

Formes Normales et Normalisation

Anomalies
Boyce-Codd BCNF
3^e Forme Normale 3NF

1

Anomalies et Redondance

Types d'anomalies
Exemple d'erreurs de conception

2

Anomalies

- ◆ Bon schéma relationnel:
 - ▶ pas de redondance
 - (redondance plusieurs fois la même information)
 - ▶ pas d'anomalies.
 - *Anomalie de mise à jour*: une occurrence d'une information est modifiée et pas les autres
 - *Anomalie de suppression*: une information pertinente est perdue en détruisant un n-uplet.

3

Exemple de schéma mal conçu

Drinkers(name, addr, beersLiked, manf, favBeer)

name	addr	beersLiked	manf	favBeer
Janeway	Voyager	Bud	A.B.	WickedAle
Janeway	???	WickedAle	Pete's	???
Spock	Enterprise	Bud	???	Bud

Les données sont redondantes:
les Dépendances Fonctionnelles
name -> addr favBeer
beersLiked -> manf.
déterminent les ???

4

Erreurs de conception et Anomalies

name	addr	beersLiked	manf	favBeer
Janeway	Voyager	Bud	A.B.	WickedAle
Janeway	Voyager	WickedAle	Pete's	WickedAle
Spock	Enterprise	Bud	A.B.	Bud

- Anomalie de Mise à jour: si Janeway part pour *l'Intrepid*, pensera-t-on à changer tous les n-uplets?
- Anomalie de Destruction: si personne n'aime Bud, on perd le fait que son fabricant Anheuser-Busch.

5

Forme Normale de Boyce Codd BCNF

Exemple
Mise en forme normale

6

Forme Normale de Boyce-Codd BCNF

- ◆ Une relation R est dite en **BCNF** ssi pour toute Dépendance Fonctionnelle non triviale (c.-à-d. X ne contient pas A) $X \rightarrow A$ sur les attributs de R X est une clé (pas forcément minimale)

7

Exemple

$Drinkers(\underline{name}, addr, \underline{beersLiked}, manf, favBeer)$
DF: $name \rightarrow addr\ favBeer, beersLiked \rightarrow manf$

- ◆ Une seule clé minimale $\{name, beersLiked\}$.
- ◆ Pour chacune DF: la partie droite n'est pas une clé
- ◆ $Drinkers$ n'est pas en BCNF (prendre l'une des deux DF Au choix)

8

Autre Exemple

$Beers(\underline{name}, manf, manfAddr)$

FD's: $name \rightarrow manf, manf \rightarrow manfAddr$

- ◆ Une seule clé minimale $\{name\}$.
- ◆ $name \rightarrow manf$ ne contredit pas BCNF, mais par contre la relation n'est pas BCNF à cause de: $manf \rightarrow manfAddr$.

9

Décomposition en BCNF

- ◆ Donnée: une relation R avec des DF F .
- ◆ Chercher les DF $X \rightarrow B$ telles que X ne soit pas une clé
 - ▶ Si R pas BCNF, il y en a au moins une.
- ◆ Calculer X^+ .
 - ▶ Qui ne contient pas tous les attributs, sinon X serait une clé.

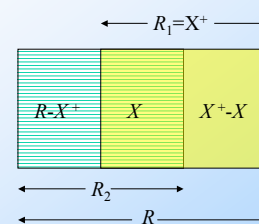
10

Décomposer R en suivant $X \rightarrow B$

- ◆ Remplacer R par deux relations dont les attributs sont:
 - 👉 $R_1 = X^+$.
 - 👉 $R_2 = R - (X^+ - X)$.
- ◆ **Projeter** les DF sur ces schémas
Attention! c'est la clôture de F , $Cl(F)$, qu'on projette sur R_1 et R_2

11

Schéma de la Décomposition



12

Exemple

$Drinkers(name, addr, \underline{beersLiked}, manf, favBeer)$

$F = name \rightarrow addr,$
 $name \rightarrow favBeer,$
 $beersLiked \rightarrow manf$

