



Base de données

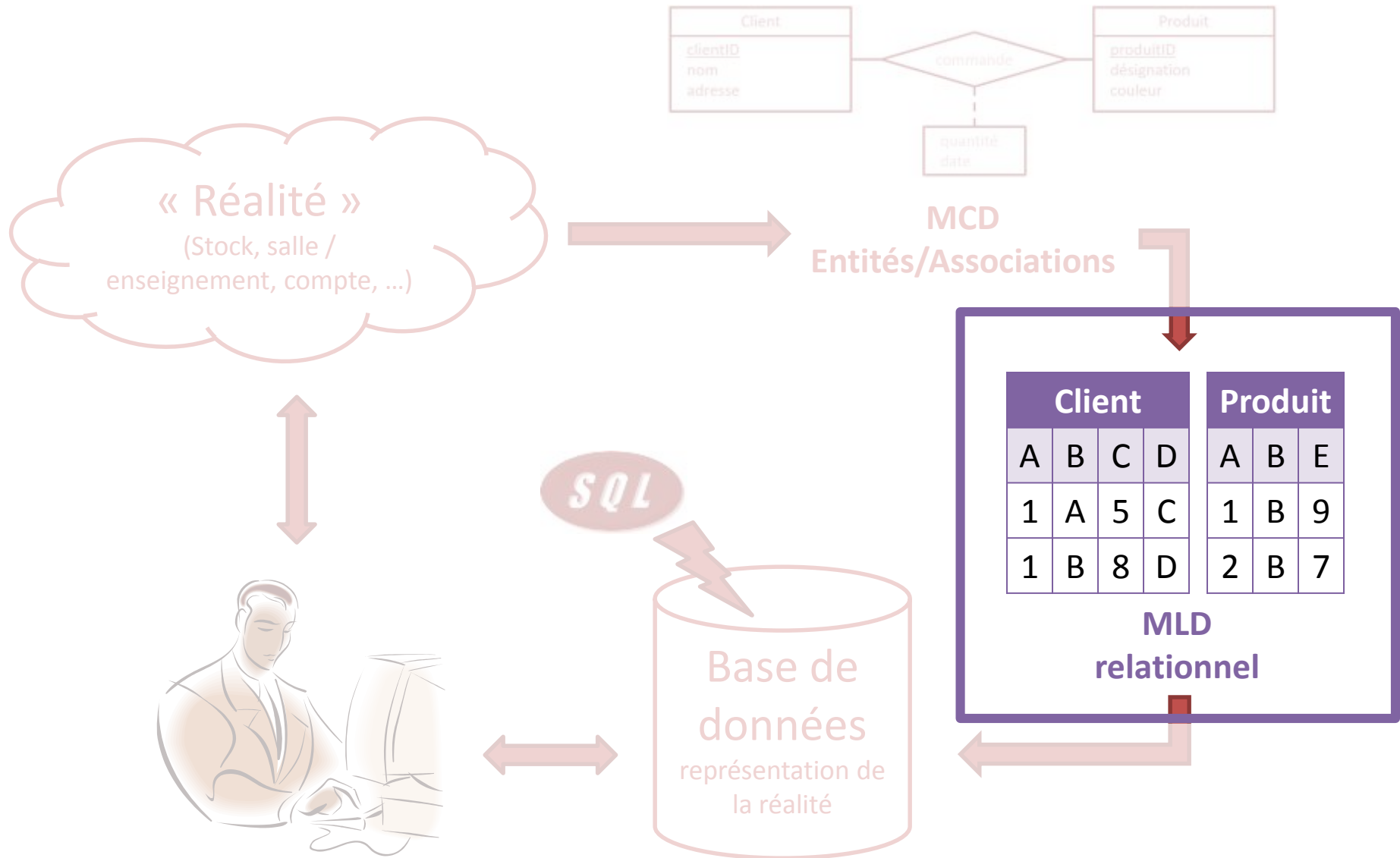
Modèle relationnel

L3 Informatique

Antoine Spicher

`antoine.spicher@u-pec.fr`

Big Picture



Plan



- Modèle logique de données
- Algèbre relationnelle
- En pratique...

Rappels

■ Modélisation et base de données

□ Modélisation des *données*

■ Système

Informations réelles à stocker dans la BD

■ Modèle

Organisation/structuration des informations

■ Formalisme

Quel formalisme ?

□ Modélisation de la *base de données*

■ Système

Stockage physique des données

■ Modèle

Abstraction réaliste du stockage des données

■ Formalisme(s)

Théorie des ensembles, diagrammes, ...

Schéma

*Modèle
de données
database model*

Rappels

■ *Modélisation logique de données (MLD)*

□ Modélisation des *données*

■ *Système*

Informations réelles à stocker dans la BD

■ *Modèle*

Organisation/structuration des informations

■ *Formalisme*

Utilisation du modèle de données

□ Modélisation de la *base de données*

■ *Système*

Stockage physique des données

■ *Modèle*

Abstraction réaliste du stockage des données

■ *Formalisme(s)*

Théorie des ensembles, diagrammes, ...

Schéma

La modélisation
des données
reflète leur
stockage
physique

*Modèle
de données
database model*

Représentation tabulaire

■ Introduction par l'exemple

- Représentation d'étudiants caractérisés par :
 - Leur numéro d'étudiant unique
 - Leur nom
 - Leur date de naissance
 - Leur niveau d'étude
- Données regroupées dans une *table* à la façon d'un fichier Excel

Etudiant	N°	Nom	Date naiss.	Niveau
	123456	Toto	25/01/93	L3
	542355	Tata	27/04/92	L2
	242643	Titi	14/08/89	M2

Représentation tabulaire

■ Vocabulaire du MLD

- **Relation** (table) : ensemble de données de même type
 - Identité *spécifique*
 - Identifiée par un **nom**
 - *Tous les étudiants peuvent être décrits de la même façon*

Etudiant	N°	Nom	Date naiss.	Niveau
	123456	Toto	25/01/93	L3
	542355	Tata	27/04/92	L2
	242643	Titi	14/08/89	M2

Relation

Représentation tabulaire

■ Vocabulaire du MLD

- **Relation** (table) : ensemble de données de même type
- **Schéma** (entête) : caractérisation des données de la table
 - Identité *qualitative*
 - *Chaque étudiant peut être caractérisé par un ensemble d'informations (qualités) de taille fixe*

Schéma

Etudiant	N°	Nom	Date naiss.	Niveau
	123456	Toto	25/01/93	L3
	542355	Tata	27/04/92	L2
	242643	Titi	14/08/89	M2

Notation textuelle : **Etudiant**(num, nom, date_naiss, niveau)

Représentation tabulaire

■ Vocabulaire du MLD

- **Relation** (table) : ensemble de données de même type
- **Schéma** (entête) : caractérisation des données de la table
- **Attribut** (colonne) : nom et domaine de valeurs d'une information
 - Chaque information est identifiée par un *nom*
 - Chaque information est prise dans un *domaine de valeurs possibles*
 - *Un étudiant possède une date de naissance valide dont le format est jj/mm/aa*

Etudiant	N°	Nom	Date naiss.	Niveau
	123456	Toto	25/01/93	L3
	542355	Tata	27/04/92	L2
	242643	Titi	14/08/89	M2

Attribut

Représentation tabulaire

■ Vocabulaire du MLD

- **Relation** (table) : ensemble de données de même type
- **Schéma** (entête) : caractérisation des données de la table
- **Attribut** (colonne) : nom et domaine de valeurs d'une information
- **Tuple** (ligne) : valeurs des attributs d'une donnée particulière
 - Chaque entrée caractérise *une et une seule entité* du monde réel
 - *La donnée du numéro, du nom, de la date de naissance et du niveau d'un étudiant permet de le distinguer de tous les autres*

Etudiant	N°	Nom	Date naiss.	Niveau
	123456	Toto	25/01/93	L3
	542355	Tata	27/04/92	L2
Tuple	242643	Titi	14/08/89	M2

Représentation tabulaire

■ Vocabulaire du MLD

- **Relation** (table) : ensemble de données de même type
- **Schéma** (entête) : caractérisation des données de la table
- **Attribut** (colonne) : nom et domaine de valeurs d'une information
- **Tuple** (ligne) : valeurs des attributs d'une donnée particulière
- **Valeur** (case) : croisement entre un tuple et un attribut
 - La *date de naissance* de l'*étudiant numéro 242643* est le *14 février 1989*

Etudiant	N°	Nom	Date naiss.	Niveau
	123456	Toto	25/01/93	L3
	542355	Tata	27/04/92	L2
	242643	Titi	14/08/89	M2

Contenu

Représentation tabulaire

■ Vocabulaire du MLD

- ❑ **Relation** (table) : ensemble de données de même type
- ❑ **Schéma** (entête) : caractérisation des données de la table
- ❑ **Attribut** (colonne) : nom et domaine de valeurs d'une information
- ❑ **Tuple** (ligne) : valeurs des attributs d'une donnée particulière
- ❑ **Valeur** (case) : croisement entre un tuple et un attribut

Schéma				
Etudiant	N°	Nom	Date naiss.	Niveau
	123456	Toto	25/01/93	L3
	542355	Tata	27/04/92	L2
Tuple	242643	Tit	14/08/89	M2

Contenu

Attribut

Relation

Représentation tabulaire



■ Faiblesses de la représentation graphique

Problématiques des représentations graphiques

- Règles de construction graphique **non-closes/incohérentes**
Représentation « dessinable » de sémantique impropre
- Règles de construction graphique **non-couvrantes**
Situations autorisées par la sémantique mais non-dessinables
- Absence de **sémantique formelle** bien définie

Représentation tabulaire

■ Faiblesses de la représentation graphique

- Ordre des tuples et des attributs
 - Absence d'ordre entre les tuples

L'ordre dans lequel les données sont entrées est oublié

Etudiant	N°	Nom	Date naiss.	Niveau
	123456	Toto	25/01/93	L3
	542355	Tata	27/04/92	L2
	242643	Titi	14/08/89	M2



Etudiant	N°	Nom	Date naiss.	Niveau
	542355	Tata	27/04/92	L2
	123456	Toto	25/01/93	L3
	242643	Titi	14/08/89	M2

Représentation tabulaire

■ Faiblesses de la représentation graphique

- Ordre des tuples et des attributs
 - Absence d'ordre entre les attributs

Différence entre les types produits \times et enregistrement

Etudiant	N°	Nom	Date naiss.	Niveau
	123456	Toto	25/01/93	L3
	542355	Tata	27/04/92	L2
	242643	Titi	14/08/89	M2



Etudiant	Date naiss.	Nom	Niveau	N°
	25/01/93	Toto	L3	123456
	27/04/92	Tata	L2	542355
	14/08/89	Titi	M2	242643

Représentation tabulaire

■ Faiblesses de la représentation graphique

1. *Travailler modulo permutations de tuples et d'attributs*

Etudiant	N°	Nom	Date naiss.	Niveau
	123456	Toto	25/01/93	L3
	542355	Tata	27/04/92	L2
	242643	Titi	14/08/89	M2



Etudiant	Date naiss.	Nom	Niveau	N°
	27/04/92	Tata	L2	542355
	25/01/93	Toto	L3	123456
	14/08/89	Titi	M2	242643

Représentation tabulaire

■ Faiblesses de la représentation graphique

□ Cas d'un attribut *non renseigné* (case vide)

- Attribut *non défini* pour le tuple en question

Ex: champ « nom de jeune fille » dans un formulaire administratif

- Attribut *non encore connu*

Ex: date de naissance/de mort, un parent, etc. dans un arbre généalogique

- ...

Etudiant	N°	Nom	Date naiss.	Niveau
	123456	Toto	25/01/93	L3
	542355	Tata		L2
	242643	Titi	14/08/89	M2

Représentation tabulaire



■ Faiblesses de la représentation graphique

1. *Travailler modulo permutations de tuples et d'attributs*
2. *Représenter l'absence d'information*

Représentation tabulaire

■ Faiblesses de la représentation graphique

□ Cas de deux tuples identiques

- Attributs insuffisants pour distinguer les entités réelles
- Identité qualitative \nRightarrow identité numérique
- ...

Etudiant	Nom	Date naiss.	Niveau
...
Toto	25/01/93	L3	
		...	
Toto	25/01/93	L3	
...

INTERDIT

Représentation tabulaire



■ Faiblesses de la représentation graphique

1. *Travailler modulo permutations de tuples et d'attributs*
2. *Représenter l'absence d'information*
3. *Respecter l'identité numérique*

Représentation tabulaire

■ Faiblesses de la représentation graphique

□ Cas de valeurs incohérentes

■ Domaines des attributs non-explicites

Chaîne de caractères pour une date (format), précision pour les nombres, ...

■ Contraintes sémantiques non représentables pour certaines valeurs

Etudiants caractérisés par leur numéro d'étudiant **unique**, etc.

■ Utilisation de valeurs non-atomiques (attributs multi-valués et composites)

Etudiant	N°	Nom	Date naiss.	Niveau
...
123456	Toto	01-25-1993	L3	
542355		27/04/92	L2	
123456	Tit	14/08/89	M2	
...

INTERDIT

Représentation tabulaire



■ Faiblesses de la représentation graphique

1. *Travailler modulo permutations de tuples et d'attributs*
2. *Représenter l'absence d'information*
3. *Respecter l'identité numérique*
4. *Représenter l'intégrité/la régularité des données*

Représentation tabulaire

■ Faiblesses de la représentation graphique

- Cas de liens sémantiques entre différentes tables
 - Résoudre les cas de valeurs non-atomiques

Etudiant	N°	Nom	Date naiss.	Niveau	
...	
123456	Toto	25/01/93	L3	GP1	
			L2	GP6	
...	

