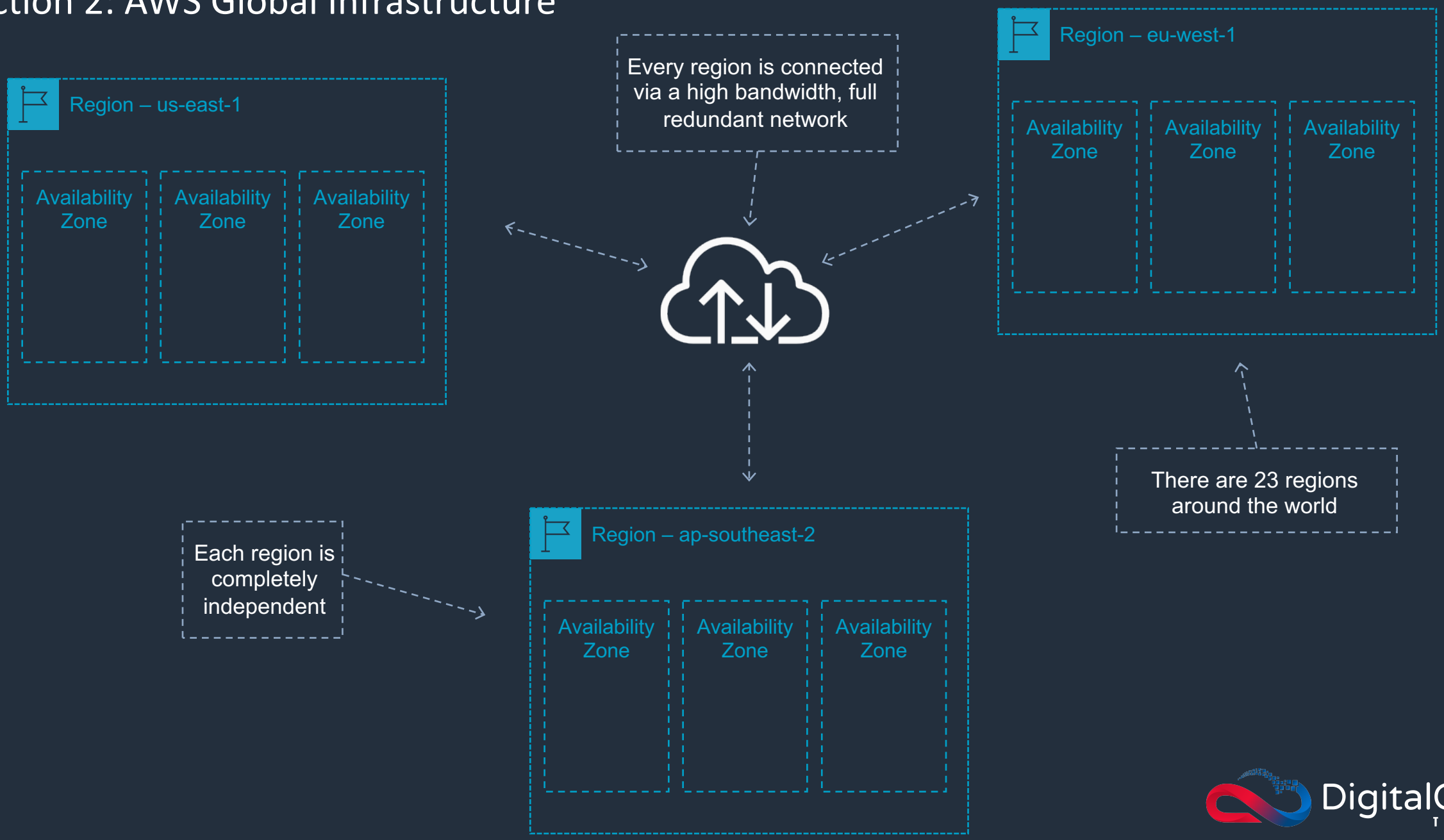
Section 2: AWS Global Infrastructure

Name	Description
Region	A geographical area with 2 or more AZs, isolated from other AWS regions
Availability Zone (AZ)	One or more data centers that are physically separate and isolated from other AZs
Edge Location	A location with a cache of content that can be delivered at low latency to users – used by CloudFront
Regional Edge Cache	Also part of the CloudFront network. These are larger caches that sit between AWS services and Edge Locations
Global Network	Highly available, low-latency private global network interconnecting every data center, AZ, and AWS region

Section 2: AWS Global Infrastructure

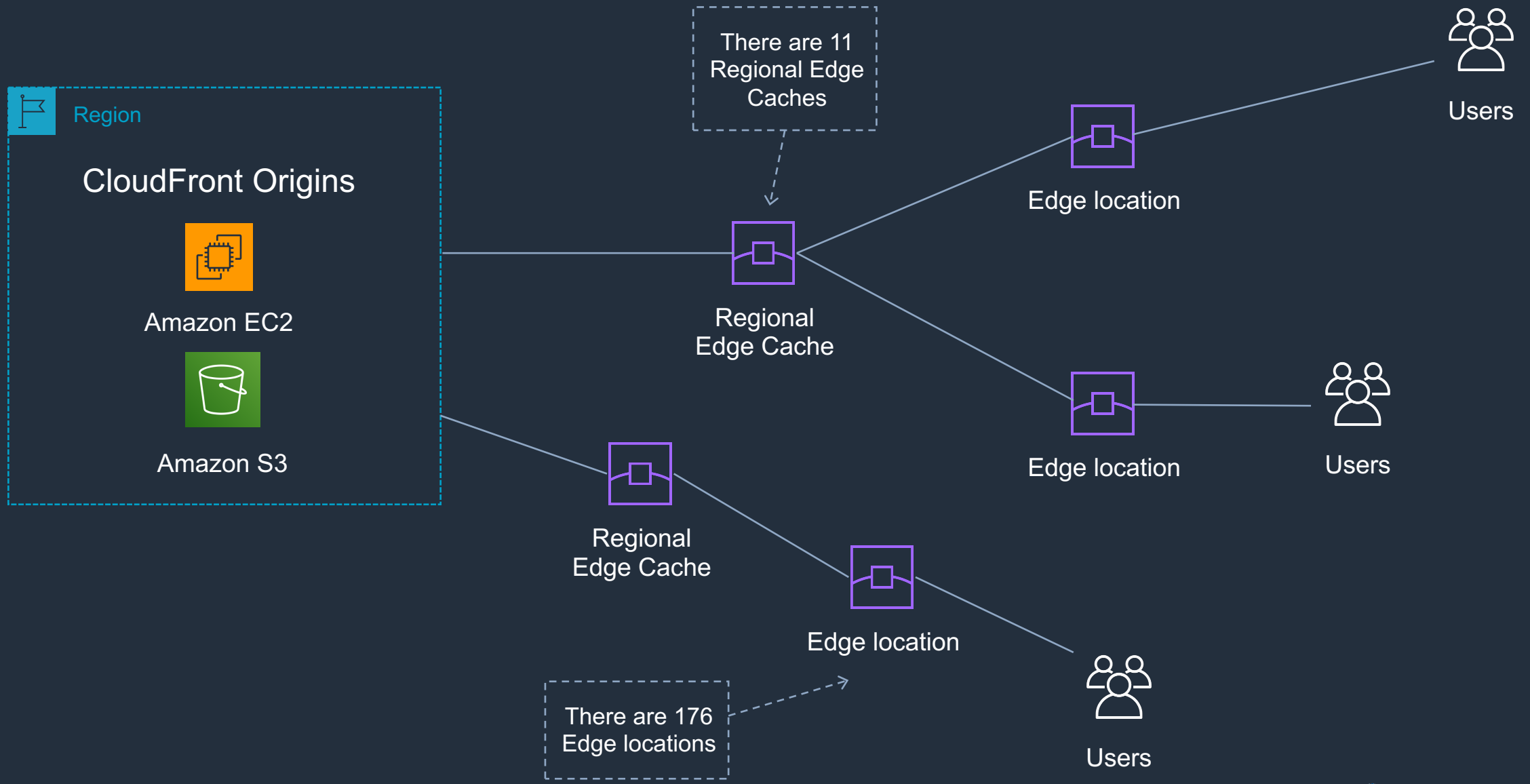


DigitalCloud TRAINING

Section 2: AWS Global Infrastructure

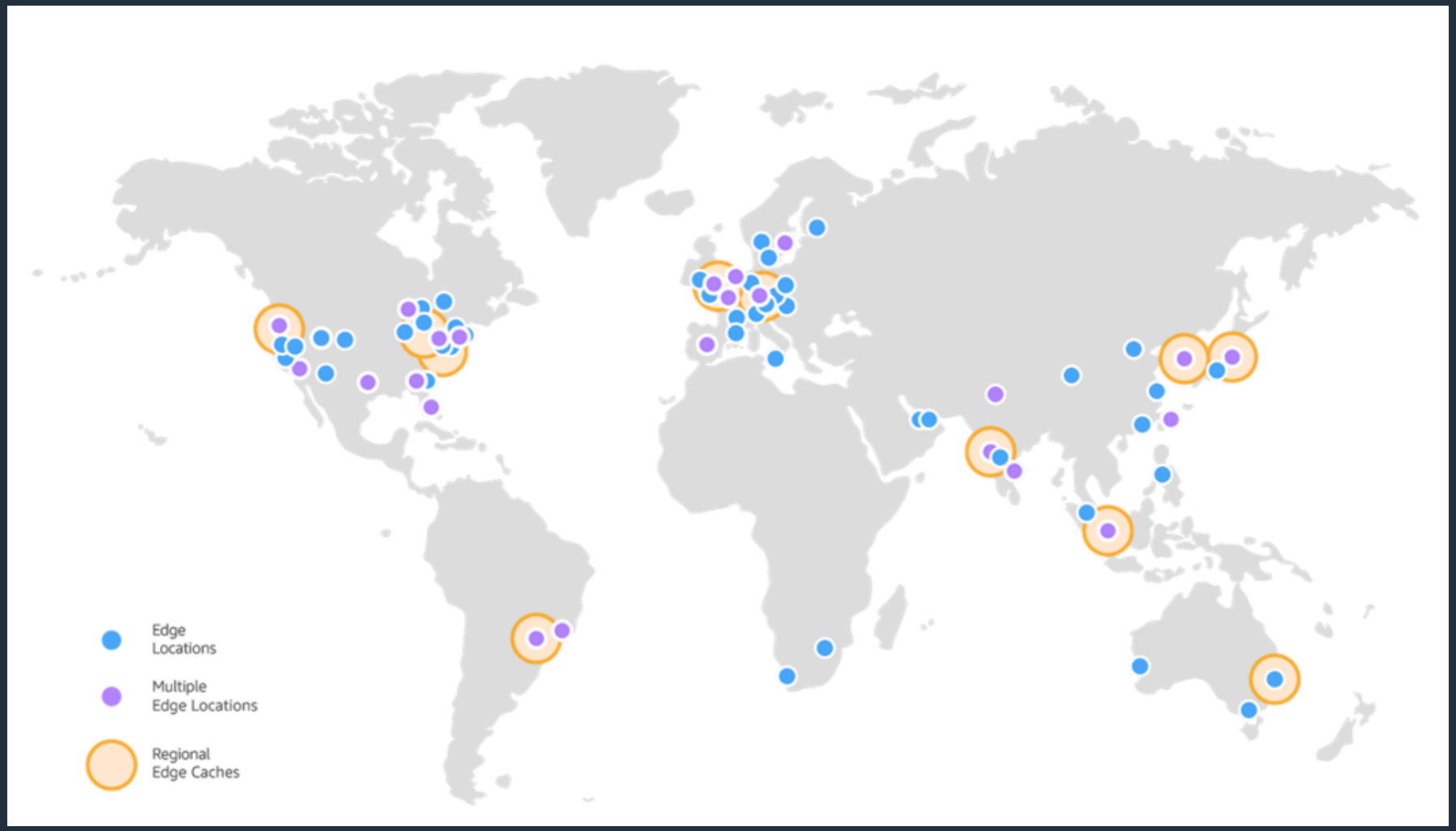
DigitalCloud TRAINING

Section 2: CloudFront Edge Locations



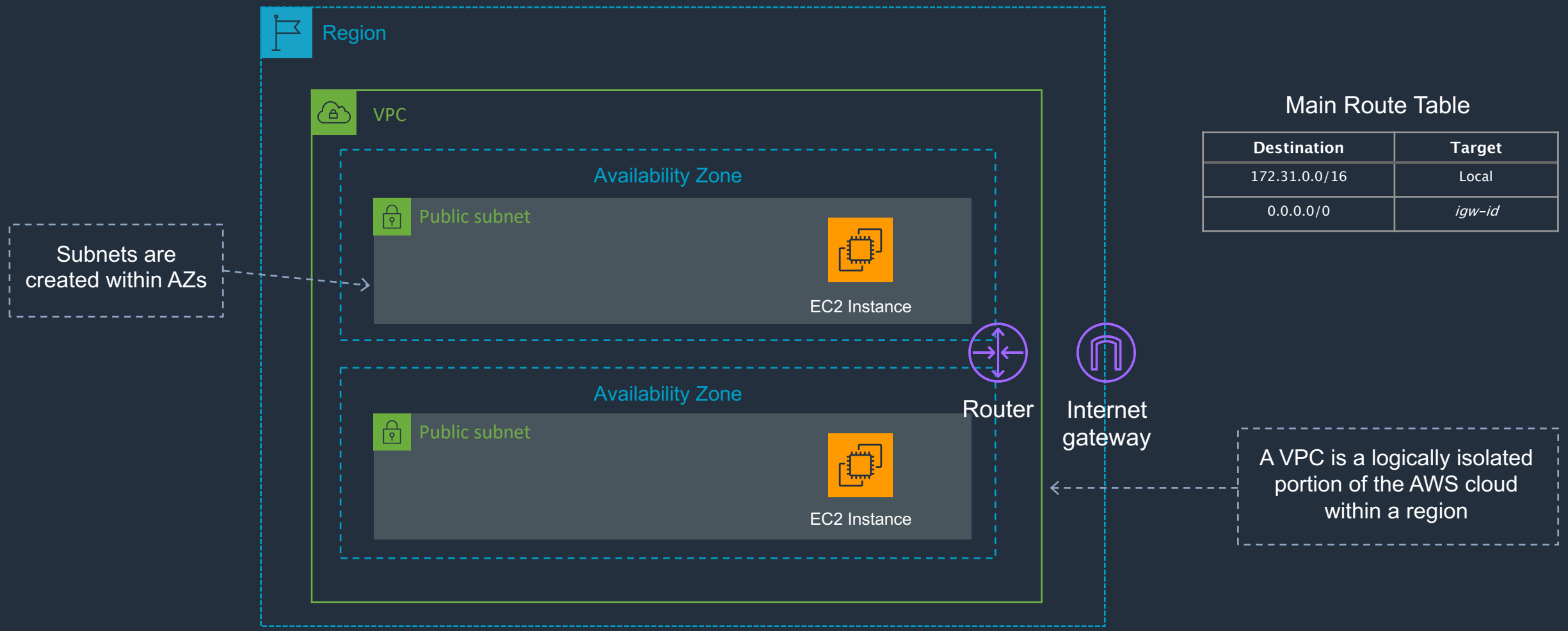
DigitalCloud TRAINING

Section 2: CloudFront Edge Locations



DigitalCloud TRAINING

Section 2: VPC Overview

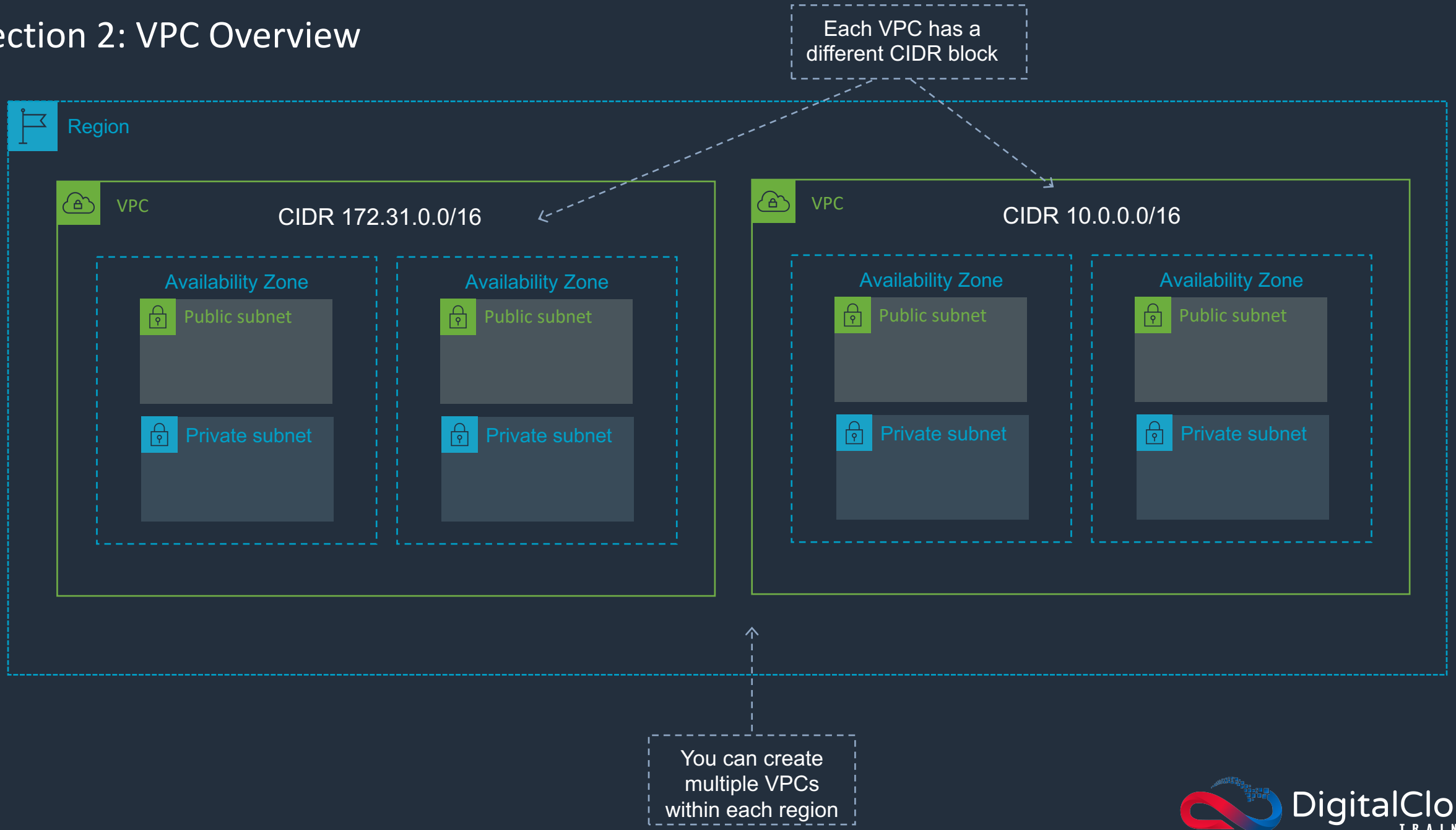


Main Route Table

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

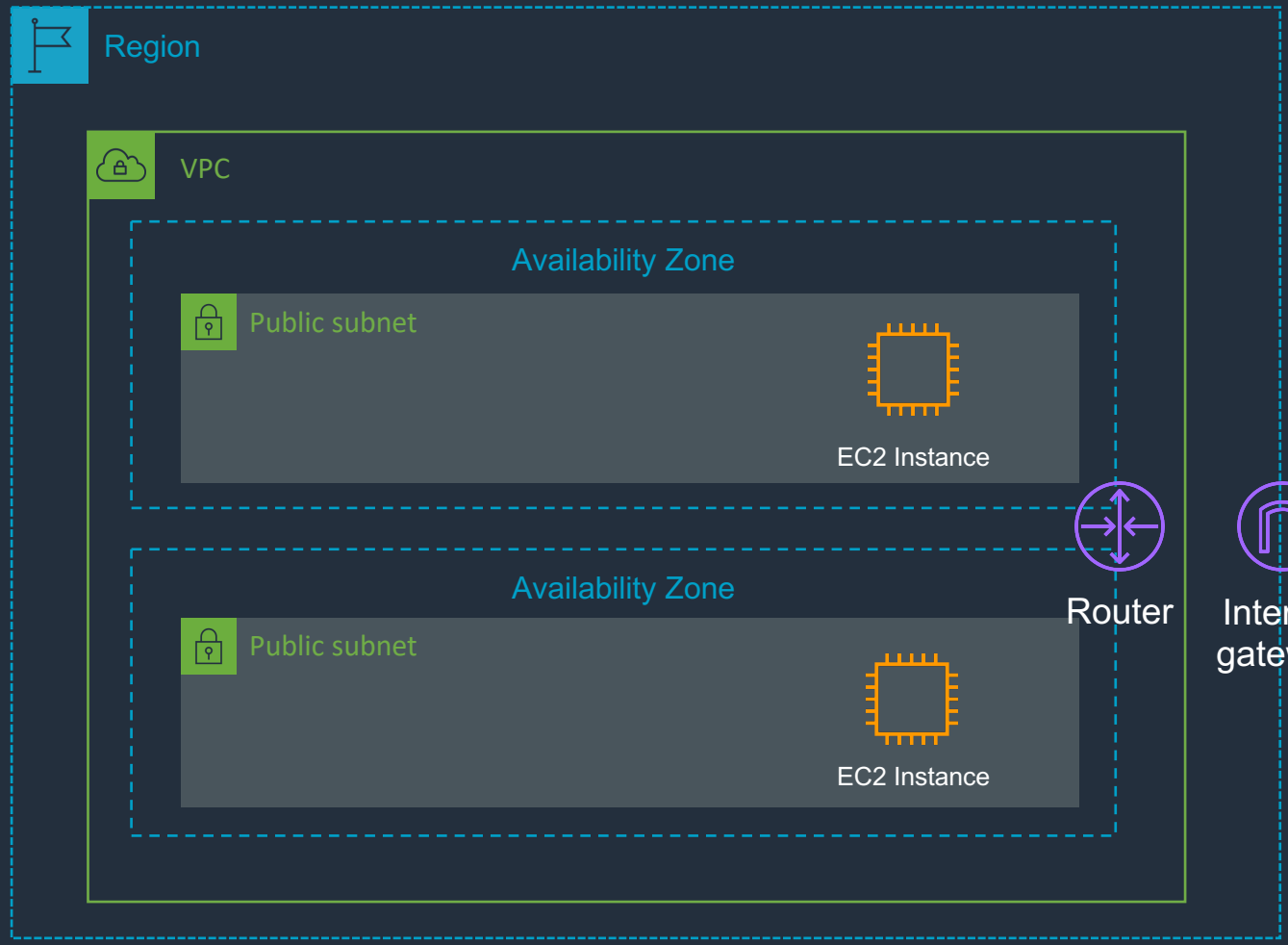
DigitalCloud TRAINING

Section 2: VPC Overview



DigitalCloud TRAINING

Section 3: Launch an EC2 Instance

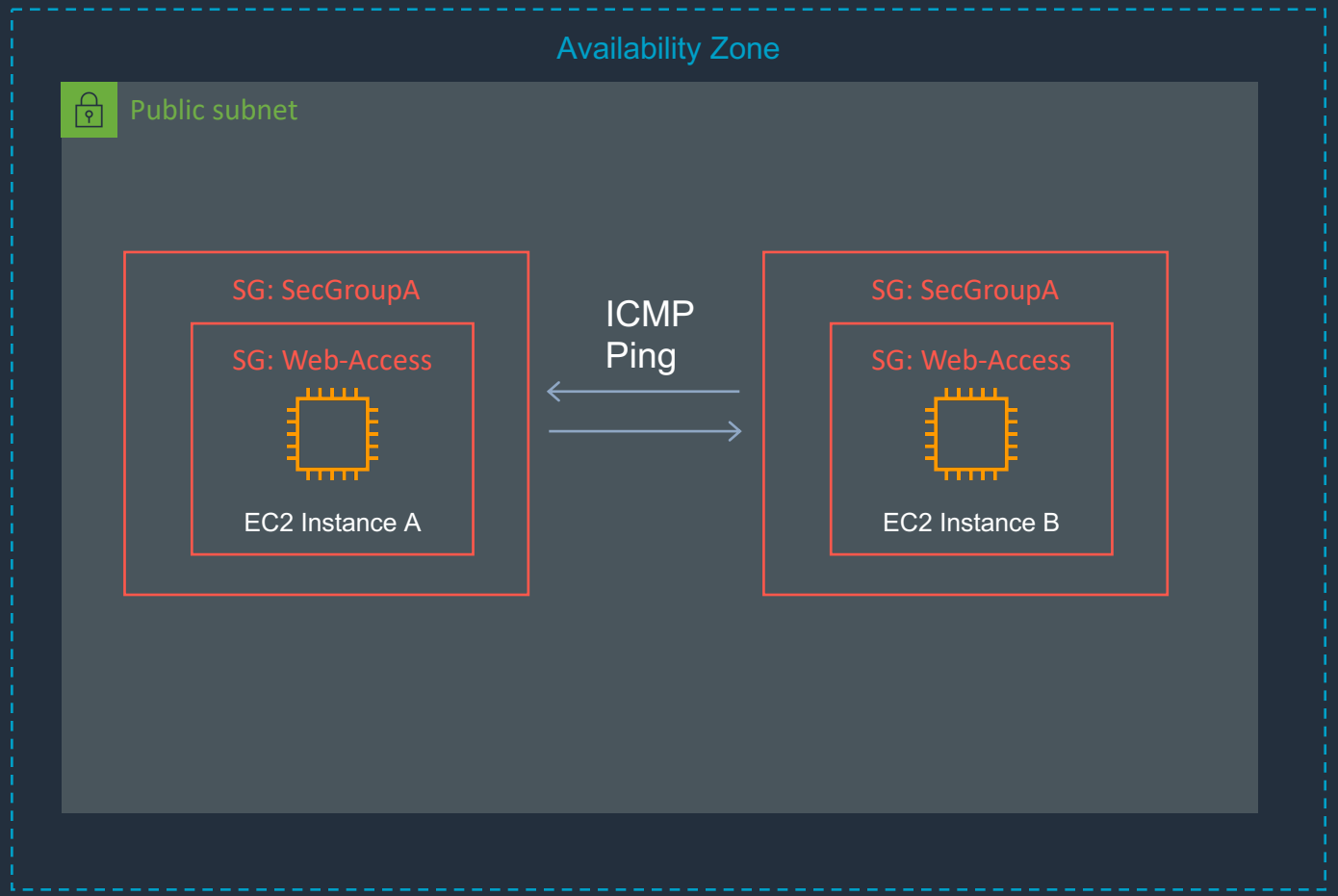


Main Route Table

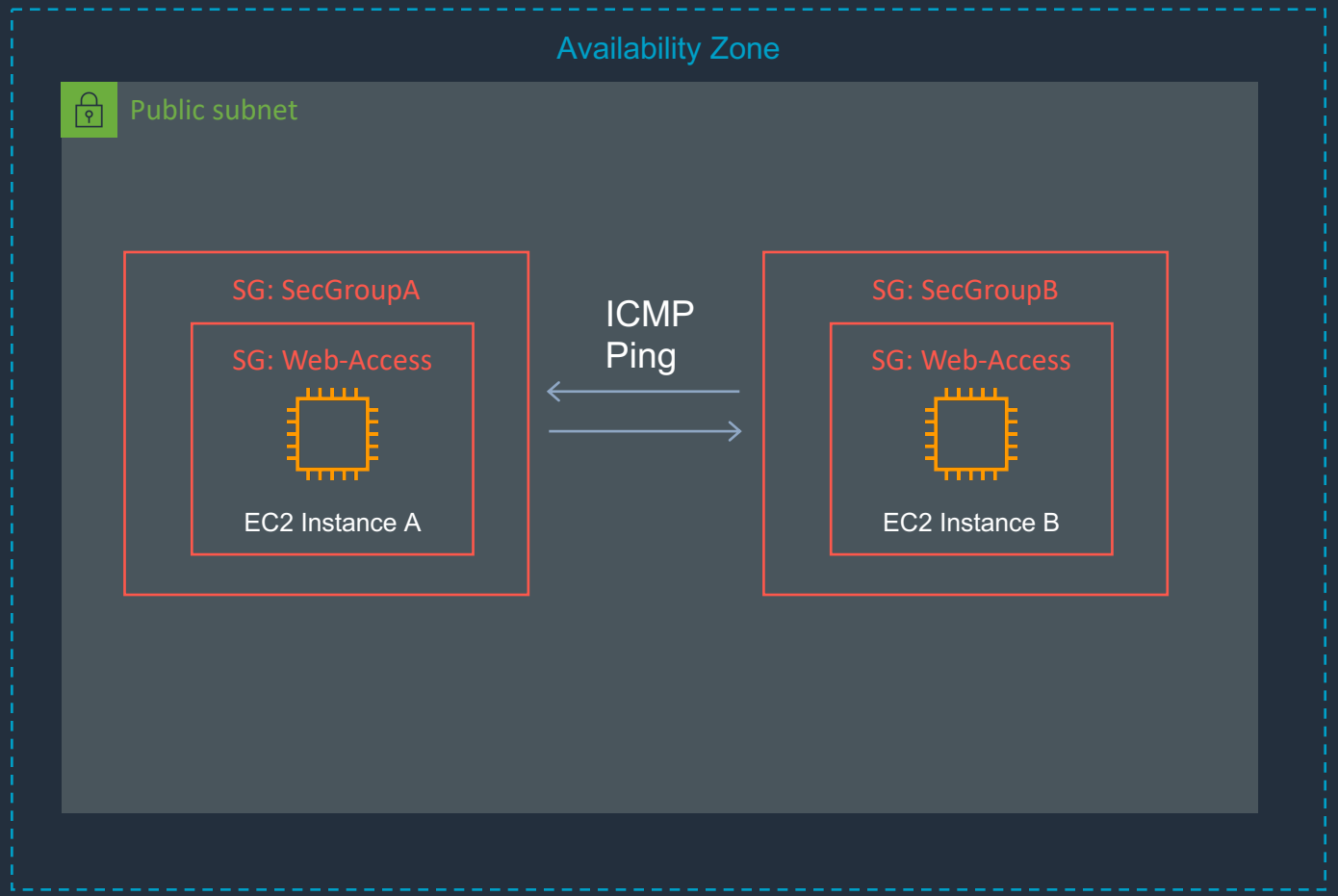
Destination	Target
172.31.0.0/16	Local
0.0.0.0/0	*igw-id*

DigitalCloud TRAINING

Section 3: Security Group Slide 1



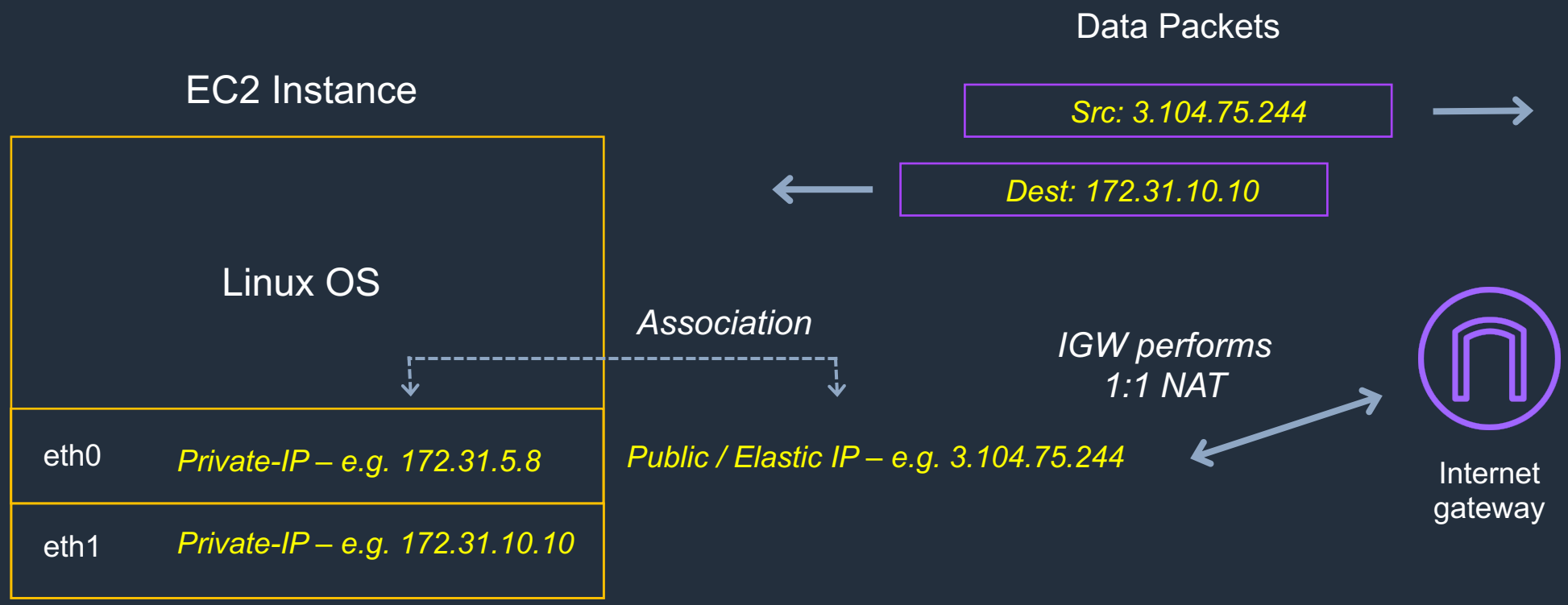
Section 3: Security Group Slide 2



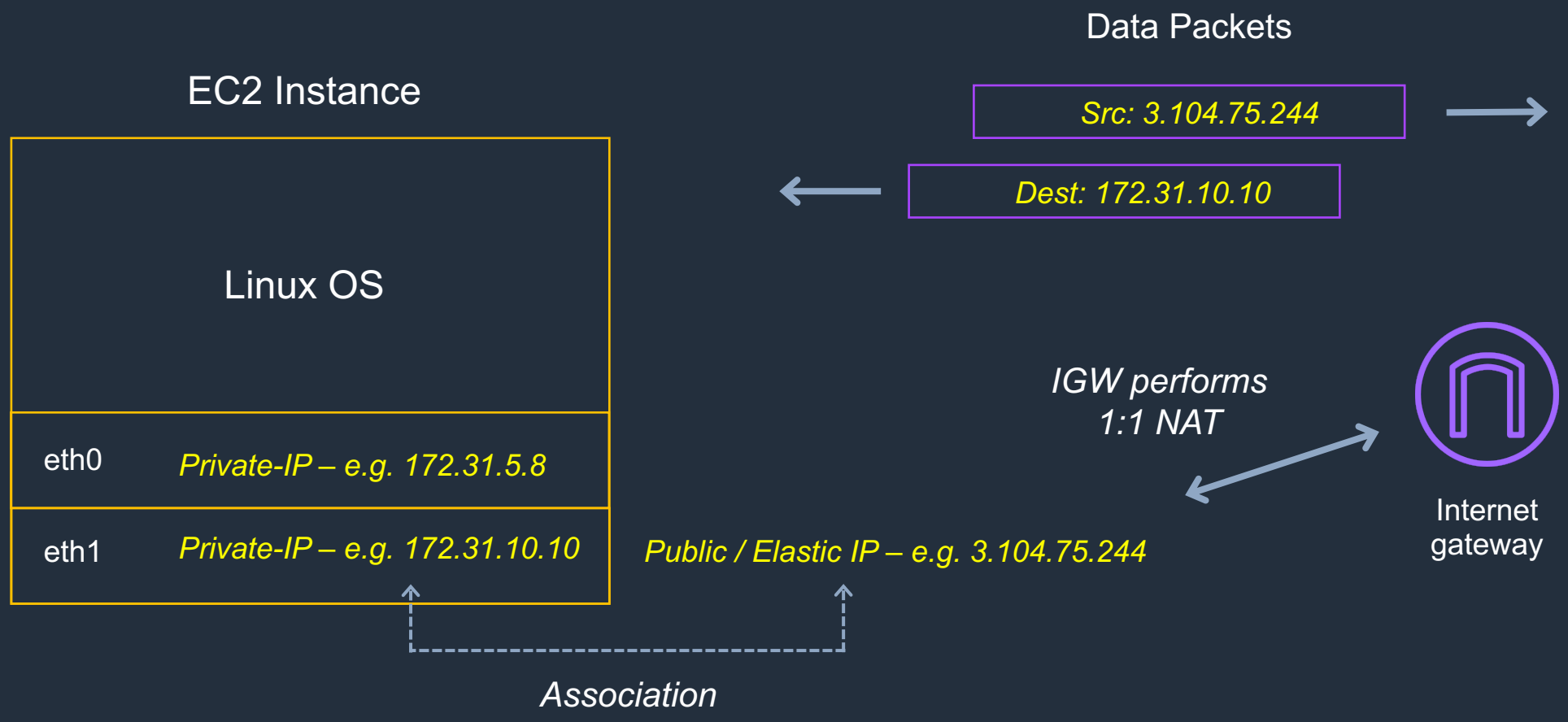
Section 3: Public, Private, and Elastic IP addresses

Name	Description
Public IP address	Lost when the instance is stopped Used in Public Subnets No charge Associated with a private IP address on the instance Cannot be moved between instances
Private IP address	Retained when the instance is stopped Used in Public and Private Subnets
Elastic IP address	Static Public IP address You are charged if not used Associated with a private IP address on the instance Can be moved between instances and Elastic Network Adapters

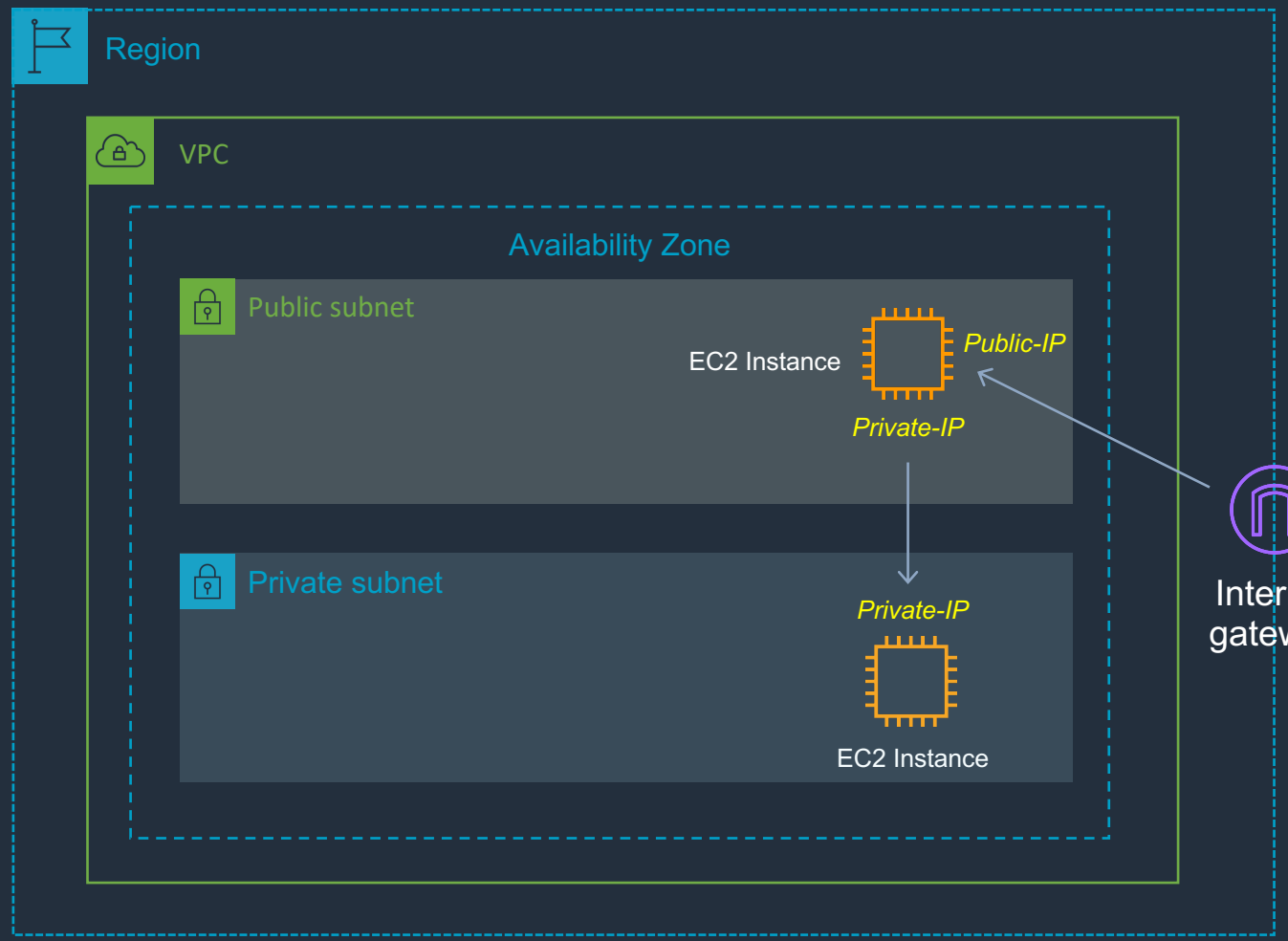
Section 3: Public, Private and Elastic IPs - Slide 1



Section 3: Public, Private and Elastic IPs - Slide 2



Section 3: Private Subnets and Bastion Hosts



Public Subnet Route Table

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

Private Subnet Route Table

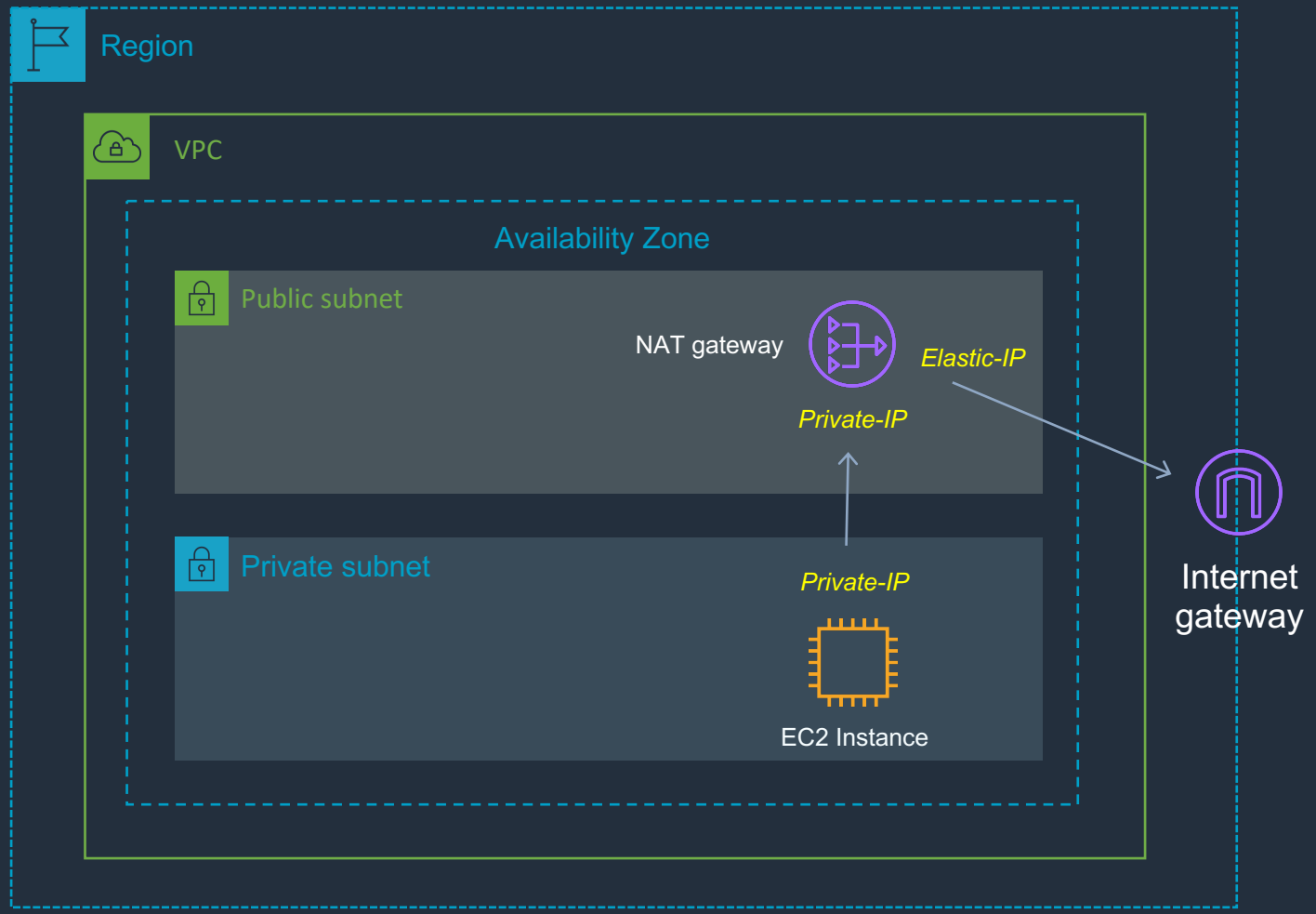
Destination	Target
172.31.0.0/16	Local

DigitalCloud TRAINING

Section 3: NAT Instance vs NAT Gateway

NAT Instance	NAT Gateway
Managed by you (e.g. software updates)	Managed by AWS
Scale up (instance type) manually and use enhanced networking	Elastic scalability up to 45 Gbps
No high availability – scripted/auto-scaled HA possible using multiple NATs in multiple subnets	Provides automatic high availability within an AZ and can be placed in multiple AZs
Need to assign Security Group	No Security Groups
Can use as a bastion host	Cannot access through SSH
Use an Elastic IP address or a public IP address with a NAT instance	Choose the Elastic IP address to associate with a NAT gateway at creation
Can implement port forwarding through manual customisation	Does not support port forwarding

Section 3: Private Subnet with NAT Gateway



Public Subnet Route Table

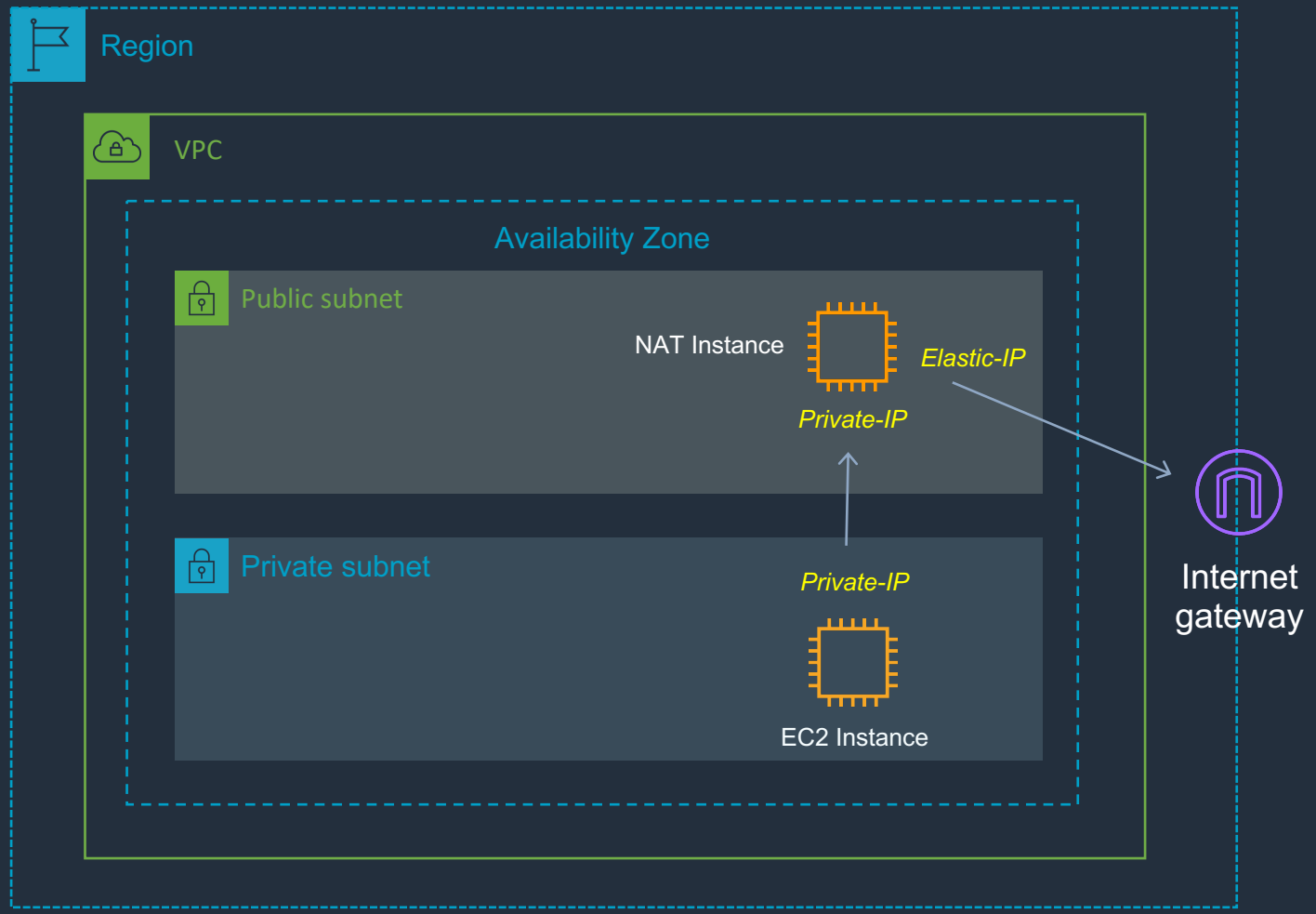
Destination	Target
172.31.0.0/16	Local
0.0.0.0/0	*igw-id*

Private Subnet Route Table

Destination	Target
172.31.0.0/16	Local
0.0.0.0/0	*nat-gateway-id*

DigitalCloud TRAINING

Section 3: Private Subnet with NAT Instance



Public Subnet Route Table

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

Private Subnet Route Table

Destination	Target
172.31.0.0/16	Local
0.0.0.0/0	*nat-instance-id*

Section 4: Amazon S3 Overview

S3 Bucket

`http://bucket.s3.aws-region.amazonaws.com`

`http://s3.aws-region.amazonaws.com/bucket`

Object

- Key
- Version ID
- Value
- Metadata
- Subresources
- Access control information

