

## ΤΕΛΙΚΟ ΔΙΑΓΩΝΙΣΜΑ ΕΞΕΤΑΣΤΙΚΗ ΣΕΠΤΕΜΒΡΙΟΥ

Διάρκεια Εξέτασης: 3 ώρες

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

### Θέμα 1 [Μονάδες 18]

Θέλουμε να κατασκευάσουμε μια βάση δεδομένων στην οποία θα διατηρούμε πληροφορίες για υποψηφιότητες ηθοποιών στα βραβεία Oscars. Συγκεκριμένα, στη βάση δεδομένων έχουμε τις παρακάτω πληροφορίες για ηθοποιούς, ταινίες και υποψηφιότητες. Για κάθε ηθοποιό, διατηρούμε το όνομά του (που είναι μοναδικό), την ημερομηνία γέννησης και την εθνικότητά του. Κάθε ηθοποιός μπορεί να παίζει παραπάνω από έναν ρόλο σε κάθε ταινία. Για κάθε ταινία έχουμε τον τίτλο της, το έτος παραγωγής της, το είδος της (πχ, κωμωδία, δράση) και τη διάρκειά της. Υποθέστε ότι μια ταινία προσδιορίζετε μοναδικά από τον τίτλο και το έτος παραγωγής της. Επίσης, για μια ταινία μπορεί να έχουμε παραπάνω από ένα είδη. Για κάθε υποψηφιότητα για ένα βραβείο, θέλουμε να ξέρουμε την κατηγορία του βραβείου (πχ, πρώτος γυναικείος ρόλος σε δραματική ταινία), το έτος, τον ηθοποιό που είναι υποψήφιος καθώς και την ταινία για την οποία είναι υποψήφιος. Διατηρούμε επίσης πληροφορία για το αν ο ηθοποιός κέρδισε ή όχι. Ένας ηθοποιός μπορεί να είναι υποψήφιος σε διαφορετικές κατηγορίες για τη συμμετοχή του σε διαφορετικές ταινίες το ίδιο έτος (πχ για πρώτο ρόλο σε μια ταινία και δεύτερο ρόλο σε κάποια άλλη), αλλά στην ίδια κατηγορία μπορεί να έχει το πολύ μια υποψηφιότητα ανά έτος. Επίσης, ένας ηθοποιός μπορεί να είναι πολλά έτη υποψήφιος στη ίδια ή άλλη κατηγορία. Τέλος, πολλοί ηθοποιοί της ίδιας ταινίας μπορεί να είναι υποψήφιοι στην ίδια ή σε διαφορετικές κατηγορίες.

(α) Σχεδιάστε ένα κατάλληλο μοντέλο Οντοτήτων/Συσχετίσεων

(β) Σχεδιάστε ένα κατάλληλο σχεσιακό μοντέλο.

(γ) Δώστε τις συναρτησιακές εξαρτήσεις που ισχύουν στο σχεσιακό μοντέλου του ερωτήματος (β) σύμφωνα με τη περιγραφή του προβλήματος και εξηγήστε σε ποια κανονική μορφή αυτό είναι.

### Θέμα 2 [Μονάδες 25]

Έστω το παρακάτω σχεσιακό σχήμα μιας βάσης δεδομένων που διατηρεί πληροφορία για τις ψήφους που πήραν πολιτικά κόμματα σε πόλεις.

CITY(Name, Population)

VOTES(City-Name, Party-Name, Vote-Number)

Η σχέση CITY έχει το όνομα (Name) και τον πληθυσμό (Population) μιας πόλης ενώ η σχέση VOTES τον αριθμό των ψήφων (Vote-Number) που πήρε ένα πολιτικό κόμμα (Party-Name) σε μια πόλη (City-Name).

(α) Δώστε μια ερώτηση σε *σχεσιακή άλγεβρα* που να δίνει ως αποτέλεσμα όλες τις πόλεις με πληθυσμό μεγαλύτερο του 100.000 στις οποίες και τα δυο κόμματα «YELLOW» και «PINK» πήραν το καθένα τουλάχιστον 5.000 ψήφους.

(β) Δώστε μια ερώτηση σε *σχεσιακή άλγεβρα* που να δίνει ως αποτέλεσμα τα ζεύγη κομμάτων που πήραν ακριβώς τον ίδιο αριθμό ψήφων σε δύο τουλάχιστον πόλεις.

(γ) Δώστε μια ερώτηση σε *σχεσιακή άλγεβρα* που να δίνει ως αποτέλεσμα τα κόμματα που πήραν σε καθεμία από τις πόλεις τουλάχιστον 200 ψήφους (δηλαδή, σε καμία από τις πόλεις δεν έχουν λιγότερες από 200 ψήφους).

