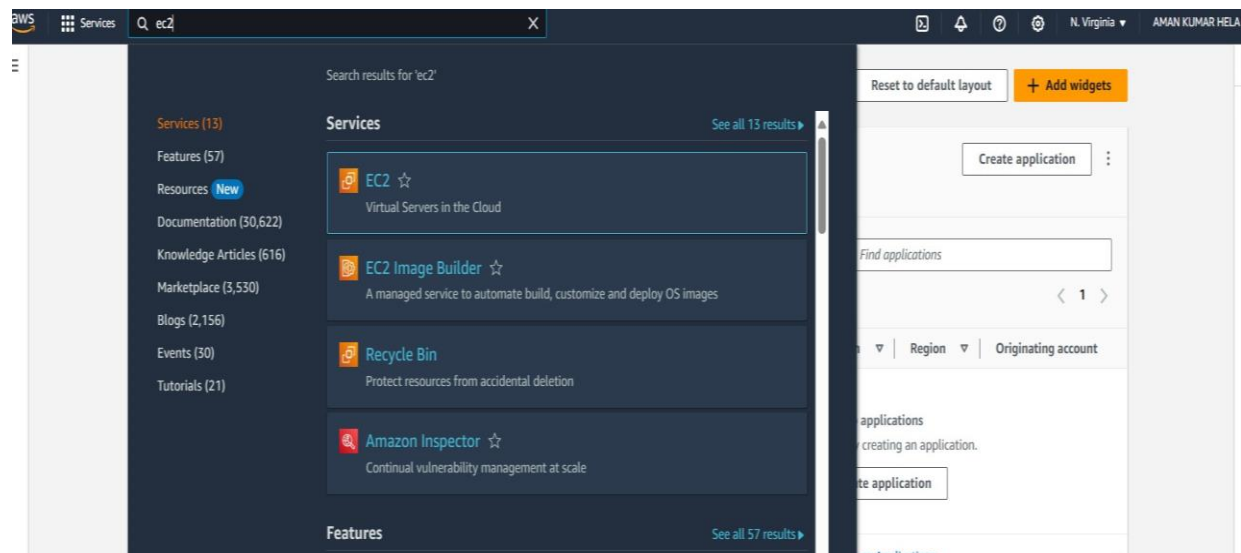


Assignment No: 11

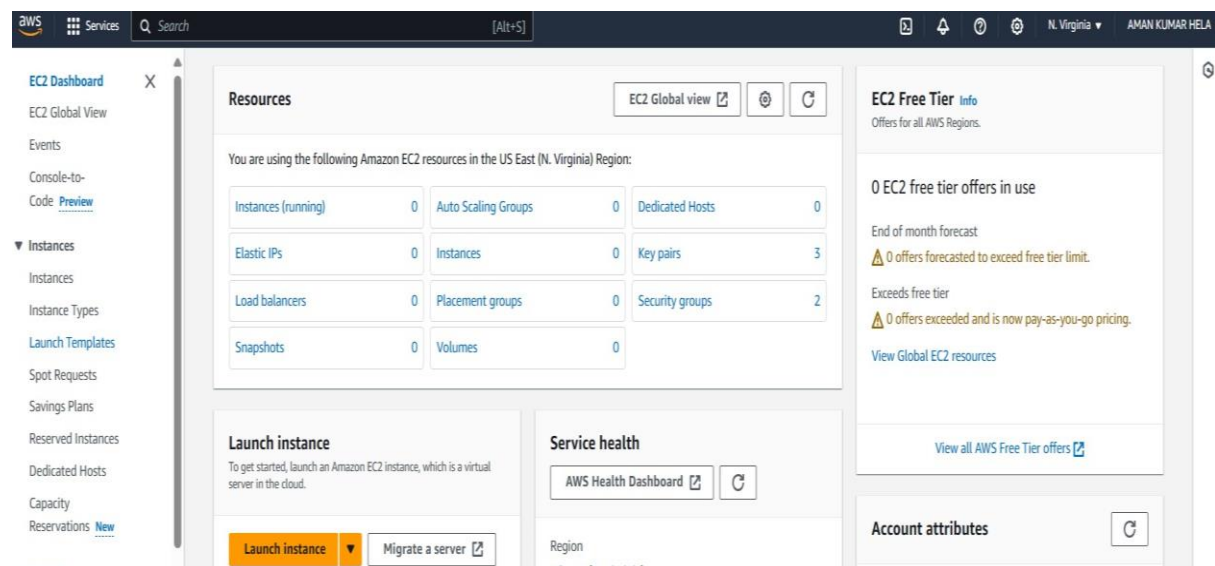
Problem Statement: Build scaling plans in AWS that balance load on different EC2 instances.

The steps are as follows: -

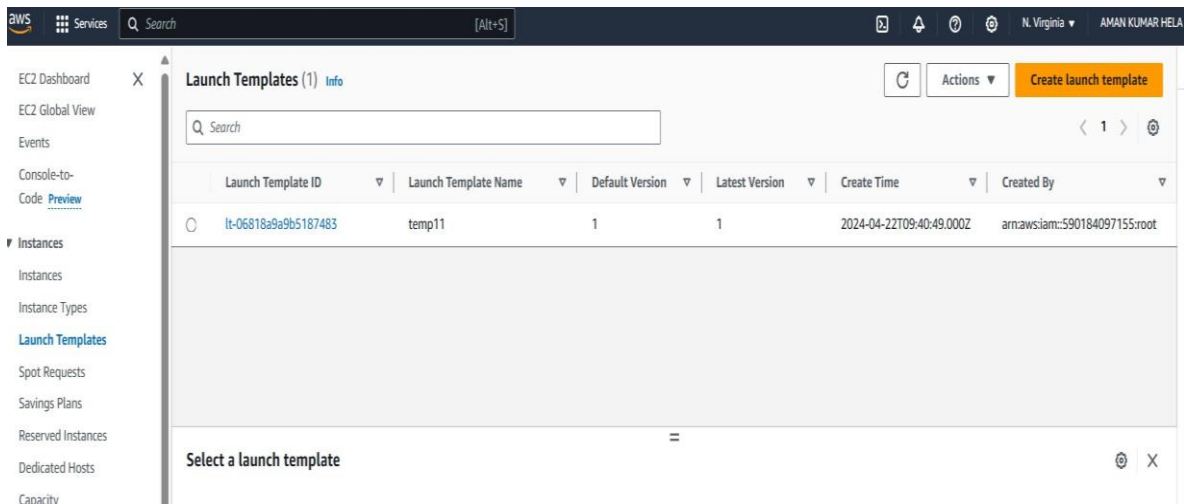
1. Please access AWS and look up EC2, then select the initial option displayed.