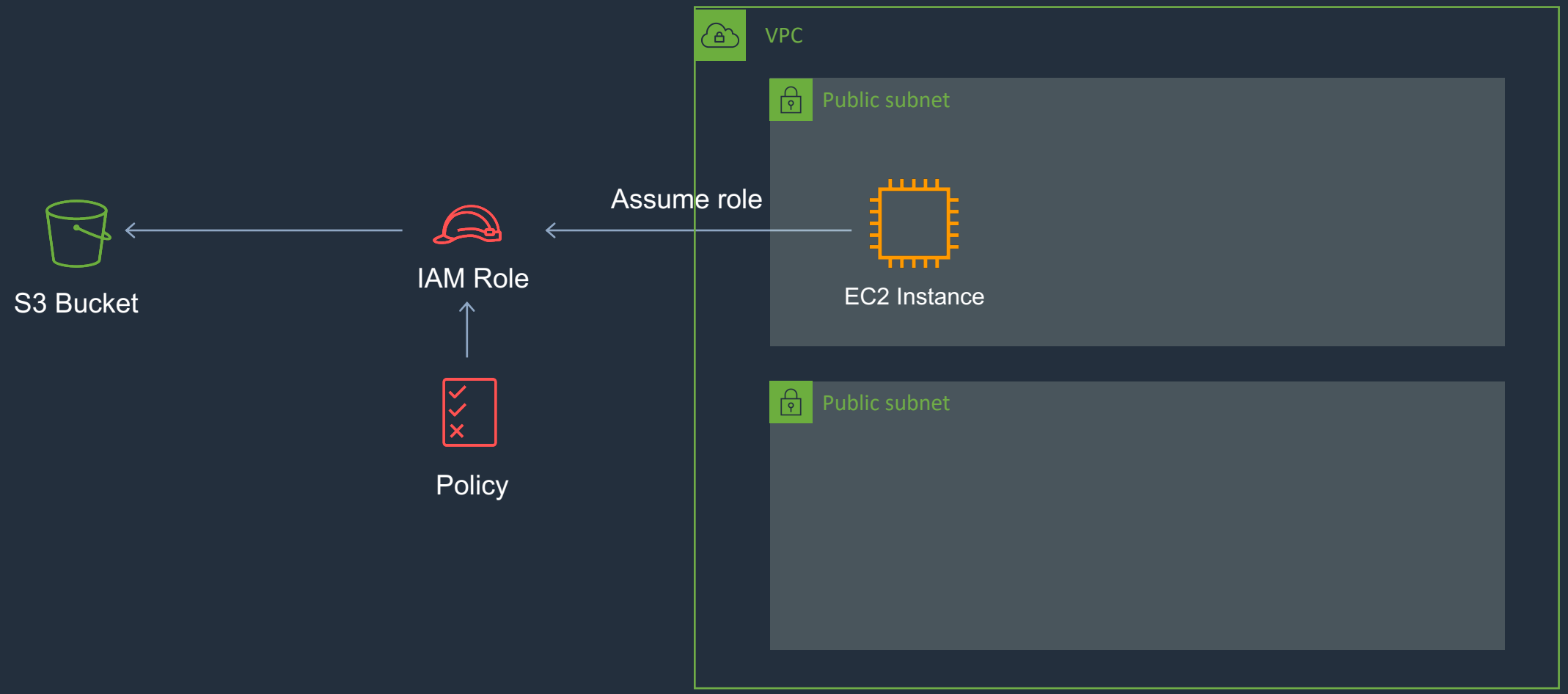
Amazon S3

Public Internet

Internet gateway

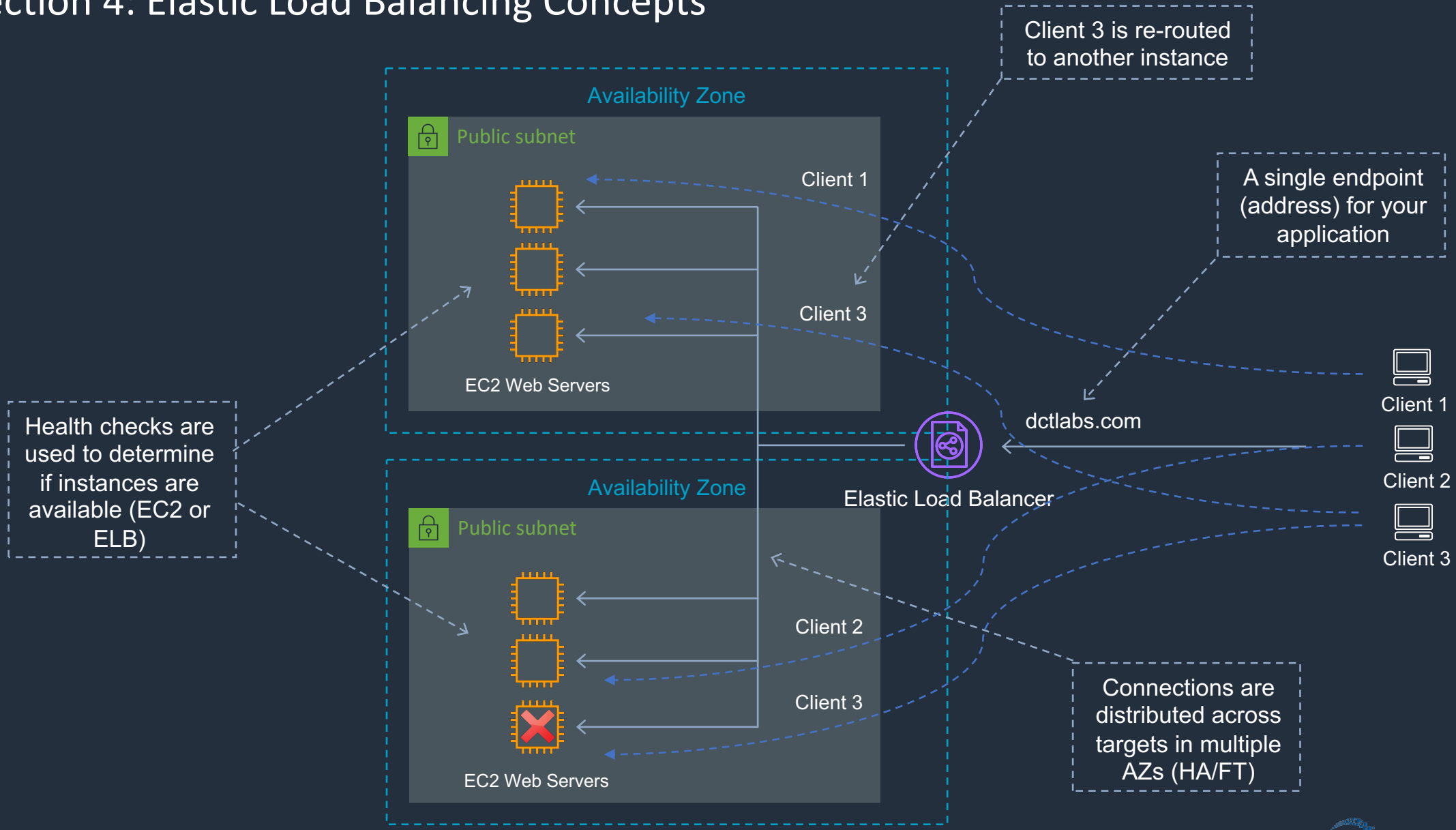
Internet Client

Section 4: IAM Roles



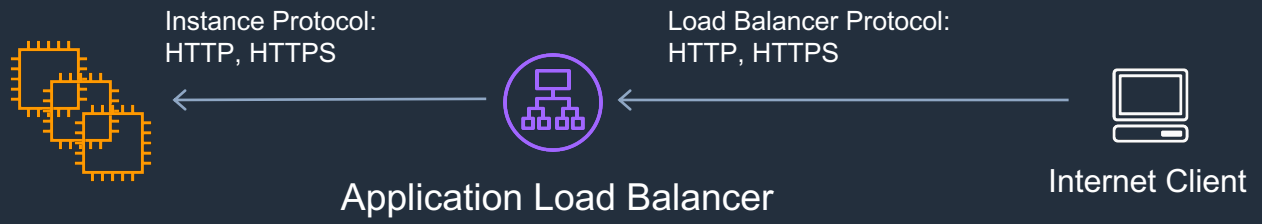
DigitalCloud TRAINING

Section 4: Elastic Load Balancing Concepts



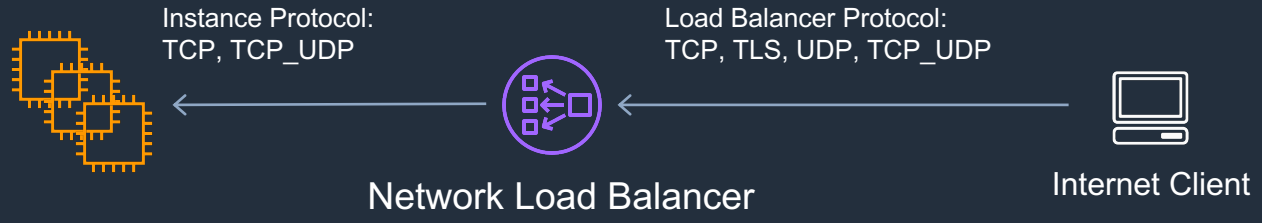
DigitalCloud TRAINING

Section 4: Elastic Load Balancing (ELB) Types



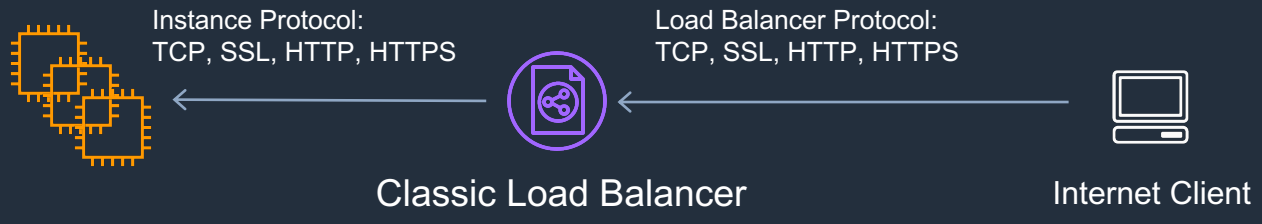
Application Load Balancer

- Operates at the request level
- Routes based on the content of the request (layer 7)
- Supports path-based routing, host-based routing, query string parameter-based routing, and source IP address-based routing
- Supports IP addresses, Lambda Functions and containers as targets



Network Load Balancer

- Operates at the connection level
- Routes connections based on IP protocol data (layer 4)
- Offers ultra high performance, low latency and TLS offloading at scale
- Can have static IP / Elastic IP
- Supports UDP and static IP addresses as targets

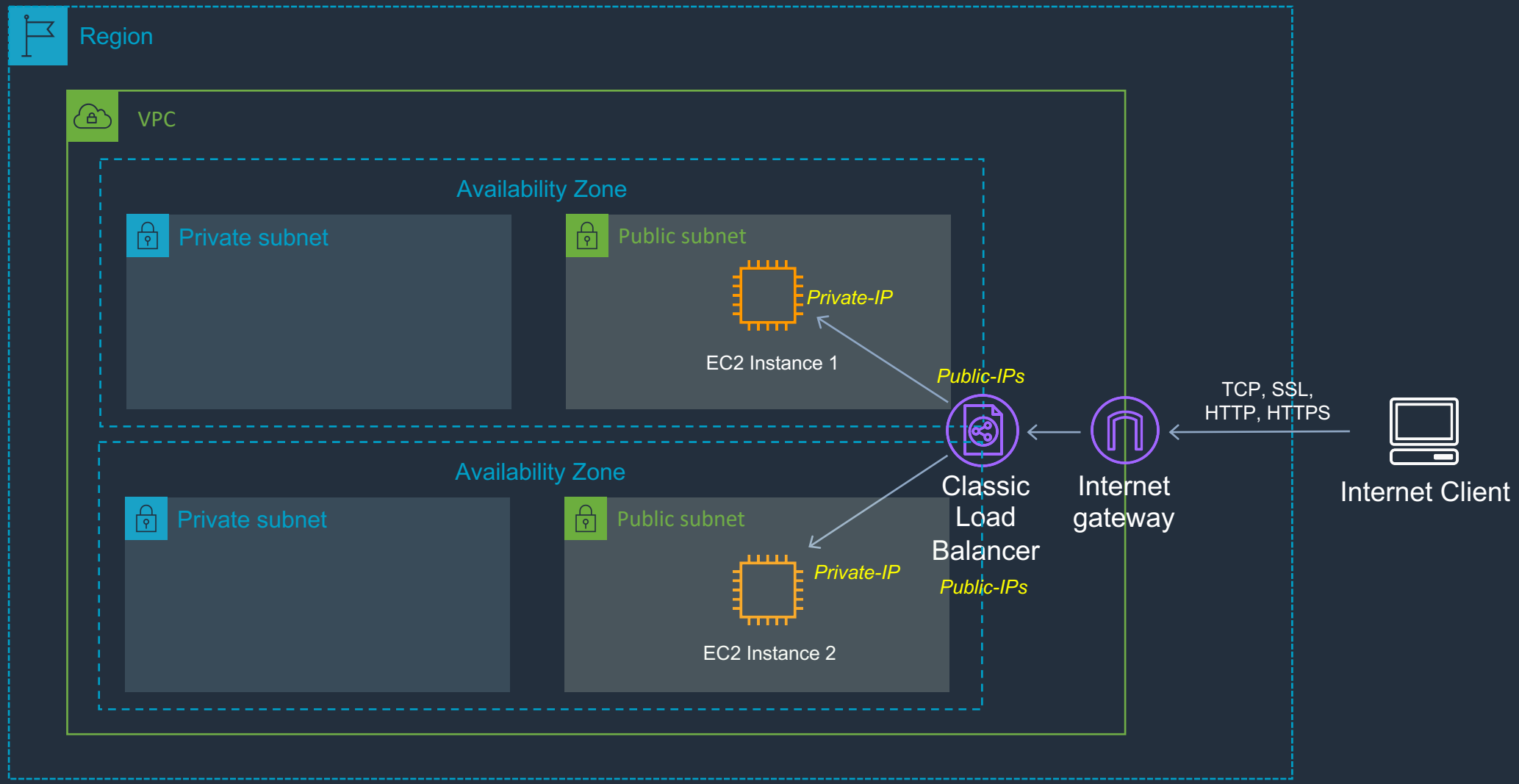


Classic Load Balancer

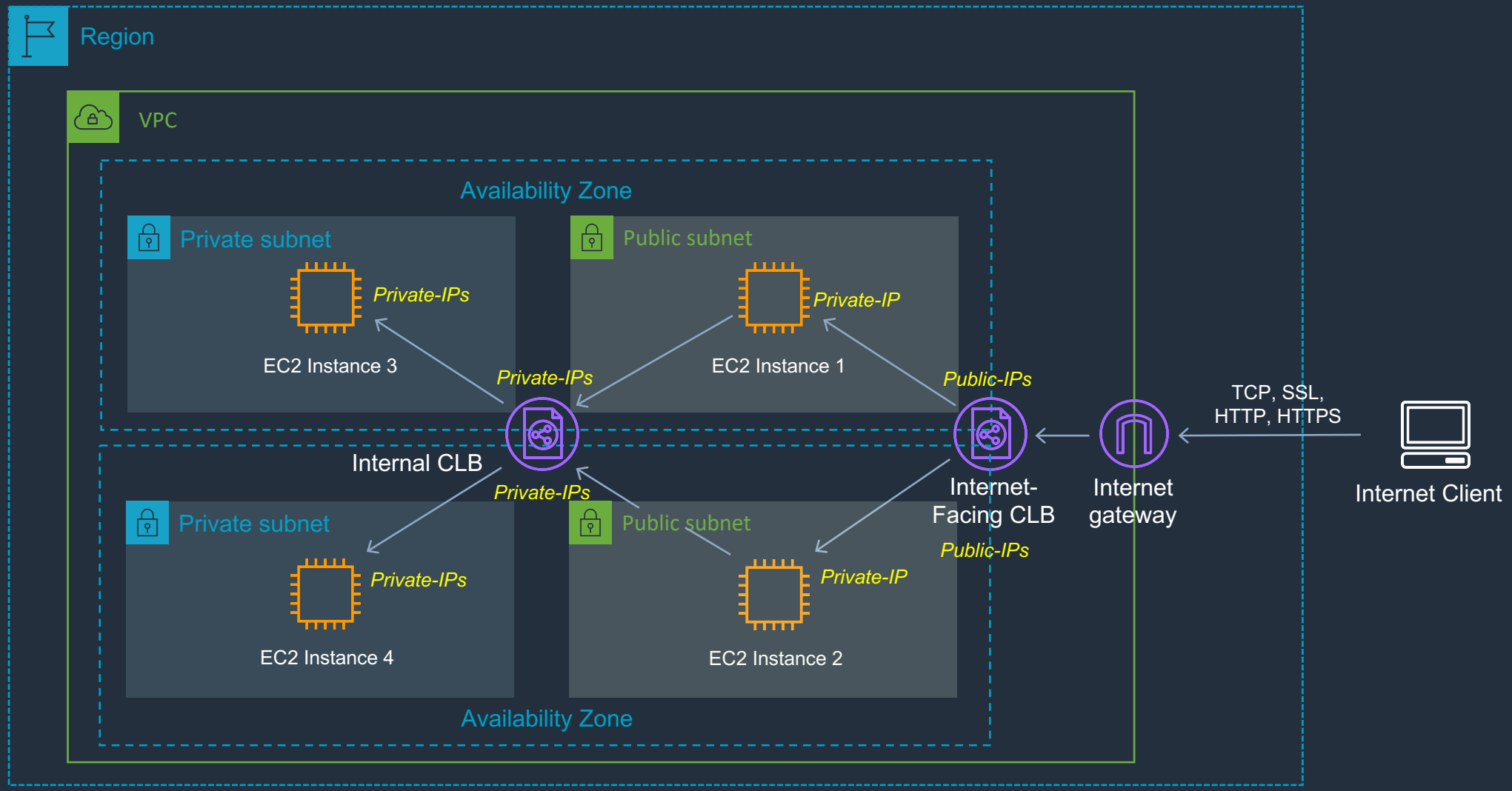
- Old generation; not recommended for new applications
- Performs routing at Layer 4 and Layer 7
- Use for existing applications running in EC2-Classic

DigitalCloud TRAINING

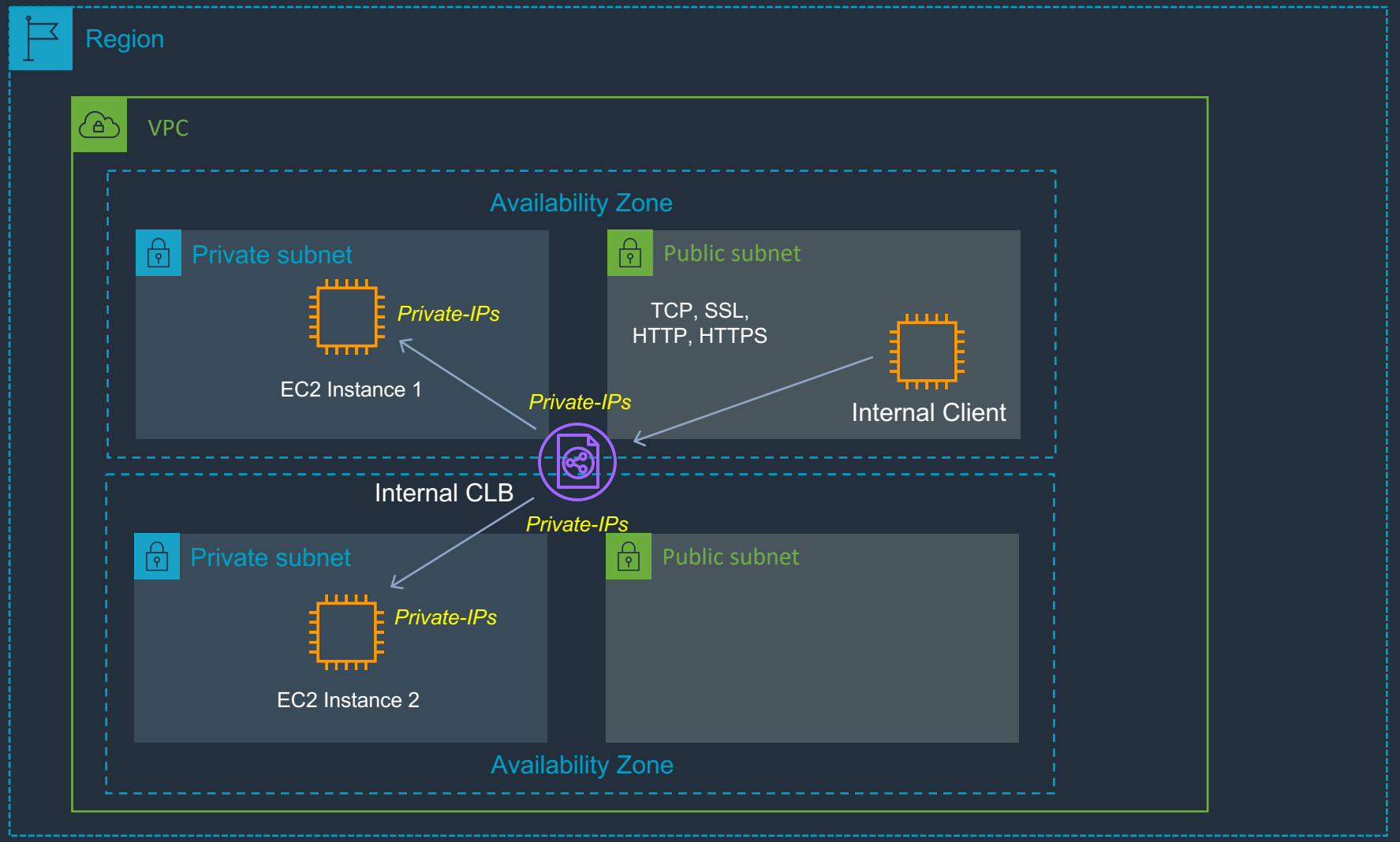
Section 4: Classic Load Balancer (Internet-Facing)



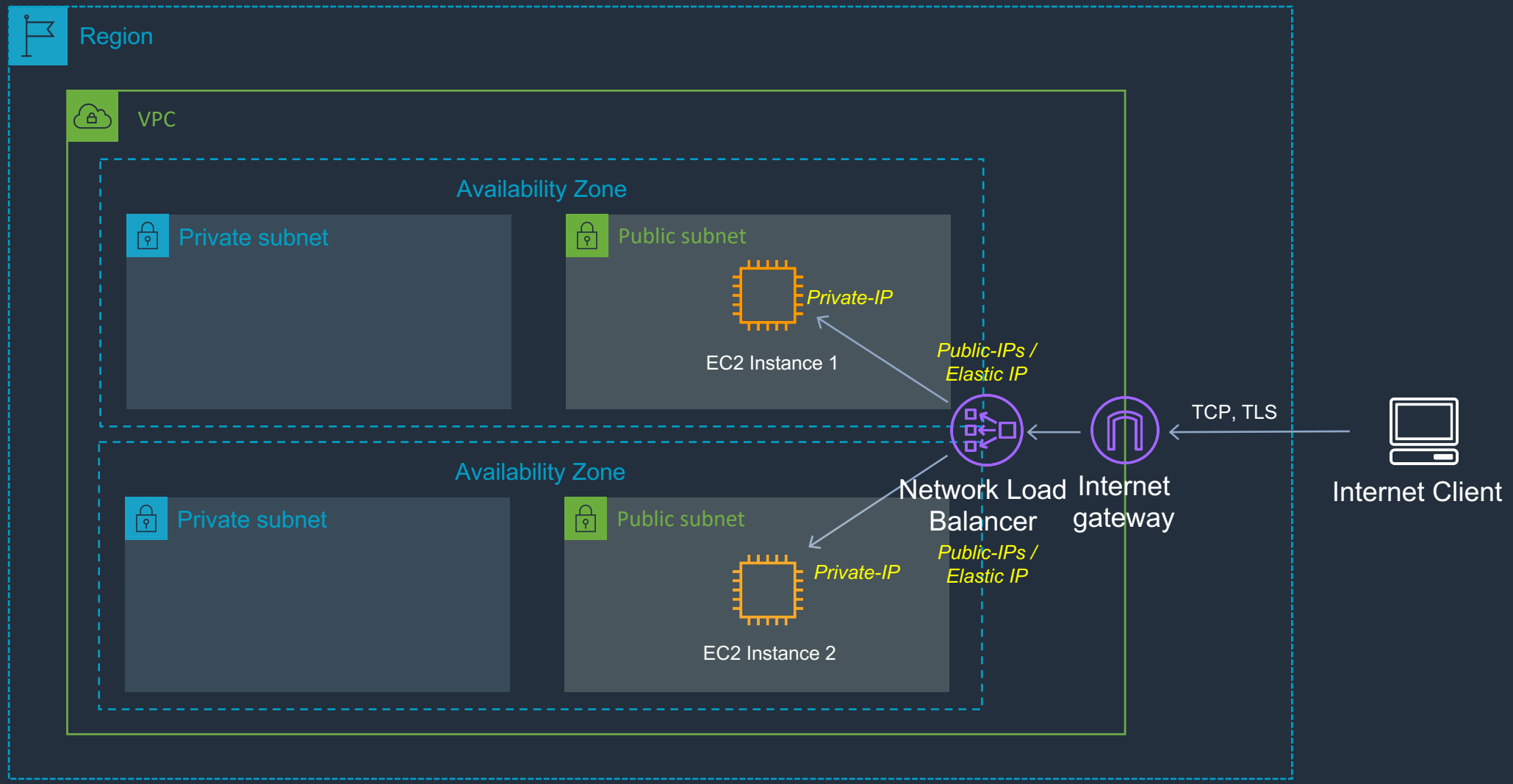
Section 4: Classic Load Balancer - Multi-tier



Section 4: Classic Load Balancer (Internal)

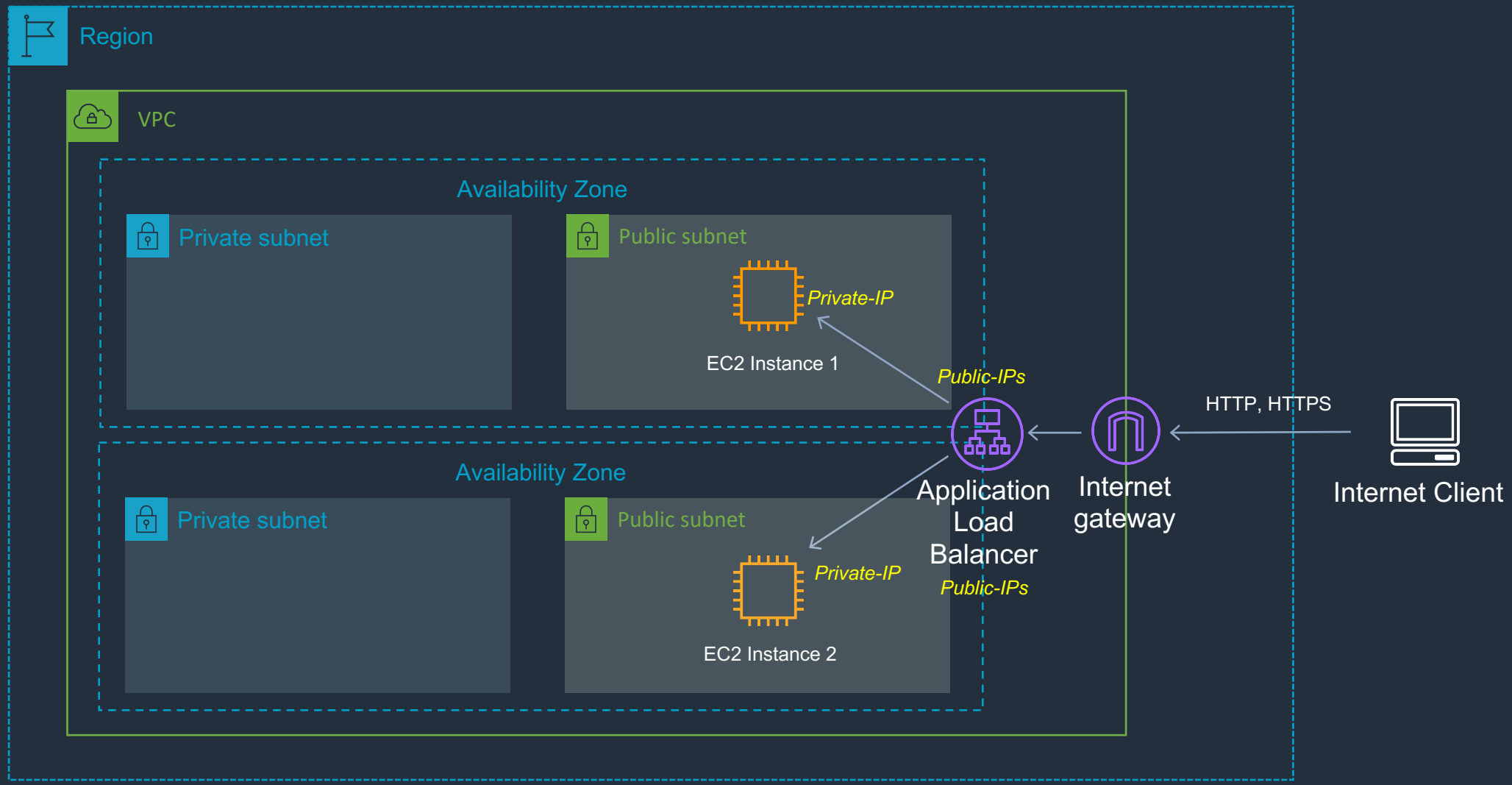


Section 4: Network Load Balancer (Internet-Facing)



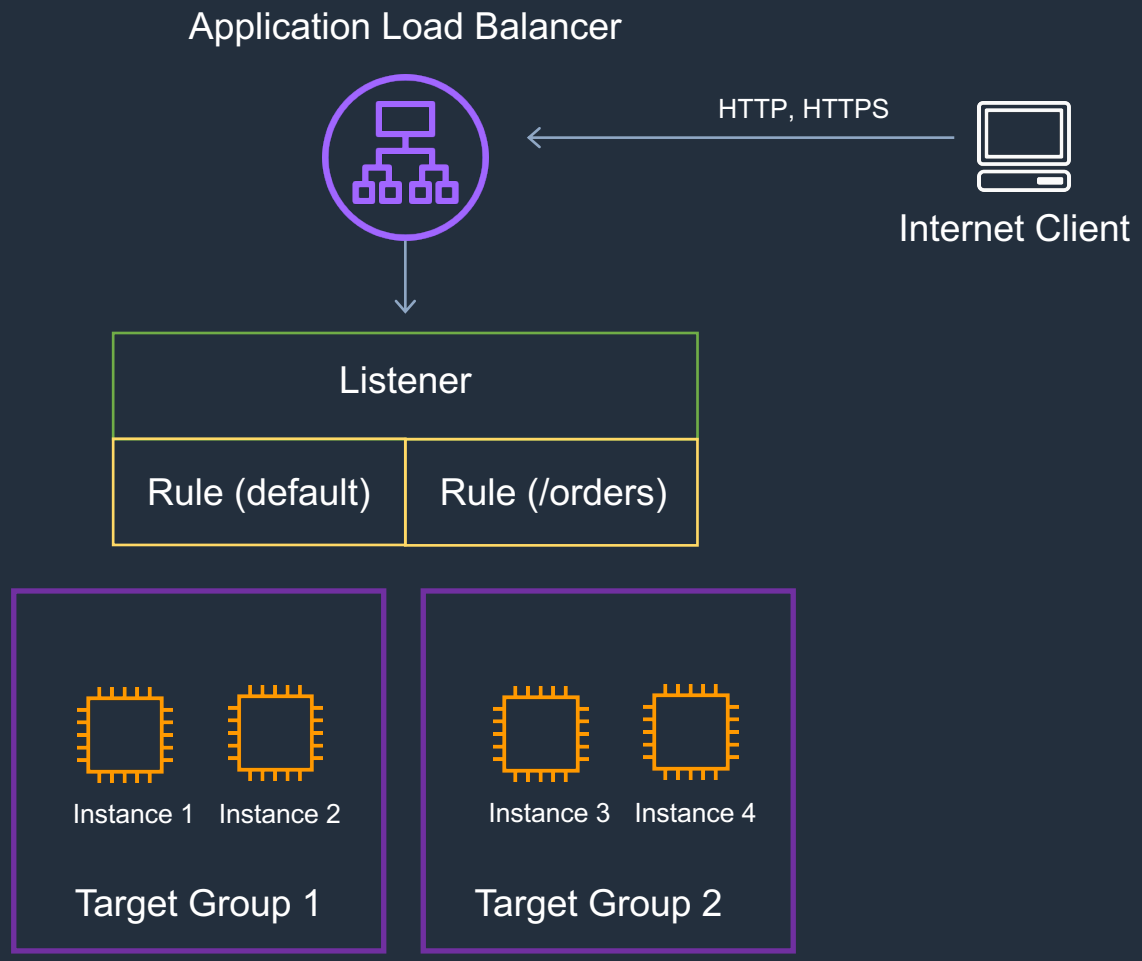
DigitalCloud TRAINING

Section 4: Application Load Balancer (Internet-Facing)

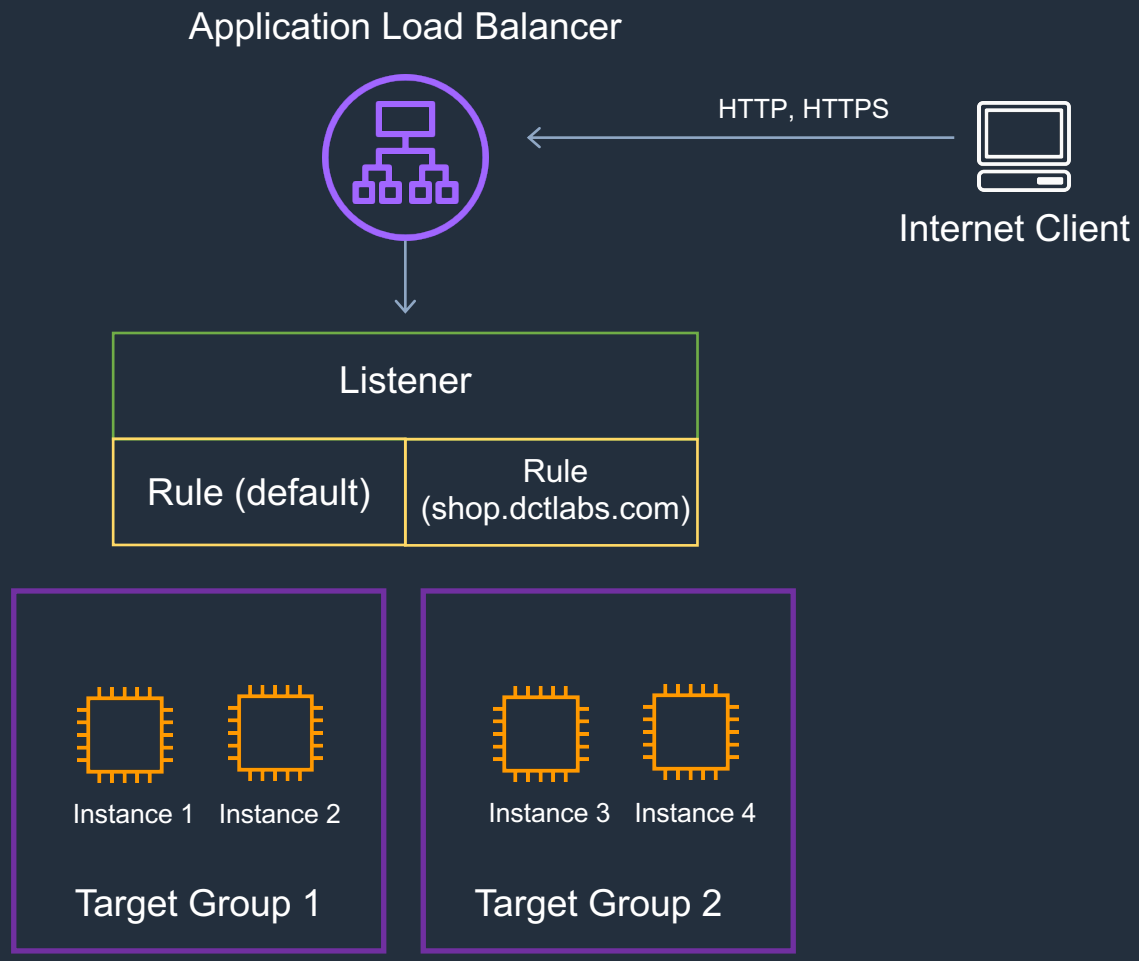


DigitalCloud TRAINING

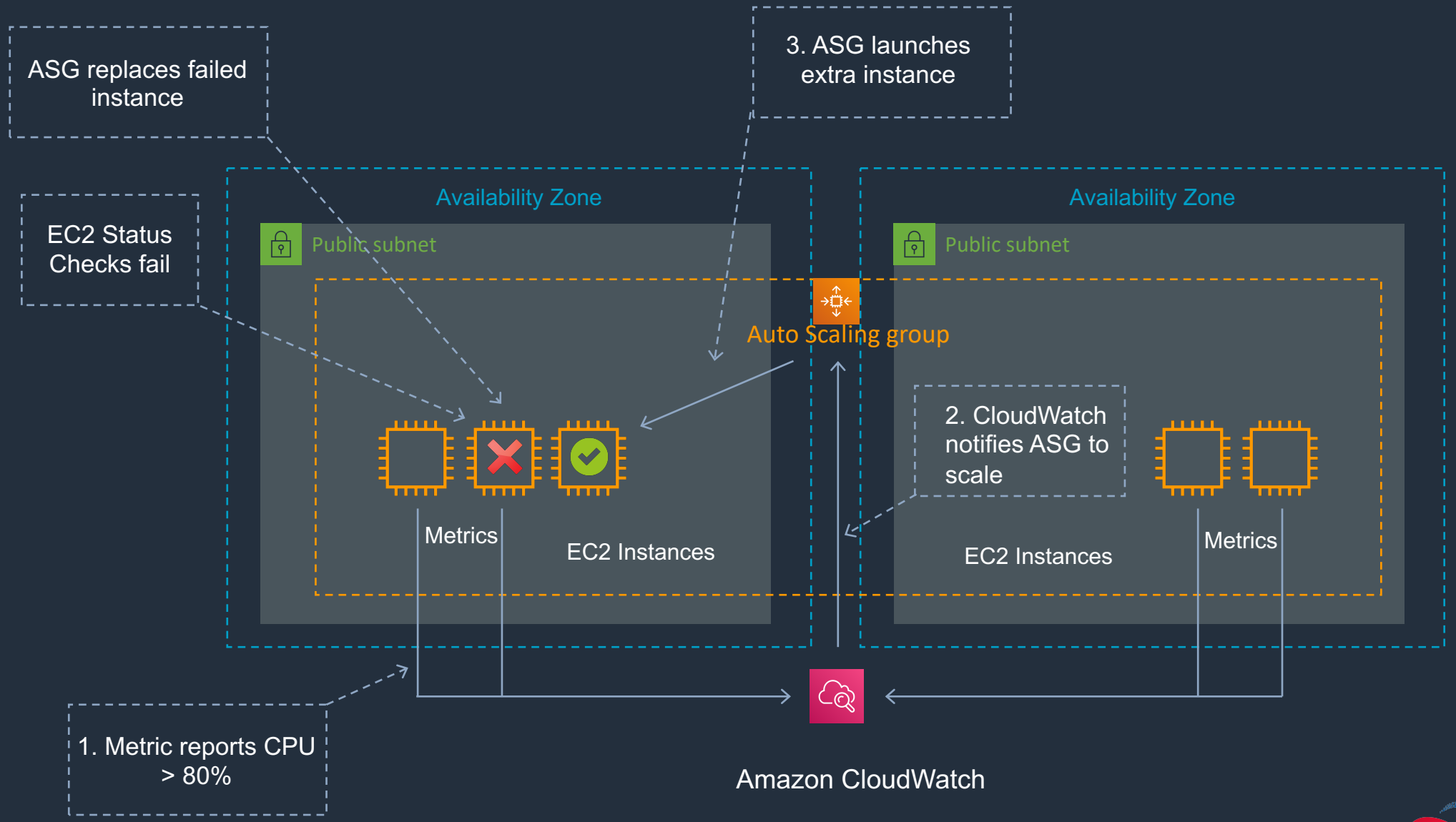
Section 4: Application Load Balancer – Path-based Routing



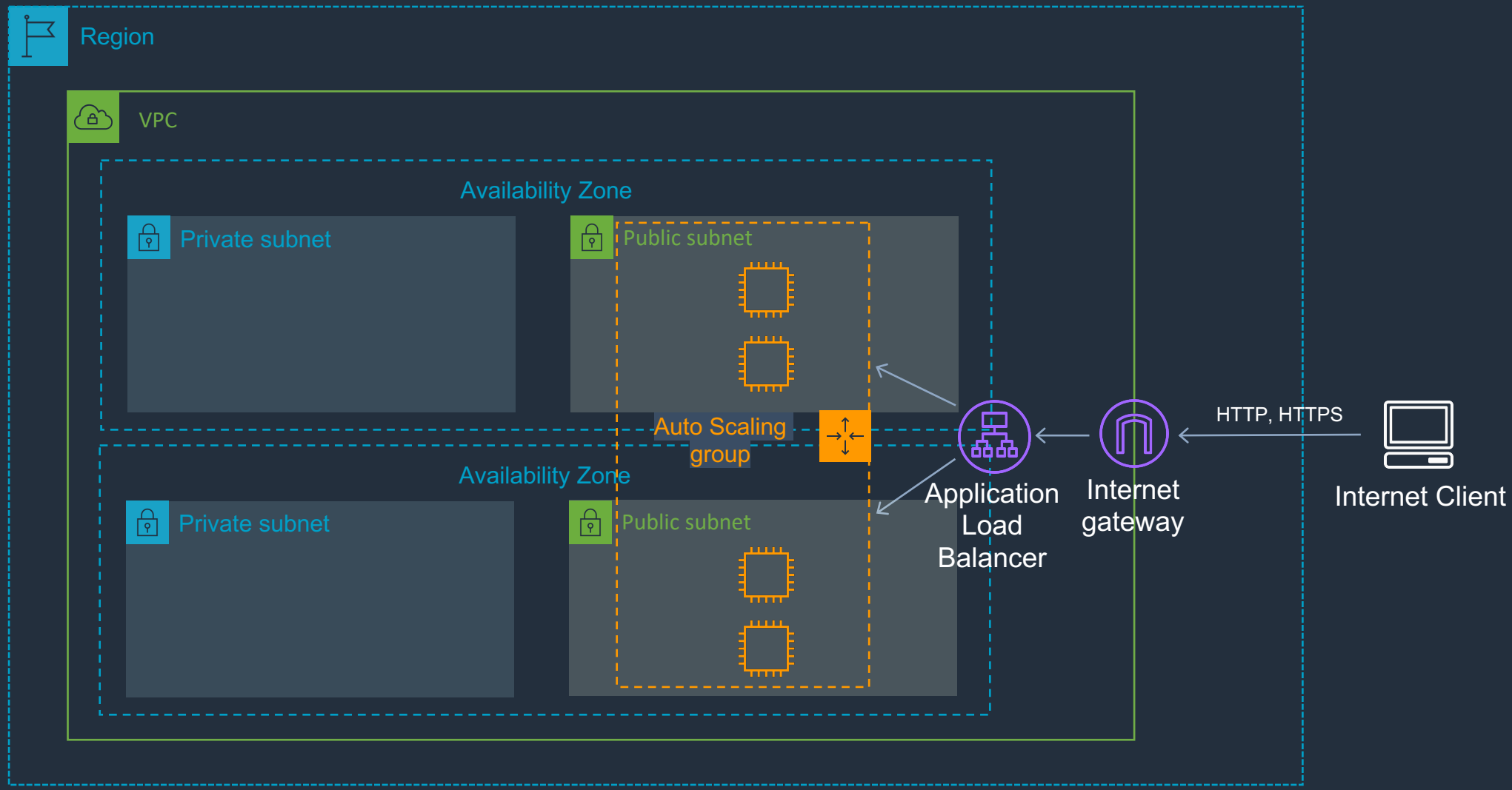
Section 4: Application Load Balancer – Host-based Routing



Section 4: Auto Scaling Overview

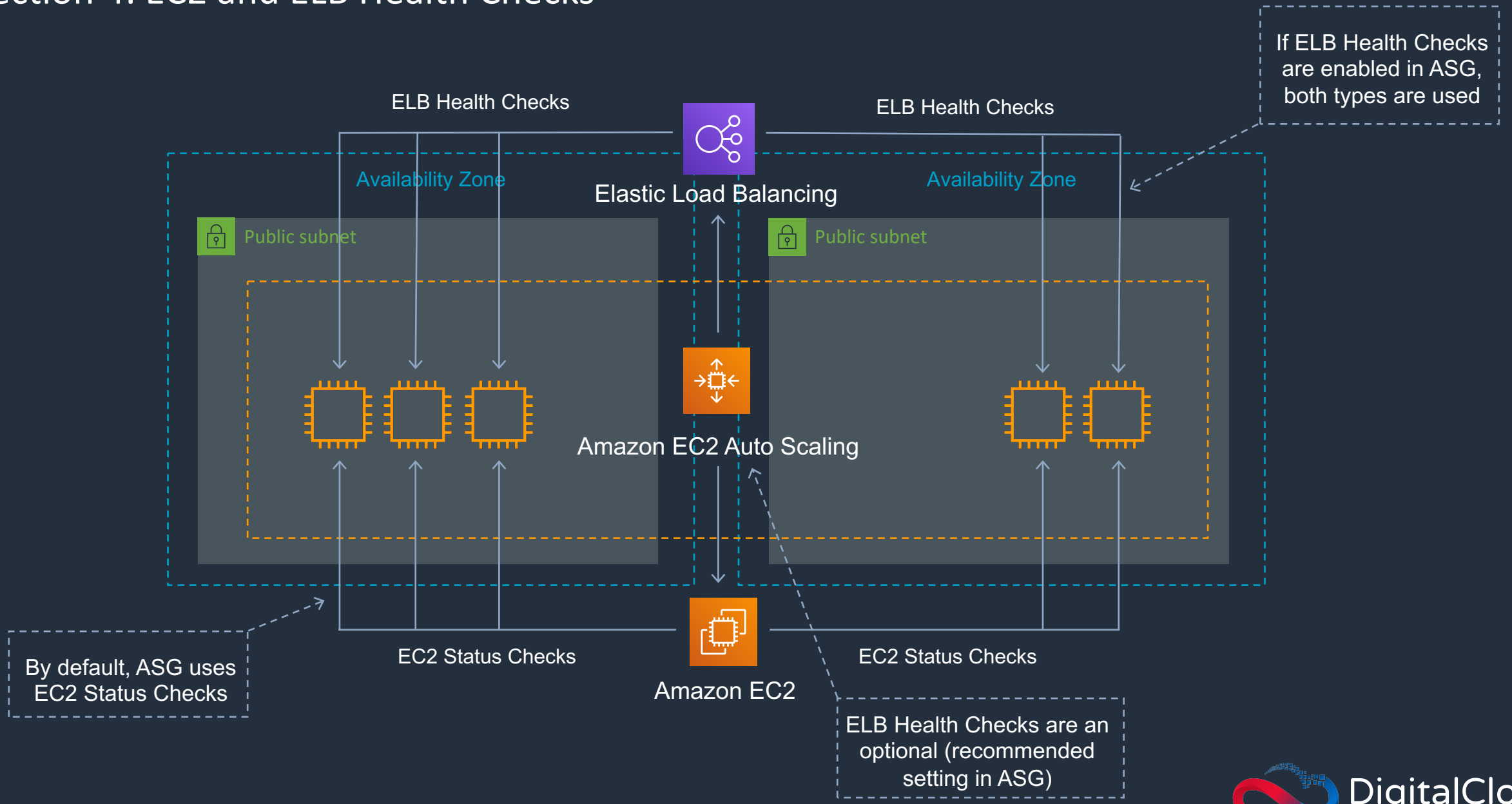


Section 4: Auto Scaling Group with ALB



DigitalCloud TRAINING

Section 4: EC2 and ELB Health Checks



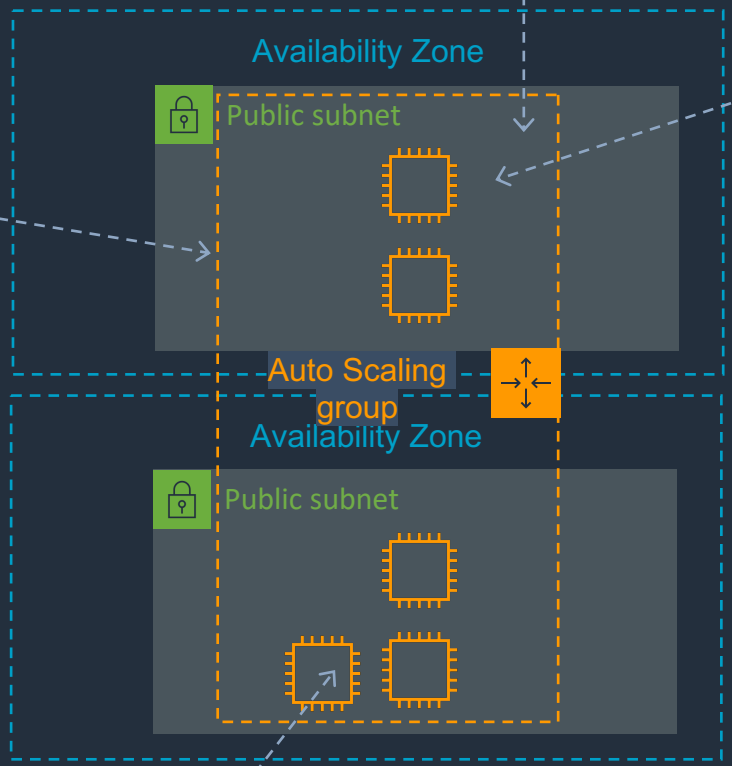
DigitalCloud TRAINING

Section 4: Auto Scaling Termination Policies – Default Policies

4. Determine whether any of the instances use the oldest launch configuration

2. Determine which instance to terminate so as to align the remaining instances to the allocation strategy for the On-Demand or Spot Instance that is terminating and your current selection of instance types

5. After applying all of the criteria in 2 through 4, if there are multiple unprotected instances to terminate, determine which instances are closest to the next billing hour

