

Σχεδιασμός Βάσεων Δεδομένων

Τοσκολλάρι Ρόναλντ

3160244

Άσκηση 1

1) Αρχικά έχουμε $45.000/100 = 450$ σελίδες στο δίσκο μας. Άρα το $B(R) = 450$. Έχουμε αρχικά $M = 4$ blocks, άρα κάθε φορά ο two-phase sort παίρνει λίστες ίσο με $M = 4$ blocks και τις ταξινομεί στην μνήμη και τα γράφει στο δίσκο. Στην φάση 1 δηλαδή θα έχω $450/4 = 113$ λίστες ταξινομημένες στο δίσκο. Επειδή $112 * 4 = 448$ έχουμε ότι οι πρώτες 112 θα έχουν μέγεθος 4 block, ενώ η τελευταία θα έχει μέγεθος ίσο με 2 block.

2) Αρχικά έχουμε απο το (1) οτι έχουμε 113 λίστες το 4 block, επίσης ισχύει οτι $M = 4$ blocks, άρα με κάνουμε 5 περάσματα διότι $113/4 = 28.25$, $28.25/4 = 7.0625$, $7.0625/4 = 1.765625$, $1.765625/4 = 0.44140625$.

3) Το κόστος είναι ίσο με $3 * B(R) = 3 * 450 = 1350$.

3) Πρέπει $M \geq B(R)^{(1/2)}$, άρα πρέπει $450^{(1/2)} = 21.2132034356 \sim 22$ blocks.

Άσκηση 2

1) Το κόστος (b) Sort-Merge Join είναι $5 * (B(R1) + B(R2)) = 5 * (150 + 90) = \mathbf{1200}$. Δεν μπορούμε να έχουμε την βελτιωμένη έκδοση του αλγορίθμου διότι έχουμε $M = 13$ και

δεν φτάνει. Για την απλή έκδοση έχουμε απαιτήσεις μνήμης $12.24 < 13$.

Το κόστος (a) Nested Loop Join είναι $B(R2) + \text{nextBigger}(B(R2)/M-1)*B(R1) = 90 + 8*150 = \mathbf{1290}$

Αν πάρω R1 εξωτερική σχέση έχω: $B(R1) + \text{nextBigger}(B(R1)/M-1)*B(R2) = 150 + 13*90 = \mathbf{1320}$

Το κόστος (c) Hash Join είναι :

Total Cost = $3*(150+90) = \mathbf{720}$ και έχω απαιτήσεις σε μνήμη $\text{sqrt}(90) \sim 9.48 = 10$. Άρα ο καλύτερος αλγόριθμος είναι το Hash Join.

1) Αν έχω $M = 100$ τότε προκύπτει:

Για το (a): $B(R2) + \text{nextBigger}(B(R2)/M-1)*B(R1) = 90 + 1*150 = \mathbf{240}$. Ομοίως αν R1 ήταν εξωτερική.

