Lesson 3
Cloud computing

Socialize key words: "Cloud computing".

Cloud Computing
On-demand access to computing resources, servers, data storage, development tools, etc., over the Internet, managed by a cloud service provider.

Cost Optimization
The process of reducing IT costs by offloading the expenses of purchasing, installing, and managing on-premises infrastructure to the cloud.

Scalability
The ability to easily adjust IT resources, such as computing power, storage, and bandwidth, based on demand, providing flexibility and cost-effectiveness.

Virtualization

The technology abstracting and grouping virtualized IT infrastructure, including servers, operating systems, and networks, independently of physical hardware boundaries.

PaaS (Platform as a Service)

A cloud service model offering developers on-demand platforms, including hardware, software stack, and development tools, to run, develop, and manage applications without maintaining on-premises platforms.

SaaS (Software as a Service)

A cloud service model where software applications are hosted in the cloud and accessed through web browsers or APIs, typically with a monthly or annual subscription fee.

IaaS (Infrastructure as a Service)

On-demand access to essential computing resources (physical and virtual servers, networks, storage) over the Internet, allowing users to scale resources based on needs.

Serverless Computing

A cloud computing model that delegates backend infrastructure management tasks to the cloud provider, allowing developers to focus on code and business logic, with automatic scaling based on demand.

FaaS (Function as a Service)

A subset of serverless technology where developers execute specific functions in response to events, with the cloud provider managing all necessary resources in real-time.

Global Public Cloud Spending

The projected global spending by end-users on public cloud services, estimated to reach nearly $600 billion in 2023 according to industry analyst firm Gartner.

Reading: "Cloud computing"

What is cloud computing?

Cloud computing is on-demand access, over the Internet, to computing resources, servers (both physical and virtual), data storage, development tools, networking capabilities, and more, hosted in a remote data center managed by a cloud service provider (or CSP). The CSP makes these resources available for a monthly subscription fee or bills them based on usage.

Compared to traditional on-premises IT, and depending on the cloud services you select, cloud computing helps you do the following:

Reduce IT costs: The cloud allows you to offload some or most of the costs and effort of purchasing, installing, configuring and managing your own on-premises infrastructure.

Improve agility and time to value: With the cloud, your organization can start using business applications in minutes, instead of waiting weeks or months for IT to respond to a request, purchase and configure supporting hardware, and install software. The cloud also allows you to empower certain users (specifically developers and data scientists) to make use of supporting software and infrastructure.

Scale Easier and More Cost-Effective: Cloud Provides Elasticity; Instead of purchasing excess capacity that remains unused during slow periods, you can increase or decrease capacity in response to spikes and dips in traffic. You can also take advantage of your cloud provider's global network to distribute your applications closer to users around the world.

The term "cloud computing" also refers to the technology that makes the cloud work. This includes some form of virtualized IT infrastructure: servers, operating system software, networks and other infrastructure that is abstracted, using special software, so that it can be grouped and divided independently of the boundaries of physical hardware. For example, a single hardware server can be divided into multiple virtual servers.

Virtualization allows cloud providers to make the most of their data center resources . Not surprisingly, many corporations have adopted the cloud delivery model for their on-premises infrastructure so they can minimize usage and costs compared to traditional IT infrastructure and offer the same self-service and agility to their end users. .

If you use a computer or mobile device at home or at work, you almost certainly use some form of cloud computing every day, whether it's a cloud app like Google Gmail or Salesforce, streaming platforms like Netflix or cloud file storage like Dropbox. Industry analyst firm Gartner recently projected that global public cloud spending by end users will reach nearly $600 billion in 2023.

Cloud computing services

IaaS (infrastructure as a service) , PaaS (platform as a service), and SaaS (software as a service) are the three most common cloud service models, and it is not uncommon for an organization to use all three.

SaaS (software as a service)

SaaS, also known as cloud-based software or cloud applications, is application software that is hosted in the cloud and accessed by users through a web browser, a dedicated desktop client, or an API that integrates with a desktop or mobile operating system. In most cases, SaaS users pay a monthly or annual subscription fee; some may offer "pay-as-you-go" pricing based on your actual usage.

