Modèle relationnel
Algèbre relationnelle

Modèle relationnel (Codd 1970)

- *On considère* D_i où $i = 1,2..n$ des ensembles, dits *domaines*
- *Un domaine = ensemble de valeurs (ex. D1 : entiers sur 10 positions, D2 : chaînes de 8 caract., …)*
- *Attribut = variable à valeurs dans un domaine*
- Une relation r sur les attributs A1, A2, ..., An de domaines respectifs D1, D2, ..., Dn, est un sous-ensemble du produit cartésien D1xD2x … x Dn.
- Schéma de la relation r = R = liste de ses attributs avec leurs domaines (souvent, on omet les domaines)

Modèle relationnel (Codd 1970)

ex.

ETUD (num-et:D1, nom-et:D2) ou ETUD (num-et, nom-et)

La relation etud ={(102, 'jacques'), (304, 'Marc')}

Représentation tabulaire :

Relation etud

num-et	nom-et
102	Jacques
304	Marc

Dépendance fonctionnelle (df)

Soit R(A1, A2, ..., An) un schéma de relation. Soit X et Y des sous ensembles de l'ensemble d'attributs {A1,A2,...An) :

On dit que Y dépend fonctionnellement de X (ou X est en df avec Y), noté X-->Y, si à chaque valeur de X correspond une valeur unique de Y.

on écrit : X $\rightarrow$ Y, on dit que X détermine Y

Exemples :

PRODUIT (num_prod, nom, prixUHT), on a :

num_prod $\rightarrow$ (nom, prixUHT)

NOTE (num_contrôle, num_élève, note), on a :

(num_contrôle, num_élève) $\rightarrow$ note

Une df est une propriété sémantique, elle correspond à une hypothèse vraie d'une contrainte du monde réel.

df élémentaire : une df X -> A, où A est un attribut unique non inclus dans X, est élémentaire, s'il n'existe pas Z inclus dans X tel que Z -> A.

Exemple : si X ={A, B} est en df avec C et si $A \rightarrow C$ alors la df $X \rightarrow C$ n'est pas élémentaire (car une partie de X est en df avec C).

df directe : une df X -> A, où A est un attribut unique non inclus dans X, est directe, s'il n'existe pas de Z tel que
$X \rightarrow Z$ et $Z \rightarrow A$ (pas de transitivité)

Exemple : soit X ={A, B} et Y={C,D,E}

Si on a $X \rightarrow Y$ et $C \rightarrow D$ alors la df $X \rightarrow Y$ n'est pas directe car on a : $X \rightarrow C$ et $C \rightarrow D$ (transitivité).

Clé d'une relation : attribut (ou groupe minimum d'attributs) qui détermine tous les autres (en df avec tous les attributs non clés).

Exemple :

PRODUIT (<u>num_prod</u>, nom, prixUHT), on a :
num_prod → (nom, prixUHT)
num_prod est une clé

- Une clé détermine tous les attributs de façon unique (par définition, dans une relation, une clé est en df avec tous les autres attributs de la relation)
- Pour trouver la clé d'une relation, il faut examiner attentivement les hypothèses sur le monde réel
- Une relation peut posséder plusieurs clés. On les appelle clés candidates.

Exemple :

Dans la relation PRODUIT, *nom* peut être une clé candidate (si dans notre système d'information, on n'a jamais deux produits de même nom)

Clé primaire : c'est une clé choisie parmi les clés candidates.

Clé étrangère : c'est un attribut (ou groupe d'attributs) qui fait référence à la clé primaire d'une autre relation

Exemple : soit les relations :

CATEG (<u>num_cat</u>, design, tva)
PRODUIT(<u>num_prod</u>, nom, marque, **num_cat**, prixUHT)
On a :
num_cat dans PRODUIT est une clé étrangère

