

Άσκηση

Έστω μια σχέση $R(A,B)$ οργανωμένη ως αρχείο κατακερματισμού στο δίσκο. Το κατακερματισμένο αρχείο καταλαμβάνει 1024 blocks (κάδους – buckets) που περιέχουν τις εγγραφές της σχέσης. Για την αποθήκευση μιας εγγραφής (a,b) εφαρμόζουμε πρώτα τη συνάρτηση κατακερματισμού h_1 στο πεδίο a λαμβάνοντας N bits. Στη συνέχεια εφαρμόζουμε τη συνάρτηση h_2 στο πεδίο b λαμβάνοντας $10-N$ bits. Τα 10 bits συνολικά ορίζουν τη διεύθυνση του block στο οποίο θα αποθηκευτεί η εγγραφή (a,b) .

Θεωρήστε ότι το 30% των επερωτήσεων που αφορούν στη σχέση R είναι της μορφής:

Q1: SELECT * FROM R WHERE A=a,

ενώ το υπόλοιπο 70% είναι της μορφής :

Q2: SELECT * FROM R WHERE B=b

όπου a και b είναι τιμές που δίνονται από τους χρήστες που υποβάλλουν τα ερωτήματα.

1. Ζητείται να προσδιορίσετε τον αριθμό των μπλοκ που πρέπει να προσπελαστούν για να απαντηθούν α) οι επερωτήσεις τύπου Q1 και β) οι επερωτήσεις τύπου Q2. Η απάντηση να δοθεί ως συνάρτηση των N bits.
2. Δώστε έναν τύπο που να εκτιμά το μέσο αριθμό των μπλοκ που πρέπει να προσπελαστούν για την απάντηση των επερωτήσεων (Q1 και Q2) στη σχέση R .

ΛΥΣΗ

1) Για τις επερωτήσεις τύπου Q1 πρέπει να λάβουμε υπόψη μας τα **10-N** bits από την εφαρμογή της συνάρτησης h_2 . Άρα θα προσπελαστούν: $2^{(10-N)}$ μπλοκ. Αντιθέτως για τις επερωτήσεις τύπου Q2 λαμβάνουμε υπόψη μας τα N bits από τη συνάρτηση h_1 , άρα θα προσπελαστούν 2^N μπλοκ.

2) Με δεδομένο ότι το 30% των επερωτήσεων είναι τύπου Q1 και το 70% τύπου Q2, ο μέσος αριθμός των μπλοκ που πρέπει να προσπελαστούν για την απάντηση των επερωτήσεων στην R δίνεται από τον τύπο: $0.3 \cdot 2^{(10-N)} + 0.7 \cdot 2^N$