GSI Global

AWS Certification Accelerator Groups

Solution Architect Associate – Session 1

AWS DXC Partner SA team

General Housekeeping Notes

- This webinar is confidential to AWS and DXC internally
- Everyone is in listen only mode until the end of the presentation
- Have a question or comment? Please:
    - Technical: type or upvote https://www.upvote.click/#CAG-SA-APAC
    - Administrative questions
        - Registration/voucher information, please work with AWSPartnerTraining@dxc.com
        - LMS portal i.e. AWS Academy portal support, refer to:
            - General: https://www.apn-portal.com/knowledgebase/
            - Portal help – Click the help button (left hand side) and chat live with a help member
            - General Support: https://support.aws.amazon.com/#/contacts/aws-training
- No E-mail support at this time from SA team for the CAG sessions

aws

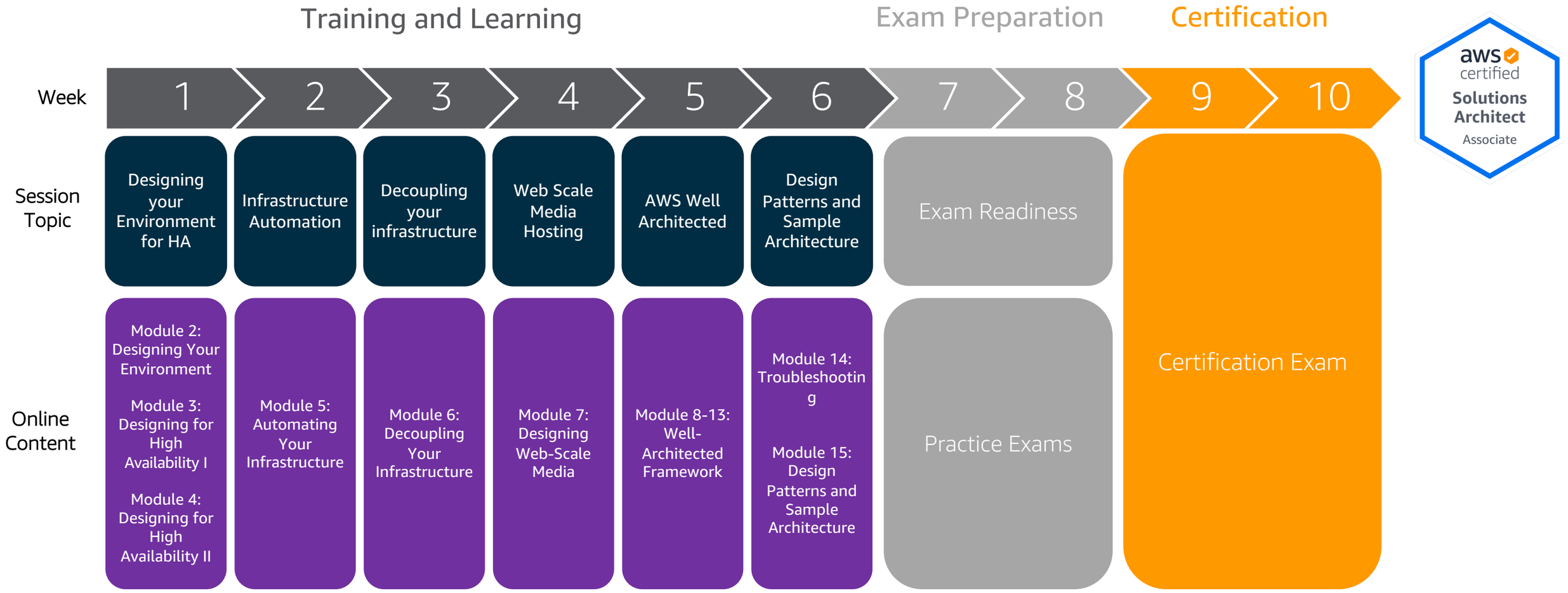
Todays Agenda

- Session 1 – Topic Recap (5 mins)
- Key Concepts (20 mins)
- Practice Questions (20 mins)
- Next Weeks Session (5 mins)
- Discussion (10 mins)

aws

Session 1 – Topic Overview

AWS CAG: Solution Architect Course Schedule



Session 1 - Overview

Session Topic - Designing your Environment for HA

Online Modules:

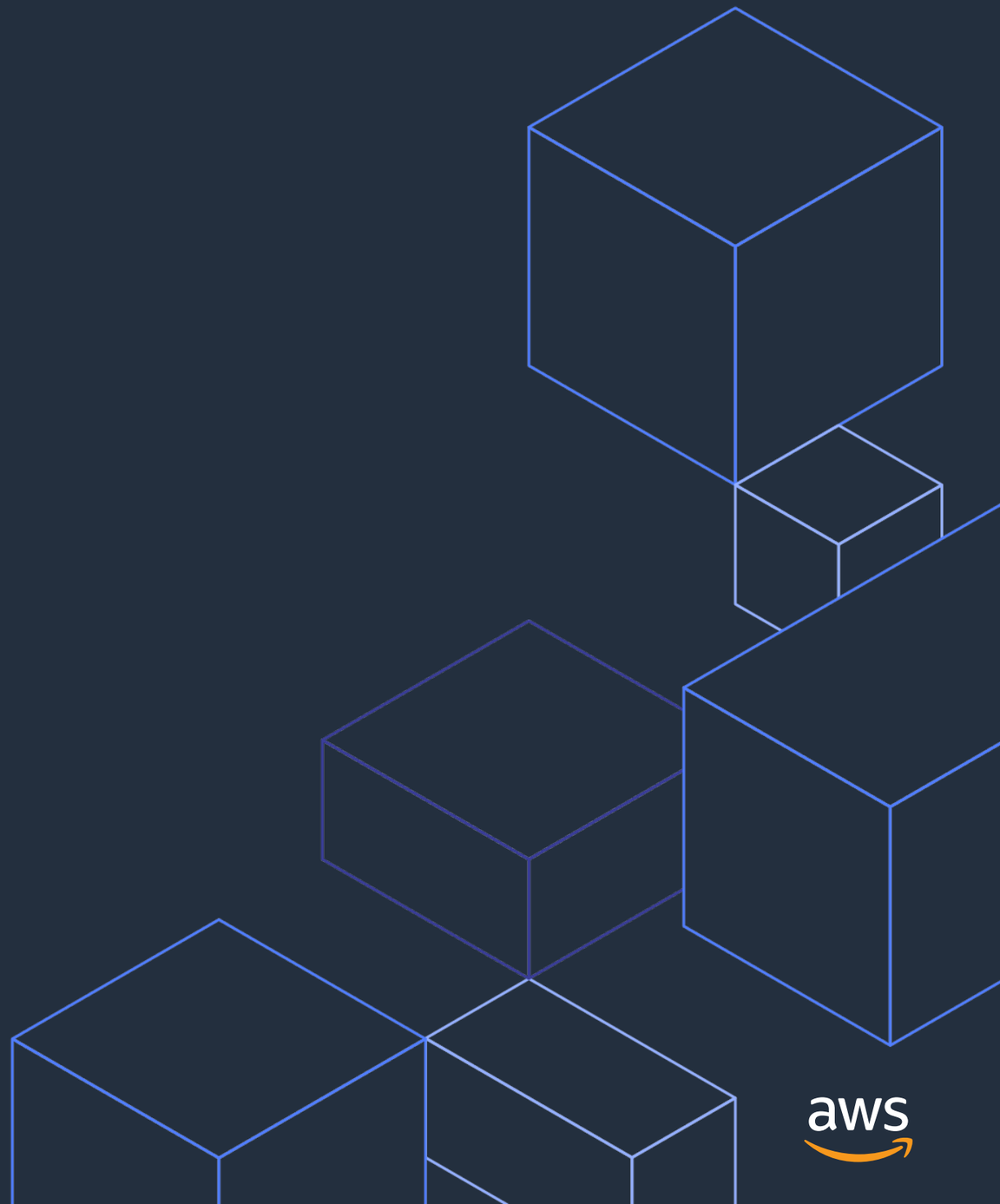
- Module 2: Designing Your Environment
- Module 3: Designing for High Availability I
- Module 4: Designing for High Availability II

