

**Εισαγωγή στις Συναρτησιακές
Εξαρτήσεις και στην
Κανονικοποίηση Σχεσιακών Βάσεων
Δεδομένων**

1. Απλές οδηγίες Σχεδιασμού Σχεσιακών Βάσεων Δεδομένων (1)

- Τι είναι ο σχεδιασμός βάσεων δεδομένων ;
 - Η κατάλληλη ομαδοποίηση γνωρισμάτων (attributes) για την δημιουργία ενός καλού σχεσιακού σχήματος
- Ποια είναι τα κριτήρια καλών σχημάτων σχέσεων ;

Απλές οδηγίες Σχεδιασμού Σχεσιακών Βάσεων Δεδομένων(2)

- Αρχικά θα αναφέρουμε απλές οδηγίες για τον σχεδιασμό καλών σχημάτων.
- Στην συνέχεια θα μελετήσουμε τυπικές έννοιες όπως αυτές των συναρτησιακών εξαρτήσεων και των κανονικών μορφών
 - - 1NF (First Normal Form) Πρώτη Κανονική Μορφή
 - - 2NF (Second Normal Form) Δεύτερη Κανονική Μορφή
 - - 3NF (Third Normal Form) Τρίτη Κανονική Μορφή
 - - BCNF (Boyce-Codd Normal Form) BC Κανονική Μορφή

1.1 Η σημασία των γνωρισμάτων θα πρέπει να είναι ξεκάθαρη

- ΟΔΗΓΙΑ 1: Κάθε πλειάδα σε μια σχέση θα πρέπει να αναπαριστά μια οντότητα ή ένα στιγμιότυπο μιας σχέσης.
 - Γνωρίσματα διαφορετικών Οντοτήτων (Εργαζόμενοι, Τμήματα, Έργα) δεν θα πρέπει να αναμειγνύονται στην ίδια σχέση.
 - Μόνο Ξένα Κλειδιά θα πρέπει να χρησιμοποιούνται για να αναφερόμαστε σε ξένες σχέσεις.
 - Γνωρίσματα Οντοτήτων και Σχέσεων θα πρέπει να διαχωρίζονται.
- Στόχος: Σχεδιάστε ένα σχήμα το οποίο να μπορεί να εξηγείται εύκολα σχέση με σχέση. Η σημασιολογία (*semantics*) των γνωρισμάτων θα πρέπει να ερμηνεύεται εύκολα.

Ένα απλό σχήμα βάσης δεδομένων για μια εταιρεία (COMPANY)

Εικόνα 10.1 σε
5^η Έκδοση

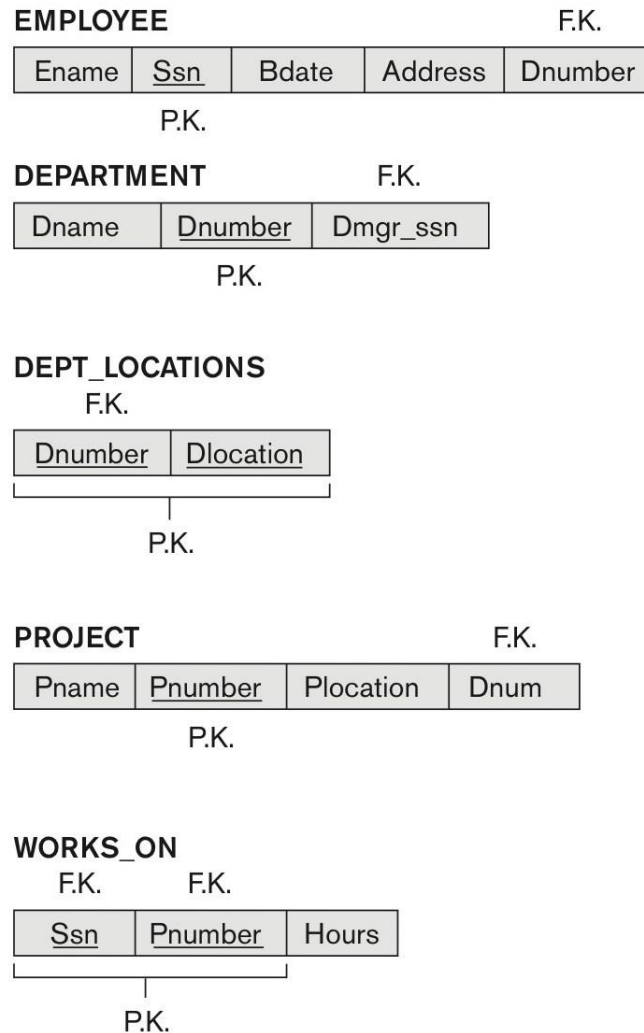


Figure 14.1 A simplified COMPANY relational database schema.

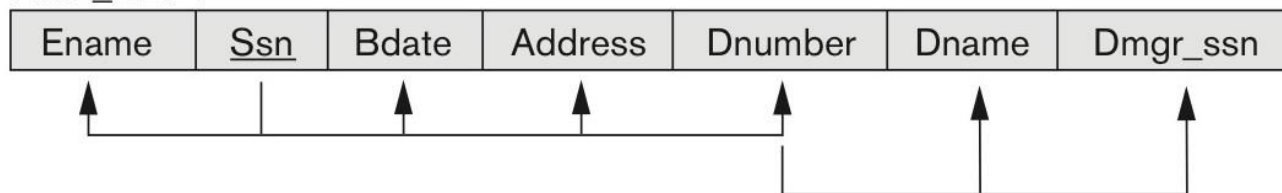
Παραδείγματα σχέσεων με ανωμαλίες τροποποίησης

Figure 14.3

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

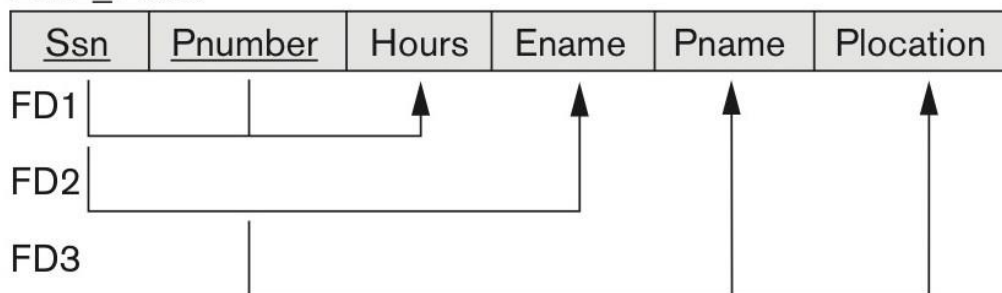
(a)

EMP_DEPT



(b)

EMP_PROJ



Εικόνα 10.3 σε
5^η Έκδοση

Δείγματα προβληματικών σχέσεων

Figure 14.4

Sample states for EMP_DEPT and EMP_PROJ resulting from applying NATURAL JOIN to the relations in Figure 14.2. These may be stored as base relations for performance reasons.

