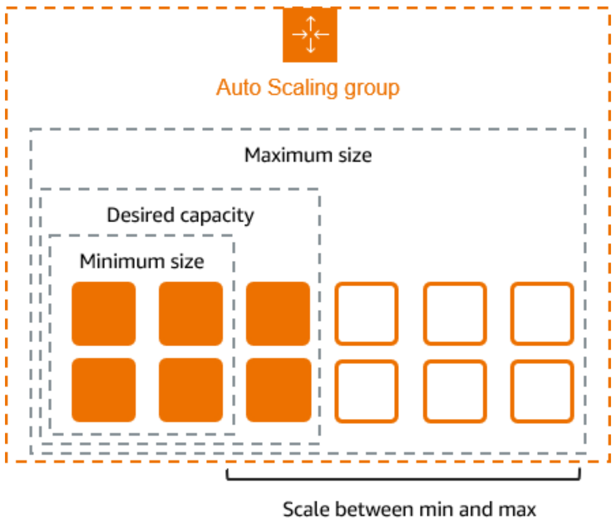
,AWS Documentation

No:-	Content
1.	Auto scaling

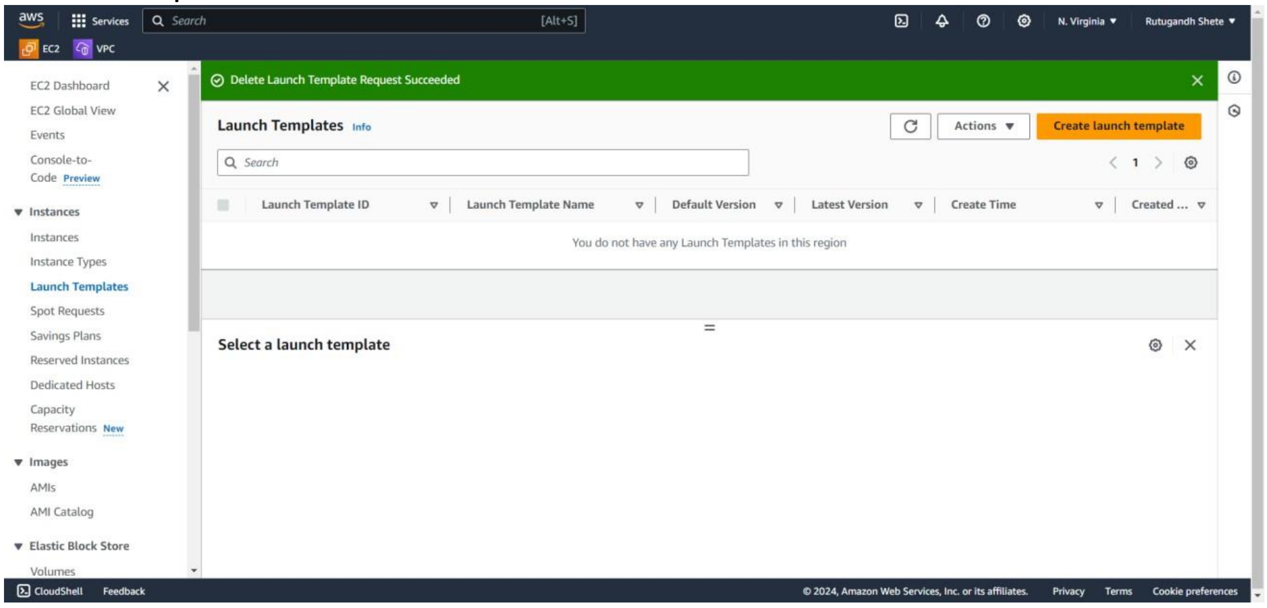
Auto Scaling:
Amazon EC2 Auto Scaling helps you ensure that you have the correct number of Amazon EC2 instances available to handle the load for your application. You create collections of EC2 instances, called *Auto Scaling groups*. You can specify the minimum number of instances in each Auto Scaling group, and Amazon EC2 Auto Scaling ensures that your group never goes below this size. You can specify the maximum number of instances in each Auto Scaling group, and Amazon EC2 Auto Scaling ensures that your group never goes above this size. If you specify the desired capacity, either when you create the group or at any time thereafter, Amazon EC2 Auto Scaling ensures that your group has this many instances. If you specify scaling policies, then Amazon EC2 Auto Scaling can launch or terminate instances as demand on your application increases or decreases.

For example, the following Auto Scaling group has a minimum size of four instances, a desired capacity of six instances, and a maximum size of twelve instances. The scaling policies that you define adjust the number of instances, within your minimum and maximum number of instances, based on the criteria that you specify.



Steps:

- Create template



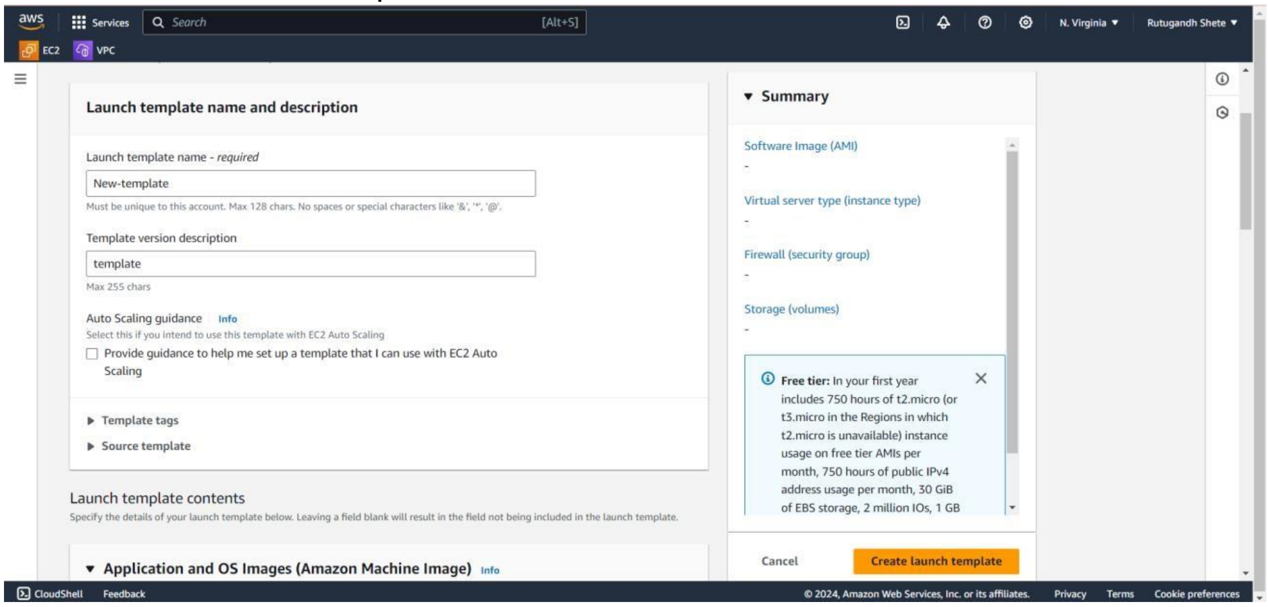
Delete Launch Template Request Succeeded

Launch Templates

You do not have any Launch Templates in this region

Select a launch template

- Provide name and description



Launch template name and description

Launch template name - *required*

New-template

Must be unique to this account. Max 128 chars. No spaces or special characters like '&', '*', '@'.

Template version description

template

Max 255 chars

Auto Scaling guidance

Select this if you intend to use this template with EC2 Auto Scaling

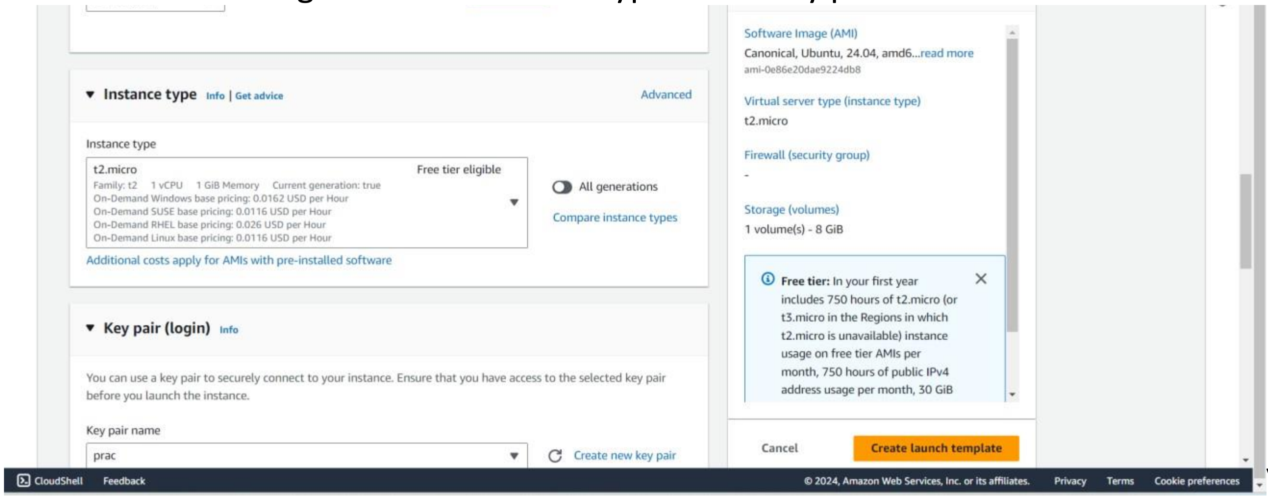
☐ Provide guidance to help me set up a template that I can use with EC2 Auto Scaling

Launch template contents

Specify the details of your launch template below. Leaving a field blank will result in the field not being included in the launch template.

Free tier: In your first year includes 750 hours of t2.micro (or t3.micro in the Regions in which t2.micro is unavailable) instance usage on free tier AMIs per month, 750 hours of public IPv4 address usage per month, 30 GiB of EBS storage, 2 million IOs, 1 GB

- Select machine image→select instance type→add key pair value



Instance type

t2.micro — Free tier eligible

Family: t2 1 vCPU 1 GiB Memory Current generation: true
On-Demand Windows base pricing: 0.0162 USD per Hour
On-Demand SUSE base pricing: 0.0116 USD per Hour
On-Demand RHEL base pricing: 0.026 USD per Hour
On-Demand Linux base pricing: 0.0116 USD per Hour

Additional costs apply for AMIs with pre-installed software

Key pair (login)

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: prac

Software Image (AMI): Canonical, Ubuntu, 24.04, amd6...read more ami-0e86e20dae9224db8

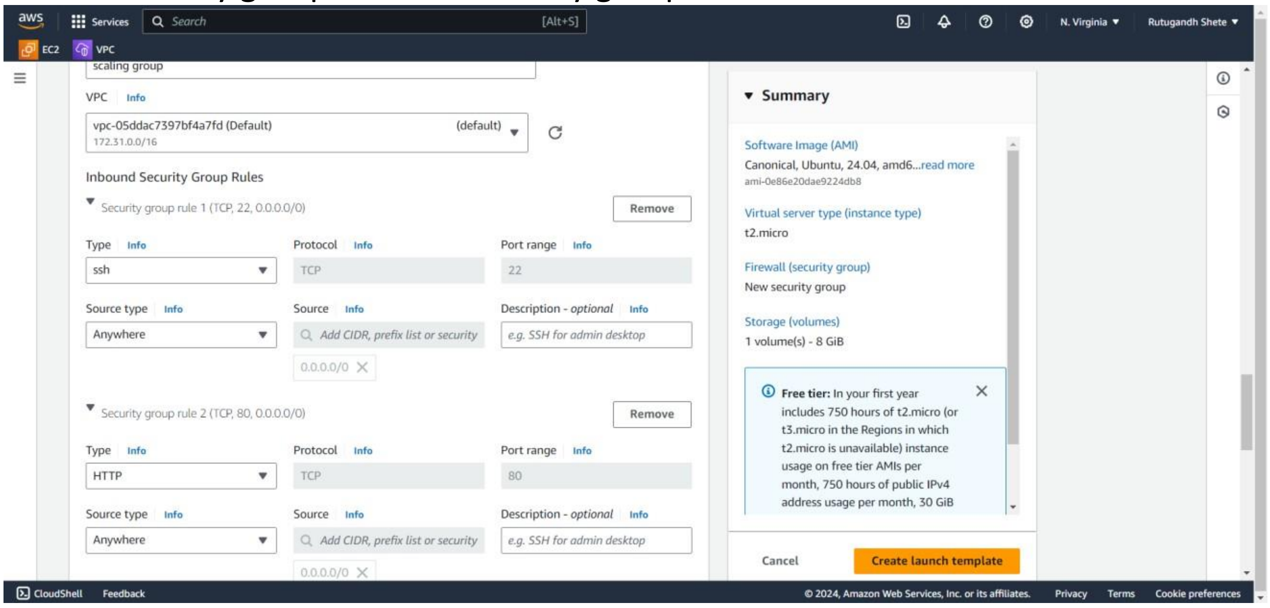
Virtual server type (instance type): t2.micro

Firewall (security group): -

Storage (volumes): 1 volume(s) - 8 GiB

\

- Create security group and add security groups rule.



scaling group

VPC Info

vpc-05ddac7397bf4a7fd (Default) (default)
172.31.0.0/16

Inbound Security Group Rules

Security group rule 1 (TCP, 22, 0.0.0.0/0) Remove

Type	Protocol	Port range
ssh	TCP	22

Source type	Source	Description - *optional*
Anywhere	Add CIDR, prefix list or security / 0.0.0.0/0	e.g. SSH for admin desktop

Security group rule 2 (TCP, 80, 0.0.0.0/0) Remove

Type	Protocol	Port range
HTTP	TCP	80

Source type	Source	Description - *optional*
Anywhere	Add CIDR, prefix list or security / 0.0.0.0/0	e.g. SSH for admin desktop

Summary

Software Image (AMI)
Canonical, Ubuntu, 24.04, amd6...read more
ami-0e86e20dae9224db8

Virtual server type (instance type)
t2.micro

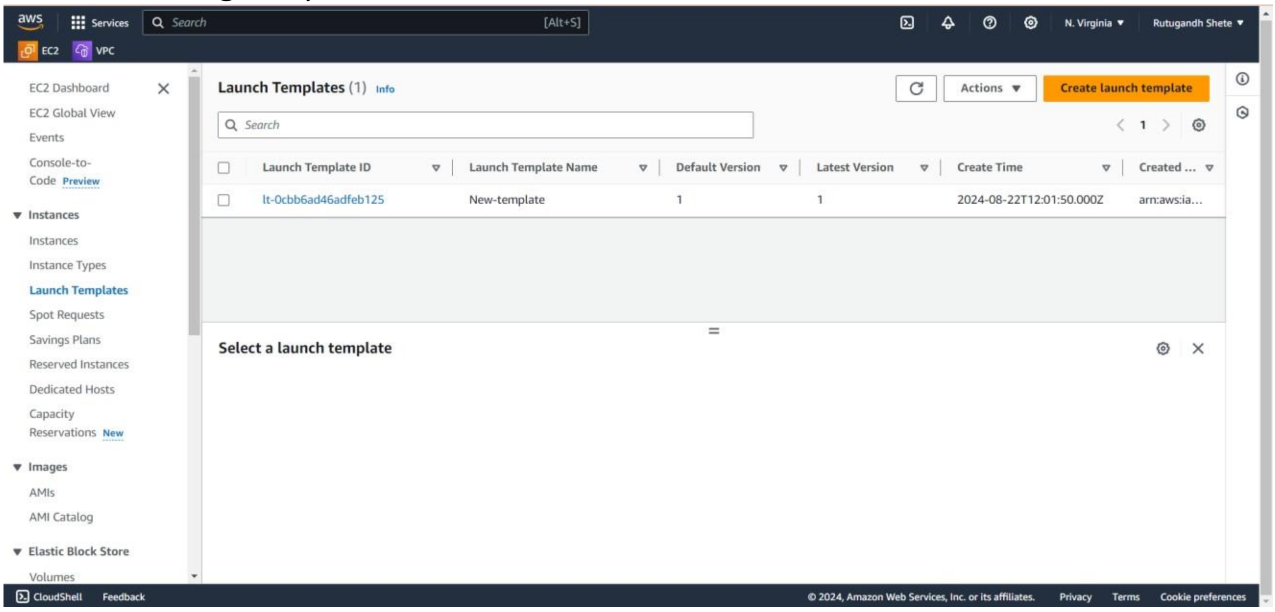
Firewall (security group)
New security group

Storage (volumes)
1 volume(s) - 8 GiB

Free tier: In your first year includes 750 hours of t2.micro (or t3.micro in the Regions in which t2.micro is unavailable) instance usage on free tier AMIs per month, 750 hours of public IPv4 address usage per month, 30 GiB

Cancel Create launch template

- After launching template.

