

Σχεδιασμός Βάσεων Δεδομένων
Διδάσκων: Ιωάννης Κωτίδης

Εαρινό εξάμηνο 2020-2021

Πρώτη Σειρά Ασκήσεων

Ανάθεση: 04-04-2021

Παράδοση: **13-04-2021 Ώρα (23:55)**

Οδηγίες

- Η πρώτη σειρά ασκήσεων είναι **ατομική** και **υποχρεωτική**.
- Η υποβολή της εργασίας πρέπει να γίνει στο *eclass*.
- Το παραδοτέο σας θα πρέπει να είναι ένα αρχείο PDF με όνομα *AM.pdf* (όπου *AM* είναι ο αριθμός μητρώου σας. π.χ. "3180001.pdf").
- Τα διαγράμματα πρέπει να είναι κατασκευασμένα σε κάποιο πρόγραμμα (της επιλογής σας) και όχι σκαναρισμένα χειρόγραφα.
- Πιθανή αντιγραφή θα τιμωρείται με μηδενισμό όλων των εμπλεκομένων.
- **Για την επίλυση των ασκήσεων να μελετήσετε τις διαφάνειες των διαλέξεων του μαθήματος.**

Η συνολική βαθμολογία των ασκήσεων ανέρχεται σε 120 (100+20 μονάδες bonus).

Άσκηση 1 [10 μονάδες]

Ένας σκληρός δίσκος έχει τα παρακάτω χαρακτηριστικά:

- Μέγεθος τομέα (sector) 4096 bytes
- 65536 ίχνη (tracks) ανά επιφάνεια (surface)
- 8 πλακέτες (platters) διπλής όψης
- 256 τομείς ανά ίχνο
- Μέσος Χρόνος Μετακίνησης Κεφαλής 10 ms
- Ταχύτητα περιστροφής 7200 rpm

Να απαντήσετε τις ακόλουθες ερωτήσεις:

- Να υπολογίσετε την χωρητικότητα σε bytes ενός ίχνους, μιας επιφάνειας και ολόκληρου του δίσκου.
- Να υπολογίσετε τον αριθμό των κυλίνδρων του δίσκου.
- Ποιά είναι η μέγιστη και η μέση καθυστέρηση περιστροφής;
- Ποιός είναι ο ρυθμός μεταφοράς (transfer rate) του δίσκου αν σε μία περιστροφή μπορεί να διαβαστεί και να μεταφερθεί στην κύρια μνήμη (RAM) ένα ίχνο;

Άσκηση 2 [20 μονάδες]

Έστω η σχέση $R(\#a,b,c,d,e)$ για την οποία ισχύουν:

- Το γνώρισμα *a* είναι το πρωτεύον κλειδί της σχέσης και έχει μέγεθος 10 bytes.
- Τα γνωρίσματα *b,c,d,e* έχουν μέγεθος 50,30,18 και 20 bytes αντίστοιχα.
- Η σχέση *R* έχει *N* εγγραφές

Θεωρείστε επίσης ότι το μέγεθος ενός δείκτη *p* προς μια εγγραφή της σχέσης *R* είναι 6 bytes και το μέγεθος μιας σελίδας είναι 1024 bytes. Υποθέστε ότι όλη η χωρητικότητα μιας σελίδας μπορεί να χρησιμοποιηθεί για την αποθήκευση των εγγραφών.

Προσδιορίστε ως συνάρτηση του N τον αριθμό των απαιτούμενων μπλοκ για την αποθήκευση:

- a) Της σχέσης R και ενός πυκνού ευρετηρίου (dense index) στο πρωτεύον κλειδί της R .
- b) Της σχέσης R και ενός αραιού ευρετηρίου (sparse index) στο πρωτεύον κλειδί της R .

Να αιτιολογήσετε τις απαντήσεις σας.

Άσκηση 3 [25 μονάδες]

Εκτελέστε τις παρακάτω λειτουργίες στο B+δέντρο της τελευταίας σελίδας. Θεωρήστε ότι οι τιμές του κλειδιού αναζήτησης είναι μοναδικές (unique index).

