

Elastic Load Balancer Overview



Steven Moran
TRAINING ARCHITECT

LESSON BREAKDOWN

Elastic Load Balancer Overview

ELB Types

Common ELB Concepts

ELB Connectivity

Request Routing

Cross-Zone Load Balancing



Steven Moran
TRAINING ARCHITECT

Load Balancers

- Single logical target for service requests
- Requests distributed across backend application servers



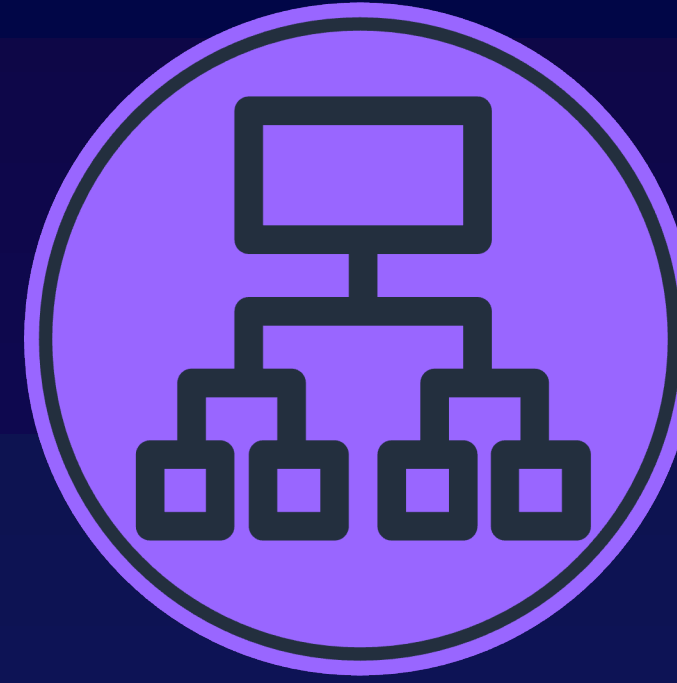


AWS PaaS Load Balancing Solution

- Highly available
- Regional resource
- Auto-scales for workload
- Handles requests from internet and within VPC
- Protected by security groups
- Can terminate HTTPS/TLS connections

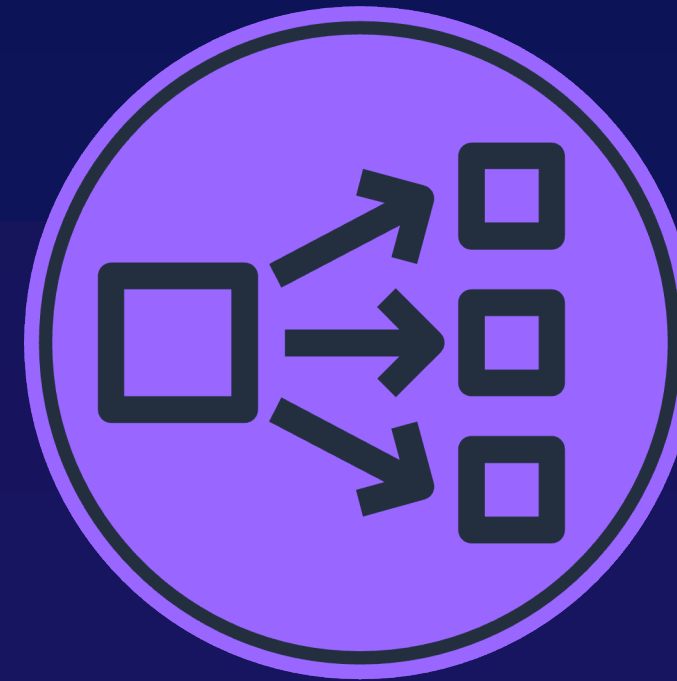
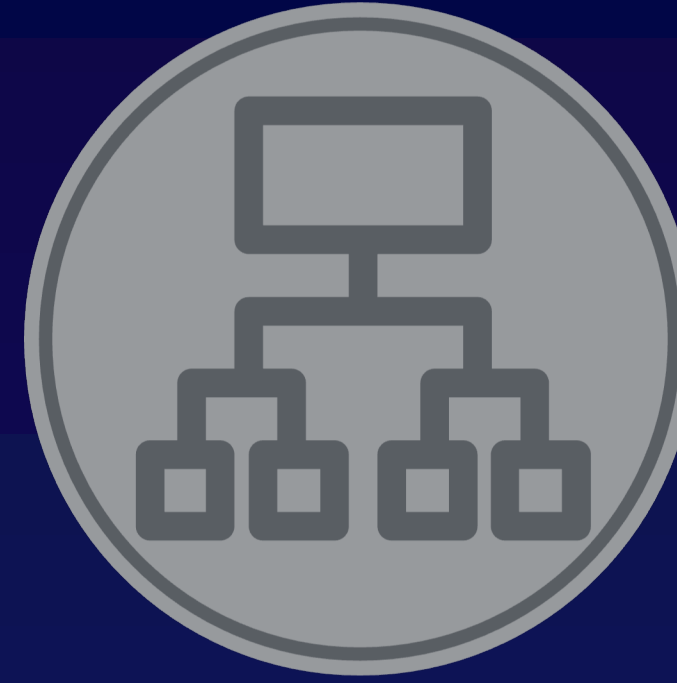
ELB Types

- Application load balancer
 - Layer 7
 - HTTP and HTTPS
 - Route requests using HTTP header content



ELB Types

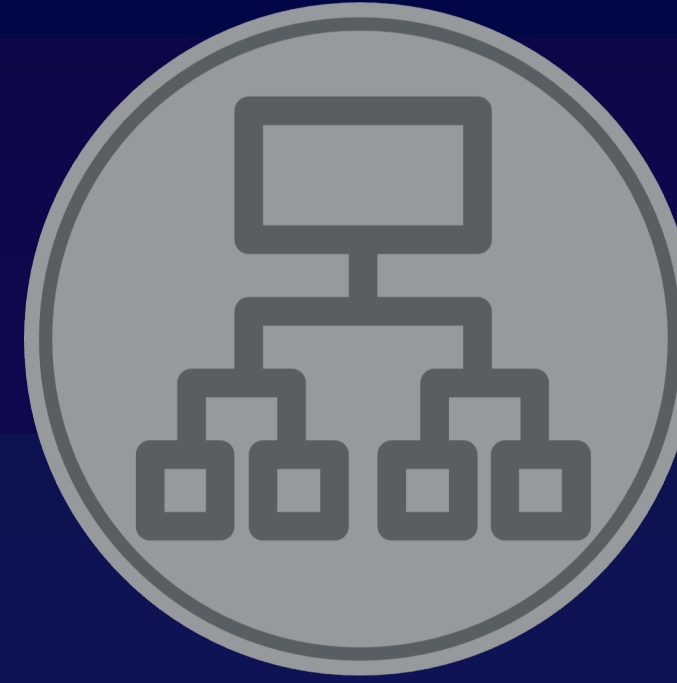
- Application load balancer
 - Layer 7
 - HTTP and HTTPS
 - Route requests using HTTP header content



