

Part 1: EC2 with ELB and ASG

Objective: Learn how to create a scalable and highly available web application environment using Amazon EC2 instances, ELB, and ASG.

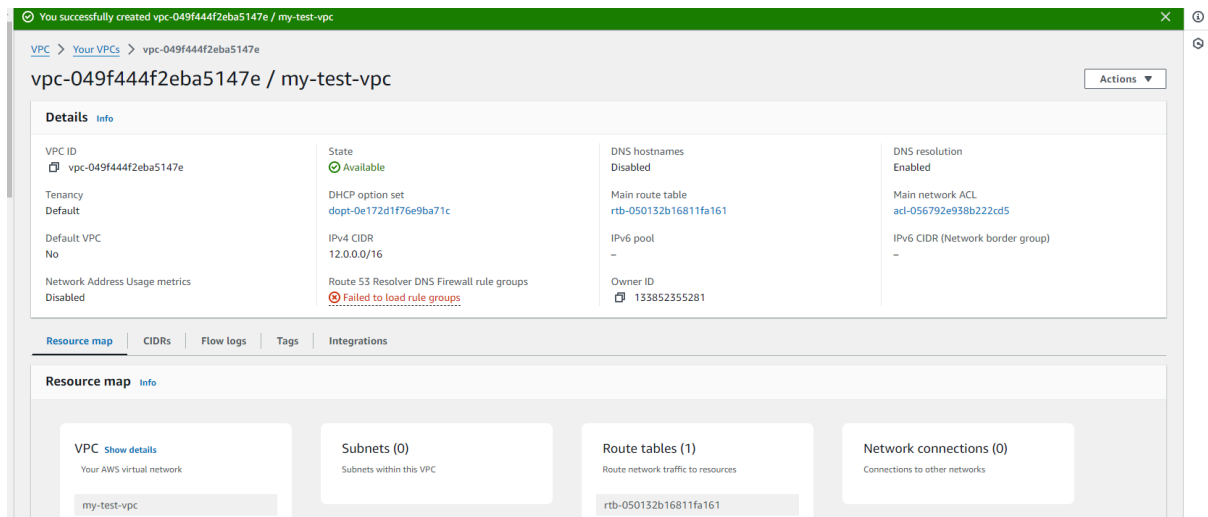
Approach:

1. **Launch EC2 Instances:** Start by launching two or more EC2 instances. These instances will run a simple web application (e.g., a "Hello World" page or any basic web service).
2. **Configure Load Balancer:** Set up an Elastic Load Balancer (ELB) to distribute incoming web traffic across your EC2 instances. This step ensures high availability and fault tolerance.
3. **Set Up Auto Scaling Group (ASG):** Create an ASG that uses the launched EC2 instances. Configure ASG policies to automatically scale the number of instances up or down based on criteria like CPU usage or network traffic.
4. **Test Your Setup:** Simulate traffic to test the scaling policies and the load balancer. Observe how ASG adds or removes instances and how ELB distributes traffic.
5. **Verify Website Functionality:** Ensure that the website hosted on EC2 instances remains accessible and functional during scaling operations.

Goal: By the end of this lab, students will have a hands-on understanding of setting up a load-balanced and auto-scaled web application using AWS services.

Solution:

1. Created my-test-vpc



2. Appropriate Internet Gateway was created

[VPC](#) > [Internet gateways](#) > Create internet gateway

Create internet gateway [Info](#)

An internet gateway is a virtual router that connects a VPC to the internet. To create a new internet gateway specify the name for the gateway below.

Internet gateway settings

Name tag
Creates a tag with a key of 'Name' and a value that you specify.

Tags - optional

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS costs.

Key	Value - optional	
<input type="text" value="Name"/>	<input type="text" value="my-test-ig"/>	<input type="button" value="Remove"/>

You can add 49 more tags.

3. Attached the internet gateway to the VPC

[VPC](#) > [Internet gateways](#) > Attach to VPC (igw-0710a8b9877967214)

Attach to VPC (igw-0710a8b9877967214) [Info](#)

VPC

Attach an internet gateway to a VPC to enable the VPC to communicate with the internet. Specify the VPC to attach below.

Available VPCs
Attach the internet gateway to this VPC.

