Case Study

Problem Statement:

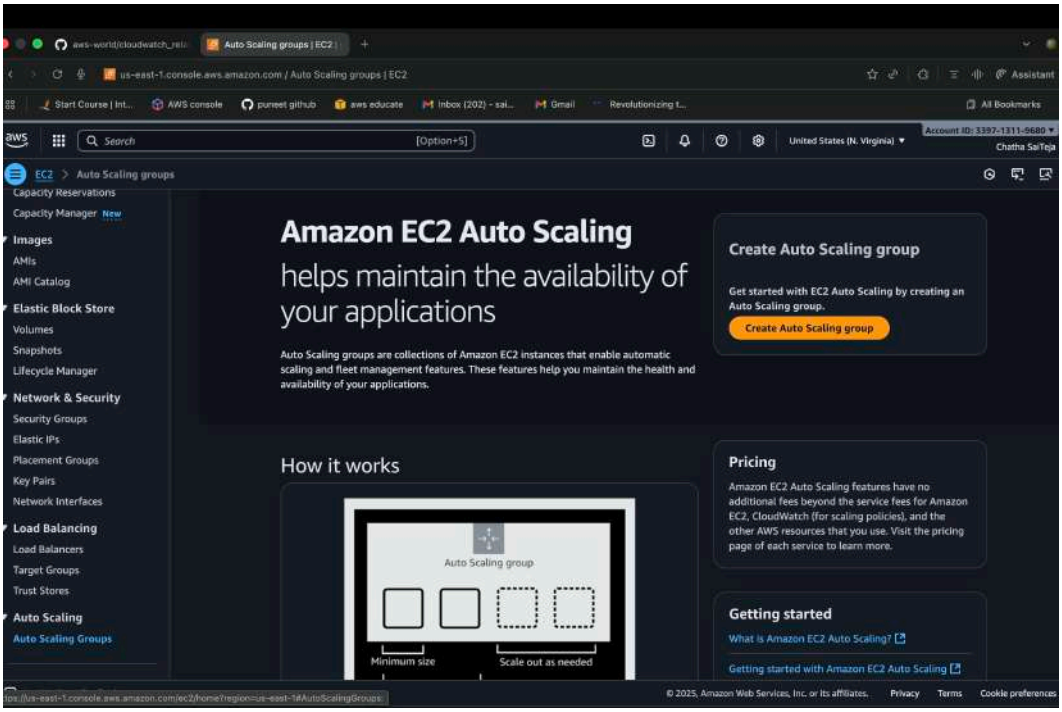
You work for XYZ Corporation that uses on premise solutions and a limited number of systems. With the increase in requests in their application, the load also increases. So, to handle the load the corporation has to buy more systems almost on a regular basis. realizing the need to cut down the expenses on systems, they decided to move their infrastructure to AWS.

Tasks To Be Performed:

1. Manage the scaling requirements of the company by:
    a. Deploying multiple compute resources on the cloud as soon as the load increases and the CPU utilization exceeds 80%
    b. Removing the resources when the CPU utilization goes under 60%
2. Create a load balancer to distribute the load between compute resources.
3. Route the traffic to the company's domain

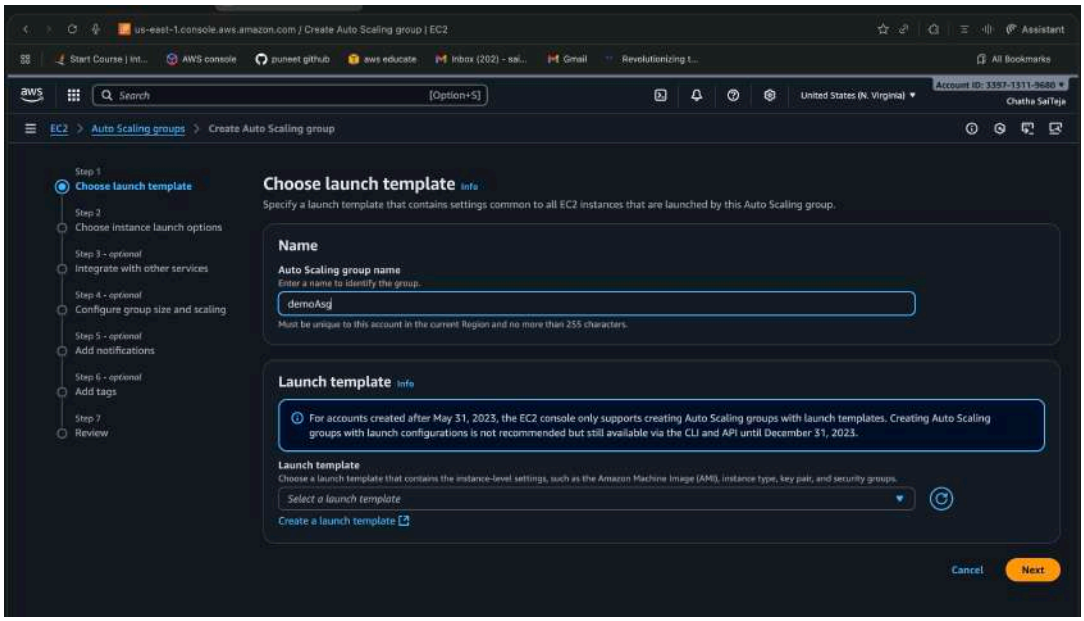
Step-by-Step Procedure:-

Step 1:- To scale the resources we have to use Auto Scaling Group (ASG) service in EC2 Dashboard ,then we can see ASG dashboard ,Click on Create Auto Scaling Group

