

High Availability Deployment Using Auto Scaling Group and Load Balancer

Introduction

This project demonstrates the setup of an Auto Scaling Group (ASG) integrated with an Application Load Balancer (ALB) on AWS. The goal is to ensure high availability, scalability, and fault tolerance for deployed applications.

Architecture Overview

Application Load Balancer (ALB) distributes incoming traffic across healthy instances.

Target Group registers instances created by the ASG.

Launch Template/Launch Configuration defines EC2 instance settings (AMI, instance type, security group, key pair).

Auto Scaling Group (ASG) manages scaling policies based on demand (CPU utilization, request count, etc).

Prerequisites

- AWS Account
- IAM role with required permissions
- Amazon VPC with public/private subnets
- Key Pair for EC2 access
- Security Groups for ALB and EC2

Steps to Deploy

Step1 - Launch template

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

Template version description
This is my home server
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

Summary

Software Image (AMI)
ami-08982f1c5bf93d976

Virtual server type (instance type)
t3.micro

Firewall (security group)
launch-wizard-1

Storage (volumes)

Create launch template

- Attach Security Group (allow SSH/HTTP).

Network settings

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.

Availability Zone | Info
Don't include in launch template | Enable additional zones

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
launch-wizard-1 sg-08b7330192296eaa4 X | Compare security group rules

Summary

Software Image (AMI)
ami-08982f1c5bf93d976

Virtual server type (instance type)
t3.micro

Firewall (security group)
launch-wizard-1

Storage (volumes)

Create launch template

- Add User Data.

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

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

Summary

Software Image (AMI)
ami-08982f1c5bf93d976

Virtual server type (instance type)
t3.micro

Firewall (security group)
launch-wizard-1

Storage (volumes)

Create launch template

- Launch Three sets of template (Home,Laptop,Mobile).

The screenshot shows the AWS EC2 Launch Templates page. On the left, there's a sidebar with navigation links for EC2, Dashboard, EC2 Global View, Events, Instances, Launch Templates, Spot Requests, Savings Plans, Reserved Instances, Dedicated Hosts, Capacity Reservations, and Images. The main area is titled "Launch Templates (3)" and displays a table with three rows. The columns are: Launch Template ID, Launch Template Name, Default Version, Latest Version, and Create Time. The data is as follows:

Launch Template ID	Launch Template Name	Default Version	Latest Version	Create Time
lt-08ee145df5a0d0b26	Laptop-LT	1	1	2025-09-25
lt-050112792998baf7	Home-LT	1	1	2025-09-25
lt-0b5edb0ba813222b0	Mobile-LT	1	1	2025-09-25

Below the table, a section titled "Select a launch template" is visible.

Step2- Create Auto Scaling Groups (ASG)

a) Home ASG (Static Scaling)

- Launch configuration / Launch template with Amazon Linux 2 AMI.

The screenshot shows the "Create Auto Scaling group" wizard, Step 2: Choose instance launch options. The left sidebar lists steps from 2 to 7. The main area has two sections: "Name" and "Launch template".

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](#) [Switch to launch configuration](#)
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-LT
[Create a launch template](#)
Version
Default (1)

- Select Availability Zone and subnets

The screenshot shows the 'Create Auto Scaling group' wizard in the AWS EC2 console. The current step is 'Configure group size and scaling'. The 'VPC' section is selected, showing a dropdown menu with 'vpc-0db04a410f35ceaba' and '172.31.0.0/16 Default'. Below it is a 'Create a VPC' button. The 'Availability Zones and subnets' section lists two subnets: 'use1-az1 (us-east-1d) | subnet-0befa7ae9863b3f42' and '172.31.0.0/20 Default', and 'use1-az2 (us-east-1a) | subnet-02a6c1780f6782510' and '172.31.80.0/20 Default'. A 'Create a subnet' button is also present. The 'Availability Zone distribution - new' section contains two options: 'Balanced best effort' (selected) and 'Balanced only'. Both options include a note about launching instances across zones.

- Fixed capacity (e.g., Desired = 2, Min = 2, Max = 2).

The screenshot shows the 'Configure group size and scaling' step of the wizard. On the left, a sidebar lists steps: 'Step 5 - optional Add notifications', 'Step 6 - optional Add tags', 'Step 7', and 'Review'. The main area shows 'Desired capacity' set to 2. Below it is a 'Scaling' section with 'Min desired capacity' at 2 and 'Max desired capacity' at 2. The 'Automatic scaling - optional' section is partially visible.

- Create Auto Scaling Group.

- Review

The screenshot shows the AWS navigation bar with the 'Home-ASG' tab selected. Other tabs include 'Home-LT | Version Default', '2', '2', '2', and '2'. The selected tab is highlighted in blue.

b) Laptop ASG (Dynamic Scaling)

- Launch configuration / Launch template.

