



# Βάσεις Δεδομένων I

## Ενότητα 12: Κανονικοποίηση

Δρ. Τιμπίρης Αλκιβιάδης

Τμήμα Μηχανικών Πληροφορικής ΤΕ



# Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



# Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο ΤΕΙ Κεντρικής Μακεδονίας» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



# Ενότητα 12

---

## Κανονικοποίηση

Δρ. Τιμπίρης Αλκιβιάδης

# Περιεχόμενα ενότητας

---

- Θεωρία Κανονικοποίησης
- Κανονικές Μορφές (normal forms)
- Πρώτη Κανονική Μορφή 1KM
- Δεύτερη Κανονική Μορφή 2KM
- Τρίτη Κανονική Μορφή 3KM

# Σκοποί ενότητας

Στην ενότητα αυτή παρουσιάζεται ένας ακόμη τρόπος σχεδιασμού βάσεων δεδομένων ο οποίος βασίζεται στην Θεωρία Κανονικοποίησης σύμφωνα με την οποία δημιουργείται ένα μεγάλος πίνακας που περιέχει όλα τα γνωρίσματα που απαιτούνται και στη συνέχεια απόδομείται. Αυτό επιτυγχάνεται ορθά με την εφαρμογή των Κανονικών Μορφών (normal forms). Η Πρώτη Κανονική Μορφή (1KM) είναι αυτή στην οποία πρέπει να βρίσκονται όλοι οι πίνακες μιας ΒΔ. Επίσης η Δεύτερη Κανονική Μορφή (2KM) και η Τρίτη Κανονική Μορφή (3KM) αντιμετωπίζουν το μεγαλύτερο μέρος των πιθανών προβλημάτων κατά τον σχεδιασμό μιας ΒΔ.

# Θεωρία Κανονικοποίησης

- Μία σχέση είναι ένα σύνολο από γνωρίσματα με τιμές για κάθε γνώρισμα τέτοιες ώστε να ισχύουν οι παρακάτω ιδιότητες
  1. Κάθε όνομα γνωρίσματος είναι μοναδικό.
  2. Όλες οι τιμές κάθε γνωρίσματος είναι ίδιου τύπου (ή πεδίου ορισμού).
  3. Κάθε τιμή γνωρίσματος είναι ατομική (μία τιμή και όχι ομάδα πολλών τιμών).
  4. Τα γνωρίσματα δεν έχουν διάταξη από τα αριστερά προς τα δεξιά..
  5. Οι συστοιχίες (σειρές) δεν έχουν διάταξη από επάνω προς τα κάτω.
  6. Δεν υπάρχουν δύο ίδιες σειρές (συστοιχίες) σε μία σχέση.

# Διαδικασία σωστού σχεδιασμού ΒΔ

Η διαδικασία που ακολουθούμε είναι :

- Συγκεντρώνουμε τις απαιτήσεις της επιχείρησης και των χρηστών.
- Σχεδιάζουμε το μοντέλο οντοτήτων-συσχετίσεων
- Μετατρέπουμε το διάγραμμα οντοτήτων-συσχετίσεων της επιχείρησης σε ένα σύνολο από σχέσεις (πίνακες) με το σχεσιακό μοντέλο.
- Κανονικοποιούμε τις σχέσεις για να απομακρύνουμε τυχόν ανωμαλίες ενημέρωσης-διαγραφής-εισαγωγής στοιχείων.
- Υλοποιούμε τη βάση δεδομένων δημιουργώντας ένα πίνακα για κάθε κανονικοποιημένη σχέση.



# Κανονικοποίηση-Κανονικές μορφές

- Κατηγορίες (ή κλάσεις) **Κανονικές Μορφές** (normal forms).
  - **Κανονική Μορφή**: Μία κλάση σχέσεων απαλλαγμένων από συγκεκριμένα προβλήματα τροποποιήσεων.
1. Πρώτη κανονική μορφή (1NF – 1KM)
  2. Δεύτερη κανονική μορφή (2NF – 2KM)
  3. Τρίτη κανονική μορφή (3NF – 3KM)
  4. Boyce-Codd κανονική μορφή (BCNF – KM BC)
  5. Τέταρτη κανονική μορφή (4NF – 4KM)
  6. Πέμπτη κανονική μορφή (5NF – 5KM)
  7. Κανονική μορφή πεδίου ορισμού κλειδιού (Domain-Key/NF)

# Κανονικοποίηση-Κανονικές μορφές

- Αυτές οι κανονικές μορφές είναι αθροιστικές. Μία σχέση που βρίσκεται σε Τρίτη κανονική μορφή είναι επίσης και σε δεύτερη και σε πρώτη.
- Οι τρεις πρώτες κανονικές μορφές (1KM, 2KM, 3KM) ορίστηκαν από τον Codd. Όλες οι κανονικοποιημένες σχέσεις είναι σε 1KM.
- Με άλλα λόγια, "κανονικοποιημένη" και "σε 1KM" σημαίνει ακριβώς το ίδιο πράγμα.
- Μερικές σχέσεις 1KM είναι επίσης σε 2KM, και μερικές σχέσεις 2KM είναι επίσης σε 3KM.

