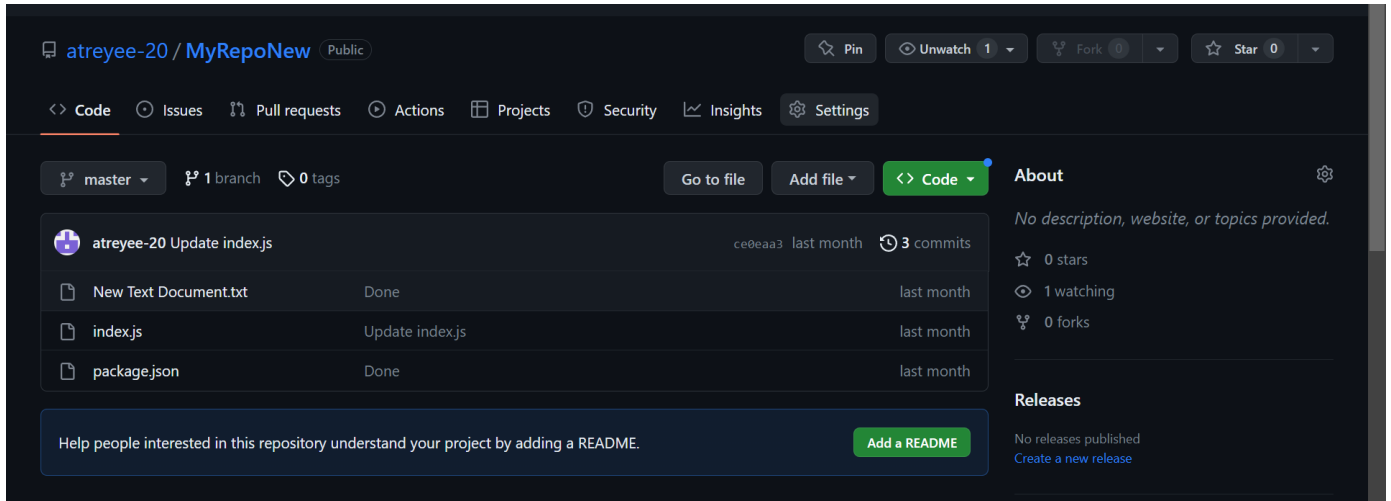


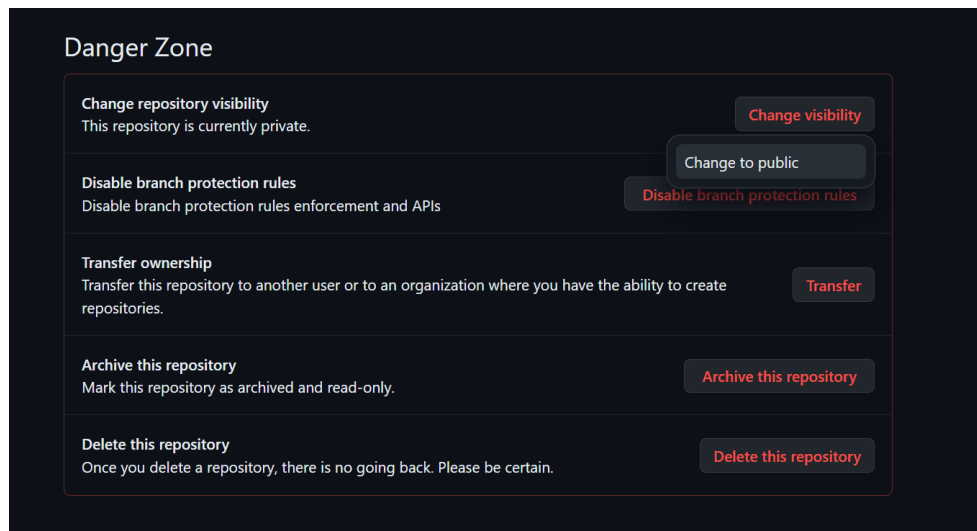
ASSIGNMENT-11

Build Scaling plans in AWS that balance load on different EC2 instances.

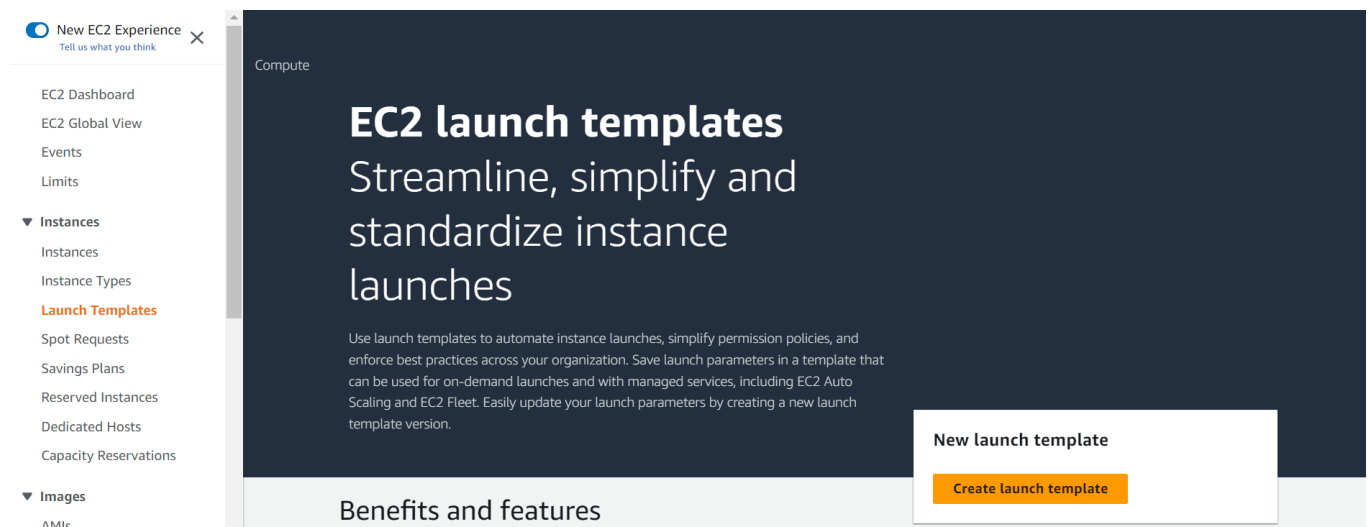
Step 1: Sign in to your GitHub account. Go to your repository. Then go to settings.



