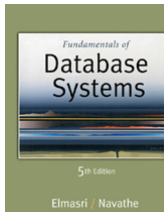


Κεφάλαιο 14

Δομές Ευρετηρίων για Αρχεία



PEARSON
Addison
Wesley

Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe Ελληνική Έκδοση,
Διαβλος, Επιμέλεια Μ. Χατζόπουλος

Θα μιλήσουμε για

- Τύποι Ταξινομημένων Ευρετηρίων ενός επιπέδου
 - Πρωτεύοντα Ευρετήρια
 - Ευρετήρια Συστάδες
 - Δευτερεύοντα Ευρετήρια
- Ευρετήρια Πολλών Επιπέδων
- Δυναμικά Ευρετήρια Πολλών Επιπέδων με χρήση Β-Δένδρων και Β+-Δένδρων
- Ευρετήρια σε Πολλαπλά Κλειδιά

Ευρετήρια σαν Μέθοδοι Προσπέλασης

- Ένα ευρετήριο ενός επιπέδου είναι ένα βοηθητικό αρχείο που κάνει πιο αποτελεσματική την αναζήτηση μιας εγγραφής σε ένα αρχείο δεδομένων.
- Το ευρετήριο συνήθως ορίζεται σε ένα πεδίο του αρχείου (αν και μπορεί να ορισθεί σε πολλά πεδία)
- Μια μορφή ευρετηρίου είναι ένα αρχείο με καταχωρήσεις <τιμή πεδίου, δείκτης στην εγγραφή>, που είναι ταξινομημένο με τιμή πεδίου
- Το ευρετήριο λέγεται δρόμος προσπέλασης στο πεδίο.

Ευρετήρια σαν Μέθοδοι Προσπέλασης(συν.)

- Το αρχείο του ευρετηρίου συνήθως καταλαμβάνει σημαντικά μικρότερο πλήθος μπλοκ από ότι το αρχείο δεδομένων επειδή οι καταχωρήσεις του είναι κατά πολύ μικρότερες
- Μια δυαδική αναζήτηση στο ευρετήριο παράγει ένα δείκτη στην εγγραφή του αρχείου
- Τα ευρετήρια μπορούν επίσης να χαρακτηρισθούν σαν πυκνά ή αραιά
 - Ένα **πυκνό ευρετήριο** έχει μια καταχώρηση ευρετηρίου για κάθε αναζήτηση τιμής κλειδιού (και επομένως κάθε εγγραφή) στο αρχείο δεδομένων.
 - Ένα **αραιό (ή μη πυκνό) ευρετήριο**, από την άλλη, έχει καταχωρήσεις ευρετηρίου μόνο για κάποιες από τις τιμές αναζήτησης.

Ευρετήρια σαν Μέθοδοι Προσπέλασης(συν.)

- Παράδειγμα: Έστω το αρχείο δεδομένων ΕΡΓΑΖΟΜΕΝΟΣ(ΟΝΟΜΑ, ΑΡ_ΤΑΥΤ, ΔΙΕΥΘΥΝΣΗ, ΕΡΓΑΣΙΑ, ΜΙΣΘΟΣ, ...)
- Υποθέστε ότι:
 - μέγεθος εγγραφής $R=150$ bytes μέγεθος μπλοκ $B=512$ bytes $r=30000$ εγγραφές
- Τότε, έχουμε:
 - παράγοντας ομαδοποίησης $Bfr = B \div R = 512 \div 150 = 3$ εγγραφές/μπλοκ
 - πλήθος μπλοκ αρχείου $b = (r / Bfr) = (30000 / 3) = 10000$ μπλοκ
- Για ένα ευρετήριο στο πεδίο ΑΡ_ΤΑΥΤ, υποθέστε μέγεθος πεδίου $V_{ΑΡΤΑΥΤ}=9$ bytes, υποθέστε μέγεθος δείκτη εγγραφής $P_R=7$ bytes. Τότε:
 - μέγεθος καταχώρησης ευρετηρίου $R_i = (V_{SSN} + P_R) = (9 + 7) = 16$ bytes
 - παράγοντας ομαδοποίησης ευρετηρίου $Bfr_i = B \div R_i = 512 \div 16 = 32$ καταχ/μπλοκ
 - πλήθος μπλοκ ευρετηρίου $b = (r / Bfr_i) = (30000 / 32) = 938$ μπλοκ
 - η δυαδική αναζήτηση απαιτεί $\log_2 b = \log_2 938 = 10$ μπλοκ προσπελάσεις
 - Σε σχέση με το κόστος της μέσης γραμμικής αναζήτησης που είναι:
 - $(b/2) = 30000/2 = 15000$ μπλοκ προσπελάσεις
 - Αν οι εγγραφές του αρχείου είναι ταξινομημένες, η δυαδική αναζήτηση θα ήταν:
 - $\log_2 b = \log_2 30000 = 15$ μπλοκ προσπελάσεις

