Université Ferhat ABBAS Sétif 1

Module: BDD (2 L / 2018-2019)

CHAPITRE 6:

THEORIE DE LA NORMALISATION

Preparé par Dr. Salim BOUAMAMA

Département d'Informatique

THEORIE DE LA NORMALISATION

Ce chapitre section justifie la nécessité d'une étape d'affinement des schémas relationnels et introduit les approches possibles pour réduire les problèmes soulevés par une mauvaise perception du réel.

raffinement de schéma $\Rightarrow$ normalisation

LA PHASE DE DESIGN D'UNE BD

- Analyse des besoins
- Design conceptuel
  - Modèle EA, UML
- Design logique
  - EA vers relations
  - raffinement de schéma: normalisation
- Design physique
  - indexes, etc.

EXEMPLE DE MAUVAISE CONCEPTION

- Considérez la relation suivante:

POSSEDE (NV, Marque, Type, Puiss, couleur, NSS, Nom, prenom, adresse, date, prix)

NV	Marque	Type	Puiss.	Coul.	NSS	Nom	Prénom	Date	Prix
672RH75	Renault	RME8	8	Rouge	142032	Martin	Jacques	10021998	70 000
800AB64	Peugeot	P206A	7	Bleue	142032	Martin	Jacques	11061999	90 000
686HK75	Citroën	BX20V	9	Verte	158037	Dupond	Pierre	200499	120 000
720CD63	Citroën	2CV8	2	Verte	158037	Dupond	Pierre	200278	5 000
400XY75	Renault	RCL5	4	Verte	275045	Fantas	Julie	110996	20 000
–	–	–	–	–	280037	Schiffer	Claudia	–	–
963TX63	Renault	P306B	7	Bleue	–	–	–	–	–

- Problèmes de conception:
  - Redondances de données
  - Anomalies de MAJ (modification, insertion, suppression)
  - Problemes des vauleurs nulles.

LES PROBLEMES DE LA REDONDA*NCE*

- La *redondance* est à la base de beaucoup de problèmes associées avec les schémas relationnels:
  - Stockage redondant, anomalies d'insertion, de suppression et de modification.
- Les contraintes d'intégrité, en particulier les *dépendances fonctionnelles*, peuvent être utilisées pour identifier les problèmes et suggérer des solutions.
- **Solution :** *décomposition* (remplacement d'une relation par des relations plus petites).
- La décomposition devrait être utilisée judicieusement:
  - Y-a-t-il une raison valable pour décomposer?
  - Quels problèmes sont causes par la décomposition?

PRINCIPES

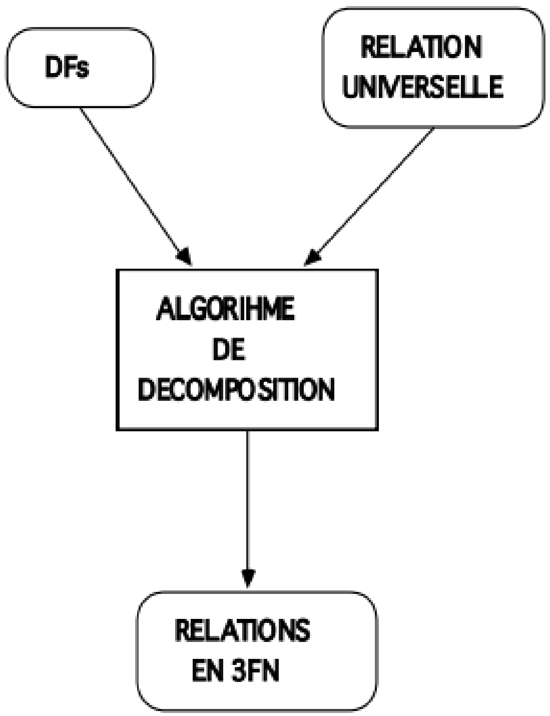
Problèmes de conception: redondances et anomalies de m.à.j. !

