

Βάσεις Δεδομένων II

Φροντιστήριο

23/3/2010

Βελτιστοποίηση Ερωτημάτων

- Διαδικασία επιλογής του πιο αποτελεσματικού πλάνου υπολογισμού ενός ερωτήματος, μεταξύ των διαφόρων στρατηγικών.
 - Επίπεδο σχεσιακής άλγεβρας
 - Επίπεδο επιλογής αλγορίθμου για την εκτέλεση μιας λειτουργίας
 - Επιλογή συγκεκριμένου ευρετηρίου

Συμβολισμός Δέντρων Επερωτήσεων

- **Δέντρο επερωτήσεων** : μία δεντρική δομή που αντιστοιχεί σε μία έκφραση της σχεσιακής άλγεβρας
 - Σχέσεις εισόδου \leftrightarrow κόμβοι φύλλα
 - Πράξεις σχεσιακής άλγεβρας \leftrightarrow εσωτερικοί κόμβοι
 - Εκτέλεση του δέντρου επερώτησης:
 - Εκτέλεση των πράξεων κάθε εσωτερικού κόμβου όταν οι σχέσεις εισόδου του είναι διαθέσιμες και...
 - Αντικατάσταση του εσωτερικού κόμβου από την σχέση που προκύπτει από την εκτέλεση της πράξης
 - Τερματισμός:
 - Εκτέλεση της πράξης στον κόμβο ρίζα
- Δείχνει την σειρά εκτέλεσης – Πολλά ισοδύναμα δέντρα

Κανόνες Ισοδυναμίας (1/5)

1. Οι συζευκτικές επιλογές μπορούν να αναλυθούν σε μια ακολουθία ξεχωριστών επιλογών

$$\sigma_{\theta_1 \wedge \theta_2}(E) = \sigma_{\theta_1}(\sigma_{\theta_2}(E))$$

2. Οι επιλογές είναι αντιμεταθετικές

$$\sigma_{\theta_1}(\sigma_{\theta_2}(E)) = \sigma_{\theta_2}(\sigma_{\theta_1}(E))$$

3. Σε μια ακολουθία από προβολές μόνο η τελευταία προβολή χρειάζεται, οι υπόλοιπες μπορούν να παραληφθούν

$$\Pi_{L_1}(\Pi_{L_2}(\dots(\Pi_{L_n}(E))\dots)) = \Pi_{L_1}(E)$$

4. Οι επιλογές μπορούν να συνδυαστούν με Καρτεσιανά γινόμενα και θήτα συνενώσεις

- a. $\sigma_{\theta}(E_1 \times E_2) = E_1 \bowtie_{\theta} E_2$

- b. $\sigma_{\theta_1}(E_1 \bowtie_{\theta_2} E_2) = E_1 \bowtie_{\theta_1 \wedge \theta_2} E_2$

Κανόνες Ισοδυναμίας (2/5)

5. Οι λειτουργίες Θ -συνένωσης (και φυσικής σύνδεσης) είναι αντιμεταθετικές

$$E_1 \bowtie_{\theta} E_2 = E_2 \bowtie_{\theta} E_1$$

6. a) Οι λειτουργίες φυσικής σύνδεσης είναι προσεταιριστικές :

$$(E_1 \bowtie E_2) \bowtie E_3 = E_1 \bowtie (E_2 \bowtie E_3)$$

(b) Οι λειτουργίες Θ -συνένωσης είναι προσεταιριστικές, με την ακόλουθη έννοια :

$$(E_1 \bowtie_{\theta_1} E_2) \bowtie_{\theta_2 \wedge \theta_3} E_3 = E_1 \bowtie_{\theta_1 \wedge \theta_3} (E_2 \bowtie_{\theta_2} E_3)$$

όπου η θ_2 περιλαμβάνει χαρακτηριστικά μόνο των E_2 και E_3

Κανόνες Ισοδυναμίας (3/5)

7. Η επιλογή αντιμετωπίζεται με Θ - συνένωση κάτω από τις ακόλουθες δύο συνθήκες:

(a) Όταν όλα τα χαρακτηριστικά της συνθήκης θ_0 προέρχονται μόνο από τα χαρακτηριστικά μιας εκ των σχέσεων που συμμετέχουν στη σύνδεση (π.χ. E_1)

$$\sigma_{\theta_0}(E_1 \bowtie_{\theta} E_2) = (\sigma_{\theta_0}(E_1)) \bowtie_{\theta} E_2$$

