

Load balancer Lab Document by Abhijeet Kumar

Lab : Create a microservice architecture using Application Load Balancer.

A load balancer is a device or service that distributes incoming network traffic across multiple servers or resources to ensure no single server is overloaded.

Different type of Load Balancer:

Application Load Balancer (ALB)

- **Features:**
 - Operates at the Layer 7 (Application) level.
 - Suited for modern, microservices-based, and containerized applications.
 - Supports routing based on host-based (domain names) and path-based routing (URLs).
 - Allows routing to different targets based on content type (e.g., HTTP headers, HTTP methods).
 - Built-in support for WebSocket and HTTP/2.
 - SSL/TLS termination and more sophisticated security features.
 - Supports containerized applications (e.g., ECS, Kubernetes).

3. Network Load Balancer (NLB)

- **Features:**
 - Operates at the Layer 4 (Transport) level.
 - Suited for applications that require high performance, low latency, and the ability to handle millions of requests per second.
 - Supports TCP and UDP traffic, which is great for real-time applications like gaming, IoT, and VoIP.
 - Handles sudden, unpredictable traffic bursts.

- Ideal for applications that require extreme performance, such as those that need to scale to millions of requests per second.

4. Gateway Load Balancer (GWLB)

- **Features:**

- Operates at Layer 3 (Network) level.
- Suited for third-party virtual appliances, like firewalls, intrusion detection systems, and deep packet inspection appliances.
- Works with appliances in the Virtual Private Cloud (VPC) to direct traffic.
- Useful for networking and security appliances that need to inspect and handle traffic at a network layer.
- Allows scaling of virtual appliances with automatic traffic distribution.

Prerequisite:

1. Create a VPC with 3 public subnets and 3 private subnets
2. Create one public server in public subnet
3. Create 3 private server in different private subnet
4. Connect to each private server using public server
5. Install webserver and add index.html in each private server

Ex: Name server 1 as home and add index.html file

Name server 2 as mobile and add index.html file in folder

`/var/www/html/mobile`

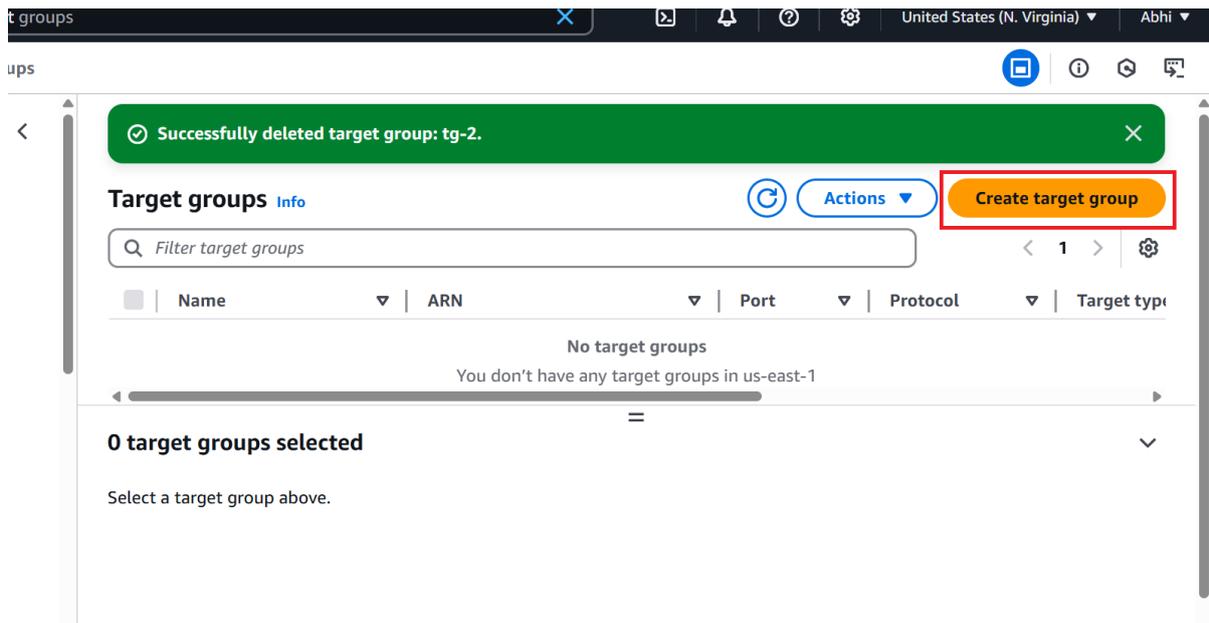
Name server 3 as electronics and add index.html file in folder