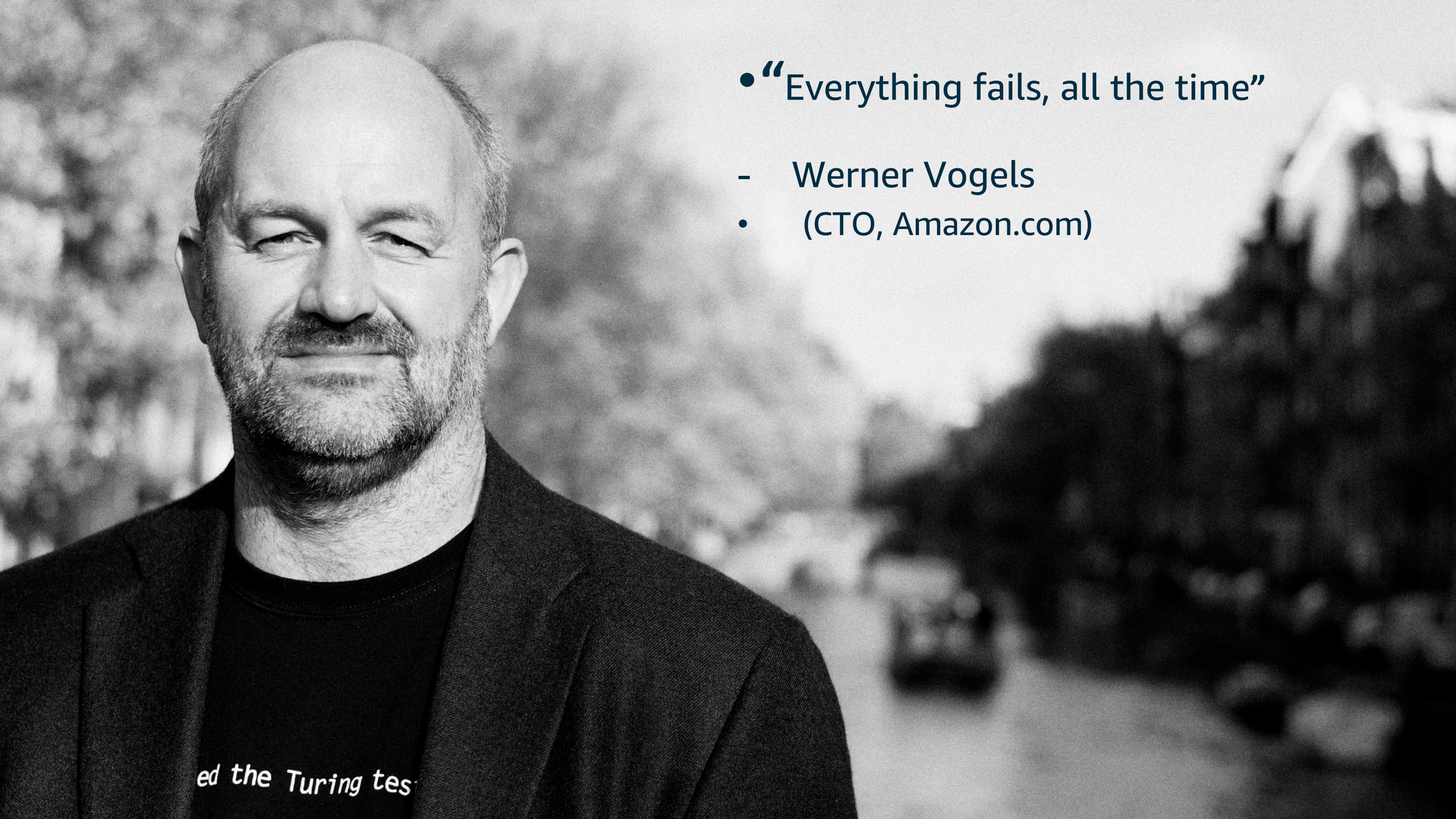
Supplementary Content:

- AWS Whitepaper - Fault-Tolerant Components on AWS - https://d1.awsstatic.com/whitepapers/aws-building-fault-tolerant-applications.pdf?did=wp_card&trk=wp_card

aws

Key Concepts





- “Everything fails, all the time”

- Werner Vogels
- (CTO, Amazon.com)

Core Principles – Designing your Environment for HA

- Automatically recover from failure
- Manage change in automation
- Scale horizontally to increase aggregate system availability
- Stop guessing capacity
- Test recovery procedures

aws

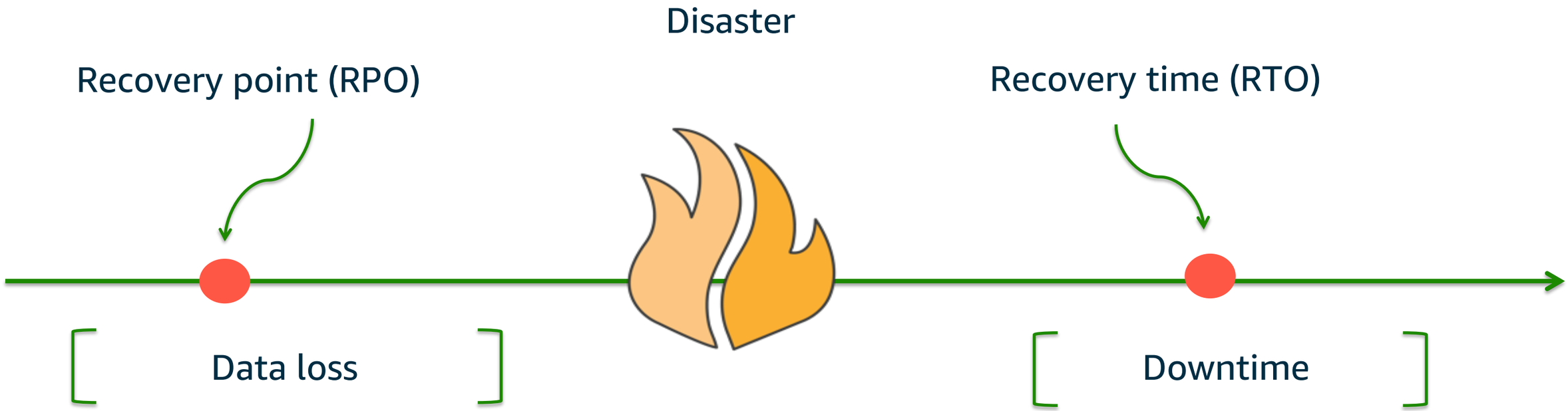
Key Concepts – Designing for High Availability

- Understand your business needs
    - Building on the core principles, architect appropriately to meet your business objectives
- Regions/Availability Zone Design
    - Leverage the AWS global infrastructure to ensure high availability in the event of failure
- Self-Healing applications
    - Automate your recovery where ever possible and fail gracefully
- Data Resiliency
    - How are you ensuring HA at the data layer and ensuring data is not lost or corrupted
- Network Design for HA
    - HA isn't just your application stack, its HA in your network topology to manage failures

Understanding business needs

How much data can you afford to recreate or lose?

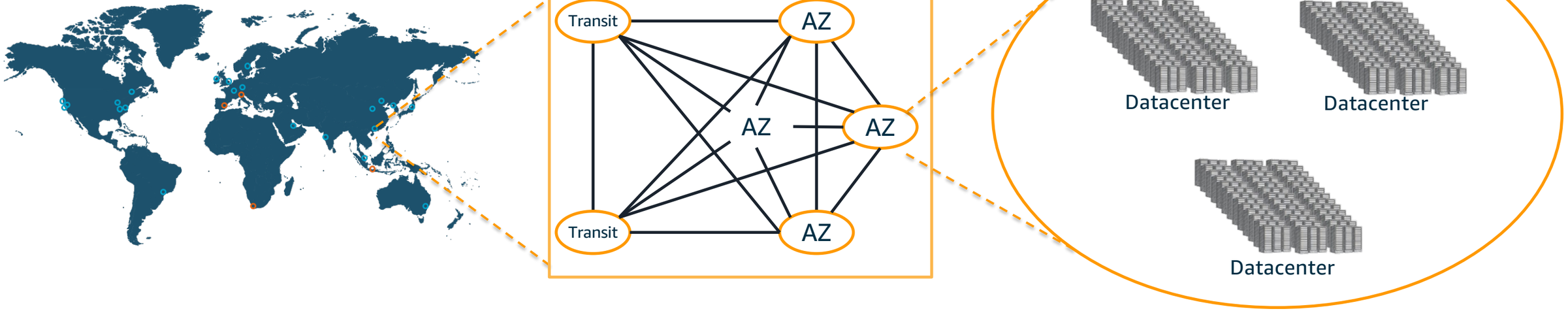
How quickly must you recover? What is the cost of downtime?



