DigitalCloud TRAINING

Auto Scaling, Amazon ELB

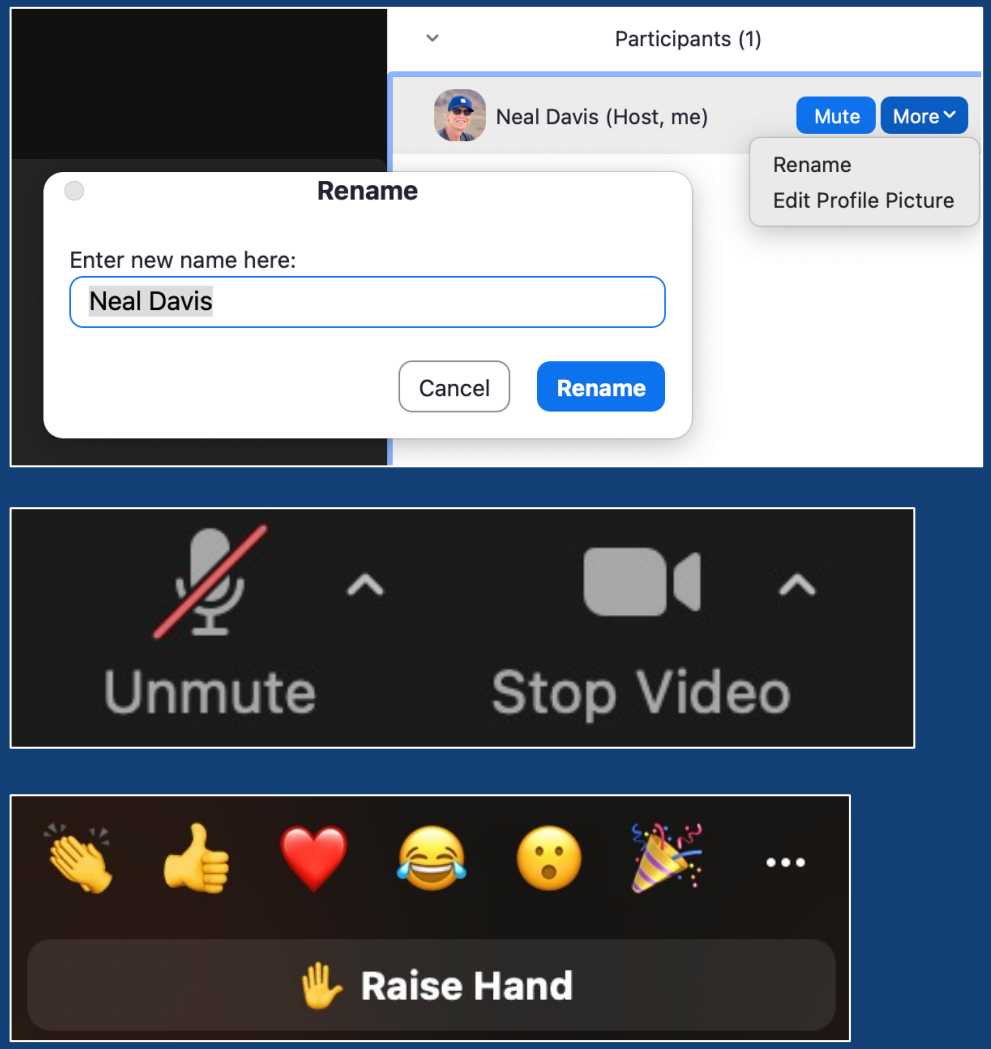
NEAL DAVIS
Founder and AWS Instructor

Virtual Classroom Requirements

1) Update your Zoom name to be your first/last name

2) Video must be always on

3) Raise your hand if you have a question and we'll let you know when to unmute



DigitalCloud TRAINING

Live Training Topics

Section 3	Public & Private Subnets
Section 3	NAT Gateways
Section 4	Scaling Strategies
Section 4	Stateful vs Stateless Applications
Section 4	Gateway Load Balancer

DigitalCloud TRAINING

Live Training Topics

Section 4	ALB/NLB Access Control
Section 4	Identifying Client IPs with ELB
Section X	Troubleshooting ELB

DigitalCloud TRAINING

Amazon EC2 in Public and Private Subnets

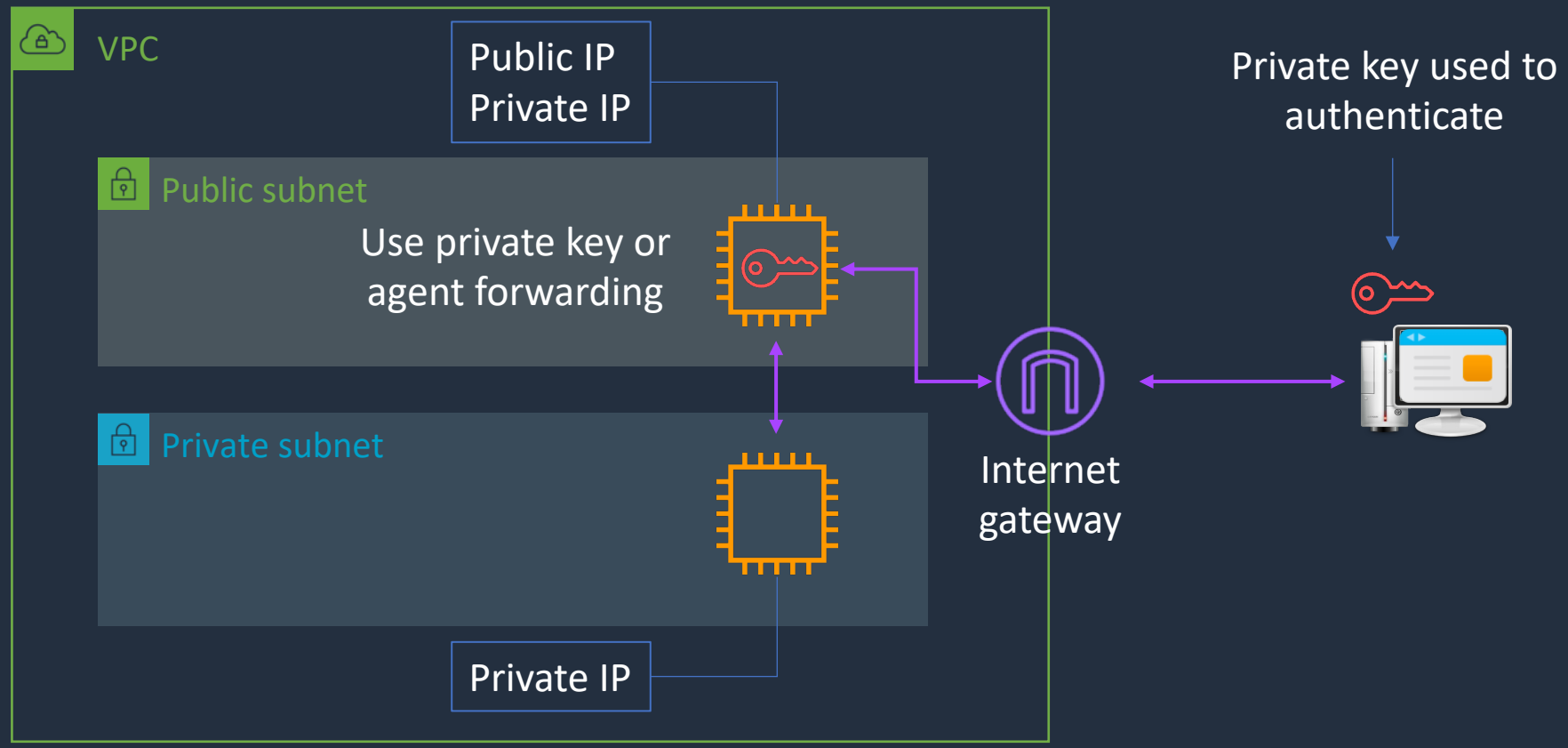
Public subnets require:

- A route to an Internet gateway
- Auto-assign public IPv4 or IPv6 address

Private subnets:

- Do not auto assign public addresses
- Do not have a route to an internet gateway

Public Subnet Route Table	
Destination	**Target**
10.0.0.0/16	Local
0.0.0.0/0	igw-id



Network Address Translation (NAT)

