



## **EPL342 –Databases**

# **Lecture 23: Functional Dependencies and Normalization Normalization and Normal Forms (Chapter 14.3-14.4, Elmasri-Navathe 7ED)**

**Demetris Zeinalipour**

<http://www.cs.ucy.ac.cy/courses/EPL342>

# Περιεχόμενο Διάλεξης



- **Κανονικοποίηση (Normalization) και Κανονικές Μορφές (Normal Forms)**
- **Ορισμοί: Πρωτεύων Γνώρισμα (Prime Attribute), Μερική/Ολική Συναρτησιακή Εξάρτηση (Partial/Full FD)**
- **Κανονικές Μορφές**
  - Πρώτη Κανονική Μορφή (1NF)
  - Δεύτερη Κανονική Μορφή (2NF)
  - Τρίτη Κανονική Μορφή (3NF)

# Εισαγωγή: Κανονικοποίηση (Normalization)



- **Κανονικοποίηση (Normalization):**

- Η διαδικασία **διάσπασης** των σχέσεων μιας βάσης για να ελαχιστοποιηθεί η **Επανάληψη Δεδομένων**.

- Η επανάληψη είναι η πηγή ανωμαλιών ενημερώσεων
- Η διάσπαση γίνεται βάσει των **FDs + Κλειδιών**.

- Η συναρτησιακή εξάρτηση **TOWN → ZIP** στο ακόλουθο σχήμα προκαλεί την επανάληψη δεδομένων (**redundancy**)
  - Π.χ., οι διευθύνσεις στην ίδια περιοχή έχουν το ίδιο κώδικα (zip)

| <i>SSN</i> | <i>Name</i> | <i>Town</i> | <i>Zip</i> |
|------------|-------------|-------------|------------|
| 1234       | Joe         | Stony Brook | 11790      |
| 4321       | Mary        | Stony Brook | 11790      |
| 5454       | Tom         | Stony Brook | 11790      |
| .....      |             |             |            |

Redundancy

# Εισαγωγή: Κανονικές Μορφές (Normal Forms, NF)



- **Κανονικές Μορφές (Normal forms, NF):**

- Είναι συνθήκες οι οποίες **επικυρώνουν (certify)** τον **Βαθμό Χρηστότητας (Goodness Degree)** ενός Σχεσιακού Σχήματος.
- Οι **συνθήκες** ορίζονται με χρήση των **κλειδιών** και των συναρτησιακών εξαρτήσεων **FDs**.

- Η Κανονική Μορφή (NF) μιας Σχέσης αναφέρεται στη **ψηλότερη δυνατή NF** που είναι **εφικτή** για ένα **σχήμα**:

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

- Το **4NF** και **5NF** δεν είναι διαδεδομένα στη πράξη και δεν μελετηθούν ενώ οι **3NF** ή **BCNF** είναι η **επιδιωκόμενη μορφή**.
- Εάν οι **FDs ΔΕΝ** μπορούν να **εντοπιστούν** εύκολα τότε η **Κανονικοποίηση** γίνεται πρακτικά **δύσκολη** διαδικασία.

# Ορισμοί: Πρωτεύων Γνώρισμα (Definitions: Prime Attribute)



- Είχαμε αναφέρει ότι εάν ένα σχήμα έχει περισσότερα από ένα κλειδί τότε κάθε κλειδί ονομάζεται **εναλλακτικό κλειδί (candidate key)**
  - Ένα από αυτά είναι το **πρωτεύων κλειδί (primary key)** και τα υπόλοιπα τα **δευτερεύοντα (secondary keys)**.
  - Π.χ., (**SSN**, **PNO**, SID, Name) ή (SSN, **PNO**, **SID**, Name)
- **Πρωτεύων Γνώρισμα (Prime ή Key Attribute):**  
Γνώρισμα το οποίο είναι μέλος ενός Candidate key
  - Π.χ., πιο πάνω το **PNO** ή **SSN** ή **SID**
- **Μη-Πρωτεύων Γνώρισμα (Non-prime ή Non-Key):**  
Γνώρισμα το οποίο ΔΕΝ είναι μέλος κανενός Candidate key
  - Π.χ., πιο πάνω το **Name**

