الجامعة الدولية للرباط
ⵜⴰⵙⴷⴰⵡⵉⵜ ⵜⴰⵅⵔⵕⵕⵉⵜ ⵏ ⵔⵕⴱⴰⵟ
Université Internationale de Rabat
L'UNIVERSITÉ INNOVANTE

Ecole Supérieure
d'Informatique et du Numérique
COLLEGE OF ENGINEERING & ARCHITECTURE

# Cloud Foundations

## Chapter 1: Introduction


**PR. OUASSIM KARRAKCHOU**

**UNIVERSITE INTERNATIONALE DE RABAT**


Academic year: 2024-2025

# The Motivations for Cloud

## Cloud is everywhere

Can you give some examples where the Cloud is used?

# The Motivations for Cloud

## Cloud is everywhere

- A startup leases cloud facilities for its website; the company can pay for additional facilities as web traffic grows.
- An enterprise company leases facilities and software for business functions, such as payroll, accounting, and billing.
- A seasonal company leases computing facilities during four peak months each year; the company doesn't pay for facilities at other times.
- A retail company leases facilities at the end of each fiscal year to run data analytics software that analyzes sales for the year.

Based on these examples, what can you say about the Cloud?

# The Motivations for Cloud

## Migration to the Cloud

- Most enterprises — not just high-tech firms and social media companies — are moving to the cloud.
- In the early 2000s, business functions, such as payroll, accounting, billing, and supply chain management were implemented with local facilities operated by an organization's IT staff.
- Now, such functions are being migrated to cloud computing.
- Enterprises that do decide to retain some local computing are shifting from a paradigm of having services run on individual computers located in various departments throughout the organization to a paradigm where the facilities are consolidated into a local cloud data center.

# The Motivations for Cloud

## Cloud is everywhere

- An individual uses a smartphone to check Internet of Things (IoT) devices in their residence.
- Students working on a project use a browser to edit a shared document.
- A patient wears a medical device that periodically uploads readings for analysis; their doctor is alerted if a medical problem is detected.
- A teenager logs into a social media site and uploads photos.
- An individual uses a streaming service to watch a movie; a copy of the movie is kept in a facility near the family's residence.
- The recipient of a package uses a tracking number to learn about the current location of the package and the expected delivery time.

Based on these examples, what can you say about the Cloud?

# The Motivations for Cloud

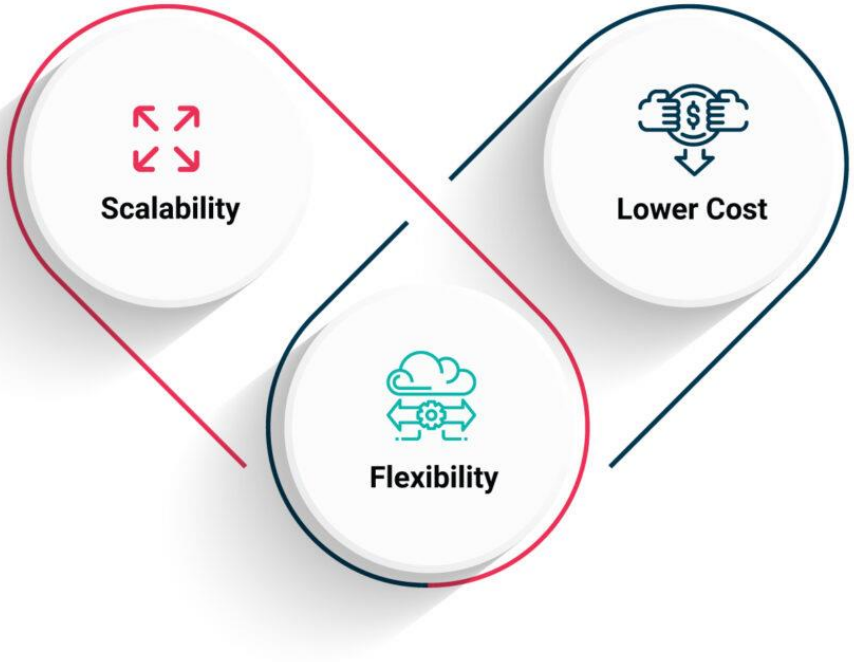
## Cloud Services

- In addition to hosting enterprise business functions, the Cloud also provides different services in the form of:
  - Communication/Networking
  - Processing/Computing
  - Storage
- These can be used to create complex consumer applications

