

ΔΙΑΧΕΙΡΙΣΗ ΔΕΔΟΜΕΝΩΝ

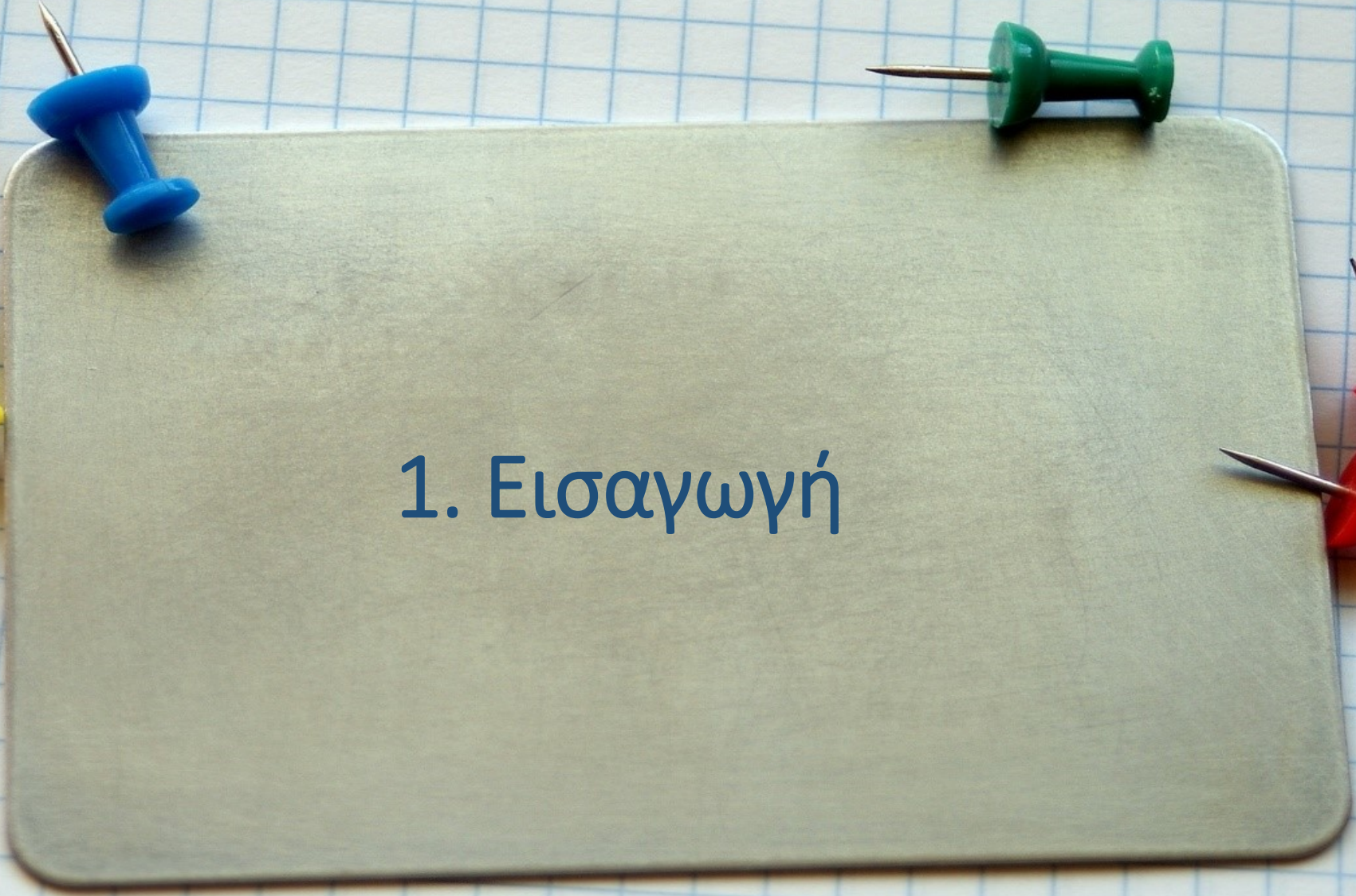
Κανονικοποίηση

Κανονικοποίηση

1. [Εισαγωγή](#)
2. [Πρακτικές οδηγίες για την ανάπτυξη του σχήματος μιας σχεσιακής ΒΔ](#)
3. [Κανονικοποίηση](#)
4. [Πρώτη κανονική μορφή](#)
5. [Δεύτερη κανονική μορφή](#)
6. [Τρίτη κανονική μορφή](#)
7. [Κανονική μορφή Boyce-Codd](#)
8. [Εξάρτηση πολλαπλών τιμών](#)

