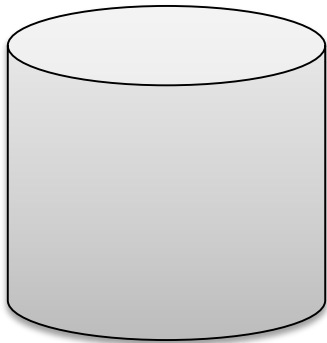


ΜΥΥ701: Βάσεις Δεδομένων

Διδάσκουσα: Ευαγγελία Πιτουρά



Εισαγωγή στα Συστήματα Βάσεων Δεδομένων

Τι θα δούμε σήμερα

- I. Τι είναι οι ΒΔ, γιατί μας ενδιαφέρουν
- II. Στόχος και περιεχόμενο του μαθήματος
- III. Ιστορία των ΒΔ
- IV. Διαδικαστικά θέματα

Βασικές Έννοιες

Τι είναι μια βάση δεδομένων;

Βάση Δεδομένων: συλλογή από σχετιζόμενα δεδομένα

Γιατί να μας ενδιαφέρουν;

Τα δεδομένα είναι παντού

Που υπάρχουν ΒΔ;



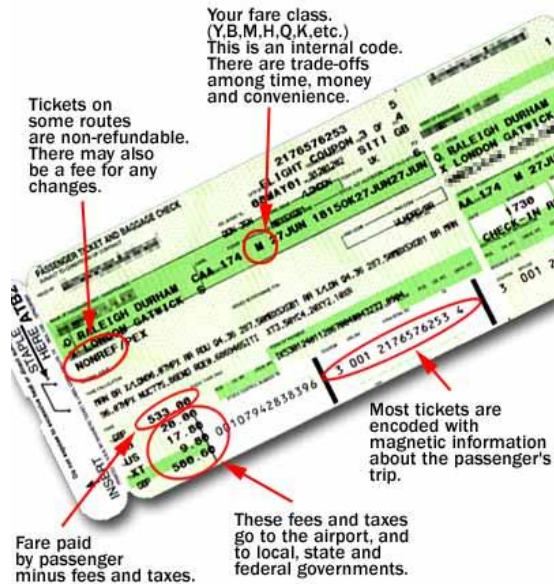
Εύδοξος

Ηλεκτρονική Υπηρεσία Ολοκληρωμένης Διαχείρισης
Συγγραμμάτων και Λοιπών Βοηθημάτων



On-Line Δήλωση Μαθημάτων
Students Web

Chronus



Source: Delta Airlines

© 2001 HowStuffWorks

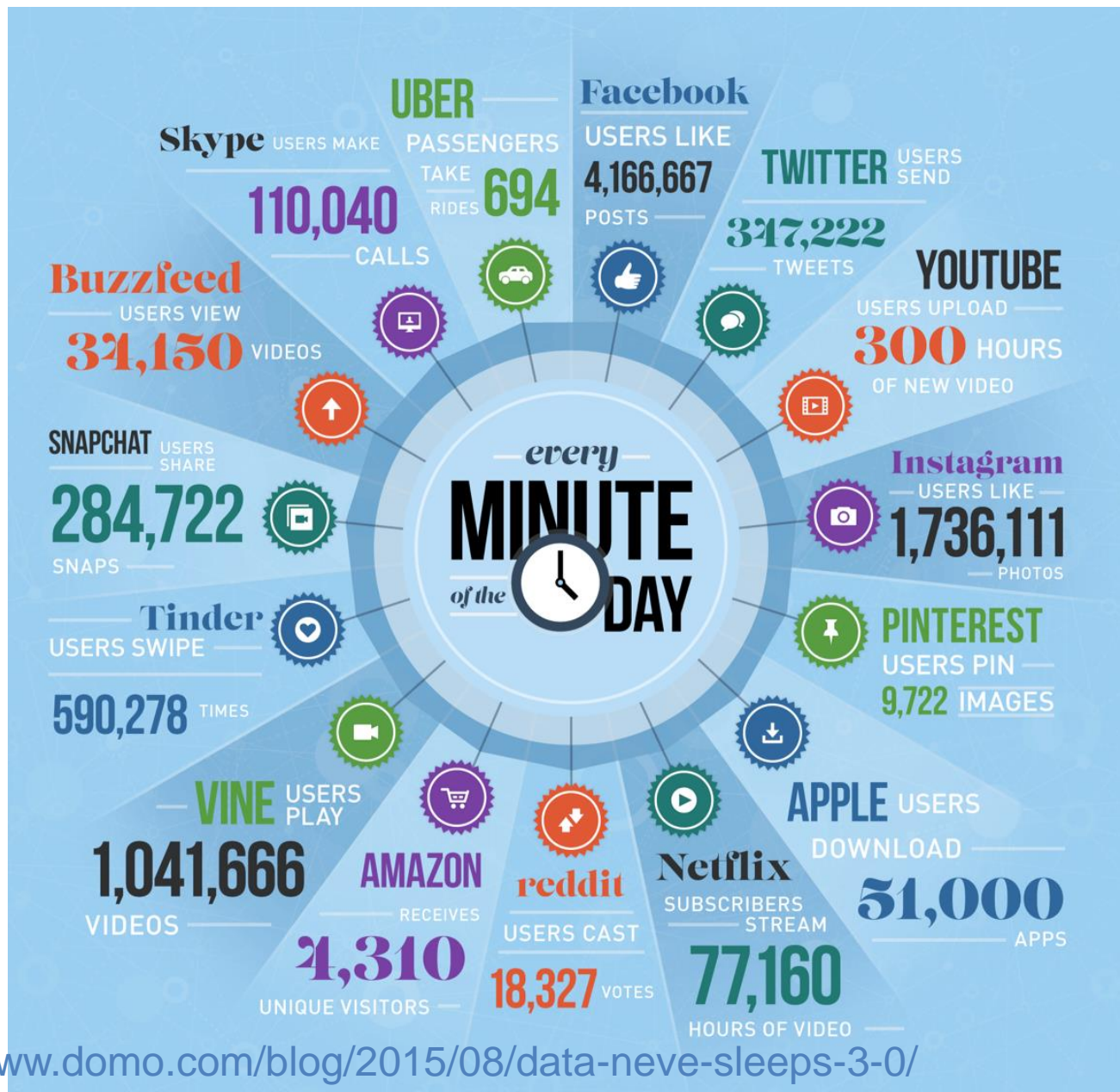


Που υπάρχουν ΒΔ;

