

# High Performance and Distributed Computing for Big Data

## Unit 3: Cloud Systems and Big Data Management

### Session 1 & 2 - Introduction to Cloud Computing + Github Pages

---

Francesc Solsona Tehas [francesc.solsona@udl.cat](mailto:francesc.solsona@udl.cat)

Universitat Rovira i Virgili and Universitat de Lleida

# What is this course about?

## High Performance and Distributed Computing for Big Data

### High Performance

Refers to using supercomputers and parallel processing techniques to solve complex computational problems.

### Distributed Computing

This concept means using multiple computers that are interconnected and communicate with each other to achieve a common goal.

### Big Data

This term leads to using large and complex data sets that are difficult to process using traditional computers and applications.

## Pre-Course Survey: Who are you?

- How familiar are you with cloud computing technology?
- What is your background, and why did you enrol in this Master's degree?
- What does cloud computing mean to you?

[Pre-Course Survey: Who are you?](#)

## **Introduction**

---

# What is Cloud Computing?

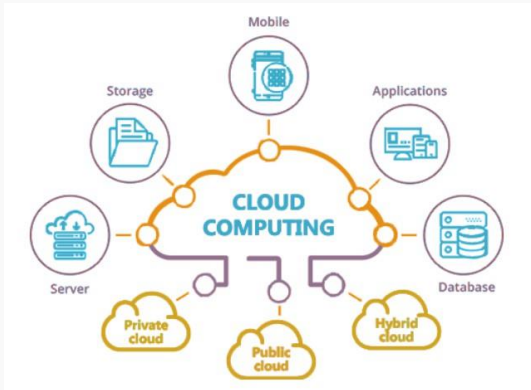


Figure 1: Overview of Cloud Computing

## Definition

Cloud computing refers to the **on-demand delivery of computing services**, including servers, *storage*, *databases*, *software*, *analytics*, and more, over the internet (*the cloud*) with **pay-per-use** pricing (Mell, Grance, et al. 2011).

## Vision

In cloud computing, users **can access these services remotely from any location with an internet connection**. Rather than owning and maintaining their own computing infrastructure or data centers, users and companies can rent access to anything from applications to storage from a cloud service provider.

# What is the idea behind Cloud Computing?



## Cloud computing

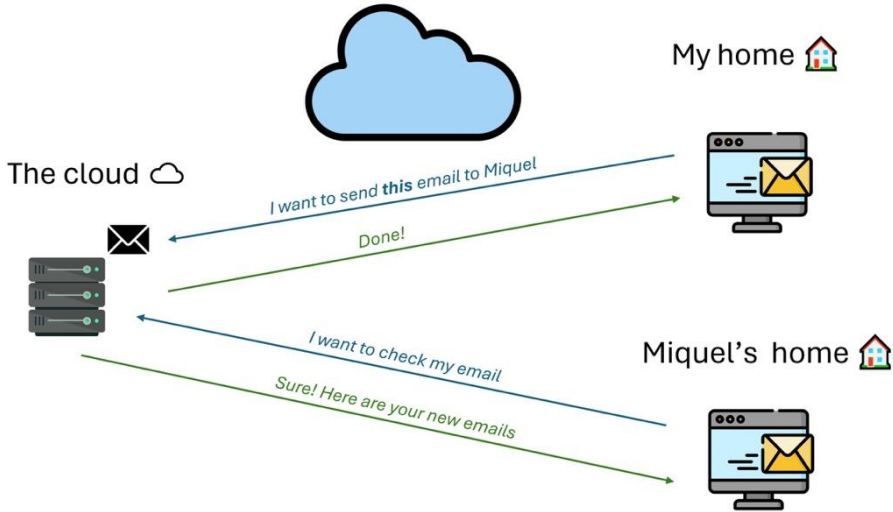
Enables companies/users to consume a compute resource, such as a virtual machine (VM), storage, or an application, as a utility



rather than having to build and maintain computing infrastructures in-house.