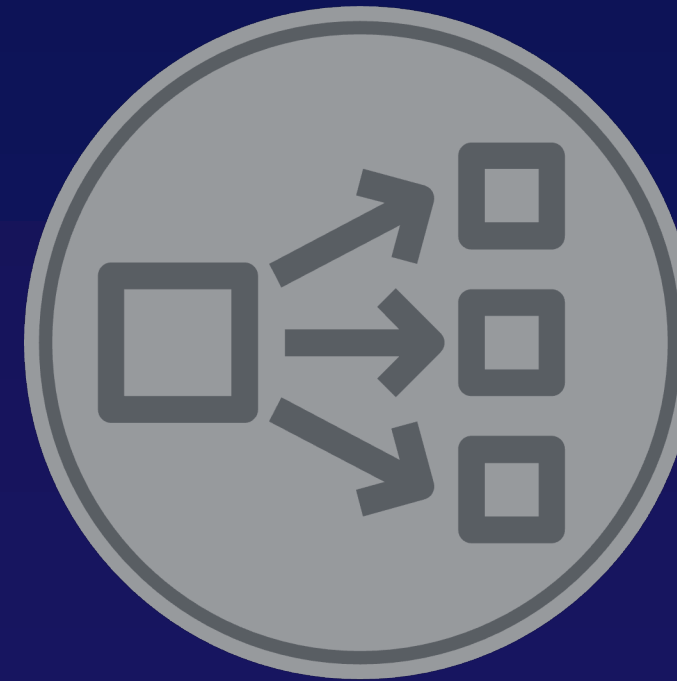
- Network load balancer
 - Layer 4
 - TCP, UDP, TLS
 - Route requests by protocol and port number

ELB Types

- Application load balancer
 - Layer 7
 - HTTP and HTTPS
 - Route requests using HTTP header content



- Classic load balancer
 - Layer 4 (and 7)
 - HTTP, HTTPS, TCP, TLS
 - Route requests by protocol and port number
 - Legacy option

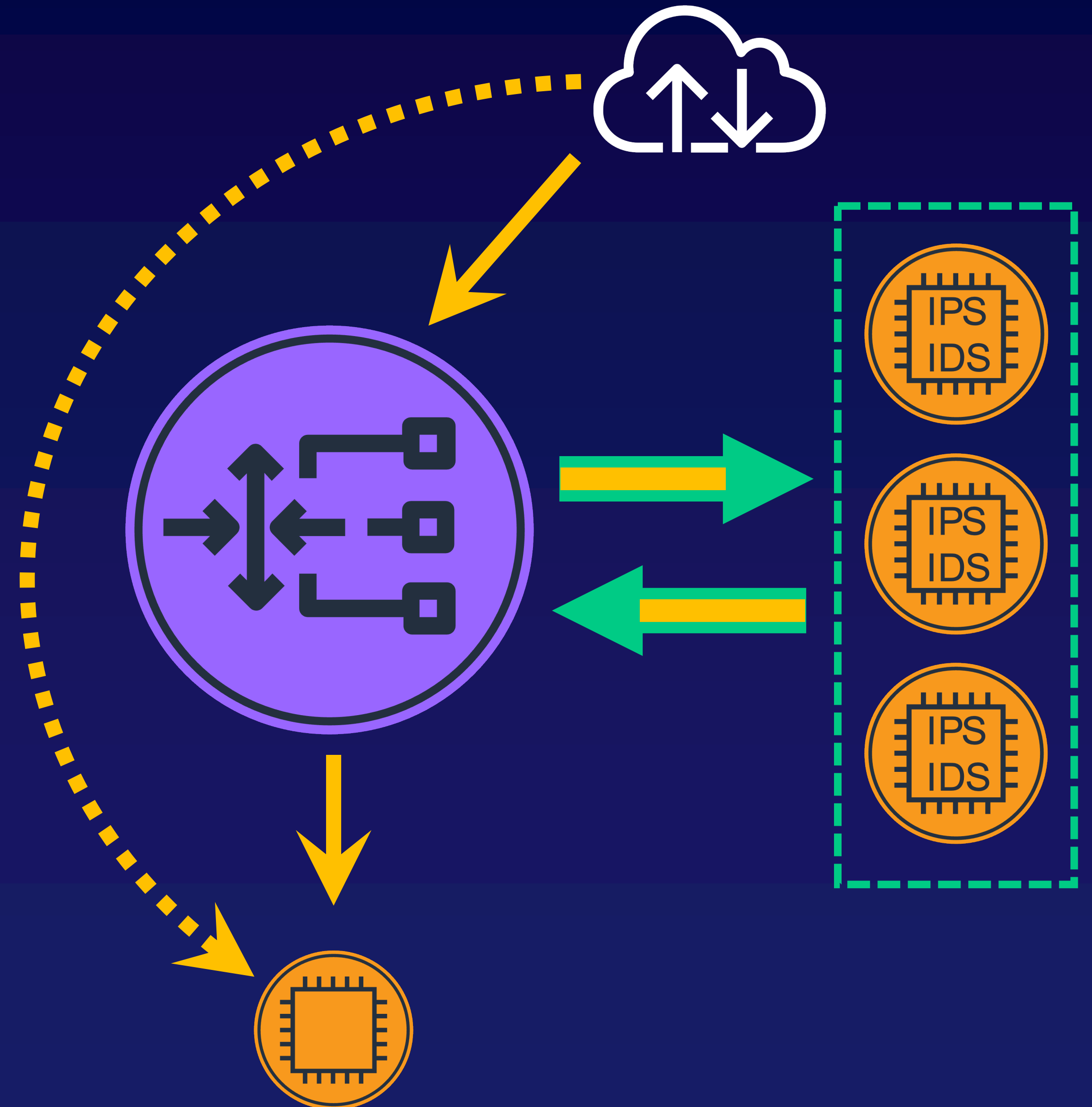


- Network load balancer
 - Layer 4
 - TCP, UDP, TLS
 - Route requests by protocol and port number



ELB Types

- Gateway load balancer
 - Works with applications that operate in-line with network traffic
 - Security
 - Analytics
 - Uses GENEVE protocol to encapsulate traffic
 - Allows original IP data to remain unchanged
 - Handles traffic to and from service targets
 - Layers 3 & 4



ELB Types – Which Is Best?