Example **without** using NAT

Company office

Private IP addresses are used within the company office / data center

192.168.0.1

192.168.0.2

Switch

192.168.0.200

Router

Public IP addresses are used on the Internet

54.200.168.152

The Internet

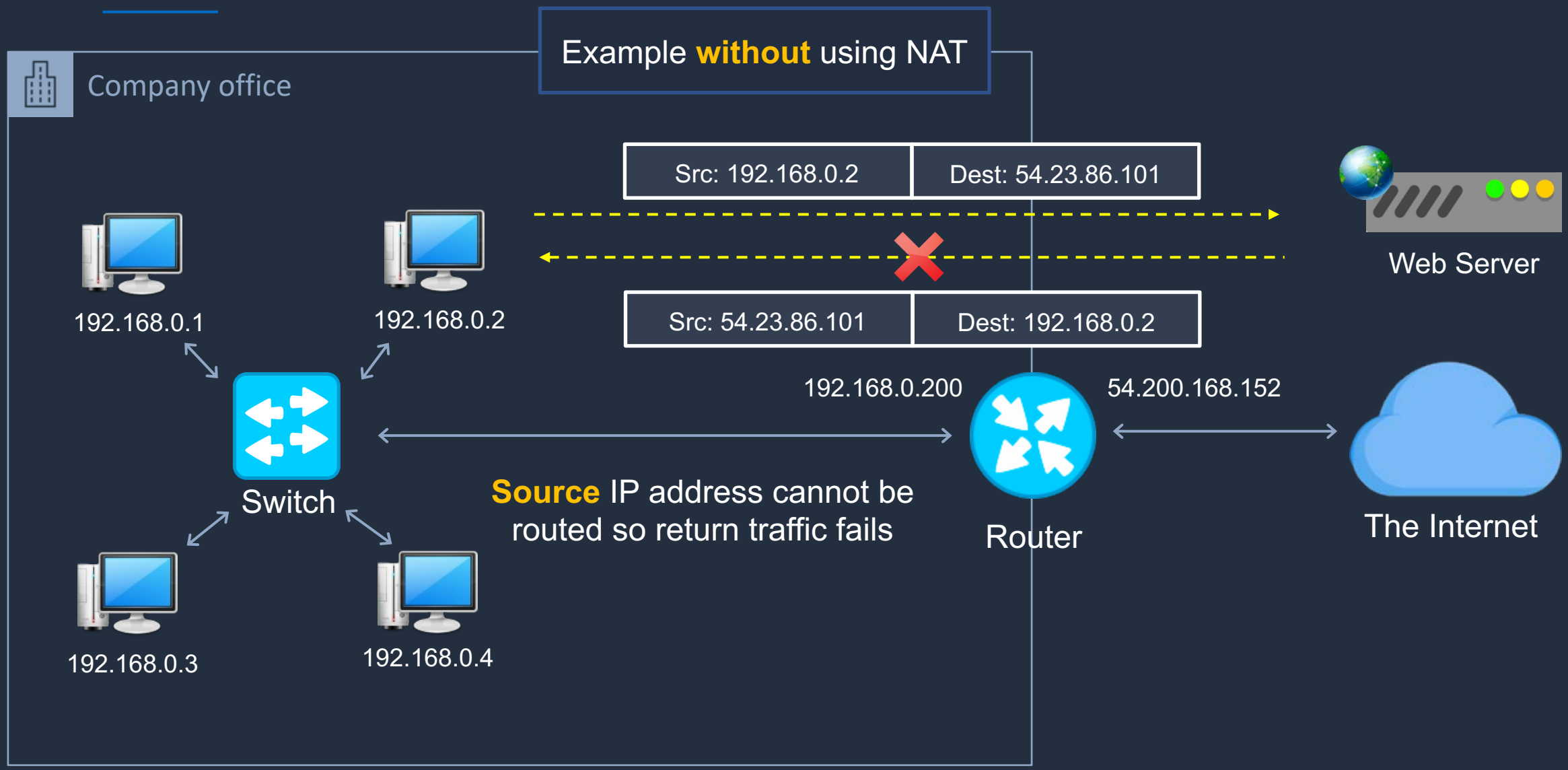
192.168.0.3

192.168.0.4

Private IP addresses are not routable on the Internet

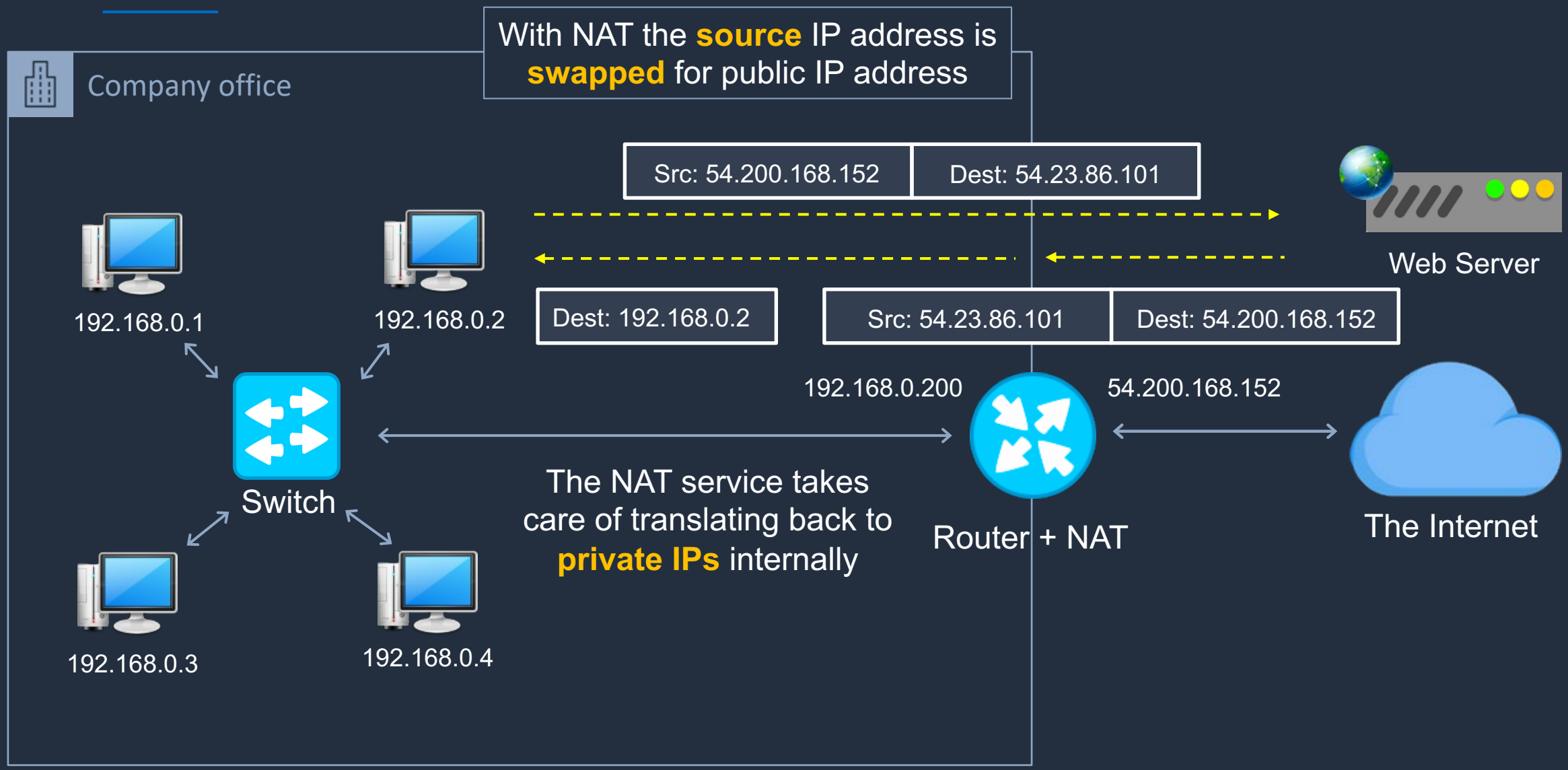
In this configuration computers with **private** addresses **cannot communicate** on the **Internet**

Network Address Translation (NAT)



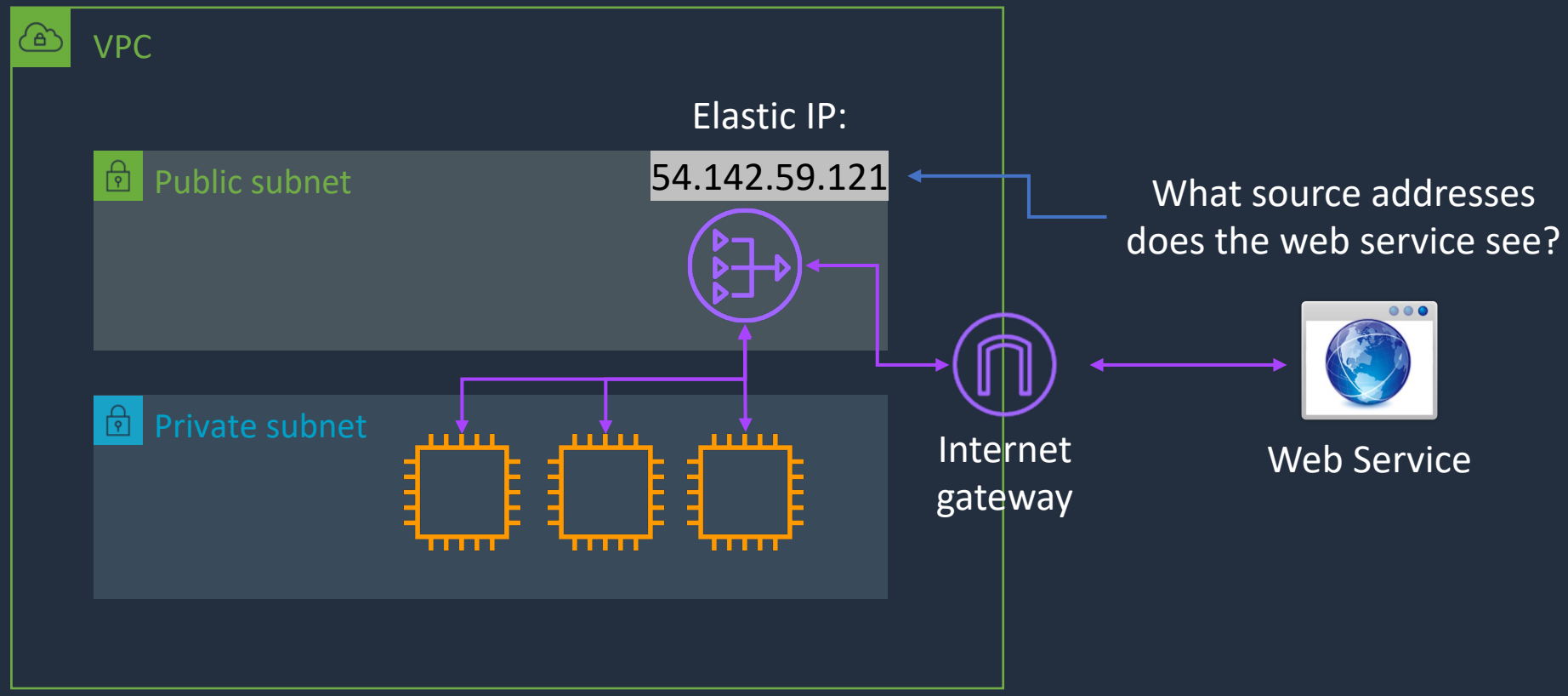
Network Address Translation (NAT)

