Αναφορά Project A

Φίλιππος Δουραχαλής, 3170045

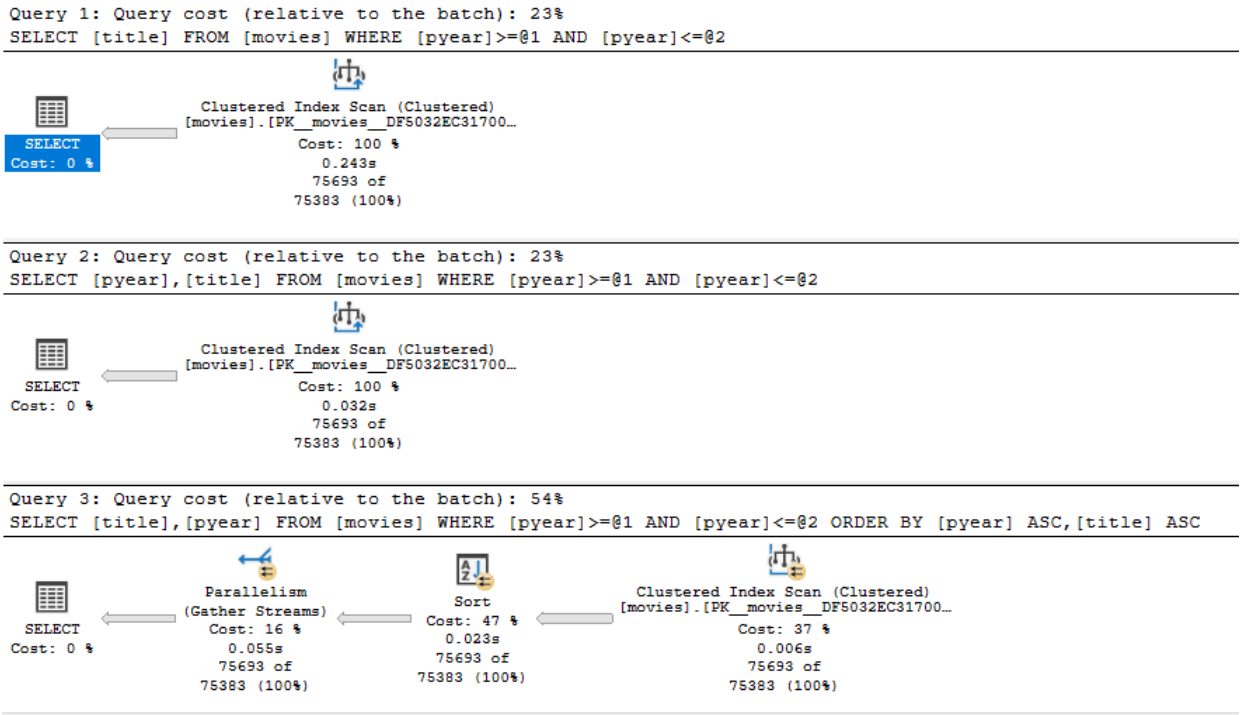
Ζήτημα 1)

1. Χρησιμοποιούμε την εντολή:

   CREATE INDEX moviesIndex ON movies (pyear) INCLUDE (title)
   WHERE pyear >= 1990 AND pyear <= 2000

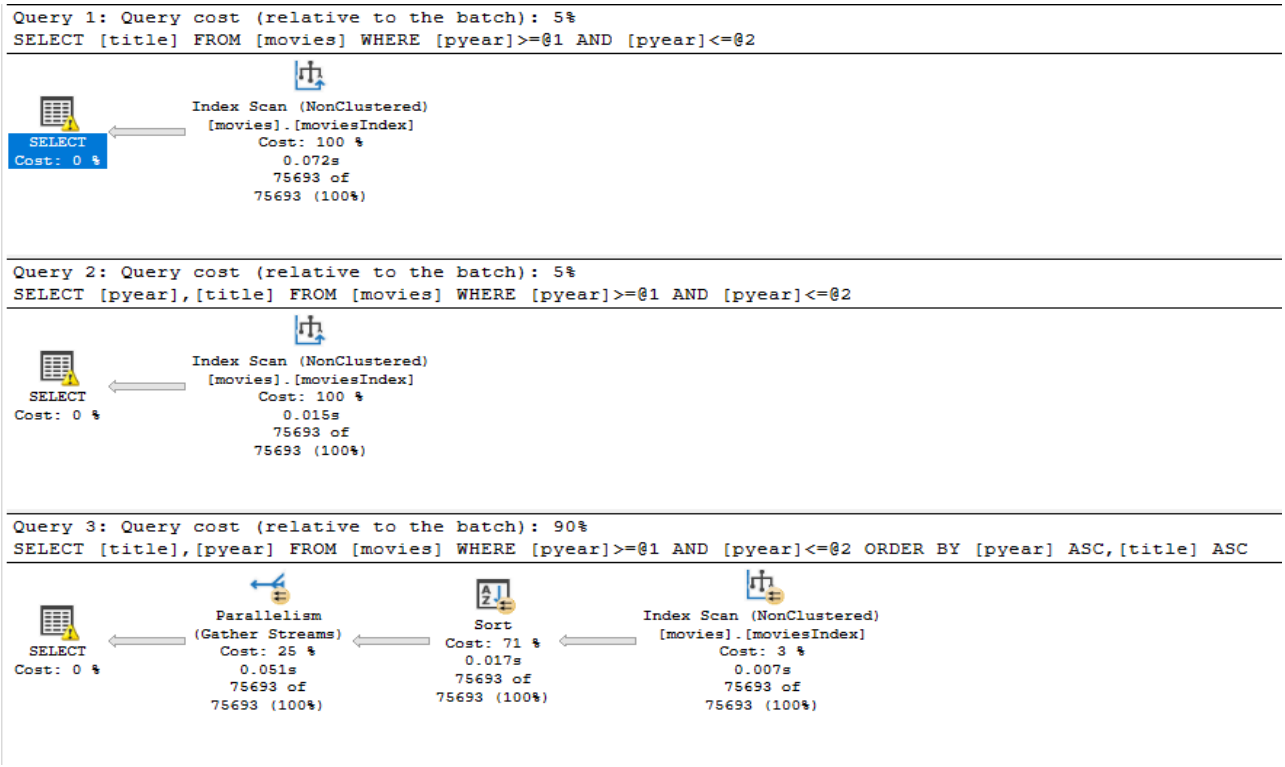
   Δημιουργούμε ένα ευρετήριο πάνω στο πεδίο pyear της σχέσης movies. Το ευρετήριο θα περιέχει μόνο τις τιμές που ικανοποιούν την πρόταση WHERE του ερωτήματος προκειμένου η ανάκτηση των αντίστοιχων εγγραφών να γίνεται πιο αποτελεσματικά. Τέλος προσθέτουμε τη στήλη title ως included column στο ευρετήριο. Η τιμή αυτή της στήλης δεν χρησιμοποιείται για την ανάκτηση, αλλά αποθηκεύεται στο ευρετήριο ώστε να μπορεί να ανακτηθεί χωρίς να γίνει αναζήτηση στη σχέση. Έτσι χρησιμοποιώντας το ευρετήριο μπορούμε να απαντήσουμε στο ερώτημα χωρίς να γίνει καθόλου αναζήτηση στη σχέση movies, όπως βλέπουμε στα παρακάτω πλάνα εκτέλεσης.

