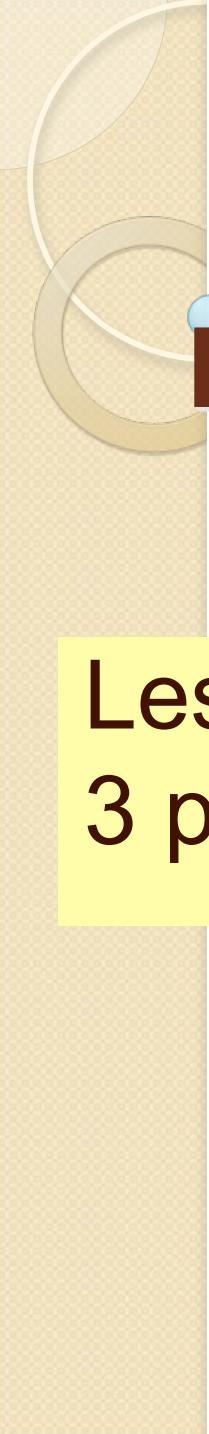


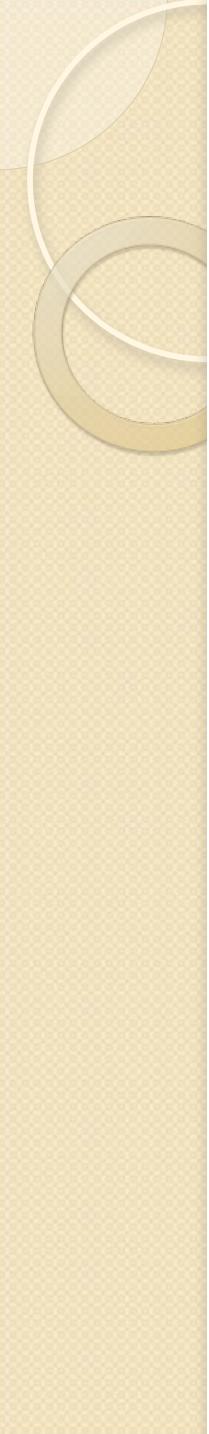
La théorie des dépendances et la normalisation

Christine Verdier



La théorie des dépendances

Les dépendances fonctionnelles et les
3 premières formes normales



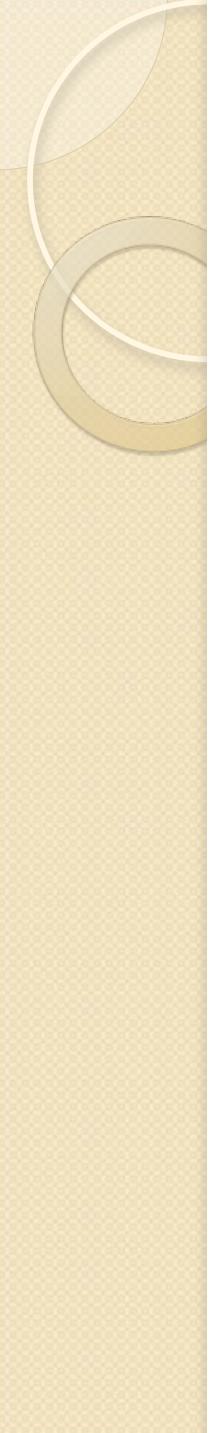
La normalisation

- Normaliser une relation consiste à la représenter sous une forme canonique qui respecte certains critères afin de s'assurer de l'intégrité des données.
- C'est un moyen fondamental mis à la disposition du concepteur de la base pour éviter bon nombre d'erreurs graves de conséquences.



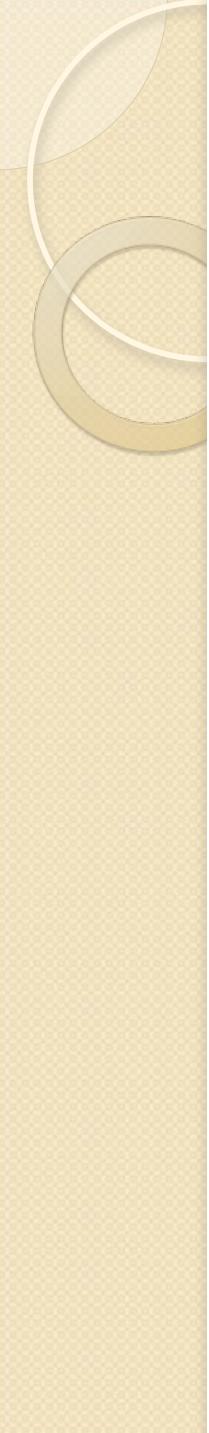
Les dépendances

- Une dépendance représente une contrainte exprimée par une règle que doivent vérifier les données pour que la base de données soit dans un état cohérent.
- Exemple : ‘ Un médecin travaille dans un et un **seul** service’



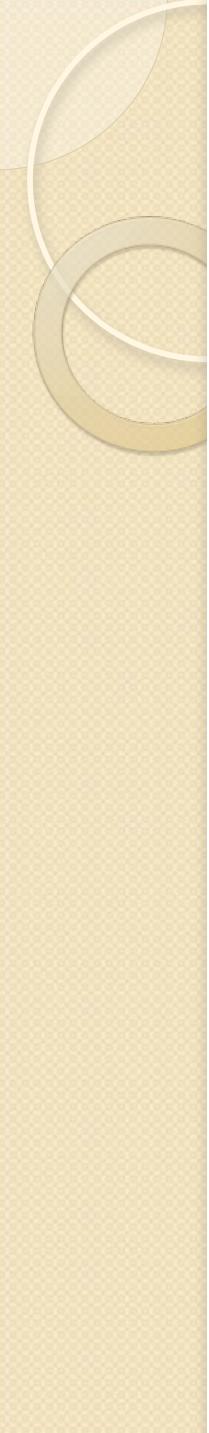
Les dépendances

Dépendances
=
Invariants
du
Système



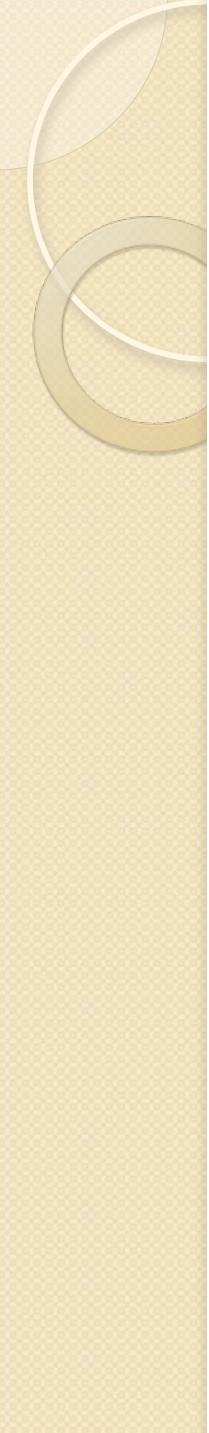
Les dépendances

- Il existe 3 sortes de dépendances qui conduisent à normaliser des relations :
 - les dépendances fonctionnelles
 - les dépendances multivaluées
 - les dépendances de jointure
- Il existe d'autres dépendances qui traduisent souvent des contraintes d'intégrité :
 - les dépendances hiérarchiques



Dépendances et normalisation

- Les 3 premières sortes de dépendances conduisent à des formes de normalisation de plus en plus élaborées.
 - Dépendances fonctionnelles : formes 1, 2, 3 et FNBC
- Dépendances multivaluées : forme 4
- Dépendances de jointure : forme 5



Dépendances fonctionnelles

- C'est de loin le type de dépendances le plus fréquemment rencontré.
- Une dépendance fonctionnelle (DF) traduit le fait qu'à une valeur d'une donnée, on associe dans une relation une valeur au plus d'une autre donnée, à un instant considéré.

