

ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ

Νίκος Αντωνιάδης

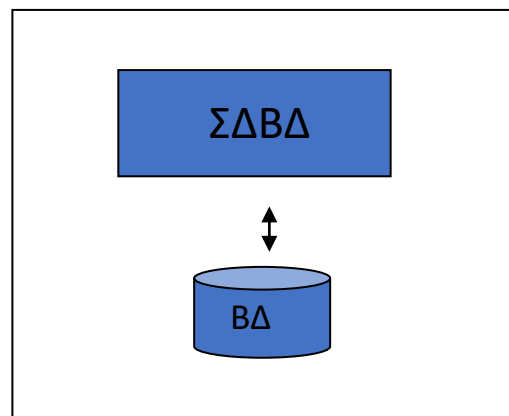
Τι είναι Βάση Δεδομένων;

- Ποιος έχει βάση δεδομένων;
 - Εφορία
 - ΕΦΚΑ
 - Γραμματεία
 - eBay
 - Ηλεκτρονικό κατάστημα
- ΟΛΟΙ έχουν ΒΔ, από πολύ μικρές (ένα ηλεκτρονικό κατάστημα με 100 προϊόντα) έως πολύ μεγάλες (taxisnet, Amazon, Google, ...)

Περιγραφή/ορισμός

- Δεδομένα αποθηκευμένα και ΟΡΓΑΝΩΜΕΝΑ
 - Χαρακτηριστικά για κάθε αντικείμενο
 - **Λογισμικό** εξειδικευμένο για αναζήτηση
-
- Σύστημα Διαχείρισης Βάσης Δεδομένων

