[Cloud Computing](https://www.oracle.com/in/cloud/what-is-cloud-computing/)-Oracle

[Cloud Computing](https://www.geeksforgeeks.org/cloud-computing/)-Geekforgeek

[Cloud Technology](https://www.ibm.com/topics/cloud-computing) by IBM

Cloud computing is the on-demand access of computing resources—physical servers or virtual servers, data storage, networking capabilities, application development tools, software, AI-powered analytic tools and more—over the internet with pay-per-use pricing.

The cloud computing model offers customers greater flexibility and scalability compared to traditional on-premises infrastructure.

Cloud computing plays a pivotal role in our everyday lives, whether accessing a cloud application like Google Gmail, streaming a movie on Netflix or playing a cloud-hosted video game.

Cloud computing has also become indispensable in business settings, from small startups to global enterprises. Its many business applications include enabling remote work by making data and applications accessible from anywhere, creating the framework for seamless omnichannel customer engagement and providing the vast computing power and other resources needed to take advantage of cutting-edge technologies like [generative AI](https://research.ibm.com/blog/what-is-generative-AI) and [quantum computing](https://www.ibm.com/topics/quantum-computing#:).

A cloud services provider (CSP) manages cloud-based technology services hosted at a remote [data center](https://www.ibm.com/topics/data-centers) and typically makes these resources available for a pay-as-you-go or monthly subscription fee.

Benefits of cloud computing

Compared to traditional on-premises IT that involves a company owning and maintaining physical data centers and servers to access computing power, data storage and other resources (and depending on the cloud services you select), cloud computing offers many benefits, including the following:

**Cost-effectiveness**

Cloud computing lets you offload some or all of the expense and effort of purchasing, installing, configuring and managing [mainframe computers](https://www.ibm.com/topics/mainframe) and other on-premises infrastructure. You pay only for cloud-based infrastructure and other computing resources as you use them.

**Increased speed and agility**

With cloud computing, your organization can use enterprise applications in minutes instead of waiting weeks or months for IT to respond to a request, purchase and configure supporting hardware and install software. This feature empowers users—specifically [DevOps](https://www.ibm.com/topics/devops) and other development teams—to help leverage cloud-based software and support infrastructure.

**Unlimited scalability**

Cloud computing provides elasticity and self-service provisioning, so instead of purchasing excess capacity that sits unused during slow periods, you can scale capacity up and down in response to spikes and dips in traffic. You can also use your cloud provider’s global network to spread your applications closer to users worldwide.

**Enhanced strategic value**

Cloud computing enables organizations to use various technologies and the most up-to-date innovations to gain a competitive edge. For instance, in retail, banking and other customer-facing industries, generative AI-powered virtual assistants deployed over the cloud can deliver better customer response time and free up teams to focus on higher-level work. In manufacturing, teams can collaborate and use cloud-based software to monitor real-time data across logistics and supply chain processes.

Origins of cloud computing

The origins of cloud computing technology go back to the early 1960s when [Dr. Joseph Carl Robnett Licklider](https://www.internethalloffame.org/inductee/jcr-licklider/) (link resides outside ibm.com), an American computer scientist and psychologist known as the "father of cloud computing", introduced the earliest ideas of global networking in a series of memos discussing an *Intergalactic Computer Network.* However, it wasn’t until the early 2000s that modern cloud infrastructure for business emerged.

In 2002, Amazon Web Services started cloud-based storage and computing services. In 2006, it introduced Elastic Compute Cloud (EC2), an offering that allowed users to rent virtual computers to run their applications. That same year, Google introduced the Google Apps suite (now called Google Workspace), a collection of SaaS productivity applications. In 2009, Microsoft started its first SaaS application, Microsoft Office 2011. Today, [Gartner predicts](https://www.gartner.com/en/newsroom/press-releases/2023-11-29-gartner-says-cloud-will-become-a-business-necessity-by-2028) worldwide end-user spending on the public cloud will total USD 679 billion and is projected to exceed USD 1 trillion in 2027 (link resides outside ibm.com).

Cloud computing components

The following are a few of the most integral components of today’s modern cloud computing architecture.

**Data centers**

CSPs own and operate remote data centers that house physical or [bare metal servers](https://www.ibm.com/topics/bare-metal-dedicated-servers), [cloud storage](https://www.ibm.com/topics/cloud-storage) systems and other physical hardware that create the underlying infrastructure and provide the physical foundation for cloud computing.