# The Motivations for Cloud

## A facility for flexible computing

- Many of the examples illustrate a key aspect of cloud computing: flexibility to accommodate both incremental growth and cyclic demand. A cloud offers flexible computing facilities (servers and software), storage facilities, and communication facilities (Internet connections).



# The Motivations for Cloud

## Incremental Growth

- The startup scenario shows why incremental growth is important. A small startup can begin by leasing minimal cloud facilities (e.g., enough to support a basic web site), and then increase its lease as the business grows.

- Similarly, the startup can begin by leasing minimal software (e.g., basic web and payment processing software), and then add database and accounting software later.

- The cloud provider will be able to satisfy computing needs even if the startup grows into a substantial enterprise business.

# The Motivations for Cloud

## Cyclic Demand

- Even if a company does not engage in seasonal business, demand for computing changes throughout the year:
  - Reports may be generated at the end of each month or each quarter as well as at the end of the year.
  - Sales activity and order processing often spike at the end of the month as sales staff work to meet their monthly quotas.
- Cloud allows companies to lease additional facilities to accommodate such demand.

# The Motivations for Cloud

## Pay for what you use

- Although the ability to lease resources as needed is attractive, the most significant aspect arises from the pricing model: a cloud provider only charges the customer for the facilities they actually use.

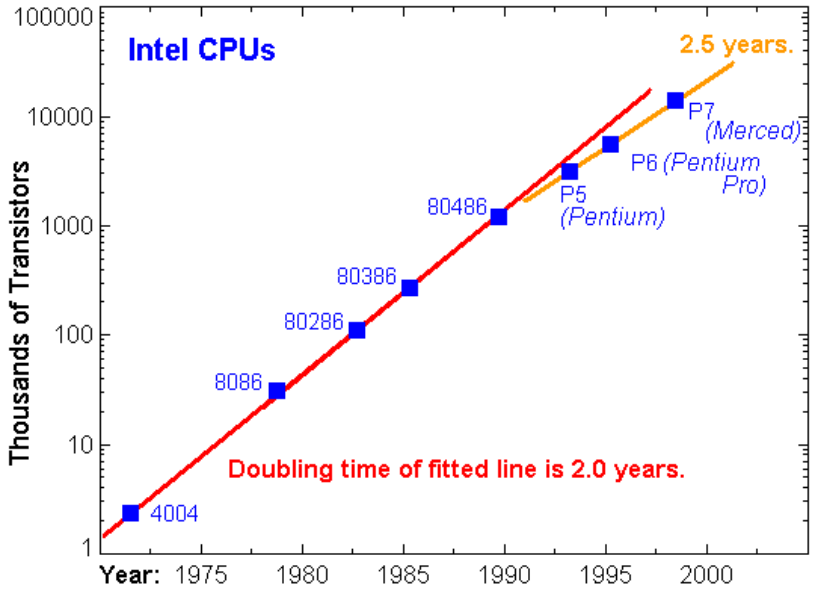
*Cloud computing allows each customer to increase or decrease their use of cloud facilities at any time; a customer only pays for the facilities they use.*

How did we get to the Cloud?

# The History of Cloud

## Moore Law

- Throughout the 1980s and early 1990s, chip manufacturers produced a series of processors that had more functionality and higher speed than the previous model. As a consequence, individual computers became more powerful each year while costs remained approximately the same.



# The History of Cloud

## The rise of individual servers

- At any time, if the processing power of a computer became insufficient for the workload, a computer could easily be upgraded to a newer, more powerful model.