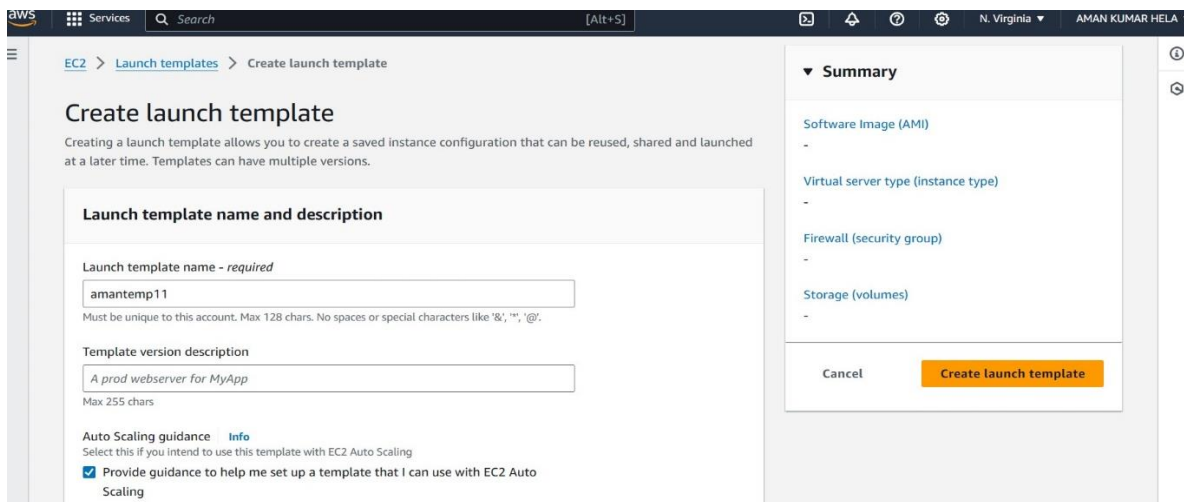
2. Next, select "Launch Template" from the menu on the left side.



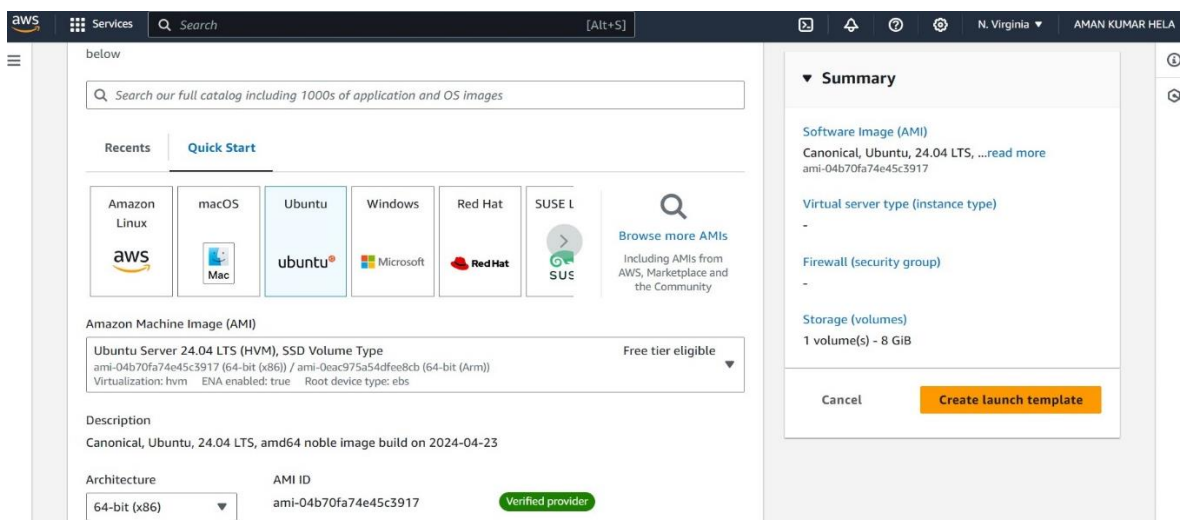
3. Now click on the "Create Template" option.



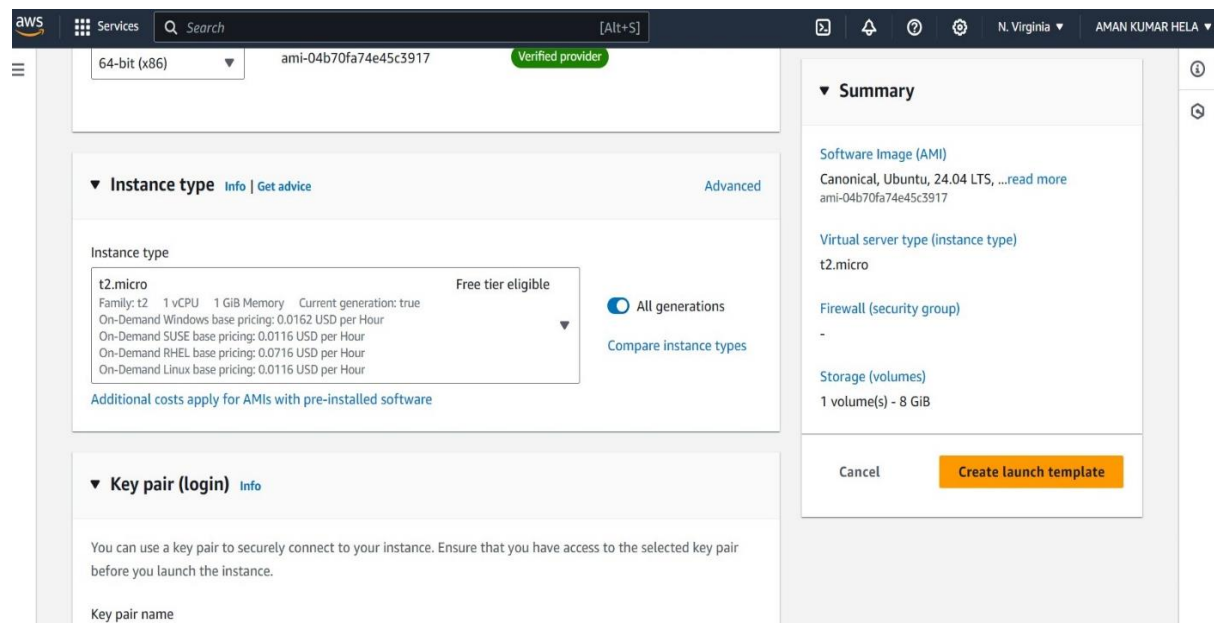
4. Enter a template name, such as "amantemp11," and check the box for autoscaling options.