Représentation tabulaire

■ Faiblesses de la représentation graphique

□ Cas de liens sémantiques entre différentes tables

■ Résoudre les cas de valeurs non-atomiques

1^{ère} solution : fusionner/fractionner

Etudiant	N°	Nom	Date naiss.	Niveau
...
123456	Toto	25/01/93	L3-GP1	
123456	Toto	25/01/93	L2-GP6	
...

Redondance introduite

Représentation tabulaire

■ Faiblesses de la représentation graphique

□ Cas de liens sémantiques entre différentes tables

■ Résoudre les cas de valeurs non-atomiques

2^{ème} solution : stocker les valeurs structurées dans une autre table

Etudiant	N°	Nom	Date naiss.
	542355	Tata	27/04/92
	123456	Toto	25/01/93
	242643	Titi	14/08/89

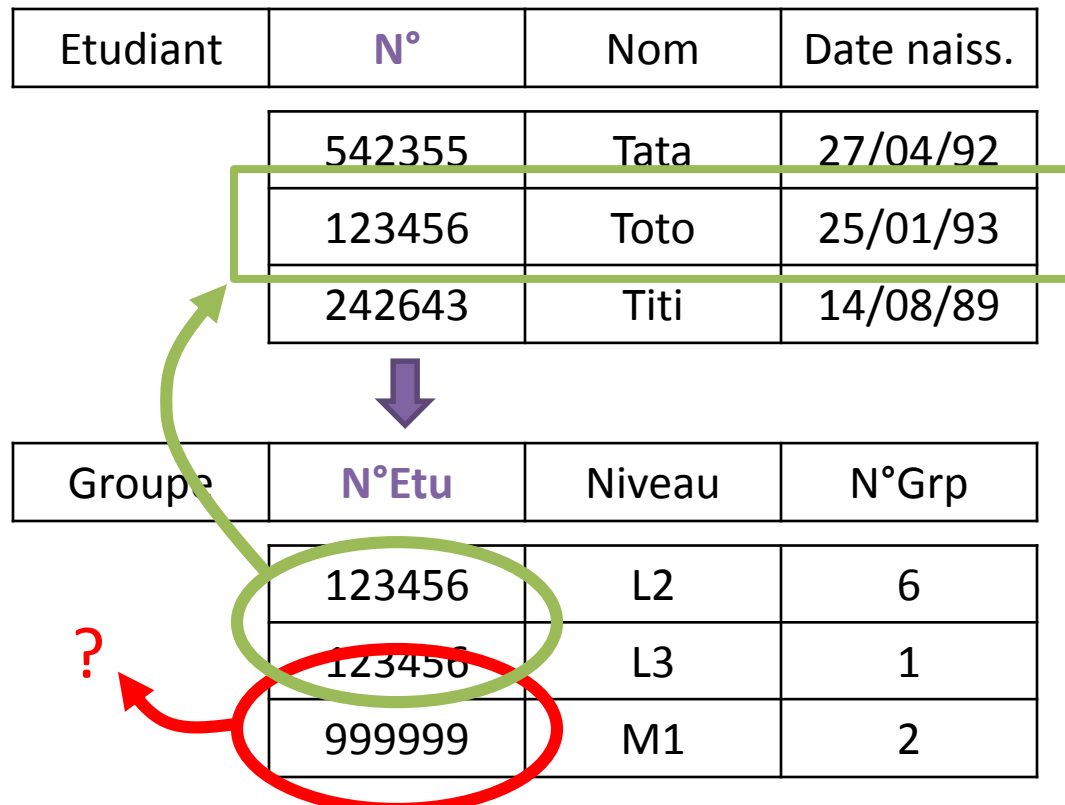
Groupe	N°Etu	Niveau	N°Grp
	123456	L2	6
	123456	L3	1
	999999	M1	2

Représentation tabulaire

■ Faiblesses de la représentation graphique

- Cas de liens sémantiques entre différentes tables
 - Résoudre les cas de valeurs non-atomiques
 - Faire référence à une entrée dans une autre table

- Les attributs **N°** et **N°Etu** ont la même sémantique
- L'attribut **N°** *seul* doit permettre d'identifier un étudiant
- Tout **N°Etu** doit correspondre à un étudiant **existant**



Représentation tabulaire



■ Faiblesses de la représentation graphique

1. *Travailler modulo permutations de tuples et d'attributs*
2. *Représenter l'absence d'information*
3. *Respecter l'identité numérique*
4. *Représenter l'intégrité/la régularité des données*
5. *Représenter les liens sémantiques entre les données*

Représentation tabulaire

■ Faiblesses de la représentation graphique

□ Effectuer des recherches

■ Ecriture de procédures de calcul *ad hoc*

Rechercher les noms des étudiants du groupe 1 de L3

```
Pour chaque tuple de Groupe Faire
| Si N°Grp = 1 et Niveau = L3 Alors
| | Pour chaque tuple d'Etudiant Faire
| | | Si N° = N°Etu Alors
| | | | Afficher Nom
| | | Fin Si
| | Fin Pour
| Fin Si
Fin Pour
```

Groupe	N°Etu	Niveau	N°Grp
--------	-------	--------	-------

...
123456	L3	1
...

Etudiant	N°	Nom	Date naiss.
----------	----	-----	-------------

...
123456	Toto	25/01/93
...

Représentation tabulaire

■ Faiblesses de la représentation graphique

□ Effectuer des recherches

- Ecriture de procédures de calcul *ad hoc*

Rechercher les noms des étudiants du groupe 1 de L3

- **Absence** de langage **dédié** à l'expression des calculs

□ *Domain Specific Language* (DSL)

□ Sémantique **formelle cohérente**

tout calcul peut être exprimé par un programme, et

tout programme décrit un calcul correct dans le cadre du modèle

- Nécessité de **fonder mathématiquement** le modèle relationnel

Représentation tabulaire



■ Faiblesses de la représentation graphique

1. *Travailler modulo permutations de tuples et d'attributs*
2. *Représenter l'absence d'information*
3. *Respecter l'identité numérique*
4. *Représenter l'intégrité/la régularité des données*
5. *Représenter les liens sémantiques entre les données*
6. *Faire du calcul sur les tables*
7. ...

Besoin d'une algèbre des relations

Plan



- Modèle logique de données
- Algèbre relationnelle
- En pratique...

Qu'est-ce qu'une algèbre

■ Origine (Wikipédia)

- Du mot arabe : **al-jabr** (réduction d'une fracture, restauration)
Latinisé en **algebra** (réunion de morceaux, reconstruction, connexion)
- **Al-Khawarizmi**
 - Mathématicien persan du IX^{ème}
Latinisé en **algoritmi**, donnant plus tard le mot **algorithme**
 - *Kitāb al-mukhtaṣar fī ḥisāb al-jabr wa-l-muqābala*
« Abrégé du calcul par la **restauration** et la comparaison »



Qu'est-ce qu'une algèbre

■ Origine (Wikipédia)

- Du mot arabe : **al-jabr** (réduction d'une fracture, restauration)
Latinisé en **algebra** (réunion de morceaux, reconstruction, connexion)
- Al-Khawarizmi
 - Mathématicien persan du IXème
Latinisé en **algoritmi**, donnant plus tard le mot *algorithme*
 - *Kitāb al-mukhtaṣar fī ḥisāb al-jabr wa-l-muqābala*
« Abrégé du calcul par la **restauration** et la comparaison »
 - Introduction du concept d'**équation**
 - Égalité entre expressions mathématiques comportant une inconnue
 - Développement des techniques de résolution
- Extension du concept
 - Évolution des nombres, résolution de systèmes d'équations, polynômes
 - Objets abstraits : structures algébriques (monoïdes, groupes, corps, ...)

Qu'est-ce qu'une algèbre

■ Définition

- Structure algébrique : *ensemble muni d'opérations*
- **Objets** : ensemble E d'éléments sous-jacent à l'algèbre
- **Opérations** ou lois : applications de $E^n \rightarrow E$
 - n est appelée l'arité de l'opération
 - On parle de constante quand $n = 0$
- **Axiomes** : identités entre compositions d'opérations
 - Algèbre libre : sans axiome
 - Ex : associativité, commutativité, distributivité, etc.
- **Calcul dans l'algèbre**
Variables, fonctions, (systèmes d') équation(s), etc.

Qu'est-ce qu'une algèbre

■ Exemple : logique classique

- **Objets** : valeurs booléennes

- **Opérations**

 - Nullaires : true et false

 - Unaire : \neg

 - Binaires : \wedge , \vee

- **Axiomes**

 - Tables de vérité

 - Associativité et commutativité du \wedge

 - Involutivité de \neg

 - Neutre, absorbant : $\text{true} \wedge p = p \wedge \text{true} = p$, $\text{false} \wedge p = p \wedge \text{false} = \text{false}$

 - Définitions : \vee (loi de de Morgan), *ou exclusif*, implication, etc.

- **Calculs**

Satisfiabilité des formules, preuve automatique

Qu'est-ce qu'une algèbre

■ Exemple : groupe commutatif

□ **Objets** : *mots* (i.e., éléments) d'un groupe (par exemple \mathbb{Z})

□ **Opérations**

- Nullaire : zero (0 pour \mathbb{Z}), etc.
- Unaire : inverse ($-n$ pour tout $n \in \mathbb{Z}$)
- Binaire : loi de composition $+$ (l'addition sur les entiers)

□ **Axiomes**

- Associativité et commutativité du $+$
- Involutivité de $-$
- Neutre : $0 + n = n + 0 = n, n + (-n) = (-n) + n = 0$

□ **Calculs**

Problème du mot

Qu'est-ce qu'une algèbre

■ Exemple : langages formels

□ **Objets** : ensembles L de mots formés sur un alphabet Σ

□ **Opérations**

- Nullaires : langage vide \emptyset , épsilon $\varepsilon: . \mapsto \{\varepsilon\}$, $a: . \mapsto \{a\}$ pour tout $a \in \Sigma$
- Unaires : complémentation \bar{L} , étoile de Kleene L^*
- Binaires : union $L_1 \cup L_2$, concaténation $L_1 \cdot L_2$

□ **Axiomes**

- Définitions : intersection $L_1 \cap L_2 = \overline{\bar{L}_1 \cup \bar{L}_2}$, étoile $L^* = \bigcup_n L^n$
- Involutivité : $\bar{\bar{L}} = L$
- Associativités : $(L_1 \cup L_2) \cup L_3 = L_1 \cup (L_2 \cup L_3)$, $(L_1 \cdot L_2) \cdot L_3 = L_1 \cdot (L_2 \cdot L_3)$
- Distributivités : $(L_1 \cup L_2) \cdot L_3 = L_1 \cdot L_3 \cup L_2 \cdot L_3$
- Neutre, absorbant : $\emptyset \cup L = L \cup \emptyset = L$, $\varepsilon \cdot L = L \cdot \varepsilon = L$, $\emptyset \cdot L = L \cdot \emptyset = \emptyset$

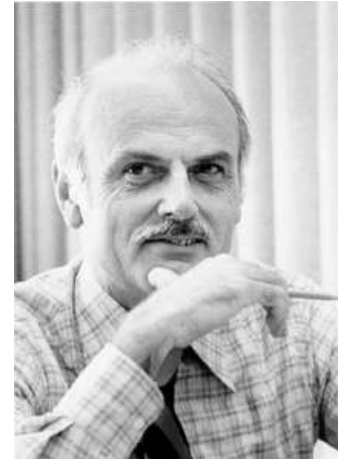
□ **Calculs**

Automates, grammaires, analyseurs LL et LR (cf. compilation)

Algèbre relationnelle

■ Origine

- Edgar Frank “Ted” Codd (1923-2003)
 - Directeur de recherche du centre IBM de San José
 - Prix Turing en 1981
- *A relational model of data for large shared data banks*
 - Article fondateur de 1970
 - **Indépendance** entre modèle et représentation interne
 - Critique de l'existant, proposition du modèle relationnel, formes normales, base d'un langage de programmation, opérateurs spécifiques



Edgar Frank "Ted" Codd
(source : Wikipédia)

Future users of large data banks must be protected from having to know how the data is organized in the machine (the internal representation). A prompting service which supplies such information is not a satisfactory solution. Activities of users at terminals and most application programs should remain unaffected when the internal representation of data is changed and even when some aspects of the external representation are changed. Changes in data representation will often be needed as a result of changes in query, update, and report traffic and natural growth in the types of stored information.

Existing noninferential, formatted data systems provide users with tree-structured files or slightly more general network models of the data. In Section 1, inadequacies of these models are discussed. A model based on n -ary relations, a normal form for data base relations, and the concept of a universal data sublanguage are introduced. In Section 2, certain operations on relations (other than logical inference) are discussed and applied to the problems of redundancy and consistency in the user's model.

Algèbre relationnelle

■ Définition

- Objets : relations

Définition à préciser

- Opérations

- Nullaires : les relations de la BD
- Unaires : ???
- Binaires : ???

- Axiomes : ???

