Το Εσωτερικό (Φυσικό) Επίπεδο του Σχεσιακού ΣΔΒΔ

... μια απλουστευμένη προσέγγιση του θέματος, με την οπτική του Τεχνολόγου Πληροφορικού

Μέρος Β!

Δημήτρης Α. Δέρβος
Τμήμα Πληροφορικής,
ΑΤΕΙ Θεσσαλονίκης

email: dad@it.teithe.gr
URL: http://aetos.it.teithe.gr/~dad

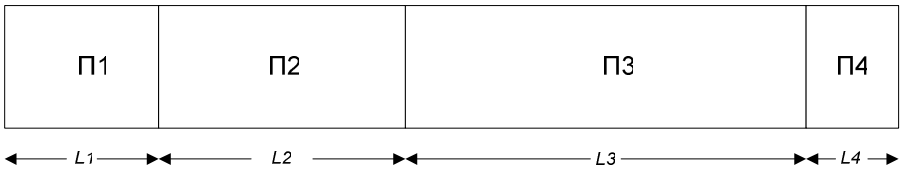
Μάϊος 2008

Τελευταία ενημέρωση: 01/05/2008

Από τους πίνακες στις εγγραφές δεδομένων

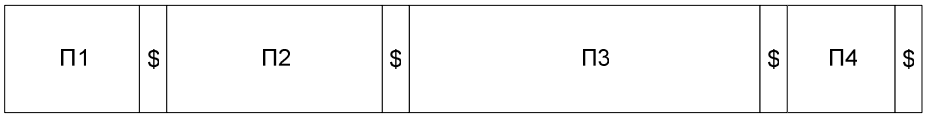
- Η οπτική της εγγραφής και στο ιδεατό επίπεδο
- Πλήθος στηλών → ο σταθερός αριθμός των επιμέρους πεδίων (λεξικό δεδομένων)
- Τύπος τιμών: επίσης παρών στο λεξικό δεδομένων
- Οπότε: εξοικονόμηση χώρου μνήμης (η πληροφορία δεν καταχωρείται στην κάθε μία εγγραφή, ξεχωριστά)

Τυπική εγγραφή δεδομένων (4 πεδία, σταθερού μήκους)

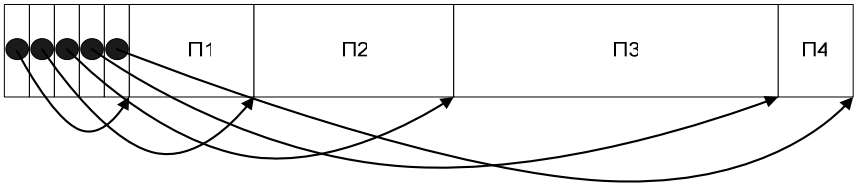


Όταν σταθερού μήκους τα πεδία: η πληροφορία στο λεξικό δεδομένων και η κίνηση στο εσωτερικό της εγγραφής με 'άλματα' γνωστού μήκους

Εγγραφή μεταβλητού μήκους με πεδία-διαχωριστές



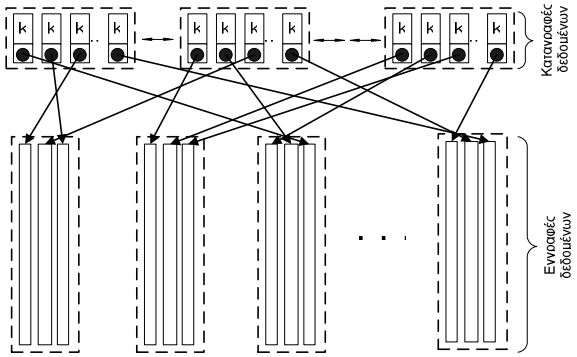
Εγγραφή μεταβλητού μήκους με εσωτερικό κατάλογο



Φυσικοί δείκτες προς την αρχή του κάθε ενός πεδίου και ένας προς το τέλος της εγγραφής. Εντοπισμός εσωτερικού πεδίου χωρίς να προηγείται σειριακή σάρωση των προηγουμένων του.

