

FONDEMENTS DES BASES DE DONNÉES

Normalisation : les formes normales basées sur les DFs

Équipe pédagogique BD



Introduction

- ▶ Les formes normales sont des propriétés que doivent vérifier les schémas pour éviter les anomalies de mises à jour, ce sont en ce sens des *bons schémas*.
- ▶ Une forme normale s'applique à un *schéma de relation*, en fonction d'un certain ensemble de contraintes d'une classe donnée.
- ▶ Concernant les DF, l'idée générale est de n'avoir **que les clés** à vérifier, et d'éliminer au maximum des DF qui ne définissent pas des clés.

Dans la suite, R est un schéma de relations
et F un ensemble de DF définies sur R .

Inclusion des formes normales

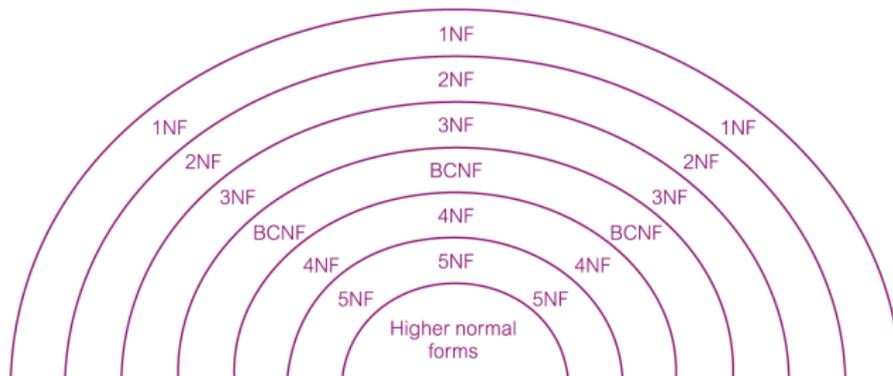


Figure 13.7

Diagrammatic illustration of the relationship between the normal forms.

Première forme normale (1FN)

Rappel : en relationnel, on est toujours en première forme normale, autrement dit toutes les valeurs des attributs sont atomiques.

Exemple

	<u>ISBN</u>	<u>Auteurs</u>
▶ Pas en 1FN :	2-212-09283-0	Gardarin
	2-7117-8645-5	{Abiteboul, Hull, Vianu}

Première forme normale (1FN)

Rappel : en relationnel, on est toujours en première forme normale, autrement dit toutes les valeurs des attributs sont atomiques.

Exemple

	ISBN	Auteurs
▶ Pas en 1FN :	2-212-09283-0	Gardarin
	2-7117-8645-5	{Abiteboul, Hull, Vianu}

	ISBN	Auteur
	2-212-09283-0	Gardarin
▶ En 1FN :	2-7117-8645-5	Abiteboul
	2-7117-8645-5	Hull
	2-7117-8645-5	Vianu

Deuxième forme normale (2FN)

Deuxième forme normale

R est en 2FN par rapport à F ssi, pour chaque DF $X \rightarrow A$ de F , l'une des deux conditions suivantes est remplie :

- ▶ A appartient à une clé de R ,
- ▶ X n'est pas un sous-ensemble *propre* d'une clé de R

Contre exemple

$\{AB \rightarrow C, B \rightarrow C\}$ est l'exemple typique qui n'est pas en 2FN.

Définition *alternative*

R est en 2FN par rapport à F ssi

- ▶ R est en 1FN
- ▶ aucun attribut non-clé ne dépend *partiellement* d'une clé candidate.

La 2FN a surtout un intérêt « historique »...

Exemple

- ▶ Soit $F = \{\{ isbn, bib. \} \rightarrow nb; isbn \rightarrow titre\}$
- ▶ La relation suivante est-elle en 2FN ?

isbn	titre	bib.	nb
2-212-09283-0	BD ...	UCBL1	5
2-7117-8645-5	Fondements des BD	UCBL1	10
2-212-09283-0	BD ...	INSA-L	3

- ▶ Soit la décomposition suivante de la relation.
- ▶ Est-elle sans perte d'information et de DFs ?

