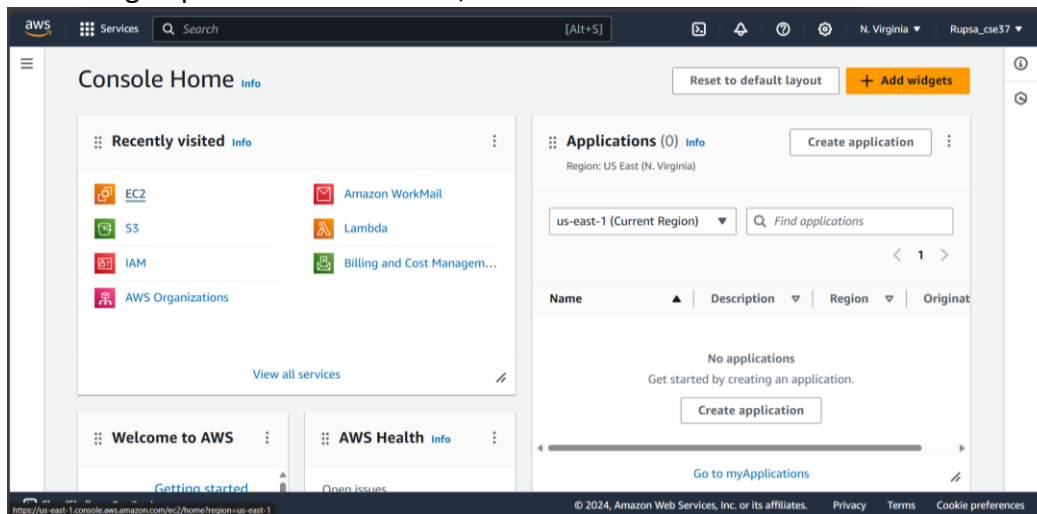


PROBLEM STATEMENT :

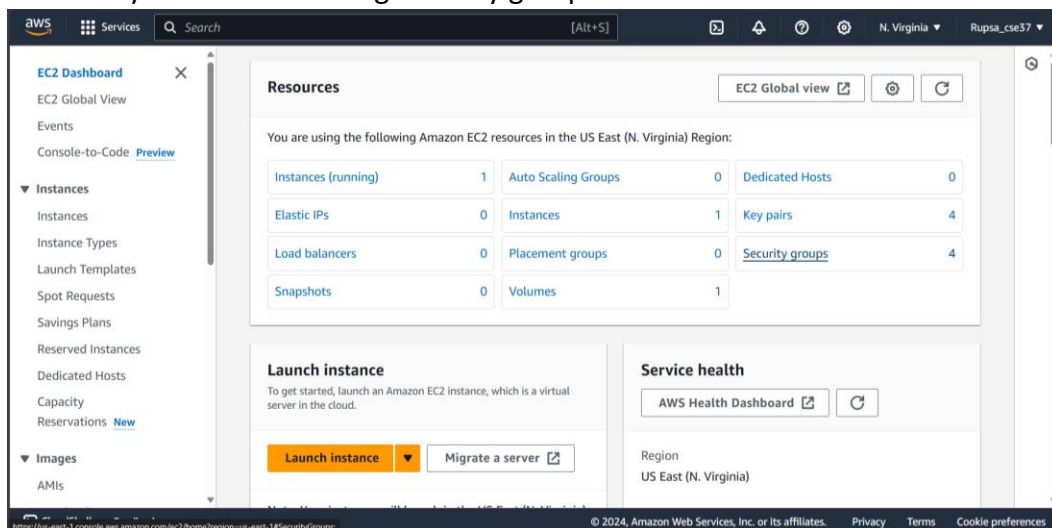
11) Build scaling plans in AWS that balance the load on different EC2 instances.

Steps to build scaling plans :

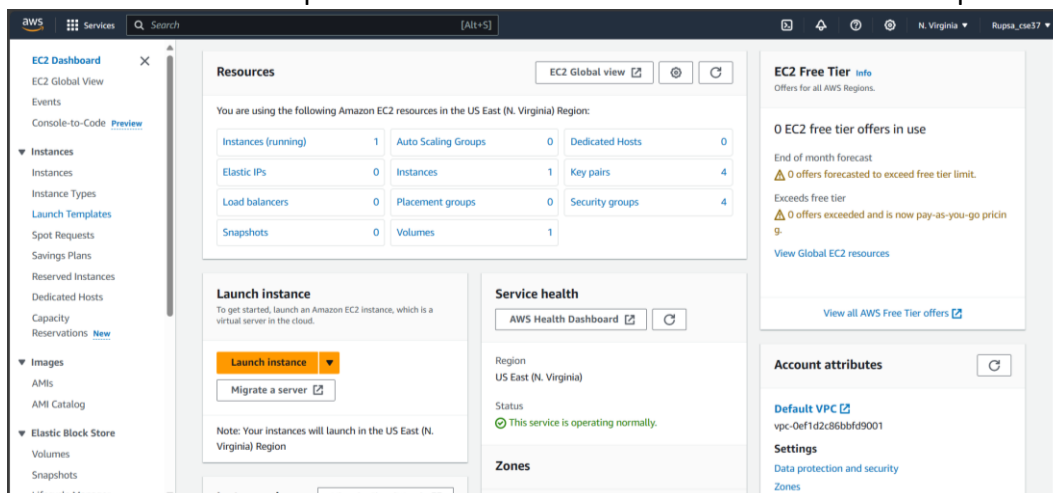
1. Sign up for an AWS account, search for 'EC2' then click on it.



