Create a simple web application with multiple web servers and implement load balancing using AWS services. The project will involve setting up EC2 instances, configuring an Elastic Load Balancer (ELB), testing the load balancing functionality, autoscaling the servers

1.Launching Two EC2 Instances

Here we are going to launch two Ec2 Instances in US-east-1a & Us-east-1b

EC2 Specifications

Name-Webserver 1

Name-Webserver2

AMI-Amazon Linux 2023

Instance type-t2.micro

Key pair- p11.pem for Webserver1

Key pair- p12.pem for Webserver2

Networking

VPC- Default VPC

Subnet – Subnet in us-east-1a for Websever1

Subnet – Subnet in us-east-1b for Websever2

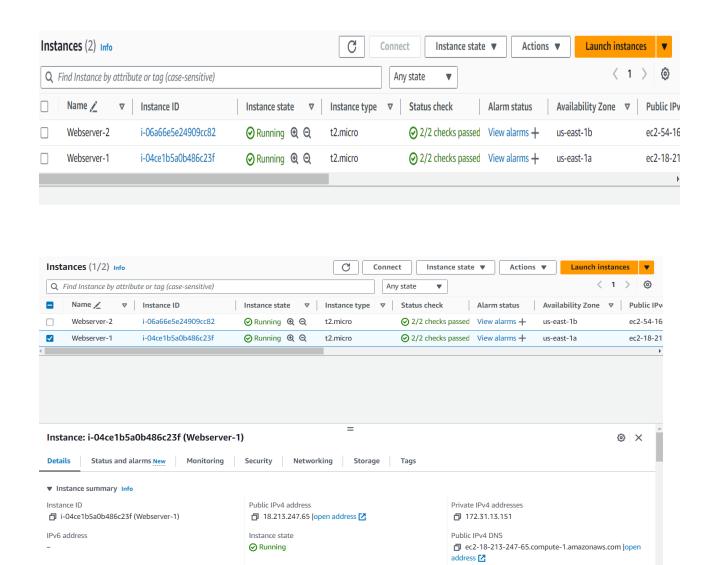
Security group

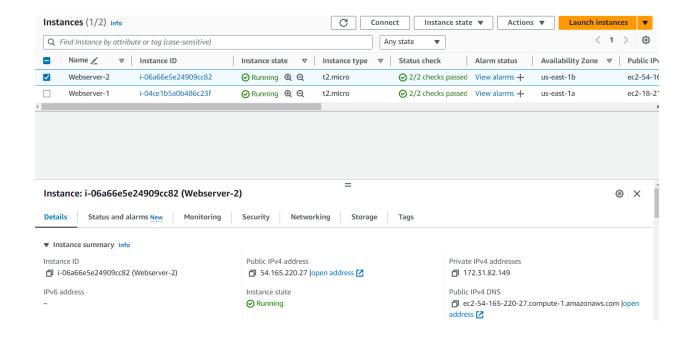
Create an SG which has SSH and HTTP access and attach it to both the instances.

Storage-

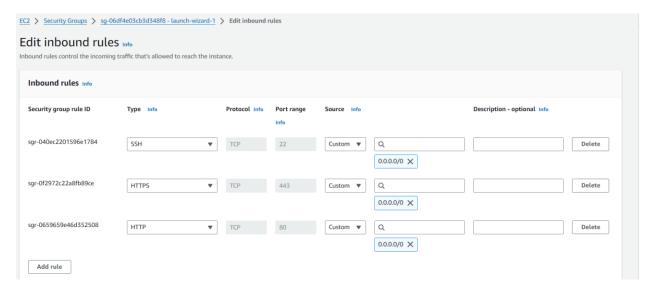
Keep default storage = 8GB as we are launching Linux instance.

