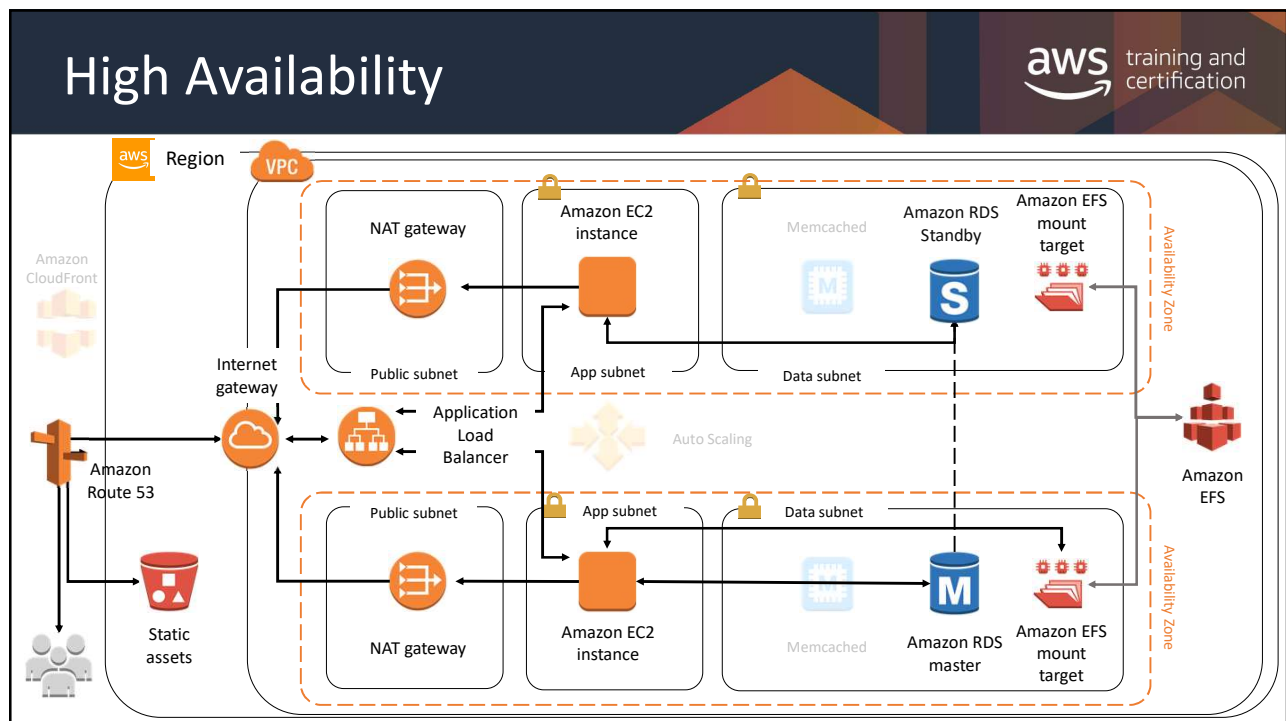




0



1

# Module 7



## The architectural need

Your application needs to support a much larger user base and variable load, and it needs to handle Availability Zone-level failures.

