# Implementing Autoscalling Group with Load Balancer

## Introduction

In this project, I designed and implemented an Application Load Balancer (ALB) in conjunction with Auto Scaling Groups (ASGs) to efficiently manage traffic distribution and the dynamic scaling of EC2 instances.

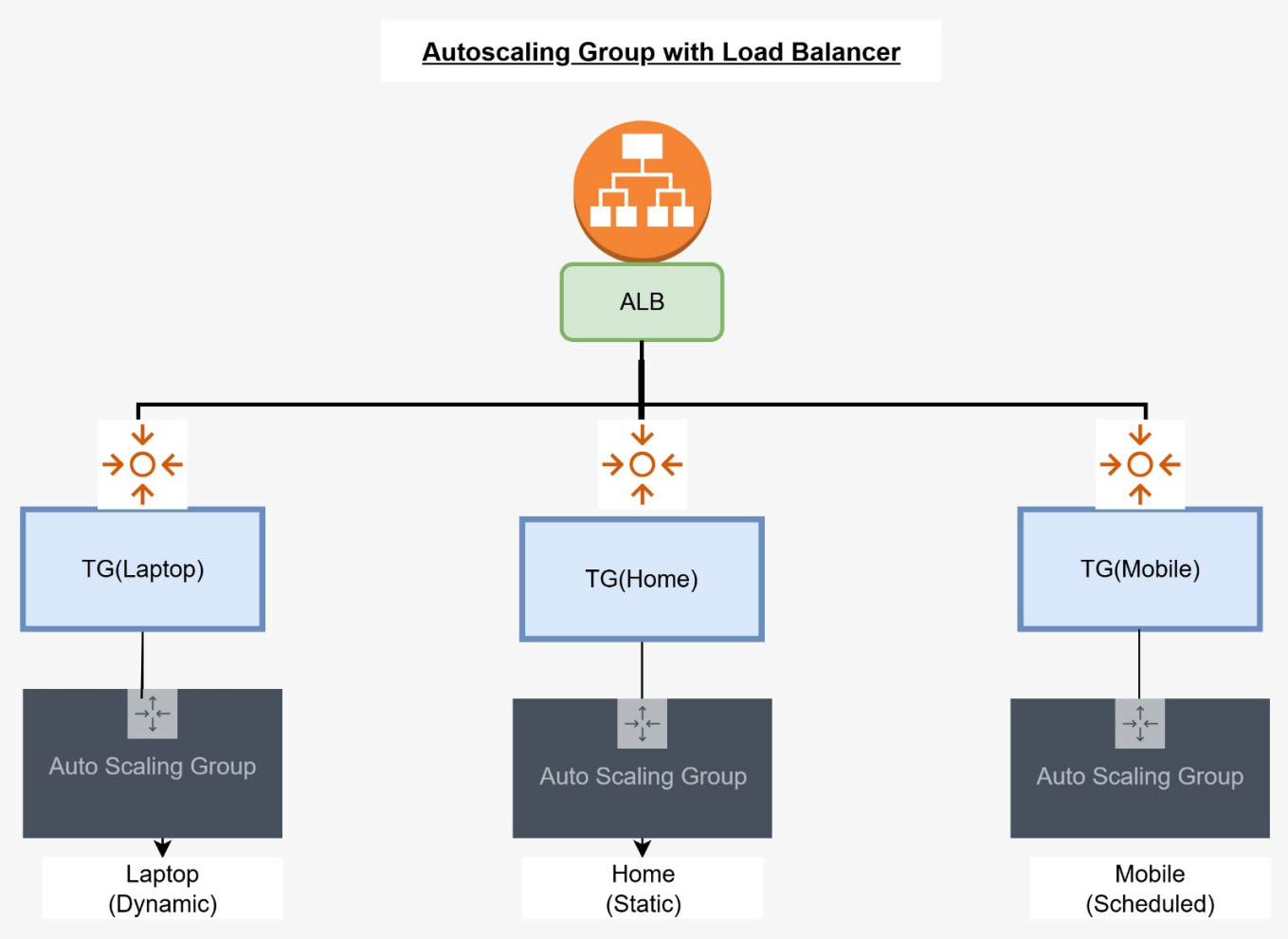
The architecture ensures high availability, fault tolerance, and scalability by routing incoming requests across multiple instances and automatically adjusting the number of instances based on real-time demand.

I configured three distinct Auto Scaling Groups to demonstrate different scaling strategies:

1. Home – a static scaling group with a fixed number of instances,
2. Laptop – a dynamic scaling group that reacts to usage metrics (e.g., CPU utilization) to scale in or out automatically.
3. Mobile – a scheduled scaling group that adjusts capacity based on predefined time-based schedules.

Each Auto Scaling Group is linked to its own Target Group, and all Target Groups are registered with the Application Load Balancer, enabling seamless request routing across the infrastructure.

## Deployment Architecture



# System Essentials

1. Application Load Balancer (ALB)

- Acts as the entry point for incoming traffic and intelligently distributes requests across multiple target groups.
- Enhances fault tolerance and ensures high availability by balancing the load among healthy instances.

2. Auto Scaling Groups (ASGs)

- To showcase various scaling techniques, I configured three distinct Auto Scaling Groups:
    - Home (Static Scaling): Maintains a fixed number of EC2 instances at all times.
    - Laptop (Dynamic Scaling): Automatically scales the number of instances up or down in response to real-time CPU utilization metrics.
    - Mobile (Scheduled Scaling): Adjusts the number of instances based on a predefined schedule (e.g., scaling out during peak traffic hours and scaling in during off-peak times).

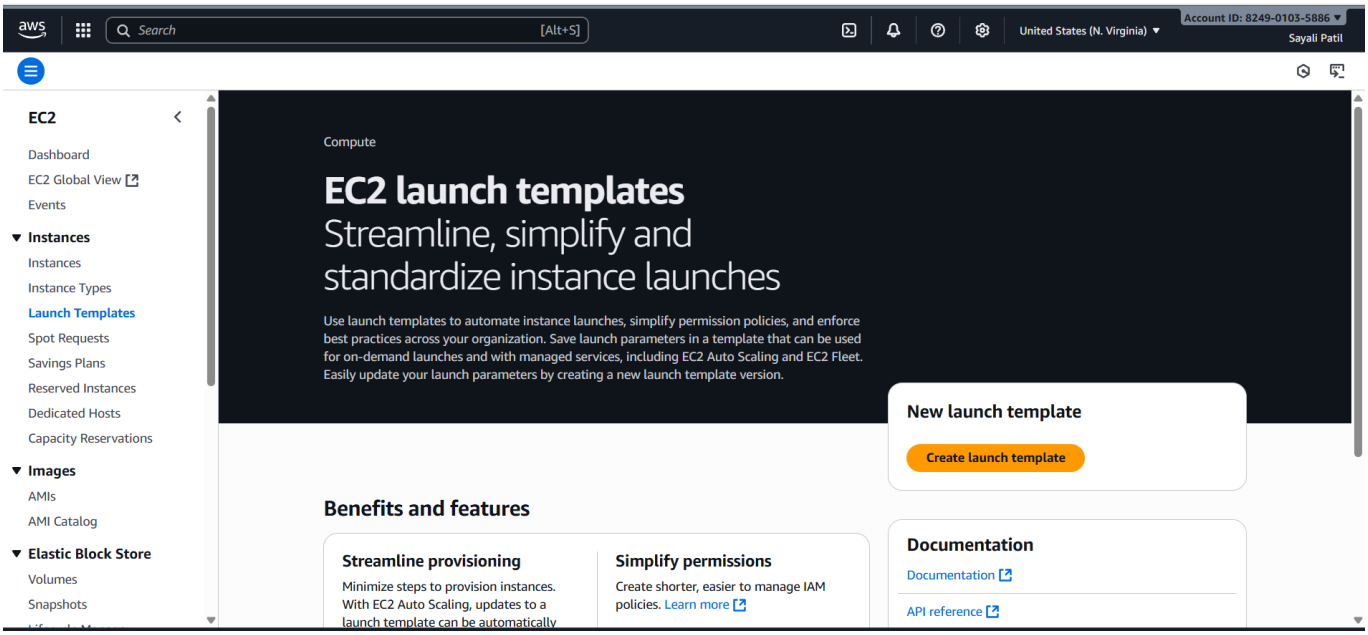
3. Target Groups

- Each ASG is linked to a dedicated target group.
- These target groups are integrated with the Application Load Balancer, enabling efficient routing of incoming requests to the appropriate set of instances.

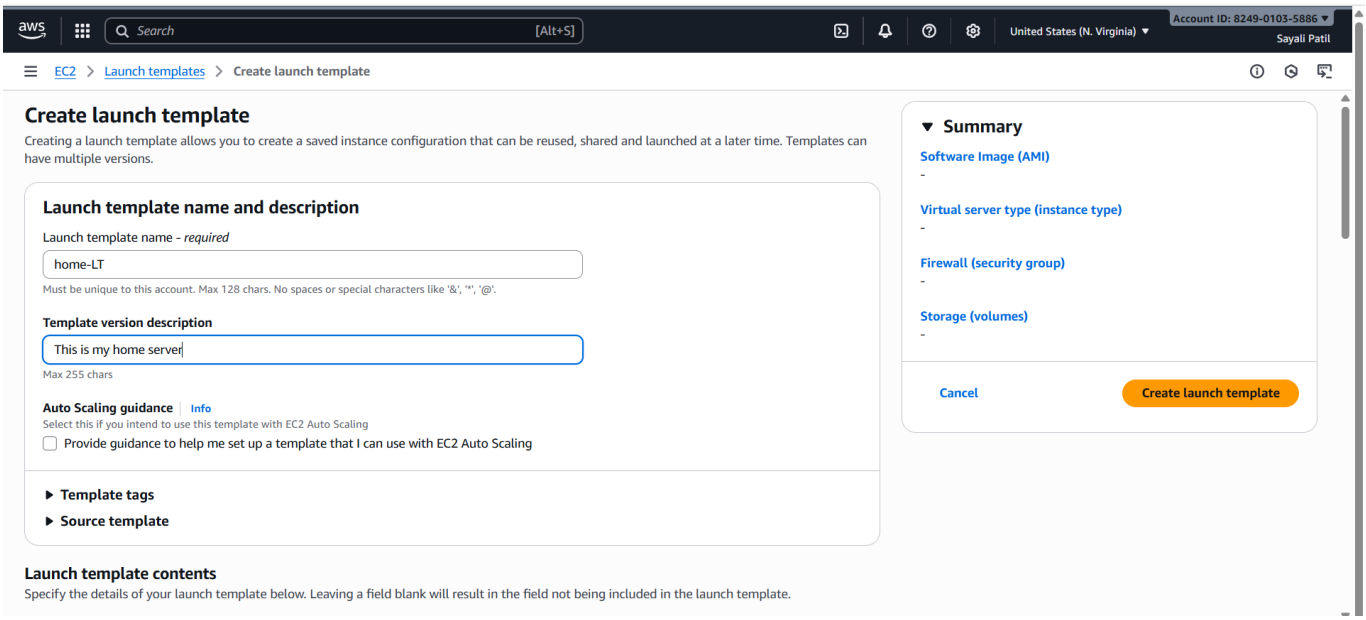
# Steps to Setup

## Step 1: Launch Template

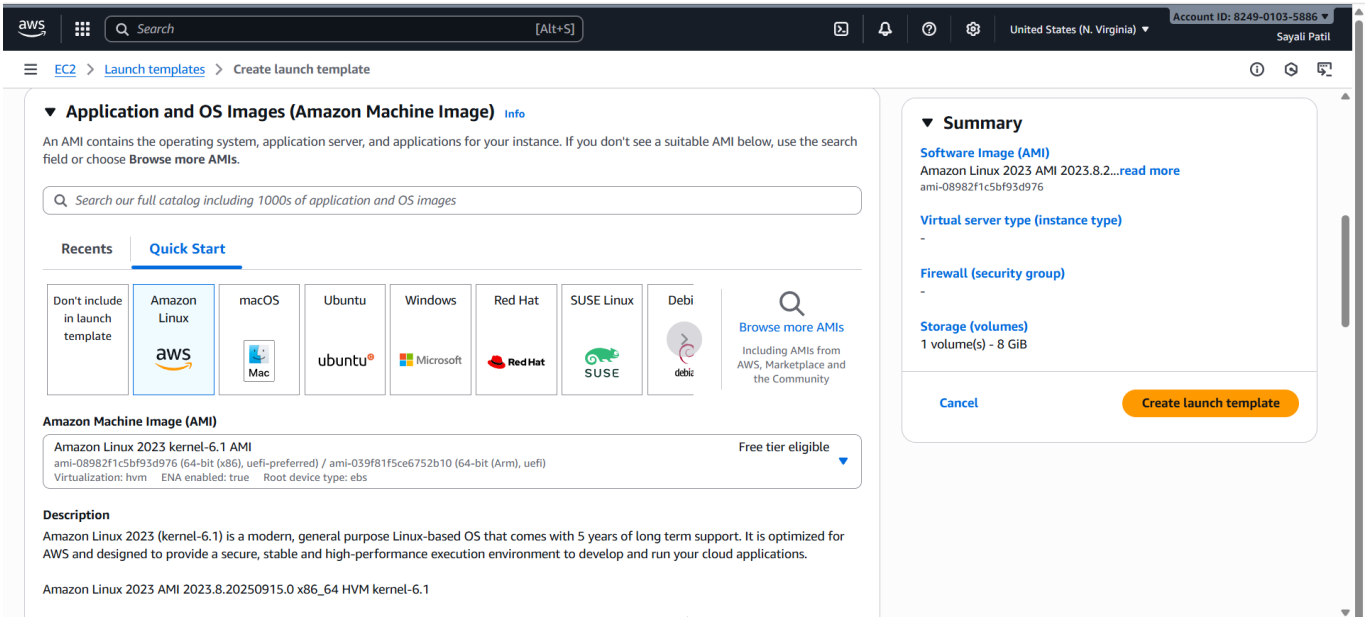
- Create 3 Launch Templates: home,laptop and mobile (each with unique User Data scripts)
- In an EC2 Instance, Navigate to the Launch Templates section in your AWS Management Console.



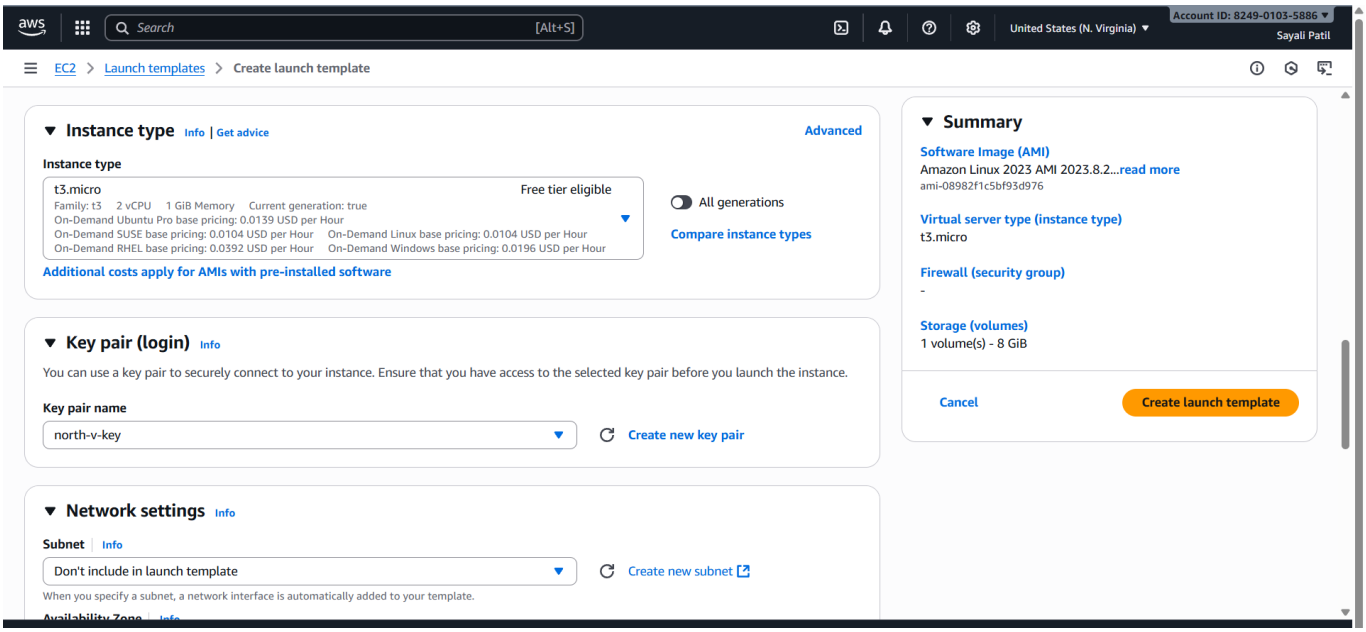
- Click on "Create Launch Template" to begin.
- Enter a Template Name (e.g., home, laptop, mobile)
- Write Disription in it.



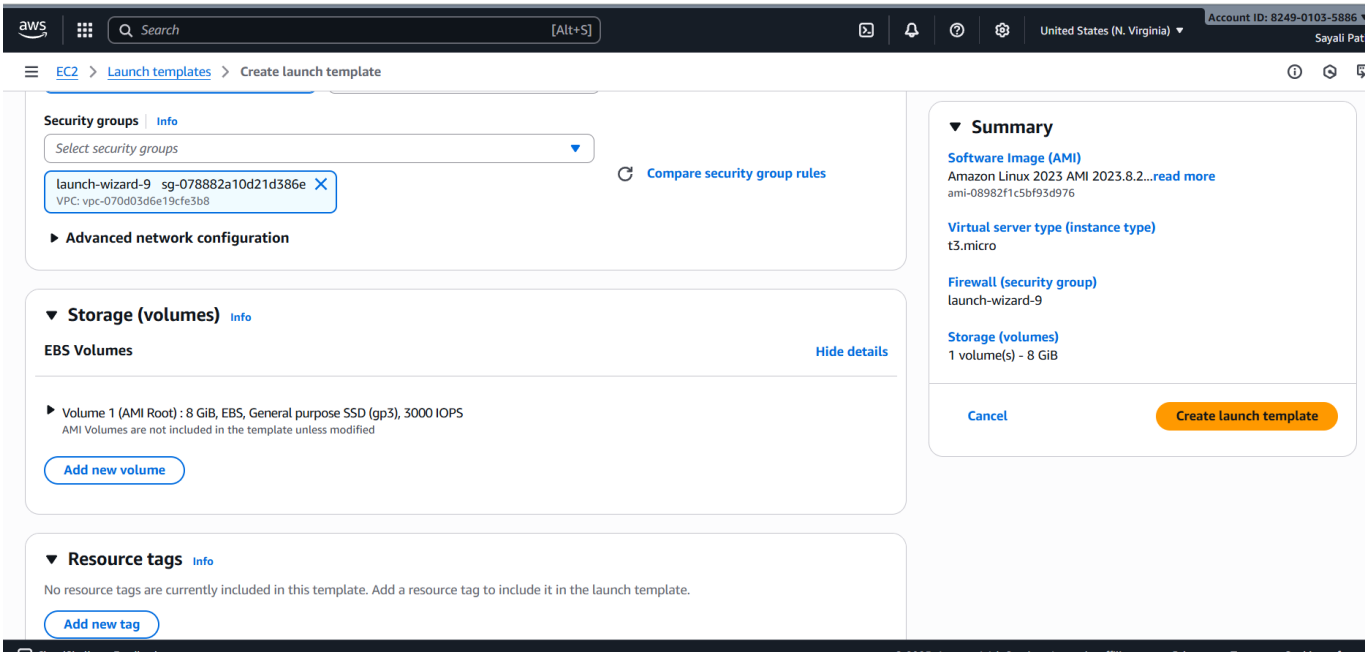

- Select an AMI from the Quick Start options (I selected Amazon Linux). Alternatively, you can use a custom AMI if you've previously created one.