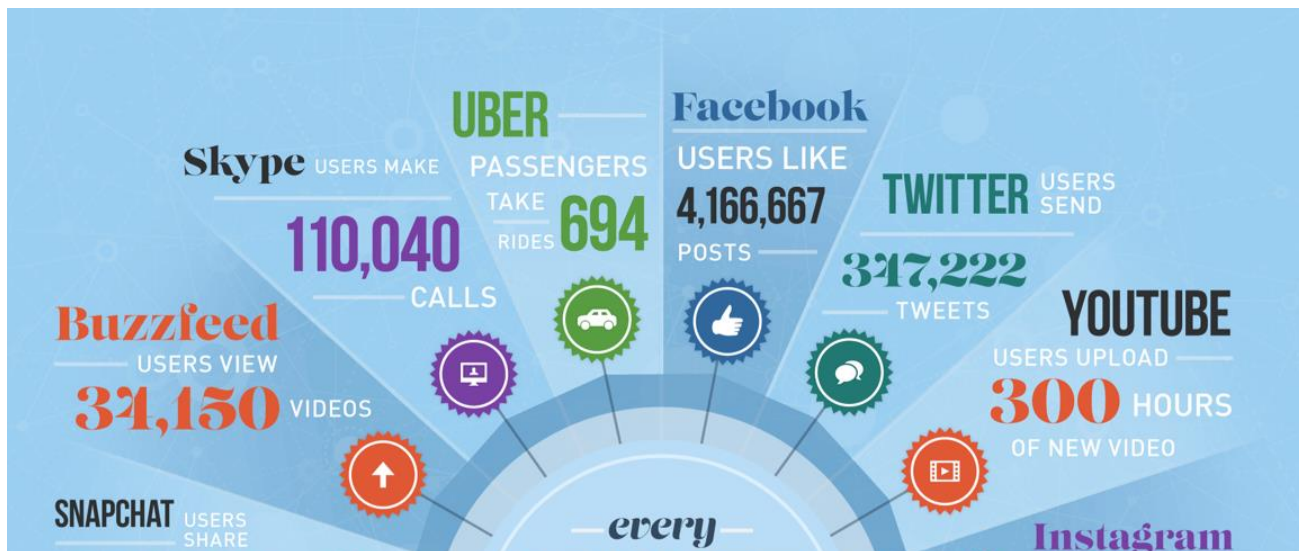
amazon.com.

 tripadvisor®





<https://www.domo.com/blog/2015/08/data-neve-sleeps-3-0/>



THE GLOBAL INTERNET POPULATION GREW **18.5%** FROM 2013-2015 AND NOW REPRESENTS

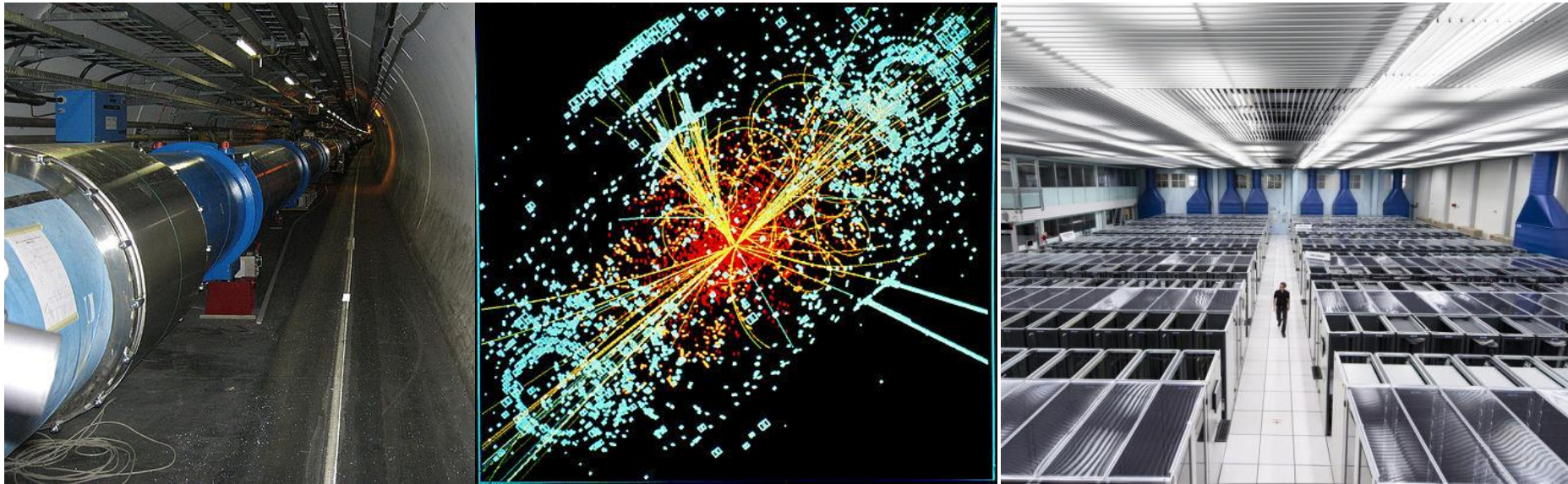
3.2 BILLION PEOPLE.



