DSC516: Cloud Computing

Marios D. Dikaiakos

http://www.cs.ucy.ac.cy/mdd

University of Cyprus
Department of Computer Science

About the Course

- Post-graduate course offered as restricted choice for the M.Sc. in Data Science, M.Sc. in Computer Science, Professional M.Sc. in Advanced Information Technologies, and Ph.D. in Computer Science programs. The course is also open to senior undergraduate students.
- Lectures: Tuesday, Friday 15:00-16:29, ΘΕΕ01-148.
- Recitation: Tuesday, 16:30-17:29, online or ΘΕΕ01-148.
- Lab: Friday, 16:30-17:29, ΘΕΕ01-148 or online.

Objective

- Provide an introduction to and understanding of advanced concepts in the field of Cloud Computing
- Enable students to design, develop, deploy, monitor and analyze applications on state-of-the-art Cloud computing platforms.
- Covers key elements and technologies of Cloud Computing Infrastructures, Services, and Applications.
- You will gain an understanding of the Cloud Computing paradigm and the technical underpinnings of Cloud services; they will be able to describe and analyze key middleware components of Cloud services, to understand the main Cloud application development paradigms, and to use state-of-the-art Cloud service offerings for Data Science-related projects.

Web Site

University of Cyprus MSc in Data Science

DSC516/EPL602: Cloud Computing

Homepage

Fall Semester 2023

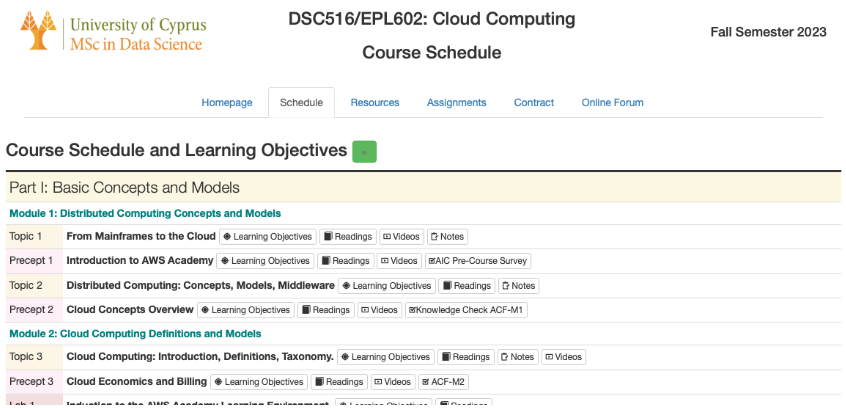
Homepage | Schedule | Resources | Assignments | Syllabus | Online Forum

Objectives

This course covers key elements and technologies of Cloud Computing Infrastructures, Services, and Applications. Students who attend this course will gain an understanding of the Cloud Computing paradigm and the technical underpinnings of Cloud services. They will be able to describe and analyze key middleware components of Cloud services, to understand the main Cloud application development paradigms, and to use state-of-the-art Cloud service offerings for Data Science-related projects. Precepts and labs will help students prepare for AWS Certification.

Instructor: Marios D. Dikaiakos, Professor. »
Teaching Assistant: Dr. Pyrros Bratskas, Special Teaching Staff. »
Prerequisites: Basic knowledge in Programming and Data Structures, Operating Systems, Networking, Parallelism.
Lectures: Tuesday, Friday, 15:00-16:29, ΘΕΕ01-148.
Recitation: Tuesday, 16:30-17:29 (online or ΘΕΕ01-148)
Labs: Friday, 16:30-17:59, ΘΕΕ01-148 or online.
Online Forum: edstem

http://www.cs.ucy.ac.cy/courses/DSC516



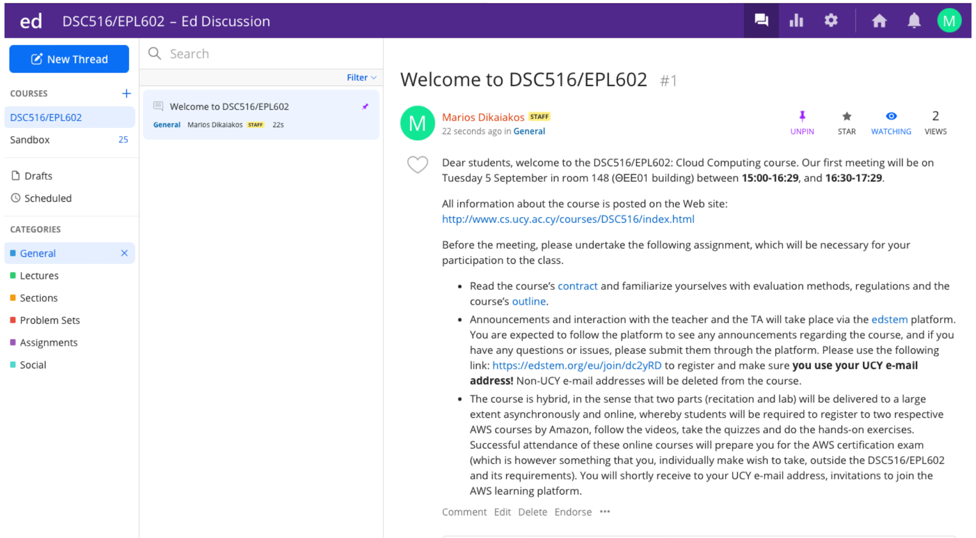
Evaluation, Grading and ECTS

- Student progress is evaluated continuously through class participation and the assessment of in-class presentations, writing assignments, group project deliverables, and final exam. The final grade is based on the following formula:
- **Final grade:**
  - Projects and Homeworks: 20%
  - Midterm Exam: 30%
  - Final Exam: 50%
- ECTS Analysis
  - One ECTS unit corresponds to 25-30 hours of work undertaken by an average student to complete successfully expected learning outcomes.
  - The total workload for successfully completing this course, is estimated to **240 hours on average**.

Textbooks

- Marinescu, D. **Cloud Computing: Theory and Practice**. Morgan Kaufmann, 2017.
- Peterson, Baker, Bavier, Williams and Davie, **Edge Cloud Operations: A Systems Approach**, version v0.2, 2022
- Luiz André Barroso, Urs Hölzle, and Parthasarathy Ranganatha, **"The Datacenter as a Computer: Designing Warehouse-Scale Machines,"** Third Edition Morgan & Claypool (2019).
- Ian Foster and Dennis B. Gannon, **"Cloud Computing for Science and Engineering"**, MIT Press (2017).
- Jeff Nickoloff and Stephen Kuenzli, **"Docker In Action,"** Manning (2019) . (First version was available on github).
- Betsy Beyer at al. **"Site Reliability Engineering,"** O'Reily (2016).
- Rajkumar Buyya and Satish Narayana Srirama, eds. **"Fog and Edge Computing,"** Wiley (2019).
- Thomas Earl and Zaigham Mahmood and Ricardo Puttini. **"Cloud Computing: Concepts, Technology and Architecture,"** Pearson (2014).

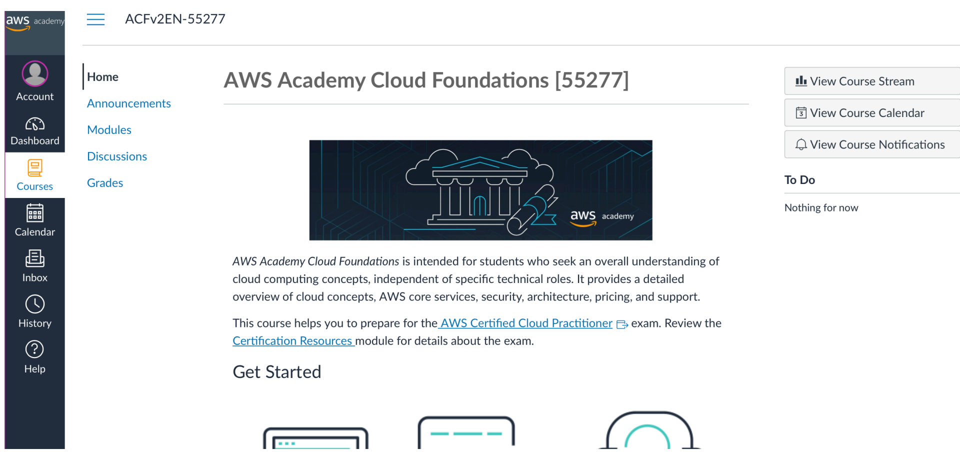
Online Forum: Edstem



Assignments

Assignment 1

- Please create an account on edstem platform using: https://edstem.org/eu/join/dc2yRD
  - You are expected to follow the platform to see any announcements regarding the course, and if you have any questions or issues, please submit them through the platform.
  - To register and make sure you use your UCY e-mail address! Non-UCY e-mail addresses will be deleted from the course.
- Use **your full name**, so that you are recognizable
- Connect and get familiar with the AWS Academy (invitation in your mailbox)



DSC516: Cloud Computing

Part I: Basic Concepts and Models

Module1: Distributed Computing Concepts and Models

Lecture 1

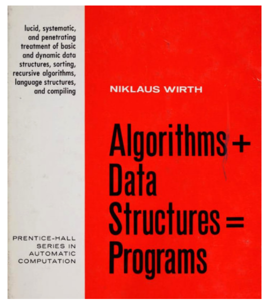
From Mainframes to the Cloud

WHERE DO YOU START FROM, WHEN YOU WANT TO THINK ABOUT A PROBLEM IN COMPUTING?

