



# Κανονικοποίηση Σχήματος



- Αποσύνθεση (διάσπαση) καθολικού σχήματος

### Επιθυμητές ιδιότητες

- διατήρηση εξαρτήσεων ( $F'^+ = F^+$ )
- όχι απώλειες στη συνένωση (τομή = κλειδί)

- όχι επανάληψη πληροφορίας λόγω ΣΕ



Κανονικές μορφές

*Έστω  $R(A, B, C)$  καμία ΣΕ, αν  $A \rightarrow B$ ?*

## Σχεδιασμός Σχισιακών Σχημάτων (επανάληψη)



### Επιθυμητές Ιδιότητες Αποσύνθεσης

#### 1. Συνενώσεις Άνευ Απωλειών

Η φυσική συνένωση των σχέσεων που προκύπτουν μας δίνει *ακριβώς* την αρχική σχέση (χωρίς επιπρόσθετες πλειάδες):  $r = \pi_{R_1}(r) * \pi_{R_2}(r) * \dots * \pi_{R_n}(r)$

$R_1 \cap R_2 \rightarrow R_1$  ή  $R_1 \cap R_2 \rightarrow R_2$  ανήκει στο  $F^+$ , δηλαδή τα κοινά γνωρίσματα των δύο σχημάτων είναι κλειδί για τουλάχιστον ένα από τα δύο

#### 2. Διατήρηση Εξαρτήσεων

Στόχος: Έλεγχος διατήρησης εξαρτήσεων όταν γίνονται τροποποιήσεις χωρίς να υπολογίζουμε τις αρχικές σχέσεις (αποφυγή των συνενώσεων)

$F' = F_1 \cup F_2 \dots \cup F_n$ , πρέπει  $F'^+ = F^+$

#### 3. Αποφυγή Επανάληψης Πληροφορίας, **πως; Κανονικές Μορφές**



Έστω  $R(A, B, C, D, E)$ ,  $F = \{A \rightarrow D, B \rightarrow E, DE \rightarrow C\}$ . Η αποσύνθεση του  $R$  σε  $S(A, B, C)$  και  $T(A, B, D, E)$  διατηρεί τις εξαρτήσεις;

Ιδιότητες της διάσπασης/αποσύνθεσης της  $R$  σε  $S$  και  $T$

- (i) Lossless join (χωρίς απώλειες στη συνένωση, μη προσθετική συνένωση)
- (ii) Διατήρηση εξαρτήσεων
  - (i) Υπολογισμός του περιορισμού της  $F$  στο  $S$
  - (ii) Υπολογισμός του περιορισμού της  $F$  στο  $T$



- **Στόχος:** Δοσμένου ενός σχήματος, αν είναι «καλό» ή χρειάζεται περαιτέρω διάσπαση.

Πως; Κανονικές μορφές.

- Ξέρουμε ότι αν ένα σχήμα είναι σε κάποια Κανονική Μορφή δεν υπάρχουν συγκεκριμένα προβλήματα
- Με φθίνουσα σειρά (από την πιο περιοριστική στη λιγότερο περιοριστική)

**BCNF 3NF 2NF 1NF**

- Βασίζεται σε Σ.Ε., οι Σ.Ε. έχουν σχέση με την επανάληψη πληροφορίας



## Πλεονασμός (επανάληψη πληροφορίας)

### Ταινία

<u>Τίτλος</u>	<u>Έτος</u>	Διάρκεια	Είδος	<u>Όνομα-Ηθοποιού</u>
---------------	-------------	----------	-------	-----------------------

*Τι συμβαίνει με το (πρωτεύον) κλειδί και τις συναρτησιακές εξαρτήσεις;*



Ένα σχεσιακό σχήμα  $R$  είναι σε **Κανονική Μορφή Boyce-Codd (BCNF)** σε σχέση με ένα σύνολο  $F$  συναρτησιακών εξαρτήσεων αν

για όλες τις ΣΕ στο  $F^+$  της μορφής  $X \rightarrow Y$  ισχύει τουλάχιστον ένα από τα παρακάτω:

- $X \rightarrow Y$  είναι μια τετριμμένη ΣΕ ή
- $X$  είναι **υπερκλειδί** (δηλαδή υποψήφιο κλειδί ή υπερσύνολο υποψήφιου κλειδιού) του σχήματος  $R$

*Δηλαδή το αριστερό μέρος κάθε μη τετριμμένης ΣΕ πρέπει να περιέχει ένα κλειδί*

Το σχήμα μιας ΒΔ είναι σε BCNF αν το σχήμα **κάθε** σχέσης της είναι σε BCNF.



