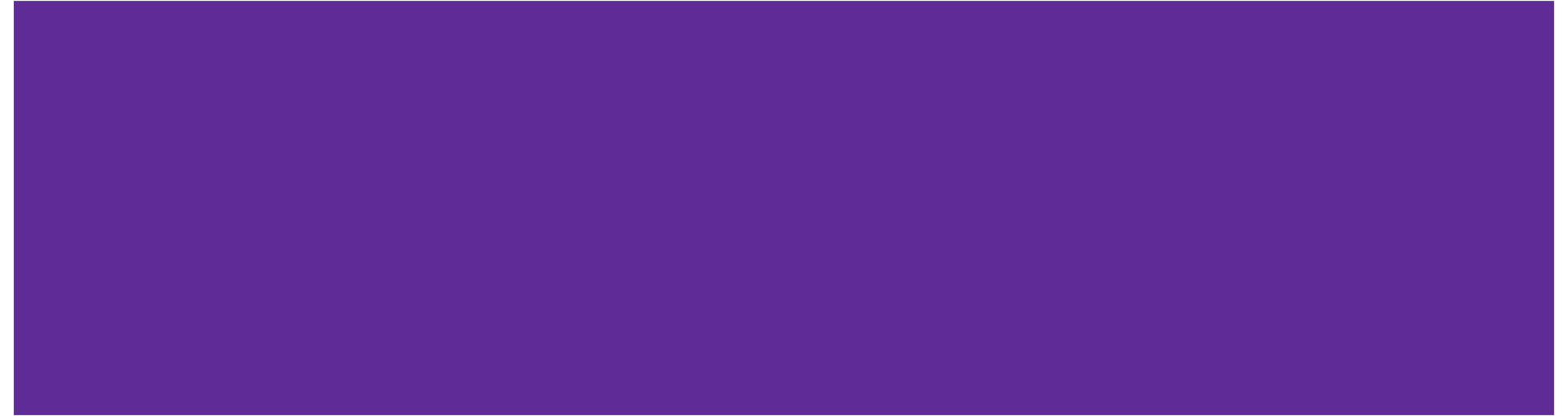


Κανονικές Μορφές



Ορισμός

Κανονικοποίηση είναι μια διαδικασία μετατροπής δεδομένων που έχουν παραπάνω από μια πιθανές αναπαραστάσεις σε μια 'τυπική' ή 'κανονική' μορφή.

Γιατί κανονικοποιούμε: Θέλουμε να μειώσουμε τον πλεονασμό των δεδομένων (data redundancy) καθώς προκύπτουν ανωμαλίες,

- ▶ Εγγραφής (*Insertion anomalies*).
- ▶ Ενημέρωσης (*Update anomalies*).
- ▶ Διαγραφής (*Deletion anomalies*).

Προσπαθούμε δηλαδή να μετασχηματίσουμε ένα σχεσιακό σχήμα σε μια νέα μορφή, η οποία είναι απαλλαγμένη από ανωμαλίες εισαγωγής, διαγραφής και τροποποίησης εγγράφων.

Ανωμαλία ΕΓΓΡΑΦΗΣ

<u>StudentID</u>	StudentName	<u>CourseID</u>	CourseName
101	Alice	100	C
102	Bob	150	C++
		252	Java

Κατά την εισαγωγή νέου μαθήματος θα πρέπει να γίνει assign και σε κάποιο φοιτητή για να είναι έγκυρη η πλειάδα.

Ανωμαλία ΕΝΗΜΕΡΩΣΗΣ

<u>StudentID</u>	StudentName	<u>CourseID</u>	CourseName
101	Alice	100	C → Python
102	Bob	150	C++
103	Carol	100	C → Python

Κατά την ενημέρωση του ονόματος ενός μαθήματος, πχ C -> Python, θα πρέπει να ελεγχθούν και ενημερωθούν όλες οι σχετικές πλειάδες. Σε περίπτωση που κάπου αποτύχει η ενημέρωση, θα έχουμε ασυνέπεια στον πίνακα.

