

Base de Données Structurées et Non Structurées

Partie 5: Normalisation (5 formes de normalisation)



Prof. Badr-Eddine BOUDRIKI SEMLALI
Email: badreddine.boudrikisemlali@uae.ac.ma

2024/2025

Plan de cours

1. Introduction aux bases de données et méthode de conception MERISE
2. Modélisation des traitements : MCC, MCT, MOT
3. Modélisation des données: (MCD) → modèle « E/A »
4. Dépendance fonctionnelles (Construction du MCD)
5. Normalisation (5 formes de normalisation)
6. Le Modèle Logique de Données (relationnel) : MLD
7. Le Modèle Physique de Données : MPD
8. Le Modèle relationnel: Concepts de bases
9. Structure de base de données
10. Règles d'intégrité structurelle
11. Algèbre relationnel

Merise: Modélisation des Données

Règles relatives au MCD (Normalisation du modèle)

Normalisation des entités

Les entités du MCD Doivent Vérifier les règles suivantes:

Première Normalisation :

- Première Forme Normale (1FN)
- Deuxième Forme Normale (2FN)
- Troisième Forme Normale (3FN)

Deuxième Normalisation :

- Dépendance multi-valuée et quatrième forme normale (4FN)
- Dépendance de jointure et cinquième forme normale (5FN)

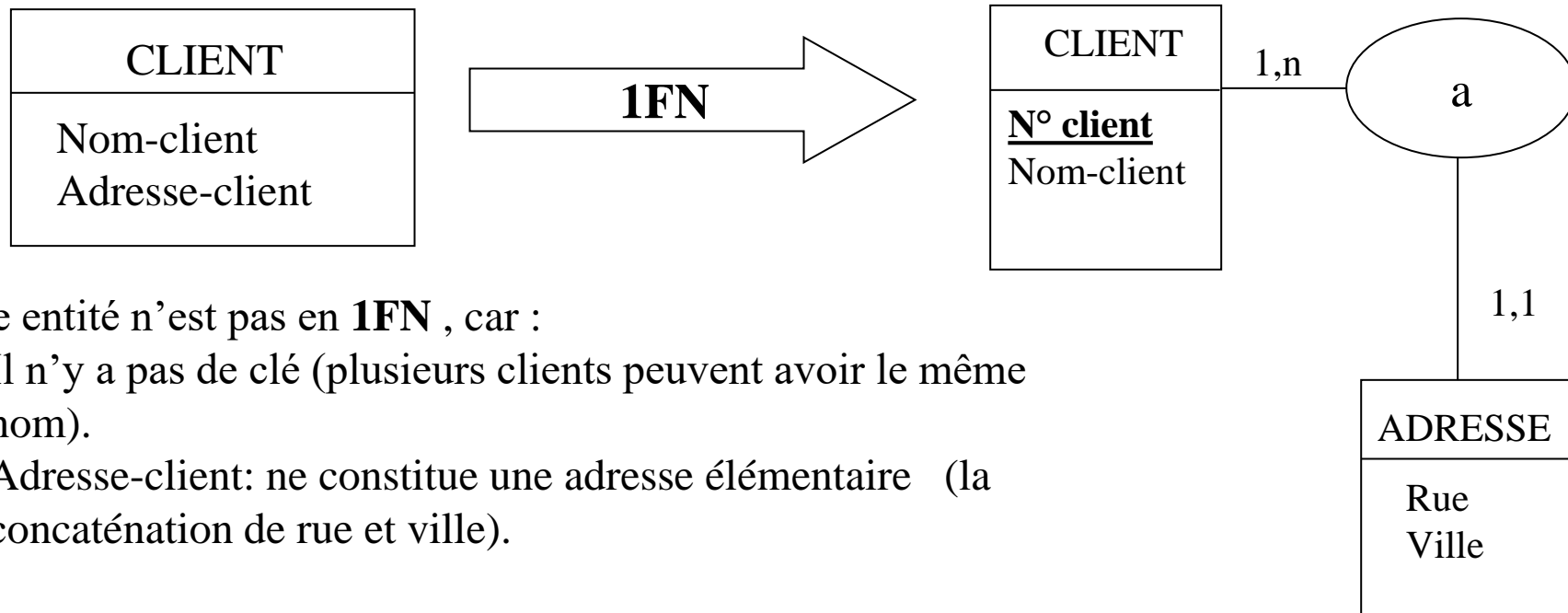
Remarque: Les formes normales sont construites selon le principe des dépendances fonctionnelles.

