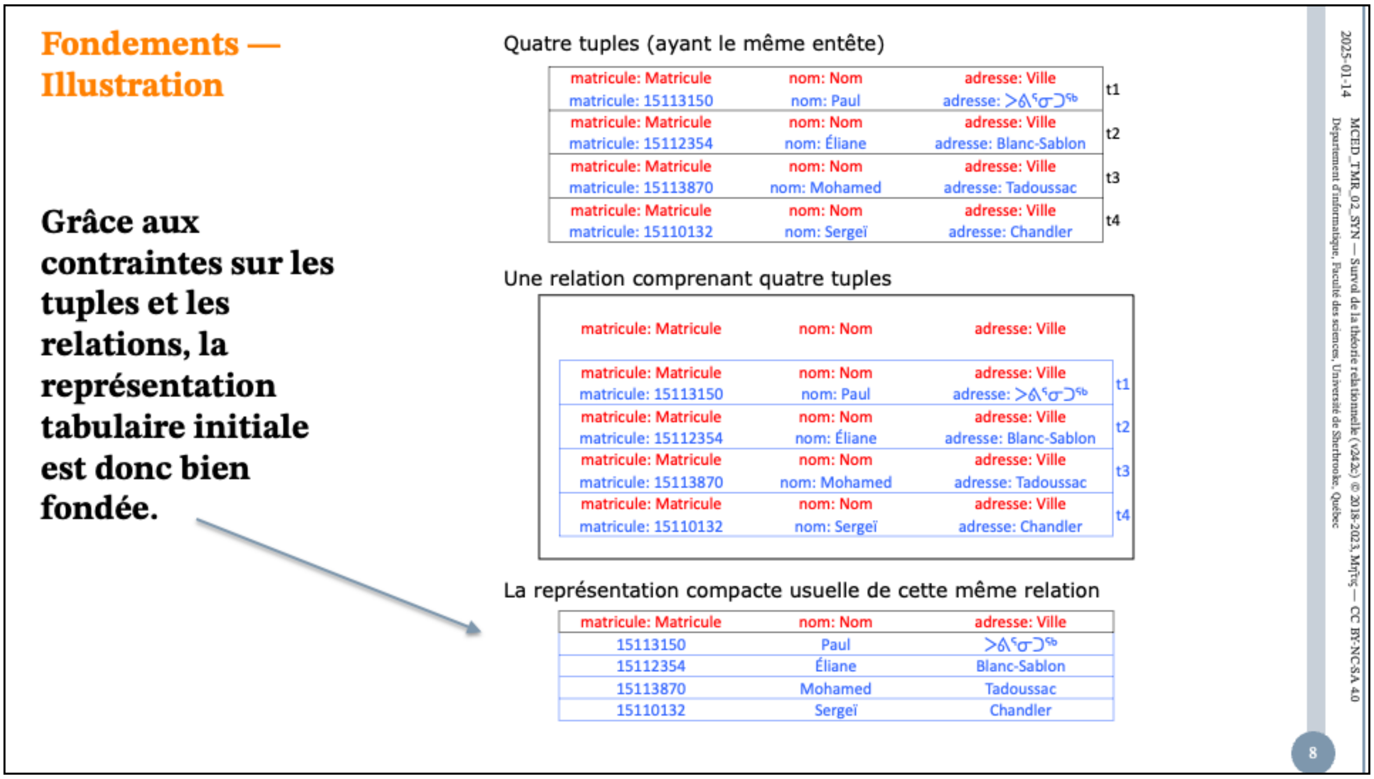
Le cas deg(R) = 0 est important.

Il existe deux relations possibles (pourquoi?):

- R0 = ({}, {})
- R1 = ({}, {t0})

et elles sont très importantes, comme le zéro et le un pour les entiers!

Finalement, la notation R.a_i désigne l'attribut a_i:D_i dans def(R).



Si on se reporte aux définitions des types de la figure du préambule.

Si on considère que l'ensemble des caractères C contient 256 caractères différents.

Combien y a-t-il de valeurs de tuple différentes?

$n = 10^8 \times 256^{60} \times 256^{40} = 10^8 \times 2^{800} \approx 6{,}67 \times 10^{248}$

Combien y a-t-il de valeurs de relation différentes?

tout tuple peut être présent ou non, donc

$m = 2^n \approx 2\hat{}(6{,}67\mathrm{x}10^{248}) \approx 10\hat{}(2\mathrm{x}10^{248})$

Combien y a-t-il d'atomes dans l'Univers?

