

2^η εργασία – PL/SQL, Ευρετήρια και Optimizer

Στόχος των εργασιών είναι η εξοικείωση με θεωρητικά και πρακτικά προβλήματα των Βάσεων Δεδομένων, μέσα από χρηστικά παραδείγματα. Στην πρώτη εργασία θα ασχοληθούμε α) με τη χρήση PL/SQL για το χειρισμό εγγραφών της ΒΔ, β) με τις εντολές εμφάνισης του πλάνου εκτέλεσης, τις πληροφορίες που μας δίνουν και τον τρόπο με τον οποίο μπορούμε να επιταχύνουμε την εκτέλεση ενός ερωτήματος.

Ερώτημα 1ο – Δημιουργία και γέμισμα σχήματος

Στη ΒΔ σας θα πρέπει να έχετε 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 με χρήση της συνάρτησης get_age_group που θα υλοποιήσετε και που θα μετατρέπει 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 χρήση της συνάρτησης get_income_level που θα υλοποιήσετε.

γ) Το πεδίο marital_status θα ομαδοποιηθεί με τη συνάρτηση fix_status που θα υλοποιήσετε η οποία θα αντιστοιχεί 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 2ο – Εντοπισμός ζημιογόνων παραγγελιών

Δημιουργώντας τις κατάλληλες συναρτήσεις ή διεργασίες σε PL/SQL να καταχωρήσετε σε ένα πίνακα τις παραγγελίες που ζημίωσαν την εταιρία ως εξής:

- i) Για κάθε παραγγελία βρείτε τη μέγιστη καθυστέρηση εκτέλεσης σε ημέρες λαμβάνοντας υπόψη ότι η καθυστέρηση μετρά μετά τις 10 πρώτες ημέρες (`days_to_process > 10` για το σύνολο των προϊόντων της παραγγελίας)
- ii) Υπολογίστε το τελικό κέρδος της αφού πρώτα αφαιρέσετε (ανά προϊόν) από την τιμή πώλησης (`price`) την τιμή κόστους (`cost`) και για κάθε ημέρα καθυστέρησης της εκτέλεσης επιπλέον το 0.0001 της τιμής καταλόγου (`list_price`)
- iii) Δημιουργήστε έναν `cursor` που να διατρέχει τις παραγγελίες και να υπολογίζει ανά παραγγελία το τελικό κέρδος (για όλα τα προϊόντα της), στη συνέχεια να ελέγχει αν το κέρδος είναι αρνητικό ή θετικό. Στην περίπτωση i) που είναι αρνητικό να καταχωρεί σε ένα πίνακα **deficit** τα (`orderid, customerid, channel, amount`) όπου η `amount` θα έχει τη ζημιά με θετικό πρόσημο, ii) που είναι θετικό να καταχωρεί σε ένα πίνακα **profit** τα αντίστοιχα `orderid, customerid, channel, amount`

Τέλος να δώσετε τα queries και τις απαντήσεις στα ακόλουθα

- iv) Ποια τα συνολικά έσοδα και ζημιές σε παραγγελίες που έγιναν από άνδρες και γυναίκες αντίστοιχα;
- v) Ποια τα συνολικά έσοδα και ζημιές ανά κανάλι παραγγελιών;

Τα πιο πάνω θα πρέπει να γίνουν με SQL ή PL-SQL και θα πρέπει να δώσετε το script με το οποίο τα δημιουργήσατε.

Ερώτημα 3ο – Βελτιστοποίηση ερωτήματος ισότητας

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

```
select order_id, price-cost, days_to_process
from products p join orders o on o.product_id=p.product_id
join customers c on o.customer_id=c.customer_id
where p.categoryname='Accessories' and o.channel='Internet' and
c.gender='Male' and c.income_level='high' and days_to_process=0;
```

- 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 – Βελτιστοποίηση ερωτήματος ανισότητας

Τι θα αλλάξει αν κάνετε το ακόλουθο ερώτημα

```
select order_id, price-cost, days_to_process
from products p join orders o on o.product_id=p.product_id
join customers c on o.customer_id=c.customer_id
where p.categoryname='Accessories' and o.channel='Internet' and
c.gender='Male' and c.income_level='high' and days_to_process>100;
```

- 1) Σύμφωνα με την EXPLAIN ποιο είναι το εκτιμώμενο συνολικό κόστος για την εκτέλεση του καλύτερου πλάνου χωρίς ευρετήρια για το ερώτημα αυτό; Ποια τα CPU_COST και IO_COST; Ποια είναι η πιο χρονοβόρα ενέργεια.
- 2) Επιχειρήστε να βελτιστοποιήσετε το ερώτημα δημιουργώντας τα κατάλληλα ευρετήρια. Ποιο το τελικό κόστος μετά τη βελτιστοποίηση της σχεδίασης;

Δώστε τις εντολές που χρησιμοποιήσατε σε κάθε βήμα

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,rpt 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 11-1-2019**

Όσες εργασίες παραδοθούν μετά το πέρας της ημερομηνίας και μέχρι τις 13-1-2019 θα έχουν μείωση 2 μονάδων στο βαθμό.