

# 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

# Βασικές Έννοιες

Γιατί να μας ενδιαφέρουν;

Τα ΒΔ και ΣΒΔ είναι παντού

# Που υπάρχουν ΒΔ;



## Εύδοξος

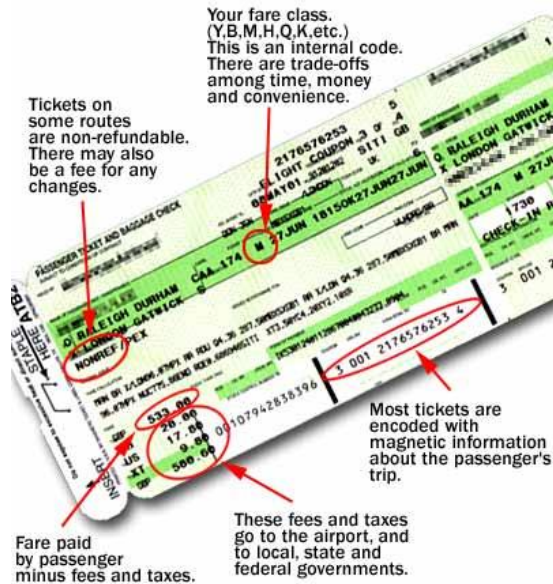
Ηλεκτρονική Υπηρεσία Ολοκληρωμένης Διαχείρισης  
Συγγραμμάτων και Λοιπών Βοηθημάτων



On-Line Δήλωση Μαθημάτων  
Students Web

Cronus

ecourse



Source: Delta Airlines

© 2001 HowStuffWorks

# Που υπάρχουν ΒΔ;



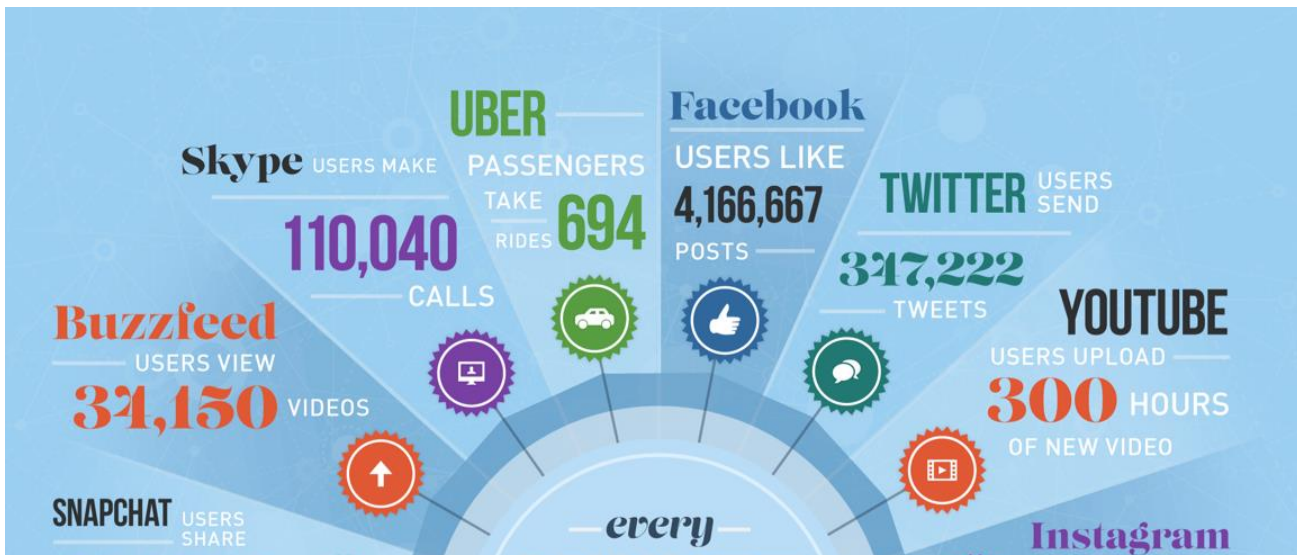
the top industries using Databases  
Software and tools

[Business Services](#)  
[Software](#)  
[Custom Software & IT Services](#)  
[Manufacturing](#)  
[Retail](#)

# Data science

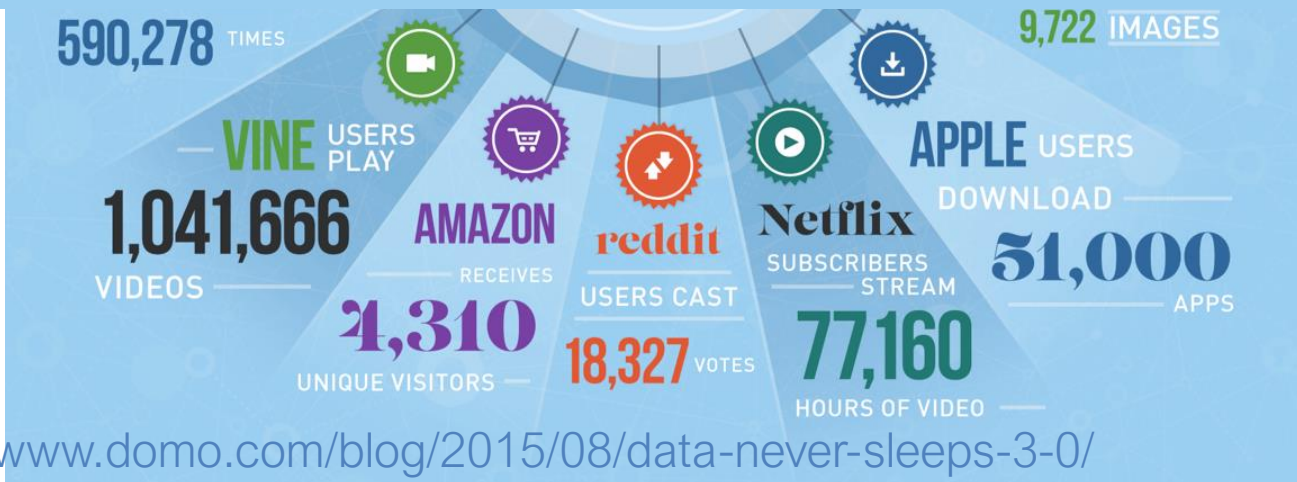
**Every two days we create as much data as much we did from the dawn of humanity to 2003.  
[Eric Schmidt, Google, 2010]**





