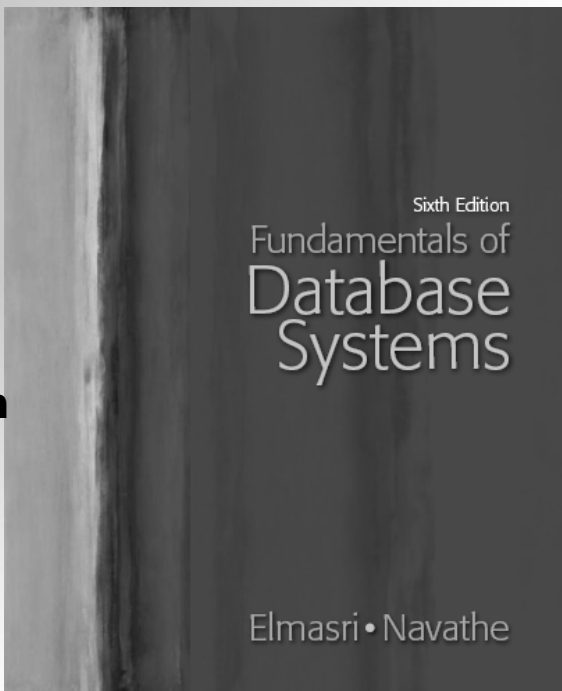


# Dépendances fonctionnelles et normalisation des BDDs

Traduction et adaptation:  
C.Kuttler



## Plan

- Principes informels pour la conception des schémas relationnels
- Dépendances fonctionnelles
  - Règles d'inférence pour les DFs
- Def des formes normales basées sur clés primaires
- Définitions générales de 2<sup>ème</sup> et 3<sup>ème</sup> forme normale
- Forme normale Boyce-Codd

## 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

## explications

La base d'une société.

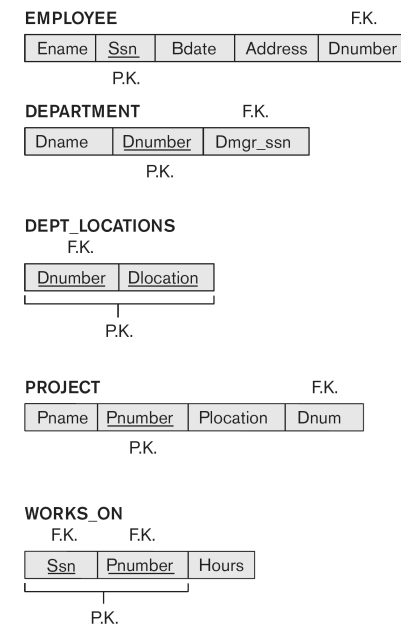
**EMPLOYEE** : chaque tuple représente un employé, avec son nom, numéro de sécu, adresse perso, et numéro du département où il travaille.

**DEPT\_LOCATIONS** : le numéro d'un département, et une de ses localisations. On note qu'il peut être réparti sur plusieurs sites. Cette relation représente un attribut multi-valué de Departments !

**WORKS\_ON** : relation M:N

**Figure 15.1**

A simplified COMPANY relational database schema.



**Figure 15.2**

Sample database state for the relational database schema in Figure 15.1.

### EMPLOYEE

Ename	Ssn	Bdate	Address	Dnumber
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4
Wallace, Jennifer S.	987654321	1941-06-20	291Berry, Bellaire, TX	4
Narayan, Ramesh K.	666884444	1962-09-15	975 Fire Oak, Humble, TX	5
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1

### DEPARTMENT

Dname	Dnumber	Dmgr_ssn
Research	5	333445555
Administration	4	987654321
Headquarters	1	888665555

### DEPT\_LOCATIONS

Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

## WORKS\_ON

Ssn	Pnumber	Hours
123456789	1	32.5
123456789	2	7.5
666884444	3	40.0
453453453	1	20.0
453453453	2	20.0
333445555	2	10.0
333445555	3	10.0
333445555	10	10.0
333445555	20	10.0
999887777	30	30.0
999887777	10	10.0
987987987	10	35.0
987987987	30	5.0
987654321	30	20.0
987654321	20	15.0
888665555	20	Null

## PROJECT

Pname	Pnumber	Plocation	Dnum
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4

# Principe 1

- Concevez un schéma de façon à ce qu'il soit facile d'en expliquer la signification
- Ne combinez pas des attributs provenant d'entités et des liaisons de différents types en une même relation