CLÉ ÉTRANGÈRE = CLÉ PRIMAIRE dans une autre relation

Formes normales et df

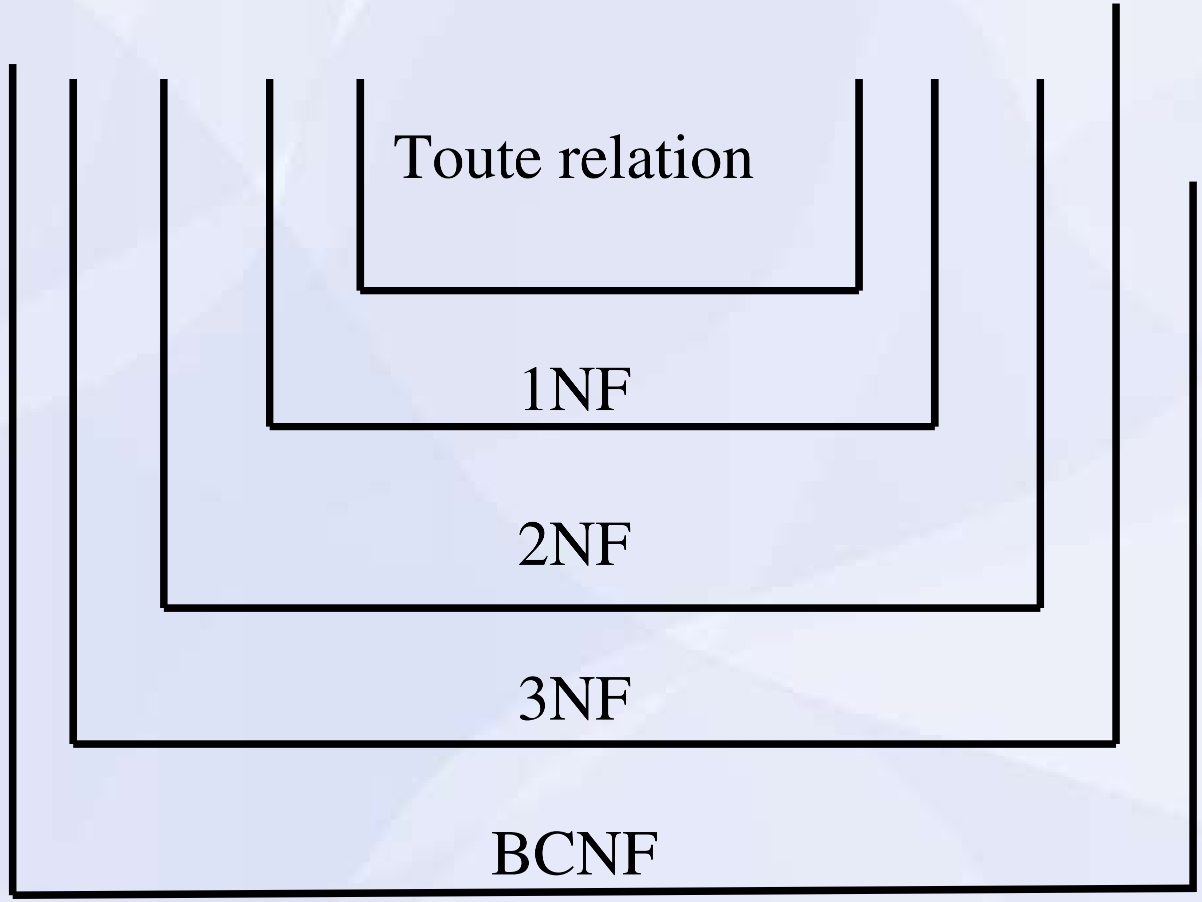
Les relations d'un schéma relationnel sont classées en plusieurs catégories selon leur « normalité » : plus la forme normale d'une relation est élevée, moins il y a de redondances dans la relation.

1. Relation en 1ère forme normale (1NF) : Si tous ses attributs sont atomiques c-à-d ne prennent pas une liste de valeurs.

2. Relation en 2ème forme normale (2NF) : Si elle est en 1NF, et si toutes les df sont élémentaires, c-à-d que chaque attribut non clé dépend de toute la clé (non d'une partie de la clé)
==> concerne les clés composées de 2 attributs ou plus.

3. Relation en 3ème forme normale (3NF) : Si elle est en 2NF, et si toutes les df sont directes, c-à-d qu'aucun attribut non clé ne dépend d'un autre attribut non clé.

4. Relation en BCNF (Boyce-Codd) (BCNF) : Si elle est en 3NF, et si aucun attribut non clé n'est en df avec la clé (ou une partie de la clé).


