

## Scalable Web Application using NLB + EC2 Auto Scaling

## 1. Create Launch Template (EC2)

Search results

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

myec2it

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

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.

Application and OS Images (Amazon Machine Image)

An AMI contains the operating system, application server, and applications for your instance. If you don't see a suitable AMI below, use the search field or choose Browse more AMIs.

Search our full catalog including 1000s of application and OS images

Summary

Software Image (AMI)

Amazon Linux 2023 kernel-6.1 A...read more

ami-05071c56x32875a8

Virtual server type (instance type)

t2.micro

Firewall (security group)

-

Storage (volumes)

1 volume(s) - 8 GiB

Free tier: In your first year of opening an AWS account, you get 750 hours per month of t2.micro instance usage (or t3.micro where t2.micro isn't available) when used with free tier AMIs, 750 hours per month of public IPv4 address usage, 30 GiB of EBS storage, 2 million I/Os, 1 GiB of snapshots, and 100 GiB of bandwidth to the internet. Data transfer charges are not included as part of the free tier allowance.

Cancel

Create launch template

Search results

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

mydebian13

Create new key pair

Network settings

Subnet

Don't include in launch template

Create new subnet

Availability Zone

Don't include in launch template

Enable additional zones

Firewall (security groups)

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

Select security groups

myec2securitygroup sg-00ae2b2e28f558b23

Compare security group rules

Advanced network configuration

Summary

Software Image (AMI)

Amazon Linux 2023 kernel-6.1 A...read more

ami-05071c56x32875a8

Virtual server type (instance type)

t2.micro

Firewall (security group)

myec2securitygroup

Storage (volumes)

1 volume(s) - 8 GiB

Free tier: In your first year of opening an AWS account, you get 750 hours per month of t2.micro instance usage (or t3.micro where t2.micro isn't available) when used with free tier AMIs, 750 hours per month of public IPv4 address usage, 30 GiB of EBS storage, 2 million I/Os, 1 GiB of snapshots, and 100 GiB of bandwidth to the internet. Data transfer charges are not included as part of the free tier allowance.

Cancel

Create launch template

Search results

Metadata response hop limit

Don't include in launch template

Allow tags in metadata

Don't include in launch template

User data - optional

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

Choose file

#!/bin/bash  
yum update -y  
amazon-linux-extras install nginx1 -y  
systemctl start nginx  
systemctl enable nginx  
echo "Served by \$(hostname) via NLB" > /usr/share/nginx/html/index.html

User data has already been base64 encoded

Summary

Software Image (AMI)

Amazon Linux 2023 kernel-6.1 A...read more

ami-05071c56x32875a8

Virtual server type (instance type)

t2.micro

Firewall (security group)

myec2securitygroup

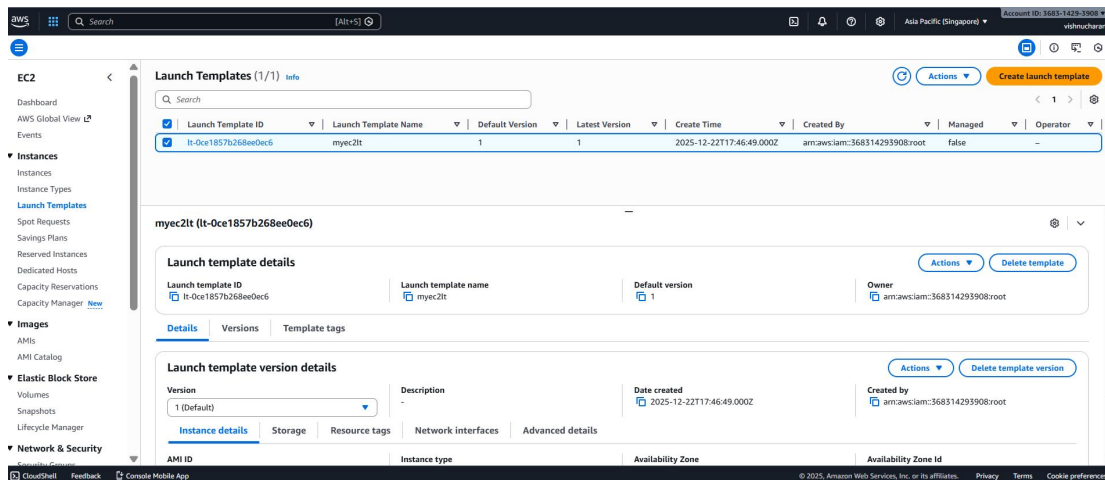
Storage (volumes)