### Module Overview

- Connecting Networks
- VPC Endpoints
- Load Balancing
- High Availability

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

2



# Connecting Networks

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

3

# Virtual Private Gateway (VGW)

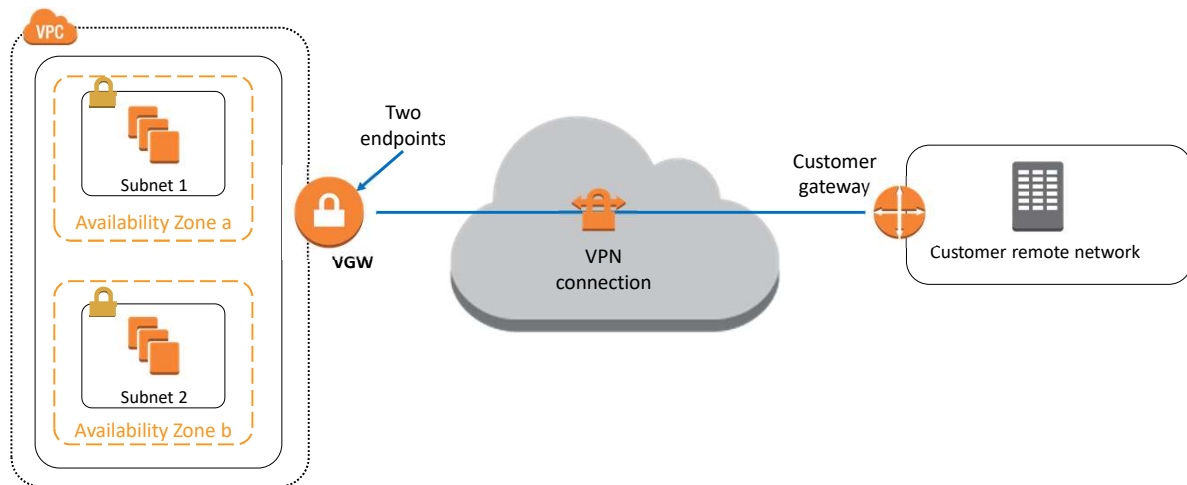


Enables you to establish private connections (VPNs) between an Amazon VPC and another network

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

4

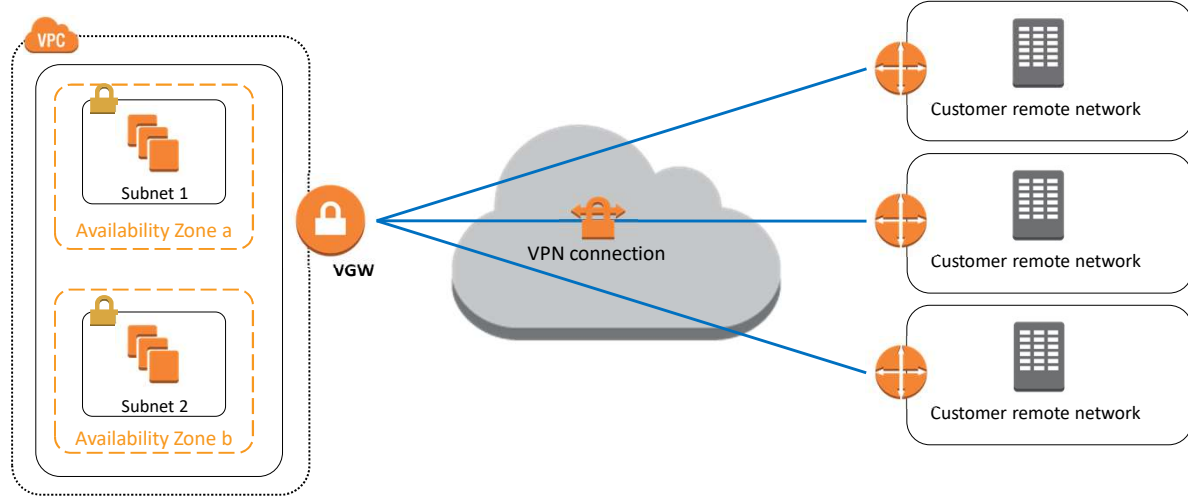
## Extending On-Premises Network to AWS: VPN Connections



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

5

## Extending On-Premises Network to AWS: Multiple VPN



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

6

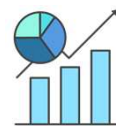
## AWS Direct Connect (DX)



AWS Direct Connect (DX) provides you with a **dedicated, private network connection** of either 1 or 10 Gbps



Reduces data transfer costs



Improve application performance with predictable metrics

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

7

## DX Use Cases



AWS Direct  
Connect

- Hybrid cloud architectures
- Continually transferring large data sets
- Network performance predictability
- Security and compliance