1. Αναζήτηση της εγγραφής με κλειδί 41.
2. Αναζήτηση της εγγραφής με κλειδί 40.
3. Αναζήτηση των εγγραφών με κλειδιά μικρότερα του 30.
4. Αναζήτηση των εγγραφών με κλειδιά στο διάστημα $[20,35]$.
5. Εισαγωγή του κλειδιού με τιμή 1.
6. Διαδοχική εισαγωγή τριών κλειδιών με τιμές 14,15 και 16 αντίστοιχα.

Για κάθε λειτουργία αναζήτησης (1,2,3,4) να περιγράψετε με ακρίβεια τους κόμβους και την σειρά με την οποία πρέπει να προσπελαστούν σε κάθε περίπτωση. Για κάθε εισαγωγή κλειδιού (λειτουργίες 5 και 6) να περιγράψετε τις αλλαγές που προκαλούνται στο δέντρο και να δώσετε την νέα μορφή των κόμβων που τροποποιούνται. Στο τέλος να παραθέσετε την τελική μορφή του δέντρου.

Άσκηση 4 [20 μονάδες]

Θεωρείστε ένα DBMS με τα εξής χαρακτηριστικά:

- Μέγεθος σελίδας 2048 bytes
- Μέγεθος δείκτη προς κόμβο δέντρου 12 bytes
- Μέγεθος δείκτη προς εγγραφή δεδομένων 12 bytes

Δημιουργούμε ένα δευτερεύον ευρετήριο στο κλειδί αναζήτησης k μιας σχέσης R . Το μέγεθος του κλειδιού αναζήτησης k είναι 8 bytes. Υπολογίστε τον μέγιστο αριθμό των εγγραφών της σχέσης R που μπορούν να ευρετηριαστούν με ένα ευρετήριο B+δέντρου **τριών επιπέδων** (ρίζα, ενδιάμεσο επίπεδο και επίπεδο με κόμβους φύλλα).

Σημείωση: Για λόγους απλούστευσης θεωρείστε ότι μπορεί να χρησιμοποιηθεί όλος ο χώρος της σελίδας (δηλαδή και τα 2048 bytes) και ότι οι κόμβοι όλων των επιπέδων του B+δέντρου μπορεί να είναι εντελώς γεμάτοι (100%).

Άσκηση 5 [25 μονάδες]

Έστω ένα αρχείο ευρετηρίου που χρησιμοποιεί την μέθοδο του γραμμικού κατακερματισμού με αρχικό μέγεθος **2** κάδους (**$m=1$**) χωρητικότητας τριών εγγραφών έκαστο. Για την κατανομή των τιμών χρησιμοποιούνται τα **$i=1$** λιγότερο σημαντικά bits. Ο αριθμός των κάδων πρέπει να αυξάνεται όταν το utilization του ευρετηρίου γίνει μεγαλύτερο ή ίσο του **70%**. Το **i** αυξάνεται μόνο όταν κρίνεται απαραίτητο. Επίσης, δεν υπάρχει όριο στον αριθμό σελίδων υπερχειλίσης. Κάθε σελίδα υπερχειλίσης χωράει και αυτή τρείς εγγραφές.

Ζητείται να εισαγάγετε τα παρακάτω κλειδιά (αναγράφεται η τιμή $h(x)$ αντί για το κλειδί x) με την σειρά που σας δίνονται ξεκινώντας από αριστερά προς τα δεξιά.

[0000, 0001, 0001, 0101, 0111,0010, 0111,0010,0011,0100]

Να εμφανίσετε την μορφή του ευρετηρίου μετά από κάθε εισαγωγή κλειδιού δείχνοντας και όσα ενδιαμέσως βήματα απαιτούνται. Κάθε πράξη εισαγωγής πρέπει να εκτελείται στο αποτέλεσμα της προηγούμενης και όχι στο αρχικό ευρετήριο.

Προς διευκόλυνσή σας ακολουθεί η εισαγωγή του πρώτου κλειδιού:

1. Εισαγωγή 0000

0000	
0	1

utilization=1/6

Άσκηση 6 [10 μονάδες]

Θεωρήστε ένα περιβάλλον επεκτατού κατακερματισμού (extendible hashing) το ολικό βάθος του ευρετηρίου (d) είναι 10 και το μέγεθος κάθε εγγραφής του ευρετηρίου είναι 4 bytes. Το μέγεθος ενός κάδου δεδομένων είναι 2400 bytes και το μέγεθος μιας εγγραφής δεδομένων είναι 400 bytes.