With NAT the **source** IP address is **swapped** for public IP address



NAT Gateways

Private Subnet Route Table	
Destination	**Target**
10.0.0.0/16	Local
0.0.0.0/0	nat-gateway-id

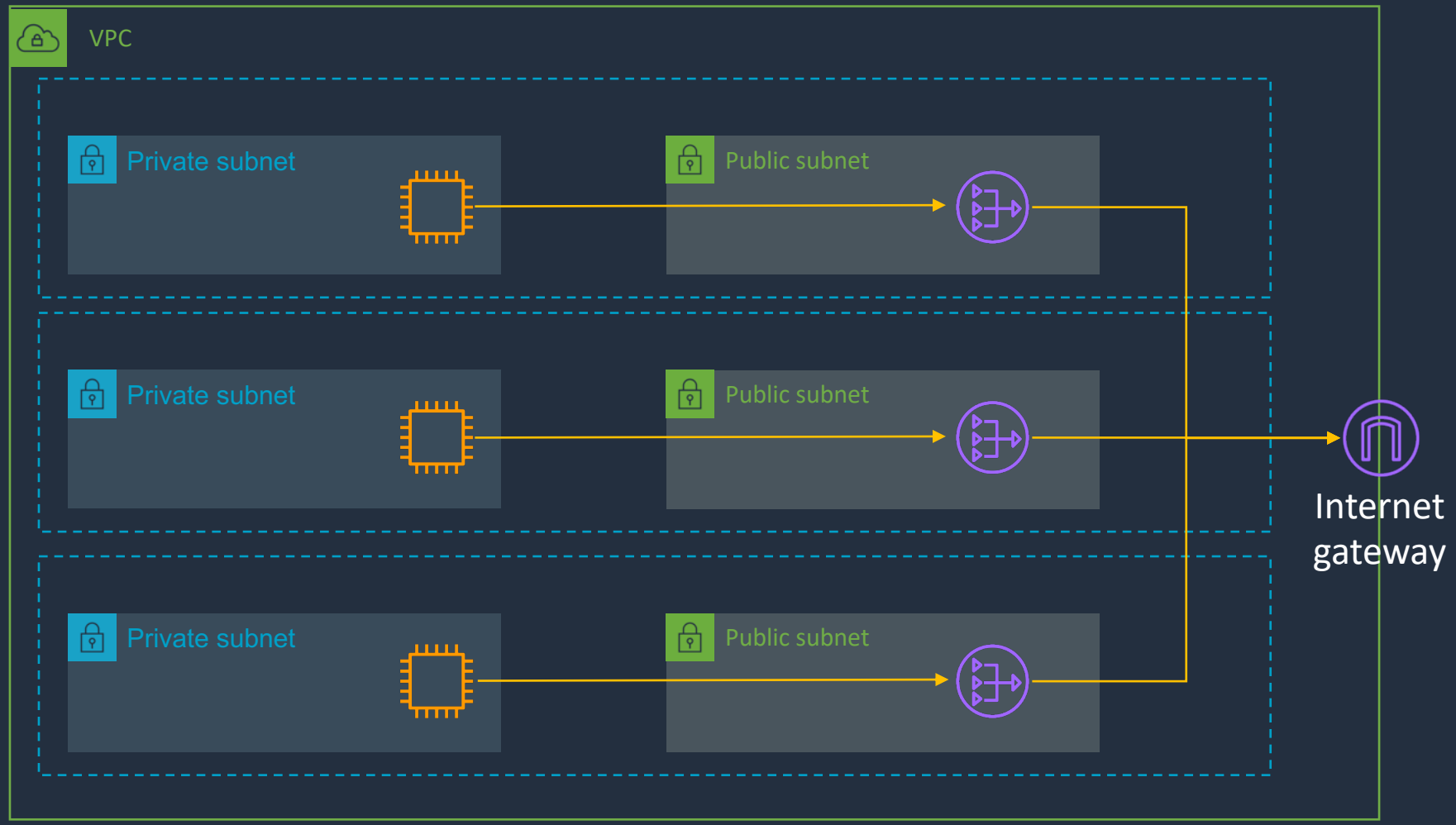


NAT Gateways deployment requirements:

- Deployed in public subnets
- Have an Elastic IP attached
- Route to the NAT gateway added to private subnet

High Availability for NAT Gateway

NAT gateways have redundancy **within an AZ**



AZ1 Private Route Table

Destination	Target
10.0.0.0/16	Local
0.0.0.0/0	nat-gw-az1

AZ2 Private Route Table

Destination	Target
10.0.0.0/16	Local
0.0.0.0/0	nat-gw-az2

AZ3 Private Route Table

Destination	Target
10.0.0.0/16	Local
0.0.0.0/0	nat-gw-az3

DigitalCloud TRAINING

Questions?

DigitalCloud TRAINING

DNS vs Load Balancing

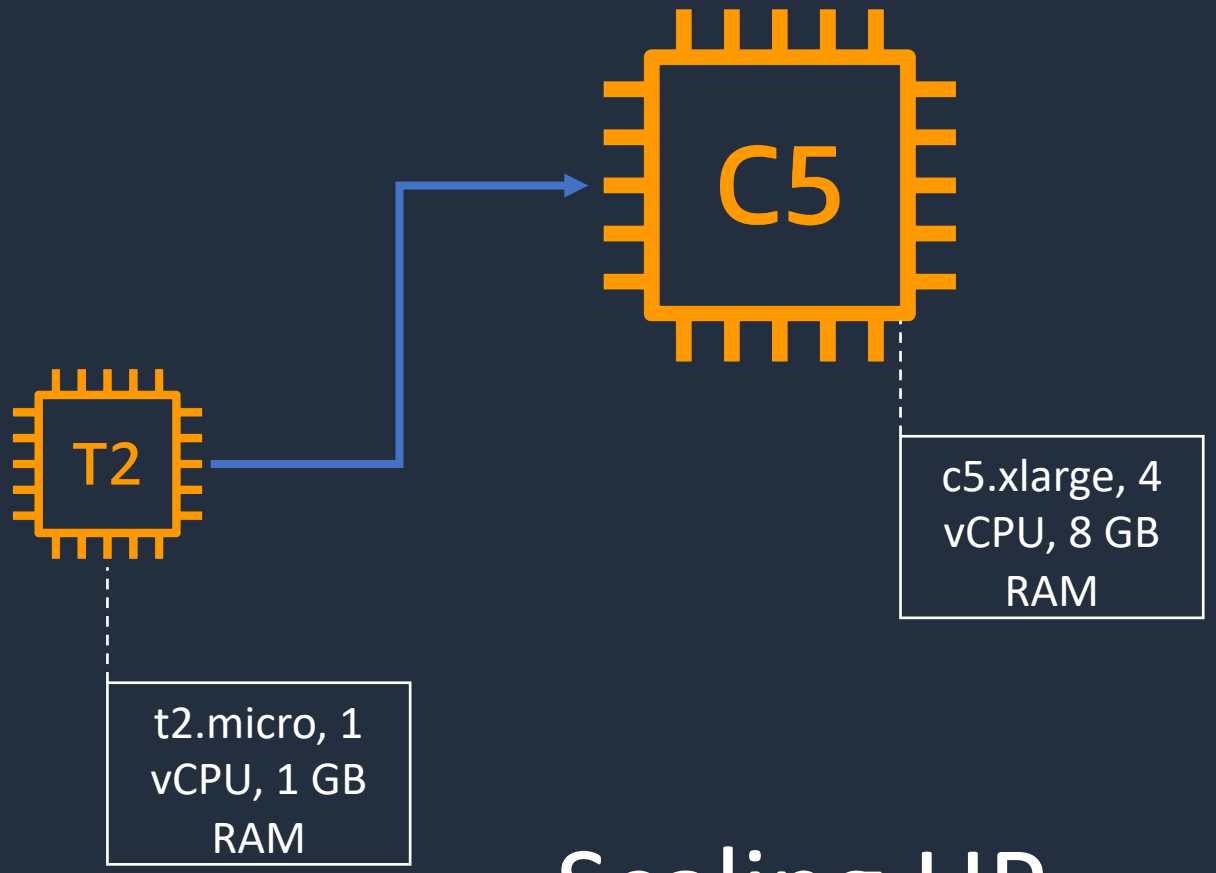
Domain Name System (DNS)