Toute relation
1NF
2NF
3NF
BCNF

Exemples

- **1NF** : ELEVE (num_e, nom_e, prenom_e, liste_notes) n'est pas en 1NF car l'attribut *liste_notes* n'est atomique.

On décompose :

ELEVE (num_e, nom_e, prenom_e) : 1NF
NOTE (num_e, num_mat, note) : 1NF

- 2NF :

COMMANDE (date, num_cli, num_pro, qte, nom_prod, prixUHT) est en 1NF et non en 2F car la clé est (date, num_cli, num_pro), et nom_prod (par ex.) ne dépend que de num_pro (on a num_pro→ nom_prod).

On décompose :

COMMANDE (date, num_cli, num_prod, qte) : 2NF
PRODUIT (num_pro, nom_prod, prixUHT) : 2NF

- 3NF :

VOITURE (matricule, marque, modèle, puissance)

→ 2NF, non 3NF car la clé est matricule, et la puissance dépend de (marque, modèle)

(donc matricule -> (marque, modèle) car c'est une clé et

(marque, modèle) -> puissance

→ donc transitivité

On décompose :

VOITURE (matricule, marque, modèle)

MODELE (marque, modèle, puissance)

Algèbre relationnelle

Exemples de relations

Relation ETUDIANT

Num-et	nom-et	dipl-et	date-nai
101	Dutois	Licence2	11/05/1992
102	Legent	DUT	16/10/1991

Relation VEHICULE

num-imm	marque	game	puiss-chev	annee-circ
AZ323CD	Peugeot	308	7	2008
FD544DE	Renault	laguna3	8	2010
AV556	Peugeot	309	5	2007

Les langages relationnels :

- sont utilisés pour effectuer des requêtes sur une BD relationnelle

- utilisent 2 approches qui expriment les mêmes opérations :
  - algèbre relationnelle
  - calcul relationnel

Algèbre relationnelle

Les opérateurs de l'algèbre relationnelle sont des opérateurs ensemblistes.

Un opérateur prend en entrée une ou deux relations (ensembles de tuples de la base) et retourne un résultat qui est également une relation

Il existe 5 opérateurs de base :

- les opérateurs unaires : selection et projection
- les opérateurs binaires : union, différence et produit cartésien

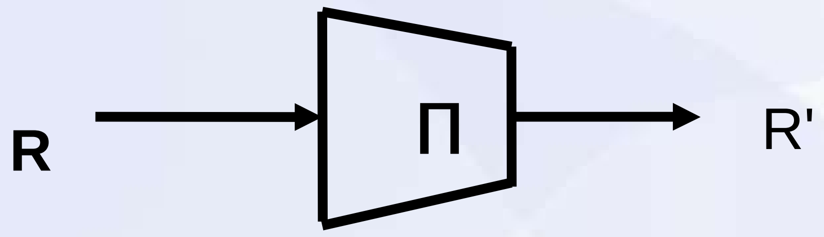
D’autres opérateurs existent qui peuvent s’exprimer à l’aide des opérateurs de base. Ce sont la jointure, la division et l ’intersection

La projection : Π

Projection des attributs d’une relation R sur un sous ensemble des attributs de R :

$\Pi_{A1, A2, \ldots, Ak}(R) = R(A1, A2, \ldots, Ak)$

où $A_1, A_2, \ldots, A_k$ sont un sous-ensemble du schéma de la relation R (égal ou inclus).



La projection sur $A_1, A_2, \ldots, A_k$ élimine tous les autres attributs de la relation et supprime les tuples dupliqués.

Exemple de projection

R

X	Y	Z
a	b	c
d	a	b
c	b	d
a	b	e
e	e	a

$\Rightarrow \pi_{X,Y}(R) =$

R'