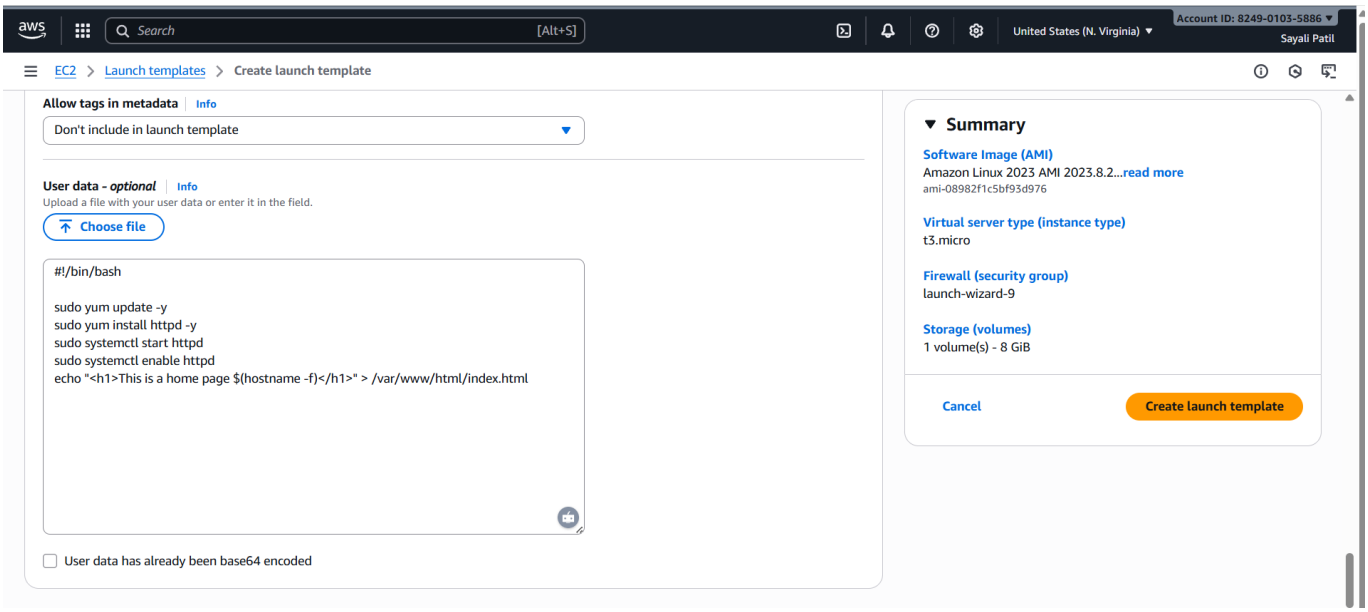

- Select a Instance type and Key Pair




- Configure the Network Settings: Since this is a basic application, only port 80 (HTTP) and port 22 (SSH) are required. You can either create a new security group with these rules or select an existing security group that already allows this traffic.




- Add user datascript in advanced options (you don't need to add this script if u hae selected your own AMI)



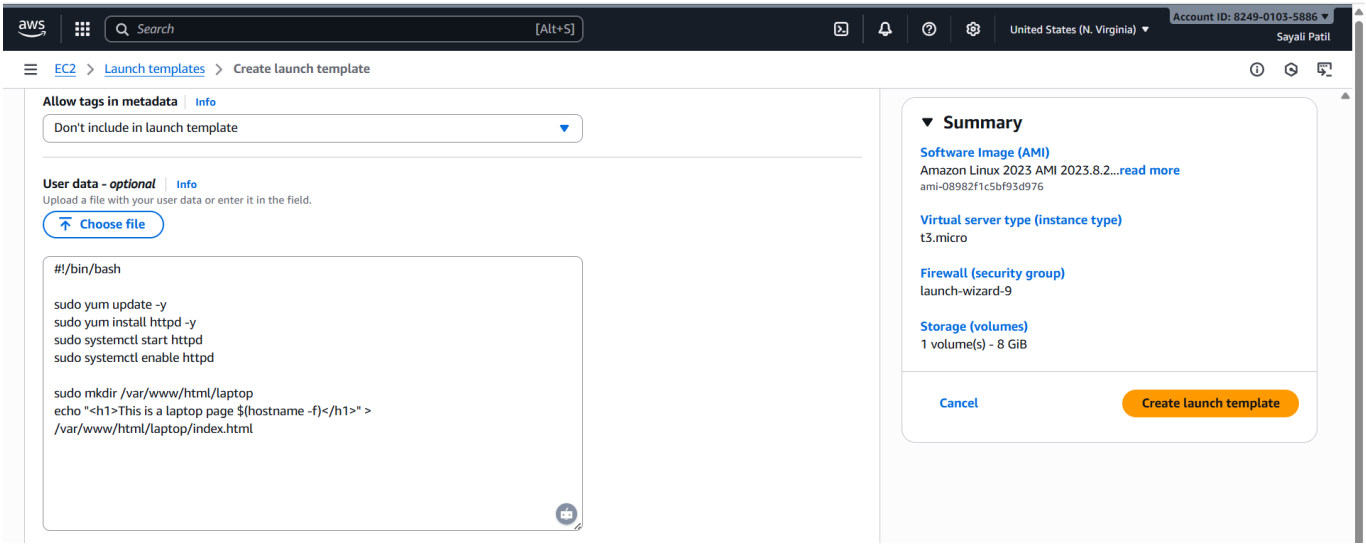

- Click on Launch Template
- Similarly, Create other templates for Mobile and Laptop

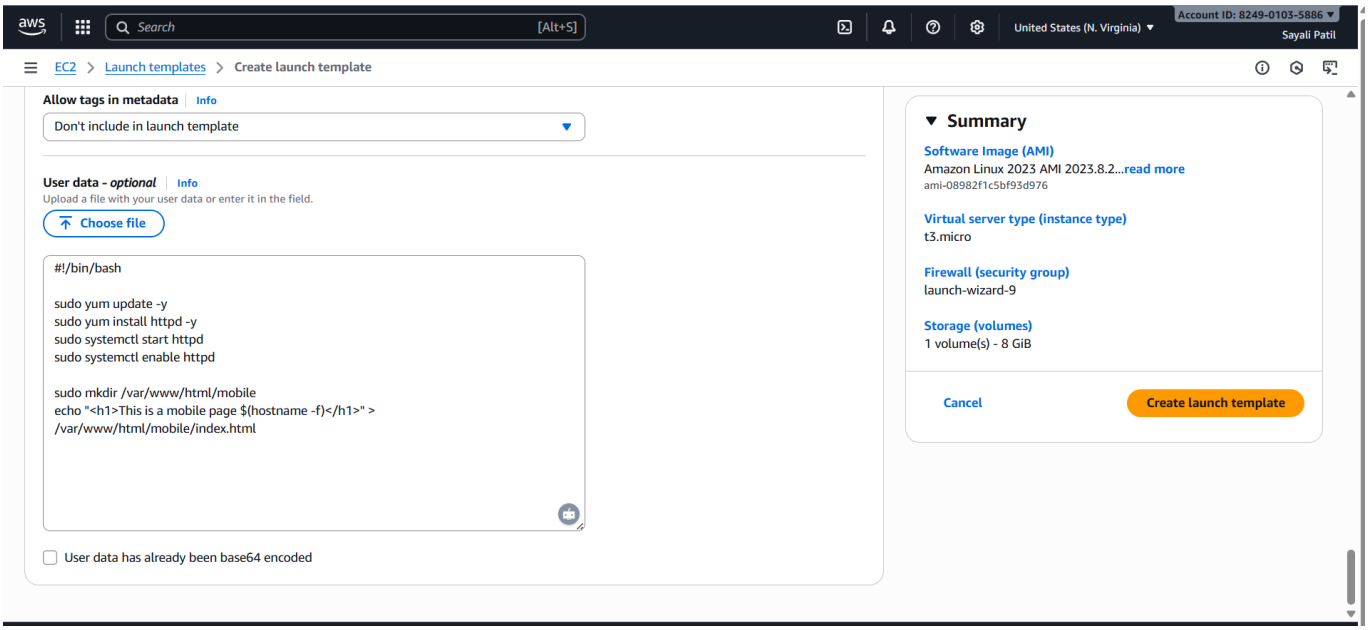
User datascript for mobile-

```
#!/bin/bash
sudo yum update -y
sudo yum install -y httpd
sudo systemctl start httpd
sudo systemctl enable httpd

sudo mkdir /var/www/html/mobile
echo "<h1>This is the mobile page $(hostname -f)</h1>" >
/var/www/html/mobile/index.html
```

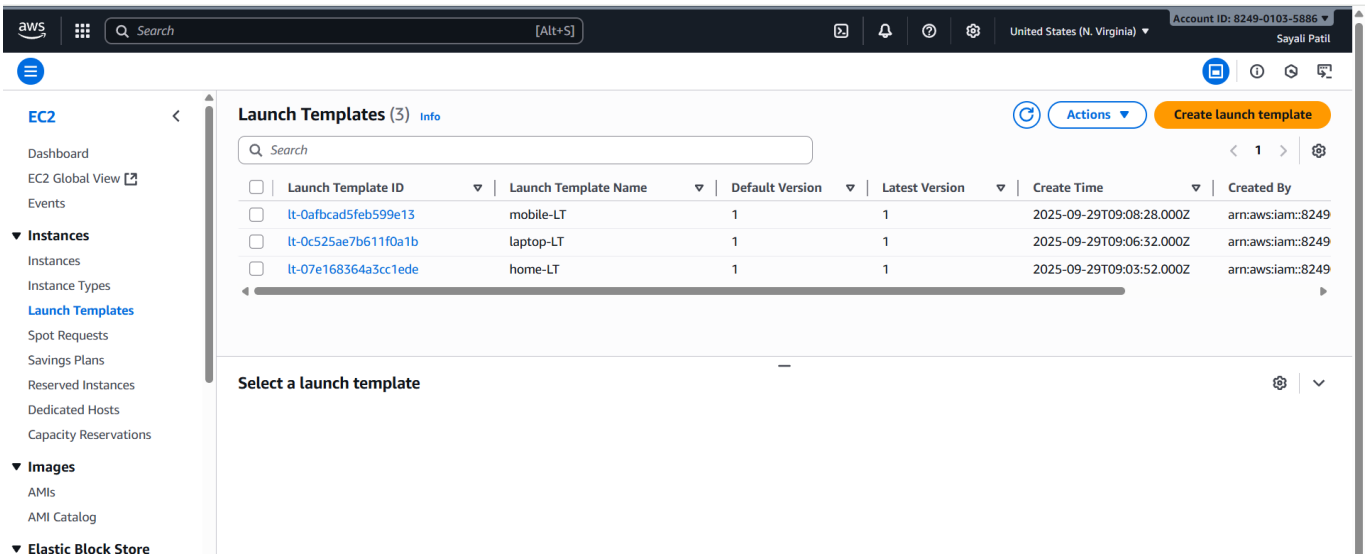
User datascript for laptop-

```
#!/bin/bash
sudo yum update -y
sudo yum install -y httpd
sudo systemctl start httpd
sudo systemctl enable httpd

sudo mkdir /var/www/html/laptop
echo "<h1>This is the laptop page $(hostname -f)</h1>" >
/var/www/html/laptop/index.html
```



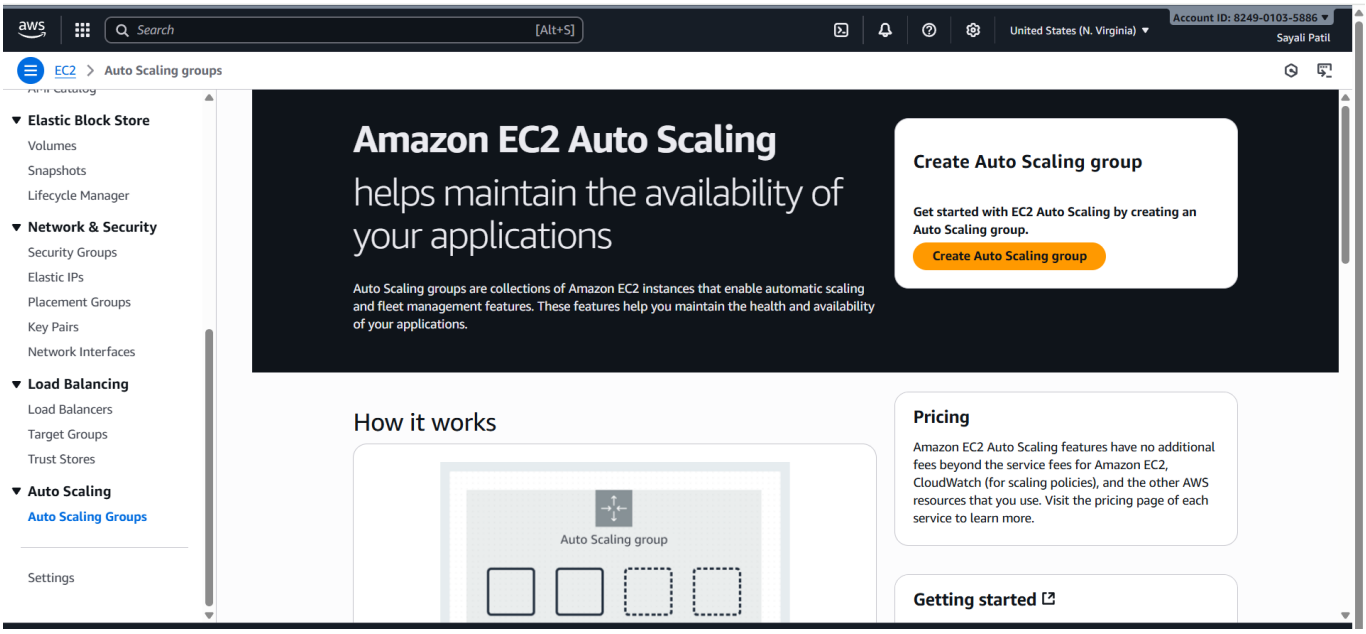

- Sucessfully, we created a Launch Template




## Step 2: Creating AutoScalling Group (ASG)

Let’s begin by creating the Auto Scaling Group for home, which uses static Scaling.

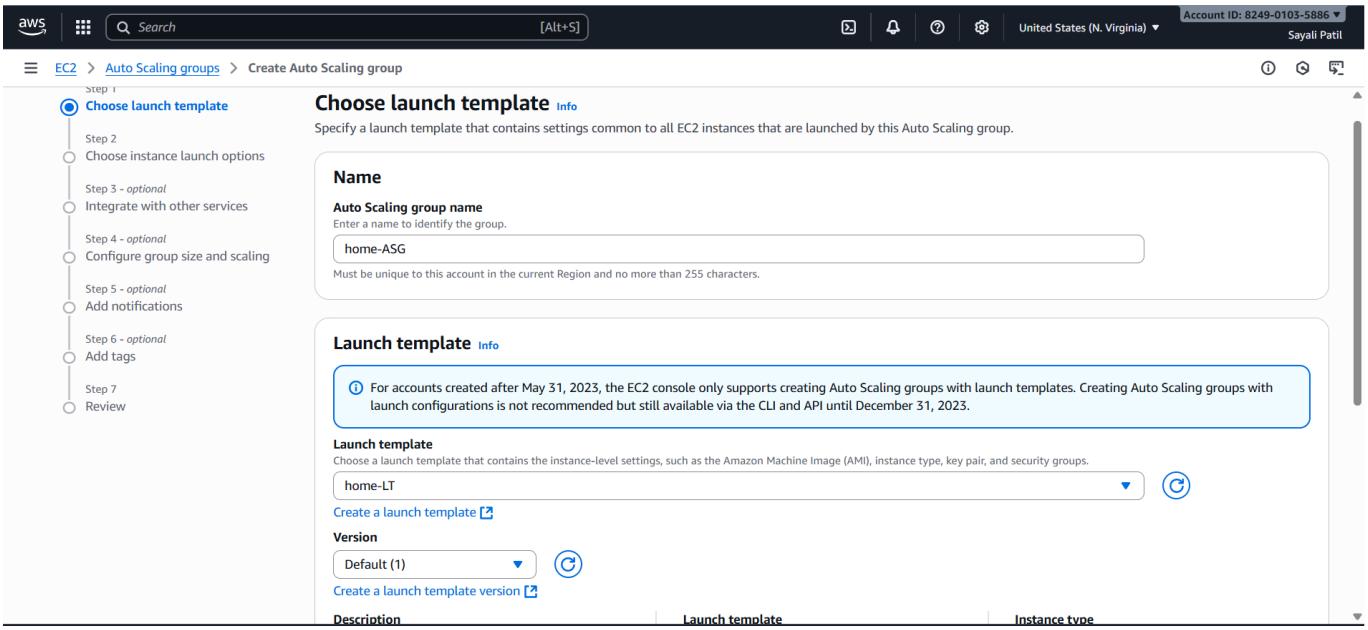
1. Navigate to the Auto Scaling Groups section in the AWS Management Console.
2. Click on “Create Auto Scaling Group” to get started.



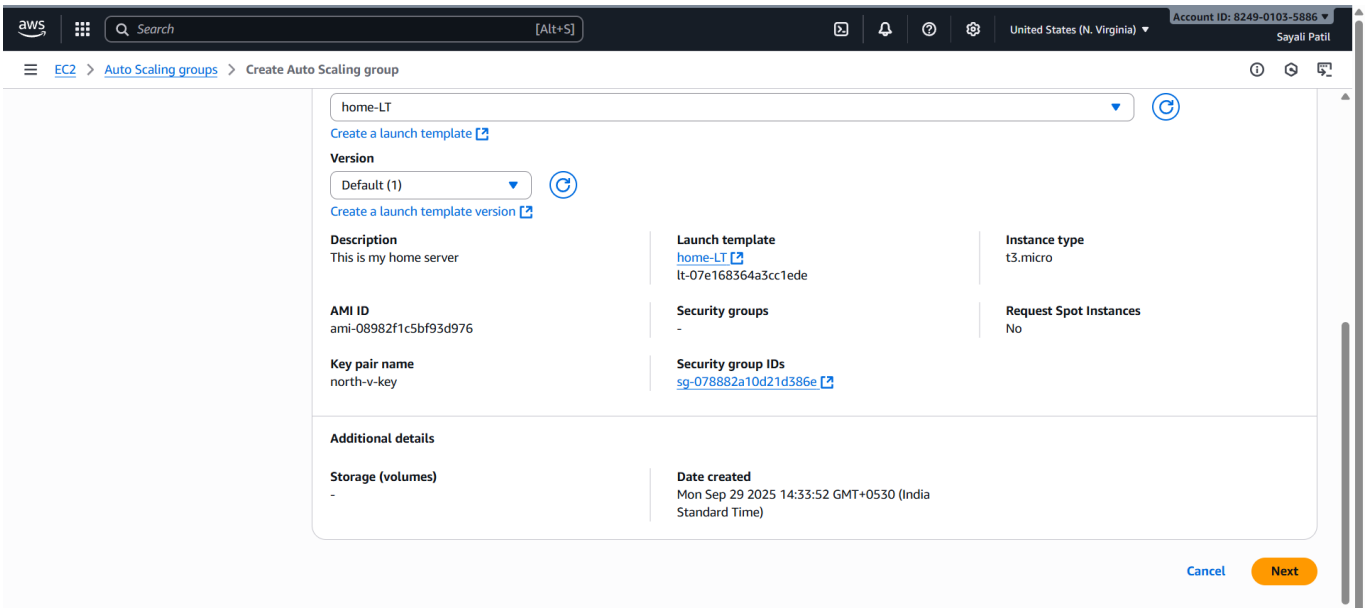

3. You’ll notice there are 7 setup steps listed on the left-hand side of the screen to guide you through the configuration process.

#Choose Launch Template

- Give Name to autoscalling group.

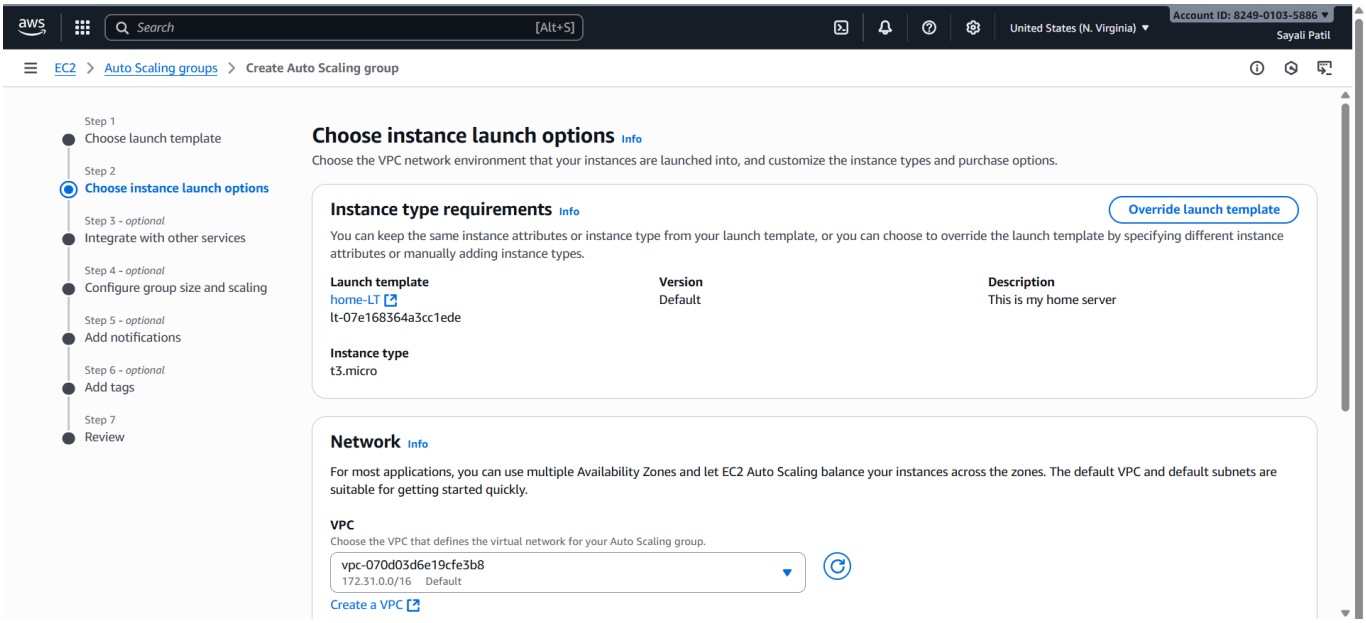


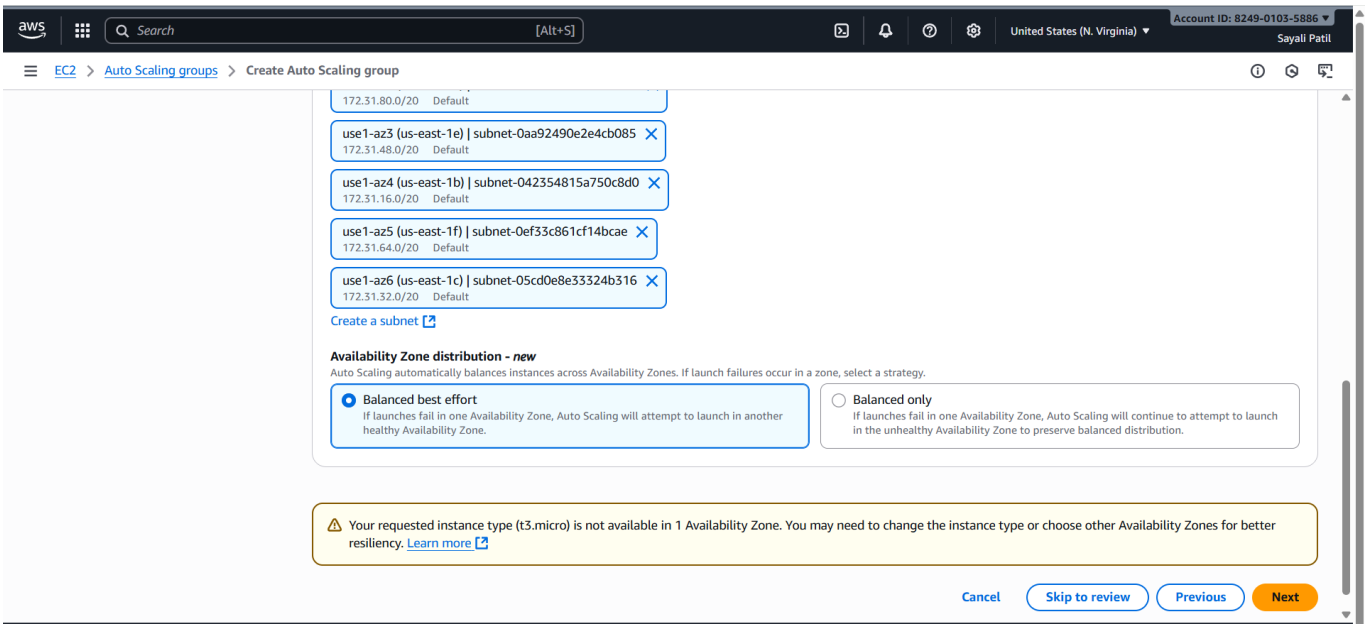

- Select a Launch Template of home for home autoscalling group.
- click on next & go to next step.




