

5η Εργασία: Βελτιστοποίηση Ερωτημάτων και Εναύσματα

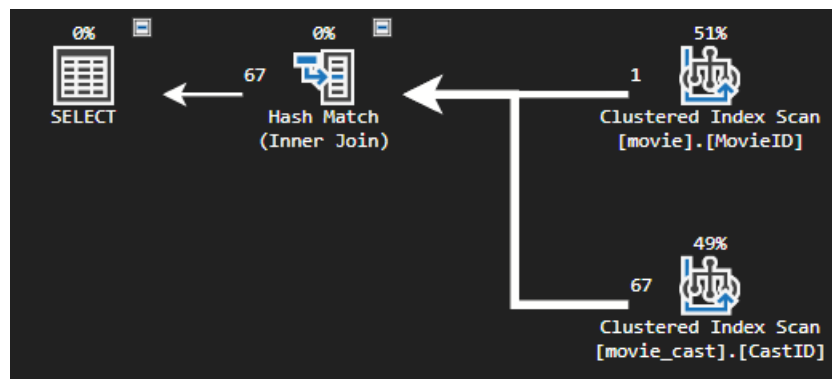
Ονοματεπώνυμο

Μητσάκης Νίκος : p3210122

Παντελίδης Ιπποκράτης : p3210150

Ερώτημα Α

Να δοθεί screenshot του πλάνου εκτέλεσης του ερωτήματος:

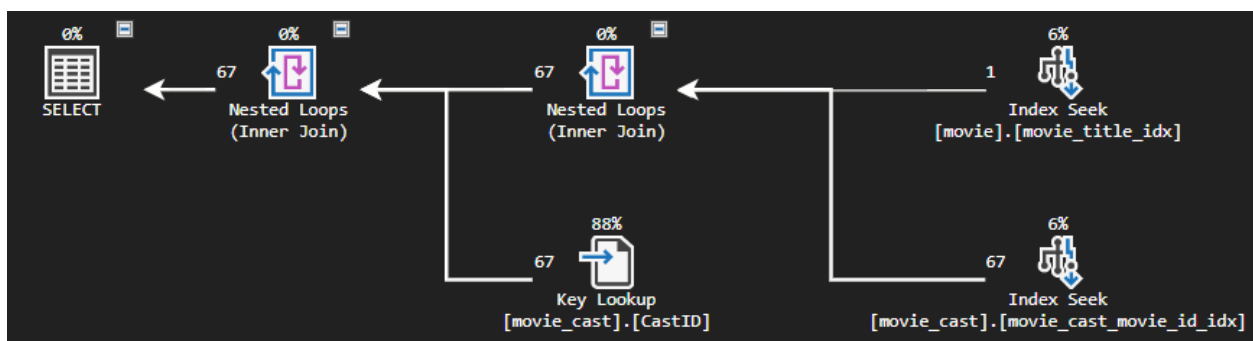


Ερώτημα Β

- a. Να φτιάξετε **2 ευρετήρια** τα οποία επιταχύνουν την εκτέλεση του ερωτήματος.

```
/* Creating index for movie table */  
CREATE INDEX movie_title_idx ON movie (title);  
  
/* Creating index for movie_cast table */  
CREATE INDEX movie_cast_movie_id_idx ON movie_cast (movie_id);
```

- b. Να δοθεί το αντίστοιχο πλάνο εκτέλεσης του ερωτήματος αφού έχουν φτιαχτεί τα ευρετήρια.



- c. Ποιές είναι οι αλλαγές που έχουν γίνει στο πλάνο εκτέλεσης;

Με την χρήση του πρώτου ευρετηρίου εντοπίζεται ταχύτατα ο τίτλος της ταινίας χωρίς να διασχιστεί ολόκληρος ο πίνακας `monie` όπως όταν δεν χρησιμοποιούμε. Επίσης το δεύτερο ευρετήριο παίρνει τις γραμμές του `monie_cast` που αντιστοιχούν στο `id` του `monie` του πρώτου βήματος. Η βασική διαφορά είναι ότι η `join` θα εκτελεστεί στις φιλτραρισμένες γραμμές και όχι σε ολόκληρους τους πίνακες. Επίσης το δεύτερο ευρετήριο αντί να δείχνει στο δίσκο δείχνει στην τιμή του πρωτεύοντος κλειδιού τις οποίες μετά θα ψάξουμε στο πρώτο ευρετήριο, γεγονός που δηλώνεται από την εμφάνιση του `Keylookup` στο δεύτερο πλάνο εκτέλεσης.

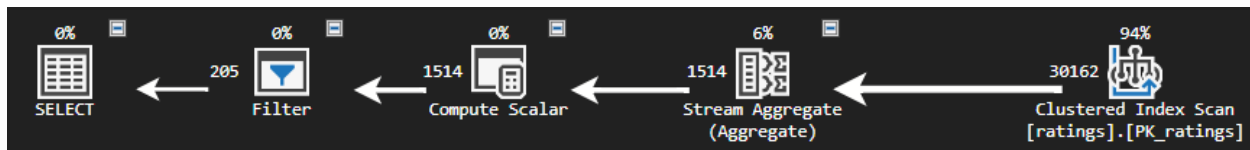
Ερώτημα C

Χρησιμοποιώντας τις εντολές `GROUP BY` και `HAVING` στον πίνακα `Ratings`:

- a. Ποιό είναι το ερώτημα το οποίο επιστρέφει τον μέσο όρο βαθμολογίας ανά ταινία για ταινίες με μέση βαθμολογία πάνω από 4;

```
SELECT movie_id, AVG(rating) as Average_Rating
FROM ratings
GROUP BY movie_id
HAVING AVG(rating) > 4;
/*OUTPUT: 205 rows.*/
```

- b. Να τρέξετε το ερώτημα και να επιστρέψετε το αντίστοιχο πλάνο εκτέλεσης.