<https://www.domo.com/blog/2015/08/data-never-sleeps-3-0/>

Scientific Data (Επιστημονικά Δεδομένα)

Large Hadron Collider (LHC)



- 150M αισθητήρες @ 40M μετρήσεις το δευτερόλεπτο: 6 Peta-γεγονότα το δευτερόλεπτο
- Μαζικό φιλτράρισμα → 700 MB το δευτερόλεπτο → 15 PB το χρόνο
- Η συλλογή δεδομένων περιορίζεται για λόγους υπολογισμού και αποθήκευσης

<https://lhcb-public.web.cern.ch/lhcb-public/en/Data%20Collection/Triggers2-en.html>

“BIG DATA”

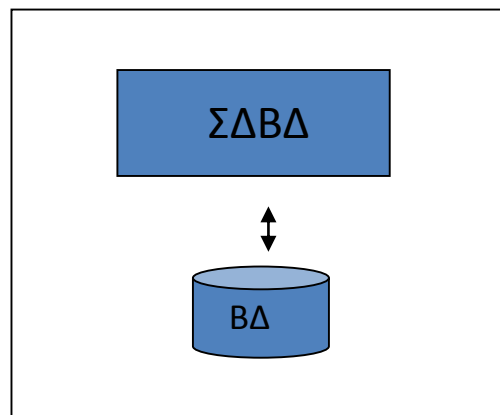
- Πανταχού παρόντες αισθητήρες και καταγραφείς
 - Κάμερες, κινητά, blogging, ...
- Μεγάλα επιστημονικά έργα (βιολογία, αστρονομία, φυσική, ιατρική,)

Αποθήκευση όλων των δεδομένων

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**

Γιατί ΣΔΒΔ;

ΒΔ υπάρχουν παντού

- Κοινά θέματα – τι προσφέρουν:
 - ✓ μοντελοποίηση, έλεγχος πλεονασμού, περιορισμοί ακεραιότητας
 - ✓ αποδοτική επεξεργασία ερωτήσεων (queries) (ευρετήρια, βελτιστοποίηση)
 - ✓ μόνιμη αποθήκευση (persistent storage)
 - ✓ ενημέρωση δεδομένων
 - ✓ 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/εξουσιοδότηση 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



Μερικά (σχεσιακά) ΣΔΒΔ

Εμπορικά

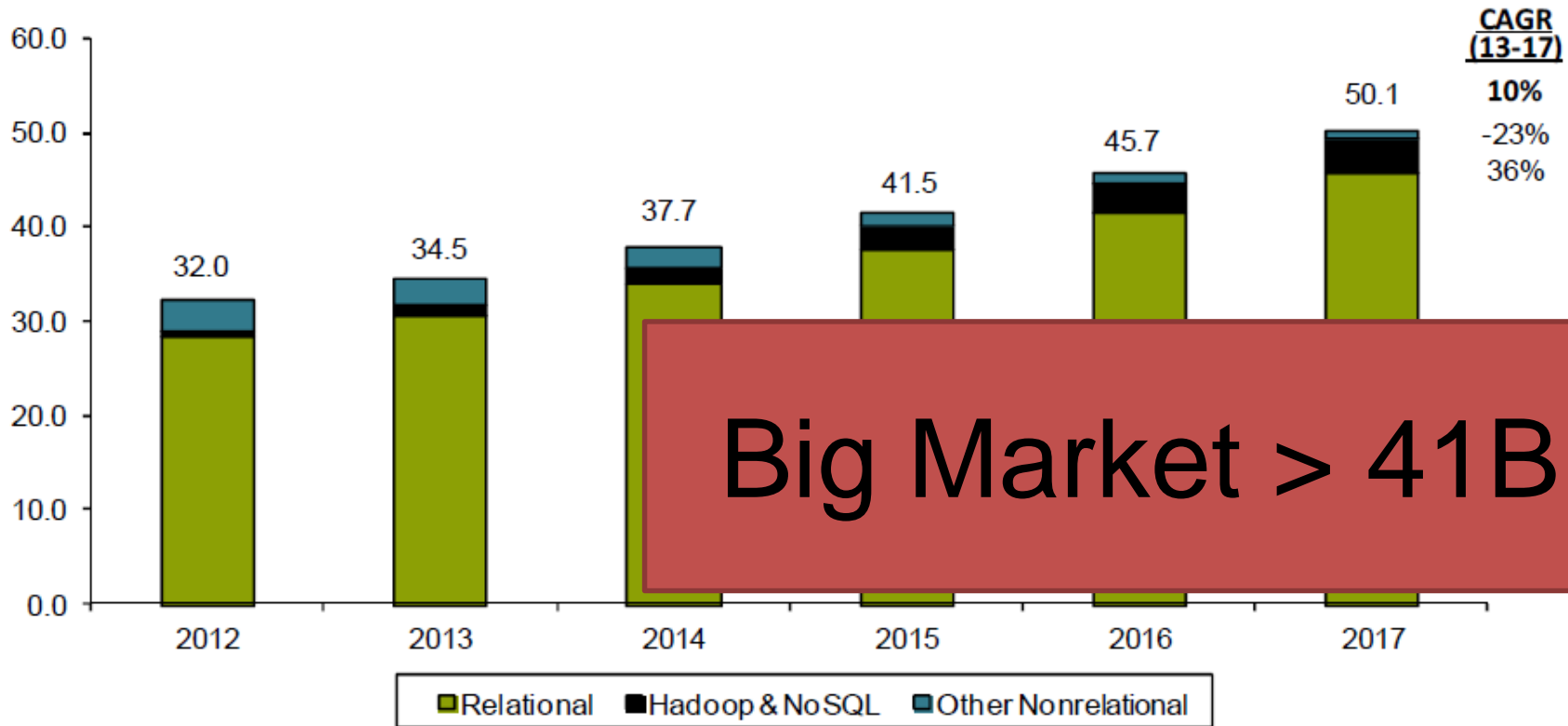
- Oracle
- IBM/DB2
- MS SQL-server
- Sybase
- Informix
- (MS Access, ...)

Ελεύθερο Λογισμικό- Open Source

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

ΣΔΒΔ

Global Database Market (\$B)



Source: IDC, Bernstein analysis

Κατάταξη τεχνολογιών ΣΔΒΔ (2017)

DB Engines Ranking

334 systems in ranking, October 2017

Rank			DBMS	Database Model	Score		
Oct 2017	Sep 2017	Oct 2016			Oct 2017	Sep 2017	Oct 2016
1.	1.	1.	Oracle +	Relational DBMS	1348.80	-10.29	-68.30
2.	2.	2.	MySQL +	Relational DBMS	1298.83	-13.78	-63.82
3.	3.	3.	Microsoft SQL Server +	Relational DBMS	1210.32	-2.23	-3.86
4.	4.	↑ 5.	PostgreSQL +	Relational DBMS	373.27	+0.91	+54.58
5.	5.	↓ 4.	MongoDB +	Document store	329.40	-3.33	+10.60
6.	6.	6.	DB2 +	Relational DBMS	194.59	-3.75	+14.03
7.	7.	↑ 8.	Microsoft Access	Relational DBMS	129.45	+0.64	+4.78
8.	8.	↓ 7.	Cassandra +	Wide column store	124.79	-1.41	-10.27
9.	9.	9.	Redis +	Key-value store	122.05	+1.65	+12.51
10.	10.	↑ 11.	Elasticsearch +	Search engine	120.23	+0.23	+21.12

Βασισμένη σε #mentions (π.χ., stack overflow), google trends, job postings, profile data στο LinkedIn, tweets ...

