



Διάλεξη 22: Κανονικοποίηση και Συναρτησιακές Εξαρτήσεις III

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

- Κανονικοποίηση (Normalization) και Κανονικές Μορφές (Normal Forms)
- Ορισμοί Γνωρισμάτων/Σχέσεων
- Κανονικές Μορφές (Πρώτη/1NF, Δεύτερη/2NF, Τρίτη/3NF)

Διδάσκων: Παναγιώτης Ανδρέου

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 στην 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

Κανονικές Μορφές

- **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**:

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

- Οι **4NF** και **5NF** δεν είναι να επιτευχθούν από τους σχεδιαστές ΒΔ γιατί οι χρήστες δύσκολα μπορούν να εκφράσουν τις συναρτησιακές εξαρτήσεις αυτού του βαθμού
- Οι **BCNF, 4NF** είναι η **επιδιωκόμενη μορφή**.
- Εάν οι **FDs ΔΕΝ** μπορούν να **εντοπιστούν** εύκολα τότε η **Κανονικοποίηση** γίνεται πρακτικά **δύσκολη** διαδικασία.

Ορισμοί: Πρωτεύων Γνώρισμα

- 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**

Πρώτη Κανονική Μορφή (1NF: First Normal Form)

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

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

- Παράδειγμα \notin 1NF

DEPARTMENT

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

- Παράδειγμα \in 1NF

Dname	Dnumber	Dmgr_ssn	Dlocation
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., δεσ Dname, DMgr_ssn)

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

Πρώτη Κανονική Μορφή (1NF: First Normal Form)

Το 1NF δεν επιτρέπονται ούτε **εμφωλευμένες σχέσεις** π.χ.,
EMP_PROJ (SSN, Ename, {PROJS(Pnumber,Hours)})

Παράδειγμα Εμφωλευμένης «Σχέσης» Κανονικοποιημένες Σχέσεις

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

∉ 1NF

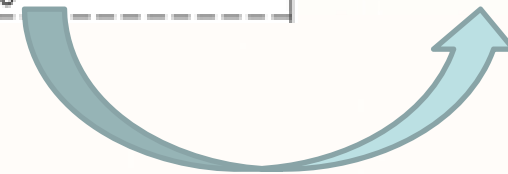
EMP_PROJ1

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

EMP_PROJ2

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

∈ 1NF



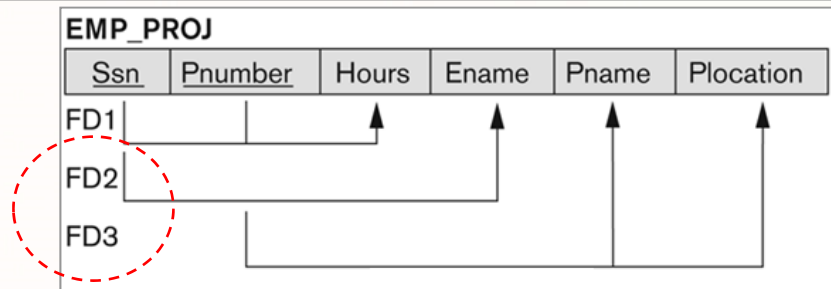
Ορισμοί: Ολική/Μερική FD (Full/Partial FD)

- **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 $SSN \rightarrow ENAME$.
- **Ο Ορισμός Χρησιμεύει στο 2NF**

Δεύτερη Κανονική Μορφή (2NF)

- 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 (points to the second row)