▶ **AWS Command Line Interface command**

- Subnets were created with proper IP address assignment

[VPC](#) > [Subnets](#) > Create subnet

Create subnet [Info](#)

VPC

VPC ID

Create subnets in this VPC.

vpc-049f444f2eba5147e (my-test-vpc) ▼

Associated VPC CIDRs

IPv4 CIDRs
12.0.0.0/16

Subnet settings

Specify the CIDR blocks and Availability Zone for the subnet.

Subnet 1 of 1

Subnet name

Create a tag with a key of 'Name' and a value that you specify.

my-test-public-subnet1

The name can be up to 256 characters long.

Availability Zone [Info](#)

Choose the zone in which your subnet will reside, or let Amazon choose one for you.

US East (N. Virginia) / us-east-1a ▼

IPv4 VPC CIDR block [Info](#)

Subnet 1 of 2

Subnet name

Create a tag with a key of 'Name' and a value that you specify.

my-test-public-subnet1

The name can be up to 256 characters long.

Availability Zone [Info](#)

Choose the zone in which your subnet will reside, or let Amazon choose one for you.

US East (N. Virginia) / us-east-1a

IPv4 VPC CIDR block [Info](#)

Choose the VPC's IPv4 CIDR block for the subnet. The subnet's IPv4 CIDR must lie within this block.

12.0.0.0/16

IPv4 subnet CIDR block

12.0.1.0/24256 IPs

<>^v

▼ Tags - optional

Key

Value - optional

Q NameX

Q my-test-public-subnet1X

Remove

Add new tag

You can add 49 more tags.

Remove

Subnet 2 of 2

Subnet name

Create a tag with a key of 'Name' and a value that you specify.

my-test-subnet2

The name can be up to 256 characters long.

Availability Zone [Info](#)

Choose the zone in which your subnet will reside, or let Amazon choose one for you.

US East (N. Virginia) / us-east-1b

IPv4 VPC CIDR block [Info](#)

Choose the VPC's IPv4 CIDR block for the subnet. The subnet's IPv4 CIDR must lie within this block.

12.0.0.0/16

IPv4 subnet CIDR block

12.0.3.0/24256 IPs

<>^v

▼ Tags - optional

Key

Value - optional

Q NameX

Q my-test-subnet2X

Remove

Add new tag

You can add 49 more tags.

5. Route table was created

VPC > Route tables > Create route table

Create route table [Info](#)

A route table specifies how packets are forwarded between the subnets within your VPC, the internet, and your VPN connection.

Route table settings

Name - optional
Create a tag with a key of 'Name' and a value that you specify.

VPC
The VPC to use for this route table.

Tags

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS costs.

Key	Value - optional	
<input type="text" value="Name"/>	<input type="text" value="my-test-route-table"/>	<input type="button" value="Remove"/>
<input type="button" value="Add new tag"/>		

You can add 49 more tags.

6. Click on “Edit subnet associations” within the created route table and assign the created subnets and then edit routes to add the internet gateway

VPC > Route tables > rtb-008137ee41b98bd35 > Edit subnet associations

Edit subnet associations

Change which subnets are associated with this route table.

Available subnets (2)

<input type="checkbox"/>	Name	Subnet ID	IPv4 CIDR	IPv6 CIDR	Route table ID
<input type="checkbox"/>	my-test-subnet2	subnet-0eda319b04e2192cb	12.0.3.0/24	-	Main (rtb-050132b16811fa161)
<input type="checkbox"/>	my-test-public-subnet1	subnet-0162194526e1124f0	12.0.1.0/24	-	Main (rtb-050132b16811fa161)

7. Now create the Target Table

VPC > Lattice: Target groups > Create target group

Step 1
Specify group details

Step 2
Register targets

Specify group details

Your service 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 instances within a specific VPC.
- ☐ **IP addresses**
 - Supports traffic to VPC resources.
 - Facilitates routing to multiple IP addresses and network interfaces on the same instance.
 - Offers flexibility with microservice based architectures, simplifying inter-application communication.
- ☐ **Lambda function**
 - Facilitates routing to a single Lambda function.
- ☐ **Application Load Balancer**
 - Facilitates routing to a single Application Load Balancer.

Target group name

8. Create security group with HTTP enabled, also create another security group with both HTTP and SSH enabled

EC2 > Security Groups > Create security group

Create security group

A security group acts as a virtual firewall for your instance to control inbound and outbound traffic. To create a new security group, complete the fields below.