**Networking capabilities**

In cloud computing, high-speed [networking](https://www.ibm.com/topics/networking) connections are crucial. Typically, an internet connection known as a wide-area network (WAN) connects front-end users (for example, client-side interface made visible through web-enabled devices) with back-end functions (for example, data centers and cloud-based applications and services). Other advanced cloud computing networking technologies, including [load balancers](https://www.ibm.com/topics/load-balancing), [content delivery networks (CDNs)](https://www.ibm.com/topics/content-delivery-networks) and [software-defined networking (SDN)](https://www.ibm.com/topics/sdn), are also incorporated to ensure data flows quickly, easily and securely between front-end users and back-end resources.

**Virtualization**

Cloud computing relies heavily on the [virtualization](https://www.ibm.com/topics/virtualization) of [IT infrastructure](https://www.ibm.com/topics/infrastructure#:)—servers, operating system software, networking and other infrastructure that’s abstracted using special software so that it can be pooled and divided irrespective of physical hardware boundaries. For example, a single hardware server can be divided into multiple [virtual servers](https://www.ibm.com/topics/virtual-server). Virtualization enables cloud providers to make maximum use of their data center resources.

Cloud computing services

IaaS (Infrastructure-as-a-Service), PaaS (Platform-as-a-Service), SaaS (Software-as-a-Service) and serverless computing are the most common models of cloud services, and it’s not uncommon for an organization to use some combination of all four.

IaaS (Infrastructure-as-a-Service)

[IaaS (Infrastructure-as-a-Service)](https://www.ibm.com/topics/iaas) provides on-demand access to fundamental computing resources—physical and virtual servers, networking and storage—over the internet on a pay-as-you-go basis. IaaS enables end users to scale and shrink resources on an as-needed basis, reducing the need for high up-front capital expenditures or unnecessary on-premises or "owned" infrastructure and for overbuying resources to accommodate periodic spikes in usage.

According to a [Business Research Company report](https://www.thebusinessresearchcompany.com/report/infrastructure-as-a-service-global-market-report) (link resides outside ibm.com), the IaaS market is predicted to grow rapidly in the next few years, growing to $212.34 billion in 2028 at a compound annual growth rate (CAGR) of 14.2%.

PaaS (Platform-as-a-Service)

[PaaS (Platform-as-a-Service)](https://www.ibm.com/topics/paas) provides software developers with an on-demand platform—hardware, complete software stack, infrastructure and development tools—for running, developing and managing applications without the cost, complexity and inflexibility of maintaining that platform on-premises. With PaaS, the cloud provider hosts everything at their data center. These include servers, networks, storage, operating system software, [middleware](https://www.ibm.com/topics/middleware) and databases. Developers simply pick from a menu to spin up servers and environments they need to run, build, test, deploy, maintain, update and scale applications.

Today, PaaS is typically built around [container*s*](https://www.ibm.com/topics/containers), a virtualized compute model one step removed from virtual servers. Containers virtualize the operating system, enabling developers to package the application with only the operating system services it needs to run on any platform without modification and the need for middleware.

[Red Hat® OpenShift](https://www.ibm.com/products/openshift)® is a popular PaaS built around [Docker](https://www.ibm.com/topics/docker) containers and [Kubernetes](https://www.ibm.com/topics/kubernetes), an open source [container orchestration](https://www.ibm.com/topics/container-orchestration) solution that automates deployment, scaling, load balancing and more for container-based applications.

SaaS (Software-as-a-Service)

[SaaS (Software-as-a-Service)](https://www.ibm.com/topics/saas), also known as cloud-based software or cloud applications, is application software hosted in the cloud. Users access SaaS through a web browser, a dedicated desktop client or an API that integrates with a desktop or mobile operating system. Cloud service providers offer SaaS based on a monthly or annual subscription fee. They may also provide these services through pay-per-usage pricing.

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

* **Automatic upgrades:** With SaaS, users use new features when the cloud service provider adds them without orchestrating an on-premises upgrade.
* **Protection from data loss:** Because SaaS stores application data in the cloud with the application, users don’t lose data if their device crashes or breaks.

SaaS is the primary delivery model for most commercial software today. Hundreds of SaaS solutions exist, from focused industry and broad administrative (for example, Salesforce) to robust enterprise database and [artificial intelligence (AI)](https://www.ibm.com/topics/artificial-intelligence) software. According to an International Data Center (IDC) survey (the link resides outside IBM), SaaS applications represent the largest cloud computing segment, accounting for more than 48% of the $778 billion worldwide cloud software revenue.

Serverless computing

[Serverless computing](https://www.ibm.com/topics/serverless), or simply serverless, is a cloud computing model that offloads all the back-end infrastructure management tasks, including provisioning, scaling, scheduling and patching to the cloud provider. This frees developers to focus all their time and effort on the code and business logic specific to their applications.

