

## 2η Σειρά Ασκήσεων

Δανοπούλου Αιμιλία - 3170033

### Άσκηση 1

45000/100 = 450 σελίδες

1.  $\lceil 450/4 \rceil = 113$  sublists

2. **6 περάσματα**

Το πρώτο (1) πέρασμα των 450 σελίδων απο τον δίσκο στην μνήμη και:

$\lceil 450/4 \rceil = 113$  (2),  $\lceil 113/4 \rceil = 29$  (3),  $\lceil 29/4 \rceil = 8$  (4),  $\lceil 8/4 \rceil = 2$  (5),  $\lceil 2/4 \rceil = 1$  (6)

(5 περάσματα στο merge στάδιο)

3.  $\text{Cost} = 2 * 450 * 6 = \mathbf{5400 \text{ I/O}}$

4. **M = 22** blocks , αφού  $\lceil 450/22 \rceil = 21, \lceil 21/22 \rceil = 1$

### Άσκηση 2

1. a.  $B(R) = 150$  ,  $B(S) = 90$

Επειδή  $T(R) > T(S)$  επιλέγουμε ως εξωτερική σχέση την S και άρα:

$\text{Cost(NLJ)} = 90 + \lceil 90/12 \rceil * 150 = \mathbf{1290 \text{ I/O}}$

b.  $\sqrt{B(R) + B(S)} \leq M \Rightarrow \sqrt{150 + 90} \leq 13 \Rightarrow 15.49 \leq 13$  , δεν ισχύει και άρα δεν μπορούμε να χρησιμοποιήσουμε τον efficient SMJ.

$\sqrt{\max(150,90)} \leq 13 \Rightarrow 12.24 \leq 13$  , ισχύει

$\text{Cost(SMJ)} = 5(B(R) + B(S)) = \mathbf{1200 \text{ I/O}}$

c.  $\text{Cost(Hash-Join)} = 3(B(R) + B(S)) = 3(240) = \mathbf{720 \text{ I/O}}$

Επομένως, ο **βέλτιστος** αλγόριθμος για το επερώτημα είναι ο **Hash Join**.

2. Για  $M = 100$

•  $\text{Cost(NLJ)} = \lceil 90/99 \rceil * 150 + 90 = \mathbf{240 \text{ I/O}}$

Ισχύει  $\sqrt{150 + 90} \leq 100$ , άρα χρησιμοποιούμε τον SMJ efficient

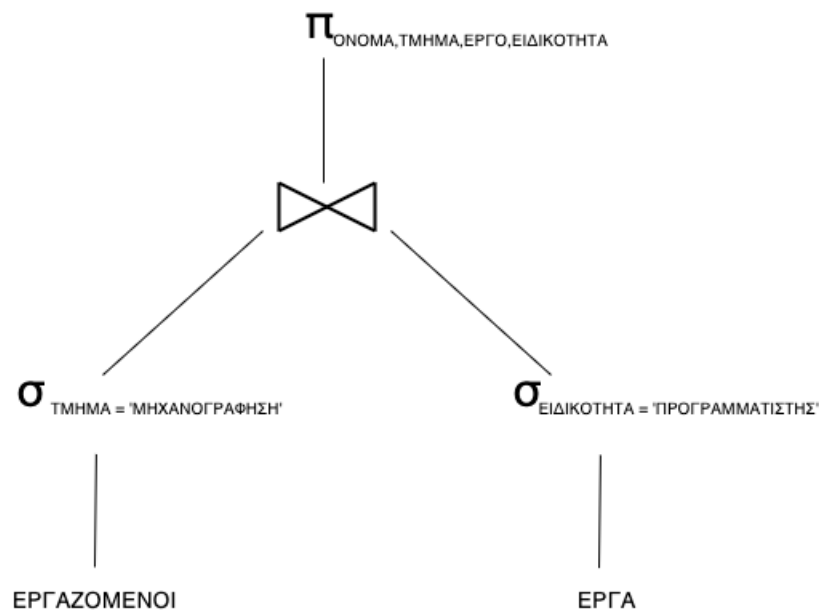
•  $\text{Cost(SMJ)} = 3(B(R) + B(S)) = 3(240) = \mathbf{720 \text{ I/O}}$

Εφόσον η S χωράει ολόκληρη στην μνήμη το κόστος θα είναι

•  $\text{Cost(Hash-Join)} = B(R) + B(S) = 150 + 90 = \mathbf{240 \text{ I/O}}$

### Άσκηση 3

#### 1. Βέλτιστο πλάνο



2.

