

Συστήματα Διαχείρισης και Ανάλυσης Δεδομένων
Διδάσκων: Ιωάννης Κωτίδης

Εαρινό εξάμηνο 2022-2023

Πρώτη Σειρά Ασκήσεων

Ανάθεση: 23-04-2023

Παράδοση: 03-04-2023 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. "3200001.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 105 μονάδες (100+5 μονάδες bonus).

Άσκηση 1 [25 μονάδες]

Ένας σκληρός δίσκος έχει τα παρακάτω χαρακτηριστικά:

- 10 επιφάνειες
- 10.000 κυλίνδρους (cylinders)
- 1.000 τομείς ανά ίχνος (sectors/track)
- Μέγεθος τομέα (sector) 512 bytes
- Μέσος Χρόνος Μετακίνησης Κεφαλής (Average Seek Time) = 8 ms
- Ταχύτητα περιστροφής 10.000 rpm
- Μέγεθος block 4096 bytes.

Ζητείται να υπολογίσετε τα ακόλουθα:

1. Τη συνολική χωρητικότητα του δίσκου.
2. Τη μέση καθυστέρηση περιστροφής (Average rotation latency).
3. Τον χρόνο μεταφοράς ενός block (Block Transfer Time).
4. Στον εν λόγω δίσκο είναι αποθηκευμένο ένα αρχείο με 1.000.000 εγγραφές ενός πίνακα της βάσης. Δεδομένου ότι κάθε εγγραφή καταλαμβάνει ακριβώς ένα block και το αρχείο αποθηκεύεται σε συνεχόμενα block του δίσκου να υπολογίσετε τον χρόνο που απαιτείται για την ανάγνωση ολόκληρου του αρχείου.
5. Έστω ότι δημιουργούμε ένα ευρετήριο B+ δέντρου στο πρωτεύον κλειδί του πίνακα. Κάθε κόμβος του δέντρου (εσωτερικός ή φύλλο) χωράει ακριβώς σε ένα block του δίσκου, αλλά οι κόμβοι του δέντρου δεν αποθηκεύονται σε συνεχόμενα block. Το δέντρο έχει 3 επίπεδα (συμπεριλαμβανομένης και της ρίζας) και κανένας κόμβος δεν αποθηκεύεται προσωρινά στην μνήμη. Υποθέστε ότι θέλετε να ανακτήσετε *N* τυχαίες εγγραφές χρησιμοποιώντας το ευρετήριο. Να υπολογίσετε τον χρόνο που απαιτείται για την ανάγνωση των *N* εγγραφών. Ο χρόνος να δοθεί ως συνάρτηση του *N*.

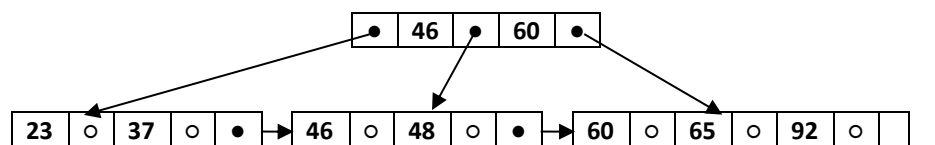
Άσκηση 2 [15 μονάδες]

Δίνεται ο πίνακας Employees(empid int, age int, salary decimal(6,2), departmentid int). Υπάρχει ένα ευρετήριο συστάδων (clustered index) στο πεδίο empid και ένα απλό ευρετήριο (non-clustered index) στο πεδίο age. Για κάθε μία από τις παρακάτω εντολές ενημέρωσης να απαντήσετε αν η εκτέλεσή της επιταχύνεται, επιβραδύνεται ή δεν επηρεάζεται από την ύπαρξη των παραπάνω ευρετηρίων. Να αιτιολογήσετε εν συντομία τις απαντήσεις σας.

