**Placement Empowerment Program**

***Cloud Computing and DevOps Centre***

**Set Up a Load Balancer in the Cloud**

Configure a load balancer to distribute traffic across multiple VMs hosting your web application.

Name: Ragul.G.V Department: ECE



**Introduction and Overview**

A Load Balancer in AWS is a crucial service that automatically distributes incoming application traffic across multiple targets, such as EC2 instances, containers, and IP addresses. Elastic Load Balancing (ELB) helps improve the availability, reliability, and performance of your applications by balancing traffic loads efficiently.

**Objectives**

* To configure an Elastic Load Balancer (ELB) in AWS.
* To distribute traffic evenly across multiple EC2 instances.
* To enhance application availability and fault tolerance.
* To improve load handling and scalability of web applications.

**Importance**

* **High Availability:** Ensures continuous application performance even if some instances fail.
* **Fault Tolerance:** Redirects traffic to healthy instances, reducing downtime.
* **Scalability:** Efficiently manages traffic spikes and dynamic workloads.
* **Improved Performance:** Optimizes resource usage and reduces response time.

**STEPS:**

**STEP 1: Launch EC2 Instances**

* Go to the AWS Management Console → EC2 → Launch Instance.
* Choose an Amazon Machine Image (AMI), e.g., Amazon Linux 2.
* Select an instance type like t2.micro.
* Configure instance details and add the following User Data script:

**#!/bin/bash**

**yum update -y**

**yum install -y httpd**

**systemctl start httpd**

**systemctl enable httpd**

**echo "<h1>Web Server from $(hostname -f)</h1>" > /var/www/html/index.html**

* Create or select an existing key pair for SSH access.
* Set up a security group allowing **HTTP (port 80)** and **SSH (port 22)**.
* Launch at least **two EC2 instances** with these configurations.

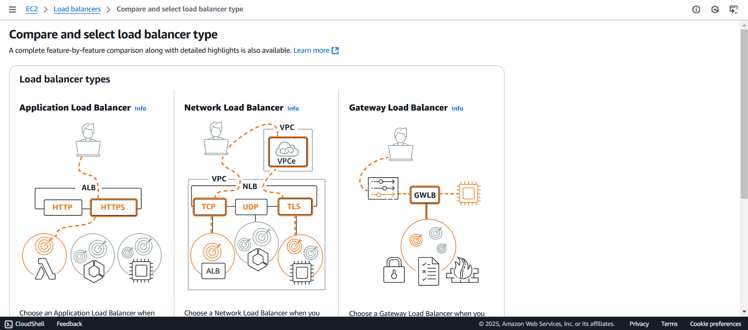


**STEP 2: Verify Web Servers**

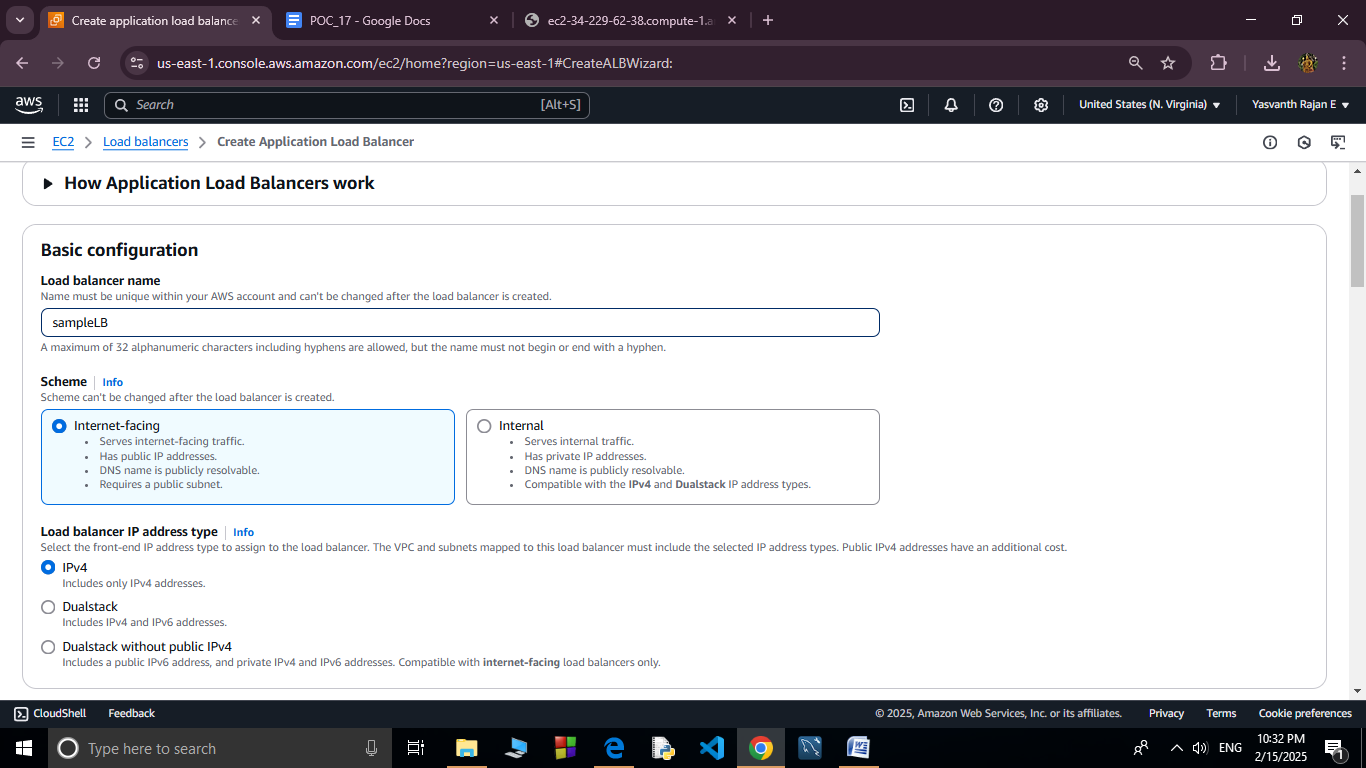
* Get the **Public IP** of each EC2 instance.
* Open the IPs in a browser to confirm the message "Web Server from [hostname]" appears.