It's *not* about the data, it's about the *mission*

Regions/Availability Zone Refresher

- AWS Regions are comprised of multiple AZs for high availability, high scalability, and high fault tolerance. Applications and data are replicated in real time and consistent in the different AZs.



A Region is a physical location in the world where we have multiple **Availability Zones.**

Availability Zones consist of one or more discrete data centers, each with redundant power, networking, and connectivity, housed in separate facilities.

Regions/Availability Zone Design

Infrastructure

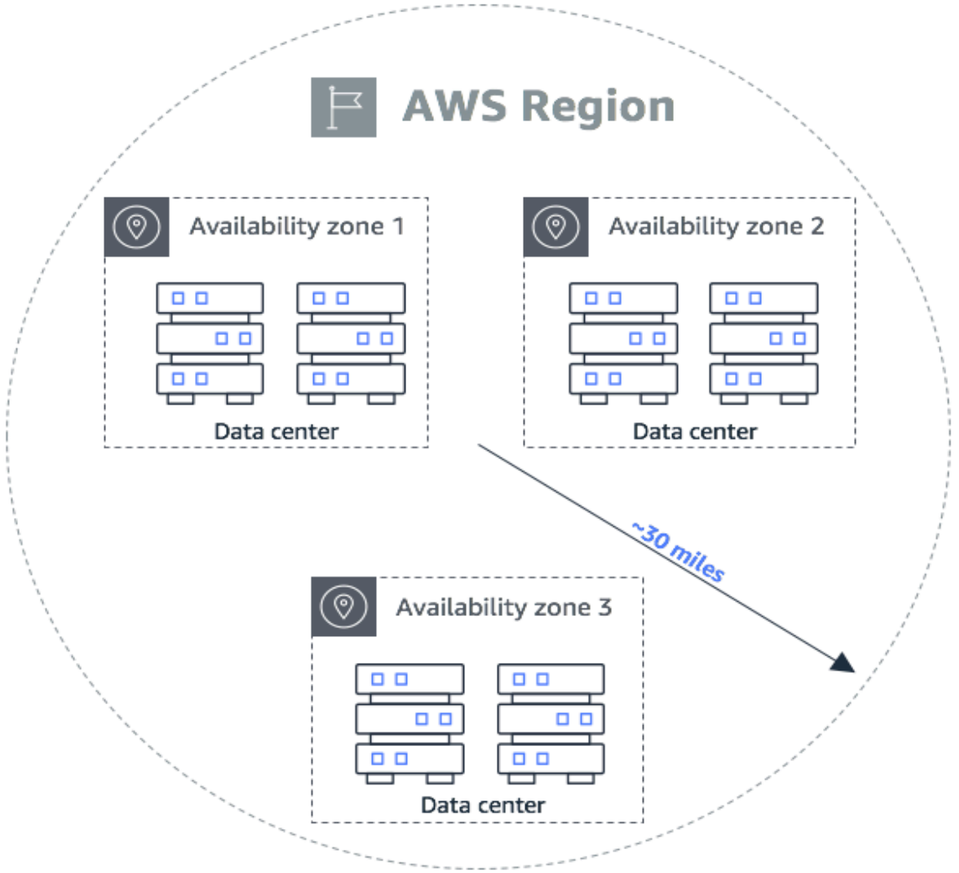
- Regions, AZs, Networking

Service Design

- Cell-based architecture
- Multi-Az architecture
- Micro-service architecture
- Distributed systems best practices

Understand the AWS Services scope

- Single AZ, Regional, Global, Cross-Reginal capability

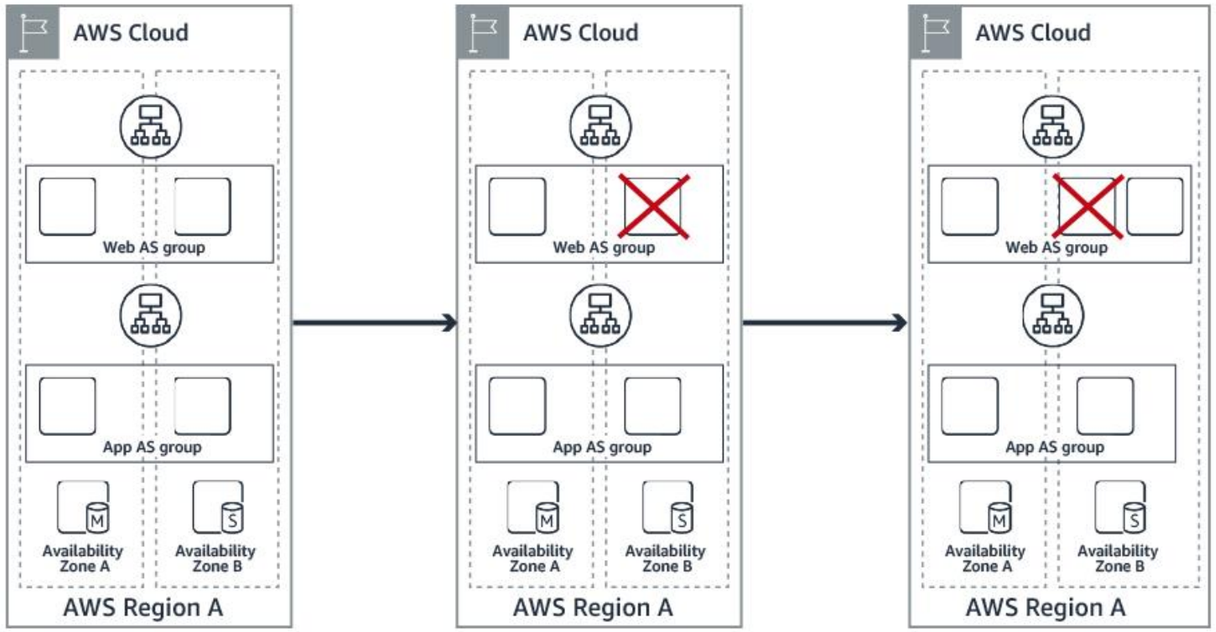


Designing Applications for HA

Highly resilient applications must be able to self-heal

How:

- Leverage Microservices app architecture
- Decouple inter-dependencies, loose coupling
- Remove state from app components



AWS services:

Elastic Load Balancing

AWS Auto Scaling

Amazon Simple Queue Service

Designing Applications for HA

Five common practices we apply to improve availability are following:

- Fault Isolation Zones
    - Make use of multiple independent components in parallel
- Redundant components
    - Avoidance of single points of failure
- Micro-service architecture
    - Differentiate the availability required of different services and reduce dependencies
- Recovery Oriented Computing
    - Minimize the disruption time when failures do occur
- Distributed systems best practices
    - Gracefully handle dependency failures

Data Resiliency

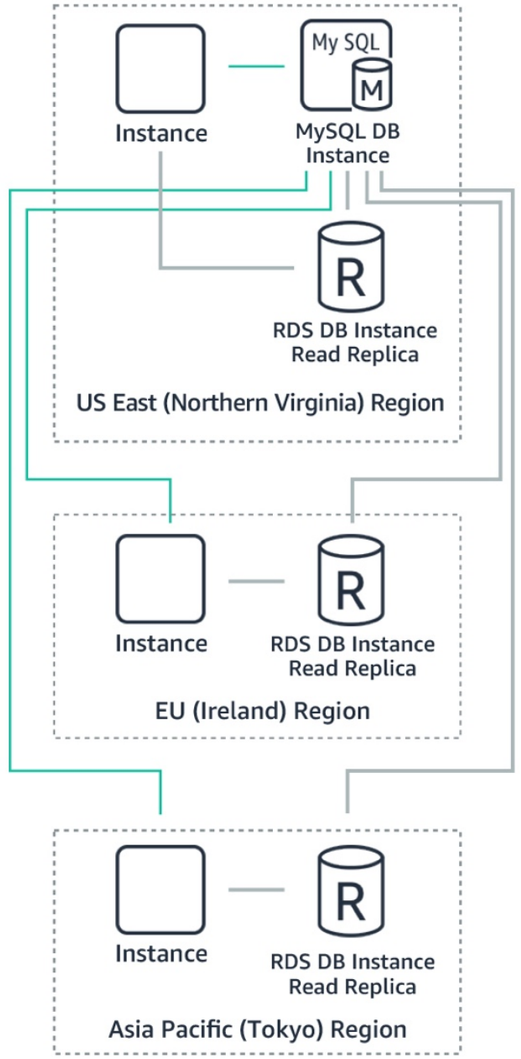
Must have confidence in the resilience of your data