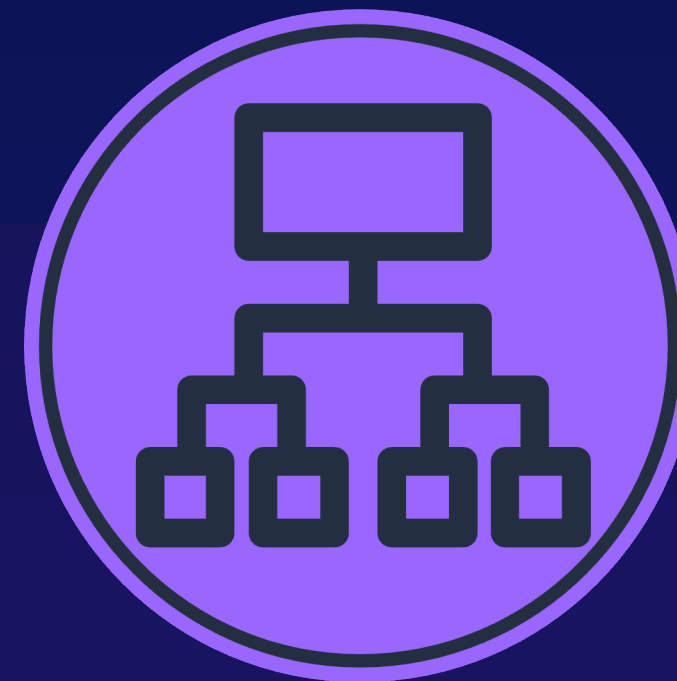


- Classic load balancer
 - Supports EC2-Classic
 - Legacy option
 - ALB and NLB offer more

ELB Types – Which Is Best?



- Classic load balancer
 - Supports EC2-Classic
 - Legacy option
 - ALB and NLB offer more

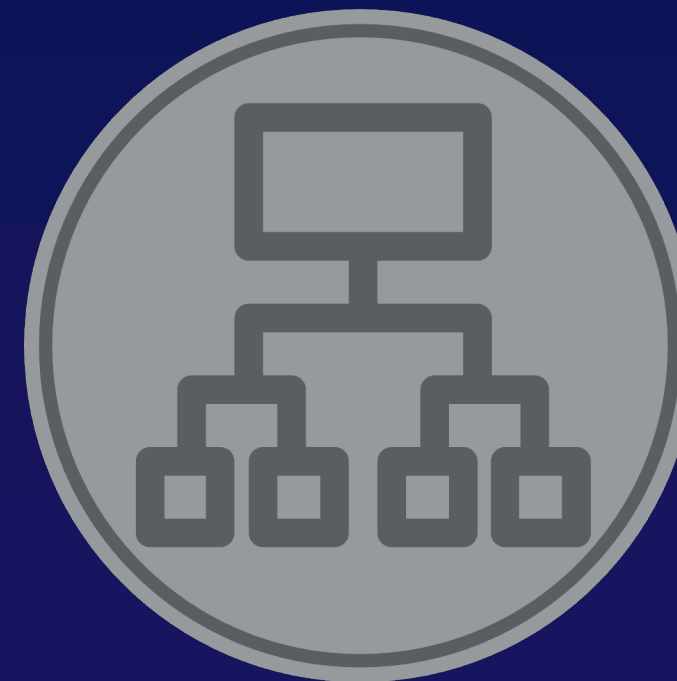


- Application load balancer
 - “Web” applications and services
 - Customizable request handling based on HTTP header content

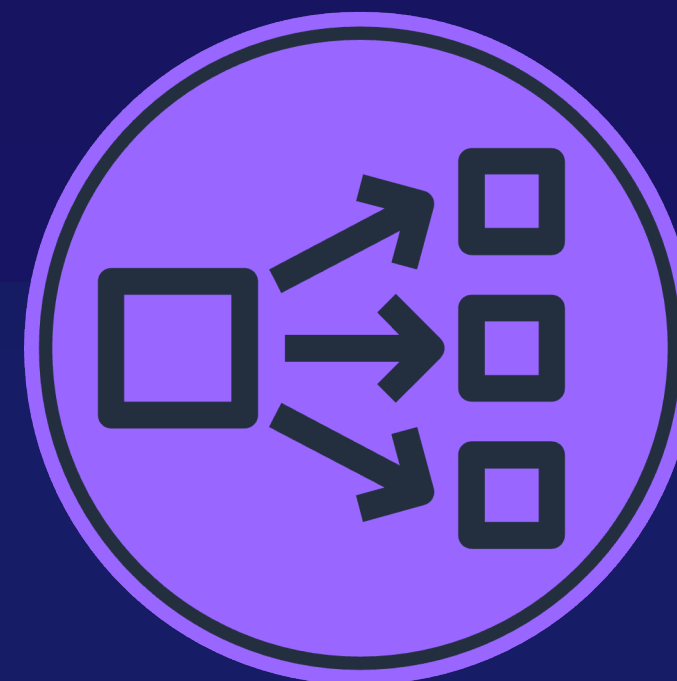
ELB Types – Which Is Best?



- Classic load balancer
 - Supports EC2-Classic
 - Legacy option
 - ALB and NLB offer more



- Application load balancer
 - “Web” applications & services
 - Customizable request handling based on HTTP header content



- Network load balancer
 - Forwards only by protocol and port
 - Better performance and scalability than ALB
 - Component of:
 - Endpoint Services
 - VPC Traffic Mirroring

Core ELB Concepts

- **Targets**
 - Backend resources that requests are forwarded to
 - Can be in different AZs
 - Health evaluated before use
 - Types vary by ELB-type
- **Target groups**
 - Logical container for targets
 - Scope of configuration for many settings
 - Health checks
 - Expected traffic
 - Not used by CLB





- Listeners
 - Processes that check for connection requests to specific protocols and ports.
 - Matching traffic forwarded to targets.
 - Settings and options vary by ELB type.

ELB Connectivity (Non-GWLB)

- Each ELB works with a single VPC.
- Requests are sent to AWS-assigned FQDN.
 - Public if ELB scheme is internet-facing
 - Private if ELB scheme is internal



