

2^η Εργασία στη Σχεδίαση ΒΔ 2022-2023

Ευρετήρια

Στόχος των εργασιών είναι η εξοικείωση με θεωρητικά και πρακτικά προβλήματα των Βάσεων Δεδομένων, μέσα από χρηστικά παραδείγματα. Στη δεύτερη εργασία θα ασχοληθούμε με τις εντολές δημιουργίας ευρετηρίων και με τη χρήση του optimizer.

A – Ερωτήματα

1^ο μέρος – Εντολές δημιουργίας ευρετηρίων

Βρείτε τη σύνταξη της εντολής CREATE INDEX στην ORACLE. Εξηγείστε τις παραμέτρους της εντολής σε σχέση με τους διαθέσιμους τύπους ευρετηρίου που έχουμε δει στο μάθημα. Αν χρησιμοποιήσατε πηγές από το δίκτυο να τις παραθέσετε (URL διευθύνσεις).

2^ο μέρος – Εντολές δημιουργίας σχήματος

Στη ΒΔ σας θα πρέπει να δημιουργήσετε τρεις σχέσεις **Customers**, **Orders**, **Products** δεδομένα για τις οποίες θα πρέπει να αντιγράψετε από τη ΒΔ **XSALES** και συγκεκριμένα από τους πίνακες **customers**, **products (and categories)**, **orders** και **order_items**.

Ανεξάρτητα από το αρχικό σχήμα, οι σχέσεις θα πρέπει να έχουν την ακόλουθη δομή:

Η Customers:

| CUSTOMER_ID | GENDER | AGEGROUP | MARITAL_STATUS | INCOME_LEVEL |
|-------------|--------|----------|----------------|--------------|
| 101542 | Female | above 70 | single | low |
| 47829 | Female | 40-50 | single | medium |
| 4940 | Female | above 70 | unknown | medium |
| 12050 | Female | 50-60 | married | medium |
| 19162 | Female | 40-50 | single | medium |
| 104407 | Female | above 70 | single | high |

α) Το πεδίο age_group θα προκύψει από το birth_date και την τρέχουσα ημερομηνία και με χρήση κατάλληλων updates που θα ετοιμάσετε και που θα μετατρέπουν την ηλικία στις ακόλουθες ομάδες: i) under 30, ii) 30-40, iii) 40-50, iv) 50-60, v) 60-70, vi) above 70, με ισότητα στο πάνω όριο.

β) Το πεδίο income level θα προκύψει με ομαδοποίηση των τιμών που έχει το αρχικό πεδίο ως εξής: i) εισόδημα ως 109.999 low, ii) εισόδημα ως 189.999 medium, iii) εισόδημα πάνω από 11000 high, iv) σε κάθε άλλη περίπτωση. Η ομαδοποίηση θα γίνει με κατάλληλα updates.

γ) Το πεδίο marital_status θα ομαδοποιηθεί με updates που θα αντιστοιχούν τα 'Married', 'Mabsent', 'married', 'Mar-AF' σε married, τα υπόλοιπα σε 'single' και τα null σε unknown.

Η products:

| PRODUCT_ID | PRODUCTNAME | CATEGORYNAME | LIST_PRICE |
|------------|--------------------------------------|-------------------|------------|
| 15 | Envoy 256MB - 40GB | Desktop PCs | 999.99 |
| 28 | Unix/Windows 1-user pack | Operating Systems | 199.99 |
| 113 | CD-R Mini Discs | Recordable CDs | 22.99 |
| 114 | Music CD-R | Recordable CDs | 18.99 |
| 115 | CD-RW, High Speed, Pack of 10 | Recordable CDs | 8.99 |
| 116 | CD-RW, High Speed Pack of 5 | Recordable CDs | 11.99 |
| 117 | CD-R, Professional Grade, Pack of 10 | Recordable CDs | 8.99 |
| 118 | OraMusic CD-R, Pack of 10 | Recordable CDs | 7.99 |
| 119 | CD-R with Jewel Cases, pack OF 12 | Recordable CDs | 6.99 |