Merise: Modélisation des Données (FN1)

1^{ère} Forme Normale (1^{ère} FN)

Une entité (ou association) est en 1^{ère} forme normale si, d'une part, elle possède un identifiant, et si, d'autre part, toutes ses propriétés sont élémentaires (non décomposable).

Exemple:



Cette entité n'est pas en **1FN**, car :

- Il n'y a pas de clé (plusieurs clients peuvent avoir le même nom).
- Adresse-client: ne constitue une adresse élémentaire (la concaténation de rue et ville).

Merise: Modélisation des Données (FN2)

2^{ème} Forme Normale (2^{ème} FN)

Une propriétés (ou une association) est en 2^{ème} forme normale si, était déjà en 1^{ère} forme normale, chacune des propriétés (ne fait pas partie de l'identifiant) est en dépendance fonctionnelle élémentaire avec l'identifiant.

Autrement dit: à une valeur de l'identifiant, il ne doit correspondre qu'une valeur de chaque propriété.

Exemple:

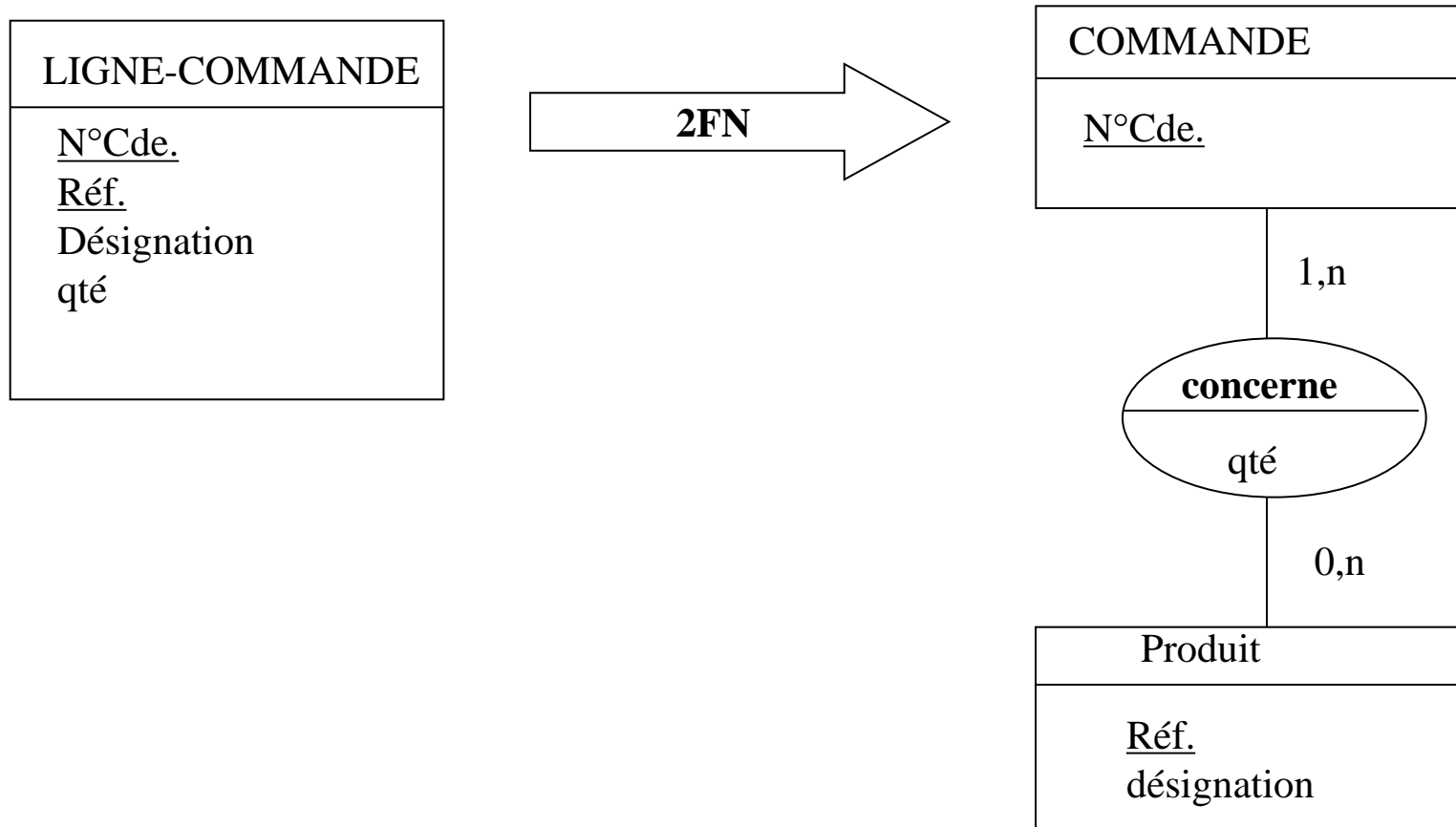
LIGNE-COMMANDE
<u>N°Cde.</u> <u>Réf.</u> Désignation qté

La clé est la concaténation N° Cde+réf.