- Responds to queries with an address
- Direct traffic to any Region or non-AWS endpoint

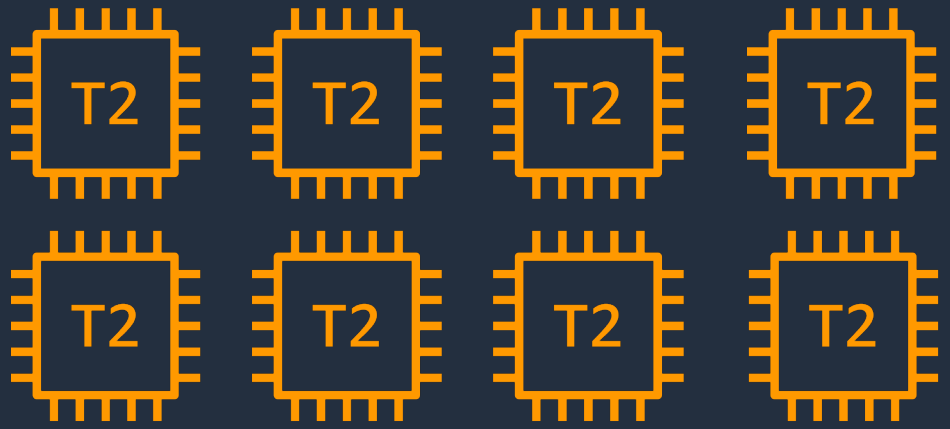
Elastic Load Balancing

- Receives connections and forwards to target
- AWS targets must be within a Region

Scaling Up vs Out

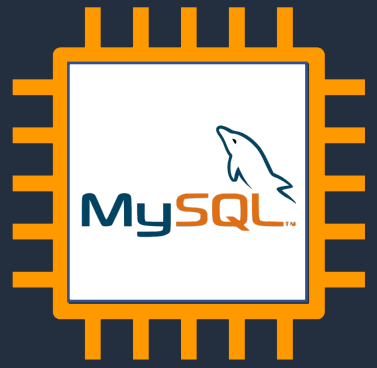


Scaling UP



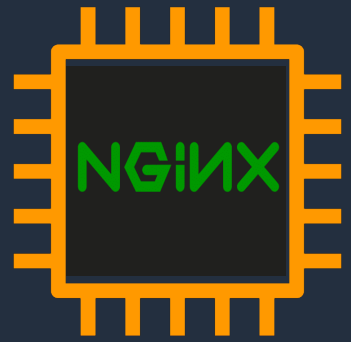
Scaling OUT

Which scaling model should be used?



EC2 with MySQL DB

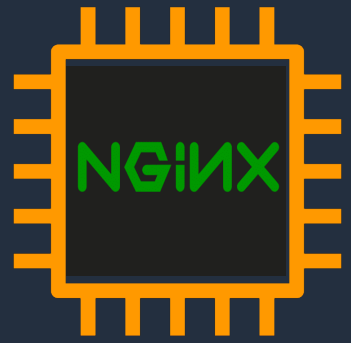
Scale UP



EC2 with **Static** Website

Scale OUT

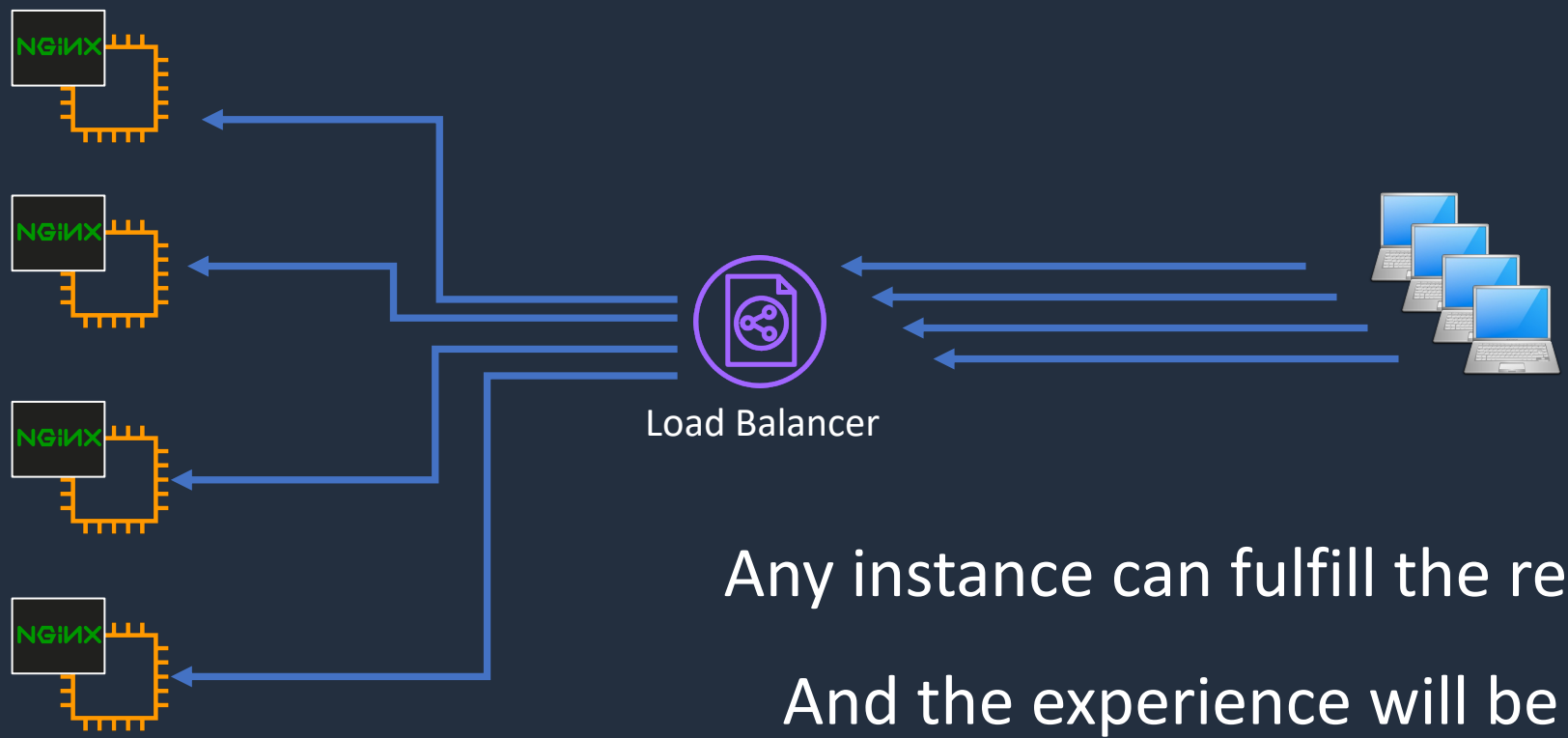
Which scaling model should be used?



...It depends

EC2 with **Dynamic** Website

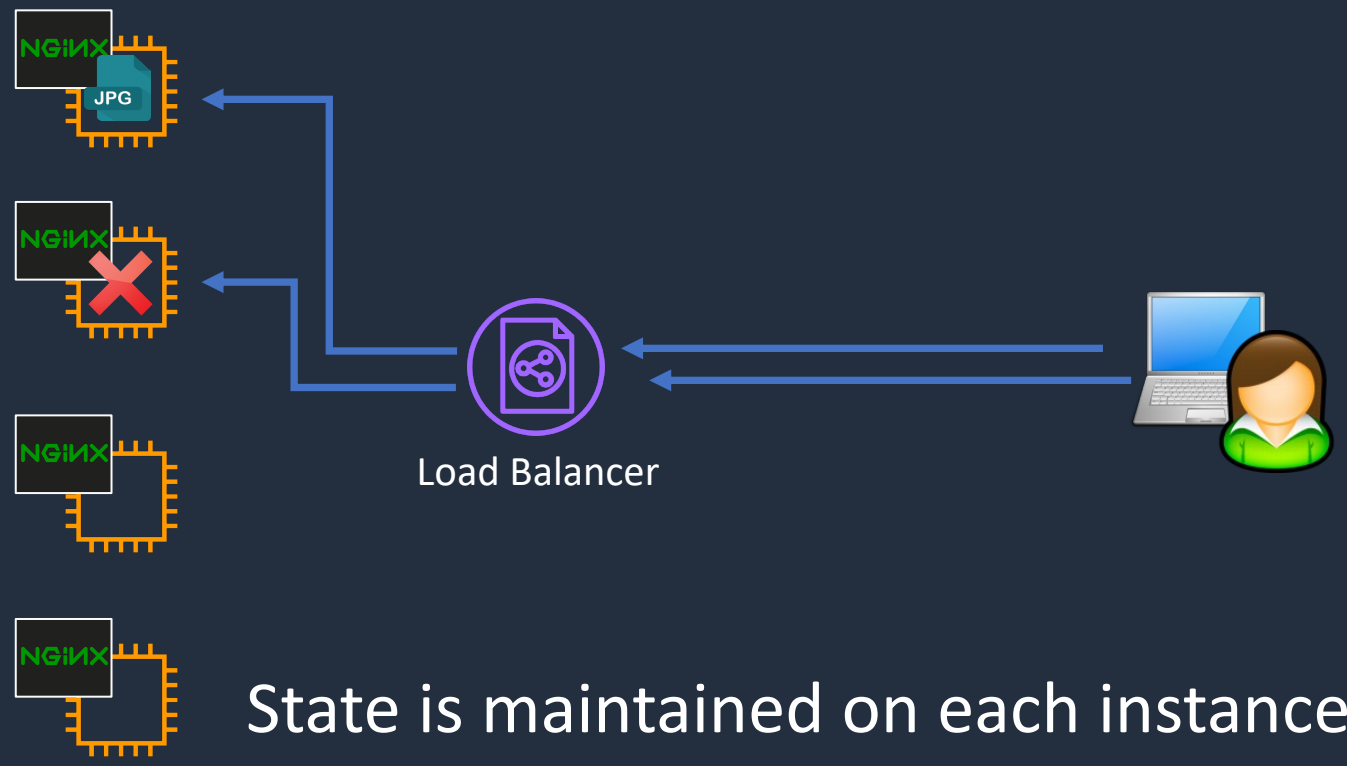
Stateless Application (static website)