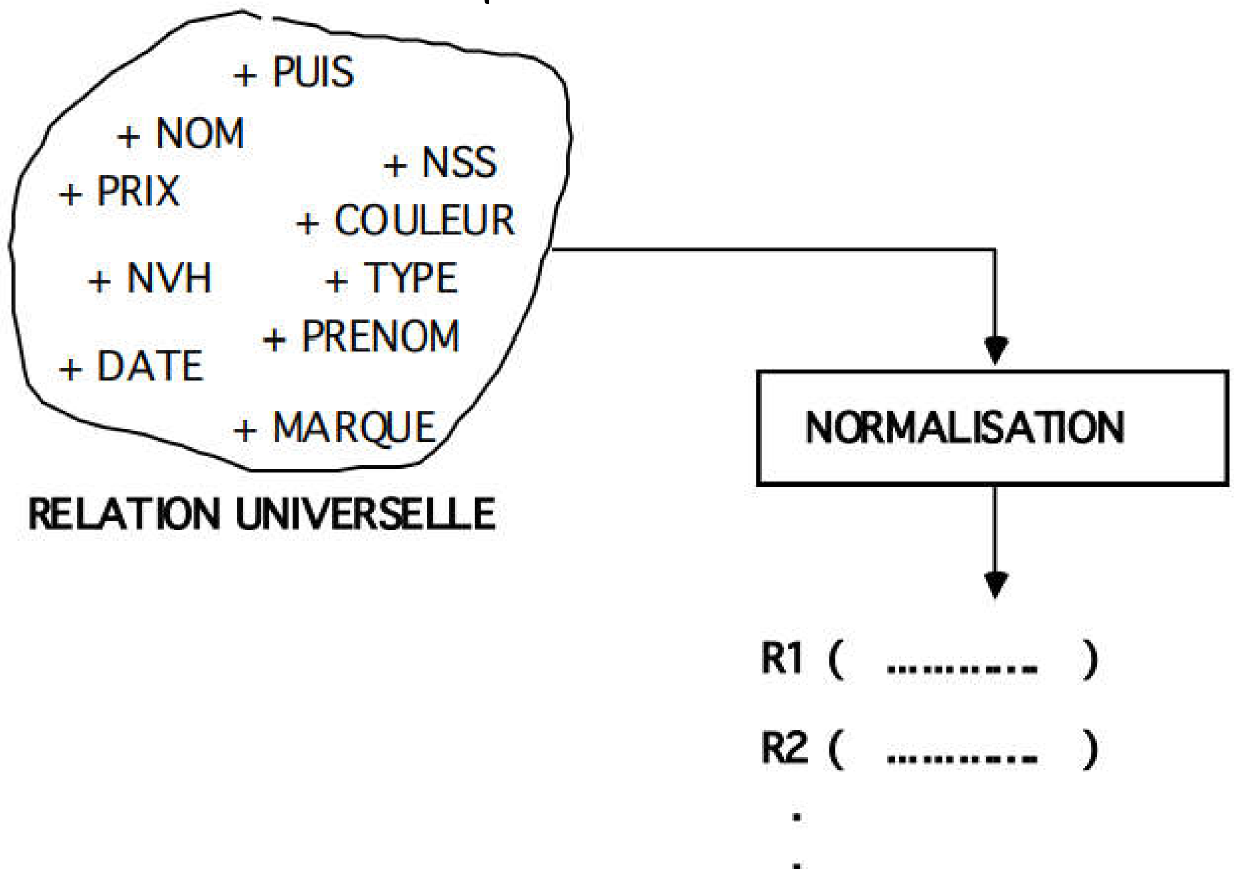
Objectif : Minimiser les redondances pour améliorer les performance de la BD.

- Requêtes d'interrogation et de mise à jour plus rapides
- Moins de valeurs NULL
- Gestion de l'intégrité des données simplifiée

Relation universelle

- Sous-ensemble du produit cartésien de la totalité des attributs de la base.
- Décomposition de cette relation universelle pour obtenir des relations normalisées-





DEPENDANCES FONCTIONNELLES (DFs)

- **Notation:**
  - R : schéma relationnel; r : instance de R
  - X, Y: sous ensemble d'attributs de R (généralement, deux sous-ensembles disjoints)
- **Définition**

  Il existe une **dépendance fonctionnelle** entre X et Y (**Notation** : $X \rightarrow Y$) sur une relation R si, pour chaque instance permise *r* de R, le fait suivant est valide:

  $\forall\ t1 \in r, \forall\ t2 \in r : \Pi_X (t1) = \Pi_x (t2) \Rightarrow \Pi_y (t1) = \Pi_y (t2)$