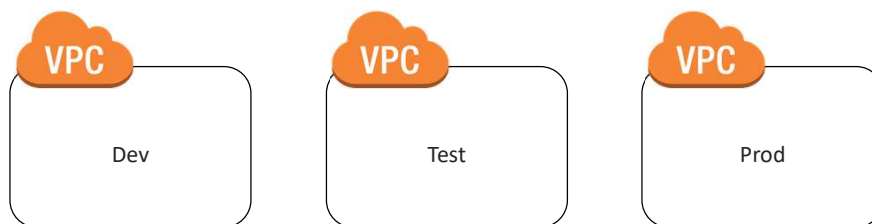
© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

8

## Connecting VPCs



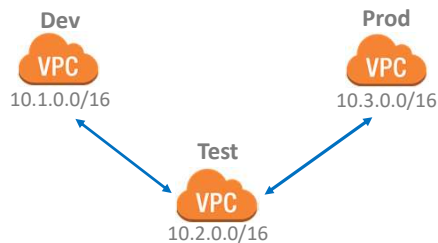
- Isolating some of your workloads is generally a good practice.
- But you may need to transfer data between two or more VPCs.



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

9

# Connecting VPCs – VPC Peering



Instances can communicate across a peering connection as if they were in the same network.

- Use **private** IP addresses
- **Intra and inter-region** support
- IP spaces **cannot overlap**
- Only **one peering resource** between any two VPCs
- **Transitive** peering relationships are **not supported**
- Can be established **between** different AWS **accounts**

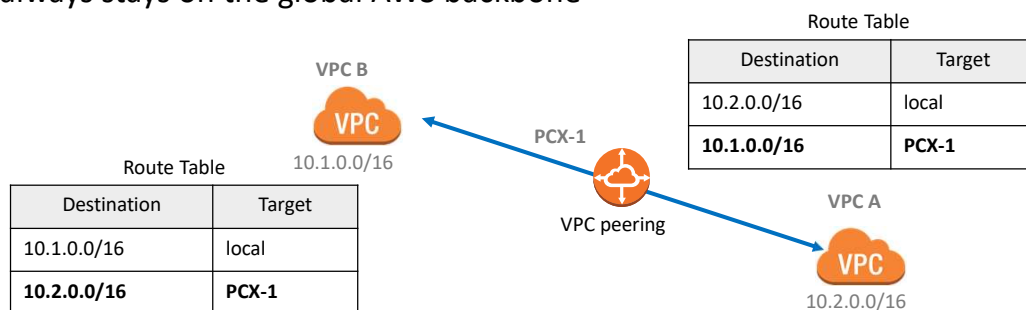
© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

10

## VPC Peering



- No internet gateway or virtual gateway required
- Highly available connections; not a single point of failure
- No bandwidth bottlenecks
- Traffic always stays on the global AWS backbone



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

11

# VPC Endpoints



Privately connect your EC2 instances to services outside your VPC **without leaving AWS**.

Don't need to use an internet gateway, VPN, network address translation (NAT) devices, or firewall proxies.