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

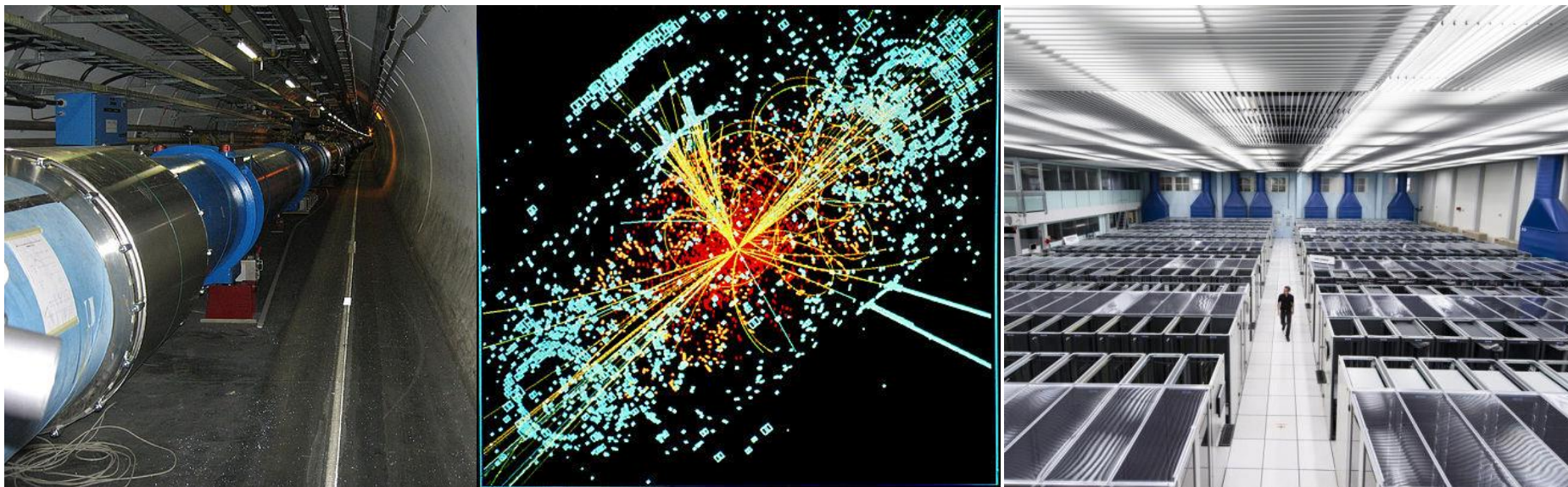
**3.2 BILLION PEOPLE.**



<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>

# DATA is the new OIL

- Πανταχού παρόντες αισθητήρες και καταγραφείς: Κάμερες, κινητά, κοινωνικά δίκτυα, μεγάλα επιστημονικά πειράματα (βιολογία, αστρονομία, φυσική, ιατρική, ....)
- Τα Συστήματα Βάσεων Δεδομένων βασική τεχνολογία για την
  - Αποθήκευση
  - Επεξεργασία, αναζήτηση με διατύπωση ερωτήσεων

# Γιατί ΣΔΒΔ;

Γιατί όχι σε απλά αρχεία;

Τι προσφέρουν:

- μοντελοποίηση, έλεγχος πλεονασμού, περιορισμοί ακεραιότητας
- δηλωτικό τρόπο διατύπωσης ερωτήσεων (queries) (SQL) αποδοτική επεξεργασία ερωτήσεων (ευρετήρια, βελτιστοποίηση) και ενημέρωση δεδομένων
- *ορθότητα λειτουργίας και μόνιμη αποθήκευση (persistent storage):* 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)

