Elastic Load Balancing Deep Dive

Narayan Subramaniam and David Pessis

aws

Elastic Load Balancing automatically distributes incoming application traffic across multiple targets, such as Amazon EC2 instances, containers, and IP addresses.

The Elastic Load Balancing Family

Application Load Balancer

HTTP & HTTPS (VPC)

Network Load Balancer

TCP Workloads (VPC)

Classic Load Balancer

Previous Generation
for HTTP, HTTPS, TCP
(Classic Network)

aws

Elastic

Secure

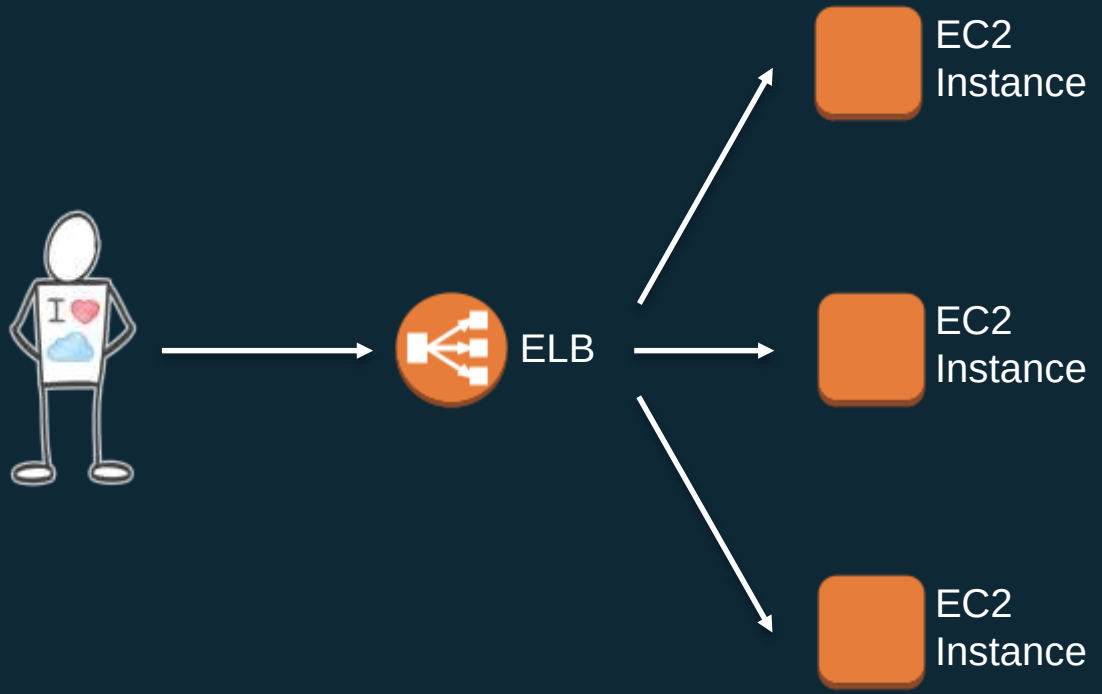
Integrated

Cost Effective

aws

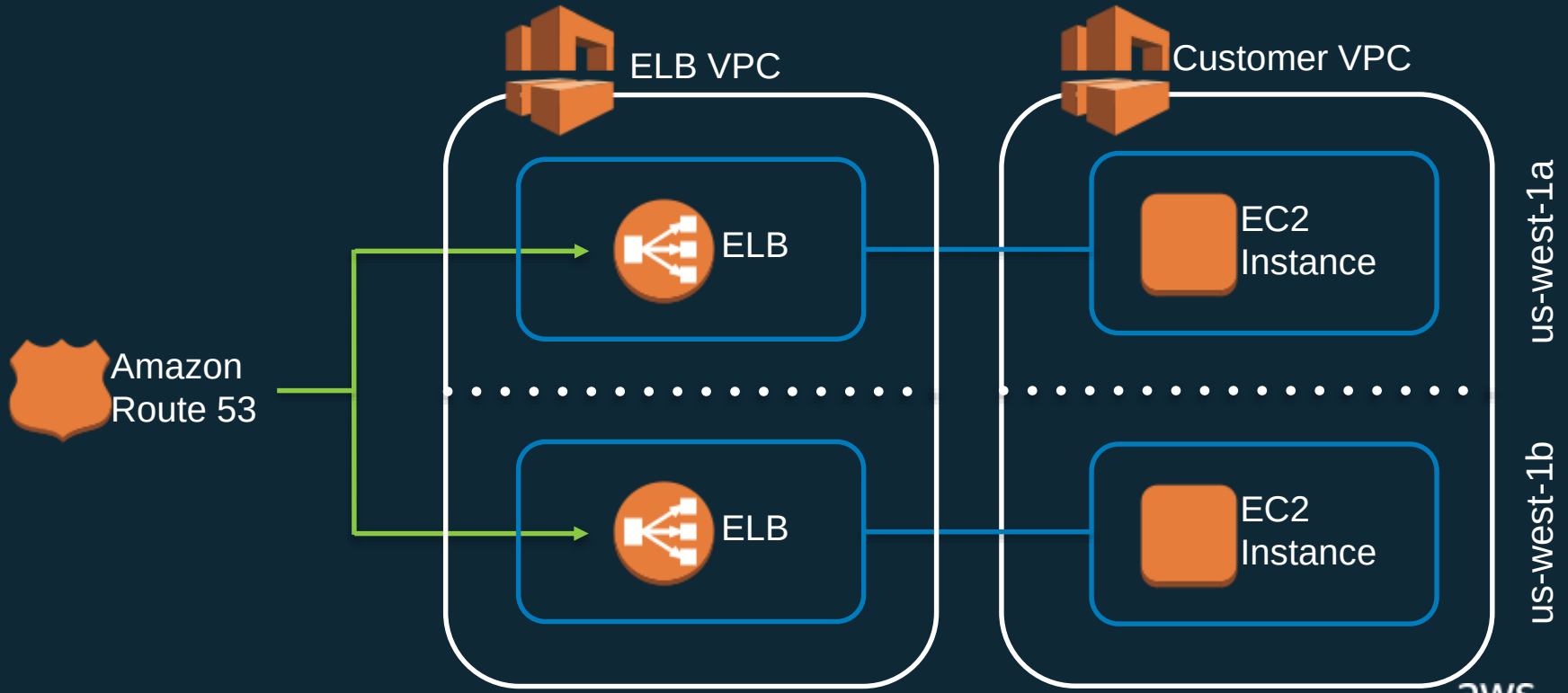
EC2
Instance

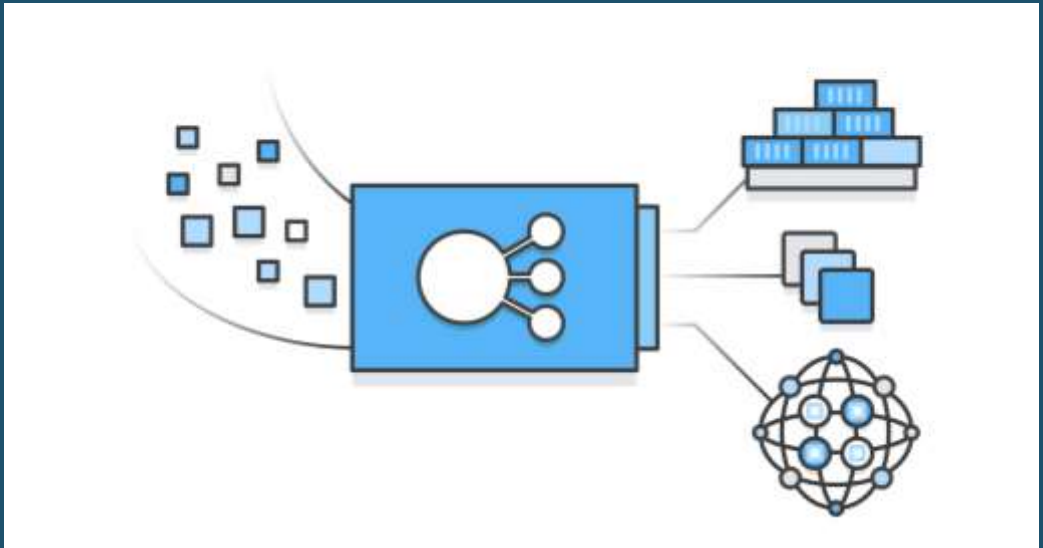
aws



Load balancer used to route incoming requests to multiple EC2 instances, Containers, or IP addresses in your VPC.

Architecture





Application Load Balancer

Advanced request routing with support for microservices and container-based applications.

aws

Application Load Balancer

New, feature rich, layer 7 load-balanced platform

Content-based routing allows requests to be routed to different applications behind a single load balancer

