Πανεπιστημιο Ιωαννινων
Τμήμα Πληροφορικής

Θέματα Βάσεων Δεδομένων
Ακαδημαϊκό Έτος 2001-2001

Ερωτήσεις Ανεκεφαλαίωσης
Ημερομηνία Παράδοσης: 16/10/2001

Θέμα 1 [Μονάδες 15] Υποθέστε ότι σας έχουν προσλάβει σε ένα τμήμα "Επιστήμης Πουλερικών" και σας ζητούν να σχεδιάστε τη βάση δεδομένων τους. Το βασικό πρόβλημα είναι η αποθήκευση πληροφορίας σχετικά με μια σειρά από πειράματα πάνω στον τρόπο εκτροφής κοτόπουλων. Κάθε κοτόπουλο έχει έναν όνομα, ένα είδος, μια ημερομηνία γέννησης και ένα μοναδικό αριθμό που ονομάζεται "ID-κοτόπουλου". Τα πειράματα έχουν ένα όνομα, ένα μοναδικό αριθμό που ονομάζεται "ID-πειράματος" μια ημερομηνία έναρξης και μια ημερομηνία περάτωσης. Για κάθε κοτόπουλο σε κάθε πείραμα, πρέπει να καταγράφετε το βάρος του πριν και μετά το πείραμα. Κάθε κοτόπουλο συμμετέχει το πολύ σε ένα πείραμα άλλα σε κάθε πείραμα συμμετέχουν πολλά κοτόπουλα.

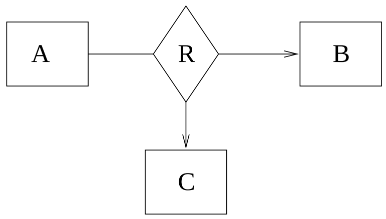
(α) Σχεδιάστε το διάγραμα Οντοτήτων/Συσχετίσεων (Ο/Σ) που αναπαριστά την παραπάνω πληροφορία.

(β) Μετατρέψτε το διάγραμμα σε σχεσιακό σχήμα. Δώστε δυο διαφορετικά σχεσιακά σχήματα, ένα κατάλληλο στην περίπτωση που σχεδόν όλα τα κοτόπουλα συμμετέχουν σε όλα τα πειράματα και ένα κατάλληλο για την περίπτωση που μόνο ένα πολύ μικρό ποσοστό συμμετέχει σε αυτά. Εξηγείστε.

(γ) Ποιες συναρτησιακές εξαρτήσεις ισχύουν στα σχεσιακά σχήματα του ερωτήματος (β); Είναι τα σχήματά σας σε BCNF;

(δ) Τώρα υποθέστε ότι ένα κοτόπουλο μπορεί να συμμετέχει σε παραπάνω από ένα πείραμα και ότι με κάθε πείραμα αντί για την ημερομηνία έναρξης και την ημερομηνία περάτωσής του, κρατάμε το διάστημα πραγματοποίησής του (πχ 5/1/2000-10/1/2000). Επιπλέον, σε κάθε διάστημα πραγματοποείται μόνο ένα πείραμα. Κάθε κοτόπουλο συμμετέχει σε ένα τουλάχιστον πείραμα.
(i) Πως τροποποιείται το διάγραμμα Ο/Σ σε αυτήν την περίπτωση;
(ii) Ποιες συναρτησιακές εξαρτήσεις ισχύουν τώρα;
(iii) Δώστε ένα σχεσιακό σχήμα για αυτήν την περίπτωση. Είναι σε BCNF;

Θέμα 2 [Μονάδες 10] Θεωρείστε το παρακάτω διάγραμμα Οντοτήτων/Συσχετίσεων. Αν η A έχει 100 οντότητες, η B 1000, και η C 10, ποιος είναι ο μέγιστος αριθμός τριάδων που μπορεί να έχει το σύνολο συσχετίσεων για την R; Εξηγείστε την απάντησή σας.



Θέμα 3 [Μονάδες 25] Έστω το παρακάτω σχεσιακό σχήμα:

ΠΡΟΤΙΜΑ($\underline{\text{Πότης}}$, $\underline{\text{Μπύρα}}$)

ΣΥΧΝΑΖΕΙ($\underline{\text{Πότης}}$, $\underline{\text{Μαγαζί}}$)

ΣΕΡΒΙΡΕΙ($\underline{\text{Μαγαζί}}$, $\underline{\text{Μπύρα}}$)

(α) Διατυπώστε σε σχεσιακή άλγεβρα (χωρίς χρήση συναθροιστικών συναρτήσεων) τις παρακάτω ερωτήσεις:

(i) Τα μαγαζιά που σερβίρουν τουλάχιστον δύο διαφορετικές μπύρες.

(ii) Τα μαγαζιά που σερβίρουν ακριβώς δύο διαφορετικές μπύρες.

(iii) Τα μαγαζιά που δε σερβίρουν καμία μπύρα που να αρέσει σε κάποιον πότη.

(iv) Τα μαγαζιά που σερβίρουν *όλες* τις μπύρες που αρέσουν σε όλους τους πότες.

(β) Διατυπώστε τις ερώτησεις (α).i και (α).ii σε SQL.

Θέμα 4 [Μονάδες 35]

(α) Δοθέντος ότι ένα σχήμα R ικανοποιεί τη συναρτησιακή εξάρτηση $X \to Y$, ποια από τα παρακάτω την ικανοποιούν επίσης (S είναι μια οποιαδήποτε άλλη σχέση). Εξηγείστε.