<http://db-engines.com/en/ranking>

Κατάταξη τεχνολογιών ΣΔΒΔ (2016)

DB Engines Ranking

292 systems in ranking, January 2016

Rank			DBMS	Database Model	Score		
Jan 2016	Dec 2015	Jan 2015			Jan 2016	Dec 2015	Jan 2015
1.	1.	1.	Oracle	Relational DBMS	1496.08	-1.47	+56.92
2.	2.	2.	MySQL	Relational DBMS	1299.26	+0.72	+21.75
3.	3.	3.	Microsoft SQL Server	Relational DBMS	1144.06	+20.90	-54.55
4.	4.	↑ 5.	MongoDB +	Document store	306.03	+4.64	+55.13
5.	5.	↓ 4.	PostgreSQL	Relational DBMS	282.40	+2.31	+27.91
6.	6.	6.	DB2	Relational DBMS	196.37	+0.24	-3.76
7.	7.	7.	Microsoft Access	Relational DBMS	134.04	-6.17	-5.10
8.	8.	8.	Cassandra +	Wide column store	130.95	+0.11	+32.20
9.	9.	9.	SQLite	Relational DBMS	103.74	+2.89	+7.54
10.	10.	10.	Redis +	Key-value store	101.16	+0.62	+6.92

Βασισμένη σε #mentions (π.χ., stack overflow), google trends, job postings, profile data στο LinkedIn, tweets ...

<http://db-engines.com/en/ranking>

Σκοπός του μαθήματος

Τρεις βασικοί στόχοι:

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

Τι σημαίνει αυτό:

- Μοντελοποίηση
- Προγραμματισμός (σε SQL)


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

Σκοπός του μαθήματος

2. Θέματα υλοποίησης ενός ΣΔΒΔ (το εσωτερικό του)
 - Βασικές δομές δεδομένων για προσπέλαση δεδομένων από το δίσκο (ευρετήρια)
 - Βασικές αρχές βελτιστοποίησης ερωτήσεων

3. Γενικές γνώσεις και δεξιότητες για τη διαχείριση δεδομένων

Τι άλλο θα δούμε σήμερα

- I. Τι είναι οι ΒΔ, γιατί μας ενδιαφέρουν
- II. Στόχος και *περιεχόμενο* του μαθήματος
- III. Ιστορία των ΒΔ 
- IV. Διαδικαστικά θέματα

Μια γενική εικόνα: δημιουργία ΣΒΔ

ΒΗΜΑ 1: Μοντελοποίηση

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

ΒΗΜΑ 2: Προγραμματισμός/Υλοποίηση

Μοντελοποίηση

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

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

Παράδειγμα (Οντότητες-Συσχετίσεις)

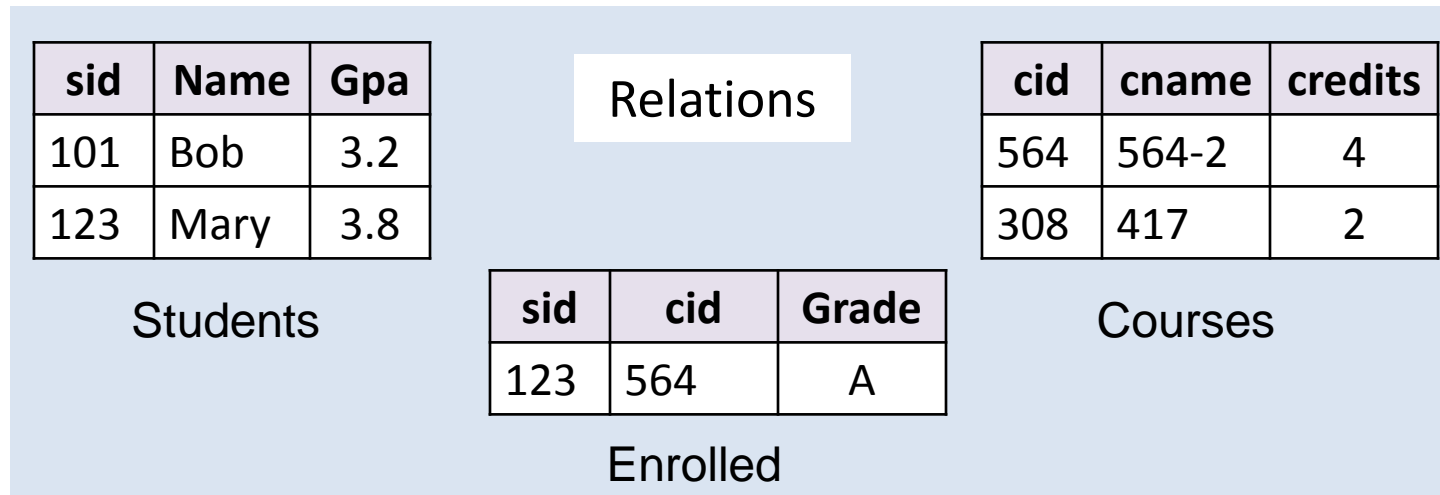
- ΣΔΒΔ για μαθήματα

- Φοιτητές
 - Μαθήματα
 - Καθηγητές
- } Οντότητες

- Ποιος φοιτητής παρακολουθεί ποιο μάθημα
 - Ποιος καθηγητής διδάσκει ποιο μάθημα
- } Συσχετίσεις

Παράδειγμα (Σχέσεις)

- Students(*sid*: *string*, *name*: *string*, *gpa*: *float*)
- Courses(*cid*: *string*, *cname*: *string*, *credits*: *int*)
- Enrolled(*sid*: *string*, *cid*: *string*, *grade*: *string*)