A Conceptual Framework

A Conceptual Framework

- What is a program?
- What is the input?
- What is the output?
- What is the computer?



Programs = Algorithms + Data Structures

Concepts & Tools

- Algorithms, data structures
- Architectures and Models
- Programming abstractions and Programming Languages + Programming Libraries
- Data encoding and storage
- Communication protocols
- Abstractions
- Performance, Reliability

WHAT IS CLOUD COMPUTING?

What is Cloud Computing?

Moving computing and data away from the desktop and the portable PC and simply displaying the results of computing that takes place in a centralised location and is then transmitted via the Internet on the user's screen

John Markoff, New York Times (circa 2007)

Starts with the premise that the data, services and architecture should be on servers. We call it cloud computing. They should be in a 'cloud' somewhere.

Eric Schmidt, Google

The practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer.

Oxford Languages Dictionary

Cloud computing is the on-demand availability of computer system resources, especially data storage (cloud storage) and computing power, without direct active management by the user.

Wikipedia

The delivery of computing services—including servers, storage, databases, networking, software, analytics, and intelligence—over the Internet ("the cloud")

Microsoft

The cloud is just another name for somebody else's computer

Graham Cluley

IN ONE WORD, WHAT IS THE COMMON THEME UNDERLYING ALL THESE DEFINITIONS?

What is Cloud Computing?

Moving computing and data away from the desktop and the portable PC and simply displaying the results of computing that takes place in a centralised location and is then transmitted via the Internet on the user's screen

John Markoff, New York Times (circa 2007)

Starts with the premise that the data, services and architecture should be on servers. We call it cloud computing. They should be in a 'cloud' somewhere.

Eric Schmidt, Google

The practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer.

Oxford Languages Dictionary

Cloud computing is the on-demand availability of computer system resources, especially data storage (cloud storage) and computing power, without direct active management by the user.

Wikipedia

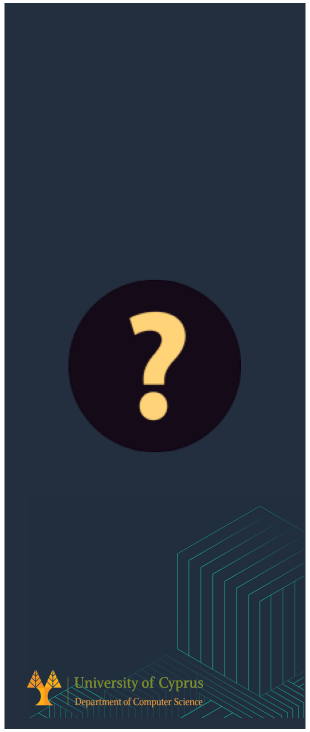
The delivery of computing services—including servers, storage, databases, networking, software, analytics, and intelligence—over the Internet ("the cloud")

Microsoft

The cloud is just another name for somebody else's computer

Graham Cluley

Location



WHY IS LOCATION SO IMPORTANT?

Location implies possession

Possession implies control

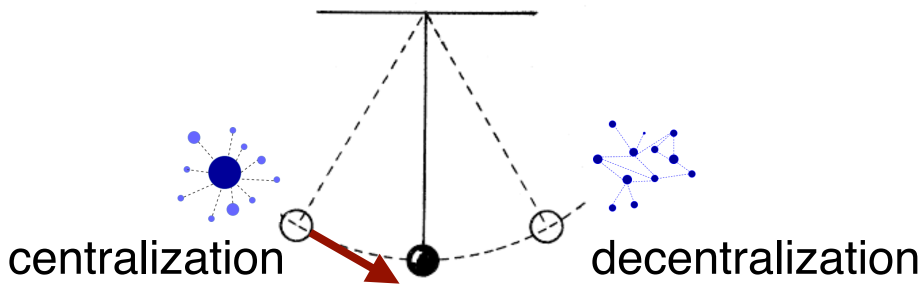
Control implies power

Source: Ed Felten, Princeton Univ.

From Mainframes to the Cloud

Historical Overview

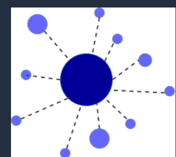
Centralization to decentralization



WHICH FORCE(S) DRIVE THE PENDULUM?

Historical Overview

Centralization and Mainframes



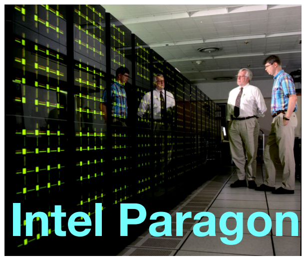
Centralization: mainframes

1960's-1980's

Centralization: time-sharing

1960's-1980's

Supercomputers and HPC

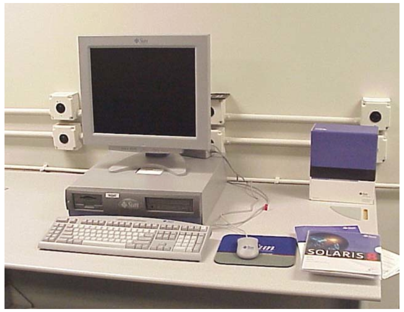


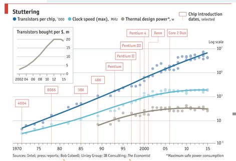
1970's-1990's

Historical Overview

Personal Computing

Personal Computing: some degree of decentralization (thanks to Moore's law)





1980's-1990's

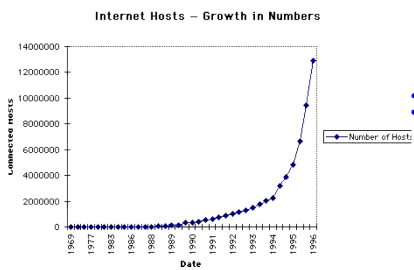
Centralized vs Decentralized

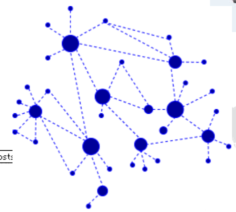
Centralized	Decentralized
• At the mainframe	• With the end-user
• High-Performance	• Cheaper
• Cost efficiency: time-sharing , amortizing building cost (economies of scale)	• Richer user interface
• Expert management	• Higher bandwidth available locally
• Programs: serial & parallel	• Greater autonomy
	• Programs: serial

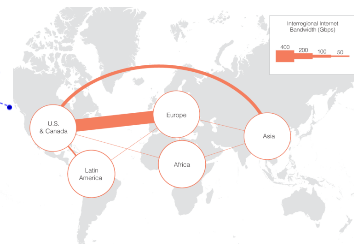
Historical Overview

Client-Server Computing

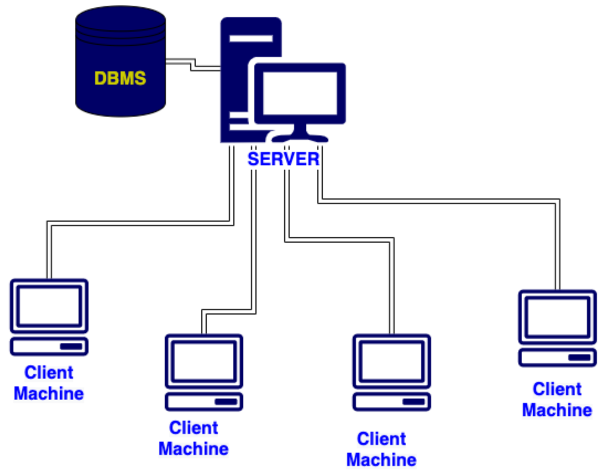
Client-Server Computing

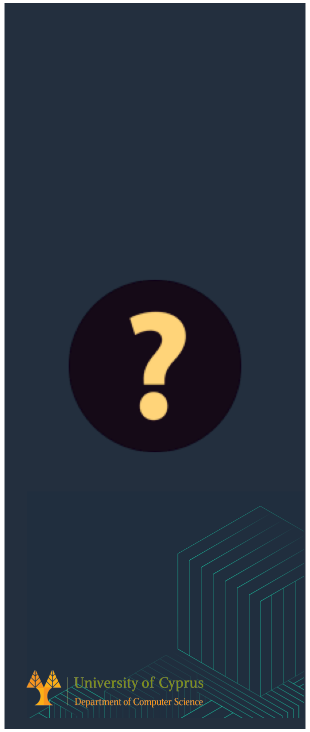






Client-Server Computing

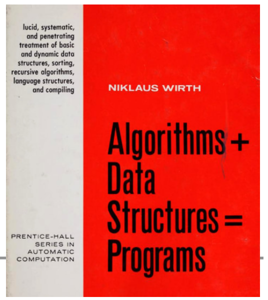




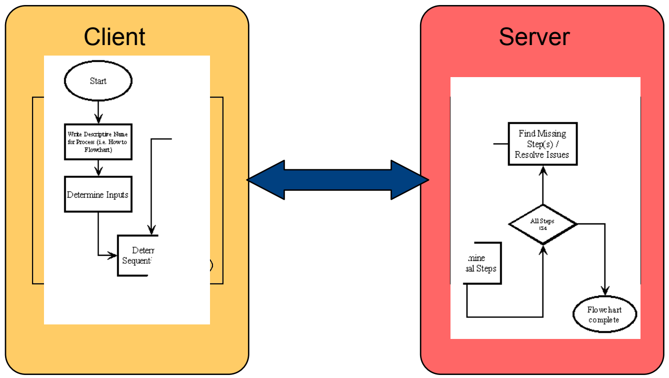
WHAT ABOUT APPLICATION DEVELOPMENT?