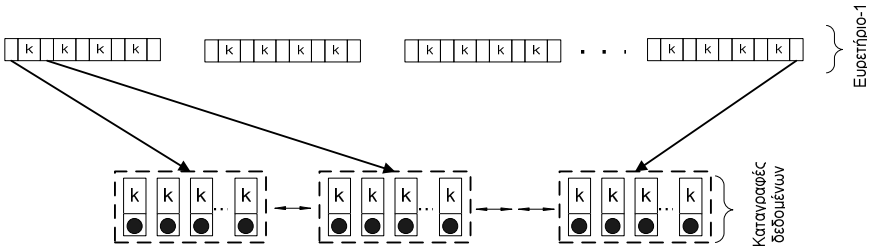
Ευρετήριο εναλλακτικής (2)

Τα δύο κατώτερα επίπεδα: καταγραφές & εγγραφές δεδομένων



Το 'αρχείο' με τις καταγραφές δεδομένων συνιστά πυκνό ευρετήριο. Το σχήμα προσφέρεται για ευρετηριοποίηση με συγκροτημένο, αραιό ευρετήριο ένα επίπεδο ψηλότερα: ευρετήριο-1

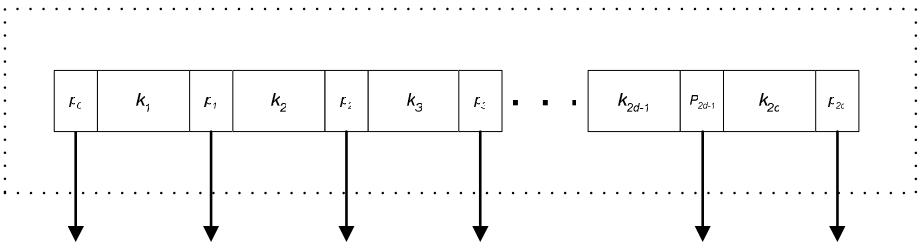
Ευρετήριο - 1



Κάθε μία σελίδα μνήμης: *2d τιμές k* (k_1 έως και k_{2d}), *2d+1* τιμές δεικτών (p_0 έως και p_{2d})

Τυπικός (εσωτερικός) κόμβος: τάξης *d*

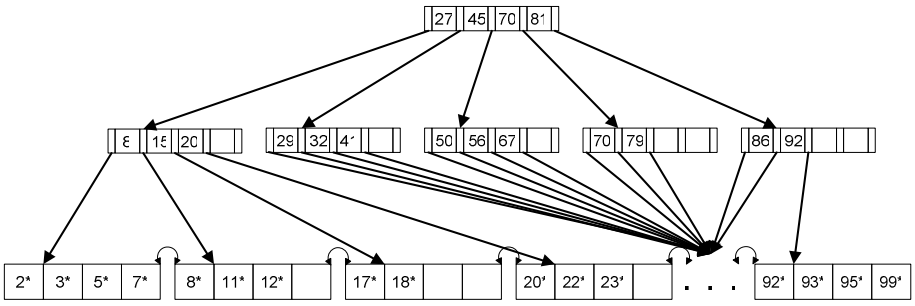
Ευρετήριο – 1: τυπική σελίδα μνήμης (τάξη *d*)



Με ανάλογο τρόπο: ευρετήριο-2 (αραιό και συγκροτημένο) επάνω στο ευρετήριο-1, κ.ο.κ.: δενδρική δομή. Στην πράξη: *d=50*, στα παραδείγματα που θα ακολουθήσουν: *d=2*

Στα φύλλα του δένδρου: καταγραφές δεδομένων *(k*)*

Στιγμιότυπο δενδρικού ευρετήριου (τάξης *d* = 2)



- Διατήρηση του δένδρου σε ισοζυγισμένη κατάσταση
- Ελάχιστος αριθμός καταχωρημένων τιμών *k* για τον τυπικό εσωτερικό κόμβο και για το κάθε ένα φύλλο: *d* (εξαίρεση: η ρίζα)

B⁺Δένδρο

B+ δένδρο τάξης *d*: διάσπαση κατά την εισαγωγή

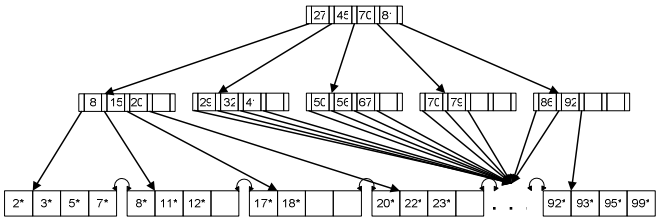
Μέγιστος αριθμός τιμών *k* (*k**) σε κόμβο/φύλλο: 2*d*