Να απαντήσετε τις ακόλουθες ερωτήσεις:

- Πόσες εγγραφές έχει το ευρετήριο;
- Ποιό είναι το μέγεθος του ευρετηρίου σε bytes;
- Ποιός είναι ο μέγιστος αριθμός εγγραφών που μπορεί να περιέχει το αρχείο των δεδομένων;

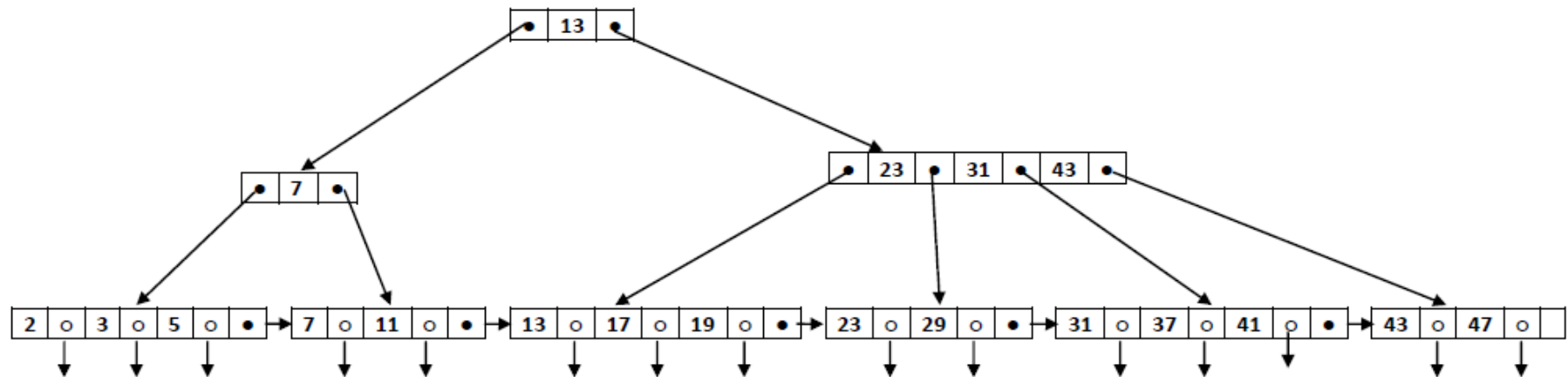
Άσκηση 7 [10 Μονάδες]

Ένας σχεδιαστής βάσεων δεδομένων του υπουργείου εργασίας θέλησε να οργανώσει τις εγγραφές όλων των ελλήνων εργαζόμενων σε ένα αρχείο κατακερματισμού με βάση το μηνιαίο μισθό τους. Ο σχεδιαστής σκέφτηκε: "θα επιλέξω μία συνάρτηση κατακερματισμού η οποία θα εκχωρεί σε κάθε κάδο, τον ίδιο αριθμό τιμών κλειδιών από το σύνολο όλων των πιθανών τιμών κλειδιών (το κλειδί στην προκειμένη περίπτωση είναι ο μισθός). Με τον τρόπο αυτό θα πετύχω την ομοιόμορφη κατανομή των εγγραφών στους κάδους". Ο σχεδιαστής επέλεξε μια συνάρτηση η οποία χωρίζει τις τιμές του μισθού σε διαστήματα με εύρος 1000 ως εξής:

Κάδος	Μισθός
1	[1000..2000)
2	[2000..3000)
3	[3000..4000)
...

Θεωρήστε ότι οι κάδοι περιέχουν τις εγγραφές των εργαζομένων. Συμφωνείτε ή όχι με το σκεπτικό του σχεδιαστή; Να αιτιολογήσετε την απάντησή σας.

B+ Δέντρο (για την Άσκηση 3)



- Δείκτης προς κόμβο του δέντρου.
- Δείκτης δεδομένων (δείκτης προς την εγγραφή)
- Δείκτης δέντρου με τιμή NULL