

The background features a complex network of thin, light gray lines connecting various points, creating a web-like structure. Scattered throughout are numerous triangles of different sizes and orientations, some with solid gray outlines and others with dashed or dotted lines. The overall aesthetic is technical and mathematical.

01

Κατανεμημένα Συστήματα

Εισαγωγή, Ορισμοί



01

TO MAGHMA

Ταυτότητα Μαθήματος

- Διαλέξεις
 - Τρίτη και Τετάρτη 09:00 – 11:00, εικονική αίθουσα R2 - <https://zoom.us/j/96592600632>, Meeting ID: 965 9260 0632.
- Σελίδα Μαθήματος:
 - <https://people.iese.edu/~lakritidis/dsys.php>
- Διδάσκοντας
 - Λεωνίδα Ακριτίδης
- Βαθμολόγηση
 - 3 εργασίες μέσα στο εξάμηνο – 1 μονάδα έκαστη.
 - Εξέταση – 7 μονάδες.



Βιβλία και Υλικό μαθήματος

- Παρουσιάσεις των διαλέξεων
 - www.cslab.ece.ntua.gr/courses/distrib
 - Βιβλίο
- Κατανεμημένα Συστήματα, G.Coulouris, J. Dollimore, T. Kindberg, G. Blair, Εκδόσεις DA VINCI, κωδικός Ευδόξου: 77112824
- Βοηθητικά βιβλία
 - Κατανεμημένα Συστήματα - Αρχές και Υποδείγματα, Andrew S. Tanenbaum, Maarten Van Steen, Εκδόσεις Κλειδάριθμος,
 - N. Lynch: Distributed Algorithms



Αντικείμενο μαθήματος

- Η δομή των κατανεμημένων συστημάτων, αρχιτεκτονικές και μοντέλα.
- Συγχρονισμός και Ρολόγια.
- Συντονισμός, αλγόριθμοι εκλογής αρχηγού, αμοιβαίος αποκλεισμός.
- Δοσοληψίες.
- Συστήματα κατανεμημένης αποθήκευσης.
- Parallelization Frameworks: MapReduce/Spark.
- Δίκτυα P2P.



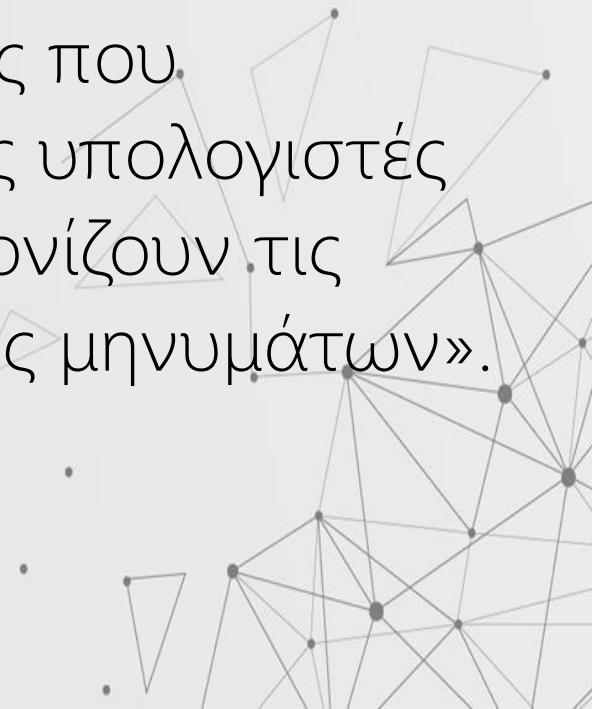
02

ΕΙΣΑΓΩΓΗ



Ορισμός Κατανεμημένου Συστήματος

- «Το κατανεμημένο σύστημα είναι μια συλλογή από αυτόνομους υπολογιστές που συνδέονται μεταξύ τους μέσω ενός δικτύου και χρησιμοποιούν ειδικά σχεδιασμένο λογισμικό για την παροχή ενοποιημένων υπολογιστικών υπηρεσιών».
- «Σε ένα τέτοιο σύστημα οι διεργασίες που εκτελούνται από τους δικτυωμένους υπολογιστές επικοινωνούν μεταξύ τους και συντονίζουν τις κινήσεις τους μόνο μέσω ανταλλαγής μηνυμάτων».