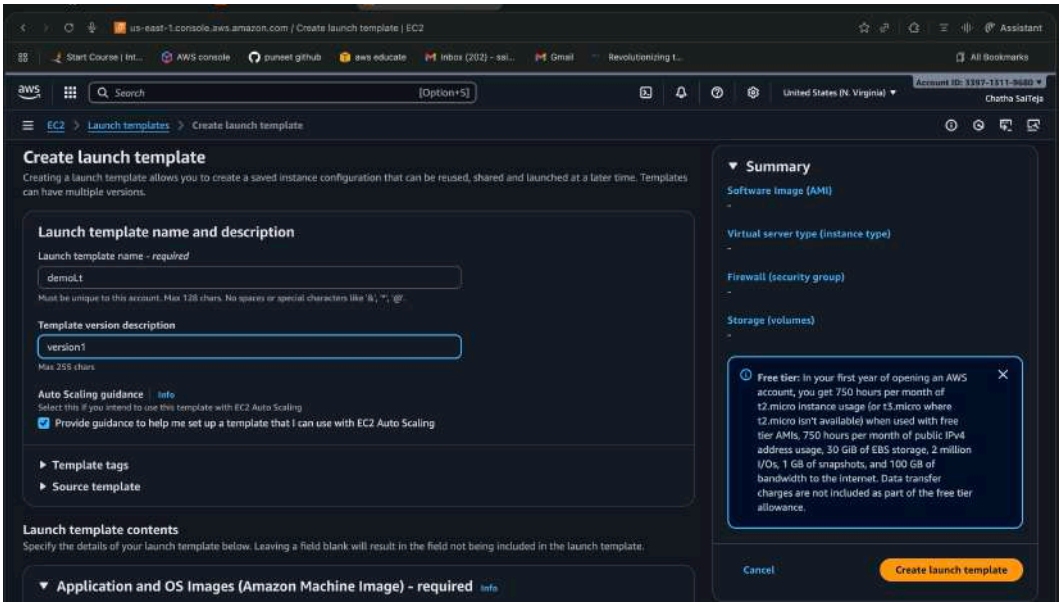


Click on Create Auto Scaling Group

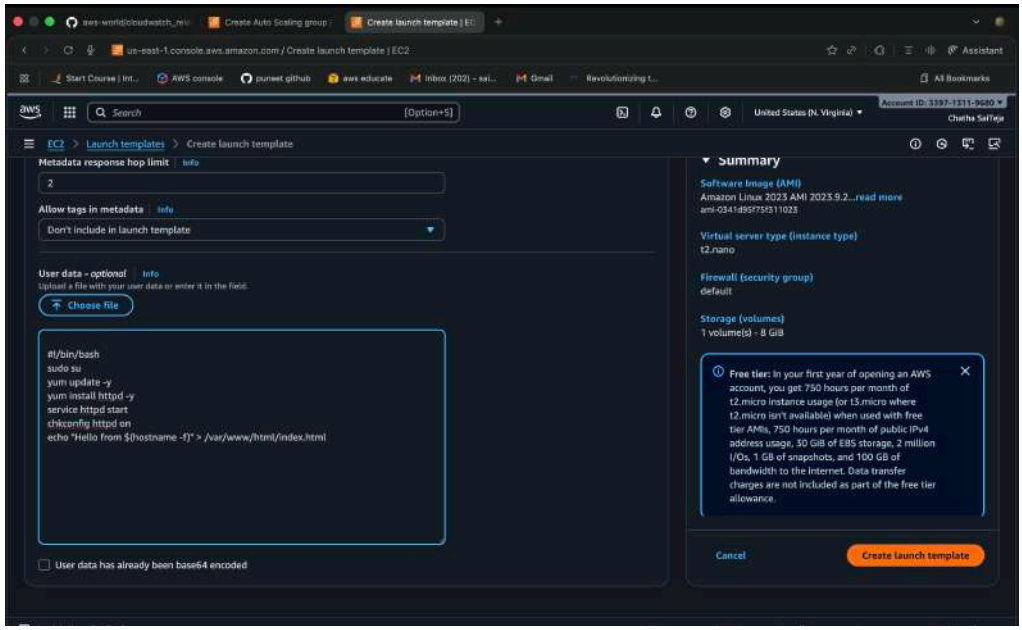
Step 2:-The first step in creating an Auto Scaling Group is to specify a name for the group. Next, you need to create a Launch Template, which defines the configuration of the EC2 instances that the ASG will launch during scale-out or scale-in events. In the Launch Template, you provide settings similar to those used when creating an EC2 instance such as the AMI, instance type, key pair, security groups, and storage configurations. After entering these details, click on Create Launch Template. Once the template is created successfully, select it to associate with your Auto Scaling Group.