Εικόνα 10.4 σε
5^η Έκδοση

EMP_DEPT							Redundancy	
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						Redundancy		Redundancy	
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				

1.2 Πλεονάζουσες πληροφορίες σε πλειάδες και Ανωμαλίες Ενημέρωσεις

- Πλεονάζουσες πληροφορίες δημιουργούν τα ακόλουθα προβλήματα
 - Προκαλούν σπατάλη αποθηκευτικού χώρου
 - Προκαλούν προβλήματα με ανωμαλίες ενημέρωσης
 - Ανωμαλίες Εισαγωγής
 - Ανωμαλίες Διαγραφής
 - Ανωμαλίες Τροποποίησης

ΠΑΡΑΔΕΙΓΜΑ ΑΝΩΜΑΛΙΑΣ ΤΡΟΠΟΠΟΙΗΣΗΣ

- Στην σχέση :
 - ΕΡΓΑΖΟΜΕΝΟΣ_ΕΡΓΟ (Αρ_Εργαζομένου, Αρ_Εργου, Ονομα_Εργαζομένου, Ονομα_εργου, Ωρες)
- Μπορεί να έχουμε την εξής Ανωμαλία Τροποποίησης:
 - Η αλλαγή του έργου με αριθμό P1 από “Χρέωση” σε “Λογιστικά Πελάτη” μπορεί να προκαλέσει μια ενημέρωση σε 100 εγγραφές των εργαζομένων στο έργο P1.

ΠΑΡΑΔΕΙΓΜΑ ΑΝΩΜΑΛΙΑΣ ΕΙΣΑΓΩΓΗΣ

- Στην σχέση:
 - ΕΡΓΑΖΟΜΕΝΟΣ_ΕΡΓΟ (Αρ_Εργαζομενου, Αρ_Εργου, Ονομα_Εργαζομένου, Ονομα_εργου, Ωρες)
- Μπορεί να έχουμε την εξής Ανωμαλία Εισαγωγής:
 - Δεν μπορούμε να εισάγουμε πληροφορίες για ένα έργο χωρίς την εισαγωγή τουλάχιστον ενός εργαζομένου.
- Ή αντίστροφα
 - Δεν μπορούμε να εισάγουμε έναν εργαζόμενο εκτός και αν το αντιστοιχίσουμε και με ένα έργο.

ΠΑΡΑΔΕΙΓΜΑ ΑΝΩΜΑΛΙΑΣ ΔΙΑΓΡΑΦΗΣ

- Στη σχέση:
 - ΕΡΓΑΖΟΜΕΝΟΣ_ΕΡΓΟ (Αρ_Εργαζομ, Αρ_Εργου, Ονομα_Εργαζ, Ονομα_εργου, Ωρες)
- Μπορεί να έχουμε την Ανωμαλία Διαγραφής:
 - Όταν διαγράφουμε ένα Έργο θα προκαλέσουμε την διαγραφή όλων των εργαζομένων στο έργο.
 - Ή αντίστροφα όταν ένας εργαζόμενος είναι ο μοναδικός εργαζόμενος σε ένα έργο, αν τον διαγράψουμε διαγράφουμε και τις πληροφορίες του έργου.

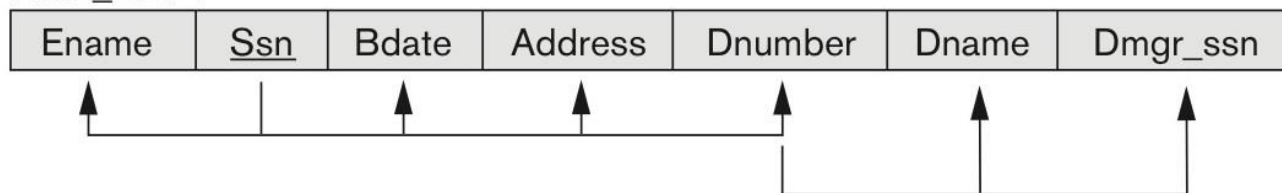
Παραδείγματα σχέσεων με ανωμαλίες τροποποίησης

Figure 14.3

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

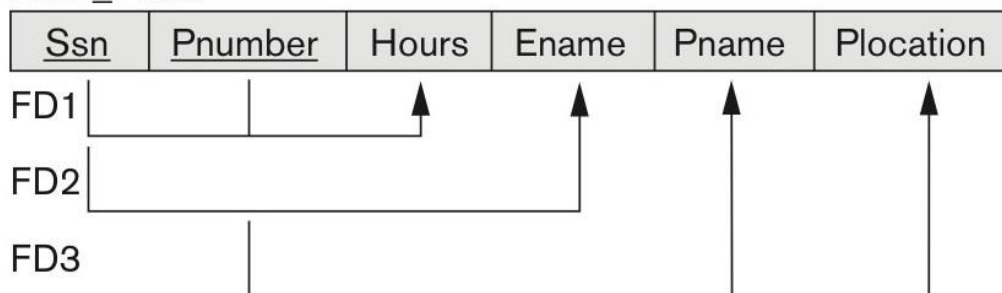
(a)

EMP_DEPT



(b)

EMP_PROJ



Εικόνα 10.3 σε
5^η Έκδοση

Δείγματα προβληματικών σχέσεων

Figure 14.4

Sample states for EMP_DEPT and EMP_PROJ resulting from applying NATURAL JOIN to the relations in Figure 14.2. These may be stored as base relations for performance reasons.

Εικόνα 10.4 σε
5^η Έκδοση

EMP_DEPT							Redundancy	
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						Redundancy		Redundancy	
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				

Οδηγία Σχετικά με πλεονάζουσες πληροφορίες και Ανωμαλίες Ενημέρωσης

■ ΟΔΗΓΙΑ 2:

- Σχεδιάστε ένα σχήμα που να μην παρουσιάζονται οι ανωμαλίες εισαγωγής, διαγραφής και τροποποίησης.
- Αν υπάρχουν ενδεχόμενα ανωμαλιών θα πρέπει να ληφθούν υπόψη από τα προγράμματα / εφαρμογές της βάσης δεδομένων.

1.3 Τιμές NULL σε πλειάδες

■ ΟΔΗΓΙΑ 3:

- Οι σχέσεις θα πρέπει να σχεδιάζονται έτσι ώστε να ελαχιστοποιείται η ύπαρξη τιμών NULL
- Γνωρίσματα με τιμές NULL θα πρέπει να τοποθετούνται σε άλλες σχέσεις (με το πρωτεύον κλειδί της αρχικής σχέσης)

■ Η χρήση τιμών NULL γίνεται συνήθως όταν:

- Κάποια γνωρίσματα δεν σχετίζονται με κάποια οντότητα
- Η τιμή του γνωρίσματος δεν είναι γνωστή (αλλά μπορεί να υπάρχει)
- Η τιμή γνωρίζουμε ότι υπάρχει αλλά δεν είναι διαθέσιμη

1.4 Δημιουργία Πλασματικών Πλειάδων – Αποφυγή πάση θυσία

- Κακός σχεδιασμός σε μια βάση δεδομένων μπορεί να προκαλέσει λανθασμένα αποτελέσματα κατά την εφαρμογή συνενώσεων (join).
- Η ιδιότητα της συνένωσης χωρίς απώλειες εγγυάται τα σωστά αποτελέσματα των συνενώσεων
- ΟΔΗΓΙΑ 4:
 - Οι σχέσεις θα πρέπει να σχεδιάζονται ώστε να ικανοποιούν την ιδιότητα της συνένωσης χωρίς απώλειες .
 - Δεν θα πρέπει να δημιουργούνται πλασματικές πλειάδες όταν εφαρμόζουμε φυσικές συνενώσεις (natural joins) στις σχέσεις μας.