# Ορισμοί: Ολική/Μερική FD (Definitions: Full/Partial FD)



- **Ολική Συναρτησιακή Εξάρτηση (Full FD):** Μια FD  $Y \rightarrow Z$  όπου το  $Z$  εξαρτάται **πλήρως** από το  $Y$ .
  - Δηλαδή εάν **αφαιρεθεί οποιοδήποτε** γνώρισμα από το  $Y$  (δηλ., απλοποίηση του  $Y$ ) τότε η **FD ΔΕΝ** ισχύει πια.
  - Π.χ.,  $\{SSN, PNUMBER\} \rightarrow HOURS$ , διότι εάν απλοποιηθεί το  $\{SSN, PNumber\}$  τότε δεν ισχύει  **$SSN \rightarrow HOURS$**  ούτε και το  **$PNUMBER \rightarrow HOURS$**

- **Μερική Συναρτησιακή Εξάρτηση (Partial FD):** Μια εξάρτηση  $Y \rightarrow Z$  που **ΔΕΝ** είναι **Ολική**
  - Δηλαδή μπορεί να **αφαιρεθεί** κάποιο γνώρισμα από το  $Y$  και να **συνεχίσει** να **ισχύει** το **FD**.
  - Π.χ.,  $\{SSN, PNUMBER\} \rightarrow ENAME$  είναι μερική FD διότι εάν αφαιρεθεί το  $Pnumber$  τότε συνεχίζει να ισχύει το  **$SSN \rightarrow ENAME$**

# Ορισμοί: Μεταβατική FD (Definitions: Transitive FD)



- **Μεταβατική Συναρτησιακή Εξάρτηση (Transitive FD):** Μια FD  $X \rightarrow Z$  η οποία μπορεί να εξαχθεί από τις **FDs**  $X \rightarrow Y$  και  $Y \rightarrow Z$
- **Παραδείγματα:**
  - **SSN**  $\rightarrow$  **LetterGrade** είναι **Μεταβατική** FD
    - Διότι **SSN**  $\rightarrow$  **NumGrade** και **NumGrade**  $\rightarrow$  **LetterGrade**
  - **SSN**  $\rightarrow$  **DMGRSSN** είναι **Μεταβατική** FD
    - Διότι **SSN**  $\rightarrow$  **DNUMBER** and **DNUMBER**  $\rightarrow$  **DMGRSSN**
  - **SSN**  $\rightarrow$  **ENAME** **ΔΕΝ** είναι **Μεταβατική** (ισχύει εξ' όρισμου)
    - Διότι δεν υπάρχει σύνολο **non-key** γνωρισμάτων  $X^*$  τέτοιο ώστε **SSN**  $\rightarrow$  **X** and **X**  $\rightarrow$  **ENAME**
      - \* Εάν **X** είναι πρωτεύων γνώρισμα (π.χ.,  $X = \text{SID}$ ) τότε η μετάβαση της μορφής **SSN**  $\rightarrow$  **SID**, **SID**  $\rightarrow$  **ENAME** δεν ισχύει.

**Ορισμός Χρησιμεύει στο 3NF**

# Πρώτη Κανονική Μορφή ( 1NF: First Normal Form )



- Άτυπος Ορισμός 1NF: Κανένα Γνώρισμα ΔΕΝ είναι **πλειότιμο (multivalued)\*** γνώρισμα.

- Παράδειγμα  $\notin$  1NF

\* ούτε σύνθετο (composite) γνώρισμα, μόνο ατομικές τιμές!