EC2 > Auto Scaling groups > Create 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 [Switch to launch configuration](#)

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](#) [Edit](#)

Version
 [Create a launch template version](#) [Edit](#)

Description **Launch template** **Instance type**

- Set Desired = 3, Min = 2, Max = 7.

EC2 > Auto Scaling groups > Create Auto Scaling group

Configure group size and 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.

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 <input type="text" value="2"/>	Max desired capacity <input type="text" value="7"/>
Equal or less than desired capacity	Equal or greater than desired capacity

- Add scaling policy:

EC2 > Auto Scaling groups > Create Auto Scaling group

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

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.

Target value

Instance warmup Info
 seconds

- Review



c) Mobile ASG (Scheduled Scaling)

- Launch configuration / Launch template.

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 [Switch to launch configuration](#)

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](#) [C](#)

Version
 [C](#) [Create a launch template version](#)

- Set Desired = 3, Min = 2, Max = 7.

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.

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 <input type="text" value="2"/>	Max desired capacity <input type="text" value="7"/>
Equal or less than desired capacity	Equal or greater than desired capacity

- Add schedule 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.

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

- Set Scheduled Action(Mobile)

- Select Mobile ASG.

Auto Scaling groups (1/3) Info

Last updated 15 minutes ago

Mobile-ASG Mobile-LT | Version Default 0 Updating capacity... 3

Laptop-ASG Laptop-LT | Version Default 3 - 3

Auto Scaling group: Mobile-ASG

Mobile-ASG Capacity overview

arn:aws:autoscaling:us-east-1:864456252120:autoScalingGroup:75fa2be0-9c63-4225-b55d-9da48567b131:autoScalingGroupName/Mobil

- Review

Auto Scaling groups (3) Info

Last updated less than a minute ago

Mobile-ASG Mobile-LT | Version Default 0 Updating capacity... 3 2 7

- Create Schedule Action

The screenshot shows the AWS Auto Scaling Groups page for the group 'Mobile-ASG'. On the left, there's a navigation sidebar with sections for Network & Security, Load Balancing, and Auto Scaling. Under Auto Scaling, 'Auto Scaling Groups' is selected. The main content area has a heading 'Predictive scaling policies use historical data to scale out your group ahead of forecasted hourly load.' Below it is a button 'Create predictive scaling policy'. The next section is titled 'Scheduled actions (0) Info' with a 'Create scheduled action' button. A table header for 'Scheduled actions' includes columns for Name, Start time, End time, Recurrence, Time zone, Desired capacity, Min, and Max. A message 'No scheduled actions are currently specified' is displayed.

- Set Desired = 8, Min = 5, Max = 15.

This screenshot shows the 'Create scheduled action' dialog box. It contains fields for 'Desired capacity' (set to 8), 'Min' (set to 5), and 'Max' (set to 15). The 'Recurrence' dropdown is set to 'Cron' with the value '0 10 21 10*'. The 'Time zone' dropdown is set to 'Etc/UTC'. Below these, there's a section for 'Specific start time' with fields for 'YYYY/MM/DD' (set to 2025/09/25) and '00:00' (set to Etc/UTC). Another section for 'End by' shows '2025/10/30' and '00:00' (both set to Etc/UTC). At the bottom right are 'Cancel' and 'Create' buttons.

The screenshot shows the same AWS Auto Scaling Groups page as before, but now it displays a single scheduled action. The 'Scheduled actions' section shows '1/1' and the details of the previously created action: 'Greatindia...', '2025 Oct...', '2025 Oct...', '0 10 21 10 *', 'Etc/UTC', '8', '5', and '15'. The message 'No predictive scaling policies have been created' is still present at the top.

- Review

The screenshot shows the AWS EC2 Auto Scaling Groups page. On the left, a sidebar lists 'Network & Security' (Security Groups, Elastic IPs, Placement Groups, Key Pairs, Network Interfaces), 'Load Balancing' (Load Balancers, Target Groups, Trust Stores), 'Auto Scaling' (Auto Scaling Groups), and 'Settings'. The main content area displays a message: 'No predictive scaling policies have been created. Predictive scaling policies use historical data to scale out your group ahead of forecasted hourly load.' Below this is a 'Create predictive scaling policy' button. A section titled 'Scheduled actions (1/1)' shows one entry: 'Greatindia...' with details: Start time: 2025 Oct..., End time: 2025 Oct..., Recurrence: 0 10 21 10 * Etc/UTC, Time zone: 8, Desired capacity: 5, Min: 5, Max: 15. There are 'Actions' and 'Create scheduled action' buttons.

- Launch Three sets of Auto Scaling Groups (Home,Laptop,Mobile).