Step 2: Scroll down to Danger Zone. Click on Change visibility. Then Change to public.



Step 3: Now Sign in to your AWS account as a root user. Then go to EC2 Dashboard. Click on Launch Templates.



Step 4: Click on Create Launch Template. Give name and description. Then check Auto Scaling guidance box.

The screenshot shows the 'Create launch template' form in the AWS console. The form is titled 'Create launch template' and includes a sub-header 'Launch template name and description'. It contains two text input fields: 'Launch template name - required' with the value 'AtreyeeTemplate' and 'Template version description' with the value 'Template for balancing'. Below these fields is a checkbox labeled 'Auto Scaling guidance' which is checked. To the right of the form is a 'Summary' panel that lists the configuration details: 'Software Image (AMI)', 'Virtual server type (instance type)', 'Firewall (security group)', and 'Storage (volumes)'. At the bottom of the summary panel, there is a 'Free tier' notification box stating 'Free tier: In your first year includes 750 hours of t2.micro (or t3.micro in the Regions in which t2.micro is available)'. The form has 'Cancel' and 'Create launch template' buttons at the bottom right.

Step 5: Select Ubuntu as OS.

▼ Application and OS Images (Amazon Machine Image) - required [Info](#)

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. Search or Browse for AMIs if you don't see what you are looking for below

Q Search our full catalog including 1000s of application and OS images

Recents

Quick Start

Amazon Linux

aws

macOS

Mac

Ubuntu

ubuntu

Windows

Microsoft

Red Hat

Red Hat

Q

Browse more AMIs

Including AMIs from AWS, Marketplace and the Community

Amazon Machine Image (AMI)

Ubuntu Server 22.04 LTS (HVM), SSD Volume Type

Free tier eligible

ami-02eb7a4783e7e9317 (64-bit (x86)) / ami-0a5dcff6fb7af3fc9 (64-bit (Arm))

Virtualization: hvm ENA enabled: true Root device type: ebs

Step 6: Select t2.micro as Instance type and provide a key pair.

▼ Instance type [Info](#)

Advanced

Instance type

t2.micro

Free tier eligible

Family: t2 1 vCPU 1 GiB Memory Current generation: true

On-Demand Linux pricing: 0.0124 USD per Hour

On-Demand Windows pricing: 0.017 USD per Hour

On-Demand RHEL pricing: 0.0724 USD per Hour

On-Demand SUSE pricing: 0.0124 USD per Hour

All generations

Compare instance types

▼ Key pair (login) [Info](#)

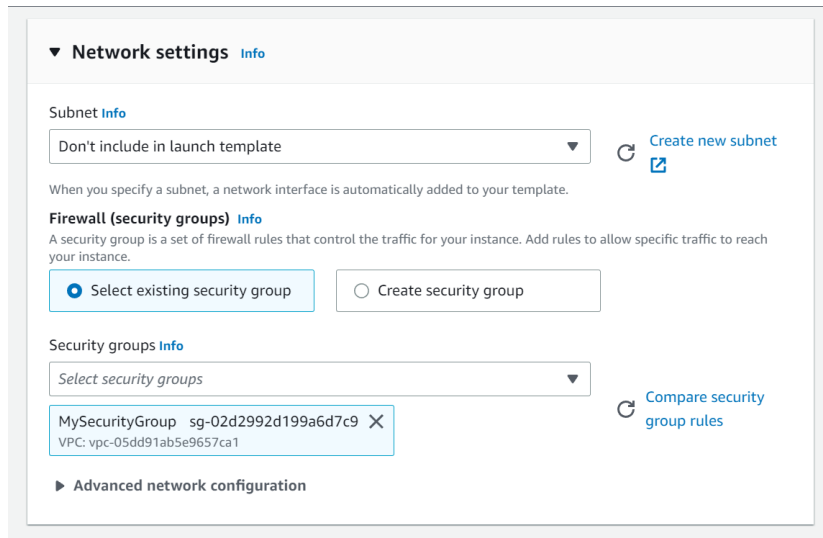
You can use a key pair to securely connect to your instance. Ensure that you have access to the selected key pair before you launch the instance.

Key pair name

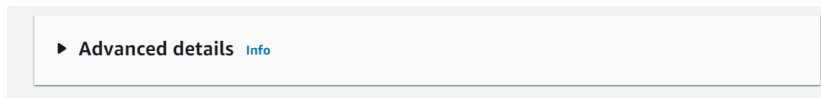
akarkey10

Create new key pair

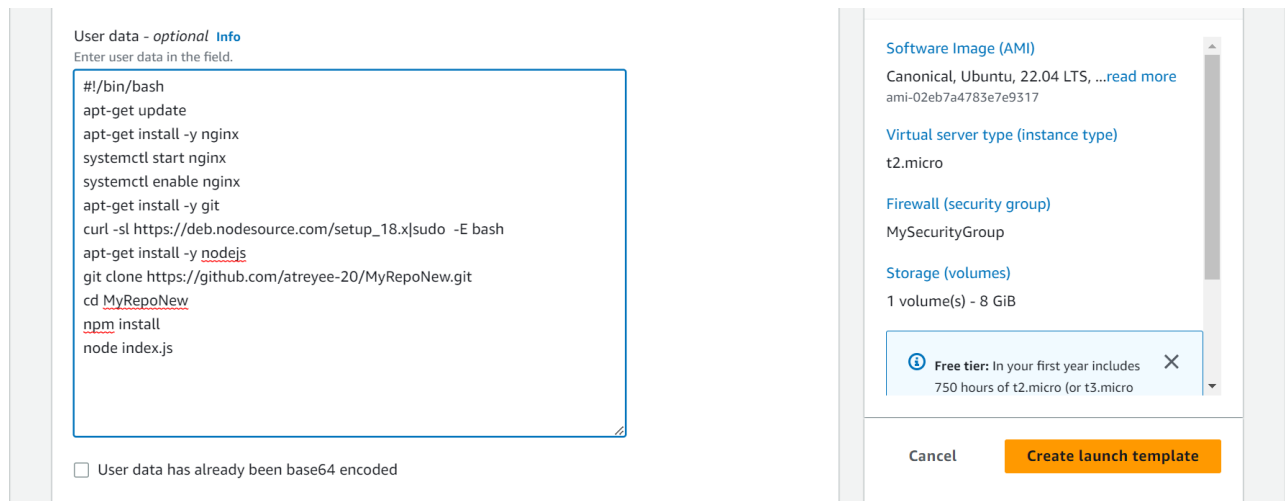
Step 7: Either select existing security group (if there is any) or create new Security Group.