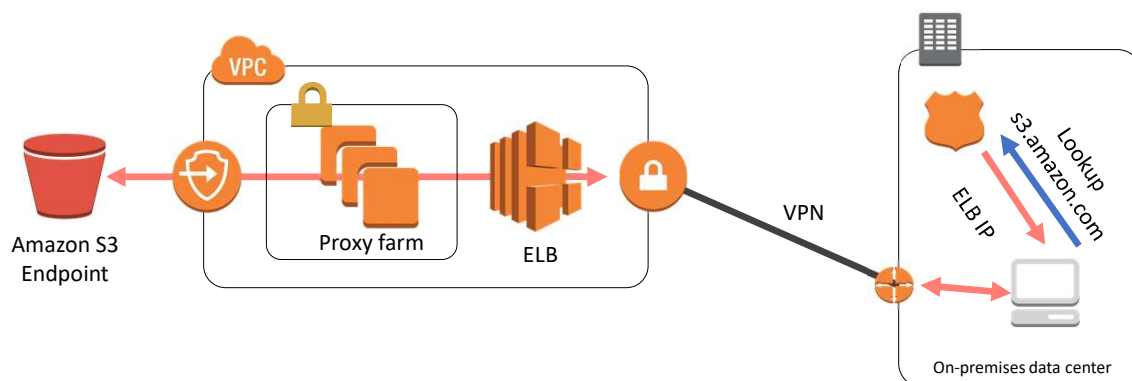


- Does not require traversal over the internet
- Must be in the same region
- They are horizontally scaled, redundant, and highly available

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

12

## Accessing VPC Endpoints from Outside the VPC



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

13

# Load Balancing on AWS

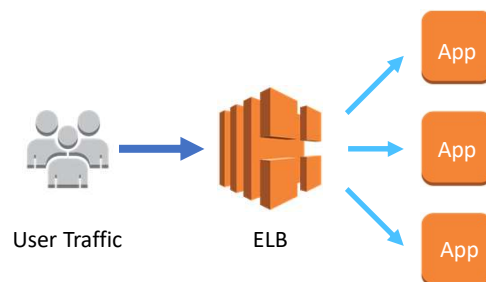
© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

14

## Elastic Load Balancing (ELB)



A **managed load balancing service** that distributes incoming application traffic across multiple Amazon EC2 instances, containers, and IP addresses.



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

15



## ELB: Features



- Uses **HTTP, HTTPS, TCP and SSL** (secure TCP) protocols.
- Can be **external or internal** facing
- Each load balancer is given a **DNS name**
- Recognizes and responds to **unhealthy instances**

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

16

## ELB: Options



### Application Load Balancer



- Flexible application management
- Advanced load balancing of HTTP and HTTPS traffic
- Operates at the request level  
(Layer 7)

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

17

# ELB: Options

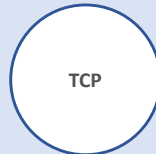


## Application Load Balancer



- Flexible application management
- Advanced load balancing of HTTP and HTTPS traffic
- Operates at the request level (Layer 7)

## Network Load Balancer



- Extreme performance and static IP for your application
- Load balancing of TCP traffic
- Operates at the connection level (Layer 4)

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

18

# ELB: Options

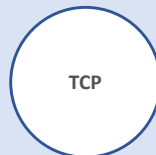


## Application Load Balancer



- Flexible application management
- Advanced load balancing of HTTP and HTTPS traffic
- Operates at the request level (Layer 7)

## Network Load Balancer



- Extreme performance and static IP for your application
- Load balancing of TCP traffic
- Operates at the connection level (Layer 4)

## Classic Load Balancer

PREVIOUS GENERATION  
for HTTP, HTTPS, and TCP

- Existing application that was built within the EC2 Classic network
- Operates at both the request level and connection level

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

19

# Why You Should Use ELB



High availability



Health checks



Security features



TLS termination

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

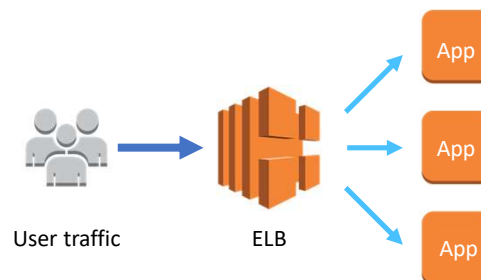
20

# Connection Draining



If you need to **remove an instance** from your production fleet, but **don't want to affect your users**:

Affected backend instances will complete requests in progress before deregistration