Many forms: filesystem, block storage, databases, in memory caches

Consider how eventual consistency impacts design

AWS services

- Amazon S3 cross-region replication
- Cross region snapshots (Amazon EBS volumes)
- Amazon RDS multi-AZ
- Amazon RDS cross region replicas
- AWS Storage & File Gateway
- Amazon FSx for Windows and Lustre



Data Resiliency – Regional versus AZ services

Service Area	AWS Service Name	Service Capabilities
Database	RDS	Regional
	Redshift	Regional
	DynamoDB	Regional
	ElastiCache	Regional
	DocumentDB	Regional
Compute	EC2 (Instances)	Availability Zone
	EC2 Auto Scaling	Regional
	EC2 (Elastic Load Balancers)	Regional
	Lambda	Regional
	ELB	Availability Zone
Storage	EBS Volumes	Availability Zone
	S3	Regional
	EFS	Regional

Network Design for HA

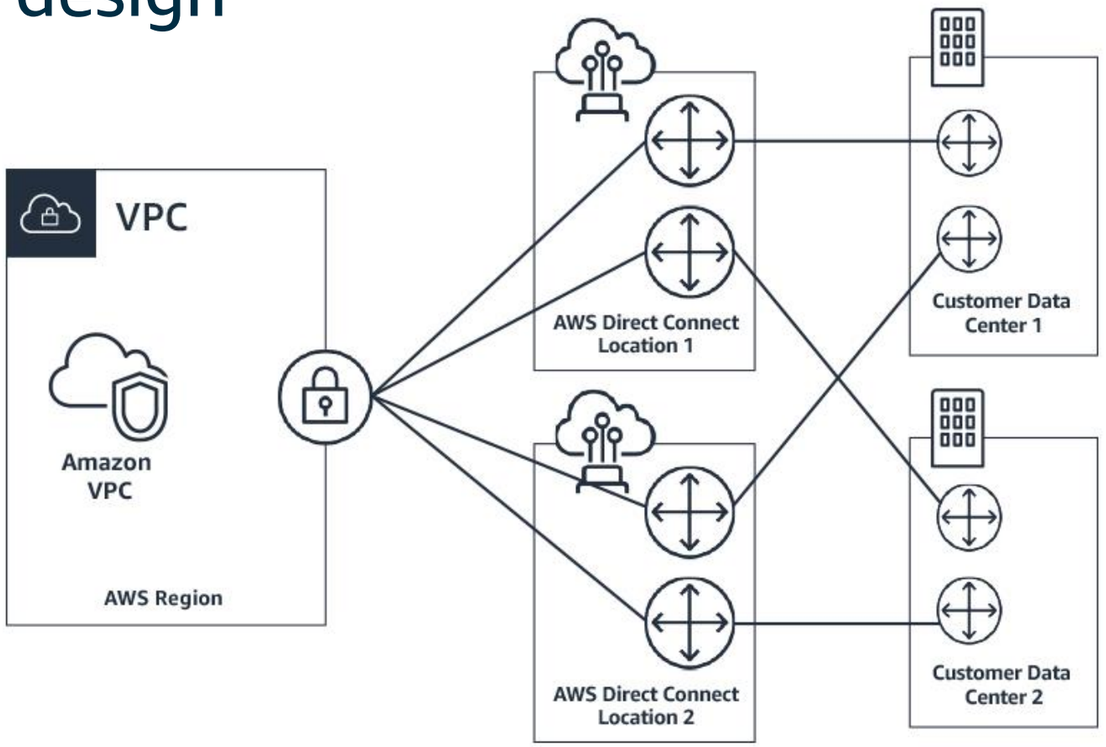
Networking is the foundation of a HA design

Packets must get from point-a to point-b!

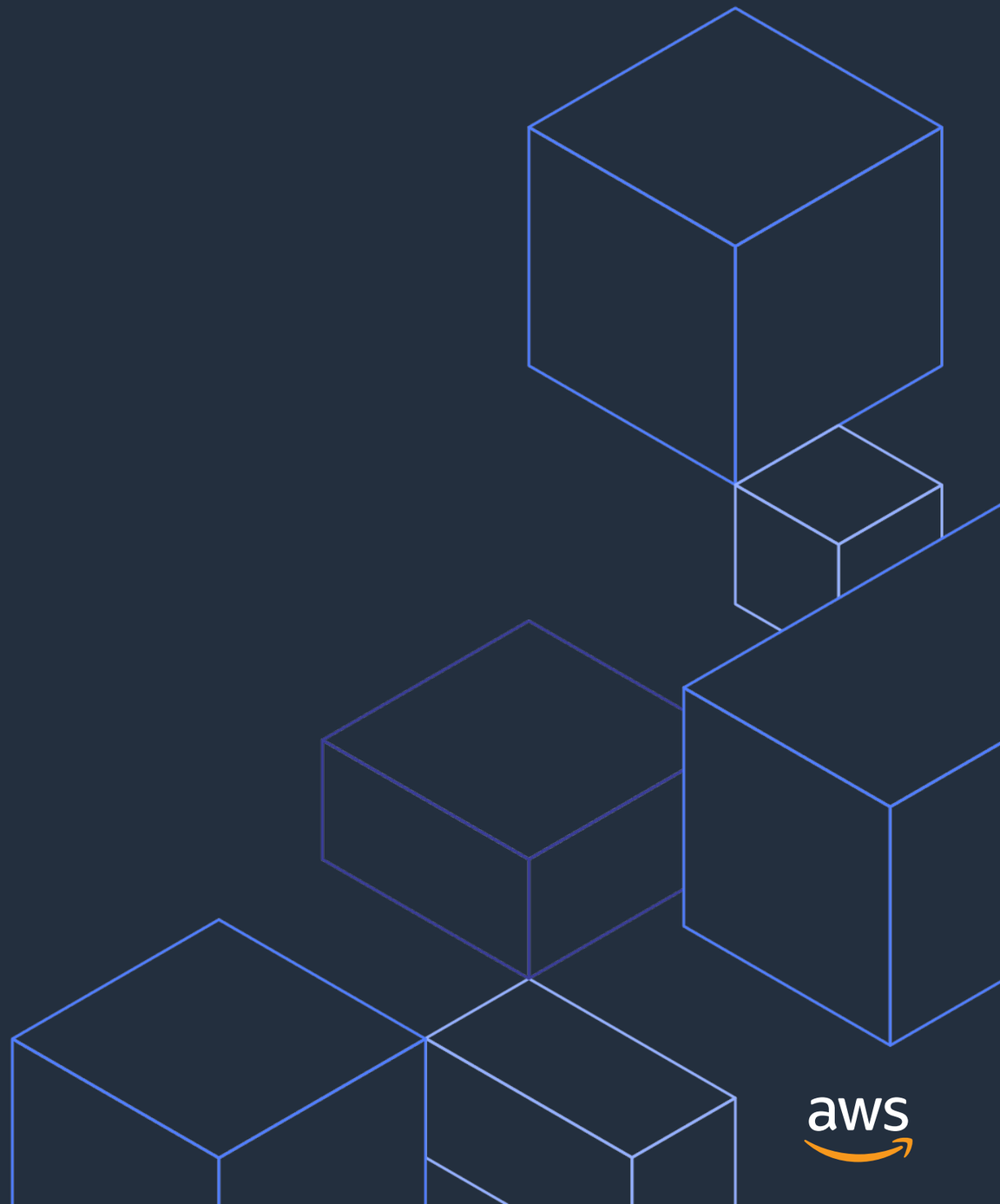
Ensure network supporting your applications is appropriately redundant, always available, and seamlessly routed.

