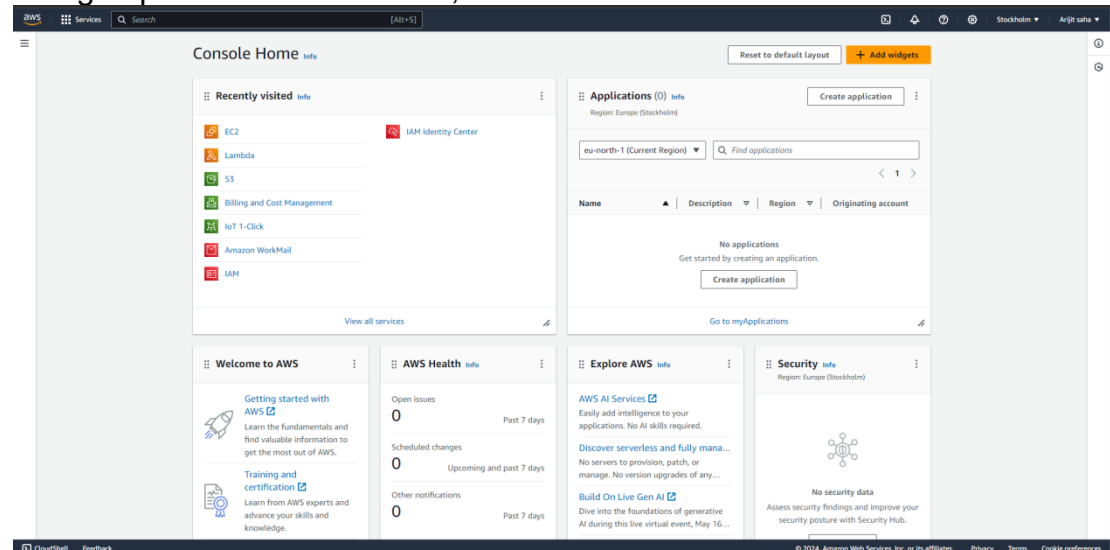


PROBLEM STATEMENT:

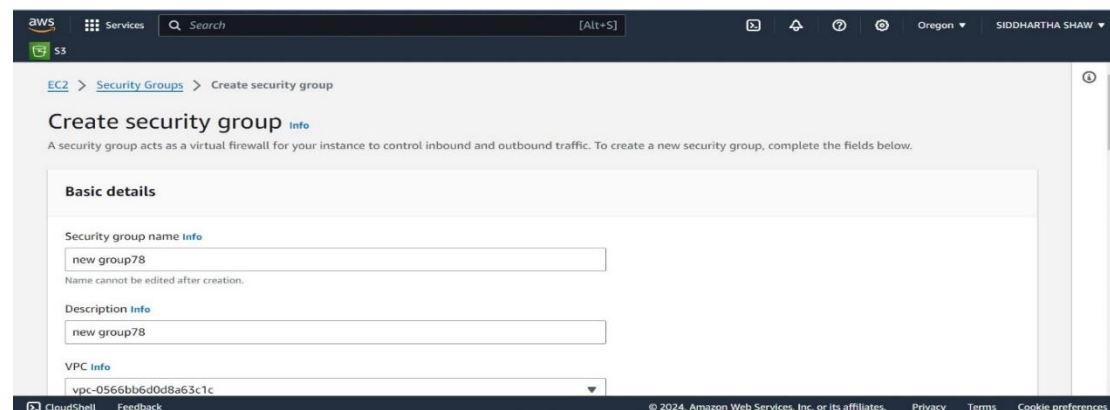
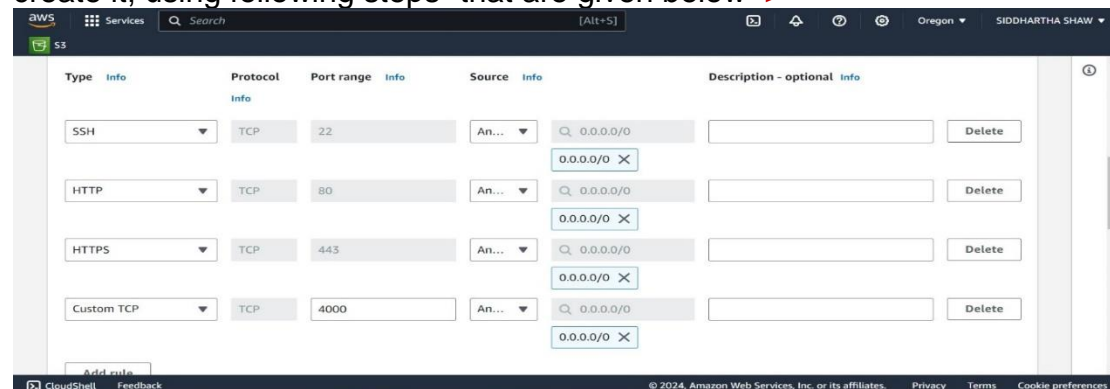
11) Build scaling plans in AWS that balance the load on different EC2 instances.