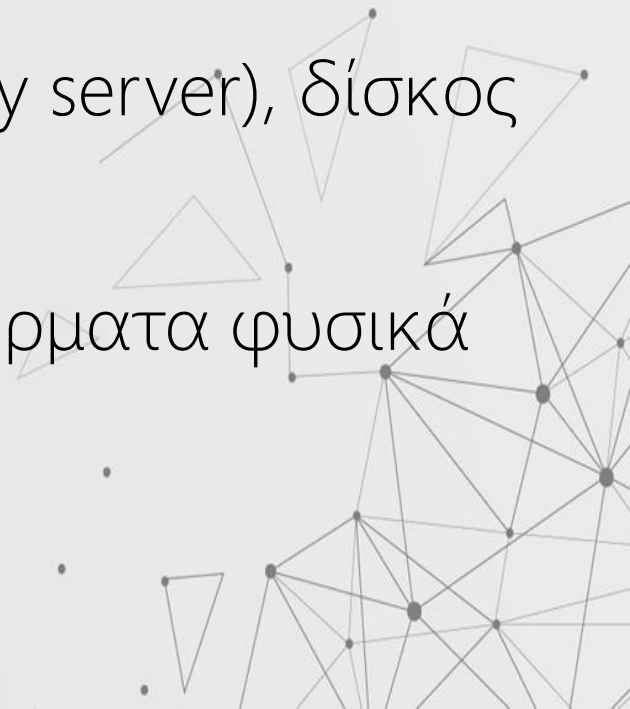
Κύρια Ωφέλη

- Διαμοιρασμός πόρων:
- Υλικό.
- Λογισμικό.
- Δεδομένα.



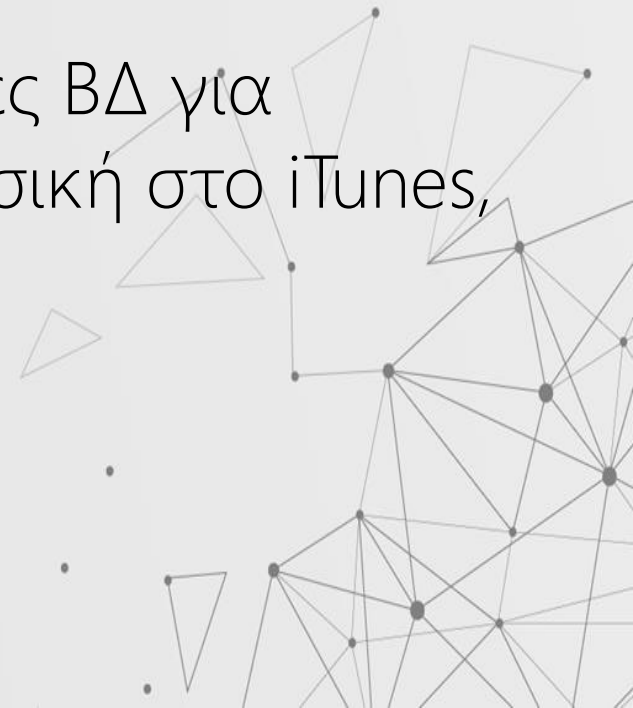
Διαμοιρασμός Υλικού

- **Υπολογιστική ισχύς (CPU):** Κάθε είδους εξυπηρετητής, εξυπηρετητές υπολογισμών σε αρχιτεκτονικές thin client, εφαρμογές τύπου SETI@home.
- **Περιφερειακά:** εκτυπωτές, scanners, επιστημονικά και πειραματικά όργανα.
- **Αποθηκευτικός χώρος:** μνήμη (proxy server), δίσκος (file/DB server).
- **Μέσο μετάδοσης:** Ασύρματα ή ενσύρματα φυσικά δίκτυα.