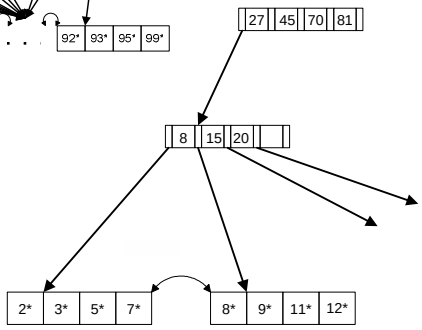
Όταν 2*d*+1: διάσπαση

Διάσπαση:

- $k_1 \ldots k_d$ στον αριστερό (υπάρχοντα) κόμβο/φύλλο
- (k_{d+1}), $k_{d+2} \ldots k_{2d+1}$ στον δεξί (νέο) κόμβο/φύλλο
- k_{d+1}: στο ανώτερο επίπεδο (αντιγραφή όταν διάσπαση φύλλου, μετακίνηση όταν διάσπαση εσωτερικού κόμβου)

Παράδειγμα: εισαγωγή *k**=9*

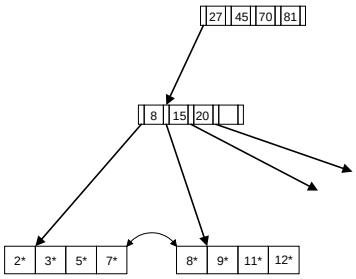


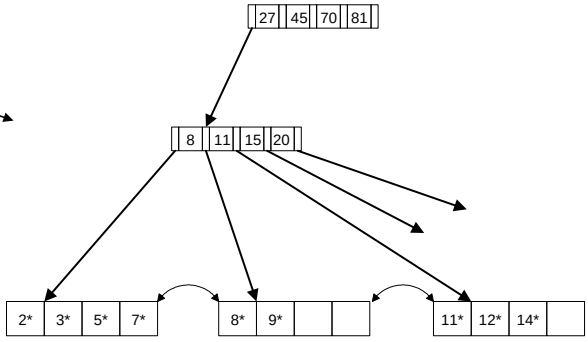


Κανένα πρόβλημα:

πρώτα 'αναζήτηση', μετά εισαγωγή

Εισαγωγή *k**=14*

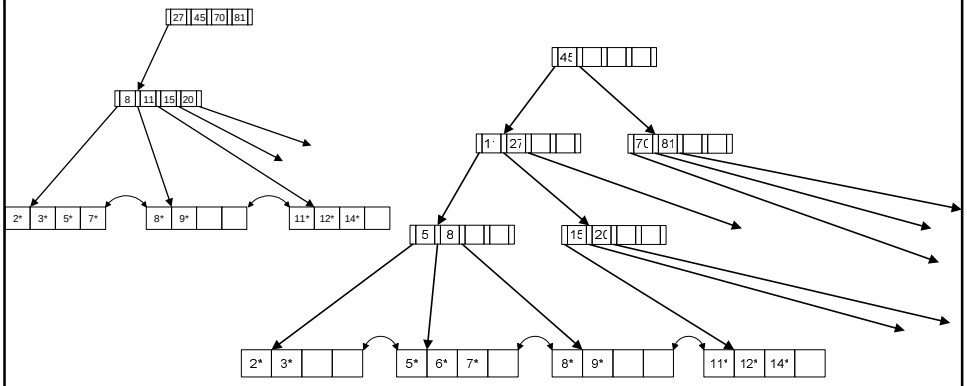




Διάσπαση (κάτω δεξιά):

- **8* και 9* αριστερά, 11*, 12* και 14* δεξιά**
- **Υπάρχει χώρος για μία ακόμη τιμή *k* στον κόμβο-πατέρα**
- **Το 11 αντιγράφεται στο αμέσως υψηλότερο επίπεδο**

