

# ***Théorie de la Normalisation***

Talel.Abdessalem@enst.fr

Supports de cours : (1) *Database Management Systems*, R. Ramakrishnan and J. Gehrke, ed. McGrawHill, 2000.  
(2) *Bases de Données*, G. Gardarin, ed. Eyrolles, 2001.

1

## *Théorie de la normalisation*

- ✍ Objectif et principes
- ✍ Dépendances fonctionnelles
- ✍ Décomposition de schéma
- ✍ Formes normales
- ✍ Automatisation de la normalisation
- ✍ Conclusion

2

## Objectif et principes

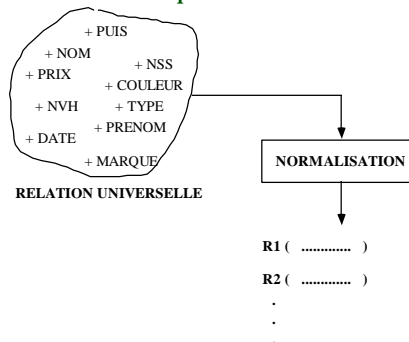
Objectif : Minimiser les redondances pour améliorer les performance de la BD.

☞ requêtes d'interrogation et de mise à jour plus rapides

☞ moins de valeurs NULL

☞ Gestion de l'intégrité des données simplifiée

Principe :



3

## Dépendances fonctionnelles

Définition :

$X \twoheadrightarrow Y$ ,  $X$  détermine  $Y$ , ou  $Y$  dépend fonctionnellement de  $X$ ,

$X$  et  $Y$  éléments de  $R$

est une assertion sur toutes les instances de la relation  $R$  telle que si deux tuples quelconques de  $R$  possèdent la même valeur sur  $X$ , alors ils doivent avoir la même valeur sur  $Y$ .

Formellement :

$X \twoheadrightarrow Y$  ssi quel que soit l'instance  $r$  de  $R$ , pour tout tuple  $t1$  et  $t2$  de  $r$  on a  $? x(t1) = ? x(t2) \implies ? y(t1) = ? y(t2)$

Utilité

Contrainte sur les données que peut contenir la relation.

Outil de contrôle d'intégrité des données (niveau schéma).

Outil mathématique permettant d'expliquer le processus de la « normalisation »

Vital lorsqu'on veut ré-étudier le schéma d'une BD qui pose problème.

4

## Axiomes d'Armstrong

### Propriétés de base

#### Réflexivité

**Y inclus dans X**  $\Rightarrow X \rightarrow Y$

#### Augmentation

$X \rightarrow Y \Rightarrow XZ \rightarrow YZ$

#### Transitivité

$X \rightarrow Y$  et  $Y \rightarrow Z \Rightarrow X \rightarrow Z$

### Propriétés déduites

#### Union

$X \rightarrow Y$   
 $X \rightarrow Z \Rightarrow X \rightarrow YZ$

#### Pseudo-transitivité

$X \rightarrow Y$   
 $YW \rightarrow Z \Rightarrow XW \rightarrow Z$

#### Décomposition

$X \rightarrow Y$   
 $Z$  inclus dans  $Y \Rightarrow X \rightarrow Z$

5

## $F^+$ et Couverture minimale

### Fermeture transitive d'un ensemble F de DF

- $F^+ = F^*$  ? {DF déduites de celles contenues dans F}

### Couverture minimale d'un ensemble F de DF

- Sous-ensemble minimum de **DF élémentaires** permettant de générer toutes les autres.
- **Théorème** : tout ensemble de DF possède une couverture minimale, en général non unique.
- Formellement, F est une CM ssi :
  - ⚡ Tout f dans F est élémentaire.
  - ⚡ Il n'existe pas de f dans F telle que  $F - \{f\}$  soit « équivalent » à F.

### DF élémentaires (les seules utiles)

- Une **DF élémentaire** est une DF de la forme  $X \rightarrow A$ , où :
  - ⚡ A est un attribut, X un ensemble d'attributs, A **n'est pas inclus dans X**
  - ⚡ **il n'existe pas X' inclus dans X** tel que  $X' \rightarrow A$

### Equivalence

- On dit que deux ensembles de DF sont équivalents s'ils possèdent la même fermeture transitive.

