

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΑΣΚΗΣΗ ΔΕΥΤΕΡΗ

ΜΑΘΗΜΑ
ΑΚΑΔ. ΕΤΟΣ

ΒΑΣΕΙΣ ΔΕΛΟΜΕΝΩΝ
2009-10

ΕΠΙΣΤΡΕΦΕΤΑΙ

21.12.2009

Η άσκηση αυτή είναι **ατομική**, δε θα γίνουν δεκτές ασκήσεις από ομάδες φοιτητών.
ΔΙΔΑΣΚΟΝΤΕΣ Ιωάννης Βασιλείου *Καθηγητής*, Τομέας Πληροφορικής
Τιμολέων Σελλής *Καθηγητής*, Τομέας Πληροφορικής

Ερώτημα 1.

Θεωρείστε επεκτατό κατακερματισμό όπου κάθε κάδος χωρά μέχρι 3 εγγραφές. Χρησιμοποιείτε τα πρώτα bits.

- (α) Δώστε το ευρετήριο μετά την εισαγωγή των παρακάτω κλειδιών: 1, 3, 6, 14, 15, 17, 21, 23, 24, 25, 29, 31. Προσδιορίστε το ολικό και το τοπικό βάθος.
- (β) Δώστε το ευρετήριο μετά την εισαγωγή των παρακάτω κλειδιών: 17, 15, 14, 6, 3, 1, 31, 29, 25, 24, 23, 21. Προσδιορίστε το ολικό και το τοπικό βάθος.

Ερώτημα 2.

Έστω ένα B^+ - δέντρο τάξης 3 αρχικά άδειο. Θεωρείστε ότι τιμή ίση με την τιμή αναζήτησης σε ένα κόμβο αποθηκεύεται στο υποδέντρο στα δεξιά του.

- (α) Θεωρείστε ότι εισάγουμε τιμές από το 1 έως το 18 με τη σειρά. Δώστε το δέντρο που προκύπτει. Ποιο είναι το μέγεθος του σε blocks;
- (β) Δώστε το δέντρο που προκύπτει αν διαγράψουμε από το δέντρο του ερωτήματος (α) όλους τους πρώτους αριθμούς (από τον μικρότερο στο μεγαλύτερο).

Ερώτημα 3.

Ποιο είναι το μεγαλύτερο μέγεθος αρχείου δεδομένων (σε αριθμό blocks) που μπορεί να δεικτοδοτήσει ένα B^+ - δέντρο τάξης $d = 28$, και που στα φύλλα χωρούν το πολύ 45 ζευγάρια (κλειδί, TID) και ύψους 3 στις παρακάτω περιπτώσεις. Θεωρείστε ότι ο παράγοντας ομαδοποίησης (σελιδοποίησης) του αρχείου είναι 35, δηλ. σε κάθε block του αρχείου δεδομένων χωρούν 35 εγγραφές). Κάντε όποιες άλλες υποθέσεις θεωρείτε αναγκαίες.

- (α) το αρχείο δεδομένων είναι ταξινομημένο ως προς το πεδίο ευρετηριοποίησης και το πεδίο ευρετηριοποίησης είναι κλειδί
- (β) το αρχείο δεδομένων δεν είναι ταξινομημένο ως προς το πεδίο ευρετηριοποίησης και το πεδίο ευρετηριοποίησης είναι κλειδί.
- (γ) το αρχείο δεδομένων είναι ταξινομημένο ως προς το πεδίο ευρετηριοποίησης αλλά το πεδίο ευρετηριοποίησης δεν είναι κλειδί και υπάρχουν 100 εγγραφές ανά τιμή του πεδίου.
- (δ) το αρχείο δεδομένων δεν είναι ταξινομημένο ως προς το πεδίο ευρετηριοποίησης, το πεδίο ευρετηριοποίησης δεν είναι κλειδί και υπάρχουν 100 εγγραφές ανά τιμή του πεδίου.

Ερώτημα 4.

Θεωρήστε μία σχέση $R(A, B, C)$ που περιέχει 100.000 εγγραφές και με κλειδί το πεδίο A που παίρνει τιμές από 0 έως 99.999.

Έχουμε δύο εναλλακτικές να οργανώσουμε τις εγγραφές της R στο δίσκο:

- (α) Σαν ένα κατακερματισμένο αρχείο, το οποίο να έχει 500 κάδους (buckets). Κάθε block στο δίσκο μπορεί να αποθηκεύσει μέχρι 10 εγγραφές – συμπεριλαμβανομένου και ενός δείκτη προς το block υπερχειλίσης (εφόσον χρειαστεί). Θεωρήστε επίσης ότι κάθε block περιέχει εγγραφές από ένα μόνο κάδο.
- (β) Σαν ένα B^+ - δέντρο στο δίσκο. Τα φύλλα του δέντρου περιέχουν τις ίδιες τις εγγραφές της σχέσης και όχι δείκτες προς αυτές. Κάθε φύλλο μπορεί να αποθηκεύσει μέχρι 10 εγγραφές – συμπεριλαμβανομένου και ενός δείκτη προς το επόμενο block στα φύλλα, ενώ κάθε εσωτερικός κόμβος μπορεί να αποθηκεύσει μέχρι 100 κλειδιά.

Για κάθε ένα από τα παρακάτω αιτήματα, προσδιορίστε πόσες λειτουργίες I/O απαιτούνται για τον καθένα από τους παραπάνω τρόπους οργάνωσης για την επεξεργασία του ερωτήματος.

1. Δώσε όλες τις πλειάδες της R .
2. Βρες όλες τις πλειάδες της R ώστε $A < 50$.
3. Βρες όλες τις πλειάδες της R ώστε $A = 50$.
4. Βρες όλες τις πλειάδες της R , ώστε $A > 50$ και $A < 100$.

Πώς θα άλλαζαν οι απαντήσεις σας στις παραπάνω ερωτήσεις αν το γνώρισμα A δεν ήταν κλειδί της R ;