

SMI6U05 : Bases de données
Conception de base de données :
DFs, décompositions, et formes normales

Meryem Billami et Luigi Santocanale
LIS, Aix-Marseille Université

14 mars 2019

Plan

Introduction

Dépendances fonctionnelles (DFs)

Décompositions

Dépendances fonctionnelles et décompositions

Formes normales

Plan

Introduction

Dépendances fonctionnelles (DFs)

Décompositions

Dépendances fonctionnelles et décompositions

Formes normales

Une table (entité), est elle bien conçue ?

(Contre)Exemple :

Voiture	nv	marque	type	puissance	couleur
	872RH75	Peugeot	P206A	7	bleue
	975AB80	Peugeot	P206A	7	rouge

- Les données sont redondantes !
- Anomalies de mise à jour :
 - ▶ introduction d'incohérences
Exemple : correction dans une seule tuple
 - ▶ introduction de valeurs nulles
Exemple : ajout d'un nouvel modèle de voiture ;
nv et couleur NULL ?

Evolution d'une BD

- Une base évolue : donnée ET schémas
- Souvent (manque de temps, pressions commerciales, arrogance) :
modification directe du Modèle Physique des Données
- Du coup :
 - ▶ incohérence,
 - ▶ absence de documentation,
 - ▶ mauvaise base.
- Evolution = nouveau MCD, puis MLD !
- Sinon : reverse engineering
- Problématique actuellement très étudiée

Plan

Introduction

Dépendances fonctionnelles (DFs)

Décompositions

Dépendances fonctionnelles et décompositions

Formes normales

Dépendances fonctionnelles

- Liens fonctionnels entre attributs
- Ces liens sont souvent laissés implicites dans le schéma
- Une bonne conception explicite ces contraintes dans le schéma

Exemple

Voiture	nv	marque	type	puissance	couleur
	872RH75	Peugeot	P206A	7	bleue
	975AB80	Peugeot	P206A	7	rouge



Notation :

`type → marque, type → puissance,`

`type → marque, puissance`

D'autres DFs :

`nv → (marque, type, puissance, couleur)`

`(type, marque) → puissance`

Exemple

Voiture	nv	marque	type	puissance	couleur
	872RH75	Peugeot	P206A	7	bleue
	975AB80	Peugeot	P206A	7	rouge



Voiture	nv	marque	type	puissance	couleur
----------------	----	--------	------	-----------	---------

Notation :

$\text{type} \rightarrow \text{marque}, \quad \text{type} \rightarrow \text{puissance},$

$\text{type} \rightarrow \text{marque}, \text{puissance}$

D'autres DFs :

$\text{nv} \rightarrow (\text{marque}, \text{type}, \text{puissance}, \text{couleur})$

$(\text{type}, \text{marque}) \rightarrow \text{puissance}$

Exemple

Voiture	nv	marque	type	puissance	couleur
	872RH75	Peugeot	P206A	7	bleue
	975AB80	Peugeot	P206A	7	rouge



Voiture	nv	marque	type	puissance	couleur
----------------	----	--------	------	-----------	---------

Notation :

$\text{type} \rightarrow \text{marque}, \quad \text{type} \rightarrow \text{puissance},$

$\text{type} \rightarrow \text{marque}, \text{puissance}$

D'autres DFs :

$\text{nv} \rightarrow (\text{marque}, \text{type}, \text{puissance}, \text{couleur})$

$(\text{type}, \text{marque}) \rightarrow \text{puissance}$

DFs (II)

- Permettent de repérer :
 - ▶ redondances
 - ▶ anomalies
- Buts :
 - ▶ affiner,
 - ▶ préciser,
 - ▶ rendre plus concis un MCD
 - ▶ data mining
- Aussi : améliorer la performance de certaines requêtes

Amélioration des requêtes

Calculer

$$A := \{(x, y, z) \mid \text{il existe } y', z' \text{ tels que } R(x, y, z') \text{ et } R(x, y', z)\}$$

Nécessaire calculer une jointure :

$$\Pi_{x,y,z'}(R \bowtie_{x=x'} R')$$

Si $y = f(x)$ (et donc $y' = f(x) = y$)

$$\begin{aligned} A &= \{(x, y, z) \mid \text{il existe } z' \text{ tels que } R(x, y, z') \text{ et } R(x, y, z)\} \\ &= \{(x, y, z) \mid R(x, y, z) \text{ et (il existe } z' \text{ tels que } R(x, y, z'))\} \\ &= \{(x, y, z) \mid R(x, y, z)\} = R. \end{aligned}$$

Pas de jointure !!!