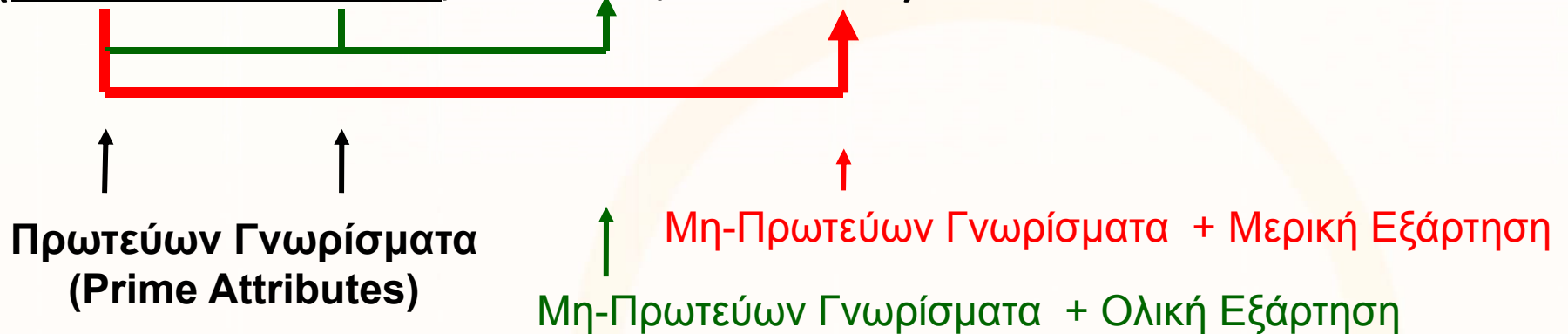
Redundancy από FD3 (points to the second and third rows)

*Σημείωση: Το Κεφάλαιο 14.4 δίνει ορισμούς με χρήση και των δυο ειδών

Δεύτερη Κανονική Μορφή (2NF)

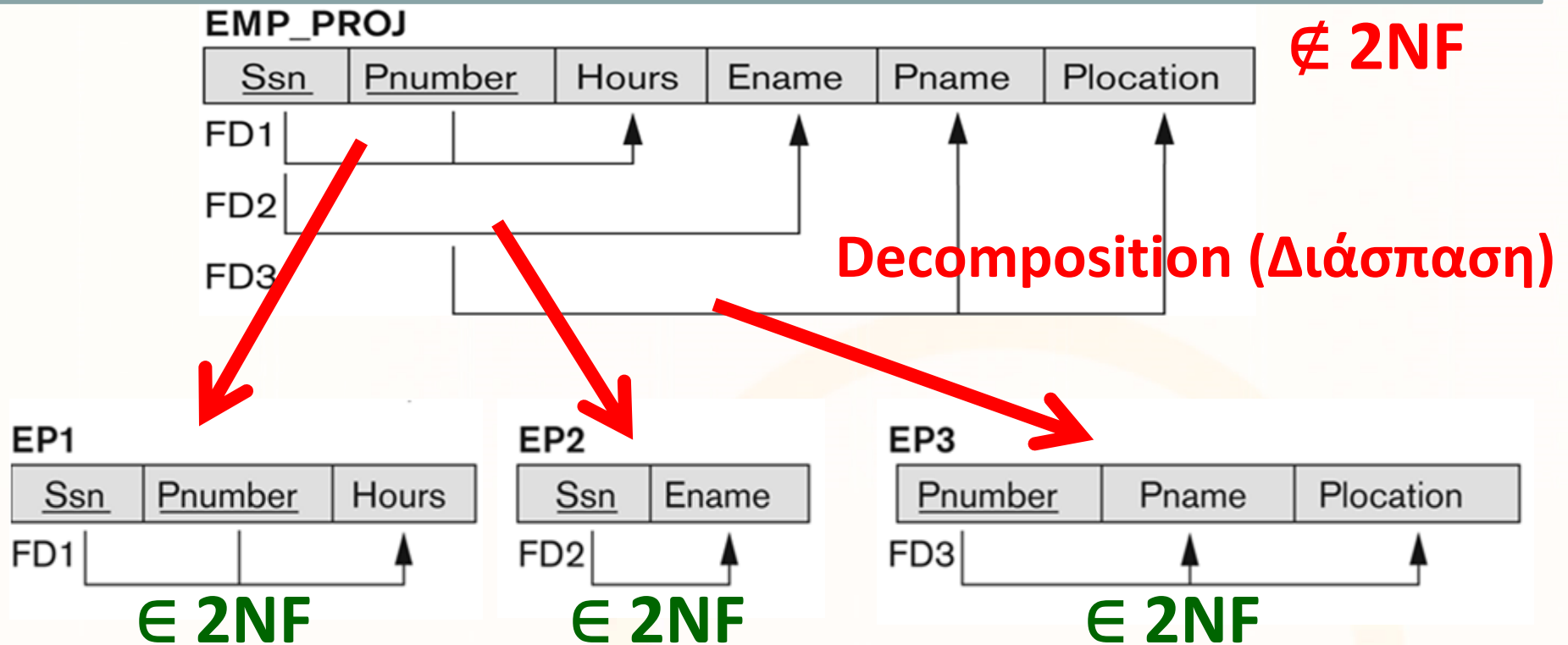
- Ας ορίσουμε την 2NF κάπως πιο τυπικά.
- **Ορισμός 2NF:** Μια σχέση R είναι σε **2NF** εάν κάθε **μη-πρωτεύων γνώρισμα** (non-prime attribute) στο R είναι **ολικά συναρτησιακά** εξαρτώμενο από το πρωτεύων κλειδί.
- **Παράδειγμα \notin 2NF**