- Calculs

Les équations sont les requêtes

Algèbre relationnelle : objets

■ Prérequis : les domaines de valeurs

□ Définition

Les ensembles de valeurs autorisées à être stockées dans la base de données

□ Synonymes : types ou types atomiques

□ Notations : $D, D', \dots, D_1, D_2, \dots$

Les valeurs seront notées $v, v', \dots, v_1, v_2, \dots, u, u', \dots, u_1, u_2, \dots$

□ Domaines classiques

■ Exemples

□ Booléens D_B

□ Numériques D_N : valeurs entières ou réelles

□ Chaînes de caractères D_S

□ Dates D_D

□ Types énumérés : tout ensemble fini de valeurs (cf. type enum de C)

■ Cf. types de base du langage SQL

Algèbre relationnelle : objets

■ Prérequis : les domaines de valeurs

□ Définition

Les ensembles de valeurs autorisées à être stockées dans la base de données

□ Cas particulier de l'absence de valeur (cases vides d'une table)

■ Définition de la valeur spéciale \perp

À l'origine absente de la théorie

■ Plusieurs sémantiques possibles

« donnée inconnue », « donnée inexistante », « donnée introuvable »

■ Valeur non typée (pas de domaine particulier : *tout domaine contient* \perp)

□ Extension des opérations classiques dans les domaines

□ Exemple 1 : domaine des chaînes de caractères

Chaines de caractères classiques avec la concaténation ($\perp \neq ""$)
toute concaténation avec \perp retourne \perp

□ Exemple 2 : domaines des booléens

Logique à trois valeurs (cf. le LID de SQL)

Algèbre relationnelle : objets

■ Prérequis : les tuples

- Définition : *Représentation formelle d'une entrée dans une BD*
- Notation : $t, t', \dots, t_1, t_2, \dots, s, s', \dots, s_1, s_2, \dots$

■ Prérequis : les attributs

- Définition : *Fonction nommée permettant d'interroger/décomposer un tuple*
- Spécification
 - Un nom (symbole) A
 - Un domaine D de valeurs
 - Un ensemble T de tuples
- Notation : $A, A', \dots, A_1, A_2, \dots, B, B', \dots, B_1, B_2, \dots$
- Compatibilité entre attributs

$$A: \begin{cases} T & \rightarrow \\ t & \mapsto A(t) = v \end{cases}$$

Deux attributs sont *compatibles* s'ils ont même *nom* et même *domaine*

Algèbre relationnelle : objets

■ Prérequis : les **tuples** et les **attributs**

Etudiant	N°	Nom	Date naiss.	Niveau
e_1	123456	Toto	25/01/93	L3
e_2	542355	Tata	27/04/92	L2
e_3	242643	Titi	14/08/89	M2
e_4	328494	Tutu	08/11/95	L1

$T =$

$\{e_1, e_2, e_3, e_4\}$

↓
Num

↓
Nom

↓
DNaiss

↓
Niv

Num : $T \rightarrow D_N$ Nom : $T \rightarrow D_S$ DNaiss : $T \rightarrow D_D$ Niv : $T \rightarrow \{L_1, \dots, M_2\}$

$e_1 \mapsto 123456$

$e_1 \mapsto \text{"Toto"}$

$e_1 \mapsto 25/01/93$

$e_1 \mapsto 25/01/93$

$e_2 \mapsto 542355$

$e_2 \mapsto \text{"Tata"}$

$e_2 \mapsto 27/04/92$

$e_2 \mapsto 27/04/92$

$e_3 \mapsto 242643$

$e_3 \mapsto \text{"Titi"}$

$e_3 \mapsto 14/08/98$

$e_3 \mapsto 14/08/98$

$e_4 \mapsto 328494$

$e_4 \mapsto \text{"Tutu"}$

$e_4 \mapsto 08/11/95$

$e_4 \mapsto 08/11/95$

Algèbre relationnelle : objets

■ Relation

□ Définition

Objets de l'algèbre relationnelle représentant les tables d'une BD

□ Formalisation : une relation $R = (...)$

... définition à compléter ...

□ Notations : $R, R', ..., R_1, R_2, ..., S, S', ..., S_1, S_2, ...$

Algèbre relationnelle : objets

■ Relation : extension

□ Définition

Objets de l'algèbre relationnelle représentant les tables d'une BD

□ Formalisation : une relation $R = (T, \dots)$

■ *Extension* : ensemble de tuples

... définition à compléter ...

□ Abus de notations

■ On identifiera souvent l'*extension* à la *relation*

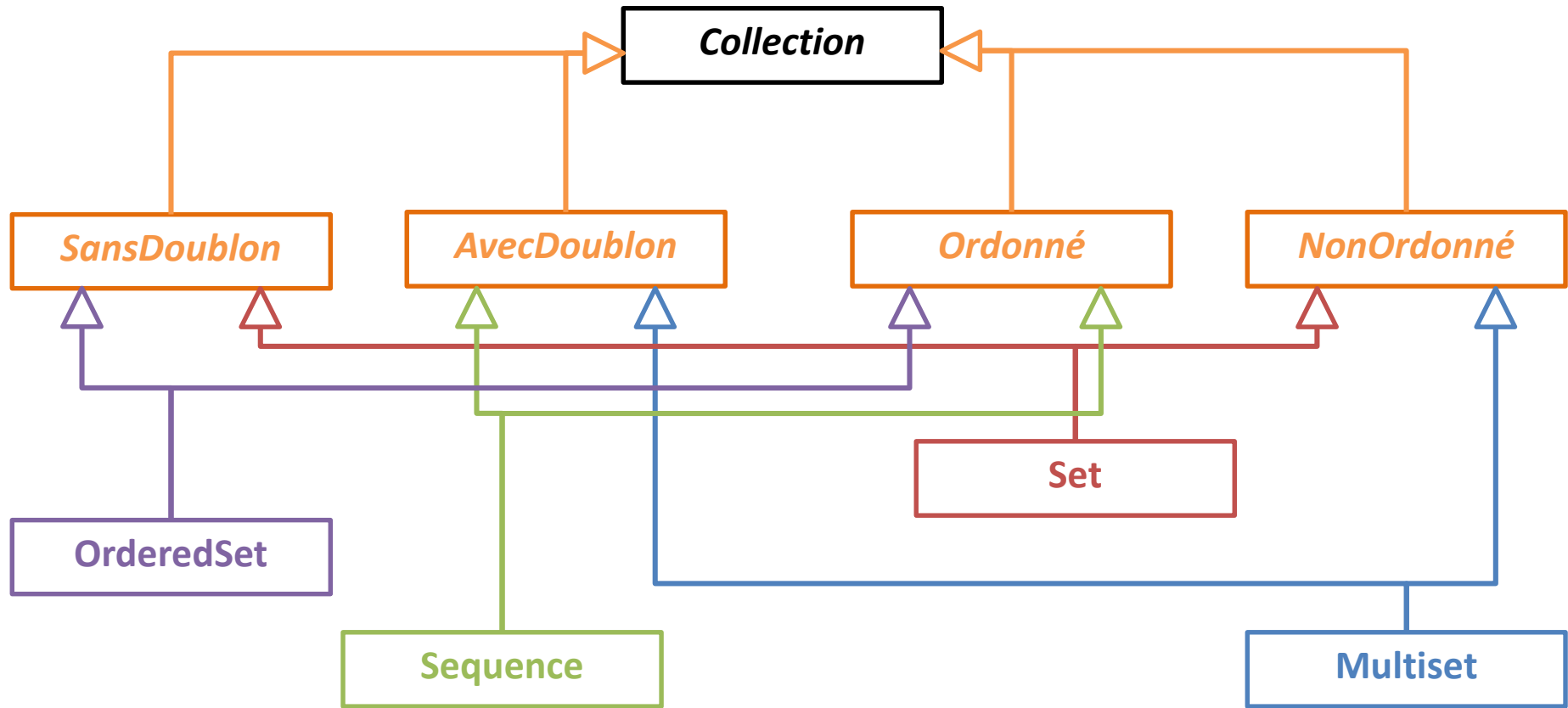
$t \in R$ désigne un tuple t appartenant à l'extension de la relation R

□ Propriétés ensemblistes de l'extension

■ Ensemble = structure non-ordonnée et sans doublon

Algèbre relationnelle : objets

■ Relation : extension



Algèbre relationnelle : objets

■ Relation : extension

□ Définition

Objets de l'algèbre relationnelle représentant les tables d'une BD

□ Formalisation : une relation $R = (T, \dots)$

■ *Extension* : ensemble de tuples

... définition à compléter ...

□ Abus de notations

■ On identifiera souvent l'*extension* à la *relation*

$t \in R$ désigne un tuple t appartenant à l'extension de la relation R

□ Propriétés ensemblistes de l'extension

■ Ensemble = structure non-ordonnée et sans doublon


■ *Il ne peut y avoir deux tuples identiques dans une même relation*

■ *L'ordre de parcours des tuples d'une relation n'est pas spécifié*

Algèbre relationnelle : objets

■ Relation : extension

Etudiant	N°	Nom	Date naiss.	Niveau
e_1 ←	123456	Toto	25/01/93	L3
e_2 ←	542355	Tata	27/04/92	L2
e_3 ←	242643	Titi	14/08/89	M2
e_4 ←	328494	Tutu	08/11/95	L1


Num Nom DNaiss Niv

$$\text{Etudiant} = (\underbrace{\{e_1, e_2, e_3, e_4\}}_{\text{extension}}, \dots)$$

Algèbre relationnelle : objets

■ Relation : schéma

□ Définition

Objets de l'algèbre relationnelle représentant les tables d'une BD

□ Formalisation : une relation $R = (T, Sc, \dots)$

■ *Extension* : ensemble de tuples

■ *Schéma* : ensemble d'attributs

... définition à compléter ...

□ Abus de notations

■ On identifiera souvent le schéma à la relation

$A \in R$ désigne un attribut A appartenant au schéma Sc de la relation R

□ Propriétés

■ Propriété ensembliste du schéma : *l'ordre des attributs n'est pas spécifié*


■ Schémas *compatibles* : si leurs attributs sont compatibles deux à deux

■ *Sous-schéma* : un sous-ensemble des attributs d'une relation

Algèbre relationnelle : objets

■ Relation : schéma

Etudiant	N°	Nom	Date naiss.	Niveau
e_1 ←	123456	Toto	25/01/93	L3
e_2 ←	542355	Tata	27/04/92	L2
e_3 ←	242643	Titi	14/08/89	M2
e_4 ←	328494	Tutu	08/11/95	L1


Num **Nom** **DNaiss** **Niv**

$$\text{Etudiant} = (\underbrace{\{e_1, e_2, e_3, e_4\}}_{\text{extension}}, \underbrace{\{\text{Num}, \text{Nom}, \text{DNaiss}, \text{Niv}\}}_{\text{schéma}}, \dots)$$

Algèbre relationnelle : objets

■ Relation : clé primaire

□ Définition

Objets de l'algèbre relationnelle représentant les tables d'une BD

□ Formalisation : une relation $R = (T, Sc, K, \dots)$

■ *Extension* : ensemble de tuples

■ *Schéma* : ensemble d'attributs

■ *Clé primaire* : clé candidate

... définition à compléter ...

□ Notion de clé

■ Super clé : sous-schéma $K \subset Sc$ distinguant l'ensemble des tuples

$$\forall t, t' \in T \quad \exists A \in K \quad A(t) \neq A(t')$$

Par définition Sc est une super clé

■ Clé candidate : Super clé minimale

$$\forall K' \subsetneq K, K' \text{ n'est pas une super clé}$$

■ Clé primaire = une des clés candidates

Algèbre relationnelle : objets

■ Relation : clé primaire

Représentée graphiquement en soulignant ses attributs

Etudiant	<u>N°</u>	Nom	Date naiss.	Niveau
e ₁	123456	Toto	25/01/93	L3
e ₂	542355	Tata	27/04/92	L2
e ₃	242643	Titi	14/08/89	M2
e ₄	328494	Tutu	08/11/95	L1

↓
Num

↓
Nom

↓
DNaiss

↓
Niv

Etudiant = ({e₁, e₂, e₃, e₄} , {Num, Nom, DNaiss, Niv} , {Num} , ...)

extension schéma clé primaire

Algèbre relationnelle : objets

■ Relation : contraintes référentielles

□ Définition

Objets de l'algèbre relationnelle représentant les tables d'une BD

□ Formalisation : une relation $R = (T, Sc, K, \mathcal{L})$

- *Extension* : ensemble de tuples
- *Schéma* : ensemble d'attributs
- *Clé primaire* : clé candidate
- *Contraintes référentielles* : ensemble de clés étrangères

□ Clé étrangère (ou *secondaire*)

- Lien sémantique avec d'autres relations
- Couple $\langle R', L \rangle$ avec $R' = (T', Sc', K', \mathcal{L}')$ et $L \subset Sc$
 - L est en bijection compatible avec K'
(on utilise une bijection car les noms des attributs peuvent différer)
 - Obligation : $\forall t \in R \quad \exists! t' \in R' \quad L(t) \text{ coïncide avec } K'(t')$
- Un même attribut peut appartenir à plusieurs contraintes référentielles

Algèbre relationnelle : objets

■ Relation : contraintes référentielles

Etudiant($\{e_1, e_2, e_3, e_4\}, \{\text{Num}, \text{Nom}, \text{DNaiss}\}, \{\text{Num}\}, \emptyset$)

Groupe = ($\{g_1, g_2, g_3\}, \{\text{NEtu}, \text{Niv}, \text{NGrp}\}, \{\text{NEtu}, \text{Niv}, \text{NGrp}\}, \{\langle \text{Etudiant}, \{\text{NEtu}\} \rangle\}$)

Etudiant	<u>N°</u>	Nom	Date naiss.
e_1	123456	Toto	25/01/93
e_2	542355	Tata	27/04/92
e_3	242643	Titi	14/08/89
e_4	328494	Tutu	08/11/95
	Num	Nom	DNaiss

Groupe	<u>N°Etu</u>	<u>Niveau</u>	<u>N°Grp</u>
g_1	123456	L2	2
g_2	123456	L3	1
g_3	328494	L2	6
	NEtu	Niv	NGrp

Algèbre relationnelle : objets

■ Tableau récapitulatif et représentation tabulaire

Vocabulaire	Représentation tabulaire	Algèbre relationnelle
Domaine	Valeur (dont l'absence)	D (avec $\perp \in D$)
Relation	Table	$R = (T, Sc, K, \mathcal{L})$
Tuple	Ligne	$t \in T$
Attribut	Colonne	$A: T \rightarrow D$
Schéma	Entête	$Sc = \{A_i \mid i \in \llbracket 1..n \rrbracket\}$
Arité	Largeur de la table	n
Clé primaire	Attribut souligné	$K \subset Sc$
Clé étrangère	-	$\langle R', L \rangle \in \mathcal{L}$ avec $L \subset Sc$

Algèbre relationnelle : objets



■ Formes normales

□ Motivations

- Définition des **relations** : qu'est qu'un modèle *correct* de données
- **Formes normales** : qu'est qu'un *bon* modèle de données
 - Éviter les redondances => décomposition des tables
 - Dépendance fonctionnelle : identifier les fonctions dans les relations

□ Cf. cours de BD avancé...