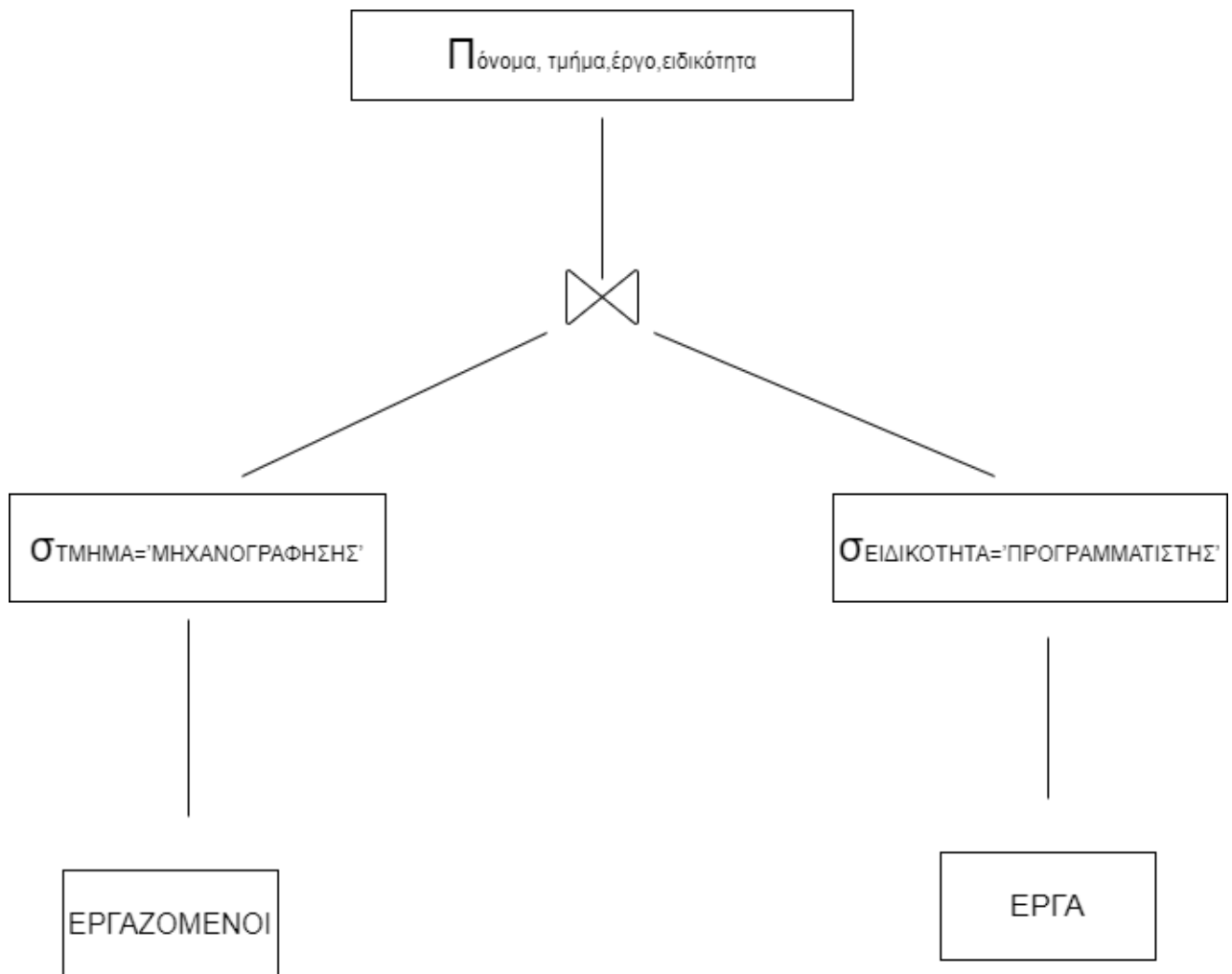
Για το (b): Τώρα μπορούμε να έχουμε την καλύτερη έκδοση του SMJ άρα κόστος = $3*240 = \mathbf{720}$.

Για το (c) : Βλέπουμε ότι μια απο τις δύο σχέσεις(R2) χωράει στην μνήμη άρα έχω κόστος = $150 + 90 = \mathbf{240}$.

Άρα οι καλύτεροι αλγόριθμοι είναι οι (a) και (c), δηλαδή ο Nested Loop Join και Hash Join.

Άσκηση 3

1)



2) Έστω $R1 = \text{Εργαζόμενοι}$ και $R2 = \text{Έργα}$.

Στην περίπτωση του Non Clustered Index έχω:

Cost: $T(\sigma_{\text{ειδικότητα}="προγραμματιστής"}) = T(R2)/V(R2, \text{Ειδικότητα}) = 2000/5 = 400$.

Στην περίπτωση του Clustered index έχω:

Cost: $T(\sigma_{\text{τμήμα}="μηχανογράφηση"}) = B(R1)/V(R1, \text{Τμήμα}) = 250/10 = 25$.

A) SMJ: $25 + 400 + 5 \cdot (25 + 400) = 2550$.

B) NLJ: $25 + 400 + \text{nextBigger}(25/7) = 1700$.

Άσκηση 4

Έστω $R1 = \text{Φοιτητές}$, $R2 = \text{Πανεπιστήμια}$, $R3 = \text{Προεγγραφές}$

1) Κόστος $\sigma_{\text{πόλη}=\text{'Αθήνα'}} = T(R1)/V(R1, \text{Πόλη}) = 2000/20 = 100$.

2) Κόστος $\sigma_{\text{κατάταξη} < 10} = T(R2)/V(R2, \text{Κατάταξη}) = 100/10 = 10$.

3) SMJ: $100 + 10 + 5 * (110) = 660$. Εφόσον $110 > 64$.

4) Εφόσον έχουμε non-clustered index στις προεγγραφές στο όρισμα που γίνεται join προκύπτει ότι ο INLJ είναι: $660 + 2000 * 2 = 4660$.

5) Κόστος $\sigma_{\text{ΜΠΣ}=\text{'DS'}} = 4660/5 = 932$

6) Το κόστος προβλής των δεν υπάρχει διότι δεν κάνω I/O.