

3^ο Σύνολο Ασκήσεων

Καταληκτική Ημερομηνία Παράδοσης: Δευτέρα 12 Ιανουαρίου 2015 (μέχρι τις 5μμ, στο γραφείο μου)
Το σύνολο θα λυθεί στη διάλεξη της Τρίτης 13 Ιανουαρίου – κατά συνέπεια δε θα γίνουν δεκτές αργοπορημένες παραδόσεις

Θεματική Ενότητα: Ευρετήρια

Για γενικές οδηγίες σχετικά με τις ασκήσεις, συμβουλευτείτε την ιστοσελίδα του μαθήματος.

Άσκηση 1

Θεωρείστε ένα ευρετήριο επεκτατού κατακερματισμού, όπου κάθε κάδος (bucket/block) μπορεί να χωρέσει έως 3 εγγραφές.

- (i) Θεωρείστε ότι κάποια στιγμή ο κατάλογος του ευρετηρίου έχει ολικό βάθος 9. Ποιος είναι ο μεγαλύτερος αριθμός θέσεων του καταλόγου που μπορεί να δείχνουν στον ίδιο κάδο και πόσες εγγραφές έχει τότε το ευρετήριο;
- (ii) Θεωρείστε ότι κάποια στιγμή το ευρετήριο έχει 200 κάδους. Ποιο είναι το μέγεθος του μικρότερου δυνατού καταλόγου;

Άσκηση 2

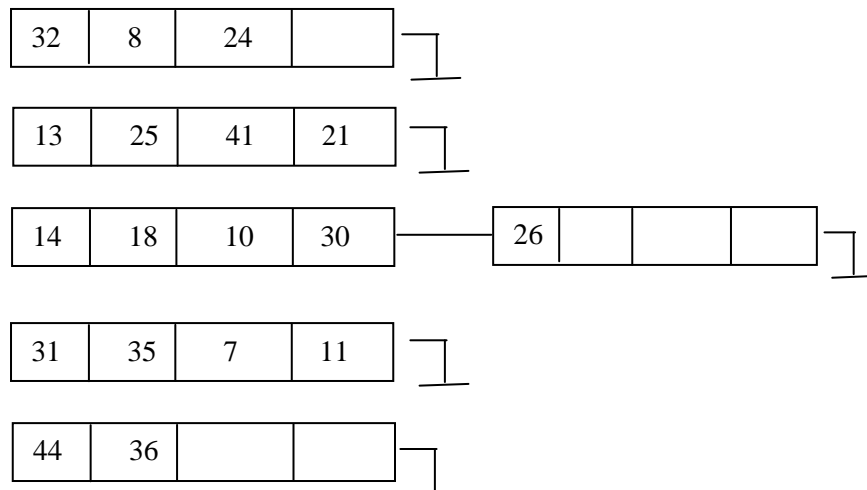
Θεωρείστε το ευρετήριο γραμμικού κατακερματισμού της Εικόνας 1, όπου υπάρχουν 5 κάδοι (και 1 κάδος υπερχείλισης) και ο επόμενος προς διάσπαση κάδος είναι ο 1 (δηλαδή, ο δεύτερος κάδος). Κάθε κάδος χωρά 4 εγγραφές.

- (α) Εισάγετε στο ευρετήριο το 4 και μετά το 15 και δώστε το αποτέλεσμα μετά από κάθε εισαγωγή.
- (β) Εισάγετε στο αρχικό ευρετήριο της Εικόνας 1 το 2 το 15 και δώστε το αποτέλεσμα μετά από κάθε εισαγωγή.
- (γ) Ποια είναι η μικρότερη αριθμητικά τιμή της οποίας η εισαγωγή στο ευρετήριο της Εικόνας 1 μπορεί να οδηγήσει στην υπερχείλιση ενός κάδου (δηλαδή, σε ένα κάδο με λίστα υπερχείλισης);
- (δ) Υποθέστε ότι μετά από έναν αριθμό από εισαγωγές, υπάρχουν 25 κάδοι (χωρίς του κάδους υπερχείλισης). Ποιος θα είναι ο επόμενος κάδος προς διάσπαση;

$$h_0(k) = k \bmod 4$$

$$h_1(k) = k \bmod 8$$

$n = 1$
(επόμενος προς
διάσπαση κάδος)



Εικόνα 1. Ευρετήριο για την Άσκηση 2.

Άσκηση 3

Θεωρείστε μια σχέση $R(A, B, C)$ αποθηκευμένη σε ένα αρχείο διατεταγμένο ως προς το γνώρισμα A . Το A είναι κλειδί για τη σχέση. Γνωρίζετε ότι το 20% των αιτημάτων προς τη βάση δεδομένων είναι αιτήματα εισαγωγής ή διαγραφής δεδομένων της μορφής **delete from R** ή **insert into R values t**, όπου t μια πλειάδα. Το υπόλοιπο 80% είναι ερωτήματα της μορφής **select from R where A = a**, όπου a μια τιμή από το πεδίο ορισμού του γνωρίσματος A .

Επίσης, γνωρίζετε ότι η σχέση R έχει 1.000.000 εγγραφές, το μέγεθος μιας πλειάδας της R είναι 200 bytes και το μέγεθος του block (σελίδας) είναι 2.048 bytes. Το γνώρισμα A καταλαμβάνει 12 bytes και ένας δείκτης (ευρετηρίου, αρχείου, δέντρου) 8 bytes.

Για την επεξεργασία των αιτημάτων, σας δίνονται οι παρακάτω τρεις επιλογές:

- (1) Χρήση ενός ευρετηρίου ενός επιπέδου στο γνώρισμα A .
- (2) Χρήση ενός B+-δέντρου στο γνώρισμα A . Θεωρίστε το δέντρο με το μικρότερο δυνατό ύψος. Υποθέστε επίσης ότι υπερχείλιση (αντίστοιχα, υποχείλιση) δημιουργεί το 10% των εισαγωγών (αντίστοιχα, διαγραφών) και ότι αυτή οδηγεί σε αύξηση (αντίστοιχα, μείωση) του ύψους. Για ευκολία στον υπολογισμό του κόστους κάθε αιτήματος, θεωρίστε ότι στην αρχή κάθε αιτήματος το δέντρο έχει το ίδιο ύψος, συγκεκριμένα, το μικρότερο ύψος για τις εισαγωγές και για τα *select* ερωτήματα και το μικρότερο ύψος + 1 για τις διαγραφές.
- (3) Απευθείας προσπέλαση στο αρχείο χωρίς χρήση ευρετηρίου.

Εκτιμήστε το κόστος επεξεργασίας για κάθε μια από τις παραπάνω επιλογές και εξηγήστε ποια είναι η πιο κατάλληλη. Αγνοήστε το κόστος για την αρχική δημιουργία των ευρετηρίων. Για ευκολία, στον υπολογισμό του κόστους κάθε αιτήματος αγνοήστε τις αυξομειώσεις στο πλήθος των εγγραφών και θεωρήστε το ίδιο πλήθος εγγραφών στην αρχή κάθε αιτήματος, δηλαδή 1.000.000 εγγραφές.