**Figure 2:** What happens when I watch Netflix?



**Figure 3:** What happens when I send an email?

## Which is the cloud computing model?

**Cloud computing** is a model for *enabling convenient, on-demand network access to a shared pool of configurable computing resources* (e.g., networks, servers, storage, applications and services) that can be **rapidly provisioned and released** with minimal management effort or service provider interaction.

Properties:

- Flexible and scalable.
- Pay what you use.
- Can change and is easy to upgrade.
- No need to buy hardware.

**Cloud computing is elastic, allowing organizations to scale resources based on demand efficiently.**

## Which is the traditional computing model?

In **traditional computing**, the user must buy the *hardware* and *software*, and then install them on the computer. **The user is responsible for the maintenance and security of the software and hardware.**

- Requires space for the servers.
- Needs a team of experts for install, configure and maintain.
- Requires physical security policies.
- Long hardware procurement process (purchase, install, configure and maintain).
- Requires prediction of the theoretical maximum load peak of the system.

If the company underestimates the peak, the system will be slow, and the users will be unhappy. If the company overestimates the peak, it will have a lot of unused resources and waste money.

**What happens if you need to change or upgrade the hardware?**

### What is the difference between living in a hotel and owning a house?

- *In a hotel, you **pay for what you use**, and you do not need to **worry about the maintenance** of the building.*
- *In a house, **you have to buy the house**, and then **you have to maintain it**.*

### What is the difference between cloud and traditional computing?

**In cloud computing, you pay for what you use, and you do not need to worry about the maintenance of the hardware.**

### Which are the steps to serve a personal website in traditional computing?

*You have to buy a server, install the operating system, install the web server, configure the web server, and then upload the website.*

### What is the advantage of using cloud computing?

*The user does not need to buy the server, the user can rent the server and the web server (cloud provider responsibility), and then upload the website (user responsibility).*

## Which are the main cloud computing services models?

- **Infrastructure as a Service (IaaS)**: Provides virtualized computing resources over the internet.
- **Platform as a Service (PaaS)**: Provides a platform allowing customers to develop, run, and manage applications without the complexity of building and maintaining the infrastructure.
- **Software as a Service (SaaS)**: Provides software applications over the internet on a subscription basis.
- **On-Premises**: refers to software and hardware that is physically located within your own building or data center, rather than being hosted by a provider like AWS.

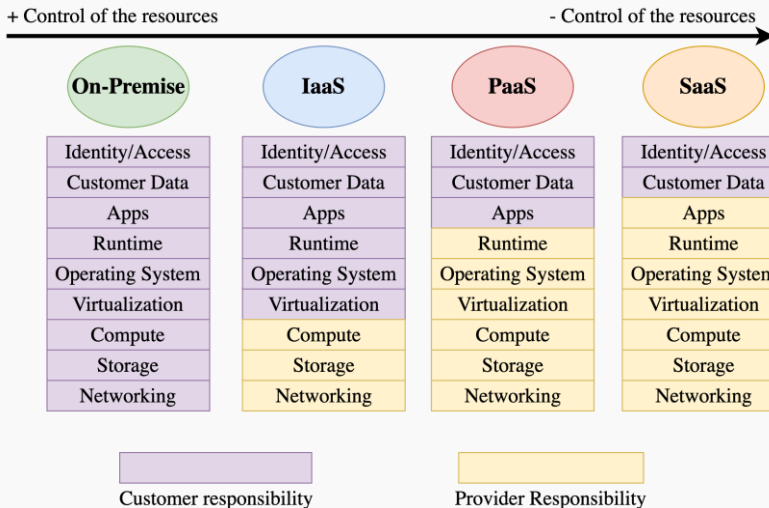
+ Control of the resources

- Control of the resources



Figure 4: Cloud Services Models

# Shared Responsibility Model



**The shared responsibility model outlines where a cloud provider's role and responsibility ends and the customer's begins.**

### The shared responsibility model is like ordering a pizza.

- **SaaS:** Take the pizza in a restaurant. The pizza provider is responsible for everything, you just need to eat the pizza.
- **PaaS:** Pizza delivery. The pizza provider is responsible for the baking the pizza and ensuring it arrives to your home. You are responsible for preparing the table and the drinks.
- **IaaS:** Take the pizza home and bake it yourself. You are responsible for the dining table, the electricity, and the oven. The pizza provider is responsible for the pizza and the box.
- **On-Premises:** You make the pizza at home from scratch. You are responsible for everything, from the ingredients to the baking and for preparing the table and the drinks too.

