

**ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ
ΕΞΕΤΑΣΗ ΜΑΡΤΙΟΥ 2006 - ΛΥΣΕΙΣ**

I. Βασιλείου

ΘΕΜΑ 1 -- ΣΧΕΣΙΑΚΟ ΜΟΝΤΕΛΟ ΚΑΙ ΓΛΩΣΣΕΣ [20]

Θεωρήστε τη βάση δεδομένων ΕΠΙΧΕΙΡΗΣΗ (COMPANY), όπως καλύπτεται και στις διαφάνειες του μαθήματος:

EMPLOYEE (SSN, Name, BirthDate, Address, Sex, Salary, SupSSN, DNumber)
DEPARTMENT (DNumber, DName, MgrSSN, MgrStartDate)

PROJECT (PNumber, PName, Location, DNumber)

DEPT_LOCATION (DNumber, DLocation)

WORKS_ON (SSN, Pnumber, HoursPW)

1Α.- Γράψτε σε Σχεσιακή Άλγεβρα (Relational Algebra) τα παρακάτω ερωτήματα (queries).

(α) Βρες τα ονόματα των υπαλλήλων που δουλεύουν σε **όλα** τα έργα (retrieve the names of employees who work on every project)

(β) Βρες τα ονόματα των υπαλλήλων που δεν δουλεύουν σε κανένα έργο (retrieve the names of employees who do not work on any project)

ΛΥΣΗ

(α)

$\gamma_{\text{PNumber}, \text{SSN}}(\text{WORKS_ON})$

$\rho_{\text{PNumber}}(\text{PROJECT})$

$\gamma_{\text{PNumber}, \text{SSN}}(\text{WORKS_ON}) \div \rho_{\text{PNumber}}(\text{PROJECT})$

Τελικό Query =

$\pi_{\text{Name}}(\text{EMPLOYEE}) \triangleright \triangleleft_{\text{SSN}=\text{SSN}} (\pi_{\text{PNumber}, \text{SSN}}(\text{WORKS_ON}) \div \rho_{\text{PNumber}}(\text{PROJECT}))$

(β)

$\pi_{\text{SSN}}(\text{EMPLOYEES})$

$\pi_{\text{SSN}}(\text{WORKS_ON})$

Τελικό query =

$\pi_{\text{Name}}(\text{EMPLOYEE}) \triangleright \triangleleft_{\text{SSN}=\text{SSN}} (\pi_{\text{SSN}}(\text{EMPLOYEES}) - \pi_{\text{SSN}}(\text{WORKS_ON}))$

1B.- Γράψτε σε SQL τις παρακάτω ερωτήσεις (queries).

(α) Για κάθε Τμήμα, βρες το όνομα του Τμήματος και το συνολικό κόστος σε μισθούς για τους εργαζομένους στο Τμήμα (for each Department, list the Department name and the total Salary - by all employees – for that Department.)

(β) Βρες τα ονόματα και διευθύνσεις υπαλλήλων που δουλεύουν τουλάχιστον σε ένα έργο που εκτελείται (βρίσκεται) στα Πατήσια, για το οποίο το τμήμα που το ελέγχει δεν βρίσκεται στα Πατήσια (retrieve the names and addresses of employees who work on at least one project located in patisia but whose department has no location in patisia)

ΛΥΣΗ

(α)

```
select Dname, sum(Salary)
from EMPLOYEE, DEPARTMENT
where EMPLOYEE.DNumber = DEPARTMENT.DNumber
group by DName
```

(β)

```
select Name, Address
from EMPLOYEE
where exists (select *
              from WORKS_ON, PROJECT
              where WORKS_ON.Pnumber = PROJECT.PNumber and
                    EMPLOYEE.SSN = WORKS_ON.SSN and
                    PLocation = 'patisia' )
and
not exists (select *
            from DEPT_LOCATION
            where DEPT_LOCATION.DNumber= EMPLOYEE.DNumber
            and DLocation = 'patisia' )
```