isbn	bib.	nb
2-212-09283-0	UCBL1	5
2-7117-8645-5	UCBL1	10
2-212-09283-0	INSA-L	3

isbn	titre
2-212-09283-0	BD ...
2-7117-8645-5	Fondements des BD

Insuffisance de la 2FN

$F = \{ isbn \rightarrow titre, editeur ; editeur \rightarrow pays \}$ En 2FN

ISBN	titre	editeur	pays
2-212-09283-0	BD ...	Eyrolles	France
2-7117-8645-5	Fondements ...	Vuibert	USA
2-212-0969-2	Internet/Intranet ...	Eyrolles	France

Il reste des redondances, notamment avec la DF $editeur \rightarrow pays$

Troisième forme normale (3FN)

Troisième forme normale (3FN)

R est en 3FN par rapport à F ssi, pour chaque DF $X \rightarrow A$ non triviale de F^+ , l'une des deux conditions suivantes est remplie :

- ▶ X est une super clé
- ▶ A appartient à une clé minimale de R .

Contre exemple

$\{A \rightarrow B, B \rightarrow C\}$ est l'exemple typique qui est en 2FN mais pas en 3FN.

Définition *alternative*

R est en 3FN par rapport à F ssi

- ▶ R est en 2FN
- ▶ toute les DFs sont directes : tout attribut non-clé dépend *directement* d'une clé.

La 3FN n'élimine pas toutes les redondances

Exemple

Soit $F = \{ isbn, bib \rightarrow nb_ex; isbn \rightarrow titre \}$

ISBN	Titre	Bib	Nb_Ex
2-212-09283-0	BD ...	UCBL1	5
2-7117-8645-5	Fondements des BD	UCBL1	10
2-212-09283-0	BD ...	INSA-L	3

Pas en 3NF (ni 2NF) et décomposée en

ISBN	Bib	Nb_Ex
2-212-09283-0	UCBL1	5
2-7117-8645-5	UCBL1	10
2-212-09283-0	INSA-L	3

$isbn, bib \rightarrow nb_ex$

ISBN	Titre
2-212-09283-0	BD ...
2-7117-8645-5	Fondements des BD

$isbn \rightarrow titre$

Cette décomposition est sans perte et préserve les dépendances fonctionnelles.

Exemple (2)

$F = \{ isbn \rightarrow titre, editeur ; editeur \rightarrow pays \}$ En 2FN, pas en 3FN

ISBN	Titre	Editeur	Pays
2-212-09283-0	BD ...	Eyrolles	France
2-7117-8645-5	Fondements ...	Vuibert	USA
2-212-0969-2	Internet/Intranet ...	Eyrolles	France

décomposée en

ISBN	Titre	Editeur	Editeur	Pays
2-212-09283-0	BD ...	Eyrolles	Eyrolles	France
2-7117-8645-5	Fondements ...	Vuibert	Vuibert	USA
2-212-0969-2	Internet/Intranet ...	Eyrolles		

$isbn \rightarrow titre, editeur$

$editeur \rightarrow pays$

Cette décomposition est sans perte et préserve les dépendances fonctionnelles.

Insuffisance de la 3NF

$rue, ville \rightarrow cp; cp \rightarrow ville$

en 3FN car :

- ▶ $\{rue, ville\}$ est une super clé ;
- ▶ $ville$ appartient à une clé candidate.

clés candidates :

- ▶ $\{rue, ville\}$
- ▶ $\{rue, cp\}$

Rue	CP	Ville
Rue J. Capelle	69100	Villeurbanne
Rue de la Doua	69100	Villeurbanne
Rue de la République	69001	Lyon
Rue de Baleine	69001	Lyon

Il reste des redondances liées à $cp \rightarrow ville$

Forme normale de Boyce-Codd

Forme normale de Boyce-Codd (FNBC)

R est en forme normale de Boyce-Codd par rapport à F ssi, pour chaque DF $X \rightarrow A$ de F , X est une superclé de R .

Autrement dit :

La forme normale de Boyce-Codd impose que toutes les parties gauches des DF sont des clés.

Contre exemple

$\{AB \rightarrow C, C \rightarrow B\}$ est l'exemple typique qui est en 3FN mais pas en FNBC.

Propriétés

- ▶ La FNBC implique la 3FN.
- ▶ Toute relation composée que de deux attributs est en FNBC.

Une forme idéale

La FNBC est, pour les DF, la forme idéale d'un schéma de bases de données.

