



**ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ**

Συστήματα Διαχείρισης Βάσεων Δεδομένων

Project

Δημήτρης Πλεξουσάκης

Τμήμα Επιστήμης Υπολογιστών

HY460 – Συστήματα Διαχείρισης Βάσεων Δεδομένων Χειμερινό Εξάμηνο 2014 Project

Δημιουργήστε τους πίνακες (tables.sql) και εισάγετε τα δεδομένα (distributedBy.sql – 143.153 πλειάδες, movie.sql – 193.781 πλειάδες, people.sql - 289.221 πλειάδες, plays.sql – 810.692 πλειάδες, producedBy.sql – 202.868 πλειάδες). Αρχικά, δεν υπάρχουν ορισμένα ευρετήρια (index), κλειδιά (keys) ή συστάδες (clusters) για τους πίνακες της άσκησης. Επίσης, δημιουργήστε στην περιοχή σας τον πίνακα PLAN_TABLE, όπου θα αποθηκευτούν τα σχέδια εκτέλεσης (execution plans) κάθε επερώτησης της άσκησης. Κατά τη διάρκεια της άσκησης, κάθε φορά που είτε δημιουργείτε έναν πίνακα και φορτώνετε σε αυτόν πλειάδες, είτε ορίζετε, είτε όχι, ένα καινούργιο ευρετήριο, θα πρέπει πάντοτε να ενημερώνετε τα στατιστικά στοιχεία του αντίστοιχου πίνακα.

Σημείωση: Η ενδιάμεση μνήμη (cache) που χρησιμοποιείται από την Oracle είναι 41.943.040 bytes, μοιρασμένη σε 5.120 buffers μεγέθους 8.192 bytes ο καθένας.

Ζήτημα 1^ο – Χρήση της ενδιάμεσης μνήμης για ελαχιστοποίηση των E/E

A) Επειδή ο μεγάλος αριθμός αιτήσεων E/E επιβραδύνουν τον χρόνο απόκρισης και αυξάνουν τις απαιτήσεις σε CPU, η απόδοση του εξυπηρετητή (server) της Oracle μπορεί να βελτιωθεί σημαντικά, όταν οι περισσότερες σελίδες από αυτές που αιτούνται βρίσκονται ήδη στην ενδιάμεση μνήμη. Το μέτρο αυτής της απόδοσης είναι το ποσοστό επιτυχίας (hit ratio) της ενδιάμεσης μνήμης (cache), και αντιστοιχεί στο λόγο του αριθμού των σελίδων που βρίσκονται στη μνήμη προς το συνολικό αριθμό των σελίδων, στις οποίες έχουν προσπελαστεί. Όταν η ενδιάμεση μνήμη του ΣΔΒΔ είναι πολύ μικρή, τότε η απόδοση του συστήματος επιβαρύνεται, εξαιτίας των πολλών αιτήσεων E/E. Θεωρούμε πως ένα σύστημα χρίζει ρύθμισης (tuning) όταν ο παραπάνω λόγος έχει τιμή μικρότερη του 80%.

Μερικές από τις προτεινόμενες λύσεις για ρύθμιση είναι και οι παρακάτω:

- Προσθήκη επιπλέον σελίδων ενδιάμεσης μνήμης.
- Χρήση πολλαπλών δεξαμενών σελίδων (buffer pools) ενδιάμεσης μνήμης, για το διαχωρισμό των σελίδων ανάλογα με το πρότυπο πρόσβασης σε αυτές.
- Μεταφορά ενός ή περισσότερων πινάκων στην ενδιάμεση μνήμη (αν φυσικά υπάρχει διαθέσιμος χώρος).

A) Για τις παρακάτω επερωτήσεις υπολογίστε τον αριθμό των αναγνώσεων (physical reads) από το δίσκο.

1^η επερώτηση:

```
SELECT    year, count(*)
FROM      movie
GROUP BY  year
ORDER BY  year desc;
```

2^η επερώτηση:

```
SELECT    companyID, count(*)
FROM      distributedBy
GROUP BY  companyID;
```

Β) Υπολογίστε το ποσοστό επιτυχίας (hit ratio) της ενδιάμεσης μνήμης (cache) για την εκτέλεση των παραπάνω ερωτήσεων. Χρειάζεται το σύστημα ρύθμιση (tuning); Δικαιολογήστε την απάντησή σας.