# Κανονικοποίηση-Κανονικές μορφές

- Ο Fagin όρισε την τέταρτη κανονική μορφή.
- Μετέπειτα και πάλι ο Fagin όρισε άλλη μία κανονική μορφή, την κανονική μορφή προβολής–σύζευξης (projection join), που αργότερα έγινε γνωστή και ως πέμπτη κανονική μορφή (5KM). ‘
- Μερικές σχέσεις που είναι σε KM-BC είναι επίσης σε 4KM, και μερικές σχέσεις που είναι σε 4KM είναι επίσης σε 5KM.

# Πρώτη Κανονική Μορφή 1KM

- Μία σχέση βρίσκεται σε πρώτη κανονική μορφή αν ικανοποιεί όλες τις 6 ιδιότητες του ορισμού της σχέσης
- Εάν υπάρχει καθορισμένο **κλειδί** για τη σχέση τότε ικανοποιείται η απαίτηση της μοναδικότητας των συστοιχιών (σειρών).
- Ένας πίνακας σε πρώτη κανονική μορφή λέγεται κανονικοποιημένος πίνακας και τότε και μόνο τότε αντιστοιχεί σε μία σχέση (οι σχέσεις του σχεσιακού μοντέλου είναι στην 1η κανονική μορφή ).

# Δεύτερη κανονική μορφή (2NF)

- Μία σχέση βρίσκεται σε δεύτερη κανονική μορφή εάν κάθε ένα από τα γνωρίσματά της που δεν είναι κλειδιά εξαρτώνται συναρτησιακά από ολόκληρο το πρωτεύων κλειδί και όχι μόνο από ένα τμήμα του.
- Οι σχέσεις που έχουν μόνο **ένα γνώρισμα σαν πρωτεύων κλειδί** βρίσκονται αυτόματα και στη **δεύτερη κανονική μορφή**.
- Αυτός είναι ένας λόγος για τον οποίο χρησιμοποιούμε συχνά τεχνητά αναγνωριστικά σαν κλειδιά.

# Τρίτη κανονική μορφή (3NF)

- Μία σχέση βρίσκεται σε Τρίτη κανονική μορφή εάν είναι σε δεύτερη και δεν περιέχει **μεταβατικές εξαρτήσεις**.
- Θεωρήστε για παράδειγμα τη σχέση R που έχει γνωρίσματα τα A, B και Γ. Εάν  $A \rightarrow B$  και  $B \rightarrow \Gamma$  τότε θα ισχύει και  $A \rightarrow \Gamma$ .

# Κανονικές μορφές

Για τις κανονικές μορφές ισχύουν οι παρακάτω κανόνες:

- Δεν πρέπει να υπάρχουν επαναλαμβανόμενα ονόματα πεδίων.
- Όσα πεδία δεν είναι κλειδιά θα πρέπει να εξαρτώνται συναρτησιακά από ολόκληρο το πρωτεύον κλειδί και όχι από μέρος του.
- Δεν πρέπει να υπάρχει συναρτησιακή μεταβατική εξάρτηση.



# Κανονικές μορφές

1KM

ΑΕΜ	ΕΠΙΘΕΤΟ	<del>ΤΗΛ</del>	<del>ΤΗΛ</del>



Τα ονόματα των πεδίων δεν είναι μοναδικά.

2KM

ΚΜ	ΑΕΜ	<del>ΟΝΟΜΑ</del>	<del>ΤΗΛ</del>



Το ΟΝΟΜΑ και το ΤΗΛ εξαρτώνται από το ΑΕΜ.

3KM

ΑΕΜ	ΕΠΙΘΕΤΟ	<del>ΟΝΟΜΑ ΠΑΤΡΟΣ</del>	<del>ΤΗΛ ΠΑΤΡΟΣ</del>



Μεταβατική Εξάρτηση



# Από-κανονικοποίηση

1KM

ΑΕΜ	ΕΠΙΘΕΤΟ	ΤΗΛ1	ΤΗΛ2



Δεν τηρεί τους κανόνες κανονικοποίησης.

Υπάρχουν περιπτώσεις που εάν δεν ακολουθήσουμε τους κανόνες κανονικοποίησης επιτυγχάνουμε καλύτερα αποτελέσματα.

Αυτό καλείται Από-Κανονικοποίηση (De-normalization)