

# Auto Scaling with Application Load Balancer (ALB)

---

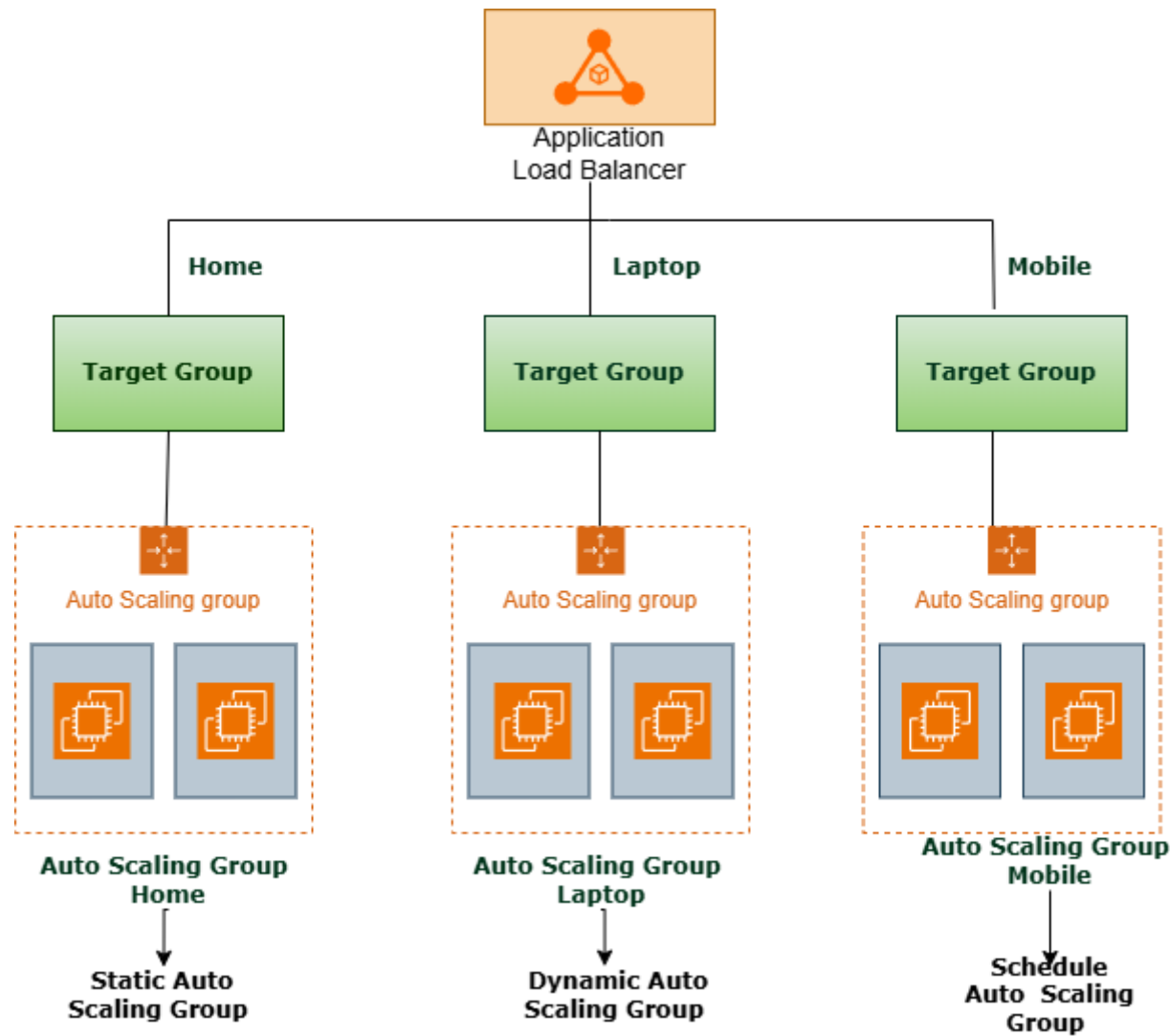
## Introduction.

Efficient traffic management and scalability are essential for modern cloud applications. By combining an **Application Load Balancer(ALB)** with multiple **Auto Scaling Groups (ASGs)**, workloads can be distributed intelligently across different user categories **Home, Laptop, and Mobile**. Each group uses a unique scaling strategy (Static, Dynamic, and Scheduled) to ensure consistent performance, high availability, and optimized resource utilization.

