

Συστήματα Διαχείρισης και Ανάλυσης Δεδομένων Διδάσκων: Ιωάννης Κωτίδης

Εαρινό εξάμηνο 2025-2026

Πρώτη Σειρά Ασκήσεων

Ανάθεση: 20-03-2026

Παράδοση: 31-03-2026 Ώρα (23:55)

Οδηγίες

- Η πρώτη σειρά ασκήσεων είναι **ατομική**.
- Η υποβολή της εργασίας πρέπει να γίνει στο *eclass*.
- Το παραδοτέο σας θα πρέπει να είναι ένα αρχείο PDF με όνομα *AM.pdf* (όπου *AM* είναι ο αριθμός μητρώου σας. π.χ. "3230001.pdf").
- Τα διαγράμματα πρέπει να είναι κατασκευασμένα σε κάποιο πρόγραμμα (της επιλογής σας) και όχι σκαναρισμένα χειρόγραφα.
- Πιθανή αντιγραφή θα τιμωρείται με μηδενισμό όλων των εμπλεκομένων.
- **Για την επίλυση των ασκήσεων να μελετήσετε τις διαφάνειες των διαλέξεων του μαθήματος.**

Άσκηση 1 [40 μονάδες]

Ένας σκληρός δίσκος έχει τα παρακάτω χαρακτηριστικά:

- 10 επιφάνειες
- 800 κυλίνδρους
- 256 τομείς ανά ίχνος (sectors/track)
- Μέγεθος τομέα 512 bytes
- Μέσος Χρόνος Μετακίνησης Κεφαλής (Average Seek Time) = 12 ms
- Ταχύτητα περιστροφής 10000 rpm
- Μέγεθος μπλοκ (block) 4096 bytes.

Στον δίσκο είναι αποθηκευμένος με τη μορφή αρχείου ο πίνακας **TRANSACTIONS**, ο οποίος περιέχει 3.000.000 εγγραφές τραπεζικών συναλλαγών. Τα πεδία του πίνακα καθώς και το μέγεθός τους σε bytes έχουν ως εξής:

TRANSACTIONS (**tid**: 8 bytes, **accno**: 12 bytes, **tdate**: 8 bytes, **amount**: 8 bytes, **type**: 4 bytes)

Να θεωρήσετε ότι:

- Το αρχείο δεδομένων καταλαμβάνει συνεχόμενα μπλοκ στον δίσκο.
- Καμία εγγραφή δεν εκτείνεται σε περισσότερα από ένα μπλοκ.
- Στο πεδίο **tdate** υπάρχει ένα δευτερεύον (non-clustered) ευρετήριο B+ δέντρου (έστω **I1**) το οποίο βρίσκεται στην κύρια μνήμη.
- Κατά την πλήρη σειριακή ανάγνωση του αρχείου δεδομένων **ο χρόνος μετακίνησης στο επόμενο ίχνος** είναι μηδαμινός (δηλαδή δεν λαμβάνεται υπόψη).

Έστω ότι εκτελείται το παρακάτω επερωτήμα το οποίο επιστρέφει **N** εγγραφές:

Q1: SELECT * FROM TRANSACTIONS WHERE tdate BETWEEN d1 AND d2

Όπου d1 και d2 είναι συγκεκριμένες ημερομηνίες.

Ένας φοιτητής ισχυρίζεται ότι η χρήση του ευρετηρίου **I1** οδηγεί σε μικρότερο **χρόνο εκτέλεσης** του επερωτήματος **Q1** από την πλήρη σειριακή ανάγνωση του αρχείου.

Να εξετάσετε τον παραπάνω ισχυρισμό και να επιλέξετε ποια από τις παρακάτω προτάσεις ισχύει:

- A. Ο παραπάνω ισχυρισμός ισχύει πάντα.
- B. Ο παραπάνω ισχυρισμός ισχύει μόνο για ορισμένες τιμές του N.

Να τεκμηριώσετε την απάντησή σας **υπολογίζοντας το χρόνο εκτέλεσης** του ερωτήματος με χρήση του ευρετηρίου και με πλήρη σειριακή ανάγνωση του αρχείου.

Άσκηση 2 [25 μονάδες]

Δίνεται η σχέση R(**a**,b,c,d) η οποία περιέχει 1800 εγγραφές και είναι αποθηκευμένη σε αρχείο στο δίσκο. Το γνώρισμα **a** είναι το πρωτεύον κλειδί της σχέσης και το αρχείο με τα δεδομένα της σχέσης R είναι διατεταγμένο κατά το γνώρισμα **a**.