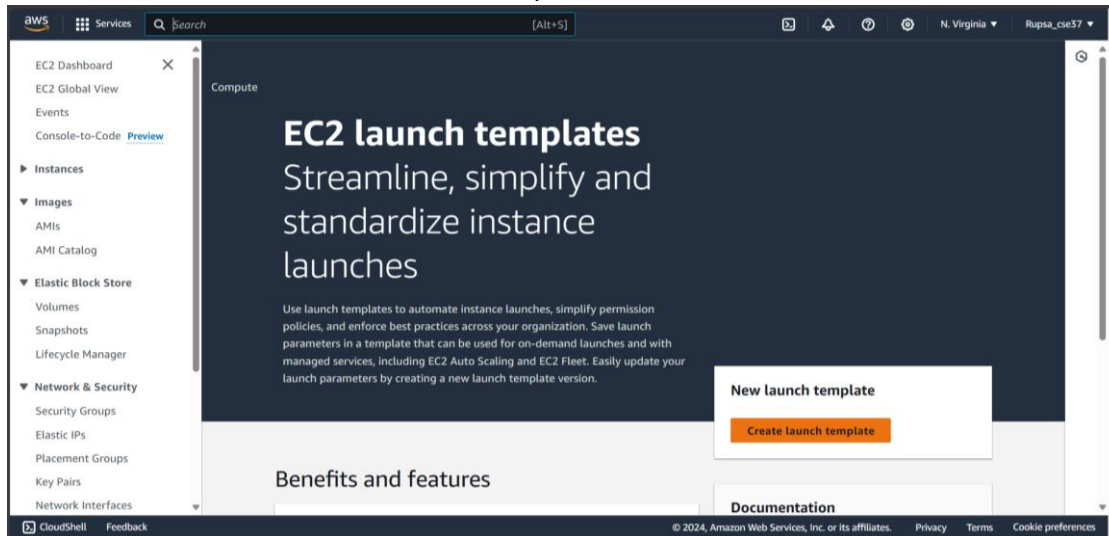
2. If you have an existing Security group then no need to create it if not then create it.



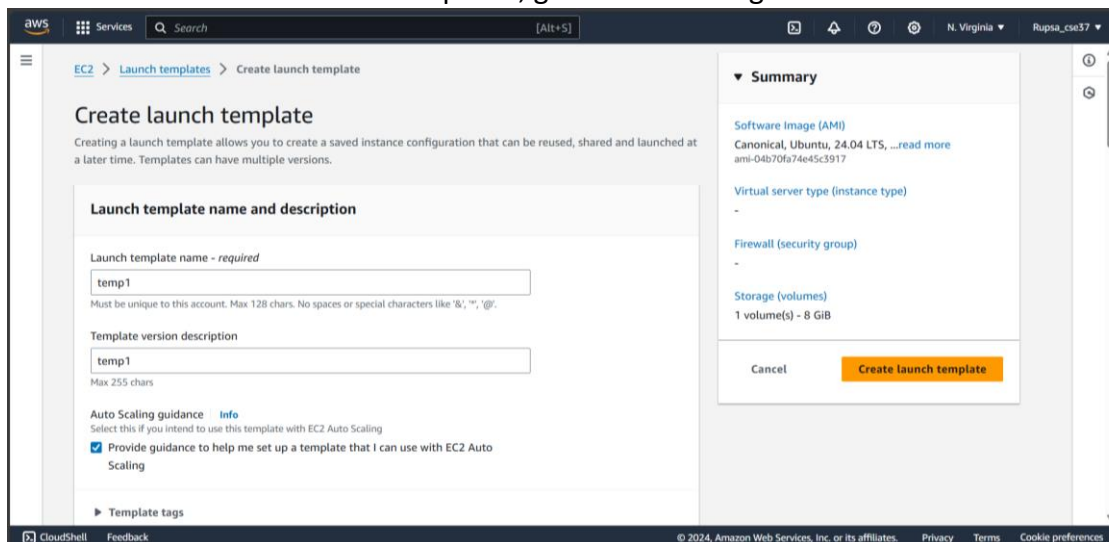
3. Create one template. Click Instances & there click on "Launch template".