   Χωρίς ευρετήριο:



```
(75693 rows affected)
Table 'movies'. Scan count 1, logical reads 1918, physical reads 2, page server reads 0, read-ahead reads 1917, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead re

(1 row affected)

(75693 rows affected)
Table 'movies'. Scan count 1, logical reads 1918, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads

(1 row affected)

(75693 rows affected)
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads
Table 'movies'. Scan count 7, logical reads 2012, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads

(1 row affected)

Completion time: 2021-05-13T16:46:40.0535782+03:00
```

Με ευρετήριο:




```
(75693 rows affected)
Table 'movies'. Scan count 1, logical reads 350, physical reads 1, page server reads 0, read-ahead reads 355, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead read

(1 row affected)

(75693 rows affected)
Table 'movies'. Scan count 1, logical reads 350, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads

(1 row affected)

(75693 rows affected)
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads
Table 'movies'. Scan count 7, logical reads 1045, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads

(1 row affected)

Completion time: 2021-05-13T17:08:26.0352526+03:00
```

Παρατηρούμε ότι με τη χρήση ευρετηρίου τόσο ο χρόνος εκτέλεσης της αναζήτησης των σωστών εγγραφών, όσο και ο συνολικός αριθμός I/Os μειώθηκε σημαντικά.

2. Για την δημιουργία του ευρετηρίου χρησιμοποιούμε την εντολή

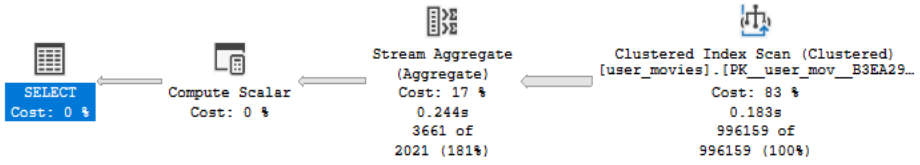
   CREATE INDEX user_index ON user_movies (mid, userid, rating)

   Το παρόν είναι ένα multi-column index το οποίο χρησιμοποιεί τις τιμές και των τριών παραπάνω στηλών για την δημιουργία των κόμβων του δέντρου

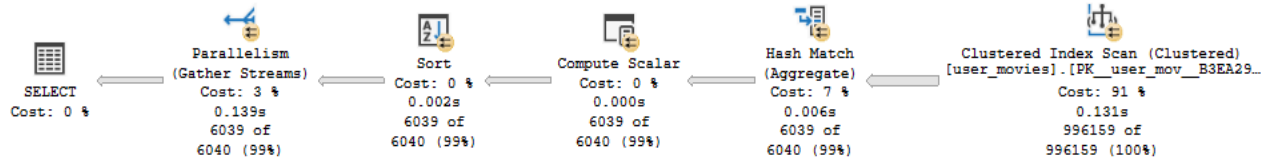
και την ευρετηρίαση. Η αναζήτηση γίνεται αρχικά μέσω της τιμής mid, καθώς η στήλη αυτή περιέχει λιγότερες μοναδικές τιμές και έτσι μειώνει το κόστος αποθήκευσης του δέντρου και της αναζήτησης. Εφόσον η τιμή του πεδίου αυτού δεν αρκεί, όπως στην περίπτωση του δεύτερου επερωτήματος, χρησιμοποιείται και η τιμή της στήλης userid για την αναζήτηση. Το πεδίο rating ευρετηριάζεται και αυτό ώστε να βοηθήσει στην εκτέλεση της συνάρτησης count. Το τελευταίο πεδίο περιλαμβάνεται επίσης για να αυξήσει τις πιθανότητες του ευρετηρίου να χρησιμοποιηθεί από τον SQL Server, αφού διαφορετικά εκείνος μπορεί να κρίνει ότι η χρήση του δεν οδηγεί σε βελτίωση επειδή χωρίς το rating θα χρειαζόταν επιπλέον να ανατρέξει στον πίνακα. Η επίδραση του ευρετηρίου φαίνεται στις παρακάτω εικόνες.

Χωρίς ευρετήριο:

```
Query 1: Query cost (relative to the batch): 59%
select mid,count(rating) from user_movies group by mid order by mid
```

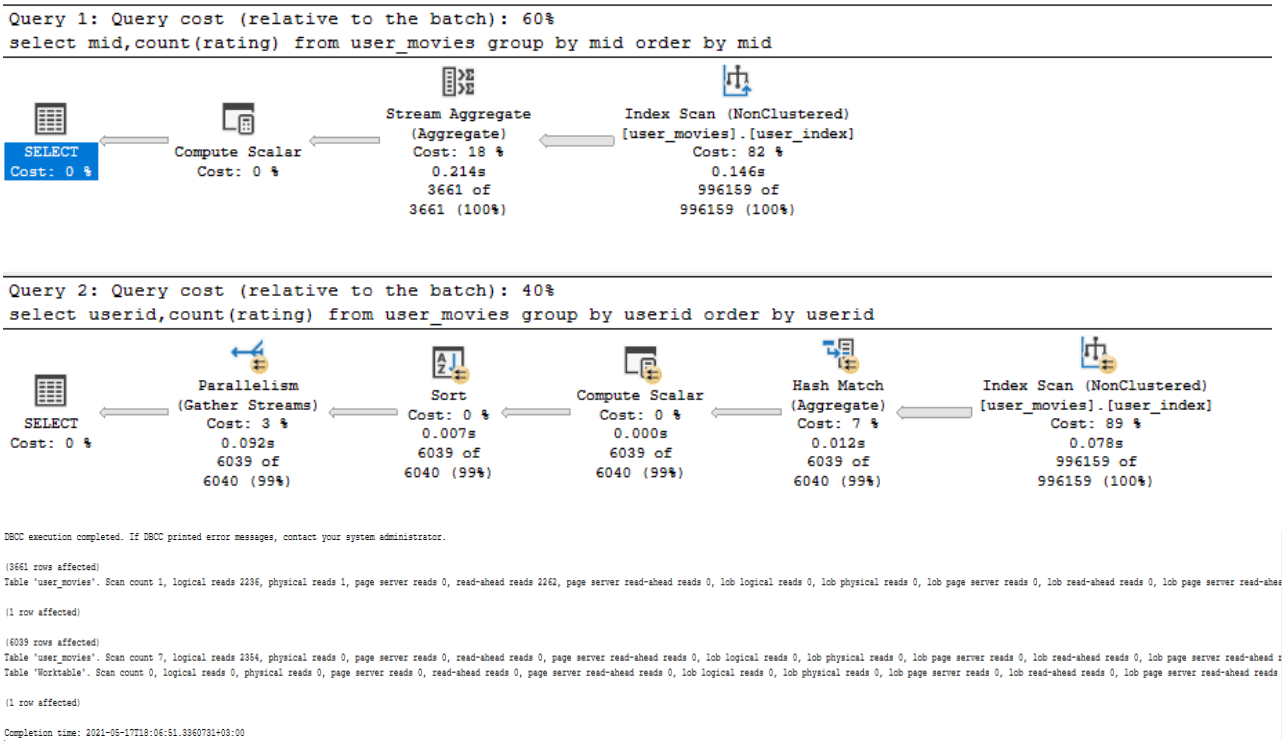



```
Query 2: Query cost (relative to the batch): 41%
select userid,count(rating) from user_movies group by userid order by userid
```




```
DBCC execution completed. If DBCC printed error messages, contact your system administrator.

(3661 rows affected)
Table 'user_movies'. Scan count 1, logical reads 2601, physical reads 2, page server reads 0, read-ahead reads 2603, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahea

(1 row affected)
DBCC execution completed. If DBCC printed error messages, contact your system administrator.

(6039 rows affected)
Table 'user_movies'. Scan count 7, logical reads 2783, physical reads 2, page server reads 0, read-ahead reads 2603, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahea
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads

(1 row affected)

Completion time: 2021-05-17T20:32:54.9999631+03:00
```