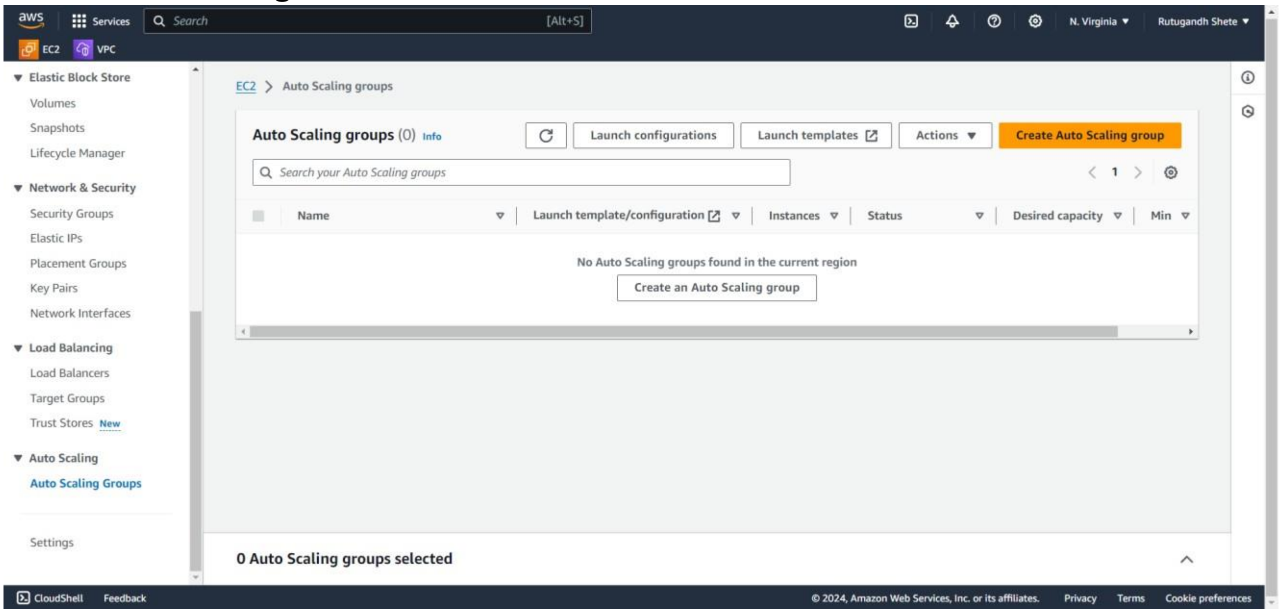


Launch Templates (1) Info

	Launch Template ID	Launch Template Name	Default Version	Latest Version	Create Time	Created ...
☐	lt-0cbb6ad46adfeb125	New-template	1	1	2024-08-22T12:01:50.000Z	arn:aws:ia...

Select a launch template

- Create auto scaling



EC2 > Auto Scaling groups

Auto Scaling groups (0) Info

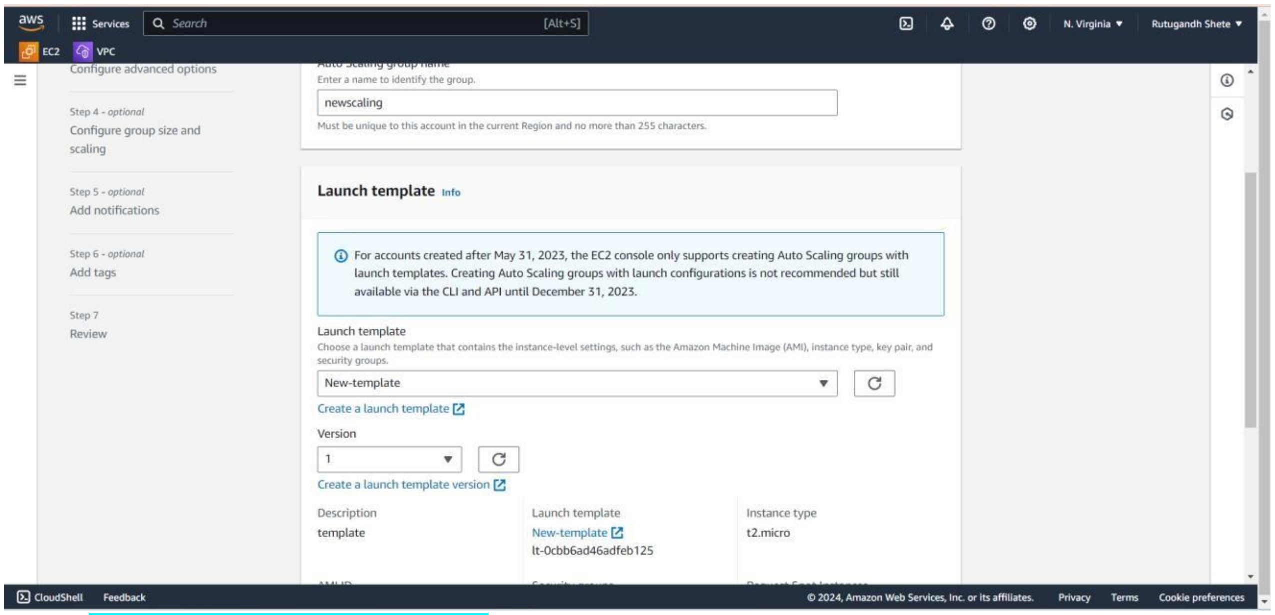
Name	Launch template/configuration	Instances	Status	Desired capacity	Min

No Auto Scaling groups found in the current region

Create an Auto Scaling group

0 Auto Scaling groups selected

- Add name→select template that we have created→select version
  Default Version:
  When you create a launch template, AWS automatically designates the first version as the "default version."
  The default version is used by the Auto Scaling group unless you specify a different version.
  Specifying a Version:
  When configuring your Auto Scaling group, you can specify which version of the launch template to use.
  If you don't specify a version, the Auto Scaling group will use the default version.



Enter a name to identify the group.

newscaling

Must be unique to this account in the current Region and no more than 255 characters.

Launch template Info

For accounts created after May 31, 2023, the EC2 console only supports creating Auto Scaling groups with launch templates. Creating Auto Scaling groups with launch configurations is not recommended but still available via the CLI and API until December 31, 2023.

Launch template
Choose a launch template that contains the instance-level settings, such as the Amazon Machine Image (AMI), instance type, key pair, and security groups.

New-template

Create a launch template

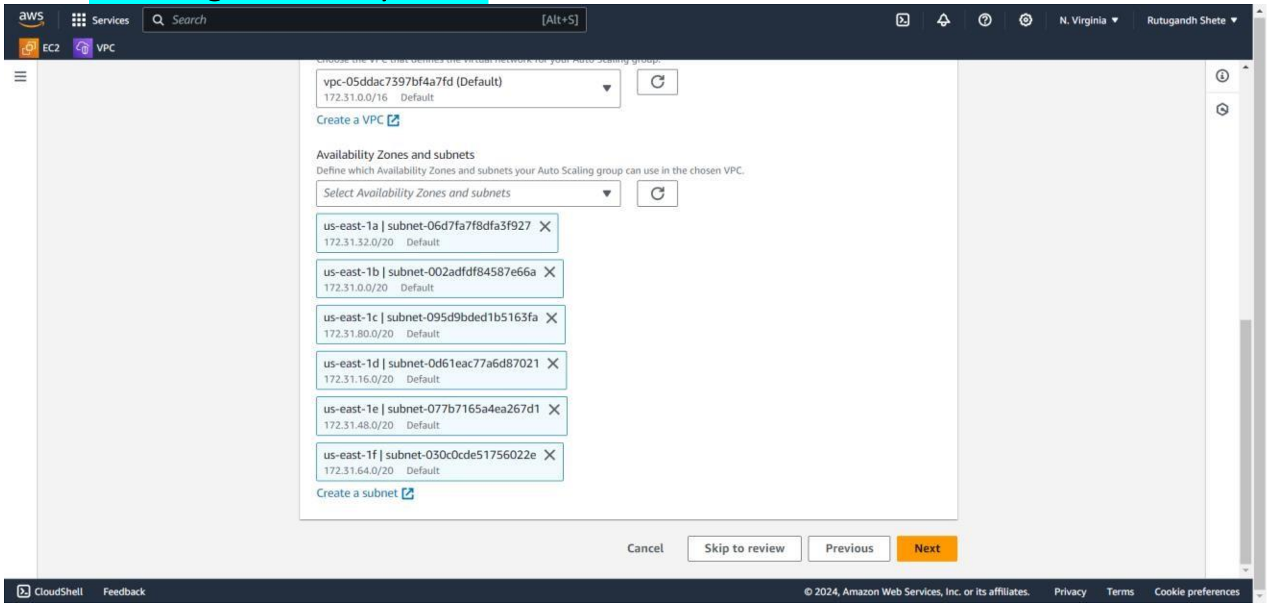
Version

1

Create a launch template version

Description	Launch template	Instance type
template	New-template lt-0cbb6ad46adfeb125	t2.micro

- After selecting availability zones→next



vpc-05ddac7397bf4a7fd (Default)
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

us-east-1a | subnet-06d7fa7f8dfa3f927
172.31.32.0/20 Default

us-east-1b | subnet-002adfdf84587e66a
172.31.0.0/20 Default

us-east-1c | subnet-095d9bded1b5163fa
172.31.80.0/20 Default

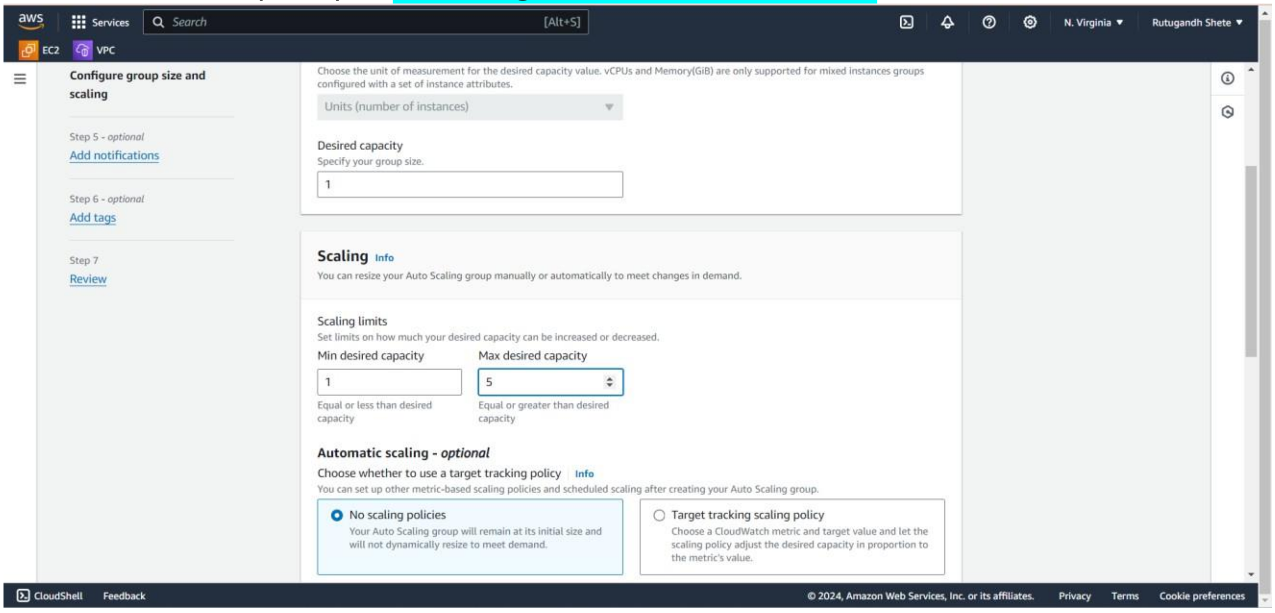
us-east-1d | subnet-0d61eac77a6d87021
172.31.16.0/20 Default

us-east-1e | subnet-077b7165a4ea267d1
172.31.48.0/20 Default

us-east-1f | subnet-030c0cde51756022e
172.31.64.0/20 Default

Create a subnet

- Select desired capacity →in scaling set min and max limits


Configure group size and scaling
Step 5 - *optional* Add notifications
Step 6 - *optional* Add tags
Step 7 Review

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.
1

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

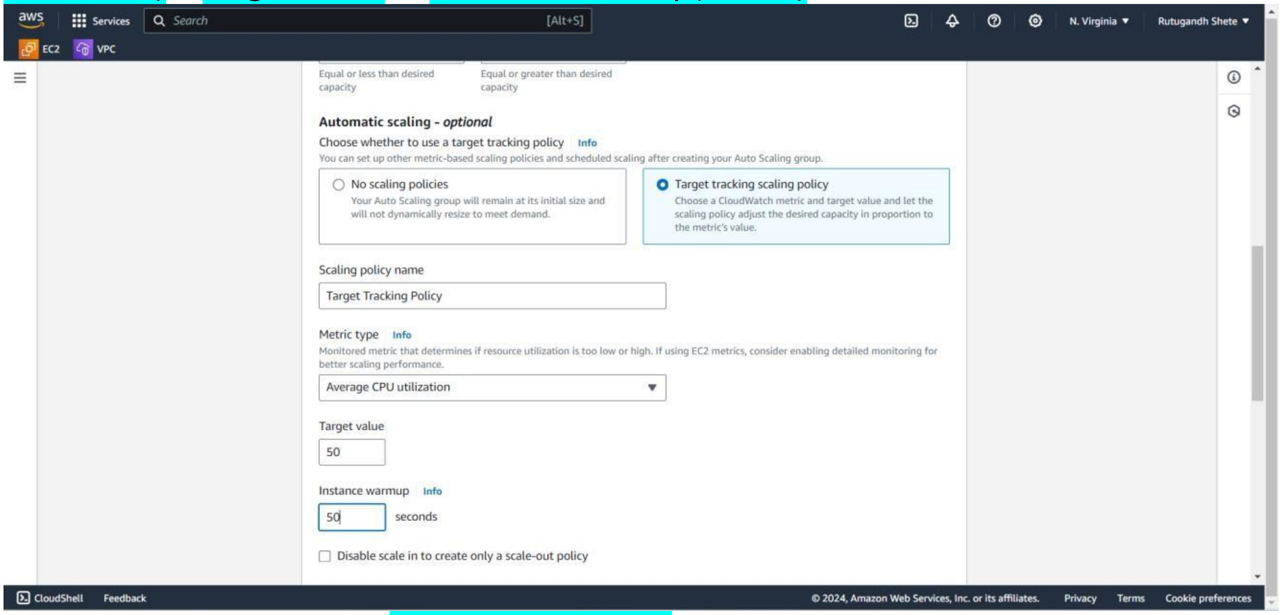
Scaling limits
Set limits on how much your desired capacity can be increased or decreased.

