Παράδειγμα υπολογισμού σύζευξης με βελτιστοποίηση

Γεώργιος Ευαγγελίδης

Σχήμα Πινάκων και Ερώτημα Σύζευξης

- **ΕΜΒΟΛΙΑΣΤΙΚΟ_ΚΕΝΤΡΟ (κωδ CHAR(10), όνομα CHAR(20), διεύθυνση CHAR(70))**
- **ΠΟΛΙΤΗΣ (ID CHAR(40), κωδΕΚ CHAR(10)),** όπου κωδΕΚ ξένο κλειδί αναφέρεται στο ΕΜΒΟΛΙΑΣΤΙΚΟ_ΚΕΝΤΡΟ(κωδ)

```
SELECT Π.ID, Ε.κωδ, Ε.όνομα
FROM ΕΜΒΟΛΙΑΣΤΙΚΟ_ΚΕΝΤΡΟ Ε, ΠΟΛΙΤΗΣ Π
WHERE Ε.κωδ = Π.κωδΕΚ
```

Μεγέθη Πινάκων, σελίδας και Ενδιάμεσης Μνήμης

- Ο Ε αποτελείται από 10000 εγγραφές των 100 bytes και ο Π από 50000 εγγραφές των 50 bytes.
- Έστω ότι το μέγεθος της σελίδας είναι 1000 bytes.
- Έστω ότι η ενδιάμεση μνήμη έχει 20 σελίδες (B=20).

Βοήθεια

- **πλήθος σελίδων Ε**:

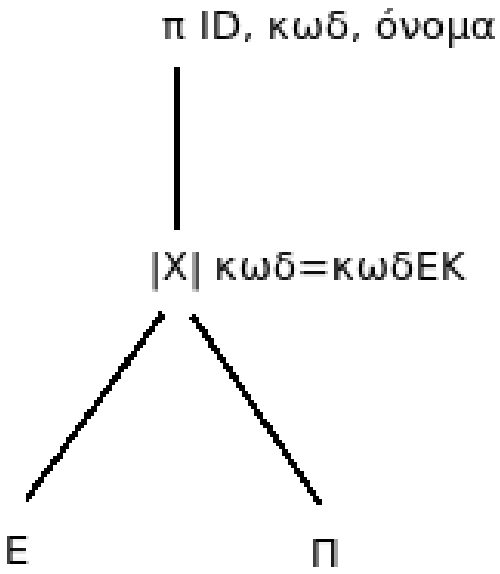
  (10000 εγγ * 100 bytes/εγγ)/1000 bytes/σελ = **1000 σελ**
- **πλήθος σελίδων Π**:

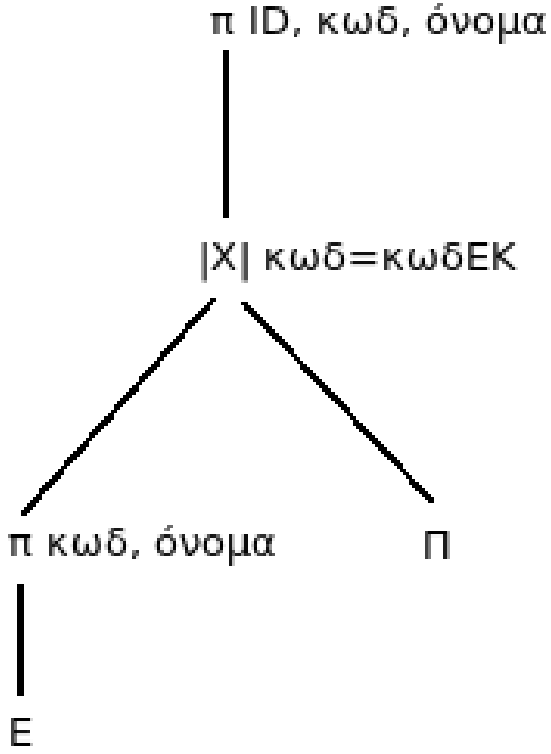
  (50000 εγγ * 50 bytes/εγγ)/1000 bytes/σελ = **2500 σελ**

Βελτιστοποίηση ερωτήματος

- Ισοδύναμη σχεσιακή άλγεβρα για το SQL ερώτημα:
  - **π Π.ID, Ε.κωδ, Ε.όνομα (σ Ε.κωδ=Π.κωδΕΚ (Ε Χ Π))**
- Για να ελαχιστοποιήσουμε το κόστος των συζεύξεων συμφέρει να “κρατήσουμε” για τον Ε μόνο τα απαραίτητα πεδία, δηλαδή τα **κωδ** και **όνομα**. Με άλλα λόγια, στο σχέδιο υπολογισμού εκτέλεσης θα “σπρώξουμε” μια προβολή για τον Ε προς τα κάτω, πριν τη σύζευξη.
- Ισοδύναμη σχεσιακή άλγεβρα
  - **π Π.ID, Ε.κωδ, Ε.όνομα (σ Ε.κωδ=Π.κωδΕΚ (π κωδ, όνομα (Ε) Χ Π))**

