

Dépendances fonctionnelles et normalisation des BDDs

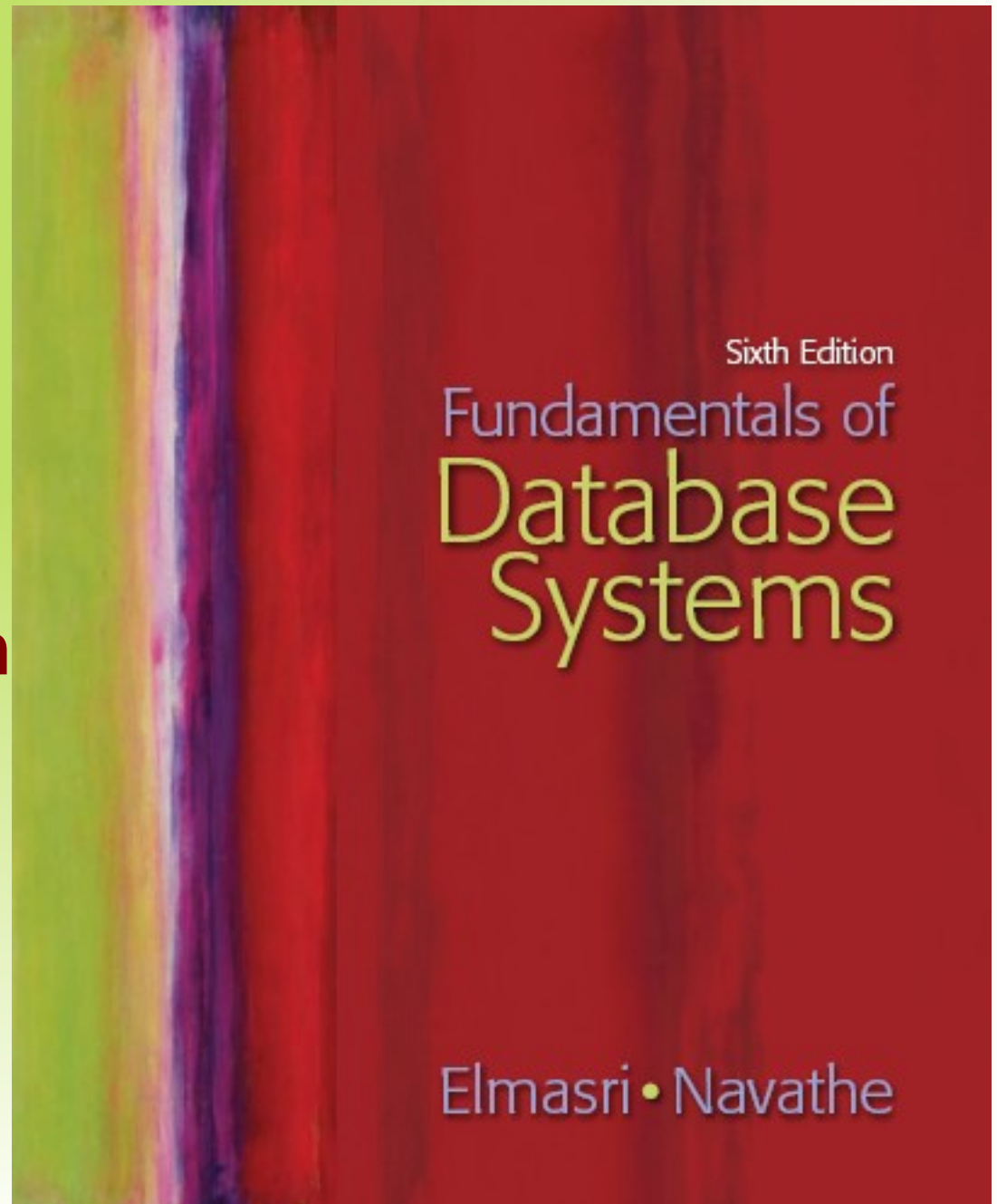
Traduction pour L3 informatique

à Lille 1: C.Kuttler

Addison-Wesley
is an imprint of

PEARSON

Copyright © 2011 Pearson Education, Inc. Publishing as Pearson Addison-Wesley



Plan

- Principes informels pour la conception des schémas relationnels
- Dépendances fonctionnelles
- Formes normales basées sur des clés primaires
- Définitions générales de la 2eme et 3eme forme normale
- Forme normale Boyce-Codd