Dépendances fonctionnelles

- Soit la relation R

CLIENT	ADRESSE	ARTICLE COMMANDE
DUPONT	LYON	LAQUE
DUPONT	LYON	LOTION
DUBOIS	MARSEILLE	LAQUE
JOHN	LONDRES	CREME
DUBOIS	MARSEILLE	CREME

Dépendances fonctionnelles

Il y a une DF entre Client et Adresse dans R

Client → Adresse

Un client n'a qu'une seule adresse.

En revanche,

Client ↗ Article commande

Il y a dépendance parce que lorsqu'une valeur est associée à une autre par une DF, par exemple (Dupont, Lyon), on ne peut pas introduire dans la base le couple (Dupont, Nice) sans supprimer le premier couple.

Client → Adresse

SOURCE

BUT

Dépendances fonctionnelles

- Formellement, une DF de A vers B existe si lorsque dans une relation $R(A, B, \dots)$

$A \longrightarrow B$ Est vraie dans R

Si et seulement si
 $\langle A_i, B_k \rangle \text{ et } \langle A_i, B_j \rangle \Rightarrow B_k = B_j$

Où $\langle A_i, B_j \rangle$ représente une occurrence de la projection sur (A, B) de la relation R

Propriétés des dépendances fonctionnelles

Soient U, V, W, X des sous-ensembles d'un ensemble d'attributs R. Les DF vérifient les règles suivantes :

Réflexivité :

$$V \subseteq U \text{ alors } U \rightarrow V$$

Augmentation :

$$\text{Si } V \rightarrow U \text{ et } W \subset X \text{ alors } V, X \rightarrow U, W$$

Transitivité :

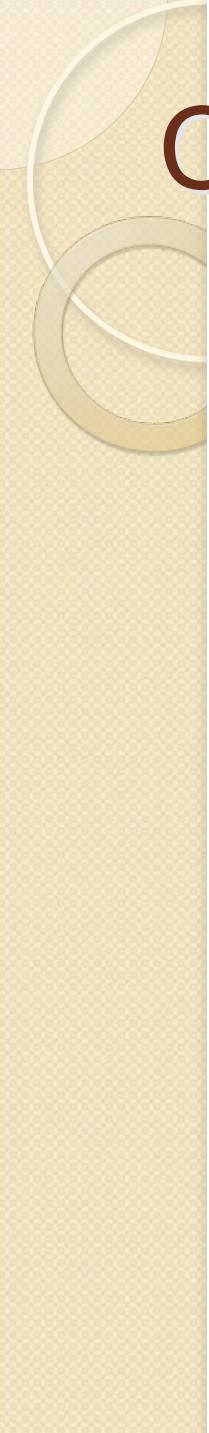
$$U \rightarrow V \text{ et } V \rightarrow W \text{ alors } U \rightarrow W$$

Pseudotransitivité :

$$\text{Si } U \rightarrow V \text{ et } V, W \rightarrow X \text{ alors } U, W \rightarrow X$$

Décomposition :

$$\text{Si } U \rightarrow V, W \text{ alors } U \rightarrow V \text{ et } U \rightarrow W$$



Clé d'une relation

On peut donner une définition plus formelle de la clé.

Lorsque, dans une relation, **un attribut** (ou un groupe d'attributs) **est source de dépendances fonctionnelles**

ayant respectivement pour but chacun des autres attributs de la relation,

cet **attribut** (ou ce groupe d'attributs) est appelé **CLE de la relation**.

DF élémentaire et DF directe

∈ DF élémentaire

On dit qu'une DF est élémentaire si un sous-ensemble de la clé n'est pas source de la DF.

Exemple :

$A, B, C \xrightarrow{} D$ Est élémentaire Ssi

$A, B \not\xrightarrow{} D$ $A, C \not\xrightarrow{} D$ $B, C \not\xrightarrow{} D$
 $A \not\xrightarrow{} D$ $B \not\xrightarrow{} D$ $C \not\xrightarrow{} D$

∈ DF directe

$A \xrightarrow{} B$

est directe s'il n'existe pas un autre attribut C tel que :

$A \xrightarrow{} C$ et $C \xrightarrow{} B$

Nécessité de la normalisation

Considérons la relation suivante :

PRODUIT	CLIENT	ADRESSE	NOMBRE PRODUIT COMMANDE	MONTANT
LOTION	MARTIN	PARIS	10	250
LAQUE	MARTIN	PARIS	250	2500
CREME	MARTIN	PARIS	20	1000
LOTION	JOHN	LONDRES	30	750
CREME	DUPONT	LYON	20	1000
LOTION	DUBOIS	PARIS	20	500

L'ensemble **PRODUIT-CLIENT** est la **seule clé** possible pour cette relation.

Nécessité de la normalisation

Une telle relation parfaitement correcte, pour un SGBD, va poser des problèmes de maintenance et risque d'engendrer des anomalies.

Cette relation n'est pas dans une « **BONNE FORME** », car on constate que cette relation est **coûteuse en place**.

- L'adresse du client est répétée autant de fois qu'il a commandé de produits.
- La mise à jour d'une adresse nécessite plusieurs accès à la base.
 - On peut avoir des anomalies de mise à jour.