A **public cloud** provides computing services over the public Internet, offered by third-party providers and available to anyone on a pay-per-use basis.

- Type of cloud computing with resources available to the general public over the internet.
- Providers own and manage infrastructure on a pay-per-use basis.
- Highly scalable and flexible, with wide range of services available.
- Popular for its affordability and scalability.

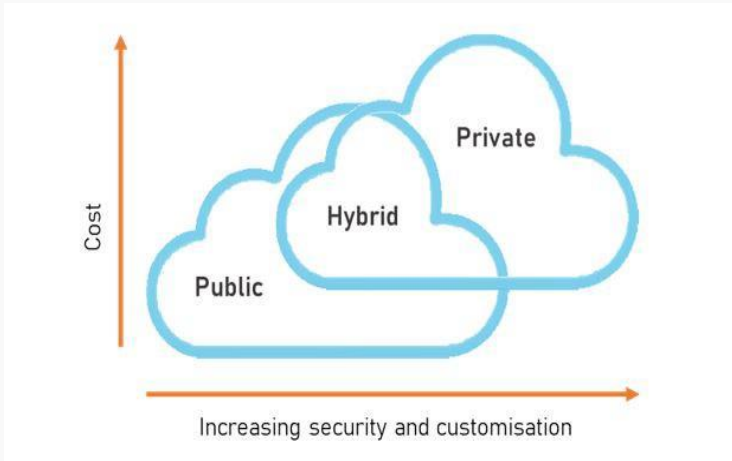
**Private cloud** refers to a *model of cloud computing where IT services are provisioned over private IT infrastructure for the dedicated use of a single organization.*

- Type of cloud computing used exclusively by a single organization.
- Provides greater control and customization over infrastructure.
- Can be hosted on-premises or by third-party provider.
- Offers enhanced security and is popular in finance, government, and healthcare.
- More expensive than public cloud but provides greater control and flexibility.

**Hybrid cloud** is a *computing environment that combines a public cloud and a private cloud by allowing data and applications to be shared between them.*

- Combines public and private cloud resources.
- Provides cost-effectiveness and scalability of public cloud, while maintaining control over sensitive data in private cloud.
- Offers greater flexibility and agility, and can optimize computing resources and costs.
- Complex to set up and manage, requires careful planning and coordination.
- Balances benefits of public and private cloud, but requires expertise to manage.

# Public vs Private vs Hybrid Cloud Models



**Figure 1:** How cost and privacy grow with different cloud paradigms

- **The global cloud computing market is expected to reach \$1.35 trillion by 2027.**
- **It grows at a CAGR of 17.5% from 2020 to 2025.**

## Cloud services everywhere

Cloud services go beyond just hosting websites. Examples include **GitHub**, **Netflix**, **iCloud**, and many more.

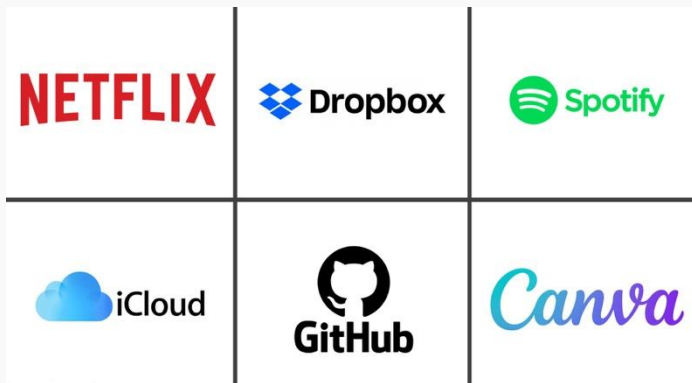


Figure 3: Popular cloud services

## Cloud services everywhere

Beyond cloud services, we have the **major cloud providers**: Azure, Google Cloud Platform (GCP), and Amazon Web Services (AWS).



