

## Chapitre 5

### Normalisation des bases de données

#### Pourquoi la normalisation ?

- Certaines **structures** de bases de données sont meilleures que d'autres
- Un certain nombre de **problèmes** peuvent corrompre la base de données
- Pour éviter ces problèmes et avoir **la structure la plus efficace** on va normaliser la base.
- La normalisation permet de construire des **schémas optimaux** en terme d'intégrité et de cohérence des données stockées
- Les conditions pour avoir une relation (table) correcte peuvent être définies **formellement** : règles de normalisation

#### Exemple

Film			Catégorie	Distributeur	
Titre	Durée	année	Code	Nom	adresse
Alien	137	1982	A	SF Movies	13, Champs Elysées, Paris
Casablanca	102	1942	CD	Constance Film	Gumpendorfsstrass, 17 A-106
Blade Runner	117	1982	SF	Classique Film	2, place Kléber, 67000 Stras
Dances with Wolves	180	1990	W	CK Movies	45, walker street, Houston



Solution avec une seul table

Code	Titre	Durée	année	Nom	adresse
A	Alien	137	1982	CK movies	45, walker street, Houston
SF	Alien	137	1982	CK Movies	45, walker street, Houston
CD	Casablanca	102	1942	Classique Film	2, place Kléber, 67000 Stras
SF	Blade Runner	117	1982	SF Movies	13, Champs Elysées, Paris
W	Dances with Wolves	180	1990	Constance Film	Gumpendorfsstrass, 17 A-106

Redondance => problèmes !!!!

#### Exemple

- **Redondances de données :**
  - Un film est stocké plusieurs fois s'il possède plus qu'une catégorie
  - L'adresse d'un distributeur est stockée plusieurs fois si (1) et/ou s'il existe plusieurs films distribués par ce distributeur
- **Anomalies de modification/insertion :**  
La modification d'un attribut doit éventuellement être effectuée à plusieurs endroits. Possibilité d'inconsistance de données.
- **Anomalies de suppression**  
La suppression d'un film peut provoquer la perte de l'adresse d'un distributeur.

## 1. Décomposition

### et Dépendances Fonctionnelles (DF)

#### Décomposition

- **Objectif** de la décomposition : décomposer la relation  $R_A$  en relations  $R_1, R_2, \dots, R_N$  tel que l'union des  $R_i$  est égale à  $R_A$ .
- Problème : **comment ?**
- La jointure des  $R_i$  doit permettre de récupérer toutes les informations enregistrées dans la table  $R_A$ .

$$R_1 \bowtie R_2 \bowtie R_3 \dots R_N = R_A$$

- Décomposition sans perte : L'égalité est validée pour **toute extension** (ensemble des tuples) de  $R_A$ .

## Décomposition : exemple 1

A		B			C	
Code	titre	Titre	Durée	année	Nom	Adresse
A	Alien	Alien	137	1982	CK Movies	A 45, walker street, Houston
CD	Casablanca	Casablanca	102	1942	Classique Film	CD 2, place Kléber, 67000 Stras
SF	Alien	Blade Runner	117	1982	SF Movies	SF 13, Champs Elysées, Paris
SF	Blade Runner	Dances with Wolves	180	1990	Constance Film	SF 45, walker street, Houston
W	Dances with Wolves					W Gumpendorfsstras, 17A106

A  $\bowtie$  B  $\bowtie$  C Jointure naturelle

Code	Titre	durée	année	Nom	adresse
SF	Alien	137	1982	CK Movies	13, Champs Elysées, Paris
SF	Alien	137	1982	CK Movies	45, walker street, Houston
A	Alien	137	1982	CK Movies	45, walker street, Houston
CD	Casablanca	102	1942	Classique Film	2, place Kléber, 67000 Stras
SF	Blade Runner	117	1982	SF Movies	13, Champs Elysées, Paris
SF	Blade Runner	117	1982	SF Movies	45, walker street, Houston
W	Dances with Wolves	180	1990	Constance Film	Gumpendorfsstras, 17A106

Erreurs !!!