In addition to the cost reduction, time-to-value, and scalability benefits of the cloud, SaaS offers the following:

Automatic updates
With SaaS, users take advantage of new features as soon as the provider adds them, without having to orchestrate an in-place update.

Protection against data loss
Since SaaS stores application data in the cloud with the application, users do not lose data if their device fails or breaks down.

SaaS is the primary delivery model for most business software today
there are hundreds of thousands of SaaS solutions available, from the most focused industrial and departmental applications, to powerful enterprise software databases and artificial intelligence (AI) software.).

PaaS (platform as a service)

PaaS provides software developers with an on-demand platform (hardware, full software stack, infrastructure, and even development tools) to run, develop, and manage applications without the cost, complexity, and inflexibility of maintaining that platform on-premises.

With PaaS, the cloud provider hosts everything – servers, networks, storage, operating system software, middleware, databases – in its data center. Developers simply choose from a menu the servers and environments they need to run, build, test, deploy, maintain, update, and scale applications.

Today, PaaS is typically built around containers , a virtualized computing model indirectly related to virtual servers. Containers virtualize the operating system, thus allowing developers to package the application with only the operating system services it needs to run on any platform, without modifying them or the need for middleware .

Red Hat OpenShift is a popular PaaS built around Docker containers and Kubernetes , an open source container orchestration solution that automates deployment, scaling, load balancing , and more for container-based applications.

IaaS (infrastructure as a service)

IaaS provides on-demand access to essential computing resources (physical and virtual servers, networks and storage) over the Internet on a pay-as-you-go basis. IaaS allows end users to scale up and down resources based on their needs, resulting in less initial capital expenditure and reducing the need for unnecessary on-premises "own" infrastructure, as well as over-purchasing resources to accommodate periodic spikes in usage.

Unlike SaaS and PaaS (and even newer PaaS computing models such as containers and serverless), IaaS provides users with lower-level control of cloud computing resources.

IaaS was the most popular cloud computing model when it emerged in the early 2010s. Although it remains the cloud model for many types of workloads, the use of SaaS and PaaS is growing at a much faster rate.
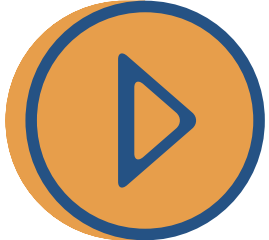
Serverless computing

Serverless computing (or simply serverless) is a cloud computing model that offloads all backend infrastructure management tasks (provisioning, scaling, scheduling, patching) to the cloud provider, freeing up developers to to focus all their time and effort on the code and business logic specific to their applications.

Additionally, serverless application code runs on a per-request basis, and the supporting infrastructure automatically scales up or down based on the number of requests. With serverless technology, clients pay only for the resources that are used when the application is running, they never pay for idle capacity.

FaaS, or Function as a Service , is often confused with serverless computing when, in fact, it is a subset of serverless technology. FaaS allows developers to execute parts of application code (called functions) in response to specific events. In addition to the code, the cloud service provider automatically provides everything needed (physical hardware, virtual machine operating system, and web server software management) in real time, as the code runs, and retires it once execution is completed. Billing starts when the run starts and stops when the run stops.

Inferential reading activity based on the text: "Cloud computing"

Inference about Cost Savings
What specific aspects of on-premises IT costs does cloud computing eliminate, and how does this contribute to cost optimization?

Inference about Time-to-Value
How does cloud computing improve agility, and what role does it play in reducing the time it takes for organizations to start using business applications?

Inference about Scalability
In what ways does cloud computing provide scalability, and how does it differ from traditional IT infrastructure in handling spikes and dips in traffic?

Inference about Virtualization

How does virtualization support the cloud delivery model, and why is it mentioned as a key aspect of the technology that makes the cloud work?

Inference about Cloud Service Models

What are the main characteristics and benefits associated with SaaS, PaaS, and IaaS, and why might an organization choose to use all three cloud service models?