(δ) Δώστε μια ερώτηση σε *SQL* που να δίνει ως αποτέλεσμα ζεύγη κομμάτων στα οποία ο συνολικός αριθμός ψήφων του πρώτου στις πόλεις με πληθυσμό κάτω του 100.000 είναι μεγαλύτερος από το συνολικό αριθμό ψήφων του δεύτερου στις πόλεις με πληθυσμό άνω του 100.000.

(ε) Δώστε μια ερώτηση σε *SQL*, που να δίνει ως αποτέλεσμα για κάθε κόμμα που πήρε ψήφους σε τουλάχιστον 10 πόλεις με πληθυσμό μικρότερο του 5.000, το συνολικό αριθμό ψήφων του σε αυτές.

### Θέμα 3 [Μονάδες 15]

Έστω δύο σχέσεις με σχήμα  $R(A, B)$  και  $S(C, D)$  και το παρακάτω στιγμιότυπό τους.

R		S	
A	B	A	C
1	6	1	1
4	4	3	6
3	2	4	1
9	6		
1	2		
4	3		

(α) Ποιο είναι το αποτέλεσμα των παρακάτω (εξηγήστε την απάντησή σας)

- (SELECT R.A from R) UNION (SELECT S.A from S)
  - (SELECT DISTINCT R.A from R) UNION ALL (SELECT S.A from S)
  - SELECT S.C, count(distinct R.A) from R, S where R.A > S.A group by S.C having sum(S.A) > 5
- (β) Διατυπώστε την ερώτηση (α)-(i) σε σχεσιακή άλγεβρα  
(γ) Διατυπώστε την ερώτηση (α)-(i) σε σχεσιακό λογισμό

### Θέμα 4 [Μονάδες 15]

(α) Έστω η σχέση  $R(A_1, A_2, A_3)$ . Υποθέστε ότι ισχύει το σύνολο συναρτησιακών εξαρτήσεων  $F = \{A_1 A_2 \rightarrow A_3\}$ .

- Δώστε όλες τις μη τετριμμένες συναρτησιακές εξαρτήσεις που **δεν** ανήκουν στο  $F^+$ .
- Δώστε ένα στιγμιότυπο στο οποίο ισχύουν μόνο οι συναρτήσεις της  $F^+$ , ενώ όλες οι άλλες συναρτησιακές του ερωτήματος (i) παραβιάζονται.

(β) Έστω η σχέση  $R(A, B, C, D, E)$  στην οποία ισχύει το σύνολο συναρτησιακών εξαρτήσεων  $F = \{AB \rightarrow C, BC \rightarrow D, CD \rightarrow E, DE \rightarrow A\}$ .

- Δώστε όλα τα υποψήφια κλειδιά
- Υπάρχουν συναρτησιακές εξαρτήσεις στο  $F$  που παραβιάζουν την BCNF; Αν ναι ποιες.
- Αν η απάντησή σας στο ερώτημα (ii) είναι θετική, δώστε μια διάσπαση της  $R$  σε σχέσεις που να είναι σε BCNF.

### Θέμα 5 [Μονάδες 27]

Έστω ένας πίνακας (σχέση)  $CITY(\underline{Name}, Population, Country)$  ο οποίος έχει πληροφορία για 800.000 πόλεις και 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) 1024 bytes. Υποθέστε ότι όλοι οι δείκτες έχουν μέγεθος 16 bytes. Θεωρείστε ότι υπάρχουν ευρετήρια σε όλα τα γνωρίσματα. Τα ευρετήρια στα  $Name$  και  $Country$  είναι απλά (μη δεντρικά) ευρετήρια, ενώ το ευρετήριο στο γνώρισμα  $Population$  είναι ένα  $B^+$ -δέντρο. Για τα παρακάτω ερωτήματα, υποθέστε για ευκολία ότι ο πίνακας έχει πληροφορία για 10.000 πόλεις από καθεμία από 80 χώρες και ότι υπάρχουν συνολικά 50.000 πόλεις με πληθυσμό πάνω από 1.000 κατοίκους από τις οποίες οι 8.000 είναι ελληνικές πόλεις.

(α) Για κάθε μία από τις παρακάτω ερωτήσεις, εξηγήστε αν συμφέρει ή όχι να χρησιμοποιηθεί το αντίστοιχο ευρετήριο, εκτιμώντας τα αντίστοιχα κόστη:

- SELECT \* FROM CITY WHERE Name = 'Ioannina'
- SELECT \* FROM CITY where Country = 'Greece'
- SELECT \* from CITY where Population > 1.000

(β) Έστω η ερώτηση SELECT \* FROM CITY WHERE Country = 'Greece' AND Population > 1.000. Περιγράψτε δυο διαφορετικούς τρόπους να υπολογιστεί η ερώτηση που να χρησιμοποιούν ευρετήρια και υπολογίστε το κόστος τους.