



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ ΣΧΟΛΗ ΗΜΜΥ

ΕΡΓΑΣΤΗΡΙΟ ΔΙΑΝΕΜΗΜΕΝΩΝ
ΠΑΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ ΚΑΙ ΕΦΑΡΜΟΓΩΝ

ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ - ΠΛΗ303 Β' ΦΑΣΗ ΕΡΓΑΣΤΗΡΙΑΚΗΣ ΕΡΓΑΣΙΑΣ – ΕΑΡΙΝΟ ΕΞΑΜΗΝΟ 2022-2023

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.

Η Εφαρμογή θα παρέχει ένα μενού επιλογών στην κονσόλα για τις παρακάτω λειτουργίες:

- Παρουσίαση βαθμολογίας φοιτητή με βάση τον Α.Μ. και τον κωδικό μαθήματος.
- Αλλαγή βαθμολογίας με βάση το Α.Μ., τον κωδικό μαθήματος και τον σειριακό αριθμό εκτέλεσης μαθήματος.
- Αναζήτηση προσώπου (Καθηγητή, Εργαστηριακό προσωπικό, Φοιτητή) με βάση τα αρχικά γράμματα του επωνύμου. Τα αποτελέσματα θα παρουσιάζονται ταξινομημένα βάσει επωνύμου σε σελίδες. Αν τα αποτελέσματα είναι πάνω από πέντε (5) θα ζητείται από τον χρήστη να επιλέξει τον αριθμό προσώπων ανά σελίδα. Η εφαρμογή θα εμφανίζει τον αριθμό των σελίδων και στη συνέχεια ο χρήστης θα μπορεί είτε να εισάγει τον αριθμό σελίδας των αποτελεσμάτων που επιθυμεί να δει είτε τον χαρακτήρα 'n' για να προχωρήσει στην επόμενη σελίδα εφόσον υπάρχει.
- Παρουσίαση αναλυτικής βαθμολογίας φοιτητή με βάση τον Α.Μ. ταξινομημένα ανά τυπικό εξάμηνο.

2. Μελέτη απόδοσης ερωτήσεων – φυσικός σχεδιασμός:

Μελετήστε το εξής αίτημα: Τύπωσε τα ονόματα των πόλεων με πληθυσμό μεγαλύτερο των 50.000 κατοίκων και το πλήθος των φοιτητών που προέρχονται από αυτές και οι οποίοι έχουν συμμετάσχει σε τουλάχιστον δύο (2) προγράμματα πενταετούς διάρκειας και η ημερομηνία εισαγωγής τους είναι μεταξύ 01-09-2040 και 30-09-2050 (πρέπει να εκτυπώνονται όλες οι πόλεις από τον πίνακα Cities, ο οποίος αναφέρεται παρακάτω). Το εύρος των ημερομηνιών είναι ενδεικτικό.

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/4IDUgpUxO>. Η βάση αυτή ακολουθεί το σχήμα που ήδη χρησιμοποιήσατε στην Α' φάση με τις εξής αλλαγές:
 - Έχει προστεθεί ο πίνακας Cities που περιγράφει την πληροφορία για τις πόλεις από τις οποίες προέρχονται τα πρόσωπα. Για την αναπαράσταση αυτής της σύνδεσης στον πίνακα Person έχει προστεθεί το πεδίο city_id (ξένο κλειδί προς τον πίνακα Cities).
 - Στους πίνακες Person, Student, Joins, Program έχει προστεθεί πεδίο text που αναπαριστά (υποτιθέμενη) περιγραφική πληροφορία (descriptions).

Η βάση αυτή περιέχει ήδη περίπου 2εκ. πλειάδες στον πίνακα Person και στον πίνακα Student, περίπου 1.96εκ πλειάδες στον πίνακα Joins, 40 πλειάδες στον πίνακα Program και 445 πλειάδες στον πίνακα Cities.

- **Βήμα 2:** Μελετήστε το πλάνο και τους χρόνους εκτέλεσης της ερώτησης 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:** Κατασκευάστε κατάλληλες συναρτήσεις δημιουργίας δεδομένων για τους πίνακες Person, Student, Joins, και Program. Χρησιμοποιήστε τις συναρτήσεις αυτές ώστε να αυξήσετε το πλήθος των δεδομένων της βάσης κατά εκατοντάδες χιλιάδες εισάγοντας νέα πρόσωπα/φοιτητές, νέα προγράμματα και συνδέοντας τυχαίους φοιτητές σε τυχαία προγράμματα. **Σβήστε τυχόν ευρετήρια που δημιουργήσατε στο Βήμα (2) πριν δημιουργήσετε τις νέες εισαγωγές.** 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: 2/06/2023