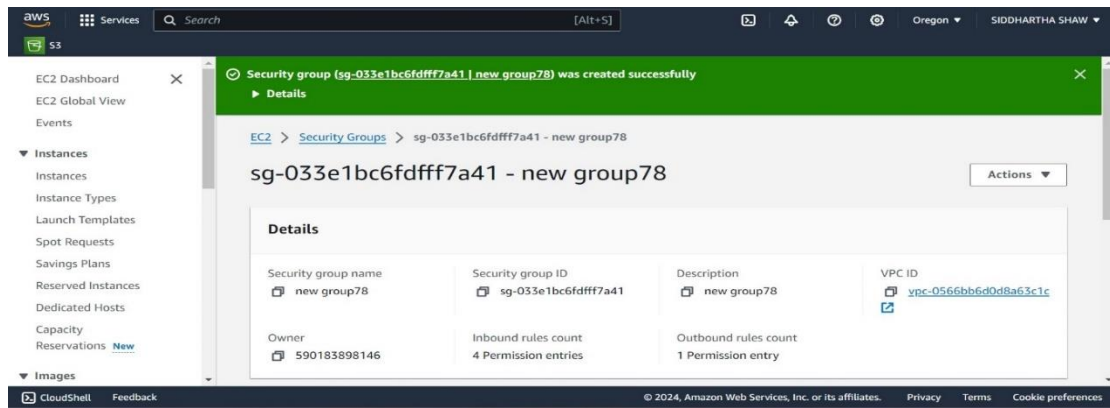
Steps to build scaling plans:

1. Sign up for an AWS account, search for 'EC2' then click on it.

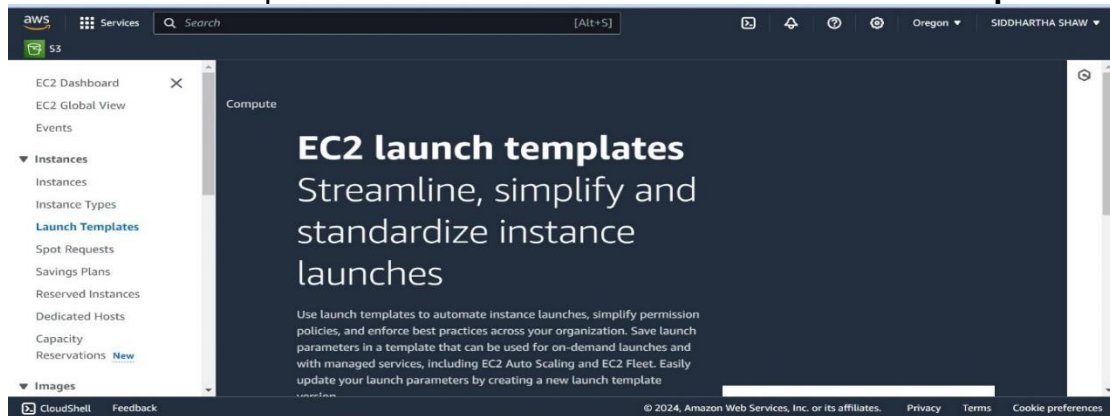


2. If you have an existing Security group then no need to create it if not then create it, using following steps- that are given below=>

