
Ενότητα 8

Σχεδιασμός Αρχείων και Βάσεων Δεδομένων

Στόχοι & αντικείμενο ενότητας

- Σχεδιασμός αρχείων
 - Δομή, οργάνωση και προσπέλαση αρχείων
 - Επεξεργασία αρχείων
 - Κριτήρια σχεδιασμού αρχείων
- Σχεδιασμός Βάσεων Δεδομένων
 - Συστήματα Διαχείρισης Βάσεων Δεδομένων
 - Δομές και μοντέλα δεδομένων
 - Προσέγγιση Οντοτήτων-Σχέσεων



Βέλτιστος σχεδιασμός αρχείων και ΒΔ του συστήματος ώστε να ικανοποιούνται οι προδιαγραφές που αντανakλούν τις απαιτήσεις των χρηστών

I. Σχεδιασμός αρχείων



• Δομή αρχείων

– Χαρακτήρας

- Ελάχιστη μονάδα προσπελάσιμης πληροφορίας που αντιστοιχεί σε ένα byte

– Πεδίο (field)

- Ένα σύνολο από χαρακτήρες που περιγράφουν μια ιδιότητα κάποιας οντότητας. Χαρακτηρίζεται από τη μορφή (αλφαβητική, αριθμητική κλπ.) και το μήκος

– Εγγραφή ή ενότητα δεδομένων (record)

- **Λογική εγγραφή (logical record):** Ένα σύνολο από πεδία που συνδέονται μεταξύ τους λογικά και περιγράφουν μια οντότητα
- **Φυσική εγγραφή (physical record) ή block:** Μονάδα μεταφοράς δεδομένων από και προς την κύρια μνήμη του Η/Υ (μπορεί να περιέχει περισσότερες από μία λογικές εγγραφές)

– Αρχείο (file)

- Ένα σύνολο από εγγραφές του ίδιου τύπου

Δομή αρχείων - Παραδείγματα

→ Πεδίο

J. Stewart (πεδίο **ΟΝΟΜΑ**)

→ Εγγραφή

ΟΝΟΜΑ **ΜΑΘΗΜΑ** **ΕΞΑΜΗΝΟ** **ΒΑΘΜΟΣ**

J. Stewart 24ΔΕ6 Φ000 6,5

→ Αρχείο

ΟΝΟΜΑ **ΜΑΘΗΜΑ** **ΕΞΑΜΗΝΟ** **ΒΑΘΜΟΣ**

K. Taylor	24ΔΕ6	Φ000	8,0
J. Stewart	24ΔΕ6	Φ000	6,5
E. Smith	24ΔΕ6	Φ000	4,5

Πεδία

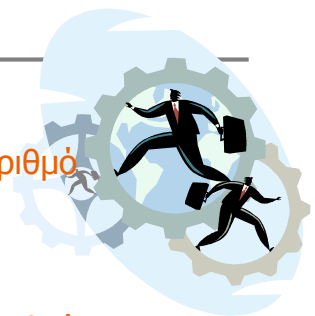
- Πρωτεύων κλειδί (primary key)
 - Προσδιορίζουν μονοσήμαντα μια οντότητα
- Δευτερεύοντα κλειδιά (secondary keys)
 - Προσδιορίζουν ένα υποσύνολο όλων των εγγραφών
- Περιγραφές (descriptions)
 - Περιγράφουν τα χαρακτηριστικά μιας οντότητας



Αριθμός παραγγελίας	Τύπος παραγγελίας	Ημερομηνία παραγγελίας	Κωδικός προϊόντος	Ποσότητα
538	Επείγουσα	05/12/00	Υ503	15
539	Κανονική	05/12/00	Ρ004	30
540	Κανονική	06/12/00	Υ178	25

Εγγραφές

- Σταθερού μήκους
 - Οι (λογικές) εγγραφές περιέχουν προκαθορισμένο αριθμό πεδίων σταθερού μήκους
- Μεταβλητού μήκους
 - Οι (λογικές) εγγραφές περιέχουν προκαθορισμένο αριθμό πεδίων μεταβλητού μήκους ή μεταβλητό αριθμό πεδίων σταθερού μήκους
 - Δεσμεύεται χώρος ίσος με το μήκος της μεγαλύτερης από τις εγγραφές
 - Αποφεύγονται γιατί τις περισσότερες φορές δεν πληρούν τους κανόνες κανονικοποίησης των δεδομένων (data normalization)
- Υπολογισμός μεγέθους αρχείων
 - Χώρος αποθήκευσης λογικών και φυσικών εγγραφών
 - Πλήθος εγγραφών και αναμενόμενο ποσοστό αύξησης