Perte d'informations : l'adresse du distributeur ne dépend pas de la catégorie du film

## Décomposition : exemple 2

A		B			C	
Code	titre	Titre	Durée	année	Nom	Adresse
A	Alien	Alien	137	1982	CK Movies	CK Movies 45, walker street, Houston
CD	Casablanca	Casablanca	102	1942	Classique Film	Classique Film 2, place Kléber, 67000 Strasbourg
SF	Alien	Blade Runner	117	1982	SF Movies	SF Movies 13, Champs Elysées, Paris
SF	Blade Runner	Dances with Wolves	180	1990	Constance Film	Constance Film Gumpendorfsstras, 17A106
W	Dances with Wolves					

A  $\bowtie$  B  $\bowtie$  C Jointure naturelle

Code	Titre	durée	année	Nom	adresse
SF	Alien	137	1982	CK Movies	45, walker street, Houston
A	Alien	137	1982	CK Movies	45, walker street, Houston
CD	Casablanca	102	1942	Classique Film	2, place Kléber, 67000 Stras
SF	Blade Runner	117	1982	SF Movies	13, Champs Elysées, Paris
W	Dances with Wolves	180	1990	Constance Film	Gumpendorfsstras, 17A106

La jointure permet de récupérer l'information sans perte

## Dépendances fonctionnelles (DF)

- Définition : Soit  $R_1$  et  $R_2$  deux sous-ensembles de  $R_A$ , on dit que  $R_1$  détermine fonctionnellement  $R_2$  (noté  $R_1 \rightarrow R_2$ ) si, à un tuple de  $R_1$ , correspond au plus un tuple de  $R_2$ .
- La DF permet de redéfinir la notion de clé : un sous-ensemble  $R_1$  est clé de  $R_A$  si et seulement si  $R_1$  détermine fonctionnellement toutes les colonnes de  $R_A$  et que  $R_1$  est minimale.
- Les dépendances fonctionnelles caractérisent un schéma relationnel. Elles sont indépendantes de ses extensions possibles.

## Exemple de DF

- Quelques DF de notre exemple (en supposant que le titre d'un film est la clé unique, ainsi que le nom d'un distributeur)
  - titre  $\rightarrow$  durée
  - nom  $\rightarrow$  adresse
  - titre, nom  $\rightarrow$  adresse
  - titre  $\rightarrow$  nom
  - titre  $\rightarrow$  adresse

Code	Titre	durée	année	Nom	adresse
SF	Alien	137	1982	CK Movies	45, walker street, Houston
A	Alien	137	1982	CK Movies	45, walker street, Houston
CD	Casablanca	102	1942	Classique Film	2, place Kléber, 67000 Stras
SF	Blade Runner	117	1982	SF Movies	13, Champs Elysées, Paris
W	Dances with Wolves	180	1990	Constance Film	Gumpendorfsstras, 17A106

## Propriétés des DF

- Axiomes d'Armstrong
  - Réflexivité
    - Si  $b$  est un sous-ensemble de  $a$  alors  $a \rightarrow b$
  - Augmentation
    - Si  $a \rightarrow b$  est une DF, alors  $a, c \rightarrow b, c$  est une DF
  - Transitivité
    - Si  $a \rightarrow b$  et  $b \rightarrow c$  sont des DF, alors  $a \rightarrow c$  est une DF
- Règles déduites des axiomes
  - Union
    - Si  $a \rightarrow b$  et  $a \rightarrow c$  sont des DF, alors  $a \rightarrow b, c$  est une DF
  - Pseudo-transitivité
    - Si  $a \rightarrow b$  et  $b, c \rightarrow d$  sont des DF, alors  $a, c \rightarrow d$  est une DF
  - Décomposition
    - Si  $a \rightarrow b, c$  est une DF, alors  $a \rightarrow b$  et  $a \rightarrow c$  sont des DF

