Create an Auto Scaling Group using the launch configuration created in the previous step. Select the VPC and the subnets in which you want the ASG to launch instances. Specify the load balancer target group (and hence the load balancer) to which your ASG needs to be attached. You can choose not to associate your ASG with any load balancer as well -> might be useful for non-web apps.

Then specify the scaling policy. For example, here we will have 2 EC2 instances launched but when average CPU utilization goes beyond 50%, max of another 2 instances could be launched. We have not enabled scale-in protection, which means when CPU utilization goes down, some of the EC2 instances could be terminated as well.