(SSN, Pnumber, Hours, Ename)



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 .

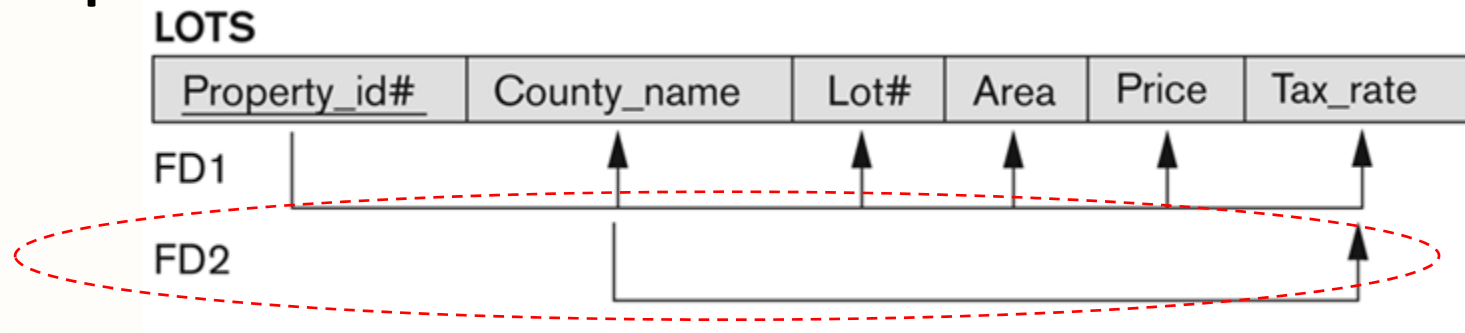


Ορισμοί: Μεταβατική FD

- **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.

Τρίτη Κανονική Μορφή (3NF)

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

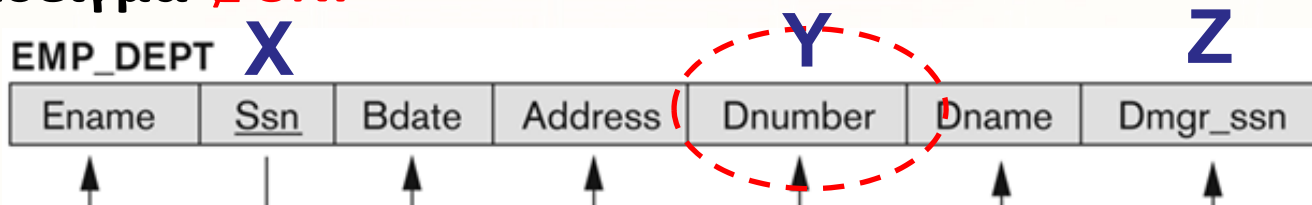
Τρίτη Κανονική Μορφή (3NF)

- Ας ορίσουμε την 3NF κάπως πιο τυπικά.
- **Τυπικός Ορισμός 3NF:** Μια σχέση R είναι σε 3NF εάν για κάθε FD $X \rightarrow Y$ που σχετίζεται με το R ισχύει ένα από τα ακόλουθα:
 - $X \supseteq Y$ (i.e., το FD είναι τετριμμένο) ή
 - Π.χ., $\{\text{County_Name}, A1, A2\} \rightarrow \text{County_Name}$
 - X είναι εναλλακτικό κλειδί (ή υπερκλειδί) της R ή
 - Π.χ., εάν $\text{County_Name} \rightarrow \{\text{Lot\#}, \text{Area}, \text{Price}, \text{Tax_Rate}\}$ άρα δεν προκύπτουν μεταβατικές εξαρτήσεις μέσω non-key (εφόσον το County_Name είναι key)
 - Κάθε γνώρ. $A \in Y$ είναι μέρος κάποιου κλειδιού του R
 - Επομένως δεν προκύπτει non-key μετάβαση μέσω του X .

