

**Συστήματα Διαχείρισης και Ανάλυσης Δεδομένων**  
**Διδάσκων: Ιωάννης Κωτίδης**

Εαρινό εξάμηνο 2025-2026

**Δεύτερη Σειρά Ασκήσεων**

Ανάθεση: 11-05-2026

Παράδοση: 24-05-2026 Ώρα (23:55)

*Οδηγίες*

- Η δεύτερη σειρά ασκήσεων είναι **ατομική**.
- 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. "3230001.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.**

**ΠΡΟΣΟΧΗ: Δεν θα αξιολογηθούν:**

- **Εργασίες που περιέχουν χειρόγραφο κείμενο ή χειρόγραφα διαγράμματα.**
- **Εργασίες οι οποίες δεν είναι σε μορφή ενός αρχείου pdf με όνομα *AM.pdf*.**
- **Εργασίες που δεν περιλαμβάνουν στην πρώτη σελίδα αριθμό μητρώου και ονοματεπώνυμο.**

**ΑΣΚΗΣΗ 1 [30 μονάδες]**

Δίνονται οι ακόλουθοι πίνακες:

- **R** ( *a* char(10), *b* char(20), *c* char(70)), με πρωτεύον κλειδί το πεδίο *a*.
- **S** ( *a* char(10), *f* char(40)), όπου το *a* είναι ξένο κλειδί που αναφέρεται στο πεδίο **R.a**

Ο πίνακας R περιέχει 3600 εγγραφές και ο πίνακας S περιέχει 7600 εγγραφές.

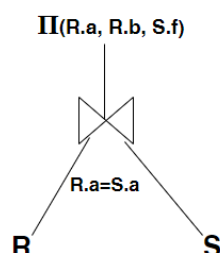
Θεωρείστε ότι ισχύουν τα ακόλουθα:

- το μέγεθος της σελίδας είναι 1000 bytes,
- το μέγεθος της διαθέσιμης μνήμης είναι 30 σελίδες (**M=30**),
- δεν υπάρχει κανένα ευρετήριο,
- τα πεδία των πινάκων R και S δεν δέχονται τιμές NULL

Έστω το ακόλουθο SQL επερώτημα:

**Q1: SELECT R.a, R.b, S.f FROM R JOIN S ON R.a=S.a**

και το αντίστοιχο λογικό πλάνο (**P1**):



Ζητείται:

- A. Να υπολογίσετε το κόστος της ισοσύνδεσης για κάθε έναν από τους παρακάτω αλγορίθμους:
1. NLJ (Block Nested Loop Join)
  2. SMJ (Sort Merge Join)
  3. Hash Join
- B. Να εξετάσετε αν υπάρχει καλύτερο λογικό πλάνο από το **P1** για το ερωτήμα **Q1**. Σε περίπτωση που υπάρχει να το σχεδιάσετε, να εξηγήσετε τους λογικούς μετασχηματισμούς που εφαρμόσατε και να επιλέξετε τον αποδοτικότερο από τους παραπάνω αλγορίθμους (NLJ, SMJ, Hash Join) για την εκτέλεσή του. Να τεκμηριώσετε την απάντησή σας υπολογίζοντας το κόστος I/O. Σε διαφορετική περίπτωση να αιτιολογήσετε γιατί το **P1** δεν μπορεί να βελτιωθεί.

## ΑΣΚΗΣΗ 2 [35 μονάδες]

Έστω οι σχέσεις:

ΑΣΘΕΝΗΣ(ΚΑ, ΟΝΟΜΑ, ΕΠΩΝΥΜΟ, ΗΛΙΚΙΑ)

ΕΠΙΣΚΕΨΗ(ΚΑ, ΚΛΙΝΙΚΗ, ΗΜΕΡΟΜΗΝΙΑ, ΔΙΑΓΝΩΣΗ)

για τις οποίες ισχύουν τα εξής:

- Η σχέση ΑΣΘΕΝΗΣ περιέχει 24000 εγγραφές και σε μία σελίδα χωράνε 40 εγγραφές.
- Η σχέση ΕΠΙΣΚΕΨΗ περιέχει 460800 εγγραφές και σε μία σελίδα χωράνε 32 εγγραφές.