ΘΕΜΑ 2. -- ΜΟΝΤΕΛΟΠΟΙΗΣΗ [30]

Μια εταιρεία μεταφορών, ονομαζόμενη ΚΟΥΒΑΛΑ, μεταφέρει αγαθά για μια μεγάλη αλυσίδα supermarkets, που ονομάζεται ΠΑΡΕΚΟΣΜΕ. Τα φορτηγά της ΚΟΥΒΑΛΑ παραλαμβάνουν αγαθά από αποθήκες και τα παραδίνουν σε συγκεκριμένα καταστήματα (supermarkets). Ένα φορτηγό είναι δυνατόν να μεταφέρει περισσότερες τις μια “αποστολές” (σύνολο αγαθών για ένα κατάστημα) σε ένα απλό ταξίδι (ο αριθμός ταξιδιού προσδιορίζει μονοσήμαντα το ταξίδι) και παραδίδει αυτές τις αποστολές σε μια σειρά από supermarkets.

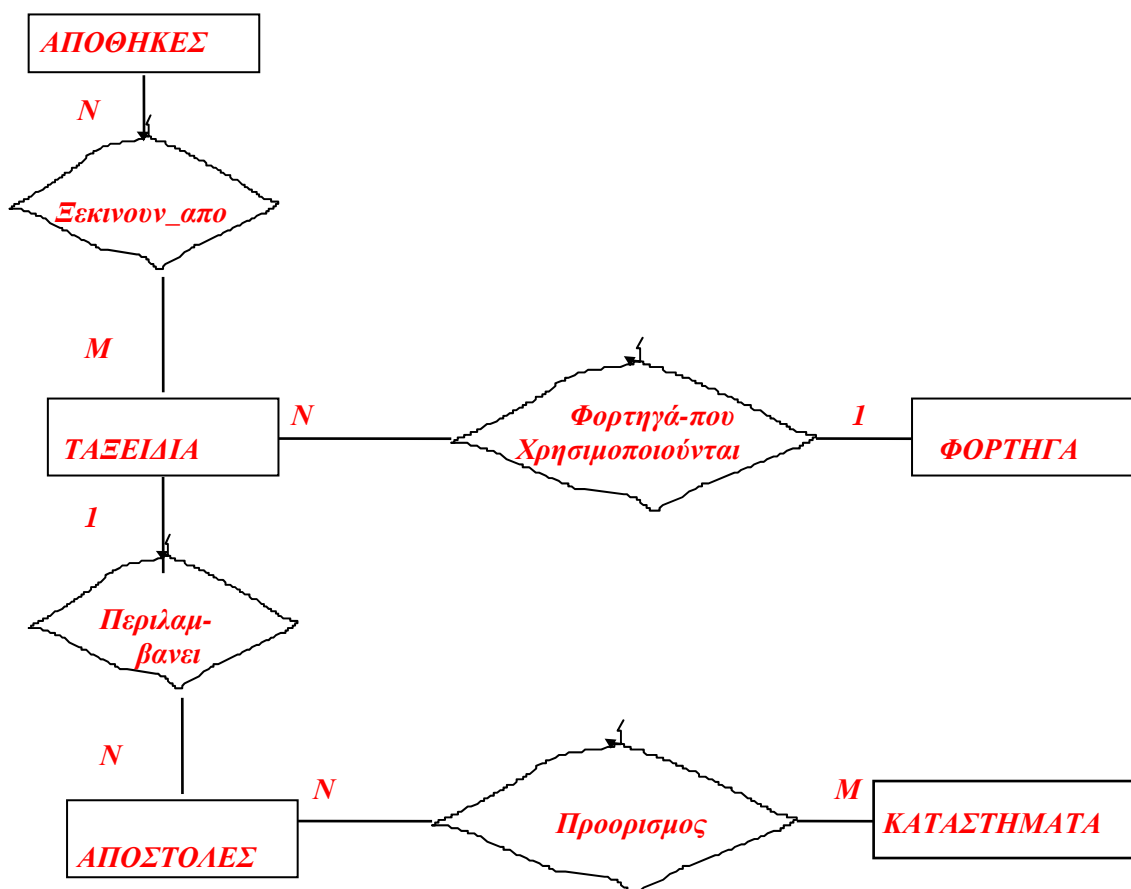
Κάθε αποστολή αναγνωρίζεται από ένα ΑριθμοΑποστολής και έχει άλλες πληροφορίες, όπως όγκο αποστολής, βάρος, προορισμός, κλπ. Τα φορτηγά έχουν διαφορετικές δυναμικότητες και όσον αφορά τον όγκο αγαθών που χωρούν, αλλά και το βάρος που μπορούν να μεταφέρουν.