Πλασματικές Πλειάδες (2)

- Υπάρχουν δύο Σημαντικές Ιδιότητες στις αποσυνθέσεις:
 - a) Μη προσθετική (ή χωρίς απώλειες των αντίστοιχων συνενώσεων)
 - b) Διατήρηση των Συναρτησιακών Εξαρτήσεων.
- Η πρώτη ιδιότητα είναι πολύ σημαντική και δεν θα πρέπει να παραβιάζεται.
- Η δεύτερη ιδιότητα δεν είναι τόσο αυστηρή και μερικές φορές θα μπορούσαμε να μην την διατηρήσουμε

2. Συναρτησιακές Εξαρτήσεις

- Συναρτησιακές Εξαρτήσεις(ΣΕ)
 - Χρησιμοποιούνται σαν ένα μέτρο ποιότητας ενός σχεδιασμού μιας βάσης δεδομένων
 - Μαζί με την έννοια των κλειδιών χρησιμοποιούνται για τον ορισμό **κανονικών μορφών** σχέσεων
 - Είναι **περιορισμοί** που προκύπτουν από τη σημασία και τις συσχετίσεις των γνωρισμάτων μιας σχέσης
- Ένα σύνολο γνωρισμάτων X καθορίζει (συναρτησιακά) ένα σύνολο γνωρισμάτων Y αν ή τιμή του X καθορίζει μοναδική τιμή για το Y

2.1 Ορισμός Συναρτησιακών Εξαρτήσεων

- Η $\Sigma E \ X \rightarrow Y$ ισχύει όταν κάθε φορά που δύο πλειάδες έχουν την ίδια τιμή για το X έχουν την ίδια τιμή και για το Y
 - Για κάθε πλειάδες $t1$ and $t2$ σε κάθε στιγμιότυπο $r(R)$: Αν $t1[X]=t2[X]$, τότε $t1[Y]=t2[Y]$
- Η $\Sigma E \ X \rightarrow Y$ στην σχέση R καθορίζει ένα περιορισμό σε όλα τα στιγμιότυπα ή καταστάσεις $r(R)$
- Οι ΣE εξάγονται από περιορισμούς που υπάρχουν στον πραγματικό κόσμο που αναπαριστούν τα γνωρίσματα.

Παραδείγματα Περιορισμών ΣΕ (1)

- Ο αριθμός ταυτότητας καθορίζει το Όνομα Εργαζομένου
 - $AP_TAYT \rightarrow EP_ONOMA$
- Ο κωδικός έργου καθορίζει το όνομα του έργου και την θέση του.
 - $K\Omega\Delta_EP\Gamma OY \rightarrow \{ONOMA_EP\Gamma OY, TO\P_EP\Gamma OY\}$
- Ο αριθμός ταυτότητας εργαζομένου και ο κωδικός έργου καθορίζουν τις ώρες την εβδομάδα που ο εργαζόμενος εργάζεται στο έργο,
 - $\{AP_TAYT, K\Omega\Delta_EP\Gamma OY\} \rightarrow \Omega P E \Sigma$

Παραδείγματα Περιορισμών ΣΕ (2)

- Μια Συναρτησιακή Εξάρτηση είναι μια ιδιότητα των γνωρισμάτων μιας σχέσης R
- Ο περιορισμός πρέπει να ισχύει σε όλα τα στιγμιότυπα $r(R)$
- Αν K είναι ένα κλειδί της R , τότε το K καθορίζει όλα τα γνωρίσματα της R ($K \rightarrow R$)
(δεν υπάρχουν δύο διαφορετικές πλειάδες με $t_1[K]=t_2[K]$)

Καθορίζοντας ΣΕ από στιγμιότυπα Σχέσεων

- Για να καθορίσουμε ποιες ΣΕ ισχύουν σε κάποια Σχέση πρέπει να κατανοήσουμε τη σημασία των γνωρισμάτων και πως σχετίζονται μεταξύ τους.
- Αν εξετάσουμε ένα στιγμιότυπο μιας σχέσης μπορούμε να συμπεράνουμε ότι ενδέχεται να ισχύει μια ΣΕ μεταξύ κάποιων γνωρισμάτων
- Μπορούμε με βεβαιότητα να συμπεράνουμε ότι κάποια ΣΕ δεν ισχύει γιατί υπάρχουν πλειάδες που την παραβιάζουν.

Αποκλείοντας ΣΕ

Στο ακόλουθο παράδειγμα της σχέσης TEACH μπορούμε να πούμε ότι η ΣΕ: *Text* → *Course* μπορεί να ισχύει. Όμως οι ΣΕ *Teacher* → *Course*, *Teacher* → *Text* and *Couse* → *Text* αποκλείεται να ισχύουν.

TEACH

Teacher	Course	Text
Smith	Data Structures	Bartram
Smith	Data Management	Martin
Hall	Compilers	Hoffman
Brown	Data Structures	Horowitz

Ποιες ΣΕ μπορεί να ισχύουν ?

- Στο ακόλουθο στιγμιότυπο της σχέση $R(A, B, C, D)$.
- Ποιες ΣΕ μπορεί ή δεν μπορεί να ισχύουν ;
- $A \rightarrow B$;
- $B \rightarrow C$;
- $C \rightarrow D$;
- $AB \rightarrow C$;
- $BC \rightarrow D$;

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

3.1 Κανονικοποίηση Σχέσεων (1)

- **Κανονικοποίηση (Normalization):**
 - Η διαδικασία αποσύνθεσης ανεπιθύμητων (προβληματικών) σχέσεων διασπώντας τα γνωρίσματά τους σε μικρότερες σχέσεις.
- **Κανονική Μορφή (Normal form):**
 - Η κανονική μορφή μιας σχέσης είναι η υψηλότερη κανονική μορφή που ικανοποιεί μια σχέση βάση των κλειδιών και των ΣΕ που ισχύουν στην σχέση.

Κανονικοποίηση Σχέσεων (2)

- 1NF
 - Απαιτείται από όλα τα σχήματα σχέσης
- 2NF, 3NF, BCNF
 - Βασίζονται στα κλειδιά και τις ΣΕ ενός σχήματος σχέσης.
- 4NF
 - Βασίζονται στα κλειδιά και τις πλειότιμες εξαρτήσεις (ΠΕ) (multi-valued dependencies - MVDs)
- 5NF
 - Βασίζονται στα κλειδιά και στις εξαρτήσεις σύνδεσης (join dependencies - JDs))
- Επιπλέον συνθήκες μπορεί να χρειάζονται για να εξασφαλίσουν ένα καλό σχεσιακό σχεδιασμό (συνένωση χωρίς απώλειες (lossless join), με διατήρηση εξαρτήσεων (dependency preservation))

3.2 Πρακτική Χρήση Κανονικών Μορφών

- Η κανονικοποίηση εφαρμόζεται στην πράξη ώστε οι σχεδιασμοί να έχουν υψηλή ποιότητα και να ικανοποιούν επιθυμητές ιδιότητες.
- Η χρησιμότητα των κανονικών μορφών μπορεί να αμφισβητηθεί όταν οι περιορισμοί στους οποίους βασίζονται είναι δύσκολο να κατανοηθούν ή να ανιχνευθούν.
- Οι σχεδιαστές δεν χρειάζεται πάντα να κανονικοποιούν στην πιο υψηλή κανονική μορφή.
 - (συνήθως μέχρι 3NF και BCNF είναι αρκετό.).
- **Αποκανονικοποίηση (Denormalization):**
 - Η τακτική να αποθηκεύεται η συνένωση (join) σχέσεων υψηλότερης κανονικής μορφής σαν η βασική σχέση (η οποία θα βρίσκεται σε χαμηλότερη κανονική μορφή)