**STEP 3: Create an Elastic Load Balancer (ELB)**

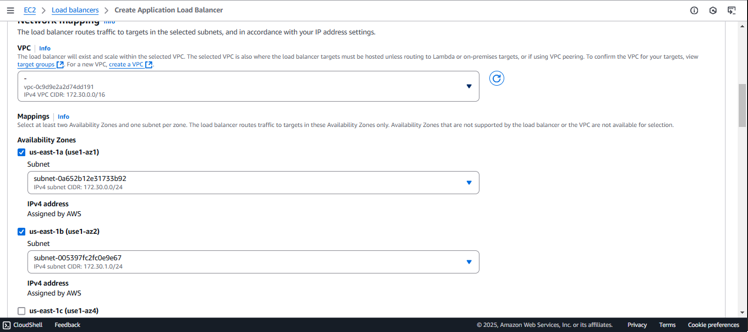
* Go to **EC2 Dashboard** → **Load Balancers** → **Create Load Balancer**.
* Choose **Application Load Balancer (ALB)**.



* Provide a name, select **internet-facing** for public access, and choose **IPv4**.



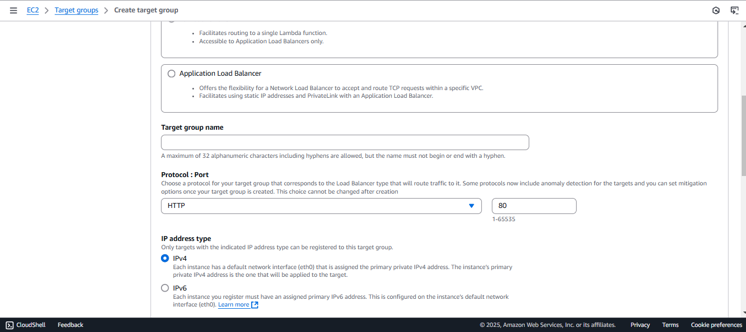
* Configure listeners: Add **HTTP (port 80)**.
* Select the VPC and **at least two subnets** for high availability.