Κανονικοποίηση

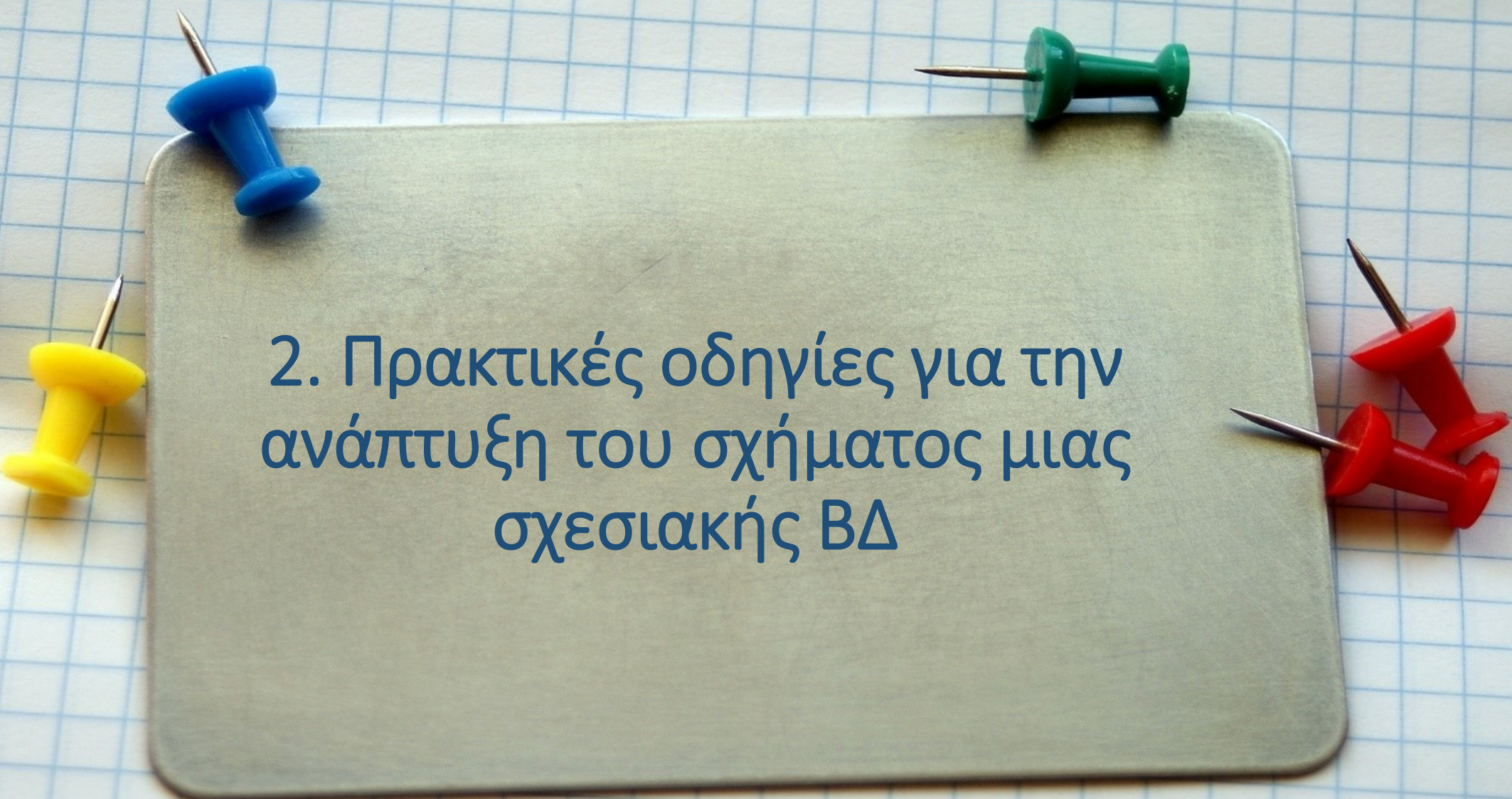




1. Εισαγωγή

Εισαγωγή

- Μέχρι τώρα παρουσιάστηκε πως το σχήμα μια σχεσιακής ΒΔ μπορεί να προκύψει με μετατροπή ενός ΔΟΣ ή ενός ΕΔΟΣ ακολουθώντας συγκεκριμένα βήματα ενός αλγορίθμου, προσέγγιση που είναι γνωστή και ως «εκ των άνω προς τα κάτω». Αυτή η προσέγγιση ξεκινάει θεωρώντας αρχικά τις οντότητες του ΔΟΣ ή του ΕΔΟΣ και τις μεταξύ τους συσχετίσεις, και καταλήγει στους πίνακες της ΒΔ και τις μεταξύ τους συσχετίσεις.
- Υπάρχουν και άλλες προσεγγίσεις και εδώ θα παρουσιαστούν άλλες δύο που προσδιορίζουν το σχήμα μιας ΒΔ.
- Η 1^η εναλλακτική προσέγγιση βασίζεται σε πρακτικές οδηγίες που προσφέρει έναν γρήγορο τρόπο, όμως μπορεί πιθανά ανασφαλή, για την ανάπτυξη του σχήματος μιας ΒΔ.
- Η 2^η προσέγγιση, γνωστή ως **κανονικοποίηση**, βασίζεται σε ένα καλά θεμελιωμένο υπόβαθρο το οποίο συστηματοποιεί τη διαδικασία προσδιορισμού του σχήματος μιας ΒΔ.



2. Πρακτικές οδηγίες για την ανάπτυξη του σχήματος μιας σχεσιακής ΒΔ

Πρακτικές οδηγίες για την ανάπτυξη του σχήματος μιας σχεσιακής βάσης δεδομένων

- Ένα σχήμα μιας σχεσιακής ΒΔ μπορεί να προσδιοριστεί με βάση την κοινή λογική βασιζόμενη σε μερικές πρακτικές οδηγίες
- Αυτός ο τρόπος ακολουθείται όταν η ΒΔ είναι απλή και έχει μικρό αριθμό πινάκων.
- Οι πρακτικές οδηγίες που εφαρμόζονται σε αυτή την περίπτωση, βασίζονται στην ευκολία κατανόησης του σχήματος της ΒΔ και την ευκολία χρήσης του.

Πρώτη πρακτική οδηγία

- Η πρώτη πρακτική οδηγία συνιστά το σχήμα μιας σχεσιακής ΒΔ να μπορεί εύκολα να εξηγηθεί και γίνει εύκολα κατανοητό από κάποιον.
- Κατά την πρακτική αυτή οδηγία θα πρέπει:
 - Το όνομα του πίνακα να είναι απλό,
 - Το νόημα όλων των πεδίων του πίνακα να είναι σε αρμονία με το όνομα του πίνακα.

Παράδειγμα πρώτης πρακτικής οδηγίας

- Έστω ότι μια εταιρεία ταχυμεταφορών διαθέτει αυτοκίνητα που τα οδηγούν οι οδηγοί της εταιρείας, όπως φαίνεται παρακάτω:

ΠΟΙΟΣ_ΟΔΗΓΟΣ_ΟΔΗΓΕΙ_ΠΟΙΟ_ΑΥΤΟΚΙΝΗΤΟ

<u>ΑριθμόςΔιπλώματος</u>	ΗμερομηνίαΈκδοσης	ΑριθμόςΚυκλοφορίας	ΗμερομηνίαΑγοράς	Χρώμα
--------------------------	-------------------	--------------------	------------------	-------

- Ωστόσο το όνομα του πίνακα δεν είναι σε αρμονία με το νόημα των πεδίων του πίνακα και επιπλέον το όνομα του πίνακα δεν είναι απλό.
- Για να αρθεί ο κακός σχεδιασμός του παραπάνω πίνακα, διασπάται σε δύο πίνακες:

ΟΔΗΓΟΣ

<u>ΑριθμόςΔιπλώματος</u>	ΗμερομηνίαΈκδοσης	ΑριθμόςΚυκλοφορίας
--------------------------	-------------------	--------------------

ΑΥΤΟΚΙΝΗΤΟ

<u>ΑριθμόςΚυκλοφορίας</u>	ΗμερομηνίαΑγοράς	Χρώμα
---------------------------	------------------	-------

Δεύτερη πρακτική οδηγία

- Κατά τον σχεδιασμό των πινάκων θα πρέπει να αποφεύγεται όσο το δυνατόν περισσότερο η εισαγωγή της τιμής NULL.
- Ο βασικότερος λόγος είναι ότι η τιμή NULL μπορεί να ερμηνευτεί με πολλούς διαφορετικούς τρόπου.
- Έτσι η ΒΔ θα πρέπει να σχεδιάζεται ώστε να ελαχιστοποιείται η χρήση της τιμής NULL.

Παράδειγμα δεύτερης πρακτικής οδηγίας

- Έστω ότι μια εταιρεία ταχυμεταφορών διαθέτει έναν πίνακα με τα στοιχεία των υπαλλήλων της, όπως φαίνεται παρακάτω:

ΥΠΑΛΛΗΛΟΣ

<u>ΑΦΜ</u>	Όνομα	Επώνυμο	ΗμερομηνίαΓέννησης	ΑριθμόςΔιπλώματος	ΗμερομηνίαΈκδοσηςΔιπλώματος
------------	-------	---------	--------------------	-------------------	-----------------------------

- Αν στη συγκεκριμένη εταιρεία οι περισσότεροι υπάλληλοι είναι νέοι και δεν έχουν δίπλωμα οδήγησης τότε θα υπάρχουν πολλές τιμές NULL στο πεδίο ΑριθμόςΔιπλώματος και ΗμερομηνίαΈκδοσηςΔιπλώματος.
- Για να αρθεί ο κακός σχεδιασμός του παραπάνω πίνακα, διασπάται σε δύο πίνακες:

ΥΠΑΛΛΗΛΟΣ

<u>ΑΦΜ</u>	Όνομα	Επώνυμο	ΗμερομηνίαΓέννησης
------------	-------	---------	--------------------

ΟΔΗΓΟΣ

<u>ΑΦΜ</u>	ΑριθμόςΔιπλώματος	ΗμερομηνίαΈκδοσηςΔιπλώματος
------------	-------------------	-----------------------------

Προβλήματα εισαγωγής, διαγραφής και ενημέρωσης

- Οι πίνακες που δεν είναι καλά σχεδιασμένοι παρουσιάζουν διάφορα προβλήματα:
 - Είναι δύσκολο να έχουν ένα απλό όνομα, ή
 - Χρησιμοποιούν συχνά την τιμή NULL.
- Εκτός όμως των παραπάνω, οι κακώς σχεδιασμένοι πίνακες παρουσιάζουν διάφορα προβλήματα κατά την εισαγωγή, διαγραφή και τροποποίηση των δεδομένων τους.
- Τα προβλήματα αυτά ονομάζονται **ανωμαλίες** και παρουσιάζονται στη συνέχεια με χρήση παραδειγμάτων.