Plan (suite)

- Dépendances multivaluée et 4ème forme normale
- Dépendances de jointure et 5ème forme normale

Introduction

- La *justesse* de la conception d'un schéma relationnel peut être appréciée au:
 - Niveau logique ou conceptuel
 - Niveau de l'implémentation (ou stockage physique)
- Deux démarches pour la conception de bases:
 - Méthodologie de conception ascendante ou descendante

Principes informels pour la conception des schémas

- Mesures de la qualité de la conception:
 - Clarté de la sémantique des attributs
 - Réduire la redondance d'information dans les tuples
 - Réduire les valeurs NULL dans les tuples
 - Ne pas permettre la génération d'attributs parasites

Sémantique des attributs des relations

- Sémantique d'une relation
 - Sa signification résulte de l'interprétation des valeurs de ses attributs
- Une sémantique de relation plus facile à expliquer
 - Indique une meilleure conception du schéma

Principe 1

- Concevez un schéma de face à ce qu'il soit facile d'en expliquer la signification
- Ne combinez pas des attributs provenant d'entités et de liaisons de différents types en une même relation
- Exemple de non-respect du Principe 1:
Figure 15.3

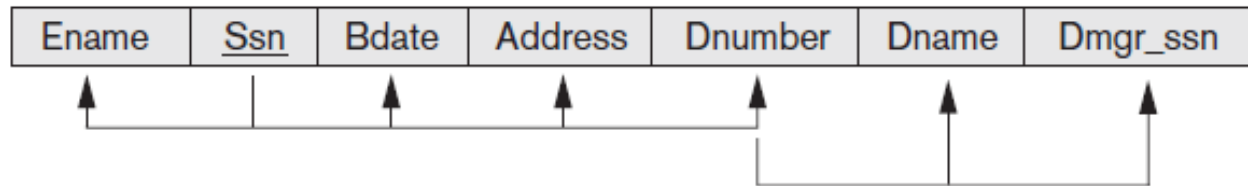
Principe 1 (cont'd.)

Figure 15.3

Two relation schemas suffering from update anomalies. (a) EMP_DEPT and (b) EMP_PROJ.

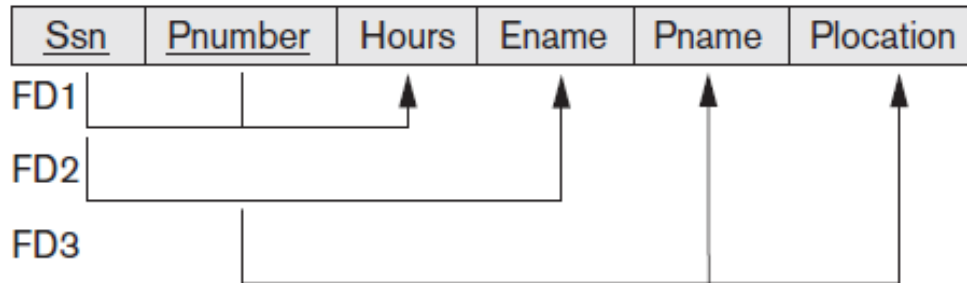
(a)

EMP_DEPT



(b)

EMP_PROJ



Informations redondantes et anomalies de mise à jour

- Regroupement d'attributs dans des schémas
 - Effet important sur l'espace de stockage
- De stocker des jointures naturelles de relations de bases peut mener a des **anomalies de mises à jour**
- Types d'anomalies de mise à jours:
 - Insertion
 - Suppression
 - Modification

