AWS Solutions Architect - Associate Certification Crash Course

SAA-C03

Chad Smith

Principal Cloud Architect

Pearson

Exam Guide and General Strategies

Exam Logistics - By the Numbers

Number of questions:	**65**
Time for exam	**130 minutes**
Answer choices	**4-6**
Score required	**720/1000**
Number of unscored questions	**15**
Partial Credit	**0**
Penalty for guessing	**0**

Exam Guide Introduction

Validates a candidate's ability to:

Use AWS technologies to ***design*** solutions based on the AWS Well-Architected Framework

Exam Guide Introduction

Validates a candidate's ability to:

Design solutions that incorporate AWS services to *meet current business requirements* and *future projected needs*

Exam Guide Introduction

Validates a candidate's ability to:

Design architectures that are *secure*, *resilient*, *high-performing*, and *cost-optimized*

Exam Guide Introduction

Validates a candidate's ability to:

Review existing solutions and ***determine improvements***

Exam Guide Target Candidate Description

At least 1 year of ***hands-on*** experience designing cloud solutions that use AWS services

Exam Guide Exam Content

Question Domains	%
Design Secure Architectures	30
Design Resilient Architectures	26
Design High-Performing Architectures	24
Design Cost-Optimized Architectures	20

Well-Architected Framework

Introduction to Pillars

Well-Architected Framework

Learn how to design, use, and manage workloads in the cloud.

Learn how to translate requirements into architecture and operations while following best practices.

Well-Architected Framework

Learn how to design, use, and manage workloads in the cloud.

Learn how to translate requirements into architecture and operations while following best practices.

- Operational Excellence
- Security
- Reliability
- Performance Efficiency
- Cost Optimization
- Sustainability

Operational Excellence

The ability to support development and run workloads effectively, gain insight into their operations, and to continuously improve supporting processes and procedures to deliver business value.

Operational Excellence

- Organize teams around business outcomes
- Implement observability for actionable insights
- Safely automate where possible
- Make frequent, small, reversible changes
- Refine operations procedures frequently
- Anticipate failure
- Learn from all operational events and metrics
- Use managed services

Performance Efficiency

The ability to use computing resources efficiently to meet system requirements, and to maintain that efficiency as demand changes and technologies evolve.

Performance Efficiency

Democratize advanced technologies

Go global in minutes

Use serverless architectures

Experiment more often

Mechanical sympathy

Security

The ability to protect data, systems, and assets to take advantage of cloud technologies to improve your security.

Security

- Implement a strong identity foundation
- Enable traceability
- Apply security at all layers
- Automate security best practices
- Protect data in transit and at rest
- Keep people away from data
- Prepare for security events

Reliability

The ability of a workload to perform its intended function correctly and consistently when it's expected to. This includes the ability to operate and test the workload through its total lifecycle.

Reliability

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

Cost Optimization

The ability to run systems to deliver business value at the lowest price point.

Cost Optimization

- Implement cloud financial management
- Adopt a consumption model
- Measure overall efficiency
- Stop spending money on undifferentiated heavy lifting
- Analyze and attribute expenditure

Sustainability

Ability to focus on environmental impacts, especially energy consumption and efficiency, since they are important levers for architects to inform direct action to reduce resource usage.

Sustainability

- Understand your impact
- Establish sustainability goals
- Maximize utilization
- Anticipate and adopt new, more efficient hardware and software offerings
- Use managed services
- Reduce the downstream impact of your cloud workloads

Question Domain 1: Design Secure Architectures

Question Domain 1: Design Secure Architectures

Implement a strong identity foundation

Principle Definition

Implement the principle of least privilege and enforce separation of duties with appropriate authorization for each interaction with your AWS resources.

Centralize identity management, and aim to eliminate reliance on long-term static credentials.

Least Privilege - RBAC and ABAC

Role-Based Access Control

Access based on identity

- IAM users or federation
- Group membership
- Instance profiles

Least Privilege - RBAC and ABAC

Attribute-Based Access Control

Access based on properties (tags)

- Policy conditions
- Principal tags
- Resource tags

Authenticating to AWS

Static identity

Includes IAM users

Authenticating to AWS

Static identity

Temporary identity

Federation and IAM roles

Authenticating to AWS

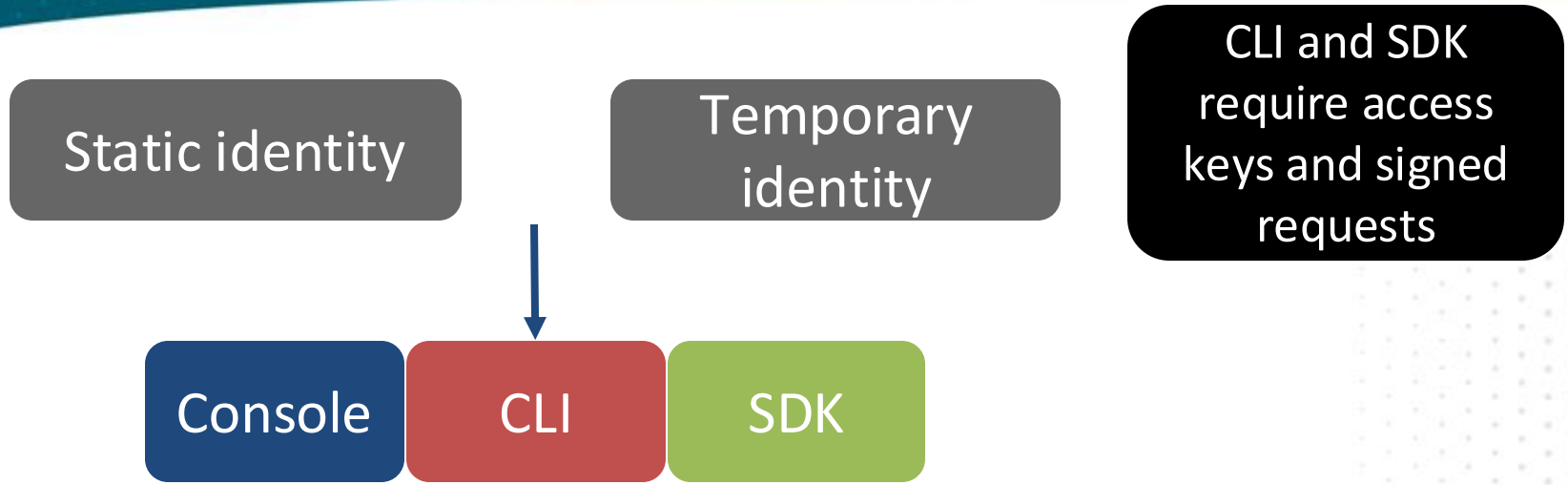
Static identity

Temporary identity

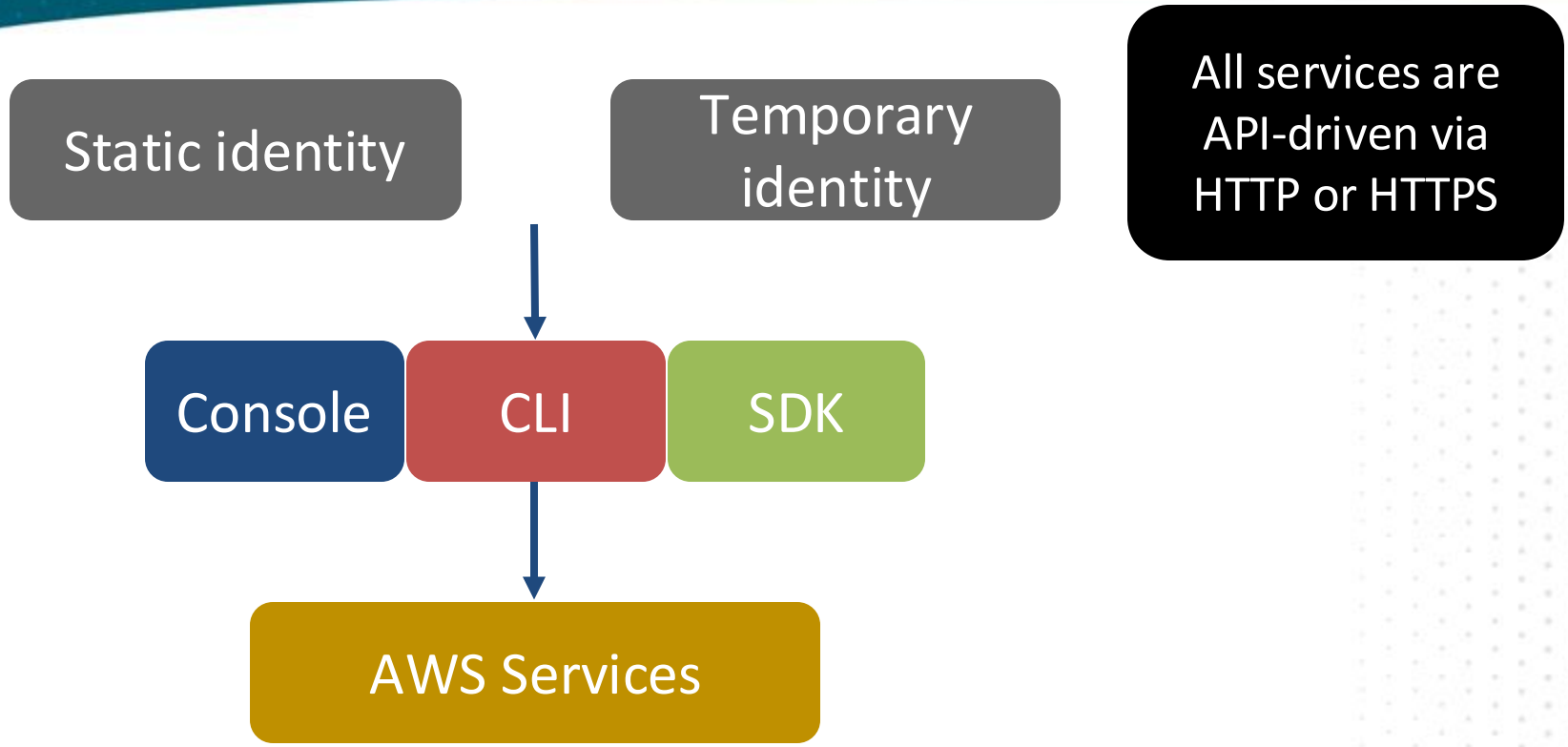
Enforce MFA for browser access, especially root account

Console

Authenticating to AWS

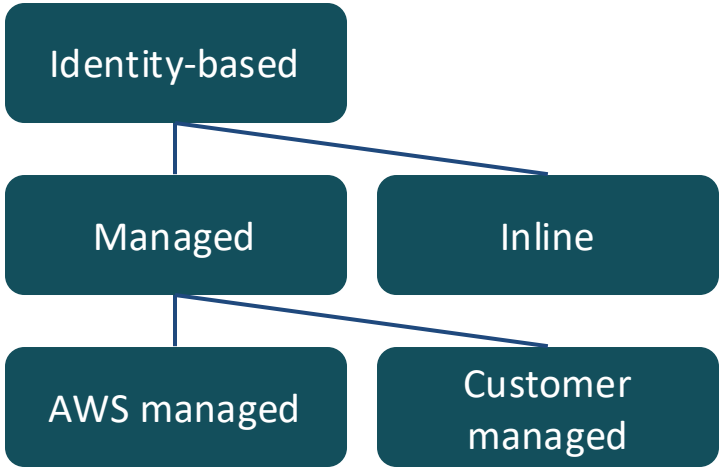


Authenticating to AWS



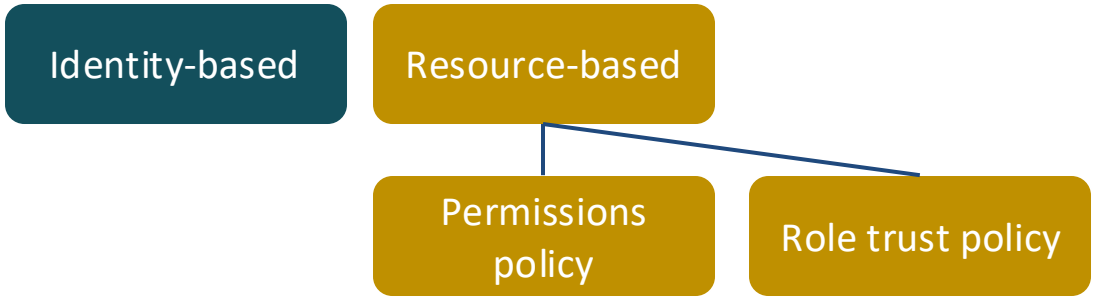
AWS Policy Types

Attached to IAM identities



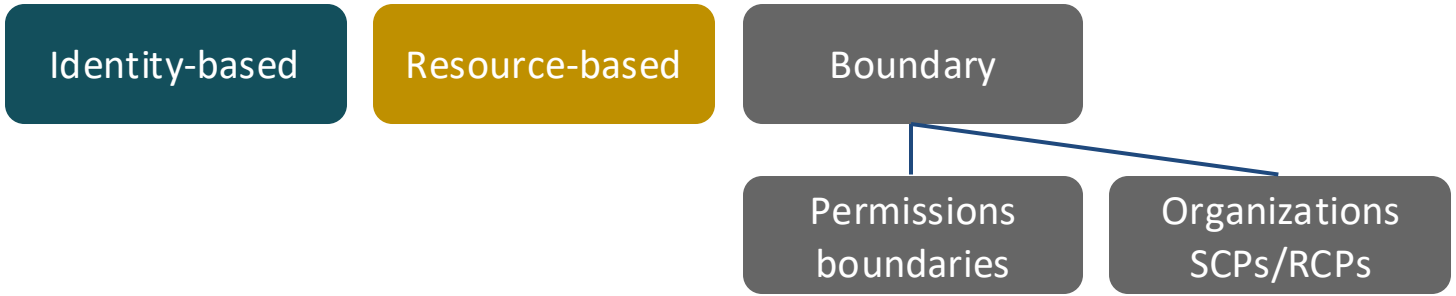
AWS Policy Types

Attached to resources (not supported by all services)



AWS Policy Types

Defines maximum permissions for a principal or account



AWS Policy Types

Limit permissions while assuming temporary credentials

Identity-based

Resource-based

Boundary

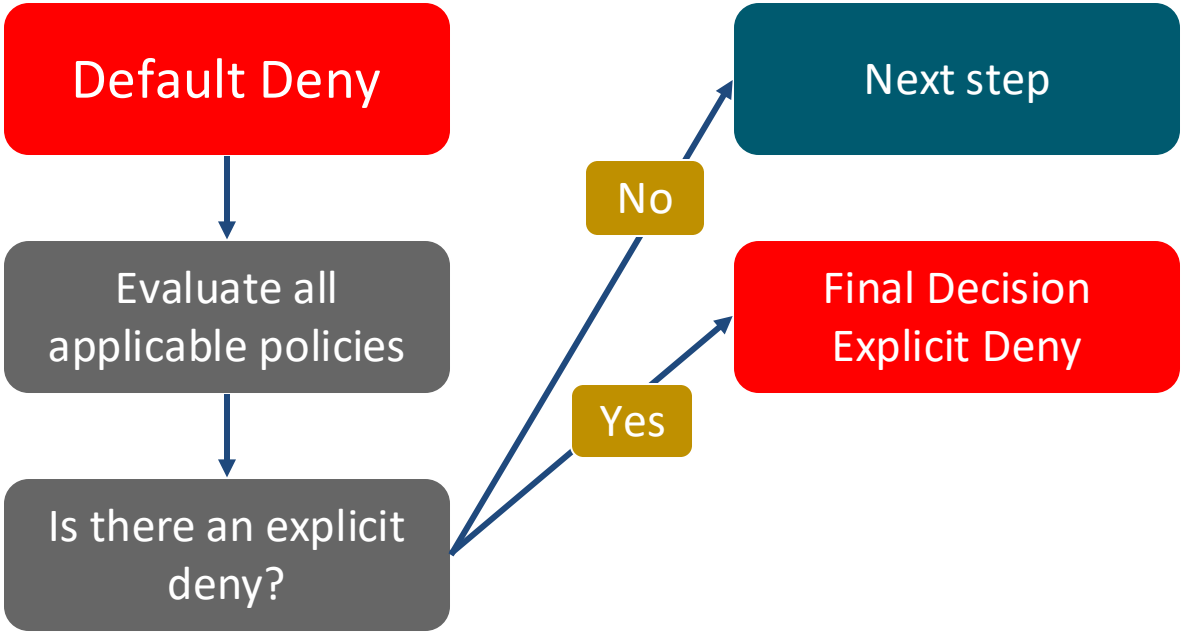
Session Policies

AWS Policy Types

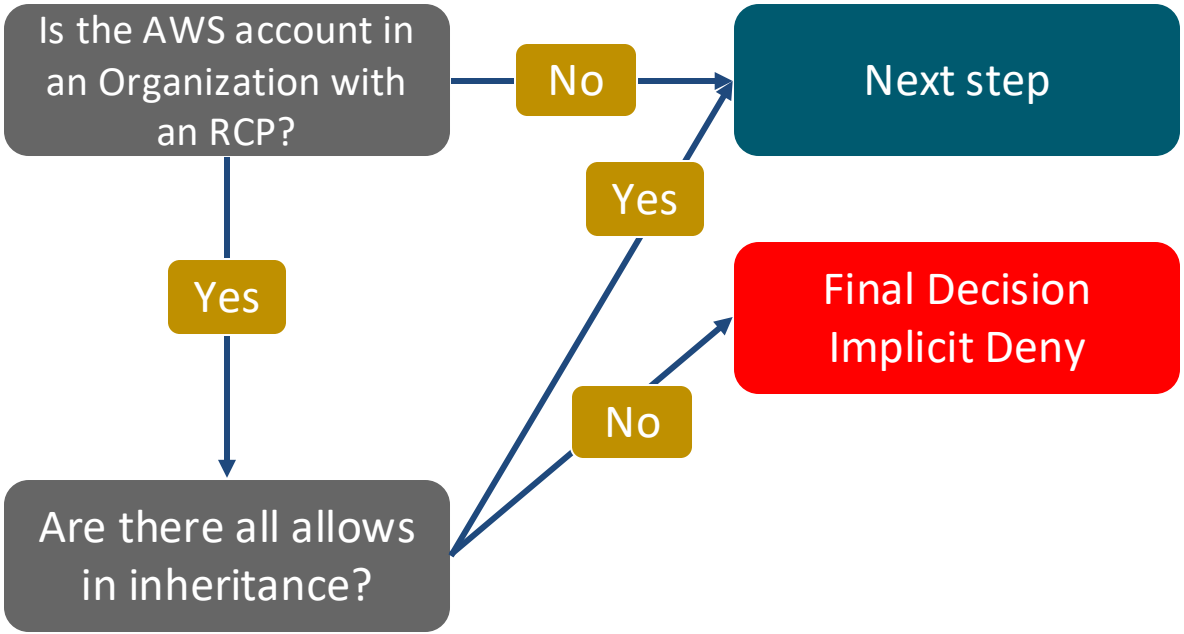
Similar to resource-based policies but does not use JSON, S3-only support

- Identity-based
- Resource-based
- Boundary
- Session Policies
- Access Control Lists (ACLs)

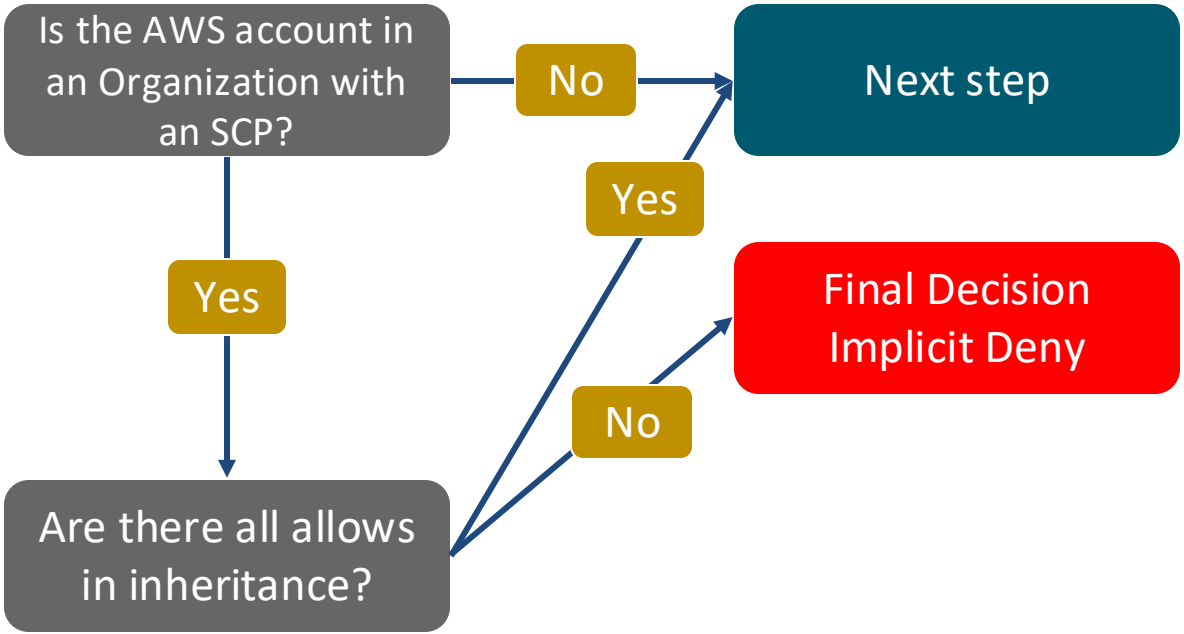
AWS Single Account Policy Evaluation Logic 1



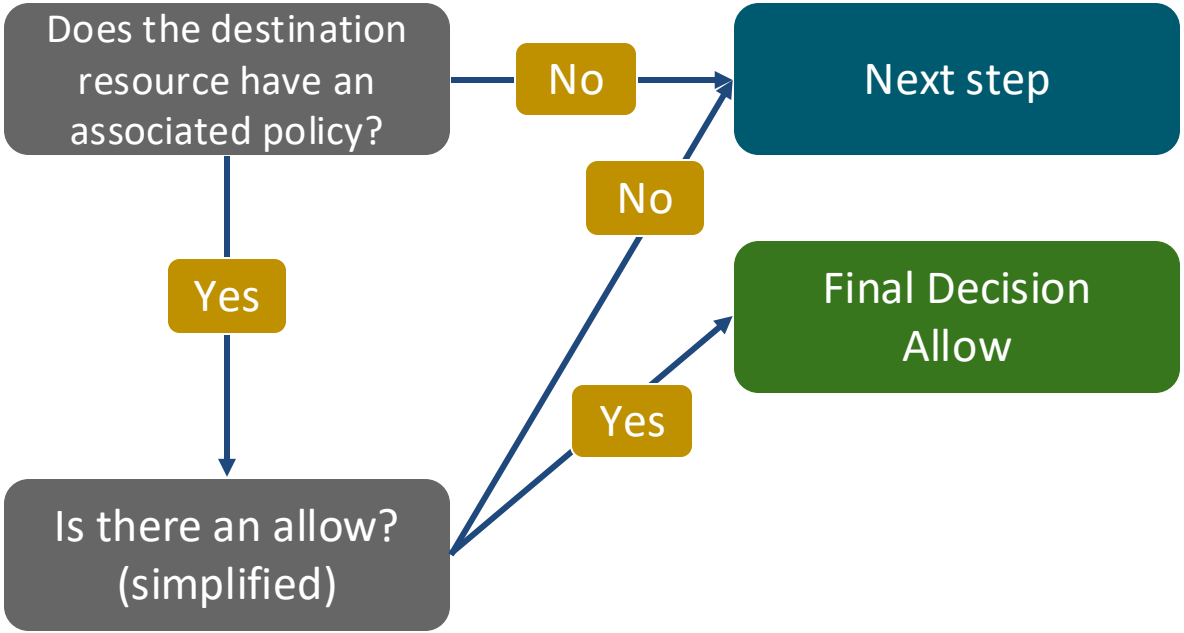
AWS Single Account Policy Evaluation Logic 2



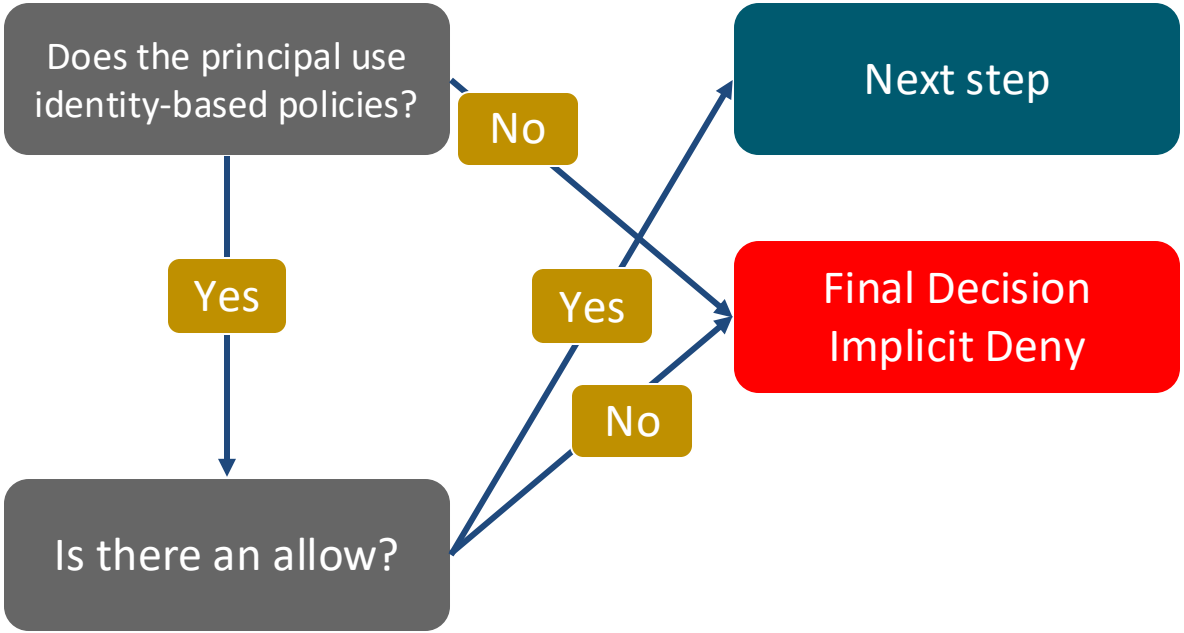
AWS Single Account Policy Evaluation Logic 3



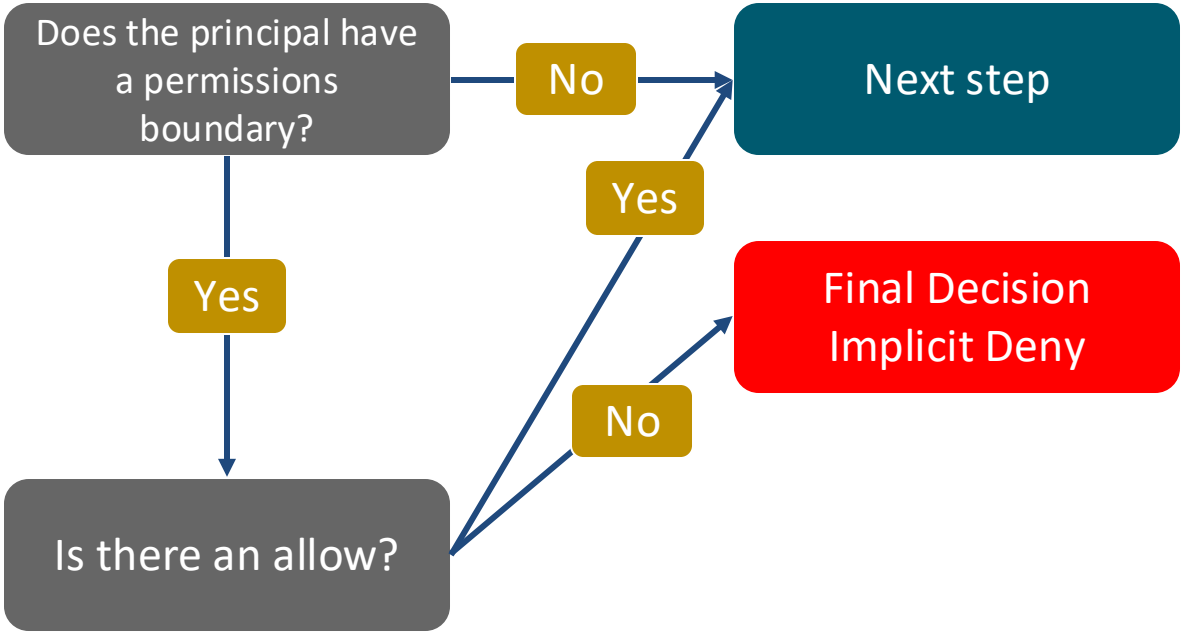
AWS Single Account Policy Evaluation Logic 4



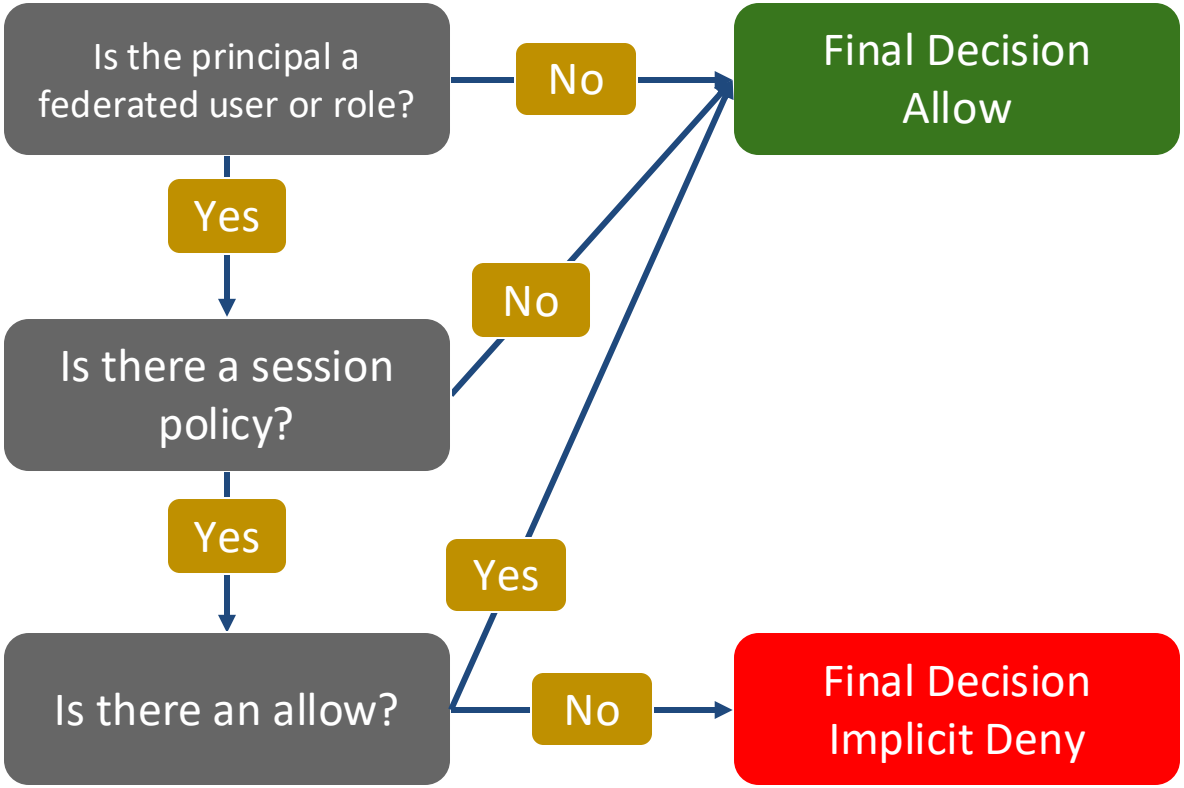
AWS Single Account Policy Evaluation Logic 5



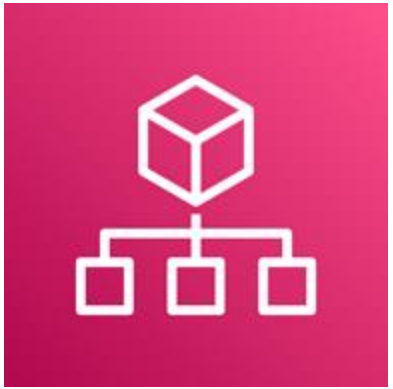
AWS Single Account Policy Evaluation Logic 6



AWS Single Account Policy Evaluation Logic 7



Multiple Accounts - AWS Organizations



- Account scope
- Multiple account management service
- Organizational Unit (OU structure)
- Central billing
- Central policy management

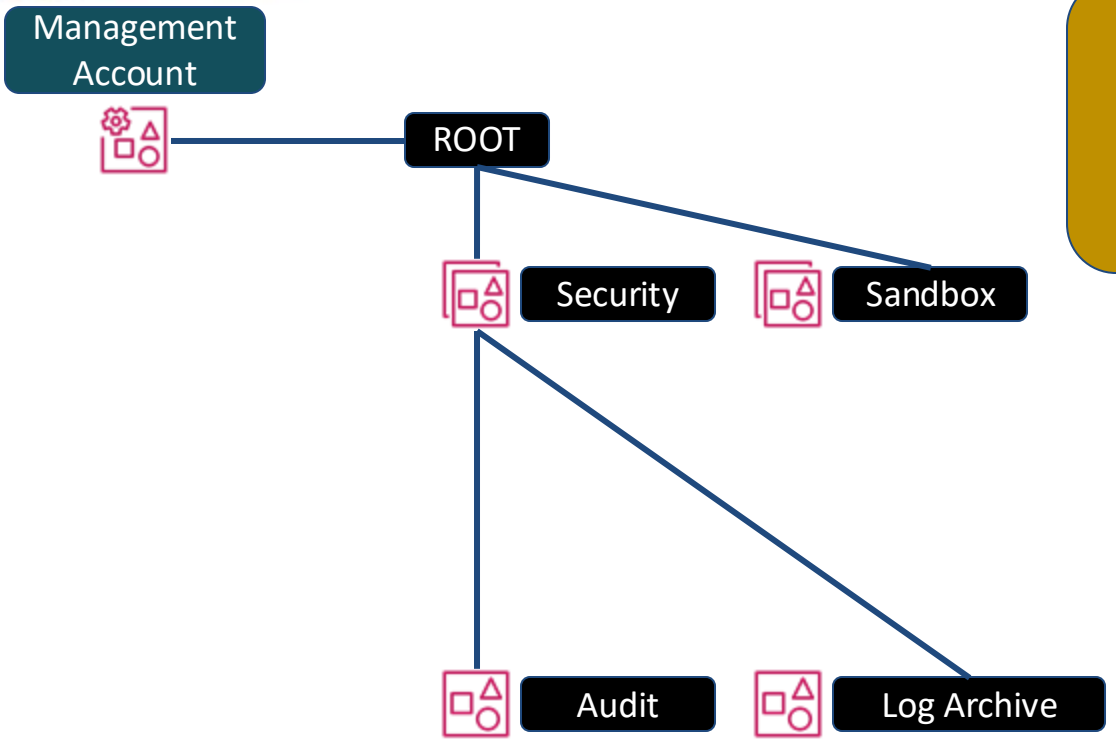
Multiple Accounts - AWS Control Tower

- All of Organizations features plus:
- Account templates
- Guardrails

Control Tower Basics

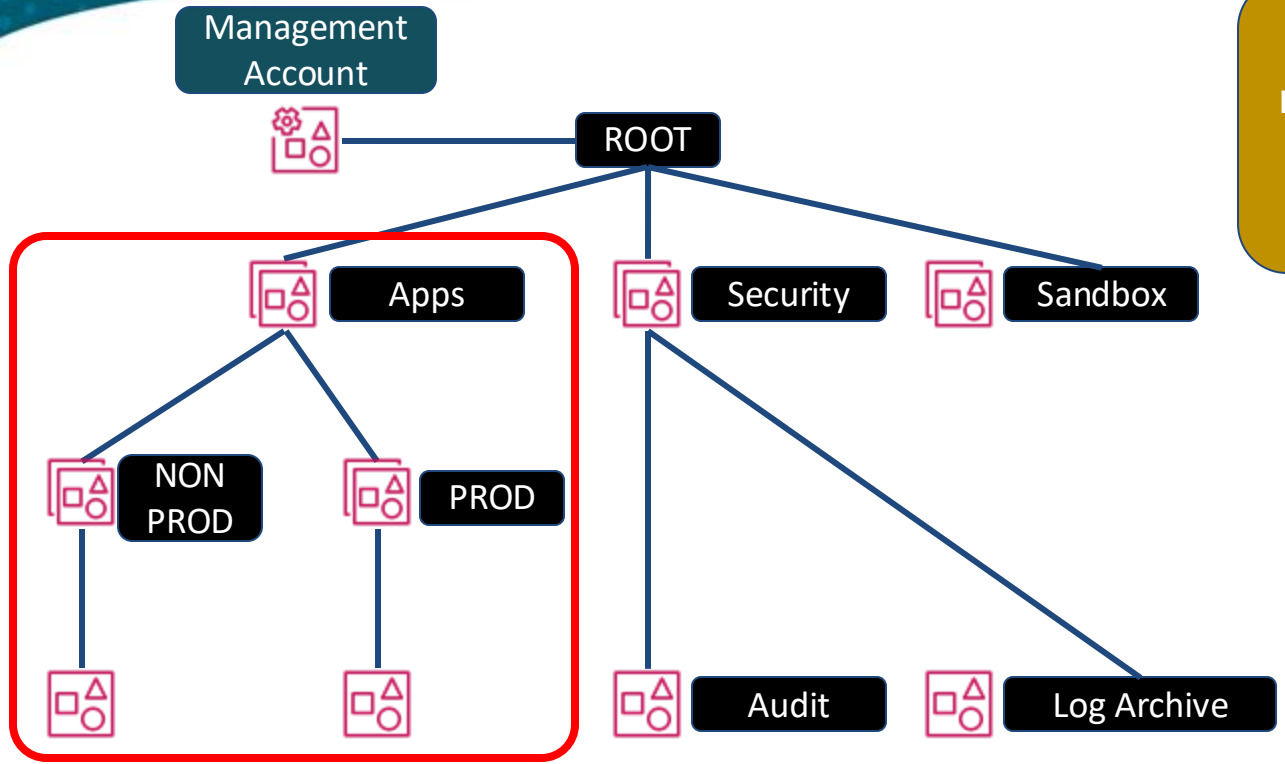
- Built on top of Organizations
- IAM Identity Center
- Landing Zone
- Account Factory
- Guardrails
- Central dashboard

Control Tower Landing Zone Resources



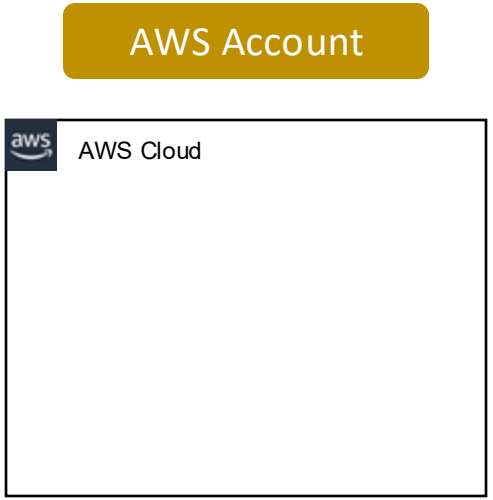
Two OUs and two accounts are created for the Landing Zone

Control Tower Landing Zone Resources



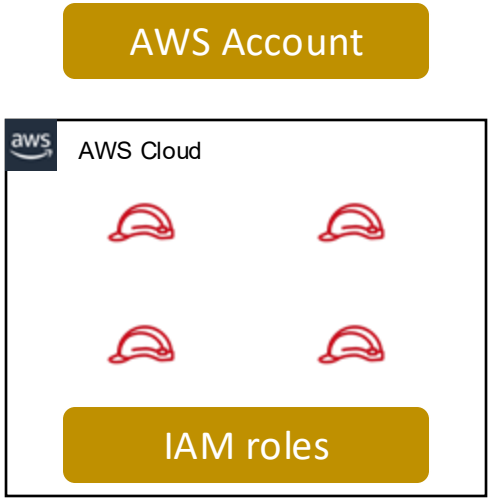
Existing Organization member accounts can be imported into the Control Tower setup

SAML Federation for Single AWS Accounts



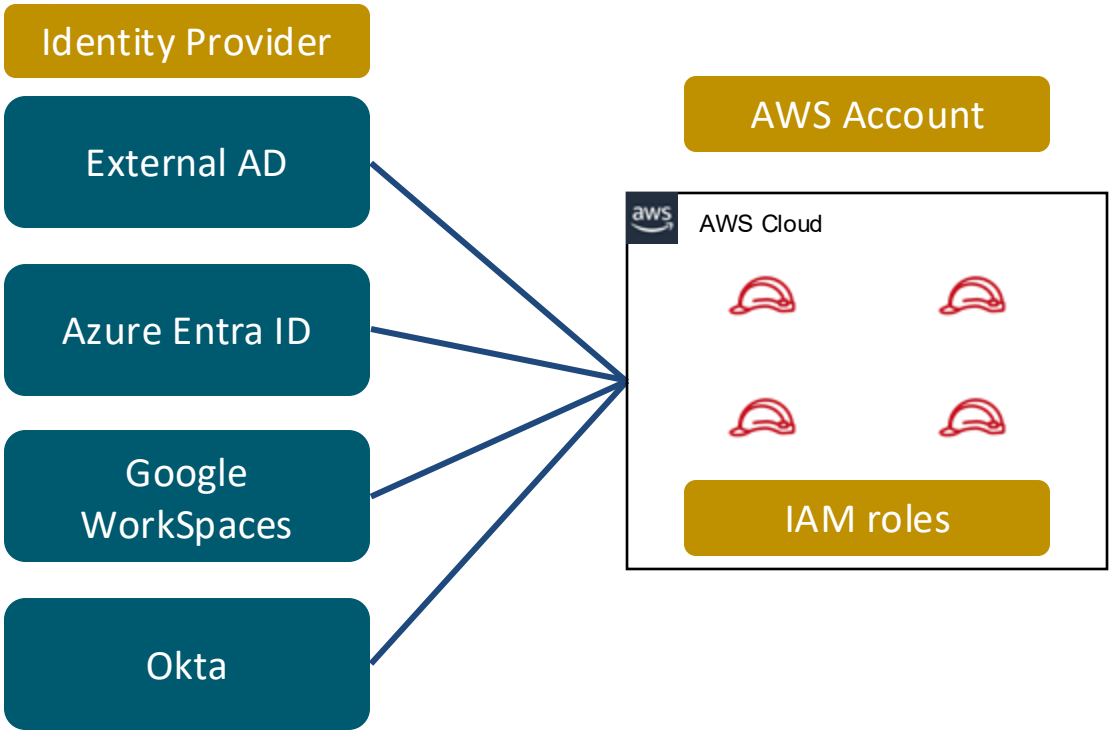
Federation can be performed at the AWS account scope

SAML Federation for Single AWS Accounts



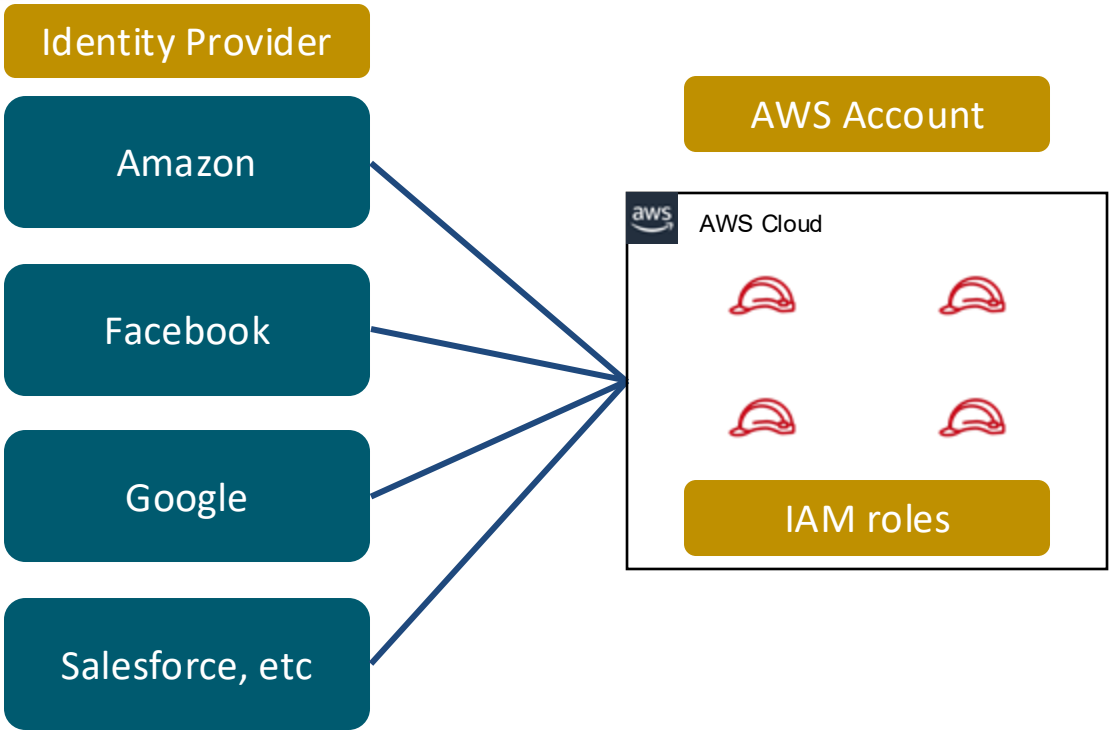
Federation uses IAM roles with permission policies

SAML Federation for Single AWS Accounts



SAML is supported for many different Identity Providers

OpenID Connect Federation for Single AWS Accounts



Federation is also supported for many different OpenID Connect providers

IAM Identity Center Federation for Multiple Accounts

Use instead of federated access to each AWS account

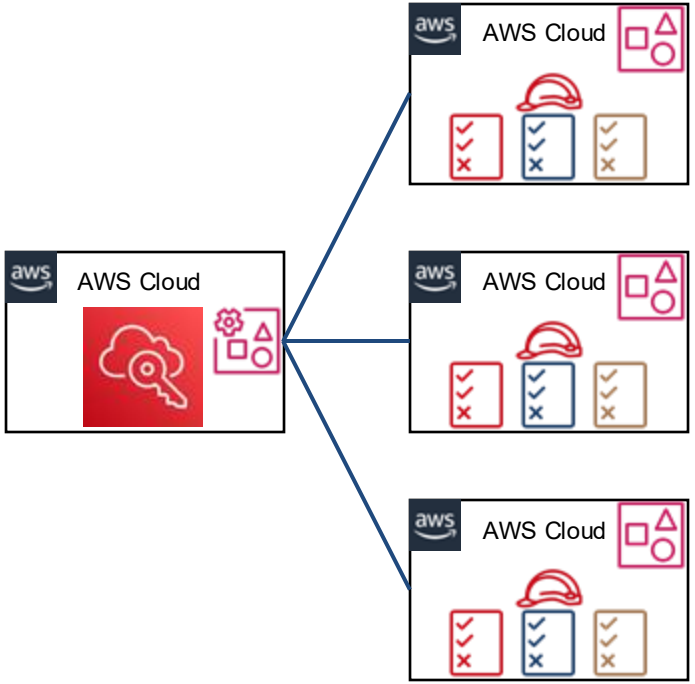
IAM Identity Center Federation for Multiple Accounts

Uses temporary credentials by assuming IAM roles in the destination accounts

IAM Identity Center Federation for Multiple Accounts

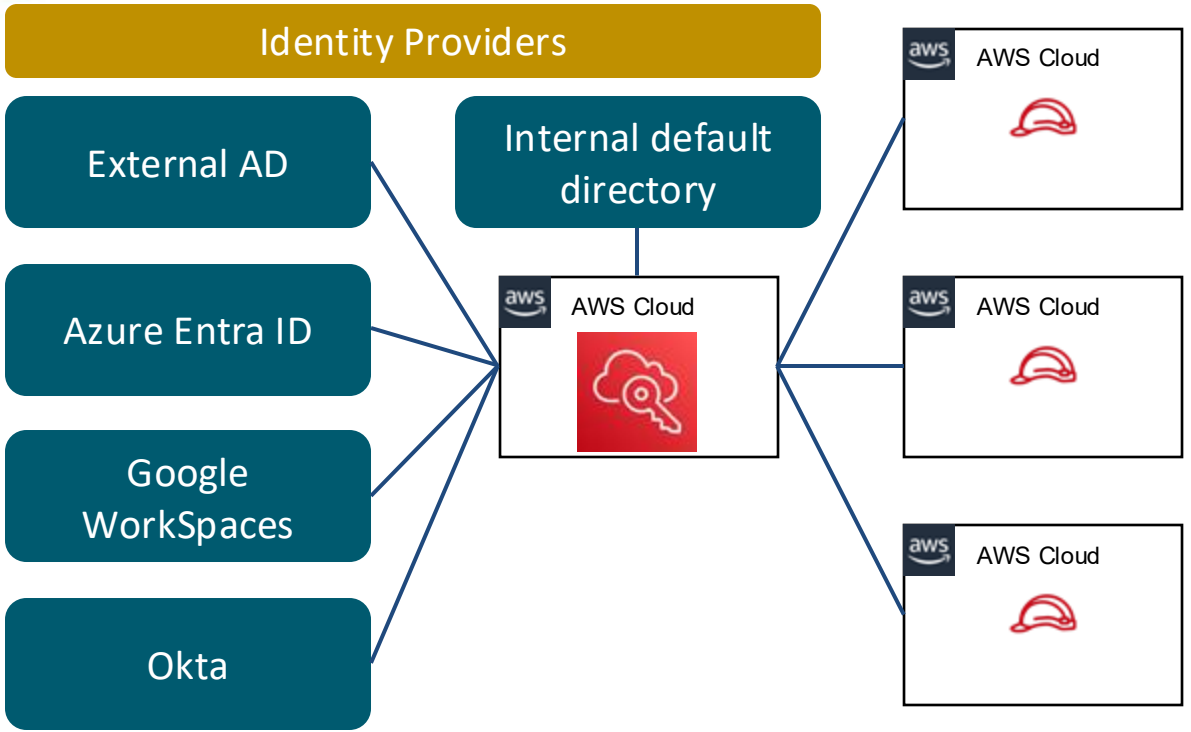
Use AWS-managed, Customer-managed, or inline policies

IAM Identity Center Federation for Multiple Accounts



All accounts must be part of the same AWS Organization

IAM Identity Center Federation for Multiple Accounts



Federation is supported with specific vendors or via SAML

Question Breakdown

Implement a strong identity foundation

Question and Answer Choices

Your company has developed a product that involves secure data transfer to individual partners. Each partner has unique, sensitive data, which is stored in dedicated, isolated S3 buckets to ensure data segregation and security.

Considering this setup, which permissions implementation would best minimize the risk of any partner accessing inappropriate data that does not belong to them?

A. S3 bucket policy applied to each bucket with cross-account permissions
B. S3 bucket ACL + object ACLS with cross-account permissions
C. IAM user in the company account, credentials delivered to partner
D. IAM role in the company account, cross-account trust with partner

Answer Choice Explanation

The S3 bucket policy is unique to each bucket, and can be customized to allow least-privilege access to a single partner without the possibility of the partner accessing any other S3 bucket.

A. **S3 bucket policy applied to each bucket with cross-account permissions**
B. **S3 bucket ACL + object ACLS with cross-account permissions**
C. **IAM user in the company account, credentials delivered to partner**
D. **IAM role in the company account, cross-account trust with partner**

Answer Choice Explanation

Like S3 bucket policies, bucket ACLs and object ACLs are specific to the scope they're configured to, but do not allow for least-privilege design, as the grantee can only be an entire 12-digit AWS account ID.

A. S3 bucket policy applied to each bucket with cross-account permissions
B. S3 bucket ACL + object ACLS with cross-account permissions
C. IAM user in the company account, credentials delivered to partner
D. IAM role in the company account, cross-account trust with partner

Answer Choice Explanation

This solution, while it can be least-privilege, allows for the possibility of "extra" permissions being associated with the IAM user.

A. S3 bucket policy applied to each bucket with cross-account permissions
B. S3 bucket ACL + object ACLS with cross-account permissions
C. IAM user in the company account, credentials delivered to partner
D. IAM role in the company account, cross-account trust with partner

Answer Choice Explanation

This solution is similar to C in that it is possible for least privilege, but also possible for extra or unintended permissions.

A. S3 bucket policy applied to each bucket with cross-account permissions
B. S3 bucket ACL + object ACLS with cross-account permissions
C. IAM user in the company account, credentials delivered to partner
D. IAM role in the company account, cross-account trust with partner

Correct Answer

Correct Answer: A

A. S3 bucket policy applied to each bucket with cross-account permissions
B. S3 bucket ACL + object ACLS with cross-account permissions
C. IAM user in the company account, credentials delivered to partner
D. IAM role in the company account, cross-account trust with partner

Question Domain 1: Design Secure Architectures

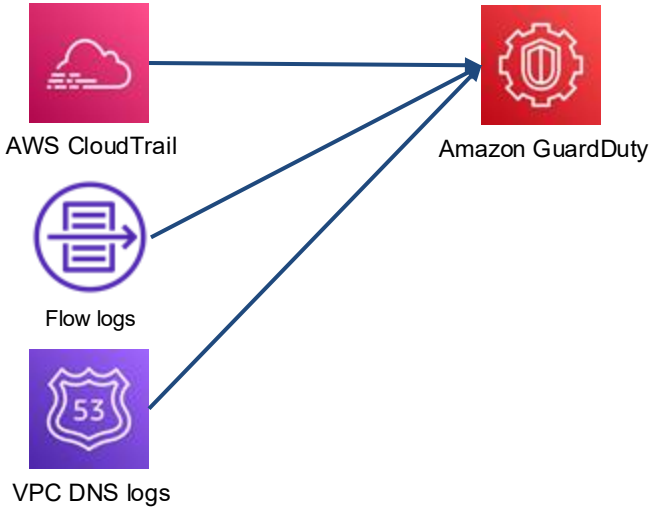
Enable traceability

Principle Definition

Monitor, alert, and audit actions and changes to your environment in real time.

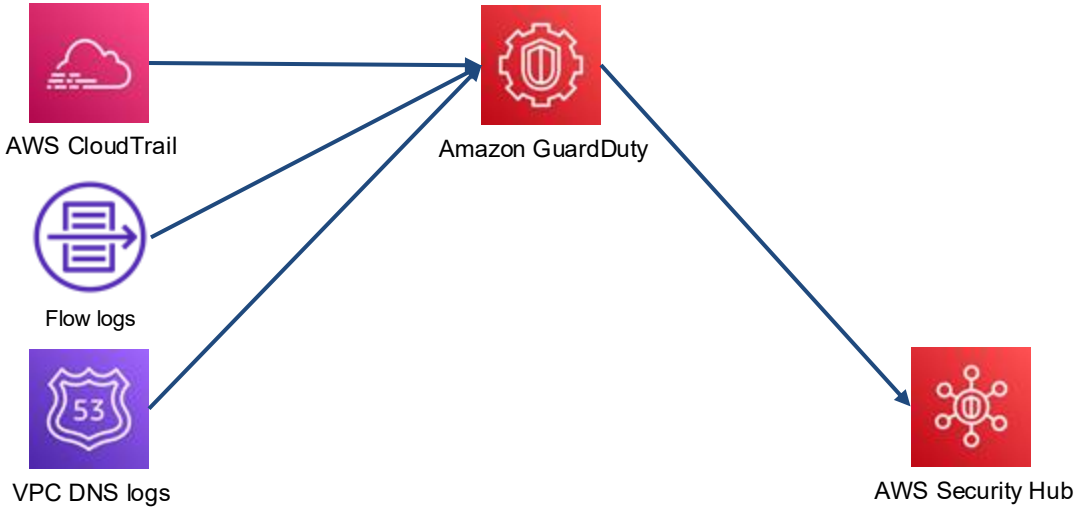
Integrate log and metric collection with systems to automatically investigate and take action.

Monitor Security Events



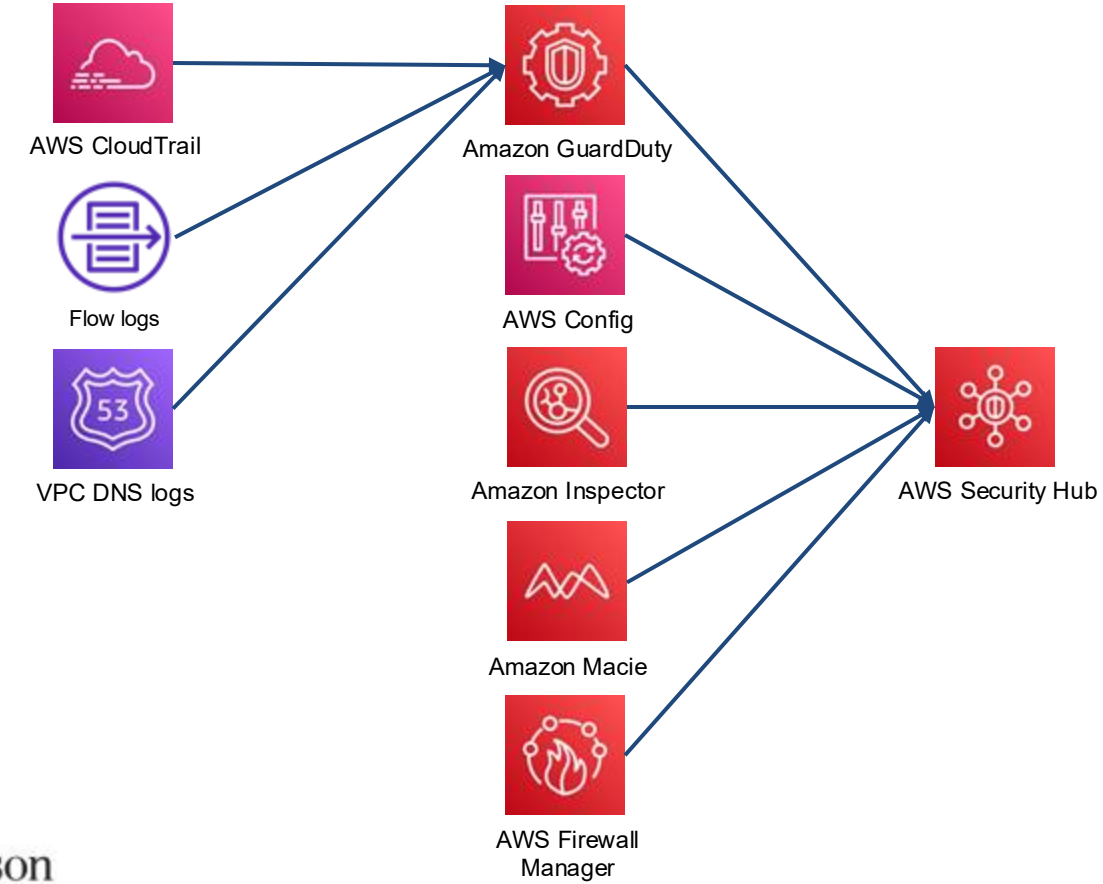
Ingest logs and generate findings of abnormal behavior based on ML

Monitor Security Events



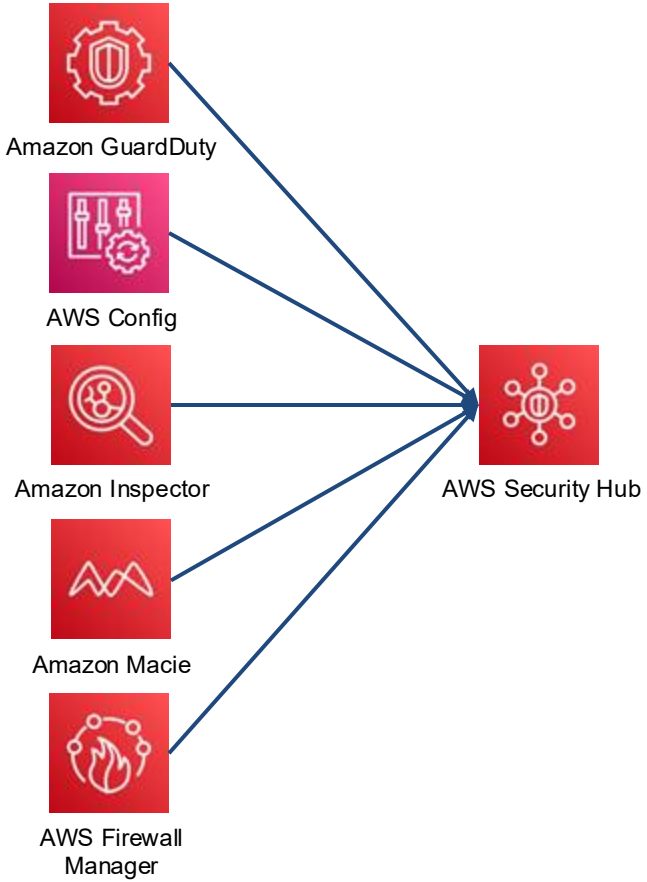
Deliver GuardDuty findings to Security Hub

Monitor Security Events



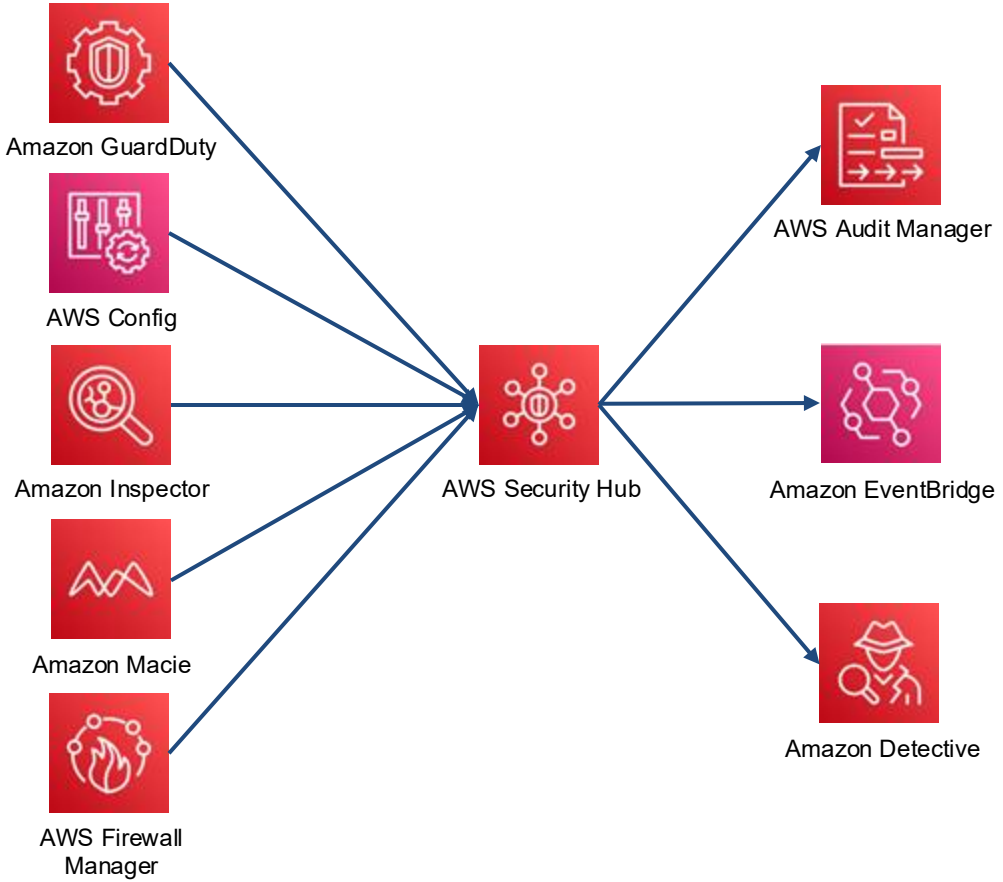
Ingest other findings to Security hub for a consolidated dashboard

Audit and Alert on Security Events



Findings are already consolidated into Security Hub

Audit and Alert on Security Events



Deliver findings for audits and mitigation

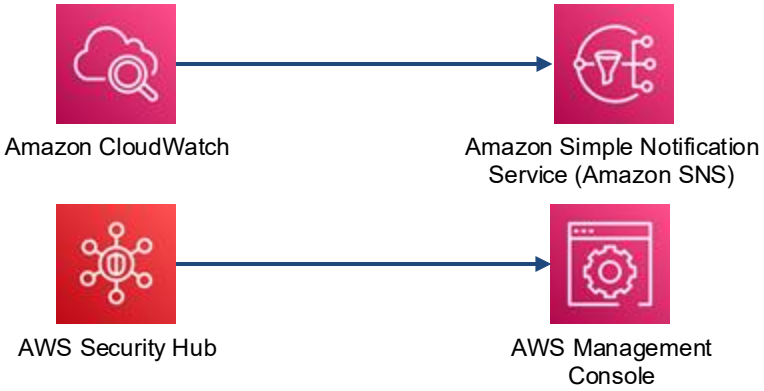
Remediate Security Events

Amazon CloudWatch

AWS Security Hub

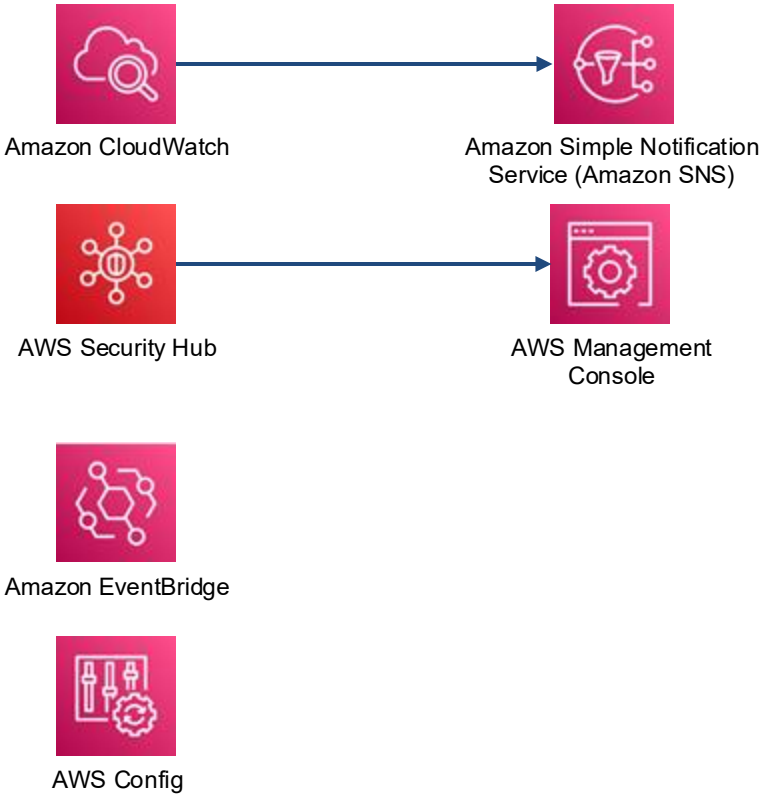
Services offer manual notifications and dashboards

Remediate Security Events



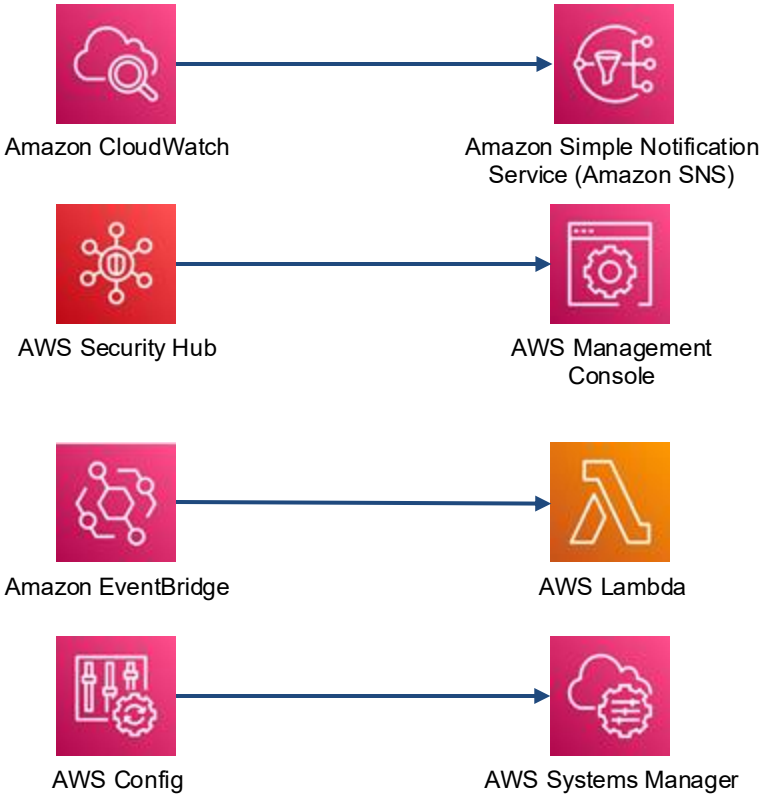
View via browser or email

Remediate Security Events



Services offer automation as events or state changes

Remediate Security Events



Automatically mitigate via multiple methods

Question Breakdown

Enable traceability

Question and Answer Choices

A company is using AWS to host its multi-tier web application. The application architecture includes Amazon EC2, Amazon RDS, and Amazon S3. The company has a strict compliance policy that requires all resources to adhere to specific standards. They are looking for a solution that can continuously monitor and evaluate the configurations of their AWS resources against the desired configurations and provide alerts when non-compliance is detected.

Which AWS service should the security team use to meet this requirement?

A. Amazon Inspector
B. AWS Config
C. Amazon CloudWatch
D. AWS Trusted Advisor

Answer Choice Explanation

Amazon Inspector is primarily used for security assessments and vulnerability management of EC2 instances and applications running on those instances. It does not provide continuous monitoring and evaluation of resource configurations against compliance standards.

A. Amazon Inspector
B. AWS Config
C. Amazon CloudWatch
D. AWS Trusted Advisor

Answer Choice Explanation

Config is designed to assess, audit, and evaluate the configurations of your AWS resources. AWS Config continuously monitors and records your AWS resource configurations and allows you to automate the evaluation of recorded configurations against desired configurations. This service can alert you whenever non-compliant resources are detected, making it the ideal solution for the company's requirement.

A. Amazon Inspector
B. AWS Config
C. Amazon CloudWatch
D. AWS Trusted Advisor

Answer Choice Explanation

While CloudWatch provides monitoring and operational data for AWS resources, it does not evaluate resource configurations against compliance standards. CloudWatch is more focused on performance monitoring and operational health.

A. Amazon Inspector
B. AWS Config
C. Amazon CloudWatch
D. AWS Trusted Advisor

Answer Choice Explanation

Trusted Advisor provides recommendations to help you follow AWS best practices. While Trusted Advisor does offer some checks related to security and compliance, it does not provide the continuous monitoring and evaluation of specific resource configurations against compliance standards like AWS Config does.

A. Amazon Inspector
B. AWS Config
C. Amazon CloudWatch
D. AWS Trusted Advisor

Correct Answer

Correct Answer: B

A. Amazon Inspector
B. AWS Config
C. Amazon CloudWatch
D. AWS Trusted Advisor

Question Domain 1: Design Secure Architectures

Apply security at all layers

Principle Definition

Apply a defense in depth approach with multiple security controls.

Apply to all layers (for example, edge of network, VPC, load balancing, every instance and compute service, operating system, application, and code).

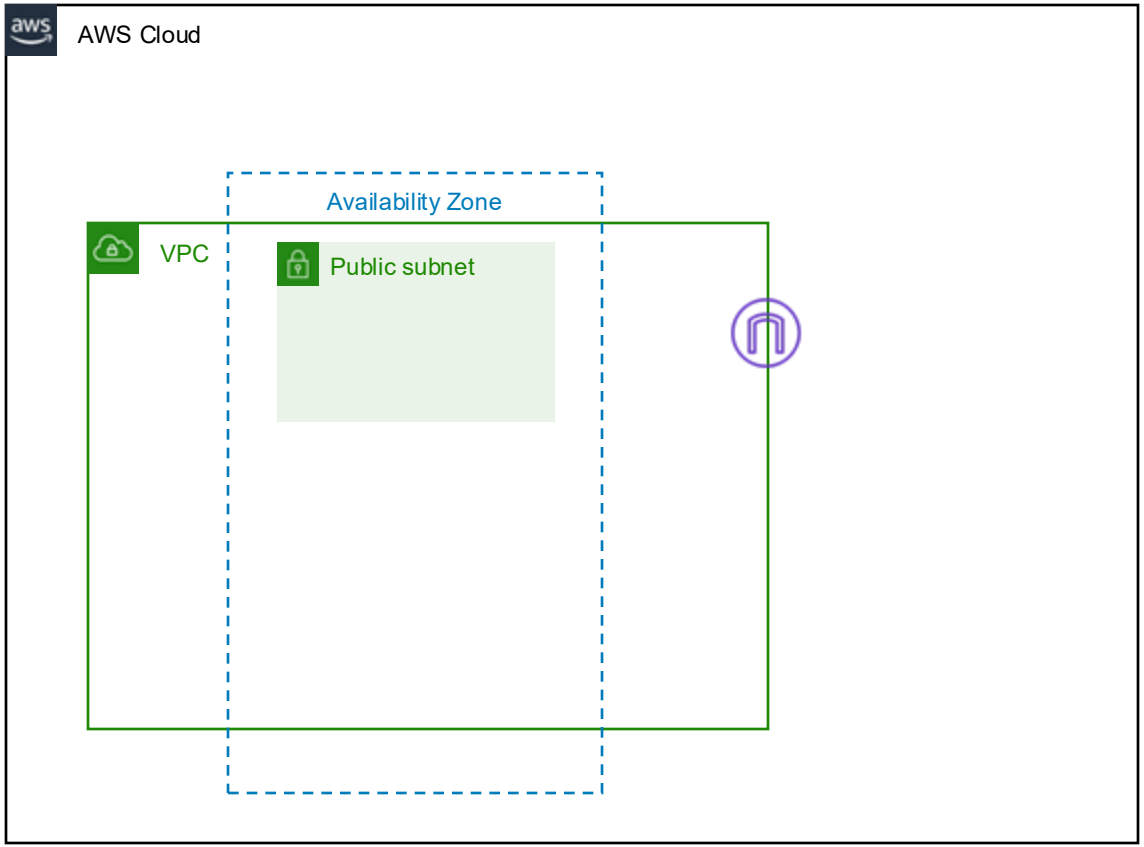
Defense In Depth Definition

- Apply security features everywhere that it is appropriate to do so
- Don't focus on protecting a single layer

Defense In Depth Scenario

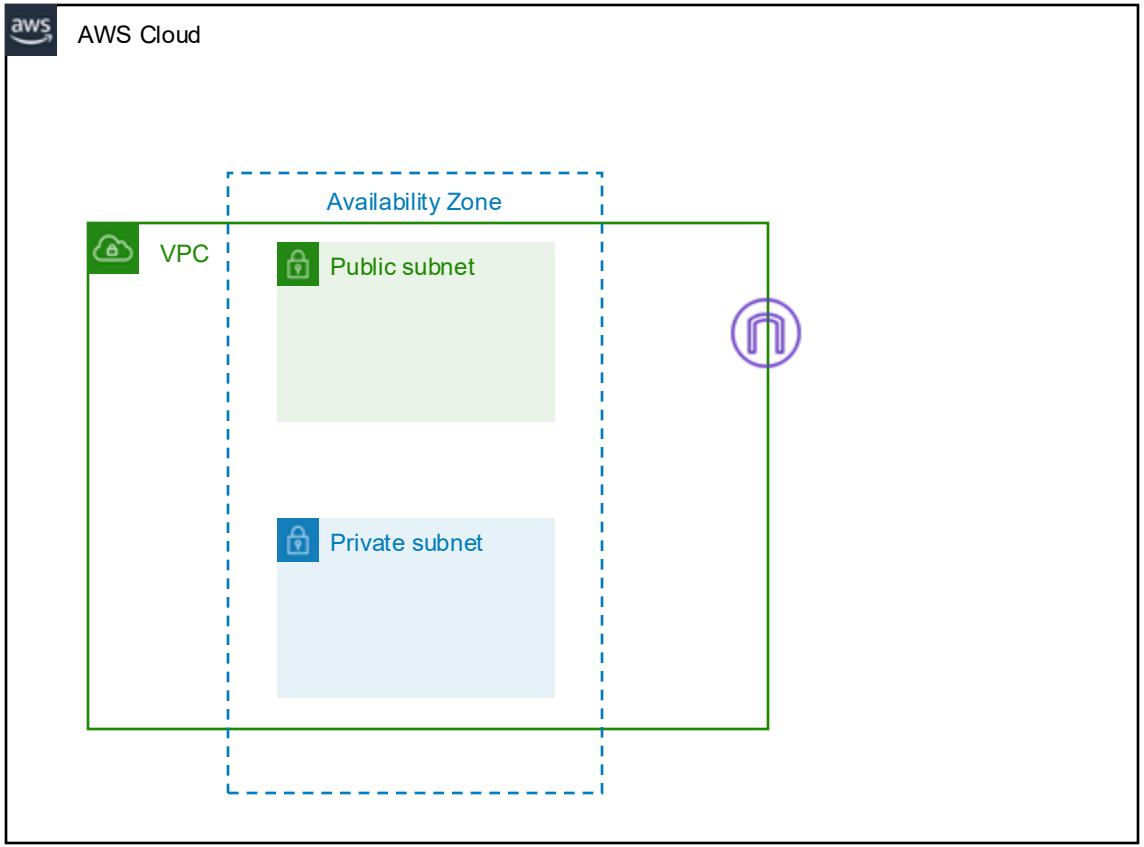
- Web-based EC2 application
- Application requires access to DynamoDB
- How do we apply defense in depth?