Δέντρο Σχεσιακής Άλγεβρας πριν/μετά τη βελτιστοποίηση





Σχέδια Εκτέλεσης χωρίς ευρετήρια

- Εκτιμήστε το κόστος του κάθε ενός σχεδίου εκτέλεσης του αιτήματος που δίνεται παρακάτω όταν δεν υπάρχουν ευρετήρια:
  - (α1) με Εμφώλευση Βρόχων κατά Σελίδα
  - (α2) με Εμφώλευση Βρόχων κατά Μπλοκ
  - (α3) με Ταξινόμηση και Συγχώνευση
  - (α4) με Κατακερματισμό

(α1) Εμφώλευση Βρόχων κατά Σελίδα [1]

- Εξ ορισμού χρησιμοποιεί μόνο **3** πλαίσια της ενδιάμεσης μνήμης.
  - **1** για να διαβάζει σελίδα-σελίδα τον εξωτερικό πίνακα,
  - **1** για να διαβάζει σελίδα-σελίδα ολόκληρο τον εσωτερικό πίνακα για κάθε σελίδα του εξωτερικού, και
  - **1** ως πλαίσιο εξόδου για τις συζευγμένες (ταιριαστές) εγγραφές των δυο πινάκων.

(α1) Εμφώλευση Βρόχων κατά Σελίδα [2]

- **Π εξωτερικός**
  - Πρώτα εκτελούμε την προβολή **Ε' = π κωδ, όνομα (Ε)**
  - Το μέγεθος του πίνακα Ε' είναι: (10000 εγγ * 30 bytes/εγγ)/1000 bytes/σελ = **300 σελ**
  - Κόστος προβολής = 1000 + 300 = **1300 I/O**
  - Κόστος σύζευξης = 2500 + 2500*300 = 2500 + 750000 = **752500 I/O**
  - Συνολικό κόστος = 1300 + 752500 = **753800 I/O**

(α1) Εμφώλευση Βρόχων κατά Σελίδα [3]

- **Ε εξωτερικός**
  - Εφόσον ο Ε θα διαβαστεί μόνο μια φορά, μπορούμε να κάνουμε την προβολή **π κωδ, όνομα (Ε)** εν πτήσει και να αποφύγουμε την αποθήκευση του προσωρινού Ε'
  - Κόστος σύζευξης = 1000 + 300*2500 = 1000 + 750000 = **751000 I/O**
  - Άρα συμφέρει να έχουμε τον Ε εξωτερικό

(α2) Εμφώλευση Βρόχων κατά Μπλοκ [1]

- Χρησιμοποιούνται **B-2=18** πλαίσια για τον εξωτερικό πίνακα, **1** για τον εσωτερικό και **1** για την έξοδο.
- **Π εξωτερικός**
  - Πρώτα εκτελούμε την προβολή **E' = π κωδ, όνομα (E)**
  - Το μέγεθος του πίνακα E' είναι: (10000 εγγ * 30 bytes/εγγ)/1000 bytes/σελ = **300 σελ**
  - Κόστος προβολής = 1000 + 300 = **1300 I/O**
  - Κόστος σύζευξης = 2500 + ceil(2500/18)*300 = 2500 + 139*300 = 2500 + 41700 = **44200 I/O**
  - Συνολικό κόστος = 1300 + 44200 = **45500 I/O**

(α2) Εμφώλευση Βρόχων κατά Μπλοκ [2]

- **Ε εξωτερικός**
  - Εφόσον ο Ε θα διαβαστεί μόνο μια φορά, μπορούμε να κάνουμε την προβολή **π κωδ, όνομα (Ε)** εν πτήσει και να αποφύγουμε την αποθήκευση του προσωρινού Ε'
  - **Κόστος σύζευξης** = 1000 + ceil(300/18)*2500 = 1000 + 17*2500 = 1000 + 42500 = **43500 I/O**
  - Άρα συμφέρει να έχουμε τον Ε εξωτερικό.

(α3) Ταξινόμηση και Συγχώνευση [1]

- Πρώτα ταξινομούμε τους Ε και Π και μετά συγχωνεύουμε με ένα πέρασμα.
- **Ταξινόμηση Ε (με ταυτόχρονη εφαρμογή της προβολής)**
  - **Πέρασμα 0**: Χρησιμοποιώ 20 πλαίσια και παράγω 15 συρμούς των 20 σελίδων (επειδή εξαιτίας της προβολής “ξεφορτώνομαι” το πεδίο **διεύθυνση**). Μετά χρειάζομαι ακόμα 1 πέρασμα που χρησιμοποιεί B-1=19 πλαίσια ως είσοδο και 1 ως έξοδο.
  - Το **Πέρασμα 1** παράγει ceil(15/19)=1 συρμό των 300 σελίδων, δηλαδή τον τελικό ταξινομημένο πίνακα Ε.