Once launched wait till passing 2/2 check





Security group



SSH to Webserver 1 & install Apache webserver on it chmod 400 "p11.pem"

ssh -i "p11.pem" <u>ec2-user@ec2-18-213-247-65.compute-1.amazonaws.com</u>

you will get access to Webserver-1

Execute Following commands on Webserver1

sudo su

yum update –y

```
_/m/'
[ec2-user@ip-172-31-13-151 ~]$ sudo su
[root@ip-172-31-13-151 ec2-user]# yum update -y
Last metadata expiration check: 0:19:12 ago on Wed Jan 24 01:27:15 2024.
Dependencies resolved.
Nothing to do.
Complete!
[root@ip-172-31-13-151 ec2-user]# |
```

yum install -y httpd

```
| State | Stat
```

```
mkdir temp

cd temp

chown ec2-user:ec2-user temp/

chmod 700 temp/

exit

exit
```

scp -i p11.pem simple.zip ec2-user@18.213.247.65:/home/ec2-user/temp/

login again to webserver

ssh -i "p11.pem" ec2-user@ec2-18-213-247-65.compute-1.amazonaws.com

cd temp

you will find sample.zip file

unzip simple.zip

cd simple

mv * /var/www/html

systemctl enable httpd

systemctl start httpd

copy public ip of webserver1 and see whether webpage is accessible or not



We successfully installed web application on webserver1

Repeat this process to install web application on Webserver2

chmod 400 "p12.pem"

ssh -i "p12.pem" ec2-user@ec2-54-165-220-27.compute-1.amazonaws.com

rest of the procedure will be same as per Webserver1

As we just started httpd without installing web app it looks like



It works!

scp -i p12.pem simple.zip ec2-user@54.165.220.27:/home/ec2-user/temp/

Log in to webserver2

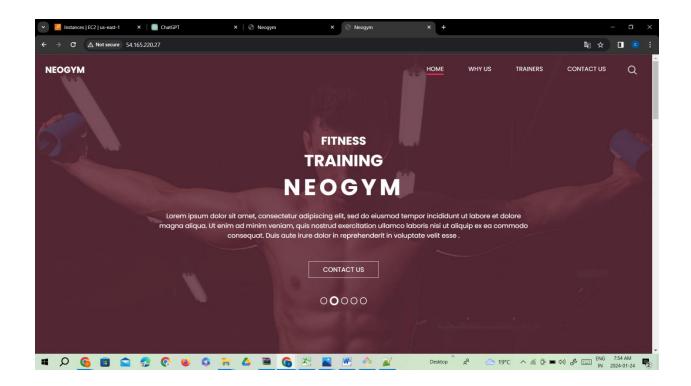
sudo su

cd temp/

unzip simple.zip

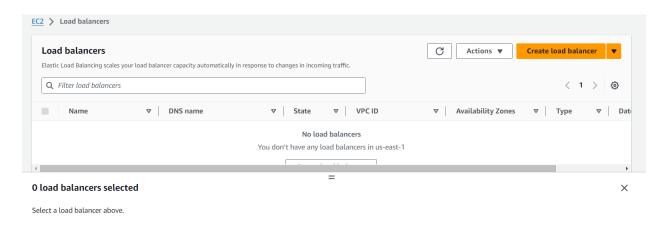
cd simple

mv * /var/www/html



Configuring load balancer

Create load balancer



Create an Application load balancer

Load balancer name- ALB-001

EC2 > Load balancers > Create Application Load Balancer

Create Application Load Balancer Info

The Application Load Balancer distributes incoming HTTP and HTTPS traffic across multiple targets such as Amazon EC2 instances, microservic on request attributes. When the load balancer receives a connection request, it evaluates the listener rules in priority order to determine which applicable, it selects a target from the target group for the rule action.

▶ How Application Load Balancers work

Basic configuration

Load balancer name

Name must be unique within your AWS account and can't be changed after the load balancer is created.

ALB-001

A maximum of 32 alphanumeric characters including hyphens are allowed, but the name must not begin or end with a hyphen.

Scheme Info

Scheme can't be changed after the load balancer is created.

