

Normalisation

Comment éviter les anomalies

Décomposition

- Définition (sans perte d'information)

Projection et jointure

$R \rightarrow R1 * R2 * \dots * Rn$

R et $R1 * R2 * \dots * Rn$ aient même schéma.

–avec préservation des DF – si les DF sont conservées

- Exemple – avec perte d'information

R	N°	nom	âge	adresse	spécialité
	1	A	30	x1	S1
	2	B	30	x2	S2
	3	c	30	x3	S2
	4	D	40	x4	S1

$R'=R1*R2*R3$	N°	nom	adresse	âge	spécialité
	1	A	x1	30	S1
	1	A	x1	30	S2
	2	B	x2	30	S1
	2	B	x2	30	S2
	3	C	x3	30	S1
	3	C	x3	30	S2
	4	D	x4	40	S1

R1	âge	Spécialité
	30	S1
	30	S2
	40	S1

R2	N	nom	adr
	1	A	x1
	2	B	x2
	3	C	x3
	4	D	x4

R3	N	âge
	1	30
	2	30
	3	30
	4	40

Introduction

- Les indices du bon schéma
 - La représentation
 - La non-redondance
 - La séparation
 - Les deux approches de définir un modèle conceptuel
 - Tous les attributs et toutes les relations sont donnés – relation universelle.
- COMPAGNIE (M_employé, N_département, responsable, type_de_contrat)**
PATIENT(Patient_No, PatNom, Diagnose, Dcoteur_No, Docteur_Nom, Service_no, Service_Nom)
- Un ensemble de relations est donné avec un ensemble de DF entre leurs attributs
- Anomalies possibles pendant la manipulation des données :
 - anomalie de mise à jour :
 - anomalie de suppression:
 - anomalie d'insertion
 - redondance

Première forme normale

Une relation est en première forme normale si tout attribut contient une valeur atomique (est **basé sur un domaine simple**) et s'il n'y a pas des attributs répétés.

No	NOM	ENFANT	
		PRENOM	AGE
500	DUPONT	ANDRE	10
501	DURAND	JEAN	11
501	DURAND	PIERRE	12
510	LEFEBVRE	PAUL	13
510	LEFEBVRE	JACQUES	14

EMPLOYE (n matricule, nom)

ENFANT (n matricule, prénom, âge)

EMPLOYE	No	NOM
	500	DUPONT
	501	DURAND
	510	LEFEBVRE

ENFANTS	No	Prénom	âge
	500	André	10
	501	Jean	11
	501	Pierre	12
	510	Paul	13
	510	Jacques	14

Première forme normale

MOVIE(No, Name, Director, Actor1, Actor2, Actor3)
 MOVIE(No, Name, Director)
 CAST(NoMovie, Actor)

MOVIE	No	Name	Director	Actor1	Actor2	Actor3
	1	The Silence of the Lambs	Jonathan Demme	Jodie Foster	Anthony Hopkins	
	2	The Sixth Sense	M. Night Shyamalan	Bruce Willis		

MOVIE	No	Name	Director	CAST	NoMovie	Actor
	1	The Silence of the Lambs	Jonathan Demme		1	Jodie Foster
	2	The Sixth Sense	M. Night Shyamalan		1	Anthony Hopkins
					2	Bruce Willis

B. Shishedjiev - Algèbre relationnelle

5

Troisième forme normale

Une relation est en 3NF ssi :

- Elle est 2NF
- *Tout attribut n'appartenant pas à une clé ne dépend pas d'un attribut non clé.*

PERSONNEL(employé, nom, prénom, service, local)

Le local ne dépend que du service.

PERSONNEL(employé, nom, prénom, service)

LOCAL(service, local)

•Caractéristiques de 3NF

- Elle préserve les DF. (la fermeture transitive des DF est la même que celle de l'union des décomposées)
- Elle est sans perte.

B. Shishedjiev - Algèbre relationnelle

7

Seconde forme normale

Une relation est dite en 2NF si et seulement si (ssi) :

- Elle est en 1 NF.
- Tous les attributs non clé sont totalement dépendants de la totalité des clés de R

STOCK (pièce, entrepot, qté, localisation_entrepot)

S'il n'y a pas de pièce, pas de localisation dans l'entrepôt.