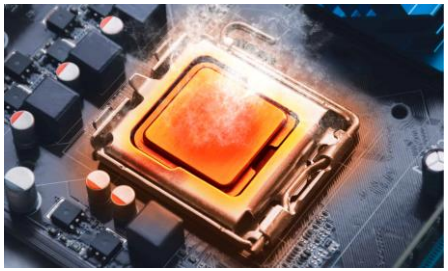
- In particular, to offer internal and external services, such as the World Wide Web, an organization could run the software on a powerful computer known as a server. When demand for a particular service became high, the organization could replace the server being used with a model that had a faster processor and more memory.

# The History of Cloud

## The power wall

- By the late 1990s, the chip industry faced a serious limitation: more transistors were squeezed together on a chip each year. When billions of transistors are squeezed together in a small space, the temperature can climb high. More important, the amount of power consumed, and therefore the temperature of the chip, is proportional to the square of the clock speed.

- Manufacturers eventually reached a critical point, and processor speeds could not be increased beyond a few Gigahertz without generating so much heat that the chip would burn out.



How can additional computational power be achieved without increasing the speed of a processor?

# The History of Cloud

## Multiple Cores

- The answer to the power wall issue lies in parallelism,
- We thus use multiple processors that each operate at a speed below the power wall instead of one processor that operates at a super high speed.
- To handle the situation, chip manufacturers devised chips that contain multiple copies of a processor, known as cores. Each core consists of a complete processor that operates at a safe speed, and software must devise a way to use multiple cores to perform computation.
- Multicore processors form one of the fundamental building blocks for cloud computing. Unlike the processors used in consumer products, however, the multicore processors used in cloud systems have many cores (e.g., 64 or 128).

# The History of Cloud

## Beyond Multiple Cores

- Although they offer increased processing power on a chip, multiple cores do not solve the problem of arbitrary scale. The cores on a chip all share underlying memory and I/O access. Unfortunately, as the number of cores increases, I/O and memory accesses become a bottleneck.

How can more powerful computer be contructed?

- The science research community was among the first groups to explore a design that provided the basis for cloud.

- As scientific instruments, such as colliders and space telescopes, moved to digital technologies, the amount of data grew beyond the capabilities of even the most powerful supercomputers. Furthermore, a supercomputer was an extremely expensive machine; few universities and laboratories could afford to purchase multiple supercomputers.

# The History of Cloud

## From Multiple Cores to Multiple Machines

- While supercomputers are expensive, personal computers had become commodity items as reflected by their low price. Despite the drop in price, personal computers had also become more powerful.

- Scientists wondered if instead of using expensive supercomputers, a new form of supercomputing could be achieved by interconnecting a large set of inexpensive personal computers.

- The resulting configuration, which became known as a cluster architecture, has a key advantage: processing power can be increased incrementally by adding additional inexpensive commodity computers.

# The History of Cloud

## Software Challenges of using Multiple Machines

- Using multiple computers for scientific computations poses a software challenge: a calculation must be divided into pieces so that each piece can be handled by one of the smaller computers in the cluster.

- Several computing paradigms were thus devised to handle the issue of distributing computation on a cluster. One important example is OpenMPI that uses message passing to synchronize distributed calculations.

# The History of Cloud

## From Clusters to Distributed Web Sites

- As the World Wide Web grew in popularity in the 1990s, the traffic to each web site increased. As in the science community, the limitation on the speed of a given processor presented a challenge to the staff responsible for running web sites, and they also considered how to use multiple personal computers to solve the problem.

