

HY 360 – Αρχεία και Βάσεις Δεδομένων

Web site: <http://www.csd.uoh.gr/~hy360>

Mailing list: hy360-list@csd.uoc.gr

E-mail: hy360@csd.uoc.gr

Δημήτρης Πλεξουσάκης

Καθηγητής

Τμήμα Επιστήμης Υπολογιστών,

Πανεπιστήμιο Κρήτης

E-mail: dp@csd.uoc.gr

Γραφείο: K 307

Επισκόπηση Μαθήματος

- Ώρες Διδασκαλίας: Δευτέρα, Τετάρτη 3-5 (Αμφ. Α)
- Ώρες Φροντιστηρίου: Παρασκευή 3-5 (Α113)
- Ώρες Γραφείου: Δευτέρα, Τετάρτη 5-6 (ή με συνεννόηση)
- Βοηθοί Μαθήματος:
 - Δέσποινα Αθανασιάδου
 - Μιχάλης Γαλλιάρης
 - Σπύρος Δουκέρης
 - Μιχάλης Καραγιαννάκης
 - Βάσια Πιτικάκη
 - Μιχάλης Σαβοριανάκης
- Ώρες Γραφείου βοηθών: θα οριστούν

Επισκόπηση Μαθήματος

- Προαπαιτούμενες Γνώσεις:

- (HY240-Δομές Δεδομένων ή HY255-Εργαστήριο Λογισμικού) και (HY118-Διακριτά Μαθηματικά ή HY180-Λογική)

- Βιβλιογραφία:

- Silberschatz, Korth & Sudarshan, «Συστήματα Βάσεων Δεδομένων», 7^η έκδοση.

- Βιβλία για συμπληρωματική μελέτη:

- C. Date, «Εισαγωγή στα Συστήματα Βάσεων Δεδομένων»
- Elmasri & Navathe, «Θεμελιώδεις Αρχές Συστημάτων Βάσεων Δεδομένων».

Επισκόπηση Μαθήματος

- Βαθμός Μαθήματος:

- Ο βαθμός στο μάθημα θα βασιστεί σε:

- Α: Ασκήσεις / Quiz (υποχρεωτικά)
- Ε: Εργασία (υποχρεωτικά, σε ομάδες)
- Τ: Τελική Εξέταση

- ... σύμφωνα με τον τύπο:

$$\text{Βαθμός} = 0,25 * A + 0,35 * E + 0,40 * T,$$

αν $T \geq 4$ και $A > 0$;

Διαφορετικά:

αν $A=0$, Βαθμός = 0;

αν $T < 4$, Βαθμός = T

Επισκόπηση Μαθήματος

- **Γενική Περιγραφή Μαθήματος:** Εισαγωγή στα Συστήματα Διαχείρισης Βάσεων Δεδομένων

- **Θέματα:**

- Μοντέλο Οντοτήτων Σχέσεων (Entity-Relationship Model)
- Σχεσιακό Μοντέλο (Relational Model)
- Γλώσσες Επερωτήσεων, SQL (Query Languages)
- Περιορισμοί Ακεραιότητας (Integrity Constraints)
- Σχεδίαση Βάσεων Δεδομένων (Database Design)
- Δομές Αρχείων, Αποθήκευση (File Structures and Storage)
- Ευρετηριασμός (Indexing)
- Δοσοληψίες (Transactions) και Έλεγχος Ταυτόχρονης Προσπέλασης (Concurrency Control)

Εισαγωγή στα Συστήματα Διαχείρισης Βάσεων Δεδομένων

- Ενα Σύστημα Διαχείρισης Βάσεων Δεδομένων (ΣΔΒΔ) (Database Management System - DBMS) αποτελείται από μία συλλογή σχετιζόμενων δεδομένων και ένα σύνολο προγραμμάτων για την προσπέλαση σε αυτά τα δεδομένα.
- Η συλλογή των δεδομένων ονομάζεται **βάση δεδομένων** (ΒΔ) (database).
- Ο κύριος σκοπός ενός ΣΔΒΔ είναι να παρέχει
 - ένα περιβάλλον για την **εύκολη** και **αποδοτική αποθήκευση**, **διαχείριση** και **ανάκτηση** μεγάλου όγκου πληροφορίας
 - μια **αφηρημένη όψη δεδομένων** αποκρύπτοντας λεπτομέρειες της αναπαράστασης και της αποθήκευσής τους στο σύστημα