Μια βάση δεδομένων, που θα χρησιμοποιηθεί και από τις δύο εταιρείες, είναι υπό σχεδιασμό για να κρατάει στοιχεία σχετικά με την χρήση των φορτηγών καθώς και τις αποστολές που γίνονται. Η βάση δεδομένων θα χρησιμοποιηθεί για τον προγραμματισμό των φορτηγών και για έγκαιρες παραδόσεις αγαθών στα πολυκαταστήματα (supermarkets).

2Α.- Κατασκευάστε ένα διάγραμμα οντότητας-σχέσης για τη μοντελοποίηση των ανωτέρω.

Το διάγραμμα πρέπει να είναι **πλήρες** υπό την έννοια ότι όλες οι ιδιότητες των οντοτήτων-σχέσεων πρέπει να δείχνονται.

ΛΥΣΗ (ΕΛΑΧΙΣΤΗ)



ΑΠΟΘΗΚΕΣ (Όνομα, Τοποθεσία, ...)

ΤΑΞΕΙΔΙΑ (ΑριθμόςΤαξιδιού, Ημερομηνία, ...)

ΑΠΟΣΤΟΛΕΣ (ΑριθμόςΑποστολής, Βάρος, Ογκος, ...)

ΦΟΡΤΗΓΑ (ΑριθμόςΦορτηγού, ΔυναμικότηταΒάρους, ΔυναμΟγκου, ...)

ΚΑΤΑΣΤΗΜΑΤΑ (Όνομα, Διεύθυνση, ...)

2B.- Κατασκευάστε την αντίστοιχη σχεσιακή βάση δεδομένων, είτε παρουσιάζοντας ένα διάγραμμα (σχεσιακού) σχήματος όπως στο Βιβλίο του μαθήματος ή με απλή παράθεση των Σχημάτων Σχέσεων (φυσικά, με όλα τα στοιχεία: ιδιότητες – γνωρίσματα, κλειδιά, κλπ.)

Σημείωση:: Για το σχεδιασμό μπορείτε να είτε να **ΜΕΤΑΤΡΕΨΕΤΕ** το E-R διάγραμμα (παραπάνω) σε ανάλογη σχεσιακή περιγραφή, ή να παρουσιάσετε το σχεσιακό μοντέλο απευθείας με τα στοιχεία περιγραφής που δίνονται στην εκφώνηση.

ΑΠΟΘΗΚΕΣ (Όνομα, Τοποθεσία, ...)

ΤΑΞΕΙΔΙΑ (ΑριθμόςΤαξιδιού, Ημερομηνία, ..ΑριθμόςΦορτηγού)

ΑΠΟΣΤΟΛΕΣ (ΑριθμόςΑποστολής, Βάρος, Ογκος, ... ΑριθμόςΤαξιδιού)

ΦΟΡΤΗΓΑ (ΑριθμόςΦορτηγού, ΔυναμικότηταΒάρους, ΔυναμΟγκου, ...)

ΚΑΤΑΣΤΗΜΑΤΑ (Όνομα, Διεύθυνση, ...)

