Amirkabir University of Technology
(Tehran Polytechnic)

Cloud Computing

Introduction

Seyyed Ahmad Javadi

sajavadi@aut.ac.ir

Spring 2024

Contact Details

- Office: CE department, 3rd floor
- Email: sajavadi@aut.ac.ir
- Home page: https://ce.aut.ac.ir/~sajavadi/

Course Introduction

- Saturday and Monday (13:30-14:45 pm)
  - Attend class on time
  - Class 001
  - No more than 3/16 absence is allowed

- Course web page
  - Check the webpage on regular basis
  - Everything will be posted on CW
  - Post All your Questions on CW Forums
    - Check forum history before posting any question

- Office hours and TA classes
  - TBD

Cell Phone and Laptop Policy

- Class use policy: Don’t!
- Cell phones should be off or silenced
- Texting is strictly prohibited in class

- Laptops and tablets may NOT be used in class: No email, browsing, Facebook, Twitter, Instagram during class lectures

- Violations may result penalties

Course Logistics

Section	Score	Considerations
Assignments	7	Four practical homework
Midterm exam	3	1402/01/26
Team project	3 + 1	In Kubernetes
Final exam	7	1402/3/27
Technical presentation	0.5 + 0.5	Topics are raised during the lectures
Total	20 + 2	Good luck ☺

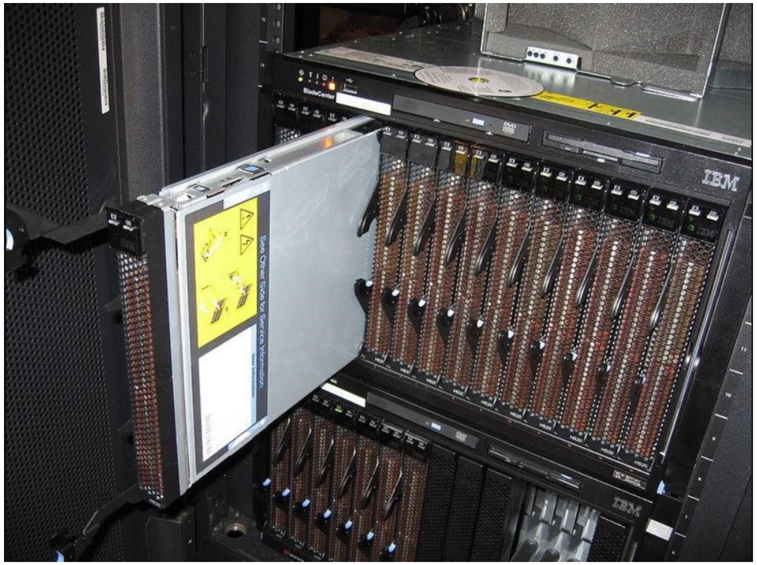
You are in a right place if you intend to do the programming assignments

Harsh penalty for plagiarism and cheating

What is a Server?

- Servers are computers that provide services to clients

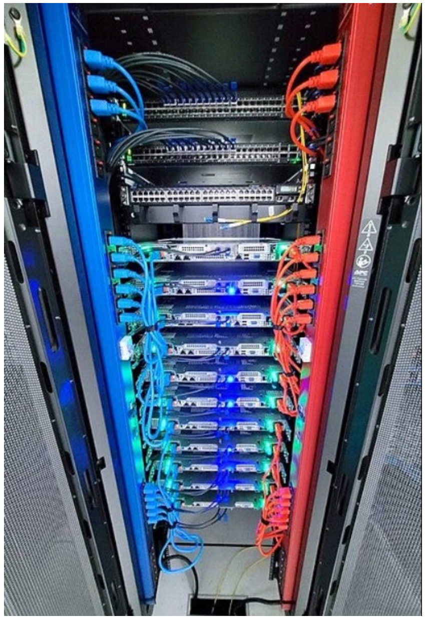
- Organizations typically require many physical servers to provide various services (Web, Email, Database, etc.)