Step 8: Go to Advanced details section and then scroll down to User data.

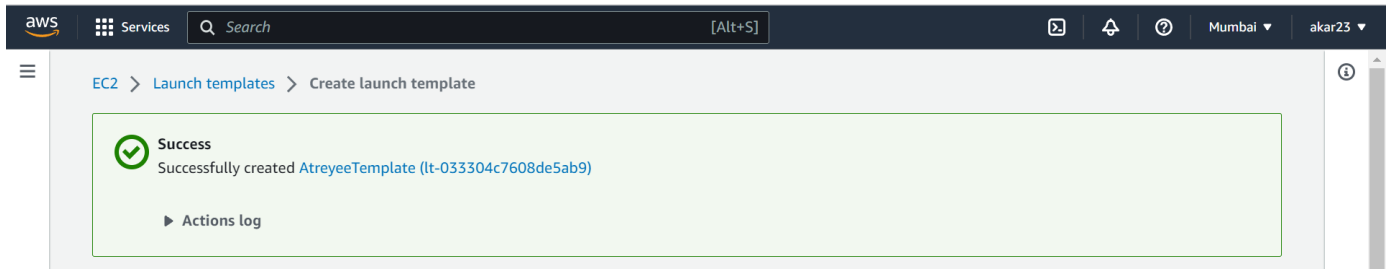


Step 9: Provide the following commands in User data as shown:

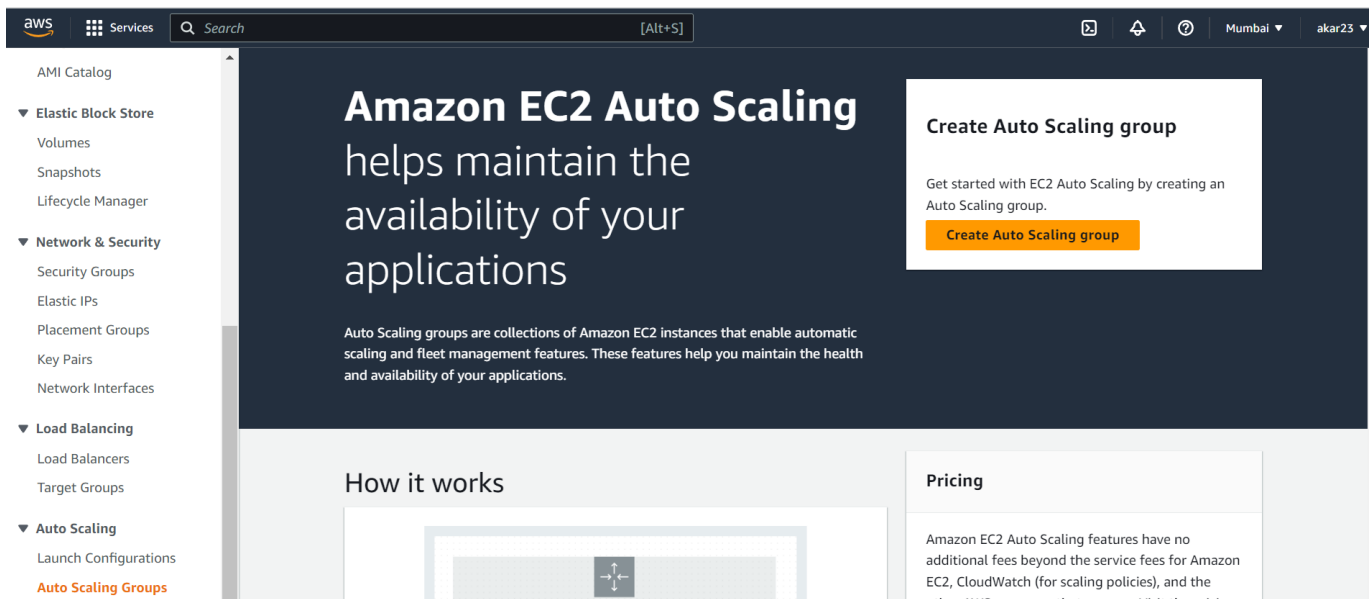


Then Click on Create launch template.

Step 10: New template is created.



Step 11: Again go to EC2 dashboard and click on Auto Scaling Groups. Then click on Create Auto Scaling group.



Step 12: Give a name. Select the template you just created and choose its latest version. Then click on Next.

A screenshot of the 'Create Auto Scaling group' wizard in the AWS console. The 'Name' section has a text input field containing 'akarscaling'. The 'Launch template' section shows a dropdown menu with 'AtreyeeTemplate' selected. The 'Version' section shows a dropdown menu with 'Latest (1)' selected. There are links to 'Create a launch template' and 'Create a launch template version'. A 'Switch to launch configuration' link is also present.

Step 13: Now go to Availability Zones and subnets and select all. Then click on Next.

Network [Info](#)

For most applications, you can use multiple Availability Zones and let EC2 Auto Scaling balance your instances across the zones. The default VPC and default subnets are suitable for getting started quickly.

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

vpc-05dd91ab5e9657ca1
172.31.0.0/16 Default

[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

- ☒ ap-south-1a | subnet-05fd3a11f7421aee6
172.31.32.0/20 Default
- ☒ ap-south-1b | subnet-0914d72889eabc512
172.31.0.0/20 Default
- ☒ ap-south-1c | subnet-08d45765a4df2b58b
172.31.16.0/20 Default

[Create a subnet](#)

Step 14: Select Attach a new load balancer.

Configure advanced options - optional [Info](#)

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 a new load balancer
Define a new load balancer to create for attachment to this Auto Scaling group.

Step 15: In Attach a new load balancer, select balancer type as Application Load Balancer and Load Balancer scheme as Internet-facing.

Attach to a new load balancer
Define a new load balancer to create for attachment to this Auto Scaling group.

Load balancer type
Choose from the load balancer types offered below. Type selection cannot be changed after the load balancer is created. If you need a different type of load balancer than those offered here, [visit the Load Balancing console](#).

☒ Application Load Balancer
HTTP, HTTPS

☐ Network Load Balancer
TCP, UDP, TLS

Load balancer name
Name cannot be changed after the load balancer is created.

akarscaling-1

Load balancer scheme
Scheme cannot be changed after the load balancer is created.