(b) Όταν όλα τα χαρακτηριστικά της συνθήκης θ_1 προέρχονται (περιλαμβάνουν) μόνο από χαρακτηριστικά της E_1 ενώ τα χαρακτηριστικά της θ_2 προέρχονται μόνο από χαρακτηριστικά της E_2

$$\sigma_{\theta_1 \wedge \theta_2}(E_1 \bowtie_{\theta} E_2) = (\sigma_{\theta_1}(E_1)) \bowtie_{\theta} (\sigma_{\theta_2}(E_2))$$

Κανόνες Ισοδυναμίας (4/5)

8. Η προβολή επιμερίζεται με τη Θ -συνένωση ως εξής:

(a) Αν η προβολή Π περιλαμβάνει χαρακτηριστικά μόνο από το σύνολο $L_1 \cup L_2$:

$$\Pi_{L_1 \cup L_2}(E_1 \bowtie_{\theta} E_2) = (\Pi_{L_1}(E_1)) \bowtie_{\theta} (\Pi_{L_2}(E_2))$$

(b) Ας θεωρήσουμε τη συνένωση $E_1 \bowtie_{\theta} E_2$

- Έστω L_1 και L_2 είναι σύνολα χαρακτηριστικών από E_1 και E_2 , αντίστοιχα
- Έστω L_3 τα χαρακτηριστικά της E_1 που συμμετέχουν στη συνθήκη θ , αλλά δεν ανήκουν στο $L_1 \cup L_2$, και
- έστω L_4 τα χαρακτηριστικά της E_2 που συμμετέχουν στη συνθήκη θ αλλά δεν ανήκουν στο $L_1 \cup L_2$

$$\Pi_{L_1 \cup L_2}(E_1 \bowtie_{\theta} E_2) = \Pi_{L_1 \cup L_2}((\Pi_{L_1 \cup L_3}(E_1)) \bowtie_{\theta} (\Pi_{L_2 \cup L_4}(E_2)))$$

Κανόνες Ισοδυναμίας (5/5)

9. Η ένωση και η τομή είναι αντιμεταθετικές

$$\begin{aligned}E_1 \cup E_2 &= E_2 \cup E_1 \\E_1 \cap E_2 &= E_2 \cap E_1\end{aligned}$$

Σημείωση: η διαφορά δεν είναι αντιμεταθετική

10. Η ένωση και η τομή είναι προσεταιριστικές

$$\begin{aligned}(E_1 \cup E_2) \cup E_3 &= E_1 \cup (E_2 \cup E_3) \\(E_1 \cap E_2) \cap E_3 &= E_1 \cap (E_2 \cap E_3)\end{aligned}$$

11. Η λειτουργία της επιλογής επιμερίζεται με τις λειτουργίες \cup , \cap και $-$

$$\sigma_\theta (E_1 - E_2) = \sigma_\theta (E_1) - \sigma_\theta (E_2)$$

Όμοια με τους τελεστές \cup και \cap

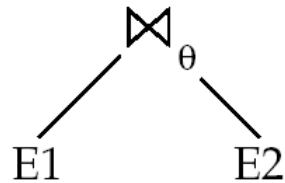
Επιπλέον: $\sigma_\theta (E_1 - E_2) = \sigma_\theta (E_1) - E_2$ αν η θ αφορά μόνο χαρακτηριστικά της E_2

Όμοια με τον τελεστή \cap αλλά δεν ισχύει με το \cup

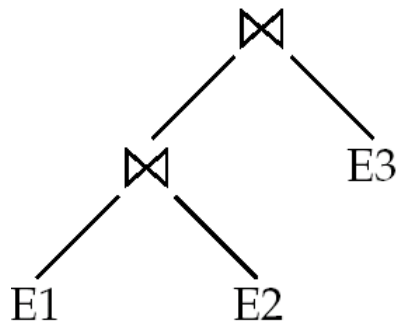
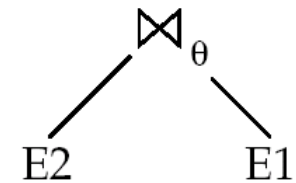
12. Η λειτουργία της προβολής κατανέμεται στη λειτουργία της ένωσης