Mais la DF N° cde+réf → Désignation n'est pas élémentaire
puisque: réf. → désignation.

Merise: Modélisation des Données (FN2)

2^{ème} Forme Normale (2^{ème} FN) (suite)

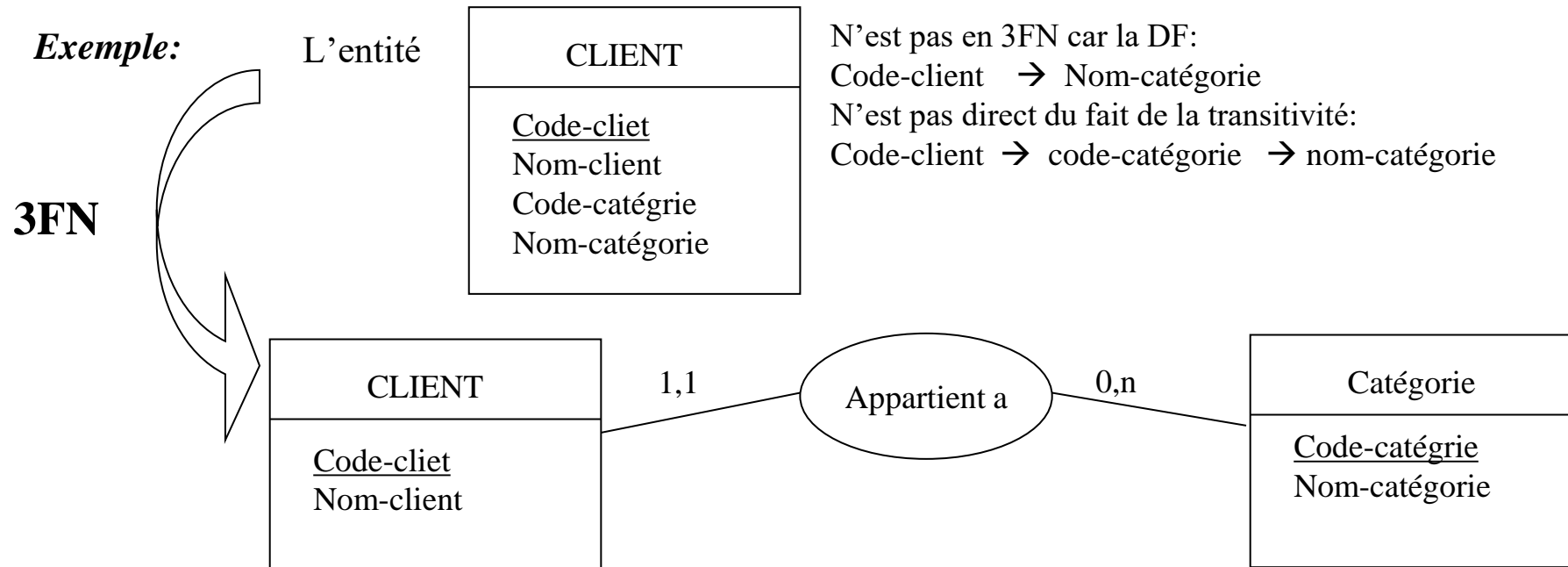


Merise: Modélisation des Données (FN3)

3^{ème} Forme Normale (3^{ème} FN)

Une entité (ou une association) est en 3^{ème} forme normale si, étant déjà en 2^{ème} FN, chacune des propriétés (ne fait pas partie de l'identifiant) est en dépendance fonctionnelle directe avec l'identifiant.

Autrement dit: les propriétés ne doivent pas dépendre d'une propriété autre que l'identifiant. Elles ne doivent pas dépendre, non plus, d'une partie de l'identifiant.