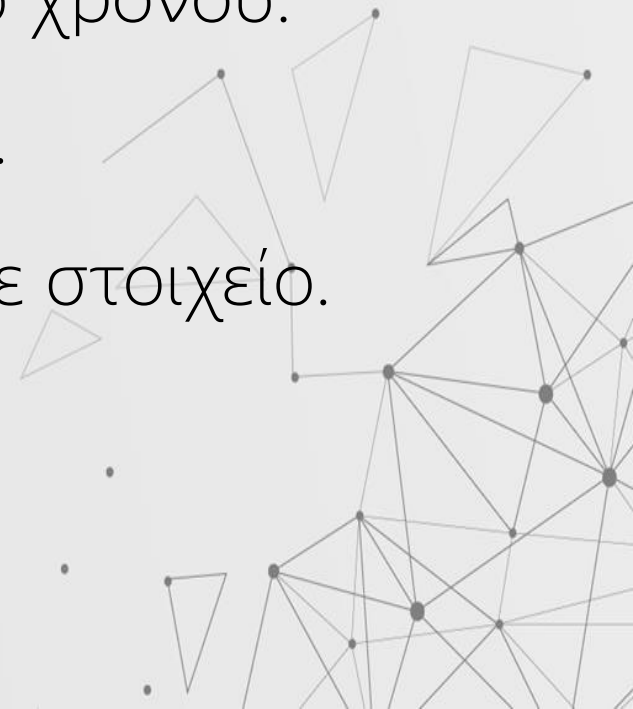
Διαμοιρασμός λογισμικού/δεδομένων

- Ιστοσελίδες, είτε στατικές (π.χ. το υλικό μιας διάλεξης), είτε δυναμικές (π.χ. για την υποστήριξη web-banking).
- Εφαρμογές, π.χ. μια μηχανή αναζήτησης στο Διαδίκτυο.
- Βάσεις Δεδομένων, π.χ. Γεωγραφικές ΒΔ για συστήματα εντοπισμού θέσης, μουσική στο iTunes, επεισόδια του Game of Thrones.
- Αρχεία, σε έναν file server.



Βασικά Χαρακτηριστικά

- Τα βασικά χαρακτηριστικά των κατανεμημένων συστημάτων είναι:
 1. Ταυτοχρονισμός των στοιχείων που συμμετέχουν.
 2. Έλλειψη καθολικής εικόνας του χρόνου.
 3. Απουσία κοινόχρηστης μνήμης.
 4. Ενδεχόμενο σφάλματος σε κάθε στοιχείο.



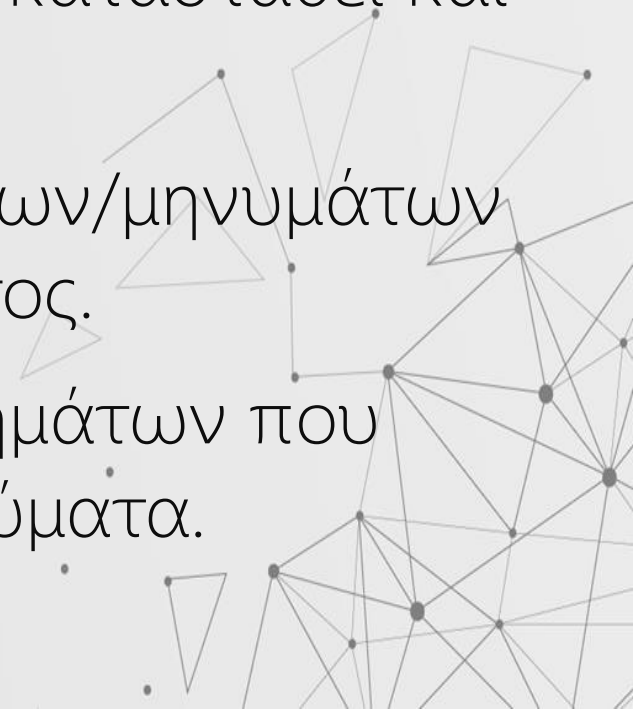
Χρησιμότητες

- Αναλογία τιμής/επίδοσης
 - Δεν είναι εύκολη/φθηνή η κλιμάκωση μέσω multi-processor.
- Κάποιες εφαρμογές είναι κατανεμημένες εκ φύσεως.
- Επίτευξη διαδραστικών επικοινωνιών.
 - π.χ. Messaging, file/photo/video sharing, gaming, telephony.
- Πρόσβαση σε απομακρυσμένο περιεχόμενο.
 - π.χ. Web browsing, BitTorrent
- Εξυπηρέτηση κινητών χρηστών:
 - Laptops, smart phones, tablets.
- Αξιοπιστία.