Algèbre relationnelle

■ Définition

□ Objets : *relation* $R = (T, Sc, K, \mathcal{L})$

Extension : (ensemble de tuples), *schéma* (ensemble d'attributs), *clé primaire* (clé candidate), *contraintes référentielles* (ensemble de clés étrangères)

□ Opérations

■ Nullaires : les relations de la BD

■ Unaires : ???

■ Binaires : ???

□ Axiomes : ???

□ Calculs

Les équations sont les requêtes

Algèbre relationnelle : opérations

■ Constantes

- Relations construites à partir des tables de BD

Toute autre relation est obtenue en appliquant des opérations sur les constantes

- Exemple

4 tables enregistrant des viticulteurs, des vins, des commandes et des clients

VITICULTEUR(NUM, NOM, PRENOM, REGION)

VIN(NVIN, CRU, MILLESIME, NVITICULTEUR, REGION)

COMMANDE(NCOMMANDE, NCLIENT, NVIN, DATE, QUANTITE)

CLIENT(NUM, NOM, PRENOM, REGION)

Algèbre relationnelle : opérations nulaires

Vin	<u>Nvin</u>	Cru	Millésime	Viticulteur	Région
-----	-------------	-----	-----------	-------------	--------

1	Saint-Emilion	2002	1	Bordeaux
2	Champagne	1996	3	Champagne
3	Pauillac	1992	1	Bordeaux
4	Chablis	2007	6	Bourgogne

Client	<u>NUM</u>	Nom	Prénom	Région
--------	------------	-----	--------	--------

Viticulteur	<u>NUM</u>	Nom	Prénom	Région
-------------	------------	-----	--------	--------

1	Barré	Henri	Bordeaux
3	Hermelin	Jean-Pierre	Champagne
4	Fort	Valérie	Languedoc
6	Schmidt	Thomas	Bourgogne

2	Voser	Armande	Alsace
3	Hermelin	Jean-Pierre	Champagne
5	Senard	Danièle	Champagne
6	Schmidt	Thomas	Bourgogne

Commande	<u>NCom</u>	NClient	<u>NVin</u>	Date	Quantité
----------	-------------	---------	-------------	------	----------

1	5	4	27/04/12	25
2	3	2	25/01/13	100
3	2	2	14/08/09	80
4	5	1	08/11/10	100

Algèbre relationnelle : opérations

■ Opérations unaires

- Opérations s'appliquant sur une seule relation

Les opérations peuvent néanmoins être paramétrées

- Opérations unaires de l'algèbre relationnelle

- *Renommage* ρ (« rho ») : changement de nom d'un attribut
- *Sélection* σ (« sigma ») : diminution du nombre de tuples
- *Projection* π (« pi ») : diminution du nombre d'attributs

Algèbre relationnelle : opérations unaires

■ Renommage ρ

□ Spécification informelle

Modifie le nom d'un ou plusieurs attributs (modification de l'entête)

□ Notation

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

□ Sémantique ensembliste

- Soit une relation $R = (T, Sc, K, \mathcal{L})$
- Soit $A = \{A_1, \dots, A_k\}$ un sous-schéma de Sc
- Soit $B = \{B_1, \dots, B_k\}$ un schéma tel que $D^{A_i} = D^{B_i}$ (domaines compatibles)

$$\rho_{A_1 \rightarrow B_1, \dots, A_k \rightarrow B_k}(R) = (T, (Sc \setminus A) \cup B, (K \setminus A) \cup B_K, \mathcal{L}')$$

$$\text{avec } \mathcal{L}' = \{ \langle R', (L \setminus A) \cup B_L \rangle \mid \langle R', L \rangle \in \mathcal{L} \}$$
$$\text{et } B_K = \{ B_i \mid A_i \in K \cap A \}$$

Algèbre relationnelle : opérations unaires

■ Renommage ρ

- Exemple : traduire en anglais les attributs de la relation viticulteur

Viticulteur	<u>NUM</u>	Nom	Prénom	Région
	1	Barré	Henri	Bordeaux
	3	Hermelin	Jean-Pierre	Champagne
	4	Fort	Valérie	Languedoc
	6	Schmidt	Thomas	Bourgogne

R	<u>ID</u>	Lastname	Firstname	Region
	1	Barré	Henri	Bordeaux
	3	Hermelin	Jean-Pierre	Champagne
	4	Fort	Valérie	Languedoc
	6	Schmidt	Thomas	Bourgogne

$$R = \rho_{\text{NUM} \rightarrow \text{ID}, \text{Nom} \rightarrow \text{Lastname}, \text{Prénom} \rightarrow \text{Firstname}, \text{Région} \rightarrow \text{Region}}(\text{Viticulteur})$$

Algèbre relationnelle : opérations unaires

■ Sélection σ

- Spécification informelle

Conserve les tuples vérifiant une propriété (suppression de lignes)

- Notation

$$\sigma_{\varphi}(R)$$

- Sémantique ensembliste

- Soit une relation $R = (T, Sc, K, \mathcal{L})$ avec $Sc = \{A_1, \dots, A_n\}$
- Soit une formule logique φ utilisant comme variables les attributs de Sc

$$\sigma_{\varphi}(R) = (T', Sc, K, \mathcal{L}) \text{ avec } T' = \{t \in T \mid \varphi[A_1 \leftarrow A_1(t), \dots, A_n \leftarrow A_n(t)]\}$$

- Remarques

- $\varphi[X \leftarrow v]$: les occurrences libres de X dans φ sont remplacées par la valeur v
Dans notre cas, les attributs A_i sont remplacés par leurs valeurs $A_i(t)$
- La sélection peut calculer une relation vide (sans ligne)

Algèbre relationnelle : opérations unaires

■ Sélection σ

□ Exemple : trouver les vins de la région « Bordeaux »

Vin	<u>Nvin</u>	Cru	Millésime	Viticulteur	Région	
1		Saint-Emilion	2002	1	Bordeaux	✓
2		Champagne	1996	3	Champagne	✗
3		Pauillac	1992	1	Bordeaux	✓
4		Chablis	2007	6	Bourgogne	✗

R	<u>NVin</u>	Cru	Millésime	Viticulteur	Région
	1	Saint-Emilion	2002	1	Bordeaux
	3	Pauillac	1992	1	Bordeaux

$$\mathbf{R} = \sigma_{\text{Région} = \text{« Bordeaux »}}(\mathbf{Vin})$$

Algèbre relationnelle : opérations unaires

■ Sélection σ

□ Exemple : trouver les vins de la région « Alsace »

Vin	<u>Nvin</u>	Cru	Millésime	Viticulteur	Région	
1		Saint Emilion	2002	1	Bordeaux	X
2		Champagne	1996	3	Champagne	X
3		Pauillac	1992	1	Bordeaux	X
4		Chablis	2007	6	Bourgogne	X

R	<u>NVin</u>	Cru	Millésime	Viticulteur	Région
----------	-------------	-----	-----------	-------------	--------

$$\mathbf{R} = \sigma_{\text{Région} = \text{« Alsace »}}(\mathbf{Vin})$$

Algèbre relationnelle : opérations unaires

■ *Projection* π

- Spécification informelle

Conserve un sous-ensemble d'attributs (suppression de colonnes)

- Notation

$$\pi_{A_1, \dots, A_k}(R)$$

- Sémantique ensembliste

- Soit une relation $R = (T, Sc, K, \mathcal{L})$
- Soit $A = \{A_1, \dots, A_k\}$ un sous-schéma de Sc

$$\pi_{A_1, \dots, A_k}(R) = (T, A, A, \mathcal{L}') \quad \text{avec} \quad \mathcal{L}' = \{\langle R', L \rangle \in \mathcal{L} \mid A \subseteq L\}$$

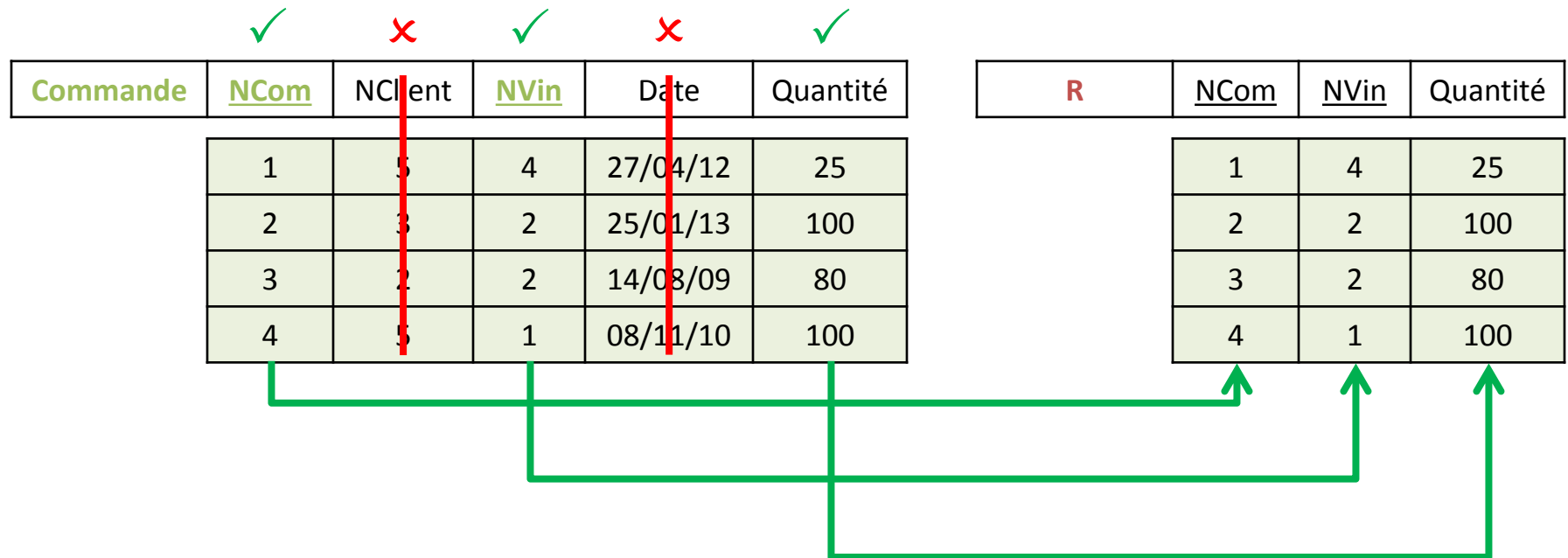
- Remarques

- Bien que les tuples ne soient pas touchés, les doublons sont supprimés
- Si $K \subseteq A$ (l'ancienne clé est toujours présente), on pourra toujours l'utiliser comme nouvelle clé primaire

Algèbre relationnelle : opérations unaires

■ Projection π

- Exemple : trouver les numéros de vins commandés ainsi que leur quantité



$$R = \pi_{NCom, NVin, Quantité} (Commande)$$

Algèbre relationnelle : opérations unaires

■ Projection π

- Exemple : trouver les viticulteurs actifs (i.e., qui produisent)

Vin	N^o vin	Cru	Millesime	Viticulteur	Région
1	1	Saint-Emilion	2002	1	Bordeaux
2	2	Champagne	1996	3	Champagne
3	3	Pauillac	1992	1	Bordeaux
4	4	Chablis	2007	6	Bourgogne

Viticulteur
1
3
1
6

$$R = \pi_{\text{Viticulteur}}(\text{Vin})$$

Algèbre relationnelle : opérations unaires

■ Projection π

- Exemple : trouver les viticulteurs actifs (i.e., qui produisent)

Vin	N^o vin	Cru	Millesime	Viticulteur	Région
1	Saint-Emilion	2002	1	Bordeaux	
2	Champagne	1996	3	Champagne	
3	Pauillac	1992	1	Bordeaux	
4	Chablis	2007	6	Bourgogne	

R	<u>Viticulteur</u>
1	
3	
6	

$$R = \pi_{\text{Viticulteur}}(\text{Vin})$$

Algèbre relationnelle : opérations

■ Opérations binaires

- Opérations s'appliquant sur deux relations
- Notation infixe
 - L'opérateur est situé entre les deux opérandes
- Opérations binaires de l'algèbre relationnelle
 - *Op. ensemblistes* \cup , \cap et \setminus : relations vues comme des ensembles de tuples
 - Conservation du schéma des relations
 - *Op. de croisement* \times , \bowtie et \div : extension et réduction de schémas