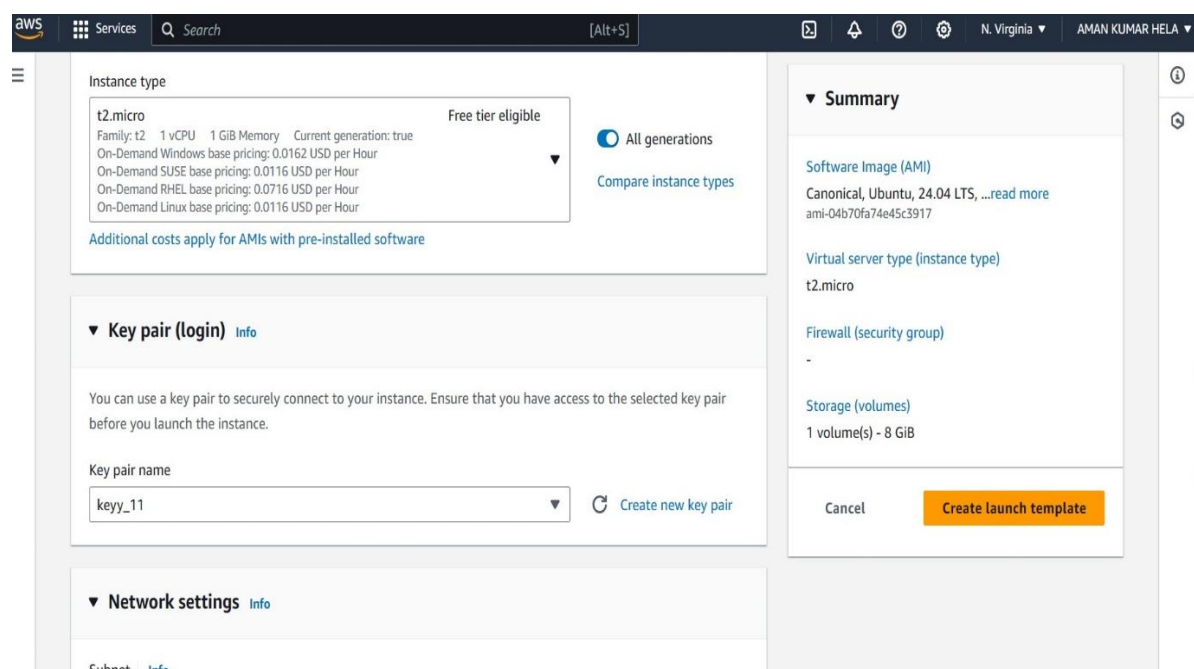
5. Navigate to "Quick Start" and choose "Ubuntu" from the list of available AMIs.



6. Next, select the instance type—either t2.micro or t3.micro, both of which are free tier eligible.



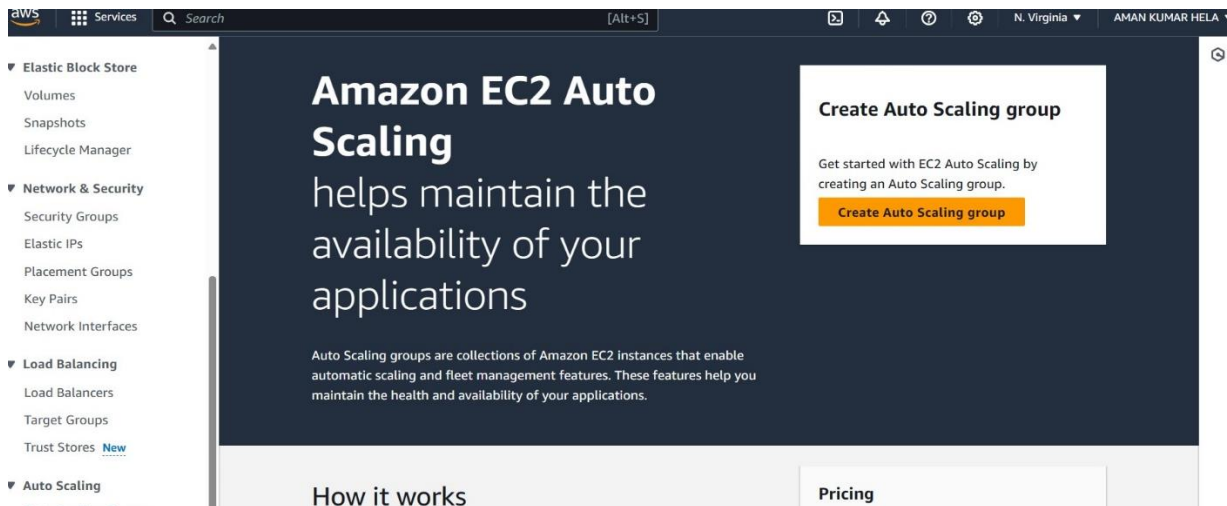
7. Choose either an existing key pair or create a new one if it doesn't exist. For example, use "key_11" as the key pair name.



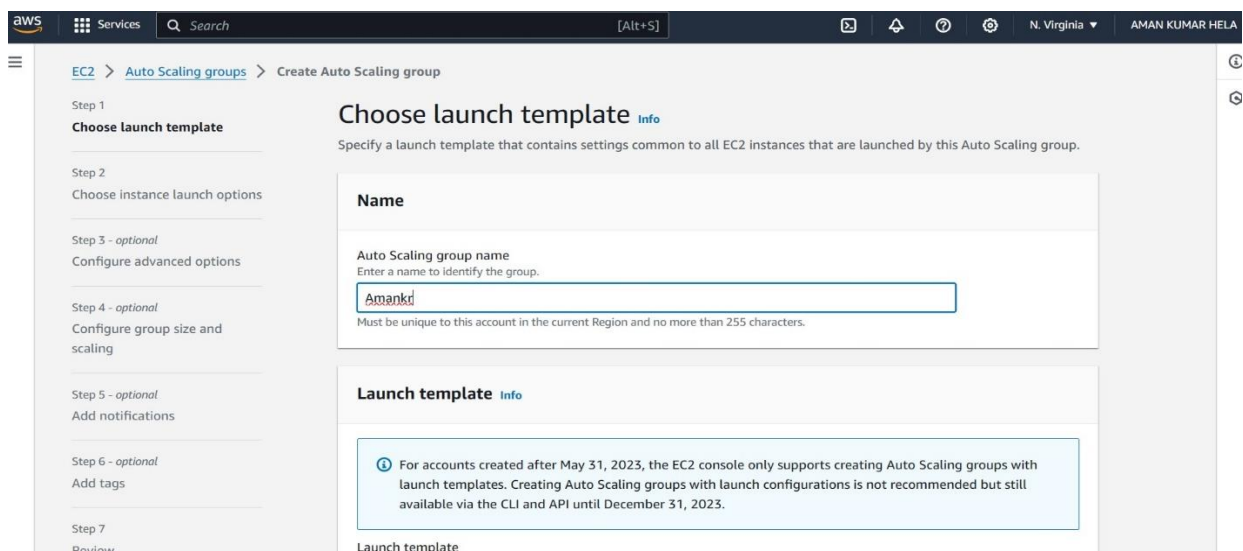
8. Select an existing security group, such as "Aman" which is already in place.