## Dépendances élémentaires

- Une dépendance est élémentaire si  $a, b \rightarrow c$  est une DF et que ni  $a \rightarrow c$  ni  $b \rightarrow c$  ne sont des DF
- Exemple

numFilm	titre	pays	duree
1	Casablanca	USA	102
3	Terminator	USA	108
4	Die Hard	USA	131
6	Les bronzés	F	87
7	Le père Noël est une ordure	F	88
14	Blade Runner	USA	117
15	Alien	USA	117

- $\text{numFilm} \rightarrow \text{titre}$  : DF élémentaire
- $\text{numFilm}, \text{pays} \rightarrow \text{titre}$  : DF Pas élémentaire (pays inutile)
- $\text{numFilm}, \text{duree} \rightarrow \text{titre}$  : DF Pas élémentaire (duree inutile)

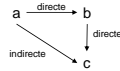
## Dépendance directe

- Une DF  $a \rightarrow c$  est directe si elle n'est pas déduite par transitivité, c'est-à-dire s'il n'existe pas de DF  $a \rightarrow b$  et  $b \rightarrow c$

- Exemple

numFilm	numDistrib	nom
15	2	SF Movies
1	1	Constance Film
6	3	Classique Film

- $\text{numFilm} \rightarrow \text{numDistributeur}$  (DF directe)
- $\text{numDistributeur} \rightarrow \text{nom}$  (DF directe)
- $\text{numFilm} \rightarrow \text{nom}$  (pas une DF directe)



## Fermeture transitive

- C'est l'ensemble de toutes les DF que l'on peut déduire par transitivité à partir d'un ensemble F de DFE donné

- Exemple

- $F = \{ NV \rightarrow \text{TYPE}, \text{TYPE} \rightarrow \text{MARQUE}, \text{TYPE} \rightarrow \text{PUISSANCE}, NV \rightarrow \text{COULEUR} \}$
- $F^+ = F \cup \{ NV \rightarrow \text{MARQUE}, NV \rightarrow \text{PUISSANCE} \}$

- Equivalence de deux ensembles de DFE : Deux ensembles de DFE sont équivalents s'ils ont la même fermeture transitive.

## La couverture minimale

- La couverture minimale est un **sous ensemble minimal** à partir duquel **toutes les DF** peuvent être reproduites (en utilisant les axiomes d'Armstrong).
- C'est un ensemble de DF élémentaires tel que :
  - Aucune DF n'est redondante
  - Toutes les DFE sont dans la fermeture transitive de la couverture minimale
  - Aucune DF ne peut être supprimée
  - Le membre droit (dépendant) de chaque DF ne contient qu'un seul attribut
  - Le membre gauche (déterminant) de chaque DF est irréductible

## Exemple de Couverture minimale

- Ensemble de DF non minimal :
  - $\text{numFilm} \rightarrow \text{titre}$
  - $\text{numFilm} \rightarrow \text{duree}$
  - $\text{numFilm} \rightarrow \text{numDistributeur}$
  - $\text{numDistributeur} \rightarrow \text{nomDistributeur}$
  - $\text{numFilm} \rightarrow \text{nomDistributeur}$
  - $\text{numFilm} \rightarrow \text{adresseDistributeur}$
  - $\text{numDistributeur} \rightarrow \text{adresseDistributeur}$
- Une couverture minimale :
  - $\text{numFilm} \rightarrow \text{titre}$
  - $\text{numFilm} \rightarrow \text{duree}$
  - $\text{numFilm} \rightarrow \text{numDistributeur}$
  - $\text{numDistributeur} \rightarrow \text{nomDistributeur}$
  - $\text{numDistributeur} \rightarrow \text{adresseDistributeur}$

Redondants ???

