

# Εισαγωγή στα Συστήματα Βάσεων Δεδομένων

# Τι θα δούμε σήμερα

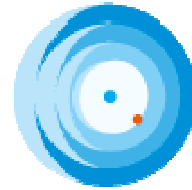
- I. Σύντομη εισαγωγή στις ΒΔ
- II. Περιγραφή σκοπού και περιεχομένου μαθήματος
- III. Ιστορία των ΣΔΒΔ
- IV. Διαδικαστικά θέματα

# Βασικές Έννοιες

Τι είναι μια βάση δεδομένων;

**Βάση Δεδομένων:** συλλογή από σχετιζόμενα δεδομένα

# Που υπάρχουν ΒΔ;



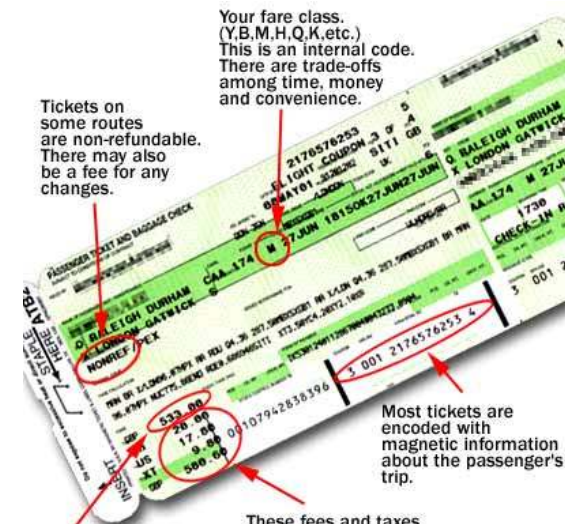
**Εύδοξος**

Ηλεκτρονική Υπηρεσία Ολοκληρωμένης Διαχείρισης  
Συγγραμμάτων και Λοιπών Βοηθημάτων



On-Line Δήλωση Μαθημάτων  
Students Web

amazon.com



Tickets on some routes are non-refundable. There may also be a fee for any changes.

Your fare class. (Y,B,M,H,Q,K, etc.) This is an internal code. There are trade-offs among time, money and convenience.

Fare paid by passenger minus fees and taxes.

These fees and taxes go to the airport, and to local, state and federal governments.

Most tickets are encoded with magnetic information about the passenger's trip.

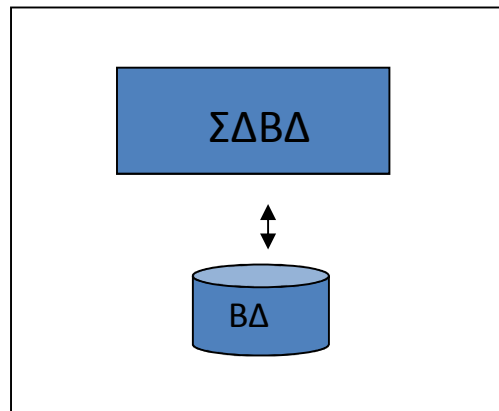
Source: Delta Airlines

© 2001 HowStuffWorks

# Βασικές Έννοιες

**Σύστημα Διαχείρισης Βάσεων Δεδομένων (ΣΔΒΔ) Database Management System (DBMS)**

λογισμικό (σύνολο από προγράμματα) για δημιουργία και χρήση μιας βάσης δεδομένων



**Σύστημα Βάσεων  
Δεδομένων**

# Μερικά ΣΔΒΔ

## Εμπορικά

- Oracle
- IBM/DB2
- MS SQL-server
- Sybase
- Informix
- (MS Access, ...)

## Ελεύθερο Λογισμικό- Open Source

- Postgres (UCB)
- MySQL, mSQL
- miniBase (Wisc)
- Predator (Cornell)
- ...

# Πολλά είδη δεδομένων

The screenshot shows a web browser window with the following elements:

- Address Bar:** `http://maps.yahoo.com/broadband#mvt=m&tp=1&tt=uc+berkeley&trf=0&lon=-122.256203&l` and a search box containing `ncbi genome database`.
- Navigation Bar:** Includes links for .Mac, News (959), Gmail, trumpet, music, p2 devel, R, Swivel, SwivelMail, ebay, PeerGuardian, hibernate, Spell with flickr, and PHI.
- Search Bar:** A search input field with a "Web Search" button.
- Header:** "YAHOO! LOCAL Maps" logo, "Sign In" and "New User? Sign Up" links, and navigation links for "Maps Home", "Dial-Up Map (Original)", and "Help".
- Left Panel:**
  - GET MAP AND DIRECTIONS:** A search input field with "Address, City, State" placeholder.
  - FIND A BUSINESS ON THE MAP:** A search input field containing "uc berkeley" and a "Search" button.
  - Browse by Category:** A list of search results for "uc berkeley" with checkboxes and a "Refine" button.
- Map Area:**
  - Map controls: "Printable Version", "Send", "Save", "Live Traffic", and "Help".
  - Map style selector: "Map", "Hybrid", "Satellite".
  - Map content: A street map of Berkeley, CA, with 24 numbered orange markers (1-24) indicating search results. The University of California Berkeley campus is visible in the background.