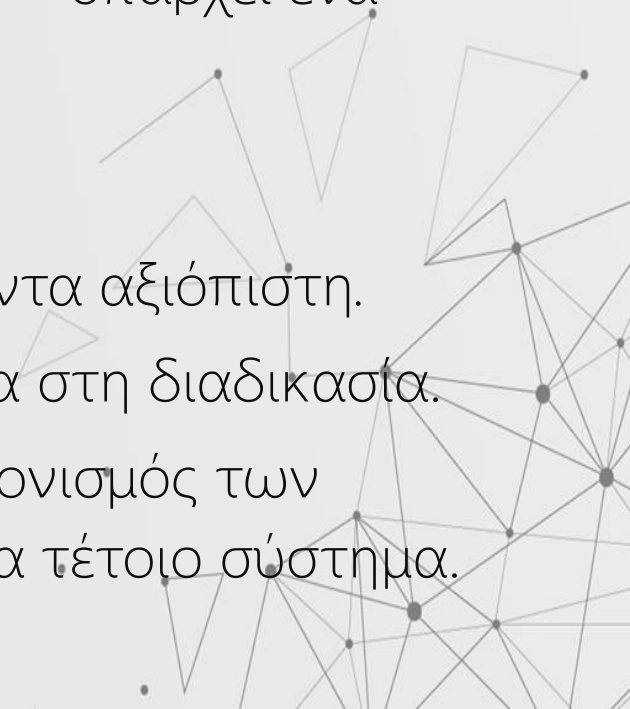
Συσχέτιση με δίκτυα υπολογιστών

- Βασικές γνώσεις δικτύωσης υπολογιστών.
- Πρωτόκολλα (TCP, UDP, IP, κλπ).
- Δρομολόγηση πακέτων και μεταφορά μηνυμάτων.
- Θεωρείται ότι το δίκτυο και οι μηχανισμοί ανταλλαγής πακέτων έχουν ήδη εγκατασταθεί και λειτουργούν.
- Ο τρόπος επεξεργασίας των πακέτων/μηνυμάτων ΔΕΝ είναι αντικείμενο του μαθήματος.
- Ενδιαφέρει όμως η σχεδίαση συστημάτων που βασίζονται σε αυτά τα πακέτα/μηνύματα.



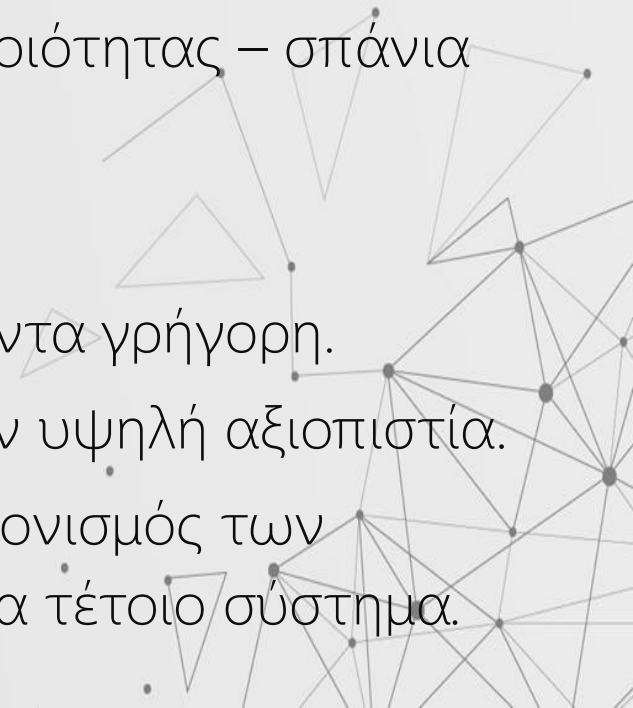
Διαφορές με λειτουργικά συστήματα

- Λειτουργικά Συστήματα:
 - Αξιόπιστοι πόροι.
 - Δεν υπάρχει πρόβλεψη για περιπτώσεις σφαλμάτων.
 - Τα σφάλματα είναι τοπικά και αντιμετωπίζονται εύκολα – π.χ. με επανεκκίνηση.
 - Δεν υπάρχουν προβλήματα συγχρονισμού – υπάρχει ένα καθολικό ρολόι.
- Κατανεμημένα Συστήματα:
 - Η επικοινωνία μέσω δικτύου δεν είναι πάντα αξιόπιστη.
 - Είναι πιθανό να έχει παρουσιαστεί σφάλμα στη διαδικασία.
 - Δεν υπάρχει ένα καθολικό ρολόι. Ο συγχρονισμός των διεργασιών είναι ζωτικής σημασίας για ένα τέτοιο σύστημα.



