

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ - ΤΜΗΥΠ

ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ Ι

Β. Μεγαλοικονόμου
Δ. Χριστοδουλάκης

Αποθήκευση Δεδομένων και Δομές αρχείων

(παρουσίαση βασισμένη εν μέρη σε σημειώσεις των Silberchatz, Korth και Sudarshan και του C. Faloutsos)

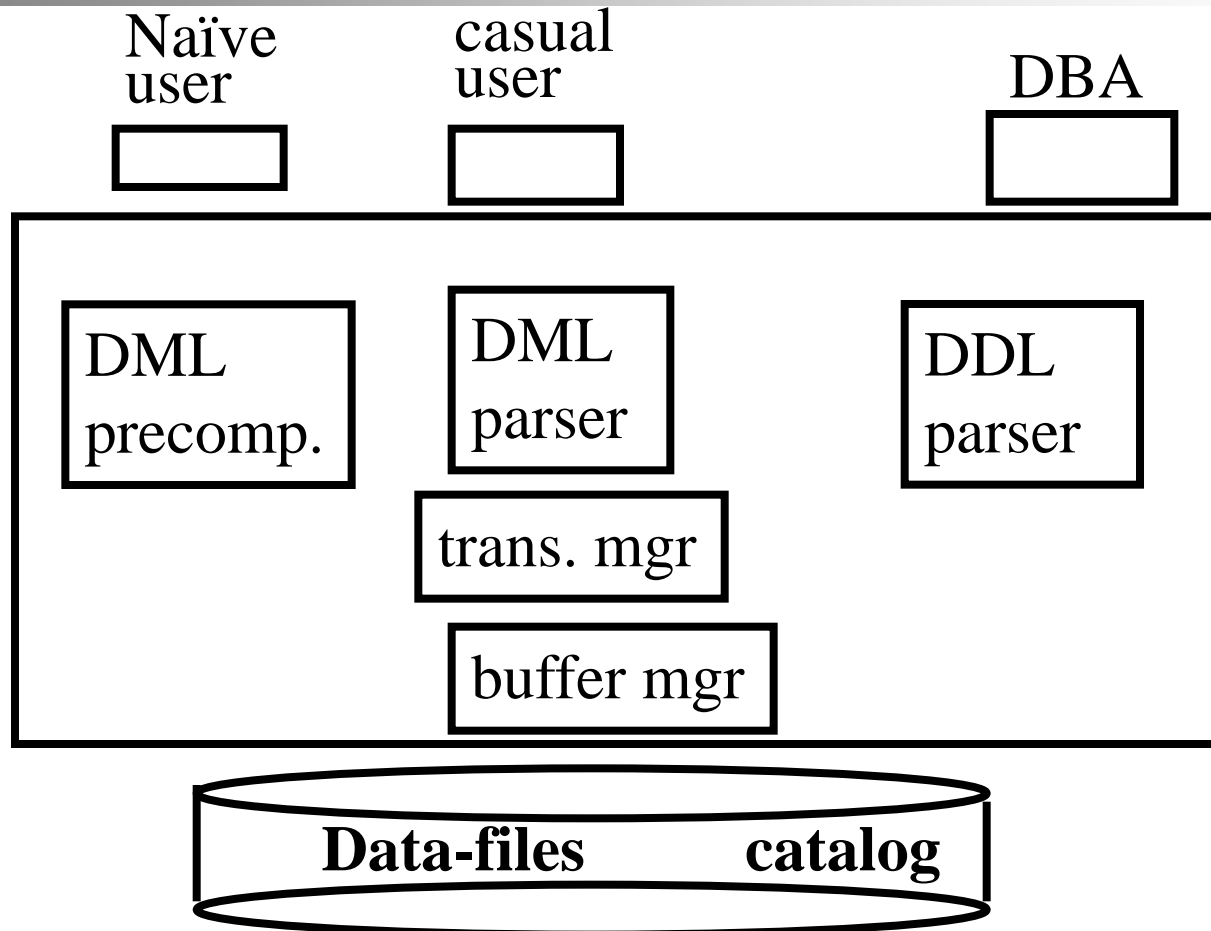


Σύνοψη Ύλης

Σχεσιακό μοντέλο (rel. model)

- **Σχεσιακό μοντέλο** (Relational model) - **SQL**
 - Επίσημες (Formal) & Εμπορικές γλώσσες ερωτήσεων (commercial query languages)
- **Συναρτησιακές Εξαρτήσεις** (Functional Dependencies)
- **Κανονικοποίηση** (Normalization)
- **Φυσικός Σχεδιασμός** (Physical Design)
- **Ευρετηριοποίηση** (Indexing)

Συνοπτική παρουσίαση ενός DBMS



DBMS : Σύστημα Διαχείρισης Βάσεων Δεδομένων



Λεπτομερής Σύνοψη

- Αποθήκευση και δομές αρχείων
 - Επισκόπηση των μέσων φυσικής αποθήκευσης
 - Χαρακτηριστικά των δίσκων αποθήκευσης
 - Τεχνολογία RAID
 - Πρόσβαση στον δίσκο (Storage Access)
 - Buffering
 - Οργανώσεις Αρχείων
 - Αποθήκευση καταλόγου

Διαχωρισμός των μέσων φυσικής αποθήκευσης

- Ταχύτητα με την οποία μπορούν να προσπελαστούν τα δεδομένα
- Κόστος ανά μονάδα δεδομένων
- Αξιοπιστία
 - Απώλεια δεδομένων σε περίπτωση διακοπής παροχής ισχύος ή σε περίπτωση πτώσης του συστήματος
 - Φυσική καταστροφή του μέσου αποθήκευσης
- Διάκριση μέσων αποθήκευσης σε:
 - **Ευμετάβλητη αποθήκευση (volatile storage):** Το περιεχόμενο χάνεται σε περίπτωση απώλειας ισχύος
 - **Μόνιμη αποθήκευση (non-volatile storage):**
 - Το περιεχόμενο διατηρείται σε περίπτωση απώλειας ισχύος
 - Περιλαμβάνει δευτερεύουσες μορφές αποθήκευσης και battery-backed up main-memory.



Μέσα Φυσικής αποθήκευσης

■ Κρυφή μνήμη (Cache)

- Το ταχύτερο μέσο αποθήκευσης
- Υψηλό κόστος
- Ευμετάβλητο – (volatile)
- Διαχειριζόμενο από το υλικό του υπολογιστικού συστήματος

■ Main memory:

- Πολύ γρήγορη πρόσβαση στα δεδομένα (10s to 100s of nanoseconds; 1 nanosecond = 10^{-9} seconds)
- Γενικά πολύ μικρή σε μέγεθος (or too expensive) ώστε να αποθηκευθεί σ' αυτήν ολόκληρη η βάση δεδομένων
 - Χωρητικότητα μέχρι μερικά Gigabytes
 - Η χωρητικότητα αυξάνεται και το κόστος ανά byte μειώνεται σταθερά (περίπου κατά έναν παράγοντα 2 κάθε 2 ή 3 χρόνια)
- **Ευμετάβλητο** μέσο αποθήκευσης
 - Τα περιεχόμενα χάνονται σε περίπτωση απώλειας ισχύος ή πτώσης του συστήματος.

Μέσα Φυσικής αποθήκευσης (συνέχεια.)

■ Μνήμη Flash

- Τα δεδομένα διατηρούνται σε περίπτωση απώλειας ισχύος
- Τα δεδομένα μπορούν να γραφτούν σε μία περιοχή (location) μόνο μια φορά, αλλά μία περιοχή μπορεί να σβηστεί και να γραφεί ξανά
 - Μπορεί να υποστηρίξει μόνο περιορισμένο κύκλο εγγραφών/διαγραφών.
 - Η διαγραφή λαμβάνει χώρα σε ολόκληρα τμήματα (Block) της μνήμης
- Οι αναγνώσεις είναι σχεδόν το ίδιο γρήγορες με αυτές στην κύρια μνήμη
- Οι εγγραφές είναι αργές (μερικά microseconds), οι διαγραφές είναι ακόμα πιο αργές
- Το κόστος ανά μονάδα αποθήκευσης είναι παρόμοιο με την κύρια μνήμη
- Χρησιμοποιείται ευρέως σε συσκευές όπως οι ψηφιακές φωτογραφικές μηχανές και τα κινητά τηλέφωνα
- ...είναι γνωστή και ως EEPROM (Electrically Erasable Programmable Read-Only Memory)

Μέσα Φυσικής αποθήκευσης (συνέχεια.)

■ Μαγνητικοί δίσκοι

- Τα δεδομένα αποθηκεύονται σε περιστρεφόμενο δίσκο ο οποίος εγγράφεται / διαβάζεται μαγνητικά
- Πρωτεύον μέσο για την αποθήκευση μεγάλης διάρκειας
- Για να προσπελαστούν τα δεδομένα θα πρέπει να μεταφερθούν από τον δίσκο στην κύρια μνήμη και να επανεγγραφούν στον δίσκο για αποθήκευση
 - Πολύ πιο αργή προσπέλαση από την κύρια μνήμη
- **Direct-access** – Δυνατότητα προσπέλασης των δεδομένων με οποιαδήποτε σειρά – σε αντίθεση με τις μαγνητικές ταινίες
- Χωρητικότητα – κυμαίνονται από μερικά GB έως μερικά TB
 - Πολύ μεγαλύτερη χωρητικότητα και πολύ μικρότερο κόστος ανά Byte σε σύγκριση με την κύρια μνήμη ή τις μνήμες Flash
 - Αυξάνεται σταθερά και ραγδαία με τεχνολογικές βελτιώσεις κατά έναν παράγοντα 2 κάθε 2 ή τρία χρόνια
- Τα δεδομένα διατηρούνται σε περίπτωση απώλειας ισχύος ή σε περιπτώσεις πτώσης του συστήματος
 - Οι αποτυχιές δίσκου μπορούν πολύ σπάνια να καταστρέψουν δεδομένα



Μέσα Φυσικής αποθήκευσης (συνέχεια.)

