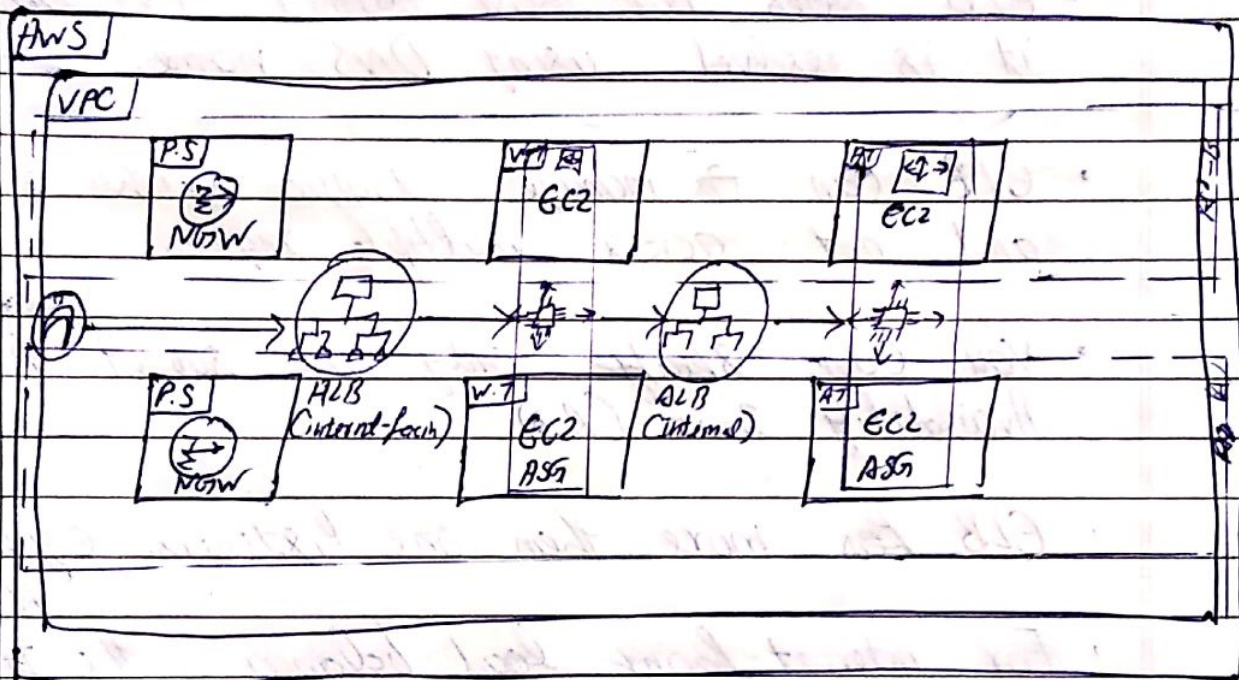


Day 09/100

P.S → Public Subnet
 W.T → Web Tier
 A.T → App Tier

AWS - Application Load Balancer (ALB)



Application Load Balancer (ALB) works at the layer-7 (Application Layer Request level) of the OSI model. ALB supports load balancing of application using HTTP and HTTPS protocols.

ALB simplifies and improves the security of your application by ensuring

Features:

- **Layer-7 Load Balancing** - You can load balance HTTP/HTTPS traffic to target - EC2 instances, microservices, and Containers based on request attributes.
- **Security Features** - You can associate security group with ALB to provide additional security options.
- **Web Application Firewall** - You can use AWS-WAF to protect your web application on your ALB.
- **HTTP Support** - ALB supports HTTPs termination w/ the client and the Load Balancer.
- **SSL/TLS offloading** - You can create HTTPS listener, which uses encrypted connections (SSL offload). ALB supports client TLS session termination. AWS Certificate Manager (ACM) or IAM can be used to manage your server certificates. You can choose from predefined security policies for your TLS listeners in order to meet compliance and security standards.