☐ Internal

☒ Internet-facing

Step 16: In Listeners and routing give the port number (For my project, it is 4000). Set the Default routing as “Create a target group”. Automatically New target group name will appear. Then click on Next.

Listeners and routing

If you require secure listeners, or multiple listeners, you can configure them from the [Load Balancing console](#) after your load balancer is created.

Protocol	Port	Default routing (forward to)
HTTP	4000	Create a target group ▼

New target group name

An instance target group with default settings will be created.

akarscaling-1

Tags - optional

Consider adding tags to your load balancer. Tags enable you to categorize your AWS resources so you can more easily manage them.

Add tag

50 remaining

Step 17: Set Group size.

Group size - optional [Info](#)

Specify the size of the Auto Scaling group by changing the desired capacity. You can also specify minimum and maximum capacity limits. Your desired capacity must be within the limit range.

Desired capacity

2

Minimum capacity

2

Maximum capacity

3

Step 18: Set scaling policy as Target tracking scaling policy and provide the time needed for instances. Then click on Next.

Scaling policies - optional

Choose whether to use a scaling policy to dynamically resize your Auto Scaling group to meet changes in demand. [Info](#)

☒ Target tracking scaling policy
Choose a desired outcome and leave it to the scaling policy to add and remove capacity as needed to achieve that outcome.

☐ None

Scaling policy name

Target Tracking Policy

Metric type

Average CPU utilization ▼

Target value

50

Instances need

300 seconds warm up before including in metric

☐ Disable scale in to create only a scale-out policy

Step 19: Again click on Next.

Add notifications - optional [Info](#)

Send notifications to SNS topics whenever Amazon EC2 Auto Scaling launches or terminates the EC2 instances in your Auto Scaling group.

Add notification

CancelSkip to reviewPreviousNext

Step 20: Again click on Next.

Add tags - optional [Info](#)

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.

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.

Tags (0)

Add tag

50 remaining

CancelPreviousNext

Step 21: Review everything and click on “Create Auto Scaling group”.

Step 6: Add tags

Edit

Tags (0)

Key	Value	Tag new instances
No tags		

CancelPreviousCreate Auto Scaling group

Step 22: Now Auto Scaling group is created.

akarscaling, 1 Scaling policy, 1 Load balancer, 1 Target group, 1 Listener created successfully. 1 new target group has been attached to ASG.

EC2 > Auto Scaling groups

Auto Scaling groups (1) [Info](#)

Edit

Delete

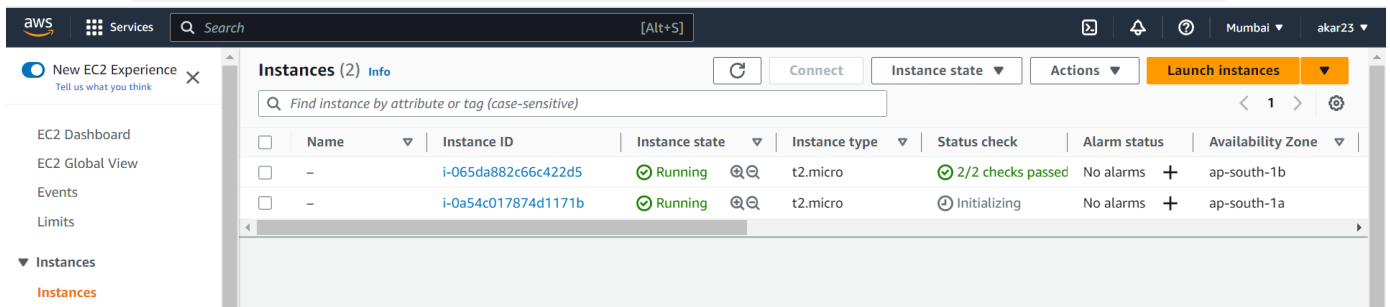
Create an Auto Scaling group

Q Search your Auto Scaling groups

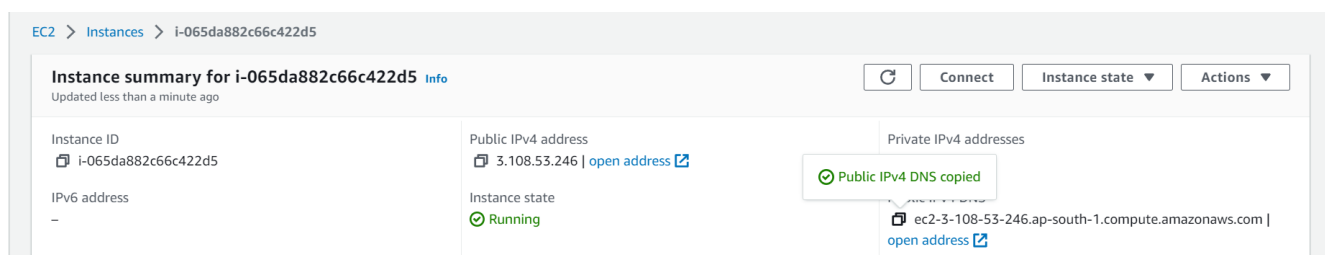
< 1 > ⚙

<input type="checkbox"/>	Name	Launch template/configuration	Instances	Status	Desired capacity	Min	Max
<input type="checkbox"/>	akarscaling	AtreyeeTemplate Version Latest	1	Updating capacity...	2	2	3

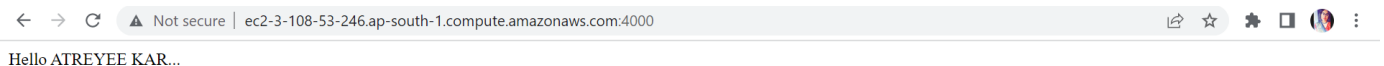
Step 23: Again go to EC2 dashboard. Then go to Instance. You can see that two instance is already created.



Step 24: Open any of the instance and copy the Public IPv4 DNS.



Step 25: Open it in a new browser. Give the port number (:4000) after the address. The content can be seen.

