Network Professional Training center

Providing Job role training in one of fastest growing IT Jobs Sector

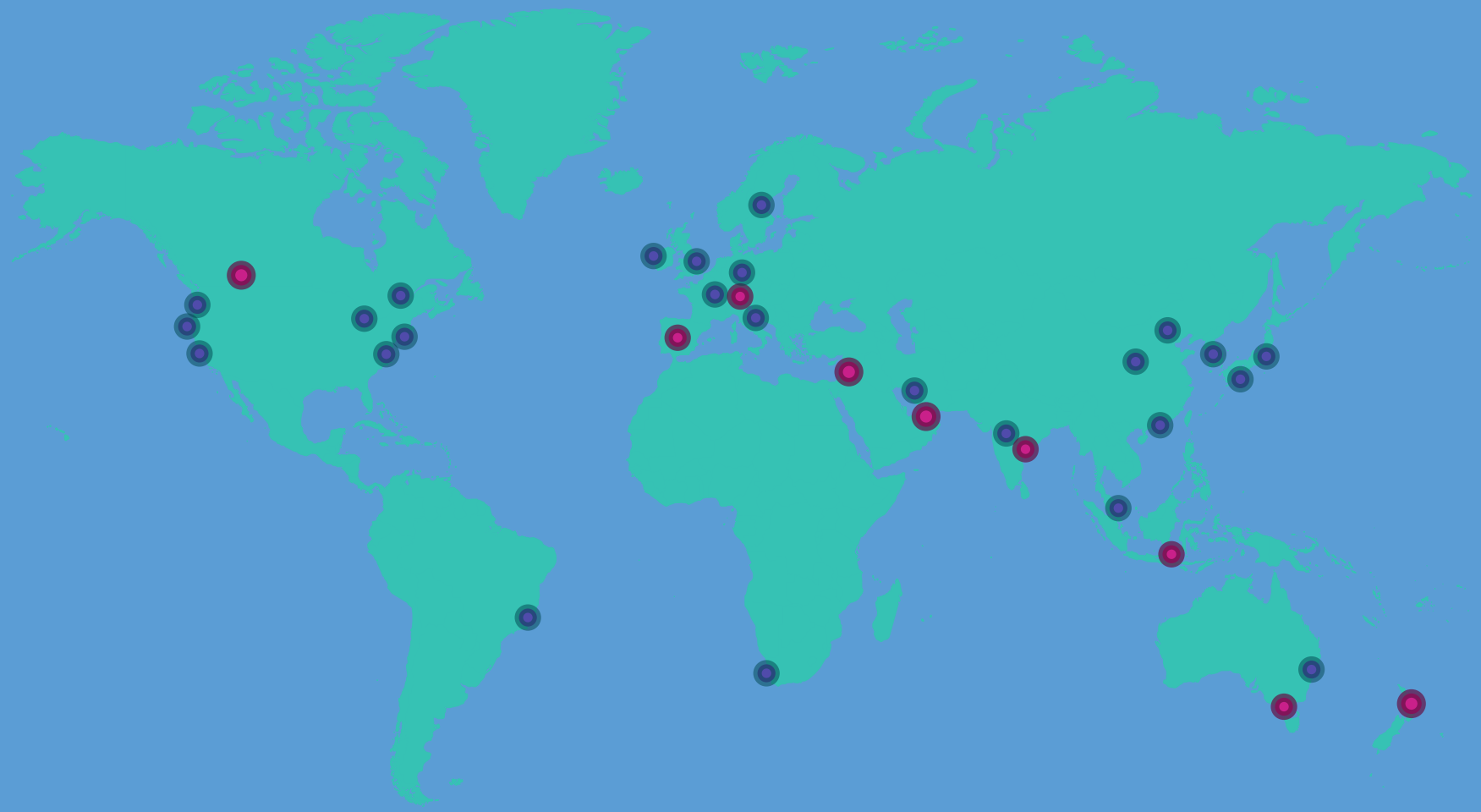
AWS Associate Guide_1

AWS services

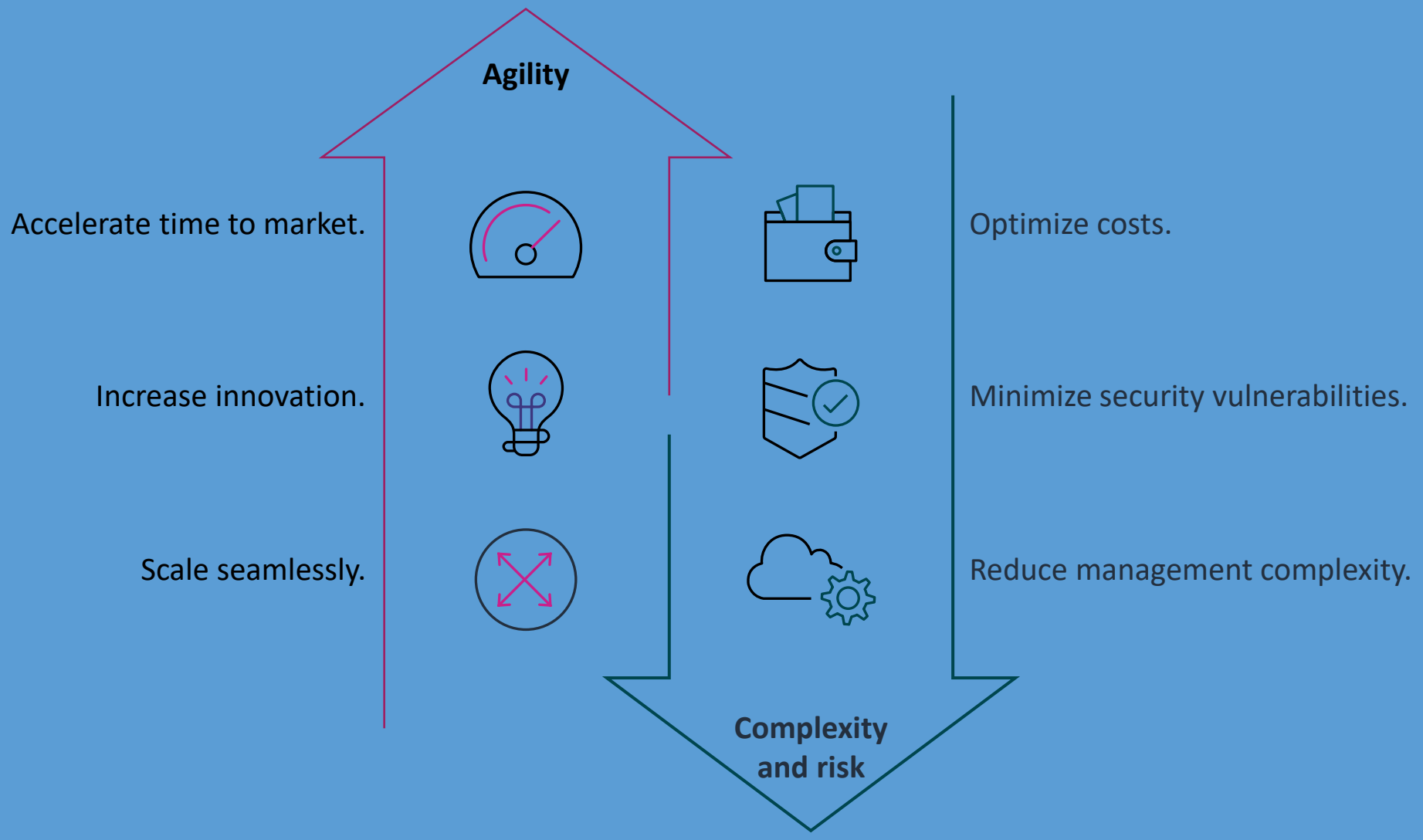
“What are the benefits of using AWS services?”

Amazon Web Services

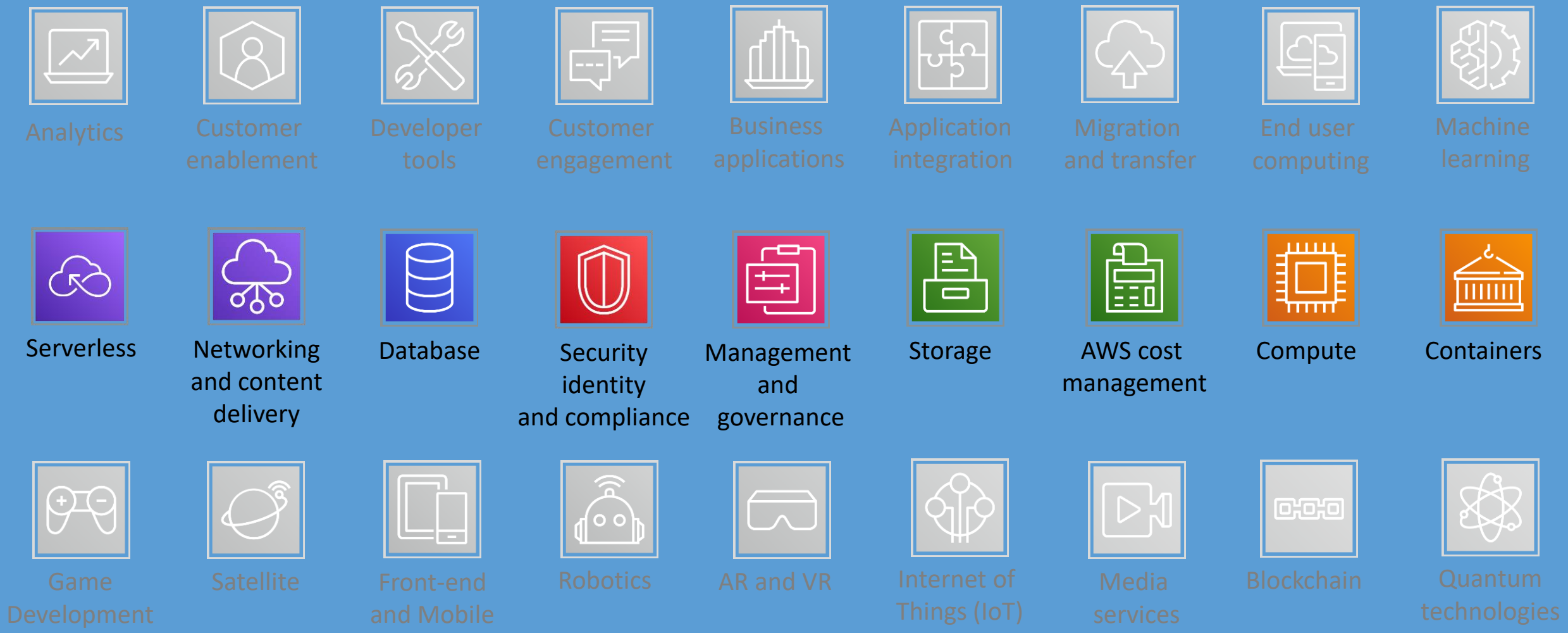
- Global data centers
- More than 200 services
- Secure and robust
- Pay as you go
- Built for business needs



Why customers move to AWS



AWS service categories



AWS infrastructure

“How is AWS global infrastructure organized?”

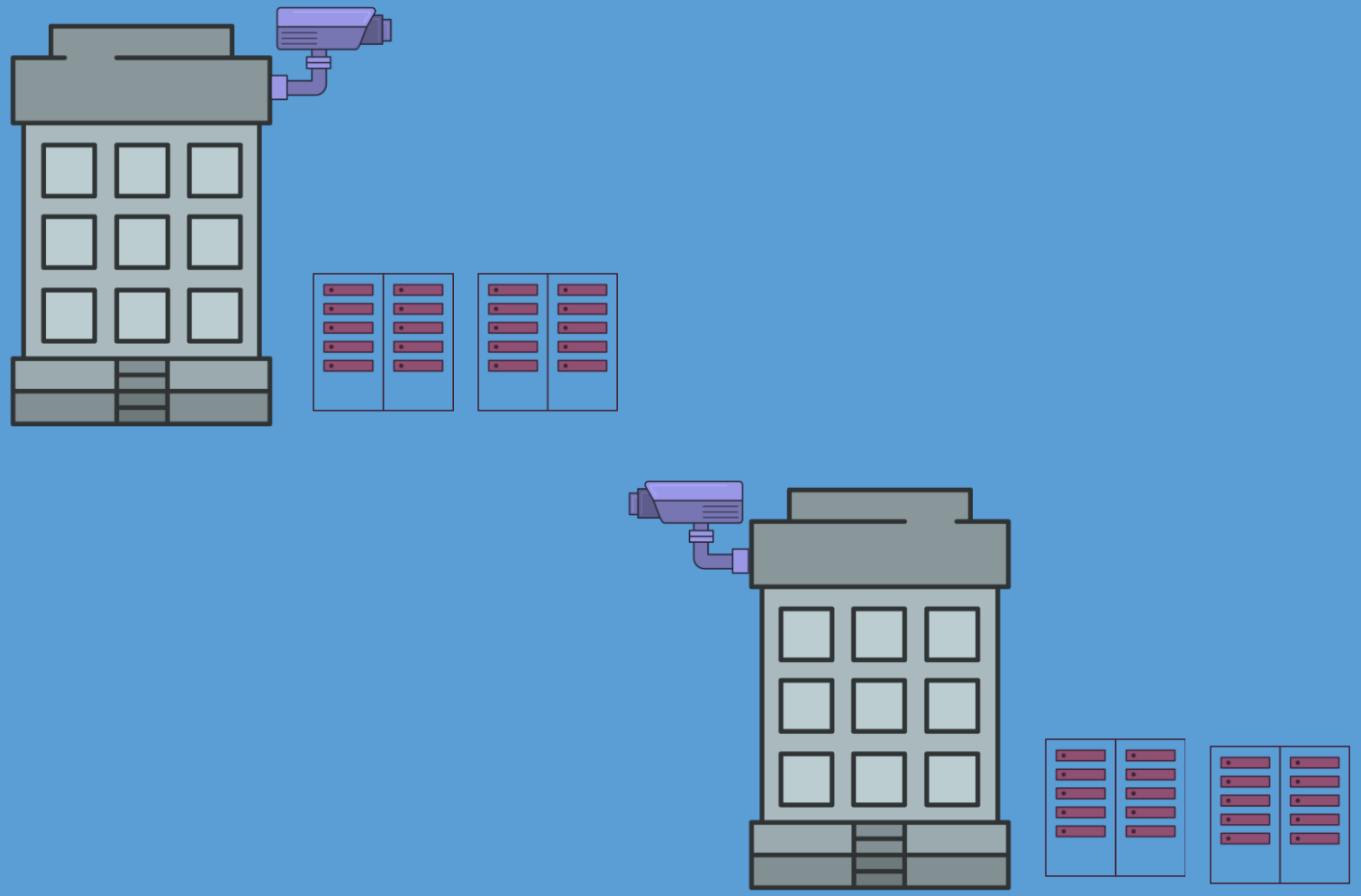
AWS infrastructure topics

AWS infrastructure

- Data centers
- Availability Zones
- Regions
- AWS Local Zones
- Edge locations

AWS data centers

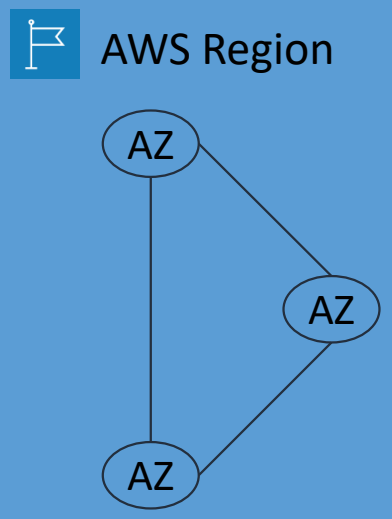
- AWS services operate within AWS data centers.
- Data centers host thousands of servers.
- Each location uses AWS proprietary network equipment.
- Data centers are organized into Availability Zones.

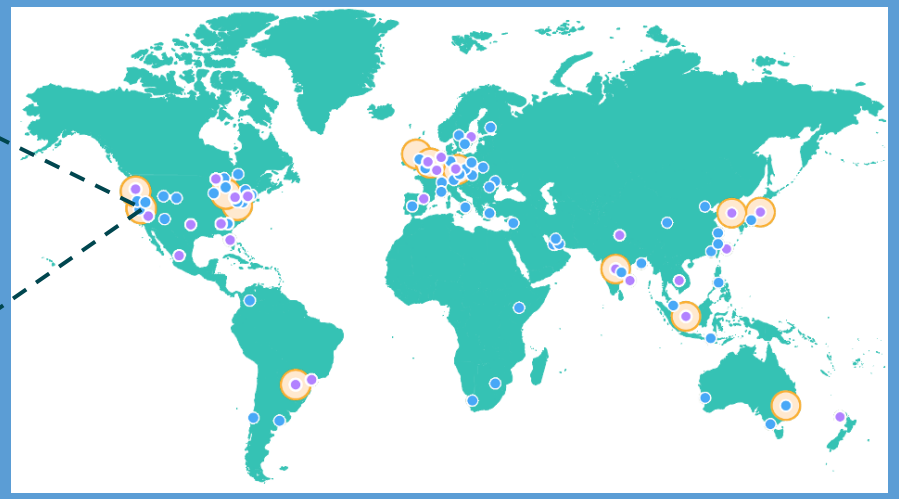


Availability Zones (AZs)

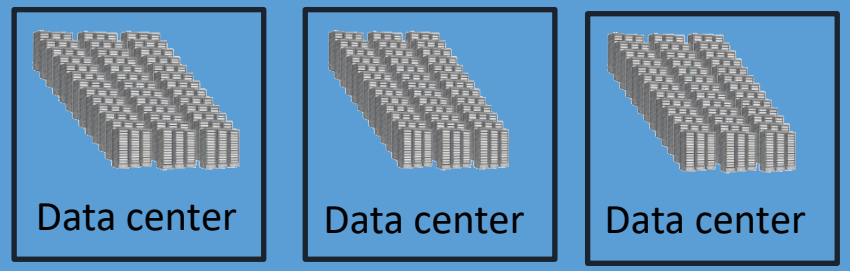
Availability Zones are:

- Data centers in a Region
- Designed for fault isolation
- Interconnected using high-speed private links
- Used to achieve high availability





Availability Zone



AWS Regions

Each Region:

- Is completely independent
- Uses AWS network infrastructure
- Has multiple Availability Zones

Factors impacting Region selection

Governance

Latency

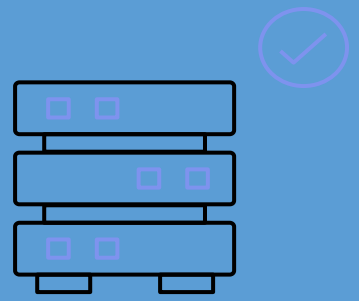
Service availability

Cost

AWS Local Zones

Use cases:

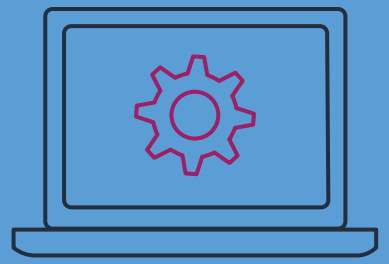
- Media and entertainment content creation
- Real-time gaming
- Machine learning inference
- Live video streaming
- Augmented reality (AR) and virtual reality (VR)



AWS infrastructure at the edge

Local compute, storage, databases, and other services

Connecting to services in AWS Regions

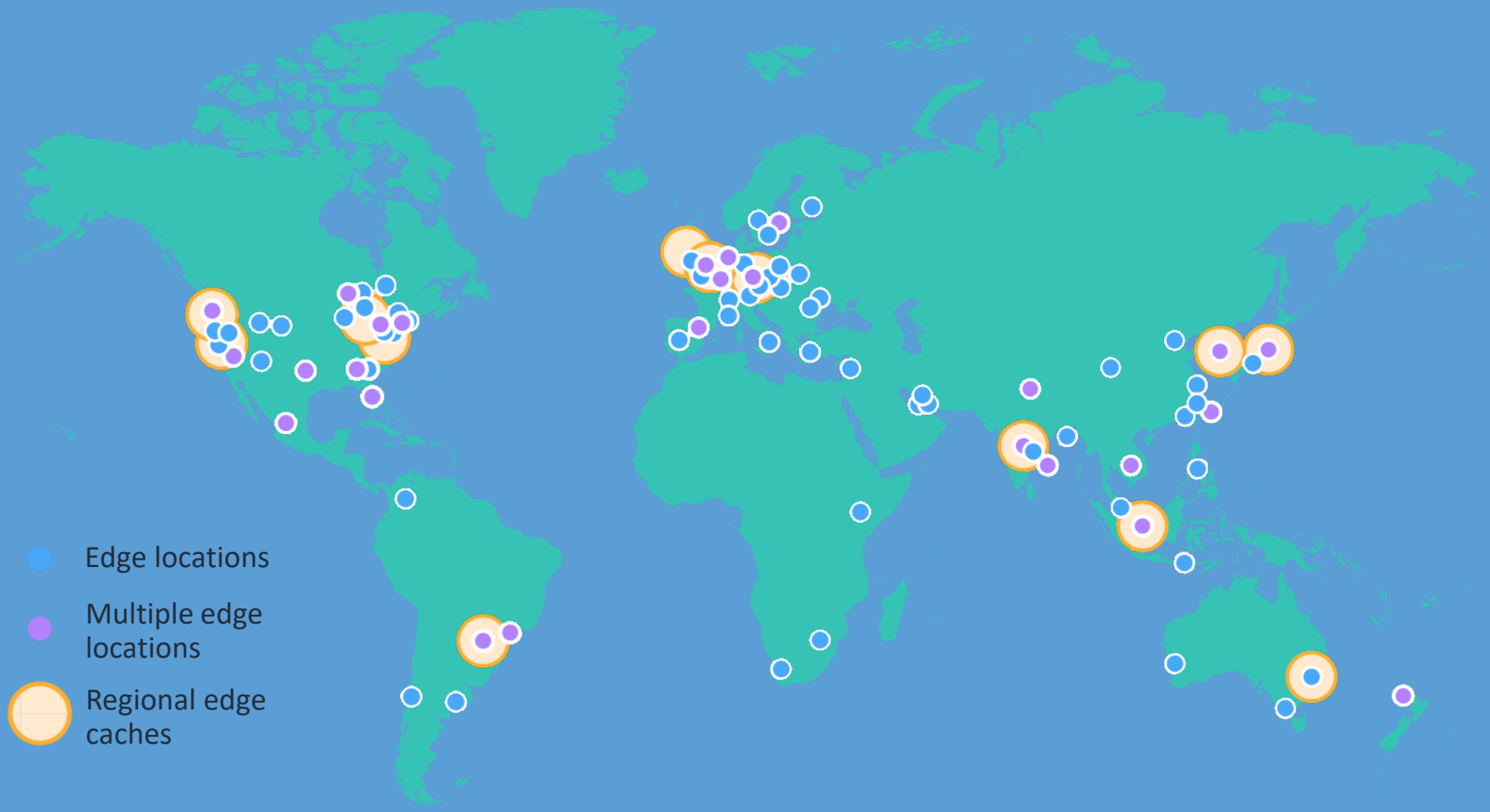


Delivering new low-latency applications

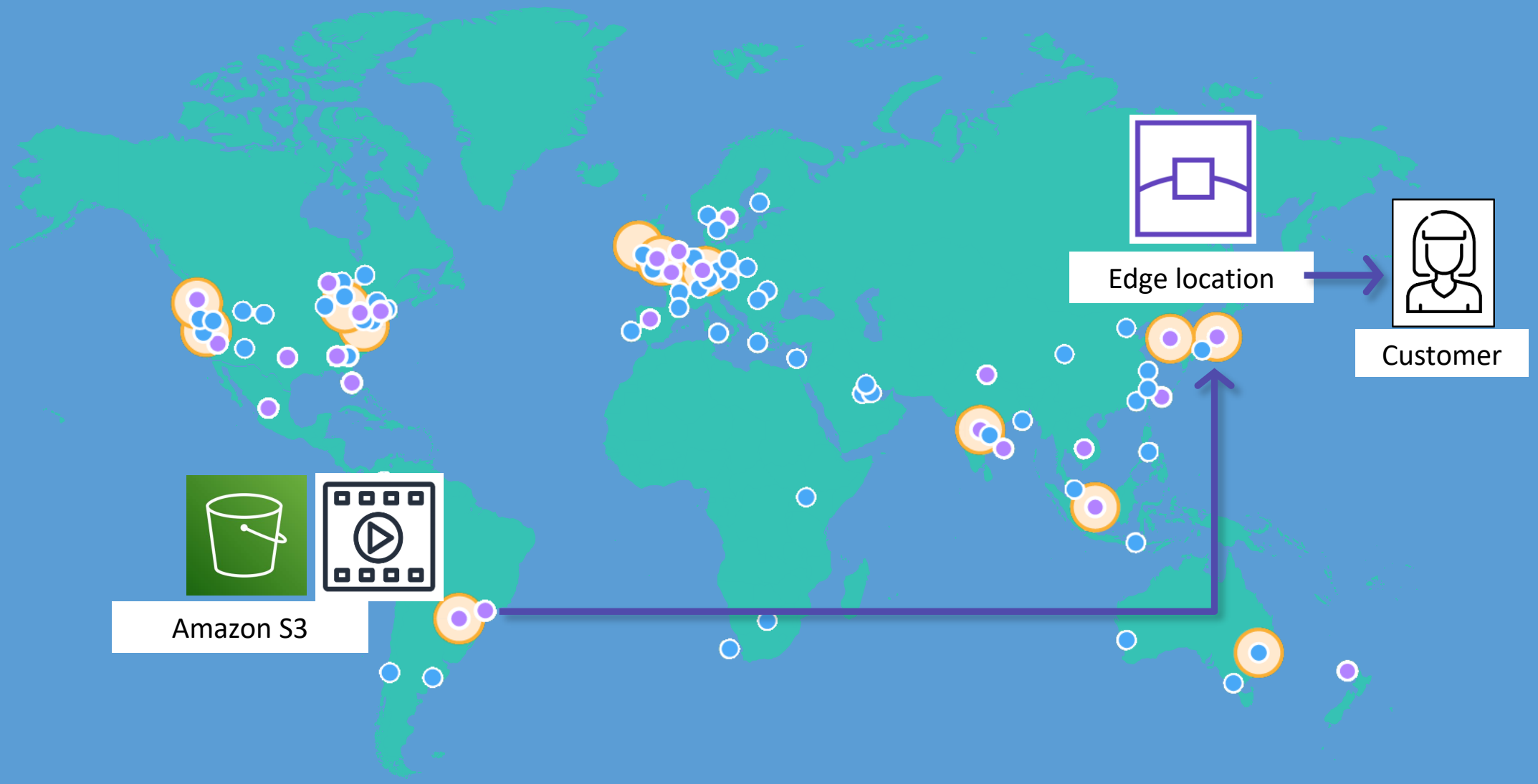
Edge locations

Edge locations:

- Run in major cities around the world
- Support AWS services like Amazon Route 53 and Amazon CloudFront



Edge location use case



AWS Local Zone and edge location features

AWS Local Zones

- Low latency
- Local data processing
- Consistent AWS experience

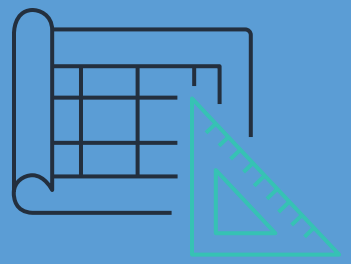
Edge Locations

- Caching of data
- Fast delivery of content
- Better user experience

AWS Well-Architected Framework

“How can we build our cloud infrastructure according to best practices?”

AWS architect responsibilities



Plan

- Set technical cloud strategy with business leads.
- Analyze solutions for business needs and requirements.

Research

- Investigate cloud services specs and workload requirements.
- Review existing workload architectures.
- Design prototype solutions.

Build

- Design the transformation roadmap with milestones, work streams, and owners.
- Manage the adoption and migration.

AWS Well-Architected Framework pillars

Security

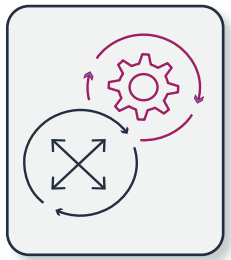
- Apply at all layers
- Enforce the principle of least privilege
- Use multi-factor authentication (MFA)

Performance Efficiency

- Reduce latency
- Use serverless architecture
- Incorporate monitoring

Cost Optimization

- Analyze and attribute expenditures
- Use cost-effective resources
- Stop guessing

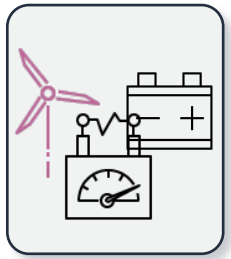


Operational Excellence

- Perform operations with code
- Test response for unexpected events

Reliability

- Recover from failure
- Test recovery procedures
- Scale to increase availability

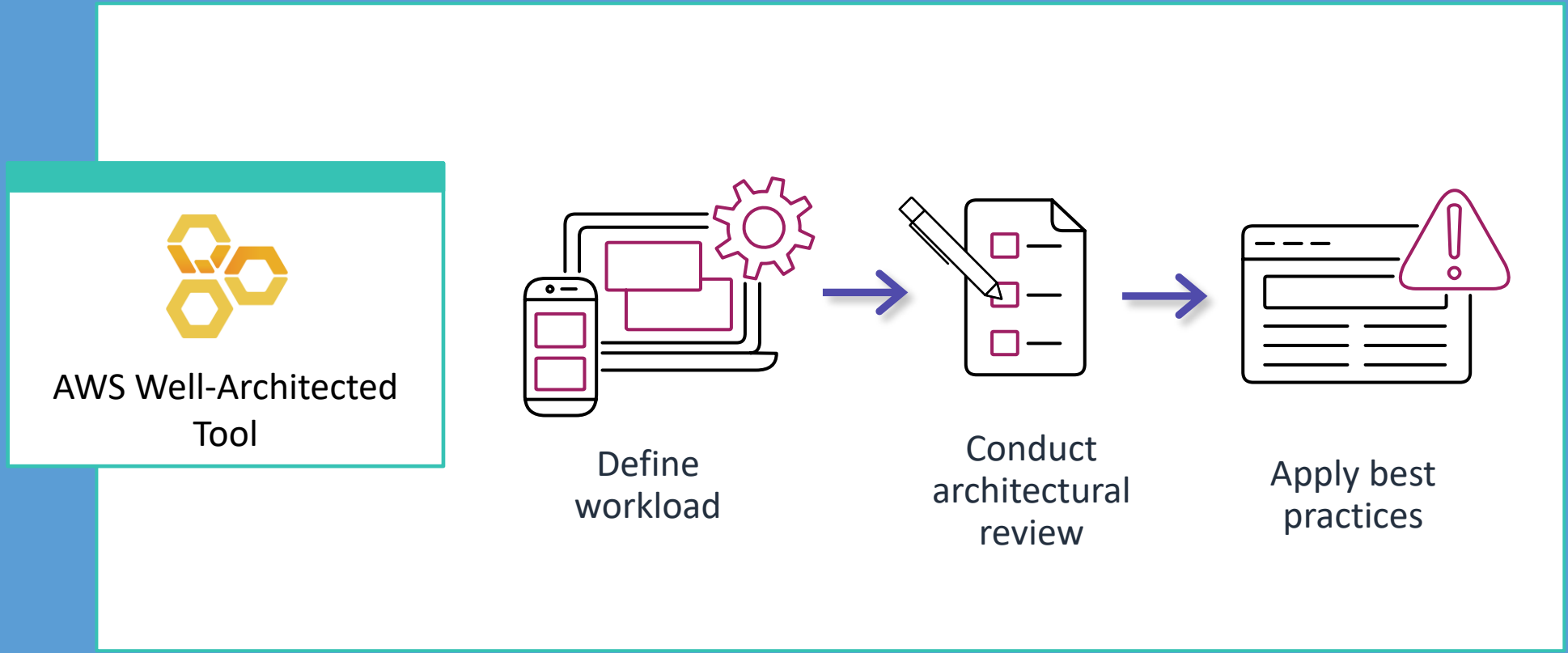


Sustainability

- Understand your impact
- Maximize utilization

AWS Well-Architected Tool

- Based on the AWS Well-Architected Framework
- Can review your applications and workloads
- Central place for best practices and guidance
- Used in tens of thousands of workload reviews



Review

Present solutions

Chief Technology Officer

Consider how you would answer the following questions:

- What are the benefits of using AWS services?
- How is the AWS global infrastructure organized?
- How can we build our cloud infrastructure according to best practices?

Module review

In this module you learned about:

- ✓ AWS services
- ✓ AWS infrastructure
- ✓ AWS Well-Architected Framework

Next, you will review:

- Knowledge check
- Lab introduction

Knowledge check

Knowledge check question 1

Which of the following is the best example of one responsibility of an AWS architect?

A	Monitor alarms for disaster response.
B	Maintain application-level code in the AWS Cloud.
C	Manage access to a group of AWS accounts.
D	Analyze solutions for business needs and requirements.

Knowledge check question 1 and answer

Which of the following is the best example of one responsibility of an AWS architect?

A	Monitor alarms for disaster response.
B	Maintain application-level code in the AWS Cloud.
C	Manage access to a group of AWS accounts.
D correct	Analyze solutions for business needs and requirements.

Knowledge check question 2

Which of the following is a cluster of data centers within a geographic location with low latency network connectivity?

A	Availability Zone
B	Region
C	Edge location
D	Outposts

Knowledge check question 2 and answer

Which of the following is a cluster of data centers within a geographic location with low latency network connectivity?

A correct	Availability Zone
B	Region
C	Edge location
D	Outposts

Knowledge check question 3

Which of the following factors do you consider when picking an AWS Region? (Select TWO.)

A	Local data regulations
B	Operating system requirements
C	Latency to end users
D	Support for hybrid networking
E	Programming language of your application

Knowledge check question 3 and answer

Which of the following factors do you consider when picking an AWS Region? (Select TWO.)

A correct	Local data regulations
B	Operating system requirements
C correct	Latency to end users
D	Support for hybrid networking
E	Programming language of your application

Knowledge check question 4

What is the primary benefit of deploying your applications into multiple Availability Zones?

A	Stronger security policies for resources
B	Decreased latency to resources
C	High availability for resources
D	There is no benefit to this design

Knowledge check question 4 and answer

What is the primary benefit of deploying your applications into multiple Availability Zones?

A	Stronger security policies for resources
B	Decreased latency to resources
C correct	High availability for resources
D	There is no benefit to this design

Knowledge check question 5

The principle of least privilege is a principle under which Well-Architected Framework pillar?

A	Operational excellence
B	Security
C	Resilience
D	Performance efficiency

Knowledge check question 5 and answer

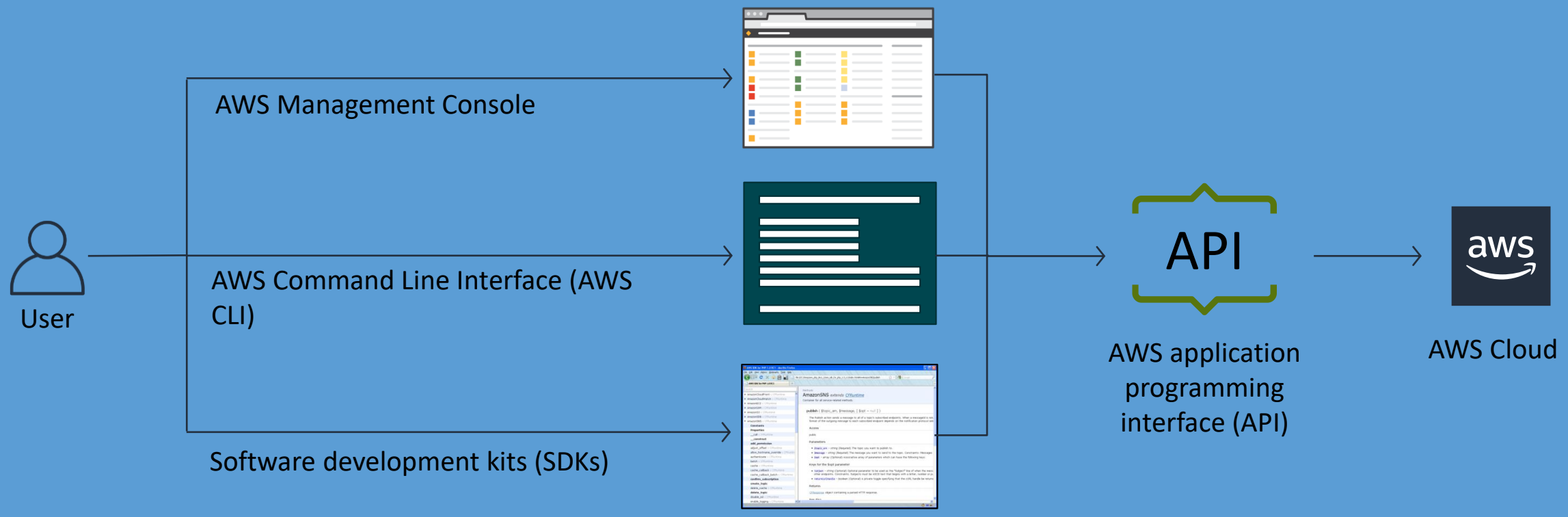
The principle of least privilege is a principle under which Well-Architected Framework pillar?

A	Operational excellence
B correct	Security
C	Resilience
D	Performance efficiency

Lab 1:

Explore and interact with the AWS Management Console and AWS Command Line Interface

Connecting to an AWS service



AWS

Account Security

Module overview

- Business requests
- Principals and identities
- Security policies
- Managing multiple accounts
- Module review
- Knowledge check

Business Requirements

Security Specialist

The security specialist needs to know:

- What are the best practices to manage access to AWS accounts and resources?
- How can we give users access to only the resources they need?
- What is the best way to manage multiple accounts?

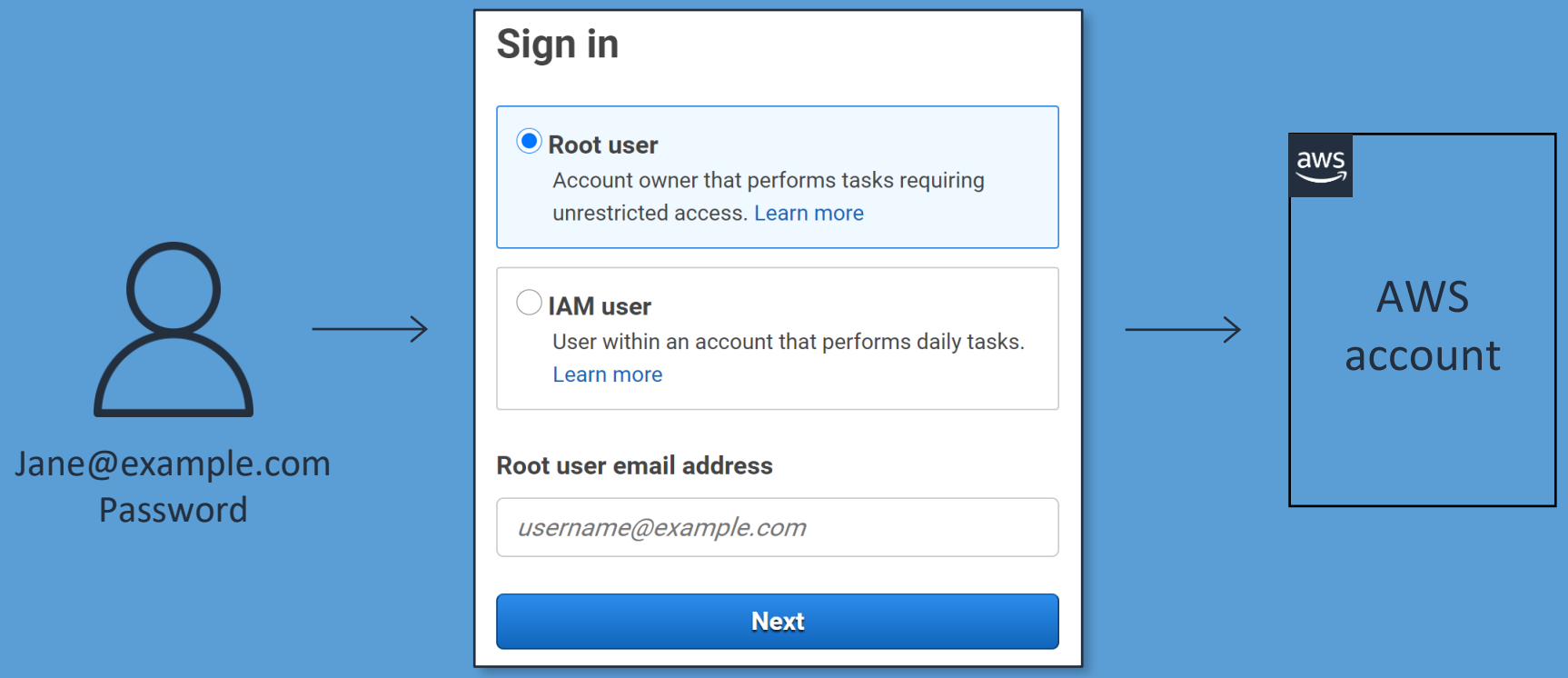
Principals and identities

“What are the best practices to manage access to AWS accounts and resources?”

AWS account root user

A root user:

- Has full access to all AWS services
- Cannot be restricted in a single account model
- Should not be used for day-to-day interactions with AWS



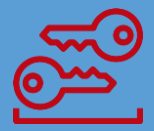
AWS Identity and Access Management (IAM)

Use IAM to:

- Create and manage users, groups, and roles.
- Manage access to AWS services and resources.
- Analyze access controls.

IAM

Authentication



Credentials

Sign in to AWS

Authorization

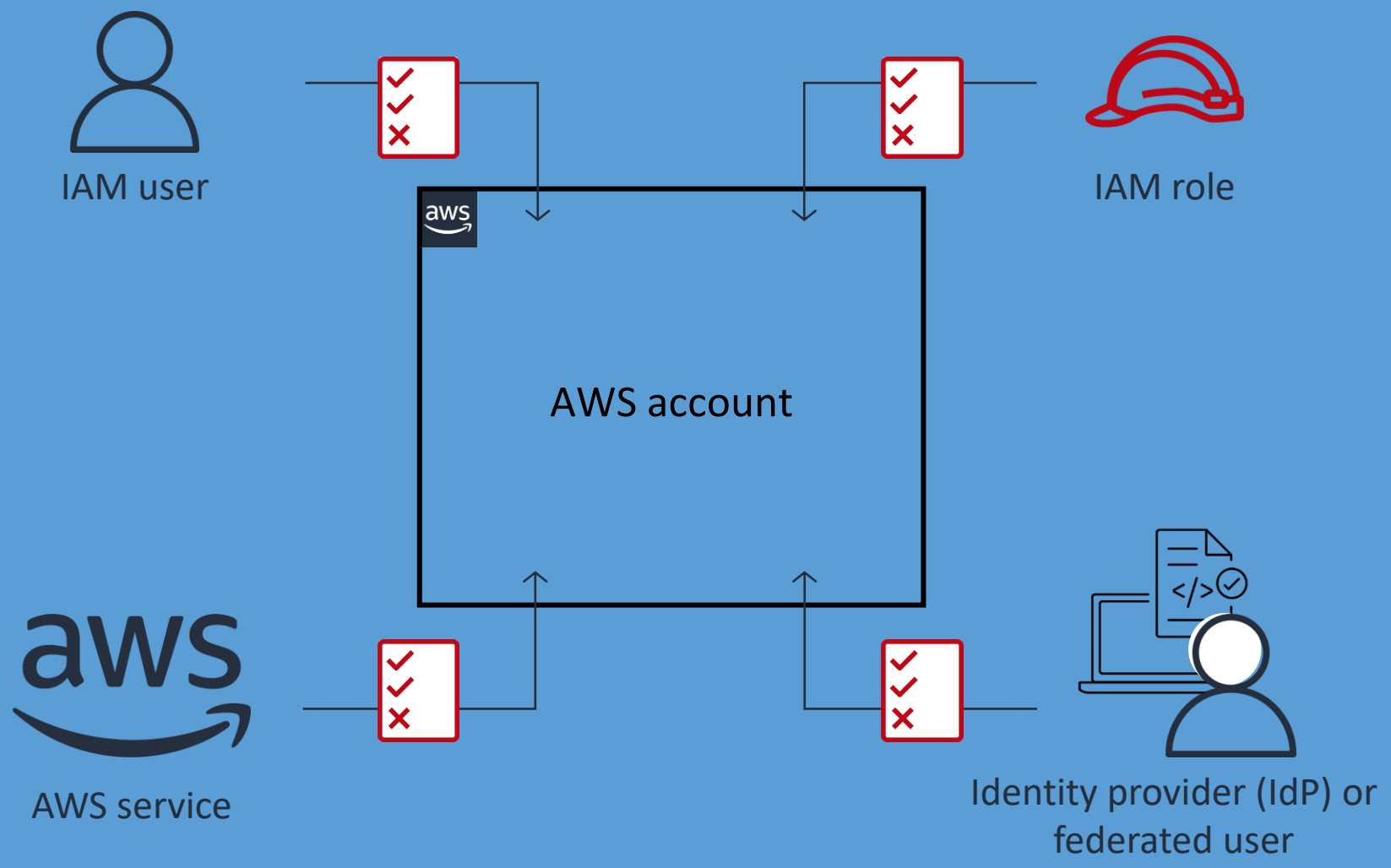
Permissions

Allowed to carry out request

Principals

A principal:

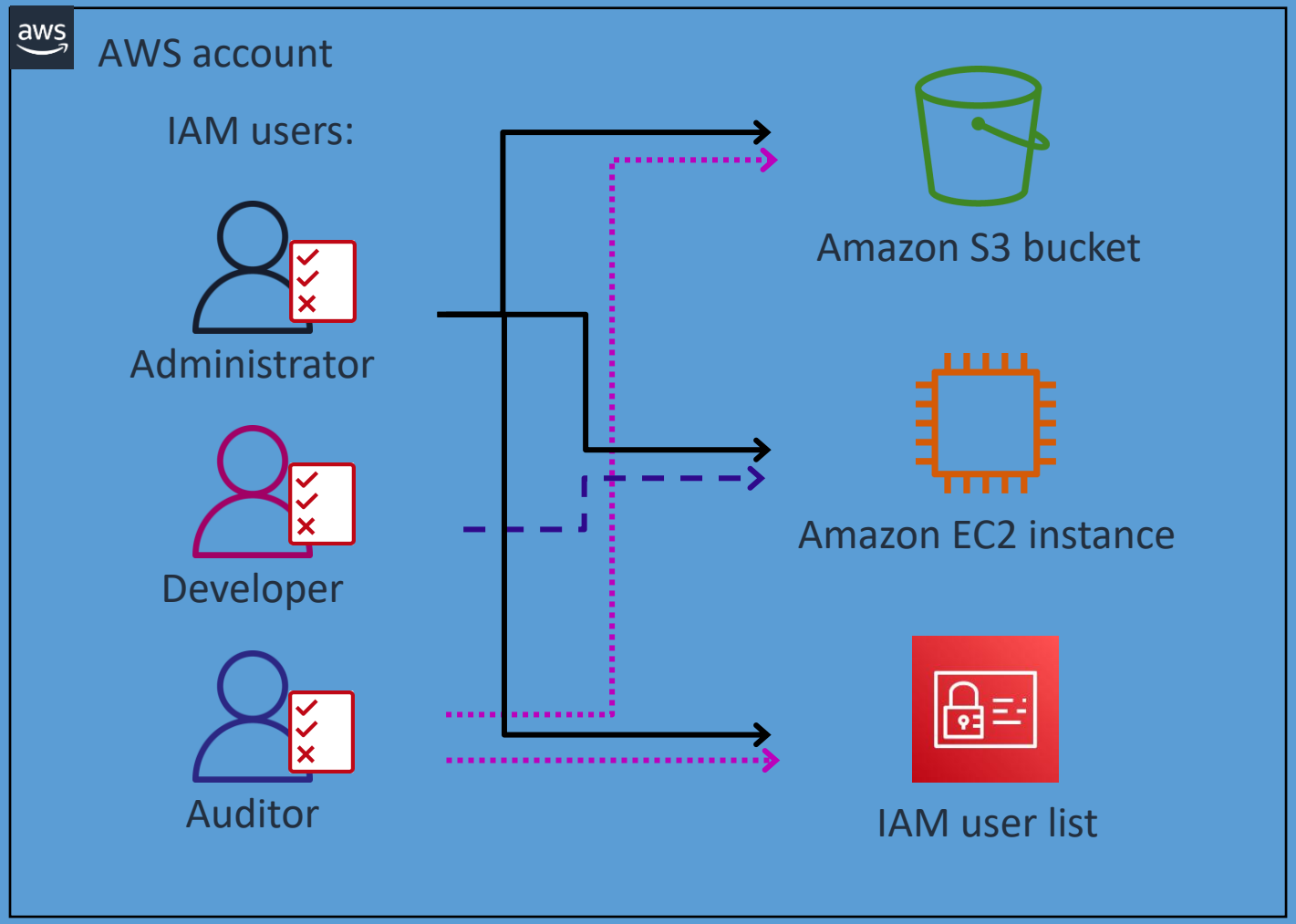
- Can make a request for an action or operation on an AWS resource
- Can be a person, application, federated user, or assumed role



IAM users

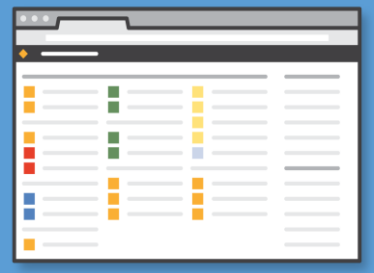
IAM users are users within an AWS account.