■ Οπτικά μέσα αποθήκευσης

- Μόνιμη αποθήκευση
- Τα δεδομένα διαβάζονται οπτικά από έναν περιστρεφόμενο δίσκο με την χρήση Laser
- Οι πιο δημοφιλείς μορφές CD-ROM (640 MB) και DVD (4.7 to 17 GB)
- Οι οπτικοί δίσκοι Write-one, read-many (WORM) χρησιμοποιούνται για αρχειακή αποθήκευση (CD-R and DVD-R)
- Εκδόσεις που επιτρέπουν πολλαπλές εγγραφές είναι διαθέσιμες (CD-RW, DVD-RW, and DVD-RAM)
- Οι αναγνώσεις και οι εγγραφές είναι περισσότερο χρονοβόρες από ότι στους μαγνητικούς δίσκους
- Τα συστήματα **Juke-box** χρησιμοποιούνται για την αποθήκευση δεδομένων πολύ μεγάλου μεγέθους και διαθέτουν: μεγάλο πλήθος αφαιρούμενων δίσκων, μικρό αριθμό οδηγών, μηχανισμό για το αυτόματο loading/unloading των δίσκων

Μέσα Φυσικής αποθήκευσης (συνέχεια.)

■ Μαγνητικές ταινίες αποθήκευσης

- Μόνιμη αποθήκευση
- Χρησιμοποιείται κυρίως για δημιουργία αντιγράφων ασφαλείας και αρχειακή αποθήκευση
- **sequential-access** – πολύ πιο αργή από τους μαγνητικούς δίσκους
- Πολύ μεγάλη χωρητικότητα (από 40 έως >300 GB)
- Οι ταινίες μπορούν να αφαιρεθούν από το drive \Rightarrow Το κόστος αποθήκευσης πολύ μικρότερο απ ότι το κόστος των μαγνητικών δίσκων – Οι οδηγοί παραμένουν ακριβοί
- Υπάρχουν jukeboxes Ταινιών για την αποθήκευση μεγάλου όγκου δεδομένων
 - Εκατοντάδες terabytes (1 terabyte = 10^9 bytes) μέχρι ακόμα και petabyte (1 petabyte = 10^{12} bytes)

Ιεραρχίες μνήμης και μονάδες αποθήκευσης

- **primary storage:** Μέσα με μεγάλη ταχύτητα πρόσβασης αλλά όχι μόνιμη αποθήκευση (κρυφή-cache, κύρια μνήμη-main memory).
- **secondary storage:** Επόμενο επίπεδο ιεραρχίας, μόνιμη αποθήκευση , σχετικά μικρός χρόνος προσπέλασης δεδομένων
 - Αποκαλείται επίσης **on-line storage**
 - Π.χ. flash memory, μαγνητικοί δίσκοι
- **tertiary storage:** Κατώτατο επίπεδο της ιεραρχίας, μόνιμη αποθήκευση, μεγάλος χρόνος πρόσβασης δεδομένων
 - Αποκαλείται επίσης **off-line storage**
 - Π.χ. Μαγνητικές ταινίες, οπτικά μέσα αποθήκευσης

Μαγνητικός Σκληρός Δίσκος

- **Seek time**

(χρόνος αναζήτησης/εντοπισμού)

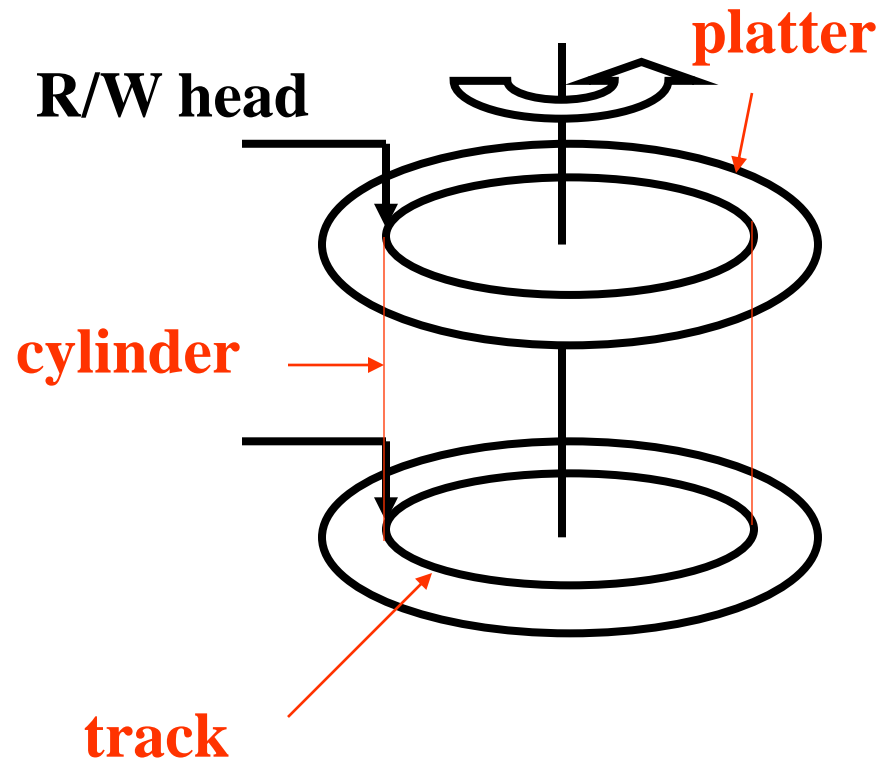
- **Rotation delay**

(Καθυστέρηση Περιστροφής)

Σχεδόν 2-10 msec vs micro/nano seconds για την κύρια μνήμη

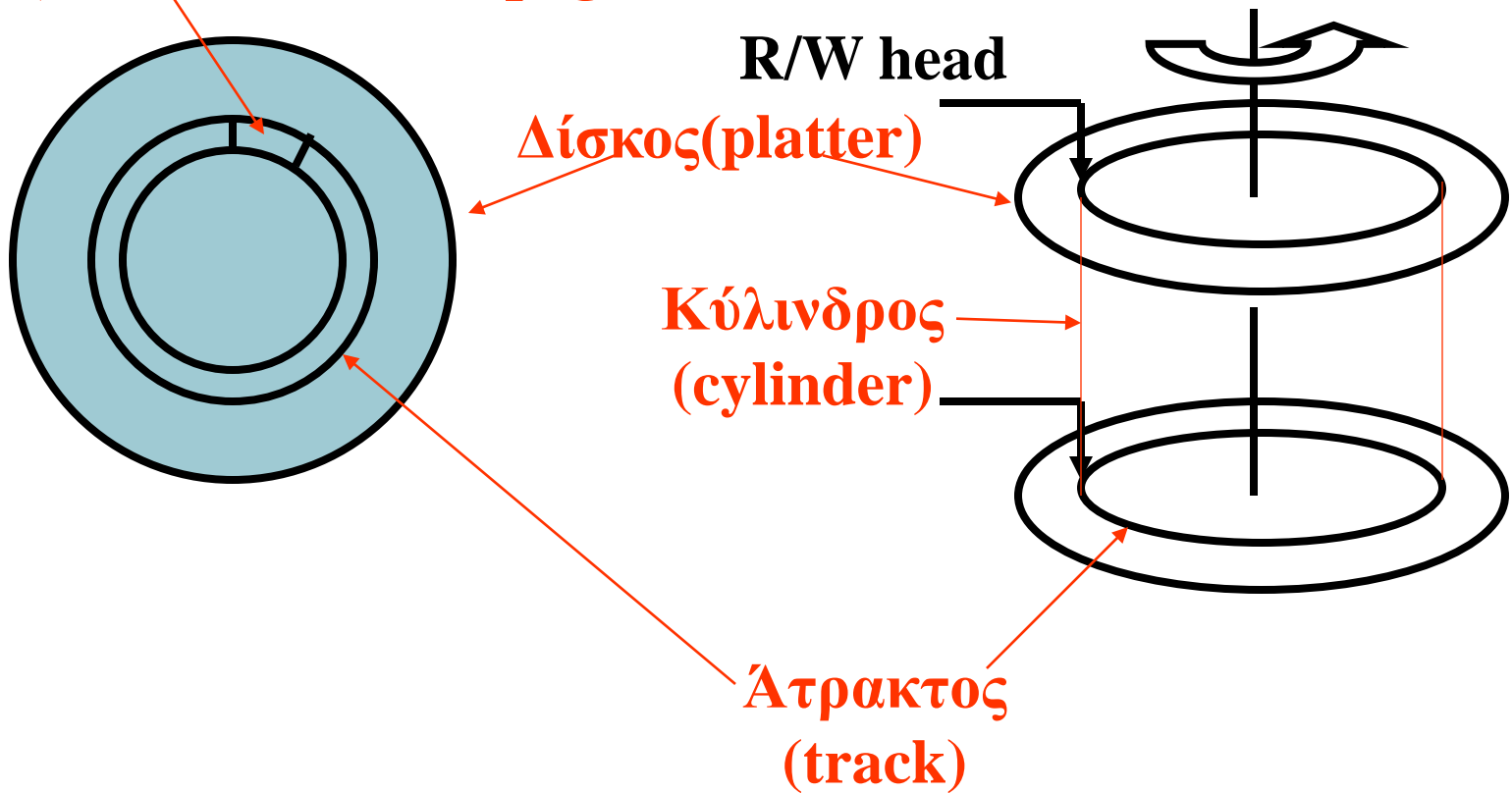
- **Transfer time**

(Χρόνος μεταφοράς)



Δίσκος

Τομέας (Sector) (= block=page)



Μαγνητικοί Δίσκοι (συνέχεια.)