## Architecture Components.

- 1. Application Load Balancer (ALB)
  - Distributes incoming traffic across multiple targets (EC2 instances) in different Auto Scaling Groups.
  - Ensures high availability and fault tolerance.
  - Routes requests based on listener rules to the appropriate Target Group (Home, Laptop, or Mobile).
- 2. Target Groups
  - Logical groups of instances that receive traffic from the ALB.
  - Each target group is associated with an Auto Scaling Group.

Target Group	Description	Scaling Strategy
Home	Handles traffic from home devices	Static Auto Scaling
Laptop	Manages traffic from laptops	Dynamic Auto Scaling
Mobile	Serves mobile users	Scheduled Auto Scaling



# Step-by-Step Deployment Guide.

Step 1: Create Launch Template.

## Launch Template (Home-TG).

☰

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*

Home-TG

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

#### Template version description

This is my Home Page

Max 255 chars

**User data - optional** | **Info**

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

↑ Choose file

```
#!/bin/bash
sudo yum update -y
sudo yum install httpd -y
sudo systemctl start httpd
sudo systemctl enable httpd
echo "<h1>This is my Home Page|$(hostname -f)</h1>" > /var/www/html/index.html
```

**Launch Template (Laptop-TG).****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*

Laptop-TG

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

**Template version description**

This is my Laptop Page|

Max 255 chars

User data - optional | Info

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

Choose file

```
#!/bin/bash
sudo yum update -y
sudo yum install httpd -y
sudo systemctl start httpd
sudo systemctl enable httpd
echo "<h1>This is my Laptop page $(hostname -f)</h1>" >
/var/www/html/laptop/index.html
```

☐ User data has already been base64 encoded

Launch Template(Mobile-TG).

aws | Search [Alt+S]

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

Mobile-TG

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

Template version description

This is my Mobile Page

Max 255 chars

4 / 16

User data - optional | Info

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

Choose file

```
#!/bin/bash
sudo yum update -y
sudo yum install httpd -y
sudo systemctl start httpd
sudo systemctl enable httpd
echo "<h1>This is my Mobile Page$(hostname -f)</h1>" >
/var/www/html/mobile/index.html
```

☐ User data has already been base64 encoded

Launch Templates (3) Info

Actions Create launch template

Search

	Launch Template ID	Launch Template Name	Default Version	Latest Version	Create Time	Created By
<input type="checkbox"/>	lt-06e01956fc442c992	Home-TG	1	1	2025-09-25T18:36:12.000Z	arn:aws:iam::9524
<input type="checkbox"/>	lt-09d8e6311fad8e716	Laptop-TG	1	1	2025-09-25T18:55:02.000Z	arn:aws:iam::9524
<input type="checkbox"/>	lt-0adfad693e0501ba5	Mobile-TG	1	1	2025-09-25T18:59:22.000Z	arn:aws:iam::9524

Step 2: Create Auto Scaling Groups (ASGs).

Home ASG: Use Static Scaling for a fixed number of instances.

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.

Home-ASG

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.

Home-TG

Create a launch template

5 / 16

### 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

use1-az1 (us-east-1b) | subnet-0a0a2f906dbf510a7  
172.31.0.0/20    Default

use1-az2 (us-east-1c) | subnet-012dbf3ca55b1eabb  
172.31.80.0/20    Default

Create a subnet

### Integrate with other services - optional

Use a load balancer to distribute network traffic across multiple servers. Enable service-to-service communications with VPC Lattice. Shift resources away from impaired Availability Zones with zonal shift. You can also customize health check replacements and monitoring.

#### Load balancing

Use the options below to attach your Auto Scaling group to an existing load balancer, or to a new load balancer that you define.

☒ No load balancer  
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.

#### Desired capacity

Specify your group size.

2

#### Scaling

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

2

Equal or less than desired capacity

##### Max desired capacity

2

Equal or greater than desired capacity

##### Automatic scaling - optional

Choose whether to use a target tracking policy

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.

Laptop ASG: Implement Dynamic Scaling based on performance metrics.

6 / 16

### 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

use1-az1 (us-east-1b) | subnet-0a0a2f906dbf510a7

172.31.0.0/20 Default

×

use1-az2 (us-east-1c) | subnet-012dbf3ca55b1eabb

172.31.80.0/20 Default

×

Create a subnet

7 / 16

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](#)

300 seconds

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

Mobile ASG: Apply Scheduled Scaling for predictable workloads.

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.

Mobile-ASG

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 group 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.

Mobile-TG

Create a launch template [?](#)

8 / 16



### Desired capacity

Specify your group size.

