What Is Cloud Computing?

Cloud computing is the on-demand delivery of compute power, database, storage, applications, and other IT resources through a cloud services platform via the Internet with pay-as-you-go pricing. Whether you are running applications that share photos to millions of mobile users or you’re supporting the critical operations of your business, a cloud services platform provides rapid access to flexible and low-cost IT resources. With cloud computing, you don’t need to make large upfront investments in hardware and spend a lot of time on the heavy lifting of managing that hardware. Instead, you can provision exactly the right type and size of computing resources you need to power your newest bright idea or operate your IT department. You can access as many resources as you need, almost instantly, and only pay for what you use.

Cloud computing is the on-demand delivery of IT resources over the Internet with pay-as-you-go pricing. Instead of buying, owning, and maintaining physical data centers and servers, you can access technology services, such as computing power, storage, and databases, on an as-needed basis from a cloud provider like Amazon Web Services (AWS).

Six Advantages of Cloud Computing

• Trade fixed expense for variable expense – Instead of having to invest heavily in data centers and servers before you know how you’re going to use them, you can pay only when you consume computing resources, and pay only for how much you consume.

• Benefit from massive economies of scale – By using cloud computing, you can achieve a lower variable cost than you can get on your own. Because usage from hundreds of thousands of customers is aggregated in the cloud, providers such as AWS can achieve higher economies of scale, which translates into lower pay as-you-go prices.

• Stop guessing capacity – Eliminate guessing on your infrastructure capacity needs. When you make a capacity decision prior to deploying an application, you often end up either sitting on expensive idle resources or dealing with limited capacity. With cloud computing, these problems go away. You can access as much or as little capacity as you need, and scale up and down as required with only a few minutes’ notice.

• Increase speed and agility – In a cloud computing environment, new IT resources are only a click away, which means that you reduce the time to make those resources available to your developers from weeks to just minutes. This results in a dramatic increase in agility for the organization, since the cost and time it takes to experiment and develop is significantly lower.

• Stop spending money running and maintaining data centers – Focus on projects that differentiate your business, not the infrastructure. Cloud computing lets you focus on your own customers, rather than on the heavy lifting of racking, stacking, and powering servers.

• Go global in minutes – Easily deploy your application in multiple regions around the world with just a few clicks. This means you can provide lower latency and a better experience for your customers at minimal cost.

**Types of Cloud Computing**

Cloud computing provides developers and IT departments with the ability to focus on what matters most and avoid undifferentiated work such as procurement, maintenance, and capacity planning. Each type of cloud service and deployment method provides you with different levels of control, flexibility, and management.

**Cloud Computing Models**

**Infrastructure as a Service (IaaS)**

infrastructure as a service (IaaS) is a type of cloud computing service that offers essential compute, storage and networking resources on demand, on a pay-as-you-go basis. IaaS is one of the four types of cloud services, along with software as a service ([SaaS](https://azure.microsoft.com/en-in/overview/what-is-saas/)), platform as a service ([PaaS](https://azure.microsoft.com/en-in/overview/what-is-paas/)) and [serverless](https://azure.microsoft.com/en-in/overview/serverless-computing/).

Migrating your organisation's infrastructure to an IaaS solution helps you reduce maintenance of on-premises data centres, save money on hardware costs and gain real-time business insights. IaaS solutions give you the flexibility to scale your IT resources up and down with demand. They also help you quickly provision new applications and increase the reliability of your underlying infrastructure.

IaaS lets you bypass the cost and complexity of buying and managing physical servers and datacentre infrastructure. Each resource is offered as a separate service component and you only pay for a particular resource for as long as you need it. A [cloud computing service provider](https://azure.microsoft.com/en-in/overview/choosing-a-cloud-service-provider/) like [Azure](https://azure.microsoft.com/en-in/overview/what-is-azure/iaas/) manages the infrastructure, while you purchase, install, configure and manage your own software—including operating systems, middleware and applications.

**Platform as a Service (PaaS)**

Platform as a service (PaaS) is a complete development and deployment environment in the cloud, with resources that enable you to deliver everything from simple cloud-based apps to sophisticated, cloud-enabled enterprise applications. You purchase the resources you need from a [cloud service provider](https://azure.microsoft.com/en-in/overview/choosing-a-cloud-service-provider/) on a pay-as-you-go basis and access them over a secure Internet connection.

Like [IaaS](https://azure.microsoft.com/en-in/overview/what-is-iaas/), PaaS includes infrastructure—servers, storage and networking—but also middleware, development tools, business intelligence (BI) services, database management systems and more. PaaS is designed to support the complete web application lifecycle: building, testing, deploying, managing and updating.

PaaS allows you to avoid the expense and complexity of buying and managing software licenses, the underlying application infrastructure and middleware, container orchestrators such as [Kubernetes](https://azure.microsoft.com/en-in/topic/what-is-kubernetes/) or the development tools and other resources. You manage the applications and services you develop and the cloud service provider typically manages everything else.

**Software as a Service (SaaS)**

Software as a service (SaaS) allows users to connect to and use cloud-based apps over the Internet. Common examples are email, calendaring and office tools (such as Microsoft Office 365).