Defense In Depth Scenario



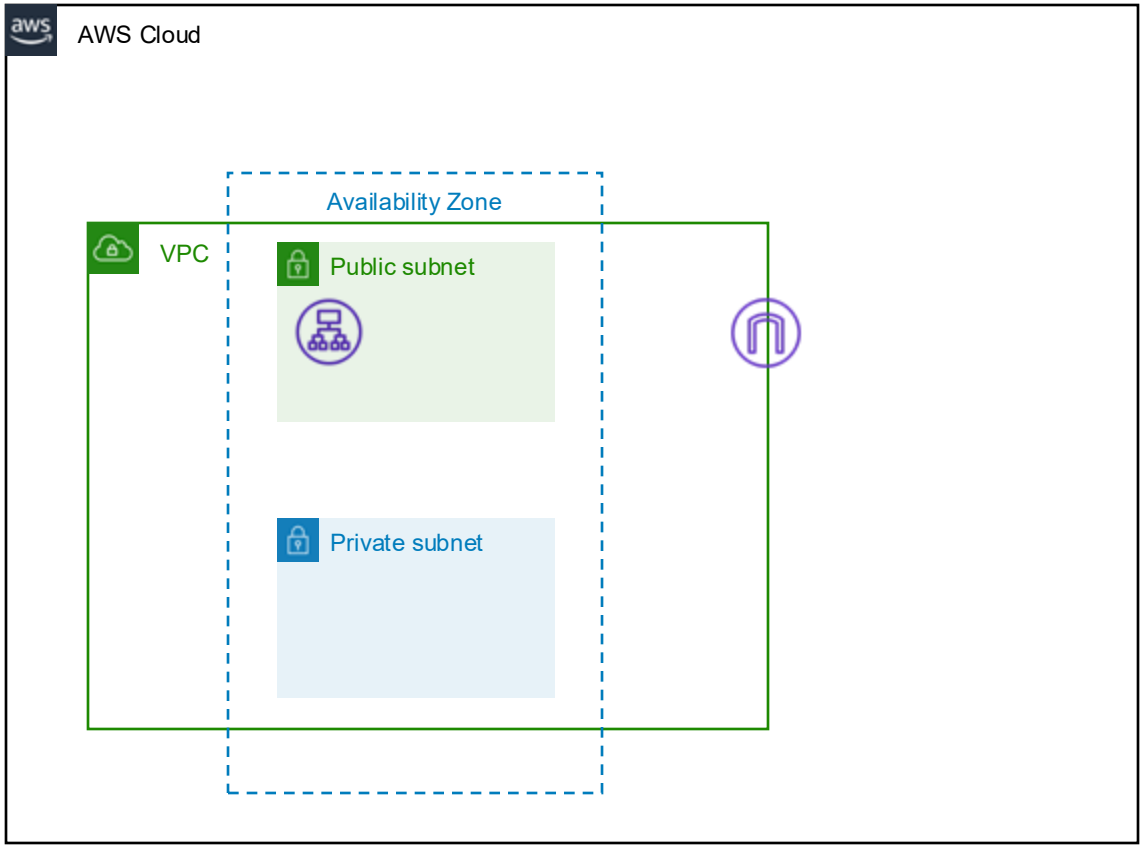
Deploy public subnet for external facing resources

Defense In Depth Scenario



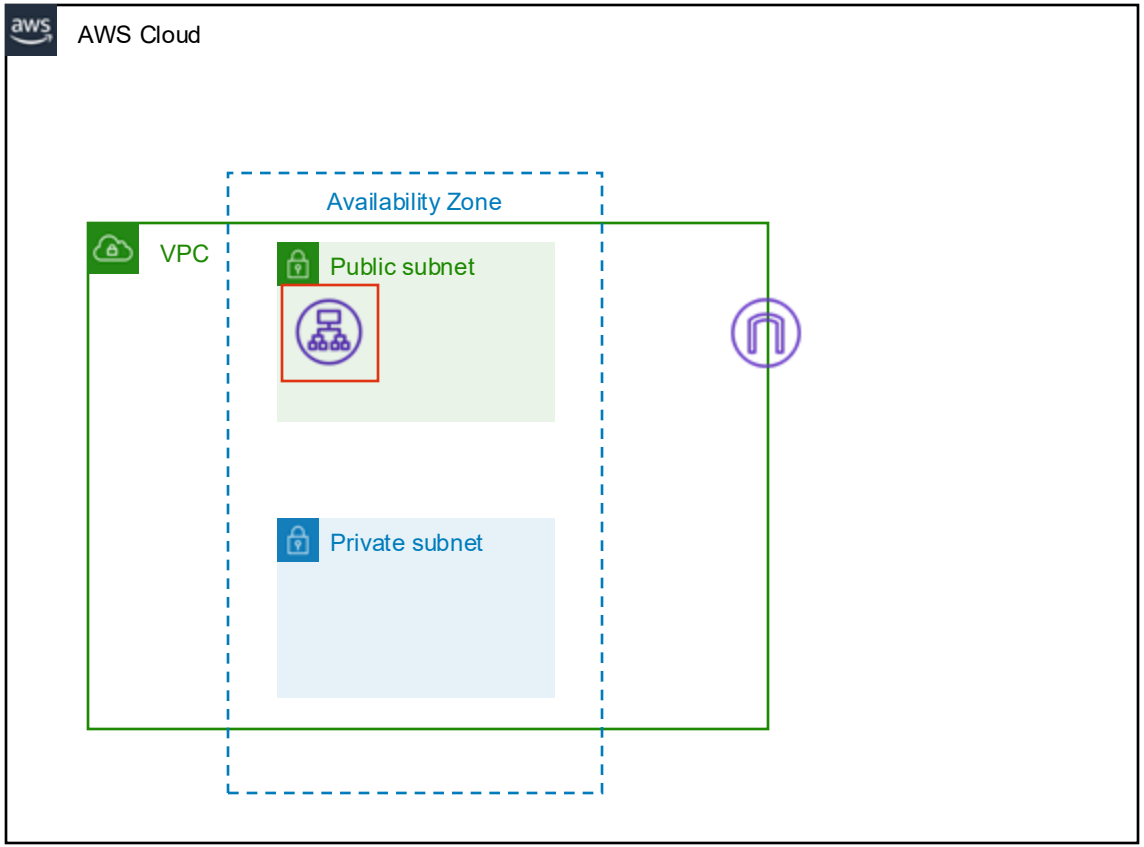
Deploy private subnet for internal resources

Defense In Depth Scenario



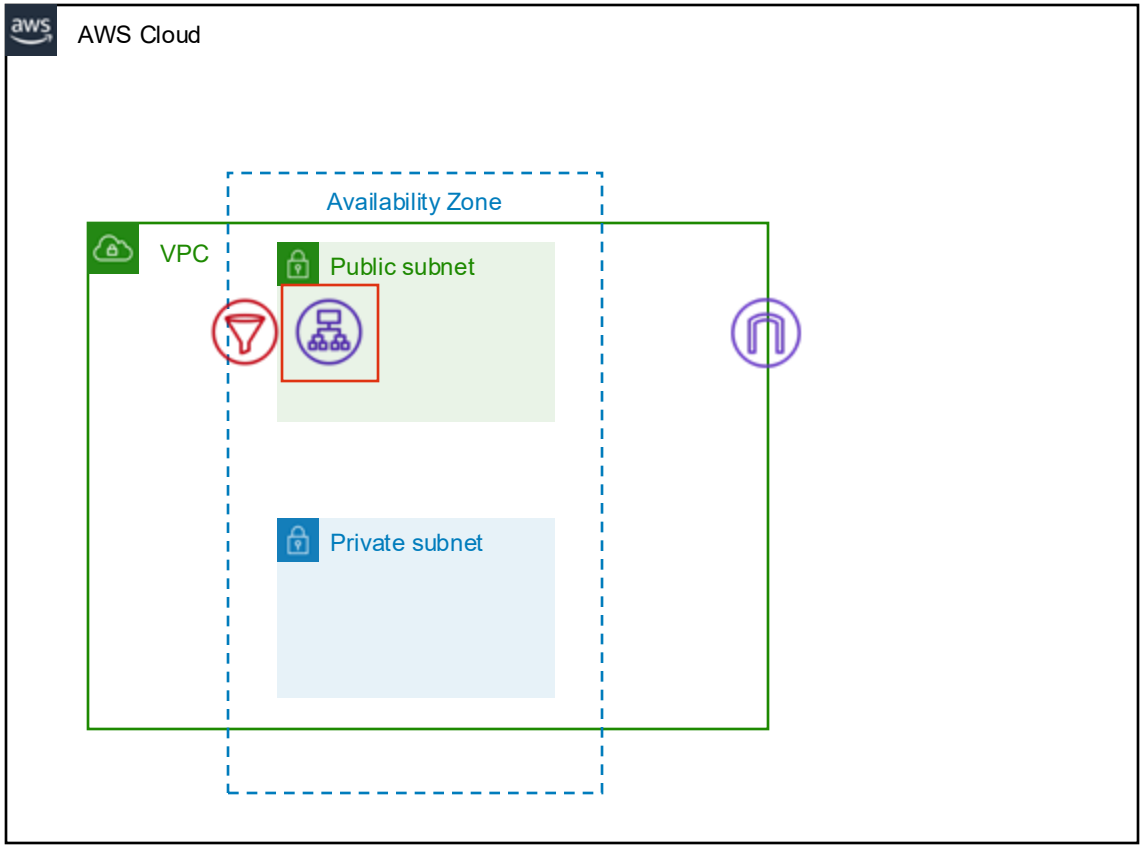
Deploy ALB to accept inbound traffic

Defense In Depth Scenario



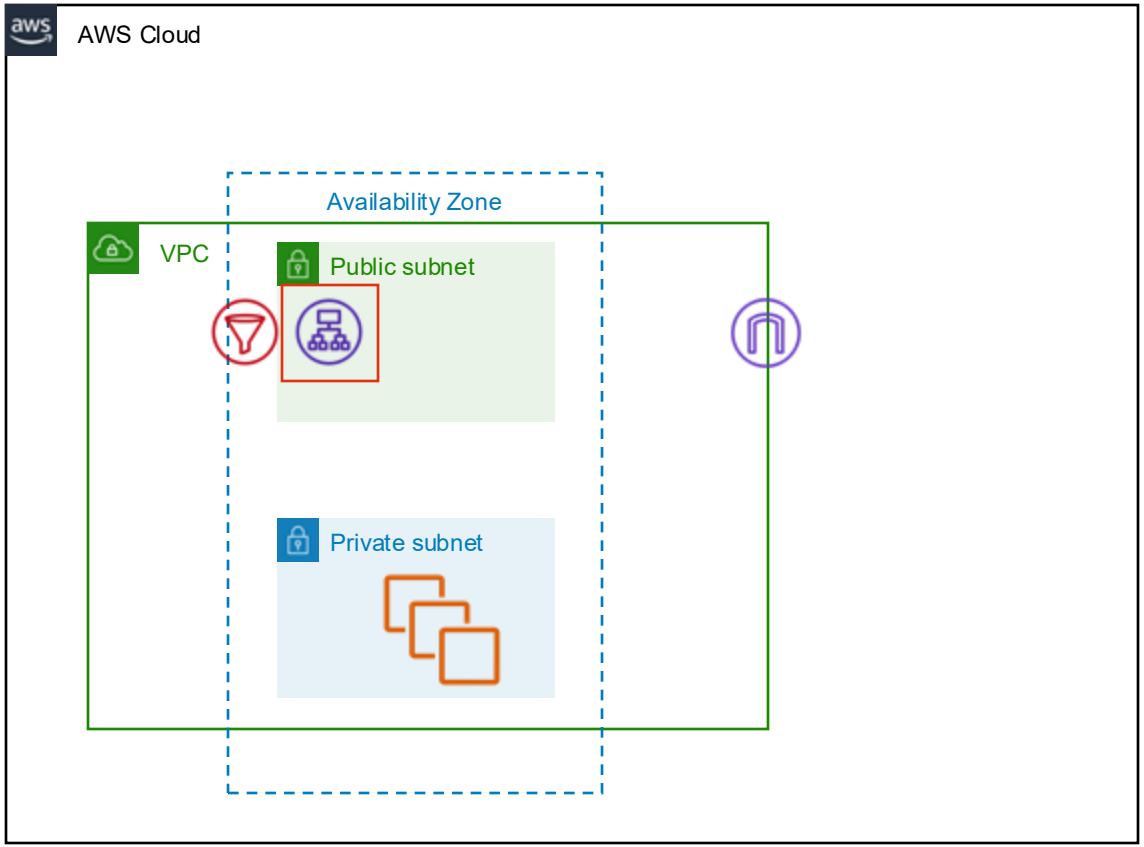
Configure security group for least privilege inbound traffic

Defense In Depth Scenario



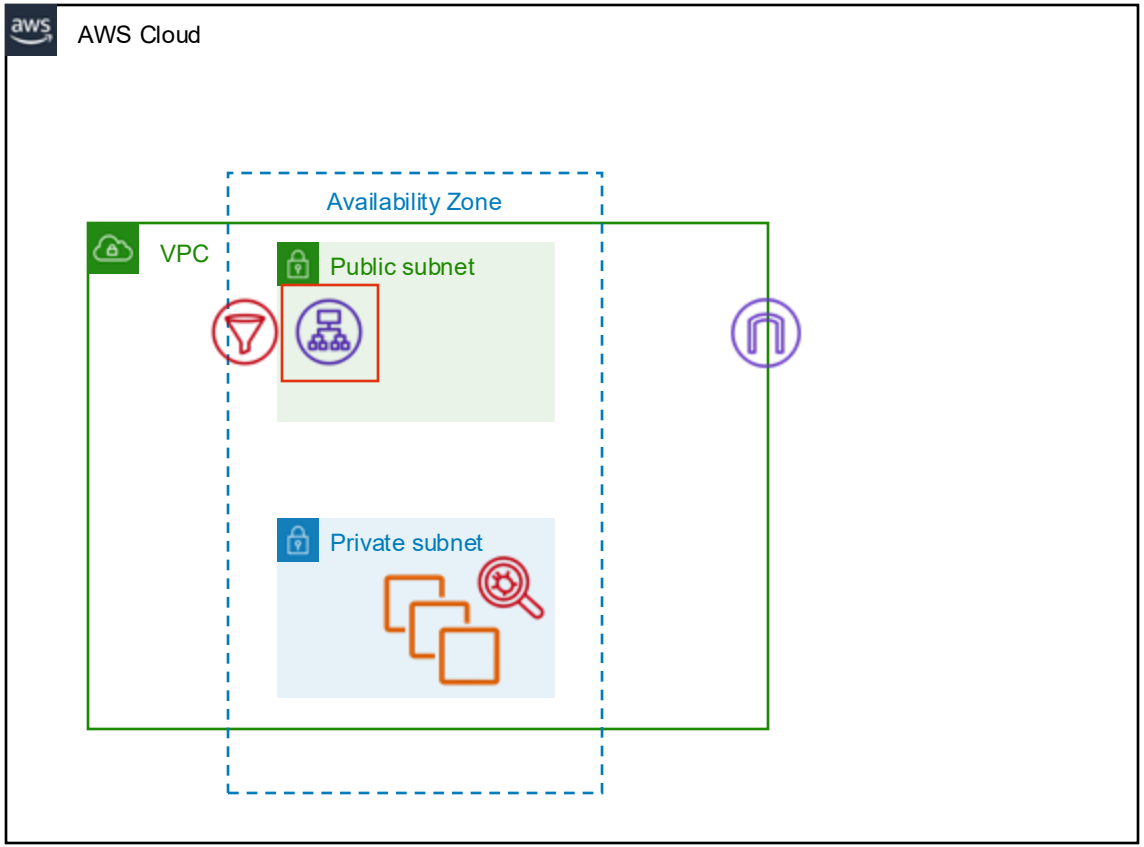
Deploy WAF Web ACL to reject improper requests

Defense In Depth Scenario



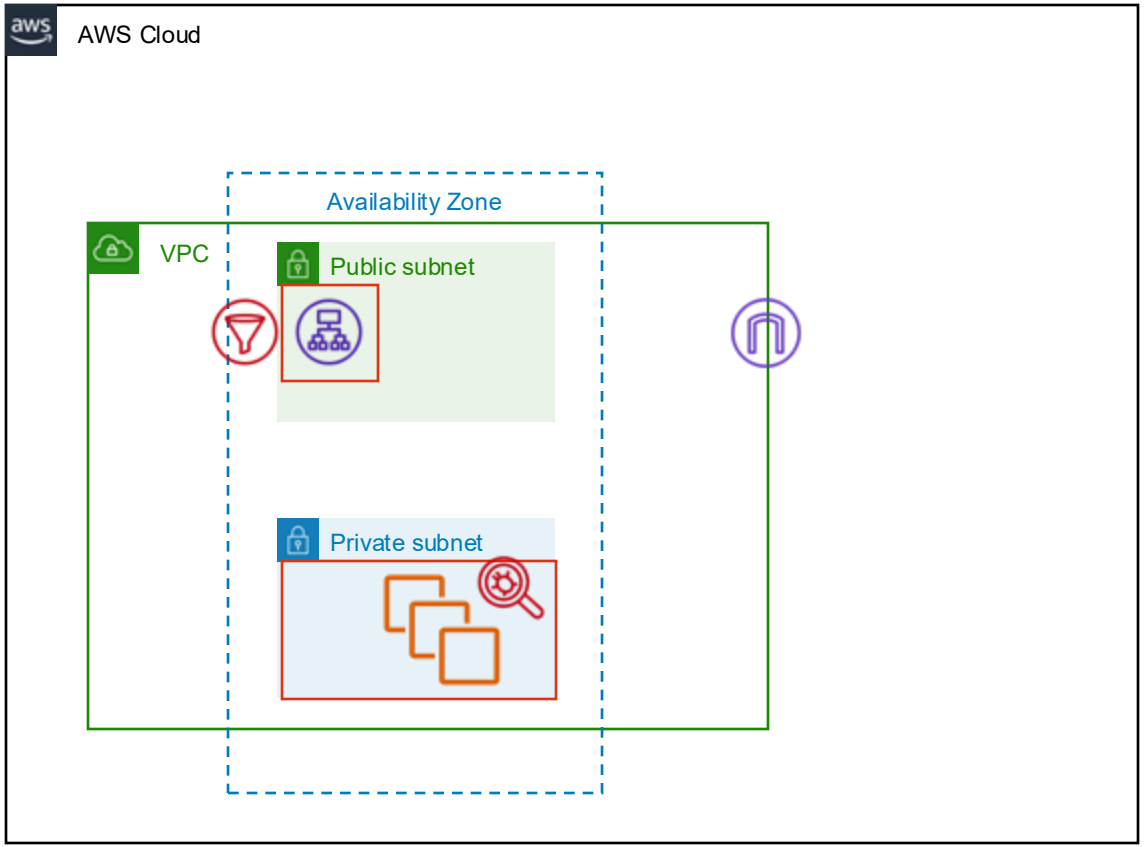
Deploy application instances into the private subnet

Defense In Depth Scenario



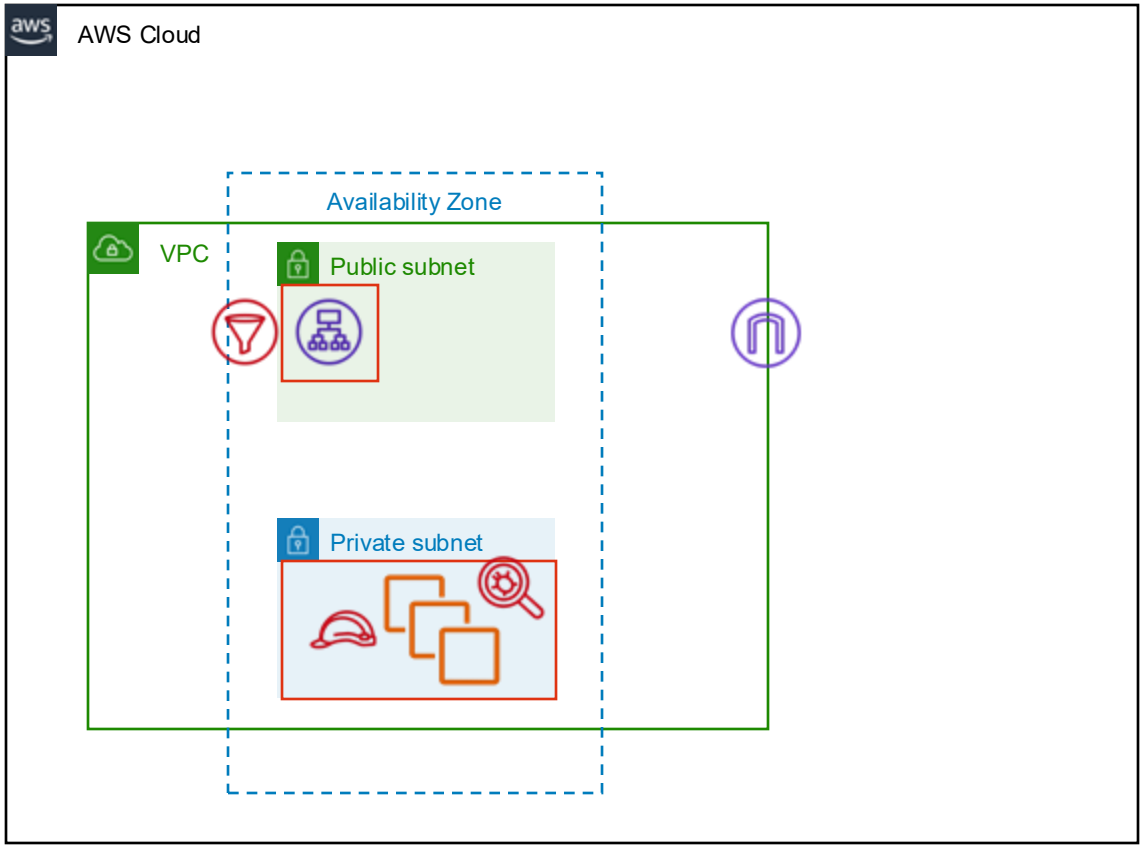
Enable Inspector to scan AMIs for vulnerabilities

Defense In Depth Scenario



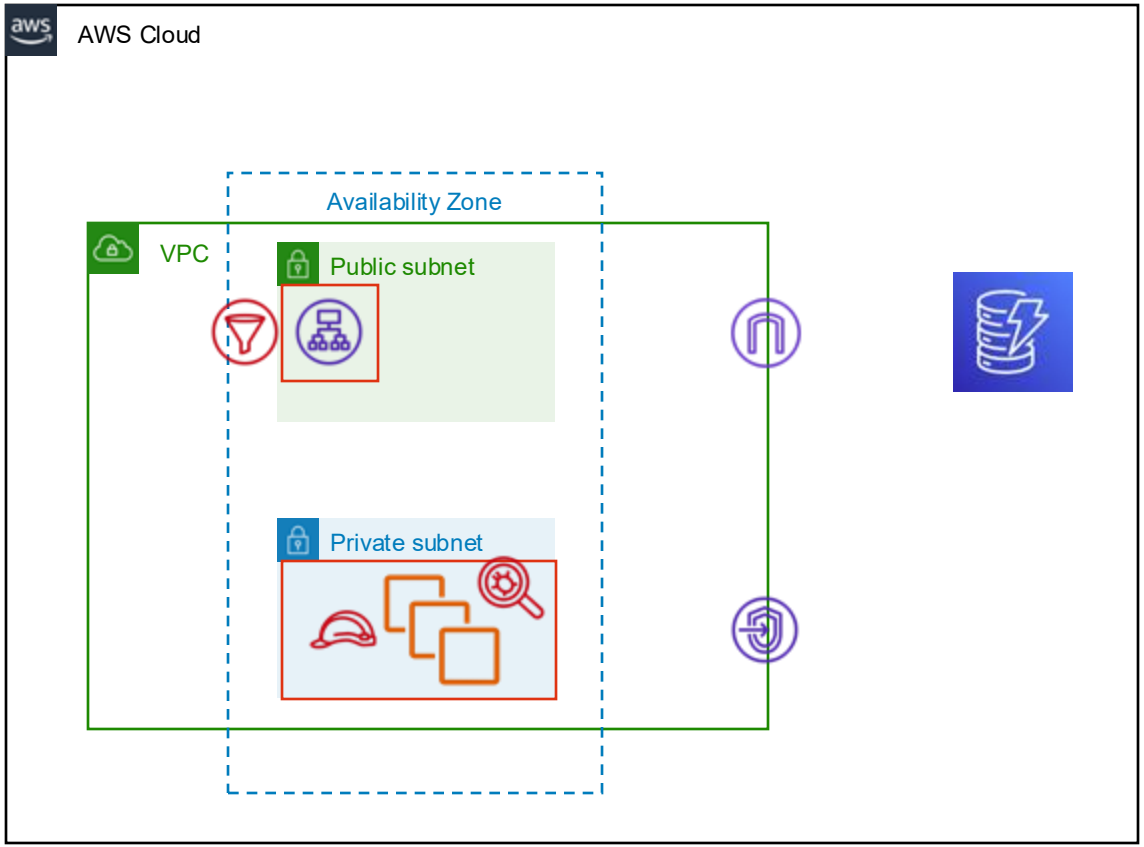
Configure security group for least privilege ingress from ALB

Defense In Depth Scenario



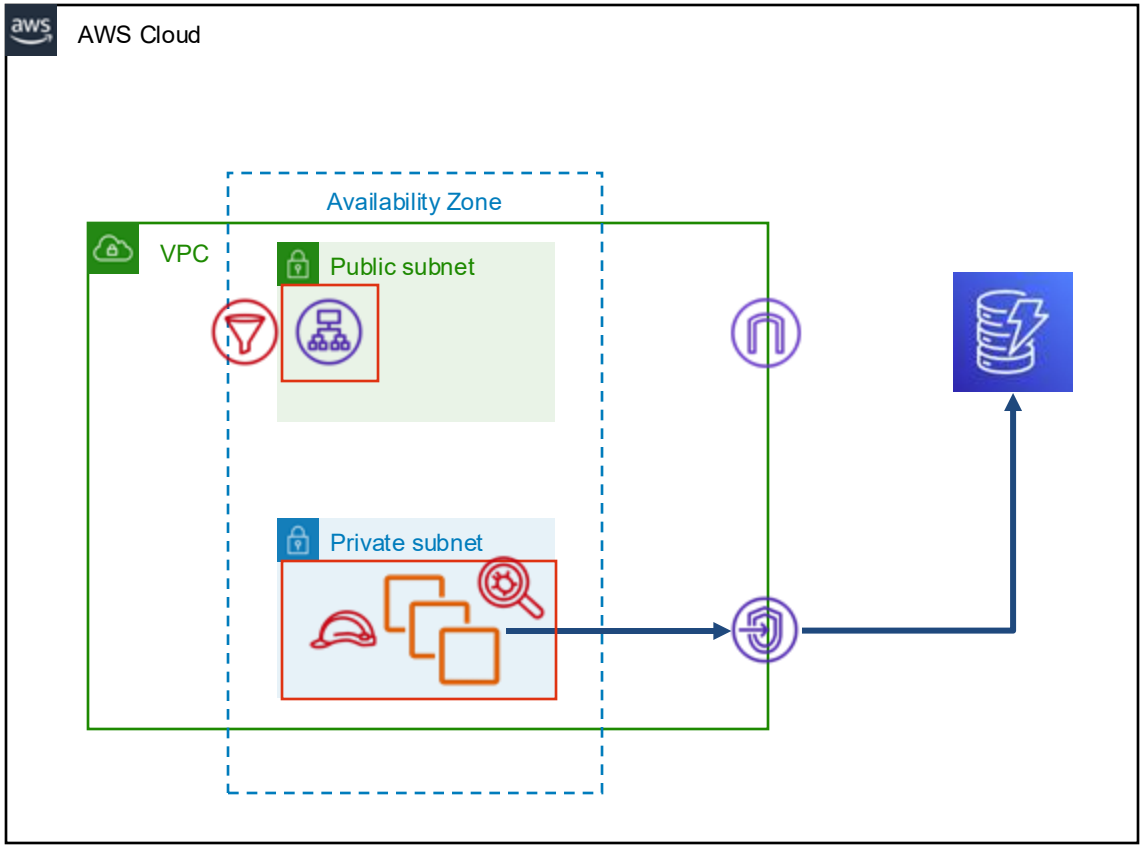
Deploy an EC2 instance profile with DynamoDB permissions

Defense In Depth Scenario



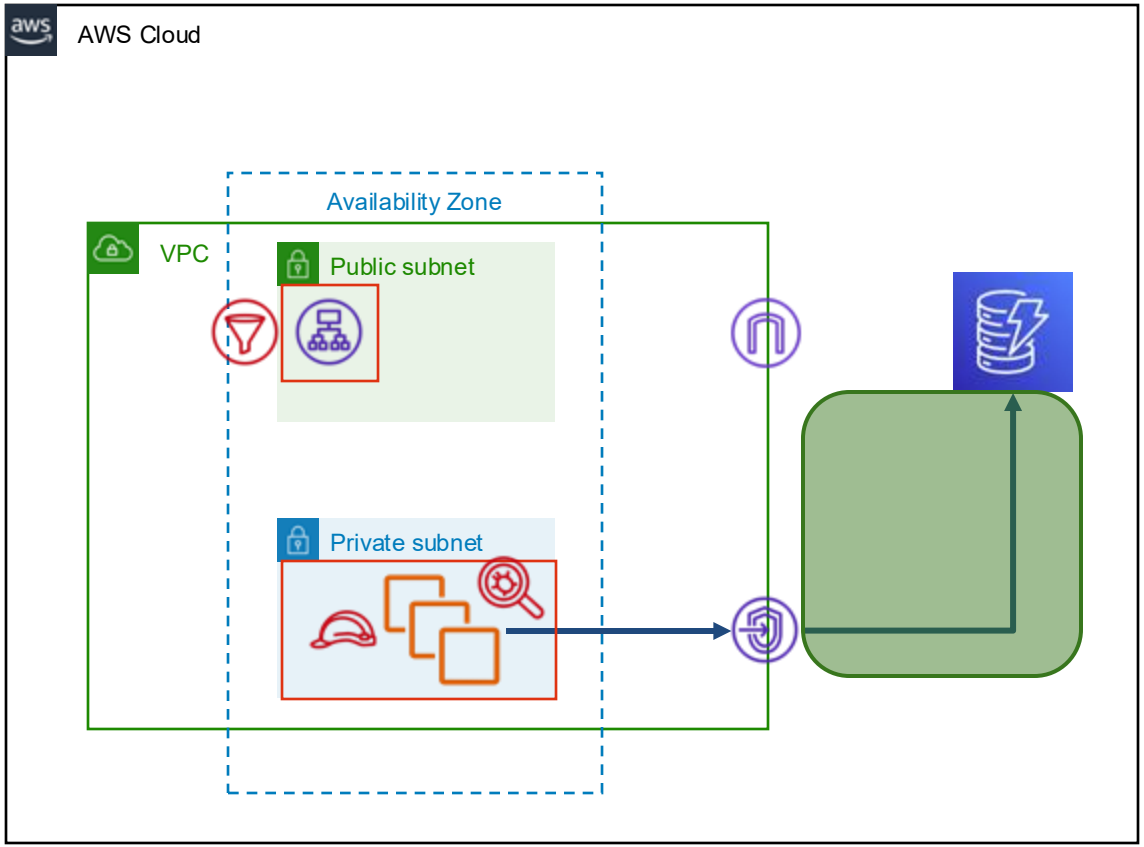
Deploy a VPC Gateway endpoint with route from private subnet

Defense In Depth Scenario



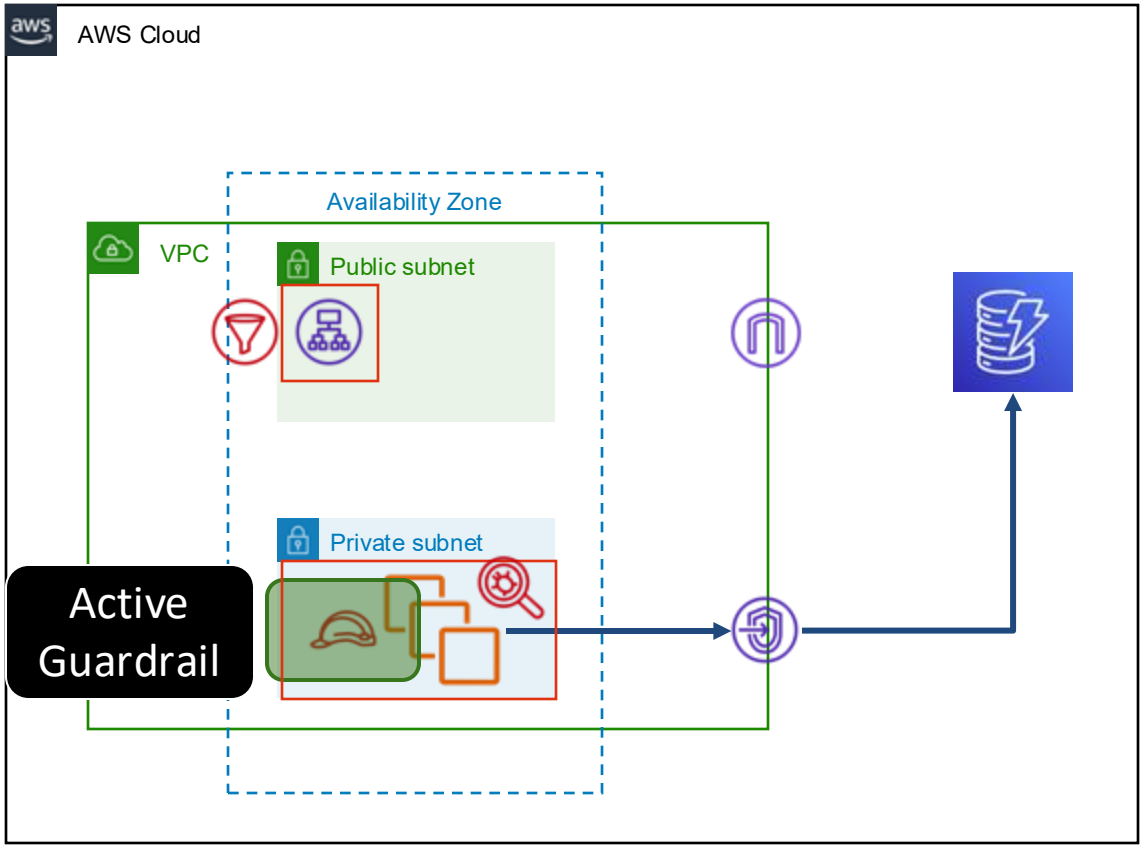
Route DynamoDB traffic through the VPC endpoint

Defense In Depth Scenario



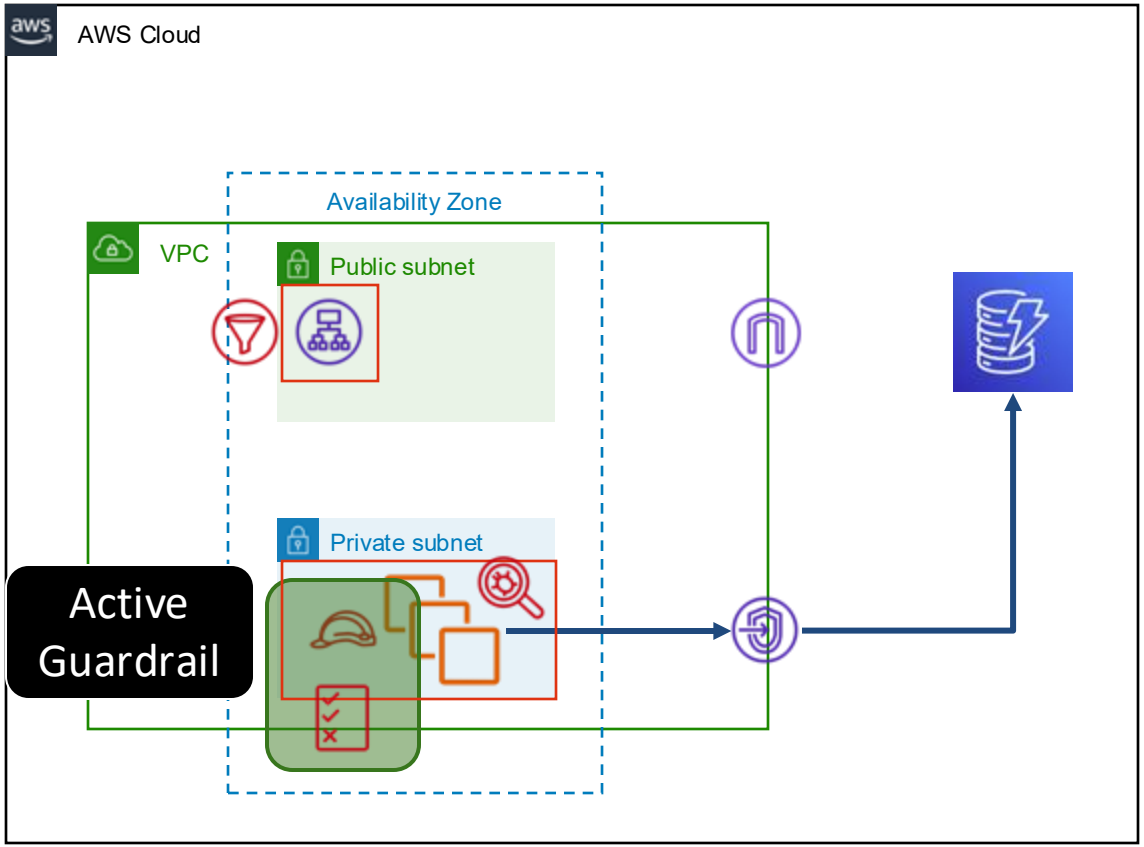
Traffic is proxied privately to DynamoDB

Defense In Depth Scenario



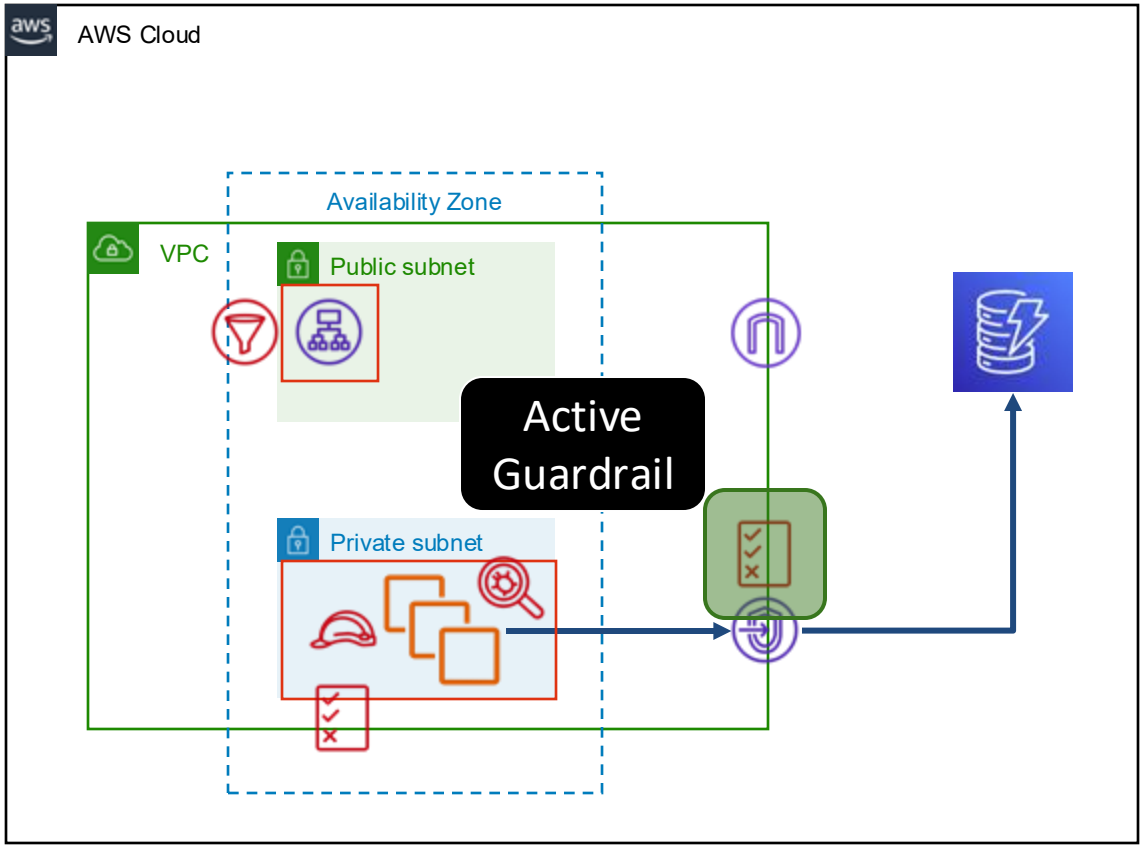
Add condition to the IAM role policy only allowing requests from VPC ID

Defense In Depth Scenario



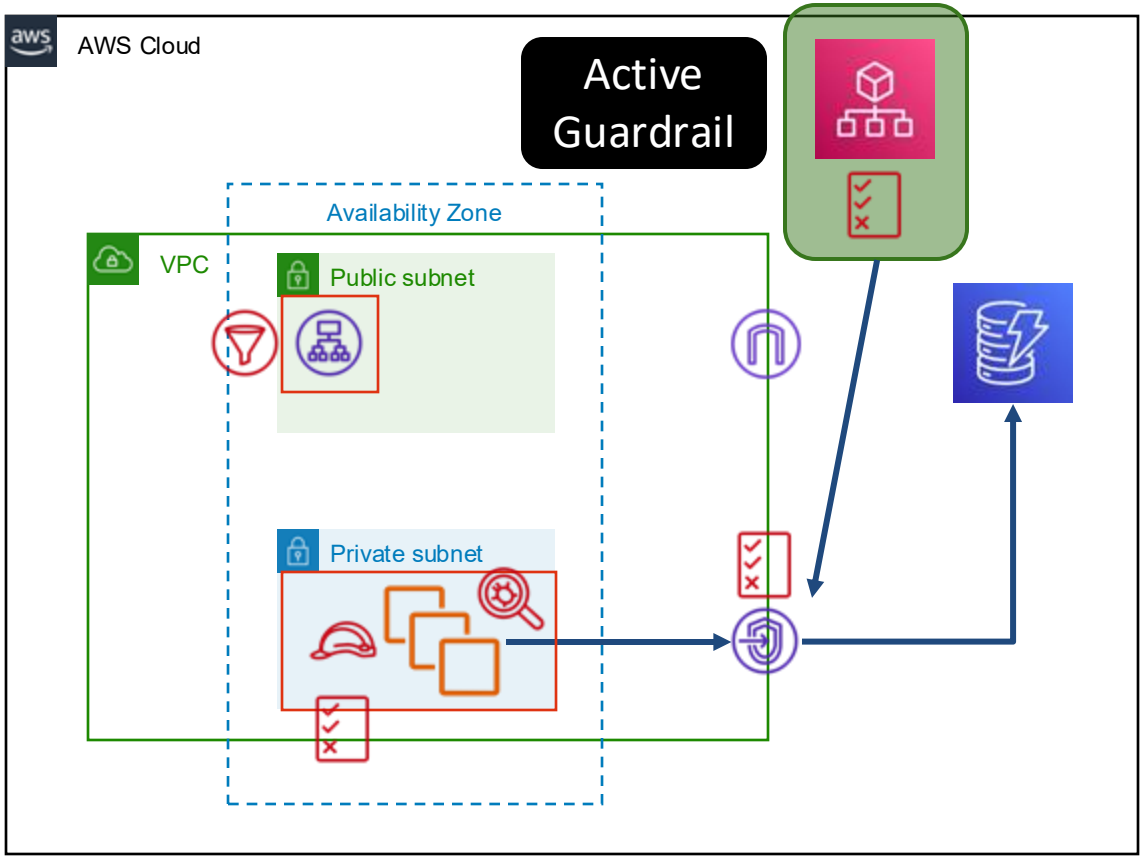
Configure a permissions boundary on the policy to ensure least privilege

Defense In Depth Scenario



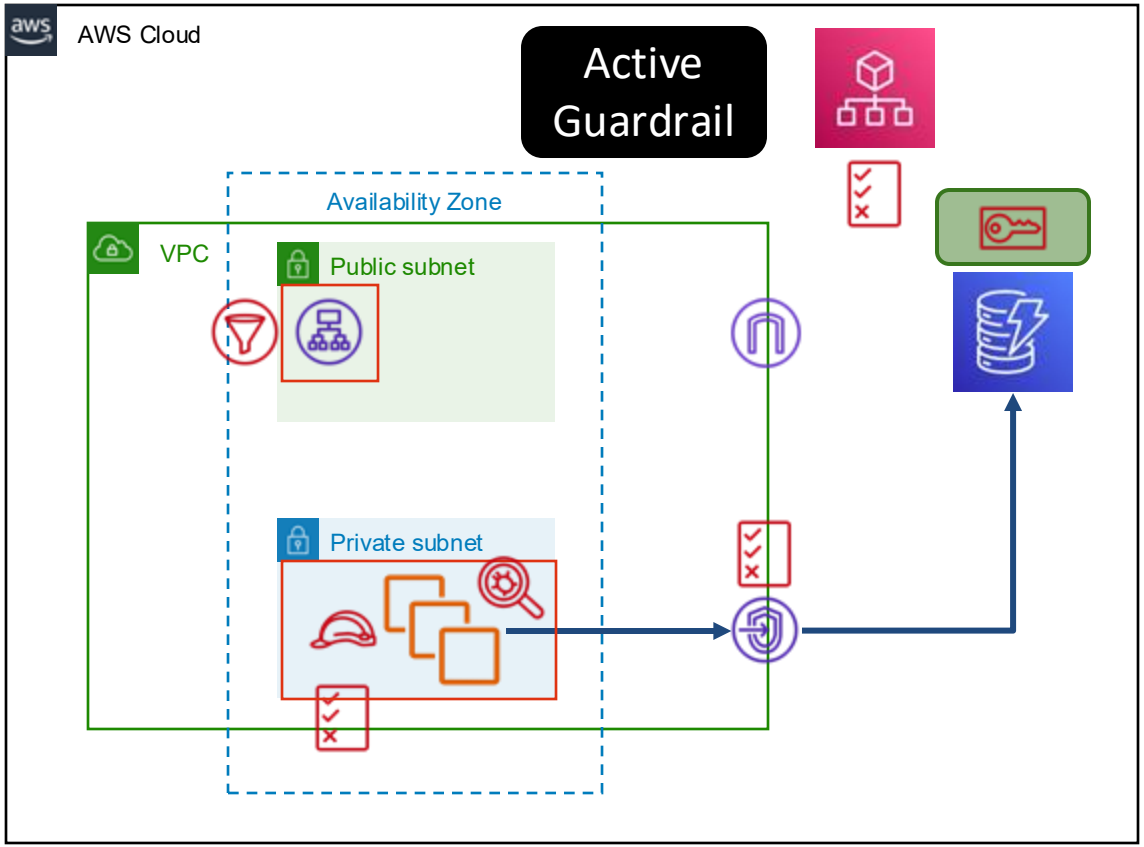
Add an endpoint policy with condition for traffic only from application subnet

Defense In Depth Scenario



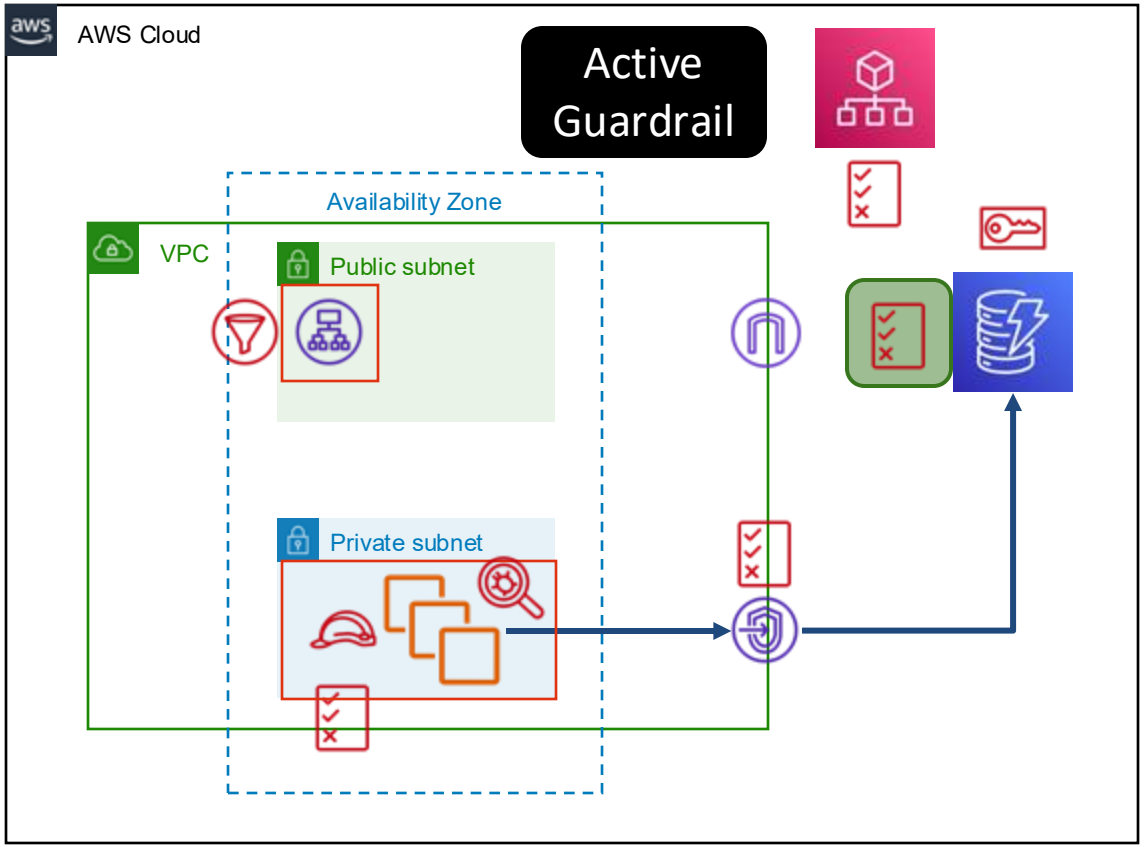
Add Organizations SCP to deny DynamoDB table access except through Gateway endpoint

Defense In Depth Scenario



Use KMS encryption on the table and allow the IAM role as the only key user

Defense In Depth Scenario



Configure resource permission policy to enforce other permissions

Question Breakdown

Apply security at all layers

Question and Answer Choices

An application is deployed to EC2 instances in a private VPC subnet in the us-west-2 region. There is a functional requirement to access S3 buckets deployed into the ap-southeast-1 region. There is a security requirement for the end to end traffic to remain private.

Which combination of steps can meet the functional and security requirements? (pick three)

A. Deploy a VPC Gateway endpoint into the us-west-2 VPC
B. Deploy a VPC in the ap-southeast-1 region with an S3 Interface endpoint
C. Configure a route table entry in the us-west-2 VPC to direct S3 traffic to the Gateway endpoint
D. Configure a VPC peering connection between the us-west-2 and ap-southeast-1 VPCs
E. Configure the application to use the ap-southeast-1 region when accessing S3 buckets
F. Configure the application to use the ap-southeast-1 Interface endpoint DNS when accessing S3 buckets

Answer Choice Explanation

This may seem like a functional solution that meets the security requirements. Unfortunately, while Gateway endpoints do proxy S3 traffic to the service API endpoint (meeting the security requirement), they cannot accept traffic for buckets deployed in any region outside that of the endpoint.

A. Deploy a VPC Gateway endpoint into the us-west-2 VPC
B. Deploy a VPC in the ap-southeast-1 region with an S3 Interface endpoint
C. Configure a route table entry in the us-west-2 VPC to direct S3 traffic to the Gateway endpoint
D. Configure a VPC peering connection between the us-west-2 and ap-southeast-1 VPCs
E. Configure the application to use the ap-southeast-1 region when accessing S3 buckets
F. Configure the application to use the ap-southeast-1 Interface endpoint DNS when accessing S3 buckets

Answer Choice Explanation

This solution is more complicated. S3 endpoints, whether they are Gateway or Interface, cannot route traffic to cross-region buckets. Furthermore, Gateway endpoints cannot accept traffic from a remote VPC via a peering connection, so the Interface endpoint is the only functional option. This creates an ENI to accept the S3 traffic, with an associated DNS entry which can be used from the remote VPC. The traffic remains private, meeting the security requirement.

A. Deploy a VPC Gateway endpoint into the us-west-2 VPC
B. Deploy a VPC in the ap-southeast-1 region with an S3 Interface endpoint
C. Configure a route table entry in the us-west-2 VPC to direct S3 traffic to the Gateway endpoint
D. Configure a VPC peering connection between the us-west-2 and ap-southeast-1 VPCs
E. Configure the application to use the ap-southeast-1 region when accessing S3 buckets
F. Configure the application to use the ap-southeast-1 Interface endpoint DNS when accessing S3 buckets

Correct Answer

Correct Answers: B,D,F

A. Deploy a VPC Gateway endpoint into the us-west-2 VPC
B. Deploy a VPC in the ap-southeast-1 region with an S3 Interface endpoint
C. Configure a route table entry in the us-west-2 VPC to direct S3 traffic to the Gateway endpoint
D. Configure a VPC peering connection between the us-west-2 and ap-southeast-1 VPCs
E. Configure the application to use the ap-southeast-1 region when accessing S3 buckets
F. Configure the application to use the ap-southeast-1 Interface endpoint DNS when accessing S3 buckets

Question Domain 1: Design Secure Architectures

Automate security best practices

Principle Definition

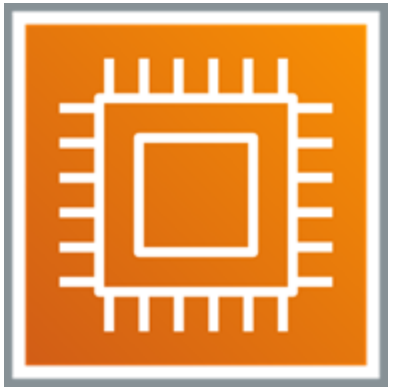
Automated software-based security mechanisms improve your ability to securely scale more rapidly and cost-effectively.

Create secure architectures, including the implementation of controls that are defined and managed as code in version-controlled templates.

What is Security Automation?

- Automate the implementation of guardrails and controls
- Consider how to scale the automation itself

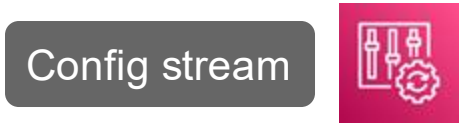
Compute Protection Automation



- Reduce network attack surface
- Implement regular vulnerability scanning
- Deploy patches in a timely manner
- Prefer immutability

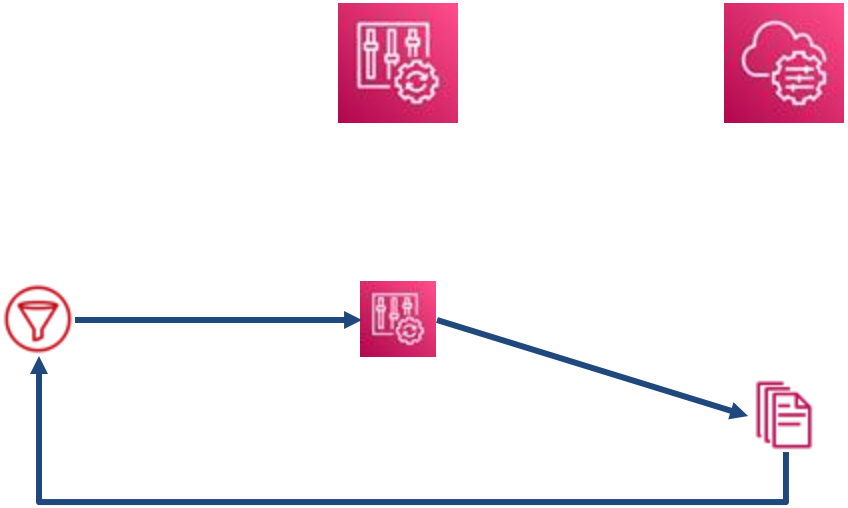
Security Automation in AWS Config

Config stream



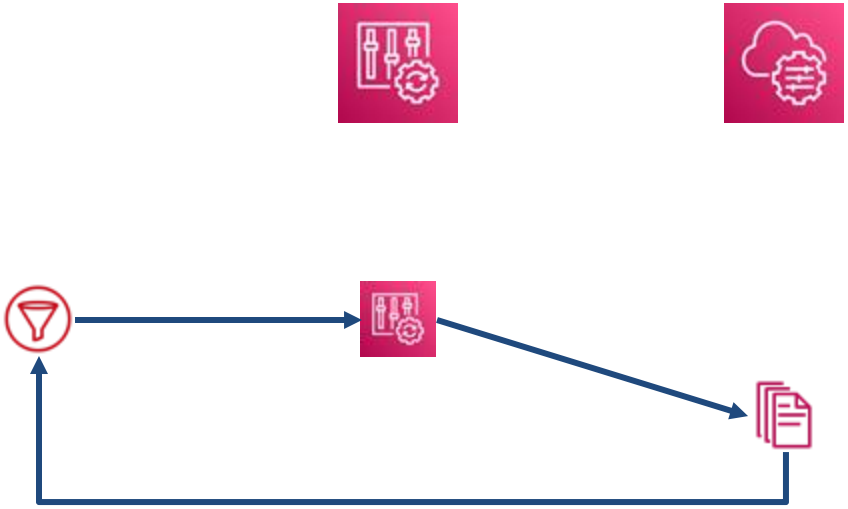
WAF Web ACL with logging disabled

Security Automation in AWS Config



Automated remediation using AWS SSM Automation document

Security Automation in AWS Config



Does this scale to all regions and accounts?

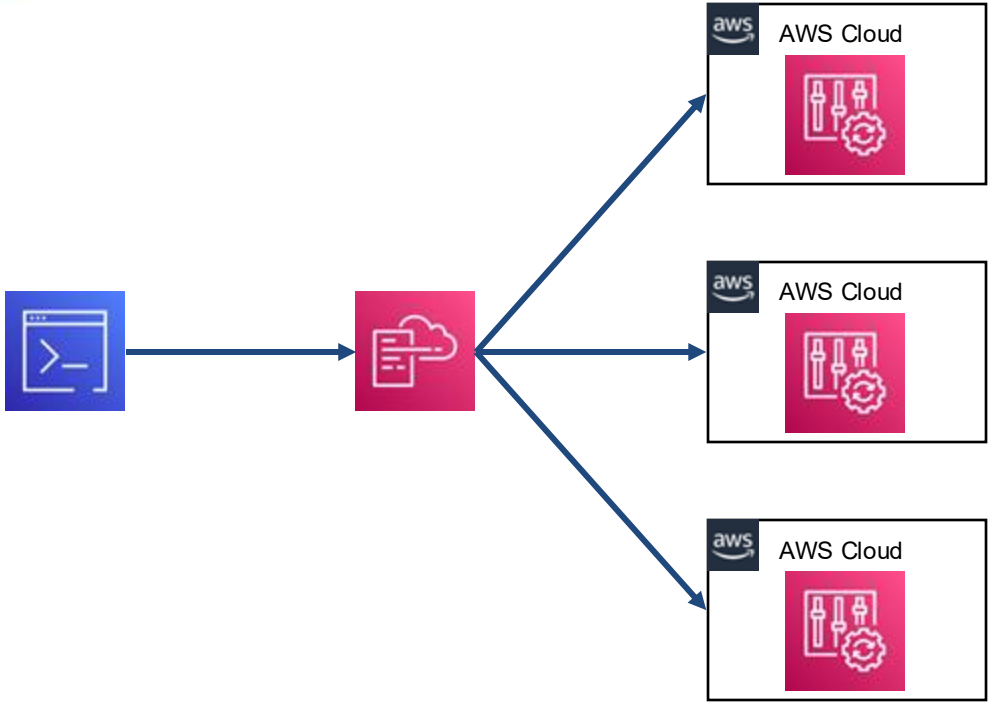
Config Automation Scaling

Create a conformance pack for WAF logging using a text editor and command line

Config Automation Scaling

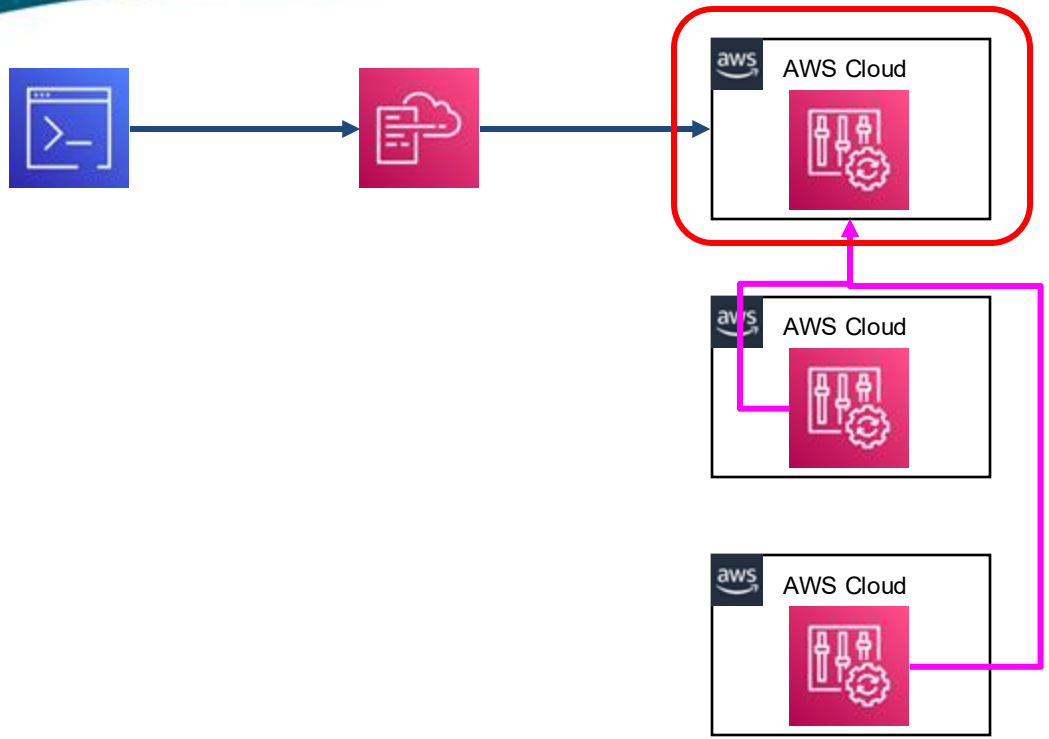
The conformance pack is deployed as a CloudFormation stack

Config Automation Scaling



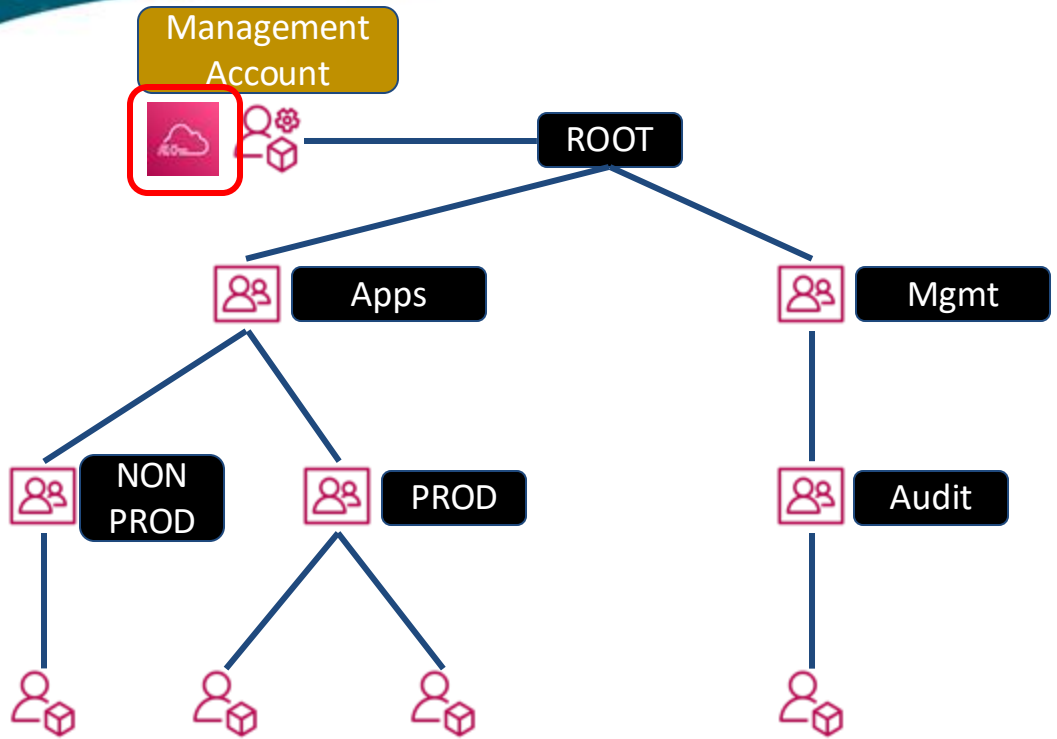
Deploy the conformance pack to all accounts and regions as required

Config Automation Scaling



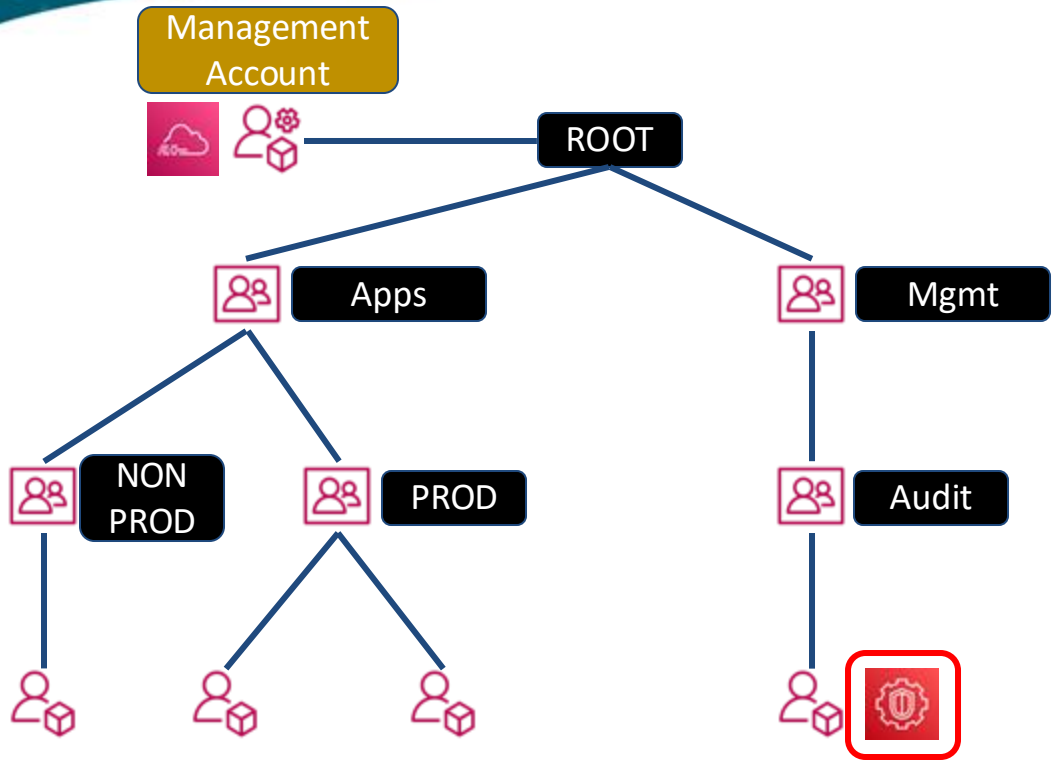
OR: deploy Config as delegated admin in one organization account and configure in one place!

Security Automation Scaling in Organizations



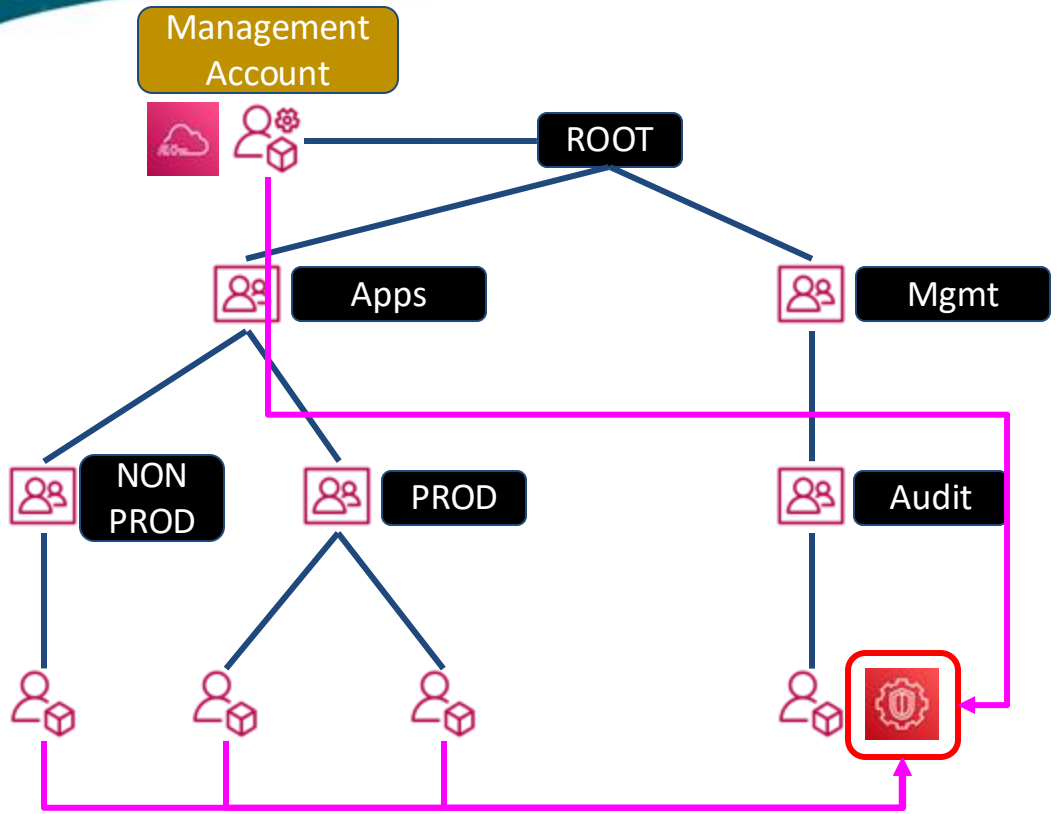
Create Organization trail in management account to centralize all CloudTrail logs

Security Automation Scaling in Organizations



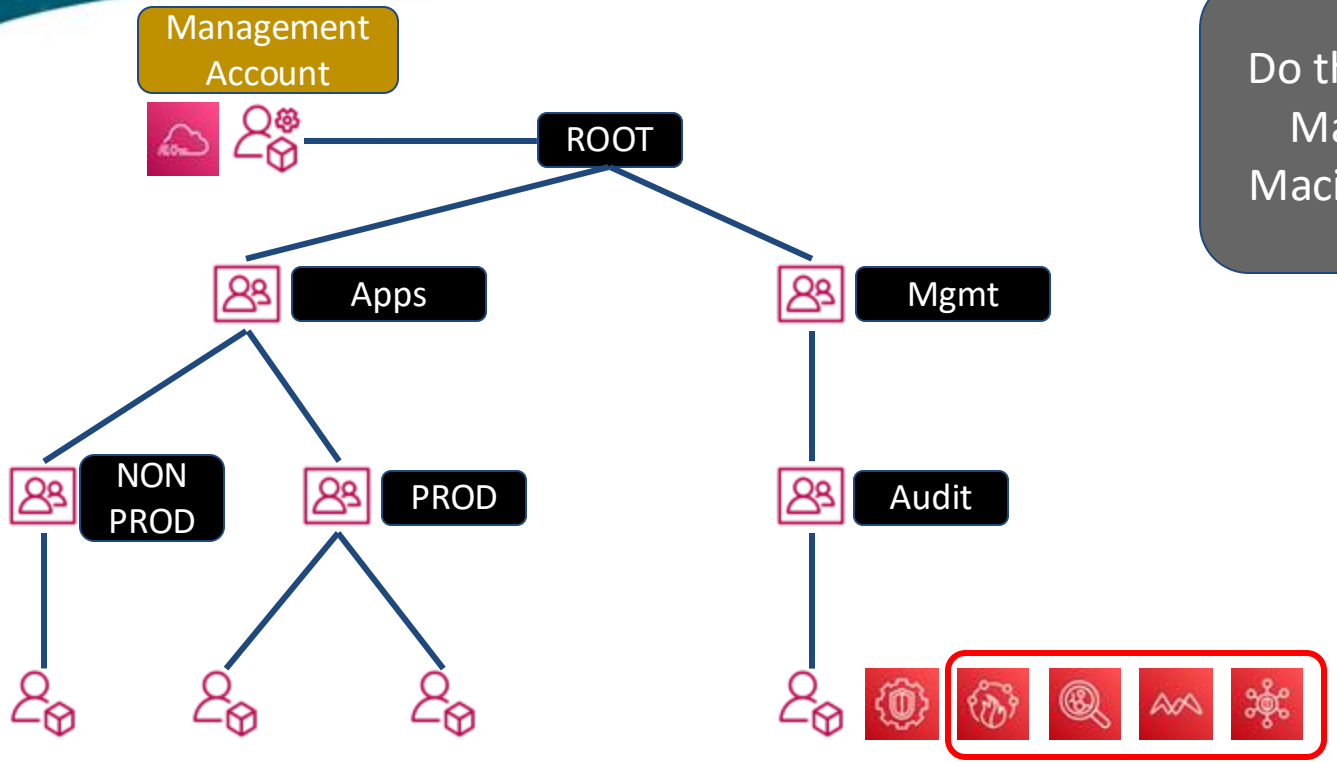
Configure GuardDuty as delegated administrator in the security audit account

Security Automation Scaling in Organizations



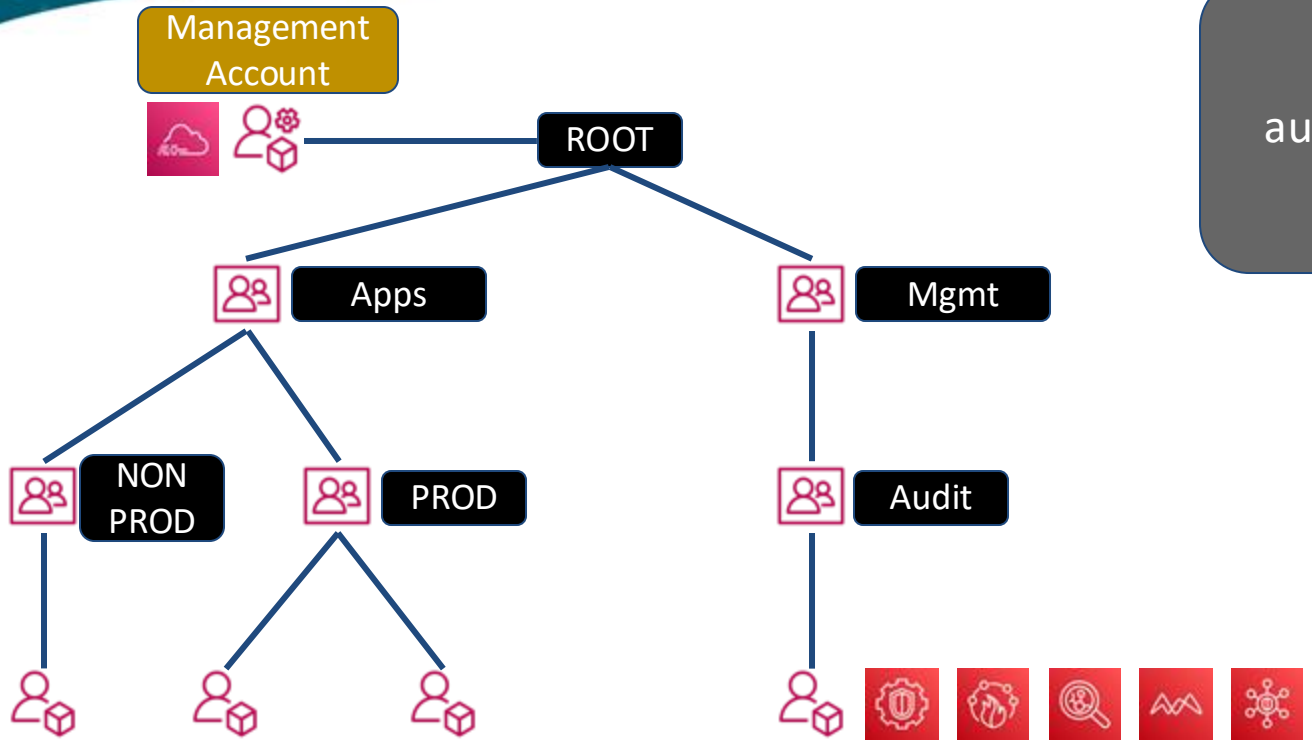
All findings are now delivered from all regions and accounts to the delegated admin account

Security Automation Scaling in Organizations



Do the same for Firewall Manager, Inspector, Macie, and Security Hub

Security Automation Scaling in Organizations



This simplifies automation of security governance

Question Breakdown

Automate security best practices

Question and Answer Choices

A financial services company is planning to deploy a new application on AWS. The application will handle sensitive customer data, including financial transactions. The company's security team wants to ensure that the application and its environment adhere to strict security best practices from the moment they are deployed. They are particularly interested in minimizing human error and ensuring consistent application of security controls across their AWS environment.

Which AWS service should the security team use?

A. AWS Security Hub
B. AWS Config
C. AWS Systems Manager
D. AWS CloudFormation

Answer Choice Explanation

AWS Security Hub provides a comprehensive view of your security state within AWS and can help you check your environment against security industry standards and best practices. However, it does not automate the deployment of security controls or resources.

A. **AWS Security Hub**
B. **AWS Config**
C. **AWS Systems Manager**
D. **AWS CloudFormation**

Answer Choice Explanation

While AWS Config can be used to assess, audit, and evaluate the configurations of your AWS resources, it is more focused on monitoring and recording the configurations of AWS resources and evaluating those configurations for compliance with desired guidelines. It does not automate the deployment of resources or security controls.

A. AWS Security Hub
B. AWS Config
C. AWS Systems Manager
D. AWS CloudFormation

Answer Choice Explanation

Although AWS Systems Manager provides visibility and control of your infrastructure on AWS, it is primarily used for operational tasks such as patch management, application deployment, and resource configuration. It does not specifically automate the deployment of security controls based on predefined security standards.

A. AWS Security Hub
B. AWS Config
C. AWS Systems Manager
D. AWS CloudFormation

Answer Choice Explanation

CloudFormation allows you to model and set up your resources so that you can spend less time managing those resources and more time focusing on your applications. You can use AWS CloudFormation to automate the deployment of resources in a secure configuration, ensuring that all resources comply with the company's security standards.

A. AWS Security Hub
B. AWS Config
C. AWS Systems Manager
D. AWS CloudFormation

Correct Answer

Correct Answer: D

A. AWS Security Hub
B. AWS Config
C. AWS Systems Manager
D. AWS CloudFormation

Question Domain 1: Design Secure Architectures

Protect data in transit and at rest

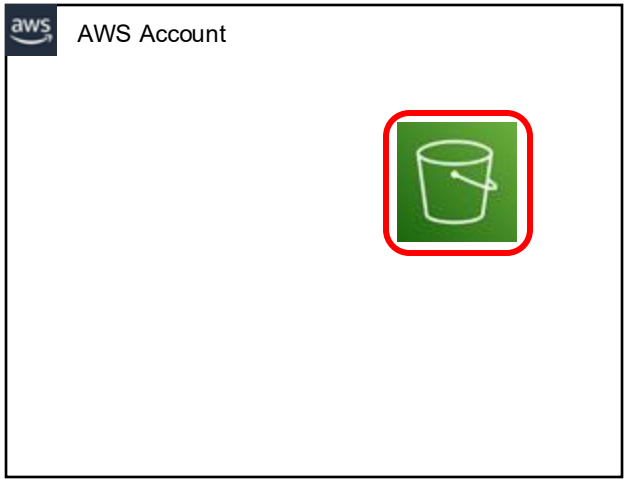
Principle Definition

Classify your data into sensitivity levels and use mechanisms, such as encryption, tokenization, and access control where appropriate.