#Choose instance launch option

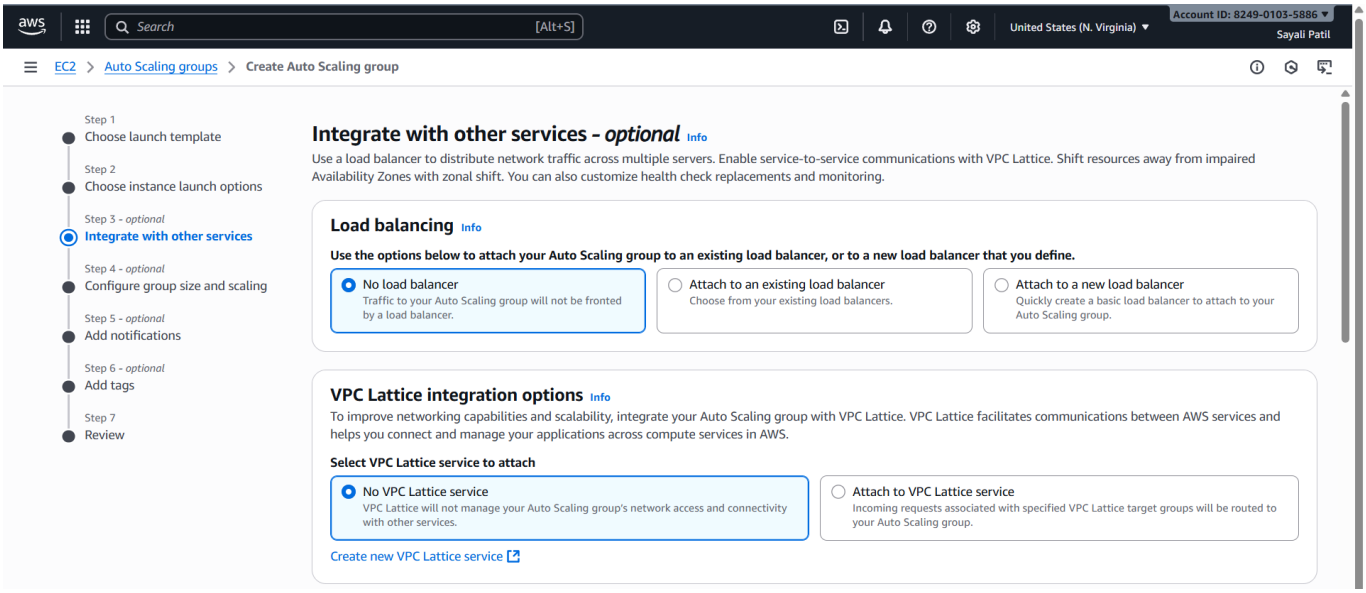
- Select Availability Zone. (Atleast 2)
- go to next step.

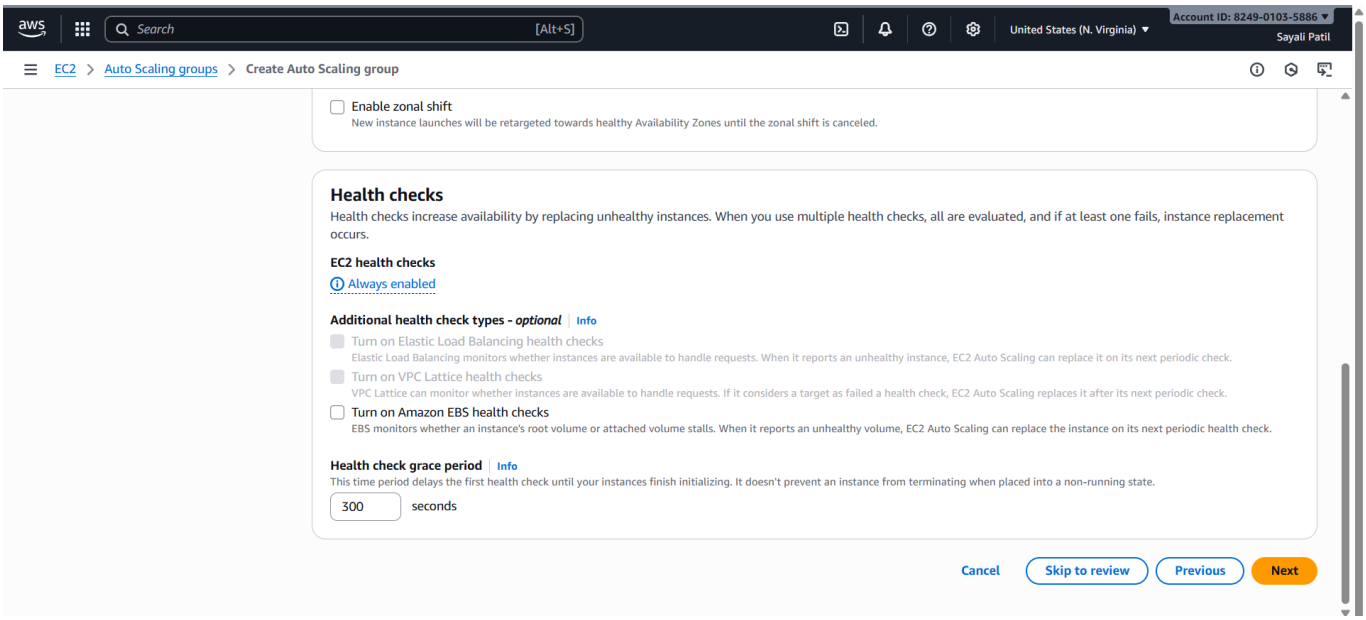





#Integrate with other services

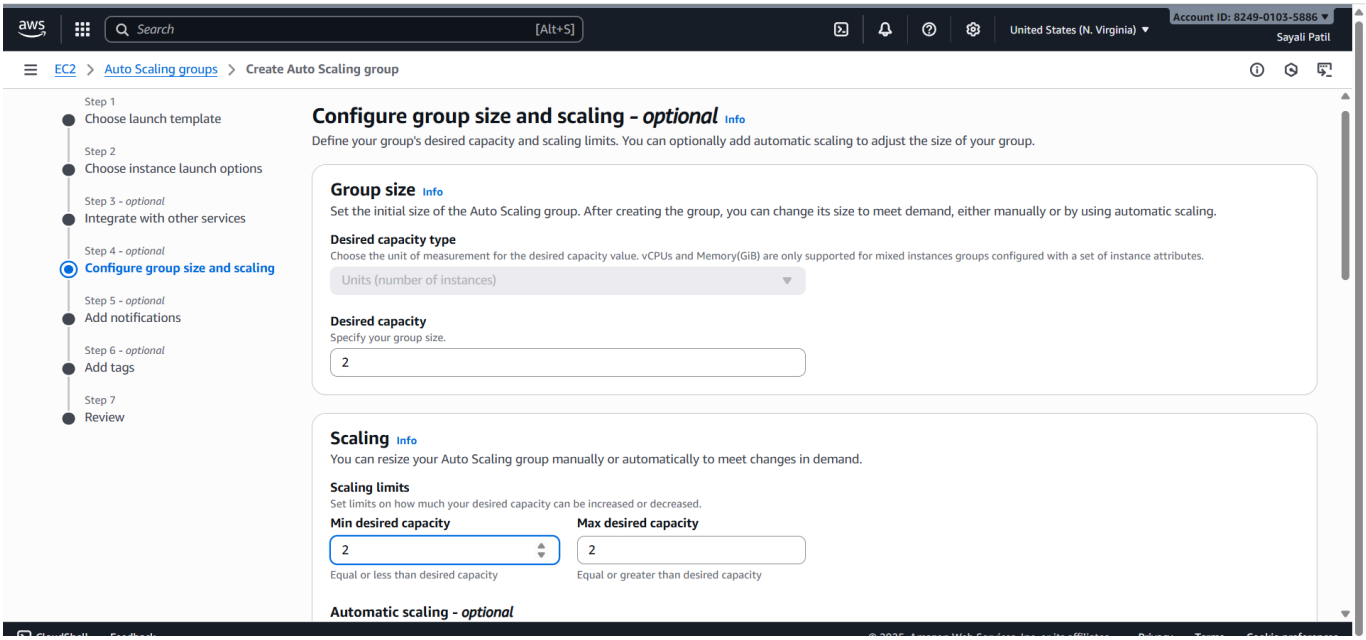
- We'll integrate these Auto Scaling Groups with the Load Balancer after all three have been created. For now, let's proceed to the next step.



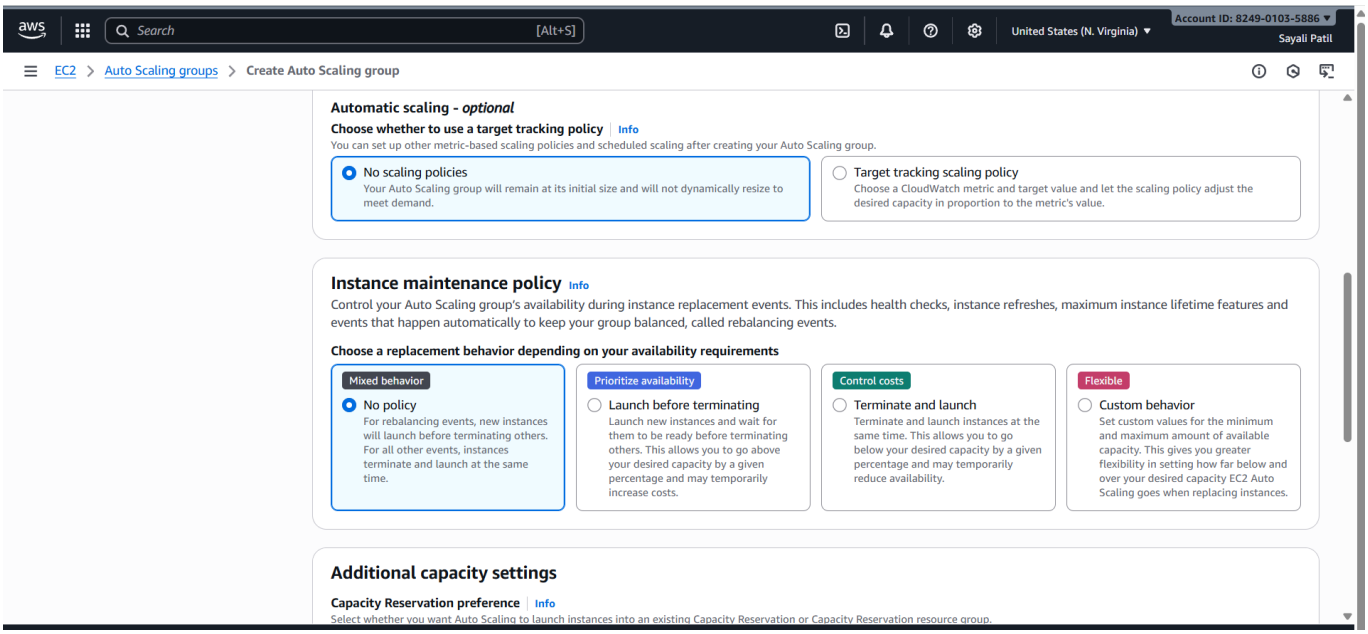



#Configure Group size and scalling

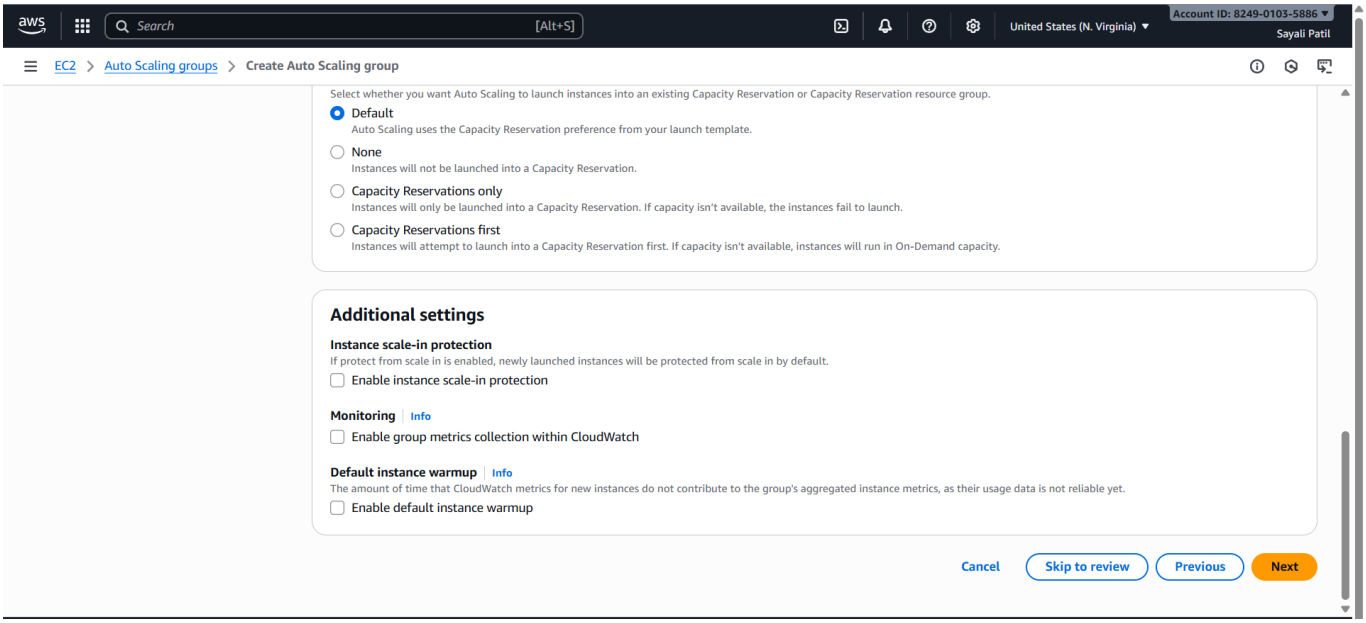
- For Home (static) ASG select : Desired = 2 , min = 2 , max = 2.
- Here for mobile we are giving : Desired = 3 , min = 2 , max = 7. (After creating mobile ASG we will make it sceduled)
- For Laptop (Dynamic) ASG select : Desired = 3 , min = 2 , max = 7.




- select target tracking scaling policy ( if your ASG is static you don't need policy so select No scaling policy)



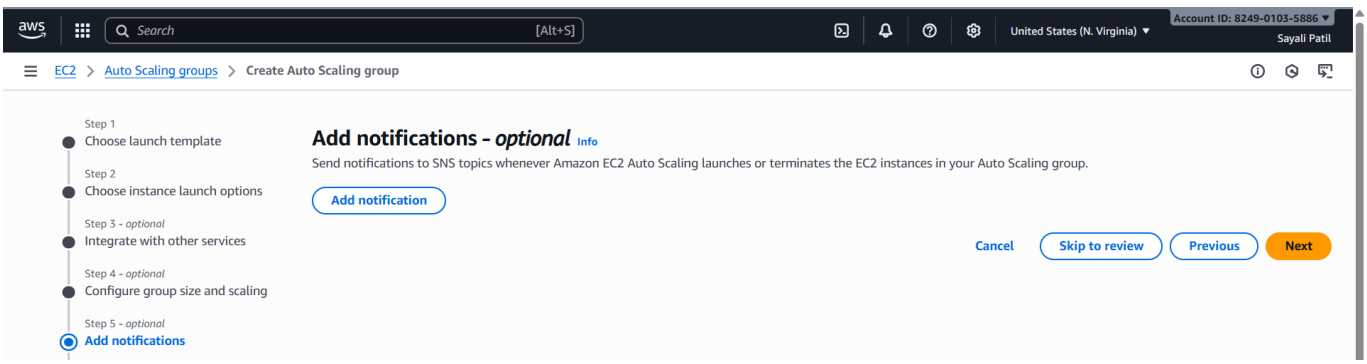

- Go to the next

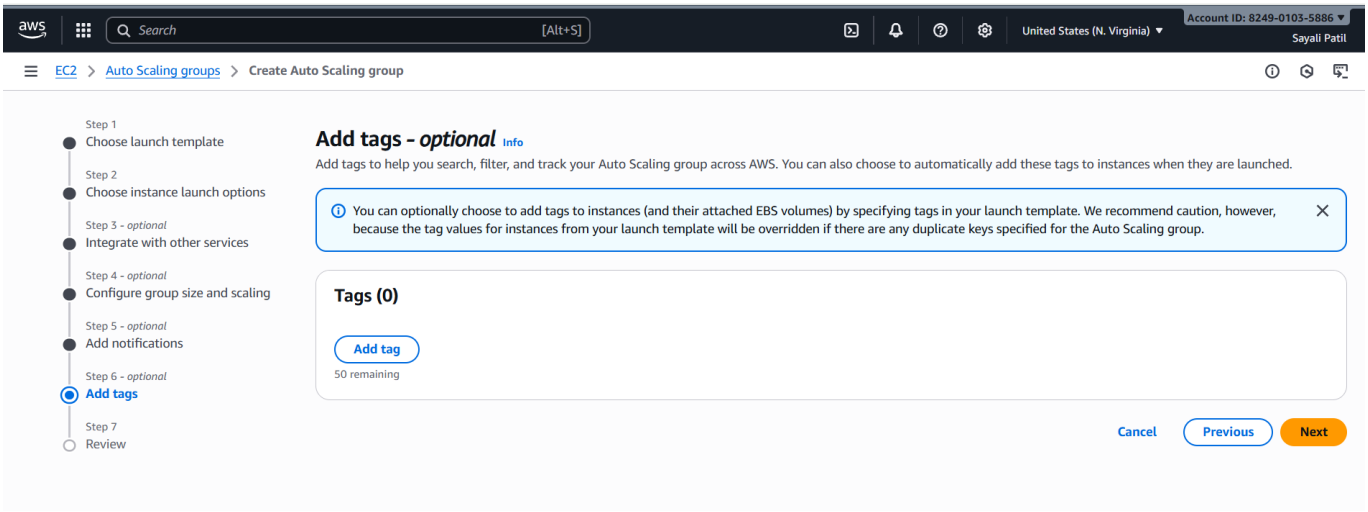



#Add notification

#Add tags

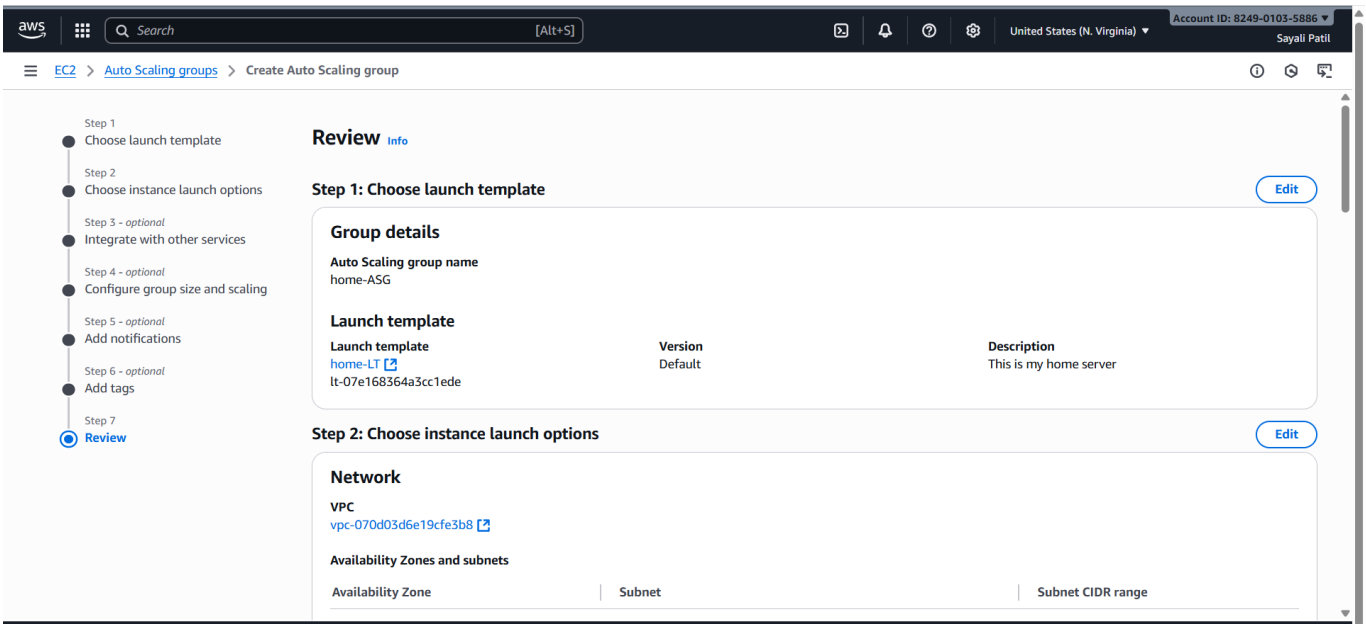
(you dont have to do anything in step 5 and 6 so just click on next)