Racks

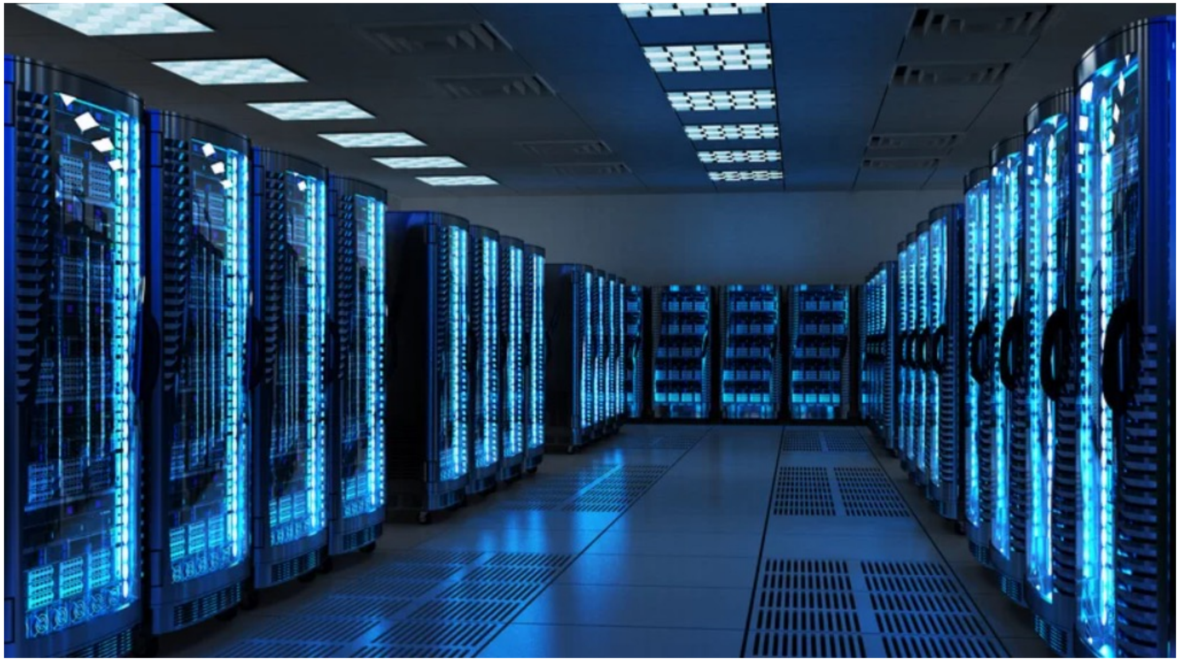
- Equipment (e.g., servers) are typically placed in **racks**

- They allow organizations to consolidate multiple servers in a single physical space, enhancing service capacity and management.



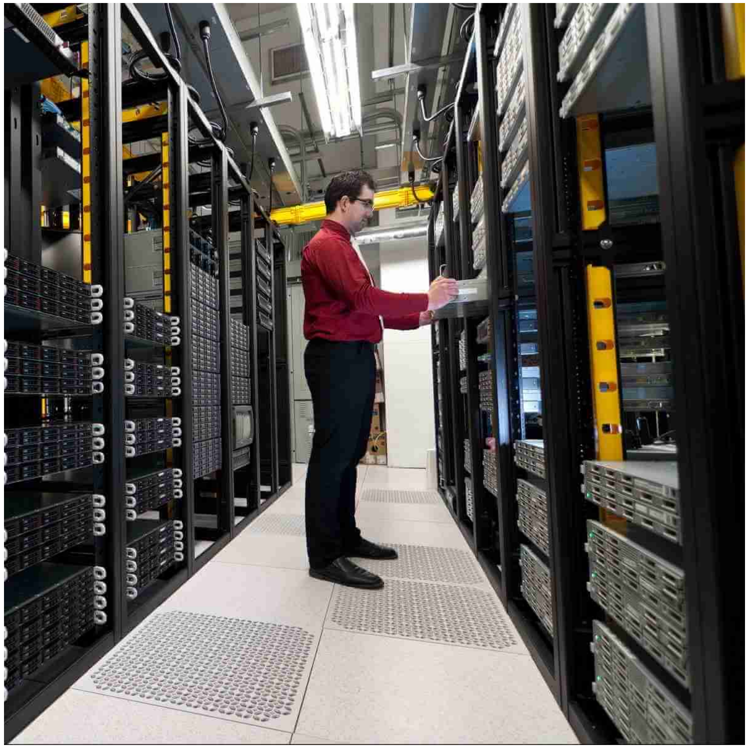
Data Center

- A data center is a facility used to house computer systems and associated components, such as networking and storage systems, cooling, uninterruptable power supply …