Εισαγωγή στα Συστήματα Διαχείρισης Βάσεων Δεδομένων

- Η διαχείριση των δεδομένων περιλαμβάνει:
 - τον ορισμό **δομών** για την **αποθήκευση** πληροφορίας
 - μηχανισμούς για τον **χειρισμό** της πληροφορίας και την **ενημέρωσή** της
 - μηχανισμούς για την **ασφάλεια** πληροφορίας που είναι προσβάσιμη από πολλούς χρήστες

Εισαγωγή στα Συστήματα Διαχείρισης Βάσεων Δεδομένων

- **Παράδειγμα:** ένα τραπεζικό σύστημα αποθηκεύει πληροφορία σχετικά με τους πελάτες της τράπεζας και τους λογαριασμούς τους.
 - Η πληροφορία αποθηκεύεται σε αρχεία του συστήματος τα οποία διαχειρίζονται προγράμματα για:
 - χρέωση ή πίστωση λογαριασμού
 - προσθήκη νέου λογαριασμού
 - εύρεση υπολοίπου
 - μηνιαίες / ετήσιες καταστάσεις κίνησης
 - κλπ
 - Καινούργια αρχεία και προγράμματα εφαρμογών προστίθενται ανάλογα με τις ανάγκες

Εισαγωγή στα Συστήματα Διαχείρισης Βάσεων Δεδομένων

- Η υλοποίηση ενός τέτοιου συστήματος μπορεί να βασιστεί σε ένα «τυπικό» σύστημα διαχείρισης και επεξεργασίας αρχείων (file processing system), πάσχει όμως από διάφορα μειονεκτήματα:
 - **Πλεονασμός** και **Ασυνέπεια** δεδομένων (data redundancy and inconsistency)
 - η ίδια πληροφορία πιθανόν να **επαναλαμβάνεται** σε διαφορετικά αρχεία με αποτέλεσμα να υπάρχει μεγαλύτερο κόστος αποθήκευσης και ανάκλησης
 - Επιπλέον, ενημέρωση ενός αρχείου πρέπει να συνοδεύεται από την ενημέρωση όλων των αρχείων στα οποία η πληροφορία επαναλαμβάνεται.

Εισαγωγή στα Συστήματα Διαχείρισης Βάσεων Δεδομένων

- Δυσκολία στην **πρόσβαση** στα δεδομένα
 - τα συστήματα διαχείρισης αρχείων δεν παρέχουν αποδοτικούς μηχανισμούς για αναζήτηση και ανάκληση πληροφορίας με διαφορετικούς τρόπους.
 - Επιπλέον, τα δεδομένα διασκορπίζονται σε αρχεία που χρησιμοποιούν διαφορετικές μορφές.
- Προβλήματα **ακεραιότητας** δεδομένων (data integrity)
 - τα δεδομένα υπόκεινται σε **περιορισμούς συνέπειας** ή **ακεραιότητας** (consistency or integrity constraints).
 - Οι περιορισμοί αυτοί εφαρμόζονται μέσω των προγραμμάτων εφαρμογών τα οποία πρέπει να αλλαχθούν για την προσθήκη νέων ή τη μεταβολή περιορισμών.

Εισαγωγή στα Συστήματα Διαχείρισης Βάσεων Δεδομένων

- Προβλήματα **ατομικότητας δοσοληψιών** (transaction atomicity)
 - το σύστημα πρέπει να εγγυάται ότι σε περίπτωση σφάλματος η πληροφορία πρέπει να μπορεί να αποκατασταθεί.
 - Π.χ. μεταφορά ενός ποσού X από λογαριασμό A σε λογαριασμό B. Αν συμβεί κάποιο σφάλμα κατά τη διάρκεια της μεταφοράς, ο λογαριασμός A μπορεί να χρεωθεί χωρίς να πιστωθεί ο λογαριασμός B.
 - Οι δοσοληψίες πρέπει να χαρακτηρίζονται από ατομικότητα, δηλαδή, είτε όλη η δοσοληψία ολοκληρώνεται είτε κανένα μέρος αυτής.