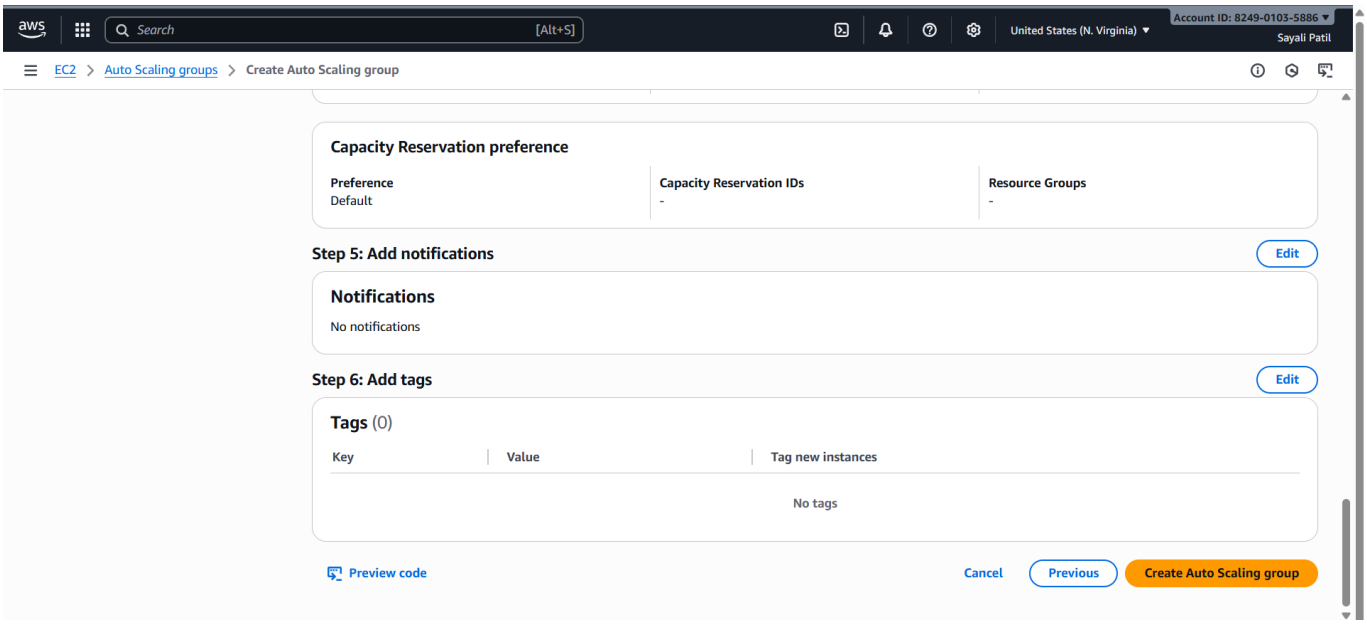




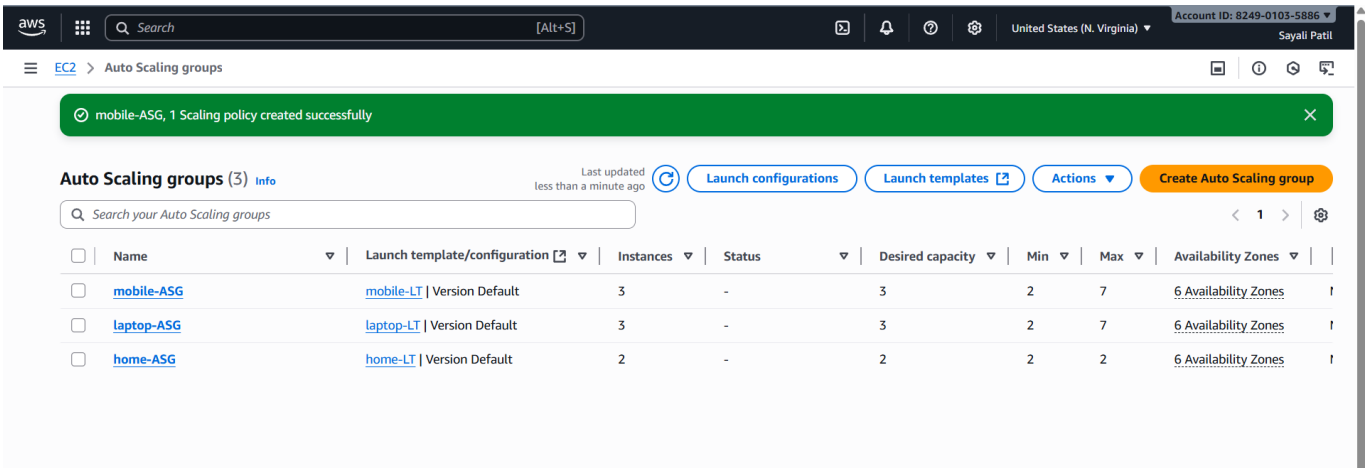

#Review

- select create autoscaling group



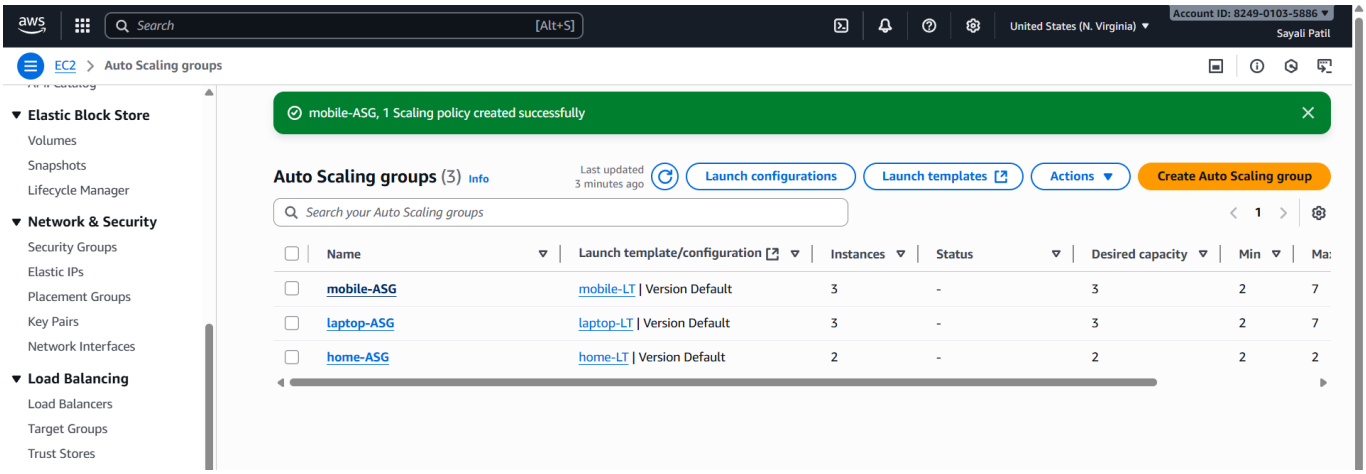



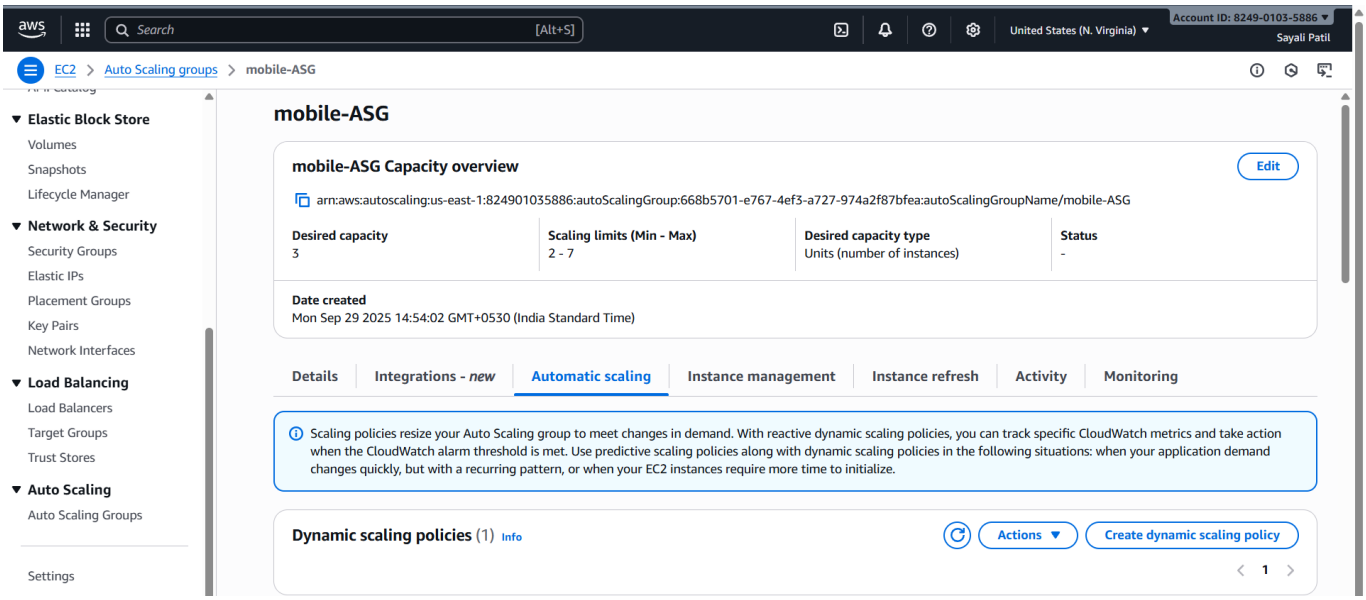
- Similarly create other 2 ASG according to their type dynamic and scheduling. Just need to change the group size as i mentioned.
- Only select target tracking scaling policy ( if your ASG is dynamic or scheduled you need policy so select target tracking scaling policy)
- Just need to change the group size as i mentioned.
    - Here, for mobile we are giving : Desired = 3 , min = 2 , max = 7. (After creating mobile ASG we will make it sceduled)
    - For Laptop (Dynamic) ASG select : Desired = 3 , min = 2 , max = 7.



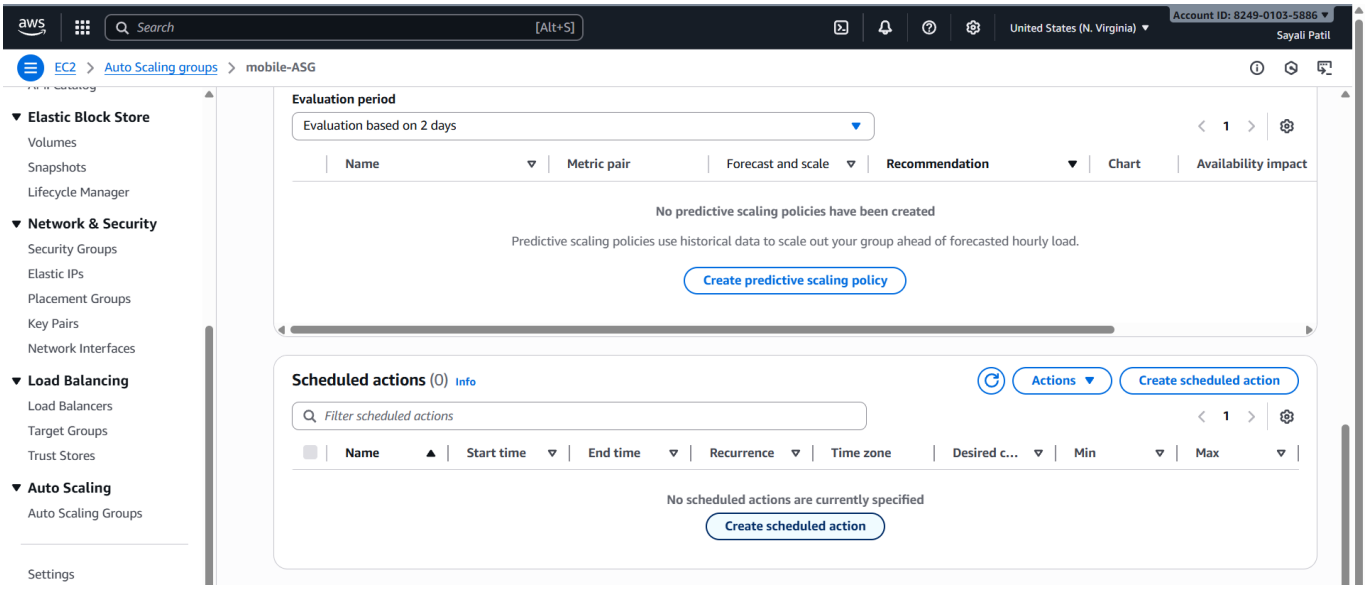

## Step 3: Now to make Mobile ASG Sceduled follow following steps

- Click on Mobile ASG and select Automatic scaling (You will see the following interface)

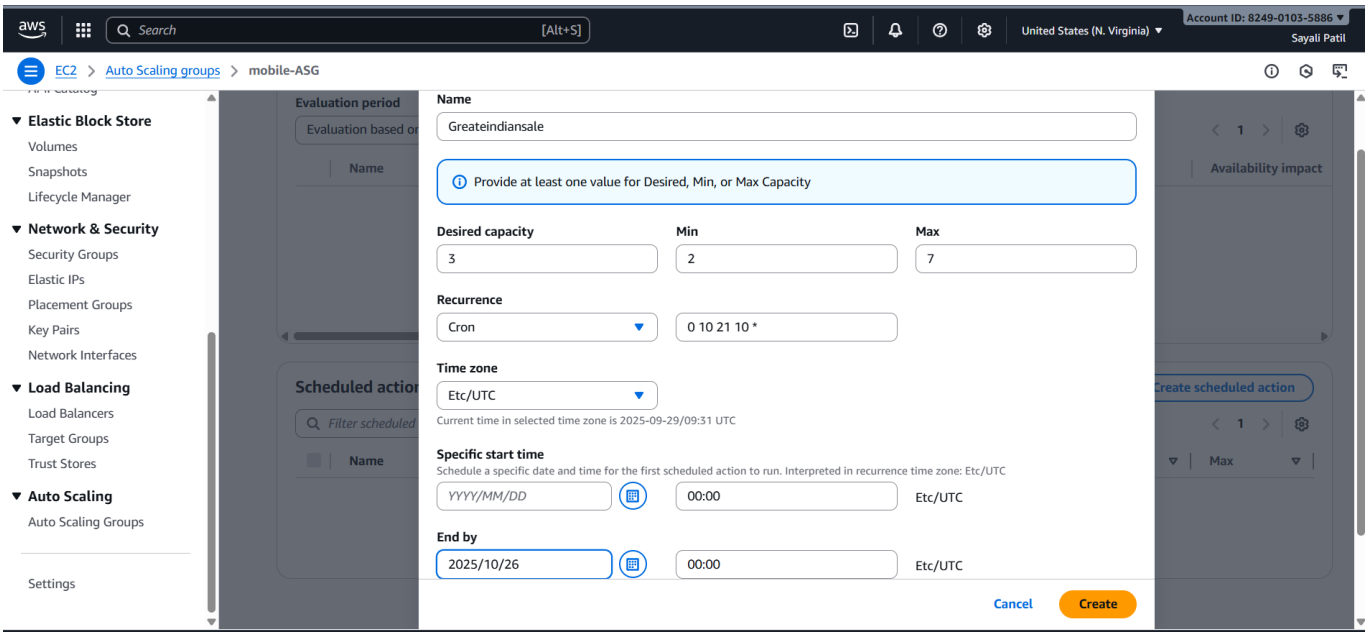


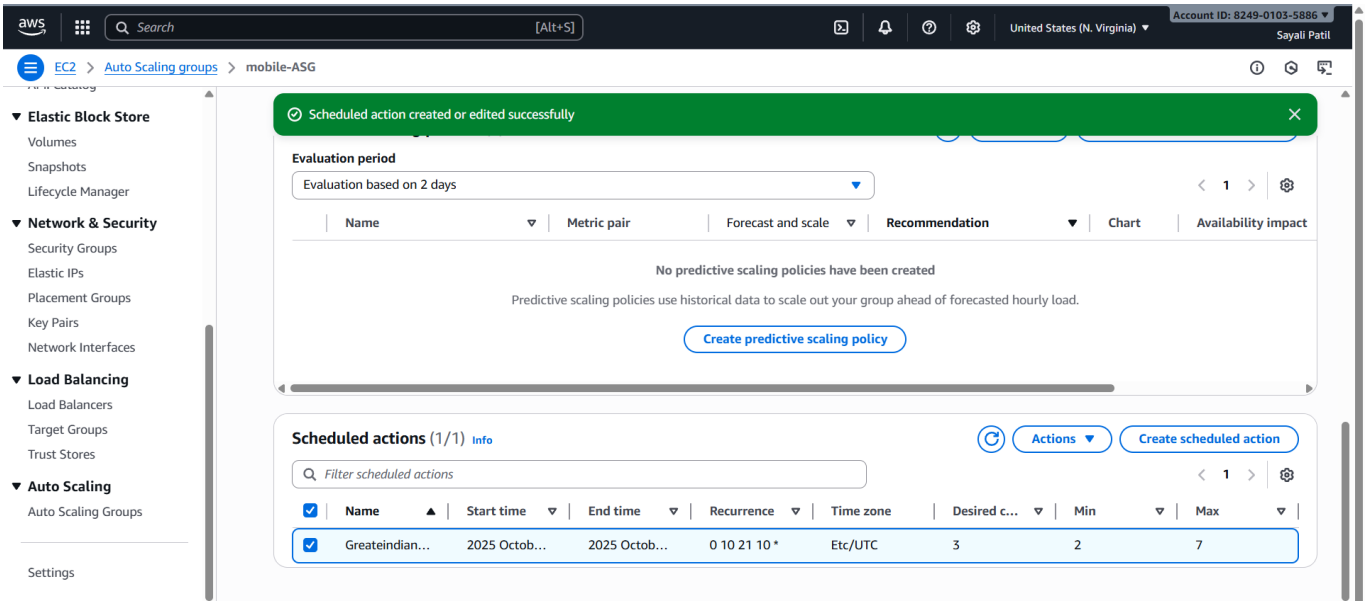



- Scroll down to bottom.
- Click on Create schedule action.