Internet-facing

An internet-facing load balancer routes requests from clients over the internet to targets. Requires a public subnet. Learn more 🔀

○ Internal

An internal load balancer routes requests from clients to targets using private IP addresses.

IP address type Info

Select the type of IP addresses that your subnets use.

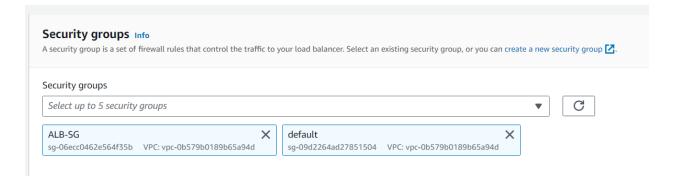
O IPv4

Decommended for internal load balancers

Mappings Info

Select at least two Availability Zones and one subnet per zone. The load balancer routes traffic to targets in these Availability Zones only. Availability Zones th balancer or the VPC are not available for selection.





Create Target group

Specify group details

Your load balancer routes requests to the targets in a target group and performs health checks on the targets.

Basic configuration

Settings in this section can't be changed after the target group is created.

Choose a target type

Instances

- Supports load balancing to instances within a specific VPC.
 Facilitates the use of <u>Amazon EC2 Auto Scaling</u> to manage and scale your EC2 capacity.

O IP addresses

- Supports load balancing to VPC and on-premises resources.
- Facilitates routing to multiple IP addresses and network interfaces on the same instance.
- Offers flexibility with microservice based architectures, simplifying inter-application communication.
 Supports IPv6 targets, enabling end-to-end IPv6 communication, and IPv4-to-IPv6 NAT.

Lambda function

· Facilitates routing to a single Lambda function.

Target group name

tg-001

A maximum of 32 alphanumeric characters including hyphens are allowed, but the name must not begin or end with a hyphen.

-

Protocol: Port

Choose a protocol for your target group that corresponds to the Load Balancer type that will route traffic to it. Some protocols now include anomaly detection for the targets and you can set mitigation options once your target group is created. This choice cannot be changed after creation



IP address type

Only targets with the indicated IP address type can be registered to this target group.

O IPv4

Each instance has a default network interface (eth0) that is assigned the primary private IPv4 address. The instance's primary private IPv4 address is the one that will be applied to the target.

O IPv6

Each instance you register must have an assigned primary IPv6 address. This is configured on the instance's default network interface (eth0). Learn more

VPC

Select the VPC with the instances that you want to include in the target group. Only VPCs that support the IP address type selected above are available in this list.



Protocol version

HTTP1

Health check path Use the default path of "/" to perform health checks on the root, or specify a custom path if preferred.		
1		
Up to 1024 characters allowed.		
▼ Advanced health check settings		
	Restore defaults	
Health check port The port the load balancer uses when performing health checks on targets. By default, the health check port group's traffic port. However, you can specify a different port as an override. Traffic port	is the same as the target	
○ Override		
Override Healthy threshold The number of consecutive health checks successes required before considering an unhealthy target healthy.		
Healthy threshold The number of consecutive health checks successes required before considering an unhealthy target healthy.		
Healthy threshold The number of consecutive health checks successes required before considering an unhealthy target healthy. 5 2-10		
Healthy threshold The number of consecutive health checks successes required before considering an unhealthy target healthy.		

Healthy threshold

The number of consecutive health checks successes required before considering an unhealthy target healthy.

5

2-10

Unhealthy threshold

The number of consecutive health check failures required before considering a target unhealthy.

2

2-10

Timeout

The amount of time, in seconds, during which no response means a failed health check.