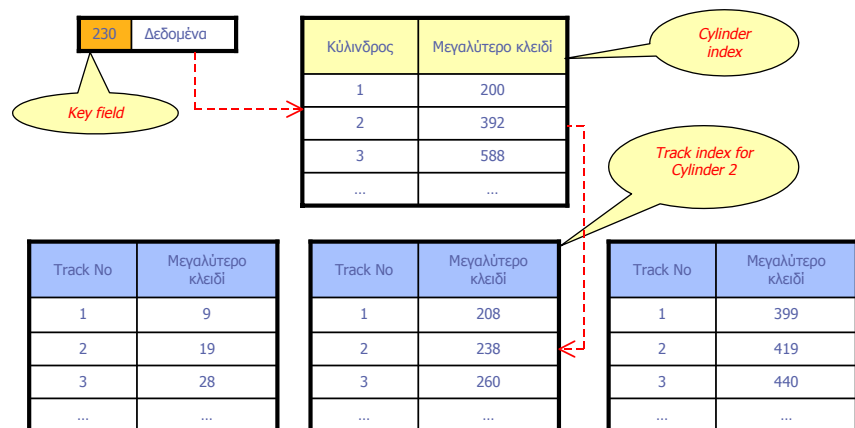
Μέθοδοι οργάνωσης και προσπέλασης αρχείων

- Σειριακή (sequential)
 - Για να διαβαστεί η v -εγγραφή, πρέπει να διαβαστούν πρώτα οι προηγούμενες ($v-1$) εγγραφές
- Σειριακή με δείκτη (indexed sequential)
 - Η μέθοδος βασίζεται σε ένα ευρετήριο (index) από πεδία-κλειδιά, το οποίο χρησιμεύει στην εύρεση μιας συγκεκριμένης εγγραφής
 - Συσχέτιση κλειδιού και φυσικής διεύθυνσης στο δίσκο
- Απευθείας (direct)
 - Το πεδίο-κλειδί χρησιμοποιείται για την εύρεση της φυσικής διεύθυνσης μιας εγγραφής, μέσω ενός ειδικού αλγορίθμου (transform or hashing algorithm)
- Εικονικής αποθήκευσης (VSAM, Virtual Storage Access Method)
 - Οι εγγραφές εισάγονται σε συνεχόμενες περιοχές σταθερού (και βέλτιστου) μεγέθους στο δίσκο (ονομάζονται διαστήματα ελέγχου και αντιστοιχούν στις φυσικές εγγραφές)

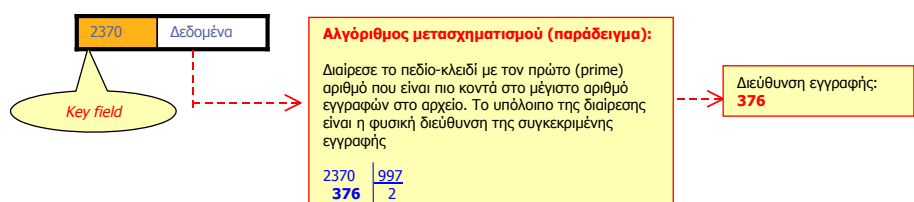


Μέθοδοι οργάνωσης και προσπέλασης (συν.)

→ Σειριακή με δείκτη



→ Απευθείας



Επεξεργασία αρχείων

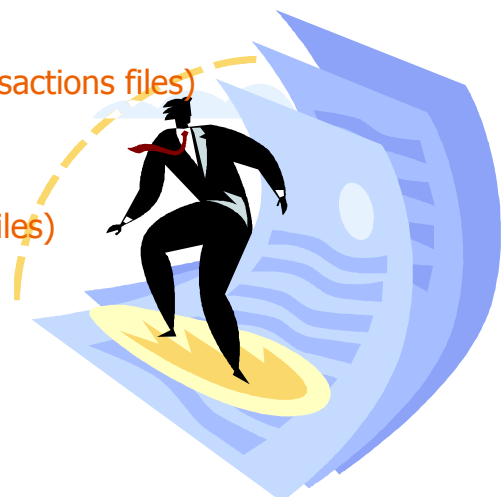
- Ανάκτηση πληροφοριών (retrieval)
 - Εντοπισμός μιας εγγραφής (συνήθως μέσω κάποιου κλειδιού)
 - Χρήση περιεχομένων της εγγραφής για απόκριση
 - Καμία μεταβολή
- Ενημέρωση (update)
 - Προσθήκη εγγραφής (addition)
 - Διαγραφή εγγραφής (deletion)
 - Τροποποίηση εγγραφής (modification)