- Each user has their own credentials.
- They are authorized to perform specific AWS actions based on permissions.



IAM users and AWS API calls

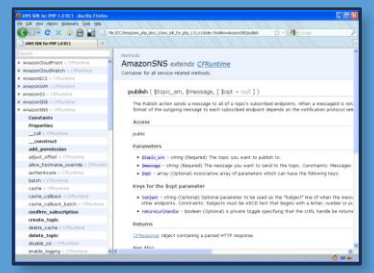
Console Access



AWS Management Console

Programmatic Access

AWS Command Line Interface (AWS CLI)



AWS SDKs

Programmatic access

IAM user

```
Access Key ID: AKIAIOSFODNN7EXAMPLE
Secret Access Key: wJalrXUtnFEMI/K7MDENG/bPxRfiCYEXAMPLEKEY
```

AWS CLI

```
$ aws configure
AWS Access Key ID [****************MPLE]:
AWS Secret Access Key [************EKEY]:
Default region name [ap-southeast-1]:
Default output format [json]:
```

AWS SDK

Java

Python

.NET

Setting permissions with IAM policies

IAM policy

Select	Policy name
	AdministratorAccess
	AmazonEC2ReadOnlyAccess
✓	AmazonS3FullAccess
	AmazonS3ReadOnlyAccess

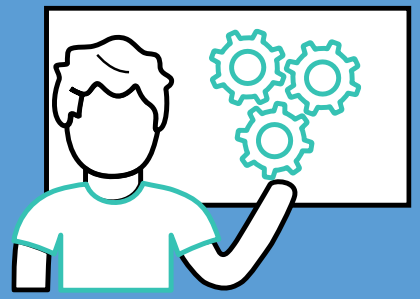
Amazon S3 administrator

Select	Policy name
	AdministratorAccess
✓	AmazonEC2ReadOnlyAccess
	AmazonS3FullAccess
✓	AmazonS3ReadOnlyAccess

Auditor

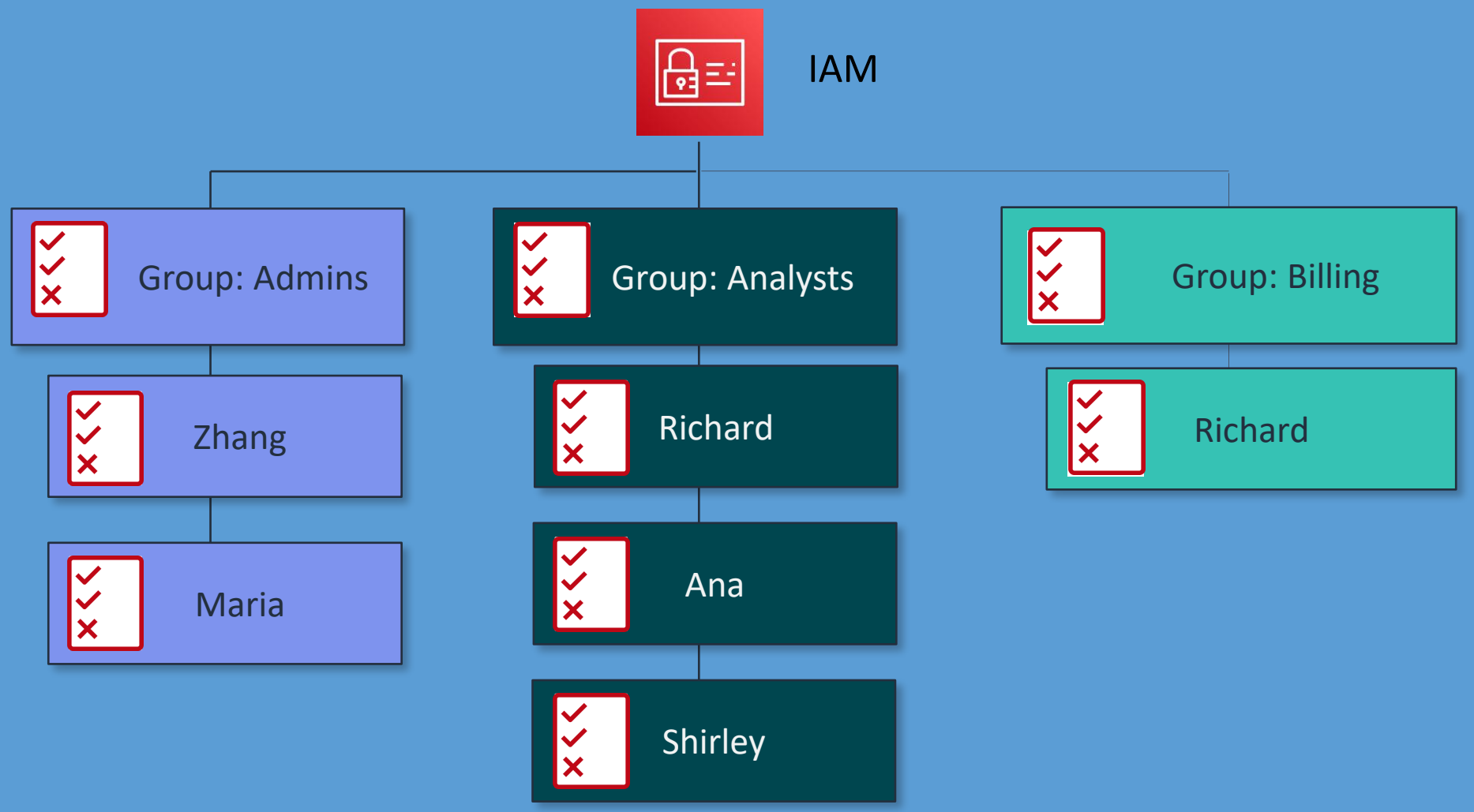
Demonstration:

Create an IAM user



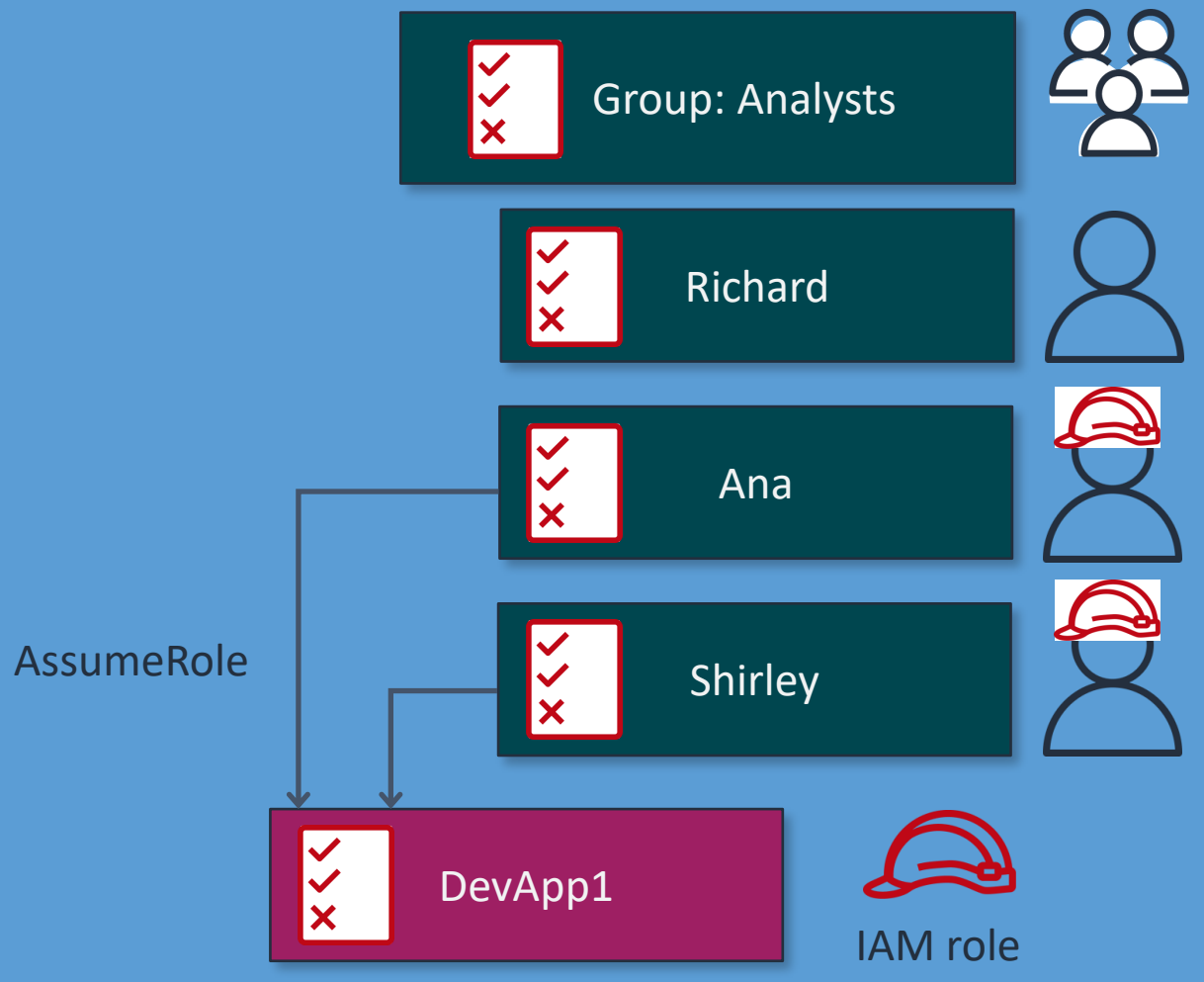
IAM user groups

- Assign IAM users to an IAM user group.
- Attach policies to an IAM user group to apply to all users within the group.



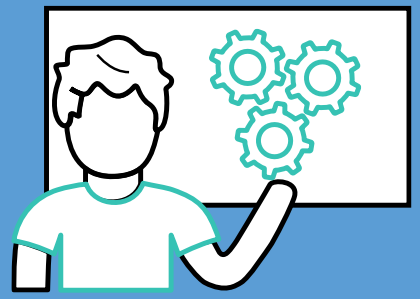
IAM roles

- Delegate set permissions to specific users or services.
- Users assume a role without sharing credentials with others.
- Permissions are only valid while operating under the assumed role.

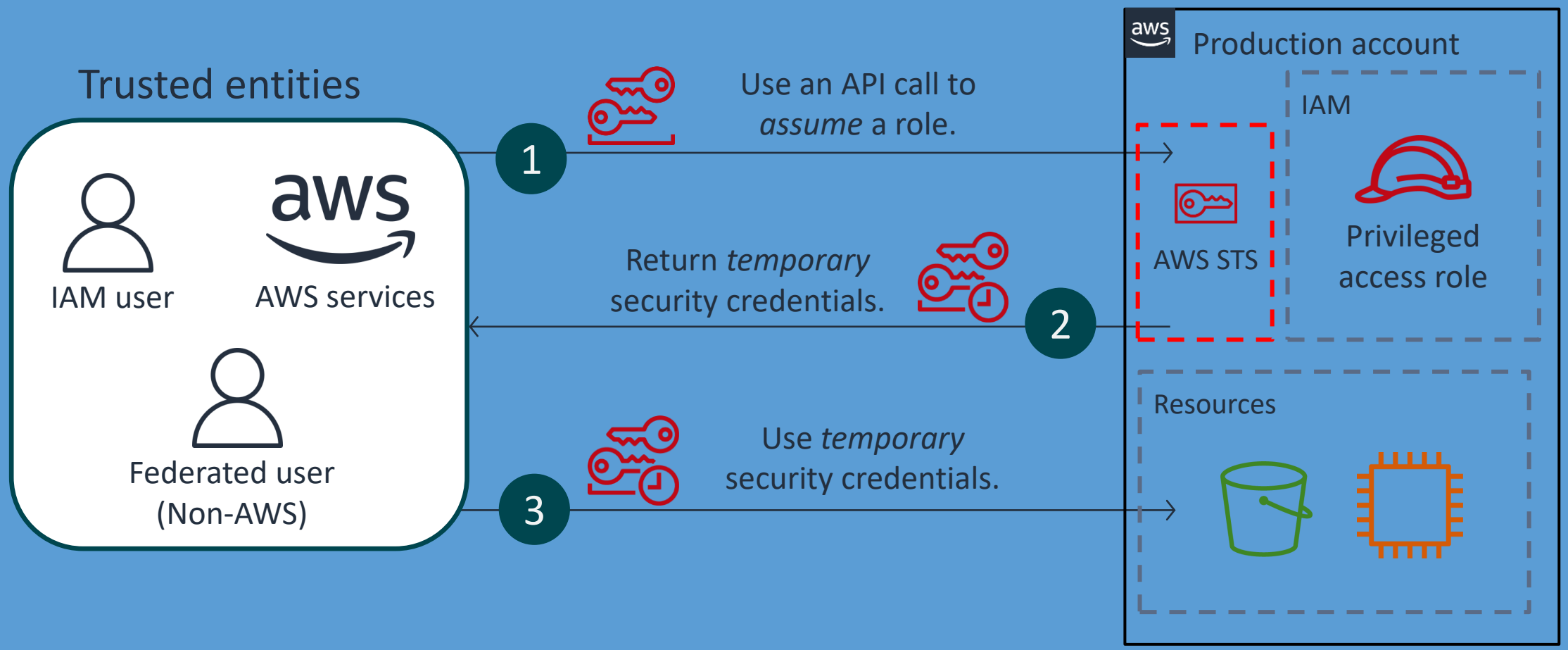


Demonstration:

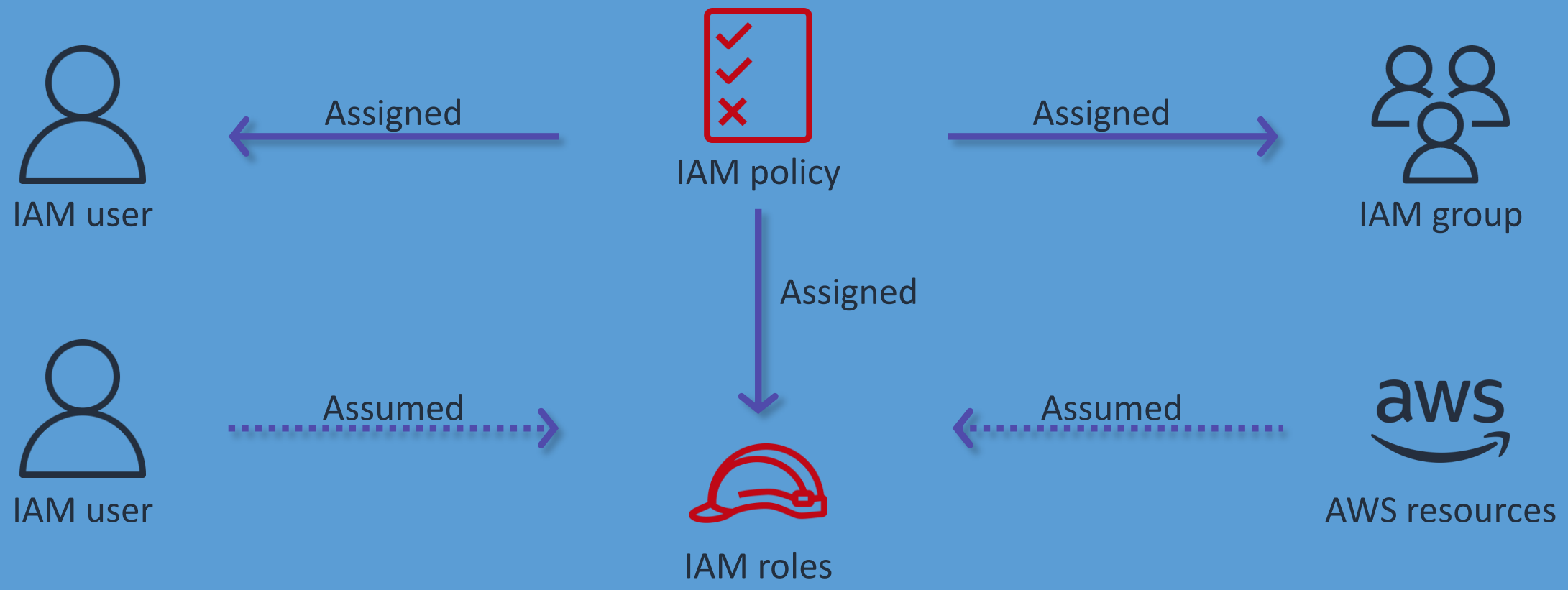
Create an IAM role



Assuming a role



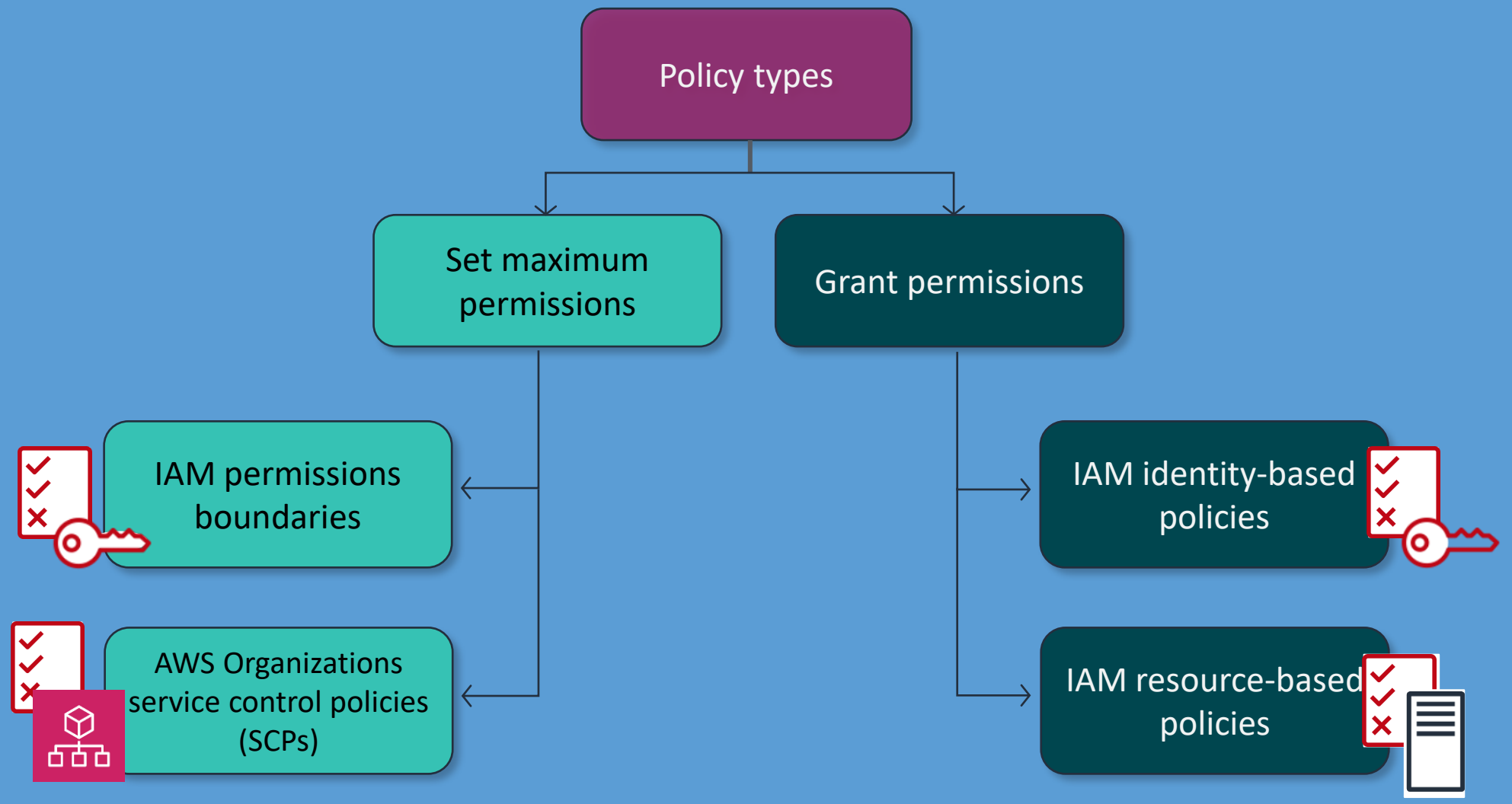
IAM policy assignments



Security policies

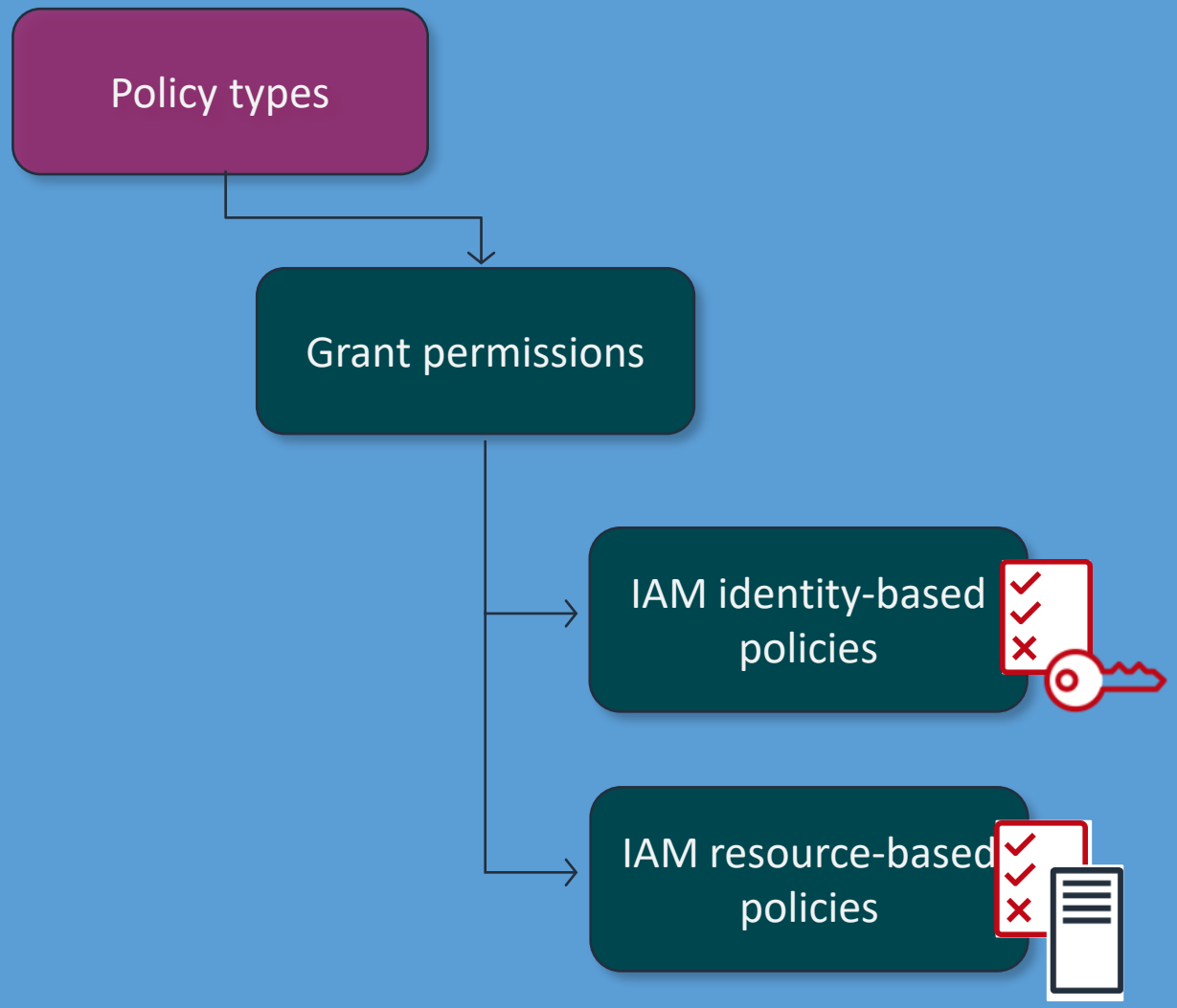
“How can we give users access to only the resources they need?”

Security policy categories



Granting permissions

- Identity-based policies are assigned to users, groups, and roles.
- Resource-based policies are assigned to resources.
- Resource-based policies are checked when someone tries to access the resource.



Types of identity-based policies

Policy elements

	Description	Required
Effect	Use Allow or Deny to indicate whether the policy allows or denies access.	✓
Principal	Indicate the account, user, role, or federated user to which you want to allow or deny access (only on resource policies).	
Action	Include a list of actions that the policy allows or denies.	✓
Resource	Specify a list of resources to which the actions apply.	✓
Condition	Specify the circumstances under which the policy grants permission.	

Identity-based policy example

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ec2:StartInstances",
        "ec2:StopInstances"
      ],
      "Resource": "arn:aws:ec2:*:*:instance/*",
      "Condition": {
        "StringEquals": {
          "ec2:ResourceTag/Owner": "${aws:username}"
        }
      }
    }
  ]
}
```

A — Use this version date to use all of the available policy features.

B — Indicate whether the policy allows or denies an action.

C — Include a list of actions that the policy allows or denies.

D — Choose a list of resources to which the effect applies.

E — Optional: Specify the conditions under which the policy applies.

Explicit allow and explicit deny

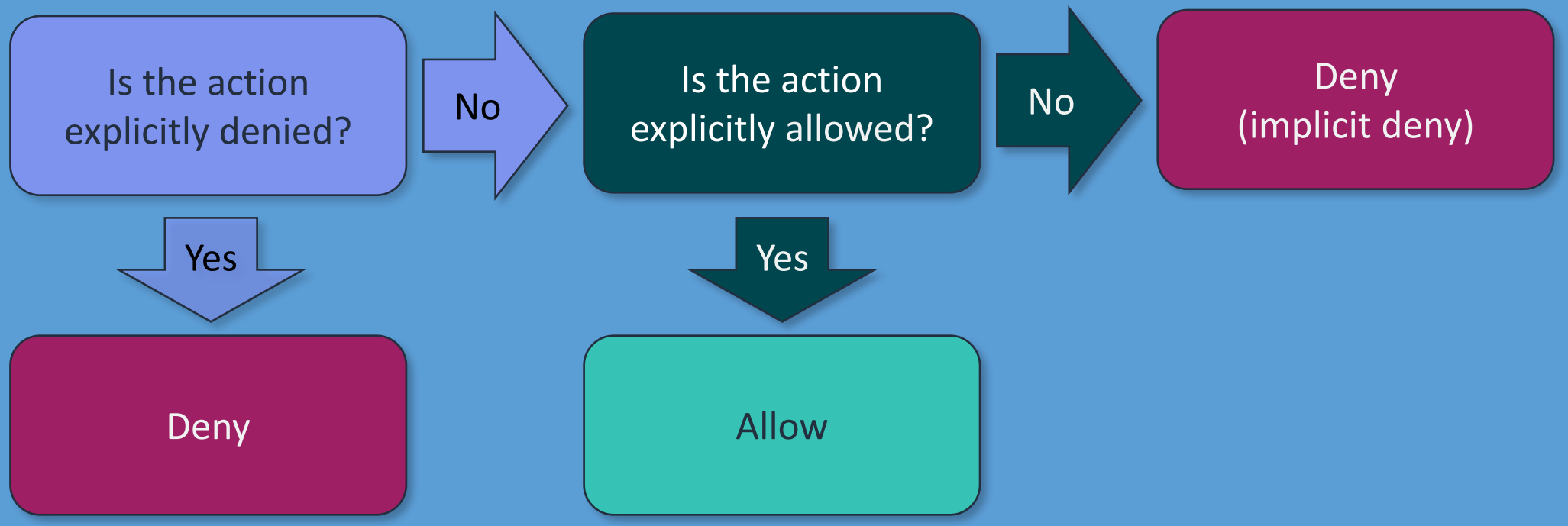
This section from a policy allows access.
This is called an *explicit allow.*