Γ) Υπολογίστε το μέγεθος των πινάκων που υπάρχουν στη βάση. Ποιος από αυτούς θα μπορούσε να φορτωθεί στην ενδιάμεση μνήμη για τους λόγους που αναπτύχθηκαν παραπάνω;

Δ) Τοποθετήστε τον πίνακα movie στην ενδιάμεση μνήμη. Εκτελέστε τις παραπάνω επερωτήσεις, με τη σειρά που δίνονται, υπολογίζοντας μετά από κάθε μία τον αριθμό των σελίδων που διαβάστηκαν από το δίσκο. Τί παρατηρείτε; Εκτελέστε ξανά την πρώτη επερώτηση και υπολογίστε στη συνέχεια τον αριθμό των σελίδων που διαβάστηκαν από το δίσκο. Τί παρατηρείτε τώρα; Καταργήστε (drop) τον πίνακα movie και ξαναδημιουργήστε τον, φορτώνοντάς του τα δεδομένα από την αρχή.

Ζήτημα 2^ο – Εναλλακτικές μορφές επερωτήσεων SQL

Αρχικά, ο optimizer_mode έχει την τιμή choose, δηλαδή ο βελτιστοποιητής (optimizer) της Oracle επιλέγει, ανάλογα με την ύπαρξη, ή μη, στατιστικών, μεταξύ μιας προσέγγισης βελτιστοποίησης (optimization) επερωτήσεων, βασισμένης σε μοντέλο κόστους, ή βασισμένης σε ευρηστικούς κανόνες.

Α) Ψάχνουμε τα IDs των ταινιών για τις οποίες δεν έχει συμπληρωθεί πληροφορία σχετικά με την εταιρεία διανομής των. Έχουμε, τις ακόλουθες δύο, ισοδύναμες ως προς το αποτέλεσμα, επερωτήσεις SQL:

1^η επερώτηση:

```
SELECT movieID
FROM movie MINUS
SELECT movieID
FROM distributedBy;
```

2^η επερώτηση:

```
SELECT movieID FROM movie
WHERE movieID NOT IN

(SELECT movieID FROM
distributedBy);
```

ι) Χρησιμοποιήστε την εντολή EXPLAIN PLAN της Oracle για να δείτε τα σχέδια εκτέλεσης (execution plans) των δύο επερωτήσεων που επιλέγει ως βέλτιστα ο εξυπηρετητής (server). Σχεδιάστε σε μια δεντρική μορφή τα δύο αυτά σχέδια και, στη συνέχεια, περιγράψτε τις λειτουργίες που περιλαμβάνει το καθένα.

ii) Πριν εκτελέσετε τις επερωτήσεις, και με βάση τα σχέδια αυτά, ταξινομήστε τις επερωτήσεις ανάλογα με το χρόνο που πιστεύετε ότι θα διαρκέσει η εκτέλεσή τους. Δικαιολογήστε την απάντησή σας. Στη συνέχεια, βρείτε τον χρόνο εκτέλεσης των επερωτήσεων και συγκρίνετε τα αποτελέσματα με την πρόβλεψη που έχετε κάνει.

iii) Προτείνετε τρόπους για να επιταχυνθούν οι παραπάνω επερωτήσεις, π.χ. με τη χρήση ευρετηρίων (indexes) και συστάδων (clusters). Εξετάστε αν όντως επιτυγχάνουν να επιταχύνουν την εκτέλεσή τους.

B) Ψάχνουμε τα IDs των ταινιών στις οποίες συμμετέχουν ηθοποιοί γεννημένοι μετά τη χρονιά χ , όπου $\chi = 1930, 1950, 1970, 1990, 1995$. Έχουμε τις ακόλουθες δύο, ισοδύναμες ως προς το αποτέλεσμα, επερωτήσεις SQL:

1^η επερώτηση:

```
SELECT pl.movieID
FROM people p, plays pl
WHERE p.personID = pl.personID and p.birthYear >  $\chi$ ;
```

2^η επερώτηση:

```
SELECT pl.movieID
FROM plays pl
WHERE pl.personID IN
  (SELECT p.personID
   FROM people p
   WHERE p.birthYear >  $\chi$ );
```

i) Βρείτε και δείξτε τα σχέδια εκτέλεσης των παραπάνω επερωτήσεων -δεν εξαρτώνται από την τιμή του χ - και στη συνέχεια συγκρίνετε τους χρόνους εκτέλεσης των επερωτήσεων για τις διάφορες τιμές του χ . Ποια από τις δύο εκτελείται πιο γρήγορα;

ii) Πώς μπορούμε να διατυπώσουμε τις παραπάνω επερωτήσεις ώστε να αξιοποιήσουμε καλύτερα τη διαθέσιμη μνήμη και να βελτιώσουμε την απόδοση εκτέλεσής τους;