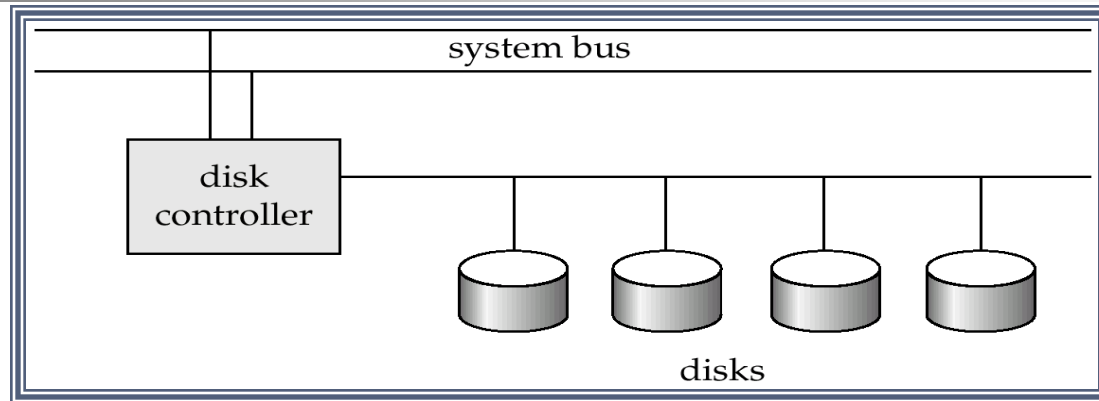
■ Μαγνητικοί δίσκοι προηγούμενης γενιάς επιρρεπείς σε αστοχίες κεφαλής

- Η επιφάνεια καλύπτεται από στρώμα μετάλλου-οξειδίου, το οποίο μπορεί να φθαρεί σε περίπτωση αστοχίας της κεφαλής ενδεχομένως προκαλώντας απώλεια δεδομένων
- Οι σύγχρονοι μαγνητικοί δίσκοι είναι λιγότερο επιρρεπείς σε τέτοιου είδους αστοχίες, δεν αποκλείεται η φθορά μεμονωμένων τομέων

■ **Disk controller** – Διεπαφή επικοινωνίας μεταξύ του υπολογιστικού συστήματος και του υλικού του μαγνητικού δίσκου

- Δέχεται εντολές υψηλού επιπέδου για ανάγνωση ή εγγραφή σε έναν τομέα
- Υπεύθυνο για την ενεργοποίηση ενεργειών όπως μετακίνηση του μηχανικού βραχίονα της κεφαλής στον σωστό τομέα και ανάγνωση ή εγγραφή δεδομένων
- Υπολογίζει και προσάπτει **ελέγχους πλεονασμού-checksums** σε κάθε τομέα ώστε να πιστοποιήσει ότι η ανάγνωση των δεδομένων γίνεται ορθά
 - Εάν τα δεδομένα έχουν αλλιωθεί τότε με μεγάλη πιθανότητα το αποθηκευμένο checksum δεν θα ταιριάζει με το checksum που θα προκύπτει από υπολογισμό
- Διασφαλίζει την επιτυχημένη εγγραφή με την επανανάγνωση του τομέα μετά την εγγραφή
- Εφαρμόζει την επανατοποθέτηση των κατεστραμμένων τομέων (**remapping of bad sectors**)

Υποσύστημα Δίσκων



- Πολλαπλοί δίσκοι συνδεδεμένοι σε ένα υπολογιστικό σύστημα μέσω ενός ελεγκτή (controller)
 - Λειτουργία ελεγκτών (checksum, bad sector remapping) συχνά επιτελείται ανεξάρτητα στους δίσκους; Μείωση του επεξεργαστικού φόρτου στους ελεγκτές
- Οικογένειες Προτύπων Interfaces Δίσκων
 - ATA (AT adaptor) σύνολο standards
 - SCSI (Small Computer System Interconnect)

Μετρικές Απόδοσης Δίσκων

■ **Access time- Χρόνος τυχαίας προσπέλασης**

Χρόνος για τον εντοπισμό ενός τυχαίου μπλοκ στο δίσκο από όταν δίνεται η διεύθυνση του μέχρι την έναρξη της μεταφοράς του μπλοκ από τον δίσκο στην μνήμη. Αποτελείται από:

- **Seek time-Χρόνος Αναζήτησης** – Ο χρόνος που απαιτείται για να τοποθετήσουν τα μηχανικά μέρη την κεφαλή στην σωστή άτρακτο
 - Μέσος χρόνος αναζήτησης: το 1/2 του χρόνου χειρότερης περίπτωσης. Από 4 έως 10 milliseconds σε τυπικούς δίσκους
- **Rotational latency-Καθυστέρηση Περιστροφής** – Ο χρόνος που απαιτείται ώστε η κεφαλή να βρεθεί πάνω από τον επιθυμητό τομέα σε μία δεδομένη άτρακτο
 - Μέση καθυστέρηση: το 1/2 της καθυστέρησης της χειρότερης περίπτωσης. 4 με 11 milliseconds σε τυπικούς δίσκους (5400 to 15000 r.p.m.)

■ **Data-transfer rate-Ρυθμός μεταφοράς δεδομένων** – Ο ρυθμός με τον οποίο τα δεδομένα μπορούν να ανακτηθούν ή να αποθηκευθούν στον δίσκο. Τυπικά 4 έως 8 MB ανά δευτερόλεπτο

- Πολλαπλοί δίσκοι μπορεί να διαμοιράζονται έναν ελεγκτή, οπότε ο ρυθμός με τον οποίο ο ελεγκτής μπορεί να διαχειριστεί είναι επίσης σημαντικός
 - Π.χ. ATA-5: 66 MB/second, SCSI-3: 40 MB/s, Fiber Channel: 256 MB/s



Μετρικές Απόδοσης (συνέχεια.)

- **Mean time to failure (MTTF)** – Μέσος χρόνος που αναμένεται να λειτουργήσει ο δίσκος χωρίς πρόβλημα.
 - Τυπικά 3 έως 5 χρόνια
 - Η πιθανότητα παρουσίασης προβλήματος σε δίσκους νέας τεχνολογίας είναι σχετικά χαμηλή
 - Αντιστοιχεί σε ένα θεωρητικό όριο μεταξύ MTTF 30,000 έως 1,200,000 ώρες λειτουργίας
 - Π.χ., Το όριο MTTF 1,200,000 ωρών λειτουργίας hours για έναν καινούργιο δίσκο μεταφράζεται ως εξής: Δεδομένων 1000 καινούργιων δίσκων κατά μέσο όρο ένα θα παρουσιάσει πρόβλημα κάθε 1200 ώρες λειτουργίας
 - Το MTTF μειώνεται με τον χρόνο χρησιμοποίησης του δίσκου

Βελτιστοποίηση πρόσβασης σε μπλοκ του δίσκου

- **Block** – μια συνεχόμενη ακολουθία από τομείς που ανήκουν σε μία άτρακτο
 - Τα δεδομένα μεταφέρονται μεταξύ δίσκου και κύριας μνήμης σε μπλοκ
 - Το μέγεθος ποικίλλει από 512 bytes μέχρι μερικά kilobytes
 - Μικρά blocks: περισσότερες προσπελάσεις στον δίσκο
 - Μεγάλα blocks: περισσότερος χώρος σπαταλιέται εξαιτίας των μερικώς γεμισμένων blocks
 - Τυπικά το μέγεθος ενός block κυμαίνεται μεταξύ 4 και 16 kilobytes
- Αλγόριθμοι χρονοπρογραμματισμού του μηχανικού βραχίονα της κεφαλής (**Disk-arm-scheduling** algorithms) μπορεί να καθυστερήσουν την προσπέλαση συγκεκριμένων ατράκτων ώστε να ελαχιστοποιηθεί η μηχανική κίνηση του βραχίονα
 - **elevator algorithm** : μετακίνησε τον μηχανικό βραχίονα στην μία κατεύθυνση (από τις εξωτερικές προς τις εσωτερικές ατράκτους και αντίστροφα) και επεξεργάσου την επόμενη αίτηση προς αυτή την κατεύθυνση, έπειτα αντίστρεψε την κατεύθυνση κίνησης και επανάλαβε

Βελτιστοποίηση πρόσβασης σε μπλοκ του δίσκου (συνέχεια.)

- **Οργανώσεις αρχείων** – Βελτιστοποίησε τον χρόνο προσπέλασης των μπλοκ με την οργάνωση των μπλοκ ώστε να αντιστοιχούν στην σειρά προσπέλασης των δεδομένων
 - Π.χ. Αποθήκευσε συσχετιζόμενες πληροφορίες στον ίδιο ή σε γειτονικούς κυλίνδρους.
 - Τα αρχεία μπορεί να «κατακερματιστούν» (get **fragmented**) με τον καιρό
 - Π.χ. εάν τα δεδομένα εισαχθούν / διαγραφούν από το αρχείο
 - Ή εάν τα ελεύθερα μπλοκ στον δίσκο είναι διασκορπισμένα, τα νέα αρχεία που δημιουργούνται θα αποτελούνται από μπλοκ που θα είναι διασκορπισμένα στον δίσκο
 - Σειριακή προσπέλαση των μπλοκ ενός κατακερματισμένου αρχείου επιφέρει μεγαλύτερη καθυστέρηση προσπέλασης εξαιτίας της αυξημένης κίνησης του μηχανικού βραχίονα
 - Κάποια υπολογιστικά συστήματα διαθέτουν εφαρμογές που αναλαμβάνουν να ανασυγκροτήσουν το σύστημα αρχείων, ώστε να επιταχύνουν την προσπέλαση δεδομένων

Βελτιστοποίηση πρόσβασης σε μπλοκ του δίσκου (συνέχεια.)