# Πολλά είδη δεδομένων

The screenshot displays the NCBI Genome browser interface. At the top, the NCBI logo is on the left, and navigation tabs for PubMed, Nucleotide, Protein, Genome, Gene, Structure, PopSet, Taxonomy, and Help are on the right. Below the tabs is a search bar with the text "Search for" and a dropdown menu set to "on chromosome(s)". To the right of the search bar are buttons for "assembly", "All", "Find", and "Advanced Search". Below the search bar are links for "Show related entries", "Help", "FTP", and "Map Viewer home".

The main content area is titled "Homo sapiens genome view" and includes a link for "Build 35.1 statistics". To the right of this title is a link for "BLAST search the human genome". Below the title is a karyotype visualization showing 22 pairs of autosomes and the X and Y chromosomes, labeled 1 through 22, X, Y, and HI.

On the left side of the interface, there is a sidebar with several sections:

- Entrez Genomes**
  - MapViewer Home
- Map Viewer**
  - Help
  - Human Maps Help
  - Mouse Maps Help
  - NCBI Handbook
- Related Resources**
  - Human Genome Guide
  - Genomic Biology
  - Gene
  - OMIM
  - UniGene
- Sequence Data**
  - Human Genome Sequencing
  - Mouse Genome Sequencing
  - RefSeq

At the bottom of the page, a taxonomic lineage is displayed: **Lineage: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo; Homo sapiens.**



# Γιατί ΣΔΒΔ;

ΒΔ υπάρχουν παντού

- Κοινά θέματα – τι προσφέρουν:

- ✓ μοντελοποίηση, έλεγχος πλεονασμού, περιορισμοί ακεραιότητας

- ✓ αποδοτική επεξεργασία ερωτήσεων (ευρετήρια, βελτιστοποίηση)

- ✓ ορθότητα λειτουργίας: Πως θα διασφαλίσουμε την ορθότητά τους κατά τη διάρκεια αποτυχιών και ταυτόχρονης προσπέλασης από πολλούς χρήστες, ανάκαμψη από σφάλματα

- ✓ Επίσης: θέματα ασφάλειας, δικαιωμάτων/εξουσιδότηση προσπέλασης προσπέλασης, ...

# Γιατί ΣΔΒΔ;

Γιατί χρειαζόμαστε ειδικό λογισμικό;

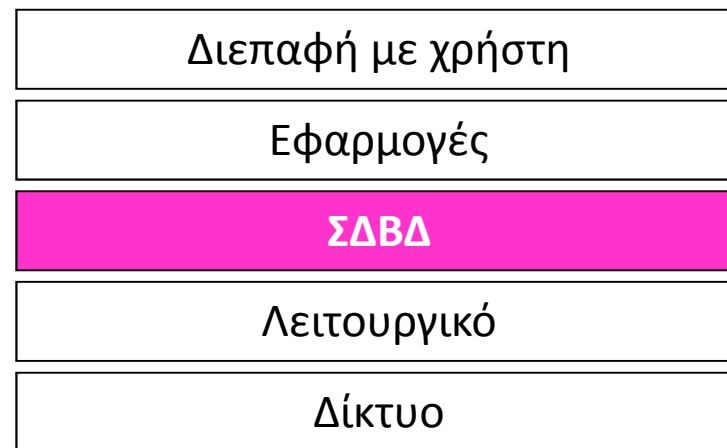
- Κοινή λειτουργικότητα ήδη υλοποιημένη
- Σωστή υλοποίηση
- Αποδοτική προσπάθεια
- Γρήγορη ανάπτυξη εφαρμογών
- Ανεξαρτησία δεδομένων (θα δούμε περισσότερα)
- Μόνιμη αποθήκευση

# Γιατί όχι ΣΔΒΔ;

- Επένδυση σε λογισμικό και υλικό, καθώς και για εκπαίδευση
- Η γενικότητα που παρέχει προκαλεί χρονική επιβάρυνση (overhead)
- Δε χρειάζονται όσα προσφέρει

# Βασικές Έννοιες

Η θέση των ΣΔΒΔ στη στοίβα του λογισμικού συστημάτων



# Σκοπός του Μαθήματος

Τρεις Βασικοί Στόχοι:

1. Να μπορείτε να σχεδιάσετε και υλοποιήσετε ένα σύστημα βάσεων δεδομένων χρησιμοποιώντας ένα ΣΔΒΔ

Τι σημαίνει αυτό:

- Μοντελοποίηση
- Προγραμματισμός (σε SQL)

Καθώς και τη *σχετική θεωρία* στην οποία τα παραπάνω βασίζονται (κανονικός μορφές, σχεσιακή άλγεβρα)

# Σκοπός του Μαθήματος

2. Θέματα υλοποίησης ενός ΣΔΒΔ (το εσωτερικό του)
3. Γενικές τεχνικές/αρχές/αλγορίθμους που διέπουν τη διαχείριση δεδομένων

*Ας δούμε τα παραπάνω λίγο πιο αναλυτικά*

# Μια γενική εικόνα: Δημιουργία ΣΒΔ

## **ΒΗΜΑ 1:** Μοντελοποίηση

- Εννοιολογικό Μοντέλο (Μοντέλο Οντοτήτων/Συσχετίσεων)
- Μοντέλο Υλοποίησης (Σχεσιακό μοντέλο)

