

5η Εργασία: Βελτιστοποίηση Ερωτημάτων και Εναύσματα

Προθεσμία: 2/6/2023

Σκοπός:

Η παρούσα εργασία αποτελείται από δύο μέρη. Στο πρώτο μέρος, θα βελτιώσουμε το χρόνο εκτέλεσης ερωτημάτων με την προσθήκη ευρετηρίων, ενώ στο δεύτερο, θα προσθέσουμε εναύσματα ώστε να εξασφαλίσουμε την ανανέωση συγκεκριμένων γνωρισμάτων της βάσης.

Προαπαιτούμενα:

Θεωρούμε πως η βάση είναι στην κατάσταση που περιγράφεται στην 4^η Εργασία.

Ζητούμενα Εργασίας:

Μέρος Α: Ευρετήρια

- A. Να επιστρέψετε το πλάνο εκτέλεσης (screenshot) για το ακόλουθο ερώτημα το οποίο αναζητά τους ηθοποιούς μιας ταινίας βάσει του ονόματός της:

```
SELECT name, character
FROM movie join movie_cast on movie_id=id
WHERE title ='Armageddon';
```

- B. Για το προηγούμενο ερώτημα:

- Να φτιάξετε **2 ευρετήρια** τα οποία επιταχύνουν την εκτέλεση του ερωτήματος.
- Να δοθεί το αντίστοιχο πλάνο εκτέλεσης του ερωτήματος αφού έχουν φτιαχτεί τα ευρετήρια.
- Ποιές είναι οι αλλαγές που έχουν γίνει στο πλάνο εκτέλεσης;

(hint: για να είναι σωστά τα ευρετήρια, θα πρέπει να εμφανίζονται στο πλάνο εκτέλεσης)

- C. Χρησιμοποιώντας τις εντολές GROUP BY και HAVING στον πίνακα Ratings:

- Ποιό είναι το ερώτημα το οποίο επιστρέφει τον μέσο όρο βαθμολογίας ανά ταινία για ταινίες με μέση βαθμολογία πάνω από 4;
- Να τρέξετε το ερώτημα και να επιστρέψετε το αντίστοιχο πλάνο εκτέλεσης.
- Να δημιουργηθεί ένα *ευρετήριο επικάλυψης* το οποίο μπορεί να χρησιμοποιηθεί για την απάντηση στο συγκεκριμένο ερώτημα.
(hint: το ευρετήριο θα πρέπει να έχει περισσότερα του ενός γνωρίσματα)
- Καταγράψτε το πλάνο εκτέλεσης με το πλάνο επικάλυψης,
- Τι παρατηρείτε σε αυτό ως προς το διάβασμα του πίνακα Ratings και γιατί συμβαίνει αυτό;

Μέρος B: Εναύσματα

- D. Θέλουμε να προσθέσουμε ένα γνώρισμα `AVG_Rating` στον πίνακα `Movie`, το οποίο θα περιέχει την μέση βαθμολογία των χρηστών σε μία ταινία. Για κάθε νέα εισαγωγή βαθμολογίας στον πίνακα `Ratings`, θα πρέπει να ανανεώνεται η τιμή του γνωρίσματος `AVG_Rating`.
- Ο SQL κώδικας για την δημιουργία του γνωρίσματος `AVG_Rating` στον πίνακα `Movie`.
 - Χρησιμοποιώντας την εντολή `UPDATE`, αρχικοποιήστε το συγκεκριμένο γνώρισμα με την μέση βαθμολογία ανά ταινία σε περίπτωση που αυτή υπάρχει και `Null` διαφορετικά. (*hint: στο `UPDATE` του πίνακα `Movie` θα πρέπει να έχω και ένα εμφωλευμένο ερώτημα το οποίο για να αλλάξει το γνώρισμα `AVG_Rating` θα πρέπει να υπολογίζει τον μέσο όρο βαθμολογίας του πίνακα `Ratings` για το `Id` της συγκεκριμένης ταινίας. Αυτό γίνεται με την χρήση εξωτερικών μεταβλητών στο εμφωλευμένο ερώτημα.*)
 - Γράψτε ένα έναυσμα (trigger) το οποίο για κάθε εισαγωγή¹ στον πίνακα `Ratings` ενημερώνει το γνώρισμα `AVG_Rating` του πίνακα `Movie` με την μέση βαθμολογία μόνο των ταινιών των οποίων η βαθμολογία άλλαξε. Ελέγξτε ότι η συνάρτησή σας λειτουργεί σωστά εισάγοντας βαθμολογίες σε κάποια/ες από τις ταινίες και βλέποντας αν το συγκεκριμένο πεδίο έχει ενημερωθεί σωστά. (*hint: Προσαρμόστε το `update` του προηγούμενου ερωτήματος με τις κατάλληλες αλλαγές οι οποίες θα διασφαλίζουν ότι δεν επανυπολογίζονται όλα τα `Movie.AVG_Rating`, αλλά μόνο αυτά για τις ταινίες στις οποίες είχαμε νέα βαθμολογία).*)
 - Ποιό από τα ευρετήρια που φτιάξατε βοηθάει στην εκτέλεση του συγκεκριμένου trigger;