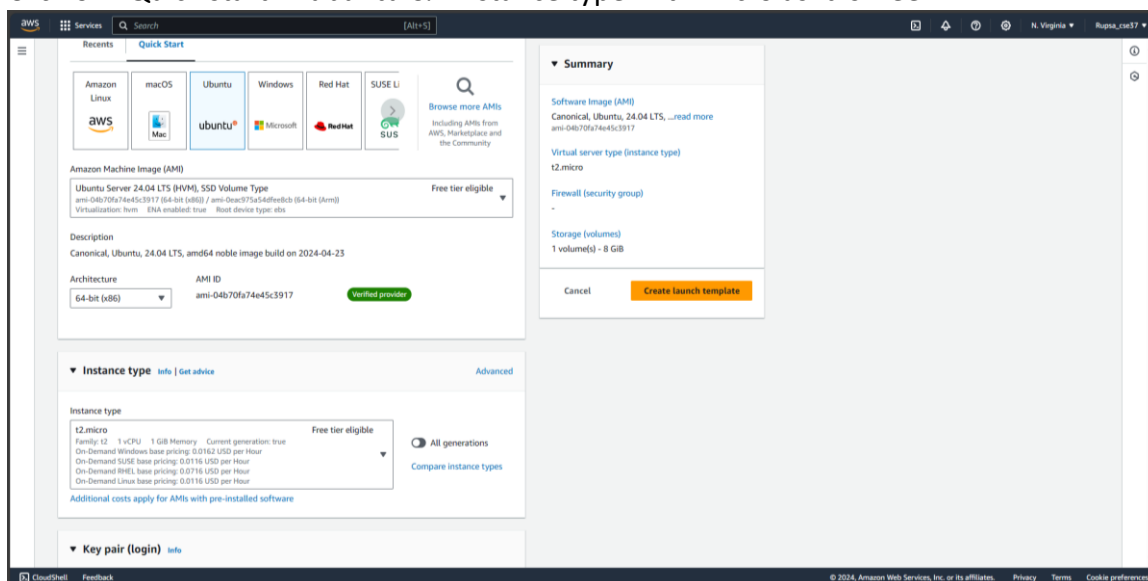
4. Now click on “Create launch template”.



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



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



Now select the key pair & the existing security group.

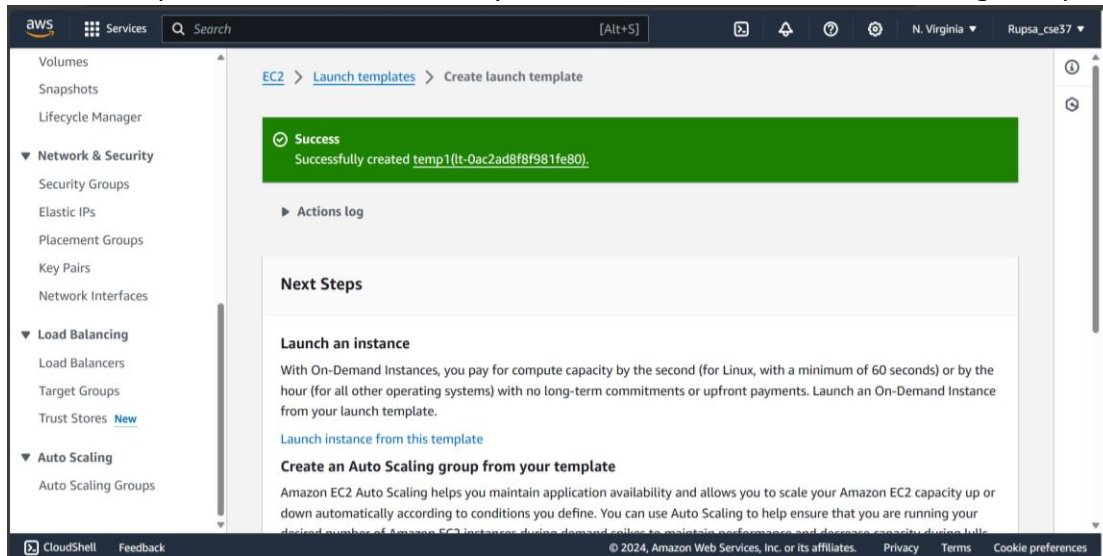
The screenshot shows the 'Create Launch Template' page in the AWS Management Console. The 'Key pair (login)' tab is active, showing a dropdown for 'Key pair name' with 'new_key' selected and a 'Create new key pair' button. Below it, the 'Network settings' tab is active, showing a dropdown for 'Subnet' with 'Don't include in launch template' selected and a 'Create new subnet' button. The 'Firewall (security groups)' section has 'Select existing security group' selected, and a dropdown for 'Security groups' with 'rup_security' selected. A 'Compare security group rules' button is also visible. The 'Summary' panel on the right shows the configuration: Software Image (AMI) Canonical, Ubuntu, 24.04 LTS, Virtual server type (instance type) t2.micro, Firewall (security group) rup_security, and Storage (volumes) 1 volume(s) - 8 GiB. At the bottom, there are 'Cancel' and 'Create launch template' buttons.

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”.

