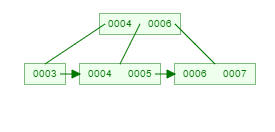
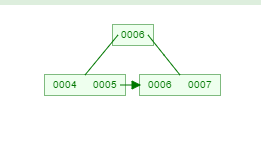
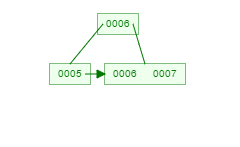
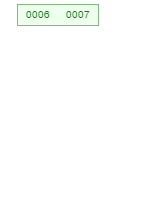
**Βάσεις Δεδομένων ΙΙ**

Εργασία 1

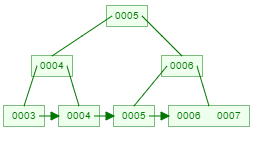
1. Ένα Β+-tree έχει κόμβους μη-φύλλα που η μέγιστη χωρητικότητα τους είναι 2 κλειδιά αναζήτησης και 3 δείκτες και κόμβους φύλλα που η μέγιστη χωρητικότητά τους είναι 2 καταχωρίσεις δεδομένων.

(α) Ξεκινώντας από ένα δέντρο με ένα μόνο κόμβο φύλλο που περιέχει τις καταχωρίσεις δεδομένων 6\* και 7\*, να εισάγετε διαδοχικά τις καταχωρίσεις δεδομένων 5\*, 4\*, και 3\*, σχεδιάζοντας κάθε φορά το δέντρο.

(β) Χρησιμοποιείστε μαζική φόρτωση για να δημιουργήσετε ένα νέο B+tree που να περιέχει όλες τις παραπάνω καταχωρίσεις δεδομένων. Θεωρείστε ότι οι σελίδες φύλλα γεμίζουν στο 100% της χωρητικότητάς τους.



b) Αρχικοποίηση: Ταξινόμησε όλες τις καταχωρίσεις δεδομένων, και βάλε έναν pointer στην πρώτη σελίδα (φύλλο) σε μια νέα σελίδα (ρίζα). Οπότε οι κομβόι θα εισαχθούν με αυτήν την σειρά 3,4,5,6,7 και θα είναι αυτής της μορφής.



1. Δίνεται στιγμιότυπο σχεσιακού πίνακα με 50.000 γραμμές/εγγραφές. Έστω ότι κατασκευάζουμε ευρετήριο τύπου Β+tree Εναλλακτικής (2).

(α) Ποιο είναι το μέγιστο ύψος (μαζί με τη ρίζα) και (β) Ποιο είναι το ελάχιστο ύψος (μαζί με τη ρίζα)

του εν λόγω Β+tree, όταν είναι γνωστό ότι η τάξη όλων των κόμβων του δέντρου είναι d=200;

Το δένδρο τάξεως 200 που σημαίνει ότι κάθε κόμβος έχει από 100 έως 199 κλειδιά εκτός της ρίζας και έως 200 παιδιά

το μέγιστο ύψος ενός Β-δένδρου ύψος ενός δένδρου με τάξεως m για Ν στοιχεία. 1+2(m/2-1)+2 1+2(ceil(m/2)-1)+2 (ceil(m/2)-1)( ceil(m/2)) +…2 (ceil(m/2)) ( ceil(m/2))h-2 =2 ( ceil(m/2))h-1-1

Όμως έχουμε h≥logm(N+1) Άρα logm(N+1)≤h≤ log ceil(m/2) ((N+1)/2)

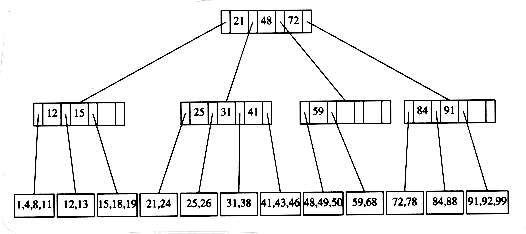
Αρα : log200(50.001)<=h<=log100(25.001)=2<=h<=3

1. Το παρακάτω Β+tree έχει τάξη d=2 (αντίθετα από ότι σας είπα στο μάθημα, στο συγκεκριμένο παράδειγμα, θεωρούμε ότι η τάξη στα μη-φύλλα αφορά το πλήθος των παιδιών/pointers και όχι των τιμών που έχει ο κόμβος). Σχεδιάστε το δέντρο ή περιγράψτε τί θα συμβεί:

(α) μετά την εισαγωγή των καταχωρίσεων δεδομένων 42 και 45, (β) μετά τη διαγραφή της καταχώρισης δεδομένων 59

(γ) μετά τη διαγραφή των καταχωρίσεων δεδομένων 59 και 50.

Για κάθε περίπτωση από τις (α), (β) (γ), υποθέστε ότι ξεκινάτε από το αρχικό δέντρο. Υποθέστε ότι όταν έχουμε περιττό αριθμό καταχωρίσεων δεδομένων (στα φύλλα) και ευρετηρίου (στα μη- φύλλα), κατά τη διάσπαση του κόμβου τα μισά + 1 μένουν αριστερά. Τέλος, υποθέστε ότι η ισότητα “πάει” στα δεξιά.



