

Pandit Deendayal Energy University, Gandhinagar

School of Technology

Department of Computer Science & Engineering

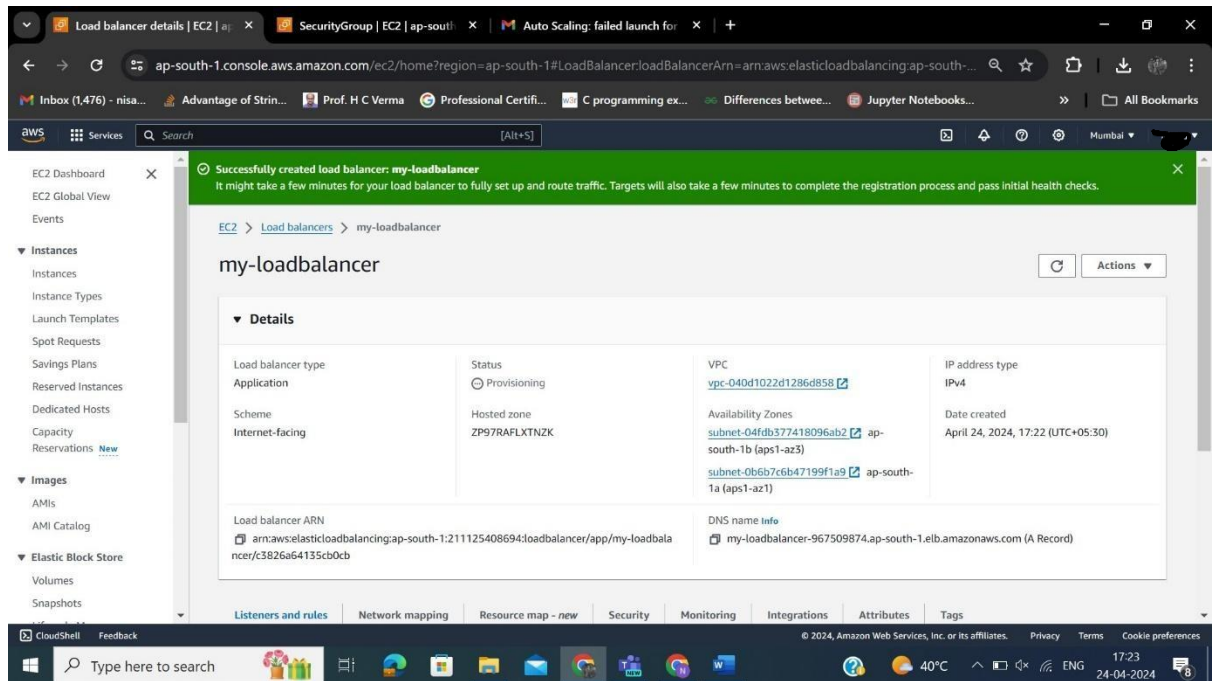
Cloud Computing Lab (20CP322P)

B.Tech-Computer Science & Engineering (Sem-VI)

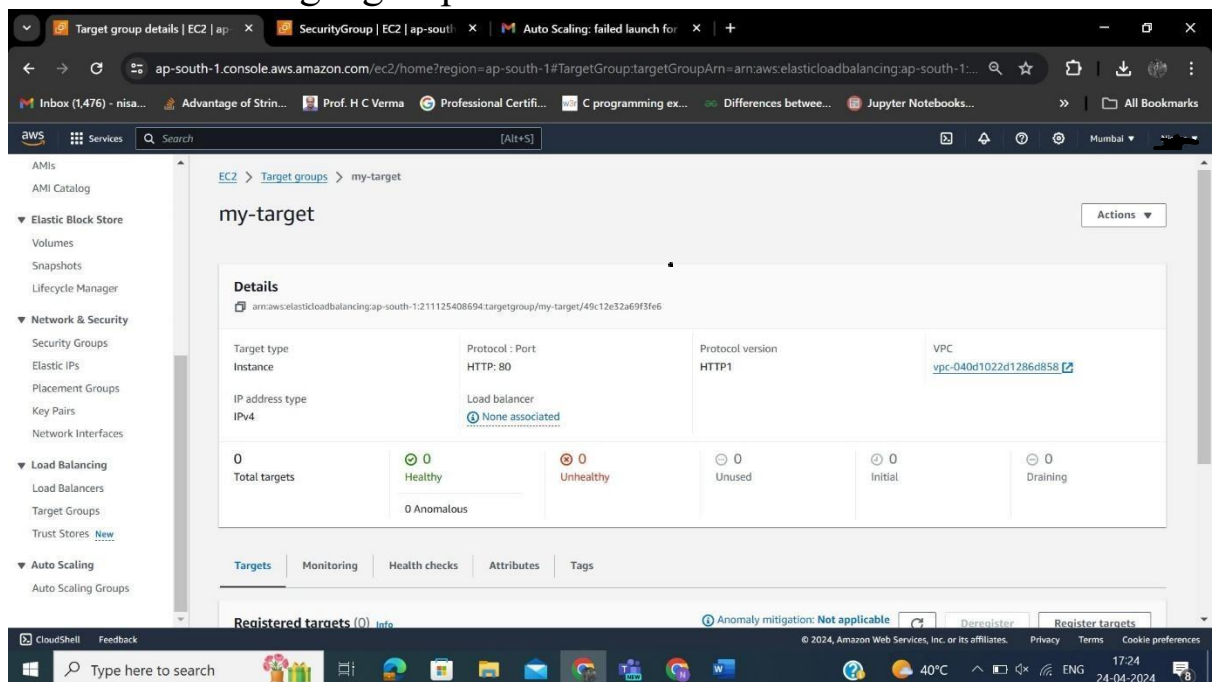
Lab 11 Assignment: Working on AWS Cloud.

Aim: Create infrastructure using auto-scaling, load balancer, SNS, cloud watch where you have to set the metrics for the CPU utilization. If CPU utilization is more than 70% for a 5 minutes then EC2 instance increases as per the requirement and if the CPU utilization goes below 40% then EC2 instances are removed one by one and send a notification to the email.

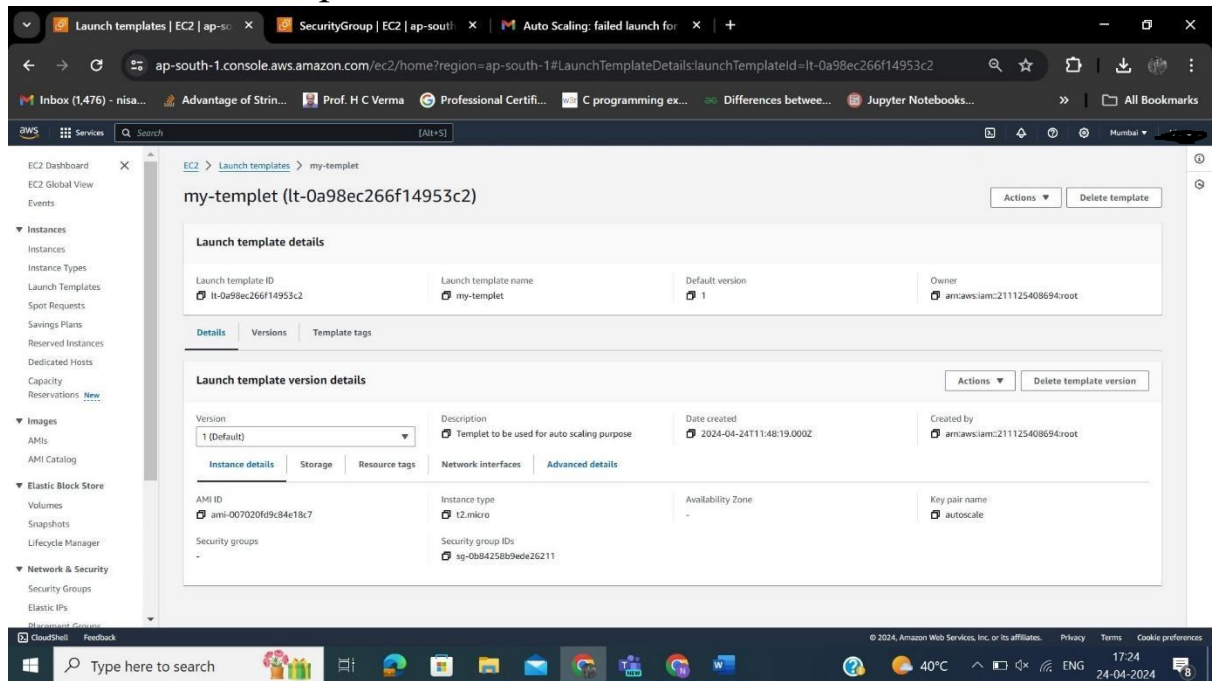
1. Create a load balancer:.