Amélioration des requêtes

Calculer

$$A := \{ (x, y, z) \mid \text{il existe } y', z' \text{ tels que } R(x, y, z') \text{ et } R(x, y', z) \}$$

Nécessaire calculer une jointure :

$$\Pi_{x,y,z'}(R \bowtie_{x=x'} R')$$

Si $y = f(x)$ (et donc $y' = f(x) = y$)

$$\begin{aligned} A &= \{ (x, y, z) \mid \text{il existe } z' \text{ tels que } R(x, y, z') \text{ et } R(x, y, z) \} \\ &= \{ (x, y, z) \mid R(x, y, z) \text{ et (il existe } z' \text{ tels que } R(x, y, z')) \} \\ &= \{ (x, y, z) \mid R(x, y, z) \} = R. \end{aligned}$$

Pas de jointure !!!

Amélioration des requêtes

Calculer

$$A := \{ (x, y, z) \mid \text{il existe } y', z' \text{ tels que } R(x, y, z') \text{ et } R(x, y', z) \}$$

Nécessaire calculer une jointure :

$$\Pi_{x,y,z'}(R \bowtie_{x=x'} R')$$

Si $y = f(x)$ (et donc $y' = f(x) = y$)

$$\begin{aligned} A &= \{ (x, y, z) \mid \text{il existe } z' \text{ tels que } R(x, y, z') \text{ et } R(x, y, z) \} \\ &= \{ (x, y, z) \mid R(x, y, z) \text{ et (il existe } z' \text{ tels que } R(x, y, z')) \} \\ &= \{ (x, y, z) \mid R(x, y, z) \} = R. \end{aligned}$$

Pas de jointure !!!

Amélioration des requêtes

Calculer

$$A := \{ (x, y, z) \mid \text{il existe } y', z' \text{ tels que } R(x, y, z') \text{ et } R(x, y', z) \}$$

Nécessaire calculer une jointure :

$$\Pi_{x,y,z'}(R \bowtie_{x=x'} R')$$

Si $y = f(x)$ (et donc $y' = f(x) = y$)

$$\begin{aligned} A &= \{ (x, y, z) \mid \text{il existe } z' \text{ tels que } R(x, y, z') \text{ et } R(x, y, z) \} \\ &= \{ (x, y, z) \mid R(x, y, z) \text{ et (il existe } z' \text{ tels que } R(x, y, z')) \} \\ &= \{ (x, y, z) \mid R(x, y, z) \} = R. \end{aligned}$$

Pas de jointure !!!

Plan

Introduction

Dépendances fonctionnelles (DFs)

Décompositions

Dépendances fonctionnelles et décompositions

Formes normales

Décompositions de relations

- Décomposition : projection
- Recomposition : jointure (naturelle).
- Une **décomposition** d'une relation $R(a_1, a_2, \dots, a_n)$ est une collection de relations R_1, R_2, \dots, R_m telles que
 - ▶ $R_i = \Pi_{A_i}(R)$, $i = 1, \dots, m$
 - ▶ $\bigcup_{i=1, \dots, m} A_i = \{a_1, \dots, a_n\}$.
- Une **décomposition** R_1, R_2, \dots, R_m de R est **sans perte d'information** si

$$R = R_1 \bowtie R_2 \bowtie \dots \bowtie R_m.$$

Perte d'information

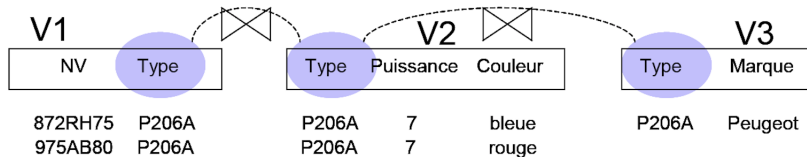
- Soit R_1, R_2, \dots, R_m une décomposition de R .
- $R_1 \bowtie R_2 \dots \bowtie R_m$ a le même schéma (les mêmes attributs) que R mais pas forcément le même contenu.
- On a, toujours,

$$R \subseteq R_1 \bowtie R_2 \bowtie \dots \bowtie R_m.$$

- Si l'inclusion est stricte : perte d'information.
La jointure est moins détaillée/moins précise/plus générale que R)
- Si égalité : décomposition sans perte d'information