`/var/www/html/electronics`

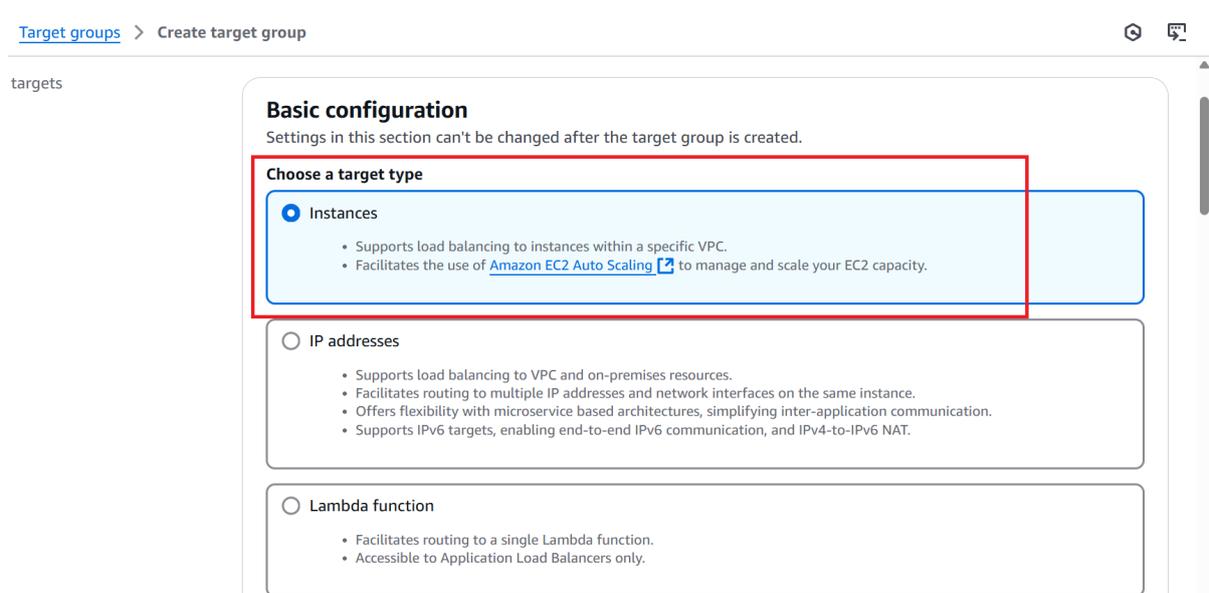
Note: Refer previous notes and doc for above steps.



Step 1: Click on create target group



Step 2: Select Instances as option



Step 3: Enter name of target group

• facilitates using static IP addresses and PrivateLink with an Application Load balancer.

Target group name

tg-1

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

Protocol : Port

Choose a protocol for your target group that corresponds to the Load Balancer type that will route traffic to it. Some protocols now include anomaly detection for the targets and you can set mitigation options once your target group is created. This choice cannot be changed after creation

HTTP

80

1-65535

IP address type

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

IPv4

Each instance has a default network interface (eth0) that is assigned the primary private IPv4 address. The instance's primary private IPv4 address is the one that will be applied to the target.