```
{
   "Effect": "Allow",
   "Action": [
       "s3:ListBucket",
       "s3:GetObject"
   ],
   "Resource": [
       "arn:aws:s3:::DOC-EXAMPLE-BUCKET",
      "arn:aws:s3:::DOC-EXAMPLE-BUCKET/*"
   ]
}
```

This section from a policy denies access.
This is called an *explicit deny*.

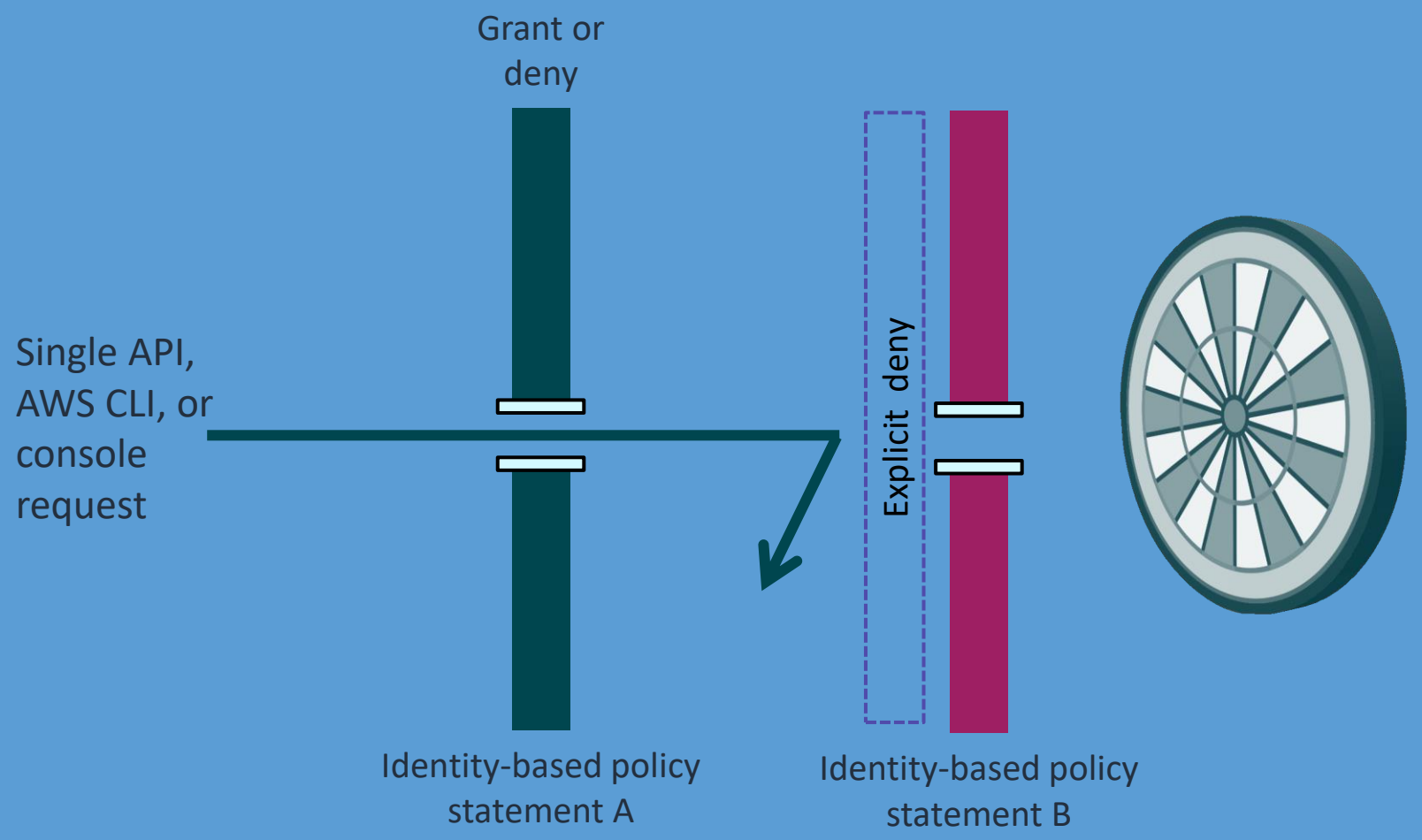
```
{
   "Effect": "Deny",
   "Action": [
     "ec2:*",
     "s3:*"
   ],
   "Resource": "*"
}
```

How IAM policies are evaluated



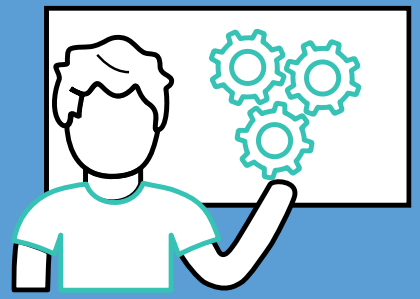
Example of IAM policy explicit deny

- Explicit deny statements override explicit allow.
- If there is no explicit deny, check for an explicit allow.
- If there is no explicit allow, then the request is denied.



Demonstration:

Create an IAM identity-based policy

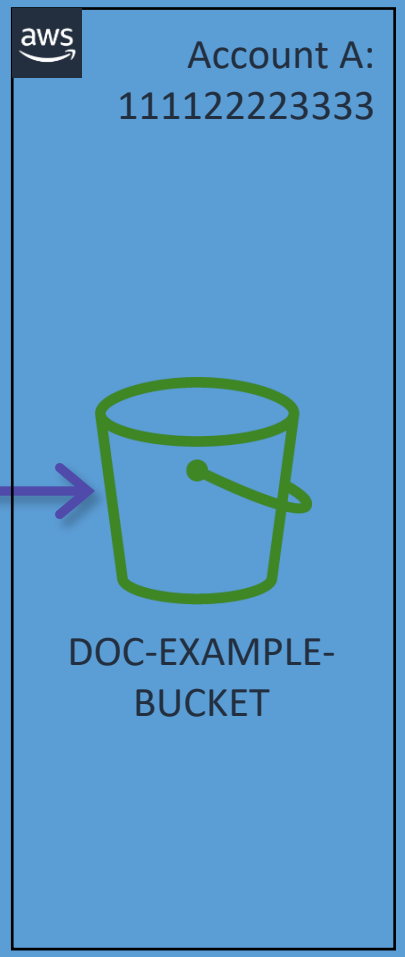


Using a resource-based policy

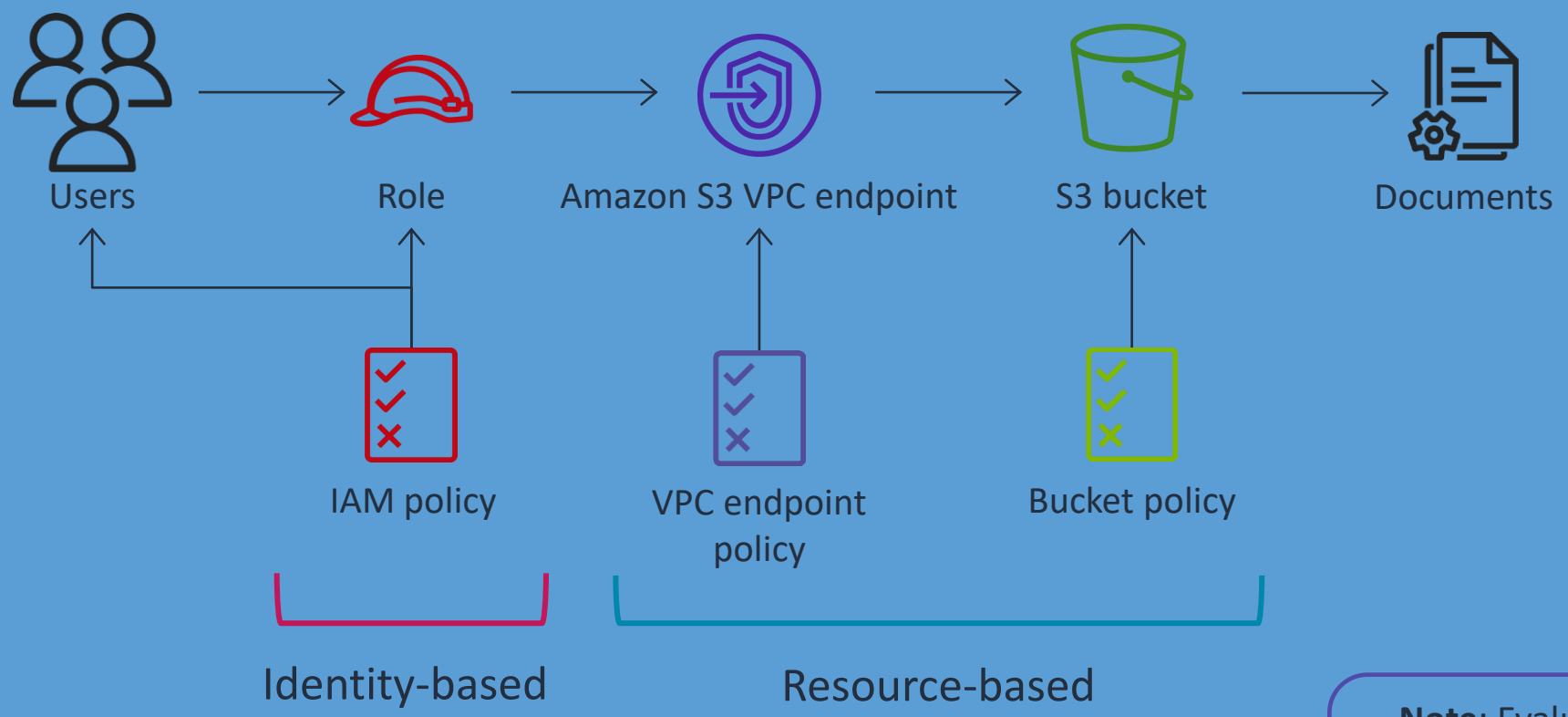
```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AccountBAccess",
      "Effect": "Allow",
      "Principal": {"AWS": "444455556666"},
      "Action": "s3:PutObject",
      "Resource": [
        "arn:aws:s3:::DOC-EXAMPLE-BUCKET/folder123/*"
      ]
    }
  ]
}
```

Required

Optional



Defense in depth

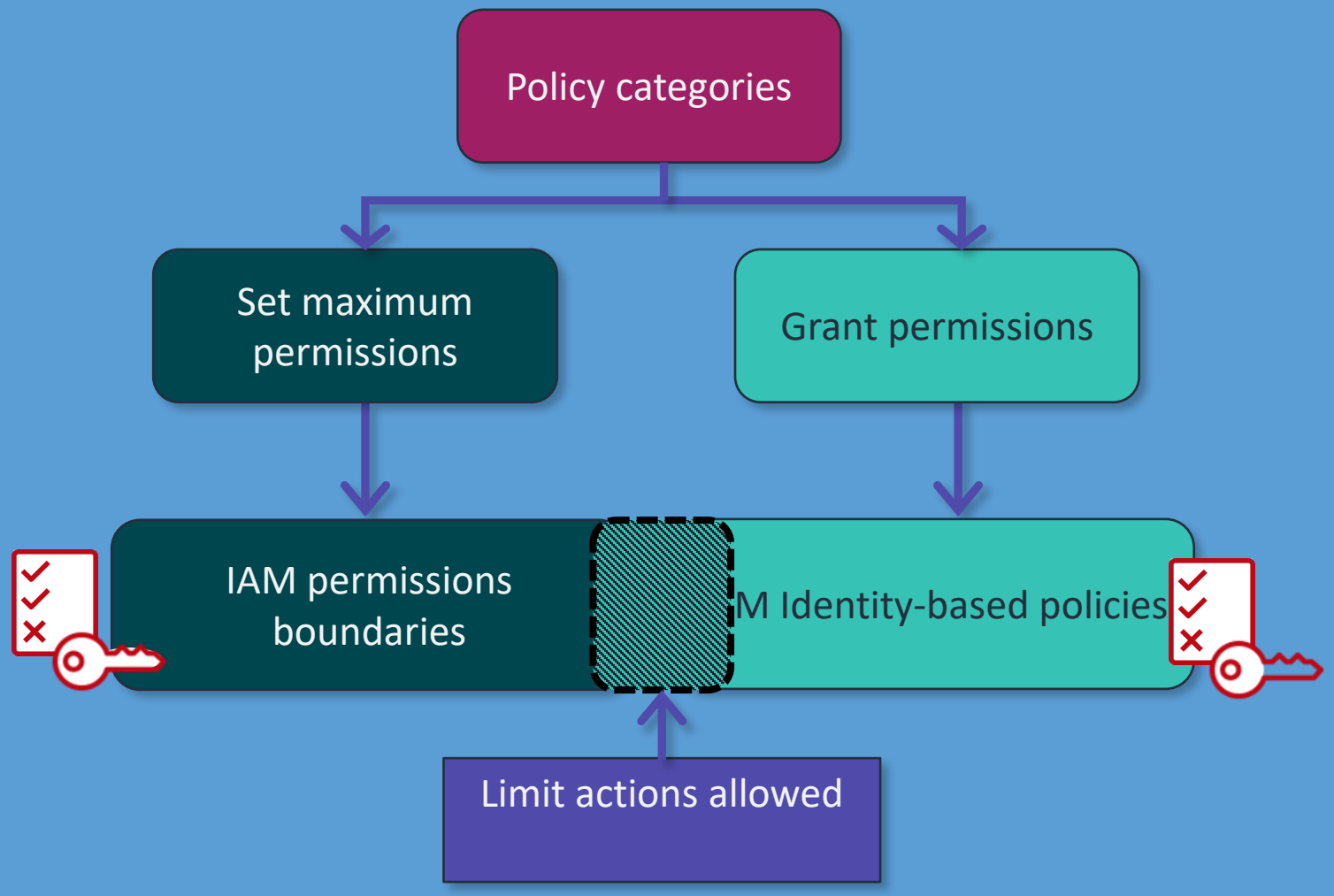


Note: Evaluate identity-based policies and resource-based policies together.

IAM permissions boundaries

IAM permissions boundaries:

- Limit the user's permissions
- Do not provide permissions on their own



Module review

In this module you learned about:

- ✓ Principals and identities
- ✓ Security policies
- ✓ Managing multiple accounts

Next, you review:

Knowledge check

Knowledge check

Knowledge check question 1

Which of the following can be attached to a user, group, or role?

A	Resource-based policies
B	AWS STS
C	Security groups
D	Identity-based policies

Knowledge check question 1 and answer

Which of the following can be attached to a user, group, or role?

A	Resource-based policies
B	AWS STS
C	Security groups
D correct	Identity-based policies

Knowledge check question 2

Which of the following sets permissions on a specific resource and requires a principal to be listed in the policy?

A	Identity-based policies
B	Service control policies (SCPs)
C	Resource-based policies
D	Permissions boundaries

Knowledge check question 2 and answer

Which of the following sets permissions on a specific resource and requires a principal to be listed in the policy?

A	Identity-based policies
B	Service control policies (SCPs)
C correct	Resource-based policies
D	Permissions boundaries

Knowledge check question 3

Which of the following are elements of an IAM user's programmatic access? (Select TWO.)

A	Username
B	Access Key ID
C	Password
D	Secret Access Key
E	MFA token

Knowledge check question 3 and answer

Which of the following are elements of an IAM user's programmatic access? (Select TWO.)

A	Username
B correct	Access key ID
C	Password
D correct	Secret access key
E	MFA token

Knowledge check question 4

True or False: The root user should be used for daily administration of your AWS account.

A	True
B	False

Knowledge check question 4 and answer

The root user should be used for daily administration of your AWS account.

A	True
B correct	False

Knowledge check question 5

Which of the following can only be managed with AWS Organizations?

A	Service control policies (SCPs)
B	Resource-based policies
C	Permissions boundaries
D	Identity-based policies

Amazon Web Services

Networking 1

Question

Which network components are you familiar with? Choose all that apply:

A. IP addressing and subnetting

B. Switching and routing

C. Network security

D. None of the above

Overview

- Business requests
- IP addressing
- Virtual Private Cloud (VPC) fundamentals
- VPC traffic security
- Present solutions
- Capstone check-in
- Knowledge check

Business Requirements

Network Engineer

The network engineer needs to know:

- How can we make sure that our network has enough IP addresses to support our workloads?
- How do we build a dynamic and secure network infrastructure in our AWS account?
- How can we filter inbound and outbound traffic to protect resources on our network?

IP addressing

“How can we make sure that our network has enough IP addresses to support our workloads?”

IPv4 32-bit address

IPv4 supports Dynamic Host Configuration Protocol (DHCP) or manual configuration.

4.3 billion addresses

Addresses must be reused.

Addresses are written in numeric dot-decimal notation.

172.31.0.0/16

Recommended: RFC1918 range

Recommended: /16 (65,536 addresses)

IPv6 128-bit address

- IPv6 has been developed to replace IPv4.
- IPv6 supports automatic configuration.

340 trillion trillion trillion addresses

Every device can have a unique address.

Addresses are written in alphanumeric hexadecimal notation.

2001:db8:1234:1a00::/56

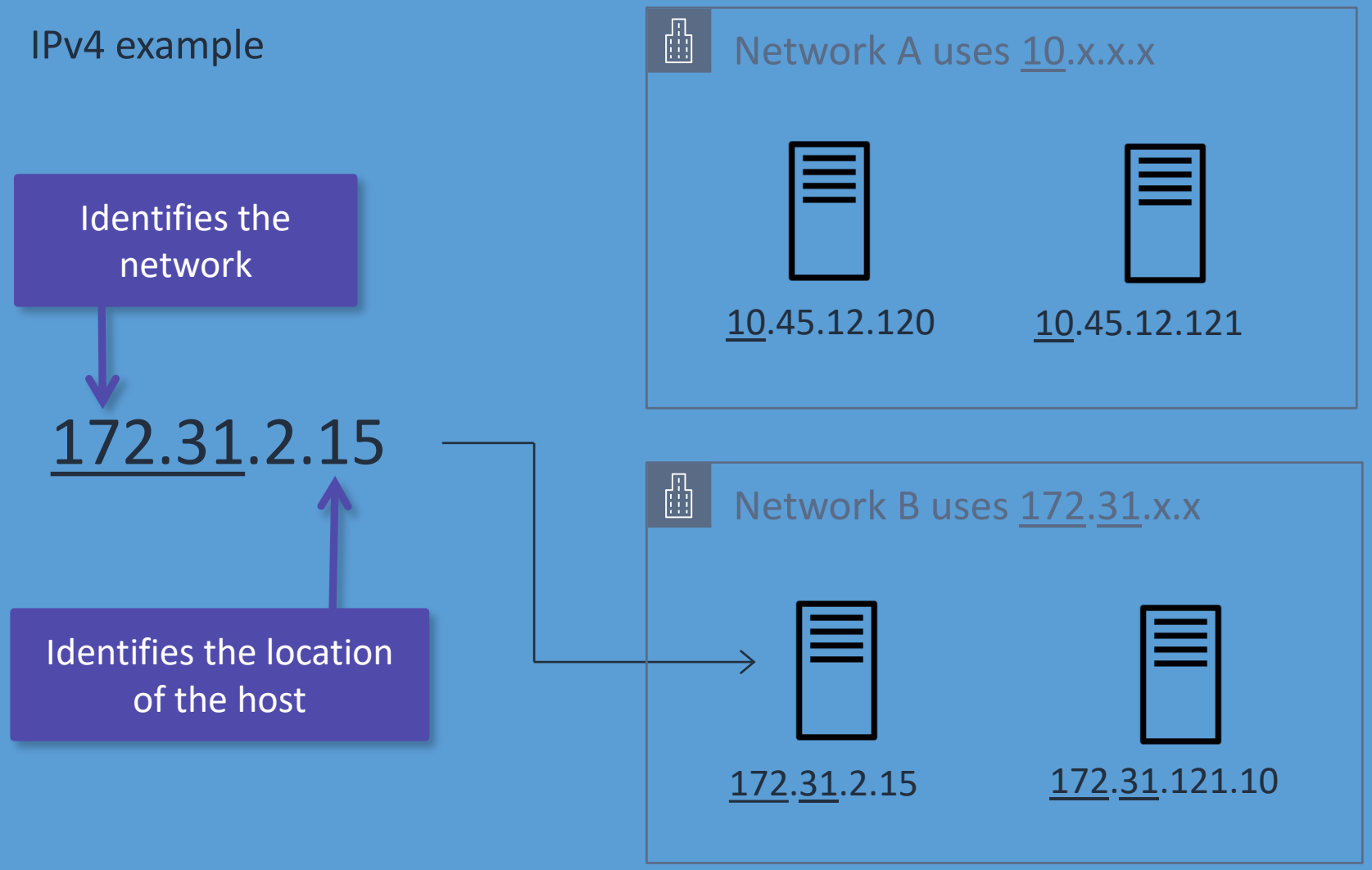
Amazon Global Unicast Addresses (GUA) – internet-routable

Associate a /56 IPv6 CIDR (automatically allocated)

IP addresses

- An IP address identifies a location within a network.
- It identifies the network and the host.
- There are two types of IP addresses:
  - IPv4
  - IPv6

IPv4 example



Classless Inter-Domain Routing (CIDR)

CIDR notation is a way of representing an IP address and its network mask.

10.22.0.0/16

Dot notation to specify the network or subnet.

Slash notation for the subnet mask

An IPv4 address is four groups of 8 bits.

10.22.0.0/16

00001010.00010110.00000000.00000000

/16 reserves the first 16 bits for network identification

The remaining bits are the host range: 10.22.0.0 to 10.22.255.255

AWS supported ranges

CIDR	Total IPs
/28	16
…	…
/20	4,096
/19	8,192
/18	16,384
/17	32,768
/16	65,536

Subnet mask

	Network		Host	
IP Address	172	16	0	0

	Network		Host	
Default Subnet Mask	255 11111111	255 11111111	0 00000000	0 00000000

Also written as “/16” where 16 represents the number of 1s in the mask.

	Network		Subnet	Host
8-bit Subnet Mask	255	255	255	0

Also written as “/24” where 24 represents the number of 1s in the mask.

VPC fundamentals

“How do we build a dynamic and secure network infrastructure in our AWS account?”

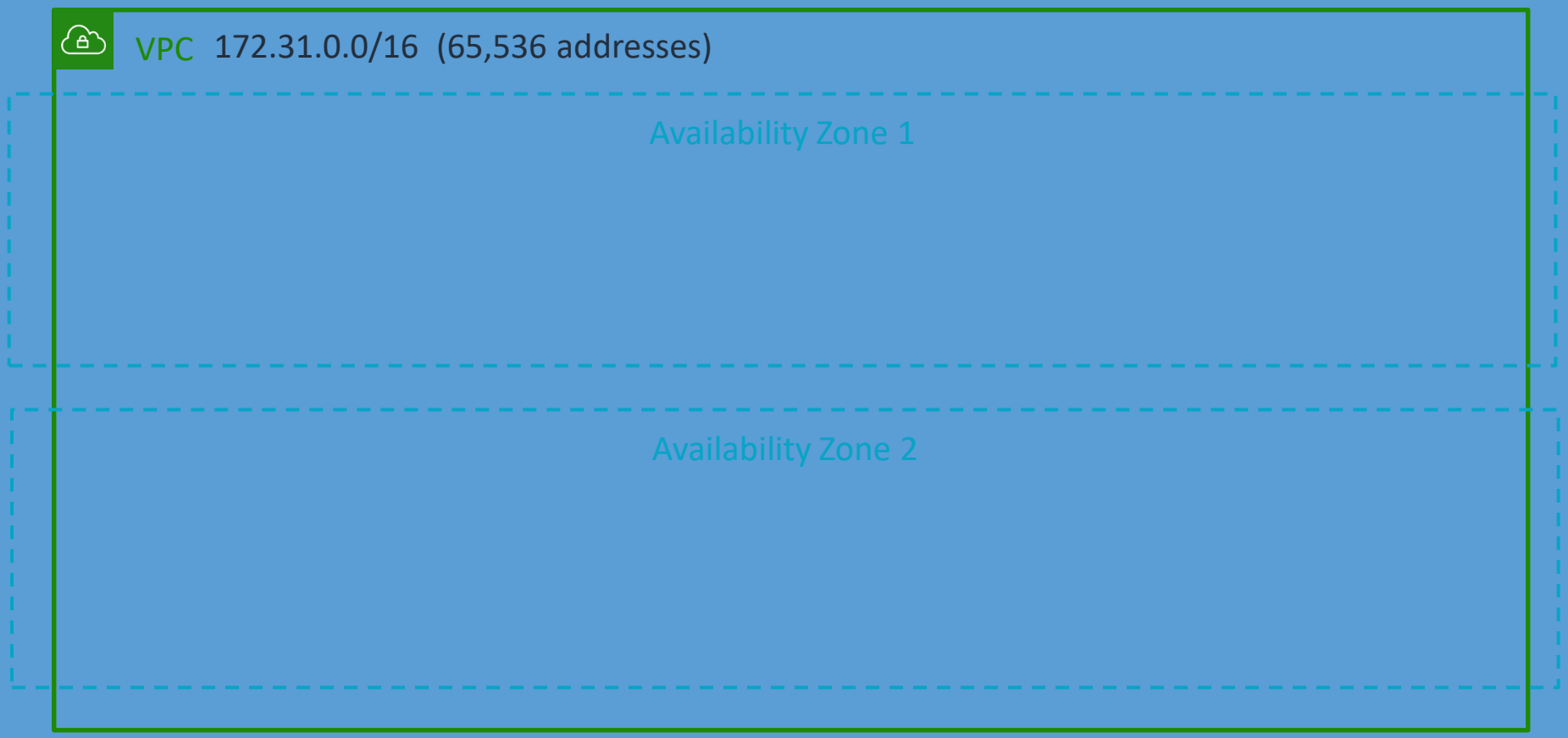
VPC fundamentals topics

VPC fundamentals

- Amazon VPC
- Subnets
- Internet gateway
- Route table
- Elastic IP address
- Elastic network interface
- NAT gateway

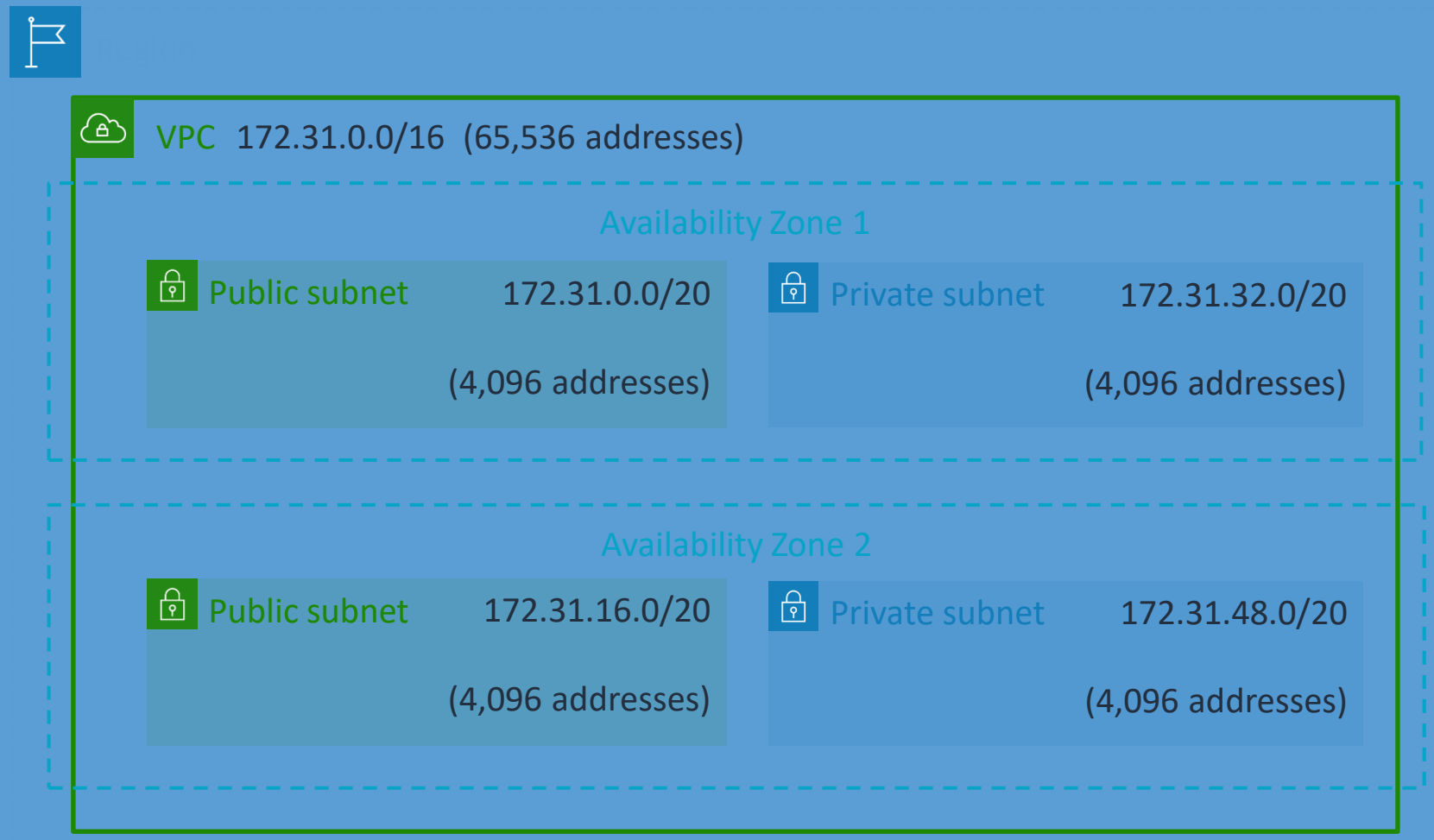
Amazon VPC

- Provides logical isolation for your workloads
- Permits custom access controls and security settings for your resources
- Is bound to a single AWS Region



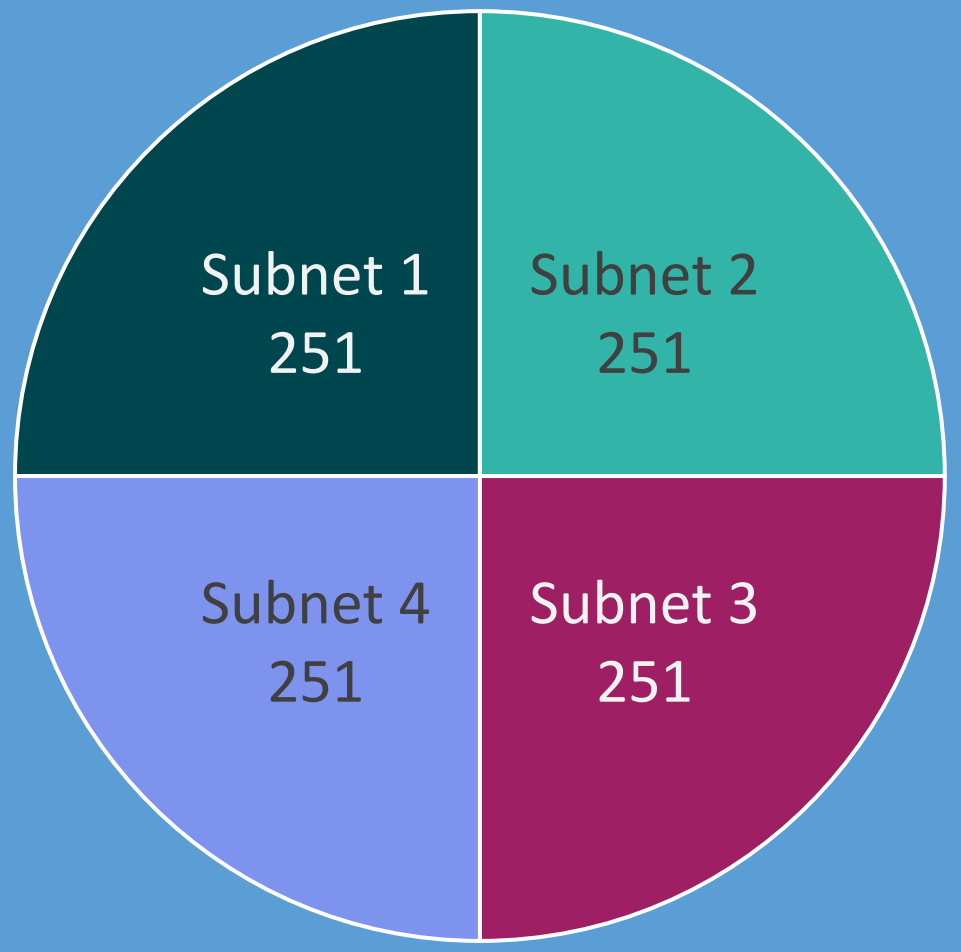
Subnets

- Subnets are a subset of the VPC CIDR block.
- Subnet CIDR blocks cannot overlap.
- Each subnet resides within one Availability Zone.
- An Availability Zone can contain multiple subnets.
- Five addresses are reserved.



Using subnets to divide your VPC

- Using subnets isolates resources for routing and security.
- AWS will reserve five IP addresses from each subnet.



A VPC with CIDR “/22” includes 1,024 total IP addresses.

Public subnets

A public subnet holds resources that work with inbound and outbound internet traffic. It requires the following:

Route table

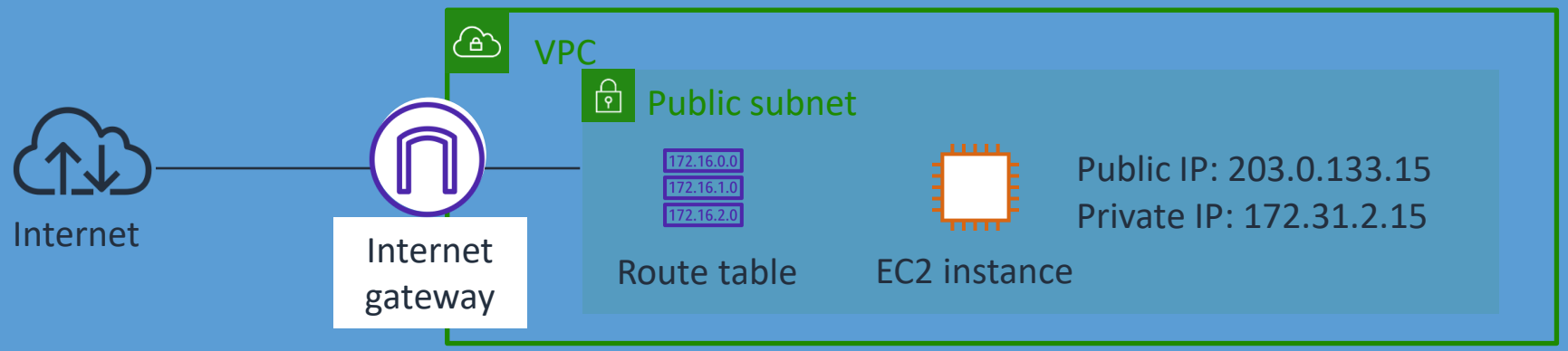
- A set of rules that the VPC uses to route network traffic
- Requires a route to the internet

Internet gateway

Allows communication between resources in your VPC and the internet

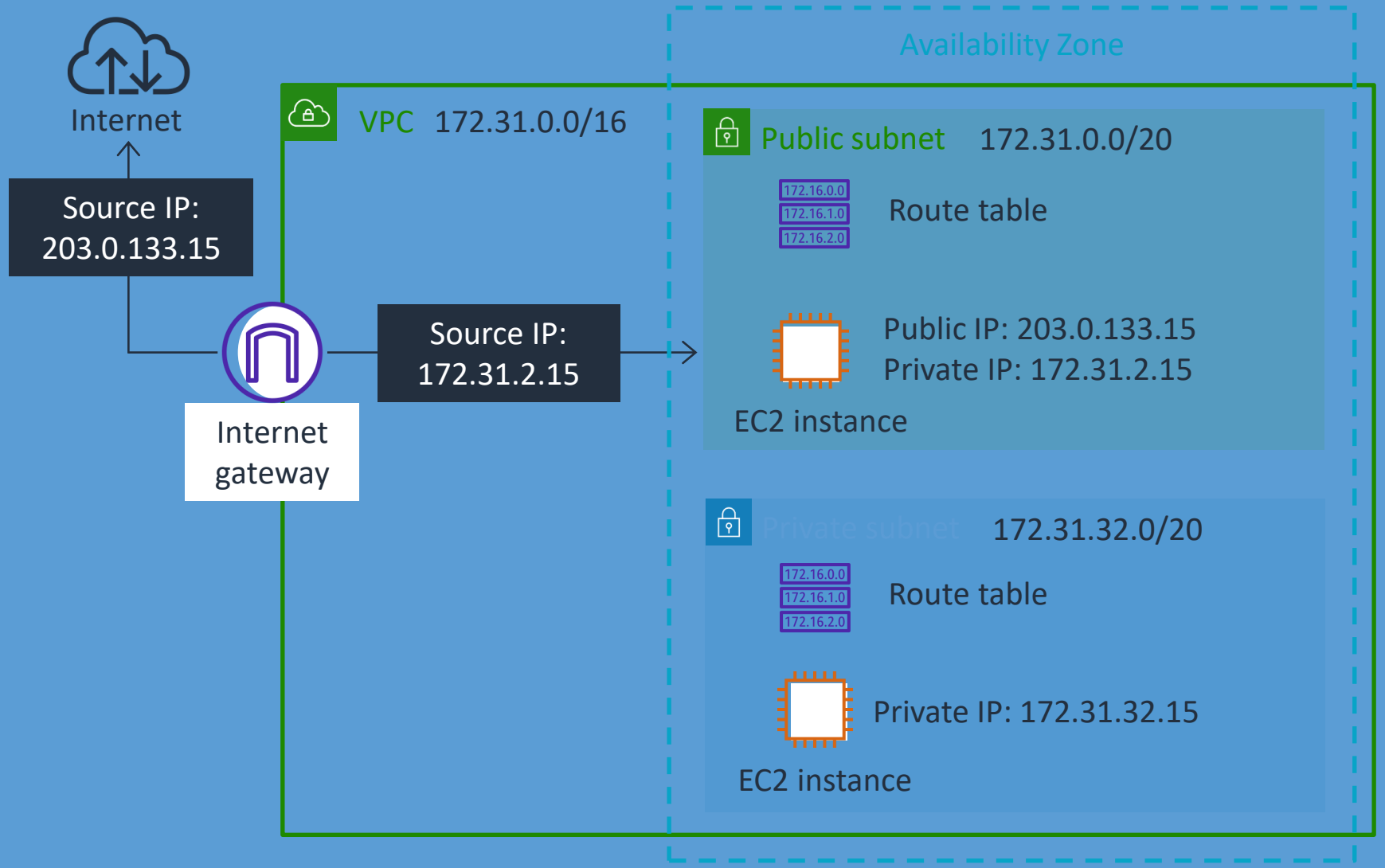
Public IP addresses

- IP addresses that can be reached from the internet
- Protects the private IP addresses only reachable on the network



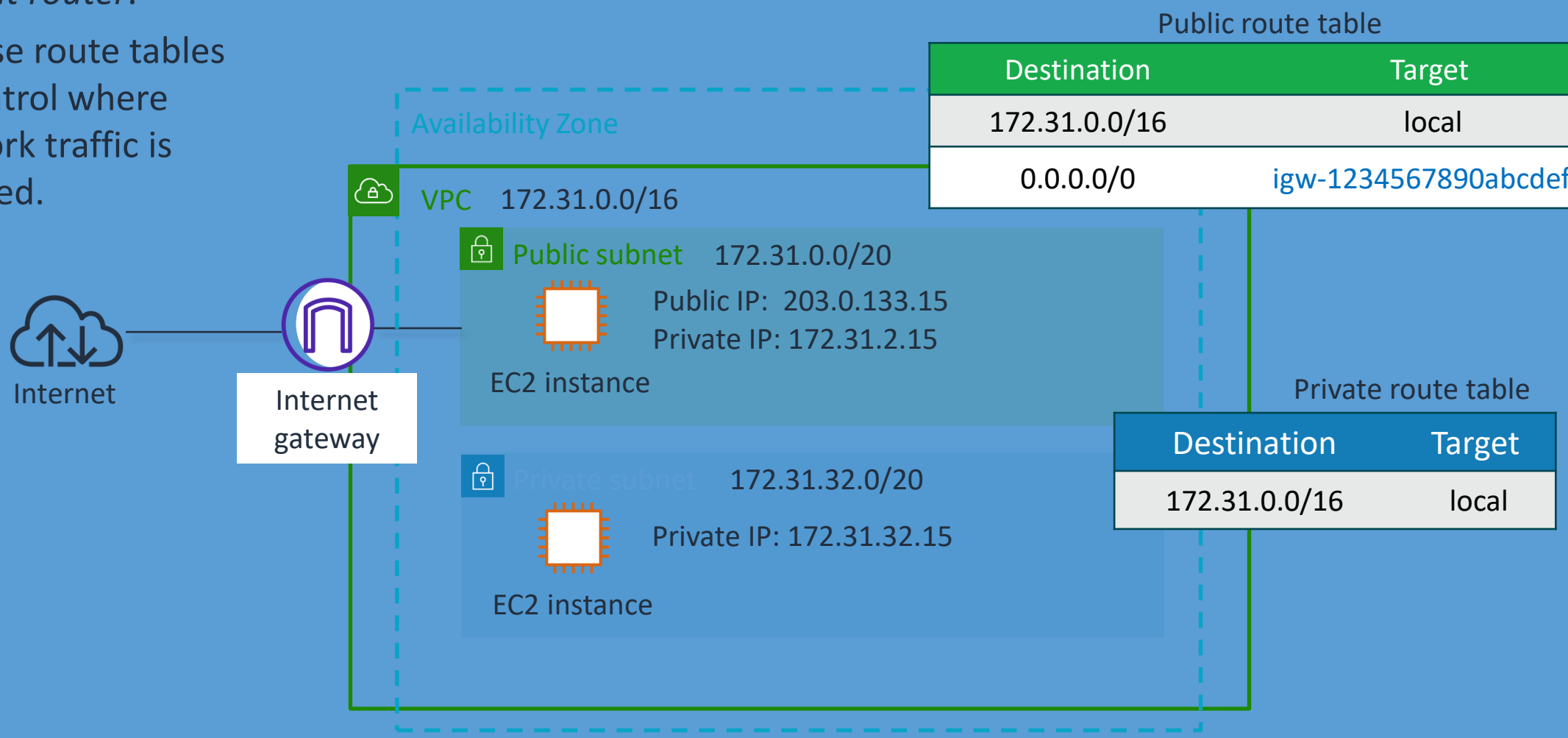
Internet gateways

- Internet gateways permit communication between instances in your VPC and the internet.
- They provide a target in your subnet route tables for internet-routable traffic.
- It protects IP addresses on your network by performing network address translation (NAT).



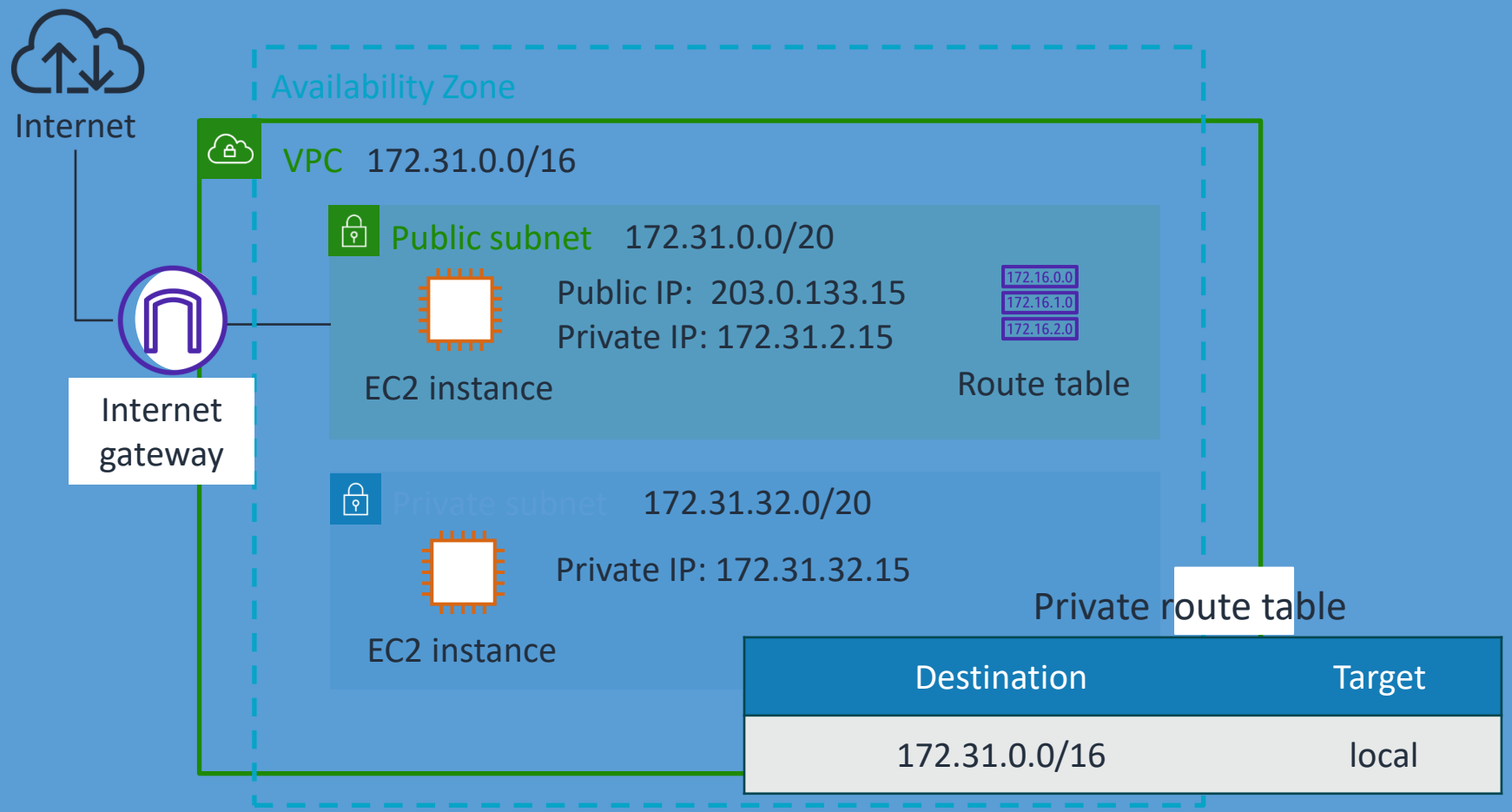
Route tables

- Your VPC has an *implicit router*.
- You use route tables to control where network traffic is directed.



Private subnets

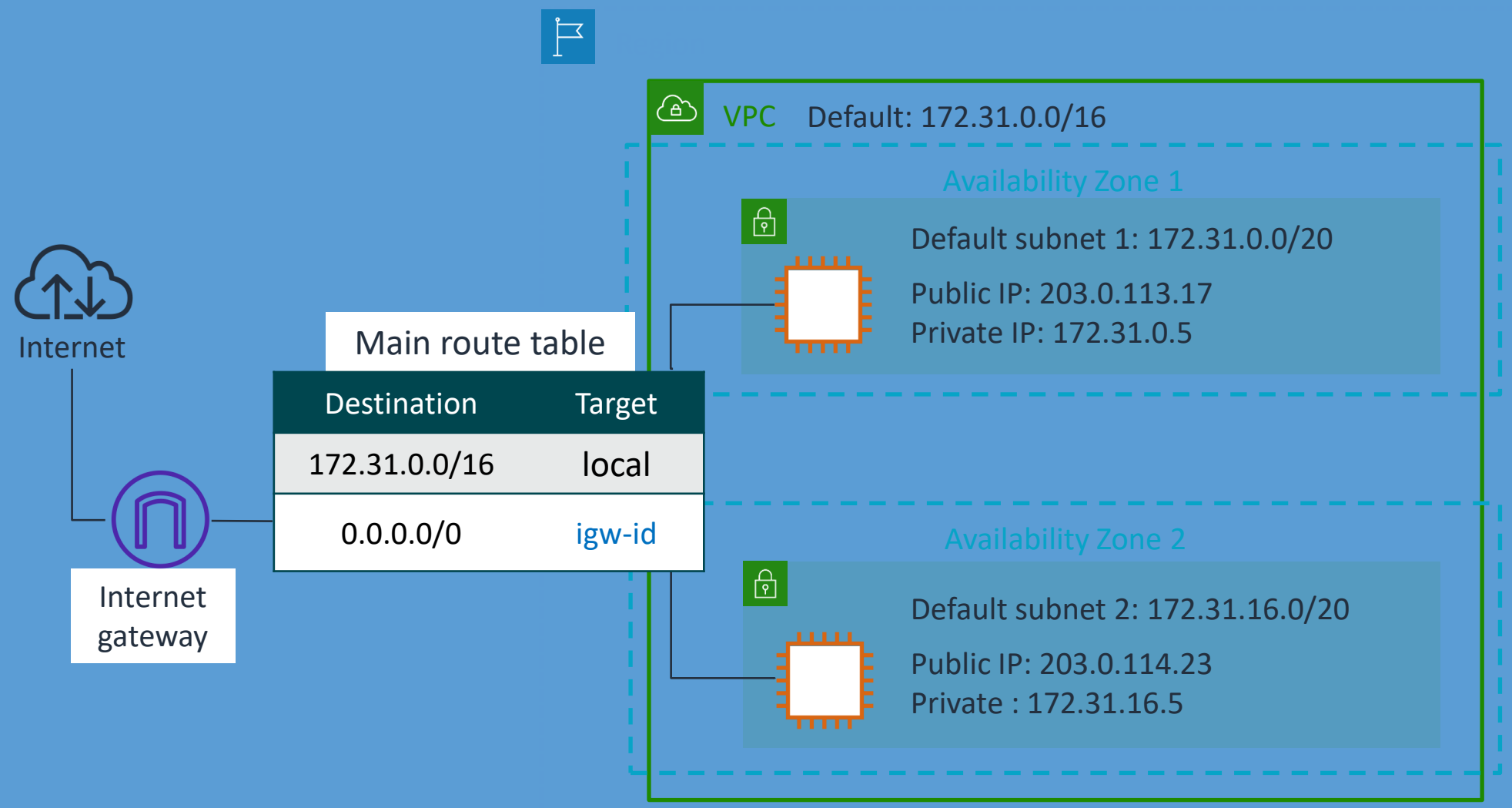
- Private subnets allow indirect access to the internet.
- The private IP address never changes.
- Traffic in the VPC stays local.



Destination	Target
172.31.0.0/16	local

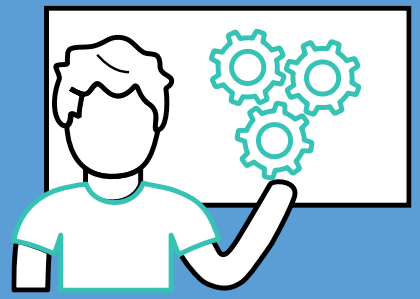
Default Amazon VPCs

- Provisioned at account creation
- Preconfigured for immediate use
- Span all Availability Zones within the Region
- Owned and controlled by the customer



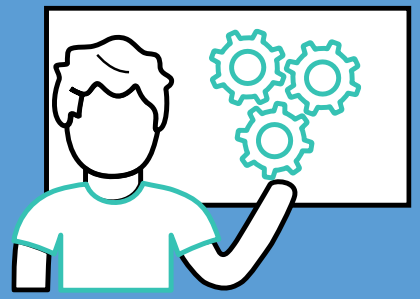
Demonstration:

How to deploy a VPC



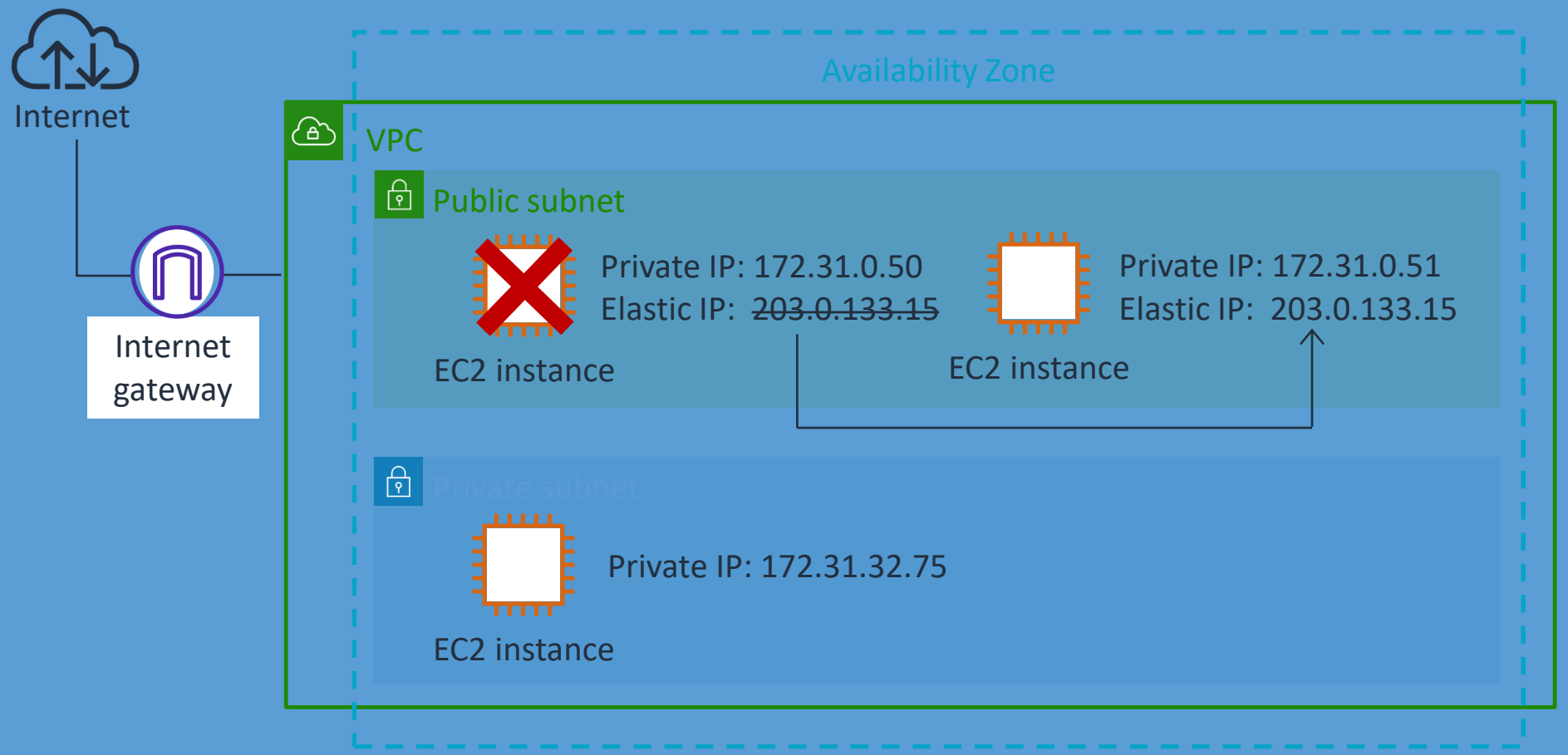
Demonstration:

Configure routing for a public subnet



Elastic IP addresses

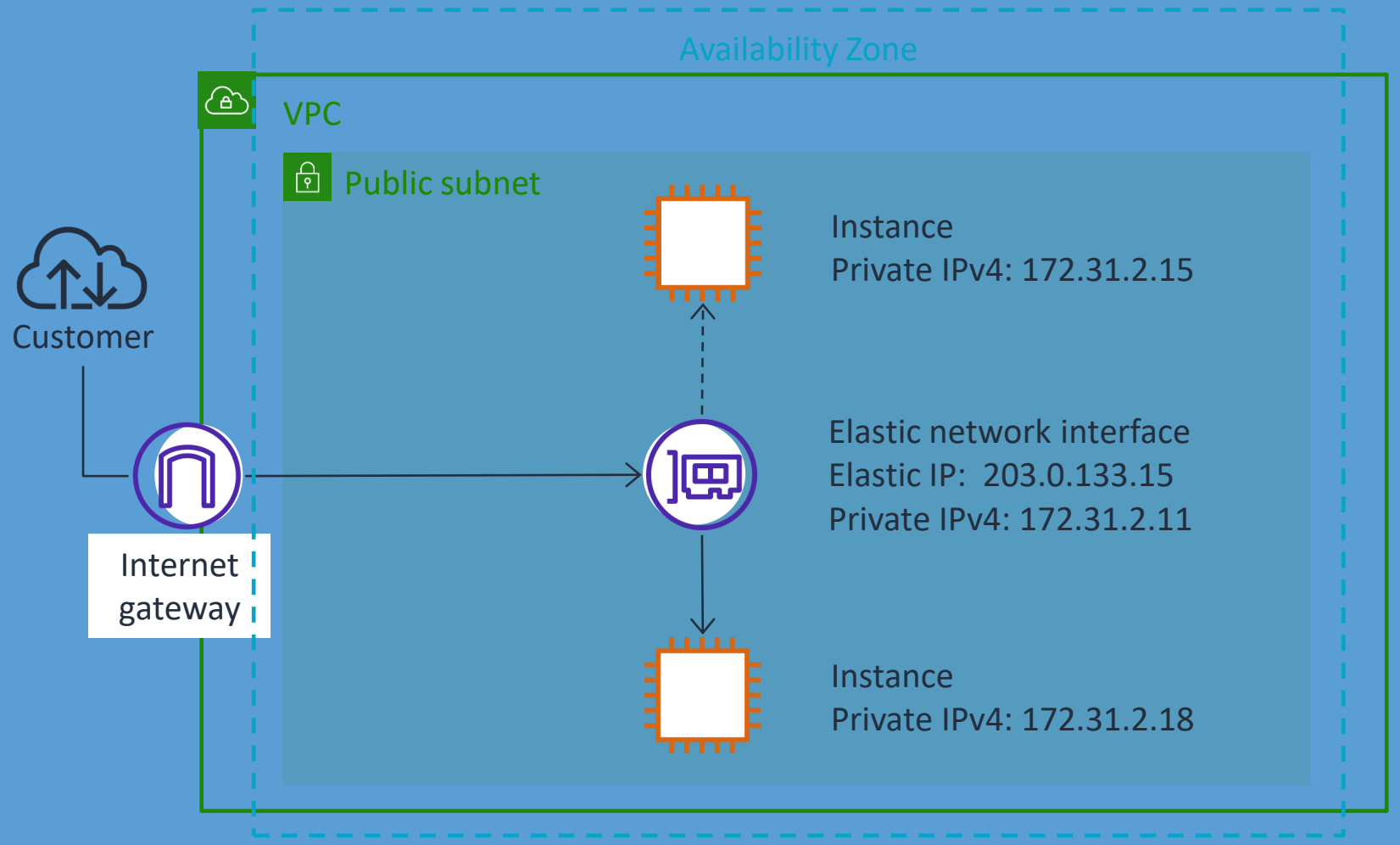
- Permit association with an instance or a network interface
- Can be reassociated and direct new traffic immediately
- Default restriction of five per Region, per account
- Support Bring Your Own IP (BYOIP)



Elastic network interface

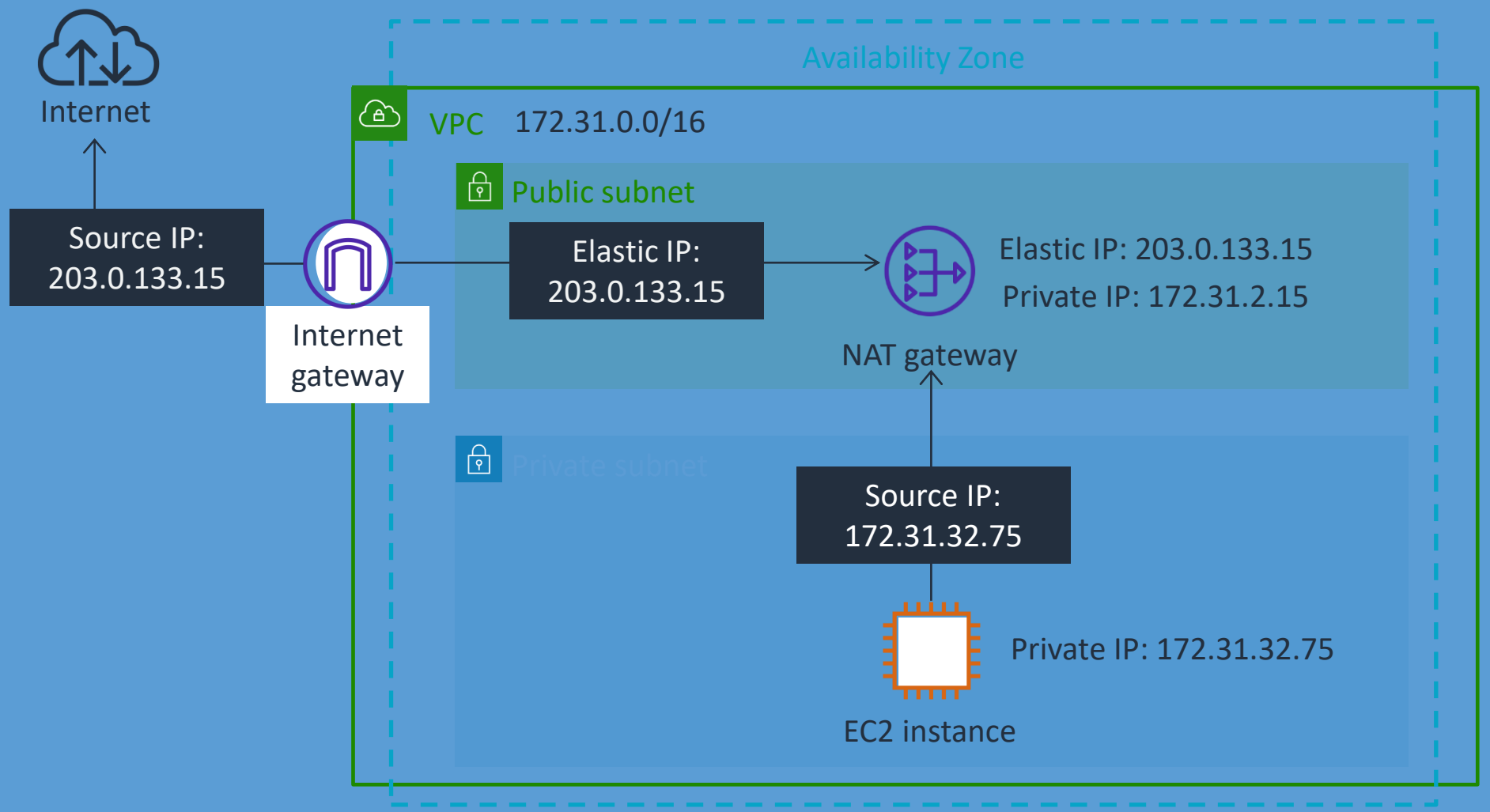
An *elastic network interface* is a logical networking component in a VPC that:

- Can be moved across resources in the same Availability Zone
- Maintains its private IP address, Elastic IP address, and MAC address



Network address translation with NAT gateways

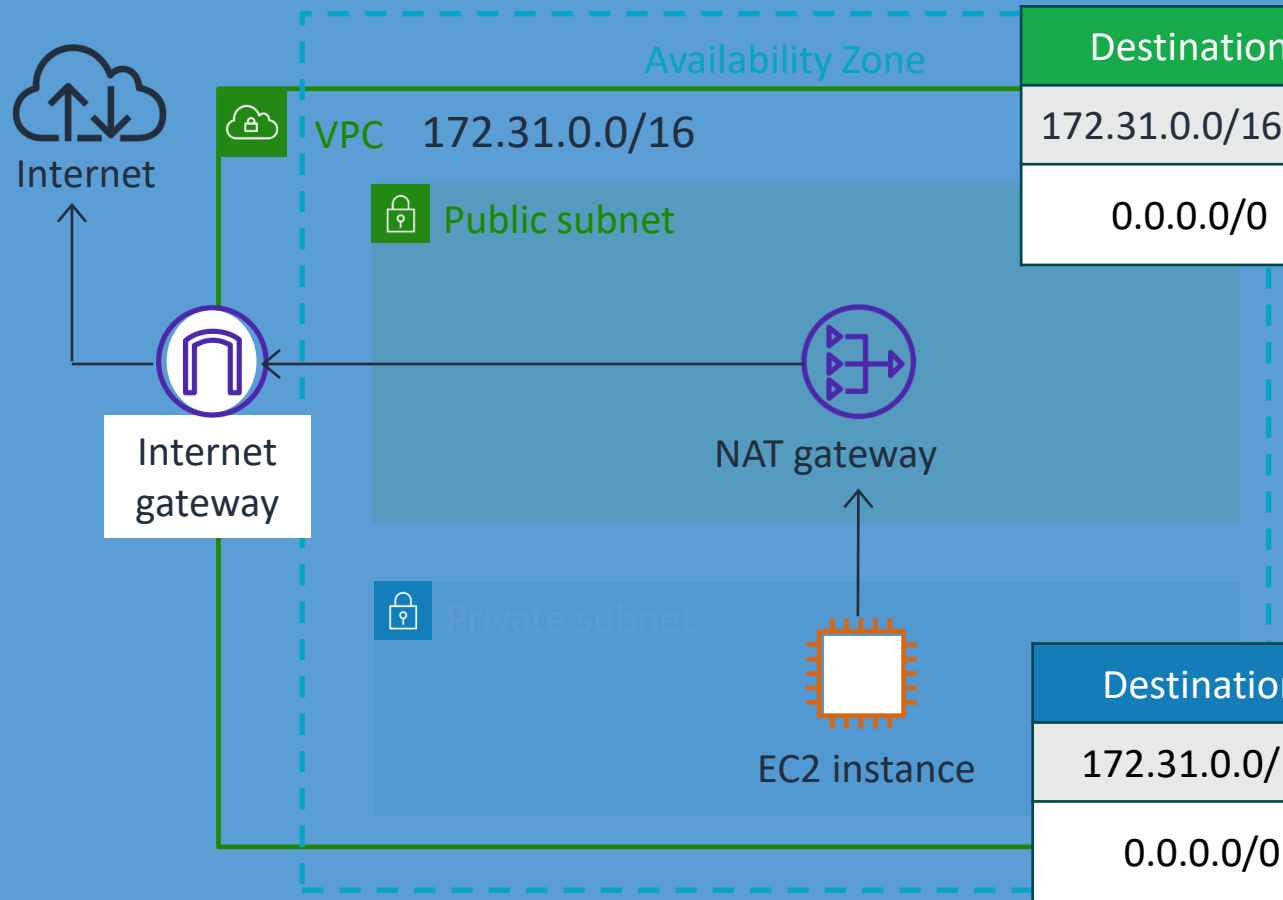
- You use NAT to protect your private IP addresses.
- A NAT gateway uses an Elastic IP address as the source IP address for traffic from the private subnet.



Connecting private subnets to the internet

NAT gateway use case: Connecting resources in a private subnet to the internet

- The route table for the private subnet sends all IPv4 internet traffic to the NAT gateway.
- The route table for the public subnet sends all internet traffic to the internet gateway.



Public route table

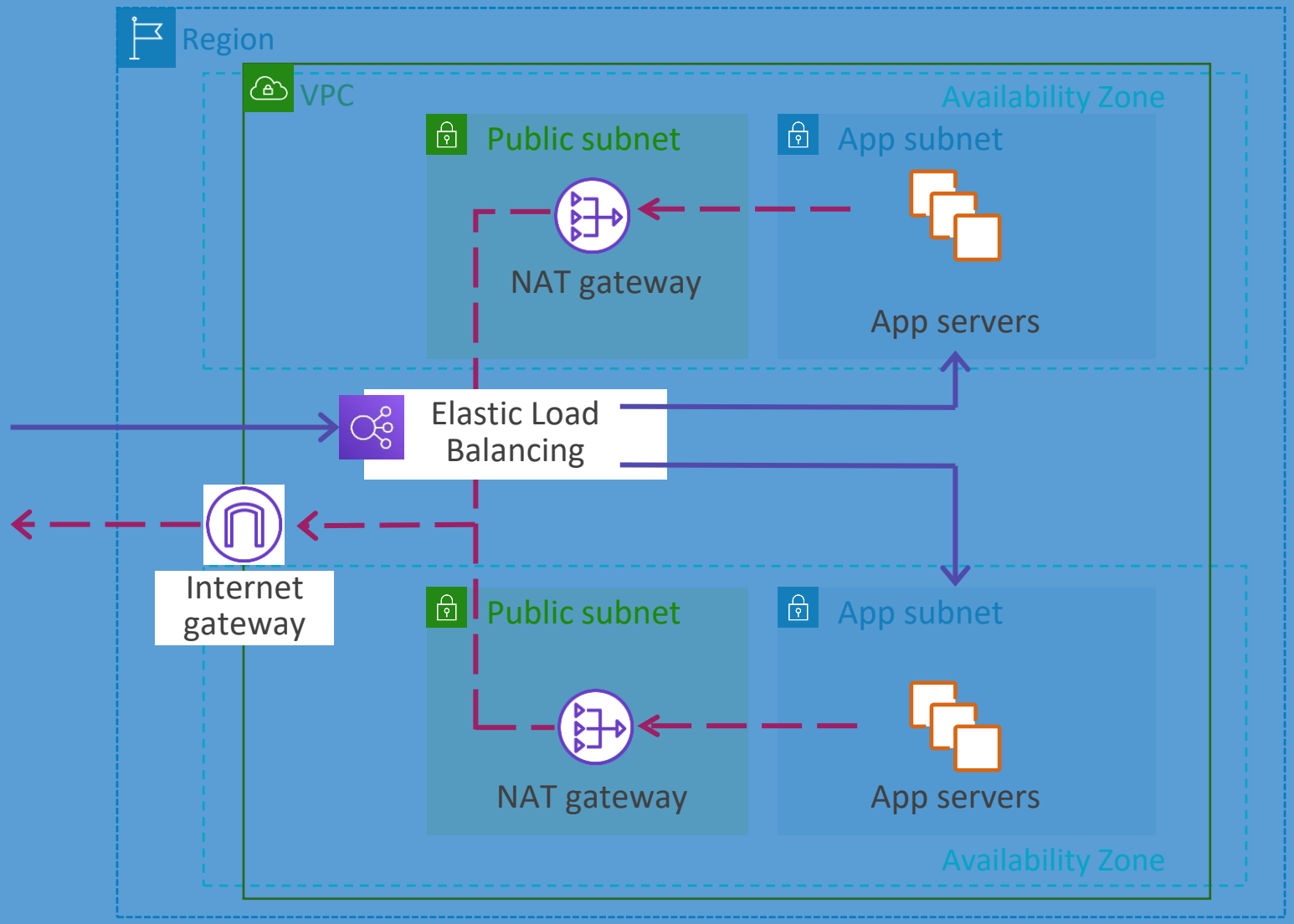
Destination	Target
172.31.0.0/16	local
0.0.0.0/0	igw-1234567890abcdef0

Private route table

Destination	Target
172.31.0.0/16	local
0.0.0.0/0	nat-021345abcdef6789

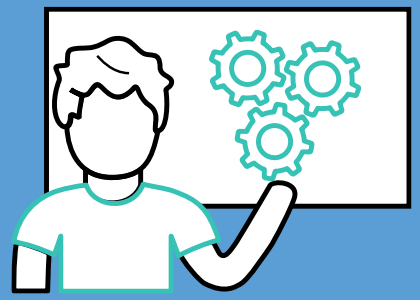
Deploy a VPC across multiple Availability Zones

- Deploy your VPCs across multiple Availability Zones to achieve high availability.
- Create subnets in each Availability Zone.
- Deploy resources in each Availability Zone.
- Distribute traffic between the Availability Zones using load balancers.



Demonstration:

Configure routing for a private subnet using a NAT gateway

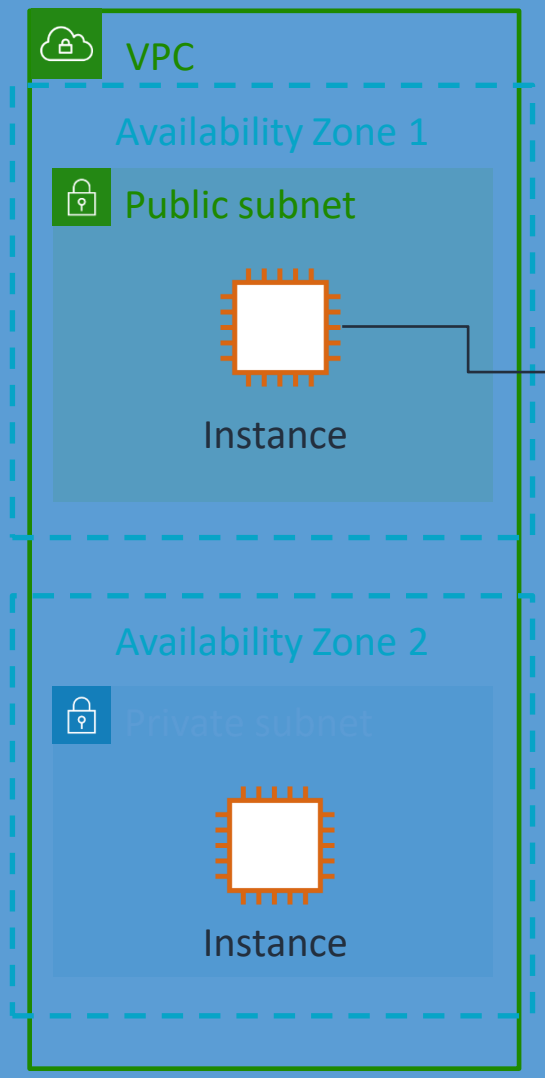


VPC traffic security

“How can we filter inbound and outbound traffic to protect resources on our network?”

Network access control lists (ACLs)

- A network ACL acts as a firewall at the subnet boundary.
- By default, it allows all inbound and outbound traffic.
- It is stateless, requiring explicit rules for all traffic.
- It evaluates rules starting with the lowest numbered rule.



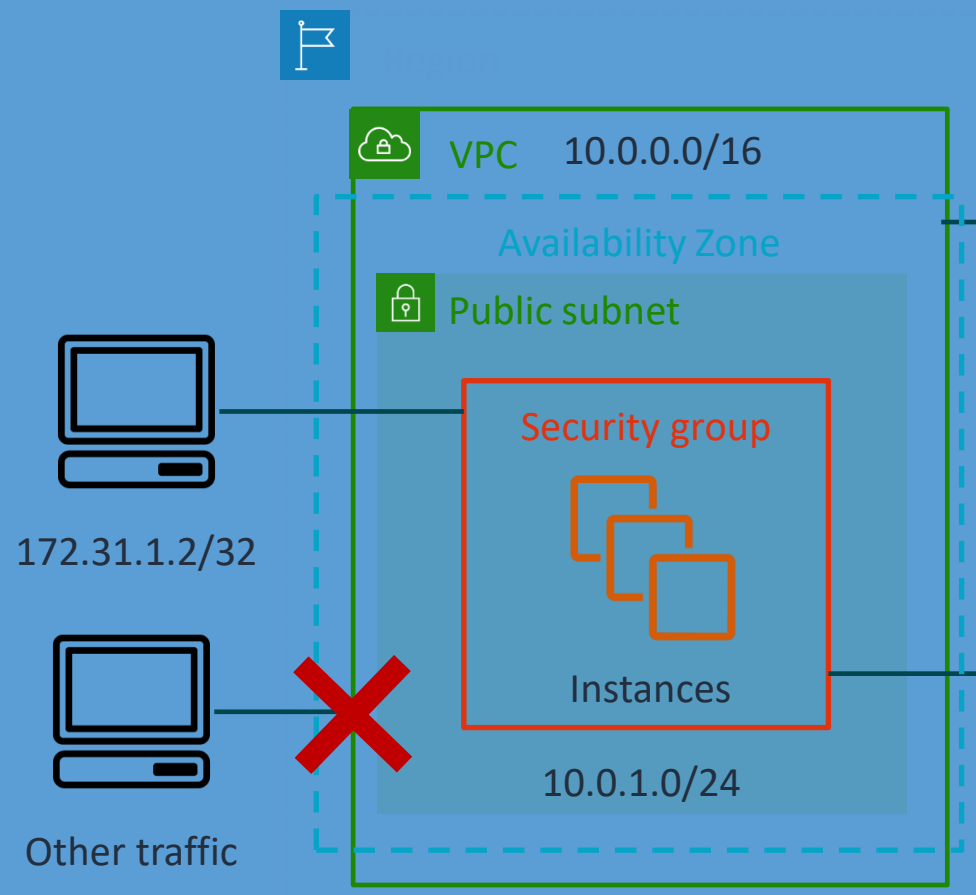
nacl-MyNACL1

Inbound					
Rule #	Type	Protocol	Port Range	Source	Allow or Deny
100	HTTP	TCP	80	0.0.0.0/0	*Allow*
101	HTTPS	TCP	443	0.0.0.0/0	*Allow*
*	ALL Traffic	ALL	ALL	0.0.0.0/0	Deny

Outbound					
Rule #	Type	Protocol	Port Range	Destination	Allow or Deny
100	Custom TCP Rule	TCP	1024-65535	0.0.0.0/0	*Allow*
