



Διάλεξη 23: Κανονικοποίηση και Συναρτησιακές Εξαρτήσεις IV

Στην ενότητα αυτή θα μελετηθούν τα εξής επιμέρους θέματα:
Εισαγωγή στις έννοιες:

- Κανονικές Μορφές (BCNF, Τέταρτη/4NF, Πέμπτη/5NF)

Διδάσκων: Παναγιώτης Ανδρέου

Περιεχόμενο Διάλεξης

- Η Κανονική Μορφή Boyce-Codd (Boyce-Codd Normal Form (BCNF))
- Η 4η Κανονική Μορφή (4NF)
- Η 5η Κανονική Μορφή (5NF)

Boyce-Codd Κανονική Μορφή (BCNF)

- Ο **Raymond F. Boyce**, ήταν ερευνητής της **IBM** όπου με άλλους συναδέλφους **Donald D. Chamberlin**, και κάτω από την καθοδήγηση του **Edgar F. Codd** (του δημιουργού του Σχεσιακού Μοντέλου), **σύγγραψε** το **1974** την πρώτη έκδοση του προτύπου **SQL** (της **SEQUEL**).
 - “**SEQUEL: A Structured English Query Language**” D.D. Chamberlin and R.F. Boyce, Proc. ACM SIGMOD Workshop on Data Description, Access and Control, Ann Arbor, Michigan (May 1974) pages 249-264.
- Μεταξύ άλλων, **υπέδειξε** επίσης με τον Codd πως μπορεί να **βελτιστοποιηθεί** η **3NF**, σε κάποιες ειδικές περιπτώσεις, **πετυχαίνονται** ακόμη **χαμηλότερα επίπεδα πλεονασμού (redundancy)** σε ένα σχεσιακό σχήμα.

Boyce-Codd Κανονική Μορφή (BCNF)

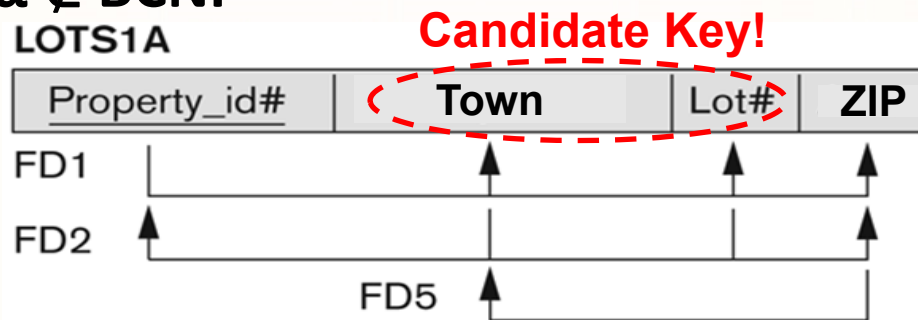
- Συγκεκριμένα, εισηγήθηκε μια **κανονική μορφή** η οποία μειώνει τον **πλεονασμό** σε περιπτώσεις που έχουμε **επικαλυπτόμενα σύνθετα κλειδιά (overlapping composite keys)**.
- Η κανονική μορφή αυτή ονομάζεται αργότερα **Boyce-Codd Normal Form** ή **BCNF**
- Η BCNF, όπως και οι **προκάτοχες κανονικές μορφές 2NF** και 3NF στηρίζονται στις έννοιες των κλειδιών και των **συναρτησιακών εξαρτήσεων**.

$$1NF \supseteq 2NF \supseteq 3NF \supseteq BCNF \supseteq 4NF \supseteq 5NF$$

Boyce-Codd Κανονική Μορφή (BCNF)

- Ορισμός BCNF: Ένα σχήμα R είναι σε **BCNF** εάν για κάθε μη-τετριμμένη FD $X \rightarrow A$ στο R, το X είναι υπερκλειδί της R

- Παράδειγμα \notin BCNF



Το ZIP δεν είναι Superkey! άρα δεν είναι BCNF
 Γιατί είναι 3NF;

- Γιατί οι \notin BCNF σχέσεις έχουν πρόβλημα;

- Γιατί εάν το X δεν είναι superkey, τότε δημιουργείται πλεονασμός

Redundancy! (circled in red dashed line)

Property_id	Town	Lot#	ZIP
1	Strovolos	1	2330
2	Lakatamia	2	2550
3	Strovolos	5	2330

Π.χ., πλεονασμός από ZIP → TOWN

Boyce-Codd Κανονική Μορφή (BCNF)

- Εάν στο προηγούμενο παράδειγμα το **ZIP** ήταν **κλειδί** (ή **υπερκλειδί**) τότε δεν θα μπορούσαμε να είχαμε δυο **πλειάδες** με τιμή **2330**
- ... συνεπώς θα **αποφεύγαμε** τον πλεονασμό που πρόκυπτε πριν λόγω του **ZIP**→**TOWN**.