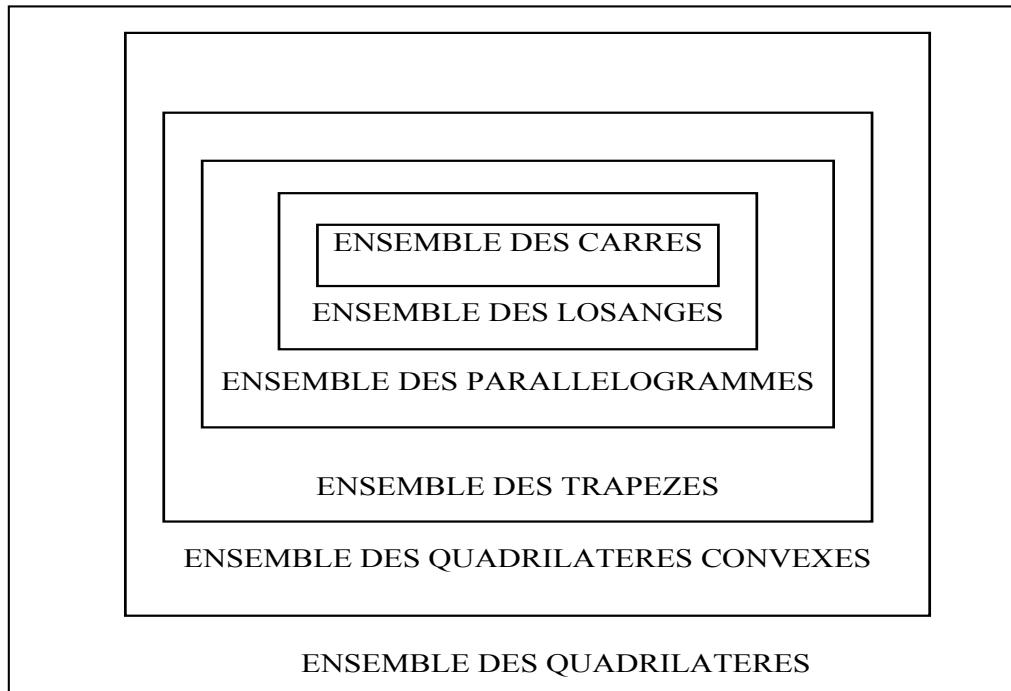
Exemple : Si on insère la ligne

<CREME,JOHN,NEW-YORK,10,500>

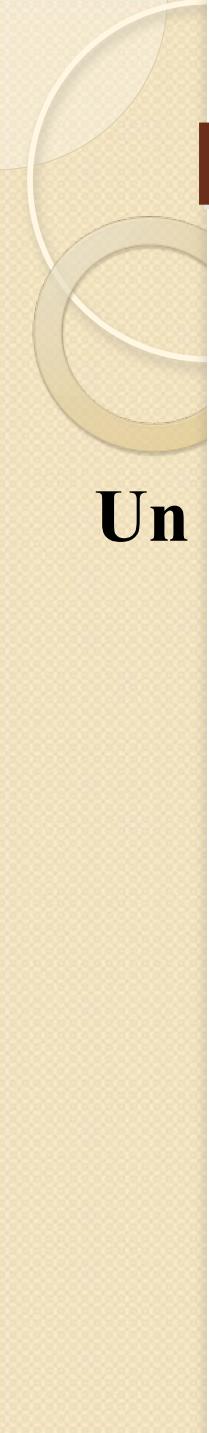
Le système ne peut pas détecter que John aura alors deux adresses, ce qui est contraire à la réalité.

- Si Dupont annule sa commande, on perd son nom et adresse.

Analogie avec la géométrie



Quadrilatère convexe	\approx	1ère forme normale
Trapèze	\approx	2ème forme normale
Parallélogramme	\approx	3ème forme normale
Losange	\approx	4ème forme normale
Carré	\approx	5ème forme normale



Relation normalisée

Une relation est normalisée si :

Un même attribut n'est pas représenté plusieurs fois

et si

Un attribut n'est pas décomposable en d'autres attributs.

(il n'est donc pas lui-même une relation).

Relation non normalisée

La relation « EMPLOYE » ci-dessous n'est pas normalisée.

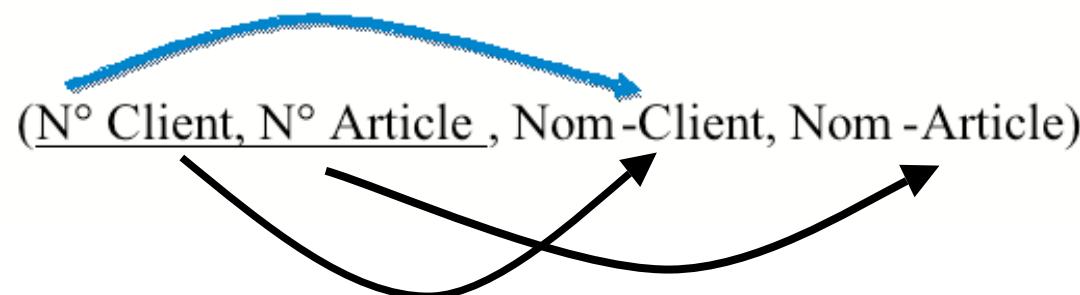
NOM EMPLOYE	ENFANTS		DIPLOMES DE L'EMPLOYE		
	Prénom	Date naissance	Nature	Examen	
				Lieu	Date
Dupont	Sophie	1971	Bac	Lyon	1967
	Luc	1973	Licence	Paris	1970
	Simon	1978			
Arnold	Jeanne	1952	Bac	Paris	1947
	Patrice	1953	Ingénieur	Paris	1952
			Doctorat	Nice	1963

1ère forme normale

Toute relation normalisée est en 1ère forme normale.

Exemple de relation en 1ère forme normale

Considérons la relation COMMANDE :



1ère forme normale

Inconvénient de la 1ère forme normale :

Seules des DF suivantes peuvent être prises en compte :

n°client, n°article —> nom client

et n°client, n°article —> nom article

On peut donc introduire sans possibilité de détection des anomalies du type suivant :

<23, A5, JEAN, CHAISE>

<23, A6, CAROLINE, TABLE>

<26, A5, CHRISTINE, TABLE>

Il convient pour mettre en évidence toutes les DF de décomposer la relation initiale en 3 relations:

CLIENT (N°Client, nom client)

ARTICLE (N°Article, nom article)

COMMANDE (N°Article,N°Client)

2ème forme normale

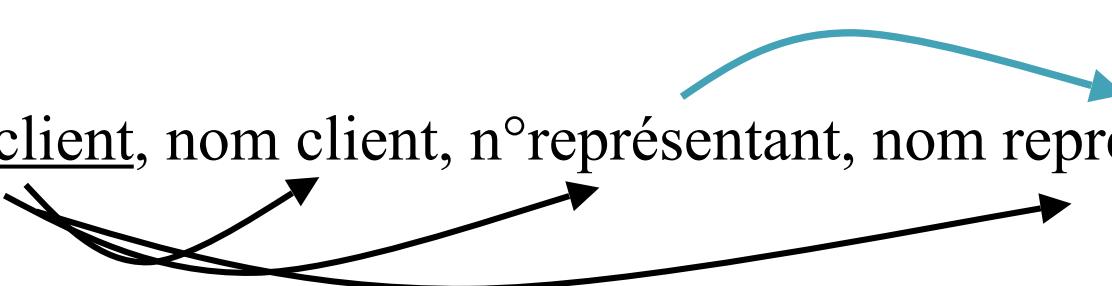
Une relation est en deuxième forme normale si :

- elle est en 1ère forme normale
- toutes les DF issues de la clé sont élémentaires.

Exemple :

Considérons la relation CLIENT

(N°client, nom client, n°représentant, nom représentant)



2ème forme normale

Cette relation présente aussi des inconvénients car la DF n°représentant → nom représentant n'est pas matérialisée.