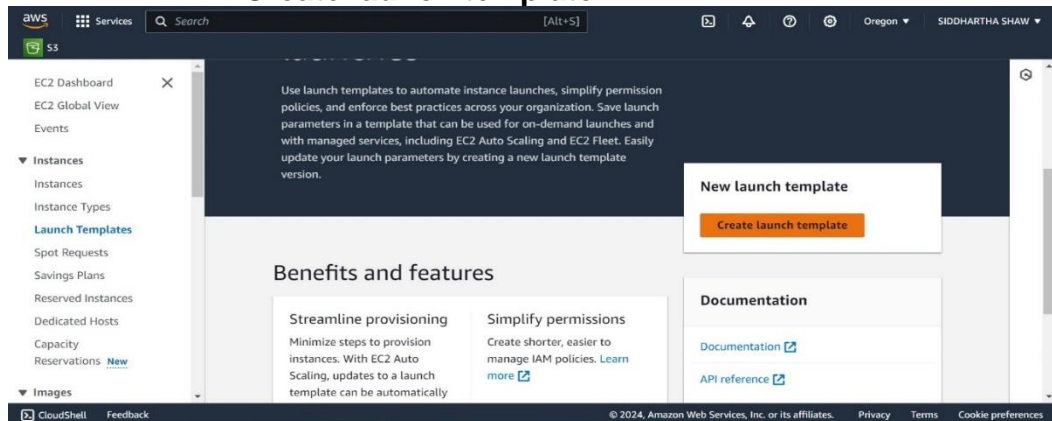




3. Create one template. Click Instances & there click on “Launch template”.

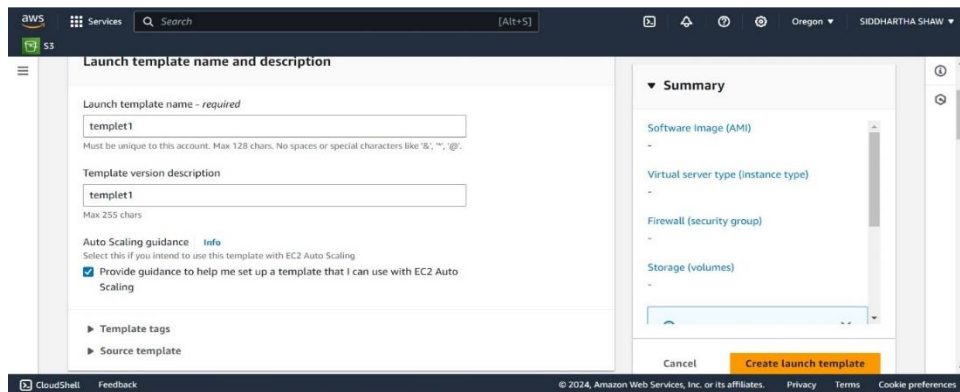


4. Now click on “Create launch template”.

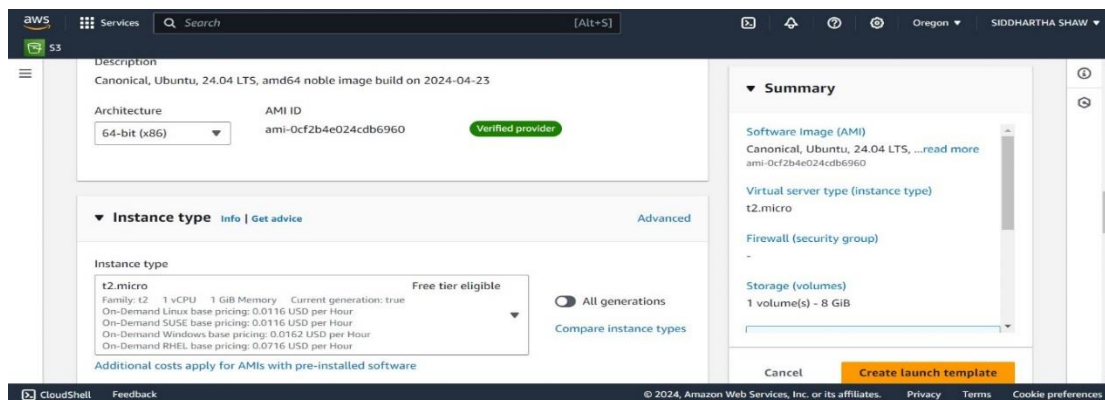
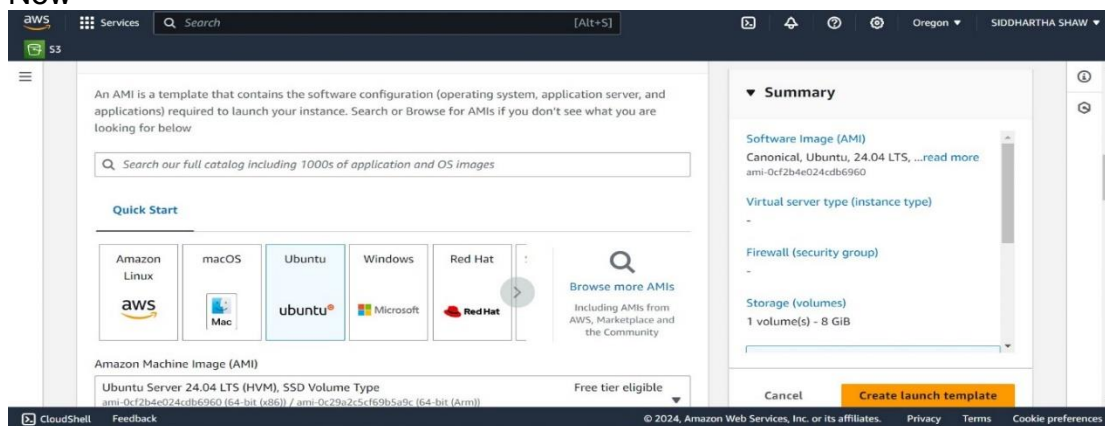


5. Under “Create launch template”, give the following details.

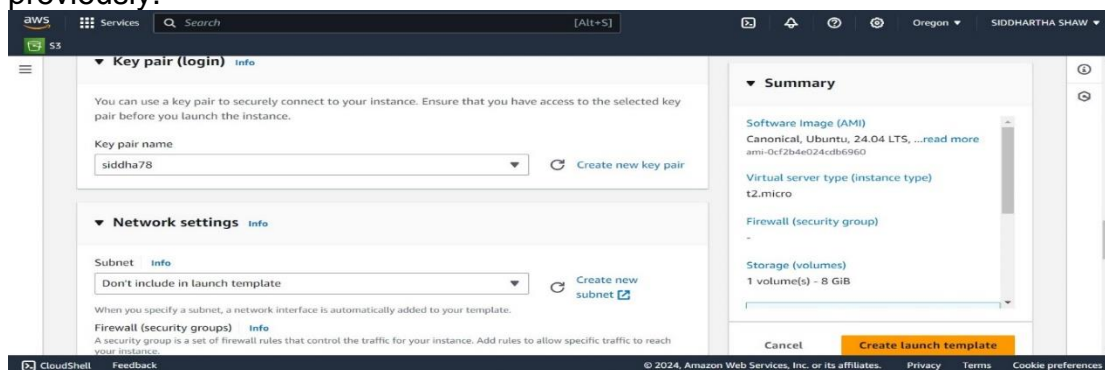
a. Firstly, we created “templet1” -that defines the launch templates name.

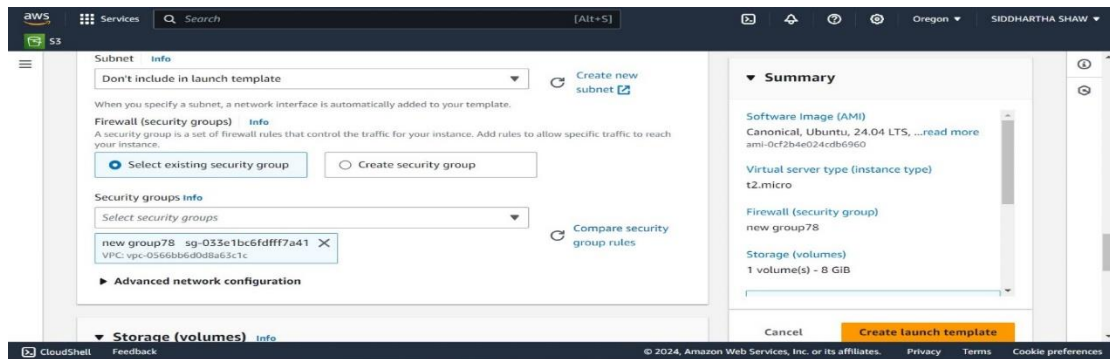


b. Click on “Quick start”->ubuntu & “Instance type”->**t2.micro** as it is free.
Now