Addison-Wesley  
is an imprint of

PEARSON

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

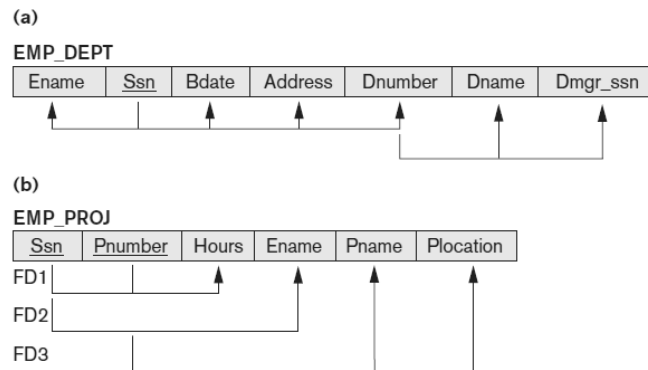
Addison-Wesley  
is an imprint of

PEARSON

Copyright © 2011 Ramez Elmasri and Shamkant Navathe

## Exemple de non respect du Principe 1

**Figure 15.3**  
Two relation schemas  
suffering from update  
anomalies. (a)  
EMP\_DEPT and (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 à des **anomalies de mises à jour**
- Types d'anomalies de mise à jours:
  - Insertion
  - Suppression
  - Modification

Addison-Wesley  
is an imprint of

PEARSON

Copyright © 2011 Ramez Elmasri and Shamkant Navathe

Addison-Wesley  
is an imprint of

PEARSON

Copyright © 2011 Ramez Elmasri and Shamkant Navathe

# Exemples d'anomalies des 3 types

## INSERTION :

Ajout d'un nouvel employé à EMP\_DEPT

Il faut ajouter les valeurs pour son département

Ou des NULLS (si il n'a pas encore de département)

P.ex. On ajoute un employé au Dept 5. Toutes les données doivent être cohérentes avec les autres du dept 5.

Avec la conception d'avant, on n'y met que le numéro du département.

Quoi faire si on veut ajouter un nouveau département qui n'a pas encore d'employés ? Il faut mettre des NULLS. Mais alors, la clé primaire Ssn est nulle.

Avec l'autre conception ce problème n'existe pas. On ajoute un dept, avec ou sans employés.

## SUPPRESSION :

Lorsqu'on supprime le dernier employé d'un département, on perd toute l'info sur ce département. Problème pas observé avec l'autre conception.

**MODIFICATION :** Si on a besoin de changer l'attribut d'un département, p.ex. Son manager, il faut le faire sur les tuples pour TOUS les employés.

EMP_DEPT						
Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

EMP_PROJ						
Ssn	Pnumber	Hours	Ename	Pname	Plocation	
123456789	1	32.5	Smith, John B.	ProductX	Bellaire	
123456789	2	7.5	Smith, John B.	ProductY	Sugarland	
666884444	3	40.0	Narayan, Ramesh K.	ProductZ	Houston	
453453453	1	20.0	English, Joyce A.	ProductX	Bellaire	
453453453	2	20.0	English, Joyce A.	ProductY	Sugarland	
333445555	2	10.0	Wong, Franklin T.	ProductY	Sugarland	
333445555	3	10.0	Wong, Franklin T.	ProductZ	Houston	
333445555	10	10.0	Wong, Franklin T.	Computerization	Stafford	
333445555	20	10.0	Wong, Franklin T.	Reorganization	Houston	
999887777	30	30.0	Zelaya, Alicia J.	Newbenefits	Stafford	
999887777	10	10.0	Zelaya, Alicia J.	Computerization	Stafford	
987987987	10	35.0	Jabbar, Ahmad V.	Computerization	Stafford	
987987987	30	5.0	Jabbar, Ahmad V.	Newbenefits	Stafford	
987654321	30	20.0	Wallace, Jennifer S.	Newbenefits	Stafford	
987654321	20	15.0	Wallace, Jennifer S.	Reorganization	Houston	
888665555	20	Null	Borg, James E.	Reorganization	Houston	

Figure 15.4

Sample states for EMP\_DEPT and EMP\_PROJ resulting from applying NATURAL JOIN to the relations in Figure 15.2. These may be stored as base relations for performance reasons.