DEPARTMENT

| Dname          | <u>Dnumber</u> | Dmgr_ssn  | Dlocations                     |
|----------------|----------------|-----------|--------------------------------|
| Research       | 5              | 333445555 | {Bellaire, Sugarland, Houston} |
| Administration | 4              | 98765432  | {Stafford}                     |
| Headquarters   | 1              | 888665555 | {Houston}                      |

- Παράδειγμα  $\in$  1NF

| 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          |

**Λογική Διάσπασης  
σε 1NF: Μετάτρεψε  
κάθε πλειότιμο σε νέα  
πλειάδα.**



# Πρώτη Κανονική Μορφή ( 1NF: First Normal Form )



- **Επισημάνσεις για 1NF**

- Πίνακες που δεν είναι σε **1NF** δεν είναι καν στο **σχεσιακό σχήμα** (στο οποίο κάθε γνώρισμα πρέπει να είναι ατομική τιμή).
- Το 1NF αναφέρεται κυρίως για **ιστορικούς λόγους** (δηλαδή ως το πρώτο βήμα των ισχυρότερων κανονικών μορφών 3NF και BCNF που χρησιμοποιούνται στην πράξη).

- **Πρόβλημα με 1NF**

- Συνεχίζουμε να έχουμε πλεονασμό πληροφορίας (π.χ., δεξ Dname, DMgr\_ssn)

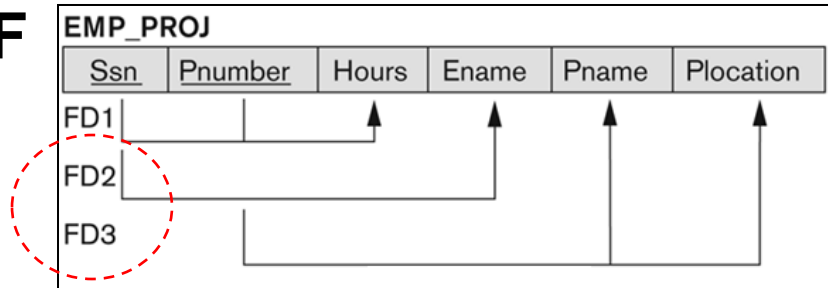
| DEPARTMENT     |                |                 |                  |
|----------------|----------------|-----------------|------------------|
| <u>Dname</u>   | <u>Dnumber</u> | <u>Dmgr_ssn</u> | <u>Dlocation</u> |
| Research       | 5              | 333445555       | Bellaire         |
| Research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

# Δεύτερη Κανονική Μορφή ( 2NF: Second Normal Form )



- Άτυπος Ορισμός 2NF: Κανένα Γνώρισμα **ΔΕΝ** εξαρτάται **μερικώς (partial dependence)** από οποιοδήποτε κλειδί (είτε είναι πρωτεύων ή δευτερεύων\*)

- Παράδειγμα  $\notin$  2NF



- Γιατί οι Non-2NF σχέσεις έχουν πρόβλημα;
  - Γιατί οι μερικές εξαρτήσεις (δηλ., FD2 και FD3) δημιουργούν πλεονασμό δεδομένων (redundancy), π.χ.,

| SSN | Pnumber | Hours | Ename    | Pname   | Plocation |
|-----|---------|-------|----------|---------|-----------|
| 1   | 1       | 3     | Costas   | Sensors | Nicosia   |
| 1   | 2       | 4     | Costas   | Web     | Limassol  |
| 2   | 2       | 5     | Christos | Web     | Limassol  |

Redundancy από FD2 → (1, 2, 4, Costas, Web, Limassol)

Redundancy από FD3 → (2, 2, 5, Christos, Web, Limassol)

\*Σημείωση: Το Κεφάλαιο 10.4 δίνει ορισμούς με χρήση και των δυο ειδών

# Δεύτερη Κανονική Μορφή ( 2NF: Second Normal Form )



- Ας ορίσουμε την 2NF κάπως πιο τυπικά.
- **Ορισμός 2NF:** Μια σχέση R είναι σε **2NF** εάν κάθε μη-πρωτεύων γνώρισμα (non-prime attribute) στο R είναι **ολικά συναρτησιακά** εξαρτώμενο από το πρωτεύων κλειδί.

- **Παράδειγμα  $\notin$  2NF**

(SSN, Pnumber, Hours, Ename)



# Δεύτερη Κανονική Μορφή (Λογική Διάσπασης σε 2NF)



**Λογική Διάσπασης σε 2NF:** Για κάθε μερική FD που παραβιάζει την **2NF** (δηλ.,  $X \rightarrow Y$ , όπου  $X$  *partial key*), δημιούργησε μια νέα σχέση  $R(X \rightarrow Y)$ , διατηρώντας στην αρχική σχέση το  $X$ .