6

## Clé d'une relation

⌘ Définition :

**Ensemble minimum d'attributs qui détermine tous les autres.**

⌘ Formellement :

Soient :  $R(A_1, A_2, \dots, A_n)$  un schéma de relation,  $F^+$  l'ensemble des DF associés et  $X$  un sous-ensemble de  $A_1, A_2, \dots, A_n$ .

$X$  est une clé de  $R$  ssi :

⌘  $X \rightarrow A_1 A_2 \dots A_n$  est dans  $F^+$

⌘ il n'existe pas de sous-ensemble  $X'$  inclus dans  $X$  tel que  $X' \rightarrow A_1 A_2 \dots A_n$

⌘ Non unicité :

⌘ Il peut y avoir plusieurs clés pour une relation (clés candidates)

⌘ Une est choisie comme clé primaire

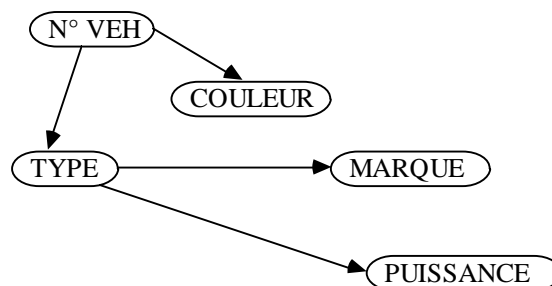
⌘ Sur-clé :

⌘ Sur-clé :  $X$  est sur-clé ssi  $X \rightarrow R$  dans  $F^+$

7

## Graphe des DF

VOITURE ( $N^{\circ}$ VEH, TYPE, COULEUR, MARQUE, PUISSANCE)

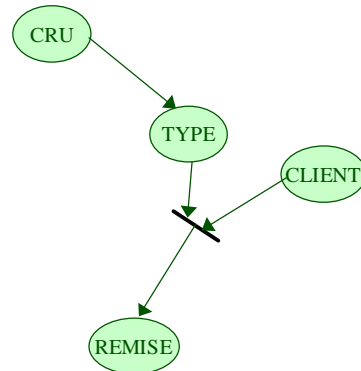


8

## Autre Exemple

CRU	TYPE	CLIENT	REMISE
CHENAS	A	C1	3%
MEDOC	A	C2	5%
JULIENAS	B	C1	4%

CRU --> TYPE  
TYPE, CLIENT --> REMISE



9

## Décomposition de schéma

Définition :

**La décomposition d'un schéma de relation  $R(A_1, \dots, A_n)$  consiste en son remplacement par une collection de relation  $(R_1, \dots, R_m)$  tel que**

$$\text{SCHEM}(R) = \text{SCHEM}(R_1) \cup \text{SCHEM}(R_2) \cup \dots \cup \text{SCHEM}(R_m)$$

But de la décomposition :

**Casser R en de plus petites relations afin d'éviter :**

**les redondances**

**les anomalies en mises à jour**

10



## Décomposition sans perte de DF

### Définition

Une décomposition ( $R_1, \dots, R_m$ ) de  $R$  préserve les DFs (SPD) si la fermeture transitive des DFs de  $R$  est la même que la fermeture transitive de l'union des Dfs des  $R_1, \dots, R_m$

### Exemple

1)  $R_1$  (NVH, TYPE, COULEUR)  $\Rightarrow F_1 = \{ NVH \rightarrow TYPE, NVH \rightarrow COULEUR \}$

$R_2$  (TYPE, MARQUE, PUIS)  $\Rightarrow F_2 = \{ TYPE \rightarrow MARQUE, TYPE \rightarrow PUIS \}$