1 volume(s) - 8 GiB

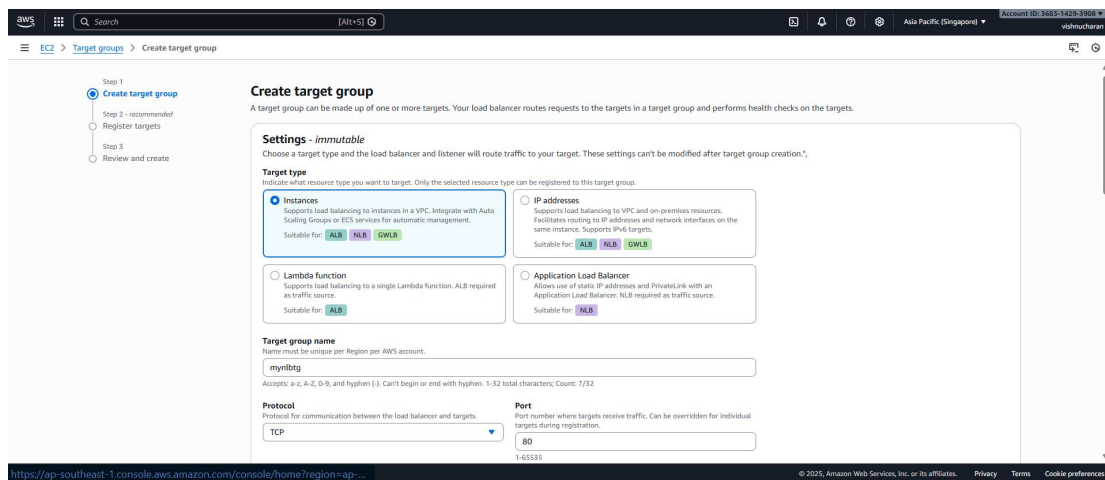
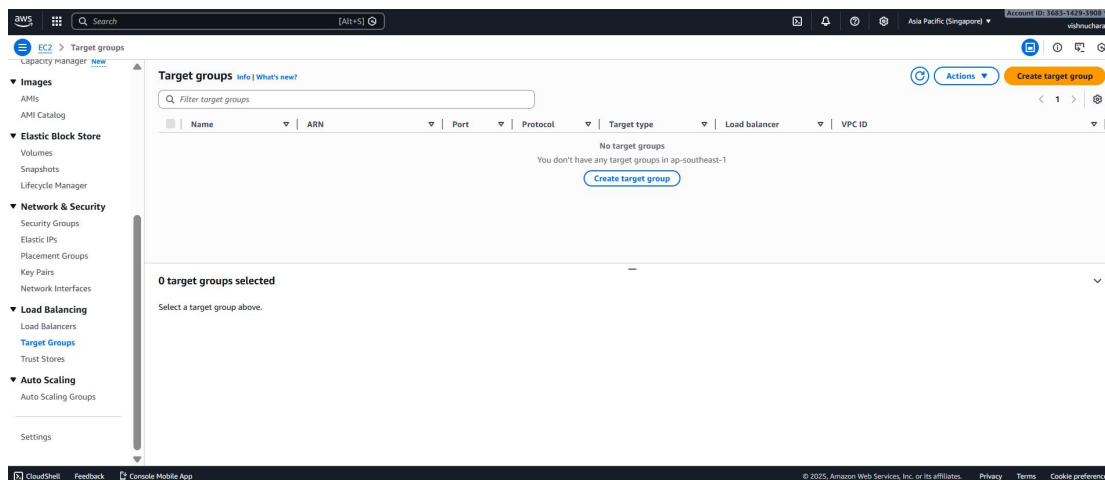
Free tier: In your first year of opening an AWS account, you get 750 hours per month of t2.micro instance usage (or t3.micro where t2.micro isn't available) when used with free tier AMIs, 750 hours per month of public IPv4 address usage, 30 GiB of EBS storage, 2 million I/Os, 1 GiB of snapshots, and 100 GiB of bandwidth to the internet. Data transfer charges are not included as part of the free tier allowance.