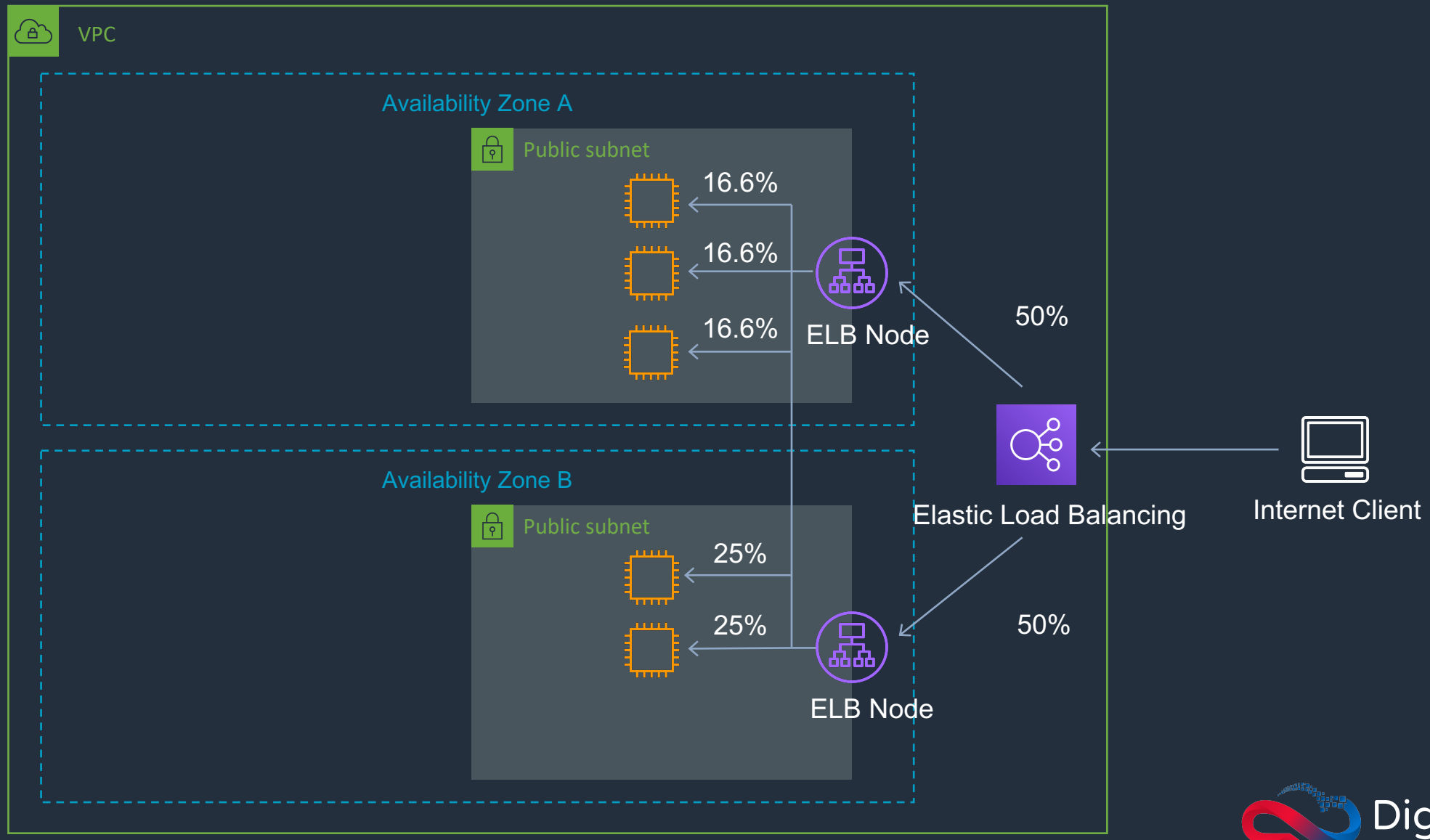


1. Determine which AZ has the most instances

3. Determine whether any of the instances use the oldest launch template

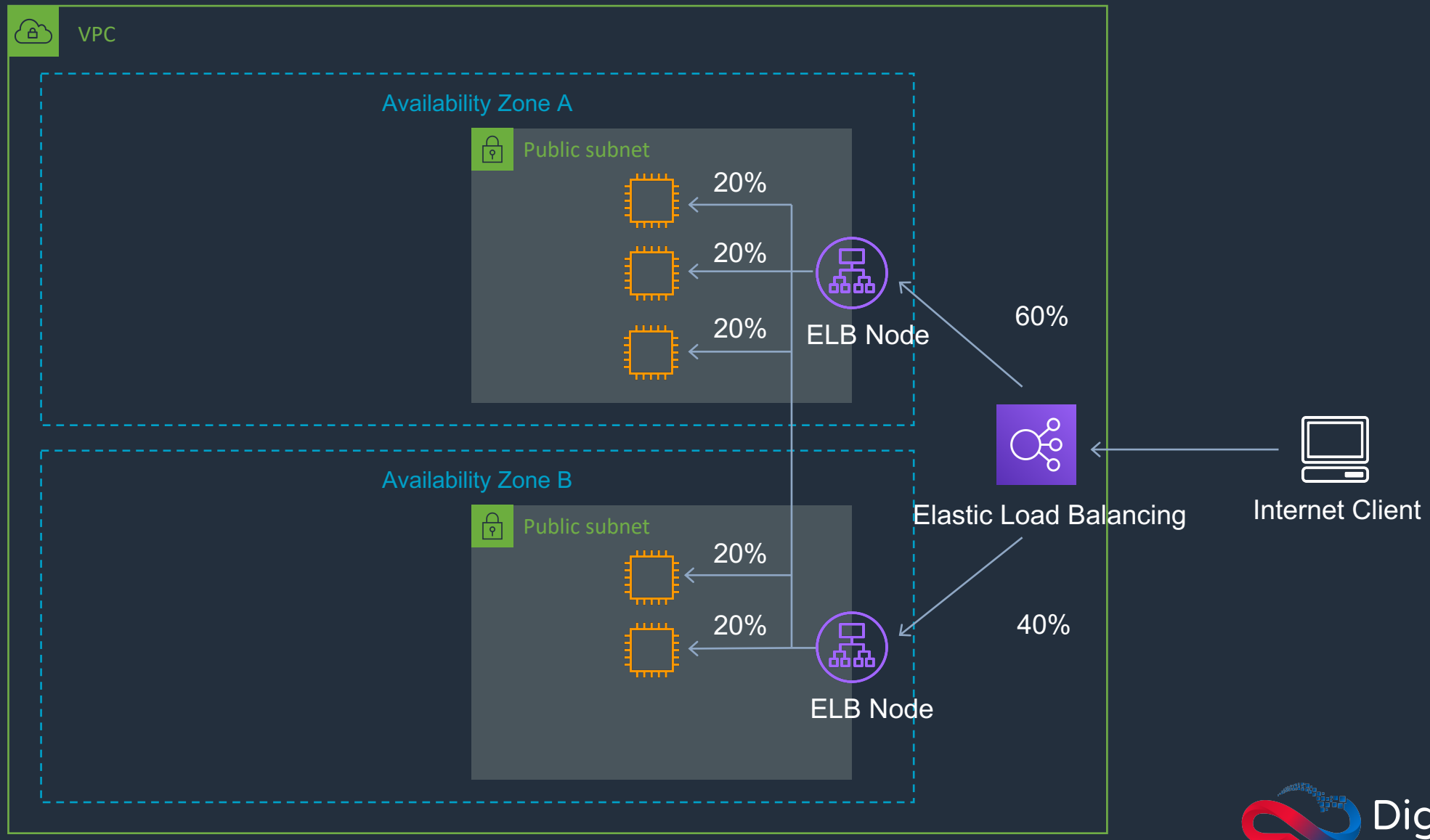
DigitalCloud TRAINING

Section 4: Cross-Zone Load Balancing - Disabled



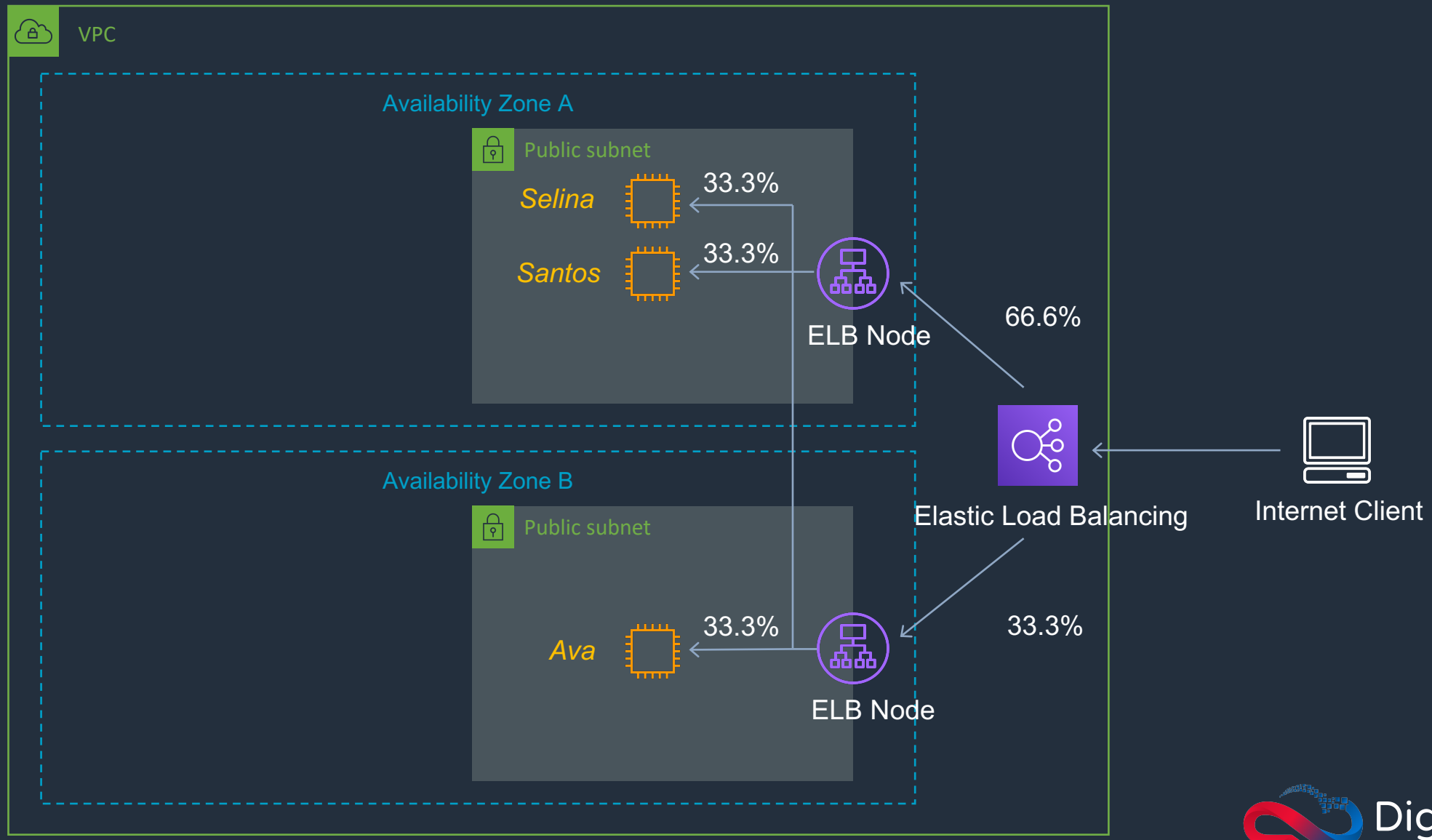
DigitalCloud TRAINING

Section 4: Cross-Zone Load Balancing - Enabled



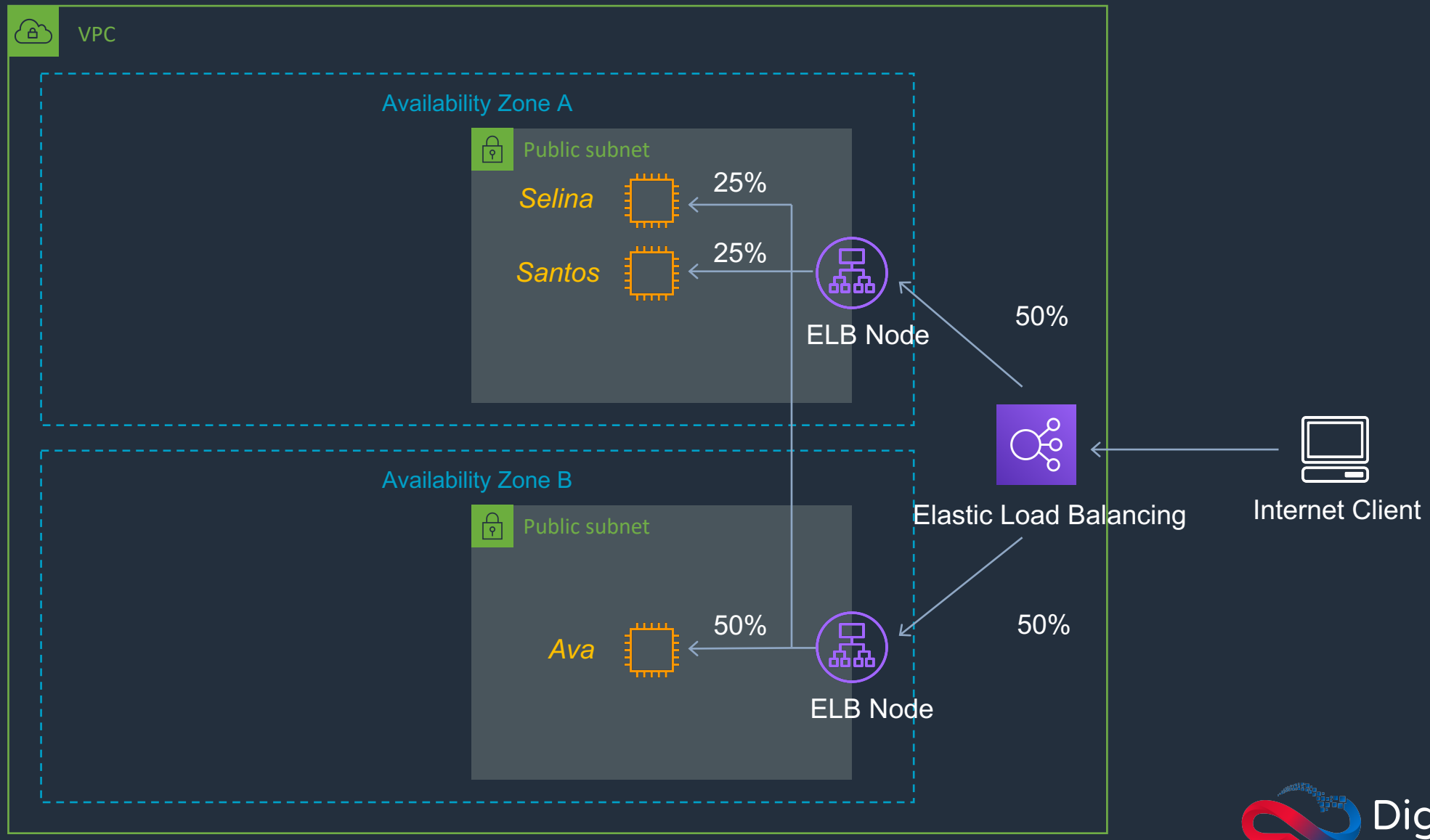
DigitalCloud TRAINING

Section 4: Cross-Zone Load Balancing - Enabled



DigitalCloud TRAINING

Section 4: Cross-Zone Load Balancing - Disabled

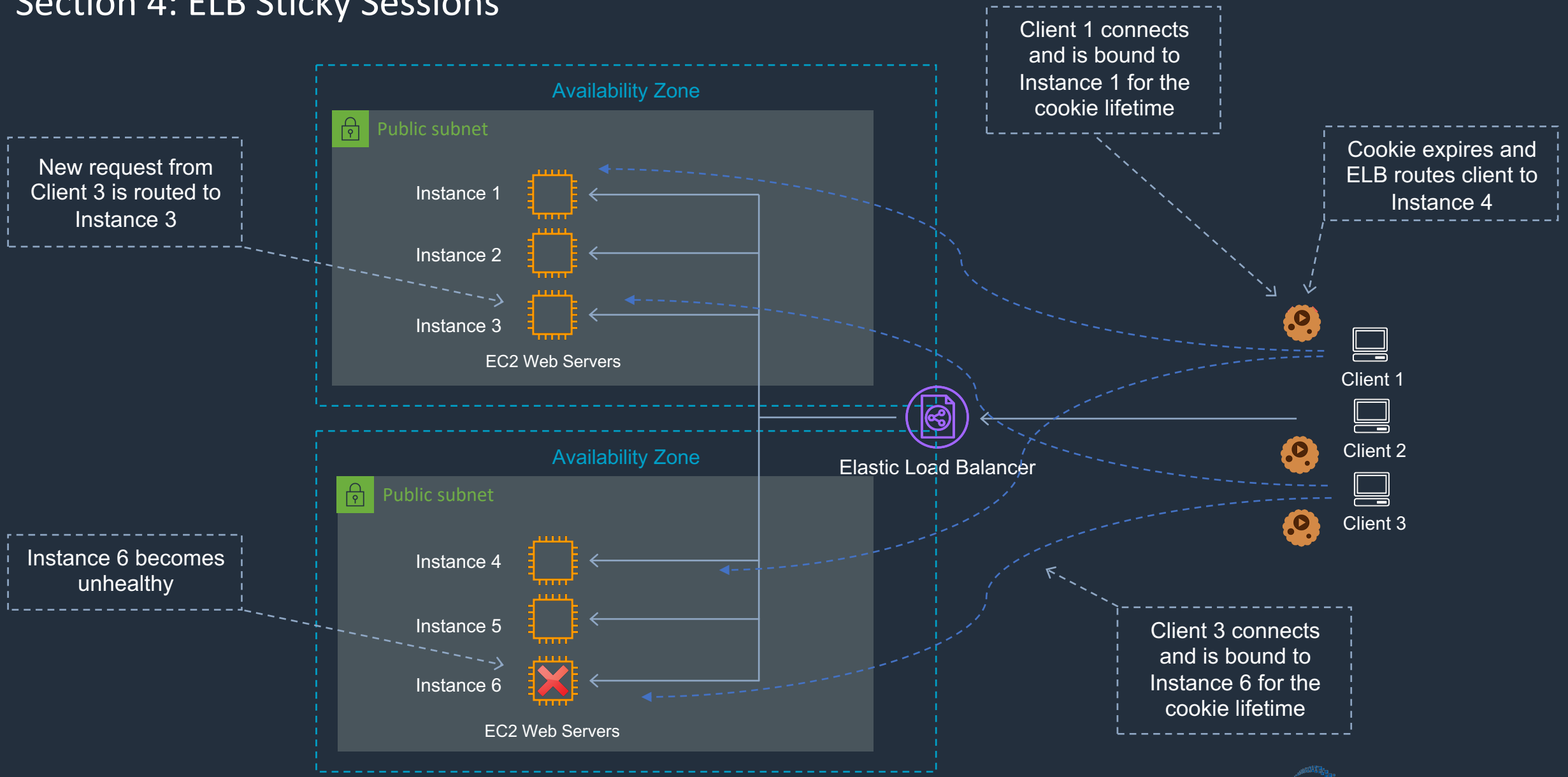


DigitalCloud TRAINING

Section 4: Cross-Zone Load Balancing

Name	Created through Console	Created through CLI/API	Can be enabled/disabled?
ALB	Enabled	Enabled	No
NLB	Disabled	Disabled	Yes
CLB	Enabled	Disabled	Yes

Section 4: ELB Sticky Sessions

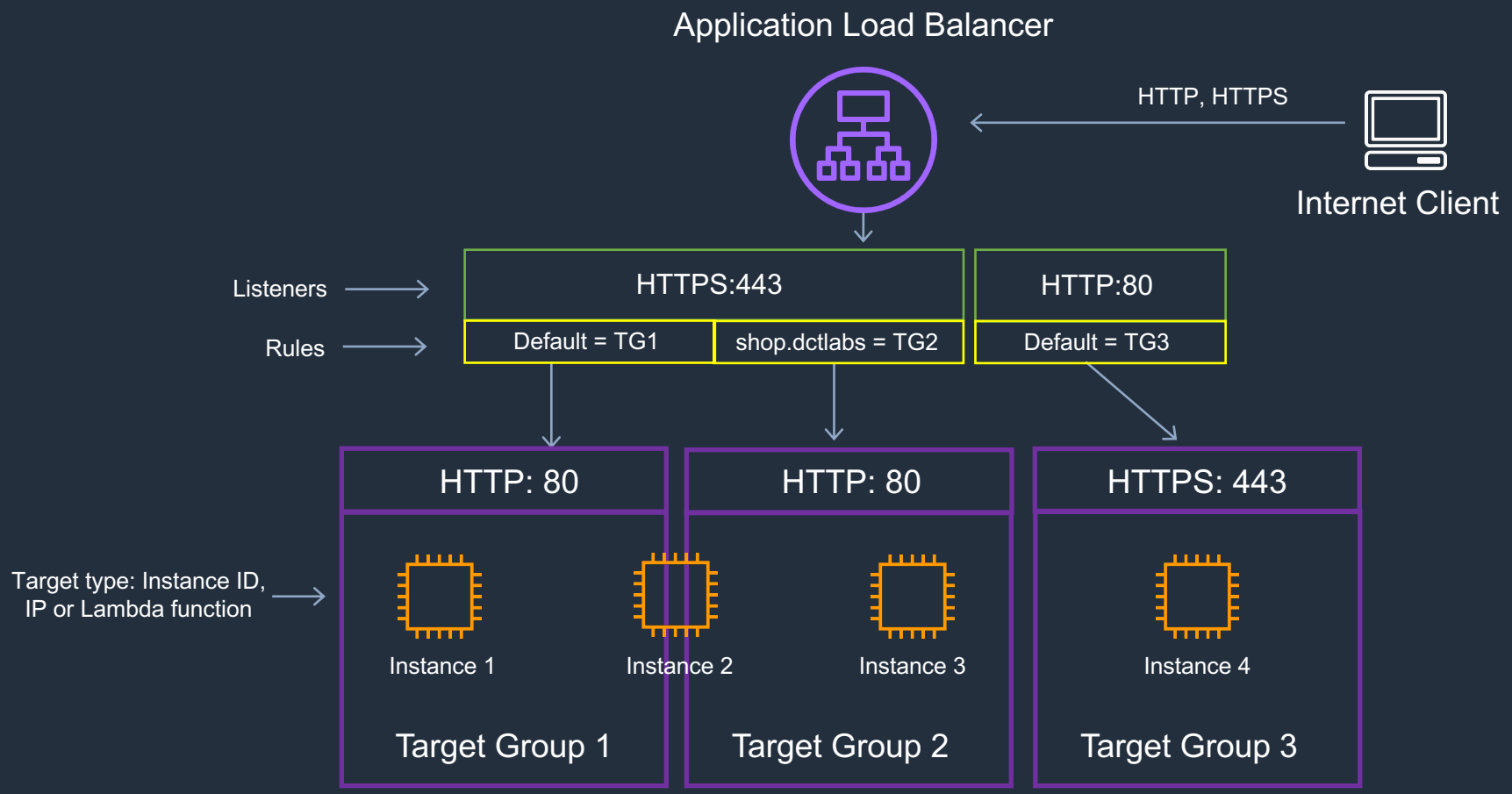


DigitalCloud TRAINING

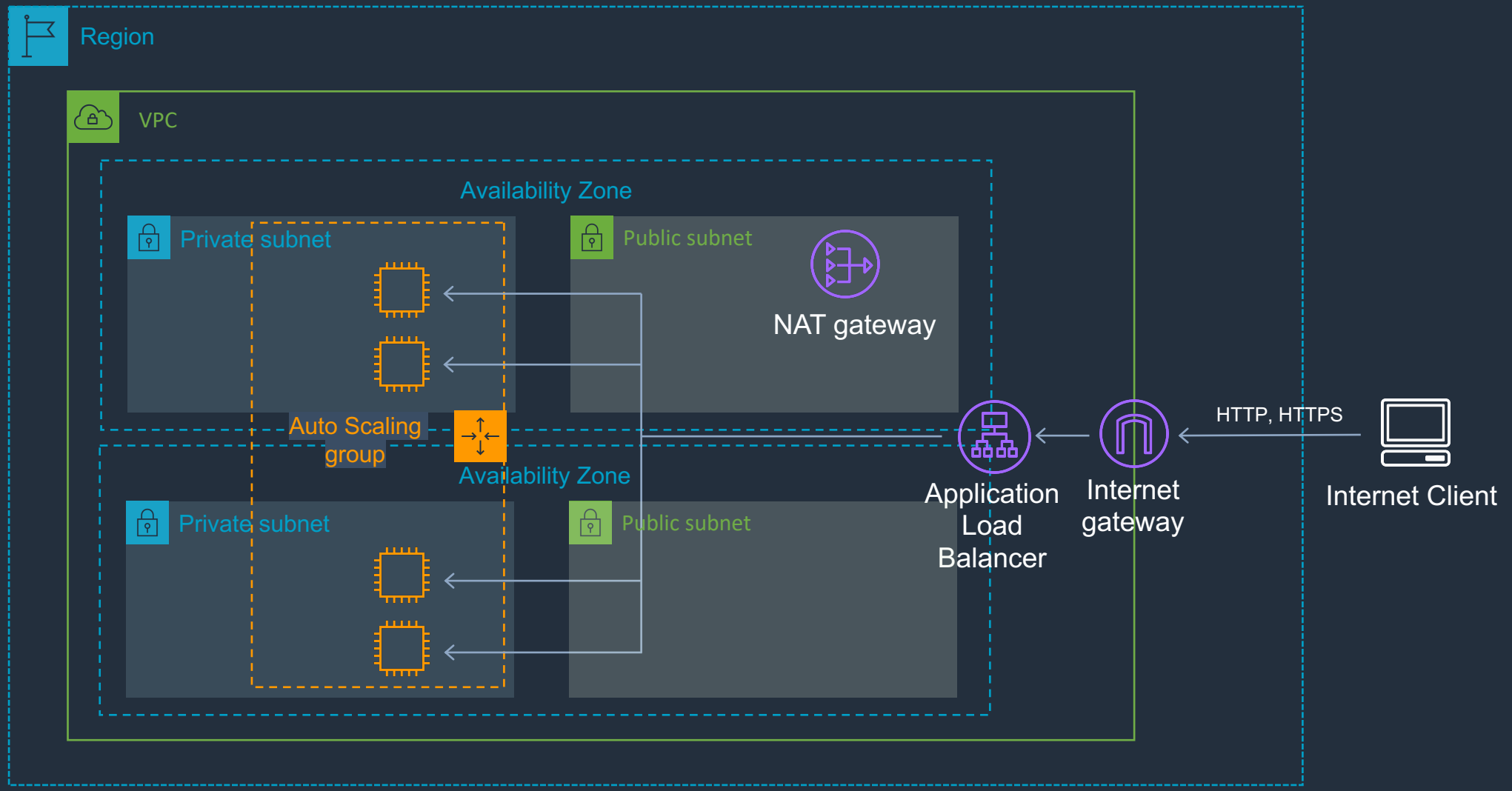
Section 4: Sticky Sessions

Name	Supported?	Load Balancer Generated Cookie	Application Generated Cookie
ALB	Yes	Yes, “AWSALB”	Not supported
NLB	No	N/A	N/A
CLB	Yes	Yes	Yes

Section 4: Application Load Balancer – Listeners and SSL/TLS

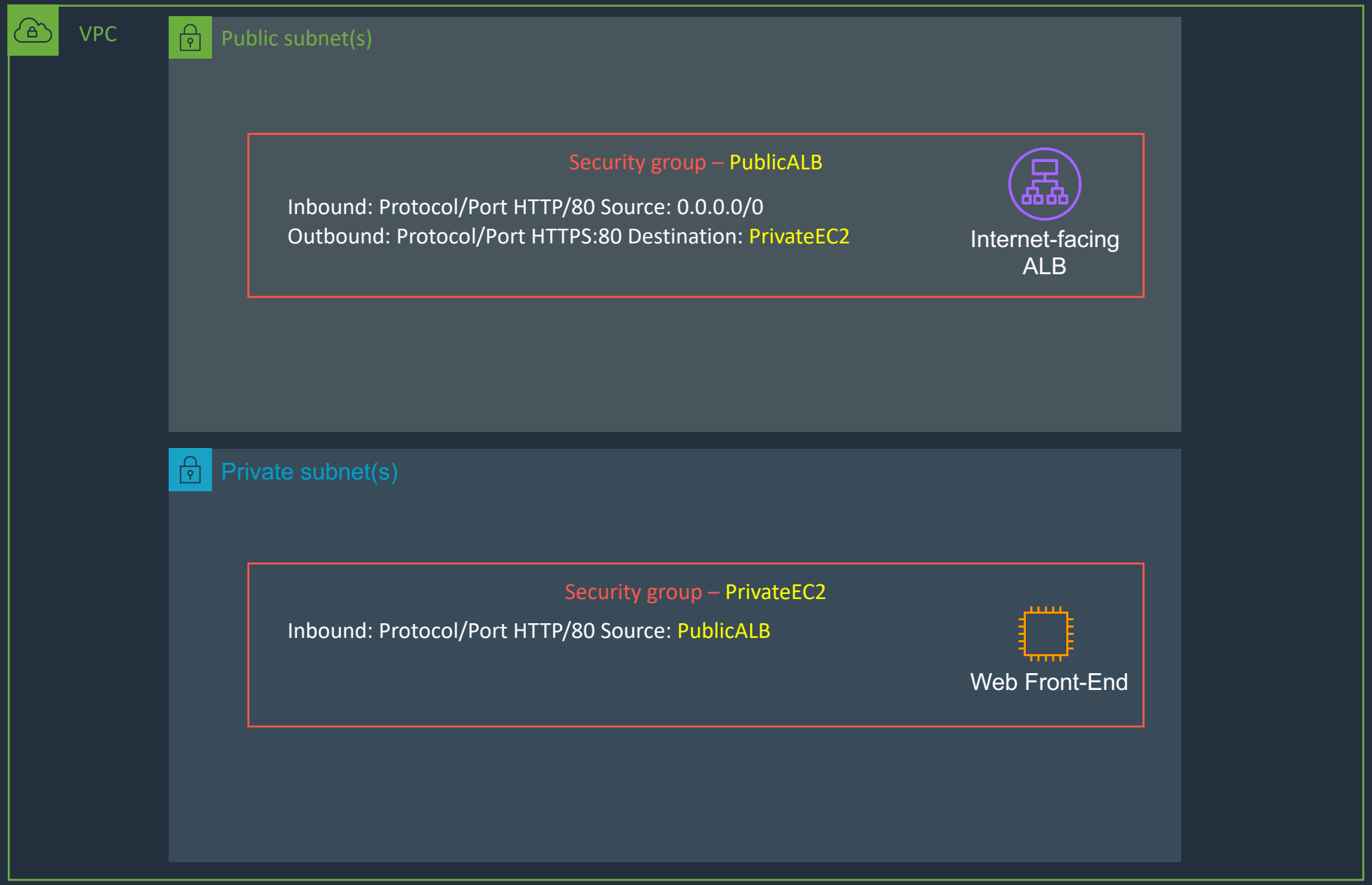


Section 4: Public ALB with Private Instances



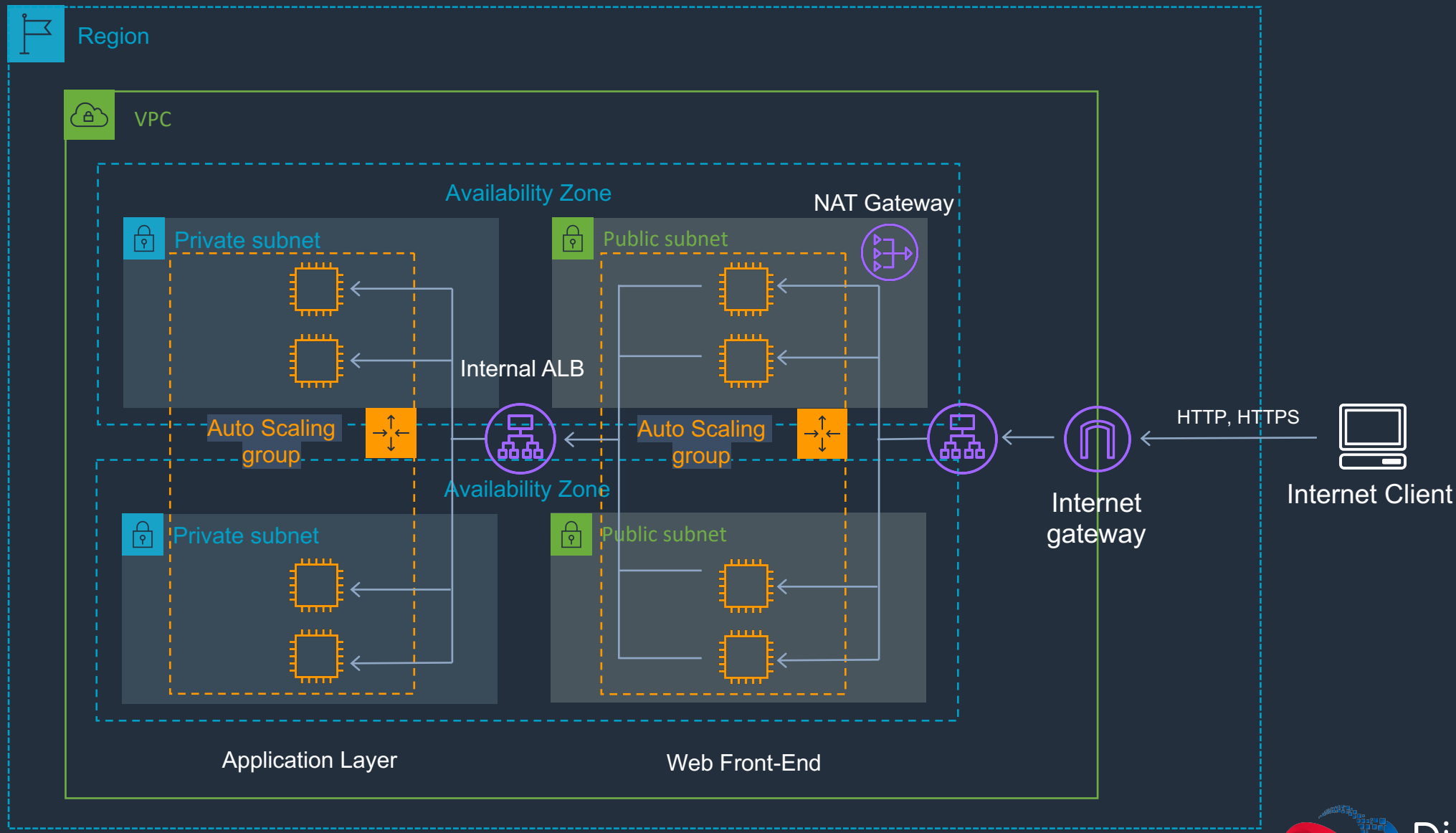
DigitalCloud TRAINING

Section 4: Public ALB with Private Instances– Security Groups



DigitalCloud TRAINING

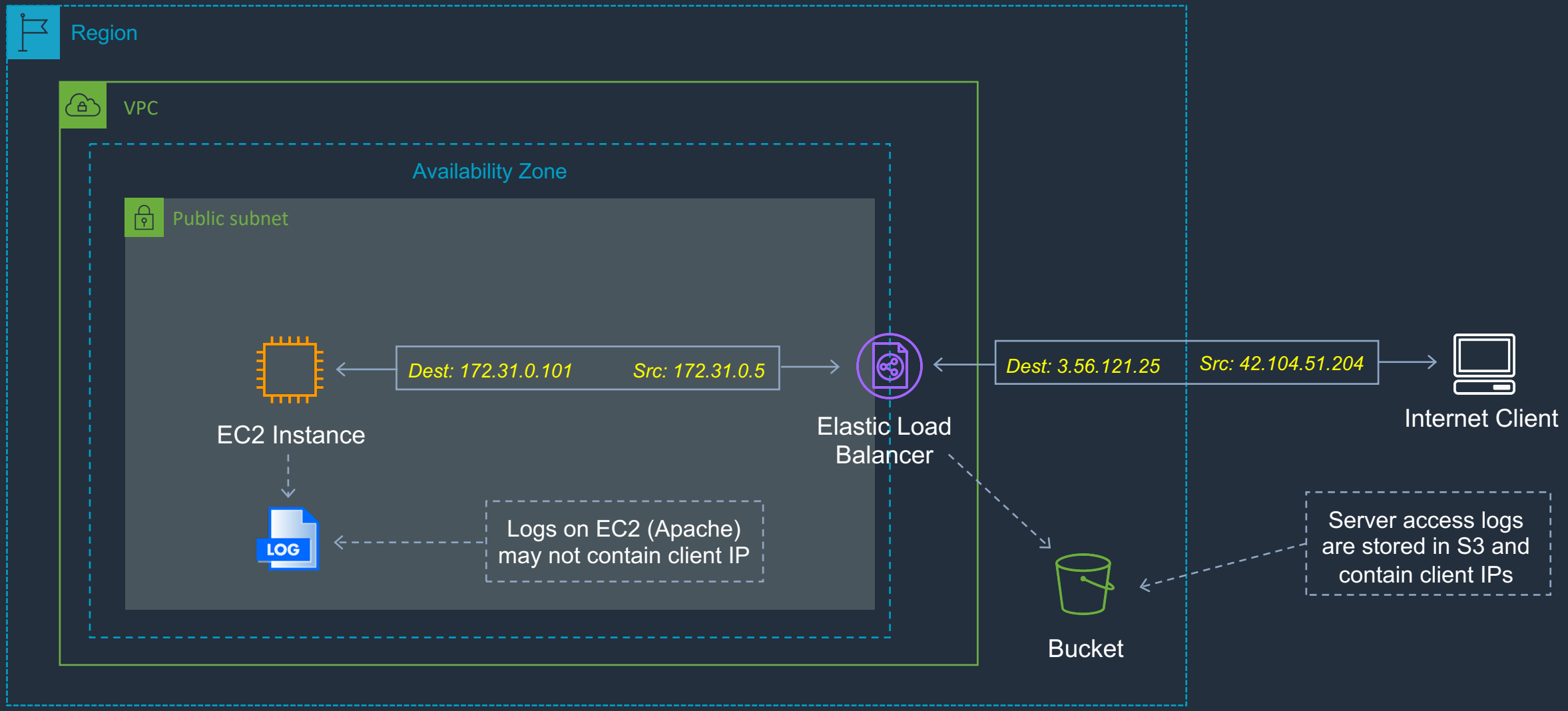
Section 4: Multi-Tier Web Architecture



DigitalCloud TRAINING

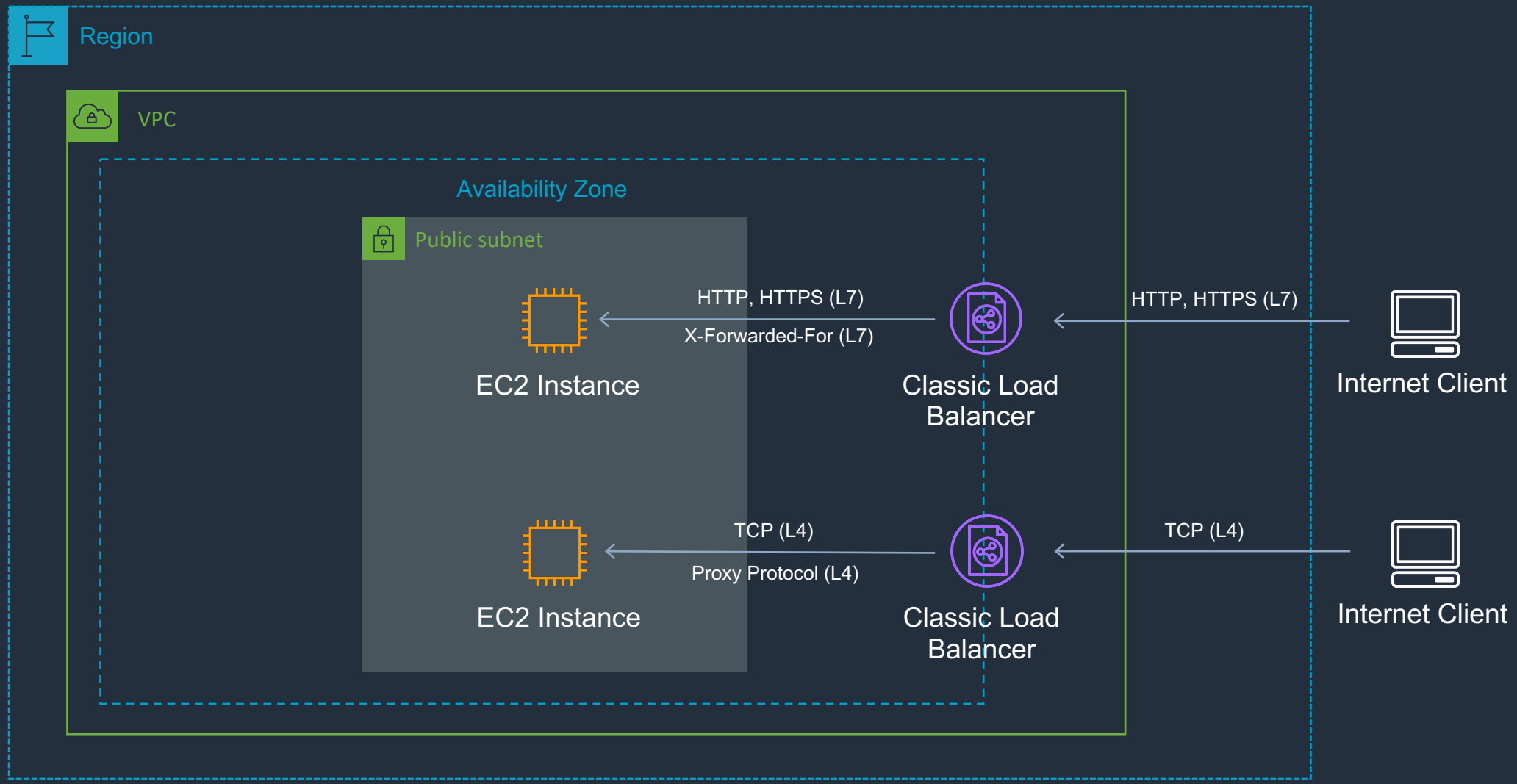
Section 4: Multi-Tier Web Architecture – Security Groups

Section 4: ELB Connections and Logging

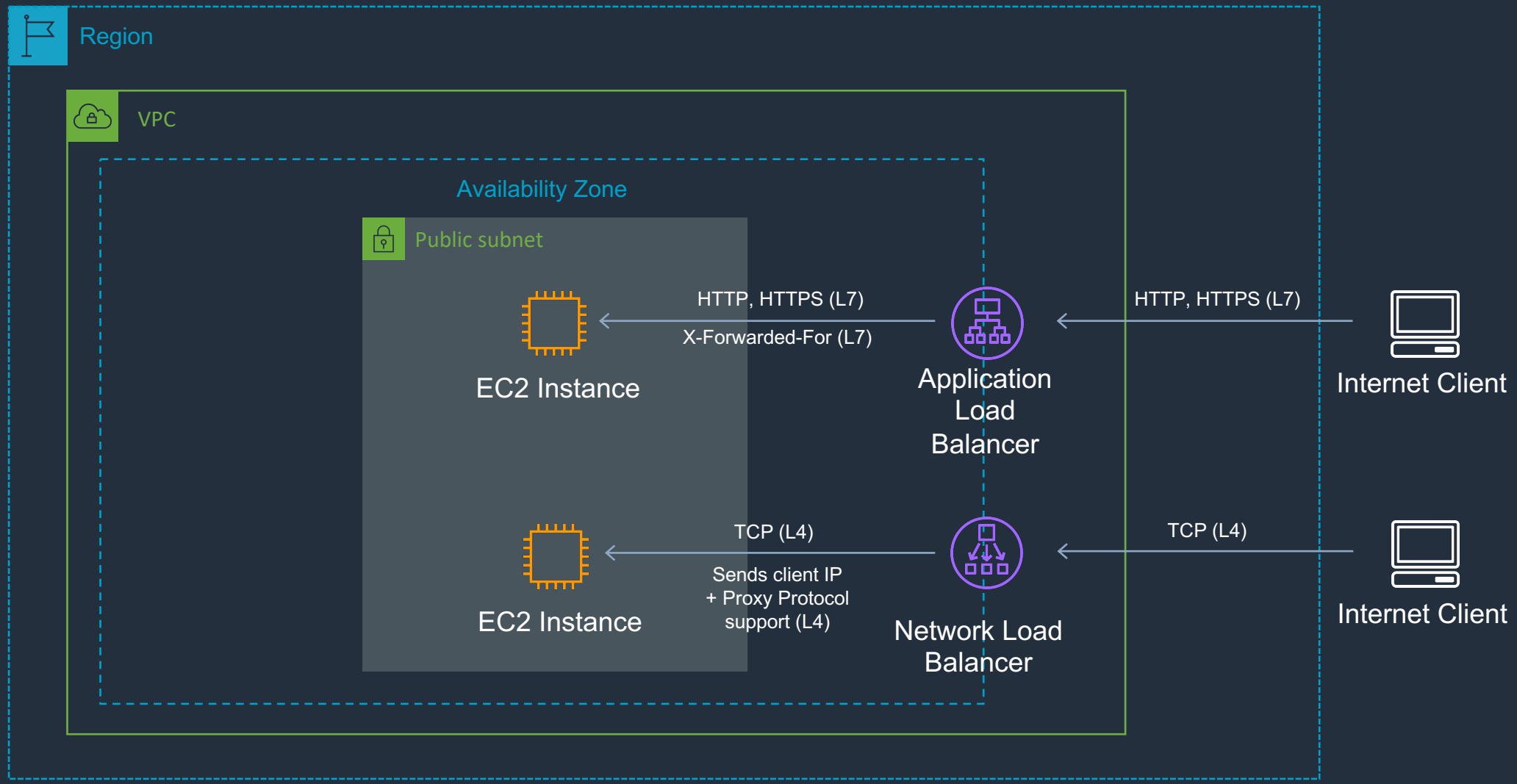


DigitalCloud TRAINING

Section 4: CLB - Proxy Protocol and X-Forwarded-For

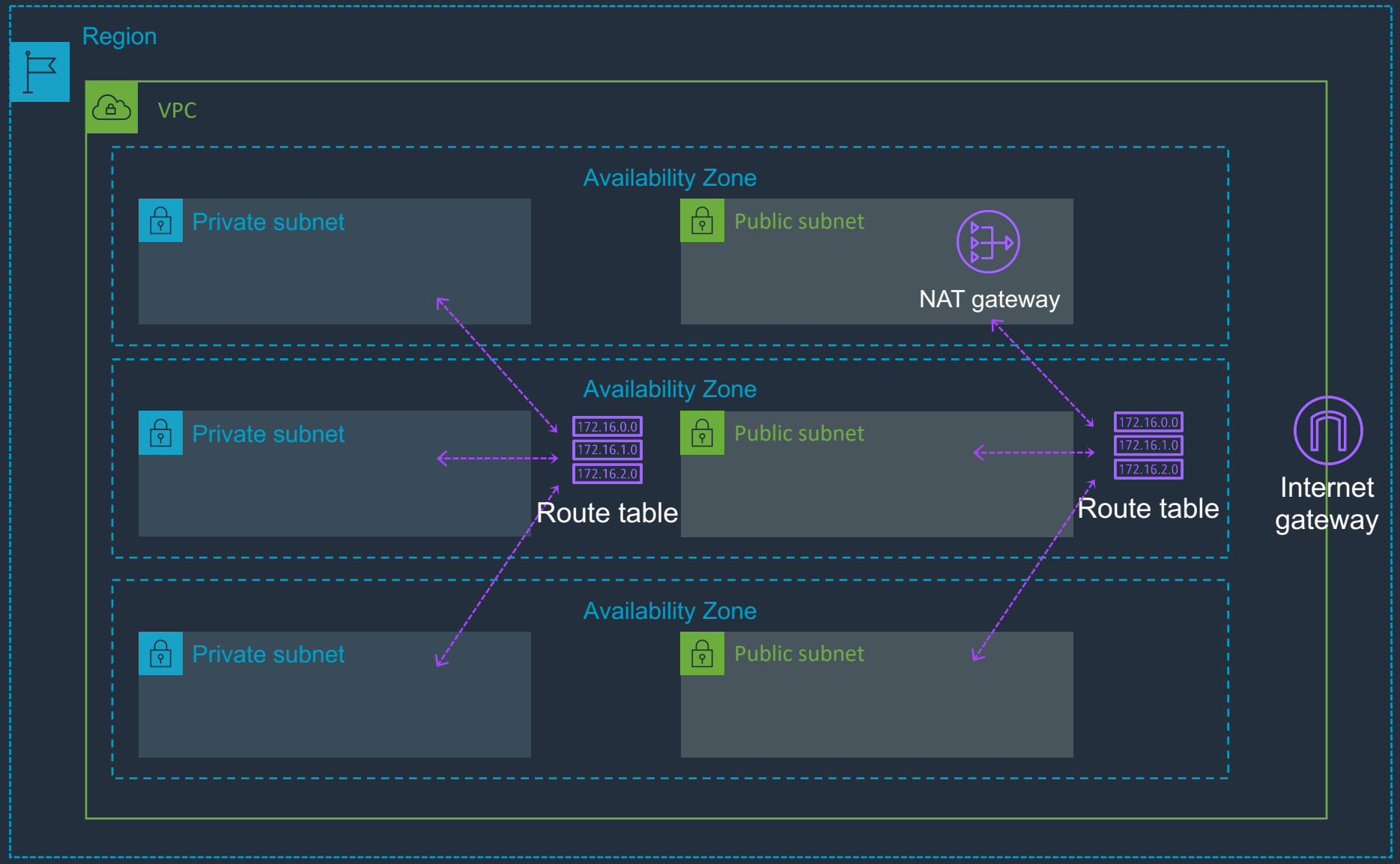


Section 4: ALB/NLB - Proxy Protocol, X-Forwarded-For and Access Logging



DigitalCloud TRAINING

Section 5: Creating a Custom VPC



Public Route Table

Destination	Target
10.0.0.0/16	Local
0.0.0.0/0	*igw-id*

Private Route Table

Destination	Target
10.0.0.0/16	Local
0.0.0.0/0	*nat-gateway-id*

DigitalCloud TRAINING

Section 5: Security Groups

Default Security Group

Inbound:

Source	Protocol	Port
Security Group ID	All	All

Outbound:

Destination	Protocol	Port
0.0.0.0/0	All	All
::/0	All	All

Custom Security Group

Inbound:

Source	Protocol	Port

Outbound:

Destination	Protocol	Port
0.0.0.0/0	All	All
::/0	All	All

DigitalCloud TRAINING

Section 5: Network Access Control Lists (NACLs)

Default NACL

Inbound:

Protocol	Port	Source	Action
All	All	0.0.0.0/0	ALLOW
All	All	::/0	ALLOW

Outbound:

Protocol	Port	Source	Action
All	All	0.0.0.0/0	ALLOW
All	All	::/0	ALLOW

Custom NACL

Inbound:

Protocol	Port	Source	Action
All	All	0.0.0.0/0	DENY
All	All	::/0	DENY

Outbound:

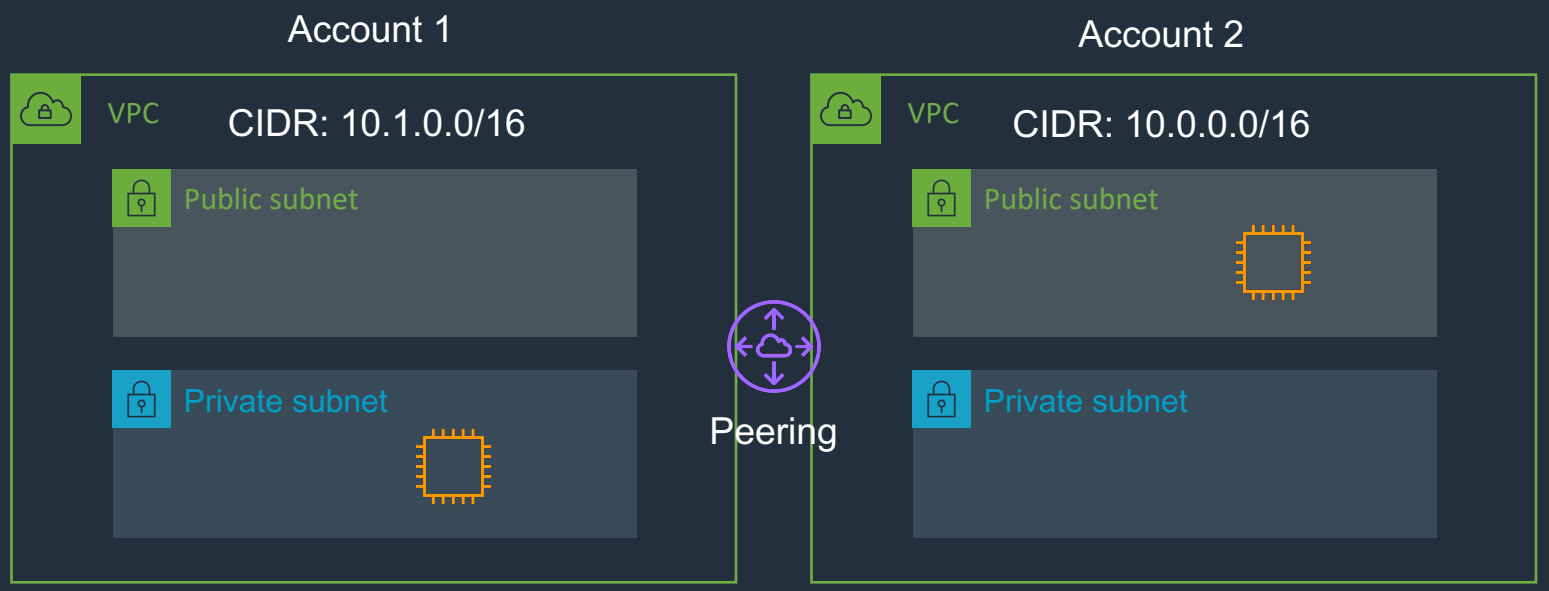
Protocol	Port	Source	Action
All	All	0.0.0.0/0	DENY
All	All	::/0	DENY