5 seconds

2-120

Interval

The approximate amount of time between health checks of an individual target

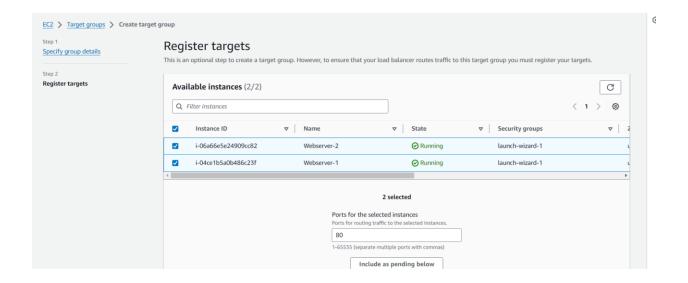
10 seconds

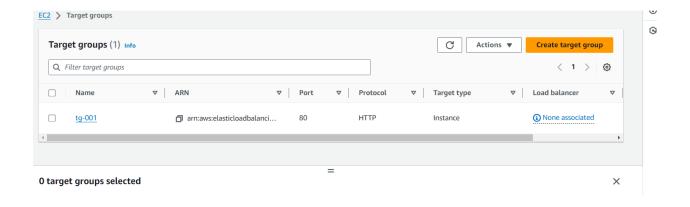
5-300

Success codes

The HTTP codes to use when checking for a successful response from a target. You can specify multiple values (for example, "200,202") or a range of values (for example, "200-299").

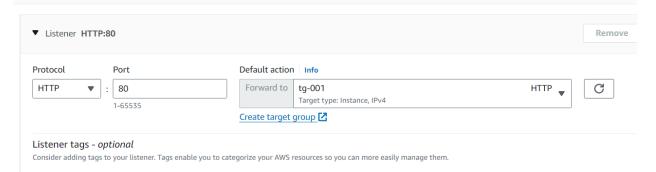
200

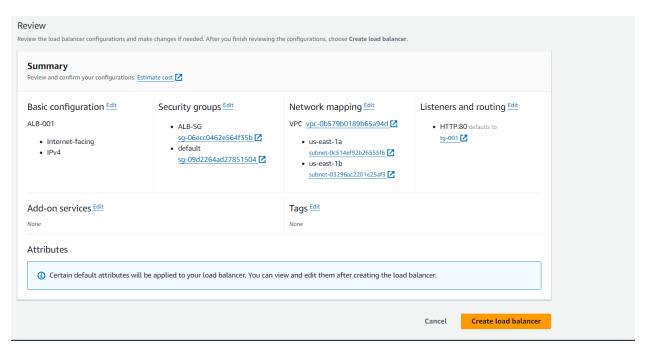




Listeners and routing Info

A listener is a process that checks for connection requests using the port and protocol you configure. The rules that you define for a listener determine how the load balancer routes reques to its registered targets.