Min desired capacity	Max desired capacity
1	5
Equal or less than desired capacity	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.

- Click on target tracking scaling policy-→add name→add metric type(CPU utilization)→target value →instance warmup(ex- 50)


Equal or less than desired capacity
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.

Scaling policy name
Target Tracking Policy

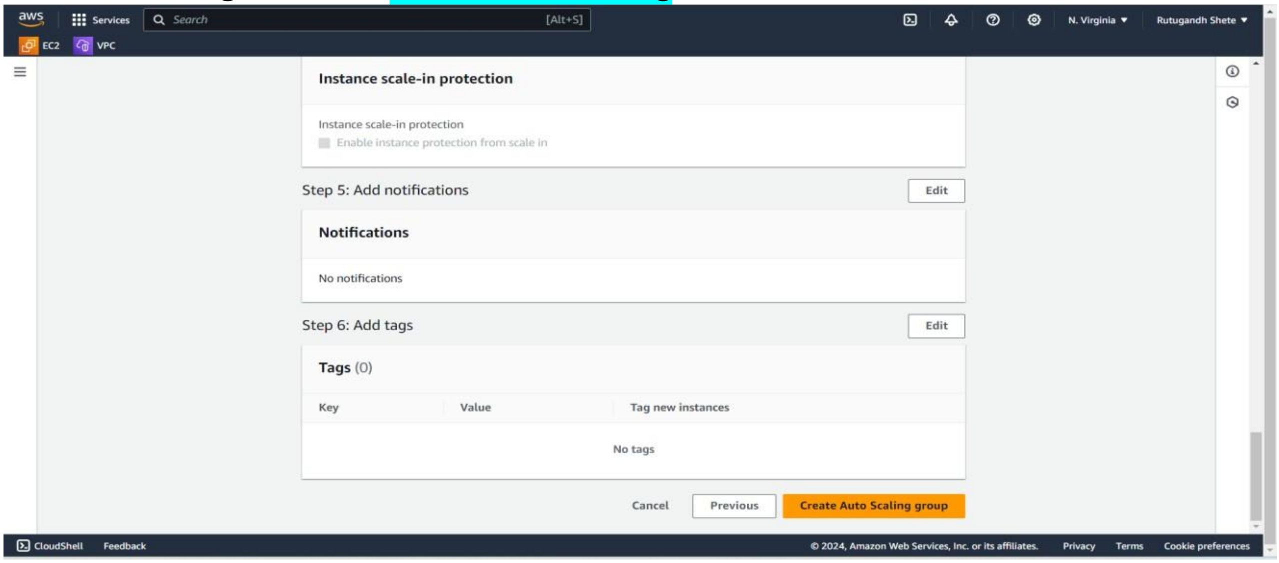
Metric type Info
Monitored metric that determines if resource utilization is too low or high. If using EC2 metrics, consider enabling detailed monitoring for better scaling performance.
Average CPU utilization

Target value
50

Instance warmup Info
50 seconds

Disable scale in to create only a scale-out policy

- After reviewing details →create auto scaling

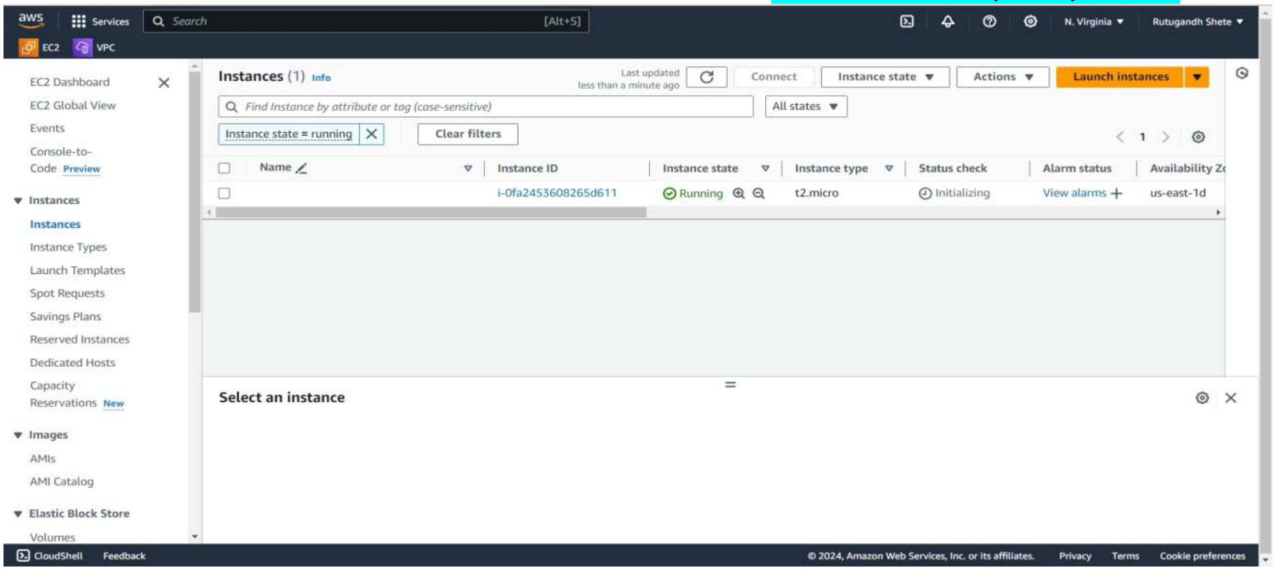

Instance scale-in protection
Instance scale-in protection
Enable instance protection from scale in

Step 5: Add notifications Edit
Notifications
No notifications

Step 6: Add tags Edit
Tags (0)

Key	Value	Tag new instances
No tags		

- As we can see one instance is created because our min desired capacity is “1”.

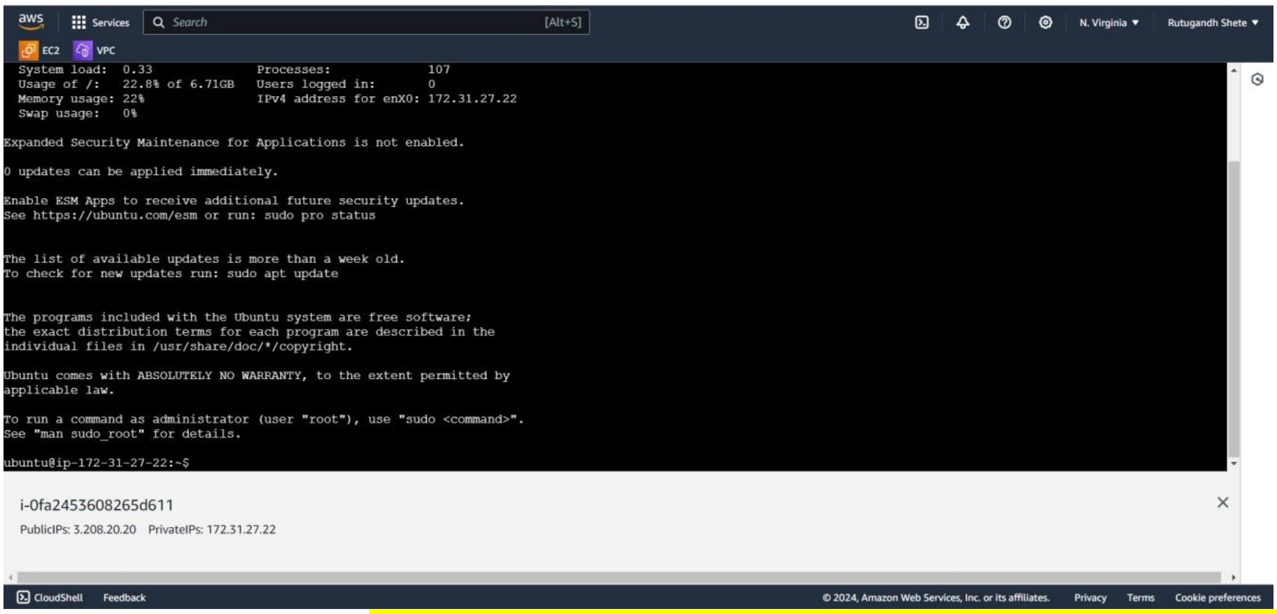


Instances (1) Info

Instance state = running

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone
	i-0fa2453608265d611	Running	t2.micro	Initializing	View alarms	us-east-1d

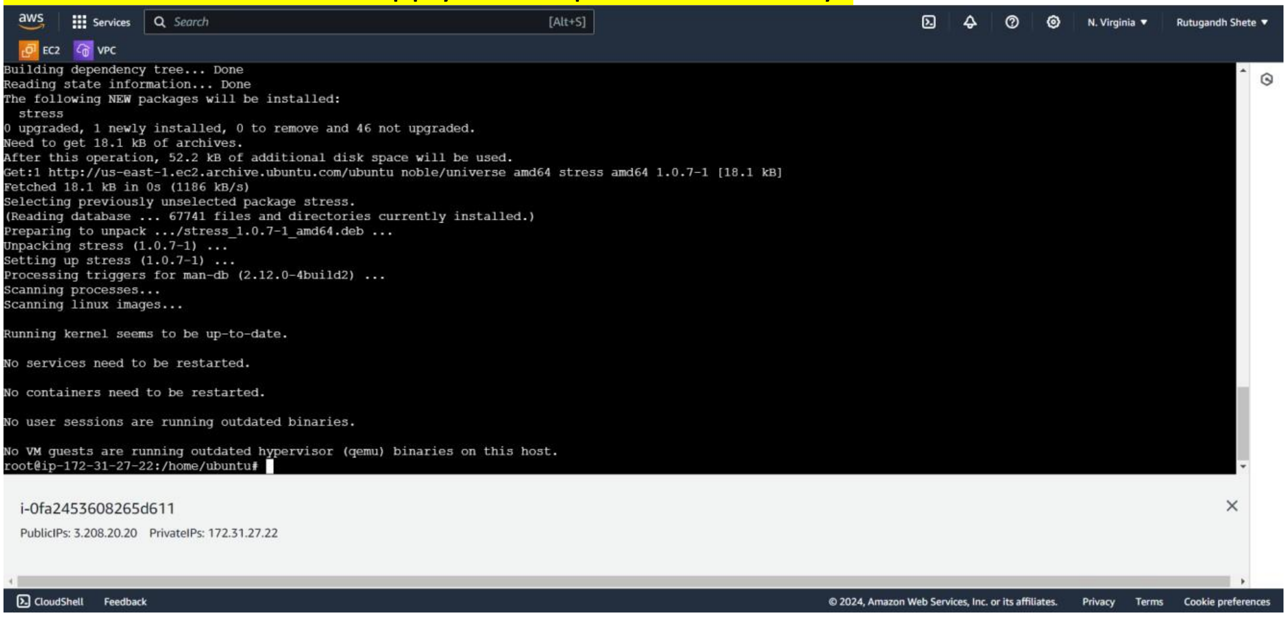
- Connect that to terminal



```
 System load:  0.33              Processes:             107
 Usage of /:   22.8% of 6.71GB   Users logged in:       0
 Memory usage: 22%               IPv4 address for enX0: 172.31.27.22
 Swap usage:   0%

Expanded Security Maintenance for Applications is not enabled.

0 updates can be applied immediately.

Enable ESM Apps to receive additional future security updates.
See https://ubuntu.com/esm or run: sudo pro status


The list of available updates is more than a week old.
To check for new updates run: sudo apt update


The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

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

ubuntu@ip-172-31-27-22:~$
```