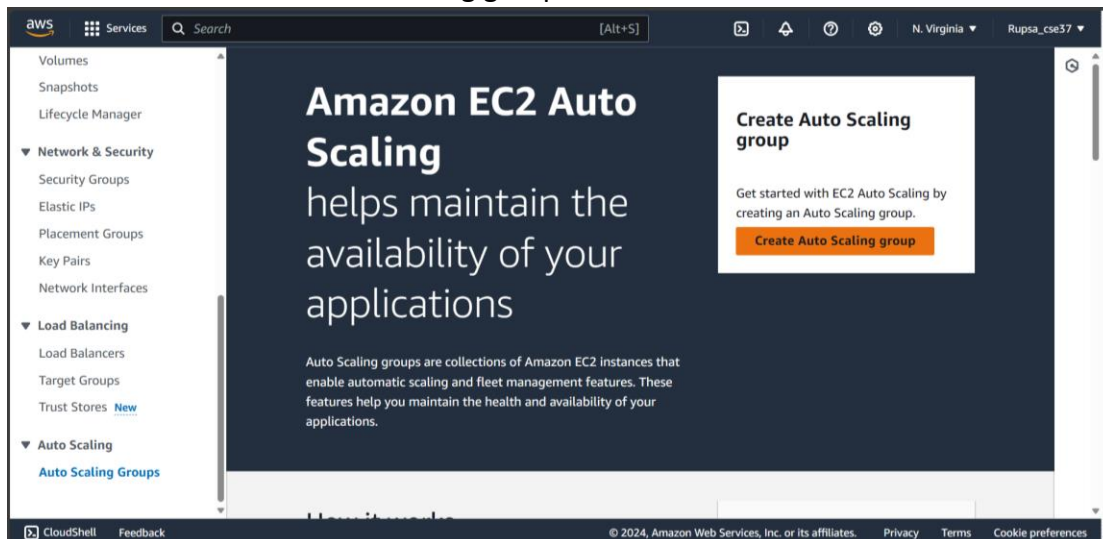
The screenshot shows the 'Create Launch Template' page in the AWS Management Console, with the 'Advanced details' tab active. The 'Metadata response hop limit' is set to 2. The 'Allow tags in metadata' dropdown is set to 'Don't include in launch template'. The 'User data - optional' section has a 'Choose file' button and a text area containing a bash script. The script includes commands for updating the system, installing nginx and git, cloning a repository, and installing npm. A checkbox for 'User data has already been base64 encoded' is present. The 'Summary' panel on the right shows the same configuration as the previous screenshot. At the bottom, there are 'Cancel' and 'Create launch template' buttons.

```
#!/bin/bash
apt-get update
apt-get install -y nginx
systemctl start nginx
systemctl enable nginx
apt-get install -y git
curl -SL https://deb.nodesource.com/setup_16.x | sudo -E bash -
apt-get install -y nodejs
git clone https://github.com/Rupsa1037/Rup_AWS1.git
cd Rup_AWS1
npm install
node index.js
```

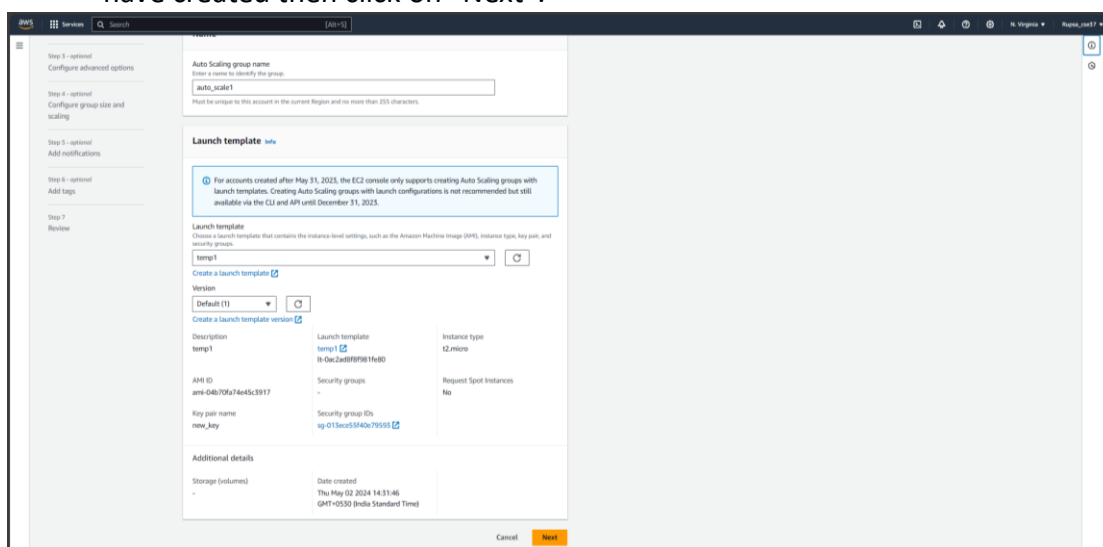
6. Template has been successfully created & now click on “Auto Scaling Groups”.