Basic details

Security group name [Info](#)
test-alb-sg
Name cannot be edited after creation.

Description [Info](#)
Allow http requests

VPC [Info](#)
vpc-049f44f2eba5147e (my-test-vpc)

Inbound rules [Info](#)

Type Info	Protocol Info	Port range Info	Source Info	Description - optional
HTTP	TCP	80	Anywhere... 0.0.0.0/0	

[Add rule](#) [Delete](#)

Security groups [Info](#)

A security group is a set of firewall rules that control the traffic to your load balancer. Select an existing security group, or you can [create a new security group](#).

Security groups

Select up to 5 security groups



test-alb-sg
sg-01b26c0e07386ece4 VPC: vpc-049f444f2eba5147e

×

default
sg-05182a6d714d9a9a6 VPC: vpc-049f444f2eba5147e

×

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 requests to its registered targets.

▼ Listener HTTP:80

Remove

Protocol

HTTP ▼

Port

80

1-65535

Default action

Forward to

new-f-test

Target type: Instance, IPv4

HTTP ▼

⌂

[Create target group](#)

Listener tags - optional

[EC2](#) > [Security Groups](#) > Create security group

Create security group [info](#)

A security group acts as a virtual firewall for your instance to control inbound and outbound traffic. To create a new security group, complete the fields below.

Basic details

Security group name [Info](#)

sg-for-temp

Name cannot be edited after creation.

Description [Info](#)

Allows SSH access to developers

VPC [Info](#)

vpc-049f444f2eba5147e (my-test-vpc) ▼

Inbound rules [Info](#)

Type Info	Protocol Info	Port range Info	Source Info	Description - optional Info	
HTTP ▼	TCP	80	Anywhere... ▼ 0.0.0.0/0 ✕		Delete
SSH ▼	TCP	22	Anywhere... ▼ Search		Delete

9. Finally Create Application Load Balancer

[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, microservices, and containers, based on request attributes. When the load balancer receives a connection request, it evaluates the listener rules in priority order to determine which rule to apply, and if 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.

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.

☒ IPv4

Recommended for internal load balancers.

☐ Dualstack

Includes IPv4 and IPv6 addresses.

Network mapping [Info](#)

The load balancer routes traffic to targets in the selected subnets, and in accordance with your IP address settings.

VPC [Info](#)

Select the virtual private cloud (VPC) for your targets or you can [create a new VPC](#). Only VPCs with an internet gateway are enabled for selection. The selected VPC can't be changed after the load balancer is created. To confirm the VPC for your targets, view your [target groups](#).

vpc-049f444f2eba5147e

IPv4: 12.0.0.0/16



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 that are not supported by the load balancer or the VPC are not available for selection.

☒ us-east-1a (use1-az4)

Subnet

my-test-public-subnet1 ▼

IPv4 address

Assigned by AWS

☒ us-east-1b (use1-az6)

Subnet

my-test-subnet2 ▼

IPv4 address

Assigned by AWS

10. But wait, you need more things to create load balancer, now create 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 at a later time. Templates can have multiple versions.

Launch template name and description

Launch template name - required

my-test-temp

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

Template version description

A prod webserver for MyApp

Max 255 chars

Auto Scaling guidance

Info

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

▶ Template tags

▶ Source template

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.

▼ Summary

Software Image (AMI)

Amazon Linux 2023 AMI 2023.3.2...read more

ami-0440d3b780d96b29d

Virtual server type (instance type)

t2.micro

Firewall (security group)

my-sg-temp

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, 30 GiB of EBS storage, 2 million IOs, 1 GB of snapshots, and 100 GB of bandwidth to the internet.

Cancel

Create launch template

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.

▼ 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

My AMIs

Quick Start

Amazon Linux

aws

macOS

Mac

Ubuntu

ubuntu

Windows

Microsoft

Red Hat

Red Hat

SUSE Li

SUS

Browse more AMIs

Including AMIs from AWS, Marketplace and the Community

Amazon Machine Image (AMI)

Free tier eligible

Amazon Linux 2023 AMI

ami-0440d3b780d96b29d (64-bit (x86), uefi-preferred) / ami-0f93c02efd1974b8b (64-bit (Arm), uefi)

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

Description

Amazon Linux 2023 AMI 2023.3.20240219.0 x86_64 HVM kernel-6.1

Architecture

Boot mode

AMI ID

Software Image (AMI)

Amazon Linux 2023 AMI 2023.3.2...read more

ami-0440d3b780d96b29d

Virtual server type (instance type)

t2.micro

Firewall (security group)

my-sg-temp

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, 30 GiB of EBS storage, 2 million IOs, 1 GB of snapshots, and 100 GB of bandwidth to the internet.