Principe 2

- Concevez les schémas des relations de telle sorte qu'il ne puisse pas survenir d'anomalies d'insertion, de suppression et de modifications des données
- S'il y a des anomalies:
 - Indiquez-les clairement
 - Assurez-vous que les programmes qui mettent à jour la base opèreront correctement

Valeurs NULL dans les tupes

- Il arrive qu'on regroupe de nombreux attributs dans une relation “bien garnie”
 - Ceci peut entraîner un grand nombre de NULLs
- Problèmes résultant des NULLs
 - Gaspillage d'espace de stockage
 - Difficultés de compréhension

Principe 3

- Evitez de placer dans une relation de base des attributs dont les valeurs sont susceptibles d'être souvent NULL
- Si les NULLs sont inévitables:
 - Faites en sorte qu'ils n'apparaissent que pour des cas exceptionnels et qu'ils ne concernent pas une majorité de tuples dans la relation

Generation de tuples parasites

- Figure 15.5(a)
 - schemas des relations EMP_LOCS and EMP_PROJ1
- NATURAL JOIN
 - Le résultat produit beaucoup plus de tuples qu'il y en avait dans EMP_PROJ
 - On appelle ces tuples supplémentaires **parasites**
 - Représentent des données parasites ou erronées qui ne sont pas valides

Principe 4

- Veillez, au cours de la définition des schémas relationnels, à ce qu'ils puissent être réunis à l'aide de conditions d'égalité spécifiées sur des attributs jouant le rôle de clé primaires ou étrangères
 - Garanti l'absence des tuples parasites
- Evitez de faire des jointures sur des attributs qui ne sont pas uniques (clés primaires ou étrangères)

Récap des principes de conception

- Les anomalies entraînent un surcroît de travail
- Gaspillage d'espace de stockage lié aux values nulles
- Difficulté de réaliser des opérations d'aggrégation et de jointures
- Production de données non valides et parasites par des jointures

Dépendances fonctionnelles

- Outil formel pour l'analyse de schémas relationnels
- Permet de détecter et de décrire plusieurs des problèmes vu dans le précédent, de manière précise
- Théorie des dépendances fonctionnelles

Définition: dépendance fonctionnelle

- Contrainte entre deux ensembles d'attributs de la base

Def: Une *dépendance fonctionnelle*, notée $X \rightarrow Y$, entre deux ensembles d'attributs X et Y qui sont des sous-ensembles de R , spécifie une contrainte sur les tuples possibles pouvant former un état de R . Selon cette contrainte, si deux tuples t_1 et t_2 sont tels que $t_1[X] = t_2[X]$, alors ceux-ci doivent être tels que $t_1[Y] = t_2[Y]$.

- Propriété de la sémantique ou signification des attributs
- **Etats relationnels légaux**
 - Satisfont les contraintes de dépendance fonctionnelles

Définition: dépendance fonctionnelle (cont'd.)

- Donnée une relation peuplée:
 - Ne peux pas inférer automatiquement quelles DFs sont valides ou non
 - Sauf si la signification et les relations entre les attributs sont connus
 - Il suffit de trouver un seul contre exemple pour réfuter une DF

Formes normales basées sur les clés primaires

- Processus de normalisation
- Approches pour les projets de conception relationnelle
 - La réalisation préalable d'un schéma conceptuel puis sa traduction en un ensemble de relations
 - La conception tenant en compte de la connaissance externe dérivée d'une implémentation de fichiers ou de formulaires ou de rapports existants

Normalisation des relations

- Fait subir à un schéma relationnel une série de tests
 - Certifier qu'il satisfait à une forme normale déterminée
 - Opère du haut vers le bas
- **Test de formes normale**
- **Définition:** la forme normale d'une relation est liée à la condition de forme normale qu'elle satisfait et indique le degré auquel celle-ci a été normalisée

Normalisation des relations (cont'd.)