Ανωμαλία ΔΙΑΓΡΑΦΗΣ

<u>StudentID</u>	StudentName	<u>CourseID</u>	CourseName
101	Alice	100	C
101	Alice	150	C++
102	Bob	150	C++

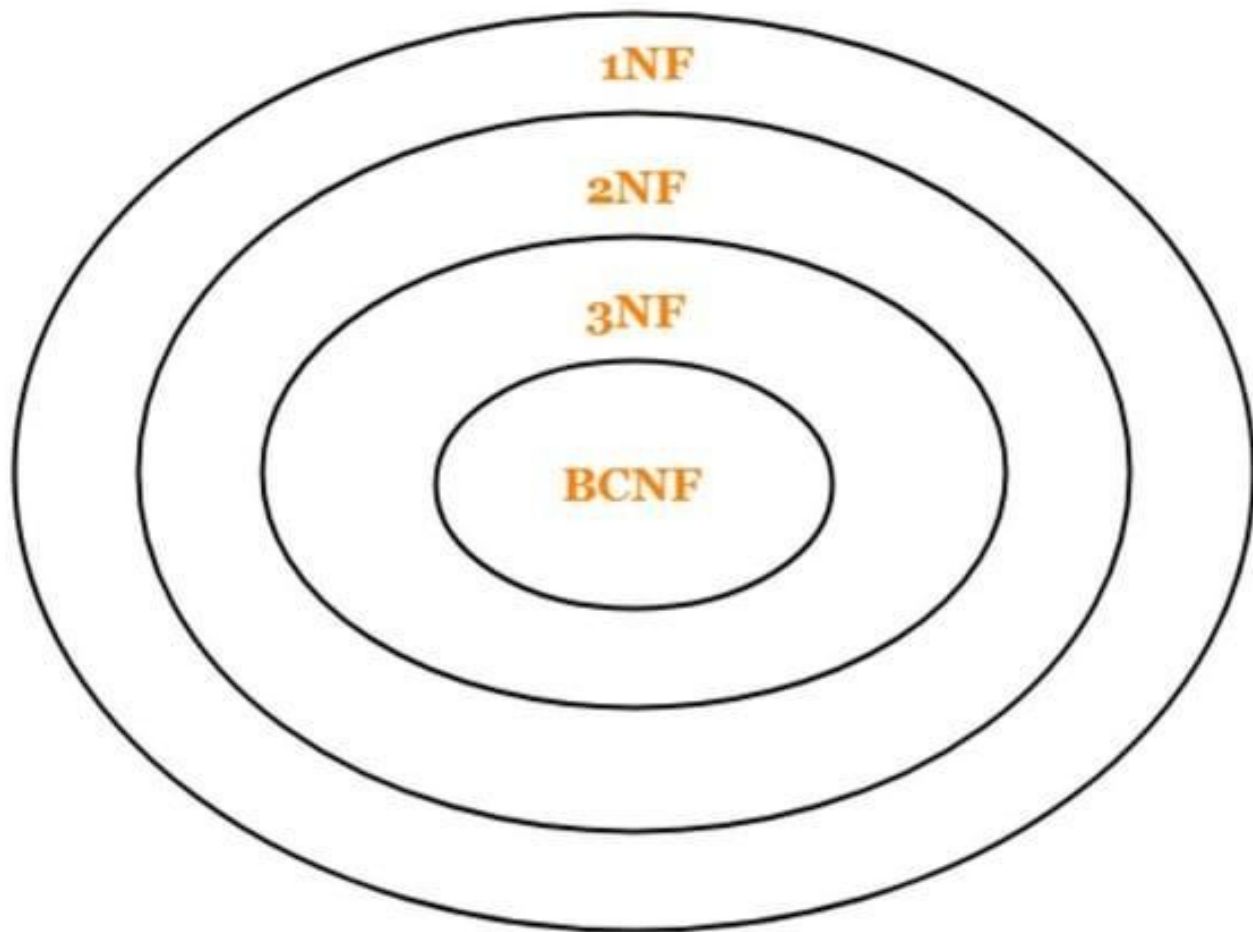
Κατά την διαγραφή κάποιου φοιτητή θα χρειαστεί να σβηστεί και το αντίστοιχο μάθημα στην πλειάδα, αυτό μπορεί να οδηγήσει στο να χαθεί από την βάση κάποιο μάθημα. Στην περίπτωση διαγραφής της Alice θα χαθεί το μάθημα 100.

Πλήρης vs Μερική Συναρτησιακή Εξάρτηση

Όταν κάποιο υποσύνολο του αριστερού μέρος μιας ΣΕ, μπορεί να προσδιορίσει συναρτησιακά το δεξιό, τότε μιλάμε για μια μερική ΣΕ, αλλιώς είναι πλήρης.

- Εστω $R = \{A,B,C,D\}$ και σύνολα εξαρτήσεων $F = \{ABC \rightarrow D\}$ και $P = \{AC \rightarrow B, A \rightarrow D, D \rightarrow B\}$.

