



# What is cloud computing?

9 minutes

Have you ever wondered what cloud computing is? It's the delivery of computing services over the internet, which is otherwise known as the cloud. These services include servers, storage, databases, networking, software, analytics, and intelligence. Cloud computing offers faster innovation, flexible resources, and economies of scale.



---

## Why is cloud computing typically cheaper to use?

Cloud computing is the delivery of computing services over the internet by using a pay-as-you-go pricing model. You typically pay only for the cloud services you use, which helps you:

- Lower your operating costs.
- Run your infrastructure more efficiently.
- Scale as your business needs change.

To put it another way, cloud computing is a way to rent compute power and storage from someone else's datacenter. You can treat cloud resources like you would your resources in your own datacenter. When you're done using them, you give them back. You're billed only for what you use.

Instead of maintaining CPUs and storage in your datacenter, you rent them for the time that you need them. The cloud provider takes care of maintaining the underlying infrastructure for

you. The cloud enables you to quickly solve your toughest business challenges, and bring cutting-edge solutions to your users.

## Why should I move to the cloud?

The cloud helps you move faster and innovate in ways that were once nearly impossible.

In our ever-changing digital world, two trends emerge:

- Teams deliver new features to their users at record speeds.
- Users expect an increasingly rich and immersive experience with their devices and with software.

Software releases were once scheduled in terms of months or even years. Today, teams release features in smaller batches that are often scheduled in days or weeks. Some teams even deliver software updates continuously--sometimes with multiple releases within the same day.

Think of all the ways you interact with devices that you couldn't do a few years ago. Many devices can recognize your face and respond to voice commands. Augmented reality changes the way you interact with the physical world. Household appliances are even beginning to act intelligently. These technologies are only a few examples, and many of them are powered by the cloud.

To power your services and deliver innovative and novel user experiences more quickly, the cloud provides on-demand access to:

- A nearly limitless pool of raw compute, storage, and networking components.
- Speech recognition and other cognitive services that help make your application stand out from the crowd.
- Analytics services that deliver telemetry data from your software and devices.

## What are some cloud computing advantages?

There are several benefits that a cloud environment has over a physical environment. For example, cloud-based applications employ a myriad of related strategies:

- **Reliability:** Depending on the service-level agreement that you choose, your cloud-based applications can provide a continuous user experience with no apparent downtime even when things go wrong.
- **Scalability:** Applications in the cloud can be scaled in two ways, while taking advantage of autoscaling:
  - *Vertically:* Computing capacity can be increased by adding RAM or CPUs to a virtual machine.

- **Horizontally:** Computing capacity can be increased by adding instances of a resource, such as adding more virtual machines to your configuration.
- **Elasticity:** Cloud-based applications can be configured to always have the resources they need.
- **Agility:** Cloud-based resources can be deployed and configured quickly as your application requirements change.
- **Geo-distribution:** Applications and data can be deployed to regional datacenters around the globe, so your customers always have the best performance in their region.
- **Disaster recovery:** By taking advantage of cloud-based backup services, data replication, and geo-distribution, you can deploy your applications with the confidence that comes from knowing that your data is safe in the event that disaster should occur.