Προβλήματα εισαγωγής

- Έστω ότι υπάρχει ο παρακάτω πίνακας:

ΠΟΙΟΣ_ΟΔΗΓΟΣ_ΟΔΗΓΕΙ_ΠΟΙΟ_ΑΥΤΟΚΙΝΗΤΟ

ΑριθμόςΔιπλώματος	ΗμερομηνίαΈκδοσης	ΑριθμόςΚυκλοφορίας	ΗμερομηνίαΑγοράς	Χρώμα
-------------------	-------------------	--------------------	------------------	-------

- Στην εταιρεία προσλαμβάνεται ένας νέος υπάλληλος, ως οδηγός, με αριθμό διπλώματος 1111134 και ημερομηνία έκδοσης διπλώματος στις 23/5/2023, αλλά δεν του έχει δοθεί ακόμη αυτοκίνητο. Άρα θα εισαχθεί η παρακάτω γραμμή:
(1111134, 23/5/2023, NULL, NULL, NULL)
- Επιπλέον στον στόλο της εταιρείας προστέθηκε ένα αυτοκίνητο με αριθμό κυκλοφορίας IYN1111, με αγορά στις 22/5/2023, κόκκινου χρώματος, αλλά ακόμα δεν έχει αποφασισθεί ποιοι οδηγοί θα το οδηγούν. Αυτά τα δεδομένα δεν μπορούν να εισαχθούν στον παραπάνω πίνακα γιατί θα υπάρχει τιμή NULL και ως εκ τούτου θα παραβιαζόταν η ακεραιότητα οντότητας.
- Αυτά τα προβλήματα που προκύπτουν με την εισαγωγή δεδομένων σε έναν πίνακα αναφέρονται ως **ανωμαλίες εισαγωγής** (insertion anomaly).

Προβλήματα διαγραφής

- Έστω ότι υπάρχει ο παρακάτω πίνακας:

ΠΟΙΟΣ_ΟΔΗΓΟΣ_ΟΔΗΓΕΙ_ΠΟΙΟ_ΑΥΤΟΚΙΝΗΤΟ

ΑριθμόςΔιπλώματος	ΗμερομηνίαΈκδοσης	ΑριθμόςΚυκλοφορίας	ΗμερομηνίαΑγοράς	Χρώμα
1111123	9/2/2003	IYN11001	1/7/2009	Λευκό
1231111	15/10/2015	IYN11001	1/7/2009	Λευκό
1222211	1/9/1995	NNA2221	5/3/2013	Κόκκινο
1322211	13/9/2008	IYN11001	1/7/2009	Λευκό

- Αν διαγραφεί η εγγραφή με αριθμό διπλώματος 1222211, τότε θα χαθούν όλα τα στοιχεία που αναφέρονται στο αυτοκίνητο με αριθμό κυκλοφορίας NNA2211.
- Το πρόβλημα ανακύπτει γιατί αυτός ο οδηγός είναι ο μόνος που οδηγεί το συγκεκριμένο αυτοκίνητο.
- Αυτά τα προβλήματα που προκύπτουν με την απώλεια δεδομένων κατά την διαγραφή μιας μοναδικής εγγραφής σε έναν πίνακα αναφέρονται ως **ανωμαλίες διαγραφής** (deletion anomaly).

Προβλήματα τροποποίησης

- Έστω ότι υπάρχει ο παρακάτω πίνακας:

ΠΟΙΟΣ_ΟΔΗΓΟΣ_ΟΔΗΓΕΙ_ΠΟΙΟ_ΑΥΤΟΚΙΝΗΤΟ

ΑριθμόςΔιπλώματος	ΗμερομηνίαΈκδοσης	ΑριθμόςΚυκλοφορίας	ΗμερομηνίαΑγοράς	Χρώμα
1111123	9/2/2003	IYN11001	1/7/2009	Λευκό
1231111	15/10/2015	IYN11001	1/7/2009	Λευκό
1222211	1/9/1995	NNA2221	5/3/2013	Κόκκινο
1322211	13/9/2008	IYN11001	1/7/2009	Λευκό

- Αν για κάποιο λόγο γίνει τροποποίηση του χρώματος του αυτοκινήτου με αριθμό κυκλοφορίας IYN1001 σε πράσινο, τότε αυτή η αλλαγή θα πρέπει να γίνει και στις τρεις εγγραφές που φαίνονται παραπάνω, μιας και το αυτοκίνητο με αριθμό κυκλοφορίας IYN εμφανίζεται σε τρεις εγγραφές.
- Αυτά τα προβλήματα που προκύπτουν με την τροποποίηση δεδομένων μιας εγγραφής σε έναν πίνακα αναφέρονται ως **ανωμαλίες τροποποίησης** (update anomaly).



3. Κανονικοποίηση

Συναρτησιακή εξάρτηση

- Η **συναρτησιακή εξάρτηση** (functional dependency), είναι μια θεμελιώδης έννοια της κανονικοποίησης και καθορίζει ένα περιορισμό ανάμεσα στα πεδία ενός πίνακα.
- Αναφέρεται ότι για δύο πεδία A και B ενός πίνακα R ισχύει η συναρτησιακή εξάρτηση $A \rightarrow B$ αν κάθε τιμή του A συνδέεται με μία μόνο τιμή του B.
- Ο συμβολισμός $A \rightarrow B$ αναφέρεται ως «το B είναι συναρτησιακά εξαρτώμενο από το A».
- Ο τυπικός ορισμός της συναρτησιακής εξάρτησης είναι ο εξής:
«Έστω ένας πίνακας R που περιέχει δύο σύνολα πεδίων: το A και το B. Λέμε ότι $A \rightarrow B$ όταν δύο οποιεσδήποτε πλειάδες γ_1 και γ_2 του R που έχουν τις ίδιες τιμές στο A, έχουν τις ίδιες τιμές και στο B».

Παράδειγμα συναρτησιακής εξάρτησης

- Έστω ο παρακάτω πίνακας το A είναι το σύνολο των πεδίων {X1, X2} και το B είναι το σύνολο {X4, X5}. Έστω ότι ισχύει $A \rightarrow B$ και ζητείται να εξακριβωθεί εάν τα δεδομένα του πίνακα είναι συνεπή με αυτή την εξάρτηση.

A			B	
x1	x2	x3	x4	x5
α	β	ζ	1	2
α	δ	ψ	1	3
α	β	ω	1	2
γ	δ	ς	1	4

- Παρατηρώντας τον πίνακα, ισχύει ότι κάθε τιμή του τμήματος A κάθε πλειάδας παρουσιάζεται πάντα με την ίδια τιμή στο τμήμα B. Συνεπώς τα δεδομένα του πίνακα δεν παραβιάζουν την εξάρτηση $A \rightarrow B$ που δόθηκε.
- Η αναπαράσταση της συναρτησιακής εξάρτησης $A \rightarrow B$ χρησιμοποιώντας τα πεδία του πίνακα είναι η εξής: $X1, X2 \rightarrow X4, X5$.