- Give name to your schedule.
- Enter Desired , min and max values of instances.
- Choose Cron option in recurrence and give specific time in side box ( sequence is - minute hour date month and day of week)
- Enter the ending time with date
- Click on Create

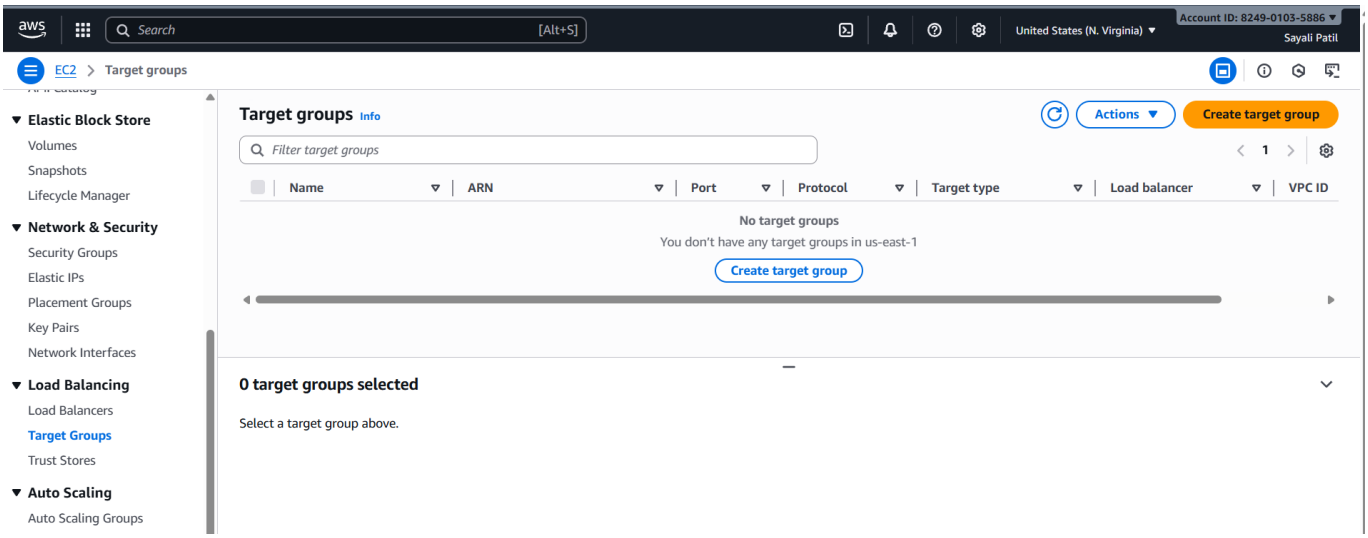


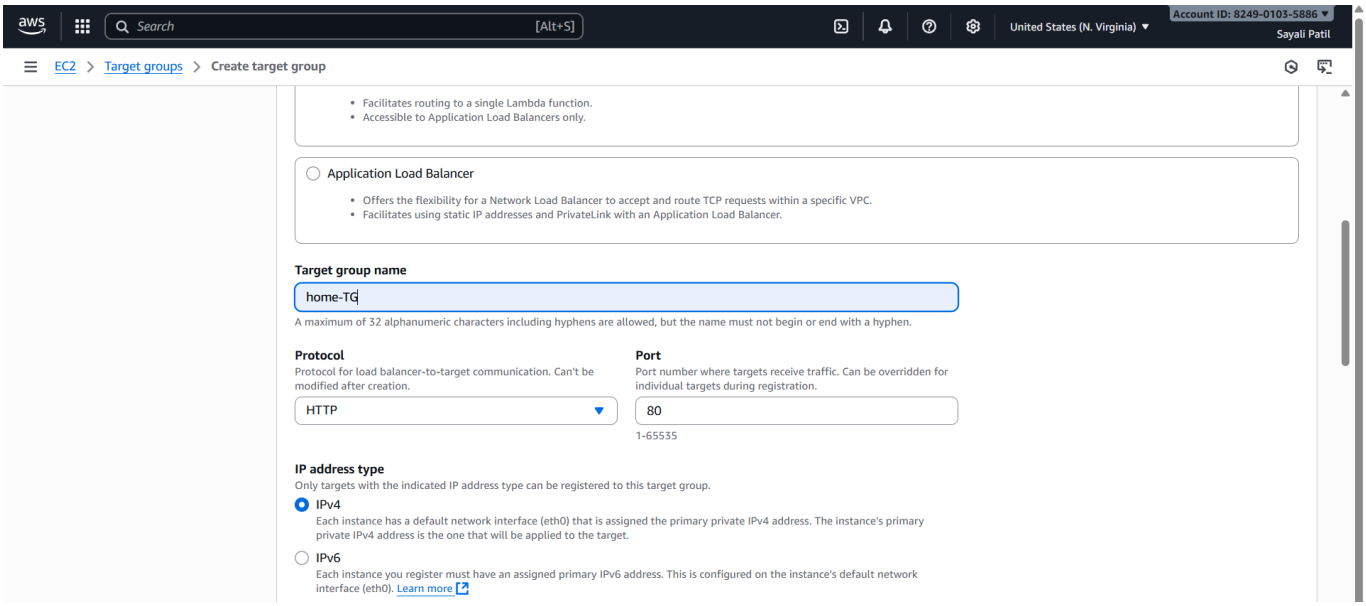

- If you click on the Mobile Auto Scaling Group and scroll down to the Scheduled Actions section, you'll see the scheduled action that has been created.

## Step 4: Creating Target Group For each Autoscalling Group

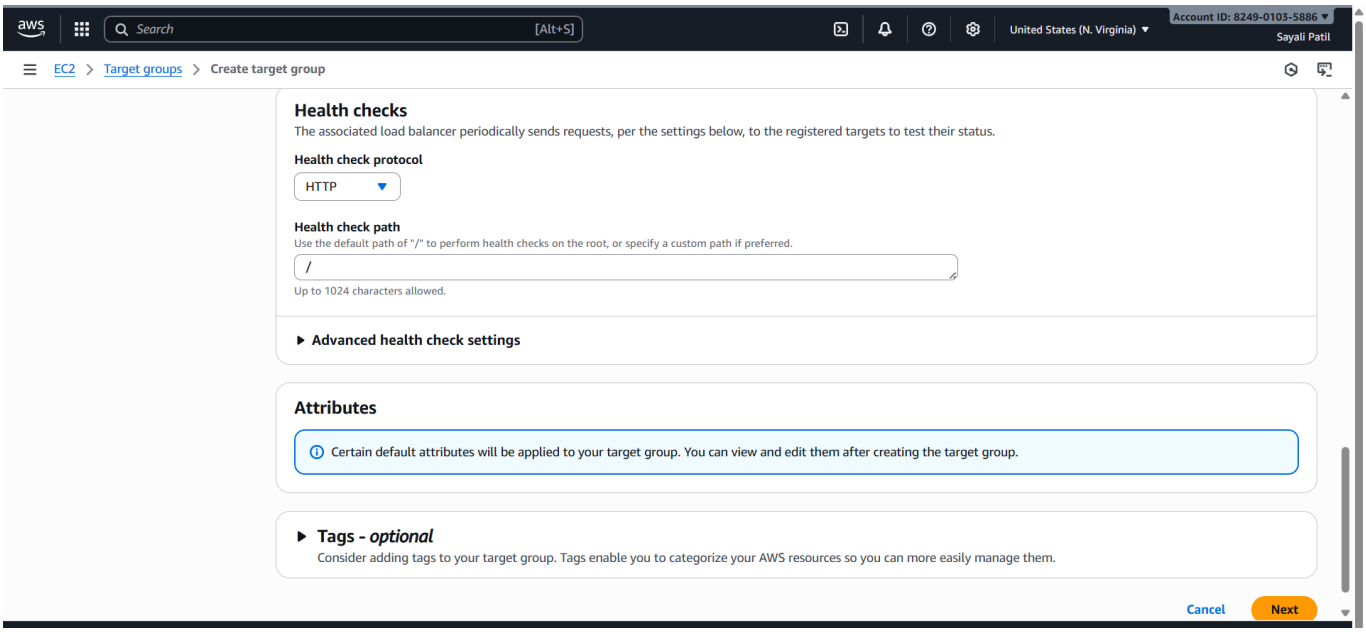
- Go to Target Groups
- Click on Create Target Group



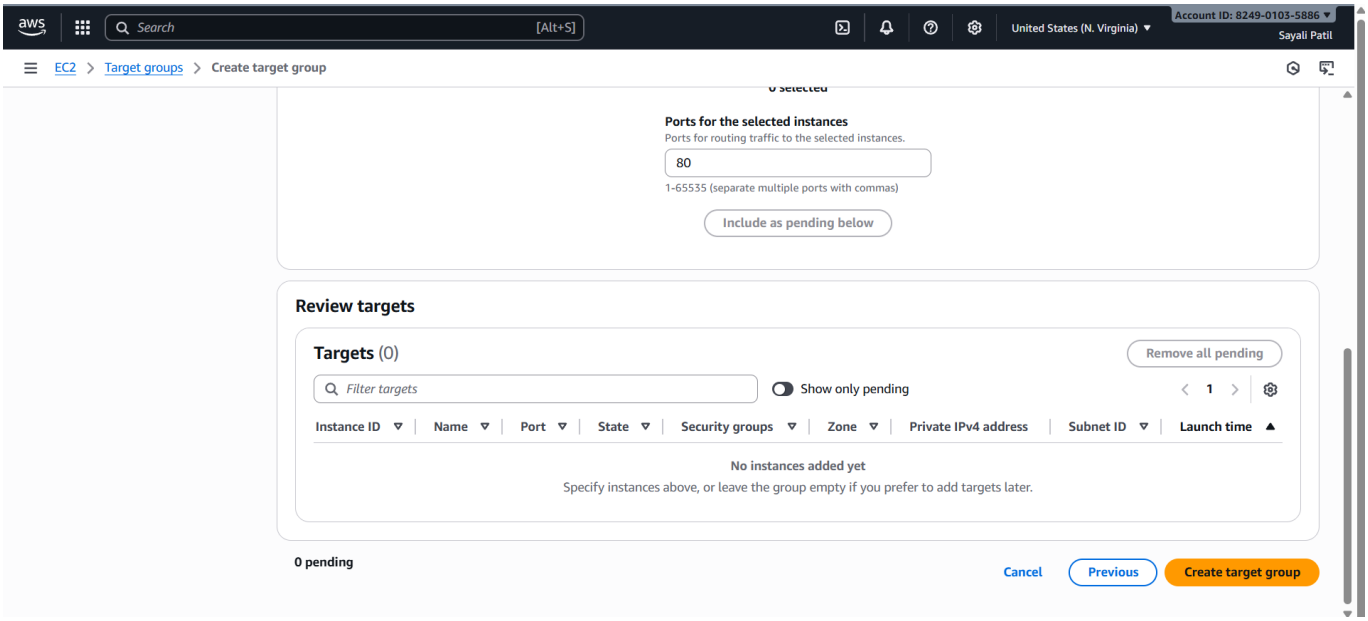

- Give a Name to your Target Group.




- Scroll down and go to health check section.
- Give a path of your application in Health Check Path section. [Note :For home, i Gave "/" as a path . Here "/ " represten the directry " /var/www/html/ " which is defualt directery of httpd]
- click on next

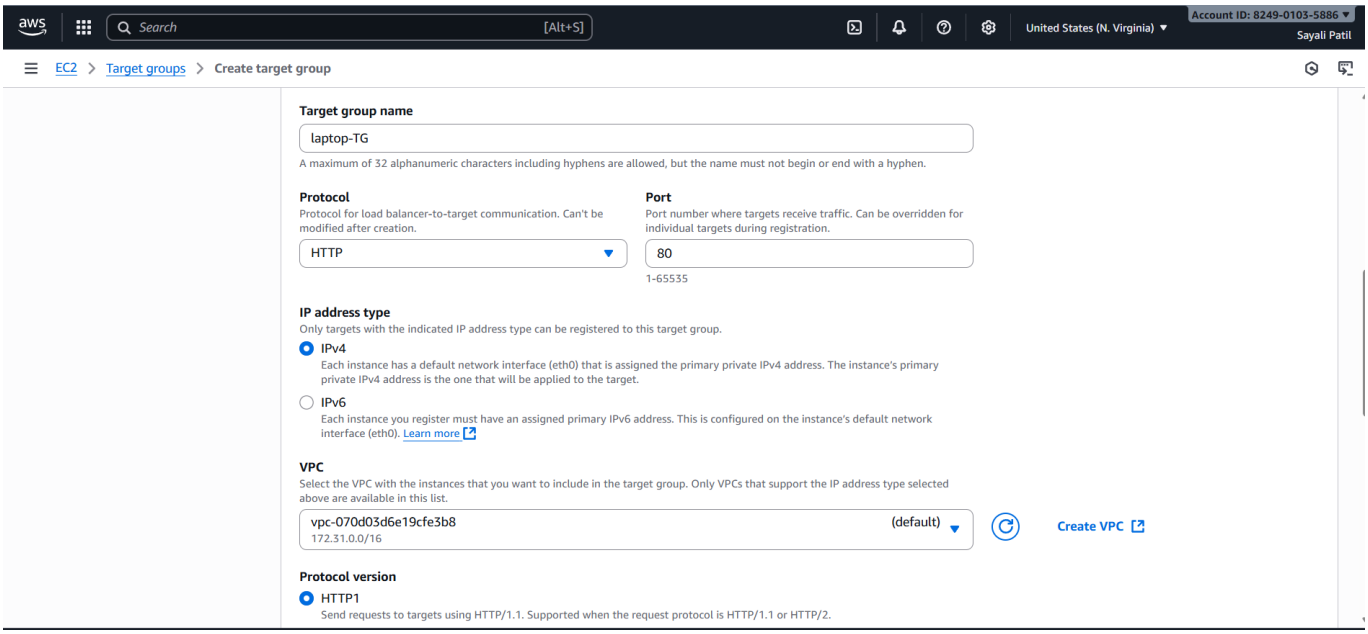


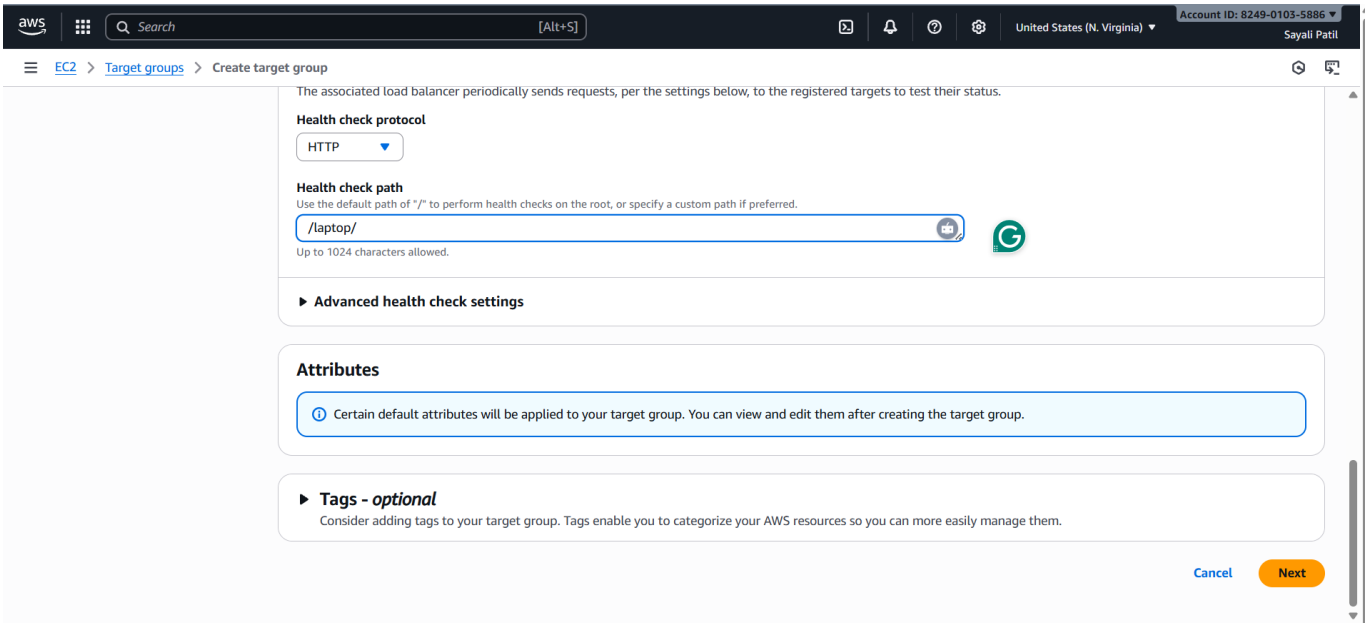

- Click on Create Traget Group (don't need to change anything else)



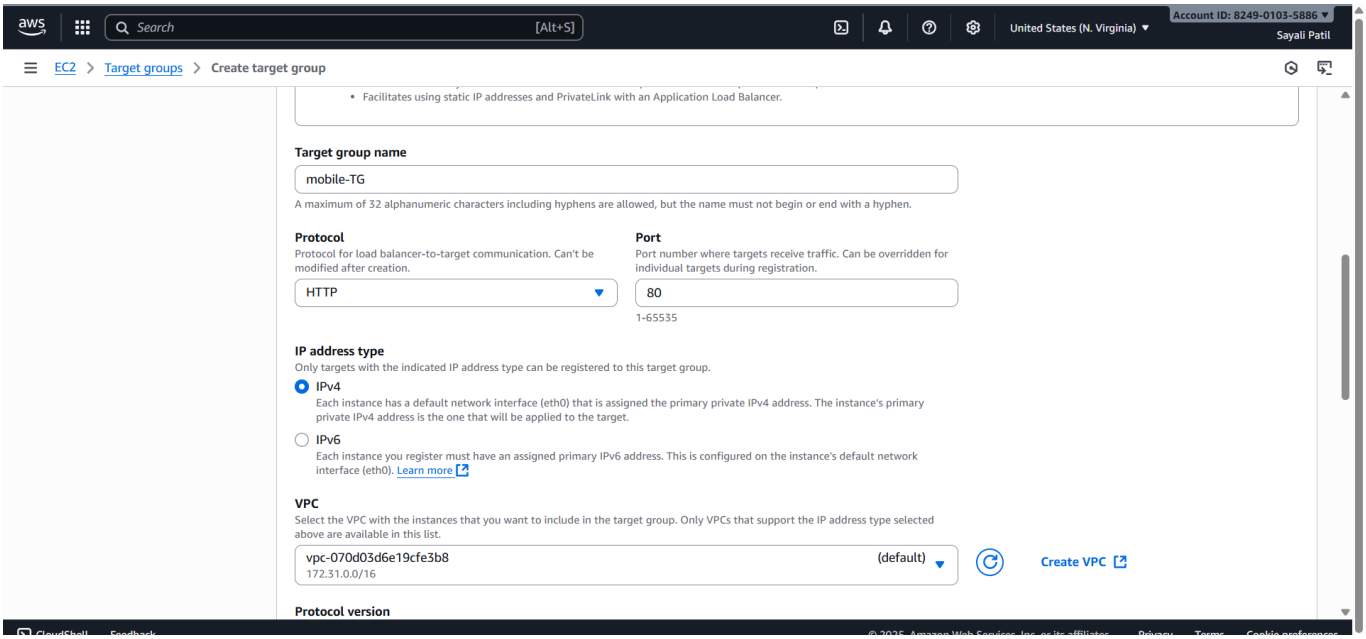

- Similarly Create Target Groups for Laptop and Mobile.
- You only need to change the path in Health check section as given below.

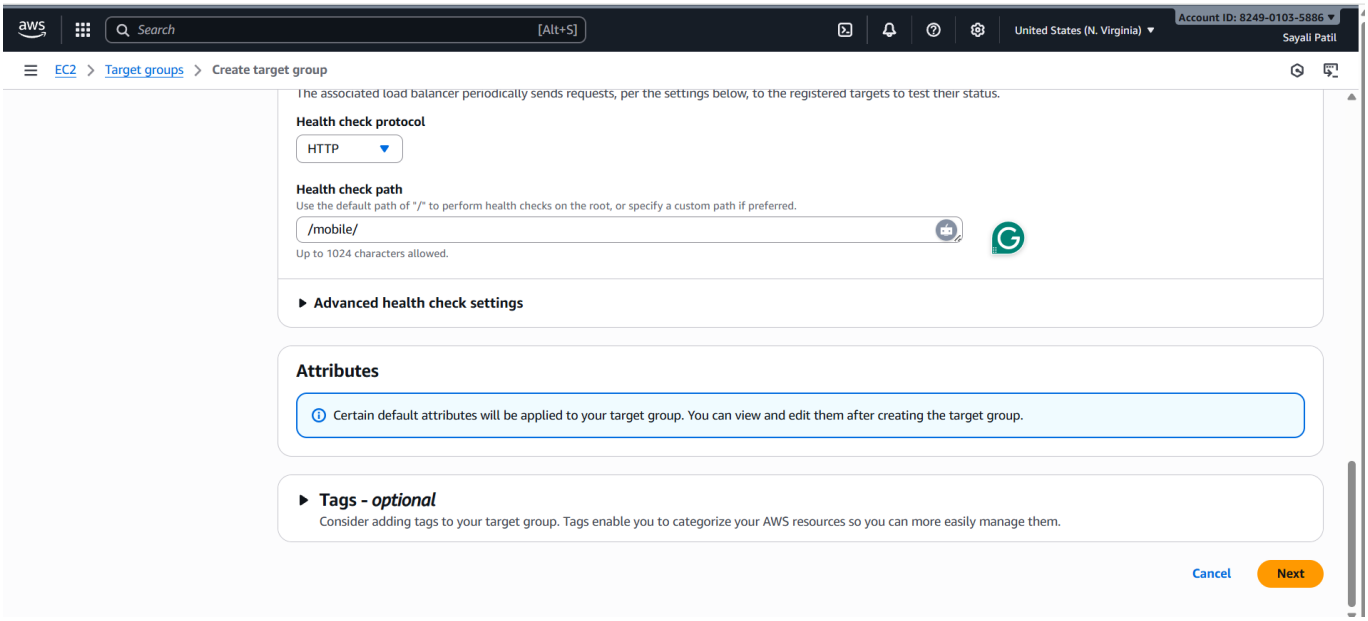
For Laptop :


Search [Alt+S]
United States (N. Virginia)
Account ID: 8249-0103-5886
Sayali Patil
EC2 > Target groups > Create target group
Target group name
laptop-TG
A maximum of 32 alphanumeric characters including hyphens are allowed, but the name must not begin or end with a hyphen.
Protocol
Protocol for load balancer-to-target communication. Can't be modified after creation.
HTTP
Port
Port number where targets receive traffic. Can be overridden for individual targets during registration.
80
1-65535
IP address type
Only targets with the indicated IP address type can be registered to this target group.
IPv4
Each instance has a default network interface (eth0) that is assigned the primary private IPv4 address. The instance's primary private IPv4 address is the one that will be applied to the target.
IPv6
Each instance you register must have an assigned primary IPv6 address. This is configured on the instance's default network interface (eth0). Learn more
VPC
Select the VPC with the instances that you want to include in the target group. Only VPCs that support the IP address type selected above are available in this list.
vpc-070d03d6e19cfe3b8
172.31.0.0/16
(default)
Create VPC
Protocol version
HTTP1
Send requests to targets using HTTP/1.1. Supported when the request protocol is HTTP/1.1 or HTTP/2.

