AWS Load Balancer – Step by Step Guide

This document explains how to create an AWS Application Load Balancer (ALB), how it works internally, and how to test it after deployment.

1. What is a Load Balancer?

A Load Balancer distributes incoming user traffic across multiple EC2 instances. It improves application availability, fault tolerance, and scalability by ensuring no single server is overloaded.

2. Types of Load Balancers in AWS

- **Application Load Balancer (ALB)** – Used for HTTP/HTTPS web applications
- **Network Load Balancer (NLB)** – Used for high performance TCP traffic
- **Gateway Load Balancer** – Used for security appliances

3. Step 1 – Launch EC2 Instances

Create at least two EC2 instances in the same VPC. Install a web server such as Nginx or Apache on port 80 so the load balancer can forward traffic.

4. Step 2 – Create a Target Group

- Go to EC2 → Target Groups → Create target group
- Target type: Instances
- Protocol: HTTP, Port: 80
- Register EC2 instances and ensure health status is Healthy

5. Step 3 – Create Application Load Balancer

- Go to EC2 → Load Balancers → Create Load Balancer
- Select Application Load Balancer
- Choose Internet-facing scheme
- Select at least two subnets in different Availability Zones

6. Step 4 – Configure Listener and Routing

Create an HTTP listener on port 80 and configure the default action to forward traffic to the previously created target group.

7. Step 5 – Security Group Configuration

Allow inbound HTTP (port 80) traffic on the Load Balancer security group. Ensure EC2 security group allows traffic only from the Load Balancer security group.

8. Step 6 – Testing the Load Balancer

After creation, AWS provides a DNS name for the load balancer. Open the DNS name in a browser. If the page loads successfully, the load balancer is working. Stopping one EC2 instance should still keep the

application accessible.

9. How Load Balancer Works Internally

User request → Load Balancer DNS → Listener → Target Group → Healthy EC2 Instance. The load balancer continuously checks health and routes traffic only to healthy targets.

10. Common Issues

- Target group not visible – VPC mismatch
- 503 error – No healthy targets
- Website not loading – Security group misconfiguration