Τύποι Ευρετηρίων ενός Επιπέδου

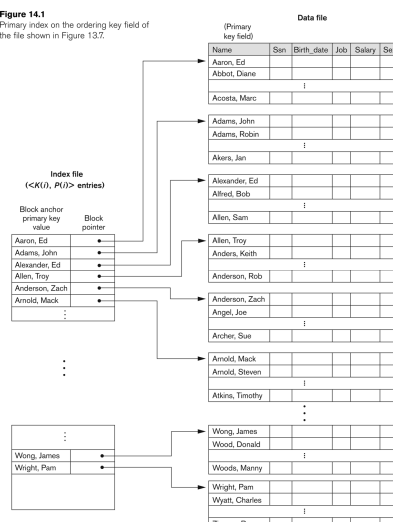
- **Πρωτεύον Ευρετήριο**
 - Ορίζεται σε ένα ταξινομημένο αρχείο δεδομένων
 - Το αρχείο δεδομένων είναι ταξινομημένο σε ένα **πεδίο κλειδί**
 - Περιλαμβάνει μια καταχώρηση ευρετηρίου για κάθε μπλοκ στο αρχείο δεδομένων· η καταχώρηση ευρετηρίου έχει την τιμή του πεδίου κλειδιού για την *πρώτη εγγραφή* στο μπλοκ, που ονομάζεται *άγκυρα του μπλοκ*
 - Ένα παρόμοιο σχήμα μπορεί να χρησιμοποιεί την *τελευταία εγγραφή* σε ένα μπλοκ.
 - Ένα πρωτεύον ευρετήριο είναι ένα μη πυκνό (αραιό) ευρετήριο, αφού έχει μια καταχώρηση για κάθε μπλοκ του αρχείου δεδομένων στο δίσκο και στα κλειδιά των εγγραφών άγκυρα παρά για κάθε τιμή αναζήτησης.

Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση, Διάβλος, Επιμέλεια Μ.Χατζόπουλος

Διαφάνεια 14- 7

Πρωτεύον ευρετήριο στο πεδίο κλειδί ταξινόμησης

Figure 14.1
Primary index on the ordering key field of the file shown in Figure 13.7.



Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση, Διάβλος, Επιμέλεια Μ.Χατζόπουλος

Διαφάνεια 14- 8

Τύποι Ευρετηρίων ενός Επιπέδου

■ Ευρετήριο Συστάδα

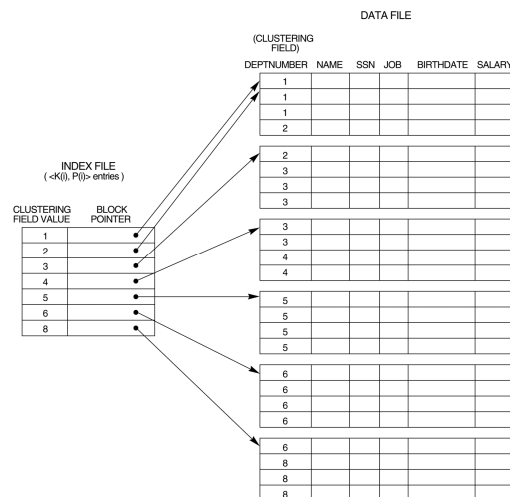
- Ορίζεται σε ένα ταξινομημένο αρχείο
- Το αρχείο δεδομένων είναι ταξινομημένο σε ένα πεδίο που δεν είναι κλειδί, σε αντίθεση από το πρωτεύον ευρετήριο, που απαιτεί ότι το πεδίο ταξινόμησης στο αρχείο δεδομένων έχει μια διακριτή τιμή για κάθε εγγραφή.
- Περιλαμβάνει μια καταχώρηση ευρετηρίου για κάθε διακριτή τιμή του πεδίου· η καταχώρηση ευρετηρίου δείχνει στο πρώτο μπλοκ δεδομένων που περιέχει εγγραφές με αυτή την τιμή πεδίου.
- Είναι ένα ακόμη παράδειγμα μη πυκνού ευρετηρίου όπου η εισαγωγή και η διαγραφή είναι σχετικά εύκολη με ένα ευρετήριο συστάδα.

Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση, Διάβλος, Επιμέλεια Μ.Χατζόπουλος

Διαφάνεια 14-9

Ένα παράδειγμα ευρετηρίου συστάδας

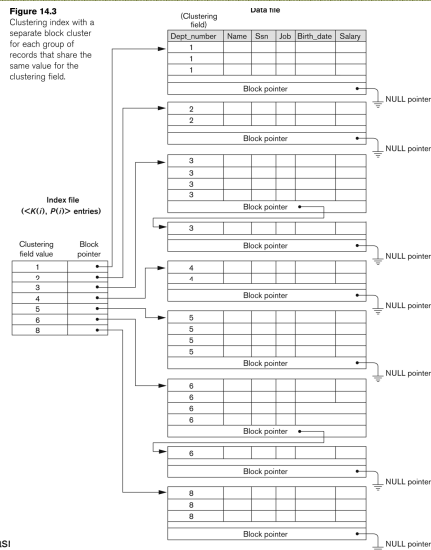
Ένα ευρετήριο συστάδα στο πεδίο ταξινόμησης ΚΩΔ_ΤΜΗΜ που δεν είναι κλειδί στο αρχείο ΕΡΓΑΖΟΜΕΝΟΣ.



Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση, Διάβλος, Επιμέλεια Μ.Χατζόπουλος

Διαφάνεια 14-10

Ένα ακόμη παράδειγμα ευρετηρίου συστάδας



Copyright © 2007 Ramez Elmasri
Διάβλος, Επιμέλεια Μ. Χατζόπουλος

Διαφάνεια 14- 11

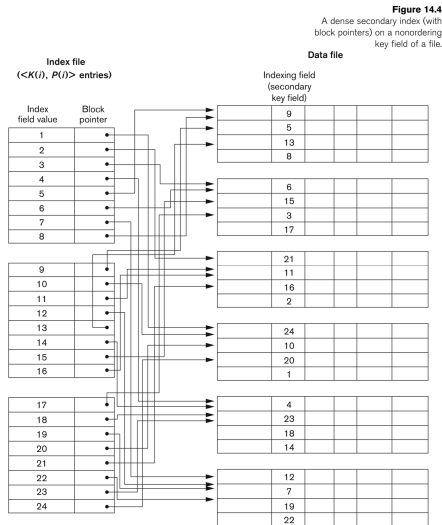
Τύποι Ευρετηρίων ενός Επιπέδου

- **Δευτερεύον Ευρετήριο**
 - Ένα δευτερεύον ευρετήριο υποστηρίζει ένα δευτερεύοντα τρόπο προσπέλασης ενός αρχείου για το οποίο υπάρχει ήδη πρωτεύουσα οργάνωση.
 - Το δευτερεύον ευρετήριο μπορεί να είναι σε ένα πεδίο που είναι υποψήφιο κλειδί και έχει μοναδική τιμή σε κάθε εγγραφή, ή ένα πεδίο που δεν είναι κλειδί με διπλές τιμές.
 - Το ευρετήριο είναι ένα ταξινομημένο αρχείο με δύο πεδία.
 - Το πρώτο πεδίο είναι ίδιου τύπου δεδομένων με κάποιο **πεδίο που δεν είναι κλειδί** του αρχείου δεδομένων και είναι το πεδίο ευρετηρίου.
 - Το δεύτερο πεδίο είναι ή δείκτης **μπλοκ** ή δείκτης εγγραφής .
 - Μπορεί να υπάρχουν *πολλά* δευτερεύοντα ευρετήρια (και επομένως, πεδία ευρετηρίασης) για το ίδιο αρχείο.
 - Περιλαμβάνει μια καταχώρηση για κάθε εγγραφή στο αρχείο δεδομένων· επομένως, είναι ένα **πυκνό ευρετήριο**.

Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση,
Διάβλος, Επιμέλεια Μ. Χατζόπουλος

Διαφάνεια 14- 12

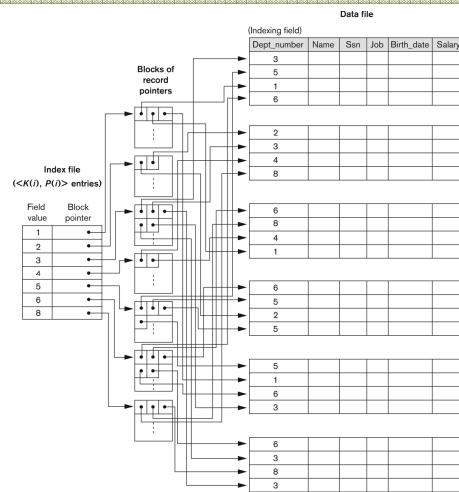
Παράδειγμα ενός Πυκνού Δευτερεύοντος Ευρετηρίου



Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση, Διαβλός, Επιμέλεια Μ.Χατζόπουλος

Διαφάνεια 14- 13

Ένα Παράδειγμα Δευτερεύοντος Ευρετηρίου



Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση, Διαβλός, Επιμέλεια Μ.Χατζόπουλος

Διαφάνεια 14- 14

Properties of Index Types

Τύπος Ευρετηρίου	Πλήθος Καταχωρήσεων του Ευρετηρίου (Πρώτου Επιπέδου)	Πυκνό η Μη Πυκνό	Άγκυρες μπλοκ στο Αρχείο Δεδομένων
Πρωτεύον Ευρετήριο	Πλήθος μπλοκ του αρχείου δεδομένων	Μη πυκνό	Ναι
Ευρετήριο Συστάδων	πλήθος διαφορετικών τιμών του πεδίου ευρετηρίασης	Μη Πυκνό	Ναι/όχι ^α
Δευτερεύον Ευρετήριο πάνω σε πεδίο-κλειδί	Πλήθος εγγραφών αρχείου δεδομένων	Πυκνό	όχι
Δευτερεύον Ευρετήριο πάνω σε πεδίο-μη κλειδί	Πλήθος εγγραφών ^β η πλήθος διαφορετικών τιμών του πεδίου ευρετηριοποίησης ^γ	Πυκνό ή μη πυκνό	όχι

Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση, Διάβλος, Επιμέλεια Μ.Χατζόπουλος

Διαφάνεια 14- 15

Ευρετήρια Πολλών Επιπέδων

- Επειδή ένα ευρετήριο ενός επιπέδου είναι ένα ταξινομημένο αρχείο, μπορούμε να δημιουργήσουμε ένα πρωτεύον ευρετήριο *στο ίδιο το ευρετήριο*.
 - Στην περίπτωση αυτή, το αρχικό αρχείο ευρετηρίου ονομάζεται *πρώτο επίπεδο ευρετηρίου* και το ευρετήριο του ευρετηρίου ονομάζεται *δεύτερο επίπεδο ευρετηρίου*.
- Μπορούμε να επαναλάβουμε τη διαδικασία αυτή, δημιουργώντας ένα τρίτο, τέταρτο ... επίπεδο, μέχρι που όλες οι καταχωρήσεις στο *ψηλότερο επίπεδο* να χωράνε σε ένα μπλοκ μνήμης.
- Ένα ευρετήριο πολλών επιπέδων μπορεί να δημιουργηθεί οποιοδήποτε τύπο ευρετηρίου πρώτου επιπέδου (πρωτεύον, δευτερεύον, συστάδα) όσο το πρώτο επίπεδο ευρετηρίου αποτελείται από *περισσότερα από ένα μπλοκ δίσκου*

Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση, Διάβλος, Επιμέλεια Μ.Χατζόπουλος