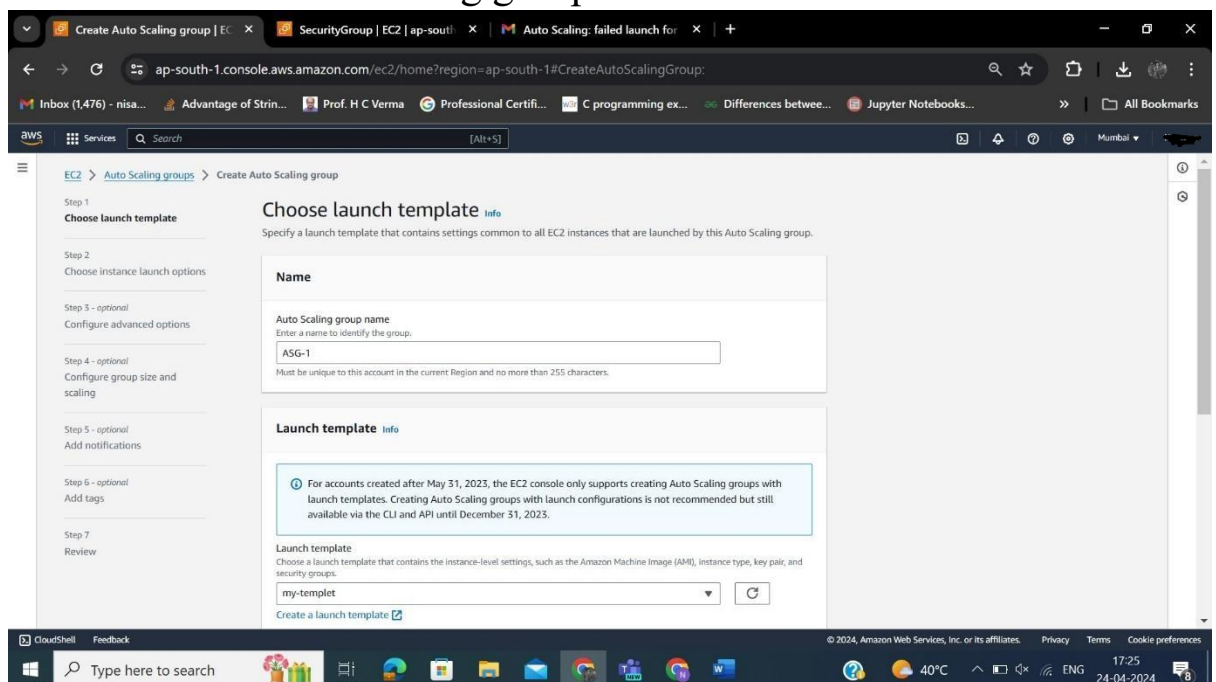
2. Create a target group:

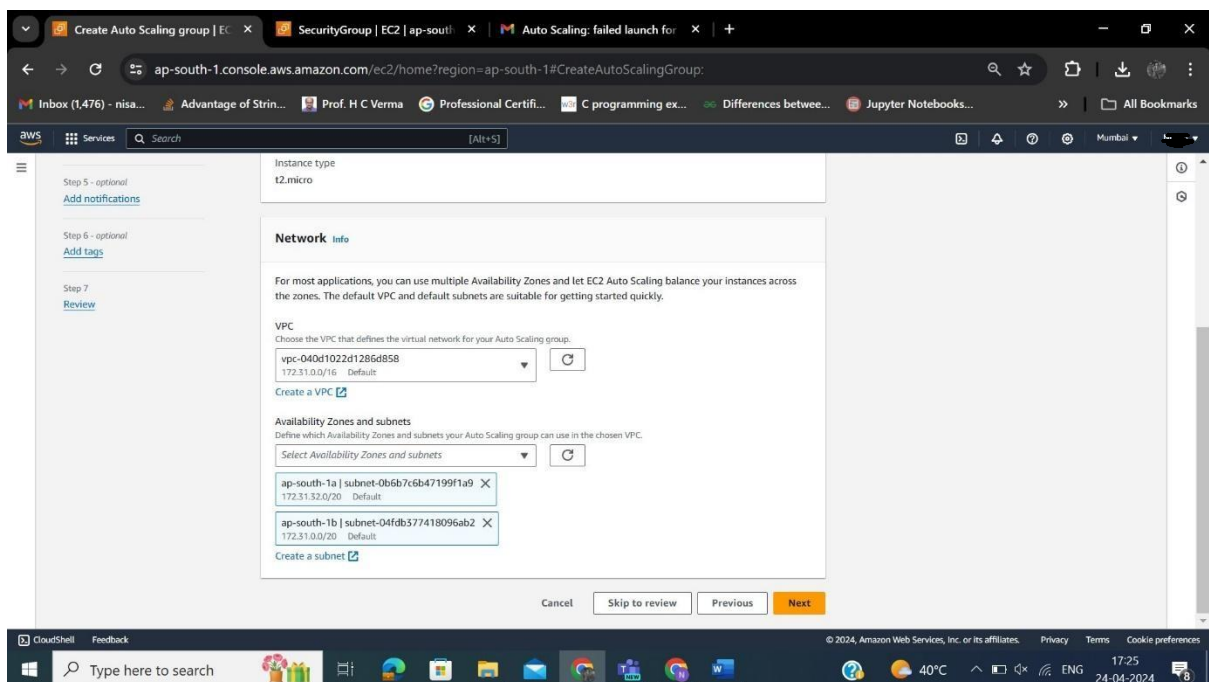
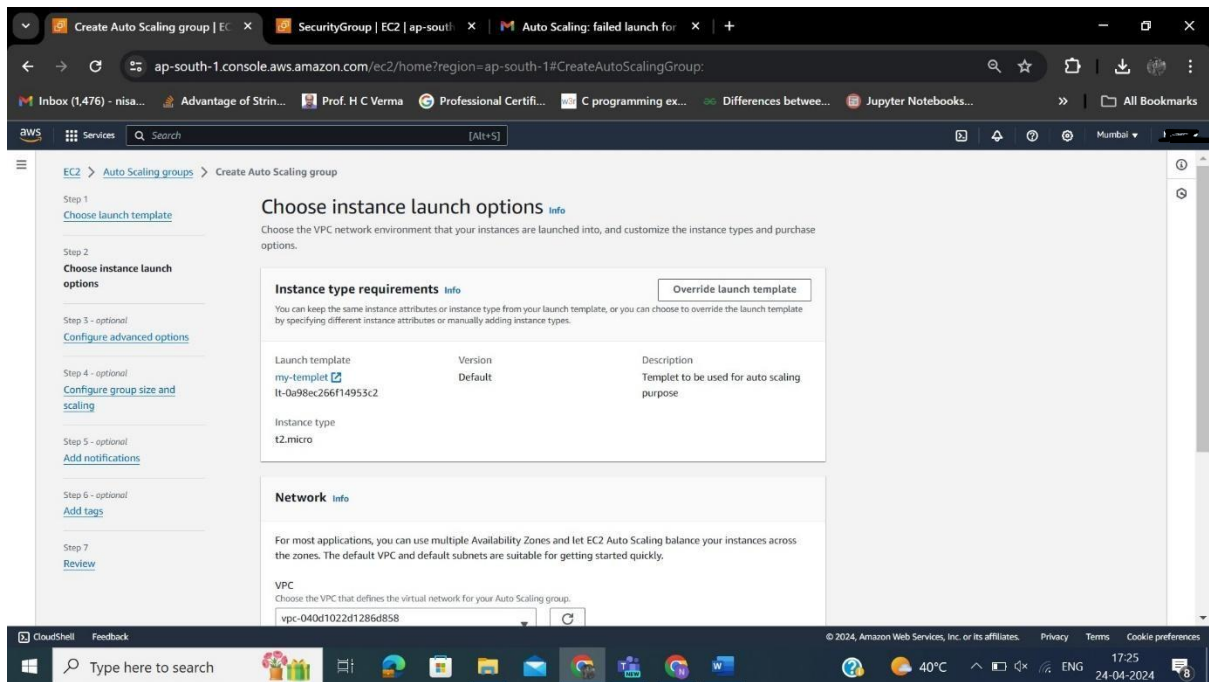