## **ΒΗΜΑ 2:** Προγραμματισμός/Υλοποίηση

# Μοντελοποίηση

**Μοντέλο Δεδομένων:** ένα σύνολο από έννοιες (δομικά στοιχεία) που μπορούν να χρησιμοποιηθούν για την περιγραφή της δομής της πληροφορίας

Σχήμα (database schema): η περιγραφή της δομής της πληροφορίας που είναι αποθηκευμένη στη βδ με τη χρήση ενός μοντέλου δεδομένων



# Μοντελοποίηση

- Υψηλού επιπέδου (εννοιολογικά) μοντέλα

Υψηλού επιπέδου, περισσότερο αφηρημένη περιγραφή της δομής

Μοντέλο Οντοτήτων/Συσχετίσεων

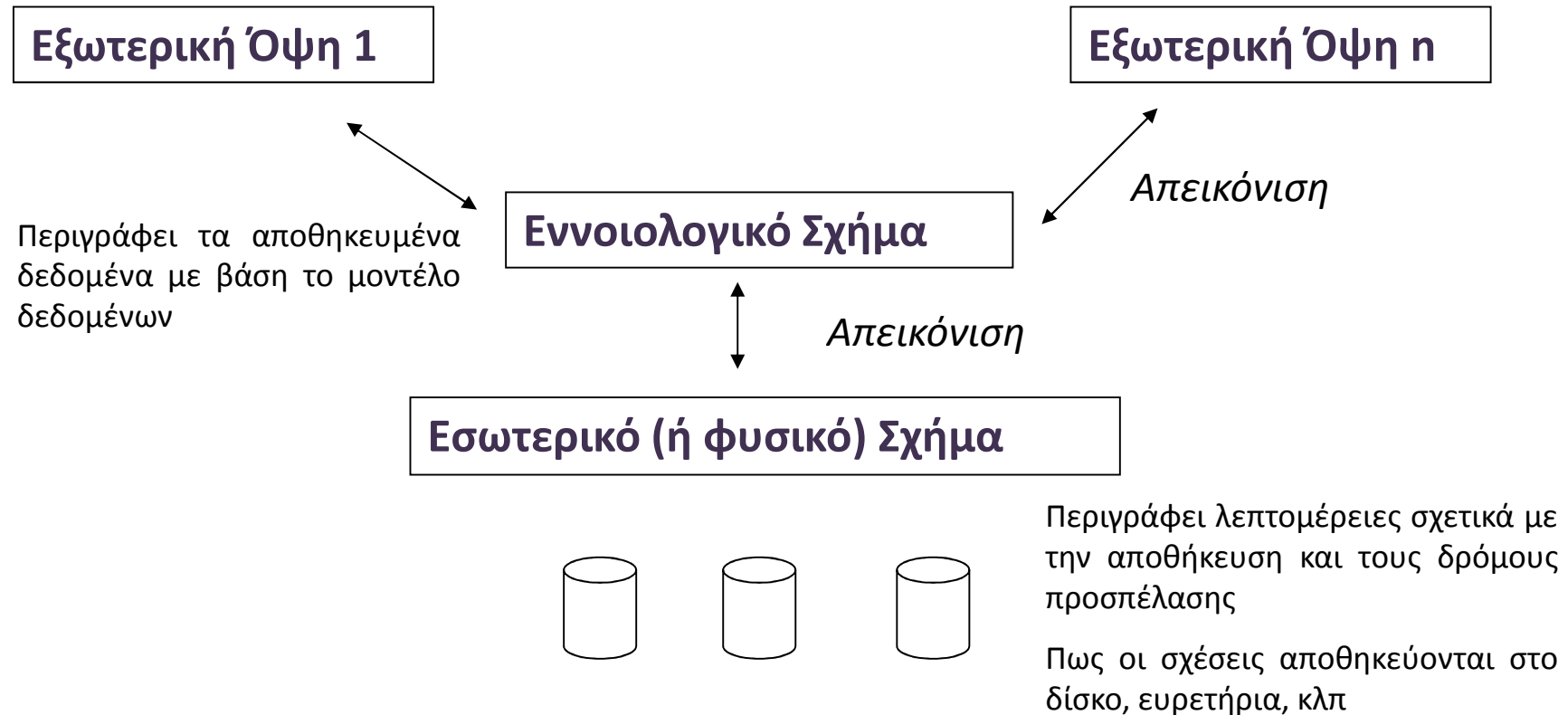
- Παραστατικά μοντέλα ή μοντέλα υλοποίησης ή λογικά μοντέλα

Σχεσιακό Μοντέλο, Ιεραρχικό Μοντέλο, Δικτυωτό Μοντέλο

- Χαμηλού επιπέδου ή φυσικά μοντέλα

Δρόμος Προσπέλασης

# Αρχιτεκτονική Τριών Επιπέδων



- Η περιγραφή της βάσης δεδομένων περιλαμβάνει ένα σχήμα για καθένα από τα επίπεδα αφαίρεσης