- **Non Volatile (Μόνιμοι) write buffers** επιταχύνουν την εγγραφή δεδομένων στον δίσκο με το να αποθηκεύουν αμέσως μπλοκ δεδομένων σε non-volatile RAM buffer:
 - Ο ελεγκτής στην συνέχεια αποθηκεύει τα δεδομένα στον δίσκο όταν δεν υπάρχουν άλλες αιτήσεις προς στον δίσκο ή οι αιτήσεις για τον δίσκο εκκρεμούν για κάποιο διάστημα
 - Εργασίες της βάσης δεδομένων που απαιτούν τα δεδομένα να αποθηκευθούν σε κάποιο ενδιάμεσο στάδιο, μπορούν να συνεχίσουν την εκτέλεσή τους χωρίς να περιμένουν να γραφτούν τα δεδομένα στον δίσκο
 - *Οι εγγραφές μπορούν να επαναδιαταχθούν ώστε να ελαχιστοποιηθούν οι κινήσεις του μηχανικού βραχίονα*
- **Δίσκος μητρώου (Log disk)** – Ένας δίσκος αφιερωμένος στο να καταγράφει το ιστορικό των ανανεώσεων των μπλοκ δεδομένων
 - Χρησιμοποιείται σαν RAM μόνιμης αποθήκευσης NV-RAM
 - Η αποθήκευση σε δίσκους μητρώου (log disk) είναι πολύ γρήγορη διαδικασία καθώς δεν απαιτούνται καθόλου αναζητήσεις στον δίσκο
- Το Σύστημα Αρχείων τυπικά αναδιατάσσει τις εγγραφές στον δίσκο ώστε να βελτιωθεί η απόδοση
 - **Journaling file systems** καταγραφή δεδομένων σε ασφαλή διάταξη σε NV-RAM ή σε log disk
 - Αναδιάταξη χωρίς journaling: κίνδυνος απώλειας δεδομένων

Ιεραρχία μονάδων αποθήκευσης

Ταχύτητα
Κόστος



- **Κρυφή μνήμη**
- **Κύρια μνήμη** – Άμεση προσπέλαση, Ευμετάβλητη αποθήκευση
- **Μαγνητικοί δίσκοι** – Άμεση προσπέλαση, μόνιμη αποθήκευση
- **Οπτικοί Δίσκοι / juke-boxes** – Άμεση προσπέλαση, μόνιμη αποθήκευση
- **Μαγνητικές ταινίες / tape juke-boxes** – σειριακή προσπέλαση, μόνιμη αποθήκευση



RAID

■ Redundant Arrays of Independent Disks (RAID)

- Τεχνικές οργάνωσης δίσκων που διαχειρίζονται μεγάλο αριθμό δίσκων και τους παρουσιάζουν σαν μια ενιαία μονάδα δίσκου, η οποία παρέχει
 - Μεγάλη χωρητικότητα και υψηλές ταχύτητες, χρησιμοποιώντας πολλαπλούς δίσκους παράλληλα και υψηλή αξιοπιστία αποθηκεύοντας πλεονάζοντα δεδομένα.

■ Η πιθανότητα να εμφανιστεί δυσλειτουργία σε έναν δίσκο από το σύνολο των N δίσκων είναι μεγαλύτερη από την πιθανότητα να εμφανιστεί δυσλειτουργία σε έναν συγκεκριμένο δίσκο

- Χρήση πολλαπλών δίσκων (redundancy) ώστε να αποτρέψουμε απώλεια δεδομένων

■ Μια οικονομική εναλλακτική στους μεγάλους και ακριβούς δίσκους

- Το **I** στο ακρώνυμο RAID αρχικά αντιστοιχούσε στον όρο “**inexpensive - χαμηλού κόστους**”
- Σήμερα οι δίσκοι RAID χρησιμοποιούνται για την μεγαλύτερη αξιοπιστία τους και τον μεγαλύτερο ρυθμό μετάδοσης
 - “**I**” → “**independent - ανεξάρτητος**”

Βελτίωση της αξιοπιστίας με χρήση πλεονασμού (Redundancy)

- **Redundancy-Πλεονασμός** – αποθήκευση πλεονάζουσας πληροφορίας η οποία μπορεί να χρησιμοποιηθεί για την ανάκτηση δεδομένων που χάθηκαν λόγω σφάλματος δίσκου.
- Π.χ., **Mirroring** (or **shadowing**) Κατοπτρισμός
 - Δημιούργησε αντίγραφο κάθε δίσκου· κάθε λογικός δίσκος (logical disk) αποτελείται από δύο φυσικούς δίσκους (physical disks)
 - Κάθε εγγραφή καταγράφεται στην επιφάνεια και των δύο δίσκων
 - Εάν υπάρξει σφάλμα στον ένα δίσκο, τα δεδομένα θα είναι διαθέσιμα στον άλλο
 - Απώλεια δεδομένων μόνο αν και τα δύο έχουν ταυτόχρονα σφάλμα
- Ο μέσος χρόνος απώλειας δεδομένων (Mean time to data loss) εξαρτάται από το μέσο χρόνο παρουσίασης σφάλματος και τον μέσο χρόνο επισκευής
 - Π.χ. MTTF 100,000 ωρών και μέσος χρόνος επισκευής 10 ωρών σημαίνει μέσος χρόνος απώλειας δεδομένων $500 \cdot 10^6$ ωρών (ή 57,000 χρόνια) για ένα «mirrored» ζεύγος δίσκων - μη λαμβάνοντας υπόψη εξαρτημένες καταστάσεις σφαλμάτων (dependent failure modes)

Βελτίωση της απόδοσης με χρήση παραλληλισμού

- Δύο σημαντικοί στόχοι παραλληλισμού σε συστήματα δίσκων:
 1. Εξισορρόπηση φορτίου πολλαπλών μικρών προσπελάσεων για αύξηση του throughput
 2. Παραλληλοποίηση μεγάλες προσπελάσεις ώστε να μειωθεί ο χρόνος απόδοσης
- Βελτίωση ρυθμού μετάδοσης με διαμοίραση δεδομένων σε πολλαπλούς δίσκους
- **Διαχωρισμό σε επίπεδο Bit** – Χώρισε κάθε bit ενός byte σε πολλαπλούς δίσκους
 - Σε ένα πίνακα 8 δίσκων, κατέγραψε το bit i κάθε byte στον δίσκο i
 - Σε κάθε προσπέλαση μπορούν να διαβαστούν τα δεδομένα 8 φορές πιο γρήγορα απ' ό τι σε έναν απλό δίσκο
 - ... αλλά ο χρόνος αναζήτησης/προσπέλασης είναι χειρότερος από τον χρόνο ενός απλού δίσκου -> Ο διαχωρισμός σε επίπεδο Bit δεν χρησιμοποιείται πλέον
- **Διαχωρισμό σε επίπεδο Block** – με n δίσκους, το block i του αρχείου εγγράφεται στον δίσκο $(i \bmod n) + 1$
 - Αιτήσεις για διαφορετικά blocks μπορούν να εξυπηρετούνται παράλληλα, εάν τα blocks ανήκουν σε διαφορετικούς δίσκους
 - Σε μία αίτηση για μια μεγάλη ακολουθία blocks μπορεί να χρησιμοποιηθούν παράλληλα όλοι οι διαθέσιμοι δίσκοι

Επίπεδο RAID

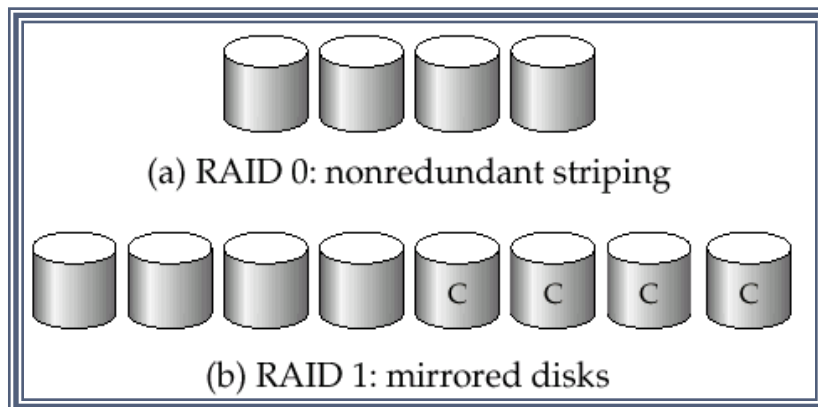
- Σχήματα που παρέχουν πλεονασμό με χαμηλότερο κόστος συνδυάζοντας διαχωρισμό δίσκων μαζί με Bit ισοτιμίας - **Διαφορετικές οργανώσεις RAID**, ή **επίπεδα RAID**, έχουν διαφορετική απόδοση και χαρακτηριστικά αξιοπιστίας
- **RAID Επίπεδο 0:** Χωρίς πλεονασμό (non-redundant) με χρήση διαχωρισμού block (block striping)

☞ Χρησιμοποιείται σε υψηλής απόδοσης εφαρμογές όπου η απώλεια δεδομένων δεν θεωρείται σημαντική

- **RAID Επίπεδο 1:** Κατοπτρισμός (Mirrored disks) με χρήση διαχωρισμού block

☞ Προσφέρει βέλτιστη απόδοση εγγραφής

☞ Δημοφιλής σε εφαρμογές αποθήκευσης ιστορικού όπως αποθήκευση αρχείων ιστορικού (log files) σε βάσεις δεδομένων



Επίπεδα RAID (συνέχεια.)

- **RAID Επίπεδο 3: Bit-Interleaved Parity - Χρήση δίσκου ισοτιμίας**
 - Ένα bit ισοτιμίας είναι αρκετό για διόρθωση σφαλμάτων, όχι απλά για ανίχνευση, καθώς γνωρίζουμε σε ποιον δίσκο συνέβη το σφάλμα
 - Κατά την καταγραφή δεδομένων, θα πρέπει να υπολογιστούν τα αντίστοιχα bit ισοτιμίας και να καταγραφούν στον δίσκο ισοτιμίας.
 - Για την ανάκτηση δεδομένων από έναν κατεστραμμένο δίσκο υπολογίζουμε το αποκλειστικό-Ή (XOR) των bits από τους υπόλοιπους δίσκους (συμπεριλαμβάνουμε τους δίσκους ισοτιμίας).
 - Ταχύτερη μεταφορά δεδομένων σε σχέση με την χρήση απλού δίσκου αλλά λιγότερες προσπελάσεις I/O ανά δευτερόλεπτο καθώς κάθε δίσκος θα πρέπει να συνεισφέρει σε κάθε I/O.
 - Υπερκεράζει το επίπεδο 2 (παρέχει όλα τα πλεονεκτήματά του με μικρότερο κόστος).



(d) RAID 3: bit-interleaved parity

Επίπεδα RAID (συνέχεια.)

- **RAID Επίπεδο 5:** Διαχωρισμοί επιπέδου μπλοκ με κατανομή ισοτιμίας στους δίσκους (Block-Interleaved Distributed Parity) Διαμοιράζει τα δεδομένα και τα bits ισοτιμίας σε όλους τους $N + 1$ δίσκους, αντί να αποθηκεύει τα δεδομένα σε N δίσκους και τα bit ισοτιμίας σε έναν δίσκο.
 - Π.χ., με 5 δίσκους, το block ισοτιμίας για το n -ιστό σύνολο των block αποθηκεύεται στο δίσκο $(n \bmod 5) + 1$, με τα block δεδομένων να αποθηκεύονται στους υπόλοιπους 4 δίσκους



(f) RAID 5: block-interleaved distributed parity

P0	0	1	2	3
4	P1	5	6	7
8	9	P2	10	11
12	13	14	P3	15
16	17	18	19	P4

Επίπεδα RAID (συνέχεια.)