- Propriétés désirées du schéma relationnel:
 - **Jointure sans perte (non-additivité)**
 - Fortement recommandée
 - Garantie de ne pas générer de tuples parasites
 - **Préservation des dépendances**
 - Désirable, mais parfois sacrifiée pour d'autres aspects

Usage pratique des formes normales

- Dans la pratique, la conception
 - est entreprise en manière à ce que les conceptions auxquelles on aboutit soient de haute qualité et satisfassent les propriétés précédemment évoquées.
 - mène la procédure de normalisation seulement jusqu'à la 3NF, BCNF, ou la 4NF
- Pas besoin de normaliser jusqu'à la forme normale la plus élevée

Def : Denormalisation : stockage de la jointure de relations normalisés à un degré plus élevé, sous forme d'une relation de base.

Définitions des clés et des attributs des clés (p 242)

- Définition de **superclé** et **clé**
- **Clé candidate**
 - Si un schéma relationnel compte plusieurs clés
 - On désigne une comme **clé primaire**
 - Les autres sont dites **clés secondaires**

Def : L'attribut d'un schéma relationnel est R est appelé *attribut primaire* s'il est membre d'une des clés candidates de R. Dans le cas contraire, càd s'il n'est pas membre d'une clé candidate, on parle d'attribut non primaire.

Première forme normale


- Fait partie de la définition formelle d'une relation du modèle relation de base (à un niveau)
- Seules valeurs permises pour les attributs sont **atomiques** (ou **indivisibles**)
- Techniques pour passer en 1NF
 - Supprimer un attribut et le placer dans une relation séparée
 - Etendre la clé
 - Utiliser plusieurs attributs atomiques

Première forme normale (cont'd.)

- Interdit les **relations imbriquées**
 - Où un tuple peut avoir une relation à l'intérieur de lui-même
- Passage en 1NF:
 - Placer des attributs multivalués, dans une nouvelle relation
 - Propager la clé primaire dans cette nouvelle relation
 - **Désimbriquer** la relation, en faire un ensemble de relations 1NF

(a)

DEPARTMENT

Dname	<u>Dnumber</u>	Dmgr_ssn	Dlocations
			

(b)

DEPARTMENT

Dname	<u>Dnumber</u>	Dmgr_ssn	Dlocations
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

(c)

DEPARTMENT

Dname	<u>Dnumber</u>	Dmgr_ssn	<u>Dlocation</u>
Research	5	333445555	Bellaire
Research	5	333445555	Sugarland
Research	5	333445555	Houston
Administration	4	987654321	Stafford
Headquarters	1	888665555	Houston

Figure 15.9

Normalization into 1NF. (a) A relation schema that is not in 1NF. (b) Sample state of relation DEPARTMENT. (c) 1NF version of the same relation with redundancy.

Deuxième forme normale

- Repose sur le concept de **dépendance fonctionnelle complète**

- **Def.** Un schéma relationnel est en 2NF si chaque attribut non primaire de R est complètement dépendant fonctionnellement de la clé primaire de R.

Opposée à **dependance partielle**.

- Décomposition en plusieurs relations 2NF
 - Attributs non primaires ne sont associés qu'à la clé primaire dont ils sont fonctionnellement complètement dépendants

Troisième forme normale

- Repose sur le concept dépendance transitive
- Selon la définition d'origine de Codd, un schéma relationnel est en 3NF s'il satisfait 2NF et qu'aucun attribut non primaire n'est transitivement dépendant de la la clé primaire
- DF problématique
 - Partie gauche fait partie de la clé primaire
 - Partie gauche est un attribut ne jouant pas le rôle de clé

Définitions générales et deuxièmes et troisièmes formes normales (p 250)

Table 15.1 Summary of Normal Forms Based on Primary Keys and Corresponding Normalization