Application Blueprint

Programs = Algorithms + Data Structures

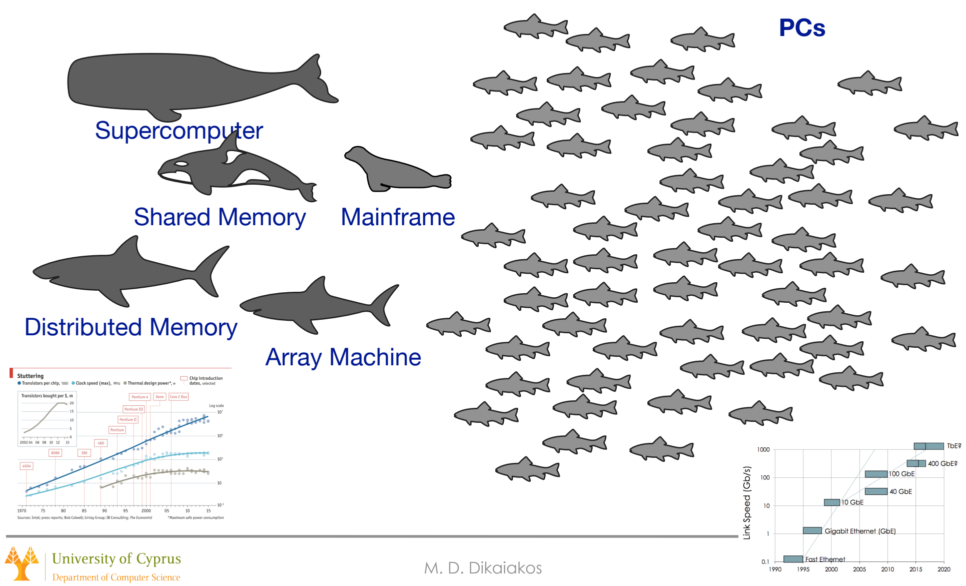


Client-Server Challenges

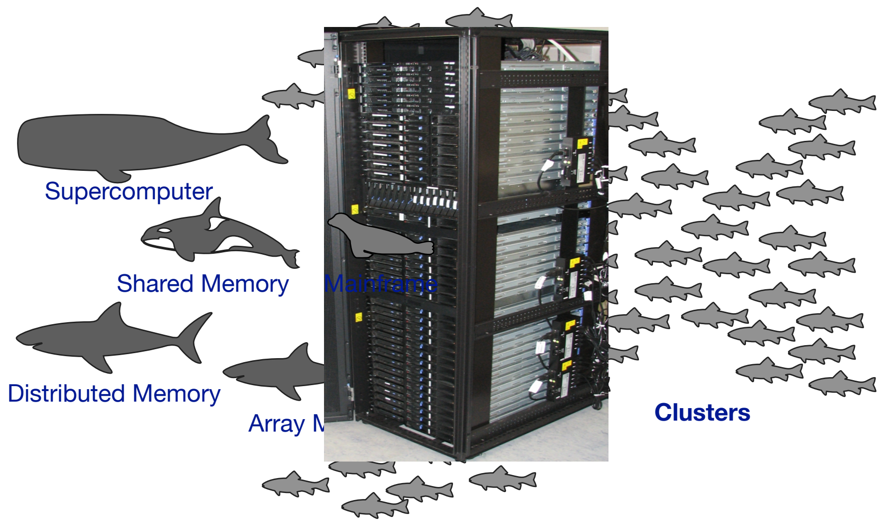


Decomposition?
Location of Data?
Transfer of Data?

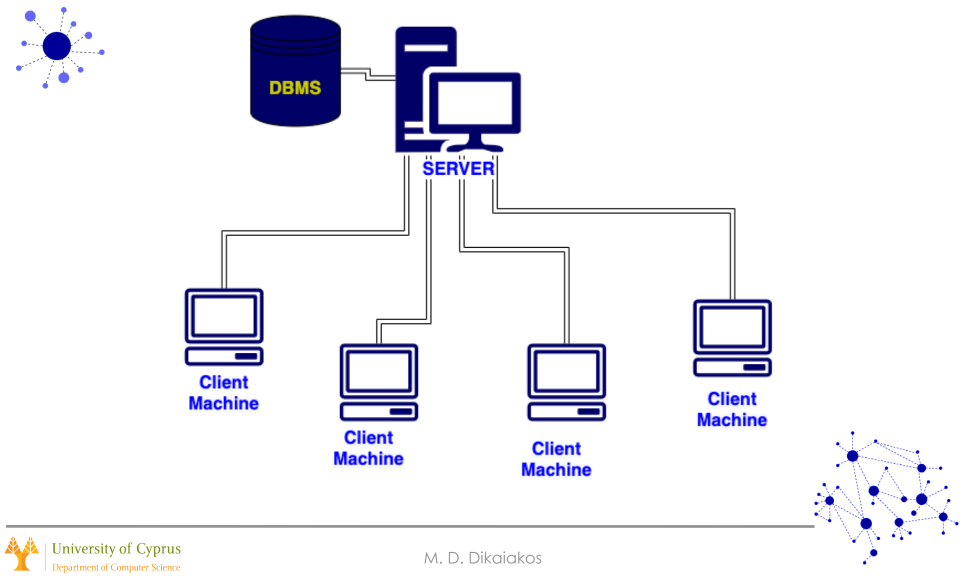
The “Killer” Micros



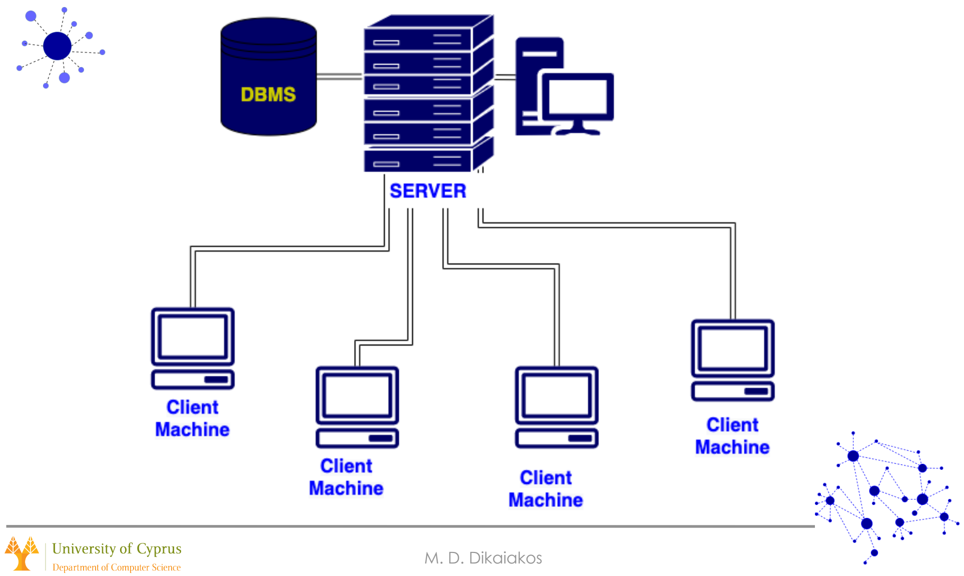
Cluster Computing



Client-Server Computing



Client-Server Computing

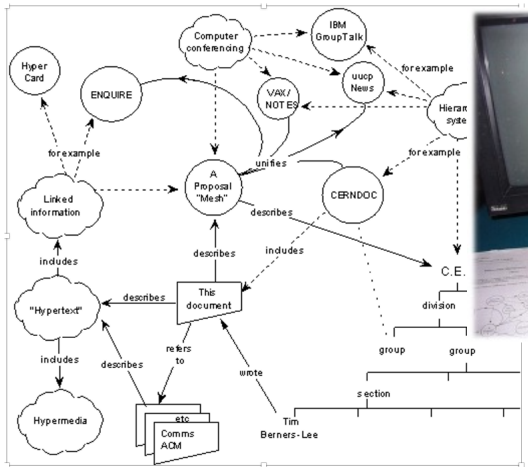


Challenges arising from client-server

- Lack of standards in hardware and software
  - Application complexity grew
  - Corporate infrastructure complexity grew
- Server nodes typically dedicated to running just one application
  - Extraordinarily low levels of capacity utilisation!
  - Peak load rarely reached
- Dense packing of computing nodes
  - Electricity consumption skyrocketed
  - Power costs comparable to hardware acquisition costs