3.3 Ορισμοί Κλειδιών και Γνωρισμάτων που συμμετέχουν σε κλειδιά (1)

- Ένα **υπερκλειδί** ενός σχήματος σχέσεως $R = \{A_1, A_2, \dots, A_n\}$ είναι ένα σύνολο γνωρισμάτων S υποσυνόλου του R με την ιδιότητα να μην υπάρχουν σε κάθε νόμιμη κατάσταση r της R δύο ξεχωριστές πλειάδες t_1 and t_2 με $t_1[S] = t_2[S]$
- Ένα **κλειδί** K είναι ένα **υπερκλειδί** όπου η αφαίρεση κάποιου γνωρίσματος από το K θα του αφαιρέσει την ιδιότητα να είναι υπερκλειδί.

Ορισμοί Κλειδιών και Γνωρισμάτων που συμμετέχουν σε κλειδιά (2)

- Εάν ένα σχήμα σχέσης έχει περισσότερα από ένα κλειδιά τότε καθένα από αυτά λέγεται **υποψήφιο κλειδί**.
 - Ένα από τα υποψήφια κλειδιά καθορίζεται να είναι το **πρωτεύων κλειδί**, και τα υπόλοιπα καλούνται **δευτερεύοντα κλειδιά**.
- Ένα **πρωτεύον γνώρισμα (primary attribute)** πρέπει να ανήκει σε κάποιο υποψήφιο κλειδί.
- Ένα **μη-πρωτεύον γνώρισμα (Nonprime attribute)** δεν είναι πρωτεύον γνώρισμα—δηλ δεν ανήκει σε κάποιο υποψήφιο κλειδί.

3.4 Πρώτη Κανονική Μορφή

- Δεν επιτρέπει
 - σύνθετα γνωρίσματα (composite attributes)
 - πλειότιμα γνωρίσματα (multivalued attributes)
 - **εμφωλευμένες σχέσεις (nested relations)**
γνωρίσματα που για κάποια πλειάδα δεν έχουν ατομικές τιμές.
- Θεωρείται ότι είναι μέρος του ορισμού μιας σχέσης.
- Παραδοσιακά τα περισσότερα ΣΔΒΔ επέτρεπαν μόνο τον ορισμό σχέσεων σε 1NF.(σημείωση)

Κανονικοποίηση σε Πρώτη Κανονική Μορφή

(a)

DEPARTMENT

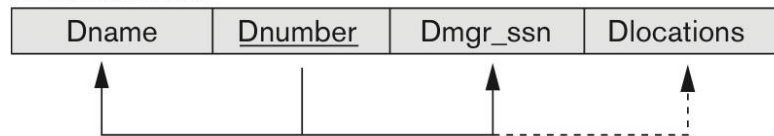


Figure 14.9

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.

(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

Κανονικοποίηση Εμφωλευμένων Σχέσεων σε Πρώτη Κανονική Μορφή.

(a)

EMP_PROJ		Projs	
Ssn	Ename	Pnumber	Hours

(b)

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 14.10

Normalizing nested relations into 1NF. (a) Schema of the EMP_PROJ relation with a *nested relation* attribute PROJS. (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.

3.5 Δεύτερη Κανονική Μορφή(1)

- Χρησιμοποιώντας τις έννοιες των **ΣΕ** και του Πρωτεύοντος Κλειδιού.
- Ορισμοί
 - **Πρωτεύον Γνώρισμα:** Ένα γνώρισμα που ανήκει στο Πρωτεύον Κλειδί.
 - **Πλήρης (full) Συναρτησιακή Εξάρτηση:** μια ΣΕ $Y \rightarrow Z$ όπου αφαίρεση κάποιου γνωρίσματος από το Y σημαίνει ότι η ΣΕ παύει να ισχύει.
- Παραδείγματα:
 - $\{ΑΡΙΘ_ΤΑΥΤ, ΚΩΔ_ΕΡΓΟΥ\} \rightarrow ΩΡΕΣ$ είναι πλήρης ΣΕ γιατί δεν ισχύει ούτε η $ΑΡΙΘ_ΤΑΥΤ \rightarrow ΩΡΕΣ$ ούτε η $ΚΩΔ_ΕΡΓΟΥ \rightarrow ΩΡΕΣ$
 - $\{ΑΡΙΘ_ΤΑΥΤ, ΚΩΔ_ΕΡΓΟΥ\} \rightarrow ΕΡ_ΟΝΟΜΑ$ δεν είναι πλήρης (καλείται μερική (partial) γιατί η $ΑΡΙΘ_ΤΑΥΤ \rightarrow ΕΡ_ΟΝΟΜΑ$ ισχύει.

Δεύτερη Κανονική Μορφή (2)

- Μια σχέση R είναι σε δεύτερη κανονική μορφή (**2NF**) εάν κάθε μη πρωτεύον γνώρισμα A της R καθορίζεται πλήρως από το πρωτεύον κλειδί.
- Η σχέση R μπορεί να διασπασθεί σε σχέσεις 2NF relations με μια διαδικασία που λέγεται κανονικοποίηση δεύτερης κανονικής μορφής.

Δείγματα προβληματικών σχέσεων

Redundancy

EMP_DEPT

Ename	<u>Ssn</u>	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

Redundancy Redundancy

EMP_PROJ

<u>Ssn</u>	<u>Pnumber</u>	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

(a)

EMP_DEPT

Ename	<u>Ssn</u>	Bdate	Address	Dnumber	Dname	Dmgr_ssn
-------	------------	-------	---------	---------	-------	----------

(b)

EMP_PROJ

<u>Ssn</u>	<u>Pnumber</u>	Hours	Ename	Pname	Plocation
------------	----------------	-------	-------	-------	-----------

FD1: Ssn → Hours, Ename, Pname, Plocation

FD2: Pnumber → Hours, Ename, Pname, Plocation

FD3: Ssn → Pnumber

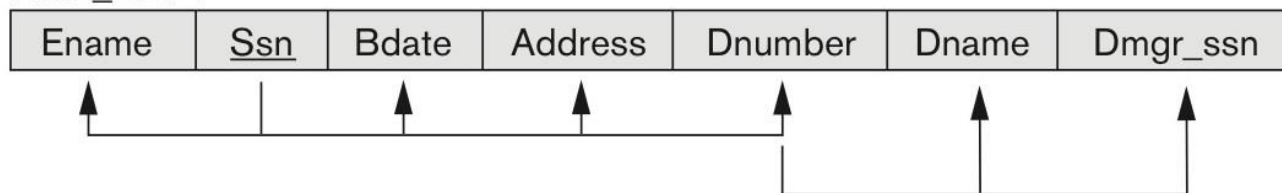
Παραδείγματα σχέσεων με ανωμαλίες τροποποίησης

Figure 14.3

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

(a)

EMP_DEPT



(b)