Διαφάνεια 14- 16

Πρωτεύον Ευρετήριο δύο επιπέδων

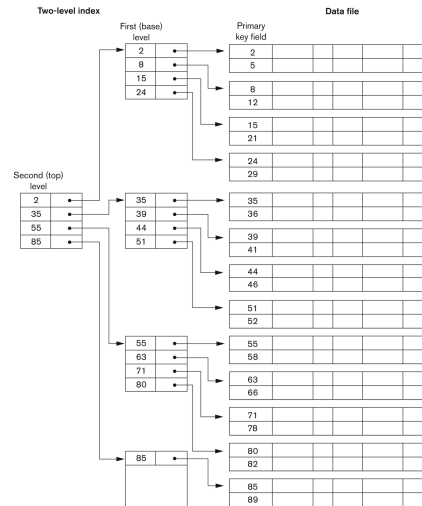


Figure 14.8
A two-level primary index resembling ISAM (Index Sequential Access Method) organization.
Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση,
Διάβλος, Επιμέλεια Μ. Χατζόπουλος

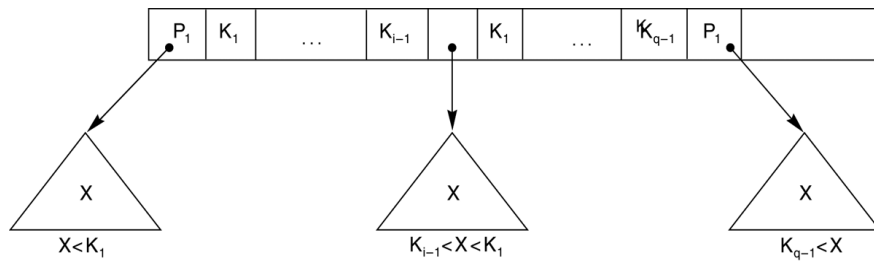
Διαφάνεια 14- 17

Ευρετήρια Πολλών Επιπέδων

- Ένα τέτοιο ευρετήριο πολλών επιπέδων είναι μια μορφή δένδρου αναζήτησης
 - Ωστόσο, η εισαγωγή και η διαγραφή καταχωρήσεων στο ευρετήριο αποτελεί σοβαρό πρόβλημα επειδή κάθε επίπεδο του ευρετηρίου είναι ένα ταξινομημένο αρχείο.

Διαφάνεια 14- 18

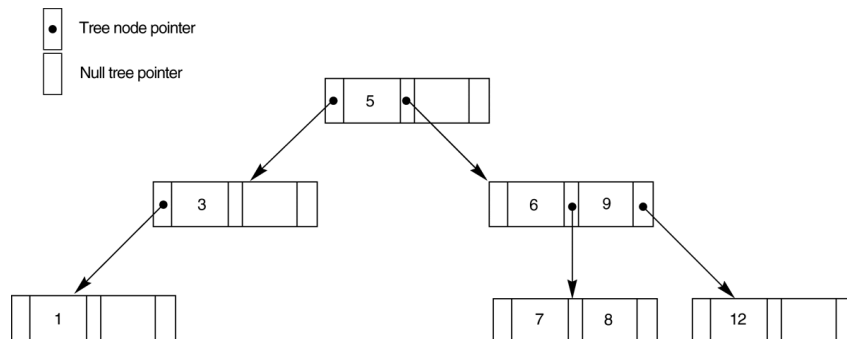
Ένας κόμβος σε ένα δένδρο αναζήτησης με δείκτες στα υποδένδρα του



Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση, Διάβλος, Επιμέλεια Μ. Χατζόπουλος

Διαφάνεια 14-19

Ένα δένδρο αναζήτησης τάξεως $p = 3$.



Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση, Διάβλος, Επιμέλεια Μ. Χατζόπουλος

Διαφάνεια 14-20

Δυναμικά Πολυεπίπεδα Ευρετήρια με Χρήση B-Δένδρων και B+-Δένδρων

- Τα περισσότερα ευρετήρια πολλών επιπέδων χρησιμοποιούν δομές δεδομένων B-δένδρων ή B+-δένδρων λόγω του προβλήματος εισαγωγής και διαγραφής
 - Αυτό αφήνει σε κάθε κόμβο του δένδρου (μπλοκ στο δίσκο) ελεύθερο χώρο για νέες καταχωρήσεις
- Αυτές οι δομές δεδομένων αποτελούν παραλλαγές των δένδρων αναζήτησης που υποστηρίζουν αποτελεσματική εισαγωγή και διαγραφή νέων τιμών αναζήτησης.
- Στις δομές δεδομένων B-δένδρων και B+-δένδρων, κάθε κόμβος αντιστοιχεί σε ένα μπλοκ του δίσκου
- Κάθε κόμβος είναι από γεμάτος ή μέχρι τη μέση

Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση, Διάβλος, Επιμέλεια Μ.Χατζόπουλος

Διαφάνεια 14- 21

Δυναμικά Πολυεπίπεδα Ευρετήρια με Χρήση B-Δένδρων και B+-Δένδρων(συν.)

- Η εισαγωγή σε ένα κόμβο που δεν είναι γεμάτος είναι πολύ αποδοτική
 - Αν ένας κόμβος είναι γεμάτος η εισαγωγή προκαλεί διάσπαση σε δύο κόμβους
- Η διάσπαση μπορεί να διαδοθεί σε άλλα επίπεδα του δένδρου
- Μια διαγραφή είναι πολύ αποδοτική αν ένας κόμβος δεν πέφτει κάτω από το μισό
- Αν μια διαγραφή οδηγήσει σε κόμβο με λιγότερες από μισές καταχωρήσεις, πρέπει να συνενωθεί με γειτονικούς κόμβους

Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση, Διάβλος, Επιμέλεια Μ.Χατζόπουλος

Διαφάνεια 14- 22

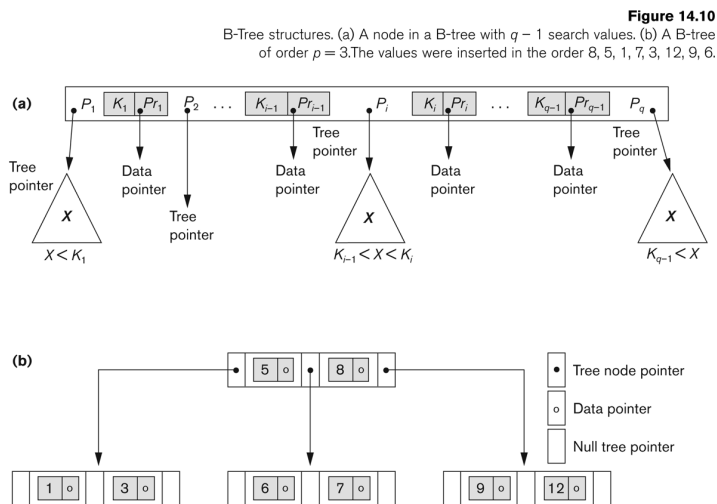
Διαφορά μεταξύ B-δένδρου και B+-δένδρου

- Σε ένα B-δένδρου, υπάρχουν δείκτες προς εγγραφές δεδομένων σε όλα τα επίπεδα του δένδρου
- Σε ένα B+-δένδρο, όλοι οι δείκτες προς τις εγγραφές δεδομένων υπάρχουν στους κόμβους φύλα
- Ένα B+-δένδρο μπορεί να έχει λιγότερα επίπεδα (ή μεγαλύτερη χωρητικότητα τιμών αναζήτησης) από το αντίστοιχο B-δένδρο

Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση, Διάβλος, Επιμέλεια Μ. Χατζόπουλος