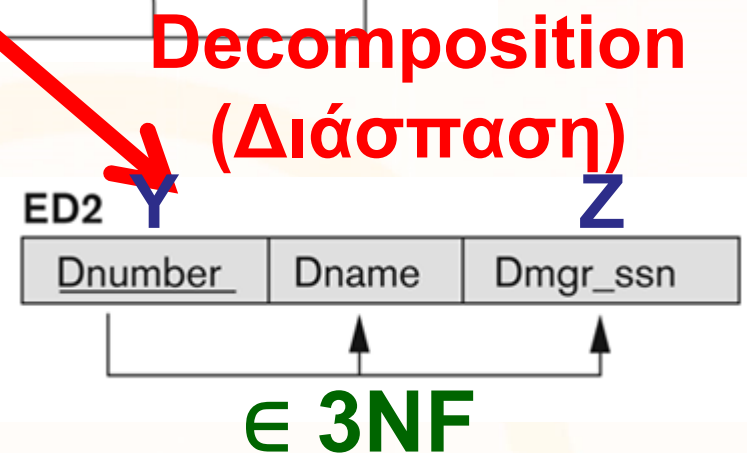
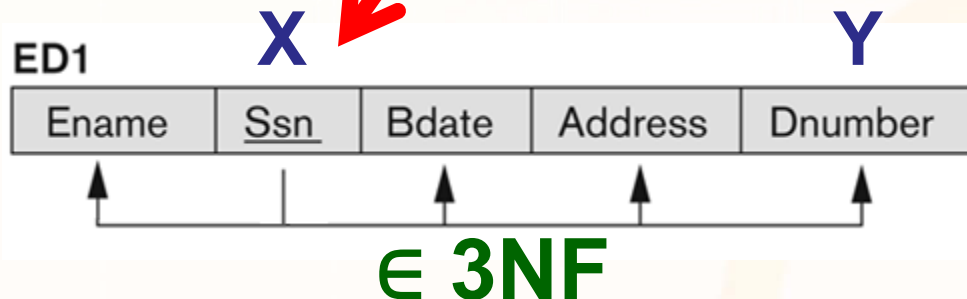
Λογική Διάσπασης σε 3NF

- **Λογική Διάσπασης σε 3NF:** Για κάθε FD που παραβιάζει την 3NF (δηλ., $X \rightarrow Y$, $Y \rightarrow Z$, όπου Y non-key), δημιουργήσε 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



Σύνοψη Κανονικών Μορφών

- **Πρώτη Κανονική Μορφή (1NF)**

- Δεν υπάρχουν **Πλειότιμα (Multi-valued)**
 - **Διαφορετικά Διατυπωμένο:** Όλα τα γνωρίσματα εξαρτώνται από το κλειδί.