Με ευρετήριο:



Και σε αυτήν τη περίπτωση παρατηρούμε μια -μικρότερη- βελτίωση στους χρόνους εκτέλεσης του ερωτήματος αλλά και μια μικρή μείωση στις σελίδες που διαβάζονται (logical reads)

Ζήτημα 2)

1. Αντικαθιστούμε την ένωση των σχέσεων που επιστρέφουν τα δύο ερωτήματα select με μια σύζευξη στην πρόταση WHERE του ερωτήματός μας. Συγκεκριμένα, επιλέγουμε τις πλειάδες της σχέσης movie_genre που στην στήλη genre περιέχουν είτε την τιμή “Adventure” ή την τιμή “Action” (η στήλη αυτή περιέχει ατομικές τιμές, οπότε αρκεί να προσδιορίσουμε αυτούσιο το string που θέλουμε να βρούμε, χωρίς τον τελεστή LIKE). Στη συνέχεια κάνουμε JOIN τη σχέση που επιστρέφεται από το εμφωλευμένο ερώτημα με την σχέση movies ώστε να πάρουμε τους τίτλους των ταινιών βάσει του mid.
   Για την σύγκριση των δύο ερωτημάτων, εκτός από το πλάνο εκτέλεσης και το πλήθος των I/Os χρησιμοποιήθηκαν επίσης στατιστικά του χρόνου εκτέλεσης και τα ερωτήματα εκτελέστηκαν διαδοχικά 10 φορές το καθένα ώστε να εξετάσουμε τον μέσο χρόνο εκτέλεσης του καθενός και να επιβεβαιώσουμε ότι το εναλλακτικό ερώτημα είναι πιο αποδοτικό. Παρακάτω βλέπουμε ότι ο μέσος χρόνος εκτέλεσης του δεύτερου ερωτήματος είναι 285ms έναντι των 342 του πρώτου. Παρατηρούμε επίσης ότι μειώνονται σημαντικά τα βήματα και οι χρόνοι εκτέλεσης του πλάνου, καθώς και ο αριθμός των σελίδων που διαβάζονται.

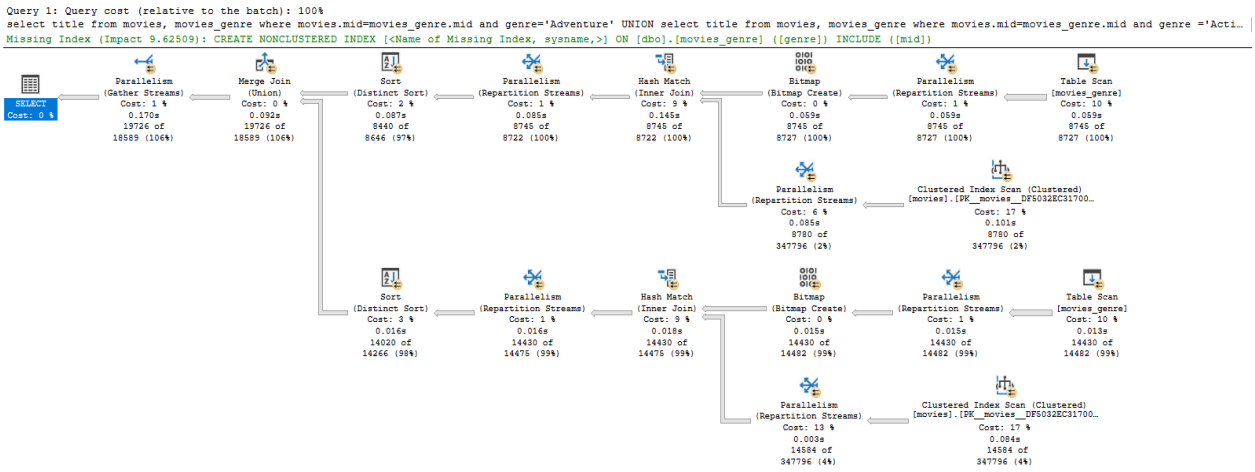
Το νέο ερώτημα είναι:

```
SELECT DISTINCT title
FROM movies_genre
INNER JOIN movies ON movies.mid = movies_genre.mid
WHERE genre = 'Action' or genre = 'Adventure'
```

Τέλος δημιουργούμε επίσης δύο ευρετήρια. Το πρώτο πάνω στις στήλες “genre” και “mid” της σχέσης movies_genre, που χρησιμοποιούνται στην πρόταση WHERE και JOIN, ώστε να βελτιώσουμε περαιτέρω το εναλλακτικό ερώτημα μας. Επίσης ορίζεται ένα ευρετήριο πάνω στο πεδίο mid της σχέσης movies, το οποίο περιλαμβάνει επιπλέον την στήλη “title”, ώστε να επιταχυνθεί η ανάκτησή των τιμών της καθώς και το JOIN με τον πίνακα movies_genre.

```
CREATE INDEX genre_index ON movies_genre (genre, mid)

CREATE INDEX movies_index ON movies (mid) INCLUDE (title)
```

Αρχικό Ερώτημα:

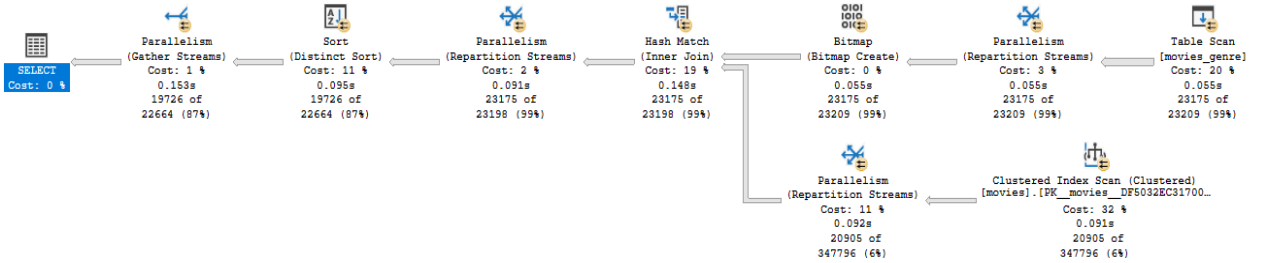



```
DBCC execution completed. If DBCC printed error messages, contact your system administrator.

(19726 rows affected)
Table 'movies_genre'. Scan count 14, logical reads 2246, physical reads 0, page server reads 0, read-ahead reads 1123, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ah
Table 'movies'. Scan count 14, logical reads 4024, physical reads 2, page server reads 0, read-ahead reads 1917, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead re
Table 'Workfile'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads 0
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads

(1 row affected)

Completion time: 2021-05-17T12:00:25.4867863+03:00
```

