



## Lesson 2 Reading:

# "Cloud computing 101: The interrelationship of scalability, reliability, and availability".

Taken from: <https://www.lucidchart.com/blog/reliability-availability-in-cloud-computing#:~:text=You%20will%20learn%20that%3A,from%20anywhere%20in%20the%20world.>




While researching reasons to migrate to the cloud, you've probably learned that the benefits include scalability, reliability, availability, and more. But what, exactly, do those terms mean?



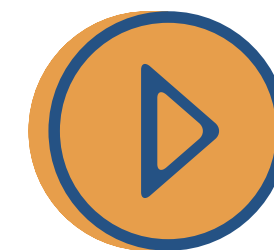

This article focuses on cloud computing scalability, cloud computing reliability, and cloud computing availability. You will learn that:

- You need cloud scalability to meet customer demand.
- You need reliability in cloud computing to ensure that your products and services work as expected.
- You need cloud availability to ensure that customers can access your cloud services whenever they need to and from anywhere in the world.
- You need to bring all three together to achieve true high availability.

# What is scalability in cloud computing?



Cloud computing scalability refers to how well your system can react and adapt to changing demands. As your company grows, you want to be able to seamlessly add resources without losing quality of service or interruptions. As demand on your resources decreases, you want to be able to quickly and efficiently downscale your system so you don't continue to pay for resources you don't need.



However, there is more to scalability in the cloud than simply adding or removing resources as needed. Let's look at some of the different types of scalability in cloud computing.





# Cloud elasticity

This refers to how well your cloud services are able to add and remove resources on demand. Elasticity is important because you want to ensure that your clients and employees have access to the right amount of resources as needed.

Cloud elasticity should be automatic and seamless. People accessing your cloud services should not be able to notice that resources are added or dropped. They should just have the confidence that they can access and use resources without interruptions.



# Vertical scaling

Vertical scaling (or “scaling up”) refers to upgrading a single resource. For example, installing more memory or storage capacity to a server. In a physical, on-premises setup, you would need to shut down the server to install the updates.



# Horizontal scaling

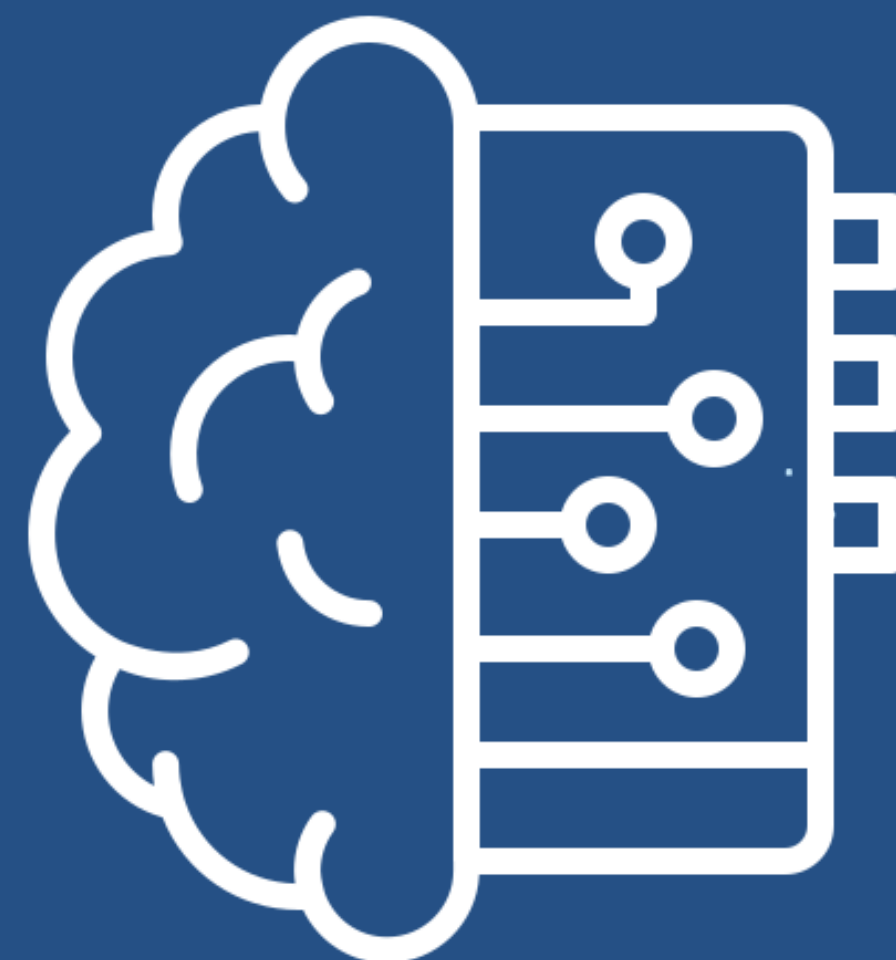
This term is used to describe “building out” a system with additional components. For example, you can add processing power or more memory to a server by linking it with other servers. Horizontal scaling is a good practice for cloud computing because additional hardware resources can be added to the linked servers with minimal impact. These additional resources can be used to provide redundancy and ensure that your services remain reliable and available.