Perte d'information : exemple

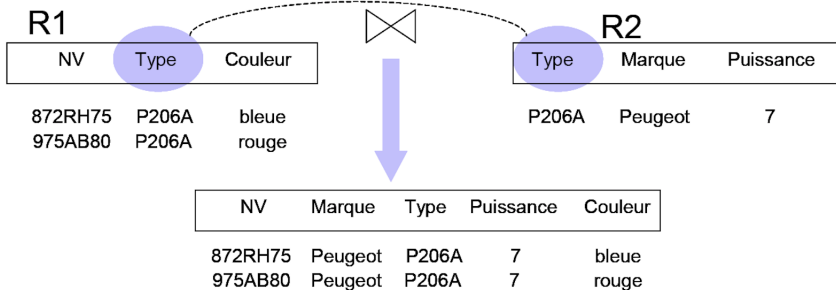
Voiture	nv	marque	type	puissance	couleur
	872RH75	Peugeot	P206A	7	bleue
	975AB80	Peugeot	P206A	7	rouge



NV	Marque	Type	Puissance	Couleur
872RH75	Peugeot	P206A	7	bleue
872RH75	Peugeot	P206A	7	rouge
975AB80	Peugeot	P206A	7	bleue
975AB80	Peugeot	P206A	7	rouge

Sans perte : exemple

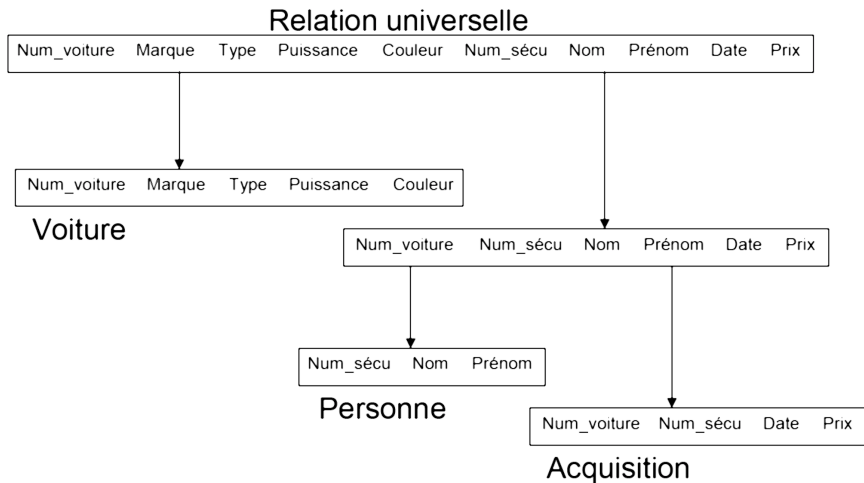
Voiture	nv	marque	type	puissance	couleur
	872RH75	Peugeot	P206A	7	bleue
	975AB80	Peugeot	P206A	7	rouge



Modélisation : approche par décomposition

- Relation universelle : relation unique dont le schéma est constitué de tous les attributs (les attributs du même concept ont le même nom, et vice versa).
- A partir de la relation universelle, isoler les entités et associations élémentaires (canoniques, ne pouvant être de re-découpées) par une processus de raffinements successifs.
- Le même approche peut être utilisé avec un schéma/table trop grand.

Approche par décomposition : exemple



Plan

Introduction

Dépendances fonctionnelles (DFs)

Décompositions

Dépendances fonctionnelles et décompositions

Formes normales

Dépendances fonctionnelles et décomposition

- Les dépendances fonctionnelles permettent de sélectionner les attributs pour une décomposition sans perte.
- Exemple : `marque`, `type` et `puissance` partagent respectivement les mêmes valeurs : ces attributs sont fonctionnellement dépendants et donc il est intéressant que la décomposition les garde ensemble.
- Gain : factorisation.

Dépendance fonctionnelle

- DF : lien sémantique entre 2 attributs.
- Lors de la conception :
peut exprimer une contrainte à mettre en œuvre dans la BD.
- Quand la BD évolue :
peut être découvert par analyse des données (lignes) de la base.

Définition formelle

Définition. Soient

- $R(a_1, a_2, \dots, a_n)$ une relation,
- $X, Y \subseteq \{a_1, a_2, \dots, a_n\}$.