$\Rightarrow$  OK

2)  $R'_1$  (NVH, TYPE)  $\Rightarrow F'_1 = \{ NVH \rightarrow TYPE \}$

$R'_2$  (TYPE, PUIS, COULEUR)  $\Rightarrow F'_2 = \{ TYPE \rightarrow PUIS \}$

$R'_3$  (TYPE, MARQUE)  $\Rightarrow F'_3 = \{ TYPE \rightarrow MARQUE \}$

$\Rightarrow$  ON A PERDU LA DF : NVH  $\rightarrow$  COULEUR

13

## Exemple

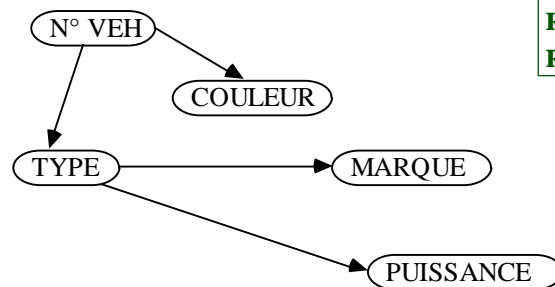
VOITURE (NVH, MARQUE, TYPE, PUIS, COULEUR)

VOITURE	NVH	MARQUE	TYPE	PUIS	COULEUR
872RH75	RENAULT	R21	7	BLEUE	
975AB80	RENAULT	R21	7	BEIGE	

$R_1$  (NVH, TYPE, COULEUR)  
 $R_2$  (TYPE, MARQUE, PUIS)

Décomposition

$R'_1$  (NVH, TYPE)  
 $R'_2$  (TYPE, PUIS, COULEUR)  
 $R'_3$  (TYPE, MARQUE)



14

## Formes normales

### ☞ Définition 1NF

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

### ☞ Exemple

PERSONNE	NOM	PROFESSION
	DUPONT	Ingénieur, Professeur
	MARTIN	Géomètre

Une telle relation doit être décomposée en répétant les noms pour chaque profession

15

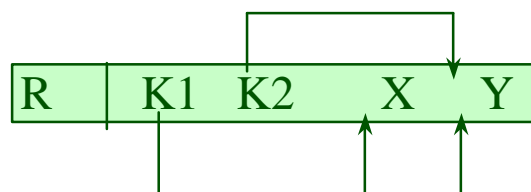
## 2NF

### ☞ Définition

une relation est en 2e forme normale ssi :

- ☞ elle est en 1ère forme
- ☞ tout attribut non clé ne dépend pas d'une partie de clé

### ☞ Schéma



Une telle relation doit être décomposée en  
 $R_1(\underline{K1}, \underline{K2}, X)$  et  $R_2(\underline{K2}, Y)$

16

## Exemple 2NF

### Exemple 1 :

- Fournisseur (nom, adresse, article, prix)
- La clé est (nom, article)
- Mais nom --> adresse : pas en 2e forme

### Exemple 2 :

- R (cru, type, client, remise)
- La clé est (cru, client)
- Mais cru --> type : pas en 2e forme

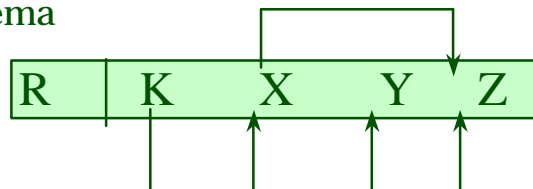
17

## 3NF

### Définition

- une relation est en 3e forme normale ssi :
  - ≠ elle est en 2e forme
  - ≠ tout attribut n'appartenant pas a une clé ne dépend pas d'un autre attribut non clé

### Schéma



Une telle relation doit être décomposée en  
 $R1(\underline{K}, X, Y)$  et  $R2(X, Z)$

18

## Exemple 3ème Forme

### Exemple

- Voiture (nvh, marque, type, puissance, couleur)
- Type --> marque
- Type --> puissance

Pas en 3ème forme !

