

An AWS Load Balancer, part of the Elastic Load Balancing (ELB) service, is essentially a traffic cop for your web application. Its job is to automatically and efficiently spread incoming internet traffic (user requests) across multiple backend resources, like your EC2 virtual servers. This prevents any single server from getting overloaded, which ensures your application remains fast (scalable) and always available (fault-tolerant).

REAL-LIFE EXAMPLE: THE RESTAURANT HOST

Imagine your web application is a very popular restaurant. The AWS Load Balancer acts as the Restaurant Host/Hostess. When customers (traffic) arrive, the Hostess checks which waiters (EC2 Instances) are free (Health Check) and directs the new customer to the emptiest one (Load Balancing Algorithm). This ensures everyone is served quickly.

LOAD BALANCER COMPONENTS

1. The Front Door (Listener): Checks for client connections on a specific port/protocol.
2. The Waiters (Target Group): A logical group of servers/resources that handle the requests.
3. The Health Check: An automatic check to see if a server is running and ready.

TYPES OF AWS LOAD BALANCERS

- * Application Load Balancer (ALB): Best for web applications (HTTP/HTTPS). Routes traffic based on content (e.g., URL path).
- * Network Load Balancer (NLB): Best for extreme performance and ultra-low latency. Routes traffic based on IP address and port (Layer 4).
- * Gateway Load Balancer (GWLB): Best for managing virtual security appliances (e.g., firewalls).

CONTENT-BASED ROUTING (ALB)

The ALB looks at the content of the HTTP request to decide where to send it.

- * Conditions (What triggers the rule?):
 - * Path: e.g., send /api/* to the API service Target Group.
 - * Host Header: e.g., send shop.example.com to the Shopping Cart Target Group.
- * Actions (What should the ALB do?):
 - * Forward: Send the request to a specific Target Group.
 - * Redirect: Send the client to a different URL.

Benefits: Enables Microservices Architecture, Cost Savings, and A/B Testing.