# 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)
- ...



# Δημοφιλή ΣΔΒΔ (2023)


415 systems in ranking, October 2023

| Rank     |          |          | DBMS                         | Database Model             | Score    |          |          |
|----------|----------|----------|------------------------------|----------------------------|----------|----------|----------|
| Oct 2023 | Sep 2023 | Oct 2022 |                              |                            | Oct 2023 | Sep 2023 | Oct 2022 |
| 1.       | 1.       | 1.       | Oracle                       | Relational, Multi-model    | 1261.42  | +20.54   | +25.05   |
| 2.       | 2.       | 2.       | MySQL                        | Relational, Multi-model    | 1133.32  | +21.83   | -72.06   |
| 3.       | 3.       | 3.       | Microsoft SQL Server         | Relational, Multi-model    | 896.88   | -5.34    | -27.80   |
| 4.       | 4.       | 4.       | PostgreSQL                   | Relational, Multi-model    | 638.82   | +18.06   | +16.10   |
| 5.       | 5.       | 5.       | MongoDB                      | Document, Multi-model      | 431.42   | -8.00    | -54.81   |
| 6.       | 6.       | 6.       | Redis                        | Key-value, Multi-model     | 162.96   | -0.72    | -20.41   |
| 7.       | 7.       | 7.       | Elasticsearch                | Search engine, Multi-model | 137.15   | -1.84    | -13.92   |
| 8.       | 8.       | 8.       | IBM Db2                      | Relational, Multi-model    | 134.87   | -1.85    | -14.79   |
| 9.       | 9.       | 10.      | SQLite                       | Relational                 | 125.14   | -4.06    | -12.66   |
| 10.      | 10.      | 9.       | Microsoft Access             | Relational                 | 124.31   | -4.25    | -13.85   |
| 11.      | 11.      | 13.      | Snowflake                    | Relational                 | 123.24   | +2.35    | +16.51   |
| 12.      | 12.      | 11.      | Cassandra                    | Wide column, Multi-model   | 108.82   | -1.24    | -9.12    |
| 13.      | 13.      | 12.      | MariaDB                      | Relational, Multi-model    | 99.66    | -0.79    | -9.65    |
| 14.      | 14.      | 14.      | Splunk                       | Search engine              | 92.37    | +0.98    | -2.28    |
| 15.      | 15.      | 16.      | Microsoft Azure SQL Database | Relational, Multi-model    | 80.93    | -1.80    | -4.03    |
| 16.      | 16.      | 15.      | Amazon DynamoDB              | Multi-model                | 80.91    | +0.00    | -7.44    |
| 17.      | 17.      | 20.      | Databricks                   | Multi-model                | 75.82    | +0.64    | +18.21   |
| 18.      | 18.      | 17.      | Hive                         | Relational                 | 69.18    | -2.65    | -11.42   |
| 19.      | 19.      | 18.      | Teradata                     | Relational, Multi-model    | 58.56    | -1.77    | -7.51    |
| 20.      | 20.      | 22.      | Google BigQuery              | Relational                 | 56.57    | +0.11    | +4.12    |
| 21.      | 21.      | 23.      | FileMaker                    | Relational                 | 53.33    | +0.00    | +0.00    |

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

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

# 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 μικρή δημοσκόπηση (στοιχεία με γκρι 2020-2021)

Σε ποιο έτος σπουδών

4<sup>ο</sup> 57

5<sup>ο</sup> 36

>5 40

Φορά που παίρνετε το μάθημα

1 94

2 20

>2 11

Πόσο ευχαριστημένοι είστε από τις σπουδές σας

Καθόλου 4

Μέτρια 90

Πολύ 21

Καλύτερα εξ αποστάσεως 69

Καλύτερα δια ζώσης 57

# Εισαγωγή στα Συστήματα Βάσεων Δεδομένων

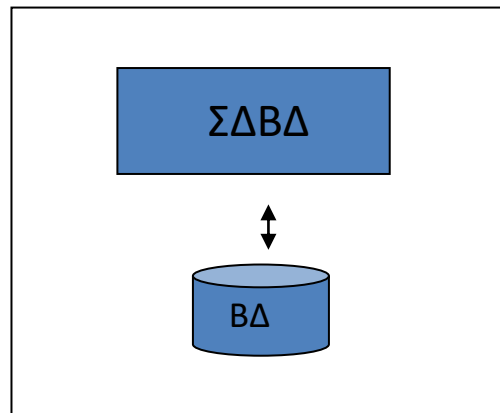
# Τι θα δούμε σήμερα

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

# 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**

# Περιεχόμενο μαθήματος

- Μέρος 1: Σχεδιασμός, υλοποίηση και χειρισμός μιας βάσης δεδομένων με χρήση ενός σχεσιακού ΣΔΒΔ
- Μέρος 2: Το εσωτερικό ενός ΣΔΒΔ

# Μέρος 1: χρήση ΣΔΒΔ

## **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 καθώς και των περιορισμών ακεραιότητας 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