## Définition d'une clé

- Une clé (primaire ou candidate) est un sous-ensemble X des attributs d'une relation R ( $A_1, A_2, \dots, A_n$ ) tel que :
  - $X \rightarrow A_1, A_2, \dots, A_n$
  - Il n'existe pas de sous-ensemble Y de X ( $Y \subset X$ ) tel que  $Y \rightarrow A_1, A_2, \dots, A_n$  (clé minimale)

## 2. Formes normales

(= différents degrés de normalisation)

## Les formes normales

- Les formes normales imposent des conditions sur les schémas.
- Objectif : formaliser et faciliter la création des schémas sans redondances et sans anomalies de mise à jour.
- => Permettre la décomposition de relations sans perte d'informations, à partir de la notion de DF.
- Règle générale : un schéma relationnel est dit bien formé si chaque relation contient de l'information sur un et un seul type d'entité



Chaque forme normale de rang  $i$  est plus restrictive que la forme normale de rang  $i-1$ .

## Décomposition des relations

- La théorie de la normalisation repose sur un principe de décomposition des relations.
- Décomposition d'une relation** : La décomposition d'un schéma de relation  $R(A_1, A_2, \dots, A_n)$  est son remplacement par une collection de schémas de relations  $(R_1, R_2, \dots, R_i)$  telle que :  $SCHEMA(R) = SCHEMA(R_1) \cup SCHEMA(R_2) \cup \dots \cup SCHEMA(R_i)$ .
- Décomposition sans perte** : Une décomposition d'une relation  $R$  en  $N$  relations  $R_1, R_2, \dots, R_N$  est sans perte si et seulement si, pour toute extension de  $R$ , on a :  

$$R = R_1 \bowtie R_2 \bowtie \dots \bowtie R_N$$

## Décomposition des relations

- Théorème de décomposition sans perte** : Une décomposition en 2 relations est sans perte si l'attribut de jointure de la recombinaison est clé d'une au moins des 2 relations.  
 ➔ On retrouve toute l'information par jointure.
- Décomposition préservant les dépendances fonctionnelles** : Une décomposition  $(R_1, R_2, \dots, R_N)$  de  $R$  préserve les dépendances fonctionnelles si la **fermeture des dépendances fonctionnelles de  $R$**  est la **même que celle de l'union** des dépendances fonctionnelles des relations  $R_1, R_2, \dots, R_N$ .  
 ➔ Autrement dit, une décomposition est dite « sans perte de DF » si on peut retrouver logiquement (par ex. par transitivité) les DF de départ à partir de la projection de l'ensemble des DF de départ sur les relations de la décomposition.

## 1NF : 1ère forme normale

- Une relation est en 1ère forme normale si :**
  - tout attribut est en dépendance fonctionnelle avec la clé
  - tout attribut contient une seule valeur
- Objectif : simplicité, esthétique
- La première forme normale a été supposée comme hypothèse dans le cadre de ce cours depuis l'introduction du modèle relationnel
- Relation ne respectant pas la 1NF :

Titre	Acteurs
Casablanca	Humphrey Bogart, Ingrid Bergman
Perfect World	Kevyn Collier, Cliff Fawcett
The Terminator	Arnold Schwarzenegger, Linda Hamilton, Michael Biehn
The Hard	Bruce Willis

## 1NF : résolution

- Création de plusieurs attributs :  
 Location (numFilm, numClient, interval)  
 ↓  
 Location (numFilm, numClient, dateDe, dateA)
- Création d'une nouvelle relation :  
 Film (titre, acteurs)  
 ↓  
 Film (numF, titre)  
 Acteurs (numF, acteur)

## 2NF : 2ème forme normale

- Une relation est en 2ème forme normale si et seulement si :**
  - elle est en 1ère forme normale
  - tout attribut n'appartenant pas à une clé ne dépend pas d'une partie d'une clé (DF élémentaire).
- Objectifs :
  - Elimination de redondances
  - Garantir que tout attribut dépend que de la clé et non pas d'une partie de la clé

## Exemple 2NF

- Exemple d'une relation violant la 2NF:

Realisation (numFilm, numArtiste, titre, salaire)

NumFilm	NumArtiste	Titre	Salaire
1	1	Alien	100000
1	3	Alien	10000
1	10	Alien	30000
2	2	Casablanca	20000
3	5	Blade Runner	40000
3	4	Blade Runner	40000
4	4	Dances with Wolves	70000