**Figure 4:** Major cloud providers

## **Advantages of Cloud Computing**

---

## Capital expenses to operational expenses

The capital expenses (*CapEx*) of building and maintaining data centers is converted to operational expenses (*OpEx*) of renting cloud computing resources.

⇒ This shift reduces the need for large investments in hardware, and provides a more predictable and manageable cost structure.

### Example: Hotel Analogy

It is like renting a room in a hotel instead of buying a house. You do not need to invest in the building, you just pay for the room.

### Be aware!

In the long term, it is more expensive to rent a room than to buy a house.

**You only pay for what you use.**

Cloud computing providers benefit from economies of scale, which allows them to offer lower prices than an individual company can achieve. This is because they can buy and manage resources at a larger scale.

### **Example: Hotel Analogy**

The hotel can buy food and cleaning products in bulk, and obtain better prices than an individual can achieve.

**Aggregated demand from multiple companies can create a more efficient use of resources.**

Cloud computing allows companies to scale up or down as needed. They can stop guessing about their infrastructure needs and scale up or down as the business changes.

### Example: Hotel Analogy

If you want a suite, you can rent a suite. If you want a single room, you can rent a single room. Each time you can rent the room that you need. You need more rooms? You can rent more rooms.

**You can scale up or down as your business needed.**

Cloud computing services are provided on demand, so vast amounts of computing resources can be provisioned in minutes. This gives businesses a lot of flexibility and taking the pressure off capacity planning.

### Example: Hotel Analogy

You can rent a room in minutes. How long does it take to buy a house?

**You can get resources in minutes.**

## Case Study: Dropbox

**Dropbox** is a file hosting service that offers cloud storage, file synchronization, personal cloud, and client software. It allows users to create a particular folder on their computers, which Dropbox then synchronizes so that it appears to be the same folder (with the same contents) regardless of which device is used to view it.

### Dropbox using Public Cloud

In its early stages, Dropbox leveraged AWS's S3 service for file storage. This decision offered Dropbox the ability to:

- Rapidly scale storage capacity as user demand increased.
- Achieve swift production deployment, enhancing their market responsiveness.

### Disadvantage: Elevated Cost

## Risks and Challenges of Cloud Computing

---

Cloud computing providers invest heavily in security, but breaches can and do happen. The cloud is not immune to security threats, and companies must carefully manage their security measures.

- Data breaches
- Data compliance
- Misconfiguration
- Insider threats

**You need to be aware of the security measures of your cloud provider and yours, it is a shared responsibility.**

## Real Cases of Security Breaches

- WannaCry ransomware attack in 2017 affected more than 200,000 computers in 150 countries, including the UK's National Health Service (NHS). The attack encrypted data on computers and demanded a ransom to unlock it.
- In 2021, Universitat Autònoma de Barcelona (UAB) suffered a data breach that exposed the personal information of students and staff. The breach occurred when a hacker gained unauthorized access to the university's computer systems and stole sensitive data.
- In 2023, the Clinic Barcelona Hospital suffered a data breach that exposed the personal information of million of patients. The breach occurred when a subcontractor accidentally uploaded patient data to the internet without proper security measures.

## How to protect your data in the cloud?

- **Data encryption:** Encrypt data at rest and in transit to protect it from unauthorized access.
- **Access control:** Implement strong access controls to ensure that only authorized users can access data.
- **Data backup:** Regularly back up data to ensure that it can be recovered in the event of data loss or corruption.
- **Security monitoring:** Monitor cloud environments for security threats and respond to incidents in a timely manner.
- **Compliance:** Ensure that cloud environments comply with relevant security standards and regulations.

Regulatory compliance is a critical consideration for companies that store and process sensitive data in the cloud. Companies need to ensure that their cloud environments comply with relevant security standards and regulations, such as GDPR, HIPAA, and PCI DSS.