Εισαγωγή στα Συστήματα Διαχείρισης Βάσεων Δεδομένων

- Ανωμαλίες ταυτόχρονης πρόσβασης (concurrent access)
 - Η ταυτόχρονη πρόσβαση και ενημέρωση της πληροφορίας από πολλούς χρήστες μπορεί να δημιουργήσει προβλήματα συνέπειας.
 - Π.χ. έστω ότι ο λογαριασμός A έχει υπόλοιπο X . Αν δύο χρήστες αφαιρέσουν ποσά $\Delta 1$ και $\Delta 2$ "ταυτόχρονα" από τον A , το αποτέλεσμα μπορεί να μην είναι ορθό.
- Προβλήματα ασφάλειας (security)
 - διαφορετικοί χρήστες θα πρέπει να έχουν διαφορετικά δικαιώματα πρόσβασης σε αρχεία ή σε συγκεκριμένα δεδομένα.

Όψεις Δεδομένων

- Ένας από τους κύριους στόχους ενός ΣΔΒΔ είναι να παρέχει μια **αφηρημένη όψη** των δεδομένων **αποκρύπτοντας** λεπτομέρειες της **αναπαράστασης** και της **αποθήκευσής** τους στο σύστημα.
- Ένα ΣΔΒΔ είναι ένα πρόγραμμα για τη διαχείριση ηλεκτρονικών αρχείων επιχειρησιακών δεδομένων με δομημένο τρόπο.
- Για παράδειγμα,
 - ΣΔΒΔ ενός πανεπιστημίου διαχειρίζεται αρχεία για τους φοιτητές, τις εγγραφές τους σε μαθήματα, βαθμολογία, κ.λ.π.
 - Εμπορικές επιχειρήσεις χρησιμοποιούν ΣΔΒΔ για την αποθήκευση στοιχείων σχετικά με πωλήσεις, αγορές, τιμολόγηση, αποθήκη κ.λ.π.

Όψεις Δεδομένων

- Αεροπορικές εταιρίες χρησιμοποιούν ΣΔΒΔ για την αποθήκευση στοιχείων σχετικά με πτήσεις, δρομολόγια, διαθεσιμότητα θέσεων, κρατήσεις, προσωπικό, κ.λ.π
- Ένα ΣΔΒΔ μιας βιβλιοθήκης καταγράφει τα διαθέσιμα βιβλία, τα άτομα που τα δανείζονται, και παρέχει θεματική κατηγοριοποίηση.
- Η συλλογή των αρχείων και εγγραφών που κρατούνται για ένα συγκεκριμένο σκοπό ονομάζεται **βάση δεδομένων** (database)
- Κατά κανόνα, οι βάσεις δεδομένων αποθηκεύονται σε δευτερεύουσα μνήμη.
- Ένα ΣΔΒΔ μπορεί να διαχειρίζεται περισσότερες από μία βάσεις δεδομένων.

Όψεις Δεδομένων

- Οι βάσεις δεδομένων μπορούν να μοιράζονται πληροφορία.
- Η πρόσβαση σε αυτές ελέγχεται από ένα κεντρικό πρόγραμμα το οποίο ονομάζεται διαχειριστής βάσεων δεδομένων (db manager).
- Η δόμηση της πληροφορίας που αποθηκεύεται στις βάσεις δεδομένων καθορίζεται από το **μοντέλο δεδομένων** (data model).

Μοντέλα Δεδομένων

- Αρχικά μοντέλα δεδομένων:
 - **ιεραρχικό** (hierarchical data model): διαφορετικοί τύποι εγγραφών δομούνται υπό μορφή ιεραρχίας (δένδρου)
 - **μοντέλο δικτύου** (network data model): γενίκευση του ιεραρχικού μοντέλου.
 - Οι εγγραφές οργανώνονται υπό μορφή κατευθυνόμενου άκυκλου γράφου.
- **Μειονέκτημα** των αρχικών μοντέλων δεδομένων:
 - η επερώτηση στις βάσεις δεδομένων γίνεται πολύπλοκη.