- Couverture minimale
  - $\text{numFilm}, \text{numArtiste} \rightarrow \text{salaire}$
  - $\text{numFilm} \rightarrow \text{titre}$

## 2NF : résolution

- Règle de décomposition pour passer en 2NF :

- $R(\underline{K1}, \underline{K2}, X, Y)$  n'est pas en 2NF :

$$K1, K2 \rightarrow X; K2 \rightarrow Y$$

- On décompose en deux relations en 2NF :

$$R1(\underline{K1}, \underline{K2}, X) \text{ et } R2(\underline{K2}, Y)$$

- Exemple :

Realisation (numFilm, numArtiste, titre, salaire)



Realisation (numFilm, numArtiste, salaire)

Film (numFilm, titre)

(Ces relations sont en 2NF)

## 3NF : 3<sup>ème</sup> forme normale

- Une relation est en 3<sup>ème</sup> forme normale si et seulement si:

- elle est en 2<sup>ème</sup> forme normale
- tout attribut n'appartenant pas à une clé ne dépend pas d'un autre attribut non clé (DF directe)

- Objectifs :
  - Elimination des redondances.
  - Garantir que seulement la clé détermine tous les attributs d'une relation, donc que les attributs font partie d'un seul « type ».
- Toute relation a au moins une décomposition en 3NF telle que :
  - La décomposition est sans perte
  - La décomposition préserve les DF

## Exemple 3NF

- Exemple d'une relation violant la 3NF :

Film (numFilm, titre, numDistributeur, nomDistributeur)

numFilm	titre	numDistributeur	nomDistributeur
1	Casablanca	3	Classique Film
2	Perfect World	2	Conscience Film
3	The Terminator	1	SF Movies
4	Die Hard	2	Conscience Film
5	Speed	2	Conscience Film
6	Les Bronzés	3	Classique Film

- Couverture minimale :
  - $\text{numFilm} \rightarrow \text{titre}$
  - $\text{numFilm} \rightarrow \text{numDistributeur}$
  - $\text{numDistributeur} \rightarrow \text{nomDistributeur}$

## 3NF : résolution

- Règle de décomposition pour passer en 3NF :

- $R(\underline{K}, X, Y, Z)$  n'est pas en 3NF :  $K \rightarrow X, Y, Z; X \rightarrow Z$

- On décompose en deux relations en 3NF :

$$R1(\underline{K}, X, Y) \text{ et } R2(\underline{X}, Z)$$

- Exemple :

Film (numFilm, titre, numDistributeur, nomDistributeur)



Film (numFilm, titre, numDistributeur)

Distributeur (numDistributeur, nomDistributeur)

## Algorithme de synthèse en 3NF

- Principe :

- En entrée : ensemble des attributs et des DF
- Construire une couverture minimale des DFE
- La couverture F est partitionnée en groupes  $F_i$  tels que les DF dans chaque  $F_i$  ont le même ensemble d'attributs à gauche
- Chaque groupe  $F_i$  produit une relation en 3NF

- Exemple : synthèse à partir du graphe de la couverture minimale des DFE pour la relation :

PROPRIETAIRE (NV, MARQUE, TYPE, PUISS, COUL, NSS, NOM, PRENOM, DATE, PRIX)

## Forme normale de Boyce-Codd (BCNF)

- Exemple introductif :  
Vins(Cru, Pays, Région) avec les dépendances fonctionnelles supposées :
  - (Cru, Pays) → Région
  - Région → Pays

Cru	Pays	Région
Chenas	France	Beaujolais
Julienas	France	Beaujolais
Morgon	France	Beaujolais
Brouilly	France	Beaujolais
Chablis	Etats-Unis	Californie

