

ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΔΟΜΕΣ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΑΡΧΕΙΩΝ

3^η άσκηση

Ημερομηνία παράδοσης: 28 Μαΐου 2021

Η άσκηση είναι ατομική

B+ Δένδρα

Υλοποιείτε ένα B+ δένδρο στο δίσκο. Οι σελίδες του δίσκου (όπως και οι κόμβοι του δένδρου) έχουν μέγεθος $N = 256$ bytes. Κάθε κλειδί συσχετίζεται με ένα πεδίο δεδομένων μεγέθους $m = 32$ bytes. Οι αριθμοί N , m είναι παράμετροι της υλοποίησης.

Οι κόμβοι είναι B+-tree αποθηκεύουν δείκτες και κλειδιά χωρίς τα δεδομένα τους (δεν περιέχουν πεδίο δεδομένων). Αποθηκεύουν πληροφορία όπως ο αριθμός των κλειδιών (πόσα κλειδιά αποθηκεύονται στο κόμβο), δείκτες στις σελίδες στο παρακάτω επίπεδο, ένα δείκτη στο παραπάνω επίπεδο. Οι εσωτερικοί κόμβοι αποτελούν δομή δεικτοδότησης (index) για τα δεδομένα. Κάθε κόμβος συνδέεται με δείκτη με τον διπλανό του κόμβο στο ίδιο επίπεδο (οι κόμβοι σε κάθε επίπεδο, όπως και στα φύλλα δημιουργούν λίστα).

Ειδικότερα, οι κόμβοι στα φύλλα περιέχουν δείκτες σε ένα εξωτερικό αρχείο δεδομένων. Δηλαδή, κάθε κλειδί (όπως όλοι οι κόμβοι) έχει ένα κλειδί που αντί να δείχνει σε κόμβο στο παρακάτω επίπεδο, δείχνει σε ένα άλλο εξωτερικό αρχείο (την θέση στην οποία αποθηκεύεται το πεδίο δεδομένων στο αρχείο).

Για συγκεκριμένο $N=256$ bytes υπολογίστε τον βαθμό n του B+tree. Ο βαθμός του δένδρου συσχετίζεται με τον αριθμό δεικτών και κλειδιών σε κόμβους, δεν εξαρτάται από το m και δεν επηρεάζει τον αριθμό των δεδομένων που αποθηκεύονται στο εξωτερικό αρχείο.

Ακολουθήστε και υλοποιήστε τα παρακάτω βήματα:

- Εισαγωγή τυχαίου κλειδιού (2 μονάδες)
- Διαγραφής τυχαίου κλειδιού (2 μονάδες)
- Αναζήτησης τυχαίας τιμής κλειδιού (2 μονάδες)
- Αναζήτηση εύρους τιμών κλειδιών (range queries) στο διάστημα K τιμών. Δηλαδή για το διάστημα $[K1, K2]$ ισχύει ότι $K = K2 - K1$ (2 μονάδες)
- Απόδοση και τεκμηρίωση αποτελεσμάτων (2 μονάδες)

Να κάνετε εισαγωγή $n = 10^5$ μοναδικών κλειδιών με τιμές 1 έως 10^6 . Συμπληρώστε τις τιμές στον παρακάτω πίνακα:

1. Μέσος αριθμός προσβάσεων δίσκου ανά εισαγωγή. Αφού κάνετε εισαγωγή όλων των κλειδιών, κάντε επιπλέον 20 εισαγωγές και μετρήστε τον μέσο αριθμό προσβάσεων δίσκου.
2. Μέσος αριθμός προσβάσεων δίσκου ανά αναζήτηση και διαγραφή. Δηλαδή για 20 αναζητήσεις και για 20 διαγραφές στοιχείων μετρήστε τον μέσο αριθμό προσβάσεων δίσκου.
3. Για 20 αναζητήσεις εύρους τιμών (δηλαδή για διαφορετικά K_1 , K_2 όπου $K=K_2-K_1$, για $K=10$ και 1000) μετρήστε τον μέσο αριθμό προσβάσεων δίσκου.