- Περιορισμοί στη χρήση αρχείων
- Ανάγνωση (read permission)
 - Γράψιμο (write permission)
 - Εκτέλεση (execute permission)

Είδη αρχείων

- Ανάλογα με το περιεχόμενό τους, τα αρχεία που χρησιμοποιούνται για αποθήκευση και επεξεργασία δεδομένων μπορούν να διακριθούν σε:
 - Κύρια αρχεία (master files)
 - Αρχεία μεταβολών ή δοσοληψιών (transactions files)
 - Αρχεία δεικτών (index files)
 - Βοηθητικά αρχεία (auxiliary files)
 - Αρχεία αναφορών (summary ή report files)
 - Εφεδρικά αρχεία (back-up files)
 - Αρχεία τροποποιήσεων
 - Προσωρινά και ενδιάμεσα αρχεία
 - Ιστορικά αρχεία



Κριτήρια σχεδιασμού αρχείων

- Αφορούν τις επιλογές του ΑΣ σχετικά με το φυσικό μέσο αποθήκευσης, την οργάνωση των εγγραφών (πεδία, κλειδιά) και την οργάνωση των αρχείων
- Βασικοί στόχοι σχεδιασμού αρχείων
 - Ορθότητα (integrity)
 - Αποφυγή πλεονασμών (redundancy)



→ Χρήση Βάσεων Δεδομένων ή όχι;

ταχύτητα επεξεργασίας ευελιξία επανάληψη δεδομένων
ευκολία σχεδιασμού και υλοποίησης

Παράδειγμα σχεδιασμού αρχείων

- Γραμμογράφηση

κωδικός	ονοματεπώνυμο	Α.Φ.Μ.	διεύθυνση	Τ.Κ.	Πόλη
9(5)	Ζ(25)	9(9)	Χ(20)	9(5)	Ζ(15)

- Λεξικό Δεδομένων

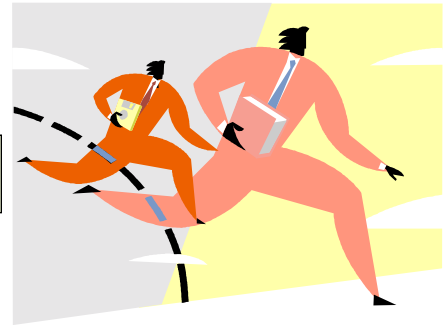
Όνομα αρχείου	PELATES
Σχεδιάστηκε από	Κ. Αναλυτή
Περιγραφή	Κύριο αρχείο πελατών εταιρίας
Μέσο	Δίσκος
Μήκος εγγραφής	Σταθερό
Μέγεθος αρχείου	3.000 εγγραφές
Περίοδος διατήρησης	Μόνιμο αρχείο
Εφεδρεία → Συχνότητα	Κάθε μέρα
→ Διατήρηση	1 εβδομάδα
Οργάνωση και προσπέλαση	Σειριακή με δείκτες

Εφεδρεία και ανάκτηση αρχείων

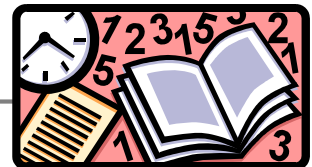
- Κίνδυνος απώλειας δεδομένων
 - Καταστροφή μαγνητικών μέσων αποθήκευσης
 - Λανθασμένος χειρισμός
 - Βλάβες λογισμικού
 - Φυσικές καταστροφές

→ Εφεδρικά αρχεία (back-up files)