(α3) Ταξινόμηση και Συγχώνευση [2]

- **Ταξινόμηση Π**
  - **Πέρασμα 0**: Χρησιμοποιώ 20 πλαίσια και παράγω 125 συρμούς των 20 σελίδων. Μετά χρειάζομαι ακόμα 2 περάσματα που χρησιμοποιούν B-1=19 πλαίσια ως είσοδο και 1 ως έξοδο.
  - Το **Πέρασμα 1** παράγει ceil(125/19)=7 συρμούς, και
  - το **Πέρασμα 2** παράγει τον τελικό ταξινομημένο πίνακα Π.
- **Συνολικό κόστος = κόστος ταξ Ε + κόστος ταξ Π + κόστος συγχώνευσης =** (**1000+300**+2*300) + (2*2500*3) + (300 + 2500) = 1900 + 15000 + 2800 = **19700 I/O**
  - [Σημείωση: στην ταξινόμηση του Ε, στο Πέρασμα 0 διαβάζουμε 1000 σελίδες αλλά λόγω της προβολής γράφουμε 300. Στα επόμενα περάσματα γράφουμε και διαβάζουμε 300 σελίδες]

(α4) με Κατακερματισμό [1]

- Κατακερματίζουμε τον Ε ενώ εφαρμόζουμε την προβολή **π κωδ, όνομα (Ε)**.
- Χρησιμοποιούμε **1** πλαίσιο εισόδου και **Β-1=19** πλαίσια εξόδου (κάδους) για τη δημιουργία των **19** διαμερισμάτων.
- Υποθέτοντας “καλή” συνάρτηση κατακερματισμού θα πάρουμε **19** διαμερίσματα των ceil(300/19)=**16** σελίδων (οι σελίδες που δημιουργούνται είναι 300 και όχι 1000 επειδή εφαρμόζουμε την προβολή).

(α4) με Κατακερματισμό [2]

- Κατακερματίζουμε τον Π.
- Χρησιμοποιούμε **1** πλαίσιο εισόδου και **Β-1=19** πλαίσια εξόδου (κάδους) για τη δημιουργία των **19** διαμερισμάτων.
- Υποθέτοντας “καλή” συνάρτηση κατακερματισμού θα πάρουμε **19** διαμερίσματα των ceil(2500/19)=**132** σελίδων.

(α4) με Κατακερματισμό [3]

- Αφού **B > sqrt(300)** ή **20 > 17,3** θα χρησιμοποιήσουμε τον Ε ως εξωτερικό πίνακα. Χρησιμοποιούμε **B-2=18** πλαίσια για να φορτώνουμε κάθε ένα διαμέρισμα του Ε και **1** πλαίσιο για να σαρώνουμε το αντίστοιχο διαμέρισμα του Π. Έτσι οι Ε και Π σαρώνονται μια φορά για τη σύζευξη.
- **Συνολικό Κόστος =** κόστος κατακερματισμού Ε + κόστος κατακερματισμού Π + κόστος σύζευξης = (1000+300) + (2500+2500) + (300+2500) = 1300 + 5000 + 2800 = **9100 I/O**

Σχέδια Εκτέλεσης με ευρετήρια

- **Εκτιμήστε το κόστος του κάθε ενός σχεδίου εκτέλεσης του αιτήματος που δίνεται παρακάτω όταν υπάρχουν ευρετήρια:**
  - (β1) Υπάρχει μη-συγκροτημένο ευρετήριο κατακερματισμού Εναλλακτικής 2 πάνω στο Π.κωδΕΚ
  - (β2) Υπάρχει συγκροτημένο ευρετήριο κατακερματισμού Εναλλακτικής 1 στο Ε.κωδ

(β1) Υπάρχει ευρετήριο πάνω στο Π.κωδΕΚ

- Σαρώνουμε τον Ε και για κάθε εγγραφή του χρησιμοποιούμε το ευρετήριο για να βρούμε τις ταιριαστές εγγραφές στον Π. Θεωρώντας ομοιόμορφη κατανομή στις τιμές του πεδίου Π.κωδΕΚ κάθε εγγραφή του Ε θα έχει 5 ταιριαστές εγγραφές στον Π (10000 εγγραφές έχει ο Ε με το Ε.κωδ κλειδί, 50000 εγγραφές έχει ο Π με το Π.κωδΕΚ ξένο κλειδί).
- Θεωρώντας ότι έχουμε ευρετήριο κατακερματισμού χρειάζονται 1,2 I/O για την εύρεση μιας καταχώρισης δεδομένων και 1 I/O για την εύρεση της αντίστοιχης εγγραφής (σύνολο 1,2 + 5*1 = 6,2 I/O ανά 5 ταιριάσματα).
- **Κόστος σύζευξης** = 1000 + 10000*6,2 = **63000 I/O**

(β2) Υπάρχει ευρετήριο πάνω στο Ε.κωδ

- Σαρώνουμε τον Π και για κάθε εγγραφή του χρησιμοποιούμε το ευρετήριο για να βρούμε την ταιριαστή εγγραφή στον Ε. Αφού το ευρετήριο είναι Εναλλακτικής 1, το κόστος εύρεσης ταιριαστής εγγραφής είναι 1,2 I/O.
- **Κόστος σύζευξης** = 2500 + 50000*1,2 = **62500 I/O**


Τέλος!