DigitalCloud TRAINING

Section 5: Security Groups vs Network ACLs

Security Group	Network ACL
Operates at the instance level	Operates at the subnet level
Supports allow rules only	Supports allow and deny rules
Stateful	Stateless
Evaluates all rules	Processes rules in order
Applies to an instance only if associated with a group	Automatically applies to all instances in the subnets its associated with

Section 5: VPC Peering



Security Group

Inbound:

Source	Protocol	Port
Security Group ID	All ICMP v4	All

Route Table

Destination	Target
10.0.0.0/16	*peering-connection-id*

Security Group

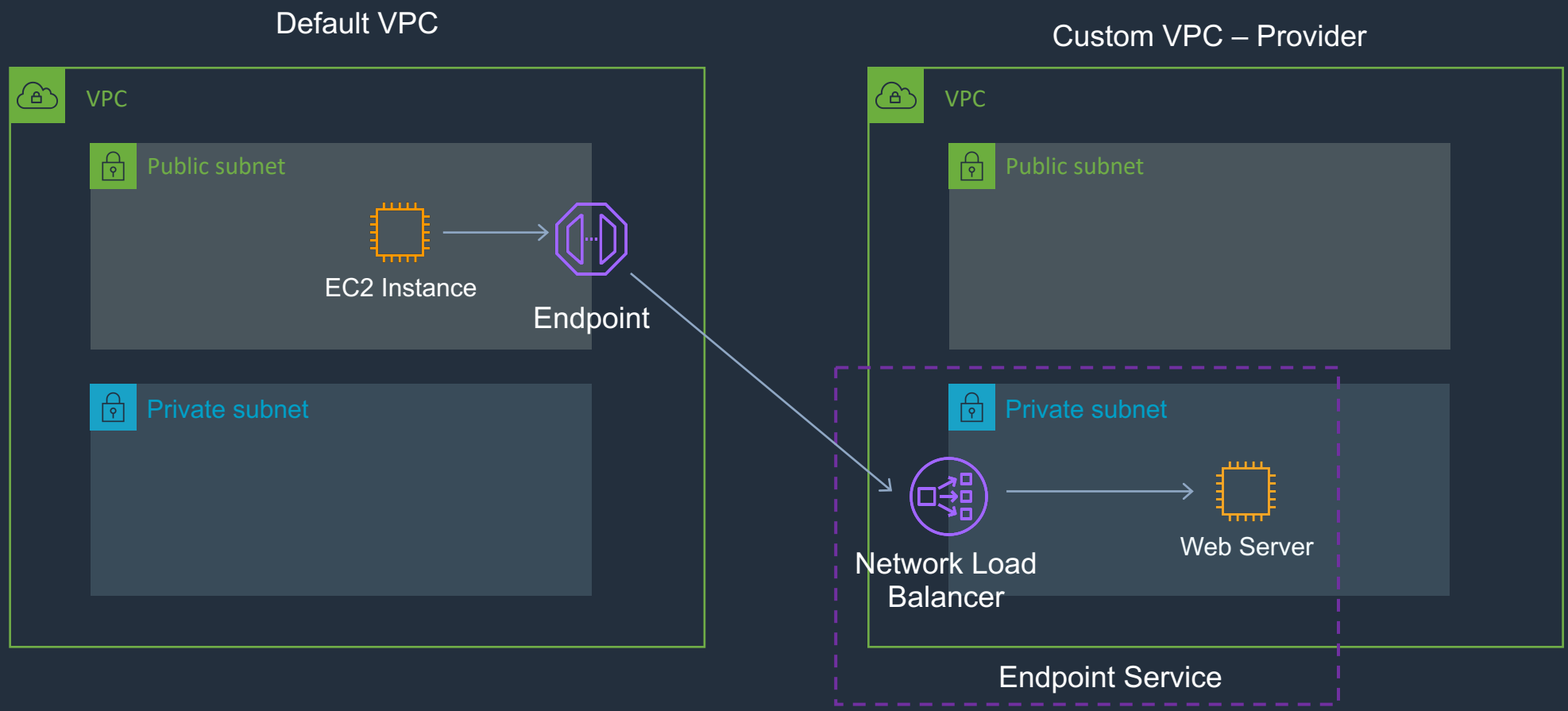
Inbound:

Source	Protocol	Port
0.0.0.0/0	TCP	22

Route Table

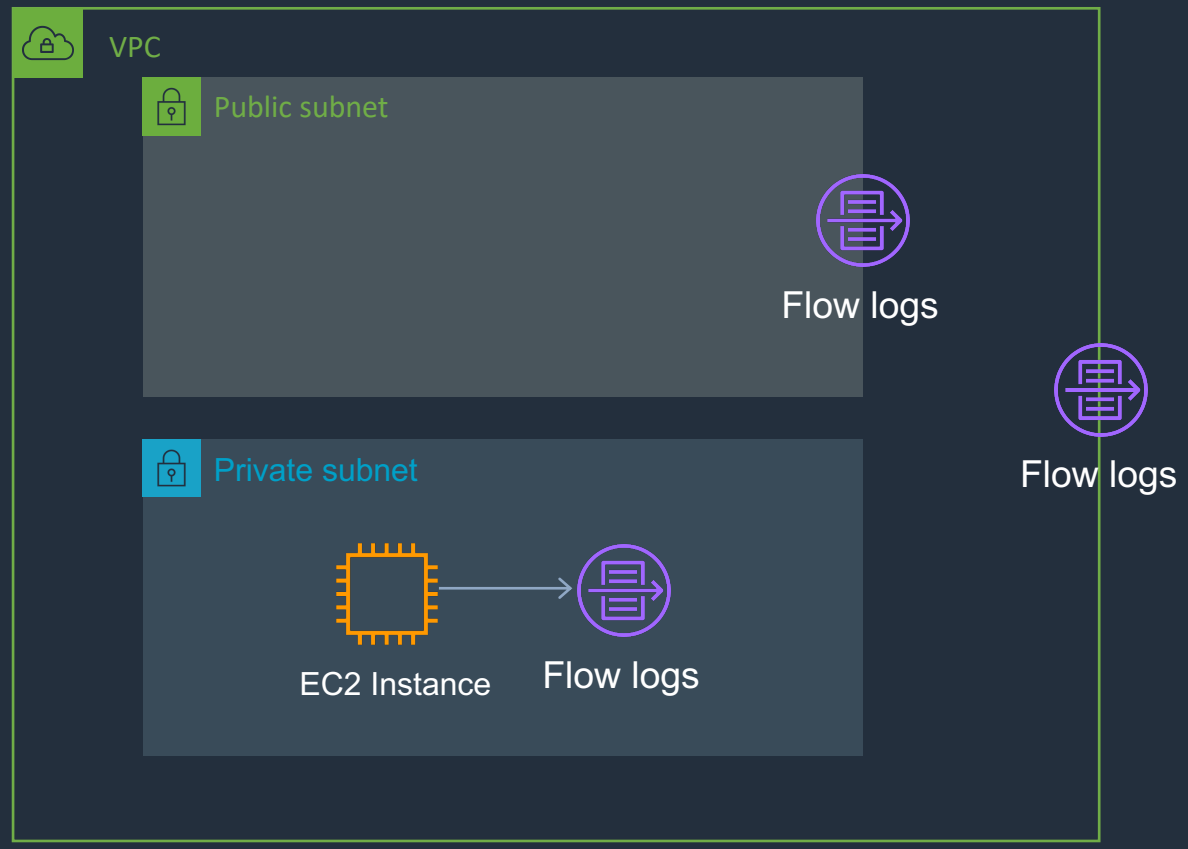
Destination	Target
10.1.0.0/16	*peering-connection-id*

Section 5: VPC Endpoint Services

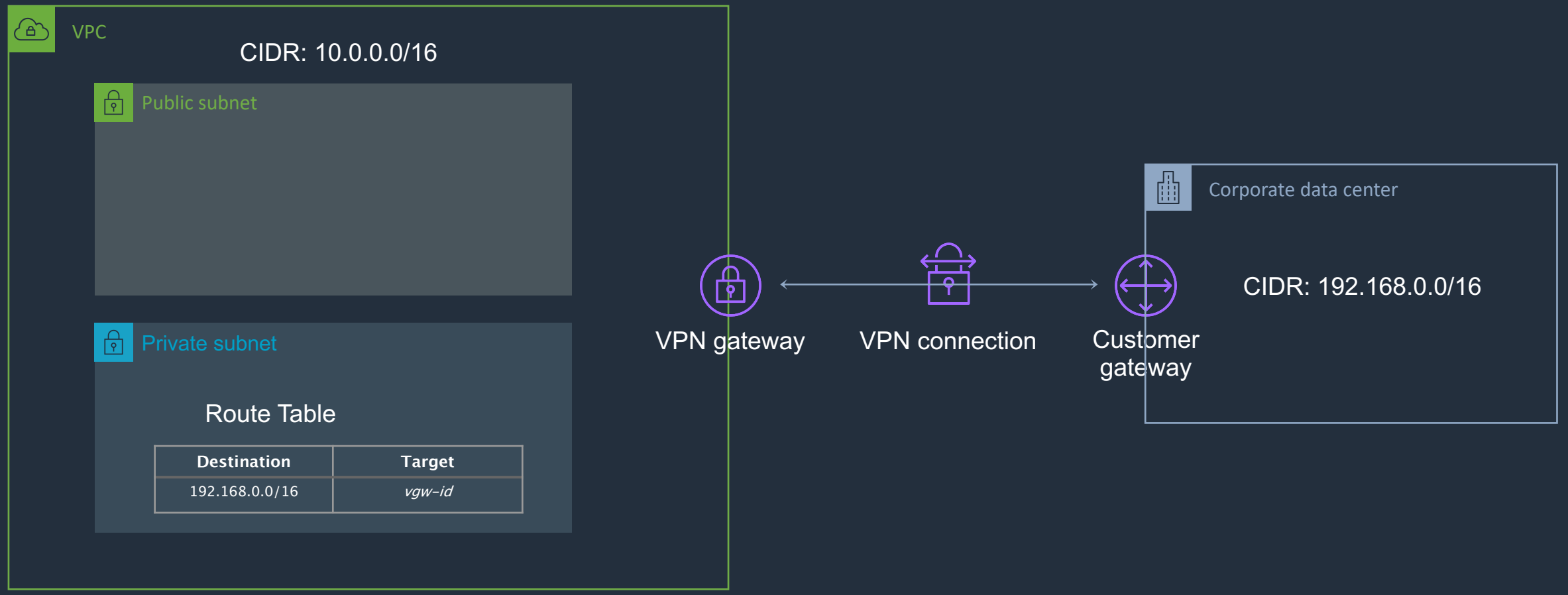


DigitalCloud TRAINING

Section 5: VPC Flow Logs

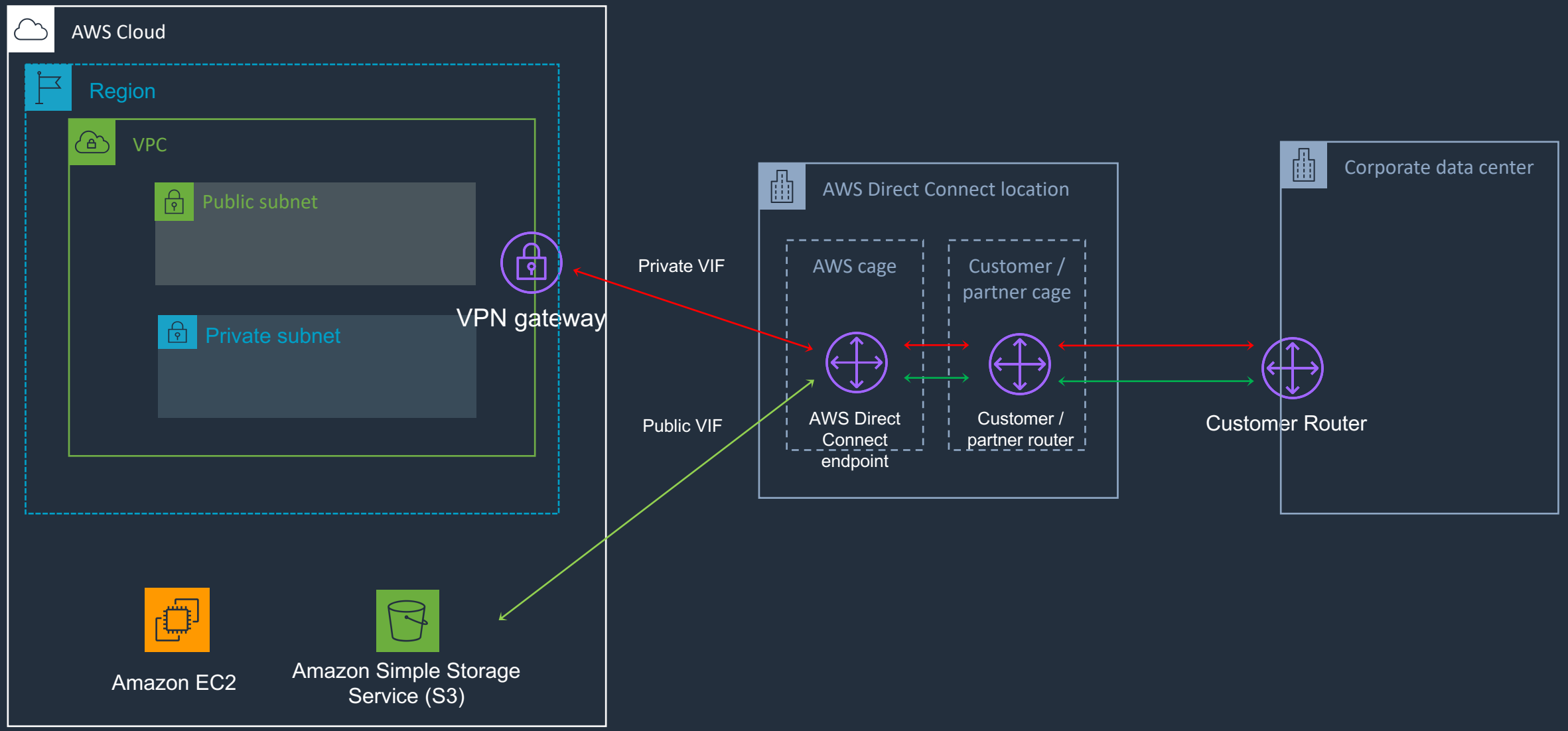


Section 5: Virtual Private Networks (VPN)

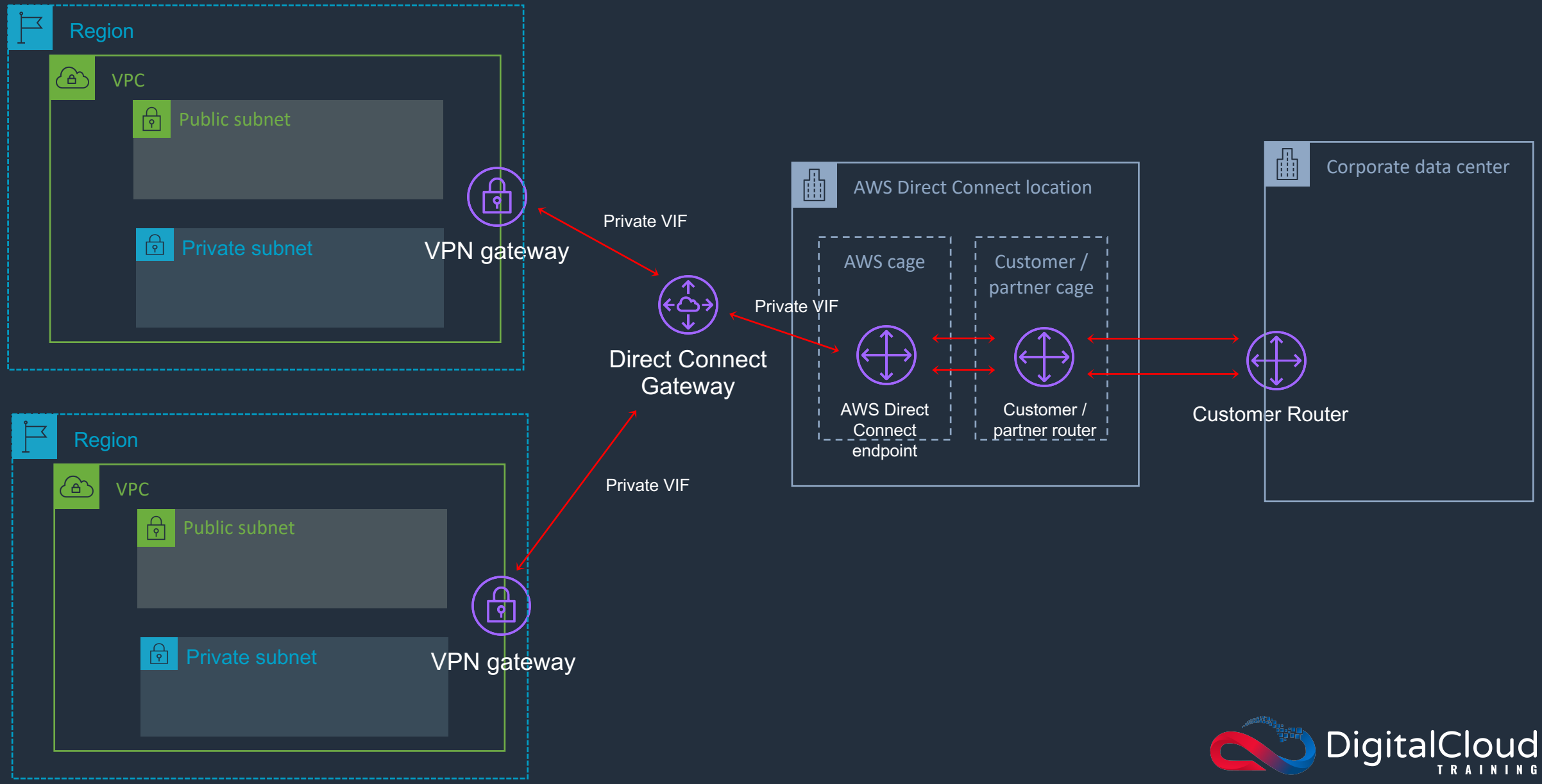


DigitalCloud TRAINING

Section 5: AWS Direct Connect



Section 5: AWS Direct Connect Gateway



DigitalCloud TRAINING

Section 6: Route 53 Overview

Amazon Route 53

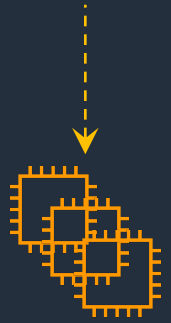
Domain Registration

.net
.com
.org

Hosted zone

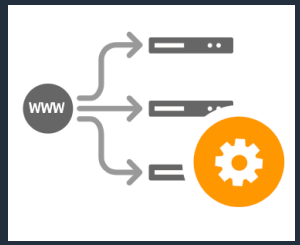
example.com
dctlabs.com

Health Checks



EC2 Instances

Traffic Flow



DigitalCloud TRAINING

Section 6: Route 53 DNS Record Types

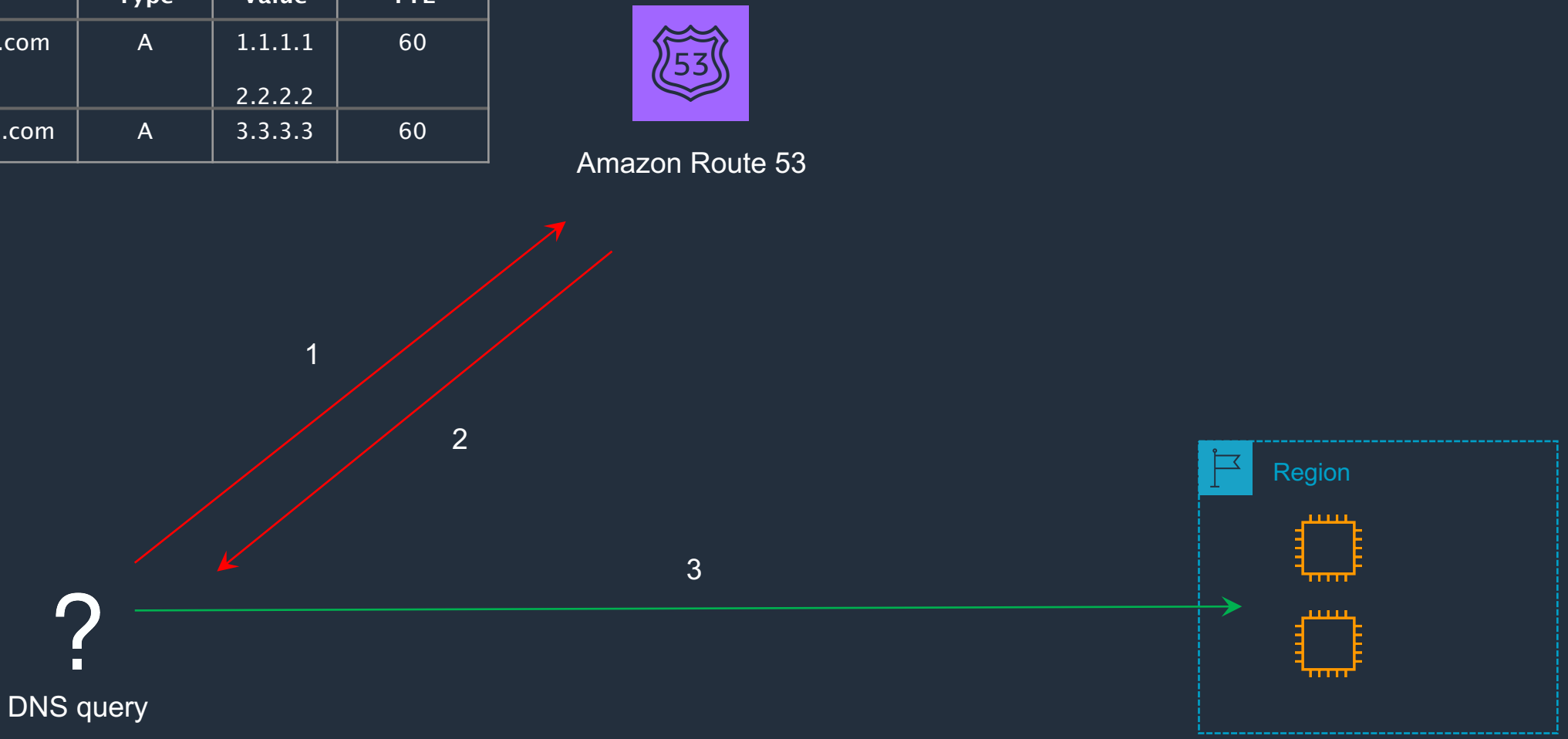
Supported DNS records

- A (address record)
- AAAA (IPv6 address record)
- CNAME (canonical name record)
- Alias (an Amazon Route 53-specific virtual record)
- CAA (certification authority authorization)
- MX (mail exchange record)
- NAPTR (name authority pointer record)
- NS (name server record)
- PTR (pointer record)
- SOA (start of authority record)
- SPF (sender policy framework)
- SRV (service locator)
- TXT (text record)

CNAME	Alias
Route 53 charges for CNAME queries	Route 53 doesn't charge for alias queries to AWS resources
You can't create a CNAME record at the top node of a DNS namespace (zone apex)	You can create an alias record at the zone apex (however you can't route to a CNAME at the zone apex)
A CNAME can point to any DNS record that is hosted anywhere	An alias record can only point to a CloudFront distribution, Elastic Beanstalk environment, ELB, S3 bucket as a static website, or to another record in the same hosted zone that you're creating the alias record in

Section 6: Route 53 - Simple Routing Policy

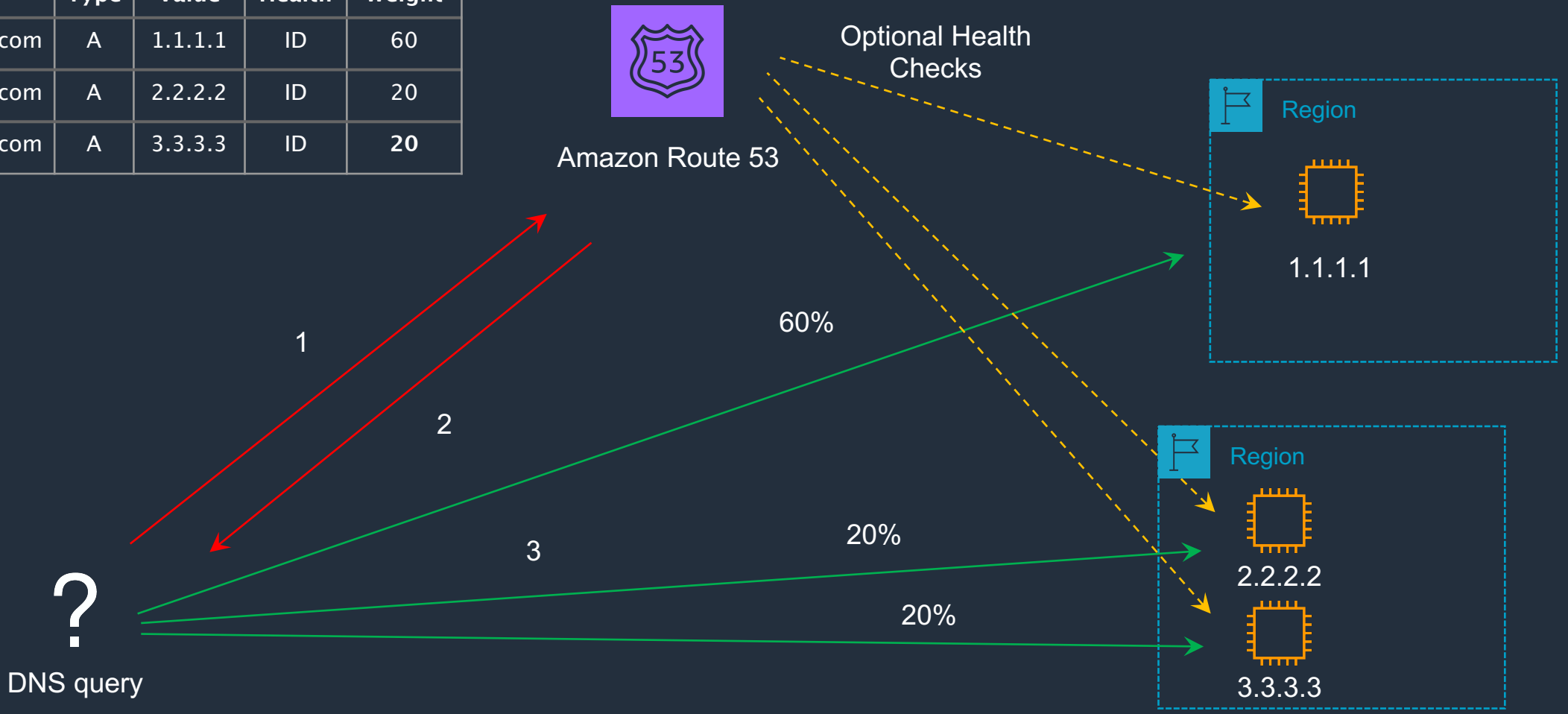
Name	Type	Value	TTL
simple.dctlabs.com	A	1.1.1.1 2.2.2.2	60
simpler.dctlabs.com	A	3.3.3.3	60



DigitalCloud TRAINING

Section 6: Route 53 - Weighted Routing Policy

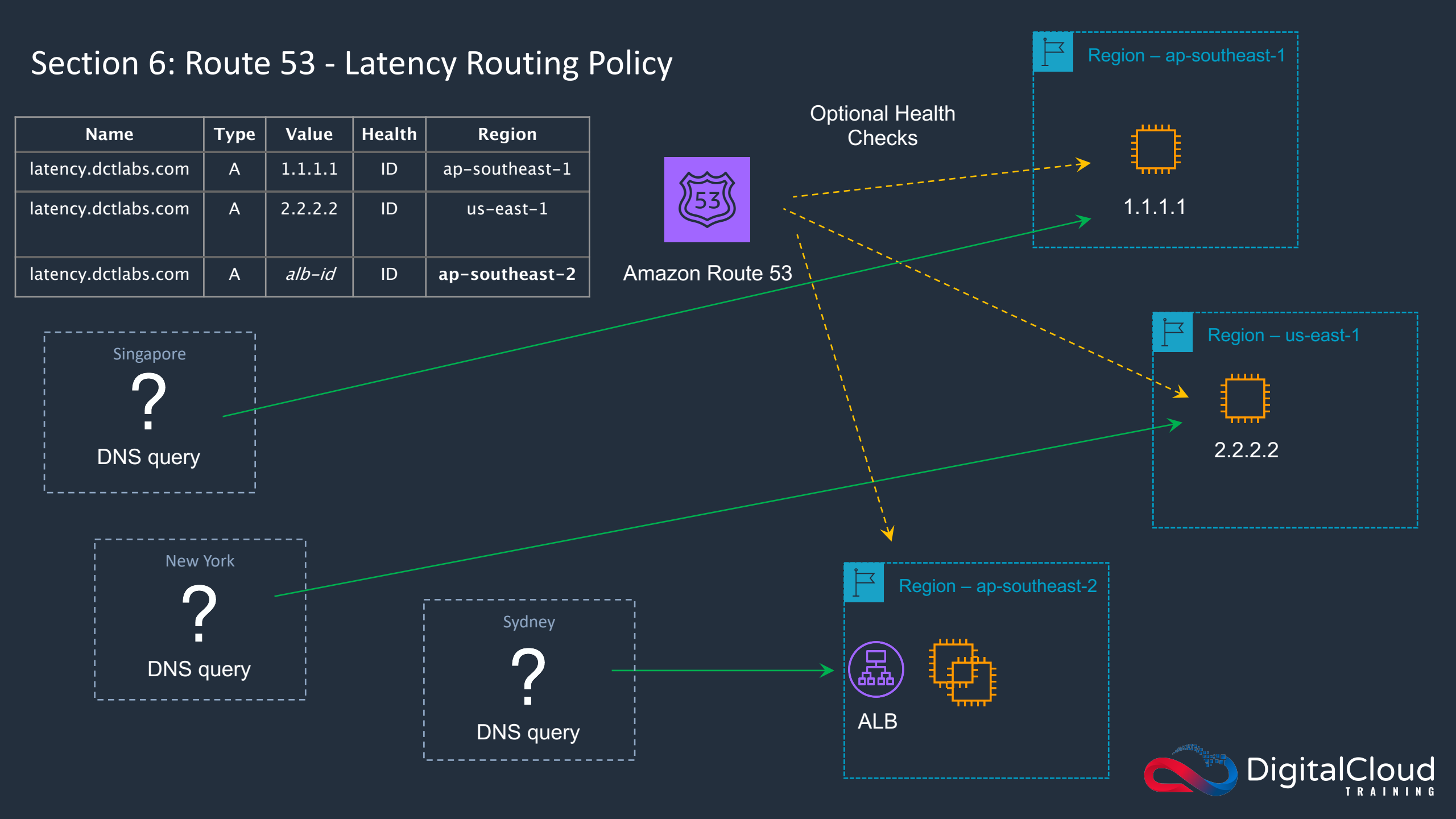
Name	Type	Value	Health	Weight
weighted.dctlabs.com	A	1.1.1.1	ID	60
weighted.dctlabs.com	A	2.2.2.2	ID	20
weighted.dctlabs.com	A	3.3.3.3	ID	**20**



DigitalCloud TRAINING

Section 6: Route 53 - Latency Routing Policy

Name	Type	Value	Health	Region
latency.dctlabs.com	A	1.1.1.1	ID	ap-southeast-1
latency.dctlabs.com	A	2.2.2.2	ID	us-east-1
latency.dctlabs.com	A	*alb-id*	ID	**ap-southeast-2**



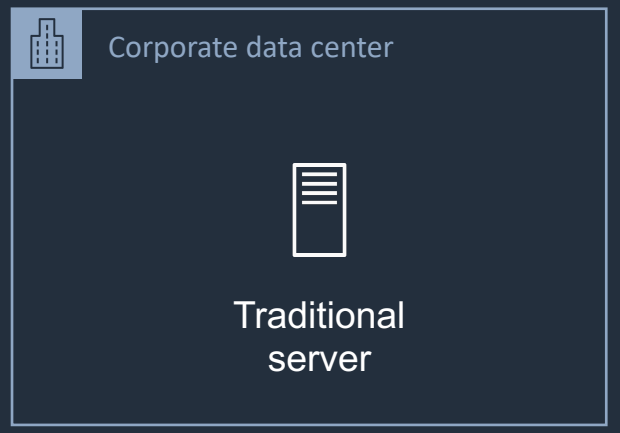
DigitalCloud TRAINING

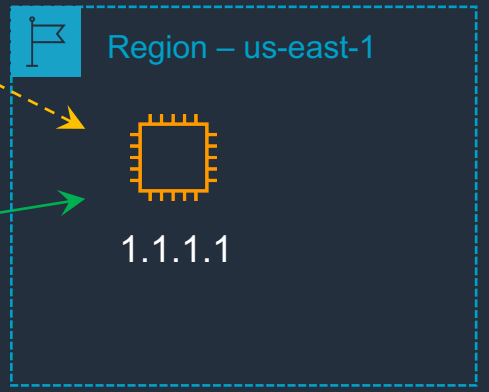
Section 6: Route 53 - Failover Routing Policy

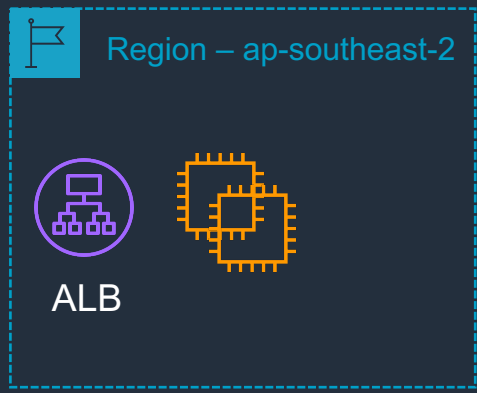
Name	Type	Value	Health	Record Type
failover.dctlabs.com	A	1.1.1.1	ID	Primary
failover.dctlabs.com	A	*alb-id*		Secondary

Amazon Route 53

Health Check required on Primary







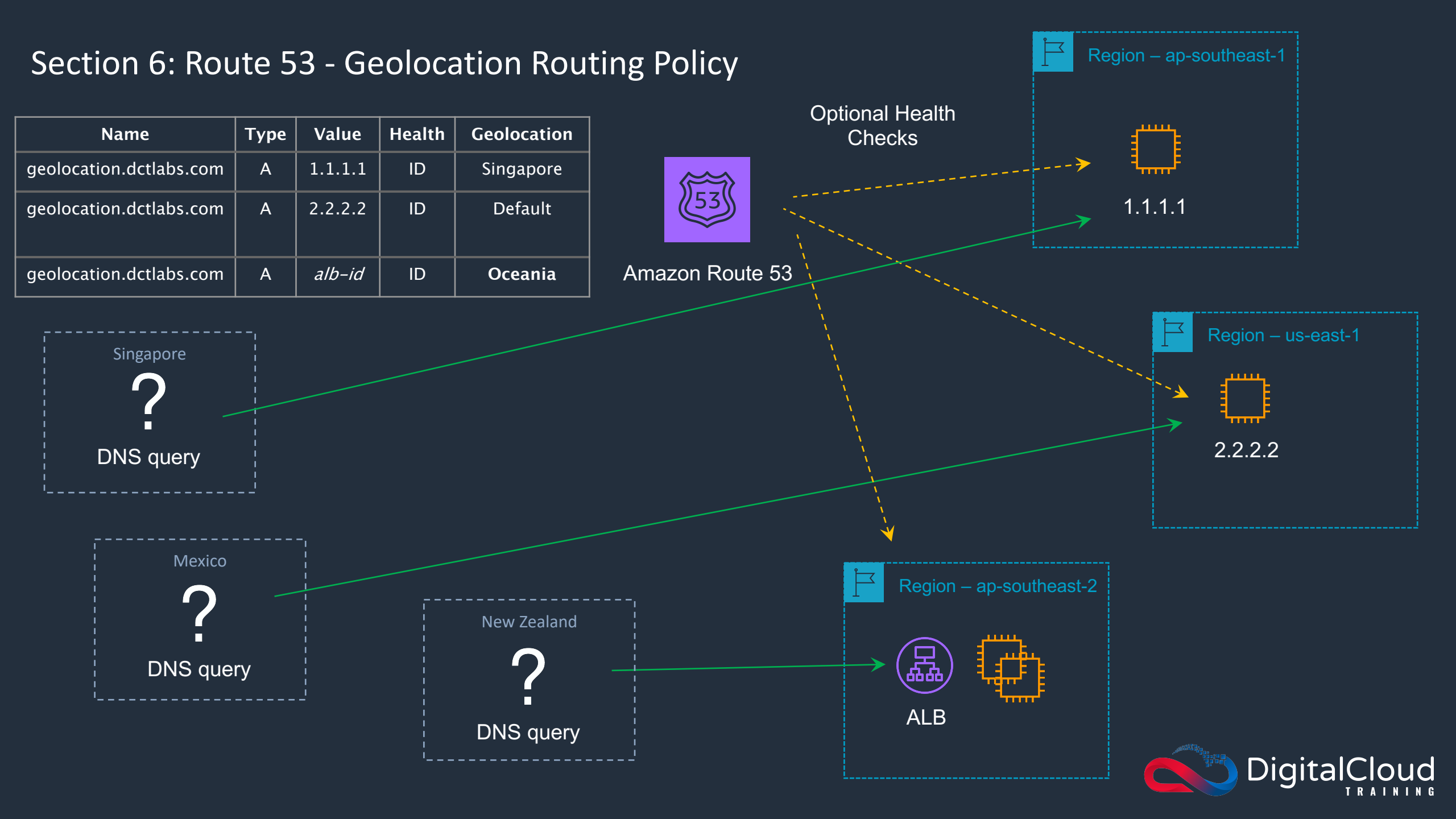
?

DNS query

DigitalCloud TRAINING

Section 6: Route 53 - Geolocation Routing Policy

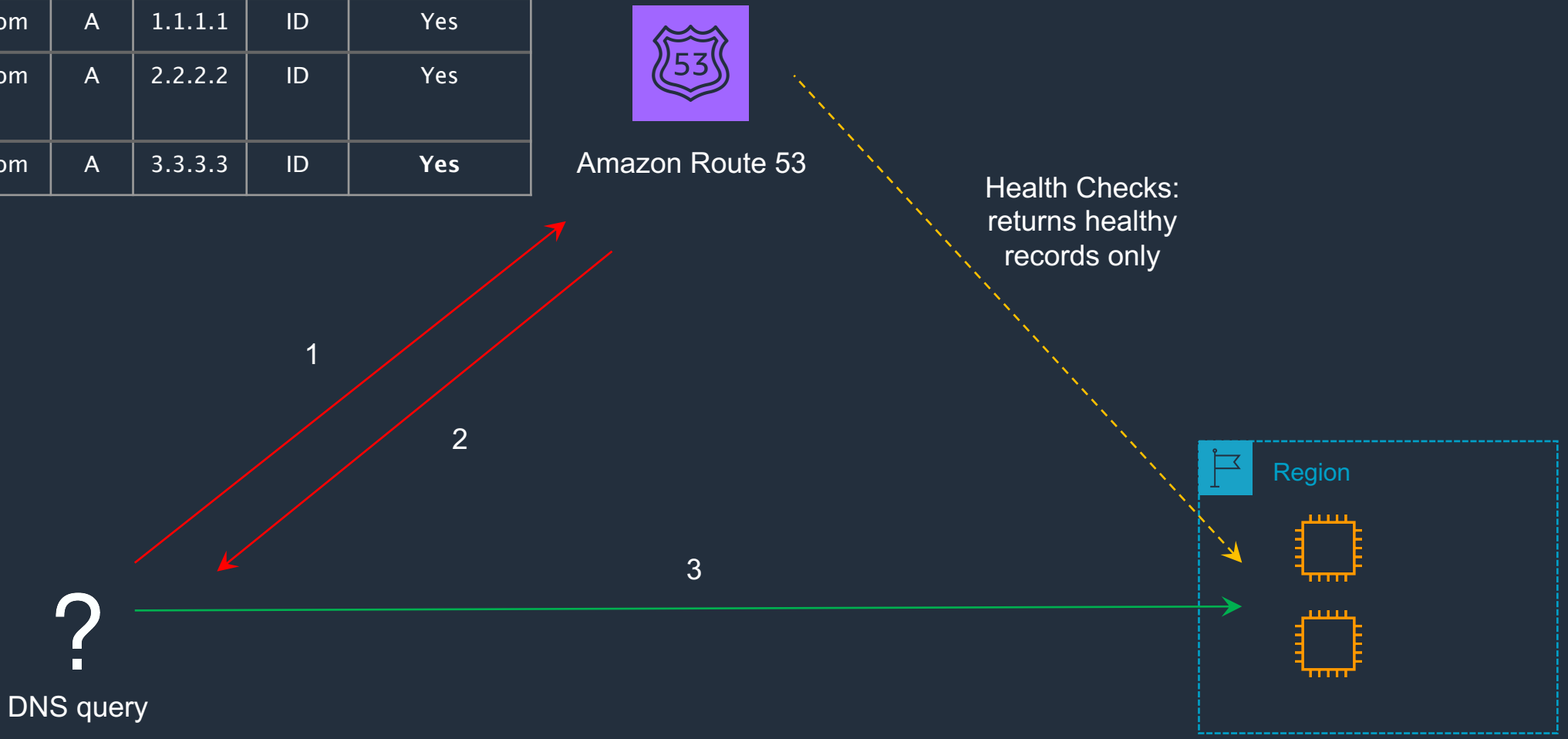
Name	Type	Value	Health	Geolocation
geolocation.dctlabs.com	A	1.1.1.1	ID	Singapore
geolocation.dctlabs.com	A	2.2.2.2	ID	Default
geolocation.dctlabs.com	A	*alb-id*	ID	**Oceania**



DigitalCloud TRAINING

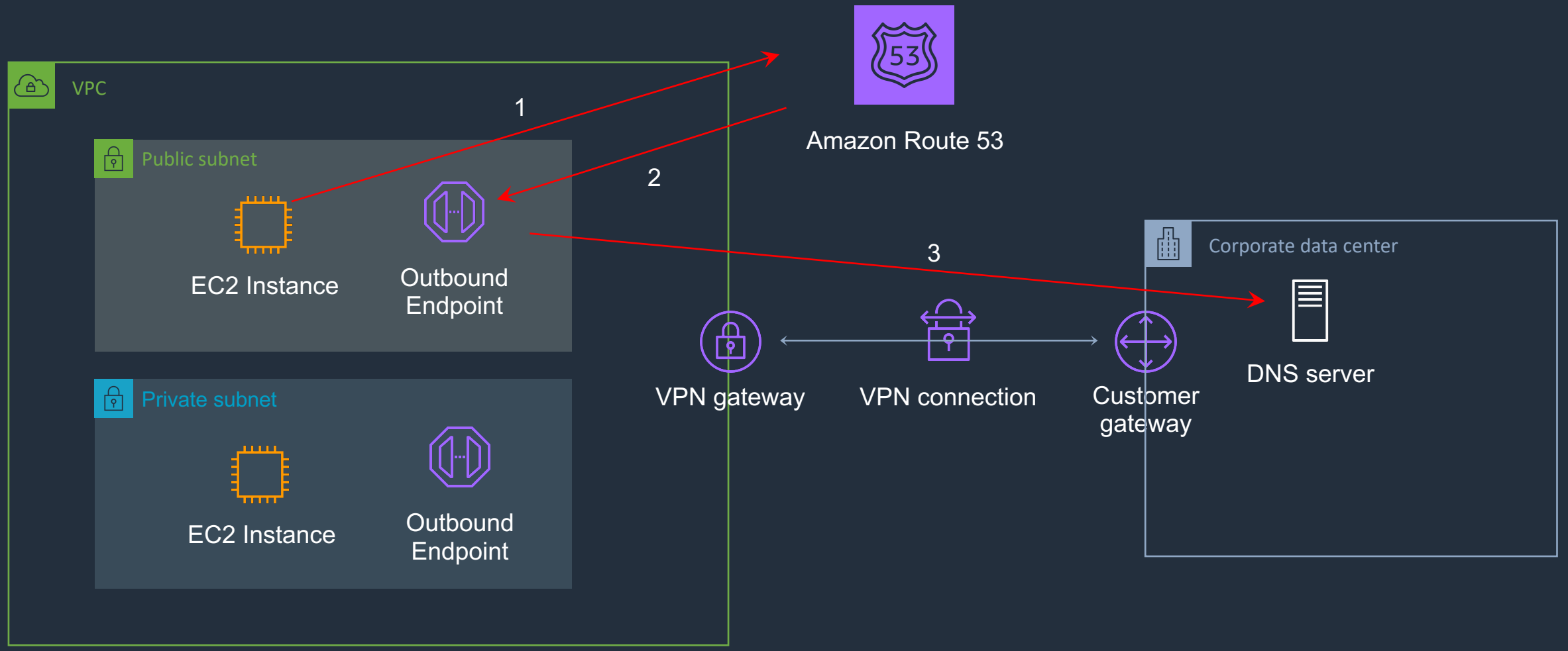
Section 6: Route 53 - Multivalue Routing Policy

Name	Type	Value	Health	Multi Value
multivalue.dctlabs.com	A	1.1.1.1	ID	Yes
multivalue.dctlabs.com	A	2.2.2.2	ID	Yes
multivalue.dctlabs.com	A	3.3.3.3	ID	**Yes**

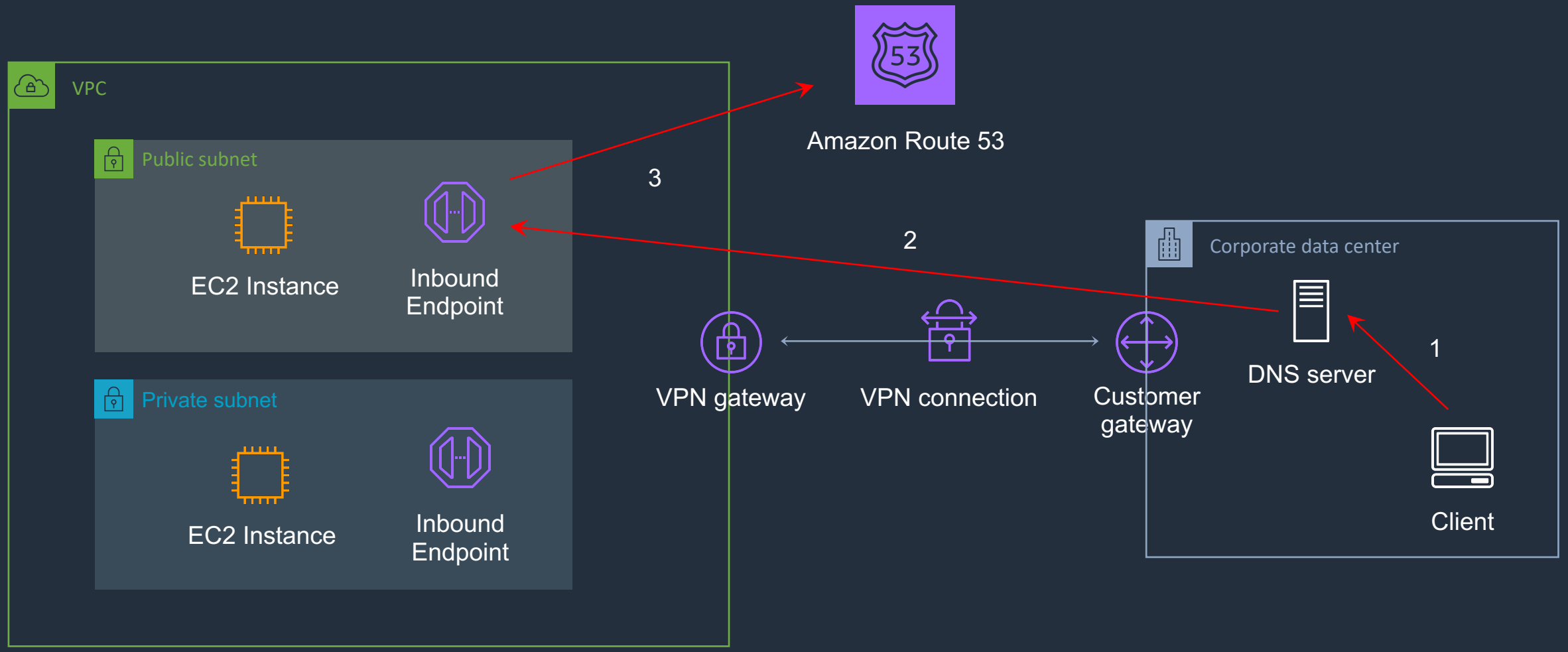


DigitalCloud TRAINING

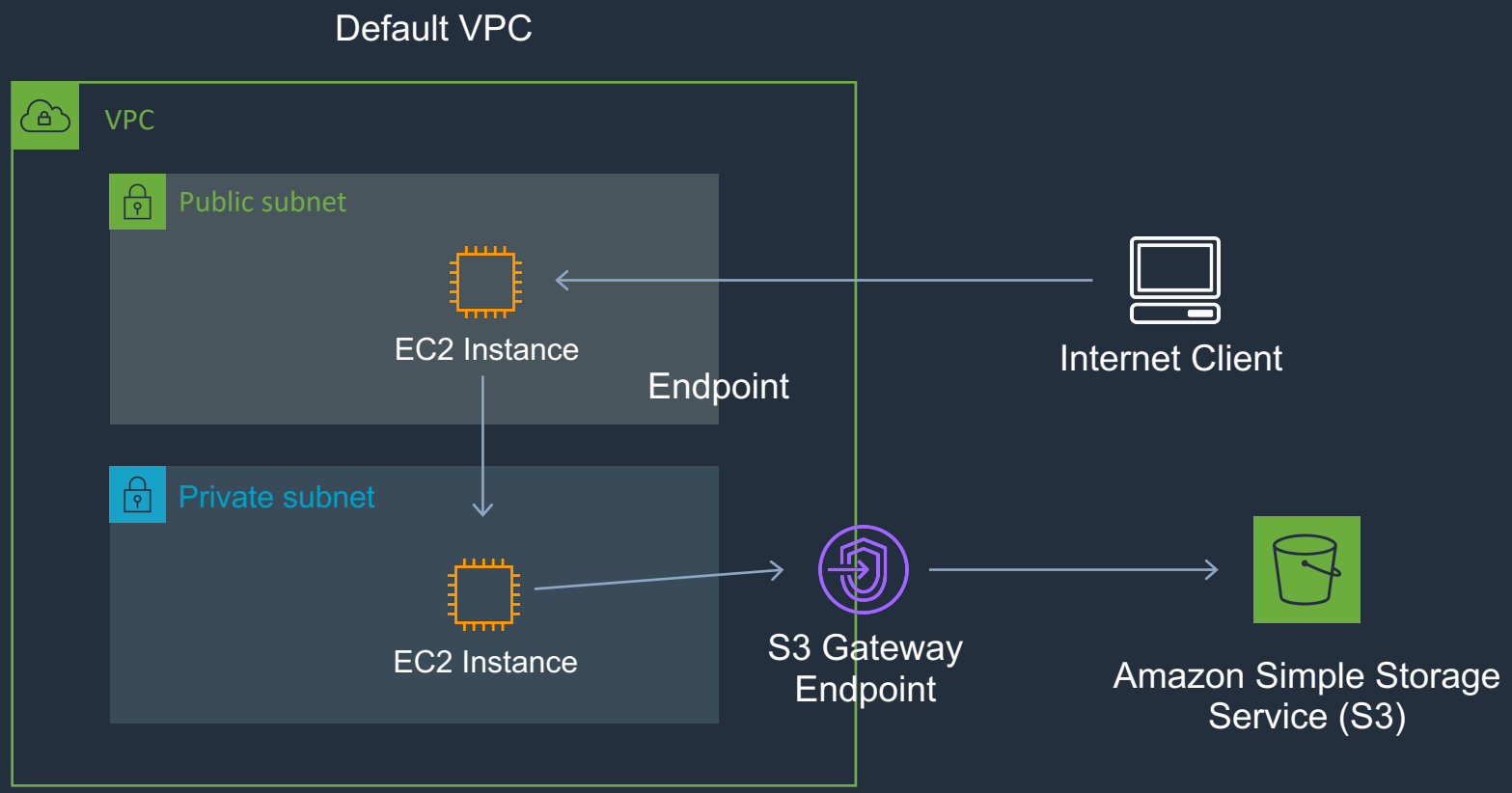
Section 6: Route 53 Resolver – Outbound Endpoints