i-0fa2453608265d611
PublicIPs: 3.208.20.20 PrivateIPs: 172.31.27.22

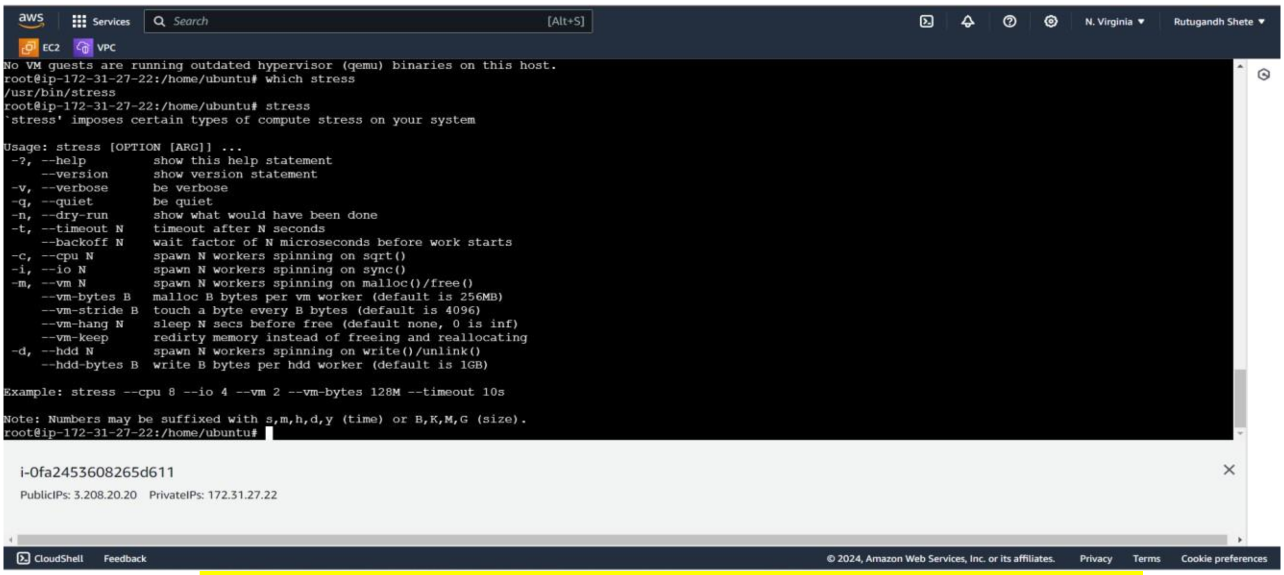
- Install stress command. The stress command has several options to specify the type and amount of stress to apply. Use “apt install stress -y”



```
Building dependency tree... Done
Reading state information... Done
The following NEW packages will be installed:
  stress
0 upgraded, 1 newly installed, 0 to remove and 46 not upgraded.
Need to get 18.1 kB of archives.
After this operation, 52.2 kB of additional disk space will be used.
Get:1 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble/universe amd64 stress amd64 1.0.7-1 [18.1 kB]
Fetched 18.1 kB in 0s (1186 kB/s)
Selecting previously unselected package stress.
(Reading database ... 67741 files and directories currently installed.)
Preparing to unpack .../stress_1.0.7-1_amd64.deb ...
Unpacking stress (1.0.7-1) ...
Setting up stress (1.0.7-1) ...
Processing triggers for man-db (2.12.0-4build2) ...
Scanning processes...
Scanning linux images...

Running kernel seems to be up-to-date.

No services need to be restarted.

No containers need to be restarted.

No user sessions are running outdated binaries.

No VM guests are running outdated hypervisor (qemu) binaries on this host.
root@ip-172-31-27-22:/home/ubuntu#
```

i-0fa2453608265d611
PublicIPs: 3.208.20.20 PrivateIPs: 172.31.27.22

- Use stress command to see details.



```
No VM guests are running outdated hypervisor (qemu) binaries on this host.
root@ip-172-31-27-22:/home/ubuntu# which stress
/usr/bin/stress
root@ip-172-31-27-22:/home/ubuntu# stress
`stress' imposes certain types of compute stress on your system

Usage: stress [OPTION [ARG]] ...
 -?, --help         show this help statement
     --version      show version statement
 -v, --verbose      be verbose
 -q, --quiet        be quiet
 -n, --dry-run      show what would have been done
 -t, --timeout N    timeout after N seconds
     --backoff N    wait factor of N microseconds before work starts
 -c, --cpu N        spawn N workers spinning on sqrt()
 -i, --io N         spawn N workers spinning on sync()
 -m, --vm N         spawn N workers spinning on malloc()/free()
     --vm-bytes B   malloc B bytes per vm worker (default is 256MB)
     --vm-stride B  touch a byte every B bytes (default is 4096)
     --vm-hang N    sleep N secs before free (default none, 0 is inf)
     --vm-keep      redirty memory instead of freeing and reallocating
 -d, --hdd N        spawn N workers spinning on write()/unlink()
     --hdd-bytes B  write B bytes per hdd worker (default is 1GB)

Example: stress --cpu 8 --io 4 --vm 2 --vm-bytes 128M --timeout 10s

Note: Numbers may be suffixed with s,m,h,d,y (time) or B,K,M,G (size).
root@ip-172-31-27-22:/home/ubuntu# 
```

i-0fa2453608265d611

PublicIPs: 3.208.20.20 PrivateIPs: 172.31.27.22

- After adding " stress --cpu 90 --io 4 --vm 2 --vm-bytes 128M --timeout 10m"

--cpu 90: This will start 90 CPU stressor processes. Each process will try to fully utilize one CPU core. This is quite intensive, especially if you have fewer than 90 cores available, as it will overburden the CPU.

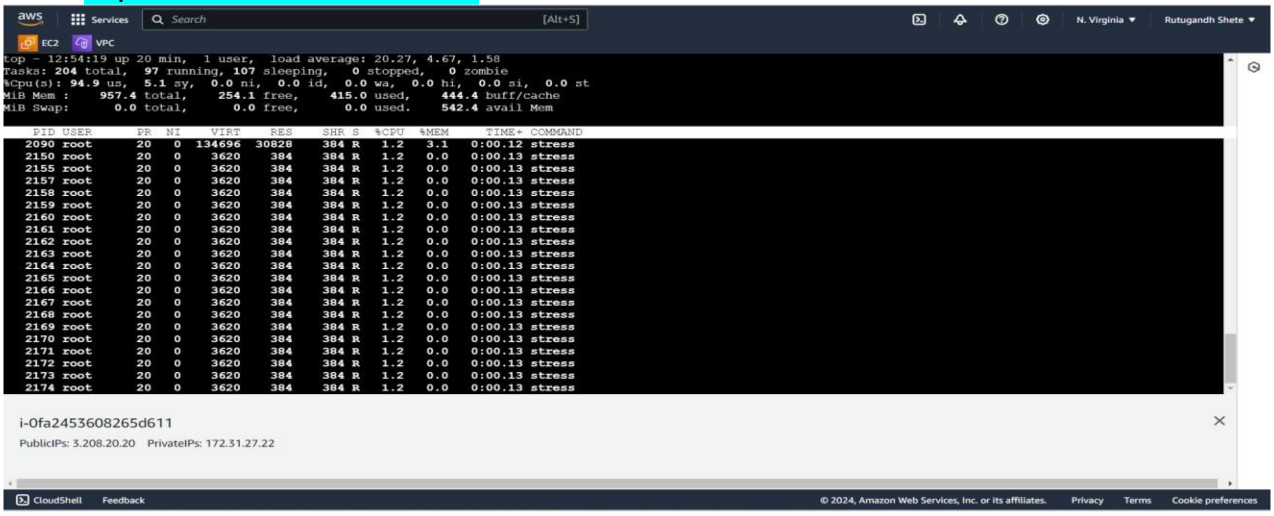
io 4: This starts 4 I/O stressor processes. These processes will perform continuous I/O operations (reading and writing data), stressing the system's input/output subsystem.

vm 2: This starts 2 virtual memory stressor processes. These processes will continuously allocate, use, and free memory.

vm-bytes 128M: This sets the amount of memory allocated by each --vm process to 128 MB. Each of the 2 --vm processes will allocate and work with 128 MB of memory.

timeout 10m: This sets the duration of the stress test to 10 minutes. After 10 minutes, the stress command will automatically stop.

Use "top" to see CPU utilization

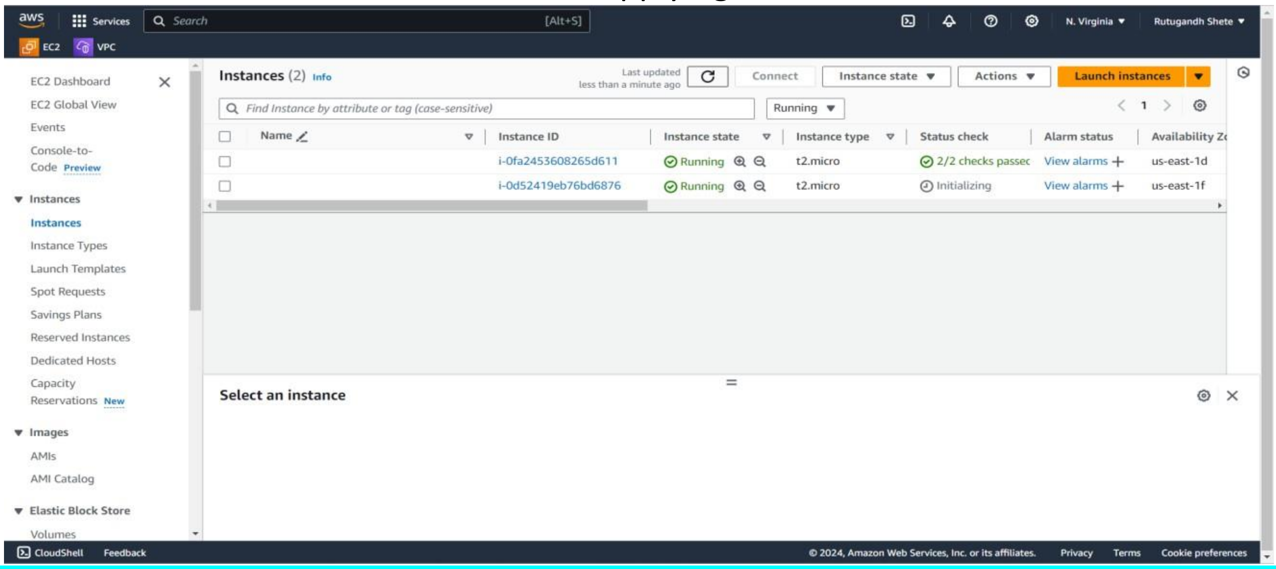


```
top - 12:54:19 up 20 min,  1 user,  load average: 20.27, 4.67, 1.58
Tasks: 204 total,  97 running, 107 sleeping,   0 stopped,   0 zombie
%Cpu(s): 94.9 us,  5.1 sy,  0.0 ni,  0.0 id,  0.0 wa,  0.0 hi,  0.0 si,  0.0 st
MiB Mem :    957.4 total,    254.1 free,    415.0 used,    444.4 buff/cache
MiB Swap:      0.0 total,      0.0 free,      0.0 used.    542.4 avail Mem

    PID USER      PR  NI    VIRT    RES    SHR S  %CPU  %MEM     TIME+ COMMAND
   2090 root      20   0  134696  30828    384 R   1.2   3.1   0:00.12 stress
   2150 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2155 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2157 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2158 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2159 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2160 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2161 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2162 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2163 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2164 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2165 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2166 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2167 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2168 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2169 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2170 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2171 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2172 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2173 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
   2174 root      20   0    3620    384    384 R   1.2   0.0   0:00.13 stress
```

i-0fa2453608265d611

PublicIPs: 3.208.20.20 PrivateIPs: 172.31.27.22

- As we can one instance is created after applying stress.



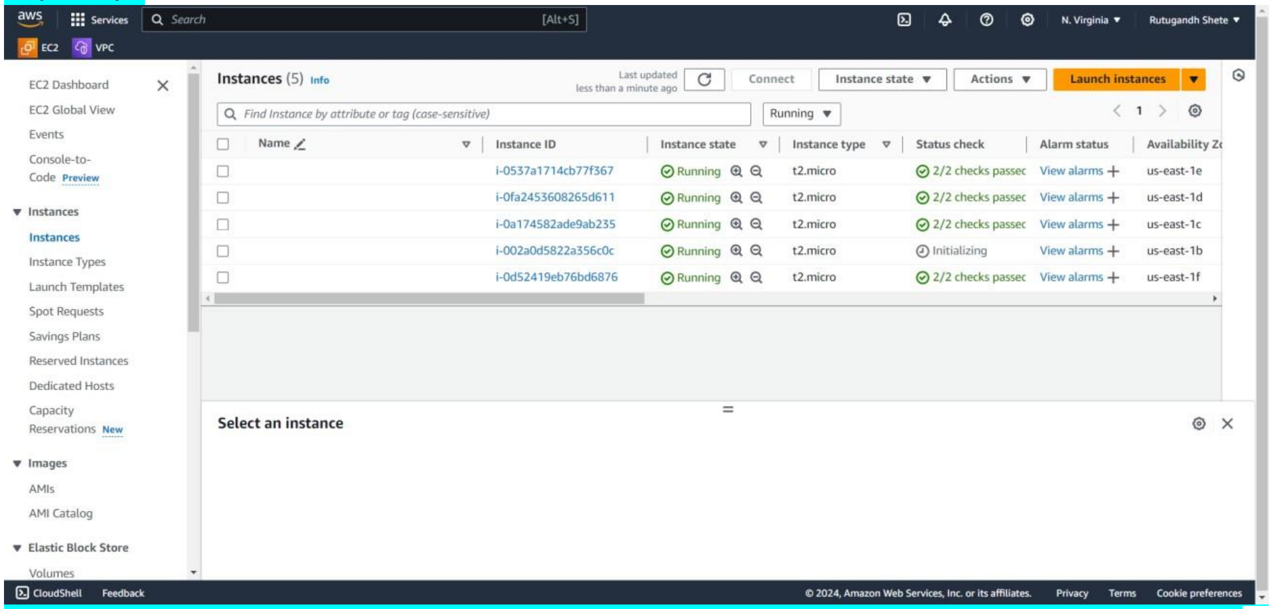
Instances (2) Info

	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone
☐		i-0fa2453608265d611	Running	t2.micro	2/2 checks passed	View alarms	us-east-1d
☐		i-0d52419eb76bd6876	Running	t2.micro	Initializing	View alarms	us-east-1f

Select an instance

© 2024, Amazon Web Services, Inc. or its affiliates.