EMP\_PROJ

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

∉ 2NF

FD1

FD2

FD3

Decomposition  
(Διάσπαση)

EP1

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

| FD1 |
|-----|
|-----|

∈ 2NF

EP2

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

| FD2 |
|-----|
|-----|

∈ 2NF

EP3

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

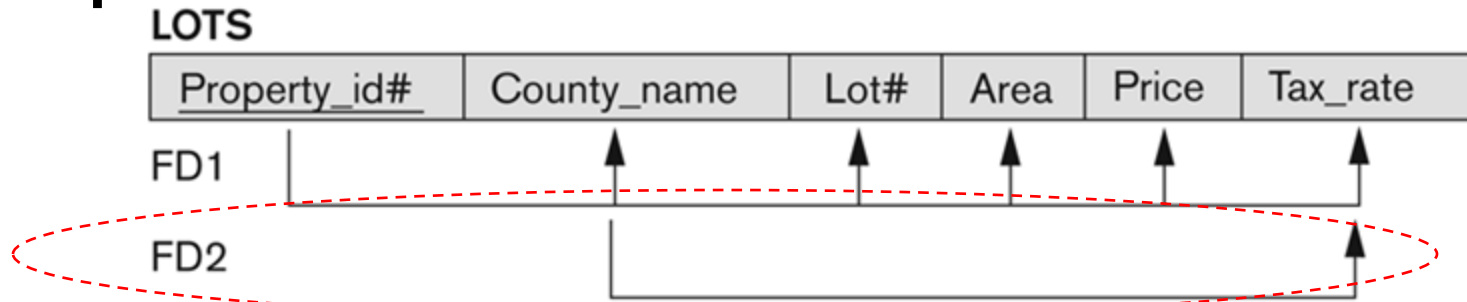
| FD3 |
|-----|
|-----|

∈ 2NF

# Τρίτη Κανονική Μορφή ( 3NF: Third Normal Form )



- **Άτυπος Ορισμός 3NF:** Κανένα Μη-Πρωτεύων Γνώρισμα **ΔΕΝ** εξαρτάται **μεταβατικά (transitive dependence)** από οποιοδήποτε **κλειδί** (είτε είναι πρωτεύων ή δευτερεύων\*)
- **Παράδειγμα ∉ 3NF**



- **Γιατί οι Non-3NF σχέσεις έχουν πρόβλημα;**
  - Γιατί και πάλι παραμένει πλεονασμός (**redundancy**), π.χ.,

| <u>Property_id</u> | County_name | Lot# | Area | Price | Tax_rate |
|--------------------|-------------|------|------|-------|----------|
| 1                  | Nicosia     | 1    | A    | 100   | 15%      |
| 2                  | Limassol    | 6    | C    | 120   | 10%      |
| 3                  | Nicosia     | 90   | F    | 130   | 15%      |

Redundancy!

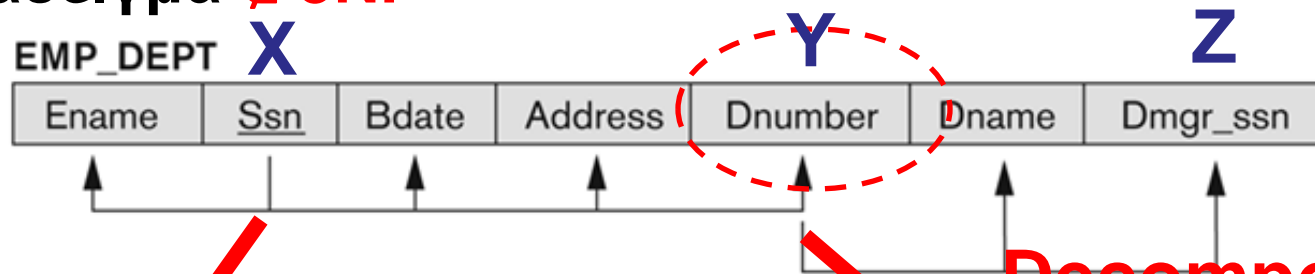
\*Σημείωση: Το Κεφάλαιο 10.4 δίνει ορισμούς με χρήση και των δυο ειδών.  
Επίσης όταν το τελικό αναφερόμενο (π.χ., Tax\_rate) είναι PRIME τότε είναι 3NF