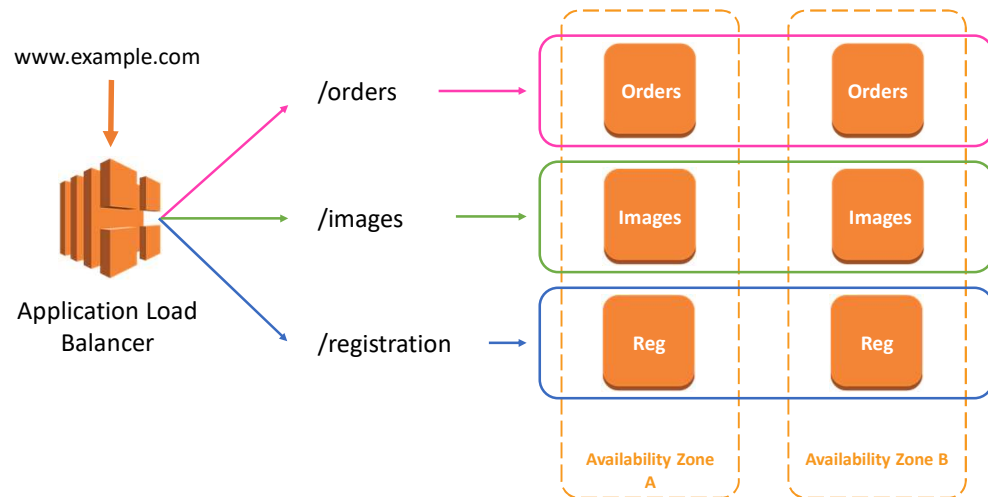


**Enable connection draining**

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

21

# Cloud Design Pattern: Application Load Balancer



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

22

## High Availability



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

23

# What is High Availability?



Your application can **recover from a failure or roll over to a secondary source** within an **acceptable** amount of degraded performance time.

Percent of Uptime	Max Downtime per Year	Equivalent Downtime per Day
90%	36.5 days	2.4 hrs
99%	3.65 days	14 min
99.9%	8.76 hrs	86 sec
99.99%	52.6 min	8.6 sec
99.999%	5.25 min	.86 sec

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

24

## High Availability Example



*Assume everything fails, and design backward*

Implement redundancy where possible in order to prevent single failures from bringing down an entire system.