IPv6

Each instance you register must have an assigned primary IPv6 address. This is configured on the instance's default network interface (eth0). [Learn more](#)

Step 4: Select your VPC

VPC

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.

my-vpc

vpc-03a4d058d1d5bbf6a
IPv4 VPC CIDR: 10.0.0.0/24

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.

Step 5: Click next

Up to 1024 characters allowed.

▶ [Advanced health check settings](#)

Attributes

ⓘ Certain default attributes will be applied to your target group. You can view and edit them after creating the target group.

▶ Tags - optional

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

Cancel

Next

Step 6: Select home instance/server from the list

Register targets

This is an optional step to create a target group. However, to ensure that your load balancer routes traffic to this target group you must register your targets.

Available instances (1/4)

Filter instances

<input type="checkbox"/>	Instance ID	Name	State	Sec
<input type="checkbox"/>	i-0613403e05da97482	electronic	Running	nlb
<input type="checkbox"/>	i-07e26778bc36fbc57	mobile	Running	nlb
<input checked="" type="checkbox"/>	i-05f008688ad230721	home	Running	nlb
<input type="checkbox"/>	i-0a83281a11517ae7f	pub-serv	Running	nlb

Step 7: Click include as pending below and create target group

80
1-65535 (separate multiple ports with commas)

Include as pending below

Review targets

Targets (0) Remove all pending

Show only pending < 1 > ⚙️

Instance ID	Name	Port	State	Security groups	Zone	Private IPv4 address	Subnet
No instances added yet Specify instances above, or leave the group empty if you prefer to add targets later.							

0 pending Cancel Previous **Create target group**

Step 8: Create another target group for mobile server

Repeat all above steps except in path add

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.

► **Advanced health check settings**

Attributes

ⓘ Certain default attributes will be applied to your target group. You can view and edit them after creating the target group.

Step 9: Select server as mobile and click include pending below

EC2 > Target groups > Create target group

Step 1
Specify group details

Step 2
Register targets

Register targets
This is an optional step to create a target group. However, to ensure that your load balancer routes traffic to this target group you must register your targets.

Available instances (1/4)

Filter instances

<input type="checkbox"/>	Instance ID	Name	State	Security groups	Zone
<input type="checkbox"/>	i-0613403e05da97482	electronic	Running	nlb-sg	us-east-1c
<input checked="" type="checkbox"/>	i-07e26778bc36fbc57	mobile	Running	nlb-sg	us-east-1b
<input type="checkbox"/>	i-05f008688ad230721	home	Running	nlb-sg	us-east-1a
<input type="checkbox"/>	i-0a83281a11517ae7f	pub-serv	Running	nlb-sg	us-east-1a

1 selected

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

80

1-65535 (separate multiple ports with commas)

[Include as pending below](#)

Step 10: Create create target group

Step 11: Create another target group for electronic

Repeat all above steps except in path add

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.