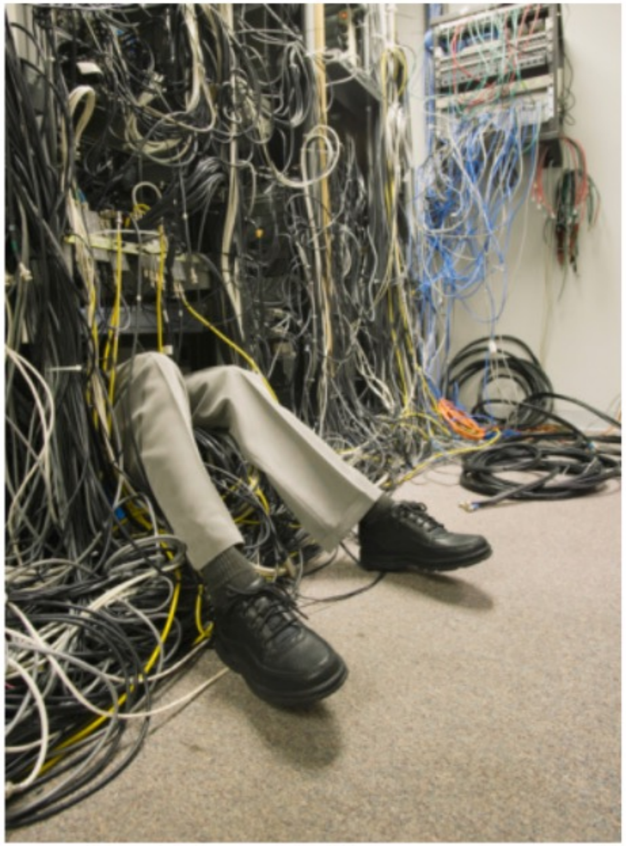


Challenges in Data Center Operations

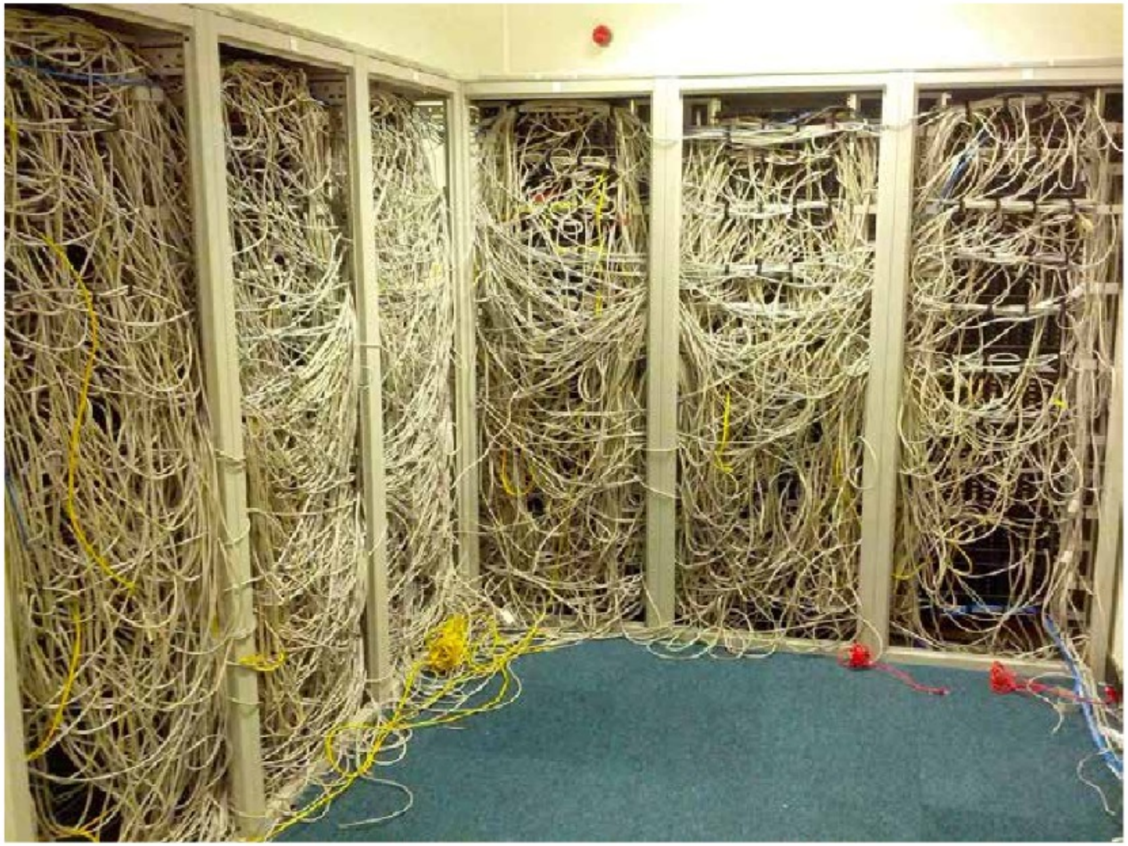
- Cooling data centers
- Servers are idle most of the time
- Managing scale and growth
- Networking at scale
- Security