# Μοντελοποίηση

- Υψηλού επιπέδου (εννοιολογικά) μοντέλα

Υψηλού επιπέδου, περισσότερο αφηρημένη περιγραφή της δομής

Μοντέλο Οντοτήτων/Συσχετίσεων

- Παραστατικά μοντέλα ή μοντέλα υλοποίησης ή λογικά μοντέλα

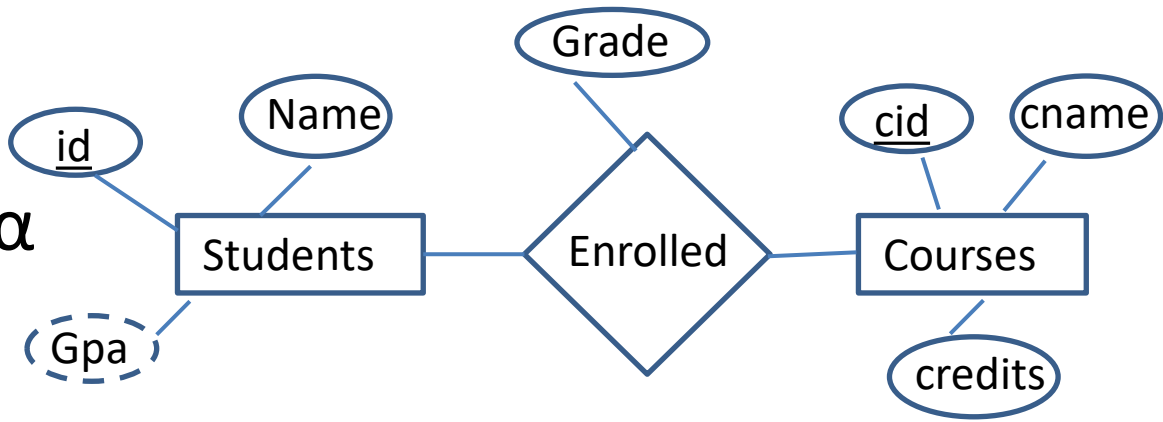
Σχισιακό Μοντέλο, Ιεραρχικό Μοντέλο, Δικτυωτό Μοντέλο, item-value, graph model, document stores (JSON)

- Χαμηλού επιπέδου ή φυσικά μοντέλα

Μοντέλα αποθήκευσης

# Παράδειγμα (Οντότητες-Συσχετίσεις)

- ΣΔΒΔ για μαθήματα



- Φοιτητές
- Μαθήματα
- Καθηγητές

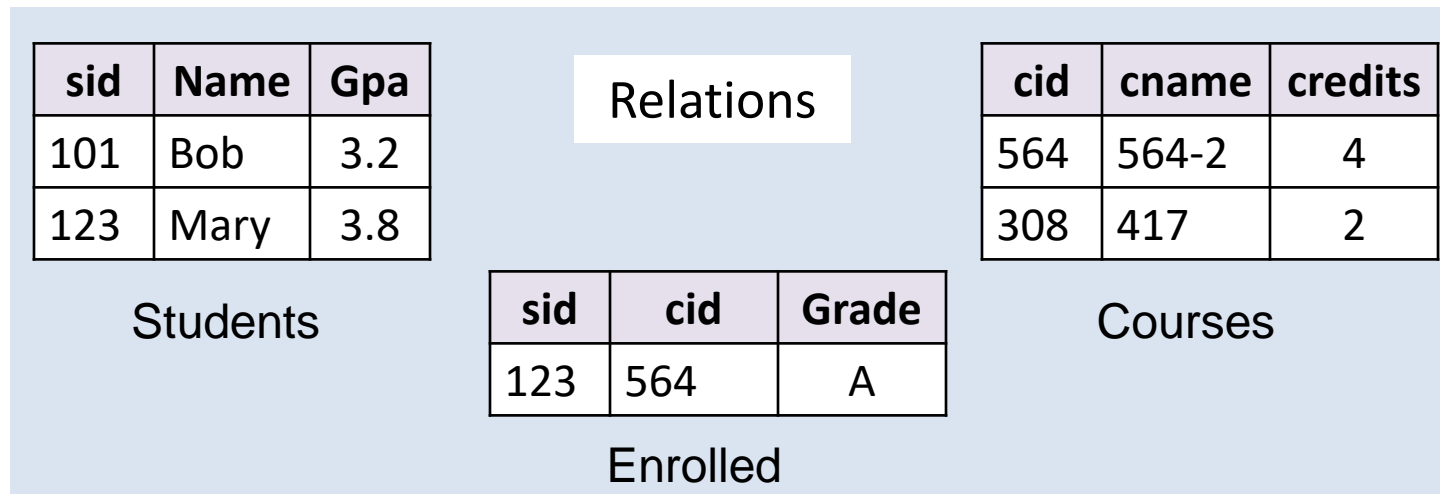
} Οντότητες

- Ποιος φοιτητής παρακολουθεί ποιο μάθημα και βαθμός
- 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*)



# 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:**

- Δημιουργία σχήματος (πινάκων)
- 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)