Merise: Modélisation des Données (FN3)

3^{ème} Forme Normale (3^{ème} FN)

Définition

Une relation $R(X, Y, Z)$ est dite décomposable en deux :

$$R_1 = R[X, Y] \text{ et } R_2 = R[Y, Z]$$

$$\text{si on a: } R = R_1 * R_2$$

(i.e. R est la jointure naturelle de R_1 et R_2)

Décompositions :

$R[X, Y]$ dénote la projection de R sur X et Y .

- On remplace R par R_1 et R_2
- Aucune information n'est perdue: On peut retrouver R par jointure (naturelle) de R_1 et R_2

Merise: Modélisation des Données (FN3)

3^{ème} Forme Normale (3^{ème} FN)

Proposition Soit $R (X, Y, Z)$

Pour que R soit décomposable en $R1 (X, Y)$ et $R2 (Y, Z)$,

il suffit que dans R , on ait

$Y \rightarrow X$ ou $Y \rightarrow Z$

i.e. Si $Y \rightarrow X$

ou $Y \rightarrow Z$

alors $R = R1 * R2$

Il suffit que l'attribut commun soit clé dans l'une des relations (ou les deux).

Le passage de la 1FN vers la 2FN ainsi que

Le passage de la 2FN vers la 3FN

sont des décompositions sans perte.

Merise: Modélisation des Données (BCNF)

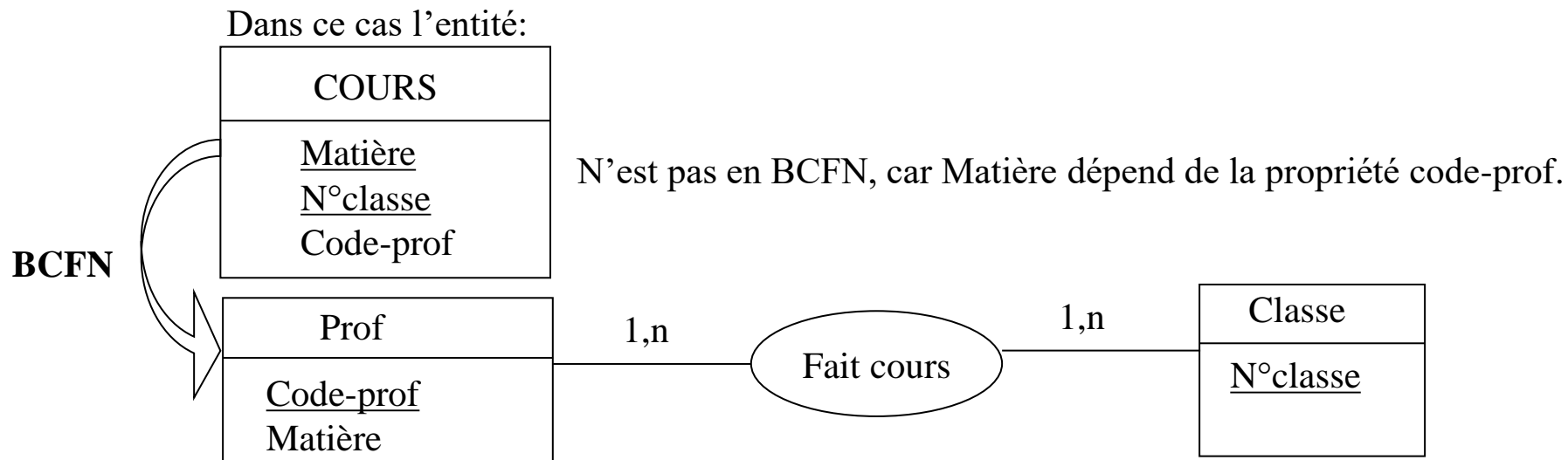
Forme normale de Boyce-Codd (BCNF)

- Le modèle est en 3^{ème} FN ;
- Les seules dépendances fonctionnelles élémentaires sont celles dans lesquelles un identifiant détermine une propriété.
 - Pour les identifiants composés de plusieurs propriétés, ces dernières ne doivent pas être dépendantes d'une autre propriété de l'entité (pour éviter les cycles de DF).

Admettant la règle de gestion:

Exemple:

Règle de gestion: tout professeur enseigne une matière et une seule.
Toute classe n'a qu'un seul Professeur par matière.



Merise: Modélisation des Données

Les formes normales (1^{ère} FN ; 2^{ème} FN ; 3^{ème} FN): Synthèse

Les normalisations ci-dessus ont pour but d'éliminer:

- **Les redondances** (inutile de répéter la désignation du produit commandé chaque commande d'un même produit).
- **Les anomalies de mise à jour** (si on annule un client on veut sans doute toutefois conserver la catégorie de ce client).