Παράδειγμα πίνακα στον οποίο ισχύουν πολλές συναρτησιακές εξαρτήσεις

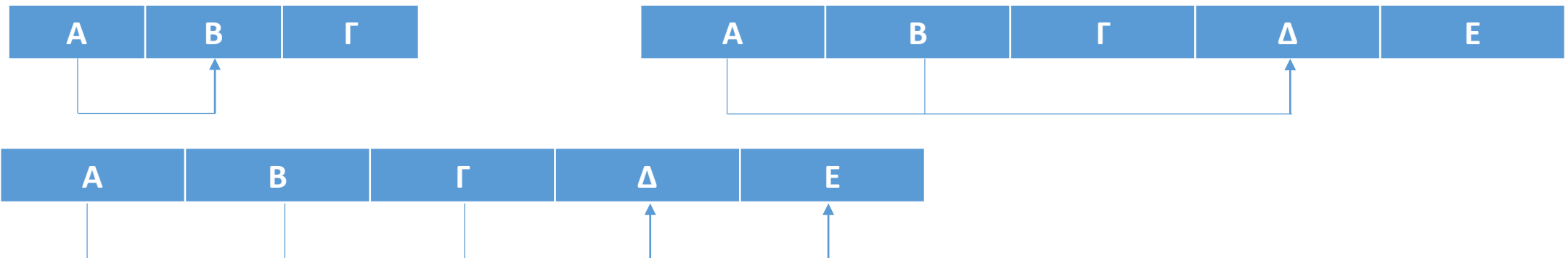
- Έστω ότι υπάρχει ο παρακάτω πίνακας:

ΕΡΓΟ	ΥΛΙΚΟ	ΠΟΣΟΤΗΤΑ	ΗΜΕΡ_ΕΝΑΡΞΗΣ	ΗΜΕΡ_ΛΗΞΗΣ	ΚΟΣΤΟΣ_ΜΟΝΑΔΑΣ
E1	X7	9	ΙΑΝ 90	ΜΑΡ 91	10
E1	X9	11	ΙΑΝ 90	ΜΑΡ 91	22
E2	X10	20	ΙΑΝ 90	ΜΑΡ 91	5
E2	X9	17	ΙΟΥΝ 90	ΙΟΥΛ 91	22
E2	X9	12	ΙΟΥΝ 90	ΙΟΥΛ 91	6

- Γι' αυτόν τον πίνακα ισχύουν οι συναρτησιακές εξαρτήσεις:
 - ΕΡΓΟ → ΗΜΕΡ_ΕΝΑΡΞΗΣ, ΗΜΕΡ_ΛΗΞΗΣ
 - ΕΡΓΟ, ΥΛΙΚΟ → ΠΟΣΟΤΗΤΑ
 - ΥΛΙΚΟ → ΚΟΣΤΟΣ_ΜΟΝΑΔΑΣ
- Εάν προστεθεί η εγγραφή (E2, X7, 17, ΙΟΥΝ 90, ΑΥΓ 91, 7) τότε παραβιάζονται οι συναρτησιακές εξαρτήσεις:
 - ΕΡΓΟ → ΗΜΕΡ_ΕΝΑΡΞΗΣ, ΗΜΕΡ_ΛΗΞΗΣ
 - ΥΛΙΚΟ → ΚΟΣΤΟΣ

Συμβολισμοί συναρτησιακών εξαρτήσεων

- Μερικές φορές είναι πιο βολικό μιας συναρτησιακή εξάρτηση $A \rightarrow B$, η οποία ισχύει για τα πεδία ενός πίνακα να αναπαριστάνεται επάνω στον πίνακα, όπως παρακάτω:



Παρατηρήσεις

- Η συναρτησιακή εξάρτηση προκύπτει από το νόημα που έχουν τα δεδομένα και έχει επίδραση στις τιμές των δεδομένων.
- Όμως δεν ισχύει το αντίθετο.
- Οι τιμές των δεδομένων δεν καθορίζουν τις συναρτησιακές εξαρτήσεις, και έτσι δεν είναι δυνατόν κοιτώντας απλώς τα δεδομένα ενός πίνακα να προσδιοριστούν οι συναρτησιακές εξαρτήσεις που ισχύουν στον πίνακα.

Η κανονικοποίηση σαν σειρά βημάτων

- Η **κανονικοποίηση** είναι μια τεχνική, η οποία επιτρέπει :
 1. την ανάλυση πινάκων, ώστε να διαπιστωθεί αν οι πίνακες είναι κακώς σχεδιασμένοι, και
 2. τον επανασχεδιασμό κακώς σχεδιασμένων πινάκων διασπώντας τους σε μικρότερους πίνακες ώστε να αρθεί η κακή σχεδίασή τους.
- Η κανονικοποίηση εκτελείται σαν μια σειρά βημάτων, όπου σε κάθε βήμα αναλύεται κάθε πίνακας της ΒΔ χρησιμοποιώντας ένα συγκεκριμένο κριτήριο-κανόνα που λέγεται **κανονική μορφή** (normal form):
 - Το πρώτο βήμα της κανονικοποίησης λέγεται πρώτη κανονική μορφή 1KM ή 1NF,
 - Το δεύτερο βήμα λέγεται δεύτερη κανονική μορφή 2KM ή 2NF,
 - Το τρίτο βήμα λέγεται τρίτη κανονική μορφή 3KM ή 3NF.
 - Εάν κάποιο κριτήριο-κανόνας παραβιάζεται, τότε ο πίνακας διασπάται σε απλούστερους πίνακες οι οποίοι ικανοποιούν το κριτήριο.



4. Πρώτη κανονική μορφή

Πρώτη κανονική μορφή

- Ένας πίνακας είναι στην πρώτη κανονική μορφή (1NF), εάν η κάθε τιμή που αποθηκεύεται στον πίνακα είναι:
 - ατομική και αδιαίρετη.
- Ένας πίνακας που δεν είναι στην 1NF μπορεί να κανονικοποιηθεί στη μορφή αυτή είτε:
 - Ορίζοντας το κατάλληλο σύνθετο πρωτεύων κλειδί, είτε
 - Διασπώντας τον πίνακα σε περισσότερους πίνακες.

Παράδειγμα ενός πίνακα αν είναι στην 1NF.

- Μια εταιρεία πουλάει διάφορα προϊόντα σε διάφορα σημεία πώλησης (περίπτερα, mini-market, super-market κ.λπ.) σε διαφορετικές πόλεις.

Πόλη	Πληθυσμός	Σοκολάτα	Σημεία Πώλησης
Πάτρα	300000	Σοκολάτα αμυγδάλου	250
		Σοκολάτα λευκή	120
		Σοκολάτα υγείας	270
Λάρισα	200000	Σοκολάτα με φουντούκια	50
		Σοκολάτα με σταφίδες	35

- Ο παραπάνω πίνακας δεν είναι στην πρώτη κατανομή 1NF, γιατί υπάρχουν τιμές που δεν είναι ατομικές, όπως για παράδειγμα στην πόλη Πάτρα υπάρχουν στα πεδία σοκολάτες και σημεία πώλησης τρεις τιμές και για την πόλη της Λάρισας για τα ίδια πεδία υπάρχουν δύο τιμές.

Παράδειγμα ενός πίνακα αν είναι στην 1NF..

- 1NF – Πρώτο τρόπος: ορίζεται ως σύνθετο πρωτεύον κλειδί του πίνακα το ζεύγος (Πόλη, Σοκολάτα).