<u>Property id</u>	Town	Lot#	ZIP
1	Strovolos	1	2330
2	Lakatamia	2	2550
3	Strovolos	5	2330

∉ BCNF

Διάσπαση

∈ BCNF

<u>Property id</u>	Lot#	ZIP
1	1	2330
2	2	2550
3	5	2330

<u>ZIP</u>	Town
2330	Strovolos
2550	Lakatamia

Boyce-Codd Κανονική Μορφή (BCNF)

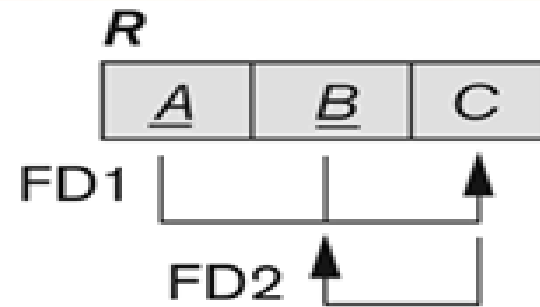
Παράδειγμα Σχέσης $\in 3NF$ αλλά $\notin BCNF$

{Student, Course} \rightarrow Instructor

Instructor \rightarrow Course

TEACH

<u>Student</u>	<u>Course</u>	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



Γιατί $R \in 3NF$;

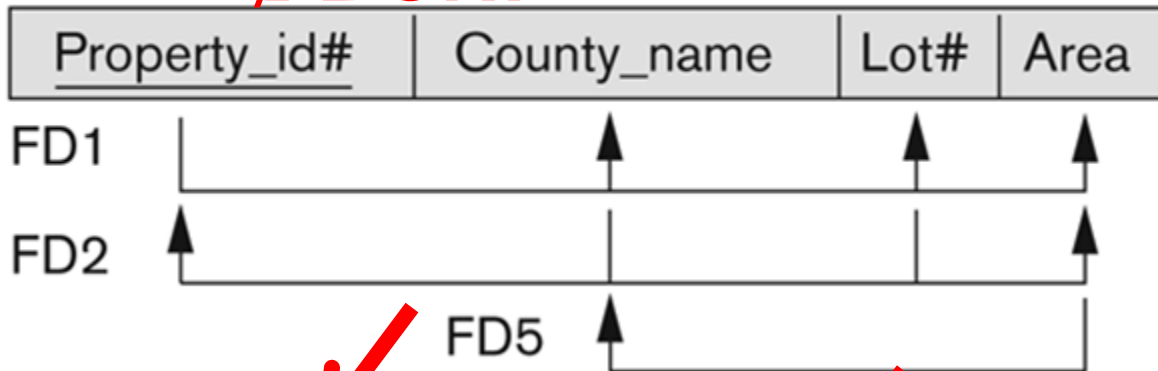
Διότι δεν υπάρχουν μεταβατικές εξαρτήσεις από το κλειδί.

Γιατί $R \notin BCNF$;
Διότι ο "Instructor" (C)
ΔΕΝ είναι superkey.

Λογική Διάσπασης σε BCNF

Λογική Διάσπασης σε BCNF: Για κάθε FD που παραβιάζει την **BCNF** (δηλ., $X \rightarrow Y$, όπου X δεν είναι υπερκλειδί), δημιουργήσε μια νέα σχέση $R(X \rightarrow Y)$, διατηρώντας στην αρχική σχέση το X ως ξένο κλειδί.