- Εφεδρεία σειριακών αρχείων
- Εφεδρεία αρχείων απευθείας προσπέλασης



Εργασίες Εξαμήνου - Συζήτηση



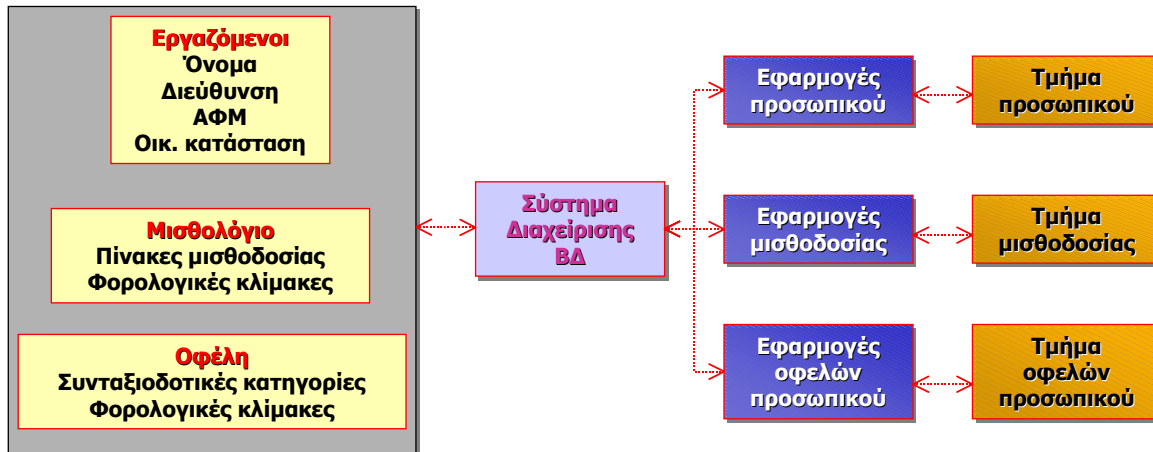
Σχεδιασμός αρχείων

- Πληροφοριακό Σύστημα Διαχείρισης Προμηθειών μιας Επιχείρησης
 - Παραδείγματα δομής αρχείων
- Σύστημα Επεξεργασίας Συναλλαγών Ηλεκτρονικού Εμπορίου (Business to Customer)
 - Χρήση Βάσης Δεδομένων;
- Σύστημα Διαχείρισης Βαθμολογίας Φοιτητών σε ένα Πανεπιστημιακό Τμήμα
 - Παραδείγματα σχεδιασμού αρχείων
- Σύστημα Διαχείρισης Γνώσης και Συλλογικής Υποστήριξης Αποφάσεων για Ιατρικές Διαγνώσεις
 - Περιορισμοί στη χρήση αρχείων

II. Σχεδιασμός Βάσεων Δεδομένων

• Βάση Δεδομένων

- Συλλογή δεδομένων που οργανώνονται έτσι ώστε να εξυπηρετούν αποδοτικά διάφορες εφαρμογές
 - Εστίαση στα δεδομένα
 - Ελαχιστοποίηση πλεοναζόντων δεδομένων



Συστήματα Διαχείρισης Βάσεως Δεδομένων

• Γενικά

- Το λογισμικό με το οποίο επιτυγχάνεται η αποθήκευση, ανάκτηση, συντήρηση και, γενικότερα, η διαχείριση των δεδομένων
- Ένα ΣΔΒΔ (Database Management System - DBMS) αποτελεί ουσιαστικά τη διασύνδεση (interface) ανάμεσα στα προγράμματα εφαρμογών και τα αρχεία δεδομένων
- Εξαλείφει σε μεγάλο βαθμό την ανάγκη για εντολές προσδιορισμού δεδομένων (data definition statements)
 - data definition statements: εντολές που προσδιορίζουν τα χαρακτηριστικά ενός αρχείου (αφορούν το όνομα, την καταχώρηση ή άλλα στοιχεία που προσδιορίζουν με ακρίβεια την ταυτότητα του αρχείου)

• Αποτελείται από:

- Γλώσσα προσδιορισμού δεδομένων (data definition language)
- Γλώσσα χειρισμού δεδομένων (data manipulation language)
- Λεξικό δεδομένων (data dictionary)

Συστήματα Διαχείρισης Βάσεως Δεδομένων (συν.)

- Γλώσσα προσδιορισμού δεδομένων
 - Είναι μια γλώσσα προγραμματισμού, η οποία καθορίζει το περιεχόμενο και τη δομή μιας ΒΔ
 - Ορίζει το πως θα εμφανίζεται στη ΒΔ κάθε στοιχείο δεδομένων (πριν αυτό μετατραπεί στη μορφή την οποία απαιτεί κάποια συγκεκριμένη εφαρμογή)
- Γλώσσα χειρισμού δεδομένων
 - Γλώσσα στενά συνδεδεμένη με ένα ΣΔΒΔ, η οποία χρησιμοποιείται από τελικούς χρήστες και προγραμματιστές για εξαγωγή δεδομένων προς ικανοποίηση ερωτήσεων, αλλά και ανάπτυξη εφαρμογών
 - Χρησιμοποιείται μαζί με μια συμβατική γλώσσα προγραμματισμού (τρίτης ή τέταρτης γενιάς)
 - SQL (structured query language): το standard για σχεσιακές ΒΔ
- Λεξικό δεδομένων
 - Είναι ένα αρχείο (ενημερώνεται αυτόματα ή από τον προγραμματιστή) στο οποίο αποθηκεύονται ο κατάλογος των αρχείων μιας ΒΔ, τα διάφορα ευρετήρια (προγραμμάτων, εφαρμογών, κλπ.), οι ορισμοί όλων των πεδίων αλλά και χαρακτηριστικά όπως η χρήση, φυσική αναπαράσταση, ιδιοκτησία, ασφάλεια κλπ. των δεδομένων