### *Παράδειγμα 1*

Ταινία (Τίτλος, Έτος, Διάρκεια, Είδος, Όνομα-Ηθοποιού)

Η σχέση Ταινία δεν είναι σε BCNF

(υποψήφιο) κλειδί: {Τίτλος, Έτος, Όνομα-Ηθοποιού}

Για παράδειγμα η ΣΕ Τίτλος Έτος → Διάρκεια





### *Παράδειγμα 2*

Ταινία2 (Τίτλος, Έτος, Διάρκεια, Είδος)

Η σχέση Ταινία2 είναι σε BCNF

### *Παράδειγμα 3*

*Οποιαδήποτε σχέση με δύο γνωρίσματα είναι σε BCNF*



## Αλγόριθμος Αποσύνθεσης σε BCNF

- Βρες μια μη τετριμμένη ΣΕ που παραβιάζει τον BCNF ορισμό, έστω  $X \rightarrow Y$  και  $X \cap Y = \emptyset$
- Αποσύνθεση του αρχικού σχήματος  $R$  σε δύο σχήματα
  - $R_1$  με γνωρίσματα  $X \cup Y$
  - $R_2$  με γνωρίσματα  $R - Y$

*Ευριστικός: στα δεξιά όσο το δυνατόν περισσότερα γνωρίσματα*

*Αποσύνθεση χωρίς απώλειες:*



*Παράδειγμα 1*

**Ταινία** (Τίτλος, Έτος, Διάρκεια, Είδος, Όνομα-Ηθοποιού)

Τίτλος Έτος → Διάρκεια Είδος

**Ταινία1**(Τίτλος, Έτος, Διάρκεια, Είδος)

**Ταινία2**(Τίτλος, Έτος, Όνομα-Ηθοποιού)



- Μπορεί να χρειαστεί παραπάνω από μία αποσύνθεση

Αποσύνθεση του αρχικού σχήματος  $R$  σε δύο σχήματα

-  $R_1$  με γνωρίσματα  $X \cup Y$  και

-  $R_2$  με γνωρίσματα  $R - Y$

Συνεχείς διασπάσεις,

αφού καταλήγουμε σε σχέσεις με αυστηρά μικρότερο αριθμό γνωρισμάτων, η διαδικασία τερματίζει



Παραβίαση του BCNF σημαίνει ότι υπάρχει  $X \rightarrow A$  όπου το  $X$  δεν είναι υπερκλειδί

**Περίπτωση 1:**  $X$  είναι γνήσιο υποσύνολο κάποιου υποψήφιου κλειδιού (μερική εξάρτηση)

**Περίπτωση 2:**  $X$  δεν είναι γνήσιο υποσύνολο κάποιου υποψήφιου κλειδιού

Τότε έστω  $K$  (υποψήφιο κλειδί)

$K \rightarrow X \rightarrow A$  (μεταβατική εξάρτηση)

Δε μπορώ να συνδυάσω μια τιμή του  $X$  με μια τιμή του  $K$  χωρίς να συνδυάσω μια τιμή  $A$  με μια τιμή  $X$

Δε μπορώ να εισάγω τιμή του  $X$ , χωρίς να ξέρω και το «σωστό»  $A$

## Συναρτησιακές Εξαρτήσεις



Παράδειγμα: Θεωρούμε ότι ένας λογαριασμός μπορεί να ανήκει σε παραπάνω από έναν πελάτη και ένας πελάτης μπορεί να έχει πολλούς λογαριασμούς.

