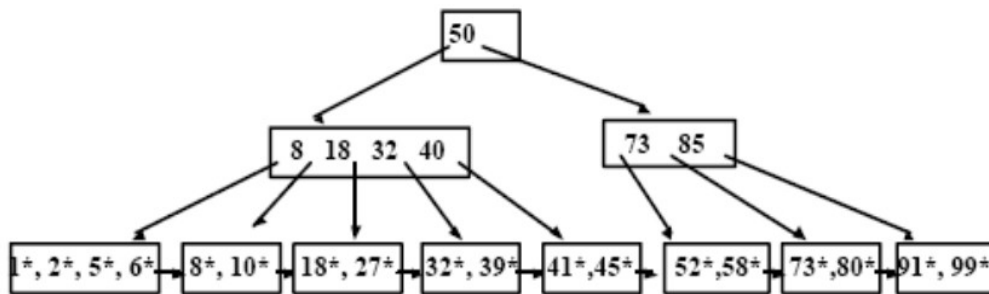


## Βάσεις Δεδομένων 2 – Εργασία 1 (B+trees και Εξωτερική Ταξινόμηση)

[Α]. Θεωρείστε το παρακάτω B+tree (κάθε κόμβος ευρετηρίου χωρά 4 καταχωρίσεις ευρετηρίου και κάθε κόμβος φύλλο χωρά 4 καταχωρίσεις δεδομένων – υποθέστε ότι όταν ένας κόμβος φύλλο δέχεται 5η καταχώριση διασπάται ώστε 3 καταχωρίσεις να μένουν στον αριστερό κόμβο και 2 στον δεξιό):



1. Ποιο είναι το μεγαλύτερο πλήθος διαγραφών που μπορούν να γίνουν για να μειωθεί το ύψος του δέντρου κατά 1; Δώστε μια τέτοια ακολουθία διαγραφών [υποθέτουμε ότι κατά τη διαγραφή πρώτα γίνεται προσπάθεια ανακατανομής με το αριστερό ή το δεξιό αδέρφι κατά σειρά, και μετά συγχώνευσης με το αριστερό ή το δεξιό αδέρφι κατά σειρά].
2. Ποιο είναι το μεγαλύτερο πλήθος εισαγωγών που μπορούν να γίνουν ώστε ο κόμβος με τις τιμές 73 και 85 να γεμίσει; Δώστε μια τέτοια ακολουθία εισαγωγών;

[Β]. Υποθέστε ότι μόλις κατασκευάσατε ένα B+ tree Εναλλακτικής (2) πάνω σε ένα αρχείο σωρού για έναν πίνακα με 100000 εγγραφές. Το πεδίο αναζήτησης του B+ tree είναι αλφαριθμητικό με μέγεθος 90 bytes και αποτελεί υποπήφιο κλειδί του πίνακα. Οι pointers (δηλαδή τα rid στις καταχωρίσεις δεδομένων στα φύλλα του B+ tree καθώς και οι διευθύνσεις των σελίδων του B+ tree στις καταχωρίσεις ευρετηρίου) είναι 10 bytes. Το μέγεθος μια σελίδας του δίσκου είναι 1000 bytes. Συνεπώς, μπορείτε να υποθέσετε ότι σε κάθε κόμβο φύλλο χωράνε max 10 καταχωρίσεις δεδομένων, ενώ σε κάθε κόμβο ευρετηρίου (ή μη-φύλλο) χωράνε max 10 τιμές και 11 pointers.

Το ευρετήριο κατασκευάστηκε με τον αλγόριθμο μαζικής φόρτωσης από κάτω προς τα επάνω και οι κόμβοι στο επίπεδο των φύλλων γέμισαν στο μέγιστο βαθμό (100%) ενώ στα υπόλοιπα επίπεδα του δέντρου είναι όλοι 50% γεμάτοι.

1. Πόσα επίπεδα έχει το δέντρο και πόσους κόμβους έχει το κάθε επίπεδο;
2. Πόσα επίπεδα θα είχε το δέντρο και πόσους κόμβους θα είχε το κάθε επίπεδο αν όλοι οι κόμβοι ήταν 70% γεμάτοι κατά μέσο όρο;

[Γ]. Υποθέστε ότι κάθε εγγραφή του πίνακα του προβλήματος Β είναι 200 bytes, συνεπώς, με σελίδες των 1000 bytes έχουμε 5 εγγραφές ανά σελίδα και το αρχείο σωρού καταλαμβάνει 20000 σελίδες στο δίσκο.

1. Αν έχουμε B=3 διαθέσιμα ενδιάμεσα πλαίσια, με πόσα περάσματα ταξινομείται ο πίνακας;
2. Πόσα περάσματα χρειάζονται αν B=101;
3. Πόσα περάσματα χρειάζονται αν B=142;
4. Πόσα περάσματα χρειάζονται αν B=19999;
5. Πόσα περάσματα χρειάζονται αν B=20000;

Καταθέστε στο openeclass ένα αρχείο εγγράφου (doc, pdf ή txt) με τις απαντήσεις σας.