3

### 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

2

Equal or less than desired capacity

#### Max desired capacity

7

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

300

 seconds

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

Auto Scaling groups (1/3) Info

Last updated less than a minute ago

Launch configurationsLaunch templates ↗Actions ▼Create Auto Scaling group

Q Search your Auto Scaling groups

<input type="checkbox"/>	Name	Launch template/configuration <small>↗</small>	Instances	Status	Desired capacity	Min	Max	Availability Zones	
<input type="checkbox"/>	Laptop-ASG	Laptop-TG   Version Default	0	⌚ Updating capacity...	3	2	7	2 Availability Zones	ⓘ
<input type="checkbox"/>	Mobile-ASG	Mobile-TG   Version Default	3	-	3	2	7	2 Availability Zones	ⓘ
<input checked="" type="checkbox"/>	Home-ASG	Home-TG   Version Default	2	-	2	2	2	2 Availability Zones	ⓘ


Step 3: Create scheduled action.

9 / 16

## Create scheduled action

### Name

GreatIndianSale

 Provide at least one value for Desired, Min, or Max Capacity

### Desired capacity

8

### Min

5

### Max

15

### Recurrence

Cron

30 10 21 10 \*

### Time zone

Etc/UTC

### End by

2025/10/30



10:30

Etc/UTC

### Scheduled actions (1/1) Info

 Filter scheduled actions



Actions

Create scheduled action

< 1 > 

<input checked="" type="checkbox"/>	Name	Start time	End time	Recurrence	Time zone	Desired capacity	Min	Max
<input checked="" type="checkbox"/>	GreatIndianSale	2025 October 30,...		30 10 21 10 *	Etc/UTC	8	5	15

## 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 [Amazon EC2 Auto Scaling](#) to manage and scale your EC2 capacity.



IP addresses

Step 4: Create Target Groups.

### Home-Target-Group

Target group name

Home-Target-Group

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

Laptop-Target-Group

Target group name

Laptop-Target-Group

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

Health checks

The associated load balancer periodically sends requests, per the settings below,

Health check protocol

HTTP ▼

Health check path

Use the default path of "/" to perform health checks on the root, or specify a custom path

/laptop/

Mobile-Target-Group

Target group name

Mobile-Target-Group

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

Health checks

The associated load balancer periodically sends requests, per the settings below, to the registered targets to test their status.

Health check protocol

HTTP ▼

Health check path

Use the default path of "/" to perform health checks on the root, or specify a custom path if preferred.

/mobile/

Up to 1024 characters allowed.

Target groups (3) Info

Filter target groups

<input type="checkbox"/>	Name	ARN	Port	Protocol	Target type	Load balancer	VPC ID
<input type="checkbox"/>	Mobile-Target-Group	arn:aws:elasticloadbalancing:us-east-1:123456789012:targetgroup/mobile-target-group/12345678901234567890123456789012	80	HTTP	Instance	None associated	vpc-0d8b1a2c
<input type="checkbox"/>	Laptop-Target-Group	arn:aws:elasticloadbalancing:us-east-1:123456789012:targetgroup/laptop-target-group/12345678901234567890123456789012	80	HTTP	Instance	None associated	vpc-0d8b1a2c
<input type="checkbox"/>	Home-Target-Group	arn:aws:elasticloadbalancing:us-east-1:123456789012:targetgroup/home-target-group/12345678901234567890123456789012	80	HTTP	Instance	None associated	vpc-0d8b1a2c

Step 5: Set Up Application Load Balancer (ALB).

aws

Search

[Alt+S]

EC2 > Load balancers > Compare and select load balancer type

A complete feature-by-feature comparison along with detailed highlights is also available. [Learn more](#)

Load balancer types

Application Load Balancer Info

Choose an Application Load Balancer when you need a flexible feature set for your

Network Load Balancer Info

Choose a Network Load Balancer when you need ultra-high performance, TLS offloading

Gateway Load

Choose a Gateway Load Balancer when you need to deploy and manage a fleet of target instances

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

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

- Serves internet-facing traffic.
- Has public IP addresses.
- DNS name resolves to public IPs.
- Requires a public subnet.

☐ Internal

- Serves internal traffic.
- Has private IP addresses.
- DNS name resolves to private IPs.
- Compatible with the IPv4 and Dualstack IP address types.

12 / 16

Listener rules (3) Info

Rule limits Actions Add rule

Traffic received by the listener is routed according to the default action and any additional rules. Rules are evaluated in priority order from the lowest value to the highest value.