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

# Valeurs NULL dans les tuples

- 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 opéreront correctement
- 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

## Principe 4

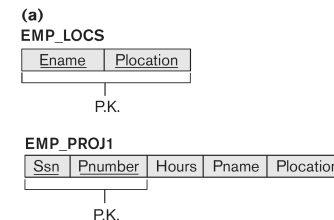
- Veillez, au cours de la définition des schémas relationnels, à ce qu'ils puissent être joints à l'aide de conditions d'égalité spécifiées sur des attributs jouant le rôle de clé primaires ou étrangères
  - Garanti l'absence des tuples parasites
- Evitez de faire des jointures sur des attributs qui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

## Génération de tuples parasites

- Figure 15.5(a)
  - relations EMP\_LOCS et EMP\_PROJ1, remplaçant EMP\_PROJ de 15.3(b)
- NATURAL JOIN (15.6)
  - Le résultat produit beaucoup plus de tuples qu'il y en avait dans EMP\_PROJ

On appelle ces tuples supplémentaires **parasites**, ils sont marqués par \*

- Représentent des données parasites ou erronées qui ne sont pas valides



**Figure 15.5**

Particularly poor design for the EMP\_PROJ relation in Figure 15.3(b). (a) The two relation schemas EMP\_LOCS and EMP\_PROJ1. (b) The result of projecting the extension of EMP\_PROJ from Figure 15.4 onto the relations EMP\_LOCS and EMP\_PROJ1.

(b)

EMP_LOCS		EMP_PROJ1				
Ename	Plocation	Ssn	Pnumber	Hours	Pname	Plocation
Smith, John B.	Bellaire	123456789	1	32.5	ProductX	Bellaire
Smith, John B.	Sugarland	123456789	2	7.5	ProductY	Sugarland
Narayan, Ramesh K.	Houston	666884444	3	40.0	ProductZ	Houston
English, Joyce A.	Bellaire	453453453	1	20.0	ProductX	Bellaire
English, Joyce A.	Sugarland	453453453	2	20.0	ProductY	Sugarland
Wong, Franklin T.	Sugarland	333445555	2	10.0	ProductY	Sugarland
Wong, Franklin T.	Houston	333445555	3	10.0	ProductZ	Houston
Wong, Franklin T.	Stafford	333445555	10	10.0	Computerization	Stafford
Zelaya, Alicia J.	Stafford	333445555	20	10.0	Reorganization	Houston
Jabbar, Ahmad V.	Stafford	999887777	30	30.0	Newbenefits	Stafford
Wallace, Jennifer S.	Stafford	999887777	10	10.0	Computerization	Stafford
Wallace, Jennifer S.	Houston	987987987	10	35.0	Computerization	Stafford
Borg, James E.	Houston	987987987	30	5.0	Newbenefits	Stafford
		987654321	30	20.0	Newbenefits	Stafford
		987654321	20	15.0	Reorganization	Houston
		888665555	20	NULL	Reorganization	Houston

Ssn	Pnumber	Hours	Pname	Plocation	Ename
123456789	1	32.5	ProductX	Bellaire	Smith, John B.
* 123456789	1	32.5	ProductX	Bellaire	English, Joyce A.
123456789	2	7.5	ProductY	Sugarland	Smith, John B.
* 123456789	2	7.5	ProductY	Sugarland	English, Joyce A.
* 123456789	2	7.5	ProductY	Sugarland	Wong, Franklin T.
666884444	3	40.0	ProductZ	Houston	Narayan, Ramesh K.
* 666884444	3	40.0	ProductZ	Houston	Wong, Franklin T.
* 453453453	1	20.0	ProductX	Bellaire	Smith, John B.
453453453	1	20.0	ProductX	Bellaire	English, Joyce A.
* 453453453	2	20.0	ProductY	Sugarland	Smith, John B.
453453453	2	20.0	ProductY	Sugarland	English, Joyce A.
* 453453453	2	20.0	ProductY	Sugarland	Wong, Franklin T.
* 333445555	2	10.0	ProductY	Sugarland	Smith, John B.
* 333445555	2	10.0	ProductY	Sugarland	English, Joyce A.
333445555	2	10.0	ProductY	Sugarland	Wong, Franklin T.
* 333445555	3	10.0	ProductZ	Houston	Narayan, Ramesh K.
333445555	3	10.0	ProductZ	Houston	Wong, Franklin T.
333445555	10	10.0	Computerization	Stafford	Wong, Franklin T.
* 333445555	20	10.0	Reorganization	Houston	Narayan, Ramesh K.
333445555	20	10.0	Reorganization	Houston	Wong, Franklin T.

Figure 15.6

Result of applying NATURAL JOIN to the tuples above the dashed lines in EMP\_PROJ1 and EMP\_LOCS of Figure 15.5. Generated spurious tuples are marked by asterisks.