  - i.e., étant donne deux tuples dans *r*, si les valeurs X sont les mêmes, alors les valeurs Y doivent aussi être les mêmes.
- **Lectures possible pour $X \rightarrow Y$:**
  - X détermine Y
  - Y dépend fonctionnellement de X

DEPENDANCES FONCTIONNELLES

- Une dépendance fonctionnelle est une assertion sur toutes les valeurs possibles et non sur les valeurs actuelles: elle caractérise une intention et non une extension de la relation.
- Les DF font partie du schéma d'une BD, en conséquence, elles doivent être déclarées par les administrateurs de la BD et être contrôlées par le SGBD.
- K est une candidate clé pour une relation R si K $\rightarrow$R
  - Cependant K $\rightarrow$ R ne requiert pas que K soit minimal!

Exemple de Dépendance Fonctionnelle

Considérez la relation suivante:

VOITURE (NVH, Type, Marque, Puis, Couleur)

- Un type de voiture peut avoir une seule puissance et une seule marque, mais différentes couleurs, numéros.

VOITURE	NVH	TYPE	MARQUE	PUIS	COULEUR
	872RH7	R21	RENAULT	7	BLEUE
	975BE5	R21	RENAULT	7	BEIGE
	975AB8	205	PEUGEOT	8	ROUGE

  - Parfois, nous nous référerons à tous les attributs d'une relation en utilisant le nom de la relation (p.ex. VOITURE au lieu de **{NVH, Marque, Type, Puis, Couleur}**)

Quelques DFs sur VOITURE

- NVH →Marque, Type, Puis, Couleur
- Type → Marque, Puis
- Type,Marque → Puissance
- Marque → ? Puis

GRAPHE DES DEPENDANCES FONCTIONNELLES

- **Dépendance fonctionnelle élémentaire**
  - $X \longrightarrow A$ tel que:
    a) $A \not\subset X$
    b) il n'existe pas de X' / $X' \subset X$ et $X' \longrightarrow A$
  - **Exemple:**
  - NVH, Puis $\longrightarrow$ Type n'est pas une DF élémentaire. Pourquoi?
- **Graphe des dépendances fonctionnelles**

  Soit un ensemble F de dépendances fonctionnelles élémentaires.
  Il est possible de visualiser cet ensemble de dépendances par un graphe orienté appelé graphe des dépendances fonctionnelles.

  Un nœud = un attribut   un arc = une DF
- **Exemple:**

F={NVH $\rightarrow$ Type, NVH $\rightarrow$ Couleur, Type $\rightarrow$ Puis, Type $\rightarrow$ Marque}

LES AXIOMES D'ARMSTRONG

Les DFs obéissent à des propriétés mathématiques particulières, dites axiomes d'Armstrong.

a) **Réflexivité** : Tout groupe d'attributs se détermine lui même et détermine chacun de ses attributs (ou sous groupe de ses attributs).

Soient X et Y des attributs : $X \supseteq Y \Rightarrow X \rightarrow Y$

$(XY \rightarrow XY$ et $XY \rightarrow X$ et $XY \rightarrow Y)$

b) **Augmentation**

Si un attribut X détermine un attribut Y, alors tout groupe composé de X enrichi avec d'autres attributs détermine un groupe composé de Y et enrichi des mêmes autres attributs.

Soient X, Y et Z des attributs : $X \rightarrow Y \Rightarrow XZ \rightarrow YZ$

c) **Transitivité**

Si un attribut X détermine un attribut Y et que cet attribut Y détermine un autre attribut Z, alors X détermine Z.

Soient X, Y et Z des attributs : $X \rightarrow Y$ et $Y \rightarrow Z \Rightarrow X \rightarrow Z$

AUTRES PROPRIÉTÉS DÉDUITES DES AXIOMES D'ARMSTRONG

Soient, W, X, Y et Z des attributs :

d) Pseudo-transitivité : $X \longrightarrow Y$ et $WY \longrightarrow Z \Rightarrow WX \longrightarrow Z$

Cette propriété est déduite de l'augmentation et de la réflexivité :

$X \longrightarrow Y$ et $WY \longrightarrow Z \Rightarrow WX \longrightarrow WY$ et $WY \longrightarrow Z \Rightarrow WX \longrightarrow Z$

e) Union : $X \longrightarrow Y$ et $X \longrightarrow Z \Rightarrow X \longrightarrow YZ$