Networking at Scale



[David Samuel Robbins, gettyimages.ch]



[@AlexCWheeler, Twitter]

Utilization in Data Centers

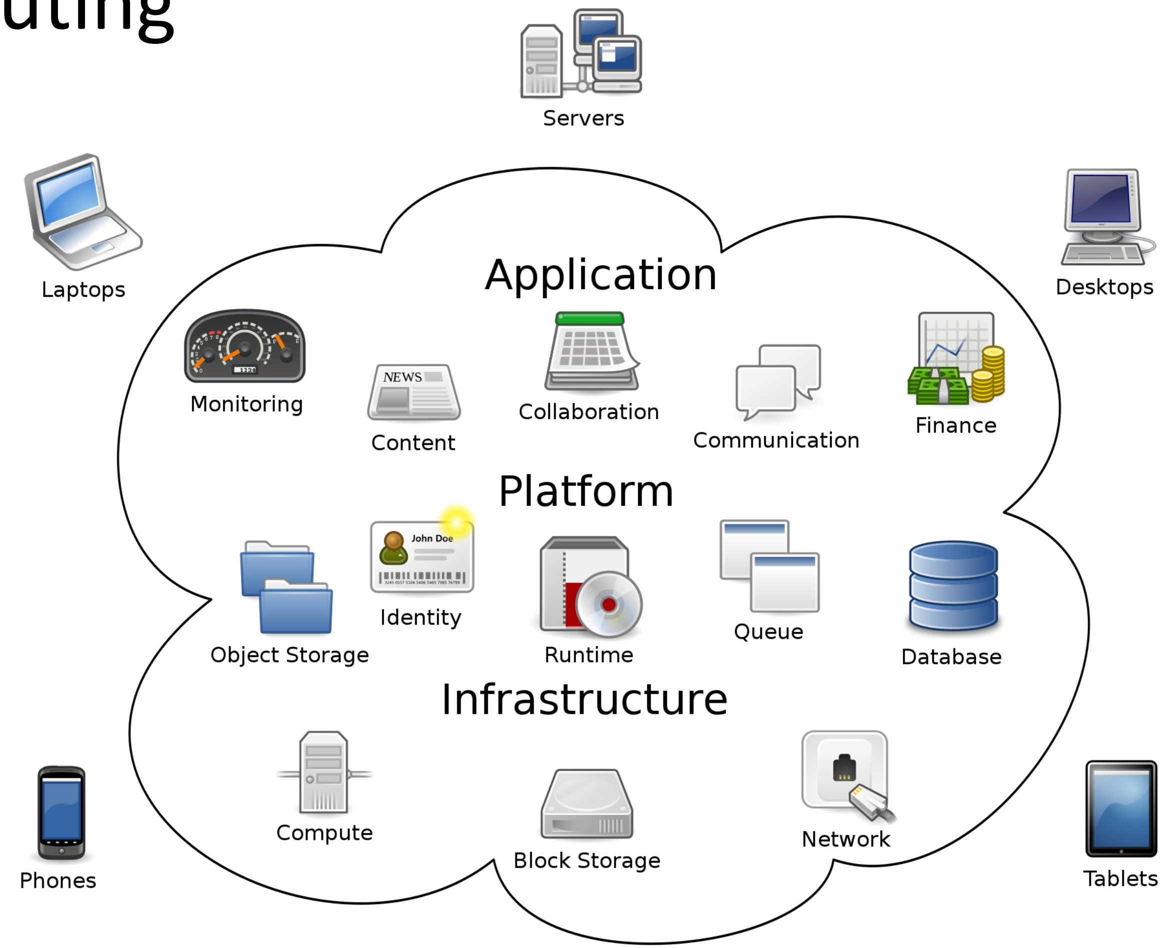
- Utilization of 10% to 30% is considered “good” in data centers
- Causes:
  - Uneven application fit:
    - Each server has CPU, memory, and disk: most applications exhaust one resource, stranding the others
  - Long provisioning timescales
  - Uncertainty in demand:
    - Demand for a new service can spike quickly
  - Risk management:
    - Not having spare servers to meet application demands lead to failure

Efficiency and Cost Optimization

- Maximize useful work per dollar; 59% of dollars are spent on servers with very low utilization (10%)
- Create a unified resource pool for services to adjust use dynamically.