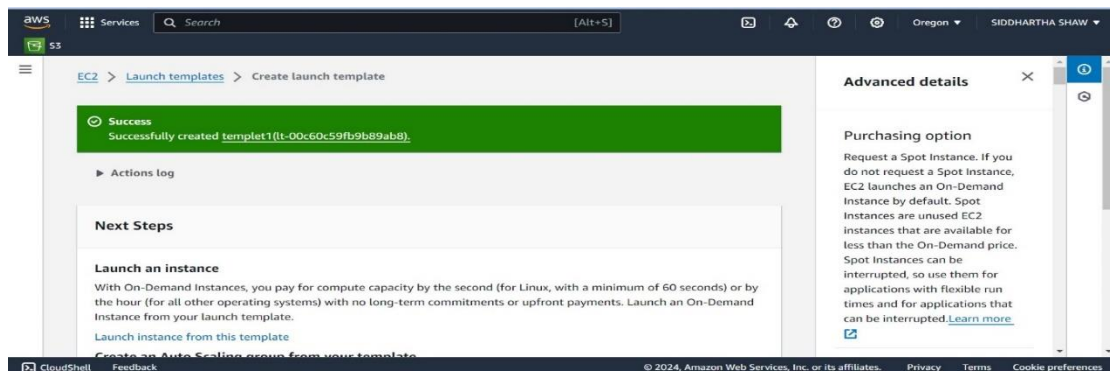


c. selects the **key pair** & the **existing** security group which you created previously.

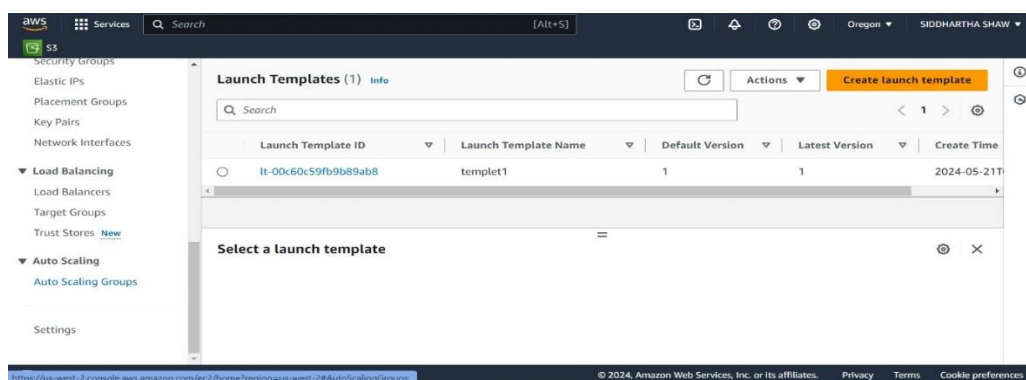




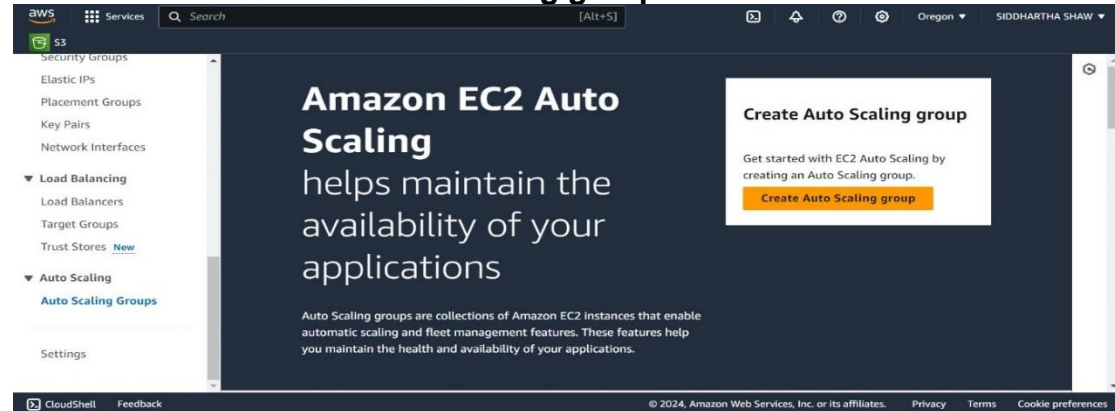
d. Expand the “Advanced Details” & scroll down to the bottom, in the bash console type the following commands, give the address & repository name from **GitHub**. Then click on “**Create Launch template**”



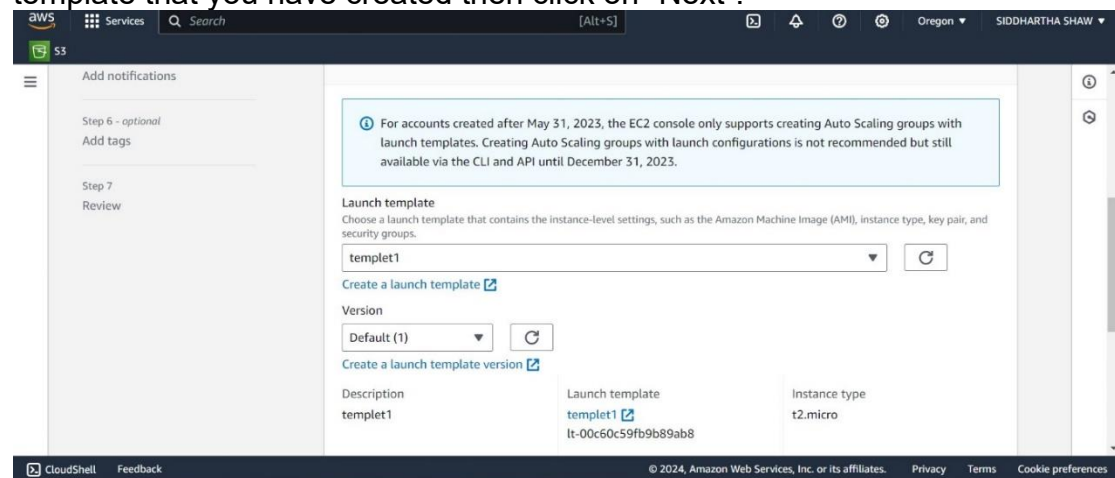
6. Template has successfully created & now go the next step.