- As we can see applying stress to 90% cpu utilization we have achieved "max desired capacity"



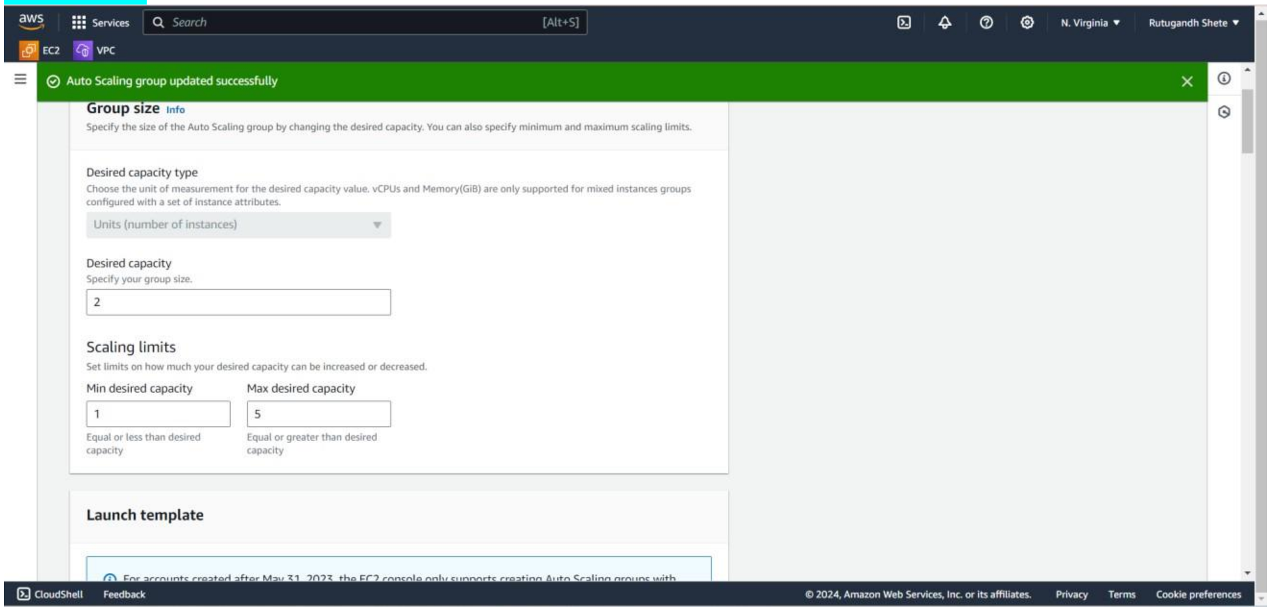
Instances (5) Info

	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone
☐		i-0537a1714cb77f367	Running	t2.micro	2/2 checks passed	View alarms	us-east-1e
☐		i-0fa2453608265d611	Running	t2.micro	2/2 checks passed	View alarms	us-east-1d
☐		i-0a174582ade9ab235	Running	t2.micro	2/2 checks passed	View alarms	us-east-1c
☐		i-002a0d5822a356c0c	Running	t2.micro	Initializing	View alarms	us-east-1b
☐		i-0d52419eb76bd6876	Running	t2.micro	2/2 checks passed	View alarms	us-east-1f

Select an instance

© 2024, Amazon Web Services, Inc. or its affiliates.

- After editing desired capacity we can scale down and number of instances will be reduced .



Auto Scaling group updated successfully

Group size Info

Specify the size of the Auto Scaling group by changing the desired capacity. You can also specify minimum and maximum scaling limits.

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

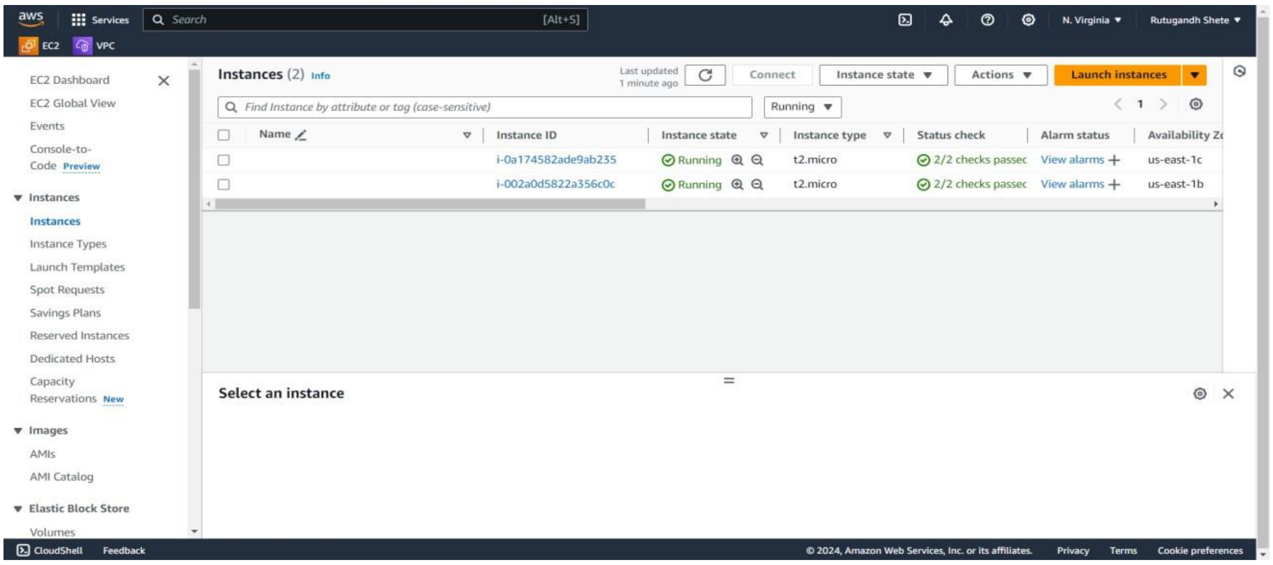
Scaling limits

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

Min desired capacity	Max desired capacity
1	5
Equal or less than desired capacity	Equal or greater than desired capacity

Launch template

© 2024, Amazon Web Services, Inc. or its affiliates.

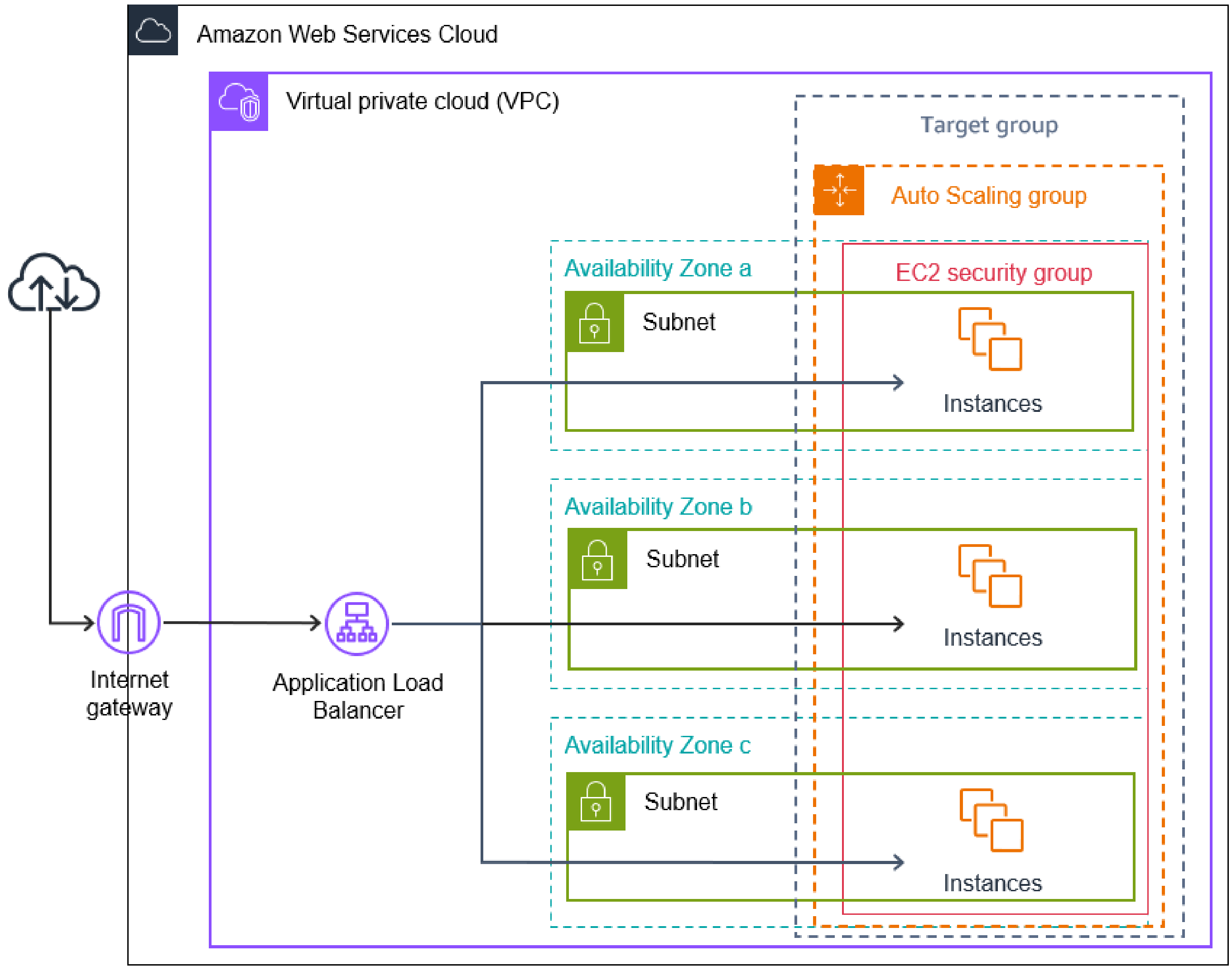


Instances (2) Info

	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zc
☐		i-0a174582ade9ab235	Running	t2.micro	2/2 checks passec	View alarms +	us-east-1c
☐		i-002a0d5822a356c0c	Running	t2.micro	2/2 checks passec	View alarms +	us-east-1b

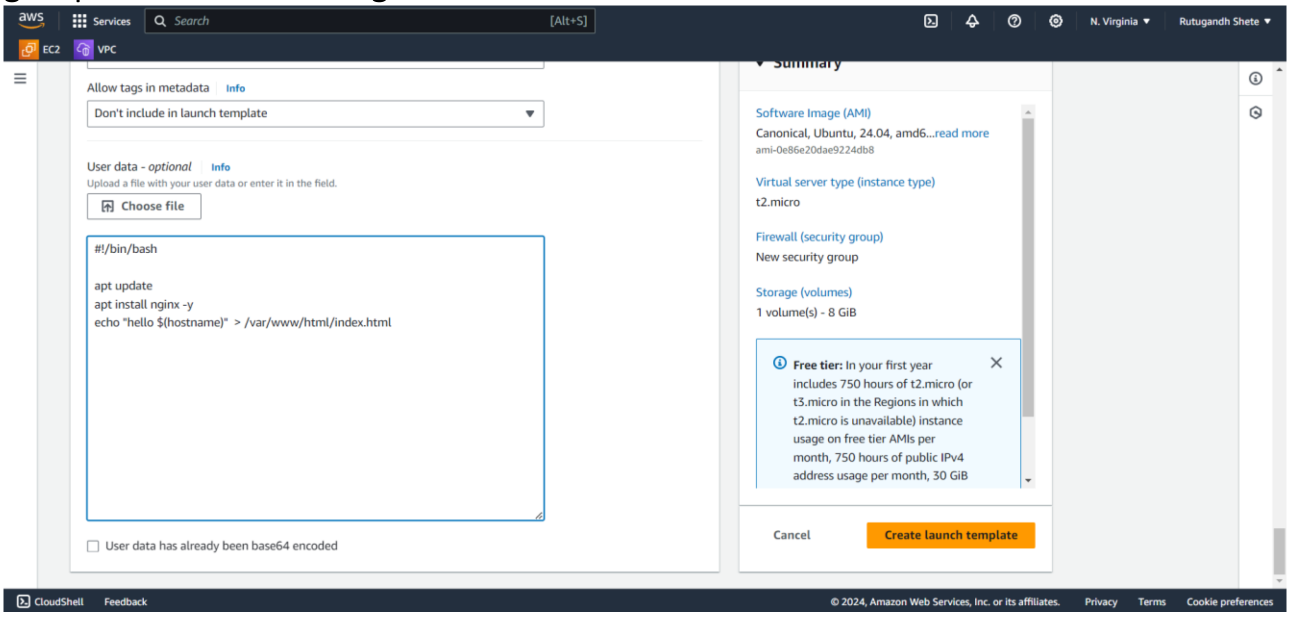
Select an instance

Auto scaling using load balancer:



Steps:

- Create template→Name→AMI→Key pair→instance type→create security group→advanced settings



Allow tags in metadata Info

Don't include in launch template

User data - *optional* Info

Upload a file with your user data or enter it in the field.

```
#!/bin/bash

apt update
apt install nginx -y
echo "hello $(hostname)" > /var/www/html/index.html
```

User data has already been base64 encoded

Summary

Software Image (AMI)
Canonical, Ubuntu, 24.04, amd6...read more
ami-0e86e20dae9224db8

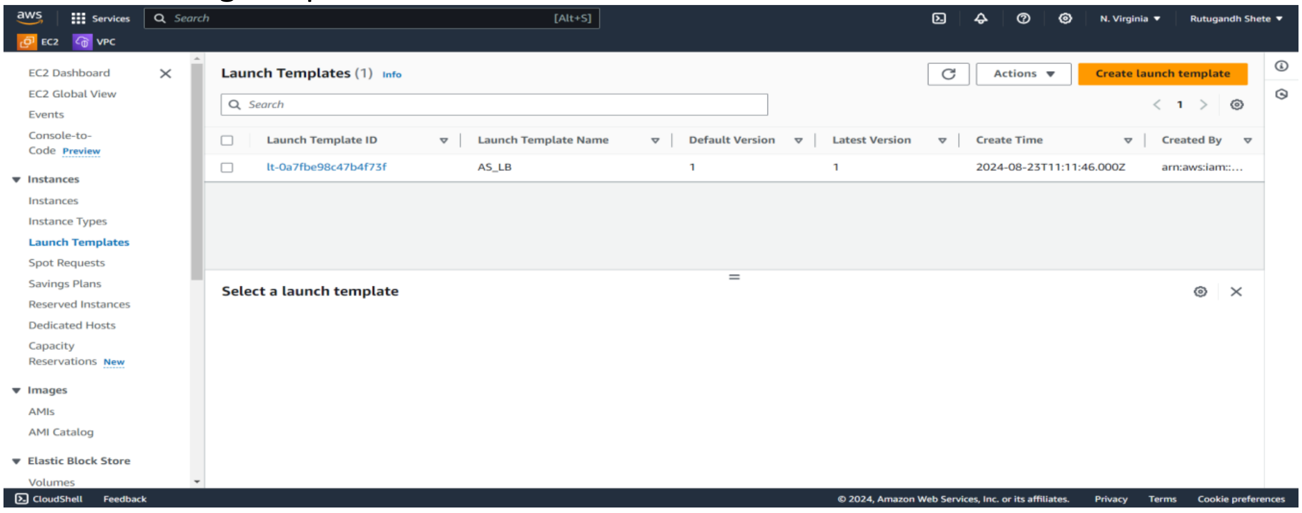
Virtual server type (instance type)
t2.micro

Firewall (security group)
New security group

Storage (volumes)
1 volume(s) - 8 GiB

Free tier: In your first year includes 750 hours of t2.micro (or t3.micro in the Regions in which t2.micro is unavailable) instance usage on free tier AMIs per month, 750 hours of public IPv4 address usage per month, 30 GiB

