

## ΤΕΛΙΚΟ ΔΙΑΓΩΝΙΣΜΑ ΕΞΕΤΑΣΤΙΚΗ ΙΑΝΟΥΑΡΙΟΥ

Διάρκεια Εξέτασης: 3 ώρες

Πριν απαντήσετε, διαβάστε την εκφώνηση κάθε θέματος προσεκτικά. Εξηγήστε τις απαντήσεις σας με σαφήνεια.

### Θέμα 1 [Μονάδες 15]

Θέλουμε να κατασκευάσουμε μια βάση δεδομένων για μια αλυσίδα καταστημάτων ρούχων με πληροφορίες για τα ρούχα που διαθέτει κάθε κατάστημα.

- Κάθε κατάστημα χαρακτηρίζεται μοναδικά από το όνομα και την πόλη στην οποία βρίσκεται (το συνδυασμό τους). Διατηρούμε επίσης το τηλέφωνο και την ταχυδρομική διεύθυνση του κάθε καταστήματος.
- Κάθε ρούχο έχει ένα αναγνωριστικό barcode. Επίσης, διατηρούμε την τιμή, σχεδιαστή, χρώμα και μέγεθός του. Ένα ρούχο με το ίδιο barcode έχει την ίδια τιμή και σχεδιαστή. Ένα ρούχο με το ίδιο barcode μπορεί να διατίθεται σε διάφορα μεγέθη και χρώματα.
- Ένα κατάστημα διαθέτει ρούχα (τουλάχιστον ένα). Μπορεί να υπάρχουν ρούχα που δε διατίθενται από κανένα κατάστημα. Ένα κατάστημα μπορεί να διαθέτει παραπάνω από ένα ρούχο με το ίδιο barcode, χρώμα και μέγεθος, δηλαδή, πολλά κομμάτια του ρούχου με το συγκεκριμένο barcode, χρώμα και μέγεθος. Θέλουμε να αποθηκεύσουμε και τον αριθμό αυτών των κομματιών. Δηλαδή, θέλουμε να ξέρουμε πόσα κομμάτια ενός ρούχου με ένα συγκεκριμένο barcode, χρώμα και μέγεθος διαθέτει κάθε κατάστημα.

(α) Δώστε ένα κατάλληλο Μοντέλο Οντοτήτων/Συσχετίσεων για την παραπάνω βάση δεδομένων.

(β) Δώστε ένα κατάλληλο Σχεσιακό Μοντέλο Οντοτήτων/Συσχετίσεων για την παραπάνω βάση δεδομένων.

### Θέμα 2 [Μονάδες 35]

Έστω το παρακάτω σχεσιακό σχήμα μιας βάσης δεδομένων που αφορά πόλεις. Το γνώρισμα Όνομα-Χώρας της σχέσης ΠΟΛΗ είναι ξένο κλειδί που αναφέρεται στο αντίστοιχο γνώρισμα της σχέσης ΧΩΡΑ.

ΠΟΛΗ(Όνομα, Πληθυσμός, Όνομα-Χώρας)

ΧΩΡΑ(Όνομα-Χώρας, Γλώσσα)

(α) Δώστε μια ερώτηση σε σχεσιακή άλγεβρα που να δίνει όλες τις πόλεις που έχουν πληθυσμό μεγαλύτερο των 100.000 και βρίσκονται σε χώρες που μια από τις γλώσσες τους είναι τα Ελληνικά.

(β) Δώστε μια ερώτηση σε σχεσιακή άλγεβρα που να δίνει όλες τις χώρες που έχουν τουλάχιστον δύο πόλεις με πληθυσμό άνω των 500.000.

(γ) Δώστε μια ερώτηση σε σχεσιακή άλγεβρα που να δίνει την πόλη με το μεγαλύτερο πληθυσμό (χωρίς χρήση συναθροιστικών συναρτήσεων).

(δ) Δώστε μια ερώτηση σε SQL, που να δίνει ως αποτέλεσμα για κάθε χώρα που έχει τουλάχιστον 5 πόλεις, το μέσο πληθυσμό αυτών των πόλεων.

(ε) Ορίστε 2 όψεις σε SQL, μία που να μπορεί να τροποποιηθεί (updateable) και μια που να μη μπορεί να τροποποιηθεί. Εξηγήστε.

(στ) Διατυπώστε σε SQL ένα παράδειγμα διαγραφής και ένα παράδειγμα εισαγωγής που το καθένα να μπορεί να οδηγήσει σε παραβίαση του ξένου κλειδιού.

(ζ) Σε αυτό το ερώτημα θεωρείστε το σχήμα

$R(A, B, C)$

$S(A, B, E)$

Διατυπώστε την ερώτηση  $\pi_{A,B}(\sigma_{A > 5}(R)) - \pi_{A,B}(S)$  σε σχεσιακό λογισμό πλειάδων.

### Θέμα 3 [Μονάδες 10]

Έστω ότι στο σχεσιακό σχήμα  $R = (A, B, C, D, E)$  ισχύει το σύνολο των συναρτησιακών εξαρτήσεων  $F = \{AB \rightarrow D, C \rightarrow E, DE \rightarrow B, D \rightarrow C\}$ .

(α) Εξετάστε αν μπορούμε να απλοποιήσουμε το σύνολο  $F$  αφαιρώντας ένα γνώρισμα από το αριστερό μέρος κάποιων από τις εξαρτήσεις του. Αν ναι, ποιο είναι το ισοδύναμο σύνολο συναρτησιακών εξαρτήσεων που προκύπτει;

(β) Ποιός είναι ο ελάχιστος αριθμός συναρτησιακών εξαρτήσεων που πρέπει να προσθέσουμε στο σύνολο F ώστε το D να είναι κλειδί; Δώστε αυτήν ή αυτές τις εξαρτήσεις.