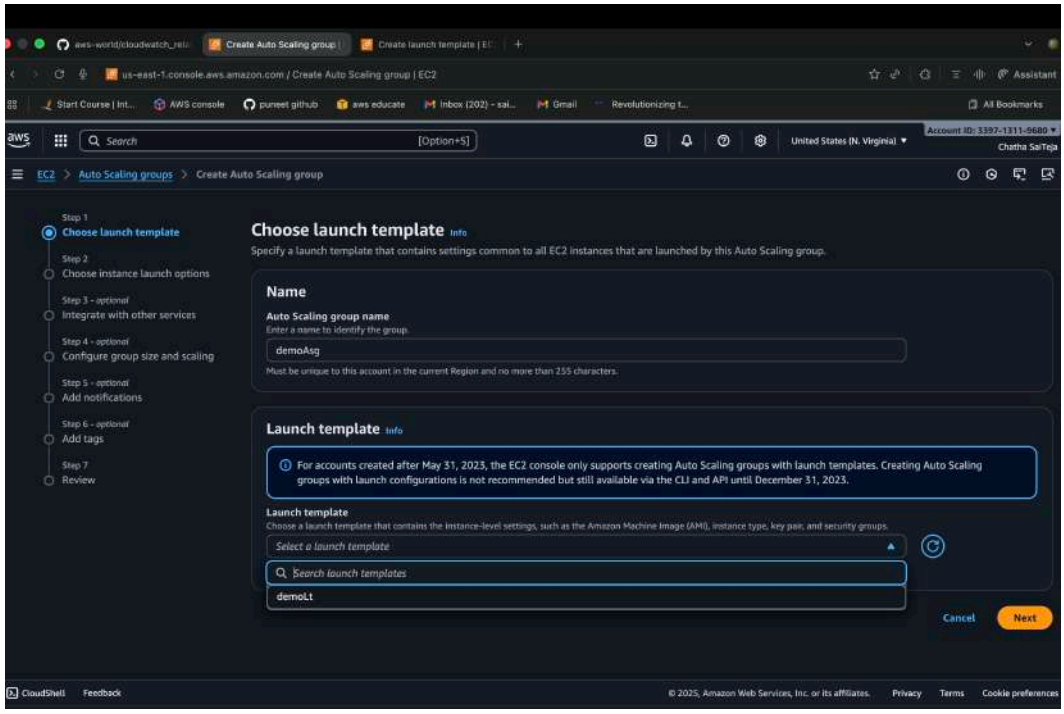
Specify the name(demoAsg) and click on create launch template



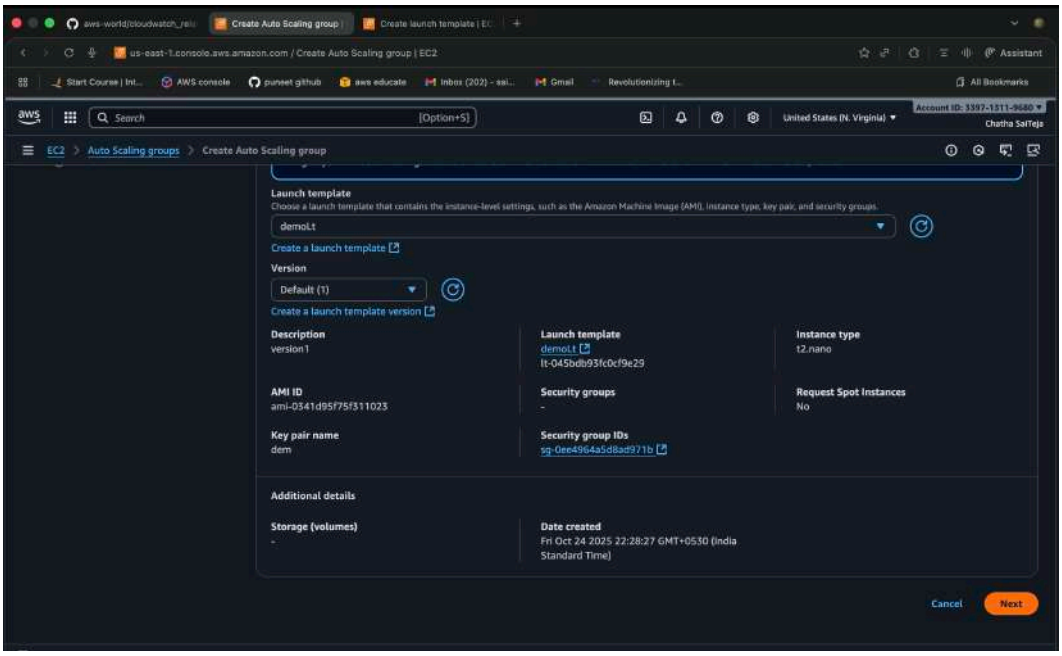
Specify the configurations and name(demoLt) of Launch Template



Click on create launch template

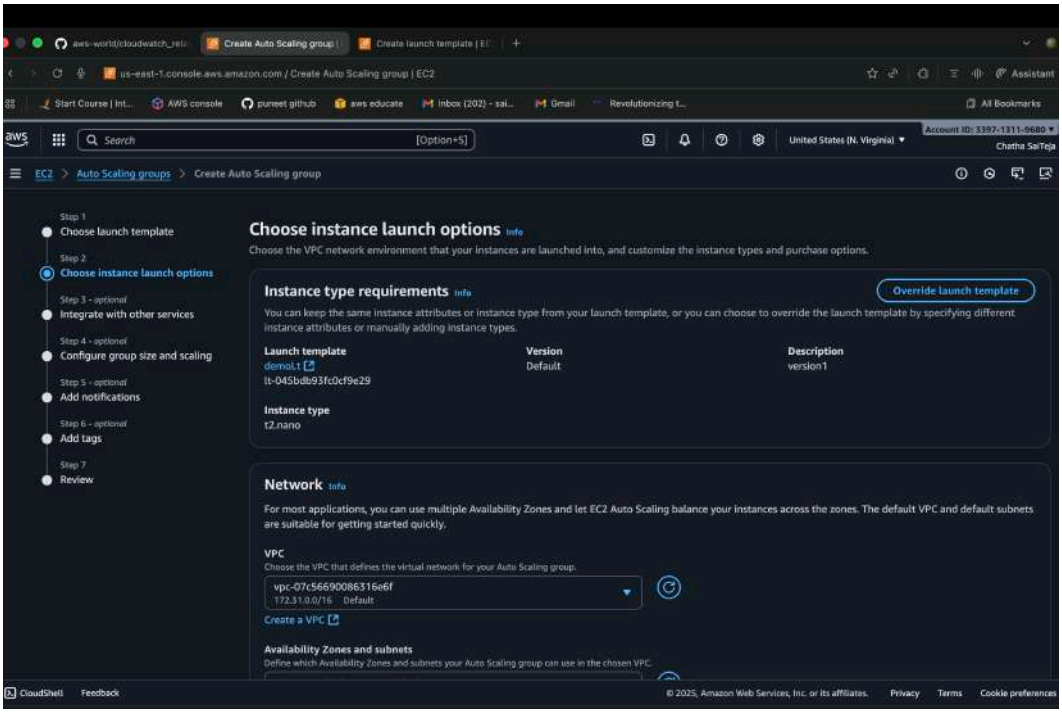


Select the Created Launch Template(demoLt)