### Λογαριασμός

<u>Όνομα-Υποκαταστήματος</u>	<u>Αριθμός-Λογαριασμού</u>	Ποσό	<u>Όνομα-Πελάτη</u>
------------------------------	----------------------------	------	---------------------

Παράδειγμα: Ένας Πελάτης πολλά δάνεια και ένα Δάνειο από παραπάνω από έναν πελάτη

### Πελάτης

<u>Όνομα-Πελάτη</u>	Οδός	Πόλη	<u>Αριθμός-Δανείου</u>
---------------------	------	------	------------------------

Διεύθυνση πελάτη

- Και στις δυο περιπτώσεις έχω μερική εξάρτηση



## Παράδειγμα 2

Ταινία-Εταιρεία (Τίτλος, Έτος, Διάρκεια, Είδος, Εταιρεία-Παραγωγής, Διεύθυνση-Εταιρείας)

Πρόβλημα: υπάρχει μια **μεταβατική** εξάρτηση

Τίτλος Έτος → Εταιρεία-Παραγωγής

**Εταιρεία-Παραγωγής → Διεύθυνση-Εταιρείας**

*Τίτλος Έτος → Διεύθυνση-Εταιρείας*

*Για να αντιστοιχήσουμε μια ταινία σε εταιρεία πρέπει να ξέρουμε τη διεύθυνση!*

Ταινία-Εταιρεία1 (Εταιρεία-Παραγωγής, Διεύθυνση-Εταιρείας)

Ταινία-Εταιρεία2 (Τίτλος, Έτος, Διάρκεια, Είδος, Εταιρεία-Παραγωγής)



Δεν είναι πάντα δυνατή η αποσύνθεση σε μια BCNF που να διατηρεί τις εξαρτήσεις

### Παράδειγμα

Έστω η σχέση Παίζει(Έργο, Κινηματογράφος, Πόλη) με τους περιορισμούς ότι

(i) δεν υπάρχουν κινηματογράφοι με το ίδιο όνομα

(ii) σε κάθε κινηματογράφο παίζονται πολλά έργα (έχει πολλές αίθουσες) αλλά κάθε έργο παίζεται μόνο σε ένα κινηματογράφο σε κάθε πόλη

Κινηματογράφος → Πόλη

Κλειδιά:

Έργο Πόλη → Κινηματογράφος

{Έργο, Πόλη}

{Κινηματογράφος, Έργο}





Παίζει(Έργο, Κινηματογράφος, Πόλη)

Κινηματογράφος → Πόλη

Έργο Πόλη → Κινηματογράφος

Κλειδιά

{Έργο, Πόλη} {Κινηματογράφος, Έργο}

Αποσύνθεση σε:  $R_1$ {Κινηματογράφος, Πόλη} και  $R_2$ {Κινηματογράφος, Έργο}

Διατηρεί τις εξαρτήσεις:

Κινηματογράφος	Πόλη	Κινηματογράφος	Έργο
Odeon-ABANA	Αθήνα	Odeon-ABANA	Vicky Cristina Barcelona
Village Center Μαρούσι	Αθήνα	Village Center Μαρούσι	Vicky Cristina Barcelona

Δε μπορώ κοιτάζοντας μόνο την  $R_2$  (ή την  $R_1$ ) να δω ότι η εισαγωγή της δεύτερης πλειάδας παραβιάζει μια ΣΕ (πρέπει να κάνω συνένωση!)



Ένα σχεσιακό σχήμα  $R$  είναι σε **τρίτη κανονική μορφή (3NF)** σε σχέση με ένα σύνολο  $F$  συναρτησιακών εξαρτήσεων αν για όλες τις ΣΕ στο  $F^+$  της μορφής  $X \rightarrow Y$  ισχύει τουλάχιστον ένα από τα παρακάτω:

- $X \rightarrow Y$  είναι μια τετριμμένη ΣΕ ή
- $X$  είναι υπερκλειδί του σχήματος  $R$
- κάθε γνώρισμα  $A$  του  $Y - X$  περιέχεται σε κάποιο **υποψήφιο κλειδί**