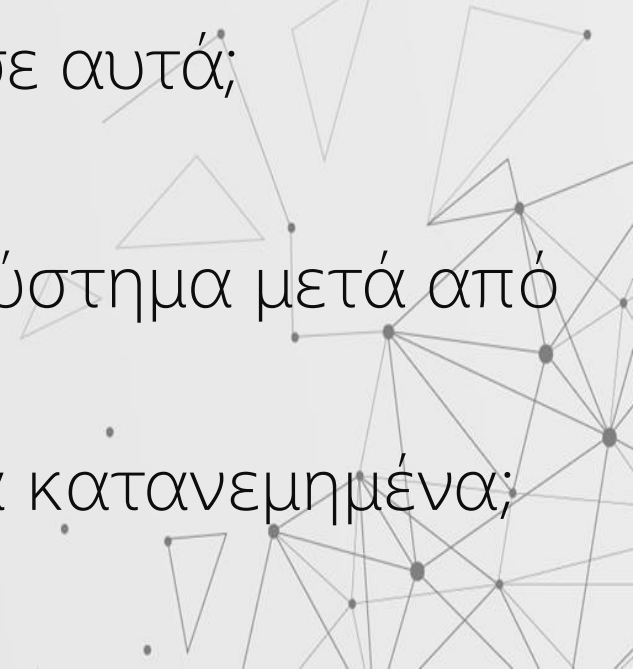
Διαφορές με παράλληλα συστήματα

- Παράλληλα Συστήματα:
 - Οι πολλαπλοί επεξεργαστές είναι τοποθετημένοι στην ίδια υπολογιστική μονάδα.
 - Η επικοινωνία μεταξύ των επεξεργαστών είναι γρήγορη.
 - Δεν υπάρχουν προβλήματα συγχρονισμού – υπάρχει ένα καθολικό ρολόι.
 - Οι υπολογιστικές μονάδες είναι υψηλής ποιότητας – σπάνια αποτυγχάνουν.
- Κατανεμημένα Συστήματα:
 - Η επικοινωνία μέσω δικτύου δεν είναι πάντα γρήγορη.
 - Οι μονάδες δεν είναι απαραίτητο να έχουν υψηλή αξιοπιστία.
 - Δεν υπάρχει ένα καθολικό ρολόι. Ο συγχρονισμός των διεργασιών είναι ζωτικής σημασίας για ένα τέτοιο σύστημα.



Βασικά θέματα κατανεμημένων

- Πώς επικοινωνούν οι κόμβοι ενός κατανεμημένου συστήματος;
- Πώς ελέγχεται η πρόσβαση σε πόρους;
- Πώς έρχονται σε συμφωνία πολλοί υπολογιστές για έναν κοινό σκοπό;
- Πώς εντοπίζει ένας κόμβος πού βρίσκονται τα δεδομένα και πώς έχει πρόσβαση σε αυτά;
- Πώς ανιχνεύονται τα σφάλματα;
- Πώς εξακολουθεί να δουλεύει το σύστημα μετά από σφάλμα;
- Με ποιον τρόπο τρέχω μια εργασία κατανεμημένα;