Σύστημα Διαχείρισης Βάσης Δεδομένων



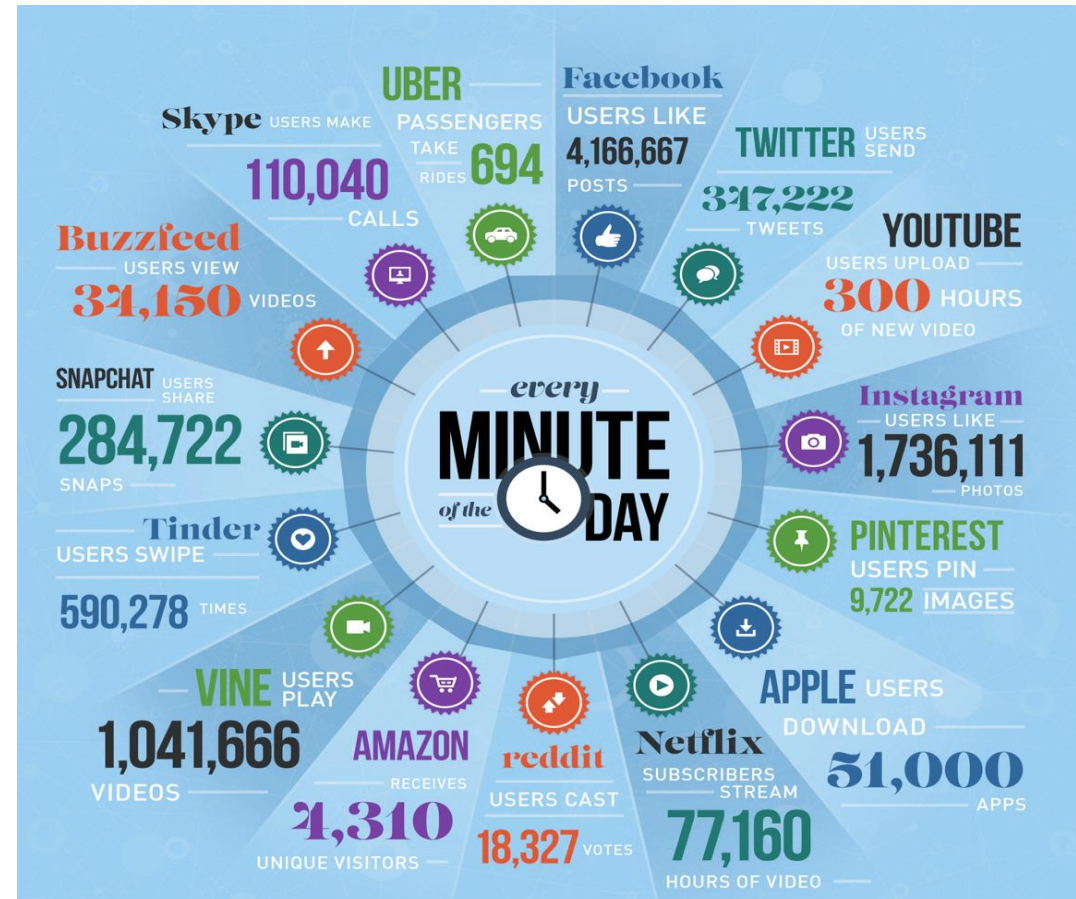
**Σύστημα Βάσεων
Δεδομένων**

Γιατί τόση φασαρία?

- Τα ΣΔΒΔ είναι παντού
- Εφορία
- Αγορές προϊόντων
- ΑΤΜ
- Πιστωτικές κάρτες
- Γραμματεία Τμήματος

Γιατί τόση φασαρία?

- Facebook
- Amazon
- Ebay
- AirBnB



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

3.2 BILLION PEOPLE.

Γιατί τόση φασαρία?

Συνεχής παραγωγή data

- Κινητή τηλεφωνία
- Συστήματα καμερών (κάμερες τροχαίας, εναέρια κίνηση, κοινωνικά δίκτυα)
- Καταγραφή στοιχείων
- Μετεωρολογικά
 - ✓ Κίνηση στους δρόμους
 - ✓ Εφαρμογές άθλησης
 - ✓ Εφαρμογές υγείας (φορέσιμες συσκευές, τηλεϊατρική)

Τι θα δούμε αυτό το εξάμηνο

- Σχεδιασμός Βάσης Δεδομένων
- Εργαλεία χειρισμού δεδομένων (SQL)
- Βέλτιστος σχεδιασμός / υλοποίηση ΒΔ

Δημιουργία ΒΔ: ένα (όχι και τόσο) σύνθετο έργο

Μοντελοποίηση

- Κατάστρωση μοντέλου
- Στο χαρτί!
- Κακό μοντέλο → αποτυχία
- Μοντέλο Οντοτήτων/Συσχετίσεων (Entity-Relationship model)

Προγραμματισμός/υλοποίηση

- Χρήση εμπορικού πακέτου
- Ανάπτυξη προγραμματιστικών εργαλείων
- Ανάπτυξη διεπαφής

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

Σκοπός ΒΔ

- Παράδειγμα: σύστημα πληροφοριών τράπεζας
 - Καταθέσεις: αριθμος λογαριασμού, υπολοιπο , κάτοχος
 - Δάνεια: αρ. δανείου, κεφάλαιο, κάτοχος, εγγυητης
 - Στοιχεία πελατών : ονομα, ΑΦΜ, τηλέφωνο, διεύθυνση, λογαριασμοί, δάνεια
- Σύστημα αρχείων (όχι ΒΔ)
- Σύνολο ρουτινών
 - Ρουτίνα πίστωσης / χρέωσης λογαριασμού
 - Ρουτίνα δημιουργίας λογαριασμού
 - Ρουτίνα δημιουργίας μηνιαίων αναφορών

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: QUEL, SEQUEL, QBE,

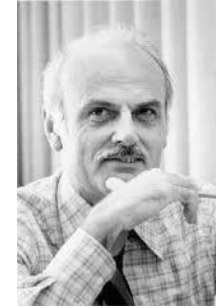
Δεκαετία του 1980

SQL (μέρος του *System R*)

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

υποσημείωση: Jim Gray gone missing

[Τάσεις: αντικειμενοστραφή, αρχιτεκτονική πελάτη-εξυπηρέτη, κατανεμημένες, έμπειρα]



Peter P. Chen



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
- AI & ML

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)

Τι είναι Βάση Δεδομένων;