Η orders:

| ORDER_ID | PRODUCT_ID | CUSTOMER_ID | DAYS_TO_PROCESS | PRICE | COST | CHANNEL |
|----------|------------|-------------|-----------------|--------|--------|--------------|
| 4631 | 27 | 2144 | 150 | 48.09 | 41.54 | Direct Sales |
| 12184 | 27 | 6228 | 51 | 48.09 | 41.54 | Direct Sales |
| 7324 | 27 | 3245 | 70 | 48.09 | 41.54 | Direct Sales |
| 14363 | 27 | 8312 | 91 | 48.09 | 41.54 | Direct Sales |
| 8594 | 27 | 3877 | 93 | 46.57 | 37.02 | Direct Sales |
| 6620 | 27 | 2935 | 31 | 46.74 | 41.24 | Internet |
| 12845 | 27 | 6794 | 31 | 46.74 | 41.24 | Internet |
| 1721 | 28 | 819 | 0 | 216.38 | 177.31 | Direct Sales |

Η στήλη price είναι η αρχική amount (του πίνακα order_items), η στήλη days_to_process είναι η διαφορά μεταξύ Order_date και Order_finished σε ημέρες.

Ερώτημα 3ο – Βελτιστοποίηση ερωτήματος

Χρησιμοποιώντας την εντολή EXPLAIN ελέγξτε πώς λειτουργεί ο optimizer για το ακόλουθο ερώτημα

```
SELECT c.name, oi.amount, oi.quantity, p.name
FROM customers c JOIN orders o ON c.id=o.customer_id
      JOIN order_items oi ON o.id=oi.order_id
      JOIN products p ON oi.product_id=p.identifier
WHERE p.name LIKE '%DVD%'
      AND oi.order_date BETWEEN '1/1/2000' AND '31/12/2000' AND
      channel='Internet'
ORDER BY c.name;
```

- 1) Σύμφωνα με την EXPLAIN ποιο είναι το εκτιμώμενο συνολικό κόστος για την εκτέλεση του καλύτερου πλάνου για το ερώτημα αυτό; Ποια τα CPU_COST και IO_COST; Ποια είναι η πιο χρονοβόρα ενέργεια.
- 2) Ποιο είναι το εκτιμώμενο πλήθος αποτελεσμάτων για το ερώτημα; Πόσες πλειάδες επιστρέφει πραγματικά το ερώτημα; Χρησιμοποιώντας συμβολισμούς σχεσιακής άλγεβρας, σχεδιάστε το πλάνο εκτέλεσης που επέλεξε ο optimizer.

Επιχειρήστε να βελτιστοποιήσετε το ερώτημα δημιουργώντας τα κατάλληλα ευρετήρια.

- 3) Ποιο το τελικό κόστος μετά τη βελτιστοποίηση της σχεδίασης;

Δώστε τις εντολές που χρησιμοποιήσατε σε κάθε βήμα

Ερώτημα 4ο – Θεωρητικός υπολογισμός κόστους

Αφού βρείτε με κατάλληλο ερώτημα προς την Oracle το μέγεθος κάθε σχέσης και κάθε ευρετηρίου που χρησιμοποιήσατε, να υπολογίσετε θεωρητικά το κόστος του πλάνου που επιλέχθηκε (ή έστω του αρχικού πλάνου χωρίς τα ευρετήρια) με βάση τις ενέργειες που πραγματοποιήθηκαν.

Συγκρίνετε την πρακτική εκτίμηση με το θεωρητικό υπολογισμό.

Β – Οδηγίες Παράδοσης

Η εργασία θα υλοποιηθεί από ομάδες των **3 ατόμων (το πολύ)**, αν και επιτρέπονται μικρότερες ομάδες. Θα πρέπει τελικά να ανεβάσετε ένα zip αρχείο με ονομασία τους ΑΜ των μελών της ομάδας: π.χ. **AM1-AM2-AM3.zip**

- Το zip θα περιλαμβάνει:
 - ένα αρχείο readme.txt
 - με τα ονοματεπώνυμα και τους ΑΜ των φοιτητών της ομάδας
 - το αρχείο word με την τελική εργασία. Τυχόν συνοδευτικά αρχεία sql, ppt κλπ.

Γ – Άλλες Οδηγίες

Όσες εργασίες δεν τηρούν τις οδηγίες παράδοσης, θα έχουν επίπτωση στο βαθμό.

Όσες εργασίες κριθούν ότι είναι **αντιγραφές θα μηδενίζονται**.

Ημερομηνία παράδοσης: **Στο e-class με οριστική τελική ημερομηνία 16-12-2022**

Όσες εργασίες παραδοθούν μετά το πέρας της ημερομηνίας και μέχρι τις 18-12-2022 θα έχουν μείωση 2 μονάδων στο βαθμό.