Μοντελοποίηση

- Υψηλού επιπέδου (εννοιολογικά) μοντέλα

Υψηλού επιπέδου, περισσότερο αφηρημένη περιγραφή της δομής

Μοντέλο Οντοτήτων/Συσχετίσεων

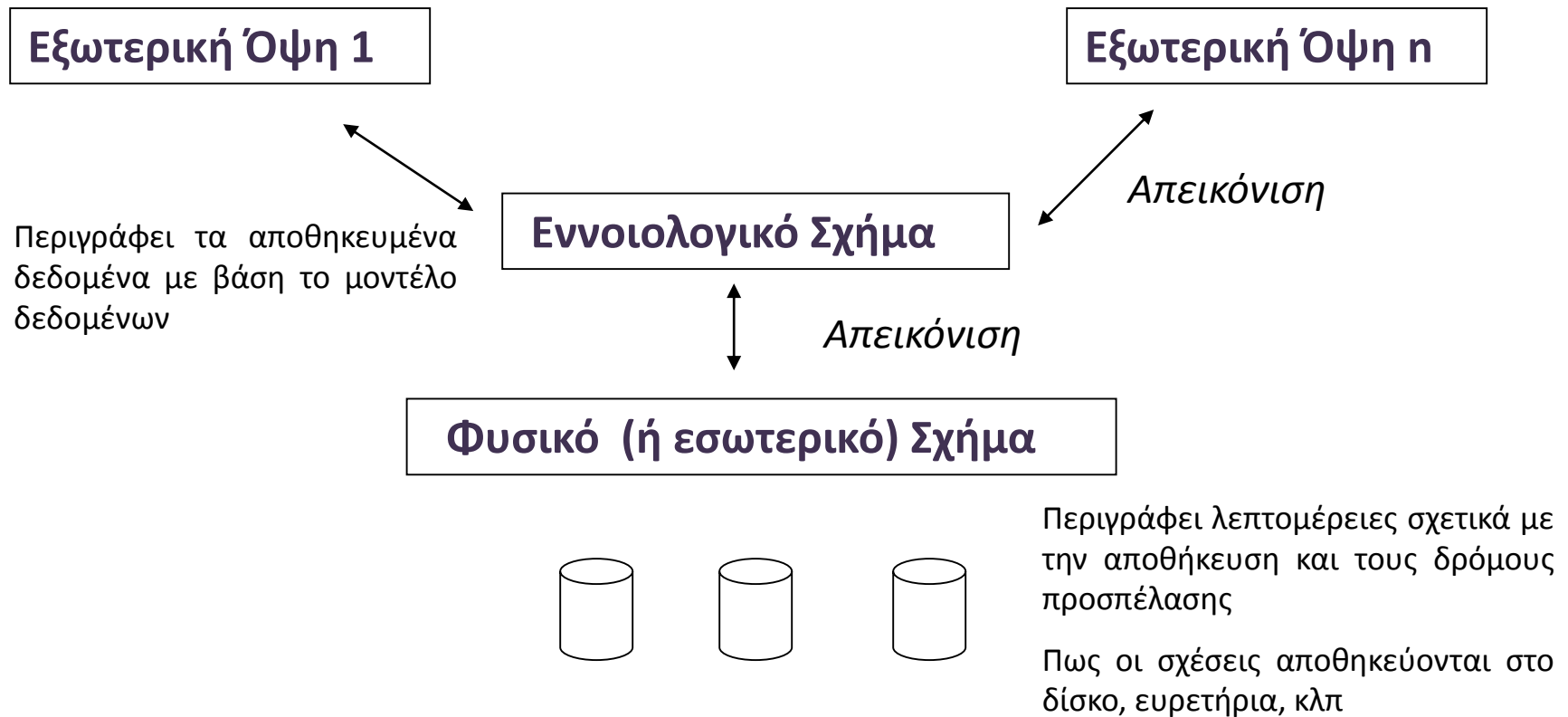
- Παραστατικά μοντέλα ή μοντέλα υλοποίησης ή λογικά μοντέλα

Σχεσιακό Μοντέλο, Ιεραρχικό Μοντέλο, Δικτυωτό Μοντέλο

- Χαμηλού επιπέδου ή φυσικά μοντέλα

Μοντέλα αποθήκευσης

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 βδ

Πρόθεση (intension)

Μοντέλο (δομικά στοιχεία + περιορισμοί ακεραιότητας)

Ανάπτυξη (extension)

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

(αρχική κατάσταση, έγκυρη κατάσταση)

Δημιουργία ΣΒΔ

ΒΗΜΑ 1: Μοντελοποίηση (ορισμός σχήματος)

ΒΗΜΑ 2: Υλοποίηση

Χρήση ΣΔΒΔ:

- Ορισμός μοντέλου
- Εισαγωγή στοιχείων (δημιουργία του αρχικού στιγμιότυπου)
- Διατύπωση ερωτήσεων
- Εισαγωγή/διαγραφή/τροποποίηση δεδομένων

Γλώσσες ΣΔΒΔ (SQL)

Γλώσσα Ορισμού

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

Γλώσσα Χειρισμού/Επεξεργασίας Δεδομένων

εισαγωγή, διαγραφή, τροποποίηση δεδομένων

insert/delete/update

διατύπωση ερωτημάτων

select *Γνωρίσματα*
from *Πίνακες*
where *Συνθήκη*

Γλώσσες ΣΔΒΔ

Γλώσσες Ερωτήσεων (Query Languages)

- δεν είναι γλώσσα προγραμματισμού (δυνατότητα εμφύτευσης σε μια γλώσσα υψηλού επιπέδου)
- **δηλωτικές** (μη διαδικαστικές)

Βασικές Έννοιες (ανασκόπηση)