Structured Query Language (SQL)

Student

Stud_id	Stud_name	Stud_email	Stud_stat
3545	K. Taylor	taylo@cmu.edu	Active
5624	J. Keen	keen@cmu.edu	Active
4235	J. Stewart	stew@yahoo.com	Active
6655	G. Best	best@aol.com	Inactive
...
4211	E. Smith	smith@aol.com	Active

Course

Course_id	Stud_id	Semester	Grade
24ΔΕ6	3545	Fall00	8,0
35ZE9	3545	Spring01	Active
24ΔΕ6	4235	Fall100	6,5
...
24ΔΕ6	4211	Fall100	4,5

Stud_name	Stud_ID	Grade
K. Taylor	3545	8,0
J. Stewart	4235	6,5
E. Smith	4211	4,5



```
SELECT Stud_name, Student.stud_id, Grade
FROM Student, Course
WHERE Student.stud_id = Course.stud_id
AND Course_id = "24ΔΕ6"
```

Περιβάλλον ΒΔ – Θεώρηση δεδομένων

- Φυσική θεώρηση δεδομένων (physical view)
 - Η αναπαράσταση των δεδομένων όπως αυτά είναι στην πραγματικότητα οργανωμένα και δομημένα στο φυσικό μέσο αποθήκευσης
- Λογική θεώρηση δεδομένων (logical view)
 - Η αναπαράσταση των δεδομένων που απαιτείται από τις συγκεκριμένες εφαρμογές και τους τελικούς χρήστες
 - Π.χ., όνομα, τύπος, μήκος πεδίου



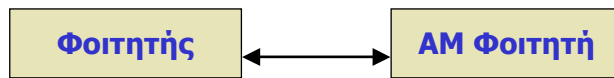
Πλεονεκτήματα ΣΔΒΔ

- Περιορισμός πλεονασμού δεδομένων (data redundancy)
- Περιορισμός αντιφατικότητας (inconsistency) δεδομένων
- Περιορισμός της πολυπλοκότητας του συστήματος
- Μείωση κόστους ανάπτυξης και συντήρησης
- Καλύτερος (κεντρικός) έλεγχος της δημιουργίας και του προσδιορισμού των δεδομένων
 - Καλύτερη πρόσβαση και διαθεσιμότητα πληροφορίας
- Περιορισμός της εξάρτησης προγραμμάτων εφαρμογών και δεδομένων
 - Αύξηση ευελιξίας συστήματος

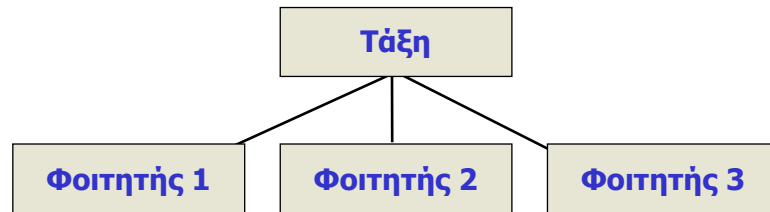


Τύποι σχέσεων σε μια ΒΔ

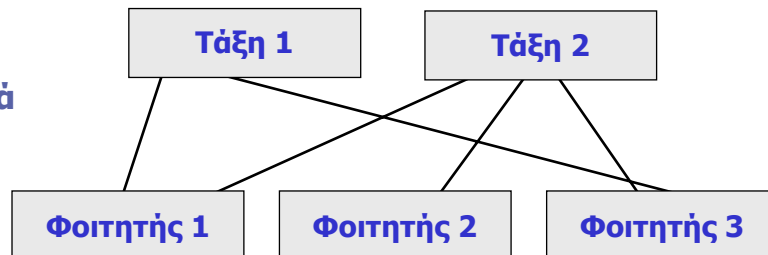
Ένα-προς-ένα
(One-to-one)



Ένα-προς-πολλά
(One-to-many)