Θα μελετήσουμε και θεωρητικές γλώσσες: σχεσιακή άλγεβρα

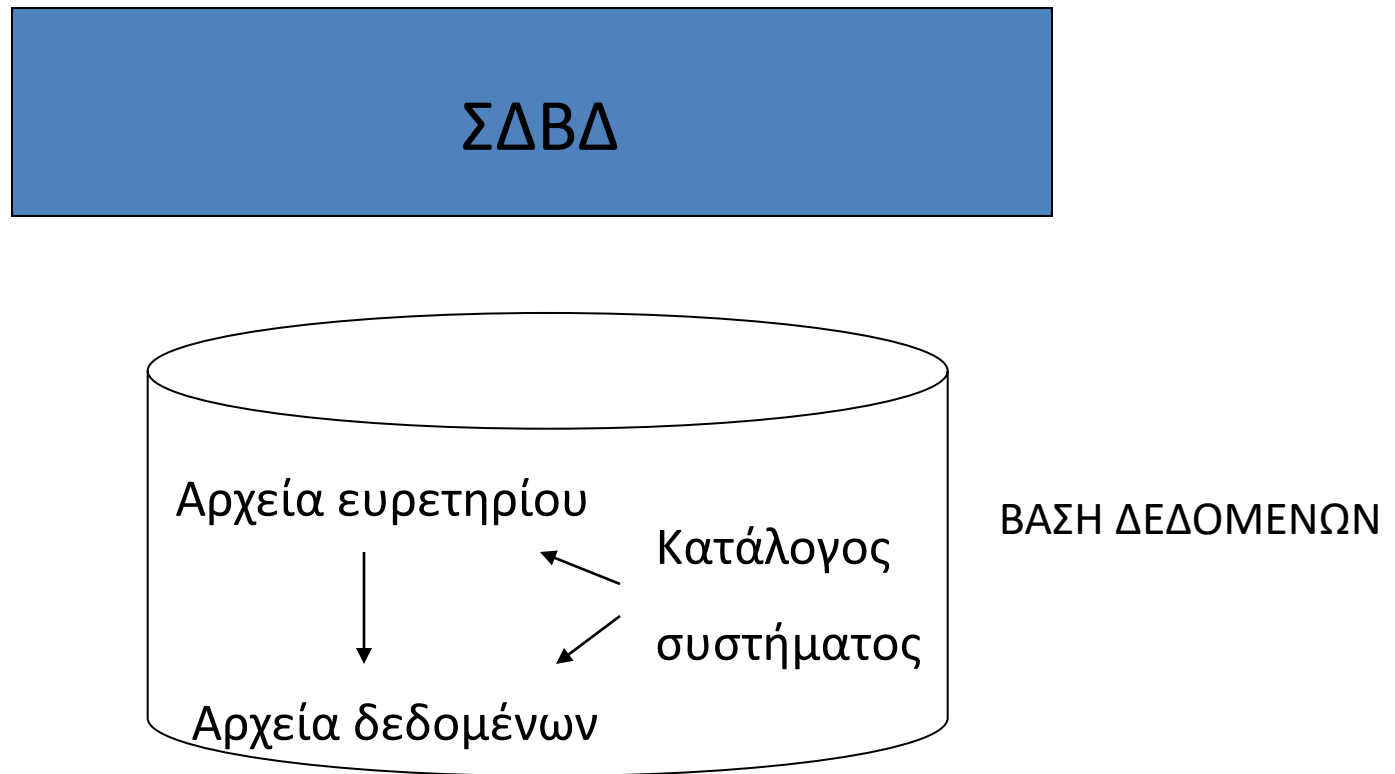


# Βασικές Έννοιες (ανασκόπηση)

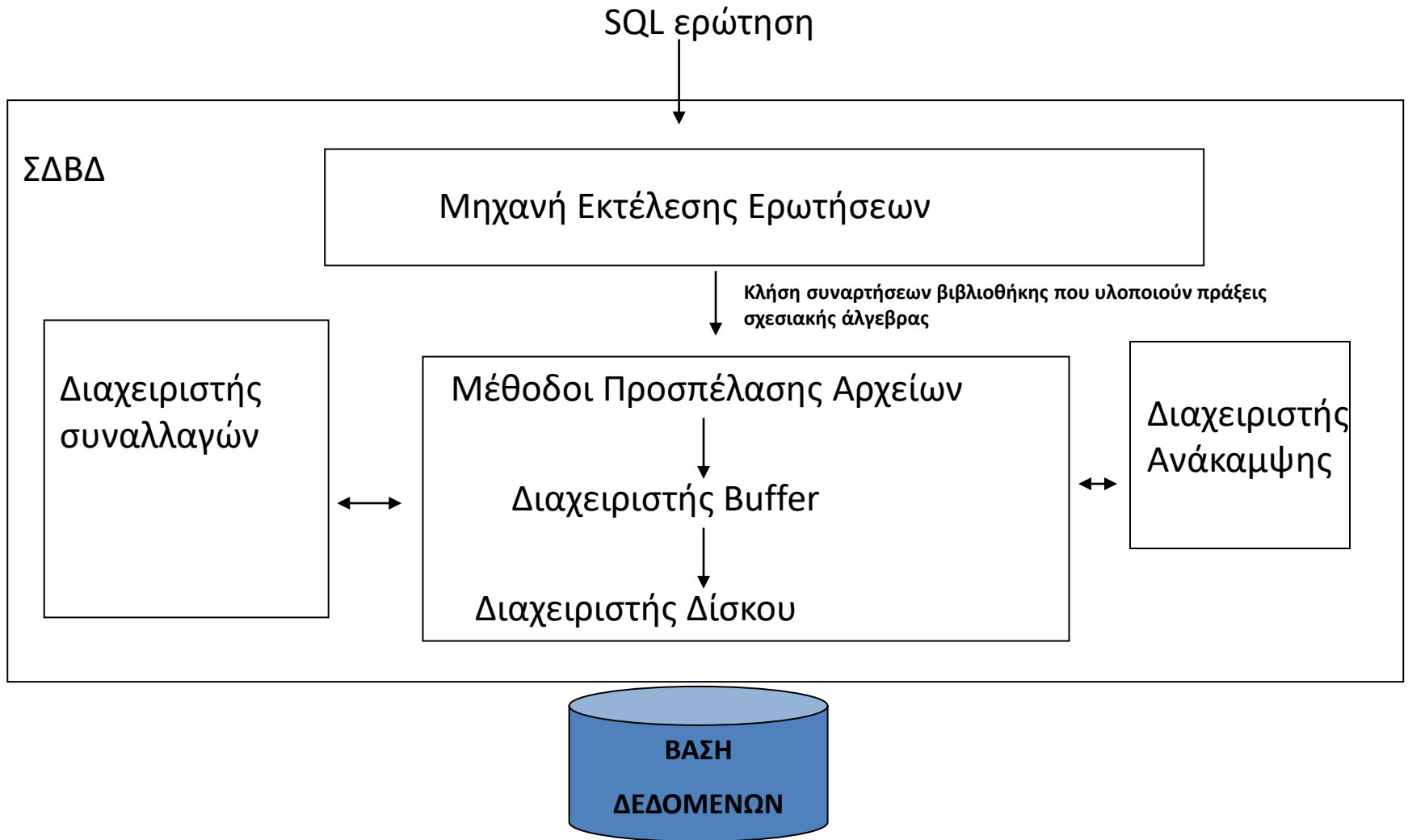
## Κάποιες λειτουργίες ενός ΣΔΒΔ

- **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)