Cancel

Create launch template



## 2. Create Target Group (Critical for NLB)



**Create target group**

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.

vpc-050738e61b42c061b  
172.31.0.0/16 (default) [Create VPC](#)

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

**Health check protocol**  
TCP

**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 is the same as the target group's traffic port. However, you can specify a different port as an override.

☒ Traffic port  
☐ Override

**Healthy threshold**  
The number of consecutive health check 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

---

**mynlbtg** [Actions](#)

**Details**  
arn:aws:elasticloadbalancing:ap-southeast-1:368314293908:targetgroup/mynlbtg/71808b2a2c8424ce

**Target type**  
Instance

**Protocol : Port**  
TCP: 80

**VPC**  
[vpc-050738e61b42c061b](#)

**IP address type**  
IPv4

**Load balancer**  
[None associated](#)

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

**Targets** | **Monitoring** | **Health checks** | **Attributes** | **Tags**

**Registered targets (0)** [Deregister](#) [Register targets](#)

Instance ID	Name	Port	Zone	Health status	Health status details	Administrative ...	Override details	Launch time
No registered targets You have not registered targets to this group yet								

[Register targets](#)

### 3. Create Network Load Balancer

**Compare and select load balancer type**

**Application Load Balancer** [info](#)

Choose an Application Load Balancer when you need a flexible feature set for your applications with HTTP and HTTPS traffic. Operating at the request level, Application Load Balancers provide advanced routing and visibility features targeted at application architectures, including microservices and containers.

[Create](#)

**Network Load Balancer** [info](#)

Choose a Network Load Balancer when you need ultra-high performance, TLS offloading at scale, centralized certificate deployment, support for UDP, and static IP addresses for your applications. Operating at the connection level, Network Load Balancers are capable of handling millions of requests per second securely while maintaining ultra-low latencies.

[Create](#)

**Gateway Load Balancer** [info](#)

Choose a Gateway Load Balancer when you need to deploy and manage a fleet of third-party virtual appliances that support GENIEVE. These appliances enable you to improve security, compliance, and policy controls.

[Create](#)

**Classic Load Balancer - previous generation**

Search

[Alt+S]

Asia Pacific (Singapore)

Account ID: 3683-1429-3908

vishnucharan

EC2

Load balancers

Create Network Load Balancer

Basic configuration

Load balancer name

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

mymlb

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

Scheme

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.

Load balancer IP address type

Select the front-end IP address type to assign to the load balancer. The VPC and subnets mapped to this load balancer must include the selected IP address types.

☒ IPv4

- Includes only IPv4 addresses.

☐ Dualstack

- Includes IPv4 and IPv6 addresses.

Network mapping

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

VPC

The load balancer will exist and scale within the selected VPC. The selected VPC is also where the load balancer targets must be hosted unless routing to on-premises targets or if using VPC peering. To confirm the VPC for your targets, view [target groups](#).

vpc-05073fe61b42c061b

172.31.0.0/16

(default)

Create VPC

Availability Zones and subnets

CloudShell

Feedback

Console Mobile App

© 2024 Amazon Web Services, Inc. or its affiliates

Privacy

Terms

Cookie preferences

EC2

Load balancers

Create Network Load Balancer

Network mapping

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

VPC

The load balancer will exist and scale within the selected VPC. The selected VPC is also where the load balancer targets must be hosted unless routing to on-premises targets or if using VPC peering. To confirm the VPC for your targets, view [target groups](#).

vpc-05073fe61b42c061b

172.31.0.0/16

(default)

Create VPC

Availability Zones and subnets

Select one or more Availability Zones and corresponding subnets. Enabling multiple Availability Zones increases the fault tolerance of your applications. The load balancer routes traffic to targets in the selected Availability Zones only. Availability Zones that are not supported by the load balancer or the VPC are not available for selection.

