

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΛΥΣΗ ΣΤΗΝ ΔΕΥΤΕΡΗ ΑΣΚΗΣΗ

ΜΑΘΗΜΑ
ΑΚΑΔ. ΕΤΟΣ
ΔΙΔΑΣΚΩΝ

ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ
2017-18
Βηρένα Καντερέ, Καθηγήτρια, Τομέας Πληροφορικής

Ερώτημα 1.

Δίνεται η σχέση $R(A, B, C, D, E, G)$ με τις παρακάτω συναρτησιακές εξαρτήσεις

$AB \rightarrow C, D \rightarrow EG, C \rightarrow A, BE \rightarrow C, BC \rightarrow D, CG \rightarrow BD, ACD \rightarrow B, CE \rightarrow AG$

Ποιες από τις παρακάτω προτάσεις είναι σωστές και ποιες λάθος? Δικαιολογήστε την απάντησή σας.

1. Η κλειστότητα του BC είναι $\{A, D, E, G\}$
2. Όλα τα γνωρίσματα της σχέσης R συμπεριλαμβάνονται στο BC^+ .
3. Η κλειστότητα του AC είναι $\{A, C\}$
4. Το ABC είναι υπερκλειδί της σχέσης R
5. Το ABC είναι υποψήφιο κλειδί της σχέσης R .
6. Το BC είναι το μόνο υποψήφιο κλειδί της σχέσης R .
7. Το CG είναι υποψήφιο κλειδί της σχέσης R .
8. Το CD είναι υποψήφιο κλειδί της σχέσης R .
9. Το AG είναι υποψήφιο κλειδί της σχέσης R
10. Το ADE είναι υποψήφιο κλειδί της σχέσης R

ΛΥΣΗ

1. Λάθος
2. Σωστό
3. Σωστό
4. Σωστό
5. Λάθος
6. Λάθος
7. Σωστό
8. Σωστό
9. Λάθος
10. Λάθος

Ερώτημα 2.

Θεωρείστε τη σχέση που φαίνεται στο ακόλουθο σχήμα.

A	B	C	D
a_1	b_2	c_3	d_1

a_1	b_2	c_3	d_3
a_2	b_3	c_2	d_2
a_3	b_4	c_3	d_1

- (a) Καταγράψτε όλες τις συναρτησιακές εξαρτήσεις που ικανοποιεί το συγκεκριμένο στιγμιότυπο της σχέσης.
 (b) Έστω ότι η τιμή του γνωρίσματος C στην τελευταία εγγραφή αλλάζει από c_3 σε c_2 . Καταγράψτε και πάλι όλες τις συναρτησιακές εξαρτήσεις που ικανοποιεί το συγκεκριμένο στιγμιότυπο της σχέσης.

ΛΥΣΗ

- A) $A \rightarrow B$, $A \rightarrow C$, $B \rightarrow C$, **$D \rightarrow C$** , $B \rightarrow A$, $AB \rightarrow C$, $BC \rightarrow A$, $AC \rightarrow B$, $AD \rightarrow B$, $AD \rightarrow C$
 B) $A \rightarrow C$, $B \rightarrow C$, $AB \rightarrow C$, $A \rightarrow B$, $B \rightarrow A$, $AC \rightarrow B$, $BC \rightarrow A$, $AD \rightarrow B$, $AD \rightarrow C$

Ερώτημα 3.

Θεωρείστε μία σχέση R με τέσσερα γνωρίσματα $ABCD$. Για κάθε ένα από τα ακόλουθα σύνολα συναρτησιακών εξαρτήσεων, θεωρώντας ότι αυτές είναι οι μόνες εξαρτήσεις που ισχύουν για την R , ζητούνται τα ακόλουθα: (a) Υπολογίστε τα υποψήφια κλειδιά της R . (b) Ορίστε την καλύτερη κανονική μορφή που επαληθεύει η R (1NF, 2NF, 3NF, ή BCNF). (c) Αν η R δεν είναι σε BCNF, επιχειρήστε μια αποσύνθεση σε ένα σύνολο από BCNF σχέσεις που διατηρούν τις εξαρτήσεις.

- $A \rightarrow D$, $A \rightarrow B$, $C \rightarrow A$.
- $C \rightarrow A$, $B \rightarrow D$.

ΛΥΣΗ

- (a) Υποψήφιο κλειδί είναι το C , διότι το BC^+ περιέχει όλα τα γνωρίσματα της R και δεν υπάρχει υποσύνολο του C που να είναι και αυτό κλειδί.

