



Autoscaling tip

When a scenario requires say, "2 as the minimum capacity", it is always best to construct an architecture that gives more than 2 instances (i.e. at least 3 instances) at all times because if an AZ outage happened, ASG will launch a new instance on the unaffected AZ.

This provisioning does not happen instantly, which means that for a certain period of time, there will only be 1 running instance left.

For example:

If a company needs to deploy at least 2 EC2 instances to support the normal workloads of its application and automatically scale up to 6 EC2 instances to handle the peak load and the architecture is processing mission-critical workloads.

Then, as the Cloud Architect of the company, create an architecture that has at least 3 instances up and running at all times.