Επιπλέον θεωρήστε ότι:

- Το νοσοκομείο διαθέτει 24 διαφορετικές κλινικές.
- Η σχέση ΕΠΙΣΚΕΨΗ περιέχει επισκέψεις για χρονικό διάστημα 400 ημερών.
- Οι επισκέψεις κατανέμονται ομοιόμορφα στις κλινικές και στις ημερομηνίες.
- Στο πεδίο ΗΛΙΚΙΑ της σχέσης ΑΣΘΕΝΗΣ το DBMS τηρεί το ακόλουθο ιστόγραμμα:

Ηλικία	Αριθμός Εγγραφών
[0..17]	1200
[18..34]	2400
[35..49]	3600
[50..64]	4800
[65..79]	6750
[80..99]	5250

- Υπάρχει ευρετήριο συστάδων (clustered index) B+ δέντρο στο πεδίο ΑΣΘΕΝΗΣ.ΗΛΙΚΙΑ.
- Υπάρχει απλό ευρετήριο (non-clustered index) B+ δέντρο στο πεδίο ΕΠΙΣΚΕΨΗ.ΚΛΙΝΙΚΗ.
- Τα ευρετήρια βρίσκονται στη μνήμη.
- Η διαθέσιμη μνήμη είναι  $M = 16$  σελίδες.
- Οι επιλογές είναι μεταξύ τους ανεξάρτητες.
- Όπου απαιτείται, υποθέστε ότι τα δεδομένα κατανέμονται ομοιόμορφα.

Δίνεται το επερώτημα:

**SELECT \***

**FROM** ΑΣΘΕΝΗΣ JOIN ΕΠΙΣΚΕΨΗ ΟΝ ΑΣΘΕΝΗΣ.ΚΑ = ΕΠΙΣΚΕΨΗ.ΚΑ

**WHERE** (ΗΛΙΚΙΑ >= 45 **AND** ΗΛΙΚΙΑ <= 72) **AND** (ΚΛΙΝΙΚΗ = 'Καρδιολογική') **AND**  
(ΗΜΕΡΟΜΗΝΙΑ >= '01-03-2025'**AND** ΗΜΕΡΟΜΗΝΙΑ <='09-04-2025')

Ζητείται:

- A. Να σχεδιάσετε το τελικό, βελτιστοποιημένο λογικό πλάνο της παρακάτω επερώτησης. Δεν χρειάζεται να δείξετε τα ενδιάμεσα βήματα.
- B. Να υπολογίσετε το ελάχιστο κόστος (σε I/O) εκτέλεσης της επερώτησης χρησιμοποιώντας του αλγορίθμους α) SMJ (Sort Merge Join) και β) NLJ (Block Nested Loop Join).

### ΑΣΚΗΣΗ 3 [35 μονάδες]

Οι παρακάτω σχέσεις αποτελούν μέρος της βάσης δεδομένων του πληροφοριακού συστήματος των γραμματειών των τμημάτων ενός πανεπιστημίου. Τα πρωτεύοντα κλειδιά των σχέσεων είναι υπογραμμισμένα.

ΠΤΥΧΙΟΥΧΟΙ(ΑΜ, ΕΠΩΝΥΜΟ, ΟΝΟΜΑ, ΤΜΗΜΑ)

ΜΑΘΗΜΑΤΑ(ΚΜ, ΤΙΤΛΟΣ, ΠΕΡΙΓΡΑΦΗ, ECTS)

ΒΑΘΜΟΙ(ΑΜ,ΚΜ, ΒΑΘΜΟΣ)

Η σχέση ΠΤΥΧΙΟΥΧΟΙ περιέχει τα στοιχεία των πτυχιούχων του πανεπιστημίου, η σχέση ΜΑΘΗΜΑΤΑ περιέχει τα προσφερόμενα μαθήματα, ενώ η σχέση ΒΑΘΜΟΙ περιέχει τα μαθήματα που έχει περάσει με επιτυχία κάθε πτυχιούχος, καθώς και τον τελικό βαθμό του σε κάθε μάθημα.