On dit que

- Y dépend fonctionnellement de X dans R , ou bien
- X détermine Y dans R , ou encore
- R satisfait $X \rightarrow Y$,

et écrit on cela par

$$R \models X \rightarrow Y,$$

si, pour tout $t, t' \in R$,

$$\pi_X(t) = \pi_X(t') \text{ implique } \pi_Y(t) = \pi_Y(t').$$

Rappel : $\pi_X(t) = t \upharpoonright_X$.

Dépendance fonctionnelle (suite)

- On souhaite manipuler des expression du type

$$X \rightarrow Y$$

sans faire référence à une relation R à un instant donné.

- Par exemple, lors de la conception, quand la table est vide.
- Y dépend fonctionnellement de X si, étant donnée une valeur de X , il lui correspond une unique valeur de Y .
- C'est une assertion sur toutes les valeurs possibles !
Non pas par rapport au contenu de la base à un instant donné.
Assertion sur le monde réel.
- Exemple : les DFs pourraient exister :

$$\begin{aligned} &Type \rightarrow Marque, & Type \rightarrow Puissance, \\ &NV \rightarrow couleur, \\ &(Type, Marque) \rightarrow Puissance. \end{aligned}$$

Schéma de relation avec DFs

Définition. Une schéma de relation avec contraintes de DFs est un couple $(R(a_1, \dots, a_n), \mathbf{F})$ tel que :

- $R(a_1, \dots, a_n)$ est un schéma de relation (nom de la relation plus noms des colonnes/attributs)
- \mathbf{F} un ensemble de DFs de la forme $X \rightarrow Y$ avec $X, Y \subseteq \{a_1, \dots, a_n\}$.

Définition formelle

Définition. On dit que

$$X_1 \rightarrow Y_1, \dots, X_k \rightarrow Y_k \quad \text{implique} \quad X_0 \rightarrow Y_0$$

si

$$R \models X_1 \rightarrow Y_1, \dots, R \models X_k \rightarrow Y_k \quad \text{implique} \quad R \models X_0 \rightarrow Y_0.$$

pour toute relation $R(a_1, a_2, \dots, a_n)$ t.q. $\bigcup_{i=0, \dots, k} X_i \cup Y_i \subseteq \{a_1, a_2, \dots, a_n\}$.

En particulier (cas $k = 0$) :

Définition. On dit/écrit

$$X \rightarrow Y$$

si

$$R \models X \rightarrow Y$$

pour toute relation $R(a_1, a_2, \dots, a_n)$ telle que $X, Y \subseteq \{a_1, a_2, \dots, a_n\}$.

Axiomatisation et propriétés des DFs

Notation : $(X, Y) = X \cup Y$.

Axiomatisation de Armstrong 1974 :

- **Réflexivité** : $X \rightarrow Y$, pour $Y \subseteq X$
- **Augmentation** : $X \rightarrow Y$ implique $(X, Z) \rightarrow (Y, Z)$,
- **Transitivité** : $X \rightarrow Y$, $Y \rightarrow Z$, implique $X \rightarrow Z$.

D'autres propriétés se dérivent des précédentes :

- **Union** : $X \rightarrow Y$, $X \rightarrow Z$ implique $X \rightarrow (Y, Z)$,
- **Pseudo-transitivité** : $X \rightarrow Y$, $(W, Y) \rightarrow Z$, implique $(W, X) \rightarrow Z$,
- **Décomposition** : si $Z \subseteq Y$, alors $X \rightarrow Y$ implique $X \rightarrow Z$.

L'axiomatisation de Armstrong est correcte et complète.

Fermeture d'un ensemble de DFs

Soit \mathbf{F} un ensemble de DFs.

Notons \mathbf{F}^+ le plus petit ensemble contenant \mathbf{F} , fermé par rapport aux règles de réflexivité, augmentation, transitivité. On appelle \mathbf{F}^+ la **fermeture** de \mathbf{F} .

Posons :

$$R \models \mathbf{F} \text{ ssi } R \models X \rightarrow Y, \text{ pour tout } X \rightarrow Y \in \mathbf{F},$$
$$\mathbf{F} \models X \rightarrow Y \text{ ssi } X \rightarrow Y \in \mathbf{F}^+$$

Clairement, si $R \models \mathbf{F}$ et $\mathbf{F} \models X \rightarrow Y$, alors $R \models X \rightarrow Y$.