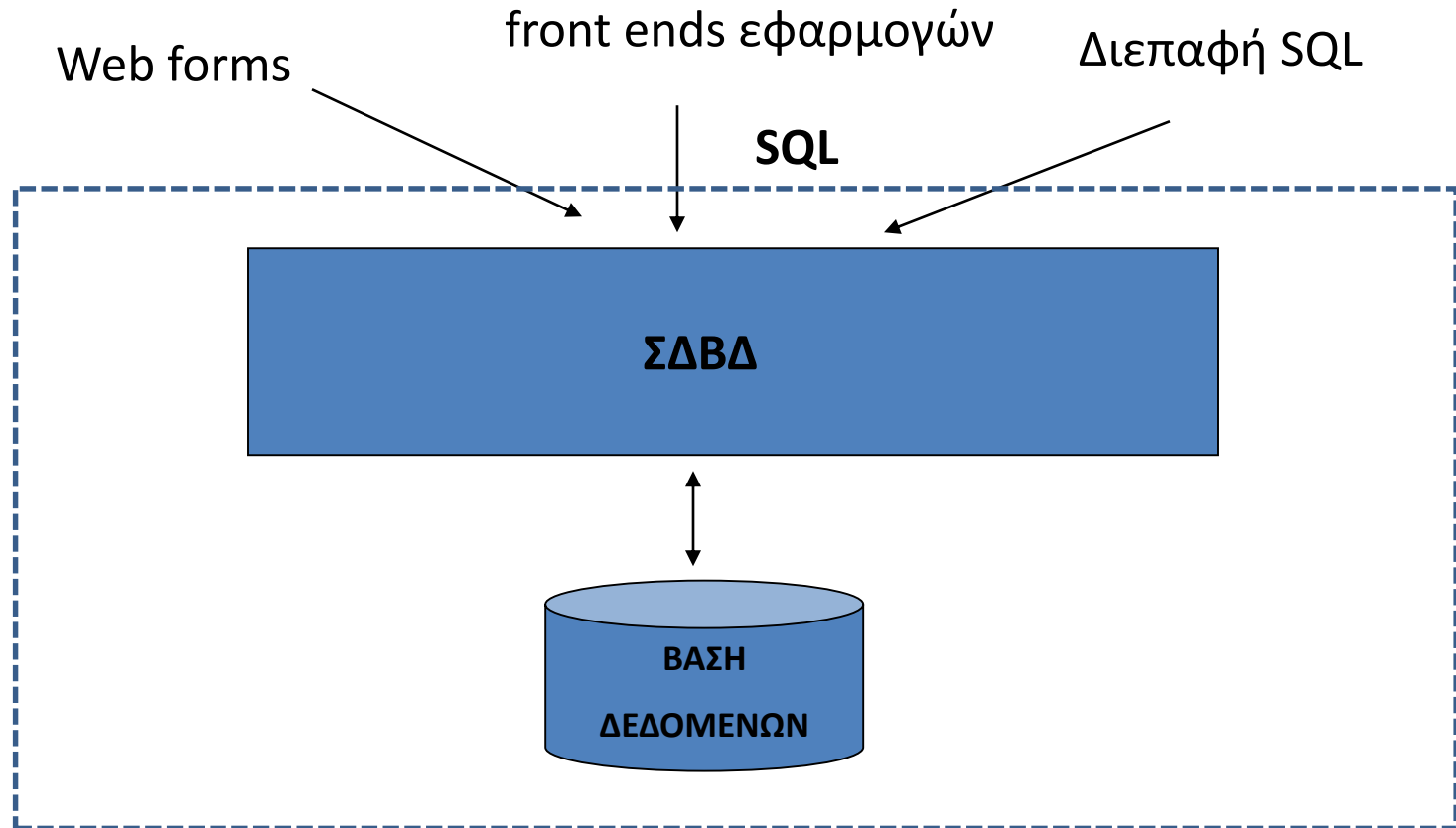
Κάποιες λειτουργίες ενός ΣΔΒΔ

- **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)

Χρήστες

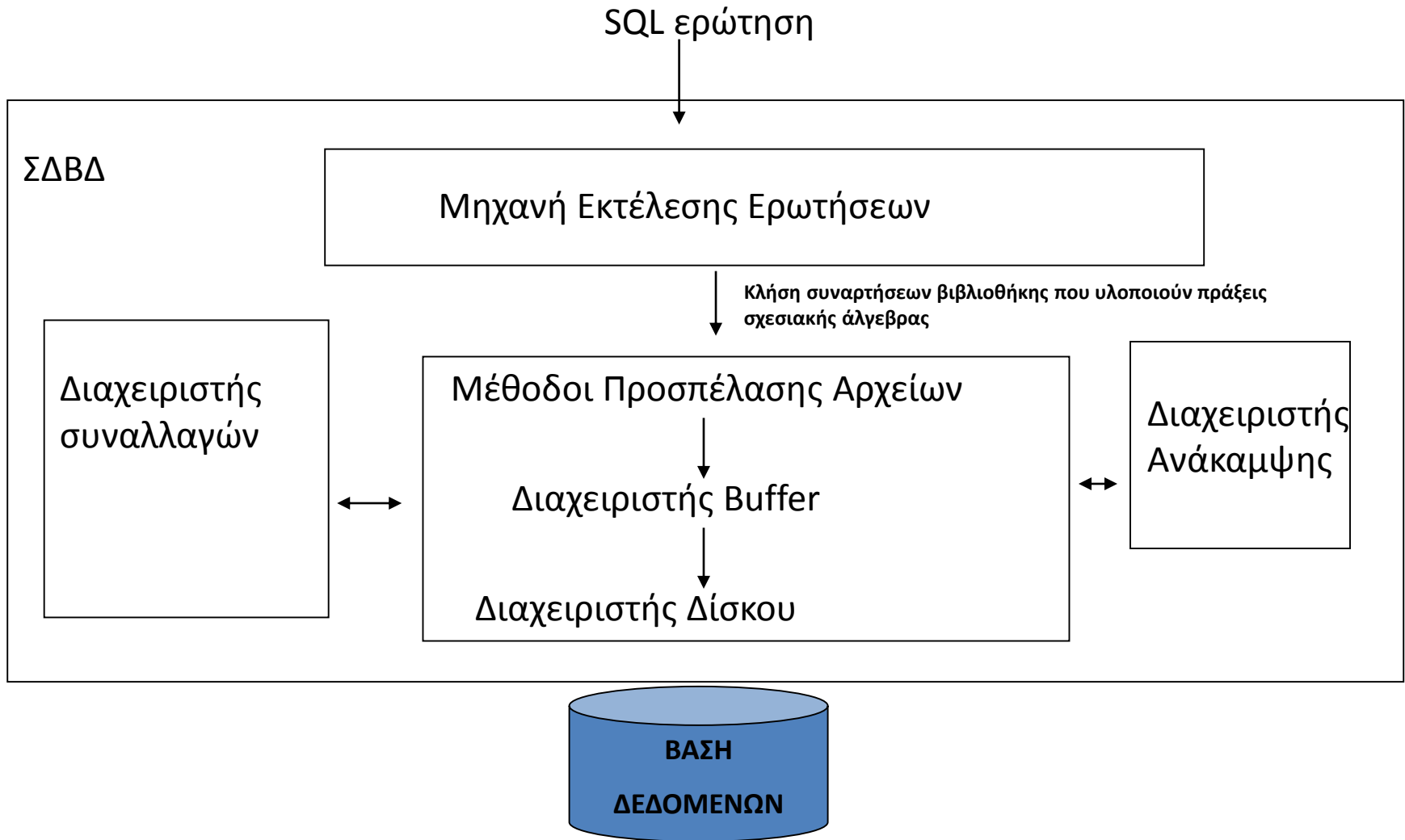
- Απλοί χρήστες
- Προγραμματιστές εφαρμογών
- Σχεδιαστές βάσεων δεδομένων
- Διαχειριστές συστήματος
- Δημιουργοί ΣΔΒΔ