Historical Overview

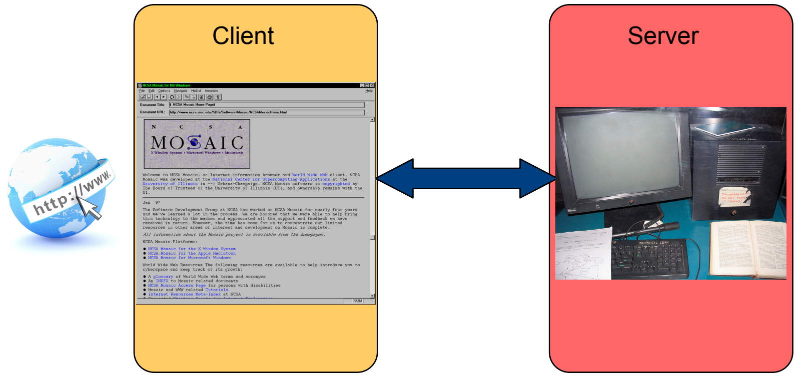
Web Computing



"Information Management: A Proposal"

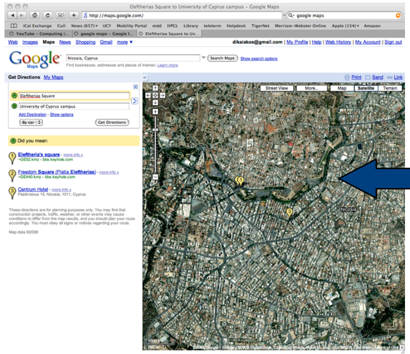
By Tim Berners-Lee, March 1989

Web 1.0 and Browsers



mid-late 90's

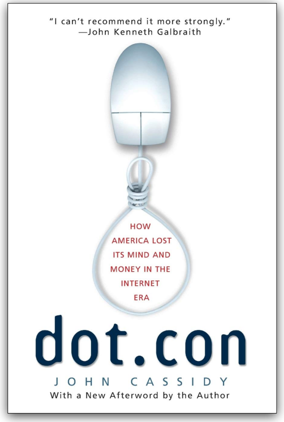
Browsers as platforms: Web 2.0



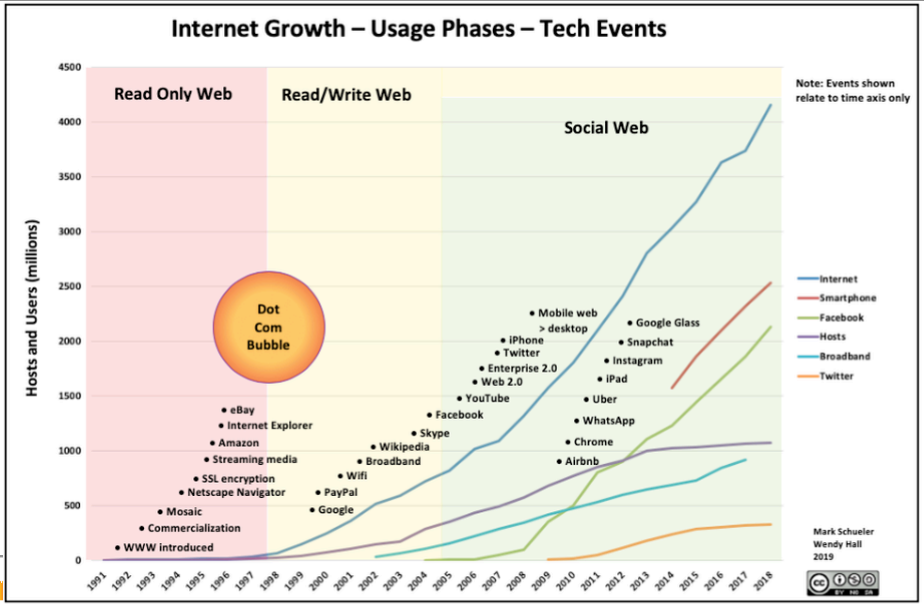
mid 90's-today

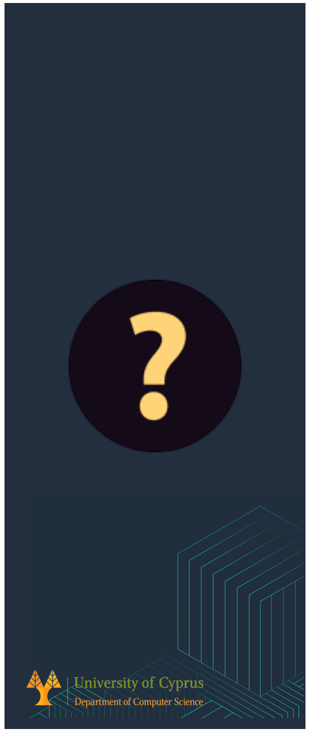
The Success of the Web

- The Web is used by:
  - billions of users
  - all kinds of users
  - for all kinds of things



A (Huge) Success Story



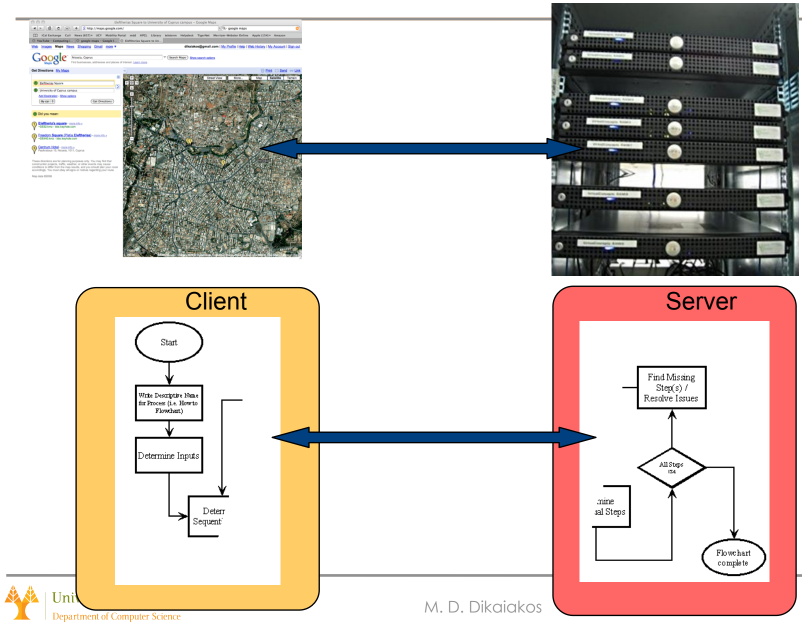


WHY DID THE WEB SUCCEED SO MUCH AGAINST PRIOR CLIENT-SERVER SYSTEMS?

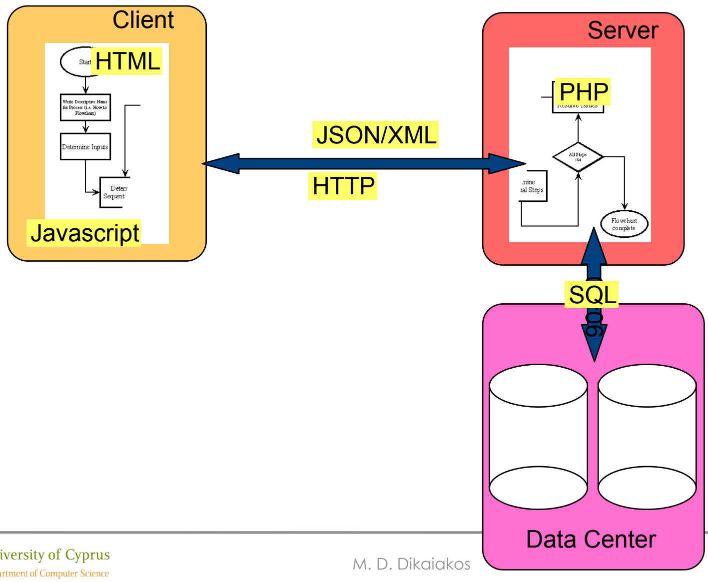
Factors of Success

- A "universal" client with a simple UI: the Web browser
- A common addressing scheme for Web resources: URI
- A common resolution/transfer mechanism: HTTP
- A common standard for encoding information: HTML & XML
- Applications that made resource discovery easy

Client-Server computing w. browsers

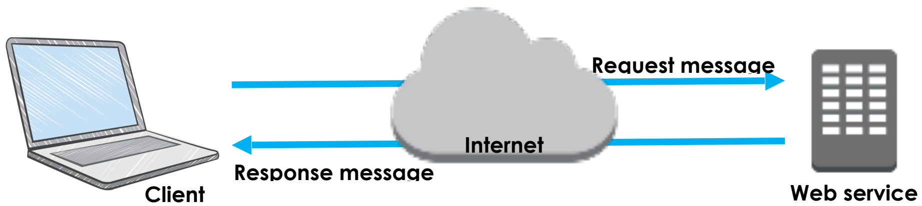


Behind the scenes

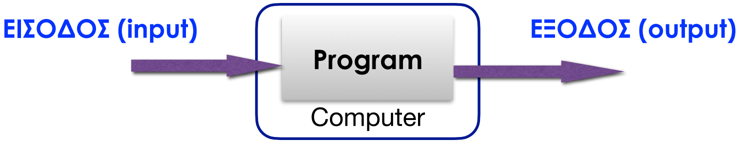


Web Services

- A web service is any piece of software that makes itself available over the internet and uses a standardized format—such as Extensible Markup Language (XML) or JavaScript Object Notation (JSON)—for the request and the response of an application programming interface (API) interaction.