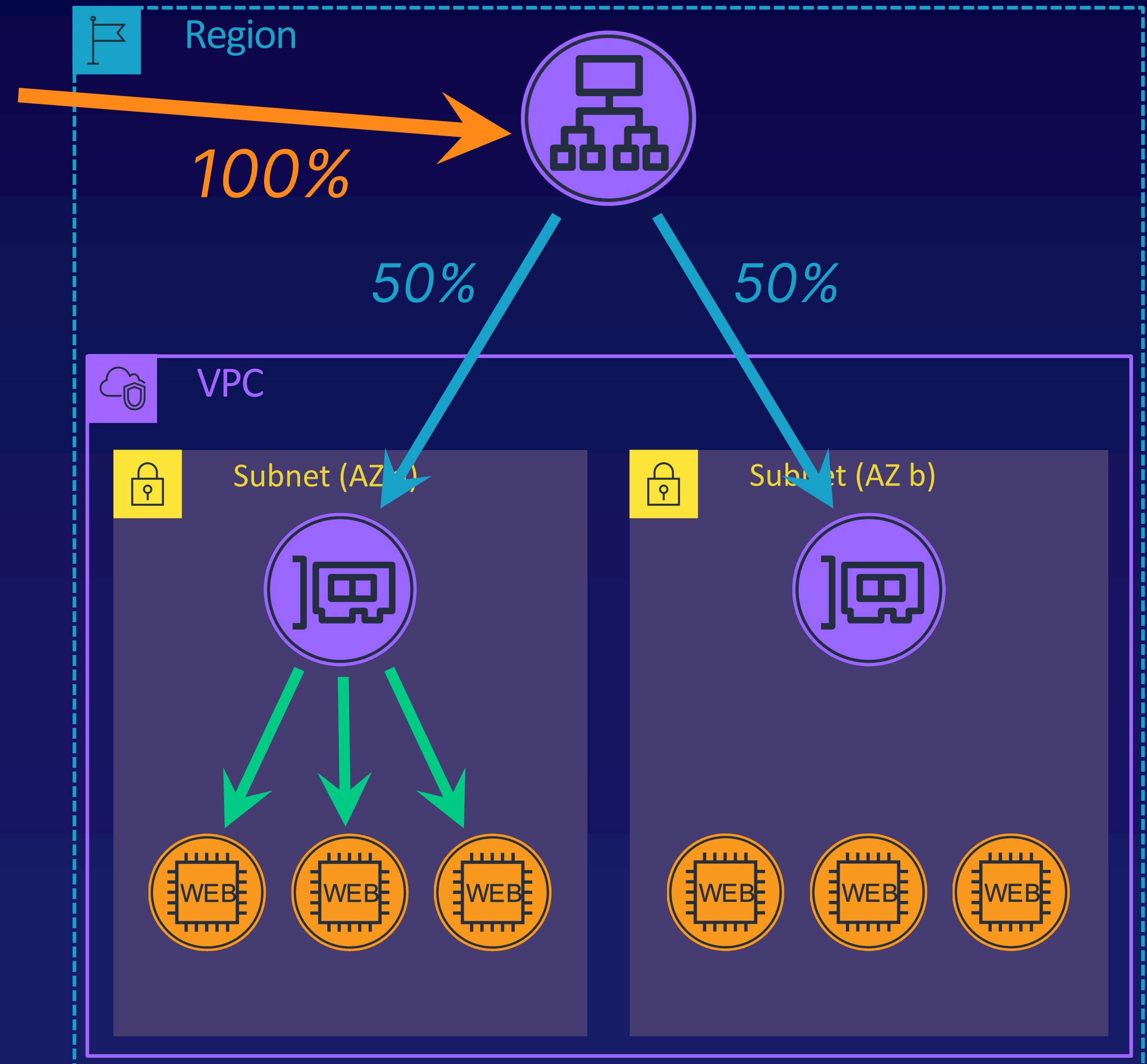
ELB Connectivity (Non-GWLB): Availability Zone Presence

- ELBs must be configured to work in at least one AZ.
 - Use at least two to support high availability
- Worker nodes created in each configured subnet.
 - Additional nodes added when scaling
- Each node accessed by IP address.
 - Private IPs for internal
 - Private and public IPs for internet-facing



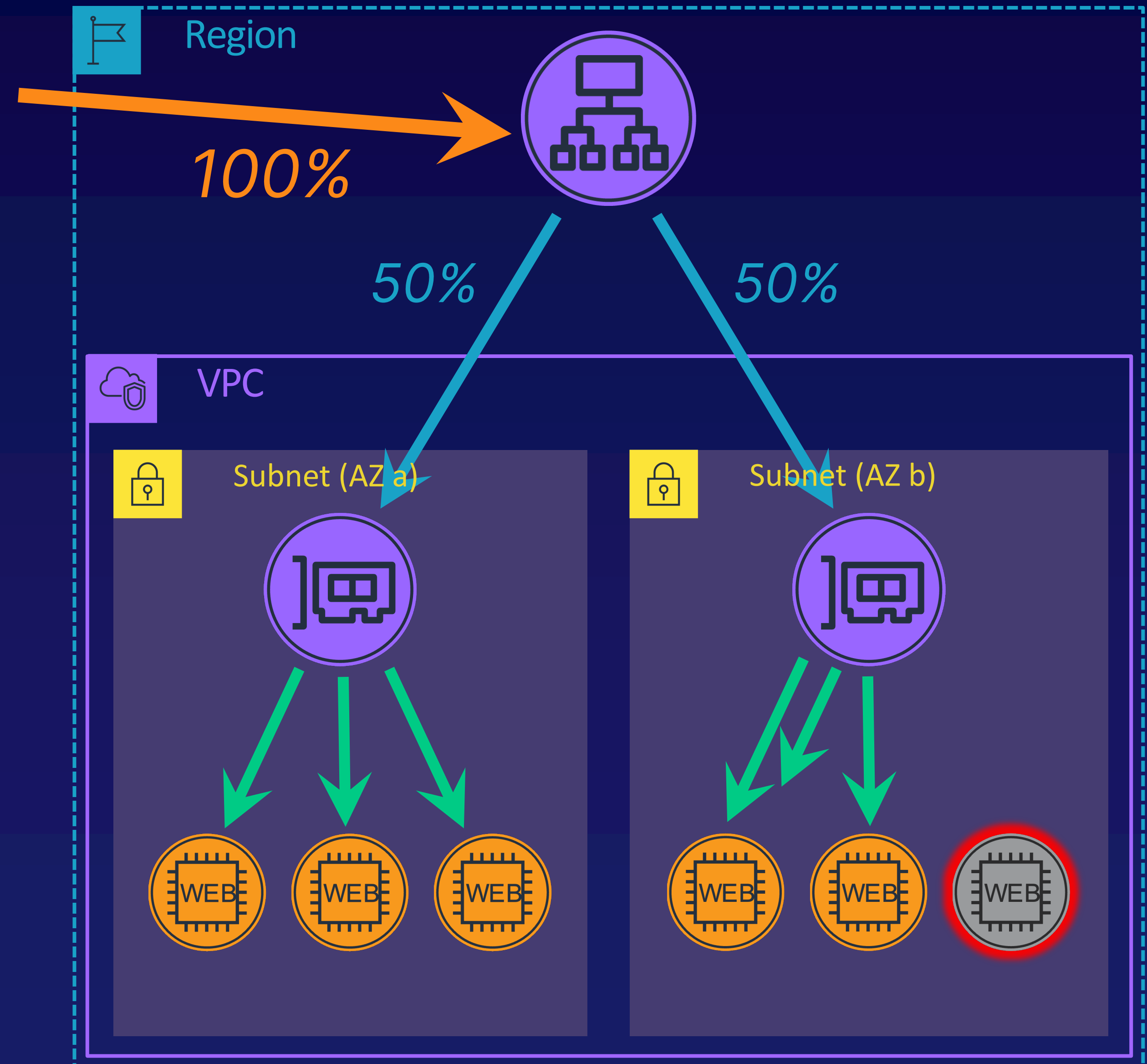
Request Routing

- Requests to ELB will be evenly distributed across all nodes.
- Nodes determine which targets receive requests.
 - Only consider healthy targets.