(i) $\sigma_P(R)$
(ii) $\pi_L(R)$
(iii) $R \cup S$
(iv) $R/\ S$
(v) $S/\ R$
(vi) $R \bowtie S$

(β) Έστω το σχήμα $R = (A, B, C, D, E)$ στο οποίο ισχύει το σύνολο συναρτησιακών εξαρτήσεων $F = \{A \to D,\ AB \to C,\ B \to E,\ D \to C,\ E \to A\ \}$. Θεωρείστε ότι διασπούμε την R και μια από τις σχέσεις που προκύπτουν είναι η $R_1(A, B, C)$. Υπολογίστε το υποψήφιο κλειδί (κλειδιά) της R_1.

(γ) Έστω το σχήμα $R = (A, B, C, D, E)$ στο οποίο ισχύει το σύνολο συναρτησιακών εξαρτήσεων $F = \{A \to B,\ B \to C,\ C \to A,\ D \to E,\ E \to D\ \}$. Δώστε ένα αντιπαράδειγμα που να δείχνει ότι η πλειότιμη εξάρτηση $AB \to\to CD$ δεν ισχύει.

(δ) Έστω μια σχέση R με γνωρίσματα A, B και C, η οποία διασπάται σε δύο σχέσεις $R_1(A, B)$ και $R_2(B, C)$. Έστω το στιγμιότυπο για την R_1: {(2, 3), (6, 3)}, και το στιγμιότυπο για την R_2: {(3, 7), (3, 9)}. Δώστε όλα τα δυνατά στιγμιότυπα της R από τα οποία μπορεί να προέκυψαν αυτά τα στιγμιότυπα. Γιατί αυτό το στιγμιότυπο δεν είναι μοναδικό;

Θέμα 5 [Μονάδες 30]

(α) Θεωρείστε δυο σχέσεις R και S (χωρίς διπλότιμα), που είναι αποθηκευμένες σε 1000 και 500 block δίσκου αντίστοιχα σε μη διατεταγμένα αρχεία.

(i) Θέλουμε να υπολογίσουμε το $S-R$ χρησιμοποιώντας έναν αλγόριθμο που διαβάζει κάθε block του R και του S στη μνήμη μόνο μια φορά. Περιγράψτε έναν αλγόριθμο που χρησμοποιεί τον ελάχιστο δυνατό αριθμό από καταχωρητές μνήμης (buffers). Ποιός είναι αυτός ο αριθμός;

(ii) Πως θα τροποιούσατε τον αλγόριθμό σας για να υπολογίζει το $R-S$ πάλι διαβάζοντας στη μνήμη κάθε block του R και του S μόνο μια φορά. χωρίς να χρησιμοποιείστε περισσότερους buffers από ότι στο ερώτημα (i).

(β) Θεωρείστε ότι έχουμε μια σχέση $R(a, b)$. 20% των ερωτήσεων σχετικά με την R έχουν την μορφή

select *
from R
where $a = x$; όπου το x είναι μια οποιαδήποτε σταθερά. Το υπόλοιπο 80% των ερωτήσεων έχει τη μορφή:

select *
from R
where $b = y$;

όπου το y είναι μια οποιαδήποτε σταθερά. Θέλουμε να κατασκευάσουμε μία "διαμοιραζόμενη συνάρτηση κατακερματισμού" όπου στον πίνακα κατακερματισμού κάποιος αριθμός k από τα bits της διεύθυνσης του κάδου (bucket) υπολογίζονται αποκλειστικά βάσει της τιμής του γνωρίσματος a κάθε πλειάδας και τα υπόλοιπα υπολογίζονται αποκλειστικά βάσει της τιμής του γνωρίσματος b της πλειάδας. Υποθέστε ότι κάθε διεύθυνση κάδου (bucket) έχει 10 bits, δηλαδή υπάρχουν 1024 κάδοι. Κάθε κάδος αποθηκεύεται σε όσο λιγότερα blocks γίνεται. Υπολογείστε την τιμή του k (δηλαδή, των αριθμών των bits της διεύθυνσης του bucket που αφιερώνονται στο a) που ελαχιστοποιεί το μέσο αριθμό προσπελάσεων blocks για μια τυχαία ερώτηση στην R. Εξηγείστε την απάντησή σας.

(γ) Θεωρήστε ότι το μέγεθος του δείκτη εγγραφής είναι 8 bytes, το μέγεθος του δείκτη block είναι 7 bytes, το μέγεθος του block 512 bytes, και το μέγεθος του γνωρίσματος $R.a$ 10 bytes. Έστω μια σχέση $R(a, b, c, d)$ με 6.000.000 εγγραφές (πλειάδες) όπου κάθε block της σχέσης έχει 10 εγγραφές (παράγοντας ομαδοποίησης του αρχείου δεδομένων). Υποθέστε ότι το $R.a$ είναι υποψήφιο κλειδί και ότι η σχέση είναι αποθηκευμένη σε ένα αρχείο ταξινομημένο με βάση το $R.a$ σε φθήνουσα διάταξη. Το $R.a$ παίρνει τιμές από 0 έως 5.999.999. Δώστε το κόστος υπολογισμού για τις παρακάτω εκφράσεις της σχεσιακής άλγεβρας: (i) $\sigma_{a=60.000}(R)$ με προσπέλαση του ταξινομημένου αρχείου απευθείας

(ii) $\sigma_{a=60.000}(R)$ με χρήση πρωτεύοντος ευρετηρίου.