Cette propriété est déduite de la réflexivité, de l'augmentation et de la transitivité :

$X \longrightarrow Y$ et $X \longrightarrow Z \Rightarrow X \longrightarrow XX$ et $XX \longrightarrow XY$ et $YX \longrightarrow YZ \Rightarrow X \longrightarrow YZ$

f) Décomposition: $X \longrightarrow YZ \Rightarrow X \longrightarrow Z$ et $X \longrightarrow Y$

Cette propriété est déduite de la réflexivité et de la transitivité :

$X \longrightarrow YZ \Rightarrow X \longrightarrow YZ$ et $YZ \longrightarrow Z \Rightarrow X \longrightarrow Z$

Fermeture d'un ensemble d'attributs

- Soit R(U) un schéma de relation et F un ensemble de dépendances fonctionnelles sur U., et X un sous-ensemble des attributs de X $\subseteq$U:

La fermeture d 'un ensemble d 'attributs X $\subseteq$Z par rapport à F, notée $[X]_F^+$, est l 'ensemble des attributs A déterminés par X.

$$[X]_F^+ = \{ A \in Z \text{ tels que } F \models X \rightarrow A \}$$

- **Exemple:**

Soit F = {A $\longrightarrow$ B,B $\longrightarrow$ C, CD $\longrightarrow$ E}. Calculer $[A]_F^+$?

$[A]_F^+$ = { A , B, C}.

- Deux ensembles de DFs F et G sont dits **équivalents** lorsqu'ils possèdent la même fermeture.

Algorithme de calcul de $[X]_F^+$

- Soit F un ensemble de dépendances et $X \subseteq U$ un ensemble d'attributs.
- L'algorithme calcule une suite d'ensembles d'attributs $[X]_F^0$, $[X]_F^1$, . . .

Algorithme

Données : X, U, F.

1. $[X]_F^0 \leftarrow X$ // Initialisation
2. $[X]_F^{i+1} \leftarrow [X]_F^i \cup \{A \mid \exists Z : (Y \rightarrow Z) \in F, A \in Z \text{ et } Y \subseteq [X]_F^i\}$.
3. Si ($[X]_F^{i+1} = [X]_F^i$) alors l'algorithme s'arrête.

- **Note :**

L'algorithme s'arrête toujours puisque $[X]_F^0 \subseteq [X]_F^1 \subseteq [X]_F^2 \cdots \subseteq U$ et U est un ensemble fini.

Algorithme de calcul de $[X]_F^+$

Algorithme

Données : X, U, F.

1. $[X]_F^0 \leftarrow X$ // Initialisation
2. $[X]_F^{i+1} \leftarrow [X]_F^i \cup \{A \mid \exists Z : (Y \rightarrow Z) \in F, A \in Z \text{ et } Y \subseteq [X]_F^i\}$.
3. Si ($[X]_F^{i+1} = [X]_F^i$) alors l'algorithme s'arrête.

❑ **Exemple:**

Soit $F = \{A \rightarrow B, B \rightarrow C, CD \rightarrow E\}$. Calculer $[A]_F^+$?

$[A]_F^0 = \{A\}$.

$[A]_F^1 = \{A, B\}$ en appliquant $A \rightarrow B$.

$[A]_F^2 = \{A, B, C\}$ en appliquant $B \rightarrow C$.

$[A]_F^3 = [A]_F^2 = \{A, B, C\}$, il ne reste pas des DFs a appliquer.

$\Rightarrow$ Arrêt d'algorithme et donc $[A]_F^+ = \{A, B, C\}$.

Utilisation de la fermeture d'un ensemble d'attributs X

- Étant donné un schéma de relation R(U) et F un ensemble de DFs défini sur U. Y et X deux sous ensembles de U.
  - $(F \models X \rightarrow Y) \Leftrightarrow (Y \subseteq [X]_F^+)$
  - $(K \subseteq U$ **superclé de** $R) \Leftrightarrow ([K]_F^+ = U)$
  - $(K \subseteq U$ clé candidate de $R) \Leftrightarrow ([K]_F^+ = U) \wedge (\forall A \in K, [K\text{-}\{A\}]_F^+ \neq U)$

Pas de $Y \subset U$ tel que $[Y]_F^+ = U$

- Une clé est un **ensemble minimal** d'attributs qui détermine tous les autres
- il peut y avoir **plusieurs clés pour une même relation**; on en choisit en général une comme **clé primaire**.