Plusieurs scientifiques estiment qu'il est de l'ordre de $10^{80} \approx 10\hat{}(8\mathrm{x}10^{1})$

(voir par exemple https://www.science-et-vie.com/galerie/sait-on-combien-il-y-a-d-atomes-dans-l-univers-6154)

Fondements — Opérations de base

Restriction

R σ cond

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

Projection

R π {A, C}

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

Jointure (naturelle)

R ⋈ S

A	B
a1	b1
a2	b1
a3	b3
a4	b4

⋈

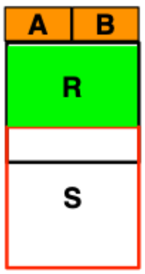
B	C
b1	c1
b2	c2
b3	c3
b3	c4

=

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

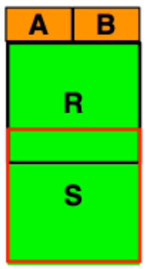
Différence

R – S



Union

R ∪ S



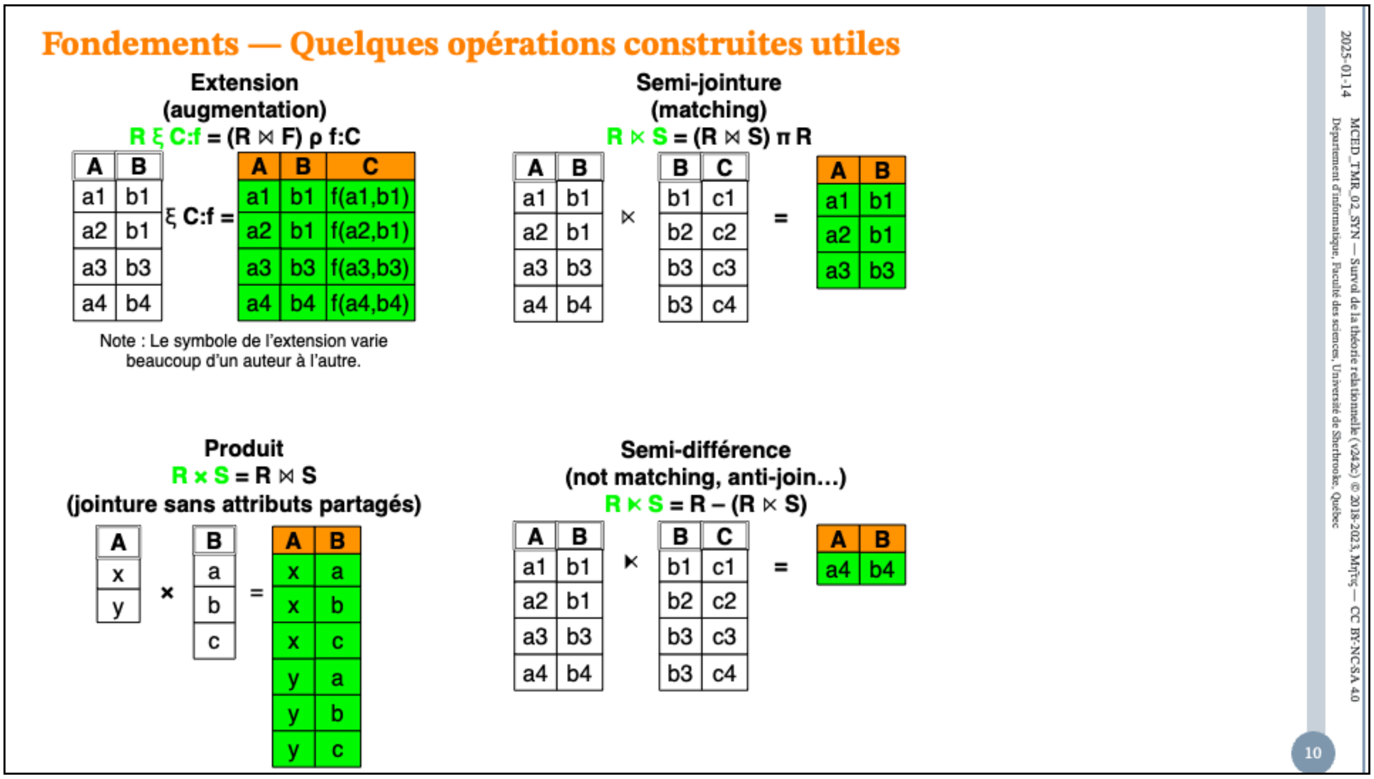
Renommage

R ρ A:C

A	B
a1	b1
a2	b2
a3	b3

ρ A:C =

C	B
a1	b1
a2	b2
a3	b3



Pour l'extension, F est la relation correspondant à la fonction f et f' est l'attribut de F correspondant à l'image de f. Au besoin les opérandes de F pourront être renommés pour correspondre aux attributs idoines de R.

Pour le produit, on suppose que R et S n'on aucun attributs partagés (sinon ceux-ci doivent être renommés au préalables).

Remarquons que (par définition)

$(R \ltimes S) \cup (R \rhd S) = R$

et

$(R \ltimes S) \cap (R \rhd S) = \varnothing$

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 définie par un prédicat représentée par une cellule (un champ) dans un tableau.
contrainte	expression logique.
relvar	variable référant une (valeur de) relation.
modèle logique	un ensemble de définitions de relvar et un ensemble de définitions de contrainte.
base de données	un ensemble de relvars conformes à un modèle logique (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.

Tuple: une proposition avérée qui représente un fait dans la réalité

https://fr.wikipedia.org/wiki/Calcul_des_prédicats
https://fr.wikipedia.org/wiki/Calcul_des_propositions

Attention! il y a variable (logique) et variable (informatique)…
Les variables dont on parle dans cette diapositive sont des variables informatiques.

Département d'informatique, Faculté des sciences, Université de Sherbrooke, Québec