Data Classification Steps

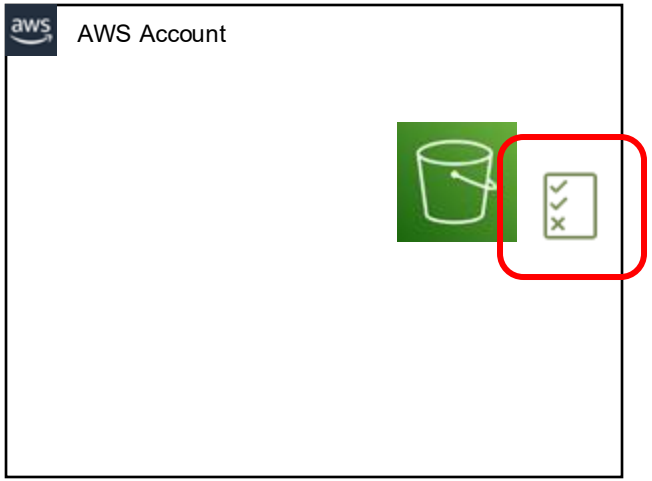
- Identify data within your workload
- Define data protection controls
- Define data lifecycle management
- Automate identification and classification

Data Protection Control ABAC Example



Objects placed in an S3 bucket can be tagged to indicate the presence of sensitive data

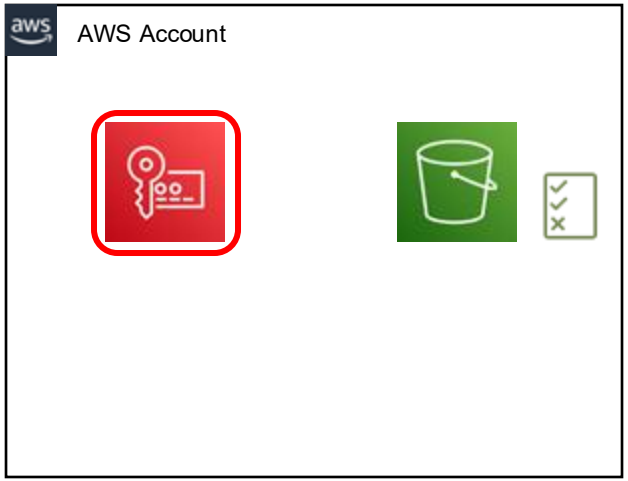
Data Protection Control ABAC Example



The S3 bucket policy can enforce object tagging as an active guardrail

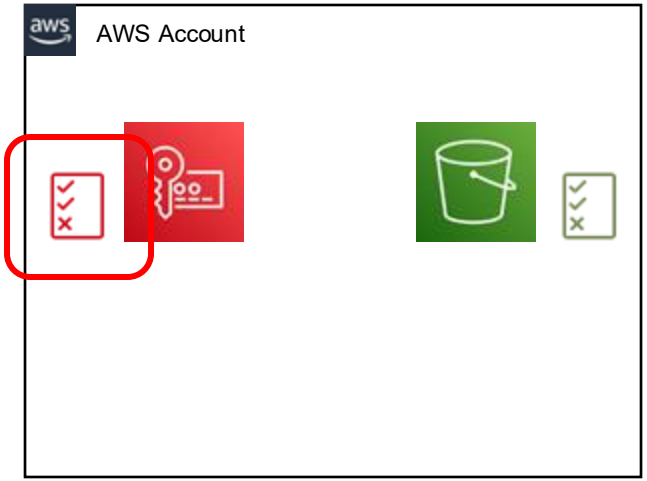
Data Protection Control ABAC Example

The object can be encrypted at rest using KMS



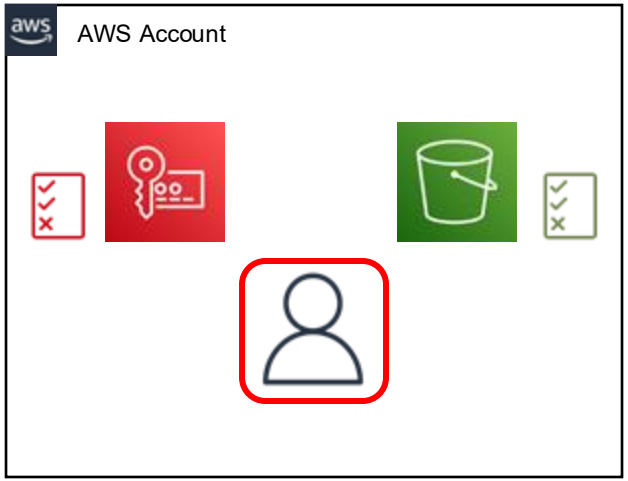
Data Protection Control ABAC Example

The KMS key policy can also enforce encryption against properly tagged objects



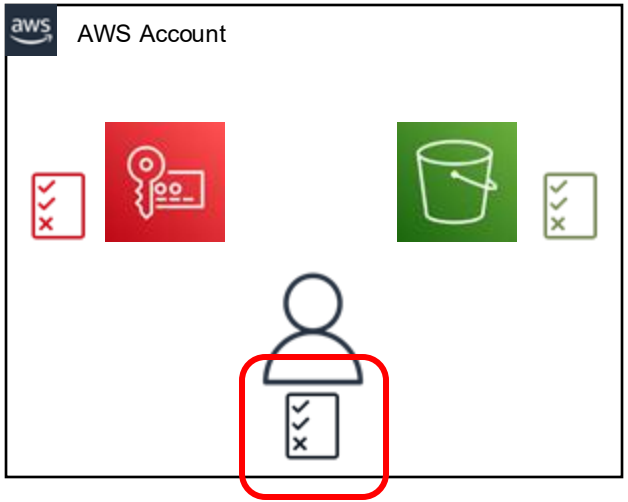
Data Protection Control ABAC Example

An IAM user can be tagged to indicate access to sensitive data



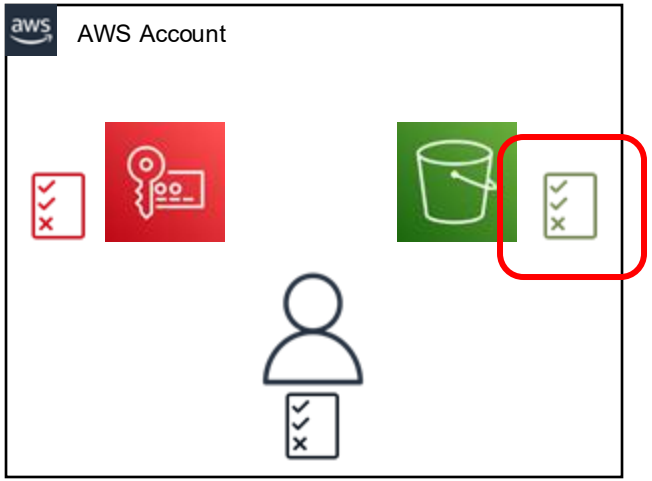
Data Protection Control ABAC Example

IAM permission policies can grant access to users and objects that have specified tags

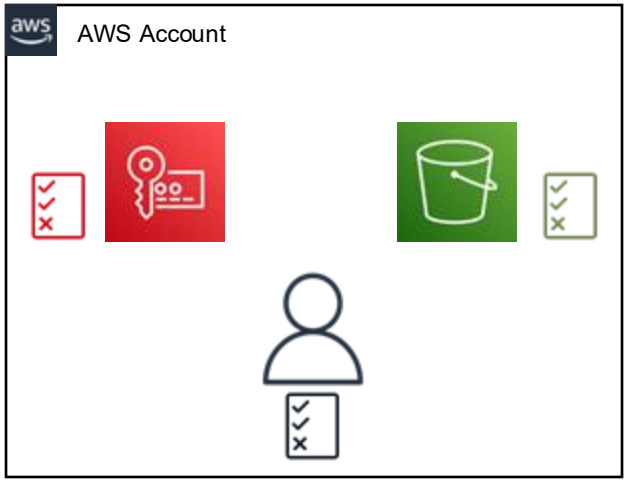


Data Protection Control ABAC Example

The bucket policy can enforce access only by properly tagged IAM principals



Data Protection Control ABAC Example



An Organizations tag policy can enforce the presence of tags on all resources and environments

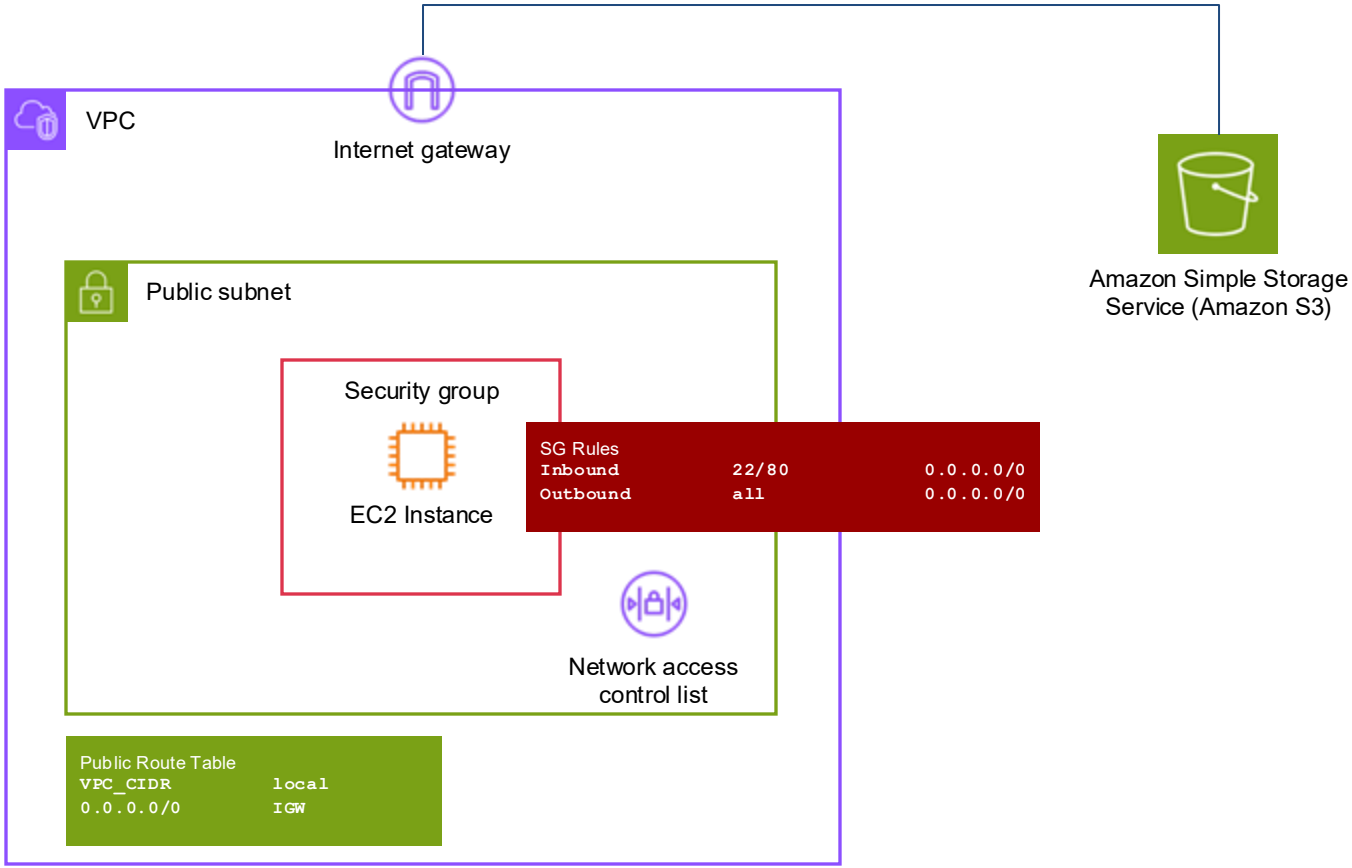
Network Security Scenario Description

A security architect has been asked to review a global network architecture. The network includes multiple VPCs in different regions within a single AWS account.

The security architect has been asked to provide recommendations to improve the overall security of the global network.

What network security improvements can be made to meet the requirements?

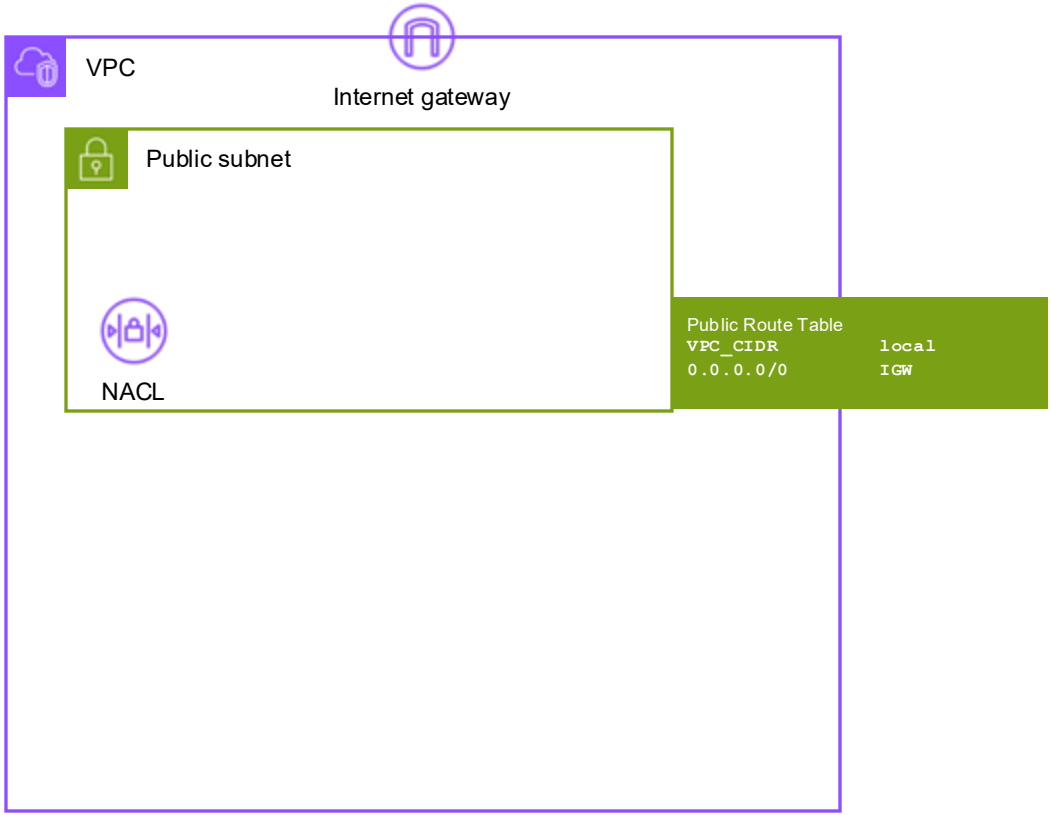
Existing Infrastructure (Single VPC)



Single VPC Security Improvement Suggestions

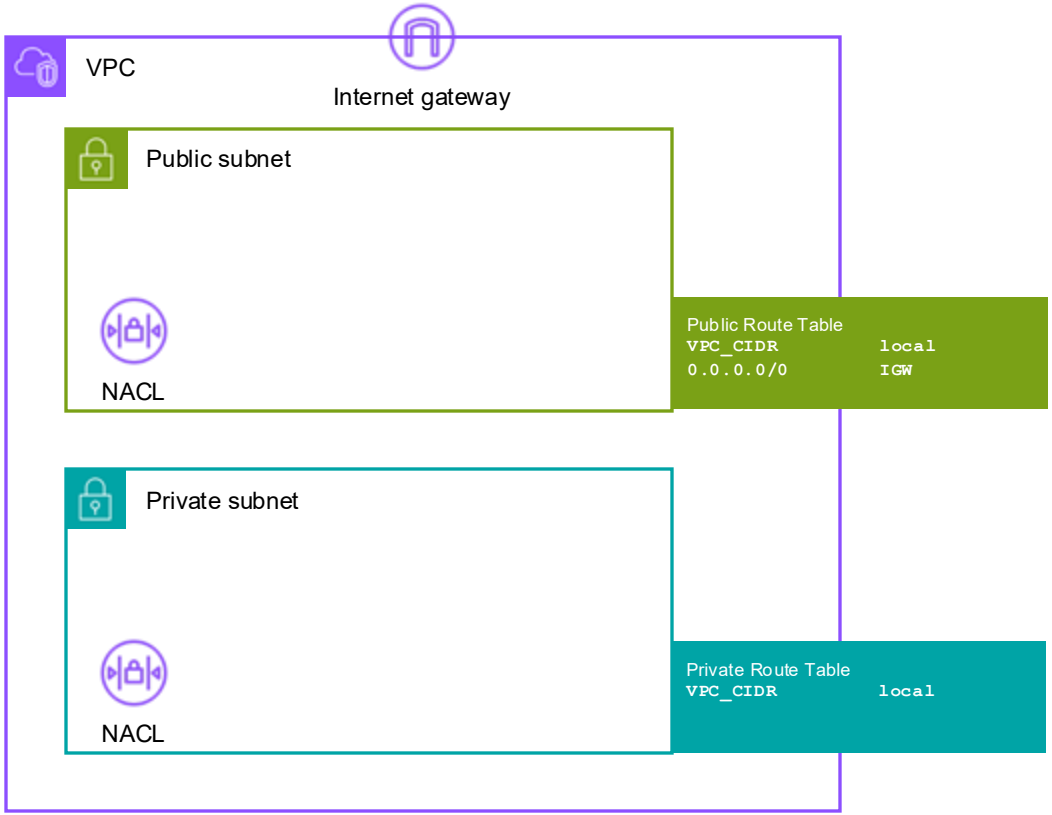
- Private subnet(s)
- NAT Gateway
- Least privilege security group rules
- S3 Gateway endpoint
- SSM Session Manager
- SSM Interface endpoint
- VPC Block Public Access(?) (NEW)

Single VPC Security Improvements



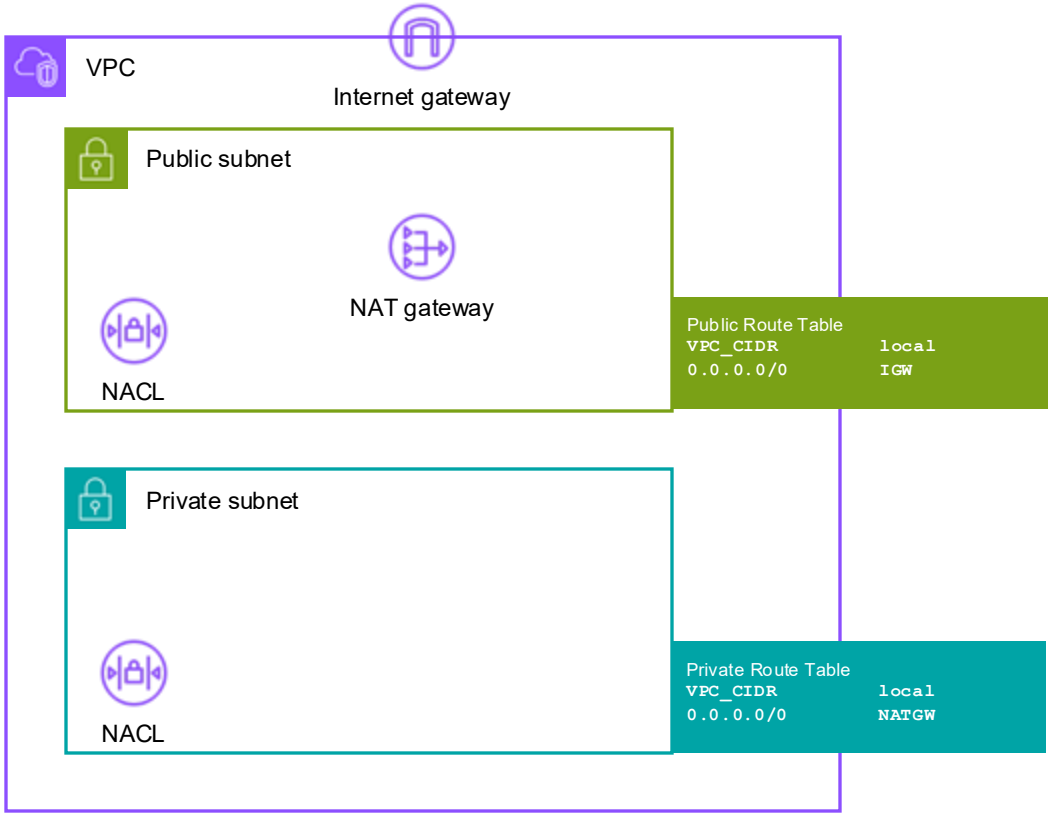
Amazon Simple Storage Service (Amazon S3)

Single VPC Security Improvements



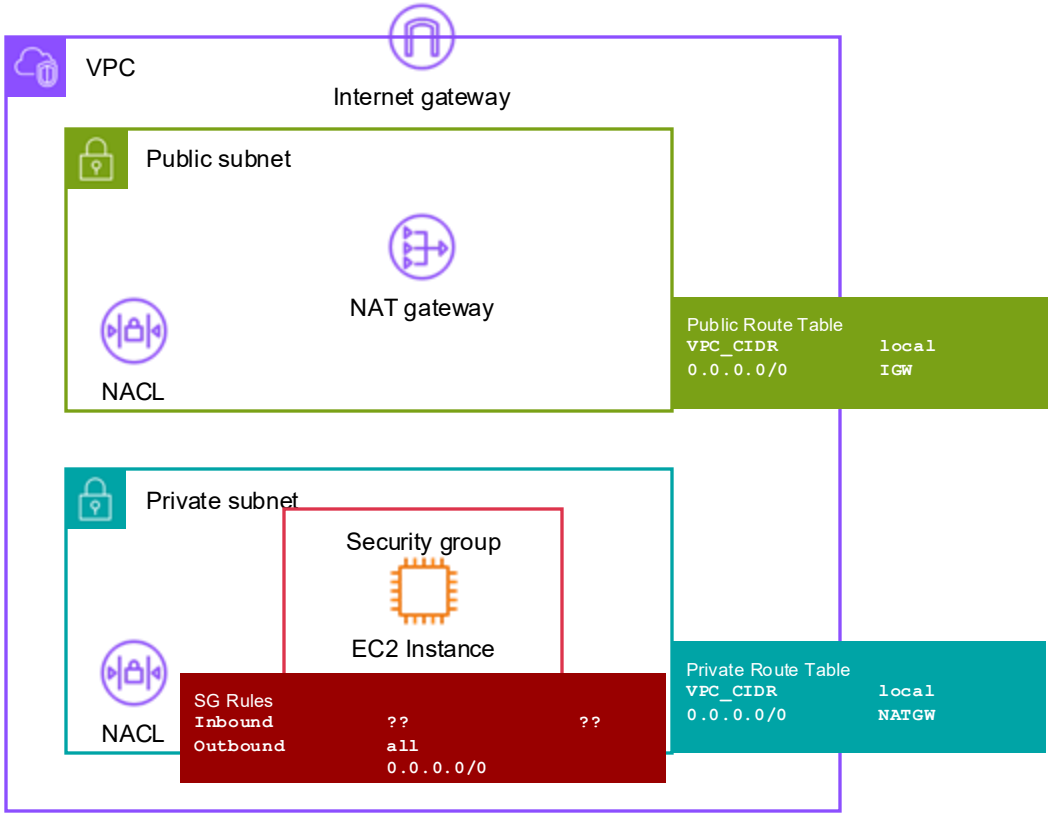
Amazon Simple Storage Service (Amazon S3)

Single VPC Security Improvements



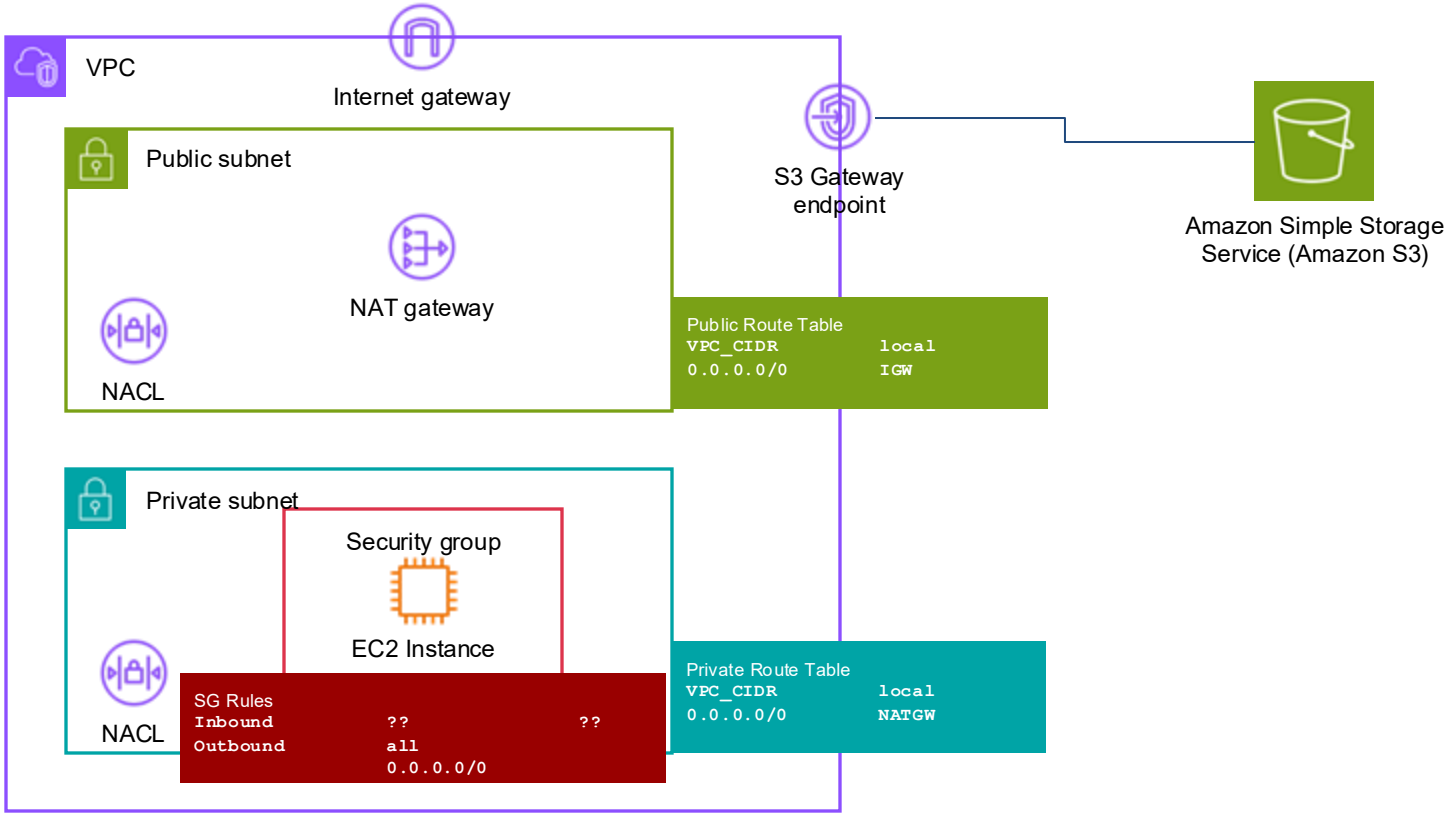
Amazon Simple Storage Service (Amazon S3)

Single VPC Security Improvements

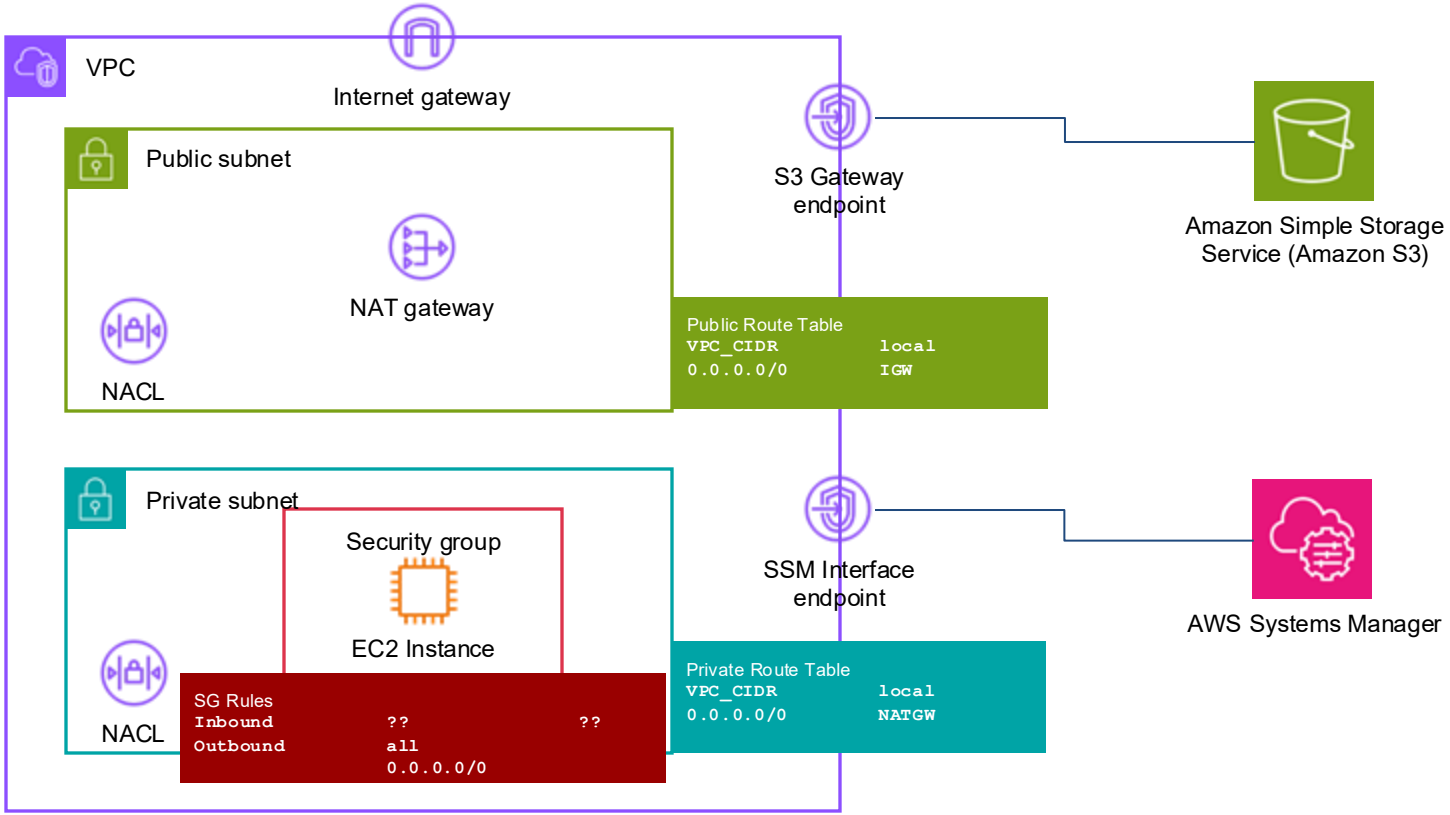


Amazon Simple Storage Service (Amazon S3)

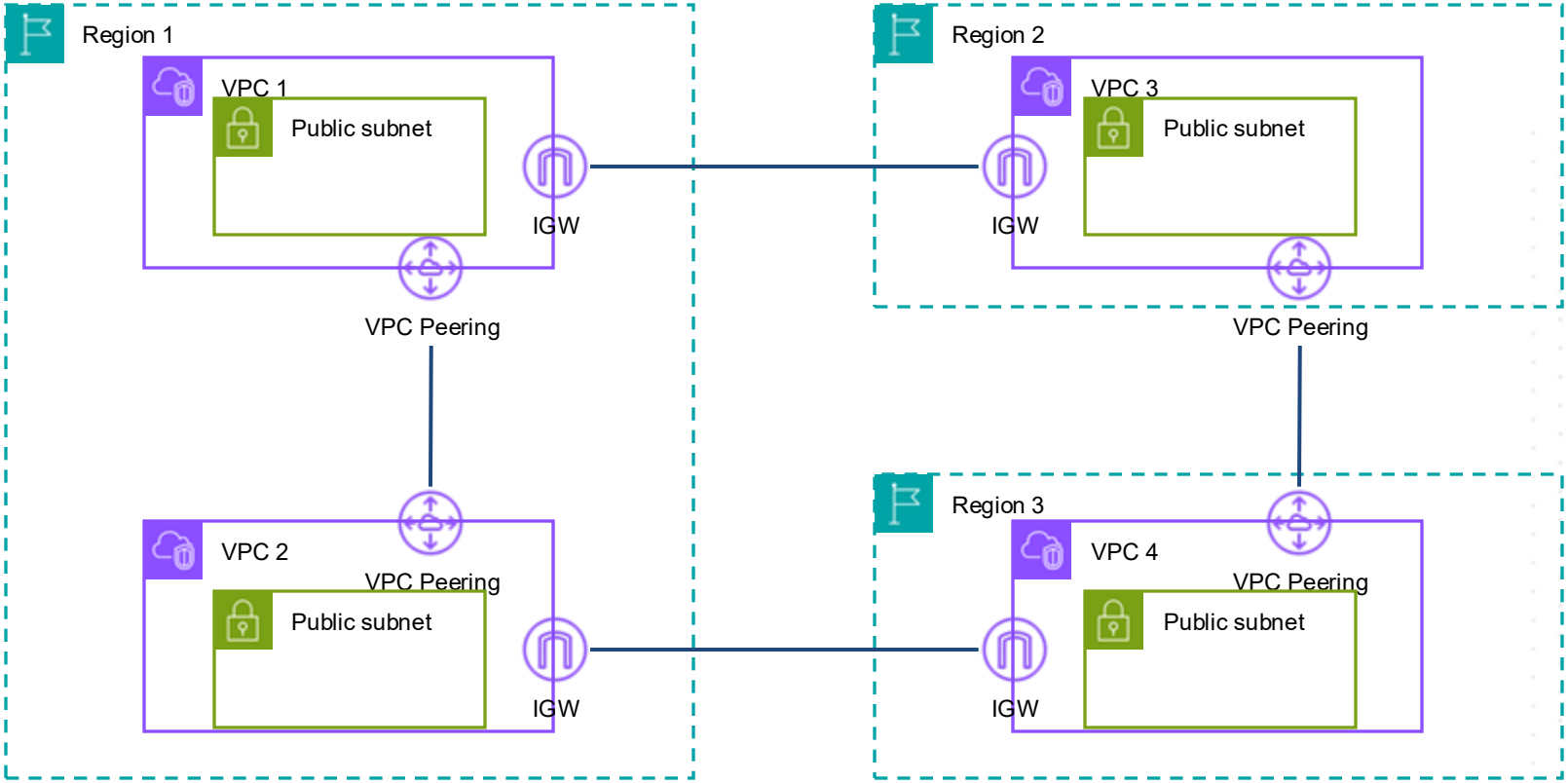
Single VPC Security Improvements



Single VPC Security Improvements



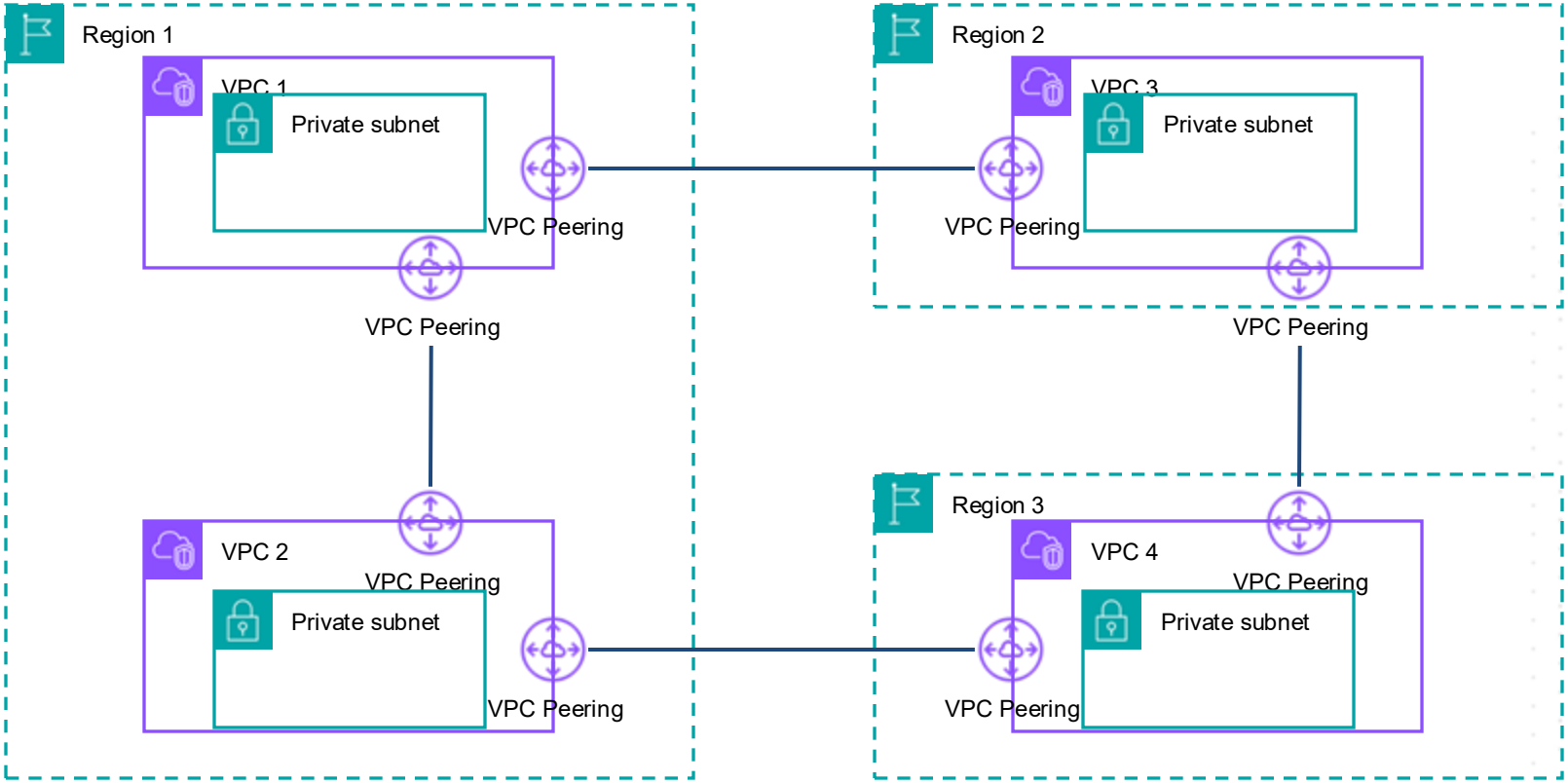
Existing Infrastructure (Global Network)



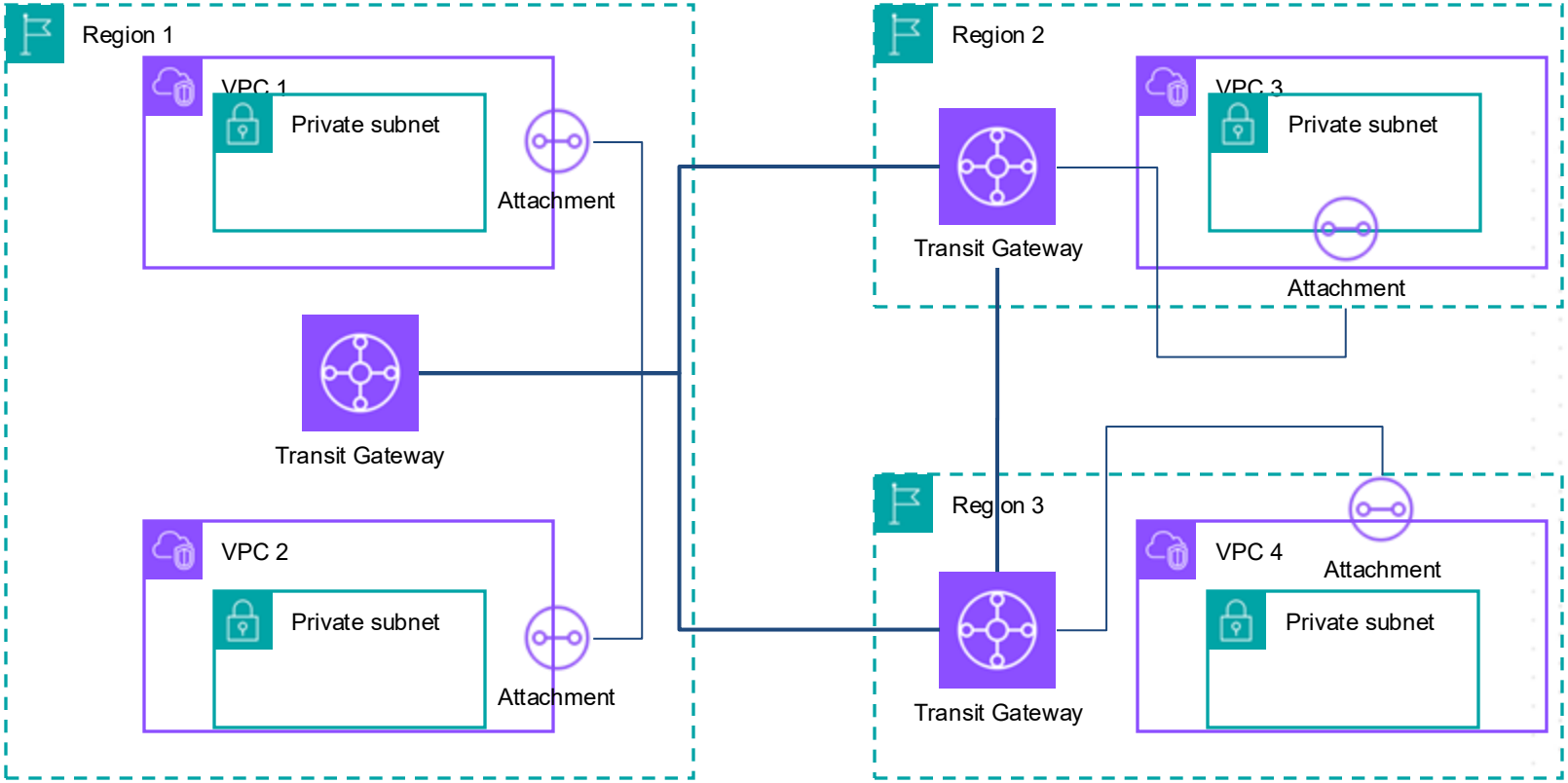
Global Network Security Improvement Suggestions

- VPC Peering instead of IGW
- Only private subnet connectivity between VPCs
- Transit Gateway (?)
- VPC Lattice(?)

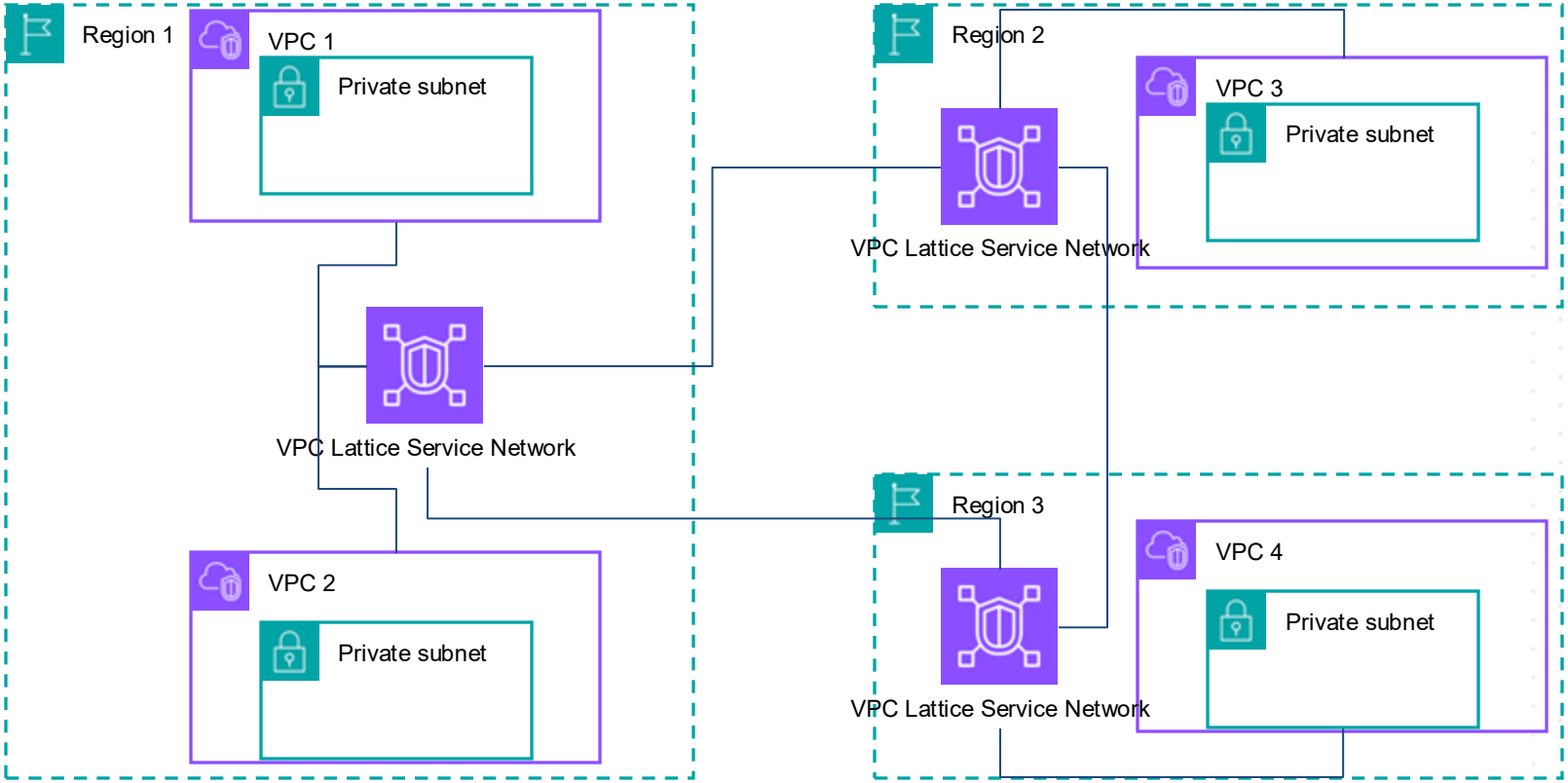
Global Network Security Improvements



Global Network Security Improvements (Alternative)



Global Network Security Improvements (Alternative 2)



VPC Connectivity Alternatives

- PrivateLink
- VPC Lattice
- On-prem + site-to-site VPN
- On-prem + Direct Connect

Question Breakdown

Protect data in transit and at rest

Question and Answer Choices

Your company securely stores sensitive data in S3. There is a strict requirement to ensure that all data is encrypted at rest using the most secure method possible.

Which of the following encryption configurations would best meet the requirement for safeguarding this sensitive information against unauthorized access?

A. SSE-S3
B. SSE-KMS
C. SSE-C
D. Client-side encryption

Answer Choice Explanation

SSE-S3 uses AES-256 with AWS owning the entire chain of trust (Root CA, master key, data key). The encryption is performed in the S3 service. This is slightly better than no encryption at all.

A. SSE-S3
B. SSE-KMS
C. SSE-C
D. Client-side encryption

Answer Choice Explanation

This is a better option than A. AWS owns the Root CA, and the customer owns the remainder of the chain of trust. The encryption is performed in the S3 service.

A. SSE-S3
B. SSE-KMS
C. SSE-C
D. Client-side encryption

Answer Choice Explanation

This is more secure than B, as this allows for the customer to own the entire chain of trust. The encryption is still performed in the S3 service using a data key provided by the customer.

A. SSE-S3
B. SSE-KMS
C. SSE-C
D. Client-side encryption

Answer Choice Explanation

This solution does not use AWS for encryption whatsoever, and, like C, allows the customer to own the entire chain of trust. This would be the most secure option.

A. SSE-S3
B. SSE-KMS
C. SSE-C
D. Client-side encryption

Correct Answer

Correct Answer: D

A. SSE-S3
B. SSE-KMS
C. SSE-C
D. Client-side encryption

Question Domain 1: Design Secure Architectures

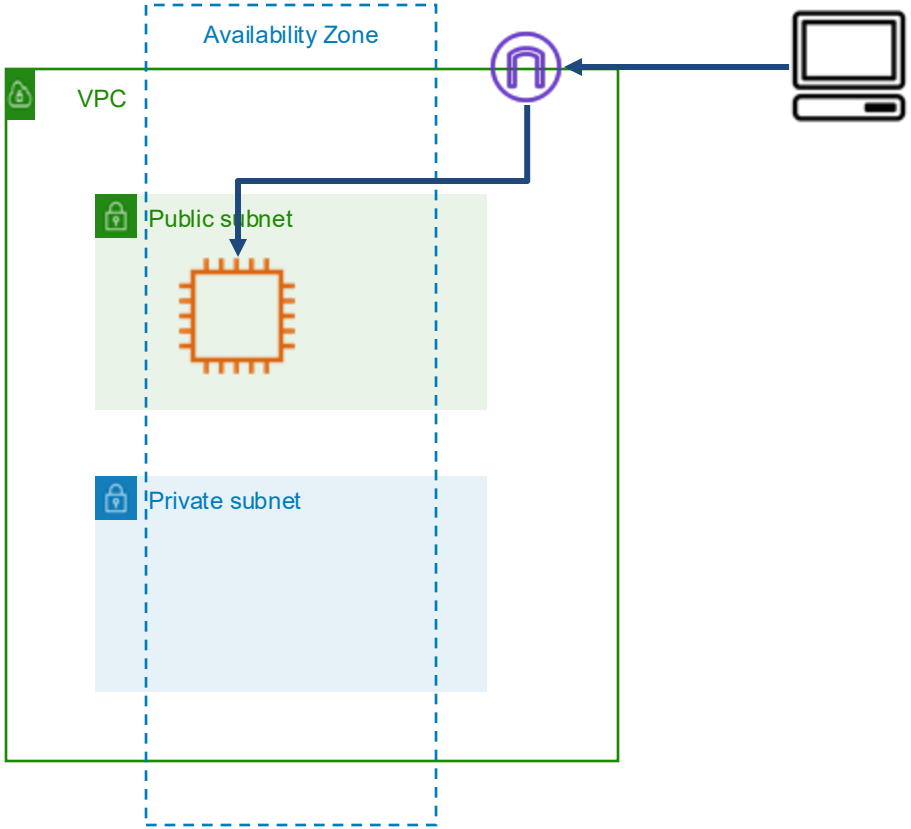
Keep people away from data

Principle Definition

Use mechanisms and tools to reduce or eliminate the need for direct access or manual processing of data.

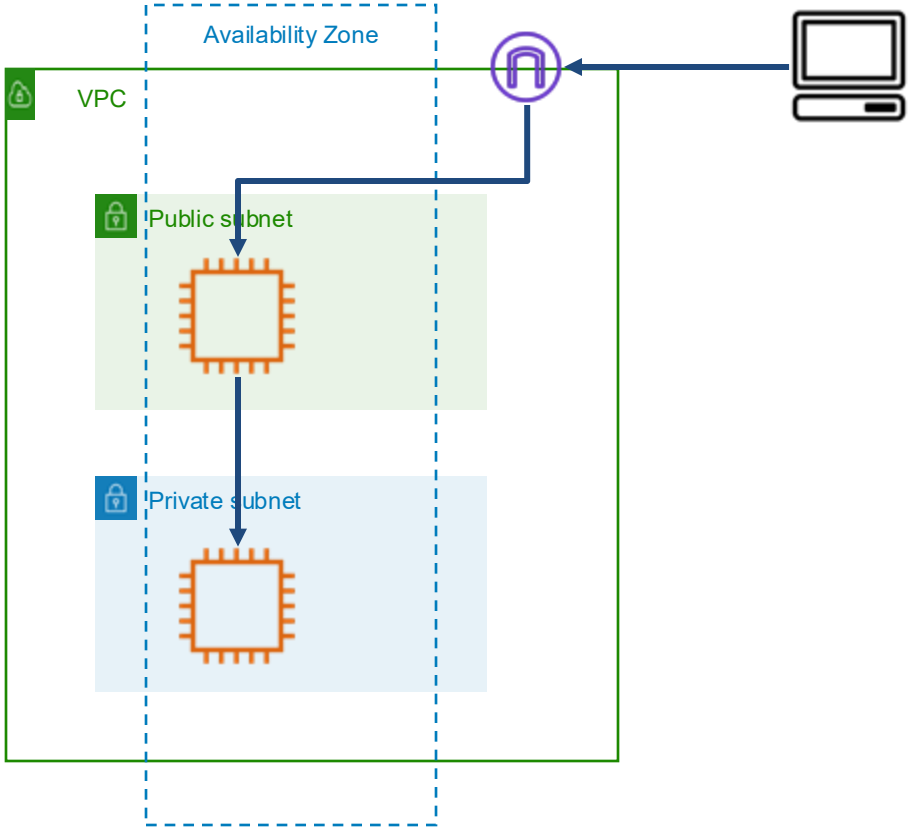
This reduces the risk of mishandling or modification and human error when handling sensitive data.

Minimize Direct EC2 Access



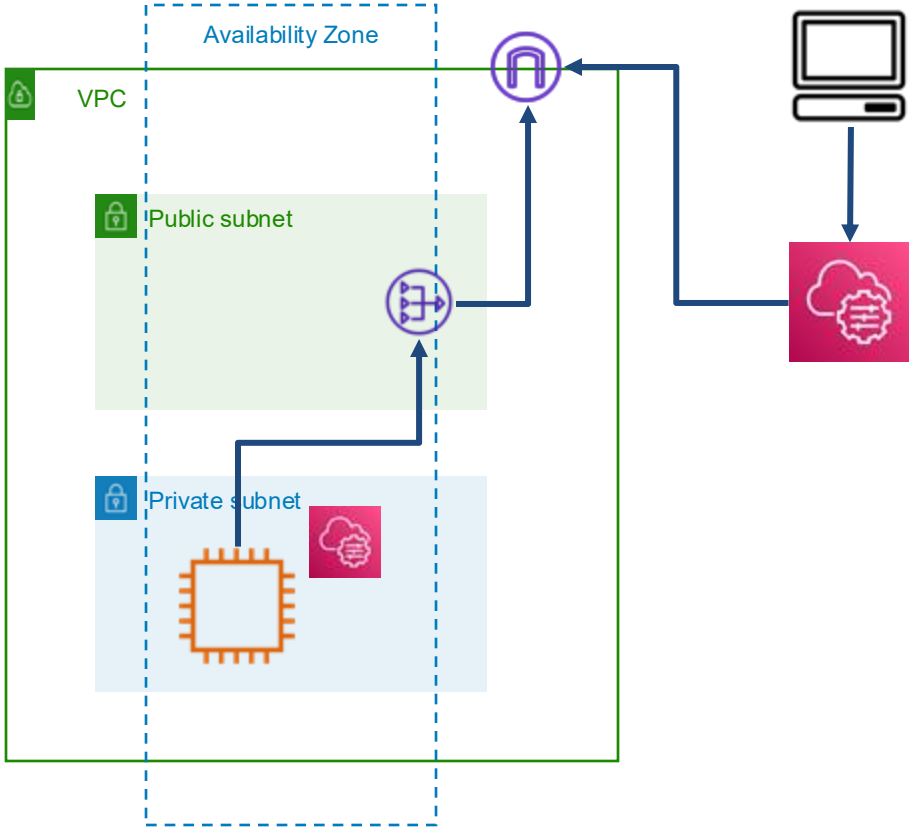
Anti pattern: direct inbound access to public subnet

Minimize Direct EC2 Access



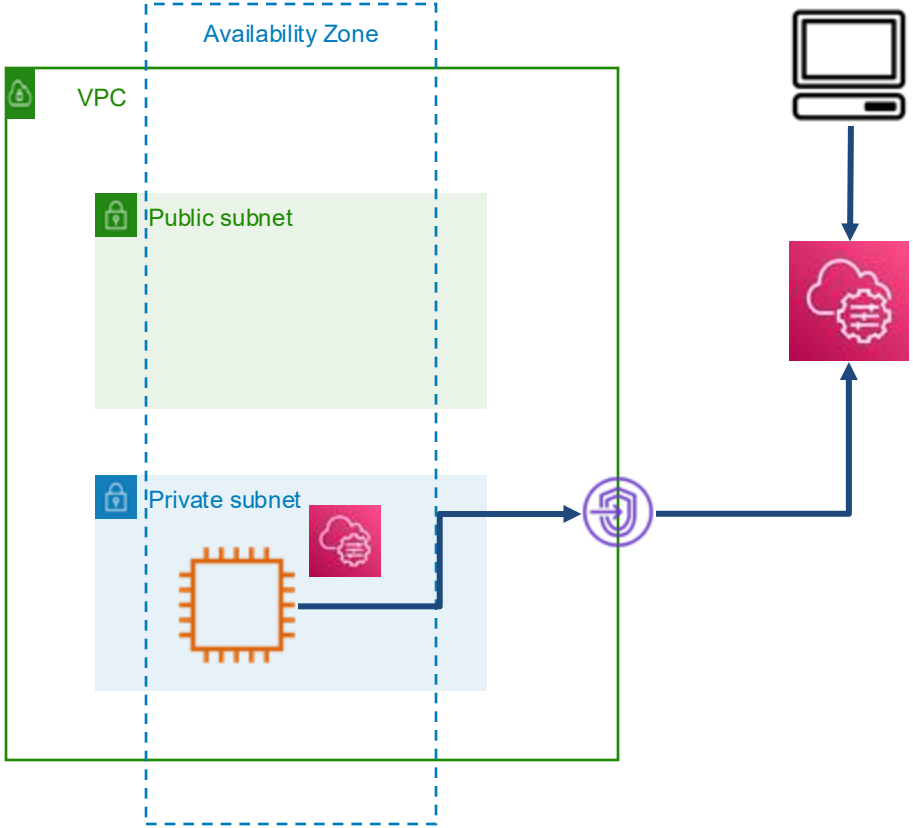
Anti pattern: inbound access to private subnet via bastion host

Minimize Direct EC2 Access



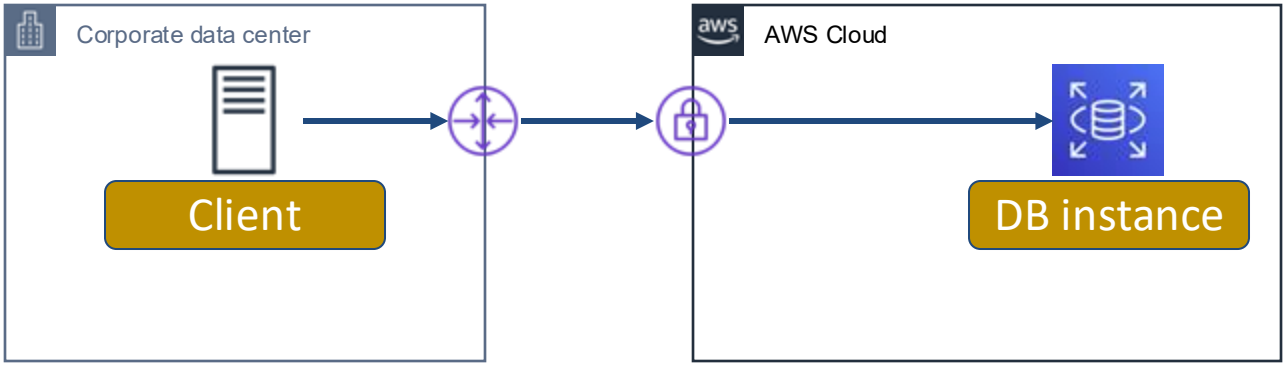
Better: use Systems Manager features to inventory, patch, execute commands, and explore

Minimize Direct EC2 Access



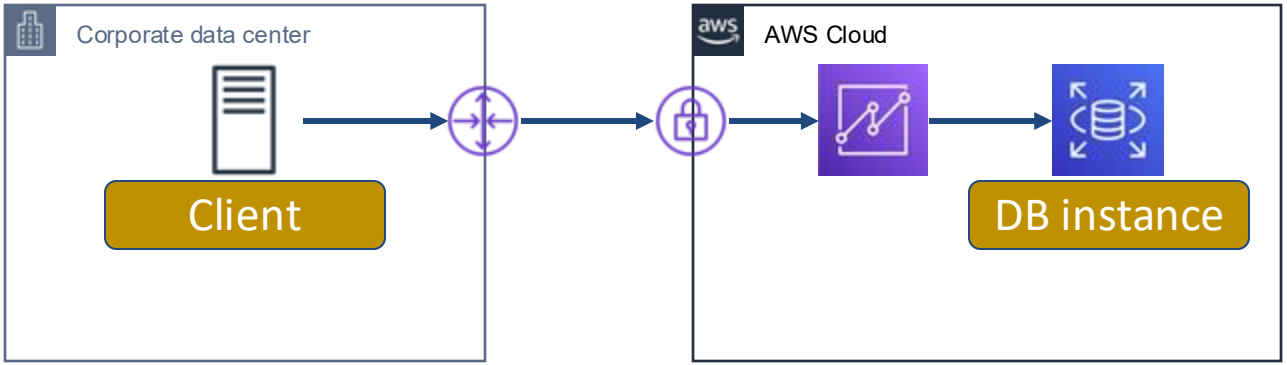
Best: use Systems Manager features via Interface Endpoint

Minimize Direct DB Access



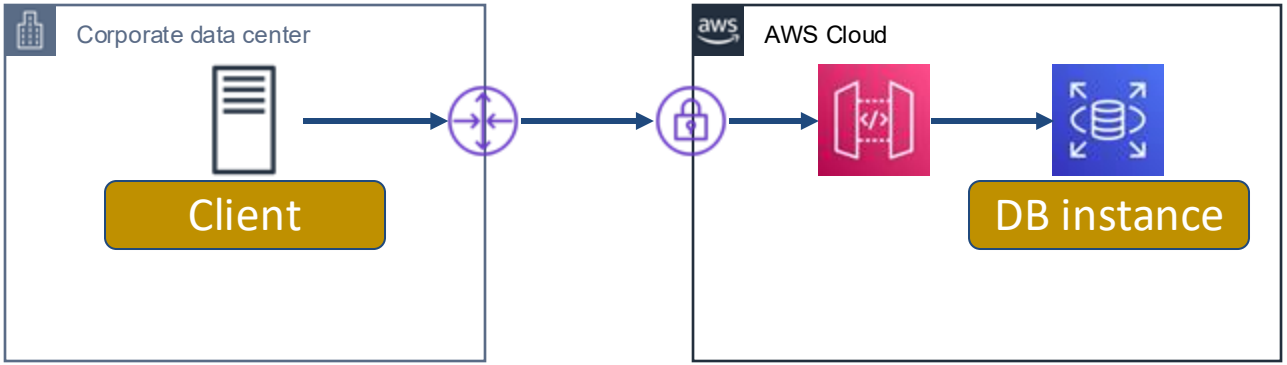
Anti pattern: direct inbound access from corporate network to database server

Minimize Direct DB Access



Better: limit client access through QuickSight dashboards

Minimize Direct DB Access



Also better: use ORM and API Gateway

Minimize Direct Data Access

- Store data in encrypted format
- Anonymize by hashing
- Pseudonymize by splitting fields into different objects
- ALWAYS use state of the art algorithms

Question Breakdown

Keep people away from data

Question and Answer Choices

An on-premises application stores data containing PII on an EFS file system, reachable through a site-to-site VPN. There is a requirement for each file to have certain fields tokenized as quickly as possible to avoid the raw data being easily accessed. The tokenization must be completed entirely within AWS.

Which of the following workflows would meet the requirements?

A. Deploy a Lambda function which accesses EFS directly and uses KMS to encrypt the fields inline in each file.
B. Deploy a Lambda function which accesses EFS directly. Generate a random string to replace each field, then extract the plaintext from each field. Encrypt the field using KMS and insert into a DynamoDB table with the random string as the partition key.
C. Use an EMR cluster to access the EFS filesystem and use KMS to encrypt the fields inline in each file.
D. Deploy an AWS Batch job which accesses EFS directly. Generate a random string to replace each field, then extract the plaintext from each field. Encrypt the field using KMS and insert into a DynamoDB table with the random string as the partition key.

Answer Choice Explanation

This is functional as far as AWS is concerned, and can be executed in near real-time. It doesn't accomplish tokenization, which is a form of pseudonymization where certain sensitive data is separated from the rest of the data to make it no longer identifiable.

A. **Deploy a Lambda function which accesses EFS directly and uses KMS to encrypt the fields inline in each file.**
B. **Deploy a Lambda function which accesses EFS directly. Generate a random string to replace each field, then extract the plaintext from each field. Encrypt the field using KMS and insert into a DynamoDB table with the random string as the partition key.**
C. **Use an EMR cluster to access the EFS filesystem and use KMS to encrypt the fields inline in each file.**
D. **Deploy an AWS Batch job which accesses EFS directly. Generate a random string to replace each field, then extract the plaintext from each field. Encrypt the field using KMS and insert into a DynamoDB table with the random string as the partition key.**

Answer Choice Explanation

Like A, this is functional and near real-time, but it also accomplishes the pseudonymization by separating some data into DynamoDB with a link back to the text string in the original data.

A. Deploy a Lambda function which accesses EFS directly and uses KMS to encrypt the fields inline in each file.
B. Deploy a Lambda function which accesses EFS directly. Generate a random string to replace each field, then extract the plaintext from each field. Encrypt the field using KMS and insert into a DynamoDB table with the random string as the partition key.
C. Use an EMR cluster to access the EFS filesystem and use KMS to encrypt the fields inline in each file.
D. Deploy an AWS Batch job which accesses EFS directly. Generate a random string to replace each field, then extract the plaintext from each field. Encrypt the field using KMS and insert into a DynamoDB table with the random string as the partition key.

Answer Choice Explanation

Similar to A, this is functional but does not meet either the requirement for tokenization or the requirement for "as soon as possible".

A. Deploy a Lambda function which accesses EFS directly and uses KMS to encrypt the fields inline in each file.
B. Deploy a Lambda function which accesses EFS directly. Generate a random string to replace each field, then extract the plaintext from each field. Encrypt the field using KMS and insert into a DynamoDB table with the random string as the partition key.
C. Use an EMR cluster to access the EFS filesystem and use KMS to encrypt the fields inline in each file.
D. Deploy an AWS Batch job which accesses EFS directly. Generate a random string to replace each field, then extract the plaintext from each field. Encrypt the field using KMS and insert into a DynamoDB table with the random string as the partition key.

Answer Choice Explanation

While this accomplishes tokenization, it is not done in near real-time fashion.

A. Deploy a Lambda function which accesses EFS directly and uses KMS to encrypt the fields inline in each file.
B. Deploy a Lambda function which accesses EFS directly. Generate a random string to replace each field, then extract the plaintext from each field. Encrypt the field using KMS and insert into a DynamoDB table with the random string as the partition key.
C. Use an EMR cluster to access the EFS filesystem and use KMS to encrypt the fields inline in each file.
D. Deploy an AWS Batch job which accesses EFS directly. Generate a random string to replace each field, then extract the plaintext from each field. Encrypt the field using KMS and insert into a DynamoDB table with the random string as the partition key.

Correct Answer

Correct Answer: B

A. Deploy a Lambda function which accesses EFS directly and uses KMS to encrypt the fields inline in each file.
B. Deploy a Lambda function which accesses EFS directly. Generate a random string to replace each field, then extract the plaintext from each field. Encrypt the field using KMS and insert into a DynamoDB table with the random string as the partition key.
C. Use an EMR cluster to access the EFS filesystem and use KMS to encrypt the fields inline in each file.
D. Deploy an AWS Batch job which accesses EFS directly. Generate a random string to replace each field, then extract the plaintext from each field. Encrypt the field using KMS and insert into a DynamoDB table with the random string as the partition key.

Question Domain 1: Design Secure Architectures

Prepare for security events

Principle Definition

Prepare for an incident by having incident management and investigation policy and processes that align to your organizational requirements.

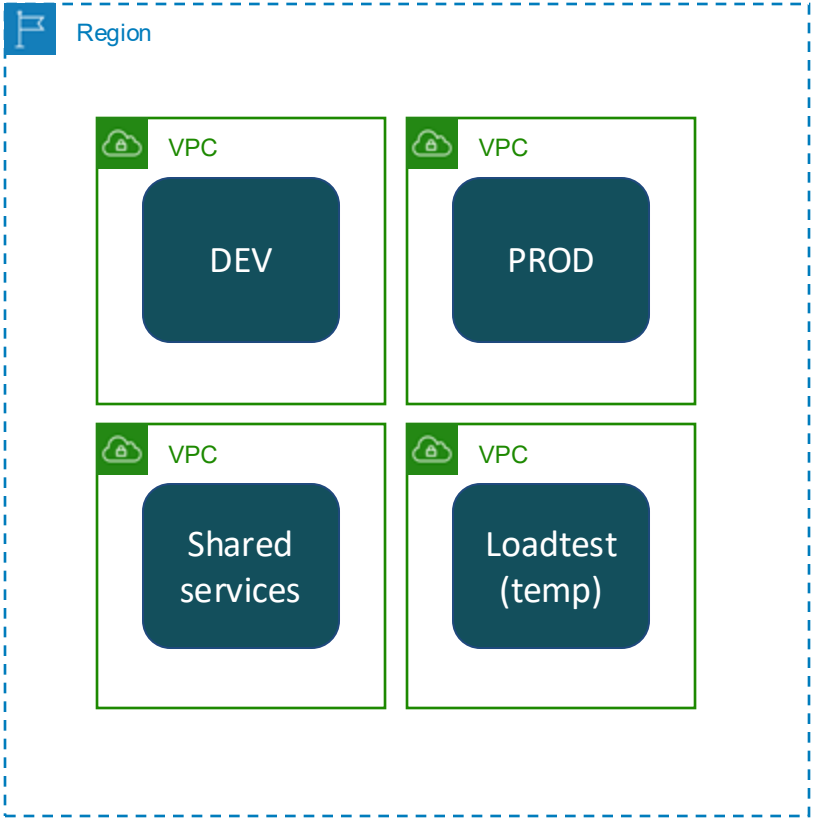
Run incident response simulations and use tools with automation to increase your speed for detection, investigation, and recovery.

Incident Response Steps

Prevent

Establish policies and guardrails, implement security features

IR Prevention Strategies



Isolate workloads into separate VPCs

IR Prevention Strategies

Isolate workloads into separate accounts

IR Prevention Strategies

AWS Config

AWS CloudFormation

AWS Systems Manager

- Self-documenting infrastructure
- Inventory and fleet management
- Explore configurations and associations

IR Prevention Strategies

Network access analyzer

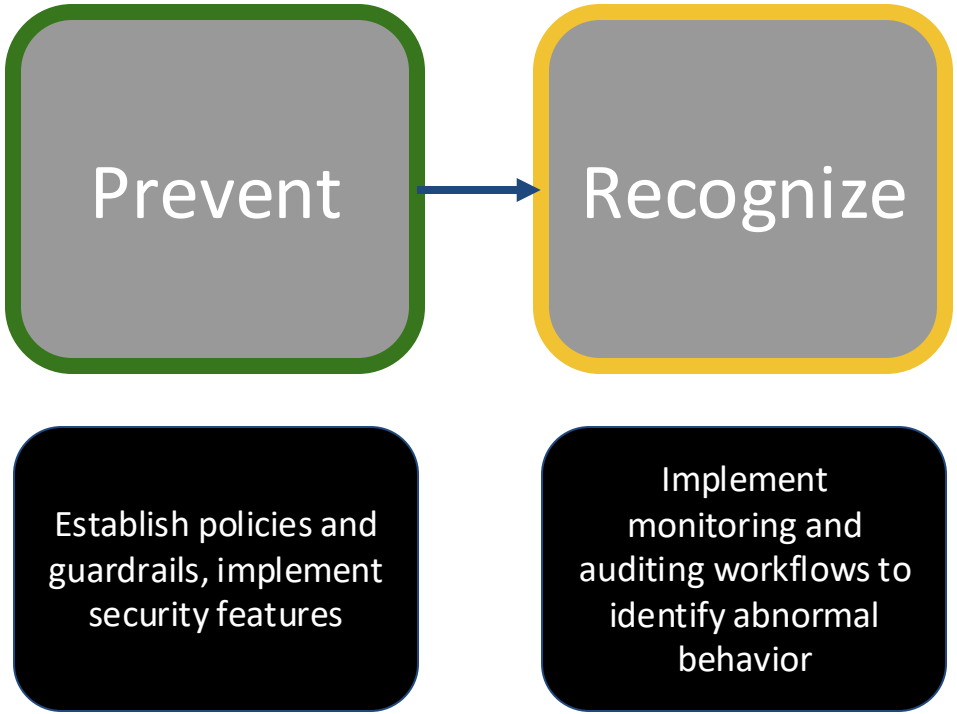
AWS IAM Access Analyzer

AWS Security Hub

AWS Config

- Identify security risks
- Identify compliance risks
- Track and prioritize all documented risks

Incident Response Steps



IR Recognition Strategies

Amazon CloudWatch

Amazon GuardDuty

AWS WAF

Flow logs

- Establish normal behavior baseline
- Configure appropriate alerts
- Detect abnormal behavior
- Detect intrusion attempts

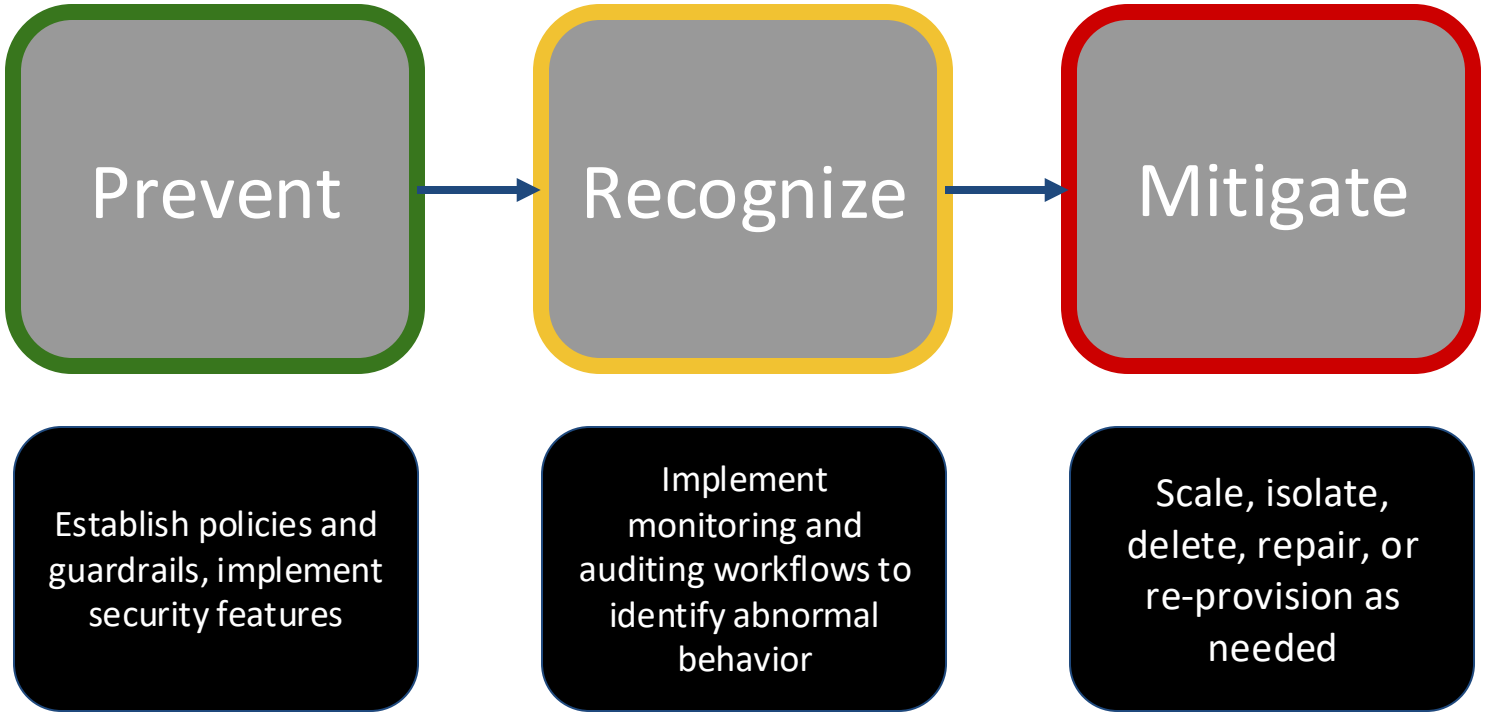
IR Recognition Strategies

AWS Chatbot

Amazon Simple Notification Service (Amazon SNS)

- Notify appropriately
- Use email or Slack integration

Incident Response Steps



IR Mitigation Strategies

Security group

AWS Identity and Access Management (IAM)

AWS Key Management Service (AWS KMS)

- Switch to restrictive Security group
- Revoke IAM role sessions or rotate access keys
- Rotate KMS backing keys

IR Mitigation Strategies

Amazon Simple Storage Service (Amazon S3)

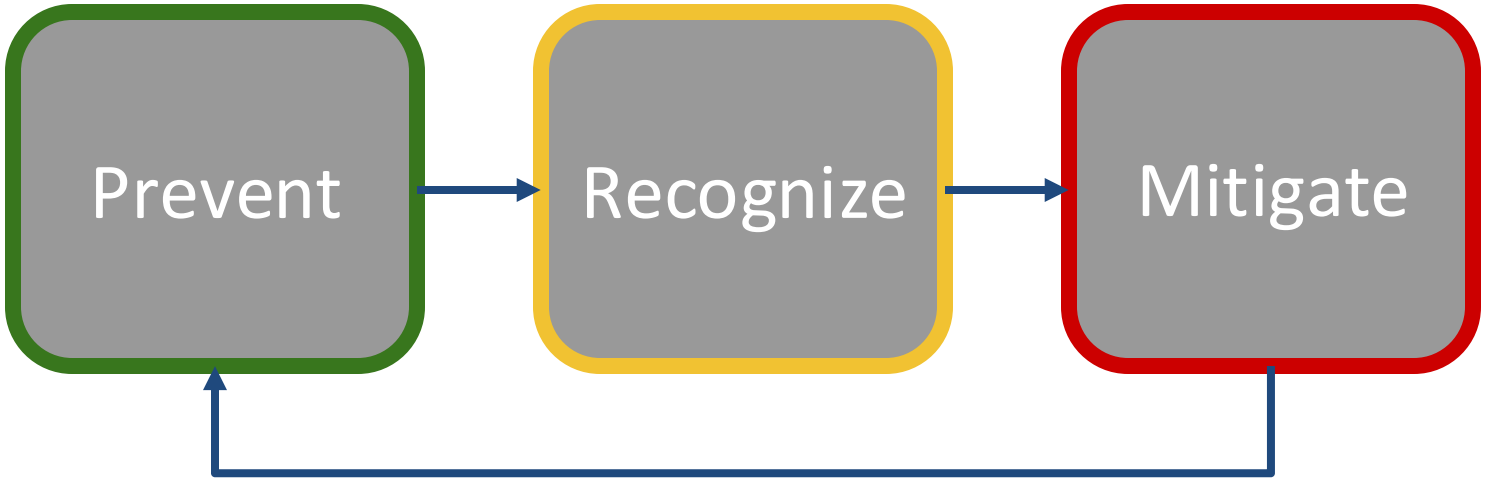
Snapshot

AMI

Logs

- Retain forensic evidence
- EBS snapshots
- EC2 AMIs
- CloudWatch logs

Incident Response Steps



These steps are not a spectrum, they are a complete cycle!

Question Breakdown

Prepare for security events

Question and Answer Choices

AWS has emailed an abuse notice to your root account identifying a compromised EC2 instance. Your security team wants to prevent the instance from performing any further activity while not affecting the running OS or other workloads.

Which of the following actions would meet the requirement?