Décomposition

STOCK(pièce, entrepôt, qté)

LOCAL(entrepôt, localisation)

B. Shishedjiev - Algèbre relationnelle

6

Troisième forme normale

• Préservation des DF

VOITURE (num, marque, type, puissance, couleur)

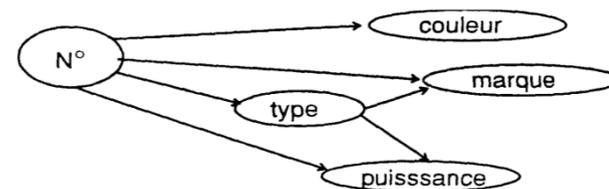
R'1(num, type)

R'2(num, puissance, couleur)

R1(num, type, couleur)

R'3(type, marque) |

R2 (type, marque, puissance).



B. Shishedjiev - Algèbre relationnelle

8

Forme normale de BOYCE-CODD (BCNF)

Une relation est BCNF ssi

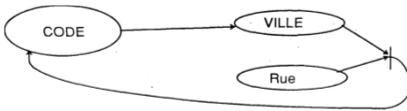
- Elle est en 3NF
- les seules DF élémentaires sont celles dans lesquelles une clé détermine un attribut (on élimine les dépendances partielles et transitives).

CODE_POSTAL(code, ville, rue)

CODE VILLE (code, ville)

CODE RUE(code, rue)

Codepostal	code	Ville	rue
59650	VA	Gambetta	
59650	VA	Jean-Jaurès	



B.Shishedjiev - Algèbre relationnelle

9

Forme normale de BOYCE-CODD (BCNF)

INSCRIPTION	CourseNum	Etudiant
	ENG101	Jones
	ENG101	Grayson
	ENG101	Samara
	MAT350	Grayson
	MAT350	Jones
	MAT350	Berg

TUTORIAL	Etudiant	Tuteur
	Jones	Clark
	Grayson	Chen
	Samara	Chen
	Grayson	Powers
	Jones	O'Shea
	Berg	Powers

TUTEURS	CourseNum	Tuteur
	ENG101	Clark
	ENG101	Chen
	MAT350	Powers
	MAT350	O'Shea

B.Shishedjiev - Algèbre relationnelle

11

Forme normale de BOYCE-CODD (BCNF)

Exemple 2:

Dans une université on veut d'avoir l'information pour l'inscription des étudiants en cours et l'affectation des tuteurs pour les TP. Supposons que les règles suivantes sont valides :

- Dans un cours il y a plusieurs étudiants ;
- Un étudiant peut s'inscrire en plusieurs cours ;
- Chaque tuteur n'est affecté qu'à un seul cours ;
- Pour chaque cour un étudiant a un seul tuteur.

COURSE_ETUD_TUT	CourseNum	Etudiant	Tuteur
	ENG101	Jones	Clark
	ENG101	Grayson	Chen
	ENG101	Samara	Chen
	MAT350	Grayson	Powers
	MAT350	Jones	O'Shea
	MAT350	Berg	Powers

Clés candidats – (CourseNum, Etudiant) et (Etudiant, Tuteur)

B.Shishedjiev - Algèbre relationnelle

10

Dépendances multivaluées

- Soit $R(A_1, A_i, A_n)$ une relation, et X, Z des sous-ensembles de clés.

On dit que $X \twoheadrightarrow Z$ (X multidétermine Z) ou qu'il existe une *multi-détermination* de Z par rapport à X , si, pour des valeurs de Y , il existe un ensemble de valeurs de X , indépendant des autres attributs (ensemble Z) de la relation.

Une MD caractérise l'interdépendance entre 2 ensembles d'attributs X et Z corrélés par un troisième Y .

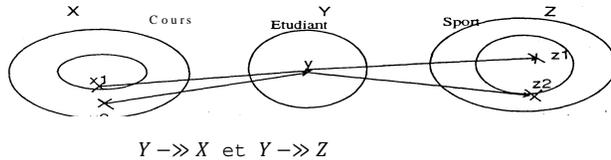
B.Shishedjiev - Algèbre relationnelle

12

Dépendances multivaluées

ETUDIANTS	N	COURS	SPORT
	100	BD	Tennis
	100	BD	Football
	200	BD	Vélo
	200	AN	Vélo

$N \twoheadrightarrow \text{COURS}$ et $N \twoheadrightarrow \text{SPORT}$



B. Shishedjiev - Algèbre relationnelle

13

Quatrième forme normale

- Une relation est 4NF ssi les seules DM élémentaires sont celles dans lesquelles une clé détermine un attribut.