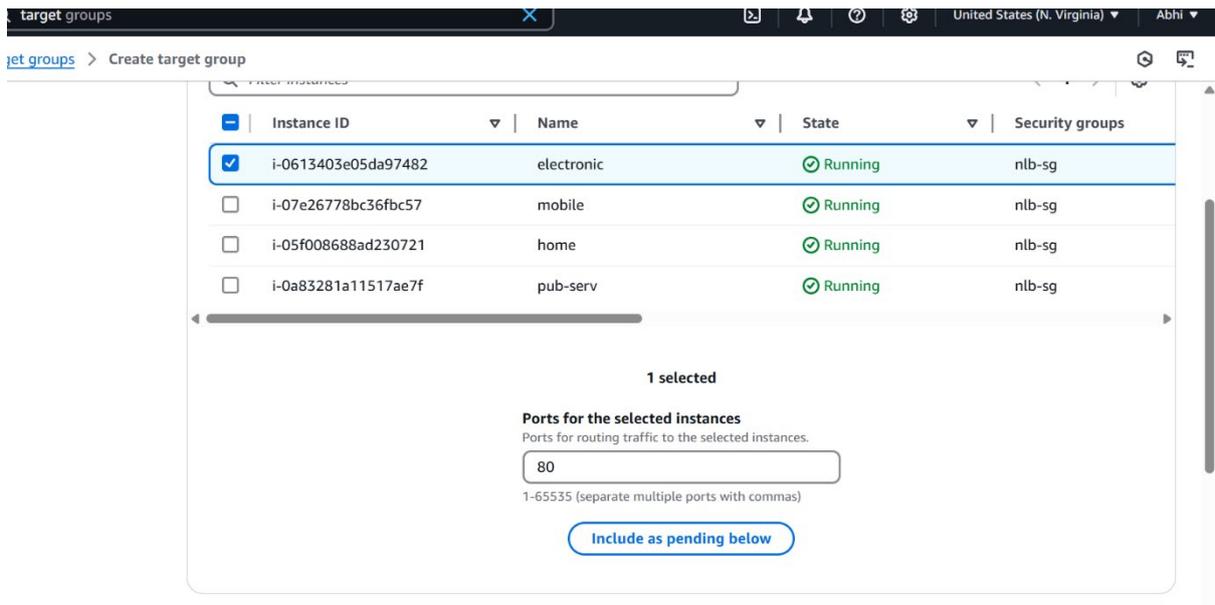
/electronic/

Up to 1024 characters allowed.

▶ **Advanced health check settings**

Attributes

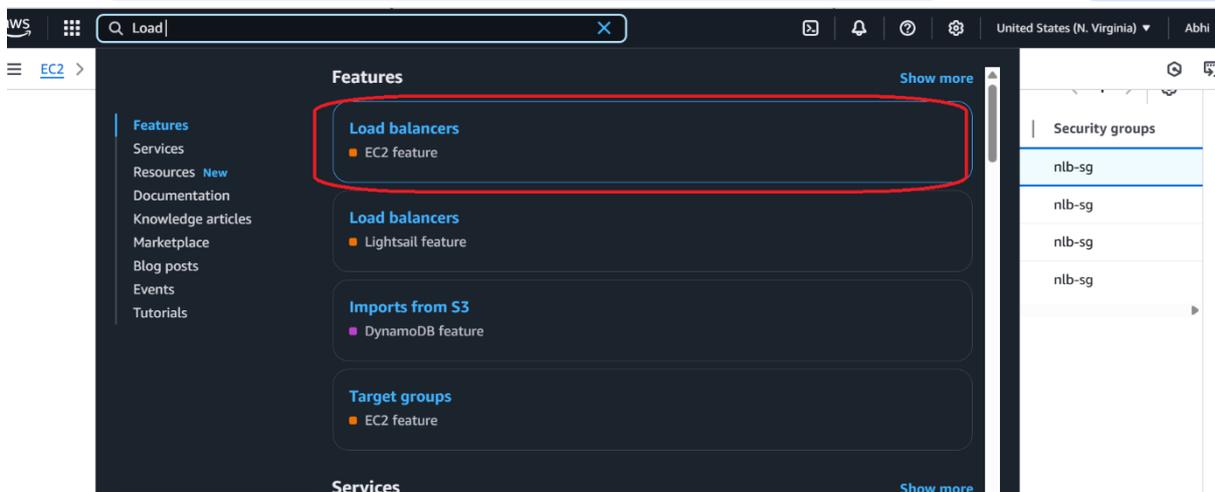
Step 12: Select server as electronic and click include pending below



