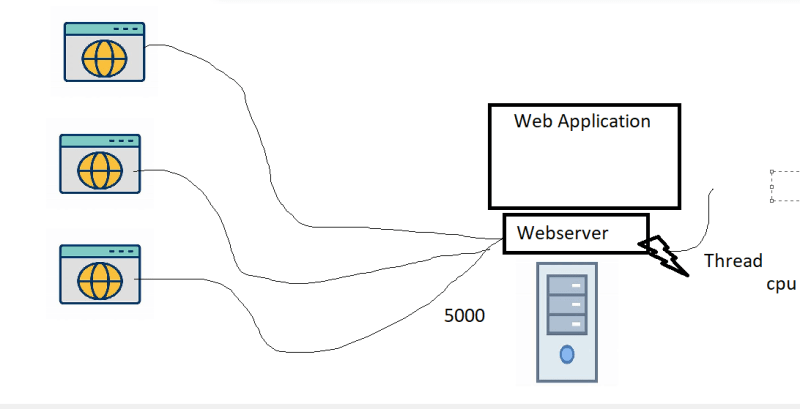
**AWS Auto Scaling Group**

Impact of parallel connections to a server. As users create parallel/concurrent connections (session) the load on cpu increases.  


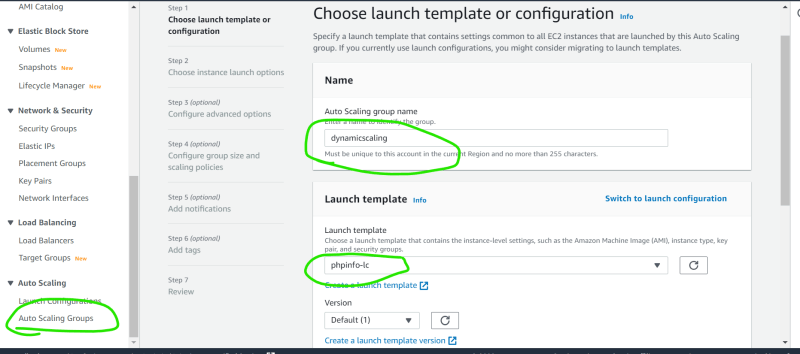
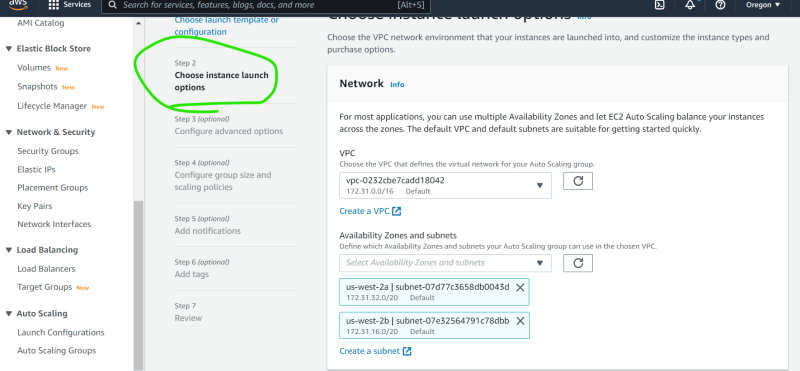
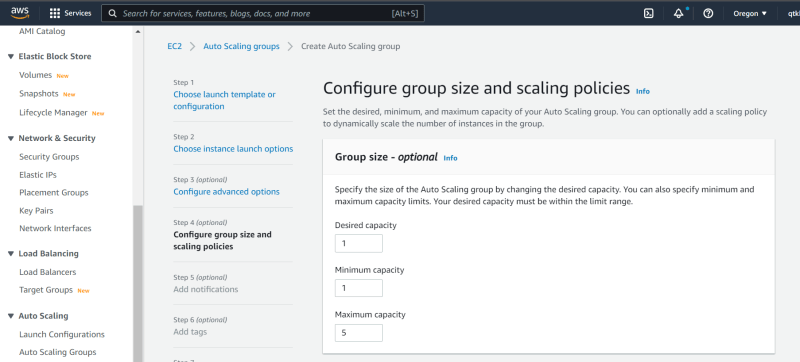
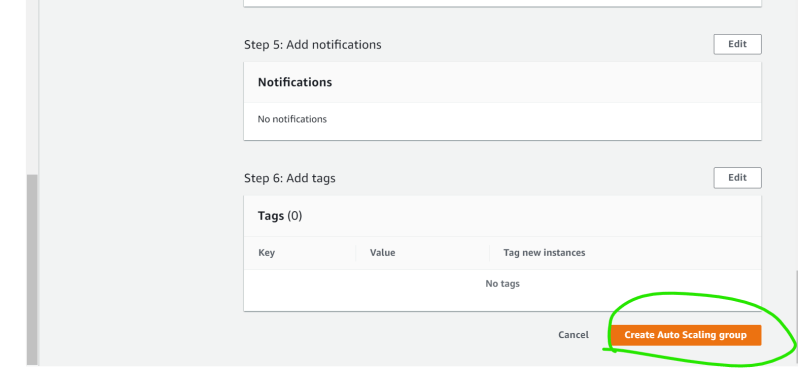
In this let’s try doing this,