Εισαγωγή *k**=6*



Διάσπαση του κάτω αριστερά, αντιγραφή του 5 στον κόμβο-πατέρα

Διάσπαση του κόμβου-πατέρα, μετακίνηση του 11 στη ρίζα

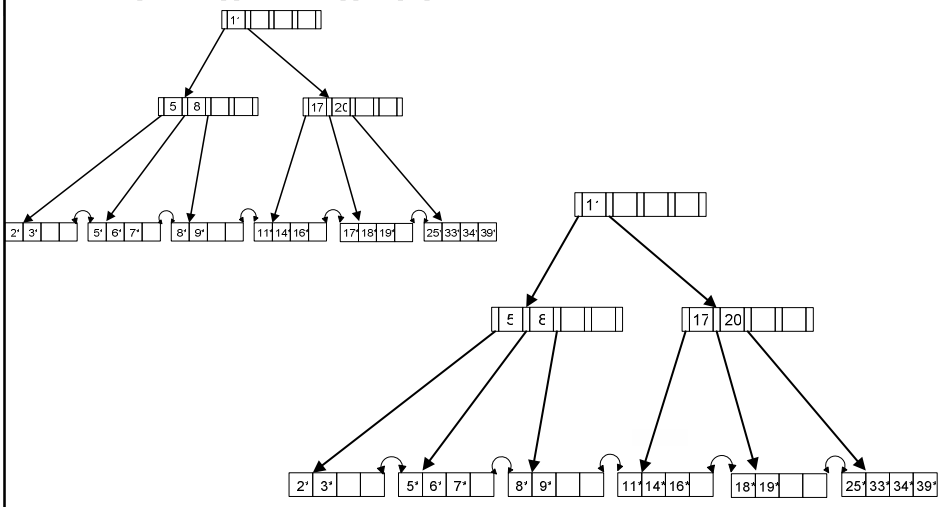
Διάσπαση της ρίζας, μετακίνηση του 45 σε νέα ρίζα

Αύξηση του ύψους του δένδρου κατά ένα επίπεδο

B+ δένδρο: διαγραφή εγγραφής

- Κάτω όριο: όχι λιγότερο από *d* τιμές *k* (*k**) ανά κόμβο/φύλλο
- Άνω όριο: όχι περισσότερες από *2d* τιμές *k* (*k**) ανά κόμβο/φύλλο. Μπορεί να παραβιάζεται στη διαγραφή; **Ναι!**
- Επανακατανομή τιμών *k* (*k**) σε κενές θέσεις (όταν υπάρχουν) γειτονικών κόμβων που έχουν κοινό πατέρα με τον κόμβο όπου συμβαίνει η διαγραφή

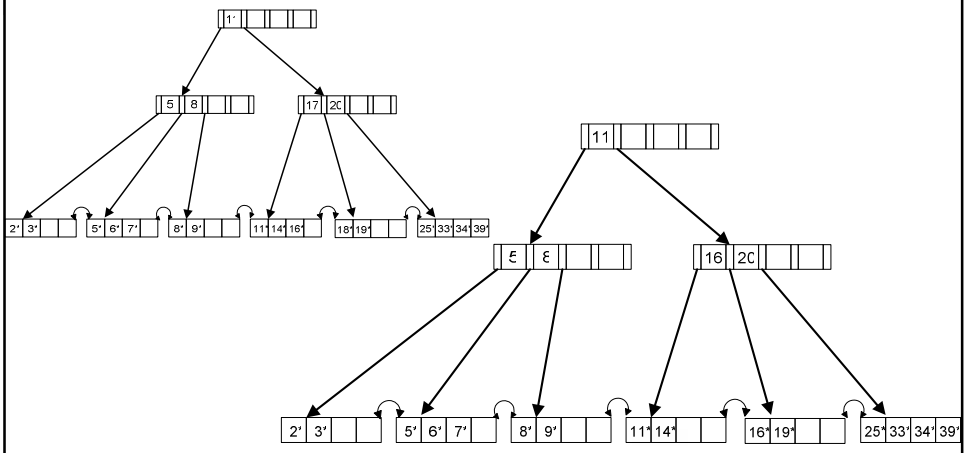
Παράδειγμα: διαγραφή του 17*



Ευσταθής η (νέα) κατάσταση του δεύτερου από κάτω δεξιά

Το '17' παραμένει στον κόμβο-πατέρα (γιατί;)

Διαγραφή του 18*



Ασταθής η νέα κατάσταση του 2ου κάτω-δεξιά

Επανακατονομή μεταξύ αυτού και του αριστερά του (γιατί;)

