

**ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ
ΕΞΕΤΑΣΗ ΦΕΒΡΟΥΑΡΙΟΥ 2005**

ΛΥΣΕΙΣ

I. Βασιλείου

Θεωρήστε την παρακάτω βάση δεδομένων

employee (person-name, street, city)
works (person-name, company-name, salary)
company (company-name, city)
manages (person-name, manager-name)

(a) Δώστε μια παράσταση στη σχεσιακή άλγεβρα για τις παρακάτω ερωτήσεις (queries).

- i.- Βρείτε τα ονόματα και πόλεις διαμονής όλων των εργαζομένων που δουλεύουν για τη BANK bla bla.
- ii.- Βρείτε τα ονόματα όλων των εργαζομένων που ζουν στην ίδια πόλη με την εταιρεία στην οποία δουλεύουν.
- iii.- Βρείτε όλες τις εταιρείες που βρίσκονται ΣΕ ΚΑΘΕ ΠΟΛΗ στην οποία βρίσκεται η BANK bla bla (υποθέτουμε ότι μια τράπεζα μπορεί να βρίσκεται σε πολλές πόλεις).

ΛΥΣΗ

$$\Pi_{person-name, city} (employee \bowtie (\sigma_{company-name = "First Bank Corporation"} (works)))$$
$$\Pi_{person-name} ((employee \bowtie manages) \bowtie (manager-name = employee2.person-name \wedge employee.street = employee2.street \wedge employee.city = employee2.city) (\rho_{employee2} (employee)))$$
$$\Pi_{company-name} (company \div (\Pi_{city} (\sigma_{company-name = "Small Bank Corporation"} (company))))$$

(b) **Χρησιμοποιώντας τη SQL**

Ορίστε μια προβολή (όψη, view) που αποτελείται από το όνομα του διευθυντή (manager-name) και το μέσο μισθό όλων των εργαζομένων υπό την εποπτεία του. Εξηγήστε γιατί το σύστημα ΔΕΝ ΘΑ ΠΡΕΠΕΙ να επιτρέπει να εκφράζονται ενημερώσεις σε σχέση με αυτή την προβολή (όψη).

ΛΥΣΗ

```
CREATE VIEW salinfo AS
SELECT manager-name, AVG(salary)
FROM manages, m, works w,
WHERE m.employee-name = w.employee-name
GROUP BY manager-name
```

Προφανώς δεν μπορούν να επιτραπούν ενημερώσεις στην προβολή αυτή, διότι δεν υπάρχει τρόπος να καθορισθεί πως θα αλλάζουν τα δεδομένα της βάσης στα οποία ορίζεται η προβολή. Για παράδειγμα αν ζητηθεί να αλλάξει ο μέσος μισθός των υπαλλήλων που έχουν διευθυντή τον SMITH σε 500, δεν υπάρχει τρόπος να βρεθεί η ανάλογη (σωστή) αλλαγή των μισθών των υπαλλήλων (ατομικά) ώστε να προκύψει το 500 ως μέσος όρος.

(c) **Σκεφτείτε το SQL ερώτημα:**

```
SELECT      p.a1
FROM        p, r1, r2
WHERE       (p.a1 = r1.a1) OR ( p.a1 = r2.a1)
```

Κάτω από ποιες συνθήκες το ερώτημα αυτό επιλέγει τιμές για το **p.a1** που είναι στο **r1** ή στο **r2** ???
Εξετάστε προσεκτικά τις περιπτώσεις όπου ένα από τα **r1, r2** είναι κενό.

ΛΥΣΗ

Το ερώτημα επιλέγει αυτές τις τιμές του p.a1 που είναι ίσες με κάποια τιμή είτε του r1.a1 ή του r2.a1 ΕΑΝ ΚΑΙ ΜΟΝΟΝ ΕΑΝ και το r1 και το r2 δεν είναι κενά. Σε περίπτωση που ένα ή και τα δύο από αυτά είναι κενά, τότε το καρτεσιανό γινόμενο των p, r1, r2 είναι κενό, άρα και το αποτέλεσμα του ερωτήματος είναι κενό. Φυσικά, αν το p είναι κενό τότε και το αποτέλεσμα του ερωτήματος είναι κενό.