Algèbre relationnelle : opérations binaires

■ *Opérateurs ensemblistes (booléennes)* \cup , \cap et \setminus

□ Spécification informelle

Calculent l'union, l'intersection et la différence entre les tuples de deux relations

□ Notation

$$R_1 \cup R_2 \quad R_1 \cap R_2 \quad R_1 \setminus R_2$$

□ Sémantique ensembliste

Soient $R_1 = (T_1, Sc, K, \mathcal{L}_1)$ et $R_2 = (T_2, Sc, K, \mathcal{L}_2)$

$$\blacksquare R_1 \cup R_2 = (T_1 \cup T_2, Sc, K, \mathcal{L}_1 \cap \mathcal{L}_2) = (\{t \mid t \in T_1 \vee t \in T_2\}, Sc, K, \mathcal{L}_1 \cap \mathcal{L}_2)$$

$$\blacksquare R_1 \cap R_2 = (T_1 \cap T_2, Sc, K, \mathcal{L}_1 \cap \mathcal{L}_2) = (\{t \mid t \in T_1 \wedge t \in T_2\}, Sc, K, \mathcal{L}_1 \cap \mathcal{L}_2)$$

$$\blacksquare R_1 \setminus R_2 = (T_1 \setminus T_2, Sc, K, \mathcal{L}_1 \cap \mathcal{L}_2) = (\{t \mid t \in T_1 \wedge t \notin T_2\}, Sc, K, \mathcal{L}_1 \cap \mathcal{L}_2)$$

□ Remarques

- Les relations doivent impérativement avoir même schéma et même clé prim.
- Toutes les contraintes référentielles ne peuvent être maintenues

Algèbre relationnelle : opérations binaires

■ Opérateurs ensemblistes (booléennes) \cup , \cap et \setminus

□ Exemple : trouver toutes les personnes référencées dans la BD

Client	<u>NUM</u>	Nom	Prénom	Région
--------	------------	-----	--------	--------

Viticulteur	<u>NUM</u>	Nom	Prénom	Région
	1	Barré	Henri	Bordeaux
	3	Hermelin	Jean-Pierre	Champagne
	4	Fort	Valérie	Languedoc
	6	Schmidt	Thomas	Bourgogne

2	Voser	Armande	Alsace
3	Hermelin	Jean-Pierre	Champagne
5	Senard	Danièle	Champagne
6	Schmidt	Thomas	Bourgogne

$$R = \text{Viticulteur} \cup \text{Client}$$

R	<u>NUM</u>	Nom	Prénom	Région
---	------------	-----	--------	--------

1	Barré	Henri	Bordeaux
2	Voser	Armande	Alsace
3	Hermelin	Jean-Pierre	Champagne
4	Fort	Valérie	Languedoc
5	Senard	Danièle	Champagne
6	Schmidt	Thomas	Bourgogne

Algèbre relationnelle : opérations binaires

■ Opérateurs ensemblistes (booléennes) \cup , \cap et \setminus

□ Exemple : trouver les viticulteurs également clients

Client	<u>NUM</u>	Nom	Prénom	Région
--------	------------	-----	--------	--------

2	Voser	Armande	Alsace
3	Hermelin	Jean-Pierre	Champagne
5	Senard	Danièle	Champagne
6	Schmidt	Thomas	Bourgogne

Viticulteur	<u>NUM</u>	Nom	Prénom	Région
	1	Barré	Henri	Bordeaux
	3	Hermelin	Jean-Pierre	Champagne
	4	Fort	Valérie	Languedoc
	6	Schmidt	Thomas	Bourgogne

$$R = \text{Viticulteur} \cap \text{Client}$$

R	<u>NUM</u>	Nom	Prénom	Région
	3	Hermelin	Jean-Pierre	Champagne
	6	Schmidt	Thomas	Bourgogne

Algèbre relationnelle : opérations binaires

■ Opérateurs ensemblistes (booléennes) \cup , \cap et \setminus

□ Exemple : trouver les viticulteurs qui ne sont pas clients

Client	<u>NUM</u>	Nom	Prénom	Région
--------	------------	-----	--------	--------

2	Voser	Armande	Alsace
3	Hermelin	Jean-Pierre	Champagne
5	Senard	Danièle	Champagne
6	Schmidt	Thomas	Bourgogne

Viticulteur	<u>NUM</u>	Nom	Prénom	Région
	1	Barré	Henri	Bordeaux
	3	Hermelin	Jean-Pierre	Champagne
	4	Fort	Valérie	Languedoc
	6	Schmidt	Thomas	Bourgogne

$$R = \text{Viticulteur} \setminus \text{Client}$$

R	<u>NUM</u>	Nom	Prénom	Région
	1	Barré	Henri	Bordeaux
	4	Fort	Valérie	Languedoc

Algèbre relationnelle : opérations binaires

■ Opérateurs ensemblistes (booléennes) \cup , \cap et \setminus

□ Exemple : trouver les clients qui ne sont pas des viticulteurs

Client	<u>NUM</u>	Nom	Prénom	Région
--------	------------	-----	--------	--------

Viticulteur	<u>NUM</u>	Nom	Prénom	Région
	1	Barré	Henri	Bordeaux
	3	Hermelin	Jean-Pierre	Champagne
	4	Fort	Valérie	Languedoc
	6	Schmidt	Thomas	Bourgogne

2	Voser	Armande	Alsace
3	Hermelin	Jean-Pierre	Champagne
5	Senard	Danièle	Champagne
6	Schmidt	Thomas	Bourgogne

$$R = \text{Client} \setminus \text{Viticulteur}$$

R	<u>NUM</u>	Nom	Prénom	Région
	2	Voser	Armande	Alsace
	5	Senard	Danièle	Champagne

Algèbre relationnelle : opérations binaires

■ *Op. de croisement : produit cartésien* \times

- Spécification informelle

Croise tous les tuples d'une relation avec tous ceux d'une autre relation

- Notation

$$R_1 \times R_2$$

- Sémantique ensembliste

Soient $R_1 = (T_1, Sc_1, K_1, L_1)$ et $R_2 = (T_2, Sc_2, K_2, L_2)$

$$R_1 \times R_2 = (T_1 \times T_2, Sc_1 \uplus Sc_2, K_1 \uplus K_2, L_1 \uplus L_2)$$

- Remarques

- La plus grand opérateur de croisement
- Le produit cartésien de relations est commutatif !!
- L'opération \uplus représente l'union disjointe de deux ensembles (cf. exemple)
 - $E_1 \uplus E_2 = (\{0\} \times E_1) \cup (\{1\} \times E_2)$
 - En algèbre relationnelle, $\{0\}$ et $\{1\}$ remplacés par les noms des relations

Algèbre relationnelle : opérations binaires

■ *Op. de croisement : produit cartésien* \times

- Exemple : croiser l'ensemble des clients avec toutes les commandes

Client	<u>NUM</u>	Nom	Prénom	Région
--------	------------	-----	--------	--------

2	Voser	Armande	Alsace
3	Hermelin	Jean-Pierre	Champagne
5	Senard	Danièle	Champagne
6	Schmidt	Thomas	Bourgogne

Commande	<u>NCom</u>	NClient	<u>NVin</u>	Date	Quantité
----------	-------------	---------	-------------	------	----------

1	5	4	27/04/12	25
2	3	2	25/01/13	100
3	2	2	14/08/09	80
4	5	1	08/11/10	100

$$R = \text{Client} \times \text{Commande}$$

Algèbre relationnelle : opérations binaires

R	<u>NCom</u>	NClient	<u>NVin</u>	Date	Quantité	<u>NUM</u>	Nom	Prénom	Région
	1	5	4	27/04/12	25	2	Voser	Armande	Alsace
	1	5	4	27/04/12	25	3	Hermelin	Jean-Pierre	Champagne
	1	5	4	27/04/12	25	5	Senard	Danièle	Champagne
	1	5	4	27/04/12	25	6	Schmidt	Thomas	Bourgogne
	2	3	2	25/01/13	100	2	Voser	Armande	Alsace
	2	3	2	25/01/13	100	3	Hermelin	Jean-Pierre	Champagne
	2	3	2	25/01/13	100	5	Senard	Danièle	Champagne
	2	3	2	25/01/13	100	6	Schmidt	Thomas	Bourgogne
	3	2	2	14/08/09	80	2	Voser	Armande	Alsace
	3	2	2	14/08/09	80	3	Hermelin	Jean-Pierre	Champagne
	3	2	2	14/08/09	80	5	Senard	Danièle	Champagne
	3	2	2	14/08/09	80	6	Schmidt	Thomas	Bourgogne
	4	5	1	08/11/10	100	2	Voser	Armande	Alsace
	4	5	1	08/11/10	100	3	Hermelin	Jean-Pierre	Champagne
	4	5	1	08/11/10	100	5	Senard	Danièle	Champagne
	4	5	1	08/11/10	100	6	Schmidt	Thomas	Bourgogne

Algèbre relationnelle : opérations binaires

■ *Op. de croisement : produit cartésien* \times

□ Exemple : croiser l'ensemble des clients avec toutes les commandes

Viticulteur	<u>NUM</u>	Nom	Prénom	Région
-------------	------------	-----	--------	--------

1	Barré	Henri	Bordeaux
3	Hermelin	Jean-Pierre	Champagne
4	Fort	Valérie	Languedoc
6	Schmidt	Thomas	Bourgogne

Vin	<u>Nvin</u>	Cru	Millésime	Viticulteur	Région
-----	-------------	-----	-----------	-------------	--------

1	Saint-Emilion	2002	1	Bordeaux
2	Champagne	1996	3	Champagne
3	Pauillac	1992	1	Bordeaux
4	Chablis	2007	6	Bourgogne

$$R = \text{Viticulteur} \times \text{Vin}$$

Algèbre relationnelle : opérations binaires

R	NUM	Nom	Prénom	Viticulteur. Région	Nvin	Cru	Millé- sime	Viticulteur	Vin.Région
	1	Barré	Henri	Bordeaux	1	Saint-Emilion	2002	1	Bordeaux
	1	Barré	Henri	Bordeaux	2	Champagne	1996	3	Champagne
	1	Barré	Henri	Bordeaux	3	Pauillac	1992	1	Bordeaux
	1	Barré	Henri	Bordeaux	4	Chablis	2007	6	Bourgogne
	3	Hermelin	Jean-Pierre	Champagne	1	Saint-Emilion	2002	1	Bordeaux
	3	Hermelin	Jean-Pierre	Champagne	2	Champagne	1996	3	Champagne
	3	Hermelin	Jean-Pierre	Champagne	3	Pauillac	1992	1	Bordeaux
	3	Hermelin	Jean-Pierre	Champagne	4	Chablis	2007	6	Bourgogne
	4	Fort	Valérie	Languedoc	1	Saint-Emilion	2002	1	Bordeaux
	4	Fort	Valérie	Languedoc	2	Champagne	1996	3	Champagne
	4	Fort	Valérie	Languedoc	3	Pauillac	1992	1	Bordeaux
	4	Fort	Valérie	Languedoc	4	Chablis	2007	6	Bourgogne
	6	Schmidt	Thomas	Bourgogne	1	Saint-Emilion	2002	1	Bordeaux
	6	Schmidt	Thomas	Bourgogne	2	Champagne	1996	3	Champagne
	6	Schmidt	Thomas	Bourgogne	3	Pauillac	1992	1	Bordeaux
	6	Schmidt	Thomas	Bourgogne	4	Chablis	2007	6	Bourgogne

Algèbre relationnelle : opérations binaires

■ Op. de croisement : jointures $\bowtie, \bowtie_{\theta}, \overset{\circ}{\bowtie}, \dots$

□ Spécification informelle

■ Spécialisation du produit cartésien

■ Différents types de jointures

- Jointure interne ou externe (à gauche, à droite, bilatérale)
- Jointure naturelle ou θ -jointure (+ équi-jointure)

LES TYPES DE JOINTURES		Naturelle	θ -jointure
Interne		\bowtie	\bowtie_{θ}
Externe	à gauche	$\overset{\circ}{\bowtie}_L$	$\overset{\circ}{\bowtie}_{\theta L}$
	à droite	$\overset{\circ}{\bowtie}_R$	$\overset{\circ}{\bowtie}_{\theta R}$
	bilatérale	$\overset{\circ}{\bowtie}$	$\overset{\circ}{\bowtie}_{\theta}$

Algèbre relationnelle : opérations binaires

■ Op. de croisement : θ -jointure et équi-jointure \bowtie_{θ}

□ Spécification informelle

Sélectionne les tuples issus d'un produit cartésien vérifiant une propriété

□ Notation

$$R_1 \bowtie_{\theta} R_2$$

□ Sémantique ensembliste

■ Soient $R_1 = (T_1, SC_1, K_1, L_1)$ et $R_2 = (T_2, SC_2, K_2, L_2)$

■ Soit une formule θ utilisant comme variables les attributs de $SC_1 \cup SC_2$

$$R_1 \bowtie_{\theta} R_2 = \sigma_{\theta}(R_1 \times R_2)$$

□ Remarque

- On parle d'**équi-jointure** lorsque θ n'utilise que les opérateurs \wedge et $=$
Conjonctions d'égalités « attribut-attribut » ou « attribut-valeur »
- Lorsque le terme est omis, une jointure est toujours interne

Algèbre relationnelle : opérations binaires

■ Op. de croisement : θ -jointure \bowtie_{θ}

- Exemple : rechercher les vins commandés à plus de 30 exemplaires

Commande	<u>NCom</u>	NClient	<u>NVin</u>	Date	Quantité
	1	5	4	27/04/12	25
	2	3	2	25/01/13	100
	3	2	2	14/08/09	80
	4	5	1	08/11/10	100