Addison-Wesley  
is an imprint of

PEARSON

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

## 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 de tuples parasites par des jointures

Addison-Wesley  
is an imprint of

PEARSON

Copyright © 2011 Ramez Elmasri and Shamkant Navathe

## Dépendances fonctionnelles

Avec les clés, les DF permettent de définir des formes normales, pour

- spécifier des mesures formelles pour la qualité d'une conception;
- détecter des problèmes vu dans le précédent, de manière précise.

Addison-Wesley  
is an imprint of

PEARSON

Copyright © 2011 Ramez Elmasri and Shamkant Navathe

## Définition: dépendance fonctionnelle

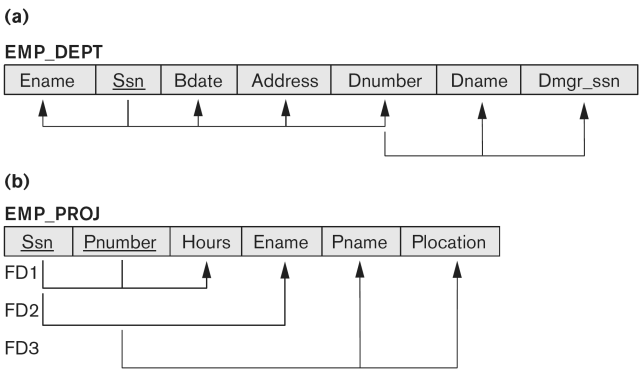
- Contrainte entre deux ensembles d'attributs de la base  $(R(A_1, A_2, \dots, A_n))$ .
- Propriété de la sémantique ou signification des attributs
- Def: Une *dépendance fonctionnelle*, notée  $X \rightarrow Y$ , entre deux ensembles d'attributs X et Y qui sont des sous-ensembles de R, spécifie une 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  $t_1[Y] = t_2[Y]$ .
- **Etats relationnels légaux** satisfont les contraintes de dépendance fonctionnelles

Addison-Wesley  
is an imprint of

PEARSON

Copyright © 2011 Ramez Elmasri and Shamkant Navathe

**Figure 15.3**  
Two relation schemas suffering from update anomalies. (a) EMP\_DEPT and (b) EMP\_PROJ.



Pour (b) :

- **{Ssn,Pnumber} → Hours**  
Le numéro de sécu et numéro de projet determinent le nombre d'heures travaillés par cet employé sur ce projet par semaine.
- **Ssn → Ename**  
Le numéro de sécu détermine le nom de l'employé.
- **Pnumber → {Pname, Location}**
- Le numéro du projet détermine le nom et l'emplacement de ce projet

# Dépendances fonctionnelles

- Donnée une relation peuplée:
  - On ne peut 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

**Figure 15.8**  
A relation  $R(A, B, C, D)$  with its extension.

A	B	C	D
a1	b1	c1	d1
a1	b2	c2	d2
a2	b2	c2	d3
a3	b3	c4	d3

## Dépendances fonctionnelles possibles :

- $B \rightarrow C$
- $C \rightarrow B$
- $\{A,B\} \rightarrow C$
- $\{C,D\} \rightarrow B$

## FD non valides (contre exemples) :

- $A \rightarrow B$  (tuples 1 et 2)
- $B \rightarrow A$  (tuples 2 et 3)
- $D \rightarrow C$  (tuples 3 et 4)

## Règles d'inférence pour DFs (1)

Etant donné un ensemble de DFs  $F$ , nous pouvons en déduire des DFs supplémentaires, qui sont vérifiées par tous les états légaux de la base.

Les règles d'inférence d'Armstrong sont:

- IR1. (**Réflexivité**) si  $Y$  est sous ensemble de  $X$ , alors  $X \rightarrow Y$
- IR2. (**Extensibilité**) Si  $X \rightarrow Y$ , alors  $XZ \rightarrow YZ$ 
  - (Notation:  $XZ$  signifie  $X \cup Z$ )
- IR3. (**Transitivité**) si  $X \rightarrow Y$  et  $Y \rightarrow Z$ , alors  $X \rightarrow Z$

IR1, IR2, IR3 constituent un ensemble de règles d'inférence correct et complet

- Ses règles sont valides et toutes les autres règles déduites d'elles également

## Règles d'inférence pour DFs (2)

Quelques règles supplémentaires sont utiles:

- **Décomposition:**  
Si  $X \rightarrow YZ$ , alors  $X \rightarrow Y$  et  $X \rightarrow Z$
- **Union:** Si  $X \rightarrow Y$  et  $X \rightarrow Z$ , alors  $X \rightarrow YZ$
- **Pseudo-transitivité:** Si  $X \rightarrow Y$  et  $WY \rightarrow Z$ , alors  $WX \rightarrow Y$  et  $WY \rightarrow Z$ , alors  $WX \rightarrow Z$

Les trois dernières règles peuvent être déduites de IR1, IR2 et IR3 (propriété de complétude)

## Règles d'inférence pour DFs (3)

**Clôture** d'un ensemble  $F$  de DFs est l'ensemble  $F^+$  de toutes les DFs pouvant en être déduites

**Clôture** d'un ensemble d'attributs  $X$  par rapport à  $F$  est l'ensemble  $X^+$  de tous les attributs qui dépendent fonctionnellement de  $X$

$X^+$  peut être calculé en appliquant IR1, IR2, IR3 avec les FDs de  $F$ , de manière itérée

### Algo: détermination de $X^+$ , la clôture de $X$ sous $F$

```
X+ := X;  
repeat  
    exX+ := X+;  
    for each dépendance fonctionnelle Y → Z dans F do  
        if X+ ⊇ Y then  
            X+ := X+ ∪ Z;  
until (X+ = exX+);
```



# Equivalence d'ensembles de DFs

Deux ensembles de DFs F et G sont **équivalents** si:

- Chaque DF dans F peut être déduite de G, et
- Chaque DF dans G peut être déduite de F
- Donc, F et G sont équivalents si  $F^+ = G^+$

Définition (**Couvrir**):

- F **couvre** G si chaque DF dans G peut être déduite de F
  - (i.e., si  $G^+$  est sous-ensemble de  $F^+$ )

F et G sont équivalents s'ils se couvrent mutuellement

Il y a un algorithme pour tester l'équivalence d'ensemble de DFs

# 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 : Dénormalisation : stockage de la jointure de relations normalisées à un degré plus élevé, sous forme d'une relation de base.

## Définition: clés et attributs primaires

**Def : Une superclé** d'un schéma relationnel  $R=\{A_1, A_{2n}, \dots, A_n\}$  est un sous-ensemble d'attributs  $S$  de  $R$  caractérisé par la propriété suivante : Dans chacun état relationnel  $r$  de  $R$ , deux tuples  $t_1$  et  $t_2$  ne seront pas tels que  $t_1[S]=t_2[S]$ .

Une **clé** est une superclé dotée de la propriété supplémentaire suivante : la suppression d'un de ses attributs annule son statut de superclé.

**Terminologie:** Si un schéma relationnel compte plusieurs clés, chacune est appelée clé candidate. 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, c'est s'il n'est pas membre d'une clé candidate, on parle d'attribut non primaire.

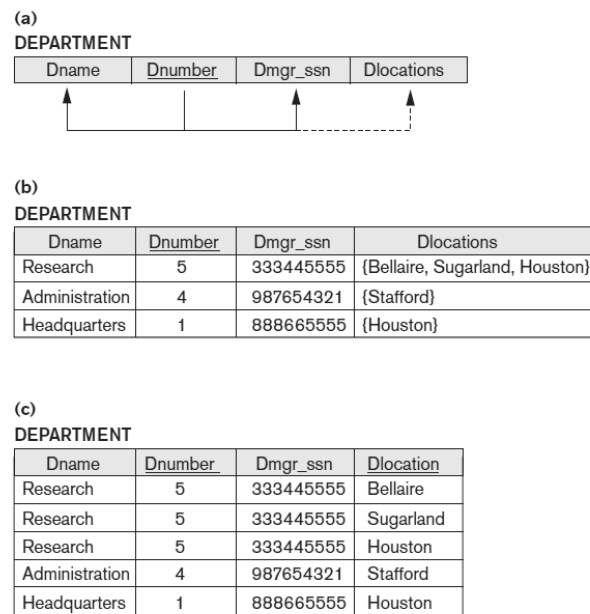
## Première forme normale

- 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**)
- **Interdit les relations imbriquées**  
Où un tuple peut avoir une relation à l'intérieur de lui-même

## Première forme normale (cont'd.)

### •Passage en 1NF:

- **Désimbriquer** la relation, en faire un ensemble de relations 1NF
- Placer des attributs multivalués, dans une nouvelle relation avec des attributs atomiques
- Propager la clé primaire dans cette nouvelle relation



**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.

Addison-Wesley  
is an imprint of

PEARSON

Copyright © 2011 Ramez Elmasri and Shamkant Navathe

## Deuxième forme normale

$X \rightarrow Y$  est dépendance fonctionnelle **complète**, si la suppression d'un des attributs A de X annule la DF. Opposée à dépendance fct **partielle**.