Any instance can fulfill the request

And the experience will be the same

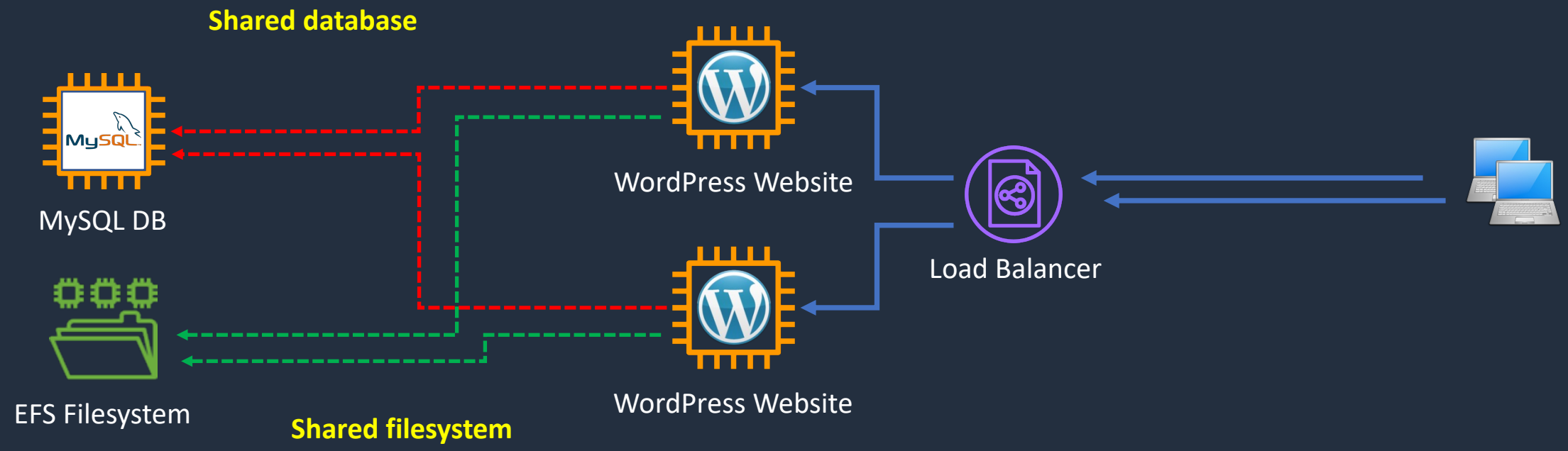
Stateful Application (dynamic website)