*	ALL Traffic	ALL	ALL	0.0.0.0/0	Deny

Network ACL use cases

- The network ACL controls access to instances in a subnet.
- The network ACL is a backup layer of defense.
- The network ACL rules apply to all instances in the subnet.



Network acl-11223344

Inbound

Rule # 100: SSH 172.31.1.2/32 **ALLOW**
Rule # *: All traffic 0.0.0.0/0 **DENY**

Outbound

Rule # 100: Custom TCP 172.31.1.2/32 **ALLOW**
Rule # *: All traffic 0.0.0.0/0 **DENY**

Sg-1a2b3c4d

Inbound

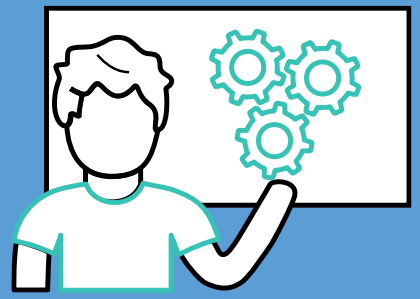
All traffic sg-1a2b3c4d SSH 172.31.1.2/32

Outbound

All traffic sg-1a2b3c4d

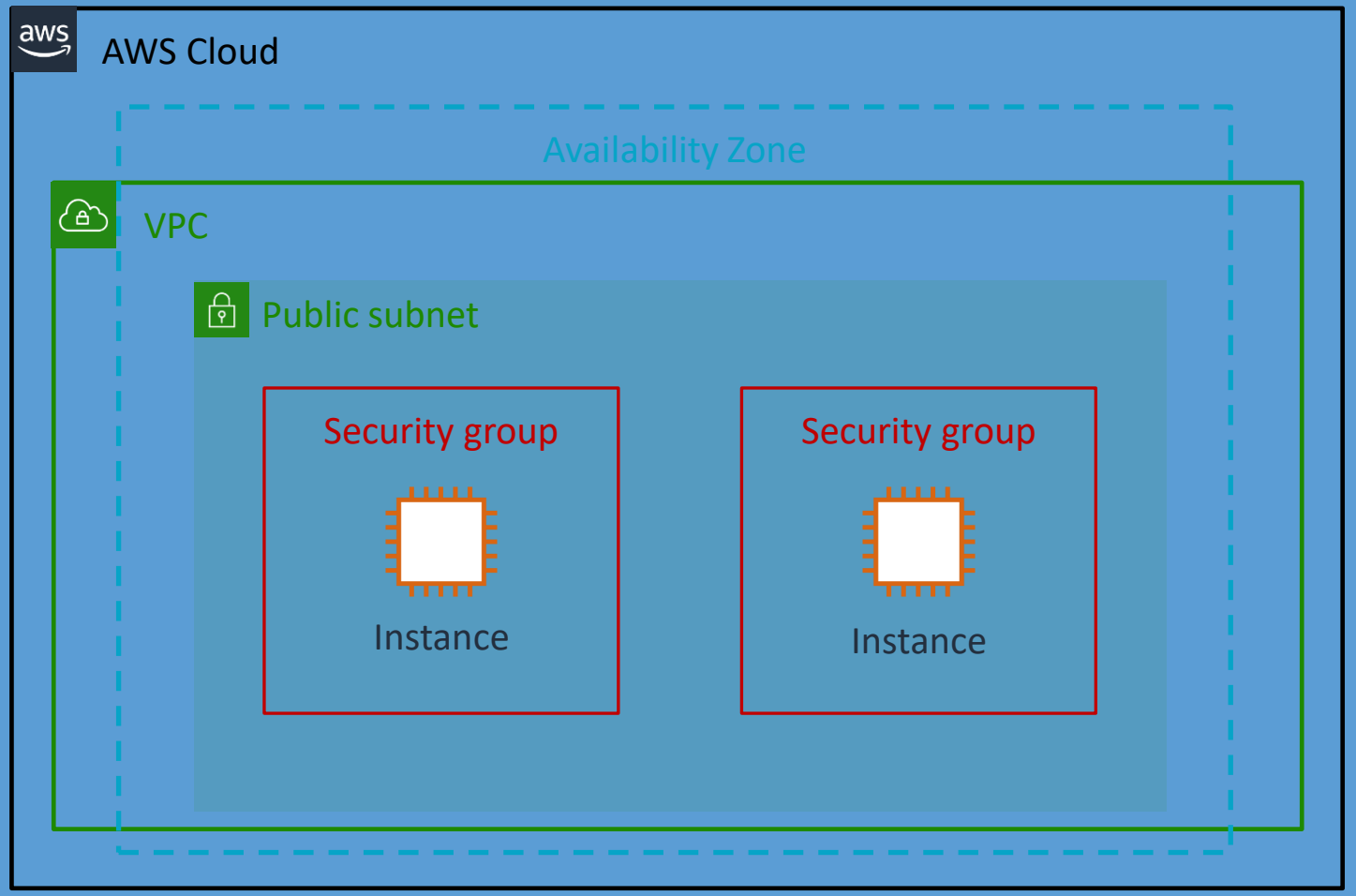
Demonstration:

Creating a network ACL rule



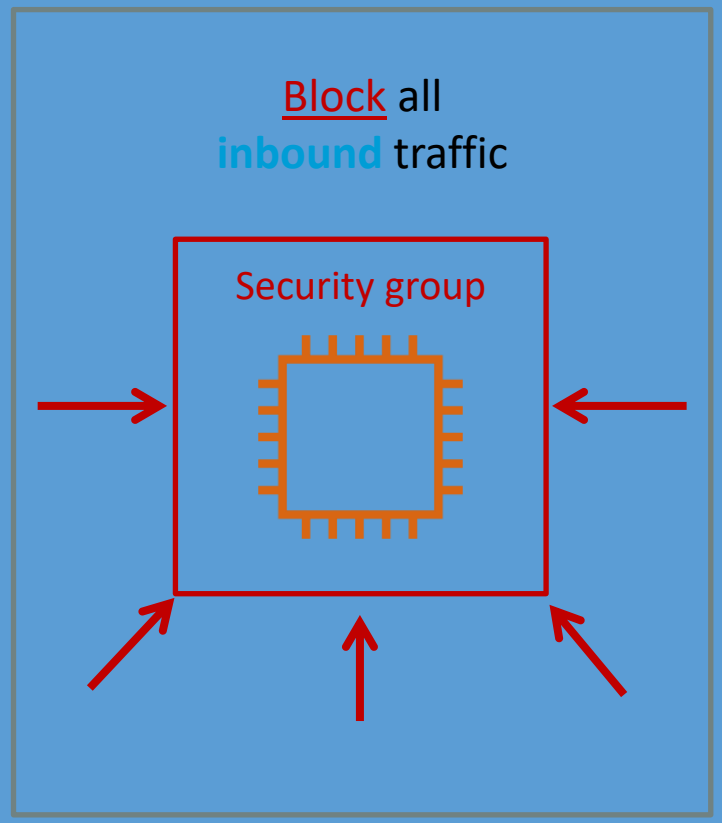
Security groups

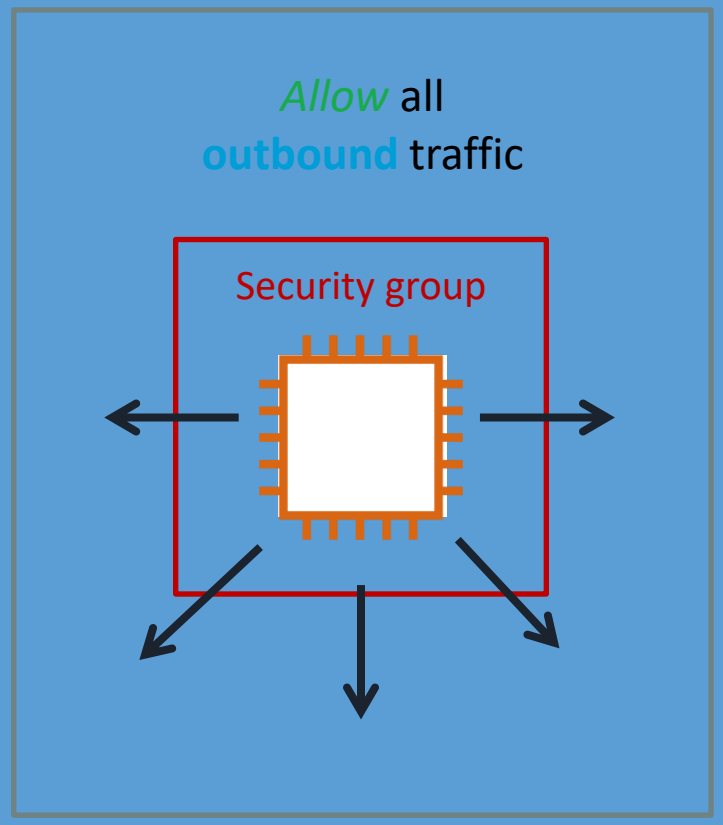
- A security group is a virtual firewall that controls inbound and outbound traffic into AWS resources.
- It allows traffic based on IP protocol, port, or IP address.
- It uses stateful rules.



Default and new security groups

- Security groups in default VPCs allow all outbound traffic.
- Custom security groups have no inbound rules and allow outbound traffic.





Custom security group rules

Inbound

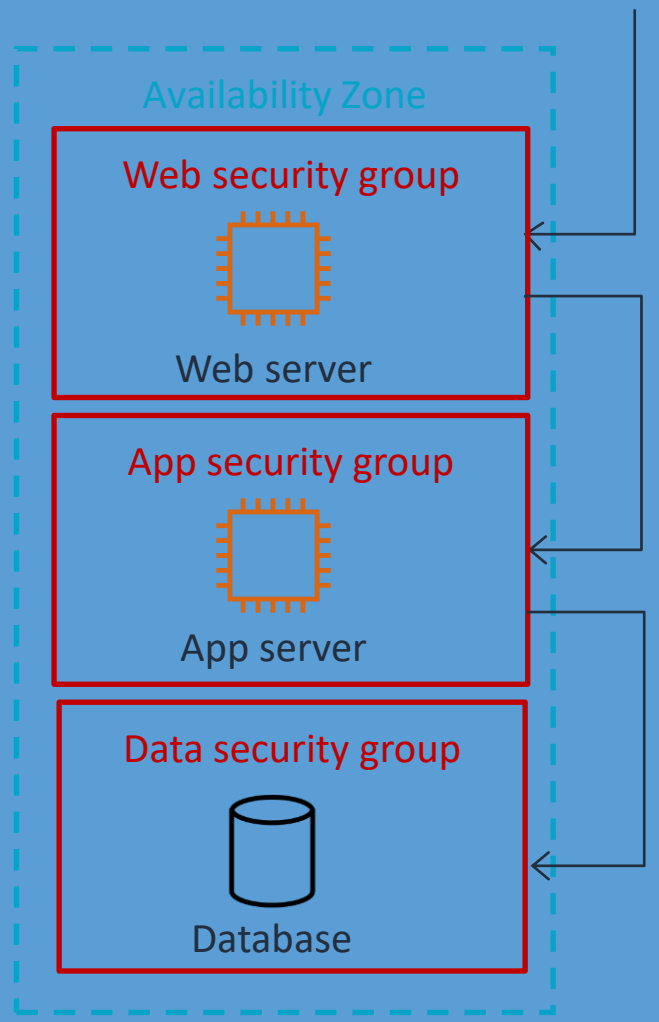
Source	Protocol	Port	Comments
0.0.0.0/0	TCP	80	Allows inbound HTTP access from all IPv4 addresses
0.0.0.0/0	TCP	443	Allows inbound HTTPS traffic from anywhere

Outbound

Destination	Protocol	Port	Comments
SG ID of DB servers	TCP	1433	Allows outbound Microsoft SQL Server access to instances in the specified security group
SG ID of MySQL servers	TCP	3306	Allows outbound MySQL access to instances in the specified security group

Security group chaining

- Inbound and outbound rules allow traffic flow from the top tier to the bottom tier.
- The security groups act as firewalls to prevent a subnet-wide security breach.

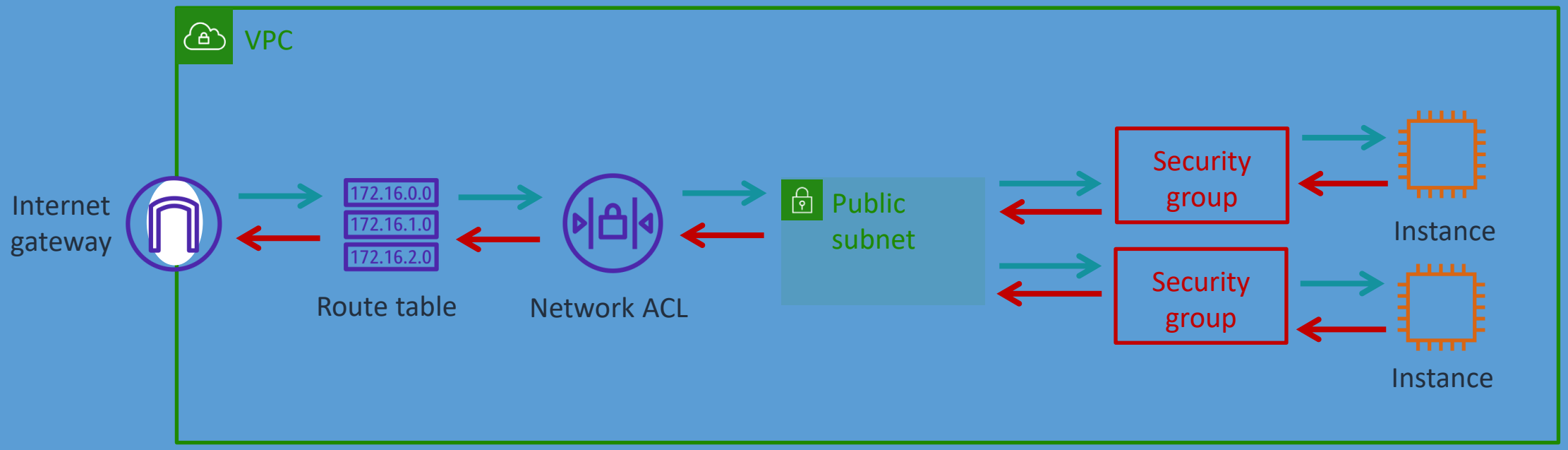


Inbound rule
Allow HTTPS port 443
Source: 0.0.0.0/0 (any)

Inbound rule
Allow HTTP port 80
Source: Web tier

Inbound rule
Allow TCP port 3306
Source: App tier

Design your infrastructure with multiple layers of defense

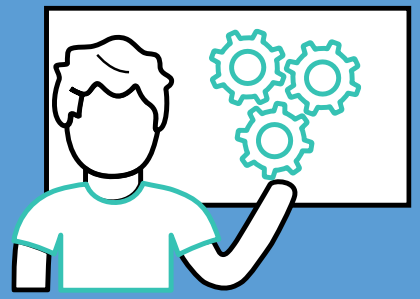


Comparing security groups and network ACLs

Security Group	Network ACL
Associated to an elastic network interface and implemented in the hypervisor	Associated to a subnet and implemented in the network
Supports Allow rules only	Supports Allow rules and Deny rules
A stateful firewall	A stateless firewall
All rules evaluated before deciding whether to allow traffic	All rules processed in order when deciding whether to allow traffic
Applies to an instance only if it is associated with the instance	Applies to all instances deployed in the associated subnet

Demonstration:

Create a security group for a public instance



Review

Present solutions

Network Engineer

Consider how you would answer the following:

- How can we make sure that our network has enough IP addresses to support our workloads?
- How do we build a dynamic and secure network infrastructure in our AWS account?
- How can we filter inbound and outbound traffic to protect resources on our network?

Module review

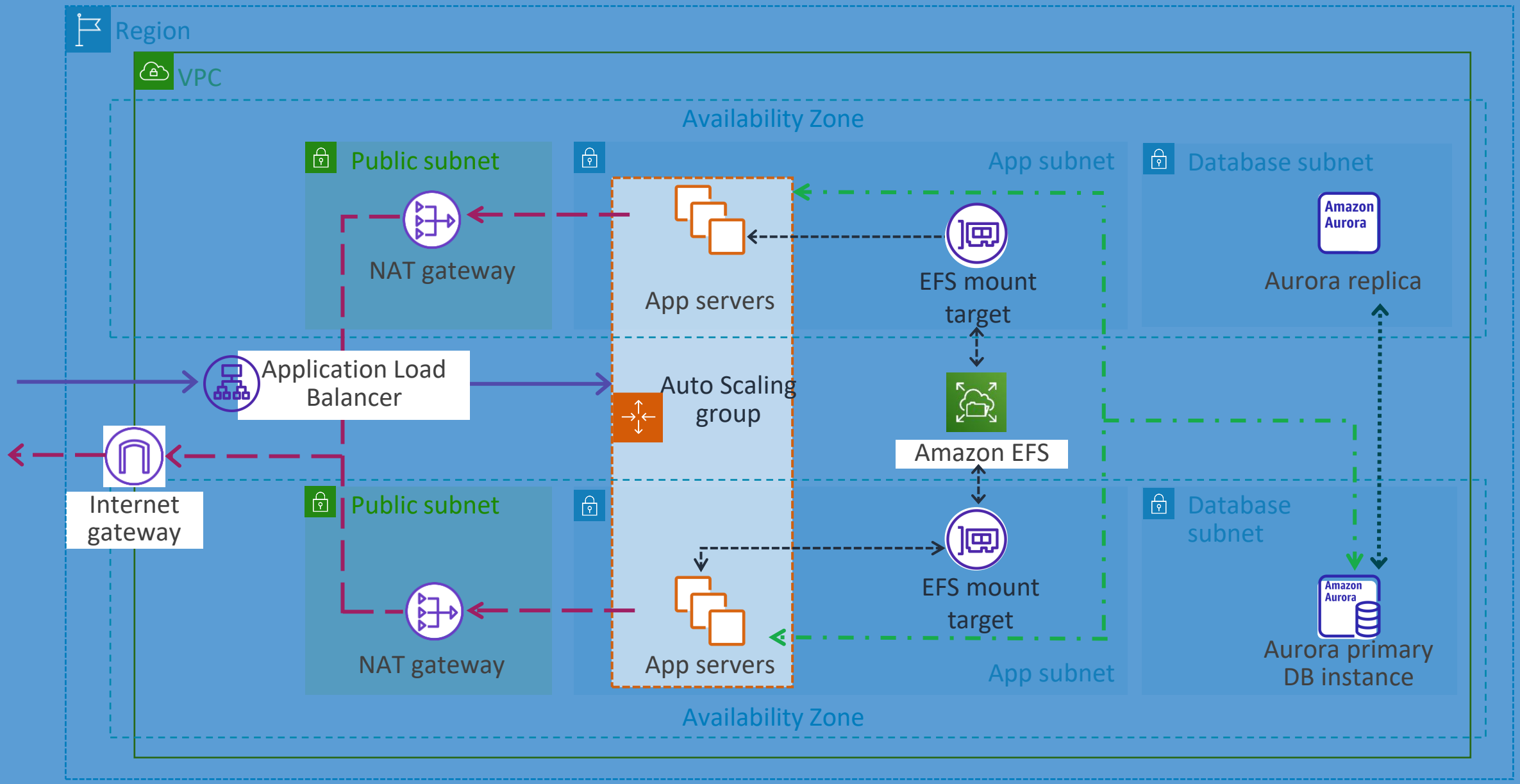
In this module you learned about:

- ✓ IP addresses
- ✓ VPC fundamentals
- ✓ VPC traffic security

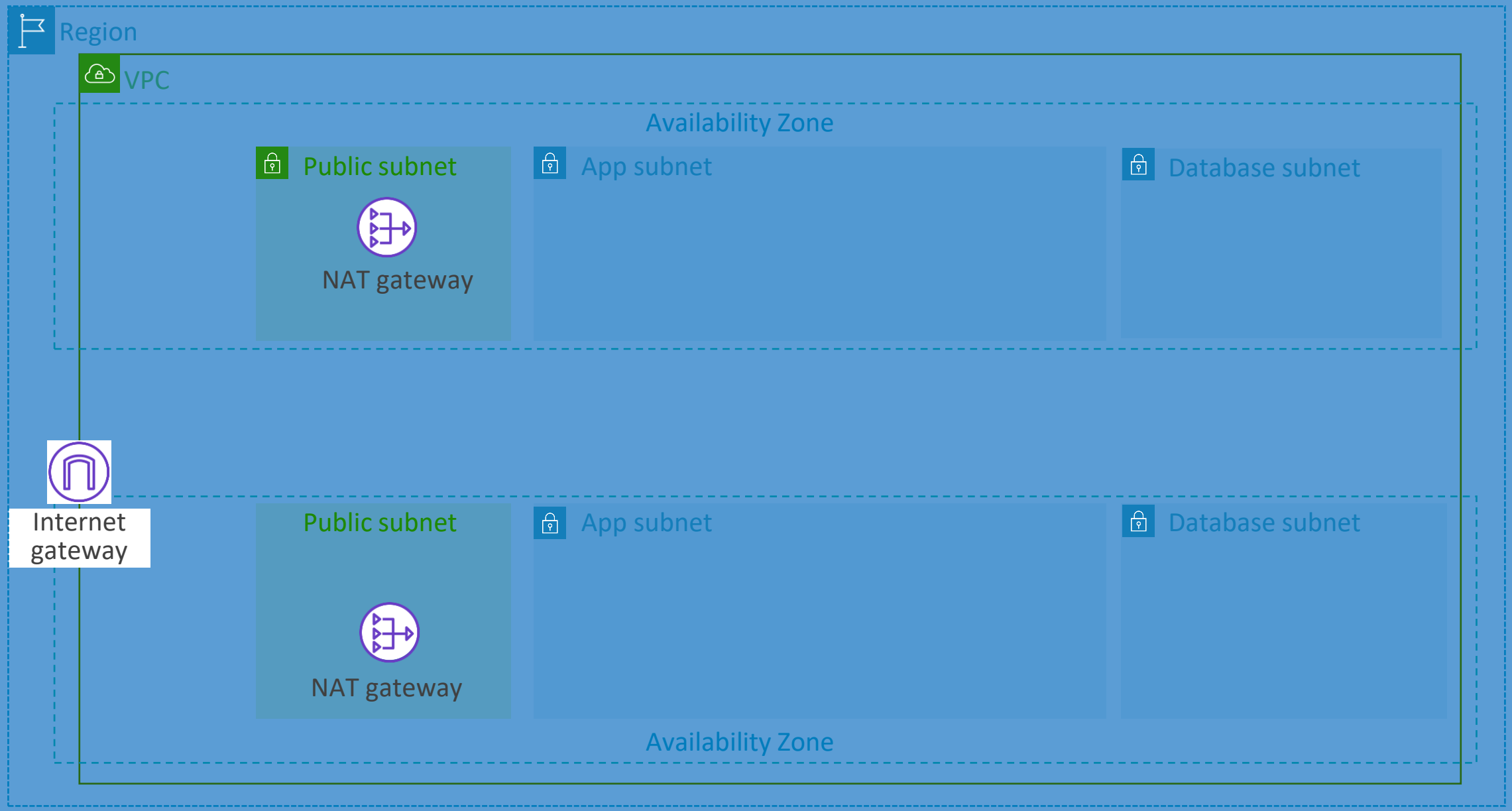
Next, you will review:

- Capstone check-in
- Knowledge check

Capstone architecture



Capstone architecture check-in



Knowledge check

Knowledge check question 1

True or False: A single Amazon VPC can span multiple Regions.

A	True
B	False

Knowledge check question 1 and answer

True or False: A single Amazon VPC can span multiple Regions.

A	True
B correct	False

Knowledge check question 2

What action must you take to make a subnet public?

A	Route outbound traffic from the subnet.
B	Route inbound traffic from the internet gateway.
C	Route outbound traffic to the internet gateway.
D	Subnets are public by default.

Knowledge check question 2 and answer

What action must you take to make a subnet public?

A	Route outbound traffic from the subnet.
B	Route inbound traffic from the internet gateway.
C correct	Route outbound traffic to the internet gateway.
D	Subnets are public by default.

Knowledge check question 3

What function does the NAT gateway serve?

A	Load balances incoming traffic to multiple instances
B	Allows internet traffic initiated by private subnet instances
C	Allows instances to communicate between subnets
D	Increases security for instances in a public subnet

Knowledge check question 3 and answer

What function does the NAT gateway serve?

A	Load balances incoming traffic to multiple instances
B correct	Allows internet traffic initiated by private subnet instances
C	Allows instances to communicate between subnets
D	Increases security for instances in a public subnet

Knowledge check question 4

What should you use to create traffic filtering rules for a subnet?

A	NAT gateway
B	Route table
C	Security group
D	Network ACL

Knowledge check question 4 and answer

What should you use to create traffic filtering rules for a subnet?

A	NAT gateway
B	Route table
C	Security group
D correct	Network ACL

Knowledge check question 5

Which ports are open by default when you create a new security group? (Select TWO.)

A	Nothing allowed inbound
B	Nothing allowed outbound
C	Anything allowed inbound
D	Anything allowed outbound
E	Inbound traffic is allowed on public subnets

Knowledge check question 5 and answer

Which ports are open by default when you create a new security group? (Select TWO.)

A correct	Nothing allowed inbound
B	Nothing allowed outbound
C	Anything allowed inbound
D correct	Anything allowed outbound
E	Inbound traffic is allowed on public subnets

AWS

Compute

Lab 2

Question

Where do you run the majority of your compute workloads?

A. On-premises physical or virtual servers
B. Cloud-based servers
C. On-premises containers
D. Cloud-based containers

Module overview

- Business request
- Compute services
- Amazon Elastic Compute Cloud (Amazon EC2) instances
- EC2 instance storage
- Amazon EC2 pricing options
- AWS Lambda
- Present solutions
- Knowledge check
- Capstone check-in
- Lab 2: Build your Amazon VPC infrastructure

Business Requirements

Compute Operations Manager

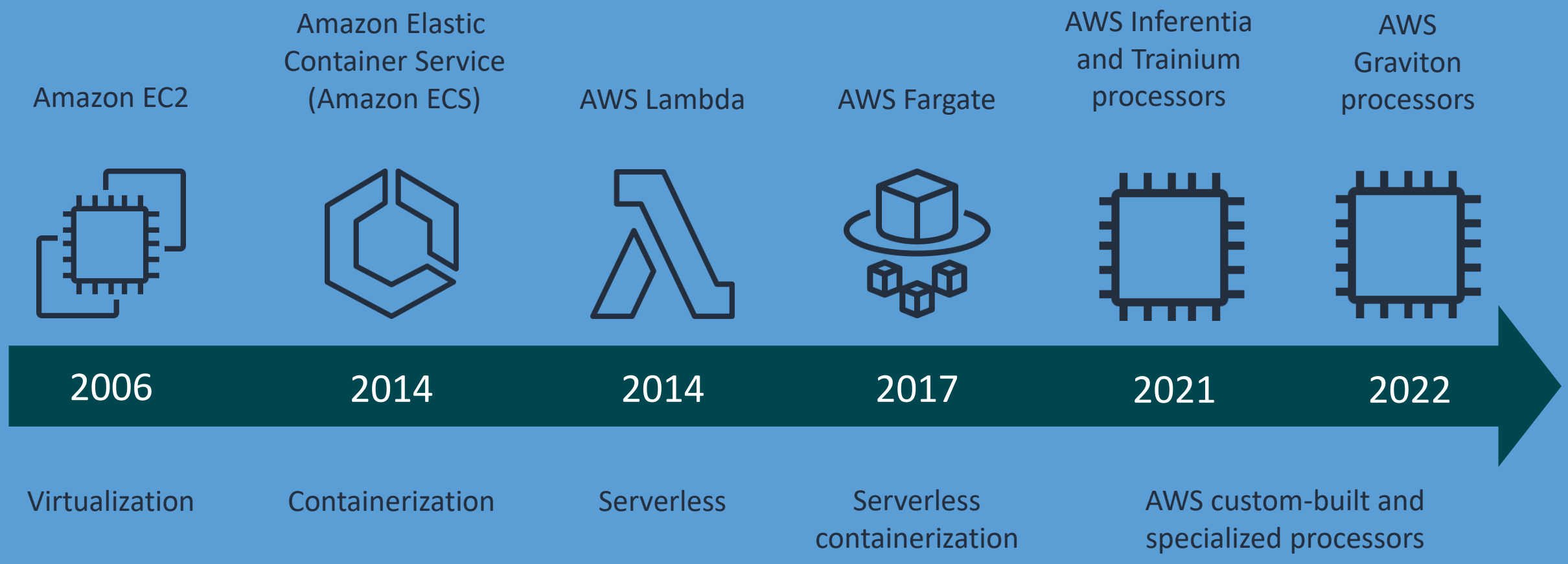
The compute operations manager wants to know:

- What AWS compute services are there?
- What should the team consider when deploying new and existing servers to Amazon EC2?
- How do we know which volume type to attach to our EC2 instances?
- How can we optimize cost for compute resources?
- Where can we start with serverless compute options?

Compute services

“What AWS compute services are there?”

Evolution of AWS compute



AWS services in this module

Amazon Elastic Compute Cloud (Amazon EC2)

Amazon Elastic Block Store (Amazon EBS)

AWS Lambda

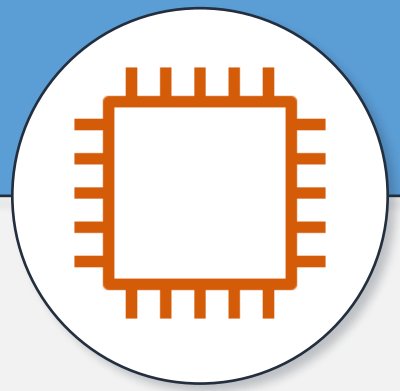
We will cover other compute-related services later in this course.

EC2 instances

“What should the team consider when deploying new and existing servers to Amazon EC2?”

EC2 instances

Physical servers host EC2 instances in AWS Regions around the world.



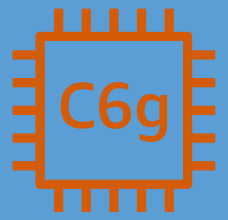
EC2 instances give you secure and resizable compute capacity in the cloud.

You can add or remove compute capacity to meet changes in demand.

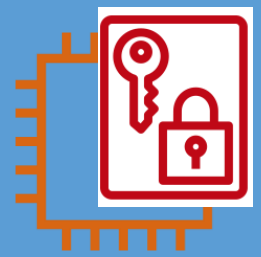
EC2 instance launch considerations

Name and tags

Application and OS image



Instance type and size

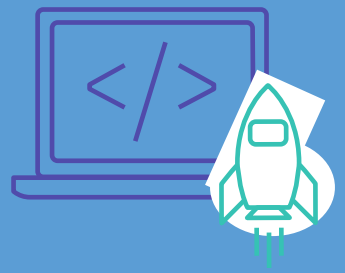


Key pair

Network and security

Storage

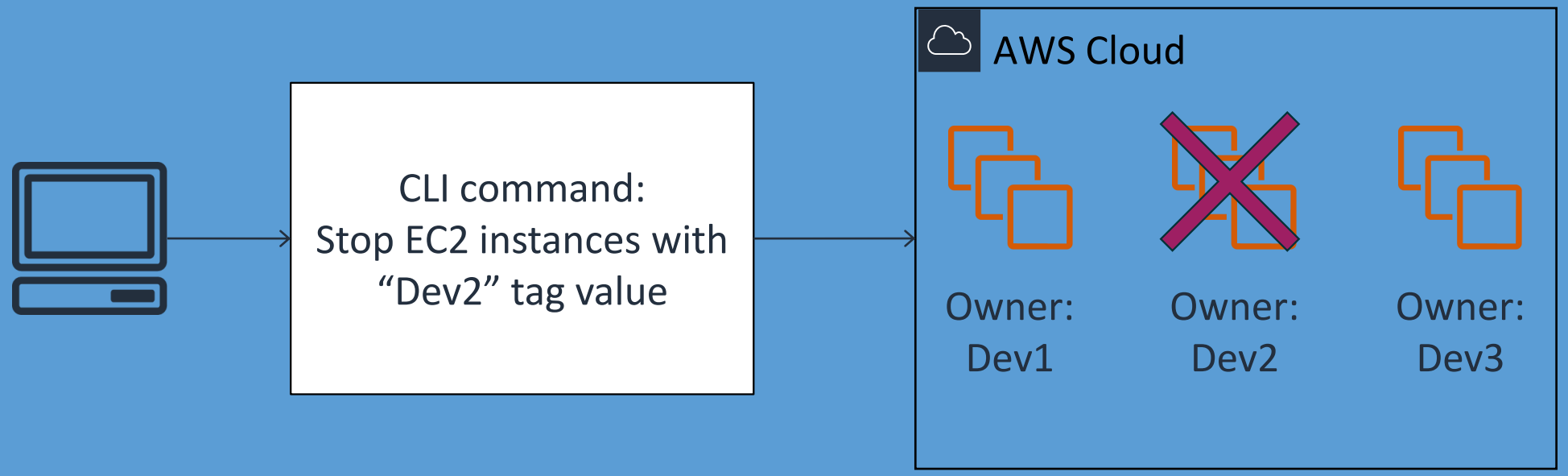
Placement and tenancy



Scripts and metadata

Tags in Amazon EC2

- Assign a name and other tags to your AWS resources.
- Manage, search, and filter resources.
- More tags are better than fewer.
- Tags are case-sensitive.



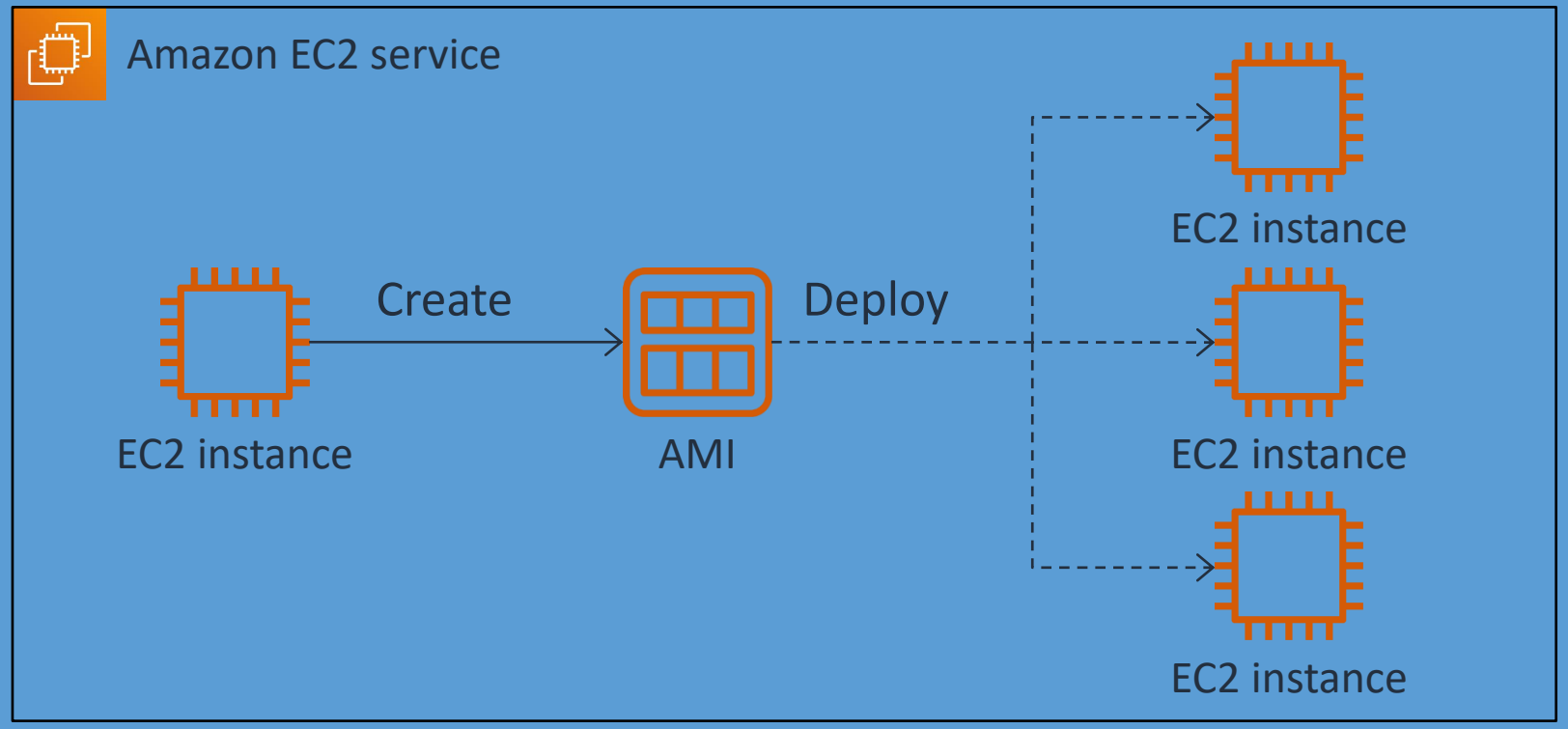
Amazon Machine Image (AMI)

AMI components:

- Template for instance volumes
- Launch permissions
- Block device mapping

Benefits:

- Repeatable
- Reusable
- Recoverable

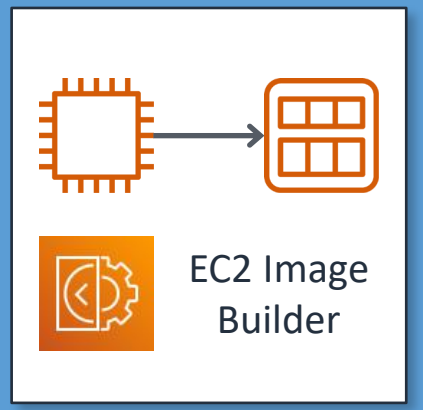


Where to get an AMI

Choose from the following:

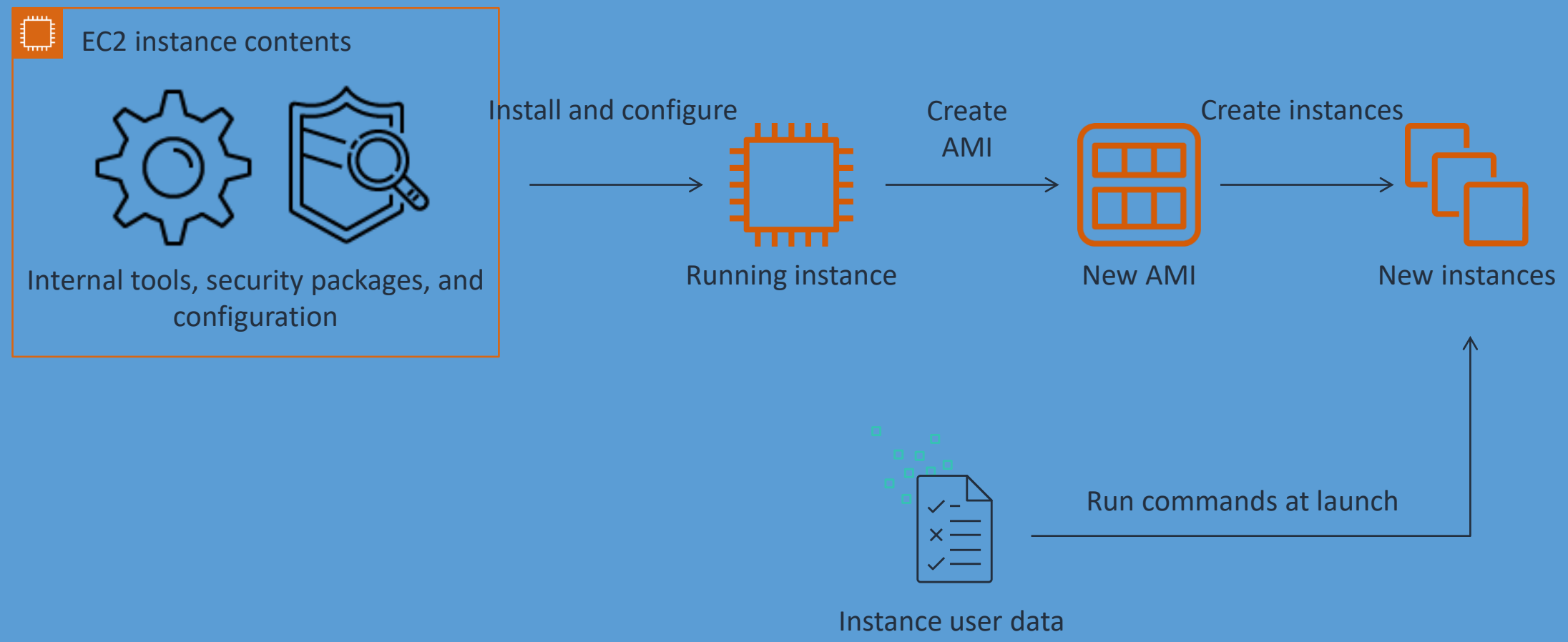
Use prebuilt AMIs offered by AWS.

Search the AWS Marketplace for a catalog with thousands of solutions.



Create your own AMIs manually, or use EC2 Image Builder

Creating custom AMIs



Amazon EC2 Image Builder

Start with a source image.

Customize software and configurations.

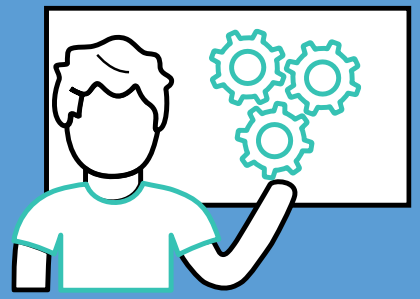
Secure the image with AWS provided or custom hardening templates.

Test the image with AWS provided or custom tests.

Distribute “golden” image to selected AWS Regions.

Demonstration

Create an AMI



Understanding instance type names

Instance generation

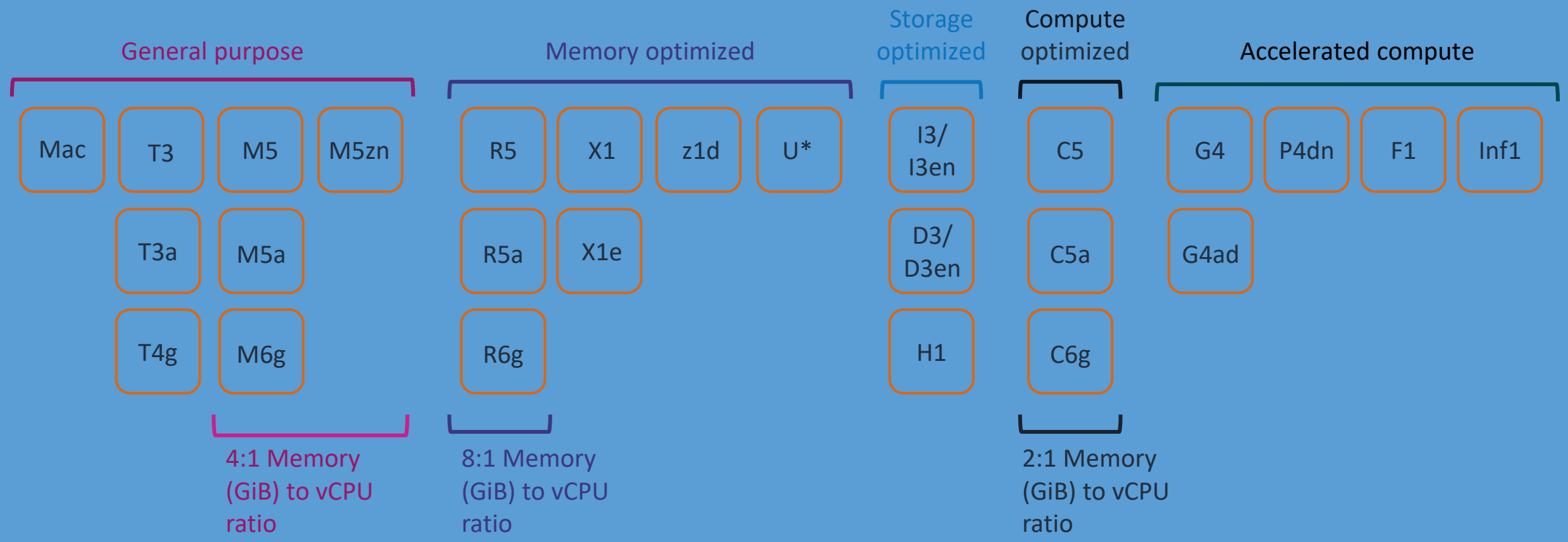
Instance size

c6g.xlarge

Instance family

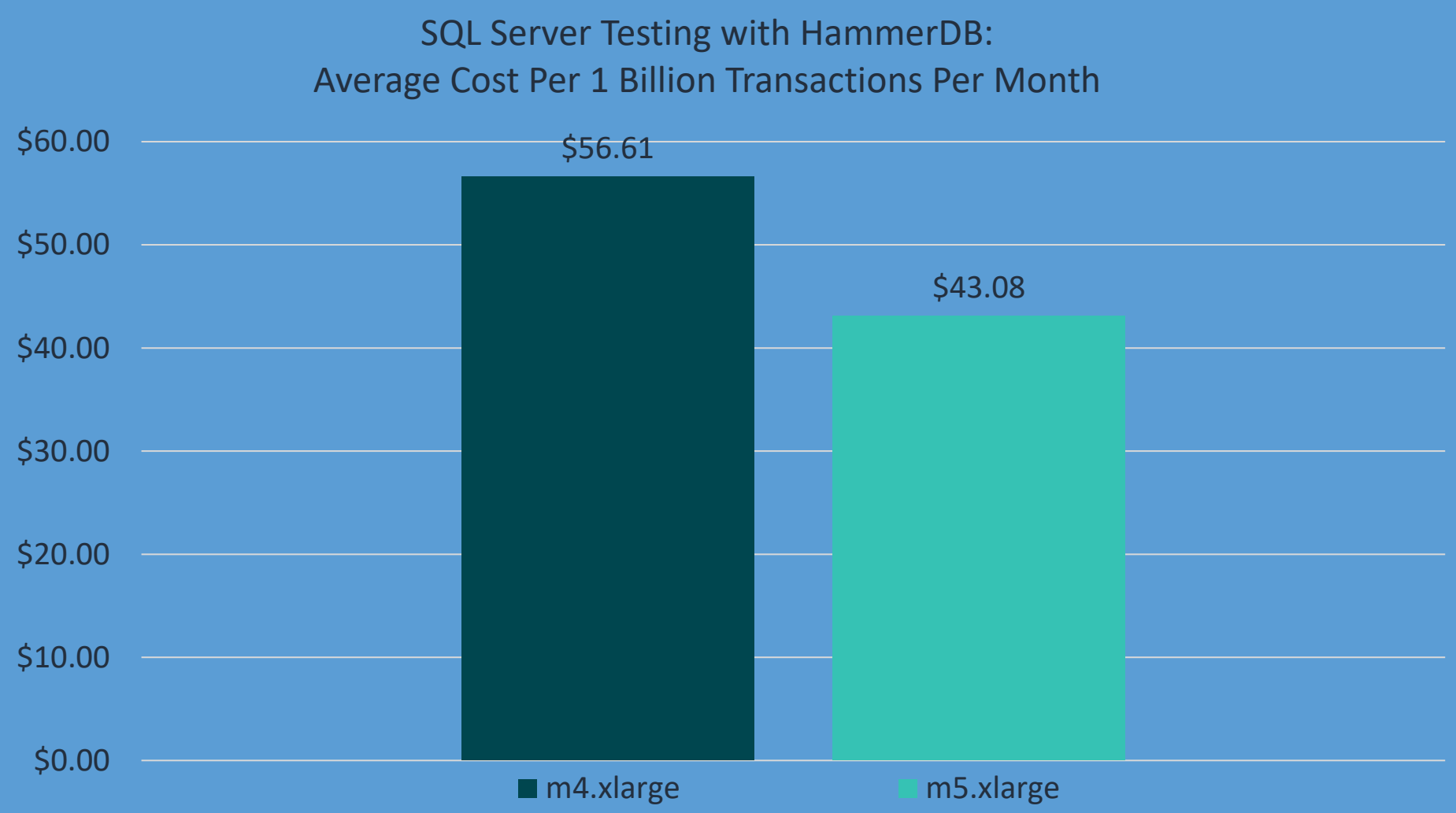
Additional properties

EC2 instance families

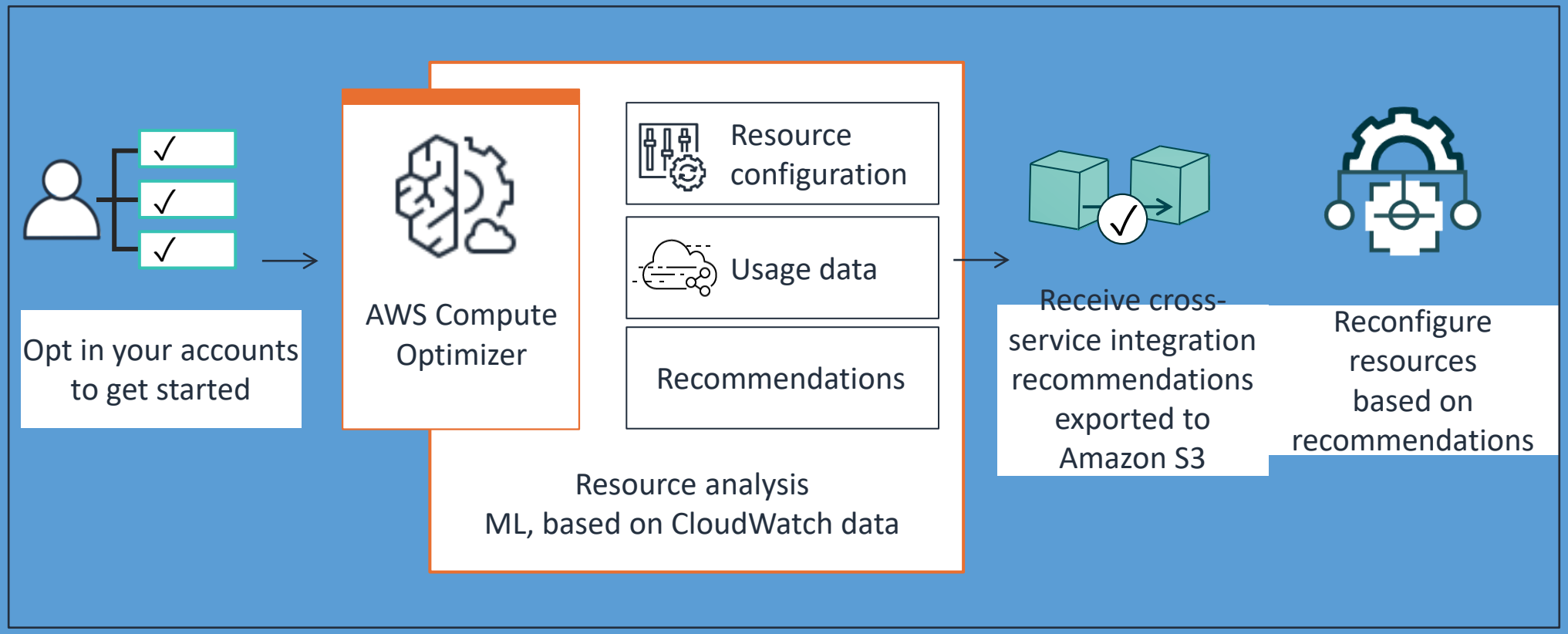


Benefits of newer generation instance types

Increase compute capabilities and reduce processing costs.



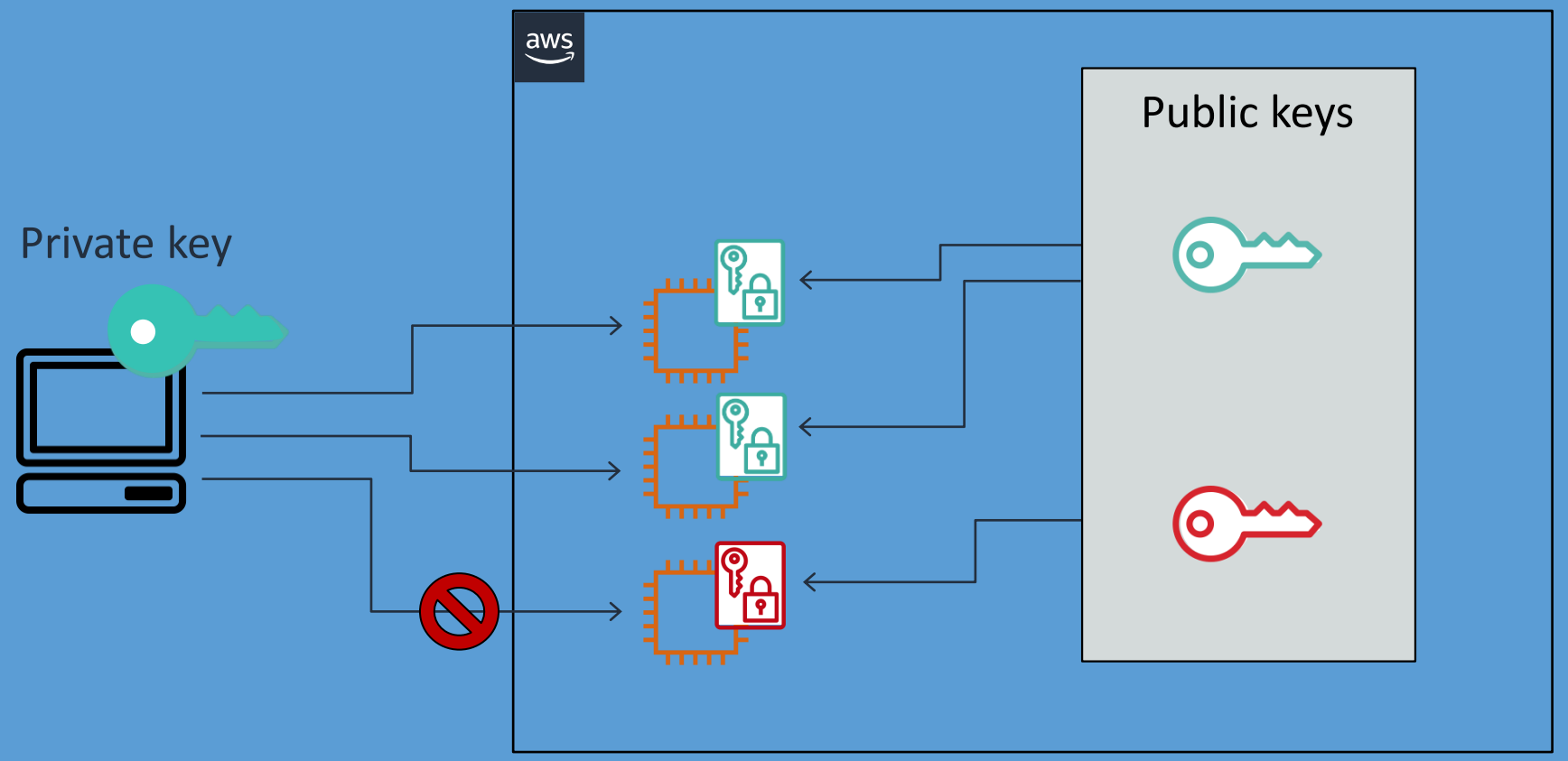
AWS Compute Optimizer



Apply insights from millions of workloads.

Save time by comparing and selecting resources.

Amazon EC2 key pairs



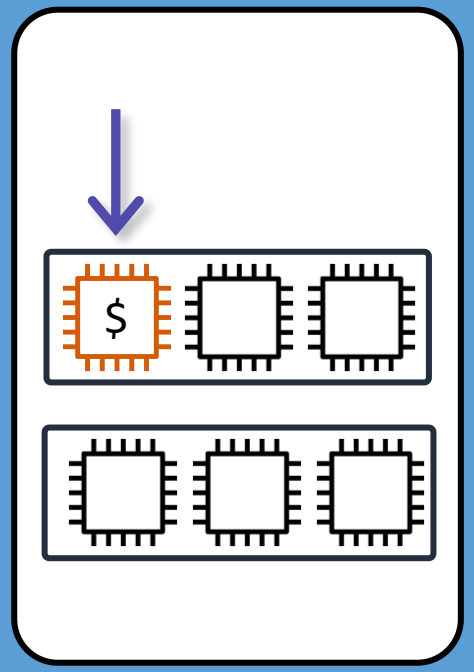
Elastic interface types

Elastic Network Interface	Elastic Network Adapter	Elastic Fabric Adapter
• Up to 10 Gbps • Virtual Machine Device Queue • TCP or IP • Multiple per instance • Traffic can traverse across subnets • VPC networking, general purpose • Default	• Up to 100 Gbps • SR-IOV • TCP or IP • Single setting per instance • Traffic can traverse across subnets • Low latency apps • Optional on supported instance type	• Up to 400 Gbps • OS bypass • Scalable Reliable Datagram • One per instance • OS bypass traffic is limited to single subnet and is not routable • HPC and ML apps • Optional on supported instance type

Tenancy

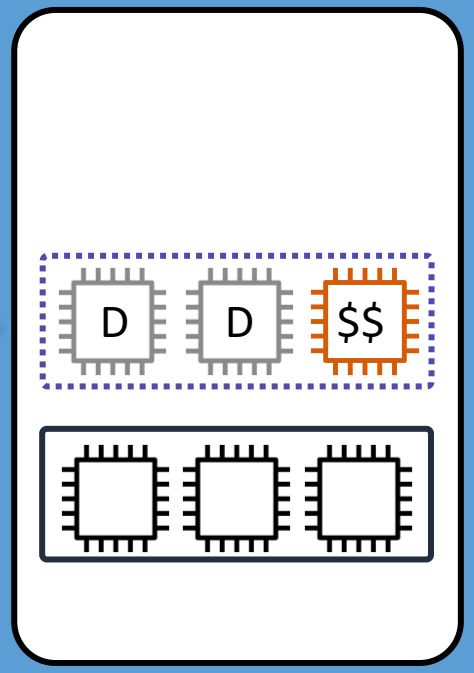
Shared tenancy

Share your hardware.



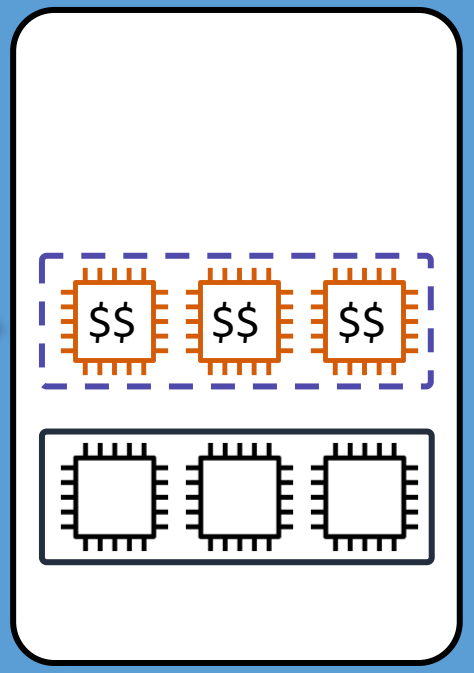
Dedicated Instance

Isolate your hardware.



Dedicated Host

Control your hardware.



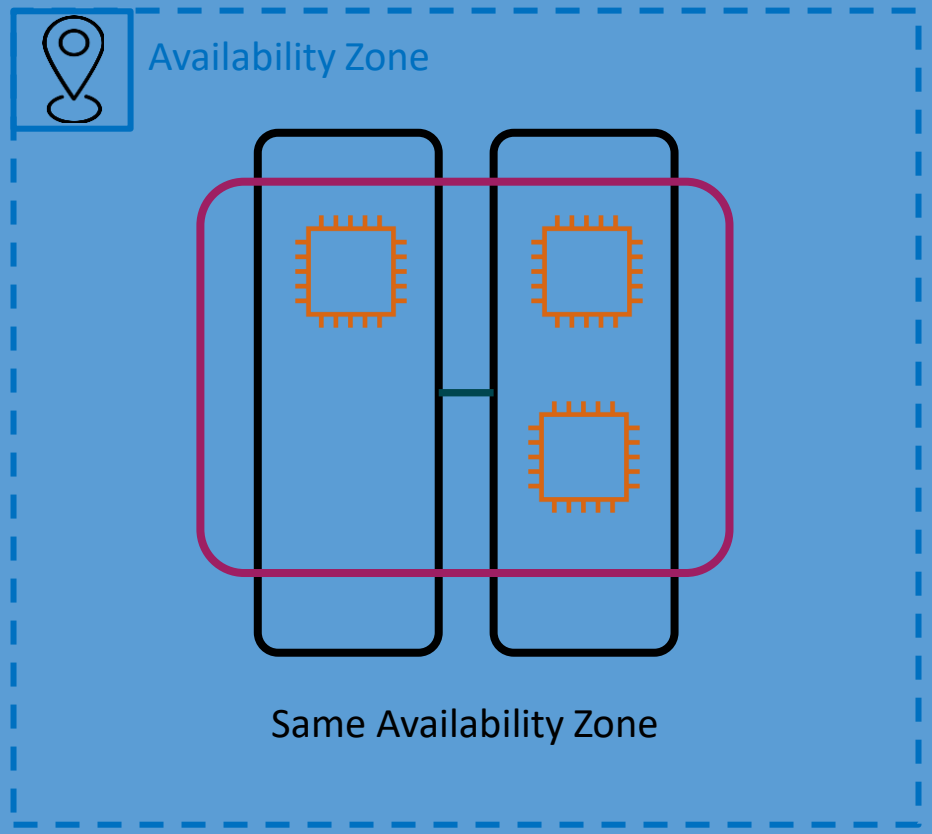
D = potential compute available
$ = purchased compute

Placement groups and use cases

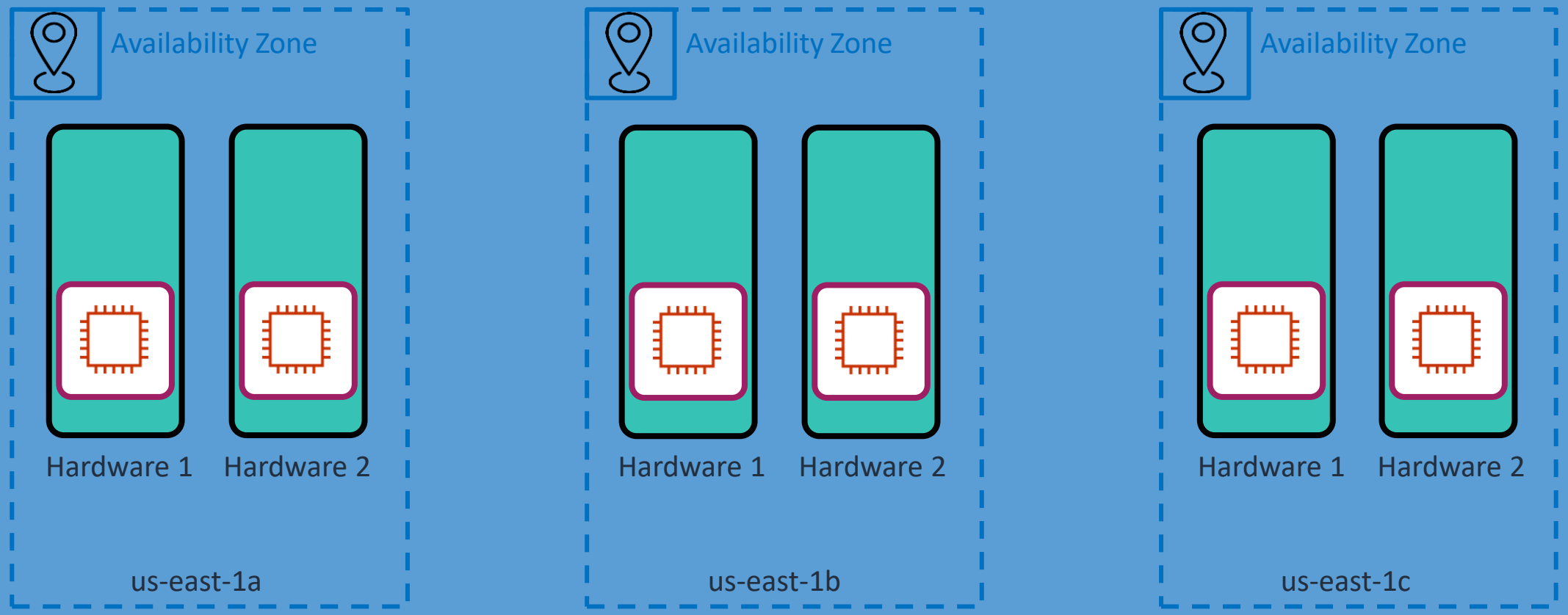
With placement groups, choose how close or far your instances are from each other.

Requirements	Solution	Example use case
Provide low network latency and high network throughput.	**Cluster** EC2 instances near each other.	High performance computing (HPC)
Critical instances must be fault-tolerant.	**Spread** across network segments and racks.	Medical health record system
Avoid correlated hardware failures.	**Partition** in logical groups on separate hardware.	Large distributed and replicated workloads like Kafka, Hadoop, and Cassandra

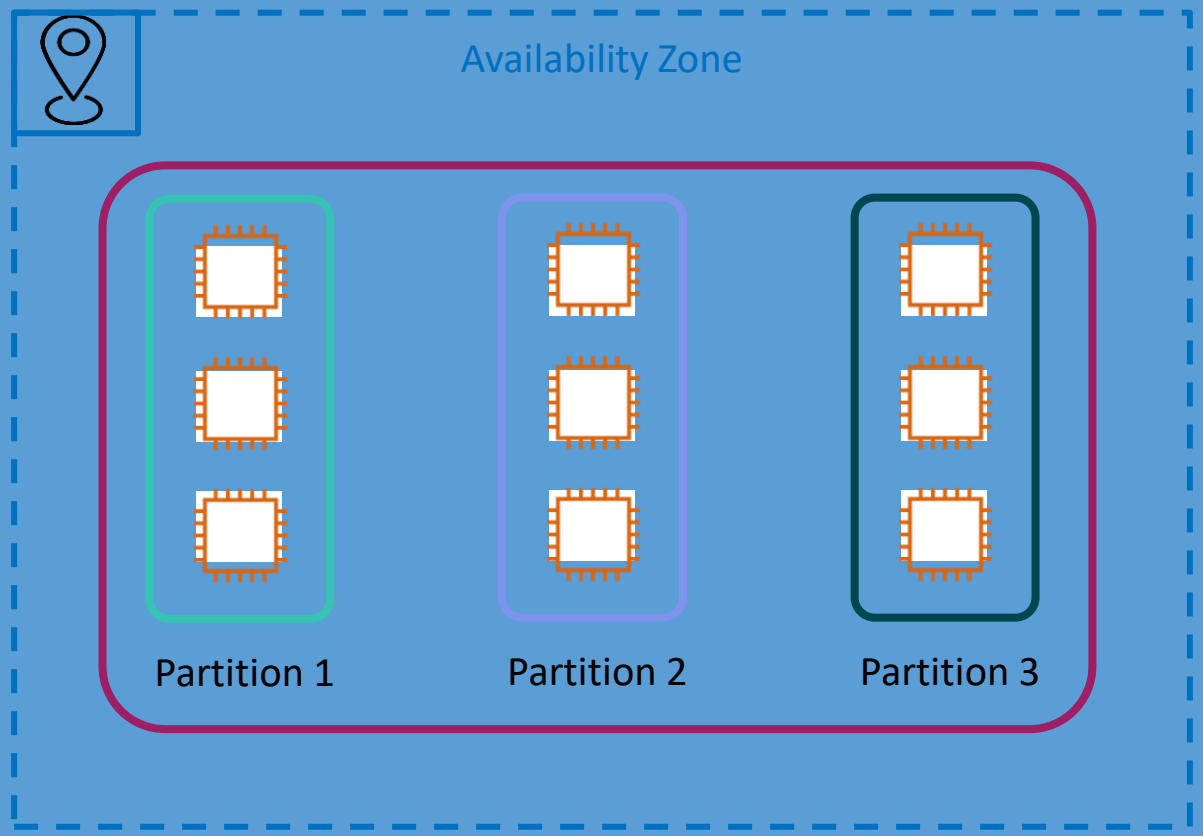
Cluster placement groups



Spread placement groups

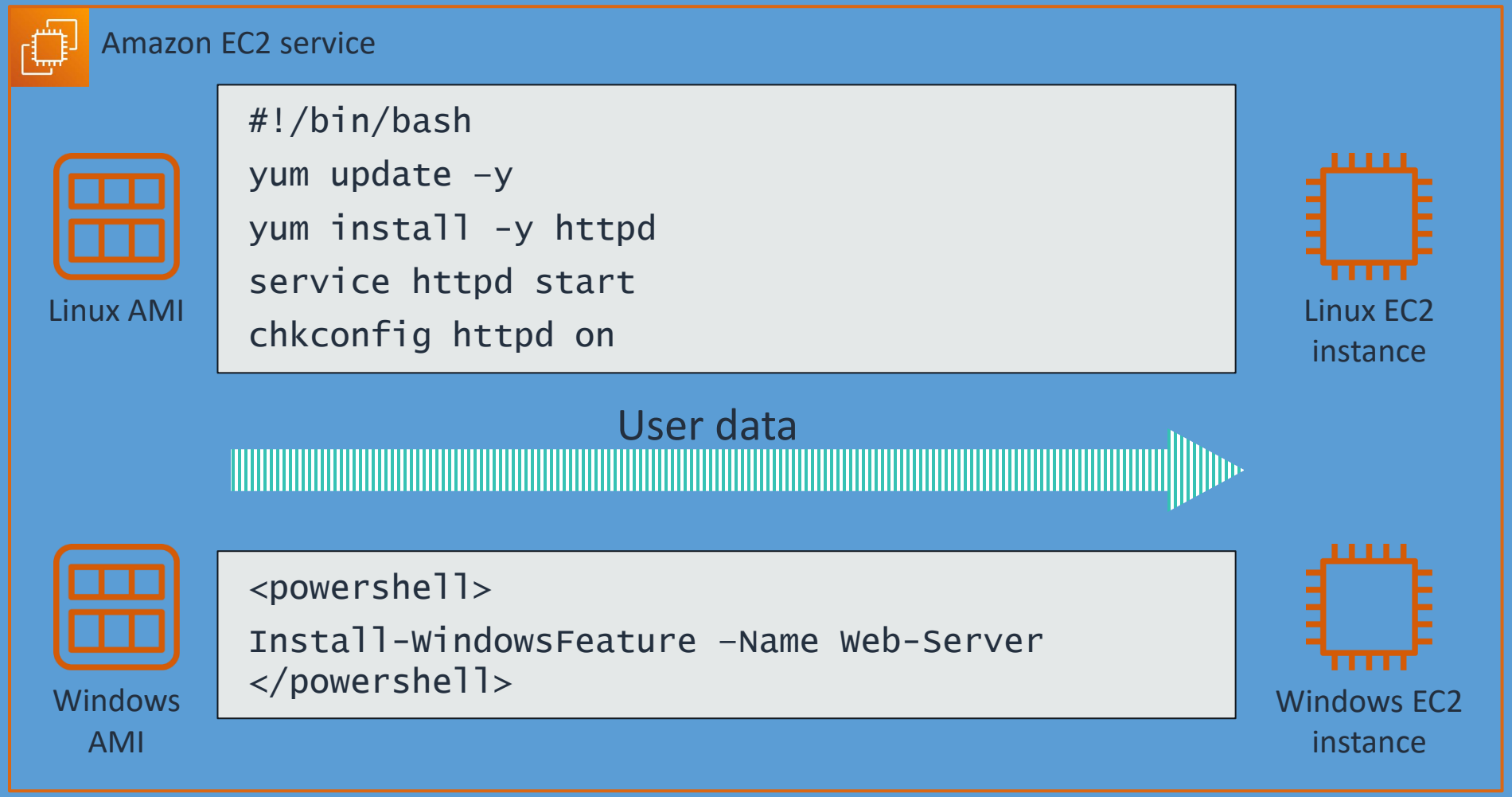


Partition placement groups

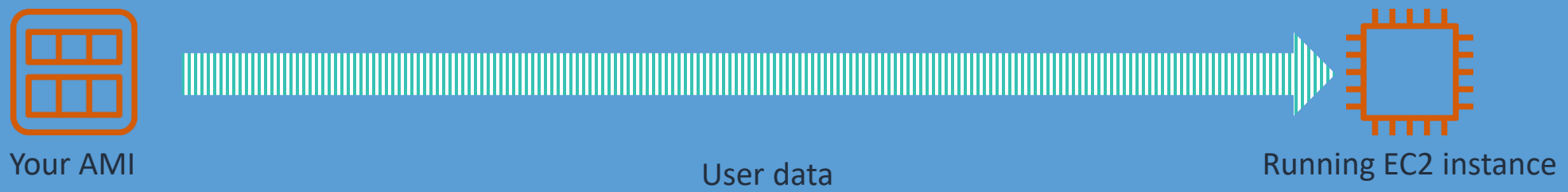


User data

- Runs scripts as root after the instance launches
- Can be used to perform common automated configuration tasks



Instance metadata




