

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΑΣΚΗΣΗ ΔΕΥΤΕΡΗ

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

ΜΑΘΗΜΑ	ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ
ΑΚΑΔ. ΕΤΟΣ	2013-14
ΕΠΙΣΤΡΕΦΕΤΑΙ	20.2.2014
ΔΙΔΑΣΚΩΝ	Ιωάννης Βασιλείου, Καθηγητής, Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Ερώτημα 1.

Θεωρήστε ένα σύστημα Βάσεων Δεδομένων που χρησιμοποιεί B+trees για την αποθήκευση των διατεταγμένων εγγραφών των σχέσεων στο δίσκο. Τα φύλλα του δέντρου περιέχουν τις ίδιες τις εγγραφές της σχέσης και όχι δείκτες προς αυτές. Ένα φύλλο καταλαμβάνει ένα block στο δίσκο και εκτός των εγγραφών της σχέσης περιέχει και ένα δείκτη στο επόμενο φύλλο. Οι εσωτερικοί κόμβοι του δέντρου καταλαμβάνουν επίσης ένα block και περιέχουν κλειδιά του δέντρου και δείκτες προς άλλους κόμβους. Επίσης, θεωρήστε τα παρακάτω χαρακτηριστικά:

- (α) Για κάθε block απαιτούνται 4096 bytes.
- (β) Κάθε εγγραφή έχει μέγεθος 300 bytes.
- (γ) Για κάθε δείκτη σε άλλο κόμβο του δέντρου απαιτούνται 10 bytes.
- (δ) Για κάθε δείκτη σε εγγραφή της σχέσης απαιτούνται 12 bytes.
- (ε) Κάθε κλειδί έχει μέγεθος 8 bytes.
- (στ) Οι κόμβοι του δέντρου είναι κατά 85% γεμάτοι. Για παράδειγμα αν ένα φύλλο μπορεί να αποθηκεύσει μέχρι 100 εγγραφές, τότε αυτό περιέχει μέχρι 85. Ομοίως αν ένας εσωτερικός κόμβος μπορεί να αποθηκεύσει μέχρι 100 κλειδιά, ο κόμβος περιέχει μέχρι 85.
- (ζ) Η σχέση R που οργανώνεται σαν B+tree περιέχει 1.000.000 εγγραφές.

Απαντήστε στα παρακάτω ερωτήματα:

- 1) Πόσες εγγραφές χωράνε σε ένα φύλλο;
- 2) Πόσα blocks καταλαμβάνουν τα φύλλα;
- 3) Πόσα κλειδιά χωράει ένας ενδιάμεσος κόμβος του δένδρου;
- 4) Ποιο είναι το ύψος του δένδρου;
- 5) Πόσα blocks χρειάζονται για την αποθήκευση των εγγραφών της σχέσης R (συμπεριλαμβανομένων των blocks του B+tree);

Ερώτημα 2.

Θεωρήστε μία σχέση R 100.000 εγγραφών. Έχουμε δύο εναλλακτικές για να οργανώσουμε τις εγγραφές της R στο δίσκο:

- (α) Ως ένα κατακερματισμένο αρχείο, το οποίο να έχει 1.000 κάδους (buckets). Κάθε block στο δίσκο μπορεί να αποθηκεύσει μέχρι 100 εγγραφές – συμπεριλαμβανομένου και ενός δείκτη προς το block υπερχειλίσης (εφόσον χρειαστεί). Θεωρήστε επίσης ότι κάθε block περιέχει εγγραφές από ένα μόνο κάδο.
- (β) Ως ένα B+tree στο δίσκο. Τα φύλλα του δέντρου περιέχουν τις ίδιες τις εγγραφές της σχέσης και όχι δείκτες προς αυτές. Κάθε φύλλο μπορεί να αποθηκεύσει μέχρι 100 εγγραφές – συμπεριλαμβανομένου και ενός δείκτη προς το επόμενο block στα φύλλα, ενώ κάθε εσωτερικός κόμβος μπορεί να αποθηκεύσει μέχρι 1.000 κλειδιά.

Σε κάθε μια από τις παραπάνω περιπτώσεις απαντήστε τις παρακάτω ερωτήσεις:

- 1) Ποιος είναι ο ελάχιστος αριθμός blocks στο δίσκο που χρειάζονται για τη σχέση R;
- 2) Ποιος είναι ο μέγιστος αριθμός blocks στο δίσκο που χρειάζονται για τη σχέση R;

Ειδικά για την περίπτωση (α), αν η σχέση περιείχε 100.099 εγγραφές, ποια θα ήταν η απάντηση στο ερώτημα 2);

Ερώτημα 3.

Υποθέστε πως μια σελίδα χωρά το πολύ τέσσερις τιμές δεδομένων και πως όλες οι τιμές είναι ακέραιοι. Χρησιμοποιώντας μόνο B+δέντρα τάξης 2, δώστε παραδείγματα για κάθε ένα από τα παρακάτω:

- 1) Ένα B+δέντρο που η εισαγωγή της τιμής 25 αυξάνει το ύψος του από 2 σε 3. Σχεδιάστε τη δομή σας πριν και μετά την εισαγωγή.
- 2) Ένα B+δέντρο που η διαγραφή της τιμής 25 προκαλεί ανακατανομή. Σχεδιάστε τη δομή σας πριν και μετά τη διαγραφή.
- 3) Ένα B+δέντρο που η διαγραφή της τιμής 25 προκαλεί τη συγχώνευση δύο κόμβων, χωρίς όμως να αλλάζει το ύψος του δέντρου.