Ζήτημα 3^ο – Βελτιστοποίηση βάσει κόστους

Χρησιμοποιήστε την εντολή ALTER SESSION για να θέσετε την τιμή της παραμέτρου OPTIMIZER_MODE σε ALL_ROWS. Με αυτή την επιλογή ο βελτιστοποιητής (optimizer) της Oracle χρησιμοποιεί μία προσέγγιση βελτιστοποίησης (optimization) επερωτήσεων, βασισμένη σε μοντέλο κόστους και προσπαθεί να επιτύχει την καλύτερη δυνατή χρήση των πόρων του συστήματος στην εκτέλεσή τους.

A) Θεωρήστε την παρακάτω επερώτηση πάνω στους πίνακες movie και plays, η οποία περιέχει το γνώρισμα movieID στις συνθήκες του WHERE. Ποιο είναι το σχέδιο επεξεργασίας της; Ποιος ο χρόνος εκτέλεσής της; Αλλάζει τίποτα αν στο SELECT προσθέσουμε την δήλωση distinct για απαλοιφή των διπλοεγγραφών;

```
SELECT      p.personID
FROM        movie m, plays p
WHERE       p.movieID = m.movieID and m.movieID = 0046790;
```

```

SELECT    DISTINCT p.personID
FROM      movie m, plays p
WHERE     p.movieID = m.movieID and m.movieID = 0046790;

```

B) Δημιουργήστε μια συστάδα κατακερματισμού (hash cluster) για το γνώρισμα movieID με 1000 κλειδιά αναζήτησης και σχετίστε τον πίνακα movie με αυτό. Θα πρέπει να καταργήσετε (drop) και να ξαναδημιουργήσετε τον πίνακα movie, για να επωφεληθείτε της συστάδας κατακερματισμού. Ποιο είναι τώρα το σχέδιο εκτέλεσης της προηγούμενης επερωτήσης και ποιος ο χρόνος εκτέλεσης;

Γ) Καταργήστε (drop) τον πίνακα plays και ξαναδημιουργήστε τον, σχετίζοντας το γνώρισμα movieID με το cluster που έχετε ορίσει από το προηγούμενο ερώτημα. Ποιο είναι το σχέδιο εκτέλεσης της ίδιας ερώτησης και ποιος ο χρόνος εκτέλεσής της;

Δ) Βρείτε το σχέδιο εκτέλεσης των παρακάτω επερωτήσεων. Παρατηρείτε κάτι; Στη συνέχεια, δημιουργήστε ένα μη συσταδοποιημένο (unclustered) B⁺-Tree ευρετήριο για το γνώρισμα year του πίνακα movie και ένα για το γνώρισμα birthYear του πίνακα people. Ποιο είναι τώρα το πλάνο εκτέλεσης των παρακάτω επερωτήσεων SQL;

```

SELECT    movieTitle
FROM      movie
WHERE     year > 1987;

```

```

SELECT    personName
FROM      people
WHERE     birthYear > 1940;

```

Ε) Βρείτε το σχέδιο εκτέλεσης των ακόλουθων επερωτήσεων SQL στον πίνακα movie που περιέχουν τα γνωρίσματα movieID και year στις συνθήκες επιλογής. Χρησιμοποιείται η συστάδα κατακερματισμού (hash cluster) στο movieID ή το B⁺-Tree ευρετήριο στο γνώρισμα year; Σε Τι οφείλεται τυχόν ύπαρξη διαφορετικών σχεδίων εκτέλεσης;

```

SELECT    movieID, movieTitle
FROM      movie
WHERE     movieID != 0046790 and year > 1983;

```

```

SELECT    movieID, movieTitle
FROM      movie
WHERE     movieID = 0046790 and year > 1983;

```

Ζήτημα 4^ο – Βελτιστοποίηση βάσει κανόνων

Σε αυτό το σημείο της άσκησης πρέπει να υπάρχει ορισμένη μια συστάδα κατακερματισμού (hash cluster) για το γνώρισμα movieID, που χρησιμοποιείται από τους πίνακες movie και plays. Επίσης, υπάρχουν δύο μη συσταδοποιημένα (unclustered) B⁺-Tree ευρετήρια για τα γνωρίσματα year και birthYear των πινάκων movie και people, αντίστοιχα.

