

2ο Σύνολο Ασκήσεων
Ημερομηνία Παράδοσης: 29/5/2002

Θεματική Ενότητα: Παράλληλες και Κατανεμημένες Βάσεις Δεδομένων.

Ποσοστό επί του Τελικού Βαθμού: 15%.

Η παράδοση των ασκήσεων δεν είναι υποχρεωτική. Αν δε παραδώσετε κάποιο σύνολο θα πάρετε βαθμό μηδέν σε αυτό. Οι ασκήσεις θα γίνονται σε ομάδες έως και 4 ατόμων. Αργοπορημένη παράδοση δε γίνεται δεκτή σε καμία περίπτωση.

1. Στα παρακάτω ερωτήματα (α) – (ε) θεωρείστε τις σχέσεις:

Εργαζόμενοι(eid, did, μισθός)
Τμήματα(did, mgrid, προϋπολογισμός)

Το πεδίο mgrid της σχέσης Τμήματα είναι το eid του διευθυντή (manager). Κάθε σχέση έχει πλειάδες των 20 bytes, και τα γνωρίσματα μισθός και προϋπολογισμός περιέχουν τιμές ομοιόμορφα κατανεμημένες μεταξύ του 0 και του 1.000.000. Η σχέση Εργαζόμενοι καταλαμβάνει 100.000 σελίδες και η σχέση Τμήματα 5.000. Κάθε επεξεργαστής έχει 100 καταχωρητές 4.000 bytes ο καθένας.

Το I/O κόστος μιας σελίδας είναι t_i και το κόστος μεταφοράς μιας σελίδας είναι t_s .

Έστω οι παρακάτω ερωτήσεις:

- (i) Το τμήμα με το μεγαλύτερο προϋπολογισμό.
- (ii) Το τμήμα με το μεγαλύτερο προϋπολογισμό ανάμεσα στα τμήματα που έχουν διευθυντές με μισθό μικρότερο του 200.000
- (iii) Τους μισθούς όλων των διευθυντών

Για τις ερωτήσεις που αναφέρονται σε καθένα από τα ερωτήματα (α) – (στ) υπολογίστε το συνολικό χρόνο εκτέλεσής τους σε όλους τους κόμβους καθώς και το “elapsed” χρόνο (δηλαδή, όταν πολλές λειτουργίες εκτελούνται ταυτόχρονα, ο χρόνος είναι ο μέγιστος ανάμεσα σε αυτές τις λειτουργίες και όχι το άθροισμα). Αν δε αναφέρεται αλγόριθμος θεωρείστε τον καλύτερο δυνατό. Επίσης, για τα ερωτήματα (α) – (δ) δώστε και το speed-up με βάσει τον καλύτερο σειριακό αλγόριθμο.

(α) Θεωρείστε ένα παράλληλο ΣΔΒΔ με 10 επεξεργαστές, ότι οι σχέσεις έχουν τμηματοποιηθεί με round-robin, και τις ερωτήσεις (i), (ii), (iii).

(β) Θεωρείστε ένα παράλληλο ΣΔΒΔ με 10 επεξεργαστές, ότι οι σχέσεις έχουν τμηματοποιηθεί με βάση περιοχές στα γνωρίσματα μισθός και προϋπολογισμός και τις ερωτήσεις (i), (ii), (iii).

(γ) Θεωρείστε ένα παράλληλο ΣΔΒΔ με 100 επεξεργαστές, ότι οι σχέσεις έχουν τμηματοποιηθεί με round-robin, και τις ερωτήσεις (i), (ii), (iii).

(δ) Θεωρείστε ένα παράλληλο ΣΔΒΔ με 100 επεξεργαστές, ότι οι σχέσεις έχουν τμηματοποιηθεί με βάση περιοχές στα γνωρίσματα μισθός και προϋπολογισμός και τις ερωτήσεις (i), (ii), (iii).

(ε) Θεωρείστε ότι οι σχέσεις τώρα είναι αποθηκευμένες σε ένα καταναμημένο ΣΔΒΔ με 2 κόμβους, τον K1 και τον K2. Όλη η σχέση Εργαζόμενος είναι στον κόμβο K1 και όλη η σχέση Τμήματα στον K2. Θεωρείστε την ερώτηση (iii). Η ερώτηση υποβάλλεται στον K2. Το ποσοστό των εργαζομένων που είναι διευθυντές είναι 1%. Θεωρείστε τους παρακάτω αλγόριθμους: (1) Μεταφορά και των δυο σχέσεων στο K2, (2) Μεταφορά της σχέσης Τμήματα στο K1, (3) Υπολογισμός της συνένωσης στο K1 χρησιμοποιώντας ημι-συνένωση, (4) Υπολογισμός της συνένωσης στο K2 χρησιμοποιώντας ημι-συνένωση. Ποιος από τους τέσσερις αλγόριθμους ελαχιστοποιεί το κόστος μεταφοράς; είναι και ο καλύτερος;

2. Εξηγήστε τις απαντήσεις σας. Θα βαθμολογηθείτε για την πληρότητα της απάντησης.

(α) Πότε είναι χρήσιμα τα αντίτυπα (replication) και πότε η τμηματοποίηση (fragmentation).

(β) Ποιο είδος παραλληλισμού (intra-operation, inter-operation, inter-query) είναι πιθανότερα πιο σημαντικό για καθένα από τα παρακάτω:

(i) αύξηση της απόκρισης ενός συστήματος με πολλές μικρές ερωτήσεις

(ii) αύξηση της απόκρισης ενός συστήματος με λίγες, μεγάλες ερωτήσεις, όταν ο αριθμός των δίσκων και των επεξεργαστών είναι μεγάλος

(γ) Περιγράψτε έναν αποδοτικό τρόπο για την παραλληλοποίηση:

(i) της πράξης count

(iii) της πράξης count distinct