- E. **Bonus²:** Δοθέντος ότι για δύο σύνολα τιμών S_1, S_2 ισχύει για την μέση τιμή ότι
- $$AVG(S_1 \cup S_2) = \frac{SUM(S_1) + SUM(S_2)}{COUNT(S_1) + COUNT(S_2)}$$
- (όπου η ένωση αναφέρεται σε πολυσύνολα), να περιγράψτε (χωρίς να υλοποιήσετε) κάποιον εναλλακτικό τρόπο με τον οποίο θα μπορούσατε να υλοποιήσετε το trigger σας, χωρίς να χρειαστεί να τρέχετε κάποια συνάθροιση σε παλαιότερες εγγραφές του πίνακα `Ratings(user_id, movie_id, rating)`. (*hint: Θεωρείστε ότι μπορείτε να προσθέσετε επιπλέον γνωρίσματα στον πίνακα `Movie`.*)



Πλάνο Εκτέλεσης

Καθώς η SQL είναι μια δηλωτική (declarative) γλώσσα, ένα ερώτημα περιγράφει ποιο θέλουμε να είναι το αποτέλεσμα του, παρά το ποια είναι ακριβώς τα βήματα με τα οποία θα φτάσει η βάση δεδομένων στο αποτέλεσμα αυτό. Συνεπώς, ένα σύστημα διαχείρισης βάσεων δεδομένων (ΣΔΒΣ) μπορεί να εκτελέσει ένα ερώτημα με πολλούς διαφορετικούς τρόπους και επιλέγοντας από ένα σύνολο αλγορίθμων για κάθε τελεστή. Το πλάνο εκτέλεσης μας δίνει την ακολουθία των βημάτων (τον τρόπο) που ακολουθήθηκε από το ΣΔΒΣ για την εκτέλεση του ερωτήματος.







¹ Για απολοποίηση θεωρείστε ότι έχουμε μόνο εισαγωγές και καθόλου διαγραφές ή ενημερώσεις.

² Το συγκεκριμένο ερώτημα είναι προαιρετικό.

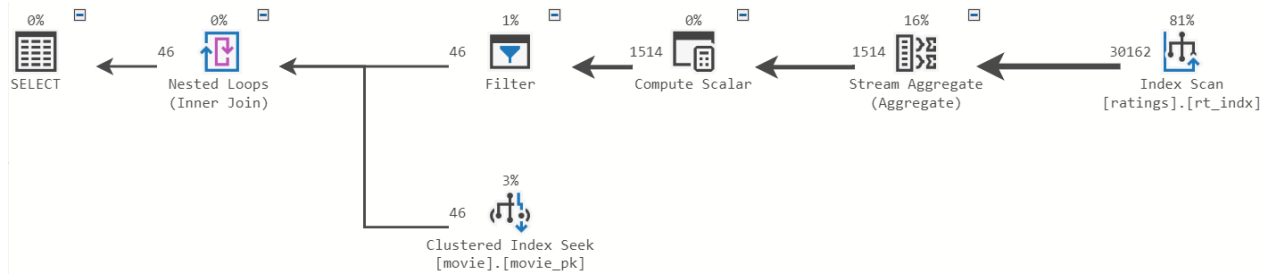
Πηγαίνοντας στο Microsoft Azure μπορείτε να οπτικοποιήσετε το πλάνο εκτέλεσης ενός ερωτήματος με τους ακόλουθους τρόπους:

- Γράφοντας το ερώτημα και πατώντας το κουμπί  **Estimated Plan** το Azure επιστρέφει το εκτιμώμενο πλάνο εκτέλεσης ενός ερωτήματος, χωρίς να τρέξει το αντίστοιχο ερώτημα.
- Πριν την εκτέλεση του ερωτήματος, πατώντας το κουμπί  **Enable Actual Plan**, ενεργοποιείται η εμφάνιση του πλάνου εκτέλεσης. Μετά την εκτέλεση κάθε ερωτήματος, εμφανίζεται και το πλάνο εκτέλεσης στο αντίστοιχο TAB.