1. UPDATE Employees SET age = age + 1
2. UPDATE Employees SET salary = salary * 1.10 WHERE empid >=1 and empid <=100
3. UPDATE Employees SET salary=salary * 1.10 WHERE departmentid = 10

Άσκηση 3 [20 Μονάδες]

Δίνεται το παρακάτω B+ δέντρο με μέγιστο αριθμό **τριών** κλειδιών (**n=3**) ανά κόμβο/φύλλο.



- Δείκτης προς κόμβο του δένδρου
- Δείκτης δεδομένων
- Δείκτης δένδρου με τιμή NULL

Ζητείται:

1. Να εισαγάγετε με την σειρά που δίνονται τις τιμές **47,100** και **50**. Σε κάθε εισαγωγή να παρουσιάσετε τη νέα μορφή του δέντρου και να εξηγήσετε πως ακριβώς προέκυψε.
2. Στο δέντρο που θα προκύψει μετά την εισαγωγή των τιμών του παραπάνω ερωτήματος **πόσοι** και **ποιοι** κόμβοι πρέπει να προσπελαστούν για να ανακτηθούν όλες οι εγγραφές με κλειδί αναζήτησης **key >= 46 AND key <= 65**; α) όταν το κλειδί αναζήτησης είναι μοναδικό και β) όταν το κλειδί αναζήτησης δεν είναι μοναδικό.

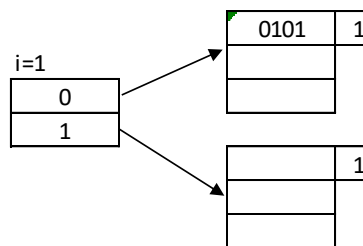
Άσκηση 4 [15 Μονάδες]

Θεωρείστε ότι σε ένα block του δίσκου χωράνε 5 εγγραφές μιας σχέσης R ή 10 κλειδιά (keys) και 11 δείκτες (pointers) ενός B+ δέντρου. Αν η σχέση R περιέχει 1000 εγγραφές, ποιος είναι ο ελάχιστος αριθμός των block που απαιτούνται για την αποθήκευση της σχέσης R και ενός αραιού (sparse) ευρετηρίου μορφής B+ δέντρου στο κλειδί της R;

Να θεωρήσετε ότι ένας δείκτης προς κόμβο του δέντρου έχει το ίδιο μέγεθος σε bytes με έναν δείκτη προς εγγραφή δεδομένων.

Άσκηση 5 [20 Μονάδες]

Θεωρείστε το παρακάτω ευρετήριο επεκτατού κατακερματισμού, με χώρο τριών κλειδιών ανά κάδο και μία συνάρτηση κατακερματισμού η οποία επιστρέφει 4 bits για κάθε κλειδί. Οι κάδοι περιέχουν τις κατακερματισμένες τιμές των κλειδιών. Στο ευρετήριο έχει γίνει εισαγωγή ενός κλειδιού με τιμή κατακερματισμού **0101**.



Ζητείται:

1. Να εισαγάγετε τις ακόλουθες τιμές με την σειρά που δίνονται: **[0111, 1111, 0110, 0000, 0100, 0011, 0001, 0010, 1000, 1010]**. Να δείξετε την μορφή του ευρετηρίου μετά την εισαγωγή τιμών που οδηγεί σε αύξηση του ολικού βάρους καθώς και την τελική του μορφή.
2. Ποιος είναι ο ελάχιστος αριθμός τιμών που θα οδηγούσε σε διπλασιασμό του ευρετηρίου; Δώστε ένα παράδειγμα αυτών των τιμών (τη δυαδική τους αναπαράσταση). Να αιτιολογήσετε την απάντησή σας.

Άσκηση 6 [10 Μονάδες]

Θεωρείστε ένα ευρετήριο επεκτατού κατακερματισμού κάθε κάδος του οποίου χωράει N εγγραφές. Ποια είναι η πιθανότητα ένας υπερχειλισμένος κάδος να πρέπει να αντιμετωπιστεί αναδρομικά; Δηλαδή όλες οι εγγραφές του κάδου να μεταφερθούν στον έναν από τους δύο νέους κάδους που θα δημιουργηθούν;