App servers



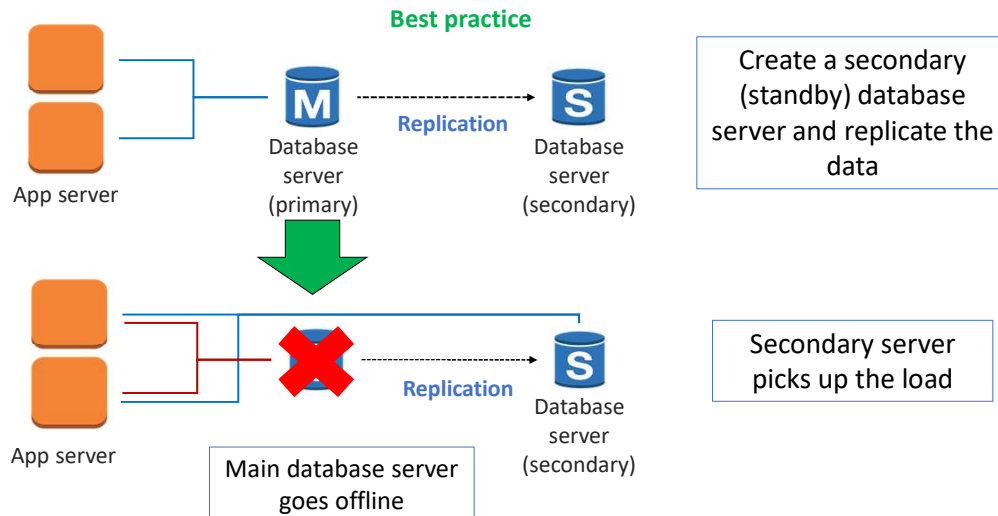
Database server

**Anti-pattern**

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

25

## High Availability Example



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

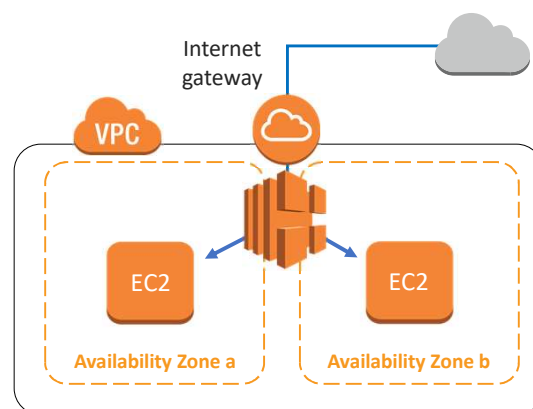
26

## How Many Availability Zones Should I Use?



Start with two Availability Zones per AWS Region.

If resources in one Availability Zone are unreachable, your application shouldn't fail.