3. Create a template:



4. Create an auto scaling group:





The screenshot shows the AWS Management Console interface for creating an Auto Scaling group. The browser tabs include 'Create Auto Scaling group | EC2', 'SecurityGroup | EC2 | ap-south-1', and 'Auto Scaling: failed launch for...'. The URL is 'ap-south-1.console.aws.amazon.com/ec2/home?region=ap-south-1#CreateAutoScalingGroup:'. The left sidebar shows the navigation menu with steps 1 through 7. The main content area is titled 'Configure advanced options - optional' and includes a description: 'Choose a load balancer to distribute incoming traffic for your application across instances to make it more reliable and easily scalable. You can also set options that give you more control over health check replacements and monitoring.'

Load balancing [Info](#)

Use the options below to attach your Auto Scaling group to an existing load balancer, or to a new load balancer that you define.

- ☐ No load balancer
Traffic to your Auto Scaling group will not be fronted by a load balancer.
- ☒ Attach to an existing load balancer
Choose from your existing load balancers.
- ☐ Attach to a new load balancer
Quickly create a basic load balancer to attach to your Auto Scaling group.

Attach to an existing load balancer

Select the load balancers that you want to attach to your Auto Scaling group.

- ☒ Choose from your load balancer target groups
This option allows you to attach Application, Network, or Gateway Load Balancers.
- ☐ Choose from Classic Load Balancers

Existing load balancer target groups

Only instance target groups that belong to the same VPC as your Auto Scaling group are available for selection.

Select target groups

my-target | HTTP
Application Load Balancer: my-loadbalancer

Health checks

Health checks increase availability by replacing unhealthy instances. When you use multiple health checks, all are evaluated, and if at least one fails, instance replacement occurs.

EC2 health checks

☒ Always enabled

Additional health check types - optional [Info](#)

☐ Turn on Elastic Load Balancing health checks **Recommended**
Elastic Load Balancing monitors whether instances are available to handle requests. When it reports an unhealthy instance, EC2 Auto Scaling can replace it on its next periodic check.

Health check grace period [Info](#)

The screenshot shows the AWS Management Console interface for creating an Auto Scaling group, specifically Step 6: 'Attach to an existing load balancer'. The browser tabs and URL are the same as the previous screenshot. The left sidebar shows steps 1 through 7, with Step 6 highlighted. The main content area is titled 'Attach to an existing load balancer' and includes a description: 'Select the load balancers that you want to attach to your Auto Scaling group.'

- ☒ Choose from your load balancer target groups
This option allows you to attach Application, Network, or Gateway Load Balancers.
- ☐ Choose from Classic Load Balancers

Existing load balancer target groups

Only instance target groups that belong to the same VPC as your Auto Scaling group are available for selection.

Select target groups

my-target | HTTP
Application Load Balancer: my-loadbalancer

Health checks

Health checks increase availability by replacing unhealthy instances. When you use multiple health checks, all are evaluated, and if at least one fails, instance replacement occurs.

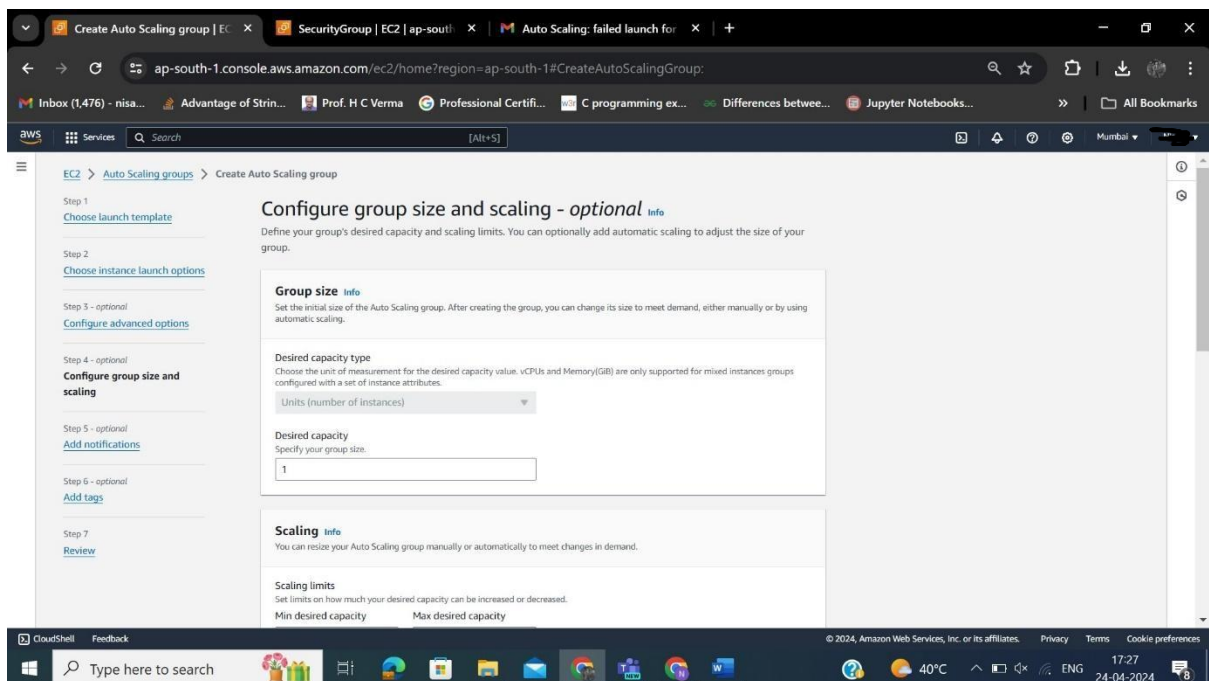
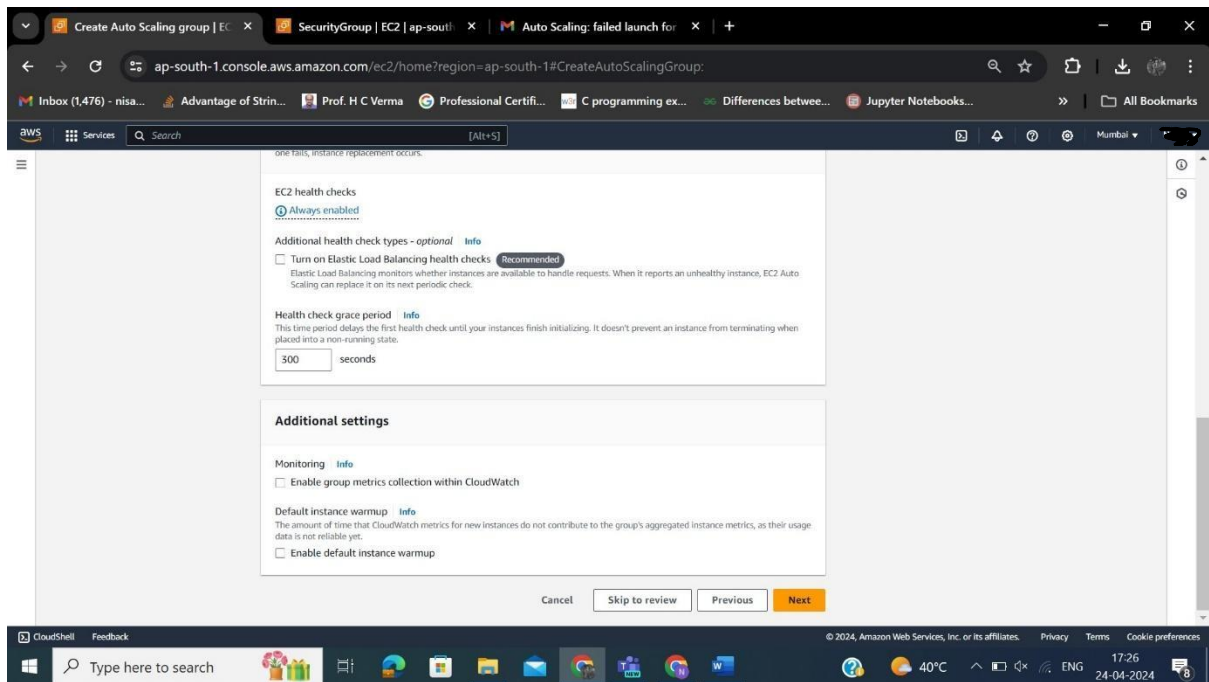
EC2 health checks

☒ Always enabled

Additional health check types - optional [Info](#)

☐ Turn on Elastic Load Balancing health checks **Recommended**
Elastic Load Balancing monitors whether instances are available to handle requests. When it reports an unhealthy instance, EC2 Auto Scaling can replace it on its next periodic check.