ΑΠΟΘΗΚΕΣ-ΤΑΞΕΙΔΙΑ(ΑριθμόςΤαξιδιού, ΌνομαΑποθήκης)

ΚΑΤΑΣΤΗΜΑΤΑ-ΤΑΞΕΙΔΙΑ(ΑριθμόςΤαξιδιού, ΌνομαΚαταστήματος)

ΘΕΜΑ 3.- ΣΧΕΣΙΑΚΟ ΜΟΝΤΕΛΟ ΚΑΙ ΚΑΝΟΝΙΚΟΠΟΙΗΣΗ [25]

3Α.- Θεωρείστε το Σχήμα R με τα γνωρίσματα (attributes): A, B, C, D, E, G, H και τις λειτουργικές εξαρτήσεις (functional dependencies) που ισχύουν στο R :

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

Για κάθε ένα από τα παρακάτω σύνολα γνωρισμάτων να κάνετε τρία πράγματα:

- (α) Υπολογίστε το σύνολο των εξαρτήσεων που ισχύουν για το συγκεκριμένο σύνολο γνωρισμάτων και διατυπώστε μια (ελάχιστη) κανονική κάλυψη του
- (β) Αναγνωρίστε αν το σύνολο γνωρισμάτων ικανοποιεί την κανονική μορφή BCNF,
- (γ) αν το σύνολο δεν είναι σε BCNF, τότε επιδιώξτε μια αποσύνθεση σε ένα σύνολο από BCNF σχήματα που διατηρούν τις εξαρτήσεις.

- (i) {A, B, C }
- (ii) {A, B, C, E, G}
- (iii) {A, C, E, H }

Υπενθυμίζεται ο ορισμός της κανονικής μορφής BCNF (εάν και μόνο εάν για κάθε $X \rightarrow Y$ που ισχύει τότε ένα εκ των δύο συμβαίνει: είτε το $X \rightarrow Y$ είναι τετριμμένη λειτουργική εξάρτηση ή το X είναι υπερκλειδί)

ΛΥΣΗ

- (i) {A, B, C }

- (α) $AB \rightarrow C, AC \rightarrow B, BC \rightarrow A$ - ήδη είναι μια ελάχιστη κάλυψη
- (β) Είναι σε BCNF διότι τα AB, AC και BC είναι όλα υποψήφια κλειδιά στη σχέση.
- (γ) Είναι ήδη BCNF.

- (ii) {A, B, C, E, G}

- (α) $AB \rightarrow C, AC \rightarrow B, BC \rightarrow A, E \rightarrow G$. - ήδη είναι μια ελάχιστη κάλυψη.
- (β) Τα ABE, ACE και BCE είναι όλα υποψήφια κλειδιά. Αλλά δεν είναι ούτε 2NF γιατί το E είναι υποσύνολο των κλειδίων και ισχύει η $E \rightarrow G$. Είναι σε 1NF.
- (γ) Η αποσύνθεση σε ABC, ABE και EG είναι BCNF.

- (iv) {A, C, E, H }

- (α) Δεν υπάρχουν λειτουργικές εξαρτήσεις - ήδη μια ελάχιστη κάλυψη.
- (β) Το κλειδί είναι η ίδια η ACEH
- (γ) Είναι σε μορφή BCNF.

3Β.- Έστω τα παρακάτω δύο σύνολα λειτουργικών εξαρτήσεων

$$F1 = \{ A \rightarrow B, AB \rightarrow C, D \rightarrow AC, D \rightarrow E \}$$

$$F2 = \{ A \rightarrow BC, D \rightarrow AE \}$$

Δείξτε ότι είναι ισοδύναμα (δηλαδή, $F1^+ = F2^+$)

Ένας τρόπος για ναδειχθεί αυτό είναι να υπολογιστούν τα δύο κλεισίματα των δύο συνόλων και να φανεί ότι είναι ίδια. Ένας άλλος ευκολότερος τρόπος είναι να γίνουν ορισμένοι μετασχηματισμοί / απλοποιήσεις - χρησιμοποιώντας τους κανόνες Armstrong - και να φανεί ότι τα δύο σύνολα εξαρτήσεων είναι ουσιαστικά τα ίδια.