LOTS1A \notin BCNF



**Decomposition
(Διάσπαση)**

LOTS1AX

Property_id#	Area	Lot#
--------------	------	------

∈ BCNF

LOTS1AY

Area	County_name
------	-------------

∈ BCNF

Λογική Διάσπασης σε BCNF

- Σας δίνεται η σχέση BOOKING η οποία περιγράφει τις κρατήσεις κάποιων γηπέδων αθλητισμού.
- Εξηγήστε εάν **BOOKING** ∈ **BCNF** δεδομένων **{Court, StartTime} → EndTime** και **{Court, EndTime} → StartTime**

Today's Court Bookings		
Court	Start Time	End Time
1	9:30	10:30
1	11:00	12:00
1	14:00	15:30
2	10:00	11:30
2	11:30	13:30
2	15:00	16:30

Απάντηση

Εφόσον το αριστερό μέρος κάθε FD είναι superkey συνεπάγεται ότι η σχέση είναι σε BCNF

Κανονικοποίηση Ψηλότερου Βαθμού

- **2NF, 3NF, BCNF:** Στηρίζονται στα κλειδιά & συναρτησιακές εξαρτήσεις (FDs)
 - **4NF:** Στηρίζεται στα κλειδιά & Multi-valued dependencies (MVDs)
 - **5NF:** Στηρίζεται σε κλειδιά & Join Depend. (JDs)
- Πρακτικά, κάθε σχήμα βάσης **είναι καλό** να κανονικοποιήται σε **3NF** ή **BCNF** έτσι ώστε:
«Κάθε γνώρισμα να **εξαρτάται από το κλειδί**, πάνω σε **ολόκληρο το κλειδί** και **όχι** σε οτιδήποτε άλλο **εκτός από το κλειδί**» W. Kent (1992)
- Για να **διασφαλίσουμε** ένα καλό **σχεσιακό σχήμα** θα πρέπει να εγγυηθούμε και άλλες ιδιότητες (lossless join, dependency preservation)

Εισαγωγή σε Multi-valued Dependencies (MVDs)

- Υποθέστε την σχέση EMP_PROJ_DEP: Κάποιος υπάλληλος (Ename) δουλεύει σε κάποιο Project (Pname) και έχει εξαρτώμενους (Dname)

EMP_PROJ_DEP		
<u>Ename</u>	<u>Pname</u>	<u>Dname</u>
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John

Υποθέσεις:

- Ένας υπάλληλος μπορεί να δουλεύει σε πολλά projects
- Ένας υπάλληλος μπορεί να έχει πολλούς εξαρτώμενους

Ερώτηση: Η σχέση EMP_PROJ_DEP είναι BCNF;

- NAI**
- Η σχέση περιέχει μόνο κλειδιά
 - Δεν έχουμε δηλώσει κάποιες συναρτησιακές εξαρτήσεις (ούτε μπορούμε). Γιατί;

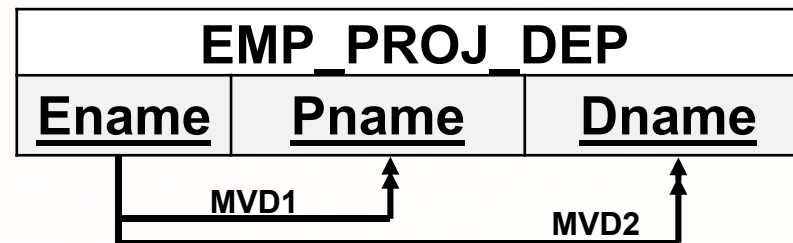
Εξαρτήσεις Πολλαπλών Τιμών (MVDs)

- Εξαρτήσεις Πολλαπλών Τιμών (MultiValued Dependencies MVDs)
- Μια Εξάρτηση Πολλαπλών Τιμών (MultiValued Dependency MVD) $X \twoheadrightarrow Y$ που ορίζεται σε κάποια σχέση R ($X \subseteq R, Y \subseteq R$) είναι ένας περιορισμός που καθορίζεται ως εξής: “Αν υπάρχουν δύο πλειάδες t_1, t_2 ώστε $t_1[X]=t_2[X]$, τότε πρέπει επίσης να υπάρχουν πλειάδες t_3, t_4 με τις ακόλουθες ιδιότητες ($Z = R - (X \cup Y)$):
 - $t_3[X]=t_4[X] = t_1[X]=t_2[X]$
 - $t_3[Y]=t_1[Y]$ και $t_4[Y]=t_2[Y]$
 - $t_3[Z]=t_2[Z]$ και $t_4[Z]=t_1[Z]$
- Θα αναφερόμαστε στα MVDs ως ακολούθως: Το X πολυπροσδιορίζει το Y
- Λόγο συμμετρίας, όταν $X \twoheadrightarrow Y$ τότε $X \twoheadrightarrow Z$ το οποίο αναγράφεται σαν $X \twoheadrightarrow Y|Z$

Τέταρτη Κανονική Μορφή (4NF)

Άτυπος Ορισμός 4NF: Όλες οι εξαρτήσεις (συναρτησιακές (FDs) ή πολλαπλών τιμών (MVDs)) πρέπει να εξαρτούνται αποκλειστικά από το κλειδί ή υπερκλειδί.