**Πρωτεύον γνώρισμα** (prime attribute): Γνώρισμα που ανήκει σε κάποιο υποψήφιο κλειδί

*BCNF πιο περιοριστική -- αν σε BCNF  $\Rightarrow$  3NF*



### Παράδειγμα

Παίζει(Έργο, Κινηματογράφος, Πόλη)

Έργο Πόλη → Κινηματογράφος

Κινηματογράφος → Πόλη

Κλειδιά {Έργο, Πόλη}  
{Κινηματογράφος, Έργο}

Υπάρχει μια μεταβατική εξάρτηση

Αλλά απαιτούμε να είναι σε  
πρωτεύον γνώρισμα

Η σχέση είναι σε 3NF



## Αλγόριθμος (Από) σύνθεσης σε 3NF

- Υπολόγισε το ελάχιστο κάλυμμα  $F_c$  του  $F$  της αρχικής  $R$
- Για κάθε α.μ.  $X$  μιας συναρτησιακής εξάρτισης του  $F_c$   
έστω  $Y$  το σύνολο όλων των γνωρισμάτων  $A_i$  που εμφανίζονται στο δ.μ. μιας ΣΕ του  $F_c$   $X \rightarrow A_i$

νέα σχέση με γνωρίσματα  $X \cup Y$

- Αν κανένα από τα σχήματα που δημιουργούνται δεν περιέχει κλειδί της  $R$ , δημιούργησε ένα σχήμα σχέσης που να περιέχει τα γνωρίσματα που σχηματίζουν κλειδί (όχι απώλεια)



## Αλγόριθμος Αποσύνθεσης σε 3NF

- Απώλειες στη συνένωση;
- Διατήρηση εξαρτήσεων;



## Παράδειγμα

Τραπεζίτης(Όνομα-Υποκαταστήματος, Όνομα-Πελάτη, Όνομα-Τραπεζίτη, Αριθμός Γραφείου)

Όνομα-Τραπεζίτη → Όνομα-Υποκαταστήματος Αριθμός-Γραφείου

Όνομα-Πελάτη Όνομα-Υποκαταστήματος → Όνομα-Τραπεζίτη

**Κλειδιά {Όνομα-Πελάτη, Όνομα-Υποκαταστήματος}**

*3NF:*

Τραπεζίτης1(Όνομα-Τραπεζίτη, Όνομα-Υποκαταστήματος Αριθμός-Γραφείου)

Τραπεζίτης2(Όνομα-Πελάτη, Όνομα-Υποκαταστήματος, Όνομα-Τραπεζίτη)

*BCNF:*



## Κανονική Μορφή Boyce-Codd

Ένα σχεσιακό σχήμα  $R$  είναι σε BCNF σε σχέση με ένα σύνολο  $F$  συναρτησιακών εξαρτήσεων αν για όλες τις ΣΕ στο  $F^+$  της μορφής  $X \rightarrow Y$  ισχύει τουλάχιστον ένα από τα παρακάτω:

- $X \rightarrow Y$  είναι μια τετριμμένη ΣΕ ή
- $X$  είναι υπερκλειδί του σχήματος  $R$

## Τρίτη Κανονική Μορφή

-- κάθε γνώρισμα  $A$  του  $Y - X$  περιέχεται σε κάποιο υποψήφιο κλειδί (είναι πρωτεύον γνώρισμα)

## Σχεδιασμός Σχισιακών Σχημάτων (Επανάληψη)



	BCNF	3NF
• Αποφυγή επανάληψης πληροφορίας	ναι	όχι πάντα
• Αποσύνθεση χωρίς απώλειες στη συνένωση	ναι	ναι
• Διατήρηση εξαρτήσεων	όχι πάντα	ναι



## Πρώτη Κανονική Μορφή



1NF, 2NF ιστορικής σημασίας

1NF (ιστορικοί λόγοι, κάθε γνώρισμα παίρνει ατομικές τιμές)

## Δεύτερη Κανονική Μορφή



$X \rightarrow Y$

$Y$  πλήρης εξάρτηση από το  $X$  αν δεν υπάρχουν περιττά γνωρίσματα στο  $X$  (στο α.μ της εξάρτησης) (αν υπάρχουν περιττά γνωρίσματα, μερική εξάρτηση)