Vin	<u>Nvin</u>	Cru	Millésime	Viticulteur	Région
	1	Saint-Emilion	2002	1	Bordeaux
	2	Champagne	1996	3	Champagne
	3	Pauillac	1992	1	Bordeaux
	4	Chablis	2007	6	Bourgogne

$$R = \text{Commande} \bowtie_{\text{Commande.NVin} = \text{Vin.NVin} \wedge \text{Quantité} > 30} \text{Vin}$$

Algèbre relationnelle : opérations binaires

R	<u>Vin.</u> <u>Nvin</u>	Cru	Millésime	Viticu- lteur	Région	<u>NCom</u>	NClient	<u>Commande.</u> <u>NVin</u>	Date	Quantité
X	1	Saint-Emilion	2002	1	Bordeaux	1	5	4	27/04/12	25
X	1	Saint-Emilion	2002	1	Bordeaux	2	3	2	25/01/13	100
X	1	Saint-Emilion	2002	1	Bordeaux	3	2	2	14/08/09	80
✓	1	Saint-Emilion	2002	1	Bordeaux	4	5	1	08/11/10	100
X	2	Champagne	1996	3	Champagne	1	5	4	27/04/12	25
✓	2	Champagne	1996	3	Champagne	2	3	2	25/01/13	100
✓	2	Champagne	1996	3	Champagne	3	2	2	14/08/09	80
X	2	Champagne	1996	3	Champagne	4	5	1	08/11/10	100
X	3	Pauillac	1992	1	Bordeaux	1	5	4	27/04/12	25
X	3	Pauillac	1992	1	Bordeaux	2	3	2	25/01/13	100
X	3	Pauillac	1992	1	Bordeaux	3	2	2	14/08/09	80
X	3	Pauillac	1992	1	Bordeaux	4	5	1	08/11/10	100
X	4	Chablis	2007	6	Bourgogne	1	5	4	27/04/12	25
X	4	Chablis	2007	6	Bourgogne	2	3	2	25/01/13	100
X	4	Chablis	2007	6	Bourgogne	3	2	2	14/08/09	80
X	4	Chablis	2007	6	Bourgogne	4	5	1	08/11/10	100

Algèbre relationnelle : opérations binaires

■ Op. de croisement : θ -jointure \bowtie_{θ}

□ Exemple : rechercher les vins commandés à plus de 30 exemplaires

R	<u>Vin.</u> <u>Nvin</u>	Cru	Millésime	Viticu- lteur	Région	<u>NCom</u>	NClient	<u>Commande.</u> <u>NVin</u>	Date	Quantité
	1	Saint-Emilion	2002	1	Bordeaux	4	5	1	08/11/10	100
	2	Champagne	1996	3	Champagne	2	3	2	25/01/13	100
	2	Champagne	1996	3	Champagne	3	2	2	14/08/09	80

$R = \text{Vin} \bowtie_{\text{Commande.NVin} = \text{Vin.NVin} \wedge \text{Quantité} > 30} \text{Commande}$

Algèbre relationnelle : opérations binaires

■ Op. de croisement : équi-jointure \bowtie_{θ}

□ Exemple : les vins et les viticulteurs de la même région

Viticulteur	<u>NUM</u>	Nom	Prénom	Région
	1	Barré	Henri	Bordeaux
	3	Hermelin	Jean-Pierre	Champagne
	4	Fort	Valérie	Languedoc
	6	Schmidt	Thomas	Bourgogne

Vin	<u>Nvin</u>	Cru	Millésime	Viticulteur	Région
	1	Saint-Emilion	2002	1	Bordeaux
	2	Champagne	1996	3	Champagne
	3	Pauillac	1992	1	Bordeaux
	4	Chablis	2007	6	Bourgogne

$R = \text{Viticulteur} \bowtie_{\text{Viticulteur.Région} = \text{Vin.Région}} \text{Vin}$

Algèbre relationnelle : opérations binaires

R	NUM	Nom	Prénom	Viticulteur. Région	Nvin	Cru	Millé- sime	Viticulteur	Vin.Région
✓	1	Barré	Henri	Bordeaux	1	Saint-Emilion	2002	1	Bordeaux
✗	1	Barré	Henri	Bordeaux	2	Champagne	1996	3	Champagne
✓	1	Barré	Henri	Bordeaux	3	Pauillac	1992	1	Bordeaux
✗	1	Barré	Henri	Bordeaux	4	Chablis	2007	6	Bourgogne
✗	3	Hermelin	Jean-Pierre	Champagne	1	Saint-Emilion	2002	1	Bordeaux
✓	3	Hermelin	Jean-Pierre	Champagne	2	Champagne	1996	3	Champagne
✗	3	Hermelin	Jean-Pierre	Champagne	3	Pauillac	1992	1	Bordeaux
✗	3	Hermelin	Jean-Pierre	Champagne	4	Chablis	2007	6	Bourgogne
✗	4	Fort	Valérie	Languedoc	1	Saint-Emilion	2002	1	Bordeaux
✗	4	Fort	Valérie	Languedoc	2	Champagne	1996	3	Champagne
✗	4	Fort	Valérie	Languedoc	3	Pauillac	1992	1	Bordeaux
✗	4	Fort	Valérie	Languedoc	4	Chablis	2007	6	Bourgogne
✗	6	Schmidt	Thomas	Bourgogne	1	Saint-Emilion	2002	1	Bordeaux
✗	6	Schmidt	Thomas	Bourgogne	2	Champagne	1996	3	Champagne
✗	6	Schmidt	Thomas	Bourgogne	3	Pauillac	1992	1	Bordeaux
✓	6	Schmidt	Thomas	Bourgogne	4	Chablis	2007	6	Bourgogne

Algèbre relationnelle : opérations binaires

■ *Op. de croisement : équi-jointure* \bowtie_{θ}

□ Exemple : les vins et les viticulteurs de la même région

R	<u>NUM</u>	Nom	Prénom	Viticulteur. Région	<u>Nvin</u>	Cru	Millé- sime	Viticulteur	Vin.Région
	1	Barré	Henri	Bordeaux	1	Saint-Emilion	2002	1	Bordeaux
	1	Barré	Henri	Bordeaux	3	Pauillac	1992	1	Bordeaux
	3	Hermelin	Jean-Pierre	Champagne	2	Champagne	1996	3	Champagne
	6	Schmidt	Thomas	Bourgogne	4	Chablis	2007	6	Bourgogne

R = **Viticulteur** \bowtie **Vin**
Viticulteur.Région = **Vin.Région**

Algèbre relationnelle : opérations binaires

■ Op. de croisement : jointure naturelle \bowtie

□ Spécification informelle

Identifie dans un produit cartésien les attributs compatibles

□ Notation

$$R_1 \bowtie R_2$$

□ Sémantique ensembliste

Soient $R_1 = (T_1, SC_1, K_1, L_1)$ et $R_2 = (T_2, SC_2, K_2, L_2)$

$$R_1 \bowtie R_2 = \pi_{SC_1 \cup SC_2} \left(R_1 \underset{\theta}{\bowtie} R_2 \right) \quad \text{avec } \theta = \bigwedge_{A \in SC_1 \cap SC_2} R_1.A = R_2.A$$

□ Remarque

- La jointure naturelle correspond à une équi-jointure identifiant les attributs partagés par les deux relations (dans l'intersection des schémas)
- La projection finale permet d'éviter les doubles colonnes identiques
- Perte des tuples défaillants

Tuples de R_1 (resp. R_2) ne coïncidant avec aucun tuple de R_2 (resp. R_1)

Algèbre relationnelle : opérations binaires

■ *Op. de croisement : jointure naturelle* ⋈

□ Exemple : les vins et les clients de la même région

Client	<u>NUM</u>	Nom	Prénom	Région
	2	Voser	Armande	Alsace
	3	Hermelin	Jean-Pierre	Champagne
	5	Senard	Danièle	Champagne
	6	Schmidt	Thomas	Bourgogne

Vin	<u>Nvin</u>	Cru	Millésime	Viticulteur	Région
1		Saint-Emilion	2002	1	Bordeaux
2		Champagne	1996	3	Champagne
3		Pauillac	1992	1	Bordeaux
4		Chablis	2007	6	Bourgogne

$$R = \text{Client} \bowtie \text{Vin}$$

Algèbre relationnelle : opérations binaires

■ Op. de croisement : jointure naturelle ⋈

□ Exemple : les vins et les clients de la même région

R'	<u>NUM</u>	Nom	Prénom	Client. Région	<u>Nvin</u>	Cru	Millé- sime	Viticulteur	Vin.Région
	3	Hermelin	Jean-Pierre	Champagne	2	Champagne	1996	3	Champagne
	5	Senard	Danièle	Champagne	2	Champagne	1996	3	Champagne
	6	Schmidt	Thomas	Bourgogne	4	Chablis	2007	6	Bourgogne

$$R' = \text{Client} \quad \text{Vin}$$

$$\text{Client.Région} = \text{Vin.Région}$$

R	<u>NUM</u>	Nom	Prénom	Région	<u>Nvin</u>	Cru	Millésime	Viticulteur
	3	Hermelin	Jean-Pierre	Champagne	2	Champagne	1996	3
	5	Senard	Danièle	Champagne	2	Champagne	1996	3
	6	Schmidt	Thomas	Bourgogne	4	Chablis	2007	6

$$R = \text{Client} \quad \text{Vin} = \pi_{\text{NUM, Nom, Prénom, Région, NVin, Cru, Millésime, Viticulteur}} (R')$$

Algèbre relationnelle : opérations binaires

■ Op. de croisement : jointures externes $\overset{\circ}{\bowtie}_L$, $\overset{\circ}{\bowtie}_R$ et $\overset{\circ}{\bowtie}$

- Spécification informelle

Conserve les tuples défailants lors d'une jointure

- Notation

$$R_1 \overset{\circ}{\bowtie}_L R_2 \quad R_1 \overset{\circ}{\bowtie}_R R_2 \quad R_1 \overset{\circ}{\bowtie} R_2$$

- Sémantique ensembliste

Soient $R_1 = (T_1, S_{C_1}, K_1, L_1)$ et $R_2 = (T_2, S_{C_2}, K_2, L_2)$

$$\blacksquare R_1 \overset{\circ}{\bowtie}_L R_2 = R_1 \bowtie R_2 \cup ((R_1 \setminus \pi_{S_{C_1}}(R_1 \bowtie R_2)) \times (\{\perp\}, S_{C_2}, K_2, L_2))$$

$$\blacksquare R_1 \overset{\circ}{\bowtie}_R R_2 = R_2 \overset{\circ}{\bowtie}_L R_1$$

$$\blacksquare R_1 \overset{\circ}{\bowtie} R_2 = R_1 \overset{\circ}{\bowtie}_L R_2 \cup R_1 \overset{\circ}{\bowtie}_R R_2$$

Algèbre relationnelle : opérations binaires

■ *Op. de croisement : jointures externes* $\overset{\circ}{\bowtie}_L$

□ Exemple : les vins et les clients de la même région

R	<u>NUM</u>	Nom	Prénom	Région	<u>Nvin</u>	Cru	Millésime	Viticulteur
	3	Hermelin	Jean-Pierre	Champagne	2	Champagne	1996	3
	5	Senard	Danièle	Champagne	2	Champagne	1996	3
	6	Schmidt	Thomas	Bourgogne	4	Chablis	2007	6

$$\mathbf{R} = \mathbf{Client} \overset{\circ}{\bowtie} \mathbf{Vin}$$

Algèbre relationnelle : opérations binaires

Tuple détaillant

R	NUM	Nom	Prénom	Client. Région	Nvin	Cru	Millé- sime	Viticulteur	Vin.Région
	2	Voser	Armande	Alsace	1	Saint-Emilion	2002	1	Bordeaux
	2	Voser	Armande	Alsace	2	Champagne	1996	3	Champagne
	2	Voser	Armande	Alsace	3	Pauillac	1992	1	Bordeaux
	2	Voser	Armande	Alsace	4	Chablis	2007	6	Bourgogne
	3	Hermelin	Jean-Pierre	Champagne	1	Saint-Emilion	2002	1	Bordeaux
	3	Hermelin	Jean-Pierre	Champagne	2	Champagne	1996	3	Champagne
	3	Hermelin	Jean-Pierre	Champagne	3	Pauillac	1992	1	Bordeaux
	3	Hermelin	Jean-Pierre	Champagne	4	Chablis	2007	6	Bourgogne
	5	Senard	Danièle	Champagne	1	Saint-Emilion	2002	1	Bordeaux
	5	Senard	Danièle	Champagne	2	Champagne	1996	3	Champagne
	5	Senard	Danièle	Champagne	3	Pauillac	1992	1	Bordeaux
	5	Senard	Danièle	Champagne	4	Chablis	2007	6	Bourgogne
	6	Schmidt	Thomas	Bourgogne	1	Saint-Emilion	2002	1	Bordeaux
	6	Schmidt	Thomas	Bourgogne	2	Champagne	1996	3	Champagne
	6	Schmidt	Thomas	Bourgogne	3	Pauillac	1992	1	Bordeaux
	6	Schmidt	Thomas	Bourgogne	4	Chablis	2007	6	Bourgogne