Moreover, serverless runs application code on a per-request basis only and automatically scales the supporting infrastructure up and down in response to the number of requests. With serverless, customers pay only for the resources used when the application runs; they never pay for idle capacity.

[FaaS, or Function-as-a-Service](https://www.ibm.com/topics/faas), is often confused with serverless computing when, in fact, it’s a subset of serverless. FaaS allows developers to run portions of application code (called functions) in response to specific events. Everything besides the code—physical hardware, virtual machine (VM) operating system and web server software management—is provisioned automatically by the cloud service provider in real-time as the code runs and is spun back down once the execution is complete. Billing starts when execution starts and stops when execution stops.

Types of cloud computing

Public cloud

A [public cloud](https://www.ibm.com/topics/public-cloud) is a type of cloud computing in which a cloud service provider makes computing resources available to users over the public internet. These include SaaS applications, individual [virtual machines (VMs)](https://www.ibm.com/topics/virtual-machines), bare metal computing hardware, complete enterprise-grade infrastructures and development platforms. These resources might be accessible for free or according to subscription-based or pay-per-usage pricing models.

The public cloud provider owns, manages and assumes all responsibility for the data centers, hardware and infrastructure on which its customers’ workloads run. It typically provides high-bandwidth network connectivity to ensure high performance and rapid access to applications and data.

Public cloud is a [multi-tenant environment](https://www.ibm.com/topics/multi-tenant) where all customers pool and share the cloud provider’s data center infrastructure and other resources. In the world of the leading public cloud vendors, such as Amazon Web Services (AWS), Google Cloud, IBM Cloud®, Microsoft Azure and Oracle Cloud, these customers can number in the millions.

Most enterprises have moved portions of their computing infrastructure to the public cloud since public cloud services are elastic and readily scalable, flexibly adjusting to meet changing workload demands. The promise of greater efficiency and cost savings through paying only for what they use attracts customers to the public cloud. Still, others seek to reduce spending on hardware and on-premises infrastructure. [Gartner predicts](https://www.gartner.com/en/newsroom/press-releases/2023-04-19-gartner-forecasts-worldwide-public-cloud-end-user-spending-to-reach-nearly-600-billion-in-2023) (link resides outside ibm.com) that by 2026, 75% of organizations will adopt a digital transformation model predicated on cloud as the fundamental underlying platform.

Private cloud

A [private cloud](https://www.ibm.com/topics/private-cloud) is a cloud environment where all cloud infrastructure and computing resources are dedicated to one customer only. Private cloud combines many benefits of cloud computing—including elasticity, scalability and ease of service delivery—with the access control, security and resource customization of on-premises infrastructure.

A private cloud is typically hosted on-premises in the customer’s data center. However, it can also be hosted on an independent cloud provider’s infrastructure or built on rented infrastructure housed in an offsite data center.

Many companies choose a private cloud over a public cloud environment to meet their regulatory compliance requirements. Entities like government agencies, healthcare organizations and financial institutions often opt for private cloud settings for workloads that deal with confidential documents, personally identifiable information (PII), intellectual property, medical records, financial data or other sensitive data.

By building private cloud architecture according to [cloud-native](https://www.ibm.com/topics/cloud-native) principles, an organization can quickly move workloads to a public cloud or run them within a hybrid cloud (see below) environment whenever ready.

Hybrid cloud

A [hybrid cloud](https://www.ibm.com/topics/hybrid-cloud) is just what it sounds like: a combination of public cloud, private cloud and on-premises environments. Specifically (and ideally), a hybrid cloud connects a combination of these three environments into a single, flexible infrastructure for running the organization’s applications and workloads.

At first, organizations turned to hybrid cloud computing models primarily to migrate portions of their on-premises data into private cloud infrastructure and then connect that infrastructure to public cloud infrastructure hosted off-premises by cloud vendors. This process was done through a packaged hybrid cloud solution like Red Hat® OpenShift® or middleware and IT management tools to create a "[single pane of glass](https://www.ibm.com/topics/single-pane-of-glass)." Teams and administrators rely on this unified dashboard to view their applications, networks and systems.

Today, [hybrid cloud architecture](https://www.ibm.com/topics/hybrid-cloud-architecture) has expanded beyond physical connectivity and cloud migration to offer a flexible, secure and cost-effective environment that supports the portability and automated deployment of workloads across multiple environments. This feature enables an organization to meet its technical and business objectives more effectively and cost-efficiently than with a public or private cloud alone. For instance, a hybrid cloud environment is ideal for DevOps and other teams to develop and test web applications. This frees organizations from purchasing and expanding the on-premises physical hardware needed to run application testing, offering faster time to market. Once a team has developed an application in the public cloud, they may move it to a private cloud environment based on business needs or security factors.