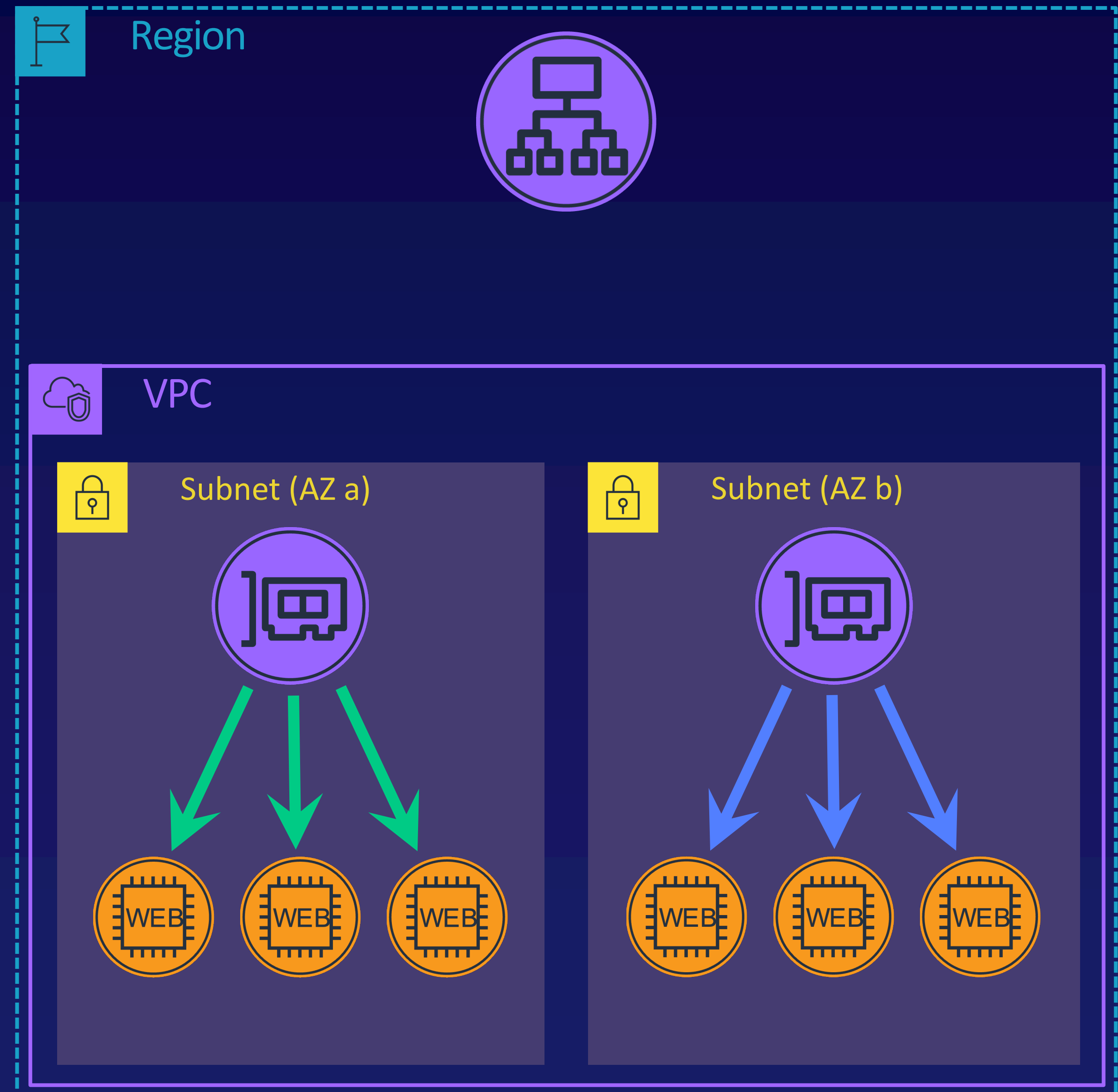
Request Routing

- Requests to ELB will be evenly distributed across all nodes.
- Nodes determine which targets receive requests.
 - Only consider healthy targets.
- Routing algorithms:
 - Round robin
 - Least outstanding requests
 - Hash (NLB)



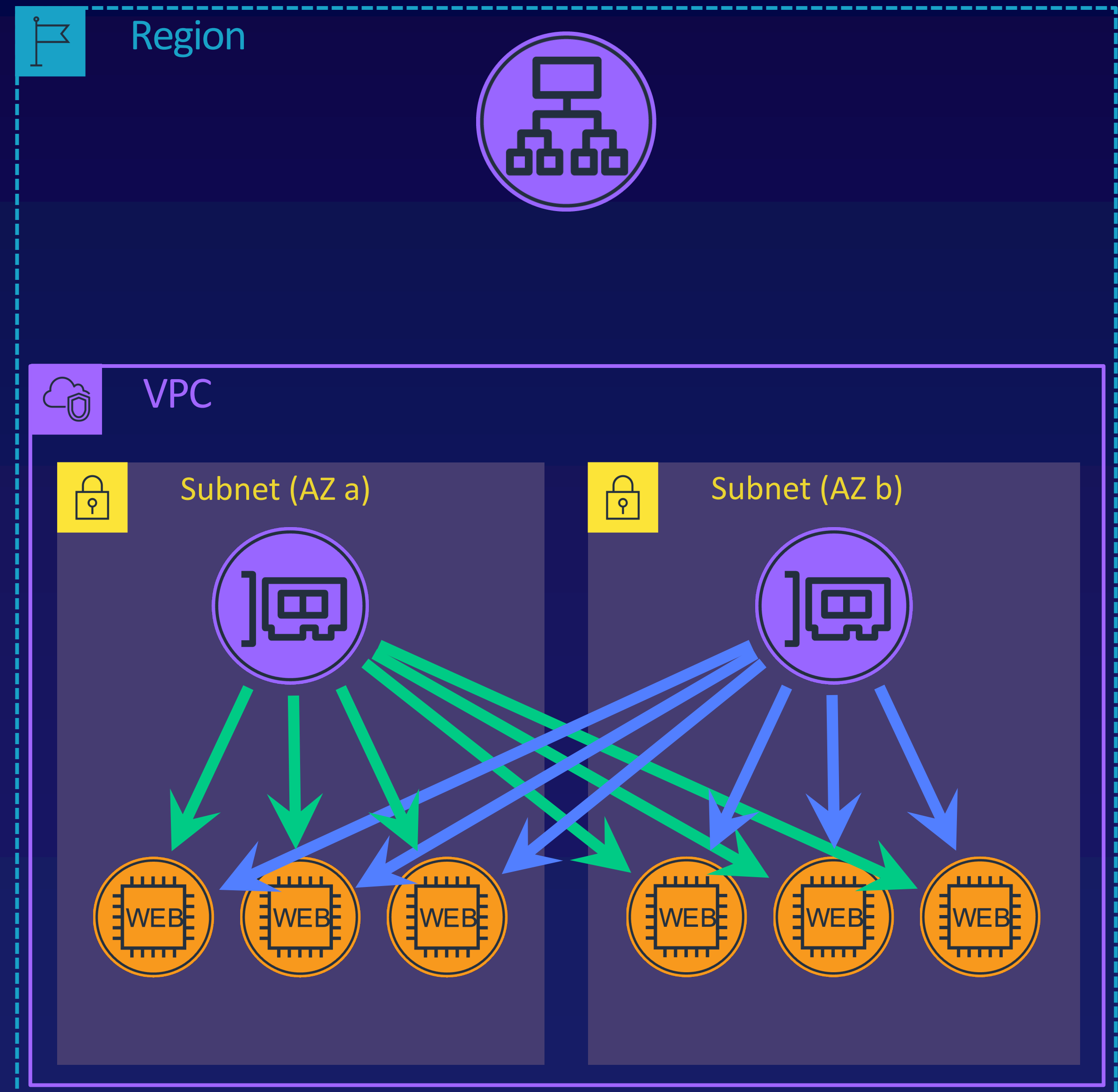
Cross-Zone Load Balancing

- If *disabled*...
 - ELB nodes only forward to targets in the same AZ.