```
#!/bin/bash
yum update -y
hostname = $(curl -s http://169.254.169.254/latest/meta-data/public-hostname)
```

Data about the EC2 instance can be used for automation.

Note: You can only get metadata with a request from your EC2 instance.

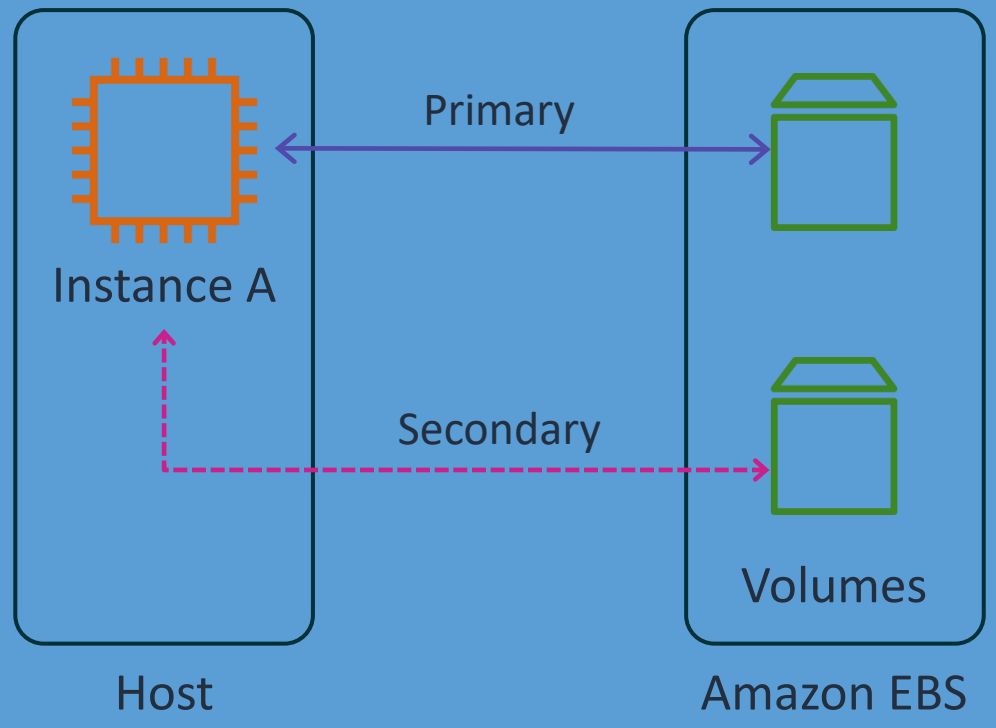
Metadata	Value
instance-id	i-1234567890abcdef0
mac	00-1B-63-84-45-E6
public-hostname	ec2-203-0-113-25.compute-1.amazonaws.com
public-ipv4	203.0.113.25
local-hostname	ip-10-251-50-12.ec2.internal
local-ipv4	10.251.50.12

Storage for EC2 instances

“How do we know which volume type to attach to our EC2 instances?”

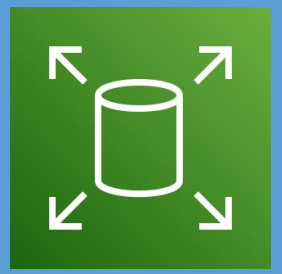
Amazon Elastic Block Store (Amazon EBS)

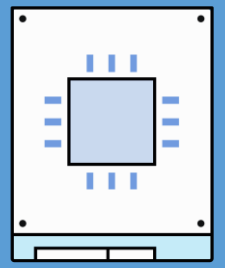
- Create block-level storage with automatic volume replication in your Availability Zone.
- Attach one or more EBS volumes to a single EC2 instance.
- Move EBS volumes between EC2 instances as needed.



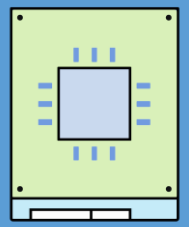
Amazon EBS volume types

- Solid state drive (SSD) is for high-performance and general-purpose workloads.
- Hard disk drive (HDD) is for big or infrequently accessed data.
- io2 includes options for Block Express.

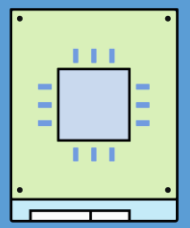




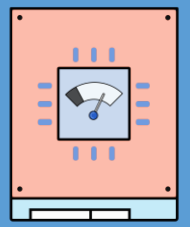
EBS SSD-backed volumes



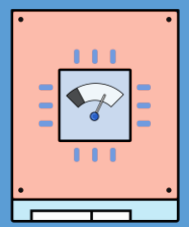
gp2



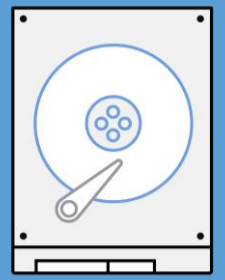
gp3



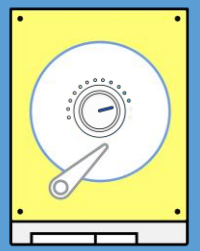
io1



io2



EBS HDD-backed volumes



st1

sc1

Amazon EBS volume characteristics (1 of 2)

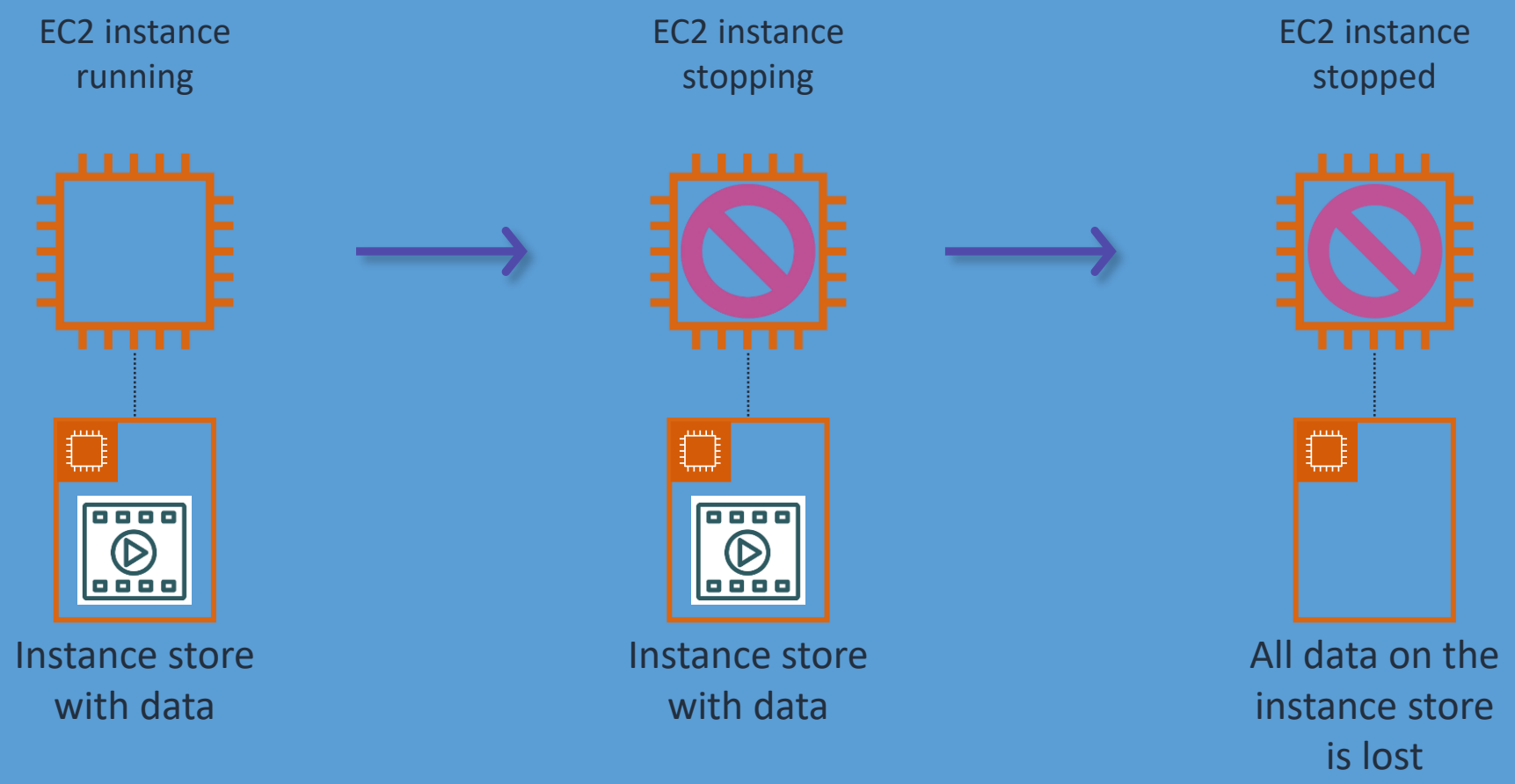
	General Purpose SSD		Provisioned IOPS SSD		
Volume type	gp2	gp3	io1	io2	io2 Block Express
Description	Volume that balances price and performance for a wide variety of workloads		Highest-performance SSD volume for mission-critical low-latency or high-throughput workloads		Next generation of Amazon EBS storage service architecture built for the cloud
Size	1 GiB to 16 TiB		4 GiB to 16 TiB		4 GiB to 64 TiB
Max IOPS	16,000 (burst)	16,000 (no burst)	64,000		256,000
Max throughput per volume	250 MiB/s	1,000 MiB/s	1,000 MiB/s		4,000 MiB/s

Amazon EBS volume characteristics (2 of 2)

	Throughput Optimized HDD	Cold HDD
Volume type	st1	sc1
Description	Low-cost HDD volume designed for frequently accessed, throughput-intensive workloads	Lowest-cost HDD volume designed for less frequently accessed workloads
Size	125 GiB to 16 TiB	125 GiB to 16 TiB
Max IOPS	500	250
Max throughput per volume	500 MiB/s	250 MiB/s

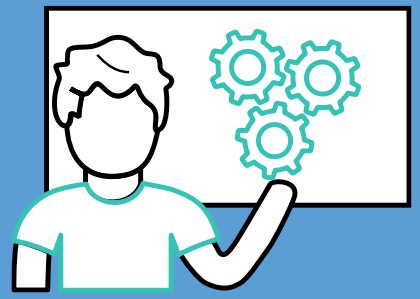
Instance store volumes

- Local to instance
- Non-persistent
- Doesn’t support snapshots
- Available in HDD, SSD, and non-volatile memory express SSD (NVMe SSD) varieties



Demonstration

Create an EBS volume



Amazon EC2 pricing options

“How can we optimize cost for compute resources?”

Amazon EC2 purchase options

On-Demand

Pay for compute capacity per second or hour with no long-term commitments

Spiky workloads or temporary needs

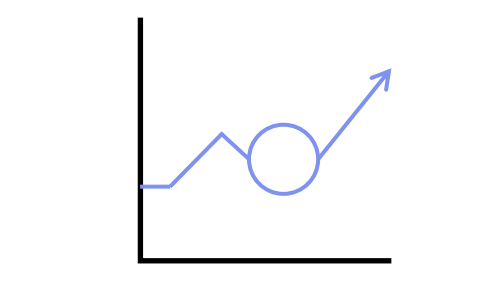
Savings Plans

1-year or 3-year commitment with varied flexibility based on type of Savings Plan

Committed flexible access to compute

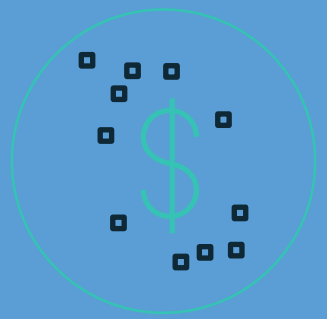
Spot Instances

Spare Amazon EC2 capacity at savings of up to 90% off On-Demand costs



Fault-tolerant, flexible, stateless workloads

Savings Plan types



Compute Savings Plans

Greatest flexibility, up to 66% off On-Demand rates, and applies to AWS Fargate and AWS Lambda usage.

Flexible across:

- Instance family
- Size
- OS
- Tenancy
- Availability Zone
- Region

EC2 Instance Savings Plans

Provide the lowest prices, up to 72% off On-Demand rates on the selected instance family in a specific AWS Region.

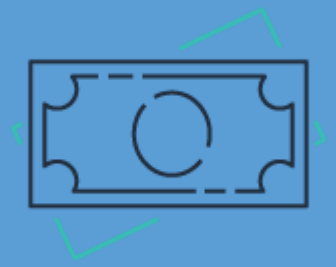
Flexible across:

- Availability Zone
- Size
- OS
- Tenancy

EC2 Spot Instances

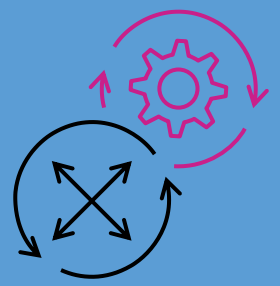
Use the same infrastructure

Run on the same hardware as On-Demand and Savings Plans.



Get the best value

Decide what you can pay for compute and save up to 90% from the On-Demand price.



Plan for interruptions

Prepare for capacity changes in your Availability Zones.

Diversify your fleet

Choose different instance types, size, and Availability Zone.

Use cases for Spot Instances

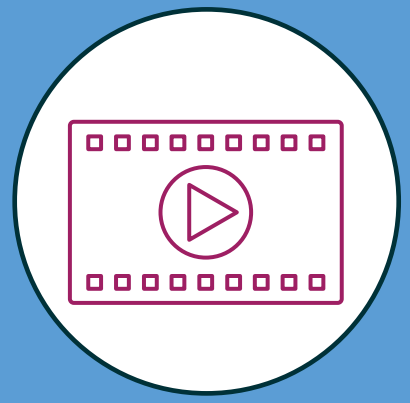


Image and media rendering

Manage rendering projects cost effectively to meet deadlines.

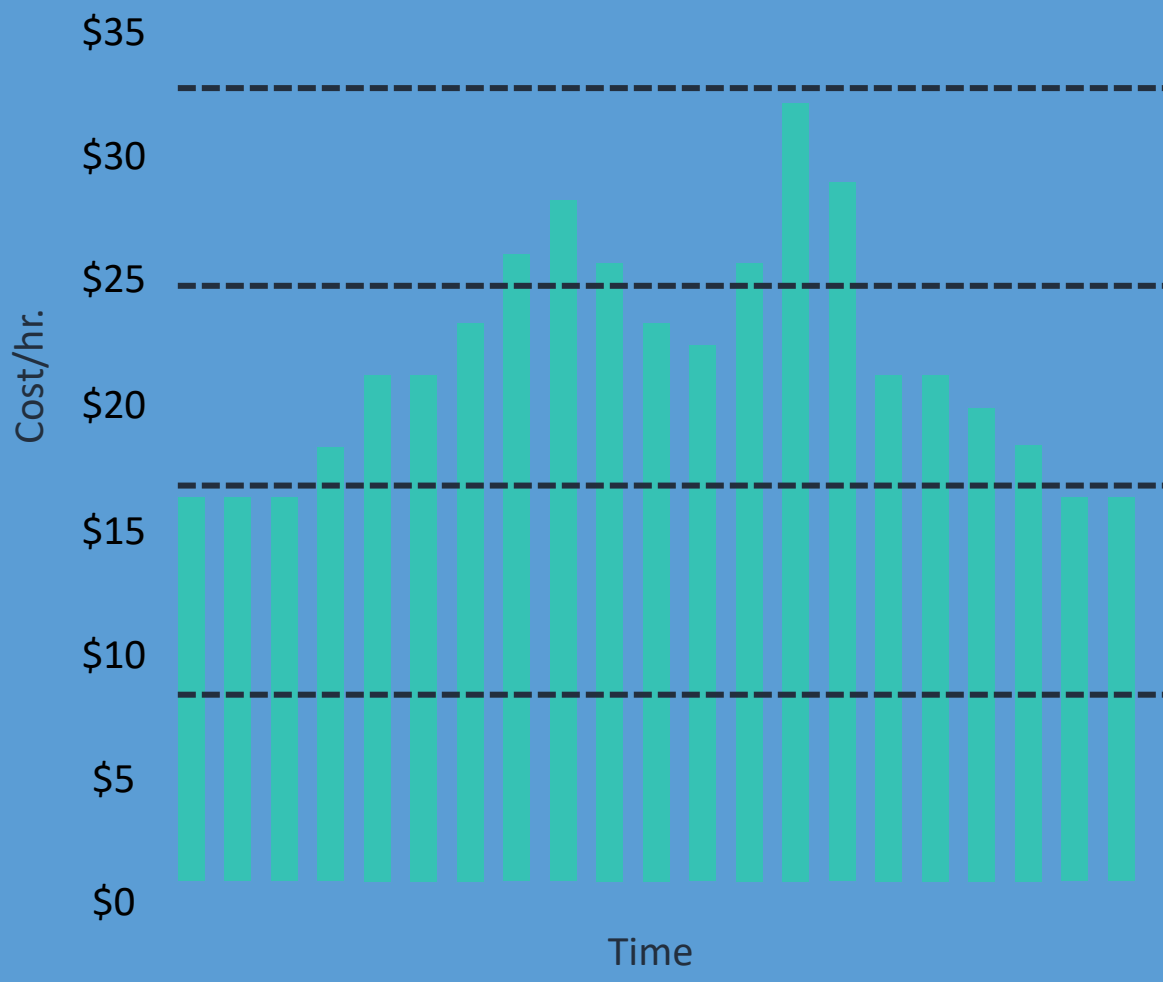
Web services

Launch Spot Instances to scale web services and applications at a lower cost.

Big data and analytics

Accelerate and scale time-critical, hyper-scale workloads.

Combining purchase options



Spiky, unpredictable Amazon EC2 usage that is not suitable for Spot	On-Demand List price
Workloads that are fault-tolerant, flexible, and stateless	Spot Instances Discounts up to 90%
Steady usage where the flexibility to change is important	Compute Savings Plans Discounts up to 66%
Steady usage unlikely to change during the commitment	EC2 Instance Savings Plans Discounts up to 72%

AWS Lambda

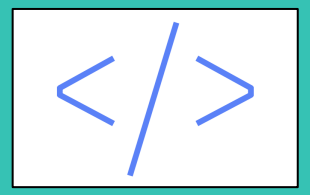
“Where can we start with serverless compute options?”

Serverless computing

- Highly available
- Fully managed by AWS

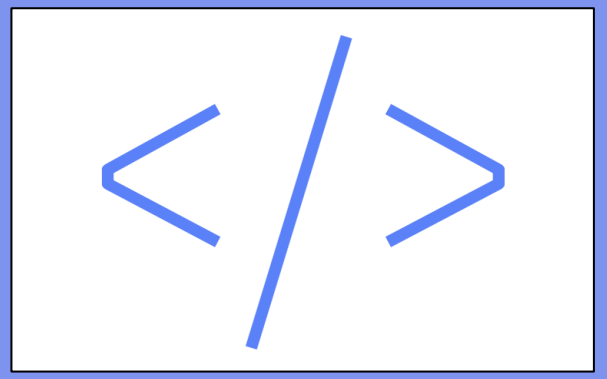
Computing with virtual servers

Servers



Code

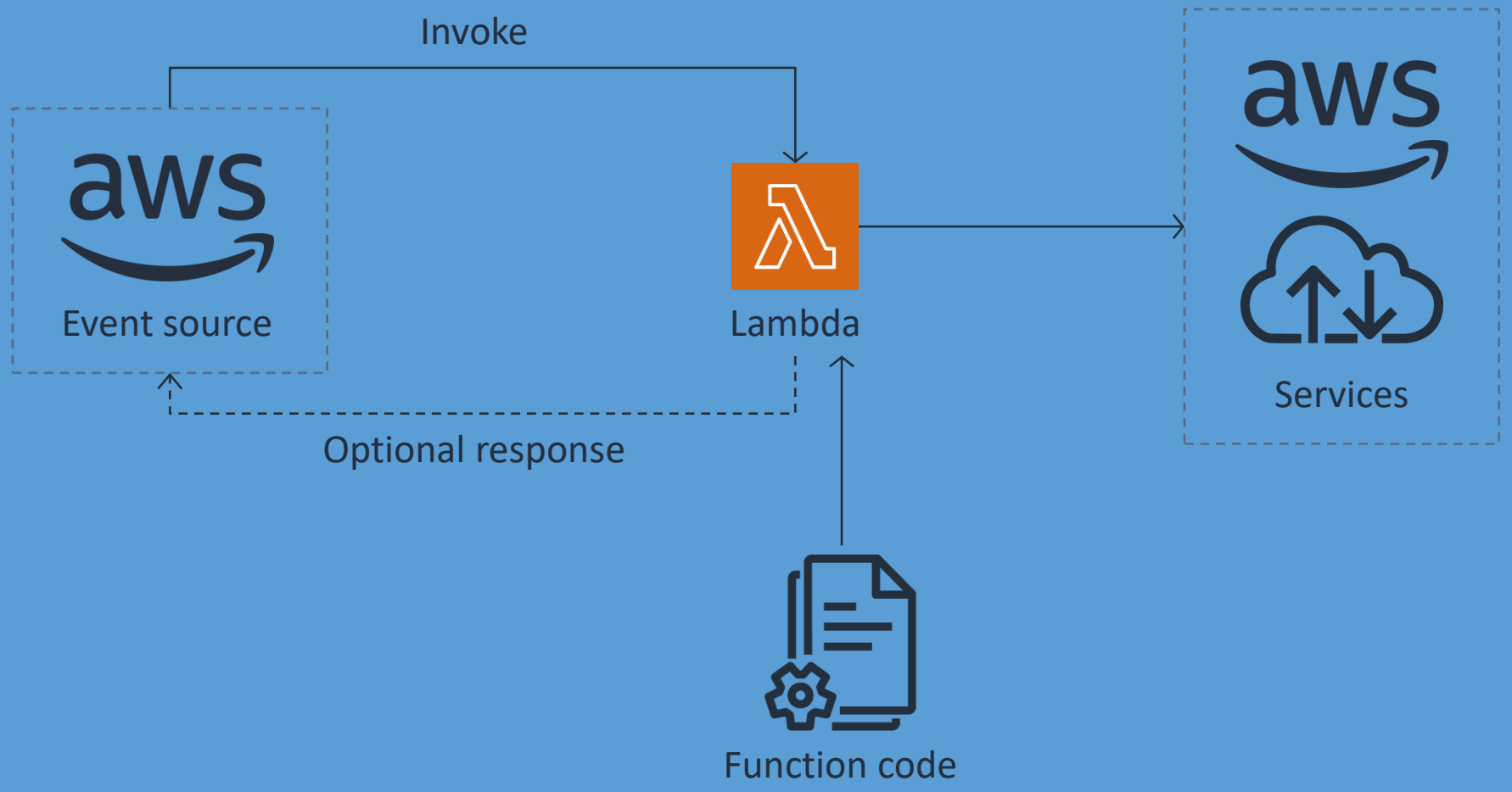
Serverless computing



Code

AWS Lambda

- Serverless compute
- Supports Node.js, Java, Python, C#, Go, PowerShell, Ruby, and more
- Runs for up to 15 minutes
- Supports up to 10 GB memory



Event source examples


Amazon DynamoDB


AWS CodeCommit


Amazon Simple Email Service (Amazon SES)


Amazon S3


AWS IoT services


Amazon Alexa


Amazon CloudWatch


AWS CloudFormation


Amazon Simple Queue Service (Amazon SQS)


Amazon Cognito


Amazon API Gateway


AWS CloudTrail


Amazon EventBridge


Amazon Simple Notification Service (Amazon SNS)


Application Load Balancer


AWS Step Functions


Amazon Kinesis

Anatomy of a Lambda function

Handler function

Function to be run upon invocation

Event object

Data sent during Lambda function invocation

Context object

Methods available to interact with runtime information (request ID, log group, more)

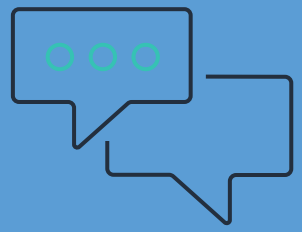
```
import json