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

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

Addison-Wesley  
is an imprint of

PEARSON

Copyright © 2011 Ramez Elmasri and Shamkant Navathe

(a)  
EMP\_PROJ

Ssn	Ename	Projs	
		Pnumber	Hours

(b)  
EMP\_PROJ

Ssn	Ename	Pnumber	Hours
123456789	Smith, John B.	1	32.5
		2	7.5
666884444	Narayan, Ramesh K.	3	40.0
453453453	English, Joyce A.	1	20.0
		2	20.0
333445555	Wong, Franklin T.	2	10.0
		3	10.0
		10	10.0
		20	10.0
999887777	Zelaya, Alicia J.	30	30.0
		10	10.0
987987987	Jabbar, Ahmad V.	10	35.0
		30	5.0
987654321	Wallace, Jennifer S.	30	20.0
		20	15.0
888665555	Borg, James E.	20	NULL

(c)  
EMP\_PROJ1

Ssn	Ename

EMP\_PROJ2

Ssn	Pnumber	Hours

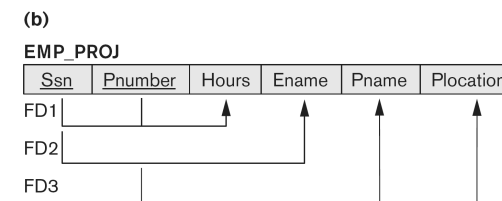
**Figure 15.10**  
Normalizing nested relations into 1NF. (a) Schema of the EMP\_PROJ relation with a nested relation attribute PROJ.S. (b) Sample extension of the EMP\_PROJ relation showing nested relations within each tuple. (c) Decomposition of EMP\_PROJ into relations EMP\_PROJ1 and EMP\_PROJ2 by propagating the primary key.

Addison-Wesley  
is an imprint of

PEARSON

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

## Exemple: DF complète vs partielle



Pour 15.3 (b) :

- **{Ssn,Pnumber} → Hours**  
complet : Ni Ssn → Hours,  
ni Pnumber → Hours sont vraies.
- **{Ssn,Pnumber} → Ename**  
partielle : Ssn → Ename est vraie.

Addison-Wesley  
is an imprint of

PEARSON

Copyright © 2011 Ramez Elmasri and Shamkant Navathe

Addison-Wesley  
is an imprint of

PEARSON

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

# Troisième forme normale

- $X \rightarrow Y$  est transitive s'il y a un ensemble d'attributs  $Z$  qui n'est ni une clé candidate ni un sous-ensemble d'une des clés de  $R$ , et que  $X \rightarrow Z$  et  $Z \rightarrow Y$  sont toutes les deux vraies.
- Def: un schéma relationnel est en 3NF s'il satisfait 2NF et qu'aucun attribut non primaire n'est transitivement dépendant de 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

Addison-Wesley  
is an imprint of  
PEARSON

Copyright © 2011 Ramez Elmasri and Shamkant Navathe

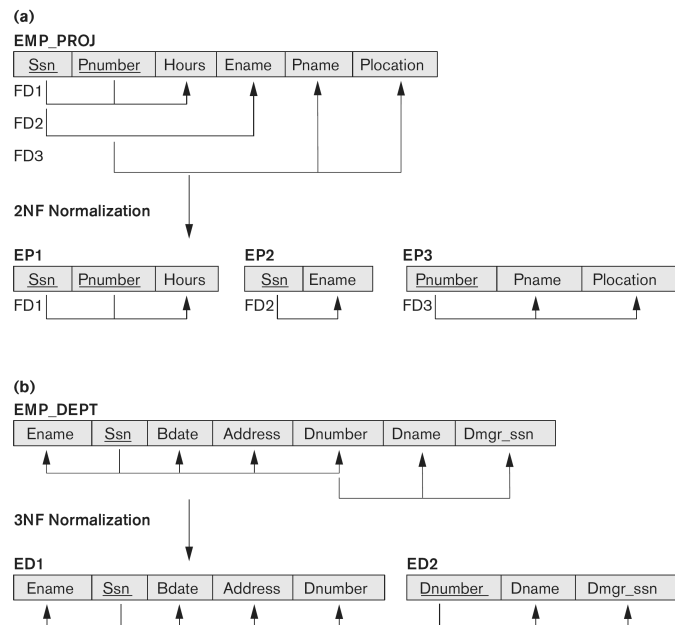


Figure 15.11

Normalizing into 2NF and 3NF. (a) Normalizing EMP\_PROJ into 2NF relations. (b) Normalizing EMP\_DEPT into 3NF relations.