Support for microservices and container-based applications, including deep integration with Elastic Container Service

aws

Application Load Balancer

Support for WebSockets and HTTP/2

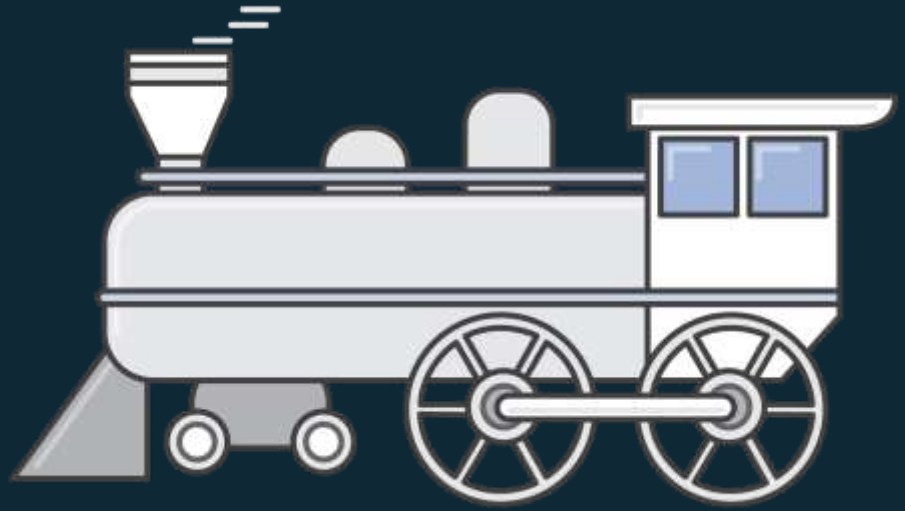
Path and Host Based Routing

Improved health checks and additional CloudWatch metrics

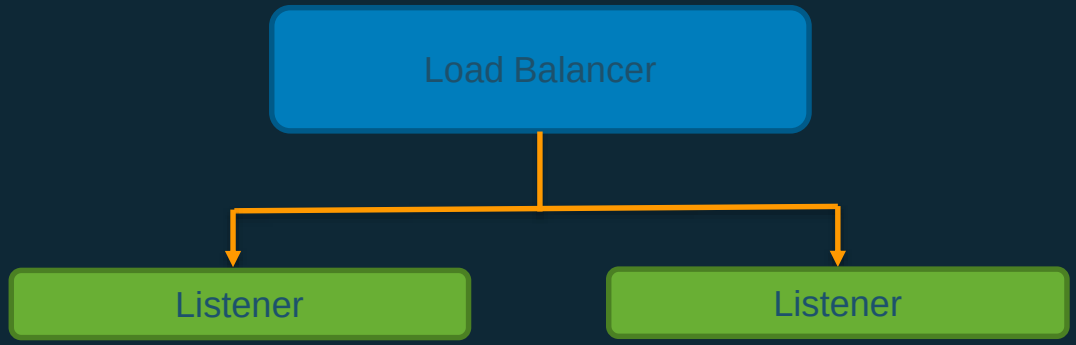
Improved performance for real-time and streaming applications

Improved Elastic Load Balancing API

Load balancer API deletion protection



aws



Listeners

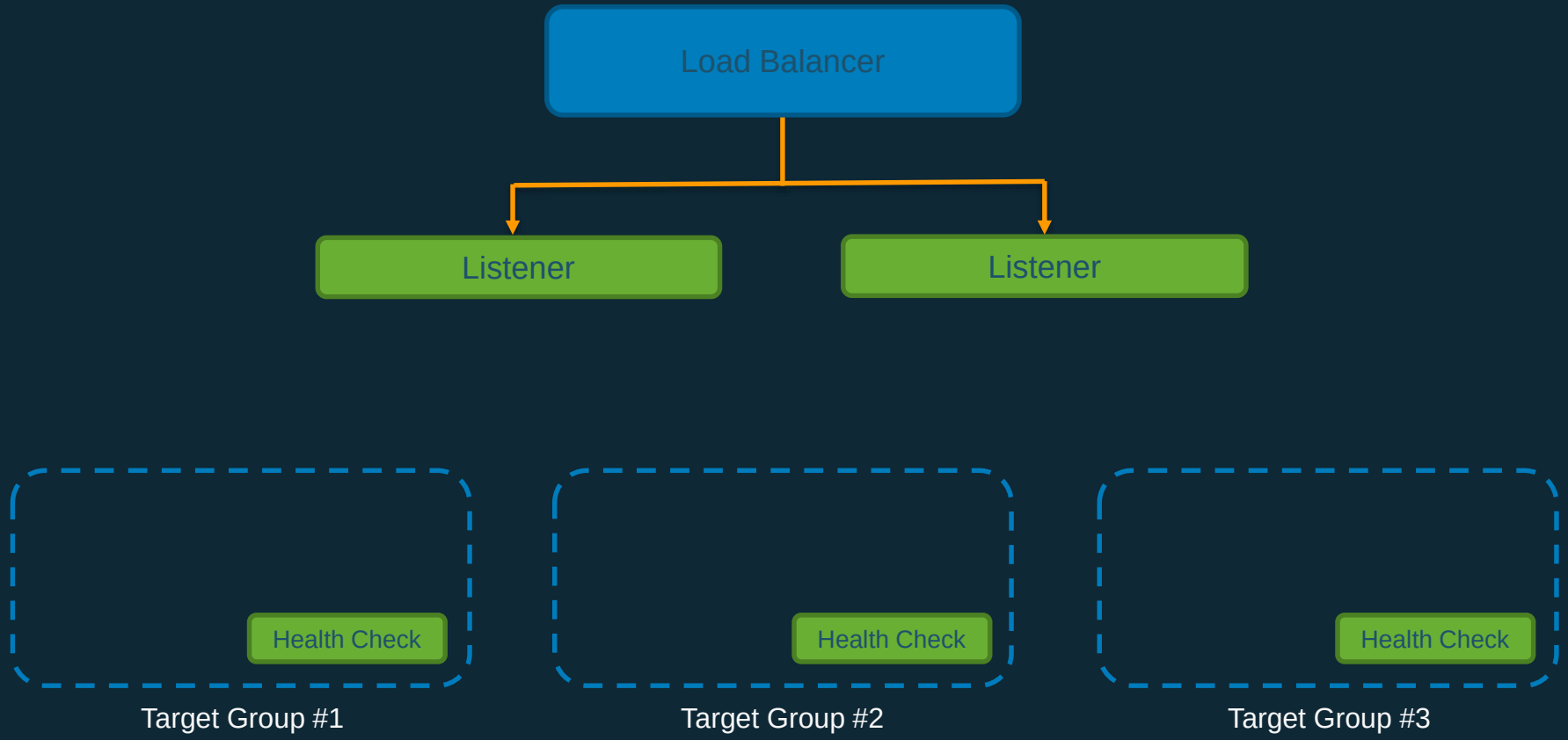
Define the port and protocol which the load balancer must listen on

Each Application Load Balancer needs at least one listener to accept traffic

Each Application Load Balancer can have up to 50 listeners

Routing rules are defined on listeners

aws



Target groups

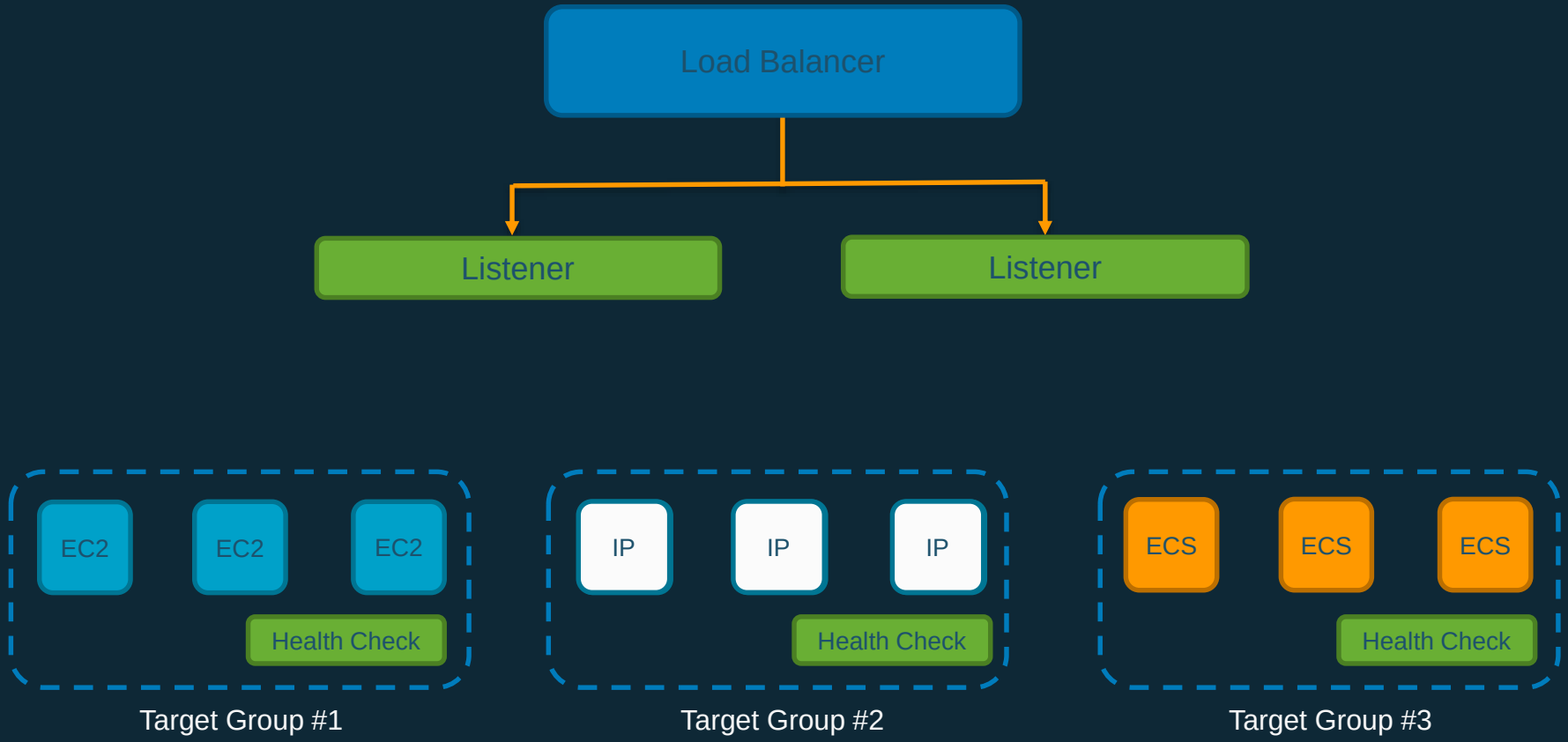
Logical grouping of targets behind the load balancer