	Trial 10	Trial 9	Trial 8	Trial 7	Trial 6	Trial 5	Trial 4	Trial 3	Trial 2	Trial 1	Average
Query Profile Statistics											
Number of INSERT, DELETE and UPDATE statements	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0.0000
Rows affected by INSERT, DELETE, or UPDATE stateme...	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0.0000
Number of SELECT statements	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4.0000
Rows returned by SELECT statements	19728 →	19728 →	19728 →	19728 →	19728 →	19728 →	19728 →	19728 →	19728 →	19728 →	19728.0000
Number of transactions	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0.0000
Network Statistics											
Number of server roundtrips	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4.0000
TDS packets sent from client	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4.0000
TDS packets received from server	134 →	134 →	134 →	134 →	134 →	134 →	134 →	134 →	134 →	134 →	134.0000
Bytes sent from client	772 →	772 →	772 →	772 →	772 →	772 →	772 →	772 →	772 →	772 →	772.0000
Bytes received from server	534542 ↑	534534 ↑	534206 ↑	534174 ↓	534478 ↑	534382 ↑	534220 ↓	534372 ↑	534332 ↓	534572 →	534381.2000
Time Statistics											
Client processing time	46 ↑	39 ↑	37 ↓	45 ↑	42 ↓	43 ↓	47 ↓	49 ↑	40 ↓	45 →	43.3000
Total execution time	269 ↑	268 ↑	236 ↓	241 ↓	1111 ↑	247 ↓	277 ↑	274 ↑	266 ↑	232 →	342.1000
Wait time on server replies	223 ↓	229 ↑	199 ↑	196 ↓	1069 ↑	204 ↓	230 ↑	225 ↓	226 ↑	187 →	298.8000

Εναλλακτικό ερώτημα χωρίς ευρετήριο:

```
Query 1: Query cost (relative to the batch): 100%
select DISTINCT title from (SELECT mid FROM movies_genre WHERE genre = 'Action' or genre = 'Adventure') Q INNER JOIN movies ON movies.mid = Q.mid
Missing Index (Impact 16.5244): CREATE NONCLUSTERED INDEX [<Name of Missing Index, sysname,>] ON [dbo].[movies_genre] ([genre]) INCLUDE ([mid])
```

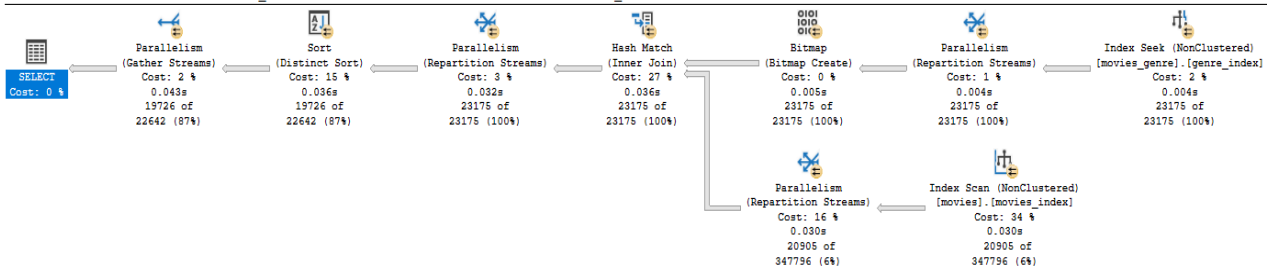



```
DBCC execution completed. If DBCC printed error messages, contact your system administrator.

(19726 rows affected)
Table 'movies_genre'. Scan count 7, logical reads 1123, physical reads 0, page server reads 0, read-ahead reads 1123, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahe
Table 'movies'. Scan count 7, logical reads 2012, physical reads 2, page server reads 0, read-ahead reads 1917, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead rea
Table 'Workfile'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads 0
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads

(1 row affected)

Completion time: 2021-05-17T13:09:34.1152339+03:00
```

	Trial 10	Trial 9	Trial 8	Trial 7	Trial 6	Trial 5	Trial 4	Trial 3	Trial 2	Trial 1	Average
Client Execution Time	13:09:34	13:09:33	13:09:32	13:09:31	13:09:29	13:09:26	13:09:23	13:09:21	13:08:59	13:08:58	
Query Profile Statistics											
Number of INSERT, DELETE and UPDATE statements	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0.0000
Rows affected by INSERT, DELETE, or UPDATE stateme...	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0.0000
Number of SELECT statements	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4.0000
Rows returned by SELECT statements	19728 →	19728 →	19728 →	19728 →	19728 →	19728 →	19728 →	19728 →	19728 →	19728 →	19728.0000
Number of transactions	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0 →	0.0000
Network Statistics											
Number of server roundtrips	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4.0000
TDS packets sent from client	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4 →	4.0000
TDS packets received from server	120 →	120 →	120 →	120 →	120 →	120 →	120 →	120 →	120 →	120 →	120.0000
Bytes sent from client	664 →	664 →	664 →	664 →	664 →	664 →	664 →	664 →	664 →	664 →	664.0000
Bytes received from server	476946 →	476946 ↑	476932 ↓	477146 ↑	477144 ↑	476954 ↓	477102 ↓	477230 ↑	477112 ↑	476948 →	477046.0000
Time Statistics											
Client processing time	41 ↓	43 ↓	47 ↑	44 ↓	48 ↑	45 ↑	39 ↓	53 ↑	51 ↑	27 →	43.8000
Total execution time	256 ↑	243 ↓	251 →	251 ↓	255 ↑	236 ↓	245 ↓	615 ↑	262 ↑	244 →	285.8000
Wait time on server replies	215 ↑	200 ↓	204 ↓	207 →	207 ↑	191 ↓	206 ↓	562 ↑	211 ↓	217 →	242.0000

Εναλλακτικό ερώτημα με ευρετήριο:

```
Query 1: Query cost (relative to the batch): 100%
select DISTINCT title FROM movies_genre INNER JOIN movies ON movies.mid = movies_genre.mid WHERE genre = 'Action' or genre = 'Adventure'
```




```
DBCC execution completed. If DBCC printed error messages, contact your system administrator.

(19726 rows affected)
Table 'movies_genre'. Scan count 13, logical reads 95, physical reads 2, page server reads 0, read-ahead reads 84, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead
Table 'movies'. Scan count 7, logical reads 1473, physical reads 3, page server reads 0, read-ahead reads 1418, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead rea
Table 'Workfile'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads 0
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads

(1 row affected)

Completion time: 2021-05-18T02:03:56.9893321+03:00
```

Βλέπουμε και πάλι ότι το ευρετήριο μειώνει δραστικά τον χρόνο εκτέλεσης, αφού πλέον το σύστημα χρησιμοποιεί IndexSeek, ενώ μειώνονται και ο αριθμός των logical reads.

2.

a. Για το πρώτο ερώτημα που χρησιμοποιούμε, εκτελούμε δύο διαφορετικά SELECT:
   - Ένα ώστε να αριθμήσουμε τους ηθοποιούς που παίζουν σε μια ταινία.
   - Ένα για να αριθμήσουμε μόνο τους άνδρες ηθοποιούς που παίζουν σε κάθε ταινία.