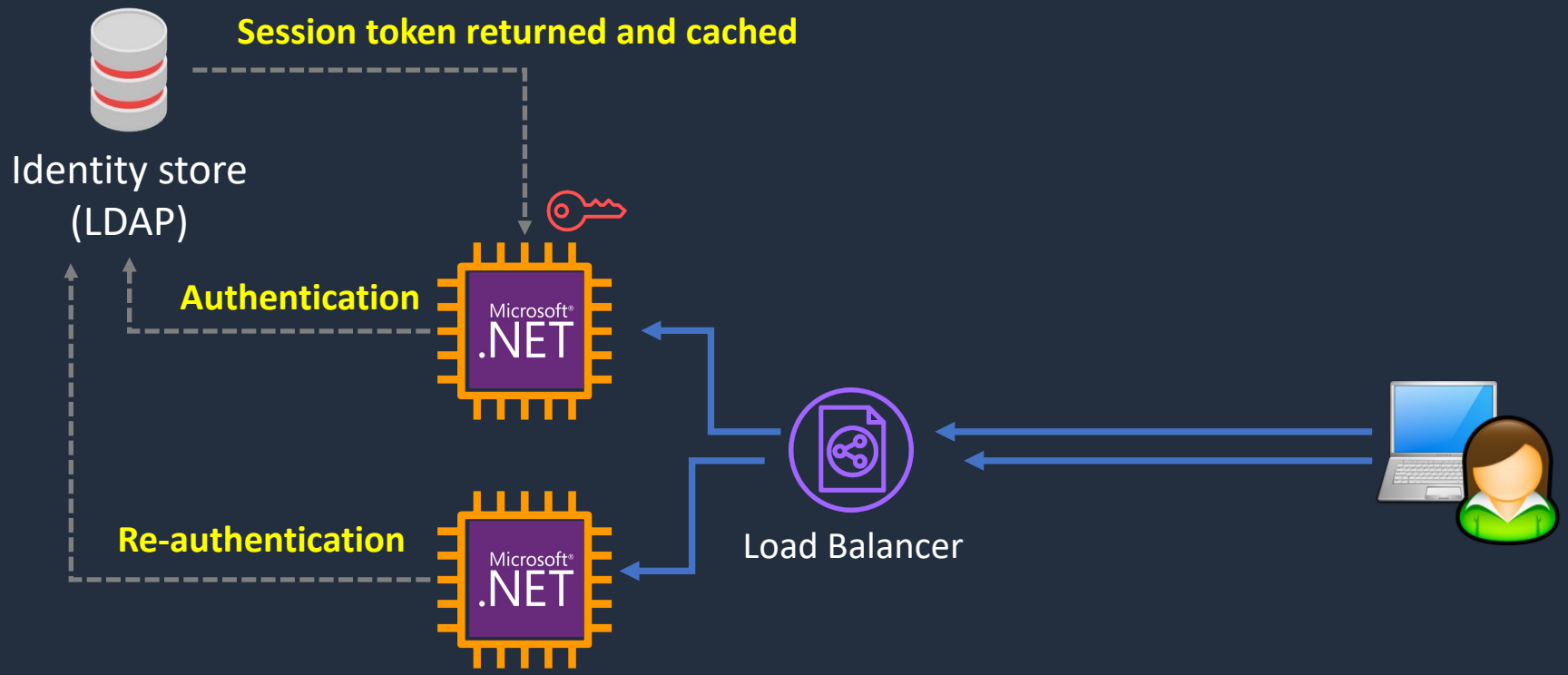
State is maintained on each instance

Experience may be inconsistent

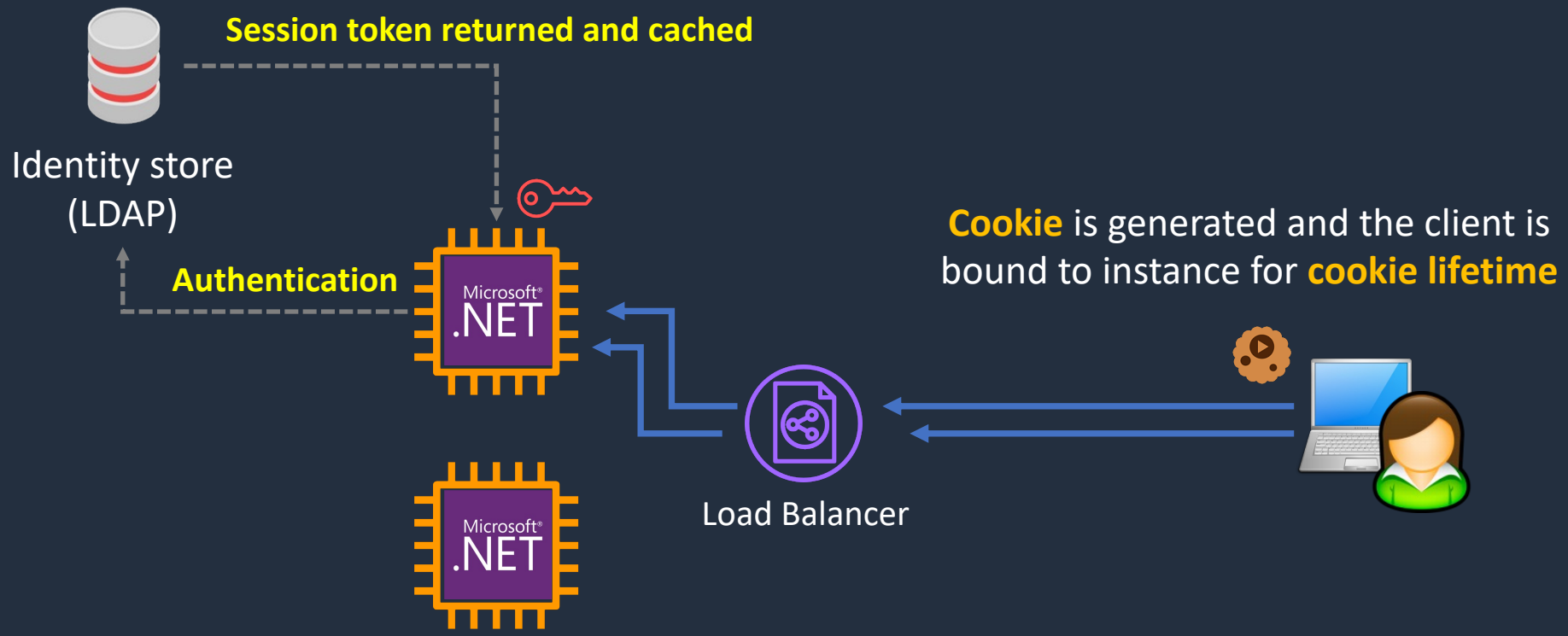
Load Balanced Dynamic Website



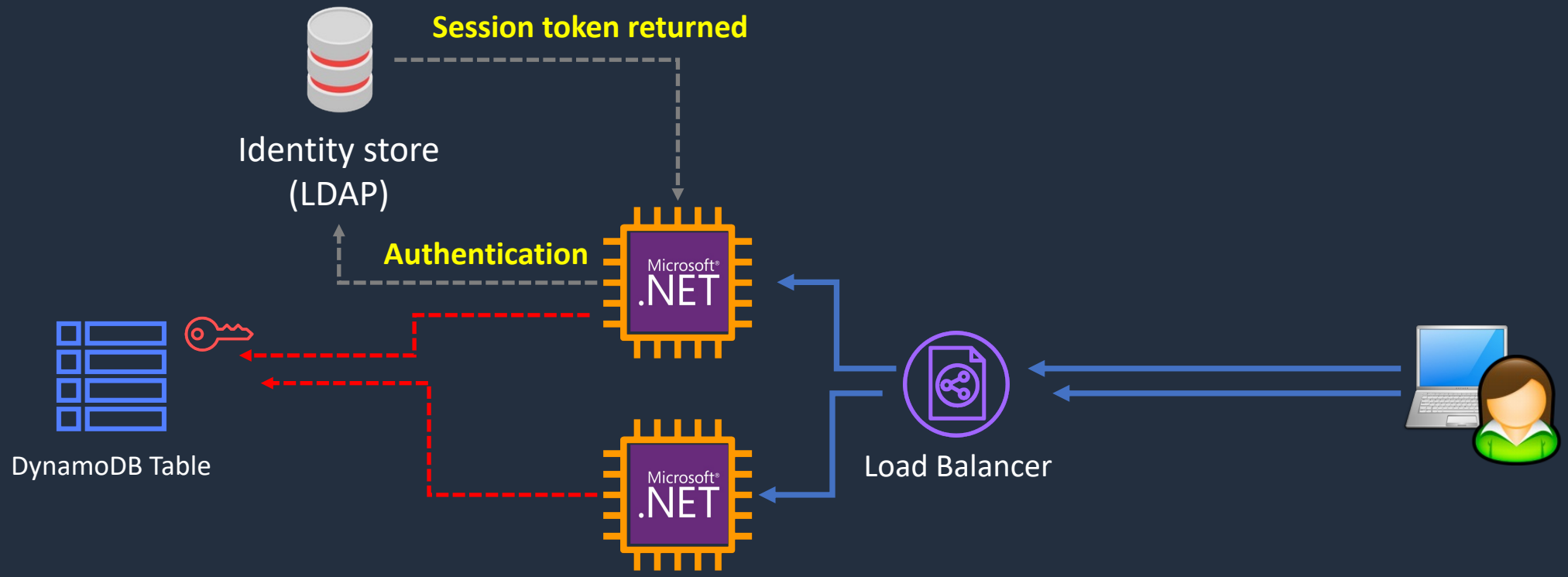
Without Sticky Sessions



With Sticky Sessions



Storing Session State



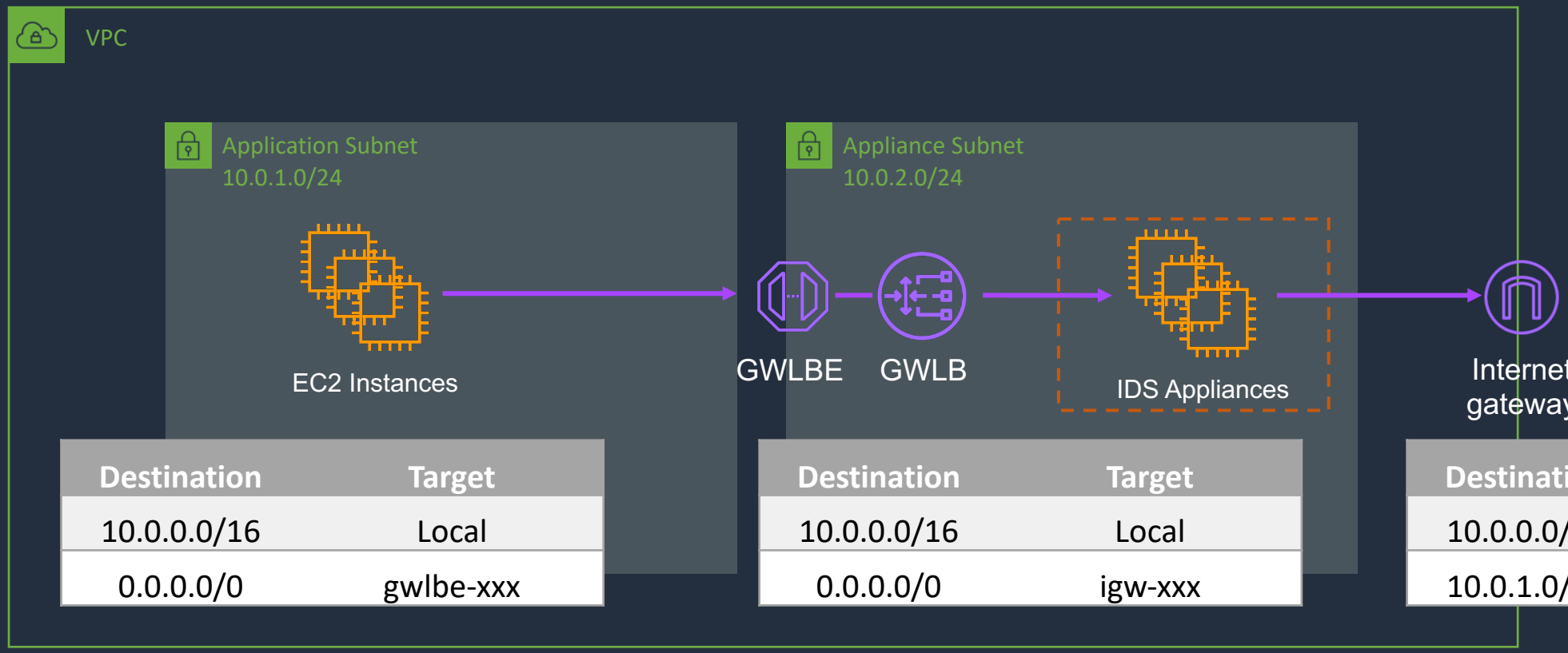
ElastiCache is also a popular solution for storing **session-state data**

Questions?