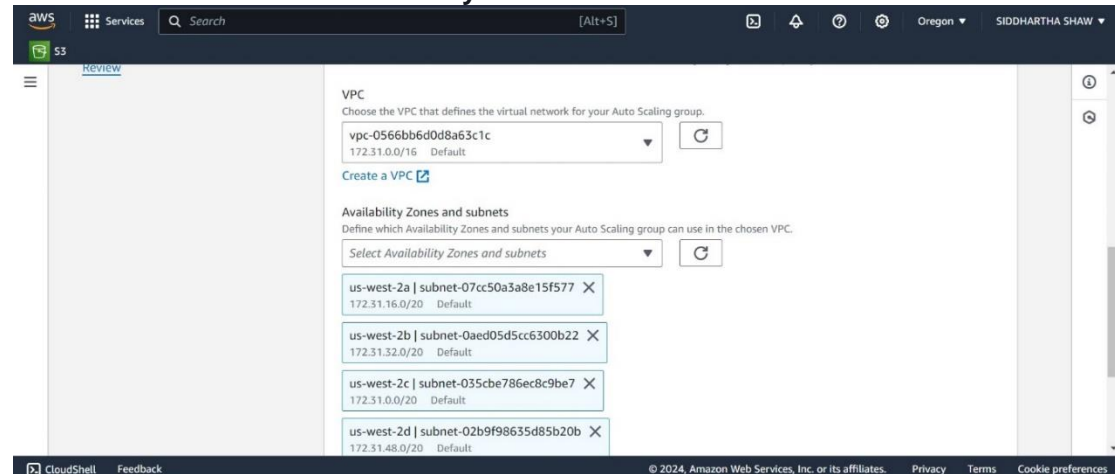
7. Now Click on “Create Auto Scaling group”.



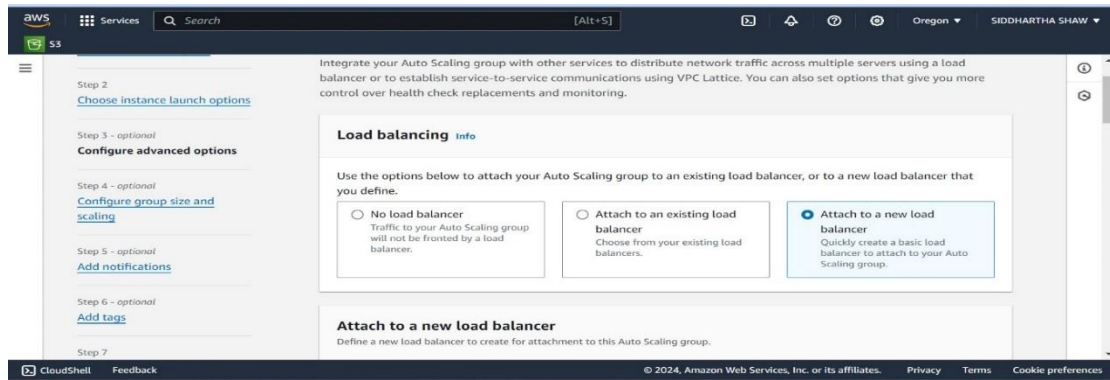
8. a. Under “Create Auto Scaling group”, give the name & choose the template that you have created then click on “Next”.



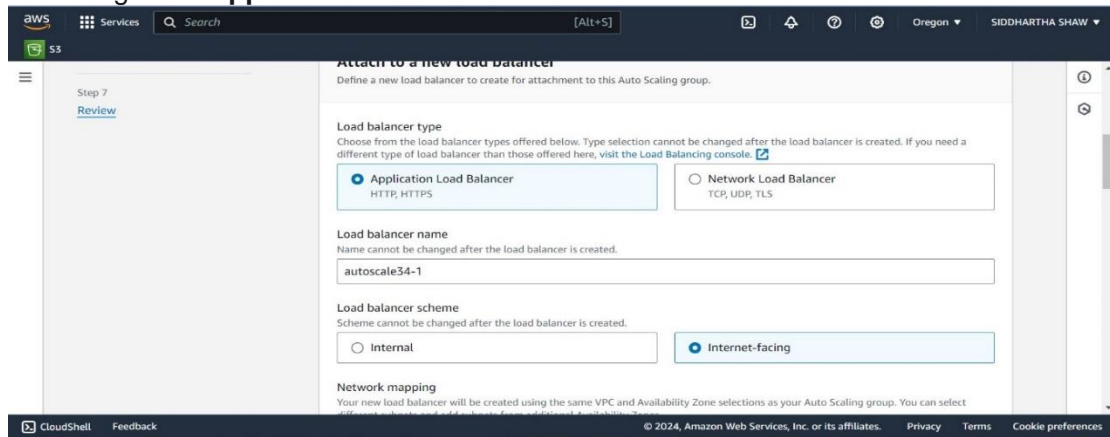
b. Then select all the “Availability Zones and subnets” & click on “Next”.