A public cloud also allows companies to quickly scale resources in response to unplanned spikes in traffic without impacting private cloud workloads, a feature known as cloud bursting. Streaming channels like Amazon use cloud bursting to support the increased viewership traffic when they start new shows.

Most enterprise organizations today rely on a hybrid cloud model because it offers greater flexibility, scalability and cost optimization than traditional on-premises infrastructure setups. According to the [*IBM Transformation Index: State of Cloud*](https://www.ibm.com/thought-leadership/institute-business-value/report/transformation-index), more than 77% of businesses and IT professionals have adopted a hybrid cloud approach.

To learn more about the differences between public, private and hybrid cloud, check out “[Public cloud vs. private cloud vs. hybrid cloud: What’s the difference?](https://www.ibm.com/think/topics/public-cloud-vs-private-cloud-vs-hybrid-cloud)”

[***Watch the IBM hybrid cloud architecture video series.***](https://www.youtube.com/watch?v=h_WE-ZFDZJQ) (link resides outside ibm.com)

[What is Hybrid Cloud? (9:45)- This link plays a video](https://www.ibm.com/topics/cloud-computing)

Multicloud

[Multicloud](https://www.ibm.com/topics/multicloud) uses two or more clouds from two or more different cloud providers. A multicloud environment can be as simple as email SaaS from one vendor and image editing SaaS from another. But when enterprises talk about multicloud, they typically refer to using multiple cloud services—including SaaS, PaaS and IaaS services—from two or more leading public cloud providers.

Organizations choose multicloud to avoid vendor lock-in, to have more services to select from and to access more innovation. With multicloud, organizations can choose and customize a unique set of cloud features and services to meet their business needs. This freedom of choice includes selecting “best-of-breed” technologies from any CSP, as needed or as they emerge, rather than being locked into offering from a single vendor. For example, an organization may choose AWS for its global reach with web-hosting, IBM Cloud for data analytics and [machine learning](https://www.ibm.com/topics/machine-learning) platforms and Microsoft Azure for its security features.

A multicloud environment also reduces exposure to licensing, security and compatibility issues that can result from "[shadow IT](https://www.ibm.com/topics/shadow-it)"— any software, hardware or IT resource used on an enterprise network without the IT department’s approval and often without IT’s knowledge or oversight.

The modern hybrid multicloud

Today, most enterprise organizations use a hybrid multicloud model. Apart from the flexibility to choose the most cost-effective cloud service, hybrid multicloud offers the most control over workload deployment, enabling organizations to operate more efficiently, improve performance and optimize costs. According to an [IBM® Institute for Business Value study](https://www.ibm.com/downloads/cas/QMRQEROB), the value derived from a full hybrid multicloud platform technology and operating model at scale is two-and-a-half times the value derived from a single-platform, single-cloud vendor approach.

Yet the modern hybrid multicloud model comes with more complexity. The more clouds you use—each with its own management tools, data transmission rates and security protocols—the more difficult it can be to manage your environment. With [over 97% of enterprises operating on more than one cloud](https://newsroom.ibm.com/IBM-Fireside-Chat-Unlocking-Deeper-Business-Transformation-with-Hybrid-Cloud) and most organizations running [10 or more clouds](https://www.ibm.com/downloads/cas/QMRQEROB), a hybrid cloud management approach has become crucial. Hybrid multicloud management platforms provide visibility across multiple provider clouds through a central dashboard where development teams can see their projects and deployments, operations teams can monitor clusters and nodes and the cybersecurity staff can monitor for threats.

Cloud security

Traditionally, security concerns have been the primary obstacle for organizations considering cloud services, mainly public cloud services. Maintaining cloud security demands different procedures and employee skillsets than in legacy IT environments. Some cloud security best practices include the following:

* **Shared responsibility for security:**Generally, the cloud service provider is responsible for securing cloud infrastructure, and the customer is responsible for protecting its data within the cloud. However, it’s also essential to clearly define data ownership between private and public third parties.
* **Data encryption:**Data should be encrypted while at rest, in transit and in use. Customers need to maintain complete control over security keys and hardware security modules.
* **Collaborative management:** Proper communication and clear, understandable processes between IT, operations and security teams will ensure seamless cloud integrations that are secure and sustainable.
* **Security and compliance monitoring:** This begins with understanding all regulatory compliance standards applicable to your industry and establishing active monitoring of all connected systems and cloud-based services to maintain visibility of all data exchanges across all environments, on-premises, private cloud, hybrid cloud and edge.