On peut alors avoir les incohérences suivantes :

<92, MARLENE, R 15, LOULOU>

<156, PHILIPPE, R 15, BABETTE>

Il convient donc de décomposer cette relation en deux autres relations :

Client (n°client, nom client, n°représentant)

Représentant (n°représentant, nom représentant)

3ème forme normale

Une relation est en troisième forme normale si :

- * elle est en 2ème forme normale
- * toutes les DF issues de la clé sont directes

R1 (N°client, nom-client, adresse-client)

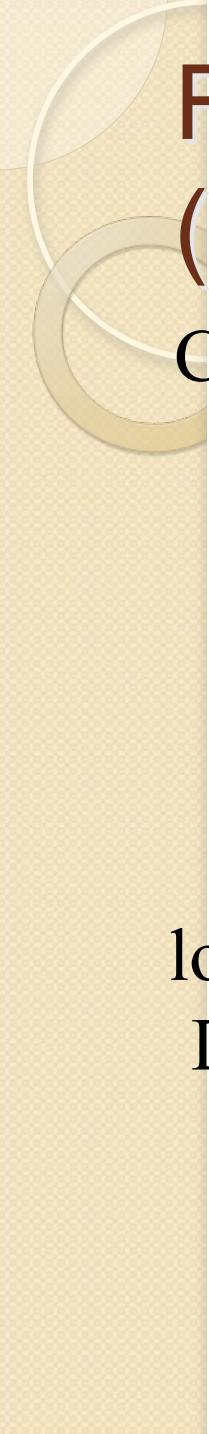
Ou encore

R2 (N°commande, n°article, qté-commandée, qté-livrée)

3ème forme normale

Lorsqu'une relation est en 3 FN

- a . Aucune DF n'est issue d'un sous-ensemble de clé
- b . Aucune DF n'est issue d'un attribut non clé vers un autre attribut non clé.

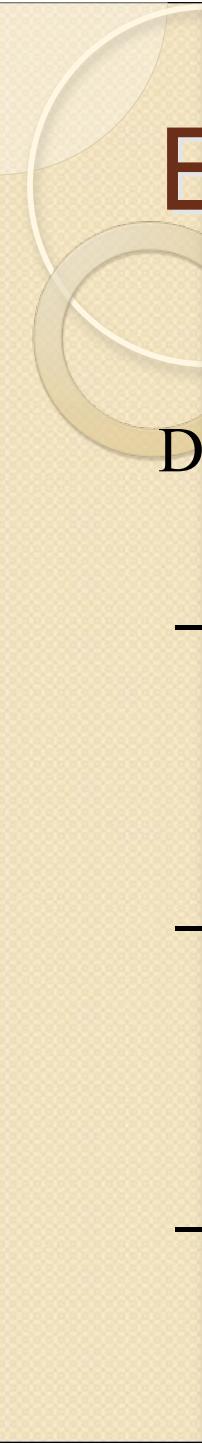


Forme normale de Boyce Codd (FNBC)

C'est une forme spécifique de la 3ème forme normale

Une relation en 3ème forme normale peut encore présenter certaines anomalies,
en particulier,

lorsqu'un attribut non clé de la relation est source d'une DF ayant pour but une partie (un sous-ensemble) de la clé.



Exemple

Dans un lycée, on a relevé que :

- Pour une matière donnée et un élève donné, on a une note annuelle.
- Pour une matière donnée et un élève donné, on a un professeur.
- Un professeur enseigne une seule matière.

Exemple

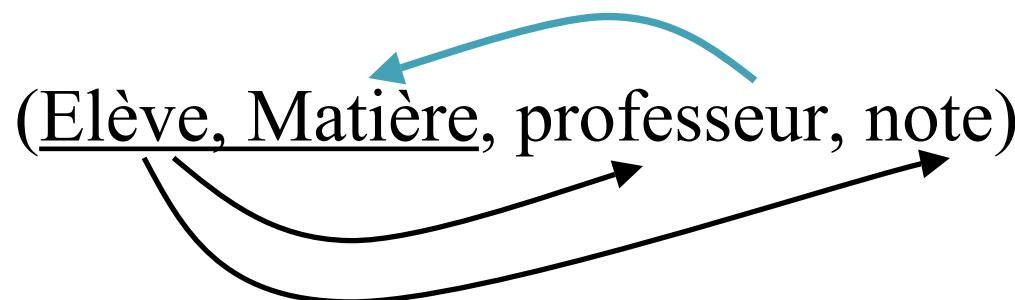
Ce qui se traduit par :

ELEVE, MATIERE --> PROFESSEUR

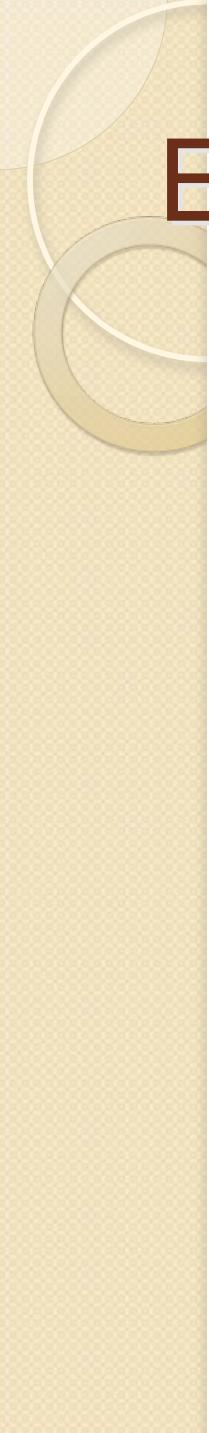
ELEVE, MATIERE --> NOTE

PROFESSEUR --> MATIERE

La relation RESULTAT



Est en 3ème forme normale mais pas en F.N.B.C



Exemple

INCONVENIENT

On ne peut pas traduire le fait qu'un professeur est associé à une unique matière.

Il est donc possible d'avoir dans la base :
<TOTO, MATH, DUPONT>

et

<JOJO, ANGLAIS, DUPONT>

Alors qu'une contrainte est violée

Exemple

La relation résultat est décomposable.

En effet :

Professeur--> matière

et **élève, matière --> note**

par pseudotransitivité :

élève, professeur --> note

On peut donc avoir :

Résultat (élève, professeur, note)

Enseignement (professeur, matière)

Cependant les DF :

Elève, matière --> professeur

Elève, matière --> note

ne sont pas représentées et ceci peut conduire à des anomalies.

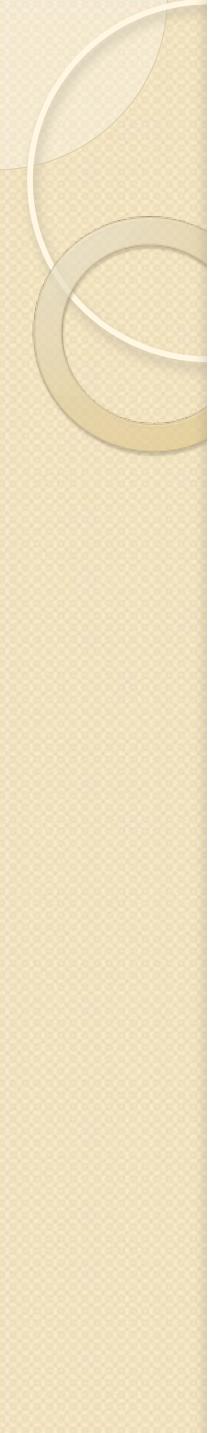
Csq : on a décomposé la relation initiale mais on n'a pu mettre en évidence toutes les contraintes.