■ RAID Επίπεδο 5 (Συνέχεια)

- Υψηλότεροι ρυθμοί I/O από το 4 επίπεδο.
 - Η εγγραφή του block συμβαίνει παράλληλα εάν το block δεδομένων και τα block ισοτιμίας βρίσκονται σε διαφορετικούς δίσκους
- Υπερκεράζει το επίπεδο 4: Παρέχει ίδια πλεονεκτήματα αποφεύγοντας καταστάσεις bottleneck των δίσκων ισοτιμίας

■ RAID Επίπεδο 6: P+Q σχήμα πλεονασμού :

- Παρόμοιο με το επίπεδο 5, με την διαφορά ότι αποθηκεύει επιπλέον πλεονάζουσα πληροφορία ώστε να διασφαλίσει τα δεδομένα από πολλαπλά σφάλματα δίσκων.
- Καλύτερη αξιοπιστία από το επίπεδο 5 με μεγαλύτερο κόστος.
- Δεν χρησιμοποιείται ευρέως.



(g) RAID 6: P + Q redundancy



Επιλογή επιπέδου RAID

- Το επίπεδο 0 χρησιμοποιείται μόνο όταν η ασφάλεια των δεδομένων δεν είναι σημαντική, Π.χ. Τα δεδομένα μπορούν να συλλεχθούν γρήγορα από άλλες πηγές
- Τα επίπεδα 2 και 4 δεν χρησιμοποιούνται καθώς έχουν καλυφθεί από τα επίπεδα 3 και 5
- Το επίπεδο 3 δεν χρησιμοποιείται πλέον καθώς η *τεχνική διαχωρισμού σε επίπεδο Bit* προκαλεί αναγνώσεις απλών Block σε όλους τους δίσκους απαιτώντας την κίνηση του μηχανικού βραχίονα (κόστος σε χρόνο), γεγονός που αποφεύγεται με την *τεχνική διαχωρισμού σε επίπεδο Block* (επίπεδο 5).
- Το επίπεδο 6 χρησιμοποιείται σπάνια καθώς τα επίπεδα 1 και 5 παρέχουν επαρκή επίπεδα ασφάλειας για σχεδόν κάθε εφαρμογή.
- Η σύγκριση λοιπόν γίνεται μεταξύ των επιπέδων 1 και 5
 - Το επίπεδο 5 προτιμάται για εφαρμογές με
 - χαμηλό ρυθμό ανανέωσης και μεγάλο πλήθος δεδομένων
 - Το επίπεδο 1 προτιμάται για όλες τις υπόλοιπες εφαρμογές



Θέματα υλικού

- **Software RAID:** υλοποιήσεις RAID εξολοκλήρου στο λογισμικό
- **Hardware RAID:** υλοποιήσεις RAID με χρήση ειδικού υλικού
 - Χρήση μόνιμης NV RAM για καταγραφή εγγραφών που εκτελούνται
- **Hot swapping:** Αντικατάσταση δίσκων ενώ το σύστημα βρίσκεται σε λειτουργία και χωρίς διακοπή παροχής ρεύματος
 - Μειώνει τον χρόνο ανάνηψης και βελτιώνει σημαντικά την διαθεσιμότητα
- Πολλά συστήματα διατηρούν εφεδρικούς δίσκους (**spare disks**) οι οποίοι είναι άμεσα διαθέσιμοι και χρησιμοποιούνται σε περίπτωση ανίχνευσης σφάλματος
 - Σημαντική μείωση του χρόνου ανάνηψης

Μαγνητικές Ταινίες

- Διατηρούν μεγάλο όγκο δεδομένων και παρέχουν υψηλούς ρυθμούς μετάδοσης
 - Μερικά GB για το πρότυπο **DAT** (Digital Audio Tape), 10-40 GB για το πρότυπο **DLT** (Digital Linear Tape), 100 GB+ για το πρότυπο **Ultrium**, και 330 GB για το πρότυπο **Ampex helical scan**
 - Ο ρυθμός μετάδοσης κυμαίνεται από μερικά MB έως μερικές δεκάδες MB ανά δευτερόλεπτο
- Μέχρι σήμερα το φθηνότερο μέσο αποθήκευσης.
 - Οι ταινίες είναι φθηνές αλλά το κόστος των οδηγών ταινιών είναι πού υψηλό
- Πολύ μεγαλύτερος χρόνος προσπέλασης σε σχέση με τους μαγνητικούς και οπτικούς δίσκους
 - Περιορίζεται εξαιτίας της σειριακής προσπέλασης των δεδομένων
- Χρησιμοποιείται κυρίως για εφεδρική αποθήκευση (Backup), για αποθήκευση πληροφορίας που δεν χρησιμοποιείται συχνά και σαν ενδιάμεσο μέσο μεταφοράς πληροφορίας από ένα σύστημα σε ένα άλλο
- Jukeboxes μαγνητικών ταινιών χρησιμοποιούνται για πολύ μεγάλη χωρητικότητα αποθήκευσης (TBs, PBs)



Αποθήκευση στον δίσκο

- Ένα αρχείο βάσης δεδομένων χωρίζεται σε δεδομένου μήκους μονάδες αποθήκευσης (**blocks**). Τα Blocks είναι μονάδες τόσο για την ανάθεση αποθηκευτικού χώρου όσο και για την μεταφορά δεδομένων
- Το σύστημα διαχείρισης μιας βάσης δεδομένων προσπαθεί να **ελαχιστοποιήσει τον αριθμό των μπλοκ που μεταφέρονται** μεταξύ του δίσκου και της μνήμης
- Μειώνουμε το πλήθος των προσπελάσεων στον δίσκο διατηρώντας όσο το δυνατόν περισσότερα blocks στην κύρια μνήμη
- **Buffer** – το μέρος της κύριας μνήμης που διατίθεται για την αποθήκευση αντιγράφων από blocks του δίσκου
- **Buffer manager – Διαχειριστής buffer** – υποσύστημα το οποίο είναι υπεύθυνο για την ανάθεση του χώρου Buffer της κύριας μνήμης



Διαχείριση Buffer

- Οι εφαρμογές απευθύνονται στον διαχειριστή buffer όταν χρειάζονται κάποιο block του δίσκου
 1. Εάν το **block βρίσκεται ήδη στον buffer**, η κατάσταση είναι εύκολη
 2. Εάν το **block δεν βρίσκεται ήδη στον buffer** τότε
 1. Ο Διαχειριστής Buffer αναθέτει χώρο buffer για το μπλοκ, και αντικαθιστά κάποια άλλα blocks εφόσον παραστεί ανάγκη
 2. Το μπλοκ που αντικαθίσταται στον χώρο buffer ξαναγράφεται στον δίσκο, μόνο εάν έχει προηγουμένως τροποποιηθεί από την τελευταία φορά που μεταφέρθηκε από τον δίσκο στην κύρια μνήμη
 3. Εφόσον ο χώρος έχει δεσμευτεί στον buffer, ο διαχειριστής buffer εκκινεί την διαδικασία μεταφοράς του Block από τον δίσκο στον buffer

Πολιτικές αντικατάστασης Buffer

- Τα περισσότερα λειτουργικά συστήματα αντικαθιστούν το Block που χρησιμοποιείται λιγότερο συχνά - **least recently used** (LRU strategy)
- Τα ερωτήματα έχουν καλά καθορισμένα πρότυπα πρόσβασης (όπως σειριακή σάρωση), και ένα ΣΔΒΔ μπορεί να χρησιμοποιήσει την πληροφορία αυτή κατά την εξυπηρέτηση ενός ερωτήματος του χρήστη ώστε να προβλέψει μελλοντικές προσπελάσεις στον δίσκο
 - LRU μπορεί να είναι κακή στρατηγική για συγκεκριμένα πρότυπα πρόσβασης συμπεριλαμβανομένου επαναλαμβανόμενων σαρώσεων δεδομένων
 - Π.χ. Όταν υπολογίζουμε την συνένωση (join) δύο σχέσεων r και s με εμφωλευμένες επαναλήψεις (nested loops)

```
for each tuple  $tr$  of  $r$  do
  for each tuple  $ts$  of  $s$  do
    if the tuples  $tr$  and  $ts$  match ...
```
 - Προτιμώνται οι συνδυασμένες στρατηγικές με έξυπνες επινοήσεις για την βελτίωση των στρατηγικών αντικατάστασης που παρέχονται από την μονάδα βελτιστοποίησης ερωτημάτων

Πολιτικές αντικατάστασης Buffer

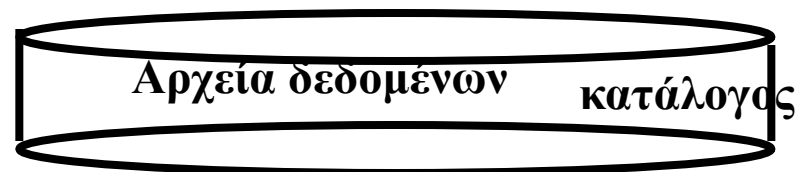
(συνέχεια.)