Target groups can exist independently from the load balancer

Regional construct that can be associated with an Auto Scaling group

Target groups can contain up to 1,000 targets

aws



Targets

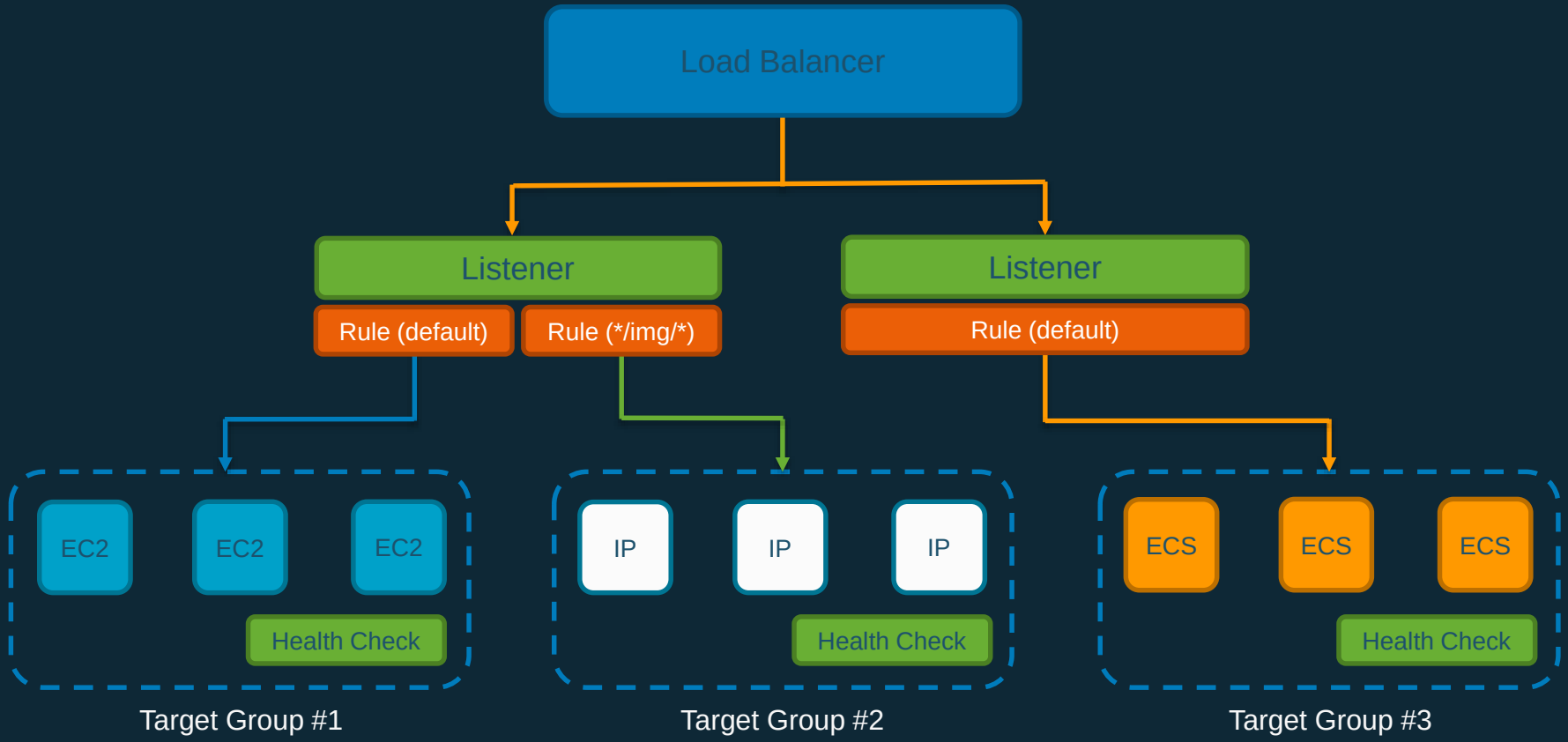
Support for EC2 instances and ECS containers, and IP Addresses.

EC2 instances can be registered with the same target group using multiple ports

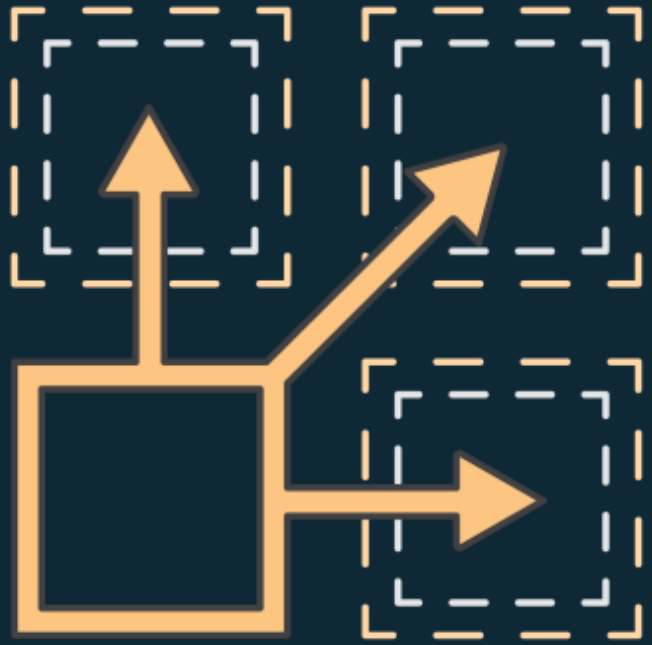
A single target can be registered with multiple target groups

IP Addresses both accessible within your VPC or via DX and VPN

aws



Rules



Each listener can have one or more rules for routing requests to target groups.

Rules consist of conditions and actions

When a request meets the condition of the rule, the action is taken

Today, rules can forward requests to a specified target group

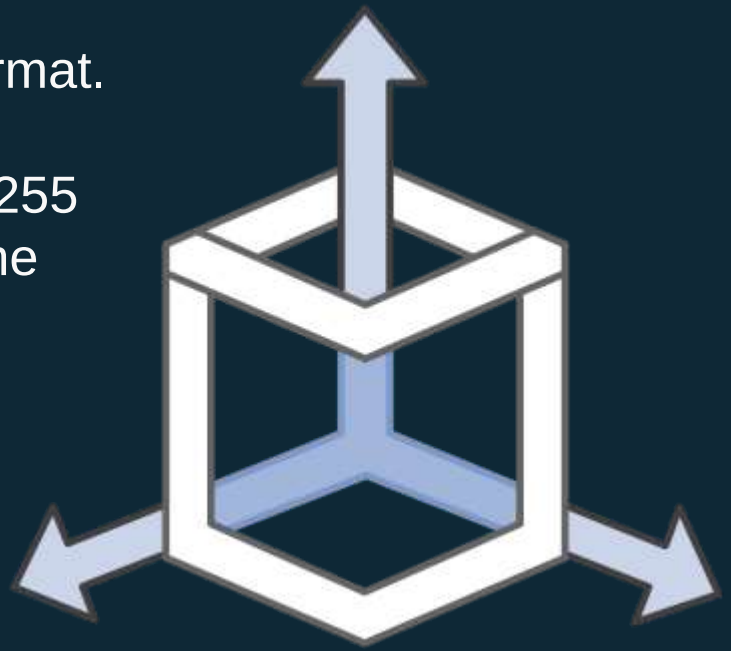
aws

Rules (continued)

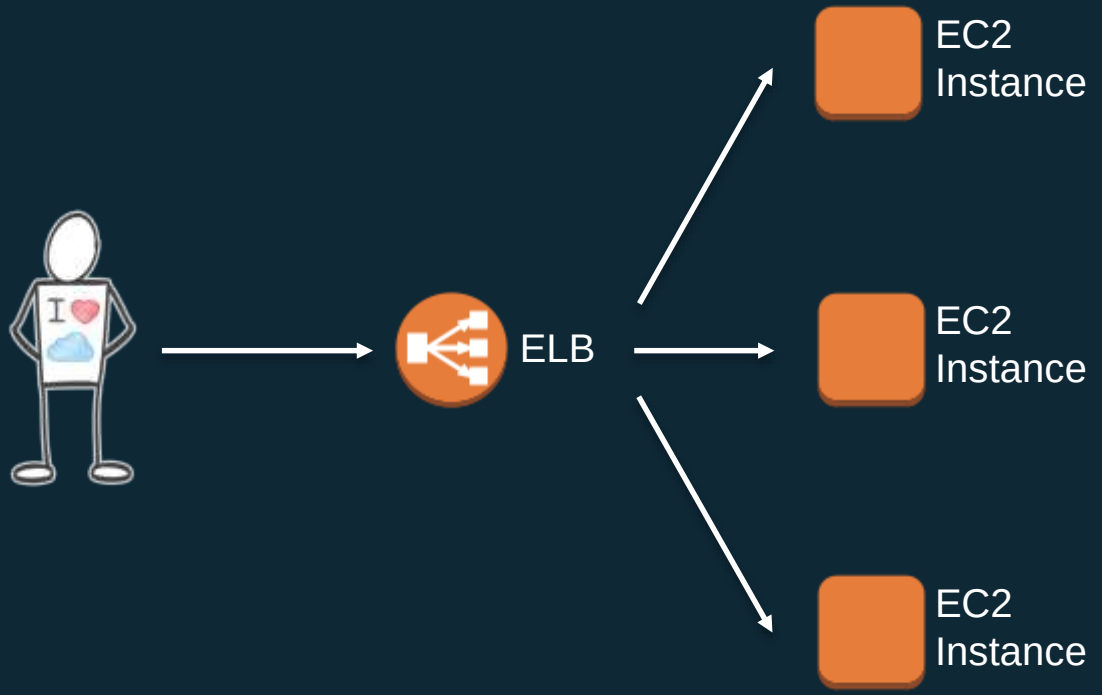
Conditions can be specified in path pattern format.