DigitalCloud TRAINING

Gateway Load Balancer Deployments

GLB endpoint is a target in the subnet route table



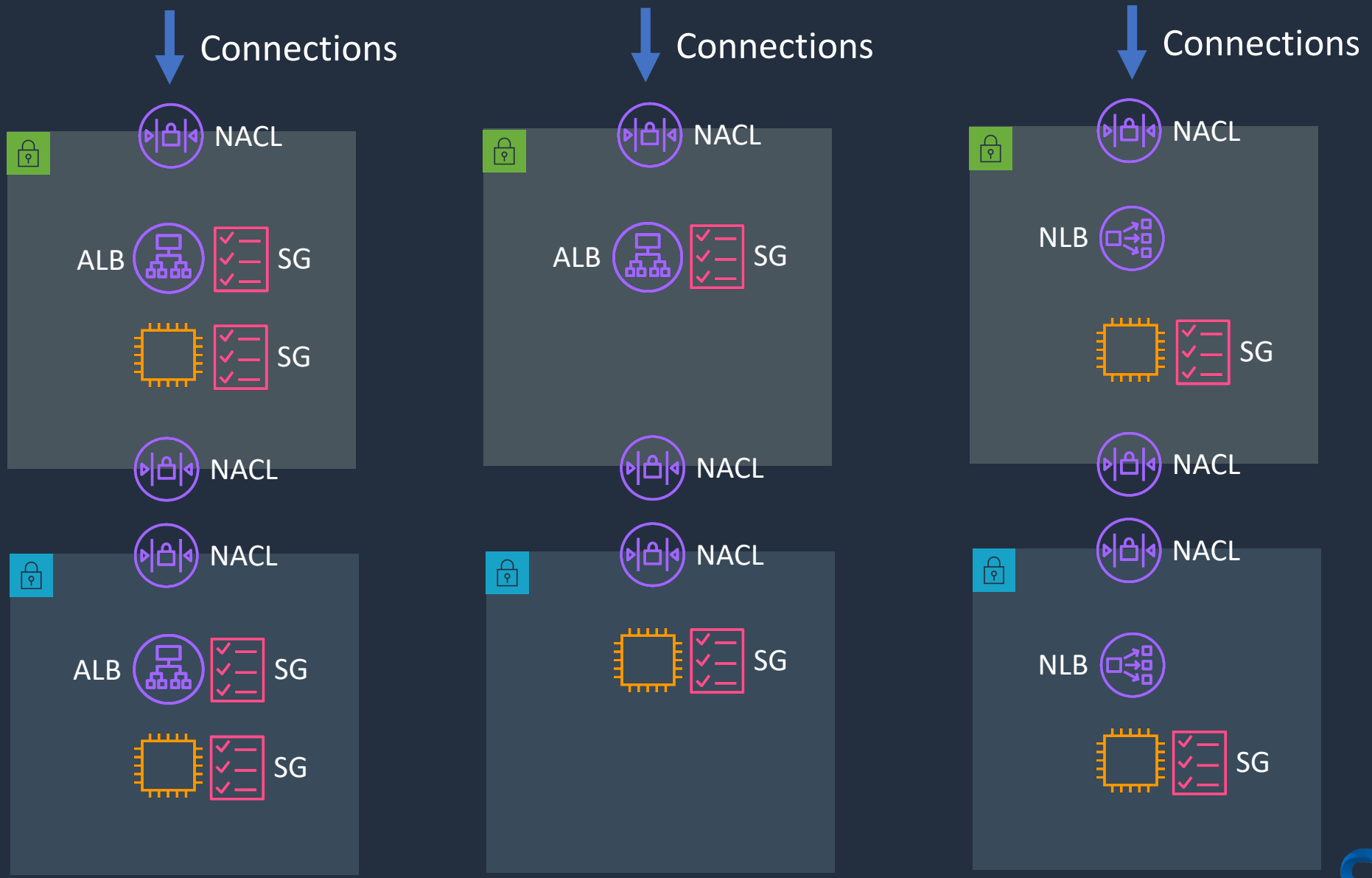
Destination	Target
10.0.0.0/16	Local
0.0.0.0/0	gwlbe-xxx

Destination	Target
10.0.0.0/16	Local
0.0.0.0/0	igw-xxx

Destination	Target
10.0.0.0/16	Local
10.0.1.0/24	gwlbe-xxx

DigitalCloud TRAINING

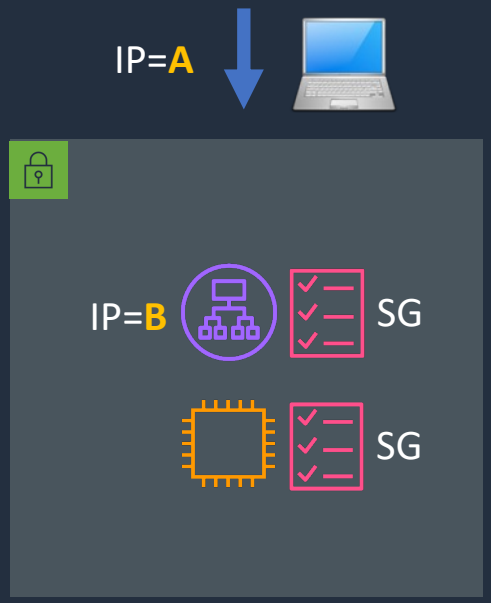
Access Control with ALB and NLB



What's the Source IP Address the App sees?

Note: X-forwarded-for can be used with ALB to capture client IPs

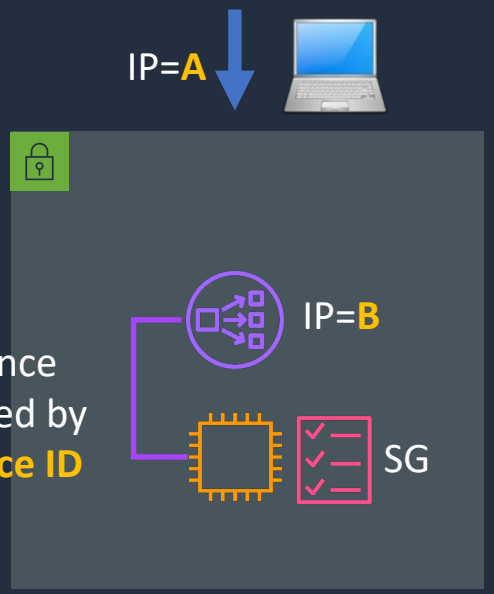
AWS ALB



CLB and ALB use **private IP** of their **ENIs** as source address

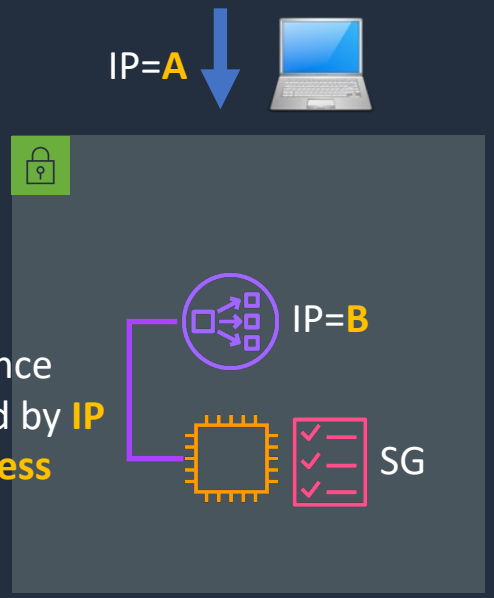
Source	Protocol	Port
IP=B	TCP	80

AWS NLB



Source	Protocol	Port
IP=A	TCP	80

AWS NLB



Source	Protocol	Port
IP=B	TCP	80

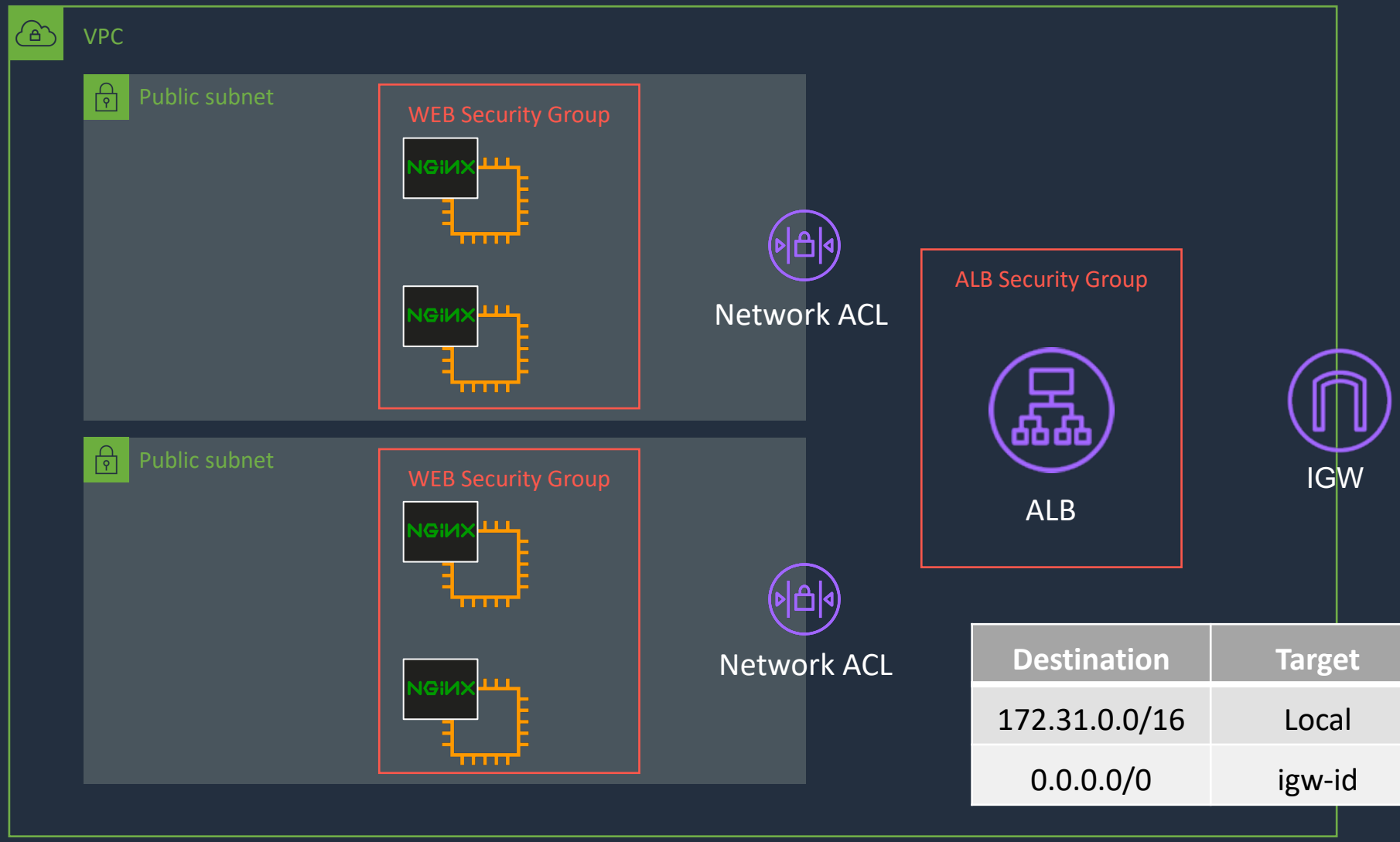
Applicable to **TCP and TLS** – for **UDP and TCP_UDP** should be IP=**A**

When using an NLB with a **VPC Endpoint** or **AWS GA** source IPs are **private IPs** of **NLB nodes**

Troubleshooting Load Balancer Deployments

What must be configured correctly?