$$\Pi_L(E_1 \cup E_2) = (\Pi_L(E_1)) \cup (\Pi_L(E_2))$$

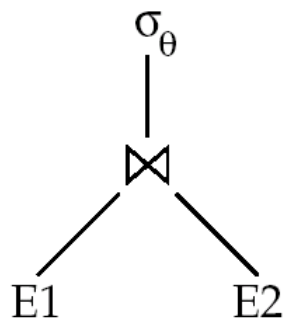
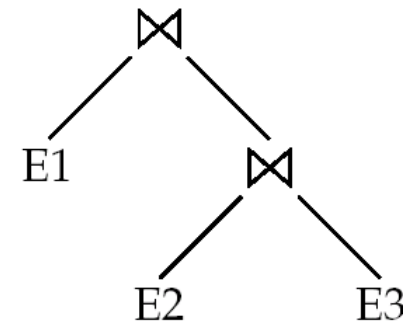
Γραφική Αναπαράσταση Ισοδυναμιών



Rule 5

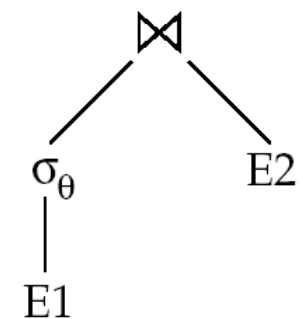


Rule 6a



Rule 7a

If θ only has
attributes from E1

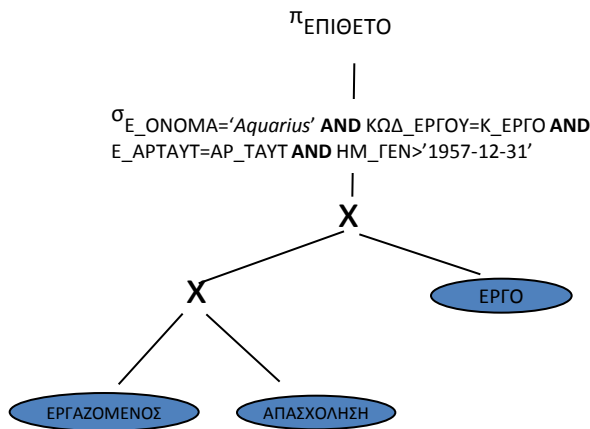


Άσκηση 1 - Παράδειγμα βελτιστοποίησης

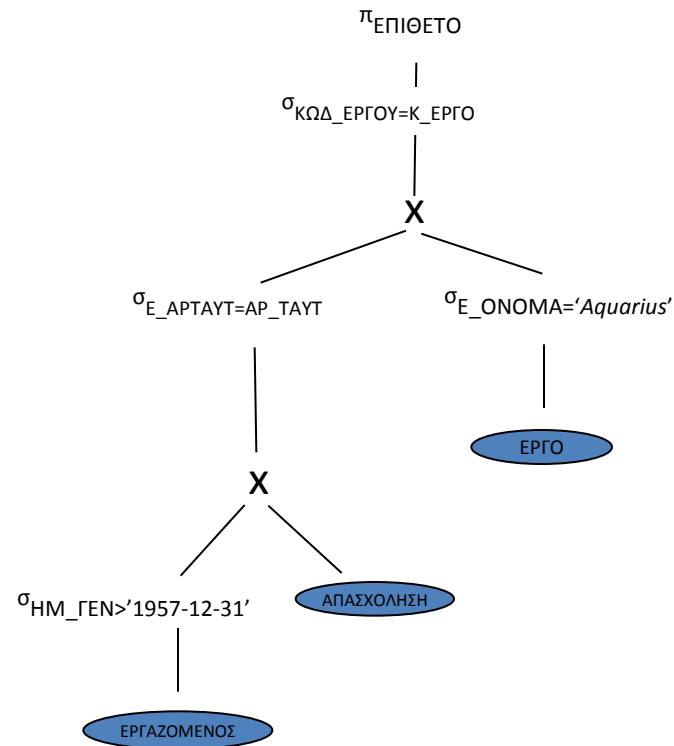
- Ερώτημα:
 - Να βρεθούν τα επίθετα των εργαζομένων που γεννήθηκαν μετά το 1957 και απασχολούνται σε ένα έργο με όνομα "Aquarius"
- SQL:
 - **SELECT** ΕΠΙΘΕΤΟ
FROM ΕΡΓΑΖΟΜΕΝΟΣ, ΑΠΑΣΧΟΛΗΣΗ, ΕΡΓΟ
WHERE Ε_ΟΝΟΜΑ='Aquarius' **AND** ΚΩΔ_ΕΡΓΟΥ=Κ_ΕΡΓΟ
AND Ε_ΑΡΤΑΥΤ_ΑΡ_ΤΑΥΤ **AND** ΗΜ_ΓΕΝ>'1957-12-31'