Cloud security is constantly changing to keep pace with new threats. Today’s CSPs offer a wide array of cloud security management tools, including the following:

* **Identity and access management (IAM):**IAMtools and services that automate policy-driven enforcement protocols for all users attempting to access both on-premises and cloud-based services.
* **Data loss prevention (DLP):** DLP services that combine remediation alerts data encryption and other preventive measures to protect all stored data, whether at rest or in motion.
* **Security information and event management (SIEM)**:[SIEM](https://www.ibm.com/topics/siem) is acomprehensive security orchestration solution that automates threat monitoring, detection and response in cloud-based environments. SIEM technology uses artificial intelligence (AI)-driven technologies to correlate log data across multiple platforms and digital assets. This allows IT teams to successfully apply their network security protocols, enabling them to react to potential threats quickly.
* **Automated data compliance platforms:** Automated software solutions provide compliance controls and centralized data collection to help organizations adhere to regulations specific to their industry. Regular compliance updates can be baked into these platforms so organizations can adapt to ever-changing regulatory compliance standards.

[Learn more about cloud security.](https://www.ibm.com/topics/cloud-security)

Cloud sustainability

[Sustainability in business](https://www.ibm.com/topics/business-sustainability#:), a company’s strategy to reduce negative environmental impact from their operations in a particular market, has become an essential corporate governance mandate.  Moreover, [Gartner predicts](https://www.gartner.com/en/newsroom/press-releases/2022-01-24-gartner-predicts-hyperscalers-carbon-emissions-will-drive-cloud-purchase-decsions-by-2025) (link resides outside ibm.com) that by 2025, the carbon emissions of hyperscale cloud services will be a top-three criterion in cloud purchase decisions.

As companies strive to advance their sustainability objectives, cloud computing has evolved to play a significant role in helping them reduce their carbon emissions and manage climate-related risks. For instance, traditional data centers require power supplies and cooling systems, which depend on large amounts of electrical power. By migrating IT resources and applications to the cloud, organizations only enhance operational and cost efficiencies and boost overall energy efficiency through pooled CSP resources.

All major cloud players have made net-zero commitments to reduce their carbon footprints and help clients reduce the energy they typically consume using an on-premises setup. For instance, IBM is driven by [sustainable procurement](https://www.ibm.com/topics/sustainable-procurement) initiatives to reach NetZero by 2030. By 2025, IBM Cloud worldwide data centers [will comprise energy procurement drawn from 75% renewable sources](https://www.ibm.com/blog/announcement/we-have-seen-the-future-and-it-is-sustainable/).

Cloud use cases

According to an [International Data Corporation (IDC) forecast](https://www.businesswire.com/news/home/20201015005069/en/Cloud-Adoption-and-Opportunities-Will-Continue-to-Expand-Leading-to-a-1-Trillion-Market-in-2024-According-to-IDC) (link resides outside ibm.com), worldwide spending on the whole cloud opportunity (offerings, infrastructure and services) will surpass USD 1 trillion in 2024 while sustaining a double-digit compound annual growth rate (CAGR) of 15.7%. Here are some of the main ways businesses are benefitting from cloud computing:

* **Scale infrastructure:**Allocate resources up or down quickly and easily in response to changes in business demands.
* **Enable business continuity and disaster recovery:**Cloud computing provides cost-effective redundancy to protect data against system failures and the physical distance required to apply [disaster recovery](https://www.ibm.com/topics/disaster-recovery) strategies and recover data and applications during a local outage or disaster. All of the major public cloud providers offer [Disaster-Recovery-as-a-Service (DRaaS)](https://www.ibm.com/topics/draas).
* **Build and test cloud-native applications**: For development teams adopting Agile, [DevOps](https://www.ibm.com/topics/devops) or [DevSecOps](https://www.ibm.com/topics/devsecops) to streamline development, the cloud offers on-demand end-user self-service that prevents operations tasks, such as spinning up development and test servers, from becoming development bottlenecks.
* **Support edge and IoT environments:**Address latency challenges and reduce downtime by bringing data sources closer to the [edge](https://www.ibm.com/topics/edge-computing). Support [Internet of Things (IoT)](https://www.ibm.com/topics/internet-of-things) devices (for example, patient monitoring devices and sensors on a production line) to gather real-time data.
* **Leverage cutting-edge technologies:** Cloud computing supports storing and processing huge volumes of data at high speeds—much more storage and computing capacity than most organizations can or want to purchase and deploy on-premises. These high-performance resources support technologies like [blockchain](https://www.ibm.com/topics/blockchain), quantum computing and [large language models (LLMs](https://www.ibm.com/topics/large-language-models#:)) that power generative AI platforms like customer service automation.