- **Pinned block** – block στην κύρια μνήμη που δεν επιτρέπονται να επανεγραφούν στον δίσκο
- **Στρατηγική Toss-immediate** – απελευθερώνει χώρο που καταλαμβάνεται από ένα block μόλις επεξεργασθεί η τελευταία εγγραφή του συγκεκριμένου Block
- Στρατηγική **Most Recently Used (MRU)** – το σύστημα μαρκάρει (pin) το block που επεξεργάζεται αυτή την στιγμή. Μόλις η τελευταία εγγραφή του συγκεκριμένου υποστεί επεξεργασία το block ξεμαρκάρεται (unpinned) και θεωρείται το πιο πρόσφατα χρησιμοποιημένο block.
- Ο διαχειριστής Buffer μπορεί να χρησιμοποιήσει στατιστικές πληροφορίες σχετικά με την πιθανότητα μια αίτηση να αναφέρεται σε συγκεκριμένη σχέση
 - Π.χ. τα δεδομένα του ευρετηρίου (data dictionary) προσπελούνται συχνά.
Ευριστική μέθοδος: κράτησε τα δεδομένα ευρετηρίου στον buffer της κύριας μνήμης
- Οι διαχειριστές Buffer υποστηρίζουν επιπλέον εξαναγκασμένη έξοδο (**forced output**) blocks σε περιπτώσεις ανάνηψης από σφάλματα

Οργανώσεις Αρχείων

Π.χ., εγγραφές «Φοιτητών» – πως θα αποθηκευθούν στον δίσκο;

Φοιτητής		
<u>ΑΜ</u>	Όνομα	Διεύθυνση
123	Σταύρου	Αιόλου
234	Αντωνίου	Θράκης





Οργανώσεις Αρχείων

- Η Βάση Δεδομένων αποθηκεύεται σαν μια συλλογή από *αρχεία*. Κάθε αρχείο είναι μια σειρά από *records*. Κάθε record είναι μια σειρά από *fields*.
 - Μια προσέγγιση:
 - Υπόθεση ότι το μέγεθος του record είναι **σταθερό**
 - Κάθε αρχείο έχει records ενός συγκεκριμένου τύπου μόνο
 - Διαφορετικά αρχεία χρησιμοποιούνται για διαφορετικές σχέσεις
- Αυτή η περίπτωση είναι πιο εύκολη για υλοποίηση – θα εξετάσουμε records μεταβλητού μήκους αργότερα.



Εγγραφές σταθερού μήκους

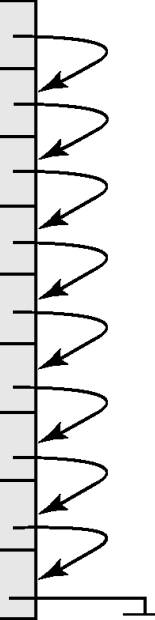
- Λύση #1: 'Σωρός' (= χωρίς διάταξη)
- Λύση #2: Σειριακά (ταξινομημένα αρχεία)

123	Σταύρου	Αιόλου
234	Αντωνίου	Θράκης

Σειριακές οργανώσεις αρχείων

- Κατάλληλες για εφαρμογές που απαιτούν σειριακή επεξεργασία ολόκληρων αρχείων
- Οι εγγραφές ταξινομούνται με βάση το κλειδί αναζήτησης (search-key)

A-217	Brighton	750	
A-101	Downtown	500	
A-110	Downtown	600	
A-215	Mianus	700	
A-102	Perryridge	400	
A-201	Perryridge	900	
A-218	Perryridge	700	
A-222	Redwood	700	
A-305	Round Hill	350	



Εγγραφές σταθερού μήκους (συνέχεια)

- Λύση #1: 'Σωρός' (= χωρίς διάταξη)
- Λύση #2: Σειριακά
 - **Αλλά: Διαγραφές? Εισαγωγές?**

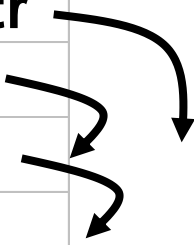
123	Σταυρου	Αιόλου
234	Αντωνίου	Θράκης

Εγγραφές σταθερού μήκους (συνέχεια)

- Σειριακά
- Διαγραφές? Εισαγωγές?

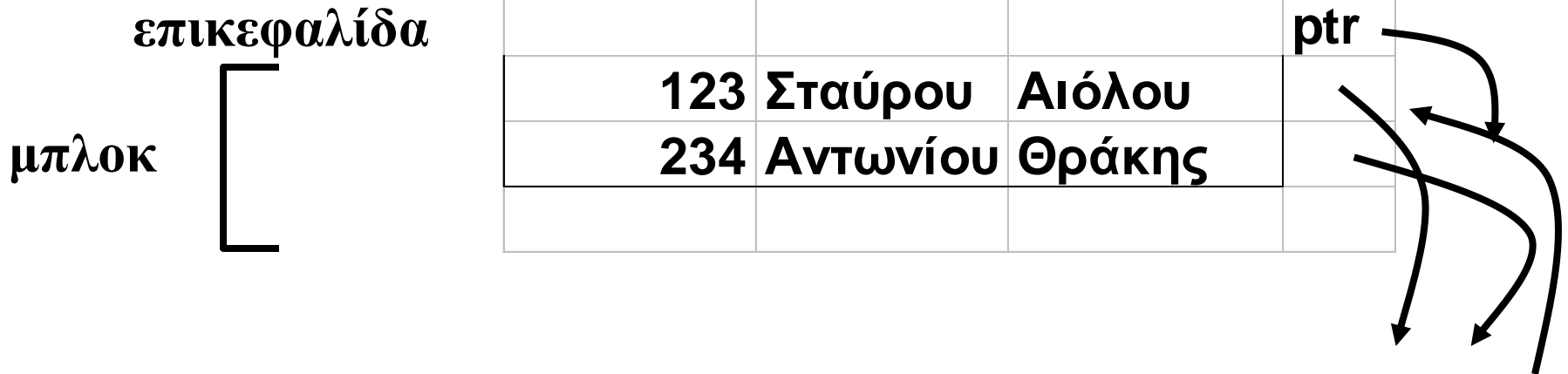
**Επικεφαλίδα
(header)**

			ptr
123	Σταύρου	Αιόλου	
234	Αντωνίου	Θράκης	

The diagram shows three curved arrows originating from the 'ptr' column of the table. The top arrow points to the first data row (123 Σταύρου Αιόλου), the middle arrow points to the second data row (234 Αντωνίου Θράκης), and the bottom arrow points to the empty row below.

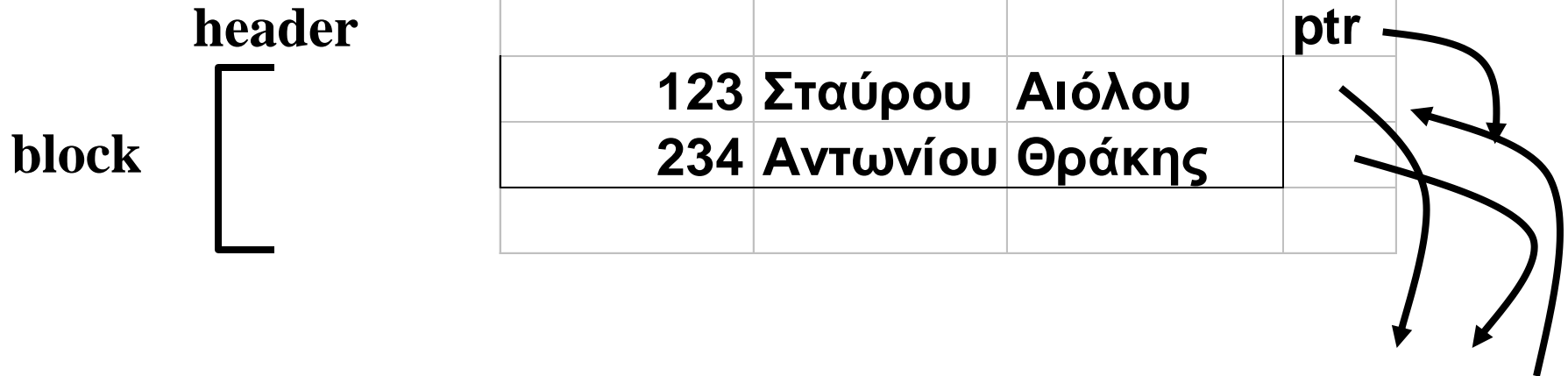
Εγγραφές σταθερού μήκους (συνέχεια)

Προβλήματα?



Εγγραφές σταθερού μήκους (συνέχεια)

Προβλήματα? Δείκτες διασχίζουν όρια των μπλοκ – Αργή σειριακή σάρωση!



Εγγραφές σταθερού μήκους (συνέχεια)

- Απλή προσέγγιση:
 - Αποθήκευσε την εγγραφή i αρχίζοντας από το byte $n * (i - 1)$, όπου n το μέγεθος της κάθε εγγραφής
 - Η προσπέλαση των εγγραφών είναι εύκολη αλλά οι εγγραφές μπορεί να διαμοιράζονται μεταξύ μπλοκ -> Τροποποίηση: Να μην επιτρέπεται να εκτείνονται οι εγγραφές πέρα από τα όρια ενός μπλοκ
- Διαγραφή της εγγραφής i :
εναλλακτικές λύσεις:
 - Μετακίνησε τις εγγραφές από τις θέσεις $i + 1, \dots, n$ στις θέσεις $i, \dots, n - 1$
 - Μετακίνησε την εγγραφή της θέσης n στην θέση i
 - Να μην γίνει μετακίνηση εγγραφών, αλλά να συνδεθούν όλες οι ελεύθερες εγγραφές σε μία *free list*

record 0	A-102	Perryridge	400
record 1	A-305	Round Hill	350
record 2	A-215	Mianus	700
record 3	A-101	Downtown	500
record 4	A-222	Redwood	700
record 5	A-201	Perryridge	900
record 6	A-217	Brighton	750
record 7	A-110	Downtown	600
record 8	A-218	Perryridge	700

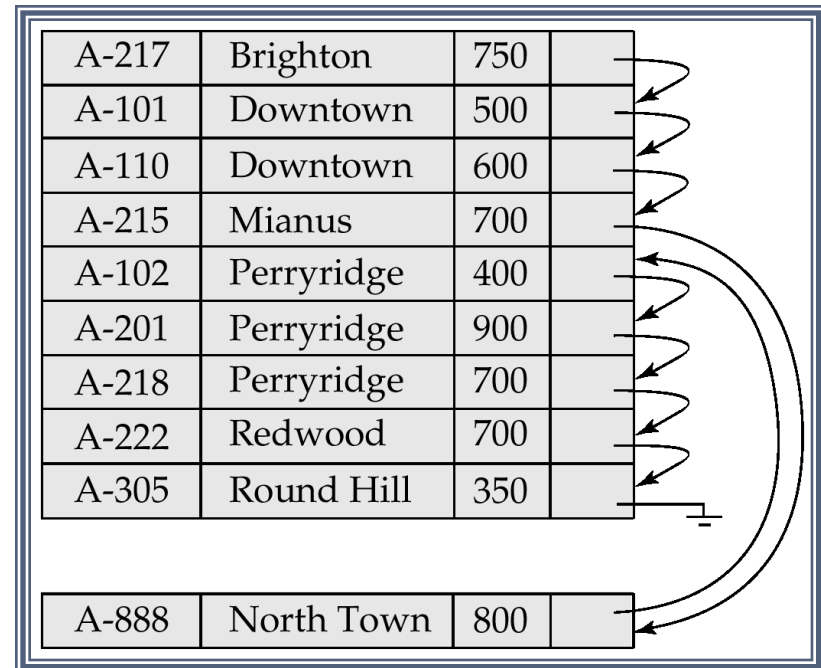
Συνδεδεμένη τοποθέτηση (Free Lists)