Addison-Wesley  
is an imprint of  
PEARSON

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

## Récap: 2NF 3NF via clés primaires

Forme normale	Test	Remède (normalisation)
Première (1FN)	La relation ne doit pas comporter d'attributs non atomiques ou de relations imbriquées.	Former de nouvelles relations pour chaque attribut non atomique ou relation imbriquée.
Deuxième (2FN)	Pour les relations dans lesquelles la clé primaire contient plusieurs attributs, il ne doit pas y avoir d'attribut non-clé fonctionnellement dépendant d'une partie de la clé primaire.	Décomposer et construire une nouvelle relation pour chaque clé partielle et son (ses) attribut(s) dépendant(s). Il faut s'assurer de préserver une relation avec la clé primaire d'origine et tous les attributs qui dépendent fonctionnellement d'elle.
Troisième (3FN)	La relation ne doit pas avoir d'attribut non-clé fonctionnellement déterminé par un autre attribut non-clé (ou par un ensemble de plusieurs attributs non-clés). Autrement dit, il ne doit pas y avoir de dépendance transitive d'un attribut non-clé vis-à-vis de la clé primaire.	Décomposer et construire une relation qui inclut le ou les attributs non-clés qui déterminent fonctionnellement le ou les autres attributs non-clés.

Addison-Wesley  
is an imprint of  
PEARSON

Copyright © 2011 Ramez Elmasri and Shamkant Navathe

## Définitions générales 2NF et 3NF

- Les définitions précédentes interdisent les dépendances partielles et transitives sur la clé primaire.
- Les dépendances fonctionnelles partielles, complètes et transitives seront désormais envisagés en tenant compte de toutes les clés candidates d'une relation
- Attribut primaire: tout attribut qui fait partie d'une des clés candidates

Addison-Wesley  
is an imprint of  
PEARSON

Copyright © 2011 Ramez Elmasri and Shamkant Navathe

# Définition générale de 2NF

**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.

Autrement dit: chaque attribut non primaire est fonct complètement dépendant de chacune des clés de R

# Définition générale de troisième forme normale

**Définition.** 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

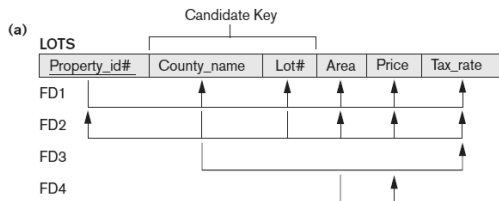
- X est une superclé de R, ou alors,
- A est un attribut primaire de R

**Définition 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 chacune des clés de R
- Il est non transitivement dépendant de chacune des clés de R

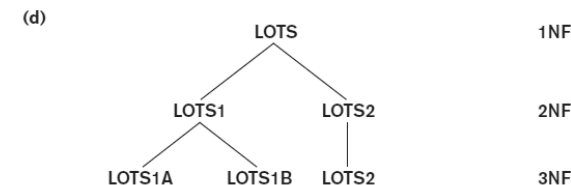
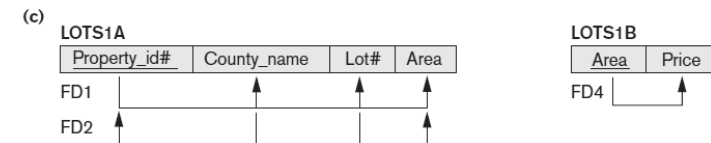
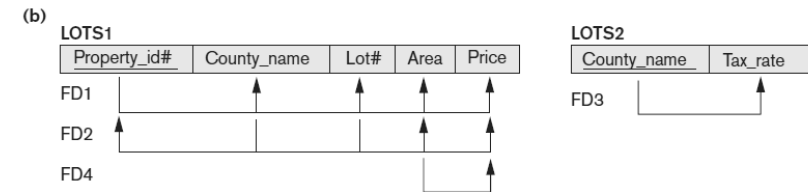
## Exemple

**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.



- LOTS décrit des parcelles de terrain en vente dans différents départements (County).
- 2 clés candidates : Property\_id# et {County\_name, Lot#}
  - Numéro de lots sont uniques à l'intérieur d'un département
  - ID de propriété est unique, globalement
- FD3 : le taux d'imposition est fixe par département
- FD4 : La superficie détermine le prix.