**You need to be aware of the data compliance regulations that apply to your business.**

Cloud providers use **proprietary technology**, which can make it difficult to migrate data and applications to other cloud providers. This can result in vendor **dependence**, where companies are dependent on a single cloud provider for their computing needs.

- **Limited flexibility and choice:** Customers may be limited in their ability to choose different vendors or solutions that better meet their evolving needs or provide better value.
- **Higher switching costs:** Customers may face high costs and complexity when migrating their data, applications, or services to a different cloud vendor or on-premises environment.
- **Data interoperability:** Customers may face challenges in ensuring data interoperability and portability across different cloud platforms or environments, which can hinder their ability to integrate or share data with other applications or systems.

## Other challenges

- **Performance:** Can be affected by factors such as network latency, bandwidth, and the location of data centers.
- **Cost management:** Cloud costs can spiral out of control if not properly managed, leading to unexpected expenses.
- **Service level agreements (SLAs):** Cloud providers may not meet their SLAs, leading to downtime or performance issues.
- **Integrity:** Ensuring the integrity of data and applications in the cloud, and protecting against unauthorized changes or tampering.
- **Real-time data processing:** Cloud environments may not be optimized for real-time data processing and analysis, which can impact the performance of applications that require low latency and high throughput.

**Cloud computing is not without its risks and challenges, and companies need to carefully manage these to ensure a successful cloud deployment.**

I want you to create a **GitHub** account. Next week we are going to create a personal website using **GitHub Pages**. Take a look at **the lab descripton** if you want.

## Case Study: Is Cloud Computing Right for Your Company? (I)

Consider a startup company **X**. The company is in a competitive market and needs to reduce costs and time to market. **The company is considering whether to adopt cloud computing or traditional infrastructure?**

*The company can leverage cloud computing to reduce costs and time to market, while also benefiting from the flexibility and scalability of cloud resources. For instance, **Amazon Web Services (AWS)** offers a pay-as-you-go model that allows startups to only pay for the services they use.*

## Case Study: Is Cloud Computing Right for Your Company? (II)

Now, consider a small company **Y** that needs to digitalize its business and reduce costs to compete effectively in the market. The company is also considering whether to adopt cloud computing or traditional infrastructure.

*The decision depends on the specific needs and resources of the company. If the company has a significant amount of data, a traditional infrastructure might be more suitable due to the increased control over data security. However, if the company values flexibility and scalability, cloud computing could be a better choice.*

## Case Study: Is Cloud Computing Right for Your Company? (III)

Fast forward a few years, and company **X** has achieved remarkable success. The company has grown significantly and its needs have evolved.

*The cost of using public cloud services increases as the company grows, but the company can now afford to invest in a private cloud. At this point, a hybrid cloud solution can help to balance between the advantages of public and private clouds. For example, **Microsoft Azure** offers hybrid cloud solutions that combine the best of both worlds.*

### Take Home Message

There's no one-size-fits-all answer when it comes to choosing between Cloud or Traditional infrastructure. The optimal solution hinges on the specific needs and resources of the company. A thorough analysis of the costs, benefits, and potential risks associated with each option is crucial to making an informed decision.

## **Cloud Computing in Health**

---

## Why Cloud Computing in Health?

- **Big Data:** Manage and store large volumes of data, including patient records, medical images, and other health information.
- **Interoperability:** Share data across different systems and platforms, improving communication and collaboration among providers.
- **Digital Transformation:** Adopt new technologies and services, such as telemedicine, remote monitoring, mobile health apps or electronic health records.

In his PhD thesis, **Management of Cloud Systems Applied to eHealth**, ([Vilaplana Mayoral et al. 2015](#)) conducts a comprehensive analysis of the adoption of cloud computing within the healthcare sector. He implemented various cloud-based tools for telemedicine and remote patient monitoring, highlighting their transformative impact on healthcare delivery.

# Which are the opportunities of Cloud Computing in Health?

## **Capital expenses savings**

Healthcare organization can rent computing resources on a pay-as-you-go basis, reducing the need for large upfront investments in expensive hardware and software.