Modèle normalisé = relations avec

- une clé, qui permet de distinguer chaque occurrence ;
- des attributs élémentaires (1^{ère} FN) ;
- en dépendance de TOUTE la clé (2^{ème} FN) ;
- et RIEN QUE de la clé (3^{ème} FN).

On parle aussi de normalisation pour un MCD. Un MCD qui donne un MCD normalisé est qualifié aussi de normalisé.

Acquérir La troisième forme normale est généralement reconnue comme la plus importante à respecter.

Merise: Modélisation des Données (FN4)

4^{ème} Forme Normale (4^{ème} FN)

- ✍ Pour être en 4FN, il faut respecter la 3^{ème} FN (et non nécessairement la FNBC).
- ✍ Pour toute relation de dimension n (avec n supérieur à 1) en forme normale de Boyce-Codd, les relations de dimension $n - 1$ construites sur sa collection doivent avoir un sens. Il ne doit pas être possible de reconstituer les occurrences de la relation de dimension n par jointure de deux relations de dimension $n - 1$.
- ✍ **Cette normalisation conduit parfois à décomposer une relation complexe en deux relations plus simples.**

Merise: Modélisation des Données

4^{ème} Forme Normale (4^{ème} FN)

Cette normalisation conduit parfois à décomposer une relation complexe en deux relations plus simples.

✍ Exemple :

PSL

<u>PERSONNE</u>	<u>SPECIALITE</u>	<u>LANGUE</u>
Ali	Cinéma	Français
Ali	Cinéma	Anglais
Ali	Cinéma	Arabe
Ali	Peinture	Français
Ali	Peinture	Anglais
Ali	Peinture	Arabe

Problèmes :

- Adjonction nouvelle spécialité (ou langue) pour Ali. Quel(s) tuple(s) insérer?
- Problème inverse pour la suppression.
- Croissance multiplicative
- Redondances

Merise: Modélisation des Données (FN4)

4^{ème} Forme Normale (4^{ème} FN)

 **Solution:**

PS

<u>PERSONNE</u>	<u>SPECIALITE</u>
Ali	Cinéma
Ali	Peinture

PL

<u>PERSONNE</u>	<u>LANGUE</u>
Ali	Français
Ali	Anglais
Ali	Arabe

Problèmes précédents résolus.

Pas de redondance. Croissance additive.

On dit qu'on est passé à la 4^e Forme Normale (4FN)

Merise: Modélisation des Données (FN4)

4^{ème} Forme Normale (4^{ème} FN)

Dépendances Multivaluées (DMV) et 4FN

Définition :

Soit $R(X, Y, Z)$ un schéma de relation.

On dit qu'il y a une dépendance Multivaluée entre X et Y

(notée $X \twoheadrightarrow Y$)

ssi :

quand les tuples $\langle x, y, z \rangle$ et $\langle x, y', z' \rangle$ apparaissent dans R , alors les tuples $\langle x, y, z' \rangle$ et $\langle x, y', z \rangle$ aussi.

Autrement dit, pour une valeur donnée x de X , les valeurs de Y associées sont indépendantes de celles de Z . x est associée d'une part à y et y' et d'autre part à z et z' , et toutes les combinaisons sont possibles.

X et Y jouant un rôle symétrique, si on a $X \twoheadrightarrow Y$ alors on a aussi $X \twoheadrightarrow Z$.

On note

$$X \twoheadrightarrow Y \mid Z$$

Merise: Modélisation des Données (FN4)

Dépendances Multivaluées (DMV) et 4FN

Dans PSL on a : PERSONNE \twoheadrightarrow SPECIALITE

et aussi PERSONNE \twoheadrightarrow LANGUE

i.e.

PERSONNE \twoheadrightarrow SPECIALITE | LANGUE

Cette DMV exprime que la relation entre une personne et la langue parlée est n'a rien à voir avec sa spécialité.

Si

<Ali, Cinéma, Arabe> et <Ali, Peinture, Anglais> sont vrais,

alors

<Ali, Peinture, Arabe> et <Ali, Cinéma, Anglais> sont vrais aussi.

Remarque Importante : Soit R (X, Y, Z) un schéma de relation

Si $X \rightarrow Y$ Alors $X \twoheadrightarrow Y$

Une DF est un cas particulier d'une DMV.

Merise: Modélisation des Données (FN4)

Dépendances Multivaluées (DMV) et 4FN

Définition : Une relation R est en 4FN ssi:

Toute Dépendance Multivaluée de R est la conséquence d'une clé (candidate).