The screenshot shows the AWS EC2 Auto Scaling groups list page. It displays three Auto Scaling groups: 'Mobile-ASG', 'Laptop-ASG', and 'Home-ASG'. Each group is associated with a launch template ('Mobile-LT', 'Laptop-LT', 'Home-LT') and has a status of 'Updating capacity...'. The desired capacity for all groups is set to 3, with minimum and maximum values of 2. The page includes a search bar, a table header with columns for Name, Launch template/configuration, Instances, Status, Desired capacity, Min, and Max, and a 'Create Auto Scaling group' button.

Step3 - Create Target Groups

- Create Target group (Home).

The screenshot shows the 'Create target group' wizard. It starts with a note: 'Facilitates routing to a single Lambda function. Accessible to Application Load Balancers only.' Then it asks 'Application Load Balancer' or 'AWS Lambda'. The 'Application Load Balancer' option is selected, with a note: 'Offers the flexibility for a Network Load Balancer to accept and route TCP requests within a specific VPC. Facilitates using static IP addresses and PrivateLink with an Application Load Balancer.' The form fields include 'Target group name' (Home-TG), 'Protocol' (HTTP), 'Port' (80), and 'IP address type' (Only targets with the indicated IP address type can be registered to this target group). At the bottom, there are links for CloudShell, Feedback, Copyright (© 2025, Amazon Web Services, Inc. or its affiliates.), Privacy, Terms, and Cookie preferences.

- Create Target Group(Laptop).

The screenshot shows the 'Create target group' wizard for an Application Load Balancer. The 'Protocol' dropdown is set to 'HTTP'. The 'Port' field contains '80'. The 'IP address type' dropdown is set to 'IPv4'. The 'Target group name' input field contains 'Laptop-TG'. A note states: 'A maximum of 32 alphanumeric characters including hyphens are allowed, but the name must not begin or end with a hyphen.' A note also specifies: 'Accessible to Application Load Balancers only.'

- Set health Checks.

The screenshot shows the continuation of the 'Create target group' wizard. Under 'Health check protocol', 'HTTP' is selected. Under 'Health check path', the value '/Laptop/' is entered. A note says: 'Up to 1024 characters allowed.' A link 'Advanced health check settings' is visible. A note in the 'Attributes' section states: 'Certain default attributes will be applied to your target group. You can view and edit them after creating the target group.'

- Create Target Group(Mobile).

Accessible to Application Load Balancers only.

Application Load Balancer

- Offers the flexibility for a Network Load Balancer to accept and route TCP requests within a specific VPC.
- Facilitates using static IP addresses and PrivateLink with an Application Load Balancer.

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

Protocol
Protocol for load balancer-to-target communication. Can't be modified after creation.
HTTP

Port
Port number where targets receive traffic. Can be overridden for individual targets during registration.
80
1-65535

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

- verify three Target Groups(Home,Laptop,Mobile).

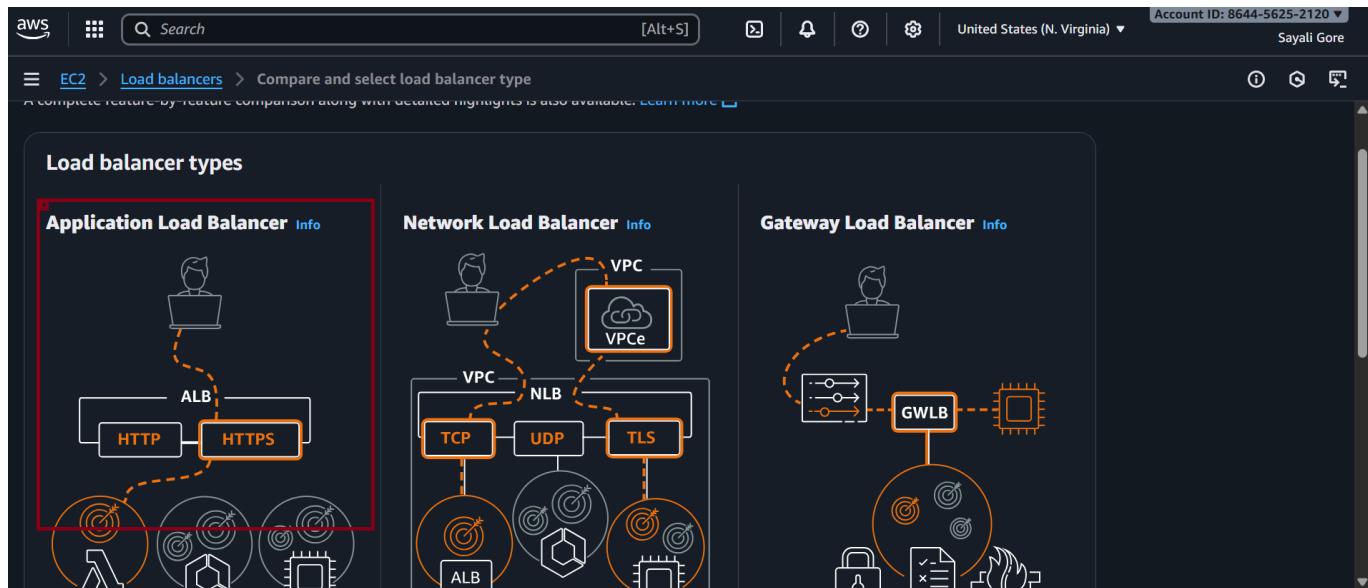
Target groups (3) <small>Info</small>						
Actions <small>Actions</small> Create target group						
Filter target groups						
	Name	ARN	Port	Protocol	Target type	Lo
<input type="checkbox"/>	Mobile-TG	arn:aws:elasticloadbalancin...	80	HTTP	Instance	<small>i</small>
<input type="checkbox"/>	Laptop-TG	arn:aws:elasticloadbalancin...	80	HTTP	Instance	<small>i</small>
<input type="checkbox"/>	Home-TG9	arn:aws:elasticloadbalancin...	80	HTTP	Instance	<small>i</small>