Πολλά-προς-πολλά
(Many-to-many)



Κανονικοποίηση σχέσεων

- Γενικά

- Κανονικοποίηση (normalization) είναι η διαδικασία που έχει σαν σκοπό την εφαρμογή κανόνων στις σχέσεις, με στόχο την ελάττωση πλεοναζόντων δεδομένων (και των επακόλουθων προβλημάτων), την εξασφάλιση συμβατότητας μεταξύ των δεδομένων, και την εξάλειψη προβλημάτων σχετικών με την εισαγωγή, διαγραφή και τροποποίηση των δεδομένων

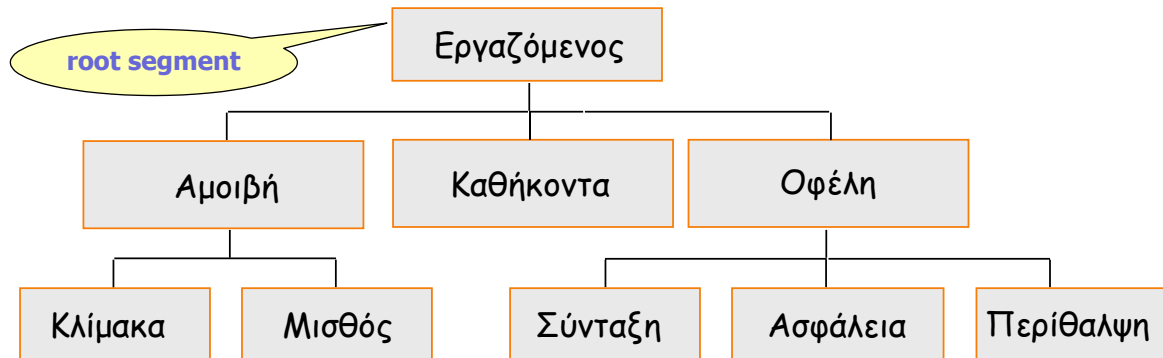


- Κανονικές Μορφές

Σχεδιασμός ΒΔ – Ιεραρχικό μοντέλο

- Γενικά

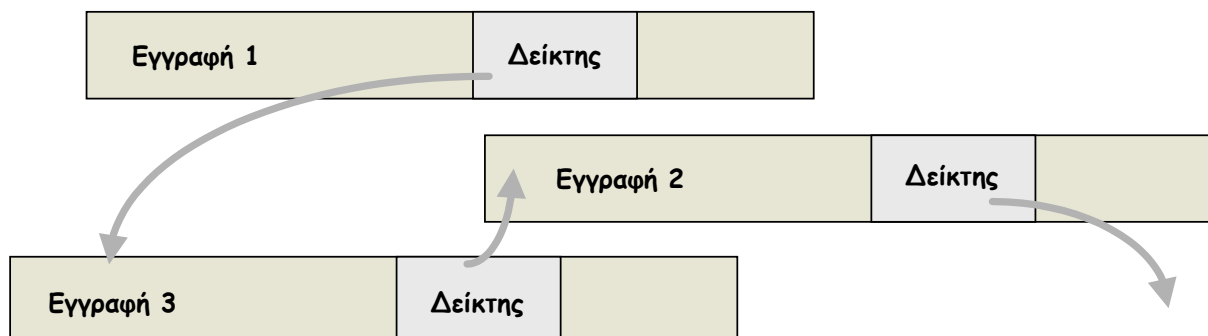
- Τα δεδομένα οργανώνονται σε μια δομή δέντρου
- Σε κάθε εγγραφή, τα δεδομένα οργανώνονται σε τμήματα (segments)
- Στον χρήστη, κάθε εγγραφή μοιάζει σαν ένα οργανόγραμμα με ένα βασικό τμήμα (root segment) στην κορυφή του
- Οργάνωση με βάση τη σχέση «πατέρας-παιδί»



Σχεδιασμός ΒΔ – Ιεραρχικό μοντέλο (συν.)