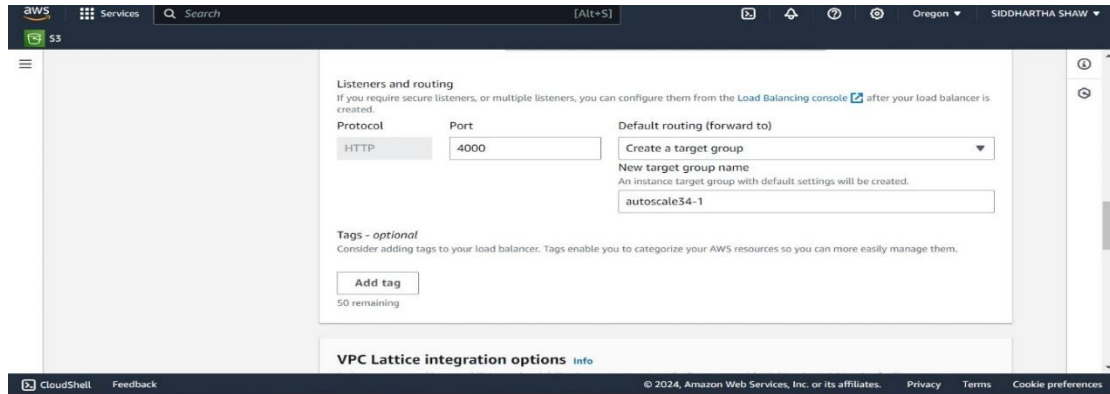
c. now go the “load balancing” part and selected “Attached to a new load balancer”.



d. then go the “Application load balancer” and create a “load balancer”.



e. now we do this.



f. Click on “No VPC Lattice service”. In “Health checks”, click on the checkbox

VPC Lattice integration options Info

To improve networking capabilities and scalability, integrate your Auto Scaling group with VPC Lattice. VPC Lattice facilitates communications between AWS services and helps you connect and manage your applications across compute services in AWS.

Select VPC Lattice service to attach

☒ **No VPC Lattice service**
VPC Lattice will not manage your Auto Scaling group's network access and connectivity with other services.

☐ **Attach to VPC Lattice service**
Incoming requests associated with specified VPC Lattice target groups will be routed to your Auto Scaling group.

[Create new VPC Lattice service](#)

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**

g. “Turn on **Elastic load ...**” Then click on “Next”.

Configure group size and scaling

Set the initial size of the Auto Scaling group. After creating the group, you can change its size to meet demand, either manually or by using automatic scaling.

Desired capacity type

Choose the unit of measurement for the desired capacity value. vCPUs and Memory(GiB) are only supported for mixed instances groups configured with a set of instance attributes.

Units (number of instances)

Desired capacity

Specify your group size.

2

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.

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

2

Equal or less than desired capacity

Max desired capacity

3

Equal or greater than desired capacity

Automatic scaling - optional

Choose whether to use a target tracking policy Info

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.

h. now click on “ **target tracking scaling policy**” and go to the further steps.

Choose whether to use a target tracking policy Info

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.

Scaling policy name

Target Tracking Policy

Metric type Info

Monitored metric that determines if resource utilization is too low or high. If using EC2 metrics, consider enabling detailed monitoring for better scaling performance.

Average CPU utilization

Target value

50

Instance warmup Info

This screenshot shows the 'Target Tracking Policy' configuration in the AWS console. The 'Metric type' is set to 'Average CPU utilization'. The 'Target value' is set to 50. The 'Instance warmup' is set to 240 seconds. There is an unchecked checkbox for 'Disable scale in to create only a scale-out policy'. Below this is the 'Instance maintenance policy' section, which includes information about controlling availability during replacement events.

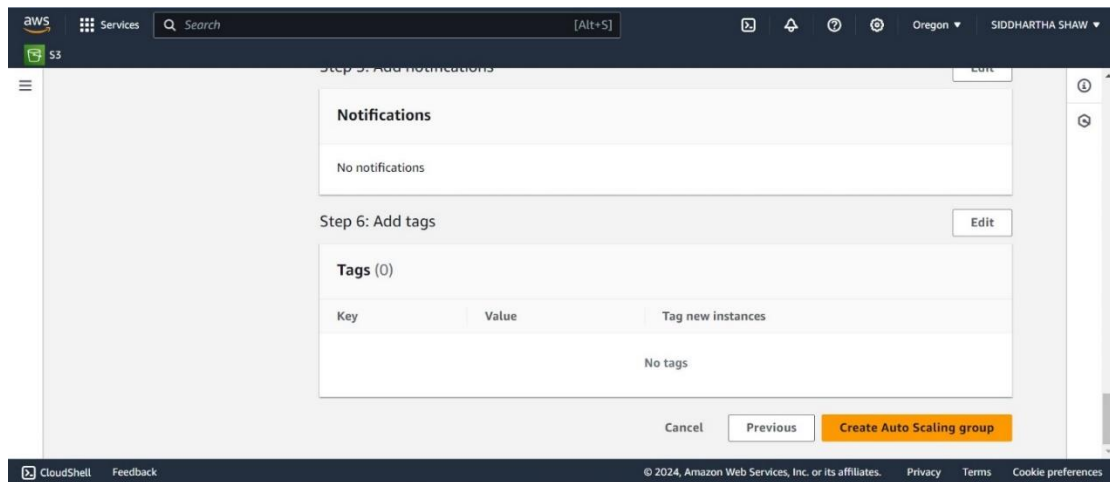
Click on “No policy” then on “NeXT”. Click on “Next”.