☒ ap-southeast-1a (apse1-az1)

Subnet

- Only CIDR blocks corresponding to the load balancer IP address type are used. At least 8 available IP addresses are required for your load balancer to scale efficiently.

subnet-04bf478a0e7c6ded7

- IPv4 subnet CIDR: 172.31.16.0/20

☒ ap-southeast-1b (apse1-az2)

Subnet

- Only CIDR blocks corresponding to the load balancer IP address type are used. At least 8 available IP addresses are required for your load balancer to scale efficiently.

subnet-06b806dabcbee64d7

- IPv4 subnet CIDR: 172.31.32.0/20

☐ ap-southeast-1c (apse1-az3)

CloudShell

Feedback

Console Mobile App

© 2024 Amazon Web Services, Inc. or its affiliates

Privacy

Terms

Cookie preferences

EC2

Load balancers

Create Network Load Balancer

Security groups

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 - recommended

Security groups support on Network Load Balancers can only be enabled at creation by including at least one security group. You can change security groups after creation. The security groups for your load balancer must allow it to communicate with registered targets on both the listener port and the health check port. For PrivateLink Network Load Balancers, security group rules are enforced on PrivateLink traffic; however, you can turn off inbound rule evaluation after creation within the load balancer's security tab or using the API.

Select up to 5 security groups

myelbsecuritygroup

vpc-05073fe61b42c061b

Listeners and routing

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: TCP/80

Protocol

TCP

Port

80

1-65535

Remove

Forward to target group

Choose a target group and specify routing weight or [create target group](#).

Target group

myelb

TCP

Weight

1

Percent

100%

0-999

Add target group

You can add up to 4 more target groups.

Target group stickiness

Enables the load balancer to bind a user's session to a specific target group. If you want to bind a user's session to a specific target, turn on the Target group attribute stickiness.

☐ Turn on target group stickiness

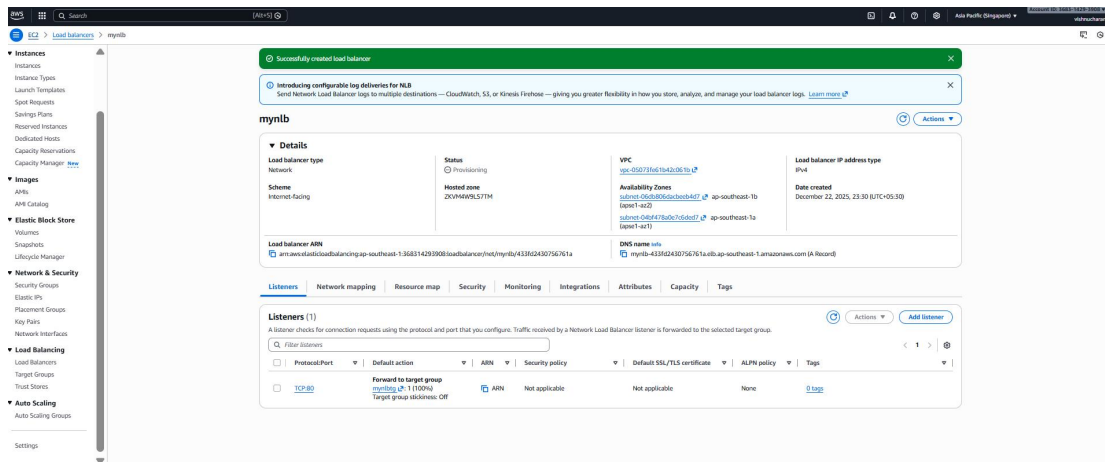
Weighted routing evaluation - recommended

You Network Load Balancer's ability to distribute requests according to assigned target group weights can be honed if your target groups don't follow best practices.