- Δείκτες (pointers)

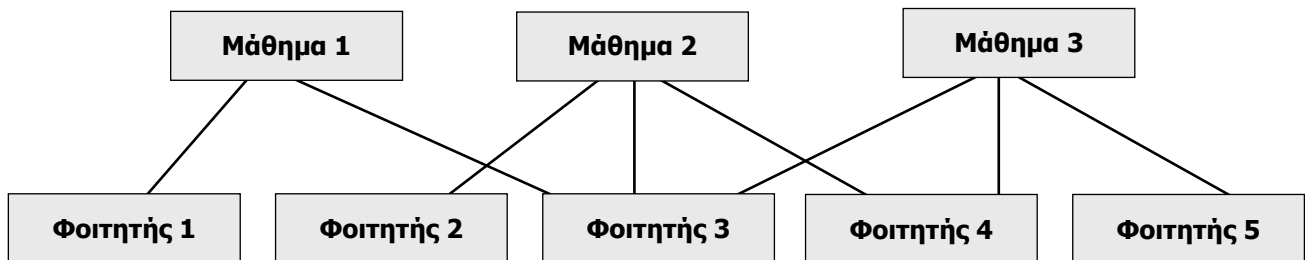
- Σε ένα ιεραρχικό ΣΔΒΔ, τα δεδομένα συνδέονται φυσικά μεταξύ τους μέσω μιας σειράς δεικτών που ορίζουν σύνολα από σχετικά τμήματα δεδομένων
- Οι δείκτες αυτοί είναι δεδομένα που επισυνάπτονται στο τέλος ενός τμήματος δεδομένων (του «πατέρα») και δίνουν τη φυσική διεύθυνση των σχετικών τμημάτων δεδομένων (των «παιδιών»)



Σχεδιασμός ΒΔ – Δικτυακό μοντέλο

- Γενικά

- Παραλλαγή του ιεραρχικού μοντέλου
- Ιδιαίτερα χρήσιμο για "many-to-many" σχέσεις
 - Ένα «παιδί» μπορεί να έχει πάνω από έναν «πατέρα»
- Πλεονεκτήματα και μειονεκτήματα
 - Ελαχιστοποίηση πλεονασμού δεδομένων και, συχνά, μείωση του χρόνου απόκρισης
 - Ραγδαία αύξηση του αριθμού των δεικτών, με αποτέλεσμα η συντήρηση και λειτουργία του συστήματος να γίνεται συχνά πολύπλοκη



Σχεδιασμός ΒΔ – Σχεσιακό μοντέλο

- Γενικά

- Αναπαράσταση όλων των δεδομένων μέσω διδιάστατων πινάκων που ονομάζονται «σχέσεις» (relations)
- Μπορεί να συσχετίσει δεδομένα τέτοιων πινάκων, αρκεί οι πίνακες αυτοί να έχουν ένα κοινό στοιχείο δεδομένων
- Λειτουργίες (εντολές) select, join, project
- IBM's DB2, Oracle, Microsoft Access

ID παραγγελίας	Ημερομηνία παραγγελίας	Ημερομηνία παράδοσης	ID εξαρτήματος	Ποσότητα
2625	05/12/00	30/12/00	A145	8
2626	06/12/00	30/12/00	A152	6
2627	09/12/00	05/01/01	B201	4

παραγγελία

εξάρτημα

προμηθευτής

ID εξαρτήματος	Περιγραφή εξαρτήματος	Τιμή μονάδας	ID προμηθευτή
A145	Πόμολο πόρτας Νο 35	2000	ΕΛ234
A152	Ασφάλεια τύπου Max	4520	ΕΞ001
B201	Βίδες μπρούντζου (x12)	160	ΕΞ029

ID προμηθευτή	Όνομα προμηθευτή	Τηλέφωνο προμηθευτή
ΕΞ001	Security Systems S.A.	+44-671-55624
ΕΞ029	Power Inc.	+33-12-987678
ΕΛ234	Σιδηρεμπορική Α.Ε.	01-9264811

Συγκριτική θεώρηση μοντέλων

Τύπος ΒΔ	Απόδοση επεξεργασίας δεδομένων	Ευελιξία	Φιλικότητα στον τελικό χρήστη	Πολυπλοκότητα στον προγραμματισμό
Ιεραρχική	Υψηλή	Χαμηλή	Χαμηλή	Υψηλή
Δικτυακή	Μέτρια προς υψηλή	Χαμηλή προς μέτρια	Χαμηλή προς μέτρια	Υψηλή
Σχεσιακή	Χαμηλότερη (βελτιώνεται)	Υψηλή	Υψηλή	Χαμηλή