- c. Να δημιουργηθεί ένα *ευρετήριο επικάλυψης* το οποίο μπορεί να χρησιμοποιηθεί για την απάντηση στο συγκεκριμένο ερώτημα.

```
CREATE INDEX rating_idx ON ratings (movie_id, rating);
```

- d. Καταγράψτε το πλάνο εκτέλεσης με το πλάνο επικάλυψης,



- e. τι παρατηρείτε σε αυτό ως προς το διάβασμα του πίνακα `Ratings` και γιατί συμβαίνει αυτό;

Παρατηρούμε ότι συνεχίζονται να διαβάζονται όλες οι γραμμές του πίνακα `ratings` όμως με την χρήση του ευρετηρίου παρατηρείται μειωμένος αριθμός επεξεργασμένων γραμμών. Αυτό

οφείλεται στο γεγονός ότι το ευρετήριο είναι πιο στοχευμένο στις γραμμές που επεξεργάζεται και αγνοεί αυτές για τις οποίες δεν ισχύει η συνθήκη.

Ερώτημα D

- a. Να δοθεί η εντολή Alter Table για την δημιουργία του γνώρισματος AVG_Rating στον πίνακα Movie.

```
ALTER TABLE movie ADD avg_rating float;
```

- b. Χρησιμοποιώντας την εντολή UPDATE, αρχικοποιήστε το συγκεκριμένο γνώρισμα με την μέση βαθμολογία ανά ταινία σε περίπτωση που αυτή υπάρχει και Null διαφορετικά.

```
UPDATE movie
SET avg_rating = (
    SELECT AVG(rating)
    FROM ratings
    WHERE ratings.movie_id = movie.id
    GROUP BY ratings.movie_id
);
```

- c. Γράψτε ένα *έναυσμα* (trigger) το οποίο για κάθε εισαγωγή στον πίνακα Ratings ενημερώνει το γνώρισμα AVG_Rating του πίνακα Movie με την μέση βαθμολογία μόνο των ταινιών των οποίων η βαθμολογία άλλαξε.

```
CREATE TRIGGER update_avg_rating_trigger
ON ratings
AFTER INSERT
AS
BEGIN
    -- Calculate the average rating for the specific movie
    UPDATE Movie
    SET avg_rating = (
        SELECT AVG(rating)
        FROM ratings
        WHERE movie_id = inserted.movie_id
        GROUP BY movie_id
    )
    FROM inserted
    WHERE movie.id = inserted.movie_id;
END;
```

- d. Ποιό από τα ευρετήρια που φτιάξατε βοηθάει στην εκτέλεση του συγκεκριμένου trigger.

Το ευρετήριο που φτιάξαμε στο ερώτημα C.c βοηθάει στην εκτέλεση του συγκεκριμένου trigger καθώς όταν γίνεται μία εισαγωγή στον πίνακα ratings κανονικά χρειάζεται να υπολογιστεί η μέση βαθμολογία της ταινίας παίρνοντας όλες τις βαθμολογίες της ταινίας. Με την χρήση όμως του ευρετηρίου μπορούν εύκολα και ταχύτατα να εντοπιστούν οι βαθμολογίες μια δοσμένης ταινίας.

Ερώτημα Ε

Περιγράψτε κάποιον εναλλακτικό τρόπο με τον οποίο θα μπορούσατε να υλοποιήσετε το trigger σας, χωρίς να χρειαστεί να τρέχετε κάποια group by συνάρτηση στα υπάρχοντα δεδομένα.

Το παραπάνω trigger θα μπορούσε να υλοποιηθεί και χωρίς την χρήση συνάρτησης GROUP BY αν χρησιμοποιούσαμε την πληροφορία που μας δίνεται στην εκφώνηση για την μέση τιμή δύο συνόλων. Εφόσον μπορούμε να προσθέσουμε επιπλέον γνωρίσματα στον πίνακα monie θα ήταν συνετό να του προσθέσουμε δύο νέα, όπου το ένα θα είναι το πλήθος των κριτικών που έχουν γίνει στην συγκεκριμένη ταινία και το άλλο η συνολική βαθμολογία που η ταινία αυτή συγκέντρωσε. Επομένως, αντί να συμπεριλαμβάνουμε αμέσως την μέση τιμή σαν πεδίο του monie περνάμε μόνο τα επιμέρους στοιχεία αυτής της πράξης και μπορούμε έτσι να τα χρησιμοποιήσουμε ώστε να πάρουμε το επιθυμητό αποτέλεσμα χωρίς την χρήση της GROUP BY. Έτσι το trigger μας θα έκανε απλά την πράξη (συνολικό άθροισμα / πλήθος) αφού πρώτα υπολόγιζε αυτές τις δύο αυξημένες τιμές, αφού άλλωστε το trigger συμβαίνει με την εισαγωγή. Δηλαδή θα υπολογίσει το αυξημένο κατά το ποσό των εισαγωγών μετρητή και το αυξημένο κατά το άθροισμα των εισαγόμενων κριτικών νέο συνολικό άθροισμα.