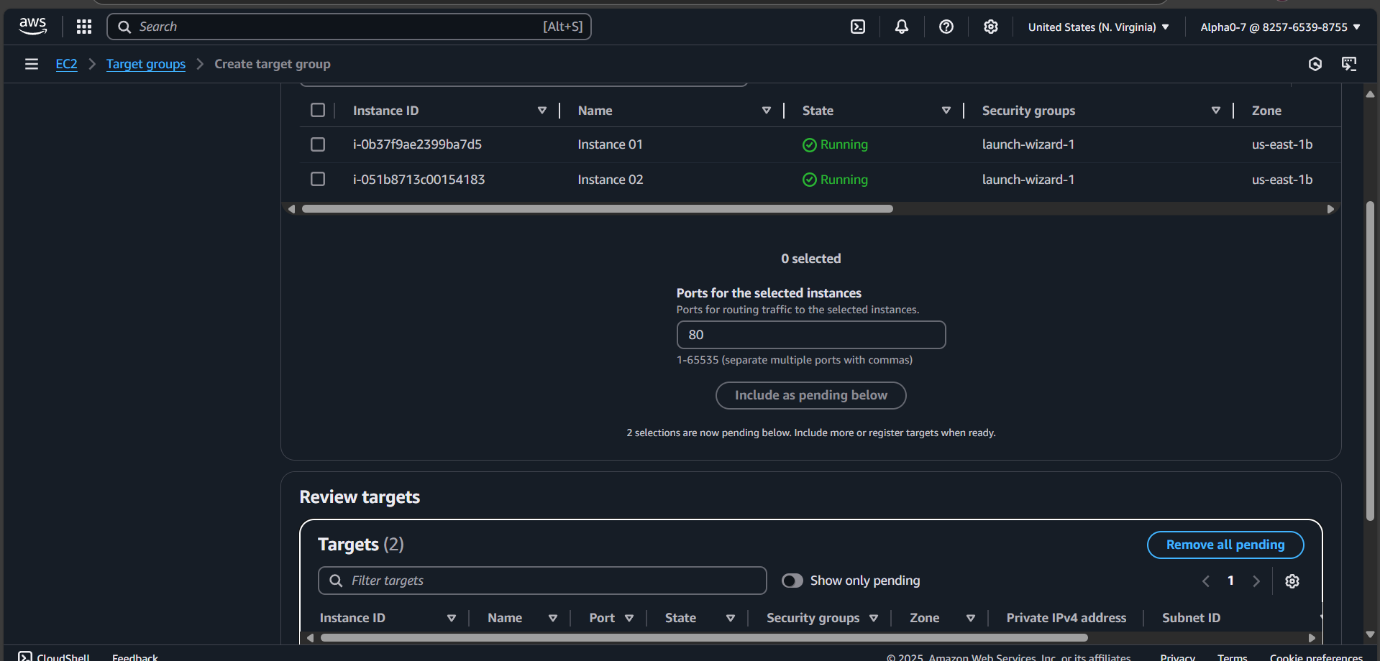
* Configure security groups to allow **HTTP (port 80)**.

**STEP 4: Set Up Target Group**

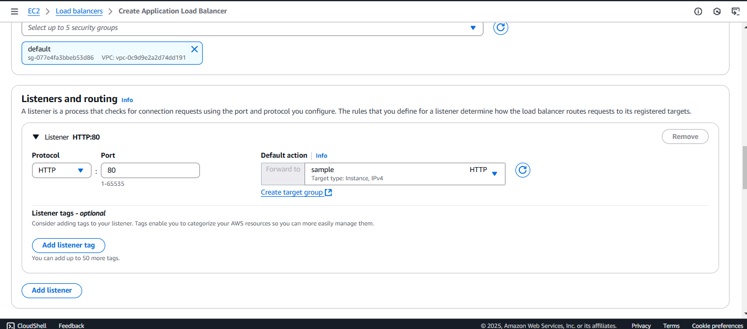
* Create a new **Target Group** with the type set to **Instances**.
* Protocol: **HTTP**, Port: **80**.

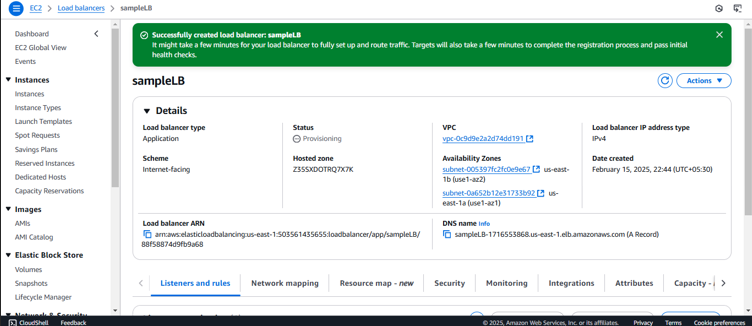


* Register the previously launched EC2 instances as targets.



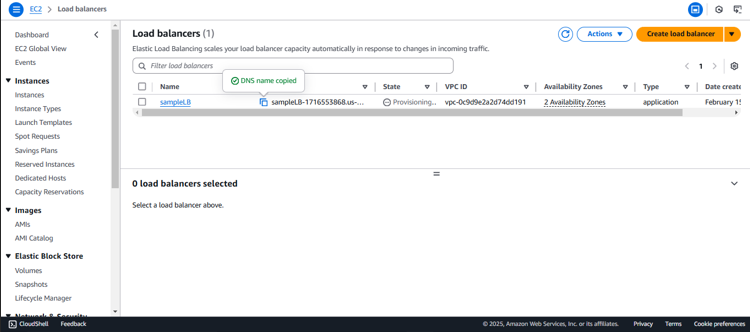
* Review and create the load balancer.





**STEP 5: Test the Load Balancer**

* Go to **EC2 Dashboard** → **Load Balancers**.



* Copy the **DNS name** of the load balancer.
* Paste it into a browser.
* Refresh multiple times to see traffic distribution between instances (hostname will change).