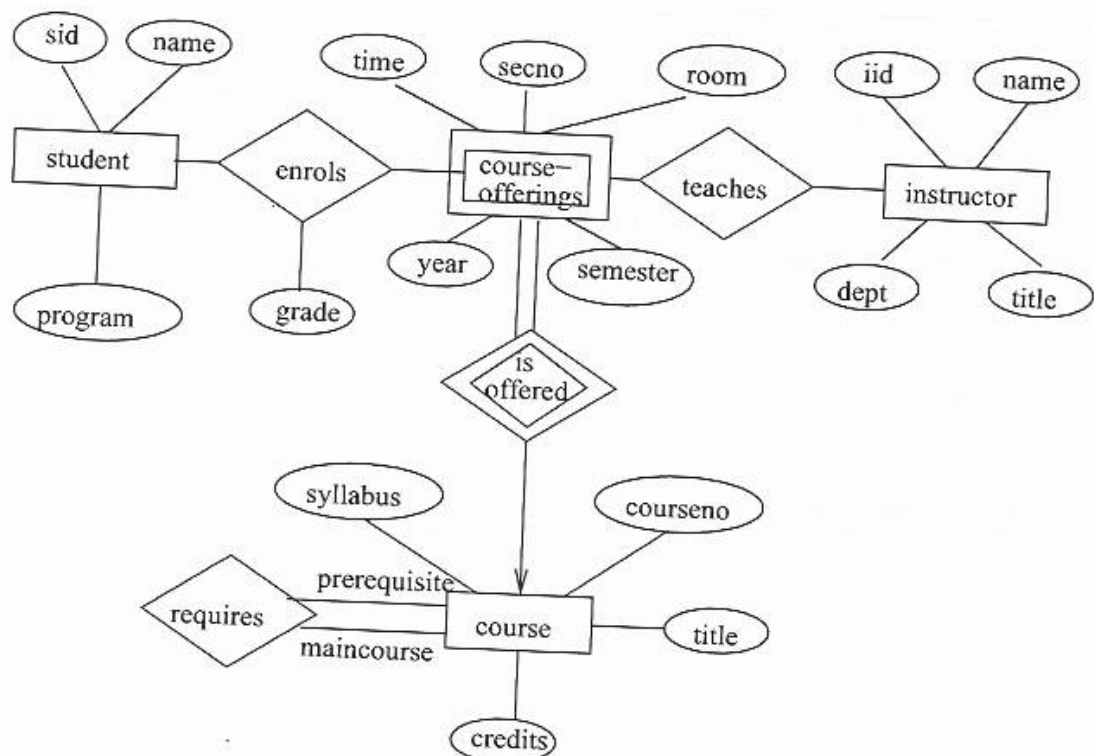
ΘΕΜΑ 2. -- ΜΟΝΤΕΛΟΠΟΙΗΣΗ [30]

Το γραφείο εγγραφών ενός Πανεπιστημίου διατηρεί δεδομένα για τις παρακάτω οντότητες:

- (α) **μαθήματα**, συμπεριλαμβανομένου του αριθμού, τίτλου, μονάδες, ύλη και προϋποθέσεων
- (β) **προσφορές μαθημάτων**, συμπεριλαμβανομένου του αριθμού, του έτους, του εξαμήνου, του αριθμού ενότητας, εισηγητών, ωρών, τάξεων κάθε μαθήματος
- (γ) **μαθητές**, συμπεριλαμβανομένου του κωδικού μαθητή, ονόματος και προγράμματος και
- (δ) **εισηγητές**, συμπεριλαμβανομένου του κωδικού, ονόματος, τμήματος και τίτλου. Επιπλέον, πρέπει να μοντελοποιηθούν κατάλληλα η ένταξη των μαθητών σε μαθήματα και οι βαθμοί που δίνονται στους μαθητές σε κάθε μάθημα.