A path pattern is case sensitive, can be up to 255 characters in length, and can contain any of the following characters:

- A-Z, a-z, 0-9
- _ - . $ / ~ " ' @ : +
- & (using &)
- * (matches 0 or more characters)
- ? (matches exactly 1 character)

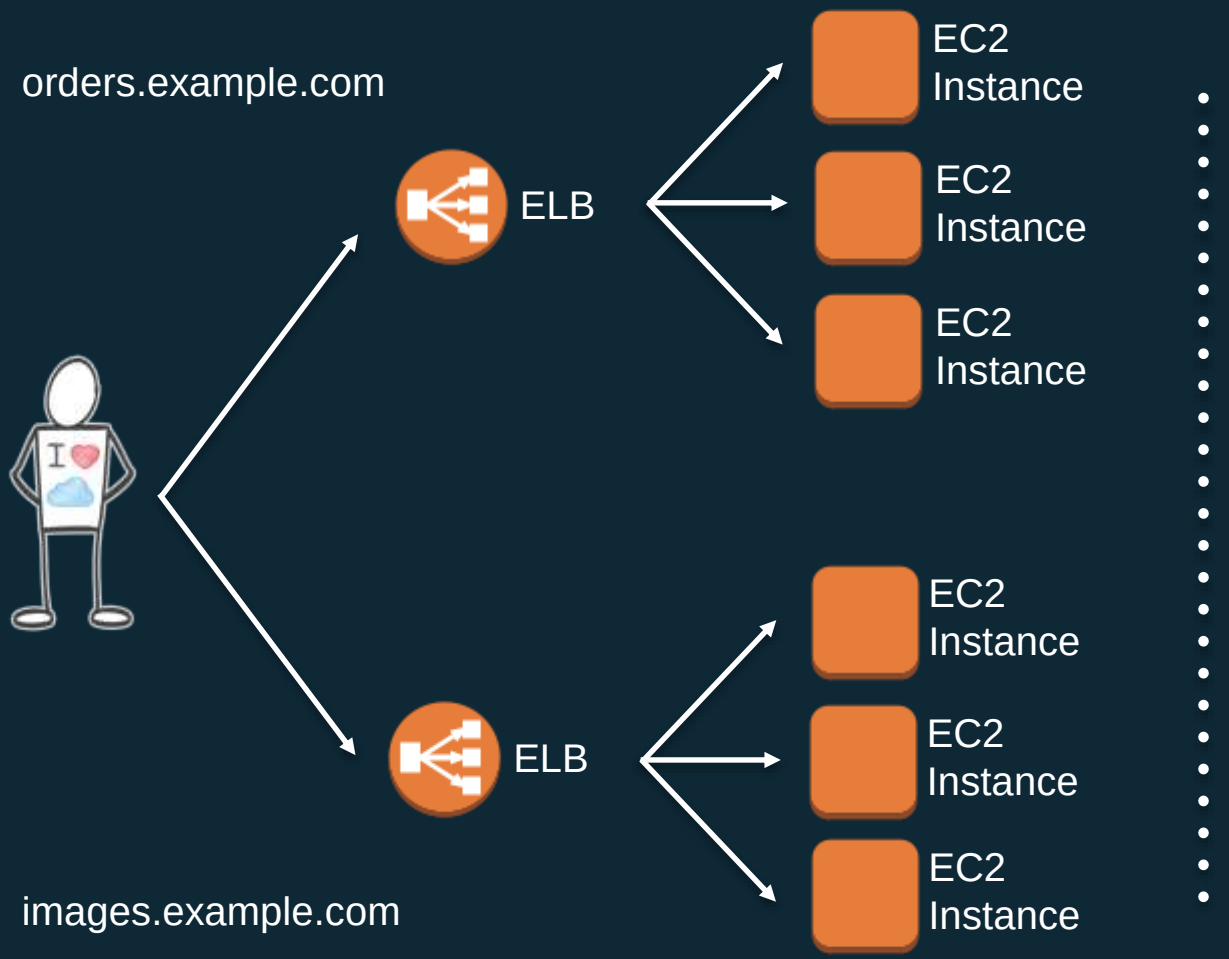


aws

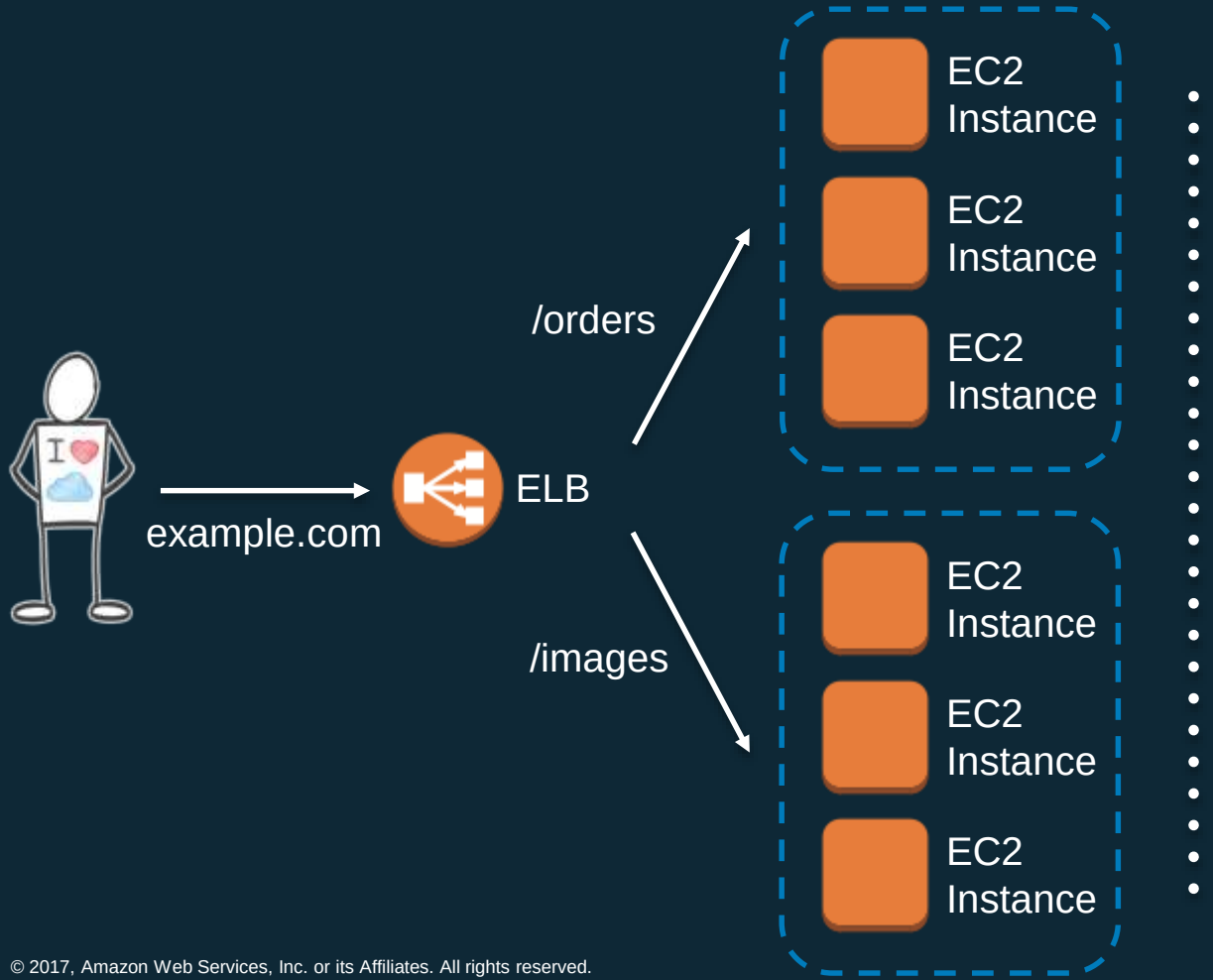


Amazon EC2 instances registered behind a Classic Load Balancer

aws



Running two separate services with Classic Load Balancer

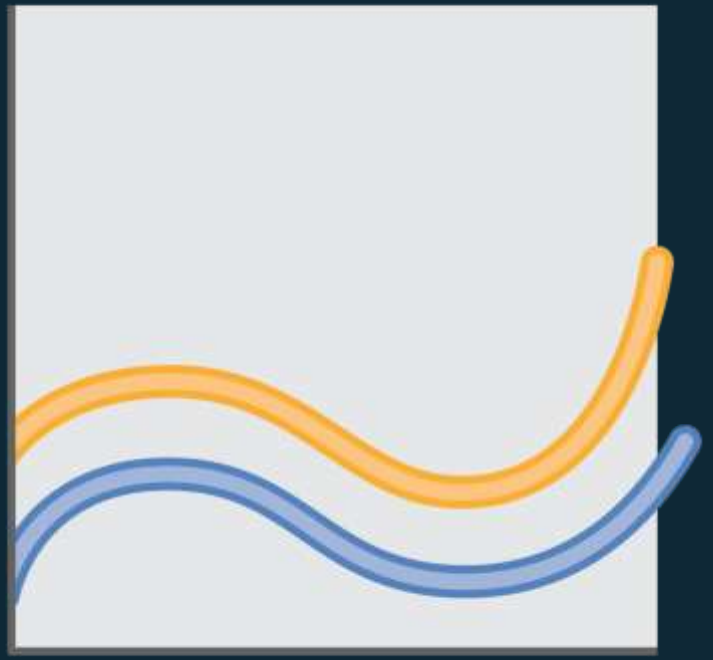


Application Load Balancer allows for multiple services to be hosted behind a single load balancer

Auto Scaling integration

Auto Scaling can now scale targets within a target group

Allows for applications to be scaled independently behind the Application Load Balancer



