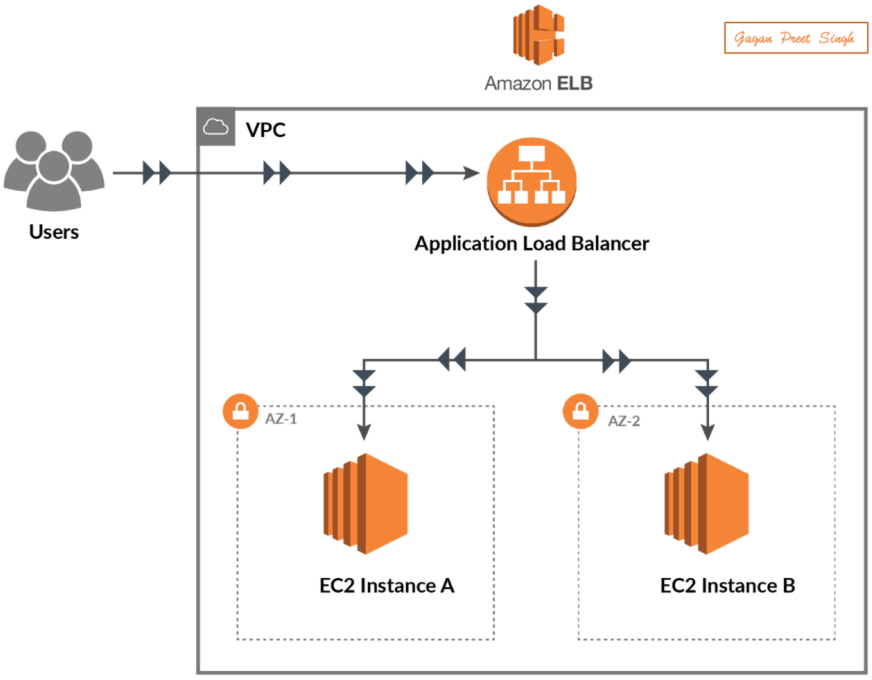
Elastic Load Balancer (ELB)

Introduction to Elastic Load Balancer (ELB)

Amazon **Elastic Load Balancer (ELB)** is a managed load balancing service that automatically distributes incoming network or application traffic across multiple **EC2 instances, containers, and IP addresses** in different **Availability Zones (AZs)**. It enhances **scalability, fault tolerance, and availability**.



Types of Load Balancers in AWS

AWS provides three types of Elastic Load Balancers:

1. Application Load Balancer (ALB)

- Operates at **Layer 7 (Application Layer)** of the OSI model.
- Routes traffic based on **HTTP headers, URL paths, hostnames, query parameters, and request methods**.
- Supports **WebSockets**, **gRPC**, and **content-based routing** (host/path-based).
- Works with **EC2 instances, containers (ECS, EKS), and Lambda functions**.

Use Cases

- Microservices architectures.
- Web applications requiring **content-based routing**.
- Applications needing **SSL termination** and **WebSockets**.

---

2. Network Load Balancer (NLB)

- Operates at **Layer 4 (Transport Layer)**.
- Handles **TCP, UDP, and TLS** traffic.
- Supports **millions of requests per second** with ultra-low latency.
- Provides **static IP addresses** and supports **Elastic IPs**.
- Best suited for applications requiring **high performance** and **low latency**.

Use Cases

- Financial services and real-time trading applications.
- Online gaming with **UDP connections**.
- Load balancing for databases and other low-latency applications.

---

3. Gateway Load Balancer (GWLB)

- Operates at **Layer 3 (Network Layer)**.
- Designed for **third-party virtual appliances** such as **firewalls, intrusion detection/prevention systems (IDS/IPS), deep packet inspection (DPI), and traffic monitoring tools**.
- **Maintains original source IP addresses**, enabling deep traffic analysis.
- Uses **Geneve encapsulation** for traffic forwarding.

Use Cases

- Deploying security appliances like **firewalls, IDS/IPS, DDoS mitigation, and traffic inspection tools**.
- Enforcing network security policies.
- Filtering or inspecting inbound and outbound traffic.