EMP_PROJ

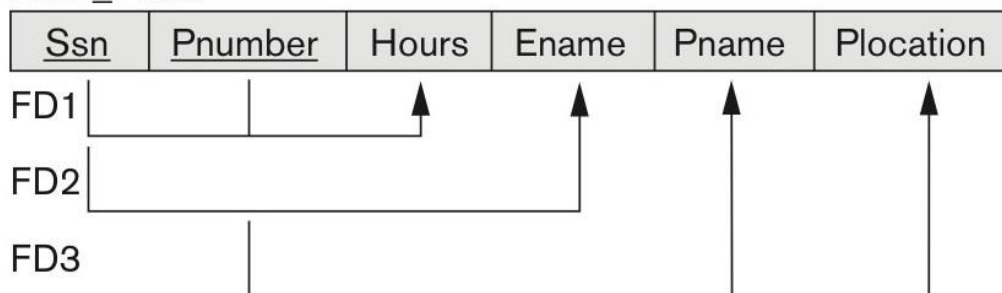
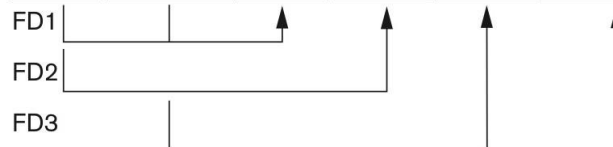


Figure 14.11 Κανονικοποίηση σε Δεύτερη Κανονική Μορφή.

(a)

EMP_PROJ

<u>Ssn</u>	<u>Pnumber</u>	Hours	Ename	Pname	Plocation
------------	----------------	-------	-------	-------	-----------



2NF Normalization

EP1

<u>Ssn</u>	<u>Pnumber</u>	Hours
------------	----------------	-------



EP2

<u>Ssn</u>	Ename
------------	-------



EP3

<u>Pnumber</u>	Pname	Plocation
----------------	-------	-----------



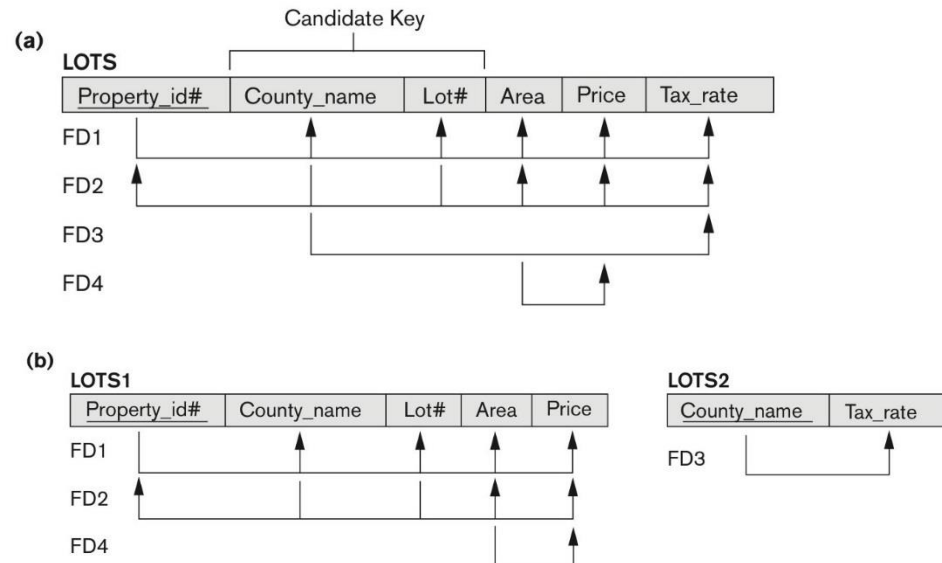
Figure 14.11

Normalizing into 2NF and 3NF.
(a) Normalizing EMP_PROJ into 2NF relations.

Figure 14.12 Κανονικοποίηση σε Δεύτερη Κανονική Μορφή (2)

Figure 14.12

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) Progressive normalization of LOTS into a 3NF design.



3.6 Τρίτη Κανονική Μορφή (1)

- Ορισμός:
 - **Μεταβατική Συναρτησιακή Εξάρτηση:** μια ΣΕ $X \rightarrow Z$ που μπορεί να προκύψει από δύο ΣΕ:
 $X \rightarrow Y$ and $Y \rightarrow Z$
- Παραδείγματα:
 - $ΑΡΙΘ_ΤΑΥΤ \rightarrow ΔΙΕΥΘΥΝΤΗΣ$ είναι μεταβατική ΣΕ γιατί
 - $ΑΡΙΘ_ΤΑΥΤ \rightarrow ΚΩΔ_ΤΜΗΜΑ$ και $ΚΩΔ_ΤΜΗΜΑ \rightarrow ΔΙΕΥΘΥΝΤΗΣ$
 - $ΑΡΙΘ_ΤΑΥΤ \rightarrow ΕΡ_ΟΝΟΜΑ$ είναι μη-μεταβατική γιατί
 - Δεν υπάρχουν γνωρίσματα X όπου $ΑΡ_ΤΑΥΤ \rightarrow X$ και $X \rightarrow ΕΡ_ΟΝΟΜΑ$

Τρίτη Κανονική Μορφή (2)

- Ένα σχήμα σχέσης είναι σε Τρίτη Κανονική Μορφή (**3NF**) εάν είναι σε 2NF και *κανένα μη πρωτεύον γνώρισμα A στην R δεν είναι μεταβατικά εξαρτώμενο από το πρωτεύον κλειδί.*
- Η σχέση R μπορεί να διασπαστεί σε σχέσης 3NF με κανονικοποίηση 3NF.
- Σημείωση:
 - Εάν $X \rightarrow Y$ and $Y \rightarrow Z$, με X το πρωτεύον κλειδί, δημιουργούν πρόβλημα μόνο όταν το Y δεν είναι υποψήφιο κλειδί.
 - Όταν το Y είναι υποψήφιο κλειδί δεν υπάρχει πρόβλημα με την μεταβατική εξάρτηση.
 - Π.Χ., Στη Σχέση ΕΡΓΑΖΟΜΕΝΟΣ (ΑΡ_ΤΑΥΤ, ΚΩΔ_ΕΡΓΑΖ, ΜΙΣΘΟΣ).
 - Ισχύει $ΑΡ_ΤΑΥΤ \rightarrow ΚΩΔ_ΕΡΓΑΖ \rightarrow ΜΙΣΘΟΣ$ και το ΚΩΔ_ΕΡΓΑΖ είναι υποψήφιο κλειδί.

Figure 14.11 Κανονικοποίηση σε Δεύτερη και Τρίτη Κανονική Μορφή.

Figure 14.11
Normalizing into 2NF and 3NF.
(a) Normalizing EMP_PROJ into 2NF relations.