Algèbre relationnelle : opérations binaires

■ *Op. de croisement : jointures externes* \bowtie_L

□ Exemple : les vins et les clients de la même région

R	NUM	Nom	Prénom	Région	Nvin	Cru	Millésime	Viticulteur
	3	Hermelin	Jean-Pierre	Champagne	2	Champagne	1996	3
	5	Senard	Danièle	Champagne	2	Champagne	1996	3
	6	Schmidt	Thomas	Bourgogne	4	Chablis	2007	6

$$R = \text{Client} \bowtie \text{Vin}$$

R	NUM	Nom	Prénom	Région	Nvin	Cru	Millésime	Viticulteur
	2	Voser	Armande	Alsace	⊥	⊥	⊥	⊥
	3	Hermelin	Jean-Pierre	Champagne	2	Champagne	1996	3
	5	Senard	Danièle	Champagne	2	Champagne	1996	3
	6	Schmidt	Thomas	Bourgogne	4	Chablis	2007	6

$$R = \text{Client} \bowtie_L \text{Vin}$$

Algèbre relationnelle : opérations binaires

■ *Op. de croisement : jointures externes* \bowtie_R

□ Exemple : les vins et les clients de la même région

R	<u>NUM</u>	Nom	Prénom	Région	<u>Nvin</u>	Cru	Millésime	Viticulteur
	3	Hermelin	Jean-Pierre	Champagne	2	Champagne	1996	3
	5	Senard	Danièle	Champagne	2	Champagne	1996	3
	6	Schmidt	Thomas	Bourgogne	4	Chablis	2007	6

$$R = \text{Client} \bowtie \text{Vin}$$

R	<u>NUM</u>	Nom	Prénom	Région	<u>Nvin</u>	Cru	Millésime	Viticulteur
	⊥	⊥	⊥	Bordeaux	1	Saint-Emilion	2002	1
	3	Hermelin	Jean-Pierre	Champagne	2	Champagne	1996	3
	5	Senard	Danièle	Champagne	2	Champagne	1996	3
	⊥	⊥	⊥	Bordeaux	3	Pauillac	1992	1
	6	Schmidt	Thomas	Bourgogne	4	Chablis	2007	6

$$R = \text{Client} \bowtie_R \text{Vin}$$

Algèbre relationnelle : opérations binaires

■ *Op. de croisement : jointures externes* \bowtie

□ Exemple : les vins et les clients de la même région

R	<u>NUM</u>	Nom	Prénom	Région	<u>Nvin</u>	Cru	Millésime	Viticulteur
2	Voser	Armande	Alsace	⊥	⊥	⊥	⊥	
⊥	⊥	⊥	Bordeaux	1	Saint-Emilion	2002	1	
3	Hermelin	Jean-Pierre	Champagne	2	Champagne	1996	3	
5	Senard	Danièle	Champagne	2	Champagne	1996	3	
⊥	⊥	⊥	Bordeaux	3	Pauillac	1992	1	
6	Schmidt	Thomas	Bourgogne	4	Chablis	2007	6	

$$\mathbf{R} = \text{Client} \overset{\circ}{\bowtie} \text{Vin}$$

Algèbre relationnelle

■ Définition

□ Objets : *relation* $R = (T, Sc, K, \mathcal{L})$

Extension : (ensemble de tuples), *schéma* (ensemble d'attributs), *clé primaire* (clé candidate), *contraintes référentielles* (ensemble de clés étrangères)

□ Opérations

■ Nullaires : les relations de la BD

■ Unaires : ρ, σ, π

■ Binaires : $\cup, \cap, \setminus, \times, \bowtie, \bowtie_{\theta}, \overset{\circ}{\bowtie}, \overset{\circ}{\bowtie}_L, \overset{\circ}{\bowtie}_R$

□ Axiomes : ???

□ Calculs

Les équations sont les requêtes

Algèbre relationnelle : Axiomes

■ Axiomatisation

- Propriétés des opérations...
- Espace des opérations couvertes
 - Définition de nouvelles opérations par composition
Autres types de jointure, auto-jointure, division, ...
 - Existence d'opérations non-définissables ?
- Jeux d'opérateurs de base
 - Union, différence, projection, sélection, produit cartésien et renommage
 - Sélection, projection, renommage, jointure naturelle, opérateurs booléens

■ Calcul relationnel

- Ensemble des calculs de la forme : $\{ t \mid P(t) \}$
avec P formule **saine** de la logique du premier ordre (FO)
- Exemple de calcul non sain (*grosso modo*, avec un complémentaire)

$$\{ t \mid t \notin R \}$$

Problème : Trouver les personnes (nom et prénom) et les vins (cru et millésime) auxquels elles sont liées soit en tant que client, soit en tant que viticulteur

Vin	<u>Nvin</u>	Cru	Millésime	Viticulteur	Région
-----	-------------	-----	-----------	-------------	--------

1	Saint-Emilion	2002	1	Bordeaux
2	Champagne	1996	3	Champagne
3	Pauillac	1992	1	Bordeaux
4	Chablis	2007	6	Bourgogne

Client	<u>NUM</u>	Nom	Prénom	Région
--------	------------	-----	--------	--------

Viticulteur	<u>NUM</u>	Nom	Prénom	Région
-------------	------------	-----	--------	--------

1	Barré	Henri	Bordeaux
3	Hermelin	Jean-Pierre	Champagne
4	Fort	Valérie	Languedoc
6	Schmidt	Thomas	Bourgogne

2	Voser	Armande	Alsace
3	Hermelin	Jean-Pierre	Champagne
5	Senard	Danièle	Champagne
6	Schmidt	Thomas	Bourgogne

Commande	<u>NCom</u>	NClient	<u>NVin</u>	Date	Quantité
----------	-------------	---------	-------------	------	----------

1	5	4	27/04/12	25
2	3	2	25/01/13	100
3	2	2	14/08/09	80
4	5	1	08/11/10	100

Bilan de l'algèbre relationnel



Travailler modulo permutations de tuples et d'attributs

Théorie des ensembles

Représenter l'absence d'information

Valeur spéciale \perp partagée par tous les domaines

Respecter l'identité numérique

Théorie des ensembles, clés primaires

Représenter l'intégrité/la régularité des données

Domaines, ???

Représenter les liens sémantiques entre les données

Clés étrangères, contraintes référentielles

Faire du calcul sur les tables

Algèbre relationnelle

Plan



- Modèle logique de données
- Algèbre relationnelle
- En pratique...

Motivations



■ Algèbre relationnel et SGBD

- Intégrité structurelle des données (contraintes référentielles)

Les SGBD permettent de spécifier des contraintes sémantiques

Contraintes d'intégrité

- *Ensemble* de tuples

- Les SGBD permettent de trier les résultats des requêtes
- Les SGBD autorisent la définition de requêtes retournant des doublons

- Tables de la BD vues comme constantes (opérations nullaires)

Les SGBD autorisent les modifications des constantes (?!#\$?!)

Contraintes d'intégrité

■ Contraintes d'intégrité

□ Définition

Propriété de cohérence des données indépendante de l'organisation de la base et liée à la sémantique du modèle de données

□ Langage de spécification

Utilisation de l'algèbre relationnel et de formules logiques du 1^{er} ordre

□ Exemple

La somme des ECTS suivis par un étudiant ne dépasse pas 40 par année

□ Affaire à suivre

■ Traduction entité/association => schéma relationnel

Génération de contraintes d'intégrité (attribut dérivé, spécialisation, ...)

■ Spécification en SQL

Utilisation des **assertions** (et des triggers)

Motivations

■ Algèbre relationnel et SGBD

- Intégrité structurelle des données (contraintes référentielles)

Les SGBD permettent de spécifier des contraintes sémantiques

Contraintes d'intégrité

- *Ensemble* de tuples

- Les SGBD permettent de trier les résultats des requêtes
- Les SGBD autorisent la définition de requêtes retournant des doublons

Algèbre relationnelle multi-ensembliste

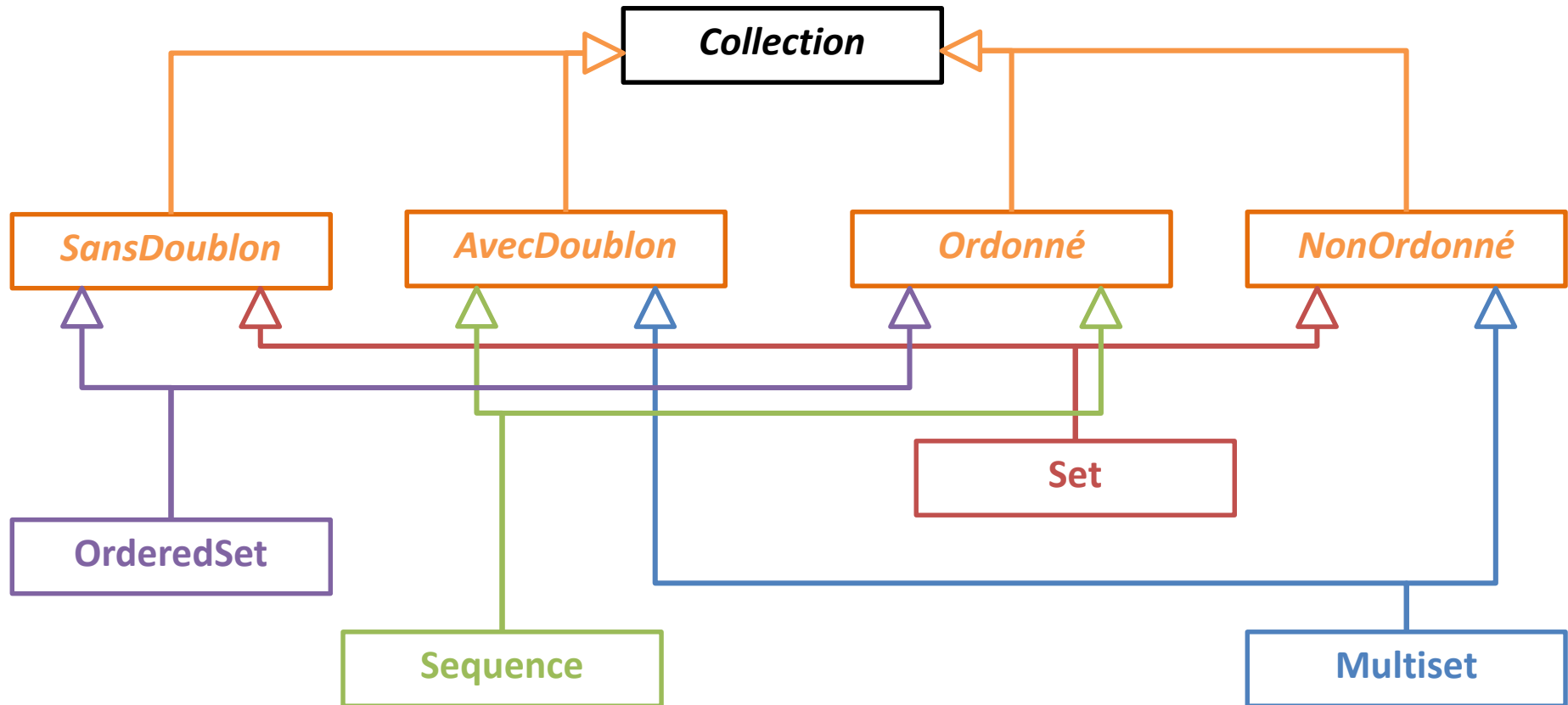
- Tables de la BD vues comme constantes (opérations nullaires)

Les SGBD autorisent les modifications des constantes (?!#\$?!)

Algèbre multi-ensembliste

■ Multi-ensemble

- Collection sans ordre avec doublon



Algèbre multi-ensembliste

■ Multi-ensemble

- Collection sans ordre avec doublon

- Multi-ensemble sur un ensemble E

- Fonction indicatrice $1_A: E \rightarrow \{0,1\}$ d'un sous-ensemble $A \subset E$

$$1_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon} \end{cases}$$

- Multi-ensemble : généralisation aux fonctions $m: E \rightarrow \mathbb{N}$

- $m(x) = 0$: x est absent du multi-ensemble m ,

- $m(x) = 1$: x est présent à un exemplaire dans m ,

- $m(x) = 2$: x est présent à deux exemplaires dans m , etc.