- ◆ Trouver une DF qui fait que $Drinkers$ n'est pas en BCNF: $name \rightarrow addr$.
- ◆ Clôture des attributs de gauche: $\{name\}^+ = \{name, addr, favBeer\}$.
- ◆ On obtient deux relations:

☞ $Drinkers1(name, addr, favBeer) = \{name\}^+$
☞ $Drinkers2(name, \underline{beersLiked}, manf)$
 $= \{name, addr, \underline{beersLiked}, manf, favBeer\}$ moins $\{adr, favBeer\}$

Exemple, suite

- ◆ Il faut maintenant voir si $Drinkers1$ et $Drinkers2$ sont en BCNF
- ◆ Projection des DF assez facile pour ce cas particulier -- calcul de la clôture $Cl(F)$
- ◆ Pour $Drinkers1(name, addr, favBeer)$, les DF pertinentes sont $name \rightarrow addr$ et $name \rightarrow favBeer$.
 - ▶ $\{name\}$ est la seule clé et $Drinkers1$ est en BCNF.

14

Exemple, suite

- ◆ Pour $Drinkers2(name, \underline{beersLiked}, manf)$, la seule DF est $beersLiked \rightarrow manf$, et la seule clé est $\{name, \underline{beersLiked}\}$.
 - ▶ Pas BCNF, on recommence.
- ◆ $beersLiked^+ = \{beersLiked, manf\}$, on décompose $Drinkers2$ en:
 - ☞ $Drinkers3(\underline{beersLiked}, manf) = X^+$
 - ☞ $Drinkers4(name, \underline{beersLiked}) = \{name, \underline{beersLiked}, manf\}$ moins $\{manf\}$

15

Exemple, fin

- ◆ Décomposition de $Drinkers$:
 - ☞ $Drinkers1(name, addr, favBeer)$
 - ☞ $Drinkers3(\underline{beersLiked}, manf)$
 - ☞ $Drinkers4(name, \underline{beersLiked})$
- ◆ Remarque:
 - ▶ $Drinkers1$ parle des personnes,
 - ▶ $Drinkers3$ parle des bières,
 - ▶ $Drinkers4$ relation entre personnes et bières

16

3e Forme Normale 3NF

Motivation
Décomposition

17

3e Forme Normale - Motivation

- ◆ Il y a des DF qui posent problème quand on décompose
 - $AB \rightarrow C$
 - $C \rightarrow B$.
 - ▶ Exemple (Paris, Marseille, Lyon, Rennes, ...)
 $A = adresse,$
 $B = ville,$
 $C = code postal.$
- ◆ Il y a deux clés, $\{A, B\}$ et $\{A, C\}$.
- ◆ $C \rightarrow B$ contredit BCNF, il faudrait décomposer en AC, BC .

18

On ne peut pas garder les DF

- ◆ Si décomposition AC et BC on ne peut pas retrouver $AB \rightarrow C$ à partir des DF projetées.
- ◆ Exemple avec
 A = adresse,
 B = ville,
 C = code postal
sur le transparent suivant.

19

Une DF non préservée

B:Ville	Code postal	Adresse	Code postal
Paris	75005	80, Bd St-Germain	75005
Paris	75006	80, Bd St-Germain	75006

Adresse	B:Ville	Code postal
80, Bd St-Germain	Paris	75005
80, Bd St-Germain	Paris	75006

Bien que toutes les DF soient vérifiées ($C \rightarrow B$) dans chacune des tables, la DF $adresse\ ville \rightarrow code_postal$ n'est pas respectée.

20

3NF: évite ce problème

- ◆ 3^e Forme Normale (3NF) assouplit la condition de BCNF pour ne pas décomposer dans ce cas.
- ◆ Un attribut est dit *premier* s'il fait partie d'une clé minimale.
- ◆ Une relation n'est pas en 3NF ssi on peut trouver une DF $X \rightarrow A$ telle que X n'est pas une clé et A ne fait pas partie d'une clé minimale.