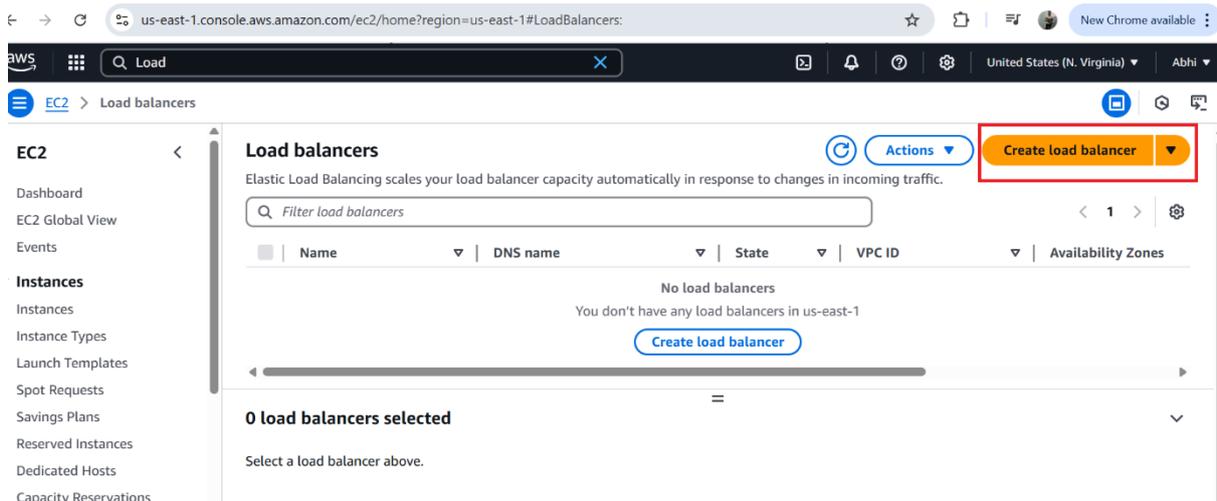
Step 13: Create create target group

Now all three target group is created with three different server and different application in it, But still if you go and check all target will show as unhealthy since there is no Load Balancer

Step 14: Search for Load balancer and select it



Step 15: Click on Create Load Balancer



Step 16: Click create on Application Load balancer

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](#)

Step 17: Enter name and select Internet facing

Load balancer name [Info](#)
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

- Serves internet-facing traffic.
- Has public IP addresses.
- DNS name is publicly resolvable.
- Requires a public subnet.

Internal

- Serves internal traffic.
- Has private IP addresses.
- DNS name is publicly resolvable.
- Compatible with the **IPv4** and **Dualstack** IP address types.

Load balancer IP address type [Info](#)
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. Public IPv4 addresses have an additional cost.

IPv4
Includes only IPv4 addresses.

Dualstack
Includes IPv4 and IPv6 addresses.

Dualstack without public IPv4
Includes a public IPv6 address, and private IPv4 and IPv6 addresses. Compatible with **internet-facing** load balancers only.

Step 18: Select your VPC and all available zones

VPC [Info](#)
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 Lambda or on-premises targets, or if using VPC peering. To confirm the VPC for your targets, view [target groups](#). For a new VPC, [create a VPC](#).

vpc-03a4d058d1d5bbf6a
IPv4 VPC CIDR: 10.0.0.0/24

IP pools - [new](#) [Info](#)
You can optionally choose to configure an IPAM pool as the preferred source for your load balancers IP addresses. Create or view [Pools in Amazon VPC IP Address Manager console](#).

Use IPAM pool for public IPv4 addresses
The IPAM pool you choose will be the preferred source of public IPv4 addresses. If the pool is depleted IPv4 addresses will be assigned by AWS.

Availability Zones and subnets [Info](#)
Select at least two Availability Zones and a subnet for each zone. A load balancer node will be placed in each selected zone and will automatically scale in response to traffic. The load balancer routes traffic to targets in the selected Availability Zones only.

us-east-1a (use1-az6)

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-0dc2cdf6e468511ed
IPv4 subnet CIDR: 10.0.0.0/27 pub-sub-1

us-east-1b (use1-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-04f554153e533aad9
IPv4 subnet CIDR: 10.0.0.32/27 pub-sub-2

us-east-1c (use1-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.