Les trois propriétés suivantes sont équivalentes

- ▶ R est en FNBC par rapport à F ;
- ▶ R n'a pas de problème de redondances par rapport à F ;
- ▶ R n'a pas de problème de mise à jour par rapport à F ;

Exemple

$$F = \{rue, ville \rightarrow cp; cp \rightarrow ville\}$$

<i>R</i>	Rue	CP	Ville
	Rue J. Capelle	69100	Villeurbanne
	Rue de la Doua	69100	Villeurbanne
	Rue de la République	69001	Lyon
	Rue de Baleine	69001	Lyon

<i>R</i> ₁	CP	Ville
	69100	Villeurbanne
	69001	Lyon

<i>R</i> ₂	Rue	CP
	Rue J. Capelle	69100
	Rue de la Doua	69100
	Rue de la République	69001
	Rue de Baleine	69001

Cette décomposition

- ▶ Est bien **sans perte d'information**, car $R = R_1 \bowtie R_2$
- ▶ Elle n'est pas **sans perte de dépendances**, car on garde $cp \rightarrow ville$ mais **on perd $rue, ville \rightarrow cp$**

Décomposition en 3FN et en FNBC

Propriété

- ▶ Toute relation a au moins une décomposition en **3FN** qui préserve les dépendances fonctionnelles **et** qui est sans perte.
- ▶ Toute relation a au moins une décomposition en **FNBC** qui est sans perte mais qui *peut ne pas préserver* les DFs.

Alors, que faire ? Un compromis

- ▶ **Accepter de ne pas être en FNBC**, et donc d'avoir des anomalies de mise à jour :
 - ▶ Se contenter de déclarer les clés.
 - ▶ Implémenter les DF qui ne sont pas des clés (avec des déclencheurs).
- ▶ **Accepter de relâcher des contraintes**, c'est à dire enlever des DF et rendre l'application « plus souple ».

Création d'attributs : codage de couple

- ▶ On peut aussi rajouter des attributs et des DF. Dans l'exemple, on peut rajouter un attribut qui "regroupe" les couples Rue/Ville. On obtiendrait :

$(Rue, Ville, Rue/Ville, CP)$

ainsi que les DF

$(Rue, Ville \rightarrow Rue/Ville)$

$(Rue/Ville \rightarrow Rue, Ville, CP)$

et

$CP \rightarrow Ville$

La décomposition en FNBC est alors possible, et aucune sémantique n'est perdue.

Exemple (1/3)

On considère les règles métier de gestion des emploi du temps suivantes que l'on modélise par deux DFs sur la relation $R = HAC$

- ▶ pour un horaire (H) et un amphi (A) donnés il n'y a qu'un seul cours (C) : $f_1 : HA \rightarrow C$
- ▶ un cours a toujours lieu dans le même amphi : $f_2 : C \rightarrow A$.

La relation R est en **3FN** mais pas en **FNBC**. On considère trois possibilités de décomposition.

Solution 1 : on accepte les redondances et on reste en 3FN

R	<i>Horaire</i>	<i>Amphi</i>	<i>Cours</i>
	Lundi 07h45	Jussieu	LIF10
	Mardi 16h00	Jussieu	LIF10
	Lundi 10h00	Jussieu	LIF8

Exemple (2/3)

Solution 2 : on décompose en 3FNBC en perdant $HA \rightarrow C$

R_1	Cours	Amphi
	LIF10	Jussieu
	LIF8	Jussieu

R_2	Horaire	Cours
	Lundi 07h45	LIF10
	Mardi 16h00	LIF10
	Lundi 10h00	LIF8
	Lundi 10h00	LIF10

Le tuple rouge est bien autorisé dans R_2 !

Exemple (3/3)

Solution 3 : on fait un codage des couples *HA* et on décompose en FNBC

- ▶ On introduit un nouvel attribut *créneau* (*Cr*) qui identifie de façon **unique** un horaire dans un amphi.
- ▶ Avec les DFs $Cr \rightarrow HA$ et $HA \rightarrow Cr$ on assure la bijection entre un créneau et un couple horaire/amphi.

R_0	<i>Horaire</i>	<i>Amphi</i>	<i>Creneau</i>
	Lundi 07h45	Jussieu	Créneau1
	Mardi 16h00	Jussieu	Créneau2
	Lundi 10h00	Jussieu	Créneau3

R_1	<i>Cours</i>	<i>Amphi</i>
	LIF10	Jussieu
	LIF8	Jussieu

R'_2	<i>Creneau</i>	<i>Cours</i>
	Créneau1	LIF10
	Créneau2	LIF10
	Créneau3	LIF8

Fin.