A. Create a security group with no inbound or outbound rules. Replace the compromised instance's security group with the newly created one
B. Create a security group with no inbound or outbound rules. Attach the newly created security group to the compromised instance
C. Modify the subnet network ACL and block all outbound traffic
D. Modify the subnet network ACL and block all inbound and outbound traffic

Answer Choice Explanation

A lack of Security group rules can block all traffic to and from the instance, effectively isolating it from the rest of the VPC network.

A. Create a security group with no inbound or outbound rules. Replace the compromised instance's security group with the newly created one
B. Create a security group with no inbound or outbound rules. Attach the newly created security group to the compromised instance
C. Modify the subnet network ACL and block all outbound traffic
D. Modify the subnet network ACL and block all inbound and outbound traffic

Answer Choice Explanation

This sounds like a functional solution, but the effective Security group rules are the union of all rules, then applied by specificity. This task accomplishes nothing.

A. Create a security group with no inbound or outbound rules. Replace the compromised instance's security group with the newly created one
B. Create a security group with no inbound or outbound rules. Attach the newly created security group to the compromised instance
C. Modify the subnet network ACL and block all outbound traffic
D. Modify the subnet network ACL and block all inbound and outbound traffic

Answer Choice Explanation

This is a functional mitigation step. However, it will affect all other instances in the subnet, while still allowing the compromised instance to explore within the subnet.

A. Create a security group with no inbound or outbound rules. Replace the compromised instance's security group with the newly created one
B. Create a security group with no inbound or outbound rules. Attach the newly created security group to the compromised instance
C. Modify the subnet network ACL and block all outbound traffic
D. Modify the subnet network ACL and block all inbound and outbound traffic

Answer Choice Explanation

This solution is effectively identical to C in that it partially stops the instance from communicating, but also impacts other resources.

A. Create a security group with no inbound or outbound rules. Replace the compromised instance's security group with the newly created one
B. Create a security group with no inbound or outbound rules. Attach the newly created security group to the compromised instance
C. Modify the subnet network ACL and block all outbound traffic
D. Modify the subnet network ACL and block all inbound and outbound traffic

Correct Answer

Correct Answer: A

A. Create a security group with no inbound or outbound rules. Replace the compromised instance's security group with the newly created one
B. Create a security group with no inbound or outbound rules. Attach the newly created security group to the compromised instance
C. Modify the subnet network ACL and block all outbound traffic
D. Modify the subnet network ACL and block all inbound and outbound traffic

Question Domain 2: Design Resilient Architectures

Question Domain 2: Design Resilient Architectures

- Automatically recover from failure

Principle Definition

By monitoring a workload for key performance indicators (KPIs), you can trigger automation when a threshold is breached.

These KPIs should be a measure of business value, not of the technical aspects of the operation of the service.

This allows for automatic notification and tracking of failures, and for automated recovery processes that work around or repair the failure.

With more sophisticated automation, it's possible to anticipate and remediate failures before they occur.

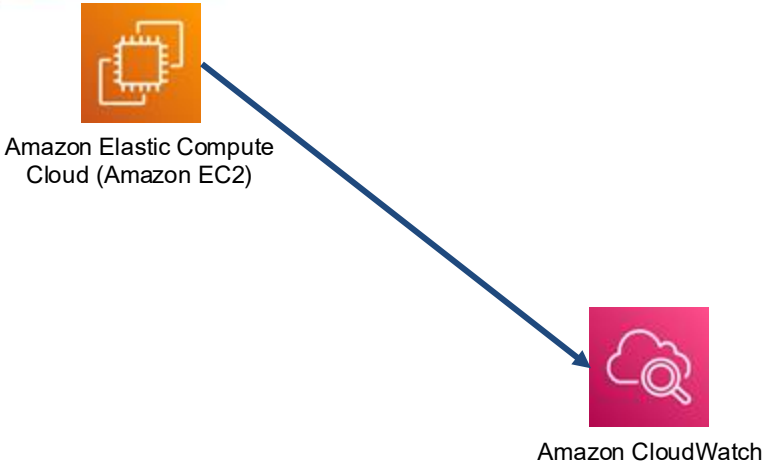
Four Phases of Monitoring

Generate

Amazon CloudWatch

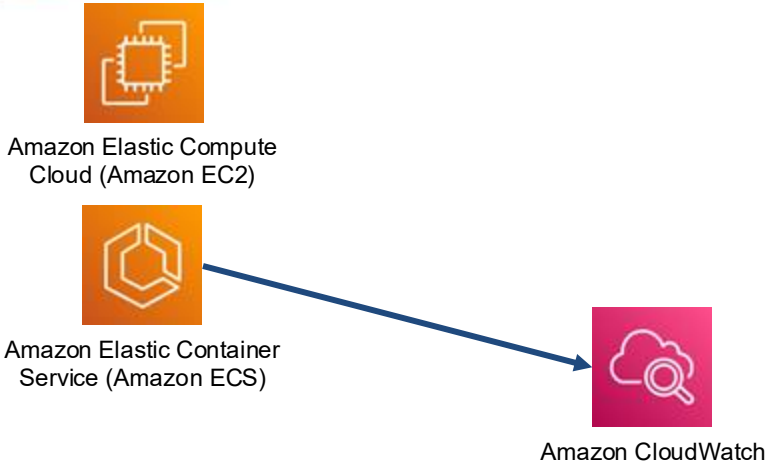
AWS Personal Health Dashboard

Metric Generation



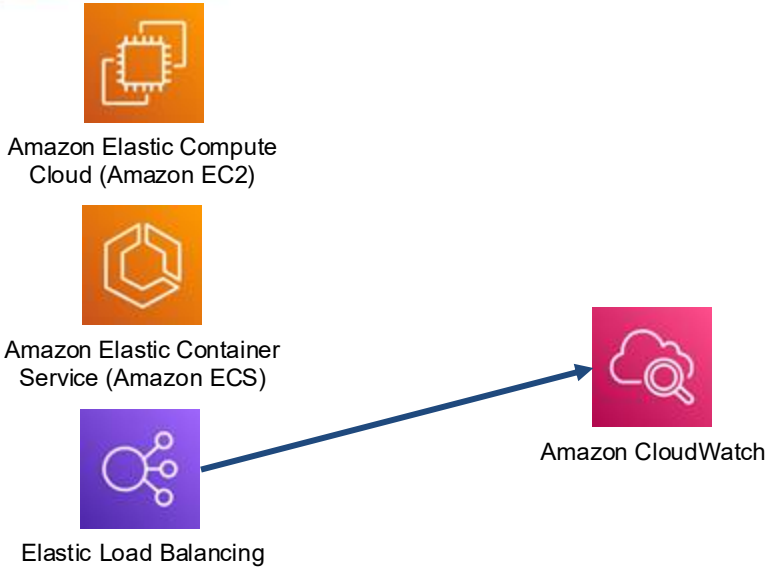
Push default and custom metrics from EC2 instances

Metric Generation



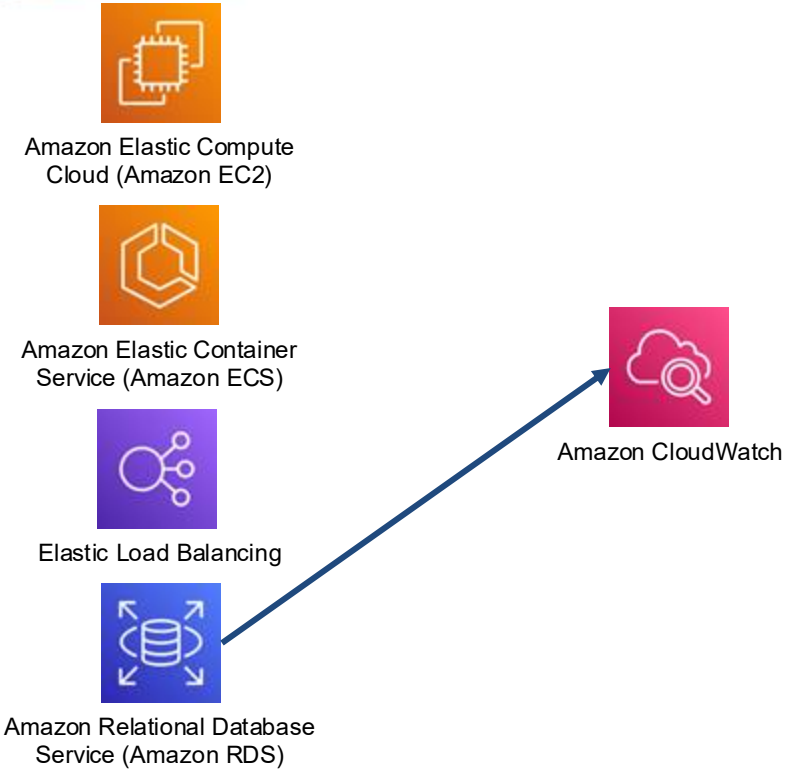
Push default and custom metrics from ECS containers

Metric Generation



Push default metrics from various ELB types

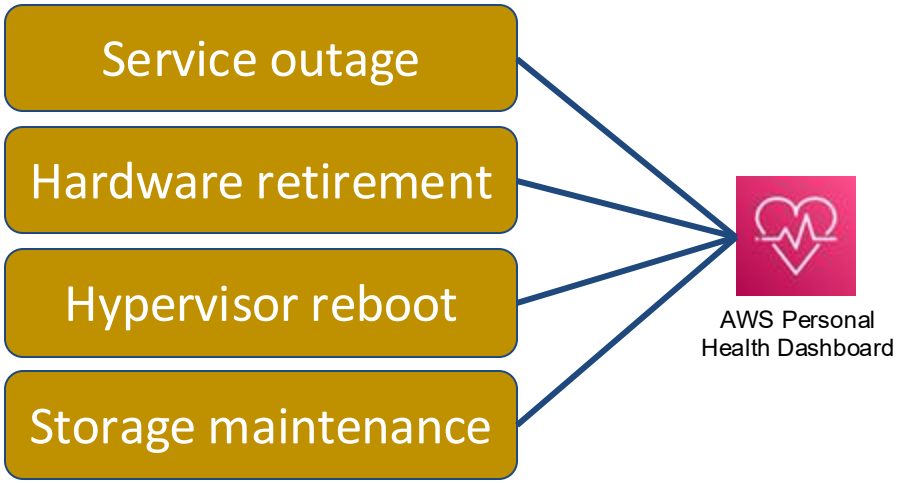
Metric Generation



Push default metrics from various RDS engines

Metric Generation

View performance and availability of AWS services underlying your resources



Four Phases of Monitoring

Generate

Amazon CloudWatch

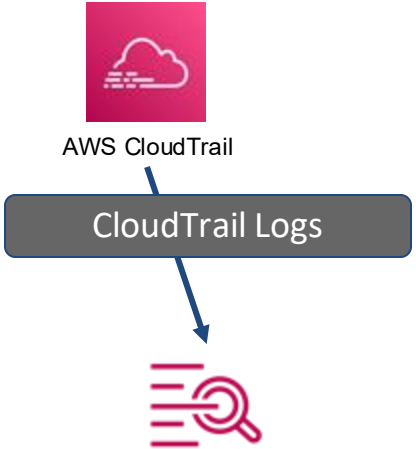
AWS Personal Health Dashboard

Aggregate

Amazon CloudWatch

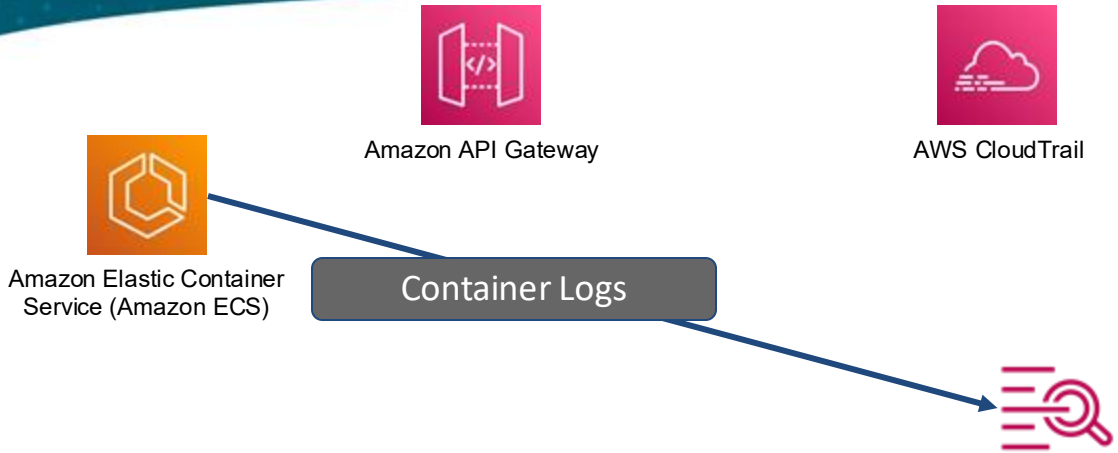
Amazon Simple Storage Service (Amazon S3)

Log Aggregation



Log Aggregation

Log Aggregation



Log Aggregation

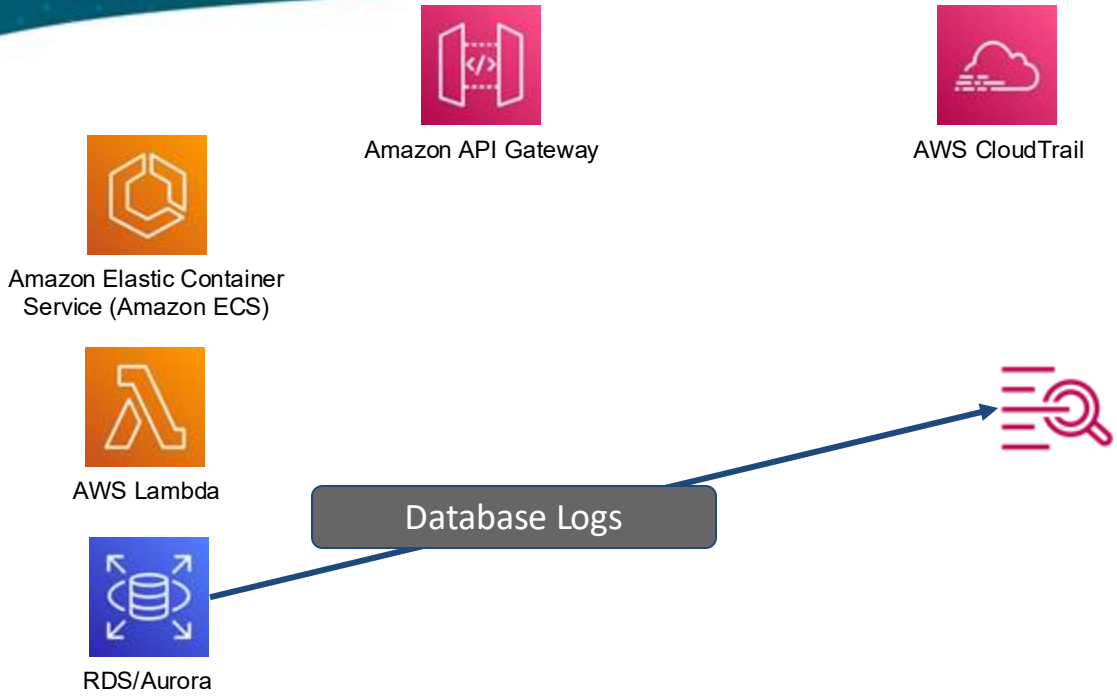
Amazon API Gateway

AWS CloudTrail

Amazon Elastic Container Service (Amazon ECS)

AWS Lambda

Log Aggregation



Log Aggregation

Amazon API Gateway

AWS CloudTrail

Amazon Elastic Container Service (Amazon ECS)

AWS Lambda

RDS/Aurora

Run-Command Output

AWS Systems Manager

Pearson

Log Aggregation

Amazon API Gateway

AWS CloudTrail

Amazon Elastic Container Service (Amazon ECS)

AWS Lambda

RDS/Aurora

AWS Systems Manager

App/OS Logs

Amazon Elastic Compute Cloud (Amazon EC2)

Four Phases of Monitoring

Generate

Amazon CloudWatch

AWS Personal Health Dashboard

Aggregate

Amazon CloudWatch

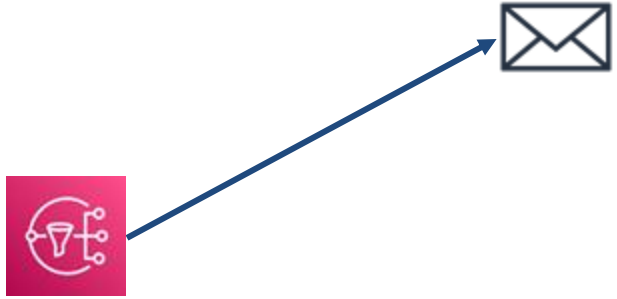
Amazon Simple Storage Service (Amazon S3)

Process

Amazon Simple Notification Service (Amazon SNS)

AWS Auto Scaling

Real-time processing - Send Notifications

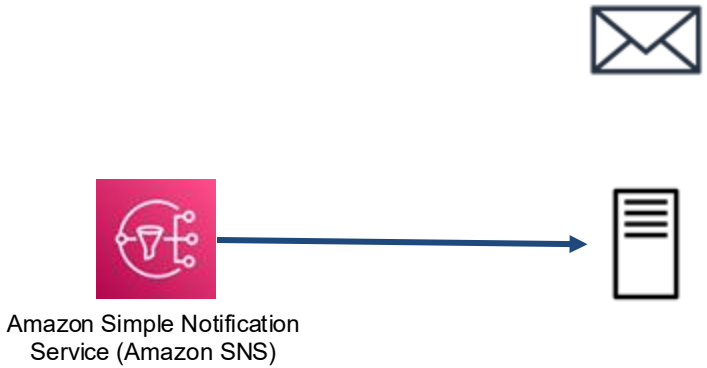


Amazon Simple Notification Service (Amazon SNS)

Use SNS to send email to a distribution list

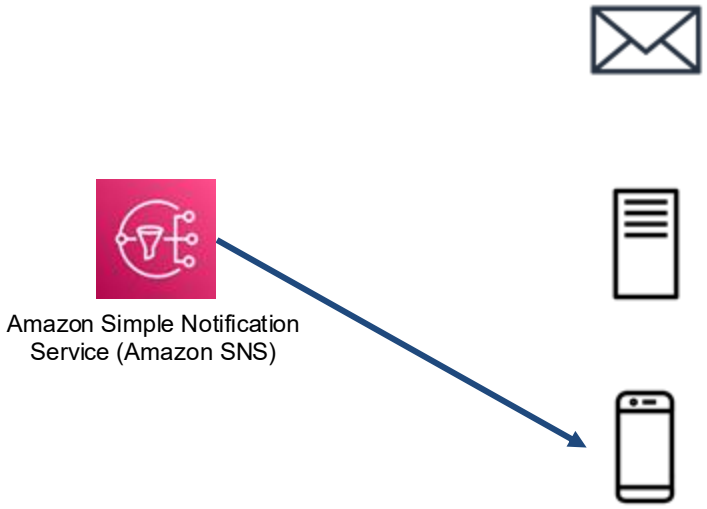
Real-time processing - Send Notifications

Use SNS to create a Jira ticket using a custom endpoint

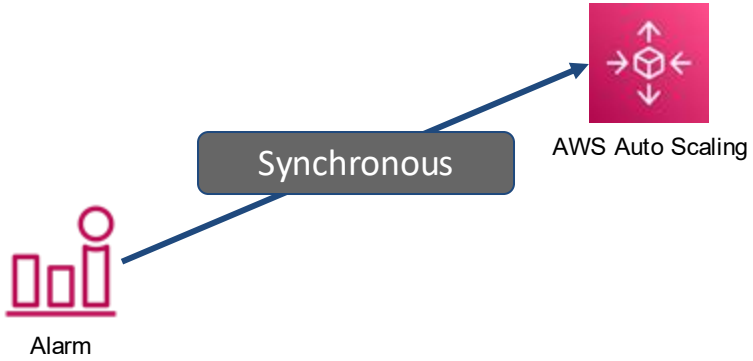


Real-time processing - Send Notifications

Use SNS to send text messages to the on-call team



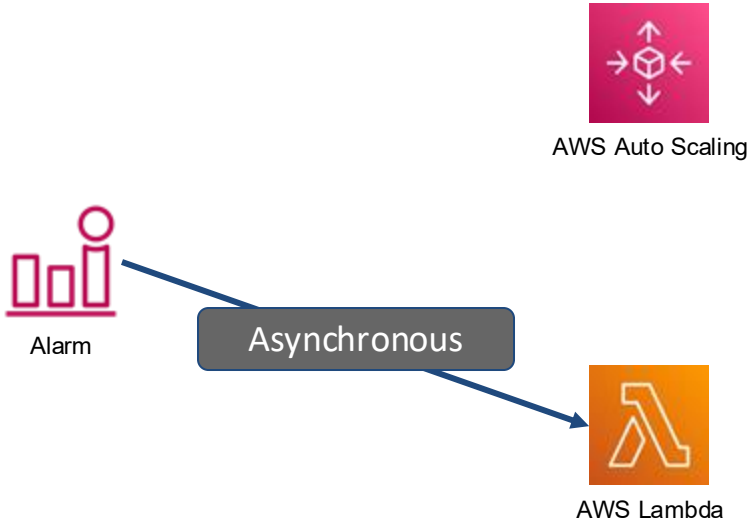
Real-time processing - Automate Responses



Automatically scale based on CloudWatch Alarms

Real-time processing - Automate Responses

Trigger other actions via Lambda function



Four Phases of Monitoring

Generate

Amazon CloudWatch

AWS Personal Health Dashboard

Aggregate

Amazon CloudWatch

Amazon Simple Storage Service (Amazon S3)

Process

Amazon Simple Notification Service (Amazon SNS)

AWS Auto Scaling

Store and Analyze

Logs

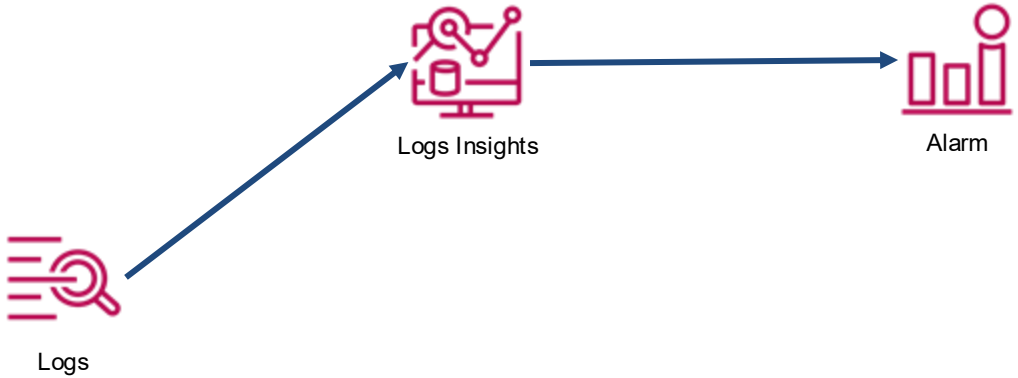
Amazon Simple Storage Service (Amazon S3)

Storage and Analytics

Retain logs directly with TTL for expiration

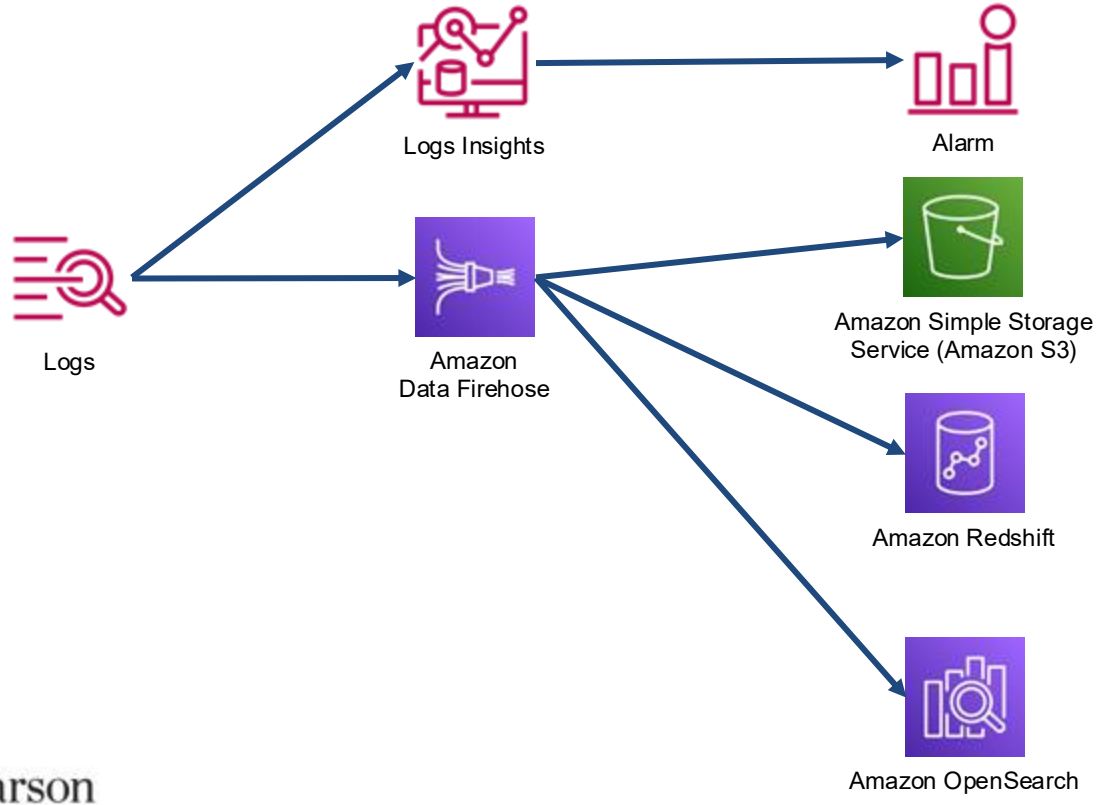
Logs

Storage and Analytics



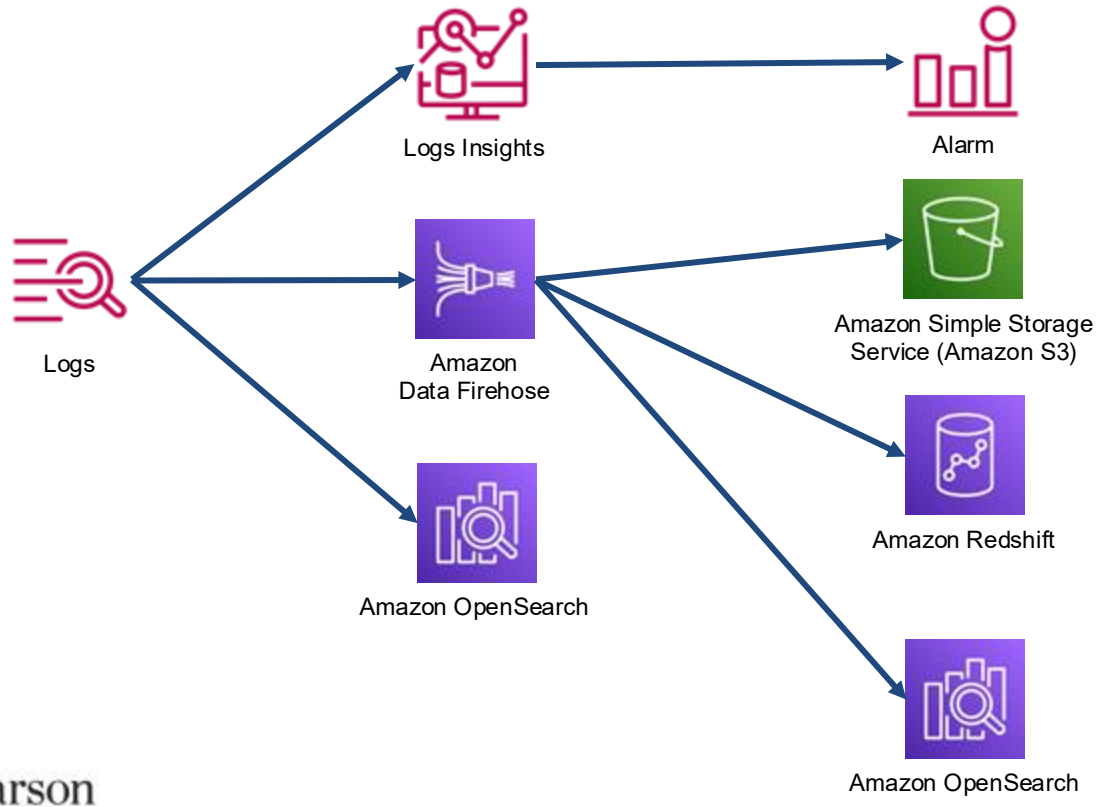
Analyze with CloudWatch Logs Insights

Storage and Analytics



Deliver optionally transformed logs to other services

Storage and Analytics



Deliver directly to OpenSearch for analysis

Question Breakdown

Automatically recover from failure

Question and Answer Choices

You've been asked to design the monitoring of an internal application which runs on an EC2 instance in a private subnet, listening on port 8080. This monitoring will be used to calculate availability of the application and be used for business continuity purposes as well as automated failover.

Which of the following recommendations will meet the requirement in the most reliable manner? (pick two)

A. Write a script to run via cron every minute which checks the listener port and uploads as a custom CloudWatch metric
B. Configure the CloudWatch agent to push a metric which tests the listener
C. Provision a CloudWatch Synthetics canary to test the listener port
D. Configure a CloudWatch alarm from the appropriate metric. Configure a Route 53 Health Check with the alarm as the source
E. Configure a Route 53 Health Check with the CloudWatch metric as the source

Answer Choice Explanation

This is a plausible choice for delivering the listener uptime metrics to CloudWatch. There is some question whether a shell script via cron is the most reliable method for delivery.

A. **Write a script to run via cron every minute which checks the listener port and uploads as a custom CloudWatch metric**
B. **Configure the CloudWatch agent to push a metric which tests the listener**
C. **Provision a CloudWatch Synthetics canary to test the listener port**
D. **Configure a CloudWatch alarm from the appropriate metric. Configure a Route 53 Health Check with the alarm as the source**
E. **Configure a Route 53 Health Check with the CloudWatch metric as the source**

Answer Choice Explanation

CloudWatch agent configuration does not support custom metrics of this type.

A. Write a script to run via cron every minute which checks the listener port and uploads as a custom CloudWatch metric
B. Configure the CloudWatch agent to push a metric which tests the listener
C. Provision a CloudWatch Synthetics canary to test the listener port
D. Configure a CloudWatch alarm from the appropriate metric. Configure a Route 53 Health Check with the alarm as the source
E. Configure a Route 53 Health Check with the CloudWatch metric as the source

Answer Choice Explanation

This is also a functional choice for delivering uptime metrics, and is likely more reliable than the cron job option described in A.

A. Write a script to run via cron every minute which checks the listener port and uploads as a custom CloudWatch metric
B. Configure the CloudWatch agent to push a metric which tests the listener
C. Provision a CloudWatch Synthetics canary to test the listener port
D. Configure a CloudWatch alarm from the appropriate metric. Configure a Route 53 Health Check with the alarm as the source
E. Configure a Route 53 Health Check with the CloudWatch metric as the source

Answer Choice Explanation

This assumes the metric has been delivered to CloudWatch. Once an alarm has been created, it can be used as a source for a Route 53 Health Check.

A. Write a script to run via cron every minute which checks the listener port and uploads as a custom CloudWatch metric
B. Configure the CloudWatch agent to push a metric which tests the listener
C. Provision a CloudWatch Synthetics canary to test the listener port
D. Configure a CloudWatch alarm from the appropriate metric. Configure a Route 53 Health Check with the alarm as the source
E. Configure a Route 53 Health Check with the CloudWatch metric as the source

Answer Choice Explanation

Route 53 Health Checks can use CloudWatch alarms as a source, but not a CloudWatch metric directly.

A. Write a script to run via cron every minute which checks the listener port and uploads as a custom CloudWatch metric
B. Configure the CloudWatch agent to push a metric which tests the listener
C. Provision a CloudWatch Synthetics canary to test the listener port
D. Configure a CloudWatch alarm from the appropriate metric. Configure a Route 53 Health Check with the alarm as the source
E. Configure a Route 53 Health Check with the CloudWatch metric as the source

Correct Answer

Correct Answers: C,D

A. Write a script to run via cron every minute which checks the listener port and uploads as a custom CloudWatch metric
B. Configure the CloudWatch agent to push a metric which tests the listener
C. Provision a CloudWatch Synthetics canary to test the listener port
D. Configure a CloudWatch alarm from the appropriate metric. Configure a Route 53 Health Check with the alarm as the source
E. Configure a Route 53 Health Check with the CloudWatch metric as the source

Question Domain 2: Design Resilient Architectures

Test recovery procedures

Principle Definition

In an on-premises environment, testing is often conducted to prove that the workload works in a particular scenario.

Testing is not typically used to validate recovery strategies. In the cloud, you can test how your workload fails, and you can validate your recovery procedures.

You can use automation to simulate different failures or to recreate scenarios that led to failures before.

This approach exposes failure pathways that you can test and fix before a real failure scenario occurs, thus reducing risk.

Business Continuity Definitions

Business Continuity - How to keep the business functional in some capacity during/after a critical incident.

Disaster Recovery - How to save and recover data and other business processes during/after a critical incident. Contributes to the Business Continuity Plan.

Business Continuity Definitions

RTO - Recovery Time Objective. The maximum acceptable delay between the interruption of service and restoration of service.

RPO - Recovery Point Objective. The maximum acceptable amount of time since the last data recovery point.

Reducing either of these metrics usually results in higher cost to implement!

Disaster Recovery Scenario

A global financial services company requires a DR strategy for its critical application to ensure business continuity. The application uses an ALB and an EC2 Auto Scaling Group for web and app, RDS and DynamoDB for the database tier, and S3 for storing sensitive financial documents.

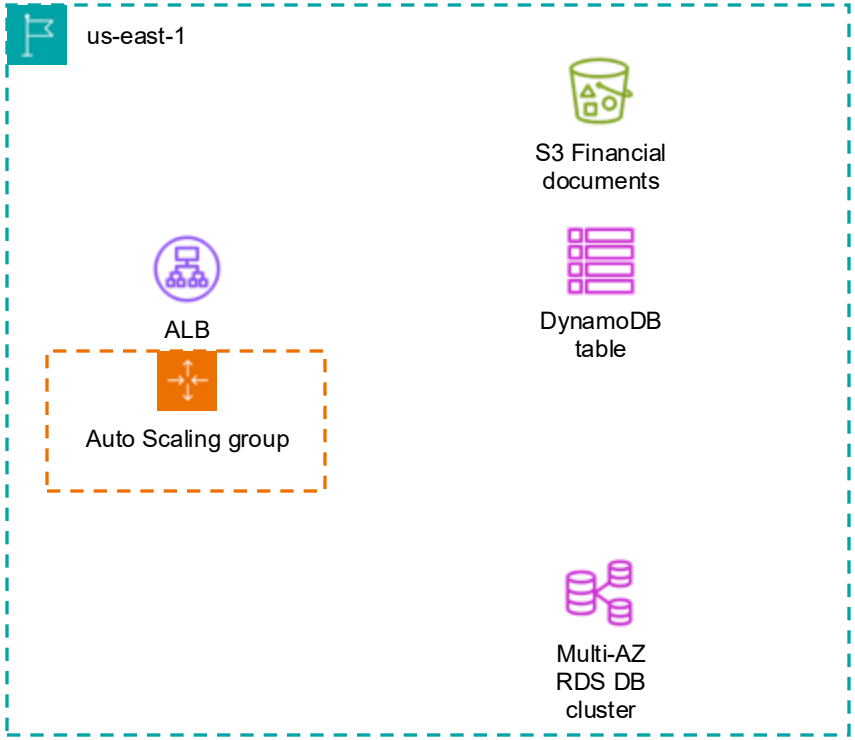
The company's primary AWS Region is us-east-1, with us-west-2 as the recovery Region.

The company aims for a Recovery Time Objective (RTO) of 1 hour and a Recovery Point Objective (RPO) of 15 minutes. What infrastructure configuration can meet these requirements?

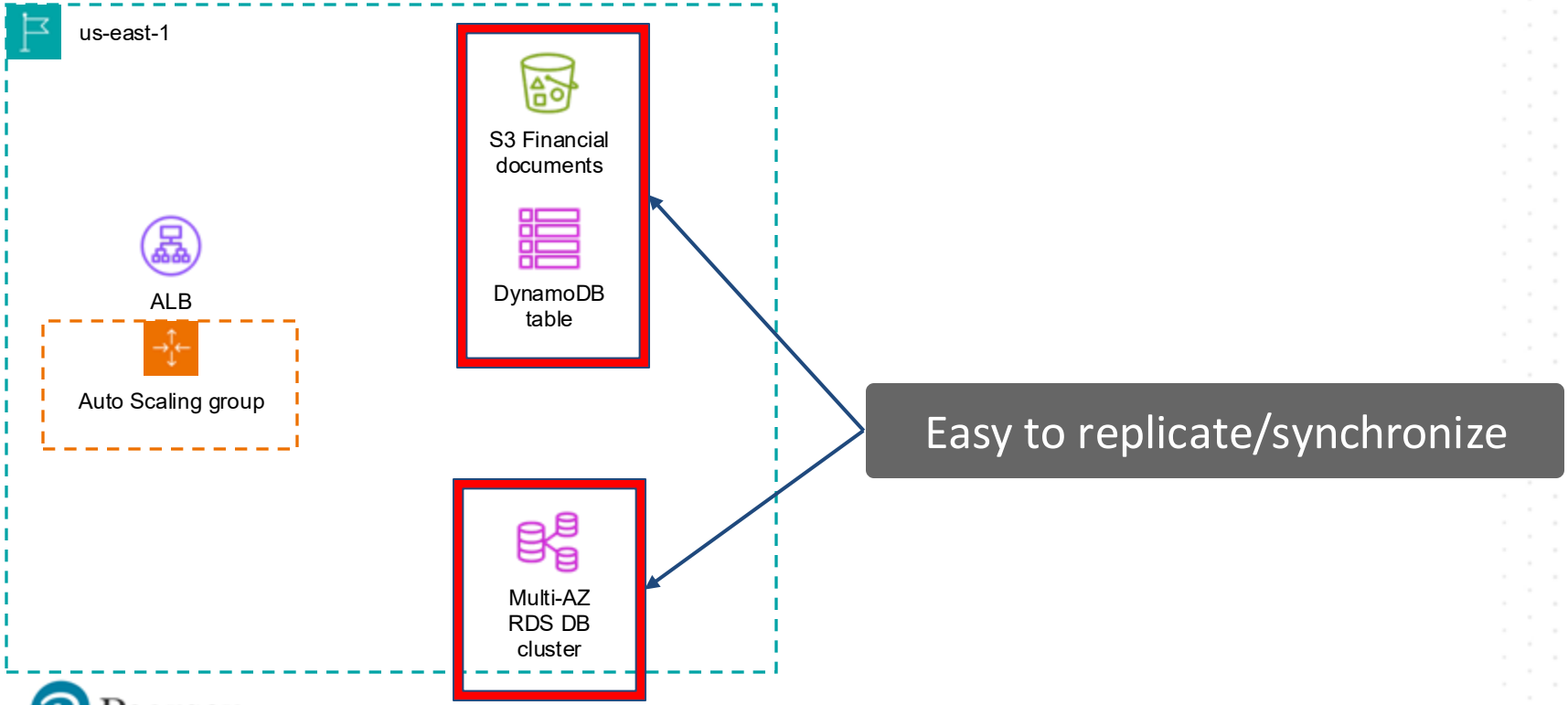
Disaster Recovery Considerations

- Replicate, synchronize, or backup/restore?
- Which resources don't support replication?
- What about DNS cutover?

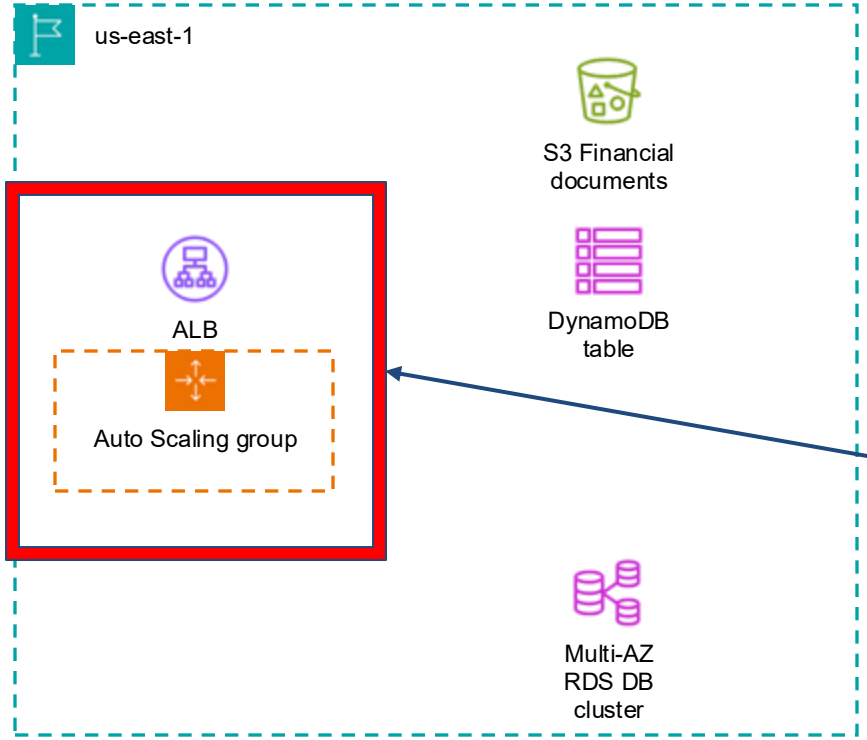
Primary Region Infrastructure



Primary Region Infrastructure

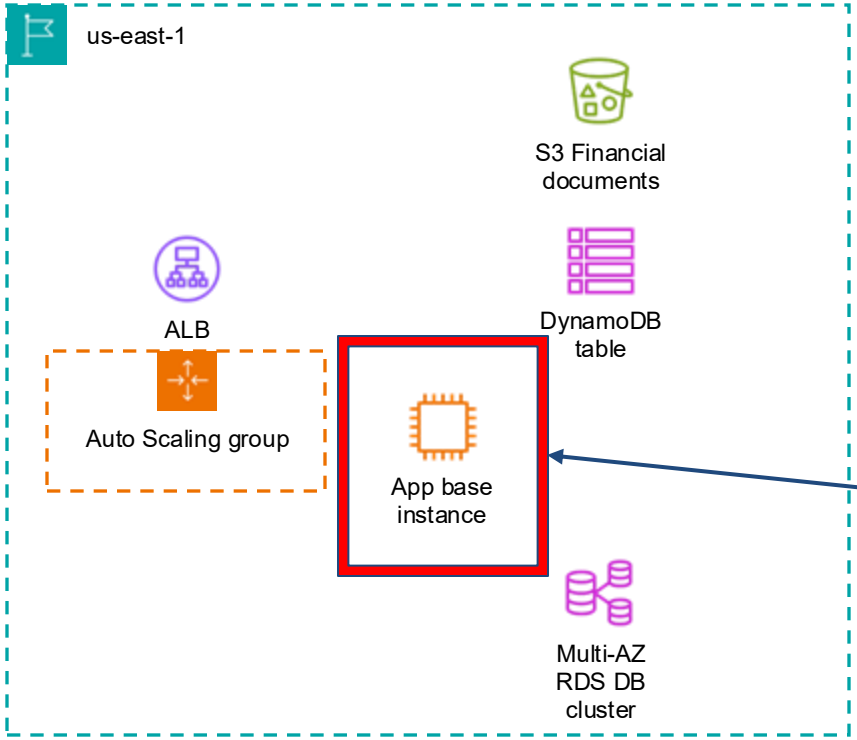


Primary Region Infrastructure



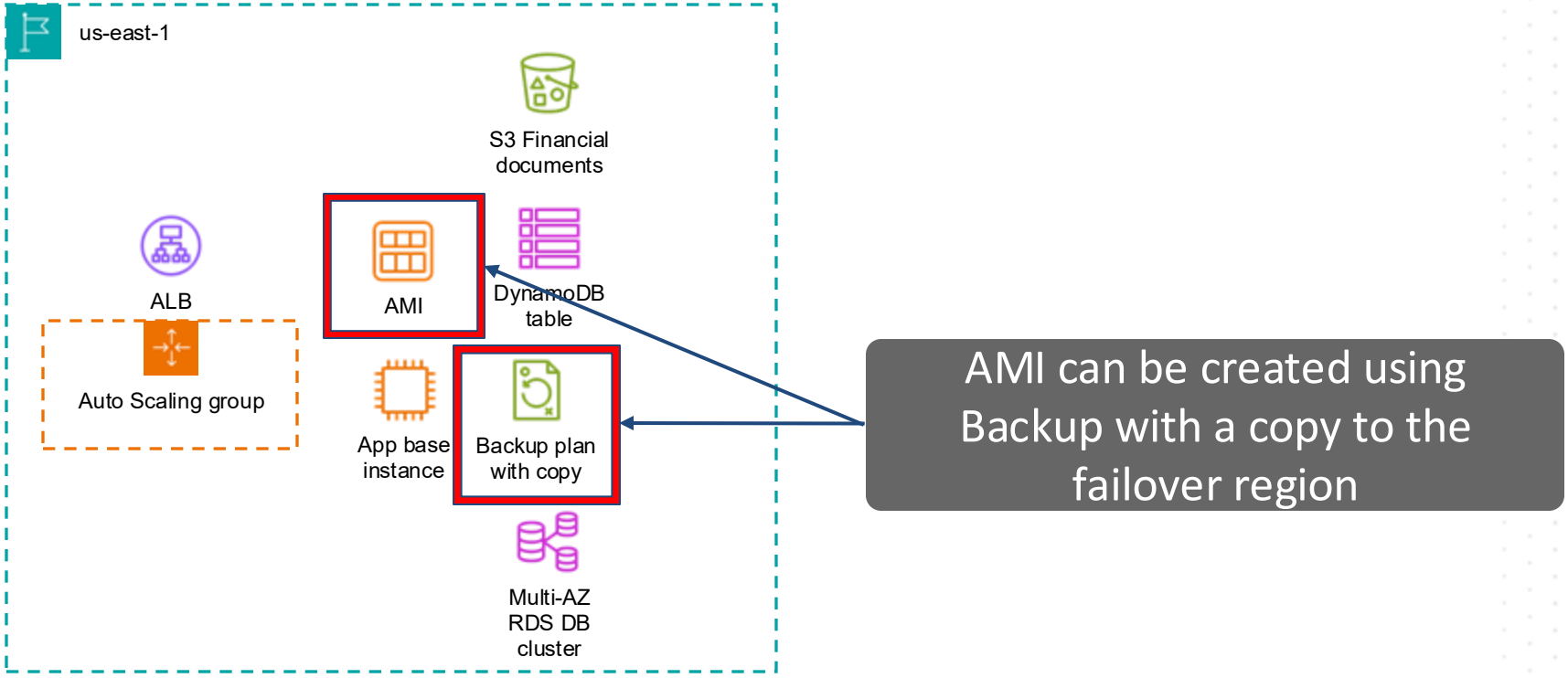
IaC can assist with these

Replication Infrastructure

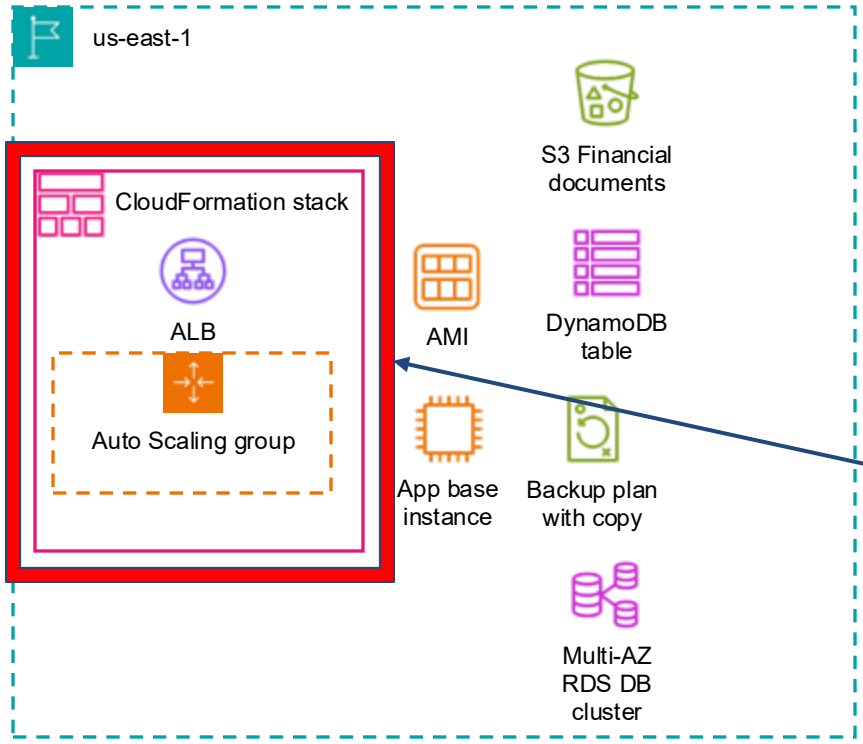


ASG requires an AMI, created from a base instance

Replication Infrastructure

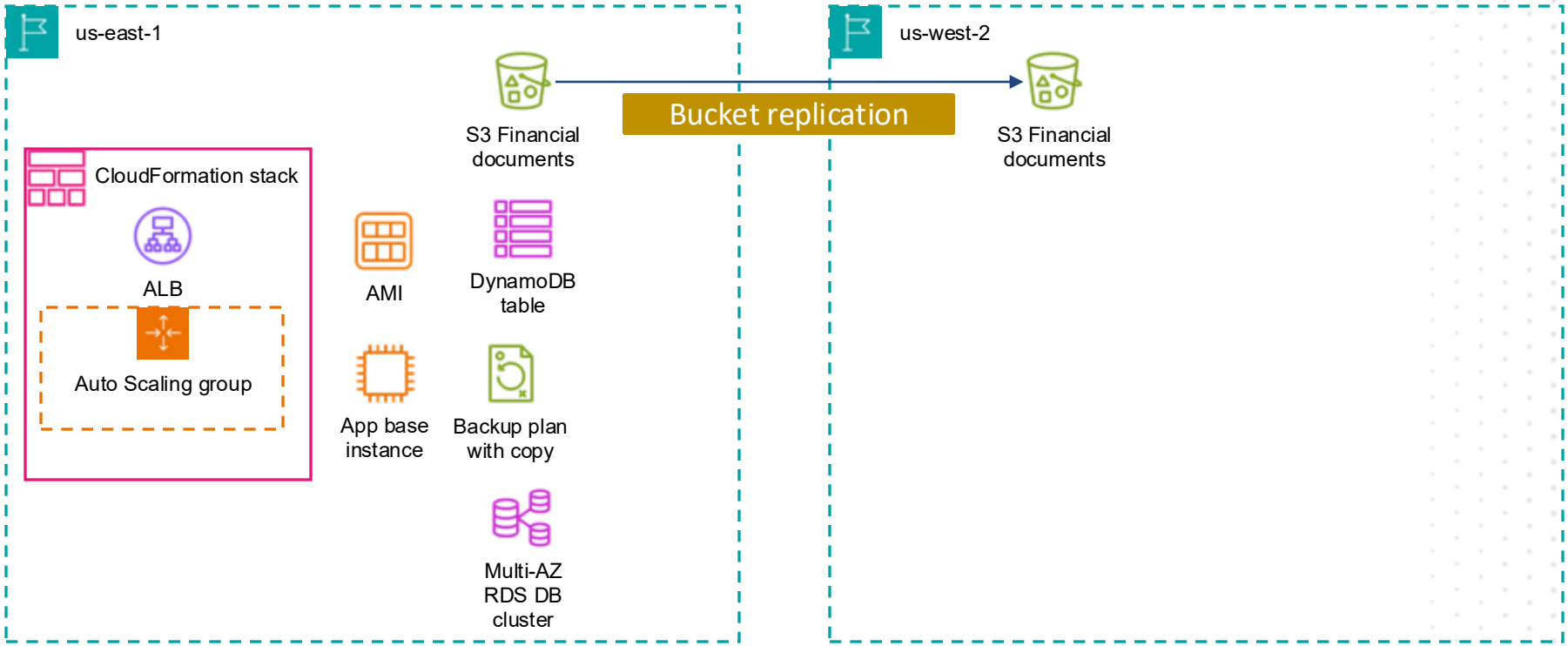


Replication Infrastructure

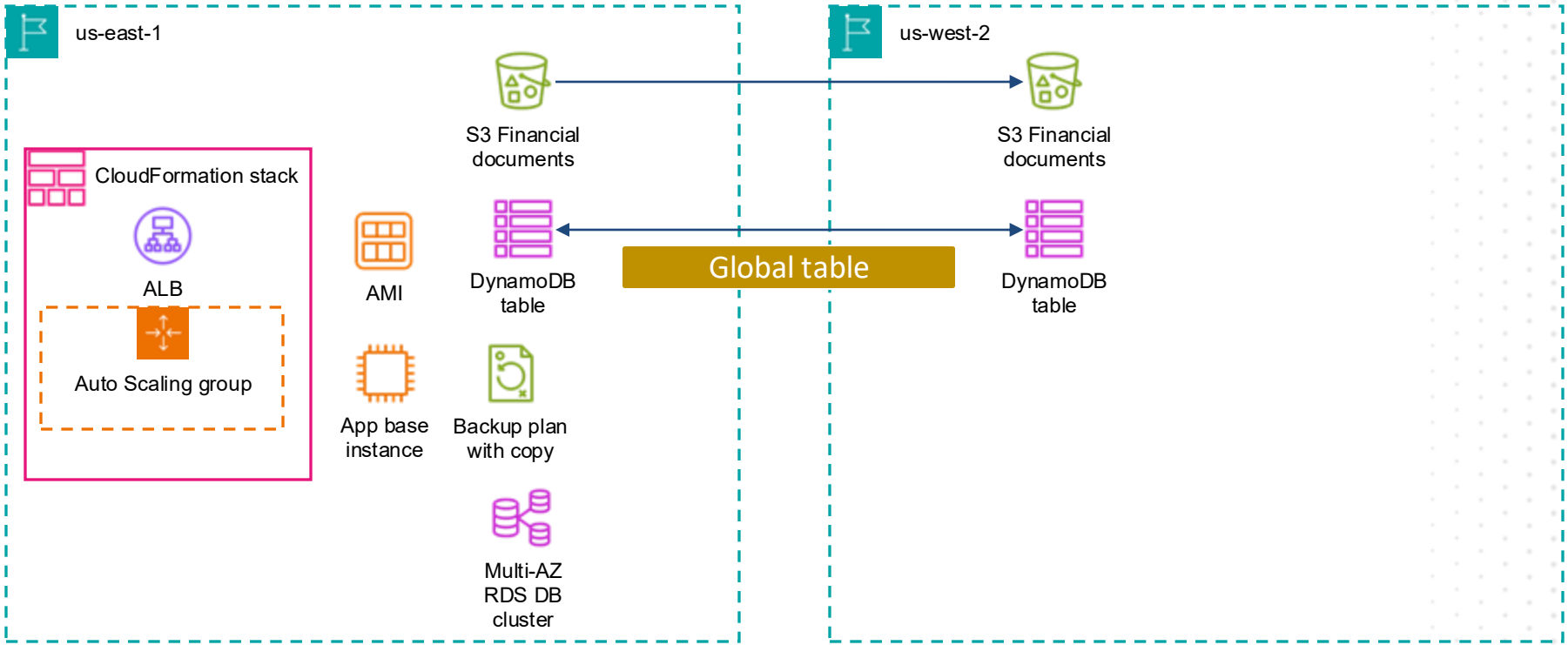


The ALB/ASG/EC2 infrastructure can be deployed using CloudFormation

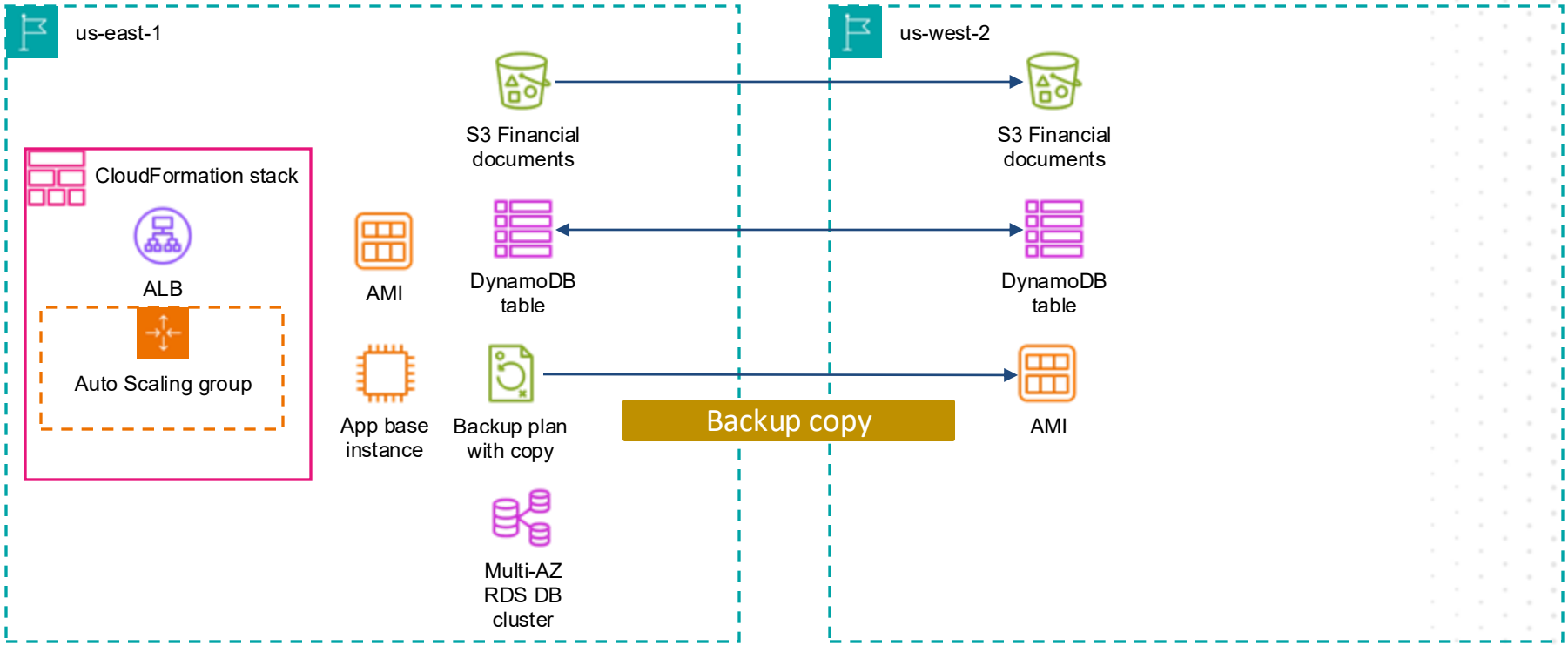
Failover Region Infrastructure



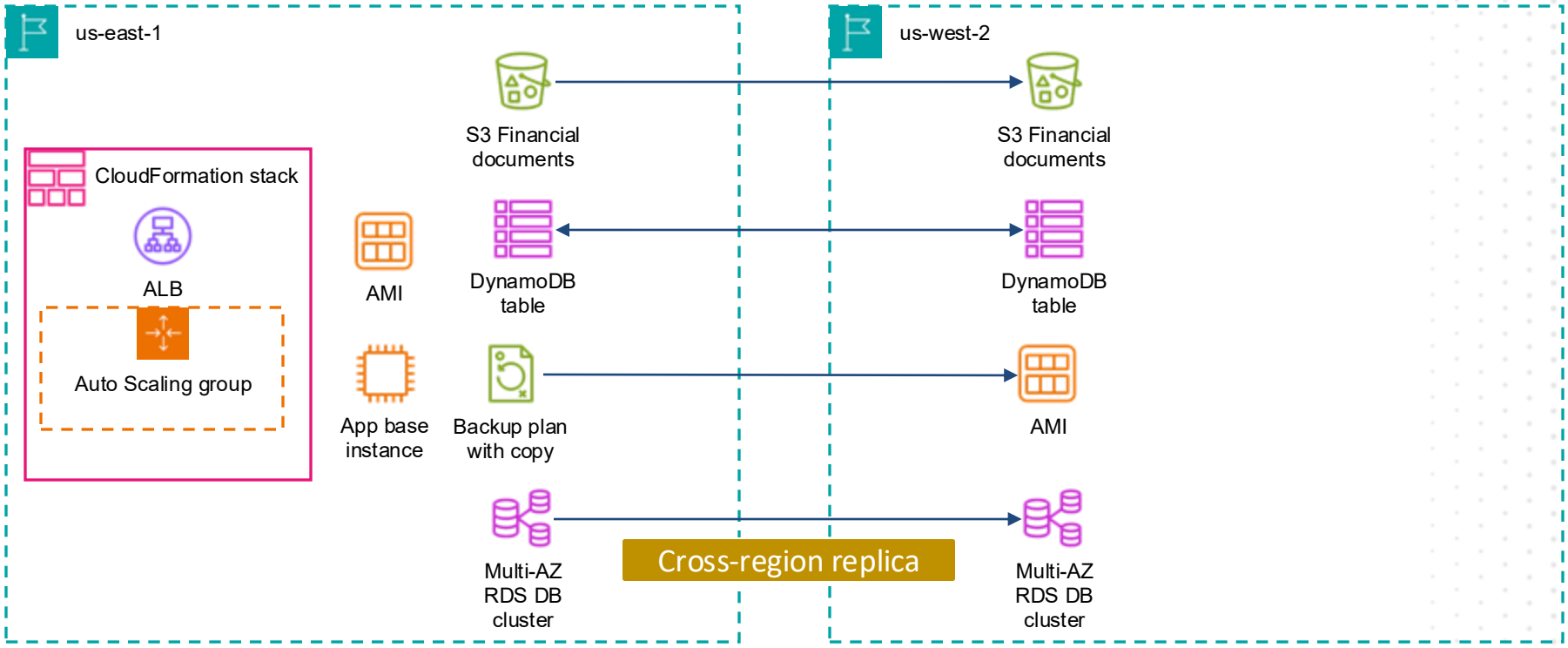
Failover Region Infrastructure



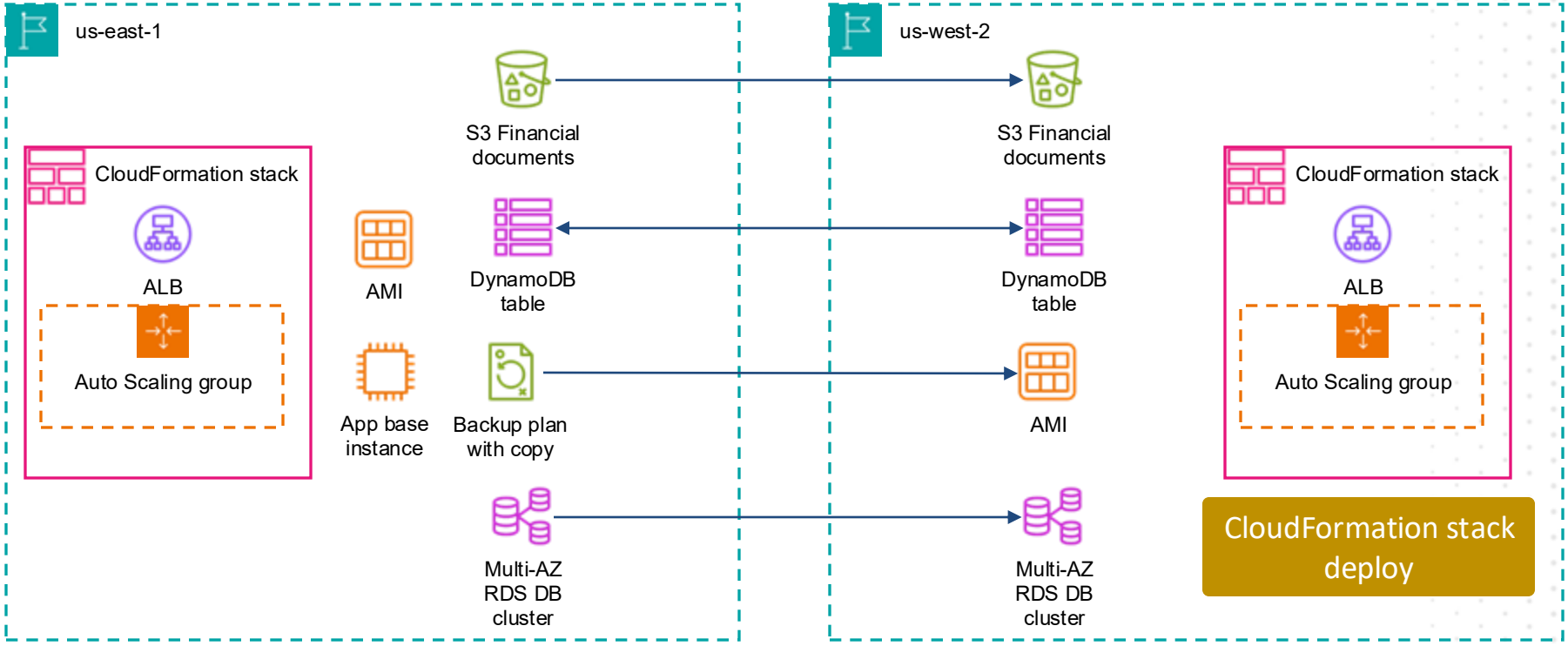
Failover Region Infrastructure

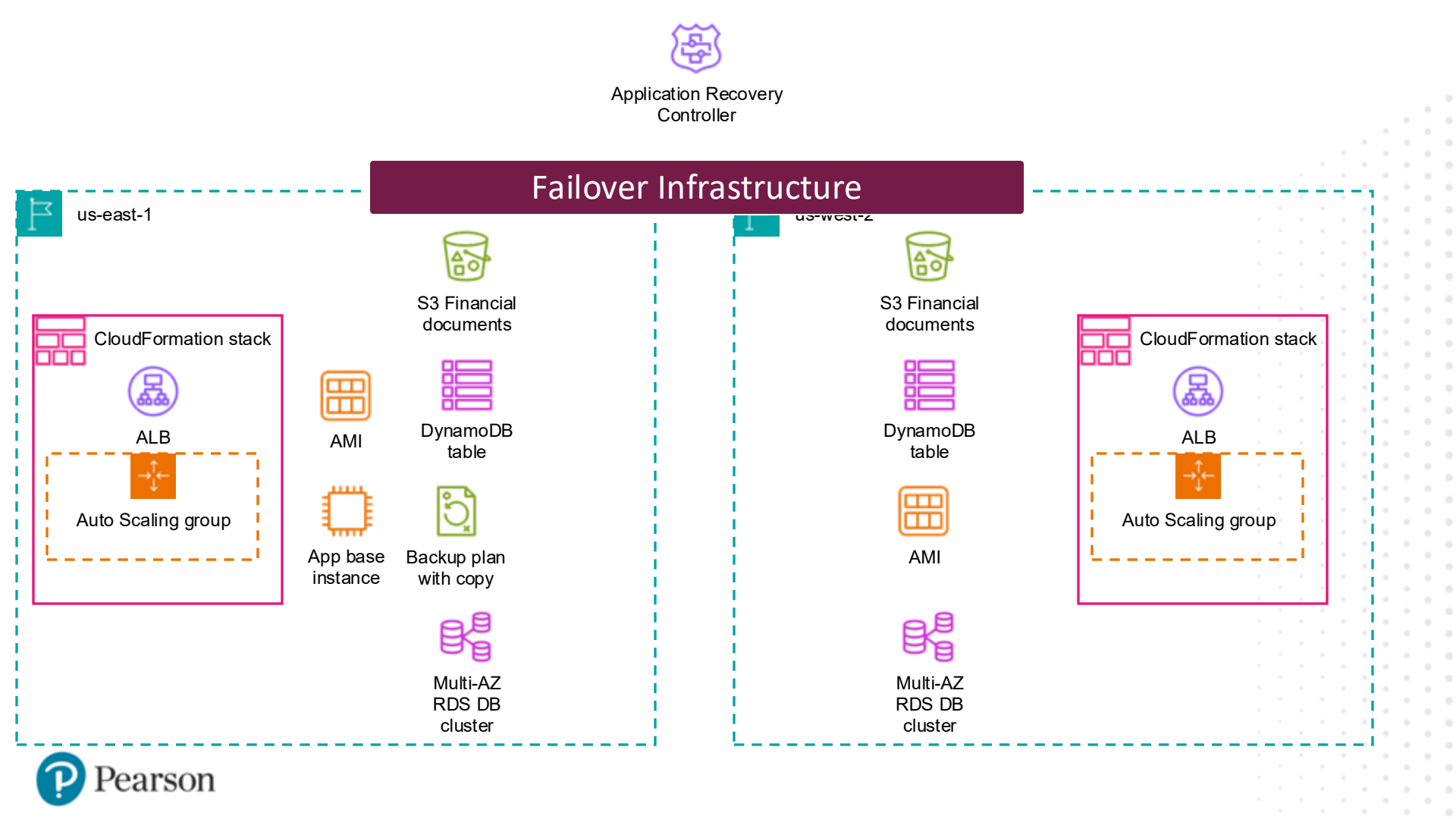


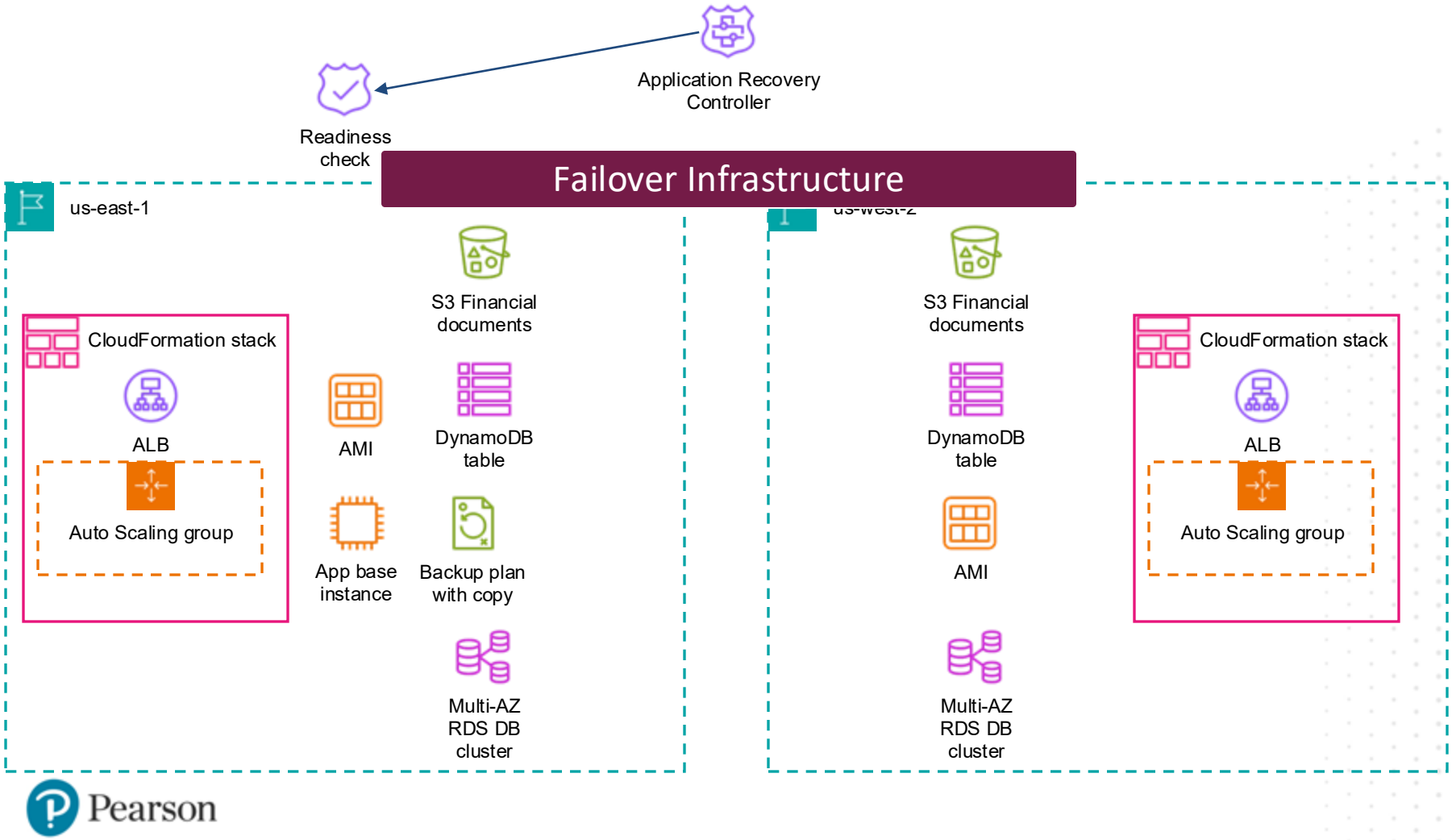
Failover Region Infrastructure

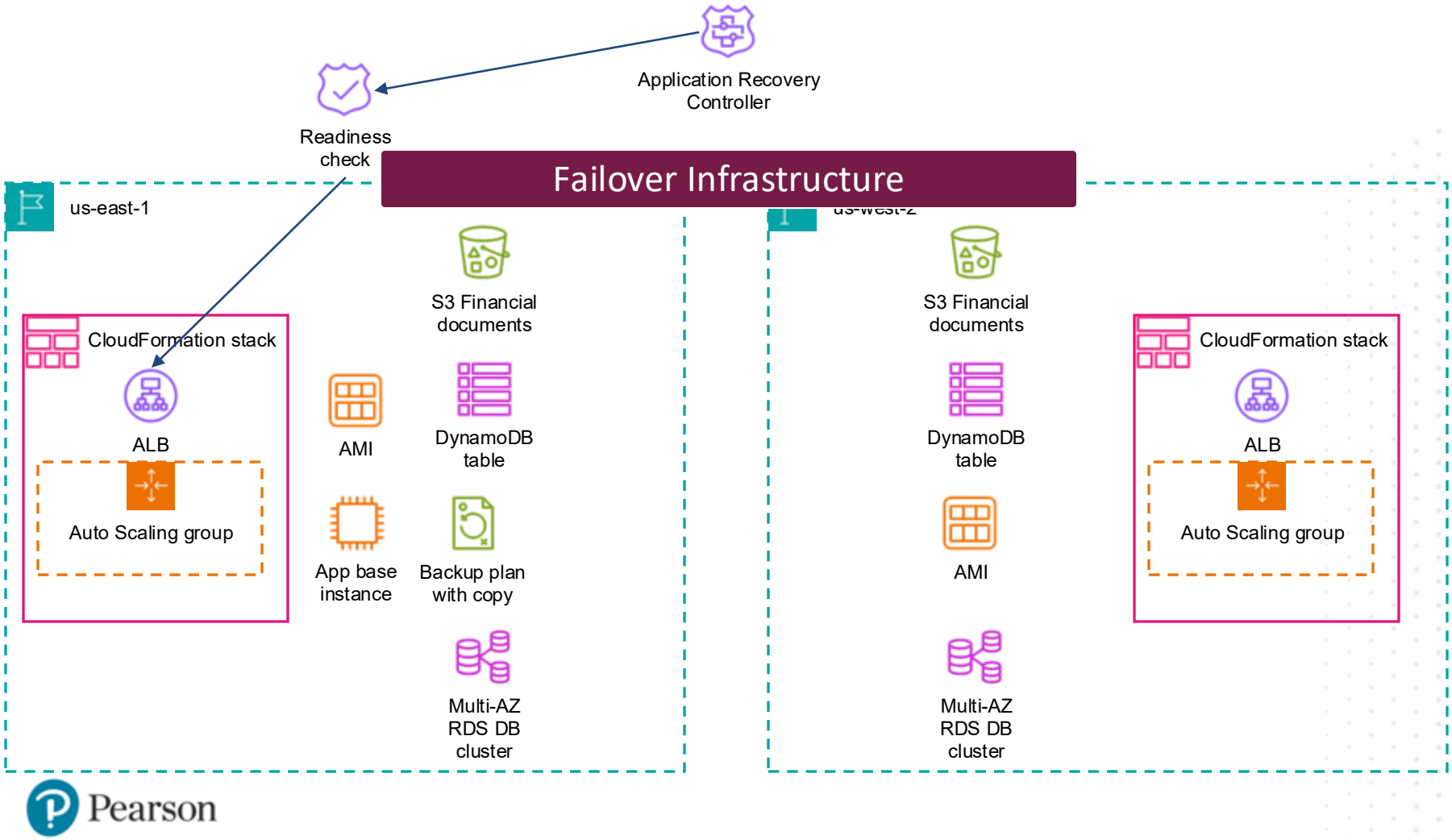


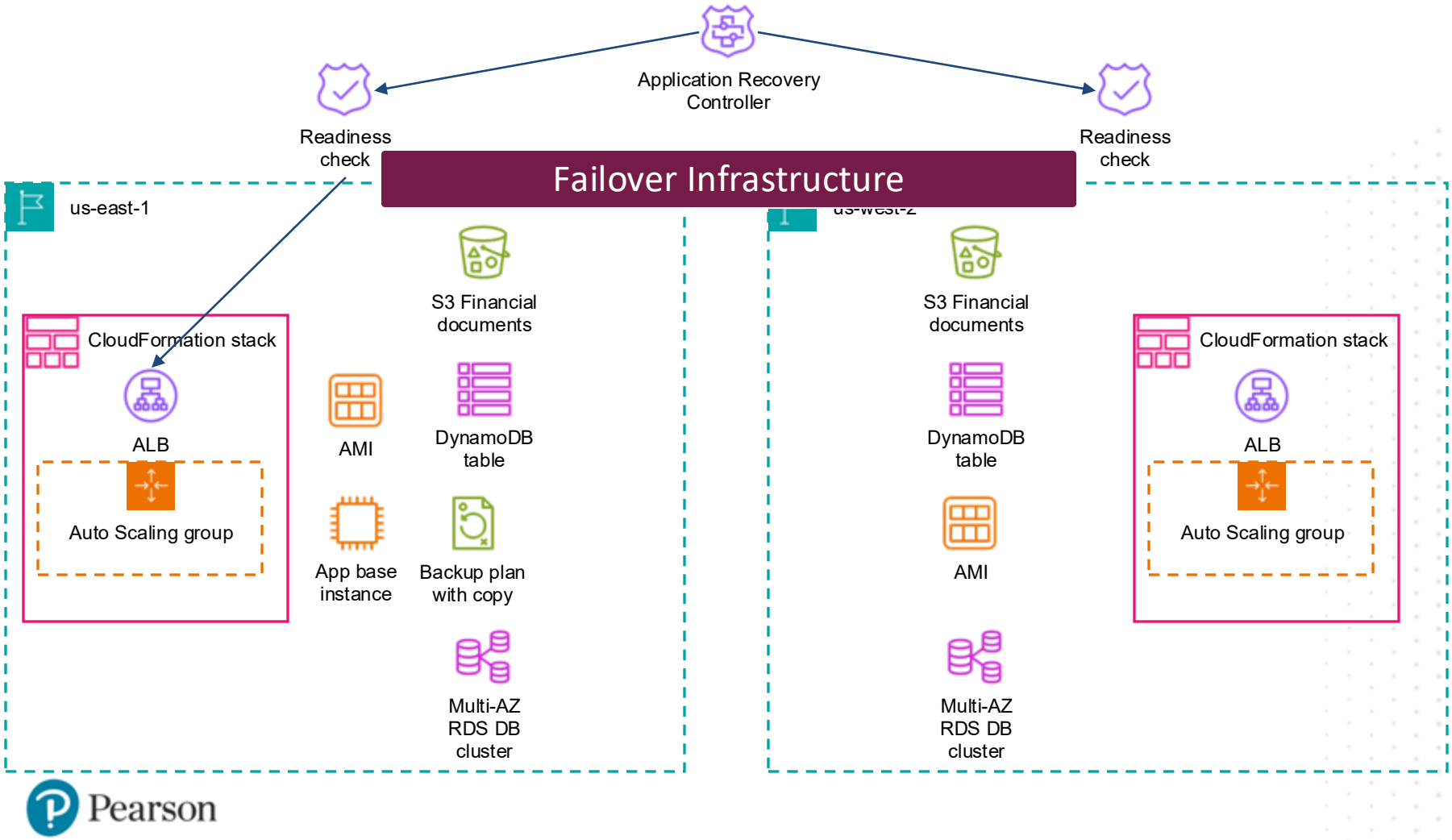
Failover Region Infrastructure

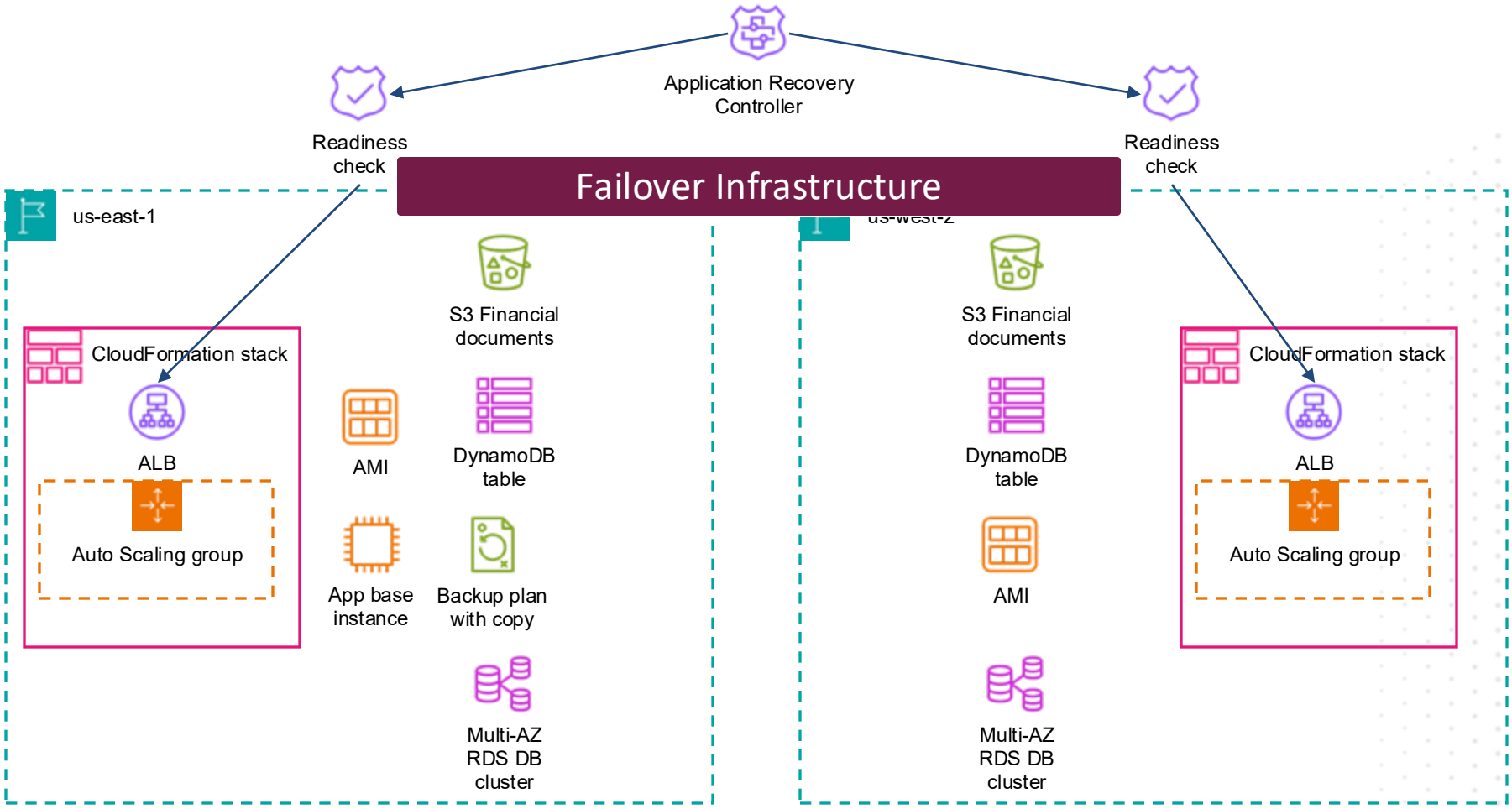



Application Recovery Controller
Failover Infrastructure
us-east-1
us-west-2
S3 Financial documents
CloudFormation stack
ALB
Auto Scaling group
AMI
DynamoDB table
App base instance
Backup plan with copy
Multi-AZ RDS DB cluster
S3 Financial documents
DynamoDB table
AMI
Multi-AZ RDS DB cluster
CloudFormation stack
ALB
Auto Scaling group