- After launching template

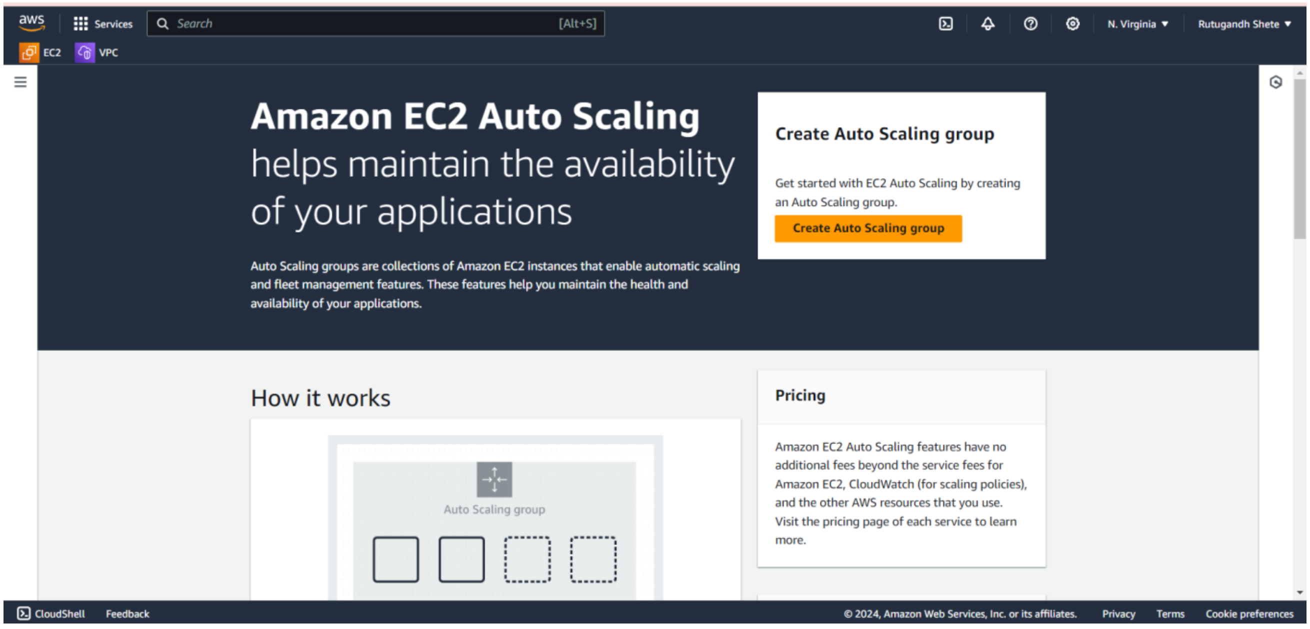


Launch Templates (1) Info

	Launch Template ID	Launch Template Name	Default Version	Latest Version	Create Time	Created By
	lt-0a7fbe98c47b4f73f	AS_LB	1	1	2024-08-23T11:11:46.000Z	arn:aws:iam:...

Select a launch template

- Create auto scaling →create



Amazon EC2 Auto Scaling helps maintain the availability of your applications

Auto Scaling groups are collections of Amazon EC2 instances that enable automatic scaling and fleet management features. These features help you maintain the health and availability of your applications.

Create Auto Scaling group

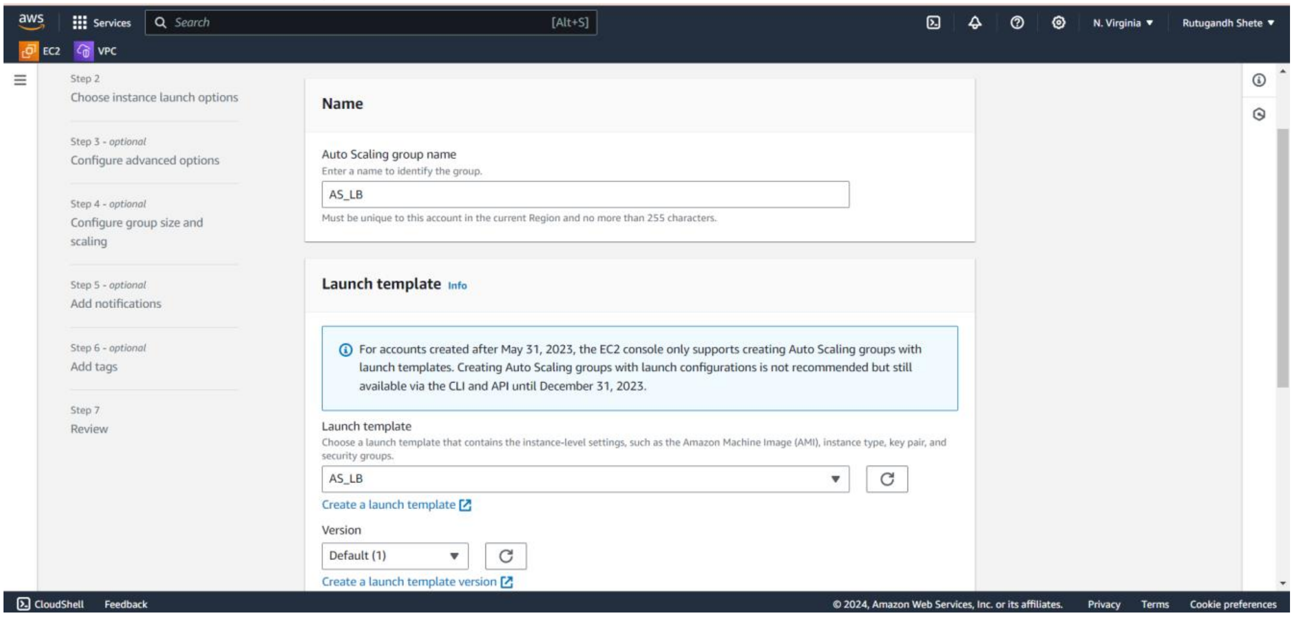
Get started with EC2 Auto Scaling by creating an Auto Scaling group.

How it works

Pricing

Amazon EC2 Auto Scaling features have no additional fees beyond the service fees for Amazon EC2, CloudWatch (for scaling policies), and the other AWS resources that you use. Visit the pricing page of each service to learn more.

- Provide name →template that we have created →select version



Name

Auto Scaling group name
Enter a name to identify the group.
AS_LB
Must be unique to this account in the current Region and no more than 255 characters.

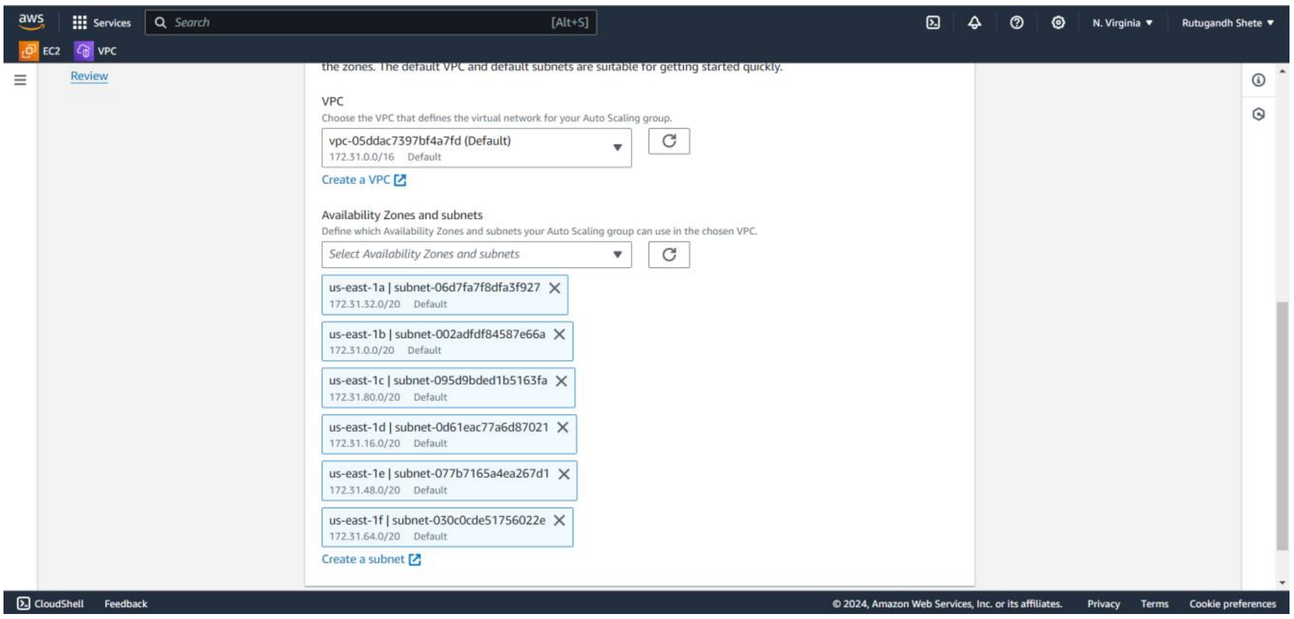
Launch template Info

For accounts created after May 31, 2023, the EC2 console only supports creating Auto Scaling groups with launch templates. Creating Auto Scaling groups with launch configurations is not recommended but still available via the CLI and API until December 31, 2023.

Launch template
Choose a launch template that contains the instance-level settings, such as the Amazon Machine Image (AMI), instance type, key pair, and security groups.
AS_LB
Create a launch template

Version
Default (1)
Create a launch template version

- Select VPC and zones



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-05ddac7397bf4a7fd (Default)
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

us-east-1a | subnet-06d7fa7f8dfa3f927
172.31.32.0/20 Default

us-east-1b | subnet-002adfdf84587e66a
172.31.0.0/20 Default

us-east-1c | subnet-095d9bded1b5163fa
172.31.80.0/20 Default

us-east-1d | subnet-0d61eac77a6d87021
172.31.16.0/20 Default

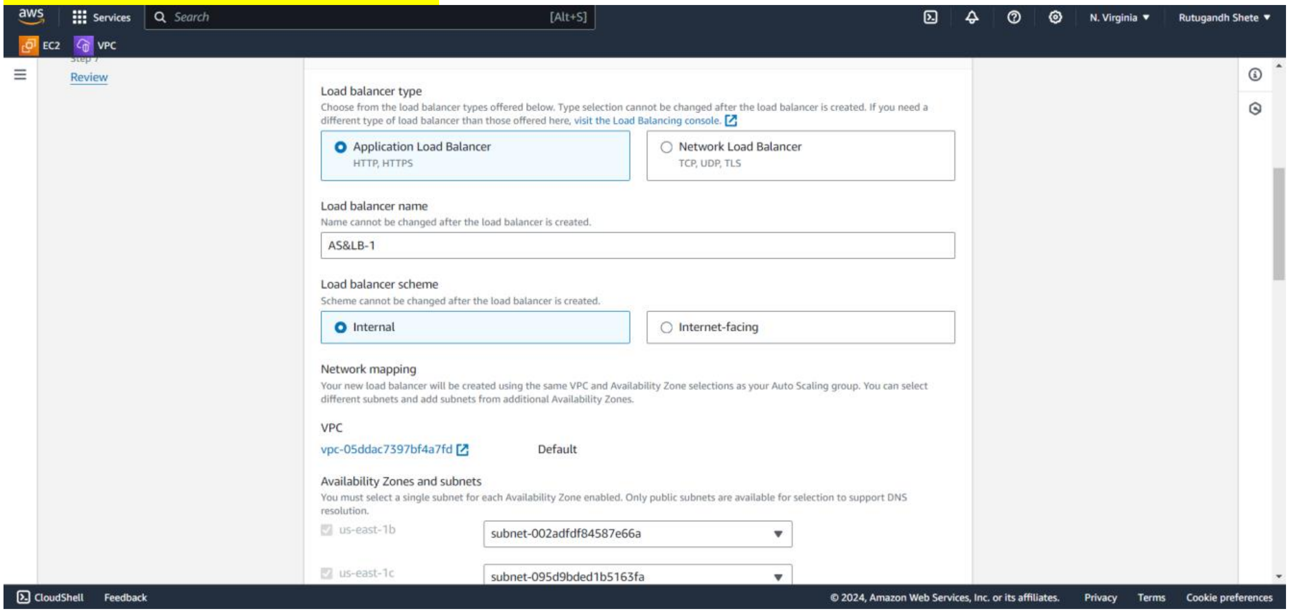
us-east-1e | subnet-077b7165a4ea267d1
172.31.48.0/20 Default

us-east-1f | subnet-030c0cde51756022e
172.31.64.0/20 Default

Create a subnet

- Load balancer→select that we have to create new load balancer (if have load balancer then we can choose)→select load balancer type→select Health check grace period (30sec).
  An internal load balancer is only accessible within your VPC. It doesn't have a publicly resolvable DNS name, which means it's not accessible from the internet.

  An internet-facing load balancer has a publicly resolvable DNS name, making it accessible from the internet.



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.

AS&LB-1

Load balancer scheme

Scheme cannot be changed after the load balancer is created.

[✓ Internal] Internet-facing

Network mapping

Your new load balancer will be created using the same VPC and Availability Zone selections as your Auto Scaling group. You can select different subnets and add subnets from additional Availability Zones.

VPC

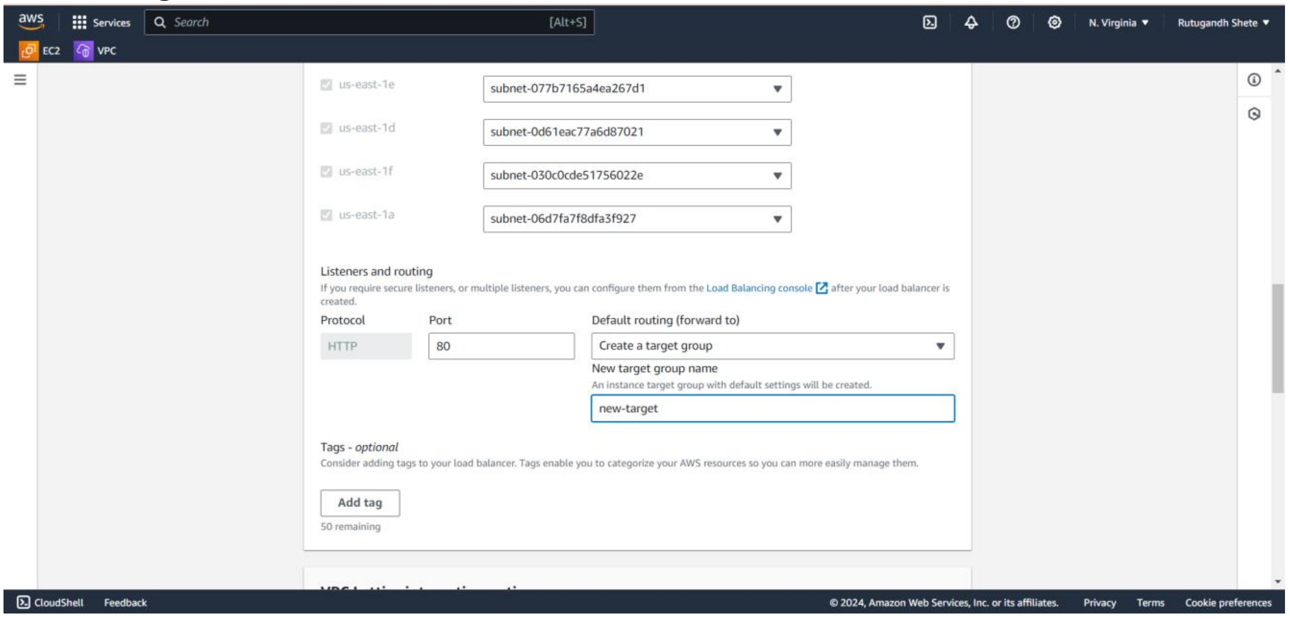
vpc-05ddac7397bf4a7fd Default

Availability Zones and subnets

You must select a single subnet for each Availability Zone enabled. Only public subnets are available for selection to support DNS resolution.