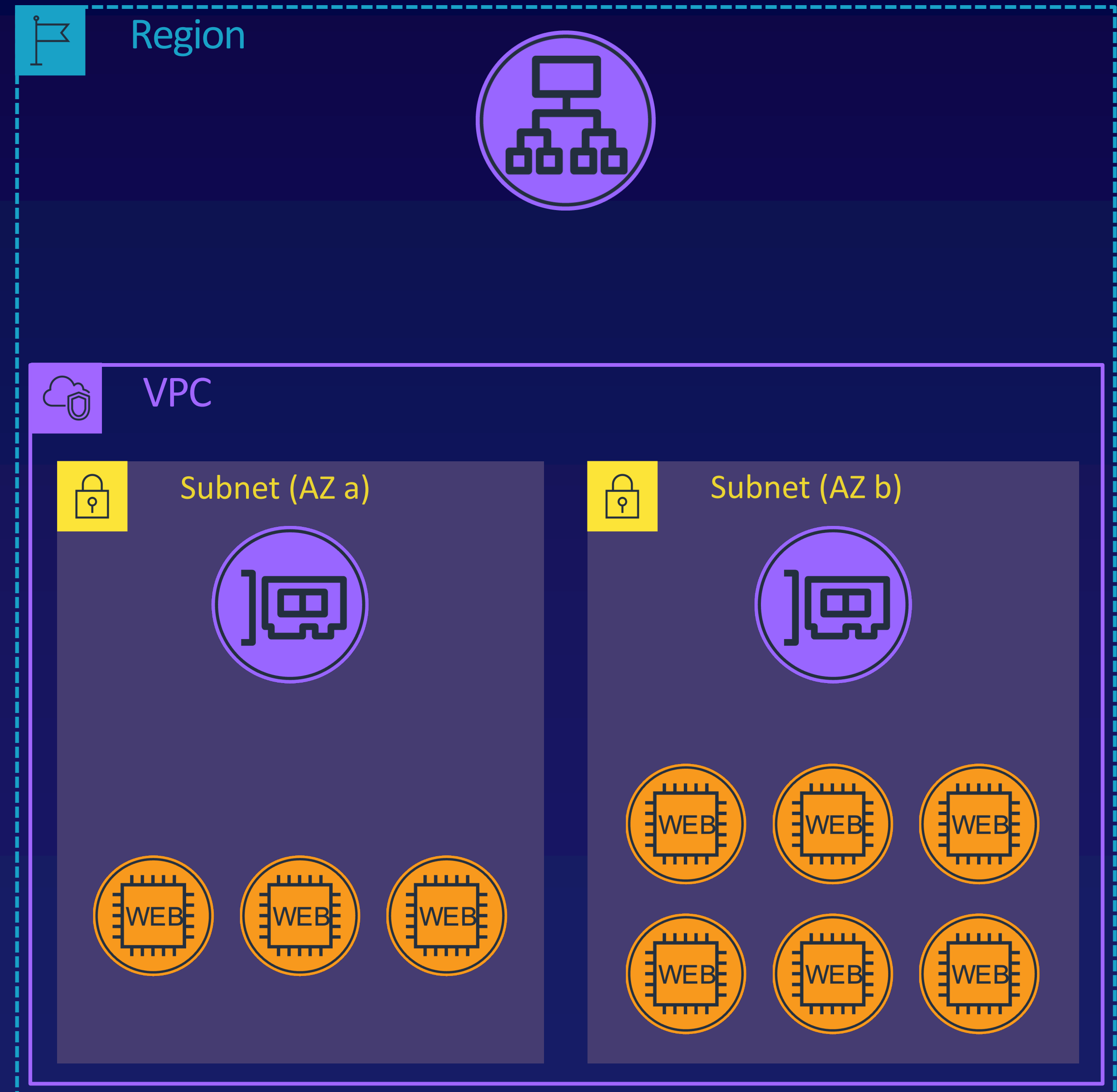
Cross-Zone Load Balancing

- If *disabled*...
 - ELB nodes only forward to targets in the same AZ.
- If *enabled*...
 - ELB nodes may forward to targets in any AZ.