Forme normale de Boyce Codd

Une relation est en F.N.B.C

si

La source de chaque DF existant entre les attributs de la relation est clé de la relation.



Forme normale de Boyce Codd

La FNBC représente la **forme la plus élaborée** fondée sur les **dépendances fonctionnelles**.

Conséquences méthodologiques

**Chaque DF doit être traduite
par
une relation.**

Dans un SGBD relationnel

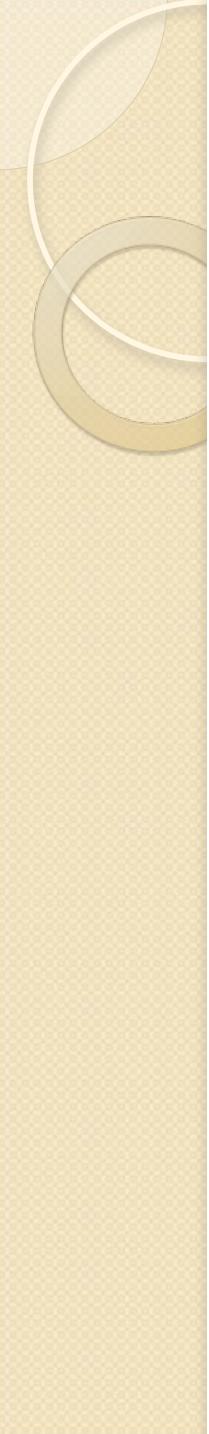
l'unique façon de prendre en compte une **DF est de déclarer sa source** comme clé d'une relation (ce qui interdit la duplication des valeurs identiques et assure ainsi la cohérence des valeurs).

(Avec SQL, on utilise soit la fonction « **PRIMARY KEY** » soit à défaut la fonction « **UNIQUE** »)

La normalisation en F.N.B.C. permet d'atteindre ce but

Conséquences méthodologiques

Lorsqu'une relation contient
AU MOINS
une dépendance fonctionnelle,
elle ne peut être
qu'en
1FN, 2FN, 3FN ou FNBC



Exercices

- Trouver les formes normales des relations suivantes et proposer une décomposition si nécessaire :
 - R1 (n°client, n°produit, qt cdée, nom produit)
 - R2 (n°cde, n° produit, qt cdée)
 - R3 (n°client, nom client, nom représentant)
 - R4 (n°produit, nom produit, n°atelier, nom chef atelier)
 - R5 (n°produit, nom produit, n°atelier, nom chef atelier)
 - R6 (n°client, nom client, n°repr, nom repr)
 - R7 (n°produit, n°fourn, prix, nom fourn)



La théorie des dépendances

Les dépendances multivaluées et les dépendances de jointure

4 FN & 5 FN



Dépendances multivaluées

- Cette forme de dépendance est plus complexe que les dépendances fonctionnelles.
- Pour qu'il y ait DM (non triviale), il faut :
 - Que l'on ait au moins 3 attributs
 - Que l'on associe à un attribut donné, plusieurs valeurs des autres attributs.

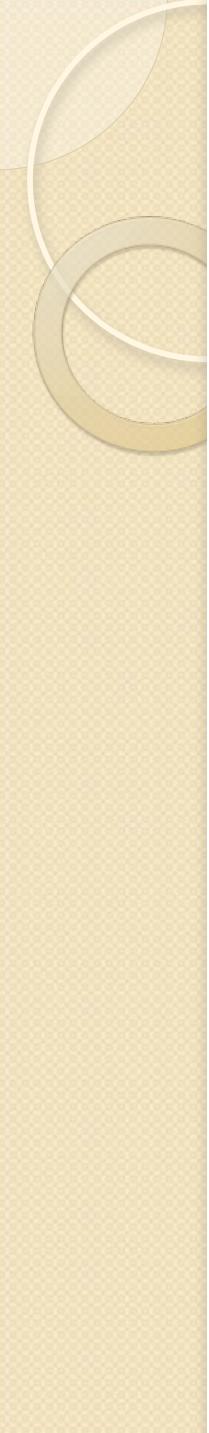
Dépendances multivaluées

- Considérons la table R suivante :

ETUDIANT	SPORT	MATIERE
Loulou	Tennis	Maths
Loulou	Ski	Physique
Loulou	Ski	Maths
Loulou	Tennis	Physique
Babette	Danse	Informatique
Babette	Danse	Maths

Les dépendances : {étudiant, sport} et {étudiant, matière} sont multivaluées

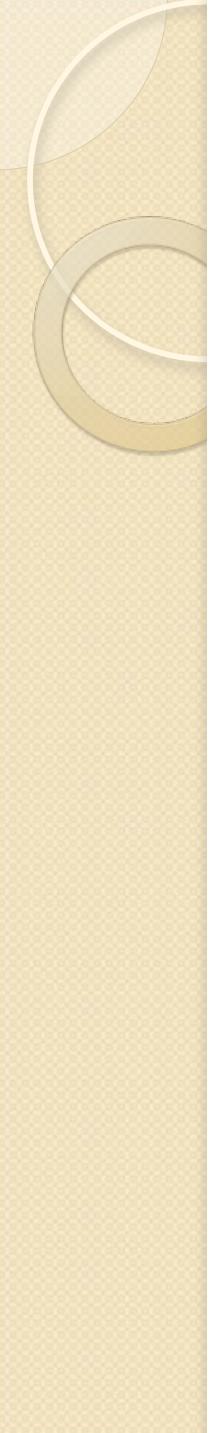
ETUDIANT $\Rightarrow\Rightarrow$ SPORT/MATIERE



Dépendances multivaluées

- Il existe une dépendance multivaluée lorsque :
 - A tout étudiant E qui pratique un sport S et qui étudie une matière M, le triplet $\langle E, S, M \rangle$ est présent dans la relation considérée.
- Formellement, la DM s'exprime de la manière suivante :

Si $A \Rightarrow\!\Rightarrow B/C$
alors $[A_i, B_j] \text{ et } [A_i, C_k] \Rightarrow [A_i, B_j, C_k]$



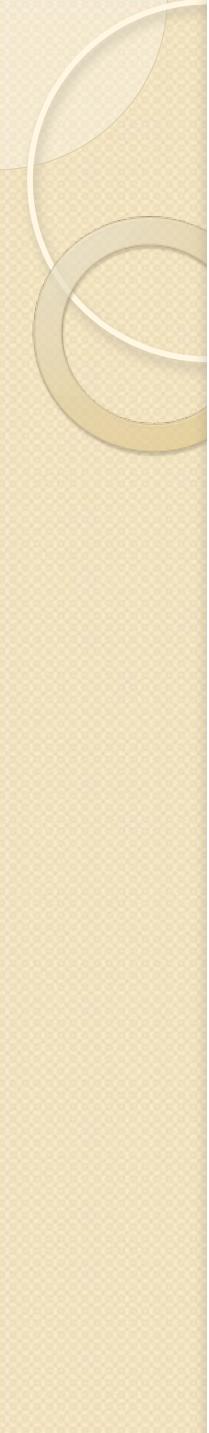
Dépendances multivaluées

- Dans la relation R, on constate que :
 - Loulou fait du ski
 - Loulou étudie des maths
- Donc, le triplet <Loulou, Ski, Maths> est présent dans la relation.



