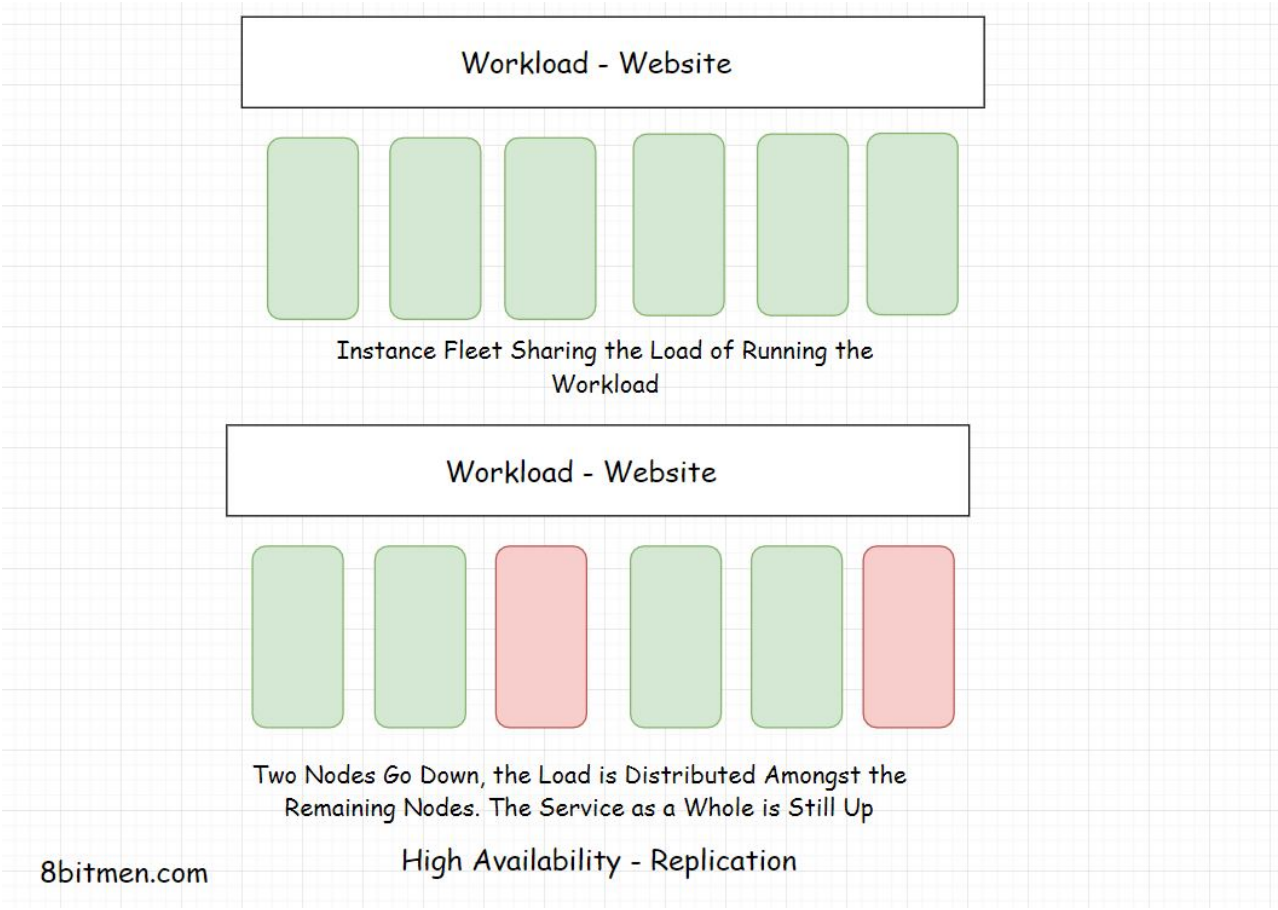
Replication

In this lesson, you 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, regional power outages, and other big-scale failures, data center workloads are spread across different data centers across the world in different geographical zones.

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

All highly available fault-tolerant design decisions are subjective to how critical the system is. Designers must ask “what are the odds that the components will fail”?

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.

← Back

Redundancy

Next →

High Availability Clustering

Mark as Completed

Report an Issue