Fermeture transitive (cloture) des DFs

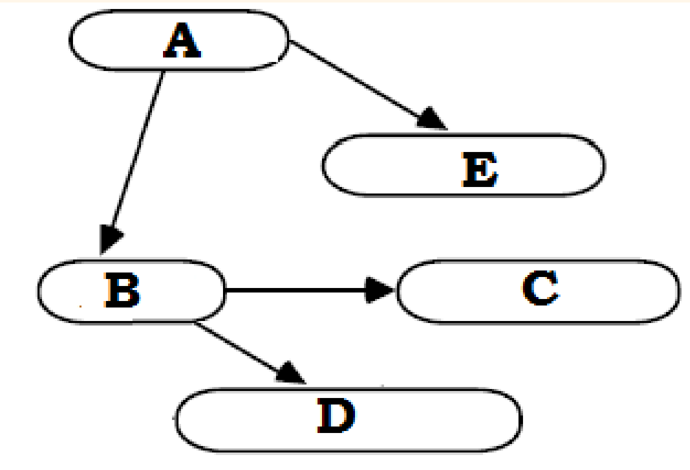
- Étant donné un ensemble F de DF, la **fermeture transitive de l'ensemble F**, notée **F^+**, est l'ensemble de toutes les DFs **élémentaires** qui peuvent être logiquement induites par application des axiomes d'Armstrong sur l'ensemble F.

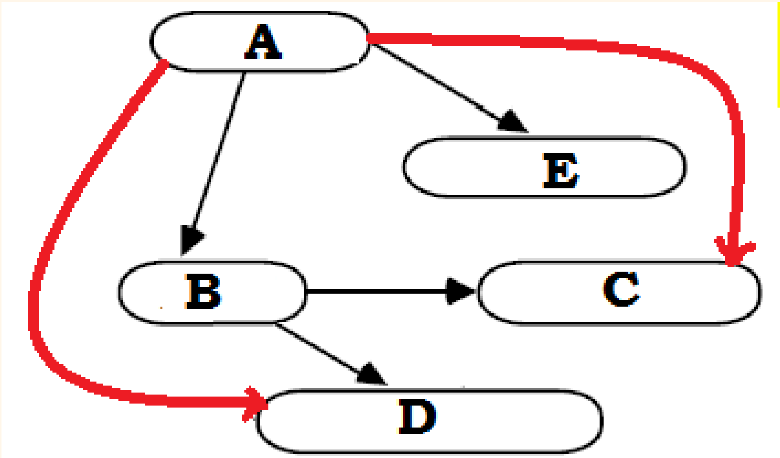
$$F^+ = \{X \rightarrow Y \mid F \models X \rightarrow Y\}$$

Exemple : Soit $F = \{A \rightarrow B, B \rightarrow C, B \rightarrow D, A \rightarrow E\}$. Calculer F^+ ?

$$F^+ = F \cup \{A \rightarrow C, A \rightarrow D\}.$$

F





F^+

Fermeture transitive des DFs

- **Définition**

Deux ensembles de DFs F et G sont dits **équivalents** si pour tout $(X \longrightarrow Y) \in F$, on a $G \models X \longrightarrow Y$ et vice-versa, ou ils possèdent la même fermeture (**Cloture**), c.à.d.,

$F^{+} = G^{+}$

$$F \equiv G \quad \Rightarrow \quad F^{+} = G^{+}$$

Couverture minimale

- La recherche de la C.M. (couverture **irredondante**, **irréductible**) d'un ensemble de DFs est un élément essential du processus de normalisation fin de décomposer un relation en plus petites relations.

- **Définition**

un ensemble de DFs F est minimal si seulement si

a) Les DFs de F sont sous forme canonique.