Διαφάνεια 14-23

Δομές B-δένδρου

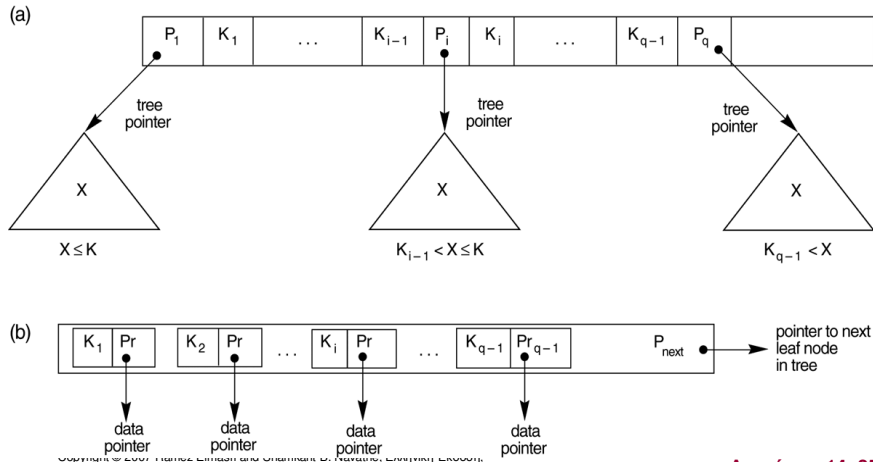


Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση, Διάβλος, Επιμέλεια Μ. Χατζόπουλος

Διαφάνεια 14-24

Οι κόμβοι ενός B+-δένδρου

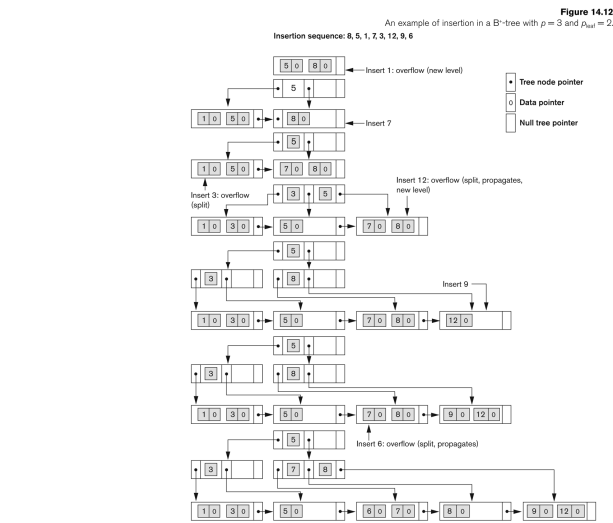
- FIGURE 14.11 The nodes of a B+-tree
 - (a) Internal node of a B+-tree with $q-1$ search values.
 - (b) Leaf node of a B+-tree with $q-1$ search values and $q-1$ data pointers.



Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση, Διαβλός, Επιμέλεια Μ. Χατζόπουλος

Διαφάνεια 14-25

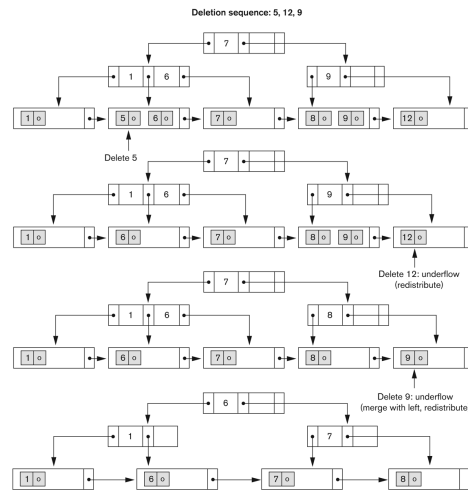
Ένα παράδειγμα εισαγωγής σε B+-δένδρο



Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση, Διαβλός, Επιμέλεια Μ. Χατζόπουλος

Διαφάνεια 14-26

Ένα παράδειγμα διαγραφής σε ένα B+-δένδρο



Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση, Διάβλος, Επιμέλεια Μ. Χατζόπουλος

Figure 14.13
An example of deletion from a B+-tree.

Διαφάνεια 14-27

Σύνοψη

- Τύποι διατεταγμένων ευρετηρίων ενός επιπέδου
 - Πρωτεύοντα ευρετήρια
 - Ευρετήρια Συστάδες
 - Δευτερεύοντα Ευρετήρια
- Ευρετήρια πολλών επιπέδων
- Δυναμικά ευρετήρια πολλών επιπέδων με χρήση B-δένδρων και B+-δένδρων
- Ευρετήρια σε πολλά κλειδιά

Copyright © 2007 Ramez Elmasri and Shamkant B. Navathe, Ελληνική Έκδοση, Διάβλος, Επιμέλεια Μ. Χατζόπουλος

Διαφάνεια 14-28