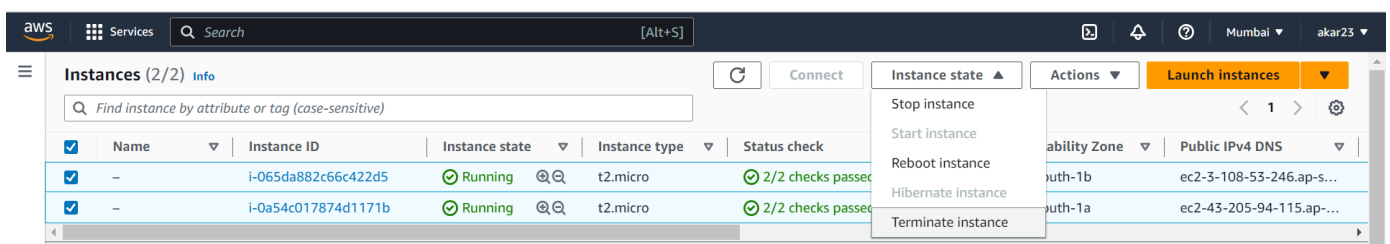


To test whether our Auto-Scaling Group actually works we need to crash or overload the existing instance servers.

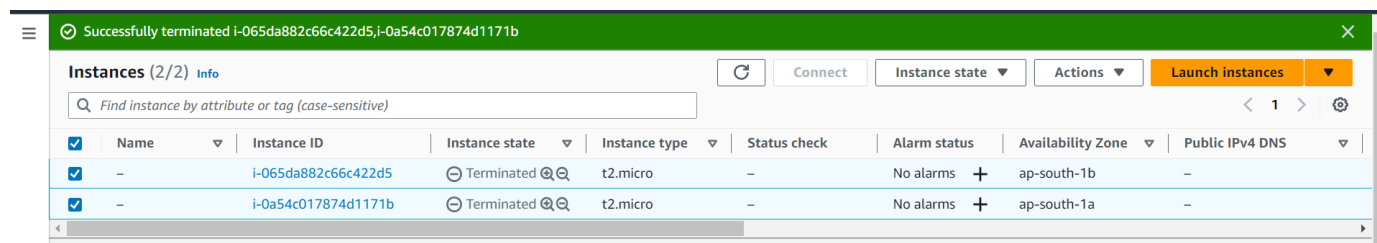
Then only our Auto-Scaling Group will provide fresh instance servers automatically in case of crash or it can provide extra servers to handle overloads.

We will now **CRASH THE SERVER INSTANCES** manually by terminating them.

Step 1: Select both the instances. Go to Instance state and select Terminate instance.



Step 2: Both the instances are terminated.



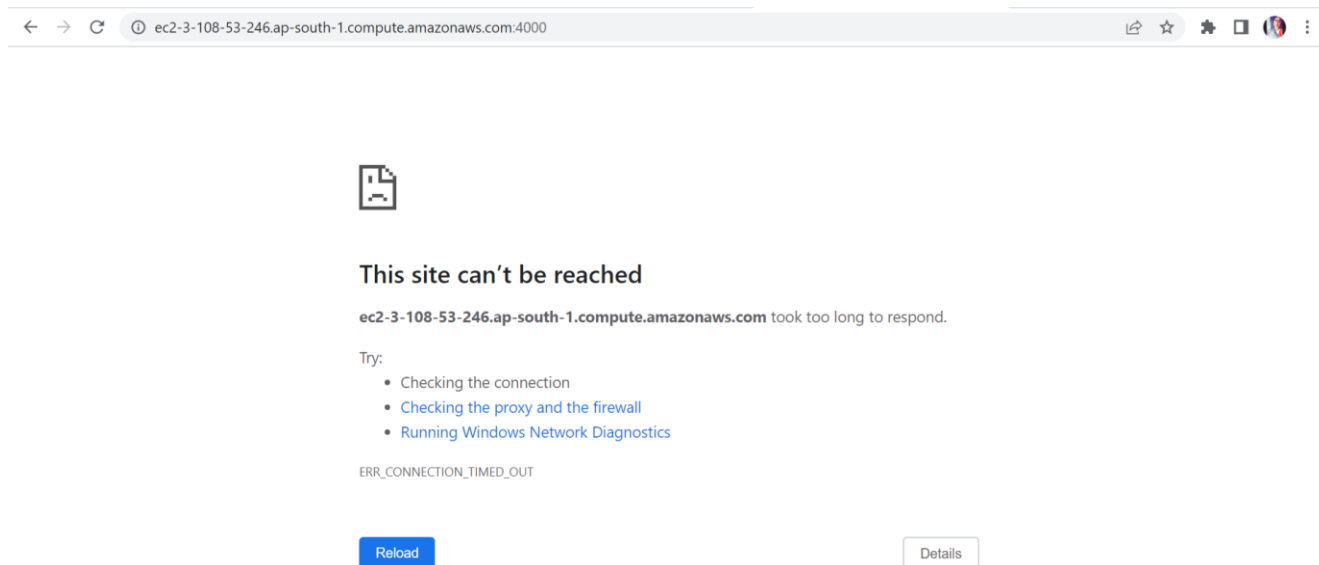
Successfully terminated i-065da882c66c422d5, i-0a54c017874d1171b

Instances (2/2) Info

Find instance by attribute or tag (case-sensitive)

	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS
<input checked="" type="checkbox"/>	-	i-065da882c66c422d5	Terminated	t2.micro	-	No alarms	ap-south-1b	-
<input checked="" type="checkbox"/>	-	i-0a54c017874d1171b	Terminated	t2.micro	-	No alarms	ap-south-1a	-

Step 3: Now refresh the browser. Now, we cannot reach the site.



ec2-3-108-53-246.ap-south-1.compute.amazonaws.com:4000

This site can't be reached

ec2-3-108-53-246.ap-south-1.compute.amazonaws.com took too long to respond.

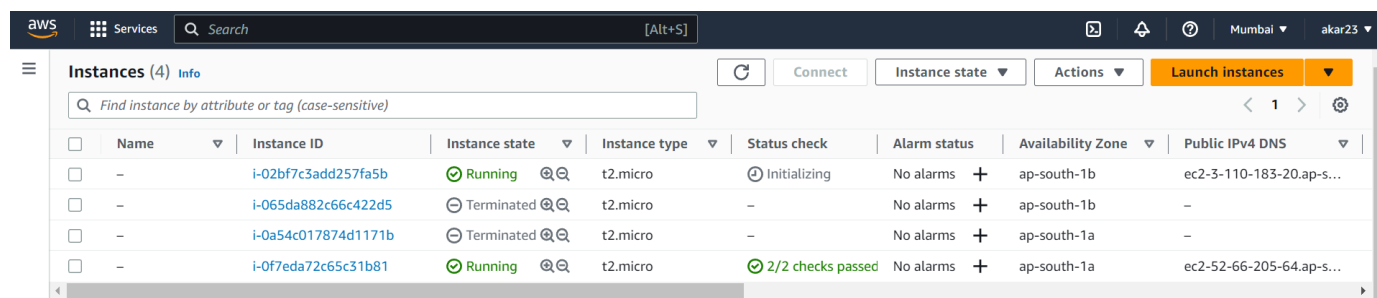
Try:

- Checking the connection
- Checking the proxy and the firewall
- Running Windows Network Diagnostics

ERR_CONNECTION_TIMED_OUT

Reload Details

Step 4: Wait for few seconds. Then refresh the Instances page. We can see two new instances are automatically created.