Dépendances multivaluées

- Existence d'une dépendance multivaluée car :
- Si on a :
 - <Babette, Danse, Maths> et <Babette, Escrime, Anglais>
 - On aura forcément :
 - <Babette, Danse, Anglais> et <Babette, Escrime, Maths>



Règles mathématiques

On suppose $X \cup Y \cup Z =$ ensemble des attributs de la relation

1. Complémentation

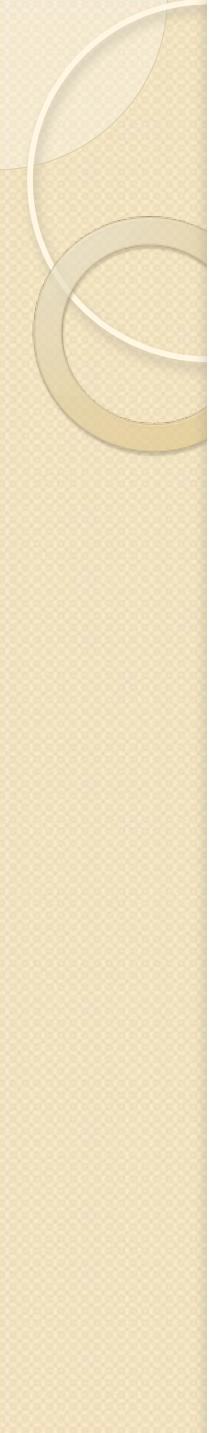
Si $X \Rightarrow\Rightarrow Y$ alors $X \Rightarrow\Rightarrow Z$

2. Réflexivité

Si $Y \subseteq X$ alors $X \Rightarrow\Rightarrow Y$

3. Augmentation

Si $V \subseteq W$ et $X \Rightarrow\Rightarrow Y$ alors $X \cup W \Rightarrow\Rightarrow Y \cup V$



Règles mathématiques

4. Transitivité

Si $X \Rightarrow\Rightarrow Y$ et $Y \Rightarrow\Rightarrow Z$ alors $X \Rightarrow\Rightarrow Z$

5. Pseudo transitivité

Si $X \Rightarrow\Rightarrow Y$ et $Y, W \Rightarrow\Rightarrow Z$ alors $X, W \Rightarrow\Rightarrow Z$

6. Union

Si $X \Rightarrow\Rightarrow Y$ et $X \Rightarrow\Rightarrow Z$ alors $X \Rightarrow\Rightarrow Y \cup Z$

7. Décomposition

Si $X \Rightarrow\Rightarrow Y$ et $X \Rightarrow\Rightarrow Z$ alors $X \Rightarrow\Rightarrow Y \cap Z$



4^{ème} forme normale

- La relation :

R(Etudiant, sport, matière) est en FNBC.

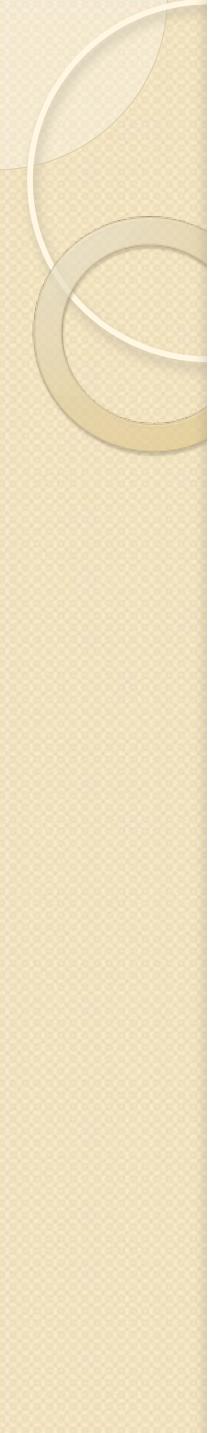
Pourtant, elle contient toujours beaucoup de redondances toujours sources d'erreurs.

Pour éliminer ces erreurs, on a fourni une forme de normalisation plus élaborée : la 4^{ème} forme normale



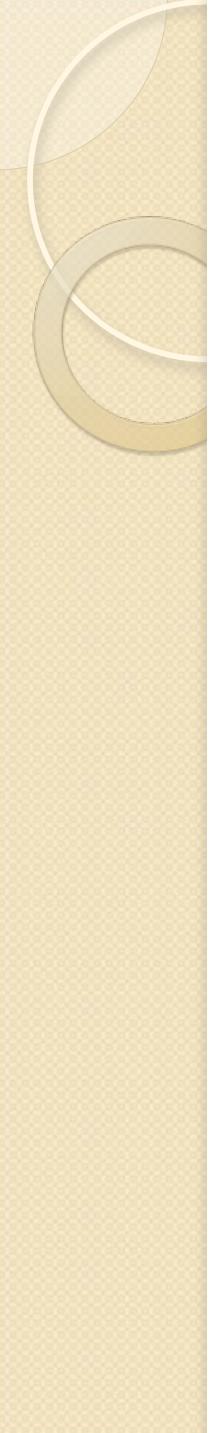
4ème forme normale

- Une relation est en 4ème FN si :
 - Elle est en FNBC
 - Elle ne possède pas de DM non triviale
- *NB : La 4 FN ne concerne que des relations réduites à leur clé*



4^{ème} forme normale

- R(Etudiant, sport, matière) n'est pas en 4^{ème} forme normale car :
Etudiant $\Rightarrow\Rightarrow$ sport/matière
- Il convient donc de la casser en deux relations :
 $R1(\underline{\text{étudiant}}, \underline{\text{sport}})$ et
 $R2(\underline{\text{étudiant}}, \underline{\text{matière}})$
- R1 et R2 sont alors en 4^{ème} forme normale



Conséquence

- On peut avoir une DM lors de 2 situations possibles :
 - Association d'entités n'ayant pas de sémantique commune entre elles → R(étudiant, matière, prof)
 - Existence de faits multivalués indépendants → R(étudiant, sport, matière)
- La plupart du temps, il s'agit d'erreurs de conception

Décomposition des relations

- L'opération de décomposition-jointure notée \bowtie est vérifiée pour :

$$R(A,B,C) = R_1(A,B) \bowtie R_2(B,C)$$

si

- a) R_1 et R_2 sont des projections de R
- b) La jointure de R_1 et R_2 est égale à R