Υπενθυμίζονται ορισμένοι κανόνες που ισχύουν:

- Αν Y υποσύνολο του X, τότε $X \rightarrow Y$ (αντανακλαστικότητα)
- Αν $X \rightarrow Y$ ισχύει, τότε $XZ \rightarrow YZ$ (προσαύξηση)
- Αν $X \rightarrow Y$ και $Y \rightarrow Z$, τότε $X \rightarrow Z$ (μεταβατικότητα)
- Αν $X \rightarrow Y$ και $X \rightarrow Z$, τότε $X \rightarrow YZ$ (ένωση)

$\text{An } X \rightarrow YZ, \text{ τότε } X \rightarrow Y \text{ και } X \rightarrow Z \text{ (αποσύνθεση)}$
 $\text{An } X \rightarrow Y \text{ και } ZY \rightarrow P, \text{ τότε } XZ \rightarrow P \text{ (ψευδο-μεταβατικότητα)}$
 $\text{An } X \rightarrow Y, \text{ τότε } X \rightarrow Y \cap Z \text{ (τομή συνόλων)}$

ΛΥΣΗ

Βάζουμε στη σειρά τις λειτουργικές εξαρτήσεις του F1.

- 1.- $A \rightarrow B$,
- 2.- $AB \rightarrow C$
- 3.- $D \rightarrow AC$
- 4.- $D \rightarrow E$

Η 3.- μπορεί να αντικατασταθεί από

- 3.- $D \rightarrow A \text{ και } D \rightarrow C \text{ (αποσύνθεση)}$

Το 2.- μπορεί να αντικατασταθεί (λόγω του 1.-) από το

- 2.- $A \rightarrow C$

Αλλά τώρα ισχύει με μεταβατικότητα ($D \rightarrow A \text{ και } A \rightarrow C \text{ άρα } D \rightarrow C$) συνεπώς το $D \rightarrow C$ μπορεί να φύγει και το 3.- γίνεται:

- 3.- $D \rightarrow A$

Το F1 τώρα έγινε:

- 1.- $A \rightarrow B$,
- 2.- $A \rightarrow C$
- 3.- $D \rightarrow A$
- 4.- $D \rightarrow E$

Που προφανώς είναι το ίδιο με το F2 (εφαρμόζοντας τον κανόνα ένωσης)

3Γ.- Αποδείξτε (χρησιμοποιώντας τους κανόνες ανωτέρω) ότι,

$\text{An } X \rightarrow Y \text{ και } Z \rightarrow P, \text{ τότε ισχύει } X(Z - Y) \rightarrow YP$

(όπου το “-” είναι η διαφορά συνόλων και το YP είναι ταυτόσημο με το $Y \cup P$ – ένωση συνόλων)

- | | |
|--|---------------------------|
| 1.- $X \rightarrow Y$ | (δεδομένο) |
| 2.- $Z \rightarrow P$ | (δεδομένο) |
| 3.- $X \rightarrow Y \cap Z$ | (τομή συνόλων και 1) |
| 4.- $(Z - Y) \rightarrow (Z - Y)$ | (αντανακλαστικότητα) |
| 5.- $X(Z - Y) \rightarrow (Y \cap Z)(Z - Y)$ | (προσαύξηση, 3 και 4) |
| 6.- $X(Z - Y) \rightarrow Z$ | (απλούστευση της 5) |
| 7.- $X(Z - Y) \rightarrow P$ | (μεταβατικότητα, 6 και 2) |
| 8.- $X(Z - Y) \rightarrow YP$ | (ένωση, 1 και 7) |

ΘΕΜΑ 4.- ΕΥΡΕΤΗΡΙΑ - ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ [25]

4Α. - Θεωρείστε το παρακάτω Σχεσιακό Σχήμα και το SQL ερώτημα

Suppliers(sid: integer, sname: char(20), city: char(20))

Supply(sid: integer, pid: integer)

Parts(pid: integer, pname: char(20), price: real)

SELECT S.sname, P.pname

FROM Suppliers S, Parts P, Supply Y
WHERE S.sid = Y.sid AND Y.pid = P.pid AND S.city = "Madison" AND P.price < 1000

(α) Τι πληροφορίες για τις παραπάνω Σχέσεις χρειάζεται ένας σοβαρός βελτιστοποιητής (όπως ο βελτιστοποιητής του System R) για να επιλέξει ένα καλό πλάνο εκτέλεσης για το συγκεκριμένο SQL Ερώτημα?