X	Y
a	b
d	a
c	b
e	e

R = (X,Y,Z) et R' $=\pi_{X,Y}(R)$ = projection de R sur les attributs X et Y

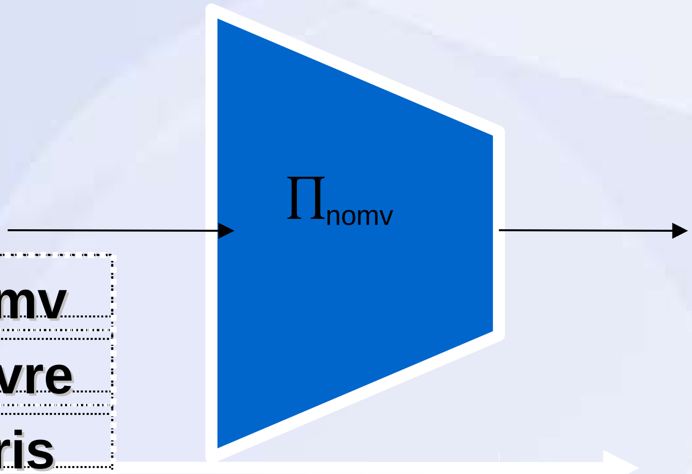
Exemple de projection (2)

- Requête : Soit la relation Ville (id, nomp, nomv)
  - Quels sont les villes de résidence des personnes de la base (projection sur l'attribut nomv)

$$\mathbf{Tab2 = \mu_{nomv}(Tab_1)}$$

Tab_1

id	nomp	nomv
1500	Marc	Havre
3333	Durand	Paris
3500	Jacques	Paris



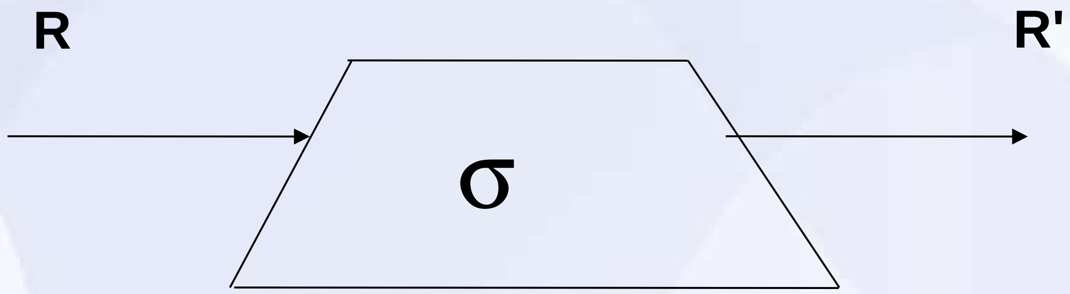
Tab_2

nomv
Havre
Paris

Selection (restriction)

La selection se fait en fonction d'une condition C portant sur des attributs de R. Le résultat est une relation dont les attributs satisfont la condition.

On note : $\sigma_C(R)$



Exemple 1 : de selection

R

X	Y	Z
a	b	1
d	a	2
c	b	3
a	b	4
e	e	5

$\Rightarrow \sigma_{Y='b'}(R) = R'$

R'

X	Y	Z
a	b	1
c	b	3
a	b	4

R

X	Y	Z
a	b	1
d	a	2
c	b	3
a	b	4
e	e	5

$\Rightarrow \sigma_{C}(R) = R'$

où C : $(X='a' \vee Y='a') \wedge Z \leq 3$

R'

X	Y	Z
a	b	1
d	a	2

EXEMPLE 2 : Projection et sélection

ETUDIANTS (Numero, Nom, Addresse)

Algèbre relationnelle : Numéro de l'étudiant 'Dumont'

$$\Pi_{Numero}\ (\sigma_{Nom='\ Dumont'}\ (ETUDIANTS))$$ => projection + sélection

Projection sur le numero

Sélection de la ligne où le nom égal 'Dumont'

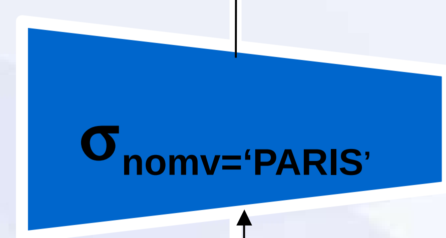
Exemple 3 : sélection

- Requête :
  - Sélectionner tous les individus habitant à Paris.

$$\mathbf{Tab1_2 = \sigma}_{\text{ville='Paris'}}(\text{Tab1})$$

P2

ID	NOM	VILLE
3333	Durant	Paris
3500	Jacques	Paris



P1

ID	NOM	VILLE
1500	Marc	Havre
3333	Durand	Paris
3500	Jacques	Paris

Les conditions de sélection

Il s'agit d 'une formule logique qui relie, par des connecteurs logiques (AND, OR, NOT), des expressions de la forme :

- A_i **op** A_j ou (A_i **op** a) ou A_i

où A_i, A_j sont des attributs de la relation R

a est un élement (une valeur) du domaine de A_i

op est un opérateur de comparaison : =, <, >, <=, >=, <>

Expressions de l 'algèbre relationnelle

- L 'algèbre relationnelle est fermée par rapport aux opérations de l 'A.R. : le résultat d 'une opération relationnelle est aussi une relation.