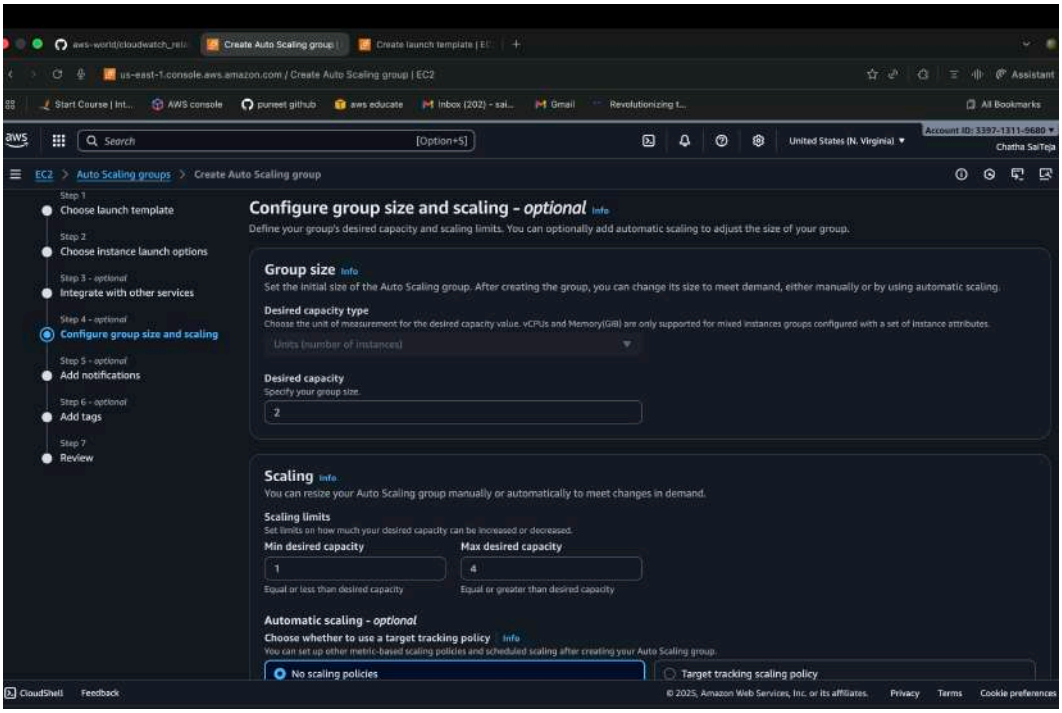
Click on Next

Step 3:- we have another steps like choosing instance options, Integrate with other services and configure group and scaling but these are optional so click on next .

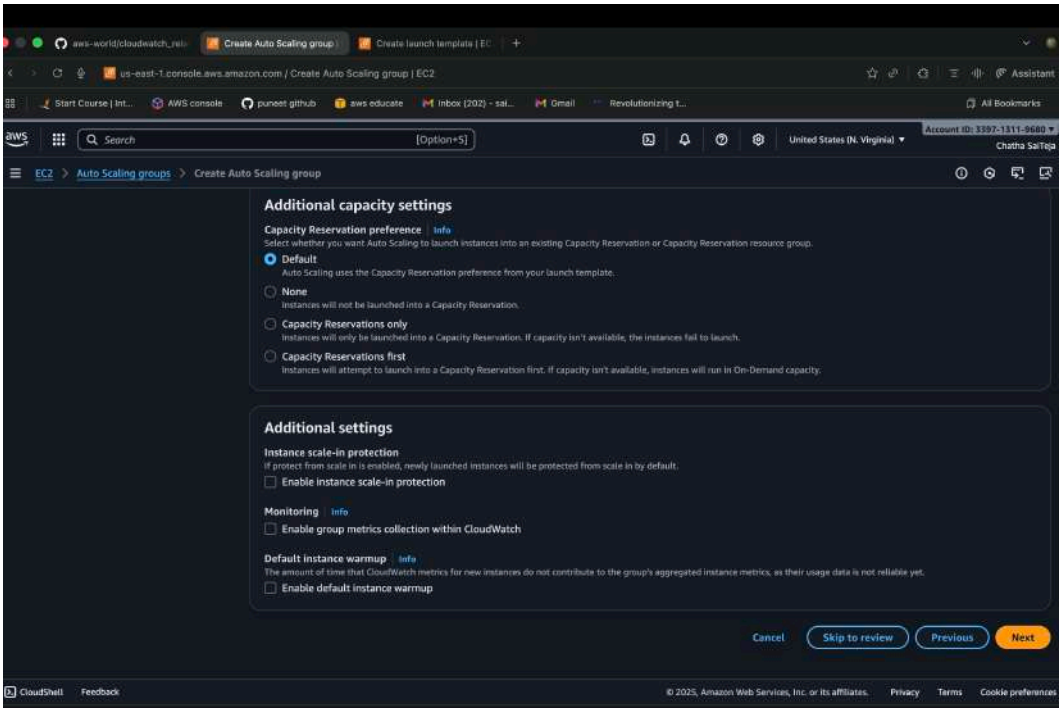


Click on next

Step 4:- In next step we have to mention the group size(desired-2, min-1, max-4) and we Can also mention the scaling policies but we don't give any policy.