Για τον υπολογισμό του C^+ ξεκινάμε με $C^+ = C$ και σε κάθε βήμα λαμβάνουμε υπόψη της συναρτησιακές εξαρτήσεις στη μορφή $\beta \rightarrow \gamma$, όπου $\beta \subseteq C^+$. Οπότε έχουμε:

$C^+ = CA$, λόγω της $C \rightarrow A$,

$C^+ = ABCD$, λόγω των $A \rightarrow D$ και $A \rightarrow B$.

(b) Η R είναι σε 2NF, αλλά όχι σε 3NF διότι το C είναι υποψήφιο κλειδί και έχουμε $C \rightarrow A$ και $A \rightarrow D$, δηλαδή το D είναι μεταβατικά εξαρτώμενο από το C , και $C \rightarrow A$ και $A \rightarrow B$, δηλαδή το B είναι μεταβατικά εξαρτώμενο από το C .

(c) Για την $R(A, B, C, D)$ οι συναρτησιακές εξαρτήσεις $A \rightarrow D$ και $A \rightarrow B$ παραβιάζουν την κανονική μορφή BCNF.

Έστω επιλέγουμε πρώτα την $A \rightarrow D$, άρα αποσυνθέτουμε σε: $R_1(A, D)$, $R'(A, B, C)$.

Η $R'(A, B, C)$ δεν είναι σε BCNF λόγω της $A \rightarrow B$, άρα αποσυνθέτουμε σε: $R_1(A, D)$, $R_2(A, B)$, $R_3(A, C)$.

Οι σχέσεις αυτές είναι σε BCNF και η αποσύνθεση είναι χωρίς απώλειες συνδέσμων. Επίσης, η εν λόγω αποσύνθεση διατηρεί τις εξαρτήσεις, διότι κάθε μέλος του αρχικού συνόλου συναρτησιακών εξαρτήσεων μπορεί να ελεγχθεί σε μία από τις σχέσεις της αποσύνθεσης. Συγκεκριμένα, η $A \rightarrow D$ στην R_1 , η $A \rightarrow B$ στην R_2 και η $C \rightarrow A$ στην R_3 .

Να σημειωθεί ότι αν επιλέγαμε πρώτα την $A \rightarrow B$ για την αποσύνθεση θα καταλήγαμε στο ίδιο αποτέλεσμα.

- (a) Υποψήφιο κλειδί είναι το BC , καθώς το BC περιέχει όλα τα γνωρίσματα της R , σύμφωνα με τον παρακάτω υπολογισμό, και δεν υπάρχει υποσύνολο του BC που να είναι και αυτό κλειδί.

$BC^+ = BC$,

$BC^+ = ABCD$, λόγω της $B \rightarrow D$ και της $C \rightarrow A$.

(b) Η R είναι σε 1NF, αλλά όχι σε 2NF διότι το A είναι μη πρωτεύον γνώρισμα και είναι μερικώς εξαρτώμενο από το BC που αποτελεί υποψήφιο κλειδί, δηλαδή ισχύει $BC \rightarrow A$ και $C \rightarrow A$ όπου $C \subset BC$.

(c) Για την $R(A, B, C, D)$ και οι δύο συναρτησιακές εξαρτήσεις $B \rightarrow D$ και $C \rightarrow A$ παραβιάζουν την κανονική μορφή BCNF.

Έστω ότι επιλέγουμε πρώτα την $B \rightarrow D$, άρα αποσυνθέτουμε σε $R_1(B, D)$, $R'(A, B, C)$.

Η $R'(A, B, C)$ δεν είναι σε BCNF λόγω της $C \rightarrow A$, άρα αποσυνθέτουμε σε: $R_1(B, D)$, $R_2(A, C)$, $R_3(B, C)$.

Οι σχέσεις αυτές είναι σε BCNF και η αποσύνθεση είναι χωρίς απώλειες συνδέσμων. Επίσης, η εν λόγω αποσύνθεση διατηρεί τις εξαρτήσεις, διότι κάθε μέλος του αρχικού συνόλου συναρτησιακών εξαρτήσεων μπορεί να ελεγχθεί σε μία από τις σχέσεις της αποσύνθεσης. Συγκεκριμένα, η $B \rightarrow D$ στην R_1 και η $C \rightarrow A$ στην R_2 .

Να σημειωθεί ότι αν επιλέγαμε πρώτα την $C \rightarrow A$ για την αποσύνθεση θα καταλήγαμε στο ίδιο αποτέλεσμα.

Ερώτημα 4.

