Replication

This lesson discusses replication as a high availability mechanism.

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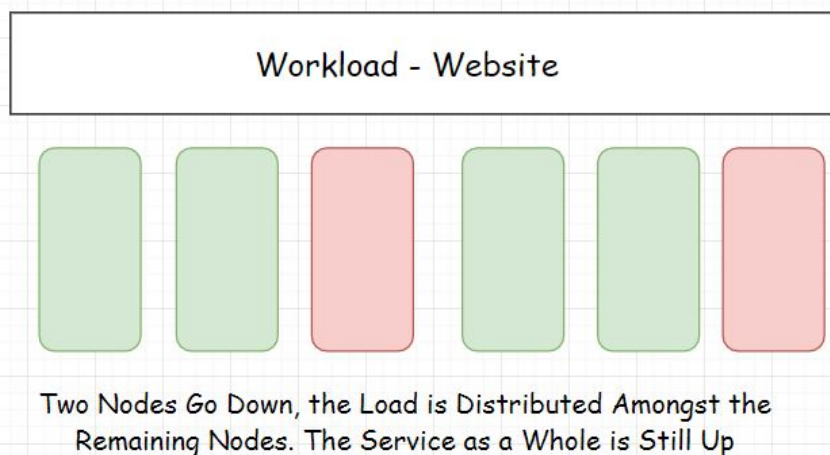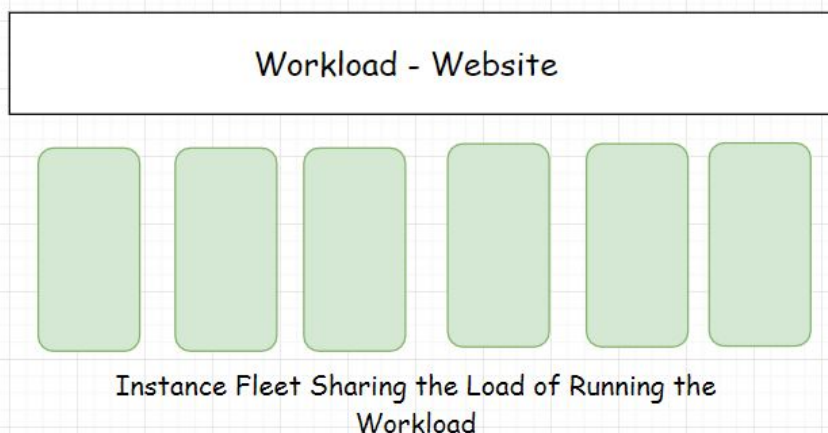
This approach is also known as the *active-active high availability* mode. In this approach, all the components of the system are active at any point in time

approach, all the components of the system are active at any point in time.

Geographical distribution of workload

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

This avoids the single point of failure with respect to a data center. Also, 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?.

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

In the next lesson, you'll take a quiz on clustering.