Μοντέλα Δεδομένων

- Το **σχεσιακό μοντέλο** (relational model) είναι το πλέον επιτυχημένο μοντέλο δεδομένων από αυτά που έχουν προταθεί.
 - Τα δεδομένα οργανώνονται σε μορφή **πινάκων εγγραφών** ή **σχέσεων** (tables of records or relations).
 - Οι σχέσεις είναι συλλογές **γνωρισμάτων** (attributes), όπου κάθε γνώρισμα (μια στήλη του πίνακα) αναπαριστά μια ιδιότητα της εγγραφής.
 - Παρέχει εύκολο τρόπο επερώτησης των δεδομένων.
 - Εμπορικά σχεσιακά ΣΔΒΔ: Ingres, Oracle, DB2, Postgres, Sybase, MySQL κλπ.

Χρήστες ΣΔΒΔ

- Ένα ΣΔΒΔ πρέπει να παρέχει εύκολη πρόσβαση στα δεδομένα σε σχετικά μη-έμπειρους χρήστες οι οποίοι λέγονται **τελικοί χρήστες** (end users).
- Η φιλικότητα προς το χρήστη και η χρήση μιας γλώσσας ερωτημάτων υψηλού επιπέδου είναι σημαντικά χαρακτηριστικά που πρέπει να διαθέτει ένα ΣΔΒΔ.
- Διαφορετικοί χρήστες έχουν διαφορετικές απαιτήσεις:
 - τελικοί χρήστες
 - προγραμματιστές εφαρμογών (application programmers)
 - διαχειριστές ΒΔ (database administrators)

Χρήστες ΣΔΒΔ

- Τελικοί χρήστες
 - **συνήθεις** ή **περιστασιακοί** χρήστες (casual users): χρησιμοποιούν μια δομημένη γλώσσα (π.χ. SQL) για την διατύπωση ερωτήσεων στη ΒΔ
 - **αδαείς χρήστες** (naive users): διατυπώνουν ερωτήσεις χρησιμοποιώντας φόρμες ή μενού
- Προγραμματιστές εφαρμογών:
 - γράφουν προγράμματα εφαρμογών για να εξυπηρετήσουν τις ανάγκες των τελικών χρηστών. Έχουν την ευθύνη της υλοποίησης ενός τρόπου επρώτησης με αποδοτικό τρόπο.
- Διαχειριστής ΒΔ: είναι υπεύθυνος για τη σχεδίαση και συντήρηση των ΒΔ