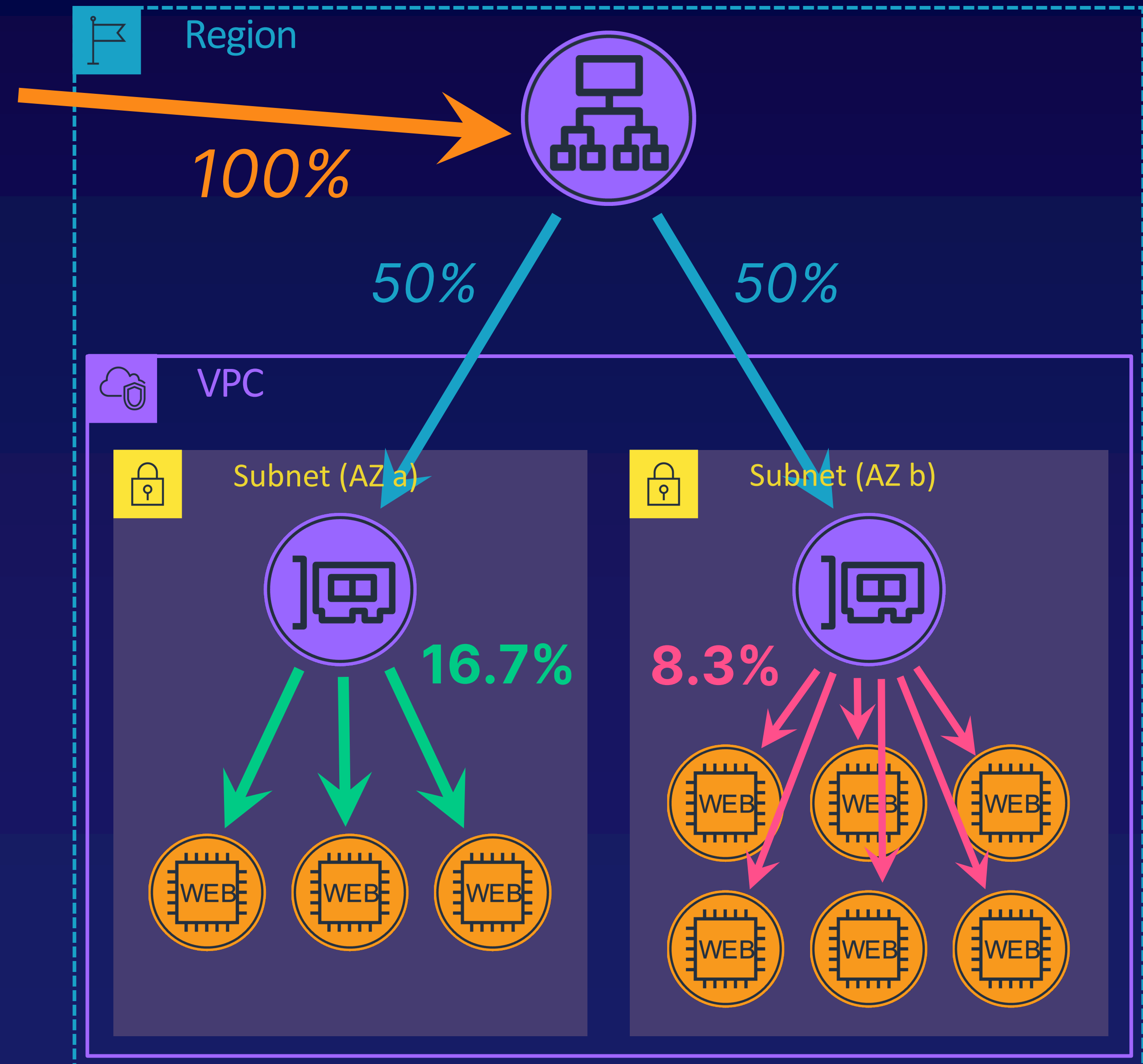
Cross-Zone Load Balancing

- If *disabled*...
 - ELB nodes only forward to targets in the same AZ.
- If *enabled*...
 - ELB nodes may forward to targets in any AZ.



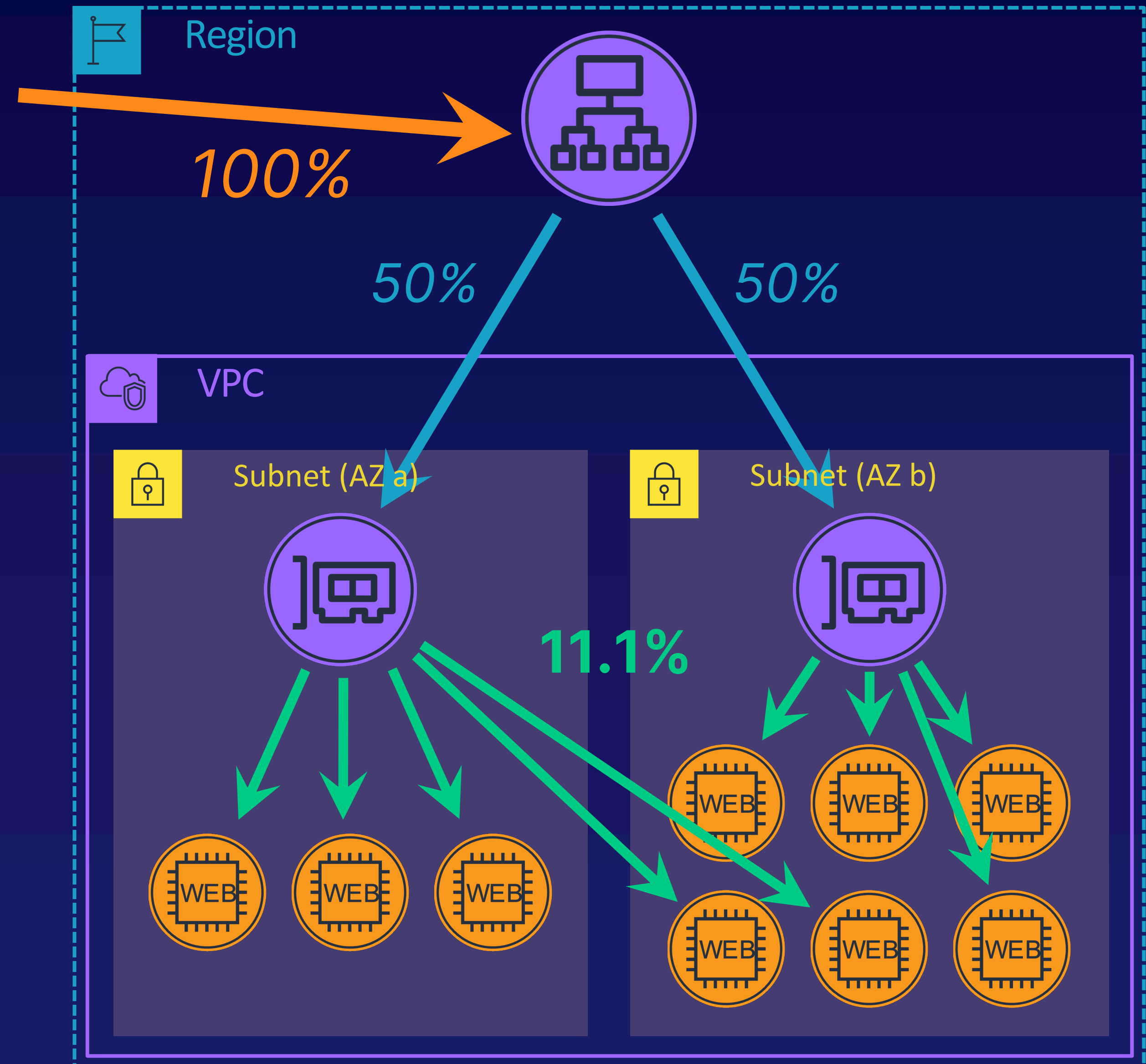
Cross-Zone Load Balancing

- If *disabled*...
 - ELB nodes only forward to targets in the same AZ.
- If *enabled*...
 - ELB nodes may forward to targets in any AZ.



