

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΛΥΣΗ ΣΤΗΝ ΔΕΥΤΕΡΗ ΑΣΚΗΣΗ

ΜΑΘΗΜΑ
ΑΚΑΔ. ΕΤΟΣ
ΔΙΔΑΣΚΩΝ

ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ
2013-14

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);

ΛΥΣΗ

1) Αφού κάθε φύλλο καταλαμβάνει ένα block, δηλαδή 4096 bytes, χρειάζεται 10 bytes για το δείκτη στο επόμενο φύλλο, και κάθε εγγραφή αντιστοιχεί σε 300 bytes έχουμε:

$$\lfloor (4096 - 10) / 300 \rfloor = 13 \text{ εγγραφές σε κάθε φύλλο}$$

Με δεδομένο ότι οι κόμβοι του δέντρου είναι κατά 85% γεμάτοι ένα φύλλο μπορεί να αποθηκεύσει μέχρι:

$$\lfloor 13 * 0.85 \rfloor = 11 \text{ εγγραφές}$$

2) Με δεδομένο ότι η σχέση R περιέχει 1.000.000 εγγραφές, χρειάζονται:

$$\lfloor 1000000 / 11 \rfloor = 90910 \text{ blocks για να τα φύλλα του δέντρου.}$$

3) Έστω ότι ένα εσωτερικός κόμβος μπορεί να αποθηκεύσει n κλειδιά. Πρέπει να ισχύει:

$$(n + 1) * 10 + n * 8 \leq 4096 \text{ άρα } n = 227$$

Με δεδομένο ότι οι κόμβοι του δέντρου είναι κατά 85% γεμάτοι ένας εσωτερικός κόμβος μπορεί να αποθηκεύσει μέχρι:
 $\lfloor 227 * 0.85 \rfloor = 192$ κλειδιά

4) Κάθε ενδιάμεσο επίπεδο μπορεί να έχει 192 κλειδιά και 193 δείκτες. Το επίπεδο των φύλλων περιέχει 90910 blocks, το αμέσως επόμενο επίπεδο περιέχει $\lceil 90910 / 193 \rceil = 472$ κόμβους, το επόμενο περιέχει 3 κόμβους και το τελευταίο 1, δηλαδή τη ρίζα του δέντρου. Άρα το δέντρο έχει 4 επίπεδα.

5) Συνολικά για την αποθήκευση χρειάζονται:
 $90910 + 472 + 3 + 1 = 91386$ blocks

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 υπερχειλίσσης (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);

ΛΥΣΗ

(α) Ως κατακερματισμένο αρχείο:

1) Στην καλύτερη περίπτωση όλα τα buckets έχουν γεμίσει με 100 εγγραφές το καθένα.

Συνεπώς, ο ελάχιστος αριθμός blocks στο δίσκο είναι $100000 / 100 = 1000$ blocks.

2) Η χειρότερη περίπτωση είναι όταν τα 999 buckets έχουν από μία εγγραφή και στο ένα bucket πέφτουν όλες οι υπόλοιπες εγγραφές. Έτσι το αρχικό block του είναι γεμάτο με 100 εγγραφές ενώ οι υπόλοιπες $100000 - 999 - 100 = 98901$ εγγραφές αποθηκεύονται σε blocks υπερχειλίσσης του ίδιου bucket. Για αυτές τις 98901 εγγραφές χρειάζονται $\lceil 98901 / 100 \rceil = 990$ blocks.

Συνεπώς, ο μέγιστος αριθμός blocks στο δίσκο είναι $1000 + 990 = 1990$ blocks.

(β) Ως ένα B+tree στο δίσκο:

1) Στην καλύτερη περίπτωση οι κόμβοι του B+tree είναι γεμάτοι. Συνεπώς, θα έχουμε 1000 blocks για τα φύλλα, καθένα από τα οποία αποθηκεύει 100 εγγραφές, και 1 block για τη ρίζα που αποθηκεύει 1000 κλειδιά.

Συνολικά, το B+tree θα έχει 2 επίπεδα και απαιτούνται 1001 blocks.

2) Στη χειρότερη περίπτωση οι κόμβοι του B+tree είναι μισό-γεμάτοι. Τα φύλλα θα αποθηκεύουν $100 * 0.5 = 50$ εγγραφές το καθένα και οι ενδιάμεσοι κόμβοι $1000 * 0.5 = 500$ κλειδιά και 501 δείκτες ο καθένας. Συνεπώς, θα έχουμε $100000 / 50 = 2000$ blocks για τα φύλλα, $\lceil 2000 / 501 \rceil = 4$ blocks για το επόμενο επίπεδο και 1 block για τη ρίζα.