Δε μπορούμε να αφαιρέσουμε γνωρίσματα από το  $X$  και η εξάρτηση να συνεχίζει να ισχύει

**2NF** κάθε μη πρωτεύον γνώρισμα (γνώρισμα που δεν ανήκει στο υποψήφιο κλειδί) είναι πλήρως εξαρτώμενο από το πρωτεύον κλειδί

Ισχύει: Πρωτεύον  $\rightarrow$  Μη πρωτεύον (το 2NF λέει ότι αυτή είναι πλήρης εξάρτηση, δεν μπορούμε στο α.μ. υποσύνολο του πρωτεύον)

Προφανώς 2NF αφορά κλειδιά με παραπάνω από ένα γνωρίσματα



**2NF** κάθε μη πρωτεύον γνώρισμα (γνώρισμα που δεν ανήκει στο υποψήφιο κλειδί) είναι πλήρως εξαρτώμενο από το πρωτεύον κλειδί

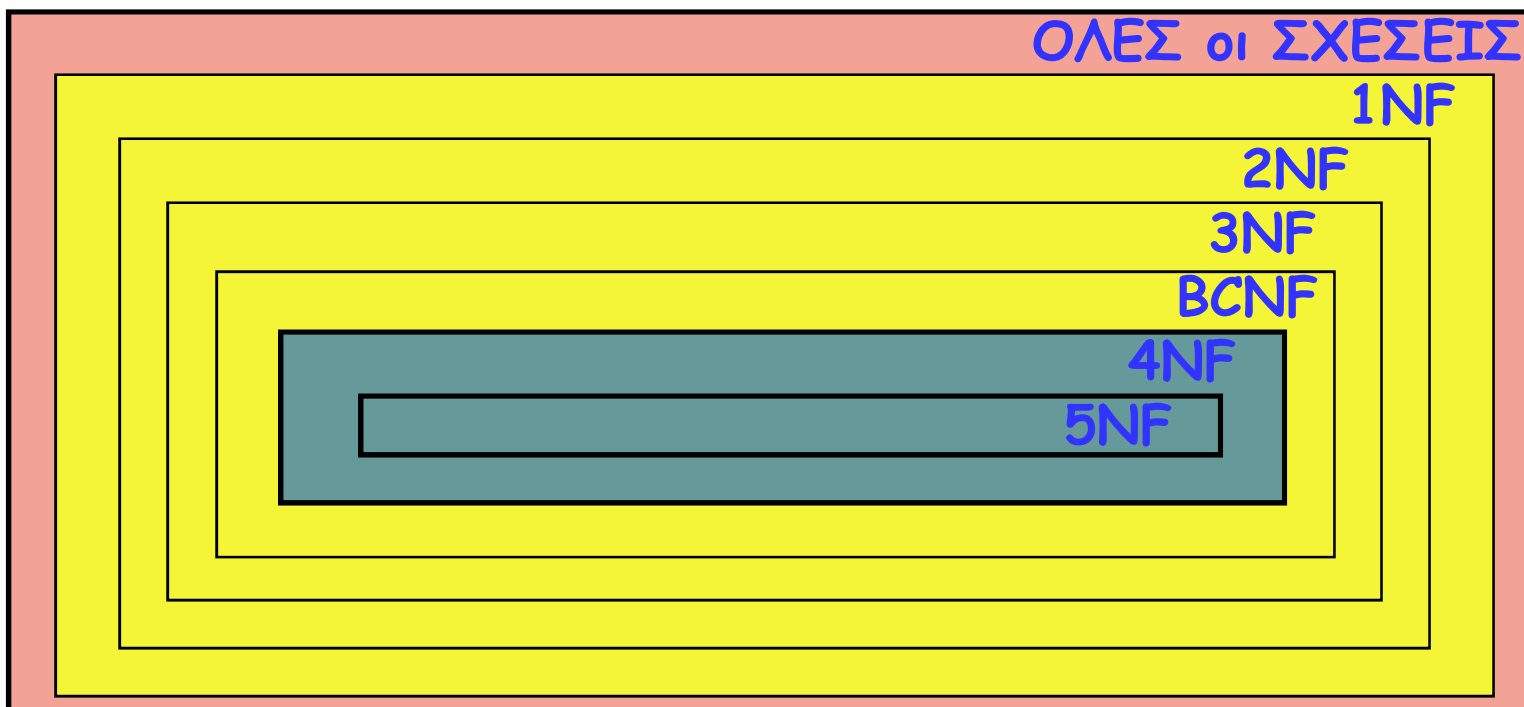
*Δηλαδή, δεν είναι σε 2NF*

*αν κάποιο μη πρωτεύον γνώρισμα είναι μερικώς εξαρτημένο από πρωτεύον, δηλαδή υπάρχει  $X \rightarrow Y$  όπου  $X$  υποσύνολο κλειδιού και  $Y$  μη πρωτεύον*

**3NF => 2NF** γιατί αν  $X \rightarrow Y$  τότε είτε  $X$  υπερκλειδί είτε  $Y$  πρωτεύον



## Κανονικές Μορφές



## Πλειότιμες Εξαρτήσεις



Υπάρχει επανάληψη πληροφορίας που δεν μπορεί να εκφραστεί με απλές ΣΕ



## Πλειότιμες Εξαρτήσεις

- Προκύπτουν όταν δυο γνωρίσματα είναι ανεξάρτητα το ένα από το άλλο

### Παράδειγμα

Ηθοποιός(Όνομα, Οδός, Πόλη, Τίτλος, Έτος)

*Ας υποθέσουμε ότι για κάθε ηθοποιό είναι πιθανόν να υπάρχουν πολλές διευθύνσεις, και ένα ηθοποιός παίζει σε πολλές ταινίες*