def lambda_handler(event, context):
    # TODO implement
    return {
        'statusCode': 200,
        'body': json.dumps('Hello World!')
    }
```

Use cases

Web applications

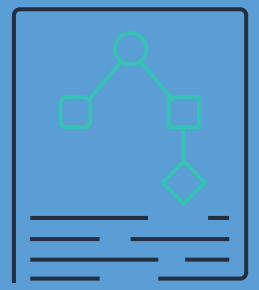
Backends

Data processing



Chatbots

Amazon Alexa



IT automation

Review

Present solutions

Compute Operations Manager

Consider how you would answer the following:

- What AWS compute services are there?
- What should the team consider when deploying new and existing servers to Amazon EC2?
- How do we know which volume type to attach to our EC2 instances?
- How can we optimize cost for compute resources?
- Where can we start with serverless compute options?

Module review

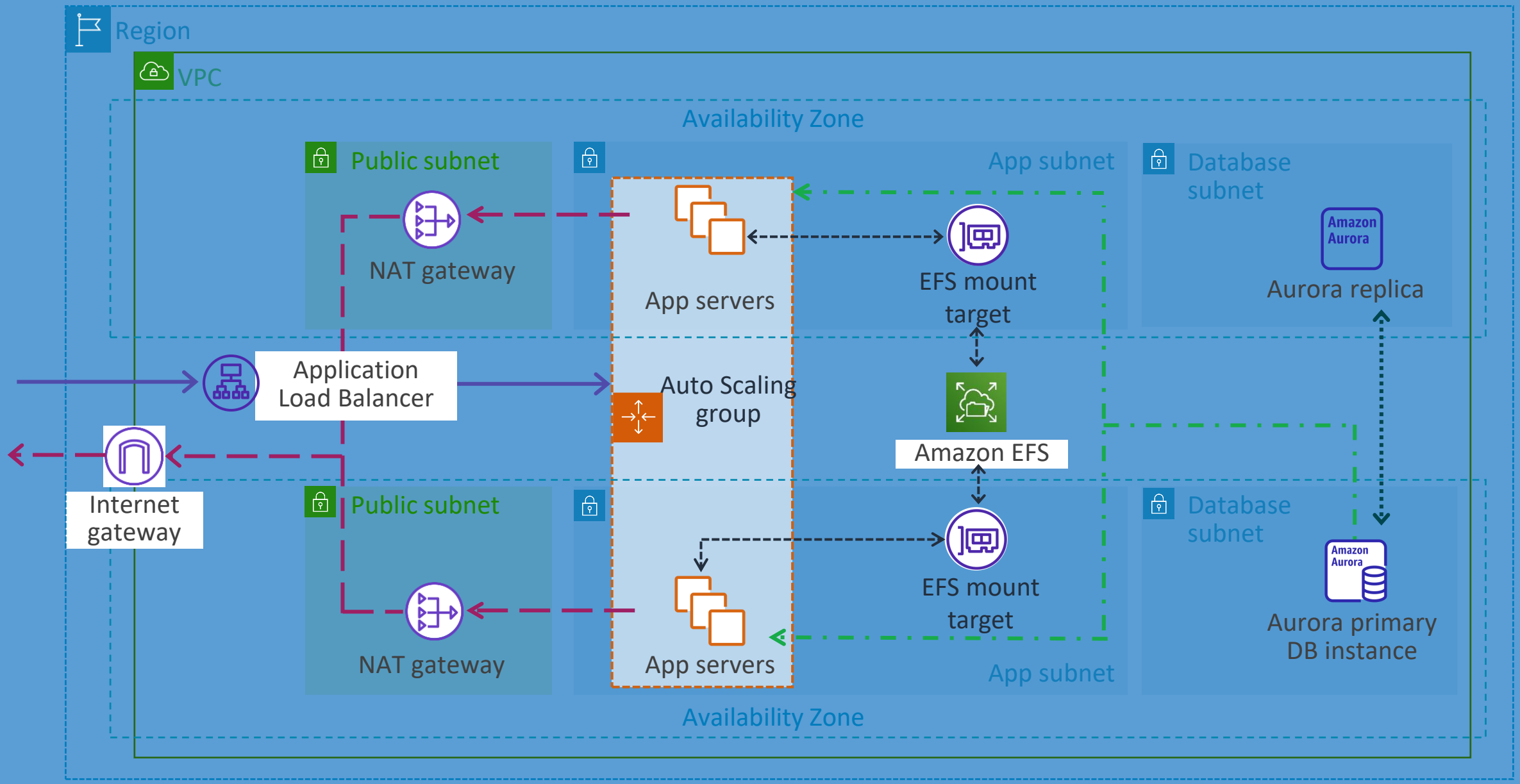
In this module you learned about:

- ✓ Compute services
- ✓ EC2 instances
- ✓ Instance storage
- ✓ Amazon EC2 pricing options
- ✓ AWS Lambda

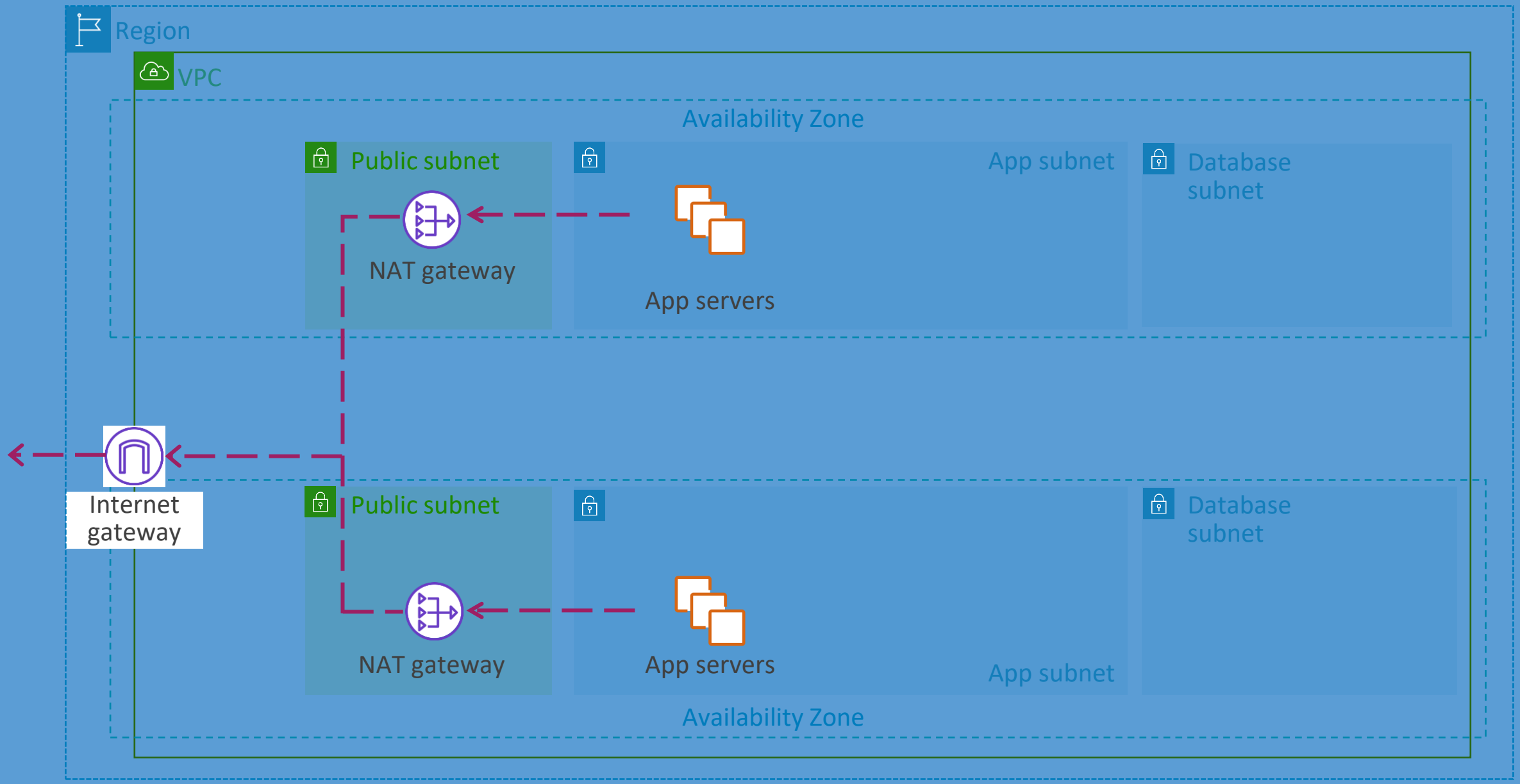
Next, you will review:

- Capstone check-in
- Knowledge check
- Lab introduction

Capstone architecture



Capstone architecture check-in



Knowledge check

Knowledge check question 1

Which of the following are true of AMIs? (Select TWO.)

A	AMIs can specify the subnets for launch.
B	AMIs can include block device mapping that specifies the volumes to attach to the Amazon EC2 instance when it is launched.
C	AMIs can only be obtained from the AWS Marketplace.
D	You can launch multiple instances from a single AMI.
E	AMIs can only be used by users within a single account.

Knowledge check question 1 and answer

Which of the following are true of AMIs? (Select TWO.)

A	AMIs can specify the subnets for launch.
B correct	AMIs can include block device mapping that specifies the volumes to attach to the Amazon EC2 instance when it is launched.
C	AMIs can only be obtained from the AWS Marketplace.
D correct	You can launch multiple instances from a single AMI.
E	AMIs can only be used by users within a single account.

Knowledge check question 2

In the instance type name m6g.2xlarge, which aspect of the name indicates the instance family and helps to determine its best use case?

A	m
B	g
C	2xlarge
D	6

Knowledge check question 2 and answer

In the instance type name m6g.2xlarge, which aspect of the name indicates the instance family and helps to determine its best use case?

A correct	m
B	g
C	2xlarge
D	6

Knowledge check question 3

Which of the following are true statements regarding Lambda? (Select TWO.)

A	Functions currently only support Python.
B	You are responsible for updating and patching Lambda servers.
C	Functions can be allocated up to 10 GB of memory.
D	Functions can run for a maximum of 15 minutes.
E	Functions require a security group.

Knowledge check question 3 and answer

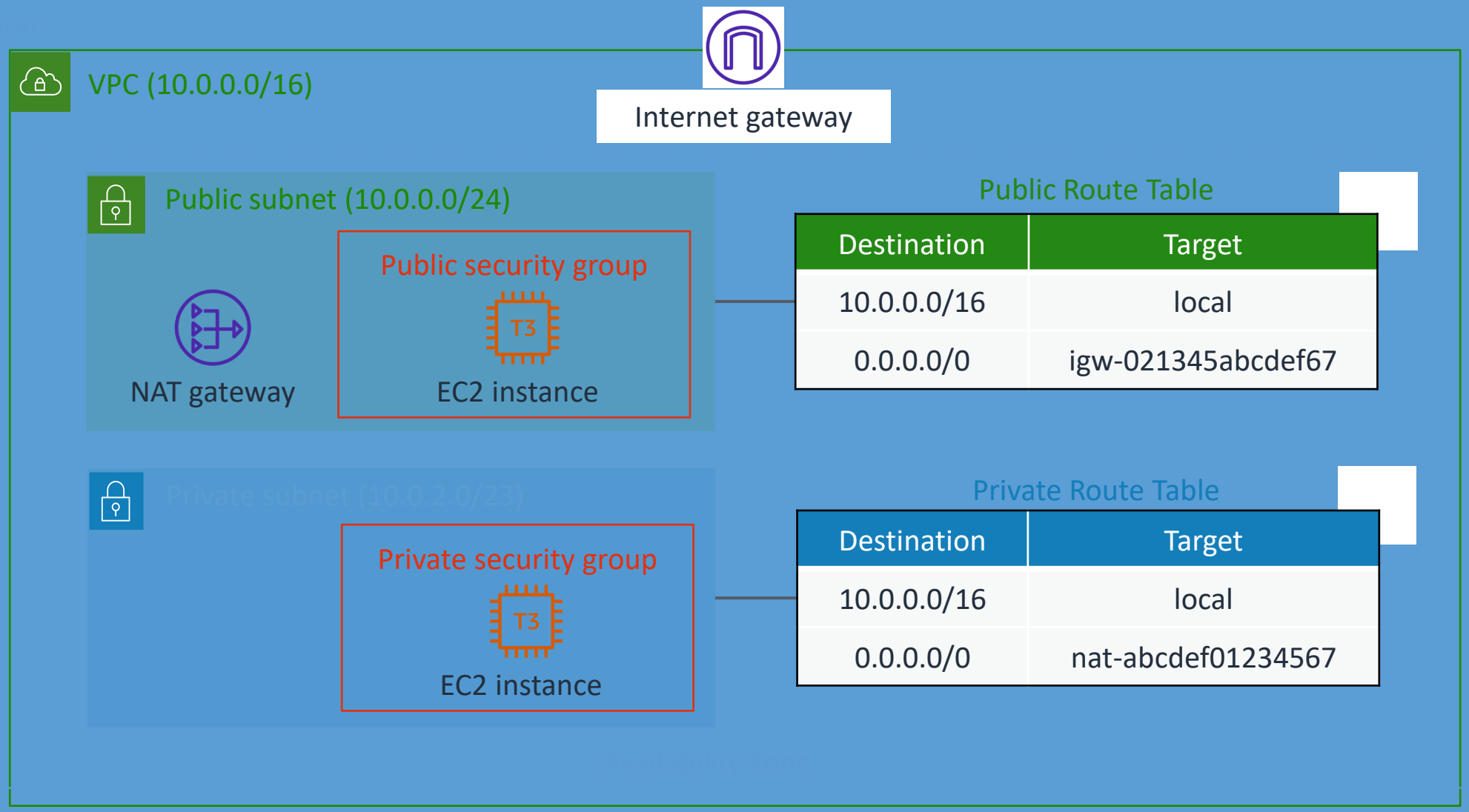
Which of the following are true statements regarding Lambda? (Select TWO.)

A	Functions currently only support Python.
B	You are responsible for updating and patching Lambda servers.
C correct	Functions can be allocated up to 10 GB of memory.
D correct	Functions can run for a maximum of 15 minutes.
E	Functions require a security group.

Lab 2:

Build your Amazon VPC infrastructure

Lab 2 diagram



Public Route Table

Destination	Target
10.0.0.0/16	local
0.0.0.0/0	igw-021345abcdef67

Private Route Table

Destination	Target
10.0.0.0/16	local
0.0.0.0/0	nat-abcdef01234567

Lab tasks

Task 1:	Create an Amazon VPC in a Region.
Task 2:	Create public and private subnets.
Task 3:	Create an internet gateway.
Task 4:	Route internet traffic in the public subnet to the internet gateway.
Task 5:	Create a public security group.
Task 6:	Launch an Amazon EC2 instance into a public subnet.
Task 7:	Connect to a public instance via HTTP.
Task 8:	Connect to the Amazon EC2 instance in the public subnet.
Task 9:	Create a NAT gateway and configure routing in the private subnet.
Task 10:	Create a security group for private resources.
Task 11:	Launch an Amazon EC2 instance into a private subnet.
Task 12:	Connect to the Amazon EC2 instance in the private subnet.
Optional Task 1:	Test connectivity to the private instance from the public instance.
Optional Task 2:	Retrieve instance metadata.

End of Module 4

AWS

Storage

Question

What type of storage are you or your teams using in your environments? Choose all that apply.

A. Block storage

B. File storage

C. Object storage

D. I don’t know

Module overview

- Business requests
- Storage services
- Amazon Simple Storage Service (Amazon S3)
- Shared file systems
- Data migration tools
- Present solutions
- Capstone check-in
- Knowledge check

Business Requirements

Storage Team Lead

The storage team lead needs to know:

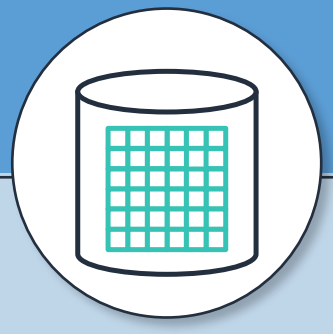
- What are some services to consider when looking at block, file and object storage?
- How do we choose the right object storage solution for my use case?
- What are some file-based options for building secure and scalable storage in the AWS Cloud?
- How can we move lots of data to the cloud in a relatively short time period?

Storage services

“What are some services to consider when looking at block, file and object storage?”

Cloud storage overview

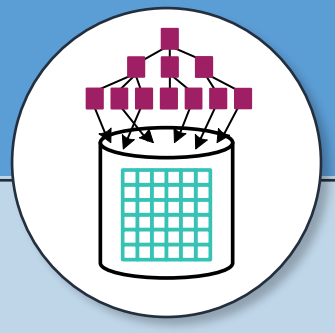
Block storage



Raw storage. Data organized as an array of unrelated blocks.

Examples: Hard disk, Storage Area Network (SAN), storage arrays

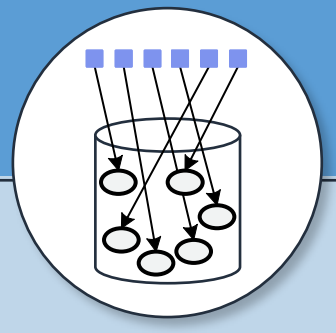
File storage



Unrelated data blocks managed by a file (serving) system. Native file system places data on disk.

Examples: Network Attached Storage (NAS) appliances, Windows file servers

Object storage



Stores Virtual containers that encapsulate the data, data attributes, metadata and Object IDs.

Examples: Ceph, OpenStack Swift

AWS data building blocks

Block Storage

Amazon Elastic
Block Store
(Amazon EBS)

File Storage

Amazon Elastic
File Service
(Amazon EFS)

Amazon
FSx

Object Storage

Amazon Simple
Storage Service
(Amazon S3)

Amazon S3
Glacier

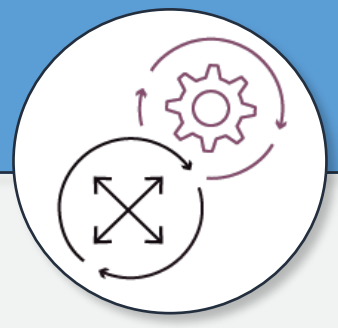
Amazon S3

“How do we choose the right object storage solution for my use case?”

Amazon S3

Amazon Simple Storage Service (Amazon S3) is a durable object storage solution.

Accelerate innovation



Increase agility

Reduce cost

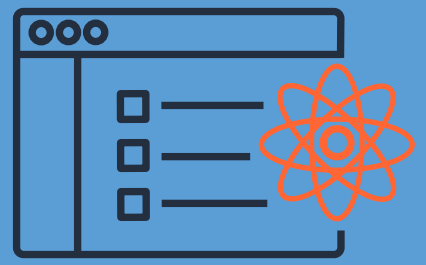
Strengthen security

Amazon S3 use cases

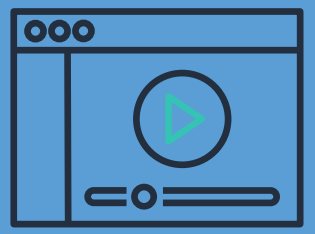
Use Amazon S3 when you have:

- Large number of users accessing your content
- Growing data sets
- Data you will write once and read many times

Backup and restore

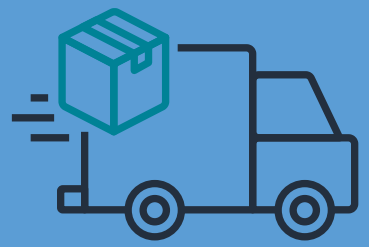


Data lakes for analytics



Media storage and streaming

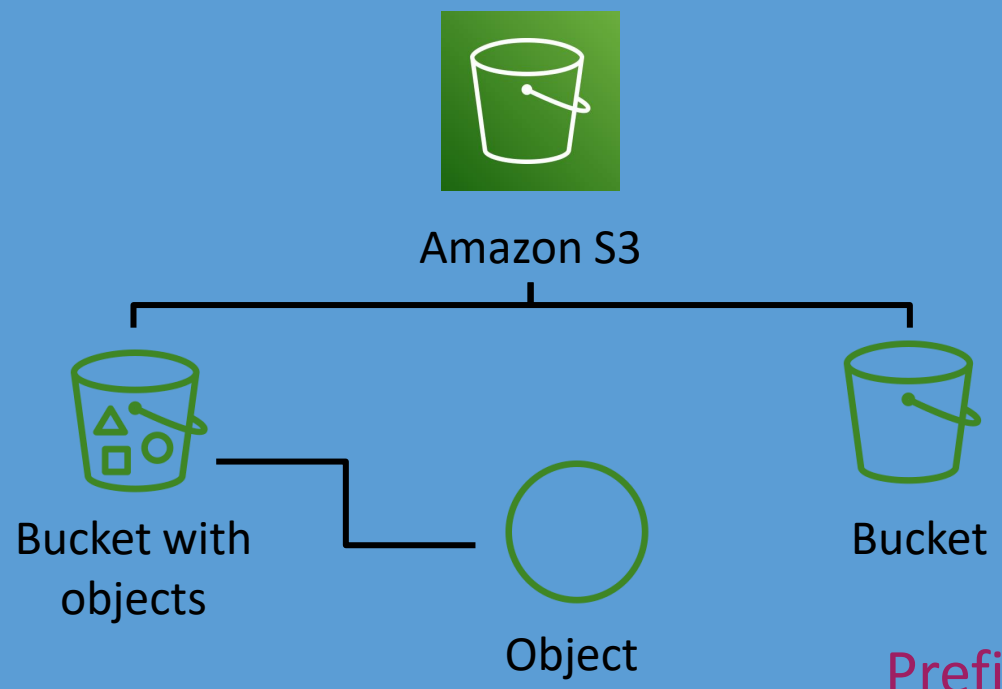
Static website



Archiving and compliance

Buckets and objects

- Amazon S3 stores data as objects within buckets.
- An object includes a file and any metadata that describes the file.
- You can control access to the bucket and its objects.



Prefix

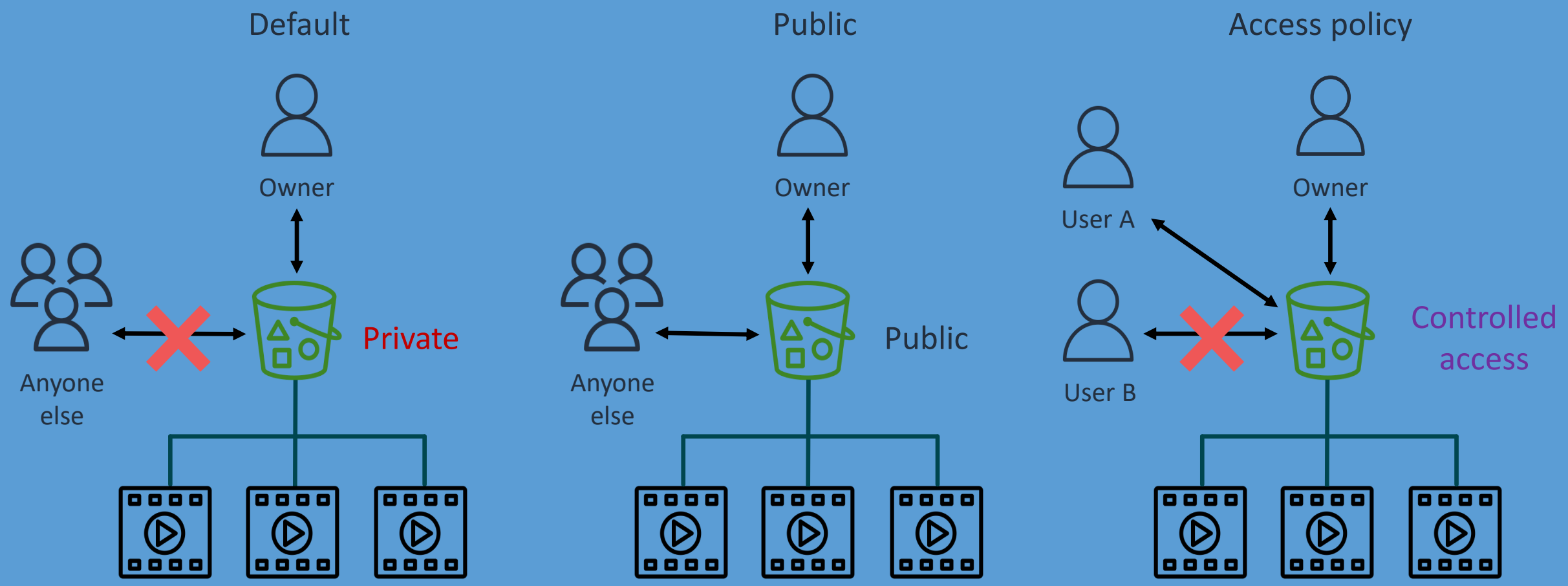
https://my-bucket.s3.amazonaws.com/2006-03-01/pup.jpg

Bucket

Object key

Securing objects

Amazon S3 access control



Amazon S3 Access Control Lists (ACLs)

- Amazon S3 access control lists (ACLs) help you manage access to buckets and objects.
- Each bucket and object has an ACL attached to it.
- The ACL names which AWS accounts or groups are granted access and the type of access.
- Only use ACLs in unusual circumstances where you need to control access for each object individually.

Bucket policies are a preferred method for controlling access to your buckets and objects.

Bucket policies

- Resource-based policy for an S3 bucket
- Controls access to a bucket without managing permissions in AWS Identity and Access Management (IAM)