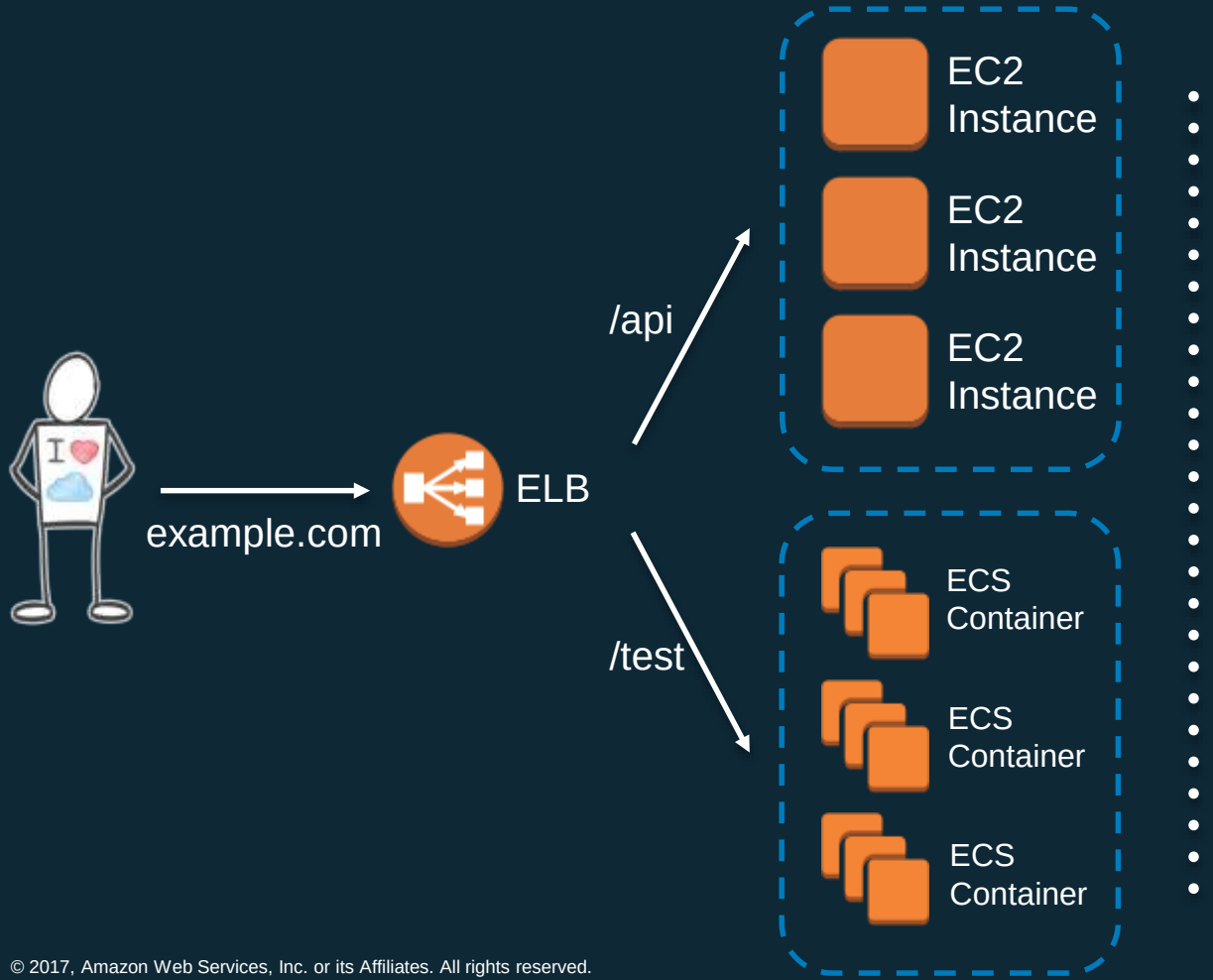
ECS integration

Application Load Balancer (ALB) is fully integrated with Amazon EC2 Container Service (Amazon ECS), managing target groups, paths, and targets

ECS will automatically register tasks with the load balancer using a dynamic port mapping

Can also be used with other container technologies

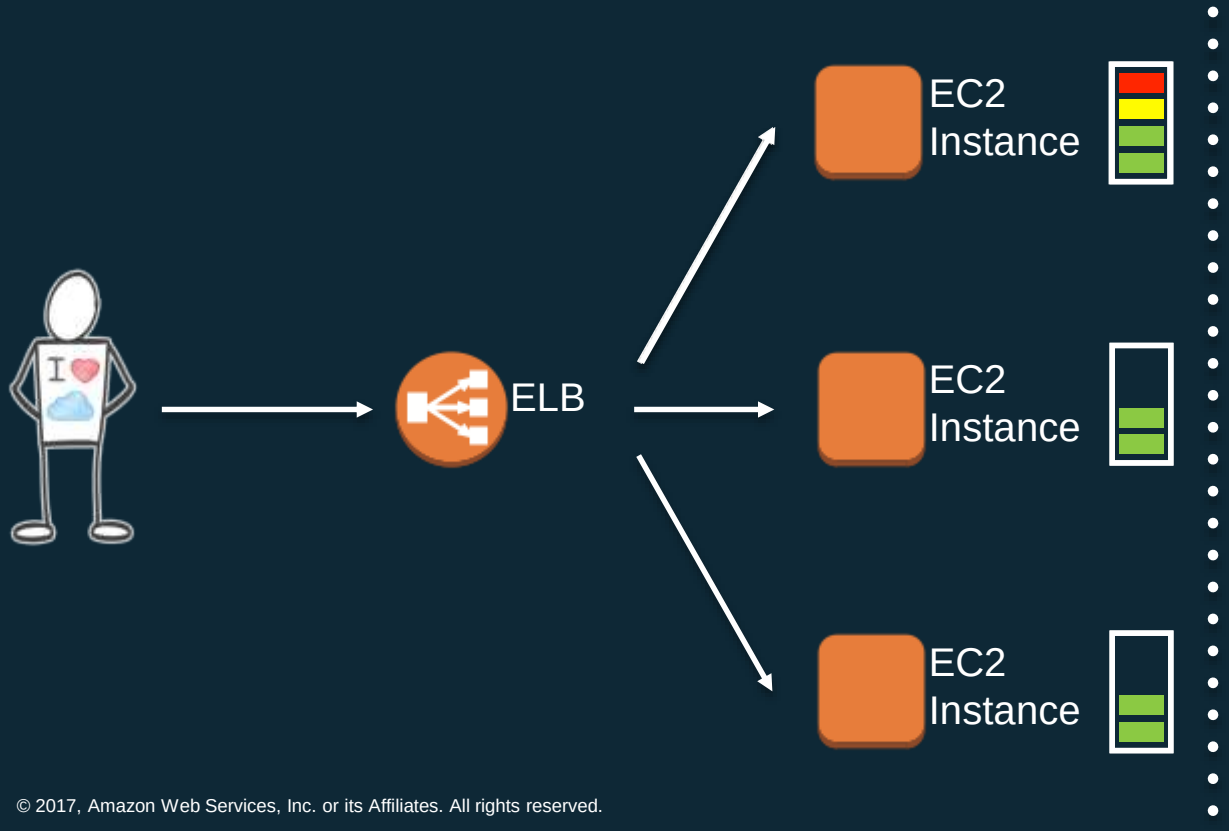
aws



Application Load Balancer allows containers to be included in the target group

Health checks allow for traffic to be shifted away from failed instances

Health checks



Health checks ensure that request traffic is shifted away from a failed instance.

Health checks

Support for HTTP and HTTPS health checks.

Customize the frequency and failure thresholds.

Consider the depth and accuracy of your health checks.

aws

Health checks

Customize list of successful response codes, for example 200-300

Details of health check failures are now returned via the API and Management Console

aws

Host-based Routing

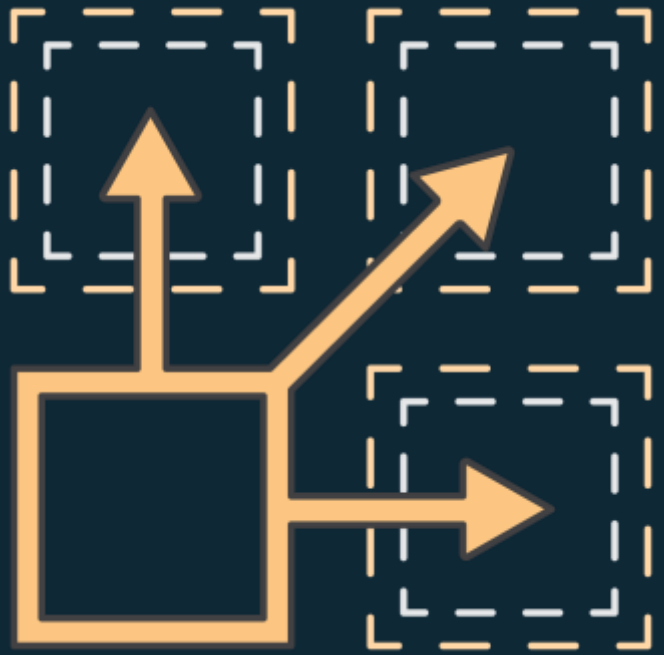
Route based on host field in the HTTP header

Support multiple domains using a single load balancer

Route each host name to a different target group

Combine host-based routing and path-based routing

- 128-character limit
- A-Z, a-z, 0-9, -, .
- * (matches 0 or more characters)
- ? (matches exactly 1 character)



Predefined Security Policies

ELBSecurityPolicy-TLS-1-1-2017-01 – Supports TLS 1.1 and above

ELBSecurityPolicy-TLS-1-2-2017-01 – Strictly supports TLS1.2

ELBSecurityPolicy-2016-08 – New default policy - same as Classic Load Balancer default policy

Windows XP Security Policy

Native IPv6 support

Application Load Balancer with WAF

Monitor web requests and protect web applications from malicious requests at the load balancer

Block or allow requests based on conditions such as IP addresses

Preconfigured protection to block common attacks like SQL injection or cross-site scripting

Set up web ACLs and rules from WAF console and apply them to the load balancer

aws

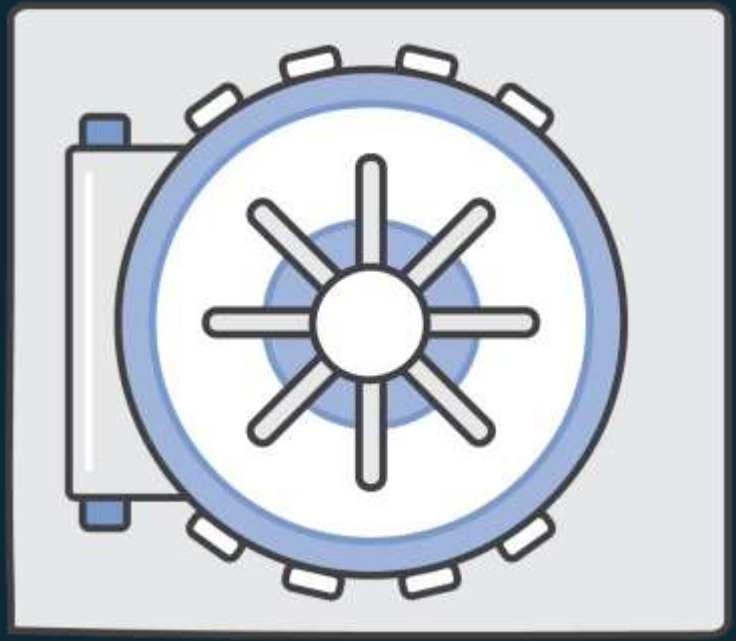
Server Name Indication (SNI)