9. Expand the **"Advanced details"** section, navigate to **"User data"**, and input the provided code. Then proceed to click on **"Create launch template"**.

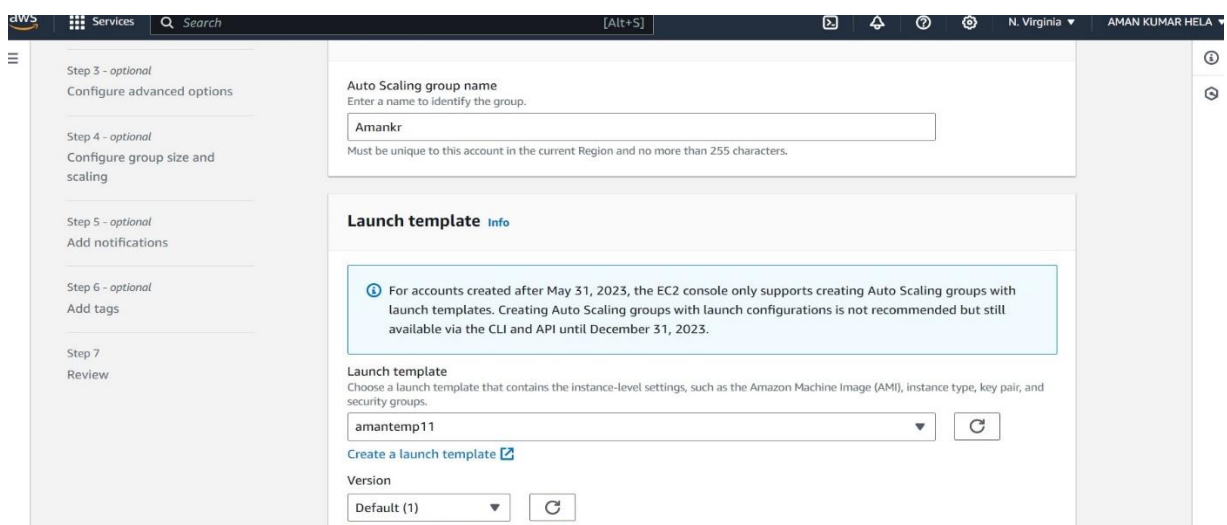
10. Once the launch template has been successfully created, navigate to the left pane and search for **"Auto Scaling Groups"**. Then, select **"Create Auto Scaling Group"**.



11. Please specify a name for the scaling group (e.g., "Amankr").



12. Select the template that was created in the preceding steps.



13. Proceed to click on **"Next"**.

aws Services Search [Alt+S] N. Virginia AMAN KUMAR HELA

Create a launch template

Version: Default (1)

Create a launch template version

Description: -

Launch template: amantemp11

Instance type: t2.micro

AMI ID: ami-04b70fa74e45c3917

Security groups: -

Request Spot Instances: No

Key pair name: keyy_11

Security group IDs: sg-0de34a12c68f20be2

Additional details

Storage (volumes): -

Date created: Wed May 01 2024 14:07:28 GMT+0530 (India Standard Time)

Cancel Next

14. In the following step, choose all available availability zones and subnets, then proceed by clicking **"Next"**.

aws Services Search [Alt+S] N. Virginia AMAN KUMAR HELA

Choose the VPC that defines the virtual network for your Auto Scaling group.

vpc-0dacc502cc085e530

Create a VPC

Availability Zones and subnets

Define which Availability Zones and subnets your Auto Scaling group can use in the chosen VPC.

Select Availability Zones and subnets

us-east-1a | subnet-0e150486a8ef61724

us-east-1b | subnet-0011be60a482f933d

us-east-1c | subnet-09762b833e7bc7922

us-east-1d | subnet-0db3af61cb13608c5

us-east-1e | subnet-0237120757929be25

us-east-1f | subnet-00f0a21d6069f0463

Create a subnet

Cancel Skip to review Previous Next

15. In the subsequent step, begin by selecting **"Attach to a new load balancer"**.

aws Services Search [Alt+S] N. Virginia AMAN KUMAR HELA

EC2 > Auto Scaling groups > Create Auto Scaling group

Step 1: Choose launch template

Step 2: Choose instance launch options

Step 3 - optional: Configure advanced options

Step 4 - optional: Configure group size and scaling

Step 5 - optional: Add notifications

Configure advanced options - optional

Integrate your Auto Scaling group with other services to distribute network traffic across multiple servers using a load balancer or to establish service-to-service communications using VPC Lattice. You can also set options that give you more control over health check replacements and monitoring.

Load balancing

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

☐ Attach to an existing load balancer

☒ Attach to a new load balancer

16. Select "**Application Load Balancer**" as the load balancer type and "**Internet-facing**" as the load balancer scheme.

The screenshot shows the 'Attach to a new load balancer' page in the AWS Management Console. The page is titled 'Attach to a new load balancer' and includes a subtitle 'Define a new load balancer to create for attachment to this Auto Scaling group.' The 'Load balancer type' section has two options: 'Application Load Balancer' (selected) and 'Network Load Balancer'. The 'Load balancer name' field is 'Amankr-1'. The 'Load balancer scheme' section has two options: 'Internal' and 'Internet-facing' (selected). The 'Network mapping' section shows the VPC 'vpc-0dacc502cc085e530' and the 'Availability Zones and subnets' section is empty.

17. Modify the HTTP port number from 80 to 4000 and designate the scaling group created for default routing.

The screenshot shows the 'Listeners and routing' page in the AWS Management Console. The 'Protocol' is 'HTTP' and the 'Port' is '4000'. The 'Default routing (forward to)' section has a dropdown menu with 'Create a target group' selected. The 'New target group name' field is 'Amankr-1'. The 'Tags - optional' section has an 'Add tag' button and a note '50 remaining'. The 'VPC Lattice integration options' section is at the bottom.

18. Enable the checkbox to activate health checks and specify a "health check grace period", set here to 224 seconds.

The screenshot shows the 'Health checks' page in the AWS Management Console. The 'EC2 health checks' section has a checkbox 'Always enabled' which is checked. The 'Additional health check types - optional' section has a checkbox 'Turn on Elastic Load Balancing health checks' which is checked. A blue box contains the text: '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 Balancer console'. The 'Health check grace period' section has a dropdown menu with '224' selected and a unit of 'seconds'.

19. Without any further modifications, proceed to click on “Next”.

The screenshot shows the 'Additional settings' section of the AWS console. It includes a 'Health check grace period' of 224 seconds. Under 'Monitoring', the checkbox 'Enable group metrics collection within CloudWatch' is unchecked. Under 'Default instance warmup', the checkbox 'Enable default instance warmup' is also unchecked. At the bottom right, there are four buttons: 'Cancel', 'Skip to review', 'Previous', and 'Next' (highlighted in orange).