SaaS provides a complete software solution which you purchase on a pay-as-you-go basis from a [cloud service provider](https://azure.microsoft.com/en-in/overview/choosing-a-cloud-service-provider/). You rent the use of an app for your organisation and your users connect to it over the Internet, usually with a web browser. All of the underlying infrastructure, middleware, app software and app data are located in the service provider’s data center. The service provider manages the hardware and software and with the appropriate service agreement, will ensure the availability and the security of the app and your data as well. SaaS allows your organisation to get quickly up and running with an app at minimal upfront cost.

**Cloud Computing Deployment Models**

What exactly is deployment?

In its IT context, deployment encompasses **all the processes involved in getting new software or hardware up and running properly in its environment, including installation, configuration, running, testing, and making necessary changes**

In cloud computing, we have access to a shared pool of computer resources (servers, storage, programs, and so on) in the cloud. You simply need to request additional resources when you require them. Getting resources up and running quickly is a breeze thanks to the clouds. It is possible to release resources that are no longer necessary.

The cloud deployment model identifies the specific type of cloud environment based on ownership, scale, and access, as well as the cloud’s nature and purpose. The location of the servers you’re utilizing and who controls them are defined by a cloud deployment model. It specifies how your cloud infrastructure will look, what you can change, and whether you will be given services or will have to create everything yourself. Relationships between the infrastructure and your users are also defined by cloud deployment types.

Different types of cloud computing deployment models are:

1. Public cloud
2. Private cloud
3. Hybrid cloud
4. Community cloud
5. Multi-cloud

**1. Public Cloud**

The public cloud makes it possible for anybody to access systems and services. The public cloud may be less secure as it is open for everyone. The public cloud is one in which cloud infrastructure services are provided over the internet to the general people or major industry groups. The infrastructure in this cloud model is owned by the entity that delivers the cloud services, not by the consumer. It is a type of cloud hosting that allows customers and users to easily access systems and services. This form of cloud computing is an excellent example of cloud hosting, in which service providers supply services to a variety of customers. In this arrangement, storage backup and retrieval services are given for free, as a subscription, or on a per-use basis. Example: Google App Engine etc.

**Advantages of the public cloud model:**

* **Minimal Investment:**Because it is a pay-per-use service, there is no substantial upfront fee, making it excellent for enterprises that require immediate access to resources.
* **No setup cost:** The entire infrastructure is fully subsidized by the cloud service providers, thus there is no need to set up any hardware.
* **Infrastructure Management is not required:**Using the public cloud does not necessitate infrastructure management.
* **No maintenance:**The maintenance work is done by the service provider (Not users).
* **Dynamic Scalability:** To fulfill your company’s needs, on-demand resources are accessible.

**2. Private Cloud**

The private cloud deployment model is the exact opposite of the public cloud deployment model. It’s a one-on-one environment for a single user (customer). There is no need to share your hardware with anyone else. The distinction between private and public cloud is in how you handle all of the hardware. It is also called the “internal cloud” & it refers to the ability to access systems and services within a given border or organization. The cloud platform is implemented in a cloud-based secure environment that is protected by powerful firewalls and under the supervision of an organization’s IT department.   
The private cloud gives the greater flexibility of control over cloud resources.

**Advantages of the private cloud model:**

* **Better Control:**You are the sole owner of the property. You gain complete command over service integration, IT operations, policies, and user behavior.
* **Data Security and Privacy:** It’s suitable for storing corporate information to which only authorized staff have access. By segmenting resources within the same infrastructure, improved access and security can be achieved.
* **Supports Legacy Systems:** This approach is designed to work with legacy systems that are unable to access the public cloud.
* **Customization:**Unlike a public cloud deployment, a private cloud allows a company to tailor its solution to meet its specific needs.

**3.  Hybrid cloud**

By bridging the public and private worlds with a layer of proprietary software, hybrid cloud computing gives the best of both worlds. With a hybrid solution, you may host the app in a safe environment while taking advantage of the public cloud’s cost savings. Organizations can move data and applications between different clouds using a combination of two or more cloud deployment methods, depending on their needs.

**Advantages of the hybrid cloud model:**

* **Flexibility and control:**Businesses with more flexibility can design personalized solutions that meet their particular needs.
* **Cost:** Because public clouds provide for scalability, you’ll only be responsible for paying for the extra capacity if you require it.
* **Security:**Because data is properly separated, the chances of data theft by attackers are considerably reduced.

**4. Community cloud**

It allows systems and services to be accessible by a group of organizations. It is a distributed system that is created by integrating the services of different clouds to address the specific needs of a community, industry, or business. The infrastructure of the community could be shared between the organization which has shared concerns or tasks. It is generally managed by a third party or by the combination of one or more organizations in the community.

**Advantages of the community cloud model:**

* **Cost Effective:**It is cost-effectivebecause the cloud is shared by multiple organizations or communities.
* **Security:** Community cloud provides better security.
* **Shared resources:**It allows you to share resources, infrastructure, etc. with multiple organizations.
* **Collaboration and data sharing:**It is suitable for both collaboration and data sharing.