This screenshot shows the 'Add notifications - optional' step in the 'Create Auto Scaling group' wizard. The left sidebar lists steps from 'Choose launch template' to 'Review'. The main area explains that notifications can be sent to SNS topics when instances are launched or terminated. There is an 'Add notification' button and a 'Next' button at the bottom right.

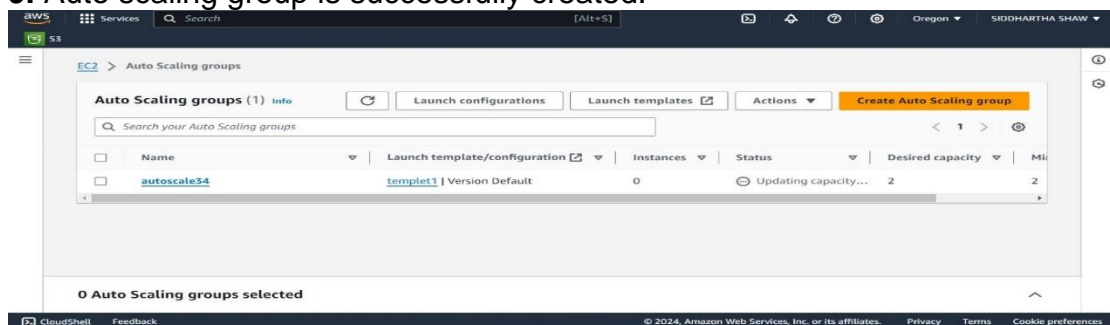
Click on “Next”.

This screenshot shows the 'Add tags - optional' step in the 'Create Auto Scaling group' wizard. A tooltip explains that tags can be added to instances and their EBS volumes, but duplicate keys will be overridden. Below the tooltip is a 'Tags (0)' section with an 'Add tag' button and a 'Next' button at the bottom right.

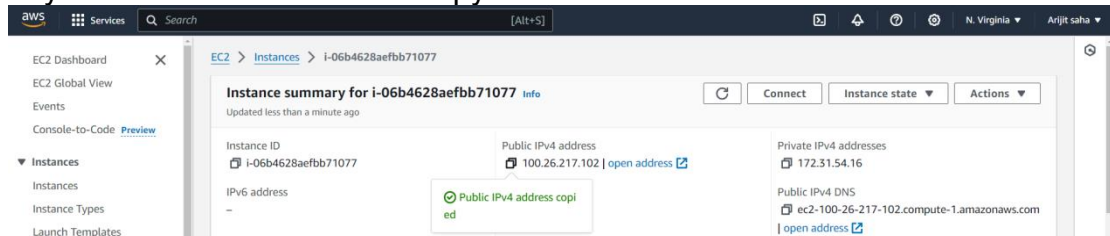
Click on “Create Auto Scaling group”.



9. Auto scaling group is successfully created.



10. Now go to "Instance" and check for running instances with no name but then click on any one of the instance ID & copy the "Public IPv4 address".



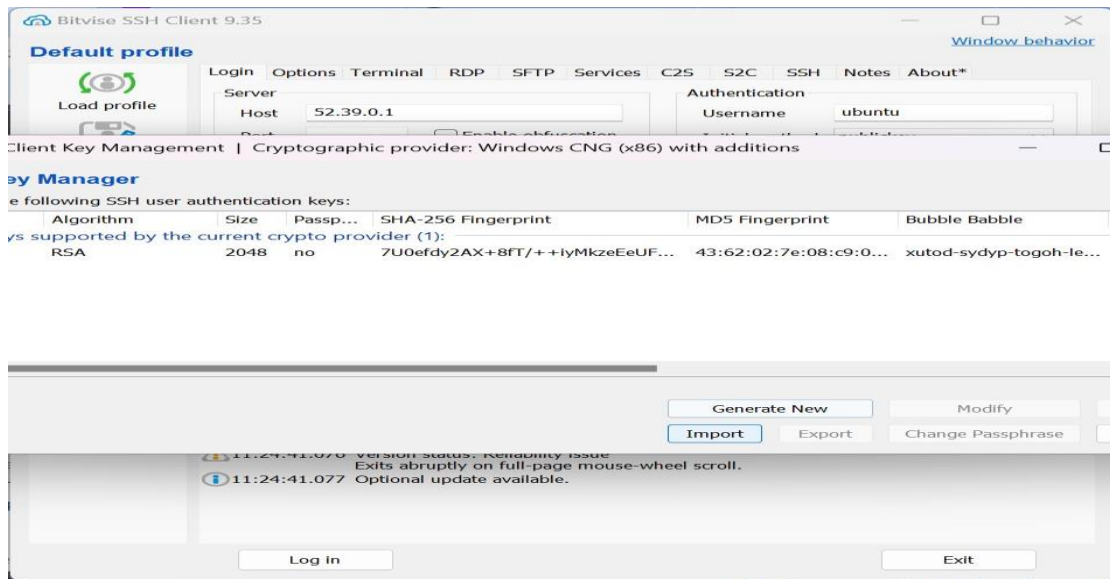
11. Paste the address in a new Window.