Application Blueprint



- More elements become crucial:
  - **Networking**: Delivering questions (input) and answers (output).
  - **Computation**: Transforming information to produce new information (algorithms+data structures).
  - **Database access**: Access to reference information needed by the computation (more input and output).
  - **Database storage**: Long-term storage of information, needed for later access (more data structures)

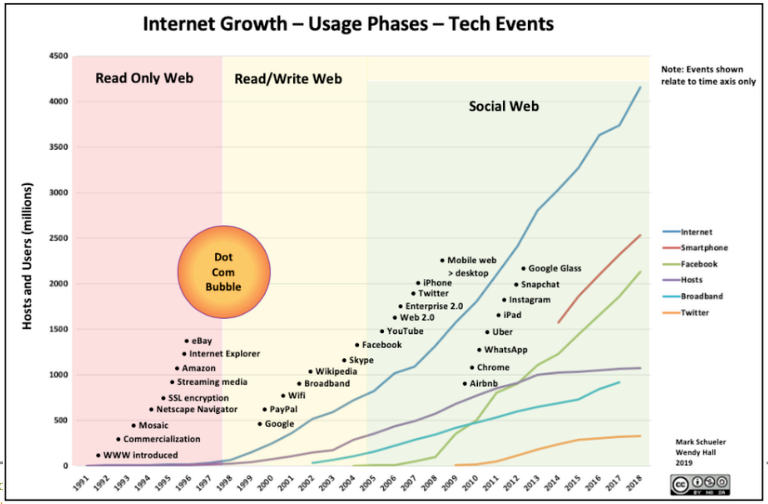
Progress and Challenges

- Widely-accepted standards for client-server communication (HTTP), resource naming & resolution (URI, DNS), data encoding (JSON, XML)
- Maturing technologies for client-server interaction (RESTful APIs), software stack (LAMP), server (PhP) & client programming (Javascript), integrated software platforms (Wordpress, Ruby on Rails, Jango, Spark).
- Multiplexing more workloads on the same cluster: improved amortization of hardware/networking/software investment
- Widely-accepted standards and best-practices for server/networking hardware, and cluster configuration, administration & management.

- High application complexity (many components & services)
- Increased need for shorter application management lifecycles (CI/CD)
- Data centers face low levels of capacity utilization. Peak load rarely reached.
- Dense packing of computing nodes - electricity consumption keeps growing
- Need to take Total-Cost of Ownership - TCO into account in procurement & service-model decisions (CAPEX, OPEX).

Exponential Age

- Driving network-effects, improved UI/UX, more users leading to critical mass formation, more investments



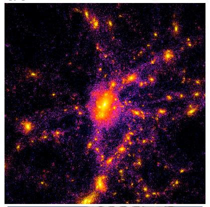
Historical Overview

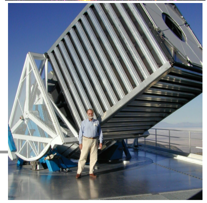
Grid Computing

e-Science: the 4th Paradigm of Scientific Exploration

- Thousand years ago science was empirical
  - describing natural phenomena
- Last few hundred years: theoretical branch
  - using models, generalizations
- Last few decades: a computational branch
  - simulating complex phenomena
- By 2007: data exploration (eScience)
  - unify theory, experiment, and simulation
  - data captured by instruments or generated by simulator
  - processed by software
  - information/knowledge stored in computer
  - scientist analyses database / files using data management and statistics
- "Computational X" and "X-Informatics" [recently: Data Science]

$$\oint \mathbf{E}\cdot d\mathbf{A} = \frac{q_{enc}}{\varepsilon_0}$$
$$\oint \mathbf{B}\cdot d\mathbf{A} = 0$$
$$\oint \mathbf{E}\cdot d\mathbf{s} = -\frac{d\Phi_B}{dt}$$
$$\oint \mathbf{B}\cdot d\mathbf{s} = \mu_0\varepsilon_0\frac{d\Phi_E}{dt} + \mu_0 i_{enc}$$

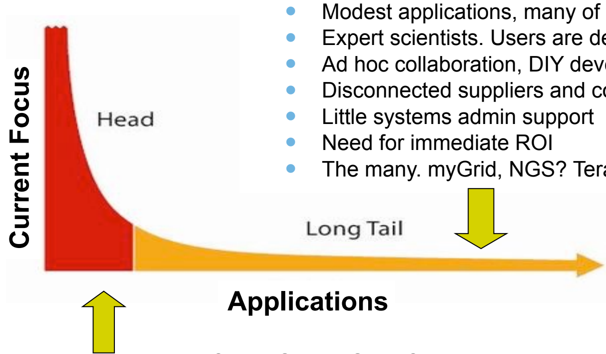




Source: **J. Gray**, talk to NRC/CSTB, "eScience - A Transformed Scientific Method." Mountain View CA, 11 January 2007.

"Democratizing e-Science"

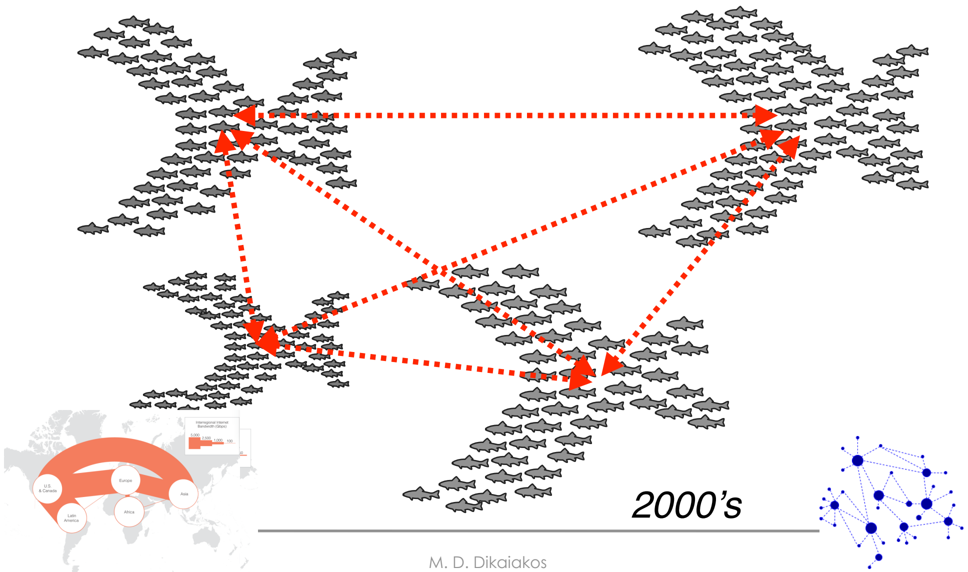
- Small labs or individuals
- Modest applications, many of them
- Expert scientists. Users are developers.
- Ad hoc collaboration, DIY development
- Disconnected suppliers and consumers.
- Little systems admin support
- Need for immediate ROI
- The many. myGrid, NGS? TeraGrid? WWG?



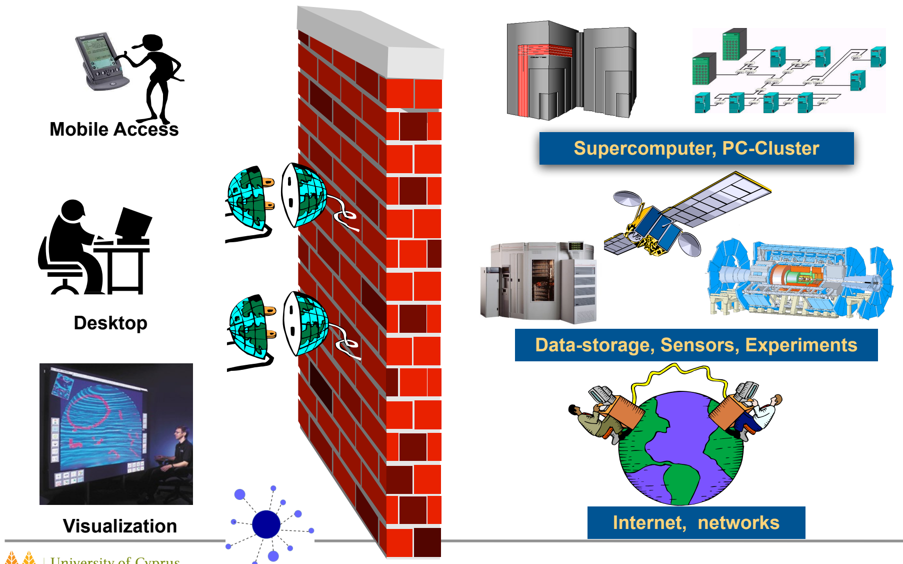
- Big enterprises (LHC, caBIG, AstroGrid). Connected communities.
- High end, specialist applications, planned collaboration, deep & narrow
- Lots of support and pain, Private, hand-rolled Grid solutions
- Prepared to invest, dedicated development. Users are not developers.
- The Few. LHC, SCEC, GEON, NeSSGrid, AstroGrid, caBIG …