Πόλη	Πληθυσμός	Σοκολάτα	ΣημείαΠώλησης
Πάτρα	300000	Σοκολάτα αμυγδάλου	250
		Σοκολάτα λευκή	120
		Σοκολάτα υγείας	270
Λάρισα	200000	Σοκολάτα με φουντούκια	50
		Σοκολάτα με σταφίδες	35
<u>Πόλη</u>	Πληθυσμός	<u>Σοκολάτα</u>	ΣημείαΠώλησης
Πάτρα	300000	Σοκολάτα αμυγδάλου	250
Πάτρα	300000	Σοκολάτα λευκή	120
Πάτρα	300000	Σοκολάτα υγείας	270
Λάρισα	200000	Σοκολάτα με φουντούκια	50
Λάρισα	200000	Σοκολάτα με σταφίδες	35

Παράδειγμα ενός πίνακα αν είναι στην 1NF...

- 1NF – Δεύτερος τρόπος: Ο πίνακας διασπάται σε δύο πίνακες που συνδέονται μεταξύ τους μέσω της κοινής στήλης Πόλη.

Πόλη	Πληθυσμός	Σοκολάτα	ΣημείαΠώλησης
Πάτρα	300000	Σοκολάτα αμυγδάλου	250
		Σοκολάτα λευκή	120
		Σοκολάτα υγείας	270
Λάρισα	200000	Σοκολάτα με φουντούκια	50
		Σοκολάτα με σταφίδες	35

		ξ.κ.			
Πόλη	Πληθυσμός	↓	Πόλη	Σοκολάτα	ΣημείαΠώλησης
			Πάτρα	Σοκολάτα αμυγδάλου	250
			Πάτρα	Σοκολάτα λευκή	120
Πάτρα	300000		Πάτρα	Σοκολάτα υγείας	270
			Λάρισα	Σοκολάτα με φουντούκια	50
			Λάρισα	Σοκολάτα με σταφίδες	35

Επισημάνσεις για την 1NF

- Η πρώτη κανονική μορφή 1NF, είναι η μόνη μορφή που είναι υποχρεωτική για όλους τους πίνακες μιας σχεσιακής ΒΔ.
- Συνεπώς, ένας πίνακας για να μπορέσει να υλοποιηθεί από ένα σχεσιακό ΣΔΒΔ, πρέπει να βρίσκεται στην πρώτη κανονική κατανομή.
- Οι άλλες κανονικές μορφές είναι προαιρετικές. Δηλαδή, ακόμη και αν ένας πίνακας δεν βρίσκεται σε κάποια από τις κανονικές μορφές μπορεί να υλοποιηθεί σε αν σχεσιακό ΣΔΒΔ παρά το γεγονός ότι δεν θα είναι ένας καλά σχεδιασμένος πίνακας.

Πλήρης συναρτησιακή εξάρτηση

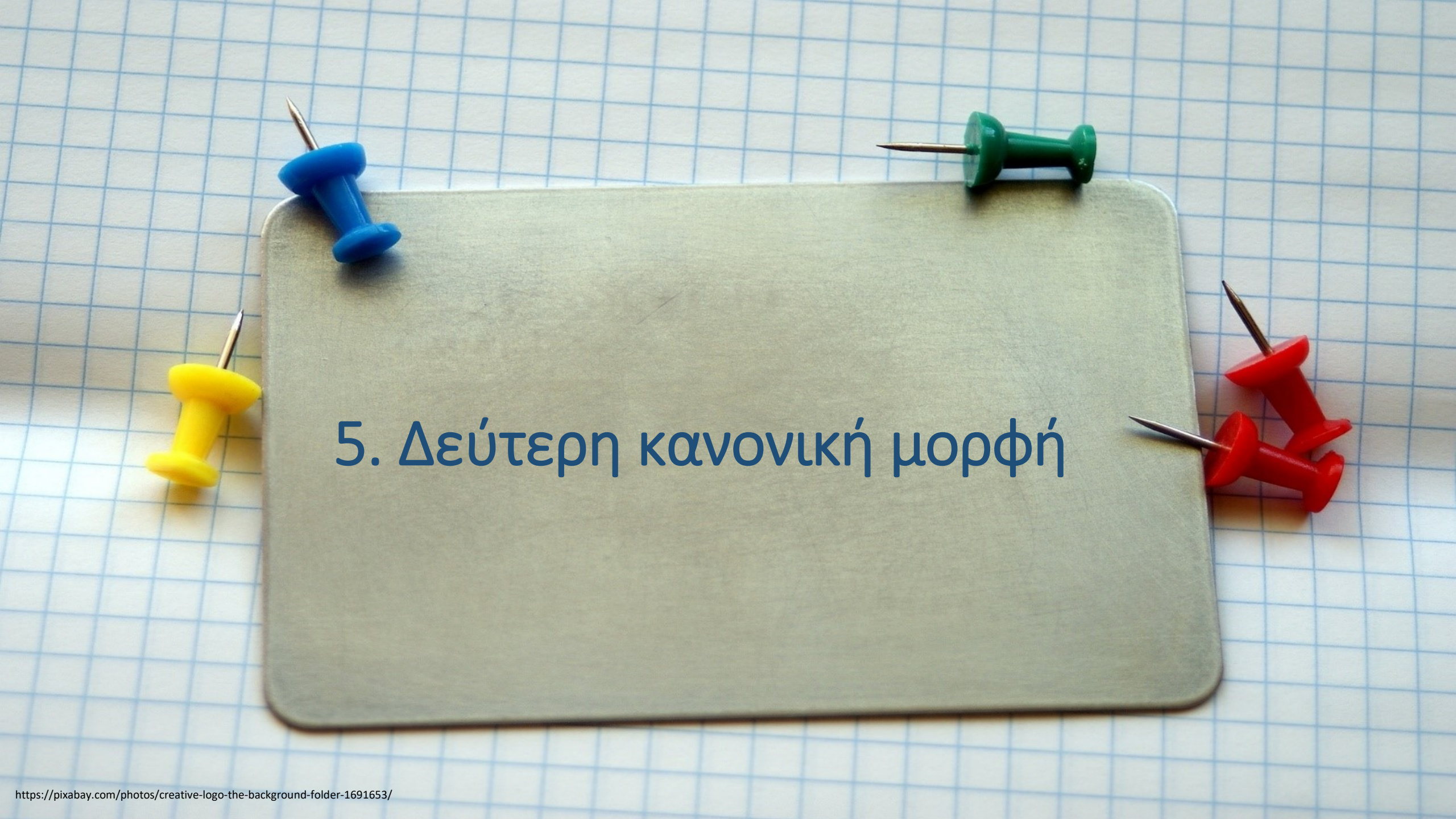
- Η δεύτερη κανονική μορφή βασίζεται στην έννοια της **πλήρους συναρτησιακής εξάρτησης** (full functional dependency).
- Μια συναρτησιακή εξάρτηση $A \rightarrow B$ είναι πλήρης, εάν όλα τα πεδία που ανήκουν στο A απαιτούνται για την ύπαρξη της εξάρτησης. Εάν αφαιρεθεί οποιοδήποτε πεδίο του A , τότε υπάρχει σαν συνέπεια η μη ισχύ της εξάρτησης $A \rightarrow B$.
- Μια συναρτησιακή εξάρτηση που δεν είναι πλήρης ονομάζεται **μερική**.

Παράδειγμα πλήρους και μερικής συναρτησιακής εξάρτησης

- Έστω ότι υπάρχει ο παρακάτω πίνακας.