Ο βελτιστοποιητής χρειάζεται πληροφορίες όπως του ποιά ευρετήρια (indexes) υπάρχουν και τι τύπων στα S.sid, Y.sid, Y.pid, P.pid, S.city, P.price. Επίσης θα χρειαστεί στατιστικά για τη βάση δεδομένων, όπως τις υψηλές / χαμηλές τιμές, κατανομή τιμών κλπ.

(β) Πόσες διαφορετικές διατάξεις συνδέσμων (joins) ένας τέτοιος βελτιστοποιητής θα εξετάσει όταν αποφασίζει πως θα εκτελεστεί το ερώτημα και ποιες? (Υπόδειξη: διατάξεις αριστερού βάθους συνδέσμων – left-deep plans)

Ο βελτιστοποιητής θα εξετάσει μόνο δύο διατάξεις (αριστερού βάθους), συγκεκριμένα ((S join Y) join P) και ((Y join P) join S).

(γ) Ποια ευρετήρια είναι ενδεχομένως χρήσιμα στον υπολογισμό του ερωτήματος. Αιτιολογείστε σύντομα την απάντησή σας.

Ένα ταξινομημένο ευρετήριο στο P.price θα ήταν χρήσιμο (range retrieval). Επίσης, B+ - δενδρικά ευρετήρια στα Y.sid, Y.pid, P.pid για τη χρήση του αλγορίθμου sort-merge.

(δ) Πώς επηρεάζονται τα πλάνα εκτέλεσης από την πρόσθεση ενός "ORDER BY sname" στο ερώτημα?

Ο βελτιστοποιητής θα πρέπει να εξετάσει πλάνα εκτέλεσης που αφήνουν το sname ταξινομημένο (παράπλευρη επιτυχία), αλλιώς θα πρέπει να έχει πλάνα εκτέλεσης που στο τέλος ταξινομούν απευθείας στο sname.

4B.- Αποδείξτε την παρακάτω ισοδυναμία:

$$X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$$

Όπου τα X, Y, Z είναι σύνολα γνωρισμάτων (ιδιοτήτων), και \cup, \cap είναι η Ένωση και Τομή αντίστοιχα.

Υπόδειξη Δείξτε ότι αν μια πλειάδα t ανήκει στο σύνολο που συγκροτεί το αριστερό σκέλος, τότε ανήκει και στο σύνολο στο δεξί σκέλος (και αντίστροφα).

ΛΥΣΗ

Έστω ότι η t ανήκει στο $X \cup (Y \cap Z)$, τότε το t ανήκει στο X ή το t ανήκει στο $(Y \cap Z)$.

Αν το t ανήκει στο X τότε ισχύει ότι το t ανήκει και στο $X \cup Y$ και στο $X \cup Z$, άρα το t ανήκει στο $(X \cup Y) \cap (X \cup Z)$

Αν το t ανήκει στο $(Y \cap Z)$, τότε ισχύει ότι το t ανήκει και στο Y και στο Z, άρα το t ανήκει και στο $(X \cup Y)$ και στο $(X \cup Z)$, συνεπώς και στην τομή αυτών.

Αντίστροφα, αν το t ανήκει στο $(X \cup Y) \cap (X \cup Z)$, συμπεραίνουμε ότι το t ανήκει και στο $(X \cup Y)$ και στο $(X \cup Z)$, άρα το t ανήκει στο X ή ανήκει και στο Y και στο Z. Συνεπώς το t ανήκει στο $X \cup (Y \cap Z)$.