# Ανεξαρτησία Δεδομένων

**Ανεξαρτησία Δεδομένων:** αλλαγή του σχήματος ενός επιπέδου χωρίς να αλλάξουμε το σχήμα του αμέσως υψηλότερου επιπέδου

- *Λογική Ανεξαρτησία Δεδομένων*

αλλαγή του εννοιολογικού δεν επηρεάζει τα εξωτερικά σχήματα ή τα προγράμματα εφαρμογών

- *Φυσική Ανεξαρτησία Δεδομένων*

αλλαγή του εσωτερικού σχήματος χωρίς να χρειάζεται αλλαγή του εννοιολογικού

αλλαγή μόνο της απεικόνισης

# Σχήμα και Στιγμιότυπο

Σχήμα της Βάσης

**Πρόθεση (intension)**

Μοντέλο (δομικά στοιχεία + περιορισμοί ακεραιότητας)

**Ανάπτυξη (extension)**

Στιγμιότυπο της Βάσης (κατάσταση ή σύνολο εμφανίσεων ή σύνολο στιγμιοτύπων)

(αρχική κατάσταση, έγκυρη κατάσταση)

# Δημιουργία ΣΒΔ

**ΒΗΜΑ 1:** Μοντελοποίηση

**ΒΗΜΑ 2:** Προγραμματισμός/Υλοποίηση

- Ορισμός Σχέσεων (πρόθεση/σχήμα)
- Εισαγωγή Στοιχείων (δημιουργία του αρχικού στιγμιότυπου)
- Διατύπωση Ερωτήσεων
- Τροποποίηση Βάσης Δεδομένων

# Γλώσσες ΣΔΒΔ (SQL)

## Γλώσσα Ορισμού

**create table R(A1 T1, A2, T2, ...)**

## Γλώσσα Χειρισμού/Επεργασίας Δεδομένων

εισαγωγή, διαγραφή, τροποποίηση δεδομένων) **και**  
**ερωτημάτων** (ανάκτηση δεδομένων)

**insert/delete/update**

**select** *Γνωρίσματα*

**from** *Πίνακες*

**where** *Συνθήκη*

# Γλώσσες ΣΔΒΔ

## Γλώσσες Ερωτήσεων (Query Languages)

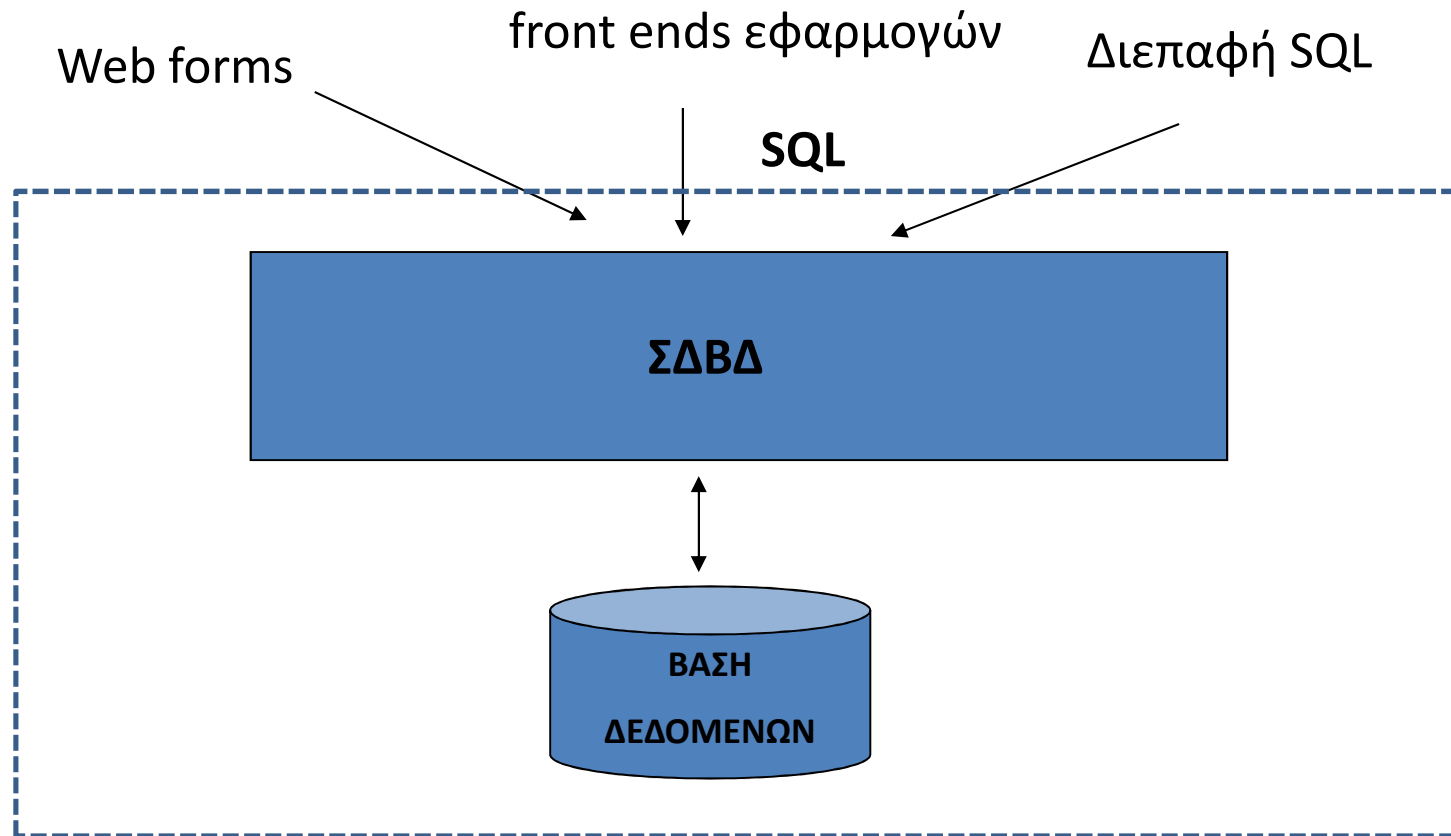
- δυνατότητα εμφύτευσης σε μια γλώσσα υψηλού επιπέδου
- μίας εγγραφής τη φορά ή συνόλου τη φορά
- διαδικαστικές και μη διαδικαστικές (δηλωτικές)