Health check grace period [Info](#)



Scaling Info
You can resize your Auto Scaling group manually or automatically to meet changes in demand.

Scaling limits
Set limits on how much your desired capacity can be increased or decreased.

Min desired capacity:
Max desired capacity:
Equal or less than desired capacity
Equal or greater than desired capacity

Automatic scaling - optional
Choose whether to use a target tracking policy. You can set up other metric-based scaling policies and scheduled scaling after creating your Auto Scaling group.

☒ No scaling policies
Your Auto Scaling group will remain at its initial size and will not dynamically resize to meet demand.

☐ Target tracking scaling policy
Choose a CloudWatch metric and target value and let the scaling policy adjust the desired capacity in proportion to the metric's value.

Instance maintenance policy
Control your Auto Scaling group's availability during instance replacement events. This includes health checks, instance refreshes, maximum instance lifetime features and events that happen automatically to keep your group balanced, called rebalancing events.

Choose a replacement behavior depending on your availability requirements.

☒ Mixed behavior
☐ No policy
☐ Prioritize availability
☐ Launch before
☐ Control costs
☐ Terminate and
☐ Flexible
☐ Custom behavior

Instance maintenance policy
Control your Auto Scaling group's availability during instance replacement events. This includes health checks, instance refreshes, maximum instance lifetime features and events that happen automatically to keep your group balanced, called rebalancing events.

Choose a replacement behavior depending on your availability requirements.

☒ Mixed behavior
For rebalancing events, new instances will launch before terminating others. For all other events, instances terminate and launch at the same time.

☐ Prioritize availability
☐ Launch before
Launch new instances and wait for them to be ready before terminating others. This allows you to go above your desired capacity by a given percentage and may temporarily increase costs.

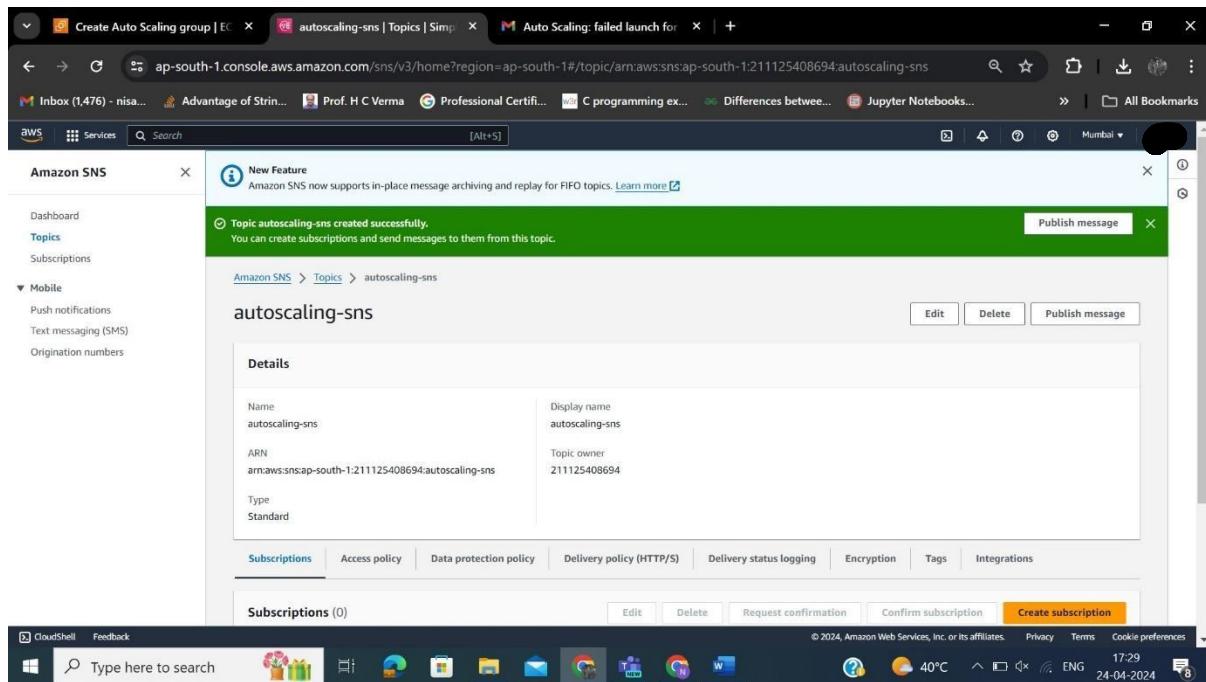
☐ Control costs
☐ Terminate and launch
Terminate and launch instances at the same time. This allows you to go below your desired capacity by a given percentage and may temporarily reduce availability.

☐ Flexible
☐ Custom behavior
Set custom values for the minimum and maximum amount of available capacity. This gives you greater flexibility in setting how far below and over your desired capacity EC2 Auto Scaling goes when replacing instances.

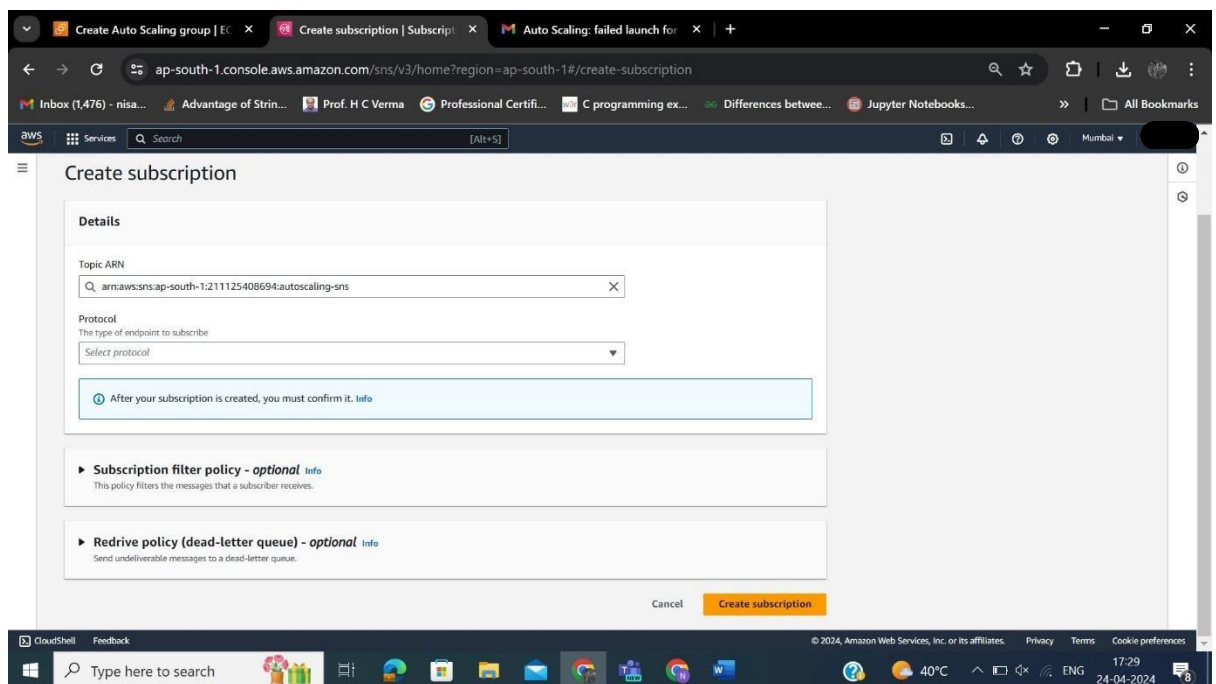
Instance scale-in protection
Scale-in protection prevents newly launched instances from being terminated by scaling activities. Make sure to remove scale-in protection for the group or individual instances when instances are ready to be terminated.