Διεπαφές

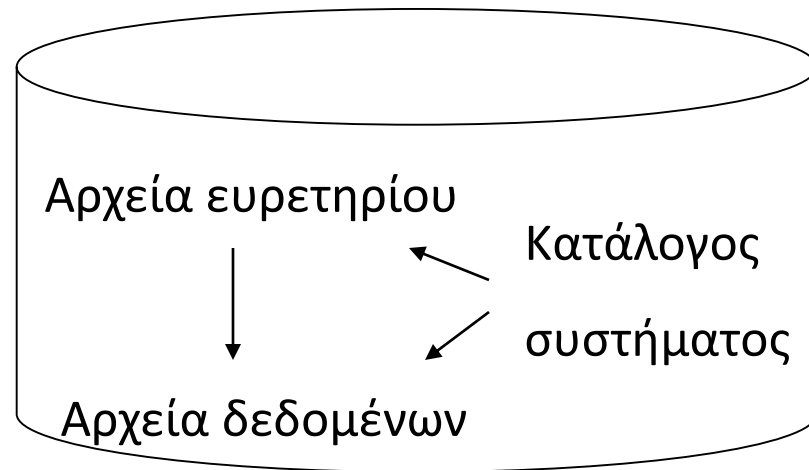


- Βασιζόμενες σε μενού (κατάλογο από επιλογές), γραφικών, Βασιζόμενες σε φόρμες, φυσικής γλώσσας, για παραμετρικούς χρήστες, για το ΔΒΔ

Το εσωτερικό ενός ΣΔΒΔ




Το εσωτερικό ενός ΣΔΒΔ



ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ

Τι άλλο θα δούμε σήμερα

- I. Τι είναι οι ΒΔ, γιατί μας ενδιαφέρουν
- II. Στόχος και περιεχόμενο του μαθήματος
-  III. Ιστορία των ΒΔ
- IV. Διαδικαστικά θέματα

Ιστορία

Δεκαετία του 1950

Κάρτες και ταινίες (σειριακή επεξεργασία) – Batch processing

Αρχή του 1960

Γενικευμένη χρήση δίσκων

πρώτο **γενικού-σκοπού** ΣΔΒΔ (διαχωρισμός της λειτουργικότητας διαχείρισης δεδομένων από τις εφαρμογές): Integrated Data Store (GE)

Charles Bachman (Recipient of the 1st **Turing Award**, 1973)

network data model (δικτυωτό)

Τέλη του 1960

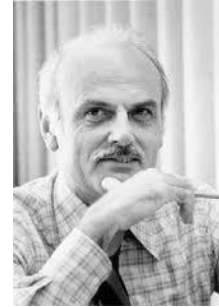
Information Management System (IMS) IBM

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)

M. Stonebraker, Turing Award 2014
Ingress, Postgress, entrepreneur, ..)
1M from Google



Ιστορία

Σήμερα, 2004 - ...

- Μεγάλος όγκος δεδομένων (**BIG DATA**)
- Αλλαγές σε υλικό (επεξεργαστές με πολλούς πυρήνες, κλπ)
- cloud computing
- data lakes

Κίνηση NoSQL

MapReduce (2004) και **Bigtable** (2006) by Google, **Dynamo** (2007) by Amazon

Hadoop (βασισμένο στο MapReduce, 2006), **Cassandra** (επηρεασμένη από Bigtable και Dynamo papers) και **MongoDB** (2009)

Κίνηση NotOnlySQL

Join! SQL interfaces πάνω στο Hadoop (και αργότερα στη Spark)