## **Improved patient care and decision making**

Healthcare professionals can store and analyze large volumes of data to build predictive and prescriptive models, improving patient care and decision making.

## **Better coordination among stakeholders**

Patients and healthcare professionals can access and share data more easily.

*More opportunities in the healthcare sector are discussed in [\(Ali et al. 2018\)](#).*

- **Sequence Analysis:** AWS offers services like AWS Lambda and Amazon S3 that can be used to run sequence analysis software like BLAST. This allows healthcare organizations to compare patient DNA sequences with known sequences to identify genetic mutations.
- **Genomic Data Storage:** Google Cloud Platform (GCP) provides a solution for storing genomic data. It offers services like Google Cloud Storage and BigQuery which can be used to store and analyze large volumes of genomic data.
- **Telemedicine:** Microsoft Azure has a service called Azure Communication Services that can be used to build telemedicine applications. It provides capabilities for video calling, chat, and SMS, enabling healthcare providers to consult with patients remotely.
- **Remote Patient Monitoring:** IBM Watson Health offers solutions for remote patient monitoring. It uses AI to analyze data from various sources such as electronic health records and wearable devices, providing healthcare professionals with insights to improve patient care.

### Among others....

- **Security and Privacy:** Healthcare organizations are obligated to adhere to stringent regulations and standards to safeguard patient data. LOPD in Spain or HIPAA in the United States mandate the protection of personal data and guarantees digital rights.
- **Data Integration:** Healthcare organizations are required to integrate data from diverse sources and systems to deliver comprehensive care. This integration is crucial in transforming isolated information units into a unified system of knowledge and action.

# Is cloud computing the future of Health?

No, cloud computing is not the future of health, it is the present and it is an essential part of the future.

## Cloud Computing is evolving continuously

- **IoT devices:** Sensors, wearables, and other devices that collect and transmit data.
- **Mobile devices:** Smartphones and tablets that can be used to access and share data.
- **Edge computing:** Facilitates real-time data processing and analysis at the source of data generation.
- **Fog computing:** Extends cloud computing to the edge of the network to improve efficiency and reduce latency.

## Big data ⇒ Quality

There is a growing demand for high-quality data in healthcare. Cloud computing can help healthcare organizations manage and analyze large volumes of data to improve patient care and outcomes. But it is not only about the **quantity of data**, it is about the **quality of the data**. Distributed Artificial Intelligence can help to analyze data and identify patterns and trends in real time.

## **Conclusion**

---

## Recap: Cloud Computing vs Traditional Computing

### Cloud Computing

- Pay what you use
- No server space needed
- No experts required for hardware and software maintenance
- Disaster recovery
- High flexibility
- Automated software updates
- Teams can collaborate from different locations
- Data can be accessed and shared from anywhere
- Rapid implementation

### Traditional

- Costly and less scalable
- Space needed for the servers
- Hardware and software team for maintenance
- Less flexibility
- No automated updates
- Less collaboration
- Data cannot be accessed remotely
- Takes long time for implementation

## Tasks

---

### HandsOnLab01: **Deploying your personal website**

The goal of this hands-on lab is to show you how to deploy a personal website using cloud computing.

For this hands-on lab, you will need to follow the instructions: [HOL01](#)

### HandsOnLab01: **Deploying your personal website**

The goal of this hands-on lab is to show you how to deploy a personal website using cloud computing.

For this hands-on lab, you will need to follow the instructions: [HOL01](#)

**Cloud Computing Fundamentals Quiz:** This quiz will test your understanding of the fundamental concepts of cloud computing. [Quiz](#).