Section 6: Route 53 Resolver – Inbound Endpoints



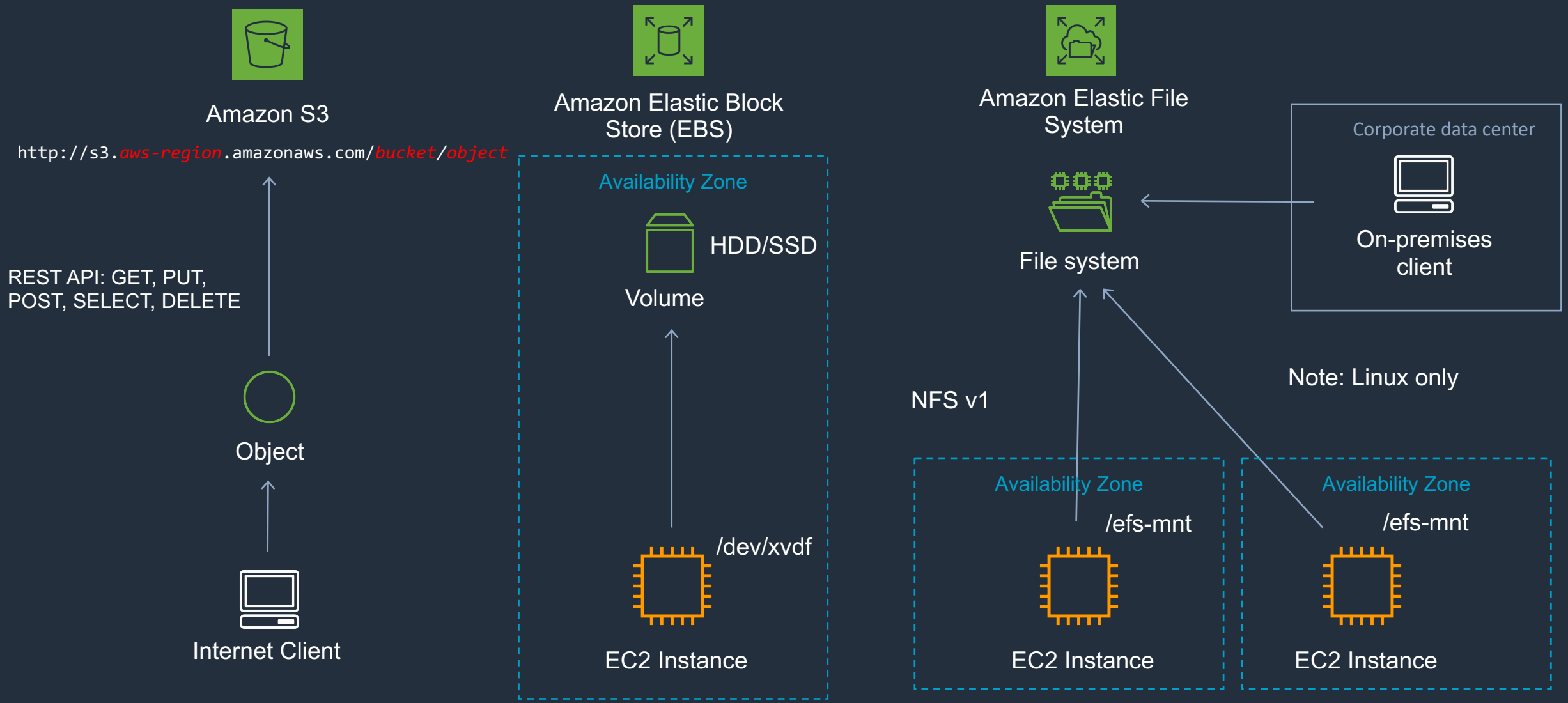
Section 7: S3 Gateway Endpoints



Route Table

Destination	Target
pl-6ca54005 (com.amazonaws.ap-southeast-2.s3, 54.231.248.0/22, 54.231.252.0/24, 52.95.128.0/21)	*vpce-ID*

Section 7: Block, Object and File Storage



DigitalCloud TRAINING

Section 7: Amazon S3 Overview

Bucket

`http://bucket.s3.aws-region.amazonaws.com`

`http://s3.aws-region.amazonaws.com/bucket`

Object

- Key
- Version ID
- Value
- Metadata
- Subresources
- Access control information

Amazon S3

Private Connection

Public Internet

S3 Gateway Endpoint

VPC

Public subnet

EC2 Instance

Internet gateway

Private subnet

EC2 Instance

Internet Client

DigitalCloud TRAINING

Section 7: Identity-Based and Resource-Based Policies

Example Policy

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
  "Sid": "SeeBucketListInTheConsole",
  "Action": ["s3:ListAllMyBuckets"],
  "Effect": "Allow",
  "Resource": ["arn:aws:s3:::*"]
    }
  ]
}
```

Identity-based policies

IAM Role Inline Policy

IAM User Inline Policy

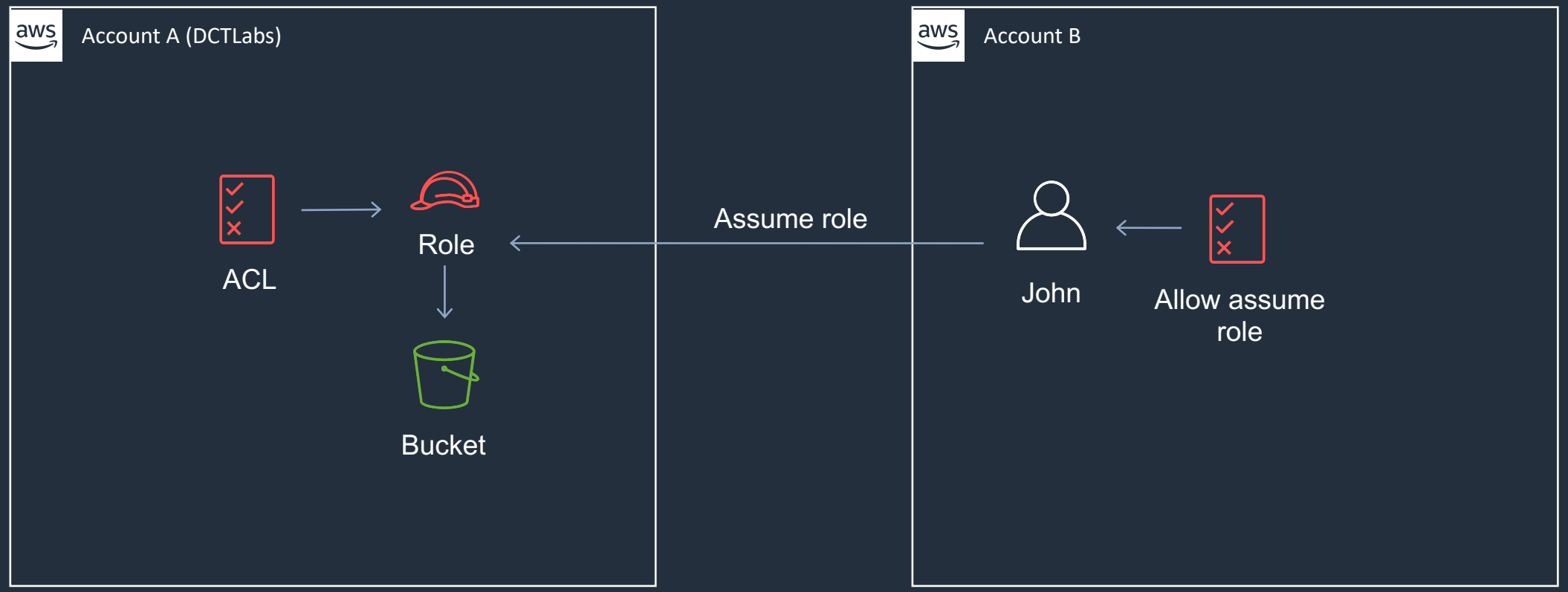
IAM Group Policy

Resource-based policy

Bucket Policy

DigitalCloud TRAINING

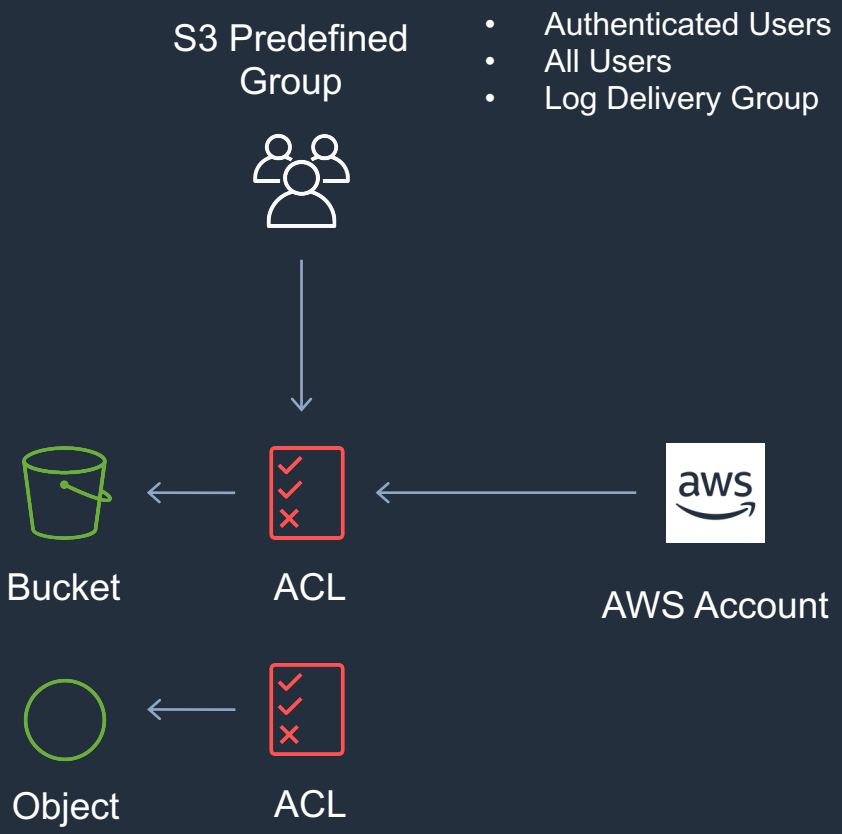
Section 7: Cross-Account Access



Section 7: Access Control Lists

Example ACL

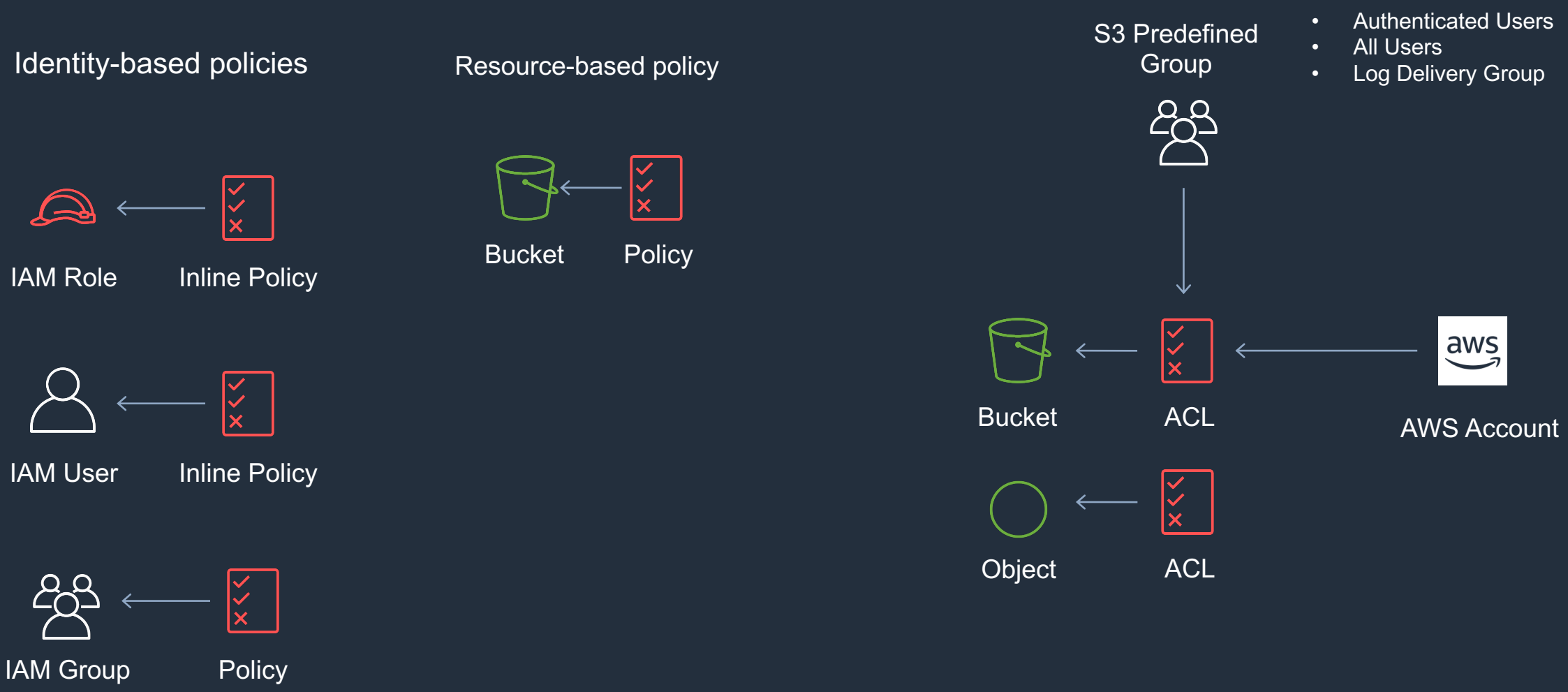
```
... <AccessControlPolicy>
 <Owner>
   <ID> AccountACanonicalUserID </ID>
  <DisplayName> AccountADisplayName </DisplayName>
 </Owner>
<AccessControlList>
 ...
   <Grant>
    <Grantee xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:type="CanonicalUser">
    <ID> AccountBCanonicalUserID </ID>
    <DisplayName> AccountBDisplayName </DisplayName>
    </Grantee>
    <Permission> WRITE </Permission>
   </Grant>
...
  </AccessControlList>
</AccessControlPolicy>
```



Section 7: Access Control List Permissions

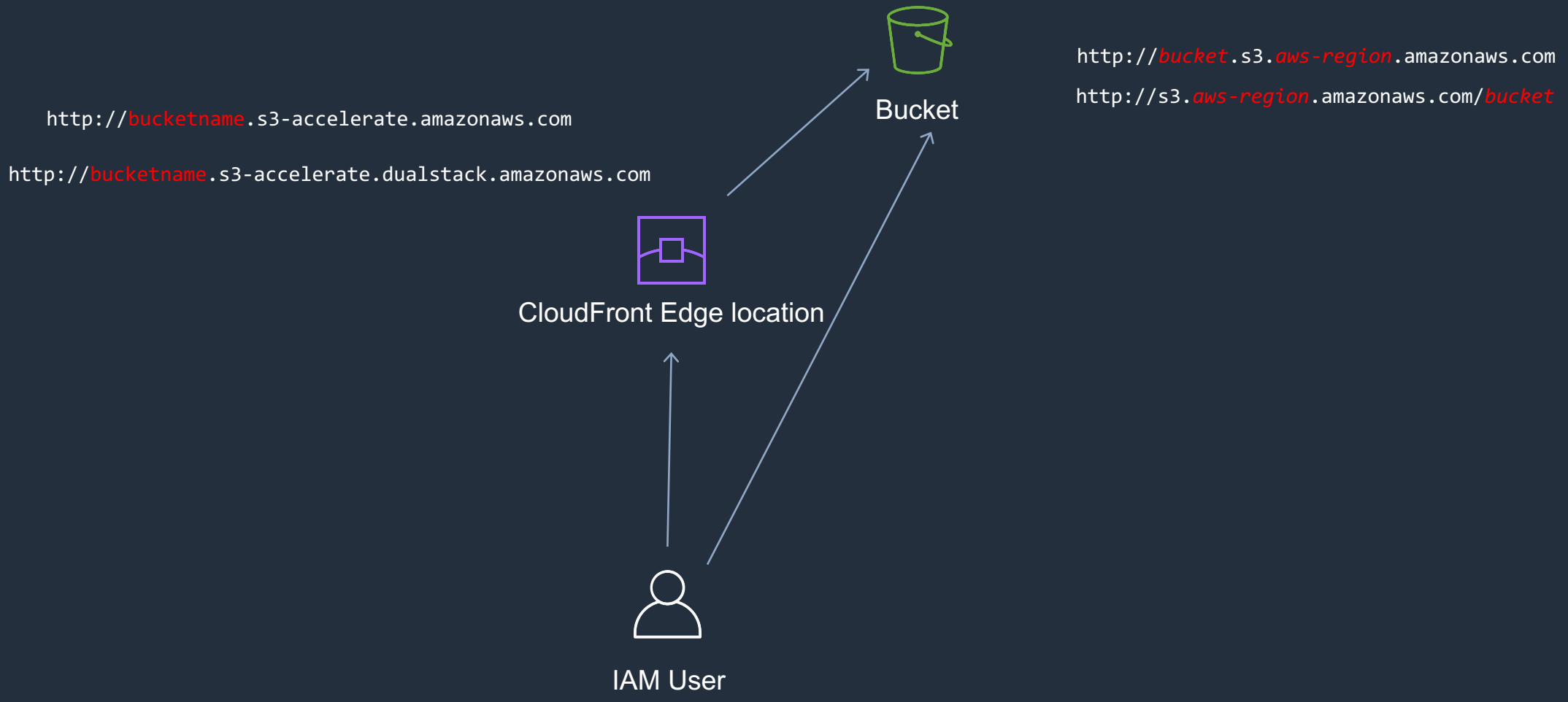
Permissions	When granted on a bucket	When granted on an object
READ	Allows grantee to list the objects in the bucket	Allows grantee to read the object data and its metadata
WRITE	Allows grantee to create, overwrite, and delete any object in the bucket	Not applicable
READ _ACP	Allows grantee to read the bucket ACL	Allows grantee to read the object ACL
WRITE_ACP	Allows grantee to write the ACL for the applicable bucket	Allows grantee to write the ACL for the applicable object
FULL_CONTROL	Allows grantee the READ, WRITE, READ_ACP, and WRITE_ACP permissions on the bucket	Allows grantee the READ, READ_ACP, and WRITE_ACP permissions on the object

Section 7: Choosing Access Control Options



DigitalCloud TRAINING

Section 7: Transfer Acceleration



DigitalCloud TRAINING

Section 7: S3 Encryption

Server-side encryption with S3 managed keys (SSE-S3)

- S3 managed keys
- Unique object keys
- Master key
- AES 256

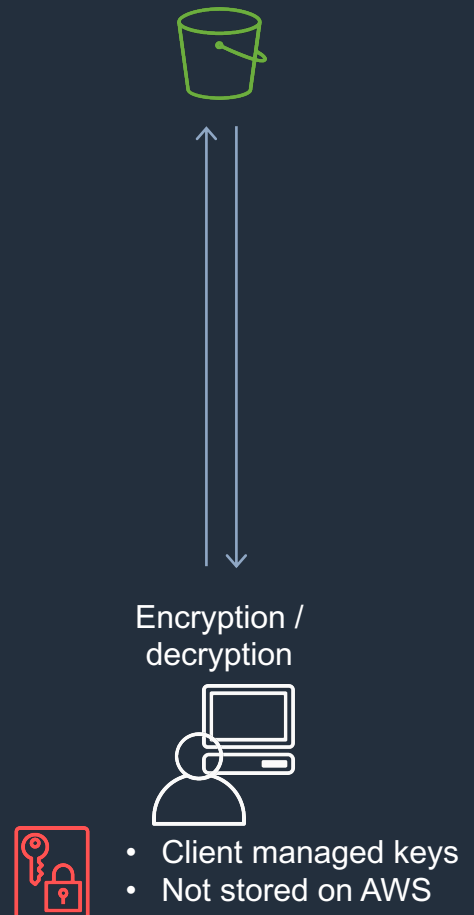
Server-side encryption with AWS KMS managed keys (SSE-KMS)

- KMS managed keys
- Customer master keys
- CMK can be customer generated

Server-side encryption with client provided keys (SSE-C)

- Client managed keys
- Not stored on AWS

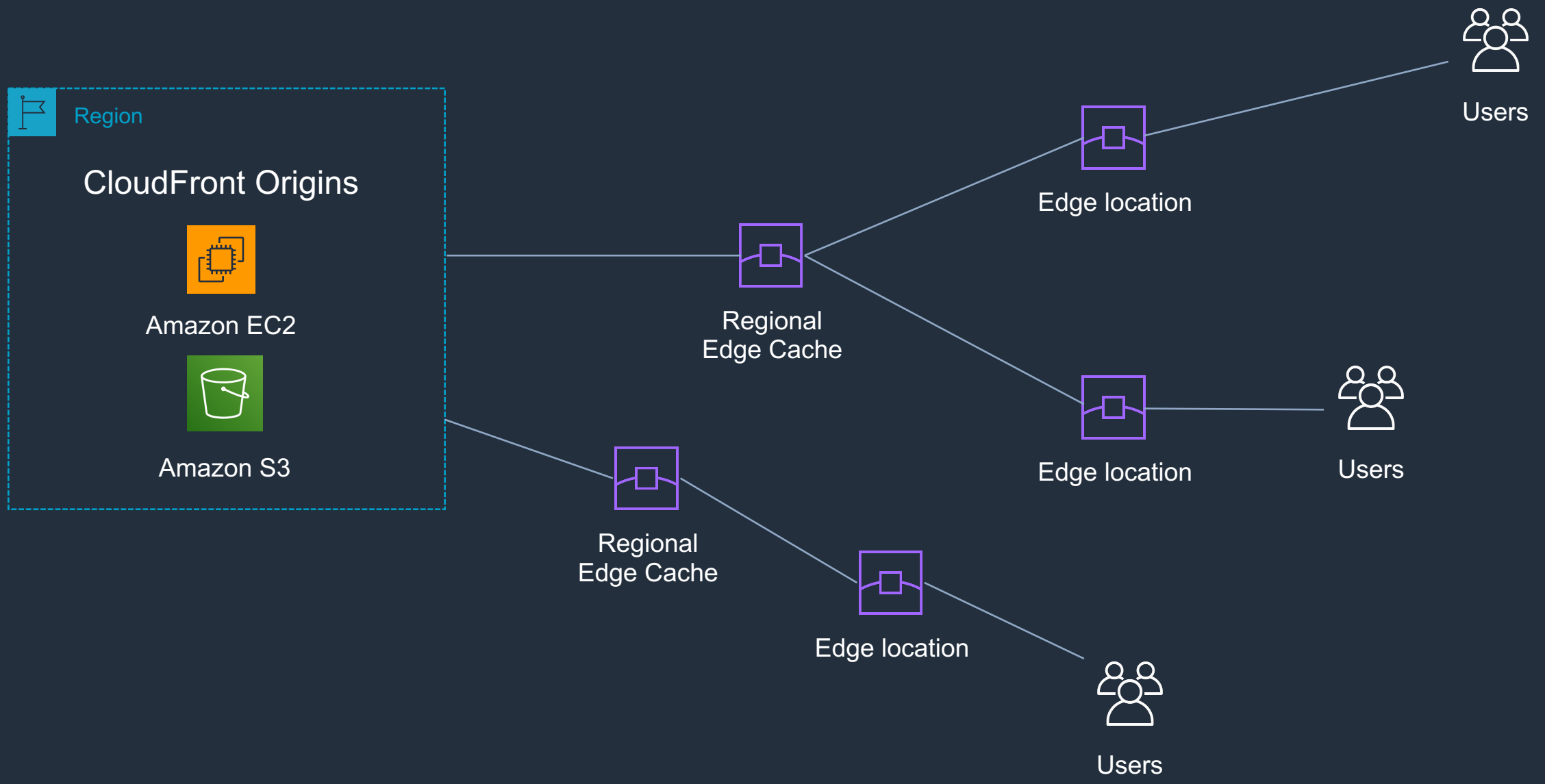
Client side encryption



- Client managed keys
- Not stored on AWS

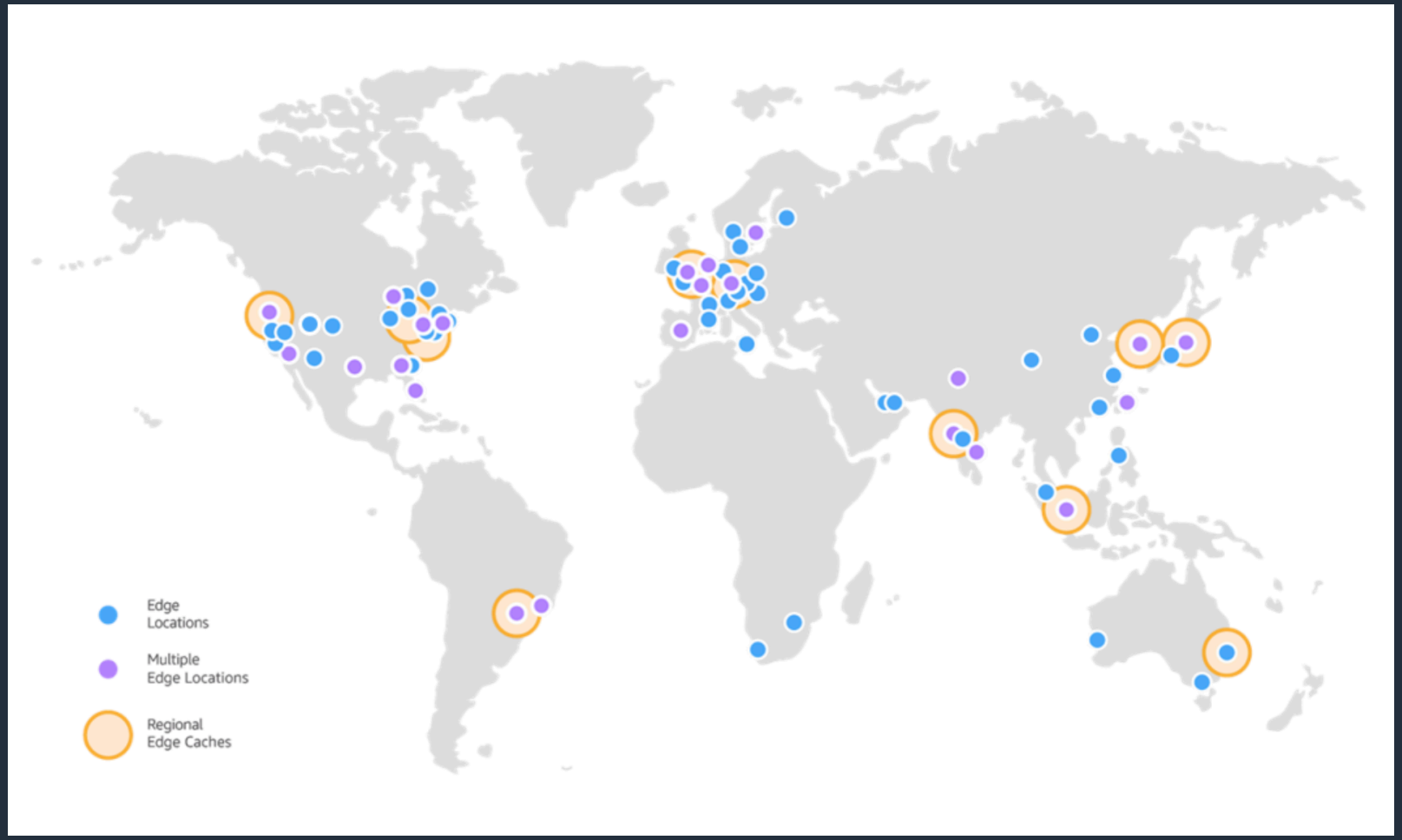
DigitalCloud TRAINING

Section 7: CloudFront Overview



DigitalCloud TRAINING

Section 7: CloudFront – Points of Presence

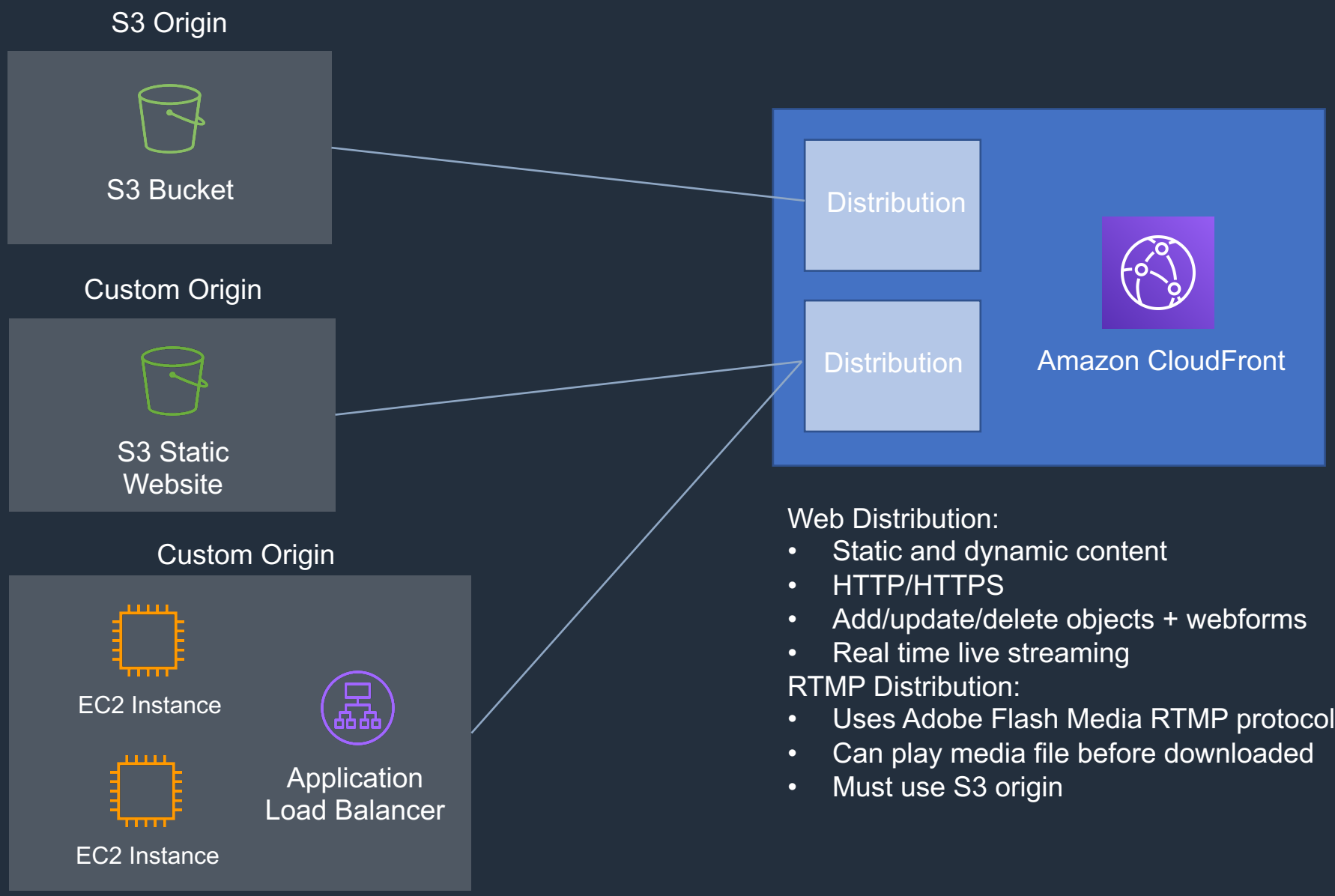


Points of Presence:

- 176 Edge Locations
- 11 Regional Edge Caches
- 69 cities
- 30 countries

DigitalCloud TRAINING

Section 7: CloudFront Distribution and Origins



Web Distribution:

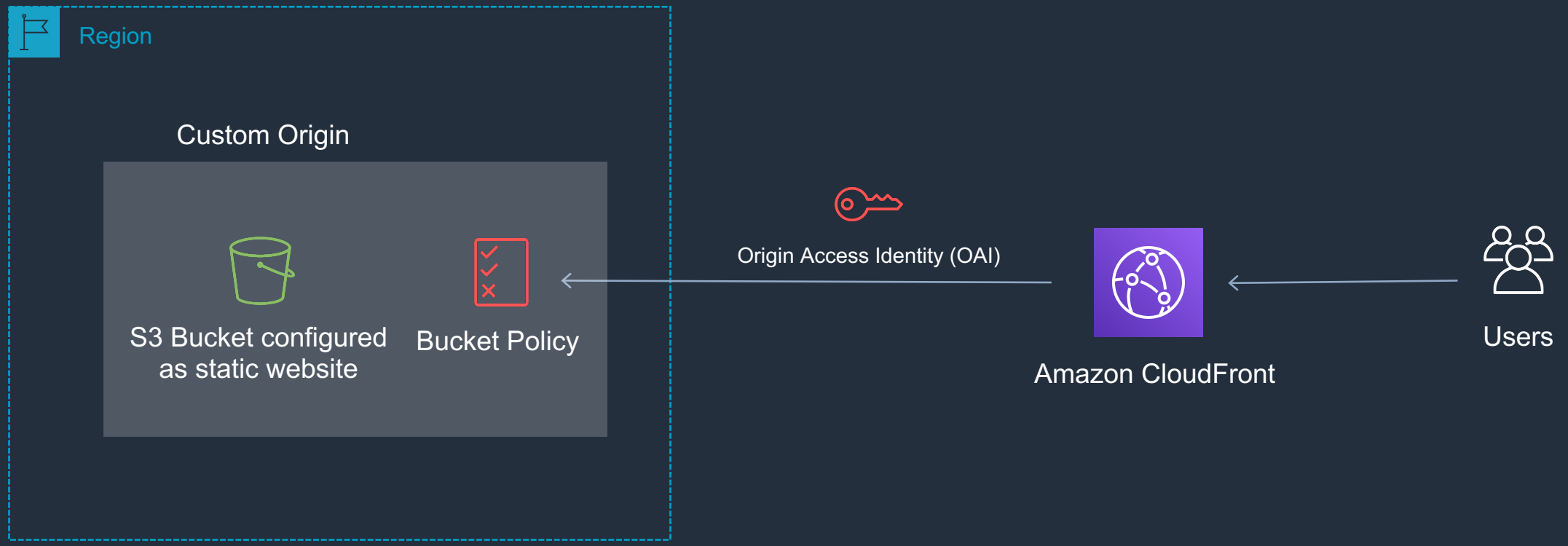
- Static and dynamic content
- HTTP/HTTPS
- Add/update/delete objects + webforms
- Real time live streaming

RTMP Distribution:

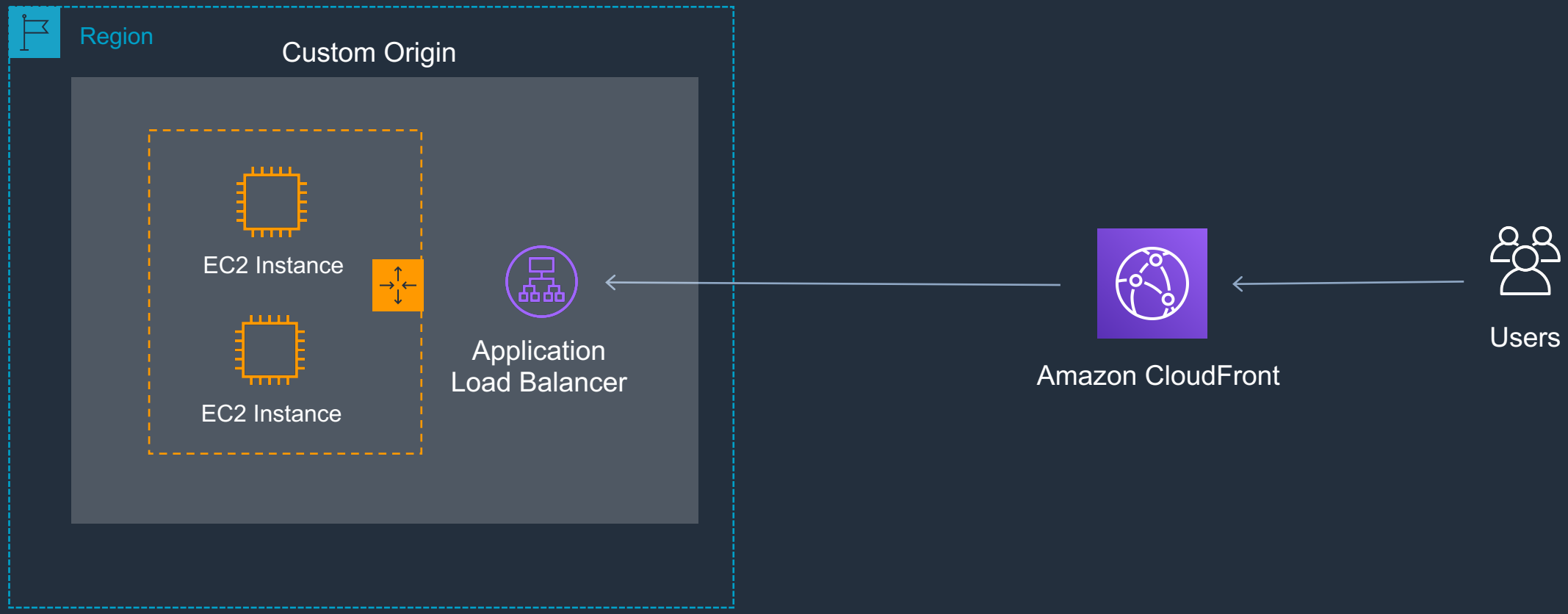
- Uses Adobe Flash Media RTMP protocol
- Can play media file before downloaded
- Must use S3 origin

DigitalCloud TRAINING

Section 7: CloudFront with S3 Static Website

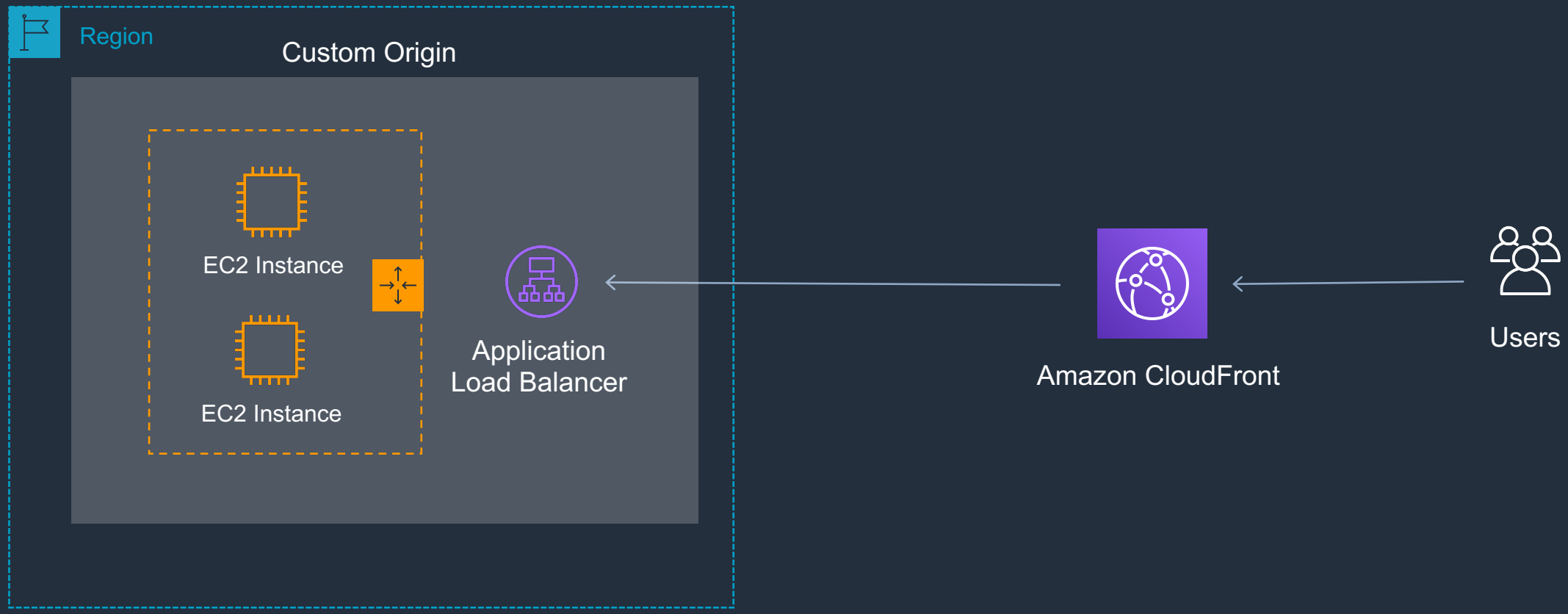


Section 7: CloudFront with ALB and EC2 Custom Origin

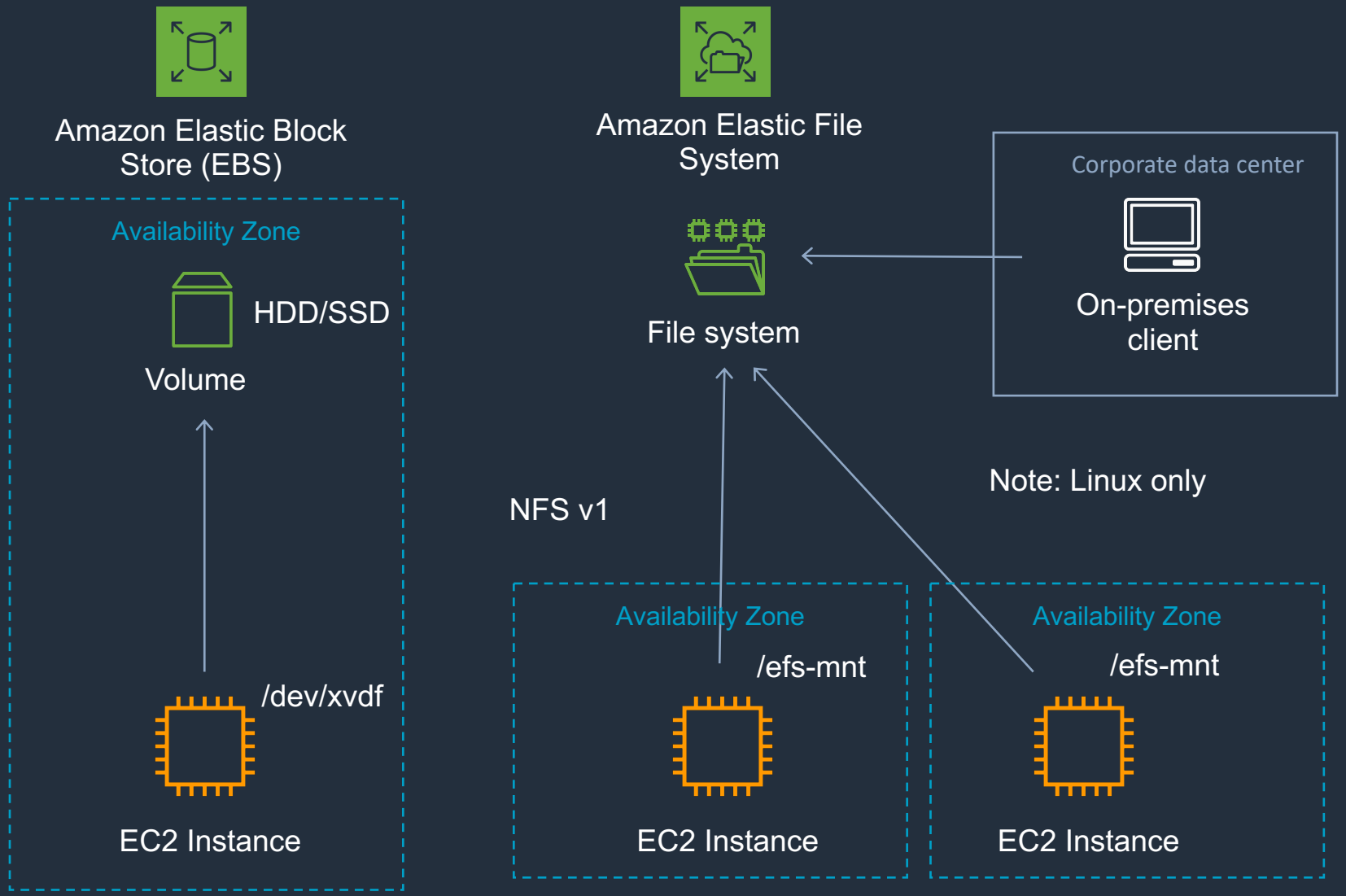


DigitalCloud TRAINING

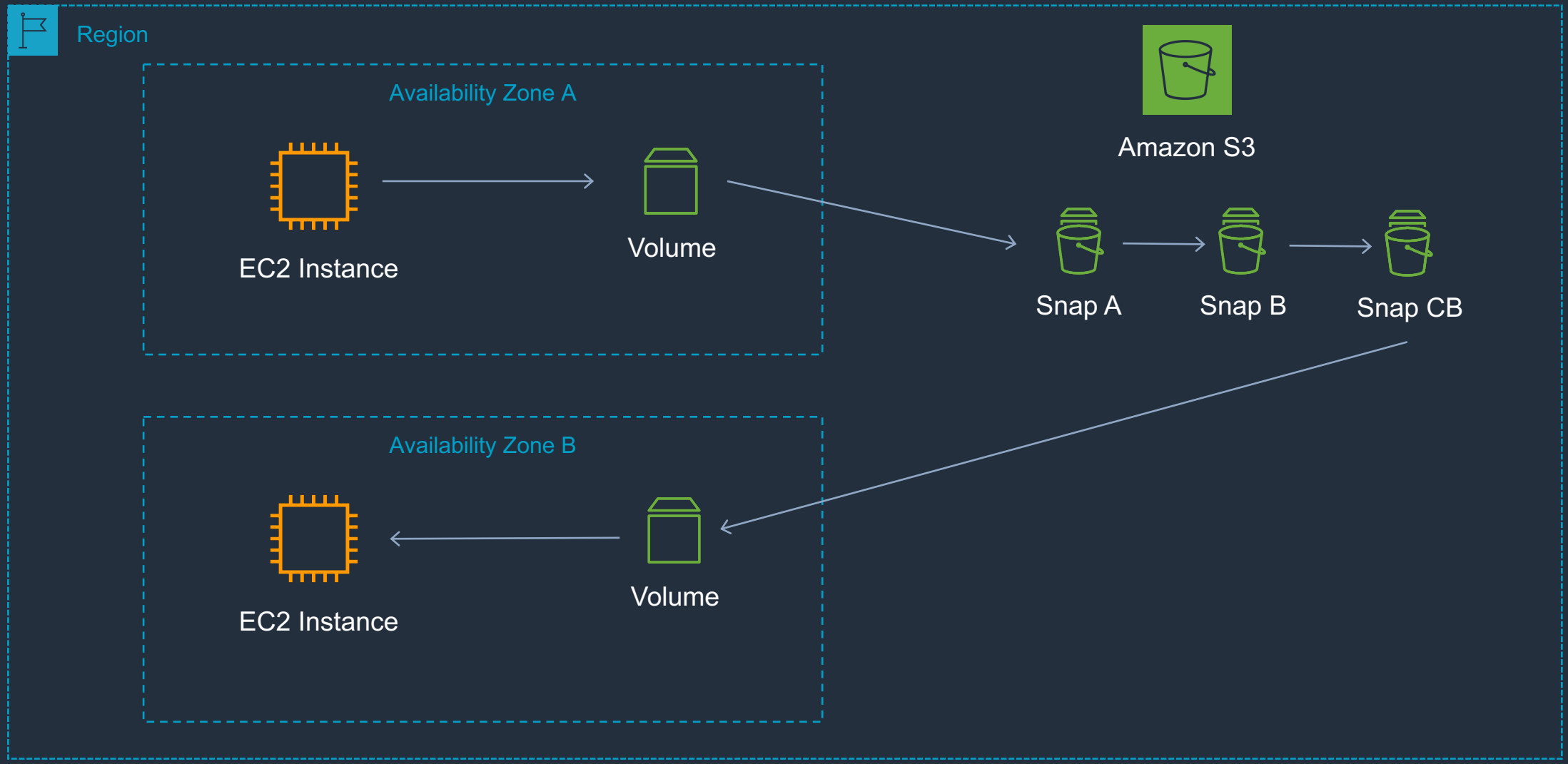
Section 7: CloudFront wit Lambda@Edge



Section 7: EBS and EFS Overview

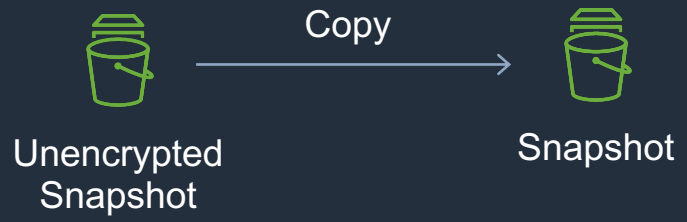


Section 7: EBS Snapshots



Section 7: EBS Copying, Sharing and Encryption

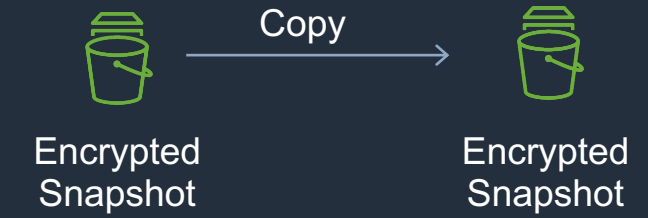
- Encryption state retained
- Same region



- Can be encrypted
- Can change regions

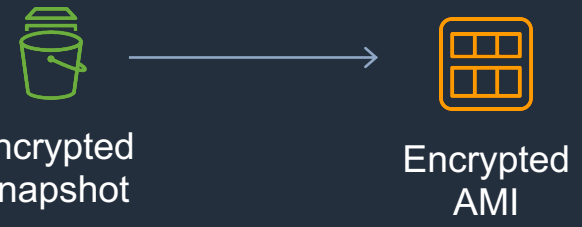
- Can be encrypted
- Can change AZ

- Cannot be encrypted
- Can be shared with other accounts
- Can be shared publicly

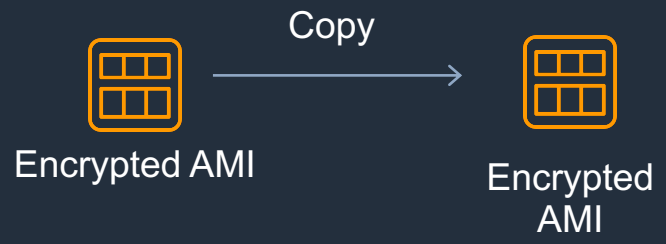


- Can change encryption key
- Can change regions

- Can change encryption key
- Can change AZ

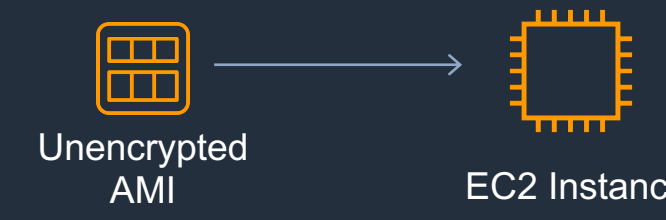


- Block devices remain encrypted
- Cannot be shared with other accounts if using AWS CMK
- Cannot be shared publicly



- Block devices remain encrypted
- Can change regions

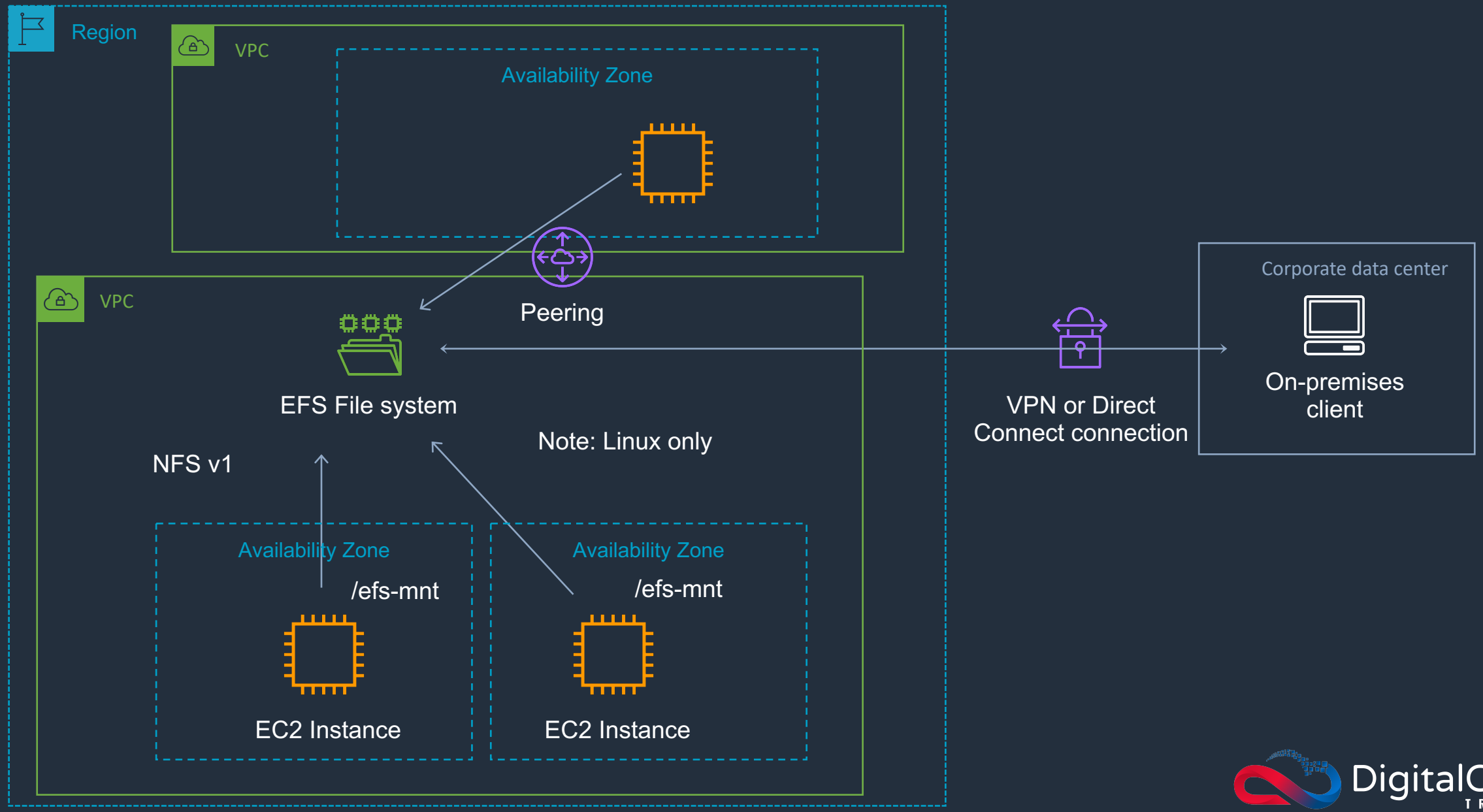
- Can change encryption key
- Can change AZ



- Can change encryption state
- Can change AZ

DigitalCloud TRAINING

Section 7: EFS Overview

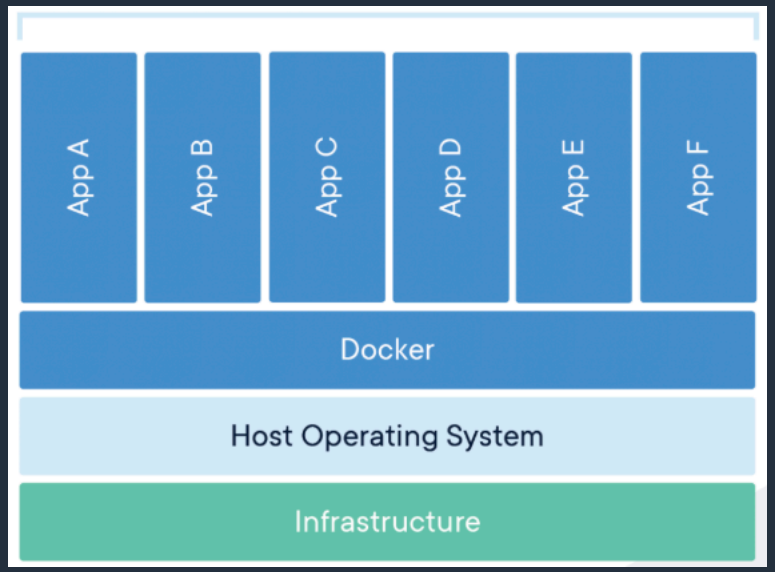


DigitalCloud TRAINING

Section 9: Elastic Container Services Overview

Definition of containers (from Docker):

A container is a standard unit of software that packages up code and all its dependencies so the application runs quickly and reliably from one computing environment to another. A Docker container image is a lightweight, standalone, executable package of software that includes everything needed to run an application: code, runtime, system tools, system libraries and settings.



Section 9: Elastic Container Services Overview

Amazon Elastic Container Service

Amazon Elastic Container Registry

Task Definition