$X \longrightarrow Y$ est sous forme canonique si Y est un singleton (Y est constitué un seul attribut)

a) F ne contient pas des DFs redondantes

$$\forall f \in F, \text{ on a } F - \{f\} \not\equiv F$$

c) F ne contient pas des DFs redondantes a gauche

$$\forall f \in F, f: XY \longrightarrow A \text{ on a } (F - \{XY \longrightarrow A\} \cup \{Y \longrightarrow A\}) \not\equiv F$$

Algorithme pour calculer Min(F)

- **Etape 1 :** F1 $\leftarrow$ F sous forme canonique.
- **Etape 2 :**

  F2 $\leftarrow$ F1

  Pour chaque f $\in$ F2 **faire**

  Si (F2 $-\{f\}$) $\equiv$ F2 **alors**

  F2 $\leftarrow$ F2 $-\{f\}$
- **Etape 3 :**

  F3 $\leftarrow$ F2

  Pour chaque f: X$\rightarrow$ A $\in$ F3 **faire**

  Pour chaque attribut B $\in$ X **faire**

  Si (F3-{X $\rightarrow$A} U {X – {B}$\rightarrow$ A}) $\equiv$ F3 **alors**

  F3 $\leftarrow$ F3 -{X $\rightarrow$A} U {X – {B} $\rightarrow$ A}
- **Etape 4 :** retourner **F3** comme **min(F)** couverture minimale de F.

Note : il peut avoir plusieurs couvertures minimales équivalents.

Exemple

Calculer min(F) telque F ={AB→C, A →BE,C →E}

a) Mise sous forme canonique

F1 ={AB→C, A →B, A → E,C →E}

b) Eliminer la redondance des DFs.

F2 ← F1

- AB→C? $[AB]^+_{F1-\{AB\rightarrow C\}}$ = {A, B, E} , on C∉ $[AB]^+_{F1-\{AB\rightarrow C\}}$ donc AB→C n'est pas redondant .

- A→ B? $[A]^+_{F1-\{A\rightarrow B\}}$ = {A, E} , on B∉ $[AB]^+_{F1-\{A\rightarrow B\}}$ donc A→B n'est pas redondant .

- A→ E? $[A]^+_{F1-\{A\rightarrow E\}}$ = {A, B, C, E} , on E∈ $[A]^+_{F1-\{A\rightarrow E\}}$ donc A→E est redondant et on doit la supprimer.

F2 ← F2-{ A→E }

- C→ E? $[C]^+_{F1-\{C\rightarrow E\}}$ = {C} , on E∉ $[C]^+_{F1-\{C\rightarrow E\}}$ donc C→E n'est pas redondant .

Exemple (suite)

c) Eliminer la redondance à gauche.

$F3 \leftarrow F2$

$F3 = \{AB \rightarrow C, A \rightarrow B, C \rightarrow E\}$

- $AB \rightarrow C$
  - Si on supprime B, $[A]^+_{F3} = \{A, B, C, E\}$, on $C \in [A]^+_{F3}$ donc $AB \rightarrow C$ redondante a gauche. On remplace $AB \rightarrow C$ par $A \rightarrow C$
  - Si on supprime A, $[B]^+_{F3} = \{B\}$, on $C \notin [B]^+_{F3}$ donc on peut pas le supprimer.

$F3 \leftarrow F3 - \{AB \rightarrow C\} \cup \{A \rightarrow C\}$

d) $Min(F) = F3 = \{A \rightarrow C, A \rightarrow B, C \rightarrow E\}$

❑ **Question:** $F=\{AB \rightarrow CD, ACE \rightarrow B, D \rightarrow C, C \rightarrow D, CD \rightarrow BE\}$, cet ensemble est t-il irréductible?

Décomposition des relations

- **Décomposition d'un schéma de relation**

  La décomposition d'un schéma de relation **R(A1, A2, ..., An)** est sa substitution par un ensemble de schéma de relations R1, **R2, ..., Rm** telles que :

schéma(R) = schéma(R1) U schéma(R2) U ... schéma(Rm)

- **But de la décomposition**: Casser R en plus petites relations afin d' éliminer :
  - redondance
  - Anomalies de mise à jour et redondance
- **Critères de bonne décomposition**
  - Décomposition sans perte d'informations.
  - Décomposition préservant les DFs.