- Listener / target ports
- Health check ports
- Web service running
- Security group for EC2
- Security group for ALB
- Network ACL
- Internet gateway
- Route tables



Destination	Target
172.31.0.0/16	Local
0.0.0.0/0	igw-id

Questions?

DigitalCloud TRAINING

Hands-On Practice Session

Load Balanced Architecture with Advanced Request Routing

Exercise 1 - Create the Red and Blue EC2 instances
Exercise 2 - Enable Path-based Routing
Exercise 3 - Enable Host-based Routing
Bonus (if time allows) **-** Add Auto Scaling

Use the Lab Guide from the resources page:

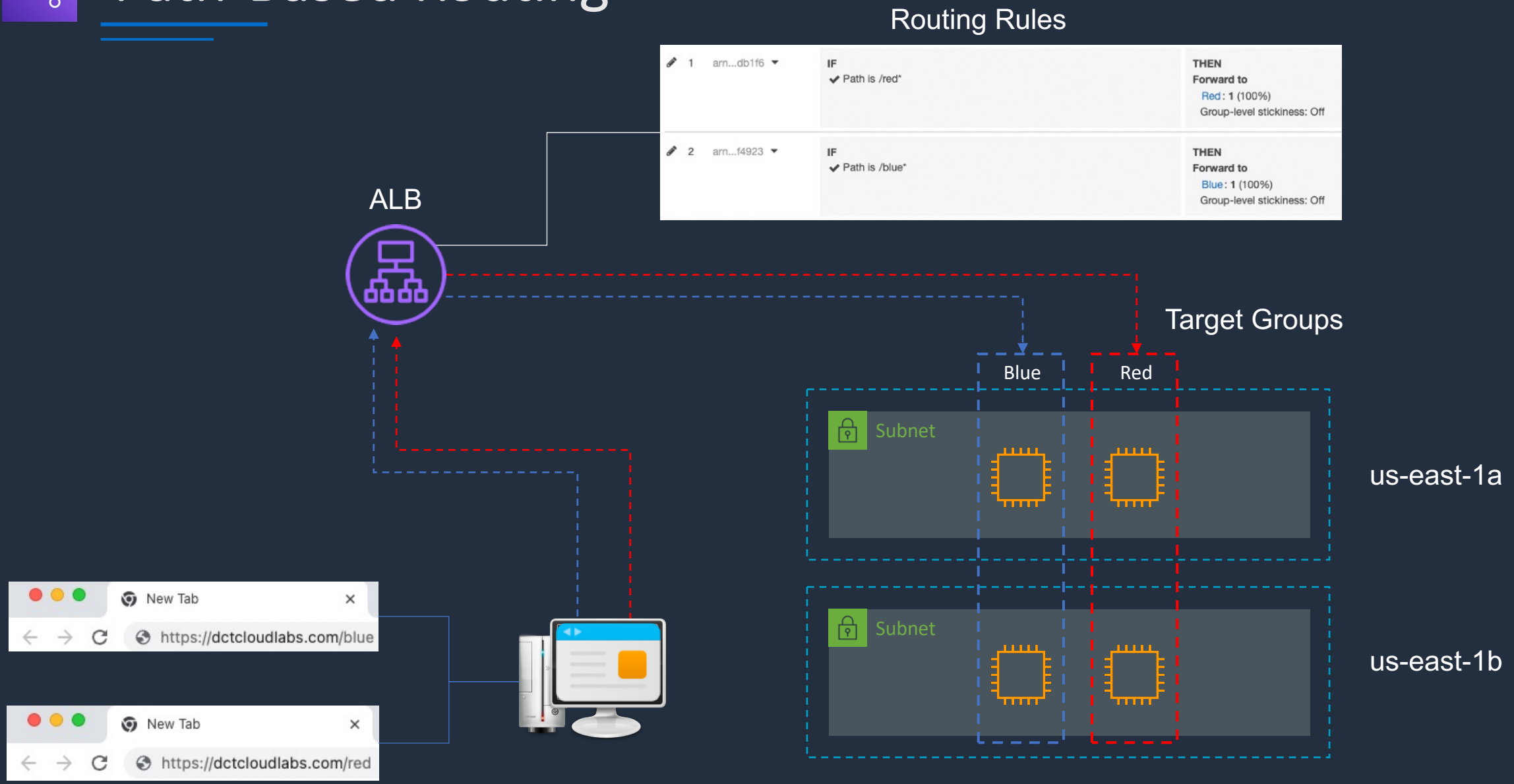
- Load Balanced Architecture with Advanced Request Routing.pdf

Download the code:

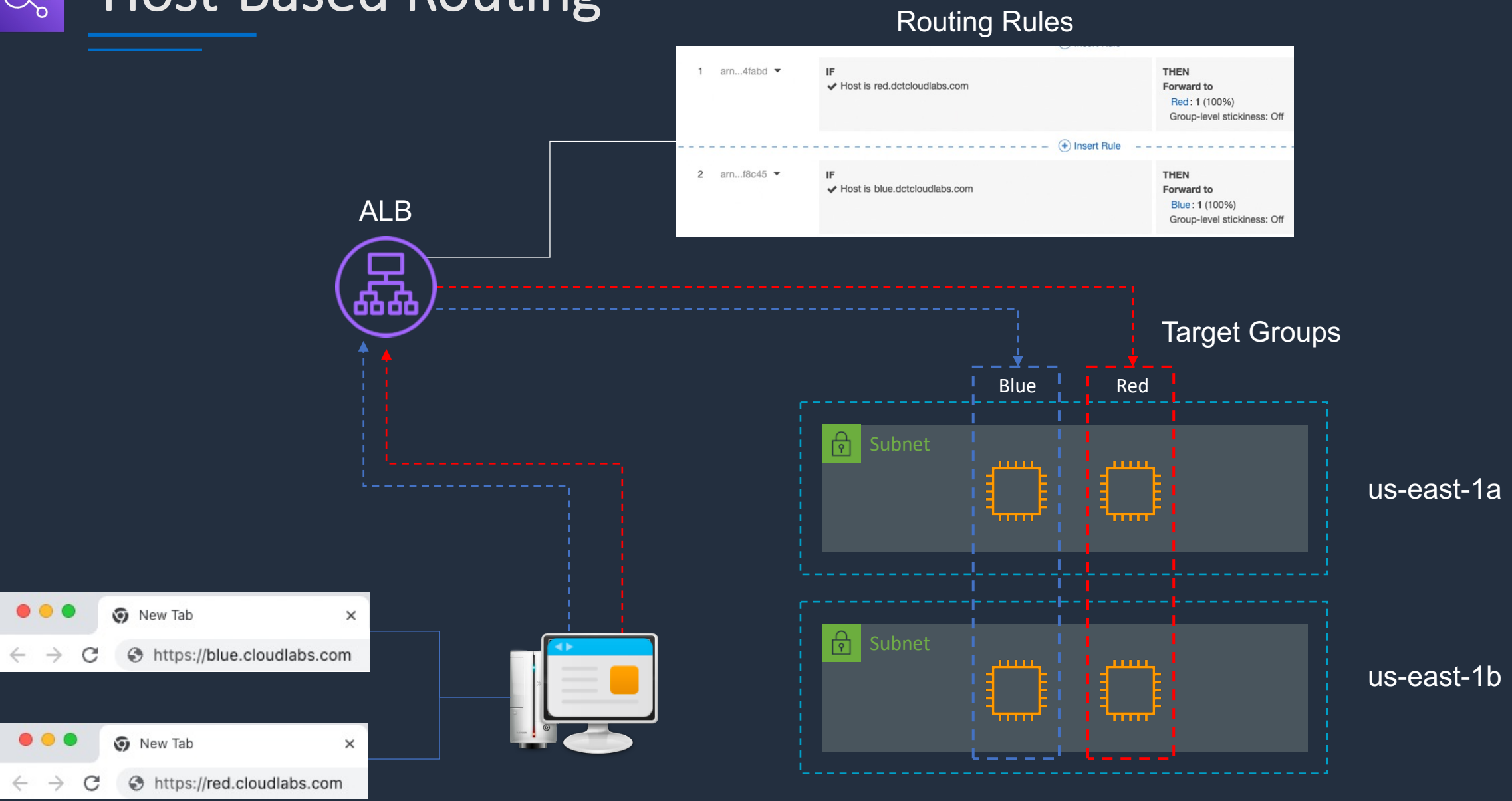
- advanced-request-routing-code.zip

DigitalCloud TRAINING

Path-Based Routing



Host-Based Routing



Final Questions?