What Is High Availability?

In this lesson, we will learn about high availability and its importance in online services.

We'll cover the following

Highly available computing infrastructure is the norm in the computing industry today. More so, when it comes to the cloud platforms, it's the key feature which enables the workloads running on them to be highly available.

This lesson is an insight into *high availability*. It covers all the frequently asked questions about it such as:

- What is it?
- Why is it so important to businesses?
- What is a *highly available cluster*?
- How do cloud platforms ensure high availability of the services running on them?
- What is *fault tolerance* & *redundancy*? How are they related to high availability?

So, without any further ado. Let's get on with it.

What Is High Availability?

> High availability also known as *HA* is the ability of the system to stay online despite having failures at the infrastructural level in real-time.

High availability ensures the uptime of the service much more than the normal time. It improves the reliability of the system, ensures minimum downtime.

The sole mission of highly available systems is to stay online & stay connected. A very basic example of this is having back-up generators to ensure continuous power supply in case of any power outages.

In the industry, HA is often expressed as a percentage. For instance, when the system is *99.99999%* highly available, it simply means *99.99999%* of the total hosting time the service will be up. You might often see this in the *SLA* (Service Level Agreements) of cloud platforms.

How Important Is High Availability To Online Services?

It might not impact businesses that much if social applications go down for a bit & then bounce back. However, there are mission-critical systems like aircraft systems, spacecrafts, mining machines, hospital servers, finance stock market systems that just cannot afford to go down at any time. After all, lives depend on it.

The smooth functioning of the mission-critical systems relies on the continual connectivity with their network/servers. These are the instances when we just cannot do without super highly available infrastructures.

Besides no service likes to go down, critical or not.

To meet the high availability requirements systems are designed to be *fault-tolerant*, their components are made *redundant*.

What is fault-tolerant & redundancy in systems designing? I’ll discuss up next;