- Est elle en 3NF ? Y a-t-il des redondances ? => BCNF

## Forme normale de Boyce-Codd (BCNF)

- Une relation est en BCNF si et seulement si** les seules DF élémentaires sont celles dans lesquelles une clé ou superclé(\*) détermine un attribut.  
(\*) une superclé est un ensemble d'attributs contenant une clé
- Objectifs : Eliminer les redondances créées par des dépendances:
  - de parties de clés entre elles
  - d'attribut non-clé vers une partie de clé
- Toute relation a une décomposition en BCNF sans perte

## BCNF : résolution

- Règle de décomposition pour passer en BCNF :
  - R (K1, K2, X, Y) n'est pas en BCNF :  
K1, K2 → X; K1, K2 → Y; Y → K2
  - On décompose en deux relations en BCNF :  
R1 (K1, Y, X) et R2 (Y, K2)
  - Perte des DF : K1, K2 → X et K1, K2 → Y
- Exemple : La relation Vins pourra être décomposée en deux relations :  
Cru (Cru, Région)  
Régions (Région, Pays)
- La dépendance fonctionnelle (Cru, Pays) → Région est perdue mais elle peut être recomposée par jointure.

## Encore des redondances...

- La décomposition en forme normale BCNF est toutefois insuffisante pour éliminer les redondances et les anomalies de mises à jour. Exemple en 3NF :

Restaurant	Pizza Variety	Delivery Area
Vincenzo's Pizza	Thick Crust	Springfield
Vincenzo's Pizza	Thick Crust	Shelbyville
Vincenzo's Pizza	Thin Crust	Springfield
Vincenzo's Pizza	Thin Crust	Shelbyville
Elite Pizza	Thin Crust	Capital City
Elite Pizza	Stuffed Crust	Capital City
Al Pizza	Thick Crust	Springfield
Al Pizza	Thick Crust	Shelbyville
Al Pizza	Thick Crust	Capital City
Al Pizza	Stuffed Crust	Springfield
Al Pizza	Stuffed Crust	Shelbyville
Al Pizza	Stuffed Crust	Capital City

- Un restaurant délivre **toutes ses variétés** de pizzas **dans toutes ses zones** de livraison
- Indépendance** entre variétés de pizzas d'un restaurant et zones de livraison du restaurant => Redondances...

## Dépendances multivaluées (DM)

- Pour approfondir la décomposition, on introduit la notion de dépendance multivaluée.
- Soit R(A1, A2, ..., An) un schéma de relation et X et Y des sous-ensembles de A1, A2, ..., An.
- On dit que X multi-détermine Y (X →> Y) si, étant données des valeurs de X, il y a un ensemble de valeurs de Y associées et cet ensemble est indépendant des autres attributs de la relation R.
- Plus formellement on a :  
 $(X \rightarrow\!\!\!\rightarrow Y) \Leftrightarrow \{ \forall (x,y,z) \text{ et } (x,y',z') \in R \Rightarrow (x,y',z) \text{ et } (x,y,z') \in R \}$   
où x,y,z,y',z' sont des occurrences des attributs X,Y,Z de la relation R.
- Une DM caractérise une indépendance entre deux ensembles d'attributs (Y et Z) corrélés par un même X.

## Dépendances multivaluées (DM)

- Propriété** : Les Dépendances fonctionnelles sont des cas particuliers de DM  
 $(X \rightarrow Y) \Rightarrow \{ \forall (x,y,z) \text{ et } (x,y',z') \in R \Rightarrow y=y' \}$   
d'où :  $(X \rightarrow Y) \Rightarrow \{ \forall (x,y,z) \text{ et } (x,y',z') \in R \Rightarrow (x,y',z) \text{ et } (x,y,z') \in R \}$   
donc :  $(X \rightarrow Y) \Rightarrow (X \rightarrow\!\!\!\rightarrow Y)$
- Axiomes d'inférence des DM** :
  - Complementation :  $(X \rightarrow\!\!\!\rightarrow Y) \Rightarrow (X \rightarrow\!\!\!\rightarrow X \cup Y)$
  - Augmentation :  $(X \rightarrow\!\!\!\rightarrow Y) \text{ et } (V \subseteq W) \Rightarrow (XW \rightarrow\!\!\!\rightarrow YV)$
  - Transitivité :  $(X \rightarrow\!\!\!\rightarrow Y) \text{ et } (Y \rightarrow\!\!\!\rightarrow Z) \Rightarrow (X \rightarrow\!\!\!\rightarrow Z)$
- Propriété** : Une dépendance multivaluée est **élémentaire** (DME) si c'est une dépendance multivaluée  $X \rightarrow\!\!\!\rightarrow Y$  d'une relation R telle que :
  - Y n'est pas vide et est disjoint de X
  - R ne contient pas une autre DM du type  $X' \rightarrow\!\!\!\rightarrow Y'$  telle que  $X' \subset X$  et  $Y' \subset Y$