- **Δεύτερη Κανονική Μορφή (2NF)**

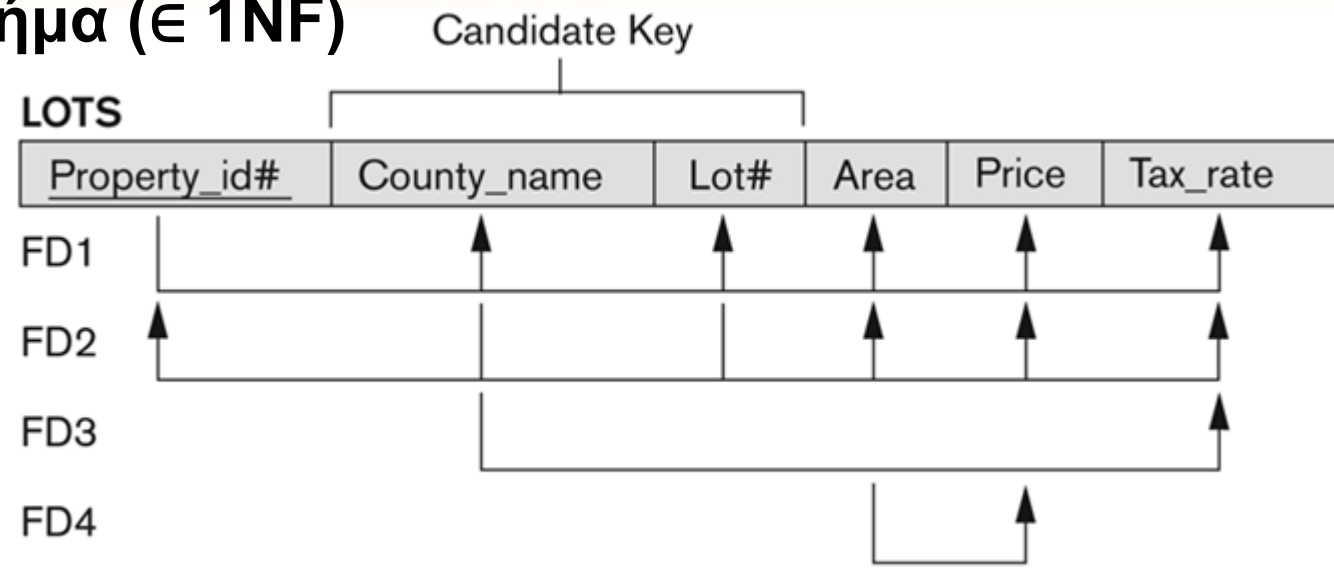
- Δεν υπάρχουν **Μερικές (Partial) Εξαρτήσεις**
 - **Διαφορετικά Διατυπωμένο:** Όλα τα γνωρίσματα εξαρτώνται από **Ολόκληρο** το κλειδί.

- **Τρίτη Κανονική Μορφή (3NF)**

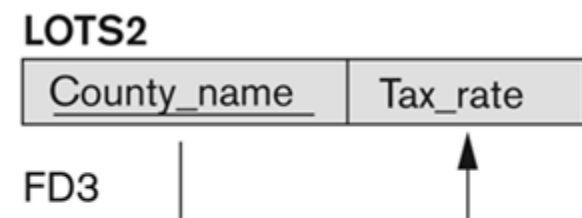
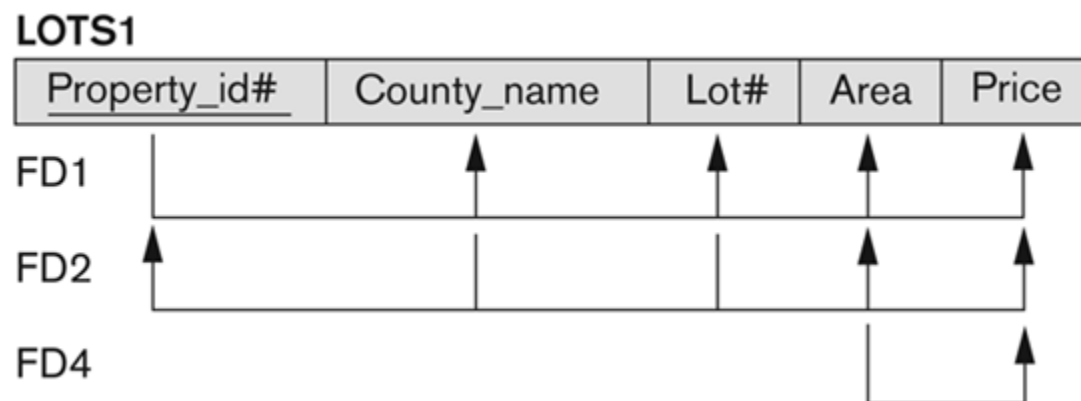
- Δεν υπάρχουν **Μεταβατικές (Transitive) Εξαρτήσεις**
 - **Διαφορετικά Διατυπωμένο:** Όλα τα γνωρίσματα δεν εξαρτώνται από **Τίποτα Άλλο** εκτός **Ολόκληρο** το κλειδί

Παράδειγμα Κανονικοποίησης σε 3NF

Αρχικό Σχήμα (ε 1NF)