# Μέρος 2: Το εσωτερικό ενός ΣΔΒΔ



# Το εσωτερικό ενός ΣΔΒΔ



# 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

# Σκοπός του μαθήματος

Τρεις βασικοί στόχοι:

1. Να μπορείτε να σχεδιάζετε και υλοποιείτε 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

# Χρήστες

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

# 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 1<sup>st</sup> **Turing Award**, 1973)

network data model (δικτυωτό)

## Τέλη του 1960

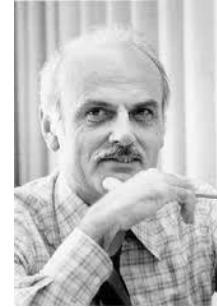
Information Management System (IMS) IBM

hierarchical data model (Ιεραρχικό)

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



# Ιστορία



Peter P. Chen

**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

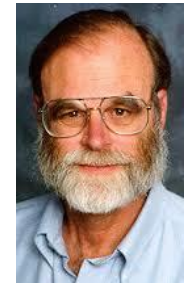
**Δεκαετία του 1980**

SQL (μέρος του **System R**)

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

υποσημείωση: Jim Gray gone missing

[*Τάσεις: αντικειμενοστραφή, αρχιτεκτονική πελάτη-εξυπηρέτη, 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

## Δεκαετία του 2010

- 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)

# Ιστορία

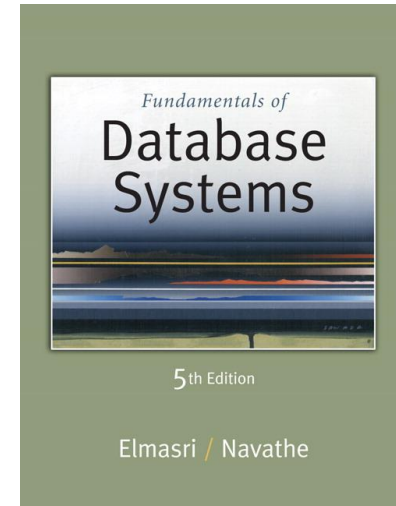
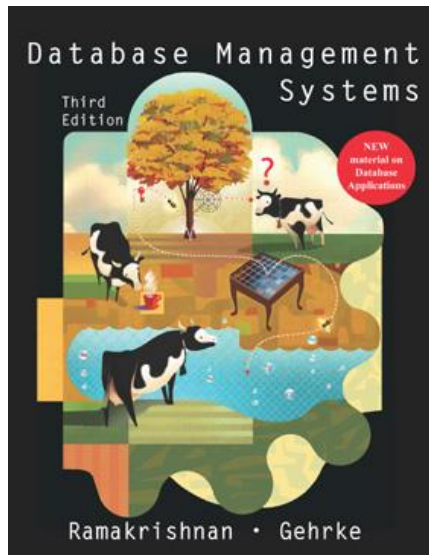
## Δεκαετία του 2020

- AI for DB
- DB for AI

# 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

## **3 ΠΡΟΑΙΡΕΤΙΚΑ** εργαστήρια στην SQL

Προκαταρκτικός σχεδιασμός (πολύ πιθανών να αλλάξει)

Εργαστήριο 1: Create table, απλές εντολές insert/update/delete

Εργαστήριο 2: SQL queries

Εργαστήριο 3: SQL queries + optimizer

**3 σύντομα προαιρετικά quiz** στο τέλος κάθε εργαστηρίου

# Διαχειριστικά Θέματα

Ο βαθμός θα προκύψει ως συνδυασμός:

- Ασκήσεων (2-3 σύνολα) με θεωρητικά και προγραμματιστικά (SQL) ερωτήματα
- Τελικού διαγωνίσματος
- Quizzes
- Απαραίτητη προϋπόθεση: Βαθμός τελικού διαγωνίσματος  $\geq 4$

$$B = \begin{cases} \max\{0.65 * E + 0.35 * A, E\} + 0.2 * Q, & \text{αν } E \geq 4 \\ E, & \text{αλλιώς} \end{cases}$$