**Viola la 2NF : Taux d'imposition est partiellement dépendant de la clé candidate à cause de DF3.**



Explications

Pour Normaliser LOTS en 2NF,  
Decompose en deux relations LOTS1 et LOTS2.  
On construit LOTS1 en supprimant l'attribut pour le taux d'imposition qui viole 2NF de LOTS, en le placant avec le nom de departement (la partie gauche de DF3 qui est la cause de la dependance partielle) dans une autre relation LOTS2.  
Notez que DF4 ne viole pas 2NF et qu'elle est transf  e    LOTS1.

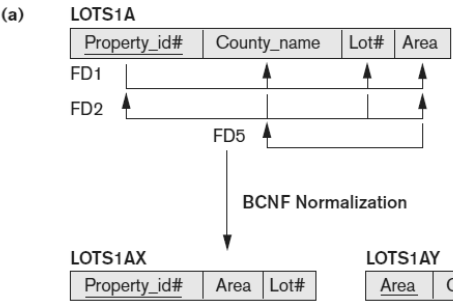
LOTS2 est en 3NF. Mais DF4 de LoTS1 viole 3NF car superficie (area) n'est pas une supercl  , et PRIX n'est pas un attribut primaire dans LOTS1. Pour normaliser LOTS1 en 3NF, on la d  compose en sch  mas relationnels LOTS1A et LOTS1B.

On construit LOTS1A en supprimant l'attribut PRIX qui viole 3NF de LOTS1 et en placant celui-ci avec SUPERFICIE (AREA) dans une autre relation LOTS1B. SUPERFICIE est la partie gauche de la DF4 qui est la cause de la dependance transitive.

# Forme normale de Boyce-Codd

**Def: Un sch  ma relationnel est en FNBC si lorsqu'une d  pendance fonctionnelle non triviale  $X \rightarrow A$  est vraie dans R, alors X est une supercl   de R.**

- Diff  rence entre FNBC et 3NF:
  - Condition 2, qui permet    A d'  tre primaire, est absente pour FNBC
- Toute relation en FNBC est aussi en 3NF. Mais, une relation en 3NF n'est pas n  cessairement en FNBC
- La plupart des sch  mas 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.

Cet exemple montre que BCNF est plus fort que 3NF .

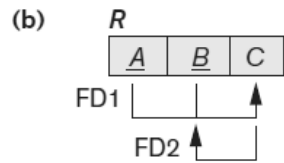
On regarde la relation LOTS1A.

Supposez maintenant qu'il y a des milliers de lignes dans cette relation, mais qu'il n'y a que deux departements, DeKalb et Fulton. Supposez que les tailles les lots en hectar (AREA) pour DeKalb sont {0.5,0.6,0.7,0.8,0.9,1.0} et celles de Fulton {1.1,1.2,...,1.9,2.0}.

Dans ce cas, cette DF supplementaire est valide:  
FD5 : Area → County\_name.  
Avec cette DF la relation reste bien en 3NF puisque County\_name est attribut prime.

On peut construire une relation separee R(Area,Contry\_name) avec les seulement 16 lignes. Ceci va r  duire la redondance.

Donc BCNF est une forme normale plus forte que 3NF, qui interdirait LOTS1A et propose comment le d  composer, en deux relations en BCNF, avec mois de redondances.



Explication

En pratique la majorité des schémas en BCNF sont également en 3NF.

Exception :

Seulement si  $X \rightarrow A$  est valide pour un schéma R, sans que X soit une superclé, et avec A attribut prime, la relation R sera en 3NF mais pas 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,<sup>15</sup> where we use  $Z$  to denote  $(R - (X \cup Y))$ :<sup>16</sup>

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

Pas vu en amphi

## 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  $R$ .

Pas vu en amphi

## 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

Pas vu en amphi

# 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  $F$ ),<sup>18</sup> every  $R_i$  is a superkey of  $R$ .

Pas vu en amphi

Addison-Wesley  
is an imprint of

PEARSON

Copyright © 2011 Ramez Elmasri and Shamkant Navathe

## 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

Addison-Wesley  
is an imprint of

PEARSON

Copyright © 2011 Ramez Elmasri and Shamkant Navathe