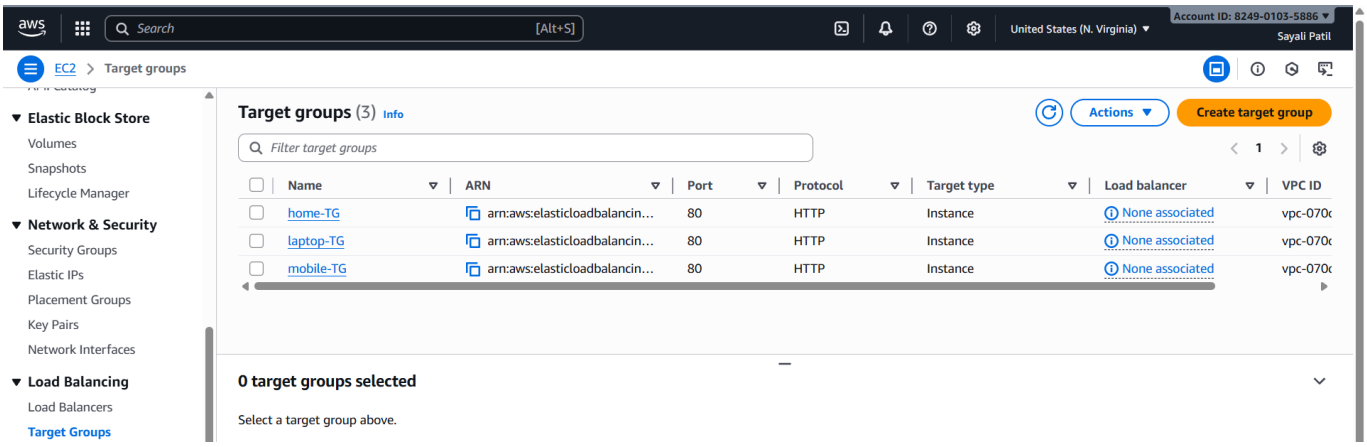

Search [Alt+S]
United States (N. Virginia)
Account ID: 8249-0103-5886
Sayali Patil
EC2 > Target groups > Create target group
The associated load balancer periodically sends requests, per the settings below, to the registered targets to test their status.
Health check protocol
HTTP
Health check path
Use the default path of "/" to perform health checks on the root, or specify a custom path if preferred.
/laptop/
Up to 1024 characters allowed.
Advanced health check settings
Attributes
Certain default attributes will be applied to your target group. You can view and edit them after creating the target group.
Tags - optional
Consider adding tags to your target group. Tags enable you to categorize your AWS resources so you can more easily manage them.
Cancel
Next

For Mobile :


Search [Alt+S]
United States (N. Virginia)
Account ID: 8249-0103-5886
Sayali Patil
EC2 > Target groups > Create target group
Facilitates using static IP addresses and PrivateLink with an Application Load Balancer.
Target group name
mobile-TG
A maximum of 32 alphanumeric characters including hyphens are allowed, but the name must not begin or end with a hyphen.
Protocol
Protocol for load balancer-to-target communication. Can't be modified after creation.
HTTP
Port
Port number where targets receive traffic. Can be overridden for individual targets during registration.
80
1-65535
IP address type
Only targets with the indicated IP address type can be registered to this target group.
IPv4
Each instance has a default network interface (eth0) that is assigned the primary private IPv4 address. The instance's primary private IPv4 address is the one that will be applied to the target.
IPv6
Each instance you register must have an assigned primary IPv6 address. This is configured on the instance's default network interface (eth0). Learn more
VPC
Select the VPC with the instances that you want to include in the target group. Only VPCs that support the IP address type selected above are available in this list.
vpc-070d03d6e19cfe3b8
172.31.0.0/16
(default)
Create VPC
Protocol version

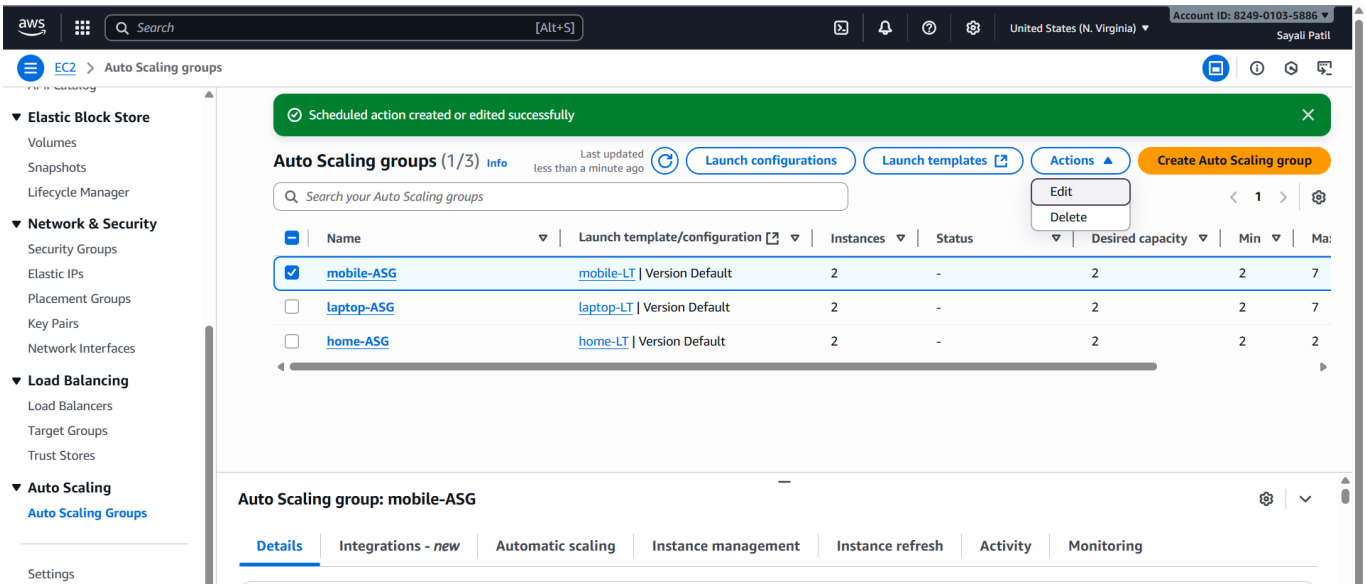



- Now, three Target Groups are created



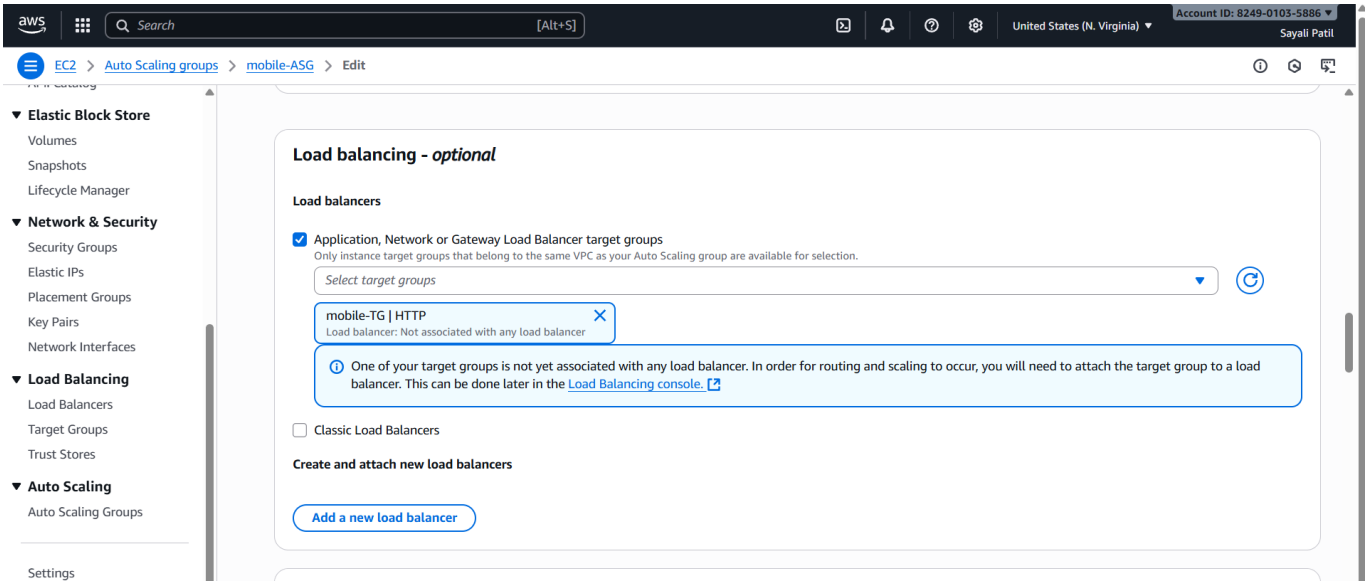

## Step 5: Connecting Autoscalling group to Target Groups

- Select an Autoscalling Group .
- Click on Action at top right corner.
- Click on Edit

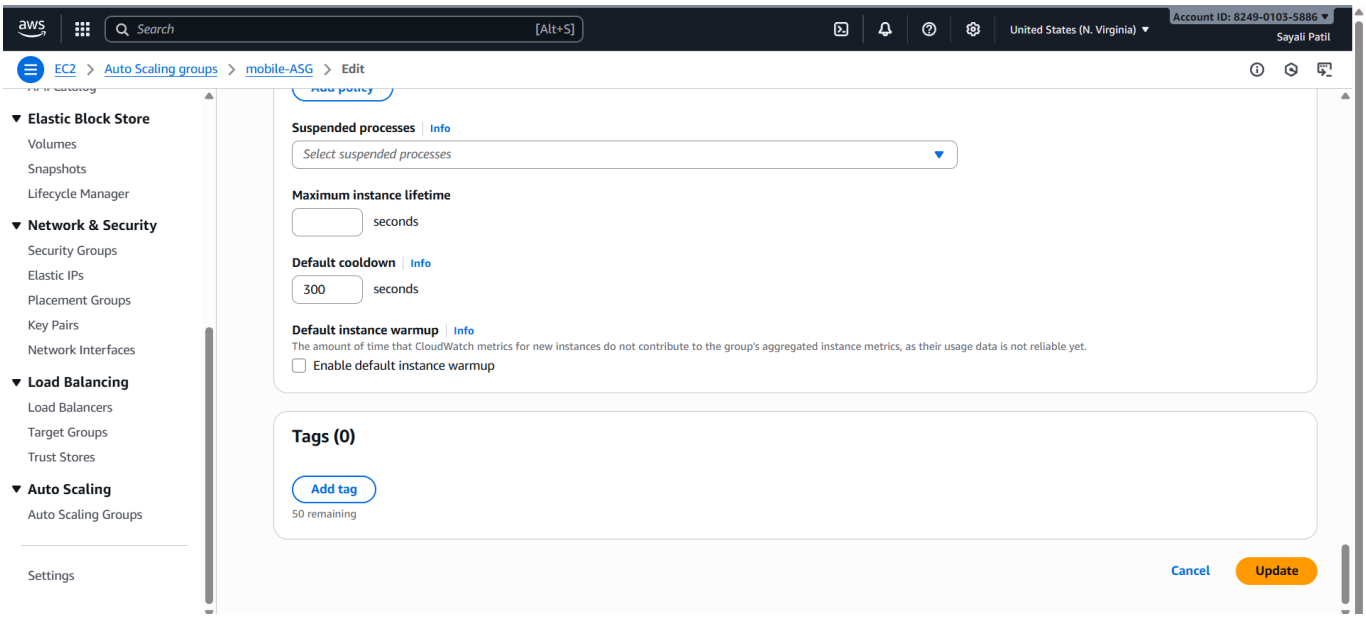


* Scroll down and you will see the Load balancing section.

- Click on Application load balancer
- Select a Target group for autoscalling group we selected. (for laptop ASG select Laptop target group, for mobile ASG select Mobile Target group and for home ASG select home target group.)



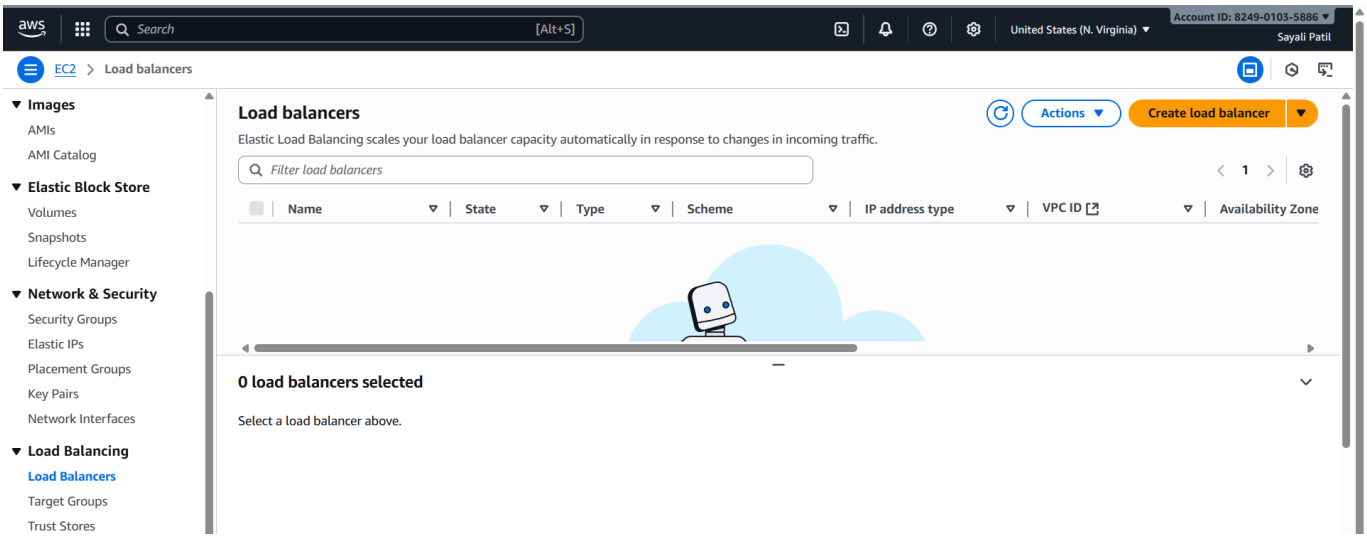

- click on update.



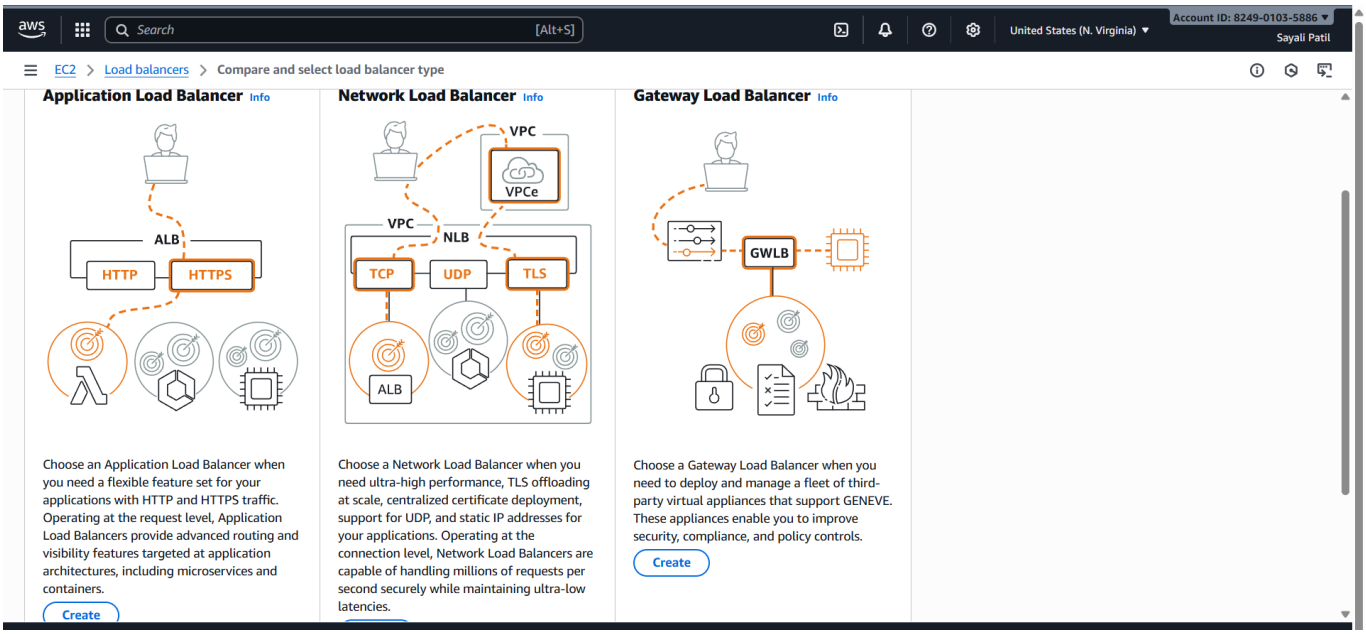

- Successfully connected the Autoscalling group and target groups.
- Repeat process to connect each autoscalling group to its target group.