Παράδειγμα βελτιστοποίησης Δέντρο επερώτησης

- Αρχικό

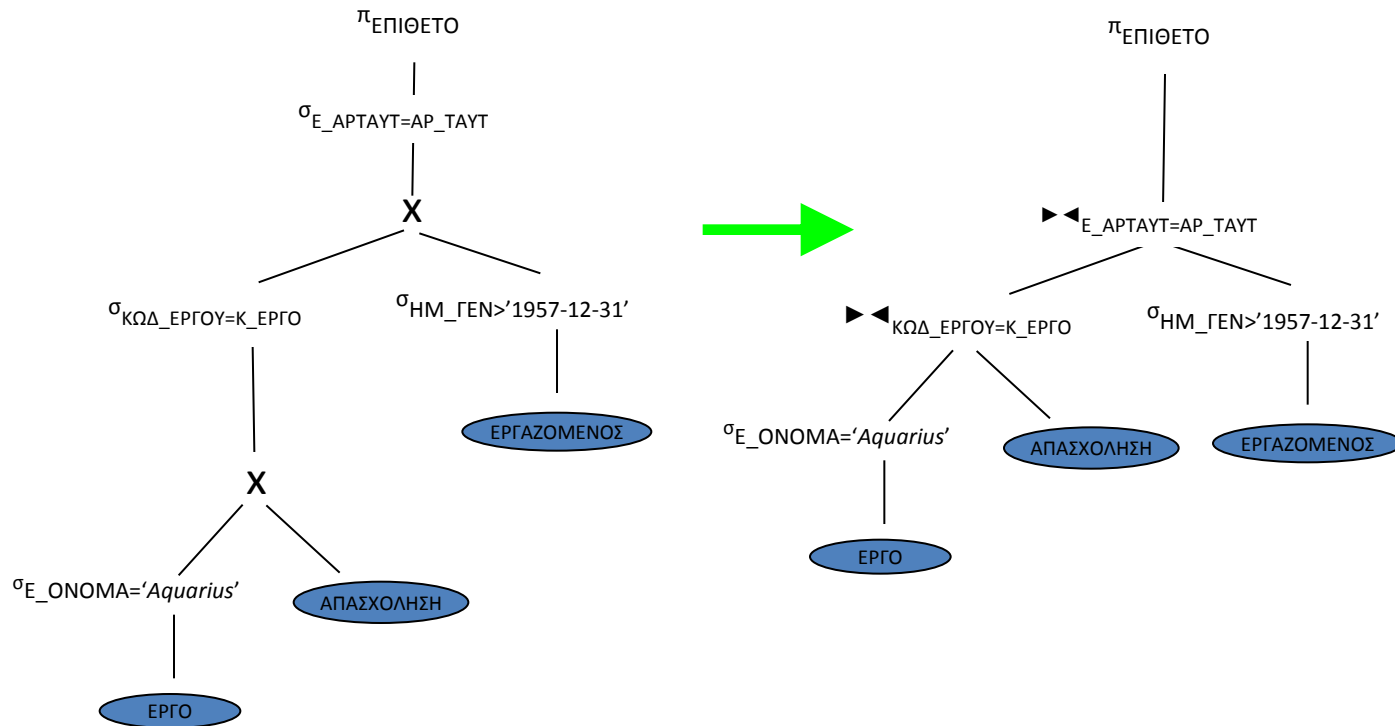


- Πρώτα πράξεις επιλογής



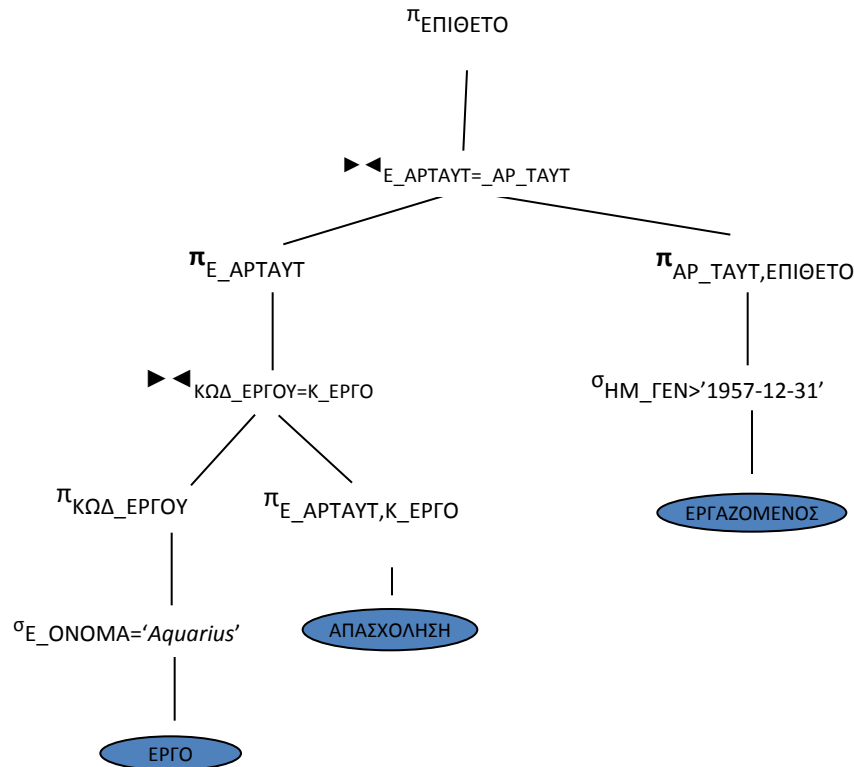
Παράδειγμα βελτιστοποίησης Δέντρο επερώτησης

- Αντιμετάθεση σχέσεων (πιο περιοριστικές πρώτα)
- Αντικατάσταση καρτεσιανού γινομένου με συνένωση



Παράδειγμα βελτιστοποίησης Δέντρο επερώτησης

- Μετακίνηση των πράξεων προβολής προς τα κάτω, στο δέντρο επερώτησης



Άσκηση 2

Έστω η πράξη επιλογής:

$\sigma_{e\text{-name}='George'}(\text{employee}),$

όπου $\text{employee}(e\text{-id}, e\text{-name}, e\text{-address}, e\text{-age}, e\text{-salary})$. Για τον υπολογισμό της επιλεκτικότητας (selection cardinality) θεωρείστε ότι έχουμε ομοιόμορφη κατανομή των τιμών.

Θεωρείστε:

$n_{\text{employee}} = 20.000$ εγγραφές,

$f_{\text{employee}} = 50$ εγγραφές ανά μπλοκ,

$V(e\text{-salary}, \text{employee}) = 100$ διακριτές εγγραφές,

$V(e\text{-age}, \text{employee}) = 40$ διακριτές εγγραφές,

$V(e\text{-name}, \text{employee}) = 100$ διακριτές εγγραφές,

- 1) Βρείτε το πλήθος των προσπελάσεων σε αριθμό μπλοκ
 - μιας ακολουθιακής σάρωσης
- 2) Αν χρησιμοποιήσουμε ένα αραιό ευρετήριο τύπου B-tree ως δομή δεικτοδότησης, για το δευτερεύον κλειδί e-name με fan-out=50:
 - a) Βρείτε το πλήθος των επιπέδων του ευρετηρίου
 - b) Βρείτε το πλήθος των προσπελάσεων σε αριθμό μπλοκ.

Λύση

1) $b_{employee} = n_{employee} / f_{employee} = 400$ disk accesses

2) a) $HT = \log(\#entries) / \log(fanout) =$
 $= \log(100) / \log(50) = 2$ επίπεδα

2) b) $HT_i + SC(ename, employee) / f_{employee} =$
 $= 2 + 200 / 50 = 6$ προσπελάσεις

$SC(ename, employee) = n_{employee} / V(e-name,$
 $employee) = 200$