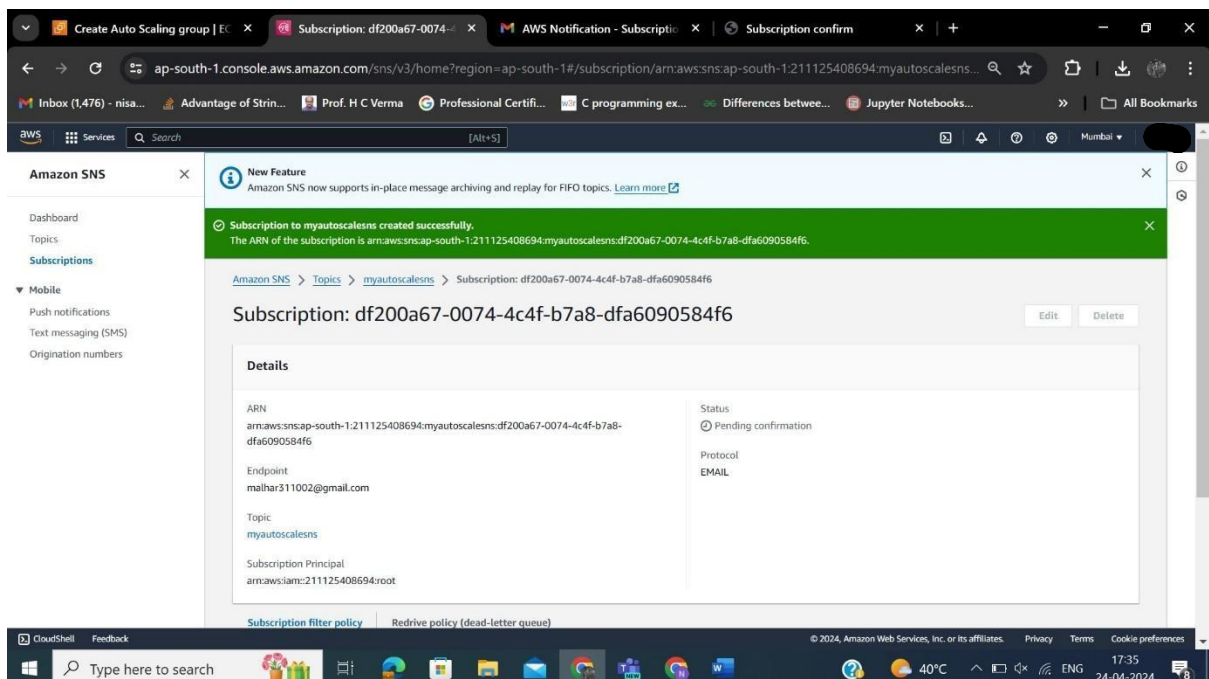
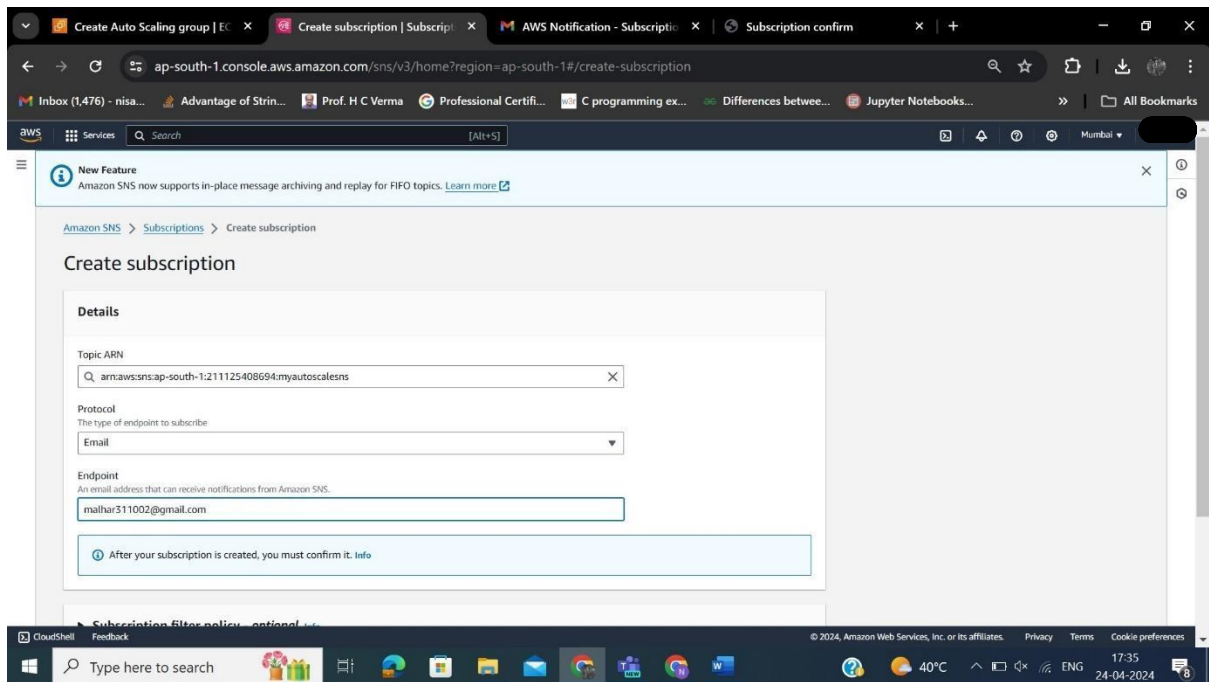
☐ Enable instance scale-in protection