- Les opérations relationnelles peuvent être combinées et former des expressions plus complexes.
- Ex : Relation R = Commande (nom, prenom, nomc, qte)

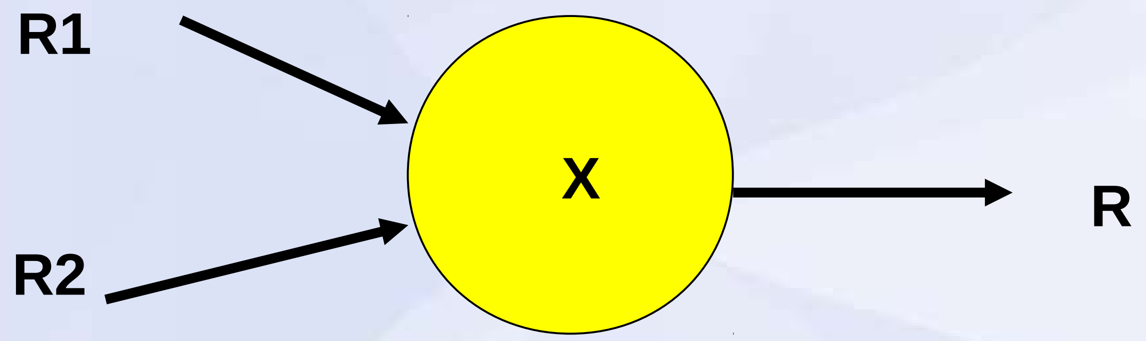
$R' = \pi_{pnom}(\sigma_{nom='paul'}(commande))$

R'' = $\sigma_{nom='paul'}$(commande)) : contient les commandes de 'paul'

Produit cartésien

- Produit cartésien de la relation R_1 par la relation R_2 : $R_1 \times R_2$
- Argument : 2 relations quelconques
  $R_1 (A_1, A_2, \ldots, A_n)$ et $R_2 (B_1, B_2, \ldots, B_k)$
- Schéma de la relation résultat T : $R_1 \times R_2$: $(A_1, \ldots, A_n, B_1, \ldots, B_k)$
- Occurrences de R = ensemble des tuples ayant n+k attributs :
  - dont les valeurs des n premiers attributs sont les tuples de R_1
  - et les k dernières sont les tuples de R_2

Produit cartésien



Exemple

R

A	B
1	1
1	2
3	4

S

C	D	E
a	b	a
a	b	c
b	a	a

R x S

A	B	C	D	E
1	1	a	b	a
1	2	a	b	a
3	4	a	b	a
1	1	a	b	c
1	2	a	b	c
3	4	a	b	c
1	1	b	a	a
1	2	b	a	a
3	4	b	a	a

Jointure naturelle

- Soient 2 relations R et S ayant des attributs en commun

  $R(A_1, \ldots, A_m, \mathbf{X_1, \ldots, X_K})$

  $S(B_1, \ldots, B_n, \mathbf{X_1, \ldots, X_k})$

- Schéma de la relation R $\bowtie$ S, jointure naturelle de R et S :

  $T\,(A_1, \ldots, A_m, B_1, \ldots, B_n, X_1, \ldots, X_K\,)$

  Un tuple de R $\bowtie$ S comporte donc (m+n+k) attributs.