Cancel

Create launch template

▼ Network settings [Info](#)

Subnet [Info](#)

Don't include in launch template ▼

 [Create new subnet](#) 

When you specify a subnet, a network interface is automatically added to your template.

Firewall (security groups) [Info](#)

A security group is a set of firewall rules that control the traffic for your instance. Add rules to allow specific traffic to reach your instance.


☒ Select existing security group

☐ Create security group

Security groups [Info](#)

Select security groups ▼


my-sg-temp sg-0b6130d6240bf5ee0 ✕
VPC: vpc-049f444f2eba5147e

 [Compare security group rules](#)

► **Advanced network configuration**

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 -y httpd
systemctl start httpd
systemctl enable httpd
echo "<h1> Hello World from $(hostname -f)<a/h1>" > /var/www/html/index.html
```


☐ User data has already been base64 encoded

Firewall (security group)

my-sg-temp

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, 30 GiB of EBS storage, 2 million IOs, 1 GB of snapshots, and 100 GB of bandwidth to the internet. ✕

[Cancel](#)

[Create launch template](#)

11. Now create auto scaling group and choose the created launch template

Choose launch template Info

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.

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.

▼↺

Create a launch template [↗](#)

Version

▼↺

Create a launch template version [↗](#)

Description	Launch template	Instance type
-	my-test-temp ↗	t2.micro
	lt-07a4584d6e88e69af	
AMI ID	Security groups	Request Spot instances
ami-0440d3b780d96b29d	-	No
Key pair name	Security group IDs	
vpckey	sg-0b6130d6240bf5ee0 ↗	

Additional details

Storage (volumes)	Date created
-	Fri Feb 23 2024 13:29:23 GMT+0545 (Nepal Time)

CancelNext

Choose instance launch options [Info](#)

Choose the VPC network environment that your instances are launched into, and customize the instance types and purchase options.

Instance type requirements [Info](#)

[Override launch template](#)

You can keep the same instance attributes or instance type from your launch template, or you can choose to override the launch template by specifying different instance attributes or manually adding instance types.

Launch template	Version	Description
my-test-temp lt-07a4584d6e88e69af	Default	-
Instance type		
t2.micro		

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-049f444f2eba5147e (my-test-vpc)

12.0.0.0/16

[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-0162194526e1124f0 (my-test-public-subnet1) X

12.0.1.0/24

us-east-1b | subnet-0eda319b04e2192cb (my-test-subnet2) X

12.0.3.0/24

[Create a subnet](#)[Cancel](#)[Skip to review](#)[Previous](#)[Next](#)

12. Attach the load balancer

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](#)

Step 6 - optional
[Add tags](#)

Step 7
[Review](#)

Configure advanced options - optional [Info](#)

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 [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 an existing load balancer

Select the load balancers that you want to attach to your Auto Scaling group.

☒ Choose from your load balancer target groups
This option allows you to attach Application, Network, or Gateway Load Balancers.

☐ Choose from Classic Load Balancers

Existing load balancer target groups
Only instance target groups that belong to the same VPC as your Auto Scaling group are available for selection.

Select target groups

new-f-test | HTTP
Application Load Balancer: my-test-lb

VPC Lattice integration options [Info](#)

To improve networking capabilities and scalability, integrate your Auto Scaling group with VPC Lattice. VPC Lattice facilitates communications between AWS services and helps you connect and manage your applications across compute services in AWS.

Select VPC Lattice service to attach

☒ No VPC Lattice service
VPC Lattice will not manage your Auto Scaling group's network access and connectivity with other services.

☐ Attach to VPC Lattice service
Incoming requests associated with specified VPC Lattice target groups will be routed to your Auto Scaling group.

[Create new VPC Lattice service](#)

Health checks