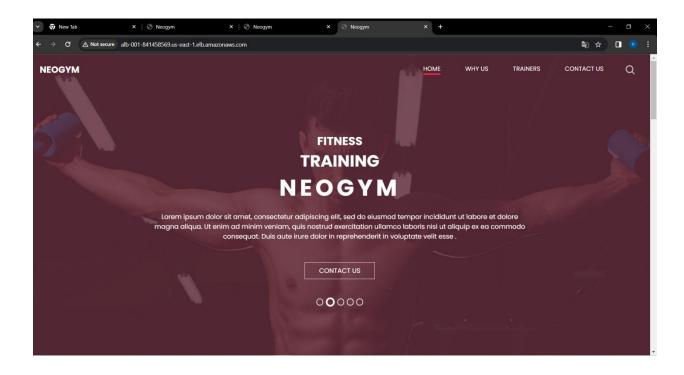


After creating load balancer wail till state is active

Once it is active

Copy ALB DNS name

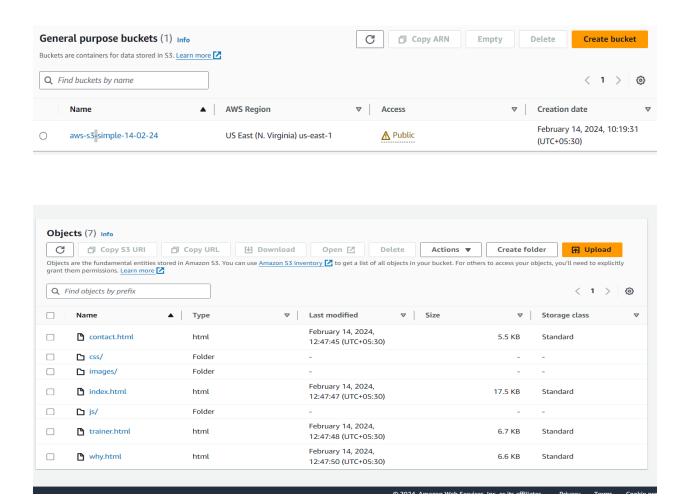
Browse



Auto Scaling group

To create Auto scaling group we have to create launch template

Here we stored application data in s3 bucket



Launch Template

EC2 > Launch templates > Create launch template

Create launch template

Creating a launch template allows you to create a saved instance configuration that can be reused, shared and launched a later time. Templates can have multiple versions.

Launch template name and description

Launch template name - required

Sample-Web-APP

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

Template version description

Template is to deploy application in production enviornment

Max 255 chars

A company of the last

▼ 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

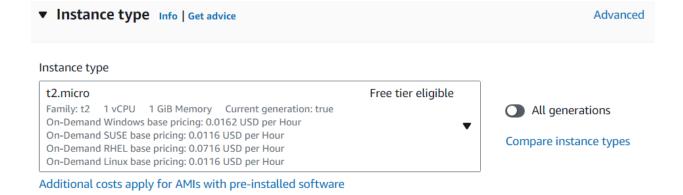


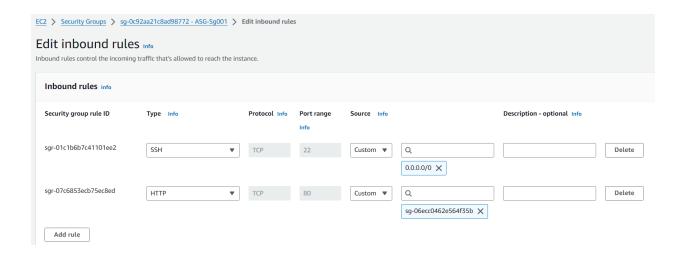
Amazon Machine Image (AMI)



Description

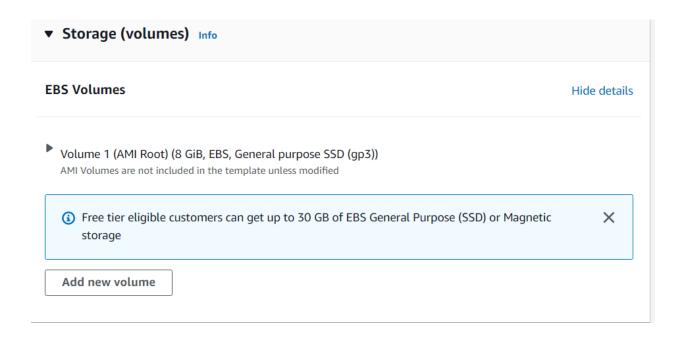
Amazon Linux 2023 AMI 2023.3.20240205.2 x86_64 HVM kernel-6.1

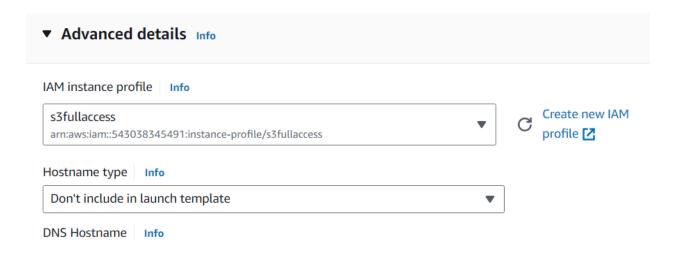




• **sg-06ecc0462e564f35b** is a security group for load balancer which allows traffic from internet.

Here we allow traffic from load balancer to our EC2 instances which are created from auto scaling group.