Στη συνέχεια παίρνουμε την τομή των σχέσεων που δημιουργούμε. Αν τα δύο πλήθη που υπολογίστηκαν για μια ταινία είναι ίσα (άρα στην ταινία παίζουν μόνο άνδρες ηθοποιοί) τότε θα επιστραφεί η πλειάδα για την ταινία αυτή, από την οποία παίρνουμε τον τίτλο κάνοντας JOIN με τη σχέση “movies”.

```
SELECT title
FROM (SELECT mid, COUNT(actors.aid) as total_actors
    FROM roles
    INNER JOIN actors ON actors.aid = roles.aid
    GROUP BY mid

    INTERSECT

    SELECT movies.mid, COUNT(actors.aid) as male_actors
    FROM roles
    INNER JOIN actors ON actors.aid = roles.aid
    INNER JOIN movies ON movies.mid = roles.mid
    WHERE gender='M'
    GROUP BY movies.mid) Q
INNER JOIN movies ON movies.mid = Q.mid
```

b. Με αντίστοιχη λογική, για το δεύτερο ερώτημα, εκτελούμε ξανά δύο ξεχωριστά SELECT:
   - Στο πρώτο επιλέγουμε όλες τις ταινίες (mid) στις οποίες παίζουν ηθοποιοί (σύνδεση με τη σχέση “roles”).
   - Στο δεύτερο επιλέγουμε τις ταινίες που παίζει έστω μια γυναίκα.

Αφού πάρουμε τα δύο αποτελέσματα, αφαιρούμε την δεύτερη σχέση που επιστρέφεται από την πρώτη με τον τελεστή EXCEPT. Με τον τρόπο αυτό παίρνουμε όλες τις ταινίες στις οποίες δεν παίζει καμία γυναίκα, δηλαδή τις ταινίες στις οποίες όλοι οι ηθοποιοί είναι άντρες.

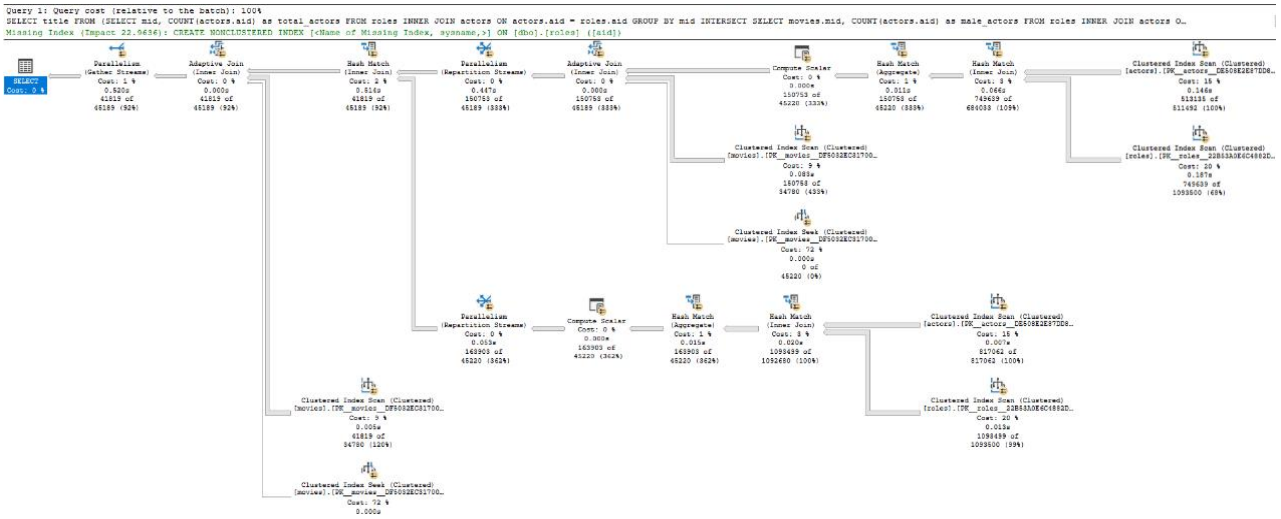
```
SELECT title
FROM movies
WHERE mid IN
    (SELECT roles.mid
    FROM movies
    INNER JOIN roles ON movies.mid = roles.mid

    EXCEPT

    SELECT mid
    FROM roles
    INNER JOIN actors ON roles.aid= actors.aid
    WHERE gender='F'
    GROUP BY mid)
```

Τα δύο παραπάνω επερωτήματα επιστρέφουν 41819 αποτελέσματα.
Αφού τρέξουμε αυτά τα ερωτήματα, δημιουργούμε τρία νέα ευρετήρια για το καθένα. Τα δύο από αυτά είναι κοινά και για τα δύο και ορίζονται πάνω στους πίνακες movie και roles για να βοηθήσουν στην ανάκτηση των πεδίων και τις πράξεις JOIN αντίστοιχα.
Το τρίτο ευρετήριο κάθε ερωτήματος ορίζεται πάνω στο πεδίο gender και aid της σχέσης actors και χρησιμοποιείται για να φιλτράρει τις εγγραφές που αντιστοιχούν στους άνδρες ηθοποιούς (για το α' ερώτημα) και τις εγγραφές που αντιστοιχούν στις γυναίκες ηθοποιούς (για το β' ερώτημα) για τις προτάσεις WHERE, ενώ το aid ευρετηριοποιείται για να βοηθήσει στα JOIN.
Οι εντολές δημιουργίας των ευρετηρίων είναι οι εξής (με τις δύο πρώτες να αφορούν τα κοινά ευρετήρια και τα επόμενα δύο να αφορούν το πρώτο και το δεύτερο ερώτημα αντίστοιχα):

```
CREATE INDEX roles_index ON roles(aid)

CREATE INDEX movies_index ON movies (mid) INCLUDE (title)

CREATE INDEX actors_index ON actors (gender, aid)
WHERE gender='M'

CREATE INDEX actors_index ON actors (gender, aid)
WHERE gender='F'
```

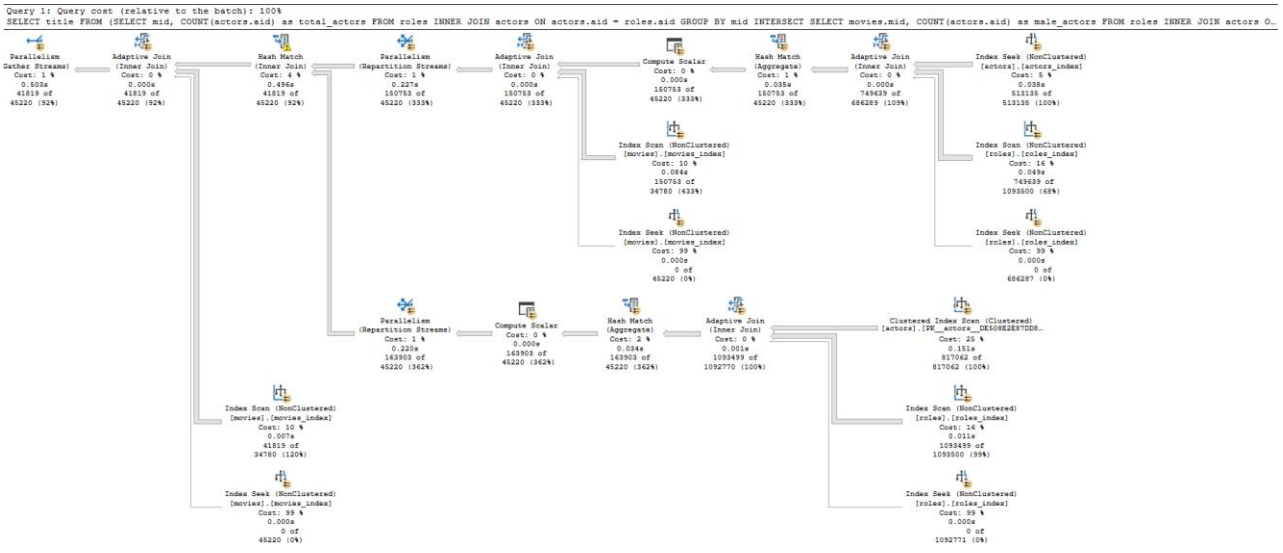
Παρακάτω βλέπου τα πλάνα εκτέλεσης και το πλήθος των I/O για κάθε ερώτημα, πριν και μετά την δημιουργία ευρετηρίων.