Host multiple TLS secured applications, each with its own TLS certificate

Bind multiple certificates to the same secure listener on your load balancer

ALB will automatically choose the optimal TLS certificate for each client

Support for both the classic RSA algorithm and the newer, faster Elliptic-curve based ECDSA algorithm



aws

IP as a Target

Use any IPv4 address from the load balancer's VPC CIDR for targets within load balancer's VPC

Use any IP address from the RFC 6598 range (100.64.0.0/10) and in RFC 1918 ranges (10.0.0.0/8, 172.16.0.0/12, and 192.168.0.0/16) for targets located outside the load balancer's VPC (this includes Peered VPC, EC2-Classic, and on-premises targets reachable over Direct Connect or VPN).

aws

Cross-zone load balancing

Requests distributed evenly across multiple Availability Zones

Load balancer absorbs impact of DNS caching

Eliminates imbalances in backend instance utilization

No additional bandwidth charge for cross-zone traffic

Enabled on all ALBs

Amazon CloudWatch metrics

CloudWatch metrics provided for each load balancer.

Provide detailed insight into the health of the load balancer and application stack.

CloudWatch alarms can be configured to notify or take action should any metric go outside the acceptable range.

All metrics provided at the 1-minute granularity

aws

Access logs

Provide detailed information on each request processed by the load balancer

Includes request time, client IP address, latencies, request path, and server responses

Delivered to an Amazon S3 bucket every 5 or 60 minutes

aws

Application Load Balancer pricing

With the Application Load Balancer, you only pay for what you use. You are charged for each hour or partial hour your Application load balancer is running and the number of Load Balancer Capacity Units (LCU) used per hour

- $0.0225 per Application Load Balancer-hour (or partial hour)
- $0.008 per LCU-hour (or partial hour)

Hourly charge is 10% cheaper than Classic Load Balancer; reducing the cost for the virtually all of our customers

Load balancer capacity units

An LCU measures the dimensions on which the Application Load Balancer processes your traffic (averaged over an hour). The three dimensions measured are:

- New connections: up to 25 new connections per second
- Active connections: up to 3,000 active connections
- Bandwidth: Up to 2.22 Mbps (1 GB per hour)

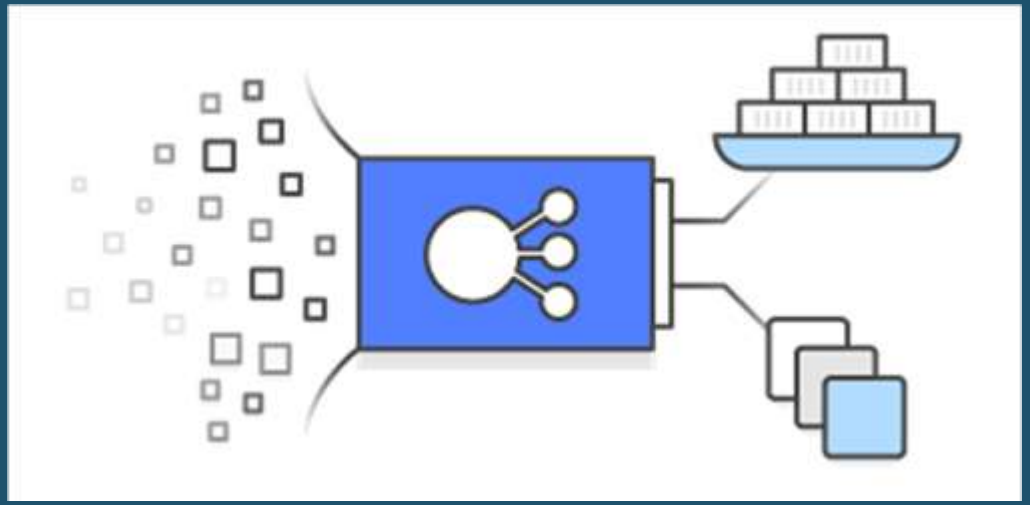
You are charged only on the dimension with the highest usage over the hour

Migrating to Application Load Balancer

Publishing LCU Metrics for Classic Load Balancer which Allows customers to estimate pricing if they migrate from Classic to ALB

Migration is as simple as creating a new Application Load Balancer, registering targets and updating DNS to point at the new CNAME.

Classic Load Balancer or Application Load Balancer migration utility: https://github.com/aws/elastic-load-balancing-tools



Network Load Balancer

Network Load Balancer

New, layer 4 load-balancing platform
- Connection-based load balancing
- TCP protocol

High Performance
- Can handle millions of requests per sec

Static IP Support

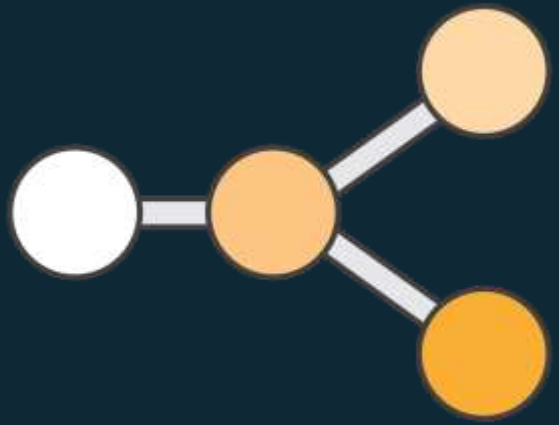
Ideal for applications with long running connections

aws

Network Load Balancer

- Extremely low latencies
- Preserves Source IP
- Same API as Application Load Balancer
- Load Balancer API Deletion Protection

Static IP

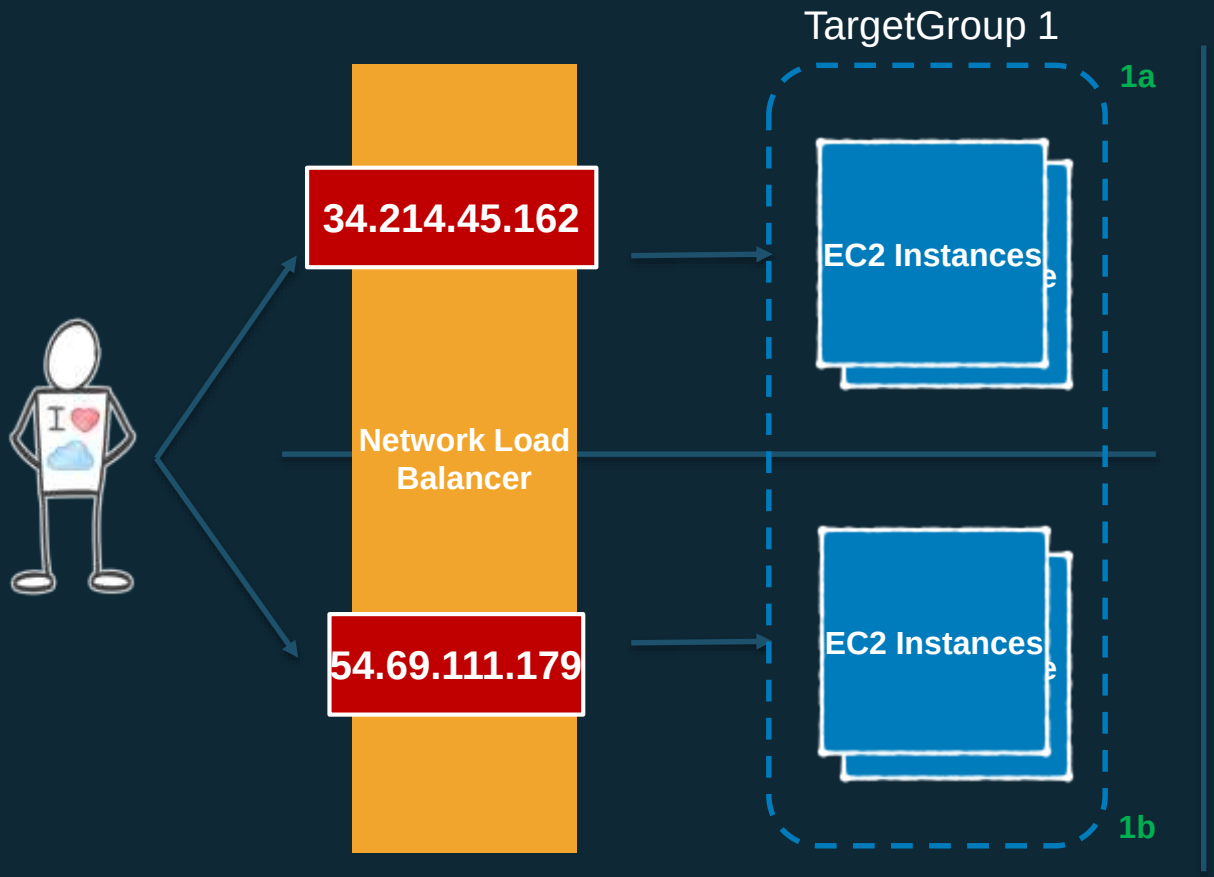


Automatically gets assigned a single IP per Availability Zone

Assign an EIP per AZ to get Static IP

Helps with white-listing for firewalls and zero dollar billing use cases

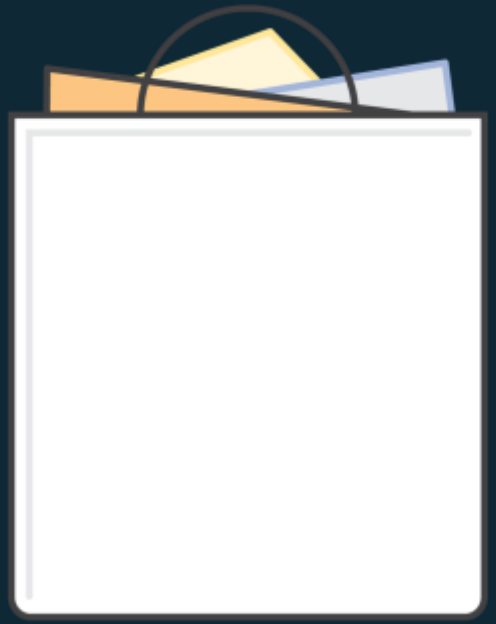
Assign Elastic IP Addresses



Assigning Elastic IP provides a single IP address per Availability Zone per load balancer that will not change.