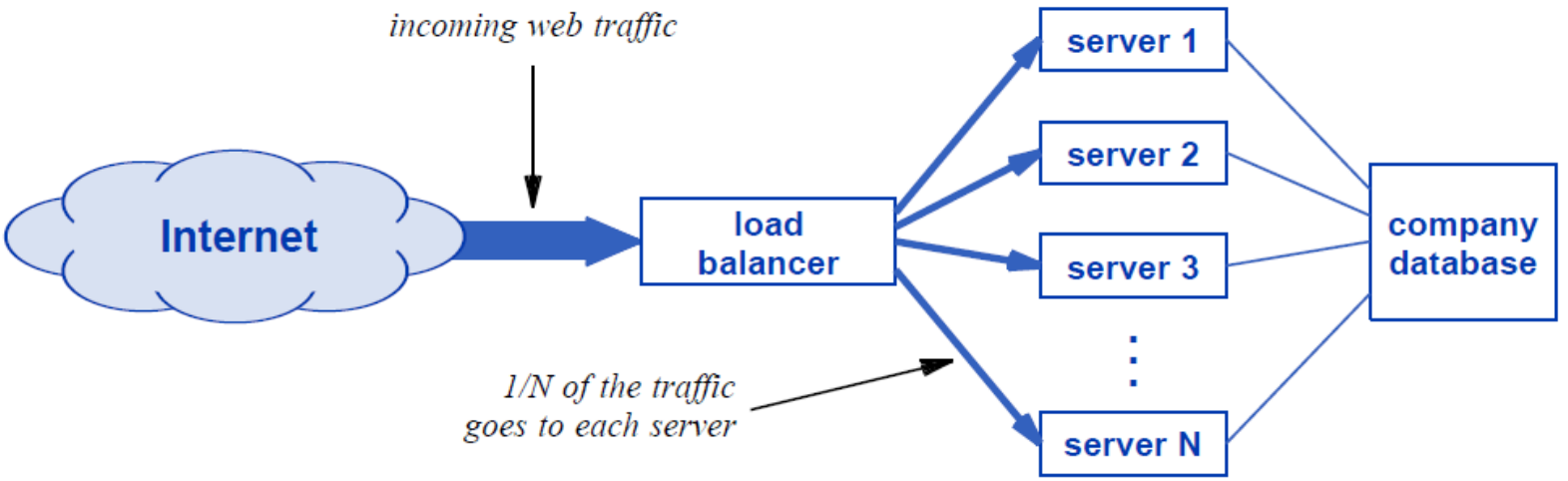
- Web sites and scientific computing systems differ in a fundamental way. Supercomputer clusters intended for scientific calculations are designed so that small computers can work together on one computation at a time. In contrast, a web site must be designed to process many independent requests simultaneously.


How can a web site scale to accommodate thousands of users?

# The History of Cloud

## Load Balancing

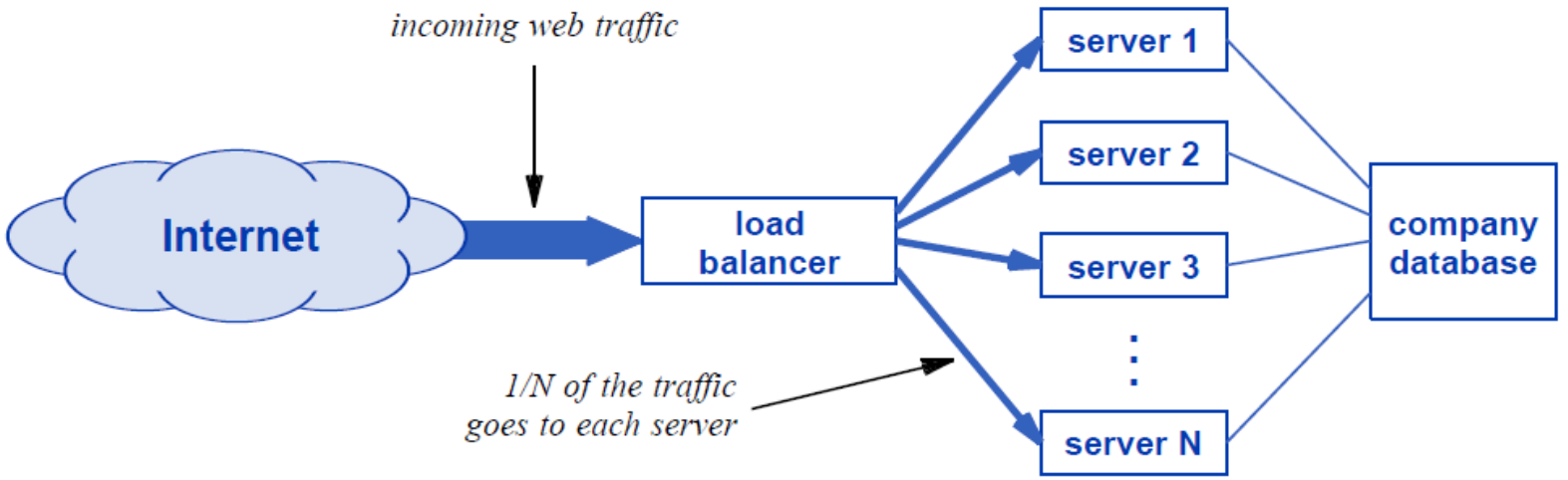
- Part of the answer to scaling websites came from a technology that has become a fundamental component in cloud computing: a load balancer.
- Typically implemented as a special-purpose hardware device, a load balancer divides incoming traffic among a set of computers that run servers.