Key Components of an Elastic Load Balancer

1. **Listeners**
    - Rules that define how ELB processes and forwards traffic.
    - Example: Listening on **port 80 (HTTP)** and forwarding traffic to **port 8080**.
2. **Target Groups**
    - Logical grouping of **EC2 instances, IPs, or Lambda functions**.
    - Health checks ensure **only healthy targets receive traffic**.
3. **Rules & Conditions** (Only for ALB)
    - **Host-based routing**: Routes `api.example.com` to API servers, `blog.example.com` to blog servers.
    - **Path-based routing**: `/login` goes to one target group, `/products` to another.
4. **Cross-Zone Load Balancing**
    - Distributes traffic evenly across **multiple Availability Zones (AZs)**.
5. **Sticky Sessions** (Session Persistence)
    - Ensures a user session stays with a specific backend instance.

How Elastic Load Balancer Works

1. **Receives Incoming Traffic** from users.
2. **Processes Listener Rules** to determine the best routing method.
3. **Performs Health Checks** to ensure only healthy targets get traffic.
4. **Distributes Load Across Targets** using various algorithms.
5. **Ensures High Availability & Auto Scaling** by dynamically adjusting traffic flow.

Load Balancing Algorithms

- **Round Robin** (default for ALB) – Distributes requests equally.
- **Least Outstanding Requests** (NLB) – Sends requests to the least busy instance.
- **Hash-based Routing** (ALB) – Routes requests based on a session, URL path, or query.

Key Features of AWS ELB

- **Auto Scaling Integration** – Automatically adjusts instances based on demand.
- **SSL Termination** – Offloads SSL decryption at ELB to reduce backend workload.
- **Access Logs & Monitoring** – Logs requests to **S3** and provides **CloudWatch metrics**.
- **Security & IAM** – Supports **AWS WAF**, **security groups**, and **IAM roles**.

Setting Up an Elastic Load Balancer

1. Creating an Application Load Balancer (ALB)

1. Open **AWS EC2 Console** → Load Balancers → Create Load Balancer.
2. Choose **Application Load Balancer**.
3. Configure **Listeners (HTTP/HTTPS)**.
4. Choose **at least two Availability Zones** for high availability.
5. Create a **Target Group** and register EC2 instances.
6. Define **Rules** for routing.
7. Review and launch.

2. Creating a Network Load Balancer (NLB)

1. Select **Network Load Balancer**.
2. Choose **TCP/UDP protocol** and specify a listener port.
3. Assign a **Static IP** or **Elastic IP** (optional).
4. Register EC2 instances or IP addresses.
5. Launch and configure.

3. Creating a Gateway Load Balancer (GWLB)

1. Select **Gateway Load Balancer**.
2. Attach it to a **security appliance (firewall, IDS/IPS, etc.)**.
3. Use **Geneve encapsulation** for routing traffic through the appliance.
4. Configure security policies and launch.

Health Checks in ELB

- Health checks verify the status of backend instances.
- By default, checks for **HTTP 200 OK responses**.
- Customizable settings:
    - **Protocol**: HTTP, TCP, HTTPS.
    - **Path**: `/health`.
    - **Interval**: Frequency of checks.
    - **Thresholds**: Number of failed checks before marking an instance unhealthy.

ELB Security Best Practices

1. **Enable HTTPS (SSL/TLS)** – Use AWS **Certificate Manager (ACM)** for SSL.

2. **Use AWS WAF** – Protect against web attacks and DDoS.
3. **Restrict Access with Security Groups** – Allow only trusted IPs.
4. **Enable Logging** – Store logs in **Amazon S3**.

ELB Pricing

AWS ELB pricing is based on:

- **Data processed** (per GB).
- **Load balancer hours** (active usage).
- **New and active connections** (for NLB).

Pricing Comparison

Feature	Application Load Balancer (ALB)	Network Load Balancer (NLB)	Gateway Load Balancer (GWLB)
OSI Layer	Layer 7 (Application)	Layer 4 (Transport)	Layer 3 (Network)
Protocols	HTTP, HTTPS, WebSockets	TCP, UDP, TLS	Geneve (traffic encapsulation)
Routing	Path-based, Host-based	Direct to target	Security appliance traffic
Performance	High	Ultra-low latency	Security & inspection
SSL Termination	Yes	No	No
Use Cases	Web apps, microservices	High-performance apps	Security & traffic filtering

---

Conclusion

- **Application Load Balancer (ALB)**: Best for **HTTP-based applications**.
- **Network Load Balancer (NLB)**: Best for **low-latency TCP/UDP traffic**.
- **Gateway Load Balancer (GWLB)**: Best for **security appliances** like **firewalls and traffic monitoring**.
- AWS ELB **improves scalability, fault tolerance, and availability**.