# Αναφορά Πρότζεκτ

Για τον καλύτερο προσδιορισμό των αποτελεσμάτων έγινε εισαγωγή εκατό χιλιάδων διαφορετικών πλειάδων σε κάθε πίνακα (Student, Professor, Labstaff).

## Ερώτημα 2:

Σε όλα τα παρακάτω ερωτήματα όπως και ζητείται γίνεται ταξινόμηση με βάση το επώνυμο και το όνομα.

## 2.1)

Στο ερώτημα αυτό γίνεται ανάκτηση των στοιχείων ενός φοιτητή με δοσμένο αριθμό μητρώου. Παρακάτω βρίσκεται το αντίστοιχο query:

**SELECT** 

amka,

NAME,

surname,

father\_name,

email

FROM "Student"

WHERE am = '2009090013'

ORDER BY surname, name:

Η επιλογή του Index με βάση το attribute που βρίσκεται στο selection στην περίπτωσή μας ο αριθμός μητρώου. Από τη θεωρία γνωρίζουμε ότι το κατάλληλο index για ένα τέτοιο point query είναι το hash παρόλαυτά, η επιλογή του έγινε σκόπιμα btree και όχι hash Καθώς η ίδια η postgres δεν υπάρχει clustering σε hash indexes. Στις παρακάτω εικόνες παρουσιάζονται τα αποτελέσματα της explain-analyse με τους αντίστοιχους χρόνους εκτελέσης με ή χωρίς index.

#### Χωρίς Index.

#### Με BTree Index πάνω στο attribute am.

```
QUERY PLAN

Sort (cost=8.45..8.45 rows=1 width=86) (actual time=0.054..0.054 rows=1 loops=1)

Sort Key: surname, name

Sort Method: quicksort Memory: 25kB

-> Index Scan using student_am_index1 on "Student" (cost=0.42..8.44 rows=1 width=86) (actual time=0.043..0.044 rows=1 loops=1)

Index Cond: (am = '2007056816'::bpchar)

Planning time: 0.227 ms

Execution time: 0.083 ms
```

Η δραματική πτώση στο execution time οφείλεται στην κατάλληλη δεικτοδότηση των απαιτούμενων tuples για το query μέσω του Index στο δίσκο. Για την περαιτέρω πτώση του execution time χρησιμοποιήσαμε clustering με σκοπό την ταξινόμηση των δεδομένων του δίσκου με βάση το btree index που μόλις χρησιμοποιήσαμε

#### Clustering on btree index

```
QUERYPLAN

1 Sort (cost=8.45..8.45 rows=1 width=86) (actual time=0.041..0.041 rows=1 loops=1)

2 Sort Key: surname, name

3 Sort Method: quicksort Memory: 25kB

4 -> Index Scan using student_am_index1 on "Student" (cost=0.42..8.44 rows=1 width=86) (actual time=0.030..0.031 rows=1 loops=1)

5 Index Cond: (am = '2007056816'::bpchar)

6 Planning time: 0.228 ms

7 Execution time: 0.066 ms
```

## 2.2)

Στο query αυτό γίνεται ανάκτηση δεδομένων ενός φοιτητή που παρακολουθεί ένα συγκεκριμένο μάθημα στο τωρινό εξάμηνο. Παρακάτω βρίσκεται το query αυτό: SELECT name, surname, am

FROM courserun

NATURAL JOIN semester

NATURAL JOIN register

NATURAL JOIN "Student"

WHERE course code = 'ΠΛΗ 302' AND semester status =

'present' :: SEMESTER\_STATUS\_TYPE

ORDER BY surname, NAME;

Στο συγκεκριμένο ερώτημα λόγω των διαφόρων natural joins ήταν αναγκαία η δημιουργία μόνο ενός index και αυτό πάνω στο course\_code του πίνακα register. Η δημιουργία ενός Index πάνω στο semester status θα άστοχη λόγω του γεγονότος ότι οι δυνατές τιμές είναι μόνο τρεις πράγμα που ο optimizer αγνοεί.