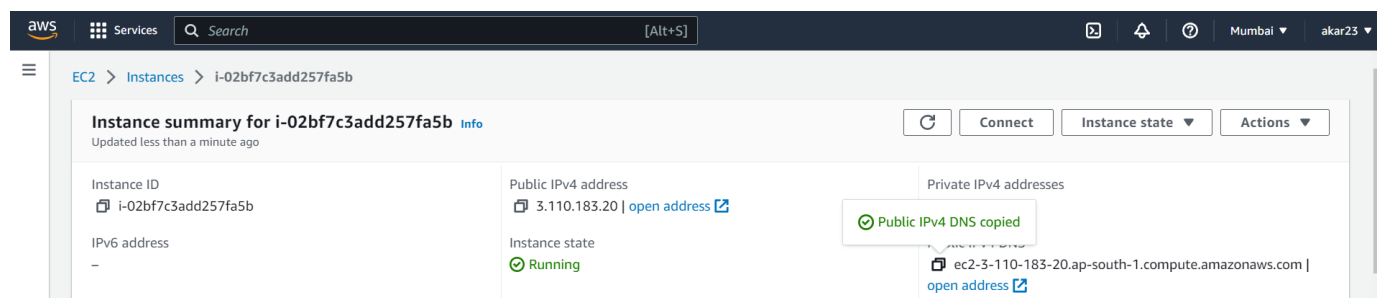


Instances (4) Info

Find instance by attribute or tag (case-sensitive)

	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS
<input type="checkbox"/>	-	i-02bf7c3add257fa5b	Running	t2.micro	Initializing	No alarms	ap-south-1b	ec2-3-110-183-20.ap-s...
<input type="checkbox"/>	-	i-065da882c66c422d5	Terminated	t2.micro	-	No alarms	ap-south-1b	-
<input type="checkbox"/>	-	i-0a54c017874d1171b	Terminated	t2.micro	-	No alarms	ap-south-1a	-
<input type="checkbox"/>	-	i-0f7eda72c65c31b81	Running	t2.micro	2/2 checks passed	No alarms	ap-south-1a	ec2-52-66-205-64.ap-s...

Step 5: Now open any of the running instance. Copy the public IPv4 DNS address.



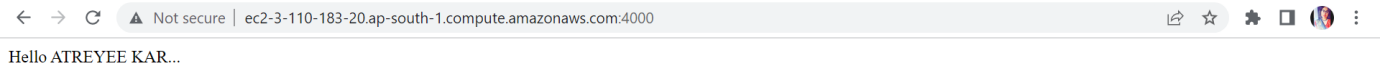
EC2 > Instances > i-02bf7c3add257fa5b

Instance summary for i-02bf7c3add257fa5b

Updated less than a minute ago

Instance ID i-02bf7c3add257fa5b	Public IPv4 address 3.110.183.20 open address	Private IPv4 addresses
IPv6 address -	Instance state Running	Public IPv4 DNS copied ec2-3-110-183-20.ap-south-1.compute.amazonaws.com open address

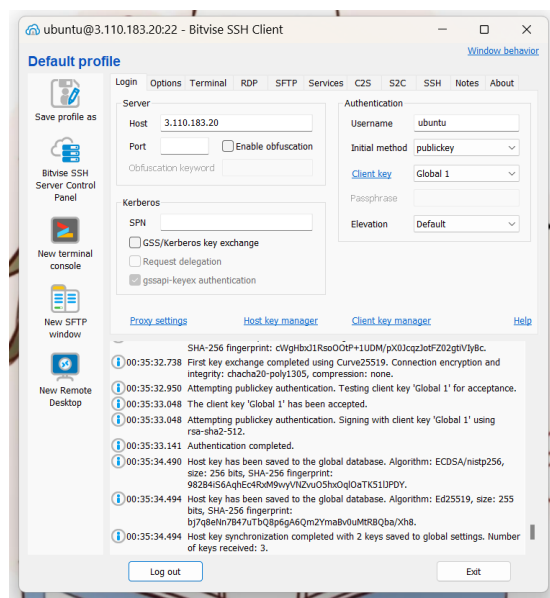
Step 6: Open it in a new browser. Give the port number after the address using ":". Again we can see the content.



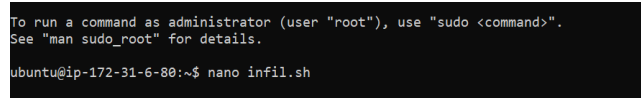
So, our Auto-Scaling Group can handle instance crashing by providing new fresh instances.

Now, we will **CRASH THE SERVER INSTANCES** by overloading them by running scripts.

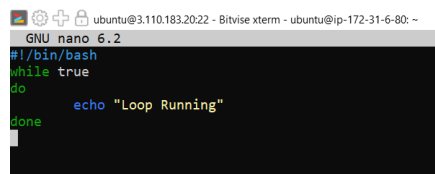
Step 1: Now Log in to Bitvise SSH Client.



Step 2: Now go to New Terminal Console and enter the command : **nano infil.sh**



Step 3: A nano Editor will open. Write the following commands in it:



Step 4: Now, to save and close the shell script we need to press the following shortcuts and keys sequentially: **Ctrl X → Y → Enter**. Then you will return to the terminal.

```
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

ubuntu@ip-172-31-6-80:~$ nano infil.sh
ubuntu@ip-172-31-6-80:~$
```

Step 5: Give execute permission to the .sh file. Then execute it.

- **chmod +x infil.sh** – To give execute permission to the file.
- **./infil.sh** – To execute the file.