# Χρήστες

- Απλοί Χρήστες
- Προγραμματιστές εφαρμογών
- Σχεδιαστές βάσεων δεδομένων
- Διαχειριστές συστήματος
- Δημιουργοί ΣΔΒΔ



# Διεπαφές



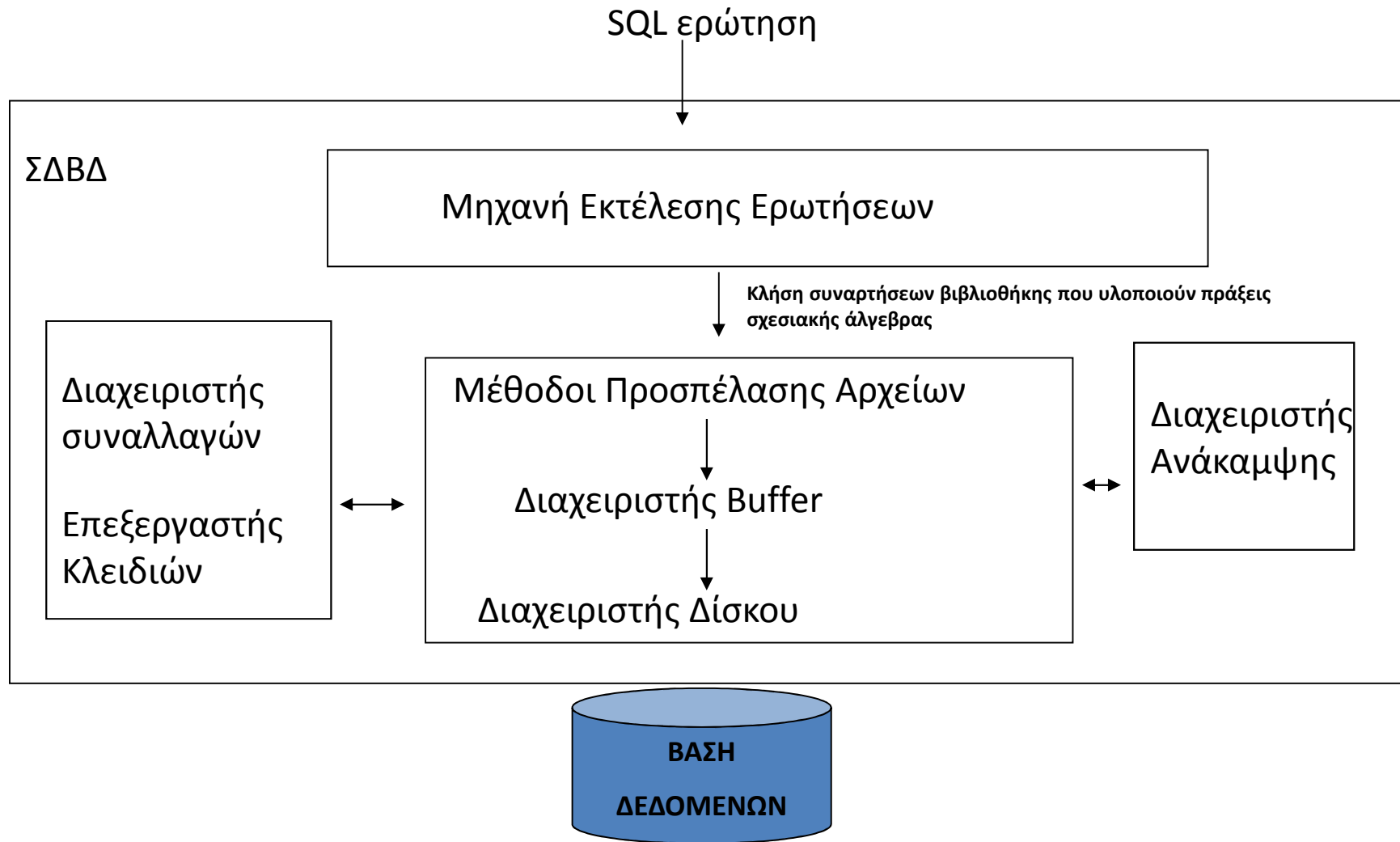
- Βασιζόμενες σε μενού (κατάλογο από επιλογές), Γραφικών, Βασιζόμενες σε φόρμες, Φυσικής γλώσσας, Για παραμετρικούς χρήστες, Για το ΔΒΔ