- Αποθήκευσε την διεύθυνση της πρώτης διαγραμμένης εγγραφής στην επικεφαλίδα του αρχείου.
- Χρησιμοποίησε την πρώτη εγγραφή για να αποθηκεύσεις την διεύθυνση της δεύτερης διαγραμμένης εγγραφής κ.ο.κ
- Αποθήκευσε διευθύνσεις σαν δείκτες (pointers) καθώς δείχνουν στην τοποθεσία που είναι αποθηκευμένη η εγγραφή.
- Αποτελεσματικότερη αναπαράσταση (more space efficient representation): επαναχρησιμοποίηση χώρου για κανονικά γνωρίσματα κενών εγγραφών για την αποθήκευση δεικτών

header				
record 0	A-102	Perryridge	400	
record 1				
record 2	A-215	Mianus	700	
record 3	A-101	Downtown	500	
record 4				
record 5	A-201	Perryridge	900	
record 6				
record 7	A-110	Downtown	600	
record 8	A-218	Perryridge	700	

Σκέψεις – σειριακές οργανώσεις αρχείων

- Διαγραφή – χρησιμοποίησε αλυσίδες δεικτών
- Εισαγωγή – Εντόπισε την θέση όπου οι εγγραφές πρόκειται να καταγραφούν
 - Εάν υπάρχει κενός χώρος να γίνει η εισαγωγή της εγγραφής εδώ
 - Εάν δεν υπάρχει κενός χώρος, η εισαγωγή της εγγραφής να γίνει σε ένα block υπερχείλισης (*overflow block*)
 - Σε κάθε περίπτωση οι αλυσίδες δεικτών πρέπει να ενημερώνονται
- Απαιτείται η επαναδιοργάνωση του αρχείου κατά καιρούς ώστε να διατηρείται η σειριακή σειρά των εγγραφών



Εγγραφές μεταβλητού μήκους

Π.χ., με τα πεδία VARCHAR:

Address VARCHAR(100).

Λύσεις?

block



			ptr
123	Σταύρου	Αιόλου	
234	Αντωνίου	Θράκης	



Εγγραφές μεταβλητού μήκους

- Μεταβλητού μήκους εγγραφές (Variable-length records) προκύπτουν σε διάφορες περιπτώσεις σε συστήματα διαχείρισης βάσεων δεδομένων :
 - Αποθήκευση πολλαπλών τύπων εγγραφών σε ένα αρχείο
 - Τύποι εγγραφών που περιέχουν ένα ή περισσότερα πεδία μεταβλητού μήκους
 - Τύποι εγγραφών που επιτρέπουν επαναλαμβανόμενα πεδία (σε χρήση σε κάποια παλαιότερα μοντέλα δεδομένων)
- Αναπαράσταση **Byte string**
 - Προσάρτηση ένα *end-of-record* (\perp) χαρακτήρα ελέγχου στο τέλος κάθε εγγραφής



Εγγραφές μεταβλητού μήκους

- Ρεύματα Byte – Byte streams (δομή slotted page!)
- Σταθερό μήκος (padding, overflow)



Εγγραφές μεταβλητού μήκους

- Ρεύματα Byte : σύμβολο end-of-record
- Χρησιμοποιείται σπάνια (Γιατί;)

123,Σταύρου,Αιόλου EOR 34,Αντωνίου,Θράκης EOR



Εγγραφές μεταβλητού μήκους

- Ρεύματα Byte : σύμβολο end-of-record
- Χρησιμοποιείται σπάνια (Γιατί;)

123,Σταύρου,Αιόλου EOR 34,Αντωνίου,Θράκης EOR

- Δυσκολία με διαγραφές
- Δυσκολία με ανάπτυξη



Εγγραφές μεταβλητού μήκους

(συνέχεια.)

- Αναπαραστάσεις σταθερού μήκους – Πως?



Εγγραφές μεταβλητού μήκους

(συνέχεια.)

Αναπαραστάσεις σταθερού μήκους – Πως?

Padding

Anchor/overflow

Εγγραφές μεταβλητού μήκους

(συνέχεια.)

- Αναπαράσταση σταθερού μήκους:
 - Αφιερωμένος χώρος (Reserved space)
 - Δείκτες
- Reserved space
 - Μπορεί να χρησιμοποιηθούν εγγραφές σταθερού μήκους μιας γνωστής μέγιστης τιμής
 - Ο αχρησιμοποίητος χώρος γεμίζεται με τιμές Null ή σύμβολα end-of-record.

0	Perryridge	A-102	400	A-201	900	A-218	700
1	Round Hill	A-305	350	⊥	⊥	⊥	⊥
2	Mianus	A-215	700	⊥	⊥	⊥	⊥
3	Downtown	A-101	500	A-110	600	⊥	⊥
4	Redwood	A-222	700	⊥	⊥	⊥	⊥
5	Brighton	A-217	750	⊥	⊥	⊥	⊥

Χρήση δεικτών

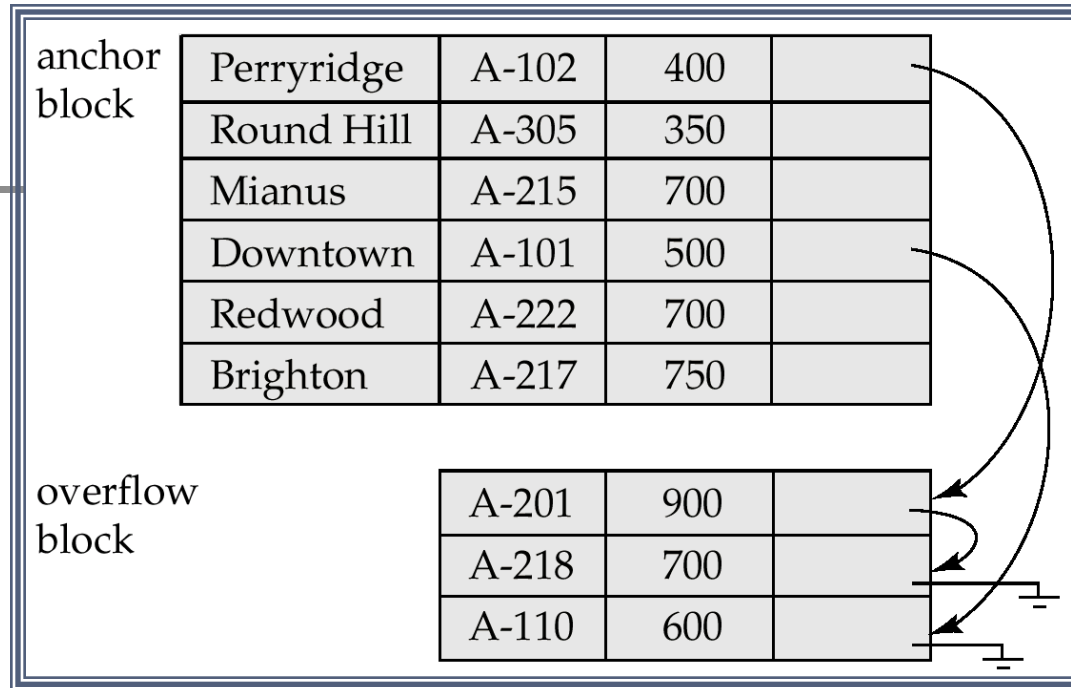
0	Perryridge	A-102	400	
1	Round Hill	A-305	350	
2	Mianus	A-215	700	
3	Downtown	A-101	500	
4	Redwood	A-222	700	
5		A-201	900	
6	Brighton	A-217	750	
7		A-110	600	
8		A-218	700	



■ Pointer method

- Μία εγγραφή μεταβλητού μήκους αναπαρίσταται σαν λίστα από εγγραφές σταθερού μήκους που συνδέονται μεταξύ τους με την χρήση δεικτών
- Μπορεί να χρησιμοποιηθεί ακόμα και αν το μέγιστο μήκος εγγραφών δεν είναι γνωστό

Χρήση δεικτών (συνέχεια.)

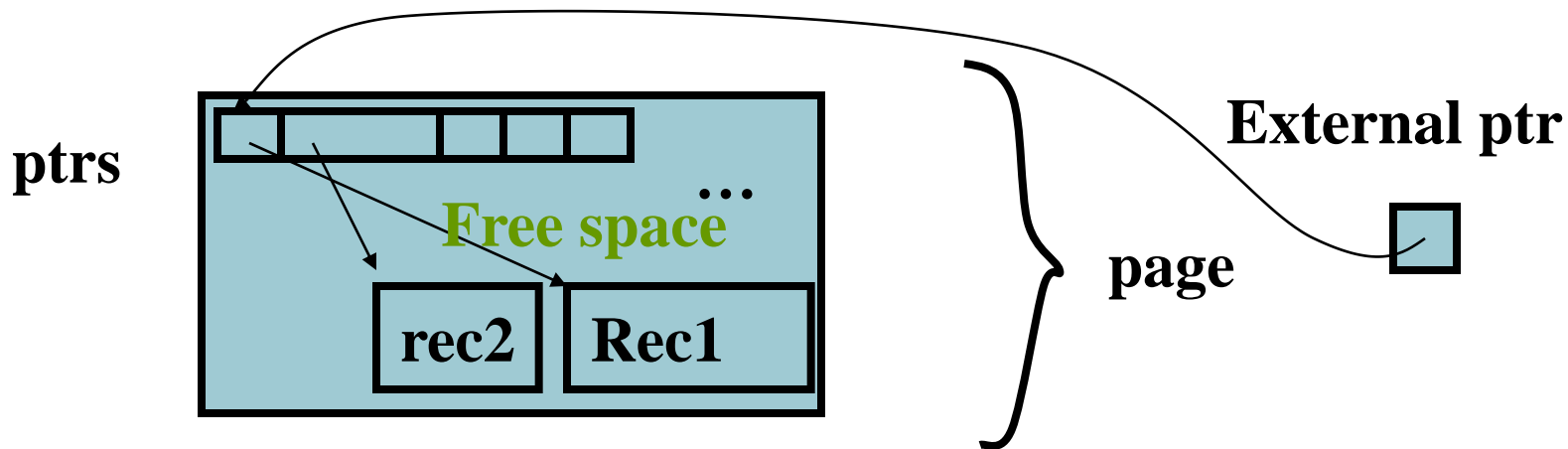


- Μειονεκτήματα της δομής δεικτών
 - Σπατάλη χώρου σε κάθε εγγραφή εκτός από την πρώτη εγγραφή κάθε αλυσίδας
- Λύση: να επιτρέπονται δύο είδη block σε ένα αρχείο:
 - Anchor block – περιέχει την πρώτη εγγραφή της αλυσίδας
 - Overflow block – περιέχει τις υπόλοιπες εγγραφές

