

## Σχεδιασμός Βάσεων Δεδομένων

Διδάσκων: Ιωάννης Κωτίδης

Εαρινό εξάμηνο 2018-2019

### Δεύτερη Σειρά Ασκήσεων

Ανάθεση: 09-05-2019

Παράδοση: 19-05-2019 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. "3170001.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.

#### Άσκηση 1

Έστω οι σχέσεις R1(a, b,c) και R2(c,d,e) για τις οποίες ισχύουν τα εξής:

- Η σχέση R1 έχει 20000 εγγραφές και σε μία σελίδα χωράνε 100 εγγραφές της σχέσης R1.
- Η σχέση R2 έχει 50000 εγγραφές και σε μία σελίδα χωράνε 100 εγγραφές της σχέσης R2.
- Το μέγεθος του διαθέσιμου ενταμιευτή μνήμης (memory buffer) είναι M=51 σελίδες.

#### Ζητείται:

1. Να εκτιμήσετε το κόστος (σε I/O) της σύζευξης R1 ⋈ R2 χρησιμοποιώντας τους παρακάτω αλγορίθμους:
  - a. NLJ (Nested Loop Join)
  - b. SMJ (Sort Merge Join)
  - c. Hash Join
2. Ποιό είναι το **ελάχιστο δυνατό** κόστος (σε I/O) για την παραπάνω σύζευξη χρησιμοποιώντας οποιονδήποτε από τους παραπάνω αλγορίθμους, και ποιο είναι το απαιτούμενο μέγεθος (M) του ενταμιευτή ώστε να επιτευχθεί το ελάχιστο δυνατό κόστος;
3. Έστω ότι στη σχέση R2 το γνώρισμα c είναι το πρωτεύον κλειδί. Ποιό είναι το μέγιστο πλήθος των εγγράφων που παράγει η παρακάτω επερώτηση και πόσες σελίδες απαιτούνται για την αποθήκευσή τους;

```
SELECT * FROM R1, R2 WHERE R1.c = R2.c
```

## Άσκηση 2

Δίνονται οι παρακάτω σχέσεις:

ΦΟΙΤΗΤΕΣ (ΑΜ, ΕΠΩΝΥΜΟ, ΟΝΟΜΑ, ΤΜΗΜΑ)  
ΜΑΘΗΜΑΤΑ (ΑΜ, ΜΑΘΗΜΑ, ΕΞΑΜΗΝΟ)

Η σχέση ΦΟΙΤΗΤΕΣ περιέχει 20000 φοιτητές οι οποίοι στο σύνολο τους φοιτούν σε 20 διαφορετικά τμήματα (κάθε φοιτητής είναι εγγεγραμμένος σε ένα τμήμα).

Η σχέση ΜΑΘΗΜΑΤΑ περιέχει 100000 εγγραφές, κάθε φοιτητής δηλώνει συνολικά 5 μαθήματα από ένα η περισσότερα εξάμηνα, ενώ υπάρχουν δέκα διαφορετικά εξάμηνα (1, 2, 3, ..., 10).

Το σύνολο των εγγραφών της σχέσης ΦΟΙΤΗΤΕΣ χωράνε σε 1000 σελίδες και το σύνολο των εγγραφών της σχέσης ΜΑΘΗΜΑΤΑ χωράνε σε 2500 σελίδες αντίστοιχα.

Τέλος θεωρείστε ότι υπάρχει ένα ευρετήριο συστάδων (cluster index) στο γνώρισμα ΦΟΙΤΗΤΕΣ.ΤΜΗΜΑ και ένα απλό ευρετήριο (Non cluster index) στο γνώρισμα ΜΑΘΗΜΑΤΑ.ΕΞΑΜΗΝΟ.

Ζητείται:

1. Να σχεδιάσετε το αρχικό λογικό πλάνο της παρακάτω επερώτησης:

```
SELECT ΕΠΩΝΥΜΟ, ΟΝΟΜΑ, ΜΑΘΗΜΑ, ΕΞΑΜΗΝΟ  
FROM ΦΟΙΤΗΤΕΣ, ΜΑΘΗΜΑΤΑ  
WHERE ΦΟΙΤΗΤΕΣ.ΑΜ=ΜΑΘΗΜΑΤΑ.ΑΜ AND  
      ΦΟΙΤΗΤΕΣ.ΤΜΗΜΑ='Πληροφορική' AND ΜΑΘΗΜΑΤΑ.ΕΞΑΜΗΝΟ=5
```

2. Να σχεδιάσετε το τελικό, βελτιστοποιημένο λογικό πλάνο της ίδιας επερώτησης. Δεν χρειάζεται να παρουσιάσετε τα ενδιάμεσα βήματα.
3. Δεδομένου ότι μπορεί να χρησιμοποιηθεί ένας ενταμιευτής μνήμης (buffer) μεγέθους  $M=16$  σελίδων, ποιά είναι το ελάχιστο κόστος (σε I/O) εκτέλεσης της παραπάνω επερώτησης χρησιμοποιώντας τους αλγορίθμους α) SMJ και β) NLJ;

## Άσκηση 3

Έστω η σχέση R(Όνομα, Πόλη, Αποθεματικό) η οποία περιέχει τα στοιχεία των υποκαταστημάτων μιας τράπεζας. Υποθέστε ότι υπάρχει ένα και μοναδικό ευρετήριο B+ δένδρου στο γνώρισμα Πόλη. Ποιός είναι ο καλύτερος τρόπος να χειριστείτε τις παρακάτω επιλογές που περιλαμβάνουν άρνηση; Περιγράψτε με χρήση φυσικής γλώσσας.

α.  $\sigma_{\neg(\text{Πόλη} < \text{"Ιωάννινα"})}(R)$

β.  $\sigma_{\neg(\text{Πόλη} = \text{"Ιωάννινα"})}(R)$

γ.  $\sigma_{\neg(\text{Πόλη} < \text{"Ιωάννινα"} \vee \text{Αποθεματικό} < 50000)}(R)$

όπου:

$\neg$  τελεστής άρνησης (NOT) και

$\vee$  τελεστής διάζευξης (OR)