ETUDIANT (num, cours, sport) n'est pas 4NF

$\text{cours} \twoheadrightarrow \text{sport}$

R1(num, cours)

R2(num, sport).

B. Shishedjiev - Algèbre relationnelle

15

Dépendances multivaluées

Axiomes d'inférence :

- COMPLEMENTATION : $X \twoheadrightarrow Y \implies X \twoheadrightarrow R-X-Y$
- MULTI-AUGMENTATION : si $(X \twoheadrightarrow Y)$ et W est un ensemble d'attributs de R , alors $XW \twoheadrightarrow YW$
- PSEUDO-TRANSITIVITE : $(X \twoheadrightarrow Y)$ et $(Y \twoheadrightarrow Z) \implies X \twoheadrightarrow Z-Y$
- REPLICATION : $X \twoheadrightarrow Y \implies X \twoheadrightarrow Y$
- COALESCENCE : $X \twoheadrightarrow Y$ et $Z \subseteq Y$ et il existe $W \subseteq R$ avec $W \cap Y = \emptyset$ et $W \twoheadrightarrow Z$, alors $X \twoheadrightarrow Z$
- UNION : $(X \twoheadrightarrow Y)$ et $(Y \twoheadrightarrow Z) \implies X \twoheadrightarrow YZ$.
- INTERSECTION : $X \twoheadrightarrow Y$ et $X \twoheadrightarrow Z \implies X \twoheadrightarrow Y \cap Z$.
- DIFFERENCE : $X \twoheadrightarrow Y$ et $X \twoheadrightarrow Z \implies X \twoheadrightarrow Y-Z$ et $X \twoheadrightarrow Z-Y$
- Une DM élémentaire $X \twoheadrightarrow Y$ de R est telle que :
 - Y n'est pas vide et est disjoint de X .
 - R ne contient pas une autre DM $X' \twoheadrightarrow Y'$ telle que $X' \in X$ et $Y' \in Y$.

B. Shishedjiev - Algèbre relationnelle

14

Dépendance de jointure

- La décomposition en 4NF ne suffit pas à éliminer les anomalies, car il peut encore y avoir des redondances.

Cette relation n'est pas 4NF, mais il n'existe pas de MD. En effet :

$(X \text{ CL } Z)$ et $(X \text{ SIO } T) \not\implies (X \text{ CL } T) \in R$ et $(X \text{ SIO } Z) \in R$

- ETUDIANT \twoheadrightarrow COURS est faux car $(X \text{ SIO } Z)$ n'existe pas et peut ne jamais exister.
- COURS \twoheadrightarrow ENSEIGNANT aussi car $(Y \text{ CL } Z)$ n'existe pas
- ENSEIGNANT \twoheadrightarrow ETUDIANT aussi car $(Y \text{ SIO } T)$ n'existe pas

R1	ETUDIANTS	COURS	ENSEIGNANT
	X	CL	Z
	X	CL	T
	X	SIO	T
	Y	SIO	T

B. Shishedjiev - Algèbre relationnelle

16

Cinquième forme normale

Dans R il y a des redondances inutiles (X CL Y y est 2 fois) mais les jointures de R1 R2, R2 R3 ou R1R3 ne redonnent pas R. Le problème est lié à la tentative de décomposition en 2

$X \rightarrow Y \implies (XY)$ et (YZ) , car Y et Z sont indépendants de X Il y a des relations non décomposables en 2, mais en N relations.

si $(a,b) \in R1$

$(a, c) \in R2 \implies (a, b, c) \in R$

$(b,c) \in R3$

alors : $R = R1 * R2 * R3$. C'est ce qu'on appelle la dépendance de jointure (notée *) dont les DM sont un cas particulier, car :

Soit $R(X, Y, Z)$ avec $X \rightarrow Y$ et $X \rightarrow Z$ Alors, la dépendance de jointure $*(XY, XZ)$ est satisfaite.