## Quatrième forme normale (4NF)

- **Définition :**  
**Une relation est en 4NF si et seulement si les seules DME sont celles où une superclé détermine un attribut**

(Superclé = ensemble d'attributs contenant une clé)

- Objectif : décomposer les relations ayant des DME
- Exemple : la relation PIZZAS (RESTAURANT, TYPE\_PIZZA, ZONE\_LIVRAISON) n'est pas en 4NF
  - RESTAURANT ->> TYPE\_PIZZA
  - RESTAURANT ->> ZONE\_LIVRAISON

## Quatrième forme normale (4NF)

- **Propriété :** Toute relation a une décomposition en 4NF sans perte.
  - Ex : (RESTAURANT, TYPE\_PIZZA) et (RESTAURANT, ZONE\_LIVRAISON) pour la relation PIZZAS
- **Remarque :** R n'est pas en 4NF s'il existe une DM  $X \twoheadrightarrow Y$  où X n'inclut pas une clé de R
  - ➔ une DF étant un cas particulier de DM, il en résulte qu'une **relation en 4NF est en BCNF et donc en 3NF**.

## Dépendances de jointure

- La 4NF n'élimine pas toutes les redondances :
  - Exemple : relation CONSOMMATION

PERSONNE CRUSTACES VILLE

Cloé	Tourteau	Binic
Cloé	Tourteau	Trouville
Cloé	Langouste	Trouville
Jean	Tourteau	Trouville
Jean	Tourteau	Binic

- Relation en 4NF ?
- Redondances ?

## Dépendances de jointure

- Redondance car "Toute personne ayant consommé un type de crustacés et ayant séjourné dans une ville les cultivant a consommé de ce type de crustacés dans cette ville."  
 $(\text{personne, crustacés}) \in C1$  et  $(\text{personne, ville}) \in C2$  et  $(\text{crustacés, ville}) \in C3$   
 $\Rightarrow (\text{personne, crustacés, ville}) \in \text{CONSOMMATION}$
- Le problème vient du fait que jusqu'alors nous avons toujours essayé de **décomposer une relation en 2 relations**. Hors, comme nous allons le montrer, certaines relations sont décomposables, non pas en 2, mais **en 3 relations**.

## Dépendances de jointure

- Soit  $R(A_1, A_2, \dots, A_n)$  un schéma de relation et  $X_1, X_2, \dots, X_m$  des sous-ensembles de  $(A_1, A_2, \dots, A_n)$ .  
On dit qu'il existe une **dépendance de jointure**  $*(X_1, X_2, \dots, X_m)$  si R est la jointure de ses projections sur  $X_1, X_2, \dots, X_m$ ,  
c'est-à-dire si  $R = \rho_{X_1(R)} \bowtie \rho_{X_2(R)} \bowtie \dots \bowtie \rho_{X_m(R)}$ .
- Par exemple, la relation Consommation obéit à la dépendance de jointure suivante :  $*(\text{PERSONNE\_CRUSTACES}, \text{PERSONNE\_VILLE}, \text{CRUSTACES\_VILLE})$
- **Propriété :** Les **dépendances multivaluées sont des cas particuliers de dépendances de jointure** :  
Une relation  $R(X, Y, Z)$  vérifiant la DM :  $X \twoheadrightarrow Y$  (et donc  $X \twoheadrightarrow Z$ ) satisfait la DJ  $*(\{XY\}, \{XZ\})$  (ex: base crustacés)