One Solution to All These Challenges

- Cloud Computing



A Cloud is …

- A data center hardware and software that the vendors use to ***offer*** the computing resources and services

Cloud Computing at a Glance

- The term ***cloud*** often denotes the infrastructure as a “cloud”
  - Businesses and users can access applications as services from *anywhere in the world and on demand.*

The Vision of Cloud Computing

- Allowing anyone with a credit card to provision virtual hardware, runtime environments, and services.
  - These are used for as long as needed, with no up-front commitments required.

Practical Examples

- Large enterprises can offload some of their activities to cloud.

Read more

Practical Examples (cont.)

- Start-ups can afford to translate their ideas into business results **more quickly**, without excessive up-front costs.

Practical Examples (cont.)

- Developers can focus on the **business logic** rather than dealing with the **complexity of infrastructure management and scalability**.

Read more

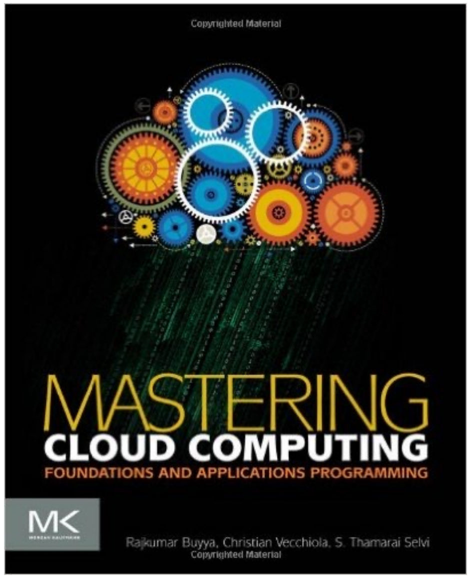
Practical Examples (cont.)

- End users can have their documents accessible from **everywhere and any device**.

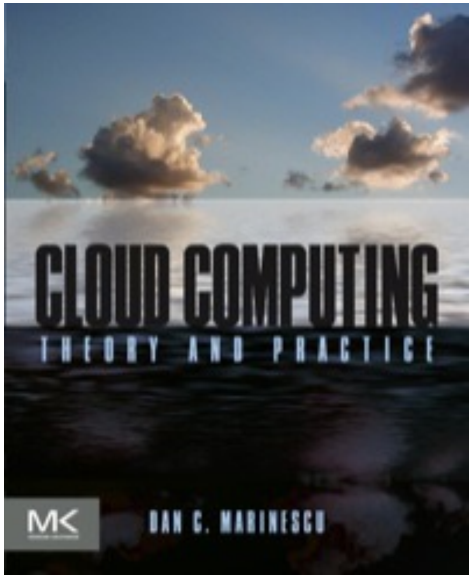
Syllabus

- Introduction to Cloud Computing
- Virtualization
- Containers
- Kubernetes
- Programming Models and MapReduce
- Hadoop Yarn and Apache Spark
- OpenStack
- Load balancing and auto-scaling

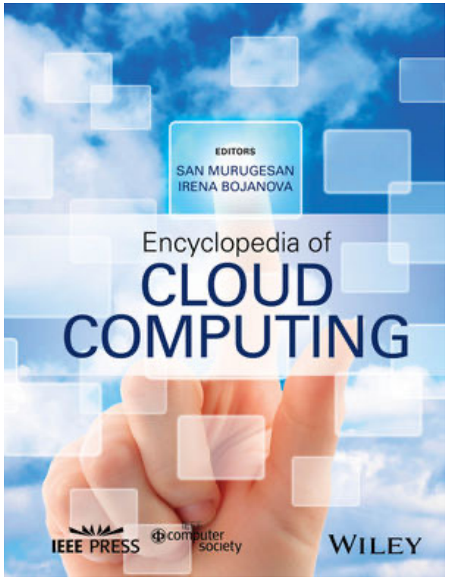
Resources


Copyrighted Material
MASTERING
CLOUD COMPUTING
FOUNDATIONS AND APPLICATIONS PROGRAMMING
MK
Rajkumar Buyya, Christian Vecchiola, S. Thamarai Selvi
Copyrighted Material

“**Mastering Cloud Computing: Foundations and Applications Programming**”, Buyya et. al.



“Cloud Computing, Theory and Practice” Marinescu et. al.



”Encyclopedia of Cloud Computing**”,** Murugesan et. al.