7. Click on “Create Auto Scaling group”.



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



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

The screenshot shows the 'Network' step of the AWS console wizard. On the left, a sidebar lists steps: 'Add notifications', 'Step 6 - optional: Add tags', and 'Step 7: Review'. The main content area is titled 'Network info' and includes a 'VPC' section with a dropdown menu showing 'vpc-0ef1d2c86bbf99001' and a 'Create a VPC' link. Below this is the 'Availability Zones and subnets' section, which has a 'Select Availability Zones and subnets' dropdown and a list of seven subnets, each with a selection box and an 'X' icon to remove it. The subnets are: 'us-east-1a | subnet-0dea201b355d8cd43', 'us-east-1b | subnet-01bf8c0100b665c89', 'us-east-1c | subnet-0935c7b3db1ef6a39', 'us-east-1d | subnet-02344c75c1b40c7e', 'us-east-1e | subnet-054e4bd8f93f6e4f', 'us-east-1f | subnet-08451054b210e3432', and 'us-east-1g | subnet-0a451054b210e3432'. At the bottom, there are buttons for 'Cancel', 'Skip to review', 'Previous', and 'Next'.

The screenshot shows the 'Load balancing' step of the AWS console wizard. The sidebar on the left shows 'Step 3 - optional: Configure advanced options' as the current step, with other steps like 'Step 4 - optional: Configure group size and scaling' and 'Step 5 - optional: Add notifications' listed below. The main content area is titled 'Load balancing info' and offers three options: 'No load balancer', 'Attach to an existing load balancer', and 'Attach to a new load balancer'. The 'Attach to a new load balancer' option is selected. Below this, the 'Attach to a new load balancer' section allows choosing a 'Load balancer type' (Application Load Balancer or Network Load Balancer), entering a 'Load balancer name' (auto_scale1-1), and selecting a 'Load balancer scheme' (Internal or Internet-facing). A 'Network mapping' section at the bottom explains that the new load balancer will use the same VPC and Availability Zone selections as the Auto Scaling group.

Click on “No VPC Lattice service”.

The screenshot shows the 'Listeners and routing' step of the AWS console wizard. The sidebar on the left shows 'Step 6 - optional: Add tags' as the current step, with other steps like 'Step 5 - optional: Add notifications' and 'Step 7: Review' listed below. The main content area is titled 'Listeners and routing' and includes a 'Protocol' dropdown set to 'HTTP' and a 'Port' input field set to '4000'. There is a 'Default routing (forward to)' dropdown set to 'Create a target group' and a 'New target group name' input field set to 'auto_scale1-1'. Below this is a 'Tags - optional' section with an 'Add tag' button and a note '50 remaining'. The 'VPC Lattice integration options' section at the bottom offers two choices: 'No VPC Lattice service' (selected) and 'Attach to VPC Lattice service'. A 'Create new VPC Lattice service' link is also present.

In “Health checks”, click on the checkbox “Turn on Elastic load ...” Then click on “Next”.

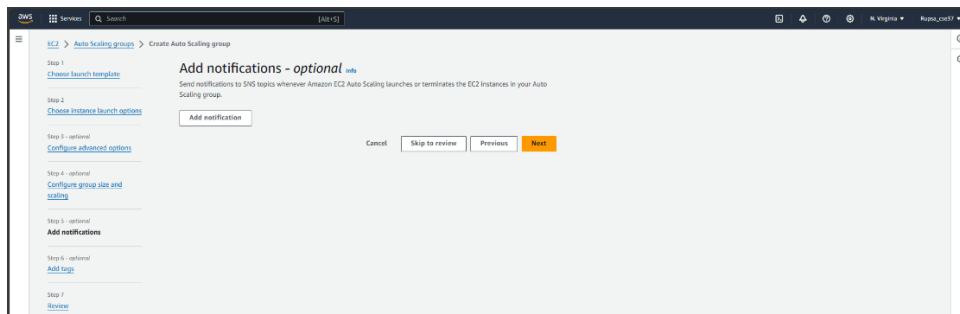
This screenshot shows the 'Health checks' configuration page in the AWS Management Console. The page is titled 'Health checks' and includes a description: 'Health checks increase availability by replacing unhealthy instances. When you use multiple health checks, all are evaluated, and if at least one fails, instance replacement occurs.' The 'EC2 health checks' section shows 'Always enabled'. Under 'Additional health check types - optional', the checkbox 'Turn on Elastic Load Balancing health checks' is checked. A warning box states: 'EC2 Auto Scaling will start to detect and act on health checks performed by Elastic Load Balancing. To avoid unexpected terminations, first verify the settings of these health checks in the Load Balancing console.' Below this, the 'Health check grace period' is set to 240 seconds. The 'Additional settings' section includes 'Monitoring' (unchecked), 'Default instance warmup' (unchecked), and 'Enable default instance warmup' (unchecked). At the bottom, there are buttons for 'Cancel', 'Skip to review', 'Previous', and 'Next'.

This screenshot shows the 'Group size' configuration page in the AWS Management Console. The page is titled 'Group size' and includes a description: '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.' The 'Desired capacity type' is set to 'Units (number of instances)'. The 'Desired capacity' is set to 2. The 'Scaling limits' section shows 'Min desired capacity' as 2 and 'Max desired capacity' as 3. Under 'Automatic scaling - optional', the 'Target tracking scaling policy' is selected. The 'Scaling policy name' is 'Target Tracking Policy'. The 'Metric type' is 'CPU usage'. At the bottom, there are buttons for 'Cancel', 'Skip to review', 'Previous', and 'Next'.