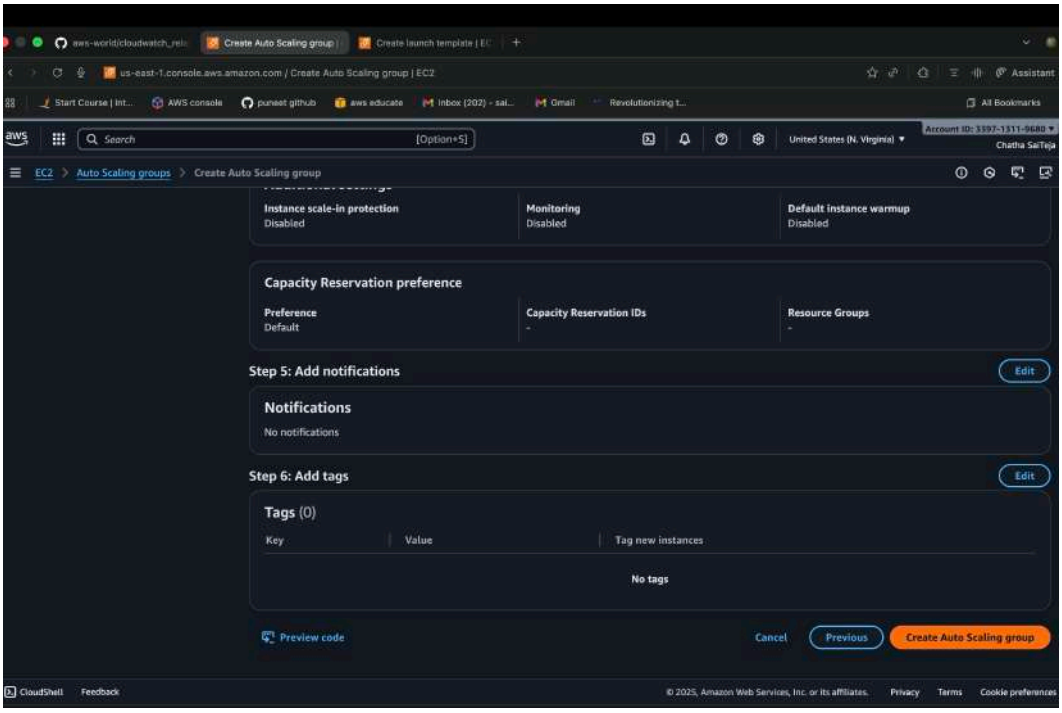


Giving group size

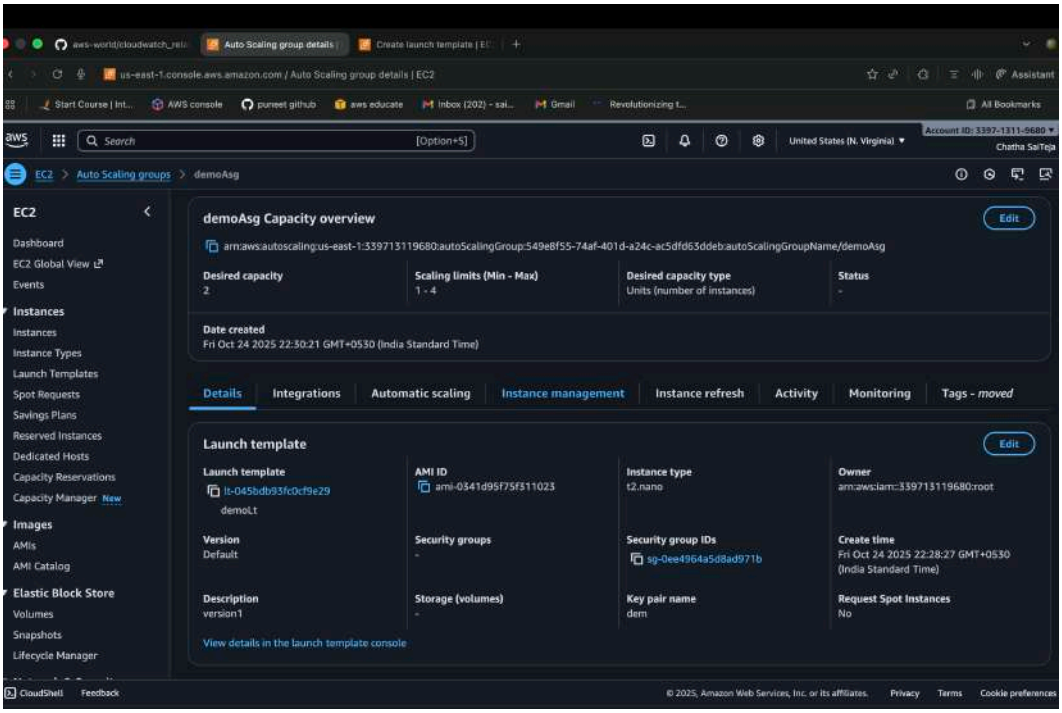


Click on next

Step 5:- In this Step we can review the total configuration of ASG and click on Create Auto Scaling Group.