aws

Preserve Source IP

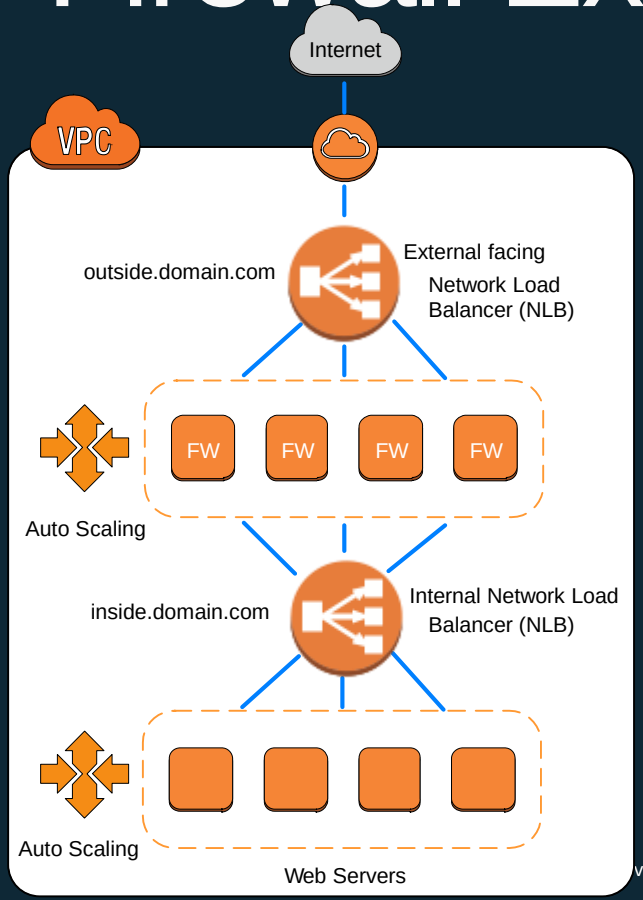


Preserves Client IP to back-ends

Can be used for logging and other applications

Removes need for Proxy Protocol

Firewall Example with NLB



External facing NLB uses less addresses
- Used for Firewalls, proxies or 3rd party load balancers

Preserves source IP helping firewalls with features like Geo-IP blocking

Internal NLB doesn't change IPs
- Allows Firewalls, WAFs and proxies to maintain a single addresses for NAT

aws

Resources same as ALB

Improved Elastic Load Balancing API

Listeners

Target Groups

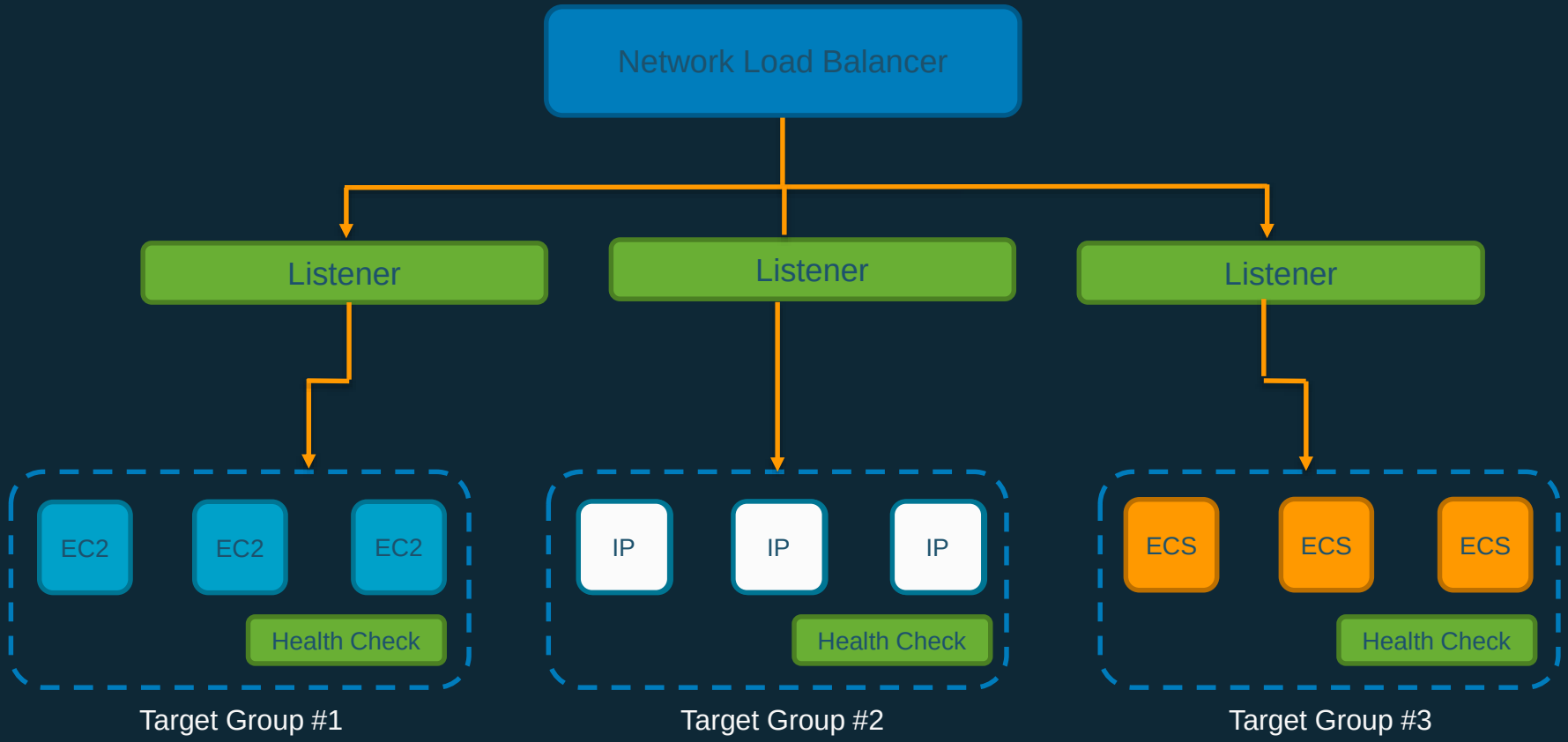
Targets

IP as a Target

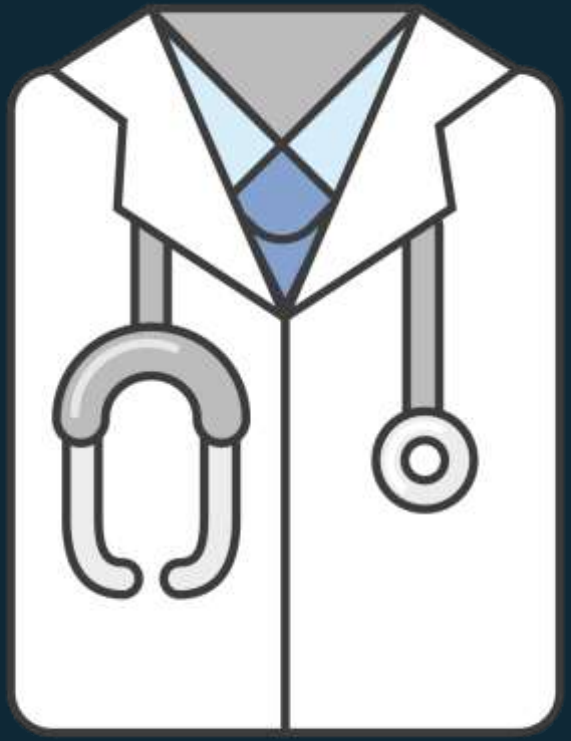
Use any IPv4 address from the load balancer's VPC CIDR for targets within load balancer's VPC