Application Recovery Controller
Readiness check
Failover Infrastructure
us-east-1
us-west-2
S3 Financial documents
S3 Financial documents
CloudFormation stack
CloudFormation stack
ALB
ALB
Auto Scaling group
Auto Scaling group
AMI
DynamoDB table
DynamoDB table
App base instance
Backup plan with copy
AMI
Multi-AZ RDS DB cluster
Multi-AZ RDS DB cluster


Application Recovery Controller
Readiness check
Failover Infrastructure
us-east-1
us-west-2
S3 Financial documents
S3 Financial documents
CloudFormation stack
CloudFormation stack
ALB
ALB
Auto Scaling group
Auto Scaling group
AMI
DynamoDB table
DynamoDB table
App base instance
Backup plan with copy
AMI
Multi-AZ RDS DB cluster
Multi-AZ RDS DB cluster


Application Recovery Controller
Readiness check
Readiness check
Failover Infrastructure
us-east-1
us-west-2
S3 Financial documents
S3 Financial documents
CloudFormation stack
CloudFormation stack
ALB
ALB
Auto Scaling group
Auto Scaling group
AMI
DynamoDB table
DynamoDB table
App base instance
Backup plan with copy
AMI
Multi-AZ RDS DB cluster
Multi-AZ RDS DB cluster

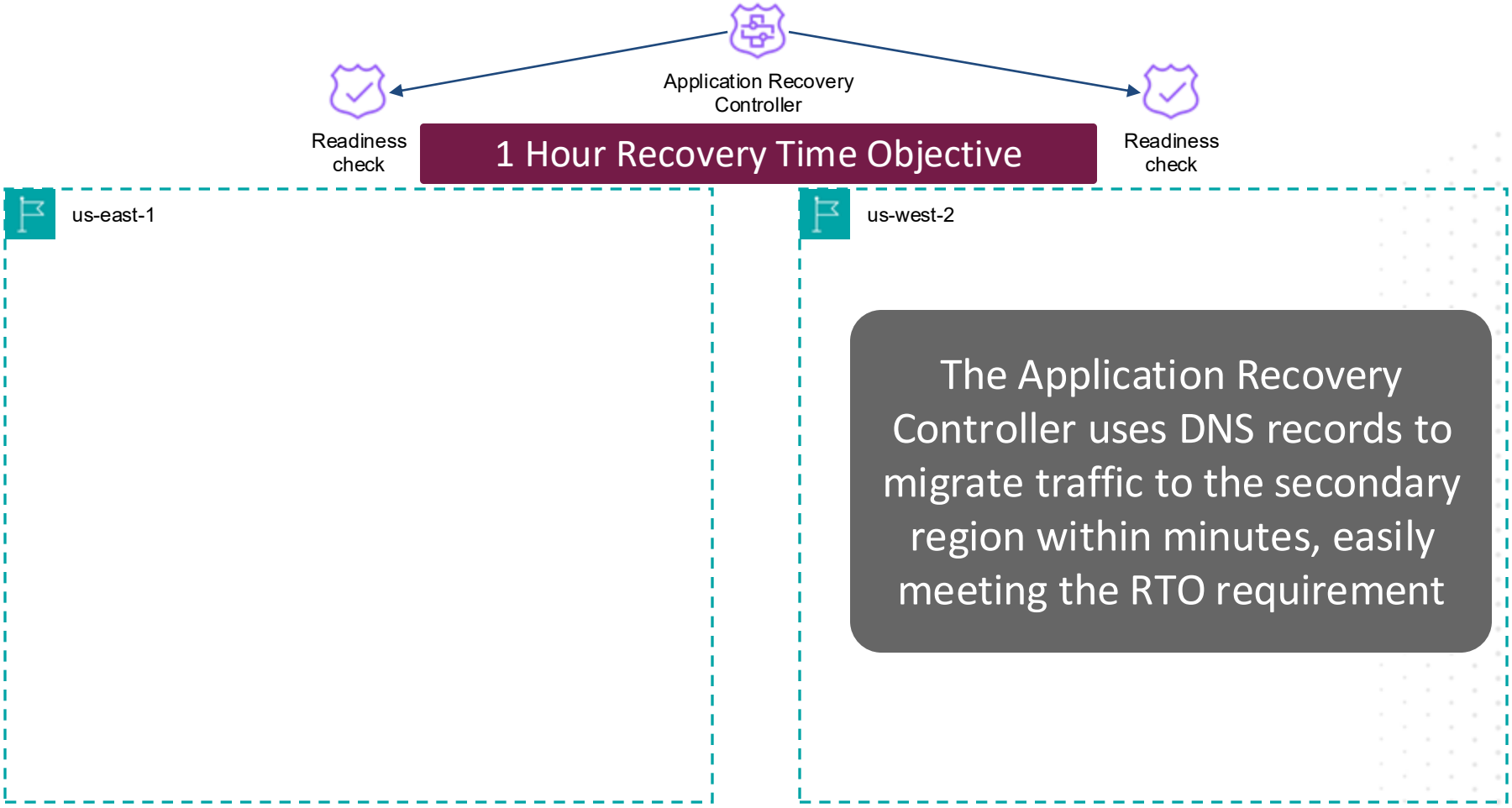

Application Recovery Controller
Readiness check
Readiness check
Failover Infrastructure
us-east-1
us-west-2
S3 Financial documents
S3 Financial documents
CloudFormation stack
CloudFormation stack
ALB
ALB
Auto Scaling group
Auto Scaling group
AMI
DynamoDB table
DynamoDB table
App base instance
Backup plan with copy
AMI
Multi-AZ RDS DB cluster
Multi-AZ RDS DB cluster

us-east-1

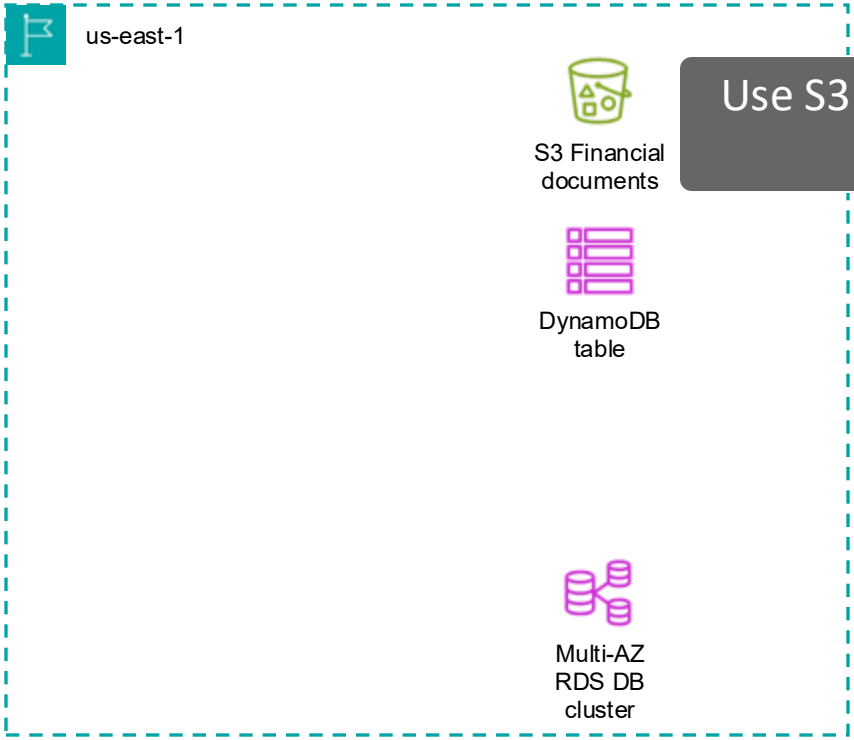
If the primary region fails it's Route 53 health check, the Readiness check also fails

us-west-2

Pearson

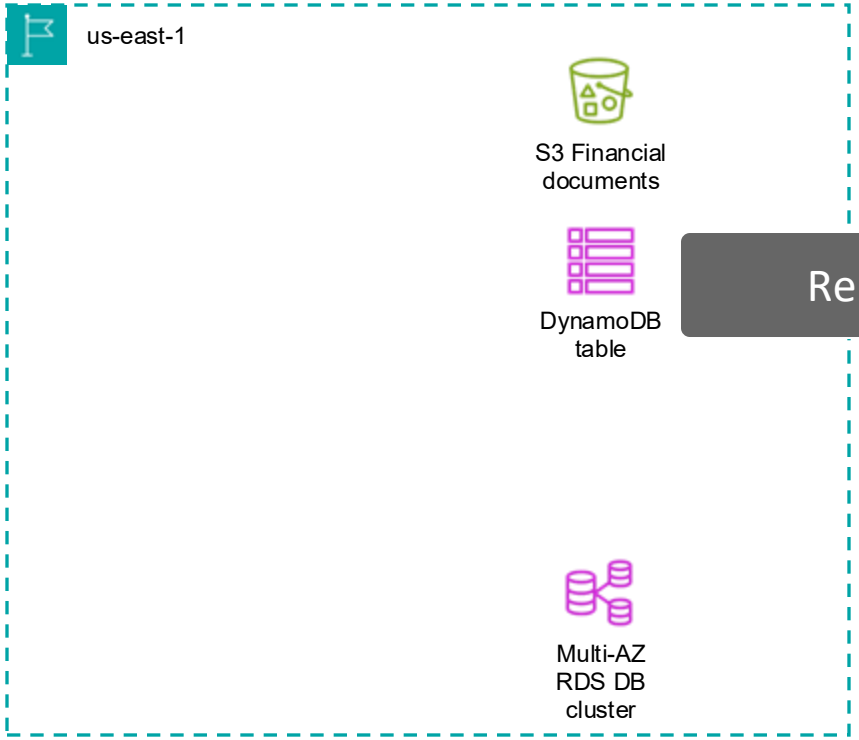

Application Recovery Controller
Readiness check
1 Hour Recovery Time Objective
Readiness check
us-east-1
us-west-2
The Application Recovery Controller uses DNS records to migrate traffic to the secondary region within minutes, easily meeting the RTO requirement

15-min Recovery Point Objective



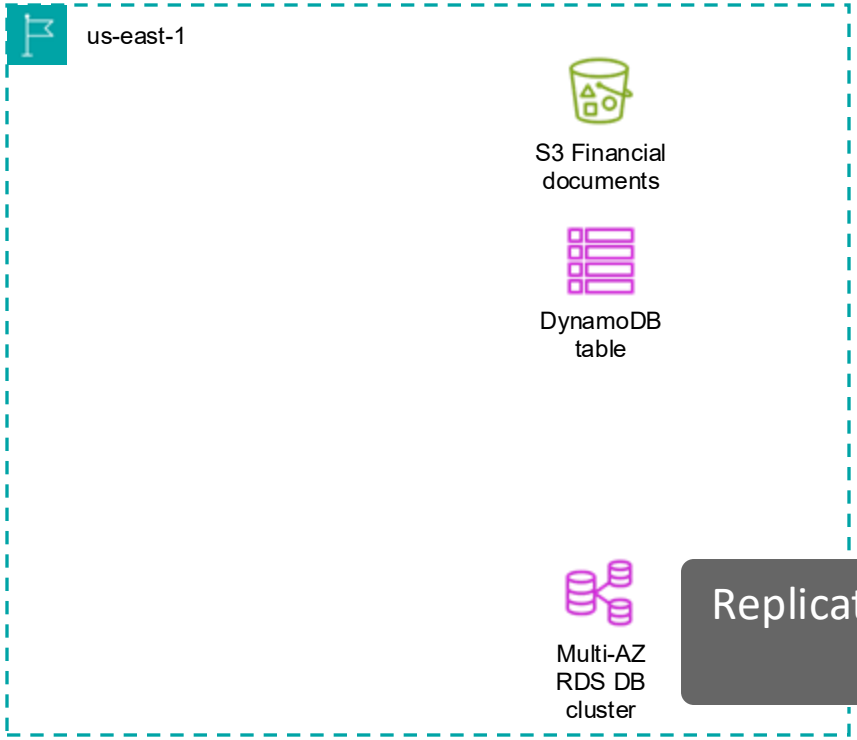
Use S3 replication time control for 15 minute SLA

15-min Recovery Point Objective



Replication typically within 1 second

15-min Recovery Point Objective



Replication speed according to resource sizing but can be less than 1 second

Question Breakdown

Test recovery procedures

Question and Answer Choices

An application has a requirement for cross-region recovery of a DynamoDB table in case of disaster. The requirements include being able to recover any deleted items. The table in the source region has DynamoDB Streams enabled.

What recommendation should be made to meet the requirement?

A. Configure the global table option and add the disaster recovery region
B. Provision a Lambda function triggered by the DynamoDB stream. Parse the stream entries and propagate Create and Update operations only to the remote table.
C. Provision a Lambda function triggered by the DynamoDB stream. Parse the stream entries and propagate all operations to the remote table.
D. Provision an EventBridge rule to capture all table events. Forward the events to a Lambda function which parses the events and applies them to the remote table.

Answer Choice Explanation

This would be a perfect solution, except that the global table feature also propagates item deletes.

A. **Configure the global table option and add the disaster recovery region**
B. **Provision a Lambda function triggered by the DynamoDB stream. Parse the stream entries and propagate Create and Update operations only to the remote table.**
C. **Provision a Lambda function triggered by the DynamoDB stream. Parse the stream entries and propagate all operations to the remote table.**
D. **Provision an EventBridge rule to capture all table events. Forward the events to a Lambda function which parses the events and applies them to the remote table.**

Answer Choice Explanation

This recommendation is functional, and accounts for the requirement to recover deleted items.

A. Configure the global table option and add the disaster recovery region
B. Provision a Lambda function triggered by the DynamoDB stream. Parse the stream entries and propagate Create and Update operations only to the remote table.
C. Provision a Lambda function triggered by the DynamoDB stream. Parse the stream entries and propagate all operations to the remote table.
D. Provision an EventBridge rule to capture all table events. Forward the events to a Lambda function which parses the events and applies them to the remote table.

Answer Choice Explanation

This solution is mostly functional, but similar to A in that item deletes would also be propagated.

A. Configure the global table option and add the disaster recovery region
B. Provision a Lambda function triggered by the DynamoDB stream. Parse the stream entries and propagate Create and Update operations only to the remote table.
C. Provision a Lambda function triggered by the DynamoDB stream. Parse the stream entries and propagate all operations to the remote table.
D. Provision an EventBridge rule to capture all table events. Forward the events to a Lambda function which parses the events and applies them to the remote table.

Answer Choice Explanation

This solution is yet another way to propagate all changes, similar to A and C, but will not meet the requirement for recovering deleted items.

A. Configure the global table option and add the disaster recovery region
B. Provision a Lambda function triggered by the DynamoDB stream. Parse the stream entries and propagate Create and Update operations only to the remote table.
C. Provision a Lambda function triggered by the DynamoDB stream. Parse the stream entries and propagate all operations to the remote table.
D. Provision an EventBridge rule to capture all table events. Forward the events to a Lambda function which parses the events and applies them to the remote table.

Correct Answer

Correct Answer: B

A. Configure the global table option and add the disaster recovery region
B. Provision a Lambda function triggered by the DynamoDB stream. Parse the stream entries and propagate Create and Update operations only to the remote table.
C. Provision a Lambda function triggered by the DynamoDB stream. Parse the stream entries and propagate all operations to the remote table.
D. Provision an EventBridge rule to capture all table events. Forward the events to a Lambda function which parses the events and applies them to the remote table.

Question Domain 2: Design Resilient Architectures

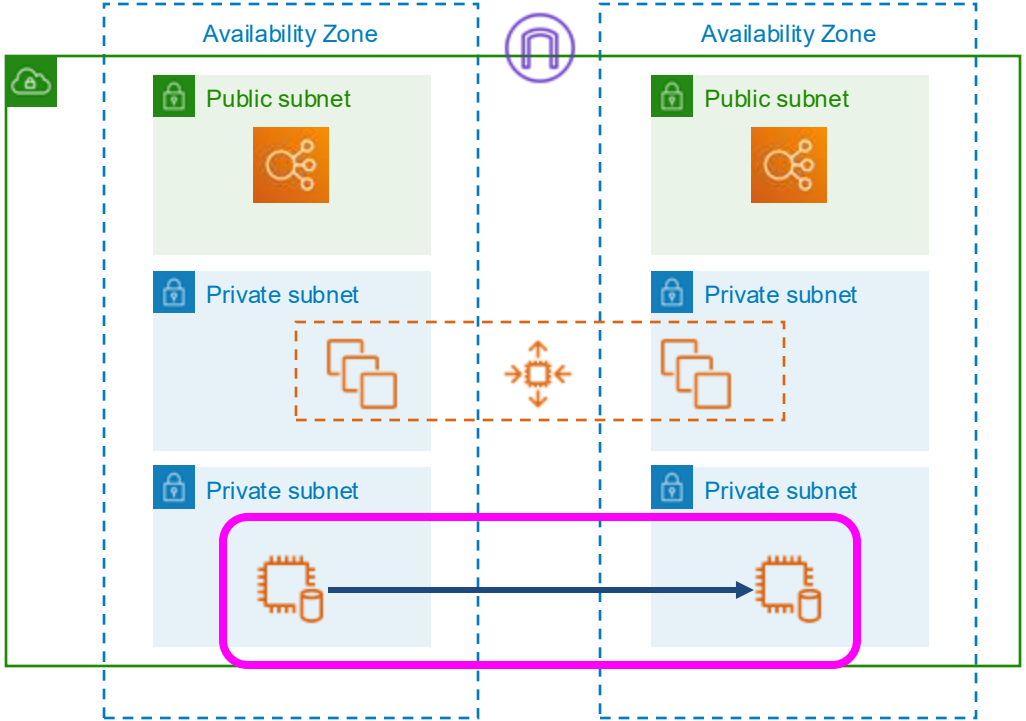
Scale horizontally to increase aggregate system availability

Principle Definition

Replace one large resource with multiple small resources to reduce the impact of a single failure on the overall workload.

Distribute requests across multiple, smaller resources to ensure that they don’t share a common point of failure.

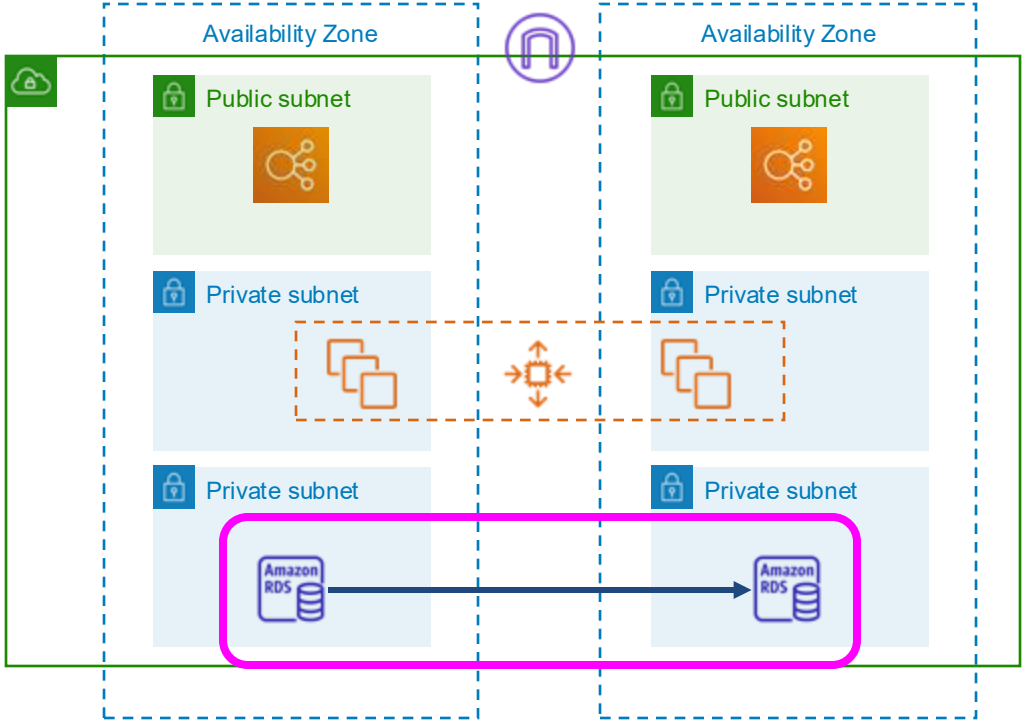
Managed Service Reliability Advantages



You can implement your DB on EC2 instances, including….

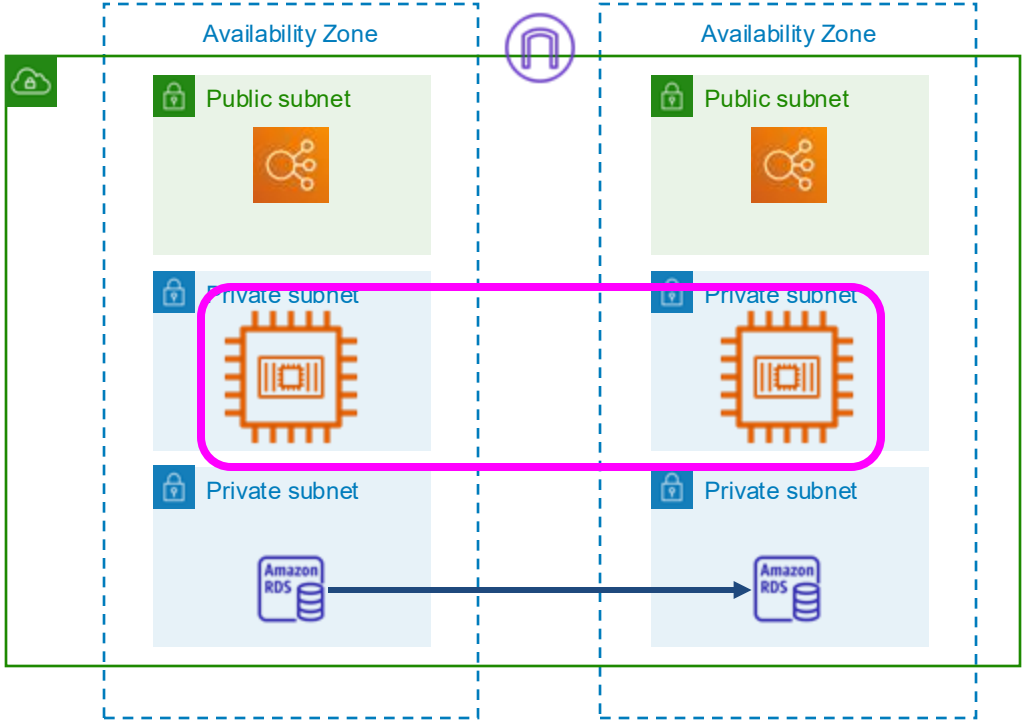
- Backups
- Replication
- Software updates
- Failover
- Restores

Managed Service Reliability Advantages



OR, you can deploy RDS and gain horizontal read scaling support

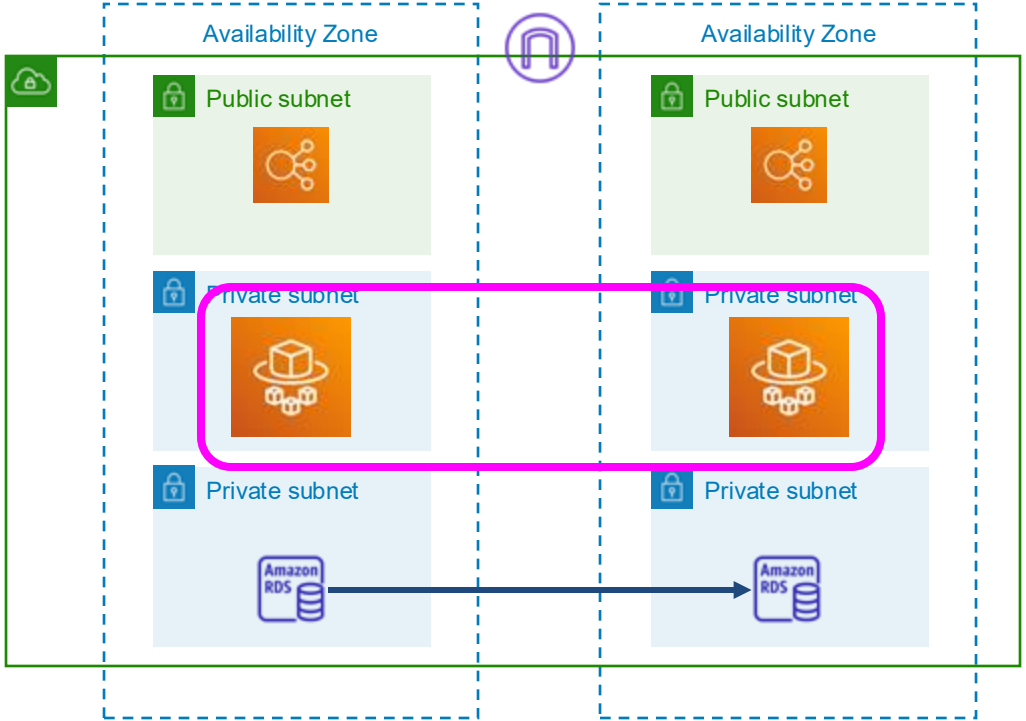
Managed Service Reliability Advantages



You can implement Docker containers on EC2 instances, including….

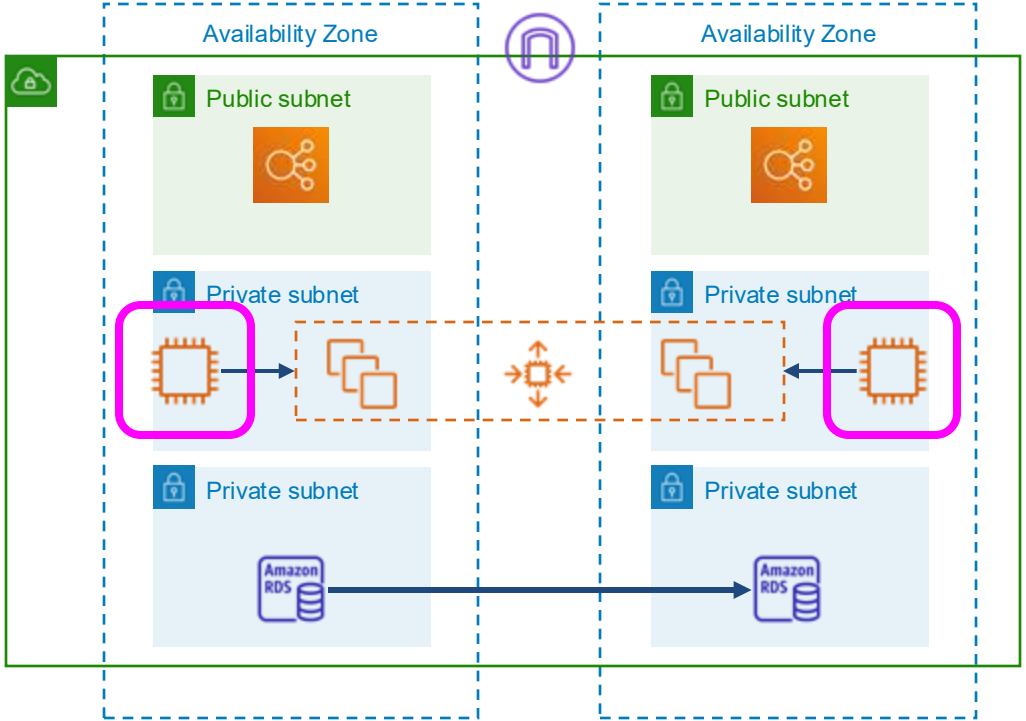
- Deployment
- Container scaling
- EC2 scaling
- OS updates
- Rollbacks

Managed Service Reliability Advantages



OR, you can deploy using Fargate and scale horizontally and serverless

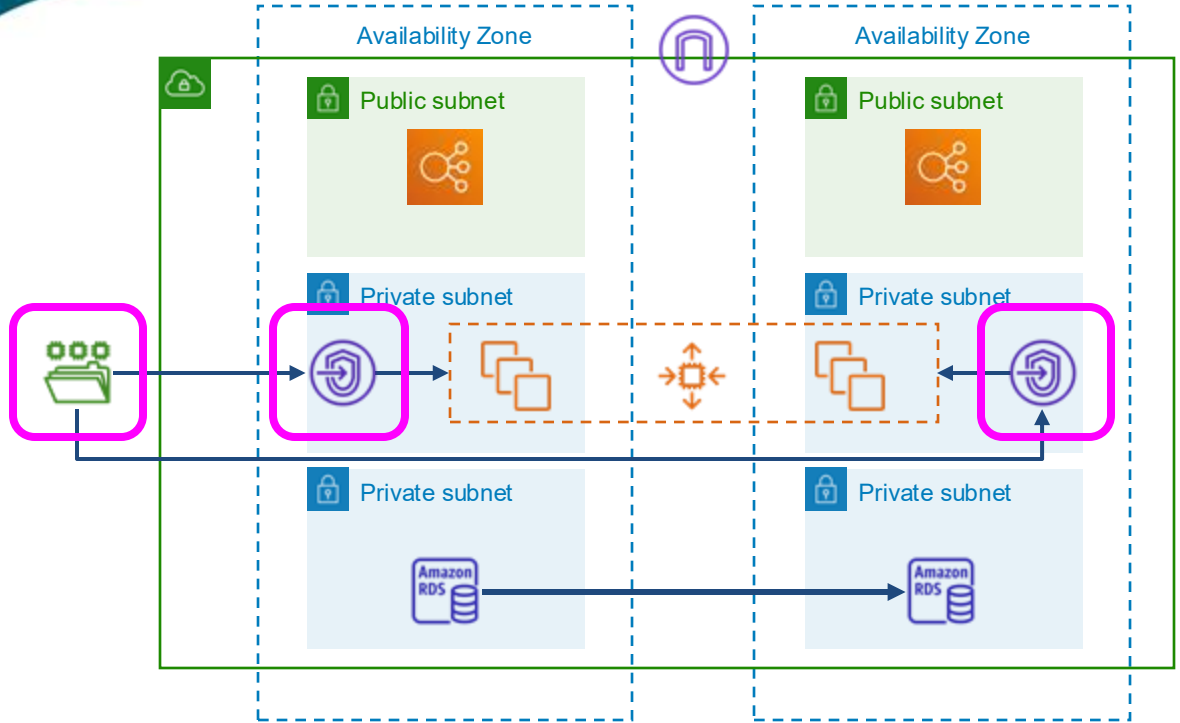
Managed Service Reliability Advantages



You can implement a shared filesystem on EC2, including….

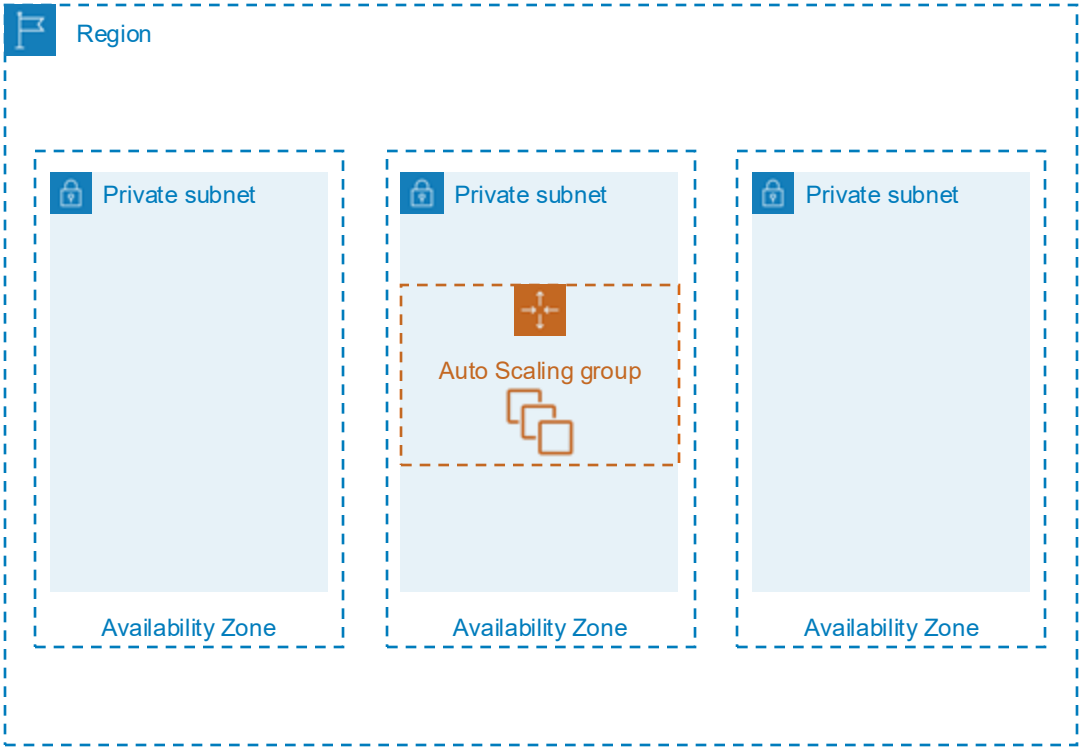
- Backups
- Replication
- Software updates
- Failover
- Restores

Managed Service Reliability Advantages



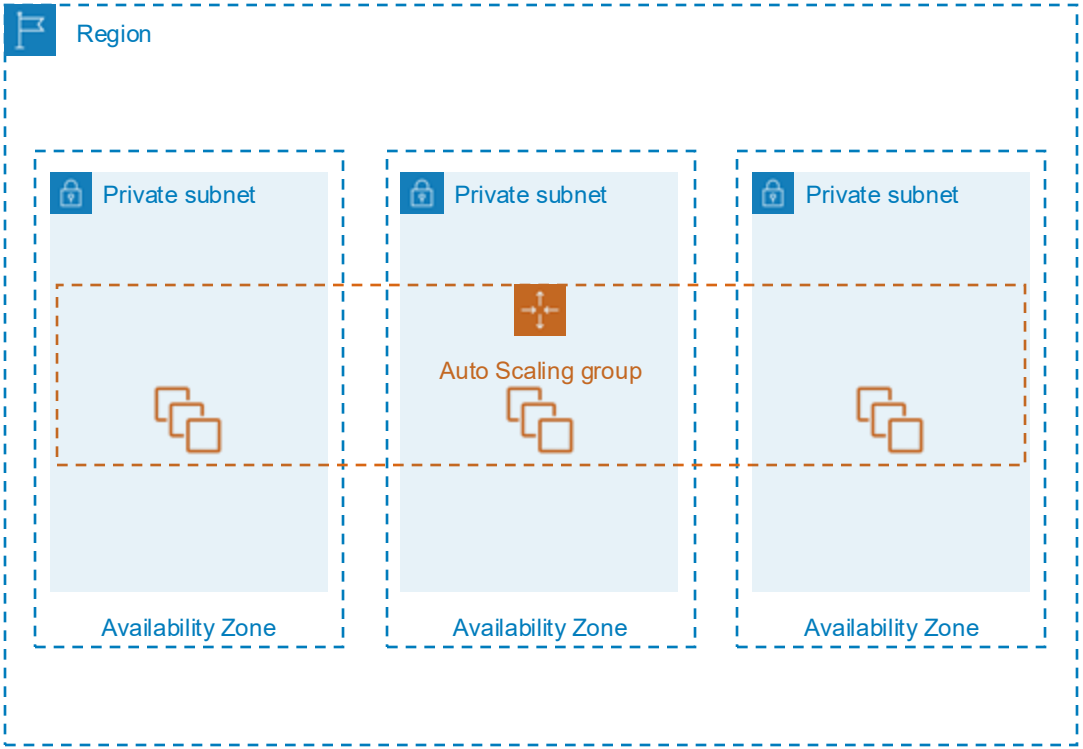
Or you can deploy a single EFS file system which is truly elastic

Scale Using AZ or Regions



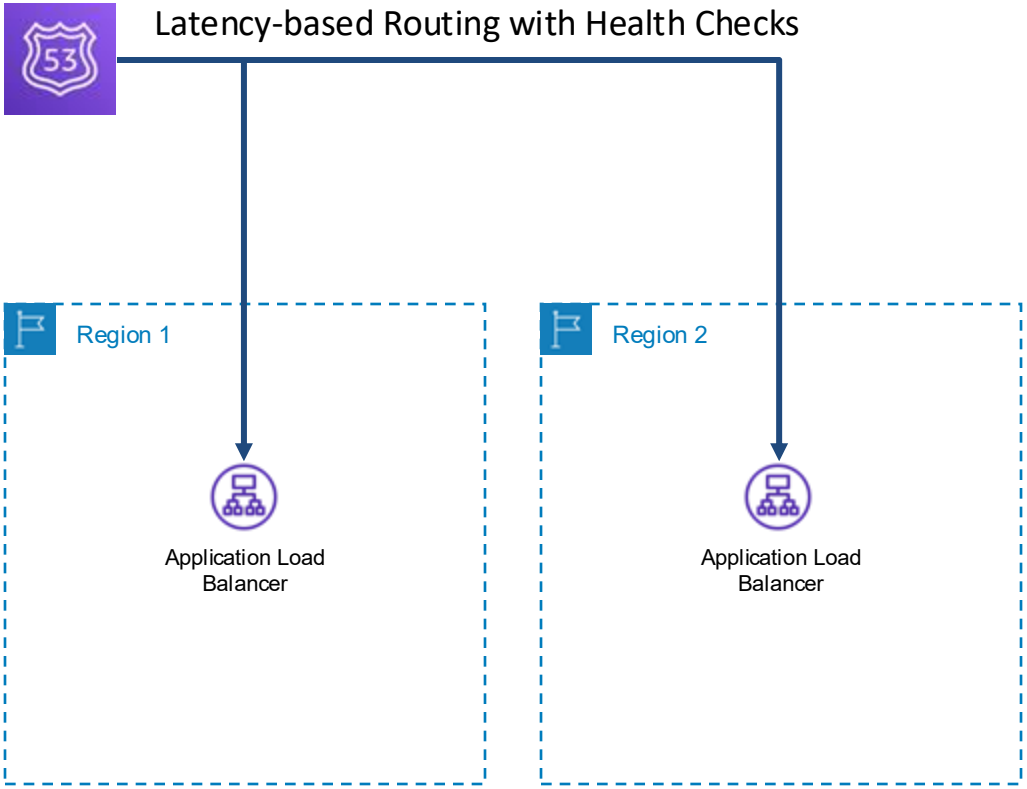
Deploy into a single AZ for performance or colocation with other resources

Scale Using AZ or Regions



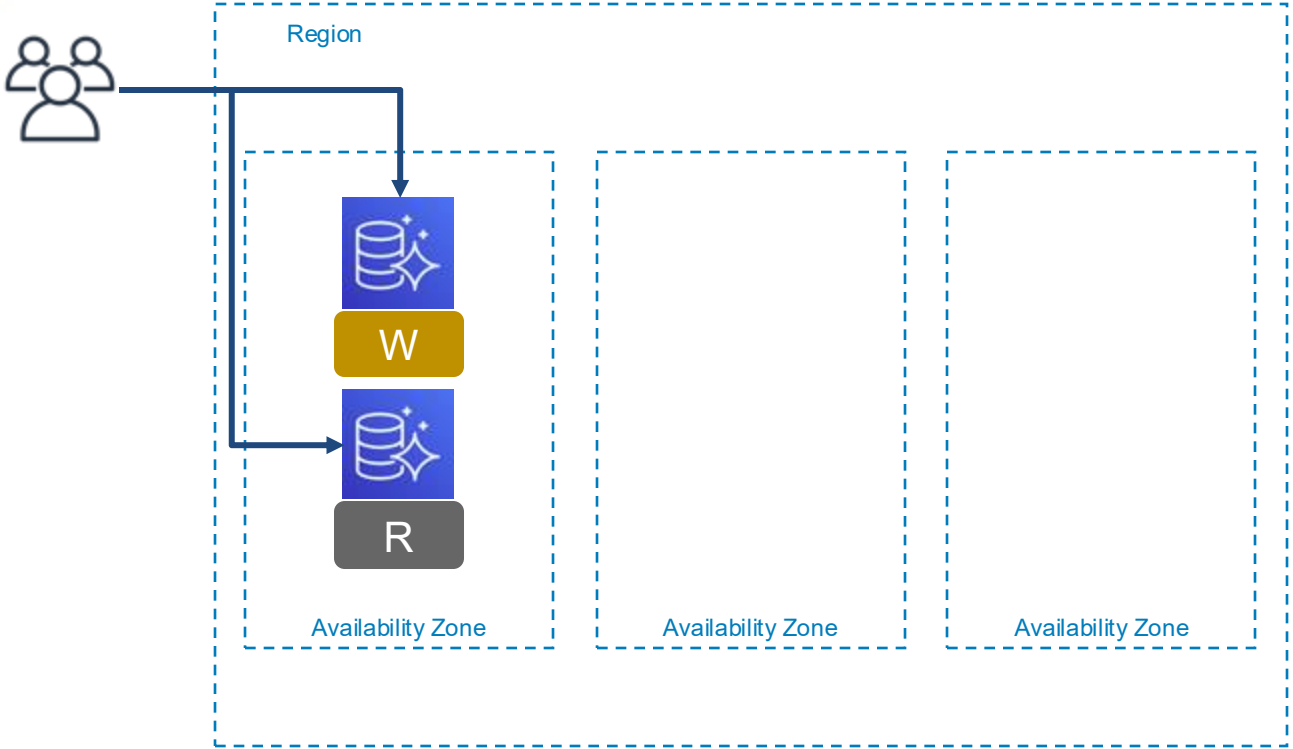
Deploy into multiple AZ for resilience

Scale Using AZ or Regions



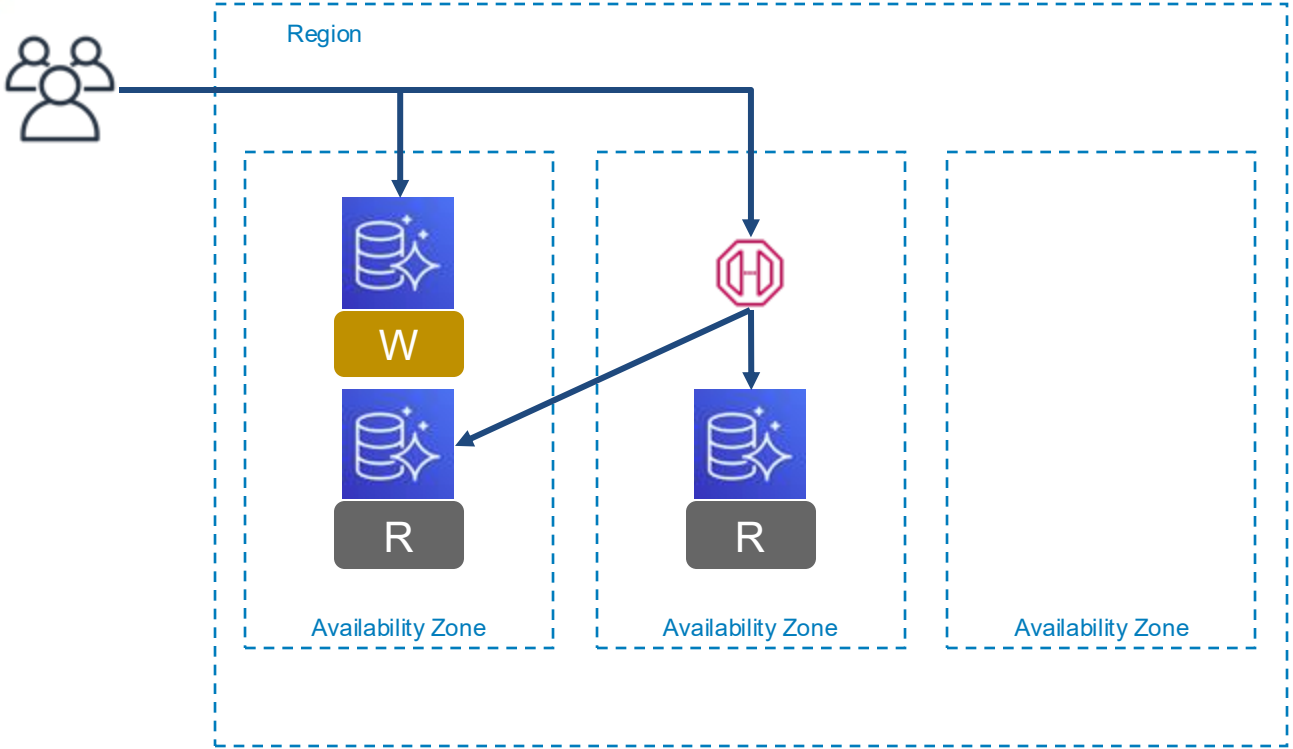
Deploy into multiple regions for extreme resilience

Avoid Monolithic Resources



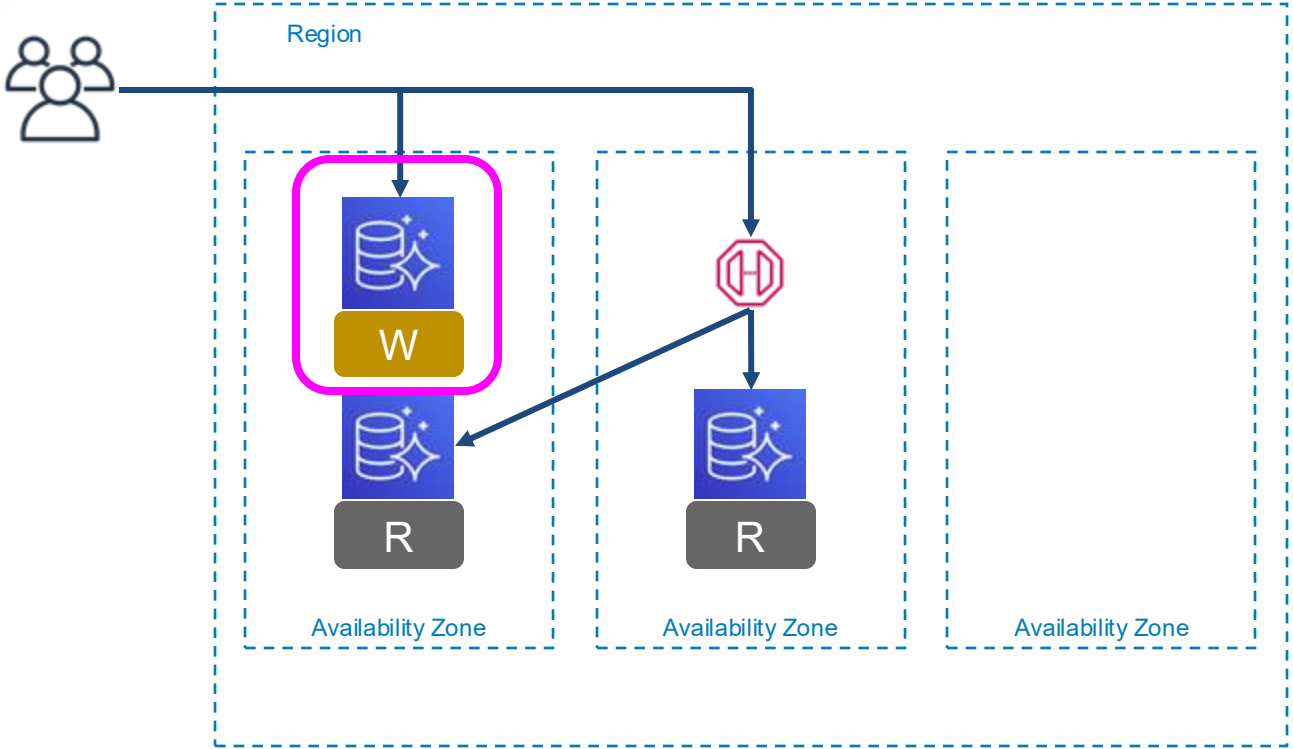
Aurora is resilient with a writer and one replica with automated failover

Avoid Monolithic Resources



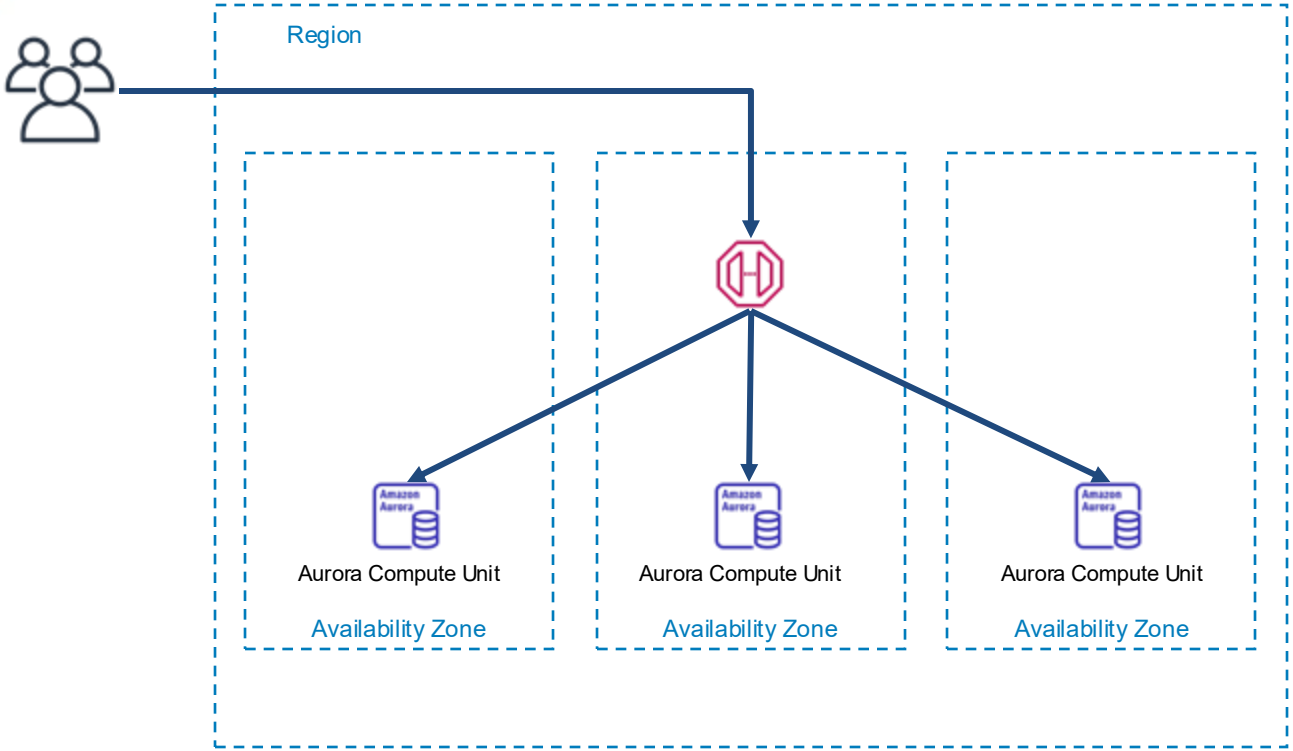
Aurora can auto scale multiple replicas using a reader endpoint

Avoid Monolithic Resources



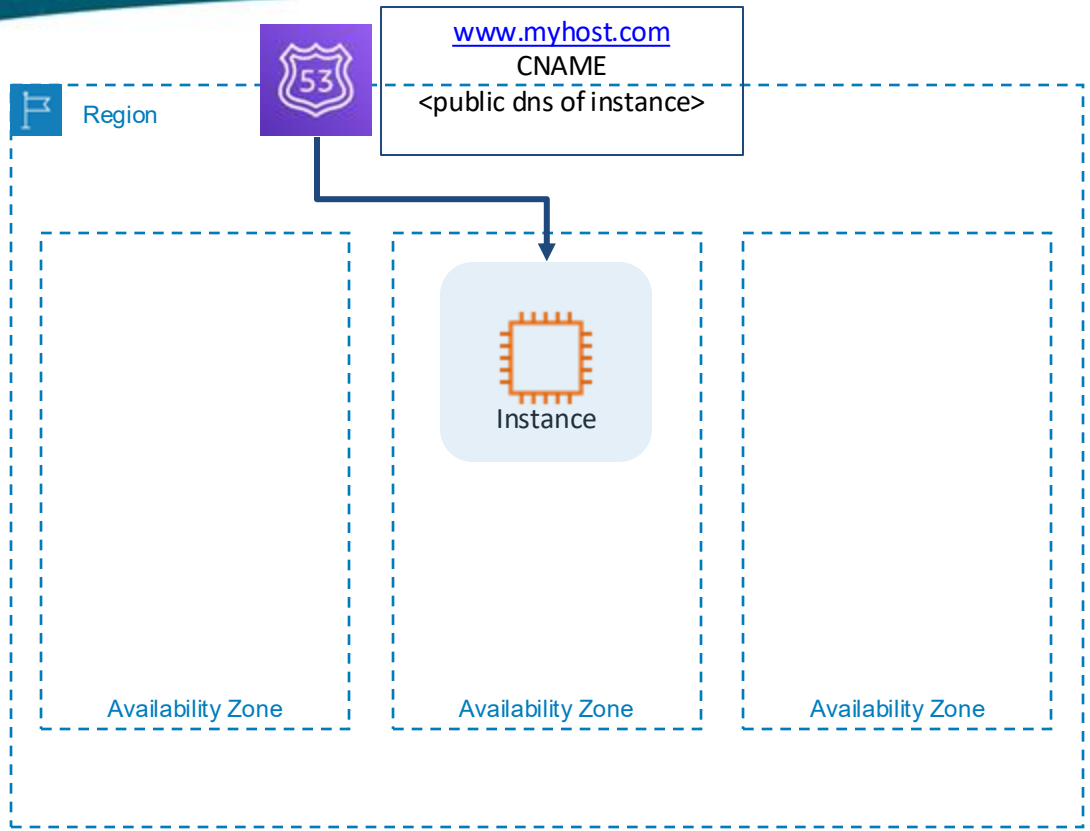
This still relies on a monolithic single point of failure for writes

Avoid Monolithic Resources



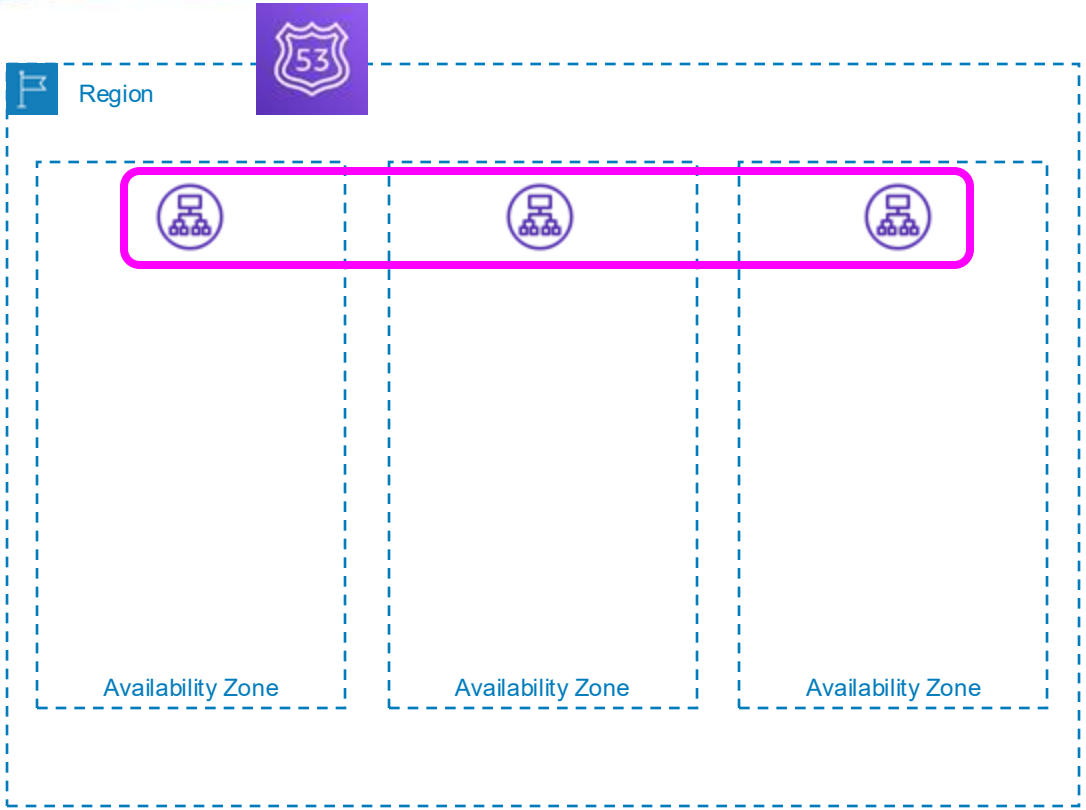
Instead, deploy Aurora Serverless to scale both reads and writes to match demand

Decouple and Scale Independently



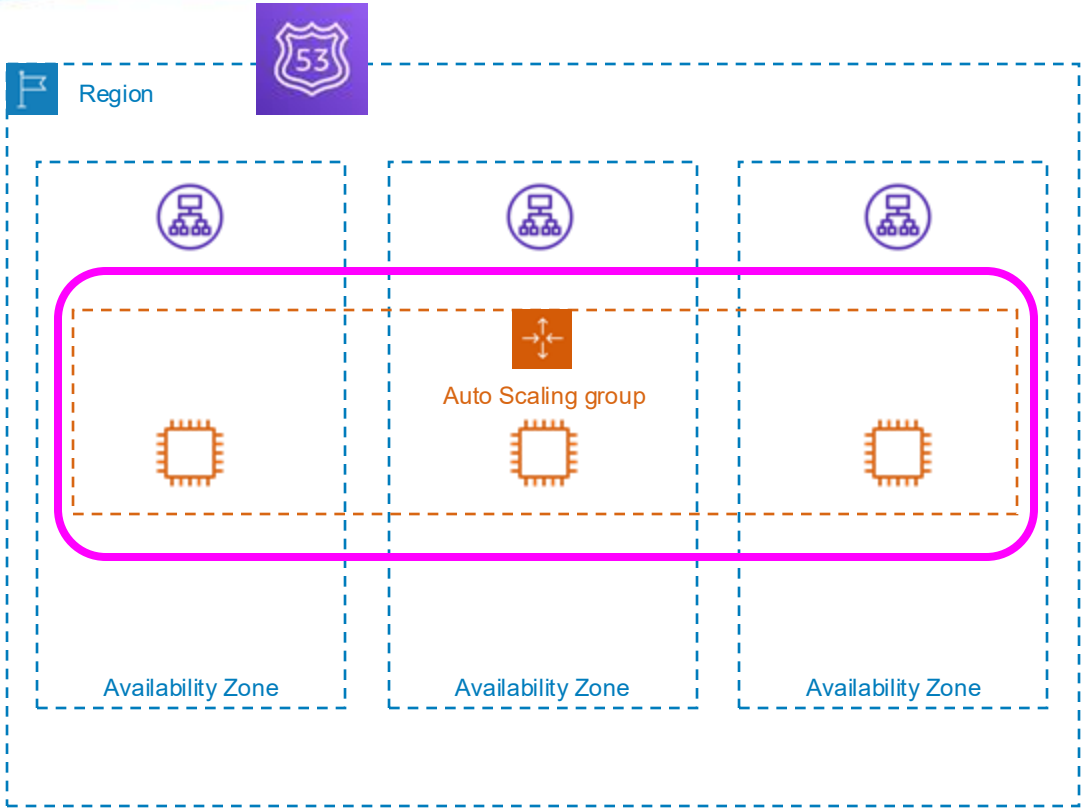
This is a tightly coupled architecture

Decouple and Scale Independently



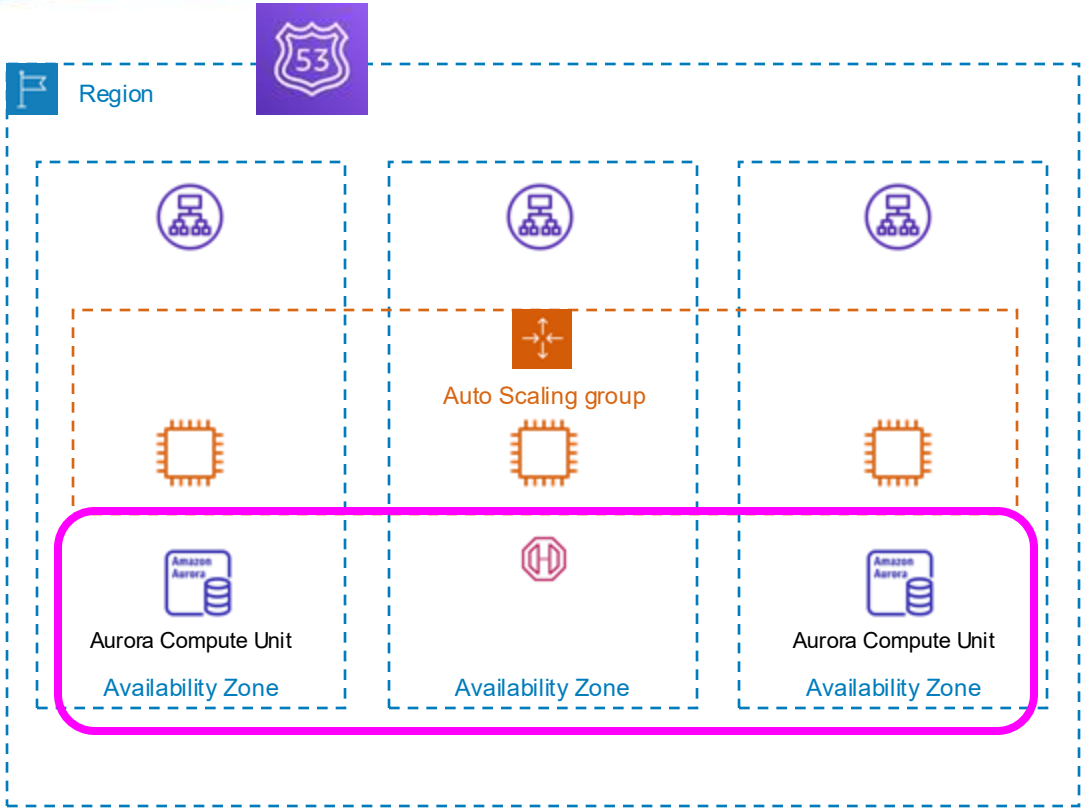
Separate the web proxy to an ALB which scales automatically

Decouple and Scale Independently



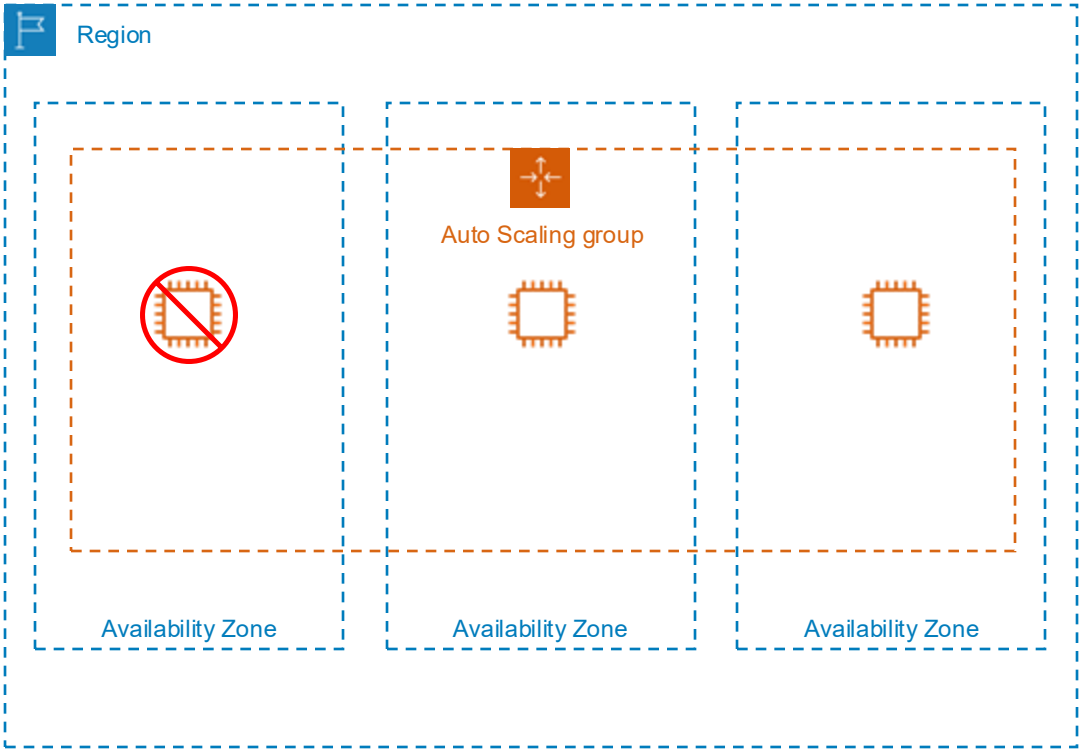
Deploy the application tier using Auto Scaling to scale horizontally

Decouple and Scale Independently



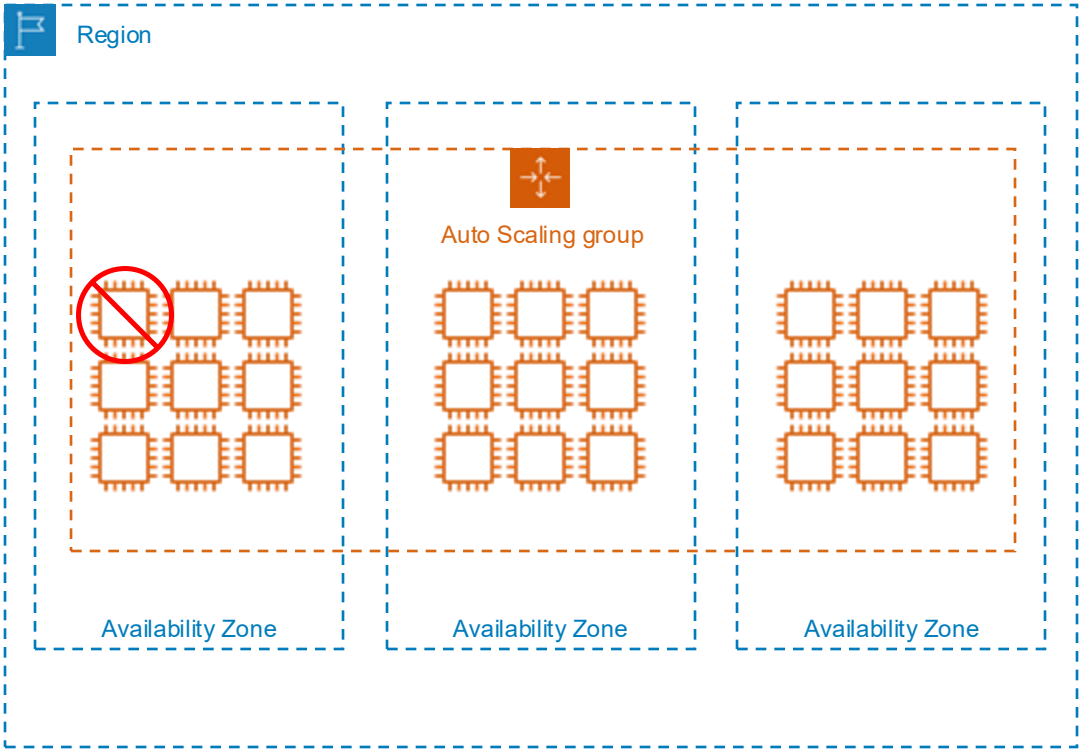
Separate the DB to Aurora Serverless to also scale horizontally according to demand

More, Smaller Resources



With 3 8xlarge instances losing 1 means 33% of requests affected

More, Smaller Resources



With 27 large instances losing 1 means 4% of requests affected

Question Breakdown

Scale horizontally to increase aggregate system availability

Question and Answer Choices

A stateful eCommerce EC2 application has been deployed using Auto Scaling behind an Application Load Balancer. During scaling, customers are complaining about their shopping carts suddenly being empty.

What steps could be taken to ensure shopping cart data is retained during EC2 scaling events?

A. Change the Auto Scaling group to a steady state group (same minimum, maximum, desired instances) to match the peak traffic load
B. Enable duration-based stickiness in the EC2 Target group to ensure all requests from the same client are routed to the same target
C. Enable application-based stickiness in the EC2 Target group to ensure all requests from the same client are routed to the same target
D. Migrate the shopping cart data to a DynamoDB table and reference the table instead of the local application cache

Answer Choice Explanation

This solution can be an incremental improvement, but the true fix for "scaling causes problems" really shouldn't be "don't scale at all", and this does not account for AWS-related events that cause instance replacement.

A. **Change the Auto Scaling group to a steady state group (same minimum, maximum, desired instances) to match the peak traffic load**
B. **Enable duration-based stickiness in the EC2 Target group to ensure all requests from the same client are routed to the same target**
C. **Enable application-based stickiness in the EC2 Target group to ensure all requests from the same client are routed to the same target**
D. **Migrate the shopping cart data to a DynamoDB table and reference the table instead of the local application cache**

Answer Choice Explanation

Enabling TG stickiness can improve the situation by ensuring that each client stays on the same target for the duration set, but does not account for scaling-in activity.

A. Change the Auto Scaling group to a steady state group (same minimum, maximum, desired instances) to match the peak traffic load
B. Enable duration-based stickiness in the EC2 Target group to ensure all requests from the same client are routed to the same target
C. Enable application-based stickiness in the EC2 Target group to ensure all requests from the same client are routed to the same target
D. Migrate the shopping cart data to a DynamoDB table and reference the table instead of the local application cache

Answer Choice Explanation

Similar to B, enabling TG stickiness can improve the situation by ensuring that each client stays on the same target for the duration set, but does not account for scaling-in activity. This could be slightly better than B if the cookie is maintained for a long period of time.

A. Change the Auto Scaling group to a steady state group (same minimum, maximum, desired instances) to match the peak traffic load
B. Enable duration-based stickiness in the EC2 Target group to ensure all requests from the same client are routed to the same target
C. Enable application-based stickiness in the EC2 Target group to ensure all requests from the same client are routed to the same target
D. Migrate the shopping cart data to a DynamoDB table and reference the table instead of the local application cache

Answer Choice Explanation

None of the other solutions truly address the root cause of the problem, which is that the application is stateful. Re-architecting the application to be stateless by moving the shopping cart persistence to a separate, reliable/durable storage service will result in the best overall solution from a resilience perspective (but could have tradeoff implications on performance due to extra latency of accessing shopping cart data).

A. Change the Auto Scaling group to a steady state group (same minimum, maximum, desired instances) to match the peak traffic load
B. Enable duration-based stickiness in the EC2 Target group to ensure all requests from the same client are routed to the same target
C. Enable application-based stickiness in the EC2 Target group to ensure all requests from the same client are routed to the same target
D. Migrate the shopping cart data to a DynamoDB table and reference the table instead of the local application cache

Correct Answer

Correct Answer: D

A. Change the Auto Scaling group to a steady state group (same minimum, maximum, desired instances) to match the peak traffic load
B. Enable duration-based stickiness in the EC2 Target group to ensure all requests from the same client are routed to the same target
C. Enable application-based stickiness in the EC2 Target group to ensure all requests from the same client are routed to the same target
D. Migrate the shopping cart data to a DynamoDB table and reference the table instead of the local application cache

Question Domain 2: Design Resilient Architectures

Stop guessing capacity

Principle Definition

A common cause of failure in on-premises workloads is resource saturation, when the demands placed on a workload exceed the capacity of that workload (this is often the objective of denial of service attacks).

In the cloud, you can monitor demand and workload utilization, and automate the addition or removal of resources to maintain the optimal level to satisfy demand without over- or under-provisioning.

There are still limits, but some quotas can be controlled and others can be managed.

Stop Guessing Capacity

Use automation when obtaining or scaling resources

Amazon Simple Storage Service (Amazon S3)

Prefix performance of 3500 writes/sec or 5500 reads/sec

Stop Guessing Capacity

Use automation when obtaining or scaling resources

Amazon Simple Storage Service (Amazon S3)

Prefix performance of 3500 writes/sec or 5500 reads/sec

AWS Lambda

Concurrency configuration set to maximum number of invocations at the same time

Stop Guessing Capacity

Use automation when obtaining or scaling resources

Amazon Simple Storage Service (Amazon S3)

Prefix performance of 3500 writes/sec or 5500 reads/sec

AWS Lambda

Concurrency configuration set to maximum number of invocations at the same time

Amazon CloudFront

Scaling per cached edge location to reduce load on expensive resources

Stop Guessing Capacity

Obtain resources upon detection of impairment to a workload

Health checks

Route 53 health checks can test AWS or outside endpoints

Stop Guessing Capacity

Obtain resources upon detection of impairment to a workload

Health checks

Route 53 health checks can test AWS or outside endpoints

Elastic Load Balancing

ELB Target Group health checks can detect failed resources and remove them from service

Stop Guessing Capacity

Obtain resources upon detection of impairment to a workload

Health checks

Route 53 health checks can test AWS or outside endpoints

Elastic Load Balancing

ELB Target Group health checks can detect failed resources and remove them from service

Alarm

CloudWatch Alarms can evaluate any number-based metric against thresholds

Stop Guessing Capacity

Obtain resources upon detection that more resources are needed for a workload

Amazon EC2
Auto Scaling

Amazon Aurora

Amazon Elastic Container
Service (Amazon ECS)

Amazon DynamoDB

AWS Auto Scaling adds or removes resources to match demand

Stop Guessing Capacity

Obtain resources upon detection that more resources are needed for a workload

Amazon EC2 Auto Scaling

Amazon Aurora

Amazon Elastic Container Service (Amazon ECS)

Amazon DynamoDB

EC2 Auto Scaling can use Predictive Scaling to schedule scaling activities in advance

Stop Guessing Capacity

Load test your workload

AWS CloudFormation

Deploy a temporary load testing environment using automation

Stop Guessing Capacity

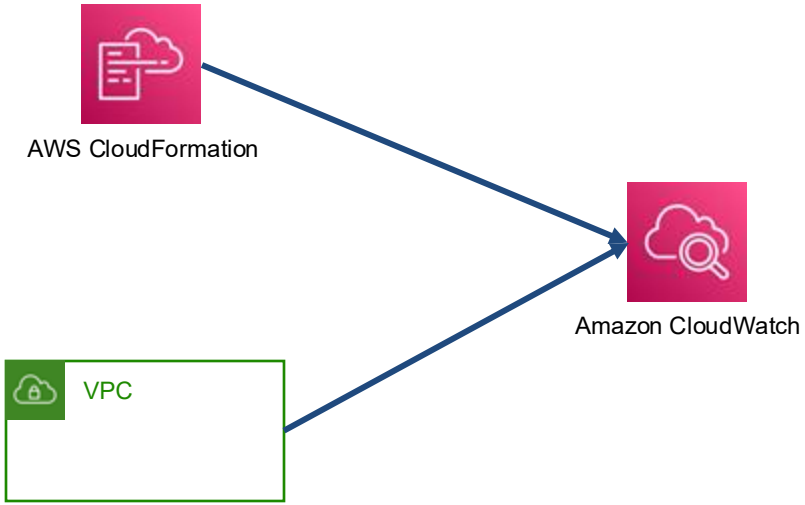
Load test your workload

AWS CloudFormation

Or use the existing production resources

Stop Guessing Capacity

Load test your workload



Evaluate metrics, learn from breakage and mitigate with improved design

Question Breakdown

Stop guessing capacity

Question and Answer Choices

A web application is being architected for deployment into AWS. There is a requirement to implement throttling of web requests when the back end latency increases beyond a specific threshold, in order to ensure an optimal user experience.