Deux ensembles \mathbf{F}_0 et \mathbf{F}_1 de dépendances fonctionnelles sont **équivalents** (on note cela par $\mathbf{F}_0 \sim \mathbf{F}_1$) si $\mathbf{F}_0^+ = \mathbf{F}_1^+$.

Décomposition sans perte et DF

Plusieurs théorèmes :

- Si $R \models X \rightarrow Y$, alors

$$R = R_1 \bowtie R_2 \quad \text{où}$$
$$R_1 = \Pi_{X,Y}(R), \quad R_2 = \Pi_{X,Z}(R),$$

avec $X \cup Y \cup Z$ tous les attributs de R .

- Soient R une relation d'attributs $X \cup Y$ et \mathbf{F} une collection de FDs tels que $R \models \mathbf{F}$. Si
 - ▶ $(X \cap Y) \rightarrow (X \setminus Y) \in \mathbf{F}^+$, ou
 - ▶ $(X \cap Y) \rightarrow (Y \setminus X) \in \mathbf{F}^+$,

alors la décomposition $\Pi_X(R), \Pi_Y(R)$ est sans perte :

$$R = \Pi_X(R) \bowtie \Pi_Y(R).$$

DF élémentaires

Soit \mathbf{F} un ensemble de DFs.

Définition. Une **dépendance fonctionnelle élémentaire** (de \mathbf{F}) est une DF de la forme

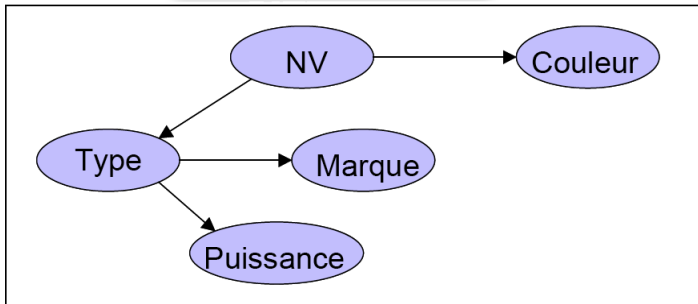
$$X \rightarrow \{a\}$$

telle que

- $a \notin X$,
 - $X' \subseteq X$ et $\mathbf{F} \models X' \rightarrow \{a\}$ implique $X' = X$.
-
- Exemple : $Type \rightarrow Puissance$
 - Contre-exemple : $(Type, Marque) \rightarrow Puissance$
 - Dépendances fonctionnelles pour amélioration de schéma : uniquement élémentaires.

Graphe de DF élémentaires

- Grâce à ce graphe la transitivité entre dépendance est facilement lisible.
- Exemple où tout dépendance $X \rightarrow a$ est telle que X singleton) :



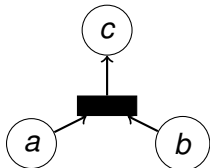
Graphe de DF élémentaires (II)

Si, dans une dépendance élémentaire $X \rightarrow a$, X n'est pas un singleton, on utilise un "graphisme" style réseaux de Pétri.

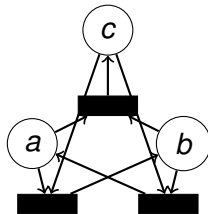
Exemple :

$$ab \rightarrow c, \quad ac \rightarrow b, \quad bc \rightarrow a.$$

On représente $ab \rightarrow c$ par :



On représente les trois DFs par :



Couverture minimale

Définition. Soit \mathbf{F} un ensemble de DFs. Une **couverture minimale** de \mathbf{F} est un ensemble \mathbf{C} de DFs tel que

- $\mathbf{C} \sim \mathbf{F}$,
- tout $X \rightarrow Y \in \mathbf{C}$ est élémentaire,
- pour tout $X \rightarrow Y \in \mathbf{C}$, $\mathbf{C} \setminus \{X \rightarrow Y\} \not\sim \mathbf{F}$.