Use any IP address from the RFC 6598 range (100.64.0.0/10) and in RFC 1918 ranges (10.0.0.0/8, 172.16.0.0/12, and 192.168.0.0/16) for targets located outside the load balancer's VPC (on-premises targets reachable over Direct Connect)

aws



Health Checks



Supports both Network and Application Target health checks

Network health checks

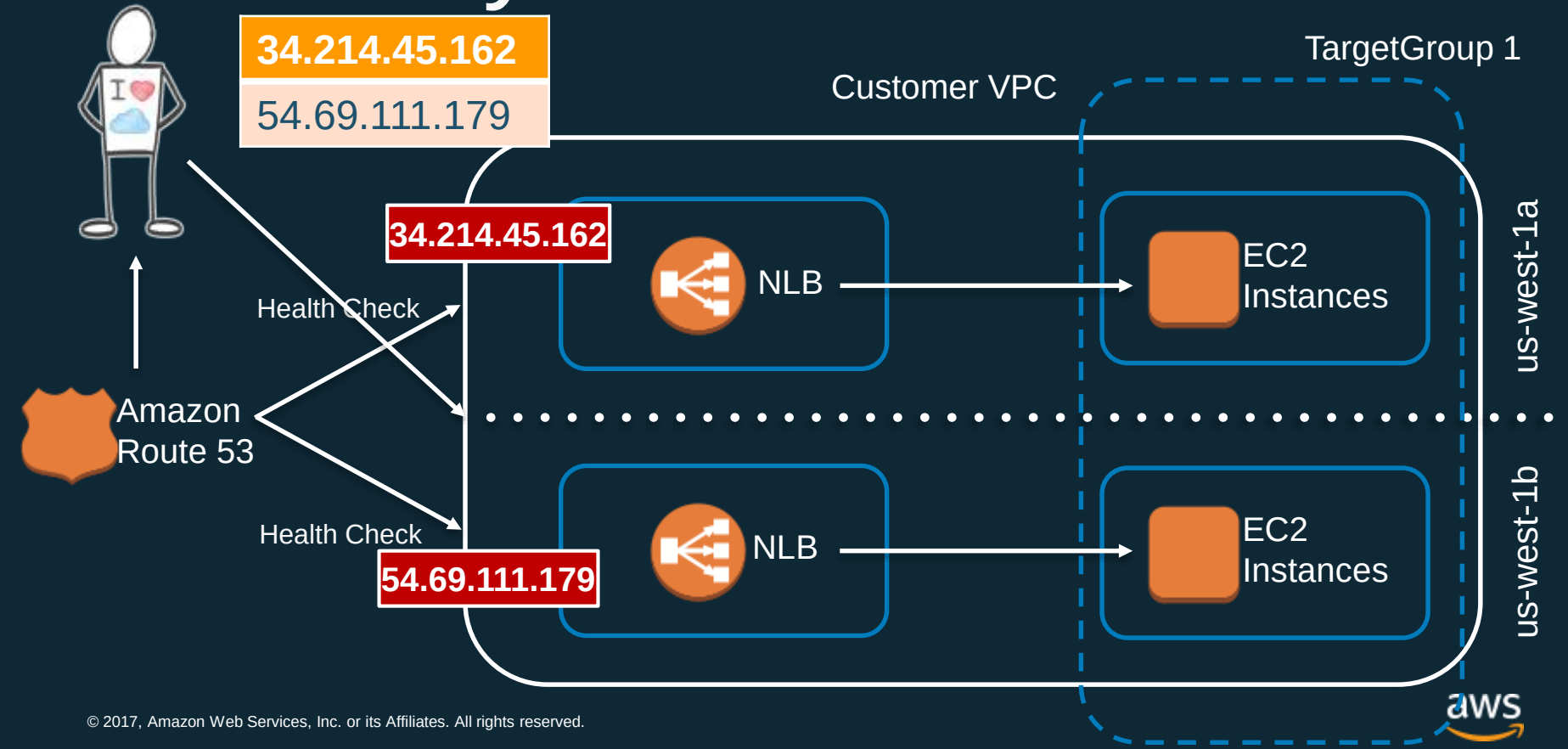
- Based on overall response of your target to normal traffic
- Will fail unresponsive targets in millisec

Application level health checks

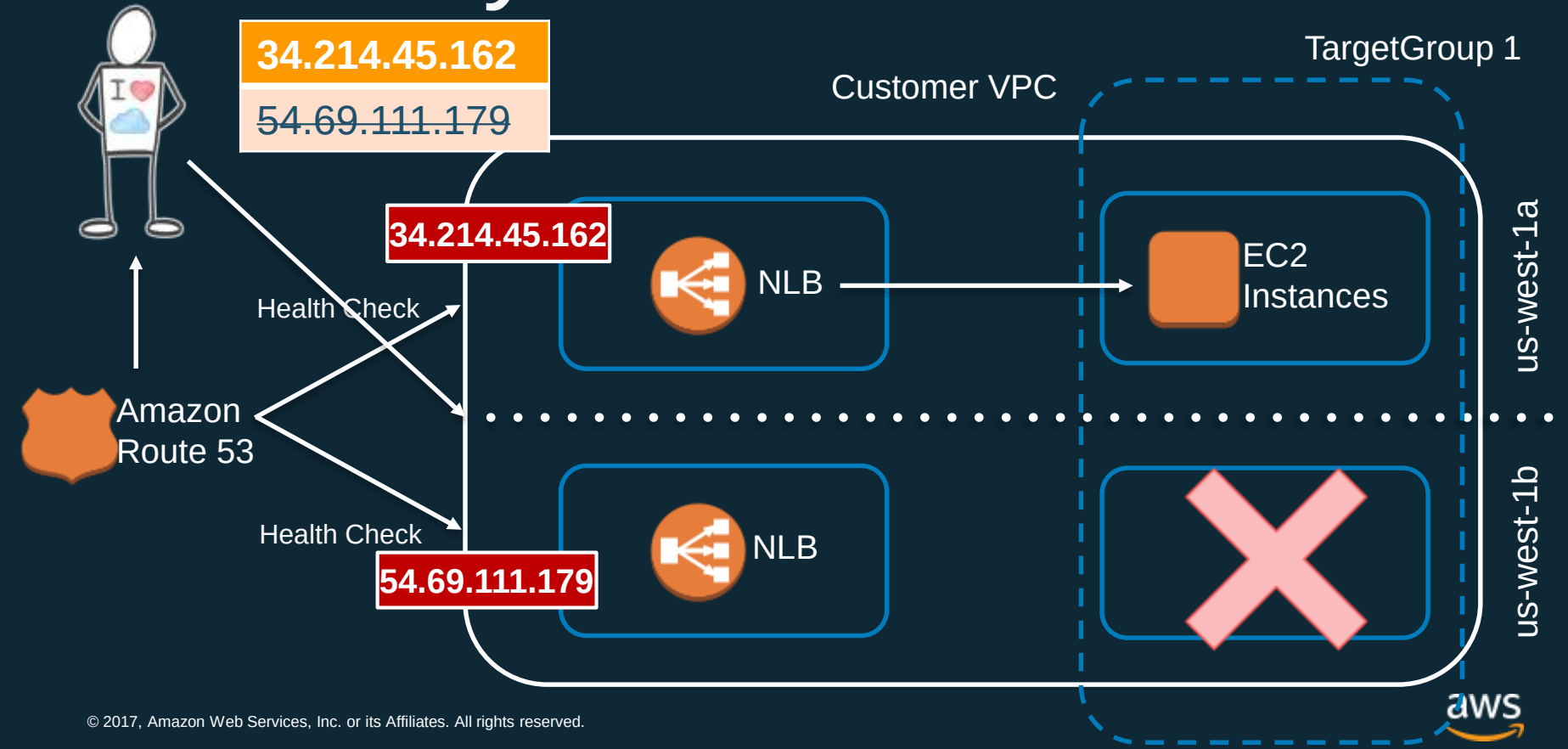
- HTTP, HTTPS and TCP HC
- Customize frequency, failure thresholds

aws

Availability Zone Fail-over



Availability Zone Fail-over



Amazon CloudWatch metrics

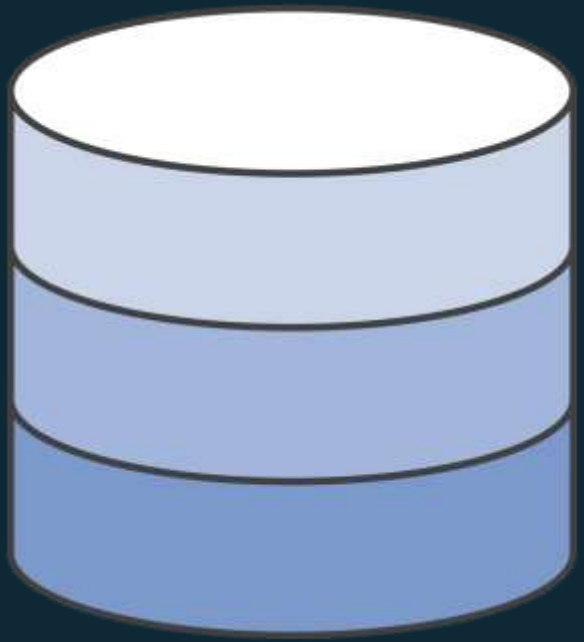
CloudWatch metrics provided for each load balancer.

Provide detailed insight into traffic and capacity, errors and back-end health for the Network Load Balancer

CloudWatch alarms can be configured to notify or take action should any metric go outside the acceptable range.

All metrics provided at the 1-minute granularity.

Flow Logs



Captures the network flow for a specific 5-tuple, for a specific capture window

- Packets
- Bytes
- Capture window start and end
- Action - Accepted or Rejected
- status
- Log Status

aws

Network Load Balancer pricing

With the Network Load Balancer, you only pay for what you use. You are charged for each hour or partial hour your Network load balancer is running and the number of Load Balancer Capacity Units (LCU) used per hour

- $0.0225 per Network Load Balancer-hour (or partial hour)
- $0.006 per LCU-hour (or partial hour)

Hourly charge is 10% cheaper than Classic Load Balancer; Data Processing charge is 25% cheaper than Classic and Application Load Balancer; reducing the cost for virtually all of our customers

Load balancer capacity units

An LCU measures the dimensions on which the Network Load Balancer processes your traffic (averaged over an hour). The three dimensions measured are:

- New connections: up to 800 new connections per second
- Active connections: up to 100,000 active connections
- Bandwidth: Up to 2.22 Mbps (1 GB per hour)

You are charged only on the dimension with the highest usage over the hour

Migrating to Network Load Balancer

Migration is as simple as creating a new Network Load Balancer, registering targets and updating DNS to point at the new CNAME.

Classic Load Balancer to Network Load Balancer migration utility: https://github.com/aws/elastic-load-balancing-tools

How do I pick the correct Load Balancer?

	Application Load Balancer	Network Load Balancer	Classic Load Balancer
Protocol	HTTP, HTTPS,HTTP/2	TCP	TCP, SSL, HTTP, HTTPS
SSL offloading	✓		✓
IP as Target	✓	✓	
Path-based routing, Host-based routing	✓		
Static IP		✓	
WebSockets	✓	✓	
Container Support	✓	✓	

aws

For TCP in VPC use Network Load Balancer

For all other use cases in VPC , use Application Load Balancer

For Classic networking use Classic Load Balancer

Questions