# Βασικές Έννοιες

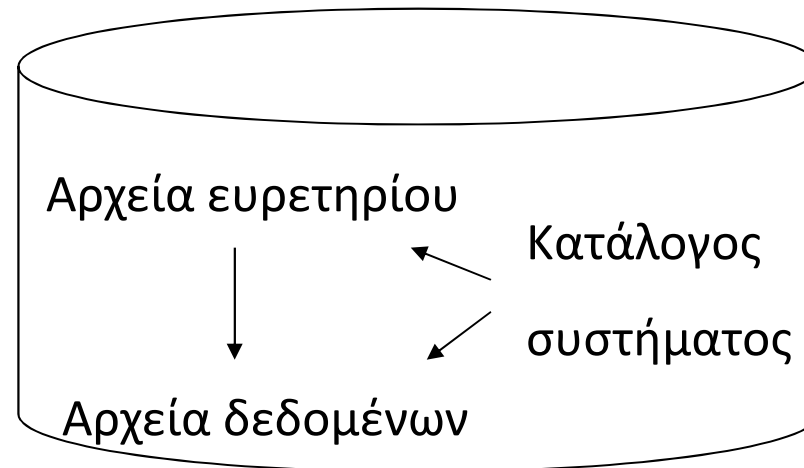
## Κάποιες λειτουργίες ενός ΣΔΒΔ

- **Ορισμός και Δημιουργία** μιας βάσης δεδομένων: προδιαγραφή των τύπων, των δομών και των περιορισμών των δεδομένων που θα αποθηκευτούν στη ΒΔ
- **Χειρισμός** (manipulation) μιας βάσης δεδομένων: υποβολή ερωτήσεων για την ανάκτηση δεδομένων, ενημέρωση (εισαγωγές, διαγραφές ή τροποποιήσεις)
- **Άλλες λειτουργίες**: Διαμοιρασμός, προστασία από αστοχίες υλικού και λογισμικού, ασφάλεια, ρύθμιση (tuning)

# Το εσωτερικό ενός ΣΔΒΔ



# Το εσωτερικό ενός ΣΔΒΔ



ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ

# Ιστορία

## **Δεκαετία του 1950**

Κάρτες και ταινίες (σειριακή επεξεργασία) – Batch processing

## **Αρχή του 1960**

Γενικευμένη χρήση δίσκων

πρώτο γενικού-σκοπού ΣΔΒΔ: Integrated Data Store (GE)

Charles Bachman (Recipient of the 1<sup>st</sup> Turing Award, 1973)

network data model (δικτυωτό)

## **Τέλη του 1960**

Information Management System (IMS) IBM

hierarchical data model (Ιεραρχικό)

SABRE Airline Reservation System (AA+IBM, travelocity!!)

# Ιστορία

## 1970

Edgar Codd (IBM, San Jose) σχεσιακό μοντέλο δεδομένων (relational data model)

(Recipient of the **Turing Award**, 1981)

Ερευνητικά Προγράμματα: System R, INGRES - Γλώσσες: SEQUEL, QBE, QUEL

## Δεκαετία του 1980

SQL (μέρος του System R)

transaction management (Jim Gray, **Turing Award**, 1999)

υποσημείωση: Jim Gray gone missing

[*Τάσεις: αντικειμενοστραφή, αρχιτεκτονική πελάτη-εξυπηρετή, κατανεμημένες, έμπειρα*]

# Ιστορία

## Δεκαετία του 1990

εμπορικά αντικειμενοστραφή συστήματα

[Τάσεις: πολυβάσεις, χωρικές & χρονικές, πολυμέσα, συμπερασματικές, αποθήκες δεδομένων (αναλυτική επεξεργασία), προγραμματισμό πόρων της επιχείρησης (ERP – Enterprise Resource Planning) και της διαχείρισης τους (MRP – Management Resource Planning), Internet]

## Δεκαετία του 2000

Σύστημα Διαχείρισης Χρωμοσωμάτων (Human Genome Project)

Σύστημα Παρατήρησης της Γης (Earth Observation System)