Ερώτημα α χωρίς ευρετήριο:



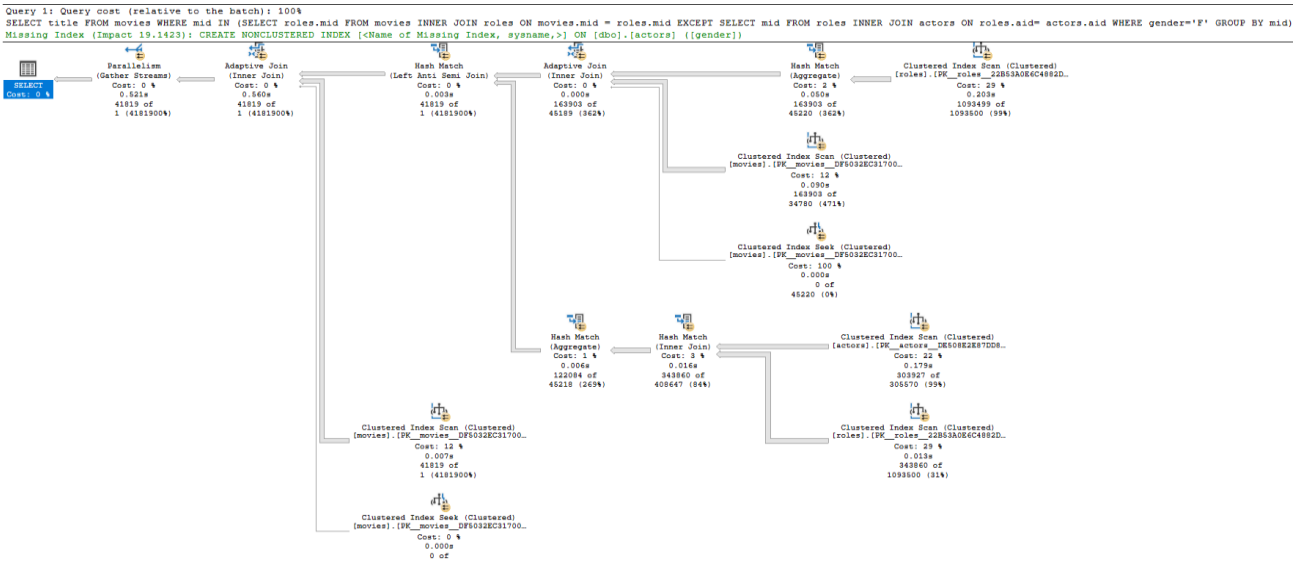

```
DBCC execution completed. If DBCC printed error messages, contact your system administrator.

(41819 rows affected)
Table 'movies'. Scan count 14, logical reads 4024, physical reads 2, page server reads 0, read-ahead reads 1917, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead re
Table 'roles'. Scan count 14, logical reads 8978, physical reads 1, page server reads 0, read-ahead reads 4424, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead rea
Table 'actors'. Scan count 14, logical reads 7060, physical reads 2, page server reads 0, read-ahead reads 3362, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead re
Table 'Workfile'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads 0
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads

(1 row affected)

Completion time: 2021-05-18T14:40:19.5668224+03:00
```

Ερώτημα α με ευρετήριο:

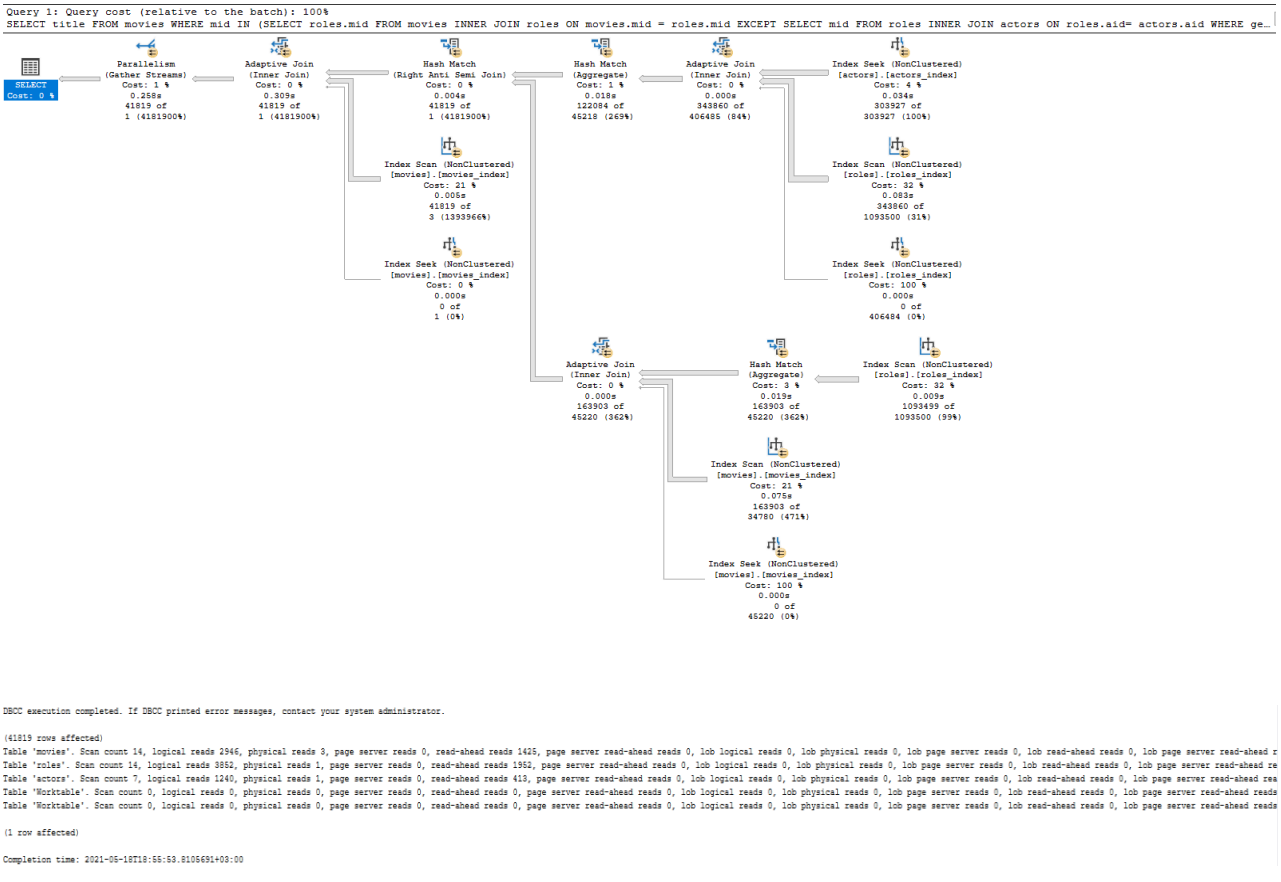



```
DBCC execution completed. If DBCC printed error messages, contact your system administrator.

(41819 rows affected)
Table 'movies'. Scan count 14, logical reads 2946, physical reads 3, page server reads 0, read-ahead reads 1418, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead r
Table 'roles'. Scan count 14, logical reads 3852, physical reads 1, page server reads 0, read-ahead reads 1944, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead re
Table 'actors'. Scan count 14, logical reads 4265, physical reads 3, page server reads 0, read-ahead reads 4060, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead r
Table 'Workfile'. Scan count 52, logical reads 624, physical reads 0, page server reads 0, read-ahead reads 624, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead r
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads

(1 row affected)

Completion time: 2021-05-18T18:03:00.8334506+03:00
```