Source: C. Coble, 2007

Grid Computing: Federated Clusters

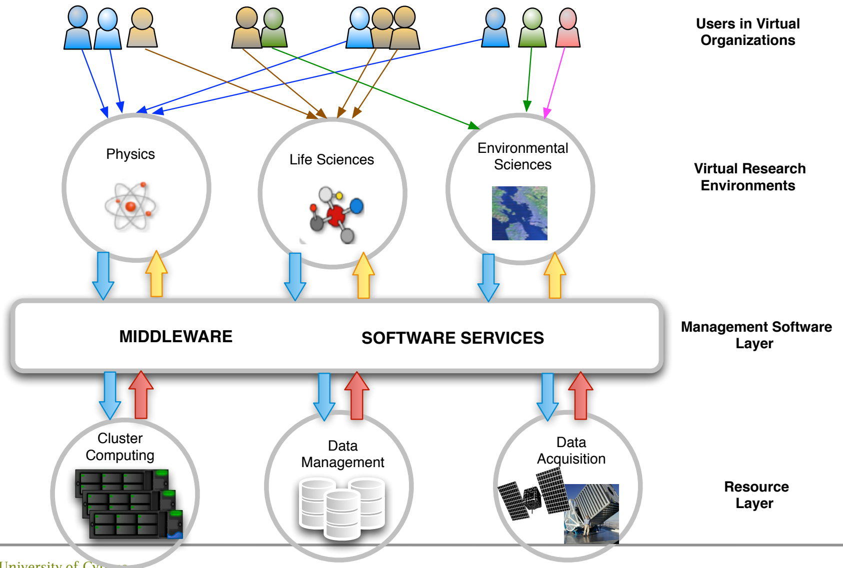


Grid Vision: Resource Virtualization

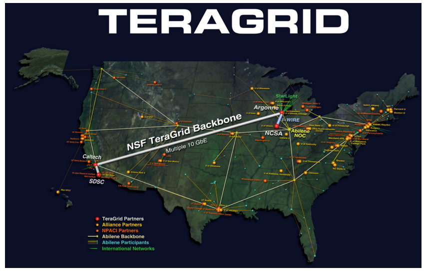


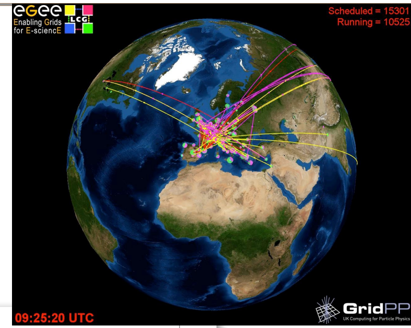
Hoffmann, Reinefeld, Putzer

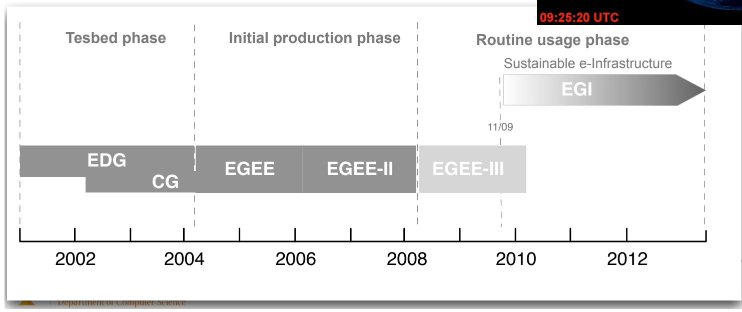
Grid Vision



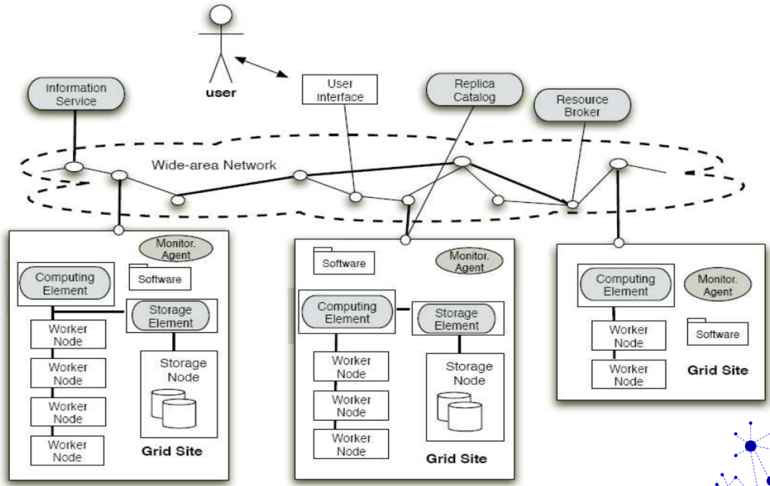
Grid Testbeds & Services







Grid Architecture



Grid Computing: Problems

- The amount of work done in Grids is still relatively small compared to the vast number of projects still being run on:
  - the desktop (81% of respondents)
  - local clusters (13% of respondents)

  Its architecture geared to HPC, high-throughput applications
- "Usability of the technology could be improved"
- From a user's point of view, most grid infrastructures of today do not yet satisfy the characteristics of reliable operation and easy access.
  - Failure management: an administrator's nightmare
- Not enough support for: VOs, Virtual Research Environments, collaborative problem solving environments and virtual labs.
- Management layers did not guarantee seamless interplay between resources, while hiding their heterogeneity.
- Grid middleware stacks: large, complex, with high operational complexity.
- Need better tools to manage the forthcoming data explosion (management, curation).

"Grid Computing's Future", K. Kroeker, CACM, 3/2011
"Perspectives on grid computing", U. Schwiegelshohn et al., FGCS (26) 2010
"Nine months in the life of EGEE: a look from the South." DaCosta, Dikaiakos, Orlando. IEEE/ACM MASCOTS 2007.
"Failure management in Grids: The case of the EGEE infrastructure", Neocleous, Dikaiakos, Fragopoulou, Markatos, PPL 2007.

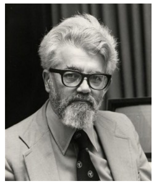
Historical Overview

Utility Computing

IT as a Public Utility: an Early Vision

"If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility...

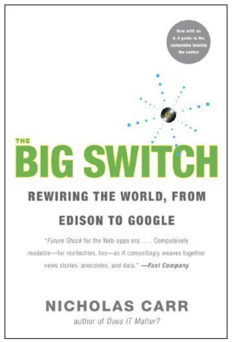
The computer utility could become the basis of a new and important industry"



John McCarthy
speaking at the MIT Centennial in 1961

"The Big Switch"

- What happened to the generation of electric power is now happening to the processing of information.
- Computing is:
  - Becoming a General Purpose Technology (GPT)
  - Delivered as a Utility
- "The **economic equations** that determine the way we work and live are being rewritten."



2008

General-Purpose Technologies

- Technologies that affect the whole economy, transforming both household life and the ways in which firms conduct business:
  - Steam, electricity, internal combustion, and information technology (IT)
- Key properties:
  - Pervasiveness: GPT should spread to most sectors.
  - Improvement : GPT should get better over time and keep lowering the costs of its users.
  - Innovation spawning: GPT should make it easier to invent and produce new products or processes.

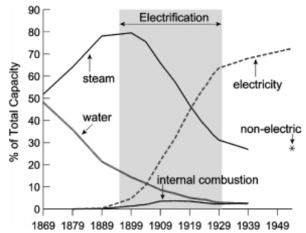


Figure 2. Shares of total horsepower generated by the main sources in U.S. manufacturing, 1869–1954.

"General Purpose Technologies"
Jovanovic & Rousseau, Handbook of Economic Growth, **2005**.

Utility Computing: Key Premises

- Resources available "on-demand:"
  - Computing, Networking, Storage
  - Software as a Service (SaaS)
- Pay-per-use:
  - Fine-grained/accurate measurement of resource consumption
  - Pricing policies that make sense
  - Payment mechanisms in place
- Frictionless use: low-barrier entry, high availability, fast access, no scarcity, minimal failures

Utility Computing: Early Obstacles

- Not enough computing capacity
- Lack of sufficient bandwidth; high latencies
- Service delivery still immature
- Lack of efficient and effective pricing schemes and trustful payment mechanisms
- Lack of critical mass and "network effects"

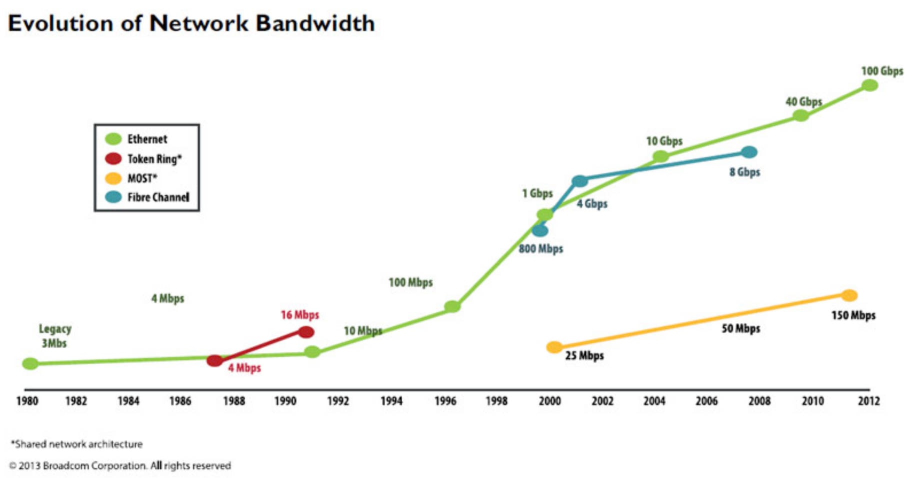
Network Effect

- The value of the network is **larger** than the sum of the value of its nodes;
- The value of the network grows "faster" than the number of its nodes (Metcalf's Law).
- Network effects occur when, from a consumer's perspective, a product becomes more useful as more people use it.