Μέσος αριθμός προσβάσεων δίσκου ανά εισαγωγή	Μέσος αριθμός προσβάσεων δίσκου ανά τυχαία αναζήτηση	Μέσος αριθμός προσβάσεων δίσκου ανά διαγραφή	Μέσος αριθμός προσβάσεων δίσκου για διάστημα K τιμών ($K=10$)	Μέσος αριθμός προσβάσεων δίσκου για διάστημα K τιμών ($K=1000$)

Σχολιάστε (ξεχωριστό αρχείο κειμένου) την απόδοση όλων των μεθόδων και προσπαθήστε να δικαιολογήσετε την απόδοση κάθε μεθόδου. **Δώστε ιδιαίτερη βαρύτητα στην ερμηνεία των αποτελεσμάτων και μην δείξετε μόνο τις τιμές.**

Συστάσεις:

- Δείτε <http://people.cs.vt.edu/~shaffer/Book/>, άλλο βιβλίο ή το WWW.
- Οι περισσότεροι κώδικες που θα βρείτε αφορούν υλοποίηση στην κεντρική μνήμη, ενώ η άσκηση ζητά υλοποίηση στον δίσκο.
- Οι μεταβλητές και σταθερές δεικτών έχουν τύπο integer (4 bytes).
- Αρχικά να κάνετε την υλοποίηση στην κεντρική μνήμη και χωρίς πεδίο δεδομένων. Στην συνέχεια μπορείτε να την μετατρέψετε για το δίσκο.
- Η υλοποίηση που ενδεχομένως θα βρείτε στο GitHub (log structured B+tree) δεν είναι ίδια με την άσκηση.
- Μην προχωρήσετε στην υλοποίηση πριν καταλάβετε καλά την λειτουργία του B+-tree και τις διαφορές από το B-tree.
- **Η υλοποίηση πρέπει να είναι B+tree και όχι B-tree. Η ασκήσεις με υλοποίηση B-tree δεν βαθμολογούνται.**
- Δεν ενδιαφέρει το περιεχόμενο του πεδίου δεδομένων (μπορεί να είναι οτιδήποτε)

Παραδοτέα: Ένα συμπιεσμένο zip αρχείο που περιέχει:

- Μία έκθεση 2 σελίδες το πολύ με τα αποτελέσματα που σας ζητούνται δηλαδή. **Δώστε ιδιαίτερη βαρύτητα στην ερμηνεία των αποτελεσμάτων** και μην δείξετε μόνο τις τιμές.
- Μία έκθεση που περιγράφει σε 1-2 σελίδες πώς φτιάχτηκε ο κώδικας (δηλ. για κάθε ερώτημα ποια είναι η γενική ιδέα της λύσης σε 3-4 προτάσεις), υπάρχουν σαφείς οδηγίες μετάφρασης από compiler και εκτέλεσης, τι λάθη έχει (αν έχει, περιπτώσεις που δεν δουλεύει το πρόγραμμα, ή περιπτώσεις που κάνει περισσότερα από όσα σας ζητεί η άσκηση, τι χρησιμοποιήσατε από έτοιμα προγράμματα ή πηγές

πληροφόρησης. Υποδείξετε ακόμα και πηγές στο WWW όπως Wikipedia ή Stackoverflow (πλήρης διεύθυνση σχετικών σελίδων).

- Ο κώδικας περιέχει συνοπτικά σχόλια που εξηγούν την υλοποίηση. Προσθέστε σχόλια σε μορφή javadoc στην αρχή της κάθε κλάσης και κάθε μεθόδου. Επίσης javadoc σχόλια πριν από κάθε member variable των κλάσεων. Και όπου απαιτείται μέσα στον κώδικά σας.
- Εκτός των παραπάνω, οι ασκήσεις βαθμολογούνται με άριστα εφόσον:
 - Το zip είναι πλήρες.
 - Οι κώδικες περνούν από compiler και εκτελούνται κανονικά και σωστά σε windows ή Linux περιβάλλον (Προσοχή: Θα πρέπει να κάνετε χρήση σχετικών directory paths και όχι απόλυτων, τα οποία ισχύουν μόνο για τους υπολογιστές σας).
- Οι ασκήσεις υποβάλλονται ηλεκτρονικά στον ιστοχώρο του μαθήματος και όχι με e-mail.
- Οι αντιγραφές (ακόμη και μέρους της υλοποίησης) μηδενίζονται.