Πόλη	Πληθυσμός	Σοκολάτα	Σημεία Πώλησης
Πάτρα	300000	Σοκολάτα αμυγδάλου	250
		Σοκολάτα λευκή	120
		Σοκολάτα υγείας	270
Λάρισα	200000	Σοκολάτα με φουντούκια	50
		Σοκολάτα με σταφίδες	35

- Η συναρτησιακή εξάρτηση {Πόλη, Πληθυσμός, Σοκολάτα} $\rightarrow$ Σημεία Πώλησης δεν είναι πλήρης, ακόμη και αν αφαιρεθεί το πεδίο Πληθυσμός από τη συναρτησιακή εξάρτηση, η εξάρτηση εξακολουθεί να ισχύει.
- Η συναρτησιακή εξάρτηση {Πόλη, Σοκολάτα} $\rightarrow$ Σημεία Πώλησης είναι πλήρης γιατί αν αφαιρεθεί το πεδίο Πόλη ή το πεδίο Σοκολάτα, η εξάρτηση παύει να ισχύει.



5. Δεύτερη κανονική μορφή

Δεύτερη κανονική μορφή

- Στη δεύτερη κανονική μορφή (2NF) ανήκει ένας πίνακας, εάν είναι στην πρώτη κανονική μορφή και κάθε μη πρωτεύον πεδίο του εξαρτάται συναρτησιακά πλήρως από το πρωτεύον κλειδί.

Παράδειγμα

- Έστω ο παρακάτω πίνακας:

ΣΗΜΕΙΑ_ΠΩΛΗΣΗΣ_ΣΟΚΟΛΑΤΑΣ



- Τότε, το πεδίο Πληθυσμός είναι μη πρωτεύον κλειδί και δεν είναι πλήρως εξαρτημένο από το πρωτεύον κλειδί.
- Ο πίνακας δεν ανήκει στην δεύτερη κανονική μορφή 2NF εξαιτίας της συναρτησιακής εξάρτησης ② .

Παράδειγμα κανονικοποίησης πίνακα στη δεύτερη κανονική μορφή

- Δίνεται ο παρακάτω πίνακας που δεν είναι στη 2^η κανονική μορφή:

ΣΗΜΕΙΑ_ΠΩΛΗΣΗΣ_ΣΟΚΟΛΑΤΑΣ

<u>Πόλη</u>	Πληθυσμός	<u>Σοκολάτα</u>	ΣημείαΠώλησης

- Τότε, πίνακας διασπάται σε δύο πίνακες που είναι στη 2^η κανονική μορφή:

ΠΟΛΗ

<u>Πόλη</u>	Πληθυσμός

ΠΩΛΗΣΗ

<u>Πόλη</u>	<u>Σοκολάτα</u>	ΣημείαΠώλησης

Μεταβατική συναρτησιακή εξάρτηση.

- Η τρίτη κανονική μορφή 3NF βασίζεται στην έννοια της **μεταβατικής συναρτησιακής εξάρτησης**.
- Μια συναρτησιακή εξάρτηση $A \rightarrow B$ ονομάζεται μεταβατική, εάν υπάρχει ένα σύνολο πεδίων του B το οποίο δεν είναι υποσύνολο κανενός κλειδιού, τέτοιο ώστε να ισχύουν οι εξαρτήσεις: $A \rightarrow B$ και $A \rightarrow C$.
- Παράδειγμα μεταβατικής συναρτησιακής εξάρτησης:

ΠΛΗΘΥΣΜΟΣ_ΚΑΙ_ΑΕΠ_ΜΙΑΣ_ΧΩΡΑΣ



Μεταβατική συναρτησιακή εξάρτηση..

- Σε αυτόν τον πίνακα υπάρχει η μεταβατική συναρτησιακή εξάρτηση:
Πόλη→ΑκαθάριστοΠροϊόνΠεριφέρειας, επειδή υπάρχει το πεδίο Περιφέρεια για το οποίο ισχύει:
Πόλη→Περιφέρεια, και Περιφέρεια→ΑκαθάριστοΠροϊόνΠεριφέρειας
- Άρα το πεδίο ΑκαθάριστοΠροϊόνΠεριφέρειας είναι μεταβατικά εξαρτώμενο από το πεδίο Πόλη.

ΠΛΗΘΥΣΜΟΣ_ΚΑΙ_ΑΕΠ_ΜΙΑΣ_ΧΩΡΑΣ





6. Τρίτη κανονική μορφή

Τρίτη κανονική μορφή

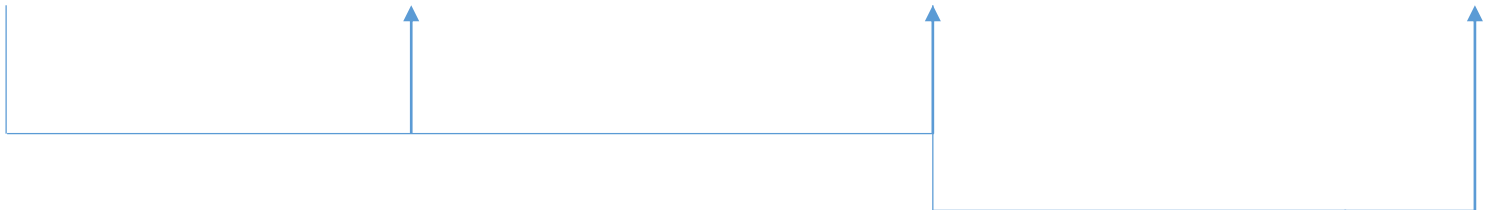
- Στην τρίτη κανονική μορφή (3NF) ανήκει ένας πίνακας, εάν ανήκει και στην πρώτη και την δεύτερη κανονική μορφή και δεν έχει μη πρωτεύον πεδίο του που να είναι μεταβατικά εξαρτώμενο από το πρωτεύον κλειδί του πίνακα.
- Ένας πίνακας που δεν είναι στην τρίτη κανονική μορφή, μπορεί να διασπαστεί σε μικρότερους πίνακες σύμφωνα με τον τρόπο που υποδεικνύουν οι λειτουργικές εξαρτήσεις.
- Οι πίνακες που προκύπτουν με αυτό τον τρόπο ανήκουν στην τρίτη κανονική μορφή (3NF).

Παράδειγμα κανονικοποίησης πίνακα στην τρίτη κανονική μορφή

- Έστω ότι υπάρχει ο παρακάτω πίνακας που δεν είναι στην τρίτη κανονική μορφή:

ΠΛΗΘΥΣΜΟΣ_ΚΑΙ_ΑΕΠ_ΜΙΑΣ_ΧΩΡΑΣ

<u>Πόλη</u>	Πληθυσμός	Περιφέρεια	Ακαθάριστο Προϊόν Περιφέρειας



- Τότε ο πίνακας διασπάται σε δύο πίνακες που είναι στην τρίτη κανονική μορφή:

ΠΟΛΕΙΣ

<u>Πόλη</u>	Πληθυσμός	Περιφέρεια



ΠΕΡΙΦΕΡΕΙΕΣ

<u>Περιφέρεια</u>	Ακαθάριστο Προϊόν Περιφέρειας





7. Κανονική μορφή Boyce-Codd

Κανονική μορφή Boyce-Codd

- Η τρίτη κανονική μορφή (3NF) δεν περιλαμβάνει μια περίπτωση κακώς σχεδιασμένων πινάκων, η οποία μπορεί σπάνια να προκύψει.
- Για να αντιμετωπιστεί αυτή η περίπτωση προτάθηκε η κανονική μορφή Boyce-Codd, η οποία είναι ισχυρότερη από την τρίτη κανονική μορφή και βασίζεται στην έννοια του υποψήφιου κλειδιού.
- Σε έναν πίνακα μπορεί να υπάρχουν περισσότερα από ένα πεδία, τα οποία να μπορούν να πάρουν διαφορετικές τιμές για κάθε εγγραφή του πίνακα και αυτά τα πεδία λέγονται **υποψήφια κλειδιά**.
- Ένα από τα υποψήφια κλειδιά γίνεται το πρωτεύον κλειδί του πίνακα.