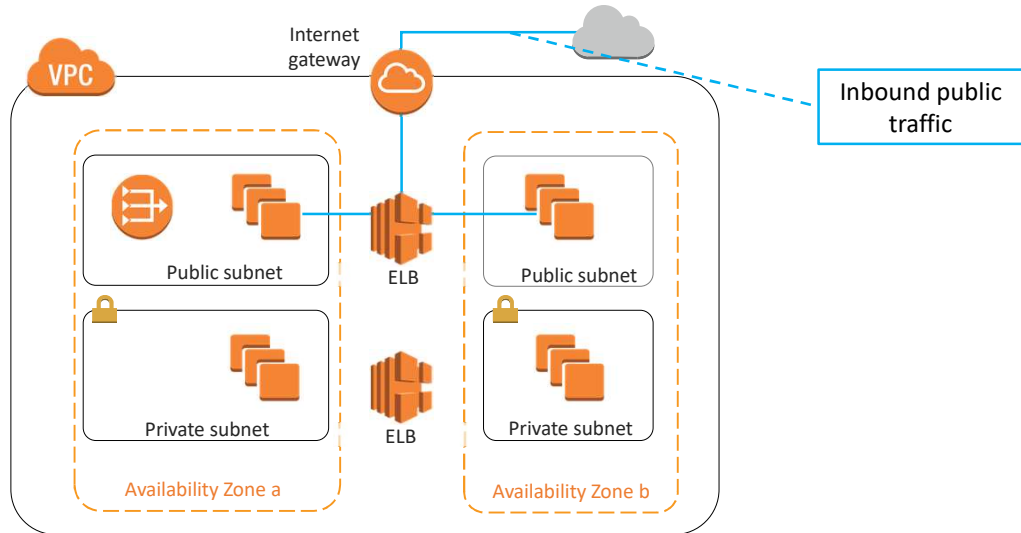


© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

27

## Example Architecture Diagram

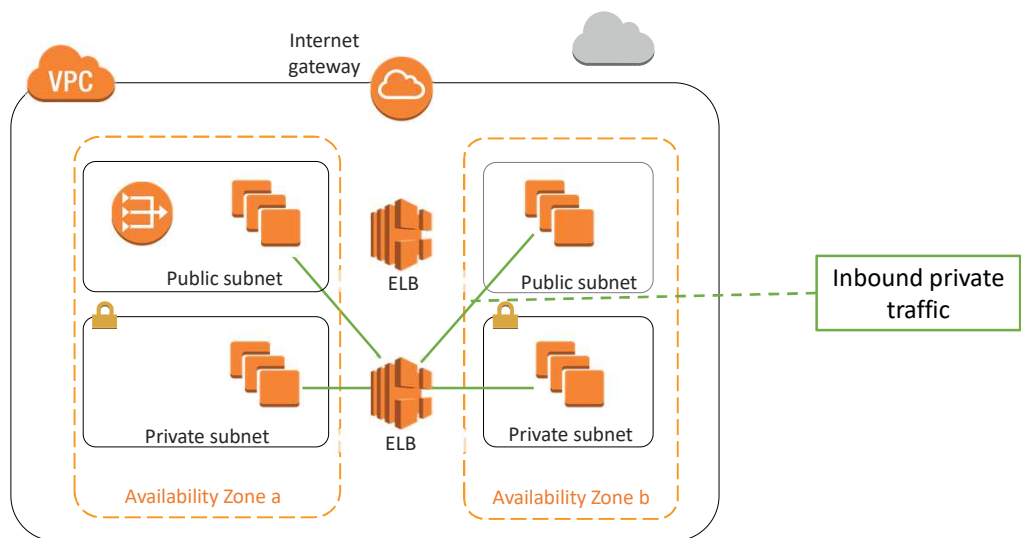
aws training and certification



28

## Example Architecture Diagram

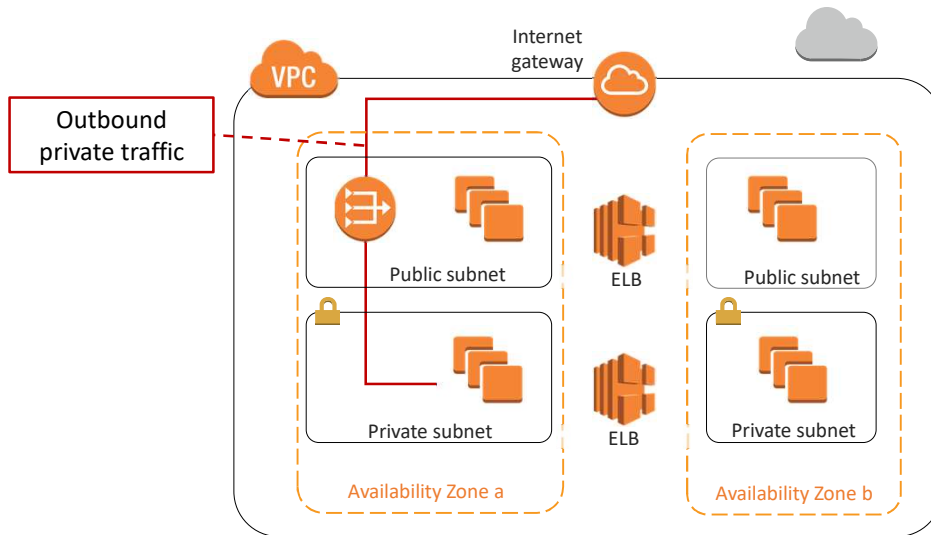
aws training and certification



29

## Example Architecture Diagram

aws training and certification

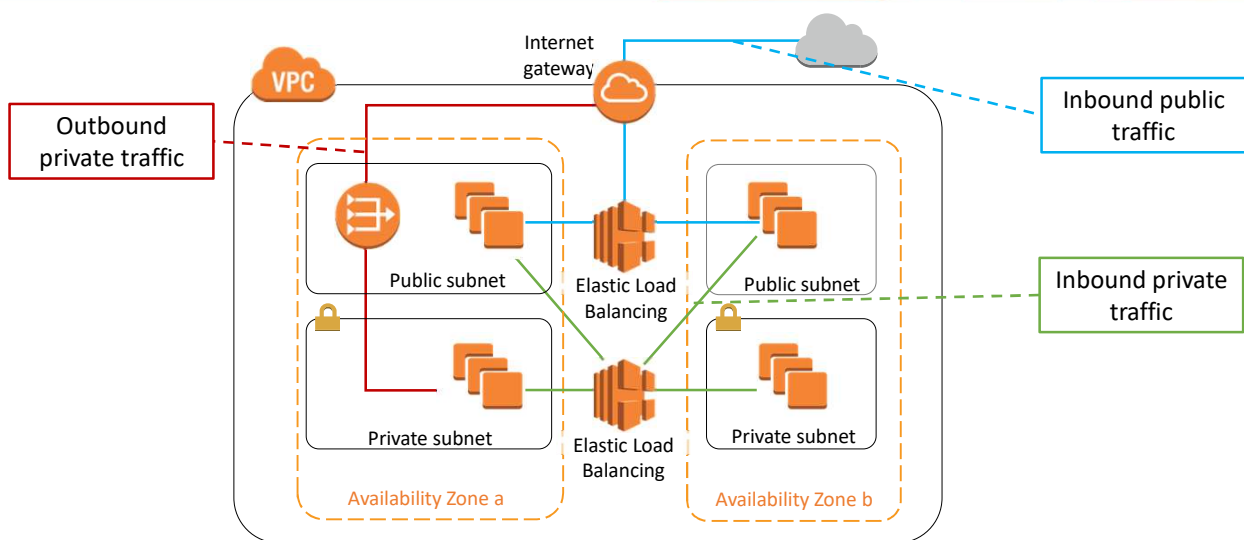


© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

30

## Example Architecture Diagram

aws training and certification



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

31

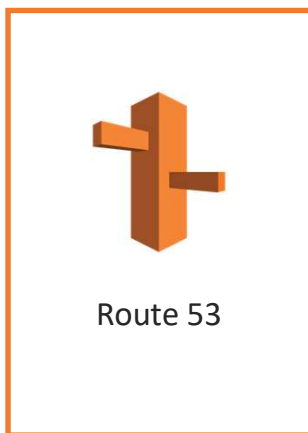


## Multi-Region High Availability and DNS

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

32

## Amazon Route 53



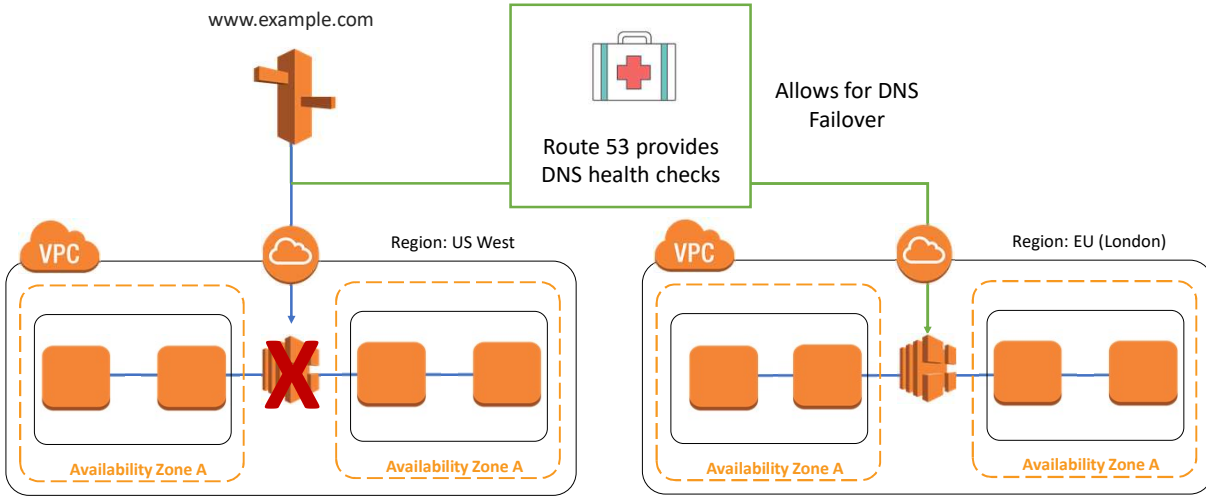
Route 53 is a highly available and scalable **cloud Domain Name System (DNS) service**.

- DNS translates domain names into IP addresses
- Able to purchase and manage domain names and automatically configure DNS settings
- Provides tools for flexible, high-performance, highly available architectures on AWS
- Multiple routing options

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

33

# How Does Route 53 Help with High Availability?



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

\* Consider using Global Accelerator for stringent SLAs

34

# Route 53 Routing Options



Simple round robin

Weighted round robin

Latency-based routing

Health check and DNS failover

Geolocation routing

Geoproximity routing with traffic biasing

Multi-value answers



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

35

