

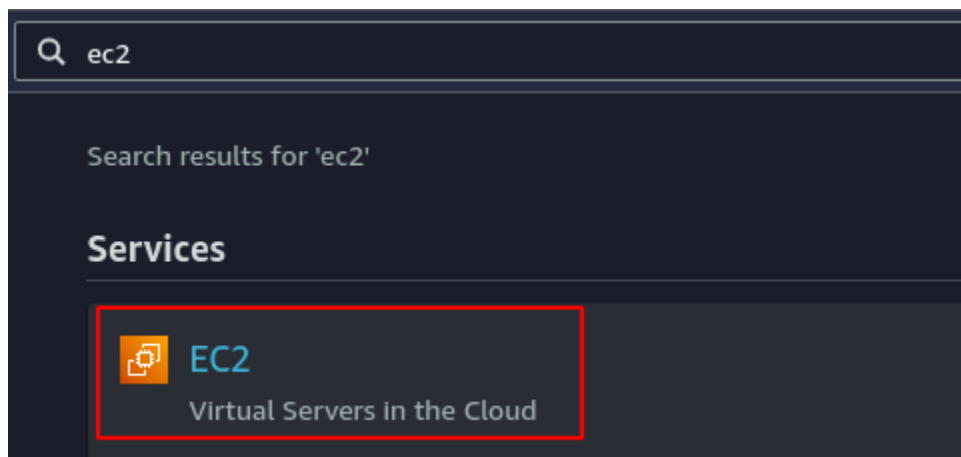
## LOAD BALANCER IN AWS (APPLICATION LOAD BALANCER)

how an application load balancer can be created and configured using an AWS management console.

### Creating target groups

Application load balancer receives traffic and forwards the traffic to the target groups. These target groups are the groups of the targets like EC2 instances in multiple availability zones.

This section will create a target group and then register the EC2 instance to the target group. First, log into the AWS management console and go to the EC2 services.



From the left side panel, go to the **Target Groups** section.

▼ Load Balancing

Load Balancers

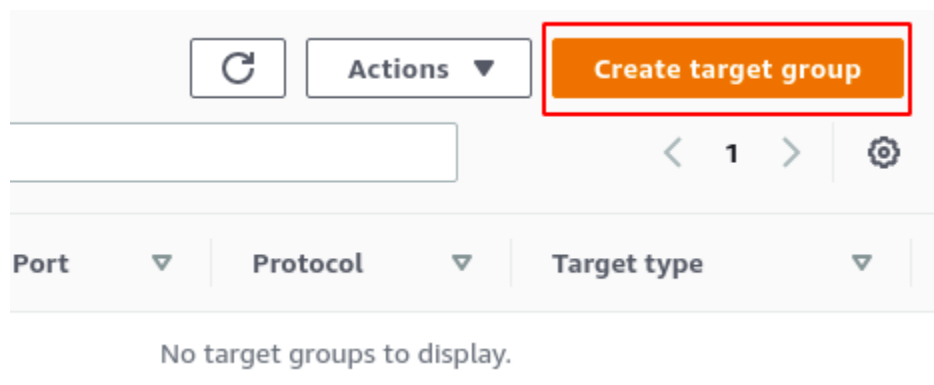
**Target Groups** New

▼ Auto Scaling

Launch Configurations

Auto Scaling Groups

From the top right corner of the console, click on the **create target group button** to create a new target group.



Now it will ask for the **target type** you want to register in this target group. For this demo, select the **Instances** as target types for this target group.

### Choose a target type

☒ **Instances**

- Supports load balancing to instances within a specific VPC.

☐ **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.

☐ **Lambda function**

- Facilitates routing to a single Lambda function.
- 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.

Now it will ask for the **target group name, port, protocol, HTTP Version, and VPC**. The target group name is a unique identifier that identifies the target group in a VPC.

The port and protocol are the port number and protocol on which the target group will listen for incoming requests from the application load balancer. For the target groups that will receive traffic from ALB, the protocol must be HTTP or HTTPS.

The VPC is the virtual private cloud in which instances are running that will be registered with the target group.

Target group name

demo-tg

A maximum of 32 alphanumeric characters including hyphens are allowed, but the name must not begin or end w

Protocol

HTTP

Port

80

VPC

Select the VPC with the instances that you want to include in the target group.

vpc-  
IPv4: 172.31.0.0/16

Protocol version

☒ HTTP1

Send requests to targets using HTTP/1.1. Supported when the request protocol is HTTP/1.1 or HTTP/2.

☐ HTTP2

Send requests to targets using HTTP/2. Supported when the request protocol is HTTP/2 or gRPC, but gRPC-specific features are not available.

☐ gRPC

Send requests to targets using gRPC. Supported when the request protocol is gRPC.

Now keep the health check settings as default and click on the **Next** button to register the targets for the target group.