- **Sticky Sessions** - Sticky sessions are a mechanism to route requests from the same client to the same target. ALB supports both duration-based cookies and application-based cookies. Sticky sessions are enabled at the target group level.
- **Native IPv6 Support** - ALB supports native IPv6 in a VPC. This will allow client to connect to the ALB via IPv4 or IPv6.
- **HTTP/2 Support** - HTTP/2 support is enabled natively on an ALB. Clients that support HTTP/2 can connect to ALB over TLS.
- **WebSockets Support** - WebSocket and Secure WebSocket support is available natively and ready for use on ALB. It allows a server to exchange real-time message with end-users without the end users having to request (or poll) the server for an update.
- **gRPC Support** - ALB can route and load balance gRPC traffic b/w microservices or b/w gRPC enabled client.

and services. gRPC uses HTTP/2 for transport and is becoming the protocol of choice for interservice communications in microservices architectures.

- **Redirects** - ALB can redirect an incoming request from one URL to another URL. For example, redirect from HTTP to HTTPS and redirecting from an old version of an application to a new version.
- **Fixed Response**: ALB can control which client request are served by your application. This enables you to respond to incoming requests with HTTP error response codes and custom error message from the load balancer itself, without forwarding the request to the application.
- **Server Name Indication (SNI)**: You can use SNI to serve multiple secure website using a single TLS listener. Meaning, you can associate multiple certificates for the same domain to a secure listener. If the hostname in the client matches multiple certificates, the load balancer selects the best certificate to use based on a smart selection algorithm.

- **IP addresses as Targets**: You can load balance any application hosted in AWS (Single VPC, Peered VPCs and EC-2 classic or on-premises (over a Direct Connect or VPC Connection) using IP address of the application backends as targets.

- **Lambda function as Targets** - You can register Lambda function as target for a load balancer and leverage the support for Content-based routing rules to route request to different Lambda functions over both HTTP & HTTPS protocol.

- **Content-based Routing** - If your application is composed of several individual services, an ALB can route a request to a service based on the Content of the request such as Host field, Path URI, HTTP Header, HTTP method, Query string or Source IP address.

- **Host-based Routing**: You can route a client request on the Host field of the HTTP header allowing you to route to multiple domains from the same load balancer.

- Path-based Routing: You can route a client request based on the URI path of the HTTP header.
- HTTP ~~method~~ ^{header}-based routing: You can route a client request based on the value of any standard or custom HTTP header.
- HTTP method-based routing: You can route a client request based on any standard or custom HTTP method.
- Query String Parameter-based routing: You can route a client request based on query string or query parameters.
- Source IP address CIDR-based routing: You can route a client request based on source IP address CIDR from where the request originates.

• Containerized Application Support - ALB provides enhanced container support by load balancing across multiple hosts on a single Amazon EC2 instance. ALB has also deep integration with Amazon EC2 Container Services (ECS).

• Slow Start Mode with Load-Balancing Algorithm -

ALB supports a slow start mode with round-robin algorithm that allows you to add new targets without overwhelming them with a flood of requests. With the slow start mode, targets warm up before accepting their fair share of request based on ramp-up period that you specify.

• **User Authentication** - You can offload the authentication functionality from your apps into ALB. ALB will securely authenticate users as they cloud application. Application Load Balancer is seamlessly integrated with Amazon Cognito.

• **Request Tracing** - ALB injects a new Custom Identifier "X-Amzn-Trace-ID" HTTP header on all requests coming in to the load balancer. Request tracing allows you to track a request by its Unique ID.

Outpost Support → ALB supports AWS Outposts a fully managed service that extends AWS infrastructure, services, and tools to virtually any data center, co-location space or ~~on~~ on-premises facility for a truly consistent hybrid experience.

Key Points:

1. ALB operates at request level.
2. You must specify two subnets from at least two Availability Zones to increase availability of ALB.
3. You can specify only one subnet per Availability Zone.
4. You can perform load balancing for the TCP ports: 1-65535.
5. Cross-Zone load balancing is already enabled by default on ALB.
6. Request tracing is enabled by default on ALB.
7. You can use AWS CloudTrail to get history of ALB API calls made on account.

8. You are charged for each hour or particular hour that an ALB is running and number of Load Balancer Capacity Units (LCU) used for hour.

Use Cases:

- Typically used for web application for advanced load balancing of HTTP & HTTPS traffic.
- When you need advanced routing capabilities (e.g. host-based, path-based routing) targeted at delivery of modern application architectures including microservices and container-based applications.
- When you need SSL offloading at load balancer level.