AWS services

- Amazon EC2 networking
- Amazon Virtual Private Cloud (VPC), VPC Peering, VP
  Sharing
- AWS Gateways for external, internal and back to on-premise routing (VPN, Transit)
- DNS (Route53)
- Elastic Load Balancer (ELB)



Design Patterns



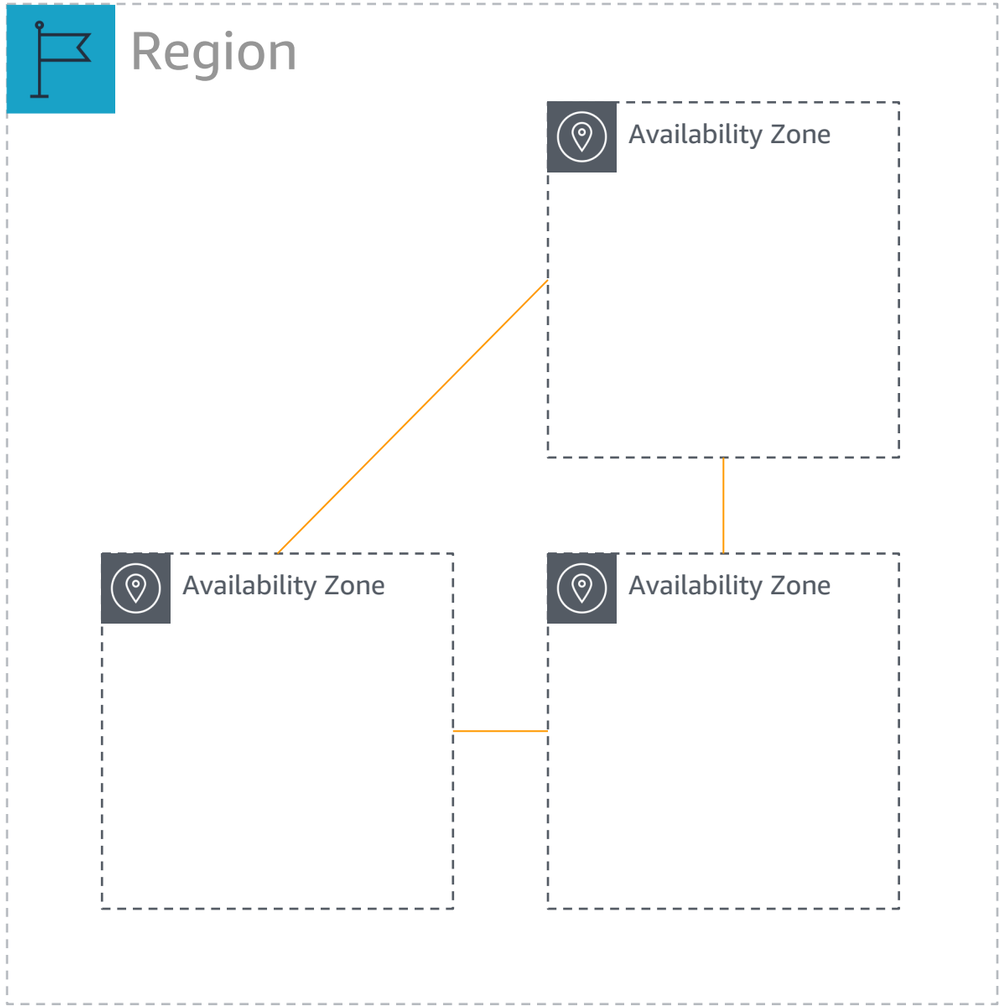
Multi-AZ architecture

- Enables fault-tolerant applications
- AWS Regional services designed to withstand AZ failures
- Leveraged in the Amazon S3 design for 99.999999999% durability

Multi-AZ → Zero blast radius!

Well-Architected Framework

AWS Shared Responsibility Model



Single Region: Multi AZ

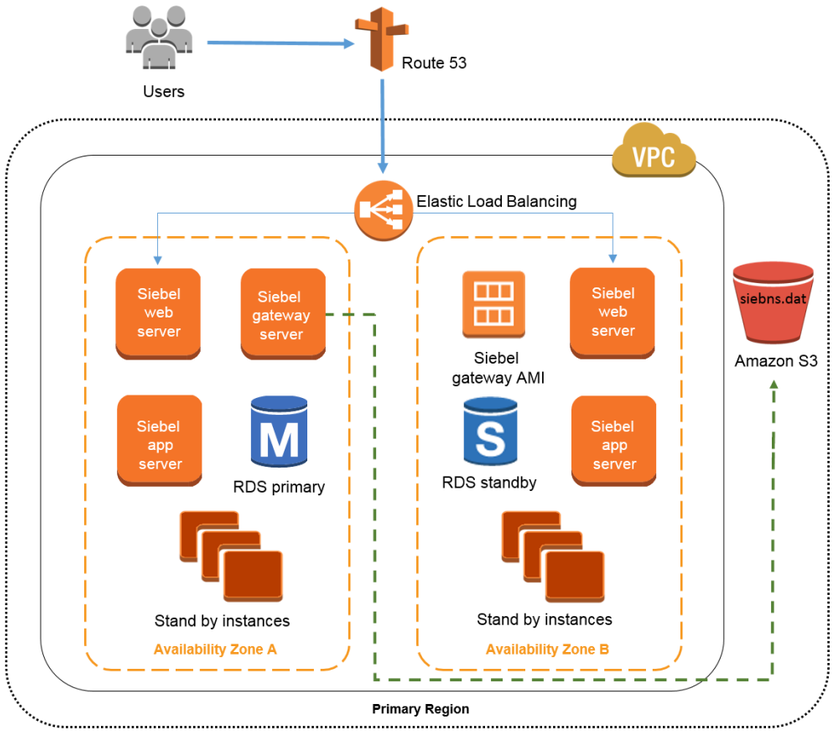
Start here before adopting more complex architecture
Only consider multi-region if requirements dictate

Pros

- Availability of AWS region-wide services include Amazon S3, Amazon DynamoDB, Amazon EFS, Amazon SQS, Amazon Kinesis
- Much less complexity in design, implementation, and operations.

Cons

- If you need >99.9% resiliency, consider multi-region.
- May not meet needs of regulators



Multi-Region: Active-Standby

Traditional DR Pattern
Backup env used in event of failure only

Pros

- For Apps which cannot use native AWS features
- Least # changes to the application