#### Εκτέλεση χωρίς Index

### Εκτέλεση με Btree Index στο register.course\_code

```
| Sort (cost=925.21..925.26 rows=21 width=44) (actual time=0.603..0.609 rows=50 loops=1)
| Sort Key: "Student".surname, "Student".name | Student".surname, "Student".surname, "Student, "Student,
```

#### Εκτέλεση με Clustering

```
Sort (cost=445.74..445.79 rows=21 width=44) (actual time=0.317..0.320 rows=50 loops=1)
     -> Nested Loop (cost=28.92..445.28 rows=21 width=44) (actual time=0.100..0.260 rows=50 loops=1)
           -> Hash Join (cost=28.63..315.60 rows=21 width=4) (actual time=0.096..0.178 rows=50 loops=1)
                Hash Cond: (register.serial_number = courserun.serial_number)
                 -> Bitmap Heap Scan on register (cost=11.50..296.71 rows=414 width=19) (actual time=0.057..0.095 rows=425 loops=1)
Recheck Cond: (course_code = '\PiAH 302'::bpchar)
                       Heap Blocks: exact=5
                 -> Hash (cost=17.12..17.12 rows=1 width=15) (actual time=0.022..0.022 rows=1 loops=1)
                       Buckets: 1024 Batches: 1 Memory Usage: 9kB
                       -> Nested Loop (cost=4.36..17.12 rows=1 width=15) (actual time=0.018..0.021 rows=1 loops=1)
                              -> Seq Scan on semester (cost=0.00..1.26 rows=1 width=4) (actual time=0.007..0.008 rows=1 loops=1)
                                   Filter: (semester_status = 'present'::semester_status_type)
                              -> Bitmap Heap Scan on courserun (cost=4.36..15.73 rows=10 width=19) (actual time=0.008..0.009 rows=10 loops=1)
                                    Heap Blocks: exact=1
                                          Index Cond: (course_code = 'ΠΛΗ 302'::bpchar)
           -> Index Scan using student_pkey on "Student" (cost=0.29..6.17 rows=1 width=48) (actual time=0.001..0.001 rows=1 loops=50)
                Index Cond: (amka = register.amka)
27 Planning time: 0.332 ms
```

Μπορούμε να παρατηρήσουμε ότι κάνοντας χρήση clustering οι διαφορές είναι υποδιπλάσιας τάξης πράγμα που οφείλεται στην ταξινόμηση των tuples του register στον δίσκο. Μετά από πολλαπλές δοκιμές ενεργοποιώντας και απενεργοποιώντας μεθόδους join (hash join, merge join, nested loop join) παρατηρήσαμε ότι οι επιλογές του optimizer είναι οι βέλτιστες. Ακολουθούν οι παραδοχές που χρησιμοποιούνται οι μεθόδοι merge join και hash join.

#### Merge Join

#### **Hash Join**

## 2.3)

Το συγκεκριμένο ερώτημα κάνει την ανάκτηση του ονοματεπωνύμου και του χαρακτηρισμού ενός ατόμου της βάσης δεδομένων. Ακολουθεί το σχετικό query:

```
SELECT*
FROM (SELECT
    "Student".surname.
    "Student".name,
    'Student' AS status
  FROM "Student"
  UNION
  SELECT
    "Labstaff".surname,
    "Labstaff".name,
    'Labstaff' AS status
  FROM "Labstaff"
  UNION
  SELECT
    "Professor".surname,
   "Professor".name,
    'Professor' AS status
  FROM "Professor") x
ORDER BY surname, name
```

Θέτοντας btree indexes στα surname και name attributes παρατηρήσαμε ότι εν τέλει δεν γίνεται η χρήση τους, πράγμα που είναι λογικό διότι το συγκεκριμένο ερώτημα κάνει ανάκτηση όλων των δεδομένων των πινάκων Student, Professor και Labstaff. Έτσι σε οποιαδήποτε περίπτωση(είτε με χρήση των indexes είτε χωρίς) το execution time παραμένει σταθερό.