It will display all the instances in the VPC, and you can register any number of instances from here. For this demo, we will register only one EC2 instance in this target group. Select the instance you want to register and enter the port on which the server is running inside the instance. Click on the **include as pending below button** to register the target in the target group.

<input checked="" type="checkbox"/>	Instance ID ▾	Name ▾	State ▾	Security groups
<input checked="" type="checkbox"/>	I- [REDACTED]	demo-server	✓ running	CustomSG

1 selected

Ports for the selected instances  
Ports for routing traffic to the selected instances.

1-65535 (separate multiple ports with commas)

After registering the target, click on the **create target group button** to create the target group with EC2 instance registered.

## Creating application load balancer

After creating the target group for the **load balancer**, now go to the Load balancers from the left side panel.

▼ Load Balancing

Load Balancers

Target Groups New

▼ Auto Scaling

Launch Configurations

Auto Scaling Groups

Click on the **create load balancer button**.

Create Load Balancer

Actions ▾

Filter by tags and attributes or search by keyword

☐

Name

▲

DNS name

It will open a new web page to select the load balancer type to create. Select the **application load balancer** and click on the **create** button.

Load balancer types

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 GENEVE. These appliances enable you to improve security, compliance, and policy controls.

Create

Now it will ask for the basic configuration of the application load balancer. Enter the **Load balancer name, scheme, and IP address type**. Load balancer name is a unique identifier for the application load balancer to be created.

Load balancer scheme defines whether load balancer will be internal or internet-facing. Internet-facing load balancers can accept connection requests from the public internet and route these requests to the target groups internally. On the other hand, internal load balancers do not have a publicly resolvable DNS name. They can only be accessed within the VPC and route the requests to the target groups internally.

The IP address type defines whether the end-users can send requests using **IPv4 or IPv6 addresses**. Setting the IP address type to IPv4 will allow the end-users to send the requests from IPv4 only. For internet-facing load balancers, it is recommended to use the **dualstack** IP address type.

For this demo, we will create an internet-facing application load balancer with **dualstack** IP address type.

### Basic configuration

**Load balancer name**

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

demo-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 cannot 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.

Now for networking, select the VPC, availability zones, and subnets from

availability zones. The VPC must be the same as selected while creating the target group.

For availability zones, AWS recommends choosing at least two availability zones with at least one public subnet in it to configure for the internet-facing application load balancer.

For this demo, select 3 availability zones of default VPC. Availability zones of default VPC have public subnets to configure it with the internet-facing application load balancer.

**VPC** [Info](#)  
Select the virtual private cloud (VPC) for your targets. Only VPCs with an internet gateway are enabled for selection. The selected VPC cannot be changed. To confirm the VPC for your targets, view your [target groups](#).

-  
vpc-[REDACTED]  
IPv4: 172.31.0.0/16

▼

**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 available for selection. Subnets cannot be removed after the load balancer is created, but additional subnets can be added.

☒ **us-east-1a**

Subnet

subnet-[REDACTED]

▼

For the security group, select a security group from the VPC that will control inbound and outbound traffic from the application load balancer.



## Security groups [Info](#)

A security group is a set of firewall rules that control the traffic to your load balancer.

### Security groups

Select security groups

[Create new security group](#)

default sg- [redacted] X  
VPC: vpc- [redacted]

After configuring VPC and security group, now configure the default listener for the application load balancer.

For this demo, we will configure port 80 of the application load balancer to receive traffic from the end-users and forward the traffic to the **demo-tg target group**.

The **demo-tg target group** is configured to equally pass the traffic to port 80 of the EC2 instances inside it.

### ▼ Listener HTTP:80

Protocol

HTTP ▼

Port

: 80  
1-65535

Default action [Info](#)

Forward to

demo-tg

Target type: Instance, IPv4

[Create target group](#)

After configuring the load balancer, now have a look at the configuration summary and click on the create load balancer to **create the load balancer**.

## Summary

Review and confirm your configurations. [Estimate cost](#)

### Basic configuration [Edit](#)

demo-alb

- Internet-facing
- Dualstack

### Security groups [Edit](#)

- default  
[sg-](#)

### Network mapping [Edit](#)

VPC [vpc-](#)

- us-east-1a  
[subnet-](#)
- us-east-1b  
[subnet-](#)
- us-east-1c  
[subnet-](#)

### Listeners and routing [Edit](#)

- HTTP:80 defaults to  
[demo-tg](#)

[Tags](#) [Edit](#)

After creating the load balancer, make sure of the following points.

- Security group attached to the load balancer have an inbound rule to allow traffic on the load balancer
- Security group attached to the EC2 instances have an inbound rule to allow traffic from the load balancer
- The port on which the application is running in the EC2 instance is configured properly in the target group
- EC2 instance and application load balancer must be in the same VPC
- The availability zone and subnet in which the EC2 instance is running must be mapped while creating the load balancer