```
{
 "Version": "2012-10-17",
 "Statement": [
  {
   "Effect": "Allow",
   "Principal": "*",
   "Action": [
     "s3:ListBucket",
      "s3:GetObject",
    ]
    "Resource": [
      "arn:aws:s3:::doc-example-bucket",
      "arn:aws:s3:::doc-example-bucket/*",
   ]
  }
 ]
}
```

Bucket policy in JSON format

Amazon S3 Block Public Access

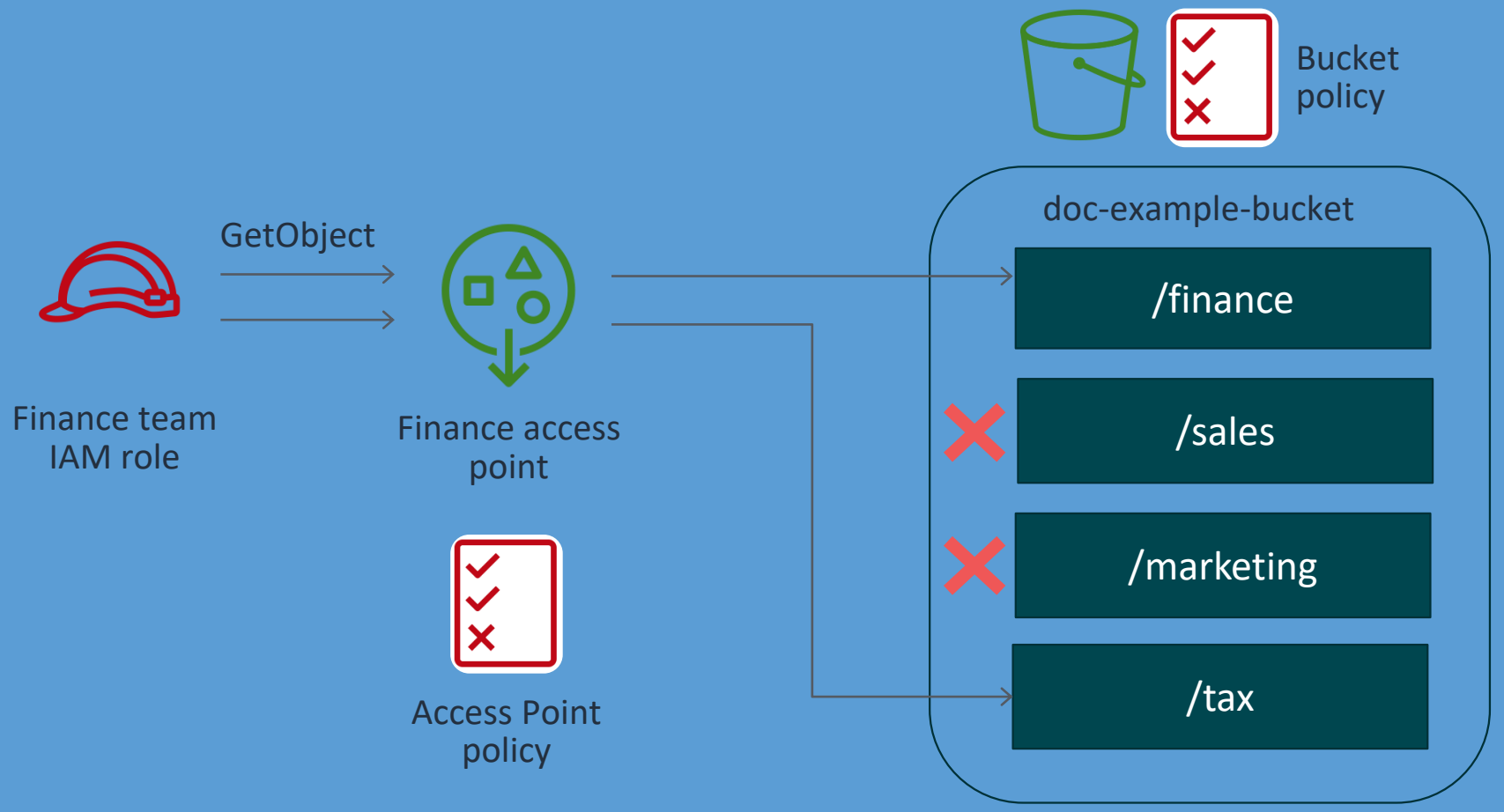
Block public …

- ❑ access to buckets and objects granted through *new* ACLs
- ❑ access to buckets and objects granted through *any* ACLs
- ❑ access to buckets and objects granted through a *new* public bucket or access point policies
- ❑ *cross-account access* to buckets and objects through *any* public bucket or access point policies

Amazon S3 Access Points

Each access point has:

- a unique DNS name and Amazon Resource Name (ARN)
- distinct permissions and network controls



ARN: arn:aws:s3:us-west-2:123456789012:accesspoint/finance

Server-side encryption key types

Choose how you encrypt objects in your S3 buckets:

Amazon S3-Managed Keys (SSE-S3)

Amazon S3 manages a primary key used to create encryption keys for each object.

AWS KMS keys (SSE-KMS)

You use AWS Key Management Service (AWS KMS) to manage your encryption keys.

Customer-Provided Keys (SSE-C)

You manage the keys and Amazon S3 manages the encryption.

Storing objects

Amazon S3 storage classes

Higher cost, frequent access ⟷ **Lower cost, infrequent access**

S3 Standard	S3 Standard-IA	S3 One Zone-IA	S3 Glacier Instant Retrieval	S3 Glacier Flexible Retrieval	S3 Glacier Deep Archive
• Active, frequently accessed data • Milliseconds to access	• Infrequently accessed objects • Milliseconds to access	• Re-creatable, less accessed data • Milliseconds to access	• Archived data that needs fast restore times • Milliseconds to restore	• Objects with unpredictable restore needs • Minutes to hours to restore	• Archive data not likely to be restored • 12 hours or less to restore

S3 Intelligent-Tiering – Data with unknown or changing access patterns. Milliseconds to access.

Amazon S3 Intelligent-Tiering

S3 Intelligent-Tiering

Delivers automatic storage cost savings in three low latency and high throughput access tiers when access patterns change.

Frequent Access tier

Objects first assigned to S3 Intelligent-Tiering

Infrequent Access tier

Objects not accessed for 30 consecutive days

Archive Instant Access tier

Objects not accessed for 90 consecutive days

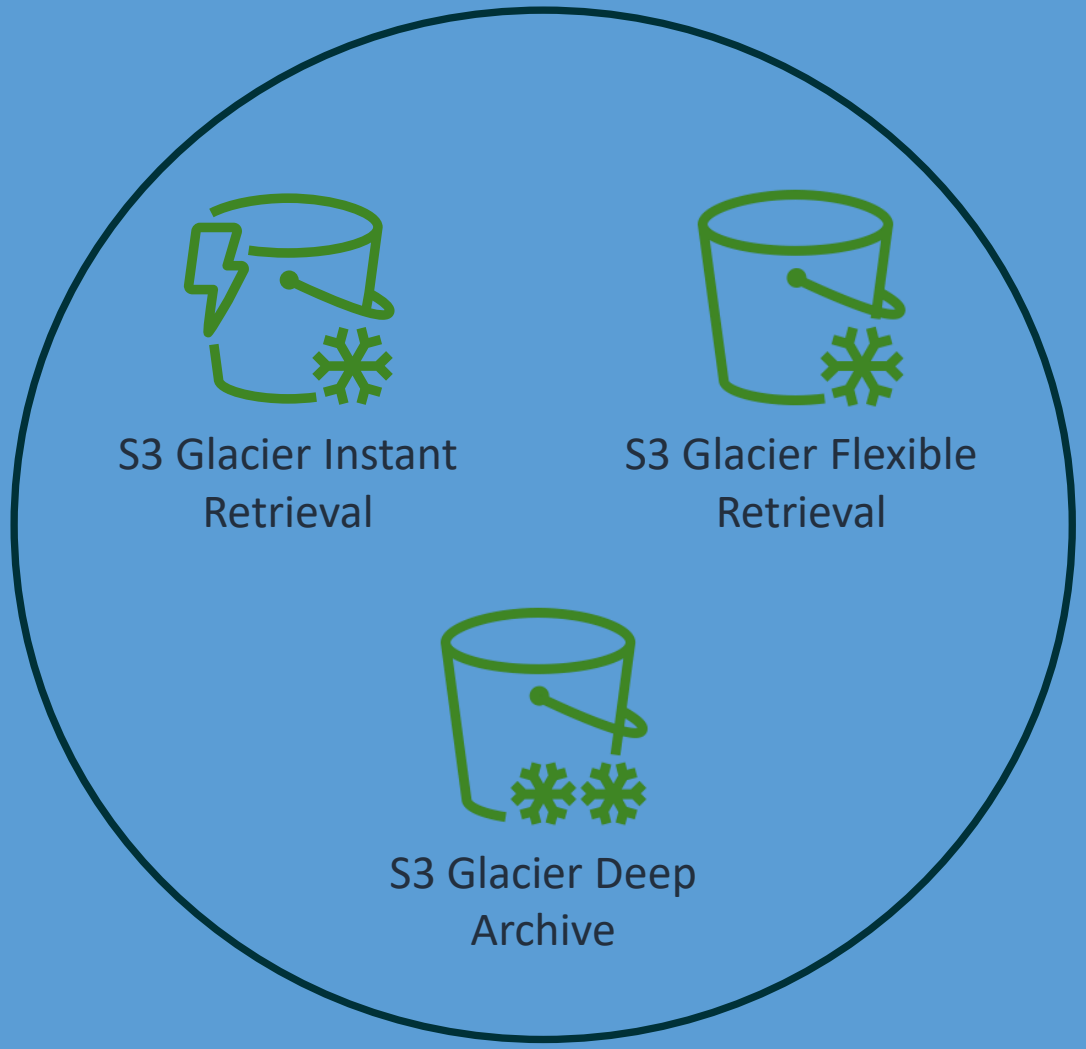
Amazon S3 Glacier storage class benefits

1. **Cost-Effective storage**
   Lowest cost for specific data access patterns.

2. **Flexible data retrieval**
   Three storage classes with variable access options.

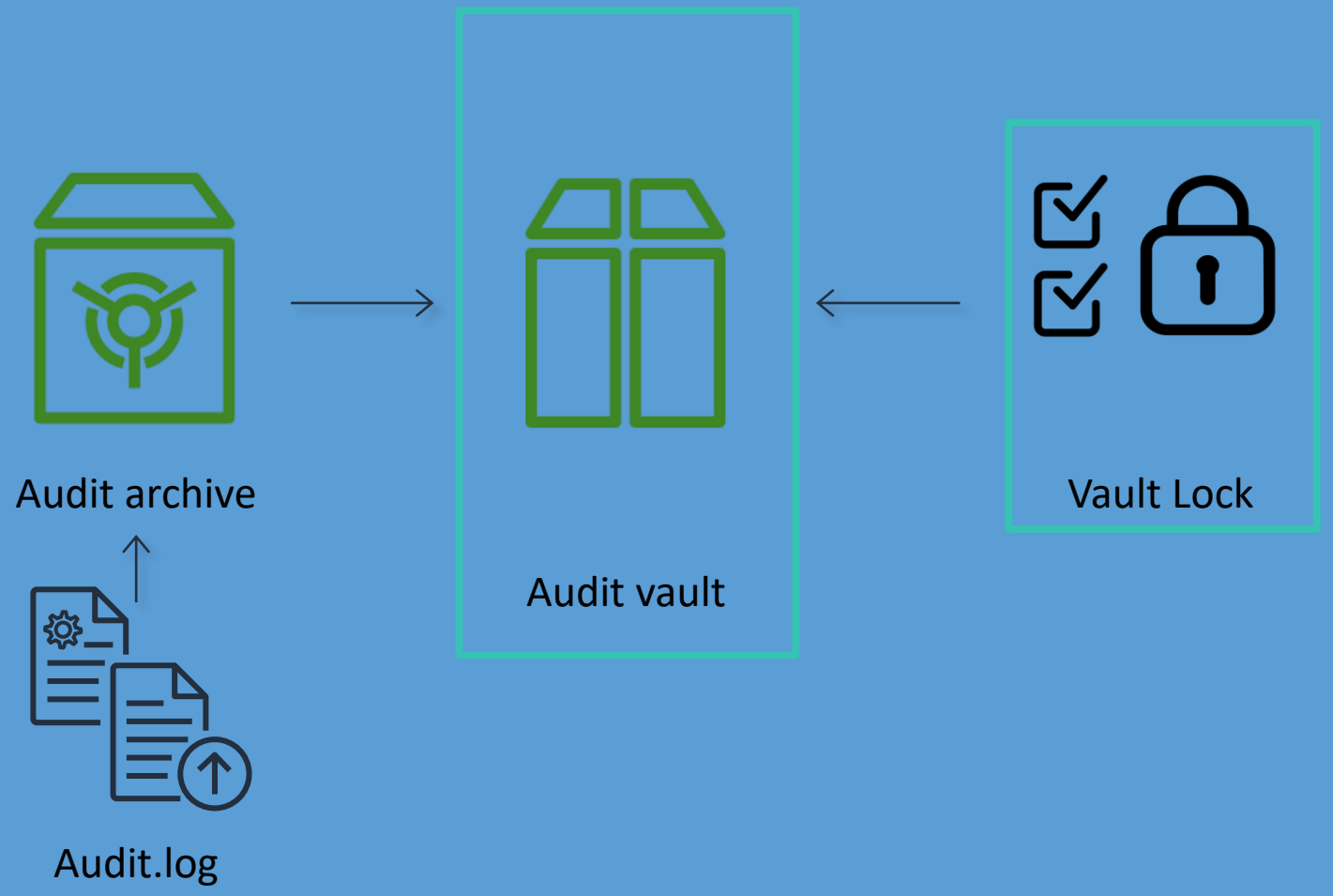
3. **Secure and compliant**
   Encryption at rest, AWS CloudTrail integration, and retrieval policies.

4. **Scalable and durable**
   Meets needs from gigabytes to exabytes with 11 9s of durability



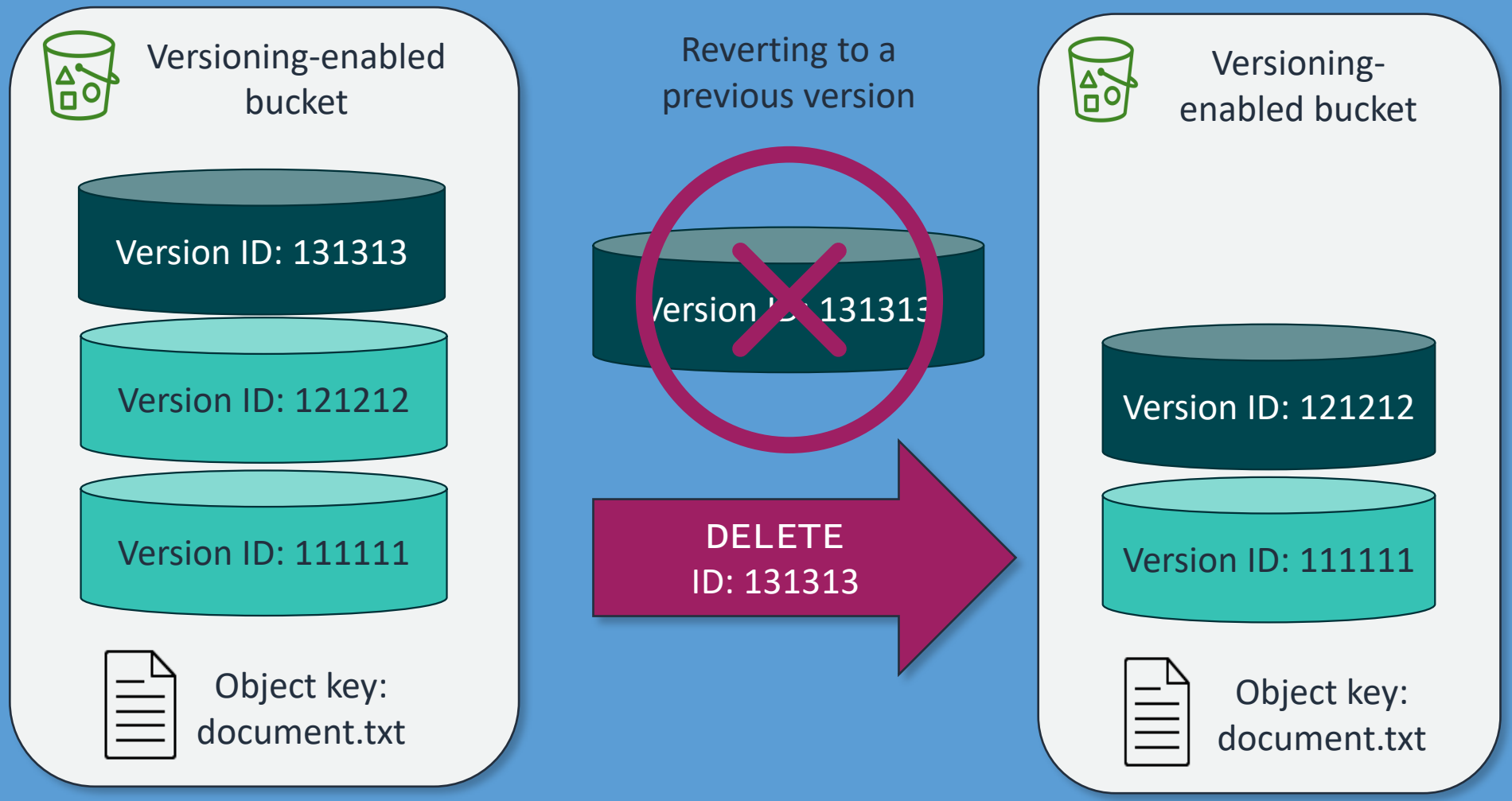
Amazon S3 Glacier archives and vaults

- Group archives together in a vault of your choice.
- Manage vaults using the AWS CLI (using the REST API) or an AWS SDK.
- Manage and protect your vaults with features like Vault Lock.



Amazon S3 Versioning

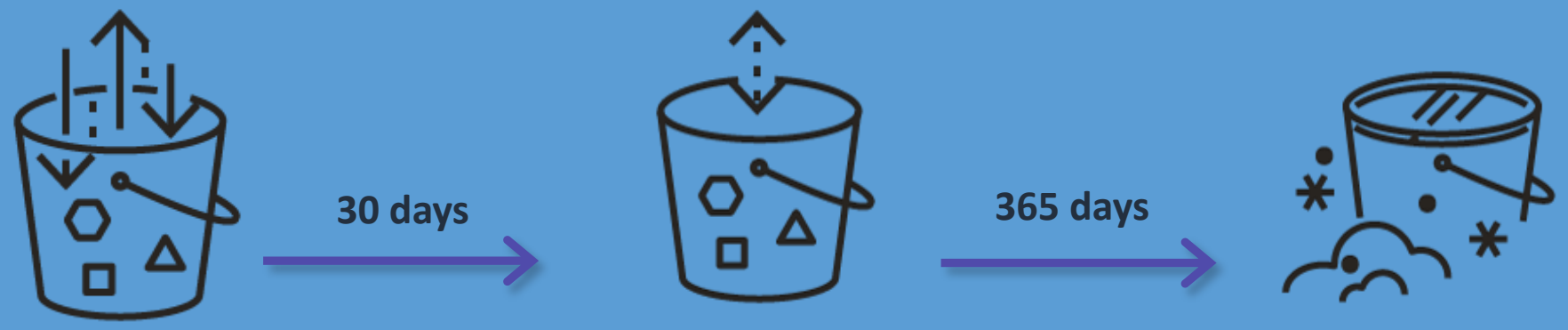
- Keep multiple variants of an object in the same bucket.
- Restore an object to a previous or specific version.
- Use S3 Object Lock for data retention or protection.



Lifecycle policies

Use S3 Lifecycle polices to transition objects to another storage class. S3 Lifecycle rules take action based on object age. Here's an example:

1. Move objects older than **30 days** to S3 Standard-IA.
2. Move objects older than **365 days** to Amazon S3 Glacier Deep Archive.

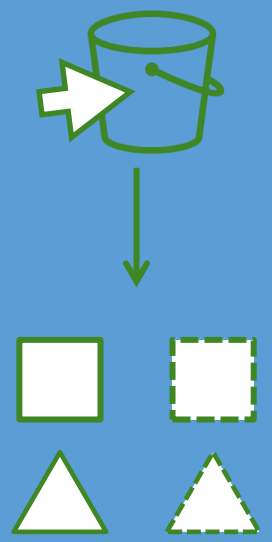


Replicating S3 objects

Amazon S3 Replication

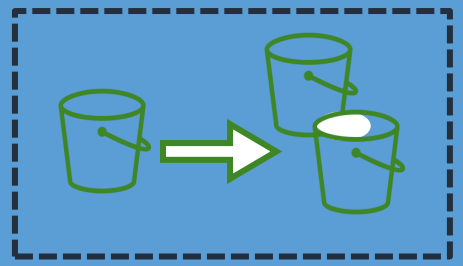
Automatic, asynchronous copying of objects across S3 buckets

Select source bucket and data set.

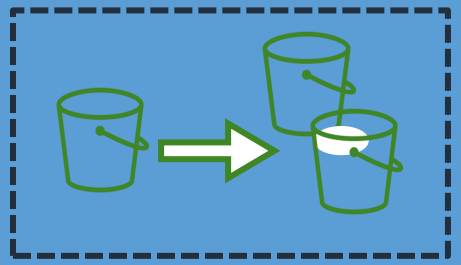


Select one or more destination buckets.

S3 Same-Region Replication (SRR)



or

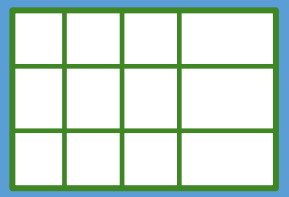


S3 Cross-Region Replication (CRR)

Option: Owner override

Change the destination account or ownership.

Option to change the destination storage class for cost optimization

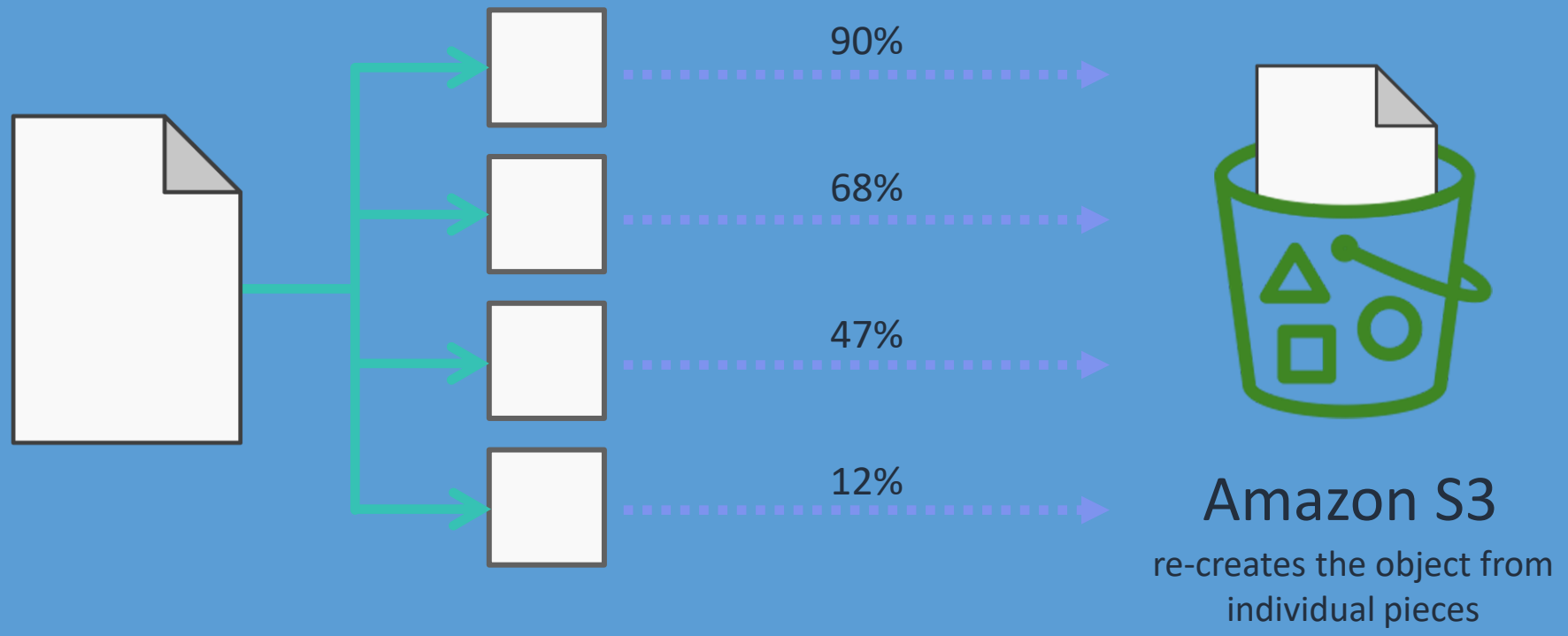


Additional Amazon S3 features

Amazon S3 multipart upload

- Initiate the upload.
- Upload the object parts.
- Complete the multipart upload.

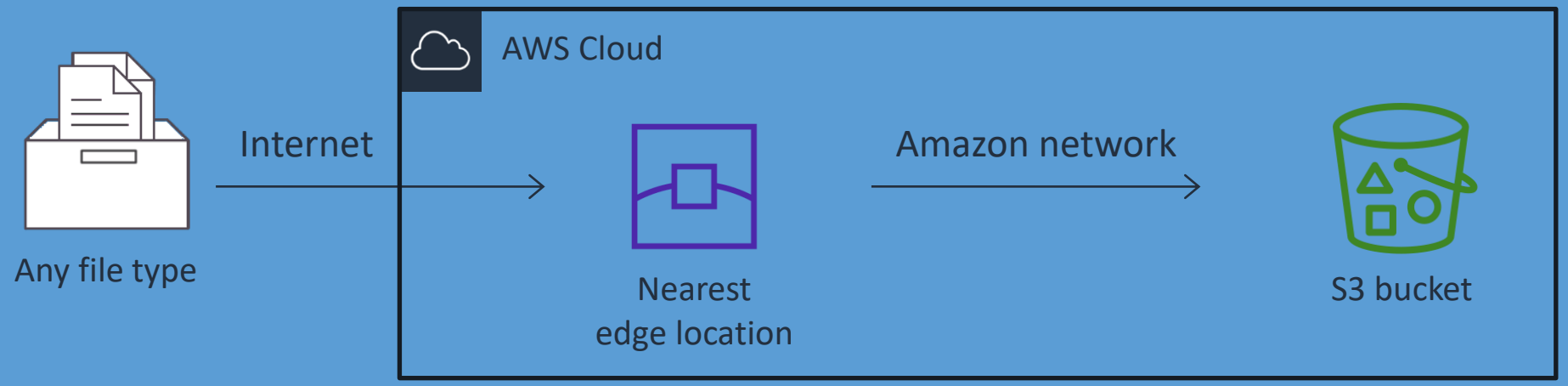
Note: You cannot perform multipart uploads manually using the AWS Management Console.



Amazon S3 Transfer Acceleration

- Move data faster over long distances.
- Reduce network variability.

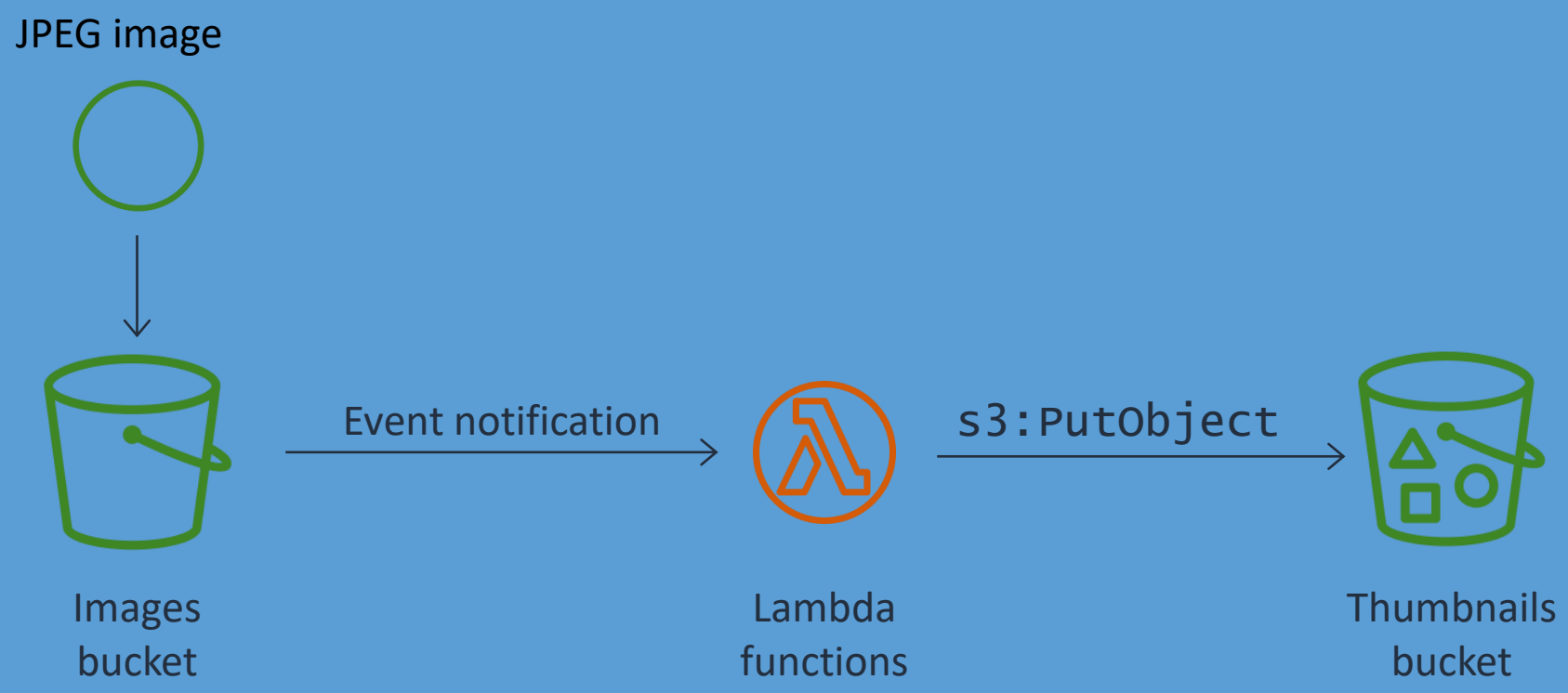
Instead…



Amazon S3 event notifications

- Get notifications sent when events happen in your S3 bucket.
- Let AWS manage event monitoring: no polling needed.

An example event notification workflow to convert images to thumbnails:



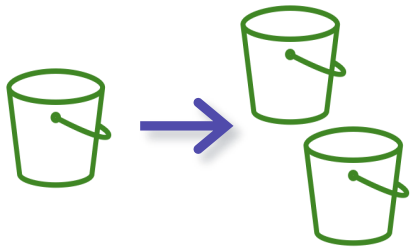
Amazon S3 cost factors

Storage type

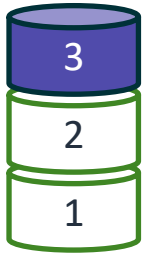
Requests and retrievals

Data transfer

Management and analytics



Replication



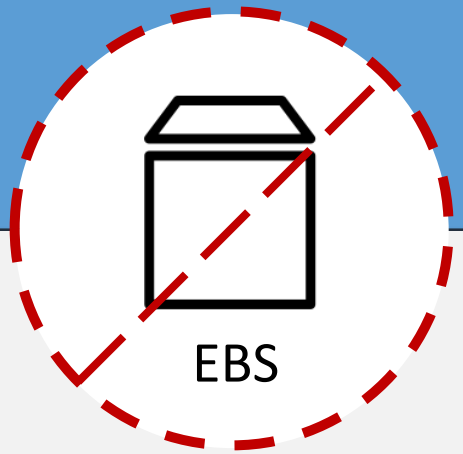
Versioning

Shared file systems

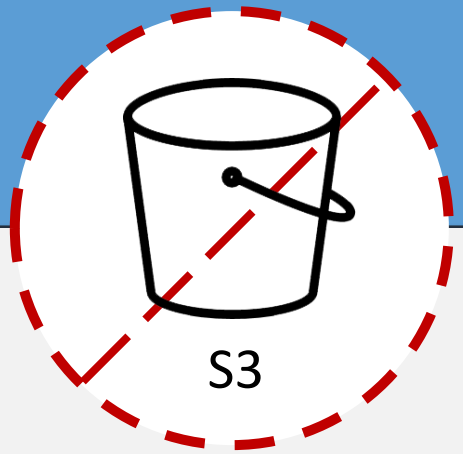
“What are some file-based options for building secure and scalable storage in the AWS Cloud?”

Shared file systems

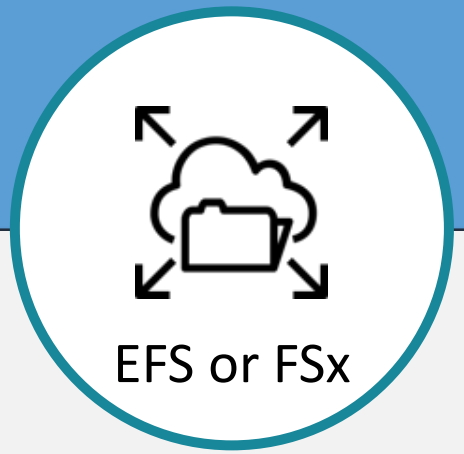
What if I have multiple instances that need to use the same storage?



Amazon EBS is usually attached to one instance



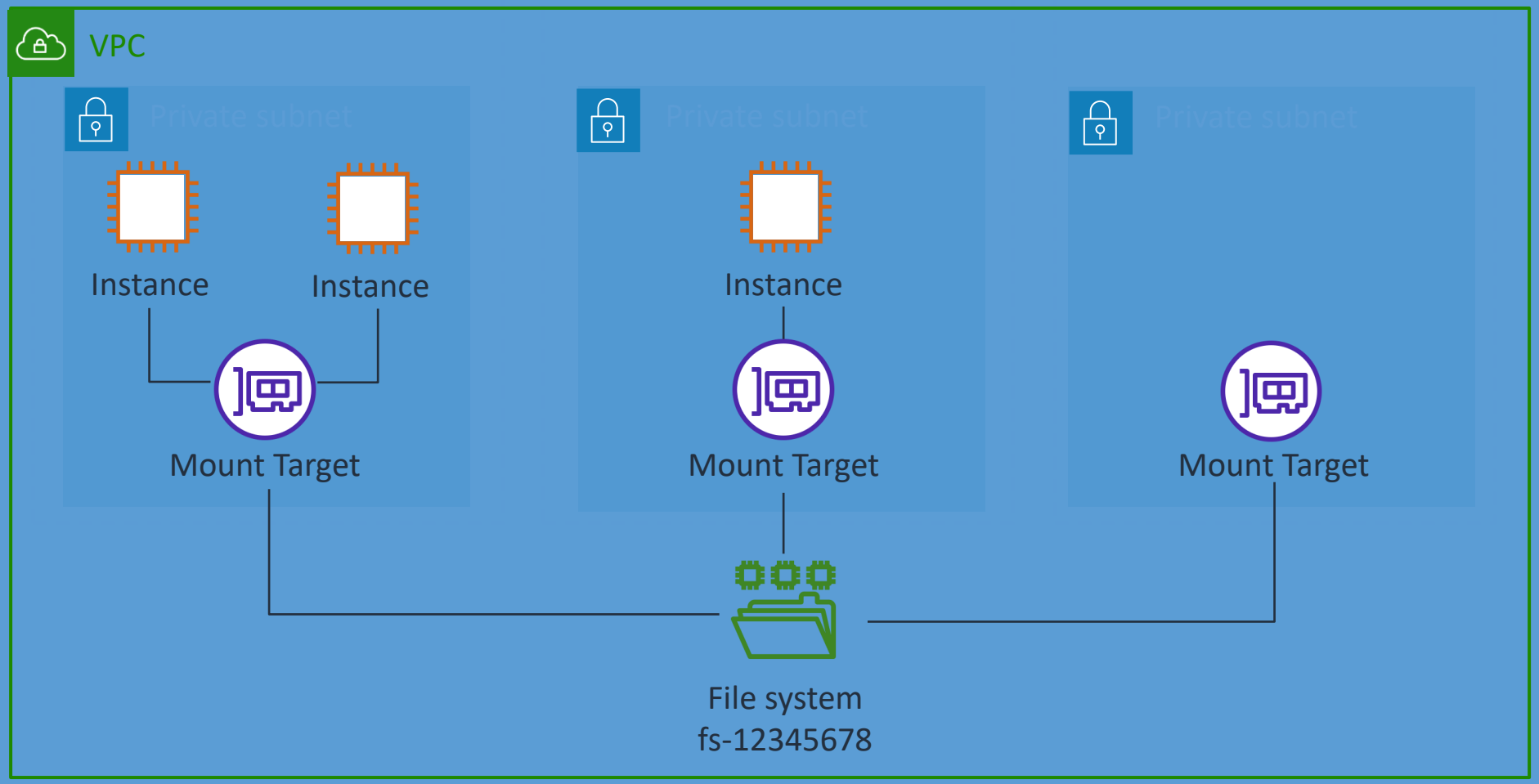
Object storage is not built for file systems



Amazon EFS and Amazon FSx are ideal for this task

Amazon EFS

- Choose Amazon EFS for a scalable and elastic file system.
- Connect using the NFSv4 protocol.
- Access file systems across EC2 instances at the same time.



Amazon EFS benefits

EFS uses burst throughput mode to scale throughput based on your storage use.

Additionally, you can provision throughput independent of storage.

EFS automatically grows and shrinks file storage without provisioning.

Monitoring is not required to avoid storage limits.

EFS managed file systems lower your total cost of ownership (TCO). Pay only for what you use.

Save on cost with EFS Infrequent Access or One Zone storage types.

Amazon FSx

- Launch, run, and scale high-performing file systems on AWS.
- Use familiar and feature-rich products without managing hardware provisioning, patching, and backups.

Amazon FSx for Windows File Server

Amazon FSx for Lustre

Amazon FSx for NETapp ONTAP

Amazon FSx for OpenZFS

Data migration tools

“How can we move lots of data to the cloud in a relatively short time period?”

AWS data migration tools

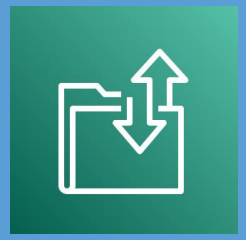
Online

AWS Storage Gateway

Sync files with SMB, NFS, and iSCSI protocols from on-premises to AWS.

AWS DataSync

Sync files from on-premises file storage to an Amazon EFS file system or S3 bucket.



AWS Transfer Family

Transfer files into and out of Amazon S3 with SFTP protocol.

Offline

AWS Snow Family

Move terabytes to petabytes of data to AWS using appliances designed for secure, physical transport.

AWS Storage Gateway

AWS Storage Gateway is a service that gives your applications seamless and secure integration between on-premises environments and AWS storage.

It provides you low-latency access to cloud data with a Storage Gateway appliance.

Storage Gateway types

Amazon S3 File Gateway

Native file access to Amazon S3 for backups, archives, and ingest for data lakes.

Amazon FSx File Gateway

Native access to Amazon FSx for Windows File Server for on-premises group file shares and home directories.

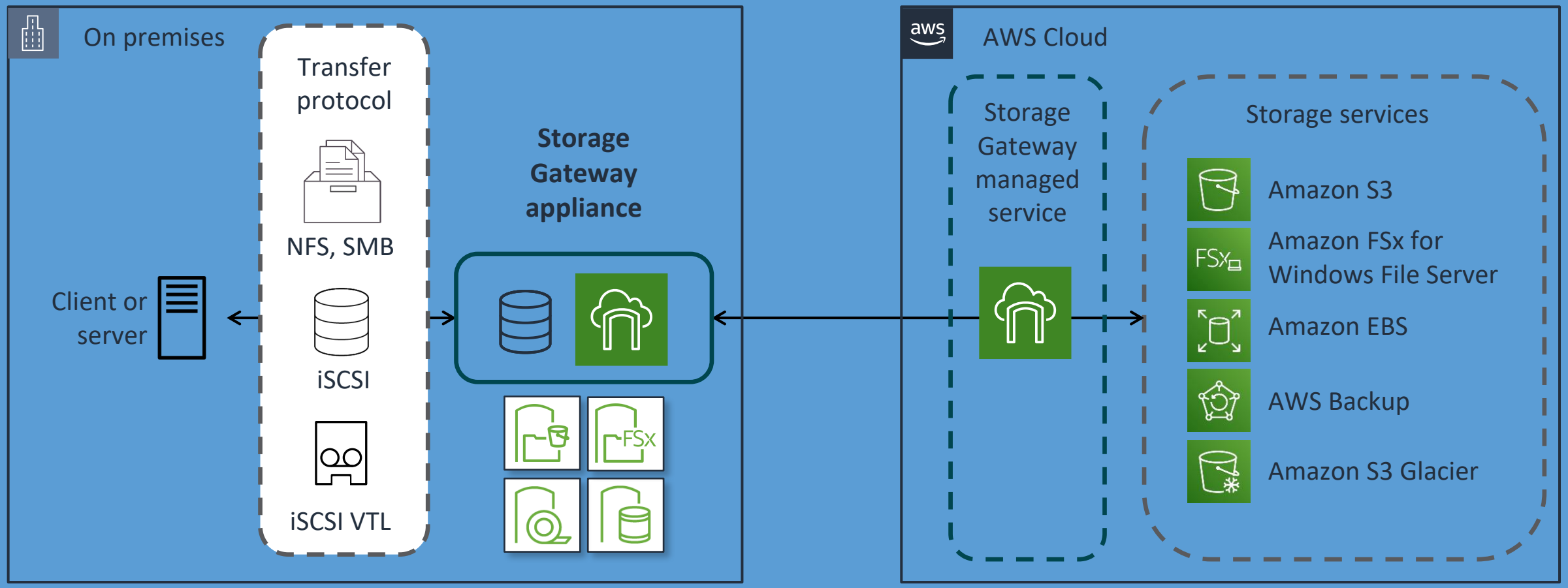
Tape Gateway

Virtual tape library using Amazon S3 archive tiers for long-term retention.

Volume Gateway

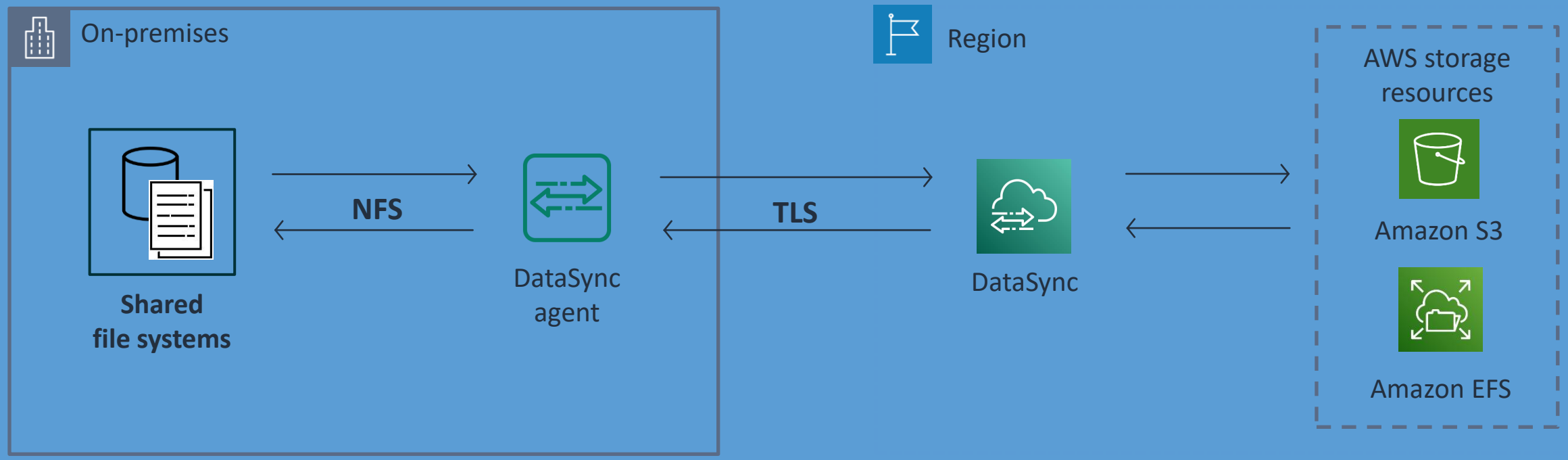
Block-level backups of volumes with Amazon EBS snapshots, AWS Backup integration, and cloud recovery.

Storage Gateway architecture



AWS DataSync

1. Deploy on-premises DataSync agent
2. Transfer data over the WAN using purpose-built protocol
3. Write or read data from AWS storage services
4. Manage from console or CLI



AWS Snow Family service models

AWS Snowcone

Snowcone is a small, rugged, edge computing and data storage product.

AWS Snowball Edge

Snowball Edge is an edge computing and data transfer device provided by the AWS Snowball service.

AWS Snowmobile

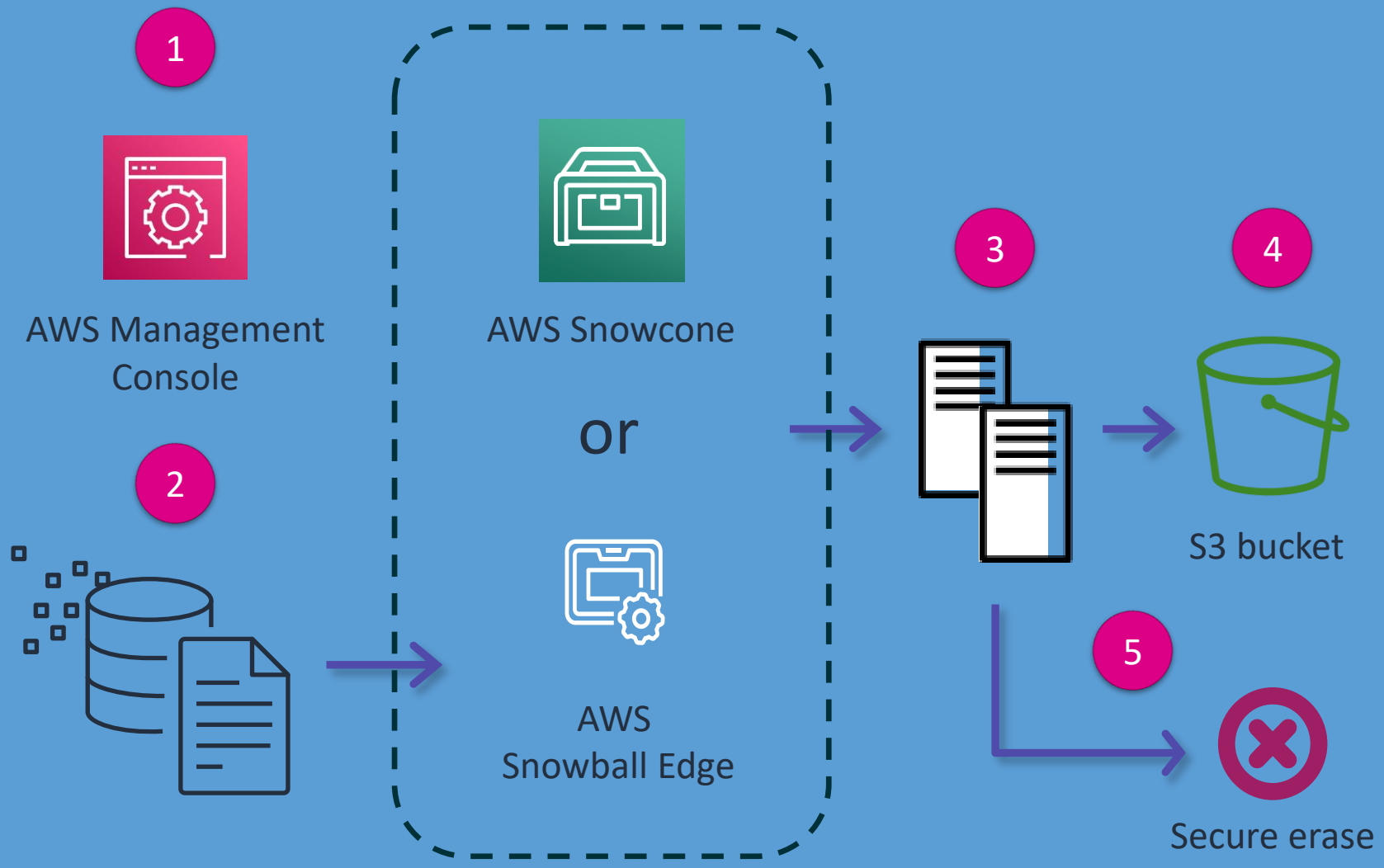
Snowmobile is the first exabyte-scale data migration service to move very large datasets from on premises to AWS.

AWS Snow Family comparison table

	Snowcone	Snowball Edge Storage Optimized	Snowball Edge Compute Optimized	Snowmobile
Migration size	Up to 24 TB, online and offline	Up to petabytes, offline		Up to exabytes, offline
Form factor	Rugged 8.5 G impact cases that are rain and dust resistant, E Ink label for shipping automation			45-foot container, scheduled delivery
Security	256-bit encryption, tamper detection			Encryption, security staff, GPS tracking, video surveillance, alarms
Usable storage	8 TB HDD 14 TB SSD	80 TB HDD 1 TB SSD	42 TB HDD 7.68 TB SSD	100 PB HDD No SSD option
DataSync agent	Pre-installed	-	-	-
Compute	4 vCPU, 4 GB RAM	40 vCPU, 80 GB RAM	52 vCPU, 208 GB RAM	-
Onboard computing options	AWS IoT Greengrass functions Amazon EC2 AMIs			
Wireless	Wi-Fi	-	-	-
Portable or mobile use	Battery-based operation	-	-	-
Clustering	-	5 to 10 nodes		-

Snowcone and Snowball Edge process

1. Create job.
2. Collect and process data.
3. Ship device to AWS.
4. Move data to Amazon S3.
5. Secure device erasure.



Review

Present solutions

Storage Team Lead

Consider how you would answer the following:

- What are some services to consider when looking at block, file and object storage?
- How do we choose the right object storage solution for my use case?
- What are some file-based options for building secure and scalable storage in the AWS Cloud?
- How can we move lots of data to the cloud in a relatively short time period?

Module review

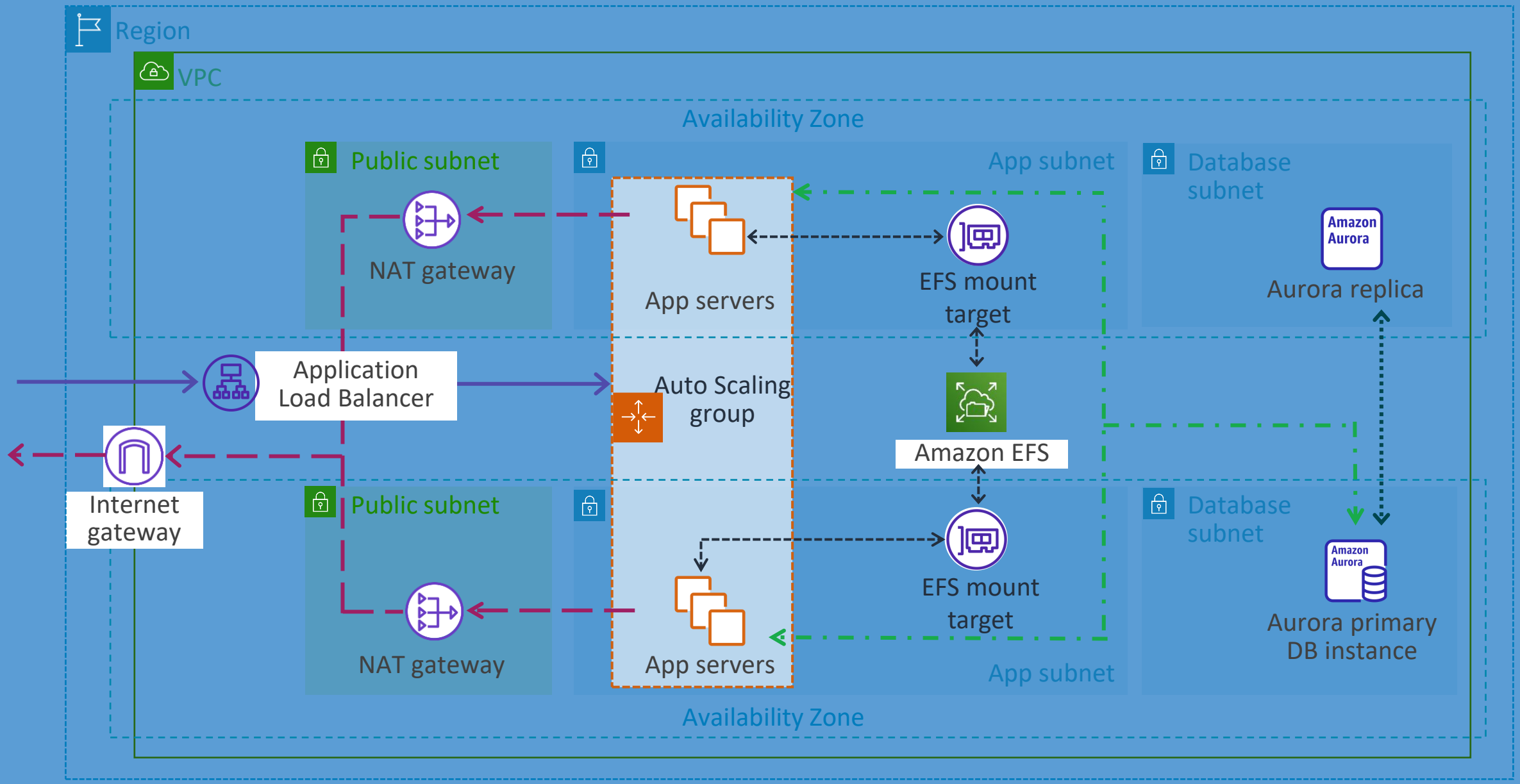
In this module you learned about:

- ✓ Storage services
- ✓ Amazon S3
- ✓ Shared file systems
- ✓ Data migration tools

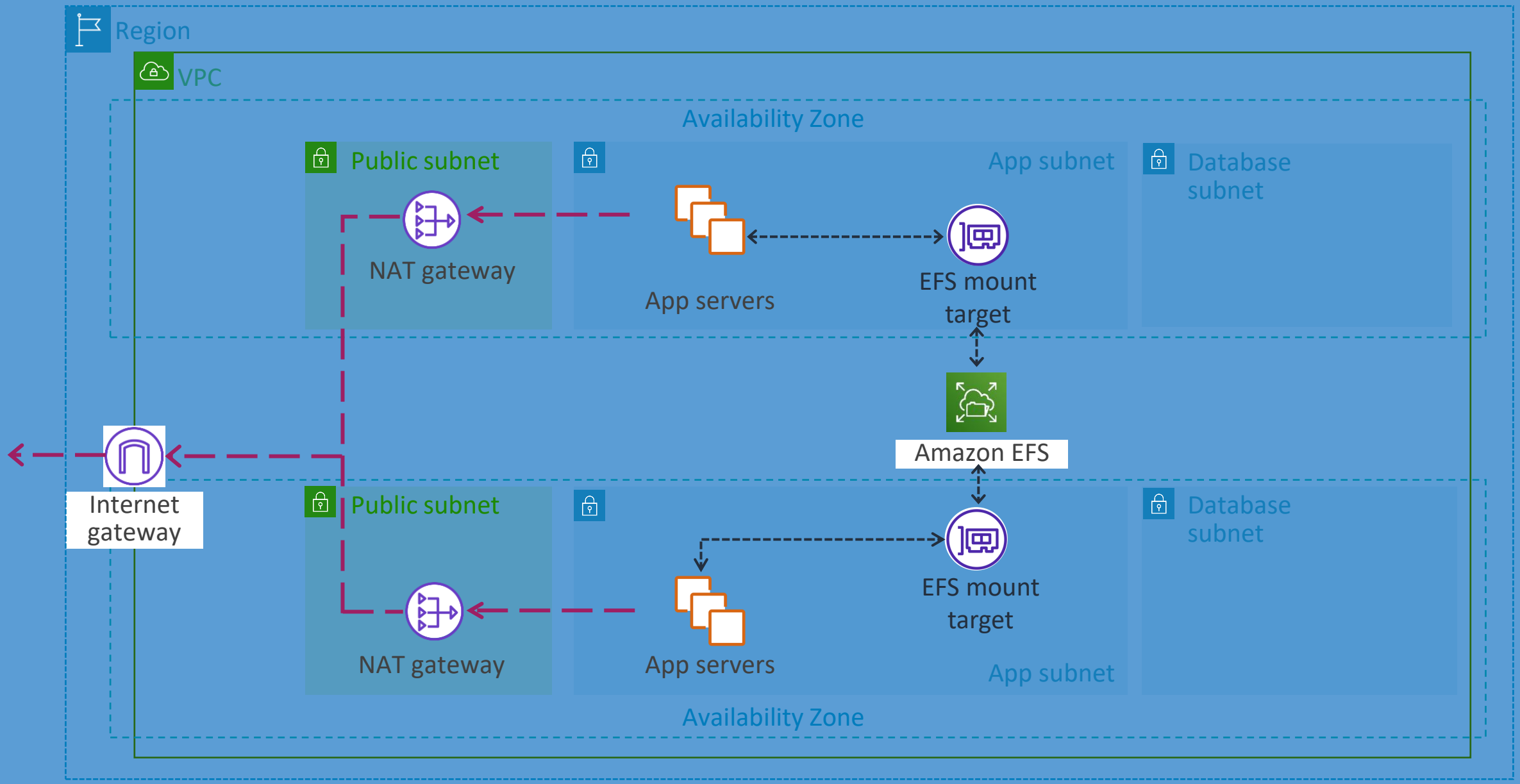
Next, you will review:

- Capstone check-in
- Knowledge Check

Capstone architecture



Capstone architecture check-in



Knowledge check

Knowledge check question 1

Which of the following Amazon S3 features would you use to automatically copy new objects to a bucket in a different AWS Region?

A	Same-Region Replication (SRR)
B	Amazon S3 Versioning
C	AWS DataSync
D	Cross-Region Replication (CRR)

Knowledge check question 1 and answer

Which of the following Amazon S3 features would you use to automatically copy new objects to a bucket in a different AWS Region?

A	Same-Region Replication (SRR)
B	Amazon S3 Versioning
C	AWS DataSync
D correct	Cross-Region Replication (CRR)

Knowledge check question 2

Which Amazon S3 feature can force an action to occur after an event takes place within a bucket?

A	Invoking
B	Event notification
C	Lambda
D	Alarm

Knowledge check question 2 and answer

Which Amazon S3 feature can force an action to occur after an event takes place within a bucket?

A	Invoking
B correct	Event notification
C	Lambda
D	Alarm

Knowledge check question 3

You have two Linux applications in different Availability Zones that must share a common file system. Which of the following is the best solution for this use case?

A	Storage Gateway
B	FSx for Windows File Server
C	Amazon S3
D	Amazon EFS

Knowledge check question 3 and answer

You have two Linux applications in different Availability Zones that must share a common file system. Which of the following is the best solution for this use case?

A	Storage Gateway
B	FSx for Windows File Server
C	Amazon S3
D correct	Amazon EFS

Knowledge check question 4

Which of the following are modes available in the Storage Gateway appliance? (Select THREE.)

A	Memory Gateway
B	Tape Gateway
C	Volume Gateway
D	Amazon EBS File Gateway
E	Amazon S3 File Gateway
F	Amazon S3 Glacier File Gateway

Knowledge check question 4 and answer

Which of the following are modes available in the Storage Gateway appliance? (Select THREE.)

A	Memory Gateway
B correct	Tape Gateway
C correct	Volume Gateway
D	Amazon EBS File Gateway
E correct	Amazon S3 File Gateway
F	Amazon S3 Glacier File Gateway

AWS

Database Services

Lab 3

Question

Which database technologies have you used in your workloads? (Select all that apply.)

A. Relational databases

B. Nonrelational databases

C. Database caching

D. Database migration tools

E. None of these

Module overview

- Business requests
- Database services
- Amazon Relational Database Service (Amazon RDS)
- Amazon DynamoDB
- Database caching
- Database migration tools
- Present solutions
- Capstone check-in
- Knowledge check
- Lab 3: Create a database layer in your Amazon VPC infrastructure

Business Requirements

Database Services Manager

The database services manager wants to know:

- What are the AWS database solutions?
- How can we more efficiently manage our relational databases in the cloud?
- How can we build a scalable key-value NoSQL database?
- How can we cache databases in the cloud to maximize performance?
- What tools are available for migrating an existing database to the AWS Cloud?

Database services

“What are the AWS database solutions?”

AWS database services

Amazon Relational Database Service (Amazon RDS)

Amazon Aurora

Amazon Redshift

Amazon DocumentDB (with MongoDB compatibility)

Amazon DynamoDB

Amazon ElastiCache

Amazon MemoryDB for Redis

Amazon Keyspaces (for Apache Cassandra)

Amazon Timestream

Amazon Neptune

Amazon Quantum Ledger Database (Amazon QLDB)

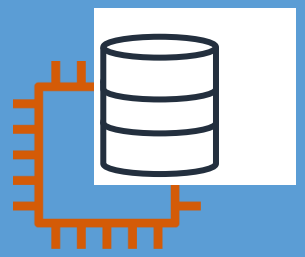
Relational and nonrelational databases

	Relational (SQL) databases	Nonrelational (NoSQL) databases
Data storage	Tables with rows and columns	Key-value, wide-column, graph, document, or other models
Schemas	Fixed	Dynamic
Example database services	Amazon RDS, Aurora	DynamoDB, ElastiCache

Choosing the right database

Relational database	Nonrelational (NoSQL) database
You require strict schema rules and data quality enforcement.	You need your database to scale horizontally.
Your database doesn’t need extreme read/write capacity.	Your data does not lend itself well to traditional schemas.
If you have a relational data set that does not require extreme performance, a relational database management system can be the best, lowest effort solution.	Your read/write rates exceed those that can be economically supported through a traditional SQL database.

Managed and unmanaged services



Database installed on EC2

You

- Application optimization
- Scaling
- High availability
- Database backups
- Database software patches
- Database software installs
- OS patches
- OS installation
- Server maintenance
- Rack and stack
- Power, HVAC, networking

You

aws

aws

Managed database on Amazon RDS

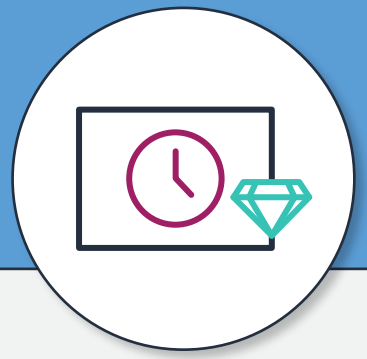
Amazon RDS

“How can we more efficiently manage our relational databases in the cloud?”

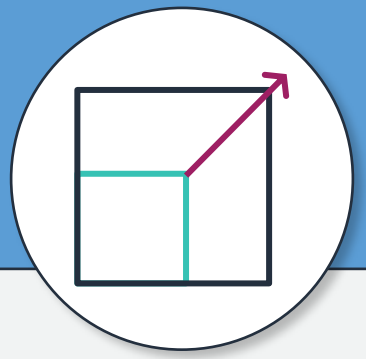
Amazon RDS features

- Hardware, OS, and database software deployment and maintenance
- Built-in monitoring

- Data encryption at rest and in transit
- Industry compliance

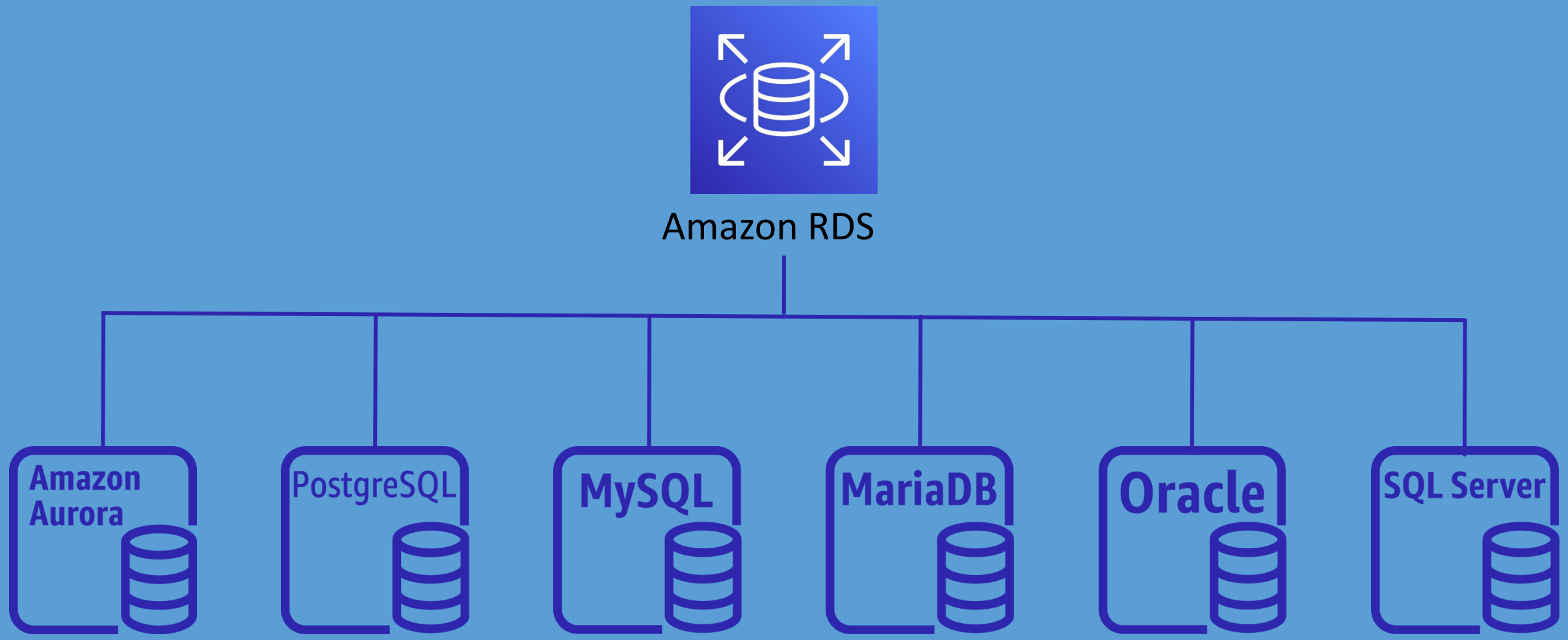


Automatic Multi-AZ data replication



- Compute and storage scaling
- Minimal application downtime

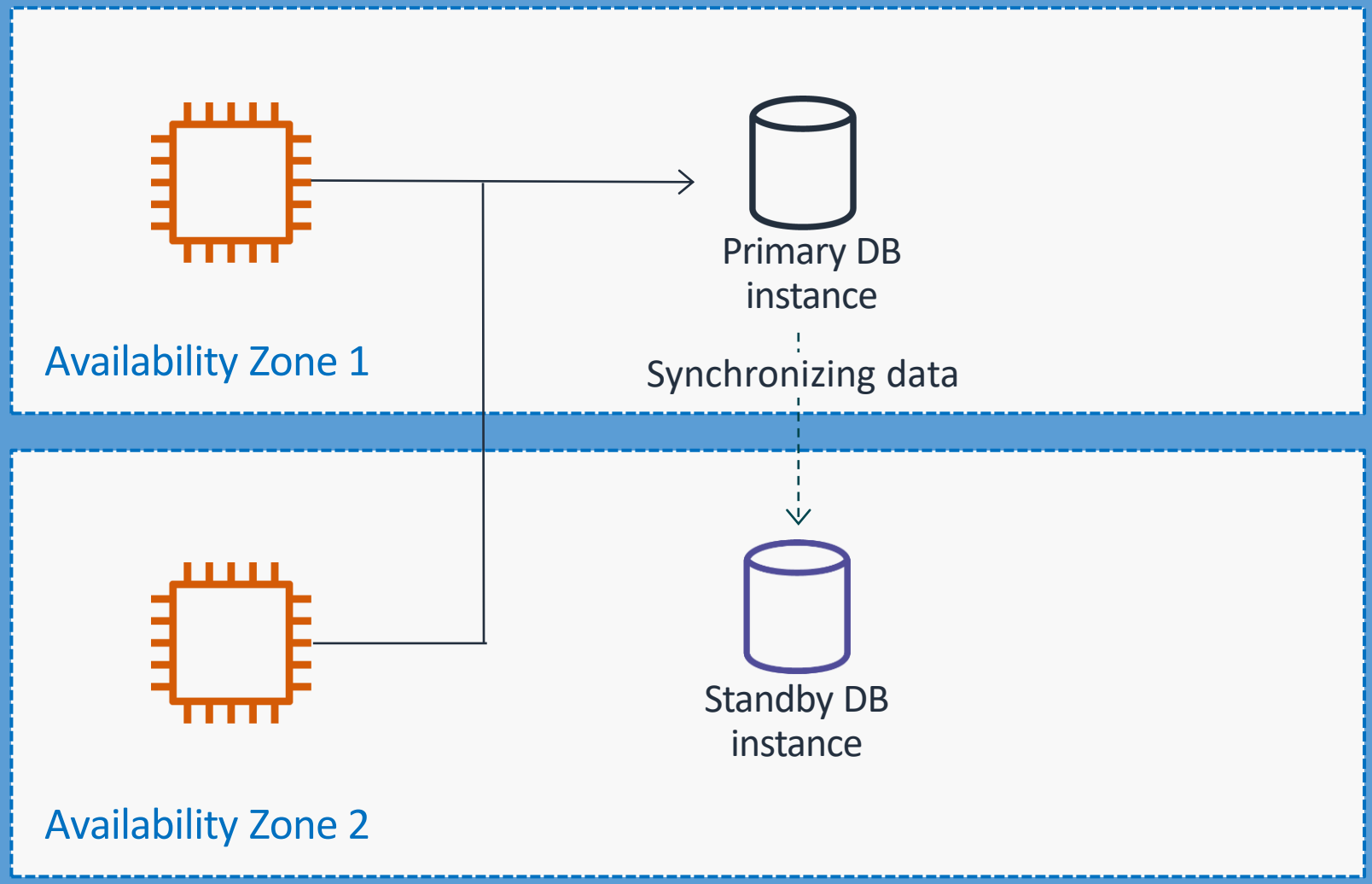
Amazon RDS database engines



Amazon RDS Multi-AZ deployments

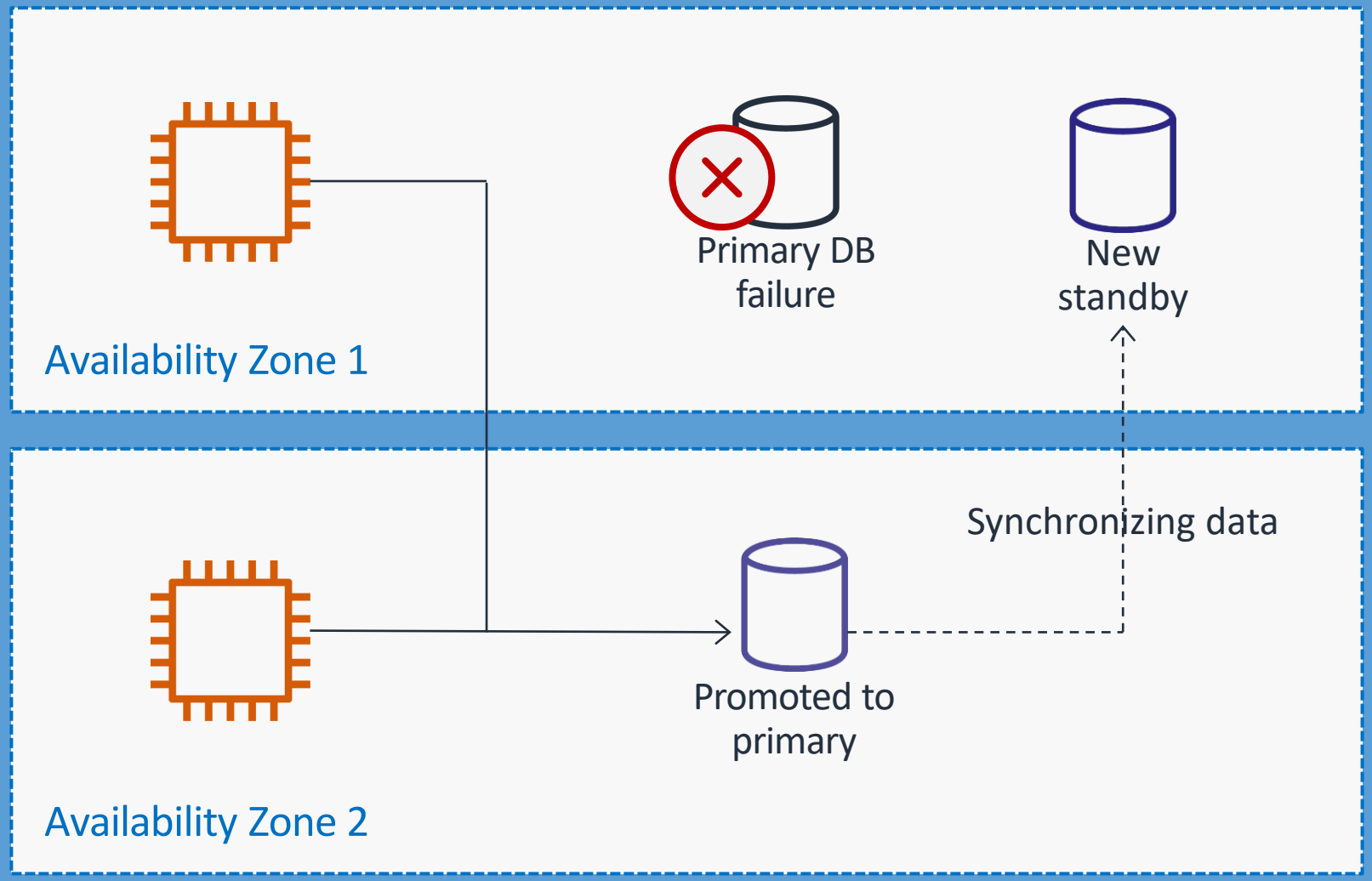
Multi-AZ deployments:

- Replicate data to a standby DB instance in another availability zone
- Not used for read-only scenarios



Amazon RDS Multi-AZ failover

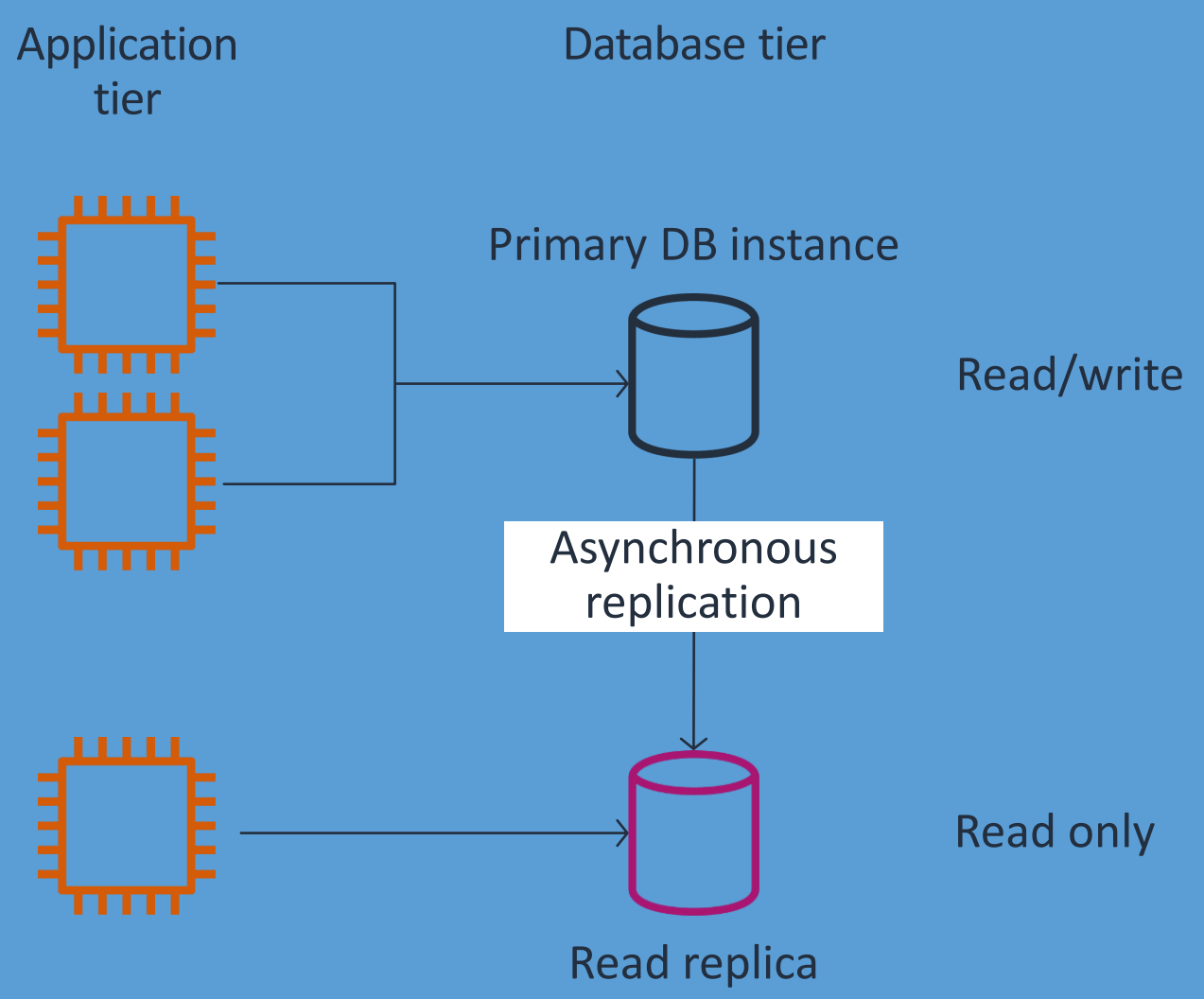
Upon failure, the standby DB instance picks up the load.



Read replicas

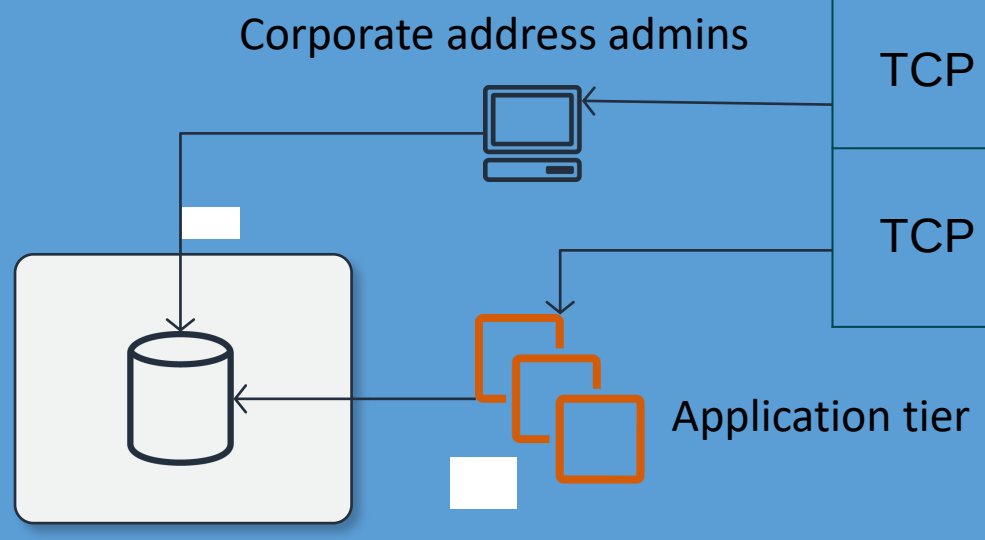
With read replicas, you can:

- Horizontally scale for read-heavy workloads
- Offload reporting
- Replicate across AWS Regions



Secure network access

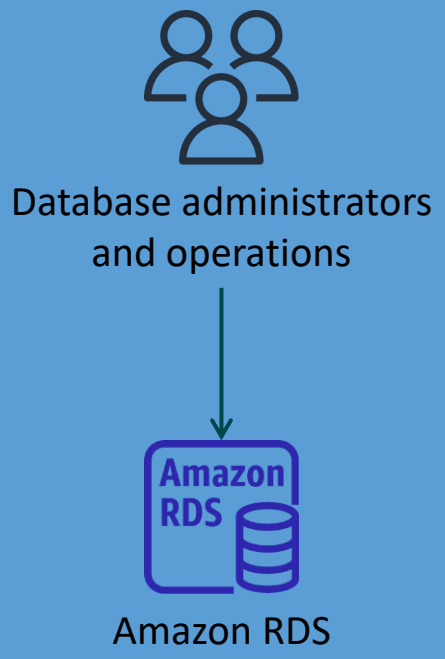
Controlled through Amazon VPC security groups



Protocol	Port range	Source
TCP	3306	172.31.0.0/16
TCP	3306	“Application security group”

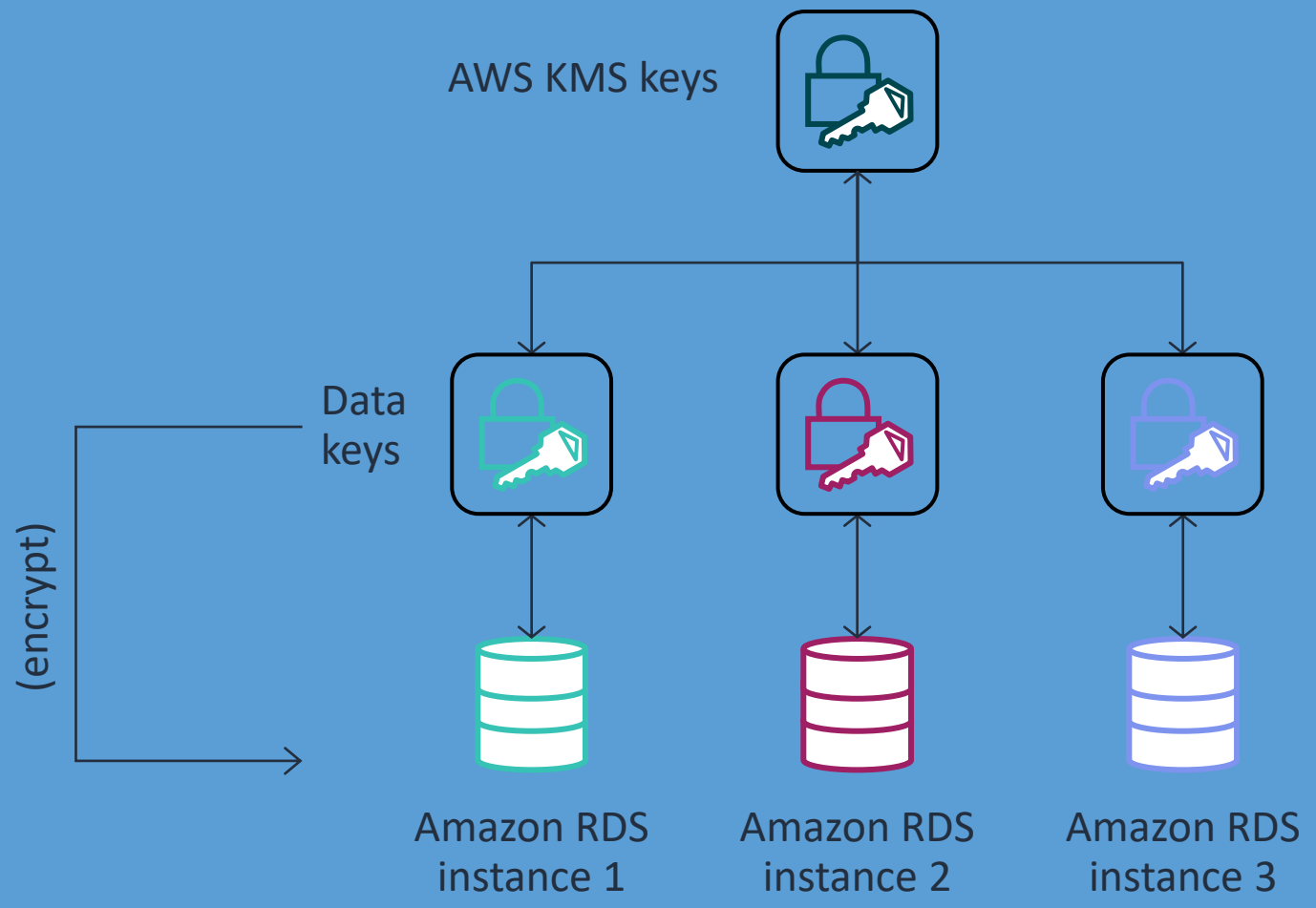
Resource-level role permissions