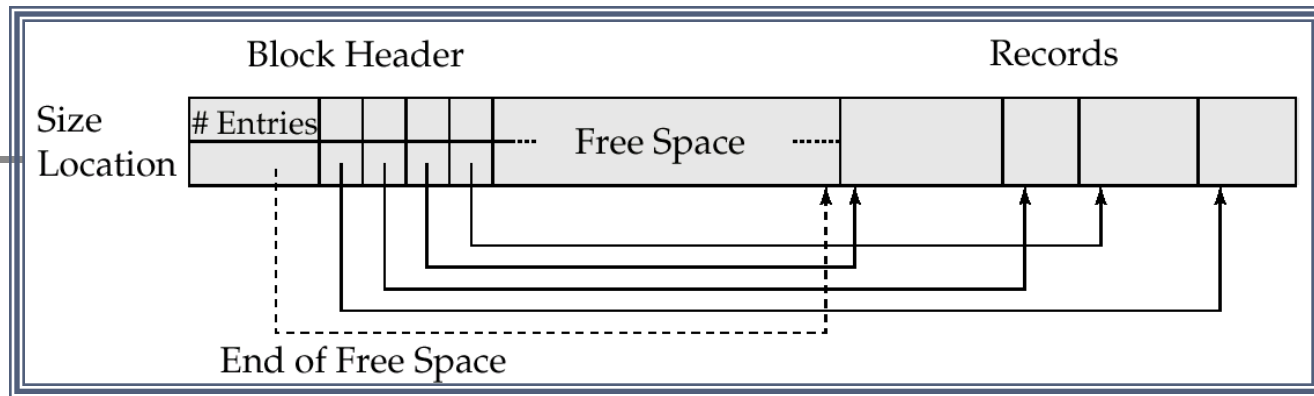
Δομή Slotted page

(Σπουδαία ιδέα – ‘page-aware’!)

- Οι εγγραφές μπορούν να μετακινούνται στα όρια της σελίδας
- Αρχή σελίδας: περιέχει δείκτες
- Εξωτερικοί δείκτες : δείχνουν μόνο σε ‘ptrs’



Εγγραφές μεταβλητού μήκους: Slotted Page Structure



- Slotted page: Η επικεφαλίδα περιέχει:
 - Αριθμό των εισαχθέντων εγγραφών
 - Τέλος του ελεύθερου χώρου στο τέλος του block
 - Θέση και μέγεθος κάθε εγγραφής
- Οι εγγραφές μπορούν να μετακινούνται μέσα σε μία σελίδα ώστε να διατηρούνται συνεχείς, χωρίς ενδιάμεσα κενά διαστήματα · Η καταχώρηση της επικεφαλίδας θα πρέπει να ενημερώνεται
- Οι δείκτες δεν δείχνουν απ' ευθείας σε εγγραφές - αντίθετα δείχνουν στην καταχώρηση για την εγγραφή στην επικεφαλίδα



Οργανώσεις αρχείων

- Σωρός (Heap) (χωρίς διάταξη, ένας πίνακας ανά αρχείο)
- Σειριακή προσπέλαση (Sequential)
- Hashing (Μια συνάρτηση hash υπολογισμένη σε κάποιο γνώρισμα κάθε εγγραφής καθορίζει το μπλοκ στο οποίο θα αποθηκευθεί η εγγραφή)
- Συσταδοποίηση - Clustering (Πολλοί πίνακες ανά αρχείο) – κίνητρο: αποθήκευσε συσχετιζόμενες εγγραφές στο ίδιο μπλοκ ώστε να ελαχιστοποιηθούν οι πράξεις I/O

Οργανώσεις αρχείων συστάδας

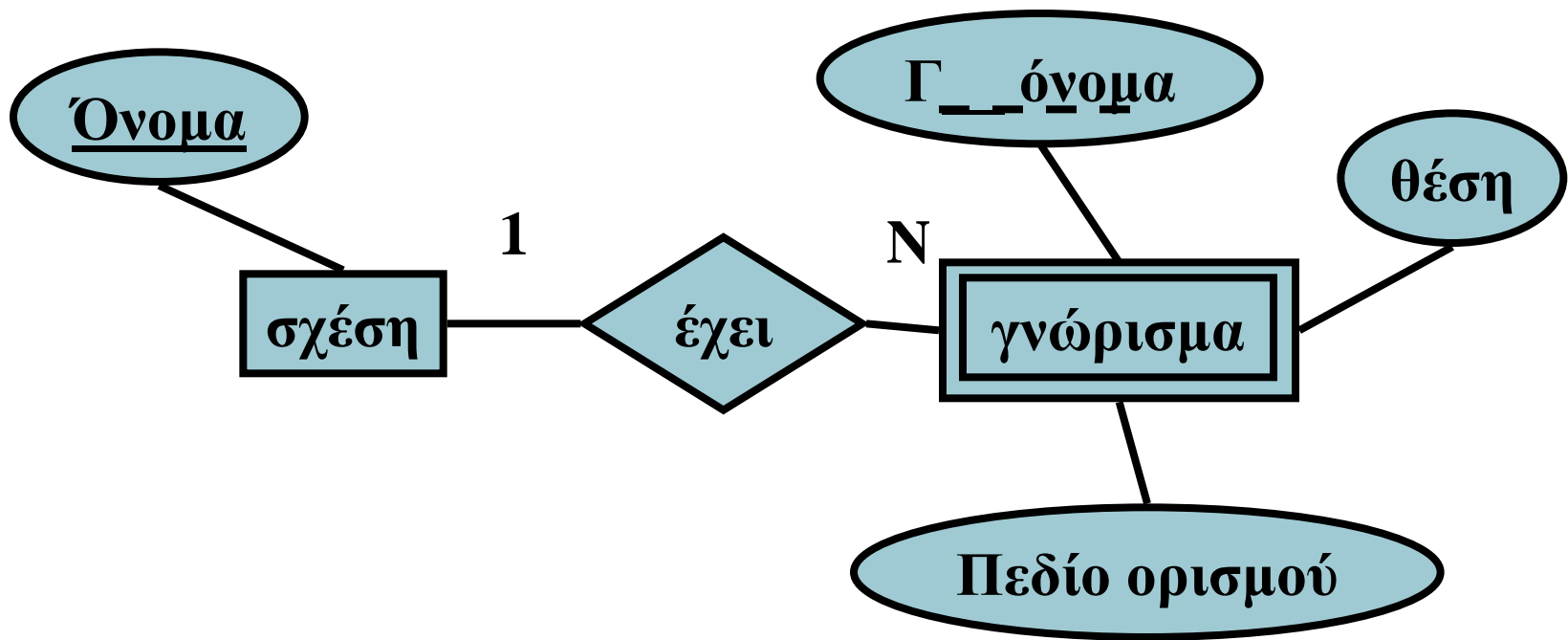
- Αντί να αποθηκεύουμε κάθε σχέση σε ξεχωριστό αρχείο, αποθηκεύουμε ικανό αριθμό σχέσεων σε ένα αρχείο χρησιμοποιώντας οργάνωση συστάδων (**clustering organization**)
- Π.χ., οργάνωση συστάδων πελάτη και καταθέτη:

Hayes	Main	Brooklyn
Hayes	A-102	
Hayes	A-220	
Hayes	A-503	
Turner	Putnam	Stamford
Turner	A-305	

- ☞ Καλό για ερωτήματα που περιλαμβάνουν συνένωση (⋈) καταθέτη πελάτη, και για ερωτήματα που περιλαμβάνουν ένα πελάτη και τους λογαριασμούς του
- ☞ Κακό για ερωτήματα που περιλαμβάνουν μόνο τον πελάτη
- ☞ Έχει ως αποτέλεσμα εγγραφές μεταβλητού μήκους

Αποθήκευση του data dictionary

- Αποθήκευση σαν πίνακες!!
- Πρόκληση: Διάγραμμα E-R?
 - Σχέσεις, γνωρίσματα, πεδία ορισμού
 - Κάθε σχέση έχει όνομα και κάποια γνωρίσματα
 - Κάθε γνώρισμα έχει όνομα, μήκος και πεδίο ορισμού
 - Επίσης, views, περιορισμοί ακεραιότητας, δείκτες
 - Πληροφορίες χρήστη (authorization, κ.α.)
 - Στατιστικά στοιχεία



Αποθήκευση του data dictionary

Πίνακες?

Sys-cat-schema (rel-name, #-attributes)

Att-schema(att-name, rel-name, domain-type, position)

User-schema(u-id, g-id, passwd)

Index-schema(i-name, rel-name, att-name, index-type)

View-schema(v-name, definition)



Συμπεράσματα

- Αποθήκευση και δομές αρχείων

- Χαρακτηριστικά των Δίσκων αποθήκευσης → blocks; Χαμηλή ταχύτητα πρόσβασης
- Τεχνολογία RAID
- Buffering
- Οργανώσεις Αρχείων : 'δομή slotted σελίδων'
- Αποθήκευση λεξικού δεδομένων: ως πίνακες!