Normal Form	Test	Remedy (Normalization)
First (1NF)	Relation should have no multivalued attributes or nested relations.	Form new relations for each multivalued attribute or nested relation.
Second (2NF)	For relations where primary key contains multiple attributes, no nonkey attribute should be functionally dependent on a part of the primary key.	Decompose and set up a new relation for each partial key with its dependent attribute(s). Make sure to keep a relation with the original primary key and any attributes that are fully functionally dependent on it.
Third (3NF)	Relation should not have a nonkey attribute functionally determined by another nonkey attribute (or by a set of nonkey attributes). That is, there should be no transitive dependency of a nonkey attribute on the primary key.	Decompose and set up a relation that includes the nonkey attribute(s) that functionally determine(s) other nonkey attribute(s).

Définitions générales et deuxièmes et troisièmes formes normales (cont'd.)

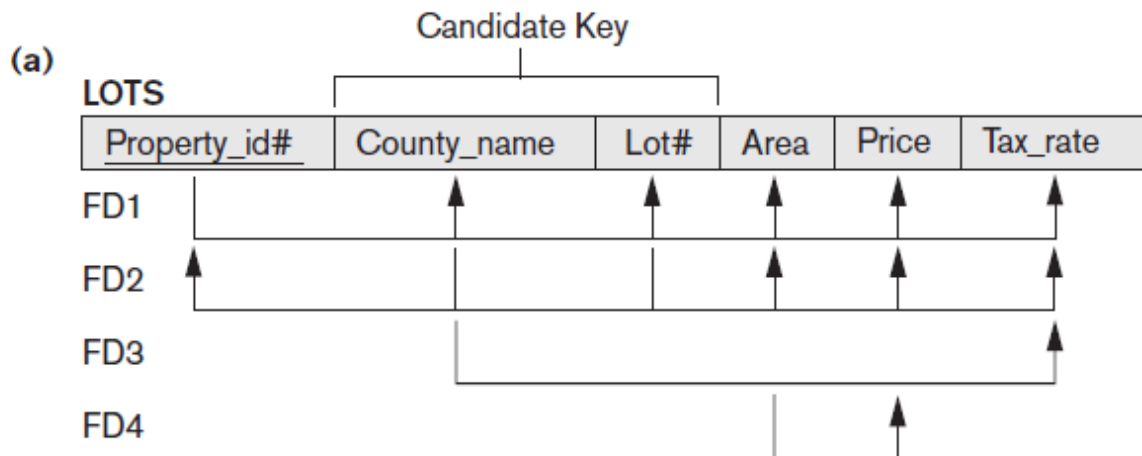
- **Attribut primaire**
 - Tout attribut qui fait partie d'une des clés primaires
- Les dépendances fonctionnelles partielles, complètes et transitives seront désormais envisagés en tenant compte de toutes les clés candidates d'une relation

Définition générale de seconde forme normale

Def: Un schéma relationnel R est en seconde forme normale (2FN) si chacun des attributs non primaires A de R n'est pas partiellement dépendant d'une des clés de R.

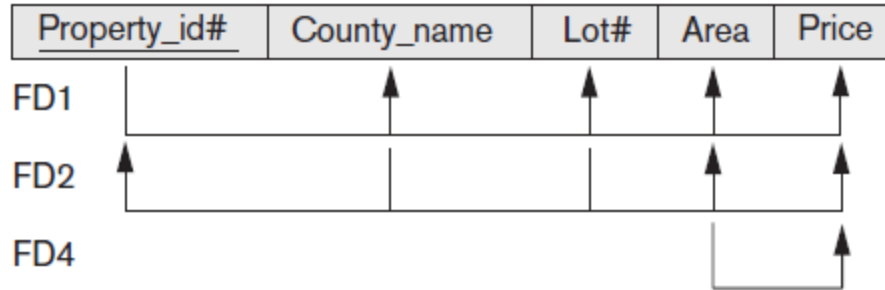
Figure 15.12

Normalization into 2NF and 3NF. (a) The LOTS relation with its functional dependencies FD1 through FD4. (b) Decomposing into the 2NF relations LOTS1 and LOTS2. (c) Decomposing LOTS1 into the 3NF relations LOTS1A and LOTS1B. (d) Summary of the progressive normalization of LOTS.