Which of the below recommendations can meet this requirement?

A. Application Load Balancer listener rules
B. CloudFront distribution behaviors
C. WAF Web ACL rate-based rule
D. API Gateway usage plans
E. A custom solution is required

Answer Choice Explanation

Listener rules have several options, including static responses, redirects or forwarding to a Target Group, but no option for rejecting requests outright.

A. Application Load Balancer listener rules
B. CloudFront distribution behaviors
C. WAF Web ACL rate-based rule
D. API Gateway usage plans
E. A custom solution is required

Answer Choice Explanation

While CloudFront distributions can be associated with a WAF, they have no intrinsic filtering capability beyond geo blocking.

A. Application Load Balancer listener rules
B. CloudFront distribution behaviors
C. WAF Web ACL rate-based rule
D. API Gateway usage plans
E. A custom solution is required

Answer Choice Explanation

This sounds like an acceptable solution, except that the throttling must be configured via a static threshold, which will not account for the busy-ness of the back end systems.

A. Application Load Balancer listener rules
B. CloudFront distribution behaviors
C. WAF Web ACL rate-based rule
D. API Gateway usage plans
E. A custom solution is required

Answer Choice Explanation

API Gateway usage plans are similar to WAF rate-based rules in that a static threshold must be set and anything beyond that is throttled. It cannot apply a dynamic rule based on an existing metric.

A. Application Load Balancer listener rules
B. CloudFront distribution behaviors
C. WAF Web ACL rate-based rule
D. API Gateway usage plans
E. A custom solution is required

Correct Answer

Correct Answer: E

A. Application Load Balancer listener rules
B. CloudFront distribution behaviors
C. WAF Web ACL rate-based rule
D. API Gateway usage plans
E. A custom solution is required

Question Domain 2: Design Resilient Architectures

Manage change in automation

Principle Definition

Changes to your infrastructure should be made using automation.

The changes that need to be managed include changes to the automation, which then can be tracked and reviewed.

Event-Driven Architecture Options

AWS Personal
Health Dashboard

AWS schedules hardware retirement due to degraded hardware

Event-Driven Architecture Options

Event is passed to EventBridge and captured via rule

Event-Driven Architecture Options

Rule target invokes a Lambda function

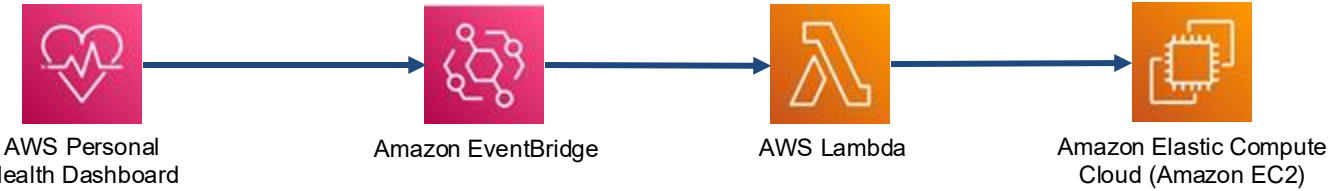
Event-Driven Architecture Options

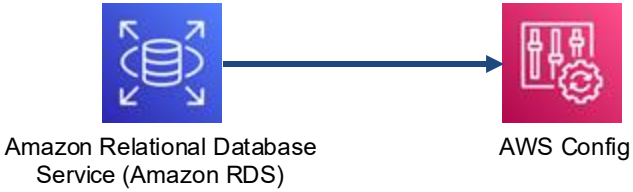
Lambda function performs a stop and start of the instance to mitigate

Event-Driven Architecture Options

RDS instance is launched without Multi-AZ enabled

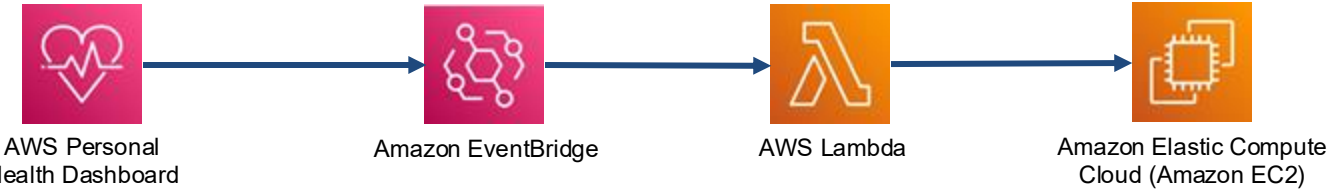
Event-Driven Architecture Options

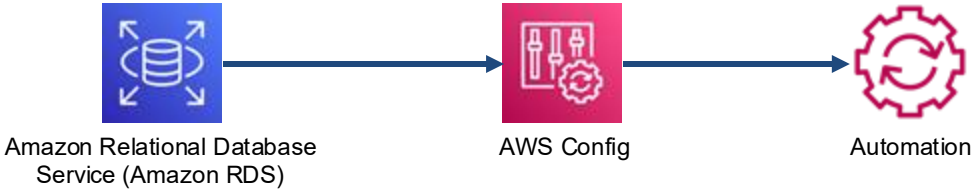




RDS instance properties are delivered to Config recording stream and captured via rule

Event-Driven Architecture Options





Config rule executes SSM automation document to enable Multi-AZ on RDS

Deployment Types

In-place

Update the application version without modifying or replacing any infrastructure components

Deployment Types

In-place

Rolling

Slowly replace previous application versions by replacing infrastructure components

Deployment Types

- In-place
- Rolling
- Blue/Green

Use two separate environments, and gradually migrate traffic from the old (blue) to the new (green)

Deployment Types

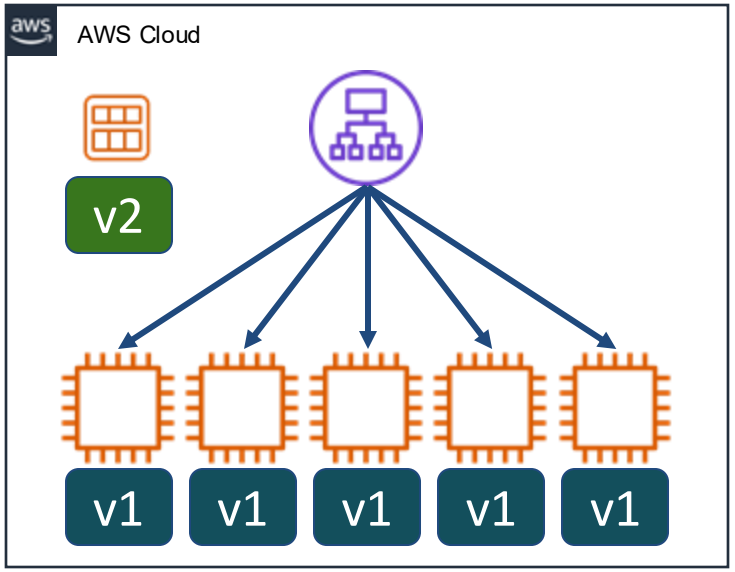
- In-place
- Rolling
- Blue/Green
- Canary

Used at the start of a deployment to test in-place application updates on one resource to determine viability

Deployment Types

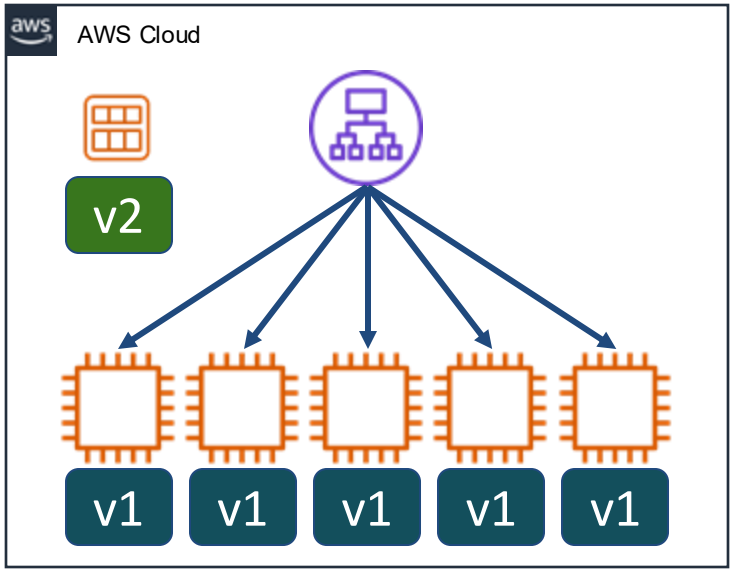
	Immutable?
In-place	
Rolling	✓
Blue/Green	✓
Canary	

Rolling Deployment



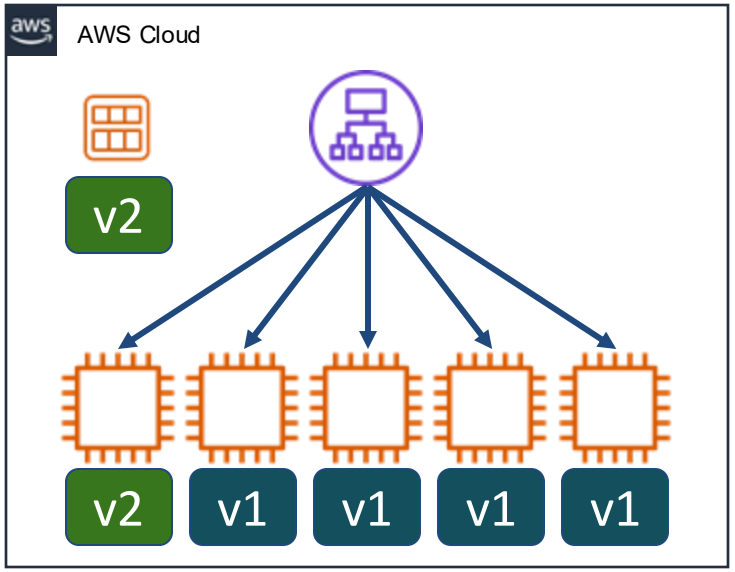
Create updated AMI

Rolling Deployment



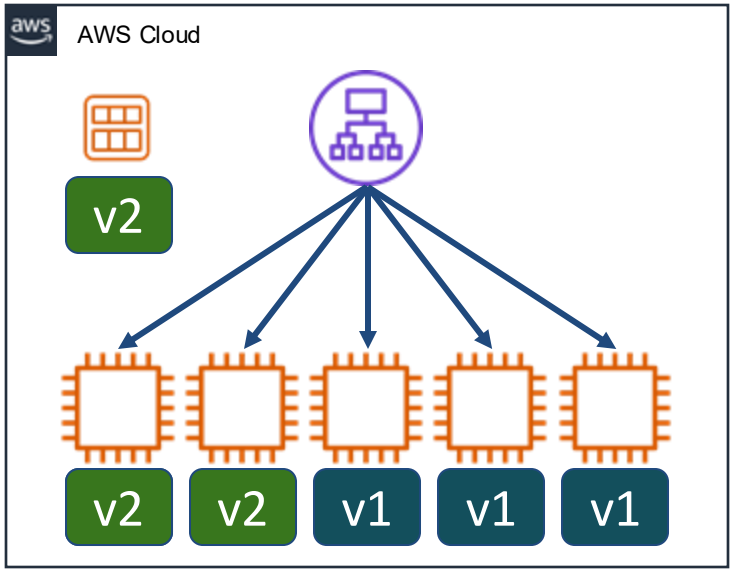
Terminate or scale in

Rolling Deployment



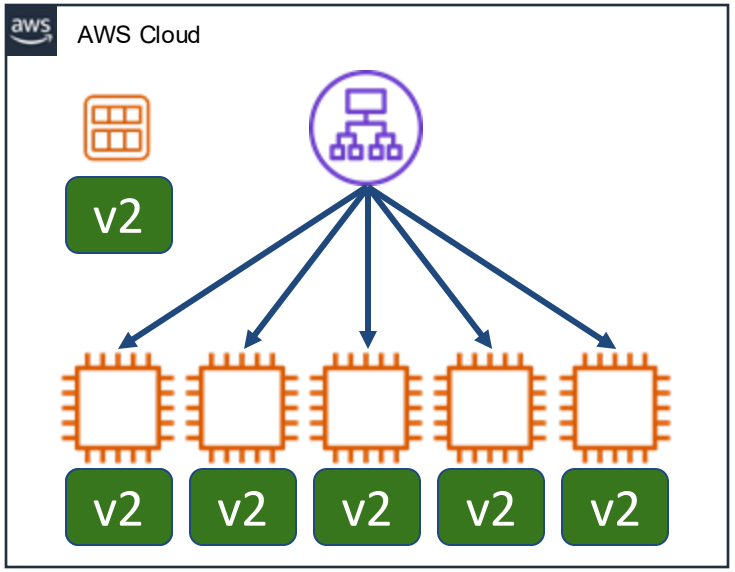
Scale out by the same increment

Rolling Deployment



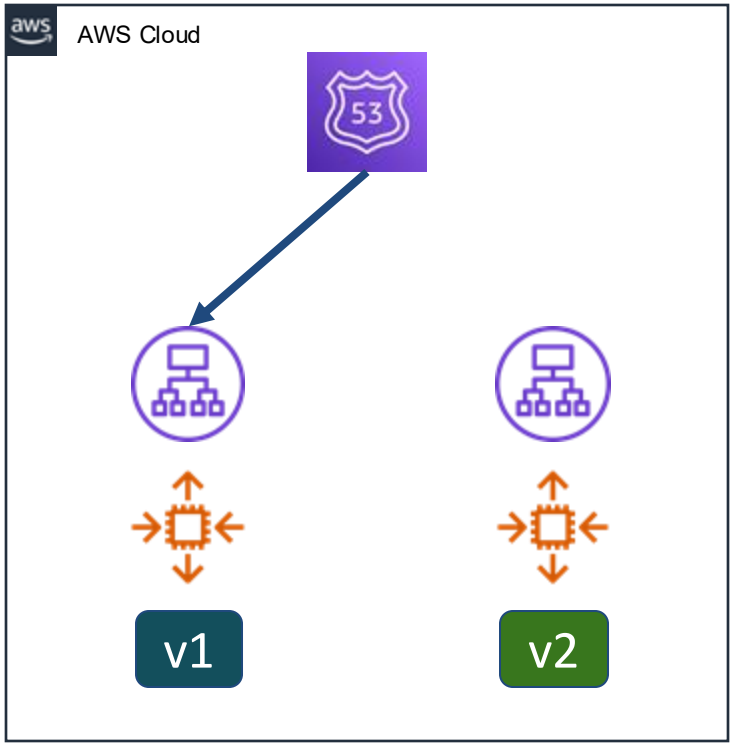
Continue replacement with equal size phases

Rolling Deployment



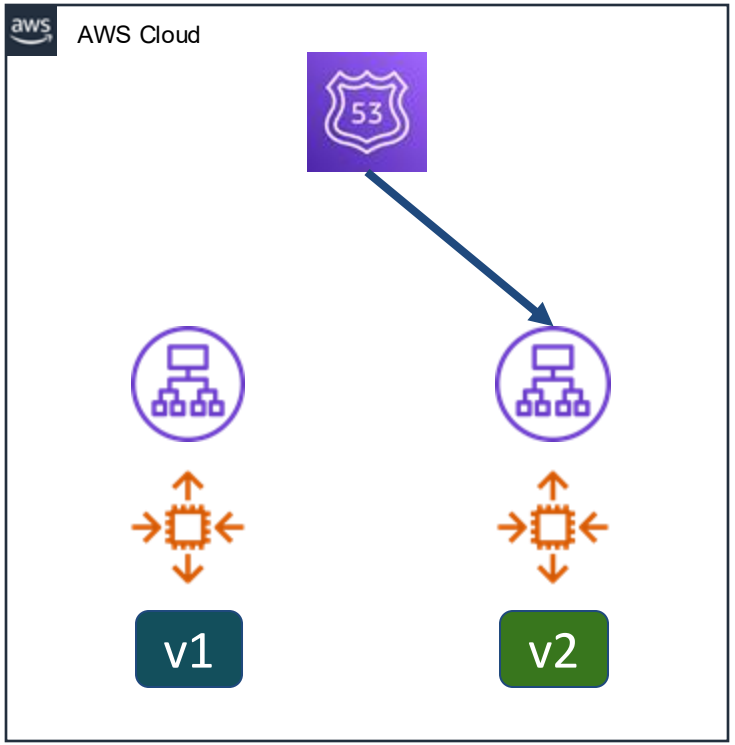
Continue replacement with equal size phases

Blue/Green Deployment Example 1



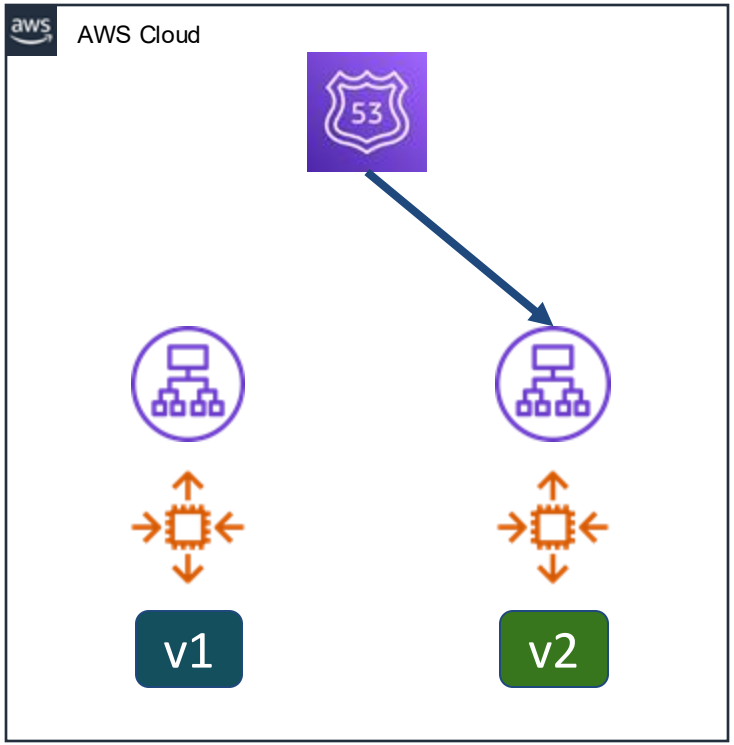
Provision the v2 infrastructure

Blue/Green Deployment Example 1



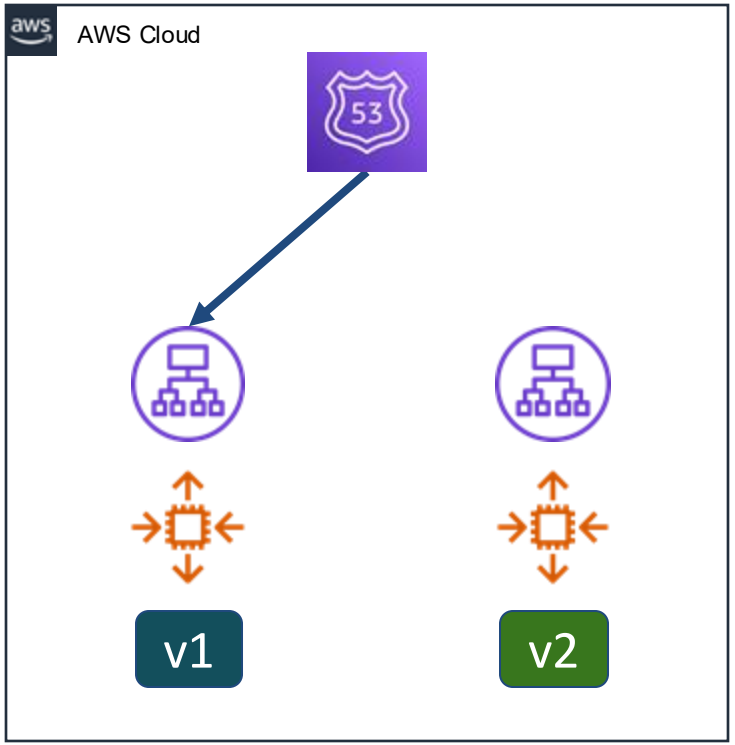
Update DNS to point to v2

Blue/Green Deployment Example 1



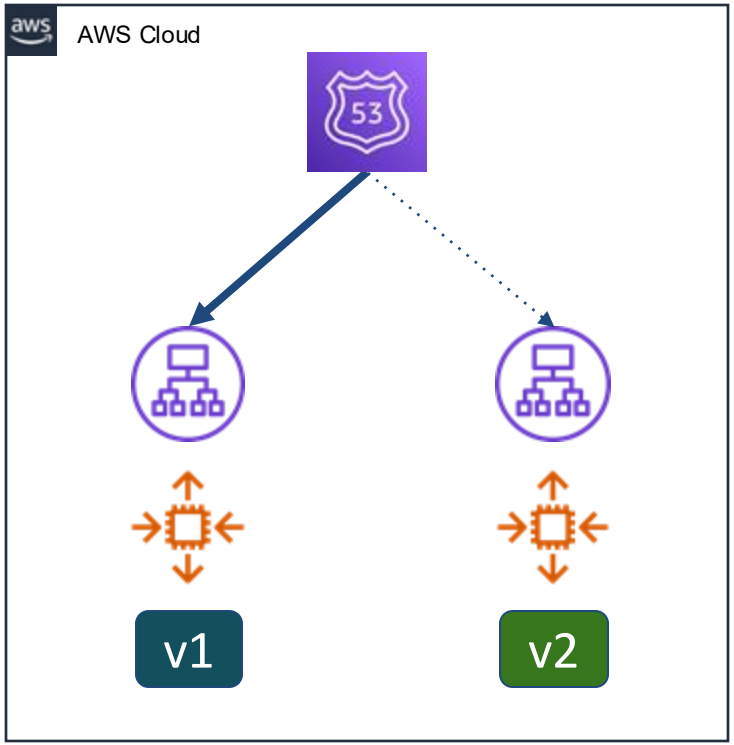
De-provision v1

Blue/Green Deployment Example 2



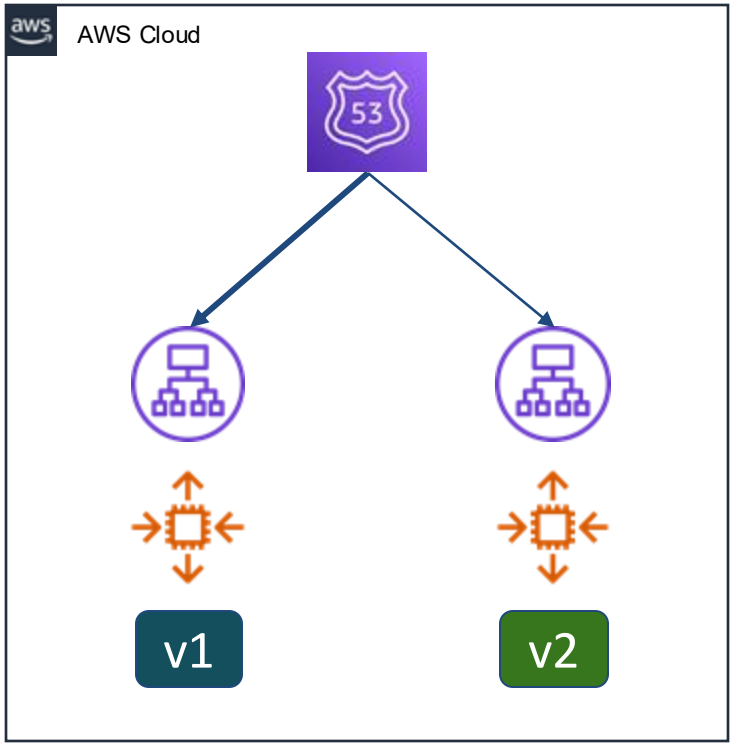
Deploy green ALB and ASG

Blue/Green Deployment Example 2



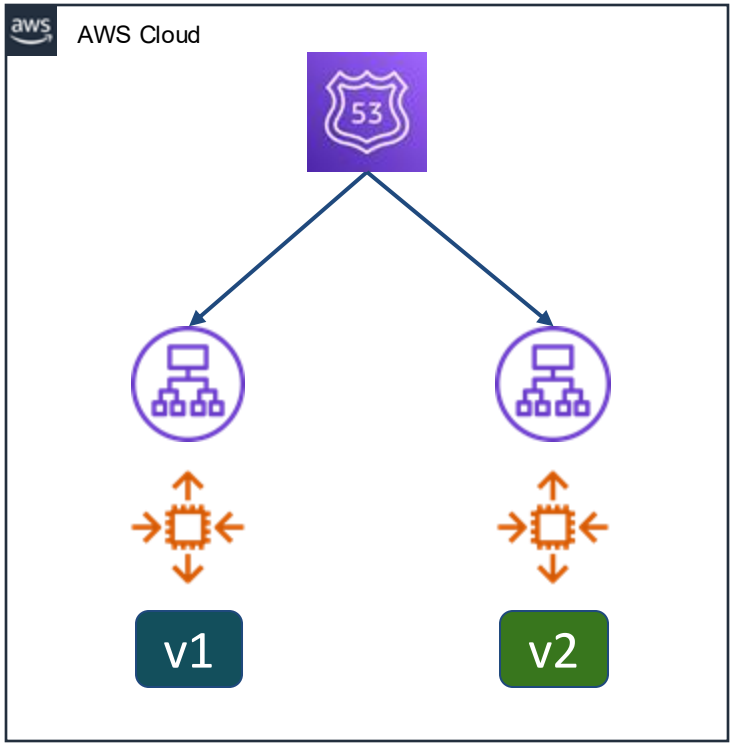
Add weighted routing record with 0 weight

Blue/Green Deployment Example 2



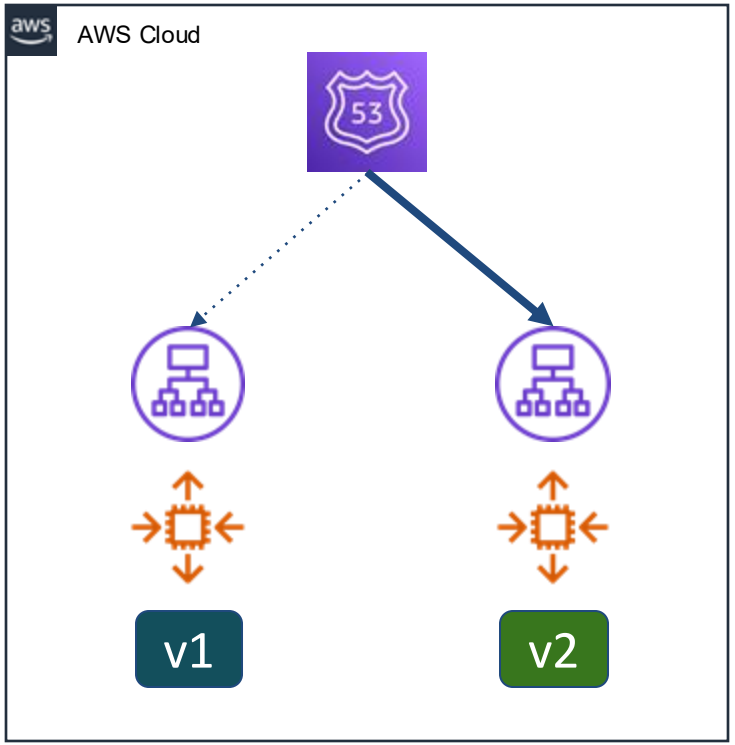
Update weights to 95 and 5

Blue/Green Deployment Example 2



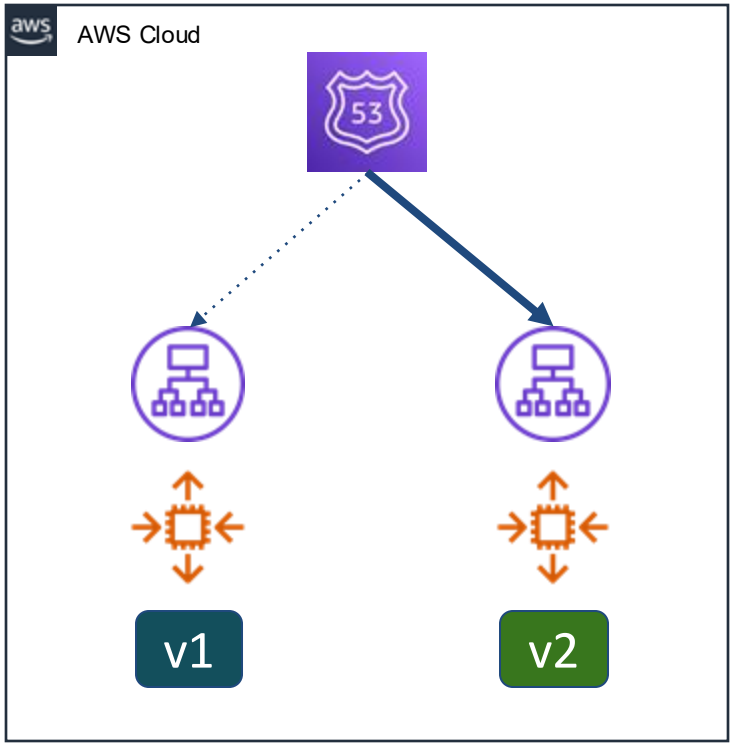
Gradually modify weights

Blue/Green Deployment Example 2



Transition until v2 takes all traffic

Blue/Green Deployment Example 2



De-provision v1

Question Breakdown

Manage change in automation

Question and Answer Choices

To improve reliability, you've recommended deploying an application using immutable infrastructure.

Which of the following is NOT a benefit of deploying application infrastructure with immutability?

A. Increased security
B. Simplified deployments
C. Increased performance
D. Increased scalability

Answer Choice Explanation

Enforcing replacement of infrastructure during every deployment limits the risk of compromise.

A. Increased security
B. Simplified deployments
C. Increased performance
D. Increased scalability

Answer Choice Explanation

Deployments become simpler when there are no inline updates to be performed or verified.

A. Increased security
B. Simplified deployments
C. Increased performance
D. Increased scalability

Answer Choice Explanation

While immutability has many advantages, there are no guarantees of improved performance.

A. Increased security
B. Simplified deployments
C. Increased performance
D. Increased scalability

Answer Choice Explanation

It is easier to scale by replacing infrastructure than to attempt inline updates, especially from a QA perspective.

A. Increased security
B. Simplified deployments
C. Increased performance
D. Increased scalability

Correct Answer

Correct Answer: C

A. Increased security
B. Simplified deployments
C. Increased performance
D. Increased scalability

Question Domain 3: Design High-Performing Architectures

Question Domain 3: Design High-Performing Architectures

Democratize advanced technologies

Principle Definition

Make advanced technology implementation easier for your team by delegating complex tasks to your cloud vendor.

Rather than asking your IT team to learn about hosting and running a new technology, consider consuming the technology as a service.

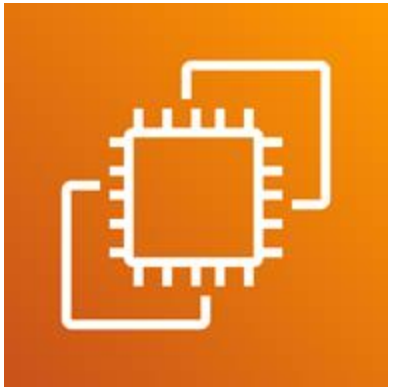
For example, NoSQL databases, media transcoding, and machine learning are all technologies that require specialized expertise.

In the cloud, these technologies become services that your team can consume, allowing your team to focus on product development rather than resource provisioning and management.

Delegate Complex Tasks Example 1

Requirement: deploy Tensorflow application using Docker containers

Tensorflow Deploy to EC2 Instance



- Launch EC2*
- Bootstrap OS
- Install prerequisites
- Deploy container*
- Configure instrumentation
- Monitor availability*

*Required for every resource

Tensorflow Deploy to ECS on Fargate

- Configure task definition
- Push Docker image to ECR
- Configure service
- Deploy task(s)
- Auto scale as required

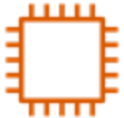
- No idle resources
- No repetitive tasks

Delegate Complex Tasks Example 2

Requirement: deploy MongoDB replica set for eCommerce shopping cart data

MongoDB Deploy to EC2

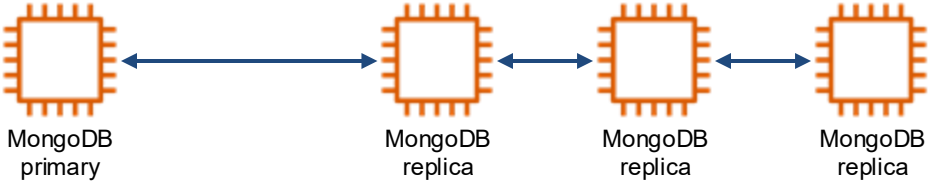
Deploy EC2 and install MongoDB, configure replica set



MongoDB primary

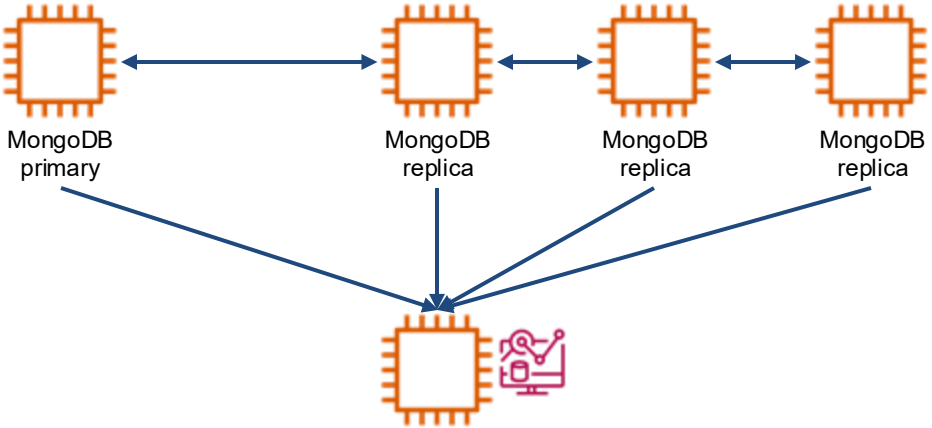
MongoDB Deploy to EC2

Deploy secondaries, manually add to replica set

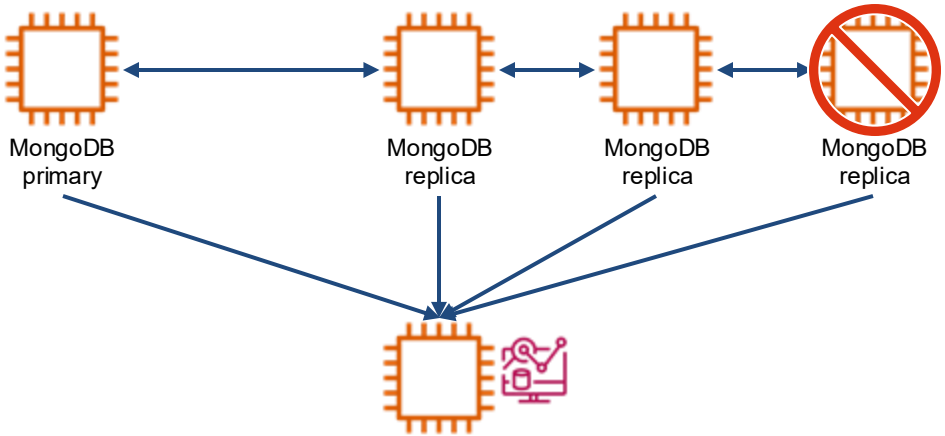


MongoDB Deploy to EC2

Manually configure performance and availability monitoring

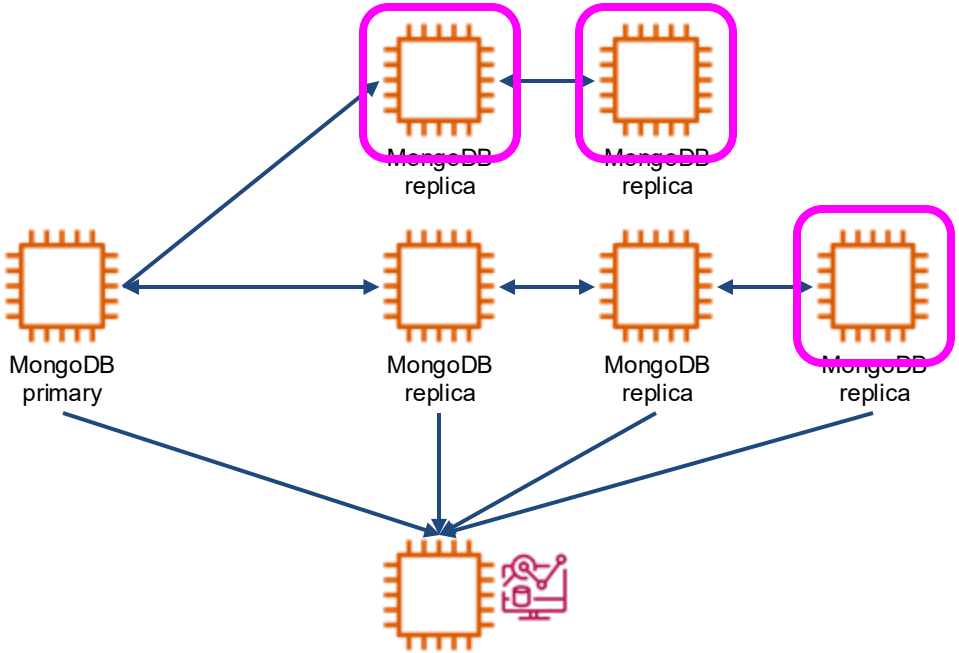


MongoDB Deploy to EC2



And when scaling or resource replacement is required?

MongoDB Deploy to EC2



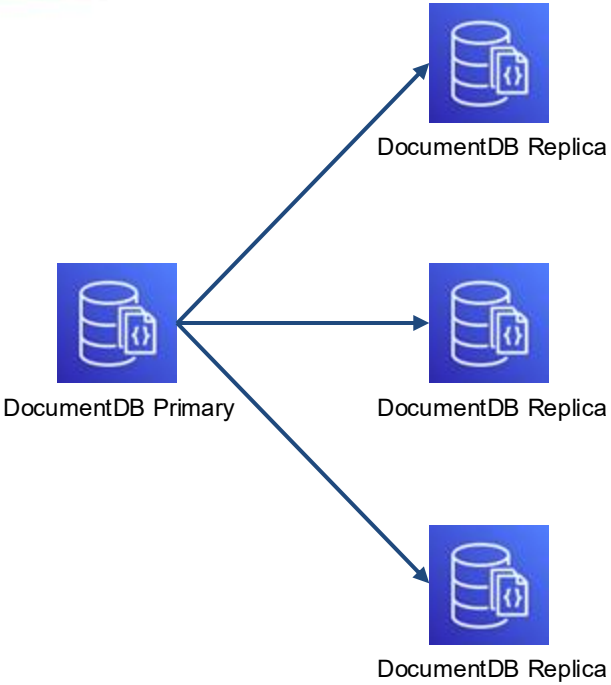
Replace and scale by hand as well

DocumentDB Deploy

Deploy DocumentDB cluster primary

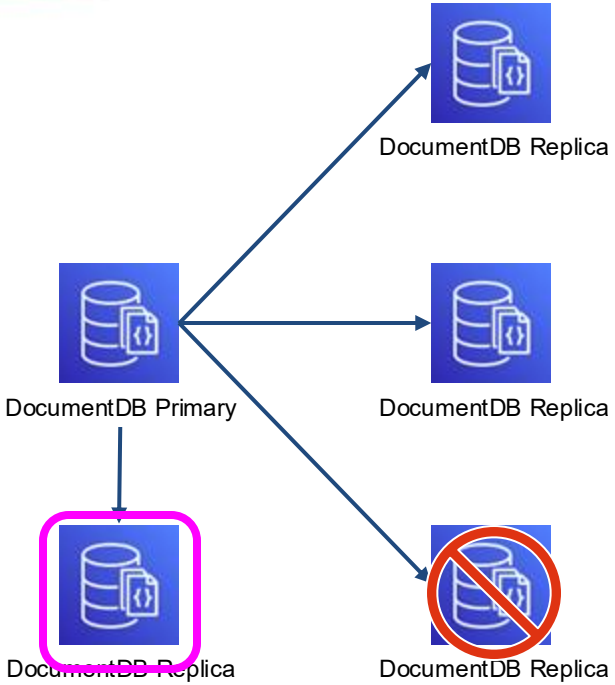
DocumentDB Primary

DocumentDB Deploy



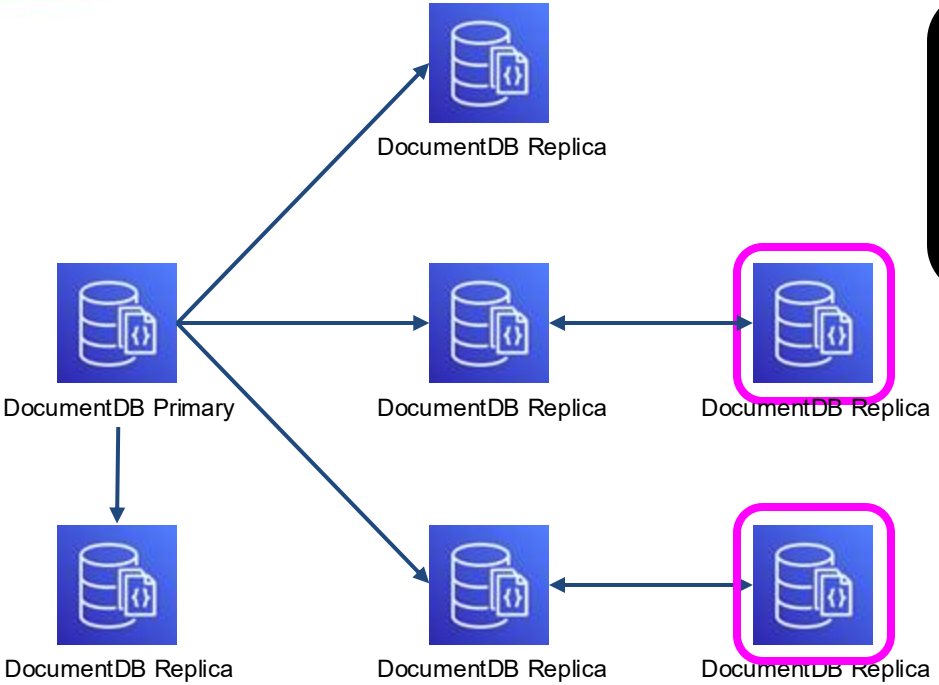
Deploy DocumentDB cluster replicas

DocumentDB Deploy



Replacement of failed instance is automatic

DocumentDB Deploy



Scaling is a simple operation as well, both vertical or horizontal

What KTLO tasks are minimized by managed services?

- Provisioning
- Scaling
- Monitoring
- Backup
- Failover
- Replication
- Disk space

This leads to more time being available for application performance tuning!

Question Breakdown

Democratize advanced technologies

Question and Answer Choices

As part of a new AWS application architecture, you've been asked to recommend a solution that can balance the needs of two requirements:

1) **Oracle Database software with minimum operational overhead**
2) **Access to the OS and software for regulatory purposes**

Which recommendation will meet both requirements?

A. **RDS with the Oracle Database engine selected**
B. **RDS with the Custom Engine selected**
C. **EC2 with a manual installation of Oracle Database**
D. **AWS does not have any service or feature that can meet both requirements**

Answer Choice Explanation

The RDS choice of the Oracle Database engine does not grant access to the underlying Operating System, so this would not meet the requirements.

A. **RDS with the Oracle Database engine selected**
B. **RDS with the Custom Engine selected**
C. **EC2 with a manual installation of Oracle Database**
D. **AWS does not have any service or feature that can meet both requirements**

Answer Choice Explanation

The Custom Engine option for RDS allows customers to access the underlying OS for either Oracle Database or Microsoft SQL Server, while providing many of the benefits of RDS as a managed service.

A. RDS with the Oracle Database engine selected
B. RDS with the Custom Engine selected
C. EC2 with a manual installation of Oracle Database
D. AWS does not have any service or feature that can meet both requirements

Answer Choice Explanation

While this is a functional solution, it does not address the requirement of low operational overhead.

A. RDS with the Oracle Database engine selected
B. RDS with the Custom Engine selected
C. EC2 with a manual installation of Oracle Database
D. AWS does not have any service or feature that can meet both requirements

Answer Choice Explanation

The RDS Custom Engine feature can indeed meet both requirements, and this is why it is important to stay current with AWS service and feature updates.

A. RDS with the Oracle Database engine selected
B. RDS with the Custom Engine selected
C. EC2 with a manual installation of Oracle Database
D. AWS does not have any service or feature that can meet both requirements

Correct Answer

Correct Answer: B

A. RDS with the Oracle Database engine selected
B. RDS with the Custom Engine selected
C. EC2 with a manual installation of Oracle Database
D. AWS does not have any service or feature that can meet both requirements

Question Domain 3: Design High-Performing Architectures

Go global in minutes

Principle Definition

Deploying your workload in multiple AWS Regions around the world allows you to provide lower latency and a better experience for your customers at minimal cost.

Multiple Region Content Access

Amazon CloudFront Origin Group (HTTP/HTTPS)

AWS Global Accelerator Endpoint Group (TCP/UDP)

S3 Multi Region Access Point

- Distribute requests to appropriate content origin
- Does not host content directly
- Favors active/active patterns

S3 Multiple Region Access

S3 Multiple Region Access

Replicate download here

S3 Multiple Region Access

S3 Multiple Region Access

Customer uses global access point and directed to closest download

Cross Region Resource Replication

RDS cross-region read replica

Aurora Global database

DocumentDB Global cluster

DynamoDB global table

Neptune Global database

- Primary region
- Replica regions
- Favors active/passive patterns

ElastiCache Global datastore

KMS Multi-region keys

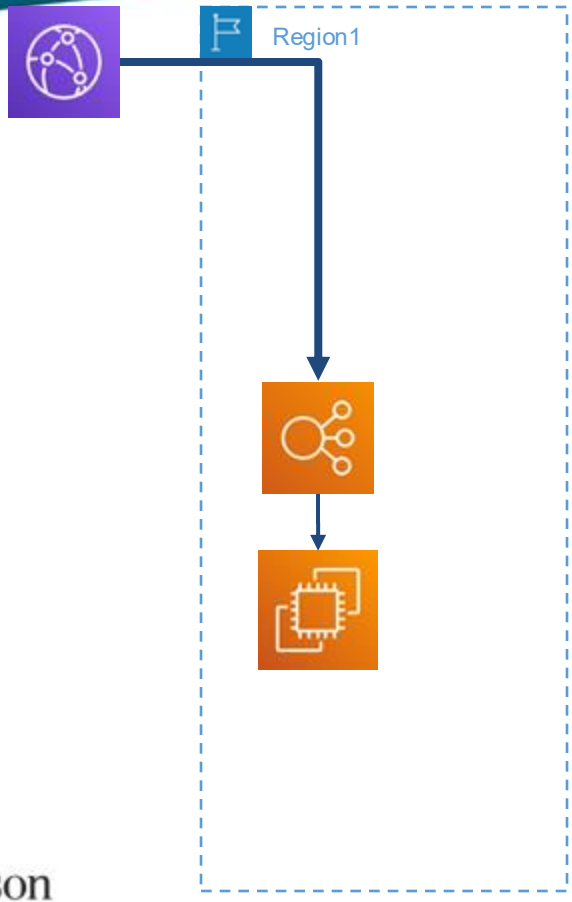
S3 Replication

Secrets Manager cross-region replica

Multiple Region Deploy Example

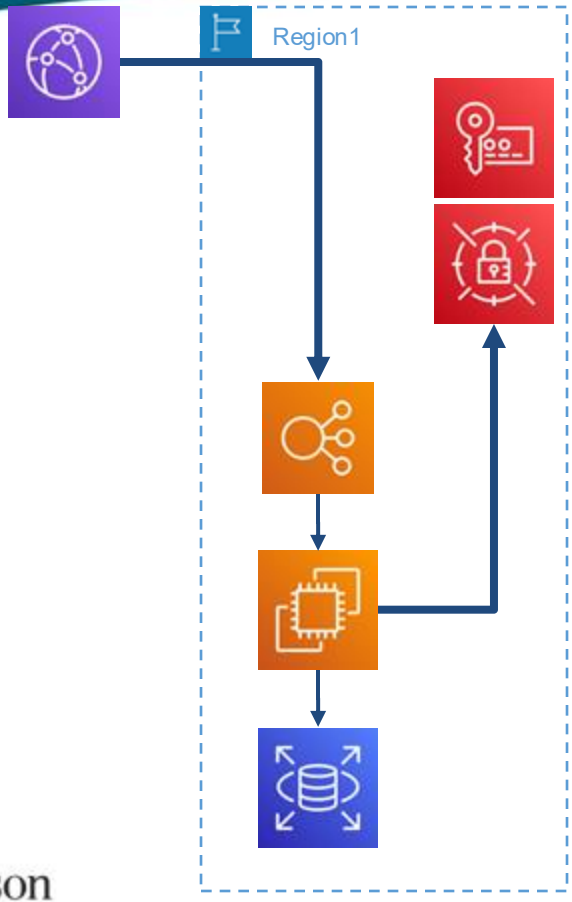
Deploy a CloudFront Distribution

Multiple Region Deploy Example



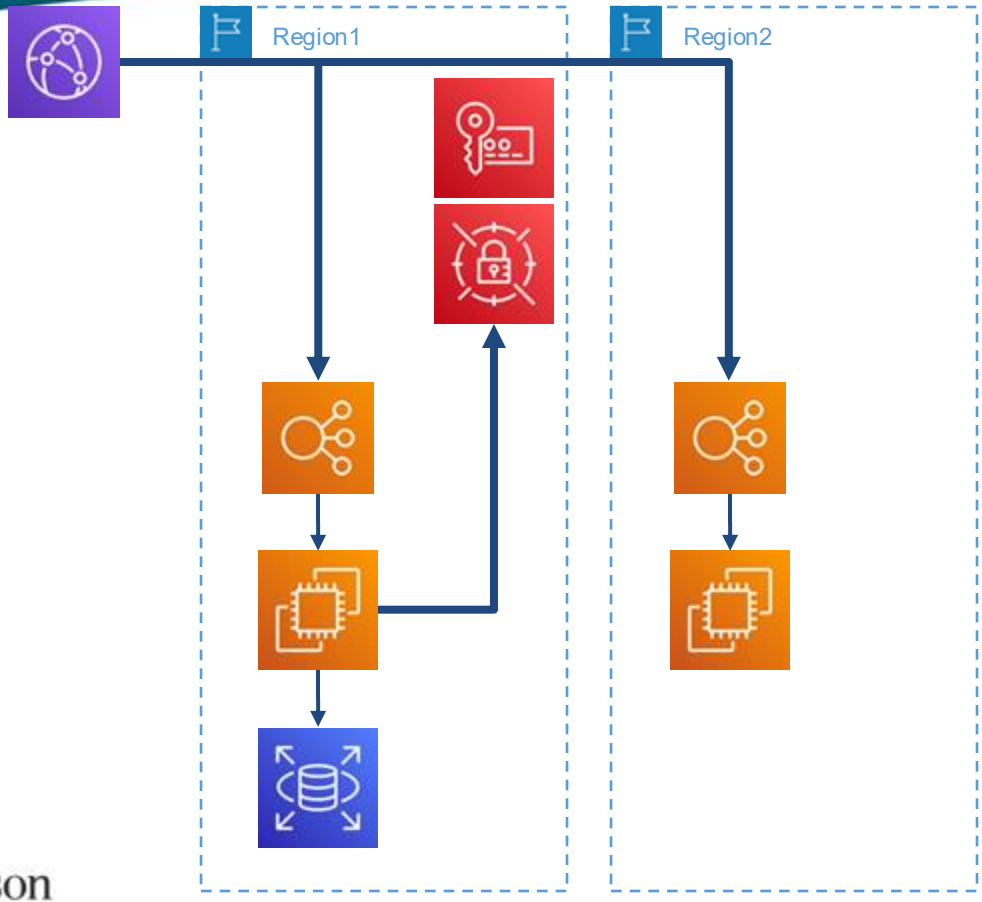
Deploy an ELB with an Auto Scaling group of EC2 instances running the application code

Multiple Region Deploy Example



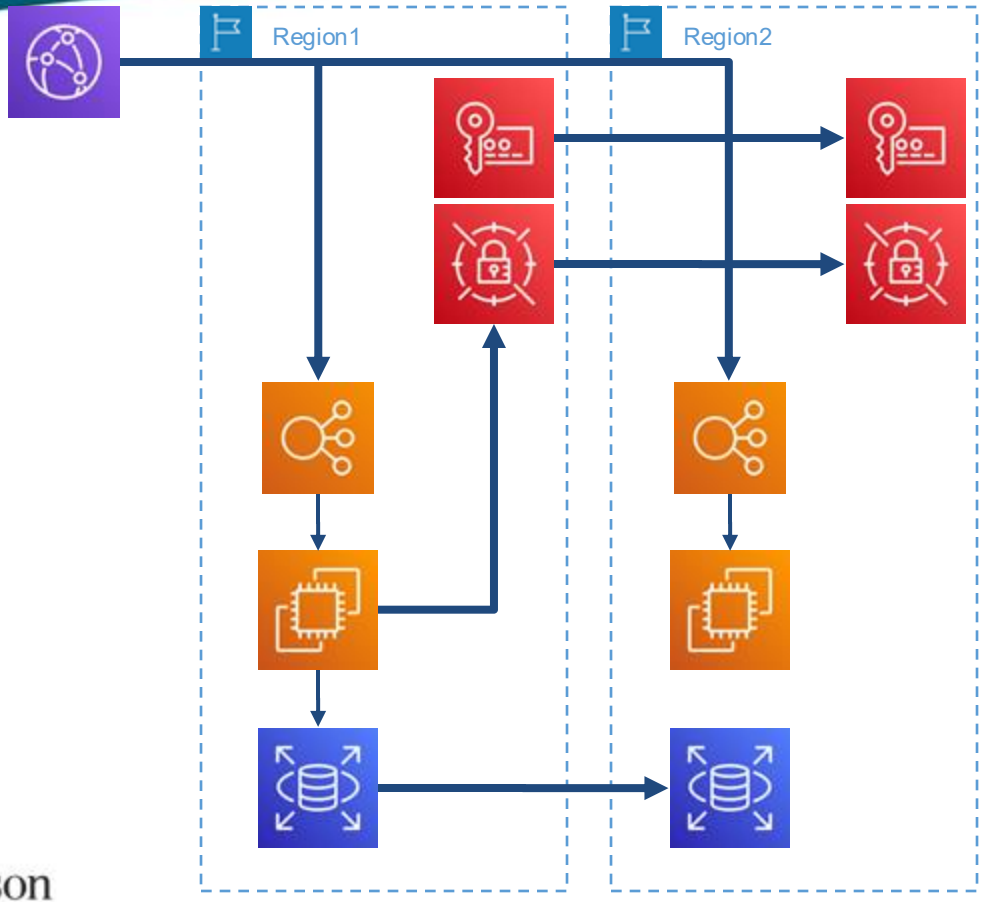
Deploy RDS, Secrets Manager, and KMS

Multiple Region Deploy Example



Deploy a parallel infrastructure into a second region and use CloudFront origin failover

Multiple Region Deploy Example



Replicate secrets, KMS CMK and database using native cross-region features

Replicate AMIs using DLM, AWS Backup, or EventBridge + Lambda

Question Breakdown

Go global in minutes

Question and Answer Choices

An IoT application is deployed using AWS IoT Core, using the MQTT protocol. There is a requirement to provide a single, static IP address, as DNS will not be used. The solution must emphasize high performance.

Which combination of steps will meet the requirements? (pick three)

A. Deploy an IoT Interface endpoint in a VPC within the same region as the IoT Core resource
B. Deploy an NLB in the VPC using the IoT Interface endpoint in a target group
C. Deploy an ALB in the VPC using the IoT Interface endpoint in a target group
D. Configure a Route 53 Alias record for the IoT Interface endpoint
E. Deploy a CloudFront distribution with the ALB endpoint in an origin group
F. Deploy a Global Accelerator with the NLB IP address in an origin group

Answer Choice Explanation

VPC Interface endpoints do indeed support the IoT Core resources, so this is a possible step.

A. Deploy an IoT Interface endpoint in a VPC within the same region as the IoT Core resource
B. Deploy an NLB in the VPC using the IoT Interface endpoint in a target group
C. Deploy an ALB in the VPC using the IoT Interface endpoint in a target group
D. Configure a Route 53 Alias record for the IoT Interface endpoint
E. Deploy a CloudFront distribution with the ALB endpoint in an origin group
F. Deploy a Global Accelerator with the NLB IP address in an origin group

Answer Choice Explanation

This step assumes that A was already performed, and is also a possibility, as NLB->Interface is a primary implementation pattern for PrivateLink.

A. Deploy an IoT Interface endpoint in a VPC within the same region as the IoT Core resource
B. Deploy an NLB in the VPC using the IoT Interface endpoint in a target group
C. Deploy an ALB in the VPC using the IoT Interface endpoint in a target group
D. Configure a Route 53 Alias record for the IoT Interface endpoint
E. Deploy a CloudFront distribution with the ALB endpoint in an origin group
F. Deploy a Global Accelerator with the NLB IP address in an origin group

Answer Choice Explanation

This step also assumes that A was completed, and is also a supported configuration.

A. Deploy an IoT Interface endpoint in a VPC within the same region as the IoT Core resource
B. Deploy an NLB in the VPC using the IoT Interface endpoint in a target group
C. Deploy an ALB in the VPC using the IoT Interface endpoint in a target group
D. Configure a Route 53 Alias record for the IoT Interface endpoint
E. Deploy a CloudFront distribution with the ALB endpoint in an origin group
F. Deploy a Global Accelerator with the NLB IP address in an origin group

Answer Choice Explanation

This step ignores the requirement to use a static IP address, and therefore does not work as a solution.

A. Deploy an IoT Interface endpoint in a VPC within the same region as the IoT Core resource
B. Deploy an NLB in the VPC using the IoT Interface endpoint in a target group
C. Deploy an ALB in the VPC using the IoT Interface endpoint in a target group
D. Configure a Route 53 Alias record for the IoT Interface endpoint
E. Deploy a CloudFront distribution with the ALB endpoint in an origin group
F. Deploy a Global Accelerator with the NLB IP address in an origin group

Answer Choice Explanation

This step is functionally possible but ignores the static IP address requirement.

A. Deploy an IoT Interface endpoint in a VPC within the same region as the IoT Core resource
B. Deploy an NLB in the VPC using the IoT Interface endpoint in a target group
C. Deploy an ALB in the VPC using the IoT Interface endpoint in a target group
D. Configure a Route 53 Alias record for the IoT Interface endpoint
E. Deploy a CloudFront distribution with the ALB endpoint in an origin group
F. Deploy a Global Accelerator with the NLB IP address in an origin group

Answer Choice Explanation

The Global Accelerator provisions a static AnyCast IP for the endpoint, which can indeed be used in the client IoT application configuration.

A. Deploy an IoT Interface endpoint in a VPC within the same region as the IoT Core resource
B. Deploy an NLB in the VPC using the IoT Interface endpoint in a target group
C. Deploy an ALB in the VPC using the IoT Interface endpoint in a target group
D. Configure a Route 53 Alias record for the IoT Interface endpoint
E. Deploy a CloudFront distribution with the ALB endpoint in an origin group
F. Deploy a Global Accelerator with the NLB IP address in an origin group

Correct Answer

Correct Answers: A,B,F

A. Deploy an IoT Interface endpoint in a VPC within the same region as the IoT Core resource
B. Deploy an NLB in the VPC using the IoT Interface endpoint in a target group
C. Deploy an ALB in the VPC using the IoT Interface endpoint in a target group
D. Configure a Route 53 Alias record for the IoT Interface endpoint
E. Deploy a CloudFront distribution with the ALB endpoint in an origin group
F. Deploy a Global Accelerator with the NLB IP address in an origin group

Question Domain 3: Design High-Performing Architectures

Use serverless architectures

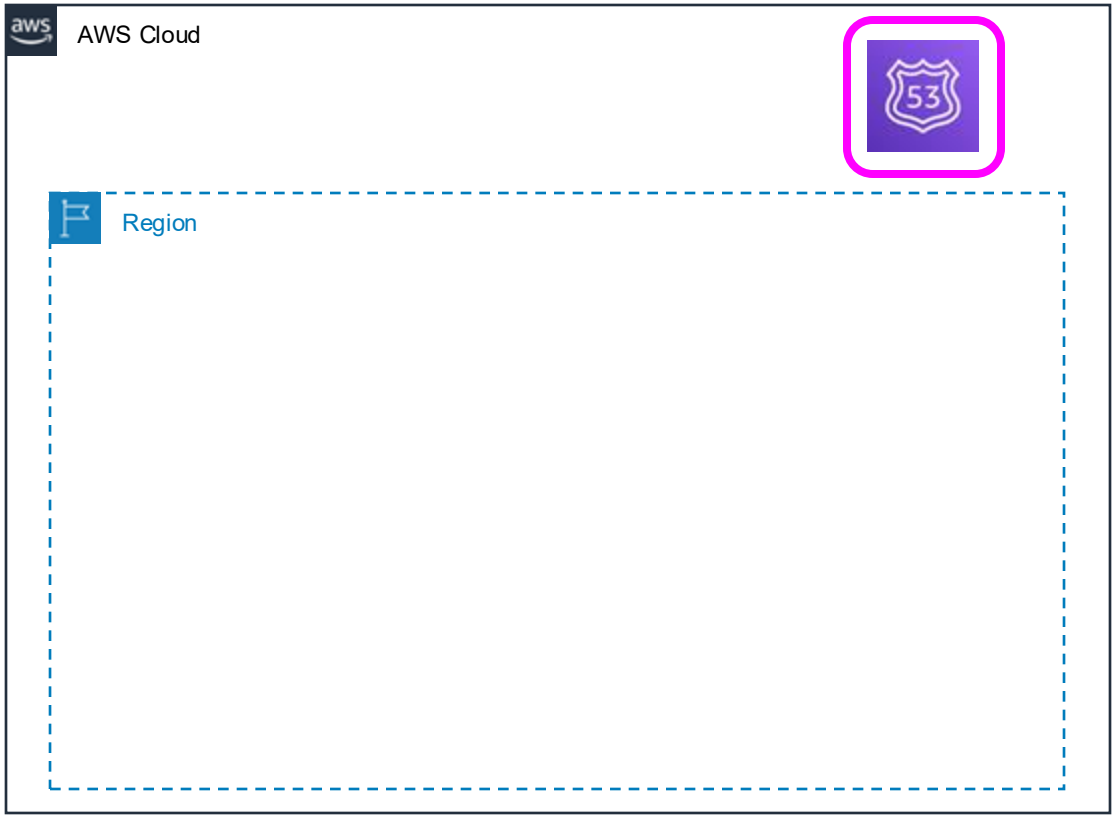
Principle Definition

Serverless architectures remove the need for you to run and maintain physical servers for traditional compute activities.

For example, serverless storage services can act as static websites (removing the need for web servers) and event services can host code.

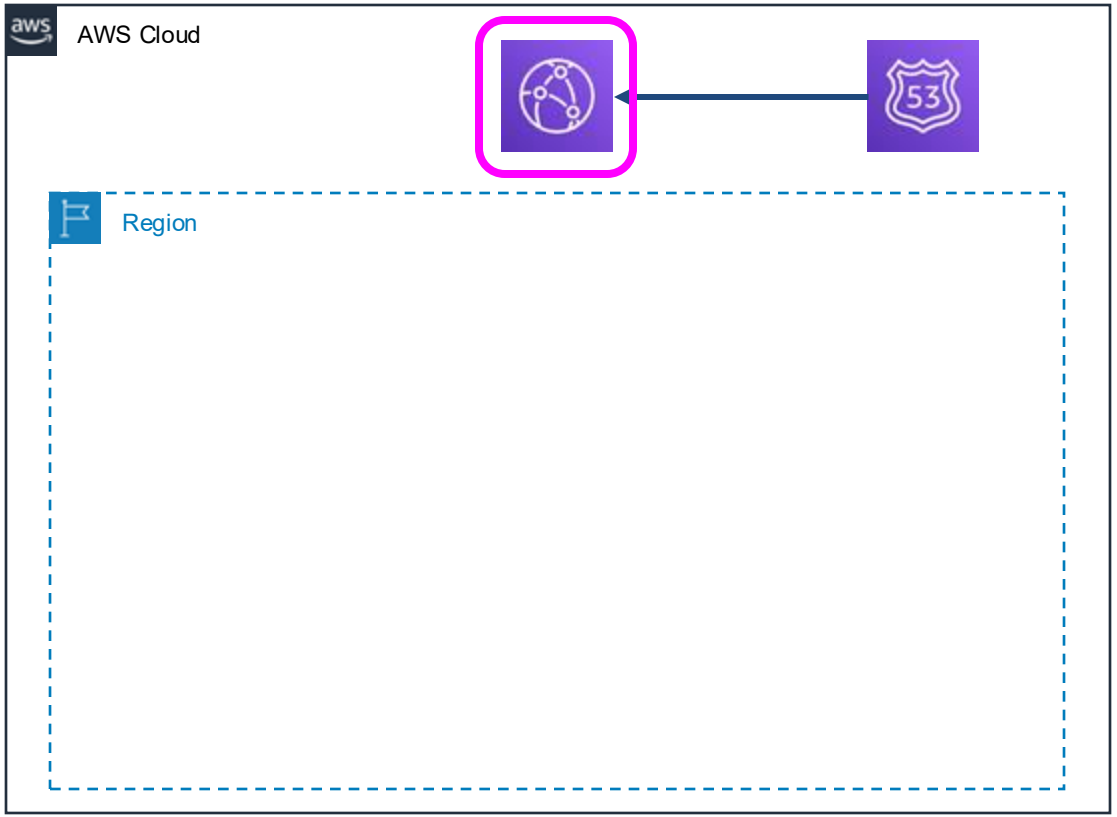
This removes the operational burden of managing physical servers, and can lower transactional costs because managed services operate at cloud scale.

Multi-Tier Serverless



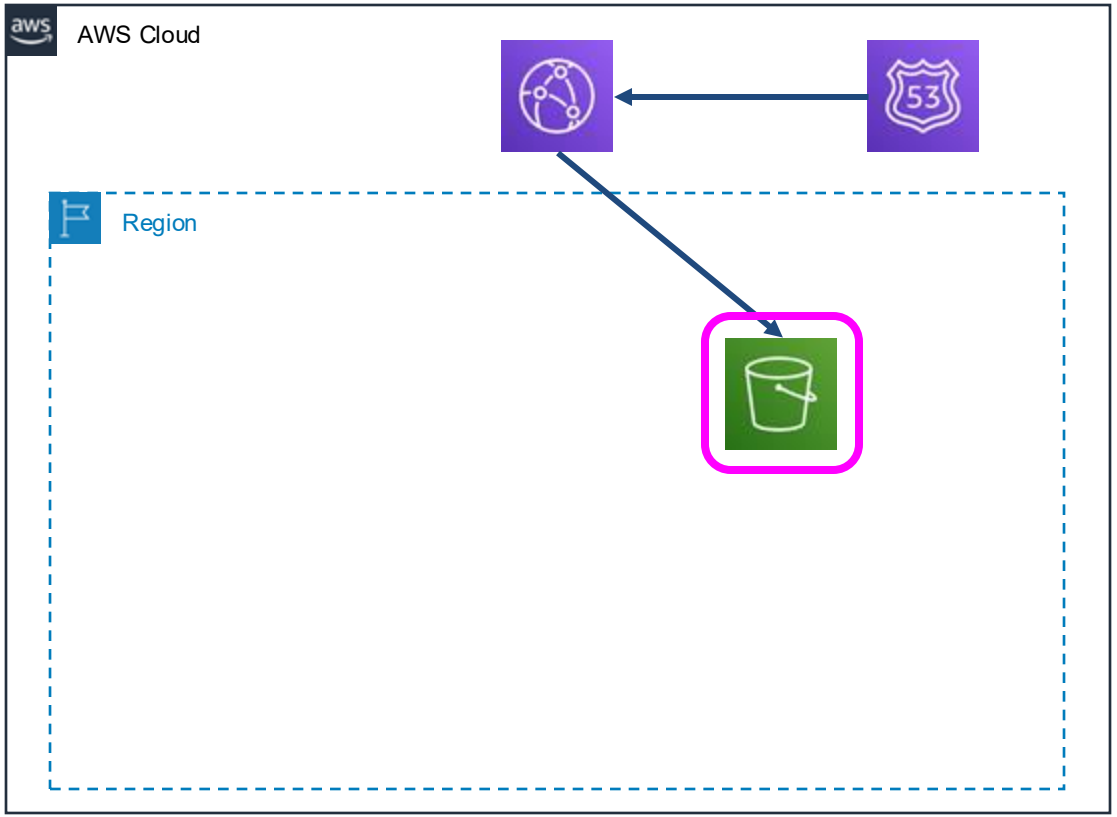
Route 53 uses Edge Locations for extreme fault tolerance of DNS lookups

Multi-Tier Serverless



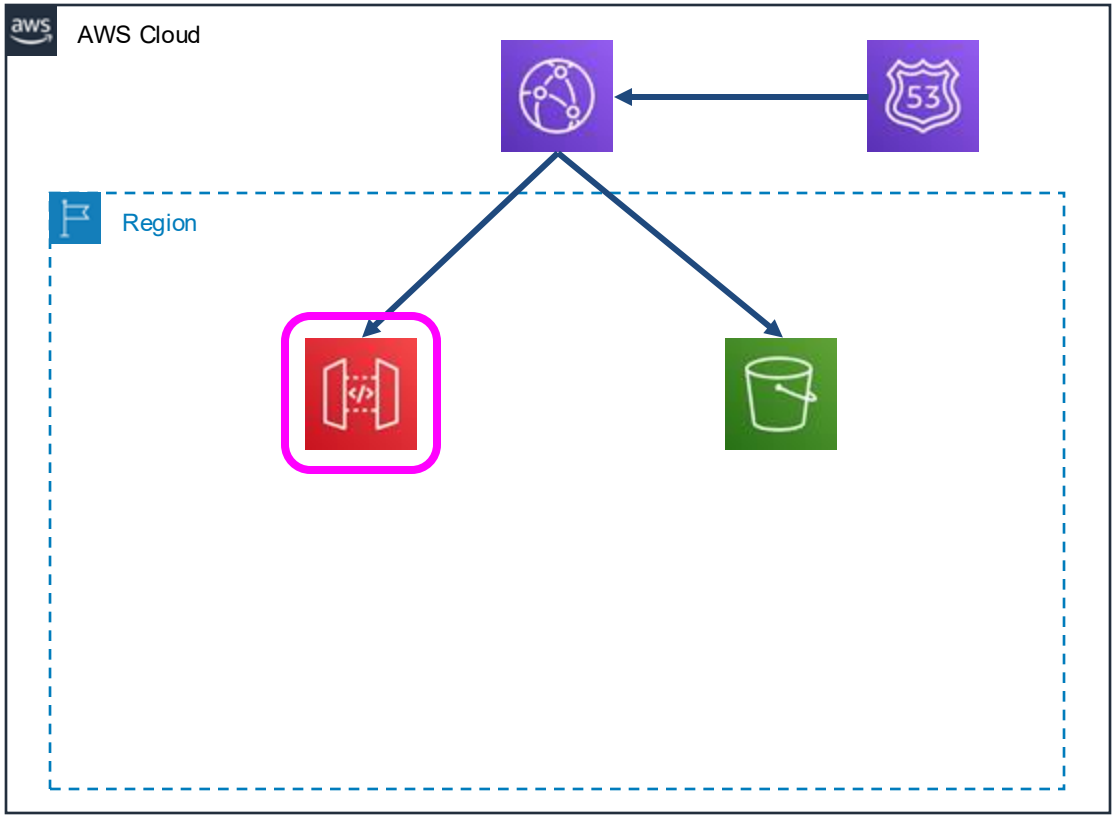
CloudFront also uses Edge Locations for fault tolerant caching and content delivery

Multi-Tier Serverless



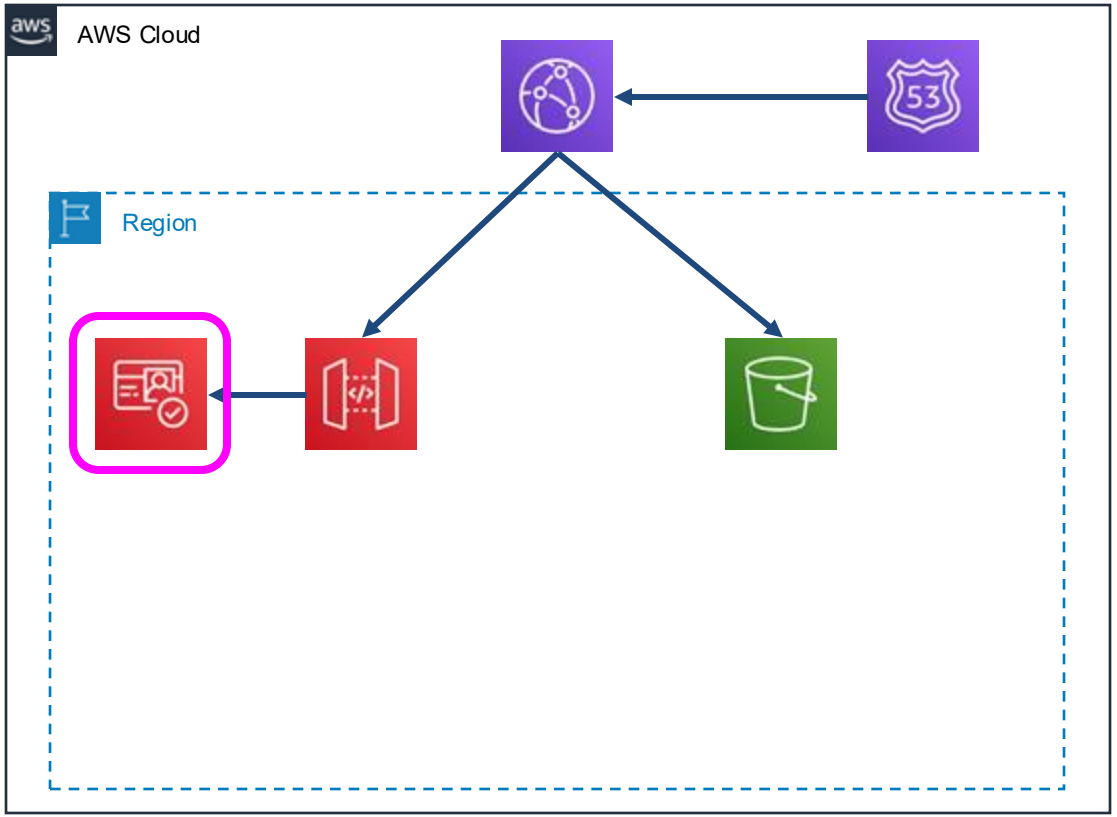
S3 is the static content origin

Multi-Tier Serverless



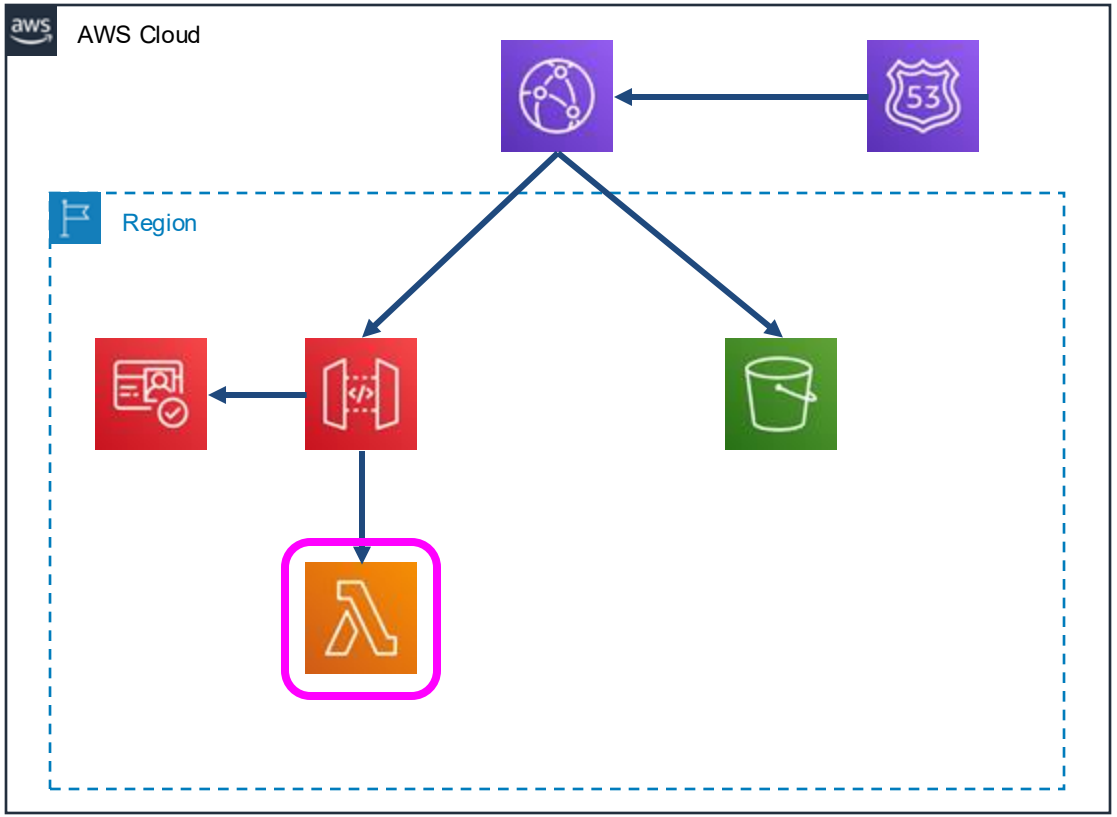
API Gateway is the dynamic front end

Multi-Tier Serverless



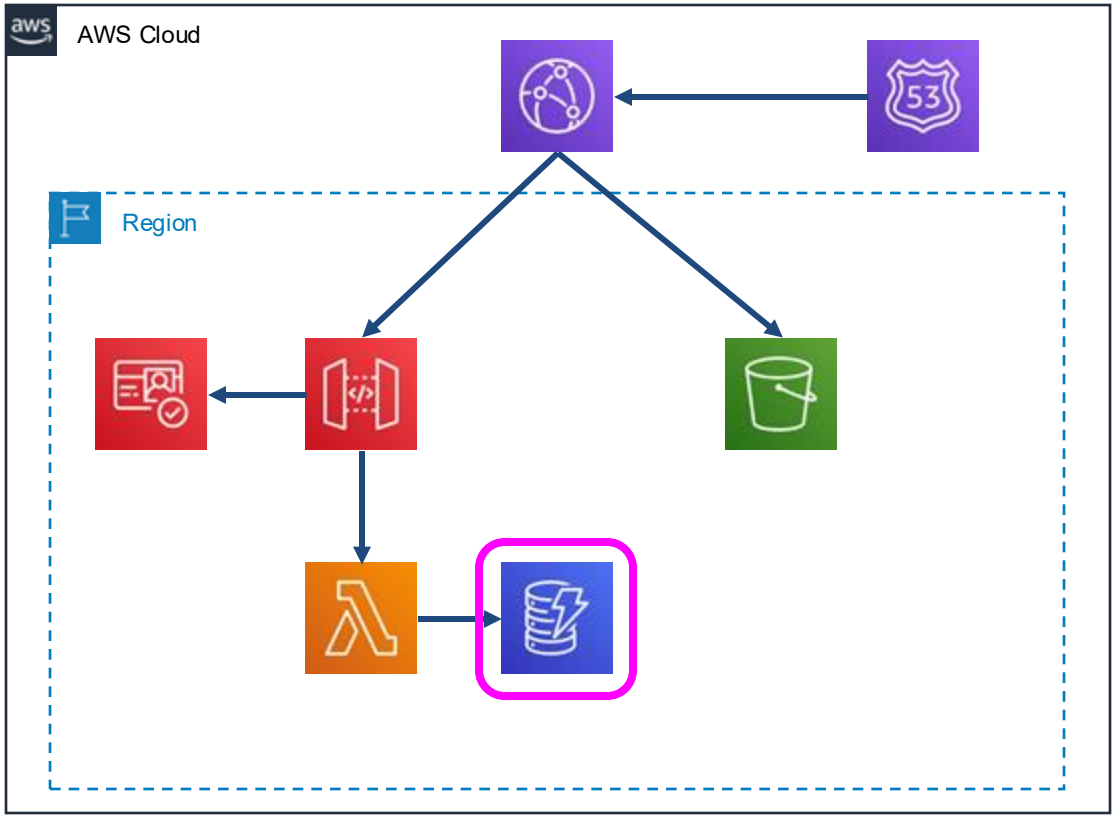
Cognito is used for authentication

Multi-Tier Serverless



Lambda offers business logic runtime execution

Multi-Tier Serverless



DynamoDB offers NoSQL data persistence and performance

Question Breakdown

Use serverless architectures

Question and Answer Choices

A workflow requires a Lambda function to be invoked as a target of an EventBridge rule, and will rarely be invoked more than once at the same time. The Lambda function must execute as quickly as possible as a highest priority, but also must optimize for cost.