Θεωρήστε ότι α) σε ένα μπλοκ του δίσκου χωράνε 6 εγγραφές της σχέσης R, ή 9 κλειδιά (keys) και 10 δείκτες (pointers) ενός B+ δέντρου και β) ένας δείκτης προς κόμβο του δέντρου έχει το ίδιο μέγεθος με έναν δείκτη προς εγγραφή δεδομένων.

a. Ζητείται να υπολογίσετε:

- Τον ελάχιστο αριθμό των μπλοκ που απαιτούνται για την αποθήκευση της σχέσης R.
- Τον ελάχιστο αριθμό των μπλοκ που απαιτούνται για ένα αραιό (sparse) ευρετήριο B+ δέντρου στο γνώρισμα **a**.
- Τον ελάχιστο αριθμό των μπλοκ που απαιτούνται για ένα πυκνό (dense) ευρετήριο B+ δέντρου στο γνώρισμα **a**.

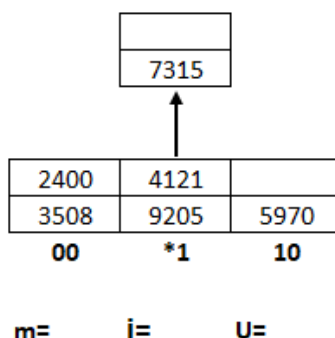
b. Για την εκτέλεση του παρακάτω επερωτήματος ποιο από τα δύο ευρετήρια (αραιό ή πυκνό) είναι αποδοτικότερο; Να θεωρήσετε ότι τα ευρετήρια είναι αποθηκευμένα στο δίσκο και δεν βρίσκονται αρχικά στην κύρια μνήμη.

SELECT * FROM R WHERE a=v (όπου v είναι μία τιμή του γνωρίσματος a)

Να αιτιολογήσετε την απάντησή σας.

Άσκηση 3 [35 μονάδες]

Έστω ένα αρχείο ευρετηρίου που χρησιμοποιεί την μέθοδο του γραμμικού κατακερματισμού με κάδους χωρητικότητας δύο εγγραφών. Για την κατανομή των τιμών χρησιμοποιούνται τα i λιγότερο σημαντικά bits ενώ η συνάρτηση κατακερματισμού είναι $h(k) = k \bmod 8$ (όπου k είναι το κλειδί). Ο αριθμός των κάδων πρέπει να αυξάνεται όταν το utilization του ευρετηρίου γίνει **μεγαλύτερο του 80%**. Το i αυξάνεται μόνο όταν κρίνεται απαραίτητο. Επίσης, δεν υπάρχει όριο στον αριθμό σελίδων υπερχειλίσης. Κάθε σελίδα υπερχειλίσης χωράει και αυτή δύο εγγραφές. Στο ευρετήριο έχουν ήδη εισαχθεί έξι κλειδιά:



Ζητείται:

- a. Να προσδιορίσετε την τιμή των i , m και U (Utilization)
- b. Να εισαγάγετε τα παρακάτω κλειδιά (k) με τη σειρά που δίνονται από αριστερά προς τα δεξιά. Να εμφανίσετε την μορφή του ευρετηρίου μετά από κάθε εισαγωγή κλειδιού δείχνοντας και τα ενδιάμεσα βήματα που απαιτούνται. Σε κάθε βήμα να εμφανίσετε την τιμή των i , m και U . Κάθε πράξη εισαγωγής πρέπει να εκτελείται στο αποτέλεσμα της προηγούμενης και όχι στο αρχικό ευρετήριο.
[5983, 6846, 9243, 6663, 8315, 8114]
- c. Να υπολογίσετε τον μέσο αριθμό προσπελάσεων για την ανάκτηση μιας εγγραφής όταν δίνεται η τιμή του κλειδιού. Να θεωρήσετε ότι:
 - Οι τιμές του κλειδιού είναι μοναδικές και το κλειδί προς αναζήτηση υπάρχει στο ευρετήριο.
 - Η αναζήτηση γίνεται στην τελική μορφή του ευρετηρίου όπως αυτή έχει προκύψει μετά την εισαγωγή των κλειδιών του ζητήματος b.
- d. Να υπολογίσετε τον μέσο αριθμό προσπελάσεων για την ανάκτηση μιας εγγραφής όταν η τιμή του κλειδιού δεν υπάρχει στο ευρετήριο. Η αναζήτηση γίνεται στην τελική μορφή του ευρετηρίου όπως αυτή έχει προκύψει μετά την εισαγωγή των κλειδιών του ζητήματος b.

Για τα ζητούμενα c και d να δείξετε τον τρόπο υπολογισμού και όχι μόνο το τελικό αποτέλεσμα.