Filter rules

	Priority	Name tag	Conditions (If)	Actions (Then)	ARN	Actions
<input type="checkbox"/>	1	Laptop-Rule	Path = /laptop/*	Forward to target group Laptop-Target-Group 1 (100%) Target group stickiness: Off		
<input type="checkbox"/>	2	Mobile-Rule	Path = /mobile/*	Forward to target group Mobile-Target-Group 1 (100%) Target group stickiness: Off		
<input type="checkbox"/>	Last (default)	Default	If no other rule applies	Forward to target group Home-Target-Group 1 (100%) Target group stickiness: Off		

Load balancers (1/1)

Actions Create load balancer

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

Filter load balancers

	Name	State	Type	Scheme	IP address type	VPC ID	Availability zones
<input checked="" type="checkbox"/>	ALB	Active	application	Internet-facing	IPv4	vpc-0d81f010b0ccfe3d1	us-east-1a (use1-az6) us-east-1b (use1-az7)

Load balancer: ALB

subnets

subnet-0074946e1e6cb0a0a us-east-1a (use1-az6)  
subnet-012dbf3ca55b1eabb us-east-1b (use1-az7)

Load balancer ARN

arn:aws:elasticloadbalancing:us-east-1:952406520450:loadbalancer/app/ALB/e4d5c4b1a8f4cf84

DNS name copied

ALB-674721536.us-east-1.elb.amazonaws.com (A Record)

Step 6: Attach ALB to ASGs.

Connect each Auto Scaling Group with its corresponding target group.

Load balancing - optional

Load balancers

☒ Application, Network or Gateway Load Balancer target groups  
Only instance target groups that belong to the same VPC as your Auto Scaling group are available for selection.

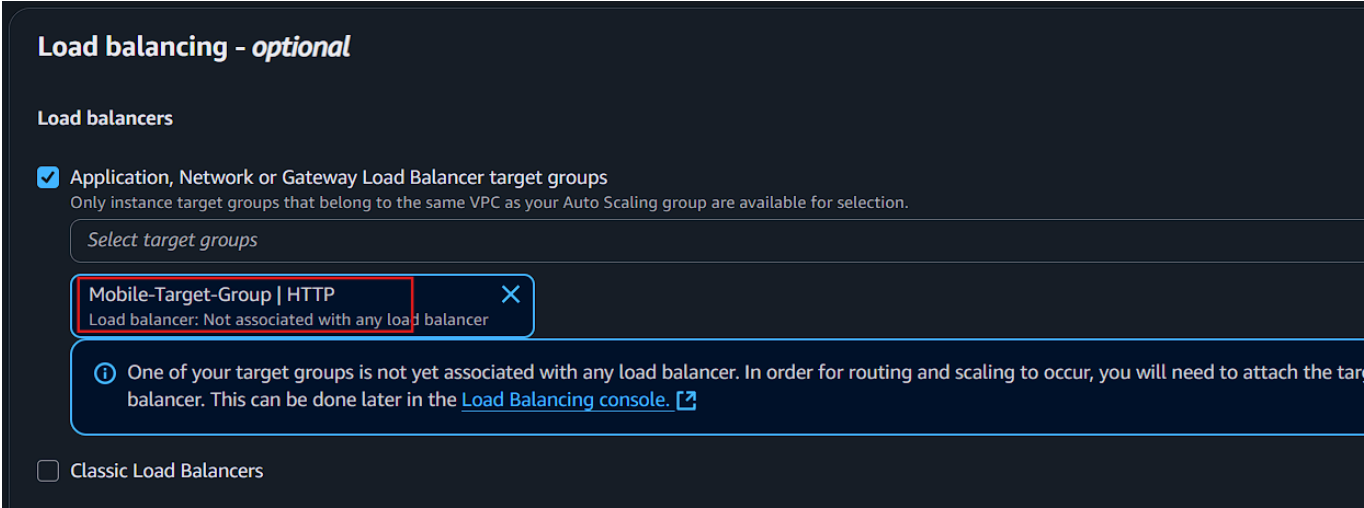
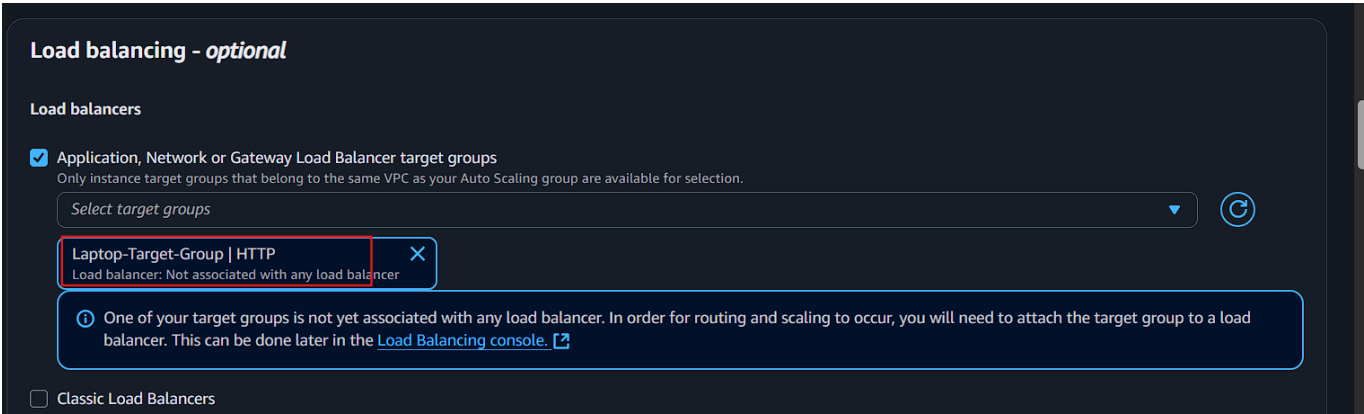
Select target groups