E: Βαθμός εξέτασης, A: βαθμός ασκήσεων Q: Βαθμός quiz

**Ασκήσεις και quizzes δεν «κρατιούνται»**



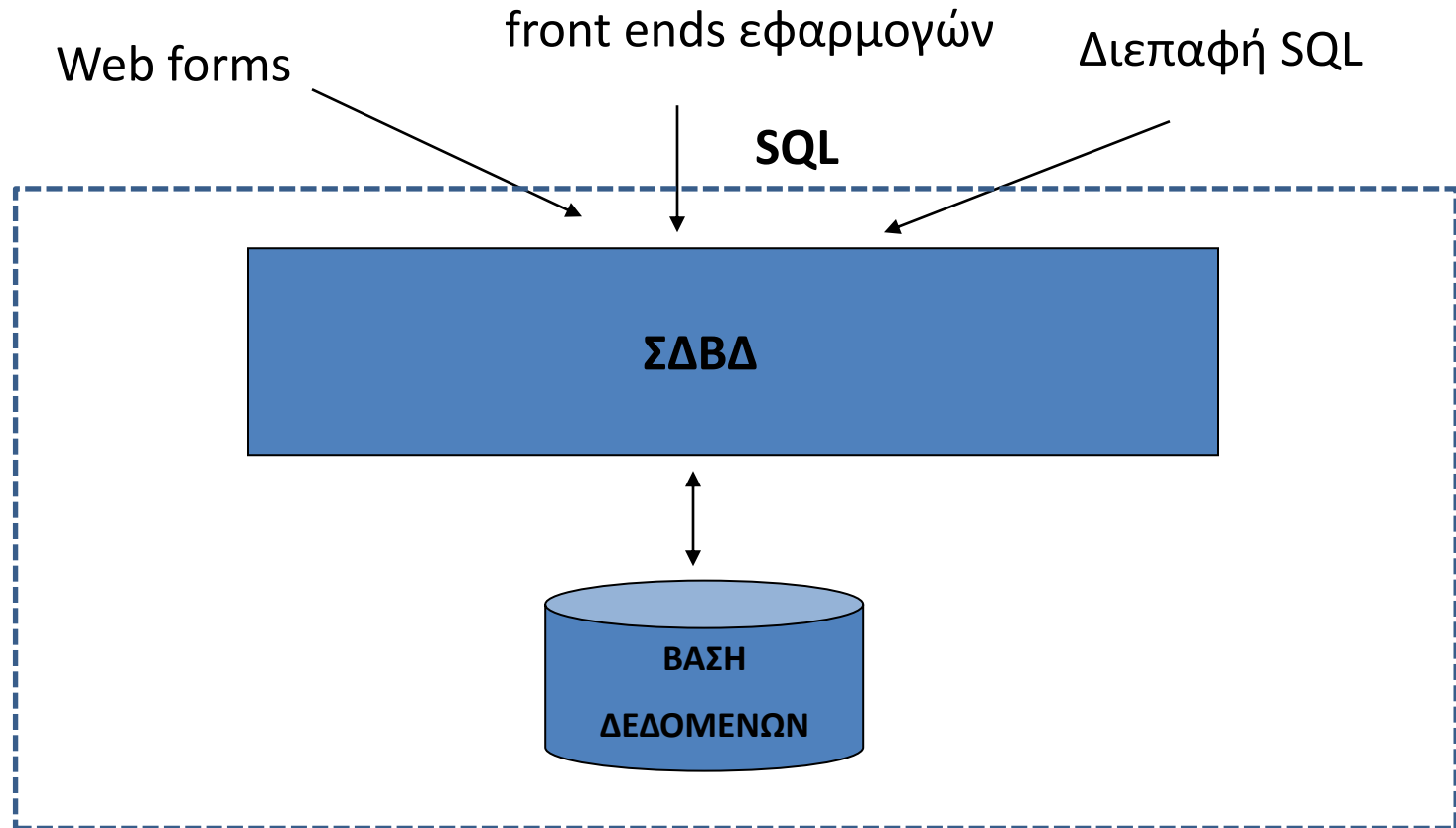
# Συμβουλές

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) ή σημειώσεις
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