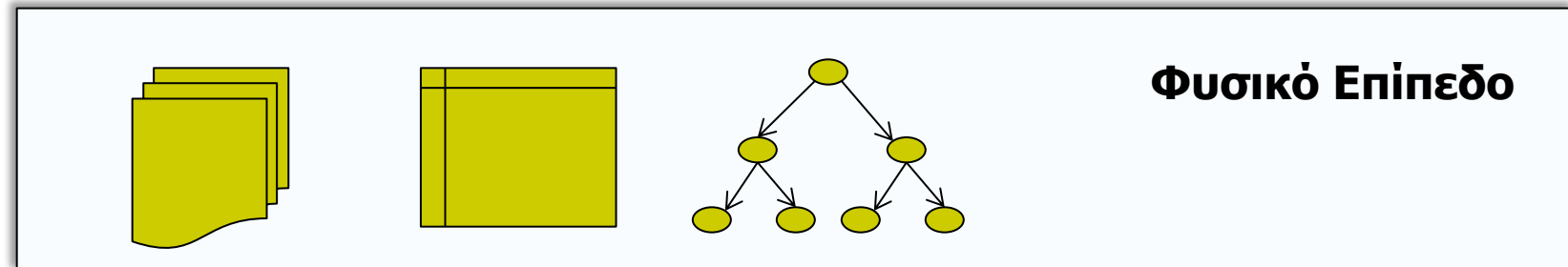
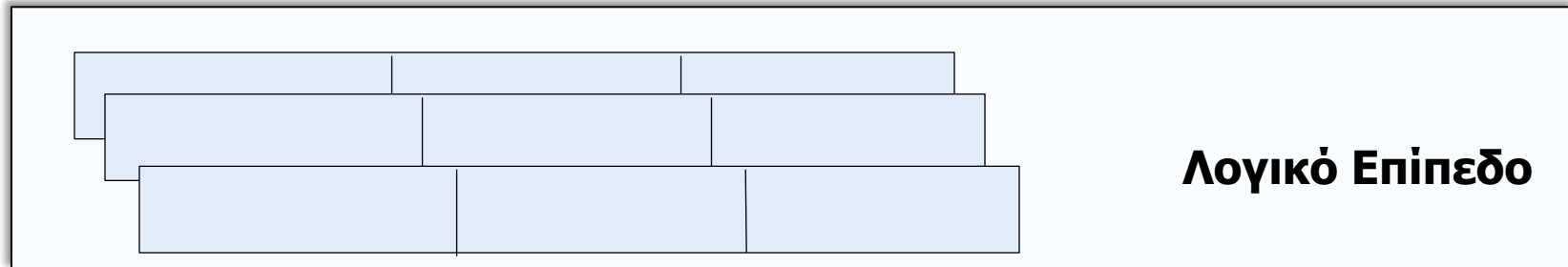
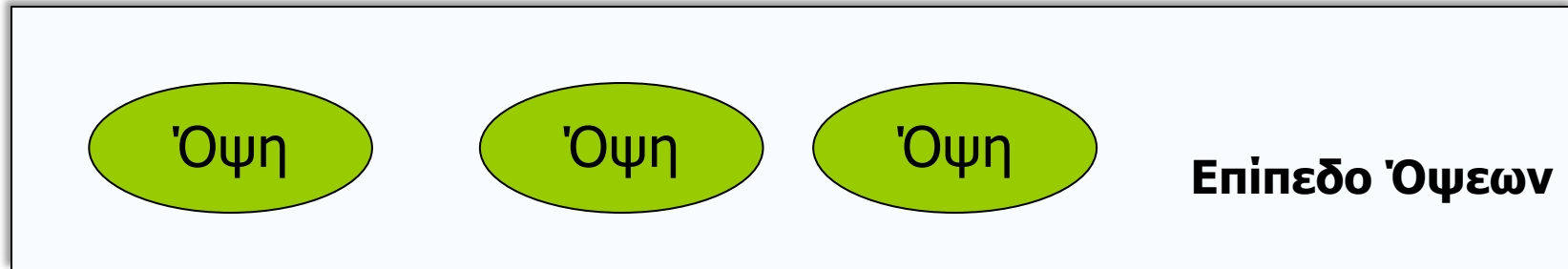
Επίπεδα Αφαίρεσης ΣΔΒΔ

- Ένας από τους κύριους στόχους ενός ΣΔΒΔ είναι να παρέχει μια **αφηρημένη όψη** των δεδομένων, αποκρύπτοντας από το χρήστη λεπτομέρειες σχετικά με την αναπαράσταση και την αποθήκευσή τους.
- Σε ένα ΣΔΒΔ, τα δεδομένα αναπαριστώνται σε 3 **επίπεδα αφαίρεσης** (abstraction levels):
 - **Φυσικό επίπεδο** (physical level):
 - το χαμηλότερο επίπεδο αφαίρεσης
 - περιλαμβάνει σύνθετες δομές και λεπτομέρειες αναπαράστασης και αποθήκευσης