Χρησιμοποιήστε την εντολή ALTER SESSION για να θέσετε την τιμή της παραμέτρου OPTIMIZER_MODE σε RULE. Με αυτή την επιλογή ο βελτιστοποιητής (optimizer) της Oracle χρησιμοποιεί μία προσέγγιση βελτιστοποίησης (optimization) επερωτήσεων, βασισμένη

αποκλειστικά σε ευρηστικούς κανόνες, αδιαφορώντας για την ύπαρξη στατιστικών.

A) Εκτελέστε εκ νέου το ερώτημα (Δ) του 3^{ου} ζητήματος. Υπάρχει κάποια διαφορά στα πλάνα εκτέλεσης. Αν ναι, συγκρίνετε τους χρόνους εκτέλεσης.

B) Ορίστε για το γνώρισμα movieTitle του πίνακα ένα μοναδικό μη-συσταδοποιημένο B⁺-Tree ευρετήριο. Βρείτε το σχέδιο εκτέλεσης της παρακάτω SQL επερώτησης. Χρησιμοποιείται η συστάδα κατακερματισμού (hash cluster) ή το ευρετήριο που μόλις ορίσατε και γιατί;

```
SELECT m.movieID, m.movieTitle, p.companyID
FROM movie m, producedBy p
WHERE m.movieID = p.movieID and m.movieID = 0046790 and m.movieTitle =
'Black Tuesday (1954)';
```

Καταργήστε (drop) το ευρετήριο στο movieTitle.

Ζήτημα 5^ο– Σύγκριση μεθόδων βελτιστοποίησης

Σε αυτό το σημείο της άσκησης πρέπει να υπάρχει ορισμένη μια συστάδα κατακερματισμού (hash cluster) στο γνώρισμα movieID που χρησιμοποιείται από τους πίνακες movie και plays. Επίσης, υπάρχουν δύο μη συσταδοποιημένα B⁺-Tree ευρετήρια για τα γνωρίσματα year και birthYear των πινάκων movie και people αντίστοιχα. Τέλος, ο optimizer_mode πρέπει να είναι RULE.

A) Για κάθε έναν από τους τρεις τύπους OPTIMIZATION_MODE (RULE, FIRST_ROWS, ALL_ROWS) βρείτε το πλάνο και το χρόνο εκτέλεσης της παρακάτω επερώτησης SQL:

```
SELECT /*+ ORDERED USE_HASH(p,pl) */ pl.movieID
FROM people p, plays pl
WHERE p.personID = pl.personID and p.birthYear > 1990;
```

B) Καταργήστε (drop) το ευρετήριο για το γνώρισμα birthYear και επαναλάβετε το ερώτημα (A). Για την ίδια παράμετρο βελτιστοποίησης (mode) παρατηρείτε κάποια διαφοροποίηση ανάμεσα στις δύο επερωτήσεις;

Γ) Δημιουργήστε ένα bitmap ευρετήριο για το γνώρισμα birthYear. Επαναλάβετε και πάλι το ερώτημα (A). Τί διαφορές παρατηρείτε σχετικά με τη χρήση bitmap ευρετηρίων;

Δ) Συγκρίνετε τους χρόνους εκτέλεσης των επερωτήσεων (A), (B) και (Γ) για κάθε παράμετρο (mode) βελτιστοποίησης. Τί διαφορές παρατηρείτε;

E) Καταργήστε (drop) τον πίνακα people και ξαναδημιουργήστε τον, ορίζοντας ένα συσταδοποιημένο (clustered) B⁺-Tree ευρετήριο στο γνώρισμα birthYear. Για κάθε έναν από τους τρεις τύπους OPTIMIZATION_MODE (RULE, FIRST_ROWS, ALL_ROWS) βρείτε το πλάνο και το χρόνο εκτέλεσης της επερώτησης του ερωτήματος (A). Παρατηρείτε κάποια διαφορά σε σχέση με τη προηγούμενη χρήση μη- συσταδοποιημένου (unclustered) B⁺-Tree ευρετηρίου;

Σημειώματα

Σημείωμα αναφοράς

Copyright Πανεπιστήμιο Κρήτης, Δημήτρης Πλεξουσάκης. «Συστήματα Διαχείρισης Βάσεων Δεδομένων. Project». Έκδοση: 1.0. Ηράκλειο 2014. Διαθέσιμο από τη δικτυακή διεύθυνση: <http://www.csd.uoc.gr/~hy460>.

Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση, Όχι Παράγωγο Έργο 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



[1] <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Ως **Μη Εμπορική** ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Κρήτης» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