- 1.- Κατασκευάστε ένα διάγραμμα οντότητας-σχέσης. Τεκμηριώστε όλες τις υποθέσεις που κάνετε για τους περιορισμούς απεικόνισης.
- 2.- Κατασκευάστε την αντίστοιχη σχεσιακή βάση δεδομένων

ΛΥΣΗ



ΣΧΕΣΙΑΚΗ ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ

student (student-id, name, program)
 course (courseno, title, syllabus, credits)
 course-offering (courseno, secno, year, semester, time, room)
 instructor (instructor-id, name, dept, title)
 enrols (student-id, courseno, secno, semester, year, grade)
 teaches (courseno, secno, semester, year, instructor-id)
 requires (maincourse, prerequisite)

ΘΕΜΑ 3.- ΣΧΕΣΙΑΚΟ ΜΟΝΤΕΛΟ ΚΑΙ ΚΑΝΟΝΙΚΟΠΟΙΗΣΗ [15]

(α) Αποδείξτε ότι ΚΑΘΕ σχέση με δύο γνωρίσματα (attributes) είναι σε κανονική μορφή BCNF (Boyce Codd Normal Form)---- **Υπόδειξη:** εξετάστε όλες τις περιπτώσεις ύπαρξης μη-τετριμμένων συναρτησιακών (λειτουργικών) εξαρτήσεων.

ΛΥΣΗ

Έστω ότι τα δύο γνωρίσματα είναι το A και το B .

Σύμφωνα με τον ορισμό του BCNF, εξετάζουμε όλες τις περιπτώσεις.

1.- ΔΕΝ υπάρχουν μη-τετριμμένες FD. Τότε προφανώς έχουμε BCNF διότι μόνο οι μη-τετριμμένες FD μπορούν να καταστρατηγήσουν τη συνθήκη του ορισμού (παρενθετικά, το μοναδικό κλειδί είναι ο συνδυασμός AB .)

2.- $A \rightarrow B$ ισχύει, αλλά δεν ισχύει $B \rightarrow A$. Τότε το A είναι το μοναδικό κλειδί και κάθε μη τετριμμένη FD περιέχει το A στο αριστερό σκέλος, άρα έχουμε BCNF.

- 3.- $B \rightarrow A$ ισχύει, αλλά δεν ισχύει η $A \rightarrow B$. Περίπτωση συμμετρική της 2.
 4.- $A \rightarrow B$ ισχύει και $B \rightarrow A$ ισχύει. Τότε και το A και το B είναι κλειδιά – χωρίς να καταστρατηγείται η συνθήκη του ορισμού της BCNF.

(β) Θεωρείστε μια σχέση $R(A, B, C, D)$ όπου ισχύουν οι παρακάτω συναρτησιακές εξαρτήσεις::
 $AB \rightarrow C, C \rightarrow D$ και $D \rightarrow A$,
 - Εντοπίστε όλα τα κλειδιά του R
 - Ποιες άλλες συναρτησιακές εξαρτήσεις προκύπτουν από τις παραπάνω εξαρτήσεις (περιοριστείτε σε εξαρτήσεις όπου η δεξιά πλευρά έχει μόνο ένα γνώρισμα)

ΛΥΣΗ

ΚΛΕΙΔΙΑ: AB, BC
ΙΣΧΥΟΥΝ ΕΠΙΣΗΣ $AB \rightarrow D, BC \rightarrow D, BC \rightarrow A, C \rightarrow A$, καθώς και οι τετριμμένες (π.χ., $AB \rightarrow A$, κλπ.)

(γ) Θεωρείστε μια σχέση $R(A, B, C, D, E, G)$ και το σύνολο των συναρτησιακών εξαρτήσεων που ισχύει στην R .

$$F = \{ AB \rightarrow C, AC \rightarrow B, AD \rightarrow E, B \rightarrow D, BC \rightarrow A, E \rightarrow G \}$$

Ποια από τις επόμενες διασπάσεις της R , όπου ισχύουν οι παραπάνω συναρτησιακές εξαρτήσεις
 (i) διατηρεί τις εξαρτήσεις? (ii) δε συνεπάγεται απώλεια πληροφορίας στις συζεύξεις – συνενώσεις.