This roles gives permission to ec2 instance to communicate with s3 & access data from s3.

```
"Version": "2012-10-17",

"Statement": [

{
    "Effect": "Allow",
    "Action": [
        "s3:*",
        "s3-object-lambda:*"
    ],
    "Resource": "*"
    }
]
```

We Have to add user data bash script

```
User data - optional Info
Upload a file with your user data or enter it in the field.

Choose file
```

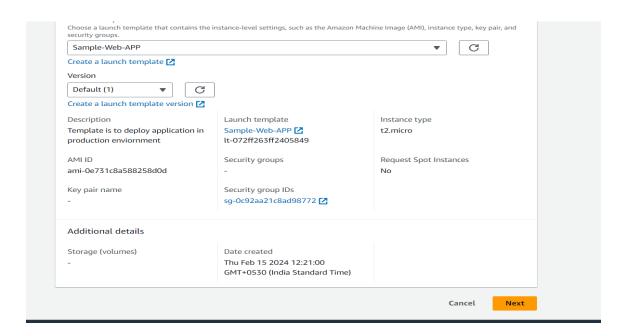
```
#!/bin/bash
yum update -y
yum install httpd -y
yum install -y aws-cli
aws s3 sync s3://aws-s3-simple-14-02-24 /var/www/html
service httpd start
```

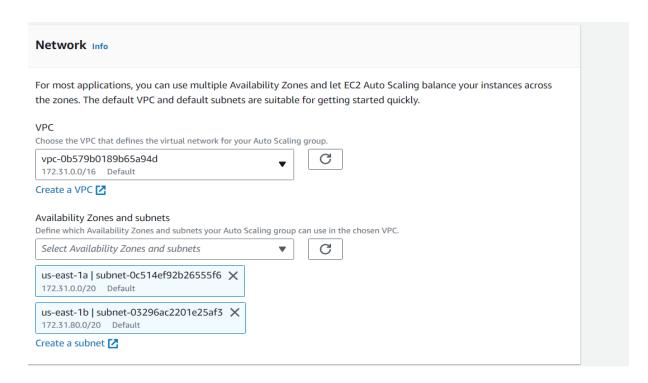
☐ User data has already been base64 encoded

```
#!/bin/bash
yum update -y
yum install httpd -y
yum install -y aws-cli
aws s3 sync s3://aws-s3-simple-14-02-24 /var/www/html
service httpd start
```