- Πλήρης (Full): Στην περίπτωση που έχουμε την F , η $ABC \rightarrow D$ είναι πλήρης, διότι το D είναι εξαρτώμενο από το ABC αποκλειστικά και δεν μπορεί να αντικαταστήσει το ABC κανένα υποσύνολο του AB, AC, A κλπ.
- Μερική (Partial): Στην περίπτωση που έχουμε την P , η $AC \rightarrow B$ είναι μερική, διότι η κλειστότητα του A είναι $\{A^+\} = ADB$. Αυτό σημαίνει ότι το A από μόνο του μπορεί να προσδιορίσει το B και δεν χρειάζεται και το C .



1NF (Πρώτη Κανονική Μορφή):

- Κάθε πεδίο περιέχει ατομικές τιμές (όχι σύνθετες, όχι πολλαπλές).
 - πχ: σε μια στήλη: Integer, String
 - πχ: στην στήλη Τηλέφωνο να έχει 2 Integers
- Δεν υπάρχουν επαναλαμβανόμενα γνωρίσματα σε μια πλειάδα. (Τηλ1, Τηλ2)
- Δεν υπάρχουν επαναλαμβανόμενες πλειάδες.

2NF (Δεύτερη Κανονική Μορφή):

- Είναι σε 1NF και δεν έχουμε μερικές ΣΕ.
- Κάθε γνώρισμα που δεν είναι μέρος ενός υποψηφίου κλειδιού εξαρτάται συναρτησιακά από ολόκληρο το κλειδί και όχι από μέρος του. ΣΕ της μορφής: Πρωτεύον γνώρισμα -> Μη πρωτεύον παραβιάζουν την 2NF.

(Πρωτεύον γνώρισμα είναι αυτό που συμμετέχει σε υποψηφιο κλειδί αλλά δεν αποτελεί μόνο του υποψήφιο κλειδί)

1NF παράδειγμα

Δίνετε ο πίνακας

ID	Επώνυμο	Όνομα	Διεύθυνση	Hobby
AH 123456	Νικολάου	Κώστας	Αλαμάνας 20, 75400	Cart, Bowling
AH 234567	Παπαδόπουλος	Γιώργος	Ελ. Βενιζέλου 36Α, 13410	Sailing
AZ 345678	Νικολάου	Μιχάλης	Λ. Δημοκρατίας 12, 71409	BasketBall, Music
AK 123456	Κανάκη	Χαρούλα	Σουλίου 36, 35410	Trekking, Canyoning

Παρατηρούμε ότι δεν είναι σε 1NF.

Κάθε πλειάδα με πλειότιμα γνωρίσματα θα διασπαστεί σε περισσότερες (1 πλειάδα για κάθε διαφορετική τιμή του γνωρίσματος). Τα σύνθετα γνωρίσματα θα διασπαστούν σε περισσότερα, απλά γνωρίσματα.

ID	Επώνυμο	Όνομα	Δ_Οδός	Δ_Αριθμός	Δ_TK	Hobby
AH 123456	Νικολάου	Κώστας	Αλαμάνας	20	75400	Cart
AH 123456	Νικολάου	Κώστας	Αλαμάνας	20	75400	Bowling
AH 234567	Παπαδόπουλος	Γιώργος	Ελ. Βενιζέλου	36A	13410	Sailing
AZ 345678	Νικολάου	Μιχάλης	Λ. Δημοκρατίας	12	71409	BasketBall
AZ 345678	Νικολάου	Μιχάλης	Λ. Δημοκρατίας	12	71409	Music
AK 123456	Κανάκη	Χαρούλα	Σουλίου	36	35410	Trekking
AK 123456	Κανάκη	Χαρούλα	Σουλίου	36	35410	Canyoning

Ο παραγώμενος πίνακας είναι 1NF.

? Όμως έχουμε επανάληψη πληροφορίας...

2NF παράδειγμα

<u>Υ_ID</u>	<u>Ε_ID</u>	ΥΕπώνυμο	ΕΌνομα	ΕΤόπος	Ώρες
AH 123456	123	Νικολάου	Γέφυρα	Πάτρα	34
AH 234567	56	Παπαδόπουλος	Γήπεδο	Πειραιάς	28
AZ 345678	123	Νικολάου	Γέφυρα	Πάτρα	5
AK 123456	34	Κανάκη	Π. Κέντρο	Ιωάννινα	76
AH 123458	56	Κούκος	Γήπεδο	Πειραιάς	56

και τις πλήρεις συναρτησιακές εξαρτήσεις

$Υ_ID, Ε_ID \rightarrow \text{Ώρες}$, $Υ_ID \rightarrow \text{ΥΕπώνυμο}$, $Ε_ID \rightarrow \text{ΕΌνομα, ΕΤόπος}$

Παρατηρούμε ότι δεν αποτελεί 2NF καθώς οι σχέσεις $Υ_ID \rightarrow \text{ΥΕπώνυμο}$, $Ε_ID \rightarrow \text{ΕΌνομα, ΕΤόπος}$ είναι της μορφής

Πρωτεύον γνώρισμα (μη-κλειδί) \rightarrow Μη πρωτεύον

Οι τρεις πλήρεις συναρτησιακές εξαρτήσεις οδηγούν στη δημιουργία τριών πινάκων.