20. In this step, specify the desired, minimum, and maximum capacities.

The screenshot shows the 'Group size' and 'Scaling' sections. In the 'Group size' section, 'Units (number of instances)' is selected, and the 'Desired capacity' is set to 2. In the 'Scaling' section, 'Scaling limits' are configured with 'Min desired capacity' at 2 and 'Max desired capacity' at 3. The left sidebar shows a navigation menu with steps 3 through 7, where 'Step 4 - optional' is highlighted as 'Configure group size and scaling'.

21. Next, opt for the "Target Tracking Scaling Policy" and configure the CPU utilization target value to 50. Additionally, set the instance warm-up time to 240 seconds.

The screenshot shows the 'Automatic scaling - optional' section. The 'Target tracking scaling policy' is selected. The 'Scaling policy name' is 'Target Tracking Policy'. The 'Metric type' is 'Average CPU utilization'. The 'Target value' is set to 50. The 'Instance warmup' is set to 240 seconds. At the bottom, there is a checkbox 'Disable scale in to create only a scale-out policy' which is unchecked.

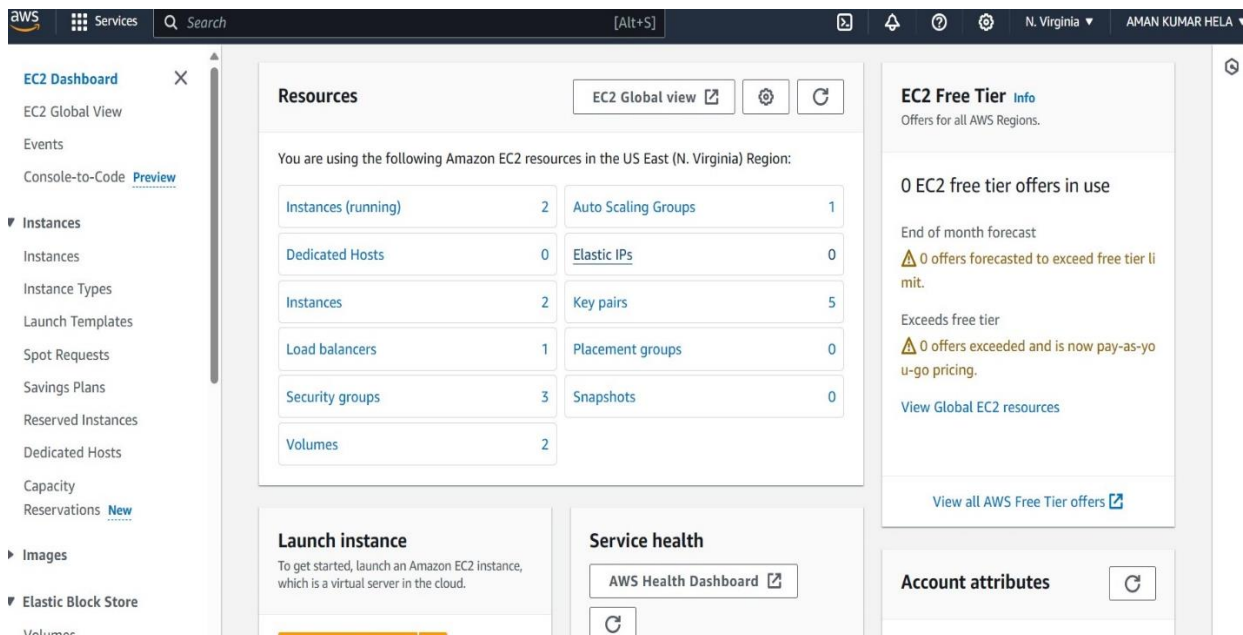
22. Proceed by clicking "**Next**" without making any changes, and finally, select "**Create auto Scaling Group**".

This screenshot shows the 'Add notifications - optional' step in the AWS console. The left sidebar lists the steps: Step 1 (Choose launch template), Step 2 (Choose instance launch options), Step 3 - optional (Configure advanced options), and Step 4 - optional. The main content area has the title 'Add notifications - optional' with an 'Info' icon. Below the title is a description: 'Send notifications to SNS topics whenever Amazon EC2 Auto Scaling launches or terminates the EC2 instances in your Auto Scaling group.' There is an 'Add notification' button. At the bottom right, there are four buttons: 'Cancel', 'Skip to review', 'Previous', and 'Next' (highlighted in orange).

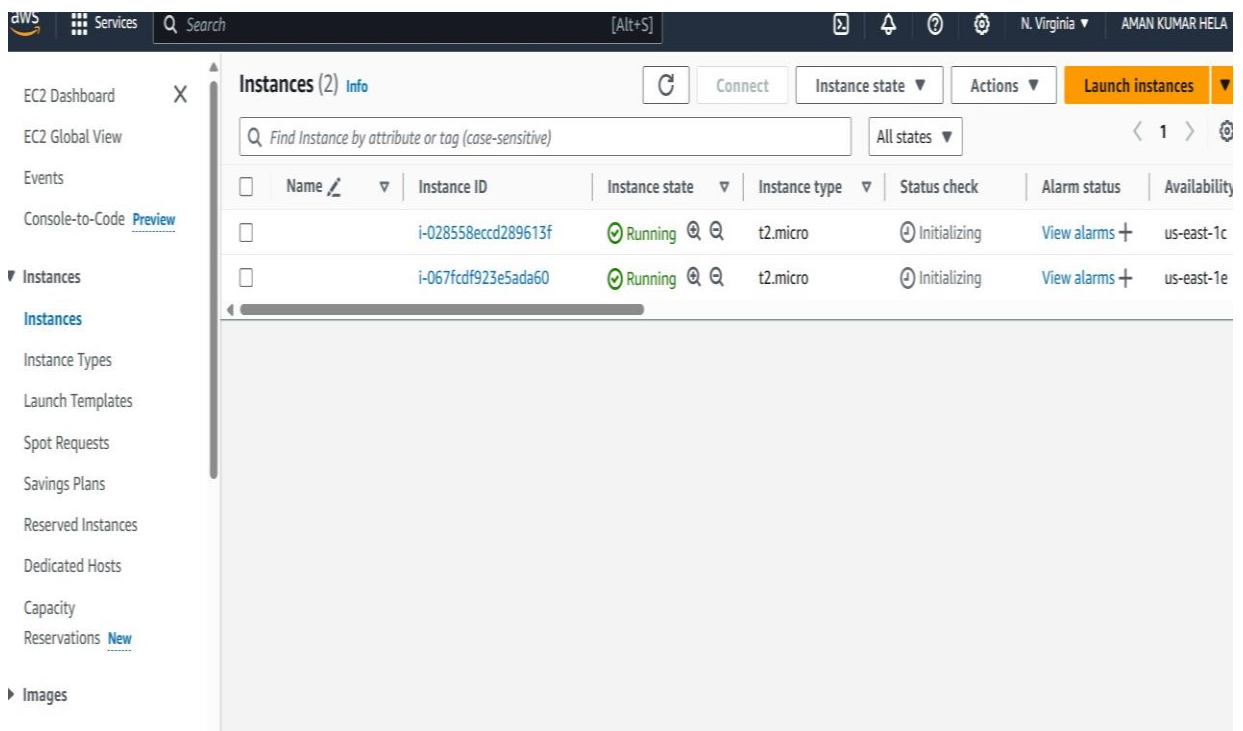
This screenshot shows the 'Add tags - optional' step in the AWS console. The left sidebar lists the steps: Step 1 (Choose launch template), Step 2 (Choose instance launch options), Step 3 - optional (Configure advanced options), Step 4 - optional (Configure group size and scaling), Step 5 - optional (Add notifications), and Step 6 - optional (Add tags). The main content area has the title 'Add tags - optional' with an 'Info' icon. Below the title is a description: 'Add tags to help you search, filter, and track your Auto Scaling group across AWS. You can also choose to automatically add these tags to instances when they are launched.' There is an information box stating: 'You can optionally choose to add tags to instances (and their attached EBS volumes) by specifying tags in your launch template. We recommend caution, however, because the tag values for instances from your launch template will be overridden if there are any duplicate keys specified for the Auto Scaling group.' Below this is a section titled 'Tags (0)' with an 'Add tag' button and '50 remaining'. At the bottom right, there are three buttons: 'Cancel', 'Previous', and 'Next' (highlighted in orange).