Συνολικά, B+tree θα έχει 3 επίπεδα και απαιτούνται 2005 blocks.

Στην περίπτωση που η σχέση R περιείχε 100.099 εγγραφές και οργανώνεται ως κατακευματισμένο αρχείο, τα blocks υπερχειλίστη του ερωτήματος (α) 2) πρέπει να αποθηκεύσουν $100099 - 999 - 100 = 99000$ εγγραφές. Συνεπώς χρειάζονται και πάλι $99000 / 100 = 990$ blocks υπερχειλίστη και συνολικά 1990 blocks στο δίσκο.

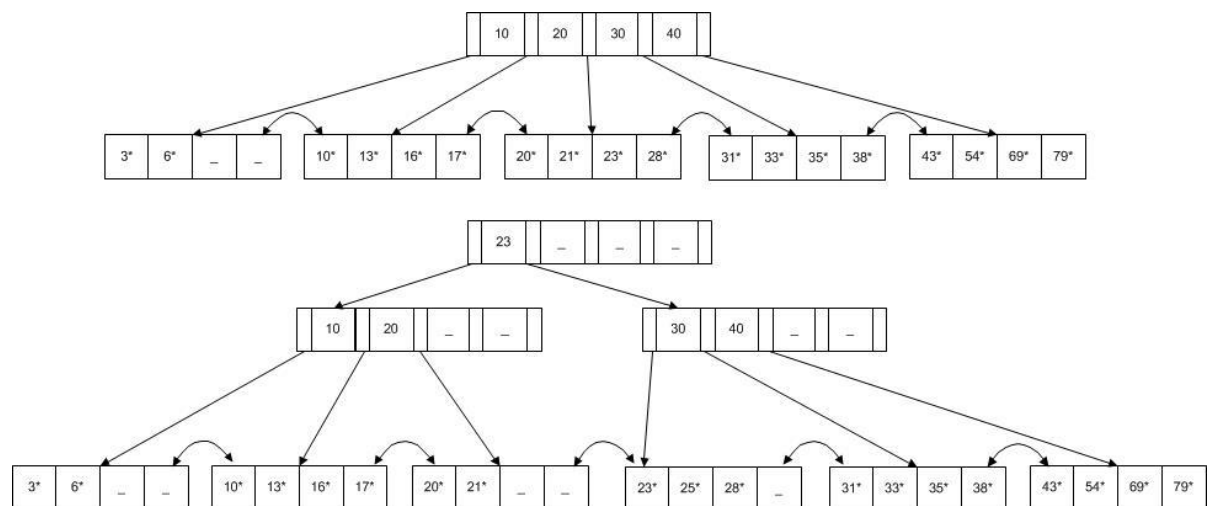
Ερώτημα 3.

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:

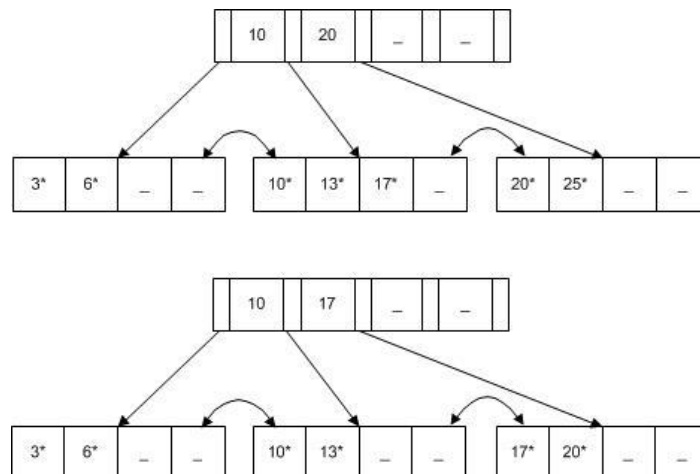
- 1) Ένα B+δέντρο που η εισαγωγή της τιμής 25 αυξάνει το ύψος του από 2 σε 3. Σχεδιάστε τη δομή σας πριν και μετά την εισαγωγή.
- 2) Ένα B+δέντρο που η διαγραφή της τιμής 25 προκαλεί ανακατανομή. Σχεδιάστε τη δομή σας πριν και μετά τη διαγραφή.
- 3) Ένα B+δέντρο που η διαγραφή της τιμής 25 προκαλεί τη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

ΛΥΣΗ

1)



2)



3)

