

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΑΣΚΗΣΗ ΔΕΥΤΕΡΗ

ΜΑΘΗΜΑ ΒΑΣΕΙΣ ΔΕΛΟΜΕΝΩΝ
ΑΚΑΔ. ΕΤΟΣ 2011-12

ΕΠΙΣΤΡΕΦΕΤΑΙ

1.2.2012

ΔΙΔΑΣΚΟΝΤΕΣ Ιωάννης Βασιλείου Καθηγητής
 Τιμολέων Σελλής Καθηγητής

Η άσκηση αυτή είναι **ατομική**, δε θα γίνουν δεκτές ασκήσεις από ομάδες φοιτητών.

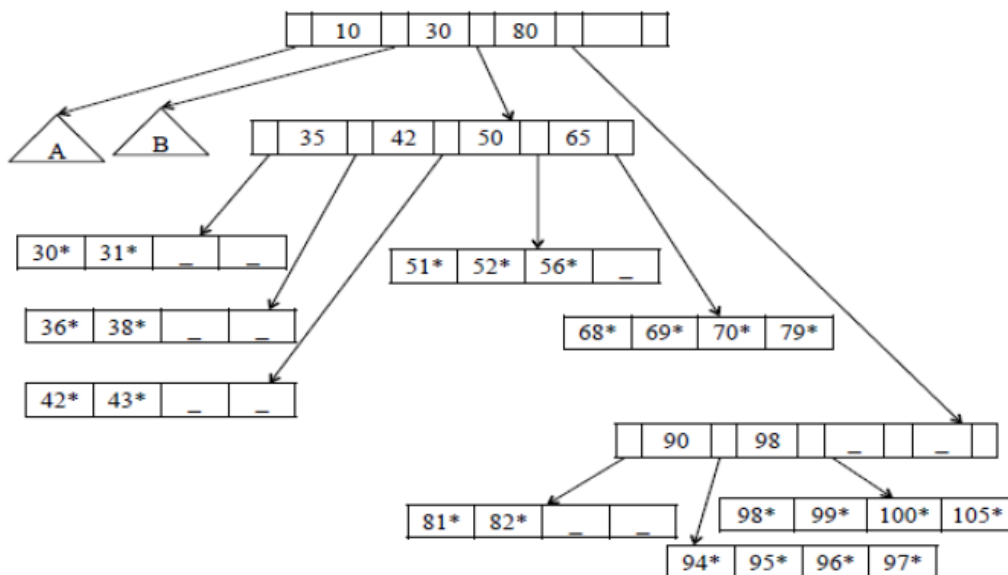
Άσκηση 1 [25 μονάδες]

Θεωρείστε ότι θέλουμε να διαγράψουμε την τιμή 43 στο B+ δέντρο της Εικόνας 1. Η διαγραφή αυτή προκαλεί μείωση της πληρότητας του φύλλου που περιέχει τα 42*, 43*, κάτω του 50%. Θεωρείστε τις παρακάτω τρεις επιλογές:

- (1) Συγχώνευση του φύλλου με τον δεξιό αδελφό του.
- (2) Συγχώνευση του φύλλου με τον αριστερό αδελφό του.
- (3) Ανακατανομή των εγγραφών του φύλλου με τον δεξιό αδελφό του.

και απαντήστε στις παρακάτω ερωτήσεις

- (α) Δώστε τα σωστά B+-δέντρα που προκύπτουν για κάθε μια από τις επιλογές.
- (β) Συγκρίνετε τις διαφορετικές επιλογές από άποψη κόστους.
- (γ) Υπάρχουν περιπτώσεις που θα προτιμούσατε τη μια από την άλλη επιλογή και γιατί. (Υπόδειξη: σκεφτείτε διαφορετικά ποσοστά εισαγωγών και διαγραφών, πχ. τι γίνεται αν έχουμε μόνο διαγραφές τιμών και τι αν έχουμε ίδιο ποσοστό διαγραφών και εισαγωγών).



