

**Άσκηση 3η**  
**(αυστηρή) Ημερομηνία Παράδοσης: 24/05/2019, 16.59**

**ΘΕΜΑ 1<sup>ο</sup>**

Έχετε την ερώτηση  $\sigma_{R.C > 40} (R \bowtie_{R.A=S.B} S)$  και τις παρακάτω πληροφορίες.

- Η σχέση R έχει 150.000 πλειάδες και 30 πλειάδες ανά σελίδα.
- Η σχέση S έχει 100 πλειάδες και 20 πλειάδες ανά σελίδα.
- Το B είναι πρωτεύον κλειδί και το A ξένο κλειδί σ' αυτό.
- Δεν υπάρχουν ευρετήρια ή ταξινόμηση των σχέσεων στο δίσκο
- Διαθέτουμε 52 buffers στην κύρια μνήμη
- Η σχέση R έχει το εξής ιστόγραμμα για την στήλη C: (α) για τιμές στο εύρος 0-50, το 40% των εγγραφών της σχέσης, (β) για τιμές από 51-65 το 45% των εγγραφών και (γ) για τιμές από 66-80 τις υπόλοιπες εγγραφές.

(α) Εκτιμήστε το μέγεθος του αποτελέσματος και τον απαιτούμενο χώρο στο δίσκο για να γραφτεί το αποτέλεσμα αυτό. Προσοχή: δεν σας δίνονται άλλες πληροφορίες (π.χ., εξωτερικά κλειδιά) και άρα πρέπει να σκεφθείτε μια λογική εκτίμηση για τη χειρότερη περίπτωση ☺

Αιτιολογήστε το ακριβές, ελάχιστο και μέγιστο μέγεθος του αποτελέσματος της ερώτησης αν

- το B είναι πρωτεύον κλειδί και το A ξένο κλειδί στο B,
- το B είναι πρωτεύον κλειδί
- τίποτε από τα προηγούμενα δεν ισχύει

(β) Δώστε δύο αλγεβρικά πλάνα για την εκτέλεση της πράξης και διαλέξτε ένα εξ' αυτών που θα χρησιμοποιήσετε για τα επόμενα ερωτήματα. Αιτιολογήστε αν μπορείτε να κάνετε pipelining αποτελεσμάτων ή αν πρέπει να αποθηκεύσετε ενδιάμεσα αποτελέσματα.

(γ) Υπολογίστε το κόστος σε I/O της εν λόγω ερώτησης, με όποιο αλγόριθμο κρίνετε σκόπιμο για την πράξη της επιλογής, αλλά για κάθε έναν από τους παρακάτω τρόπους εκτέλεσης της πράξης σύνδεσης:

- Page-based nested loops join
- Block nested loops join
- Sort-Merge join
- Hash join

(δ) Υπολογίστε το ελάχιστο δυνατό κόστος για την εκτέλεση της ερώτησης και υπό ποιες συνθήκες αυτό μπορεί να επιτευχθεί.

**ΘΕΜΑ 2<sup>ο</sup>**

Έστω η παρακάτω βάση δεδομένων που αφορά το προσωπικό και τα οικονομικά μιας εταιρείας εξορύξεως πετρελαίου στη θάλασσα.

```
EMP (EID, ENAME, ECID, ESAL, E_PREV_SERVICE)
CREW (CID, CNAME, CTASK, CPID)
PLATFORM (PID, PLOCATION, PBUDGET)
```

Τα EMP.ECID και CREW.CPID είναι foreign keys στους αντίστοιχους πίνακες. Υποθέστε ότι η εταιρεία έχει 5 τοποθεσίες εργασιών, 20 πλατφόρμες, 5 είδη καθηκόντων συνεργείου, 100 συνεργεία και 1000 υπαλλήλους. Επίσης, η προϋπηρεσία κάποιου είναι μεταξύ 0 και 39 χρόνων και οι μισθοί κυμαίνονται από 20000 ως και 60000. Όλες οι τιμές κυμαίνονται ομοιόμορφα και είναι ακέραιες. Θεωρήστε 100 εγγραφές ανά σελίδα για τις σχέσεις CREW & PLATFORM και 50 εγγραφές ανά σελίδα για τη σχέση EMP.

1. Γράψτε σε SQL και σχεσιακή άλγεβρα το ερώτημα: «Βρες το όνομα και το μισθό, καθώς και το ID και το budget της πλατφόρμας στην οποία εργάζεται κάθε υπάλληλος για τις πλατφόρμες Κύπρου και Ιονίου και τους υπαλλήλους που έχουν μισθό πάνω από 20.000 και προϋπηρεσία μικρότερη των 20 ετών».
2. Δώστε το τυπικό δέντρο σχεσιακής άλγεβρας που περιγράφει την ερώτηση, όπως θα προέκυπτε με «τυφλή» μετάφραση της SQL σε σχ. άλγεβρα. Δώστε δύο άλλα πλάνα που θα εξέταζε ένας βελτιστοποιητής.

3. Διαλέξτε ένα από τα δύο πλάνα αυτά και προσδιορίστε για κάθε τελεστή του δέντρου, τι εκτίμηση θα κάνει ο βελτιστοποιητής για τον αριθμό των εγγραφών που θα παράγει.
4. Προσδιορίστε το κόστος του πλάνου που δώσατε, θεωρώντας ότι οι συνδέσεις γίνονται με hash-join? Θεωρείστε ότι έχετε διαθέσιμες  $\sqrt{|R|}$  buffers ( $|R|$  το μέγεθος της μέγιστης σχέσης σε σελίδες, από αυτές που συμμετέχουν στην ερώτηση) σελίδες στη μνήμη. Ζωγραφίστε την ανάθεση στη μνήμη.

Οι εργασίες εξασκήσεως είναι **ατομικές** (εκτός κι αν ρητά αναφέρεται κάτι άλλο) και παραδίδονται **εκτυπωμένες** την προκαθορισμένη ημερομηνία και ώρα (αν δεν υπάρχει ένδειξη ώρας, υπονοείται η ώρα έναρξης του μαθήματος). **Ασκήσεις που υποβάλλονται ηλεκτρονικά ή σε χειρόγραφη μορφή ΔΕΝ θα γίνονται αποδεκτές.**