# Ερωτήσεις;

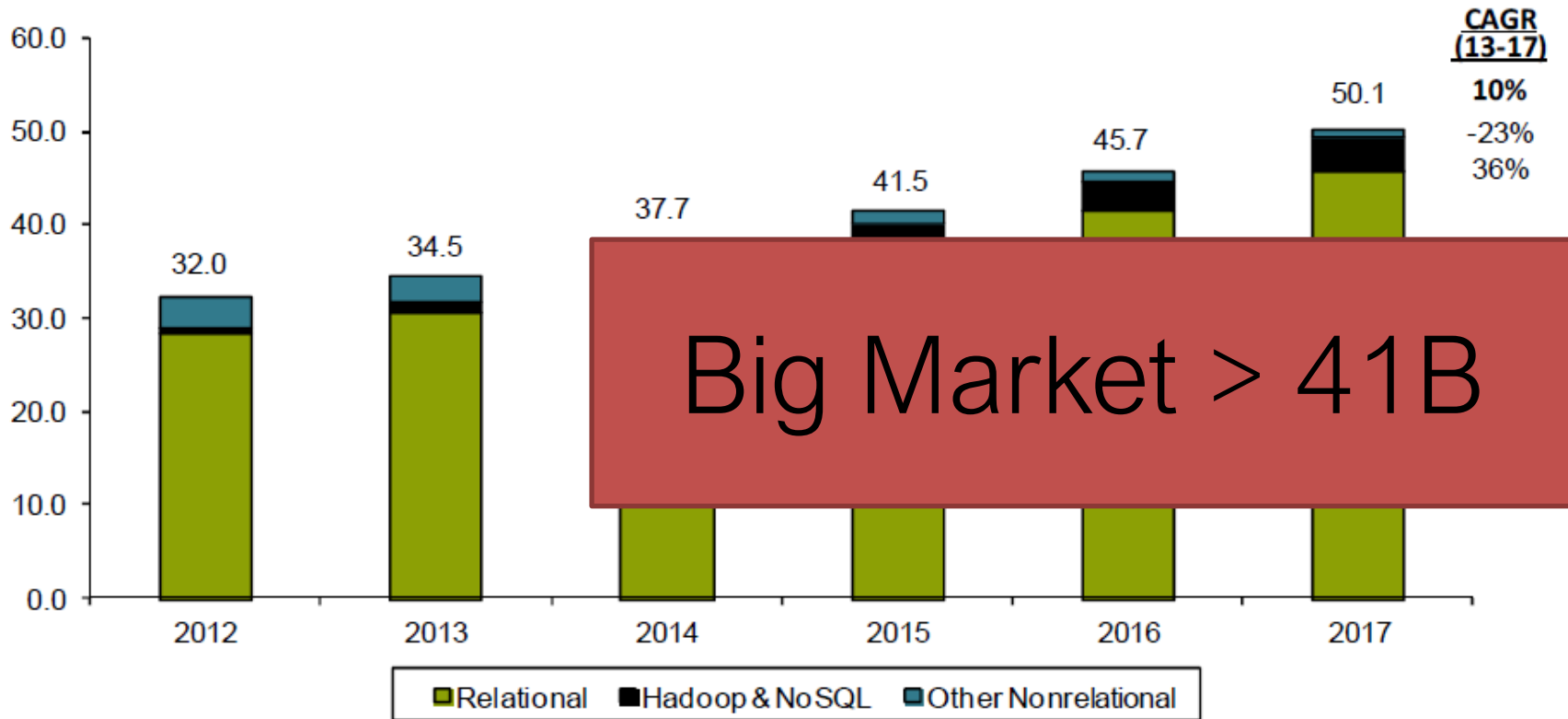
# Διεπαφές



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

# ΣΔΒΔ

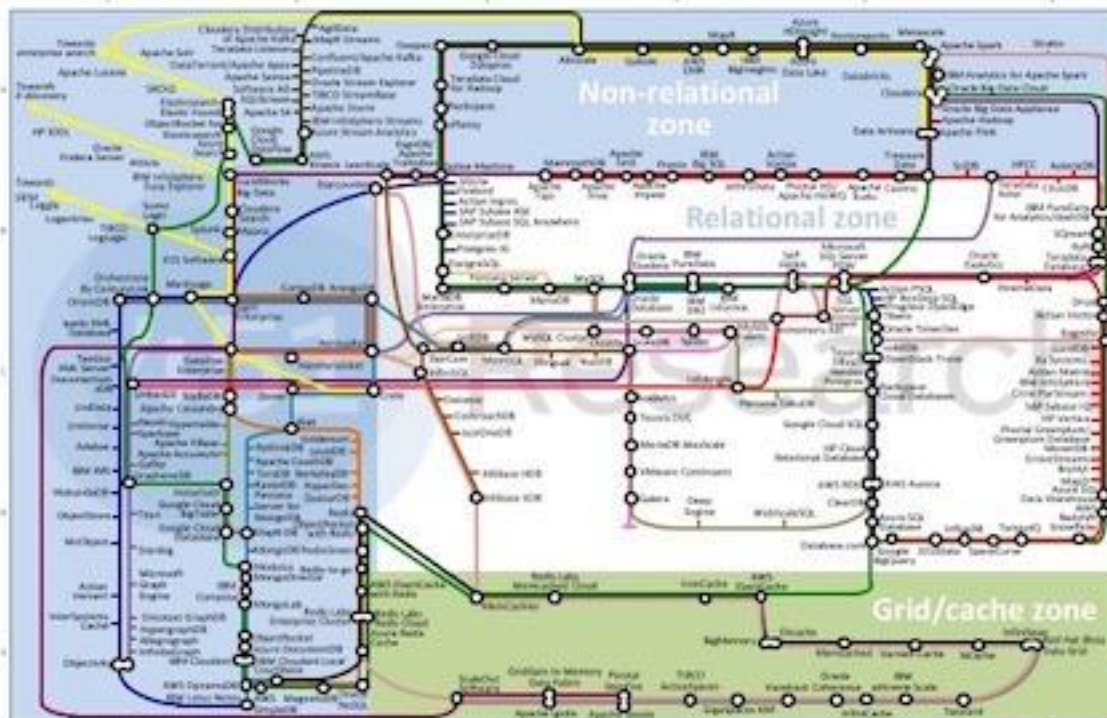
## Global Database Market (\$B)



Source: IDC, Bernstein analysis

The DBMS market saw strong growth of 18.4% in 2018, driven by new investment going primarily to cloud dbPaaS offerings. The nonrelational DBMS segment continued its strong above-market-rate growth at 55.9%, while RDBMS grew a healthy 16.2%. Prerelational era DBMS continued its decline. (\*)

(\*) Gartner report, June 2019



451 Research

# Data Platforms Map January 2016

- Map**
- General purpose
  - Analytics analysis
  - ad-hoc queries
  - NoSQL
  - Graph
  - Document
  - Key value stores
  - Key value stores
  - Index
  - Feeding
  - Analytics ingestion
  - Advanced
  - Ingestion/Exporting
  - New SQL databases
  - Data writing
  - Data grid
  - Search
  - MapReduce
  - In-memory
  - In-memory processing

<https://451research.com/state-of-the-database-landscape>

© 2016 by 451 Research LLC. All rights reserved.