R

A	**B**	**C**
a	b	c
d	b	c
b	b	f
c	a	d

S

B	**C**	**D**
b	c	d
b	c	e
a	d	b

R $\bowtie$ S

A	**B**	**C**	**D**
a	b	c	d
a	b	c	e
d	b	c	d
d	b	c	e
c	a	d	b

Jointure naturelle (suite)

- La jointure naturelle correspond à un produit cartésien, suivi d'une sélection.
- Soient 2 relations R et S ayant des attributs en commun

  $R(A_1, \ldots, A_m, \mathbf{X_1, \ldots, X_K})$

  $S(B_1, \ldots, B_n, \mathbf{X_1, \ldots, X_k})$. Soit $V = \{X_1, \ldots, X_k\}$

$R \bowtie S = \Pi_U (\sigma_{\forall A \in V : R.A=S.A} (R \times S))$

où U = l 'ensemble des attributs de R et S et

R.A est l'attribut A de R

Jointure naturelle - algorithme

- Début

  Pour tout tuple a de R et tout tuple b de S :

  1. Concaténer a et b. On obtient un tuple avec comme attributs a | b, c-à-d :

     $A_1, \ldots, A_m, X_1, \ldots, X_k, B_1, \ldots, B_n, X_1, \ldots, X_k$

  2. Ne garder ce tuple que si chaque attribut X_i de a est égal à l'attribut X_i de b : $\forall_{i=1::..k}\ a.X_i = b.X_i$

  3. Eliminer les valeurs (colonnes) dupliquées. On obtient, pour la jointure naturelle, un tuple avec comme attributs :

$$\underbrace{A_1, \ldots, A_m}_{a}, \underbrace{B_1, \ldots, B_n}_{b}, \underbrace{X_1, \ldots, X_k}_{a+b}$$

θ-jointure

- Notée : $R \bowtie_{Ai\,\theta\,Bj} S$, où $\theta \in \{=, , <>, \leq, \geq, \neq\}$
- C 'est une jointure entre 2 relations R et S avec :
- $R = (A_1, \ldots, A_m)$, $S = (B_1, \ldots, B_n)$
- Schéma de $T = R \bowtie_{Ai\,\theta\,Bj} S = (A_1, \ldots, A_m, B_1, \ldots, B_n)$
- La valeur de T est : $\sigma_{Ai\,\theta\,Bj}$ (RxS) : sélection des tuples de RxS tels que $A_i\ \theta\ B_j$
- Equijointure : on parle de équijointure quand l'opérateur θ est l'égalité.

Exemple de θ-jointure

R

A	B
1	a
1	b
3	a

S

C	D	E
1	b	a
2	b	c
4	a	a

RxS

A	B	C	D	E
1	a	1	b	a
1	a	2	b	c
1	a	4	a	a
1	b	1	b	a
1	b	2	b	c
1	b	4	a	a
3	a	**1**	b	a
3	a	**2**	b	c
3	a	4	a	a

A > C →

A > C →

$R \bowtie_{A \leq C} S$

A	B	C	D	E
1	a	1	b	a
1	a	2	b	c
1	a	4	a	a
1	b	1	b	a
1	b	2	b	c
1	b	4	a	a
3	a	4	a	a

Exemple d 'équijointure

R

A	B
1	a
1	b
3	a

S

C	D	E
1	b	a
2	b	c
4	a	a

RxS

A	B	C	D	E
1	**a**	1	**b**	a
1	**a**	2	**b**	c
1	a	4	a	a
1	b	1	b	a
1	b	2	b	c
1	**b**	4	**a**	a
3	**a**	1	**b**	a
3	**a**	2	**b**	c
3	a	4	a	a

B $\neq$ D

$R \bowtie_{B=D} S$

A	B	C	D	E
1	a	4	a	a
1	b	1	b	a
1	b	2	b	c
3	a	4	a	a

$= \sigma_{B=D}(RxS)$

UNION

- Soient 2 relations $R(A_1, \ldots, A_m)$ et $S(A_1, \ldots, A_m)$
- Le schéma de $T = R \cup S$ est : $T(A_1, \ldots, A_m)$
- Les tuples de T : union ensembliste sur $D_1 x \ldots x D_m$ où D_i est le domaine de A_i. Les doublons sont éliminés.

