**Elastic Load Balancing**

* + Elastic Load Balancing automatically distributes incoming application traffic across multiple targets, such as EC2 instances, containers and IP addresses
  + 3 types of ELB's
    - Classic LB
      * Oldest, provides basic LB at both layer 4 and 7
    - Application LB
      * Layer 7 - routes connections based on the content of the request
    - Network LB
      * Layer 4 - routes connections based on IP protocol data
  + Provides fault tolerance for your applications by automatically balancing traffic across targets(ec2 instances, containers and IP addresses) and AZ's while ensuring only healthy targets receive traffic
  + An ELB can distribute incoming traffic across your EC2 instances in multiple AZ's
  + Route53 can be used for region load balancing
  + ELBs can be internet facing or internal-only
    - internet facing:
      * ELB nodes have public IPs
      * Routes traffic to the private IP addresses of the EC2 instances
      * Need one public subnet in each AZ where the ELB is defined
    - Internal-only ELB
      * Private IPs
      * Routes traffic to the private IP of the EC2 instances
      * do not need Internet Gateway
  + An ELB forwards traffic to eth0 (primary IP address)
  + An ELB listener is the process that checks for connection requests
    - CLB options for TCP and HTTP/HTTPS
    - ALB options for HTTP and HTTPS
    - NLB options for TCP
  + ELB SG's
    - Provide control of ports and protocols that can reach the front end listener
  + DDOS
    - ELB automatically distributes incoming traffic across multiple targets
    - ELB, like CF, only support valid TCP requests, so DDoS attacks such as UDP and SYN floods are not able to reach the EC2 instances
    - ELB also offers a single point of management and can serve as a line of defense between the internet and your backed
  + Monitoring
    - CW - every 1 minutes
  + Access Logs
    - Disabled by default
    - Includes information about the clients
  + CloudTrail
    - Can be used to capture API calls to the ELB
    - Can be stored in S3 bucket
  + CLB
    - Basic LB at both request and connection level
    - Operates at layer 4 and 7
    - Supported protocols: TCP, SSL, HTTP, HTTPS
    - Cross-Zone load balancing
      * enabled by default for CLB and ALB but not for NLB
    - Supports single X.509 certificate
  + ALB
    - Layer 7
    - You can LB HTTP/HTTPS applications and use layer 7 specific features, such as X-Forwarded-For headers
    - Native IPv6 support
    - Better support for **real-time streaming**
    - Request tracing allows you to track a request by its unique IDHealth checks with CW metrics
    - User authentication with Cognito
    - Monitoring
      * CT can be used to capture API calls. Only pay for S3 charges
      * Access logs are used to monitor actions such as request received, client's IP address, request paths
        + Access logging is disabled by default
        + Only charged for S3 storage
    - Target groups - a logical grouping of targets(EC2 instances or ECS)
      * targets are endpoints and can be EC2 instances, ECS containers or IP addresses
      * Target groups can exist independently from the ALB
      * Target groups can have up to 1000 targets
      * A single target can be in multiple target groups
      * Only 1 protocol and 1 port can be defined per target group
      * Cannot use public IP addresses as targets
      * Cannot use instance IDs and IP address targets within the same target group
      * A target group can only be associated with one load balancer
      * TG's are regional
      * ASG's can scale each target group individually
    - Listeners and Rules
      * Listeners
        + Each ALB needs at least one listener and can have up to 10
        + Listeners define the port and protocol to listen on
        + Can add one or more listeners
        + Cannot have the same port in multiple listeners
      * Rules
        + Rules determine how the load balancer routes requests to the targets in one or more target groups
        + Each rule consists of a piority, one or more actions, an optional host condition and an optional path condition
        + Only one action can be configured per rule
        + At least one rule is required
        + Up to 100 rules per ALB
        + Rules are defined on listeners
        + Each listener has a default rule
        + If no rules are found the default rule will be followed which directs draffic to the default target groups
        + Rule priority

Each rule has a pioritiy and are evaluated lowest to highest

default rule is evaluated last

* + - * + Rule action

Only one target group per action

Each rule has a type and a target group

* + - * + Rule conditions

There are two types of rule condition: host and path

When the conditions for a rule are met an action is take

* + - * + Request routing

After the LB receives a request it evaluates the listener rules in priority in order to determine which rule to apply and then selects a target from the group for the rule action using the round robin routing algorithm

* + - Content-based routing
      * ALB can route requests based on the content of the request in the host field: host-based or path-based
        + Host-based is domain based routing

example.com

app1.example.com

* + - * + Path-based is URL based

example.com/images, example.com/app1

* + - ALB and ECS
      * ECS service maintains the 'desired count' of instances
      * ALB integrates with EC2 container service using service load balancing
      * Federated authentication:
        + ALB now supports authentication from OIDC comnpliant identity providers such as Google, FB and Amazon
        + Implemented through an authentication action on a listener rule that integrates with Cognito to create user pools
        + AWS SAM can also be used with Cognito
  + Network Load Balancer(NLB)
    - Operates at layer 4
    - Architected to handle millions of requests/sec, sudden volatile traffic patterns and provides extremely low latencies
    - NLB supports these features:
      * WebSockets
      * TLS termination
      * Preserves the source IP of the clients
      * Long-running connections are very useful for websocket type applications
    - High throughput - designed to handle traffic as it grows and can load balance millions of requests/second
    - Uses static IP
      * Provides single IP address for each AZ
    - Integration with Route53 enables the removal of a failed load balancer IP address from service and subsequent redirection of traffic to an alternate NLB in another region
    - Supports cross-zone load balancing
    - Enhanced logging - can use the Flow Logs feature to record all requests sent to your LB\

Graphical user interface, text, application, email

Description automatically generated

Diagram

Description automatically generated