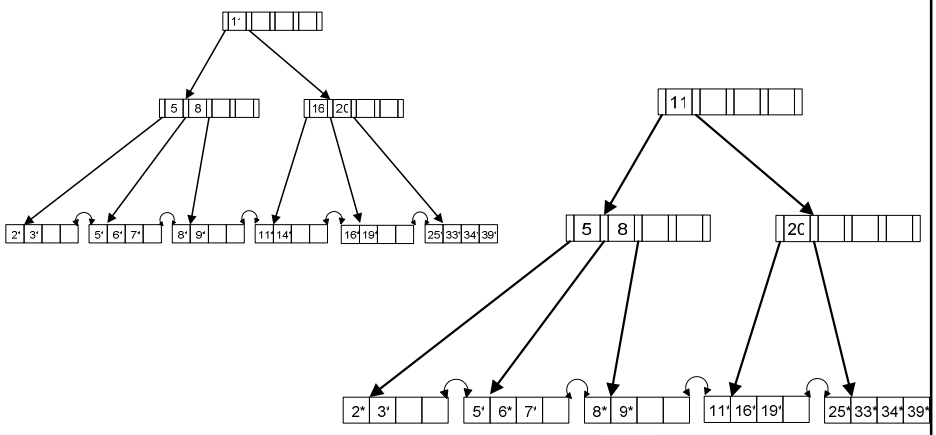
11* και 14* μένουν, 16* και 19* δεξιά, το 16* αντιγράφεται επάνω

Επανακατανομή τιμών *k** κατά τη διαγραφή

Εφόσον αυτή γίνεται ανάμεσα σε κόμβους οι οποίοι έχουν κοινό πατέρα-κόμβο, η επαναδιοργάνωση συνιστά ‘διαταραχή’ τοπικού χαρακτήρα. Επηρρεάζει μόνον τους εν λόγω γειτονικούς κόμβους και (το πολύ) τον κοινό τους πατέρα-κόμβο.

Δεν προκαλείται το λεγόμενο ‘φαινόμενο της χιονοστιβάδας’.

Διαγραφή του 14*



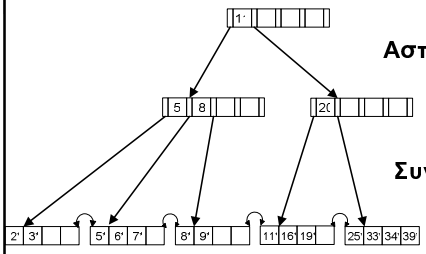
Ο τρίτος απο δεξιά κάτω: ασταθής

Γειτονικός μόνο στα δεξιά του: η επανακατανομή δεν αποδίδει

Συγχώνευση: το 16 διαγράφεται από τον κοινό τους πατέρα

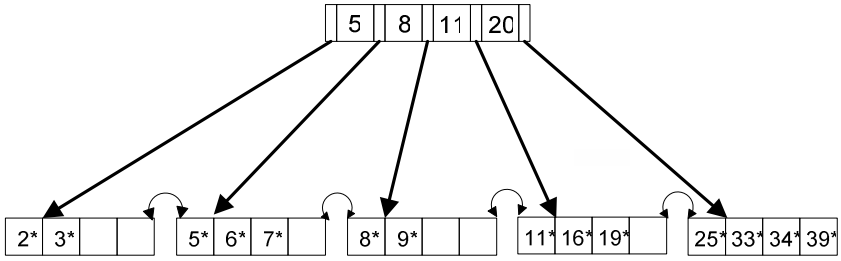
Ο πατέρας-κόμβος γίνεται ασταθής...

Διαγραφή του 14* (συνέχεια)