us-east-1b	subnet-002adfdf84587e66a
us-east-1c	subnet-095d9bded1b5163fa

- Create target →name



us-east-1e	subnet-077b7165a4ea267d1
us-east-1d	subnet-0d61eac77a6d87021
us-east-1f	subnet-030c0cde51756022e
us-east-1a	subnet-06d7fa7f8dfa3f927

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	80	Create a target group

New target group name

An instance target group with default settings will be created.

new-target

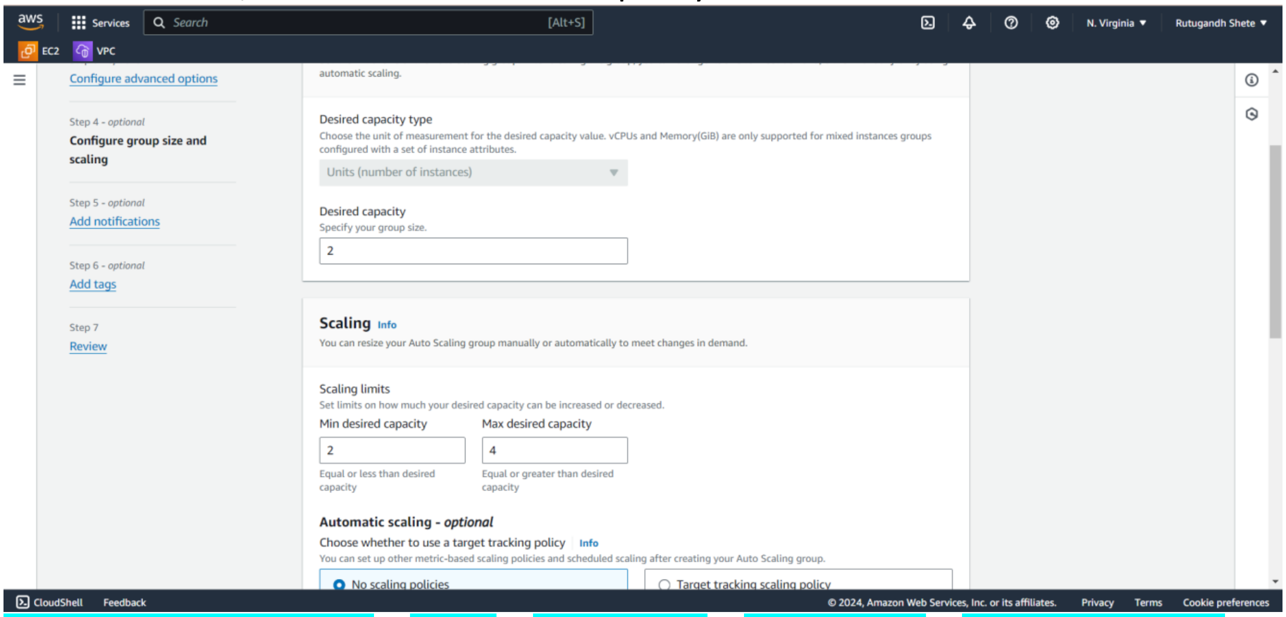
Tags - *optional*

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

- Provide desired, min and max desired capacity.



Configure advanced options

Step 4 - *optional*
Configure group size and scaling

Step 5 - *optional*
Add notifications

Step 6 - *optional*
Add tags

Step 7
Review

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

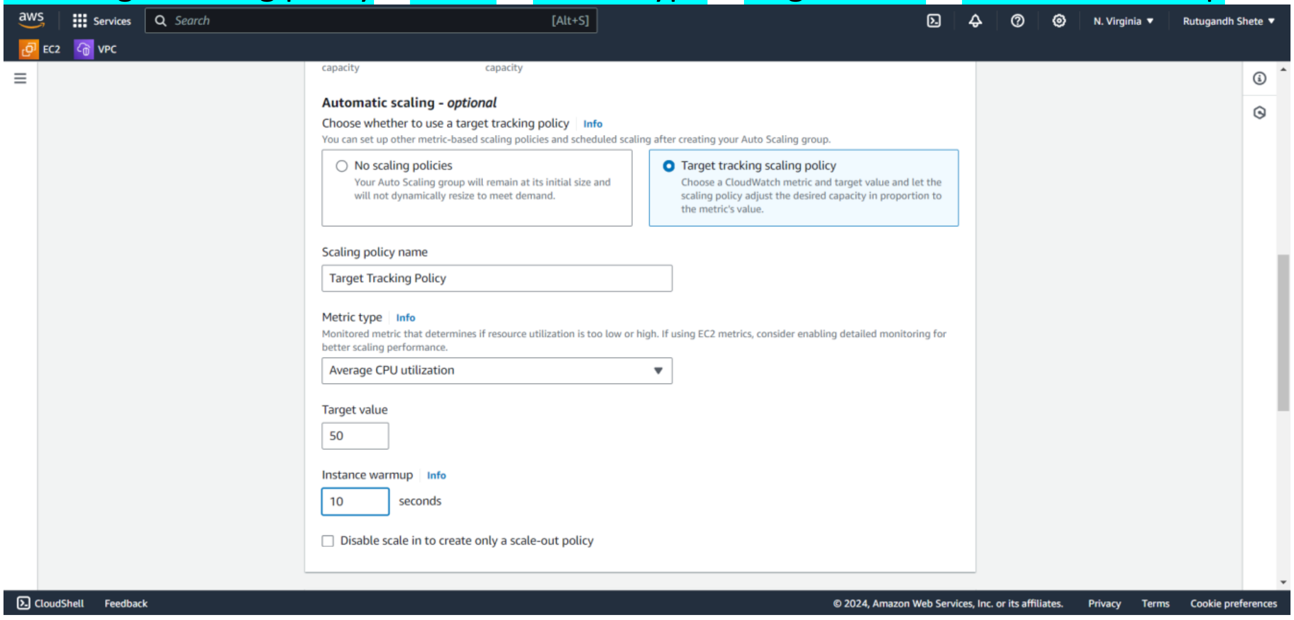
Scaling Info
You can resize your Auto Scaling group manually or automatically to meet changes in demand.

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

Min desired capacity	Max desired capacity
2	4
Equal or less than desired capacity	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] Target tracking scaling policy

- Add target scaling policy→name→metric type→target value→instance warmup.



capacity 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.

Scaling policy name
Target Tracking Policy

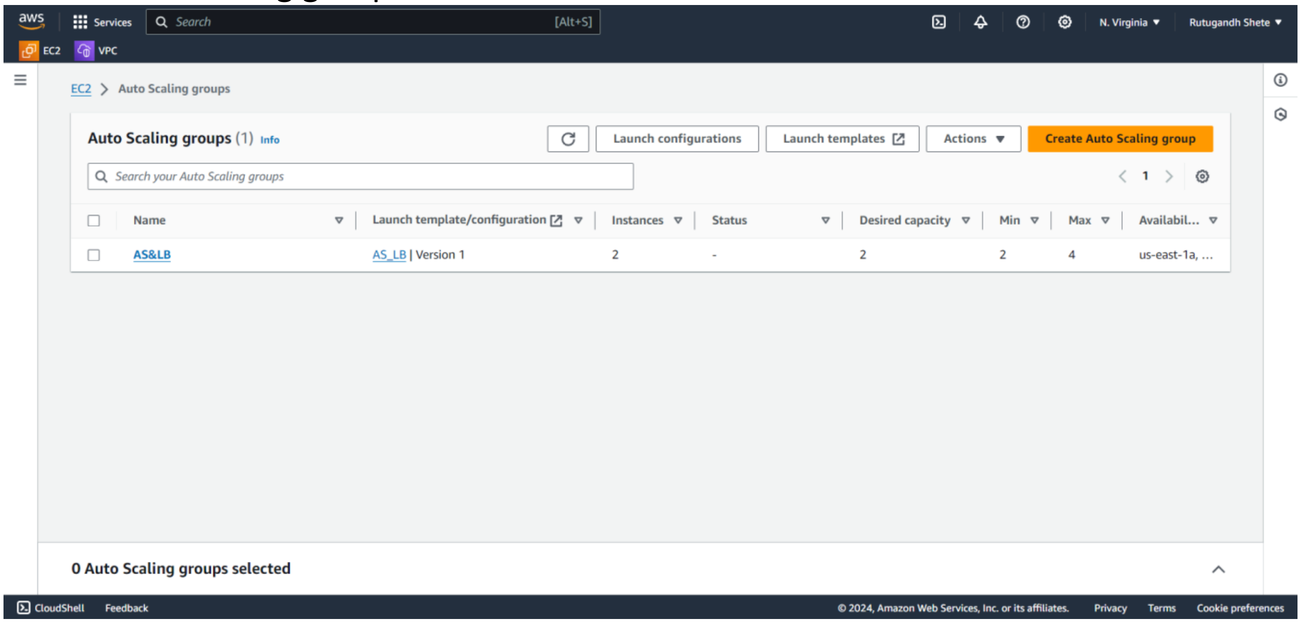
Metric type Info
Monitored metric that determines if resource utilization is too low or high. If using EC2 metrics, consider enabling detailed monitoring for better scaling performance.
Average CPU utilization

Target value
50

Instance warmup Info
10 seconds

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

- Create auto scaling group



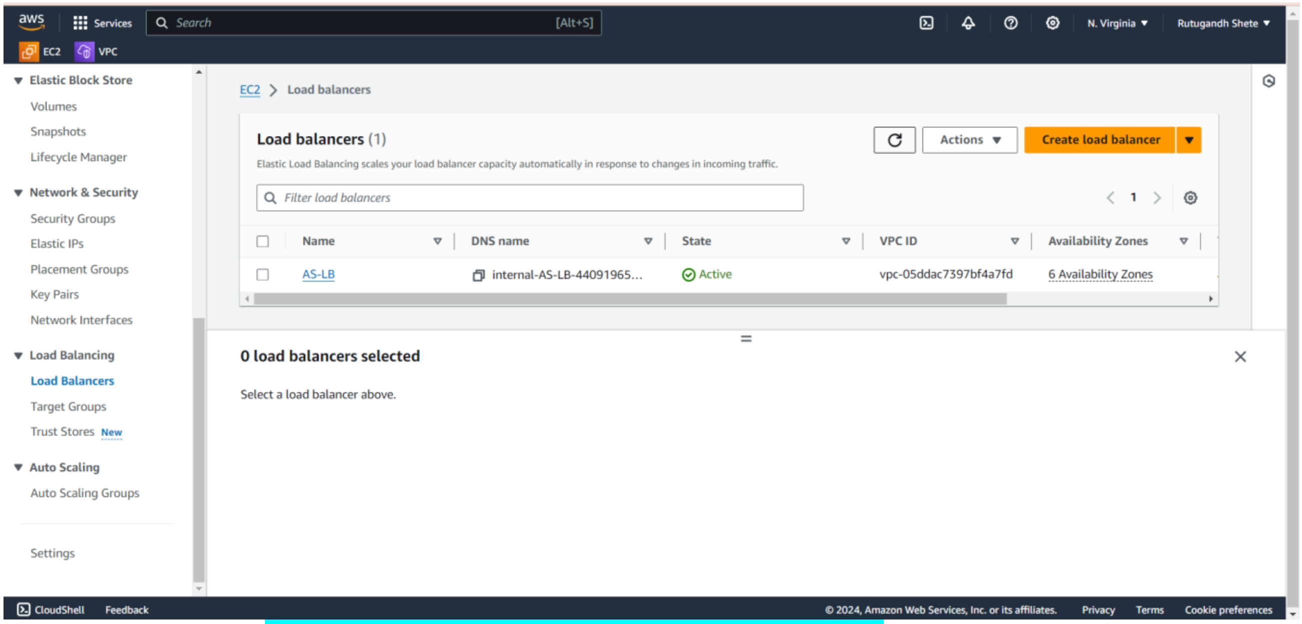
EC2 > Auto Scaling groups

Auto Scaling groups (1) Info

Name	Launch template/configuration	Instances	Status	Desired capacity	Min	Max	Availabil...
AS&LB	AS_LB | Version 1	2	-	2	2	4	us-east-1a, ...

0 Auto Scaling groups selected

- As we can see that our load balancer is in active state and paste DNS onto browser



EC2 > Load balancers

Load balancers (1)

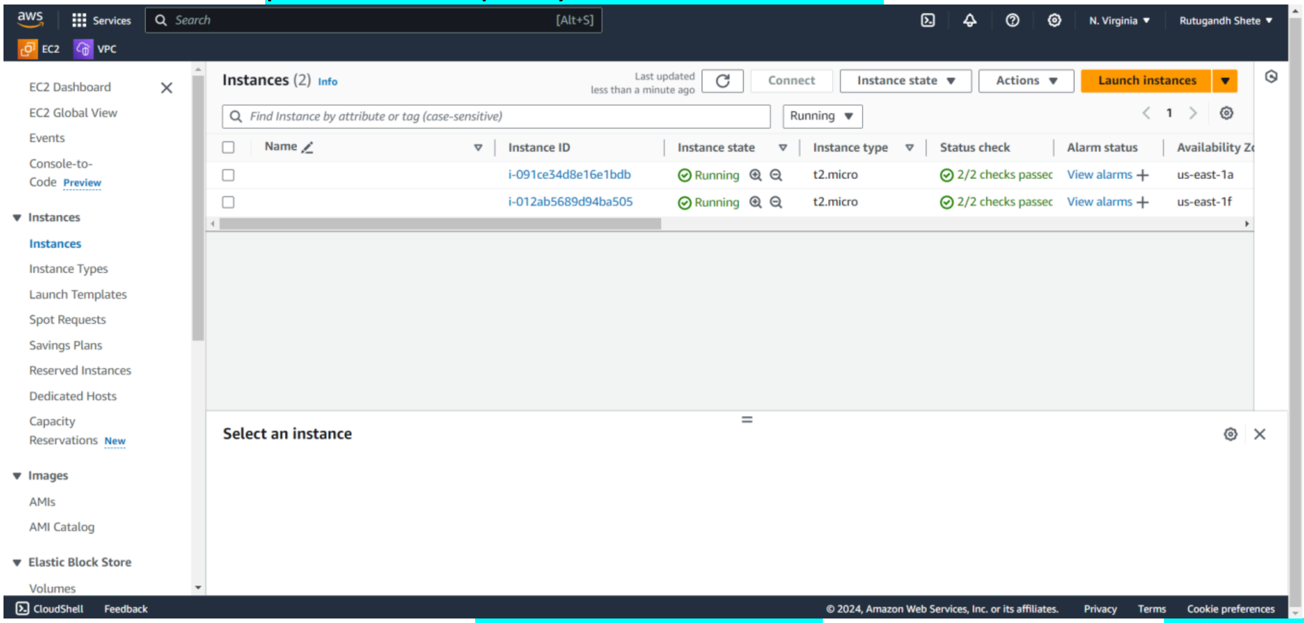
Elastic Load Balancing scales your load balancer capacity automatically in response to changes in incoming traffic.

