



ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ II

ΚΑΘΗΓΗΤΗΣ : Β. ΜΕΓΑΛΟΟΙΚΟΝΟΜΟΥ

ΟΝΟΜΑΤΕΠΩΝΥΜΟ: _____

ΑΜ: _____ ΕΤΟΣ: _____

Θέμα 1^ο Γενικές ερωτήσεις (30%)

1. Ποια είναι τα πλεονεκτήματα των ευρετηρίων με την χρήση B-trees σε σχέση με τα ευρετήρια που χρησιμοποιούν κατακερματισμό; (10 μονάδες)
2. Γιατί η μέθοδος των space filling curves (SFC) είναι σημαντική στην δεικτοδότηση χωρικών και γενικότερα πολυδιάστατων δεδομένων; Περιγράψετε δύο γνωστούς τύπους καμπυλών. Αναφέρετε έναν τρόπο αξιολόγησης μιας SFC. (10 μονάδες)
3. Συμπληρώστε τον παρακάτω πίνακα δηλώνοντας με X την συμμετοχή κάθε γραμμής στην αντίστοιχη στήλη. (10 μονάδες)

	OLTP	OLAP
Μικρής διάρκειας συναλλαγές		
Σύνθετα ερωτήματα		
επαναλαμβανόμενες ενέργειες		
Μεγάλης διάρκειας ερωτήματα		
διερευνητικά ερωτήματα		

Θέμα 2^ο Βελτιστοποίηση Επερωτήσεων (30%)

Θεωρείστε τη σχέση:

GAMES (Title, Price, Publisher, Platform, Description)

και τα ακόλουθα χαρακτηριστικά:

$N_{\text{GAMES}} = 32768$ εγγραφές

$F_{\text{GAMES}} = 64$ εγγραφές ανά μπλοκ

$V(\text{Price}, \text{GAMES}) = 512$ διακριτές εγγραφές

$V(\text{Platform}, \text{GAMES}) = 16$ διακριτές εγγραφές

$V(\text{Publisher}, \text{GAMES}) = 64$ διακριτές εγγραφές

Για τον υπολογισμό της επιλεκτικότητας (selection cardinality) θεωρείστε ότι έχουμε ομοιόμορφη κατανομή τιμών.

A. Θεωρείστε την πράξη επιλογής $\sigma_{\text{Title} = \text{'Pacman'}}(\text{GAMES})$

1. Βρείτε το πλήθος των προσπελάσεων σε αριθμό μπλοκ (στη μέση περίπτωση), αν εφαρμόσουμε σειριακή αναζήτηση.
2. Αν το αρχείο που περιέχει τις εγγραφές είναι διατεταγμένο ως προς το πεδίο **Title**, ποιο είναι το αντίστοιχο πλήθος προσπελάσεων;

B. Θεωρήστε την πράξη επιλογής $\sigma_{\text{Publisher} = \text{'EA Sports'}}(\text{GAMES})$

1. Βρείτε το πλήθος των προσπελάσεων σε αριθμό μπλοκ, αν εφαρμόσουμε ακολουθιακή σάρωση.
2. Αν το αρχείο είναι διατεταγμένο ως προς το πεδίο **Publisher**, ποιο είναι το πλήθος των προσπελάσεων σε αριθμό μπλοκ;

Θέμα 3^ο Δοσοληψίες (20%)

Έστω το χρονοπρόγραμμα:

S₁: W₁(X); R₂(Y); R₃(X); W₁(Y); W₂(X); R₄(Y); R₄(X);

Όπου:

**R_i(X): η δοσοληψία T_i διαβάζει (READ) το στοιχείο X και
w_i(X): η δοσοληψία T_i γράφει (WRITE) το στοιχείο X**

1. Είναι το χρονοπρόγραμμα σειριοποιήσιμο; Να δικαιολογήσετε την απάντησή σας.
2. Ακολουθεί το χρονοπρόγραμμα το πρωτόκολλο κλειδώματος δύο φάσεων (two phase locking); Να δικαιολογήσετε την απάντησή σας.

Θέμα 4^ο Indexing (20%)

Να σχεδιάσετε το point quad-tree που προκύπτει από τη διαδοχική εισαγωγή των παρακάτω σημείων (με συντεταγμένες που δίνονται στις παρενθέσεις):

1. **A** (5, 7)
2. **B** (8, 9)
3. **C** (6, 4)
4. **D** (2, 5)
5. **E** (7, 8)
6. **F** (4, 1)

Καλή επιτυχία!