This screenshot shows the 'Create Auto Scaling group' step in the AWS console. The left sidebar lists the steps: Step 1 (Choose launch template), Step 2 (Choose instance launch options), Step 3 - optional (Configure advanced options), Step 4 - optional (Configure group size and scaling), Step 5 - optional (Add notifications), and Step 6 - optional (Add tags). The main content area shows the 'Instance scale-in protection' section with a checkbox 'Enable instance protection from scale in'. Below this is 'Step 5: Add notifications' with an 'Edit' button and a 'Notifications' section showing 'No notifications'. Below that is 'Step 6: Add tags' with an 'Edit' button and a 'Tags (0)' section. The 'Tags (0)' section has a table with columns 'Key', 'Value', and 'Tag new instances'. The table is currently empty, showing 'No tags'. At the bottom right, there are three buttons: 'Cancel', 'Previous', and 'Create Auto Scaling group' (highlighted in orange).

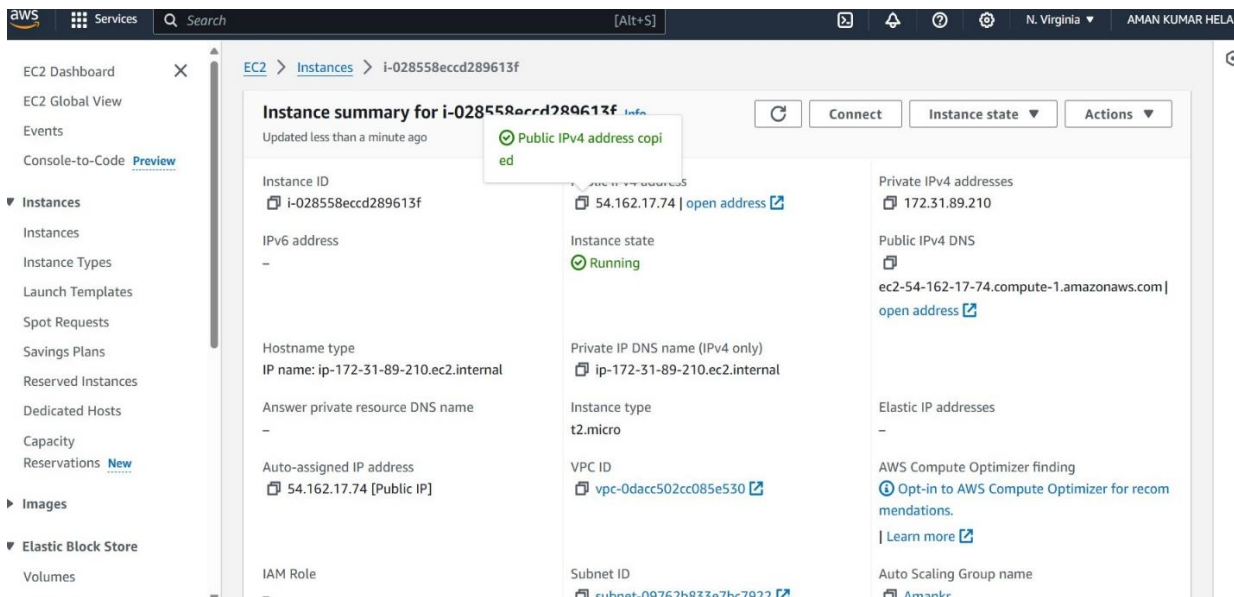
23. After the auto-scaling group is created, return to the EC2 dashboard and navigate to the **"Instances"** section for running instances.



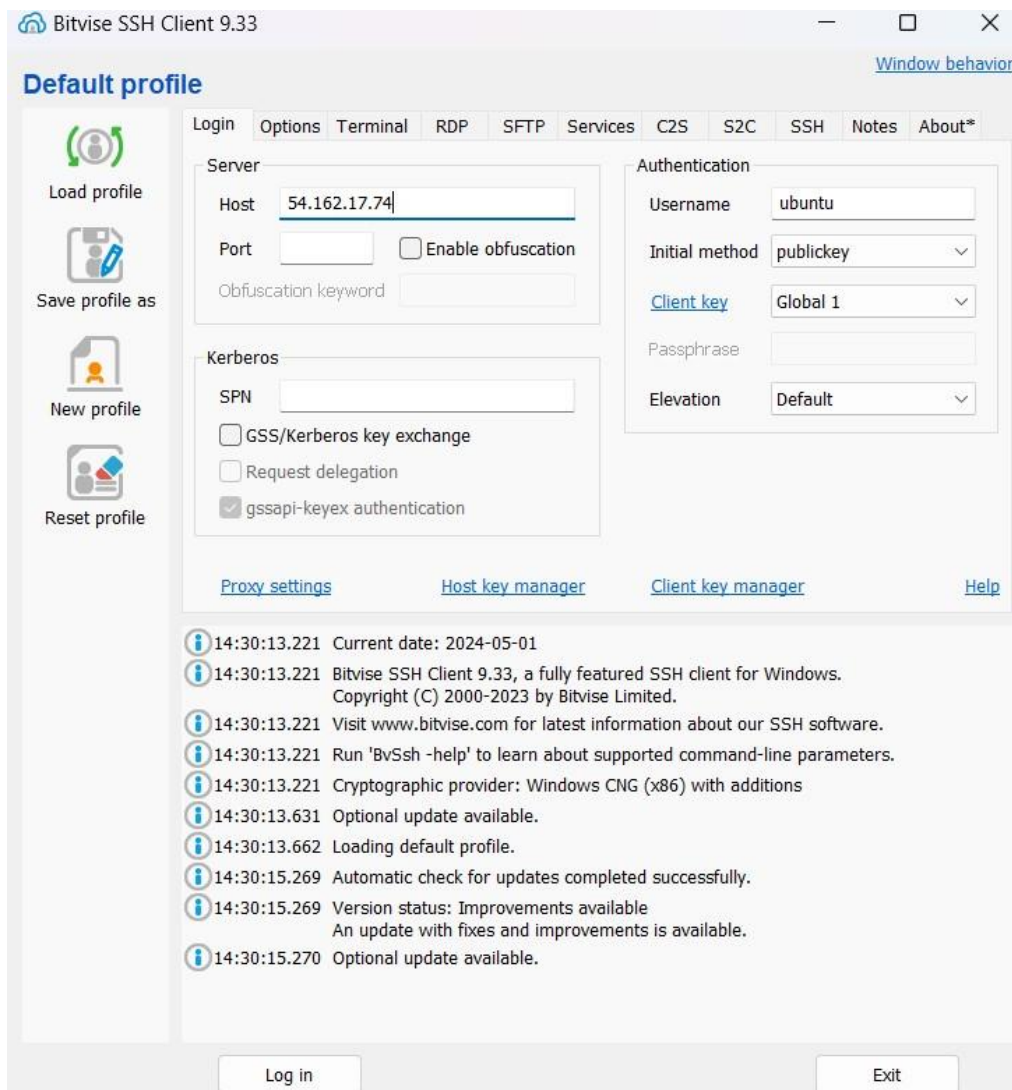
24. Since the minimum capacity chosen was 2, two instances have been created.