Décomposition SPI

- La décomposition d'un schéma de relation R(A1, A2, ..., An) par un ensemble de schéma de relations R1, R2, ..., Rm est **sans perte d'informations (SPI)** par rapport à un ensemble de DF si et seulement si, quelle que soit I une instance de R, alors

$$I = I_1 \bowtie I_2 \bowtie \dots \bowtie I_m \quad \text{où } I_j = \Pi_{Rj}(I)$$

- **Théorème:**

- Une décomposition de R en R1 et R2 est SPI ssi R1 ∩ R2 est **super-clé de R1 et/ou R2**.
- Une décomposition R = {R1, R2} est sans perte d'informations par rapport à un ensemble de DFs F ssi

F |= |R1∩ R2| → (R1-R2) ou F |= | R1∩ R2| → (R2-R1)

Décomposition SPI (Suite)

F = { NVH →Marque, Type, Puis, Couleur, Puis; Type → Marque, Puis}

VOITURE	NVH	TYPE	MARQUE	PUIS	COULEUR
	872RH7	R21	RENAULT	7	BLEUE
	975BE5	R21	RENAULT	7	BEIGE
	975AB8	205	PEUGEOT	8	ROUGE

- **Décomposition 1 :** R1(NVH, Type, Couleur),
  R2(Type, Marque, Puis)
- **Décomposition 2 :** R1(NVH, Type),
  R2(Type, Puis, Couleur)
  R3(Type, marque)

Décomposition SPI (Suite)

VOITURE	NVH	TYPE	MARQUE	PUIS	COULEUR
	872RH7	R21	RENAULT	7	BLEUE
	975BE5	R21	RENAULT	7	BEIGE
	975AB8	205	PEUGEOT	8	ROUGE

R1	NVH	TYPE	COULEUR
	872RH7	R21	BLEUE
	975BE5	R21	BEIGE
	975AB8	205	ROUGE

R2	TYPE	MARQUE	PUIS
	R21	RENAULT	7
	205	PEUGEOT	8

Décomposition 1 :

R1(NVH, Type, Couleur)

R2(Type, Marque, Puis)

R1	NVH	TYPE
	872RH7	R21
	975BE5	R21
	975AB8	205

R2	TYPE	PUIS	COULEUR
	R21	7	BLEUE
	R21	7	BEIGE
	205	8	ROUGE

R3	TYPE	MARQUE
	R21	RENAULT
	205	PEUGEOT

Décomposition 2 :

R1(NVH, Type)

R2(Type, Puis, Couleur)

R3(Type, marque)

Les formes normales

- **Définition 1NF**

  Une relation est en **1ère forme normale** si tout attribut contient une valeur atomique (unique)

- **Exemple**

  PERSONNE(NoSS, nom, prenom,diplomes)

  Cette relation n'est pas en 1 FN , elle doit être décomposé en 2 relations afin d' éliminer l'attribut multivalué.

  PERSONNE(<u>NoSS</u>, nomn, prenom,diplome)

  DIPLOME(<u>NoSS, diplome</u>)

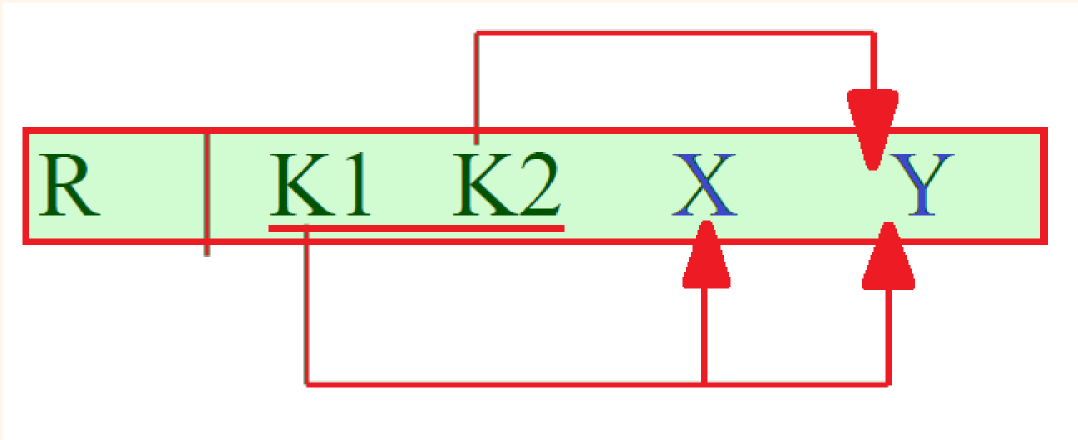
Les formes normales

- **Définition 2NF**

une relation est en 2e forme normale ssi :