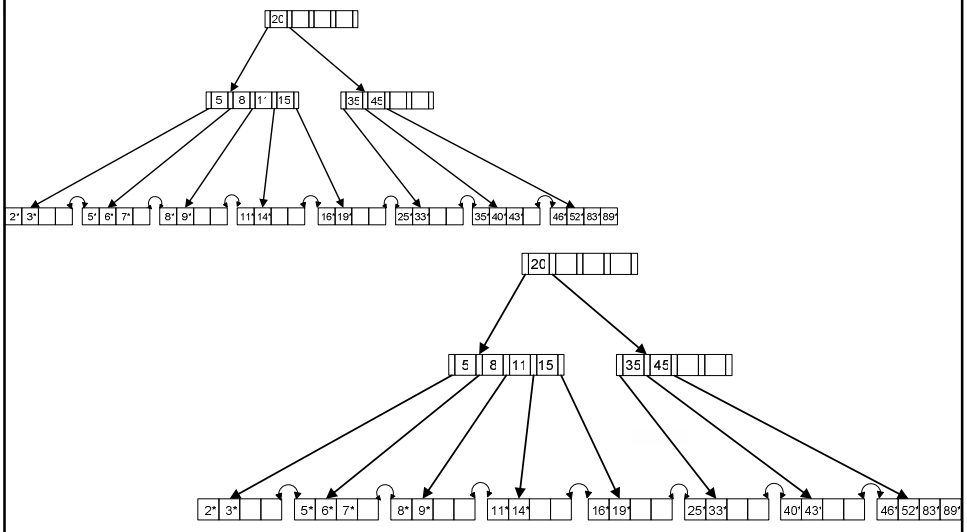
Ασταθής ο δεξιά κόμβος στο μεσαίο επίπεδο

Συγχωνεύεται με τον (μοναδικό) γειτονικό του και (αναγκαστικά) 'καταρρέει' η ρίζα



Το ύψος του δένδρου ελαττώνεται κατά ένα

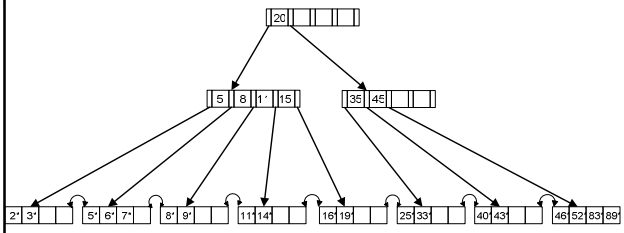
Διαγραφή του 35*



Ευσταθής κατάσταση

Το '35' παραμένει στον εσωτερικό κόμβο (γιατί;)

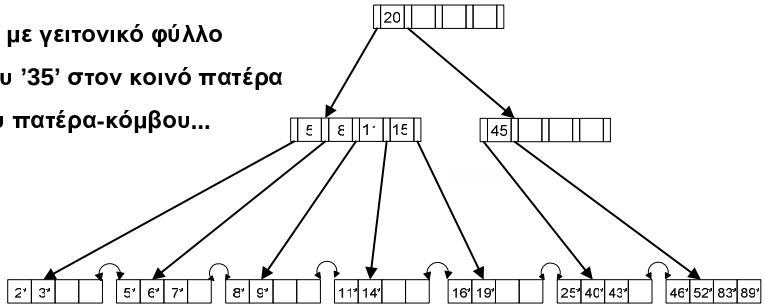
Διαγραφή του 33*



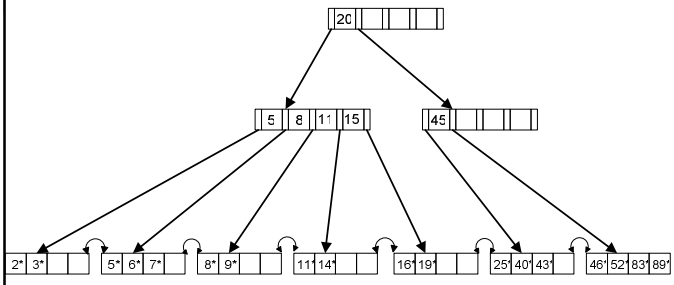
Συγχώνευση με γειτονικό φύλλο

Διαγραφή του ’35’ στον κοινό πατέρα

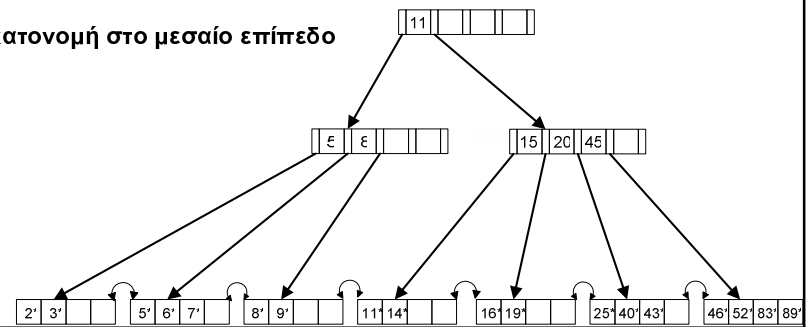
Αστάθεια του πατέρα-κόμβου...