21

Exemple

- ◆ Dans notre exemple avec les DF $AB \rightarrow C$ et $C \rightarrow B$, les clés minimales sont AB et AC .
- ◆ Chaque attribut A , B , ou C est premier.
- ◆ Bien que $C \rightarrow B$ contredise BCNF, le schéma A, B, C est en 3NF.

22

Décomposition SPI

Test pour SPI
Théorème de Heath

23

Décomposition SPI Sans Perte d'Information

$R(A_1, \dots, A_n)$
Décomposition de R suivant
 X_1, \dots, X_k ensembles d'attributs $\subset \{A_1, \dots, A_n\}$
 $R_i = \pi_{X_i}(R)$ pour $1 \leq i \leq k$
 $\mu(R) = R_1$ jointure R_2 jointure ... jointure R_k

a-t-on: $R = \mu(R)$?

24

Décomposition SPI

Quelques propriétés:

1. $\mu(R)=R_1$ jointure R_2 jointure ... jointure $R_k \supset R$
2. $\pi_{X_i}(\mu(R))=R_i$
3. $\mu(\mu(R))=\mu(R)$

Démonstration au tableau de 1.

25

Décomposition SPI

- $R(A,B,C,D,E)$
- Dépendances fonctionnelles:
 - $A \rightarrow C$
 - $B \rightarrow C$
 - $C \rightarrow D$
 - $DE \rightarrow C$
 - $CE \rightarrow A$
- Décomposition:
 - $X_1=AD$
 - $X_2=AB$
 - $X_3=BE$
 - $X_4=CDE$
 - $X_5=AE$

26

Décomposition SPI

	A	B	C	D	E
X1	A ₁	B ₁₂	B ₁₃	A ₄	B ₁₅
X2	A ₁	A ₂	B ₂₃	B ₂₄	B ₂₅
X3	B ₃₁	A ₂	B ₂₃	B ₂₄	B ₂₅
X4	B ₄₁	A ₂	B ₃₃	B ₃₄	A ₅
X5	A ₁	B ₅₂	B ₅₃	B ₅₄	A ₅

27

Décomposition SPI

- Remplir la table
- Forcer les égalités jusqu'à ce que les DF soient vraies
- Si une ligne est $A_1 \dots A_n$ décomposition SPI
- Sinon, il y a perte d'information, c.-à-d. $\mu(R) \not\subset R$

28

Décomposition SPI

- Si la ligne $A_1 \dots A_n$ n'apparaît pas, on a une relation qui satisfait les DFs et $A_1 \dots A_n$ est dans la jointure des projections mais pas dans R
- Sinon, il n'y a pas de perte d'information, c.-à-d. $\mu(R) = R$ (explication au tableau)

29

Décomposition SPI

- ◆ Théorème de Heath: une décomposition en X_1 et X_2 est SPI ssi
 - ▶ $(X_1 \cap X_2) \rightarrow X_1 - X_2$
 - ou
 - ▶ $(X_1 \cap X_2) \rightarrow X_2 - X_1$
- ◆ Application du test précédent

30

Décomposition 3NF SPI et sans perte de DF

Algo de base
Raffinement

31

Décomposition 3NF

- ◆ Calculer une couverture minimale
- ◆ Créer une table par DF + une par attribut isolé
 - ▶ Le résultat est 3NF
- ◆ Si on veut une décomposition SPI ajouter il faut une table avec les attributs d'une clé minimale
 - ▶ Découle du test SPI

32

Décomposition 3NF

- ◆ Correction de la décomposition:
- ◆ DF préservées, évidemment.
- ◆ 3NF. EN effet, si pb avec $X \rightarrow A$ dans YB (de $Y \rightarrow B$)
 - ▶ $A=B$
 X pas clé donc X strictement inclus Y
 $X \rightarrow A$ peut remplacer $Y \rightarrow B$
couverture pas minimale
 - ▶ $A \neq B$
soit Z clé minimale de YB incluse dans Y
 A pas dans Z car A premier et Z clé de YB
 $Z \rightarrow B$ et Z strictement inclus dans Y
couverture pas minimale

33

Décomposition 3NF

- ◆ $C \rightarrow P$
- ◆ $HP \rightarrow S$
- ◆ $HE \rightarrow S$
- ◆ $HS \rightarrow C$
- ◆ $CE \rightarrow N$
- ◆ $HC \rightarrow P$

34

Décomposition efficace en BCNF SPI

Le premier algo donné
impose de calculer les projections des DFs,
et donc leur clôture ce qui est coûteux.