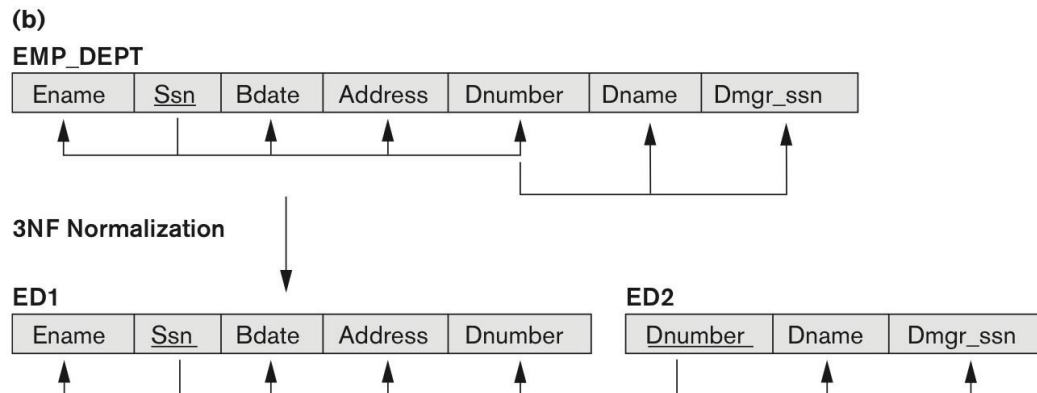
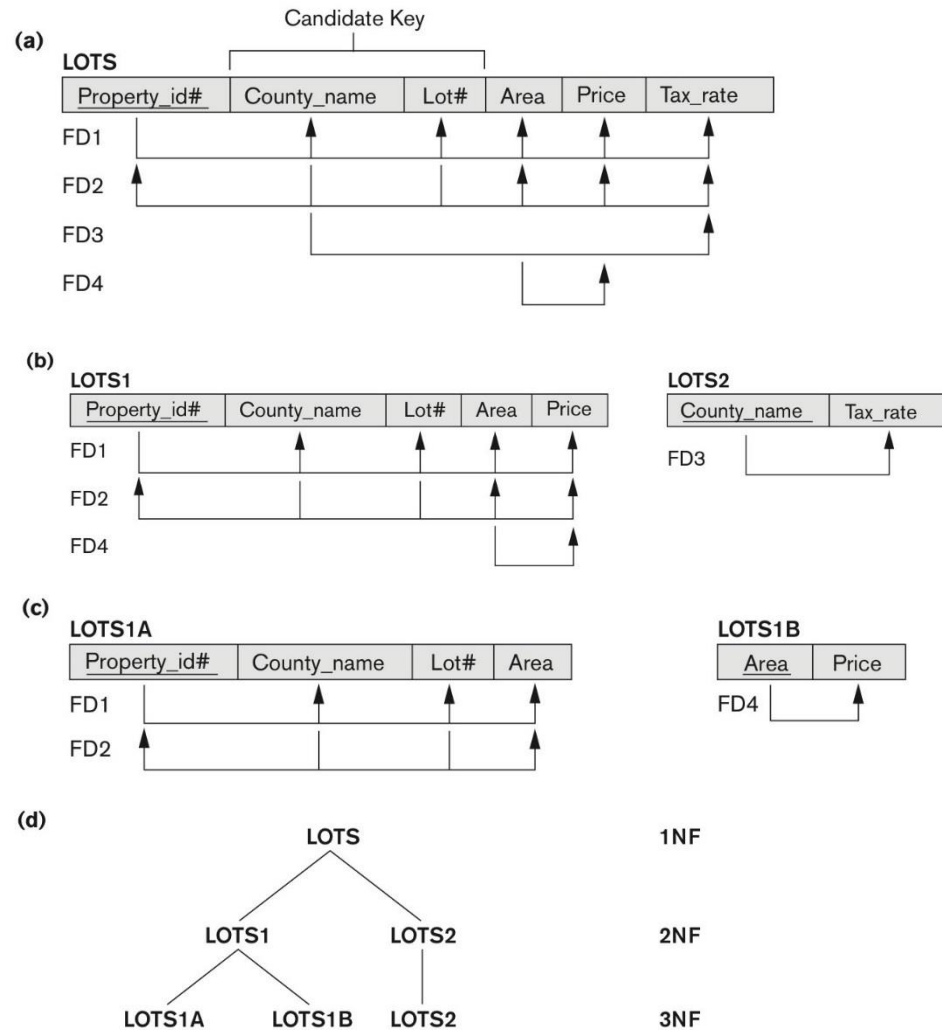


Figure 14.12 Κανονικοποίηση σε 2NF και 3NF

Figure 14.12

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) Progressive normalization of LOTS into a 3NF design.



Άτυποι ορισμοί κανονικών μορφών

- Πρώτη Κανονική Μορφή (1st normal form)
 - Όλα τα γνωρίσματα εξαρτώνται από **το κλειδί**
 - All attributes depend on **the key**
- Δεύτερη Κανονική Μορφή (2nd normal form)
 - Όλα τα γνωρίσματα εξαρτώνται από **όλο το κλειδί**
 - All attributes depend on **the whole key**
- Τρίτη Κανονική Μορφή (3rd normal form)
 - Όλα τα γνωρίσματα εξαρτώνται **μόνο** από το **κλειδί**
 - All attributes depend on **nothing but the key**

4. Γενικοί Ορισμοί Κανονικών Μορφών (Για πολλαπλά κλειδιά) (1)

- Οι προηγούμενοι ορισμοί έλαβαν υπόψη μόνο το πρωτεύον κλειδί
- Οι ακόλουθοι ορισμοί λαμβάνουν υπόψη σχέσεις με περισσότερα από ένα υποψήφια κλειδιά.
- Κάθε γνώρισμα που συμμετέχει σε κάποιο υποψήφιο κλειδί είναι ένα πρωτεύον γνώρισμα (prime attribute)
- Όλα τα άλλα γνωρίσματα ονομάζονται μη-πρωτεύοντα γνωρίσματα (non-prime attributes)

4.1 Γενικοί Ορισμοί για 2NF (Για πολλαπλά υποψήφια κλειδιά)

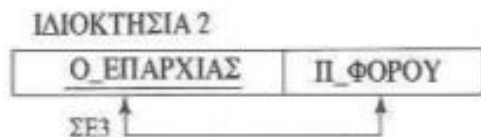
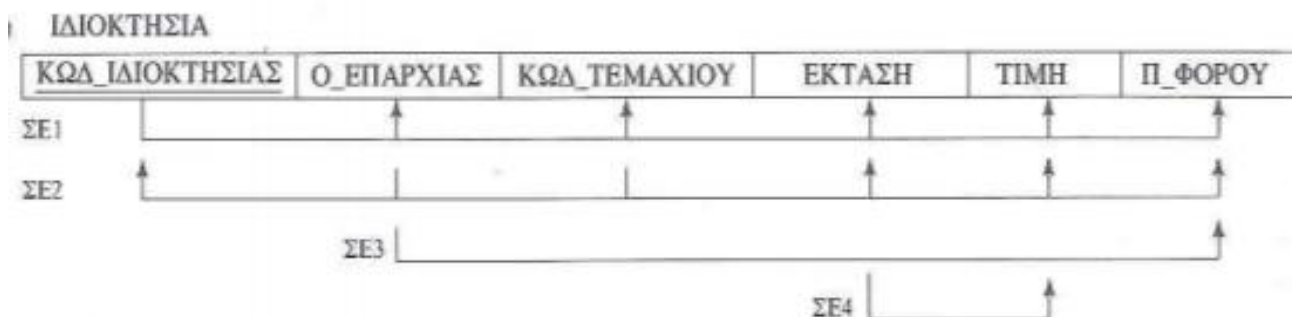
- Ένα σχήμα σχέσης R είναι σε δεύτερη κανονική μορφή (**2NF**) όταν κάθε μη-πρωτεύον γνώρισμα A στο R είναι πλήρως συναρτησιακά εξαρτώμενο από κάθε κλειδί του R
- Στο ακόλουθο σχήμα

Ο_ΕΠΑΡΧΙΑΣ $\rightarrow$ Π_ΦΟΡΟΥ παραβιάζει την 2NF.