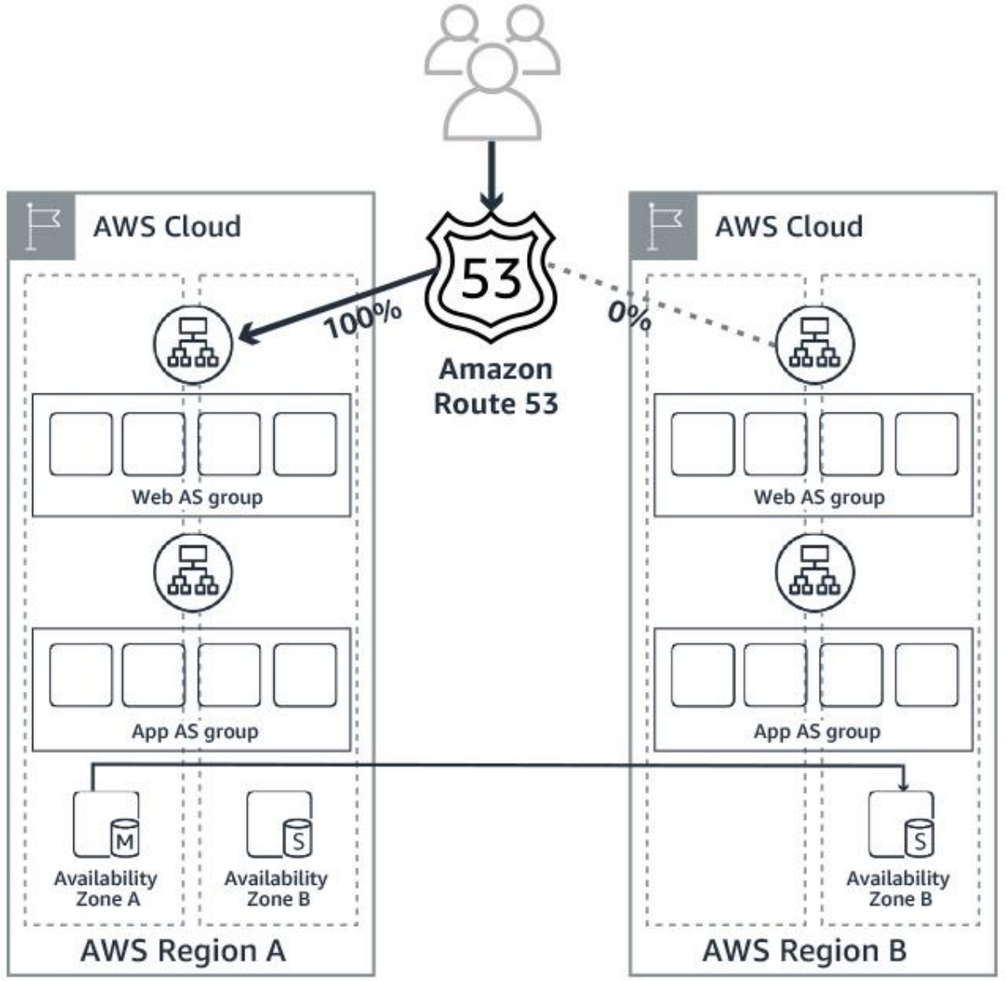
Cons

- Delays while Standby becomes Active (hrs)
- RPO limited by replication lag

AWS Services

Amazon RDS

Amazon Route 53



Multi-Region: Active-active

Both stacks active, traffic distributed
Data replication critical, must consider latency impacts

Pros

- Zero RTO
- Works well for apps that can partition users

Cons

- Data replication must be handled by Applications

AWS Services

- Storage replication from APN partners

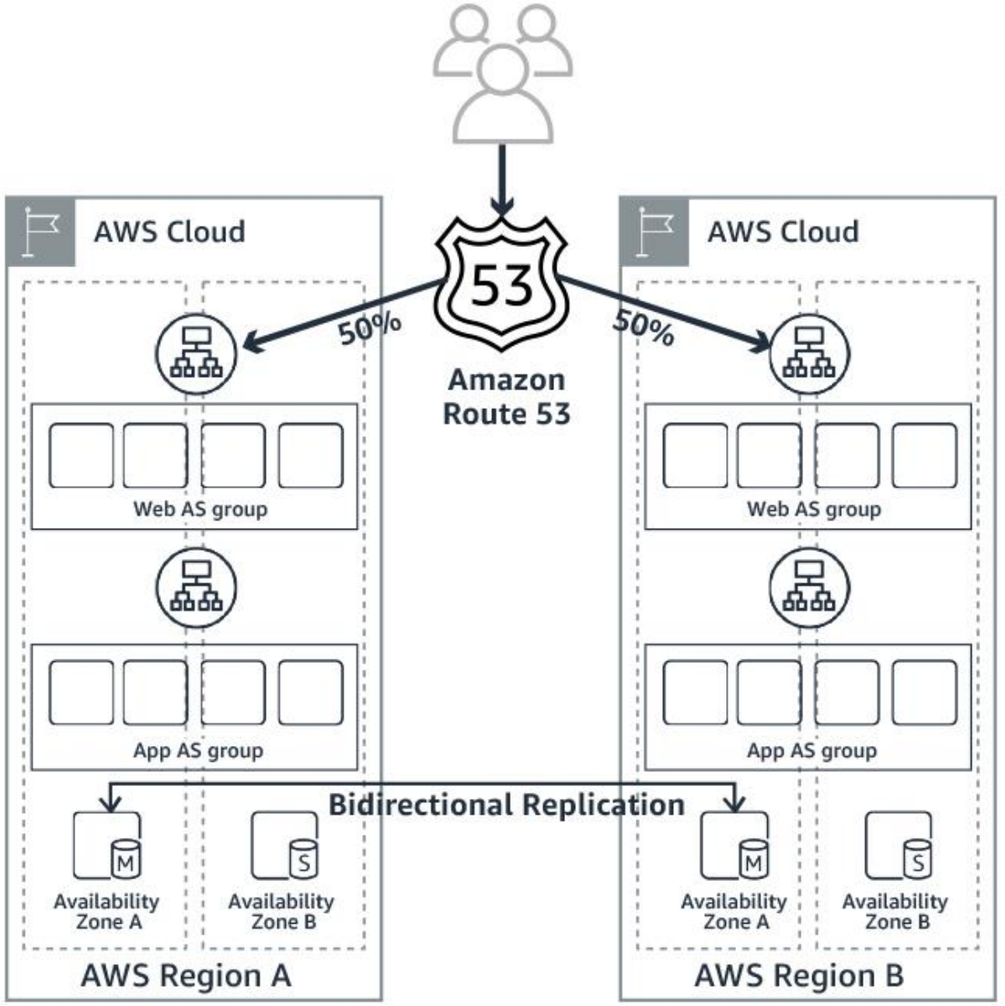

Amazon RDS


Amazon DynamoDB


Amazon Aurora


AWS Database Migration Service


Amazon Route 53



Multi-Region: Dual-write

Shared nothing architecture – all TX processed in duplicate/parallel

Good for legacy applications

Pros

- Zero RPO
- Little/No change to apps in each region

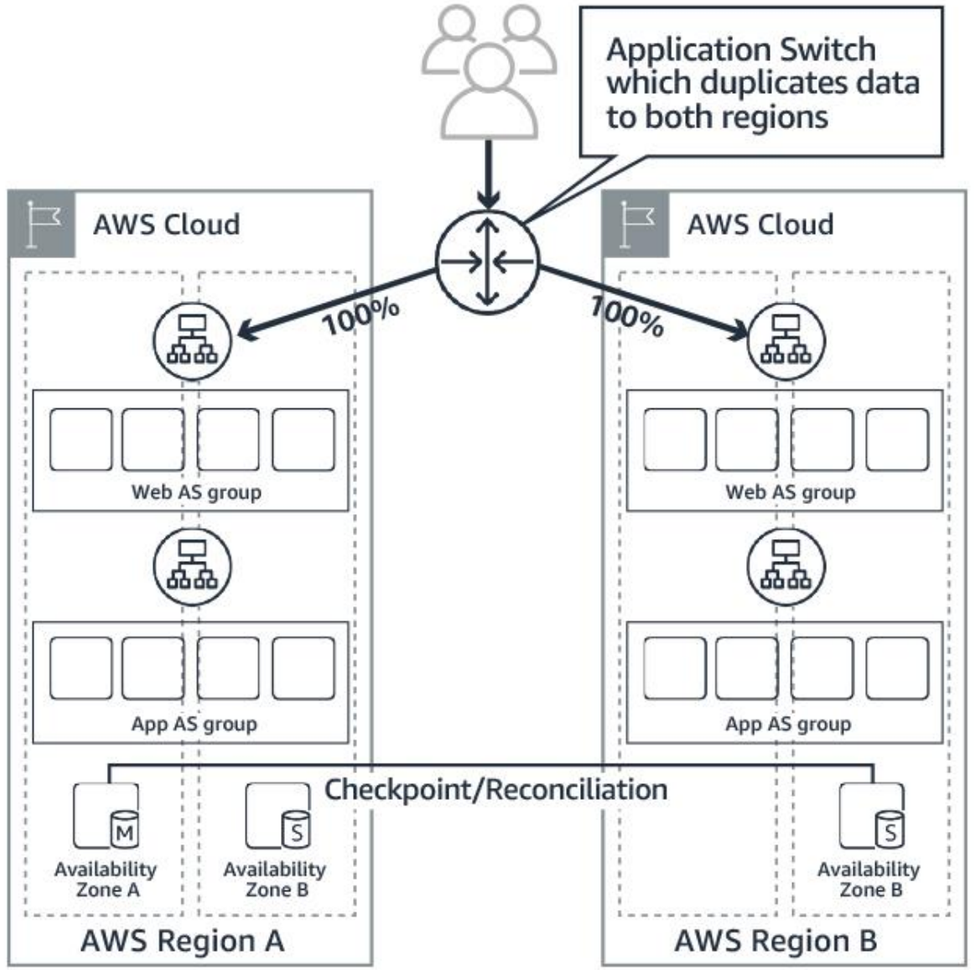
Cons

- Requires checkpointing
- Reconciliation jobs to ensure sites in sync
- Downstream apps must avoid duplicates

