

3^ο Σύνολο Ασκήσεων

Θεματική Ενότητα: Αποθήκευση, ευρετήρια, εισαγωγή στην επεξεργασία ερωτήσεων.

Θα λυθεί στο μάθημα της Τρίτης 8/1/2019

Άσκηση 1

- (α) Θεωρείστε 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 Έστω ένας πίνακας R(A, B, C, D, E) με 720.000 γραμμές (πλειάδες). Το γνώρισμα B παίρνει 12.000 διαφορετικές τιμές ομοιόμορφα κατανομημένες στις γραμμές του πίνακα. Υποθέστε ότι όλα τα γνώρισμα έχουν μέγεθος 16 bytes, όλοι οι δείκτες 32 bytes και ένα block (σελίδα) 1024 bytes. Ο πίνακας είναι αποθηκευμένος σε ένα διατεταγμένο αρχείο ως προς το γνώρισμα B. Υπάρχει επίσης ένα ευρετήριο τύπου B+-δέντρο στο γνώρισμα B και ένα (μη δεντρικό) ευρετήριο στο γνώρισμα A. Θεωρείστε μη εκτεινόμενη οργάνωση.

- (α) Πόσος χώρος χάνεται εξαιτίας της μη-εκτεινόμενης αποθήκευσης του αρχείου;
- (β) Έστω η ερώτηση $\sigma_{A>a}(R)$, όπου a μια τιμή από το πεδίο ορισμού του A. Υποθέστε ότι το αποτέλεσμα έχει 10 πλειάδες. Εκτιμήστε το κόστος υπολογισμού της ερώτησης με και χωρίς χρήση του κατάλληλου ευρετηρίου.
- (γ) Έστω η ερώτηση $\sigma_{B=b}(R)$, όπου b μια τιμή από το πεδίο ορισμού του B. Εκτιμήστε το κόστος υπολογισμού της ερώτησης με και χωρίς χρήση του κατάλληλου ευρετηρίου.
- (δ) Έστω η ερώτηση $\sigma_{A>a \text{ AND } B=b}(R)$, όπου a και b όπως στα προηγούμενα ερωτήματα. Χρησιμοποιήστε τις απαντήσεις σας στα προηγούμενα ερωτήματα (β) και (γ) για να προτείνετε έναν κατάλληλο τρόπο υπολογισμού της ερώτησης καθώς και μια εκτίμηση του κόστους αυτού του υπολογισμού.
- (ε) Θέλουμε να κατασκευάσουμε ένα ευρετήριο επεκτατού κατακερματισμού στο γνώρισμα E. Θεωρείστε ότι το γνώρισμα E είναι επίσης υποψήφιο κλειδί. Ποιο θα είναι το μικρότερο ολικό βάθος για ένα τέτοιο ευρετήριο;

Άσκηση 3 Θεωρείστε δύο σχέσεις (πίνακες) R(A, B, C) και S(D, E). Οι δύο πίνακες είναι αποθηκευμένοι σε αρχεία σωρού (μη διατεταγμένα αρχεία). Ο πίνακας R έχει 800.000 πλειάδες και καταλαμβάνει 3.000 σελίδες (blocks) ενώ ο πίνακας S έχει 500.000 πλειάδες και καταλαμβάνει 1.500 σελίδες (blocks). Το γνώρισμα C μπορεί να πάρει 200 διαφορετικές τιμές. Έστω η παρακάτω SQL ερώτηση.

```
SELECT R.A  
FROM R, S  
WHERE R.C = S.D
```

- (α) Δώστε έναν κατάλληλο αλγόριθμο υπολογισμού της ερώτησης και μια εκτίμηση του κόστους του.
- (β) Υποθέστε ότι υπάρχει ένα ευρετήριο κατακερματισμού στο C. Δώστε έναν αλγόριθμο υπολογισμού της ερώτησης που να χρησιμοποιεί αυτό το ευρετήριο και δώστε μια εκτίμηση του κόστους του