## Step 6: Setting up an Application Load Balancer and linking it to target groups.

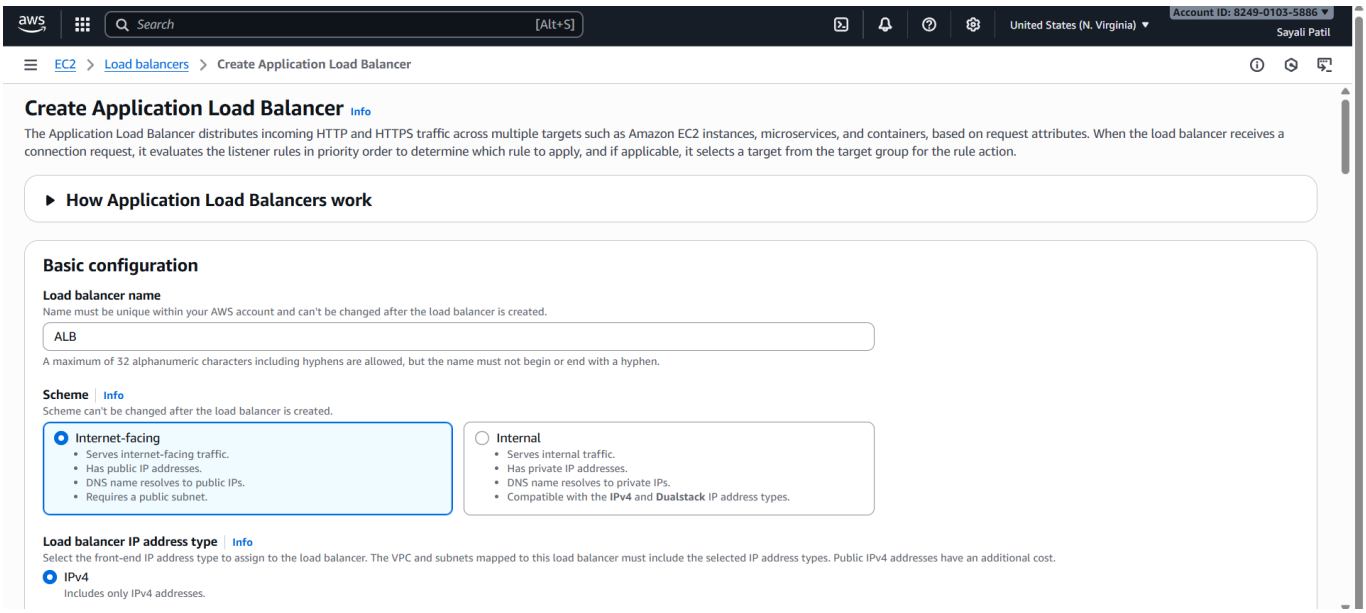
- Go to Load Balancers and Click on Create Load Balancer



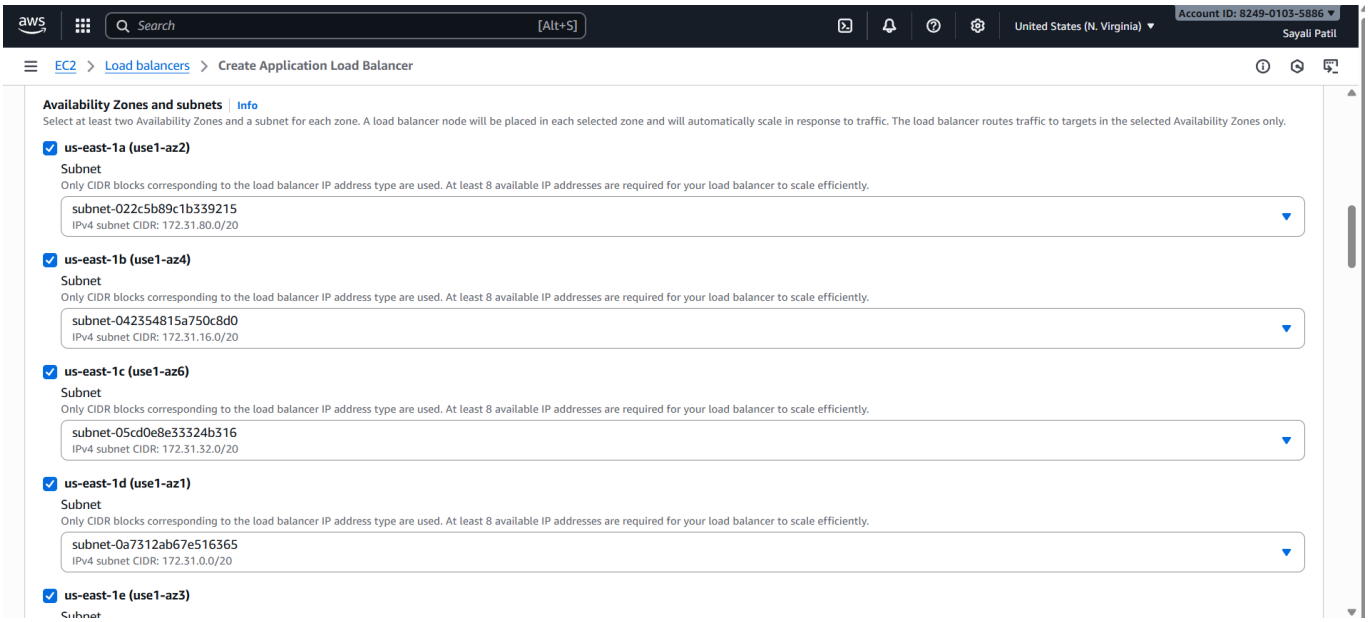

- Select Application Load Balancer & Click on Create



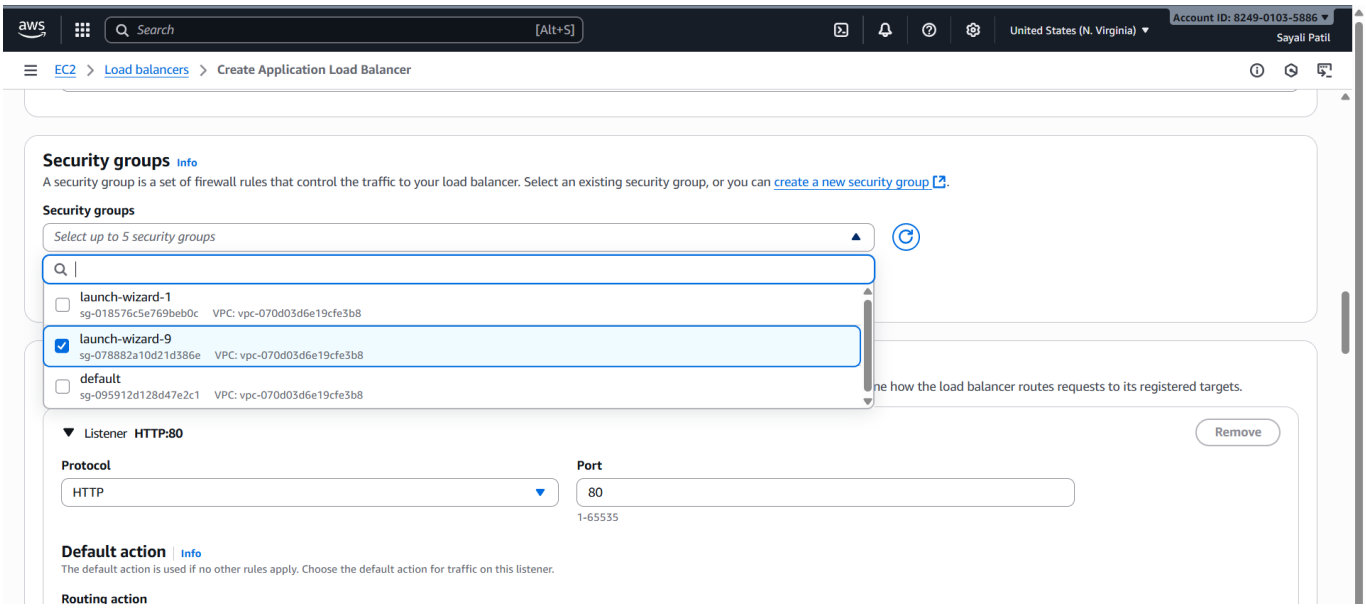

- Give Load balancer name Like ALB



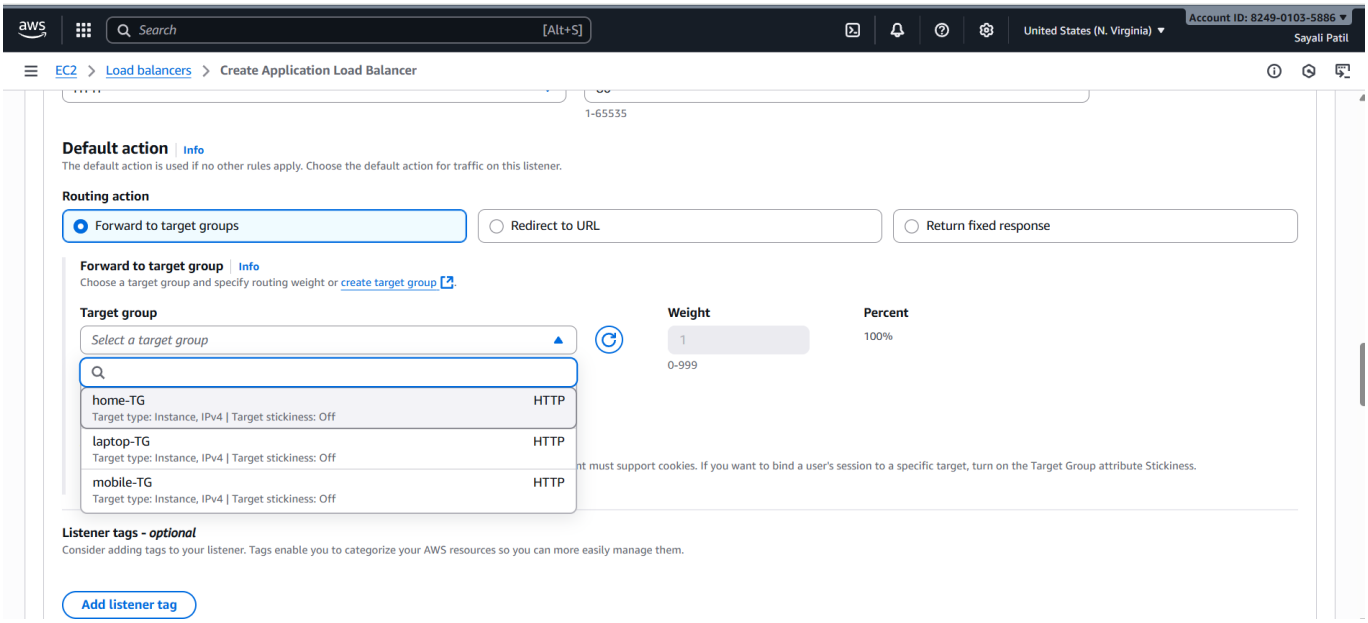
- Select all AZ (choose AZ same as the AZ we selected while creating autoscalling group.)



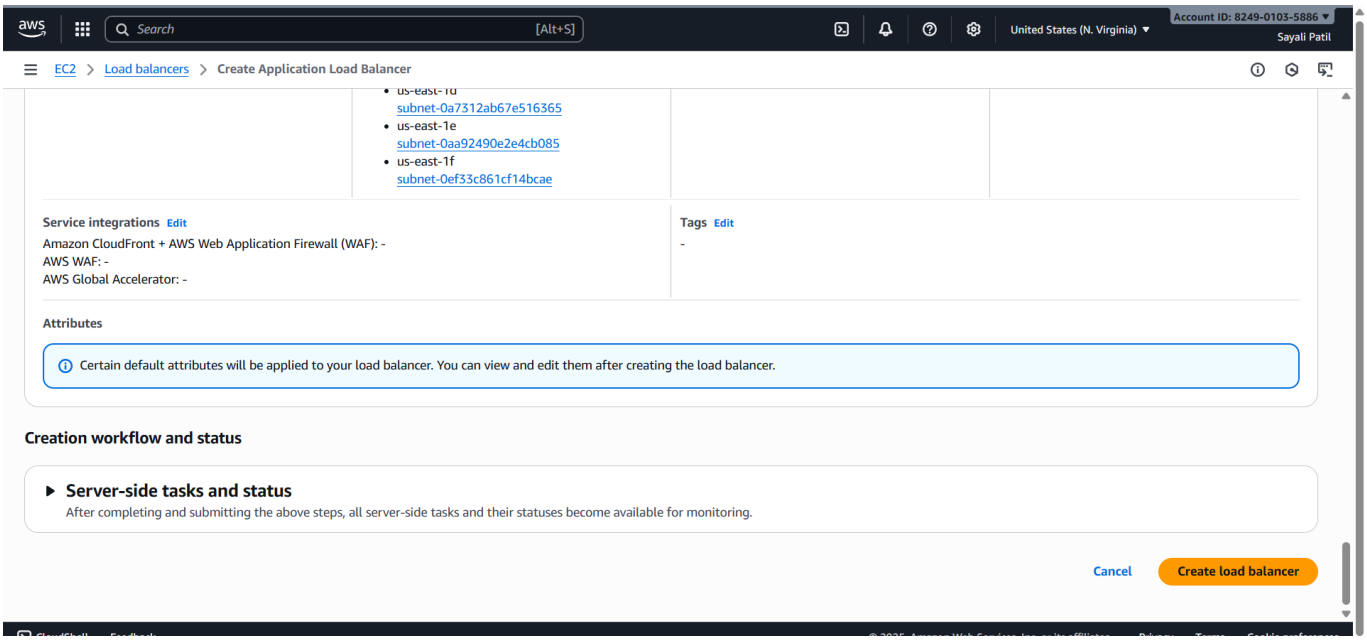

- select a existing security group. (Select same security group we selected while creating autoscalling group)




- In Listners and Routing section,
    - Select forward to target group
    - Choose a default target group. (Here default target group is home-TG)

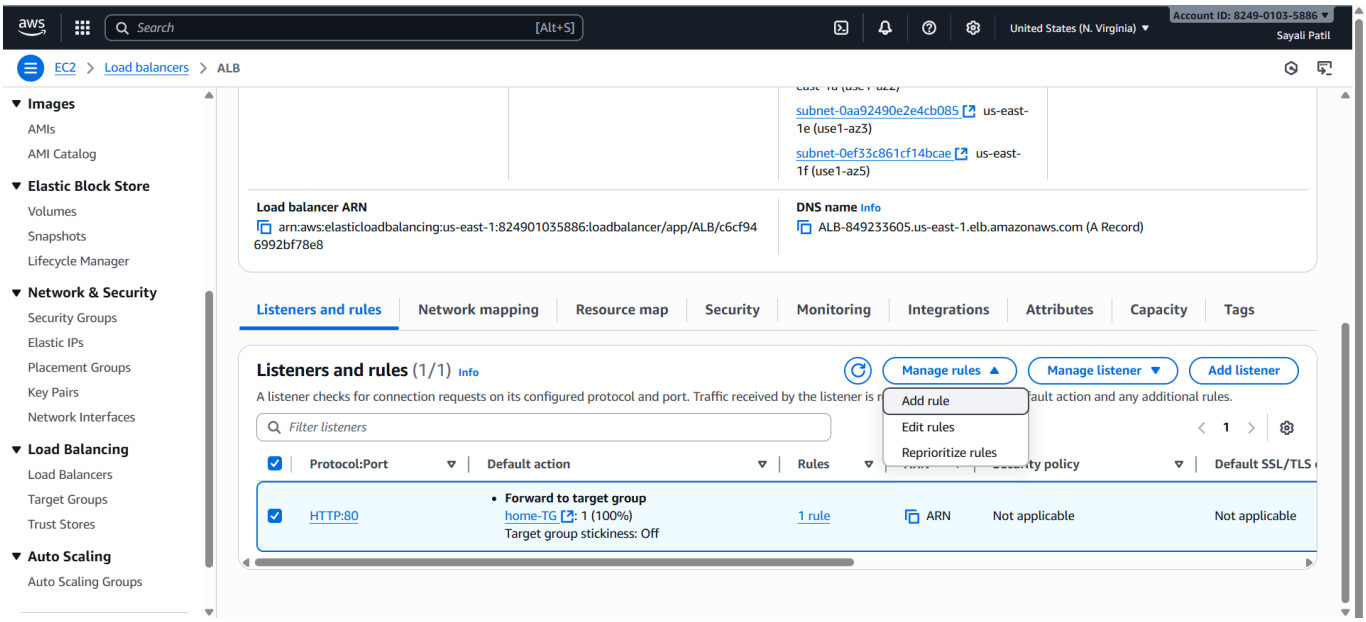



- Click on Create Load Balancer

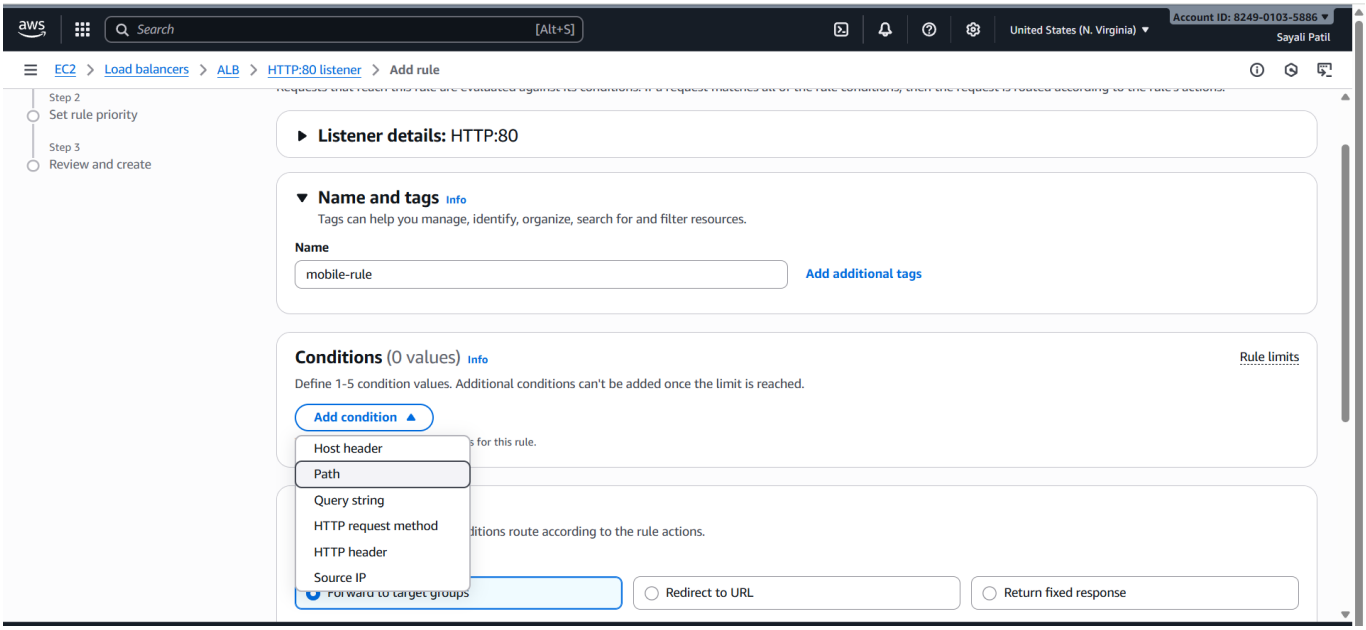



- Successfully created Application loadbalancer
- If you scroll down , you will see the HomeTG as defualt target group in Listners and Rules section.
- To add other Target groups..

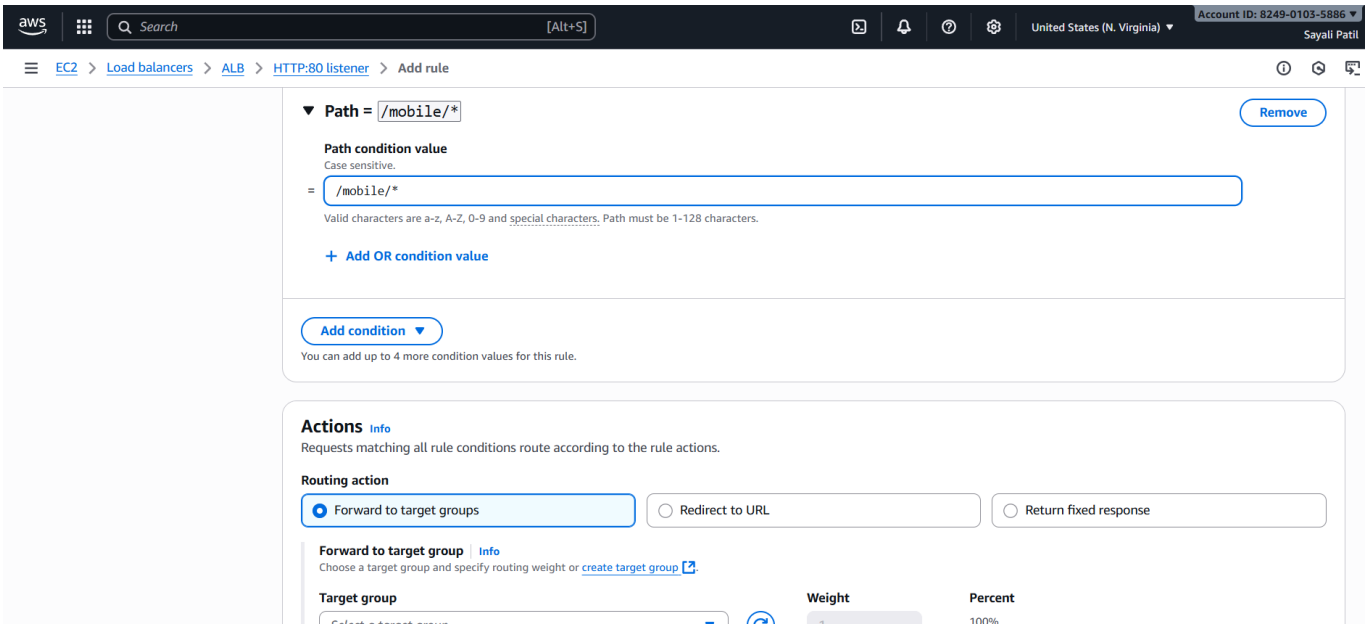
```
* Select HTTP80 i.e., home-TG

* Click on Manage Rules

* Click on Add Rule
```



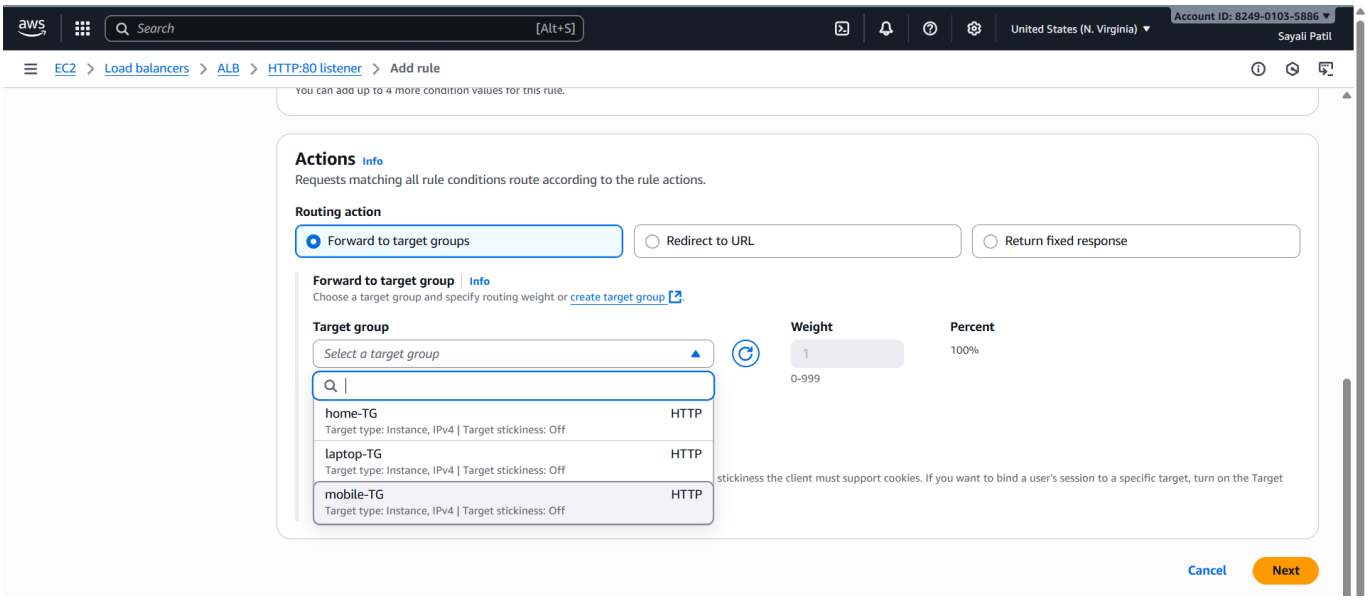

- Give name to the rule and click on add condition and select path.