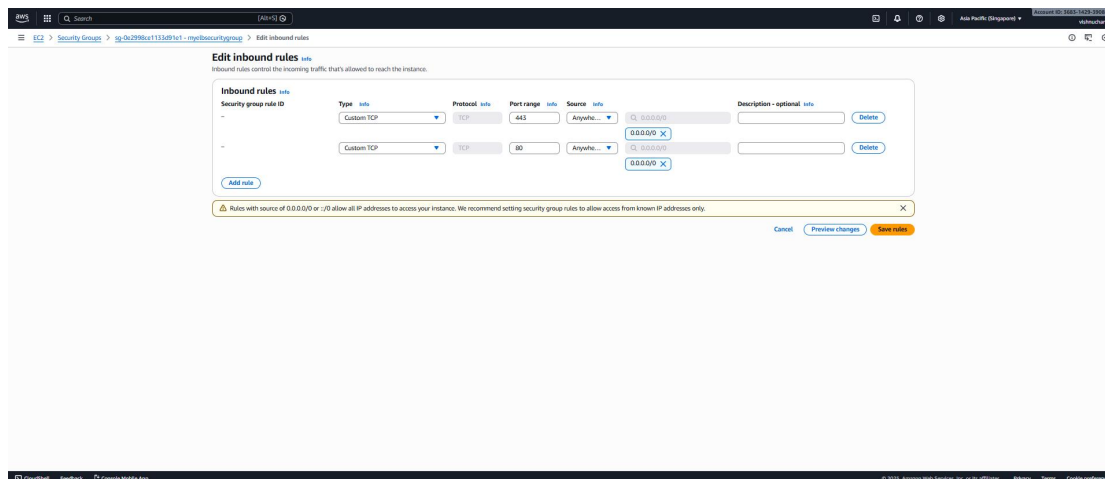
Evaluate

Listener tags - optional

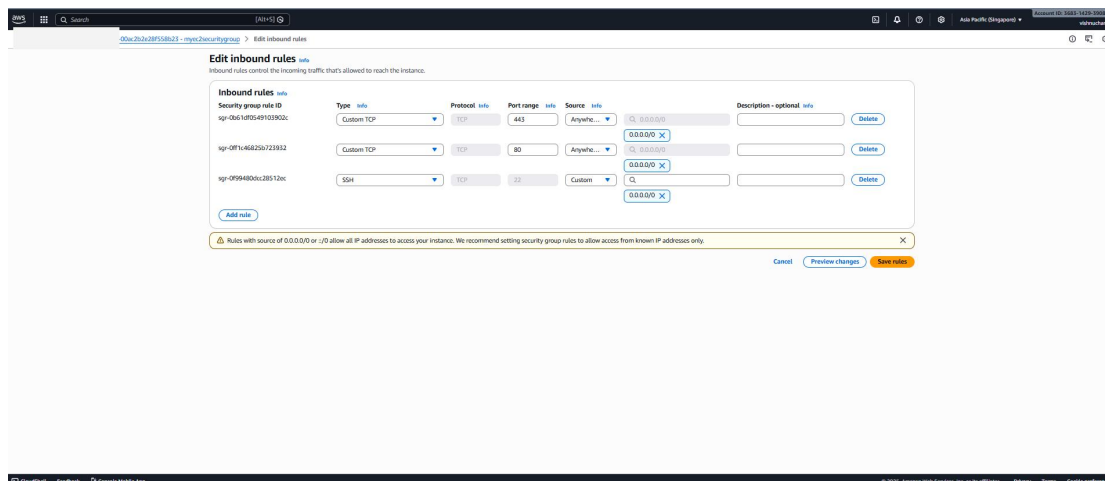
Consider adding tags to your listener. Tags enable you to categorize your AWS resources so you can more easily manage them.



## NLB Security group(Inbound rules):-



## EC2 Security group(Inbound rules):-



## 4. Create Auto Scaling Group

Step 1

Choose launch template or configuration

Step 2

Choose instance launch options

Step 3 - optional

Integrate with other services

Step 4 - optional

Configure group size and scaling

Step 5 - optional

Add notifications

Step 6 - optional

Add tags

Step 7

Review

### Choose launch template or configuration [info](#)

Specify a launch template that contains settings common to all EC2 instances that are launched by this Auto Scaling group. If you currently use launch configurations, you might consider migrating to launch templates.

**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](#)

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

Launch template

[Create a launch template](#)

Version

[Create a launch template version](#)

Description

AMI ID

Key pair name

Launch template

[Switch to launch configuration](#)

Instance type

Security groups

Security group IDs

Request Spot Instances

Additional details

Storage (volumes)

Date created

Step 1

Choose launch template or configuration

Step 2

Choose instance launch options

Step 3 - optional

Integrate with other services

Step 4 - optional

Configure group size and scaling

Step 5 - optional

Add notifications

Step 6 - optional

Add tags

Step 7

Review

### 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](#)

Launch template

Version

Description

**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

[Create a VPC](#)

Availability Zones and subnets

[Create a subnet](#)

apn1-az1 (ap-southeast-1) | subnet-04b478a0c7d5d0e77

[Create a subnet](#)

apn1-az2 (ap-southeast-1) | subnet-06c9d8056a0eeb4d7

[Create a subnet](#)

**Availability Zone distribution - new**

Auto Scaling automatically balances instances across Availability Zones, if launch failures occur in a zone, select a strategy.

