Replication

In this lesson, we will learn about Replication as a High Availability mechanism.

We'll cover the following

Replication – Active-Active HA Mode

Replication means having a number of similar nodes running the workload together. There are no standby or passive instances. When a single or a few nodes go down, the remaining nodes bear the load of the service. Think of this as load balancing.

This approach is also known as the *Active-Active High Availability* mode. In this approach, all the components of the system are active at any point in

time.

Geographical Distribution of Workload

As a contingency for natural disasters, data centre regional power outages & other big-scale failures, workloads are spread across different data centres across the world in different geographical zones.

This avoids the single point of failure thing in context to a data centre. Also, the latency is reduced by quite an extent due to the proximity of data to the user.

All the highly available fault-tolerant design decisions are subjective to how critical the system is? What are the odds that the components will fail? Etc.

Businesses often use multi-cloud platforms to deploy their workloads which ensures further availability. If things go south with one cloud provider, they have another to fail back over.