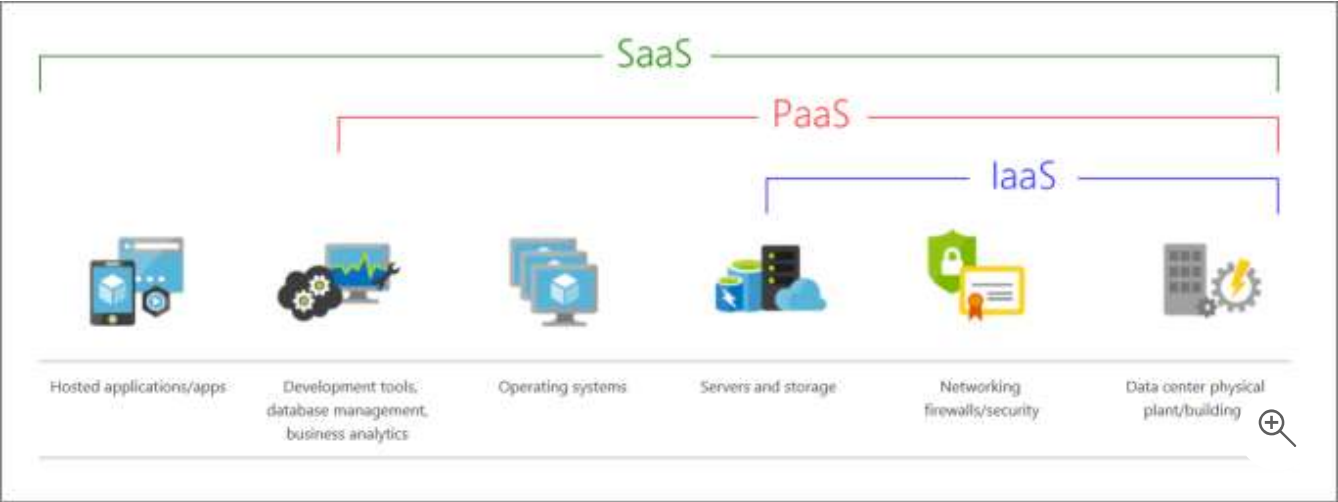
## What are cloud service models?

Cloud computing falls into one of the following computing models. If you've been around cloud computing for a while, you've probably seen the terms infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS) for the different cloud service models. These models define the different level of shared responsibility that a cloud provider and cloud tenant are responsible for.

Computing model	Description
IaaS	This cloud service model is the closest to managing physical servers. A cloud provider keeps the hardware up to date, but operating system maintenance and network configuration is left to the cloud tenant. For example, Azure virtual machines are fully operational virtual compute devices running in Microsoft's datacenters. An advantage of this cloud service model is rapid deployment of new compute devices. Setting up a new virtual machine is considerably faster than procuring, installing, and configuring a physical server.
PaaS	This cloud service model is a managed hosting environment. The cloud provider manages the virtual machines and networking resources, and the cloud tenant deploys their applications into the managed hosting environment. For example, Azure App Services provides a managed hosting environment where developers can upload their web applications without having to deal with the physical hardware and software requirements.

Computing model	Description
SaaS	In this cloud service model, the cloud provider manages all aspects of the application environment, such as virtual machines, networking resources, data storage, and applications. The cloud tenant only needs to provide their data to the application managed by the cloud provider. For example, Office 365 provides a fully working version of Office that runs in the cloud. All that you need to do is create your content, and Office 365 takes care of everything else.

The following illustration demonstrates the services that might run in each of the cloud service models.



The following chart illustrates the various levels of responsibility between a cloud provider and a cloud tenant.

	You Manage	Cloud Provider Manages	
On-premises (Private Cloud)	Infrastructure (as a Service)	Platform (as a Service)	Software (as a Service)
Data & Access	Data & Access	Data & Access	Data & Access
Applications	Applications	Applications	Applications
Runtime	Runtime	Runtime	Runtime
Operating System	Operating System	Operating System	Operating System
Virtual Machine	Virtual Machine	Virtual Machine	Virtual Machine
Compute	Compute	Compute	Compute
Networking	Networking	Networking	Networking
Storage	Storage	Storage	Storage

# What is serverless computing?

Overlapping with PaaS, serverless computing enables developers to build applications faster by eliminating the need for them to manage infrastructure. With serverless applications, the cloud service provider automatically provisions, scales, and manages the infrastructure required to run the code. Serverless architectures are highly scalable and event-driven. They use resources only when a specific function or trigger occurs.