Step 19: Select security group and target group (target group select as home server in it i.e tg-1)

Security groups

Select up to 5 security groups

nlb-sg
sg-0162aec1c3afd6b38 VPC: vpc-03a4d058d1d5bbf6a

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 tg-1 (Target type: Instance, IPv4) HTTP

[Create target group](#)

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

[Add listener tag](#)

Step 20: Click on create load balancer

pub-sub-2
• us-east-1c
[subnet-018e95ee3e15a0c8c](#)
pub-sub-3

Service integrations [Edit](#)

Amazon CloudFront + AWS Web Application Firewall (WAF): -
AWS WAF: -
AWS Global Accelerator: -

Tags [Edit](#)

-

Attributes

ⓘ Certain default attributes will be applied to your load balancer. You can view and edit them after creating the load balancer.

Operation workflow and status

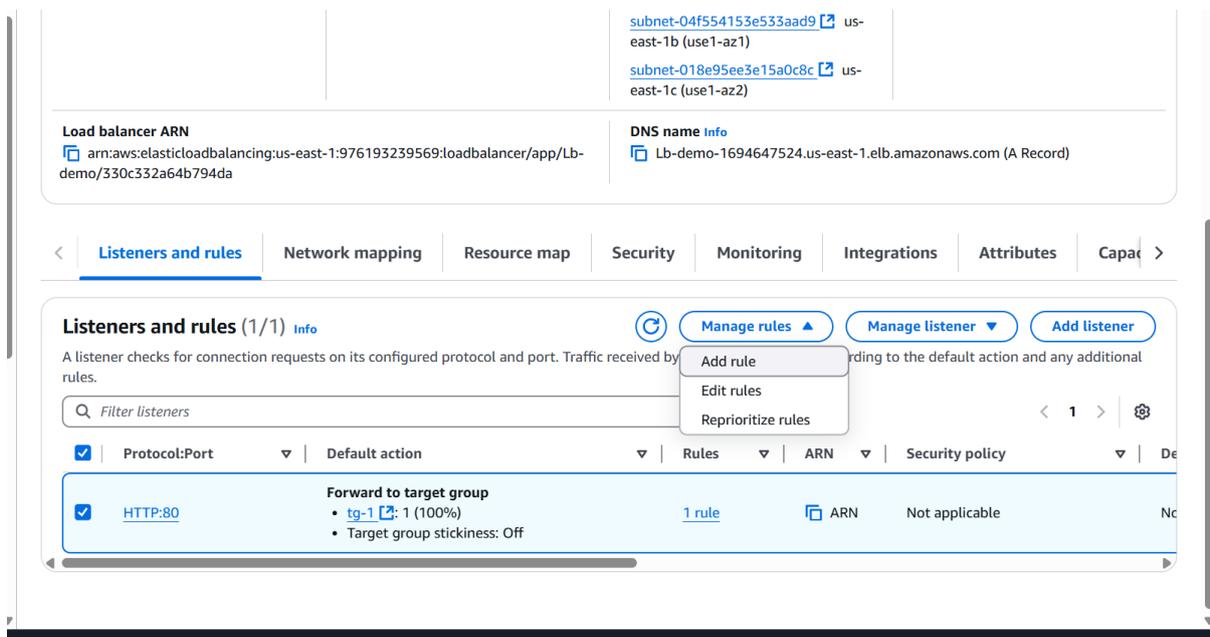
► **Server-side tasks and status**

After completing and submitting the above steps, all server-side tasks and their statuses become available for monitoring.

[Cancel](#) [Create load balancer](#)

Load balancer will be created successfully, scroll below and go to bottom

Select the listener and rule and click on manage rule and select add rule from that



Step 21: enter name and click next

Add rule [Info](#)

Define the rule and then review it in the context of the other rules on this listener.

