

3^ο Σύνολο Ασκήσεων

Καταληκτική Ημερομηνία Παράδοσης: Τρίτη 10 Ιανουαρίου, 2017

Θεματική Ενότητα: Αποθήκευση, Ευρετήρια, Εισαγωγή στην επεξεργασία ερωτήσεων.

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

Άσκηση 1 [B-δέντρα, B+-δέντρα]

- (α) Θεωρήστε ένα B-δέντρο τάξης 3 στο οποίο εισάγετε τις τιμές 1, 2, ... Ποια είναι η πρώτη τιμή που θα οδηγήσει σε αύξηση του ύψους του δέντρου από 2 σε 3; Δώστε το δέντρο μετά από αυτήν την εισαγωγή.
- (β) Επαναλάβετε το ερώτημα (α) για ένα B+-δέντρο τάξης εσωτερικών κόμβων 3 και φύλλων 2.

Άσκηση 2 [Επεικτός κατακερματισμός]

Θεωρείστε επεικτός κατακερματισμό ως ευρετήριο στο γνώρισμα A μιας σχέσης R(A, B, C, D, E, F) που είναι αποθηκευμένη σε ένα μη ταξινομημένο αρχείο. Το A είναι υποψήφιο κλειδί. Το μέγεθος του γνωρίσματος A είναι 16 bytes και του δείκτη σε αρχείο 32 bytes. Το μέγεθος σελίδας (block) είναι 1024 bytes. Η σχέση έχει 100.000 εγγραφές και 10.000 blocks (σελίδες).

- (α) Ποιο είναι το μικρότερο ολικό βάθος που χρειάζεται για να δεικτοδοτηθεί το αρχείο;
- (β) Πόσες σελίδες το πολύ χρειάζεται να προσπελαστούν για να απαντηθεί η ερώτηση $\sigma_{A=5000}(R)$ και γιατί.
- (γ) Εξηγήστε για ποιες από τις παρακάτω ερωτήσεις συμφέρει να χρησιμοποιηθείτε το ευρετήριο και γιατί.

(i) $\sigma_{A=50000} \text{ AND } B=100(R)$

(ii) $\sigma_{A=50000} \text{ OR } B=100(R)$

(iii) $\sigma_{A>50000} \text{ AND } B=100(R)$

(iv) $\pi_A(R)$

(v) $R * S$, όπου * η φυσική συνένωση και S μια σχέση με σχήμα S(A, G, H)

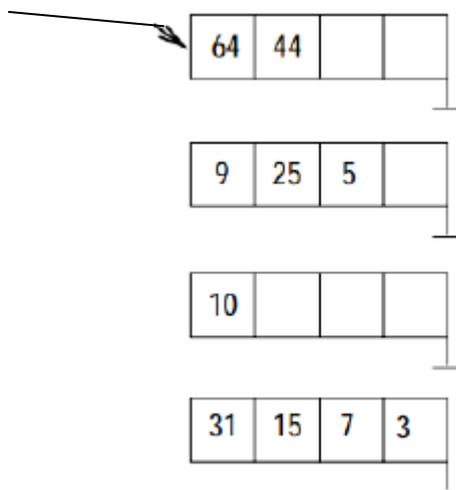
(vi) $R * S$, όπου * η φυσική συνένωση και S μια σχέση με σχήμα S(B, G, H)

Άσκηση 3 [Γραμμικός Κατακερματισμός]

Θεωρείστε τον πίνακα γραμμικού κατακερματισμού της Εικόνας 1. Η αρχική συνάρτηση κατακερματισμού είναι $k \bmod 4$ και σε κάθε κάδο χωρούν 4 εγγραφές.

- (α) Δώστε τη μικρότερη τιμή που η εισαγωγή της θα οδηγήσει σε διάσπαση κάδου. Δώστε τον πίνακα μετά την εισαγωγή αυτής της τιμής.
- (β) Δώστε τον πίνακα μετά την εισαγωγή των 17, 21 και 19 στον πίνακα της Εικόνας 1.
- (γ) Εισάγετε τις τιμές της Εικόνας 1 σε ένα πίνακα κατακερματισμού με συνάρτηση κατακερματισμού $k \bmod 3$ όπου σε κάθε κάδο χωρούν 2 εγγραφές με τη σειρά που εμφανίζονται, δηλαδή 64, 44, 9, 25, 5, 10, 31, 15, 7, 3. Εκτός από τον τελικό πίνακα, δώστε και τους πίνακες μετά από κάθε διάσπαση.

Επόμενος κάδος για διάσπαση



Εικόνα 1: Πίνακας γραμμικού κατακερματισμού

Άσκηση 4 [Ευρετήρια]

Έστω ένας πίνακας (σχέση) CITY(Name, Population, Country) ο οποίος έχει πληροφορία για 100.000 πόλεις που είναι ομοιόμορφα κατανεμημένες σε 1.000 χώρες. Ο πίνακας είναι αποθηκευμένος σε ένα διατεταγμένο αρχείο ως προς το γνώρισμα Name. Τα γνώρισμα Name και Country έχουν μέγεθος 16 bytes, το γνώρισμα Population 32 bytes και ένα block (σελίδα) 2048 bytes. Υποθέστε ότι όλοι οι δείκτες έχουν μέγεθος 16 bytes.

Θεωρείστε ότι υπάρχει ένα (μη δεντρικό) ευρετήριο στο γνώρισμα Country και ένα πυκνό B+-δέντρο (χωρίς ενδιάμεσο επίπεδο) στο γνώρισμα Population.

Έστω η ερώτηση

```
SELECT * FROM CITY WHERE Country = 'Greece' AND Population > 50.000.
```

Θεωρείστε τους παρακάτω διαφορετικούς τρόπους για να υπολογιστεί η ερώτηση

(α) με χρήση του B+-ευρετήριο στο γνώρισμα Population

(β) με χρήση του ευρετηρίου στο γνώρισμα Country

(γ) με χρήση και των δύο ευρετηρίων

(δ) χωρίς χρήση ευρετηρίου.

Υπολογίστε το κόστος καθενός από τα παραπάνω. Υποθέστε ότι υπάρχουν 100 ελληνικές πόλεις και ότι υπάρχουν συνολικά 200 πόλεις με πληθυσμό πάνω από 50.000 κατοίκους από τις οποίες ελληνικές είναι 20.