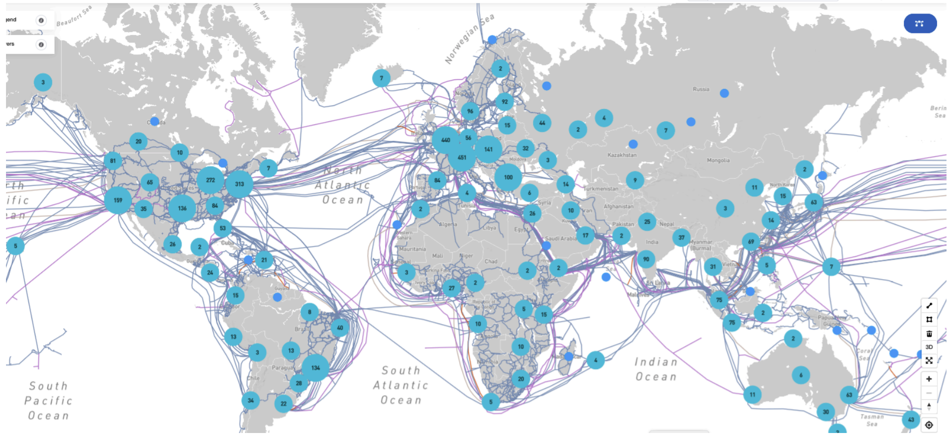
Reading Assignment

- Check Wikipedia's lemma on **Network Effect**
- If you wish to look deeper into this, you can look at the synopsis given by Sinan Aral, "**The Hype Machine**" book (chapter 5 "A Network's Gravity is Proportional to its Mass)