Εικόνα 1. B+ δέντρο για τις Ασκήσεις 1 και 2. Ο μέγιστος αριθμός δεικτών για τους εσωτερικούς κόμβους είναι 5 και για τα φύλλα 4. Τα φύλλα είναι διπλά δια-συνδεδεμένα μεταξύ τους (δε φαίνεται στην εικόνα). Τα A, B είναι υπο-δέντρα στα οποία δείχνουν οι αντίστοιχοι δείκτες.

Άσκηση 2 [15 μονάδες]

Θεωρείστε το B+ δέντρο της Εικόνας 1 και υποθέστε ότι τα φύλλα των υπό-δέντρων A και B είναι όσο το δυνατόν πιο άδεια. Υποθέστε επίσης ότι το δέντρο ευρετηριοποιεί μια σχέση R, το πεδίο ευρετηριοποίησης είναι υποψήφιο κλειδί και κλειδί διάταξης για την R και ότι κάθε block της R χωράει 20 εγγραφές της. Δώστε μια εκτίμηση για το πόσες πλειάδες έχει η R.

Άσκηση 3 [40 μονάδες]

Θεωρείστε μια σχέση r με σχήμα R(A, B, C, D, E, F) η οποία έχει 6.000.000 πλειάδες (εγγραφές) και κάθε block (σελίδα) της χωράει 10 εγγραφές. Η σχέση είναι αποθηκευμένη σε ένα αρχείο διατεταγμένο σε αύξουσα διάταξη με πεδίο διάταξης το γνώρισμα A. Το γνώρισμα A είναι υποψήφιο κλειδί για τη σχέση και παίρνει τιμές από το 0 έως 5.999.999. Για κάθε μια από τις παρακάτω τέσσερις ερωτήσεις σε σχεσιακή άλγεβρα

Ερώτηση

- (1) $\sigma_{A < 60.000}(r)$
- (2) $\sigma_{A = 60.000}(r)$
- (3) $\sigma_{A > 60.000 \text{ AND } A < 60.010}(r)$
- (4) $\sigma_{A <> 60.000}(r)$

δώστε το κόστος (σε αριθμό προσπελάσεων σελίδων) για τις παρακάτω τρεις διαφορετικές μεθόδους υπολογισμού τους και πείτε ποια μέθοδος είναι η καλύτερη σε κάθε περίπτωση.

Μέθοδος υπολογισμού

- (i) Προσπέλαση του αρχείου δεδομένων απευθείας (χωρίς χρήση ευρετηρίου)
- (ii) Χρήση ενός B+ δέντρου στο γνώρισμα A. Υποθέστε ότι κάθε εσωτερικός κόμβος έχει μέγιστο αριθμό δεικτών 100 και κάθε φύλλο έχει μέγιστο αριθμό κλειδιών 100.
- (iii) Χρήση ευρετηρίου κατακερματισμού στο γνώρισμα A. Θεωρείστε μη δυναμικό κατακερματισμό (δεν υπάρχουν υπερχειλίσεις).

Άσκηση 4 [20 μονάδες]

Έστω ότι έχετε μια σχέση με 100.000 εγγραφές. Επιθυμείτε να κατακερματίσετε τη σχέση σε ένα πίνακα κατακερματισμού με 1.000 κάδους. Ένα μπλοκ δίσκου μπορεί να αποθηκεύσει 100 εγγραφές (μαζί με ένα επιπρόσθετο δείκτη προς ένα μπλοκ υπερχειλίσης). Θεωρείστε ότι ένα μπλοκ δίσκου δεν μπορεί να αποθηκεύσει εγγραφές από δύο διαφορετικούς κάδους.

- (α) Ποιος είναι ο μέγιστος αριθμός από μπλοκ δίσκου που απαιτούνται για την αποθήκευση της σχέσης;
- (β) Ποιος είναι ο ελάχιστος αριθμός από μπλοκ δίσκου που απαιτούνται για τη σχέση;
- (γ) Ποια είναι η απάντηση στο ερώτημα (α) αν η σχέση είχε 100.099 εγγραφές;