Στην συνέχεια απεικονίζεται το πλάνο εκτέλεσης ενός ερωτήματος ως μια δεντρική δομή όπου οι *ενδιάμεσοι κόμβοι αντιστοιχούν σε πράξεις* οι οποίες γίνονται στους πίνακές μας, ενώ οι *κόμβοι φύλλα* αντιστοιχούν σε σάρωση ή αναζήτηση των εγγραφών που υπάρχουν στους πίνακες ή τα ευρετήρια της βάσης μας. Ένας *κόμβος φύλλο*, έχει μόνο εξερχόμενες και καθόλου εισερχόμενες ακμές. Στο παρακάτω πλάνο εκτέλεσης αρχικά το ΣΔΒΔ εκτελεί τις ακόλουθες ενέργειες:

-  Σαρώνει το ευρετήριο `rt_idx` του πίνακα `Ratings`, επιστρέφοντας τις εγγραφές του στον επόμενο τελεστή.
-  Υπολογίζει μία συνάρτηση συνάθροισης στις ήδη ταξινομημένες ως προς τα γνωρίσματα του `GROUP BY` βάσει του ευρετηρίου, εγγραφές (λόγω του ευρετηρίου).
-  Αλλάζει τον τύπο του αποτελέσματος της συνάθροισης σε σχέση με τον αρχικό τύπο του πεδίου.
-  Επιλέγει συγκεκριμένα αποτελέσματα ως έξοδο του προηγούμενου τελεστή, βάσει μιας συνθήκης `HAVING` που υπάρχει στο αποτέλεσμα της συνάθροισης.
-  Κάνει σύζευξη του προηγούμενου αποτελέσματος με τον πίνακα `Movie`, χρησιμοποιώντας το πρωτεύον ευρετήριο το οποίο έχει χτιστεί στο πρωτεύον κλειδί του πίνακα.
-  Κάνει προβολή του τελικού αποτελέσματος στα γνωρίσματα που υπάρχουν στο `SELECT` κομμάτι ενός ερωτήματος.

Κάνοντας κλικ σε κάθε τελεστή, μπορούμε να δούμε επιπλέον πληροφορίες που σχετίζονται με την εκτέλεση του συγκεκριμένου τελεστή όπως το εκτιμώμενο κόστος διαβάσματος από τον δίσκο, το κόστος χρήσης της CPU, οι πλειάδες που διαβάζονται από τον τελεστή. Πάνω από κάθε τελεστή βλέπουμε το κόστος εκτέλεσης του τελεστή ως προς το συνολικό κόστος εκτέλεσης του ερωτήματος (π.χ. η πράξη `filter` αποτελεί το 1% του συνολικού κόστους), ενώ στην έξοδο κάθε τελεστή έχουμε τον εκτιμώμενο αριθμό των πλειάδων που προκύπτουν μετά την εκτέλεση του τελεστή (π.χ., ο τελεστής `filter` έχει σαν έξοδο 46 πλειάδες).



Συμβουλές για την υλοποίηση:

- Επιβεβαιώστε ότι ο χρήστης της βάσης του οποίου μας στέλνετε τα credentials όντως έχει πρόσβαση στη βάση σας και μπορεί να τρέχει SELECT queries.
- Στις συζητήσεις του μαθήματος θα βρείτε χρήσιμες πληροφορίες στα πεδία “Συχνές ερωτήσεις” και “Χρήσιμες Εντολές”

Χρήσιμα links:

Βασική εντολή για δημιουργία ευρετηρίου:

<https://learn.microsoft.com/en-us/sql/t-sql/statements/create-index-transact-sql?view=sql-server-ver16>

Δημιουργία εναύσματος:

<https://learn.microsoft.com/en-us/sql/t-sql/statements/create-trigger-transact-sql?view=sql-server-ver16>

Προσθήκη Γνωρίσματος:

<https://learn.microsoft.com/en-us/sql/t-sql/statements/alter-table-transact-sql?view=sql-server-ver16>

Εντολή Ανανέωσης:

<https://learn.microsoft.com/en-us/sql/t-sql/queries/update-transact-sql?view=sql-server-ver16>

Παραδοτέα:

- Δημιουργήστε ένα .txt αρχείο στο οποίο θα αναγράφονται το endpoint του Azure instance σας (Server name στο Overview tab του Azure), το όνομα της βάσης σας και το username και το password ενός χρήστη με read-only δικαιώματα, ώστε να μπορούμε να δούμε τους πίνακες της βάσης σας. Το .txt αρχείο θα πρέπει να έχει την παρακάτω μορφή:
Endpoint: <name_of_the_endpoint>
Username: <username>
Password: <password>
Database: <name_of_the_database>
- Απαντήστε στα επιμέρους ερωτήματα χρησιμοποιώντας το **πρότυπο για την απάντηση που σας δόθηκε**.
- Το όνομα του φακέλου πρέπει να αποτελείται από τους αριθμούς μητρώου σας χωρισμένους με παύλα, δηλαδή αριθμός_μητρώου_1-αριθμός_μητρώου_2. Δημιουργήστε ένα .zip αρχείο αυτού του φακέλου, το οποίο θα έχει το ίδιο όνομα με τον φάκελο.
- Κάντε υποβολή το .zip αρχείο στο eclass στην ενότητα Εργασίες / 5η Εργασία.