Chaque fois qu'on a $X \twoheadrightarrow Y$ dans R , X est (ou contient) une clé.
i.e. les DMVs que R contient sont en réalité des DFs.

En l'occurrence, les DFs que la clé détermine tout autre attribut.

Exemple : PL et PS sont en 4FN.

PSL n'est pas en 4FN.

PL et PS sont des projections de PSL.

Merise: Modélisation des Données (FN4)

Dépendances Multivaluées (DMV) et 4FN

Proposition :

Une relation $R (X, Y, Z)$ est décomposable en $R1 (X, Y)$ et $R2 (X, Z)$,
ssi: $X \twoheadrightarrow Y | Z$

Noter le si et seulement si :

$$R = R1 * R2 \Rightarrow X \twoheadrightarrow Y | Z \text{ et}$$

$$R = R1 * R2 \Leftarrow X \twoheadrightarrow Y | Z$$

Dans notre cas (exemple)on a:

$$\text{PERSONNE} \twoheadrightarrow \text{SPECIALITE} | \text{LANGUE}$$

donc (et inversement),

$$\text{PSL} = \text{PS} * \text{PL}$$

Autrement dit, une relation 3FN qui n'est pas 4FN peut toujours être remplacée par des relations 4FN.

Merise: Modélisation des Données (FN5)

Dépendances Jointures et 5^{ème} FN

ACP

<u>AGENT</u>	<u>COMPAGNIE</u>	<u>PRODUIT</u>
Ali	Ford	Voiture
Ali	Ford	Camion
Rim	Ford	Voiture
Ali	Toyota	Voiture

4FN

Un agent représente une compagnie pour vendre des produits.

Soit la contrainte supplémentaire :

« Si un agent vend un produit, et s'il représente une compagnie faisant ce produit, alors il vend ce produit pour cette compagnie. »

Ce n'est pas le cas de Ali par exemple sans la ligne 1.

ACP est en 4FN mais non en 5FN.

Merise: Modélisation des Données (FN5)

Dépendances Jointures et 5^{ème} FN

ACP

<u>AGENT</u>	<u>COMPAGNIE</u>	<u>PRODUIT</u>
Ali	Ford	Voiture

Solution : En 3 relations !

AC

<u>AGENT</u>	<u>COMPAGNIE</u>
Ali	Ford
Rim	Ford
Ali	Toyota

AP

<u>AGENT</u>	<u>PRODUIT</u>
Ali	Voiture
Ali	Camion
Rim	Voiture

CP

<u>COMPAGNIE</u>	<u>PRODUIT</u>
Ford	Voiture
Ford	Camion
Toyota	Voiture

On dit qu'on est passé à la 5^e Forme Normale (5FN)

Merise: Modélisation des Données (FN5)

Dépendances Jointures et 5^{ème} FN

La contrainte précédente signifie :

si $\langle a, c \rangle$	<i>figure dans AC</i>	<i>(a travaille pour c)</i>
et $\langle a, p \rangle$	<i>figure dans AP</i>	<i>(a vend p)</i>
et $\langle c, p \rangle$	<i>figure dans CP</i>	<i>(c fabrique p)</i>

alors $\langle a, c, p \rangle$ *figure dans ACP*

(a travaille pour c pour vendre p)

Autrement dit :

$$\mathbf{ACP = AC * AP * CP}$$

On dit que ACP contient une dépendance jointure

Merise: Modélisation des Données (FN5)

Dépendances Jointures et 5^{ème} FN

Définition:

Une relation $R (X_1, X_2, \dots, X_n)$ satisfait la Dépendance Jointure (DJ)

$$\begin{array}{l} R_1 * R_2 * \dots * R_n \\ ssi \\ R = R_1 * R_2 * \dots * R_n \end{array}$$

où R_i est la projection de R sur les attributs X_i , i dans $[1..n]$

Autrement dit : R est décomposable en R_1, R_2, \dots, R_n

Exemple : $ACP = AC * AP * CP$

Merise: Modélisation des Données (FN5)

Dépendances Jointures et 5^{ème} FN

Définition :

Une relation R est en 5FN ssi:

Toute Dépendance Jointure de R est la conséquence d'une clé (candidate).

Exemple :

AC, AP, CP sont en 5FN

ACP n'est pas 5FN (Absence de DJ) (**ACP = AC*AP*CP**)

Remarque :

Une relation 4FN qui n'est pas 5FN peut toujours être remplacée par des relations 5FN. Remplacement sans perte