▶ Listener details: HTTP:80

Name and tags [Info](#)
Tags can help you manage, identify, organize, search for and filter resources.

Name
 [Add additional tags](#)

[Cancel](#)

[Next](#)

Step 22: Click on add condition

Define rule conditions [Info](#)

Requests reaching this rule must match all specified conditions for the rule to apply. At least 1 condition is required.

itions

ins

te

► Listener details: HTTP:80

Conditions (0)

[Rule limits](#)

No conditions
No conditions to display.

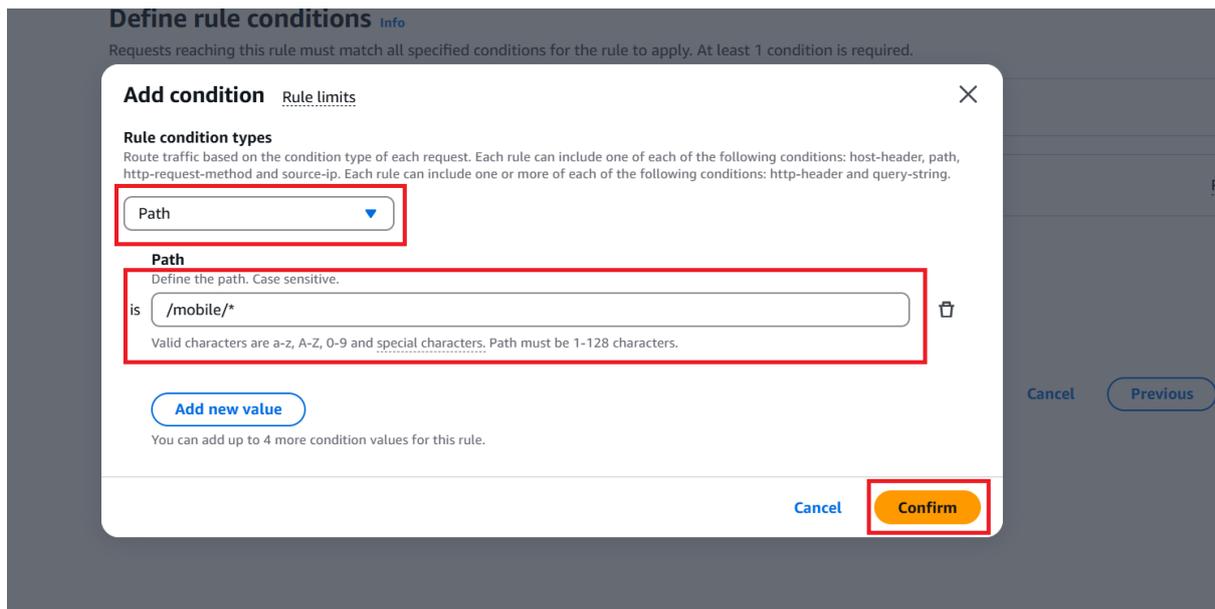
[Add condition](#)

[Cancel](#)

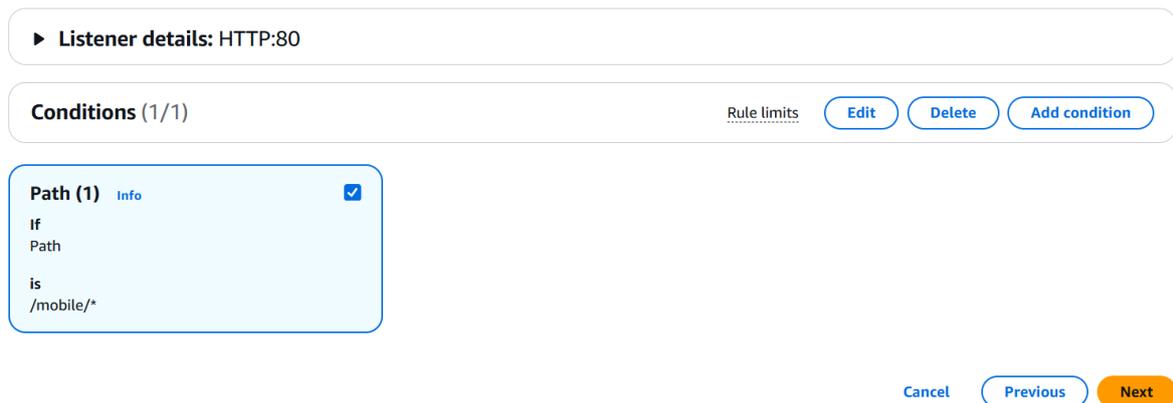
[Previous](#)