## GitHub – *The platform*



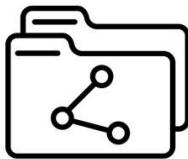
## GitHub – *The services*

---

GitHub Pages



GitHub Repositories



GitHub Copilot



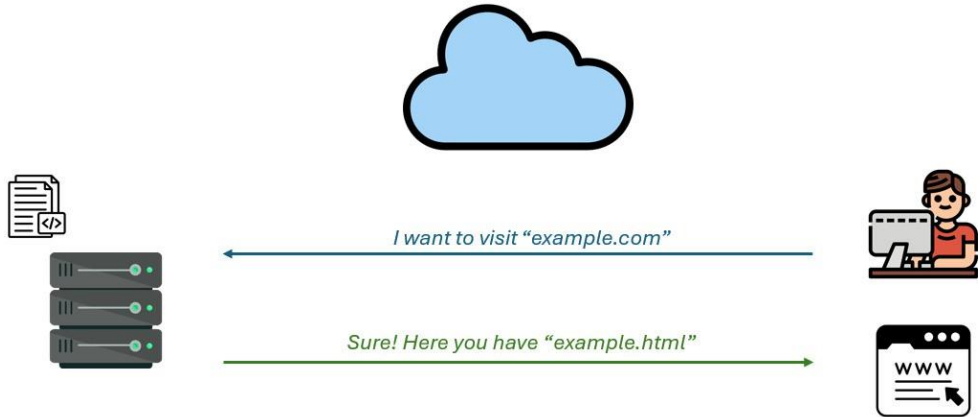
Figure 7: Some of the services GitHub offers

## Is GitHub a PaaS, IaaS, or SaaS?

- **GitHub Repositories:** A GitHub repository (often called a repo) is basically a storage space on GitHub where you keep all the files and history of a project.
- **GitHub Pages:** GitHub Pages is a static website hosting platform. You put HTML, CSS, JavaScript, Python ... files in a GitHub repo. You can use it for personal websites, project documentation, portfolios, blogs, class notes / course materials, simple static web apps.
- **GitHub Copilot:** GitHub Copilot is an AI-powered coding assistant created by GitHub + Microsoft + OpenAI. It helps programmers write code faster and with fewer mistakes by suggesting code in real time.

**Figure 7:** Some of the services GitHub offers

# What does it mean to host a website?



**Figure 8:** What happens when I want to visit a website?

## Example. Website deployment

GitHub as a PaaS to deploy our own website and compare it to a local deployment.

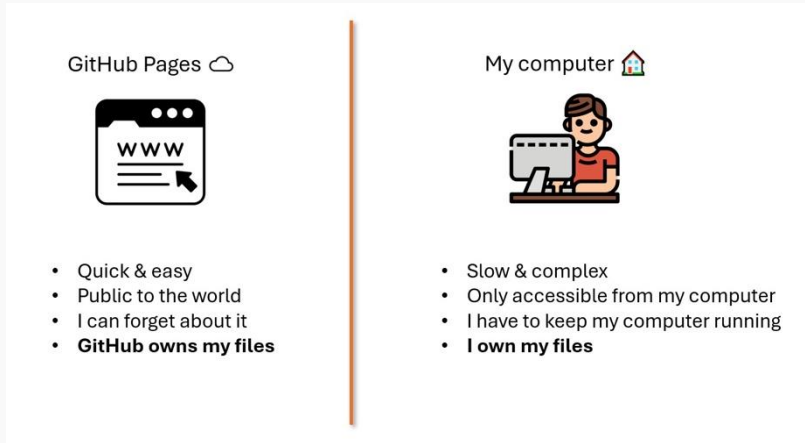


Figure 2: Differences between GitHub Pages and local deployment

### On my computer

- I have to download a repository
- I have to install software to build the website
- I have to install software to serve the website
- I have to edit the files
- I have to build the website
- I have to serve the website

Result? A website accessible only from my computer only while the software is running

### With GitHub Pages

- Fork a repository
- Tweak some settings
- Edit the files
- Commit the changes

Result? A free website accessible from anywhere

- Ali, Omar, Anup Shrestha, Jeffrey Soar, and Samuel Fosso Wamba. 2018. "Cloud Computing-Enabled Healthcare Opportunities, Issues, and Applications: A Systematic Review." *International Journal of Information Management* 43: 146–58.
- Mell, Peter, Tim Grance, et al. 2011. "The NIST Definition of Cloud Computing."
- Vilaplana Mayoral, Jordi et al. 2015. *Management of Cloud Systems Applied to eHealth*. Universitat de Lleida.