Διαγραφή του 33* (συνέχεια)



Επανακατονομή στο μεσαίο επίπεδο



Ευρετήριο και SQL

```
>>-CREATE----+---------+--INDEX--index-name---------------------->
             '-UNIQUE--'

                                .-,------------------------.
                      (1)       V                .-ASC--.  |
>----ON--+-table-name------+--(-----column-name--+------+--+---)->
         |          (2)    |                    '-DESC-'
         '-nickname--------'

>-----+--------------------+----------------------------------->
      '-SPECIFICATION ONLY--'

>----*--+-------------------------------------------+---------->
        |                    .-,---------------.    |
        |          (3)       V                 |    |
        '-INCLUDE---------(-----column-name---+---)--'

>----*--+-------------------------------------------------------------------------
        +-CLUSTER---------------------------------------------------------------
        '-EXTEND USING--index-extension-name--+-------------------------------
                                              |    .-,-----------------------.
                                              |    V                         |
                                              '-(-----constant-expression---+---)

     .-PCTFREE 10--------.
>----+-------------------+---*----+----------------------+--*--->
     '-PCTFREE--integer--'        '-MINPCTUSED--integer--'

     .-DISALLOW REVERSE SCANS--.
>----+-------------------------+---*---------------------------><
     '-ALLOW REVERSE SCANS-----'
```


Πίνακας: telcosts

TELNO	AMT
2310717322	62300
2310737322	14000
2107417322	27000
2107276382	12300
2107173272	62300
2107173229	1000
2207173272	80000
2807173272	51000
3307173272	1500
2207146272	22000
2155171172	11000
1310717322	62300
2210737322	14000
2137417322	27000
2104276382	12300
2107573272	62300
2107163229	1000
2207177272	80000
2807173872	51000
3307173292	1500

Δημιουργία ευρετηρίου

CREATE INDEX aname ON telcosts(telno)

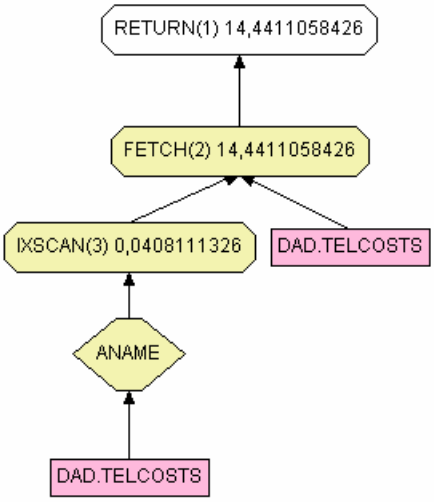
…Το RDBMS ‘αποφασίζει’ πότε θα χρησιμοποιείται

Αναζήτηση - 1

select *

from telcosts

where telno like '21%'



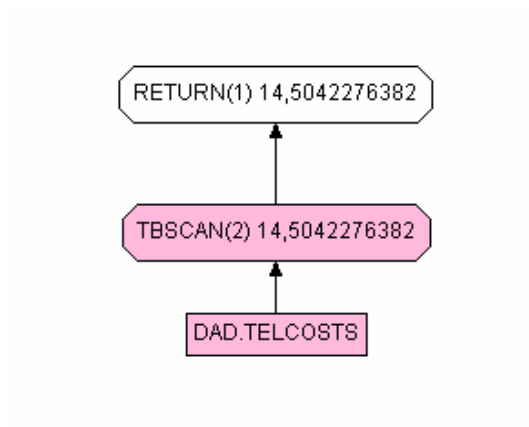
Το ευρετήριο χρησιμοποιείται

Αναζήτηση - 2

select *

from telcosts

where telno like '%21%'



Το ευρετήριο ΔΕΝ χρησιμοποιείται

Τέλος

Τέλος μίζοντας με τα του Β! Μέρους

- Δενδρικό ευρετήριο: Β+ δένδρο
- Διατηρείται ισοζυγισμένο
- Αλγόριθμος διατήρησης του ισοζυγισμένου κατά την εισαγωγή
- Αλγόρι ... υ κατά τη διαγρα
- Εντολέ ... κατάργηση του ευρ ... για τη χρήση του ... ηρίου κατά την αναζήτηση (γιατί;)

ΤΕΛΟΣ
με τα του φυσικού επιπέδου
στο σχεσιακό ΣΔΒΔ