```
{
"containerDefinitions": [
                {
                  "name": "wordpress",
                  "links": [
                    "mysql"
                ],
                "image": "wordpress",
                "essential": true,
                "portMappings": [
                  {
                    "containerPort": 80,
                    "hostPort": 80
                     }
                  ],
                  "memory": 500,
                  "cpu": 10
                }
```

Availability Zone

Availability Zone

ECS Cluster

ECS Service

Auto Scaling group

ECS Container instance

Task

Task

ECS Container instance

Task

Task

Registry

Image

Image

DigitalCloud TRAINING

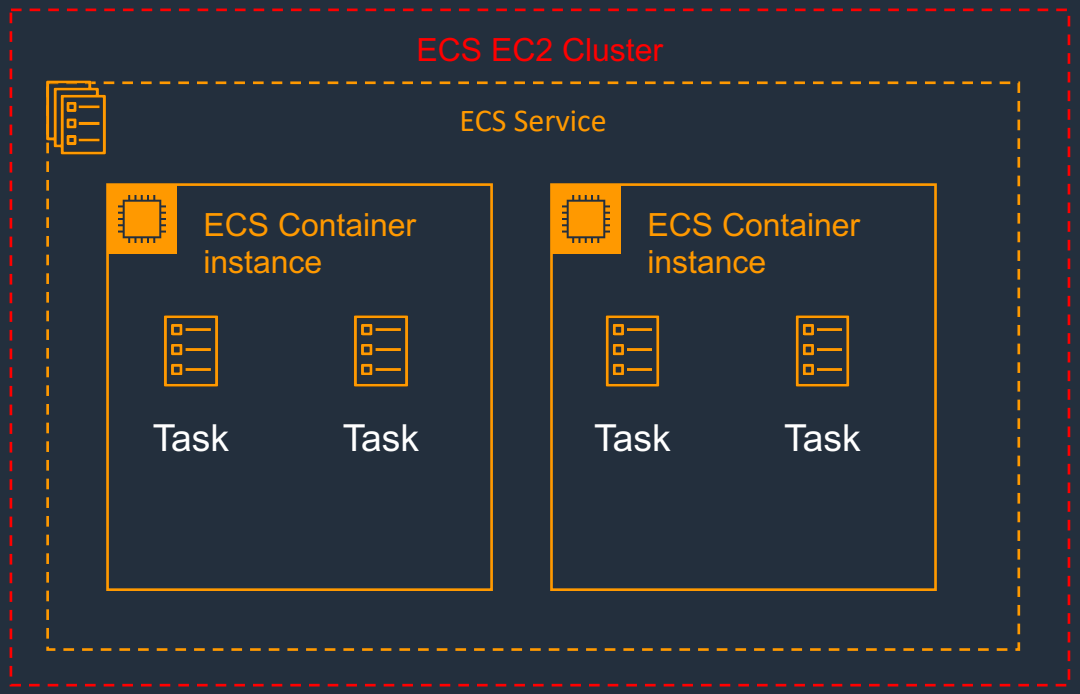
Section 9: ECS Terminology

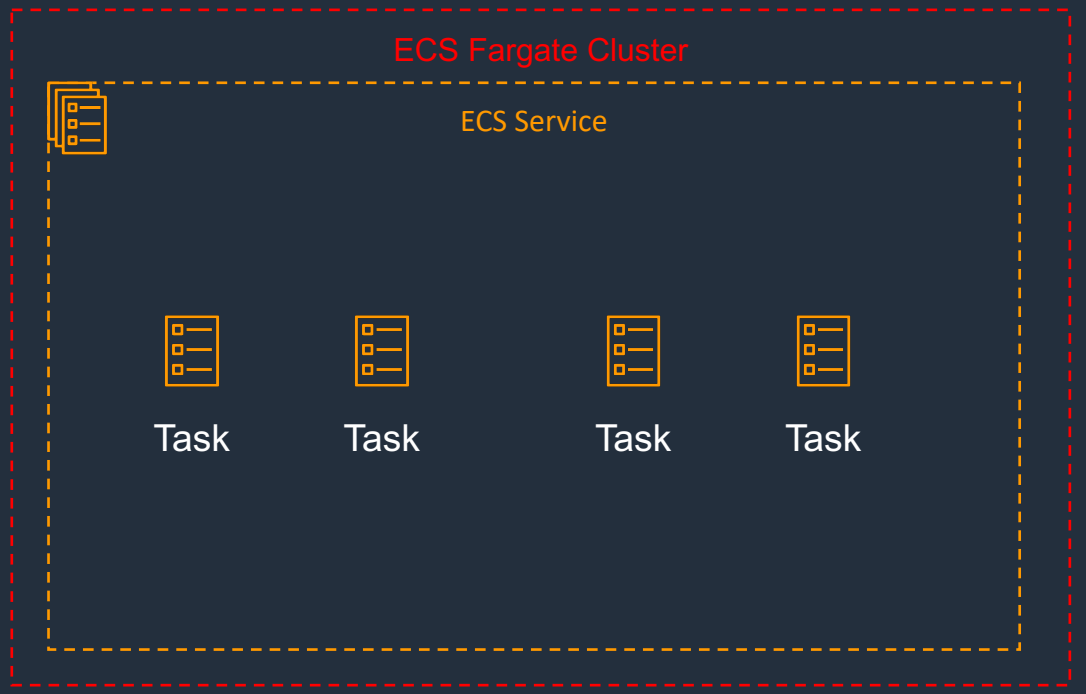
Elastic Container Service (ECS)	Description
Cluster	Logical grouping of EC2 instances
Container instance	EC2 instance running the the ECS agent
Task Definition	Blueprint that describes how a docker container should launch
Task	A running container using settings in a Task Definition
Service	Defines long running tasks – can control task count with Auto Scaling and attach an ELB

Section 9: Launch Types – EC2 and Fargate

Registry:
ECR, Docker Hub, Self-hosted

Registry:
ECR, Docker Hub





EC2 Launch Type

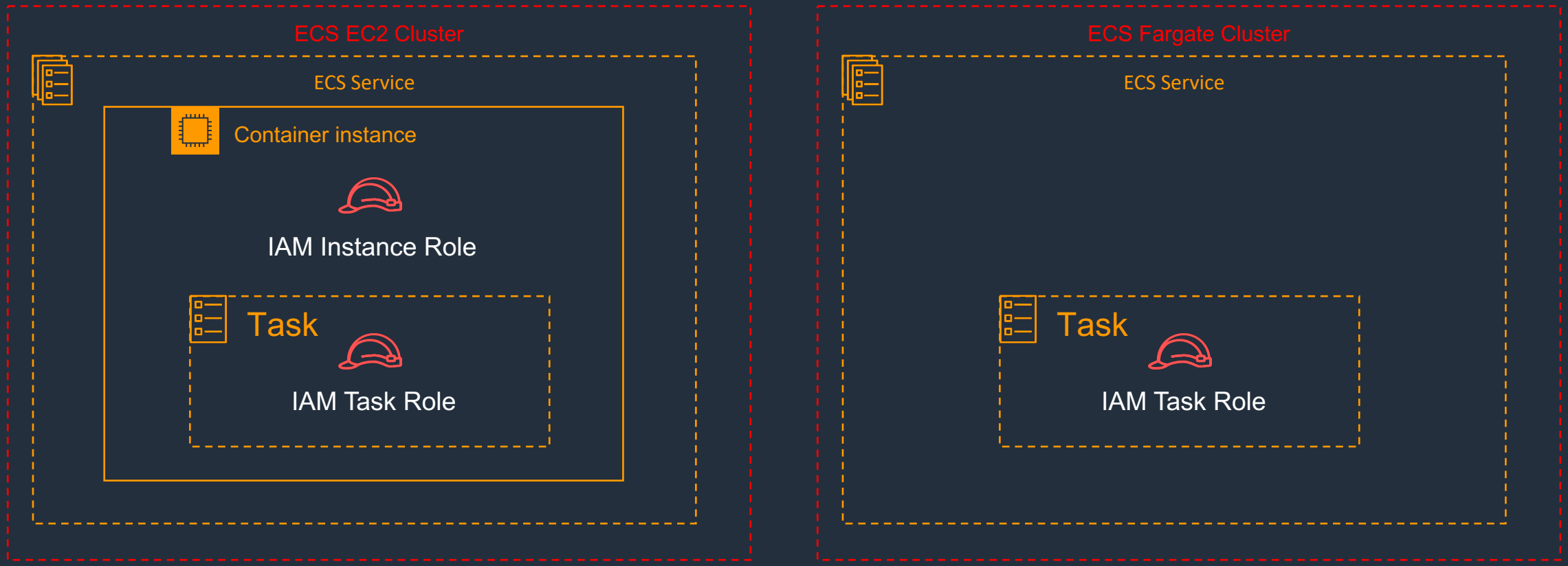
- You explicitly provision EC2 instances
- You're responsible for managing EC2 instances
- Charged per running EC2 instance
- EFS and EBS integration
- You handle cluster optimization
- More granular control over infrastructure

Fargate Launch Type

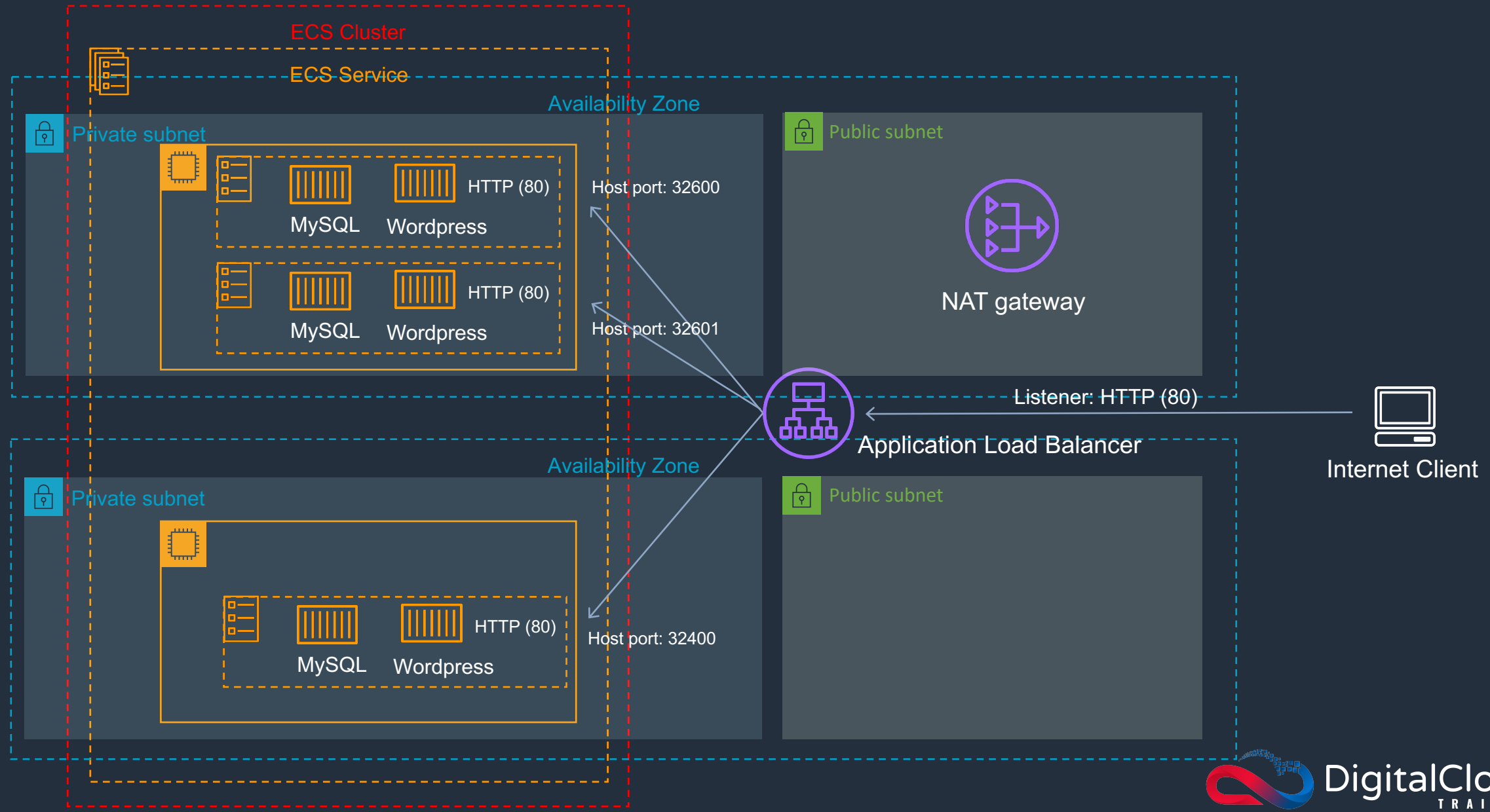
- Fargate automatically provisions resources
- Fargate provisions and manages compute
- Charged for running tasks
- No EFS and EBS integration
- Fargate handles cluster optimization
- Limited control, infrastructure is automated

DigitalCloud TRAINING

Section 9: IAM Roles

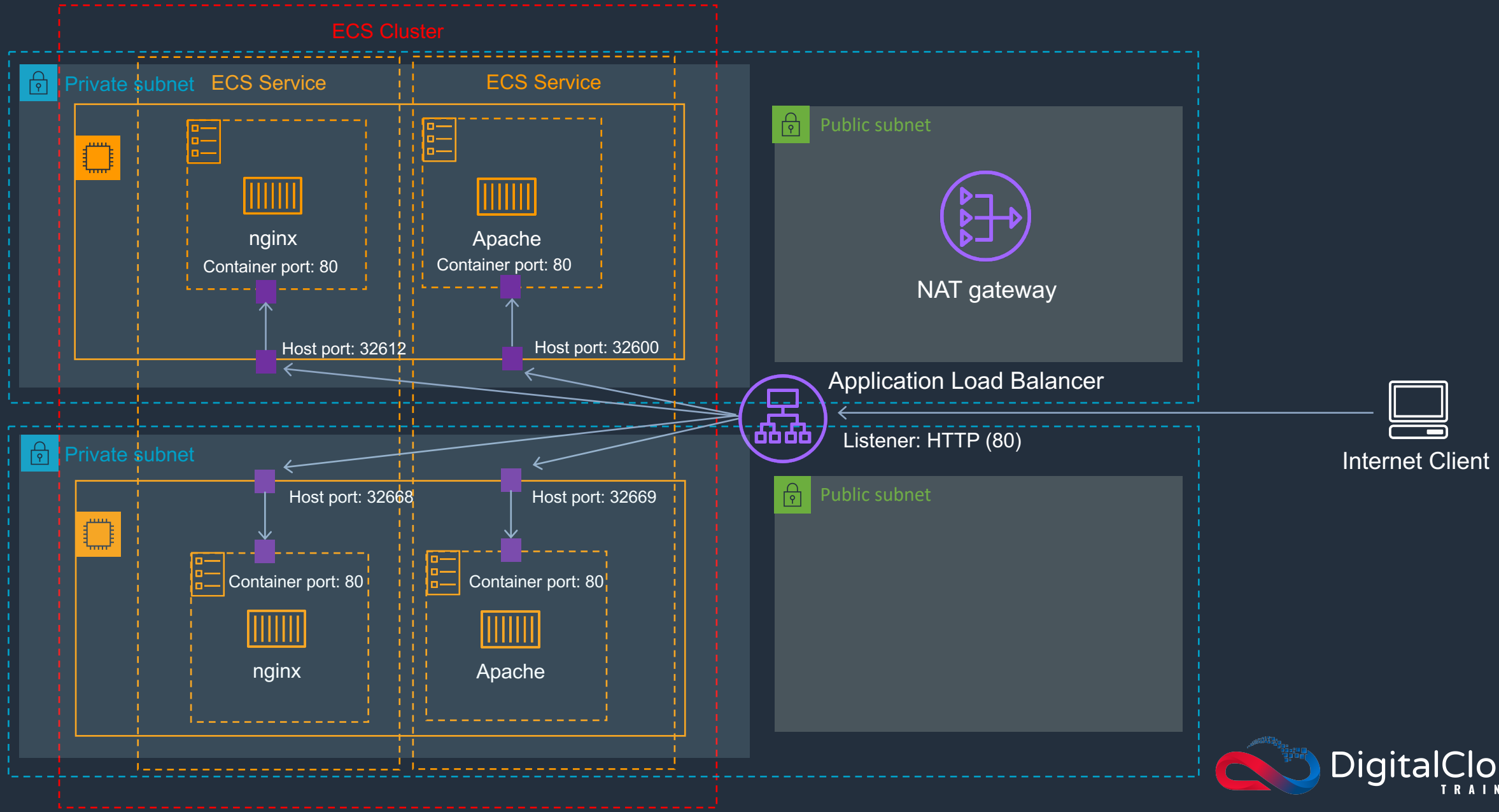


Section 9: ECS with Application Load Balancer (ALB)



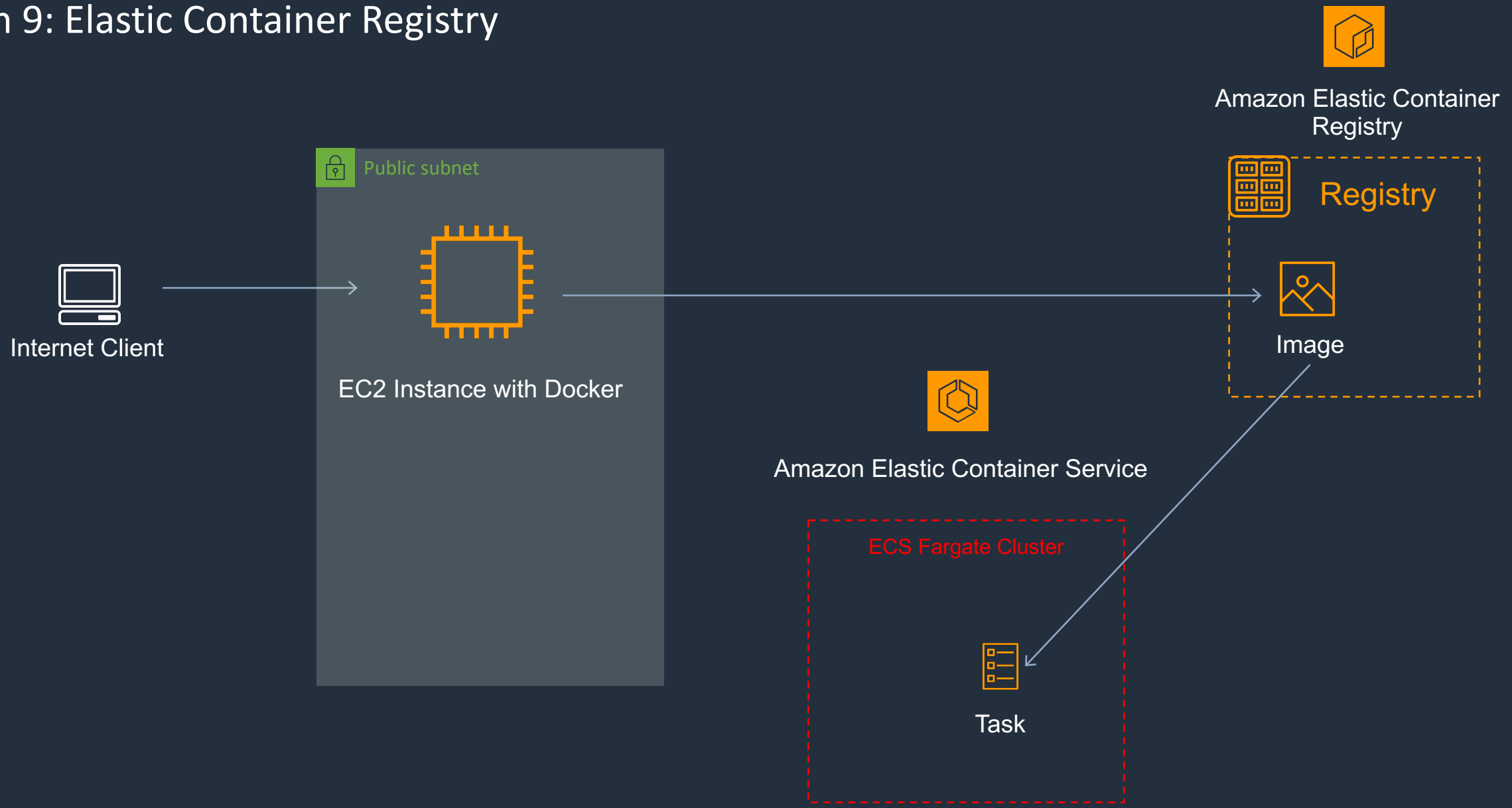
DigitalCloud TRAINING

Section 9: ECS with Application Load Balancer (ALB)

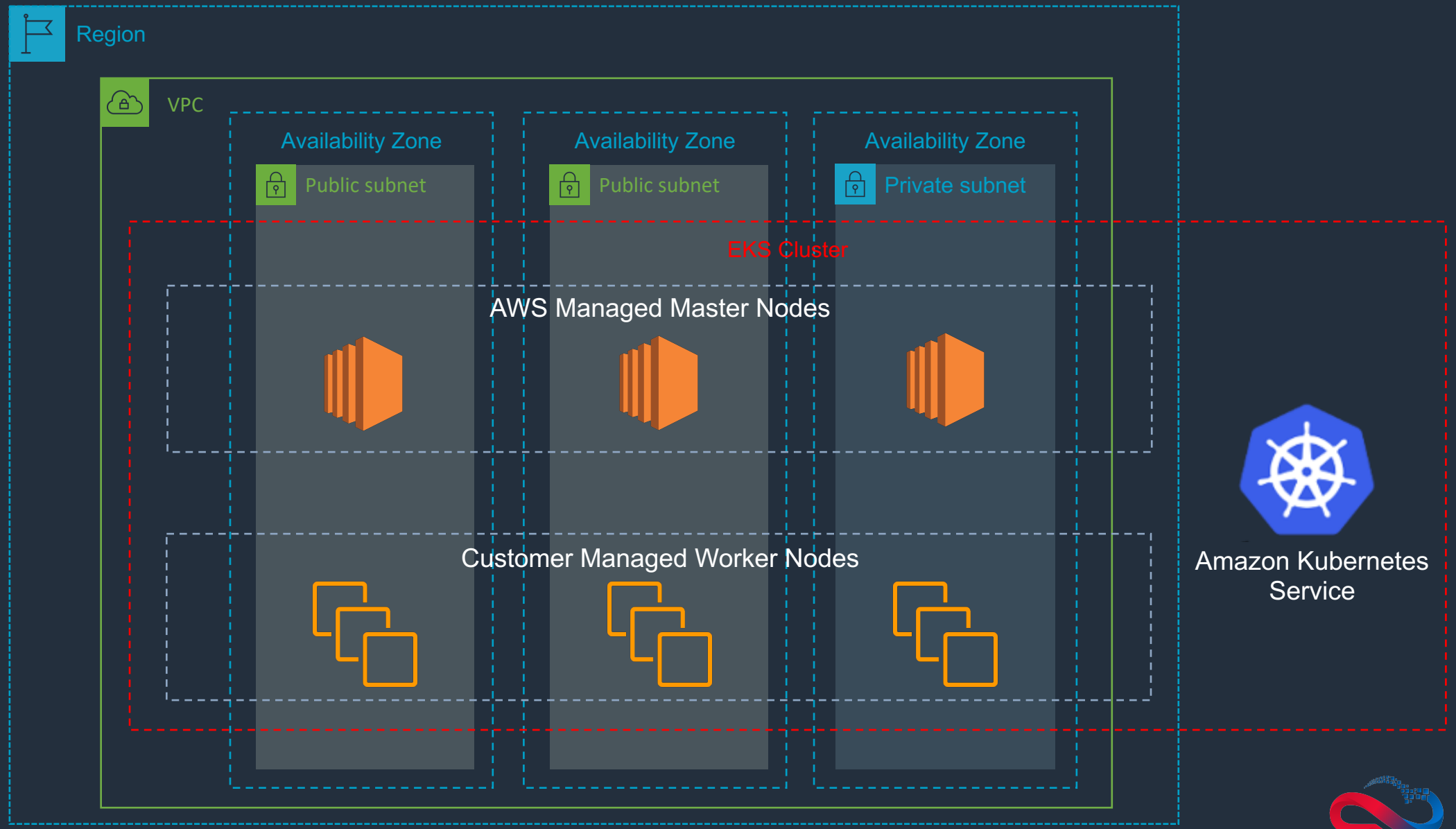


DigitalCloud TRAINING

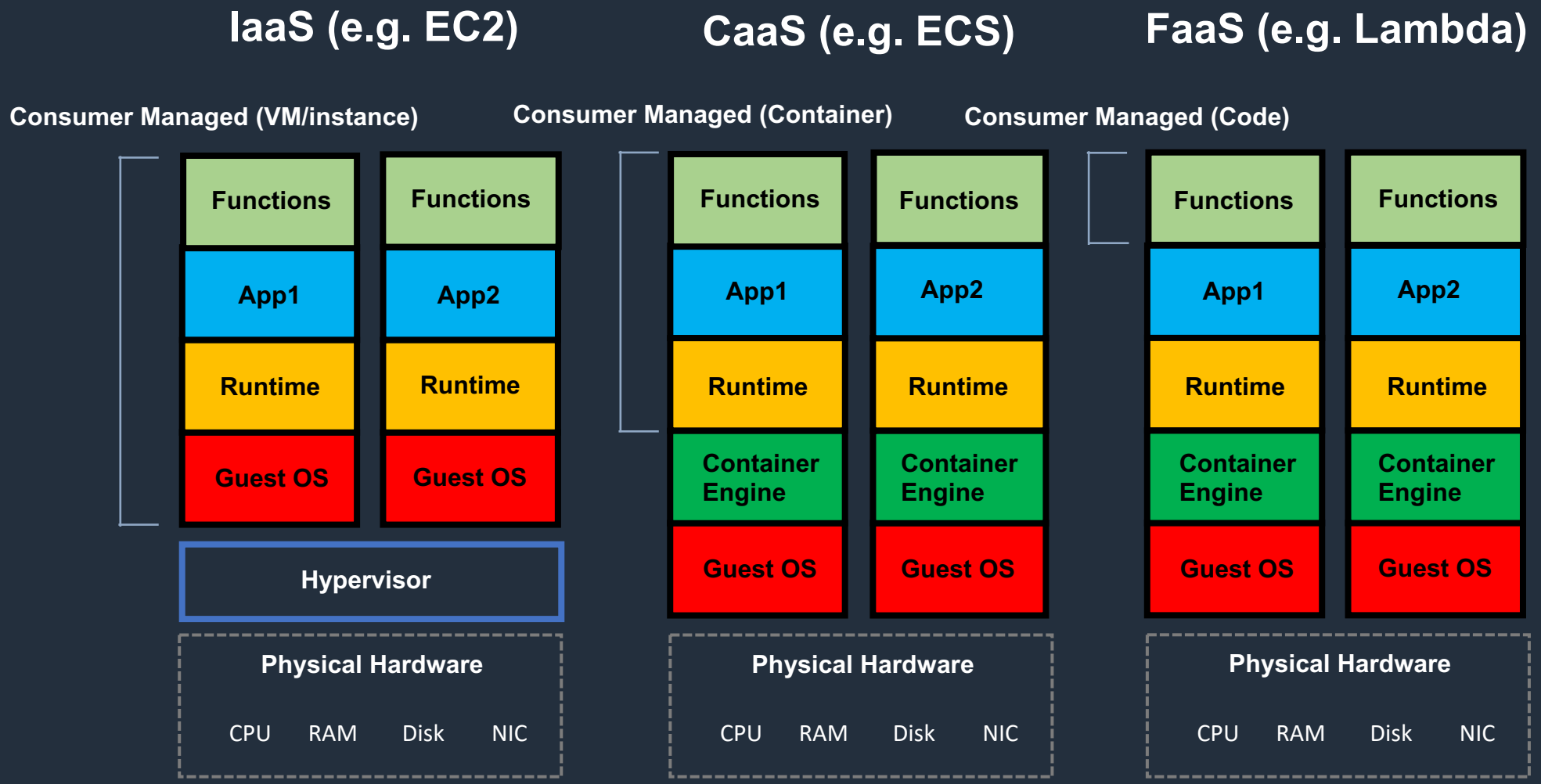
Section 9: Elastic Container Registry



Section 9: Elastic Kubernetes Service



Section 10: Comparing IaaS, CaaS, and FaaS

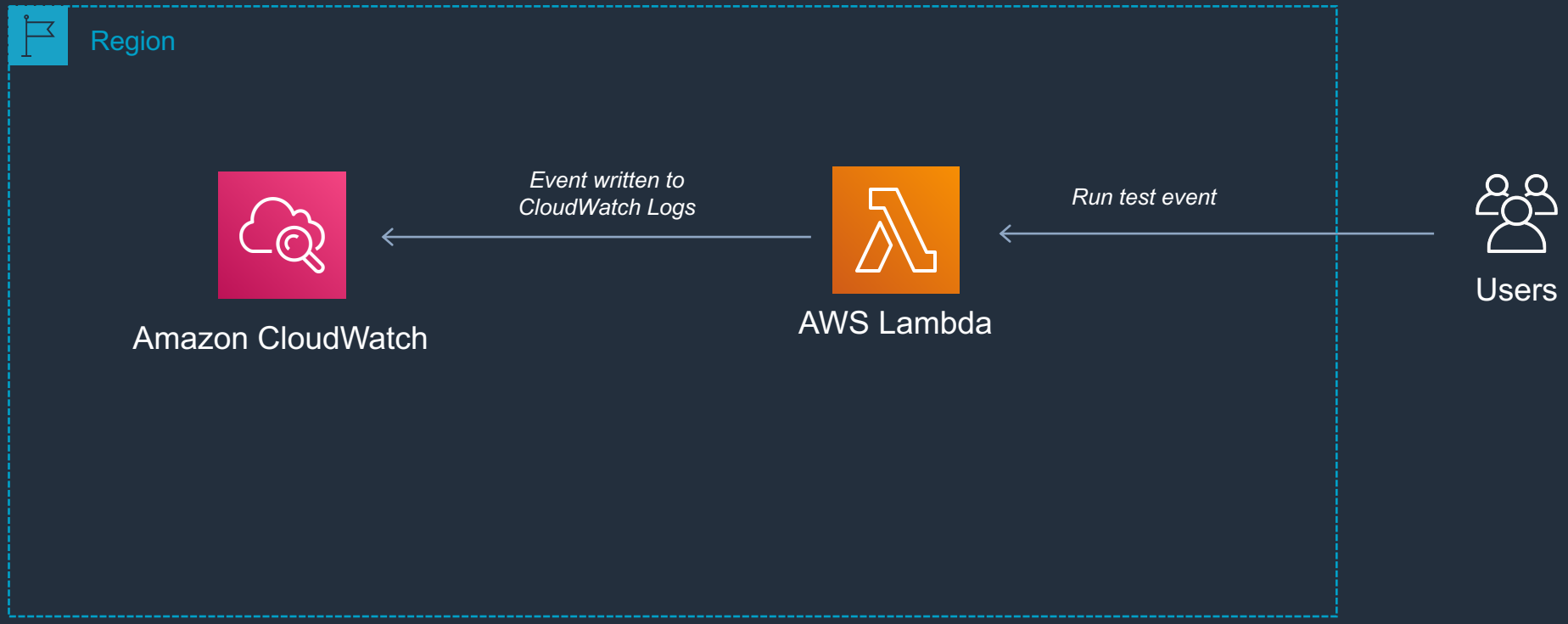


DigitalCloud TRAINING

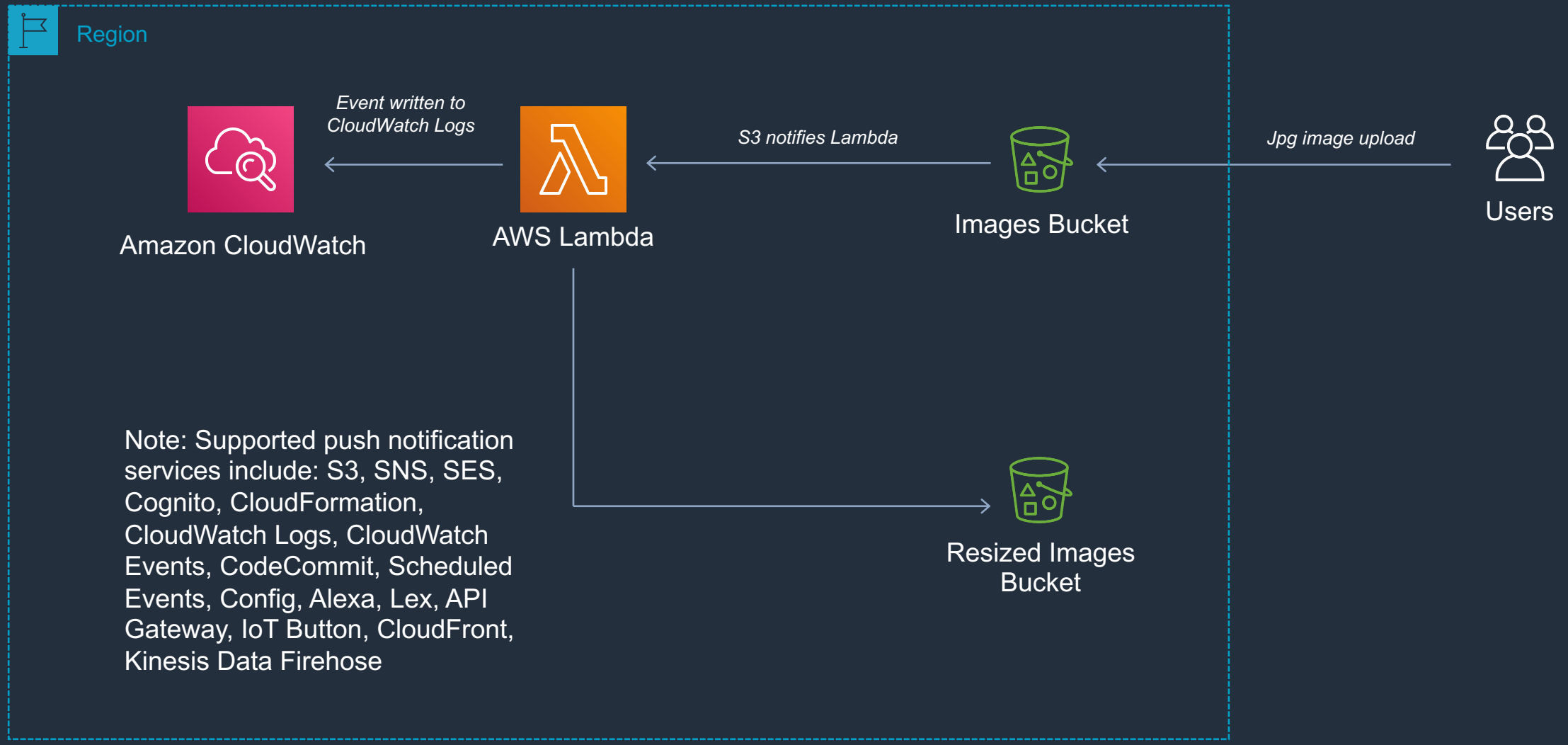
Section 10: Comparing Compute Options

EC2	ECS (EC2 Launch Type)	ECS (Fargate Launch Type)	Lambda
You manage the operating system	You manage container instance (EC2) and the containers (tasks)	You manage the containers (tasks)	You manage the code
Scale vertically – more CPU/Mem/HDD or scale horizontally (automatic) with Auto Scaling	Manually add container instances or use ECS Services and EC2 Auto Scaling	AWS scales the cluster automatically	Lambda automatically scales concurrent executions up to default limit (1000)
Use for traditional applications and long running tasks	Use for microservices and batch use cases where you need containers and need to retain management of underlying platform	Use for microservices and batch use cases	Use for ETL, infrastructure automation, data validation, mobile backends
No timeout issues	No timeout issues	No timeout issues	Limited to 900 seconds execution time for single execution (3 second default)
Pay for instance run time based on family/type	Pay for instance run time based on family/type	Pay for container run time based on allocated resources	Pay only for execution time based on memory allocation

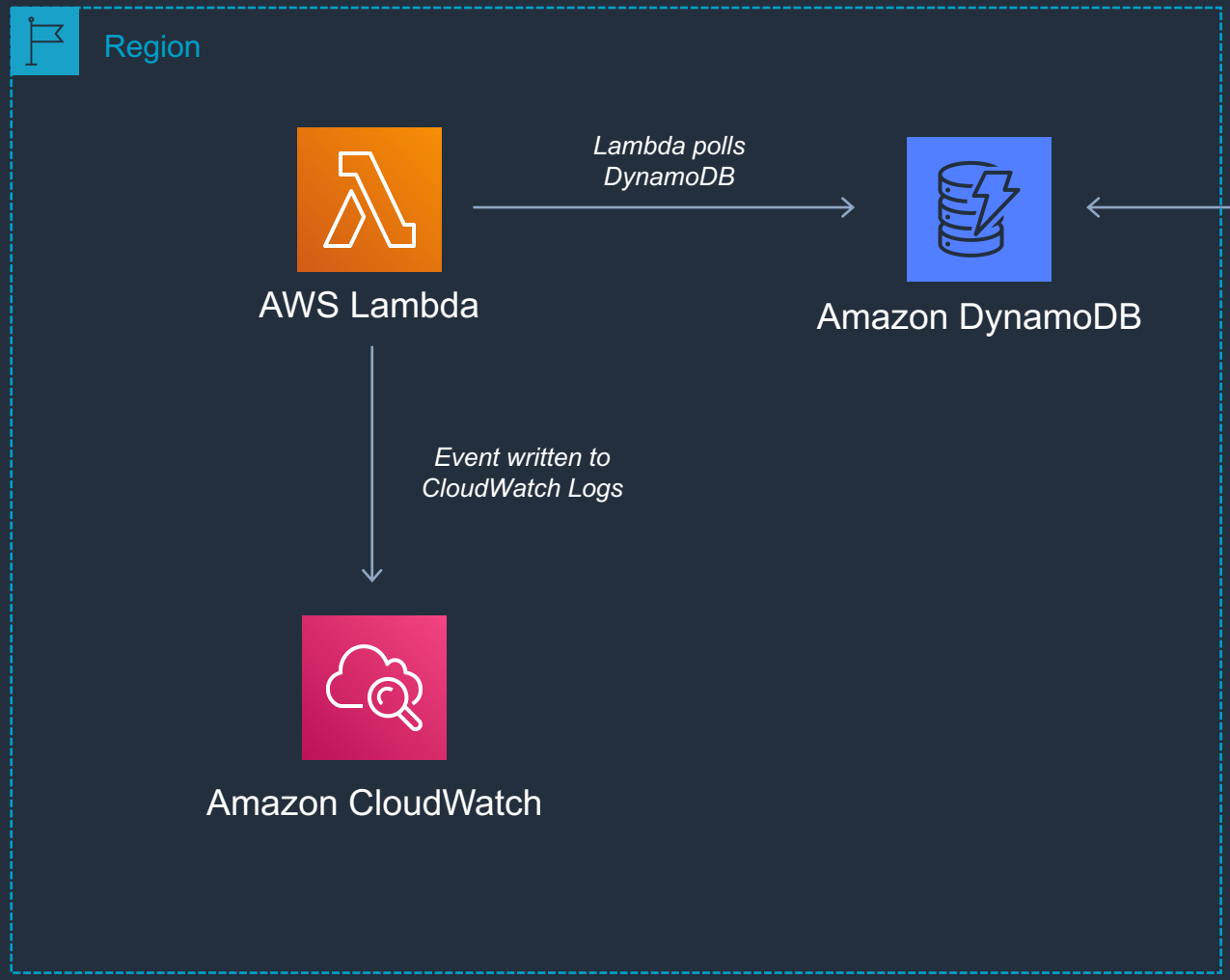
Section 10: AWS Lambda – Hello World



Section 10: AWS Lambda – S3 Event Source Mapping

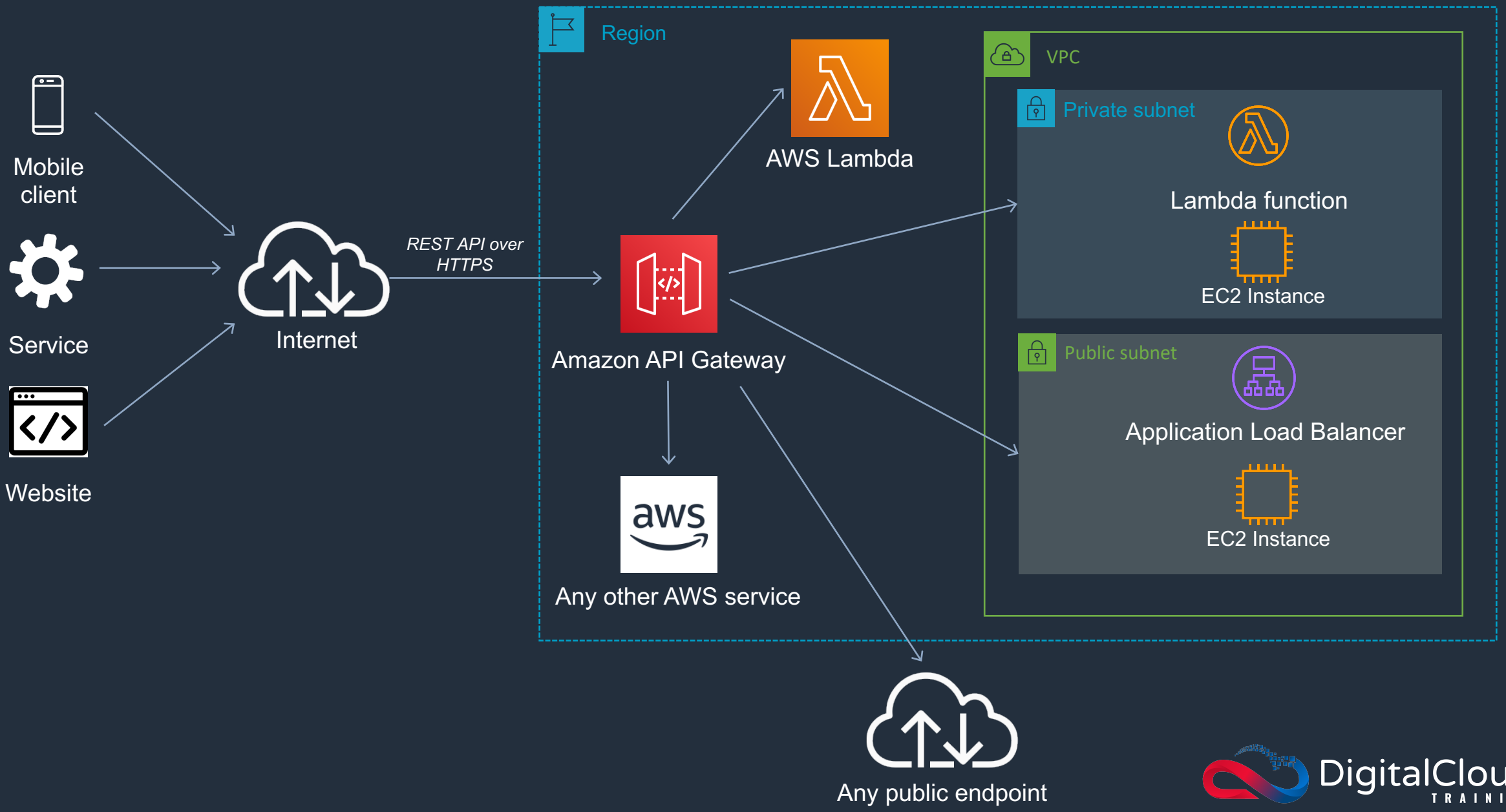


Section 10: AWS Lambda – DynamoDB Event Source Mapping



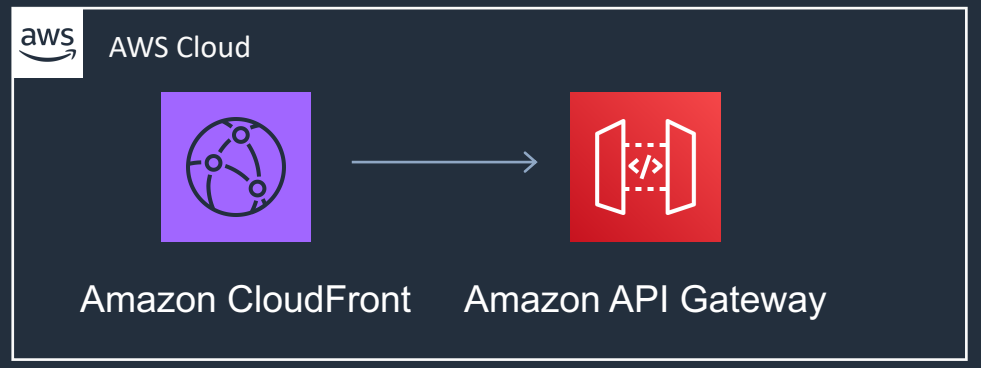
Note: Supported poll-based services are DynamoDB, Kinesis, and SQS

Section 10: API Gateway Overview



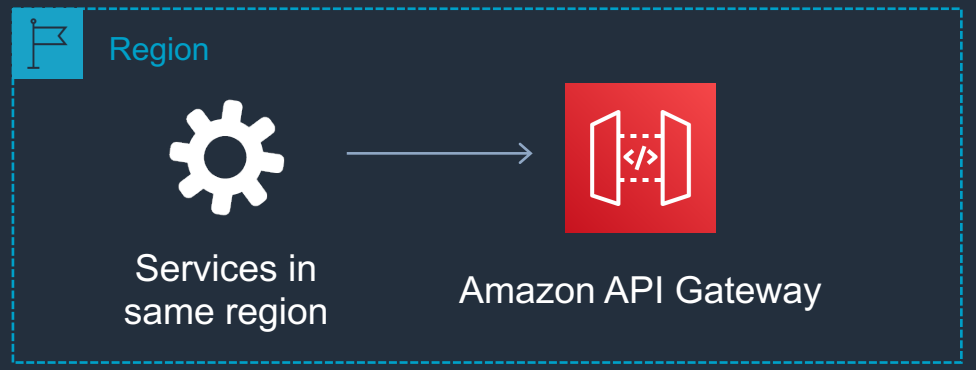
Section 10: API Gateway Endpoint Types

Edge-optimized endpoint:



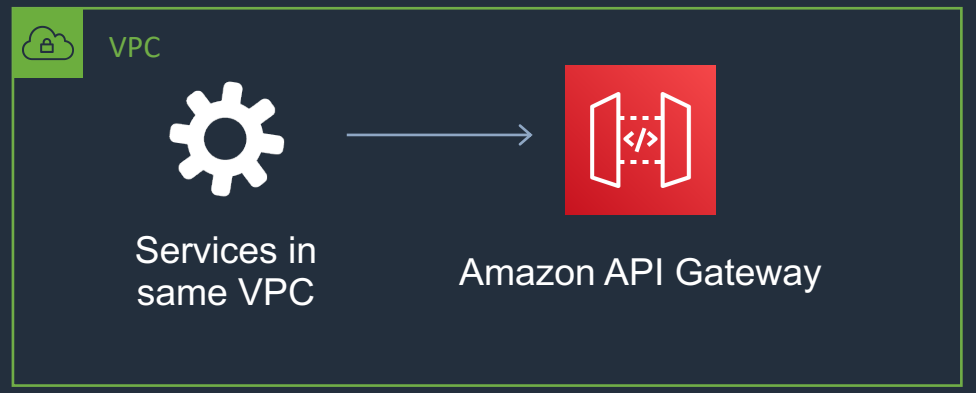
Key benefits:
- Reduced latency for requests from around the world

Regional endpoint:



Key benefits:
- Reduced latency for requests that originate in the same region
- Can also configure your own CDN and protect with WAF

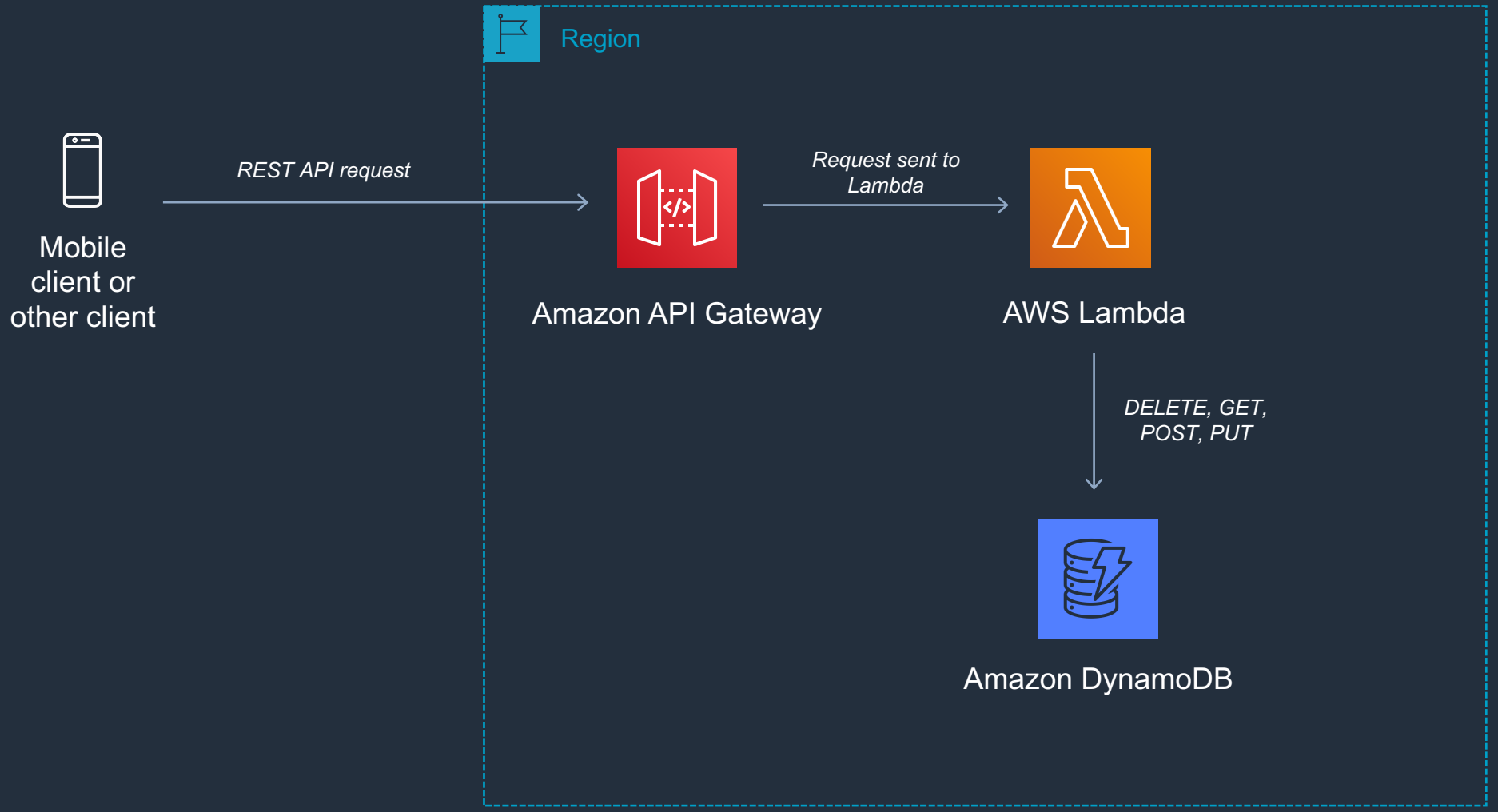
Private endpoint:



Key benefits:
- Securely expose your REST APIs only to other services within your VPC or connect via Direct Connect

DigitalCloud TRAINING

Section 10: AWS Lambda – Microservice with Lambda, API Gateway and DynamoDB



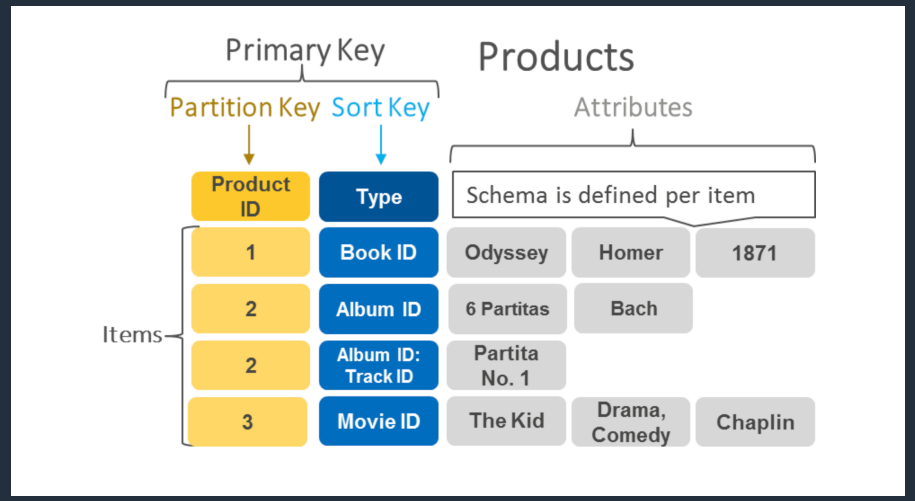
Section 11: Database Types – Relational vs Non-Relational

Key differences are how data are ***managed*** and how data are ***stored***

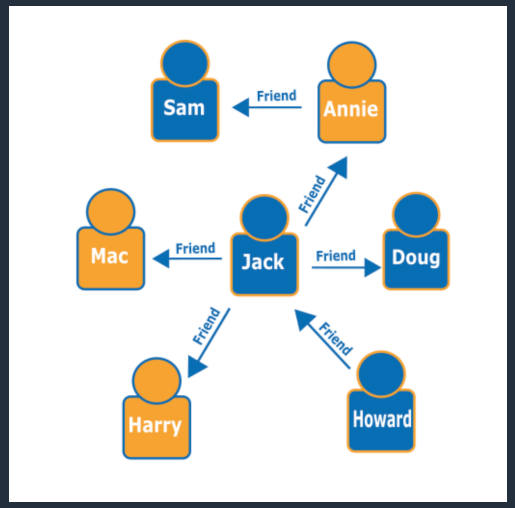
Relational	Non-Relational
Organized by tables, rows and columns	Varied data storage models
Rigid schema (SQL)	Flexible schema (NoSQL) – data stored in key-value pairs, columns, documents or graphs
Rules enforced within database	Rules can be defined in application code (outside database)
Typically scaled vertically	Scales horizontally
Supports complex queries and joins	Unstructured, simple language that supports any kind of schema
ACID (Atomicity, Consistency, Isolation, Durability) compliance typically enforced	Performance is typically prioritised
Amazon RDS, Oracle, MySQL, IBM DB2, PostgreSQL	Amazon DynamoDB, MongoDB, Redis, Neo4j

Section 11: Types of Non-Relational DB

Key-value – e.g. Amazon DynamoDB



Graph – e.g. Amazon Neptune



Document – e.g. MongoDB