Health checks increase availability by replacing unhealthy instances. When you use multiple health checks, all are evaluated, and if at least one fails, instance replacement occurs.

EC2 health checks

☒ Always enabled

Additional health check types - optional [Info](#)

☐ Turn on Elastic Load Balancing health checks **Recommended**
Elastic Load Balancing monitors whether instances are available to handle requests. When it reports an unhealthy instance, EC2 Auto Scaling can replace it on its next periodic check.

☐ Turn on VPC Lattice health checks
VPC Lattice can monitor whether instances are available to handle requests. If it considers a target as failed a health check, EC2 Auto Scaling replaces it after its next periodic check.

Health check grace period [Info](#)
This time period delays the first health check until your instances finish initializing. It doesn't prevent an instance from terminating when placed into a non-running state.

300

seconds

Additional settings

13. Configure group size as below:

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.

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

Equal or less than desired capacity

Max desired capacity

Equal or greater than desired capacity

Automatic scaling - optional [Info](#)

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.

Instance maintenance policy - new [Info](#)

Control your Auto Scaling group's availability during instance replacement events. This includes health checks, instance refreshes, maximum instance lifetime features and events that happen automatically to keep your group balanced, called rebalancing events.

**Control availability and cost during replacement events** ✕

An instance maintenance policy determines how much availability your application has when EC2 Auto Scaling replaces instances. It also establishes guardrails that limit the amount of capacity that can be added or removed when replacing instances.

Choose a replacement behavior depending on your availability requirements

Mixed behavior

☒ No policy
For rebalancing events, new instances will launch before terminating others. For all other events, instances terminate and launch at the same time.

Prioritize availability

☐ Launch before terminating
Launch new instances and wait for them to be ready before terminating others. This allows you to go above your desired capacity by a given percentage and may temporarily increase costs.

Control costs

☐ Terminate and launch
Terminate and launch instances at the same time. This allows you to go below your desired capacity by a given percentage and may temporarily reduce availability.

Flexible

☐ Custom behavior
Set custom values for the minimum and maximum amount of available capacity. This gives you greater flexibility in setting how far below and over your desired capacity EC2 Auto Scaling goes when replacing instances.

Instance scale-in protection

Scale-in protection prevents newly launched instances from being terminated by scaling activities. Make sure to remove scale-in protection for the reason or individual instances when instances are nearly to be terminated.

14. Created the Auto Scaling Group Successfully: What a relief!!, here are the created instances by ASG

my-test-asg

Details | Activity | Automatic scaling | **Instance management** | Monitoring | Instance refresh

Instances (2)

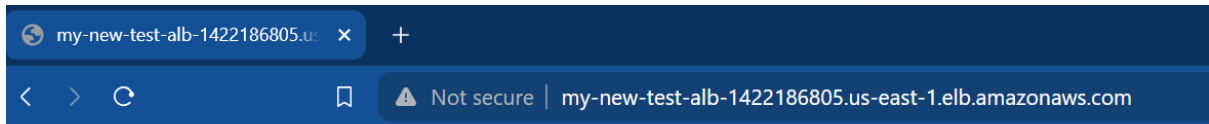
Instance ID	Lifecycle	Instance type	Weighted capacity	Launch template/co...	Availability Zone	Health status	Protected from
i-00c48dbcd5a0d0356	Pending	t2.micro	-	my-test-temp Version 1	us-east-1b	Healthy	
i-08579632ebd0cc25a	Pending	t2.micro	-	my-test-temp Version 1	us-east-1a	Healthy	

Lifecycle hooks (0)

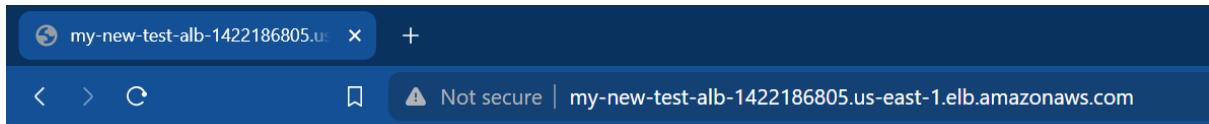
No lifecycle hooks are currently configured.

Lifecycle hooks help you perform custom actions on instances as they launch and before they terminate.

15. After fourth attempt, finally got this result: each refresh will point to new EC2 instance, (dns name was copied from load balancer and then pasted as below)



Hello World from ip-12-0-3-39.ec2.internal



Hello World from ip-12-0-1-52.ec2.internal