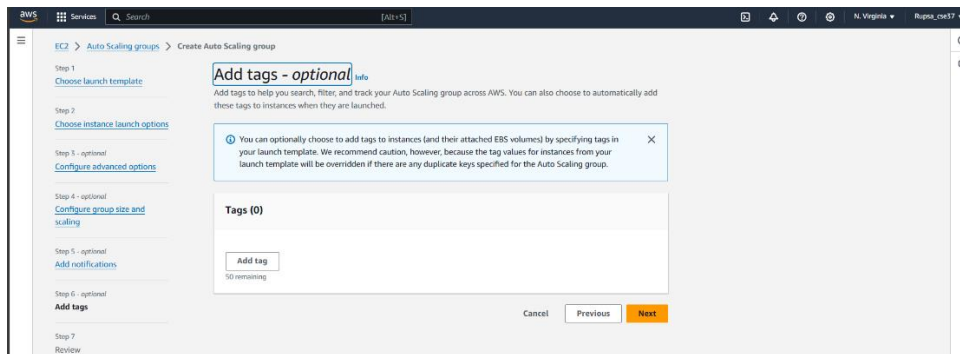
Click on “No policy” then on “Next”.

This screenshot shows the 'Target value' configuration page in the AWS Management Console. The page is titled 'Target value' and includes a description: '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.' The 'Target value' is set to 50. The 'Instance warmup' is set to 240 seconds. The 'Disable scale in to create only a scale-out policy' checkbox is unchecked. Under 'Choose a replacement behavior depending on your availability requirements', the 'No policy' option is selected. The 'Instance scale-in protection' section shows 'Enable instance scale-in protection' unchecked. At the bottom, there are buttons for 'Cancel', 'Skip to review', 'Previous', and 'Next'.

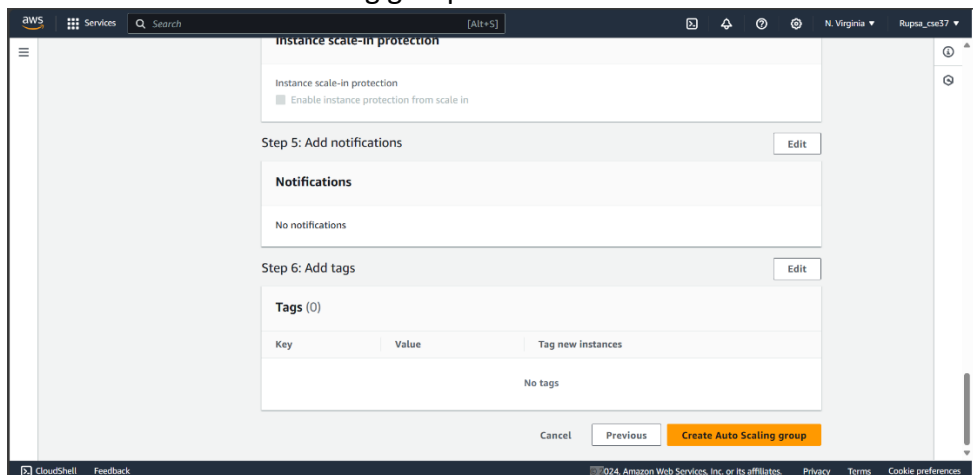
Click on “Next”.



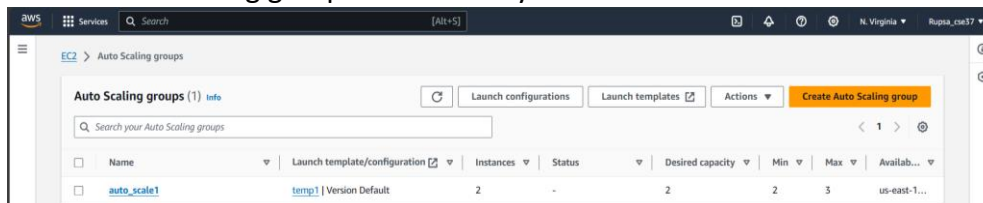
Click on “Next”.



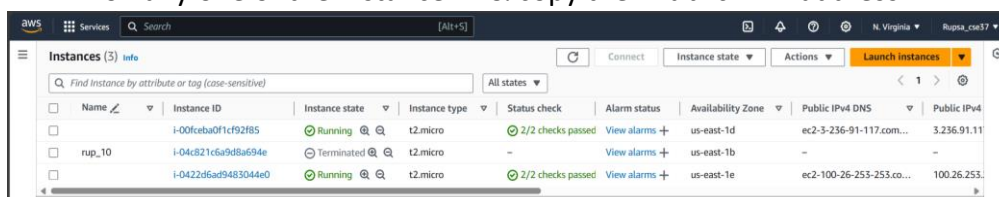
Click on “Create Auto Scaling group”.