What recommendations should be made to meet these requirements? (pick three)

A. Configure the Lambda function with more memory
B. Configure the Lambda function with less memory
C. Configure the Lambda function with Provisioned Concurrency at a low number
D. Configure the Lambda function with a Provisioned Concurrency at a high number
E. Configure the Lambda function with a timeout at the maximum execution time
F. Configure the Lambda function with a timeout at 95th percentile of maximum execution time

Answer Choice Explanation

More memory = more vCPU, which improves performance, meeting the top priority.

A. **Configure the Lambda function with more memory**
B. **Configure the Lambda function with less memory**
C. **Configure the Lambda function with Provisioned Concurrency at a low number**
D. **Configure the Lambda function with a Provisioned Concurrency at a high number**
E. **Configure the Lambda function with a timeout at the maximum execution time**
F. **Configure the Lambda function with a timeout at 95th percentile of maximum execution time**

Answer Choice Explanation

This optimizes for cost, but not for performance.

A. Configure the Lambda function with more memory
B. Configure the Lambda function with less memory
C. Configure the Lambda function with Provisioned Concurrency at a low number
D. Configure the Lambda function with a Provisioned Concurrency at a high number
E. Configure the Lambda function with a timeout at the maximum execution time
F. Configure the Lambda function with a timeout at 95th percentile of maximum execution time

Answer Choice Explanation

Provisioned Concurrency ensures there is a prewarmed copy of the function ready to execute, which reduces the startup latency, thus decreasing execution time and improving performance. A low number here would match the expected performance profile.

A. Configure the Lambda function with more memory
B. Configure the Lambda function with less memory
C. Configure the Lambda function with Provisioned Concurrency at a low number
D. Configure the Lambda function with a Provisioned Concurrency at a high number
E. Configure the Lambda function with a timeout at the maximum execution time
F. Configure the Lambda function with a timeout at 95th percentile of maximum execution time

Answer Choice Explanation

Like C, this improves performance, but since the function is not executed concurrently often, this is a cost risk.

A. Configure the Lambda function with more memory
B. Configure the Lambda function with less memory
C. Configure the Lambda function with Provisioned Concurrency at a low number
D. Configure the Lambda function with a Provisioned Concurrency at a high number
E. Configure the Lambda function with a timeout at the maximum execution time
F. Configure the Lambda function with a timeout at 95th percentile of maximum execution time

Answer Choice Explanation

This would optimize the reliability of the function, but would end up costing more for failed function executions.

A. Configure the Lambda function with more memory
B. Configure the Lambda function with less memory
C. Configure the Lambda function with Provisioned Concurrency at a low number
D. Configure the Lambda function with a Provisioned Concurrency at a high number
E. Configure the Lambda function with a timeout at the maximum execution time
F. Configure the Lambda function with a timeout at 95th percentile of maximum execution time

Answer Choice Explanation

This is generally a good recommendation to optimize for cost by assuming that the high runtime outliers are likely in an error condition, so why pay for that extra execution time?

A. Configure the Lambda function with more memory
B. Configure the Lambda function with less memory
C. Configure the Lambda function with Provisioned Concurrency at a low number
D. Configure the Lambda function with a Provisioned Concurrency at a high number
E. Configure the Lambda function with a timeout at the maximum execution time
F. Configure the Lambda function with a timeout at 95th percentile of maximum execution time

Correct Answer

Correct Answers: A,C,F

A. Configure the Lambda function with more memory
B. Configure the Lambda function with less memory
C. Configure the Lambda function with Provisioned Concurrency at a low number
D. Configure the Lambda function with a Provisioned Concurrency at a high number
E. Configure the Lambda function with a timeout at the maximum execution time
F. Configure the Lambda function with a timeout at 95th percentile of maximum execution time

Question Domain 3: Design High-Performing Architectures

Experiment more often

Principle Definition

With virtual and automatable resources, you can quickly carry out comparative testing using different types of instances, storage, or configurations.

Automated Performance Testing Challenge 1

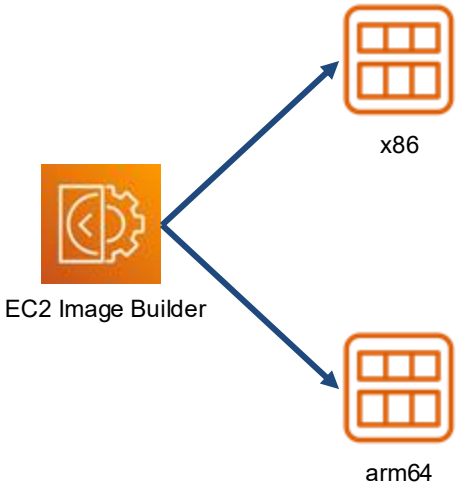
Establish application performance baseline on candidate EC2 instance types for analysis and decision

Evaluate Compute Options For Performance

Implement pipeline to bootstrap images with appropriate software baseline

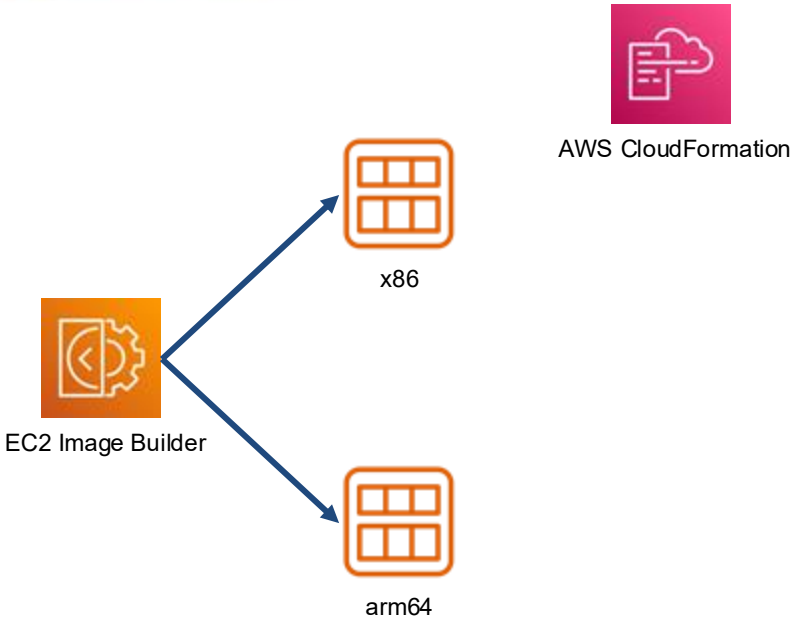
EC2 Image Builder

Evaluate Compute Options For Performance



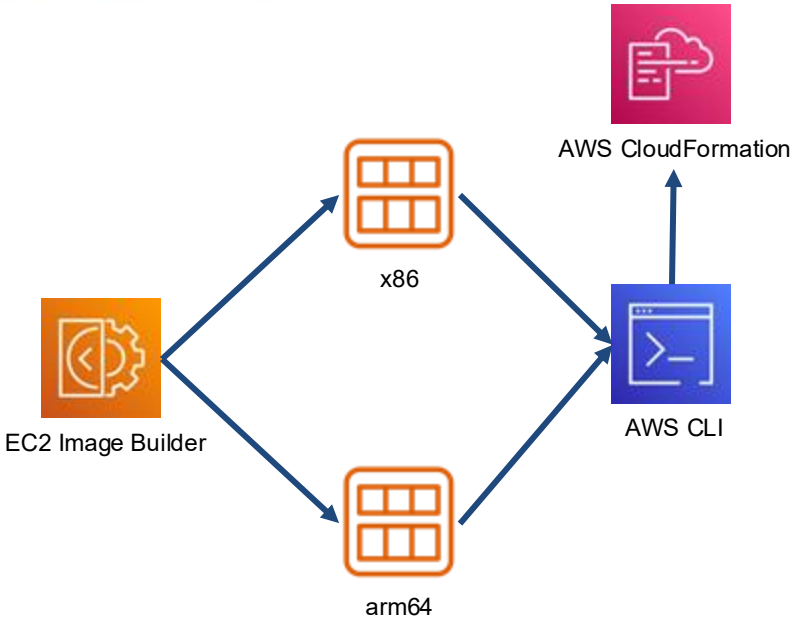
Deploy cpu-appropriate images to match candidate instance types

Evaluate Compute Options For Performance



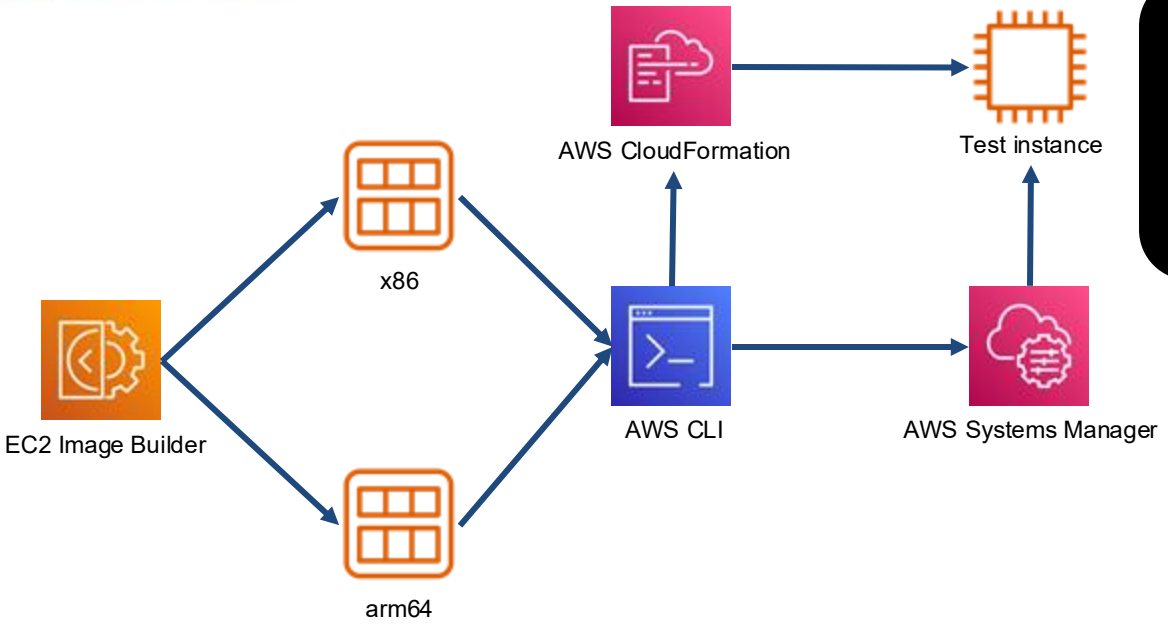
Compose CloudFormation template with mapping for instance AMIs

Evaluate Compute Options For Performance



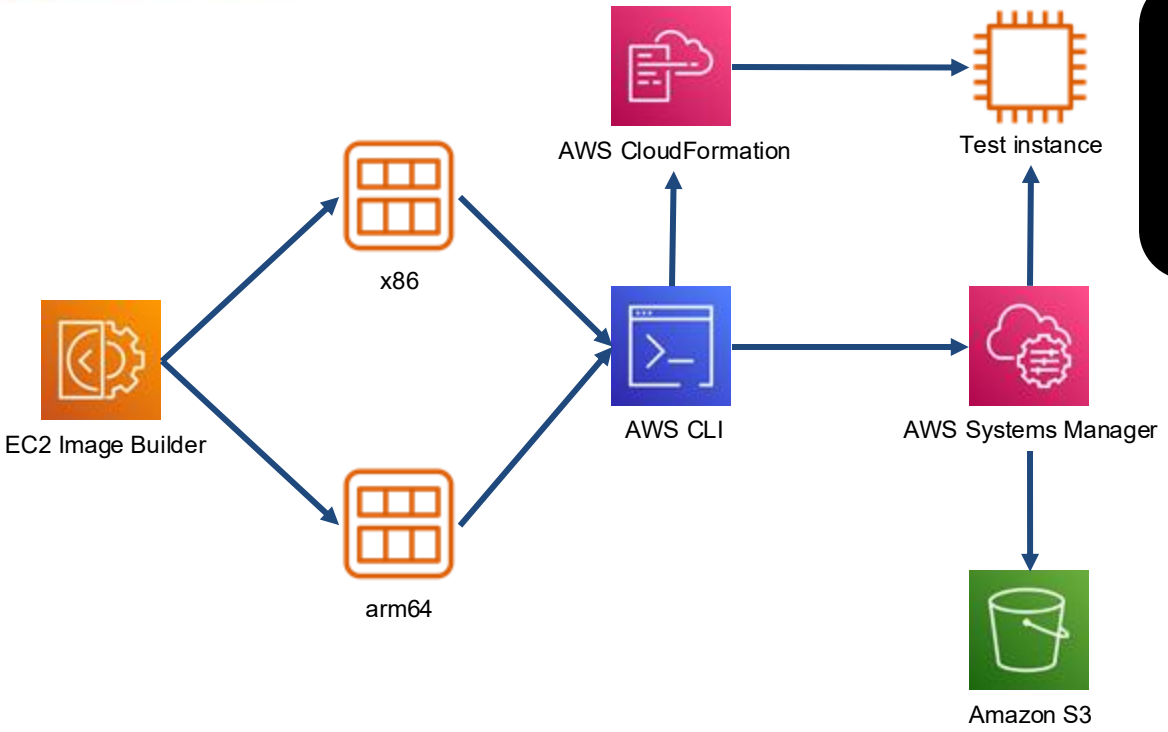
Compose shell script to loop through instance type launches using template

Evaluate Compute Options For Performance



During loop, execute performance baseline test with SM Run command

Evaluate Compute Options For Performance



Store results in S3 for later analysis and decisions

Automated Performance Testing Challenge 2

Establish throughput and IOPS performance baseline on candidate EBS volume types for analysis and decision

Extra Credit: view performance test results in near real time

Bonus: execute parallel testing if possible!

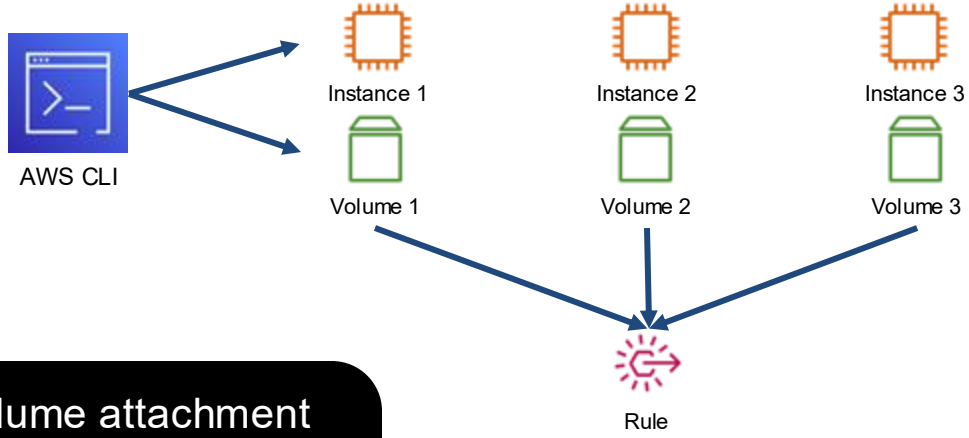
Evaluate Storage Options For Performance

Compose a shell script to launch multiple identical EC2 instances

Evaluate Storage Options For Performance

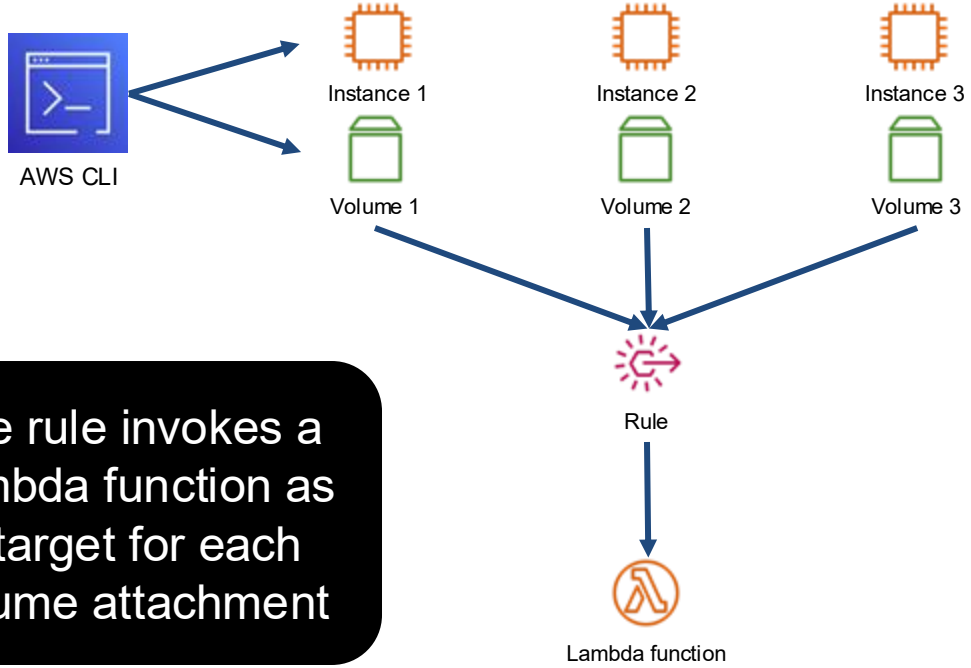
Use the same script to attach a different EBS volume type to each instance

Evaluate Storage Options For Performance



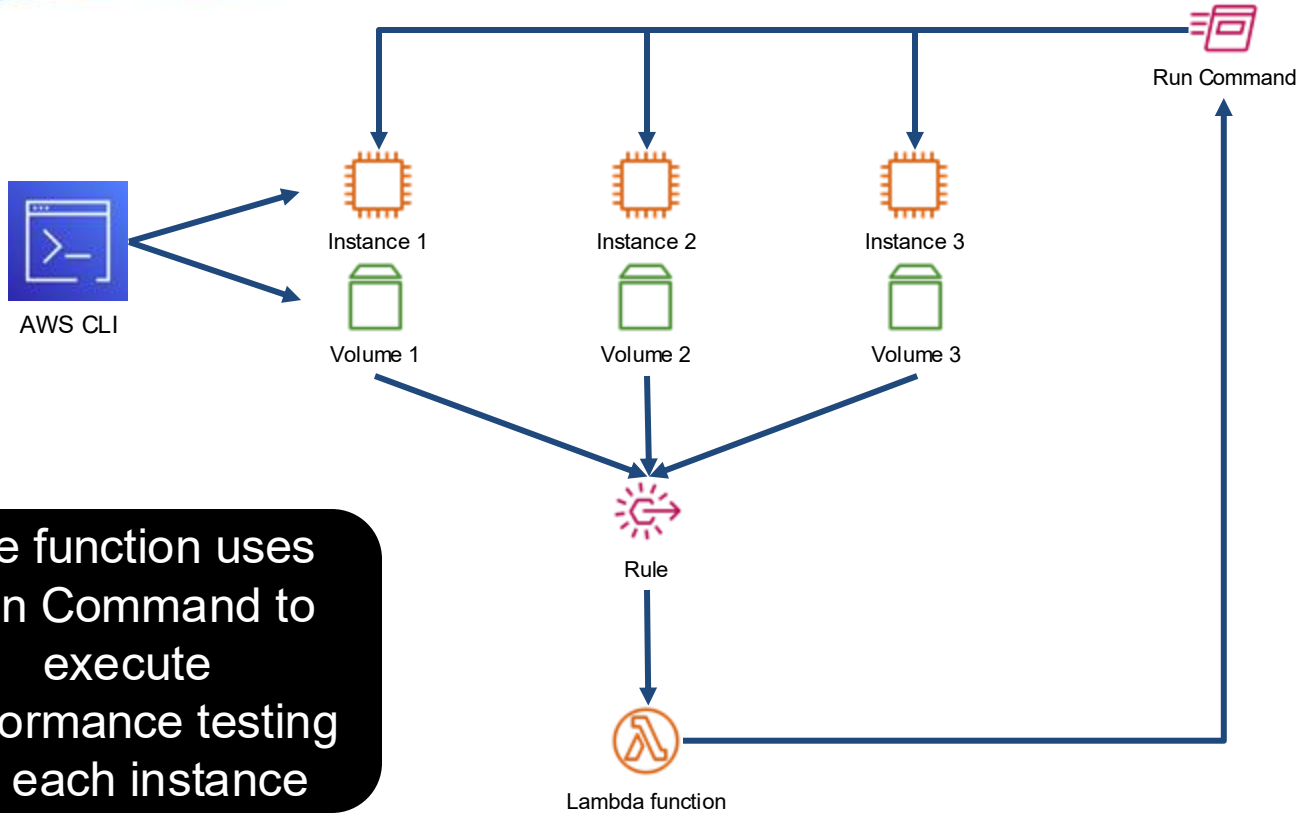
Volume attachment generates a CloudTrail log, which is captured by an EventBridge Rule

Evaluate Storage Options For Performance



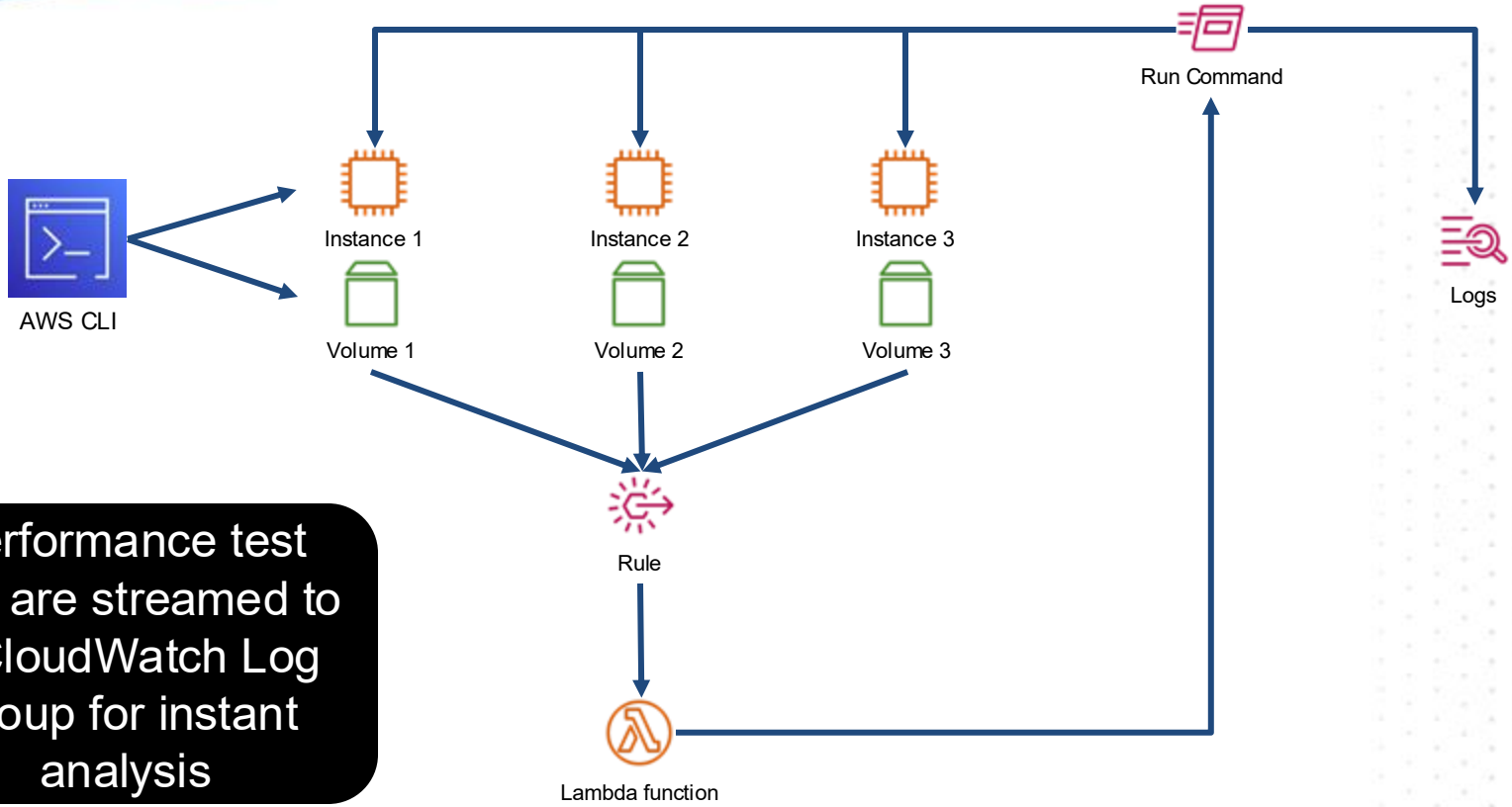
The rule invokes a Lambda function as a target for each volume attachment

Evaluate Storage Options For Performance



The function uses Run Command to execute performance testing on each instance

Evaluate Storage Options For Performance



Performance test logs are streamed to a CloudWatch Log group for instant analysis

Question Breakdown

Experiment more often

Question and Answer Choices

An application has a requirement for high storage IOPS, in order to accommodate parallel reads and writes of many small data files (approximately 4Kb each).

Which storage solution would NOT be appropriate for a direct performance test of the workload?

A. FSx for Lustre provisioned with default configuration
B. EBS Provisioned IOPS io2 Block Express volume
C. EC2 instance provisioned with an NVMe SSD Instance Store volume
D. EFS provisioned with General Purpose performance mode

Answer Choice Explanation

FSx is widely regarded as high-performing with low latency. However, the default configuration will place all files below 1GiB in the same stripe, reducing the overall performance capacity.

A. FSx for Lustre provisioned with default configuration
B. EBS Provisioned IOPS io2 Block Express volume
C. EC2 instance provisioned with an NVMe SSD Instance Store volume
D. EFS provisioned with General Purpose performance mode

Answer Choice Explanation

The io2 Block Express volume is the next generation of EBS and supports up to 256,000 provisioned IOPS at 16k write blocks.

A. FSx for Lustre provisioned with default configuration
B. EBS Provisioned IOPS io2 Block Express volume
C. EC2 instance provisioned with an NVMe SSD Instance Store volume
D. EFS provisioned with General Purpose performance mode

Answer Choice Explanation

NVMe SSD is very high-performing for IOPS (at 4k write blocks, matching the data size) and throughput, but dependent on the size of the EC2 instance. There will be no network-related latency for this storage option.

A. FSx for Lustre provisioned with default configuration
B. EBS Provisioned IOPS io2 Block Express volume
C. EC2 instance provisioned with an NVMe SSD Instance Store volume
D. EFS provisioned with General Purpose performance mode

Answer Choice Explanation

General Purpose performance mode actually has lower per-operation latency than Max I/O, which is why it was selected as a candidate for the test.

A. FSx for Lustre provisioned with default configuration
B. EBS Provisioned IOPS io2 Block Express volume
C. EC2 instance provisioned with an NVMe SSD Instance Store volume
D. EFS provisioned with General Purpose performance mode

Correct Answer

Correct Answer: A

A. FSx for Lustre provisioned with default configuration
B. EBS Provisioned IOPS io2 Block Express volume
C. EC2 instance provisioned with an NVMe SSD Instance Store volume
D. EFS provisioned with General Purpose performance mode

Question Domain 3: Design High-Performing Architectures

Consider mechanical sympathy

Principle Definition

Use the technology approach that aligns best with your goals. For example, consider data access patterns when you select database or storage approaches.

Mechanical Sympathy Considerations - Database

Data shape

The type of data being stored, such as NoSQL or relational

Mechanical Sympathy Considerations - Database

Data structure

Within the shape, more specific parameters such as key/value or graph data

Mechanical Sympathy Considerations - Database

Data size

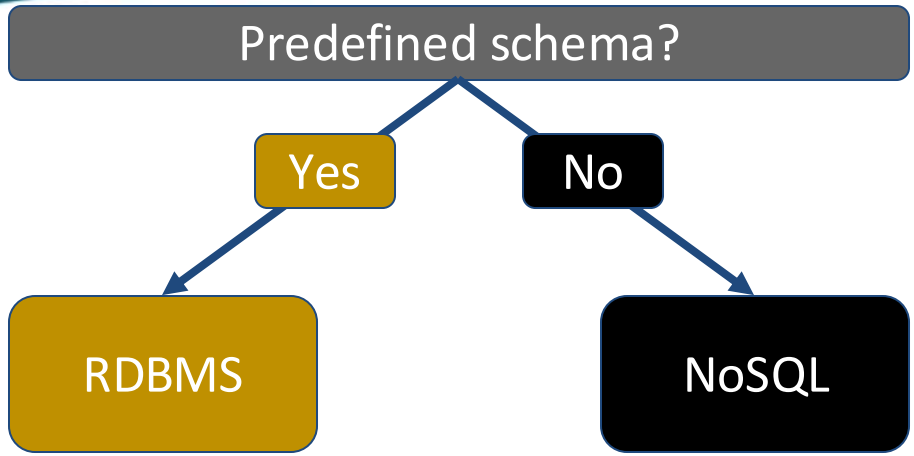
How many individual records, and how much data is contained in each record

Mechanical Sympathy Considerations - Database

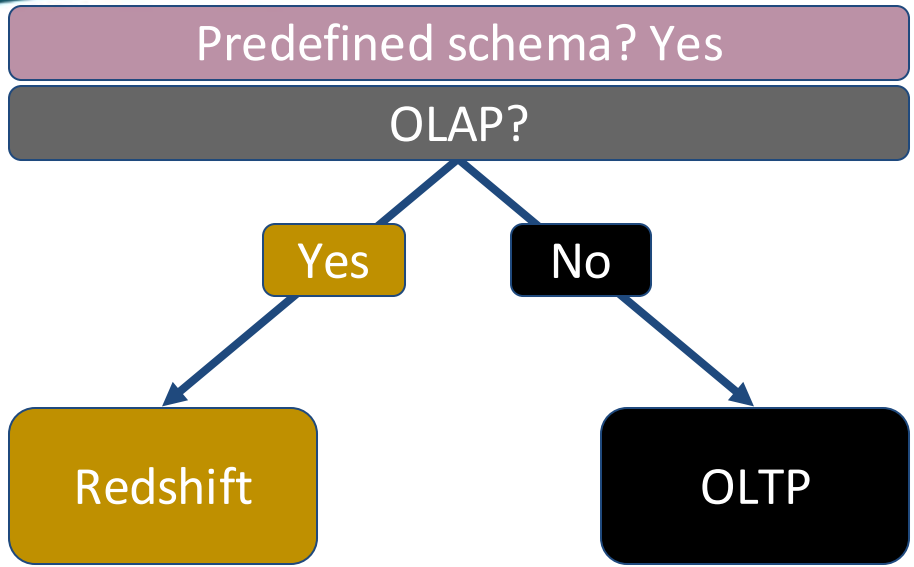
Access patterns

Ratio of reads vs writes, and amounts of each

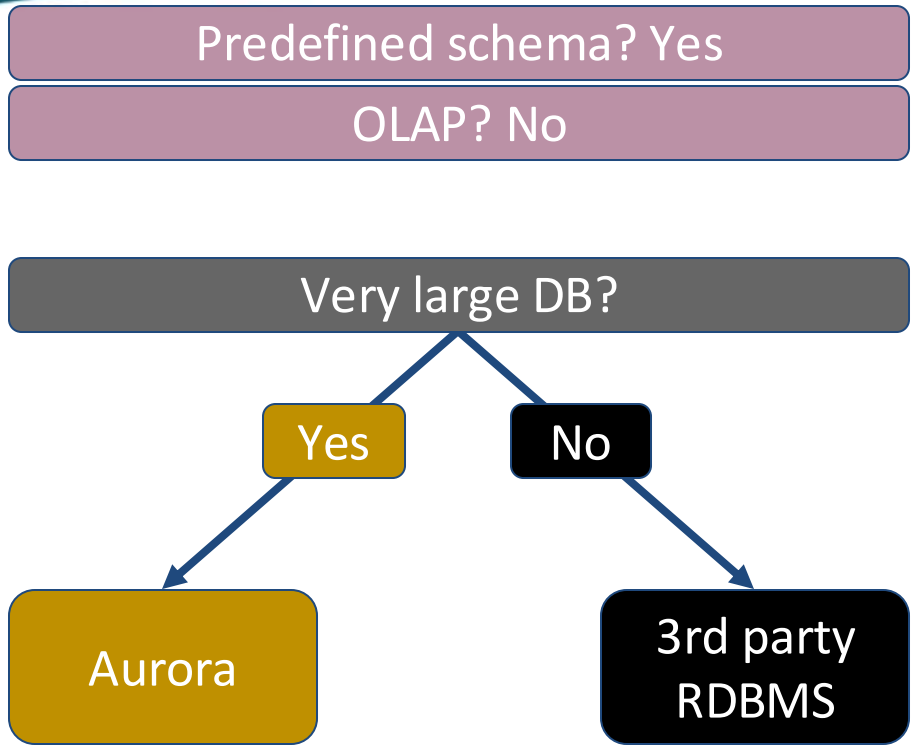
Data Structure Decision Tree



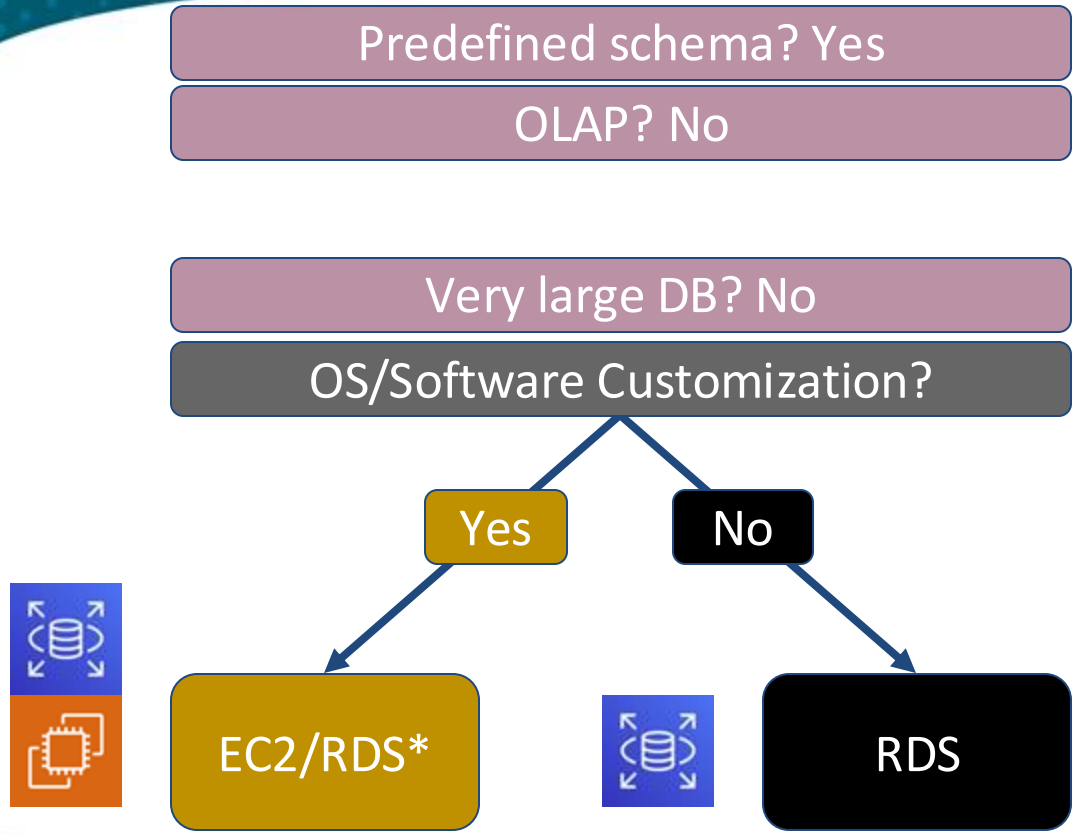
Data Structure Decision Tree



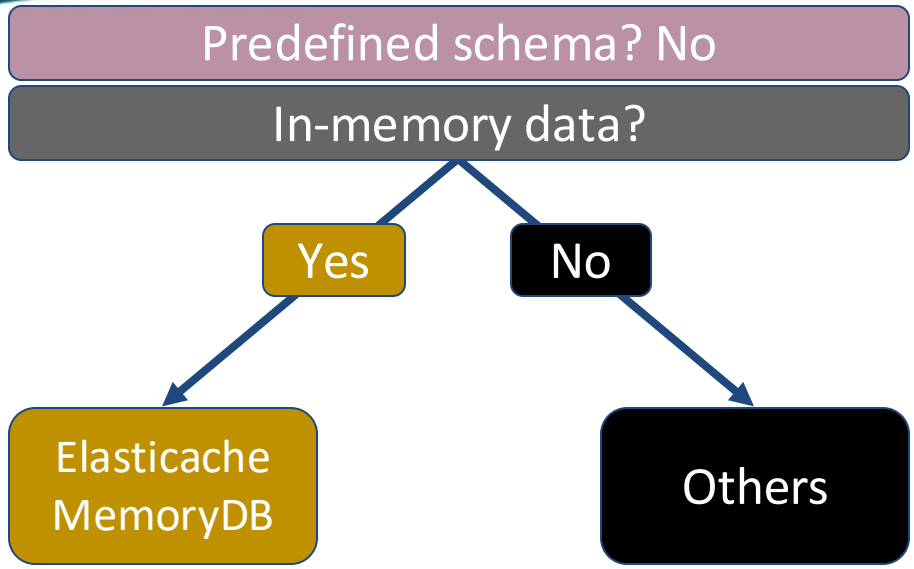
Data Structure Decision Tree



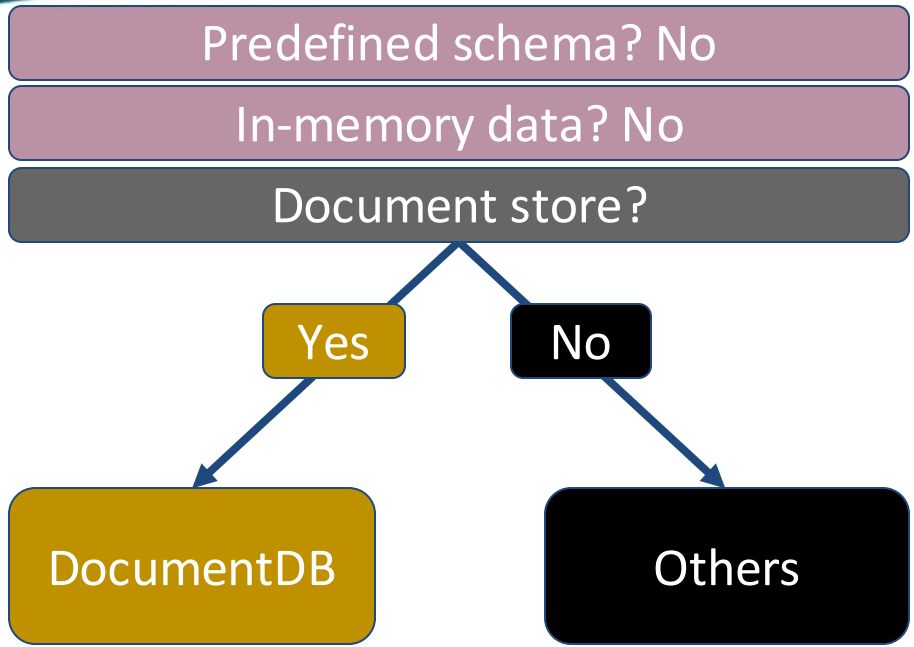
Data Structure Decision Tree



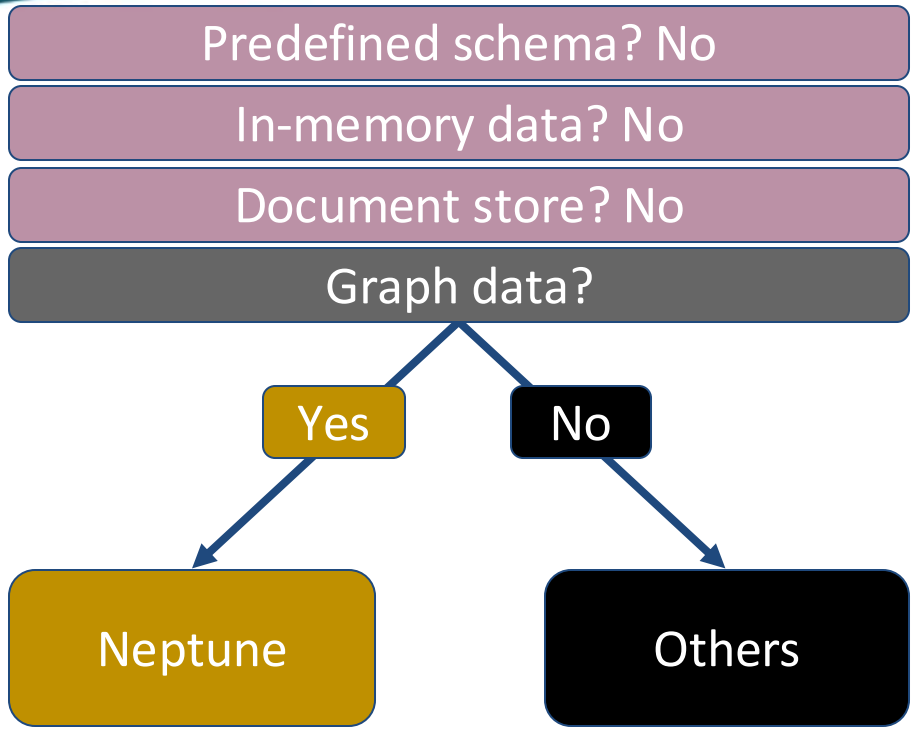
Data Structure Decision Tree



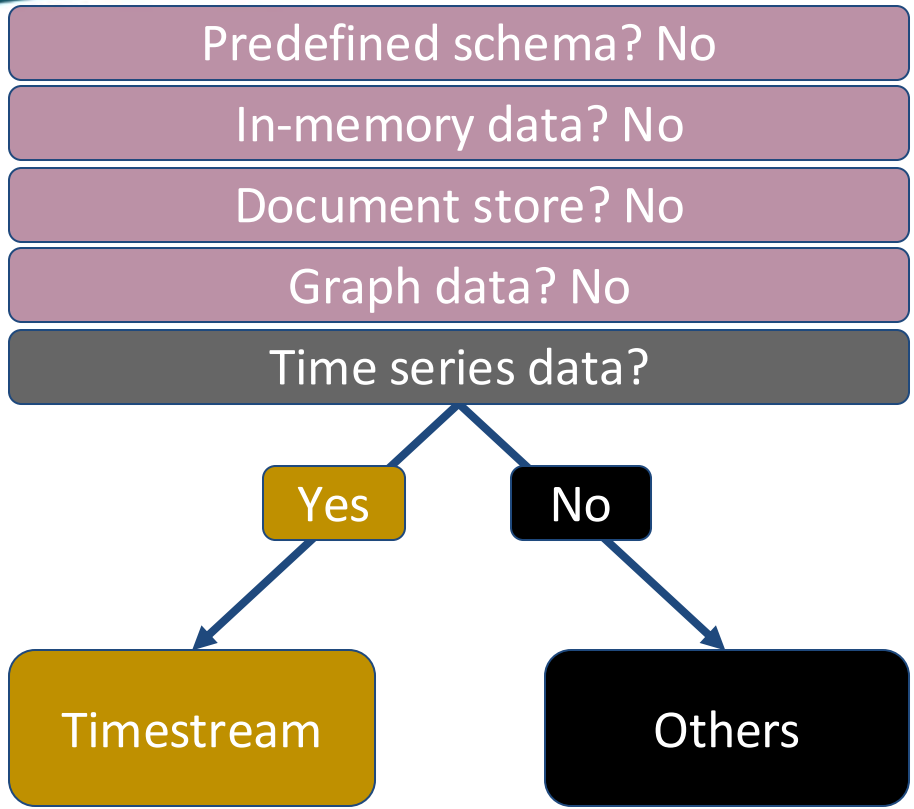
Data Structure Decision Tree



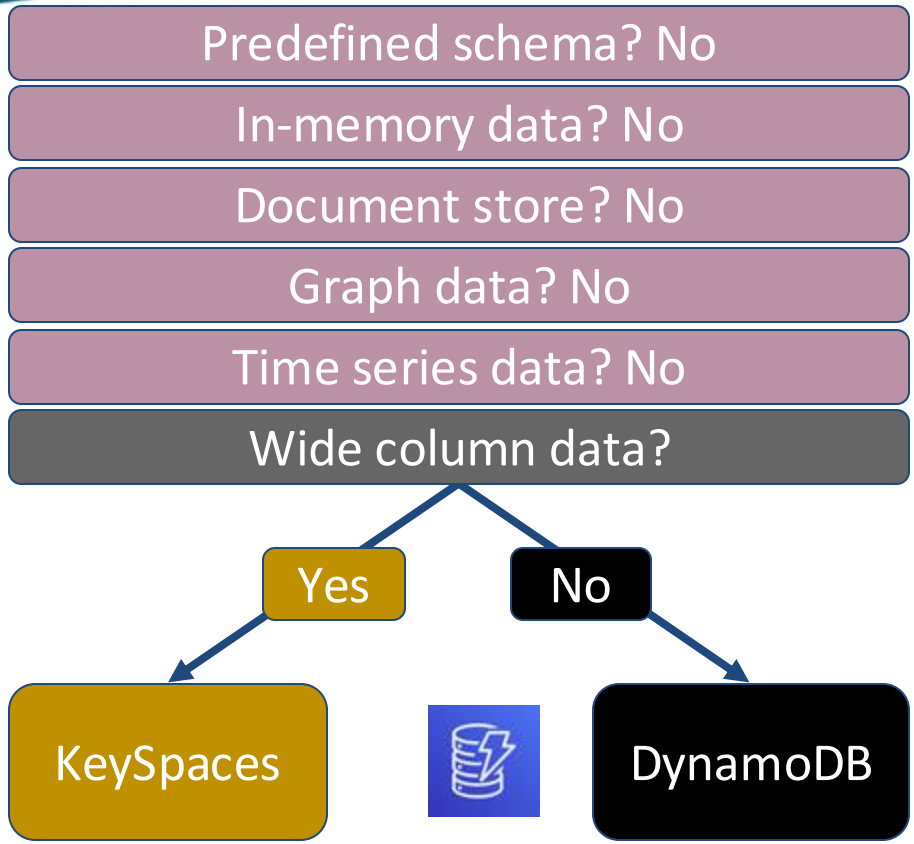
Data Structure Decision Tree



Data Structure Decision Tree



Data Structure Decision Tree



Question Breakdown

Consider mechanical sympathy

Question and Answer Choices

A gaming company is developing a high-performance, real-time multiplayer game. The game requires low latency and high throughput to ensure a seamless experience for players worldwide. The development team is focused on optimizing the game's architecture. The team needs to select an AWS service that allows them to deploy their game servers in a way that maximizes performance for players, while accounting for the worldwide distribution of the player base.

Which AWS service should the development team use?

A. Amazon EC2 Auto Scaling
B. Amazon CloudFront
C. AWS Global Accelerator
D. Amazon RDS

Answer Choice Explanation

While Auto Scaling allows applications to automatically adjust capacity to maintain steady, predictable performance at the lowest possible cost, it primarily focuses on scaling resources rather than optimizing for low latency and high throughput across geographically distributed users.

A. **Amazon EC2 Auto Scaling**
B. **Amazon CloudFront**
C. **AWS Global Accelerator**
D. **Amazon RDS**

Answer Choice Explanation

CloudFront is a content delivery network (CDN) service that securely delivers data, videos, applications, and APIs to customers globally with low latency and high transfer speeds. While it is useful for delivering static and dynamic content, it is not specifically designed for optimizing game server performance and player latency in real-time multiplayer games.

A. Amazon EC2 Auto Scaling
B. Amazon CloudFront
C. AWS Global Accelerator
D. Amazon RDS

Answer Choice Explanation

The Global Accelerator improves the availability and performance of applications with local or global users by directing traffic to optimal endpoints over the AWS global network. It leverages the AWS backbone network to route traffic efficiently and reduces latency by providing static IP addresses that act as a fixed entry point to application endpoints in one or more AWS Regions. This service is ideal for real-time multiplayer games that require low latency and high throughput, aligning with the concept of mechanical sympathy.

A. Amazon EC2 Auto Scaling
B. Amazon CloudFront
C. AWS Global Accelerator
D. Amazon RDS

Answer Choice Explanation

RDS is a managed relational database service that makes it easy to set up, operate, and scale a relational database in the cloud. It provides cost-efficient and resizable capacity while automating time-consuming administration tasks such as hardware provisioning, database setup, patching, and backups. While important for data storage and management, Amazon RDS does not directly contribute to minimizing latency or maximizing throughput for real-time multiplayer games.

A. Amazon EC2 Auto Scaling
B. Amazon CloudFront
C. AWS Global Accelerator
D. Amazon RDS

Correct Answer

Correct Answer: C

A. Amazon EC2 Auto Scaling
B. Amazon CloudFront
C. AWS Global Accelerator
D. Amazon RDS

Question Domain 4: Design Cost-Optimized Architectures

Implement cloud financial management

Principle Definition

To achieve financial success and accelerate business value realization in the cloud, you must invest in Cloud Financial Management.

Your organization must dedicate the necessary time and resources for building capability in this new domain of technology and usage management.

Similar to your Security or Operations capability, you need to build capability through knowledge building, programs, resources, and processes to help you become a cost efficient organization.

Pay As You Go

- Adapt to changing business needs
- Stop wasting time on forecasting
- No need to overprovision

Save When You Commit

- Reservations
- Savings Plans
- 1- or 3-year commitments

Pay Less By Using More

- Volume-based discounts
- Tiered pricing
- Mostly storage and network traffic

What is CapEx?

- Up front payment
- Maintenance contracts
- Amortize value over time
- Own the product
- Predictable cost

What is OpEx?

- Subscriptions
- Pay as you go
- Operations have their own cost
- Variable and often unpredictable

Static Cost Reporting

AWS Cost Explorer

View current and historical charges

AWS Budgets

Create budgets with set monthly thresholds

Dynamic Cost Reporting

AWS Cost Explorer

View predicted costs by day or month, up to 1 year in advance

AWS Budgets

Create dynamic budgets using ML based on historical cost data

Cloud Financial Management Strategy

Functional ownership

Cost optimization function with executive sponsorship

Cloud Financial Management Strategy

Functional ownership

Finance and technology partnership

Finance must understand cloud economics, engineers must understand cloud cost structure

Cloud Financial Management Strategy

Functional ownership

Finance and technology partnership

Cloud budgets and forecasts

There should never be "sticker shock". Ever.

Cloud Financial Management Strategy

Functional ownership

Finance and technology partnership

Cloud budgets and forecasts

Cost-aware processes

Build cost awareness into regular reports and change management

Cloud Financial Management Strategy

Functional ownership

Finance and technology partnership

Cloud budgets and forecasts

Cost-aware processes

Cost-aware culture

Gamify and reward cost efficiency

Cloud Financial Management Strategy

- Functional ownership
- Finance and technology partnership
- Cloud budgets and forecasts
- Cost-aware processes
- Cost-aware culture
- Quantify business value delivered through cost optimization

Cost optimization is an ongoing, forever activity - not a "project"

Question Breakdown

Implement cloud financial management

Question and Answer Choices

A company has the following AWS migration goals for compute resources:

1. Evolve technology when appropriate
2. Start with familiar technology and learn from there
3. Reduce KTLO (Keep The Lights On) activities over time
4. Agility is top priority, followed by cost optimization

Which of the following compute cost strategies would be appropriate for the company?

A. Purchase 1 year EC2 reservations
B. Purchase 3 year EC2 reservations
C. Purchase 1 year EC2 savings plans
D. Purchase 1 year Compute savings plans

Answer Choice Explanation

1 year EC2 reservations will help reduce cost, but are limited to a single EC2 instance type.

A. **Purchase 1 year EC2 reservations**
B. **Purchase 3 year EC2 reservations**
C. **Purchase 1 year EC2 savings plans**
D. **Purchase 1 year Compute savings plans**

Answer Choice Explanation

3 year EC2 reservations provide a larger discount than 1 year, but lock you into a single instance type with some flexibility to change within the instance family, but not much.

A. Purchase 1 year EC2 reservations
B. Purchase 3 year EC2 reservations
C. Purchase 1 year EC2 savings plans
D. Purchase 1 year Compute savings plans

Answer Choice Explanation

A savings plan is less discount than a reservation, but more flexible as it covers an entire instance family automatically.

A. Purchase 1 year EC2 reservations
B. Purchase 3 year EC2 reservations
C. Purchase 1 year EC2 savings plans
D. Purchase 1 year Compute savings plans

Answer Choice Explanation

A Compute savings plan covers all EC2, Fargate and Lambda, and while the discount is less than EC2 savings plans, it absolutely allows the organization to be more agile in the choice of compute services.

A. Purchase 1 year EC2 reservations
B. Purchase 3 year EC2 reservations
C. Purchase 1 year EC2 savings plans
D. Purchase 1 year Compute savings plans

Correct Answer

Correct Answer: D

A. Purchase 1 year EC2 reservations
B. Purchase 3 year EC2 reservations
C. Purchase 1 year EC2 savings plans
D. Purchase 1 year Compute savings plans

Question Domain 4: Design Cost-Optimized Architectures

Adopt a consumption model

Principle Definition

Pay only for the computing resources you consume, and increase or decrease usage depending on business requirements.

For example, development and test environments are typically only used for eight hours a day during the work week.

You can stop these resources when they're not in use for a potential cost savings of 75% (40 hours versus 168 hours).

Question Breakdown

Adopt a consumption model

Cost-Optimization Scenario

A software development company is planning to deploy a new web application that is expected to have fluctuating traffic, with high usage during the weekdays and significantly lower traffic on weekends. The application will be hosted on EC2 instances within an Auto Scaling group.

The company aims to optimize their AWS costs while maintaining high availability and performance. They are evaluating different EC2 pricing options to find the most cost-effective solution for their use case.

Which EC2 pricing options should the company choose to optimize costs for their application's expected usage pattern?

EC2 Pricing Options

Spot Instances

- No guaranteed pricing
- Pay for unused capacity
- Volatile
- +Attribute selection
- +Multiple instance types
- +Multiple AZ

EC2 Pricing Options

Spot Instances	RIs/SPs
• No guaranteed pricing • Pay for unused capacity • Volatile • +Attribute selection • +Multiple instance types • +Multiple AZ	• Guaranteed pricing for 1-3 years • Variable up-front for more discount • EC2 Savings Plans for more flexibility • Compute Savings Plans for even more flexibility!

EC2 Pricing Options

Spot Instances	RIs/SPs	On Demand Instances
• No guaranteed pricing • Pay for unused capacity • Volatile • +Attribute selection • +Multiple instance types • +Multiple AZ	• Guaranteed pricing for 1-3 years • Variable up-front for more discount • EC2 Savings Plans for more flexibility • Compute Savings Plans for even more flexibility!	• Pay as you go • No discount

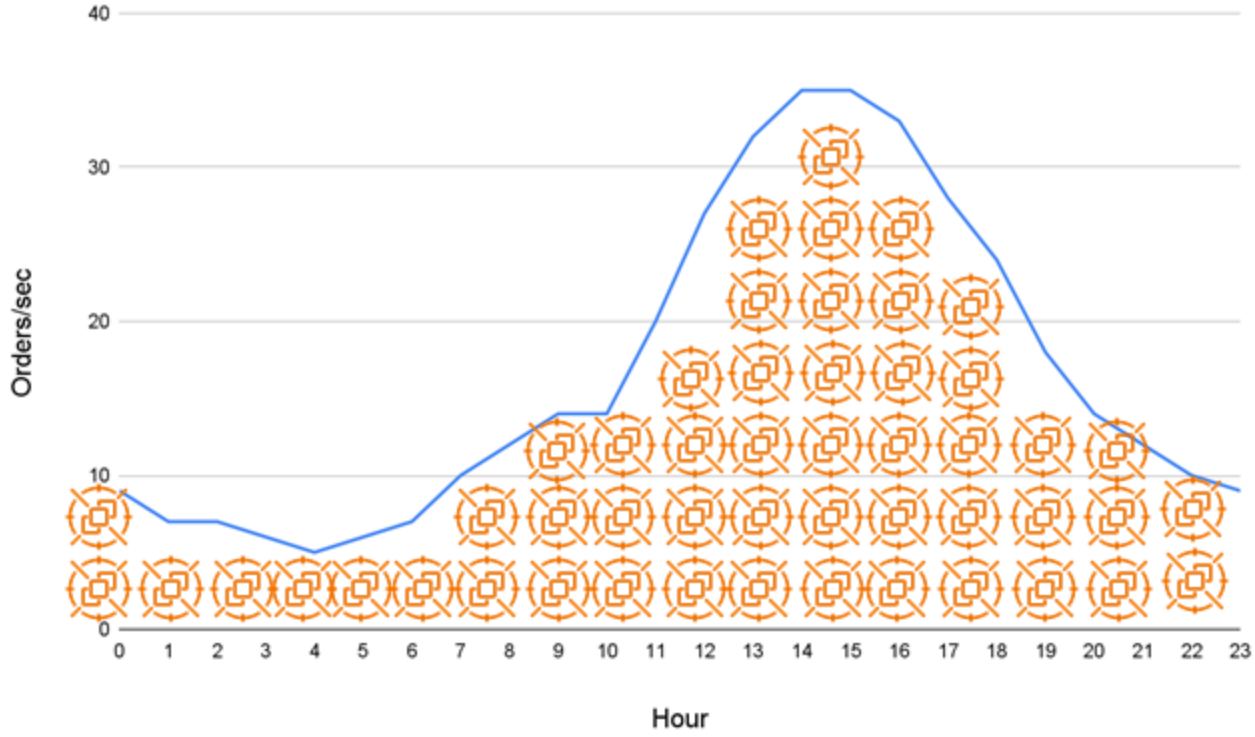
EC2 Pricing Options

Spot Instances	RIs/SPs	On Demand Instances	Dedicated Instances
• No guaranteed pricing • Pay for unused capacity • Volatile • +Attribute selection • +Multiple instance types • +Multiple AZ	• Guaranteed pricing for 1-3 years • Variable up-front for more discount • EC2 Savings Plans for more flexibility • Compute Savings Plans for even more flexibility!	• Pay as you go • No discount	• Dedicated hardware • Can share with non-dedicated VMs • Per-region fee • +Spot • +Reservations • +On Demand

EC2 Pricing Options

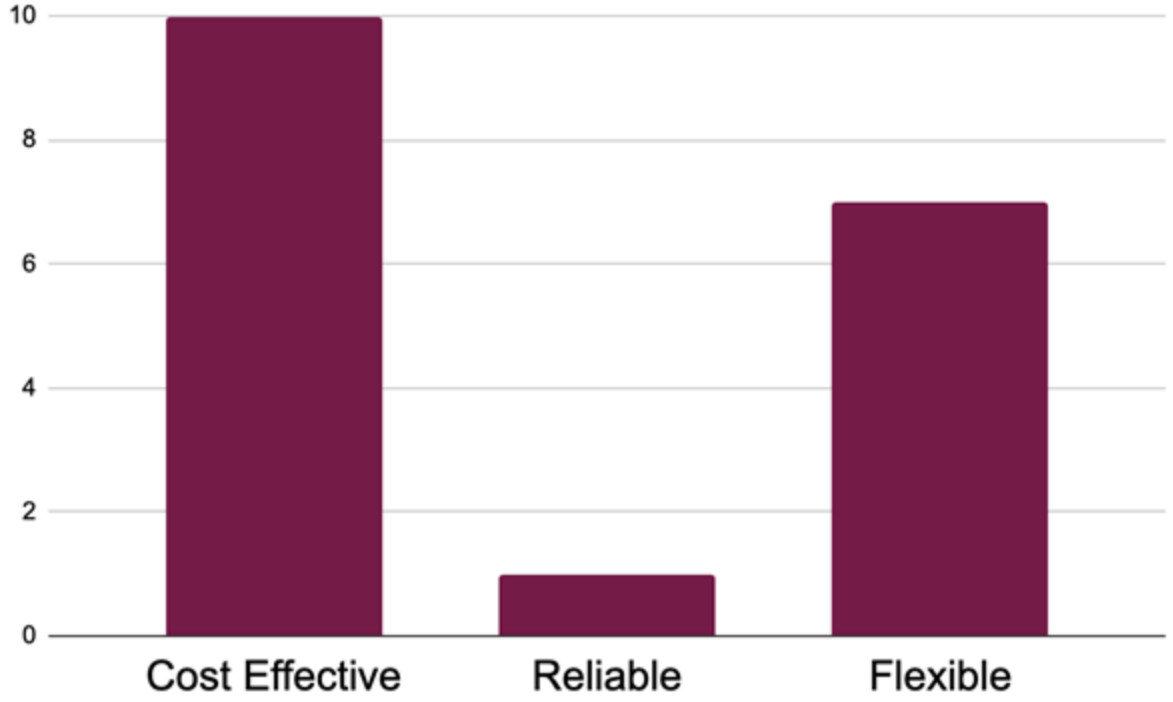
Spot Instances	RIs/SPs	On Demand Instances	Dedicated Instances	Dedicated Hosts
• No guaranteed pricing • Pay for unused capacity • Volatile • +Attribute selection • +Multiple instance types • +Multiple AZ	• Guaranteed pricing for 1-3 years • Variable up-front for more discount • EC2 Savings Plans for more flexibility • Compute Savings Plans for even more flexibility!	• Pay as you go • No discount	• Dedicated hardware • Can share with non-dedicated VMs • Per-region fee • +Spot • +Reservations • +On Demand	• Dedicated hardware • Single instance type • Pay for host capacity, not instance • +Reservations • +On Demand

Solution 1: Spot Instances

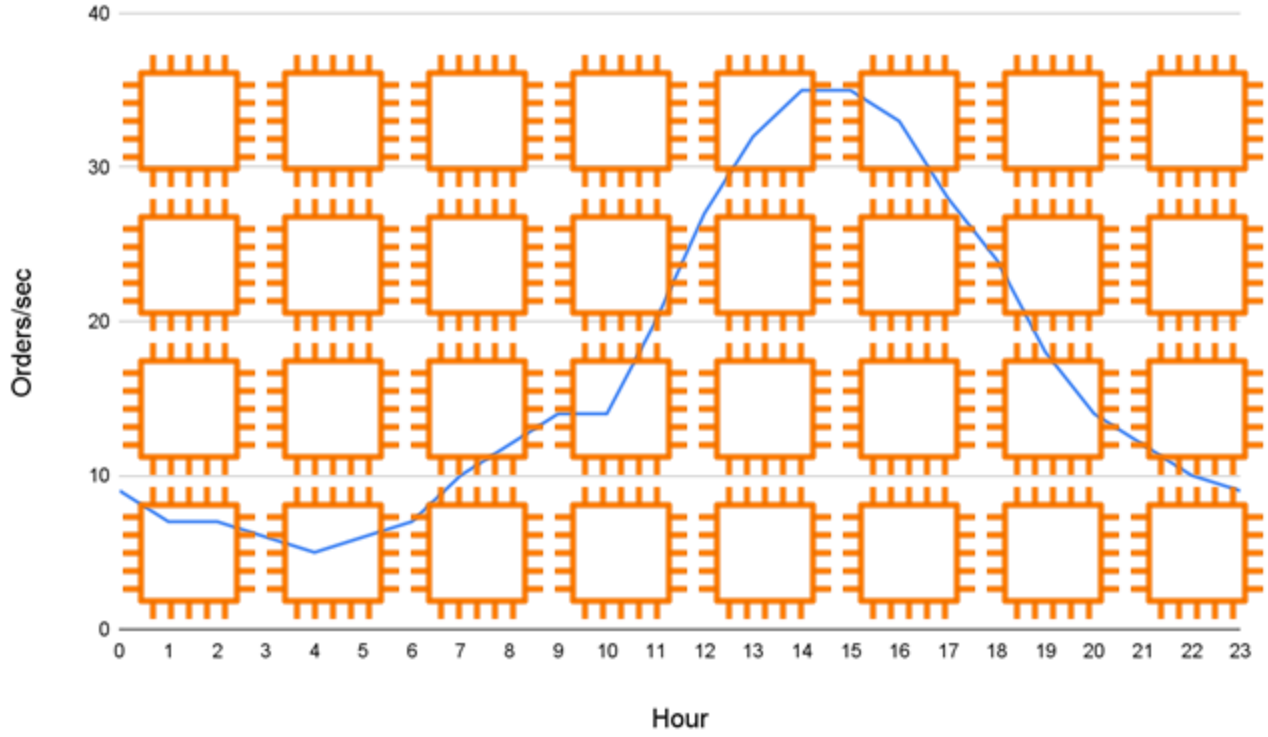


Pearson

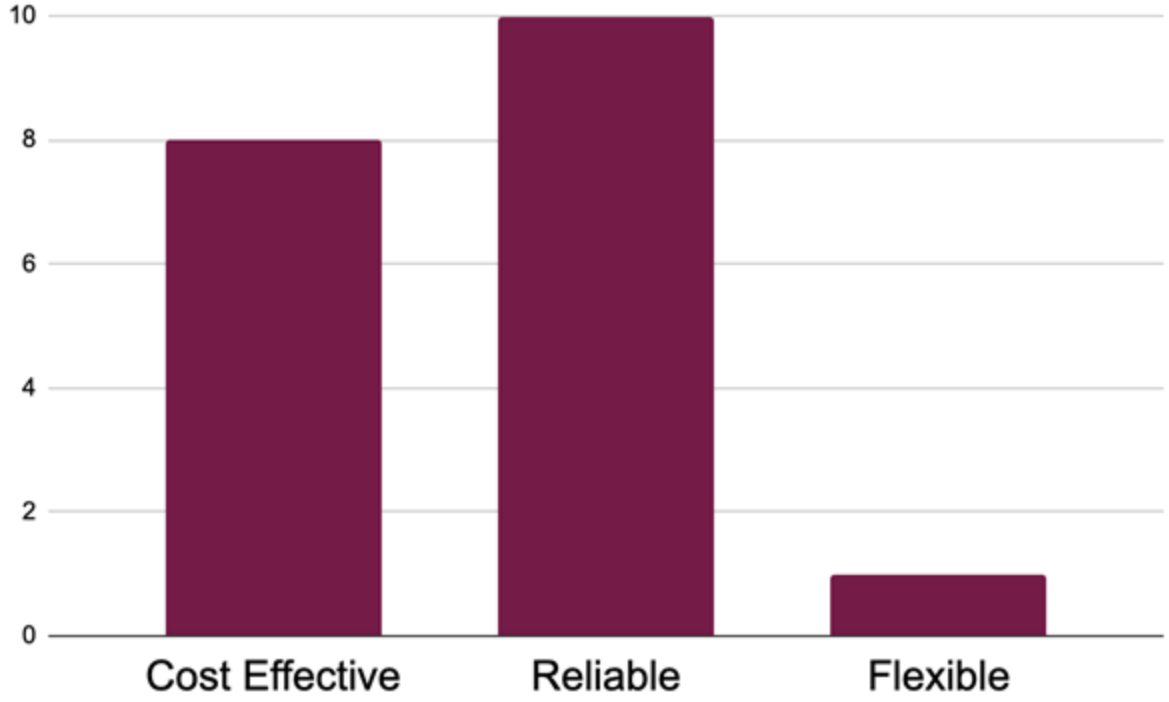
Solution 1 Considerations



Solution 2: Reservations

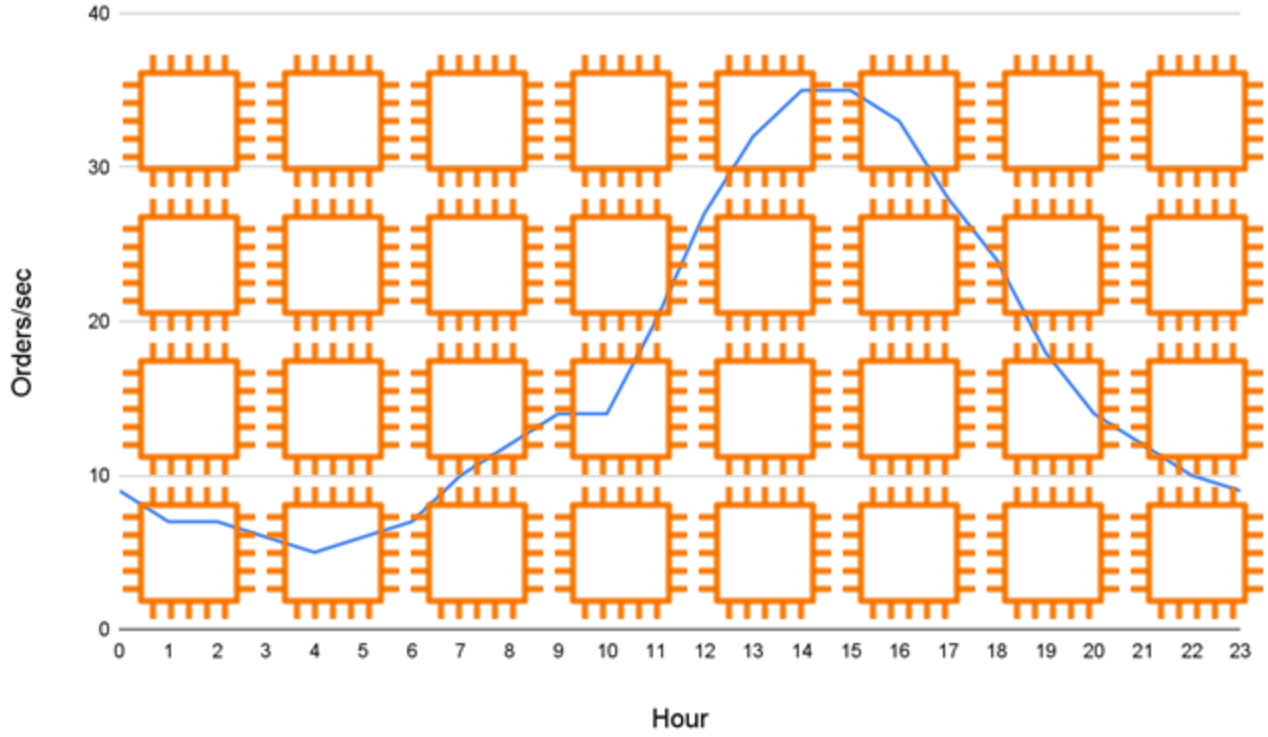


Solution 2 Considerations

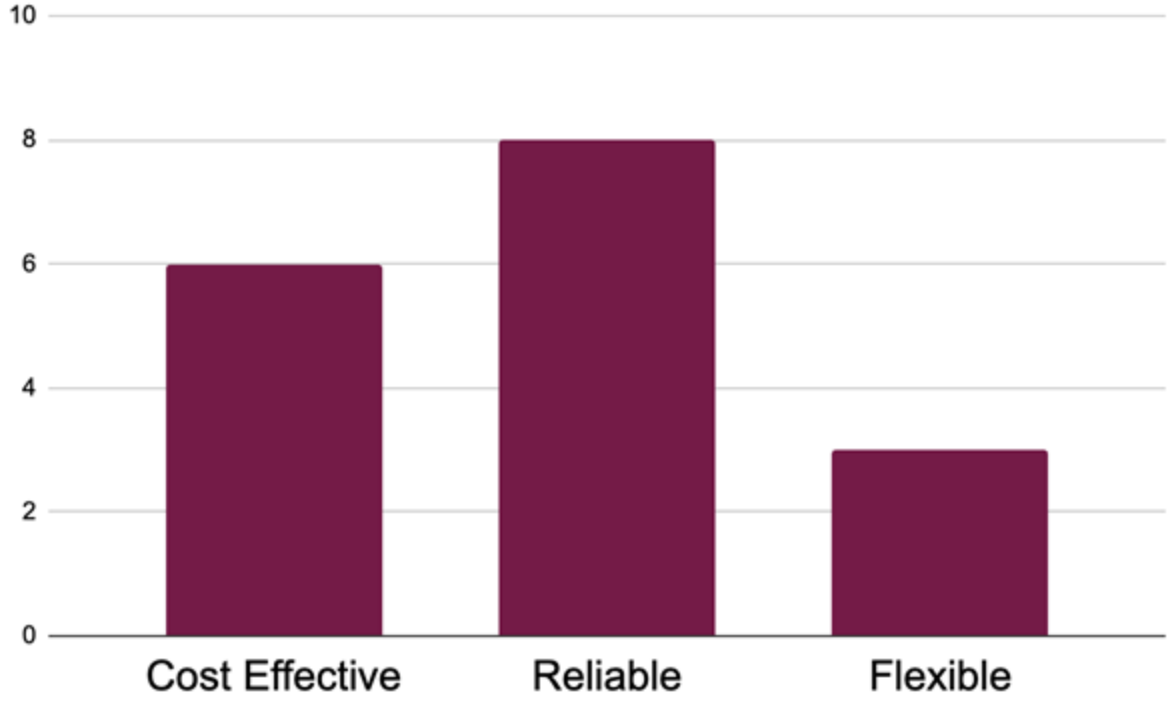


Pearson

Solution 3: EC2 Savings Plans

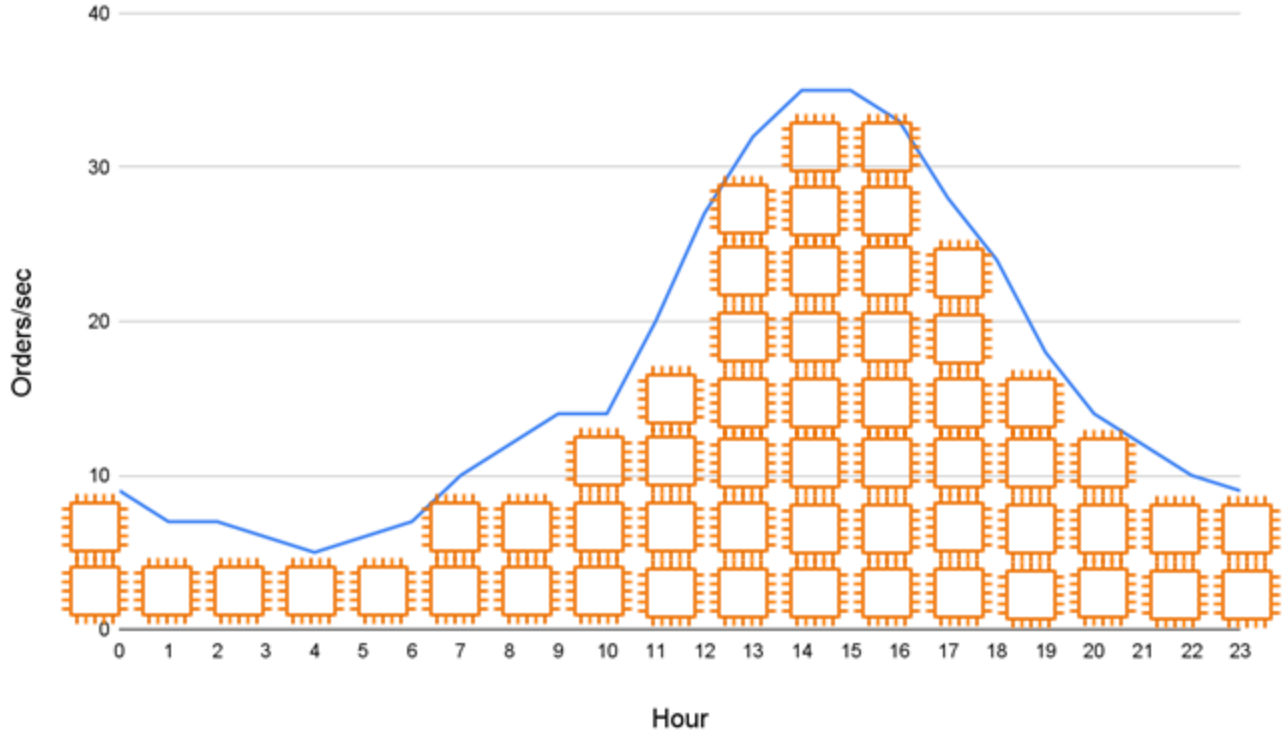


Solution 3 Considerations



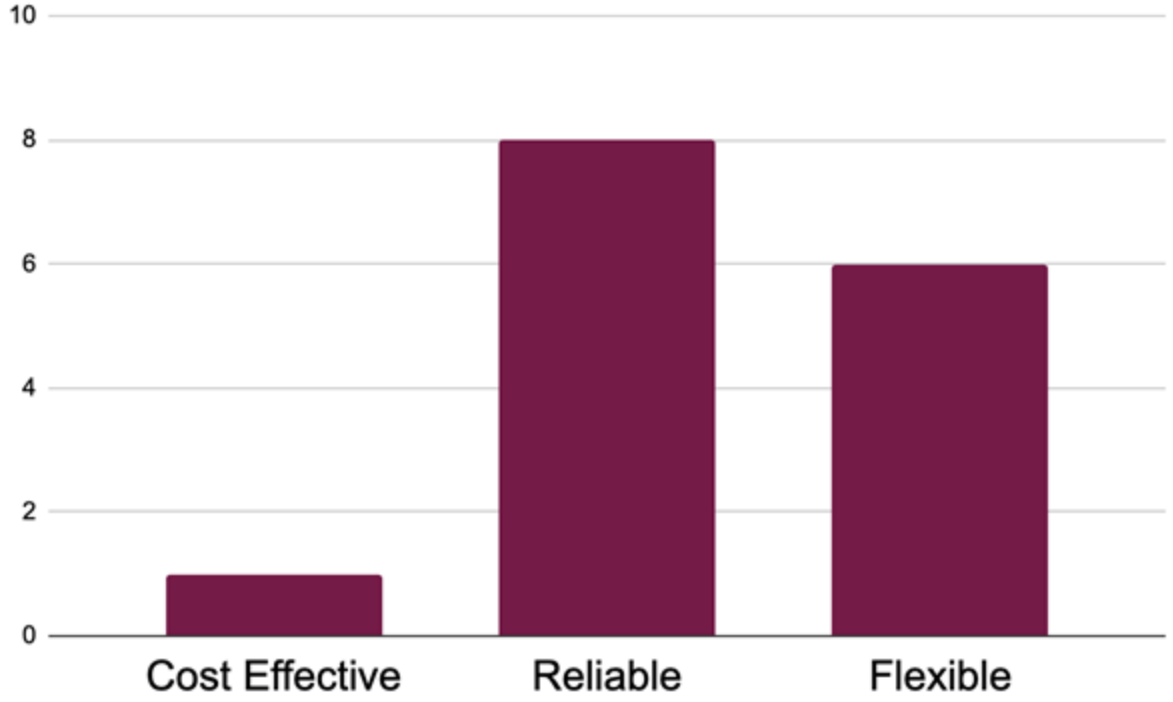
Pearson

Solution 4: On-demand Instances



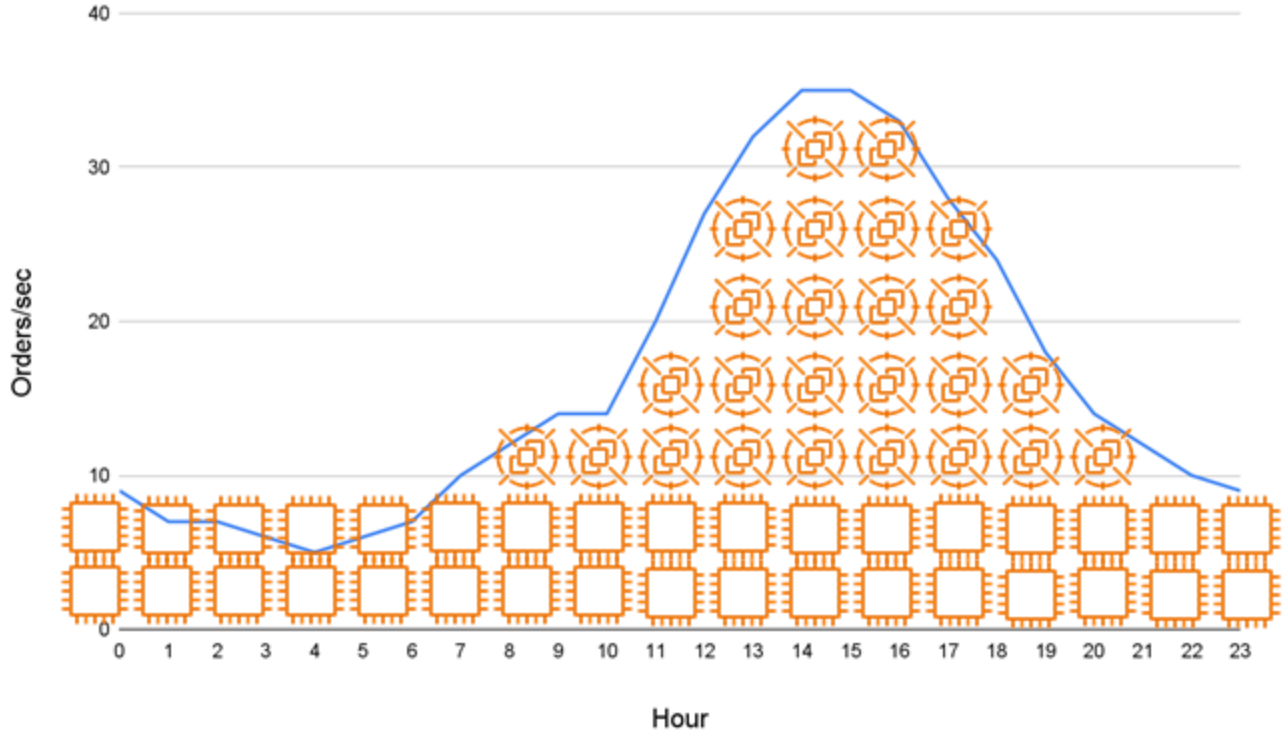
Pearson

Solution 4 Considerations



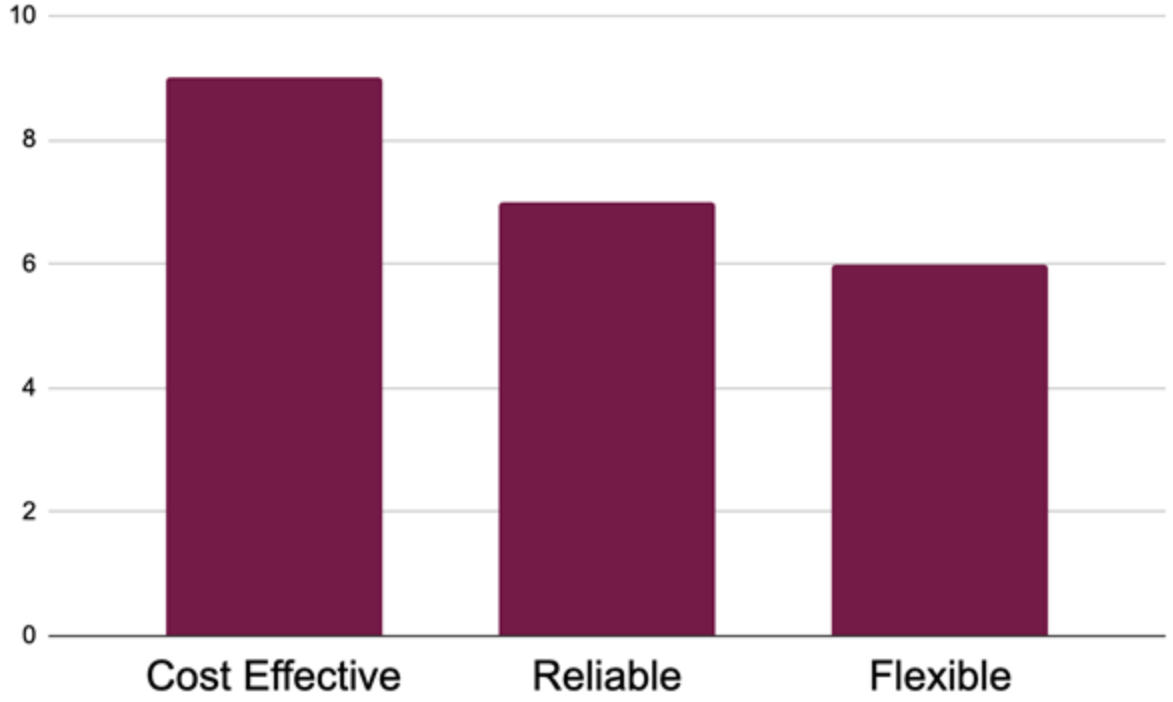
Pearson

Solution 5: Hybrid Configuration



Pearson

Solution 5 Considerations



Pearson

Question and Answer Choices

When designing a multi-tier application, all non-production environments must be architected to optimize for cost after hours and on weekends. The application tiers include a web proxy, application runtime, and a database.