## Χωρίς Indexes

```
QUERYPLAN

Unique (cost=107055.05..110059.80 rows=300475 width=228) (actual time=1190.086..1903.317 rows=300475 loops=1)

-> Sort (cost=107055.05..107806.23 rows=300475 width=228) (actual time=1190.084..1845.873 rows=300475 loops=1)

Sort Key: "Student".surname, "Student".name, ('Student'::text)

Sort Method: external merge Disk: 15752kB

-> Append (cost=0.00..13986.50 rows=300475 width=228) (actual time=0.011..73.291 rows=300475 loops=1)

-> Seq Scan on "Student" (cost=0.00..2681.75 rows=100275 width=65) (actual time=0.010..20.327 rows=100275 loops=1)

-> Seq Scan on "Labstaff" (cost=0.00..4150.00 rows=100100 width=66) (actual time=0.008..16.835 rows=100100 loops=1)

Seq Scan on "Professor" (cost=0.00..4150.00 rows=100100 width=65) (actual time=0.886..17.208 rows=100100 loops=1)

Planning time: 0.129 ms

Execution time: 1917.216 ms
```

#### Mε Index

## Ερώτημα 3:

Στο συγκεκριμένο ερώτημα ζητείται η δημιουργία και η σύγκριση των χρόνων εκτέλεσης του ίδιου query σε virtual view και σε materialized view. Η ζητούμενη SQL ερώτηση είναι η εξής:

Εύρεση και παρουσίαση των στοιχείων των φοιτητών που το επίθετό τους περιέχει την συμβολοσειρά "αλ", τον αριθμό των μαθημάτων για τα οποία έχει εγκριθεί η εγγραφή τους και τον συνολικό αριθμό διδακτικών μονάδων. Παρακάτω παρουσιάζεται το σχετικό query: SELECT

```
s.amka,
s.surname,
s.father_name,
s.name,
s.entry date,
s.email,
s.am,
count(r.course_code),
CASE WHEN sum(r.units) IS NULL
 THEN 0
ELSE sum(r.units) END
FROM "Student" s
LEFT OUTER JOIN ((SELECT *
          FROM register
          WHERE register_status =
              'approved' :: REGISTER_STATUS_TYPE) r1
 NATURAL JOIN "Course") r ON S.amka = r.amka
WHERE surname LIKE '%ΑΛ%'
GROUP BY S.amka:
```

Παρακάτω παρουσιάζονται οι εκτελέσεις των δύο όψεων:

#### **Virtual View**

#### **Materialized View**

```
QUERY PLAN

1 Seq Scan on student_details_materialized_view (cost=0.00..312.41 rows=10741 width=117) (actual time=0.007..1.043 rows=10741 loops=1)

2 Planning time: 0.030 ms

3 Execution time: 1.434 ms
```

Από τα παραπάνω μπορούμε να παρατηρήσουμε ότι ο χρόνος εκτέλεσης των δύο όψεων είναι δραματικά καλύτερος στην περίπτωση του materialized view. Αυτό οφείλεται στο γεγονός ότι το materialized view έχει ήδη εκτελεσθεί και αποθηκευτεί στην βάση με την μορφή πίνακα με απότοκο κατά την εκτέλεση να γίνεται η απευθείας ανάκτηση των δεδομένων εν αντιθέσει με το virtual view το οποίο εκτελείται realtime.

## Ερώτημα 4:

Μετά από την εκτέλεση των δύο εισαγωγών στους πίνακες Student, Professor και Labstaff τόσο με indexes όσο και χωρίς οι χρόνοι εκτέλεσης που παρατηρήθηκαν έχουν ελάχιστες διαφορές πράγμα που οφείλεται στο γεγονός ότι τα χρησιμοποιούμενα indexes είναι σχετικά λίγα για να γίνουν εμφανείς οι διαφορές. Παρόλαυτα μπορούμε να παρατηρήσουμε μία ελαφριά αύξηση του χρόνου εκτέλεσης στην περίπτωση όπου γίνεται η χρήση των indexes. Παρακάτω παρατίθενται οι σχετικοί χρόνοι εκτέλεσης.

#### Με χρήση Indexes

```
QUERY PLAN

1 Result (cost=0.00..0.26 rows=1 width=4) (actual time=10538.044..10538.045 rows=1 loops=1)

2 Planning time: 0.015 ms

3 Execution time: 10538.053 ms
```

### Χωρίς Χρήση Indexes

```
QUERY PLAN

1 Result (cost=0.00..0.26 rows=1 width=4) (actual time=8863.569..8863.570 rows=1 loops=1)

2 Planning time: 0.027 ms

3 Execution time: 8863.586 ms
```