Elastic Load Balancing (ELB)

Elastic Load Balancing

- Automatically distribute traffic across EC2 instances
- Health Check – route traffic to healthy instances
- Achieve fault tolerance protect against:
  - EC2 instance failure
  - Availability Zone failure
- Fully managed – automatically scales load balancing capacity needed to route application traffic
- Single point-of-contact for your clients

Elastic Load Balancing

- Add or Remove EC2 instances without impacting clients

- Protection against Denial-Of-Service Attacks – absorb certain type of DOS attacks

- Reduce attack surface – for internet facing services, keep your EC2 instances in private subnet

- Use as an Internal Load Balancer or Internet Facing Load Balancer

- SSL Offload

Elastic Load Balancer - Concepts

[Internet Facing Load Balancer](#)

[Internal Load Balancer](#)

Load Balancer Types

- Classic Load Balancer
    - Simple load balancing across EC2 instances
    - Supports HTTP, HTTPS, TCP, SSL (Secure TCP)
    - OSI Layer 4 (Transport), 7 (Application) Load Balancer
- Application Load Balancer
    - Path based routing
    - Route traffic to multiple services
    - Route traffic to different ports on the same EC2 instance
    - Ideal for microservices and container based architectures
    - Supports HTTP, HTTPS, HTTP/2, WebSocket
    - OSI Layer 7 (Application) Load Balancer

amazon
web services

WebSocket

- [WebSocket](#) provides Full Duplex communication over a single TCP connection

- Facilitates near real-time data transfer by sending data back and forth keeping the connection open

- Ideal for frequent interaction between Web browser and Web Server

HTTP/2

- HTTP/2 (HTTP version 2) is an improvement over HTTP protocol

- More efficient use of network resources

- Reduced perception of latency – multiple concurrent exchanges on same connection

- Server to Client Push notification support – Web server can send newer information to client without requiring the client to request for data ['Note that you can't use the server-push feature of HTTP/2' in ELB]

Cross Zone Load Balancing

- For fault tolerance, EC2 instances should be distributed across two or more availability zones

- Cross Zone Load Balancing controls how traffic is distributed across Availability Zones and Instances in each Availability Zone

Disabled - Cross Zone Load Balancing

- Distribute traffic evenly across Availability Zones
  - Happens when Cross Zone load balancing is "Disabled"
  - Two availability zones 'A' and 'B' would each receive 50% of the traffic irrespective of number of EC2 instances in each Availability Zone
  - May cause higher loading if one Availability Zone has fewer EC2 instances
  - Default mode in classic load balancer

Enabled - Cross Zone Load Balancing

- Distribute traffic evenly across EC2 instances in all Availability zones

    - Happens when Cross Zone load balancing is "Enabled"

    - Availability Zone 'A' has 3 instances and Availability Zone 'B' has 2 instances. Each instance would receive 20% of the traffic (1/5th)

    - Default mode in application load balancer

Internet Facing Load Balancer

- One public subnet in every enabled Availability Zone
- Load Balancer nodes are deployed in the public subnet with public IP addresses
- Public DNS host name resolves to public IP Addresses
- Ensure subnet has at least 8 IP addresses left for load balancer use
- Your EC2 instances can be deployed in private subnet
- Load Balancer talks to your EC2 instances using Private IP Addresses

amazon
web services

Internal Load Balancer

- One private or public subnet in every enabled Availability Zone

- Load Balancer nodes are deployed in the specified subnet with private IP addresses

- Public DNS host name resolved to private IP Addresses

- Ensure subnet has at least 8 IP addresses left for load balancer use

- Your EC2 instances can be deployed in private subnet

- Load Balancer talks to your EC2 instances using Private IP Addresses

Classic Load Balancer

- Create Classic Load Balancer
- Specify Listener Configuration
    - Frontend - Protocol and Port where load balancer listens for connections from client
    - Backend - Protocol and Port where EC2 instances listen. Load balancer forwards the requests to the Backend
- Specify Subnets - One Subnet per Availability Zone for Load Balancer nodes
- Assign Security Groups for Load Balancer
- Configure Health Checks
- Register EC2 Instances

Classic Load Balancer Demo

Internet Facing Load Balancer

Internal Load Balancer

Classic Load Balancer Demo

- Internet Facing Load Balancer

- VPC with 2 Availability Zones

- Each AZ contains one public and one private subnet

- Load Balancer nodes in public subnet

- EC2 instances in private subnet

- Bastion Host in public subnet (maintenance access to webserver instances)

- Health Check and simulate errors

Classic Load Balancer

[Listener Configurations](#)

Monitoring

[CloudWatch Metrics](#)

[Access Logs](#)

[CloudTrail Logs](#)

Application Load Balancer

- [Application Load Balancer](#) – Single Point of contact for clients

- Listener Configuration
    - Frontend - Protocol and Port where load balancer listens for connections from client
    - Rules – Routing rules based on condition, priority and target group. When condition is met, request is forwarded to the specified Target Group
    - Backend / Target Group – Each Target Group routes request to registered EC2 instances at specified protocol and port
    - EC2 Instance – Can be registered many target groups

Application Load Balancer

- Health Checks – Configured for every Target Group
  - Health checks are performed on every registered instance in the target group
- Specify Subnets - One Subnet per Availability Zone for Load Balancer nodes
- Assign Security Groups for Load Balancer

Application Load Balancer Benefits

- Path based routing – route based on URL
  - Microservices
  - Containerized application
- Route to multiple services on a Single EC2 instance by registering with different target groups and ports
- CloudWatch metrics at Target Group level
- Improved load balancer performance

Elastic Load Balancer Types

Table: Features of Classic and Application Load Balancer

Routing Algorithm

Classic Load Balancer

- HTTP/HTTPS Listener - Least Outstanding Requests routing

- TCP Listener - Round Robin

Application Load Balancer

- Round Robin within Target Group

Billing

[Classic Load Balancer Billing](#)

- Hourly Charge for load balancer ($0.025 per hour)
- Per GB data processed charge ($0.008 per GB)

[Application Load Balancer Billing](#)

- Hourly Charge for load balancer ($0.0225 per hour)
- Hourly Load Balancer Compute Unit (LCU) ($0.008 per LCU hour)