Balanced load effort

☒

Auto Scaling will attempt to search in another healthy Availability Zone.

Balanced only

☐

If a specific fail in one Availability Zone, Auto Scaling will continue to attempt to search in the unhealthy Availability Zone to preserve balanced distribution.

Step 1

Choose launch template or configuration

Step 2

Choose instance launch options

Step 3 - optional

Integrate with other services

Step 4 - optional

Configure group size and scaling

Step 5 - optional

Add notifications

Step 6 - optional

Add tags

Step 7

Review

### Integrate with other services - optional [info](#)

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 arn shift. You can also customize 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.

Select Load balancing options

☐ No load balancer

☒ Attach to an existing load balancer

☐ Attach to a new load balancer

Choose from your existing load balancers.

Choose from Classic Load Balancers

**Attach to an existing load balancer**

Select the load balancers to attach

Choose from your load balancer target groups

Existing load balancer target groups

**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

☐ Attach to VPC Lattice service

Create new VPC Lattice service

**Application Recovery Controller (ARC) zonal shift - new** [info](#)

During an Availability Zone impairment, target instance launches towards other healthy Availability Zones.

Step 1: Choose launch template or configuration

Step 2: Choose instance launch options

Step 3 - optional: Integrate with other services

Step 4 - optional: Configure group size and scaling

Step 5 - optional: Add notifications

Step 6 - optional: Add tags

Step 7: Review

### 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 MemoryGiB are only supported for mixed instance groups configured with a mix of instance attributes.

100% (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

1

2

Equal or less than desired capacity

Equal or greater than desired capacity

#### Automatic scaling - optional [info](#)

Choose whether to use a target tracking policy [info](#).

Amazon set up an [Amazon CloudWatch](#) target tracking policy 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 will not dynamically resize to meet demand.

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 [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.

#### Choose a replacement behavior depending on your availability requirements

☒ Replaced instance

☐ Preserve availability

☐ Continue with

☐ Preserve

☒ No policy

Launch before terminating

Terminate and launch

Custom behavior

Auto Scaling groups (1/1) [info](#)

Search your Auto Scaling groups

Last updated less than a minute ago

[Launch configurations](#)

[Launch templates](#)

[Actions](#)

[Create Auto Scaling group](#)

Name	Launch template/configuration	Instances	Status	Desired capacity	Min	Max	Availability Zones	Creation time
mytmbag	mytmbag1   Version Latest	2	-	2	1	3	2 Availability Zones	Mon Dec 22 2025 23:41:16 GMT+0530 (India Standard Time)

Auto Scaling group: mytmbag

Details

Integrations

Automatic scaling

Instance management

Instance refresh

Activity

Monitoring

Tags - moved

#### mytmbag Capacity overview

Desired capacity

Scaling limits

Desired capacity type

Status

Date created

Mon Dec 22 2025 23:41:16 GMT+0530 (India Standard Time)

#### Launch template

Launch template

mytmbag1 | Version Latest

AMI ID

ami-05a0716c4e52875a8

Instance type

t2.micro

Owner

amazon:arn:346314293908:root

Security groups

sg-0963b263a29558b29

Security group ID

sg-0963b263a29558b29

Create time

Mon Dec 22 2025 23:16:49 GMT+0530 (India Standard Time)

Description

-

Key pair name

mytmbag13

Request Spot instances

No

Instances (2) [info](#)

Find instance by attribute or tag (case sensitive)

All status

[Joining](#) [Clear filters](#)

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 ...	Elastic IP	IPv6 IP	Monitoring	Security group name	Key name
	i-0504a77b2d65c3b1	Running	t2.micro	Initializing	View alarms	ap-southeast-1a	ec2-10-212-144-114.ap...	13.12.144.114	-	-	disabled	mytmbag2securitygroup	mytmbag13
	i-04a055a39ac7389d	Running	t2.micro	Initializing	View alarms	ap-southeast-1a	ec2-3-9-1-222-40.ap-sou...	3.1.222.40	-	-	disabled	mytmbag2securitygroup	mytmbag13

Select an instance