$$T = R \cup S = \{t \,/\, t \in R \vee t \in S\}$$

R

A	B
a	b
a	c
d	e

S

A	B
a	b
a	e
d	e
f	g

R ∪ S

A	B
a	b
a	c
d	e
a	e
f	g

DIFFERENCE

- Soient 2 relations $R(A_1, \ldots, A_m)$ et $S(A_1, \ldots, A_m)$
- Le schéma de $T = R - S$ est : $T(A_1, \ldots, A_m)$
- Les tuples de T : différence ensembliste sur $D_1 x \ldots x D_m$ où D_i est le domaine de A_i

$$T = R - S = \{t \,/\, t \in R \wedge t \notin S\}$$

R

A	B
a	b
a	c
d	e

S

A	B
a	b
a	e
d	e
f	g

R - S

A	B
a	c

S - R

A	B
a	e
f	g

INTERSECTION

- Soient 2 relations $R(A_1, \ldots, A_m)$ et $S(A_1, \ldots, A_m)$
- Le schéma de $T = R \cap S$ est : $T\,(A_1, \ldots, A_m)$
- Les tuples de T : intersection ensembliste sur $D_1 x \ldots x\ D_m$ avec

D_i domaine de A_i

R $\cap$ S = R - (R - S)

(intersection)

$$T = R \cap S = \{t \,/\, t \in R \wedge t \in S\}$$

R

A	B
→ a	b
a	c
→ d	e

S

A	B
→ a	b
a	e
→ d	e
f	g

R - S

(différence)

A	B
a	c

A	B
a	b
d	e

Semijointure

- La semijointure $R \ltimes S$ correspond à une projection sur les attributs de R de la jointure naturelle entre R et S : $\Pi_U (R \bowtie S)$
- Si $R = (A_1, \ldots, A_m, \mathbf{X_1, \ldots, X_k})$ et $S = (B_1, \ldots, B_n, \mathbf{X_1, \ldots, X_k})$ alors le schéma de T : $R \ltimes S = (A_1, \ldots, A_m, \mathbf{X_1, \ldots, X_k})$
- **Exemple :**

R

A	B	C
a	b	c
d	b	c
b	b	f
c	a	d

S

B	C	D
b	c	d
b	c	e
a	d	b

=> $R \bowtie S$

A	B	C	D
a	b	c	d
a	b	c	e
d	b	c	d
d	b	c	e
c	a	d	b

jointure naturelle

=> $R \ltimes S$

A	B	C
a	b	c
d	b	c
c	a	d

semijointure

Division

R

A	B	C	D
a	b	x	m
a	b	y	n
a	b	z	o
b	c	x	o
b	d	x	m
c	e	x	m
c	e	y	n
c	e	z	o
d	a	z	p
d	a	y	m

S

C	D
x	m
y	n
z	o

R ÷ S

A	B
a	b
c	e

Division : exemple

COMM

Num	nom	p-nom	qte
1	**Jean**	**briques**	**100**
2	**Jean**	**ciment**	**10**
3	**Jean**	**plâtre**	**5**
4	Paul	briques	300
5	Paul	platre	8
5	Vincent	platre	15

PROD

P-nom
briques
ciment
plâtre

COMM ÷ PROD

nom
Jean

Client(s) qui ont commandé tous les produits

Division : formellement

- Soient $R(A_1, \ldots, A_m, X_1, \ldots, X_k)$ et $S (X_1, \ldots, X_k)$, c-à-d les attributs de S sont inclus dans R. Le résultat T a pour schéma : $T (A_1, \ldots, A_m)$
- Tuples de $T = R \div S$

$$=\{(a_1, \ldots, a_m) / \forall (x_1, \ldots, x_k) \in S : (a_1, \ldots, a_m, x_1, \ldots, x_k) \in R$$

- La division peut s'exprimer en utilisant les opérateurs *produit cartésien, projection et différence* : $R \div S = R1 - R2$
  où : $R1 = \Pi_{A1, \ldots, Am}(R)$ **et** $R2 = \Pi_{A1, \ldots, Am}((R1 x S) - R)$

Opération de renommage

- Notée : ρ , c ’est une opération unaire.
- Consiste à remplacer le nom d’un attribut par un autre.
- Soit R (A_1, …, A_m), on a :

  $\rho_{Ai \rightarrow Bi}$ R : T (A_1, …, $\mathbf{A_{i-1}}$, $\mathbf{B_i}$, $\mathbf{A_{i+1}}$, ..., A_m)