

Ασκήσεις στην Επεξεργασία Ερωτήσεων

Ημερομηνία Παράδοσης: 14 Μαΐου 2003

Άσκηση 1

Θεωρείστε μια βάση δεδομένων με δύο σχέσεις.

Employees(eid, did, salary)

Departments(did, location, budget)

Κάθε σχέση περιέχει πλειάδες μεγέθους 20 bytes. Το μέγεθος της σελίδας (block) είναι 2.000 bytes. Τα πεδία sal και budget παίρνουν ομοιόμορφα κατανεμημένες τιμές από 0 έως 1.000.000. Η σχέση Employees περιέχει 200.000 σελίδες και η σχέση Departments 5.000 σελίδες. Το κόστος I/O για μια σελίδα είναι t_d και το κόστος της μεταφοράς μιας σελίδας είναι t_s . Οι πλειάδες μεταφέρονται σε μονάδες της μιας σελίδας.

Θεωρείστε ένα κατανεμημένο ΣΔΒΔ με 20 κόμβους. Οι πλειάδες της σχέσης Departments διαμερίζονται ομοιόμορφα (με οριζόντια κατάτμηση) ανάμεσα στις 20 τοποθεσίες βάσει του did, με ίσο αριθμό πλειάδων σε κάθε τοποθεσία χωρίς κάποια συγκεκριμένη σειρά. Οι πλειάδες της σχέσης Employee διαμερίζονται (πάλι με οριζόντια κατάτμηση) με βάση τα διαστήματα τιμών του sal, με τις πλειάδες με $sal \leq 50.000$ να αποθηκεύονται στον πρώτο κόμβο, τις πλειάδες με $50.000 < sal \leq 100.000$ στο δεύτερο κ.ο.κ. Για τα παρακάτω διατυπώστε την ερώτηση σε SQL και σχεσιακή άλγεβρα. Υπολογίστε τον ολικό χρόνο και το χρόνο απόκρισης για το καλύτερο πλάνο εκτέλεσης.

(α) Υπολογίστε τη φυσική συνένωση των σχέσεων Employees και Departments μεταφέροντας όλα τα κομμάτια της μικρότερης σχέσης σε όλους τους κόμβους που υπάρχουν πλειάδες της μεγαλύτερης.

(β) Βρείτε τον υπάλληλο με το μεγαλύτερο μισθό από 450.000 έως 550.000.

(γ) Βρείτε τον διευθυντή με το μεγαλύτερο μισθό.

Άσκηση 2

Θεωρείστε μια βάση δεδομένων με δύο σχέσεις.

ACTORS(name, address, age)

MOVIES(title, leading-actor, studio)

Το πεδίο leading-actor είναι το name του ηθοποιού που πρωταγωνιστεί στην ταινία. Μόνο ένα 5% των ηθοποιών είναι πρωταγωνιστές.

Έστω η ερώτηση

```
select *  
from ACTORS A, MOVIES M  
Where M>leading-actor = A.name
```

Κάθε σχέση περιέχει πλειάδες μεγέθους 40 bytes. Το μέγεθος της σελίδας (block) είναι 4000 bytes. Τα πεδία name και leading-actor έχουν μέγεθος 10 bytes. Η σχέση ACTORS περιέχει 100.000 σελίδες και η σχέση MOVIES 4.000 σελίδες. Το κόστος I/O για μια σελίδα είναι t_d και το κόστος της μεταφοράς μιας σελίδας είναι t_s . Θεωρείστε $t_s = 2t_d$. Οι πλειάδες μεταφέρονται σε μονάδες της μιας σελίδας.

Θεωρείστε ένα καταναμημένο ΣΔΒΔ με 3 κόμβους. Οι πλειάδες της σχέσης ACTORS βρίσκονται στον κόμβο 1 και τις σχέσεις MOVIES στον κόμβο 2. Το αποτέλεσμα το θέλουμε στον κόμβο 3. Υπολογίστε τον ολικό χρόνο και το χρόνο απόκρισης για τον υπολογισμό της ερώτησης για καθένα από τα παρακάτω.

(α) Υπολογίστε την ερώτηση στον κόμβο 1 με ημι-συνένωση και τελική μεταφορά του αποτελέσματος στον κόμβο 3

(β) Υπολογίστε την ερώτηση στον κόμβο 2 με ημι-συνένωση και τελική μεταφορά του αποτελέσματος στον κόμβο 3

(γ) Ποιος πλάνο από τα (α) και (β) είναι καλύτερο;

(δ) Περιγράψτε μια μέθοδο που επεκτείνει την έννοια της ημι-συνένωσης για τον υπολογισμό της ερώτησης στον κόμβο 3 και υπολογίστε το κόστος της.

Άσκηση 3

Θεωρείστε μια βάση δεδομένων με τις ακόλουθες σχέσεις:

```
PROJ(PNO, PNAME, BUDGET)  
ASG(ENO, PNO, RESP, DUR)
```

Για τη σχέση PROJ θεωρούμε οριζόντια κατάτμηση σε $PROJ1 = \sigma_{PNO \leq P2} (PROJ)$ και $PROJ2 = \sigma_{PNO > P2} (PROJ)$. Το ίδιο και για τη σχέση ASG σε $ASG1 = \sigma_{PNO \leq P2} (ASG)$, $ASG2 = \sigma_{P2 < PNO \leq P3} (ASG)$ και $ASG3 = \sigma_{PNO > P3} (ASG)$.

Μετασχηματίστε την παρακάτω ερώτηση σε μια ερώτηση πάνω σε τμήματα και εφαρμόστε τις κατάλληλες απλοποιήσεις.

```
select RESP, BUDGET  
from ASG, PROJ  
where ASG.PNO = PROJ.PNO and PNAME = "CAD/CAM"
```