Η κανονικοποίηση 2NF διασπά την ΙΔΙΟΚΤΗΣΙΑ σε
ΙΔΙΟΚΤΗΣΙΑ1 (ΚΩΔ_ΙΔΙΟΚΤΗΣΙΑΣ, Ο_ΕΠΑΡΧΙΑΣ, ΚΩΔ_ΤΕΜΑΧΙΟΥ,
ΕΚΤΑΣΗ, ΤΙΜΗ)

ΙΔΙΟΚΤΗΣΙΑ2 (ΚΩΔ_ΕΠΑΡΧΙΑΣ, Π_ΦΟΡΟΥ)

Κανονικοποίηση σε Δεύτερη Κανονική Μορφή



4.2 Γενικός Ορισμός Τρίτης Κανονικής Μορφής

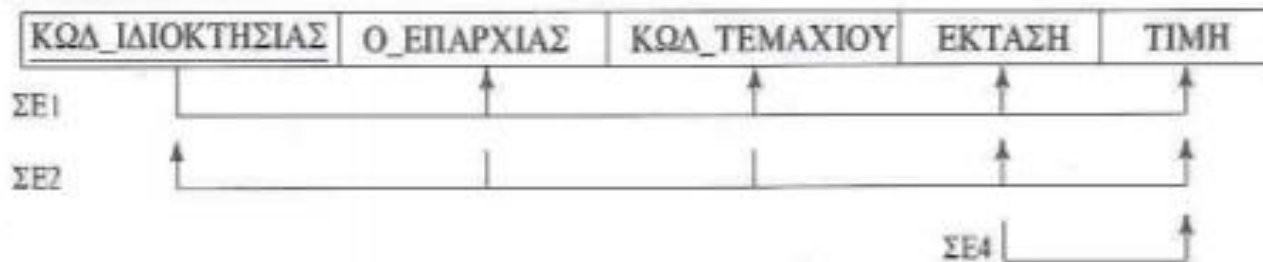
- Ορισμός:
 - **Υπερκλειδί** ενός σχήματος σχέσης R είναι ένα σύνολο γνωρισμάτων S του R που περιέχει ένα κλειδί του R
 - Μια συναρτησιακή εξάρτηση $A \rightarrow B$ λέγεται τετριμμένη όταν το B είναι υποσύνολο του A (η ΣΕ ισχύει πάντα)
 - Ένα σχήμα σχέσης R είναι σε Τρίτη Κανονική Μορφή όταν για κάθε μη τετριμμένη ΣΕ $X \rightarrow A$ που ισχύει στο R
 - (α) το X είναι υπερκλειδί του R , ή
 - (β) το A είναι πρωτεύον γνώρισμα του R
- Η σχέση ΙΔΙΟΚΤΗΣΙΑ1 δεν είναι σε 3NF γιατί $ΕΚΤΑΣΗ \rightarrow ΤΙΜΗ$ και το $ΕΚΤΑΣΗ$ δεν είναι υπερκλειδί της ΙΔΙΟΚΤΗΣΙΑ1. (see Figure 14.12).

4.3 Ερμηνεία του Γενικού Ορισμού της Τρίτης Κανονικής Μορφής.

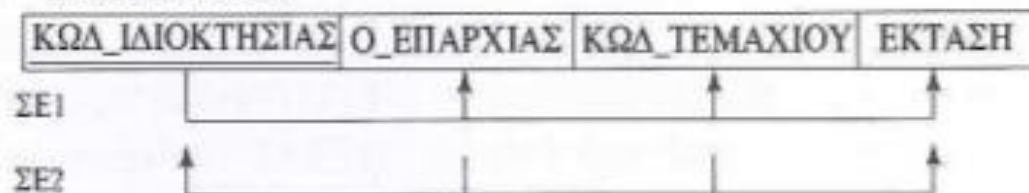
- Υπάρχουν 2 συνθήκες στον ορισμό της 3NF:
 - Για οποιαδήποτε μη τετριμμένη ΣΕ $X \rightarrow A$ στο R:
 - (α) το X είναι υπερκλειδί του R, ή
 - (β) το A είναι ένα πρωτεύον γνώρισμα του R
- Η πρώτη συνθήκη προστατεύει από δύο περιπτώσεις :
 - όταν ένα πρωτεύον γνώρισμα καθορίζει ένα μη πρωτεύον γνώρισμα. Προστατεύει από παραβίαση της 2NF.
 - όταν ένα μη πρωτεύον γνώρισμα καθορίζει ένα μη πρωτεύον γνώρισμα. Προστατεύει από παραβιάσεις λόγω μεταβατικών εξαρτήσεων.

Κανονικοποίηση σε Τρίτη Κανονική Μορφή

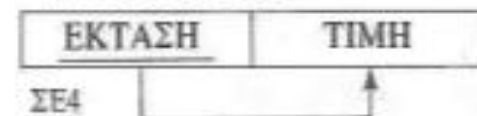
(β) ΙΔΙΟΚΤΗΣΙΑ 1



ΙΔΙΟΚΤΗΣΙΑ 1Α



ΙΔΙΟΚΤΗΣΙΑ 1Β



5. BCNF (Boyce-Codd Normal Form)

- Ένα σχήμα σχέσης R βρίσκεται σε κανονική μορφή **Boyce-Codd (BCNF)** αν για οποιαδήποτε (μη τετριμμένη) ΣΕ $X \rightarrow A$ του R το X είναι ένα υπερκλειδί στο R
- Κάθε κανονική μορφή είναι πιο περιοριστική από την προηγούμενη
 - Κάθε 2NF σχέση βρίσκεται σε 1NF
 - Κάθε 3NF σχέση βρίσκεται σε 2NF
 - Κάθε BCNF σχέση βρίσκεται σε 3NF
- Υπάρχουν σχέσεις που είναι σε 3NF αλλά όχι σε BCNF
- Η BCNF είναι πιο περιοριστική από την 3NF
- Ο στόχος είναι να έχουμε σχέσεις σε BCNF (ή σε 3NF)