Name	DNS name	State	VPC ID	Availability Zones
AS-LB	internal-AS-LB-44091965...	Active	vpc-05ddac7397bf4a7fd	6 Availability Zones

0 load balancers selected

Select a load balancer above.

- As we can see as per desired capacity we have 2 instances.



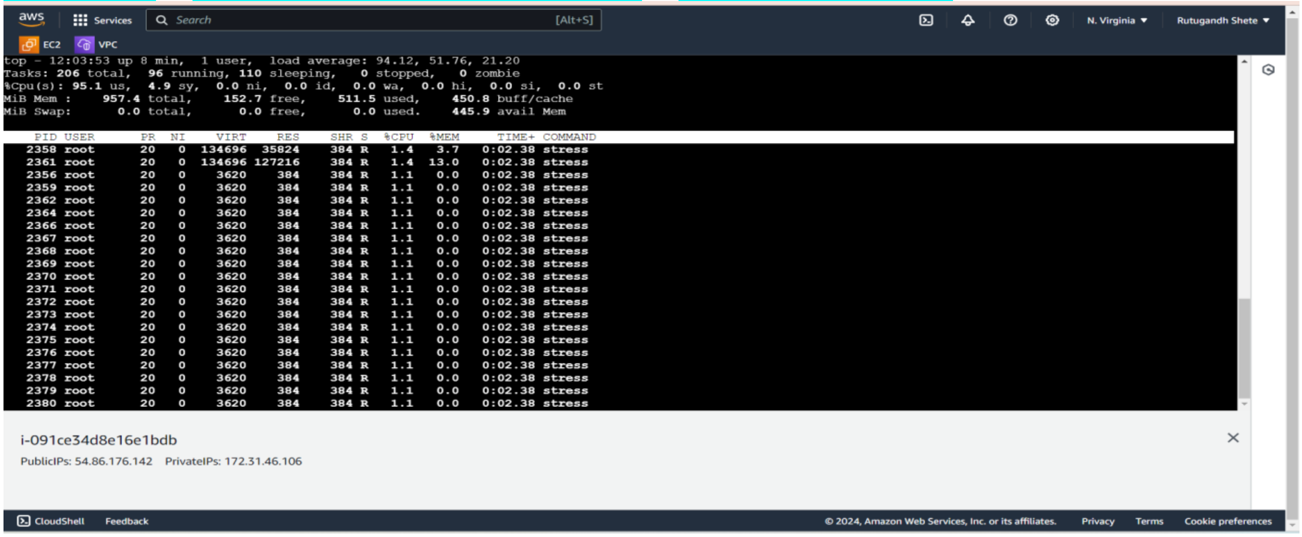
Instances (2) Info

Last updated less than a minute ago

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone
	i-091ce34d8e16e1bdb	Running	t2.micro	2/2 checks passed	View alarms +	us-east-1a
	i-012ab5689d94ba505	Running	t2.micro	2/2 checks passed	View alarms +	us-east-1f

Select an instance

- Connect instance to teminal →install stress command(apt install stress -y)→use stress command→after pasting stress command→use top command

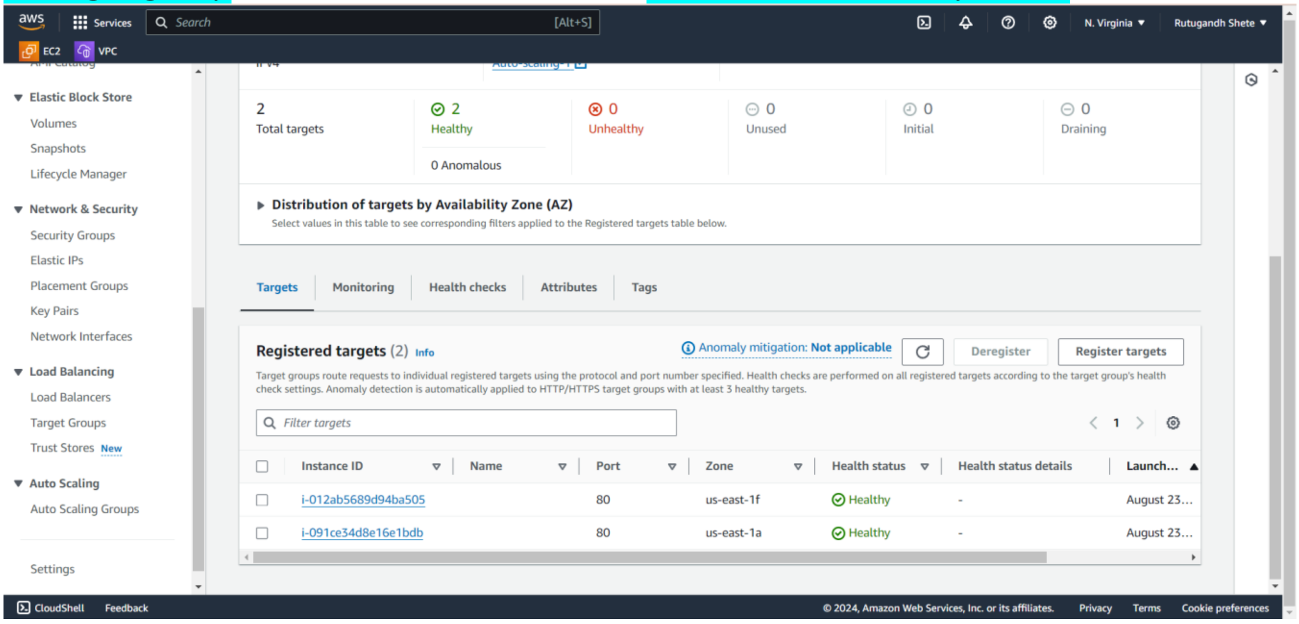


```
top - 12:03:53 up 8 min,  1 user,  load average: 94.12, 51.76, 21.20
Tasks: 206 total,  96 running, 110 sleeping,   0 stopped,   0 zombie
%Cpu(s): 95.1 us,  4.9 sy,  0.0 ni,  0.0 id,  0.0 wa,  0.0 hi,  0.0 si,  0.0 st
MiB Mem :    957.4 total,    152.7 free,    511.5 used,    450.8 buff/cache
MiB Swap:      0.0 total,      0.0 free,      0.0 used.    445.9 avail Mem

    PID USER      PR  NI    VIRT    RES    SHR S  %CPU  %MEM     TIME+ COMMAND
   2358 root      20   0  134696  35824    384 R   1.4   3.7   0:02.38 stress
   2361 root      20   0  134696 127216    384 R   1.4  13.0   0:02.38 stress
   2356 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
   2359 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
   2362 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
   2364 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
   2366 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
   2367 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
   2368 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
   2369 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
   2370 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
   2371 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
   2372 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
   2373 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
   2374 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
   2375 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
   2376 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
   2377 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
   2378 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
   2379 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
   2380 root      20   0    3620    384    384 R   1.1   0.0   0:02.38 stress
```

i-091ce34d8e16e1bdb

PublicIPs: 54.86.176.142 PrivateIPs: 172.31.46.106

- In target group we can check whether the instances are healthy or not.



2 Total targets	2 Healthy	0 Unhealthy	0 Unused	0 Initial	0 Draining
	0 Anomalous				

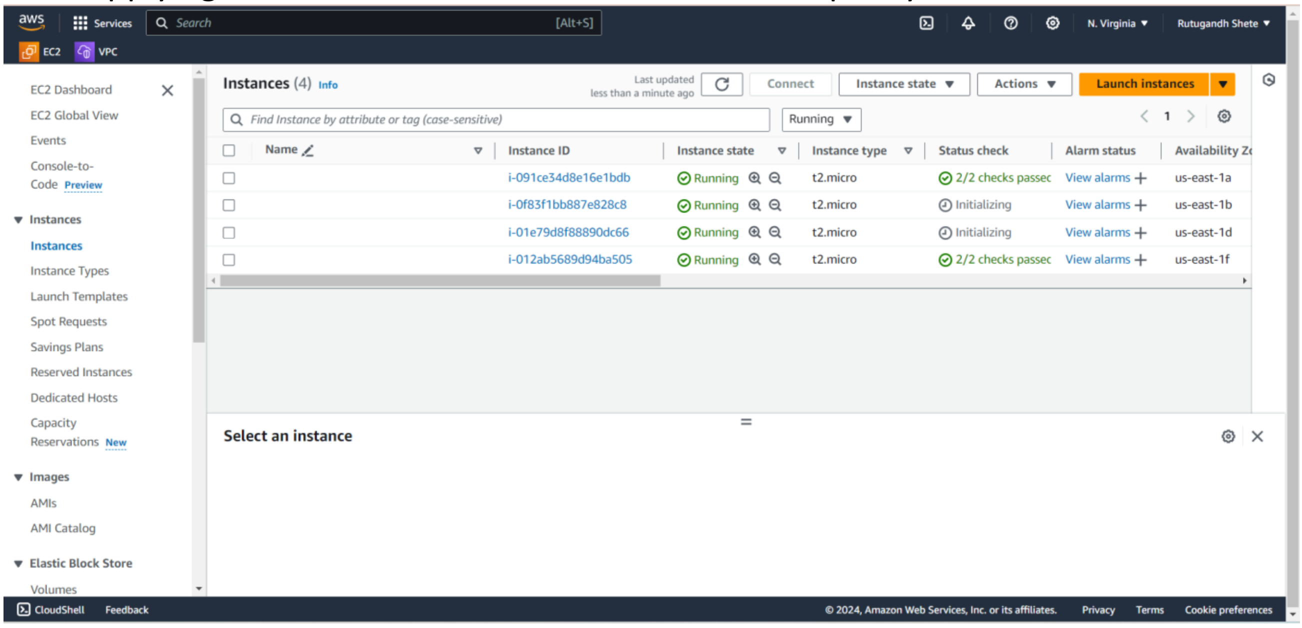
Distribution of targets by Availability Zone (AZ)
Select values in this table to see corresponding filters applied to the Registered targets table below.

Registered targets (2) Info

Target groups route requests to individual registered targets using the protocol and port number specified. Health checks are performed on all registered targets according to the target group's health check settings. Anomaly detection is automatically applied to HTTP/HTTPS target groups with at least 3 healthy targets.

Instance ID	Name	Port	Zone	Health status	Health status details	Launch...
i-012ab5689d94ba505		80	us-east-1f	Healthy	-	August 23...
i-091ce34d8e16e1bdb		80	us-east-1a	Healthy	-	August 23...

- After applying stress we have achieved max desired capacity.

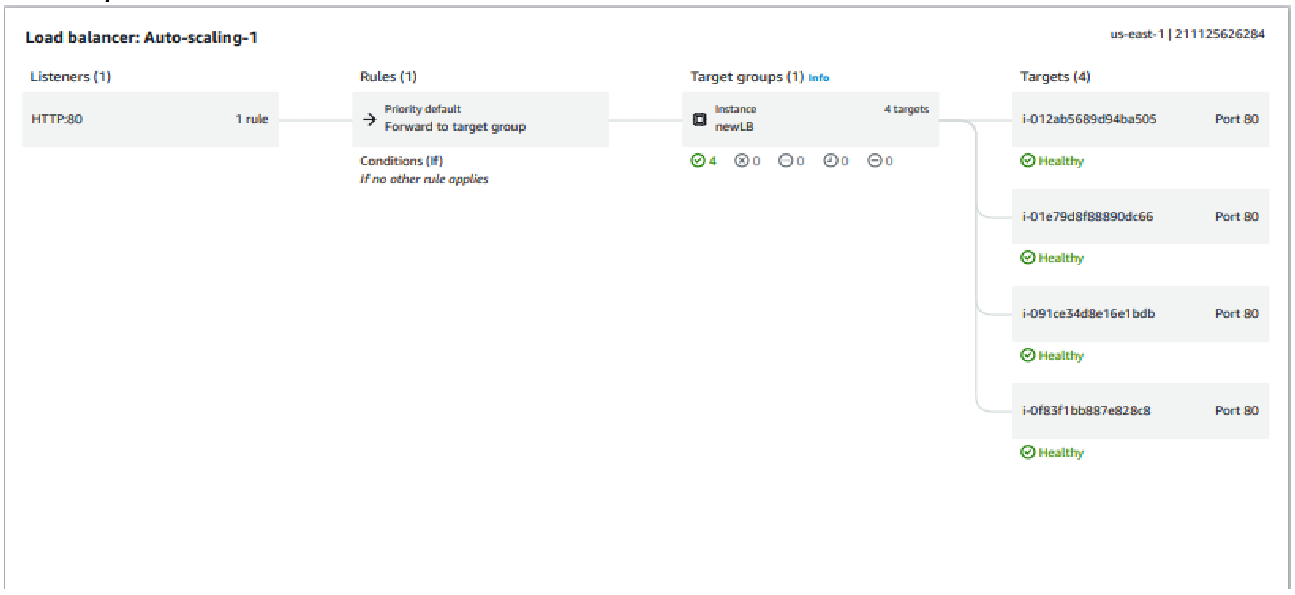


Instances (4) Info

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zo
	i-091ce34d8e16e1bdb	Running	t2.micro	2/2 checks passec	View alarms +	us-east-1a
	i-0f83f1bb887e828c8	Running	t2.micro	Initializing	View alarms +	us-east-1b
	i-01e79d8f88890dc66	Running	t2.micro	Initializing	View alarms +	us-east-1d
	i-012ab5689d94ba505	Running	t2.micro	2/2 checks passec	View alarms +	us-east-1f

Select an instance

- We can see that after achieving max desired capacity all instances generated are in healthy state .


Load balancer: Auto-scaling-1

us-east-1 | 211125626284

Listeners (1)	Rules (1)	Target groups (1) Info	Targets (4)
HTTP:80 1 rule	Priority default Forward to target group	Instance newLB 4 targets	i-012ab5689d94ba505 Port 80 Healthy
	Conditions (If) If no other rule applies	✓4 ⊗0 ⊖0 ⊙0 ⊖0	i-01e79d8f88890dc66 Port 80 Healthy
			i-091ce34d8e16e1bdb Port 80 Healthy
			i-0f83f1bb887e828c8 Port 80 Healthy