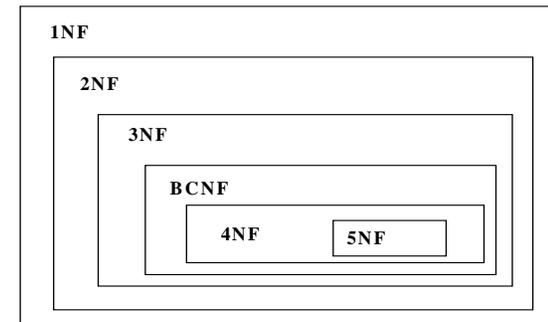
Les DM sont utilisées pour exprimer l'interdépendance de 2 attributs. Les dépendances de jointure expriment une interdépendance de plusieurs attributs entre eux et expriment une notion plus large que celle de DM.

R1	Etudiant	cours	R2	Etudiant	enseignant	R3	cours	enseignant
	X	CL		X	Z		CL	Z
	X	SIO		X	T		CL	T
	Y	CL		Y	T		SIO	T

B.Shishedjiev - Algèbre relationnelle

17

Le formes normales



B.Shishedjiev - Algèbre relationnelle

19

Cinquième forme normale

• La 5NF est la forme normale de protection-jointure

Une relation est 5NF si toute dépendance de jointure est impliquée par les clés candidates de R. Soit $R(A1, A2, A3, A4)$ A1 et A2 étant clés candidates. une décomposition sans perte est :

$*(A1A2, A1A3, A1A4)$ ou encore :

$*(A1A2, A2A3, A3A4)$

Une relation 5NF ne peut plus être décomposée sans perte d'information (avec projection et recomposition par jointure). On ne peut donc trouver d'autres décompositions.

B.Shishedjiev - Algèbre relationnelle

18

Exemple

Le domaine réel étudié ici est celui des vols assurés par différentes compagnies aériennes.

Les données sont modélisées par différents attributs de la façon suivante :

les noms des compagnies par l'attribut C

les numéros de vol par l'attribut V

les numéros d'avion par l'attribut A

les modèles d'avion par l'attribut M

la capacité des avions par l'attribut K

les noms des pilotes par l'attribut P

les heures de départ des vols par l'attribut H

les jours de départ des vols par l'attribut J

les villes de départ par l'attribut D

les villes d'arrivées par l'attribut E

Les DF sont :

(1) $M \rightarrow K$

(2) $A \rightarrow M, C$

(3) $V \leftrightarrow H, D, E$

(4) $J, V \rightarrow P, A$

(5) $J, H, P \rightarrow D, E, A$

(6) $J, H, A \rightarrow D, E, P$

$R(C,A,V,M,K,P,H,J,D,E)$ avec sur-clés $[VJ], [HDEJ], [PJH], [AJH]$

De DF (3) R n'est pas dans 2NF:

$R1(V,H,D,E)$ avec sur-clés $[V]$ et $[HDE]$

$R2(C,A,V,M,K,P,J)$ avec sur-clés $[VJ]$

De DF (2) R2 n'est pas dans 3NF

$R21(V,J,A,P)$ $[VJ]$

$R22(A,C,M,K)$ $[A]$

De DF (1) R22 n'est pas dans 3NF

$R221(A,C,M)$ $[A]$

$R222(M,K)$ $[M]$

Le schéma normalisé est :

$R1(\underline{V},H,D,E)$

$R21(\underline{V},\underline{J},A,P)$

$R221(\underline{A},C,M)$

$R222(\underline{M},K)$

B.Shishedjiev - Algèbre relationnelle

20

Exemple 2

Trouver le DF et normaliser le schéma dont l'extension est :

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
a1	b1	c1	d1	e1
a2	b1	c1	d1	e4
a1	b1	c1	d2	e5
a1	b3	c3	d1	e2
a1	b1	c1	d1	e7
a1	b2	c1	d1	e3
a3	b1	c3	d1	e6
a4	b1	c1	d2	e7

$DE \rightarrow A$; $EA \rightarrow D$; $E \rightarrow B$; $E \rightarrow C$; $AB \rightarrow C$

Réponse : $T1(A, D, E)$, $T2(A, B, C)$, $T3(B, E)$

Ou $T1(D, E, A)$, $T2(E, B, C)$

B. Shishedjiev Normalisation

21