# Τρίτη Κανονική Μορφή (Λογική Διάσπασης σε 3NF)

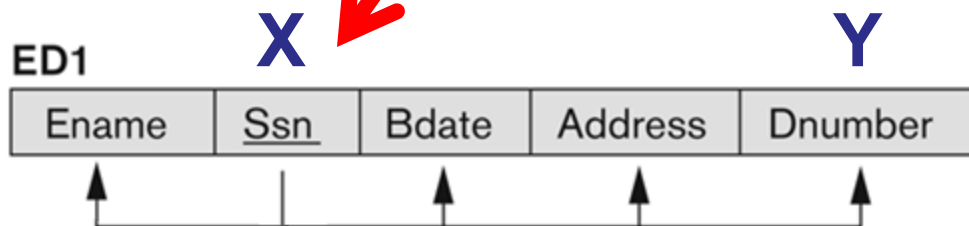


- **Λογική Διάσπασης σε 3NF:** Για κάθε FD που παραβιάζει την 3NF (δηλ.,  $X \rightarrow Y$ ,  $Y \rightarrow Z$ , όπου  $Y$  non-key), δημιούργησε δυο σχέσεις  $R1(X \rightarrow Y)$ ,  $R2(Y \rightarrow Z)$ , τοποθετώντας στην  $R1$  ως ξένο κλειδί το  $Y$ .

- Παράδειγμα  $\notin$  3NF

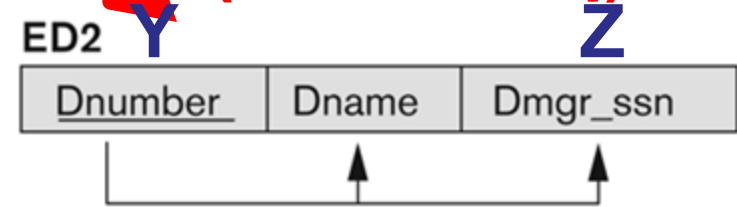


- Ίδιο Παράδειγμα  $\in$  3NF



**$\in$  3NF**

**Decomposition  
(Διάσπαση)**



**$\in$  3NF**

# Σύνοψη Κανονικών Μορφών

(Όπως θα πρέπει να τα θυμάστε...)