- Παράδειγμα \notin 4NF
- MVD1: $Ename \twoheadrightarrow Pname$
- MVD2: $Ename \twoheadrightarrow Dname$



- **Πότε συμβαίνει το πιο πάνω;**
- **Γιατί οι Non-4NF σχέσεις έχουν πρόβλημα;**
 - Γιατί και πάλι παραμένει πλεονασμός (**redundancy**)
 - Είναι πιο δύσκολο όμως να ανιχνευτεί

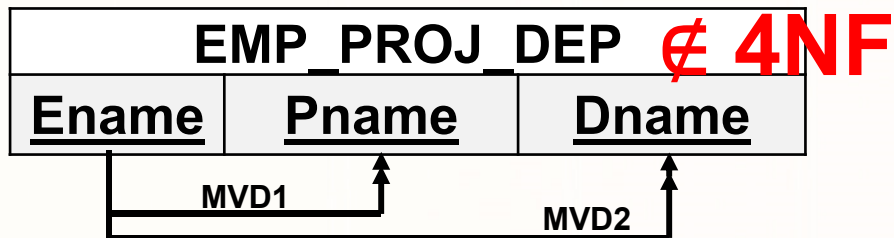
<u>Ename</u>	<u>Pname</u>	<u>Dname</u>
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John

Τέταρτη Κανονική Μορφή (4NF)

- Ας ορίσουμε την 4NF κάπως πιο τυπικά.
- **Τυπικός Ορισμός 4NF:** Μια σχέση R είναι σε 4NF εάν για κάθε **Εξάρτηση (FD ή MVD) $X \twoheadrightarrow Y$** που σχετίζεται με το R ισχύει ότι το X είναι υπερκλειδί της R.
- Ισχύουν τα ακόλουθα σημεία:
 - *Κάθε σχέση που αποτελείται μόνο από γνωρίσματα κλειδιά είναι σε BCNF αφού δεν αποτελείται από FDs*
 - *Αν μία σχέση όπως αναφέρεται πιο πάνω περιέχει μία MVD τότε δεν είναι σε 4NF*
- **Πότε συμβαίνει αυτό; Όταν αναπαριστούμε 2 ή περισσότερες 1:M σχέσεις στον ίδιο πίνακα**

Τέταρτη Κανονική Μορφή (4NF)

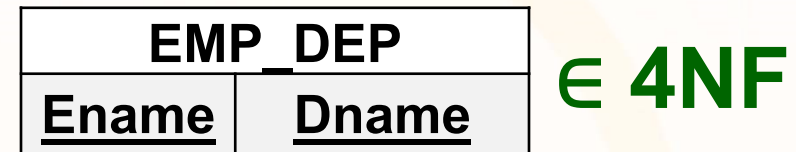
Λογική Διάσπασης σε 4NF: Για κάθε non-trivial MVD που παραβιάζει την 4NF (δηλ., $X \twoheadrightarrow Y$ όπου X δεν είναι υπερκλειδί), δημιουργήσε μια νέα σχέση $R(X \twoheadrightarrow Y)$ χωρίς να διατηρήσεις την αρχική σχέση.



MVD1: $\text{Ename} \twoheadrightarrow \text{Pname}$

MVD2: $\text{Ename} \twoheadrightarrow \text{Dname}$

**Decomposition
(Διάσπαση)**



Πέμπτη Κανονική Μορφή (5NF)

- Η **5NF** ορίζει ένα καινούριο είδος εξάρτησης το οποίο ονομάζεται **Join Dependency (JD)**
 - Για να ανιχνευθούν απαιτείται μεγάλη κατανόηση από τους σχεδιαστές/προγραμματιστές ΒΔ
 - Μπορούν να εκφραστούν πιο εύκολα με περιορισμούς προγραμματισμένους στην ΒΔ
- Τα JDs συμβολ. με $JD(R_1, R_2, \dots, R_n)$ και ορίζουν περιορισμούς σε καταστάσεις r μίας σχέσης R .
 - Προϋποθέτουν m -αδικές σχέσεις με $m \geq 3$
- Μπορούν να εκφραστούν με non-additive-join των πλειάδων κάθε $R_x \subset R: * (\pi_{R_1}(r), \pi_{R_2}(r), \dots, \pi_{R_n}(r)) = r$. Γι' αυτό ονομάζεται επίσης και **Project-Join Normal Form-PJNF**.