**Θέμα 4 [Μονάδες 20]**

Έστω ένα B+-δέντρο τάξης 16 για τους εσωτερικούς κόμβους και 19 για τα φύλλα που χρησιμοποιείται ως ευρετήριο πάνω σε ένα γνώρισμα A. Το γνώρισμα A είναι κλειδί μιας σχέσης R που είναι αποθηκευμένη σε κάποιο αρχείο. Υποθέστε ότι το δέντρο έχει 4 επίπεδα (συμπεριλαμβανομένου του επιπέδου της ρίζας και των φύλλων) και ότι είναι όσο πιο άδειο (λιγότερο γεμάτο) γίνεται. Ο παράγοντας ομαδοποίησης (blocking factor) για το αρχείο δεδομένων είναι 230.

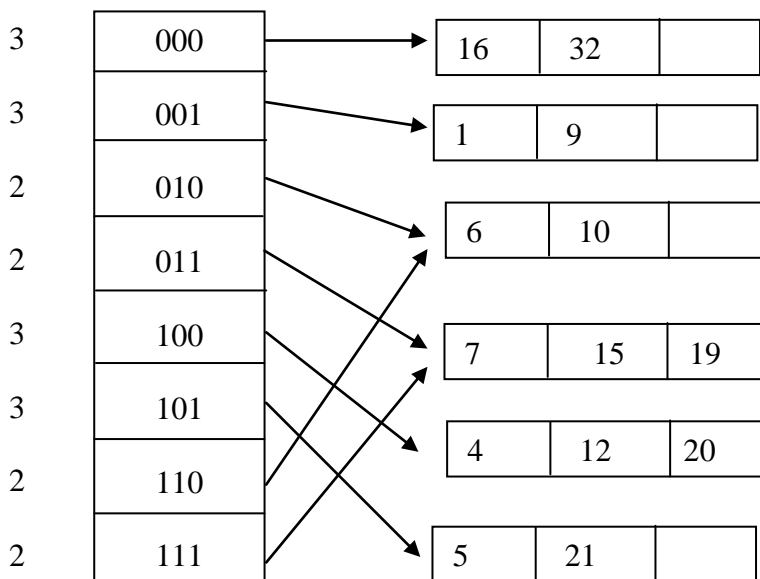
- (α) Πόσα blocks καταλαμβάνει το ευρετήριο; Εξηγήστε την απάντησή σας.
- (β) Εκτιμήστε τον αριθμό των blocks που καταλαμβάνει το αρχείο αν το A είναι πεδίο διάταξης του αρχείου; Εξηγήστε την απάντησή σας.
- (γ) Εκτιμήστε τον αριθμό των blocks που καταλαμβάνει το αρχείο αν το A δεν είναι πεδίο διάταξης του αρχείου; Εξηγήστε την απάντησή σας.
- (δ) Υπολογίστε το κόστος μιας ερώτησης διαστήματος στο γνώρισμα A (δηλαδή, μια ερώτησης της μορφής:  $\sigma (x_1 \leq A \text{ and } A \leq x_2)$  (R)) όταν αυτή η ερώτηση ικανοποιείται από το 10% των εγγραφών για την περίπτωση (γ). Δώστε με απλά λόγια το σχετικό αλγόριθμο.

**Θέμα 5 [Μονάδες 10]**

Θεωρείστε το ευρετήριο επεκτατού κατακερματισμού της Εικόνας 1.

- (α) Ποιος είναι ο μικρότερος αριθμός τιμών που πρέπει να διαγραφούν από το ευρετήριο της Εικόνας 1 για να μειωθεί το ολικό βάθος του; Δώστε ένα παράδειγμα τέτοιων τιμών (ή τιμές, αν αρκεί μία) και το ευρετήριο ολικού βάθους 2 που προκύπτει.
- (β) Ποιος είναι ο μικρότερος αριθμός τιμών των οποίων η εισαγωγή στο ευρετήριο της Εικόνας 1 μπορεί να οδηγήσει σε αύξηση του ολικού βάθους του; Δώστε ένα παράδειγμα τέτοιων τιμών (ή τιμές, αν αρκεί μία) και το ευρετήριο ολικού βάθους 4 που προκύπτει.

**Τοπικό Βάθος**



Εικόνα 1. Επεκτατός Δυναμικός Κατακερματισμός

**Θέμα 6 [Μονάδες 10]**

- (α) Έστω η ερώτηση  $\sigma (B = b \text{ OR } A = a)$  (R) όπου A και B είναι γνώρισμα της σχέσης R. Η σχέση R είναι αποθηκευμένη σε αρχείο διατεταγμένο ως προς το A και υπάρχει ευρετήριο στο A. Εξηγήστε γιατί η χρήση του ευρετηρίου δε βελτιώνει το κόστος αυτής της ερώτησης.
- (β) Θεωρείστε ένα ευρετήριο γραμμικού κατακερματισμού. Σε ποια περίπτωση χρησιμοποιούμε στην αναζήτηση μιας τιμής και τις δύο συναρτήσεις και σε ποιες αρχεί η χρήση της μίας συνάρτησης;
- (γ) Πόσα μη πυκνά ευρετήρια μπορούμε να έχουμε σε ένα αρχείο δεδομένων και γιατί;