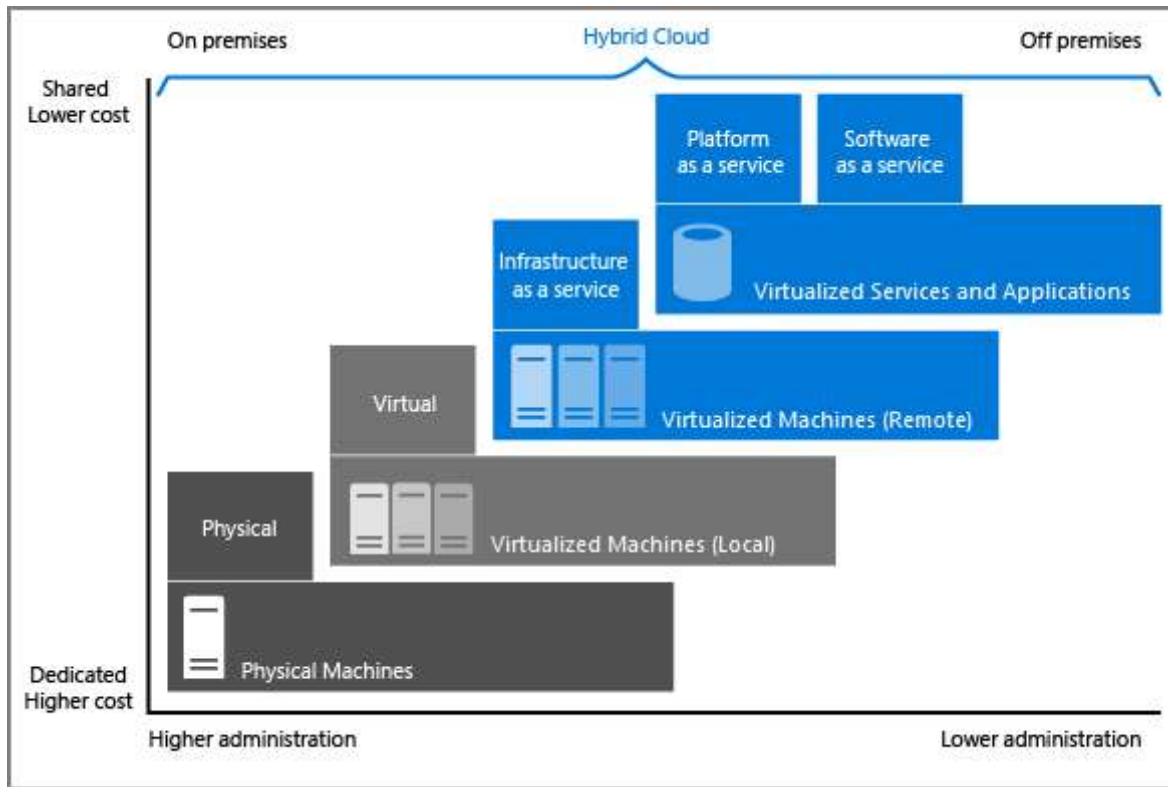
In understanding the definition of serverless computing, it's important to note that servers are still running the code. The serverless name comes from the fact that the tasks associated with infrastructure provisioning and management are invisible to the developer. This approach enables developers to increase their focus on the business logic and deliver more value to the core of the business. Serverless computing helps teams increase their productivity and bring products to market faster. It allows organizations to better optimize resources and stay focused on innovation.

## What are public, private, and hybrid clouds?

There are three deployment models for cloud computing: *public cloud*, *private cloud*, and *hybrid cloud*. Each deployment model has different aspects that you should consider as you migrate to the cloud.

Deployment model	Description
Public cloud	Services are offered over the public internet and available to anyone who wants to purchase them. Cloud resources like servers and storage are owned and operated by a third-party cloud service provider and delivered over the internet.
Private cloud	Computing resources are used exclusively by users from one business or organization. A private cloud can be physically located at your organization's on-site datacenter. It also can be hosted by a third-party service provider.
Hybrid cloud	This computing environment combines a public cloud and a private cloud by allowing data and applications to be shared between them.

The following image illustrates several of the cloud computing concepts that are presented in this unit. In this example, several factors are demonstrated when you're considering where to deploy a database server in a hybrid cloud environment. As your resources move from on-premises to off-premises, your costs are reduced, and your administration requirements decrease.



## Next unit: What is Azure?

[Continue >](#)