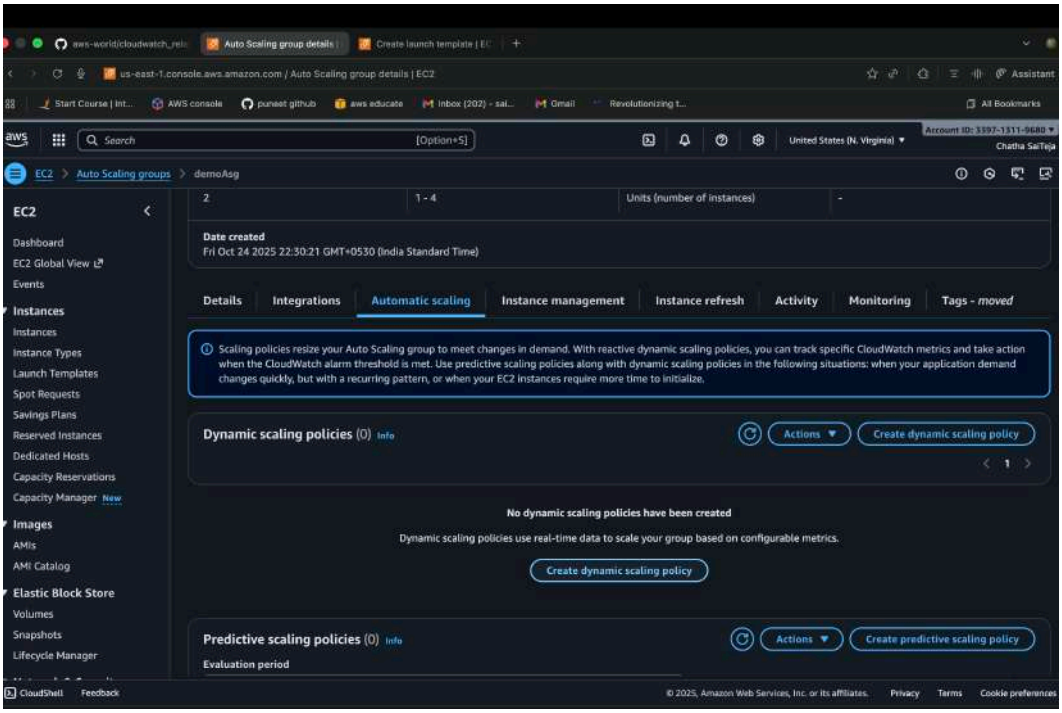
Click on Create Auto Scaling Group.



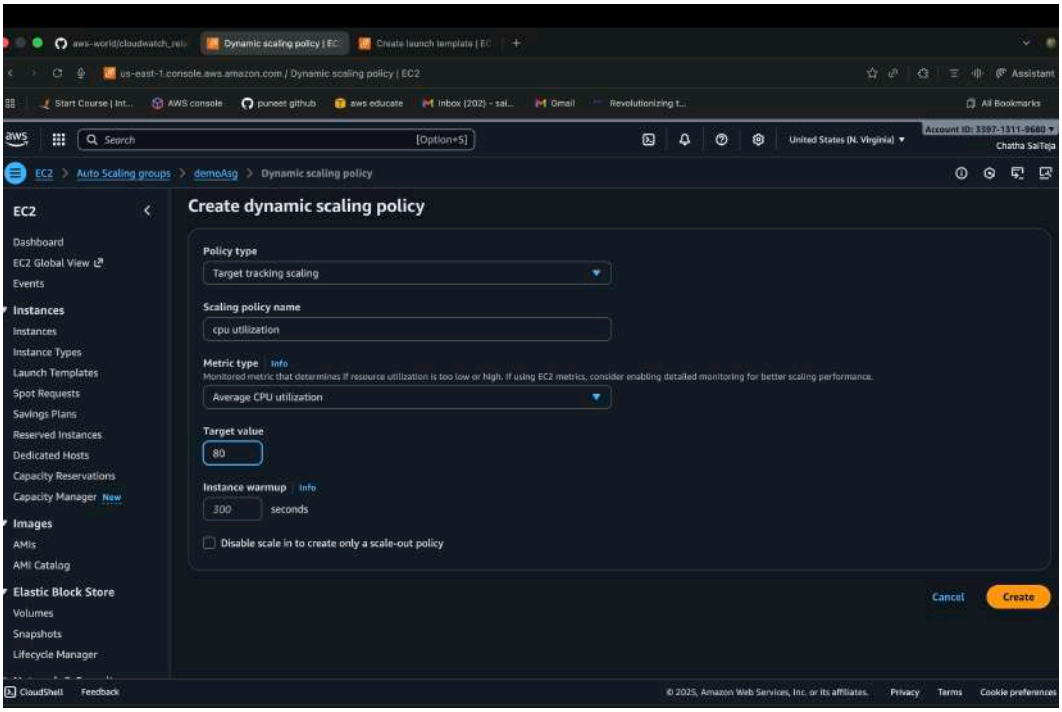
demoAsg Summary

Step 6:- To manage scaling requirements efficiently, Auto Scaling automatically adds or removes compute resources based on demand.you can configure it to add new instances when average Cpu utilization reaches 80% and terminate instances when utilization drops below 60%. This behaviour is implemented through Automatic Scaling, which supports three types of dynamic scaling policies: Simple, Step, and Target Tracking.