<u>Υ_ID</u>	ΥΕπώνυμο
AH 123456	Νικολάου
AH 234567	Παπαδόπουλος
AZ 345678	Νικολάου
AK 123456	Κανάκη
AH 123458	Κούκος

<u>Υ_ID</u>	<u>E_ID</u>	Ώρες
AH 123456	123	34
AH 234567	56	28
AZ 345678	123	5
AK 123456	34	76
AH 123458	56	56

<u>E_ID</u>	ΕΌνομα	ΕΤόπος
123	Γέφυρα	Πάτρα
56	Γήπεδο	Πειραιάς
34	Π. Κέντρο	Ιωάννινα

3NF (Τρίτη Κανονική Μορφή):

Είναι σε 2NF και δεν υπάρχουν μεταβατικές εξαρτήσεις.

- Δεν είναι μεταβατική όταν: Για μια $X \rightarrow Y$, δεν υπάρχει Z τ.ω. $X \rightarrow Z$ και $Z \rightarrow Y$.
- Δεν υπάρχουν ΣΕ της μορφής Μη-πρωτεύον \rightarrow Μη-πρωτεύον.

BCNF (Κανονική Μορφή Boyce-Codd):

- Είναι σε 3NF και όλες οι ΣΕ είναι ή τετριμμένες ή το αριστερό μέλος είναι υποψήφιο κλειδί.

π.χ. Έστω $R = \{A, B, C, D, E, F\}$ με κλειδί AB . Για να είναι BCNF επιτρέπονται μόνο οι εξαρτήσεις: $AB \rightarrow AB$, $AB \rightarrow A$, $AB \rightarrow B$, $ABC \rightarrow D$, $ABCD \rightarrow F$,
Απεναντίας δεν επιτρέπονται εξαρτήσεις, όπως: $A \rightarrow CD$, $DE \rightarrow F$, $BCD \rightarrow EF$,

3NF παράδειγμα

Έχουμε τον πίνακα

ΑριθμόςΠαραγγελίας	ΚωδικόςΠελάτη	ΠόληΠελάτη
124/2010	C142	Ηράκλειο
138/2010	C139	Θεσσαλονίκη
221/2010	C127	Πάτρα
253/2010	C139	Θεσσαλονίκη
391/2010	C142	Ηράκλειο

με τις συναρτησιακές εξαρτήσεις

ΑριθμόςΠαραγγελίας → ΚωδικόςΠελάτη, ΚωδικόςΠελάτη → Πόλη Πελάτη

και υπάρχει η μεταβατικότητα

ΑριθμόςΠαραγγελίας → ΠόληΠελάτη

Η μετατροπή σε 3NF οδηγεί στην διάσπαση

ΚωδικόςΠελάτη	ΠόληΠελάτη
C142	Ηράκλειο
C139	Θεσσαλονίκη
C127	Πάτρα

ΑριθμόςΠαραγγελίας	ΚωδικόςΠελάτη
124/2010	C142
138/2010	C139
221/2010	C127
253/2010	C139
391/2010	C142

BCNF παράδειγμα 1

1) Έχουμε το σχήμα σχέσης που περιγράφεται από τον πίνακα

IDΈργου	<u>IDΣυμβουλος</u>	<u>IDΜηχανικός</u>	Αντικείμενο
E12	Σ12	M13	A111
E12	Σ34	M13	A111
E12	Σ65	M45	A111
E28	Σ28	M67	A678
E28	Σ71	M67	A678
E67	Σ43	M35	A904

με κλειδί: {IDΣύμβουλος, IDΜηχανικός} και συναρτησιακές εξαρτήσεις:

{IDΣύμβουλος, IDΜηχανικός} → IDΈργου, IDΈργου → Αντικείμενο

Παρατηρούμε ότι δεν είναι BCNF, μάλιστα δεν είναι καν 3NF καθώς η εξάρτηση

ID Έργου \rightarrow Αντικείμενο

παραβιάζει τον ορισμό. Το διασπάμε σε δύο σχέσεις
{ID Σύμβουλος, ID Μηχανικός, ID Έργου} και
{ID Έργου, Αντικείμενο}