Παράδειγμα Κανονικοποίησης Boyce-Codd

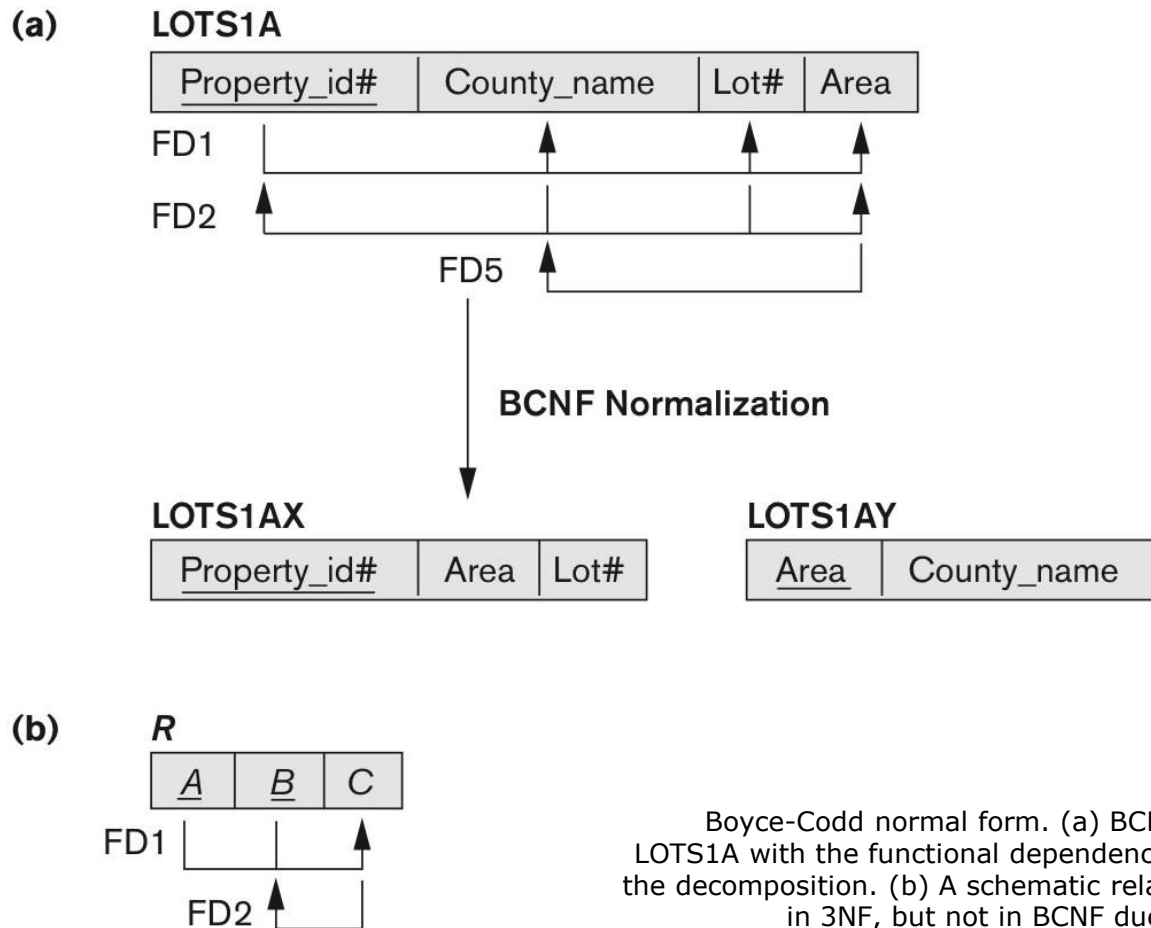


Figure 14.13
 Boyce-Codd normal form. (a) BCNF normalization of LOTS1A with the functional dependency FD2 being lost in the decomposition. (b) A schematic relation with FDs; it is in 3NF, but not in BCNF due to the f.d. $C \rightarrow B$.

Παράδειγμα σχέσης που βρίσκεται σε 3NF αλλά όχι σε BCNF

TEACH

Student	Course	Instructor
Narayan	Database	Mark
Smith	Database	Navathe
Smith	Operating Systems	Ammar
Smith	Theory	Schulman
Wallace	Database	Mark
Wallace	Operating Systems	Ahamad
Wong	Database	Omiecinski
Zelaya	Database	Navathe
Narayan	Operating Systems	Ammar

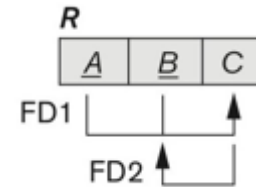
Συν. Εξαρτήσεις:
 $S, C \rightarrow I$
 $I \rightarrow C$



Figure 14.14
A relation TEACH that is in 3NF
but not BCNF.

Επιτυγχάνοντας BCNF με Αποσύνθεση (1)

- Υπάρχουν δύο ΣΕ στην σχέση TEACH:
 - fd1: { student, course} -> instructor
 - fd2: instructor -> course
- {student, course} είναι υποψήφιο κλειδί και οι εξαρτήσεις ακολουθούν το πρότυπο στο σχήμα.



- Αυτή ή σχέση είναι σε 3NF αλλά όχι σε BCNF
- Μια σχέση που δεν είναι σε BCNF πρέπει γενικά να αποσυντίθεται αν και μερικές φορές μπορεί να μην διατηρηθούν όλες οι εξαρτήσεις.

Επιτυγχάνοντας BCNF με Αποσύνθεση(2)

- Υπάρχουν τρεις πιθανές αποσυνθέσεις της σχέσης TEACH
 - D1: {student, instructor} και {student, course}
 - D2: {course, instructor} και {course, student}
 - D3: {instructor, course} και {instructor, student} ✓
- Και οι τρεις θα χάσουν την fd1.
 - Μπορούμε να αποδεχθούμε την απώλεια μιας συναρτησιακής εξάρτησης. Αλλά δεν θα πρέπει να θυσιάσουμε την μη προσθετική ιδιότητα (συνένωση χωρίς απώλειες).
- Από τις τρεις περιπτώσεις μόνο η τρίτη δεν θα δημιουργήσει πλασματικές πλειάδες μετά από συνένωση (join) και έτσι διαθέτει την μη προσθετική ιδιότητα.
- Ένας έλεγχος για το αν μια δυαδική αποσύνθεση (αποσύνθεση σε δύο σχέσεις) είναι μη προσθετική (χωρίς απώλειες) παρουσιάζεται σαν η ιδιότητα LJ1 (ή NBJ) .

Έλεγχος μη προσθετικής ιδιότητας Δυαδικής Αποσύνθεσης Σχέσης

- Έλεγχος για μη προσθετική ιδιότητα Δυαδικής Αποσύνθεσης Σχέσης:
 - **ΙΔΙΟΤΗΤΑ NJB (LJ1) (non-additive join test for binary decompositions):** Μια αποσύνθεση $D = \{R1, R2\}$ της R έχει την ιδιότητα συνένωσης χωρίς απώλειες ως προς ένα σύνολο συναρτησιακών εξαρτήσεων F στο R αν και μόνο αν είτε
 - Η ΣΕ $((R1 \cap R2) \rightarrow (R1 - R2))$ είναι στο F^+ , ή
 - Η ΣΕ $((R1 \cap R2) \rightarrow (R2 - R1))$ είναι στο F^+ .
- F^+ , είναι το σύνολο όλων των ΣΕ που μπορεί να προκύψουν από το F . Λέγεται κλειστότητα του F .

Έλεγχος μη προσθετικής ιδιότητας Δυαδικής Αποσύνθεσης Σχέσης

Αν εφαρμόσετε τον έλεγχο NJB (LJ1) στις τρεις αποσυνθέσεις 3 της σχέσης TEACH:

- D1 δίνει **Student** → Instructor ή **Student** → Course, καμία από τις οποίες δεν ισχύει.
- D2 δίνει **Course** → Instructor ή **Course** → Student, καμία από τις οποίες δεν ισχύει.
- Όμως στην D3 παίρνουμε **Instructor** → Course ή **Instructor** → Student.

Η **Instructor** → Course ισχύει και έτσι η ιδιότητα NJB ικανοποιείται και η D3 είναι μη προσθετική (επιθυμητή) αποσύνθεση.