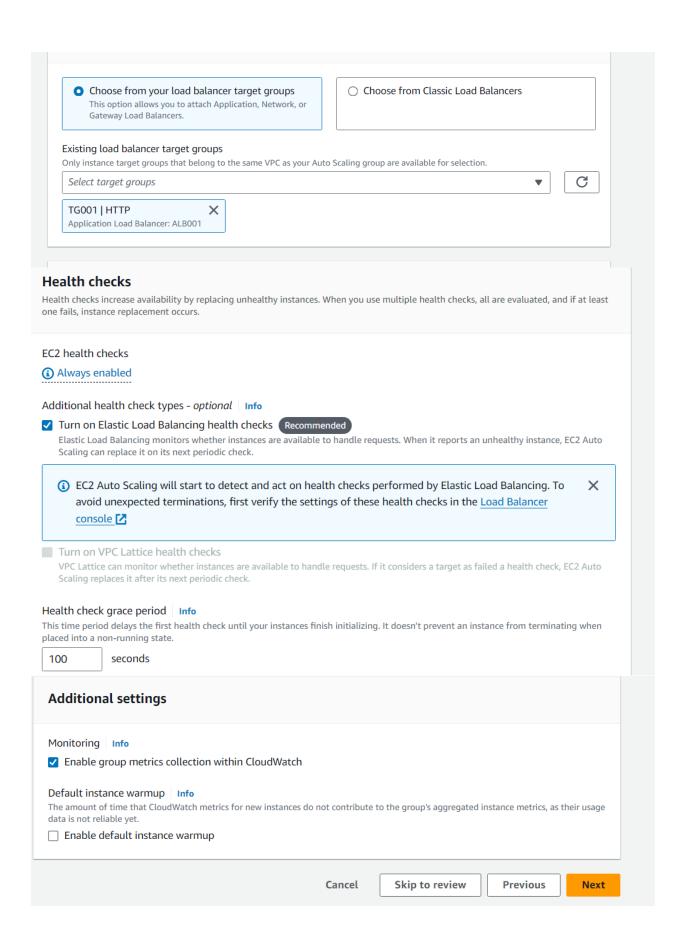


Specify a launch template that contains settings common to all EC2 instances that are launched by this Auto Scaling group. Name Auto Scaling group name Enter a name to identify the group. ASG-Sample-Web-app Must be unique to this account in the current Region and no more than 255 characters. Launch template Info 3 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. C Sample-Web-APP Create a launch template <a>I Version Default (1) Create a launch template version <a>Z





Step 1 Choose launch template	Configure advanced options – optional Info Integrate your Auto Scaling group with other services to distribute network traffic across multiple servers using a load
Step 2 Choose instance launch options	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.
Step 3 - optional Configure advanced options	Load balancing Info
Step 4 - <i>optional</i> Configure group size and	Use the options below to attach your Auto Scaling group to an existing load balancer, or to a new load balancer that you define.
scaling	O No load balancer Traffic to your Auto Scaling group will not be fronted by a load balancer Attach to an existing load balancer balancer balancer
Step 5 - optional Add notifications	will not be fronted by a load balancer. Choose from your existing load balancer to attach to your Auto Scaling group.
itep 6 - optional	Attach to an existing load balancer
tad tags	Select the load balancers that you want to attach to your Auto Scaling group.
Step 7	
Review	 Choose from your load balancer target groups This option allows you to attach Application, Network, or Gateway Load Balancers.



Configure group size and scaling - optional Info

Define your group's desired capacity and scaling limits. You can optionally add automatic scaling to adjust the size of your group.

Group size Info

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

Desired capacity type

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

Units (number of instances) ▼



Desired capacity
Specify your group size.

2

Scaling Info

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

Scaling limits

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

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

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.

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 Your Auto Scaling group will remain at its initial size and Choose a CloudWatch metric and target value and let the will not dynamically resize to meet demand. scaling policy adjust the desired capacity in proportion to the metric's value. Scaling policy name Monitor CPU utilization metric 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 80 Instance warmup Info 100 seconds ☐ Disable scale in to create only a scale-out policy 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. ▼ Notification 1 Remove Send a notification to my-sns-topic-1 With these recipients cpandharpure@gmail.com Use existing topic Event types Notify subscribers whenever instances Launch Terminate Fail to launch ✓ Fail to terminate

Details

Type Info

Topic type cannot be modified after topic is created

O FIFO (first-in, first-out)

- · Strictly-preserved message ordering
- Exactly-once message delivery
- High throughput, up to 300 publishes/second
- Subscription protocols: SQS

Standard

- Best-effort message ordering
- · At-least once message delivery
- Highest throughput in publishes/second
- Subscription protocols: SQS, Lambda, HTTP, SMS, email, mobile application endpoints

Mamo

My-SNS-001

Maximum 256 characters. Can include alphanumeric characters, hyphens (-) and underscores (_).