Επίπεδα Αφαίρεσης ΣΔΒΔ

- **Λογικό επίπεδο** (logical level): περιγράφει τα δεδομένα και τις μεταξύ τους σχέσεις χρησιμοποιώντας ένα σχετικά μικρό αριθμό απλών δομών και εννοιών
 - Οι χρήστες οι οποίοι βλέπουν το σύστημα στο λογικό επίπεδο δεν χρειάζεται να γνωρίζουν τις λεπτομέρειες του φυσικού επιπέδου
- **Επίπεδο όψεων** (view level):
 - το υψηλότερο επίπεδο αφαίρεσης
 - περιγράφει κάποια **μέρη** της αποθηκευμένης πληροφορίας, καθώς κάποιοι χρήστες χρειάζονται πρόσβαση μόνο σε μέρος των ΒΔ
 - αποτελεί συγχρόνως και **μηχανισμό ασφάλειας**

Επίπεδα Αφαίρεσης ΣΔΒΔ



Επίπεδα Αφαίρεσης ΣΔΒΔ

- Η διάκριση μεταξύ των διαφορετικών επιπέδων αφαίρεσης είναι ανάλογη με τη διάκριση μεταξύ επιπέδων αφαίρεσης στις γλώσσες προγραμματισμού
- Για παράδειγμα, σε μια γλώσσα προγραμματισμού, οι δηλώσεις:

```
type customer = record
    customer-name : string;
    customer-id : string;
    customer-address: string;
end;
type account=record
    account-number: integer;
end;
```

Επίπεδα Αφαίρεσης ΣΔΒΔ

```
type employee=record  
    employee-name: string;  
    salary: integer;  
end;
```

- αντιστοιχούν στο λογικό επίπεδο.
- Στο φυσικό επίπεδο, οι τύποι αυτοί αναπαριστώνται σαν συνεχόμενες θέσεις αποθήκευσης.
- Η λεπτομέρεια αυτή αποκρύπτεται από τον προγραμματιστή.

Σχήμα και Στιγμιότυπα (Schema and Instances)

- Οι ΒΔ μεταβάλλονται με την πάροδο του χρόνου καθώς πληροφορία προστίθεται, αφαιρείται ή τροποποιείται.
- Το σύνολο της πληροφορίας το οποίο βρίσκεται αποθηκευμένο σε μια ΒΔ σε μια συγκεκριμένη χρονική στιγμή λέγεται **στιγμιότυπο** (instance or snapshot) της ΒΔ.
- Το **σχήμα** (schema) της ΒΔ είναι το σύνολο των εννοιών που περιγράφουν τη δομή της και είναι – σχεδόν πάντα – σταθερό.
- **Αναλογία** : record types $\leftrightarrow$ schema, variables $\leftrightarrow$ instances
- Μια ΒΔ έχει ένα (ή περισσότερα) σχήματα **σε κάθε** επίπεδο αφαίρεσης:
 - Φυσικό σχήμα
 - Λογικό σχήμα
 - Σχήματα όψεων ή υποσχήματα

Ανεξαρτησία Δεδομένων (Data Independence)

- Ο όρος **ανεξαρτησία δεδομένων** (data independence) χαρακτηρίζει την ικανότητα μεταβολής του σχήματος σε κάποιο επίπεδο αφαίρεσης, χωρίς να επηρεάζεται ο ορισμός του σχήματος στο αμέσως υψηλότερο επίπεδο
 - **φυσική ανεξαρτησία** δεδομένων:
 - το φυσικό σχήμα μπορεί να μεταβάλλεται χωρίς να χρειάζεται να γραφούν ξανά τα προγράμματα εφαρμογών που έχουν πρόσβαση στη ΒΔ.
 - μεταβολές στο φυσικό σχήμα είναι συχνά απαραίτητες για λόγους βελτιστοποίησης της απόδοσης του συστήματος

Ανεξαρτησία Δεδομένων (Data Independence)

■ λογική ανεξαρτησία δεδομένων:

- είναι η ικανότητα μεταβολής του λογικού σχήματος χωρίς την ανάγκη μεταβολής των προγραμμάτων εφαρμογών
- τέτοιες μεταβολές είναι απαραίτητες όταν μεταβάλλεται η εννοιολογική δομή μιας ΒΔ
- Η λογική ανεξαρτησία είναι δυσκολότερο να επιτευχθεί από τη φυσική ανεξαρτησία, καθώς τα προγράμματα εφαρμογών εξαρτώνται σε μεγάλο βαθμό από τη λογική δομή μιας ΒΔ.
- Η έννοια της ανεξαρτησίας είναι συγγενής με αυτή των αφηρημένων τύπων δεδομένων (abstract data types).