Πέμπτη Κανονική Μορφή (5NF)

- Παράδειγμα $\in 4NF$, $\notin 5NF$
- Περιορισμός: “Όταν κάποιος προμηθευτής s προμηθεύει κάποιο εξάρτημα p , και κάποιο project j χρησιμοποιεί το p , και ο s προμηθεύει τουλάχιστον ένα εξάρτημα στο j , τότε ο s προμηθεύει το εξάρτημα p στο project j ”
- Ο $JD(R1, R2, R3)$ μεταφράζεται σε 3 projections:
 $R1(Sname, PTname)$, $R_2(Sname, Pjname)$, $R_3(PTname, Pjname)$
- Προβλήματα: α) πλεονασμός, β) λάθη εισαγ.

SUPPLY		
<u>Sname</u>	<u>PTname</u>	<u>Pjname</u>
Smith	Bolt	PjX
Smith	Nut	PjY
Adam	Bolt	PjY
Walton	Nut	PjZ
Adam	Nail	PjX
Adam	Bolt	PjX
Smith	Bolt	PjY

Λογική Διάσπασης σε 5NF

Λογική Διάσπασης σε 5NF: Για κάθε non-trivial $JD(R_1, R_2, \dots, R_n)$ που παραβιάζει την **5NF** (δηλ., R_i δεν είναι υπερκλειδί της R), $\forall R_i$ δημιούργησε μια νέα σχέση, χωρίς να διατηρήσεις την R .


SUPPLY \notin 5NF		
<u>Sname</u>	<u>PTname</u>	<u>PJname</u>

$R1(Sname, PTname)$

$R2(Sname, PJname)$


$R3(PTname, PJname)$

**Decomposition
(Διάσπαση)**




R1	
<u>Sname</u>	<u>PTname</u>

\in 5NF



R2	
<u>Sname</u>	<u>PJname</u>

\in 5NF



R3	
<u>PTname</u>	<u>PJname</u>

\in 5NF

Διάσπαση σε 5NF

SUPPLY		
<u>Sname</u>	<u>PTname</u>	<u>PJname</u>
Smith	Bolt	PjX
Smith	Nut	PjY
Adam	Bolt	PjY
Walton	Nut	PjZ
Adam	Nail	PjX
Adam	Bolt	PjX
Smith	Bolt	PjY

R1	
<u>Sname</u>	<u>PTname</u>
Smith	Bolt
Smith	Nut
Adam	Bolt
Walton	Nut
Adam	Nail

R2	
<u>Sname</u>	<u>PJname</u>
Smith	PjX
Smith	PjY
Adam	PjY
Walton	PjZ
Adam	PjX

R3	
<u>PTname</u>	<u>PJname</u>
Bolt	PjX
Nut	PjY
Bolt	PjY
Nut	PjZ
Nail	PjX

- Η διάσπαση απαλείφει τον πλεονασμό
- Προσοχή:
 - Η **συνένωση μερικών projection** R_i μπορεί να δημιουργήσει **σποραδικές πλειάδες** (Παράδειγμα;)
 - Μόνο με την **συνένωση όλων των projection** R_i μπορούμε να δημιουργήσουμε την σχέση SUPPLY
 - Η **συνένωση πρέπει να δημιουργεί όλες τις πλειάδες της αρχικής σχέσης R**

Πέμπτη Κανονική Μορφή (5NF)

- Για να ανιχνευθούν τα JDs απαιτείται μεγάλη κατανόηση των δεδομένων από τους σχεδιαστές/προγραμματιστές ΒΔ
- Τα **MVDs** είναι ειδική περίπτωση των **JDs** με **n=2**
- **Βασικός στόχος** μας είναι η σχεδίαση ΒΔ σε **4NF**
- Με την εμπειρία σχεδιασμού και μελέτη των δεδομένων μπορούμε να επιτύχουμε την 5NF
- Πρακτικά, τα JDs μπορούν να εκφραστούν πιο εύκολα με περιορισμούς προγραμματισμένους στην ΒΔ (π.χ., check constraints βασισμένα σε UDFs ή SPs)