```json
[
    {
        "year" : 2013,
        "title" : "Turn It Down, Or Else!",
        "info" : {
            "directors" : [ "Alice Smith", "Bob Jones"],
            "release_date" : "2013-01-18T00:00:00Z",
            "rating" : 6.2,
            "genres" : ["Comedy", "Drama"],
            "image_url" : "http://ia.media-imdb.com/images/N/O9ERWAU7FS797AJ7LU8HN09AMUP908RLlo5JF90EWR7LJKQ7@@._V1_SX400_.jpg",
            "plot" : "A rock band plays their music at high volumes, annoying the neighbors.",
            "actors" : ["David Matthewman", "Jonathan G. Neff"]
        }
    },
    {
        "year": 2015,
        "title": "The Big New Movie",
        "info": {
            "plot": "Nothing happens at all.",
            "rating": 0
        }
    }
]
```

DigitalCloud TRAINING

Section 11: Database Types – Operational vs Analytical

Key differences are ***use cases*** and how the database is ***optimized***

Operational / transactional	Analytical
Online Transaction Processing (OLTP)	Online Analytics Processing (OLAP) – the source data comes from OLTP DBs
Production DBs that process transactions. E.g. adding customer records, checking stock availability (INSERT, UPDATE, DELETE)	Data warehouse. Typically, separated from the customer facing DBs. Data is extracted for decision making
Short transactions and simple queries	Long transactions and complex queries
Relational examples: Amazon RDS, Oracle, IBM DB2, MySQL	Relational examples: Amazon RedShift, Teradata, HP Vertica
Non-relational examples: MongoDB, Cassandra, Neo4j, HBase	Non-relational examples: Amazon EMR, MapReduce

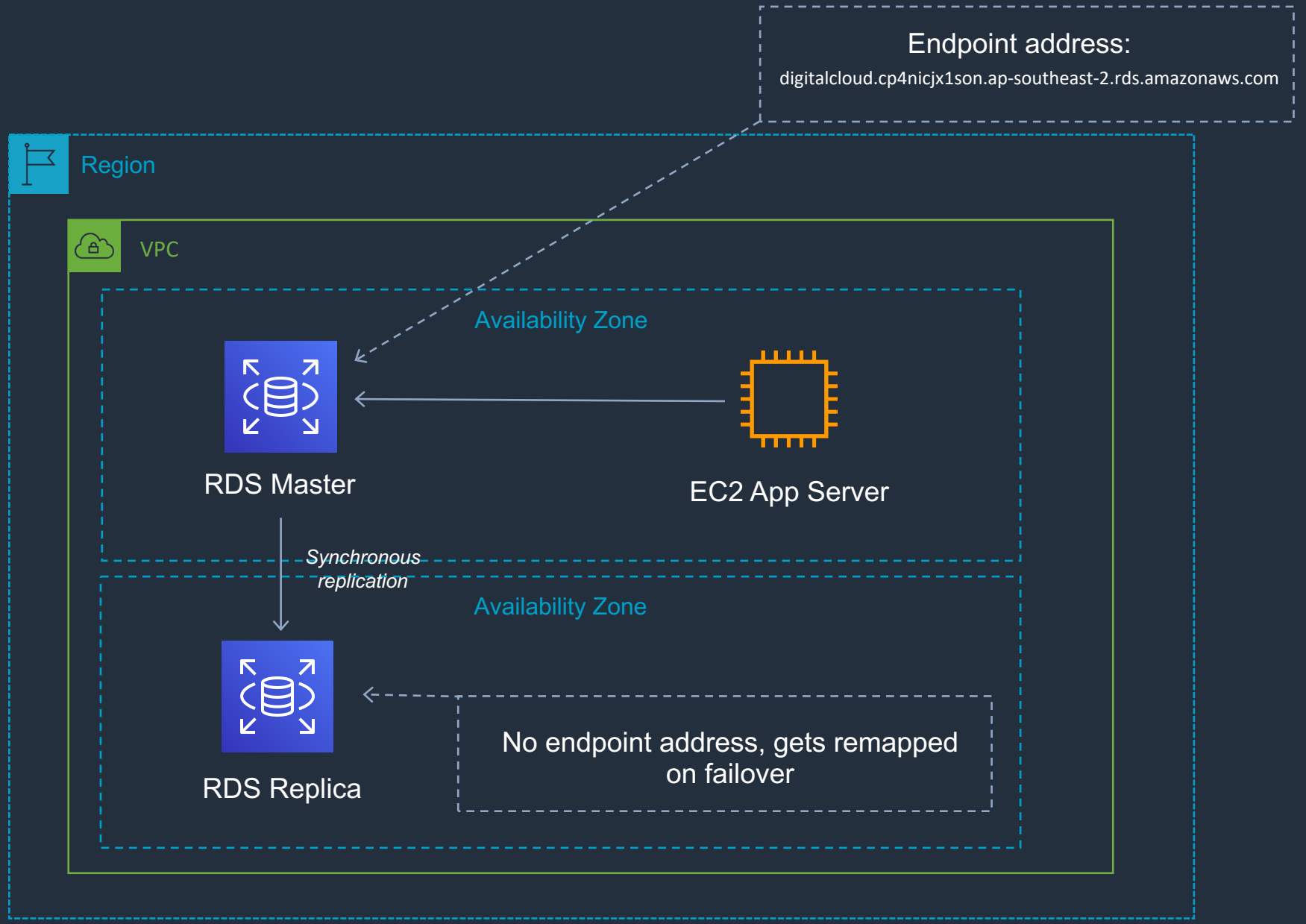
Section 11: Databases –Architecture Discussion

Data Store	When to Use
Database on EC2	• Full control over instance and database • Preferred DB not available under RDS
Amazon RDS	• Need traditional relational database for OLTP • Your data is well-formed and structured
Amazon DynamoDB	• Name/value pair data • Unpredictable data structure • In-memory performance with persistence • High I/O needs • Require dynamic scaling
Amazon RedShift	• Data warehouse for large volumes of aggregated data • Primarily OLAP workloads
Amazon Neptune	• Relationships between objects are of high value
Amazon ElastiCache	• Fast temporary storage for small amounts of data • Highly volatile data (non-persistent)

Section 11: Amazon RDS – Multi-AZ and Read Replicas

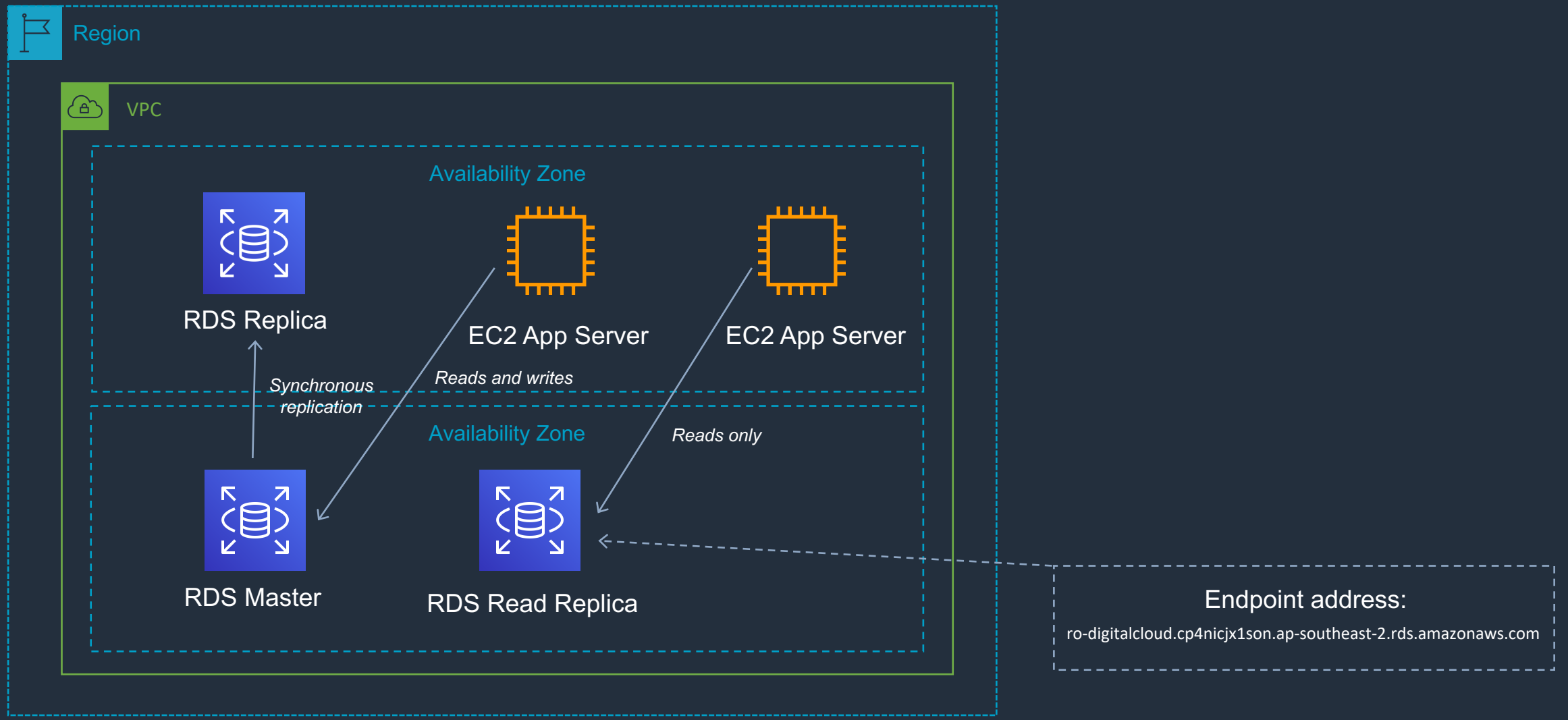
Multi-AZ Deployments	Read Replicas
Synchronous replication – highly durable	Asynchronous replication – highly scalable
Only database engine on primary instance is active	All read replicas are accessible and can be used for read scaling
Automated backups are taken from standby	No backups configured by default
Always span two Availability Zones within a single Region	Can be within an Availability Zone, Cross-AZ, or Cross-Region
Database engine version upgrades happen on primary	Database engine version upgrade is independent from source instance
Automatic failover to standby when a problem is detected	Can be manually promoted to a standalone database instance

Section 11: Amazon RDS Multi-AZ



DigitalCloud TRAINING

Section 11: Amazon RDS Read Replicas



DigitalCloud TRAINING

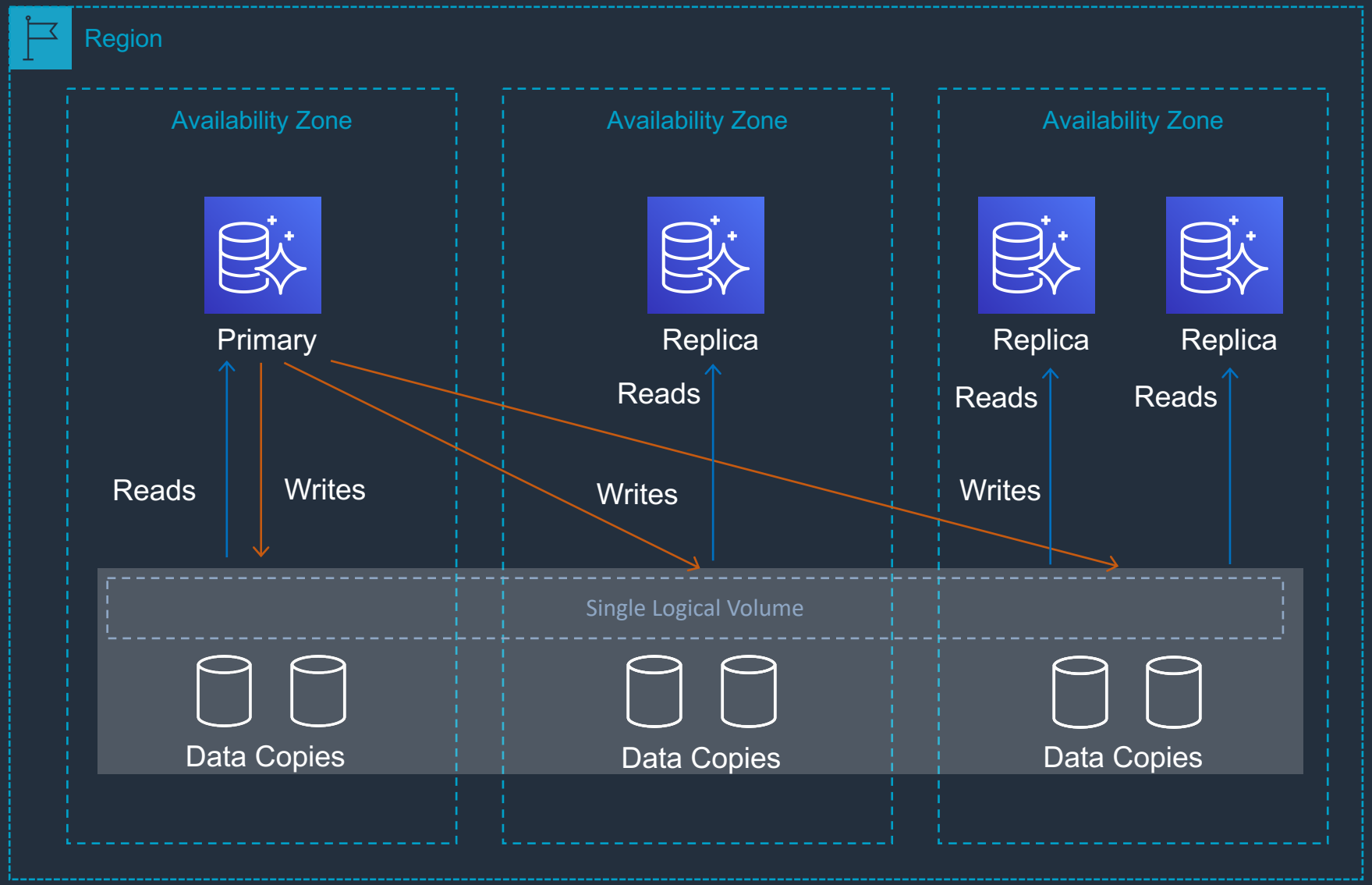
Section 11: Amazon RDS Aurora Key Features

Aurora Feature	Benefit
High performance and scalability	Offers high performance, self-healing storage that scales up to 64TB, point-in-time recovery and continuous backup to S3
DB compatibility	Compatible with existing MySQL and PostgreSQL open source databases
Aurora Replicas	In-region read scaling and failover target – up to 15 (can use Auto Scaling)
MySQL Read Replicas	Cross-region cluster with read scaling and failover target – up to 5 (each can have up to 15 Aurora Replicas)
Global Database	Cross-region cluster with read scaling (fast replication / low latency reads). Can remove secondary and promote
Multi-Master	Scales out writes within a region. In preview currently and will not appear on the exam
Serverless	On-demand, autoscaling configuration for Amazon Aurora - does not support read replicas or public IPs (can only access through VPC or Direct Connect - not VPN)

Section 11: Amazon RDS Aurora Replicas

Feature	Aurora Replica	MySQL Replica
Number of replicas	Up to 15	Up to 5
Replication type	Asynchronous (milliseconds)	Asynchronous (seconds)
Performance impact on primary	Low	High
Replica location	In-region	Cross-region
Act as failover target	Yes (no data loss)	Yes (potentially minutes of data loss)
Automated failover	Yes	No
Support for user-defined replication delay	No	Yes
Support for different data or schema vs. primary	No	Yes

Section 11: Aurora Fault Tolerance and Aurora Replicas

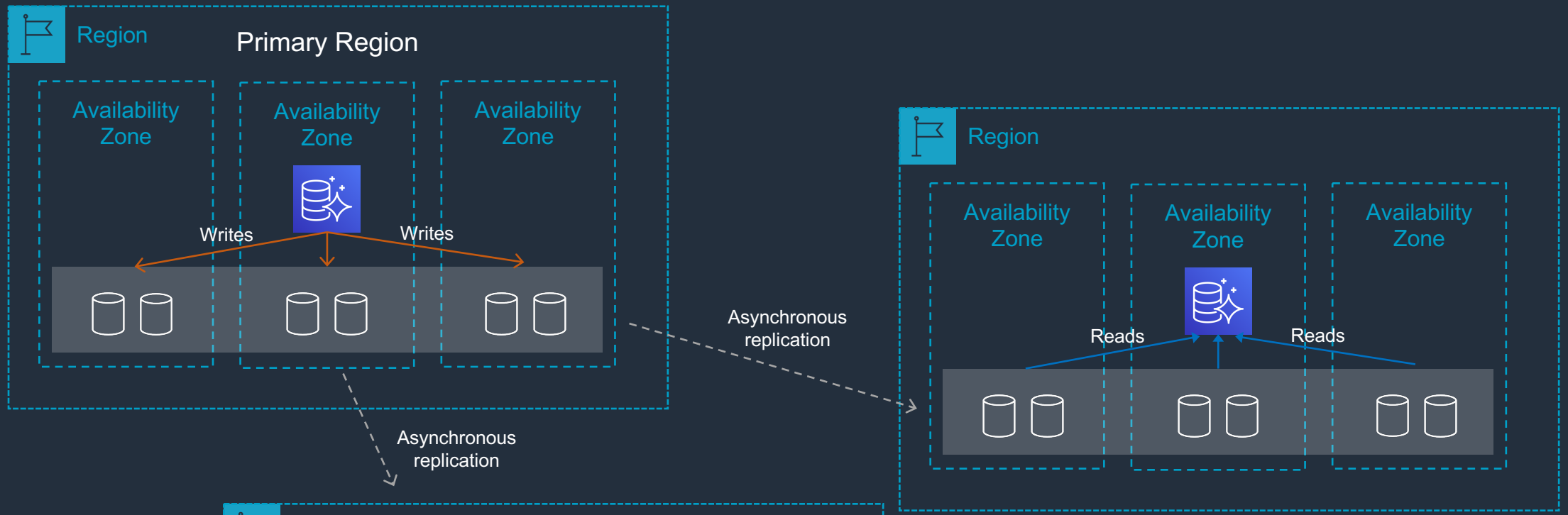


Aurora Fault Tolerance

- Fault tolerance across 3 AZs
- Single logical volume
- Aurora Replicas scale-out read requests
- Up to 15 Aurora Replicas with sub-10ms replica lag
- Aurora Replicas are independent endpoints
- Can promote Aurora Replica to be a new primary or create new primary
- Set priority (tiers) on Aurora Replicas to control order of promotion
- Can use Auto Scaling to add replicas

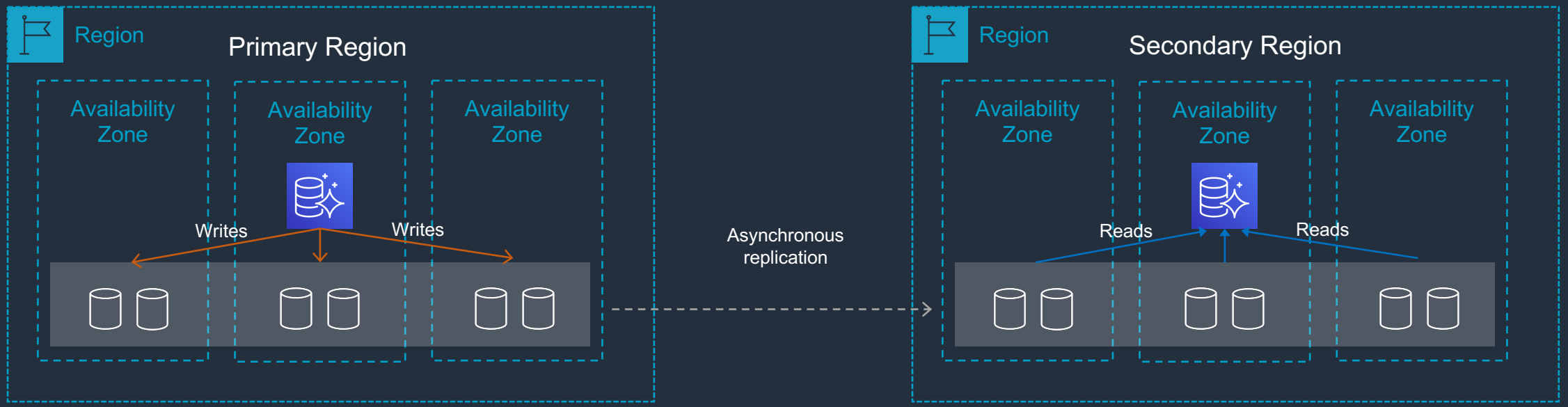
DigitalCloud TRAINING

Section 11: Cross-Region Replica with Aurora MySQL



DigitalCloud TRAINING

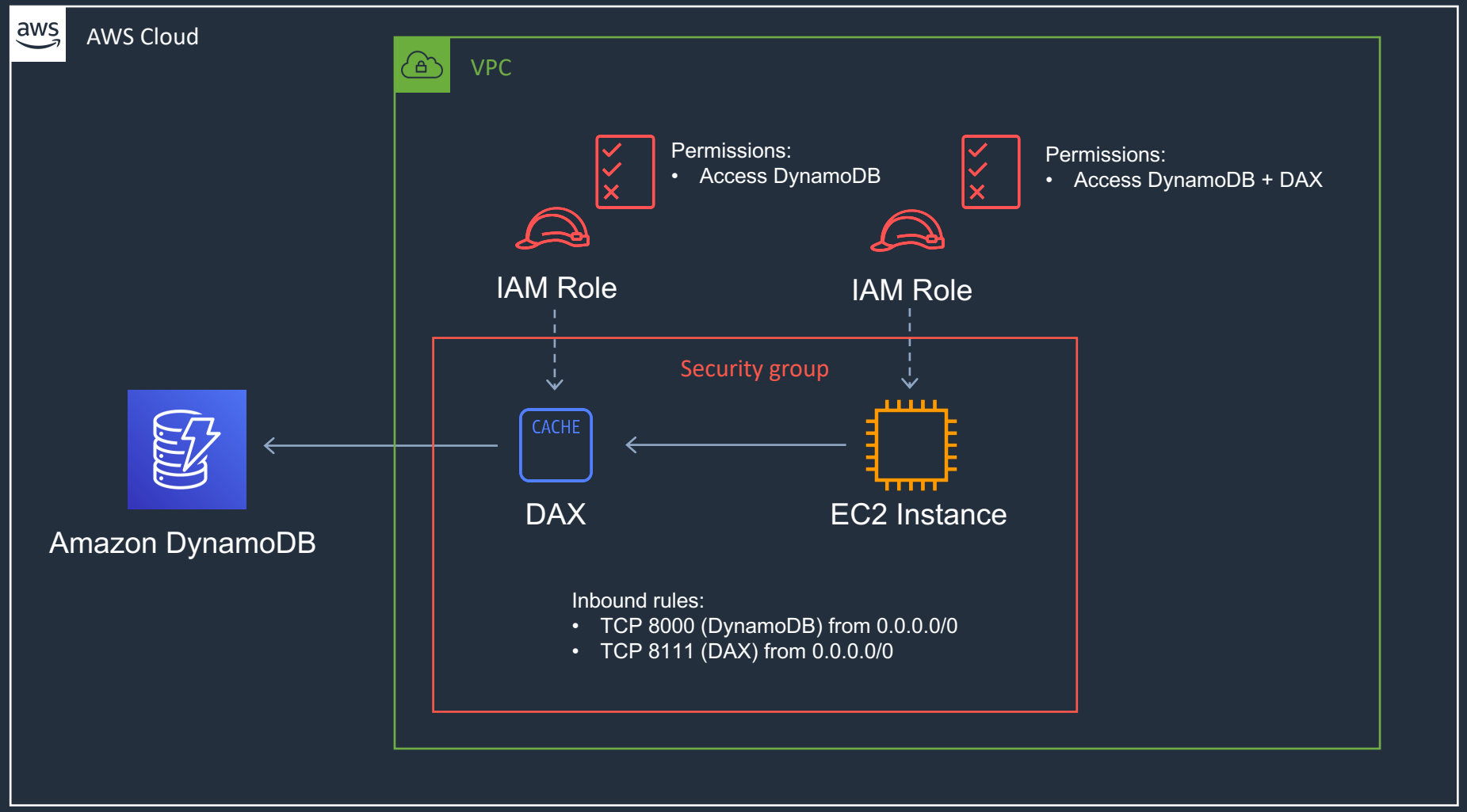
Section 11: Aurora Global Database



Section 11: DynamoDB Overview

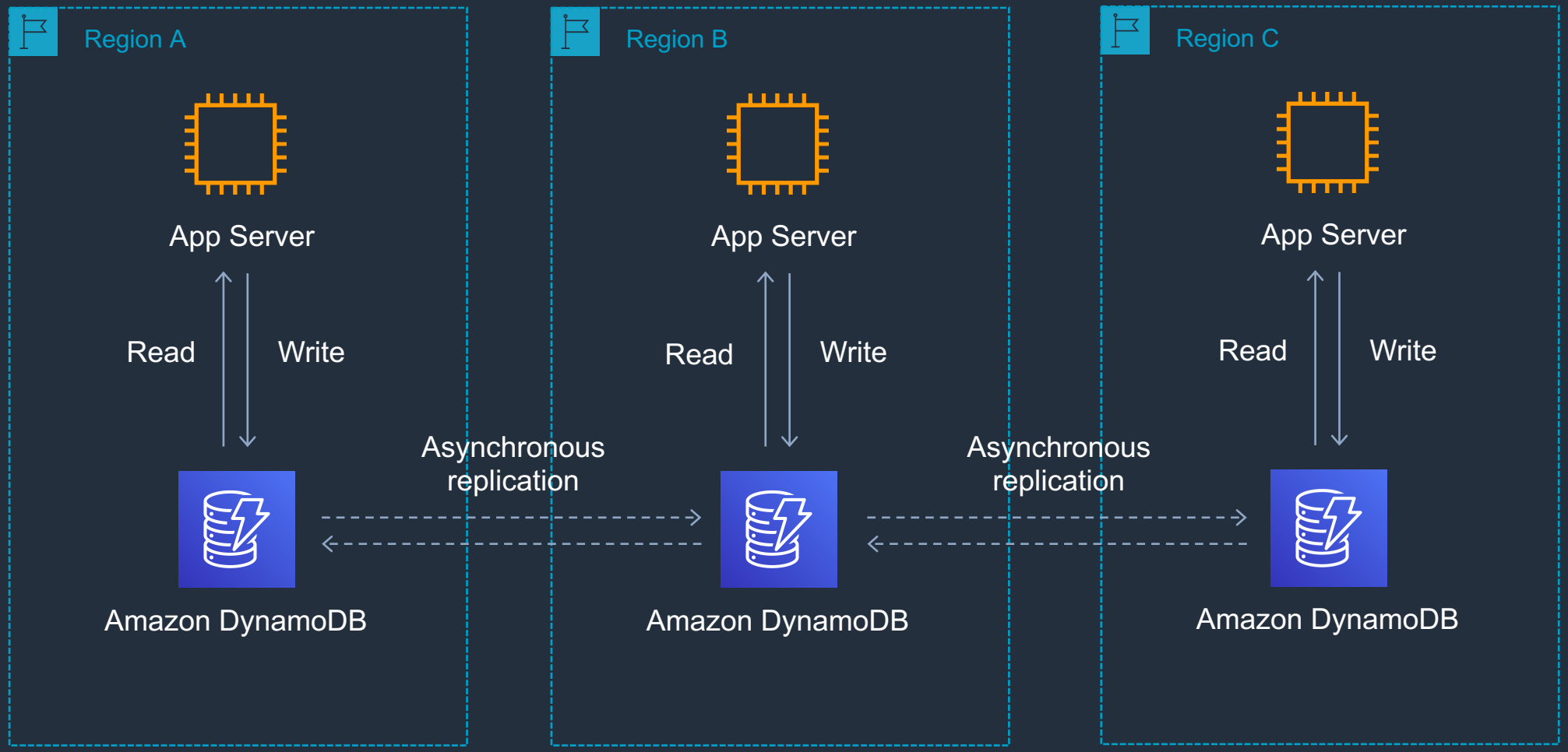
DynamoDB Feature	Benefit
Serverless	Fully managed, fault tolerant, service
Highly available	99.99% availability SLA – 99.999% for Global Tables!
NoSQL type of database with Name / Value structure	Flexible schema, good for when data is not well structured or unpredictable
Horizontal scaling	Seamless scalability to any scale with push button scaling or Auto Scaling
DynamoDB Streams	Captures a time-ordered sequence of item-level modifications in a DynamoDB table and durably stores the information for up to 24 hours. Often used with Lambda and the Kinesis Client Library (KCL)
DynamoDB Accelerator (DAX)	Fully managed in-memory cache for DynamoDB that increases performance (microsecond latency)
Transaction options	Strongly consistent or eventually consistent reads, support for ACID transactions
Backup	Point-in-time recovery down to the second in last 35 days; On-demand backup and restore
Global Tables	Fully managed multi-region, multi-master solution

Section 11: DynamoDB Accelerator (DAX)



DigitalCloud TRAINING

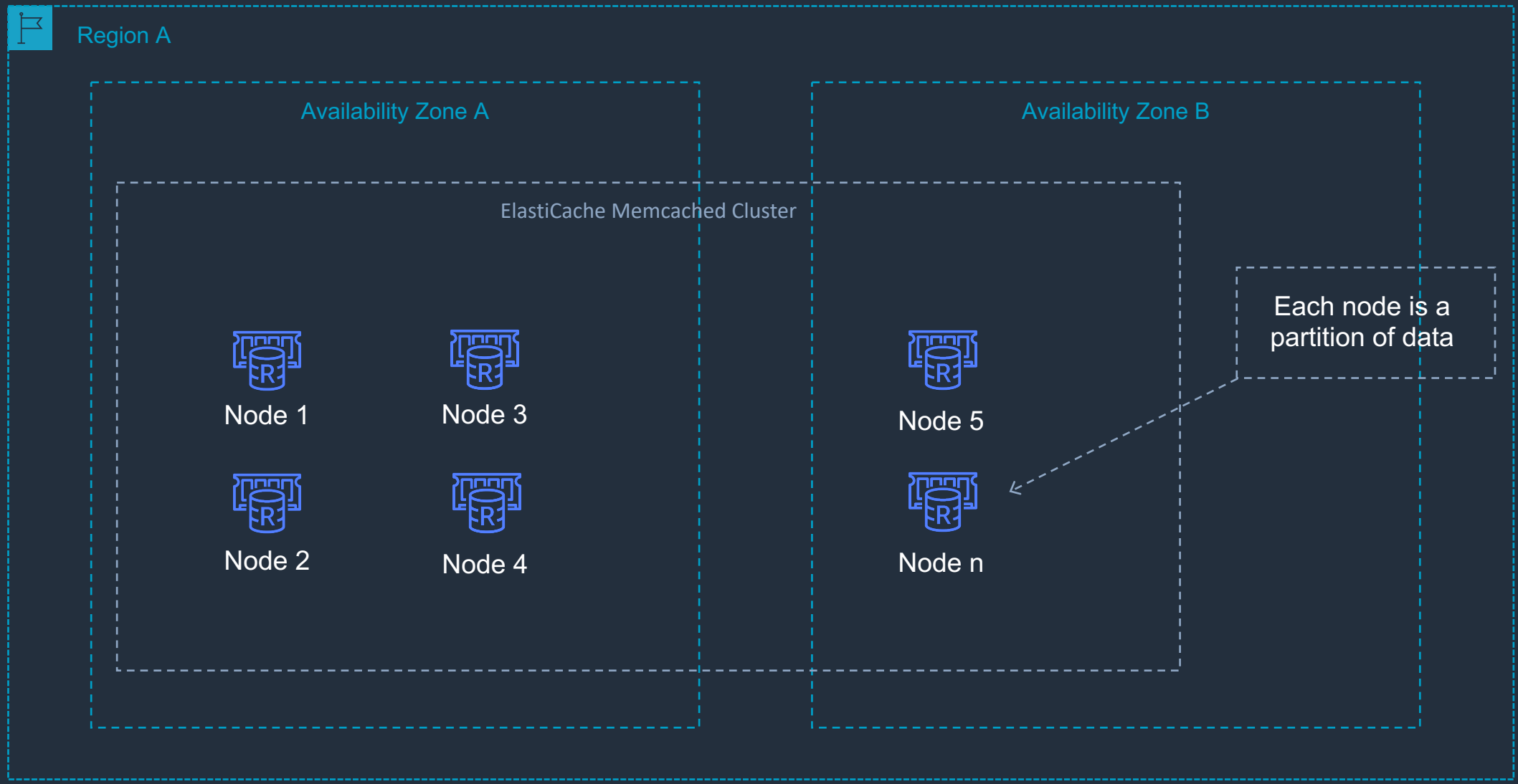
Section 11: DynamoDB Global Tables



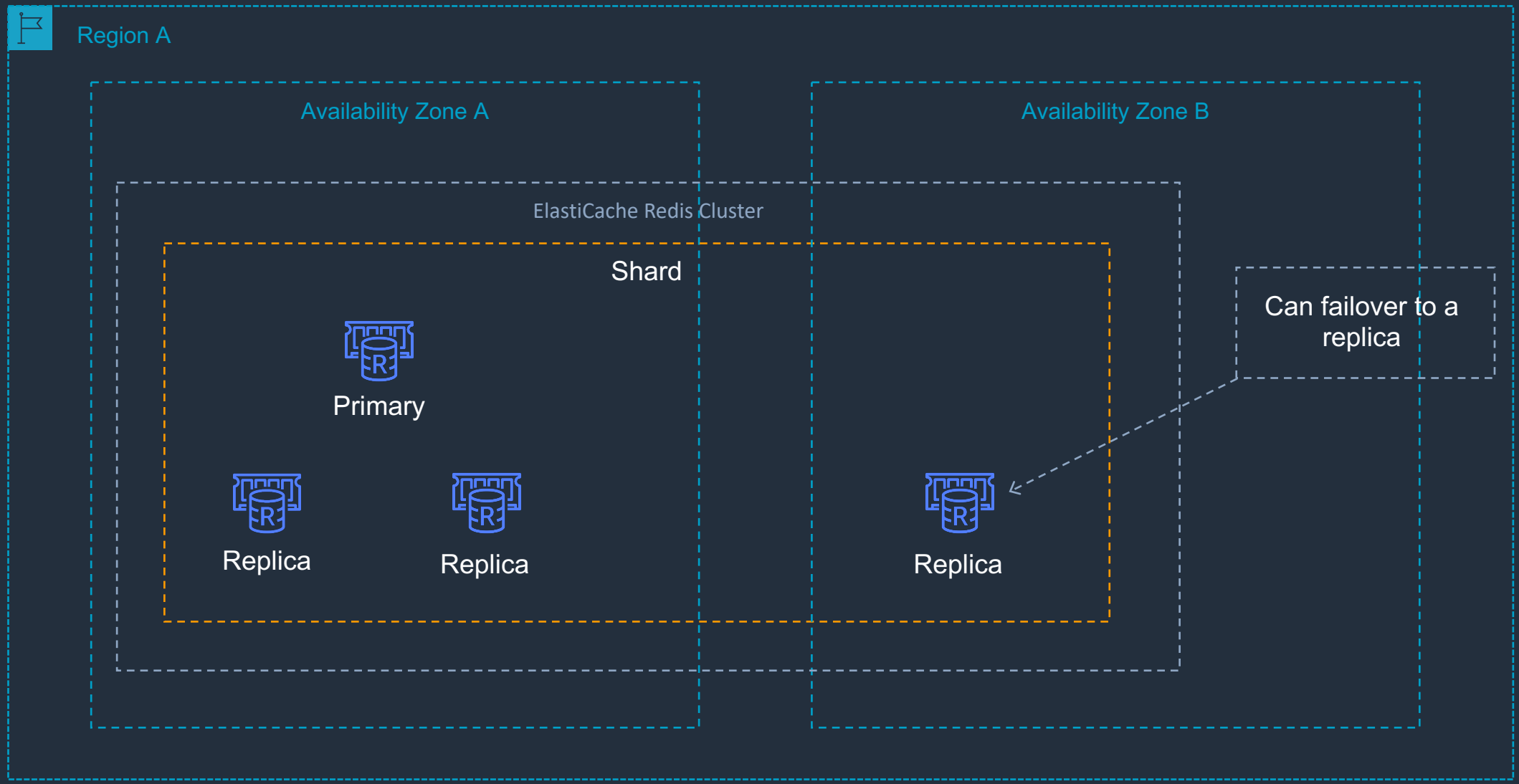
Section 11: ElastiCache Overview

Feature	Memcached	Redis (cluster mode disabled)	Redis (cluster mode enabled)
Data persistence	No	Yes	Yes
Data types	Simple	Complex	Complex
Data partitioning	Yes	No	Yes
Encryption	No	Yes	Yes
High availability (replication)	No	Yes	Yes
Multi-AZ	Yes, place nodes in multiple AZs. No failover or replication	Yes, with auto-failover. Uses read replicas (0-5 per shard)	Yes, with auto-failover. Uses read replicas (0-5 per shard)
Scaling	Up (node type); out (add nodes)	Single shard (can add replicas)	Add shards
Multithreaded	Yes	No	No