Δίνεται το παρακάτω SQL επερώτημα, το οποίο εμφανίζει τον αριθμό μητρώου, το ονοματεπώνυμο των πτυχιούχων του τμήματος Πληροφορικής, καθώς και τους τίτλους των μαθημάτων που έχουν περάσει με βαθμό 10 (Δέκα).

**SELECT** ΠΤΥΧΙΟΥΧΟΙ.ΑΜ, ΟΝΟΜΑ, ΕΠΩΝΥΜΟ, ΤΙΤΛΟΣ

**FROM** ΠΤΥΧΙΟΥΧΟΙ, ΜΑΘΗΜΑΤΑ, ΒΑΘΜΟΙ

**WHERE** ΠΤΥΧΙΟΥΧΟΙ.ΑΜ = ΒΑΘΜΟΙ.ΑΜ **AND** ΒΑΘΜΟΙ.ΚΜ = ΜΑΘΗΜΑΤΑ.ΚΜ **AND**  
(ΤΜΗΜΑ = 'Πληροφορική' **AND** ΒΑΘΜΟΣ = 10);

Έστω ότι ισχύουν τα ακόλουθα:

1. T(ΠΤΥΧΙΟΥΧΟΙ)=1600, B(ΠΤΥΧΙΟΥΧΟΙ)=80, T(ΜΑΘΗΜΑΤΑ)=500, B(ΜΑΘΗΜΑΤΑ)=250, T(ΒΑΘΜΟΙ)=64000, B(ΒΑΘΜΟΙ)=1600
2. Το πανεπιστήμιο έχει 20 διαφορετικά τμήματα.
3. Κάθε πτυχιούχος έχει περάσει κατά μέσο όρο 40 μαθήματα.
4. Ο βαθμός 10 (Δέκα) εμφανίζεται στο 5% των βαθμολογιών των μαθημάτων των πτυχιούχων.
5. Το πεδίο ΒΑΘΜΟΙ.ΑΜ είναι ξένο κλειδί που αναφέρεται στο πεδίο ΠΤΥΧΙΟΥΧΟΙ.ΑΜ.
6. Το πεδίο ΒΑΘΜΟΙ.ΚΜ είναι ξένο κλειδί που αναφέρεται στο πεδίο ΜΑΘΗΜΑΤΑ.ΚΜ.

7. Υπάρχει ένα απλό ευρετήριο (non-clustered) B+ δέντρο στο πεδίο ΜΑΘΗΜΑΤΑ.ΚΜ. Αυτό είναι το μόνο ευρετήριο που υπάρχει και βρίσκεται στην κύρια μνήμη.
8. Το μέγεθος της διαθέσιμης μνήμης είναι 10 σελίδες (M=10).
9. Όπου απαιτείται, υποθέστε ότι τα δεδομένα κατανέμονται ομοιόμορφα.

Ζητείται να υπολογίσετε το ελάχιστο δυνατό κόστος σε I/O του φυσικού πλάνου εκτέλεσης που ακολουθεί. Συγκεκριμένα:

1. Να υπολογίσετε το κόστος σε I/O (εφόσον υφίσταται) για κάθε μία από τις 4 επιμέρους λειτουργίες του πλάνου και να δείξετε πως αυτό προκύπτει.
2. Να υπολογίσετε το πλήθος των εγγραφών στην έξοδο για κάθε μία από τις 4 επιμέρους λειτουργίες του πλάνου και να δείξετε πως αυτός προκύπτει.
3. Να υπολογίσετε το συνολικό κόστος του φυσικού πλάνου εκτέλεσης σε I/O.

### ΦΥΣΙΚΟ ΠΛΑΝΟ ΕΚΤΕΛΕΣΗΣ