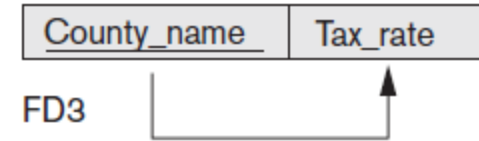


(b)

LOTS1

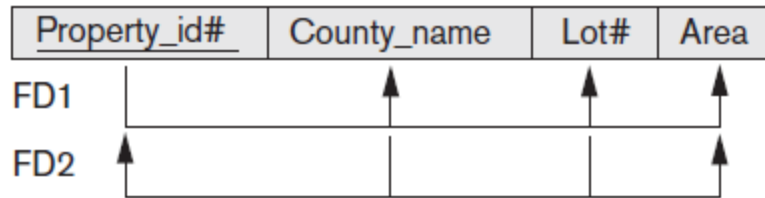


LOTS2

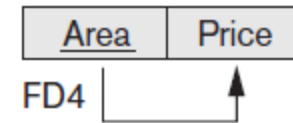


(c)

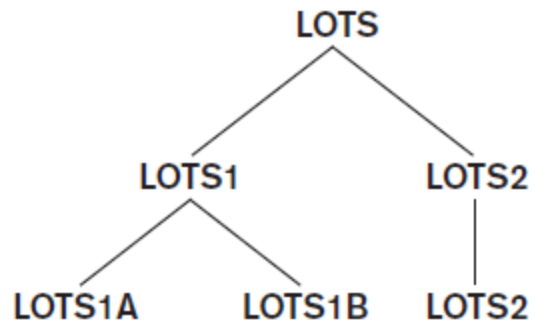
LOTS1A



LOTS1B



(d)



1NF

2NF

3NF

Définition générale de troisième forme normale

Definition. Un schéma relationnel R est en troisième forme normale (3NF) si, lorsqu'une dépendance fonctionnelle non triviale $X \rightarrow A$ est vraie dans R , soit

- X est une superclé de R , ou alors,
- A est un attribut primaire de R

Definition alternative: un schéma relationnel R est en 3NF si chacun des attributs non primaires de R satisfait les deux conditions suivantes:

- Il est fonctionnellement dépendant de chacun des clés de R
- Il est non transitivement dépendant de chacun des clés de R

Forme normale de Boyce-Codd

- Toute relation en FNBC est aussi en 3NF
 - Mais une relation en 3NF n'est pas nécessairement en FNBC

Def: Un schéma relationnel est en FNBC si lorsqu'une dépendance fonctionnelle non triviale $X \rightarrow A$ est vraie dans R, alors X est une superclé de R.

- Difference entre FNBC et 3NF:
 - Condition 2, qui permet à A d'être primaire, est absente pour FNBC
- La plupart des schémas relationnels qui sont en 3NF, sont aussi en FNBC.