9. Auto scaling group is successfully created.



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



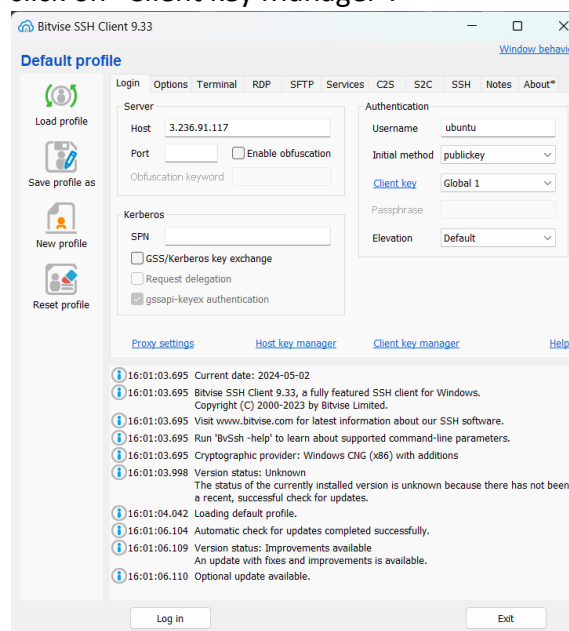
11. Paste the address in a new Window.



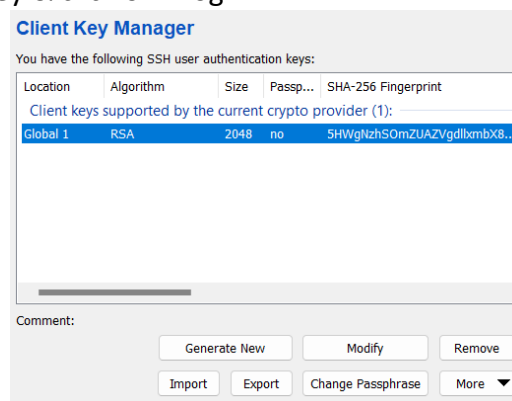
12. Now add “:4000” at the end of the IPv4 address and press enter.



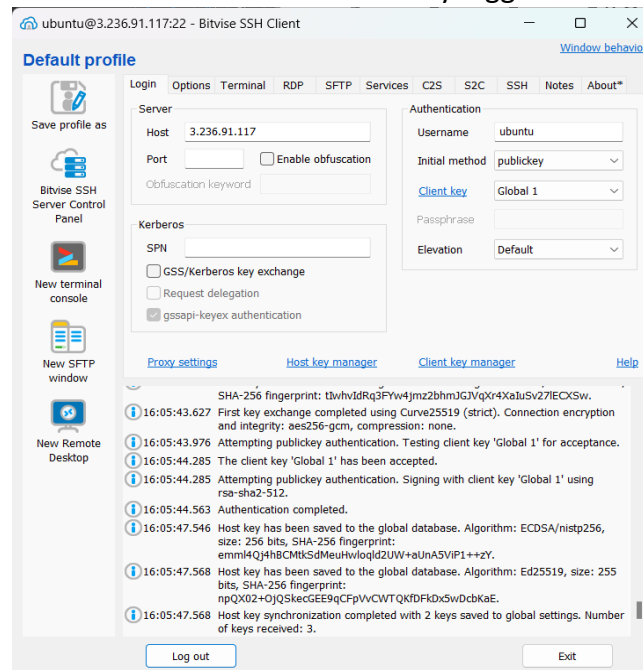
13. Then copy the address of any one of the instance & paste in the host of the “Bitvise SSH Client” then click on “Client key Manager”.



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.

```
ubuntu@ip-172-31-2-45:~$ pwd
/home/ubuntu
ubuntu@ip-172-31-2-45:~$ sudo nano infy.sh
```

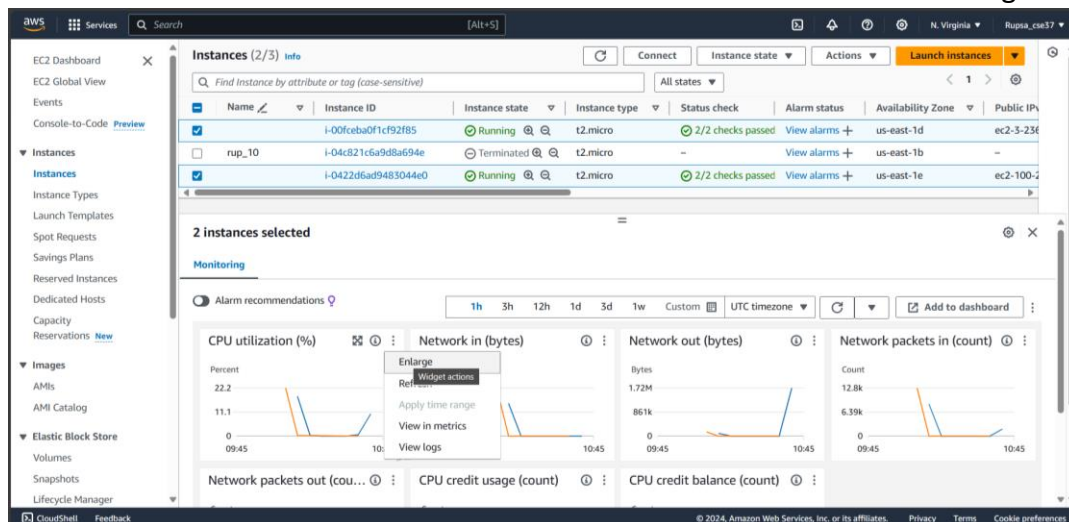
ubuntu@3.236.91.117:22 - Bi

```
GNU nano 7.2
#!/bin/bash
while(true)
do
    echo "Inside loop"
done
```