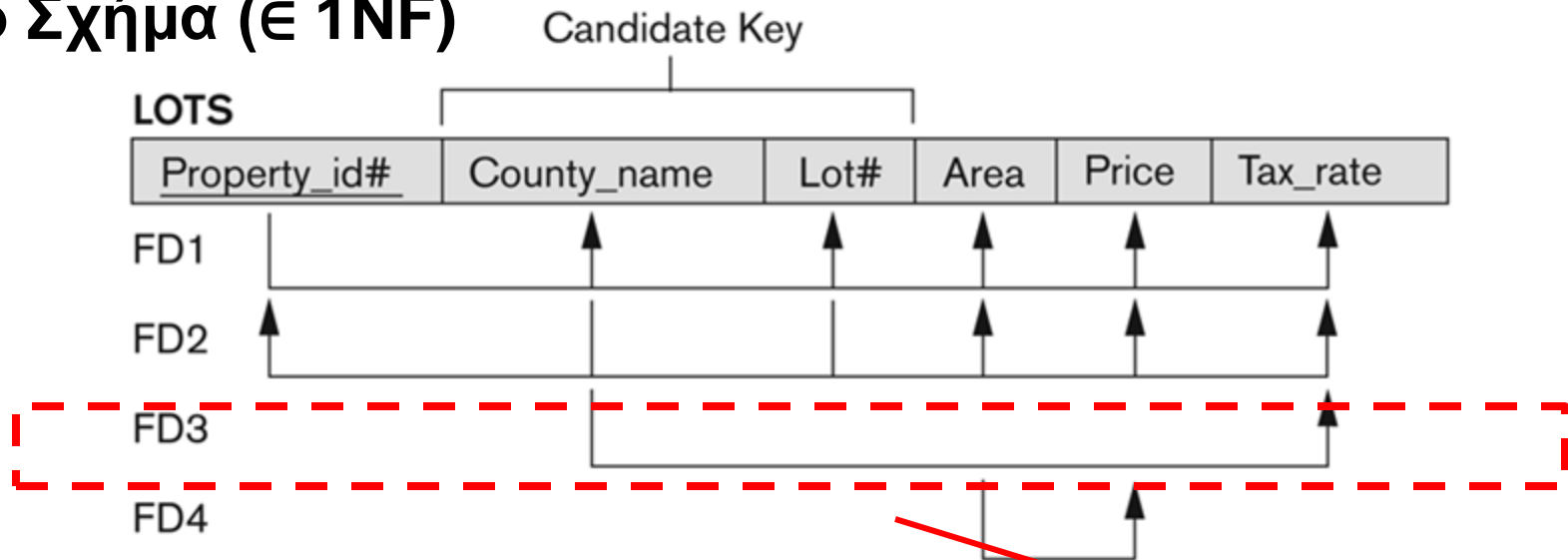


- **Πρώτη Κανονική Μορφή (1NF)**
  - Δεν υπάρχουν **Πλειότιμα (Multivalued)**
    - **Διαφορετικά Διατυπωμένο:** Όλα τα γνωρίσματα εξαρτώνται από το κλειδί.
- **Δεύτερη Κανονική Μορφή (2NF): επιπλέον 1NF:**
  - Δεν υπάρχουν **Μερικές (Partial) Εξαρτήσεις** από κλειδιά
    - **Διαφορετικά Διατυπωμένο:** Όλα τα γνωρίσματα εξαρτώνται από **Ολόκληρο** το κλειδί.
- **Τρίτη Κανονική Μορφή (3NF): επιπλέον 2NF:**
  - Δεν υπάρχουν **Μεταβατικές (Transitive) Εξαρτήσεις** από non-keys σε non-keys.