Algorithme de calcul de la couverture minimale

Entrée : un ensemble \mathbf{F} de DFs

Sortie : une couverture minimale \mathbf{C} de \mathbf{F}

$\mathbf{C} := \mathbf{F}$

Pour chaque $X \rightarrow \{a_1, \dots, a_n\} \in \mathbf{C}$ avec $n \geq 2$:

..... $\mathbf{C} := \mathbf{C} \setminus \{X \rightarrow \{a_1, \dots, a_n\}\} \cup \{X \rightarrow a_1, X \rightarrow a_2, \dots, X \rightarrow a_n\}$

Pour chaque $X \rightarrow a \in \mathbf{C}$, pour chaque $b \in X$:

..... soit $\mathbf{D} := \mathbf{C} \setminus \{X \rightarrow a\} \cup \{X \setminus \{b\} \rightarrow a\}$

..... si $\mathbf{C} \sim \mathbf{D}$:

..... $\mathbf{C} := \mathbf{D}$

Pour chaque $X \rightarrow a \in \mathbf{C}$:

..... soit $\mathbf{D} := \mathbf{C} \setminus \{X \rightarrow a\}$

..... si $\mathbf{C} \sim \mathbf{D}$:

..... $\mathbf{C} := \mathbf{D}$

Algorithmme : exemple

$$\begin{aligned} \mathbf{F} = & ab \rightarrow c, \quad c \rightarrow a, \quad bc \rightarrow d, \\ & acd \rightarrow b \quad d \rightarrow \mathbf{eg}, \quad be \rightarrow c \\ & cg \rightarrow \mathbf{bd}, \quad ce \rightarrow \mathbf{ag} \end{aligned}$$

Etape 1 :

$$\begin{aligned} \mathbf{C} = & ab \rightarrow c, \quad c \rightarrow a, \quad bc \rightarrow d, \\ & acd \rightarrow b, \quad d \rightarrow e, \quad be \rightarrow c, \\ & \quad \quad \quad d \rightarrow g, \\ & cg \rightarrow b, \quad ce \rightarrow a, \\ & cg \rightarrow d, \quad ce \rightarrow g. \end{aligned}$$

Algorithmme : exemple

Etape 1

$$\begin{aligned} \mathbf{C} = & ab \rightarrow c, \quad c \rightarrow a, \quad bc \rightarrow d, \\ & \mathbf{acd} \rightarrow \mathbf{b}, \quad d \rightarrow e, \quad be \rightarrow c, \\ & \quad \quad \quad d \rightarrow g, \\ & cg \rightarrow b, \quad \mathbf{ce} \rightarrow \mathbf{a}, \\ & cg \rightarrow d, \quad ce \rightarrow g. \end{aligned}$$

Au final :

$$\begin{aligned} \mathbf{C} = & ab \rightarrow c, \quad c \rightarrow a, \quad bc \rightarrow d, \\ & ce \rightarrow g, \quad cg \rightarrow d, \quad d \rightarrow e, \\ & d \rightarrow g, \quad be \rightarrow c, \quad cg \rightarrow b \end{aligned}$$

DFs et clés

- Clé primaire : intuitif jusqu'à maintenant.
- Une clé d'une relation $R(a_1, \dots, a_n)$ est un sous-ensemble X des attributs de la relation R tel que les deux conditions ci-dessous sont réunies :
 1. $X \rightarrow \{a_1, \dots, a_n\}$,
 2. il n'existe pas de $Y \subseteq X$, $Y \neq X$, tel que $Y \rightarrow \{a_1, \dots, a_n\}$
- Moins formellement : une clé est un ensemble minimal d'attributs qui détermine tous les autres.
- Clé candidate (à être primaire) : toute clé.

Plan

Introduction

Dépendances fonctionnelles (DFs)

Décompositions

Dépendances fonctionnelles et décompositions

Formes normales

Formes normales

- Permettent de faciliter (et guider) la décomposition de relation sans perte d'information.
La conception est
 - ▶ précise,
 - ▶ cohérente,
 - ▶ sans redondance.
- Plusieurs formes normales :
 - ▶ les trois premières :-) pour la décomposition sans perte
 - ▶ la forme normal de Boyce-Codd
 - ▶ d'autres...

La première forme normale

Définition. Un schéma de relation est en **première forme normale** si

- tout attribut contient une valeur atomique (non composé).

C'est-à-dire : pas d'attribut multi-valué.

- Autre formulation : il n'y a pas dans la relation d'attribut dont la valeur est, par exemple, une liste/ensemble de valeurs.
- Or, être d'atomique dépend du domaine !
Exemple : chaînes de caractères.
- Facile à mettre en oeuvre : transformer toute relation $R(K, A)$ où A prend n valeurs en n relations $R_i(K, A)$.

Première forme normale, exemple

Personne(nom, prenom)

~>

Personne1(nom, prenom1)

Personne2(nom, prenom2)

Ou

~>

Personne(nom)

Prenom(nom, prenom)

Deuxième forme normale

Définition. Un schéma de relation est en **deuxième forme normale** si

- il est en première forme normale,
 - tout attribut n'appartenant pas à une clé ne dépend pas d'une partie (propre) de la clé.
-
- Élimine les redondances.