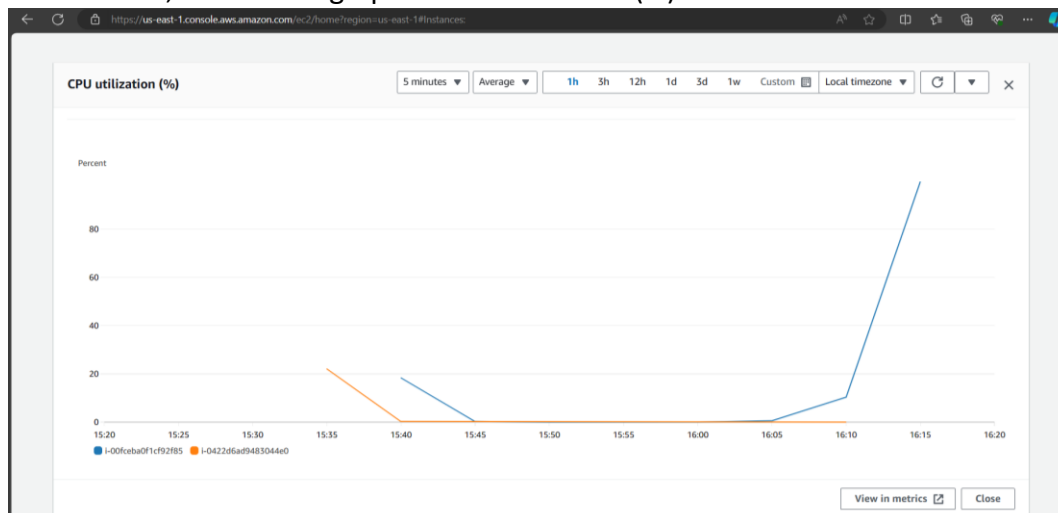
```
ubuntu@ip-172-31-2-45:~$ sudo chmod +x infy.sh
ubuntu@ip-172-31-2-45:~$ sh infy.sh
```

```
Inside loop
Inside loop
Inside loop
Inside loop
Inside loop
Inside loop
Inside loop
Inside loop
Inside loop
```

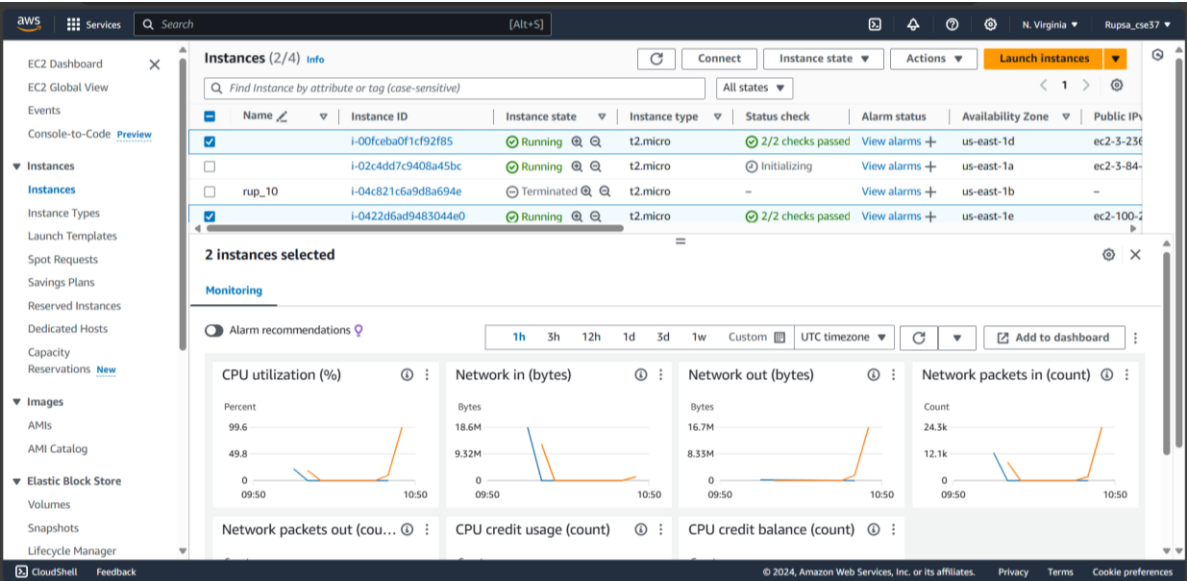
17. Now In Instances select both the unnamed instances & click on 'Enlarge'.



18. Here, we see the graph of CPU utilization(%) & select "Local timezone".



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



CloudShell

Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)