Ανάπτυξη μιας Βάσης Δεδομένων

- **Εννοιολογικός ή λογικός σχεδιασμός (conceptual or logical design)**
 - Γενικευμένο (abstract) μοντέλο της ΒΔ, με βάση τις απαιτήσεις της Διοίκησης
 - Ομαδοποίηση δεδομένων
 - Καθορισμός σχέσεων μεταξύ δεδομένων
 - Καθορισμός αναγκών τελικού χρήστη
- **Φυσικός σχεδιασμός (physical design)**
 - Αναλυτικό μοντέλο της ΒΔ
 - Διαγράμματα οντοτήτων-συσχετίσεων (entity-relationship, ER)
 - Κανονικοποίηση (normalization)
 - Δημιουργία κατάλληλων (αποδοτικών και μικρών σε έκταση) δομών δεδομένων από περίπλοκα σύνολα δεδομένων
 - Σχεδιασμός με βάση καθορισμένο hardware και software

Διαγράμματα οντοτήτων-συσχετίσεων

- Γενικά

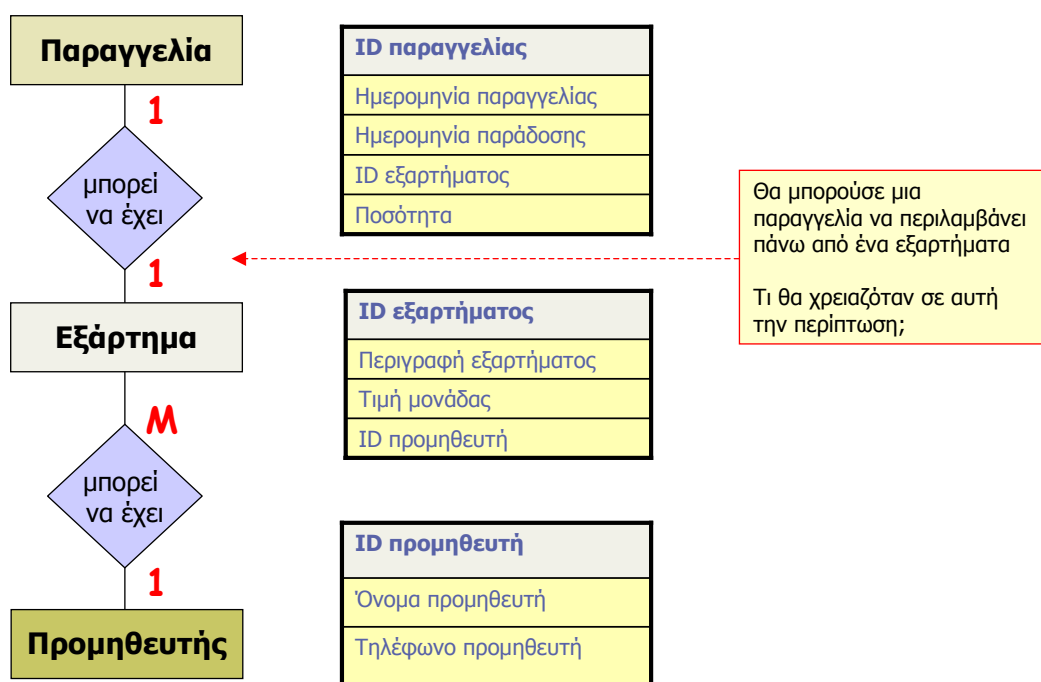
- Εργαλείο ανάλυσης και σχεδιασμού ΠΣ
- Προσδιορισμός οντοτήτων και ιδιοτήτων τους, καθώς και των μεταξύ τους συσχετίσεων
- Ανεξάρτητα από το ΣΔΒΔ που θα χρησιμοποιηθεί
- ER (Entity-Relationship)

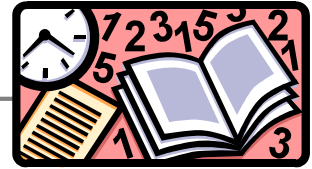


- Συμβολισμοί



Διαγράμματα οντοτήτων-συσχετίσεων (συν.)





- Πληροφοριακό Σύστημα Διαχείρισης Προμηθειών μιας Επιχείρησης
 - Σχεδιασμός ΒΔ: Ιεραρχικό μοντέλο
- Σύστημα Επεξεργασίας Συναλλαγών Ηλεκτρονικού Εμπορίου (Business to Customer)
 - Διαγράμματα Οντοτήτων-Συσχετίσεων
- Σύστημα Διαχείρισης Βαθμολογίας Φοιτητών σε ένα Πανεπιστημιακό Τμήμα
 - Τύποι σχέσεων στη ΒΔ
- Σύστημα Διαχείρισης Γνώσης και Συλλογικής Υποστήριξης Αποφάσεων για Ιατρικές Διαγνώσεις
 - Σχεδιασμός ΒΔ: Σχεσιακό μοντέλο