Home-Target-Group | HTTP  
Load balancer: Not associated with any load balancer

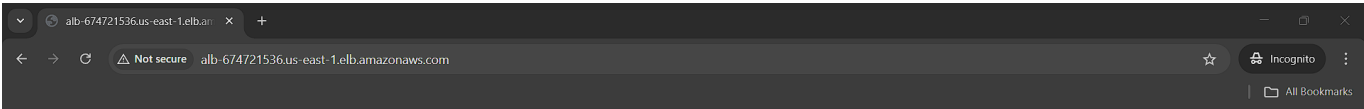
One of your target groups is not yet associated with any load balancer. In order for routing and scaling to occur, you will need to attach the target group to a load balancer. This can be done later in the Load Balancing console.

☐ Classic Load Balancers

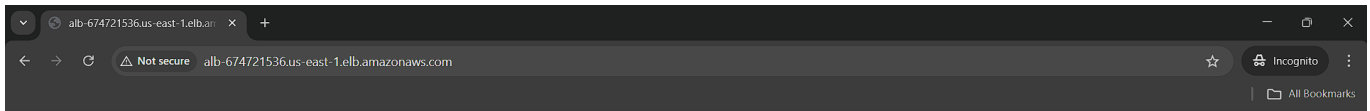
Create and attach new load balancers



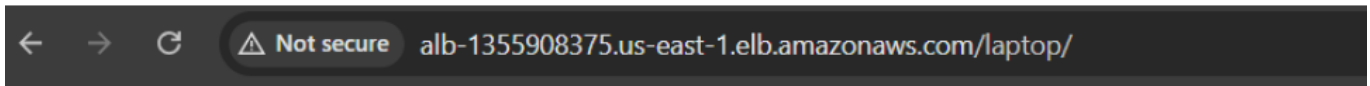
Step 7: Test and Verify.



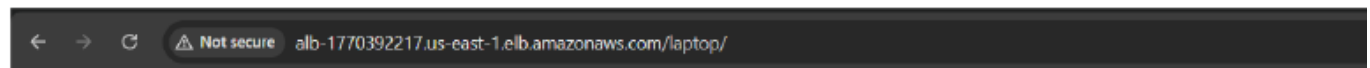
This is my Home Page ip-172-31-2-23.ec2.internal



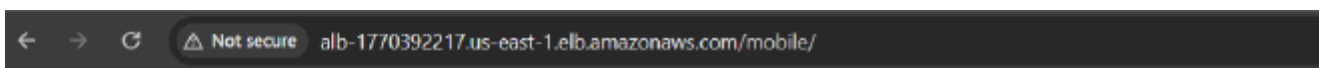
**This is my Home Page ip-172-31-81-153.ec2.internal**



**this is laptop page ip-172-31-27-238.ec2.internal**



**this is laptop page ip-172-31-24-241.ec2.internal**



**this is mobile page ip-172-31-20-81.ec2.internal**



## Conclusion

This setup provides a flexible and robust cloud architecture combining the power of **AWS Auto Scaling** and **Application Load Balancer**. By using different scaling strategies for various workloads, organizations can achieve both performance optimization and cost efficiency across all user categories.