Παράδειγμα ελέγχου στην κανονική μορφή Boyce-Codd.

- Έστω μία εταιρία διαχειρίζεται έναν μεγάλο αριθμό έργων από ευρωπαϊκά προγράμματα και ισχύουν τα παρακάτω:
 - Κάθε έργο αποτελείται από πολλές δραστηριότητες,
 - Διαφορετικά έργα μπορεί να έχουν τις ίδιες δραστηριότητες, αλλά σε κάθε έργο μια δραστηριότητα εκτελείται μία μόνο φορά,
 - Κάθε δραστηριότητα ενός έργου επιβλέπεται από έναν εργαζόμενο της εταιρίας που ορίζεται ως υπεύθυνος δραστηριότητας,
 - Κάθε εργαζόμενος στην εταιρία εξειδικεύεται να επιβλέπει μια συγκεκριμένη δραστηριότητα.
- Το ζεύγος (Έργο, Δραστηριότητα) είναι ένα υποψήφιο κλειδί, γιατί σε ένα έργο κάθε δραστηριότητα εμφανίζεται μία μόνο φορά.
- Το ζεύγος (Έργο, Υπεύθυνος Δραστηριότητας) είναι επίσης υποψήφιο κλειδί επειδή ένας υπεύθυνος δραστηριότητας χειρίζεται ένα μόνο είδος δραστηριότητας και σε ένα έργο μια δραστηριότητα εμφανίζεται μία μόνο φορά.

Παράδειγμα ελέγχου στην κανονική μορφή Boyce-Codd..

- Στον παρακάτω πίνακα φαίνονται τα δεδομένα που αφορούν τις δραστηριότητες των έργων και τους υπεύθυνους κάθε δραστηριότητας.

ΕΡΓΑ_ΚΑΙ_ΥΠΕΥΘΥΝΟΙ_ΤΩΝ_ΕΡΓΩΝ

Έργο	Δραστηριότητα	ΥπεύθυνοςΔραστηριότητας
ΕΡΓ1	Προδιαγραφές	Αναστασίου
ΕΡΓ1	Προκήρυξη	Ιωάννου
ΕΡΓ1	Κατακύρωση διαγωνισμού	Πέτρου
ΕΡΓ2	Προδιαγραφές	Αναστασίου
ΕΡΓ2	Προκήρυξη	Ιωάννου
ΕΡΓ2	Κατακύρωση διαγωνισμού	Πέτρου

- Επιπλέον ισχύουν οι συναρτησιακές εξαρτήσεις:
 - ΥπεύθυνοςΔραστηριότητας → Δραστηριότητα
 - {Έργο, Δραστηριότητα} → ΥπεύθυνοςΔραστηριότητας
 - {Έργο, ΥπεύθυνοςΔραστηριότητας} → Δραστηριότητα

Παράδειγμα ελέγχου στην κανονική μορφή Boyce-Codd...

- Με βάση τα δεδομένα που αναφέρθηκαν προκύπτουν οι συναρτησιακές εξαρτήσεις του παρακάτω πίνακα:

ΕΡΓΑ_ΚΑΙ_ΥΠΕΥΘΥΝΟΙ_ΤΩΝ_ΕΡΓΩΝ

	Έργο	Δραστηριότητα	ΥπεύθυνοςΔραστηριότητας
①			
②			
			③

- Ο παραπάνω πίνακας δεν ανήκει στην κανονική μορφή Boyce-Codd, επειδή στην συναρτησιακή εξάρτηση ③ η ορίζουσα της εξάρτησης (ΥπεύθυνοςΔραστηριότητας) δεν είναι υποψήφιο κλειδί του πίνακα.

Διάσπαση πίνακα που δεν ανήκει στην κανονική μορφή Boyce-Codd

- Ένας πίνακας που δεν ανήκει στην κανονική μορφή Boyce-Codd, διασπάται σε μικρότερους πίνακες και ο καθένας από αυτούς ανήκει στην κανονική μορφή Boyce-Codd.

Αρχικός Πίνακας

Έργο	Δραστηριότητα	ΥπεύθυνοςΔραστηριότητας
ΕΡΓ1	Προδιαγραφές	Αναστασίου
ΕΡΓ1	Προκήρυξη	Ιωάννου
ΕΡΓ1	Κατακύρωση διαγωνισμού	Πέτρου
ΕΡΓ2	Προδιαγραφές	Αναστασίου
ΕΡΓ2	Προκήρυξη	Ιωάννου
ΕΡΓ2	Κατακύρωση διαγωνισμού	Πέτρου



Προκύπτουν δύο πίνακες
από την διάσπαση του Αρχικού Πίνακα

Έργο	ΥπεύθυνοςΔραστηριότητας
ΕΡΓ1	Αναστασίου
ΕΡΓ1	Ιωάννου
ΕΡΓ1	Πέτρου
ΕΡΓ2	Αναστασίου
ΕΡΓ2	Ιωάννου
ΕΡΓ2	Πέτρου

Δραστηριότητα	ΥπεύθυνοςΔραστηριότητας
Προδιαγραφές	Αναστασίου
Προκήρυξη	Ιωάννου
Κατακύρωση διαγωνισμού	Πέτρου



8. Εξάρτηση πολλαπλών τιμών

Εξάρτηση πολλαπλών τιμών.

- Μία εξάρτηση πολλαπλών τιμών υπάρχει στα πεδία A, B, Γ ενός πίνακα, όταν για κάθε τιμή του A υπάρχει ένα σύνολο τιμών του B και ένα ανεξάρτητο σύνολο τιμών του Γ.
- Η εξάρτηση πολλαπλών τιμών ανάμεσα στα A, B, Γ συμβολίζεται ως εξής: $A \twoheadrightarrow B$ και $A \twoheadrightarrow \Gamma$.
- *Παράδειγμα*

Πόλη	Σοκολάτα	Περιοδικό
Πάτρα	Σοκολάτα αμυγδάλου Σοκολάτα λευκή Σοκολάτα υγείας	Δυτική Ελλάς Πάτραι Νέα της Πάτρας
Λάρισα	Σοκολάτα με φουντούκια Σοκολάτα με σταφίδες	Θεσσαλία

- Οπότε ισχύουν οι εξαρτήσεις:
 - Πόλη $\twoheadrightarrow$ Σοκολάτα
 - Πόλη $\twoheadrightarrow$ Περιοδικό

Εξάρτηση πολλαπλών τιμών..

- Μία εξάρτηση πολλαπλών τιμών $A \twoheadrightarrow B$ που ισχύει σέ έναν πίνακα R λέγεται τετριμμένη αν:
 - το B είναι υποσύνολο του A ή αν,
 - $A \cup B = R$
- Μια εξάρτηση πολλαπλών τιμών $A \twoheadrightarrow B$ που ισχύει σε έναν πίνακα λέγεται μη τετριμμένη αν:
 - το B δεν είναι υποσύνολο του A και
 - $A \cup B \neq R$

Παράδειγμα μη τετριμμένης εξάρτησης πολλαπλών τιμών

- Έστω η εξάρτηση πολλαπλών τιμών Πόλη → Σοκολάτα, όπως φαίνεται παρακάτω:

Πόλη	Σοκολάτα	Περιοδικό
Πάτρα	Σοκολάτα αμυγδάλου Σοκολάτα λευκή Σοκολάτα υγείας	Δυτική Ελλάς Πάτραι Νέα της Πάτρας
Λάρισα	Σοκολάτα με φουντούκια Σοκολάτα με σταφίδες	Θεσσαλία

- Για την εξαίρεση αυτή: η Σοκολάτα δεν είναι υποσύνολο της Πόλης και η ένωση της Πόλης με την Σοκολάτα δεν δημιουργεί τον πίνακα ΣΟΚΟΛΑΤΕΣ_ΚΑΙ_ΠΕΡΙΟΔΙΚΑ, επειδή απουσιάζει το πεδίο Περιοδικό.
- Συνεπώς, η εξάρτηση Πόλη → Σοκολάτα είναι μη τετριμμένη.