# The History of Cloud

## Load Balancing

- Load balancing technology ensures that all communication from a given customer goes to the same server. The scheme has a key advantage: successive requests from a customer go back to the server that handled earlier requests, making it possible for the server to retain information and use it for a later request.

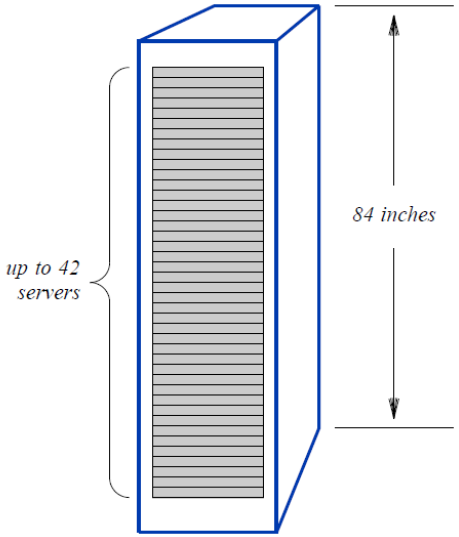


# The History of Cloud

## Racks of servers

- As demand increased for facilities composed of many smaller computers, computer vendors responded by changing the shape of computers to make it easier to store many computers in a small space.

- Instead of large enclosures that had significant amounts of empty space inside, designers focused on finding ways to reduce the size. Furthermore, they redesigned the enclosures to fit into tall metal equipment cabinets called racks:



# The History of Cloud

## Centralized Data Centers

- The availability of low-cost servers and the ability to collect multiple servers into a rack may seem insignificant. From the point of view of IT management, however, collecting servers into a small place has an important advantage: lower cost. There are two aspects:
  - Operating expenses (opex): lower recurring cost
  - Capital expenses (capex): lower equipment cost
- These economic considerations led to the creation of centralized data centers.

What economic advantages of centralized data centers can you think of in terms of opex and capex?

# The History of Cloud

## Opex Advantages of Centralized Data Centers

- Cheap servers that take small spaces mean that every group or department of an organization can have their own IT infrastructure.
- However, each group then needs to hire a dedicated IT staff to manage these servers, which actually caused much higher costs than anticipated:
  - Cheap computers have turned into a major expense.
- The availability of high-speed computer networks allows an organization to optimize costs by consolidating server equipment into a single physical location.
- Instead of locating a server in each department, the organization places all servers in racks in a central facility, and hires a small, centralized staff to maintain the servers. Employees in departments can access the servers over a network. Such a centralized facility has become known as a data center

# The History of Cloud

## Capex Advantages of Centralized Data Centers

- The data center approach has an advantage of reducing overall equipment cost.
- If an organization consolidates servers into a data center, the organization can choose a uniform configuration for all servers.
- Furthermore, the organization can upgrade many servers at once (e.g., upgrade one-third of all servers every year). Consequently, when an upgrade occurs, the organization will purchase dozens or hundreds of servers at the same time, making it possible to negotiate a quantity discount.

# The History of Cloud

## Summary

- *For decades, the low cost of computers encouraged decentralization.*
- *The power wall and cost of IT staffing favor a return to a centralized model that consolidates computing facilities into data centers.*
- *In the next chapter, we will see how public clouds like AWS and Azure further increase the centralization by creating data centers with servers for multiple organizations.*