# Ιστορία

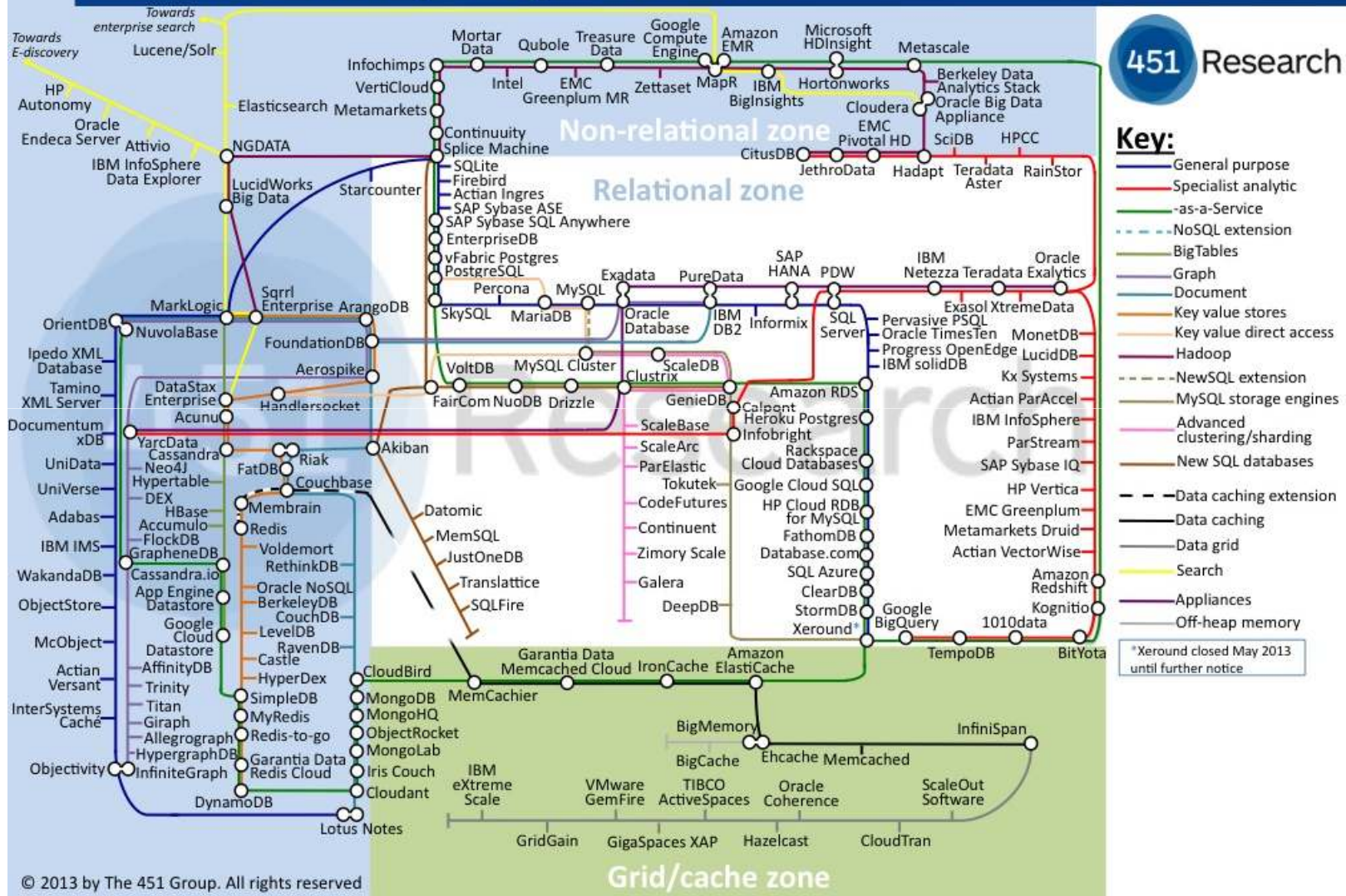
## Σήμερα

- Μεγάλος όγκος δεδομένων (BIG DATA)
- Αλλαγές σε υλικό (επεξεργαστές με πολλούς πυρήνες, ιεραρχία αποθήκευσης)
- cloud computing, software as service