Backup and restore	No (and no snapshots)	Yes, automatic and manual snapshots	Yes, automatic and manual snapshots

Section 11: ElastiCache Memcached

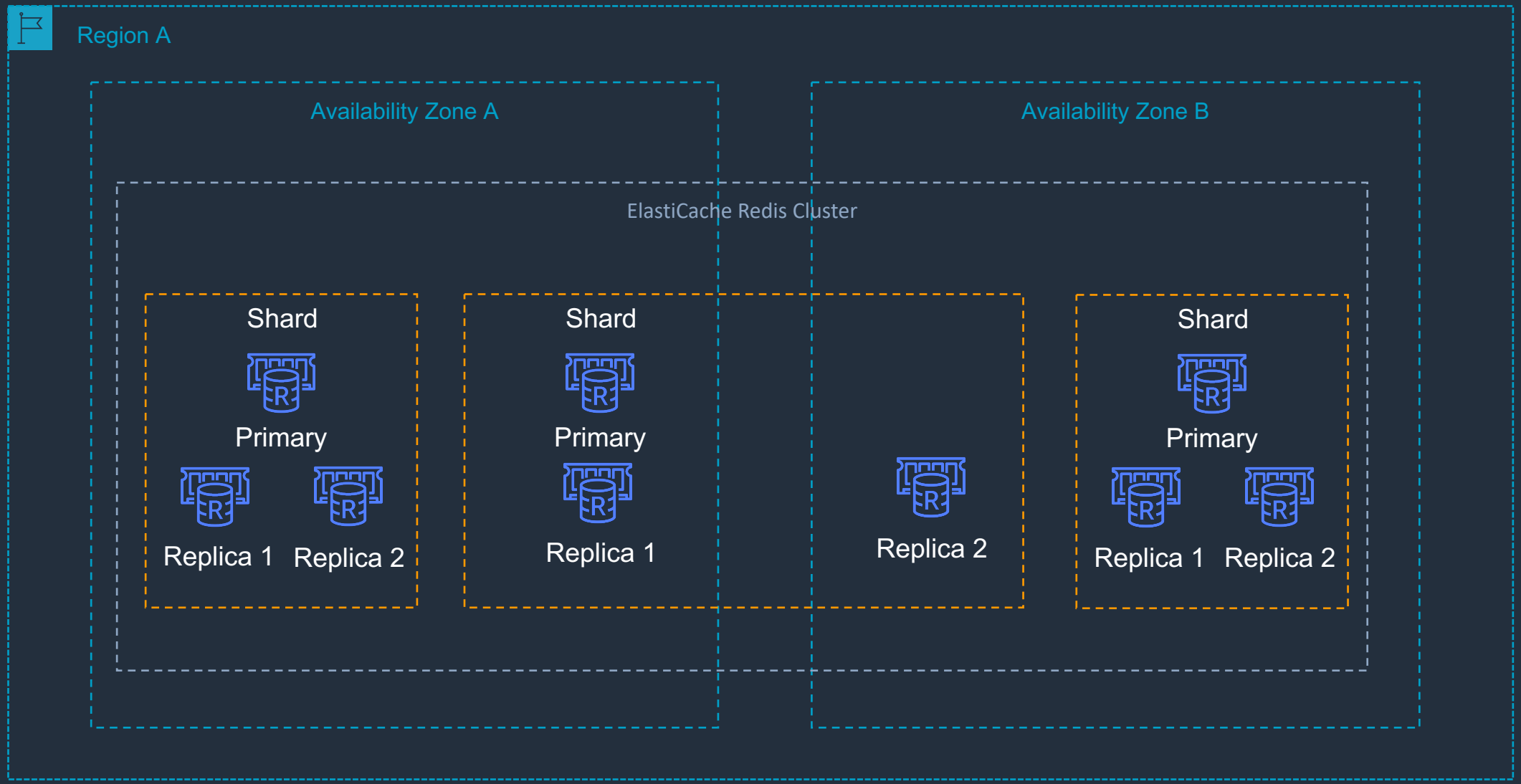


Section 11: ElastiCache Redis (Cluster mode disabled)



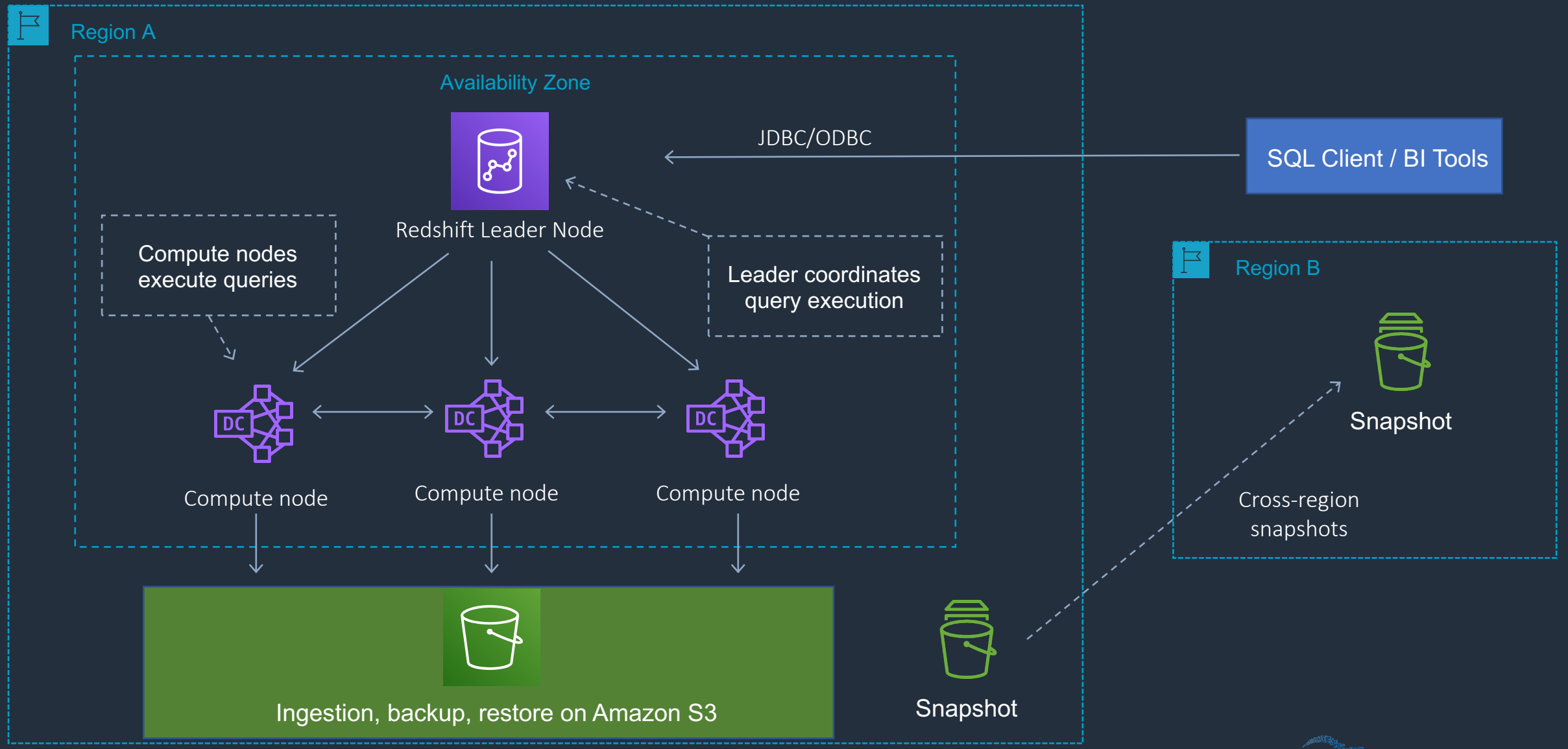
DigitalCloud TRAINING

Section 11: ElastiCache Redis (Cluster mode enabled)

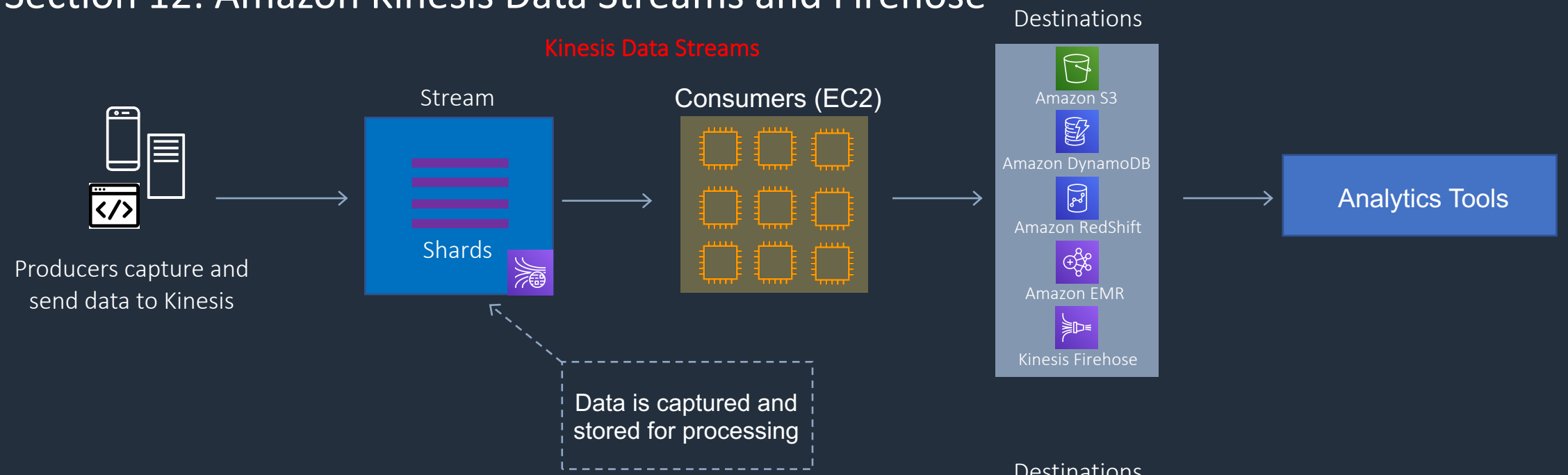


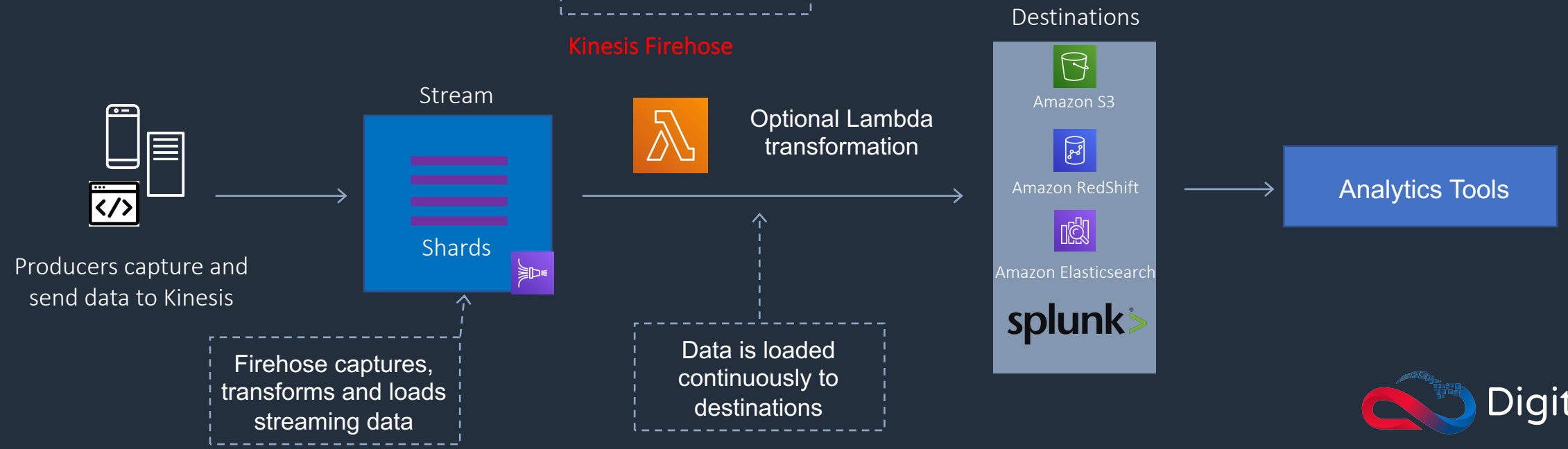
DigitalCloud TRAINING

Section 11: Amazon RedShift



Section 12: Amazon Kinesis Data Streams and Firehose

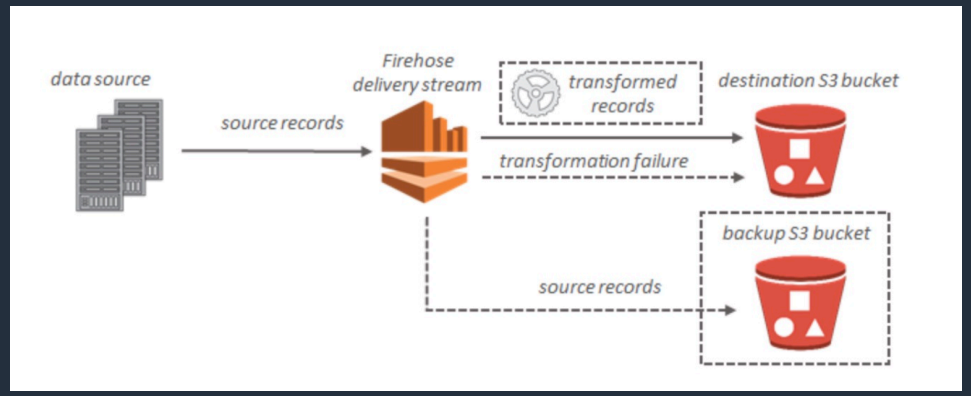




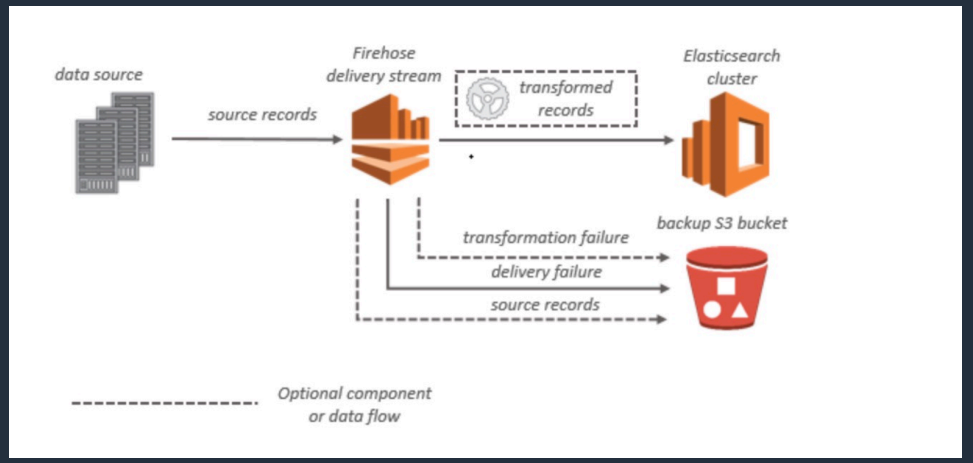
DigitalCloud TRAINING

Section 12: Amazon Kinesis Firehose Destinations

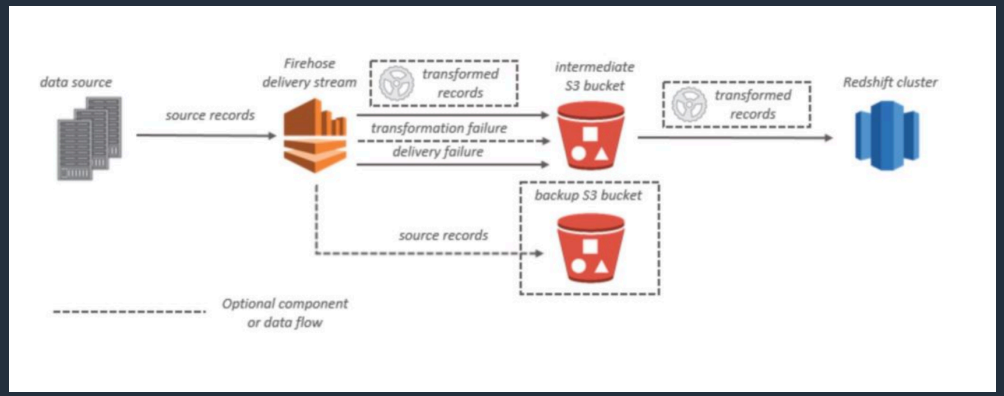
Amazon S3: delivered to S3 bucket, optional backup



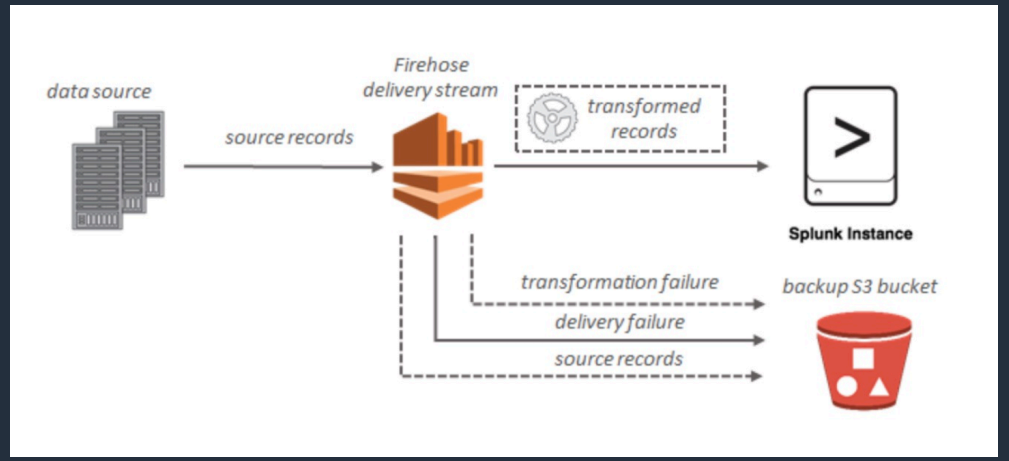
Amazon Elasticsearch: delivered to ES and optionally to S3



Amazon RedShift: delivered to S3 bucket first, then RedShift

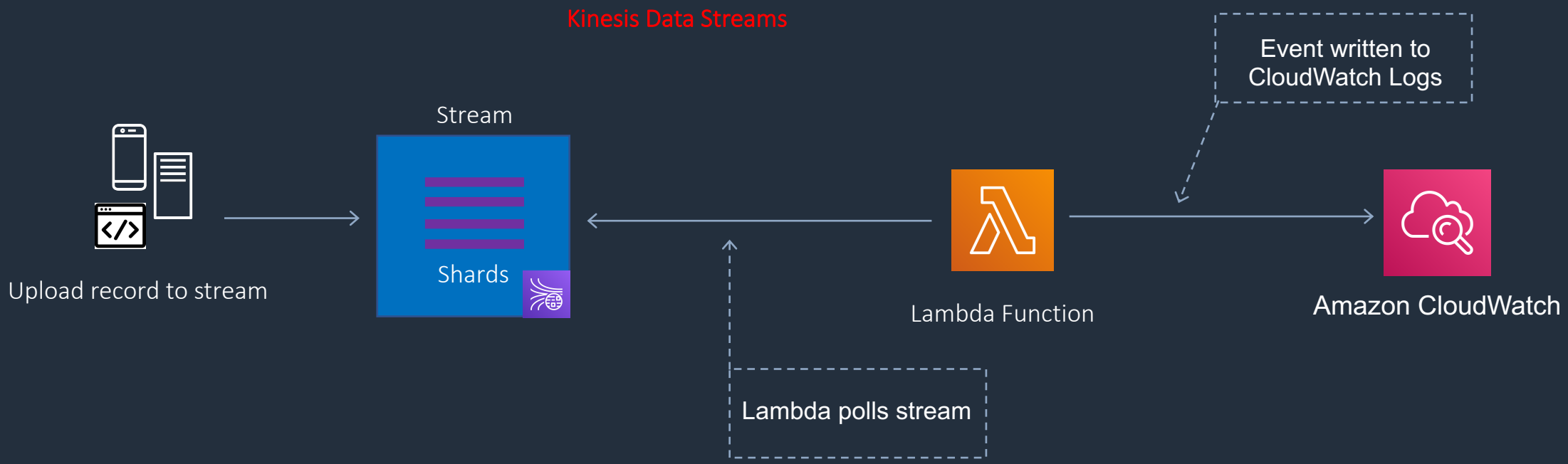


Splunk: delivered to Splunk and optionally to S3

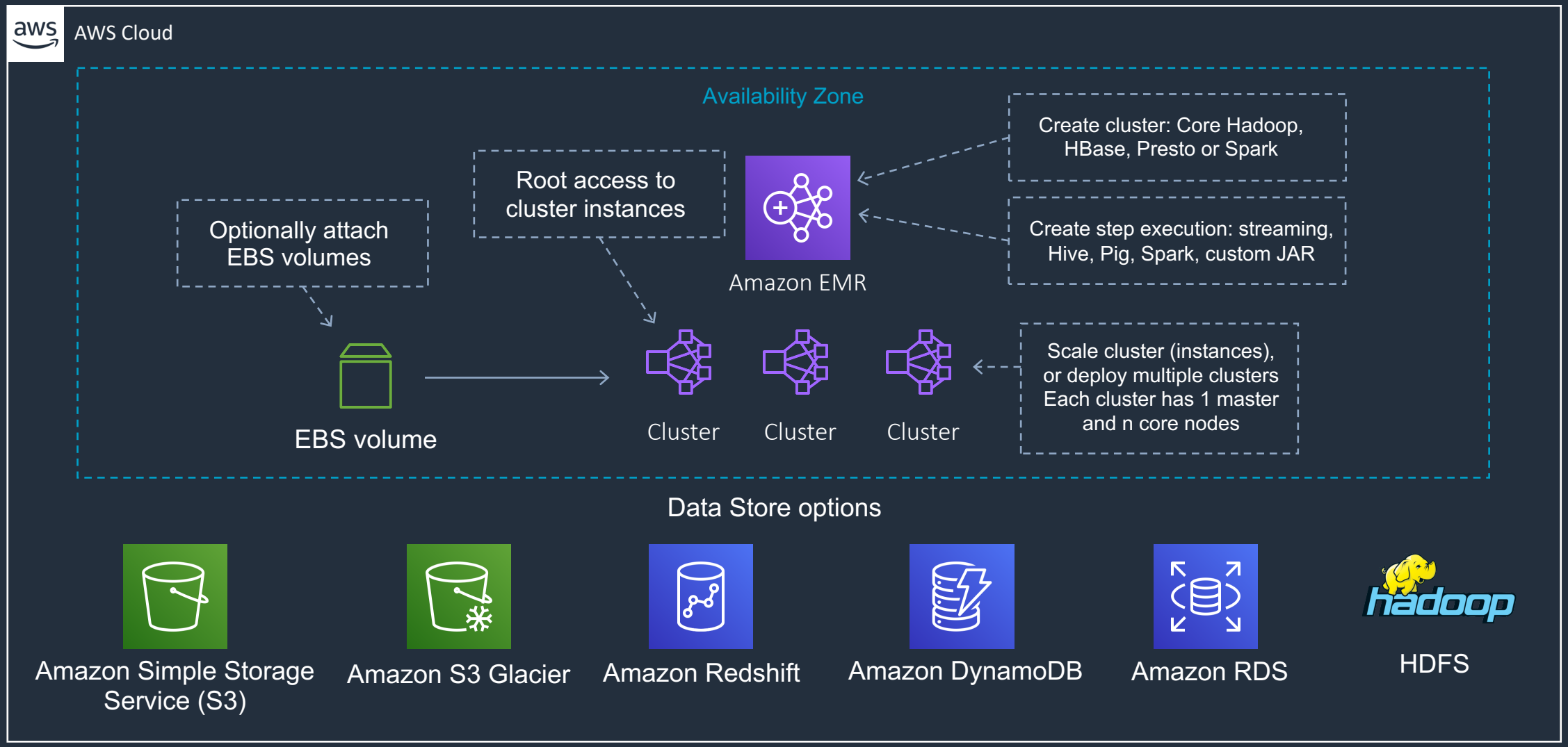


DigitalCloud TRAINING

Section 12: AWS Lambda and Kinesis Stream



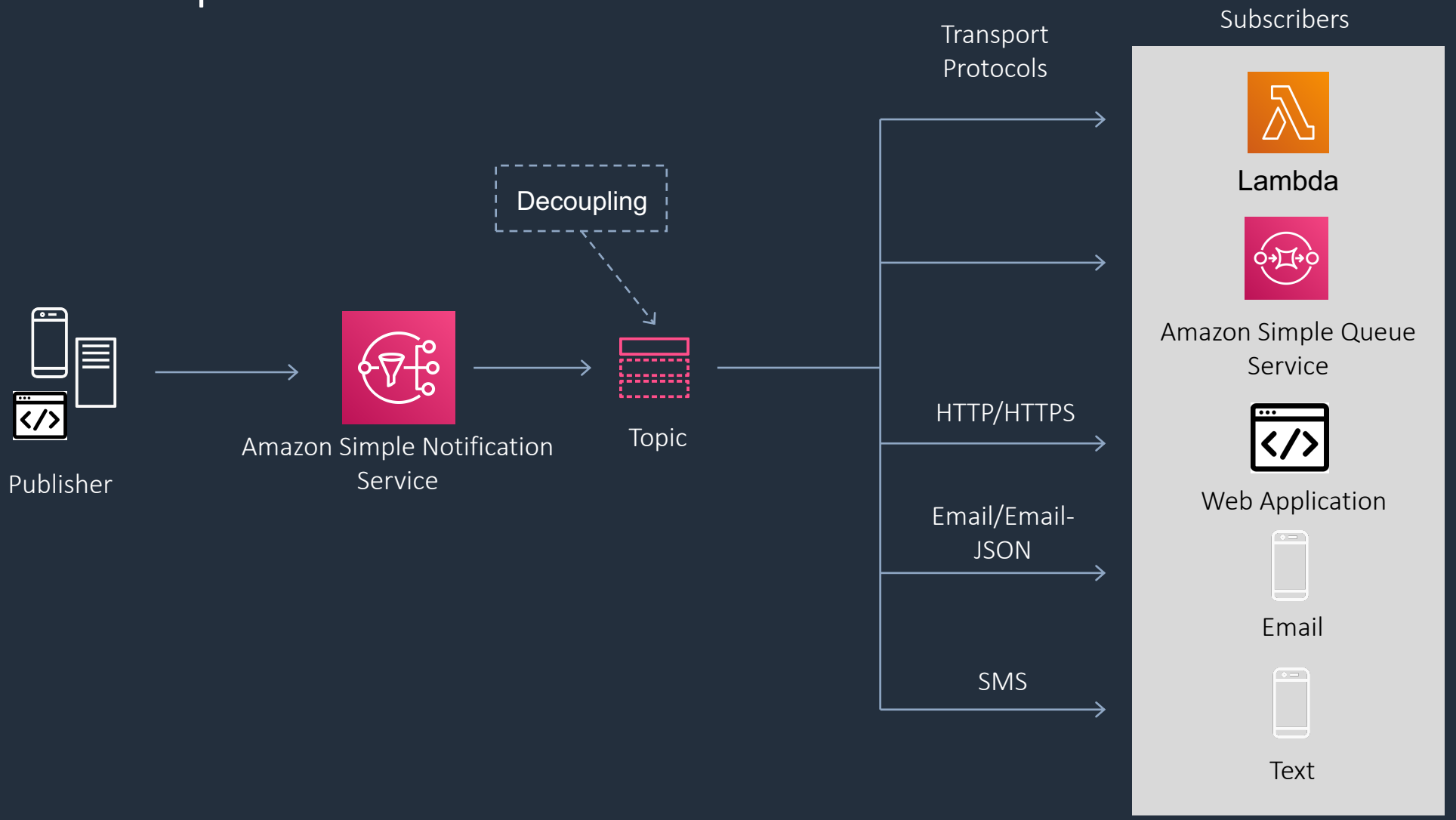
Section 12: Amazon EMR



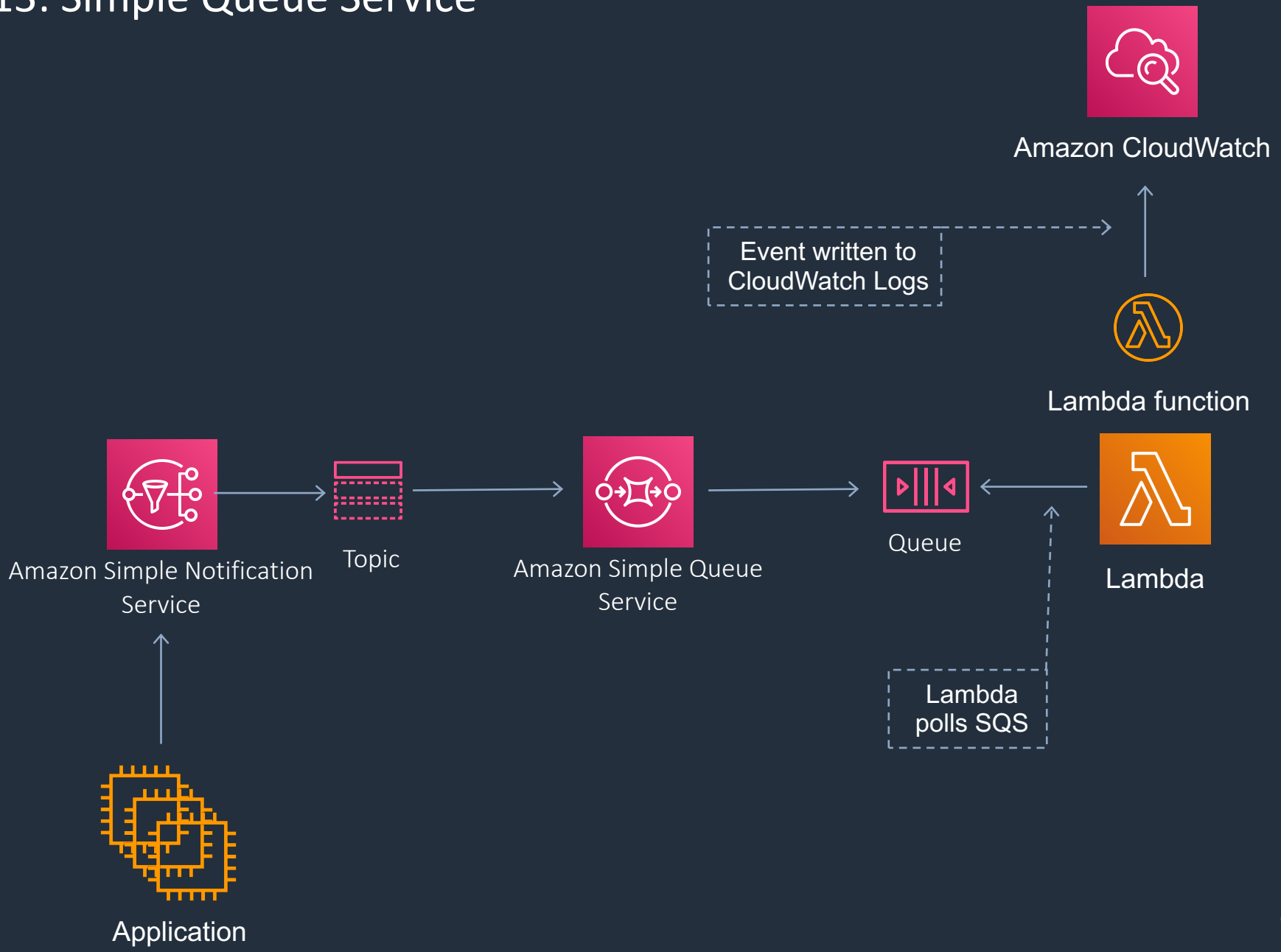
Section 13: Application Integration Services

Service	What it does	Example use cases
Simple Notification Service	Set up, operate, and send notifications from the cloud	Send email notification when CloudWatch alarm is triggered
Step Functions	Out-of-the-box coordination of AWS service components with visual workflow	Order processing workflow
Simple Workflow Service	Need to support external processes or specialized execution logic	Human-enabled workflows like an order fulfilment system or for procedural requests AWS recommends that for new applications customers consider Step Functions instead of SWF
Simple Queue Service	Messaging queue; store and forward patterns	Building distributed / decoupled applications
Amazon MQ	Managed message broker based on Apache MQ	Easy low-hassle path to migrate from existing message brokers to AWS

Section 13: Simple Notification Service



Section 13: Simple Queue Service



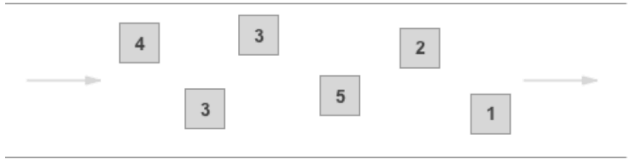
Section 13: Simple Queue Service Queue Types

Standard Queues

Unlimited Throughput: Standard queues support a nearly unlimited number of transactions per second (TPS) per API action.

At-Least-Once Delivery: A message is delivered at least once, but occasionally more than one copy of a message is delivered.

Best-Effort Ordering: Occasionally, messages might be delivered in an order different from which they were sent.



You can use standard message queues in many scenarios, as long as your application can process messages that arrive more than once and out of order, for example:

- Decouple live user requests from intensive background work: Let users upload media while resizing or encoding it.
- Allocate tasks to multiple worker nodes: Process a high number of credit card validation requests.
- Batch messages for future processing: Schedule multiple entries to be added to a database.

FIFO Queues

High Throughput: By default, FIFO queues support up to 300 messages per second (300 send, receive, or delete operations per second). When you batch 10 messages per operation (maximum), FIFO queues can support up to 3,000 messages per second. To request a limit increase, file a support request.

Exactly-Once Processing: A message is delivered once and remains available until a consumer processes and deletes it. Duplicates aren't introduced into the queue.

First-In-First-Out Delivery: The order in which messages are sent and received is strictly preserved (i.e. First-In-First-Out).

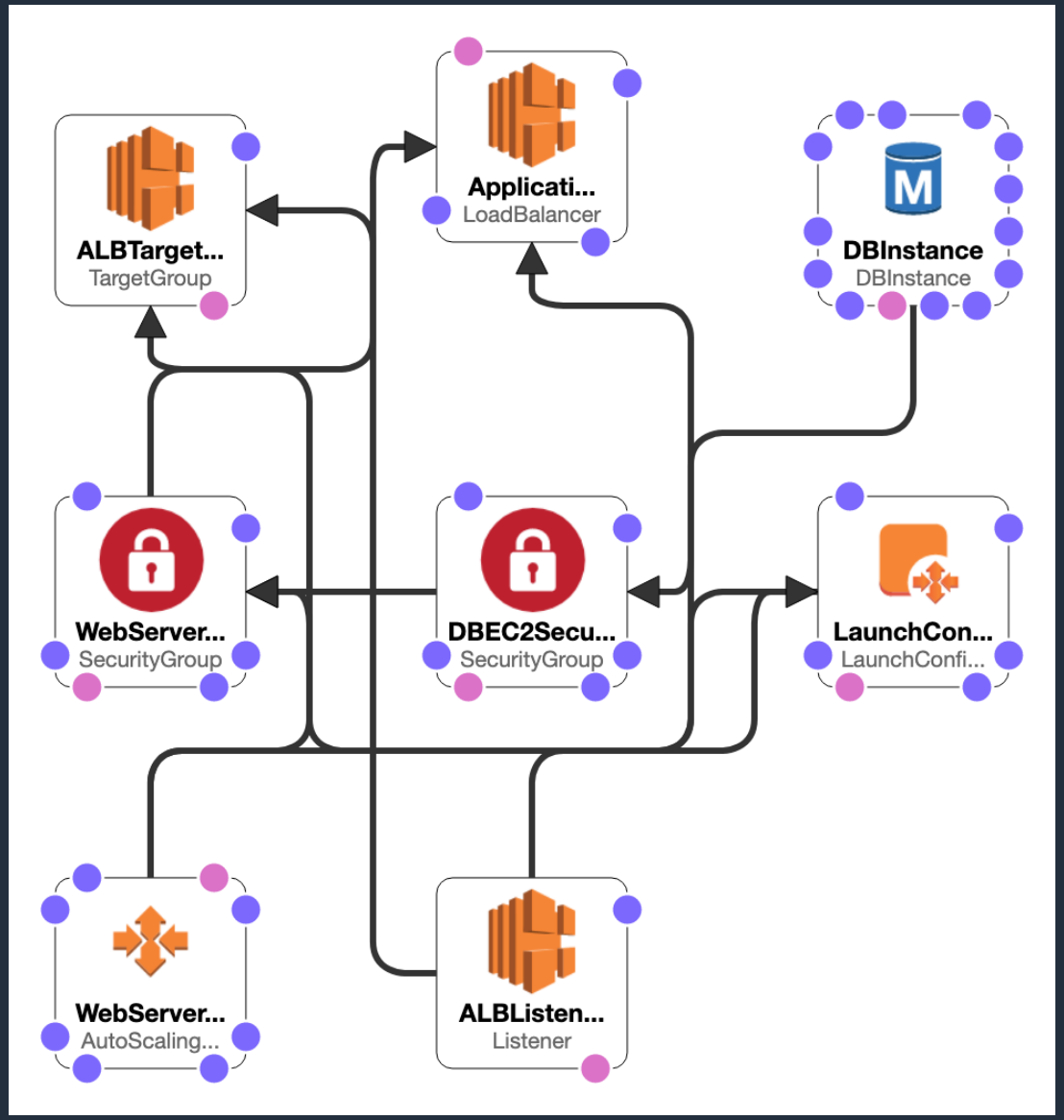
FIFO queues are designed to enhance messaging between applications when the order of operations and events is critical, or where duplicates can't be tolerated, for example:

- Ensure that user-entered commands are executed in the right order.
- Display the correct product price by sending price modifications in the right order.
- Prevent a student from enrolling in a course before registering for an account.

Section 14: Infrastructure as Code and PaaS

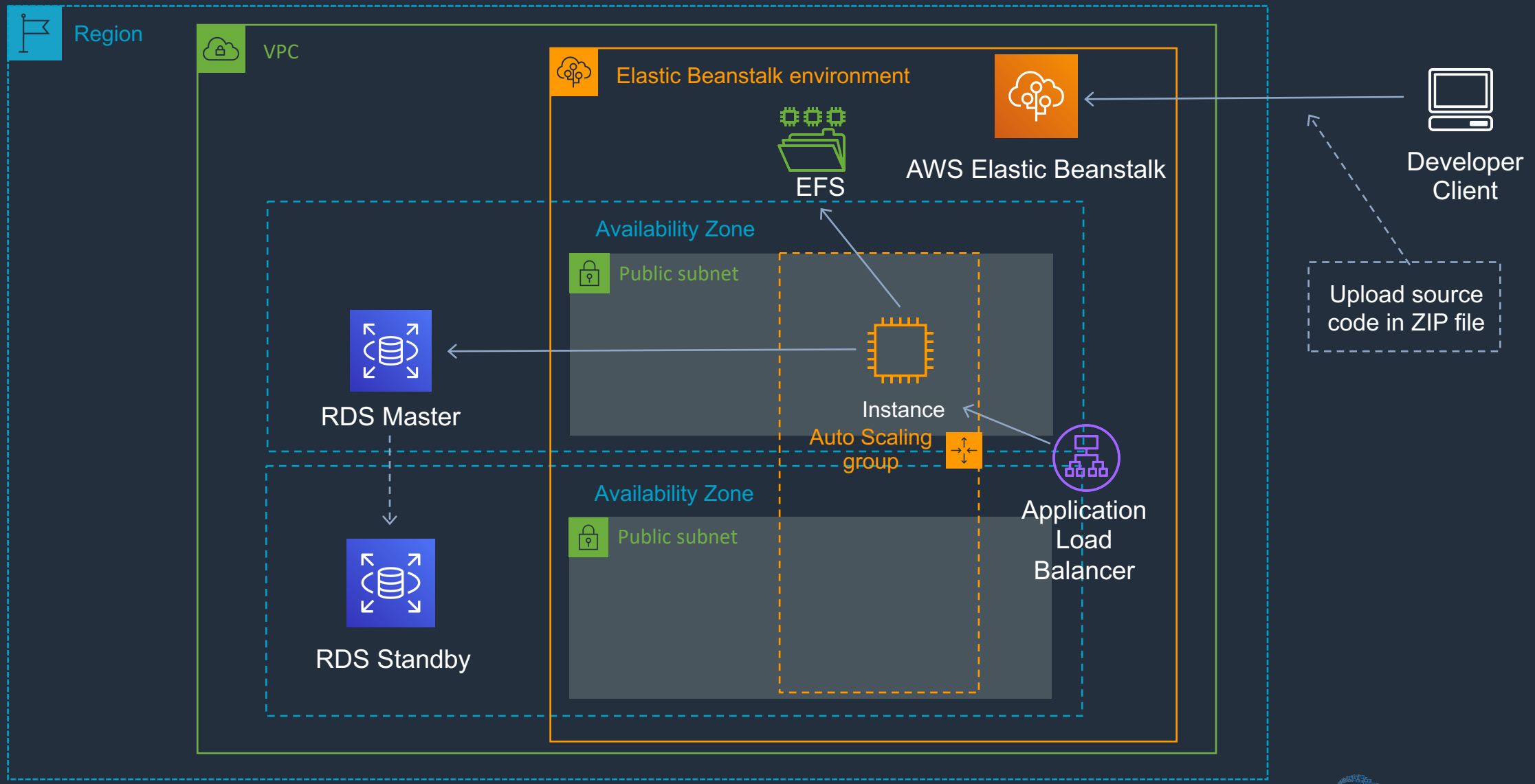
CloudFormation	Elastic Beanstalk
“Template-driven provisioning”	“Web apps made easy"
Deploys infrastructure using code	Deploys applications on EC2 (PaaS)
Can be used to deploy almost any AWS service	Deploys web applications based on Java, .NET, PHP, Node.js, Python, Ruby, Go, and Docker
Uses JSON or YAML template files	Uses ZIP or WAR files (or Git)
CloudFormation can deploy Elastic Beanstalk environments	Elastic Beanstalk cannot deploy using CloudFormation
Similar to Terraform	Similar to Google App Engine

Section 14: HA Wordpress using CloudFormation



DigitalCloud TRAINING

Section 14: HA WordPress with Elastic Beanstalk and RDS

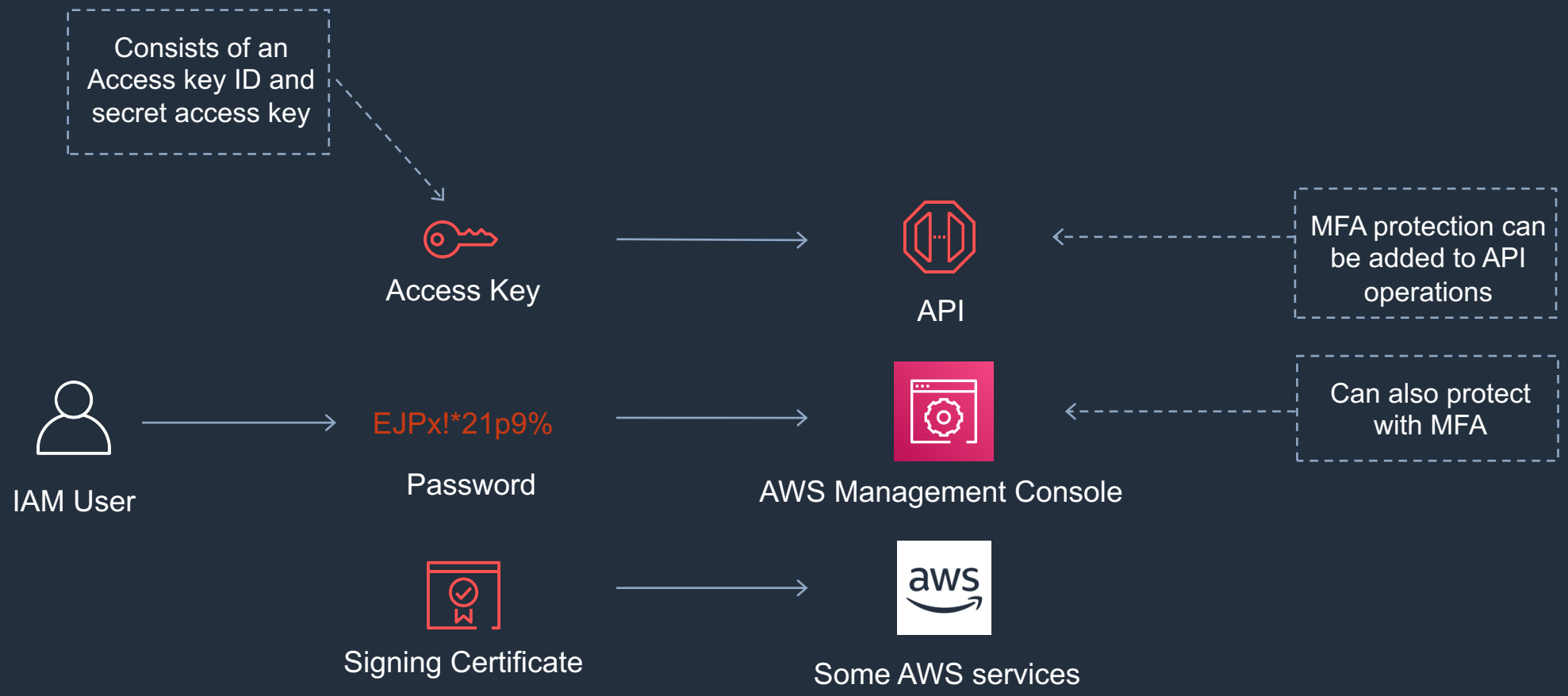


DigitalCloud TRAINING

Section 15: Monitoring and Logging Overview

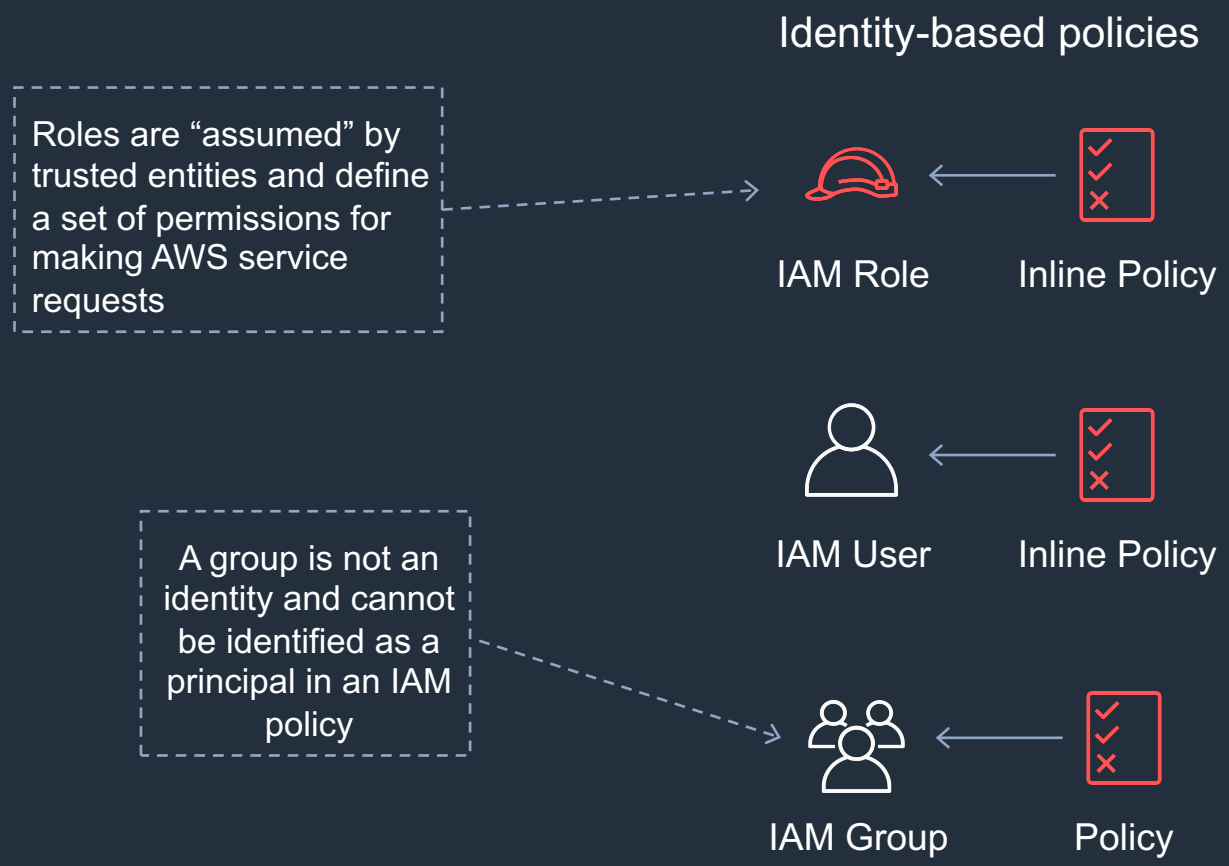
CloudWatch	CloudTrail
Performance monitoring (operations)	Auditing (security)
Log events across AWS services – think operations	Log API activity across AWS services – think activities
Higher-level comprehensive monitoring and eventing	More low-level granular
Log from multiple accounts	Log from multiple accounts
Logs stored indefinitely	Logs stored to S3 or CloudWatch indefinitely
Alarms history for 14 days	No native alarming; can use CloudWatch alarms

Section 16: IAM Authentication Methods



DigitalCloud TRAINING

Section 16: IAM Policies – Roles, Users and Groups



IAM Policies:

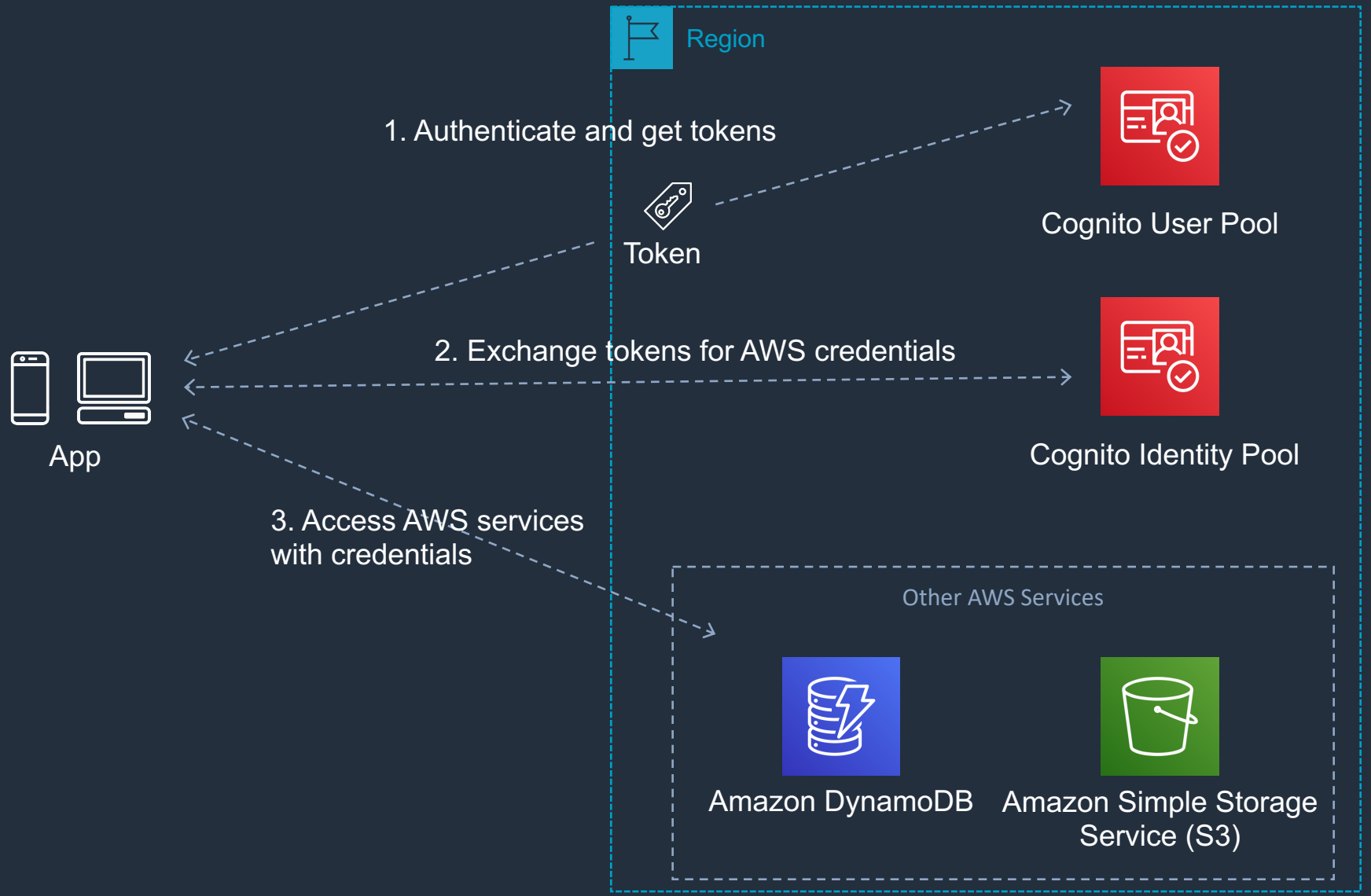
- Policies are documents that define permissions and can be applied to users, groups and roles
- Written in JSON
- All permissions are implicitly denied by default
- With multiple policies the most restrictive policy is applied
- The Condition element can be used to apply further conditional logic

DigitalCloud TRAINING

Section 16: IAM Best Practices

- Lock away the AWS root user access keys
- Create individual IAM users
- Use AWS defined policies to assign permissions whenever possible
- Use groups to assign permissions to IAM users
- Grant least privilege
- Use access levels to review IAM permissions
- Configure a strong password policy for users
- Enable MFA for privileged users
- Use roles for applications that run on AWS EC2 instances
- Delegate by using roles instead of sharing credentials
- Rotate credentials regularly
- Remove unnecessary credentials
- Use policy conditions for extra security
- Monitor activity in your AWS account

Section 16: Amazon Cognito



Section 16: KMS Customer Master Keys (CMKs)

Type of CMK	Can view	Can manage	Used only for my account
Customer managed CMK	Yes	Yes	Yes
AWS managed CMK	Yes	No	Yes
AWS owned CMK	No	No	No

Section 16: Old and New CloudHSM

	”Classic” CloudHSM	Current CloudHSM
Device	safeNET Luna SA	Proprietary AWS
Pricing	Upfront cost required ($5000)	No upfront cost, pay per hour
High Availability	Have to buy a second device	Clustered
FIPS 140-2	Level 2	Level 3