- 1.- $R1(A, B), R2(B, C), R3(A, B, D, E), R4(E, G)$
- 2.- $R1(A, B, C), R2(A, C, D, E), R3(A, D, G)$

ΛΥΣΗ

1. ΔΕΝ διατηρεί τις εξαρτήσεις
 (π.χ., τη $AB \rightarrow C$).

Συνεπάγεται απώλεια πληροφορίας (ΔΕΝ είναι lossless).

Για να αποδειχθεί αυτό παίρνουμε δυο τυχαίες πλειάδες της αρχικής σχέσης, έστω $(a1, b, c1, d1, e1, g1)$ και $(a2, b, c2, d2, e2, g2)$. Εφόσον ισχύουν $AB \rightarrow C$ και $BC \rightarrow A$, τότε $a1 \neq a2$ αν και μόνο αν $c1 \neq c2$. Είναι εύκολο να δει κανείς ότι η συνένωση μεταξύ των AB και BC περιέχει 4 πλειάδες: $(a1, b, c1), (a1, b, c2), (a2, b, c1), (a2, b, c2)$. Συνεπώς, η συνένωση μεταξύ όλων των σχέσεων της διάσπασης θα περιέχει ΤΟΥΛΑΧΙΣΤΟΝ 4 πλειάδες, άρα έχουμε απώλεια πληροφορίας.

2. ΔΕΝ διατηρεί τις εξαρτήσεις
 (π.χ., τη $B \rightarrow D$, που προέρχεται από τις αρχικές εξαρτήσεις με τους κανόνες Armstrong.)

ΔΕΝ συνεπάγεται απώλεια πληροφορίας (είναι lossless)

Αν κανείς πάρει τη συνένωση των $R1$ και $R2$, τα κοινά γνωρίσματα AC αποτελούν κλειδί για τη $R2$ (lossless join). Συνεχίζοντας, αν το αποτέλεσμα της ανωτέρω συνένωσης το συνενώσουμε με το $R3$, τότε τα κοινά γνωρίσματα AD αποτελούν κλειδί της $R3$ (lossless join).

ΘΕΜΑ 4.- ΑΡΧΕΙΑ - ΕΥΡΕΤΗΡΙΑ - ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ [25]

- (α) Κατασκευάστε ένα $B+$ δένδρο για το παρακάτω σύνολο από τιμές κλειδιών:
 2, 3, 5, 7, 11, 17, 19, 23, 29, 31

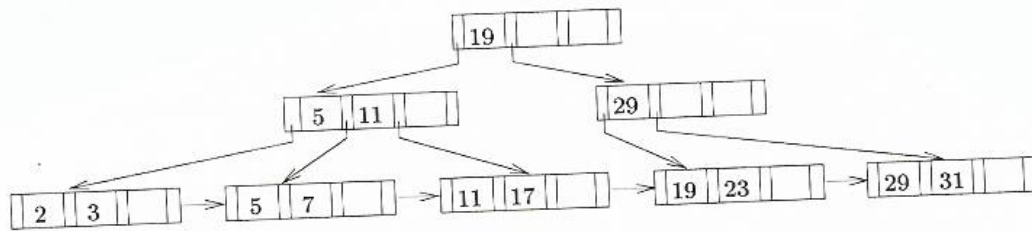
Θεωρείστε ότι ο αριθμός των δεικτών που χωρούν σε ένα κόμβο είναι 4.