1. Elle est en 1ère forme
2. Tout attribut non clé ne dépend pas d'une partie de clé

- **Schéma**



- Une telle relation doit être décomposée en

  R1(<u>K1,K2,</u>X) et R2(<u>K2</u>,Y)

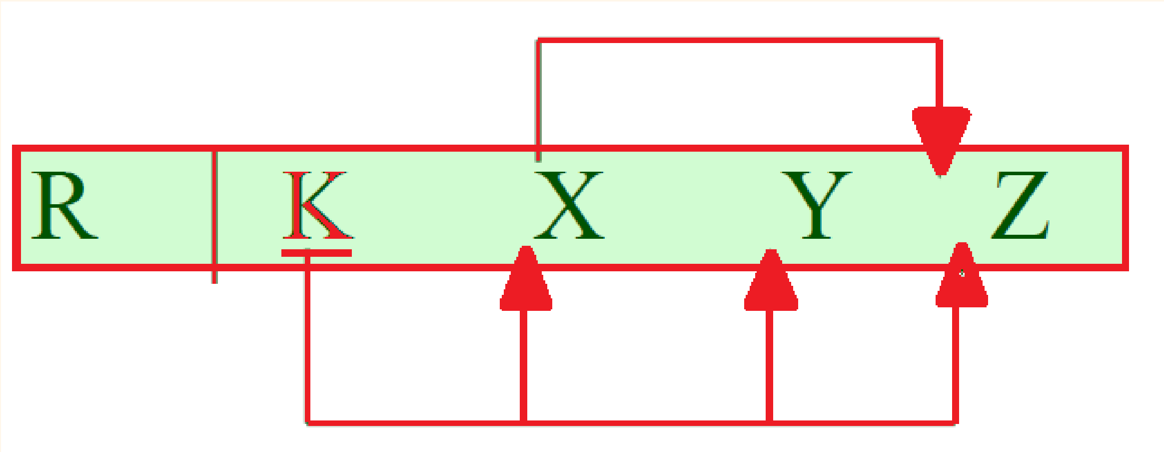
Les formes normales (suite)

- **Définition 3NF**

une relation est en 3e forme normale ssi :

1. Elle est en 2 ème FN
2. tout attribut n'appartenant pas a une clé ne dépend pas d'un ensemble d'attributs qui ne sont pas clé.

- **Schéma**



- Une telle relation doit être décomposée en

R1(<u>K1,</u>X, Y) et R2(<u>X</u>,Z)

- **Note:** la 3FN permet d'assurer l' élimination des redondances dues aux DFs transitives.

Comment calculer une forme normale

Soit R(A1, ..., An) et F ensemble des DFs associé.

- calculer F^+ (Dessiner le graphe des DFs)
- déterminer la (les) clé (s) de R
- Partitionner les attributs en attributs clés (ils appartiennent à au moins une clé) et attributs non clés
- appliquer les définitions de forme normale (depuis la 1ère)

Propriétés

- Toute relation R admet au moins une décomposition en 3FN qui préserve l'information et les DF .
- Deux approches pour la calculer :
  a) **décomposition** : on s'arrête dès que les relations dérivées sont en 3FN (mais pas de garantie sur les DF)

  b) **synthèse** à partir de la couverture minimale

Algorithme de synthèse

Principe de l'algorithme :

Données: la relation universelle + ensemble F des DFs

1) A partir du graphe G des DF, calculer une couverture minimale **min(F) .**
2) Editer l'ensemble des attributs isolés dans une même relation (tous sont clés)
3) Recherche le plus grand ensemble X d'attributs qui détermine d'autres attributs
4) Editer la relation R(X, A1, ..., An)
5) Supprimer les DF $X \rightarrow A1$, .., $X \rightarrow An$ du graphe de couverture minimale **min(F) .**
6) Supprimer les attributs isolés de **min(F) .**
7) Reprendre l'opération à partir de l'étape **3** jusqu'à ce que **min(F)** soit vide.

Algorithme de décomposition

Soit R(A1, ..., An) et F ensemble des DFs associé.

- Trouver min(F) (Dessiner le graphe des DFs)
- déterminer la (les) clé (s) de R
- Partitionner les attributs en attributs clés (ils appartiennent à au moins une clé) et attributs non clés
- appliquer les définitions de forme normale (depuis la 1ère)