ID Έργου	ID Σύμβουλος	ID Μηχανικός
E12	Σ12	M13
E12	Σ34	M13
E12	Σ65	M45
E28	Σ28	M67
E28	Σ71	M67
E67	Σ43	M35

ID Έργου	Αντικείμενο
E12	A111
E28	A678
E67	A904

BCNF παράδειγμα 2

Έχουμε το σχήμα σχέσης που περιγράφεται από τον πίνακα

Πελάτης	Συναλλαγή	ΠοσόΣ	ΥποΚατ	ΟνομαΥΚ	ΑποθεμαΥΚ
Π12	Σ1312	4060	ΥΚ13	Καβάλα	450
Π14	Σ1434	230	ΥΚ14	Καλαμάτα	720
Π78	Σ4565	4600	ΥΚ45	Ξάνθη	316
Ε26	Σ6728	2670	ΥΚ67	Ρέθυμνο	412
Ε67	Σ1371	13784	ΥΚ13	Καβάλα	450

με κλειδί: {Πελάτης, Συναλλαγή} και συναρτησιακές εξαρτήσεις

$\text{ΥποΚατ} \rightarrow \{\text{ΟνομαΥΚ}, \text{ΑπόθεμαΥΚ}\}$, $\text{Συναλλαγή} \rightarrow \{\text{ΥποΚατ}, \text{ΠοσόΣ}\}$

Παρατηρούμε ότι δεν είναι BCNF καθώς και οι δύο εξαρτήσεις παραβιάζουν τον BCNF ορισμό. Μετά από διάσπαση η εξάρτηση $\text{ΥποΚατ} \rightarrow \{\text{ΟνομαΥΚ}, \text{ΑπόθεμαΥΚ}\}$ οδηγεί στην διάσπαση του R σε R_1 και R_2 ,

R_1

ΥποΚατ	ΟνομαΥΚ	ΑποθεμαΥΚ
ΥΚ13	Καβάλα	450
ΥΚ14	Καλαμάτα	720
ΥΚ45	Ξάνθη	316
ΥΚ67	Ρέθυμνο	412

R_2

Πελάτης	Συναλλαγή	ΠοσόΣ	ΥποΚατ
Π12	Σ1312	4060	ΥΚ13
Π14	Σ1434	230	ΥΚ14
Π78	Σ4565	4600	ΥΚ45
Ε26	Σ6728	2670	ΥΚ67
Ε67	Σ1371	13784	ΥΚ13

Όμως η εξάρτηση Συναλλαγή \rightarrow {ΥποΚατ, ΠοσόΣ} παραβιάζει τον BCNF ορισμό στην R_2 . Επομένως διασπάμε το R_2 , την εξάρτηση Συναλλαγή \rightarrow {ΥποΚατ, ΠοσόΣ} σε R_3 και R_4 και καταλήγουμε στις

R_1

ΥποΚατ	ΌνομαΥΚ	ΑποθεμαΥΚ
ΥΚ13	Καβάλα	450
ΥΚ14	Καλαμάτα	720
ΥΚ45	Εάνθη	316
ΥΚ67	Ρέθυμνο	412

R_3

Συναλλαγή	ΠοσόΣ	ΥποΚατ
Σ1312	4060	ΥΚ13
Σ1434	230	ΥΚ14
Σ4565	4600	ΥΚ45
Σ6728	2670	ΥΚ67

R_4

Πελάτης	Συναλλαγή
Π12	Σ1312
Π14	Σ1434
Π78	Σ4565
Ε28	Σ6728
Ε67	Σ1371

Sum up

Μια σχέση σε 3NF διασπάται σε ένα σύνολο από σχέσεις της BCNF όταν αφαιρεθούν από τον πίνακα όσα χαρακτηριστικά δεν εξαρτώνται πλήρως από κλειδιά.

Απλά

1NF: Όχι σύνθετα και πλειότιμα γνωρίσματα.

2NF: Μόνο πλήρεις συναρτησιακές εξαρτήσεις.

3NF: Όχι μεταβατικές συναρτησιακές εξαρτήσεις.

BCNF: Μόνο περιορισμοί κλειδιού.

Παρατηρήσεις

- ▶ Οδηγούμενοι σε πιο περιοριστική κανονική μορφή, αναγκαζόμαστε να διασπάσουμε την αρχική σχέση σε περισσότερες.
- ▶ Οι περισσότερες 3NF είναι και BCNF.
- ▶ Οι 1NF και 2NF δεν είναι ικανοποιητικές για να χαρακτηρίσουμε την βάση ως καλά σχεδιασμένα.
- ▶ Οι 3NF και BCNF εξαλείφουν προβλήματα που σχετίζονται με ενημερώσεις.