SMJ

Έστω  $X = \sigma(\text{τμήμα} = \text{'μηχανογράφηση'})$

Και  $Y = \sigma(\text{ειδικότητα} = \text{'προγραμματιστής'})$

Δεδομένου ότι υπάρχουν 10 διαφορετικά τμήματα, 5 διαφορετικές ειδικότητες και τα δεδομένα κατανέμονται ομοιόμορφα:

- $T(X) = 5000/10 = 500$  records
- $B(X) = 500 (5000/250) = 25$  σελίδες
- $T(Y) = 2000/5 = 400$  records
- $B(Y) = 400 / (2000/50) = 10$  σελίδες

Επειδή υπάρχει ευρετήριο συστάδων (clustered index) στο πεδίο  $\text{ΕΡΓΑΖΟΜΕΝΟΙ.ΤΜΗΜΑ}$ , το κόστος για το  $X$  θα είναι:

$$\text{Cost}(X) = B(X) = 25$$

Επειδή υπάρχει απλό ευρετήριο (non- clustered index) στο πεδίο ΕΡΓΑ.ΕΙΔΙΚΟΤΗΤΑ, το κόστος για το Y θα είναι:

$$\text{Cost}(Y) = T(Y) = 400$$

### SMJ

$$\text{Ισχύει } \sqrt{B(X) + B(Y)} \leq 8 \Rightarrow \sqrt{35} \leq 8$$

Επομένως χρησιμοποιήσουμε τον αποδοτικό SMJ

$$\text{Cost}(\text{SMJ}) = \text{Cost}(X) + \text{Cost}(Y) + 2(B(X)+B(Y)) = 25 + 400 + 2(25+10) = \mathbf{495 \text{ I/O}}$$

### NLJ

Επειδή  $T(X) = 500$  και  $T(Y) = 400$  επιλέγουμε ως εξωτερική σχέση την Y

$$\text{Cost}(\text{NLJ}) = \text{Cost}(Y) + \lceil B(Y)/7 \rceil \text{Cost}(X) = 400 + \lceil 10/7 \rceil 25 = \mathbf{450 \text{ I/O}}$$

### Άσκηση 4

Έστω  $X = \sigma(\text{πόλη} = \text{'Αθήνα'})$  και  $Y = \sigma(\text{κατάταξη} < 10)$

1.  $\text{Cost}(X) = B(\text{ΦΟΙΤΗΤΕΣ}) = \mathbf{100}$  (αφού δεν υπάρχει κάποιο ευρετήριο)

- $T(X) = 2000/20 = 100$  εγγραφές
- $B(X) = 100/(2000/100) = 5$  σελίδες

2.  $\text{Cost}(Y) = B(\text{ΠΑΝΕΠΙΣΤΗΜΙΑ}) = \mathbf{10}$  (αφού δεν υπάρχει κάποιο ευρετήριο)

3.

$$T(Y) = 9 \text{ records}$$

$$B(Y) = 1 \text{ σελίδα (αφού } 100/10 = 10 \text{ εγγραφές ανα σελίδα)}$$

$$\sqrt{B(X) + B(Y)} \leq 8 \Rightarrow \sqrt{6} \leq 8, \text{ ισχύει, επομένως χρησιμοποιούμε SMJ efficient:}$$

$$\text{Cost}(\text{SMJ}) = \text{Cost}(X) + \text{Cost}(Y) + 2(B(X)+B(Y)) = 100 + 10 + 2(6) = \mathbf{122 \text{ I/O}}$$

4.  $T(\text{ΦΟΙΤΗΤΕΣ})/V(\text{ΚΠ}) = 2000/100 = 20$  εγγραφές για κάθε κατάταξη στον πίνακα φοιτητές

- $V(\text{ΚΠ}) = 100$ , αφού το πεδίο ΦΟΙΤΗΤΕΣ.ΚΠ είναι ξένο κλειδί το οποίο αναφέρεται στο πεδίο ΠΑΝΕΠΙΣΤΗΜΙΑ.ΚΠ

Άρα για τις 9 πρώτες κατατάξεις οι εγγραφές στον πίνακα φοιτητές θα είναι:

- $9 \cdot 20 = 180$  εγγραφές
- $T(X) = 100$  εγγραφές για την πόλη Αθήνα στον πίνακα φοιτητές

Για να βρω πόσες εγγραφές είναι η τομή αυτών των δύο :

$P(A \cap B) = P(A)P(B) = \frac{100}{2000} * \frac{180}{2000} = 0.0045$  η πιθανότητα ένας φοιτητής να μένει στην Αθήνα και να πηγαίνει σε ένα απο τα 9 καλύτερα πανεπιστήμια.  
(ανεξάρτητα ενδεχόμενα)

$$2000 \cdot 0.0045 = 9 \text{ εγγραφές}$$

- $T(\text{New}) = 9$  records
  - $B(\text{New}) = 1$  page
  - $T(\text{ΠΡΟΕΓΓΡΑΦΕΣ}) = 3000$
  - $B(\text{ΠΡΟΕΓΓΡΦΕΣ}) = 200$
- $$\text{Prob} = \lceil \log_{200} 3000 \rceil = 2$$

$$\text{Cost}(\text{INLJ}) = 1 + 9 \cdot \text{prob} = 1 + 18 = \mathbf{19 \text{ I/O}}$$

$$5. Z = \mathbf{\sigma}(\text{ΜΠΣ} = \text{'DS'})$$

$$\text{Cost}(Z) = B(\text{New}) = \mathbf{1}$$

Το 6 δεν έχει κάποιο επιπλέον κόστος.