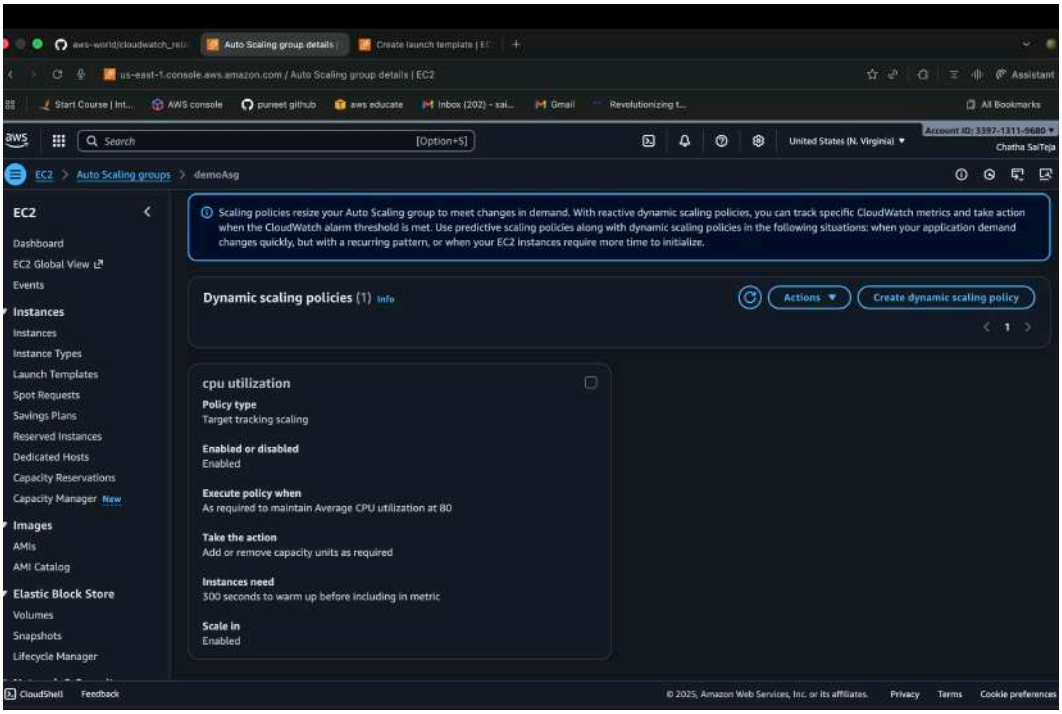
Using a Target Tracking policy, you simply specify a target metric for instance, an Average Cpu utilization of 80%. Auto Scaling then automatically adjusts the capacity of your group by launching or terminating instances to maintain this target level.



Click on Automatic Scaling and Select dynamic Scaling policy

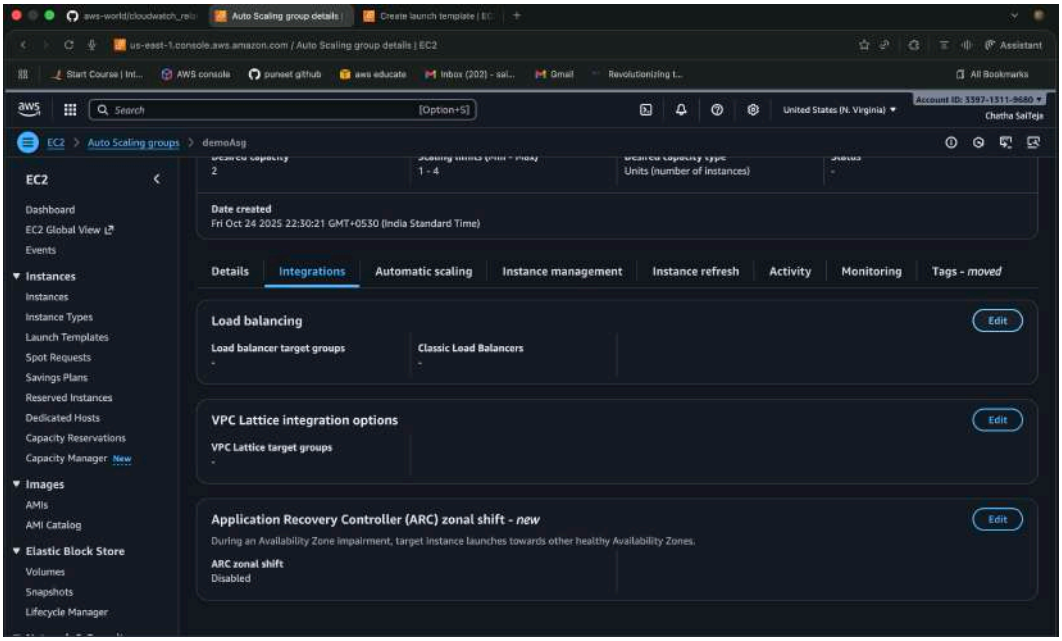


Specifying the Metric



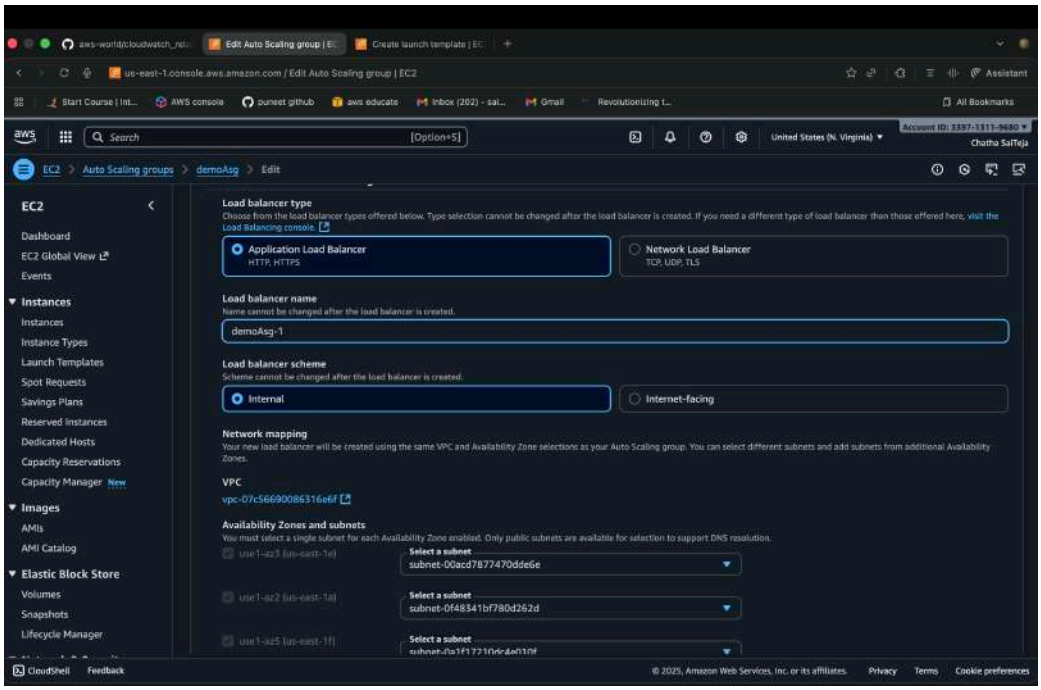
Target Tracking policy with 80% Avg cpu utilization metric

Step 7:- We Can Create a load balancer to distribute the load between the resources by selecting Integrations In the AutoScaling Group and Click on Edit.