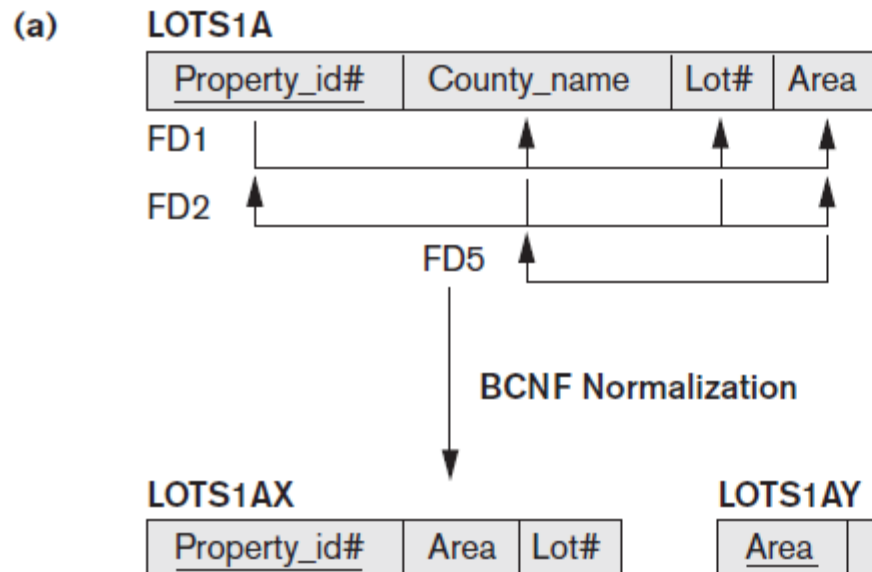
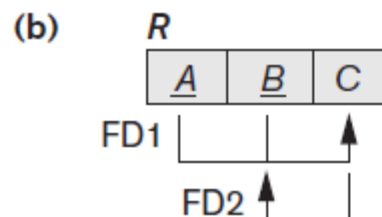


Figure 15.13

Boyce-Codd normal form. (a) BCNF normalization of LOTS1A with the functional dependency FD2 being lost in the decomposition. (b) A schematic relation with FDs; it is in 3NF, but not in BCNF.



Dépendances multi-valuées et quatrième forme normale

- Dépendance multi-valuée (DMV)
 - Conséquence de la 1NF

Definition. A multivalued dependency $X \twoheadrightarrow Y$ specified on relation schema R , where X and Y are both subsets of R , specifies the following constraint on any relation state r of R : If two tuples t_1 and t_2 exist in r such that $t_1[X] = t_2[X]$, then two tuples t_3 and t_4 should also exist in r with the following properties,¹⁵ where we use Z to denote $(R - (X \cup Y))$:¹⁶

- $t_3[X] = t_4[X] = t_1[X] = t_2[X]$.
- $t_3[Y] = t_1[Y]$ and $t_4[Y] = t_2[Y]$.
- $t_3[Z] = t_2[Z]$ and $t_4[Z] = t_1[Z]$.

Dépendances multi-valuées et quatrième forme normale (suite)

- Relations contenant des DMV non triviales
 - Relations tout-clés
- **Quatrième forme normale (4NF)**
 - Violée quand une relation a des DMV non désirables

Definition. A relation schema R is in 4NF with respect to a set of dependencies F (that includes functional dependencies and multivalued dependencies) if, for every *nontrivial* multivalued dependency $X \twoheadrightarrow Y$ in F^{+17} X is a superkey for R .

Join Dependencies and Fifth Normal Form

- **Join dependency**
- Multiway decomposition into fifth normal form (5NF)
- Very peculiar semantic constraint
 - Normalization into 5NF is very rarely done in practice

Join Dependencies and Fifth normal form (cont'd.)

Definition. A join dependency (JD), denoted by $JD(R_1, R_2, \dots, R_n)$, specified on relation schema R , specifies a constraint on the states r of R . The constraint states that every legal state r of R should have a nonadditive join decomposition into R_1, R_2, \dots, R_n . Hence, for every such r we have

$$* (\pi_{R_1}(r), \pi_{R_2}(r), \dots, \pi_{R_n}(r)) = r$$

Definition. A relation schema R is in **fifth normal form (5NF)** (or **project-join normal form (PJNF)**) with respect to a set F of functional, multivalued, and join dependencies if, for every nontrivial join dependency $JD(R_1, R_2, \dots, R_n)$ in F^+ (that is, implied by F),¹⁸ every R_i is a superkey of R .

Résumé

- Principes informels pour une conception de qualité
- Dépendance fonctionnelle
 - Outil de base pour l'analyse des schémas relationnels
- Normalisation:
 - 1NF, 2NF, 3NF, FNBC, 4NF, 5NF