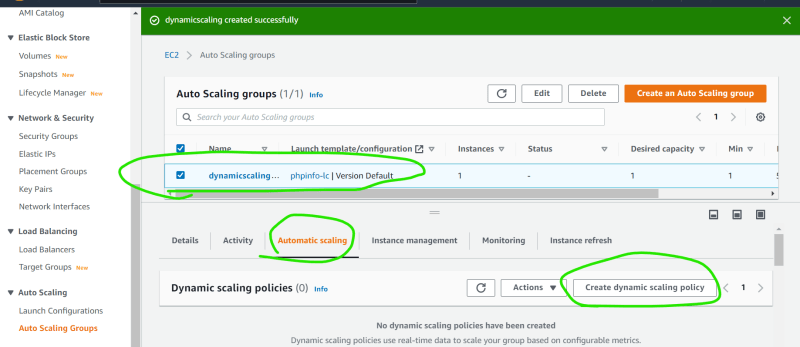
increase number of ec2 instances by 1 when average cpu utilization for last 5 minutes > 70

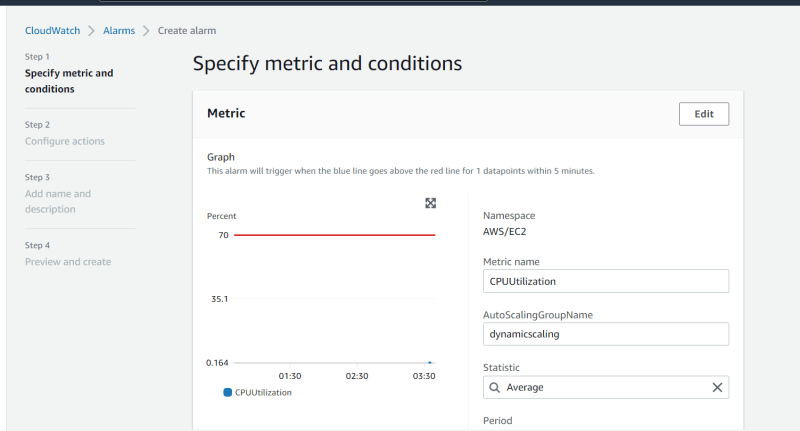
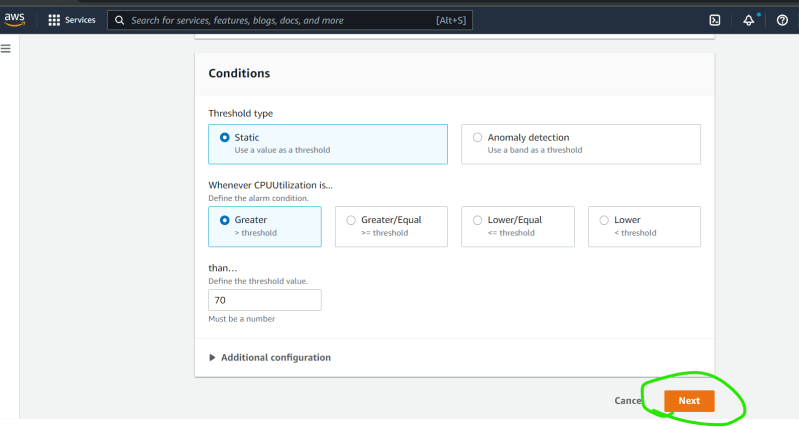
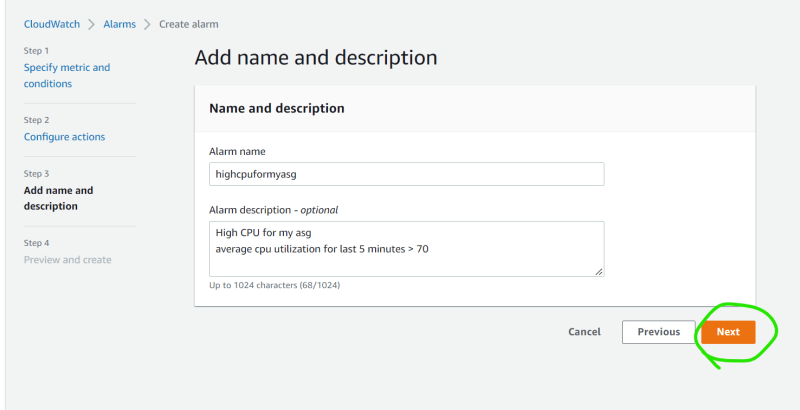
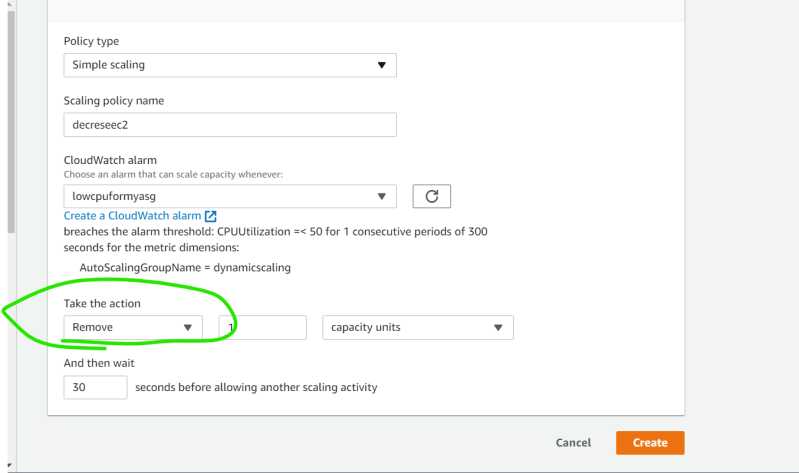
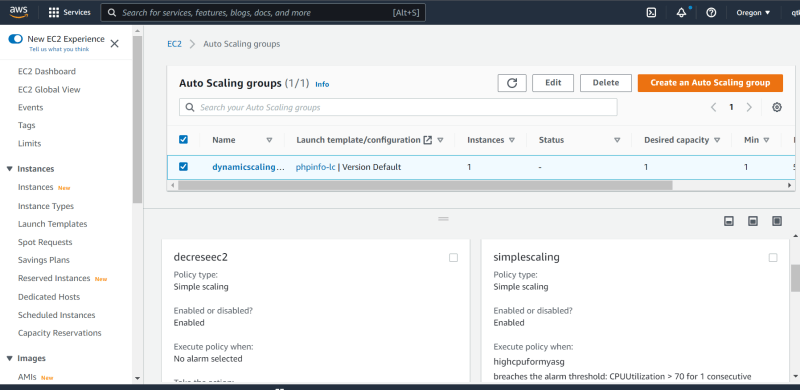
decrease number of ec2 instances by 1 when average cpu utilization for last 5 minutes < 50

minimum instances = 1, max instances = 5

To create artificial stress on the ec2 instance we will be using stress tool.