### Caractéristiques de la 3NF :

- redondance limitée
- On arrive toujours à une décomposition en 3NF SPI et SPD

19

## Exemples de Décomposition

### Voiture (nvh, marque, type, puissance, couleur)

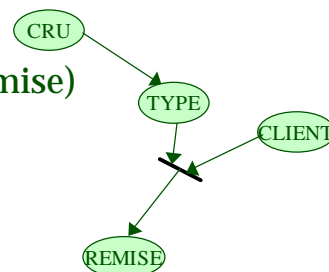
véhicule (nvh, type, couleur)

Modèle (type, marque, puissance)

### Réduction (cru, type, client, remise)

Remise (cru, client, remise)

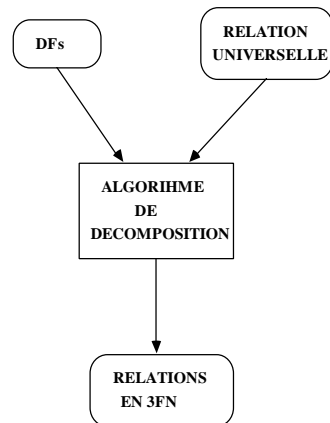
Type (cru, type)



20

# Automatisation de la décomposition

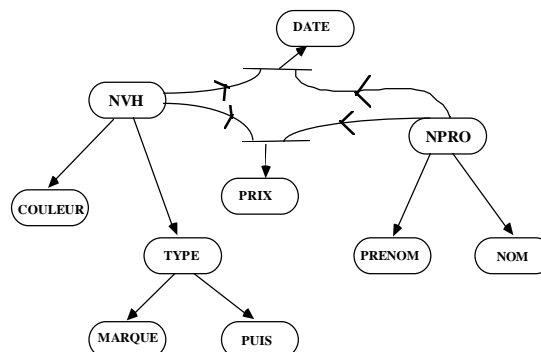
**AUTOMATISATION DU  
PROCESSUS DE DECOMPOSITION  
EN 3NF, A PARTIR DE L'ENSEMBLE  
DES ATTRIBUTS ET DES DFs**



21

## Algorithme de décomposition

1. A partir du graphe G des DF, calculer une couverture minimale C
2. Editer l'ensemble des attributs isolés dans une même relation (tous les attributs sont clés)
3. Rechercher le plus grand ensemble X d'attributs qui détermine d'autres attributs A1, ... An
4. Editer la relation (X, A1, ... , An) qui est en 3FN
5. Supprimer les DFs (X->A1, ... X->An) du graphe de couverture minimale C
6. Supprimer les attributs isolés de C (i.e. les attributs non-cible ni source de DF)
7. Répéter l'opération de réduction du graphe C à partir de l'étape 3, jusqu'à ce que C soit vide.
8. Ajouter une relation dont les attributs forment une clé de la relation universelle



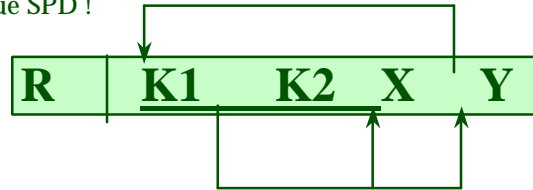
22

## Encore moins de redondances : la forme BCNF

### ☞ Définition

une relation est en BCNF (Boyce-Codd Normal Form) ssi :  
les seules dépendances fonctionnelles sont du type clé détermine attribut non clé

Plus simple que 3NF, un peu plus fort, mais ne peut pas toujours être obtenue SPD !



Une telle relation peut être décomposée en

$R_1(\underline{K1}, \underline{K2}, X)$  et  $R_2(\underline{Y}, K1)$

23

## Exemple

$CRU, PAYS \twoheadrightarrow REGION$

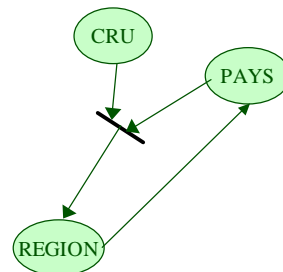
$REGION \twoheadrightarrow PAYS$

**R (CRU, PAYS, REGION)**

**Décomposition BCNF :**

**R1(REGION, PAYS)**

**R2(CRU, REGION)**



24

## *Conclusion*

### ✍ Normalisation utile

- pour obtenir une représentation canonique des entités du monde réel
- pour améliorer la performance de la BD

### ✍ Normalisation automatisable

- Algos 3NF
- outils du commerce (Power Designer, ER-Win, etc.)