# Παράδειγμα Κανονικοποίησης σε 3NF

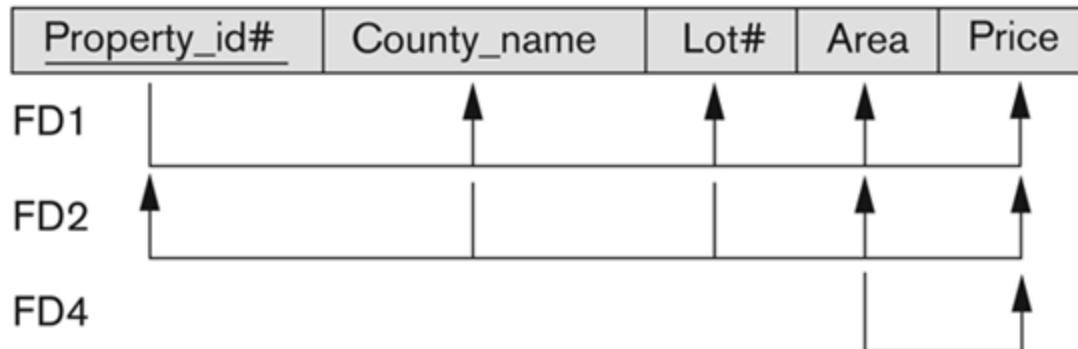


## Αρχικό Σχήμα (ε 1NF)

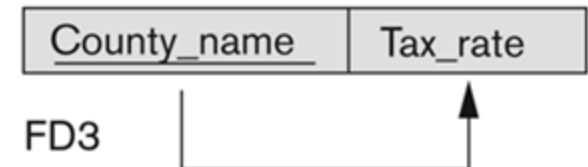


## Βήμα 1: Κανονικοποίηση σε 2NF

LOTS1



LOTS2



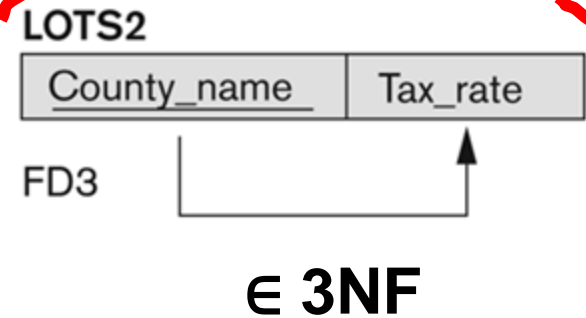
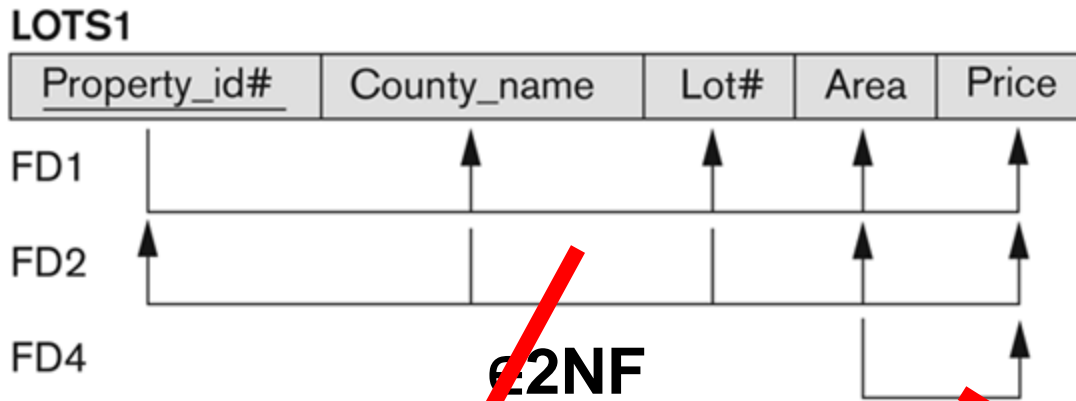
Μερική εξάρτηση από το Candidate Key



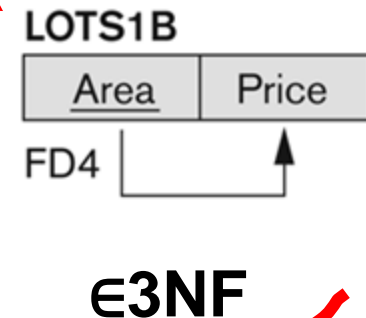
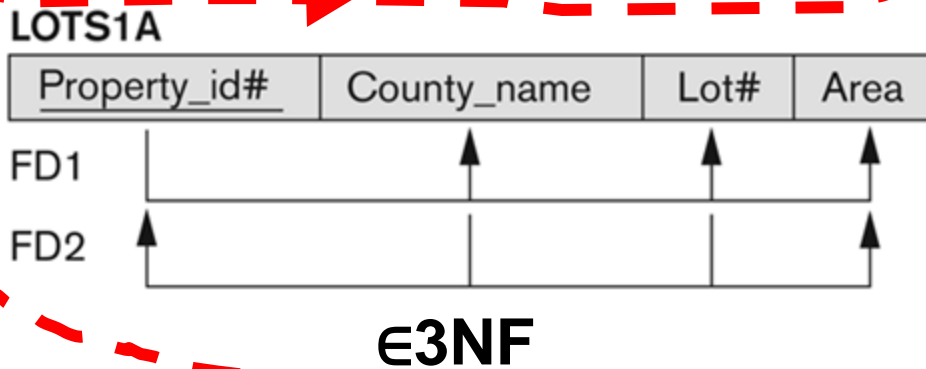
# Παράδειγμα Κανονικοποίησης σε 3NF



... Από προηγούμενη διαφάνεια (€ 2NF)



**Βήμα 2: Κανονικοποίηση σε 3NF**



# Παράδειγμα Κανονικοποίησης σε 3NF



## Η Αναδρομική Εκτέλεση της Κανονικοποίησης (Από πάνω προς τα κάτω)

