



Σχεδιασμός Βάσεων Δεδομένων

Διδάσκων: Ιωάννης Κωτίδης

Εαρινό εξάμηνο 2020-2021

Δεύτερη Σειρά Ασκήσεων

Ανάθεση: 10-05-2021

Παράδοση: 23-05-2021 Ώρα (23:55)

Οδηγίες

- Η δεύτερη σειρά ασκήσεων είναι **ατομική** και **υποχρεωτική**.
- Η υποβολή της εργασίας πρέπει να γίνει στο *eclass*.
- Το παραδοτέο σας θα πρέπει να είναι ένα αρχείο PDF με όνομα *AM.pdf* (όπου *AM* είναι ο αριθμός μητρώου σας. π.χ. "3180001.pdf").
- Τα διαγράμματα πρέπει να είναι κατασκευασμένα σε κάποιο πρόγραμμα (της επιλογής σας) και όχι σκαναρισμένα χειρόγραφα.
- Πιθανή αντιγραφή θα τιμωρείται με μηδενισμό όλων των εμπλεκομένων.
- Για την επίλυση των ασκήσεων να μελετήσετε τις διαφάνειες των διαλέξεων του μαθήματος.

ΑΡΙΣΤΕΙΔΗΣ ΧΡΟΝΟΠΟΥΛΟΣ 3160194

Ασκηση 1)

1. a)

- i) Έχουμε clustered index άρα $B(R) * 1/V(R,a) = 20.000 * 1/n = 20.000 / n$ I/Os.
- ii) Έχουμε non clustered index άρα $T(R) * 1/V(R,a) = 1.000.000 * 1/n = 1.000.000/n$ I/Os.

b)

i) $B(R) * n/10 = 2000n$ I/Os

ii) $T(R) * n/10 = 100.000n$ I/Os.

2. a) Table scan = $B(R) = 20.000$ I/Os.

Όταν το n είναι πολύ μικρό δεν συμφέρει να κανουμε index.

b) Table scan = $B(R) = 20.000$ I/Os.

Ασκηση 2)

1. Κοινές τιμές έχουμε στα διαστήματα $[21,40]$ $[41,60]$ $[61,80]$

Επομένως για το διάστημα $[21,40]$ έχουμε:

Για τον πίνακα R-> 80 τιμές σε διάστημα 20 τιμών. Άρα το κάθε ένα θα έχει 4

Για τον πίνακα S->100 τιμές σε διάστημα 20 τιμών. Άρα 5 το κάθε ένα.

Επομένως για τους 2 πίνακες στο διάστημα $[21,40]$ έχουμε $4*5*20$ (διάστημα)
= 400.

Για το διάστημα $[41,60]$ έχουμε:

Για τον πίνακα R-> Με την ίδια λογική 100 για 20. Άρα 5 το κάθε ένα.

Για τον πίνακα S->Αντίστοιχα εδώ 60 για 20. Άρα 3 το κάθε ένα.

Επομένως για τους 2 πίνακες στο διάστημα $[41,60]$ έχουμε $3*5*20 = 300$.

Για το διάστημα $[61,80]$ έχουμε:

Για τον πίνακα R-> Με την ίδια λογική 20 για 20.Άρα 1 το κάθε ένα.

Για τον πίνακα S->Αντίστοιχα εδώ 60 για 20.Άρα 3 το κάθε ένα.

Επομένως για τους 2 πίνακες στο διάστημα [61,80] έχουμε $1*3*20 = 60$.

Άρα συνολικά θα επιστραφούν 760 πλειάδες.

2. Εδώ υποθέτουμε ομοιόμορφη κατανομή έτσι έχουμε

$$\frac{T(R) \times T(S)}{\max(V(R,b), V(S,b))}$$

Δηλαδή $(230*230)/100=529$.

Άσκηση 3)

1.

Block Nested Loop Join:

Έχουμε $B(R) + [B(R)/(M-1)] * B(S) = 800 + (800/(41-1)) * 1500 = 3.800$ I/Os.

Sort-Merge Join:

Δεν μπορούμε να έχουμε την αποδοτική μορφή του αλγορίθμου διότι το μέγεθος του ενταμιευτή της μνήμης είναι μικρό.

Έχουμε $800/41 = 20$ sublists και $1500/41 = 37$ sublists, $20+35 = 57 > 41$

Άρα $5 * (B(R) + B(S)) = 5(800+1500)=11.500$ I/Os.

Hash Join:

Το μέγεθος του ενταμιευτή μνήμη είναι μικρό,καμία από τις σχέσης δεν χωράει στην μνήμη. $M < \min(B(R),B(S))$.

Όμως $M < \sqrt{\min(B(R),B(S))}$ και εφαρμόζουμε hash join στην μνήμη.

Άρα $3(B(R) + B(S)) = 3(800 + 1500) = 6.900$ I/Os.

2.

Δύο τρόποι που θα μπορούσαν να να μειώσουν το κόστος είναι να μειώσουμε το μέγεθος της μνήμης M ή να είναι ήδη ταξινομημένες οι τιμές ώστε να αποφύγουμε το sort.

Άσκηση 4)

1. Έχουμε index scan άρα $T(\text{BIBΛΙΟ}) / V(\text{BIBΛΙΟ}, \text{ΕΚΔΟΤΗΣ}) = 50.000/500 = 100$ I/O. Άρα συμπαιρνούμε και απο τους πίνακες οτι έχουμε $10(50.000/5000)*100 = 1000$ εγγραφές.
2. Στο 2 έχουμε INLJ οπότε σύμφωνα με τον τύπο έχουμε $100(\text{σελιδες}) + 1000(\text{εγγραφές})*(2/20) = 200$ I/O.Επίσης έχουμε 300.000 εγγραφες απο το πινακα ΔΑΝΕΙΖΟΜΕΝΟΙ / $500(\text{εκδοτες}) = 600$ εγγραφές.
3. Εδώ έχουμε NLJ όπου έχουμε $\text{Cost}(\text{outer}) + [B(x)/(M-1)]*\text{Cost}(\text{inner}) = 200(\text{κόστος απο το 2}) + [10 / (20-1)] * 1000(\text{table scan αρα διαβάζει ολες τις σελίδες}) \approx 700$ I/O. Εδώ για τις εγγραφές παραμένουν 600 αφού θεωρούμε οτι καθε δανειζόμενος αντιστοιχεί ακριβώς σε μια απο τις 600 εγγραφες απο το ερώτημα 2.
Το $B(x)$ δλδ τις σελίδες τις υπέθεσα 10 σύμφωνα με τους πίνακες που έχουμε ως την ελάχιστη τιμή που θα μπορούσε να έχει.
4. 0 I/O.Δεν επηρεάζει το κόστος αφού όλες οι σελίδες χωράνε στην μνήμη και απλά τις εμφανίζει.Οσων αφορά τις εγγραφές έχουμε τα εξης στοιχεία.Έχουμε οτι οι δανειζόμενοι κυμαίνονται απο 7 έως 24 έτη.Αυτες είναι 18 τιμές.Και εφόσον απο το ερώτημα 3 έχουμε 600 εγγραφές προκύπτει οτι $T(x)/V(\text{ΔΑΝΕΙΖΟΜΕΝΟΙ}, \text{ΗΛΙΚΙΑ}) = 600 / 18 \approx 34$.
Το 34 είναι όμως το αποτέλεσμα για μία τιμή ηλικίας απο το διάστημα που έχουμε.Εμείς θέλουμε $\sigma_{12 < \text{ηλικια} < 20}$, δηλαδή για 7 τιμές.

Επομένως $34 * 7 = 238$ εγγραφές.