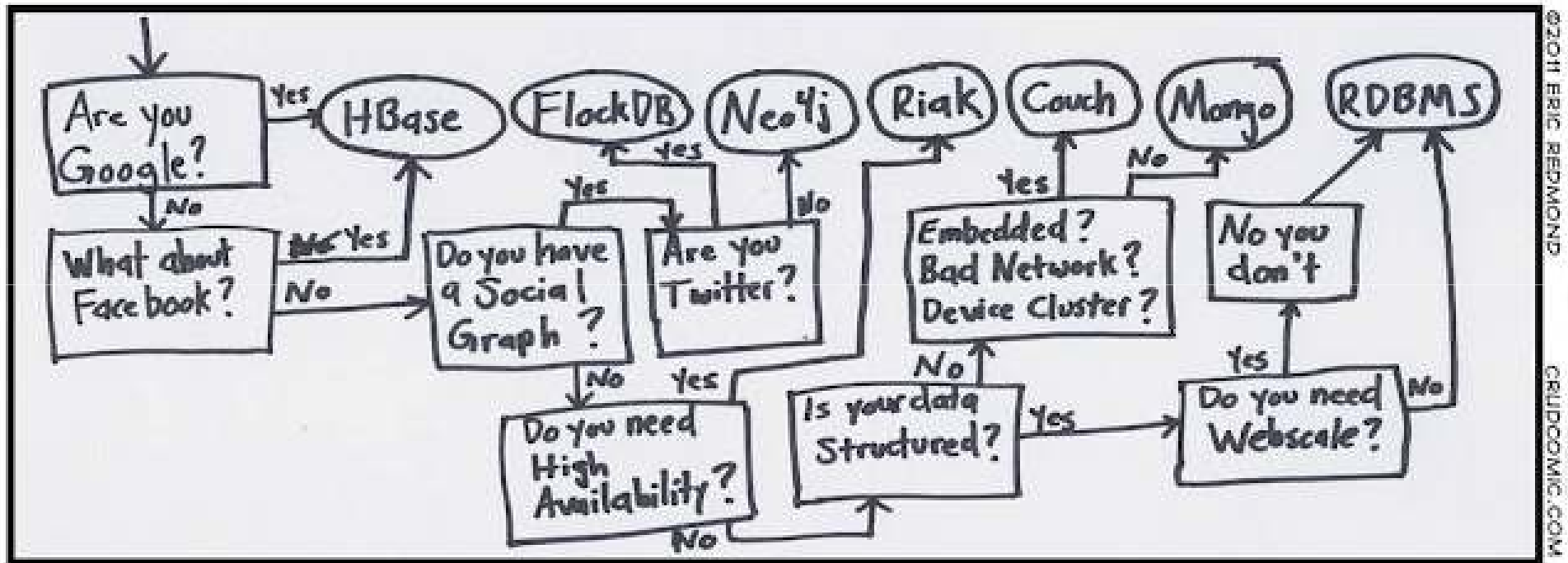
## *Κίνηση NoSQL*



# Database Landscape Map – June 2013



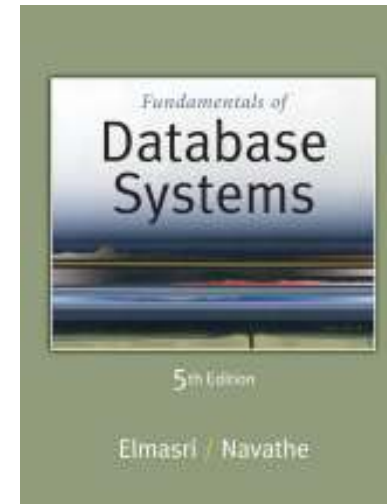
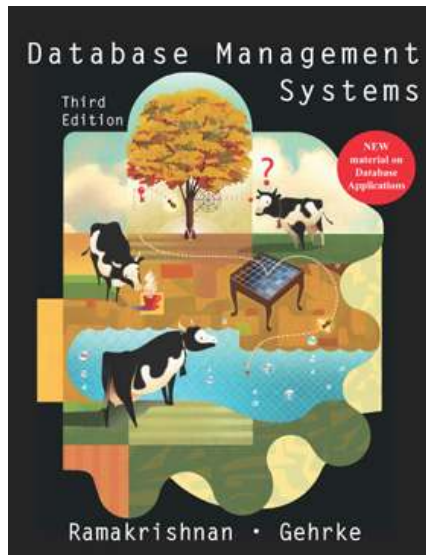
# Τι ΣΔΒΔ να χρησιμοποιήσω;



# Διαχειριστικά Θέματα

- web σελίδα <http://www.cs.uoi.gr/~pitoura>

«Θεμελιώδεις Αρχές Συστημάτων Βάσεων Δεδομένων», Elmasri&Navathe



«Συστήματα Διαχείρισης Βάσεων Δεδομένων»  
Ramakrishnan&Gehrke

Hank Korth, Avi Silberschatz, and S. Sudarshan, *Database System Concepts*, 5th Edition, McGraw-Hill, 2005.

# Διαχειριστικά Θέματα

## Βαθμός

- Ασκήσεις (3 - 4 σύνολα) + Μια προγραμματιστική άσκηση ( $\geq 4.0$ )
- Τελικό διαγώνισμα ( $\geq 4.0$ )
- Τελικός Βαθμός ( $\geq 5.0$ )
- 35% \* (Βαθμός Ασκήσεων+Προγραμματιστικής) +  
65% \* Βαθμός Τελικού Διαγωνίσματος

Όσοι έδωσαν ασκήσεις το προηγούμενο Ακαδημαϊκό Έτος ( Ακαδ. Έτος 2012-2013) μπορούν να «κρατήσουν» το βαθμό τους

Οι υπόλοιποι **πρέπει** να τις επαναλάβουν

1. ΔΕΝ υπάρχει «μεθοδολογία ασκήσεων»
2. Καλό θα είναι να παρακολουθείτε το μάθημα (τις διαλέξεις, αλλά και το «ρυθμό» του)
3. Η ύλη/σειρά στο βιβλίο μπορεί να διαφέρει από το μάθημα – αλλά ό,τι πούμε στο μάθημα και ό,τι υπάρχει στα σχετικά κεφάλαια του βιβλίου αρκεί για να «περάσετε» το μάθημα – αυτό ισχύει και για τα δύο προτεινόμενα βιβλία
4. Οι διαφάνειες ΔΕΝ αντικαθιστούν το βιβλίο (είναι συμπληρωματικές σε αυτό)
5. Κάποιες ασκήσεις λύνονται ΜΟΝΟ στον πίνακα σκόπιμα

Και αν δεν περάσω  
το μάθημα;

## HOW TO WRITE A CV



Leverage the NoSQL boom

# Ερωτήσεις;