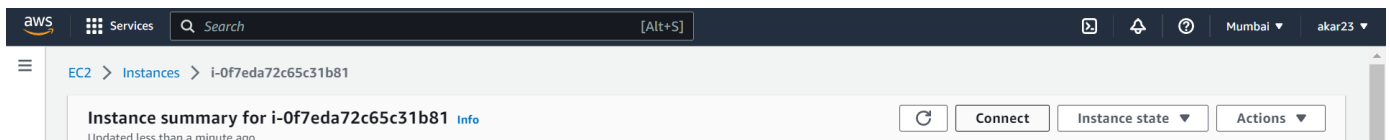
```
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

ubuntu@ip-172-31-6-80:~$ nano infil.sh
ubuntu@ip-172-31-6-80:~$ chmod +x infil.sh
ubuntu@ip-172-31-6-80:~$ ./infil.sh
```

Step 6: The file is executing. Therefore, the first instance is running infinitely. Keep it as it is.

```
ubuntu@3.110.183.2022 - Bitvise xterm - ubuntu@ip-172-31-6-80: ~
Loop Running
Loop Running
Loop Running
Loop Running
Loop Running
Loop Running
Loop Running
Loop Running
Loop Running
Loop Running
Loop Running
Loop Running
Loop Running
Loop Running
Loop Running
Loop Running
Loop Running
Loop Running
Loop Running
Loop Running
```

Step 7: Now go to the other running instance. And click on Connect. Then again click on Connect.



Connect to instance [Info](#)

Connect to your instance i-0f7eda72c65c31b81 using any of these options


EC2 Instance Connect

Session Manager


SSH client

EC2 serial console

Instance ID


 i-0f7eda72c65c31b81

Public IP address

 52.66.205.64

User name

Enter the user name defined in the AMI used to launch the instance. If you didn't define a custom user name, use the default user name, ubuntu.

 **Note:** In most cases, the default user name, ubuntu, is correct. However, read your AMI usage instructions to check if the AMI owner has changed the default AMI user name.

Cancel

Connect

Step 8: A connect terminal will open. Repeat Step-2 to Step-5 here.

```
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

ubuntu@ip-172-31-34-9:~$ nano infil.sh
ubuntu@ip-172-31-34-9:~$ chmod +x infil.sh
ubuntu@ip-172-31-34-9:~$ ./infil.sh
```

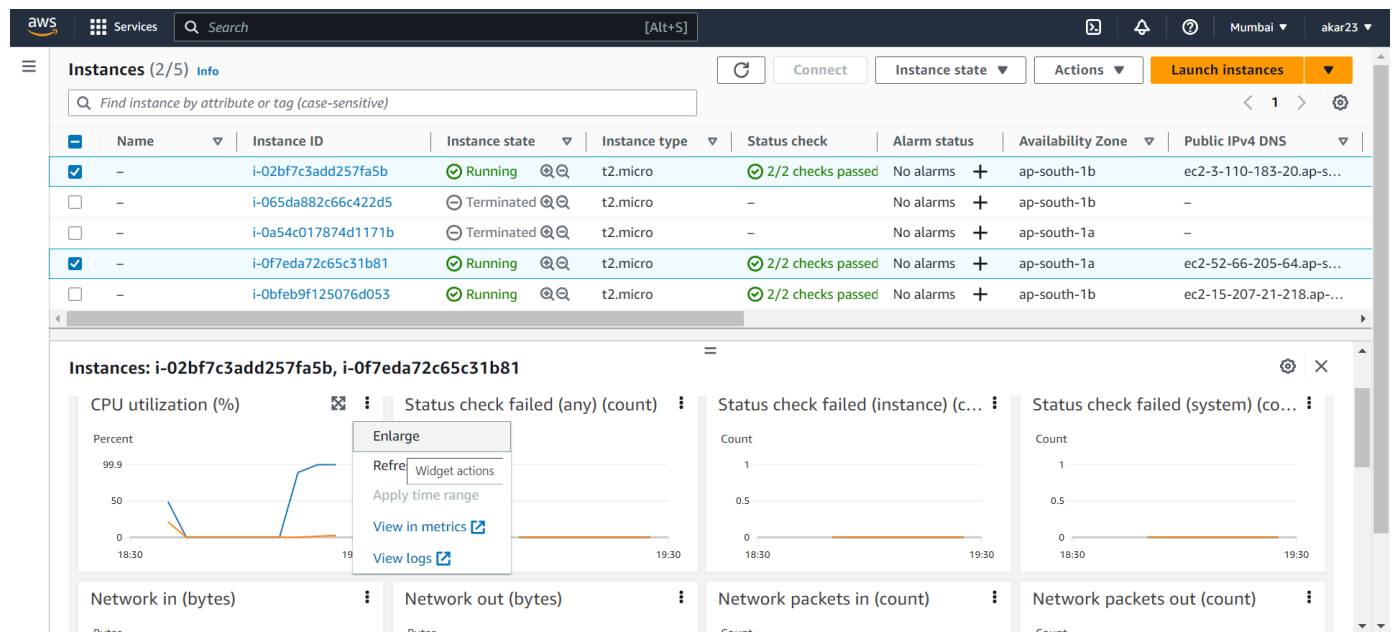
i-Of7eda72c65c31b81

PublicIPs: 52.66.205.64 PrivateIPs: 172.31.34.9

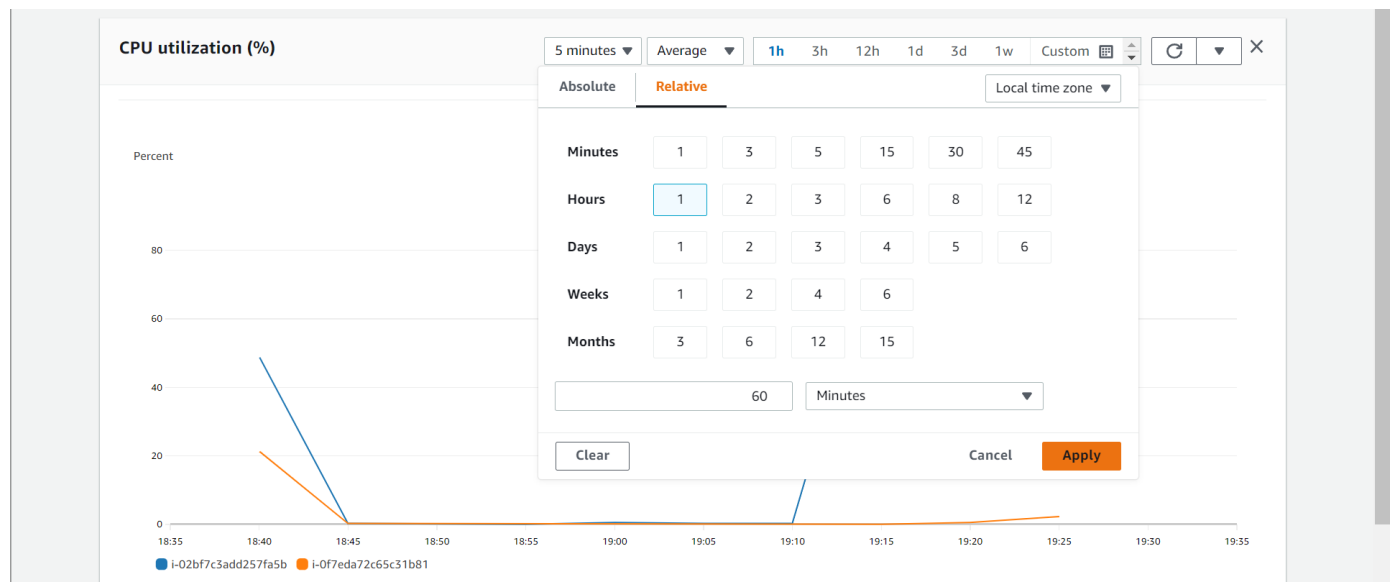
Step 9: Now the second instance is also running infinitely. Keep it as it is.