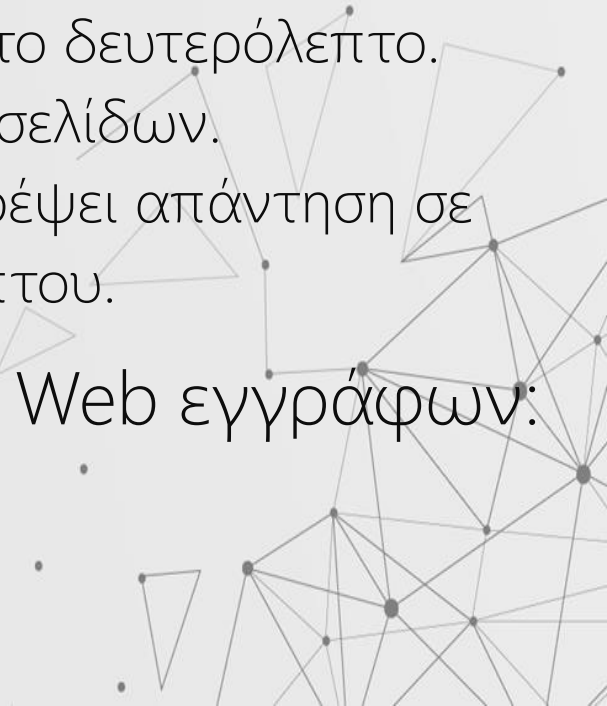


03

ΠΡΟΚΛΗΣΕΙΣ ΚΑΙ ΠΑΡΑΔΕΙΓΜΑΤΑ



Παραδείγματα

- Movie rendering:
 - Η ταινία Disney Cars 2 χρειαζόταν 11.5 ώρες rendering για κάθε frame.
 - Το Monsters University απαιτούσε 29 ώρες για κάθε frame.
 - Συνολικός χρόνος: πάνω από 100 εκατομμύρια CPU hours, 5K AMD processors.
 - Google:
 - Εξυπηρετεί πάνω από 50,000 αναζητήσεις το δευτερόλεπτο.
 - Αναζητήσεις επί περισσότερων από 70 δις σελίδων.
 - Χρησιμοποιεί χιλιάδες servers για να επιστρέψει απάντηση σε χρόνους μικρότερους του ενός δευτερολέπτου.
 - Για τη δεικτοδότηση (indexing) 5×10^7 Web εγγράφων:
 - 1999: ένας μήνας
 - 2012: ένα λεπτό
- 

Προκλήσεις

- Η σχεδίαση και η υλοποίηση ενός κατανεμημένου συστήματος δεν είναι απλή γιατί:
- Εμπλέκεται μεγάλος αριθμός υπολογιστών (10^3 - 10^6).
- Οι υπολογιστές και το υλικό δικτύωσης (routers, switches, κλπ) εμφανίζουν βλάβες:
 - Yahoo!: 50 μηχανήματα αποτυγχάνουν κάθε μέρα (στα 20K)
 - Google: 1 δίσκος αποτυγχάνει ανά 6 ώρες (στους 16K)
- Οι ενημερώσεις του υλικού και του λογισμικού απαιτούν ιδιαίτερους χειρισμούς.
- Η κατανάλωση ενέργειας είναι τεράστια.
- Η απαγωγή θερμότητας είναι δισεπίλυτο πρόβλημα.

Μηχανή αναζήτησης μεγάλης κλίμακας

- Αποτελείται από τρία βασικά κατανεμημένα υποσυστήματα:
- Crawlers: Σάρωσή του (ορατού) Web, λήψη και αποθήκευση αντιγράφων Web σελίδων στο αποθετήριο (repository, Doc Servers) της μηχανής.
- Indexers: Επεξεργασία του περιεχομένου των σελίδων (αφαίρεση HTML tags, κλπ), εξαγωγή του κειμένου, επεξεργασία κειμένου (casefolding, stemming, κλπ), και κατασκευή υποστηρικτικών δομών δεδομένων (αντεστραμμένο ευρετήριο, Index Servers).
- Query processors: Απάντηση ερωτημάτων χρηστών.

Αντεστραμμένο Ευρετήριο

- Λεξικό: Όλες οι λέξεις της συλλογής εγγράφων.
- Αντεστραμμένο αρχείο: Για κάθε λέξη, δημιουργείται μία λίστα με τα έγγραφα που την περιέχουν.
- Κάθε έγγραφο λαμβάνει μοναδικό ακέραιο αναγνωριστικό (document ID).
- Επιπλέον πληροφορία μπορεί να περιλαμβάνει συχνότητες εμφάνισης, καθώς και τη θέση μίας λέξης μέσα σε ένα έγγραφο.
- Οι λίστες σε ένα repository 7×10^{10} είναι τεράστιες.
- Ανάγκη διαμοιρασμού τους σε πολλαπλά μηχανήματα (index servers).

Αντεστραμμένο Ευρετήριο

Vocabulary	n_i	Occurrences as full inverted lists
to	2	[1,4,[1,4,6,9]], [2,2,[1,5]]
do	3	[1,2,[2,10]], [3,3,[6,8,10]], [4,3,[1,2,3]]
is	1	[1,2,[3,8]]
be	4	[1,2,[5,7]], [2,2,[2,6]], [3,2,[7,9]], [4,2,[9,12]]
or	1	[2,1,[3]]
not	1	[2,1,[4]]
I	2	[2,2,[7,10]], [3,2,[1,4]]
am	2	[2,2,[8,11]], [3,1,[5]]
what	1	[2,1,[9]]
think	1	[3,1,[2]]
therefore	1	[3,1,[3]]
da	1	[4,3,[4,5,6]]
let	1	[4,2,[7,10]]
it	1	[4,2,[8,11]]