## Cinquième forme normale (5NF) ou forme normale de projection-jointure

- Soit une relation R et  $\{R_1, R_2, \dots, R_p\}$  une dépendance de jointure. On dit qu'une telle DJ est **triviale** si l'une des relations  $R_i$  est la relation R elle-même.
- **Définition :**  
**Une relation R est en 5NF si et seulement si toute dépendance de jointure est triviale ou tout  $R_i$  est une superclé de R (i.e. chaque  $R_i$  contient une clé de R).**
- Objectif : Eliminer des redondances dans une relation en 4NF, grâce à la notion de dépendance de jointure.
- La 5NF est une généralisation de la 4NF qui nécessite de prendre en compte les dépendances de jointure induites par la connaissance des clés d'une relation.

## Cinquième forme normale (5NF)

- **Propriété** : La décomposition d'une relation en 5NF suit les DJ et est sans perte de données.
- Mais elle ne préserve en général pas les DF (comme la BCNF).
- Exemple : relation Consommation (Personne, Crustaces, Ville)
  - Cette relation n'est pas en 5NF car la DJ :  
 $\{ (Personne, Crustaces), (Personne, Ville), (Crustaces, Ville) \}$   
n'est pas triviale, et aucun des Ri n'est une superclé :
    - la seule superclé est (Personne, Crustaces, Ville)
    - On la décompose donc en ces trois relations

## Cinquième forme normale (5NF)

- **Propriété** : Toute relation en 5NF ne peut plus être décomposée par projection et recomposée par jointure sans perte d'informations. C'est l'ultime étape de la décomposition par projection-jointure, d'où son nom de « **forme normale de projection-jointure** » (JDNF).

## Pertes de données

- Définition :
  - Une décomposition en plusieurs relations est dite « sans perte de données » si la jointure naturelle des relations obtenues dans la décomposition redonne la relation de départ
- Une décomposition correcte est sans perte de données
- Exemple de décomposition avec pertes :  
Relation EMP\_DEPT (NumE, NumD, NomE, Salaire, NomD)  
 $NumE \rightarrow NomE; NumD \rightarrow NomD; NumE \rightarrow NumD$   
Peut-on décomposer dans les deux relations suivantes :  
EMP(NumE, NomE, Salaire) et DEPT(NumD, NomD) ?

## Pertes de données

- Pour décomposer une relation sans perte de données en plusieurs relations qui sont dans une Xe NF ( $2 \leq X \leq 5$ ), on extrait les DF qui font que la relation de départ n'est pas en XNF (voir « Théorème de décomposition sans perte » en début de sous section 2).
- Dans l'exemple précédent, la DF  $NumD \rightarrow NomD$  doit être extraite et donne lieu à la création d'une nouvelle relation :  
EMP(NumE, NomE, Salaire, NumD)  
DEPT(NumD, NomD)

## Perte de dépendances

- Définition :  
Soit une relation R qui est décomposée en deux relations R1 et R2. La projection de l'ensemble des DF de R sur la relation R1 est composée des DF de R dont les attributs (source et but) sont des attributs de R1.  
Une décomposition est dite « sans perte de DF » si on peut retrouver logiquement (par ex. par transitivité) les DF de départ à partir de la projection de l'ensemble des DF de départ sur les relations de la décomposition
- Si on décompose avec perte de DF, on ne peut travailler indépendamment dans chacune des relations de la décomposition. Lors des mises à jour il faudra vérifier que les DF perdues sont respectées en consultant les relations concernées.

## Conclusion

- La perte de DF est moins grave que la perte de données
  - On peut décomposer jusqu'en 3NF sans perte de DF
  - A partir de la BCNF il n'est pas toujours possible de conserver les DF
- Une normalisation trop forte d'une base peut entraîner des pertes d'informations. Il faut faire un compromis entre intégrité des données de la base et respect d'une forme normale.