# Auto-scaling

This term refers to a cloud computing feature that lets you automatically manage the different types of cloud scalability automatically. Cloud providers such as Amazon Web Services offer auto-scaling to enable consistent performance regardless of the current demand on resources.



# Implementing and managing a cloud scaling strategy is:

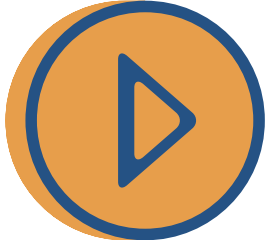
## Convenient: Cost-effective: Time-saving:



You can easily increase or decrease storage capacity as needed.

You don't have to pay for expensive hardware or provide the space to store it.

Upgrading existing hardware and installing new hardware on-site can be very time-consuming.



## Flexible and fast:

You can quickly respond to changing demands to keep customers up and running without delays in service.

## Fault-tolerant:

Resources can automatically be scaled to accommodate redundancies and to facilitate disaster recovery.

Cloud computing can take care of the scaling for you. This frees you up to focus on innovation and process improvement rather than troubleshooting errors and other issues.


Cloud computing is so scalable because the cloud service providers have the necessary hardware and software in place. They also use virtual machines (VMs) to scale up or down because:

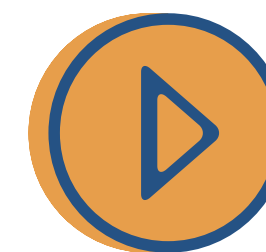
- You can easily add resources to VMs at any time with minimal impact.
- You can easily move VMs to a different server that has more resources.
- You can host VMs on a server cluster to share resources and balance the load.



# What is reliability in cloud computing?

When you access an app or service in the cloud, you can reasonably expect that:

- 
- The app or service is up and running.
  - You can access what you need from any device at any time from any location.
  - There will be no interruptions or downtime.
  - Your connection is secure.
  - You will be able to perform the tasks you need to get your job done.







Factors like these measure the reliability of your cloud offerings. In a perfect world, your system would be 100% reliable. But that is probably not an attainable goal. In the real world, things will go wrong. You will see faults from things such as server downtime, software failure, security breaches, user errors, and other unexpected incidents.





Proper planning and cloud visualization can help you address faults quickly so that they don't become huge problems that keep people from accessing your cloud offerings. The cloud makes it easy to build fault-tolerance into your infrastructure. You can easily add extra resources and allocate them for redundancy.

## Employing measures that make your cloud system more reliable ensures that:

- Redundant resources kick in automatically when the system experiences a fault.
- There is no downtime and products and services remain available.
- Employees keep doing their jobs without knowing that something went wrong.





Reliability in cloud computing is important for businesses of any size. Buggy software can cause lost productivity, lost revenue, and lost trust in your brand. Before you deploy your applications to the cloud, make sure they are thoroughly tested against a variety of real-world scenarios. This helps to ensure that they are reliable and will meet customer expectations.



## What is availability in cloud computing?

High availability is the ultimate goal of moving to the cloud. The idea is to make your products, services, and tools available to your customers and employees at any time from anywhere using any device with an internet connection.



Cloud availability is related to cloud reliability.





For example, let's say you have an online store that is available 24/7. But sometimes clicking the "checkout" button kicks customers out of the system before they have completed the purchase. So, your store may be available all the time, but if the underlying software is not reliable, your cloud offerings are basically useless.



## Bringing it all together

Cloud availability, cloud reliability, and cloud scalability all need to come together to achieve high availability. This means that your products and services are accessible anytime and anywhere, function reliably and as expected, and that the system can seamlessly scale up or down to accommodate customer demand without suffering a loss in performance.






Cloud service providers offer an Infrastructure as a Service (IaaS) model that gives you access to storage, servers, and other resources. IaaS provides automation and scalability on demand so that you can spend your time managing and monitoring your applications, data, and other services.




Because IaaS provides scalability based on a pay-as-you-go model, this saves you money and frees you up to track down and address problems that may come up with the software. Having more time to monitor can help you find areas that need improvement so you can do a better job consistently deploying reliable products and services.





To survive in today's global market, it's inevitable that your company will need to move to the cloud. It won't happen overnight and will require a lot of planning. As you plan what and how you will make solutions available in the cloud, remember that it is important that your products and services and cloud infrastructure are scalable, reliable, and available when and where they are needed.





# Answer the multiple - choice questions based on the previous reading text.

Para consolidar las respuestas ingrese al cuestionario online.

INICIO