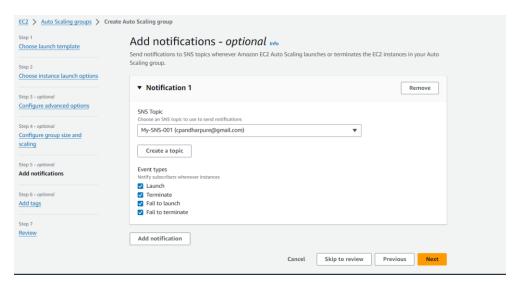
Display name - optional | Info

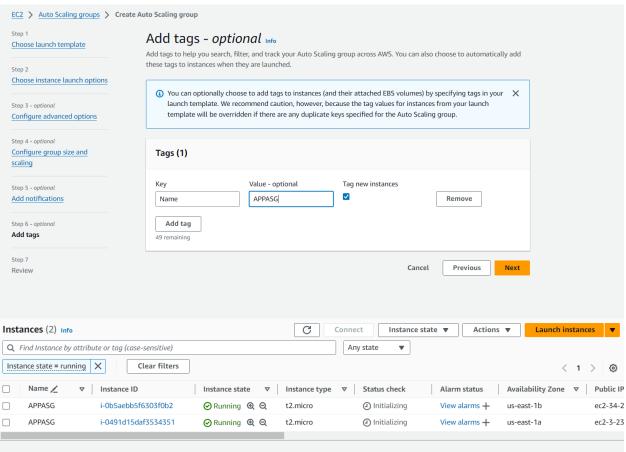
To use this topic with SMS subscriptions, enter a display name. Only the first 10 characters are displayed in an SMS message.

Notifications from the Autoscaling group

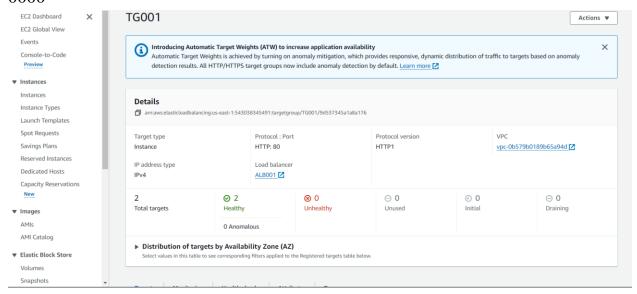
Maximum 100 characters.

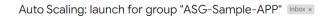
Create subscription Details Topic ARN Q arm:aws:sns:us-east-1:543038345491:My-SNS-001 Protocol The type of endpoint to subscribe Email Endpoint An email address that can receive notifications from Amazon SNS. cpandharpure@gmail.com ④ After your subscription is created, you must confirm it. Info















Notifications from the Autoscaling group

Service: AWS Auto Scaling Time: 2024-02-15T07:23:45.513Z Requestld: 140636ae-97a8-3bb9-e892-068db67512d3 Event: autoscaling: EC2_INSTANCE_LAUNCH Accountld: 54303



Notifications from the Autoscaling group <no-reply@sns.amazonaws.com>

12:53 PM (6 minutes ago)



☆

Service: AWS Auto Scaling

Time: 2024-02-15T07:23:45.649Z

Requestld: f6a636ae-97a9-71c3-7fd3-ccc662794ab6

Event: autoscaling:EC2_INSTANCE_LAUNCH

Accountld: 543038345491

AutoScalingGroupName: ASG-Sample-APP

AutoScalingGroupARN: arn:aws:autoscaling:us-east-1:543038345491:autoScalingGroup:cd936eeb-3543-4bae-8683-84dde6a70b89:autoScalingGroupName/ASG-Sample-APP

ActivityId: f6a636ae-97a9-71c3-7fd3-ccc662794ab6

Description: Launching a new EC2 instance: i-0b5aebb5f6303f0b2

Cause: At 2024-02-15T07:23:08Z a user request created an AutoScalingGroup changing the desired capacity from 0 to 2. At 2024-02-15T07:23:11Z an instance was started in response to a

difference between desired and actual capacity, increasing the capacity from 0 to 2.

StartTime: 2024-02-15T07:23:13.884Z

EndTime: 2024-02-15T07:23:45 6497

StatusCode: InProgress

StatusMessage:

Progress: 50 EC2Instanceld: i-0b5aebb5f6303f0b2

Details: {"Subnet ID":"subnet-03296ac2201e25af3","Availability Zone":"us-east-1b"}

Instance: i-0b5aebb5f6303f0b2 (APPASG)

4