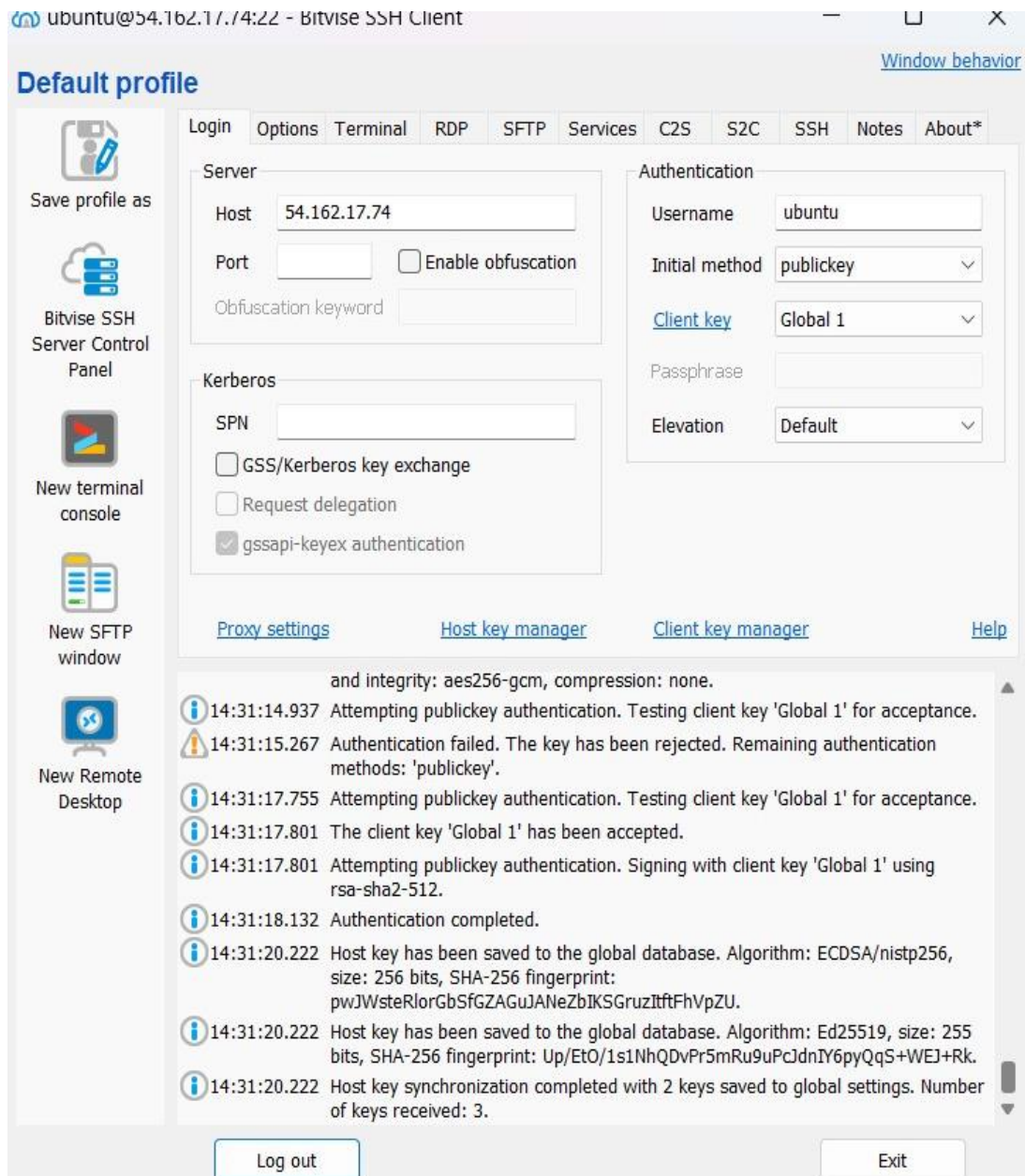
25. Choose any one of the instance IDs and copy the public IPv4 address.



26. Launch the Bitwise SSH Client and log in using the IPv4 address you copied earlier.



27. After successfully logging in, open a new terminal console from the left pane.



28. Now write the commands in the terminal as follows: -

→ ***sudo nano infi.sh*** (creates a .sh file)