Ερώτημα β χωρίς ευρετήριο:



Ερώτημα β με ευρετήριο:



Και στις δύο περιπτώσεις παρατηρούμε πως υπάρχει σημαντική βελτίωση στον χρόνο αναζήτησης και το πλήθος σελίδων που διαβάζονται με την χρήση ευρετηρίων. Ωστόσο από τα στατιστικά και τα πλάνα εκτέλεσης μπορούμε να συμπεράνουμε πως το δεύτερο επερώτημα είναι πιο αποδοτικό καθώς όχι μόνο είναι ταχύτερο, αλλά διαβάζει και σημαντικά λιγότερες σελίδες, επομένως επιλέγουμε αυτό.

Ζήτημα 3)

1.
    a. «Εμφάνισε για κάθε σκηνοθέτη την ταινία με τις υψηλότερες κριτικές που έχει σκηνοθετήσει»

       Για το συγκεκριμένο ερώτημα υπολογίζουμε αρχικά τον μέσο όρο των βαθμολογιών κάθε ταινίας. Στη συνέχεια κάνουμε GROUP BY ώστε να υπολογίσουμε μετά για κάθε σκηνοθέτη την μέγιστη από τις μέσες τιμές των ταινιών που έχει σκηνοθετήσει, τις οποίες βρίσκουμε μέσω JOIN.

```
SELECT did, title as best_movie, MAX(ratings) rating
FROM (SELECT mid, ROUND(AVG(CAST(rating AS FLOAT)), 2) as ratings
        FROM user_movies
        GROUP BY mid) avg_ratings
INNER JOIN movie_directors ON movie_directors.mid = avg_ratings.mid
INNER JOIN movies ON movies.mid = movie_directors.mid
GROUP BY did, title
ORDER BY did
```

    b. «Εμφάνισε τις 100 καλύτερες κωμωδίες που έχουν περισσότερες από 150 κριτικές»
       Για την εκτέλεση του ερωτήματος επιλέγουμε αρχικά όλες τις κωμωδίες από τη σχέση movies_genre ώστε να μειώσουμε το πλήθος τον εγγραφών που θα χρειαστεί να γίνουν JOIN με τον πίνακα user_movies. Στη συνέχεια υπολογίζουμε το μέσο όρο βαθμολογιών κάθε ταινίας και φιλτράρουμε τα αποτελέσματα ώστε να μείνουν μόνο οι εγγραφές που έχουν περισσότερες από 150 κριτικές.

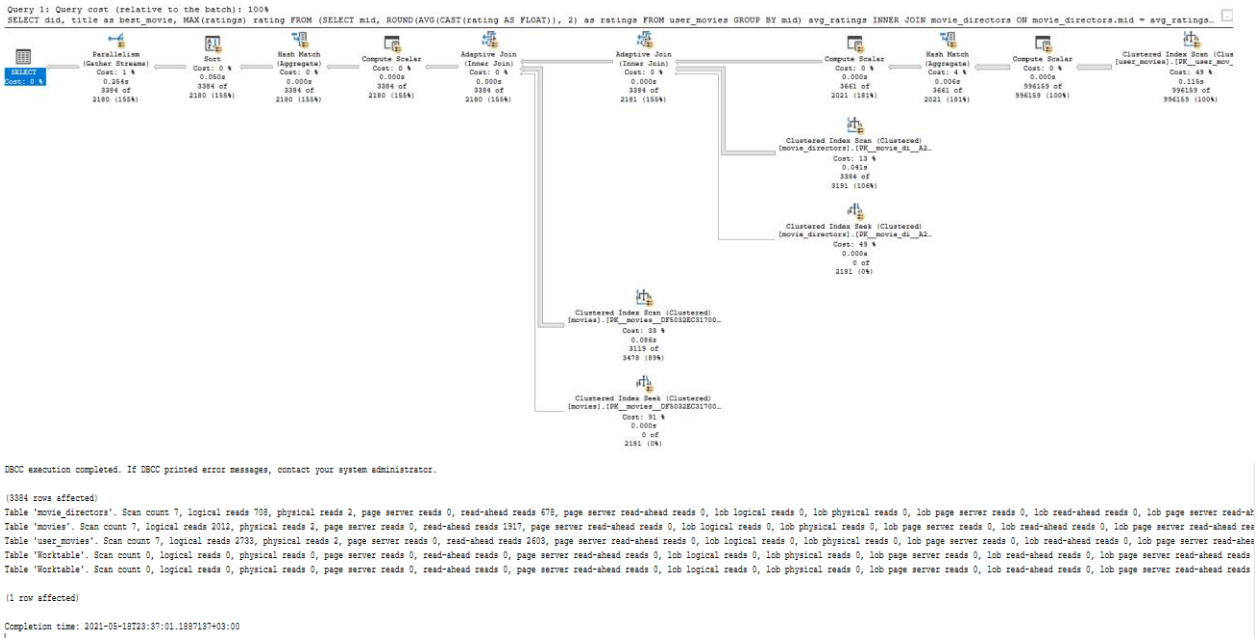
```
SELECT TOP 100 comedies.mid, ROUND(AVG(CAST(rating AS FLOAT)), 2) as rating
FROM (SELECT mid
        FROM movies_genre
        WHERE genre LIKE 'Comedy') comedies
INNER JOIN user_movies ON comedies.mid = user_movies.mid
GROUP BY comedies.mid
HAVING COUNT(rating) > 150
ORDER BY rating DESC
```

2.

a. Για το ερώτημα δημιουργούμε ένα ευρετήριο πάνω στο πεδίο mid της σχέσης movies, κάνοντας include το πεδίο title. Επίσης χρειαζόμαστε ένα ευρετήριο στο πεδίο mid της σχέσης movie_directors ώστε να εκτελείται πιο γρήγορα η αναζήτηση της συγκεκριμένης τιμής και να επιταχυνθεί το JOIN, μιας και στο clustered ευρετήριο αποτελεί το δεύτερο πεδίο ευρετηρίασης. Τα αποτελέσματα πριν και μετά τη χρήση των ευρετηρίων φαίνονται παρακάτω:

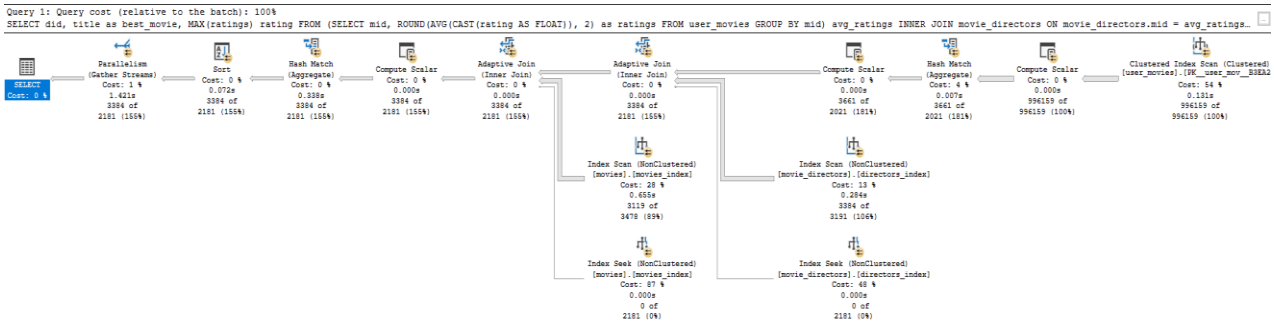
```
CREATE INDEX movies_index ON movies(mid) INCLUDE (title)

CREATE INDEX directors_index ON movie_directors(mid)
```

Χωρίς ευρετήριο:




```
DBCC execution completed. If DBCC printed error messages, contact your system administrator.

(3384 rows affected)
Table 'movie_directors'. Scan count 7, logical reads 708, physical reads 2, page server reads 0, read-ahead reads 678, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ah
Table 'movies'. Scan count 7, logical reads 2012, physical reads 2, page server reads 0, read-ahead reads 1917, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead rea
Table 'user_movies'. Scan count 7, logical reads 2733, physical reads 2, page server reads 0, read-ahead reads 2603, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahea
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads

(1 row affected)

Completion time: 2021-05-18T23:37:01.1887137+03:00
```

Με ευρετήριο:



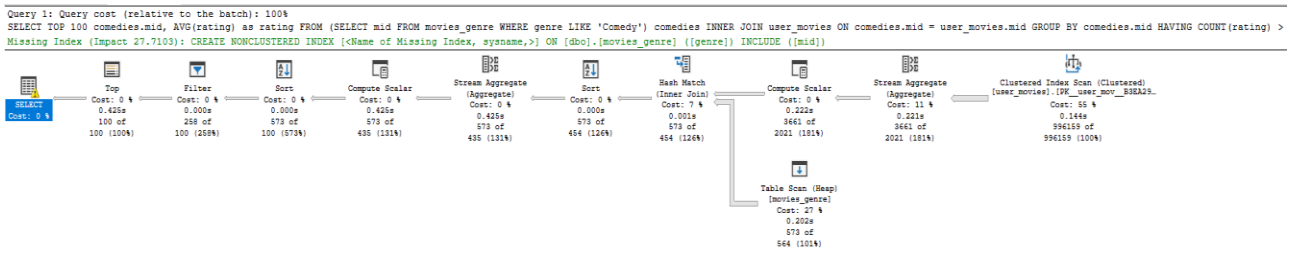

```
DBCC execution completed. If DBCC printed error messages, contact your system administrator.

(3384 rows affected)
Table 'movies'. Scan count 7, logical reads 1473, physical reads 3, page server reads 0, read-ahead reads 1432, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead rea
Table 'movie_directors'. Scan count 7, logical reads 586, physical reads 3, page server reads 0, read-ahead reads 555, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ah
Table 'user_movies'. Scan count 7, logical reads 2733, physical reads 2, page server reads 0, read-ahead reads 2603, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahea
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads

(1 row affected)

Completion time: 2021-05-18T23:49:21.3542987+03:00
```

b. Στο δεύτερο ερώτημα χρειαζόμαστε αρχικά ένα ερώτημα πάνω στο genre σχέσης movies_genre ώστε να επιταχυνθεί η αναζήτηση των κωμωδιών. Προσθέτοντας το πεδίο mid ως included field μπορούμε να εξαλείψουμε εντελώς την αναζήτηση στη σχέση και οι εγγραφές ανακτώνται από το ευρετήριο.

```
CREATE INDEX genre_index ON movies_genre (genre) INCLUDE (mid)
WHERE genre = 'Comedy'
```

Όπως φαίνεται και από τις παρακάτω εικόνες το index βελτιώνει το χρόνο που χρειάζεται για την ανάκτηση των πλειάδων και μειώνει τον αριθμό των I/Os

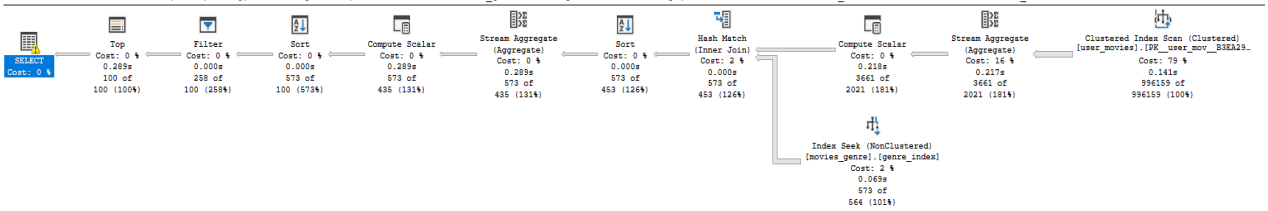
Χωρίς ευρετήριο:




```
DBCC execution completed. If DBCC printed error messages, contact your system administrator.

(100 rows affected)
Table 'movies_genre'. Scan count 1, logical reads 1412, physical reads 0, page server reads 0, read-ahead reads 1412, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahe
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads
Table 'user_movies'. Scan count 1, logical reads 2601, physical reads 2, page server reads 0, read-ahead reads 2603, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahea

(1 row affected)

Completion time: 2021-05-18T22:59:04.5826227+03:00
```

Με ευρετήριο:

```
Query 1: Query cost (relative to the batch): 100%
SELECT TOP 100 comedies.mid, AVG(rating) as rating FROM (SELECT mid FROM movies_genre WHERE genre LIKE 'Comedy') comedies INNER JOIN user_movies ON comedies.mid = user_movies.mid GROUP BY comedies.mid HAVING
```




```
DBCC execution completed. If DBCC printed error messages, contact your system administrator.

(100 rows affected)
Table 'movies_genre'. Scan count 1, logical reads 198, physical reads 1, page server reads 0, read-ahead reads 196, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, page server reads 0, read-ahead reads 0, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahead reads
Table 'user_movies'. Scan count 1, logical reads 2601, physical reads 2, page server reads 0, read-ahead reads 2603, page server read-ahead reads 0, lob logical reads 0, lob physical reads 0, lob page server reads 0, lob read-ahead reads 0, lob page server read-ahea

(1 row affected)

Completion time: 2021-05-18T23:00:03.0325892+03:00
```