- Είναι αρχείο;
- Είναι κάτι σαν το Word;
- Είναι κάτι σαν το Excel;

- Ποια είναι η διαφορά με ένα αρχείο (οποιοδήποτε) που έχει τα ίδια δεδομένα ;

Αρχεία αποθήκευσης

- Πλεονασμός δεδομένων

Πίνακας στοιχείων

Όνομα	ΑΔΤ	ΑΦΜ	Διεύθυνση
Παππάς	AB475859	012345678	Βενιζέλου 12, Αρτα
Πέτρου	AE069584	320195849	
Νικολάου	AB392020	039283742	Δωδώνης 34, Ιωάννινα

Πίνακας λογαριασμών

Όνομα	ΑΔΤ	ΑΦΜ	Διεύθυνση	Αρ. λογ	Υπόλοιπο
Παππάς	AB475859	012345678	Βενιζέλου 12, Αρτα	A123456	324.00
Πέτρου	AE069584	320195849		B324354	1258,99

Υπάρχει πρόβλημα;

- Τα δεδομένα επαναλαμβάνονται

Πλεονασμός (redundancy)

- Αν θέλουμε να συνδέσουμε τους πίνακες, πρέπει να γίνει προγραμματιστικά
- Αν αλλάξει κάτι, πχ μια διεύθυνση, πρέπει η αλλαγή να γίνει σε όλα αρχεία αυτή υπάρχει → λάθη (σκόπιμα ή τυχαία)

Ασυνέπεια (inconsistency)

Δυσκολίες πρόσβασης

- Ζητείται η λίστα των πελατών της Αθήνας → ρουτίνα
- Ζητείται η λίστα των πελατών της Λάρισας και του Βόλου
→ ρουτίνα ή τροποποίηση της προηγούμενης
- Ζητείται η λίστα των πελατών όλης της Ελλάδας → ρουτίνα
- Ζητείται η λίστα λογαριασμών με υπόλοιπο πάνω από 10000 → ρουτίνα
- Ζητείται η λίστα λογαριασμών με υπόλοιπο μεταξύ 5000 και 15000 → ρουτίνα

Δυσκολίες πρόσβασης

- Μειονεκτήματα
 - Συνεχής ανάπτυξη νέου λογισμικού
 - Τροποποιήσεις από διάφορους προγραμματιστές
 - Αδύνατον να προβλεφθούν όλα τα πιθανά ερωτήματα

Ανάγκη για καλύτερη οργάνωση των δεδομένων

Ακεραιότητα

- Τύποι δεδομένων
 - Αριθμοί
 - Δεκαδικοί
 - Δυαδικές (true/false)
 - Ημερομηνία
- Format
 - Δεν υπάρχει δυνατότητα πρόβλεψης συγκεκριμένης μορφοποίησης

Ακεραιότητα

- Περιορισμοί:
 - Υπόλοιπο λογαριασμού τουλάχιστον 0
 - Ημερομηνίες μόνο νόμιμες (όχι 30/2/2020)
 - ...
- Το αρχείο **δεν μπορεί** να κάνει τέτοιους ελέγχους

Ακεραιότητα (integrity)

Αναγκαιότητα η ΒΔ

- Τα παραδοσιακά συστήματα έχουν πολύ λίγες δυνατότητες
- Ανάγκη διασύνδεσης
- Ανάγκη βέλτιστης αποθήκευσης δεδομένων
- Ανάγκη γρήγορης και αξιόπιστης συνεργασίας