To do is to be.
To be is to do.

d_1

To be or not to be.
I am what I am.

d_2

I think therefore I am.
Do be do be do.

d_3

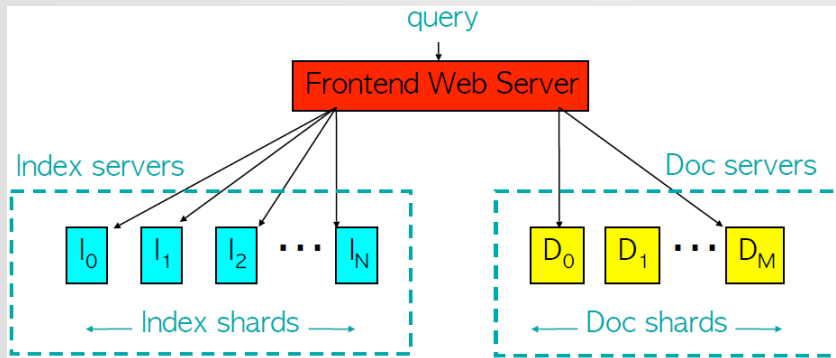
Do do do, da da da.
Let it be, let it be.

d_4

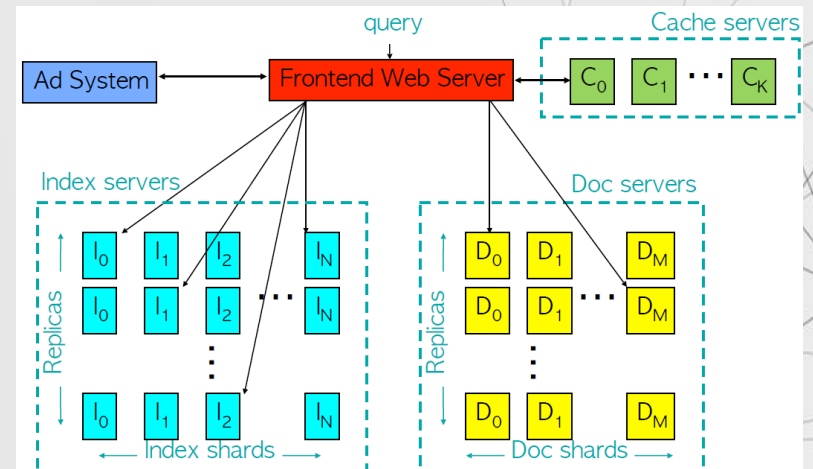
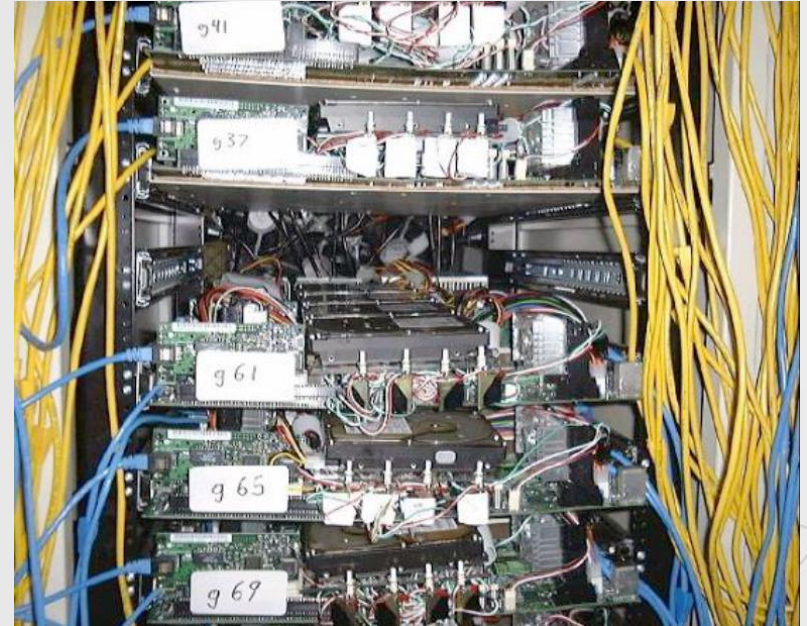
- Πώς κατασκευάζεται?
- Πώς διανέμεται?