A)με την εισαγωγή του 42 ο κόμβος (41,43,46) θα διασπαστεί γιατί δεν υπάρχει άλλος χώρος υπερχειλίζει και το 43 ανεβαίνει πάνω υπερχειλίζει τον γονέα του με αποτέλεσμα και αυτός να διασπαστεί και να ανεβεί το 41 ένα επίπεδο πιο πάνω ( στην ριζά) το οποίο υπερχειλίζει και τον δικό του πατερά( την ριζά) με αποτέλεσμα να ψηλώσει κιάλο το δέντρο το 48 θα είναι η νέα ριζά θα έχει παιδιά το (21,43) και το (72) ενώ το 43 θα έχει παιδιά τους κόμβους (25,31) και (41,43) οπού το 43 θα έχει παιδιά τους κόμβους ( 41,42) και (43,46) . η εισαγωγή του 45 απλά θα προστεθεί στον κόμβο (43,46).

B) μετα την διαγραφή του 59 ο κόμβος υποχειλιζει αλλά το αριστερό αδέλφι έχει να του δανείσει ένα παιδί οπότε ανεβαίνει πάνω το 49 και δίνεται διπλά το 50 οπότε και σχηματίζεται ο κόμβος (48,49) και ( 50,68).

Γ) στην περίπτωση που διαγραφεί και το 50 λόγο υποχειλησης τα 2 παιδιά συγχωνεύονται στον κόμβο ( 48,49,68) το 41 θα ανέβει πάνω και το 48 θα κατέβει κάτω δεξιά και θα έχει παιδιά τους κόμβους ( 41,43,46) και κόμβο ( 48,49,68).

1. Έχετε έναν πίνακα που αποτελείται από 100.000 σελίδες.

(α) Με πόσα περάσματα ταξινομείται ο πίνακας αν έχουμε Β=100 διαθέσιμα πλαίσια στην ενδιάμεση μνήμη;

ΠΕΡΑΣΜΑ 0: Διαβάζω το αρχείο σε δόσεις των 100 σελίδων και δημιουργώ ταξινομημένους συρμούς των 100 σελίδων. Πόσους συρμούς; 100.000/100=1000

ΠΕΡΑΣΜΑ 1: Με 99 πλαίσια εισόδου, θα διαβάσω τους 1000 συρμούς σε 11 δόσεις (99+99+99+99+99+99+99+99+99+99+10). Σε κάθε δόση (εκτός της τελευταίας) θα πράξω ένα συρμό των 99\*1000=99.900 σελίδων.

ΠΕΡΑΣΜΑ 2: Συγχωνεύω τους 11 συρμούς στον ένα τελικό συρμό.

πλήθος\_περασμάτων = 1+⌈logB−1 (⌈ N / B⌉)⌉ = 1+⌈log99 (1000)⌉ = 1 + 2 = 3 Κόστος: 2 \* 100.000 \* 3 = 300.000Ι/Ο

(β) Με πόσα περάσματα ταξινομείται ο πίνακας αν έχουμε Β=1000 διαθέσιμα πλαίσια στην ενδιάμεση μνήμη;

ΠΕΡΑΣΜΑ 0: Διαβάζω το αρχείο σε δόσεις των 1000 σελίδων και δημιουργώ ταξινομημένους συρμούς των 1000 σελίδων. Πόσους συρμούς; 100.000/1000=100

ΠΕΡΑΣΜΑ 1: Χρησιμοποιώ τα 999 πλαίσια ως πλαίσια εισόδου και το 1 πλαίσιο ως πλαίσιο εξόδου. Έχω (μόνο) 100 συρμούς, άρα τους συγχωνεύω όλους με ένα πέρασμα. Ως έξοδο παίρνω 1 συρμό των 100.000 σελίδων.

1. Έχουμε Β=20 διαθέσιμα πλαίσια ενδιάμεσης μνήμης για να ταξινομήσουμε έναν πίνακα. 1) Ποιο είναι το μέγεθος Ν (σε σελίδες) του μεγαλύτερου πίνακα που μπορούμε να ταξινομήσουμε με 2 περάσματα; 2) Δώστε την απάντηση αν Β=1000. Υποθέστε ότι στο Πέρασμα 0 παράγονται συρμοί μήκους Β.

1)⌈log19(20)⌉=2

πλήθος\_περασμάτων = 1+⌈logB−1 (⌈ N / B⌉)⌉ = 1+⌈log19 (N/20)⌉ = 2 =>⌈log19 (N/20)⌉=1=>

⌈log19 (N) ⌉- log19(20)=1=>⌈log19 (N) ⌉=2 (1)

αx = θ ⇔ x = logαθ (2)

από (1),(2) =>19^2=Ν=361

2)

πλήθος\_περασμάτων = 1+⌈logB−1 (⌈ N / B⌉)⌉ = 1+⌈log999 (N/1000)⌉ = 2 =>⌈log999 (N/1000)⌉=1=>

⌈log999 (N) ⌉- log999(1000)=1=>⌈log999 (N) ⌉=2 (1)

από (1),(2) =>999^2=Ν= 998.001περιπου 998.000

ΠΕΡΑΣΜΑ 0: Διαβάζω το αρχείο σε δόσεις των 1000 σελίδων και δημιουργώ ταξινομημένους συρμούς των 1000 σελίδων. Πόσους συρμούς; 998.000/1000=998

ΠΕΡΑΣΜΑ 1: Χρησιμοποιώ τα 999 πλαίσια ως πλαίσια εισόδου και το 1 πλαίσιο ως πλαίσιο εξόδου. Έχω (μόνο) 998 συρμούς, άρα τους συγχωνεύω όλους με ένα πέρασμα. Ως έξοδο παίρνω 1