Cancel Skip to review Previous Next

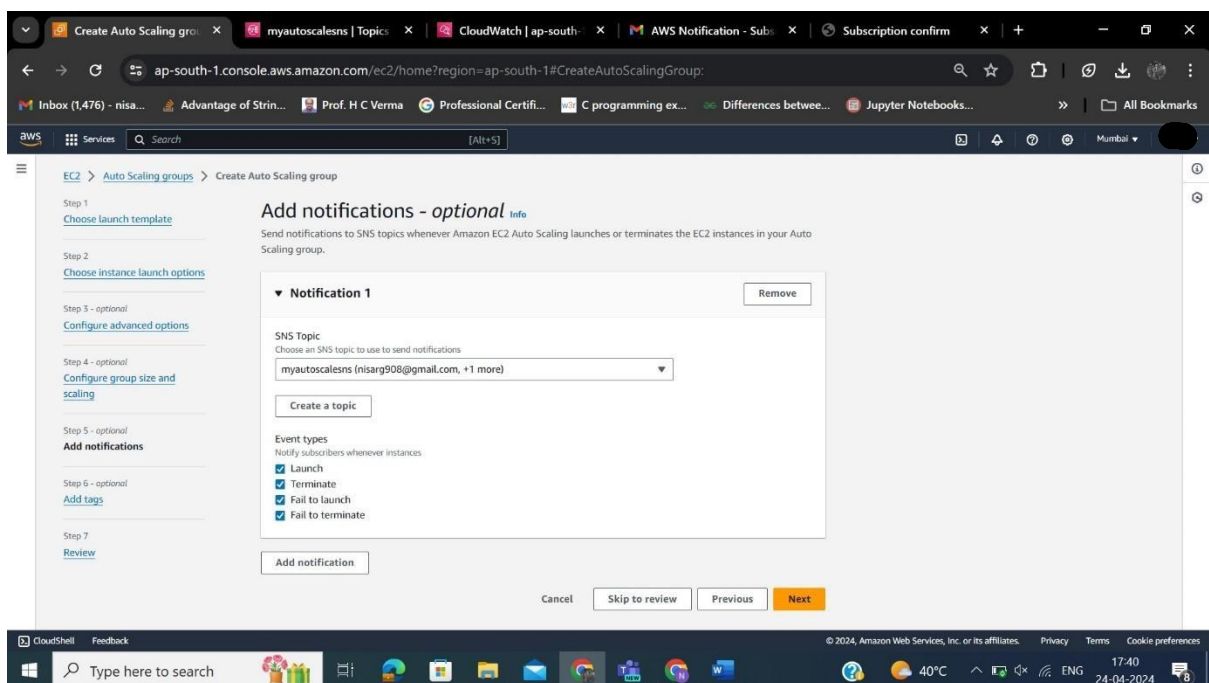
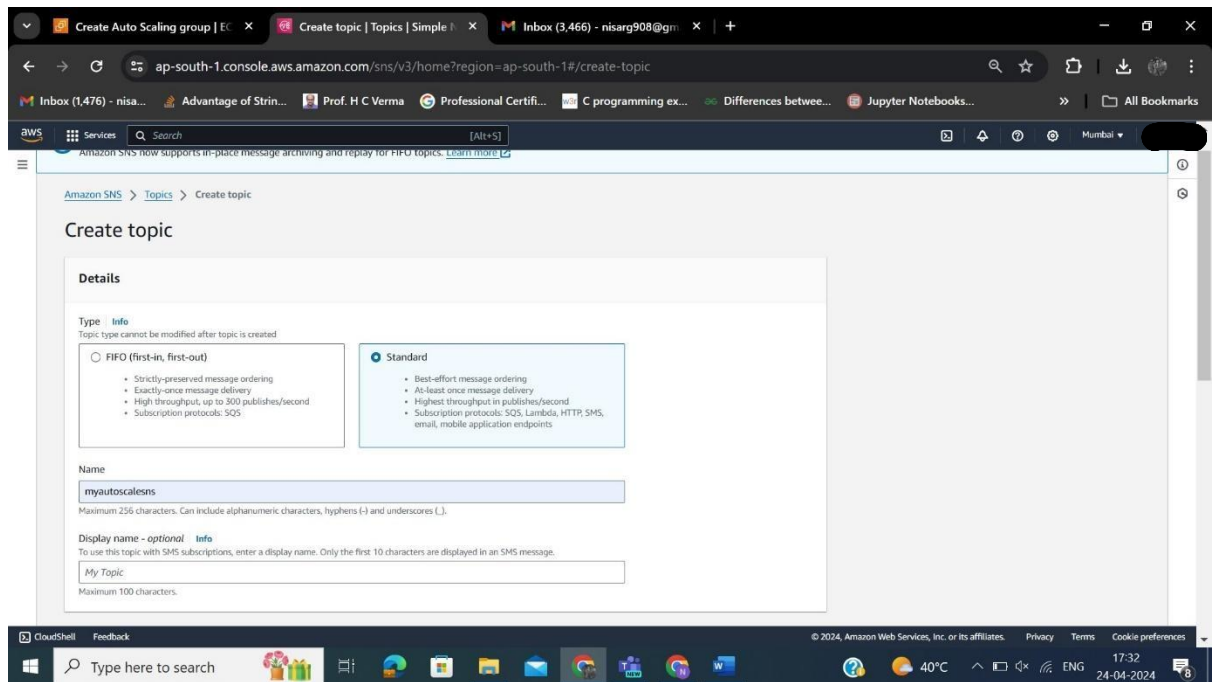


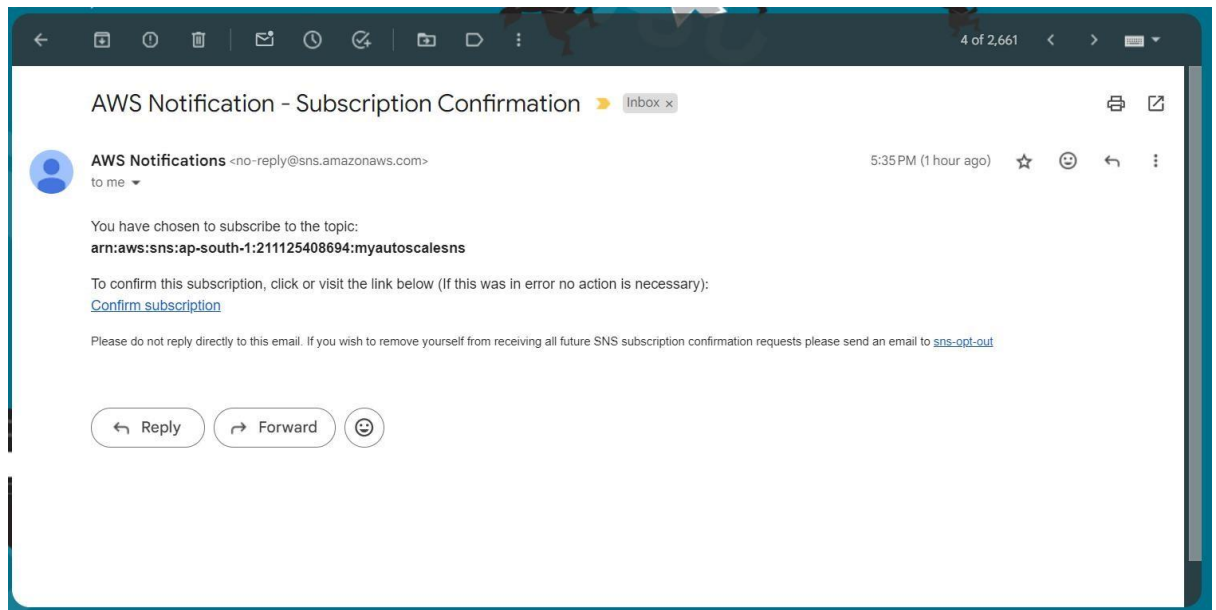
5. Create an SNS:



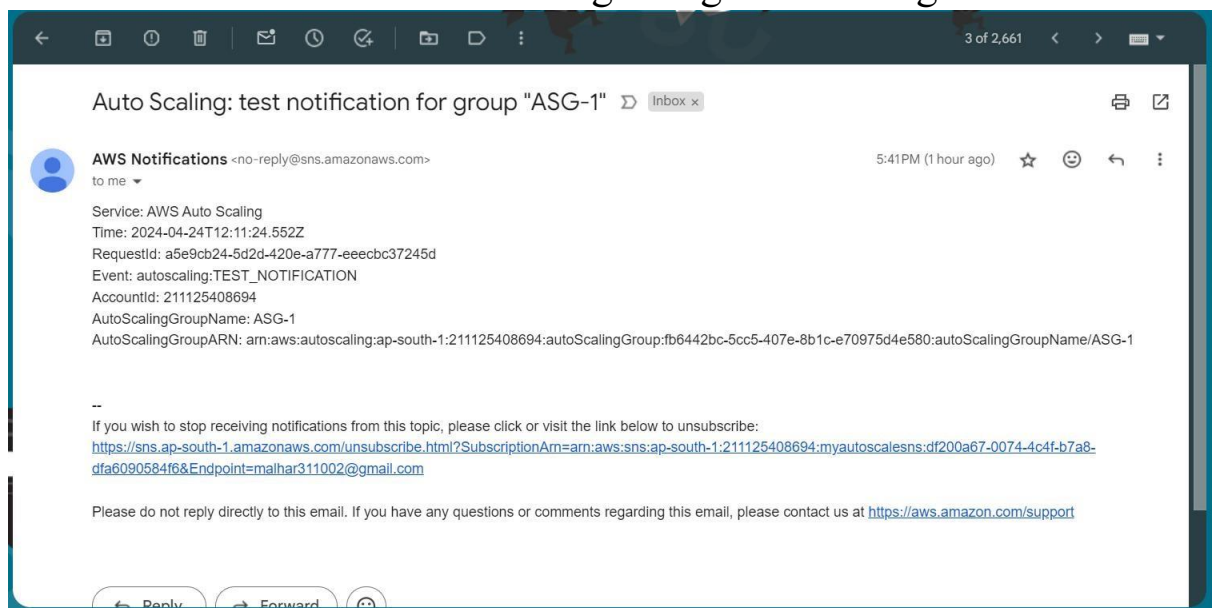


6. Create an SNS topic:

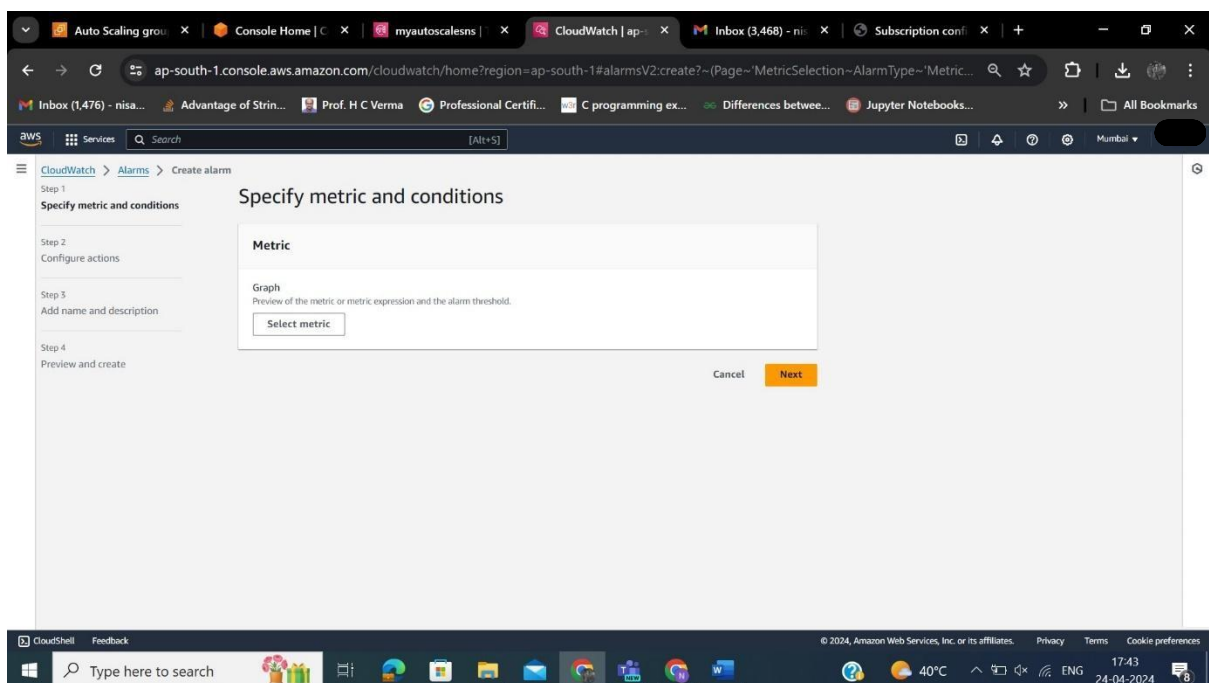
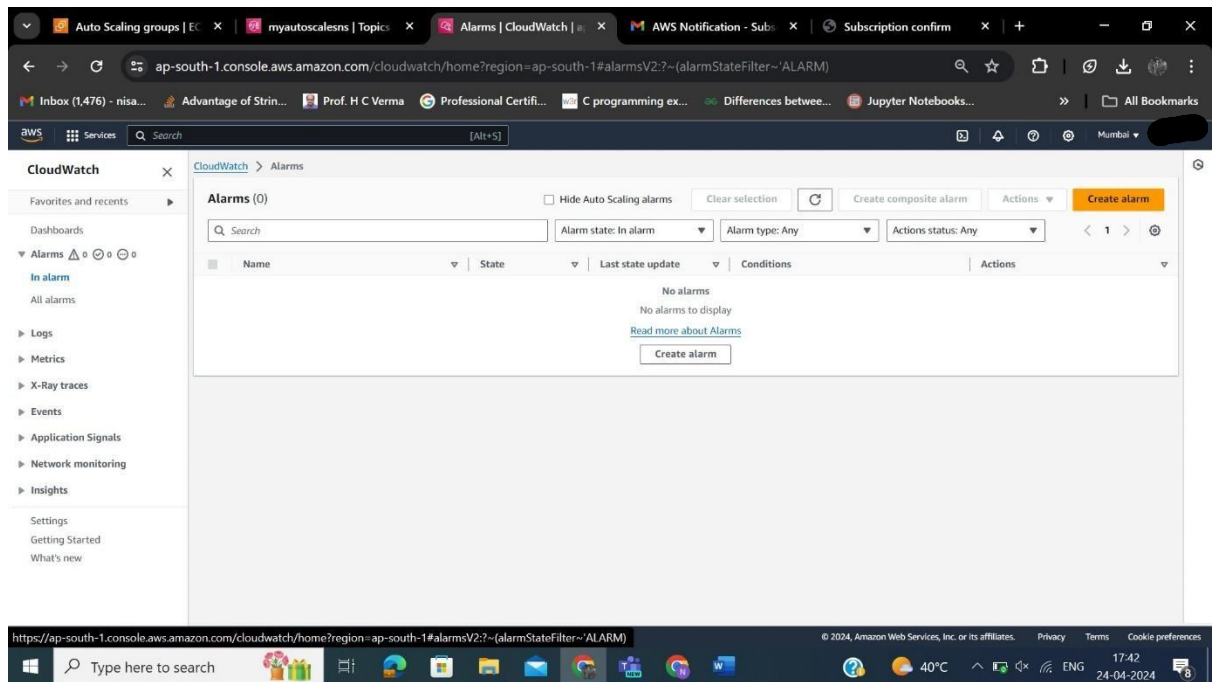