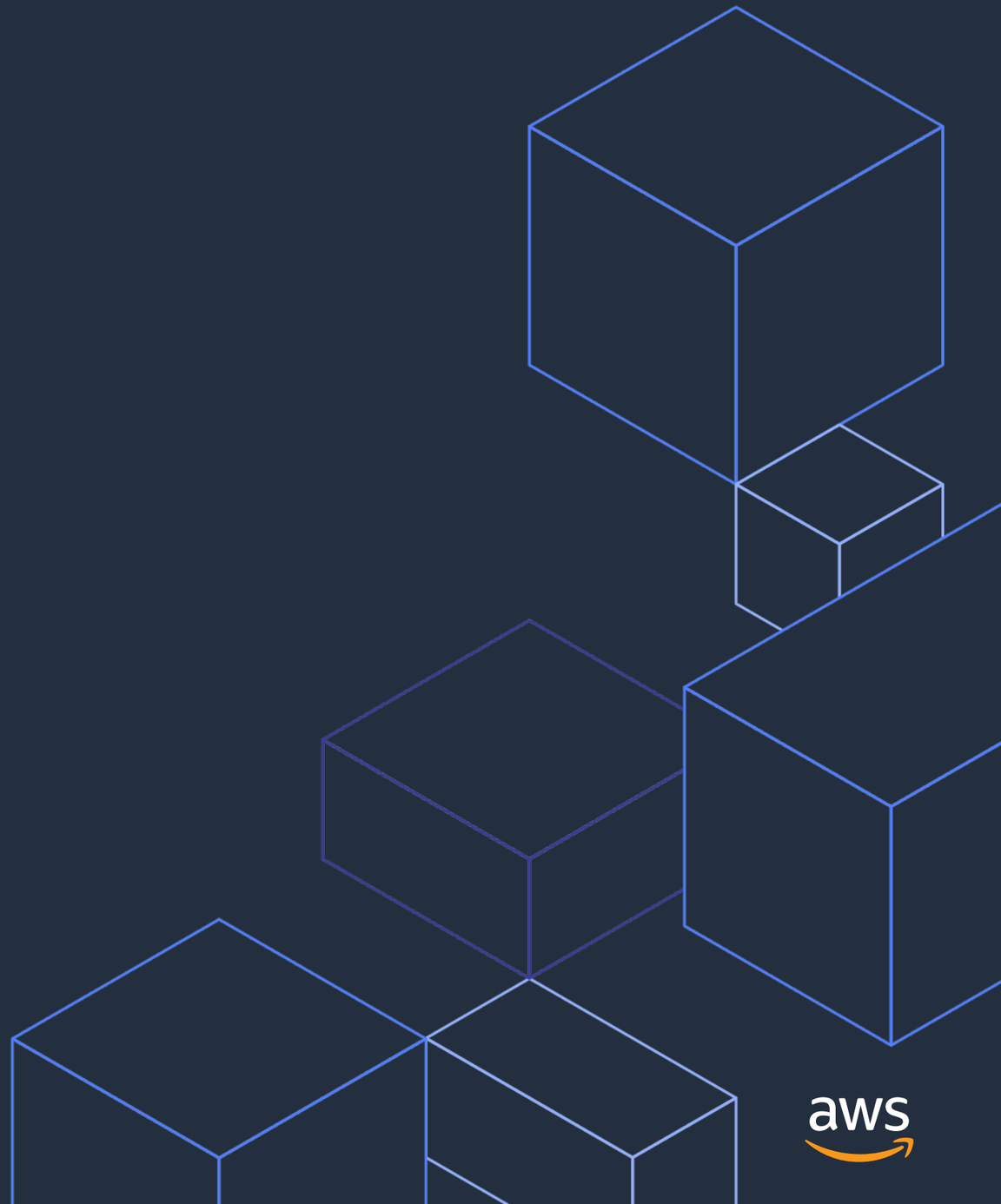
AWS Service

AWS Lambda

Amazon Route 53



Practice Time!



A company is developing a highly available web application using stateless web servers.

Which services are suitable for storing session state data? (Select TWO.)

A. CloudWatch
B. DynamoDB
C. Elastic Load Balancing
D. ElastiCache
E. Storage Gateway

aws

A company is developing a highly available web application using stateless web servers.

Which services are suitable for storing session state data? (Select TWO.)

~~A. CloudWatch~~

B. DynamoDB

~~C. Elastic Load Balancing~~

D. ElastiCache

~~E. Storage Gateway~~

aws

A company is developing a highly available web application using stateless web servers.

Which services are suitable for storing session state data? (Select TWO.)

B. DynamoDB
D. ElastiCache

B, D – Both DynamoDB and ElastiCache provide high performance storage of key-value pairs. CloudWatch and ELB are not storage services. Storage Gateway is a storage service, but it is a hybrid storage service that enables on-premises applications to use cloud storage.

aws

Company salespeople upload their sales figures daily. A Solutions Architect needs a durable storage solution for these documents that also protects against users accidentally deleting important documents.

Which action will protect against unintended user actions?

A. Store data in an EBS volume and create snapshots once a week.
B. Store data in an S3 bucket and enable versioning.
C. Store data in two S3 buckets in different AWS regions.
D. Store data on EC2 instance storage.

aws

Company salespeople upload their sales figures daily. A Solutions Architect needs a durable storage solution for these documents that also protects against users accidentally deleting important documents.

Which action will protect against unintended user actions?

~~A. Store data in an EBS volume and create snapshots once a week.~~

B. Store data in an S3 bucket and enable versioning.

~~C. Store data in two S3 buckets in different AWS regions.~~

~~D. Store data on EC2 instance storage.~~

aws

Company salespeople upload their sales figures daily. A Solutions Architect needs a durable storage solution for these documents that also protects against users accidentally deleting important documents.

Which action will protect against unintended user actions?

B. Store data in an S3 bucket and enable versioning.

B – If a versioned object is deleted, then it can still be recovered by retrieving the final version. Response A would lose any changes committed since the previous snapshot. Storing the data in 2 S3 buckets would provide slightly more protection, but a user could still delete the object from both buckets. EC2 instance storage is ephemeral and should never be used for data requiring durability.

aws

A customer relationship management (CRM) application runs on Amazon EC2 instances in multiple Availability Zones behind an Application Load Balancer.

If one of these instances fails, what occurs?

A) The load balancer will stop sending requests to the failed instance.
B) The load balancer will terminate the failed instance.
C) The load balancer will automatically replace the failed instance.
D) The load balancer will return 504 Gateway Timeout errors until the instance is replaced

aws

A customer relationship management (CRM) application runs on Amazon EC2 instances in multiple Availability Zones behind an Application Load Balancer.

If one of these instances fails, what occurs?

A) The load balancer will stop sending requests to the failed instance.

~~B) The load balancer will terminate the failed instance.~~

~~C) The load balancer will automatically replace the failed instance.~~

~~D) The load balancer will return 504 Gateway Timeout errors until the instance is replaced~~

aws

A customer relationship management (CRM) application runs on Amazon EC2 instances in multiple Availability Zones behind an Application Load Balancer.

If one of these instances fails, what occurs?

A) The load balancer will stop sending requests to the failed instance.

A – An Application Load Balancer (ALB) sends requests to healthy instances only. An ALB performs periodic health checks on targets in a target group. An instance that fails health checks for a configurable number of consecutive times is considered unhealthy. The load balancer will no longer send requests to the instance until it passes another health check.

aws

A company runs a public-facing three-tier web application in a VPC across multiple Availability Zones. Amazon EC2 instances for the application tier running in private subnets need to download software patches from the internet. However, the instances cannot be directly accessible from the internet.

Which actions should be taken to allow the instances to download the needed patches? (Select TWO.)

A) Configure a NAT gateway in a public subnet.