[Next](#)

Step 23: Select option as path and enter path as `/mobile/*` and click on confirm



Step 24: select path checkbox and click next



Step 25: Select target group in which mobile server is placed

Actions

Action types

Routing actions

Forward to target groups
 Redirect to URL
 Return fixed response

Forward to target group [Info](#)
 Choose a target group and specify routing weight or [Create target group](#).

Target group

tg-2 Target type: Instance, IPv4 HTTP ▼

[Add target group](#)

You can add up to 4 more target groups.

Target group stickiness [Info](#)
 Enables the load balancer to bind a user's session to a specific target group. To use stickiness the client must support cookies. 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

Weight: 1 (0-999) Percent: 100%

[Cancel](#)
[Previous](#)
[Next](#)

Step 26: Enter priority as 1000 and click next

Listener rules (2) [Info](#) [Rule limits](#) [Reset priorities](#) [Add gap between priorities](#)

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.

Name tag	Priority	Conditions (If)	Actions (Then)	ARN
mobile	1000 <small>Priority value must be 1-50,000.</small>	Path Pattern is /mobile/*	Forward to target group <ul style="list-style-type: none"> tg-2: 1 (100%) Target group stickiness: Off 	Pending
Default	Last (default)	If no other rule applies	Forward to target group <ul style="list-style-type: none"> tg-1: 1 (100%) Target group stickiness: Off 	ARN

[Cancel](#)
[Previous](#)
[Next](#)

Step 27: Click create

Priority 1000	Conditions (If) If request matches all: Path Pattern is /mobile/*	Actions (Then) Forward to target group <ul style="list-style-type: none"> • tg-2: 1 (100%) • Target group stickiness: Off
-------------------------	---	--

Rule ARN
Pending

Rule tags (1) Edit

Tags can help you manage, identify, organize, search for and filter resources.

Key	Value
Name	mobile

► **Server-side tasks and status**
After completing and submitting the above steps, all server-side tasks and their statuses become available for monitoring.

Cancel Previous Create

Repeat the same step for target group 3 ie electronic

Go to load balancer and copy the DNS name

> Lb-demo 🔍

Lb-demo 🔄 Actions ▾

▼ **Details**

Load balancer type Application	Status 🔄 Provisioning	VPC vpc-03a4d058d1d5bbf6a	Load balancer IP address type IPv4
Scheme Internet-facing	Hosted zone Z35SXDOTRQ7X7K	Availability Zones subnet-0dc2cdbfe468511ed us-east-1a (use1-az6) subnet-04f554153e533aad9 us-east-1b (use1-az1) subnet-018e95ee3e15a0c8c us-east-1c (use1-az2)	Date created March 29, 2025, 14:23 (UTC+05:30)

Load balancer ARN
[arn:aws:elasticloadbalancing:us-east-1:976193239569:loadbalancer/app/Lb-demo/330c332a64b794da](#)

DNS name Info
[Lb-demo-1694647524.us-east-1.elb.amazonaws.com](#) (A Record)

< Listeners and rules Network mapping Resource map Security Monitoring Integrations Attributes Capacity >

Go to the browser and paster the dns name in browser as URL

-You will see the home page

Now append /mobile/ and the end of DNS URL and you will see mobile page

Similarly for electronics