On peut faire mieux...

35

Quelques propriétés

Prop. 1

Si on décompose R en $R_1 R_2 \dots R_k$ SPI
et si on décompose R_1 en S_1 et S_2 SPI
alors la décomposition de R en
 $S_1 S_2 R_2 \dots R_k$ est SPI

Raison: la jointure est associative.

36

Quelques propriétés

Prop. 2

Toute schéma à deux attributs est BCNF
Si R n'est pas BCNF alors
il existe des attributs A et B tels que
(R-AB) \rightarrow A

Réciproque fausse:

R(A,B,C)

C \rightarrow A

C \rightarrow B

ABC-AB \rightarrow A et pourtant R est BCNF

37

Quelques propriétés

Explication Prop. 2

pas BCNF à cause de X \rightarrow A

Si pas d'autres attributs que X et A, X clé.

Donc il existe B tq

R-AB \rightarrow X (réflexivité)

X \rightarrow A

et donc (R-AB) \rightarrow X

38

Quelques propriétés

Prop. 3

SI

- ◆ on projette un ensemble de DF F (sa clôture) sur $R_1 \subset R$ et qu'on obtient G
- ◆ et si on projette G (sa clôture) sur $R_2 \subset R_1$

ALORS

On obtient la même chose que si on projette F (sa clôture) sur R_2

39

Algo décomposition BCNF SPI

Z: le schéma à décomposer

Init: Y := Z

Pour chaque paire d'attributs A B de Y
si Y-AB \rightarrow A

faire Y := Y-B et C := A

Si aucune paire ne marche ou si |Y|=2

Y est BCNF

Décomposer Z en Y et Z-C.

Recommencer avec Z-C.

40

Algo décomposition BCNF SPI

- ◆ Cours
- ◆ Heure
- ◆ Salle
- ◆ Prof
- ◆ Etudiant
- ◆ Note
- ◆ C \rightarrow P
- ◆ HP \rightarrow S
- ◆ HE \rightarrow S
- ◆ HS \rightarrow C
- ◆ CE \rightarrow N

Attention les DFs ne sont pas conservées

41

Algo décomposition BCNF SPI

Correction de l'algo:

Le schéma extrait est BCNF (Prop. 2)

On décompose Z en XA et Z-A avec X \rightarrow A
(pas de perte par jointure, th. de Heath)

La combinaison des décomposition SPI est SPI
(Prop. 1)

Les X⁺ sont calculés avec F (Prop. 3)

42

Algo décomposition BCNF SPI

Peut décomposer un schéma déjà en BCNF

Identifiant
Nom
Département
Salaire

I->N
I->D
I->S

Décomposé en IN, ID, IS mais déjà BCNF!

43

Comparaison 3NF BCNF

Décompositions

- sans perte d'information
- avec préservation des DF

44

Propriétés de 3NF et de BCNF

- ◆ Il y a deux propriétés importantes des décompositions:
 - ✎ *Décomposition SPI: Sans perte d'information* : on peut projeter la relation de départ sur chacune des composantes et reconstruire la relation de départ.
 - ✎ *Préservation des dépendances fonctionnelles* : on peut vérifier dans les relations projetées que les dépendances originales sont préservées.

45

3NF et BCNF, fin

- ◆ On peut toujours mettre en BCNF sans perte d'information.
- ◆ On peut toujours mettre en 3NF
 - ▶ Sans perte d'information
 - ▶ En préservant les Dépendances Fonctionnelles
- ◆ Par contre il n'y a pas toujours de forme BCNF sans perte d'information et sans perte de dépendance fonctionnelle
 - ▶ Adresse rue ville est un exemple.

46