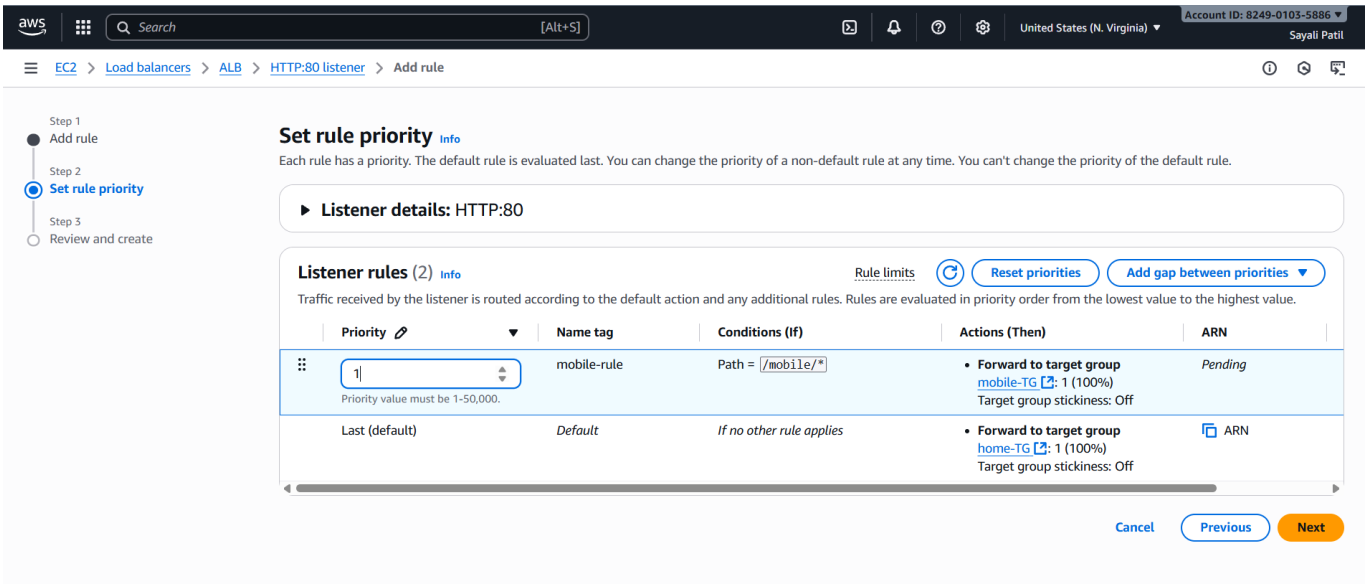

- Give path for mobile application



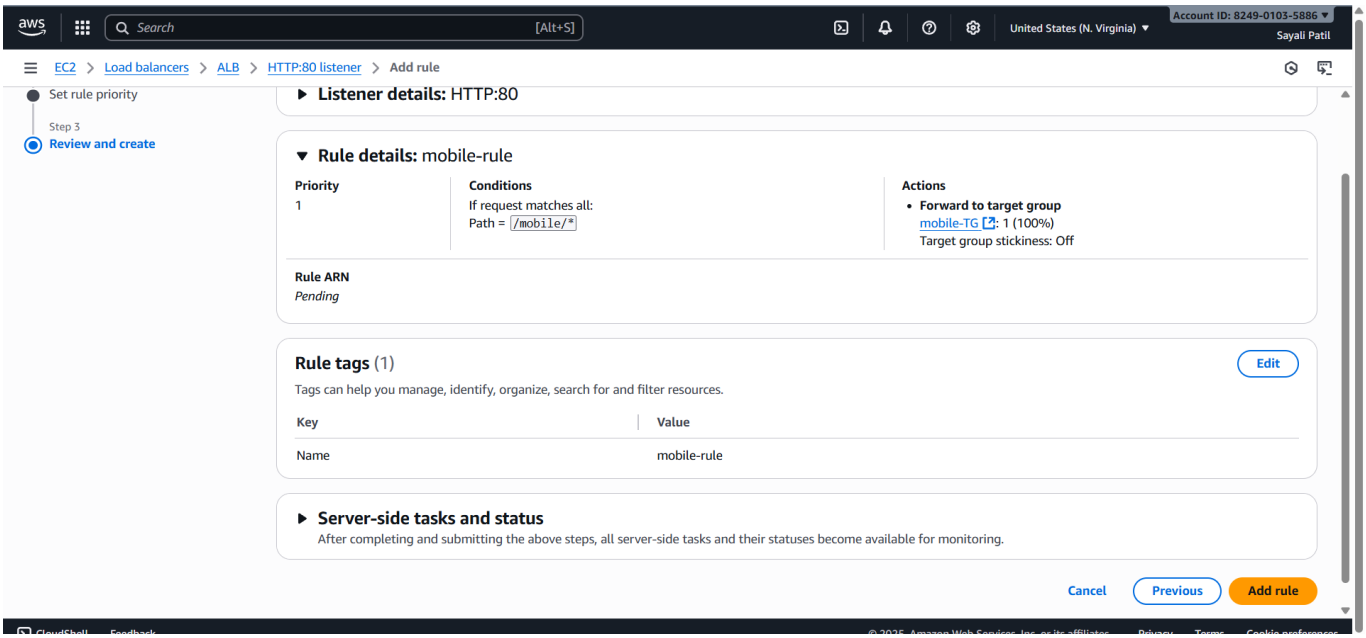

- In Actions section
  - select forward to target group
  - select the target group i.e., mobile-TG
  - Go to next step.



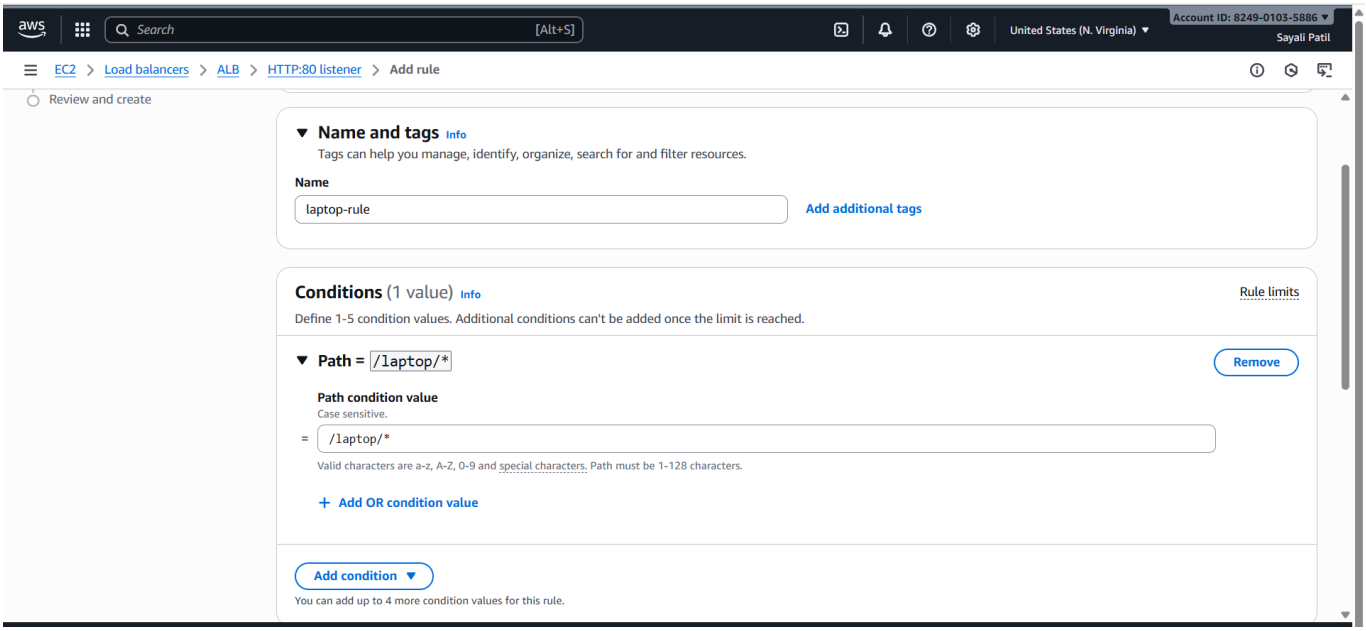

- In Listners rule , give priority 1 for mobileTG
- Click on next

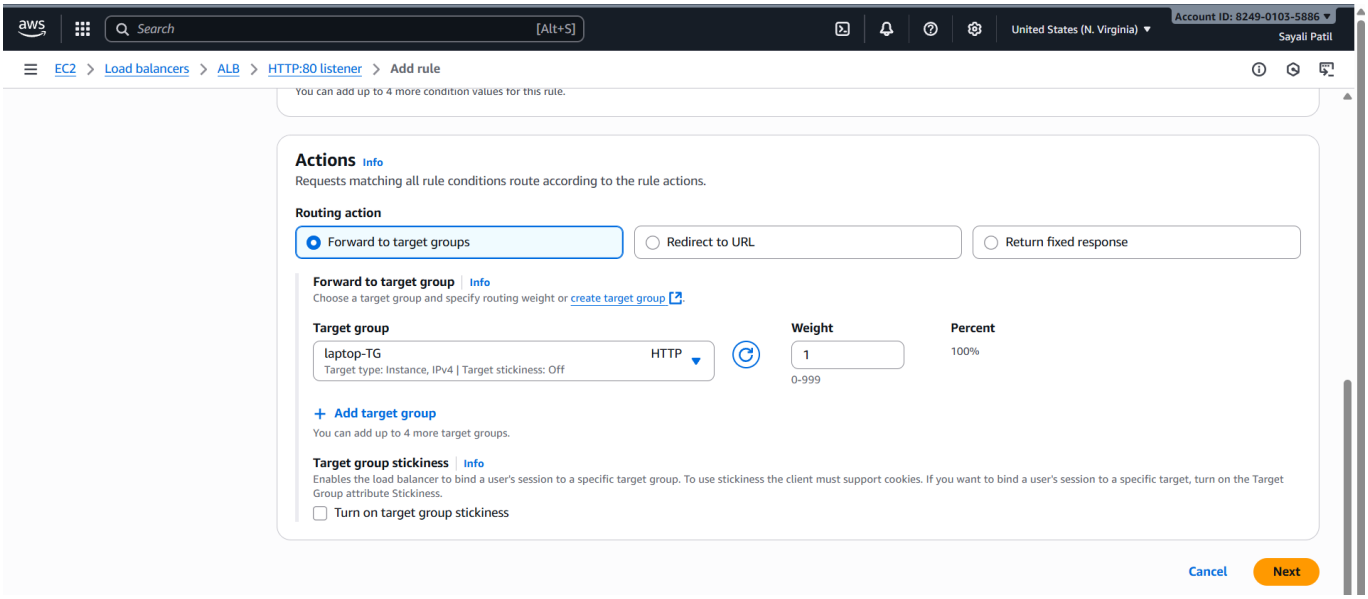


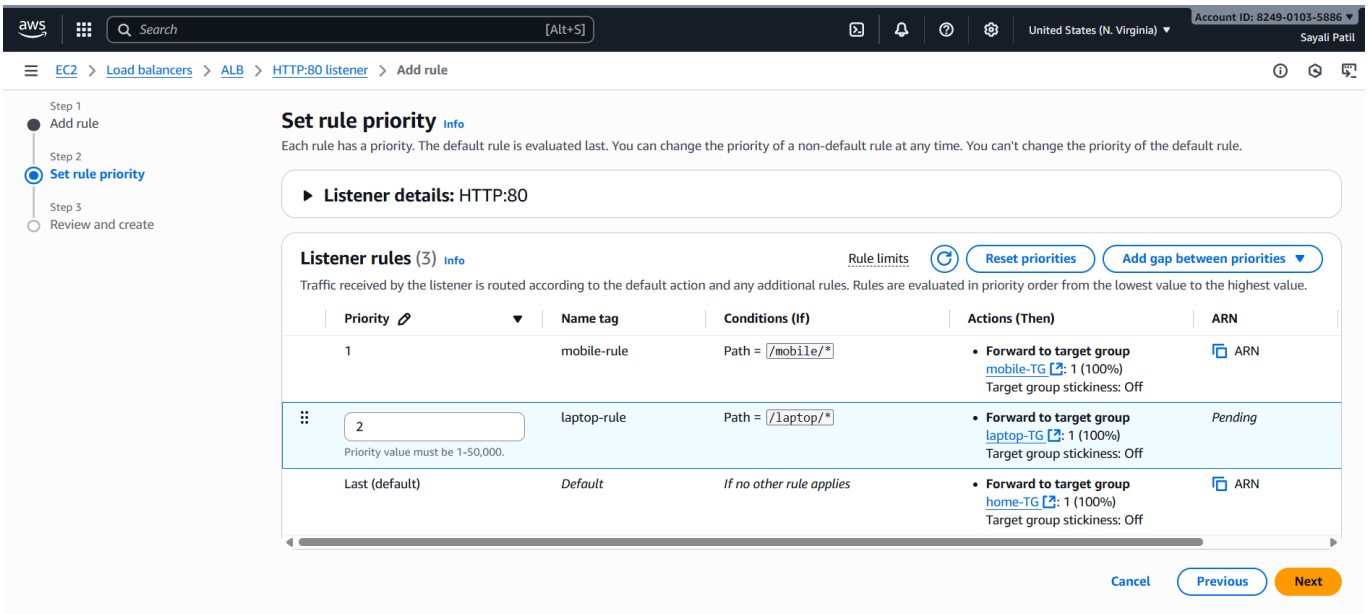

- Click on add rule



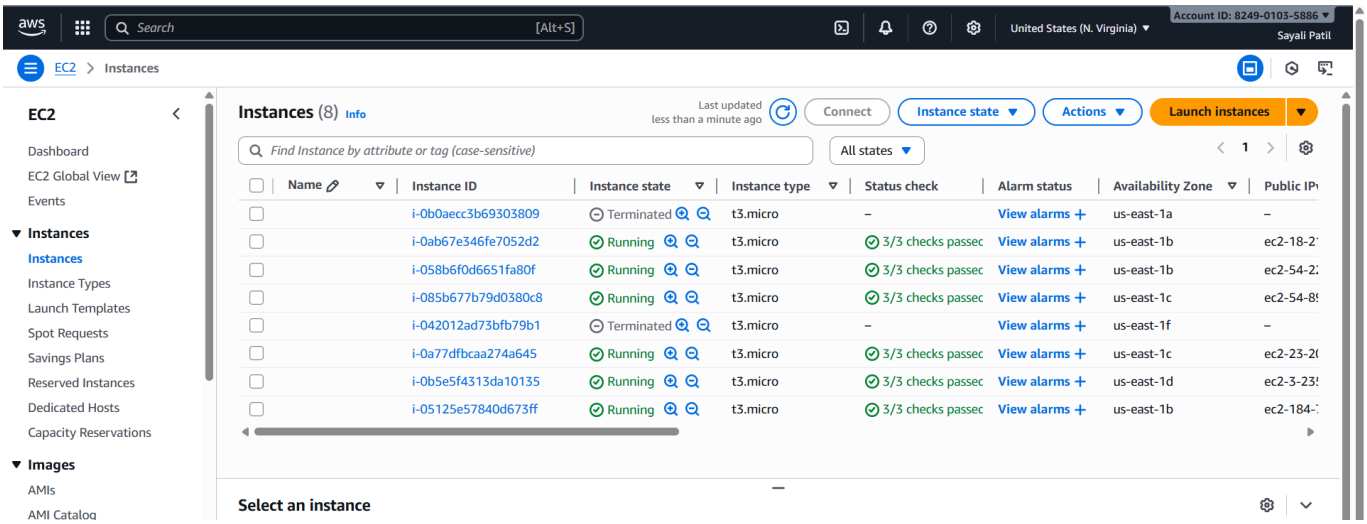

- Successfully connected Mobile Target Group to Loadbalancer
- Similarly connect Laptop target group to load balancer with priority 2



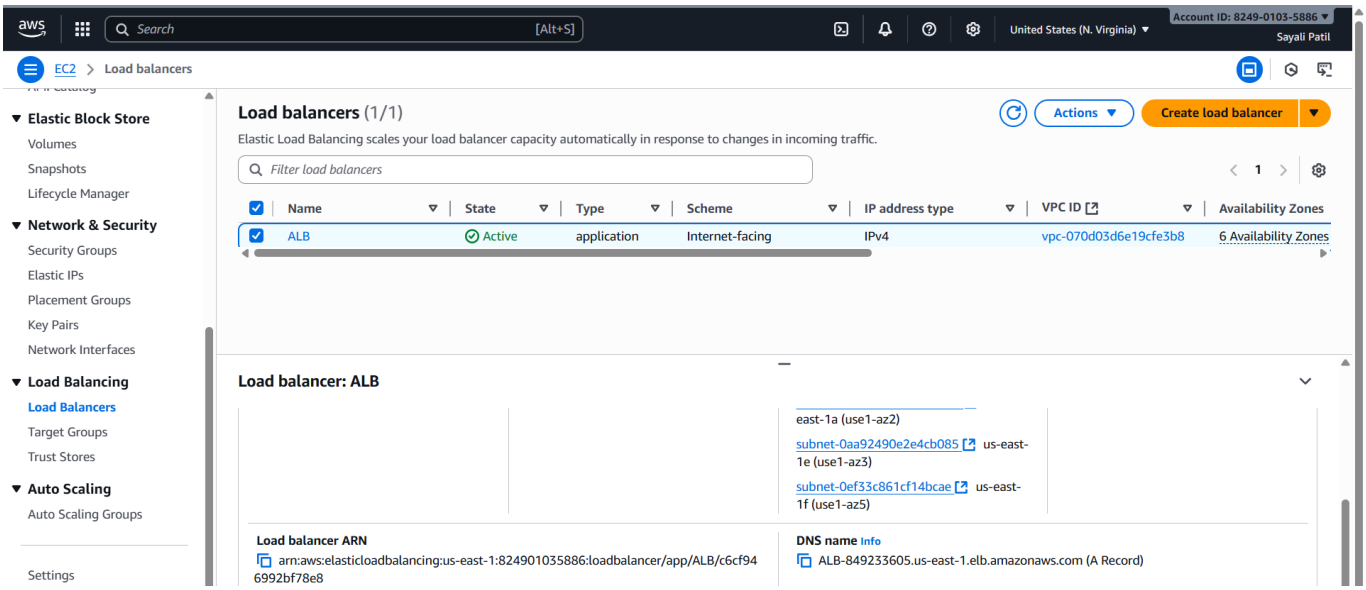




- You can now see that the active instances have been automatically created in the instance dashboard




- To verify if it's functioning correctly:
  - Navigate to the Load Balancers section.
  - Select the load balancer you just created.
  - Scroll down to find and copy the DNS name.
  - Paste the DNS name into your web browser's address bar and press Enter




## Step 7: Result

- ALB DNS successfully routed traffic to all three target groups.

Output of home:

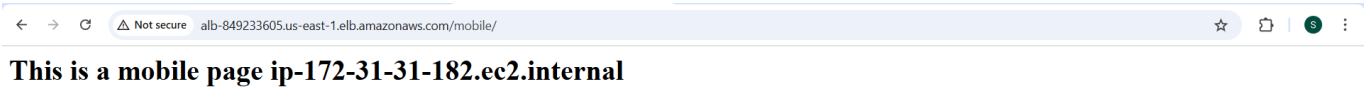

**This is a home page ip-172-31-30-177.ec2.internal**


**This is a home page ip-172-31-42-38.ec2.internal**

Output of laptop:



Output of mobile:


## Technologies Used

1. Amazon EC2 – Virtual machines used to run the application.
2. Application Load Balancer (ALB) – Distributes incoming traffic to different servers.
3. Launch Templates – Predefined settings used to quickly launch EC2 instances.
4. Auto Scaling Groups (ASG) – Automatically adds or removes instances based on demand (supports fixed, automatic, or scheduled scaling).
5. Target Groups – Directs traffic to the correct group of servers.
6. Amazon Linux – The operating system installed on the EC2 instances.
7. Security Groups – Act like a firewall to control incoming and outgoing traffic to the servers.

## Conclusion

This project showcases how integrating an Application Load Balancer (ALB) with multiple Auto Scaling Groups (ASGs) can effectively deliver high availability, fault tolerance, and scalability within the AWS environment.

By working with Static, Scheduled, and Dynamic scaling methods, I gained practical experience in configuring and managing various scaling techniques. The setup efficiently distributed incoming traffic across target groups, automatically adjusted the number of instances based on demand, and served application requests through a unified DNS endpoint.

In summary, this project strengthened my understanding of AWS load balancing, automated scaling, and cloud infrastructure management.