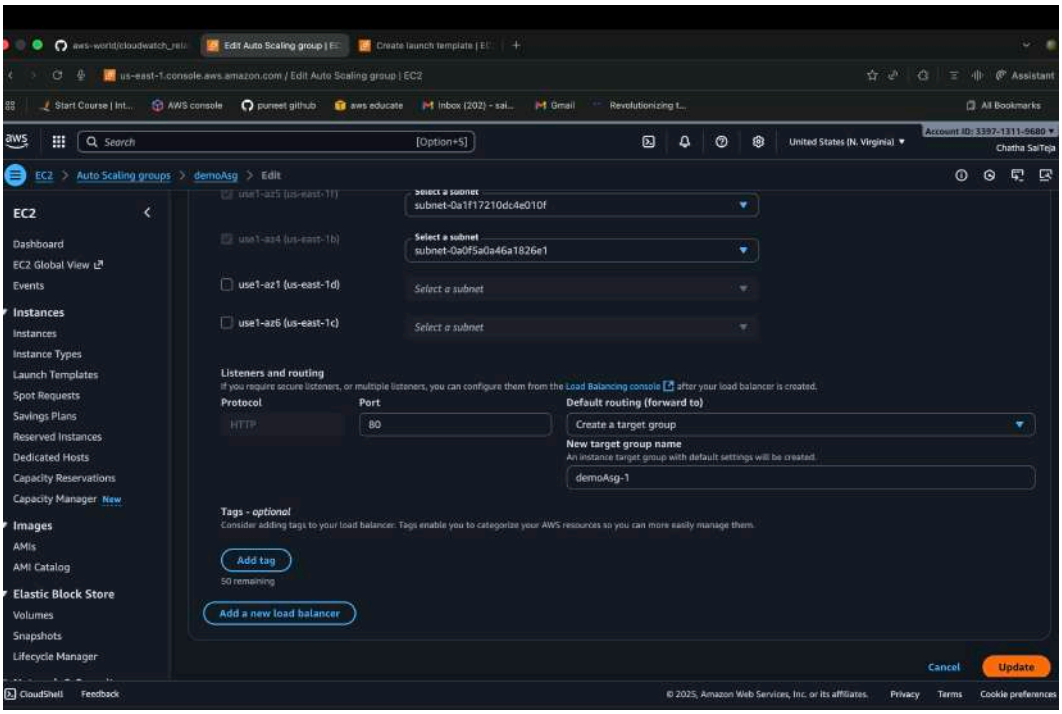


Click on Integrations

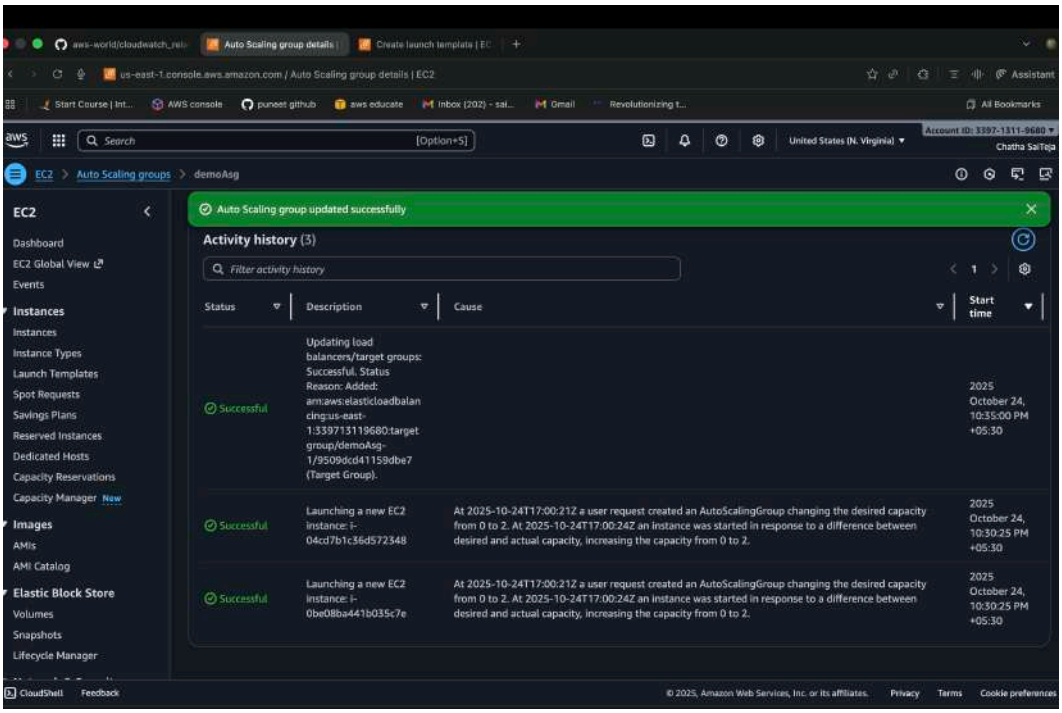
Step 8:- we select the load balancer type (ALB,NLB) and give the name to LB and create a target group it automatically creates a TG and then click on update then load balancer will be created.



Specify the name



Click on Update



Load balancer added successfully

Step 9:- The next one is we have to route the traffic to domain for that we have to buy a domain which is expensive so we ignoring the Route 53 task.