Τέταρτη κανονική μορφή.

- Ένας πίνακας ανήκει στην τέταρτη κανονική μορφή (4NF), αν ανήκει στην κανονική μορφή Boyce-Codd και δεν περιέχει μη τετριμμένες εξαρτήσεις πολλαπλών τιμών.

Παράδειγμα ελέγχου τέταρτης κανονικής μορφής

- Στον παρακάτω πίνακα ΣΟΚΟΛΑΤΕΣ_ΚΑΙ_ΕΦΗΜΕΡΙΔΕΣ δεν ανήκει στην τέταρτη κανονική μορφή, γιατί αν και ανήκει στη μορφή Boyce-Codd περιέχει δύο μη τετριμμένες εξαρτήσεις πολλαπλών τιμών:
 - Πόλη → Σοκολάτα και
 - Πόλη → Περιοδικό

Πόλη	Σοκολάτα	Περιοδικό
Πάτρα	Σοκολάτα αμυγδάλου Σοκολάτα λευκή Σοκολάτα υγείας	Δυτική Ελλάς Πάτραι Νέα της Πάτρας
Λάρισα	Σοκολάτα με φουντούκια Σοκολάτα με σταφίδες	Θεσσαλία

Τέταρτη κανονική μορφή..

- Ένας πίνακας που δεν ανήκει στην τέταρτη κανονική μορφή 4NF, μπορεί να κανονικοποιηθεί στην 4NF με διάσπαση.
- Για κάθε εξάρτηση πολλαπλών τιμών δημιουργείται ένας νέος πίνακας που περιέχει το πεδίο πολλαπλών τιμών μαζί με ένα αντίγραφο της ορίζουσας της εξάρτησης πολλαπλών τιμών.

Τέταρτη κανονική μορφή...

Ο αρχικός πίνακας με εξαρτήσεις πολλαπλών τιμών

Πόλη	Σοκολάτα	Περιοδικό
Πάτρα	Σοκολάτα αμυγδάλου Σοκολάτα λευκή Σοκολάτα υγείας	Δυτική Ελλάς Πάτραι Νέα της Πάτρας
Λάρισα	Σοκολάτα με φουντούκια Σοκολάτα με σταφίδες	Θεσσαλία

Οι εξαρτήσεις που ισχύουν στον αρχικό πίνακα:

- Πόλη → Σοκολάτα και
- Πόλη → Περιοδικό



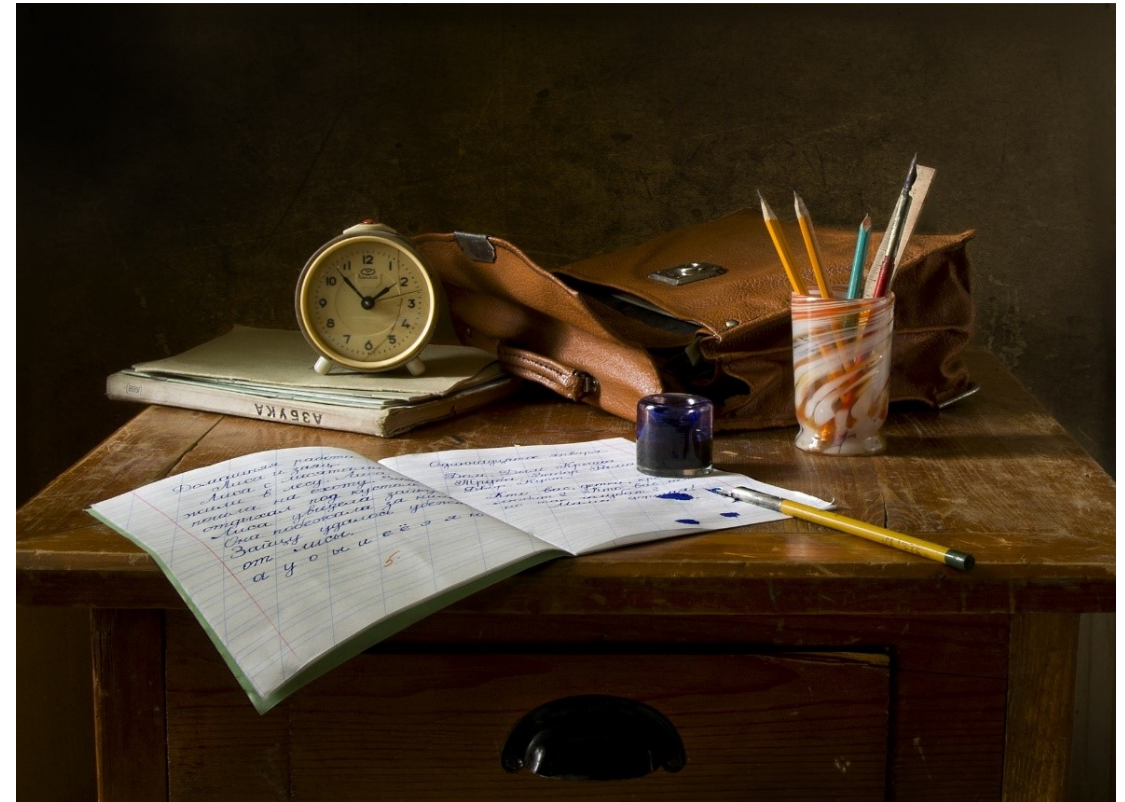
Οι δύο πίνακες τέταρτης κανονικής μορφής που προκύπτουν από τη διάσπαση του αρχικού πίνακα

Πόλη	Σοκολάτα
Πάτρα	Σοκολάτα αμυγδάλου
Πάτρα	Σοκολάτα λευκή
Πάτρα	Σοκολάτα υγείας
Λάρισα	Σοκολάτα με φουντούκια
Λάρισα	Σοκολάτα με σταφίδες

Πόλη	Περιοδικό
Πάτρα	Δυτική Ελλάς
Πάτρα	Πάτραι
Πάτρα	Νέα της Πάτρας
Λάρισα	Θεσσαλία

ΜΕΛΕΤΗ

- [1] **Σχεσιακές Βάσεις Δεδομένων (3^η έκδοση)**
(2021), Ε. Κεχρής.
7^ο Κεφάλαιο, σελ. 281-308



Ερωτήσεις Κατανόησης

1. Οι πίνακες που δεν είναι καλά σχεδιασμένοι τι προβλήματα μπορεί να παρουσιάζουν;
2. Τι είναι η κανονικοποίηση;
3. Πότε ένας πίνακας είναι στην πρώτη κανονική μορφή (1NF);
4. Ένας πίνακας που δεν είναι στην πρώτη κανονική μορφή (1NF) πως μπορεί να κανονικοποιηθεί σε αυτή τη μορφή;
5. Πότε ένας πίνακας ανήκει στην δεύτερη κανονική μορφή (2NF);
6. Πότε ένας πίνακας ανήκει στην Τρίτη κανονική μορφή (3NF);
7. Πότε ένας πίνακας ανήκει στην τέταρτη κανονική μορφή (4NF)