- Par généralisation, on dira que :

$$R(A,B,C) = R_1(A,B) \bowtie R_2(B,C) \bowtie R_3(A,C)$$

Décomposition des relations

- Considérons la table R :

Nom Etudiant	Faculté	Sport	Prof Sport
Loulou	Droit	Ski	José
Fifi	Gestion	Danse	Sylvie
Riri	Lettres	Ski	Sam
Babette	Sciences	Natation	Sylvie
Toto	Droit	Danse	Sylvie
Jojo	Gestion	Ski	José

Décomposition des relations

A

Nom Etudiant	Faculté	Sport
Loulou	Droit	Ski
Fifi	Gestion	Danse
Riri	Lettres	Ski
Babette	Scienc	Natation
Toto	Droit	Danse
Jojo	Gestion	Ski

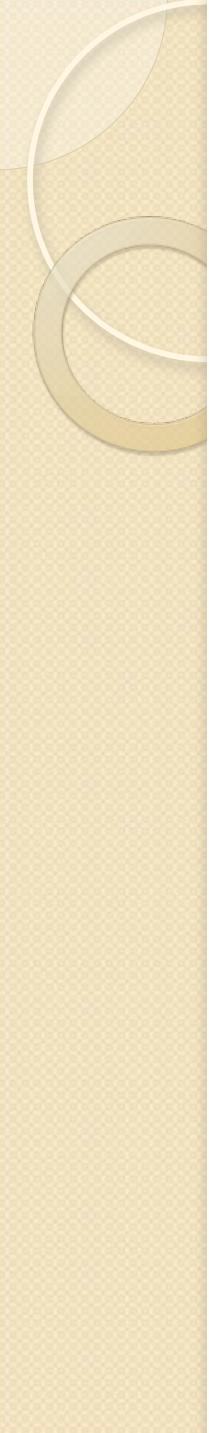
B

Sport	Prof sport
Ski	José
Danse	Sylvie
Ski	Sam
Natation	Sylvie

Nom Etudiant	Faculté	Sport	Prof Sport
Loulou	Droit	Ski	José
Loulou	Droit	Ski	Sam
Fifi	Gestion	Danse	Sylvie
Riri	Lettres	Ski	José
Riri	Lettres	Ski	Sam
....

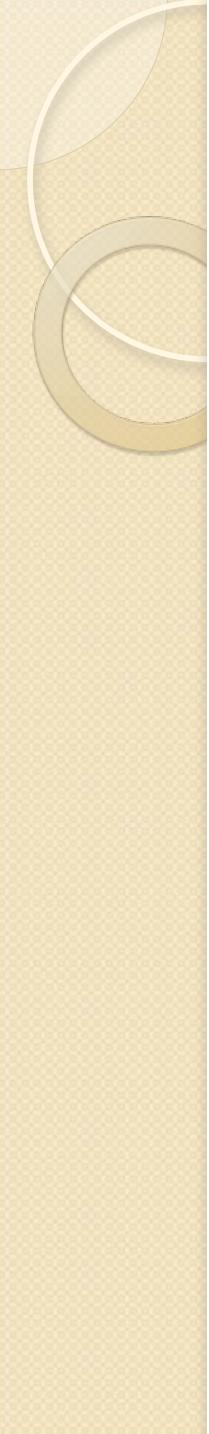
Jointure S=A \bowtie B

Jointure avec perte



Décomposition des relations

- On remarque que $S \bowtie R$
- On dit que la jointure est avec perte (d'informations)
- En effet, $A \bowtie B$ conduit à introduire dans la base des informations erronées.
- Dans S , on peut constater qu'il existe l'occurrence <Loulou, droit, ski, Sam> or cette occurrence n'existe pas dans R .



Décomposition des relations

- Lorsqu'il est possible d'effectuer une opération de décomposition-jointure sur une relation, on dit que cette relation est **DECOMPOSABLE**.

Dépendances de jointure

- Soit la relation Antibiothérapie suivante :

Maladie	Germe	Antibiotique
M1	G1	A2
M2	G2	A2
M3	G2	A2
M2	G2	A4
M3	G1	A3
M3	G1	A2

- On constate qu'il n'existe pas de DM

Dépendances de jointure

- La relation Antibiothérapie contient beaucoup de redondances (G1 est associé plusieurs fois à M1, A2 est associé plusieurs fois à G1, etc.)
- Il existe une dépendance de jointure

$$R(M, G, A) = R_1(M, G) \bowtie R_2(G, A)$$

$$R_3(M, A)$$

R(M,G,A)

Dépendances de jointure

- Projetons Antibiothérapie sur les 3 ta

Maladie	Germe	Antibiotique
M1	G1	A2
M2	G2	A2
M3	G2	A2
M2	G2	A4
M3	G1	A3
M3	G1	A2

R1(M,G)

Maladie	Germe
M1	G1
M2	G2
M3	G2
M3	G1

R2(G,A)

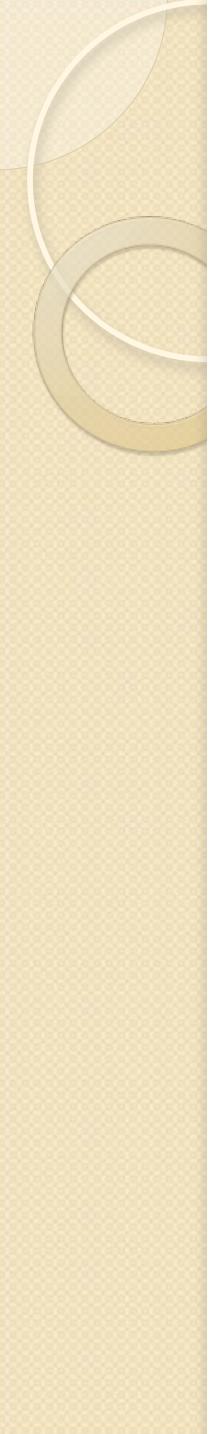
Maladie	Germe
G1	A2
G2	A2
G2	A4
G1	A3

R3(M,A)

Maladie	Antibiotique
M1	A2
M2	A2
M3	A2
M2	A4
M3	A3

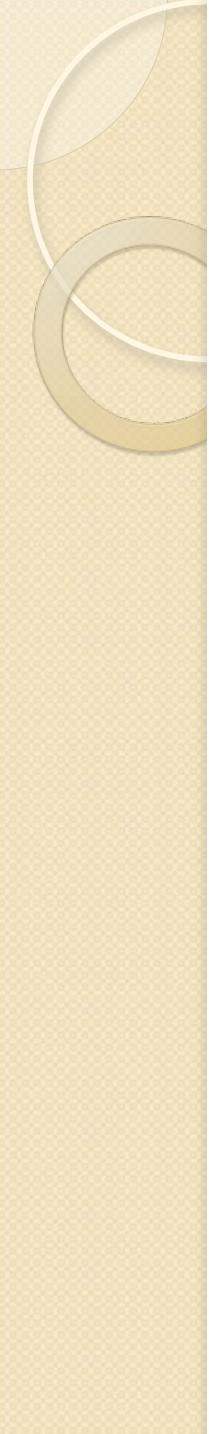
On réalise la jointure sur les 2 premières tables