Για το δένδρο που κατασκευάσατε, δείξτε ποια είναι η μορφή του μετά από:

- 1.- εισαγωγή του 9,
- 2.- εισαγωγή του 10

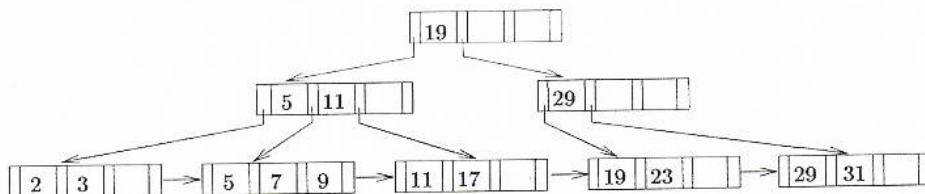
- 3.- διαγραφή του 23
4.- διαγραφή του 19

ΛΥΣΗ

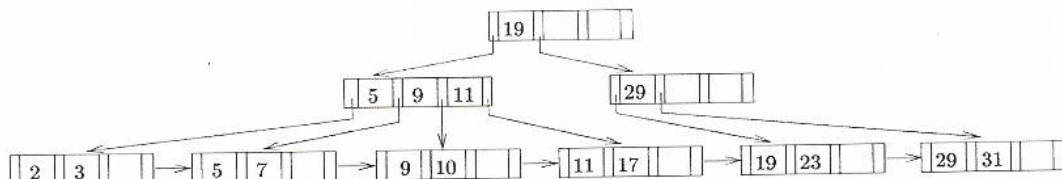


Σημειώνεται ότι εδώ δείχνουμε τα δέντρα μετά τις «διαδοχικές» αλλαγές – αλλά και αν κάθε αλλαγή γινόταν στο «αρχικό» δένδρο ξεχωριστά, θεωρήθηκε ΣΩΣΤΗ.

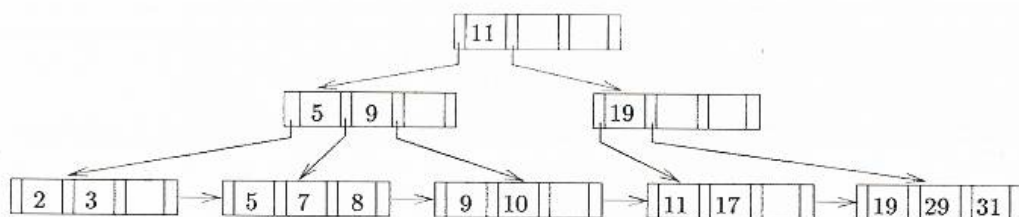
Insert 9:



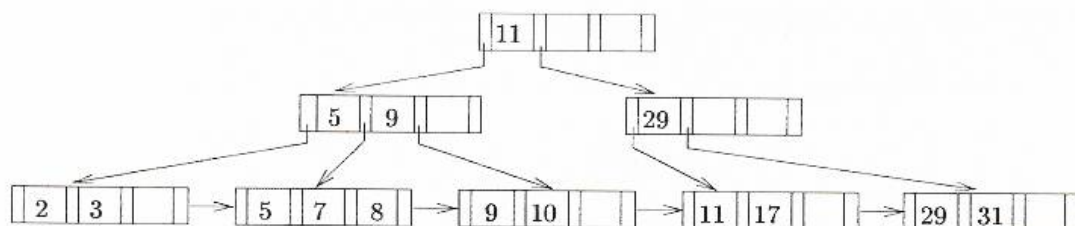
Insert 10:



Delete 23



Delete 19:



(b) Ας πούμε ότι οι σχέσεις $R1(A, B, C)$ και $R2(C, D, E)$ έχουν τις παρακάτω ιδιότητες: το $R1$ έχει 20.000 εγγραφές, το $R2$ έχει 45.000 εγγραφές, οι 25 εγγραφές του $R1$ χωρούν σε ένα μπλοκ και 30

εγγραφές του R2 χωρούν σε ένα μπλοκ. Εκτιμήστε τον αριθμό των προσπελάσεων μπλοκ που απαιτούνται, χρησιμοποιώντας καθεμία από τις στρατηγικές συνδέσμου για το $R1 \bowtie R2$

- σύνδεσμος ένθετου βρόχου (nested loop join)
- σύνδεσμος μπλοκ ένθετου βρόχου (block nested loop join)
- σύνδεσμος συγχώνευσης (merge join)
- σύνδεσμος HASH (hash join)

ΛΥΣΗ

Σύμφωνα με την εκφώνηση, η r1 χρειάζεται 800 μπλοκ και η r2 χρειάζεται 1500 μπλοκ για αποθήκευση.