Networking Evolution

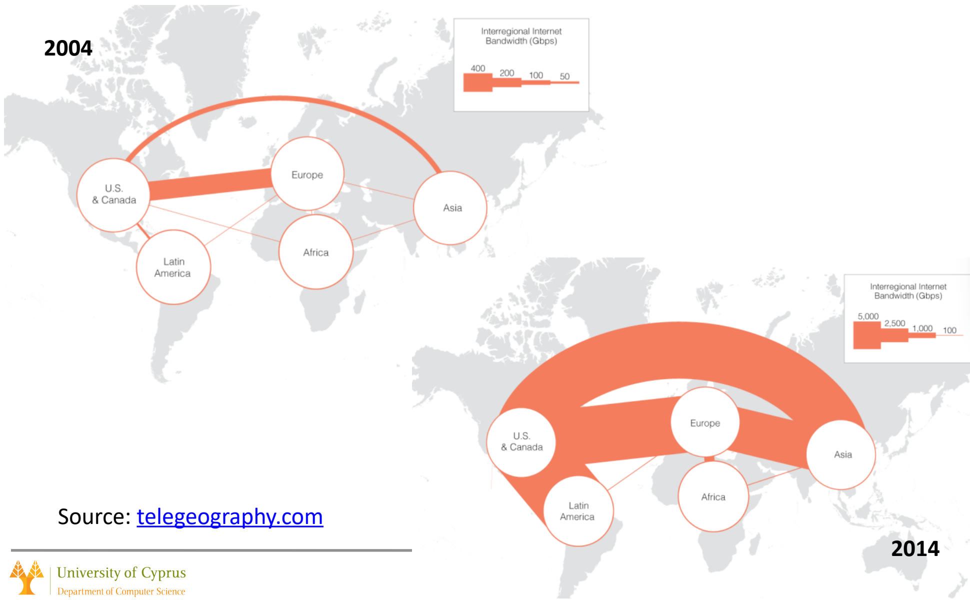


Networking Evolution



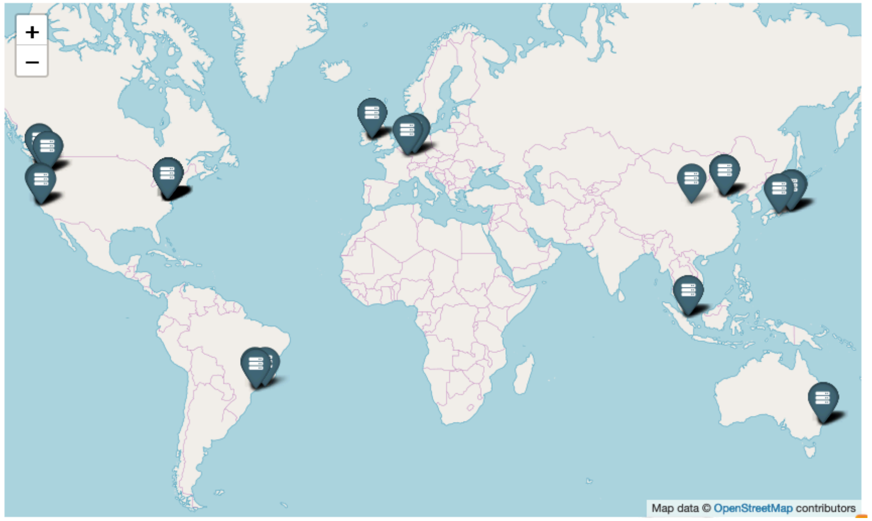
Source: https://www.networkatlas.org/

Networking Evolution



Source: telegeography.com

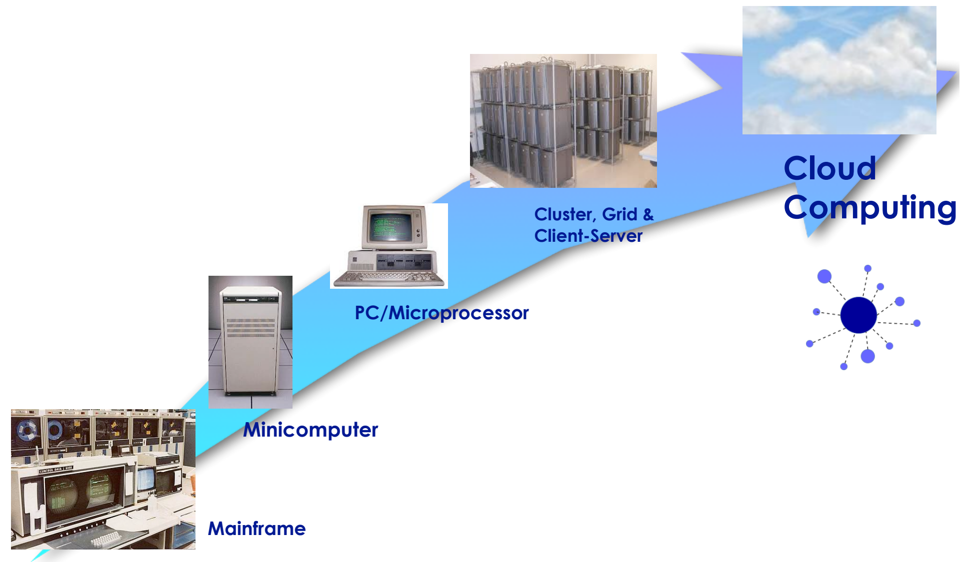
Proliferation of Data Centers

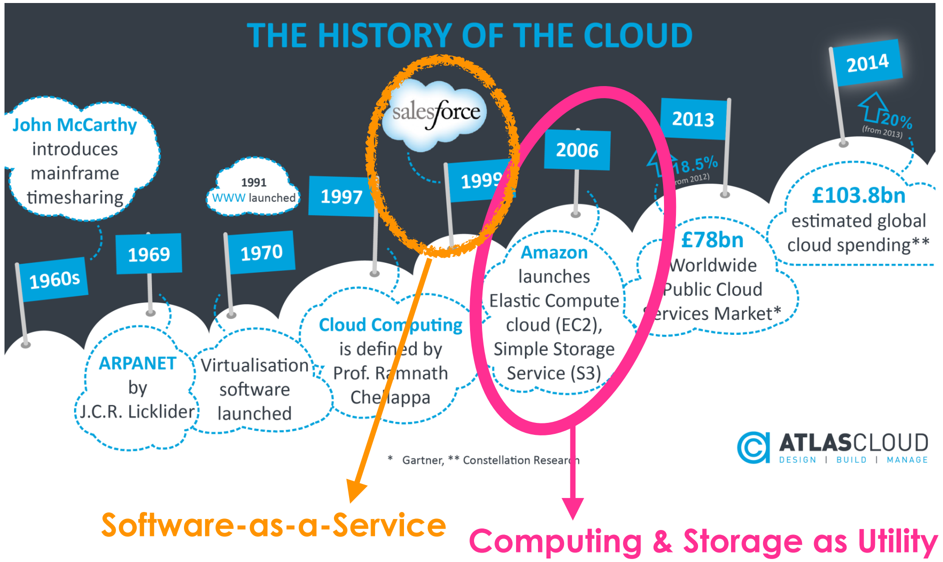


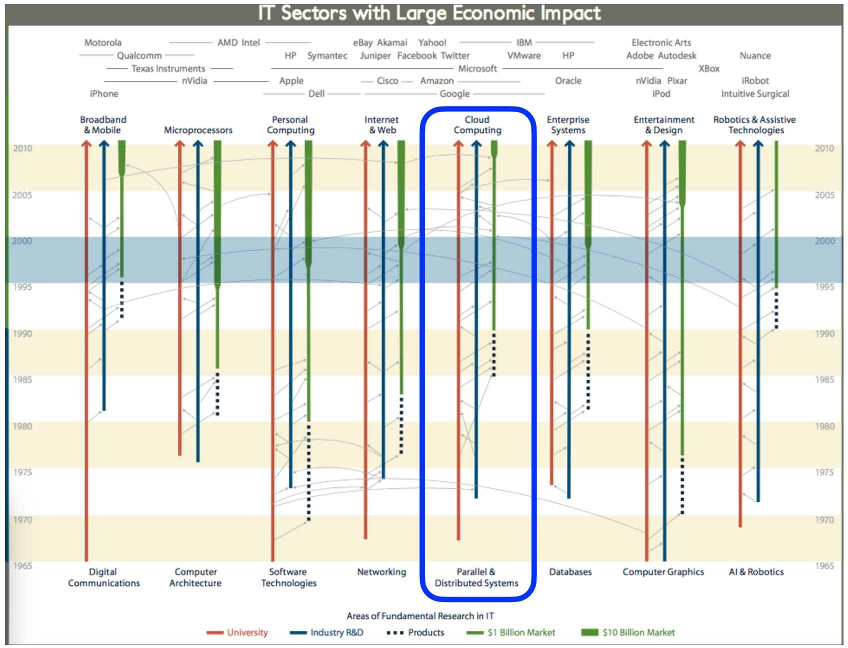
Amazon's Data Centers, Source: Wikileaks, 2015

Historical Overview

The emergence of Cloud Computing







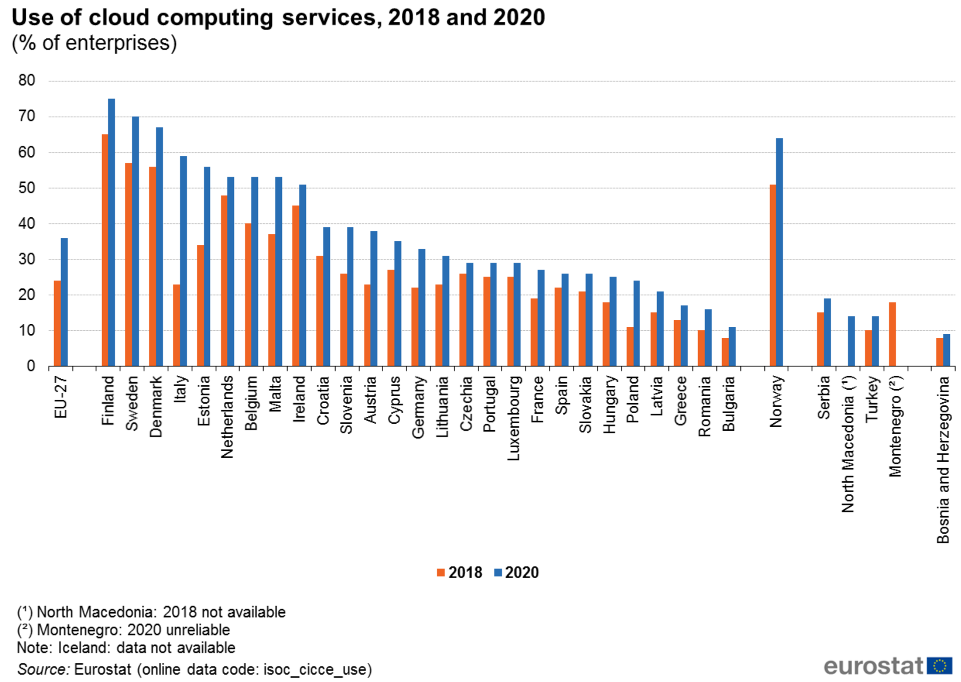
Source: **Continuing Innovation in Information Technology**
U.S. National Academies Press, 2012

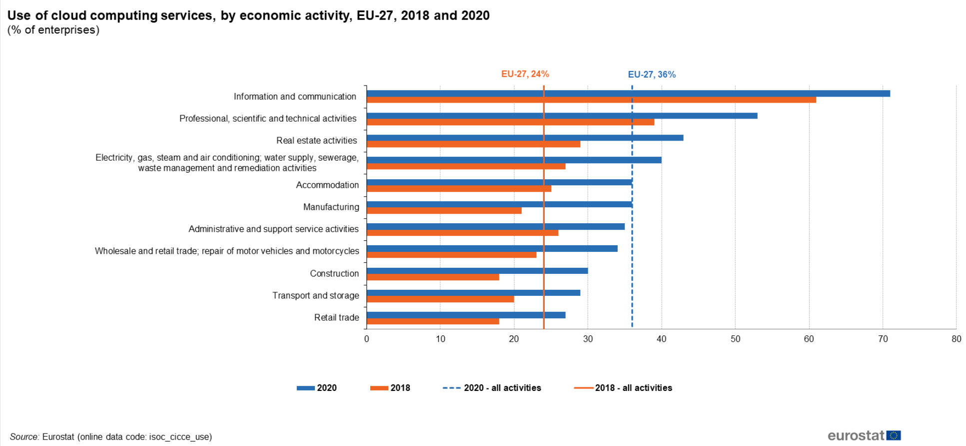
The Importance of the Cloud

There is no doubt that the cloud is one of the most significant platform shifts in the history of computing.

- Cloud impacted hundreds of billions of dollars of IT spend
- Still growing rapidly on a base of over $100B of annual public cloud spend (circa 2021).
- Growth driven by an incredibly powerful **value proposition**:
  - Infrastructure available immediately
  - At exactly the scale needed by the business
  - Driving efficiencies both in operations and economics.
  - Helping cultivating innovation as company resources freed up to focus on new products and growth.

Source: The Cost of Cloud, a Trillion Dollar Paradox, S. Wang & M. Casado, Andreessen Horowitz, 27/5/2021





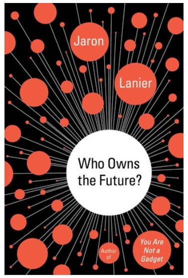
Location implies possession

Possession implies control

Control implies power

Source: Ed Felten, Princeton Univ.

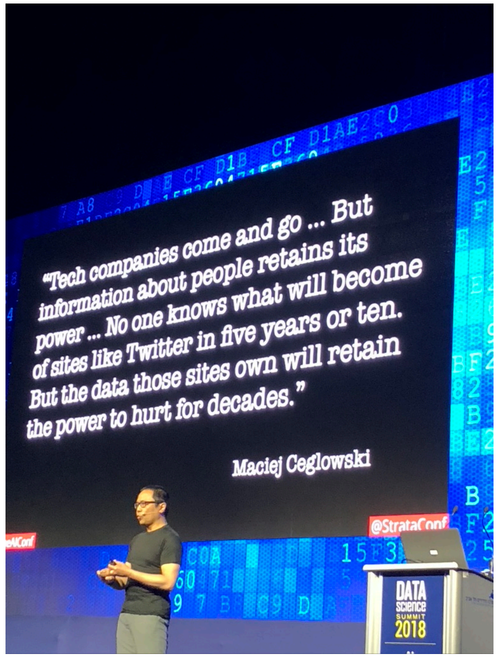
Control implies Power



- "Siren servers" accrue large amounts of data and monetize on their holding of the data and computation
  ["Who owns the future?"Jaron Lanier, 2013]
- ➡ Far reaching implications for the economy, society, science

The cloud is just another name for somebody else's computer

Graham Cluley



Source: Data Science Summit, Tel Aviv, 2018

Main Concepts

- Moore's Law
- Mainframes
- PCs and Client-Server
- Cluster Computing
- Web Computing
- Internet-scale Services
- Exponential Phenomena
- Network Effects
- Grid Computing
- Utility Computing
- Software-as-a-Service

University of Cyprus
Department of Computer Science

Main Concerns

- Privacy
- Security
- Cost and Performance
- Application Development and Evolution
- Portability and preservation of data
- Data Ownership
- Possession of Computation and its implications
- Ex post regulation
- Ex ante agreements

University of Cyprus
Department of Computer Science

M. D. Dikaiakos

Related Readings

University of Cyprus
Department of Computer Science

M. D. Dikaiakos

Lecture 1

- Can the Grid be considered as a GPT? Explain.
- Can the Cloud be considered to be a GPT? Explain.
- Explain what Network Effects are and give an example of a Cloud service that has benefited from Network Effects.
- Describe the core value proposition of the Cloud.
- Why is location important in modern computing? Name three key reasons.

University of Cyprus
Department of Computer Science

M. D. Dikaiakos