Κανένα από τα 5 γνωρίσματα δεν εξαρτάται συναρτησιακά από τα άλλα τέσσερα  $\Rightarrow$  δεν υπάρχουν μη μη τετριμμένες εξαρτήσεις  $\Rightarrow$  κλειδί ?

π.χ., Όνομα Οδός Τίτλος Έτος  $\rightarrow$  Πόλη δεν ισχύει



*Παράδειγμα (συνέχεια)*

Ηθοποιός(Όνομα, Οδός, Πόλη, Τίτλος, Έτος)

Όλες οι εξαρτήσεις είναι τετριμμένες

Το σχήμα είναι σε BCNF αλλά υπάρχει επανάληψη πληροφορίας που δεν οφείλεται όμως σε συναρτησιακές εξαρτήσεις



*Παράδειγμα*

Ηθοποιός(Όνομα, Οδός, Πόλη, Τίτλος, Έτος)

Όνομα →→ Οδός Πόλη

Ο ηθοποιός C. Fisher έχει 2 διευθύνσεις:

Όνομα	Οδός	Πόλη	Τίτλος	Έτος
C. Fisher	123 Mapple Str	Hollywood	Star Wars	1977
C. Fisher	5 Locust Ln	Malibu	Empire Strikes Back	1980
?				
?				





$$X \twoheadrightarrow Y$$

Για κάθε ζεύγος πλειάδων  $t_1$  και  $t_2$  της σχέσης  $R$  που συμφωνούν σε όλα τα γνωρίσματα του  $X$  μπορούμε να βρούμε στο  $R$  δυο πλειάδες  $t_3$  και  $t_4$  τέτοιες ώστε

- και οι δυο συμφωνούν με τις  $t_1$  και  $t_2$  στο  $X$ :

$$t_1[X] = t_2[X] = t_3[X] = t_4[X]$$

- η  $t_3$  συμφωνεί με την  $t_1$  στο  $Y$ :  $t_3[Y] = t_1[Y]$
- η  $t_3$  συμφωνεί με την  $t_2$  στο  $R - X - Y$ :  $t_3[R - X - Y] = t_2[R - X - Y]$
- η  $t_4$  συμφωνεί με την  $t_2$  στο  $Y$ :  $t_4[Y] = t_2[Y]$
- η  $t_4$  συμφωνεί με την  $t_1$  στο  $R - X - Y$ :  $t_4[R - X - Y] = t_1[R - X - Y]$