A screenshot of an AWS CloudShell terminal window. The top bar shows the AWS logo, 'Services' button, a search bar with 'Search' text, and a keyboard shortcut '[Alt+S]'. On the right, there are icons for help, refresh, and a location dropdown set to 'Mumbai', along with a user profile 'akar23'. The terminal area is black with white text, displaying a continuous list of 'Loop Running' messages. At the bottom, a white status bar shows the instance ID 'i-0f7eda72c65c31b81' and IP addresses 'PublicIPs: 52.66.205.64' and 'PrivateIPs: 172.31.34.9'.

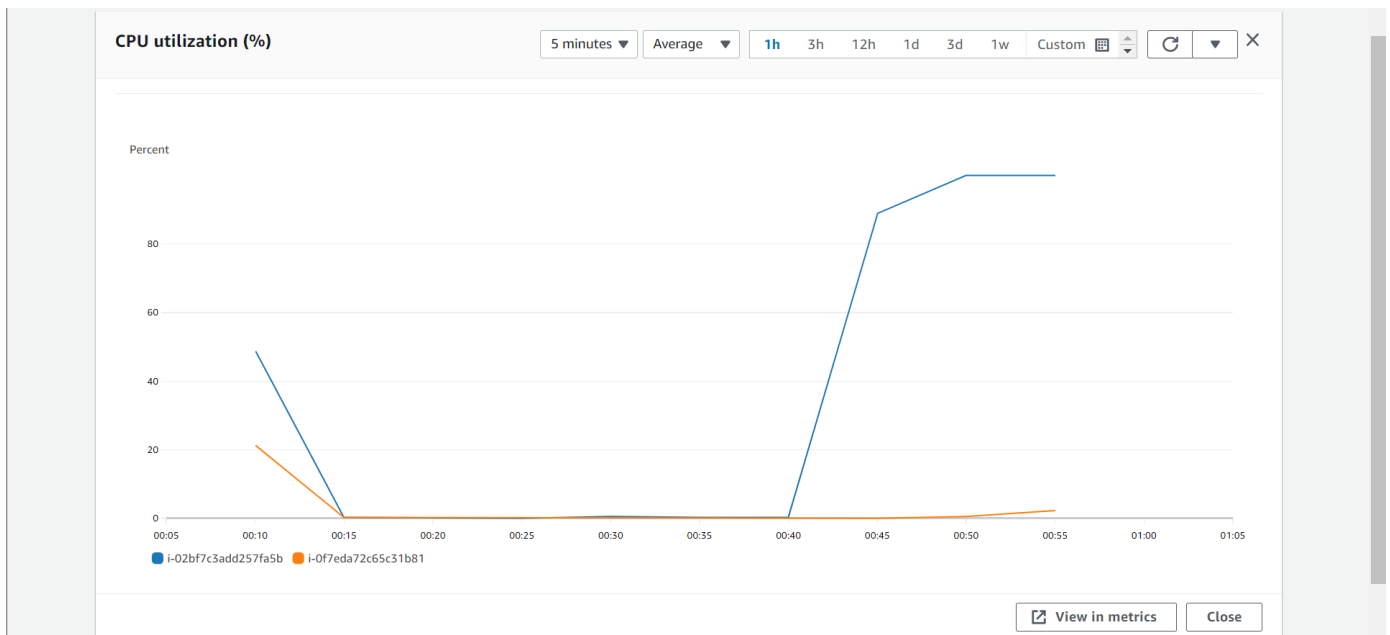
Step 10: Now go to Instance page. Select both the instances. Then scroll down in the “Monitoring” section. Go to CPU Utilization. Click on the three dots. Click on Enlarge.



Step 11: Go to Custom. Select Local time zone. Then click on Apply.



Step 12: We can see our first instance has already reached 50% utilization.



So, our Auto-scaling group has created another instance to compensate for the overload.

Instances (5) Info								
<input type="text" value="Find instance by attribute or tag (case-sensitive)"/>								
<input type="checkbox"/>	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	
<input type="checkbox"/>	-	i-0bfeb9f125076d053	Running	t2.micro	2/2 checks passed	No alarms	ap-south-1b	+
<input type="checkbox"/>	-	i-02bf7c3add257fa5b	Running	t2.micro	2/2 checks passed	No alarms	ap-south-1b	+
<input type="checkbox"/>	-	i-065da882c66c422d5	Terminated	t2.micro	-	No alarms	ap-south-1b	+
<input type="checkbox"/>	-	i-0a54c017874d1171b	Terminated	t2.micro	-	No alarms	ap-south-1a	+
<input type="checkbox"/>	-	i-0f7eda72c65c31b81	Running	t2.micro	2/2 checks passed	No alarms	ap-south-1a	+

Our webpage was not at all disconnected in this whole process. You can check it by refreing the webpage.

If we want to delete the instances permanently, we have to follow the sequence given below. Otherwise, the loop will continue and the instances will be created automatically.

Deleting Sequence:

- Delete Auto-Scaling groups
- Delete Load Balancers
- Delete Target groups
- Delete EC2 Instances

Thus, we have built Scaling plans in AWS that balance load on different EC2 instances.