Μοντέλα Δεδομένων (Data Models)

- Ενα **μοντέλο δεδομένων** (data model) είναι ένα σύνολο εννοιών για την περιγραφή των δεδομένων, των σχέσεων μεταξύ αυτών, τη σημασιολογία τους και τους περιορισμούς στους οποίους υπόκεινται.
- Διακρίνονται σε 3 κατηγορίες:
 - **οντοκεντρικά** λογικά μοντέλα (object-oriented)
 - **πλειαδικά** λογικά μοντέλα (record-based)
 - **φυσικά** μοντέλα
- Τα οντοκεντρικά μοντέλα χρησιμοποιούνται για την περιγραφή δεδομένων στο λογικό επίπεδο και το επίπεδο όψεων.
- Τέτοια μοντέλα είναι τα:
 - **μοντελο οντοτήτων-σχέσεων** (entity-relationship model), **σημασιολογικό μοντέλο** (semantic data model), **συναρτησιακό μοντέλο** (functional model)

Μοντέλα Δεδομένων (Data Models)

- Τα πλειαδικά λογικά μοντέλα αναπαριστούν τη λογική δομή μιας ΒΔ μέσω συγκεκριμένων τύπων που αναπαριστώνται ως **πλειάδες** (tuples) ή **εγγραφές** (records).
- Τέτοια μοντέλα είναι τα:
 - **σχεσιακό μοντέλο** (relational model), **ιεραρχικό μοντέλο** (hierarchical model), **μοντέλο δικτύου** (network model)
- Τα φυσικά μοντέλα χρησιμοποιούνται για την περιγραφή των δεδομένων στο φυσικό επίπεδο.
- Τέτοια μοντέλα είναι τα:
 - **ενοποιημένο μοντέλο** (unifying data model)
 - **μοντέλο πλαισίων μνήμης** (frame-memory model)

Γλώσσες Βάσεων Δεδομένων

- Γλώσσες για τον **ορισμό δεδομένων** (data definition languages)
 - χρησιμοποιούνται για τον ορισμό του σχήματος
 - εντολές της γλώσσας ορισμού δεδομένων μεταφράζονται σε ένα σύνολο δομών που αποθηκεύονται στο **λεξικό ή ευρετήριο δεδομένων** (data dictionary or directory)
 - το ευρετήριο δεδομένων περιέχει **μετα-δεδομένα** (metadata), δηλαδή δεδομένα για τα δεδομένα. Το ΣΔΒΔ συμβουλεύεται το ευρετήριο πριν από οποιαδήποτε πρόσβαση στη ΒΔ.
 - μέθοδοι αποθήκευσης και πρόσβασης καθορίζονται σε μια ειδική γλώσσα ορισμού δεδομένων η οποία λέγεται **γλώσσα ορισμού αποθήκευσης δεδομένων** (data storage definition language)

Γλώσσες Βάσεων Δεδομένων

- Γλώσσες για το **χειρισμό δεδομένων** (data manipulation languages)
 - υποστηρίζουν λειτουργίες για **εισαγωγή**, **ανάκληση**, **μεταβολή** και **διαγραφή** πληροφορίας
 - διακρίνονται σε **διεργασιακές** (procedural) και **δηλωτικές** (declarative)
 - διεργασιακές γλώσσες καθορίζουν το «**πως**» γίνεται ο χειρισμός των δεδομένων, ενώ οι δηλωτικές καθορίζουν «**ποιο**» είναι το επιθυμητό αποτέλεσμα
 - οι δηλωτικές γλώσσες είναι ευκολότερες στη χρήση τους αλλά όχι τόσο αποδοτικές
 - η **γλώσσα επερωτήσεων** (query language) είναι υποσύνολο της γλώσσας χειρισμού δεδομένων

Τυπική Αρχιτεκτονική ΣΔΒΔ