Παράδειγμα: Σύστημα γραμματείας

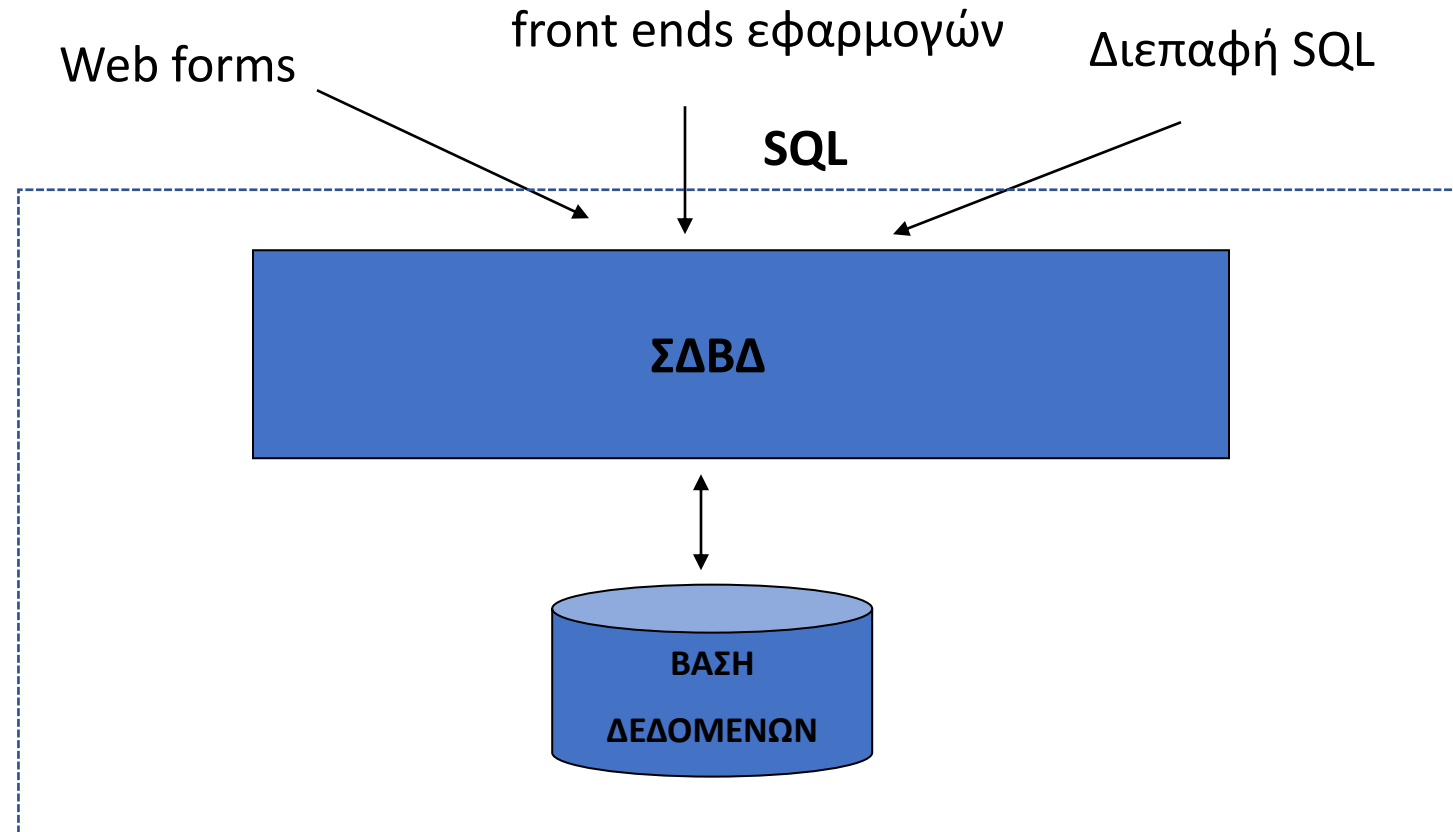
Απαιτήσεις χρήστη (user requirements)

- Καθηγητές: όνομα, αντικείμενο, τηλέφωνο, φοιτητές που επιβλέπει, μαθήματα
- Φοιτητές: μαθήματα, όνομα, τηλέφωνο, εξάμηνο, βαθμοί
- Μαθήματα: κωδικός, όνομα, ώρες, φοιτητές

Για ποιον υπάρχει η ΒΔ?

- Καθημερινοί χρήστες (εκατοντάδες εκατομμύρια)
- Προγραμματιστές
- Σχεδιαστές βάσεων δεδομένων
- Admins
- Δημιουργοί ΣΔΒΔ

Διεπαφές



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