**5. Multi-cloud**

We’re talking about employing multiple cloud providers at the same time under this paradigm, as the name implies. It’s similar to the hybrid cloud deployment approach, which combines public and private cloud resources. Instead of merging private and public clouds, multi-cloud uses many public clouds. Although public cloud providers provide numerous tools to improve the reliability of their services, mishaps still occur. It’s quite rare that two distinct clouds would have an incident at the same moment. As a result, multi-cloud deployment improves the high availability of your services even more.

**Advantages of a multi-cloud model:**

* You can mix and match the best features of each cloud provider’s services to suit the demands of your apps, workloads, and business by choosing different cloud providers.
* **Reduced Latency:** To reduce latency and improve user experience, you can choose cloud regions and zones that are close to your clients.
* **High availability of service:** It’s quite rare that two distinct clouds would have an incident at the same moment. So, the multi-cloud deployment improves the high availability of your services.

**Why Pay Attention to Scalability?**

Sometimes scalability is erroneously used as a synonym for growth. But the true definition of scalability has to do with meeting demand. Demand can change seasonally, weekly, and hourly. In a real-world IT environment, demand isn’t steady. Even a thriving business might encounter times when there is more or less demand.

In a data center world, reducing capacity was almost never practical, so companies were left provisioning enough resources to cover their expected peak demand. In other words, an eCommerce site would need enough computing resources to handle Black Friday levels of traffic every single day. Utilization rates were very low, especially because most companies would provision resources based on expected peak demand, plus some.

The alternative is to provision just enough resources for daily use and not for peak traffic. Yet the consequences of not having enough compute or storage resources are dire. First come performance issues, then users start getting error messages and getting locked out of the application. In a business setting, that equals lost revenue. Conversely, resources are not free. Over-provisioning can lead to ballooning IT costs.

#### What are the differences between horizontal and vertical scaling in the cloud?

**Horizontal scaling** refers to provisioning additional servers to meet your needs, often splitting workloads between servers to limit the number of requests any individual server is getting. Horizontal scaling in cloud computing means adding additional instances instead of moving to a larger instance size.

**Vertical scaling** refers to adding more or faster CPUs, memory, or I/O resources to an existing server, or replacing one server with a more powerful server. In a data center, administrators traditionally achieved vertical scaling by purchasing a new, more powerful server and discarding or repurposing the old one. Today’s cloud architects can accomplish [AWS](https://cloudcheckr.com/partners/aws/) vertical scaling and [Microsoft Azure](https://cloudcheckr.com/partners/azure-cloud/) vertical scaling by changing instance sizes. AWS and Azure cloud services have many different instance sizes, so vertical scaling in cloud computing is possible for everything from EC2 instances to RDS databases.

### Horizontal vs. Vertical Scaling Pros and Cons

#### Pros and cons of horizontal scaling:

**Pros:** Horizontal scaling is much easier to accomplish without downtime. Horizontal scaling is also easier than vertical scaling to manage automatically. Limiting the number of requests any instance gets at one time is good for performance, no matter how large the instance. Provisioning additional instances also means having greater redundancy in the rare event of an outage.

**Cons:** Depending on the number of instances you need, your costs may be higher. Additionally, without a load balancer in place, your machines run the risk of being over-utilized, which could lead to an outage. However, with public cloud platforms, you can pay attention to discounts for Reserved Instances (RIs) if you’re able to predict when you require more compute power. Following cloud [cost management best practices](https://cloudcheckr.com/solutions/cloud-cost-optimization/) can help you efficiently scale in or out.

#### Pros and cons of vertical scaling:

**Pros:** In the cloud, vertical scaling means changing the sizes of cloud resources, rather than purchasing more, to match them to the workload. This process is known as right sizing. For example, [right sizing in AWS](https://aws.amazon.com/aws-cost-management/aws-cost-optimization/right-sizing/) can refer to the CPU, memory, storage, and networking capacity of instances and storage classes. Right sizing is one of the most effective ways to control cloud costs. When done correctly, right sizing can help lower costs of vertically scaled resources.

**Cons:** In general, vertical scaling can cost more. Why is vertical scaling expensive? When resources aren’t right sized correctly — or at all — costs can skyrocket. There’s also downtime to consider. Even in a cloud environment, scaling vertically usually requires making an application unavailable for some amount of time. Therefore, environments or applications that can’t have downtime would typically benefit more from horizontal scalability by provisioning additional resources instead of increasing capacity for existing resources.

### Which Is Better: Horizontal or Vertical Scaling?

The decision to scale horizontally or vertically in the cloud depends upon the requirements of your data. Remember that scaling continues to be a challenge, even in cloud environments. All parts of your application need to scale, from the compute resources to database and storage resources. Neglecting any pieces of the scaling puzzle can lead to [unplanned downtime](https://cloudcheckr.com/webinar/five-strategies-for-preventing-cloud-downtime/) or worse. The best solution might be a combination of vertical scaling in order to find the ideal capacity of each instance and then horizontal scaling to handle spikes in demand, while ensuring uptime.