12. Now add ":4000" at the end of the IPv4 address and press enter.



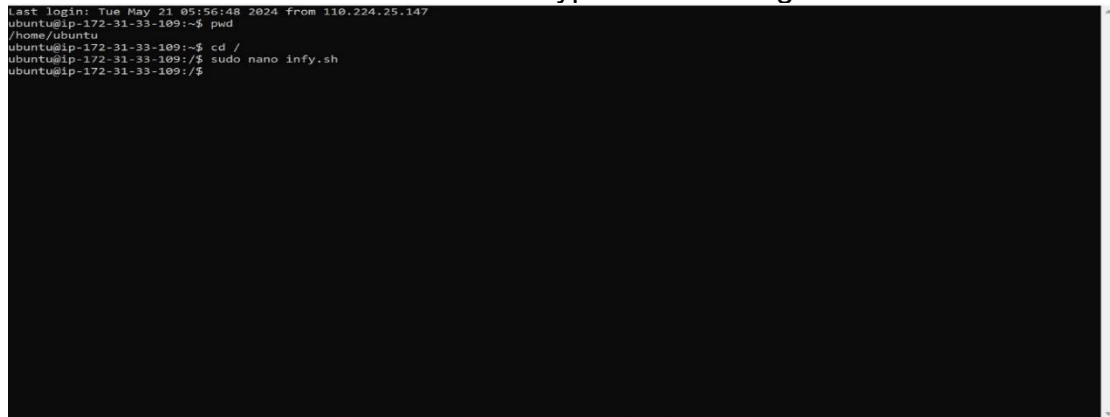
13. Then copy the address of any one of the instances & paste in the host of the "Bitvise SSH Client" then click on "Client key Manager".



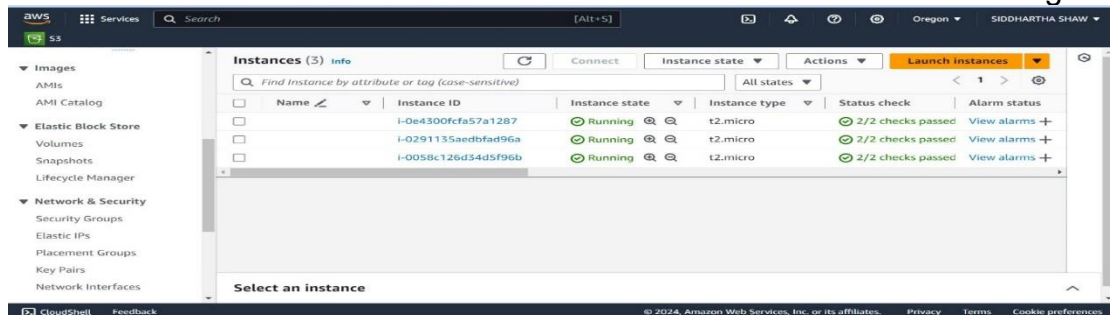
14. Then import the key & click on “Log in”.

15. The “Log out” came means that is is successfully logged in.

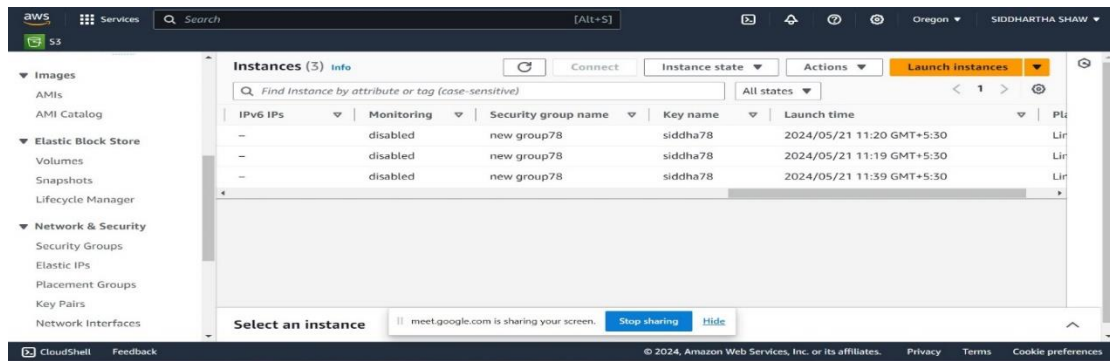
16. Now in “New terminal console and type the following commands.



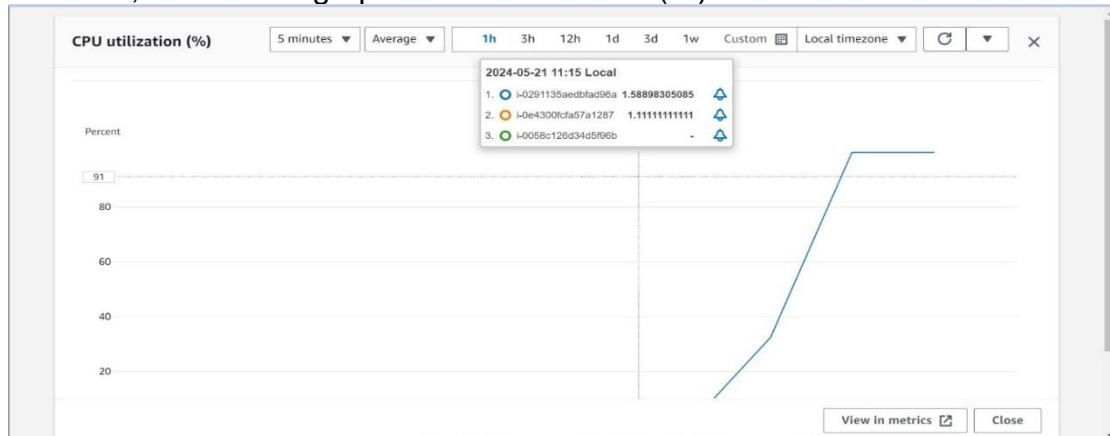
17. Now in Instances select both the unnamed instances & click on ‘Enlarge’.



See the “time zone”



18. Here, we see the graph of CPU utilization (%) & select “Local time zone”.



19. Here, we see that two instances are running along with the initialization of the third instance.