A. Θεωρείστε μία σχέση R που έχει 11.000 εγγραφές (πλειάδες) και είναι αποθηκευμένη σε ένα αρχείο σωρού. Θεωρείστε ένα πυκνό B+-tree ευρετήριο πάνω στο αρχείο σωρού. Το πεδίο ευρετηριοποίησης για το εν λόγω ευρετήριο είναι ένα αλφαριθμητικό 45 bytes και είναι υποψήφιο κλειδί. Για κάθε δείκτη απαιτούνται 12 bytes. Το μέγεθος μιας σελίδας δίσκου είναι 1024 bytes. Το ευρετήριο έχει δημιουργηθεί με bottom-up λογική χρησιμοποιώντας bulk-loading. Οι κόμβοι σε κάθε επίπεδο είναι όσο το δυνατόν γεμάτοι.

1. Πόσα επίπεδα έχει το παραπάνω B+-tree ευρετήριο?
2. Πόσοι κόμβοι υπάρχουν σε κάθε επίπεδο του δέντρου?

B. Έστω τώρα ένα πεδίο A της σχέσης R που δεν είναι κλειδί και μπορεί να πάρει 100 διαφορετικές τιμές. Το μέγεθος του πεδίου A είναι 8 bytes. Θεωρείστε ότι χρησιμοποιείται ένα επιπλέον επίπεδο έμμεσων δεικτών, εφόσον το κλειδί αναζήτησης (A) δεν είναι υποψήφιο κλειδί.

(i) Δώστε το συνολικό μέγεθος του αρχείου ευρετηρίου (συμπεριλαμβανομένου και του επιπλέον επιπέδου) σε αριθμό blocks.

(ii) Υπολογίστε το κόστος σε αριθμό I/O (δηλαδή, σε αριθμό προσπελάσεων σελίδων) της αναζήτησης $A = a$, όπου a μία τιμή από το πεδίο ορισμού του A .

ΛΥΣΗ

1. Κάθε κόμβος του ευρετηρίου καταλαμβάνει ένα block-σελίδα στον δίσκο. Έστω ότι κάθε κόμβος του ευρετηρίου μπορεί να αποθηκεύσει n κλειδιά. Θα πρέπει να ισχύει:

$$(n + 1) * 12 + n * 45 \leq 1024 \text{ άρα } n = 17$$

Εφόσον το B+-tree είναι όσον το δυνατόν πιο γεμάτο κάθε κόμβος του δέντρου μπορεί να αποθηκεύσει μέχρι 17 κλειδιά και 28 δείκτες.

Εφόσον το B+-tree είναι πυκνό ευρετήριο στο επίπεδο των φύλλων περιέχει ένα κλειδί και έναν δείκτη για κάθε μία από τις τιμές του πεδίου ευρετηριοποίησης της σχέσης. Άρα, το επίπεδο των φύλλων χρειάζεται $\lceil 11000/17 \rceil = 648$ block.

Το αμέσως επόμενο επίπεδο υπάρχουν $648/18 = 36$ block, το επόμενο $\lceil 36/18 \rceil = 2$ block και το τελευταίο 1, δηλαδή η ρίζα του δέντρου.

Συνεπώς, το B+-tree έχει 4 επίπεδα.

2. Από τον υπολογισμό στο προηγούμενο ερώτημα στα φύλλα υπάρχουν 648 κόμβοι, στο αμέσως επόμενο επίπεδο 36 κόμβοι, στο επόμενο 2 κόμβοι και η ρίζα 1 κόμβος.

- i. $8 + 12 = 20$ bytes για κάθε εγγραφή του ευρετηρίου

$$\lfloor 1024/51 \rfloor = 51 \text{ εγγραφές/block}$$

$$\lfloor 1000/51 \rfloor = 2 \text{ block δίσκου}$$

$$\lfloor 11000/100 \rfloor = 110 \text{ εγγραφές ανά τιμή του A}$$

$$\lfloor 1024/12 \rfloor = 86 \text{ δείκτες/block}$$

$$\lfloor 110/86 \rfloor = 2 \text{ block ανά τιμή του A}$$

Σύνολο $2 \cdot 100 + 2$ block δίσκου για το ευρετήριο.

ii. Αρχικά θα αναζητηθεί η εγγραφή με τιμή κλειδιού αναζήτησης $A = a$ στο ευρετήριο. Με δυαδική αναζήτηση (εφόσον το αρχείο είναι ταξινομημένο) το κόστος είναι $\lceil \log_2 2 \rceil = 1$ block. +2 block που περιέχουν τους δείκτες προς τις εγγραφές με $A = a + 110$ εγγραφές ανά τιμή του A. max 113 I/O