Which of the following recommendations will balance cost optimization with low operational effort? (pick three)

A. Application Load Balancer
B. EC2 with nginx using an Auto Scaling group
C. Application deployed on EC2 instance using Jenkins
D. Application deployed on EC2 using an Auto Scaling group
E. RDS database instance
F. Aurora Serverless

Answer Choice Explanation

The ALB, while it is a managed service, does have a minimum footprint on cost even with zero traffic, but the operational overhead of managing this solution is also low/zero.

A. Application Load Balancer
B. EC2 with nginx using an Auto Scaling group
C. Application deployed on EC2 instance using Jenkins
D. Application deployed on EC2 using an Auto Scaling group
E. RDS database instance
F. Aurora Serverless

Answer Choice Explanation

It may seem counterintuitive, but if there is no traffic, the EC2-based web proxy solution can scale better to 0 total cost, by setting the number of desired instances to 0. The overhead of managing this solution is relatively high.

A. Application Load Balancer
B. EC2 with nginx using an Auto Scaling group
C. Application deployed on EC2 instance using Jenkins
D. Application deployed on EC2 using an Auto Scaling group
E. RDS database instance
F. Aurora Serverless

Answer Choice Explanation

Deploying the application manually using a single EC2 instance is not going to scale to 0 well, if at all.

A. Application Load Balancer
B. EC2 with nginx using an Auto Scaling group
C. Application deployed on EC2 instance using Jenkins
D. Application deployed on EC2 using an Auto Scaling group
E. RDS database instance
F. Aurora Serverless

Answer Choice Explanation

An Auto Scaling group (similar to B) can scale to zero with the correct configuration.

A. Application Load Balancer
B. EC2 with nginx using an Auto Scaling group
C. Application deployed on EC2 instance using Jenkins
D. Application deployed on EC2 using an Auto Scaling group
E. RDS database instance
F. Aurora Serverless

Answer Choice Explanation

RDS has no intrinsic features for scaling to 0. It is possible to stop a database instance, but only for a 1 week period, and not automatically.

A. Application Load Balancer
B. EC2 with nginx using an Auto Scaling group
C. Application deployed on EC2 instance using Jenkins
D. Application deployed on EC2 using an Auto Scaling group
E. RDS database instance
F. Aurora Serverless

Answer Choice Explanation

Aurora Serverless can be configured to reduce the number of ACUs to 0.5 during periods of idle activity, which is close to zero, and entirely automated, so no operational overhead involved.

A. Application Load Balancer
B. EC2 with nginx using an Auto Scaling group
C. Application deployed on EC2 instance using Jenkins
D. Application deployed on EC2 using an Auto Scaling group
E. RDS database instance
F. Aurora Serverless

Correct Answer

Correct Answers: B,D,F

A. Application Load Balancer
B. EC2 with nginx using an Auto Scaling group
C. Application deployed on EC2 instance using Jenkins
D. Application deployed on EC2 using an Auto Scaling group
E. RDS database instance
F. Aurora Serverless

Question Domain 4: Design Cost-Optimized Architectures

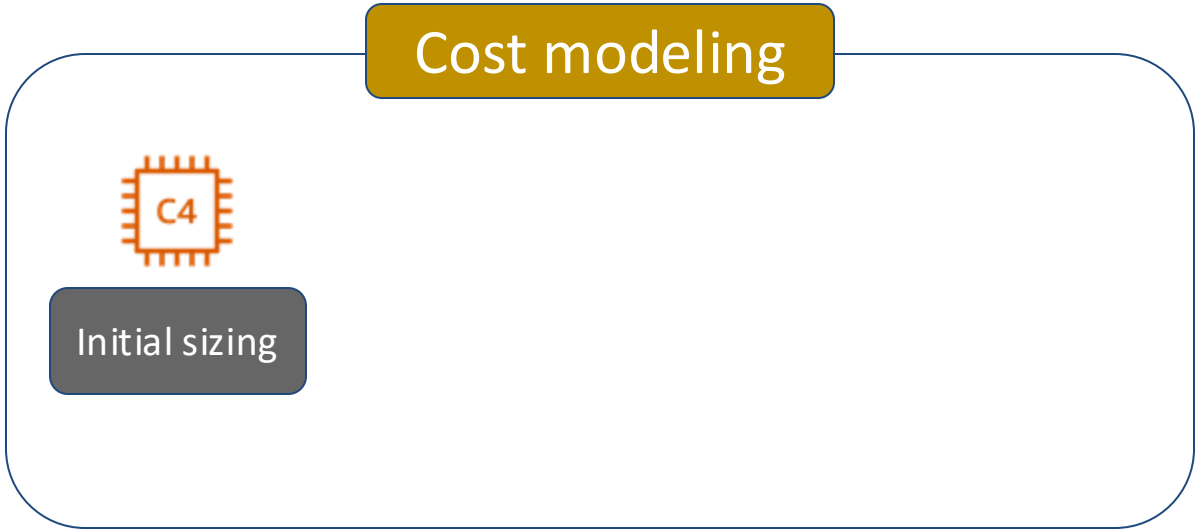
Measure overall efficiency

Principle Definition

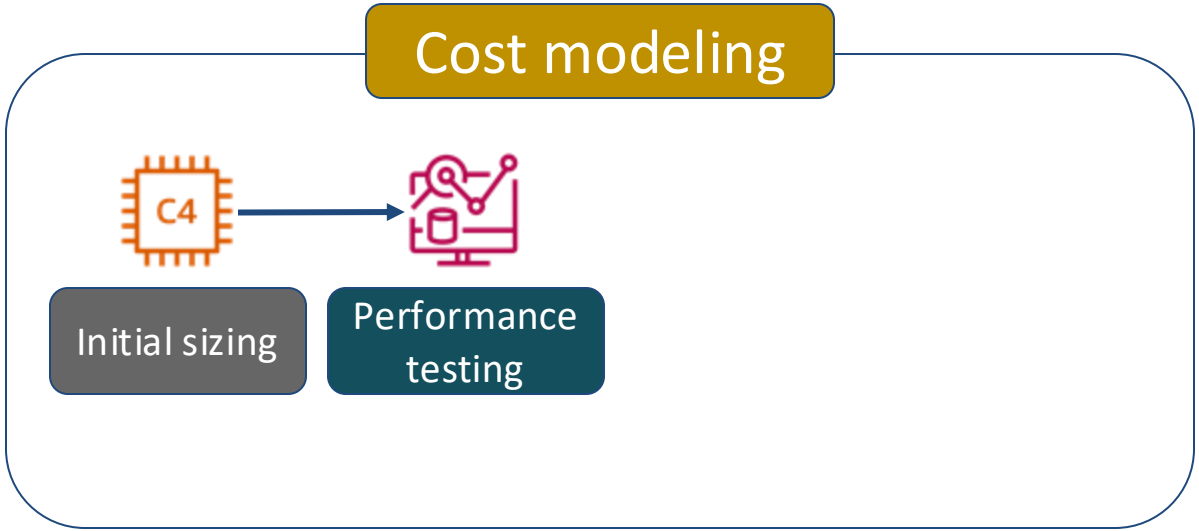
Measure the business output of the workload and the costs associated with delivery.

Use this data to understand the gains you make from increasing output, increasing functionality, and reducing cost.

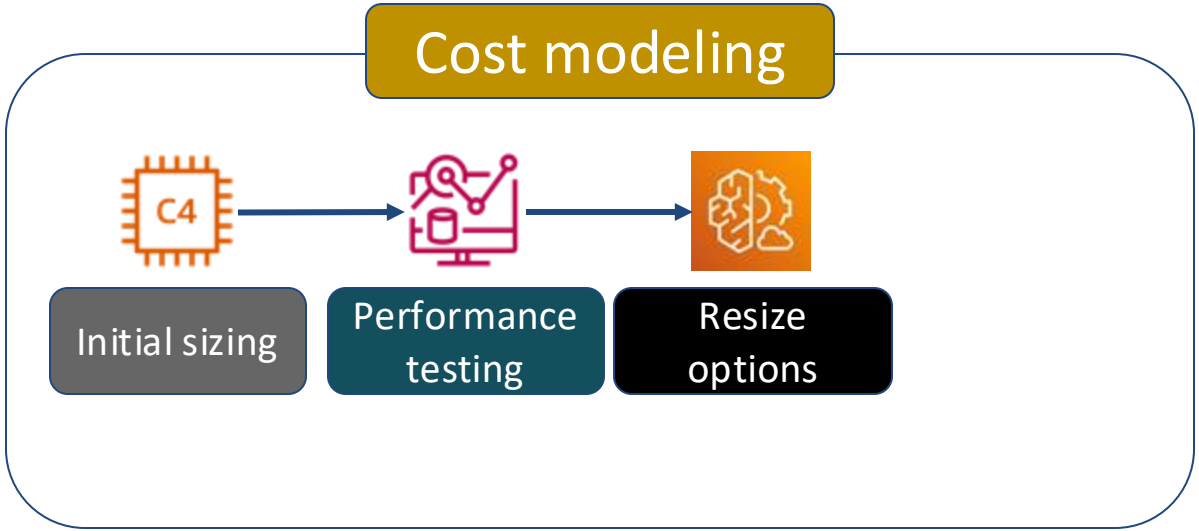
Select the Correct Resource Type, Size, and Number



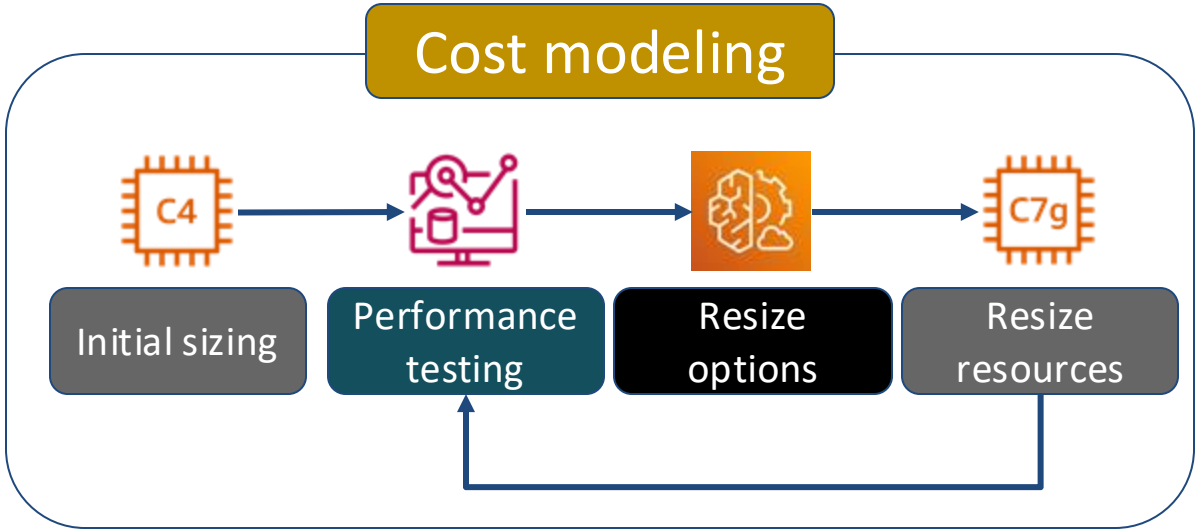
Select the Correct Resource Type, Size, and Number



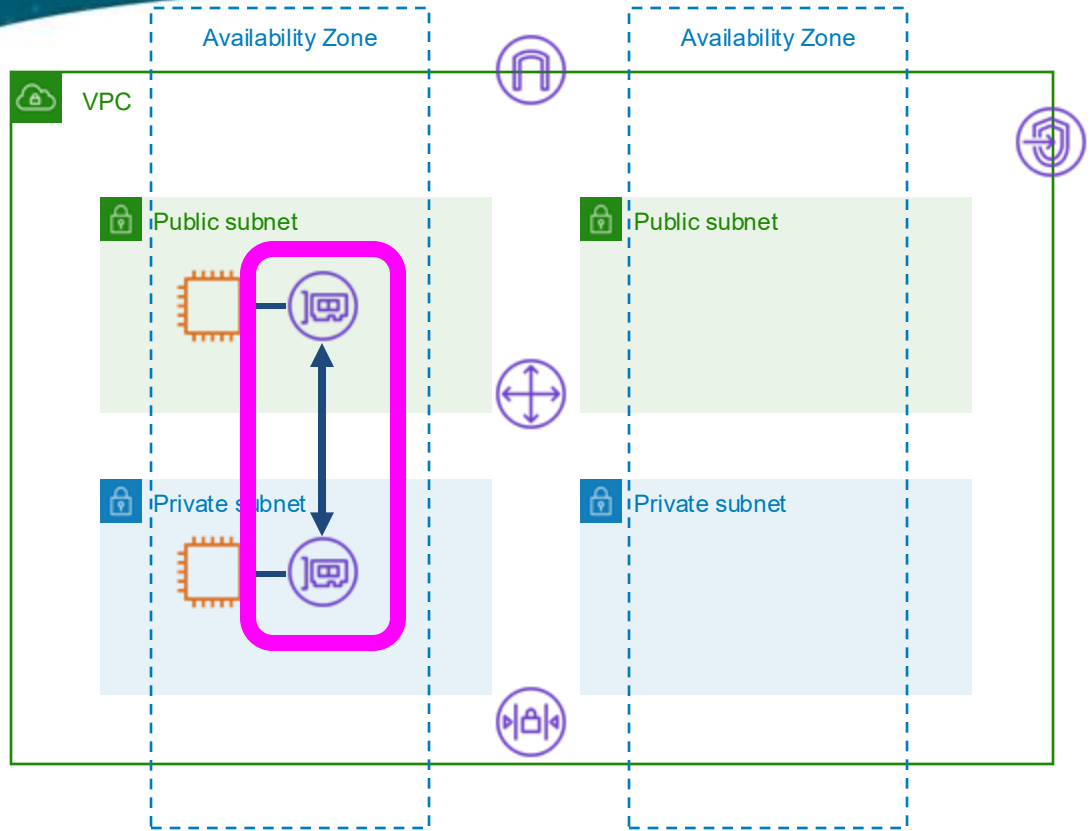
Select the Correct Resource Type, Size, and Number



Select the Correct Resource Type, Size, and Number

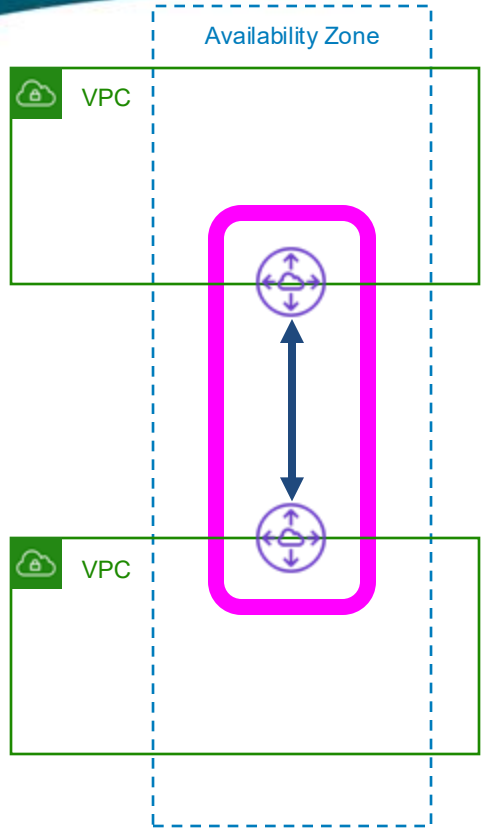


Plan for data transfer



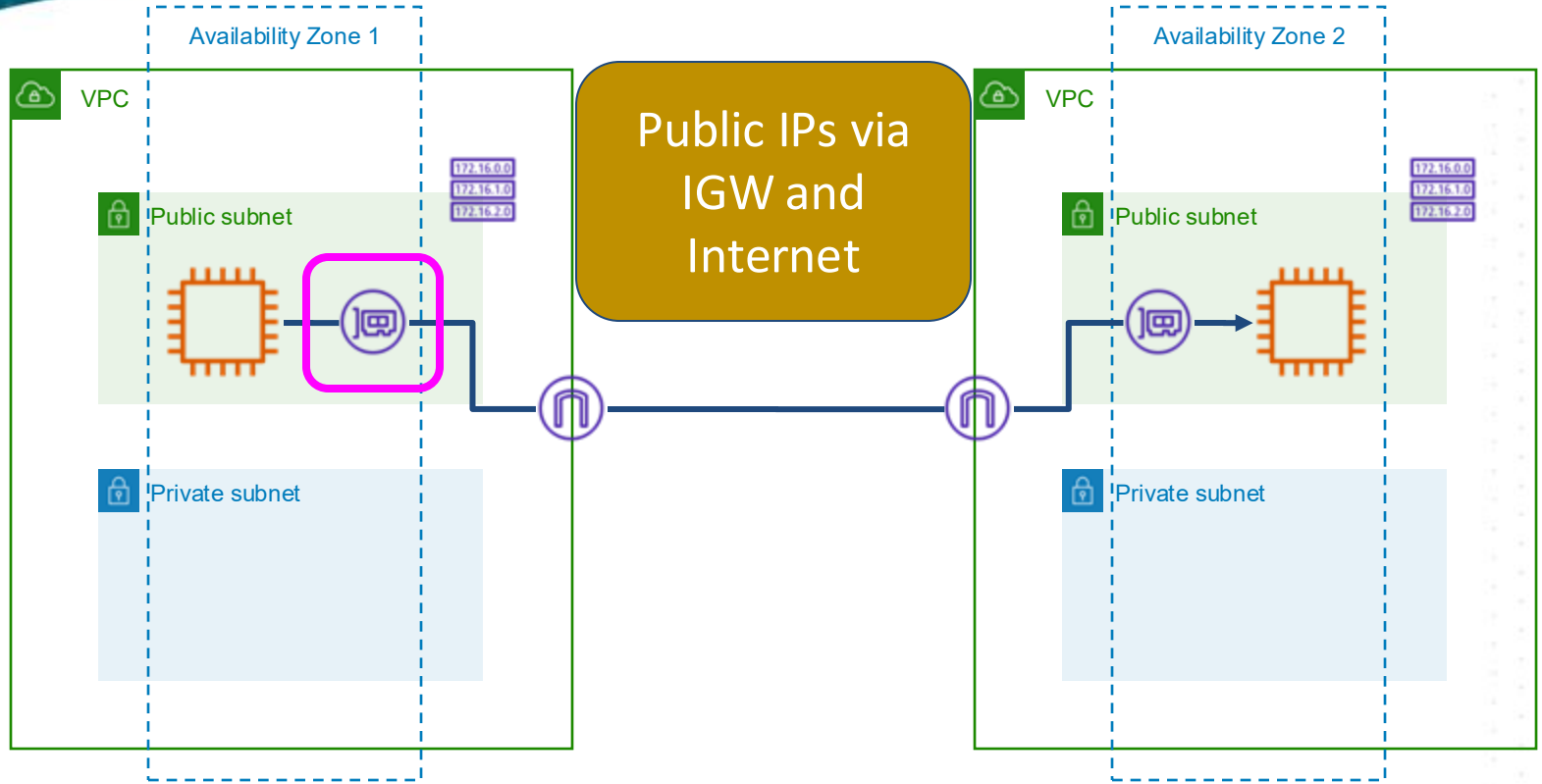
Same-AZ network traffic is free EXCEPT if a public IP is the destination

Plan for data transfer

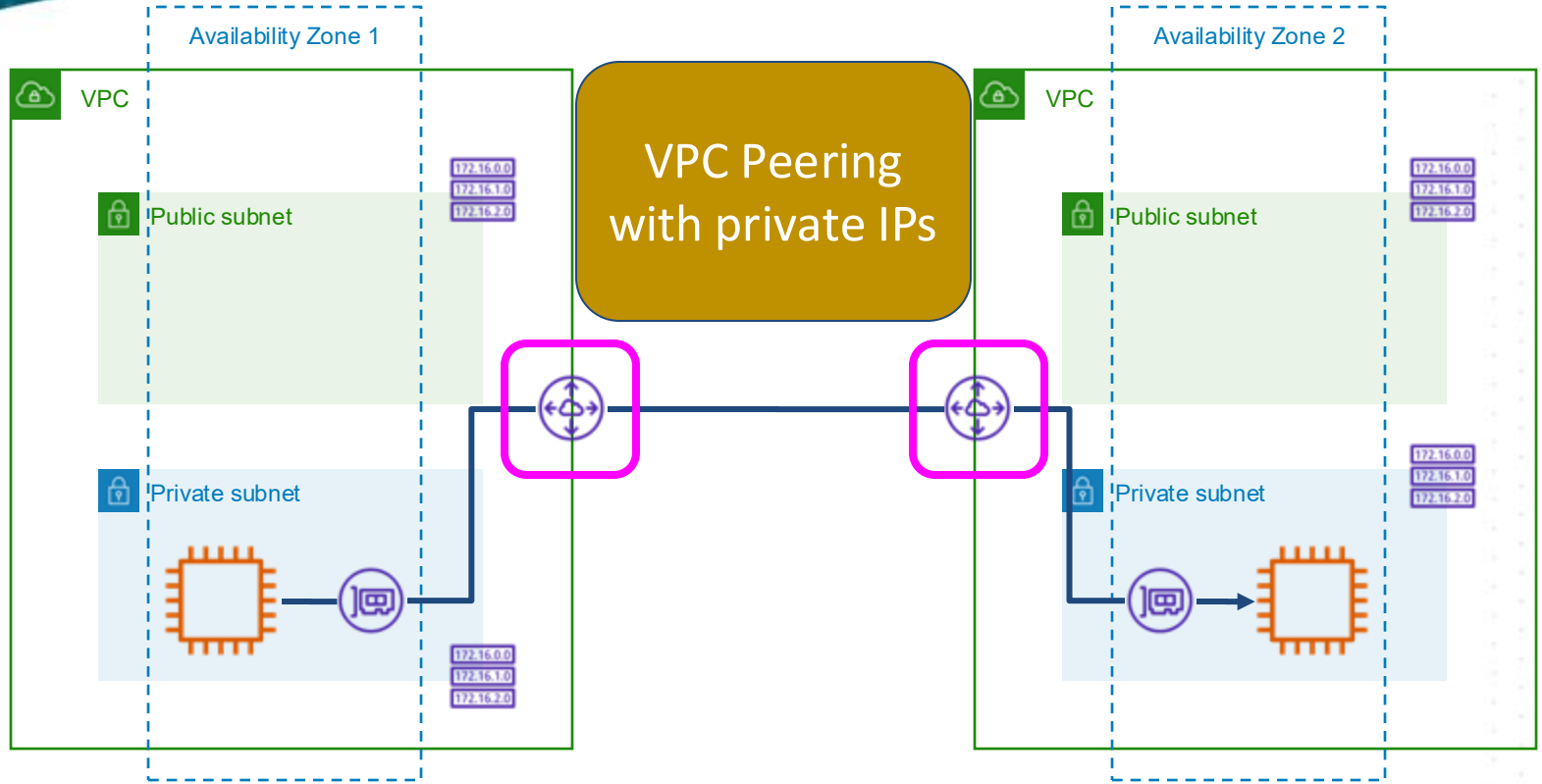


Same-AZ VPC Peering traffic is free, as long as the public IP of the destination is not used

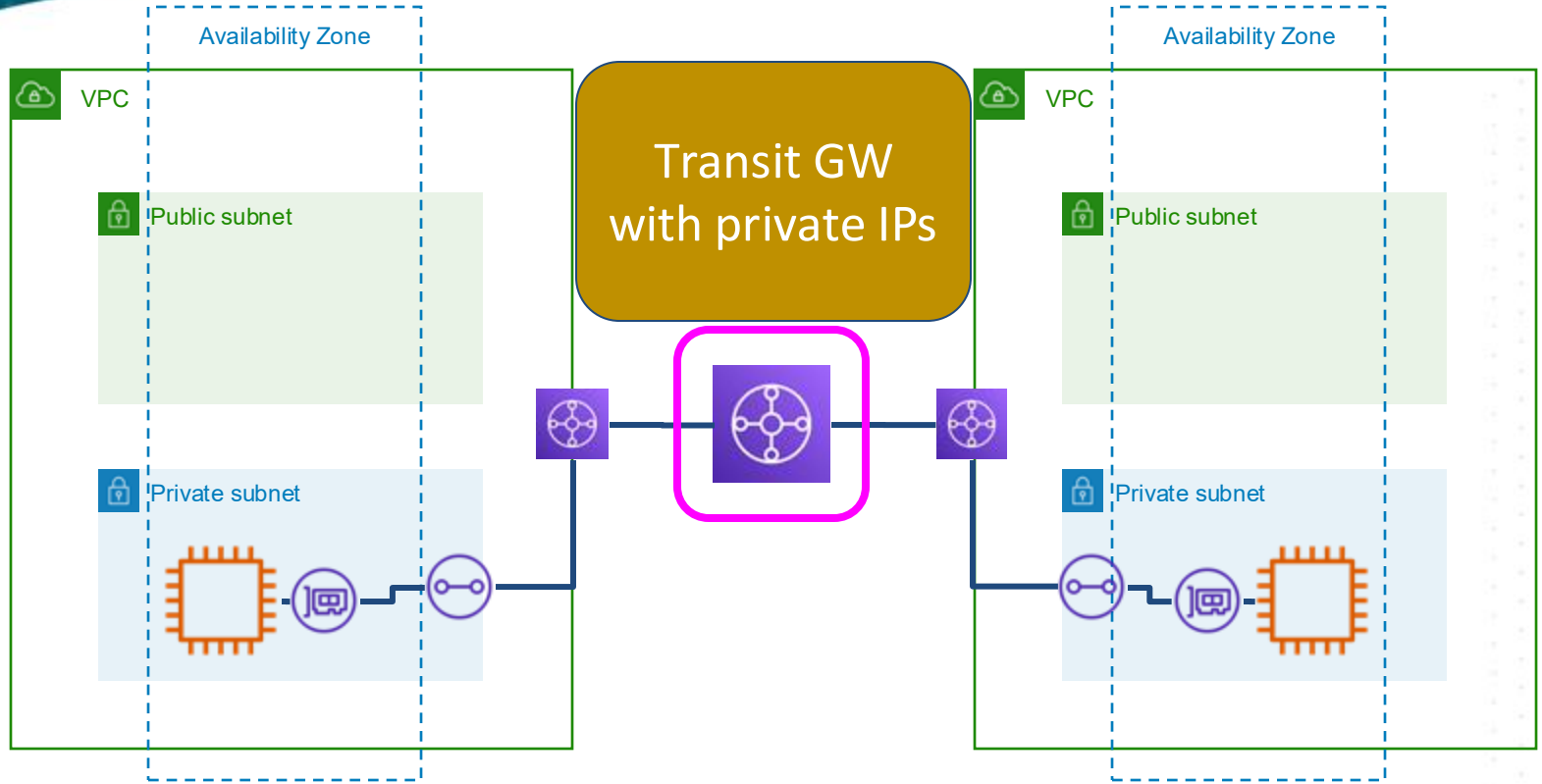
Plan for data transfer



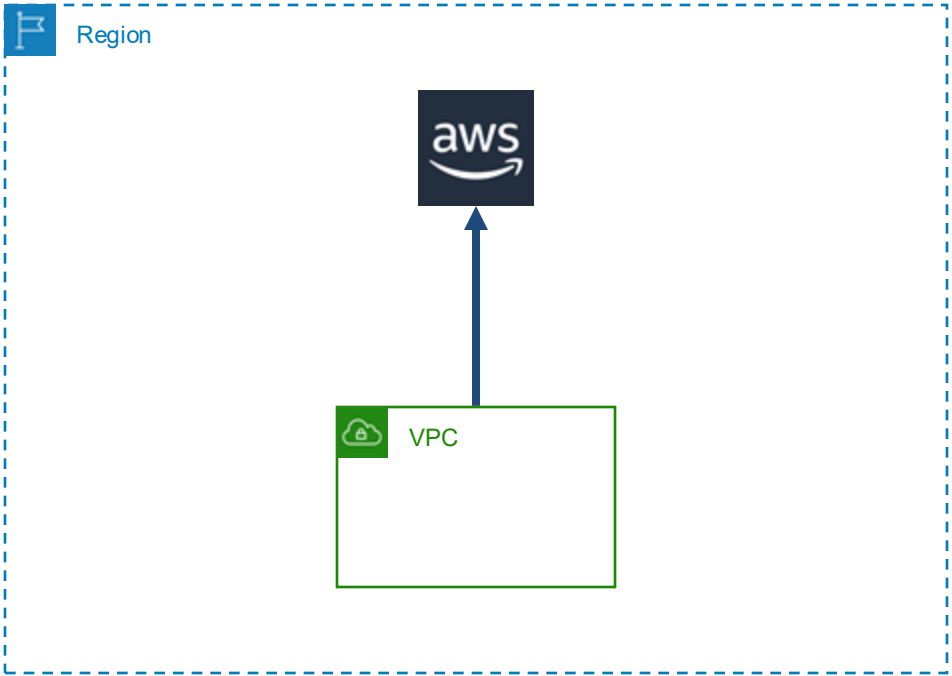
Plan for data transfer



Plan for data transfer charges

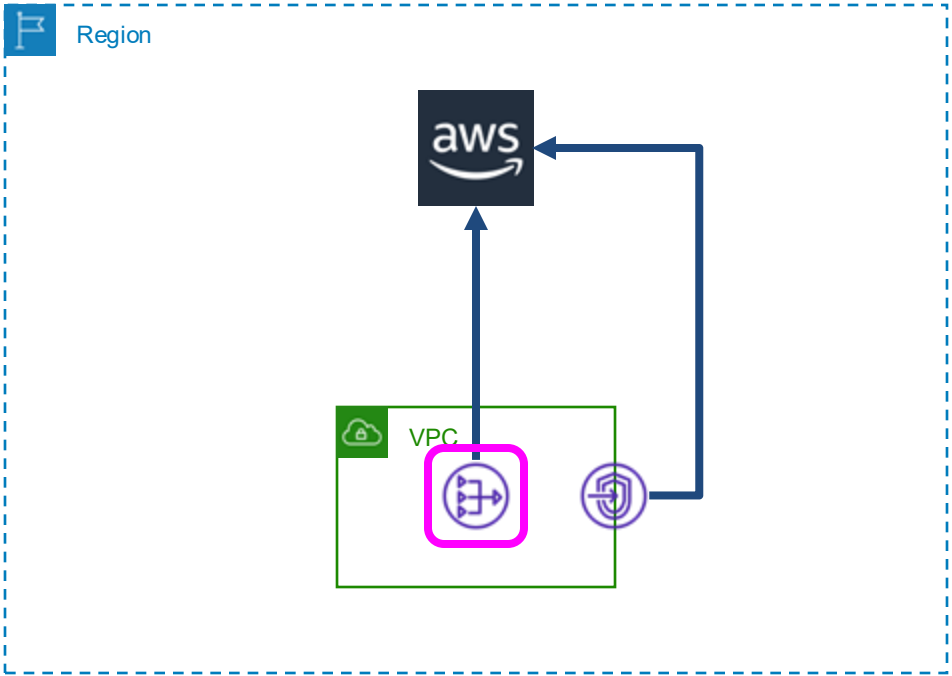


Plan for data transfer



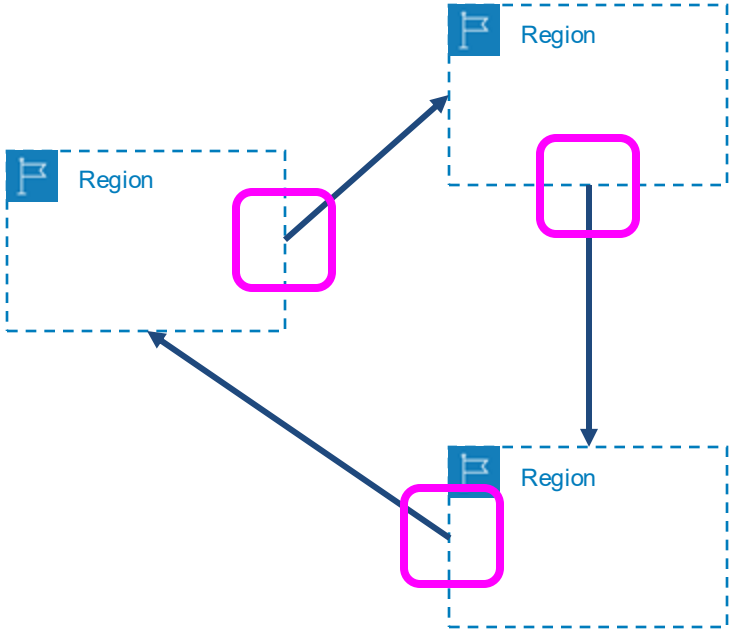
Most same-region traffic from VPC to AWS services will be free, such as S3 bucket access, unless otherwise noted

Plan for data transfer



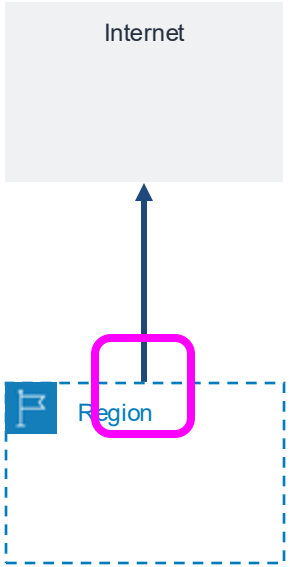
NAT Gateway data transfer is always chargeable in the same region, while Gateway VPC endpoint data transfer is free

Plan for data transfer



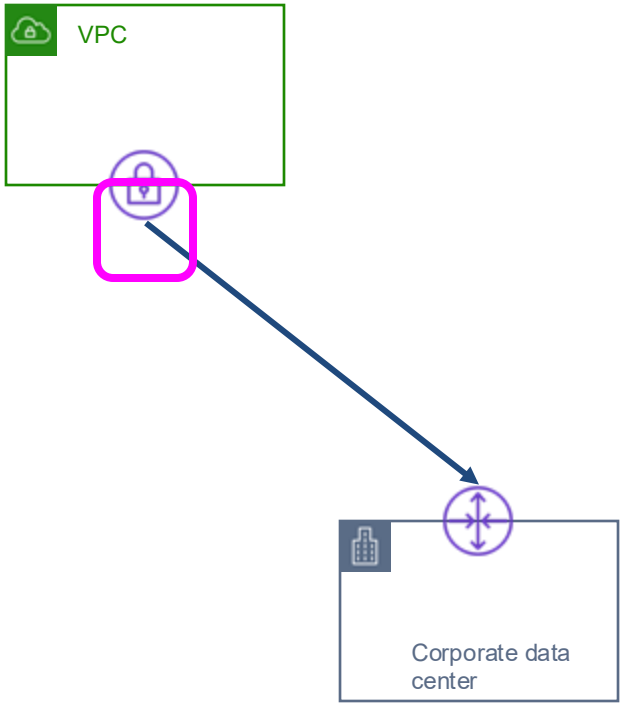
All outbound cross-region service data transfer is charged

Plan for data transfer



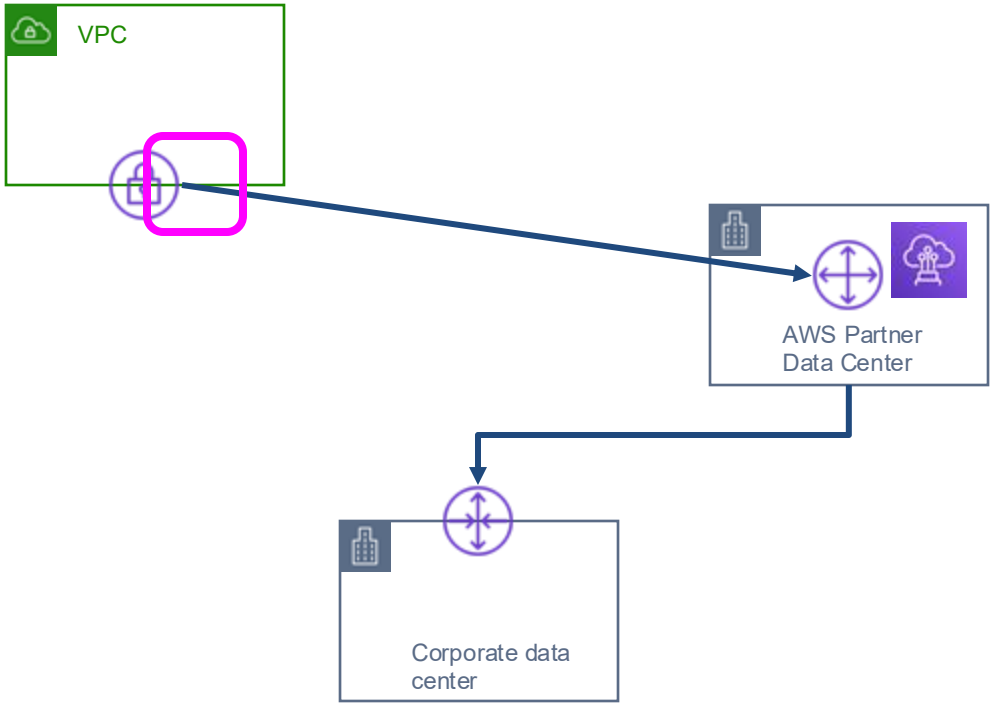
All outbound Internet traffic is charged, and there can be additional fees based on the gateway used

Plan for data transfer



Site-to-site VPN data transfer is charged with Internet outbound rates

Plan for data transfer



Direct Connect data transfer charges are similar to cross-region rates and applied to outbound traffic only

Question Breakdown

Measure overall efficiency

Question and Answer Choices

One of your memory-bound applications needs to have a performance baseline established. The application runs on EC2 instances, and you've been asked to design performance baseline tests that include memory usage metrics from the EC2 Operating System. The memory usage metrics will then be used to recommend the appropriate instance size to ensure cost efficient resource usage.

How can you ensure the memory metrics are collected?

A. Use the default CloudWatch metrics for EC2
B. Enable detailed monitoring on the EC2 instance
C. Install/configure the CloudWatch Agent on the EC2 instance
D. Install/configure the CloudWatch Logs Agent on the EC2 instance

Answer Choice Explanation

The default CloudWatch metrics are pushed from the hypervisor, and do not include any metrics from inside the guest OS, such as memory usage, or cpu usage breakdown into individual metrics.

A. **Use the default CloudWatch metrics for EC2**
B. **Enable detailed monitoring on the EC2 instance**
C. **Install/configure the CloudWatch Agent on the EC2 instance**
D. **Install/configure the CloudWatch Logs Agent on the EC2 instance**

Answer Choice Explanation

Detailed monitoring changes the default CloudWatch period from 5 minutes to 1 minute, but does not include any new metrics.

A. Use the default CloudWatch metrics for EC2
B. Enable detailed monitoring on the EC2 instance
C. Install/configure the CloudWatch Agent on the EC2 instance
D. Install/configure the CloudWatch Logs Agent on the EC2 instance

Answer Choice Explanation

The CloudWatch Agent has the capability to deliver memory usage metrics to CloudWatch with a configurable polling period.

A. Use the default CloudWatch metrics for EC2
B. Enable detailed monitoring on the EC2 instance
C. Install/configure the CloudWatch Agent on the EC2 instance
D. Install/configure the CloudWatch Logs Agent on the EC2 instance

Answer Choice Explanation

The CloudWatch Logs Agent is a previous-generation utility for streaming OS or application logs to CloudWatch, and would not be appropriate for memory usage metric delivery.

A. Use the default CloudWatch metrics for EC2
B. Enable detailed monitoring on the EC2 instance
C. Install/configure the CloudWatch Agent on the EC2 instance
D. Install/configure the CloudWatch Logs Agent on the EC2 instance

Correct Answer

Correct Answer: C

A. Use the default CloudWatch metrics for EC2
B. Enable detailed monitoring on the EC2 instance
C. Install/configure the CloudWatch Agent on the EC2 instance
D. Install/configure the CloudWatch Logs Agent on the EC2 instance

Question Domain 4: Design Cost-Optimized Architectures

Stop spending money on undifferentiated heavy lifting

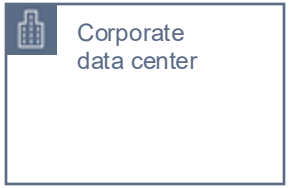
Principle Definition

AWS does the heavy lifting of data center operations like racking, stacking, and powering servers.

It also removes the operational burden of managing operating systems and applications with managed services.

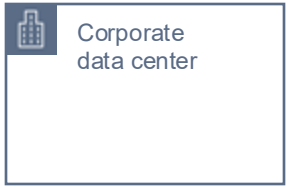
This allows you to focus on your customers and business projects rather than on IT infrastructure.

TCO - Total Cost of Ownership



Data Center

TCO - Total Cost of Ownership



Data Center

Hardware

TCO - Total Cost of Ownership

Data Center

Hardware

Storage

TCO - Total Cost of Ownership

Data Center | Hardware | Storage | Network

What do many organizations consider to be Total Cost of Ownership?

TCO - Total Cost of Ownership

Data Center | Hardware | Storage | Network

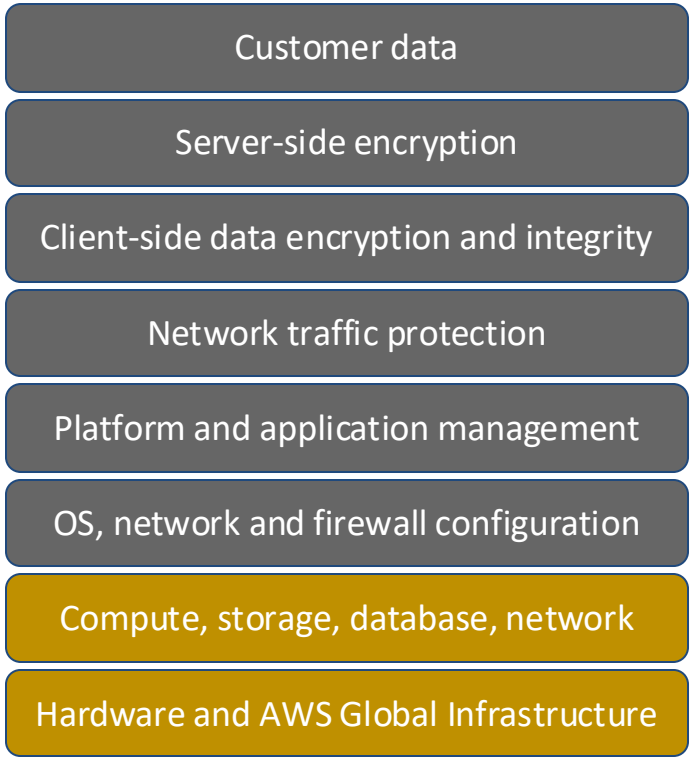
What do many organizations consider to be Total Cost of Ownership?

What is missing from this list?

KTLO - Keep The Lights On

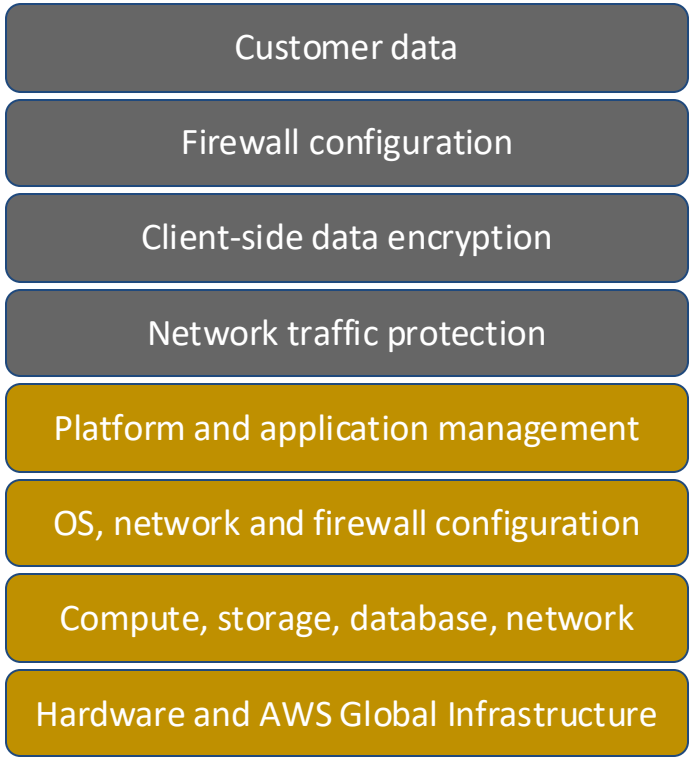
- Any zero-sum game operation
- Proportional to unmanaged resources
- More OS-based resources = more operations
- Cannot improve agility
- Does not scale

Revisiting the Shared Responsibility Model



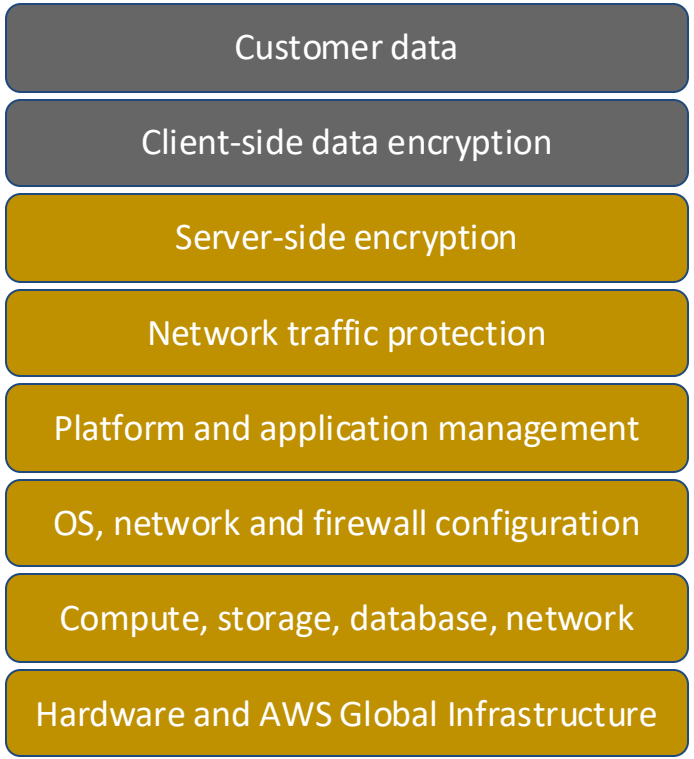
Infrastructure services provide a lot of flexibility but also a lot of KTLO tasks

Revisiting the Shared Responsibility Model



Container services (PAAS) assume quite a lot of KTLO tasks but still rely on the customer to schedule them

Revisiting the Shared Responsibility Model



Managed services (SAAS) are much less flexible than other models but reduce operational overhead significantly

Question Breakdown

Stop spending money on undifferentiated heavy lifting

Question and Answer Choices

After migrating an on-premises application to EC2, a company needs to identify the method for accessing the instance operating systems as the fleet scales. The method must also scale with the number of instances and optimize for operational complexity.

Which recommendation would be appropriate to meet the requirement?

A. **Use Security Group rules and SSH**
B. **Use EC2 Serial Console**
C. **Use Systems Manager Run Command**
D. **Use Systems Manager Session Manager**

Answer Choice Explanation

This is functionally equivalent to the on-premises infrastructure, which also does not scale as it requires network connectivity to all instances.

A. Use Security Group rules and SSH
B. Use EC2 Serial Console
C. Use Systems Manager Run Command
D. Use Systems Manager Session Manager

Answer Choice Explanation

The EC2 Serial Console is only usable on a single instance, acting as a console connection. This may be useful for a non-bootable instance, but does not scale.

A. Use Security Group rules and SSH
B. Use EC2 Serial Console
C. Use Systems Manager Run Command
D. Use Systems Manager Session Manager

Answer Choice Explanation

Run Command can be deployed against an arbitrary number of instances whether they are on-premises or EC2. It also scales with the number of instances.

A. Use Security Group rules and SSH
B. Use EC2 Serial Console
C. Use Systems Manager Run Command
D. Use Systems Manager Session Manager

Answer Choice Explanation

Session Manager can replace SSH, without the same network connectivity requirements, which scales better, but it is not functionally scalable.

A. Use Security Group rules and SSH
B. Use EC2 Serial Console
C. Use Systems Manager Run Command
D. Use Systems Manager Session Manager

Correct Answer

Correct Answer: C

A. Use Security Group rules and SSH
B. Use EC2 Serial Console
C. Use Systems Manager Run Command
D. Use Systems Manager Session Manager

Question Domain 4: Design Cost-Optimized Architectures

Analyze and attribute expenditure

Principle Definition

The cloud makes it easier to accurately identify the cost and usage of workloads, which then allows transparent attribution of IT costs to revenue streams and individual workload owners.

This helps measure return on investment (ROI) and gives workload owners an opportunity to optimize their resources and reduce costs.

Configure detailed data sources

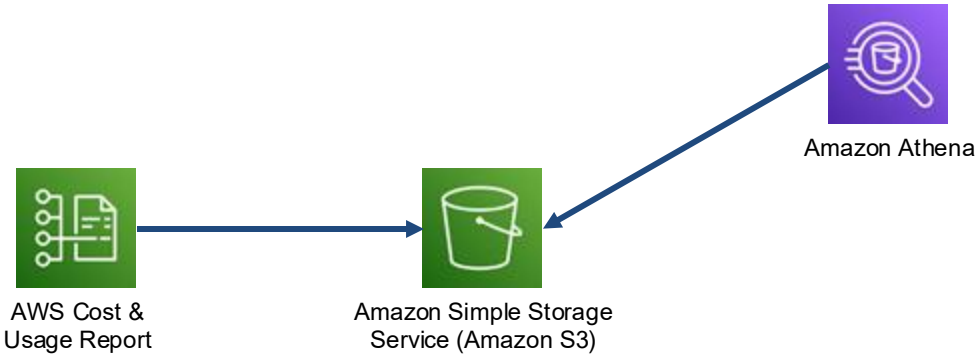
Generate CUR hourly with resource IDs

AWS Cost & Usage Report

Configure detailed data sources

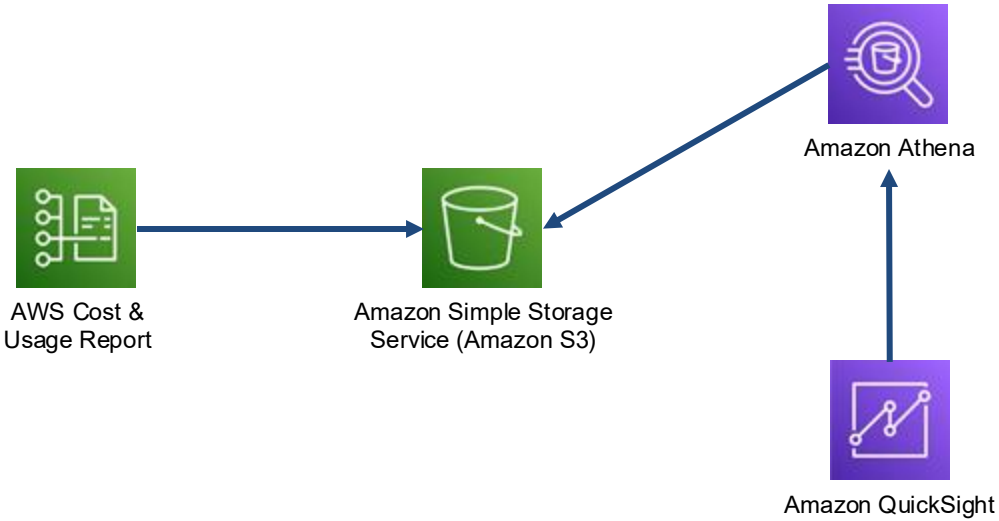
Deliver to S3 bucket for hourly CUR updates

Configure detailed data sources



Execute SQL queries against CUR using Athena

Configure detailed data sources



Generate visualization and reports using QuickSight

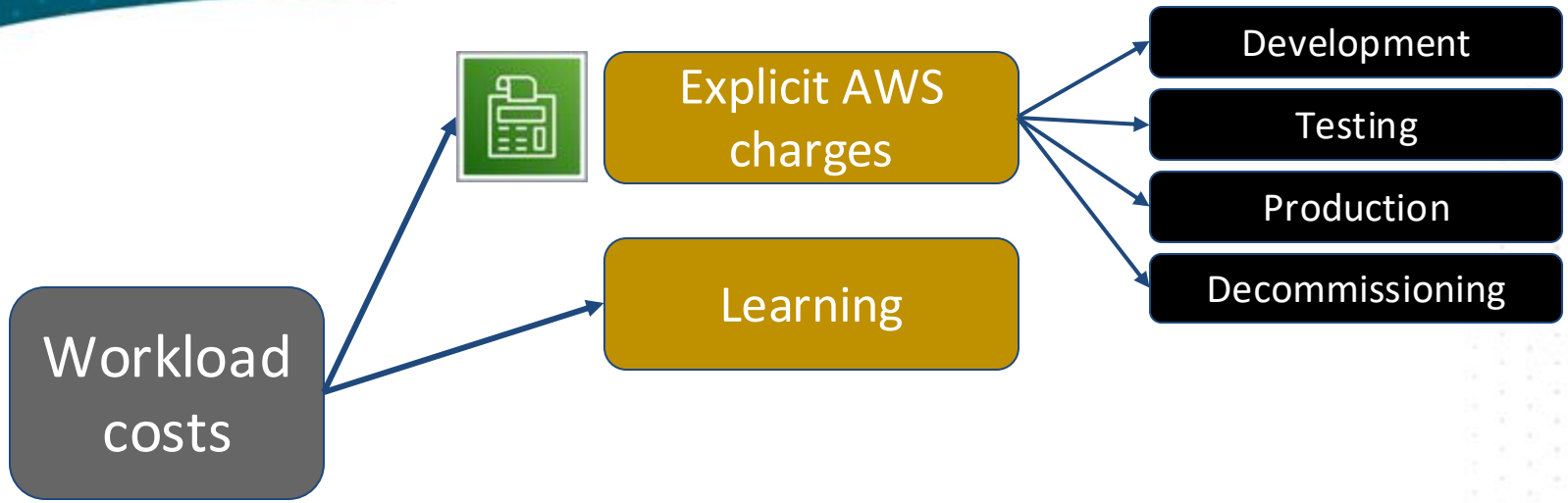
Identify cost attribution categories

Explicit AWS charges

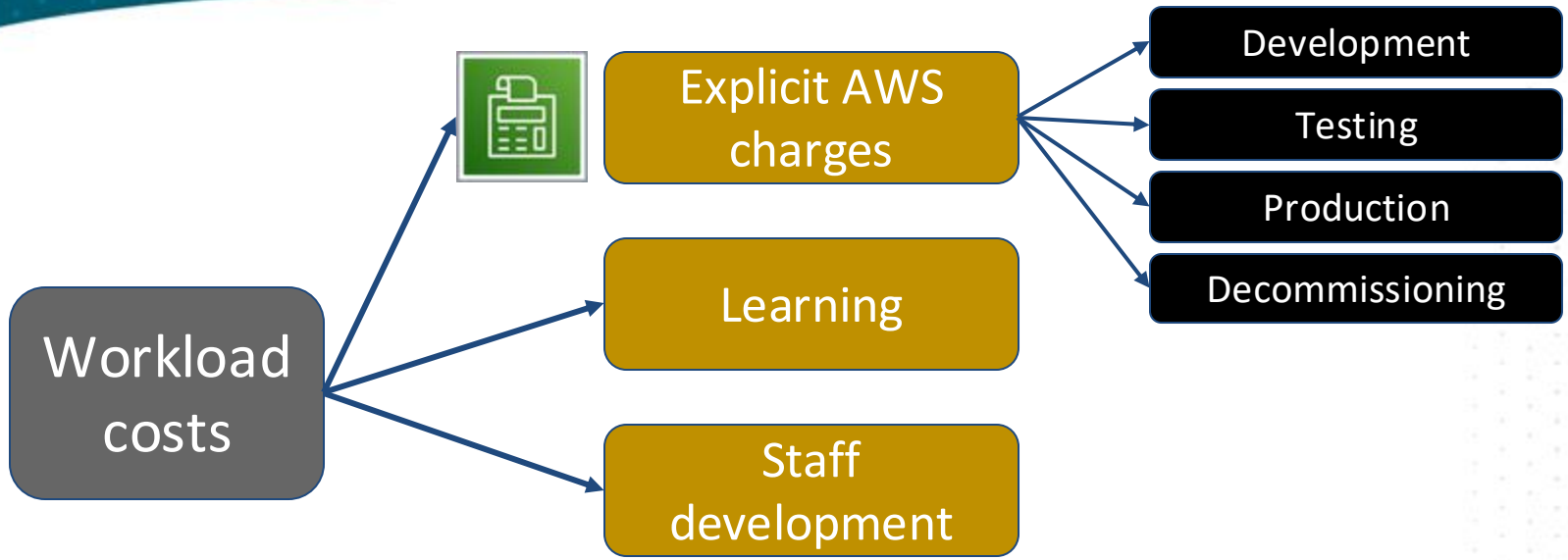
Workload costs

Identify cost attribution categories

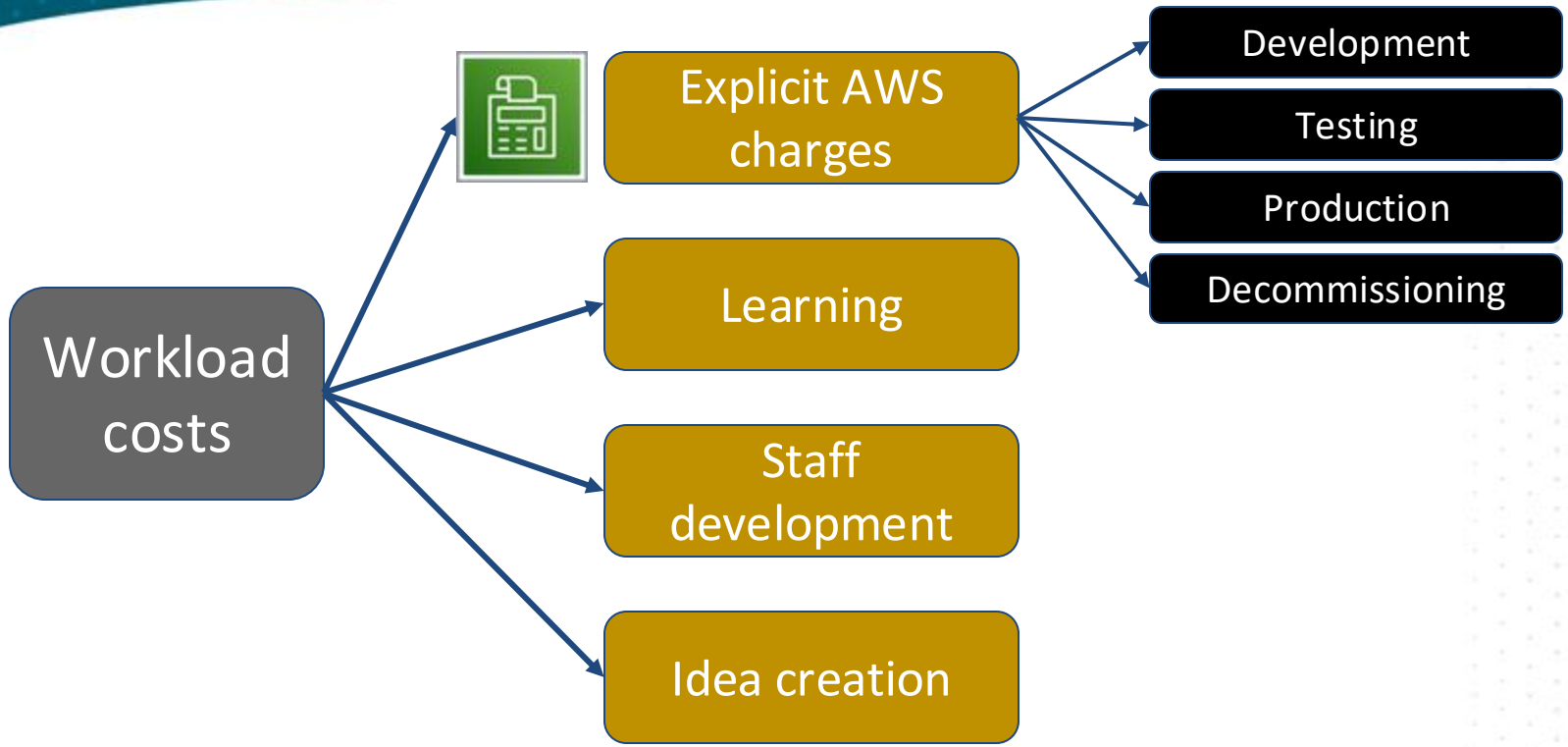
Identify cost attribution categories



Identify cost attribution categories



Identify cost attribution categories



Establish workload metrics

Bad KPI

- Average CPU
- Memory usage
- Network traffic

Establish workload metrics

Bad KPI

- Average CPU
- Memory usage
- Network traffic

Good KPI

- LB latency
- Overall uptime
- Requests/second

Establish workload metrics

Bad KPI

- Average CPU
- Memory usage
- Network traffic

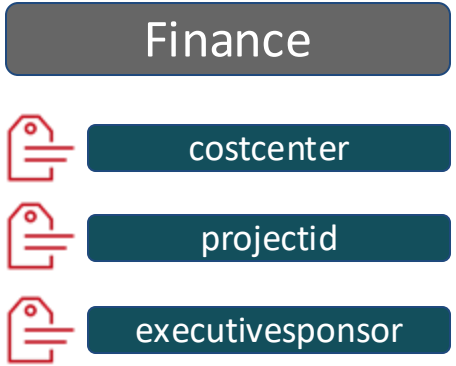
Good KPI

- LB latency
- Overall uptime
- Requests/second

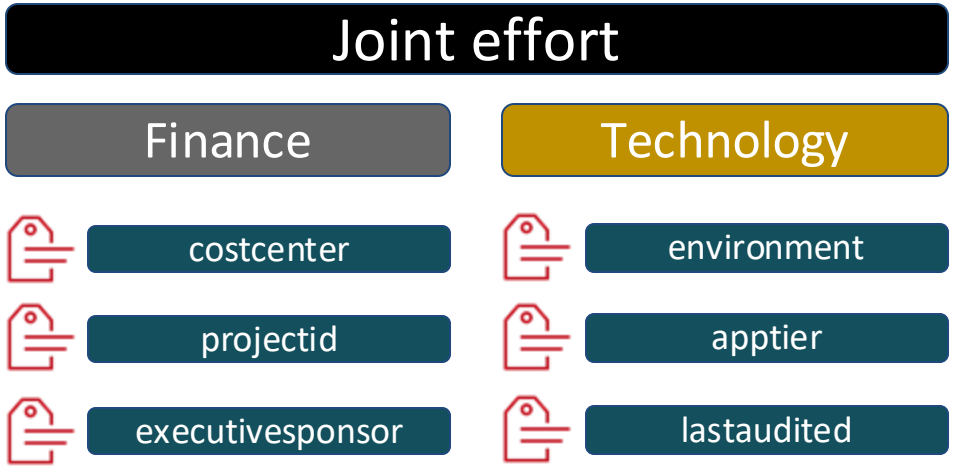
Better KPI

- Failed workflows
- Conversion rate
- Revenue per transaction

Assign organization meaning to cost and usage



Assign organization meaning to cost and usage



Assign organization meaning to cost and usage

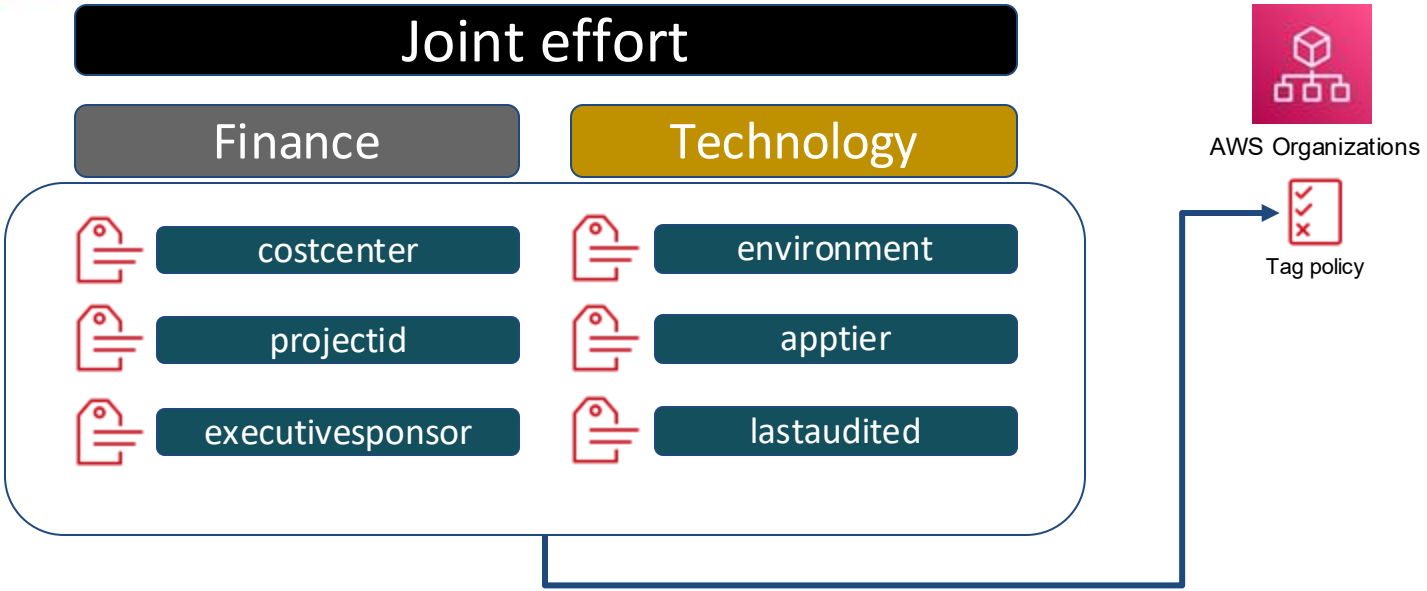
Joint effort

Finance	Technology
costcenter	environment
projectid	apptier
executivesponsor	lastaudited

AWS Organizations

Tag policy

Assign organization meaning to cost and usage



Configure billing and cost optimization tools

Reports	Summary of all cost and usage information

Configure billing and cost optimization tools

Reports	Summary of all cost and usage information
Notifications	Provide notifications when cost or usage is outside of defined limits

Configure billing and cost optimization tools

Reports	Summary of all cost and usage information
Notifications	Provide notifications when cost or usage is outside of defined limits
Current State	Configure a dashboard showing current levels of cost and usage

Configure billing and cost optimization tools

Reports	Summary of all cost and usage information
Notifications	Provide notifications when cost or usage is outside of defined limits
Current State	Configure a dashboard showing current levels of cost and usage
Trending	Show the variability in cost and usage over the required period of time

Configure billing and cost optimization tools

Reports	Summary of all cost and usage information
Notifications	Provide notifications when cost or usage is outside of defined limits
Current State	Configure a dashboard showing current levels of cost and usage
Trending	Show the variability in cost and usage over the required period of time
Forecasts	Provide the capability to show estimated future costs

Configure billing and cost optimization tools

Reports	Summary of all cost and usage information
Notifications	Provide notifications when cost or usage is outside of defined limits
Current State	Configure a dashboard showing current levels of cost and usage
Trending	Show the variability in cost and usage over the required period of time
Forecasts	Provide the capability to show estimated future costs
Tracking	Show the current cost and usage against configured goals or targets

Configure billing and cost optimization tools

Reports	Summary of all cost and usage information
Notifications	Provide notifications when cost or usage is outside of defined limits
Current State	Configure a dashboard showing current levels of cost and usage
Trending	Show the variability in cost and usage over the required period of time
Forecasts	Provide the capability to show estimated future costs
Tracking	Show the current cost and usage against configured goals or targets
Analysis	Provide the capability for team members to perform custom and deep analysis down to the hourly granularity with appropriate dimensions

Question Breakdown

Analyze and attribute expenditure

Question and Answer Choices

Your company's application is using several custom AWS tags that your accounting department would like to generate reports from. The application runs in several AWS accounts that are all part of the same AWS Organization.

What action can ensure that the AWS tags are available as part of the AWS monthly bill?

A. Enable User-defined Cost Allocation Tags for each of the AWS tags in the AWS Organizations Management account
B. Enable User-defined Cost Allocation Tags for each of the AWS tags in all of the AWS Organizations member accounts
C. Enable AWS-managed Cost Allocation Tags for each of the AWS tags in the AWS Organizations Management account
D. No action is required, all tags are automatically made available to the Cost Explorer tool and Cost and Usage Reports

Answer Choice Explanation

AWS tags are not automatically converted into Cost Allocation Tags, and so this task must be done manually, which makes this task not only appropriate, but required.

A. Enable User-defined Cost Allocation Tags for each of the AWS tags in the AWS Organizations Management account
B. Enable User-defined Cost Allocation Tags for each of the AWS tags in all of the AWS Organizations member accounts
C. Enable AWS-managed Cost Allocation Tags for each of the AWS tags in the AWS Organizations Management account
D. No action is required, all tags are automatically made available to the Cost Explorer tool and Cost and Usage Reports

Answer Choice Explanation

Similar to A, the task must be performed manually, but it is not required for all member accounts, just the Management account.

A. Enable User-defined Cost Allocation Tags for each of the AWS tags in the AWS Organizations Management account
B. Enable User-defined Cost Allocation Tags for each of the AWS tags in all of the AWS Organizations member accounts
C. Enable AWS-managed Cost Allocation Tags for each of the AWS tags in the AWS Organizations Management account
D. No action is required, all tags are automatically made available to the Cost Explorer tool and Cost and Usage Reports

Answer Choice Explanation

The custom tags are not available as AWS-managed Cost Allocation Tags, and so this solution would not be plausible.

A. Enable User-defined Cost Allocation Tags for each of the AWS tags in the AWS Organizations Management account
B. Enable User-defined Cost Allocation Tags for each of the AWS tags in all of the AWS Organizations member accounts
C. Enable AWS-managed Cost Allocation Tags for each of the AWS tags in the AWS Organizations Management account
D. No action is required, all tags are automatically made available to the Cost Explorer tool and Cost and Usage Reports

Answer Choice Explanation

Tags must be enabled as Cost Allocation Tags manually, and so this solution will not work.

A. Enable User-defined Cost Allocation Tags for each of the AWS tags in the AWS Organizations Management account
B. Enable User-defined Cost Allocation Tags for each of the AWS tags in all of the AWS Organizations member accounts
C. Enable AWS-managed Cost Allocation Tags for each of the AWS tags in the AWS Organizations Management account
D. No action is required, all tags are automatically made available to the Cost Explorer tool and Cost and Usage Reports

Correct Answer

Correct Answer: A

A. Enable User-defined Cost Allocation Tags for each of the AWS tags in the AWS Organizations Management account
B. Enable User-defined Cost Allocation Tags for each of the AWS tags in all of the AWS Organizations member accounts
C. Enable AWS-managed Cost Allocation Tags for each of the AWS tags in the AWS Organizations Management account
D. No action is required, all tags are automatically made available to the Cost Explorer tool and Cost and Usage Reports

AWS Solutions Architect - Associate Certification Crash Course

Q&A