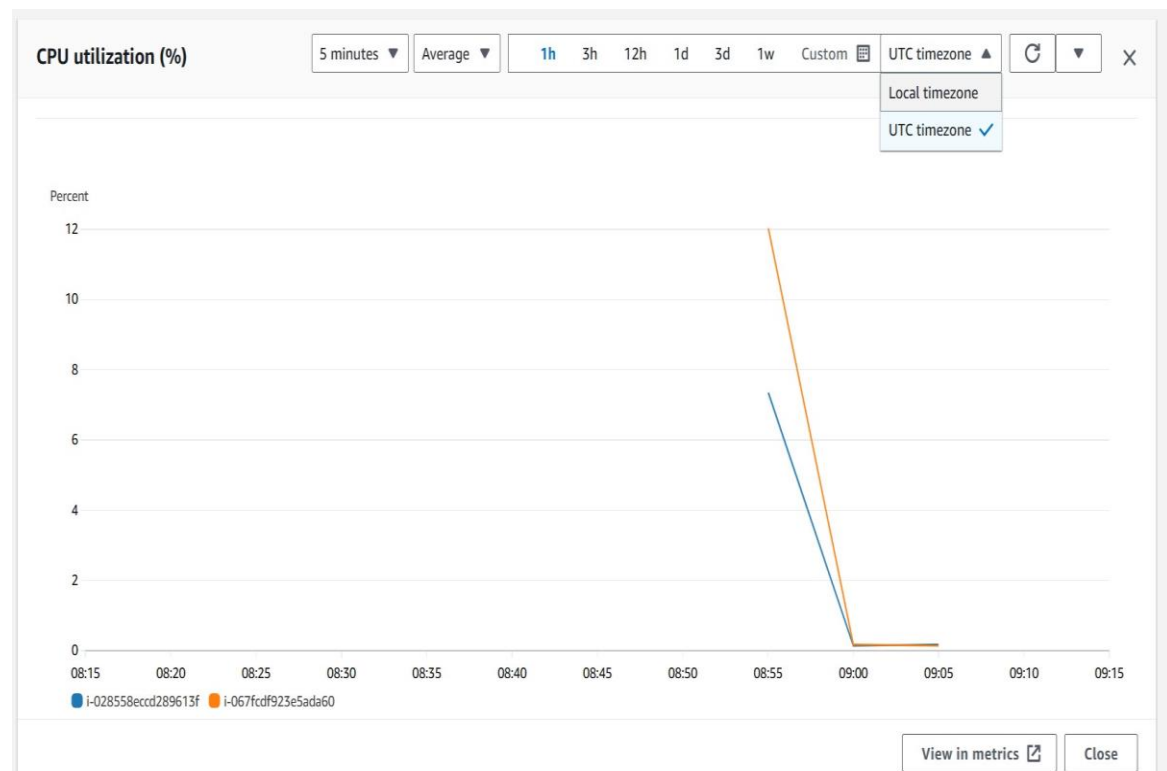
```
ubuntu@ip-172-31-89-210:~$ sudo nano infi.sh
```

→ Write this code in the file "infi.sh" to run an infinite loop.

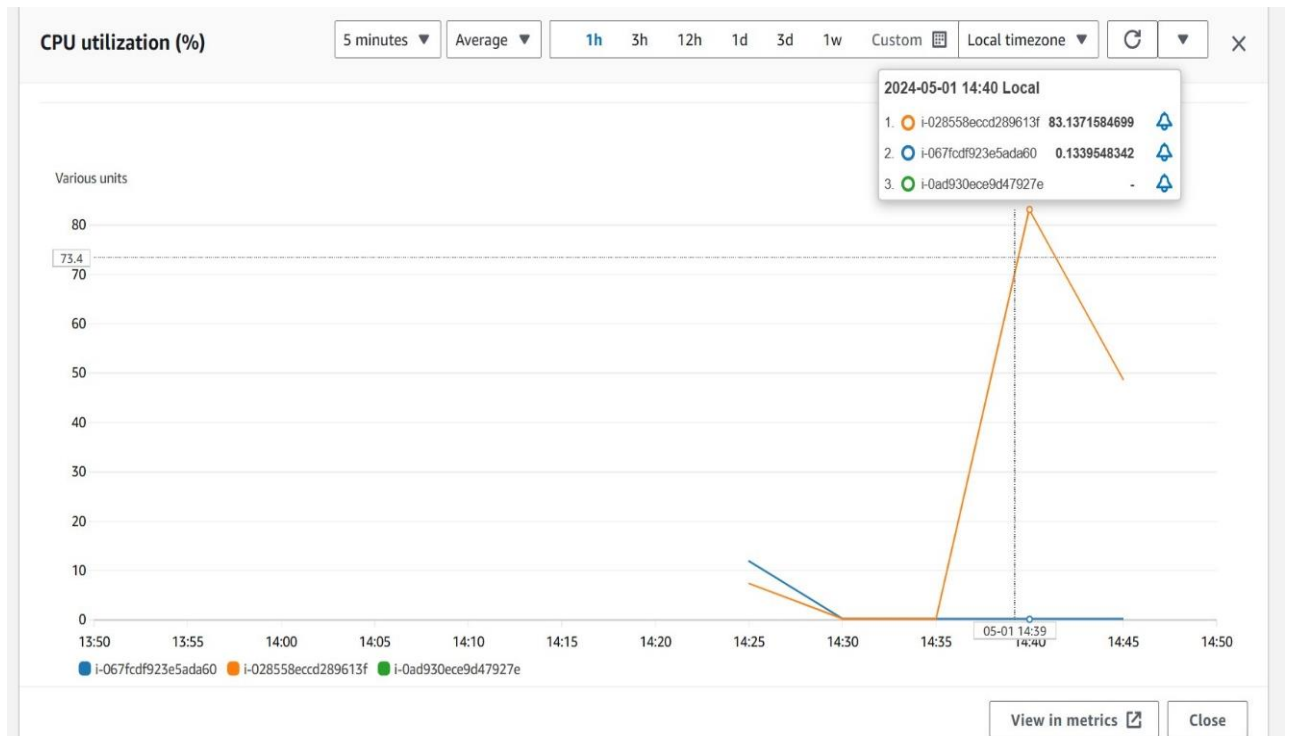
29. Return to AWS and select both running instances. Below, locate the monitoring options, and choose "CPU utilization". Then, enlarge the view.

The screenshot shows the AWS Management Console interface. On the left, the navigation pane lists various services, with 'Instances' selected. The main content area shows 'Instances (2/2) Info'. Two instances are listed: i-028558eccd289613f and i-067fcd923eSada60, both in a 'Running' state. Below the list, the 'Monitoring' tab is active, showing four metrics: CPU utilization, Network in (bytes), Network out (bytes), and Network packets in. The CPU utilization graph is highlighted, and an 'Enlarge' button is visible.

30. From the panel above, select "Local timezone."



31. The graph displays CPU utilization for both instances.



When the CPU utilization exceeds the limit for both instances, another instance is created, as we have set the maximum capacity to 3.

aws Services Search [Alt+S] N. Virginia AMAN KUMAR HELA

EC2 Dashboard X

EC2 Global View

Events

Console-to-Code Preview

Instances

Instances

Instance Types

Launch Templates

Spot Requests

Savings Plans

Reserved Instances

Dedicated Hosts

Instances (3/3) Info

Find Instance by attribute or tag (case-sensitive) All states

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone
i-028558eccd289613f	i-028558eccd289613f	Running	t2.micro	2/2 checks passed	View alarms +	us-east-1
i-0ad930ece9d47927e	i-0ad930ece9d47927e	Running	t2.micro	2/2 checks passed	View alarms +	us-east-1
i-067fcd923e5ada60	i-067fcd923e5ada60	Running	t2.micro	2/2 checks passed	View alarms +	us-east-1

3 instances selected

Monitoring

Alarm recommendations

3h 1d 1w 1h UTC timezone Add to dashboard