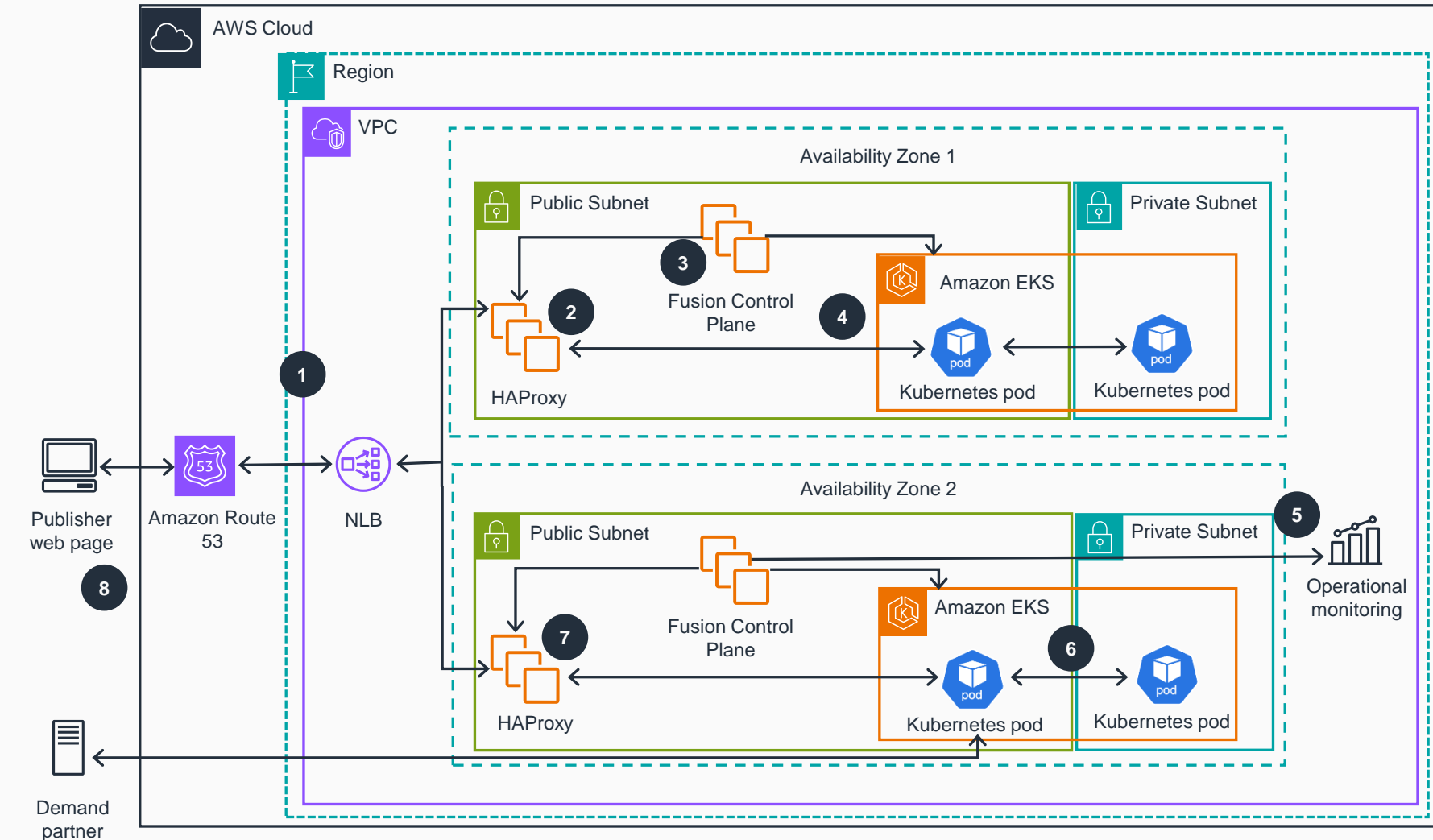# Guidance for Load Balancing Advertising Real-Time Bidding Traffic with HAProxy on AWS

This architecture diagram shows how to implement reliable, high-performance traffic distribution using a network load balancer (NLB) and HAProxy Enterprise for advertising RTB workloads.

AWS Reference Architecture

1. Publisher web pages generate ad requests that route through **Amazon Route 53** to network load balancers (NLBs) at the supply side platform (SSP). NLBs operate in transmission control protocol (TCP) mode with proxy protocol v2 and, overall, provide consistent connection management and high availability.

2. HAProxy Enterprise deployed on **Amazon Elastic Compute Cloud (Amazon EC2)** instances across multiple Availability Zones (AZs) receive ad requests from NLBs, running on **AWS Graviton** for best cost-performance. They provide advanced routing, preserve client IP information, and perform TLS termination using both Elliptic Curve Cryptography (ECC) and Rivest-Shamir-Adleman (RSA) certificates.

3. The HAProxy Fusion Control Plane runs on three **EC2** instances for high availability, managing service discovery, configuration, and real time monitoring of HAProxy fleet.

4. SSP auction services run on **Amazon Elastic Kubernetes Service (Amazon EKS)** pods which filter and enrich ad requests before sending bids to demand partners. Service discovery is configured through the Kubernetes API integration, enabling direct pod routing and health checks for backend auction services with real time service discovery updates.

5. The monitoring infrastructure logs hundreds of thousands of requests per second for high volume traffic and metrics collection at the control plane, including alerts for key performance indicators as well as integration with operational monitoring tools.

6. The networking layer leverages public subnets for services requesting bids to demand partners and private subnets for internal services, blending cost efficiency with properly configured security groups and networking features for optimal performance.

7. HAProxy auto scaling uses either Auto Scaling Groups or Karpenter with hardened Amazon Machine Images (AMIs) and bootstrap processes managed by the control plane for rapid scaling.

8. Ad responses are routed to publishers through HAProxy and returned to the browser, completing the RTB process in under 500 milliseconds.