Deuxième forme normale, contre-exemple et exemple

Contre-exemple :

Livraison(ref_client, ref_article, quantite, adresse_livraison)
ref_client → adresse_livraison

Pas en 2NF, car :

$$\{\text{ref_client}\} \subsetneq \{\text{ref_client}, \text{ref_article}\}.$$

Exemple :

Casting(ref_film, ref_acteur, personnage, cachet)

Deuxième forme normale (suite)

- Transformation en DFN :
 - ▶ $R(\underline{K1}, \underline{K2}, X, Y)$ avec $K1 \rightarrow X$:
 - ▶ Décomposition en 2 relations $R1(\underline{K1}, \underline{K2}, Y)$ et $R2(\underline{K1}, X)$
- Contre-exemple précédent :

Livraison(ref_client, ref_article, quantité, adresse_livraison)
 $\text{ref_client} \rightarrow \text{adresse_livraison}$

~>

Livraison1(ref_client, ref_article, quantité)
Livraison2(ref_client, adresse_livraison)

- Souvent : changer les noms.
- Rejoint les critères de normalisation des attributs d'association de Merise.

Troisième forme normale

Définition. Un schéma de relation est en **troisième forme normale** si :

1. il est en deuxième forme normale
 2. tout attribut n'appartenant pas à une clé ne dépend pas d'un autre attribut non-clé.
- Élimine les redondances liées aux dépendances fonctionnelles transitives.
 - Doit être vérifiée pour chaque clé.
 - Attention : 1. est redondante, car 2. implique 1.

(Contre-)Exemple

Contre-exemple :

Voiture(nv, marque, type, puissance, couleur)
type \rightarrow marque, puissance

est en 2FN :

- nv est la seule clé et elle est un singleton,

mais pas en 3FN car

- type \rightarrow marque et type \notin {nv}.

Exemple :

\rightsquigarrow
Voiture(nv, type, couleur)
Modele(type, marque, puissance)

est en 3FN.

Préservation des DFs

DF(R) : DFs associées au schéma de R .

Définition. Une décomposition ($R_1, R_2 \dots R_n$) d'un schéma de relation R **préserve les dépendances fonctionnelles** si

$$\mathbf{DF}(R)^+ = \left(\bigcup_{i=1, \dots, n} \mathbf{DF}(R_i) \right)^+.$$

C'est-à-dire : la fermeture de l'union des **DF**(R_i) est et la même que celle de **DF**(R).

- Toute relation a au moins une décomposition en 3FN telle que :
 - ▶ préservation des dépendances fonctionnelles
 - ▶ décomposition sans perte.

Algorithme de décomposition en 3FN

Décomposition($R(a_1, \dots, a_n), \mathbf{F}$) =

1. Rechercher une couverture minimale \mathbf{C} de \mathbf{F} .
2. Partitionner \mathbf{C} en groupes de dépendances \mathbf{C}_i ayant la même partie gauche.
Chaque \mathbf{C}_i est de la forme $X_i \rightarrow Y_i$.
3. Fusionner les groupes $\mathbf{C}_i = X_i \rightarrow Y_i$ et $\mathbf{C}_j = X_j \rightarrow Y_j$ si $Y_i = X_j$ et $Y_j = X_i$.
4. Associer à chaque groupe $\mathbf{C}_i = X_i \rightarrow Y_i$ un schéma de relation R_i dont la clé est X_i et les constituants non clés est Y_i .
5. Si aucune des relations précédentes ne contient une clé candidate K de R , créer une relation supplémentaire K qui la contient.

Forme normale de Boyce-Codd

Définition. Un schéma de relation est en **BCFN** si pour toute dépendance $K \rightarrow a$ avec $a \notin K$, K est une clé.

- Toute relation a une décomposition en FNBC.
- Cette décomposition ne préserve pas en général les dépendances fonctionnelles.
- Idée :
ne pas représenter la même information plusieurs fois.

(Contre-)Exemples

- Contre-exemple :

```
Movie(title, director, actor)
title → director
```

n'est pas en FNBC. Car title n'est pas une clé.
Anomalies de mise à jour de la colonne directeur.

- La schéma Movie peut se décomposer en

```
Movie_director(title,director)
title → director

Movie_actor(title,actor)
```