

### 3<sup>ο</sup> (Προαιρετικό) Σύνολο Ασκήσεων

Καταληκτική Ημερομηνία Παράδοσης: **Παρασκευή 10 Ιανουαρίου 2020**, 5μμ (στο κουτί έξω από το γραφείο μου)

**Θεματική Ενότητα:** Αποθήκευση, Ευρετήρια, Επεξεργασία Ερωτήσεων

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

Ο βαθμός σε αυτό το σύνολο θα μετρήσει θετικά στο βαθμό ασκήσεων για το μάθημα.

#### Άσκηση 1

Έστω ένας πίνακας (σχέση) CITY(Name, Population, Country) ο οποίος έχει πληροφορία για 100.000 πόλεις που είναι ομοιόμορφα κατανεμημένες σε 1.000 χώρες (δηλαδή 100 πόλεις ανά χώρα). Ο πίνακας είναι αποθηκευμένος σε ένα διατεταγμένο αρχείο ως προς το γνώρισμα Name. Τα γνωρίσματα 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 στο γνώρισμα Population. **Δώστε μια εκτίμηση του κόστους για καθένα από τα παρακάτω.**

Έστω η ερώτηση

```
SELECT * FROM CITY WHERE Country = 'Greece' AND Population BETWEEN 50000 AND 60000;
```

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.

#### Άσκηση 2

(α) Θεωρείστε γραμμικά κατακερματισμό με συνάρτηση κατακερματισμού  $k \bmod 5$  όπου σε κάθε κάδο χωρούν 3 εγγραφές. Δώστε μια ακολουθία τιμών που να οδηγεί σε κάδο υπερχείλισης με 3 τιμές. **(Υποθέστε ότι γίνεται διάσπαση κάδου την πρώτη φορά που γίνεται εισαγωγή τιμής σε γεμάτο κάδο, δηλαδή όταν έχουμε υπερχείλιση.)**

(β) Θεωρείστε επεκτατό κατακερματισμό ως ευρετήριο σε ένα γνώρισμα A μιας σχέσης R που είναι αποθηκευμένη σε ένα μη ταξινομημένο αρχείο. Το A είναι υποψήφιο κλειδί και η σχέση έχει 1.000.000 εγγραφές

(i) Υποθέστε ότι σε κάθε κάδο χωρούν 50 τιμές. Ποιο είναι το μικρότερο ολικό βάθος που χρειάζεται για να δεικτοδοτηθεί το αρχείο και σε ποια περίπτωση είναι αυτό εφικτό. Ποιο είναι το μέγεθος (σε block) του ευρετηρίου.

(ii) Πόσα blocks το πολύ χρειάζεται να προσπελαστούν για να απαντηθεί η ερώτηση  $\sigma_{A=500}(R)$  και γιατί.

(iii) Εξηγήστε για ποιες από τις παρακάτω ερωτήσεις συμφέρει να χρησιμοποιήσετε το ευρετήριο και γιατί.

(i)  $\sigma_{A=500000 \text{ AND } B=100}(R)$

(ii)  $\sigma_{A=500000 \text{ OR } B=100}(R)$

(iii)  $\sigma_{A>500000 \text{ AND } B=100}(R)$

(iv)  $\pi_A(R)$

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

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