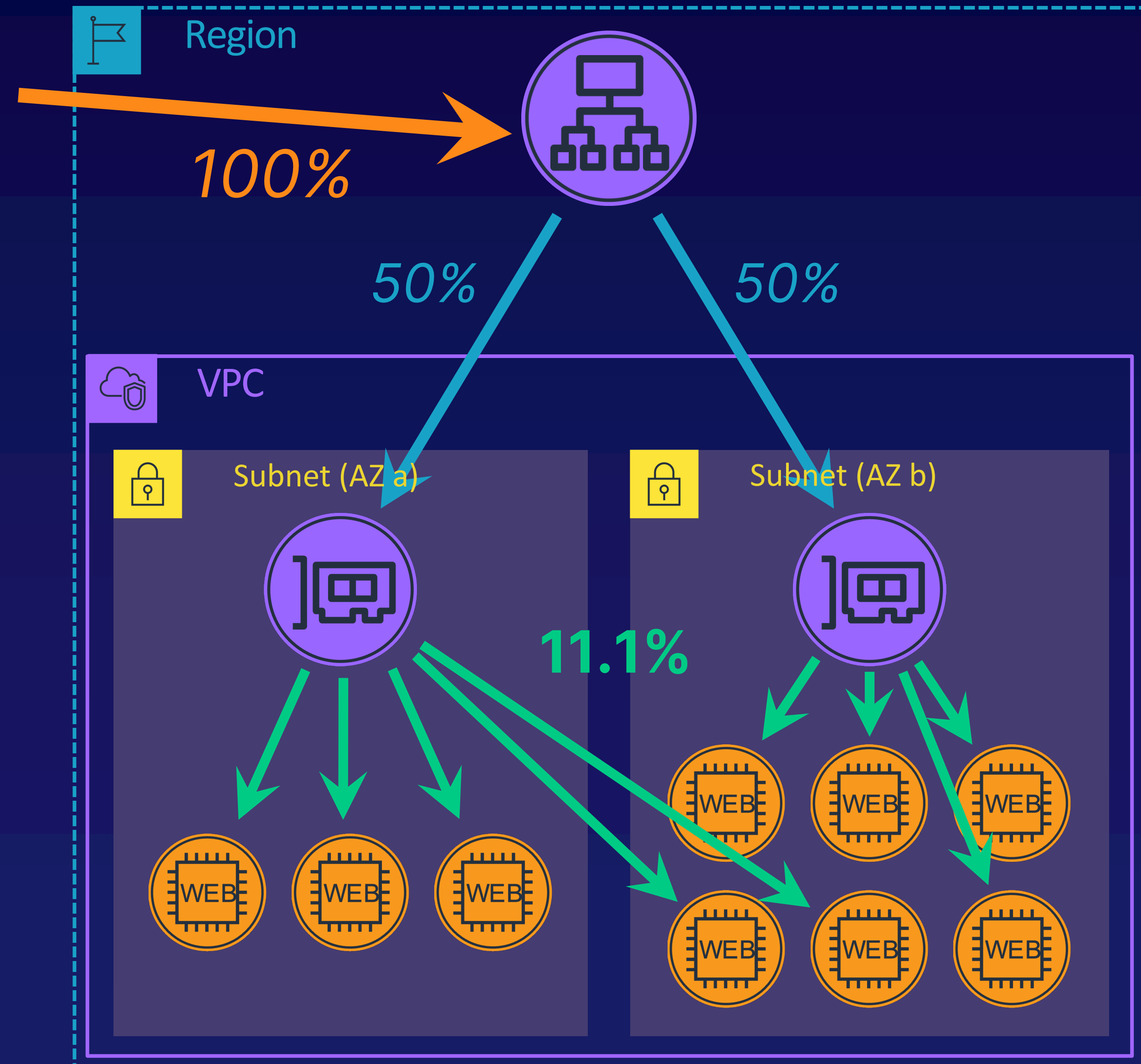
Cross-Zone Load Balancing

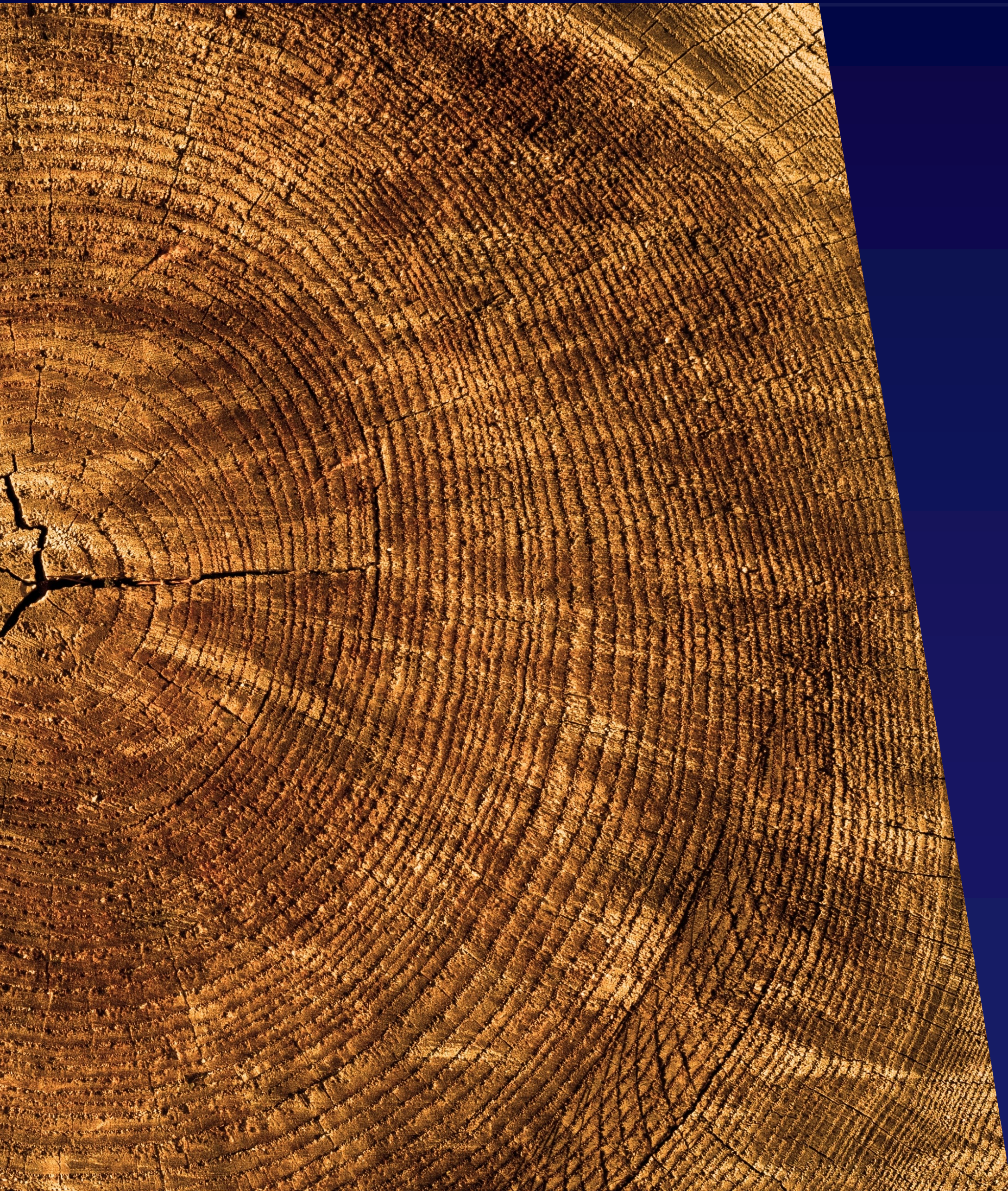
- If *disabled*...
 - ELB nodes only forward to targets in the same AZ.
- If *enabled*...
 - ELB nodes may forward to targets in any AZ.



Cross-Zone Load Balancing

- Always enabled for ALB
- Enabled by default for CLB created using AWS console
- Disabled by default for:
 - NLB
 - GWLB
 - CLB created using CLI/API
- Incurs no data transfer charges for ALB and CLB





- Disabled by default.
- Captures information about requests.
- Records stored in S3 bucket.
 - Best effort
 - Eventually consistent
- Supported for:
 - ALB
 - CLB
 - NLB for TLS protocol only
- Data fields vary by ELB type.

Fast Takeaways

Elastic Load Balancers are an important part of highly available application architecture within AWS.

Optimized for different types of workloads.

Listeners define what kind of traffic load balancers should process.

Target groups identify possible targets and the type of traffic they are expecting.