7. You will receive a mail regarding auto scaling

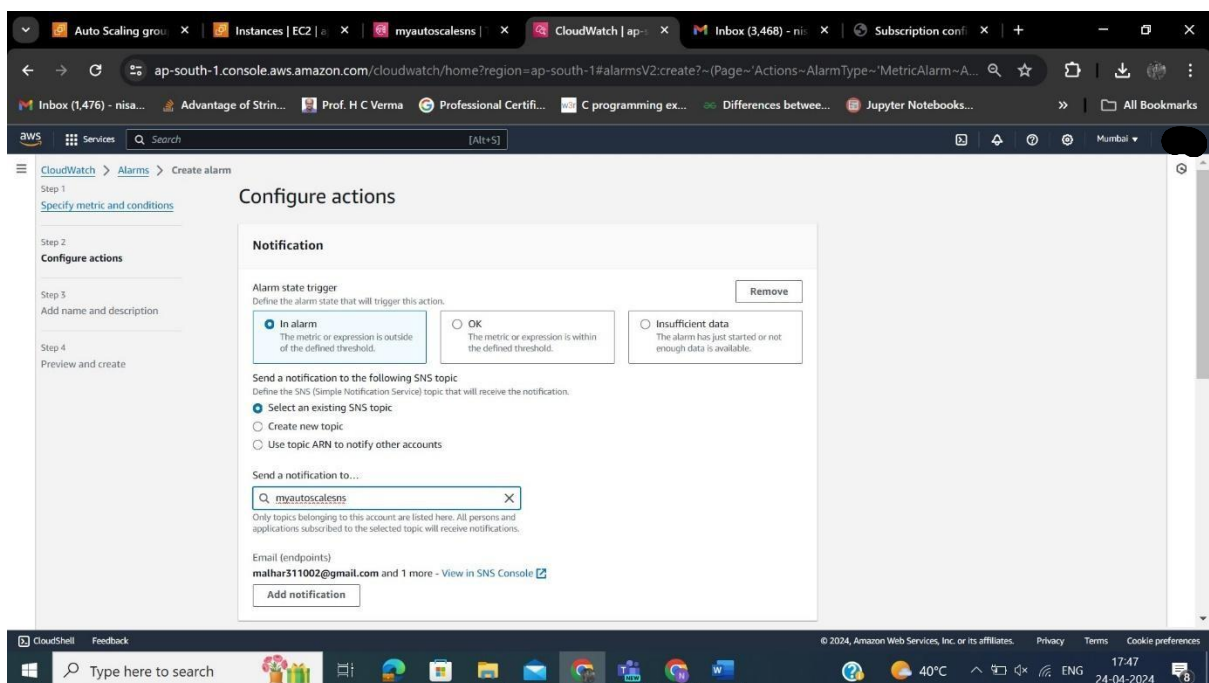
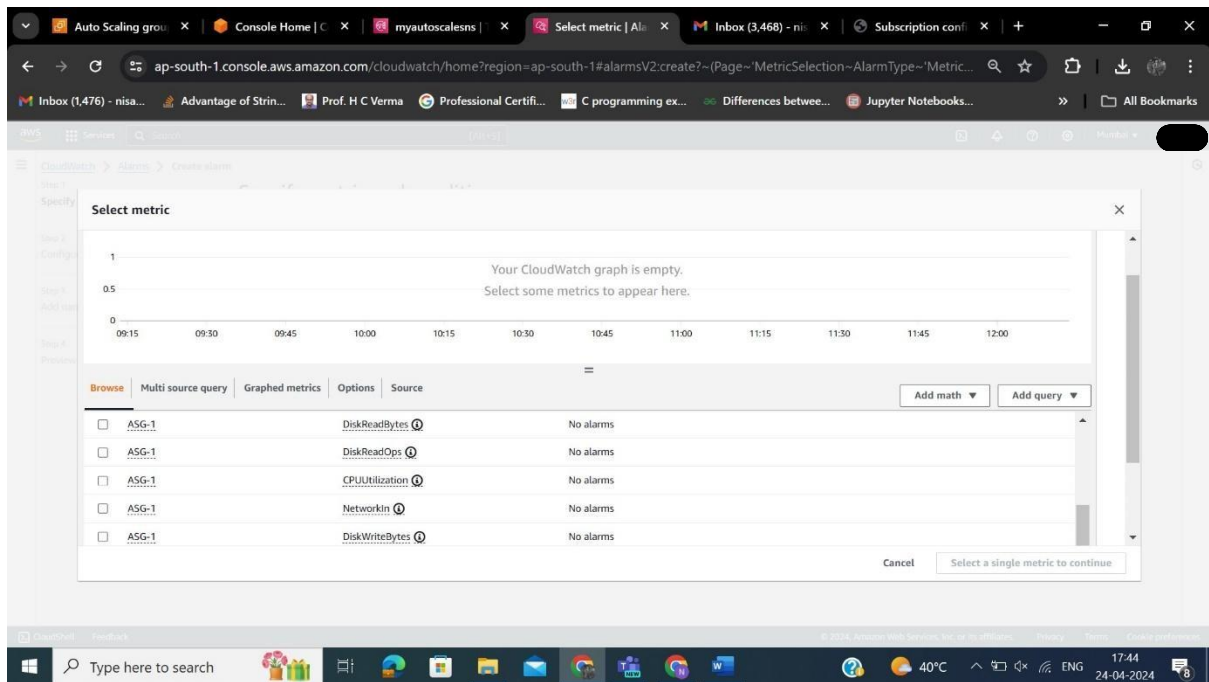


8. Create an alarm



The screenshot shows the AWS CloudWatch console with the 'Select metric' dialog open. The dialog displays a graph area with the text 'Your CloudWatch graph is empty. Select some metrics to appear here.' Below the graph, there are tabs for 'Browse', 'Multi source query', 'Graphed metrics', 'Options', and 'Source'. The 'Browse' tab is selected, showing a list of metrics for the 'Mumbai' region. The list includes 'ApplicationELB' (17), 'EBS' (9), 'EC2' (28), and 'Events' (1). There are buttons for 'Add math', 'Add query', 'Alarm recommendations', 'Graph with SQL', and 'Graph search'. At the bottom, there are 'Cancel' and 'Select a single metric to continue' buttons.

The screenshot shows the AWS CloudWatch console with the 'Select metric' dialog open. The dialog displays a graph area with the text 'Your CloudWatch graph is empty. Select some metrics to appear here.' Below the graph, there are tabs for 'Browse', 'Multi source query', 'Graphed metrics', 'Options', and 'Source'. The 'Browse' tab is selected, showing a list of metrics for the 'Mumbai' region. The list is filtered to show 'By Auto Scaling Group' (14) and 'Per-Instance Metrics' (14). There are buttons for 'Add math', 'Add query', 'Alarm recommendations', 'Graph with SQL', and 'Graph search'. At the bottom, there are 'Cancel' and 'Select a single metric to continue' buttons.



The screenshot shows the AWS CloudWatch Alarms console. A green banner at the top indicates 'Successfully created alarm Below-40-utilization.' The left sidebar shows the 'Alarms' section. The main content area displays a table of alarms:

Name	State	Last state update	Conditions	Actions
Below-40-utilization	Insufficient data	2024-04-24 12:23:12	CPUUtilization <= 40 for 1 datapoints within 5 minutes	Actions enabled
Above-75-utilization	OK	2024-04-24 12:22:04	CPUUtilization >= 75 for 1 datapoints within 5 minutes	Actions enabled

9. Check the auto scaling groups

The screenshot shows the AWS Auto Scaling Groups console. The left sidebar shows the 'Auto Scaling groups' section. The main content area displays a table of auto scaling groups:

Name	Launch template/configuration	Instances	Status	Desired capacity	Min	Max	Availability Zones
ASG-1	my-template Version Default	0	Updating capacity...	1	1	3	ap-south-1b, ap-south-1a

ASG-1

Details Activity **Automatic scaling** Instance management Monitoring Instance refresh

Scaling policies resize your Auto Scaling group to meet changes in demand. With reactive dynamic scaling policies, you can track specific CloudWatch metrics and take action when the CloudWatch alarm threshold is met. Use predictive scaling policies along with dynamic scaling policies in the following situations: when your application demand changes quickly, but with a recurring pattern, or when your EC2 instances require more time to initialize.

Dynamic scaling policies (0) [Info](#) [Refresh](#) [Actions](#) [Create dynamic scaling policy](#)

No dynamic scaling policies have been created

Dynamic scaling policies use real-time data to scale your group based on configurable metrics.

[Create dynamic scaling policy](#)

Predictive scaling policies (0) [Info](#) [Refresh](#) [Actions](#) [Create predictive scaling policy](#)

Evaluation period

CloudWatch

Successfully created alarm Below-40-utilization. [View alarm](#)

CloudWatch > Alarms

Alarms (2) ☐ Hide Auto Scaling alarms [Clear selection](#) [Refresh](#) [Create composite alarm](#) [Actions](#) [Create alarm](#)

<input type="checkbox"/>	Name	State	Last state update	Conditions	Actions
<input type="checkbox"/>	Below-40-utilization	In alarm	2024-04-24 12:23:31	CPUUtilization <= 40 for 1 datapoints within 5 minutes	Actions enabled
<input type="checkbox"/>	Above-75-utilization	OK	2024-04-24 12:22:04	CPUUtilization >= 75 for 1 datapoints within 5 minutes	Actions enabled

The first screenshot shows the 'Dynamic scaling policies (2)' page for an Auto Scaling Group. It lists two policies:

- Above75utilization**: Simple scaling, Enabled. Breaches the alarm threshold: CPUUtilization ≥ 75 for 1 consecutive periods of 300 seconds for the metric dimensions: AutoScalingGroupName = ASG-1. Action: Add 1 capacity units.
- Below40utilization**: Simple scaling, Enabled. Breaches the alarm threshold: CPUUtilization ≤ 40 for 1 consecutive periods of 300 seconds for the metric dimensions: AutoScalingGroupName = ASG-1. Action: Remove 1 capacity units.

The second screenshot shows the 'Instances (3)' page for the same Auto Scaling Group. It displays a table of instances:

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4
	i-05b5afd10428adb76	Shutting-d...	t2.micro	-	View alarms +	ap-south-1b	ec2-65-2-150-95.ap-so...	65.2.150.95
	i-098b51c0f08bb6259	Shutting-d...	t2.micro	-	View alarms +	ap-south-1a	ec2-3-110-215-221.ap-...	3.110.215.2
	i-089b29dkt98b76899	Running	t2.micro	2/2 checks passed	View alarms +	ap-south-1a	ec2-3-108-191-30.ap-s...	3.108.191.3

10. The metrics are as follows:
11. As you can see, the auto scaling group has been applied in accordance to the given metrics and limits.