Step4 - Create Application Load Balancer (ALB)

- Load Balancers → Create Load Balancer.

Load balancers						
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
0 load balancers selected						
Select a load balancer above.						

- Choose Application Load Balancer.



Configure:

- Listeners: HTTP on port 80

The screenshot shows the 'Create Application Load Balancer' configuration page. Under 'Security groups', it lists two selected security groups: 'default' (sg-02441e6338684683d) and 'launch-wizard-1' (sg-08b7330192296eaa4). Under 'Listeners and routing', it shows a single listener configuration for 'Listener HTTP:80' with 'Protocol' set to 'HTTP' and 'Port' set to '80'.

- Under Listeners and Routing:

Forward /home path → Home-TG

Forward /laptop path → Laptop-TG

Forward /mobile path → Mobile-TG

The screenshot shows the AWS Lambda console with the path: EC2 > Load balancers > ALB > HTTP:80 listener. On the left, there's a sidebar with sections like Network & Security, Load Balancing, Auto Scaling, and Settings. The main area is titled "Listener rules (3) Info". It displays three rules:

- Rule 1:** Priority 1, Name tag Laptop, Path = [/Laptop/*]. Action: Forward to target group Laptop-TG [2]: 1 (100%). Target group stickiness: O.
- Rule 2:** Priority 2, Name tag Mobile-rule, Path = [/Mobile/*]. Action: Forward to target group Mobile-TG [2]: 1 (100%). Target group stickiness: O.
- Default Rule:** Last (default), Name tag Default, Condition: If no other rule applies. Action: Forward to target group Home-TG9 [2]: 1 (100%). Target group stickiness: O.

✓ Verification

- Get ALB DNS name from Load Balancer.

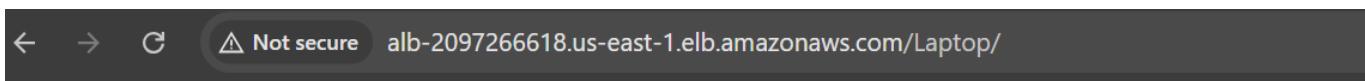
The screenshot shows the AWS Lambda console with the path: EC2 > Load balancers. The sidebar includes sections like Network & Security, Load Balancing, Auto Scaling, and Settings. The main area is titled "Load balancers (1/1)". It shows one active application load balancer named "ALB". Below it, under "Load balancer: ALB", the "Load balancer ARN" is listed as arn:aws:elasticloadbalancing:us-east-1:864456252120:loadbalancer/app/ALB/63f98402f563b4c1. A tooltip indicates that the DNS name has been copied.

- Test in browser:
- http://home → Routes to Home ASG instances.

A screenshot of a web browser window. The address bar shows the URL: alb-2097266618.us-east-1.elb.amazonaws.com. A tooltip indicates that the connection is "Not secure".

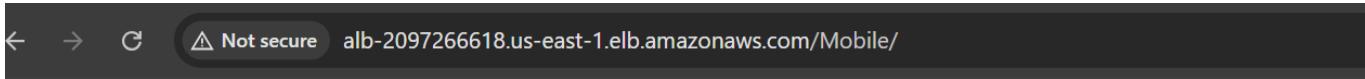
This is Home Page ip-172-31-82-32.ec2.internal

- http://laptop → Routes to Laptop ASG instances.



This is Laptop page ip-172-31-83-22.ec2.internal

- http://mobile → Routes to Mobile ASG instances.



This is Mobile page ip-172-31-10-5.ec2.internal

Summary

ALB distributes traffic to different Target Groups.

Home ASG → Static scaling.

Laptop ASG → Dynamic scaling.

Mobile ASG → Scheduled scaling.

This setup ensures high availability, cost optimization, and performance across different application services.