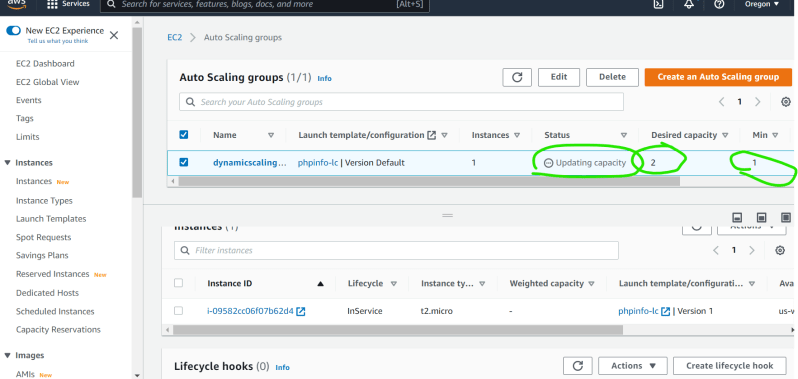
Steps followed:  
  
  
  


Wait for the auto scaling group to be created with no scaling policy. Now lets create a dynamic scaling policy  


Now Select Metric and Create Cloudwatch Alarm  
  
  
  
  


Now lets generate artifical load. Login into ec2 instance

stress --cpu 8 --io 4 --vm 2 --vm-bytes 128M --timeout 60m -v

Now observe for the changes in desired capacity in the period of 5-10 mins  


Now stop the stress tool (ctrl + c) and wait for 5-10 mins the desired capacity will be back to 1