B) Define a custom route table with a route to the NAT gateway for internet traffic and associate it with the private subnets for the application tier.

C) Assign Elastic IP addresses to the application instances.

D) Define a custom route table with a route to the internet gateway for internet traffic and associate it with the private subnets for the application tier.

E) Configure a NAT instance in a private subnet.

aws

A company runs a public-facing three-tier web application in a VPC across multiple Availability Zones. Amazon EC2 instances for the application tier running in private subnets need to download software patches from the internet. However, the instances cannot be directly accessible from the internet.

Which actions should be taken to allow the instances to download the needed patches? (Select TWO.)

A) Configure a NAT gateway in a public subnet.

B) Define a custom route table with a route to the NAT gateway for internet traffic and associate it with the private subnets for the application tier.

~~C) Assign Elastic IP addresses to the application instances.~~

~~D) Define a custom route table with a route to the internet gateway for internet traffic and associate it with the private subnets for the application tier.~~

~~E) Configure a NAT instance in a private subnet.~~

aws

A company runs a public-facing three-tier web application in a VPC across multiple Availability Zones. Amazon EC2 instances for the application tier running in private subnets need to download software patches from the internet. However, the instances cannot be directly accessible from the internet.

Which actions should be taken to allow the instances to download the needed patches? (Select TWO.)

A) Configure a NAT gateway in a public subnet.

B) Define a custom route table with a route to the NAT gateway for internet traffic and associate it with the private subnets for the application tier.

A,B - A NAT gateway forwards traffic from the instances in the private subnet to the internet or other AWS services, and then sends the response back to the instances. After a NAT gatew

aws

A company plans to run a monitoring application on an Amazon EC2 instance in a VPC. Connections are made to the instance using its private IPv4 address. A solutions architect needs to design a solution that will allow traffic to be quickly directed to a standby instance if the application fails and becomes unreachable.

Which approach will meet these requirements?

A) Deploy an Application Load Balancer configured with a listener for the private IP address and register the primary instance with the load balancer. Upon failure, de-register the instance and register the secondary instance.

B) Configure a custom DHCP option set. Configure DHCP to assign the same private IP address to the secondary instance when the primary instance fails.

C) Attach a secondary elastic network interface (ENI) to the instance configured with the private IP address. Move the ENI to the standby instance if the primary instance becomes unreachable.

D) Associate an Elastic IP address with the network interface of the primary instance. Disassociate the Elastic IP from the primary instance upon failure and associate it with a secondary instance.

aws

A company plans to run a monitoring application on an Amazon EC2 instance in a VPC. Connections are made to the instance using its private IPv4 address. A solutions architect needs to design a solution that will allow traffic to be quickly directed to a standby instance if the application fails and becomes unreachable.

Which approach will meet these requirements?

~~A) Deploy an Application Load Balancer configured with a listener for the private IP address and register the primary instance with the load balancer. Upon failure, de-register the instance and register the secondary instance.~~

~~B) Configure a custom DHCP option set. Configure DHCP to assign the same private IP address to the secondary instance when the primary instance fails.~~

C) Attach a secondary elastic network interface (ENI) to the instance configured with the private IP address. Move the ENI to the standby instance if the primary instance becomes unreachable.

~~D) Associate an Elastic IP address with the network interface of the primary instance. Disassociate the Elastic IP from the primary instance upon failure and associate it with a secondary instance.~~

aws

A company plans to run a monitoring application on an Amazon EC2 instance in a VPC. Connections are made to the instance using its private IPv4 address. A solutions architect needs to design a solution that will allow traffic to be quickly directed to a standby instance if the application fails and becomes unreachable.

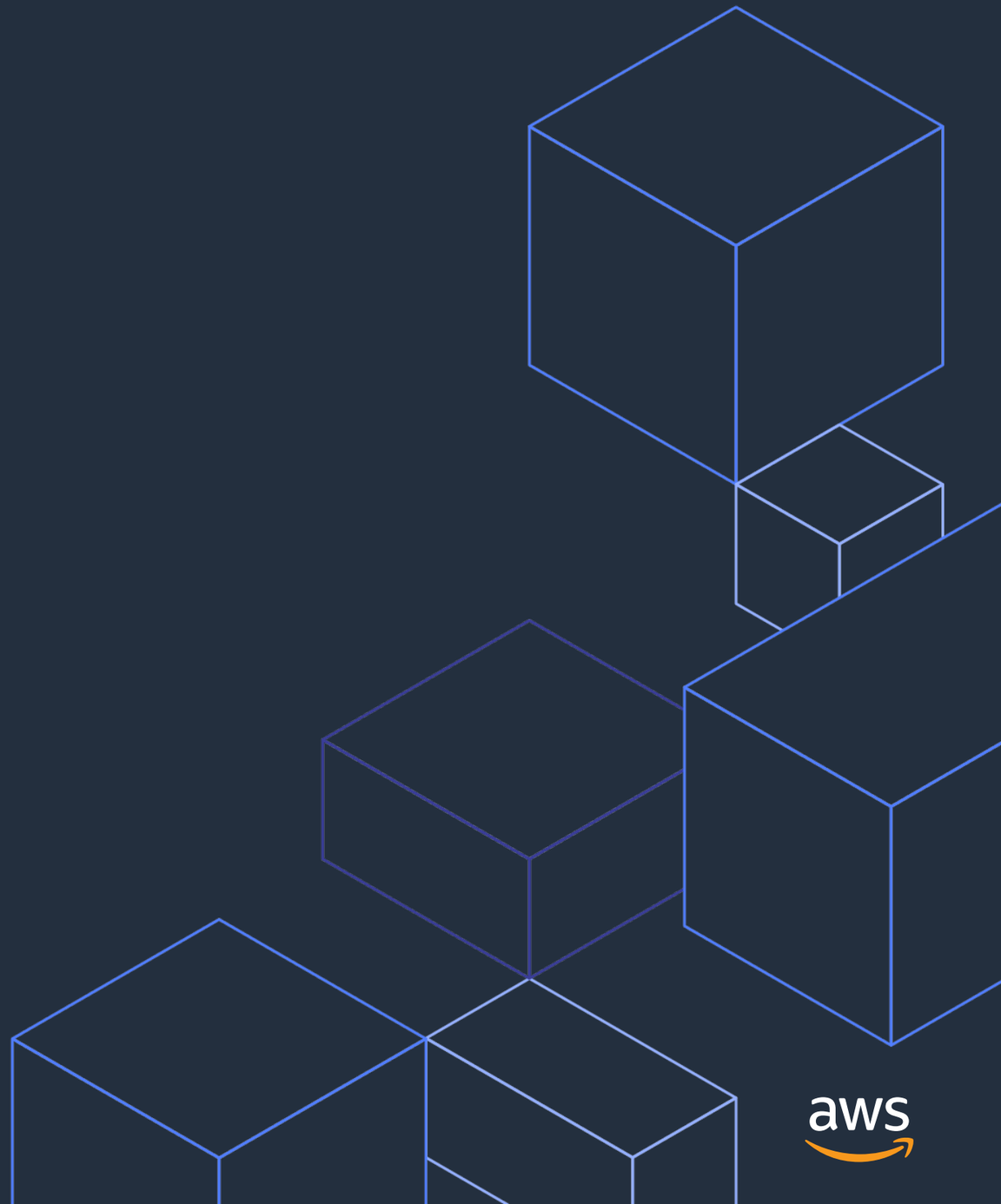
Which approach will meet these requirements?

C) Attach a secondary elastic network interface (ENI) to the instance configured with the private IP address. Move the ENI to the standby instance if the primary instance becomes unreachable.

C – A secondary ENI can be added to an instance. While primary ENIs cannot be detached from an instance, secondary ENIs can be detached and attached to a different instance.

aws

Next Week



Next Week Sessions

Session Topic – Infrastructure Automation

Online Modules:

Module 5: Automating Your Infrastructure

Supplementary Content:

- AWS Whitepaper - Practicing Continuous Integration and Continuous Delivery on AWS https://d0.awsstatic.com/whitepapers/DevOps/practicing-continuous-integration-continuous-delivery-on-AWS.pdf
- AWS Samples – Cloudformation Examples https://aws.amazon.com/cloudformation/templates/aws-cloudformation-templates-us-west-1/

aws

Questions?