



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ ΣΧΟΛΗ ΗΜΜΥ

ΕΡΓΑΣΤΗΡΙΟ ΔΙΑΝΕΜΗΜΕΝΩΝ
ΠΑΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ ΚΑΙ ΕΦΑΡΜΟΓΩΝ

ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ - ΠΛΗ303 Β' ΦΑΣΗ ΕΡΓΑΣΤΗΡΙΑΚΗΣ ΕΡΓΑΣΙΑΣ – ΕΑΡΙΝΟ ΕΞΑΜΗΝΟ 2021-2022

ΔΙΔΑΣΚΩΝ: Δεληγιαννάκης Αντώνιος
ΕΡΓΑΣΤΗΡΙΑΚΟ ΠΡΟΣΩΠΙΚΟ: Καζάσης Φώτιος, Παππάς Νικόλαος

Η δεύτερη φάση της εργαστηριακής εργασίας περιλαμβάνει τα παρακάτω ζητούμενα:

1. Επικοινωνία Java εφαρμογής με το σύστημα διαχείρισης βάσεων δεδομένων PostgreSQL μέσω JDBC.

Για την υλοποίηση της εφαρμογής θα χρησιμοποιήσετε το περιβάλλον Eclipse και τον οδηγό JDBC που παρέχεται μαζί με τον εξυπηρετητή postgresQL.

Θα πρέπει να υλοποιήσετε μια κλάση η οποία θα παρέχει μεθόδους για τα ακόλουθα:

1. Σύνδεση/Αποσύνδεση σε/από συγκεκριμένη βάση σε εξυπηρετητή postgresQL (Διεύθυνση IP, όνομα βάσης, όνομα χρήστη, κωδικός πρόσβασης)
2. Προβολή της βαθμολογίας στις εργαστηριακές εργασίες ενός φοιτητή για ένα εξάμηνο. Θα δίνονται ως ορίσματα εισόδου ο κωδικός του φοιτητή (amka) και το εξάμηνο (academic_year, academic_season). Η κλήση της μεθόδου θα πρέπει να έχει σαν αποτέλεσμα την προβολή των στοιχείων του φοιτητή και μια λίστα που θα περιλαμβάνει τον κωδικό μαθήματος, τον τίτλο εργασίας και τη βαθμολογία.
3. Εισαγωγή τυχαίων εργαστηριακών εργασιών και ομάδων εργασίας για ένα συγκεκριμένο μάθημα εξαμήνου. Η μέθοδος αυτή θα δέχεται σαν ορίσματα εισόδου τον κωδικό μαθήματος (course_code), τη σειριακή εκτέλεση (serial_number), τον αριθμό των εργαστηριακών εργασιών (LabModule) και τον αριθμό των ομάδων εργασίας που θα σχηματιστούν. Κατά την εισαγωγή δεδομένων θα λάβετε υπόψη τα παρακάτω:
 - i. Ο καθορισμός του τύπου της εργασίας θα γίνεται τυχαία με πιθανότητα 0.2 για τύπο project και 0.8 για τύπο lab_exercise.
 - ii. Ο μέγιστος αριθμός συμμετοχών ανά ομάδα για κάθε εργασία θα επιλέγεται τυχαία από 2 ως 6.
 - iii. Ο βαθμός αξιολόγησης θα επιλέγεται ακολουθώντας κανονική κατανομή στην κλίμακα 0-10, με μέση τιμή 6 και τυπική απόκλιση 2.

Υπόδειξη: Για την παραγωγή τυχαίων τιμών αριθμών που ακολουθούν τα παραπάνω θα κάνετε χρήση των μεθόδων της κλάσης java.util.Random: nextDouble (i), nextInt (ii) και nextGaussian (iii).

4. Εισαγωγή τυχαίων εργαστηριακών εργασιών και ομάδων εργασίας για όλα τα εξαμηνιαία εργαστηριακά μαθήματα (CourseRun) ενός συγκεκριμένου γνωστικού αντικείμενου. Η μέθοδος αυτή θα δέχεται σαν ορίσματα εισόδου το γνωστικό αντικείμενο (ΠΛΗ, ΤΗΛ, ΑΡΧ, ΕΚΠ, κτλ.), τον αριθμό των εργαστηριακών εργασιών (LabModule) και τον αριθμό των ομάδων εργασίας που θα σχηματιστούν για κάθε εκτέλεση μαθήματος. **(Σημείωση:** Θα πρέπει να αξιοποιηθεί η μέθοδος που υλοποιήσατε στο ζητούμενο 3)

2. Μελέτη απόδοσης ερωτήσεων – φυσικός σχεδιασμός:

Μελετήστε το εξής αίτημα: Βρες τα ονοματεπώνυμα φοιτητών με ημερομηνία εισαγωγής (entry_date) μετά την 1/9/2010 και πριν την 1/9/2011 οι οποίοι έχουν συμμετάσχει σε ομάδες για την εκπόνηση εργαστηριακής εργασίας τύπου project με βαθμό μεγαλύτερο του 8. Ακολουθήστε τα βήματα:

- **Βήμα 1:** Χρησιμοποιήστε το αρχείο ασφαλείας (backup) της βάσης δεδομένων που συνοδεύει την εκφώνηση και το οποίο μπορείτε να κατεβάσετε από το σύνδεσμο <http://147.27.41.88:5000/sharing/NJ75GuyeN>. Η βάση αυτή περιέχει ήδη περισσότερες από 53.000 πλειάδες στους πίνακες Person και Student, περίπου 1,5εκ. πλειάδες στον πίνακα Joins, περισσότερες από 370.000 πλειάδες στον πίνακα Workgroup και περίπου 1.800 πλειάδες στον πίνακα LabModule.
- **Βήμα 2:** Μελετήστε το πλάνο και τους χρόνους εκτέλεσης χρησιμοποιώντας την EXPLAIN ANALYSE και σημειώστε τα αποτελέσματα και τις παρατηρήσεις σας στην αναφορά σας. Στη συνέχεια δοκιμάστε διαδοχικά να δημιουργήσετε κάποιο ή κάποια κατάλληλα ευρετήρια που θεωρείτε ότι μπορεί να επιταχύνει την εκτέλεση του αιτήματος και μελετήστε εκ νέου το πλάνο εκτέλεσης. Τι ευρετήριο(α)

επιλέγετε, τι τύπου και γιατί; Δοκιμάστε και σημειώστε στην αναφορά σας τις διαφορές μεταξύ διαφορετικών τύπων ευρετηρίων όπως προκύπτουν από τα αποτελέσματα που σας δίνει η EXPLAIN ANALYSE. Τι παρατηρείτε; Ποιο ευρετήριο(α) θεωρείτε ότι είναι καταλληλότερο; Δοκιμάστε επίσης να επιταχύνετε ενδεχομένως περαιτέρω το αίτημα αξιοποιώντας τη δυνατότητα ομαδοποίησης (clustering). Τι παρατηρείτε; Σημειώστε όλες τις παρατηρήσεις σας στην αναφορά σας. Επιπλέον δοκιμάστε να απενεργοποιήσετε επιλεκτικά αλγορίθμους υπολογισμού συνδέσεων (εφόσον χρησιμοποιούνται) και καταγράψτε στην αναφορά σας τις παρατηρήσεις σας όσον αφορά στην αλλαγή των πλάνων εκτέλεσης και της απόδοσής τους. Αυτό να το κάνετε τόσο πριν όσο και μετά τη δημιουργία ευρετηρίων που θα αποφασίσετε. Τέλος δοκιμάστε, τόσο πριν όσο και μετά τη δημιουργία ευρετηρίων, να αλλάξετε τη σειρά των συνδέσεων και αναφέρατε τις αλλαγές που παρατηρείτε στα πλάνα εκτέλεσης και τους χρόνους.

- **Βήμα 3:** Χρησιμοποιήστε την υλοποίηση του ζητούμενου (1.4) ώστε να αυξήσετε το πλήθος των δεδομένων της βάσης κατά εκατοντάδες χιλιάδες εισάγοντας τυχαίες εργαστηριακές εργασίες και ομάδες εργασίας για όλα τα εξαμηνιαία εργαστηριακά μαθήματα (CourseRun) των γνωστικών αντικειμένων HPY, ΣΥΣ, ΕΝΕ. **Σβήστε τυχόν ευρετήρια που δημιουργήσατε στο Βήμα (2) πριν δημιουργήσετε τις νέες πλειάδες φοιτητών.** Αφού τις δημιουργήσετε, μελετήστε εκ νέου τα πλάνα εκτέλεσης πριν και μετά τη δημιουργία ευρετηρίων ακολουθώντας τα προηγούμενα βήματα και σημειώστε για κάθε βήμα τις παρατηρήσεις δικαιολογώντας τις επιλογές σας στην αναφορά σας

Παρατηρήσεις:

- Πριν την χρήση της EXPLAIN ή της EXPLAIN ANALYSE απενεργοποιήστε τη δυνατότητα δημιουργίας παράλληλων πλάνων εκτέλεσης χρησιμοποιώντας την εντολή:

SET max_parallel_workers_per_gather = 0;

- Να εκτελείτε την EXPLAIN ANALYSE μερικές φορές πριν αποφανθείτε για τον χρόνο εκτέλεσης μιας ερώτησης καθώς οι χρόνοι αλλάζουν κάθε φορά. Στην αναφορά σας βάλτε τα λεπτομερή αποτελέσματα της EXPLAIN ANALYSE για τουλάχιστον τρεις (3) εκτελέσεις κάθε ερώτησης.

Παραδοτέο

Θα παραδώσετε ένα παραδοτέο σε μορφή zip που θα περιλαμβάνει:

1. Ένα Backup της Βάσης δεδομένων της Β' φάσης.
2. Τον κώδικα για όσα από τα ζητούμενα με (*) της Α' φάσης παραδώσατε με τη Β' Φάση καθώς επίσης και συνοπτική περιγραφή στην οποία θα αναφέρετε την αντιστοιχία συναρτήσεων κ.λπ. που υλοποιήσατε με τα ζητούμενα.
3. Το Eclipse project της εξωτερικής εφαρμογής (export σε .zip) που υλοποιήσατε στο ζητούμενο 1 της Β' φάσης.
4. Λεπτομερή αναφορά υλοποίησης σε μορφή pdf για το ζητούμενο 2 της Β' φάσης στην οποία θα αναφέρετε αιτιολογημένα τα ευρετήρια που δημιουργήσατε και όλες τις άλλες ενέργειες που κάνατε και τα αποτελέσματα από την χρήση της εντολής EXPLAIN ANALYSE και τη διαφορά στους χρόνους εκτέλεσης που παρατηρήσατε ακολουθώντας τις οδηγίες της εκφώνησης.

Ημερομηνία παράδοσης: 30/05/2022