- Appartenance : $x \in m \Leftrightarrow m(x) > 0$

- Opérations classiques

- Union : $(m_1 \cup m_2)(x) = m_1(x) + m_2(x)$

- Intersection : $(m_1 \cap m_2)(x) = \min(m_1(x), m_2(x))$

- Différence : $(m_1 \setminus m_2)(x) = \max(0, m_1(x) - m_2(x))$

Algèbre multi-ensembliste

■ Redéfinition des relations et des opérations

□ Relation $(m_T, Sc, K, \mathcal{L})$

L'extension est donnée par un multi-ensemble de tuples

□ Définition des opérations classiques dans le cadre multi-ensembliste

$$(m'_T, \{A_1, \dots, A_n\}, K, \mathcal{L}) = \tilde{\sigma}_\varphi(m_T, \{A_1, \dots, A_n\}, K, \mathcal{L})$$

$$m'_T(t) = \begin{cases} m'_T(t) & \text{si } \varphi[A_1 \leftarrow A_1(t), \dots, A_n \leftarrow A_n(t)] \\ 0 & \text{sinon} \end{cases}$$

$$(m_{T_1 \times T_2}, Sc, K, \mathcal{L}) = (m_{T_1}, Sc_1, K_1, \mathcal{L}_1) \tilde{\times} (m_{T_2}, Sc_2, K_2, \mathcal{L}_2)$$

$$m_{T_1 \times T_2}(t_1, t_2) = m_{T_1}(t_1) * m_{T_2}(t_2)$$

Algèbre multi-ensembliste

■ Opération de **groupement** γ

- Spécification informelle

Identifier les tuples de mêmes valeurs pour un sous-ensemble d'attributs

- Notation

$$\gamma_{A_1, \dots, A_k}(R)$$

- Sémantique multi-ensembliste

- Soit une relation $R = (m_T, Sc, K, \mathcal{L})$
- Soit $A = \{A_1, \dots, A_k\}$ un sous-schéma de Sc

$$\gamma_{A_1, \dots, A_k}(R) = (m''_T, A, A, \mathcal{L}')$$

$$\text{avec } \tilde{\pi}_{A_1, \dots, A_k}(R) = (m'_T, A, A, \mathcal{L}')$$

$$\text{et } m''_T(t) = \begin{cases} 1 & \text{si } m'_T(t) > 0 \\ 0 & \text{sinon} \end{cases}$$

Algèbre multi-ensembliste

■ Opération de **grouperment** γ avec agrégation

□ Notation

$$\gamma_{A_1, \dots, A_k, op_1(B_1) \rightarrow C_1, \dots, op_l(B_l) \rightarrow C_l}(R)$$

□ Sémantique multi-ensembliste

- Soit une relation $R = (m_T, Sc, K, \mathcal{L})$
- Soit $A = \{A_1, \dots, A_k\}$ et $B = \{B_1, \dots, B_l\}$ deux sous-schémas de Sc
- Soit $C = \{C_1, \dots, C_l\}$ un nouveau schéma
- Soient les op. d'agrégation $\{Op_i\}_{i \in \llbracket 1, l \rrbracket}$ parmi $\{SUM, AVG, MIN, MAX, CNT\}$

$$\gamma_{A_1, \dots, A_k, op_1(B_1) \rightarrow C_1, \dots, op_l(B_l) \rightarrow C_l}(R) = (m'_T, A \cup C, A, \mathcal{L}')$$

$$\text{avec } \gamma_{A_1, \dots, A_k}(R) = (m'_T, A, A, \mathcal{L}')$$

$$\text{et } C_i(t) = Op_i\{m_T(t') * B_i(t') \mid A(t') = A(t)\}$$

Motivations

■ Algèbre relationnel et SGBD

- Intégrité structurelle des données (contraintes référentielles)

Les SGBD permettent de spécifier des contraintes sémantiques

Contraintes d'intégrité

- *Ensemble* de tuples

- Les SGBD permettent de trier les résultats des requêtes
- Les SGBD autorisent la définition de requêtes retournant des doublons

Algèbre relationnelle multi-ensembliste

- Tables de la BD vues comme constantes (opérations nullaires)

Les SGBD autorisent les modifications des constantes (?!#\$?!)

Variables relationnelles

Variable relationnelle



■ ... ou relvar

- Distinction dans une de relation entre
 - Valeur : extension
 - Type : schéma, clé primaire et clés étrangères
- Mise à jour d'une relation
 - Relation vue comme une variable
 - Ajout, modification ou retrait de tuple
 - Modification de la valeur
 - Respect du type
- Origine et controverse
 - Introduit par C. J. Date (et H. Darwen)
 - Équivalent au concept de table en SQL



-- FIN --

Plan



■ Algèbre relationnel dans la pratique

- Contraintes d'intégrité
- Table multi-tuple
 - Cadre ensembliste limitant passage aux multi-ensembles
 - En fait les relations stockées sont ensembles, celles calculées sont multi-ens.
 - Redéfinition des opérations dans le cadre multi-ensemblistes
 - Nouvelles opérations : groupement/agrégation
- Relvar
 - Relation telle que vue dans algèbre = constante
 - Quid de l'update d'une relation ?
 - Relvar ensembliste, retour de requête/vue multi-ensembliste

Algèbre relationnelle : opérations binaires

■ Op. de croisement : quotient \div

- Spécification informelle

Eviter de perdre des tuples défectueux en utilisant \perp pour les attributs non-définis

- Notation

$$R_1 \overset{\circ}{\bowtie}_L R_2 \quad R_1 \overset{\circ}{\bowtie} R_2 \quad R_1 \overset{\circ}{\bowtie}_R R_2$$

- Sémantique ensembliste

Soient $R_1 = (T_1, SC_1, K_1, L_1)$ et $R_2 = (T_2, SC_2, K_2, L_2)$

- $R_1 \overset{\circ}{\bowtie}_L R_2 = (R_1 \bowtie R_2) \cup \left(\left(R_1 \setminus \pi_{SC_1}(R_1 \bowtie R_2) \right) \times R_{\perp} \right)$

- $R_1 \overset{\circ}{\bowtie}_R R_2 = R_2 \overset{\circ}{\bowtie}_L R_1$

- $R_1 \overset{\circ}{\bowtie} R_2 = R_1 \overset{\circ}{\bowtie}_L R_2 \cup R_1 \overset{\circ}{\bowtie}_R R_2$

- Remarque

- En l'absence de mention supplémentaire, une jointure externe est naturelle
- R_{\perp} : relation singleton contenant le tuple (\perp, \dots, \perp) sur le schéma $SC_2 \setminus SC_1$

Plan

■ Algèbre relationnelle

- Qu'est qu'une algèbre
 - Al-jabr => (système d')équation(s), définition et résolution
 - Exemple avec les entiers
- Objets
 - Domaines et relations
 - Attribut
 - Clé (primaire et étrangère/secondaire)
 - Lien avec la représentation tabulaire (sur l'exemple précédent)
- Opérations
 - Opérations à schéma fixe (ensemblistes/booléennes, renommage)
 - Opérations à schéma variable (réduction, augmentation)
- Sur la sémantique $\{ t \text{ in } R \mid P(t) \}$
 - Axiomatisation minimale (tout en théorie des ensembles)
 - Calcul relationnel et équivalence avec algèbre relationnel

Algèbre relationnelle : opérations

Union

R	<u>NUM</u>	Nom	Prénom	Région
	1	Barré	Henri	Bordeaux
	2	Voser	Armande	Alsace
	3	Hermelin	Jean-Pierre	Champagne
	4	Fort	Valérie	Languedoc
	5	Senard	Danièle	Champagne
	6	Schmidt	Thomas	Bourgogne

Intersection

R	<u>NUM</u>	Nom	Prénom	Région
	3	Hermelin	Jean-Pierre	Champagne
	6	Schmidt	Thomas	Bourgogne

Différence

R	<u>NUM</u>	Nom	Prénom	Région
	1	Barré	Henri	Bordeaux
	4	Fort	Valérie	Languedoc

Algèbre relationnelle : opérations

age

R	<u>ID</u>	Lastname	Firstname	Region
	1	Barré	Henri	Bordeaux
	3	Hermelin	Jean-Pierre	Champagne
	4	Fort	Valérie	Languedoc
	6	Schmidt	Thomas	Bourgogne

sélection : les vins de bordeaux, les vins d'alsace

R	<u>NVin</u>	Cru	Millésime	Viticulteur	Région
	1	Saint-Emilion	2002	1	Bordeaux
	3	Pauillac	1992	1	Bordeaux

R	<u>NVin</u>	Cru	Millésime	Viticulteur	Région
---	-------------	-----	-----------	-------------	--------

n

R	<u>NCom</u>	NVin	Quantité
	1	4	25
	2	2	100
	3	2	80
	4	1	100

Pour chaque vin la liste des commandes avec qt

R	Viticulteur
	1
	3
	1
	6

Liste des viticulteurs actifs

Algèbre relationnelle : opérations

theta join : les vins commandé plus lors d'une commande à plus de 30 exemplaires

R	<u>Nvin</u>	Cru	Millésime	Viticulteur	Région	<u>NCom</u>	NClient	Date	Quantité
	1	Saint-Emilion	2002	1	Bordeaux	4	5	08/11/10	100
	2	Champagne	1996	3	Champagne	3	2	14/08/09	80
	2	Champagne	1996	3	Champagne	2	3	25/01/13	100

Équi join :vins et viticulteurs

R	<u>Nvin</u>	Cru	Millésime	Viticulteur	Vin.Région	<u>NUM</u>	Nom	Prénom	Viticulteur. Région
	1	Saint-Emilion	2002	1	Bordeaux	1	Barré	Henri	Bordeaux
	2	Champagne	1996	3	Champagne	3	Hermelin	Jean-Pierre	Champagne
	3	Pauillac	1992	1	Bordeaux	1	Barré	Henri	Bordeaux
	4	Chablis	2007	6	Bourgogne	6	Schmidt	Thomas	Bourgogne

R	<u>Nvin</u>	Cru	Millésime	Viticulteur	Région	<u>NUM</u>	Nom	Prénom
	2	Champagne	1996	3	Champagne	3	Hermelin	Jean-Pierre
	2	Champagne	1996	3	Champagne	5	Senard	Danièle
	4	Chablis	2007	6	Bourgogne	6	Schmidt	Thomas

ner

Algèbre relationnelle : opérations

R	<u>Nvin</u>	Cru	Millésime	Viticulteur	Région	<u>NUM</u>	Nom	Prénom
	1	Saint-Emilion	2002	1	Bordeaux	⊥	⊥	⊥
	2	Champagne	1996	3	Champagne	3	Hermelin	Jean-Pierre
	2	Champagne	1996	3	Champagne	5	Senard	Danièle
	3	Pauillac	1992	1	Bordeaux	⊥	⊥	⊥
	4	Chablis	2007	6	Bourgogne	6	Schmidt	Thomas

R	<u>Nvin</u>	Cru	Millésime	Viticulteur	Région	<u>NUM</u>	Nom	Prénom
	⊥	⊥	⊥	⊥	Alsace	2	Voser	Armande
	2	Champagne	1996	3	Champagne	3	Hermelin	Jean-Pierre
	2	Champagne	1996	3	Champagne	5	Senard	Danièle
	4	Chablis	2007	6	Bourgogne	6	Schmidt	Thomas

R	<u>Nvin</u>	Cru	Millésime	Viticulteur	Région	<u>NUM</u>	Nom	Prénom
	1	Saint-Emilion	2002	1	Bordeaux	⊥	⊥	⊥
	⊥	⊥	⊥	⊥	Alsace	2	Voser	Armande
	2	Champagne	1996	3	Champagne	3	Hermelin	Jean-Pierre
	2	Champagne	1996	3	Champagne	5	Senard	Danièle
	3	Pauillac	1992	1	Bordeaux	⊥	⊥	⊥

Algèbre relationnelle : opérations

Vin	<u>Nvin</u>	Cru	Millésime	Viticulteur	Région
1		Saint-Emilion	2002	1	Bordeaux
2		Champagne	1996	3	Champagne
3		Pauillac	1992	1	Bordeaux
4		Chablis	2007	6	Bourgogne

Commande	<u>NCom</u>	NClient	NVin	Date	Quantité
1		5	4	27/04/12	25
2		3	2	25/01/13	100
3		2	2	14/08/09	80
4		5	1	08/11/10	100

Algèbre relationnelle : opérations

Client	<u>NUM</u>	Nom	Prénom	Région
	2	Voser	Armande	Alsace
	3	Hermelin	Jean-Pierre	Champagne
	5	Senard	Danièle	Champagne
	6	Schmidt	Thomas	Bourgogne

Commande	<u>NCom</u>	NClient	NVin	Date	Quantité
	1	5	4	27/04/12	25
	2	3	2	25/01/13	100
	3	2	2	14/08/09	80
	4	5	1	08/11/10	100

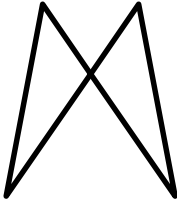
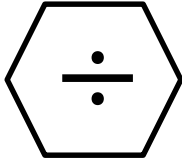
Algèbre relationnelle : opérations

Vin	<u>NVin</u>	Cru	Millésime	Viticulteur	Région
	1	Saint-Emilion	2002	1	Bordeaux
	2	Champagne	1996	3	Champagne
	3	Pauillac	1992	1	Bordeaux
	4	Chablis	2007	6	Bourgogne

Client	<u>NUM</u>	Nom	Prénom	Région
	2	Voser	Armande	Alsace
	3	Hermelin	Jean-Pierre	Champagne
	5	Senard	Danièle	Champagne
	6	Schmidt	Thomas	Bourgogne

Récapitulatif

■ Algèbre relationnel : opérations

Vocabulaire	Notation graphique	Algèbre relationnelle
θ -jointure		$R_1 \bowtie_{\theta} R_2$
Jointure naturelle		$R_1 \bowtie R_2$
Jointure externe		$R_1 \bowtie_{LR} R_2$
Division		$R_1 \div R_2$


Récapitulatif

■ Algèbre relationnel : objets

Vocabulaire	Représentation tabulaire	Algèbre relationnelle
Domaine	Valeur (dont NULL)	D (avec $\perp \in D$)
Relation	Table	$R \subset D_1 \times \dots \times D_n$
Arité	Largeur de la table	n
Tuple	Ligne	$(v_1, \dots, v_n) \in R$
Attribut	Colonne	$A_i: R \rightarrow D_i$ avec $i \in \llbracket 1..n \rrbracket$
Schéma	Entête	$D_1 \times \dots \times D_n$
Clé	Attribut souligné	$k \subset \llbracket 1..n \rrbracket$

Récapitulatif

■ Algèbre relationnel : opérations

Vocabulaire	Notation graphique	Algèbre relationnelle
Union		$R_1 \cup R_2$
Intersection		$R_1 \cap R_2$
Différence		$R_1 \setminus R_2$
Renommage		$\rho_{(att \rightarrow att')}(R)$
Sélection		$\sigma_{\varphi}(R)$
Projection		$\pi_{(att_1, att_2, \dots)}(R)$
Produit cartésien		$R_1 \times R_2$