# Πλειότιμες Εξαρτήσεις



$$A_1 A_2 \dots A_n \longrightarrow B_1 B_2 \dots B_m$$

Όνομα				Πόλη Οδός				Τίτλος Έτος					
X				Y				R - X - Y					
←-----→				←-----→				←-----→					
$A_1$	$A_2$	...	$A_n$	$B_1$	$B_2$	...	$B_m$	$C_1$	$C_2$	...	$C_k$		
$a_1$	$a_2$	...	$a_n$	$b_1$	$b_2$	...	$b_m$	$c_1$	$c_2$	...		←	$t_1$
$c_k$	$a_2$	...	$a_n$	$b'_1$	$b'_2$	...	$b'_m$	$c'_1$	$c'_2$	...	$c'_k$	←	$t_2$
$a_1$												←	$t_3$
												←	$t_4$



ΕΡΓΑΖΟΜΕΝΟΣ(ΕΡ-ΟΝΟΜΑ, ΑΤ, ΗΜ-ΓΕΝ, ΔΙΕΥΘΥΝΣΗ, ΚΩΔ-ΤΜΗΜΑΤΟΣ)

ΤΜΗΜΑ(ΤΜ-ΟΝΟΜΑ, ΚΩΔ-ΤΜΗΜΑΤΟΣ)

Κάποιο εργαζόμενοι δεν έχουν ανατεθεί σε τμήματα (null στο ΚΩΔ-ΤΜΗΜΑΤΟΣ που είναι **ξένο κλειδί**)

Θέλουμε να βρούμε τα ζεύγη (ΕΡ-ΟΝΟΜΑ, ΤΜ-ΟΝΟΜΑ)

Φυσική συνένωση?



- Η διαδικασία Κανονικοποίησης έχει και *μειονεκτήματα*:
  - ο Δεν είναι δημιουργική
  - ο **Συνήθως η κανονικοποίηση γίνεται αφού έχουμε κάποιο σχήμα (μας λέει αν είναι «καλό» ή «κακό»)**
  - ο Δεν προσφέρει ένα εννοιολογικό σχήμα (ασχολείται μόνο με σχέσεις και γνωρίσματα)

Όμως, είναι μια ενδιαφέρουσα και πρακτικά χρήσιμη προσπάθεια να γίνουν με τυπικό και συστηματικό τρόπο πράγματα που τα κάνουμε συνήθως διαισθητικά.



- Ένας μεγάλος αριθμός από **εμπορικά εργαλεία**, δοθέντων ενός συνόλου Σχημάτων Σχέσεων/Γνωρισμάτων και ενός συνόλου συναρτησιακών εξαρτήσεων **δημιουργούν αυτόματα** σχήματα σχέσεων σε μορφή *3NF* (σπάνια πάνε σε *BCNF*, *4NF* και *5NF*)
- Μια άλλη χρήση τέτοιων εργαλείων είναι να **ελέγχουν το επίπεδο κανονικοποίησης** μιας σχέσης - γενικά, η χρήση ως ευριστικό εργαλείο επιλογής ενός σχεδιασμού έναντι κάποιου άλλου
- Υπάρχουν **πρακτικά αποτελέσματα** της θεωρίας που επιτρέπουν σε έναν σχεδιαστή να κάνει ανάλυση της μορφής:

*Αν μια σχέση είναι σε 3NF και κάθε υποψήφιο κλειδί αποτελείται ακριβώς από ένα γνώρισμα, τότε είναι και σε 5NF (Fagin, 1991)*



## Η Διαδικασία Σχεδιασμού

1. Συλλογή και ανάλυση απαιτήσεων
2. Εννοιολογικός σχεδιασμός
3. Επιλογή ΣΔΒΔ
4. Απεικόνιση στο μοντέλο δεδομένων (λογικός σχεδιασμός)
5. Φυσικός σχεδιασμός
6. Υλοποίηση