Έστω ότι έχουμε M σελίδες διαθέσιμες στη μνήμη (buffer). Αν το $M > 800$, τότε χρησιμοποιούμε τις 800 σελίδες της μνήμης και κάνουμε ένα απλό σύνδεσμο ένθετου βρόχου με συνολικό κόστος (βέλτιστο) το $1500 + 800$ μπλοκ προσβάσεις.

Σε περίπτωση όπου το $M < 800$.

- σύνδεσμος ένθετου βρόχου (nested loop join)

Με το r1 εξωτερική σχέση, απαιτούνται $20000 * 1500 + 800 = 30000800$ προσβάσεις. Με το r2 εξωτερική σχέση, απαιτούνται $45000 * 800 + 1500 = 36001500$ προσβάσεις.

- σύνδεσμος μπλοκ ένθετου βρόχου (block nested loop join)

Με το r1 εξωτερική σχέση, απαιτούνται $800 / (M-1) * 1500 + 800$. Με το r2 εξωτερική σχέση, απαιτούνται $1500 / (M-1) * 800 + 1500$.

- σύνδεσμος συγχώνευσης (merge join)

Απαιτούνται $1500 + 800 + B$ προσβάσεις, όπου B είναι το κόστος ταξινόμησης στο γνώρισμα της συνένωσης ($B = 1500 * (2 \log_{(m-1)}(1500 / M)) + 800 * (2 \log_{(m-1)}(800 / M))$)

- σύνδεσμος HASH (hash join)

Υποθέτοντας ότι δεν έχουμε υπερχείλιση και με αρκετά μεγάλο M (για να μη χρειάζονται διαχωρισμοί), το κόστος είναι $3 * (1500 + 800) = 6900$ προσβάσεις

(γ) Θεωρείστε την παρακάτω ερώτηση σε SQL.

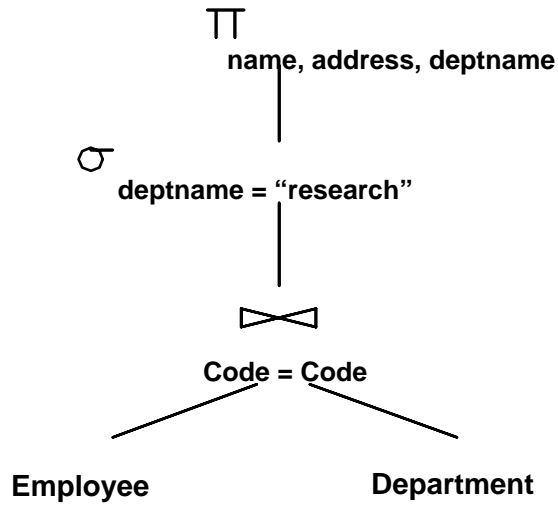
```
SELECT      E.name, E.Address, D.deptname
FROM        Employee E, Department D
WHERE       D.deptname= "Research" AND D.code=E.code
```

1.- Δώστε τουλάχιστον δύο δένδρα ερώτησης που μπορούν να την αναπαραστήσουν. Κάτω από ποιες συνθήκες θα χρησιμοποιούσατε καθεένα από τα δένδρα που σχηματίσατε?

2 – Δείξτε ένα δένδρο ερώτησης σε μια όσο το δυνατόν πιο βελτιστοποιημένη μορφή (κάνοντας όποιες παραδοχές θέλετε)

ΛΥΣΗ

Πρώτο Δένδρο (χωρίς βοηθητικές δομές)



Δεύτερο Δένδρο (που είναι και καλύτερο από το πρώτο)