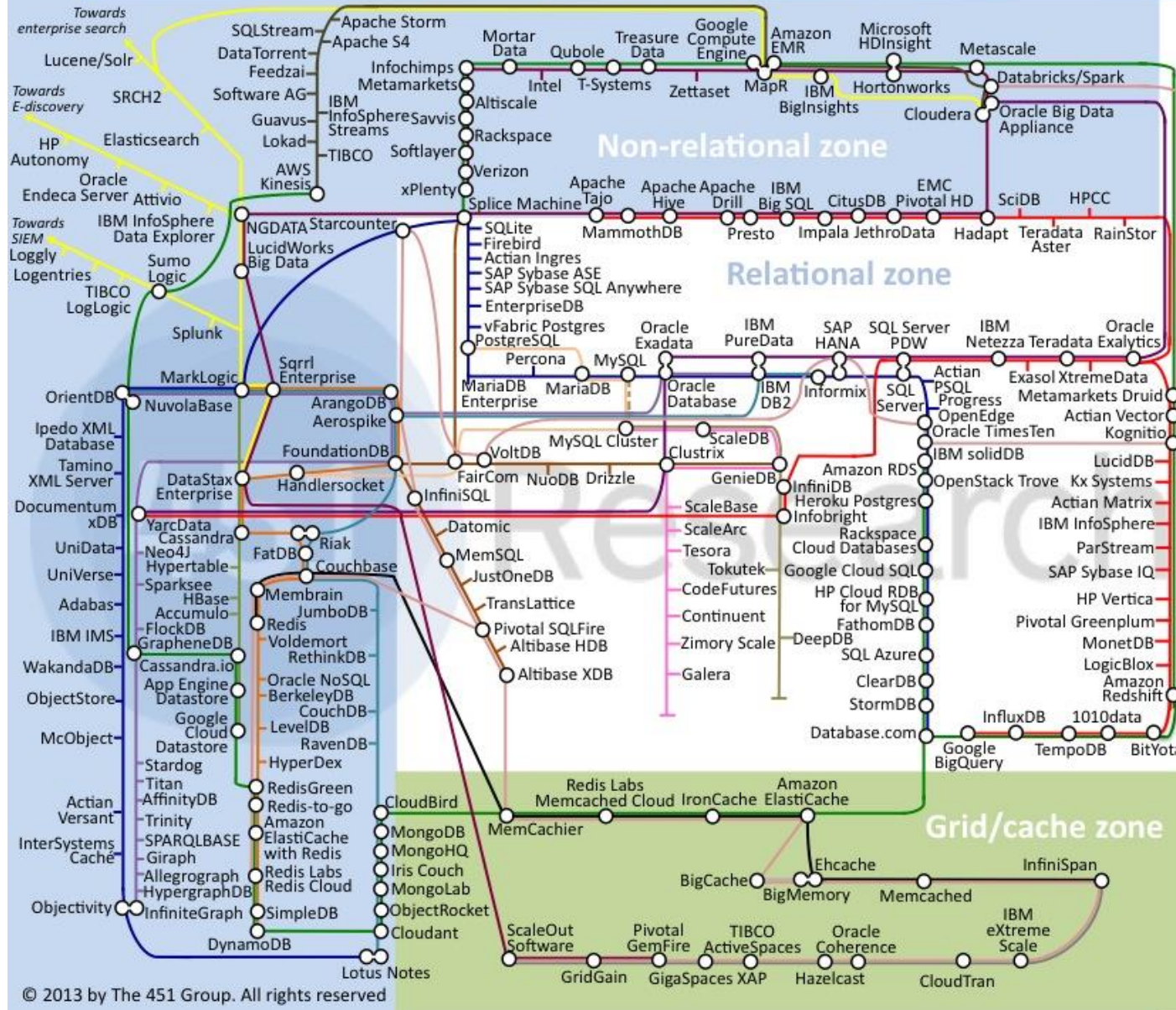
Κίνηση NewSQL

Spanner (2012) by Google **PostgreSQL 10** (native support for JSON, κλπ)

Database Landscape Map – February 2014

451 Research

- Key:**
- General purpose
 - Specialist analytic
 - as-a-Service
 - - - NoSQL extension
 - Big Tables
 - Graph
 - Document
 - Key value stores
 - Key value direct access
 - Hadoop
 - - - NewSQL extension
 - Advanced clustering/sharding
 - New SQL databases
 - Data caching
 - Data grid
 - Search
 - Appliances
 - Off-heap memory
 - In-memory
 - Stream processing

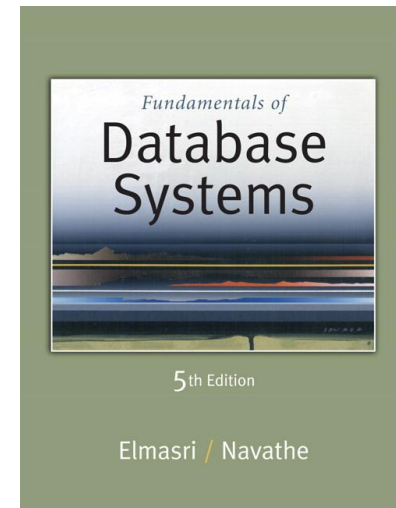
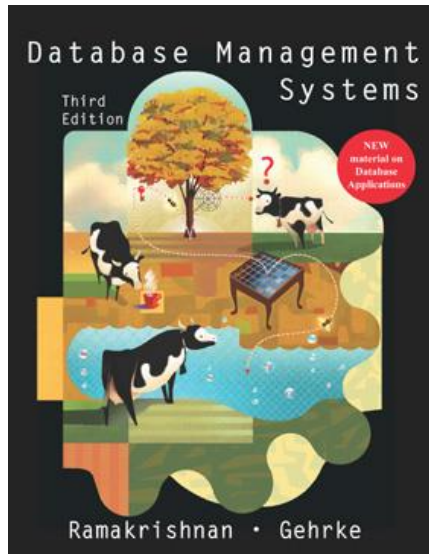


© 2013 by The 451 Group. All rights reserved

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 (2-3 σύνολα) + μια προγραμματιστική άσκηση (≥ 4.0)
- Τελικό διαγώνισμα (≥ 4.0)
- Τελικός Βαθμός (≥ 5.0)
- $35\% * (\text{Βαθμός Ασκήσεων} + \text{Προγραμματιστικής}) +$
 $65\% * \text{Βαθμός Τελικού Διαγωνίσματος}$

Όσοι έδωσαν ασκήσεις τα 2 προηγούμενα ακαδημαϊκά έτη μπορούν να «κρατήσουν» το βαθμό τους

Οι υπόλοιποι **πρέπει** ή να τις επαναλάβουν ή να γράψουν ≥ 5.0

1. Καλό θα είναι να παρακολουθείτε το μάθημα (τις διαλέξεις, αλλά και το «ρυθμό» του)
2. Οι διαφάνειες ΔΕΝ αντικαθιστούν το βιβλίο (είναι συμπληρωματικές σε αυτό)
3. Κάποιες ασκήσεις λύνονται ΜΟΝΟ στον πίνακα σκόπιμα
4. Κάνετε ερωτήσεις (και μια «άσχετη» ερώτηση είναι καλύτερη από καμία ερώτηση)
5. Διαβάζω – κατανοώ – μαθαίνω

ΔΕΝ υπάρχει «μεθοδολογία ασκήσεων»

Ερωτήσεις;