Maladie	Germe	Antibiotique
M1	G1	A2
M1	G1	A3
M2	G2	A2
M2	G2	A4
M3	G2	A2
M3	G2	A4
M3	G1	A2
M3	G1	A3



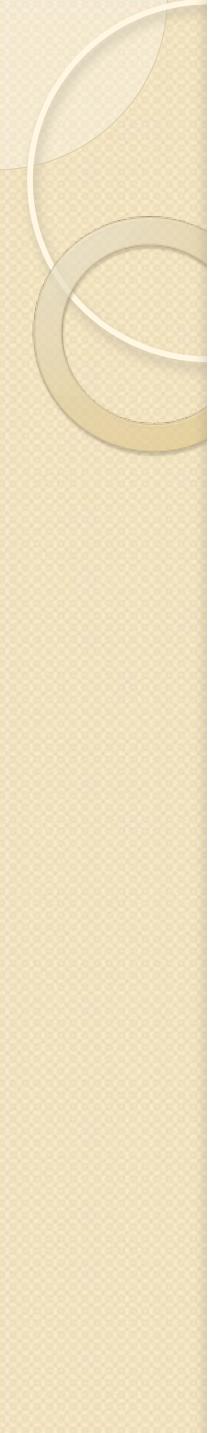
Dépendances de jointure

- Si ensuite, on réalise la jointure entre cette table-résultat et R3, on éliminera les occurrences fausses (en rouge) et on obtiendra la relation initiale.



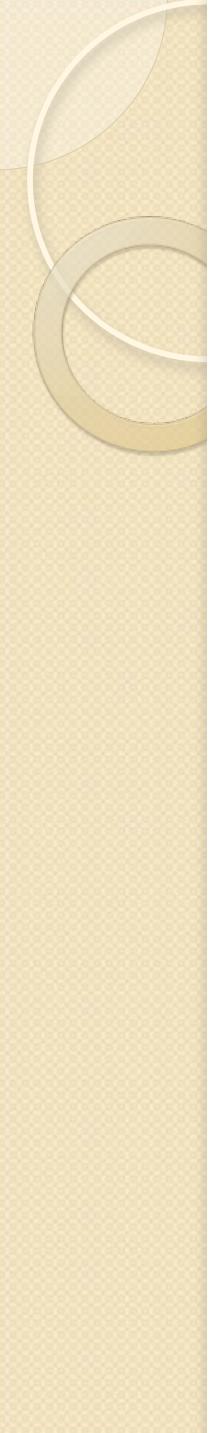
5ème forme normale

- Si un germe G est présent dans la maladie M, si un antibiotique A agit sur le germe G et si l'antibiotique A agit sur la maladie M, alors le triplet $\langle M, G, A \rangle$ est présent dans la relation Antibiothérapie.



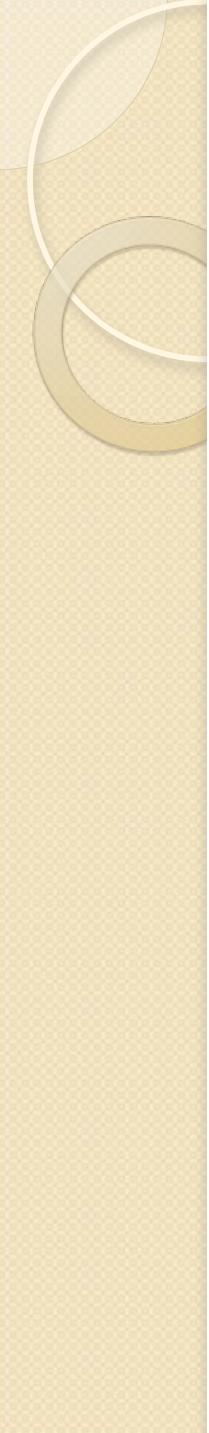
5^{ème} forme normale

- On constate que la relation Antibiothérapie est bien en 4^{ème} forme normale puisqu'il n'existe pas de DM.
- Pourtant, il existe encore beaucoup de redondances.
- La 5^{ème} forme normale est une forme qui permet de prendre en compte les dépendances de jointure.



5ème forme normale

- Une relation est en 5ème forme normale lorsqu'il n'existe pas de dépendance de jointure.
- La relation $R(M,G,A)$ n'est pas en 5ème forme normale.
- Elle doit être décomposée en
 - $R1(M,G)$
 - $R2(G,A)$
 - $R3(M,A)$



5ème forme normale

- Une relation est en 5ème forme normale si elle ne possède pas de dépendances de jointure non triviales.

Récapitulatif

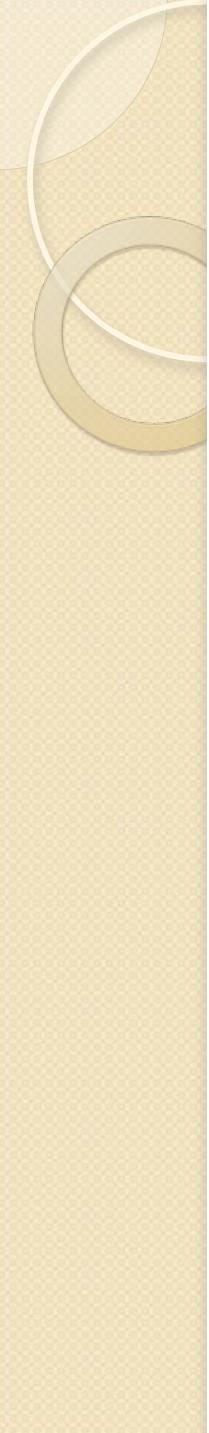
- Soit $R(\underline{C_1}, \underline{C_2}, \underline{C_3})$
- 1er cas : Cas de la 4ème FN
 - A) Il existe une DM entre C_1 et C_2/C_3 alors $C_1 \Rightarrow\Rightarrow C_2/C_3$ donc on peut décomposer la relation R :

$$R(\underline{C_1}, \underline{C_2}, \underline{C_3}) = R_1(\underline{C_1}, \underline{C_2}) \bowtie R_2(\underline{C_1}, \underline{C_3})$$

Récapitulatif

- B) Il n'existe pas de DM
Alors $R(\underline{C_1}, \underline{C_2}, \underline{C_3})$ est non décomposable et
 R est en 4ème FN.
- 2ème cas : Cas de la 5ème FN
 - A) Il existe une DJ entre C_1 , C_2 et C_3 alors

$$R(\underline{C_1}, \underline{C_2}, \underline{C_3}) = R_1(\underline{C_1}, \underline{C_2}) \times R_2(\underline{C_1}, \underline{C_3}) R_3(\underline{\cancel{C_2}} \times \underline{C_3})$$



Récapitatif

- B) Il n'existe pas de DJ alors R (C1,C2,C3) est non décomposable et R est en 5ème FN

Algorithme de recherche d'une forme normale