Το σύστημα αναζήτησης της Google

Το σύστημα το 1997



Το σύστημα το 1999

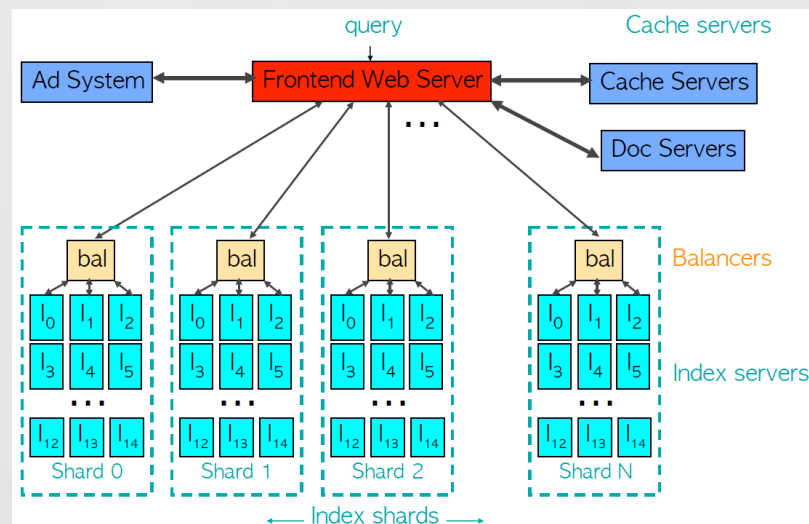


Το σύστημα αναζήτησης της Google

Data Center 2000



Data Center 2001



Το σύστημα αναζήτησης της Google

- Δεκάδες data centers ανά τον πλανήτη.
- Σε κάθε ένα από αυτά, εκατομμύρια υπολογιστικοί πυρήνες σε χιλιάδες μηχανήματα.
- 20000 μηχανήματα για ένα task.





04

ΥΛΗ ΜΑΘΗΜΑΤΟΣ

Ύλη μαθήματος (1)

- Εισαγωγή.
 - Αρχιτεκτονικά μοντέλα.
- Συγχρονισμός.
 - Φυσικά Ρολόγια.
 - Λογικά Ρολόγια.
 - Συγχρονισμός φυσικών και λογικών ρολογιών.
 - Καθολικές Καταστάσεις.
- Κατανεμημένος Συντονισμός.
 - Ομαδική επικοινωνία.
 - Αλγόριθμοι Εκλογής Αρχηγού.
 - Αλγόριθμοι Αμοιβαίου Αποκλεισμού.
 - Κατανεμημένος αλγόριθμος ομοφωνίας Paxos.



Ύλη μαθήματος (2)

- Δοσοληψίες.
 - Ιδιότητες ACID.
 - Έλεγχος ταυτοχρονισμού.
 - Κλείδωμα 2 φάσεων.
 - Διάταξη χρονοσφραγίδων.
 - Αισιόδοξος έλεγχος ταυτοχρονισμού.
- Κατανεμημένες Δοσοληψίες.
 - Ατομικές δοσοληψίες.
 - 2 phase commit.
 - 3 phase commit.
 - Έλεγχος ταυτοχρονισμού κατανεμημένων δοσοληψιών.
 - Αλγόριθμοι εντοπισμού αδιεξόδων.
 - Ανάνηψη από σφάλματα.



Ύλη μαθήματος (3)

- Αντίγραφα δεδομένων και διαχείρισή τους.
 - Το θεώρημα CAP.
 - Μοντέλα συνέπειας.
 - Πρωτόκολλο gossip.
- Δίκτυα Ομότιμων Κόμβων (P2P).
 - Κατηγορίες Δικτύων P2P.
 - Κατανεμημένοι Πίνακες Κατακερματισμού (DHT).
 - Βασικές λειτουργίες DHT.
 - Παράδειγμα: το σύστημα Chord.
- Κατανεμημένα Συστήματα Αρχείων.
 - HDFS, Google File System, κλπ.



Ύλη μαθήματος (4)

- Μοντέλα Κατανεμημένης Επεξεργασίας.
 - Προγραμματιστικό πλαίσιο εργασίας MapReduce (Apache Hadoop).
 - Προγραμματιστικό πλαίσιο εργασίας Apache Spark.
- Cloud computing.
 - Οι πόροι και οι υποδομές ως υπηρεσίες.
 - Απομακρυσμένοι αποθηκευτικοί χώροι.
 - Απομακρυσμένοι πόροι επεξεργασίας.