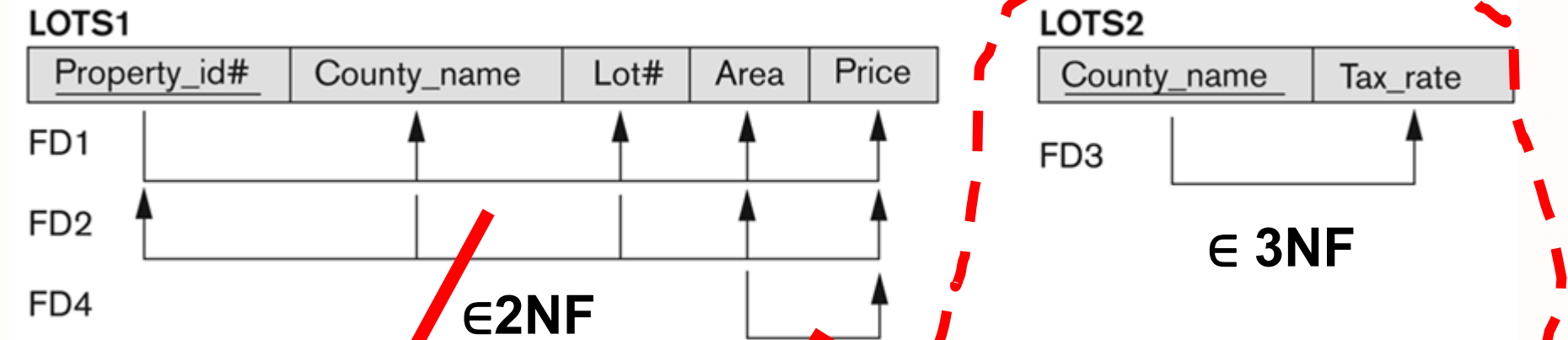


Βήμα 1: Κανονικοποίηση σε 2NF

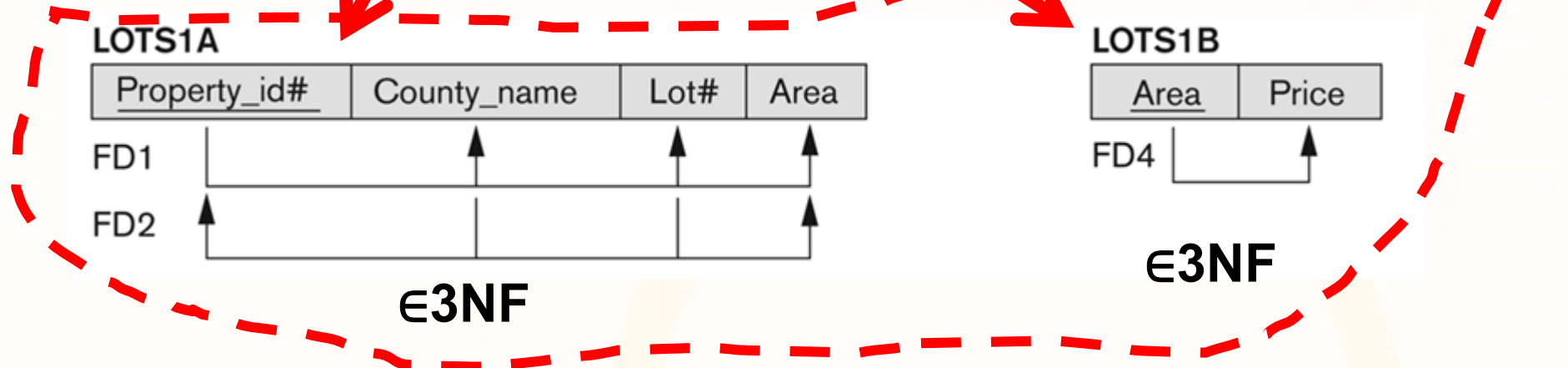


Παράδειγμα Κανονικοποίησης σε 3NF

... Από προηγούμενη διαφάνεια (€ 2NF)



Βήμα 2: Κανονικοποίηση σε 3NF



Παράδειγμα Κανονικοποίησης σε 3NF

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