```
{  "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowCreateDBInstanceOnly",
            "Effect": "Allow",
            "Action": "rds:CreateDBInstance",
            "Resource": [
                "arn:aws:rds:*:123456789012:db:test*",
                "arn:aws:rds:*:123456789012:og:default*",
                "arn:aws:rds:*:123456789012:pg:default*",
                "arn:aws:rds:*:123456789012:subgrp:default*"
            ],
            "Condition": {
                "StringEquals": {
                    "rds:DatabaseEngine": "mysql",
                    "rds:DatabasedClass": "db.t2.micro"
                } ...
```



Data encryption at rest

- Managed by AWS KMS
- Unique data key encrypts your data
- AWS KMS key encrypts data keys
- Available for all RDS engines



Amazon Aurora

A MySQL and PostgreSQL compatible relational database built for the cloud

Performance
and scalability

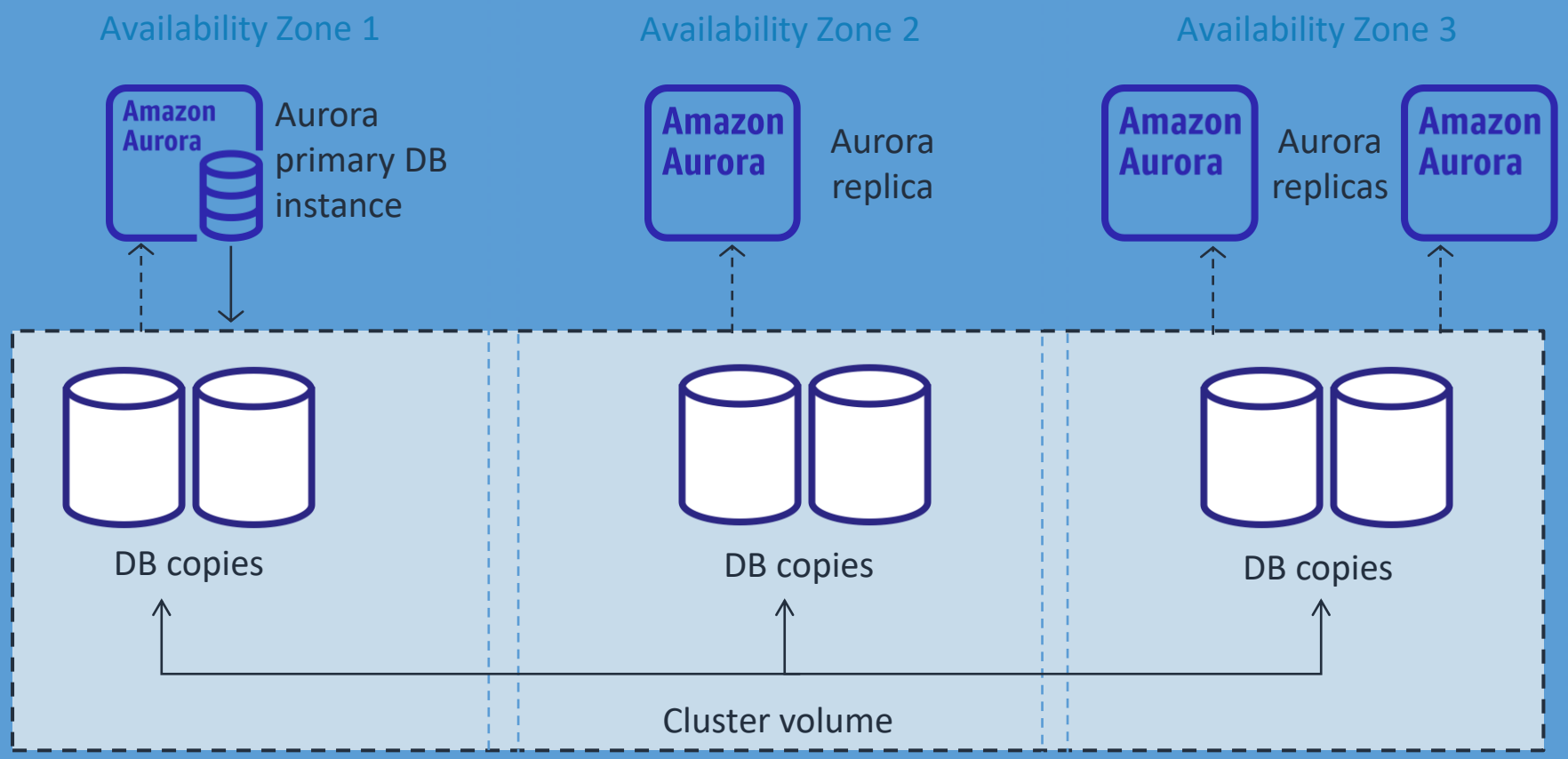
Availability
and durability

Highly
secure

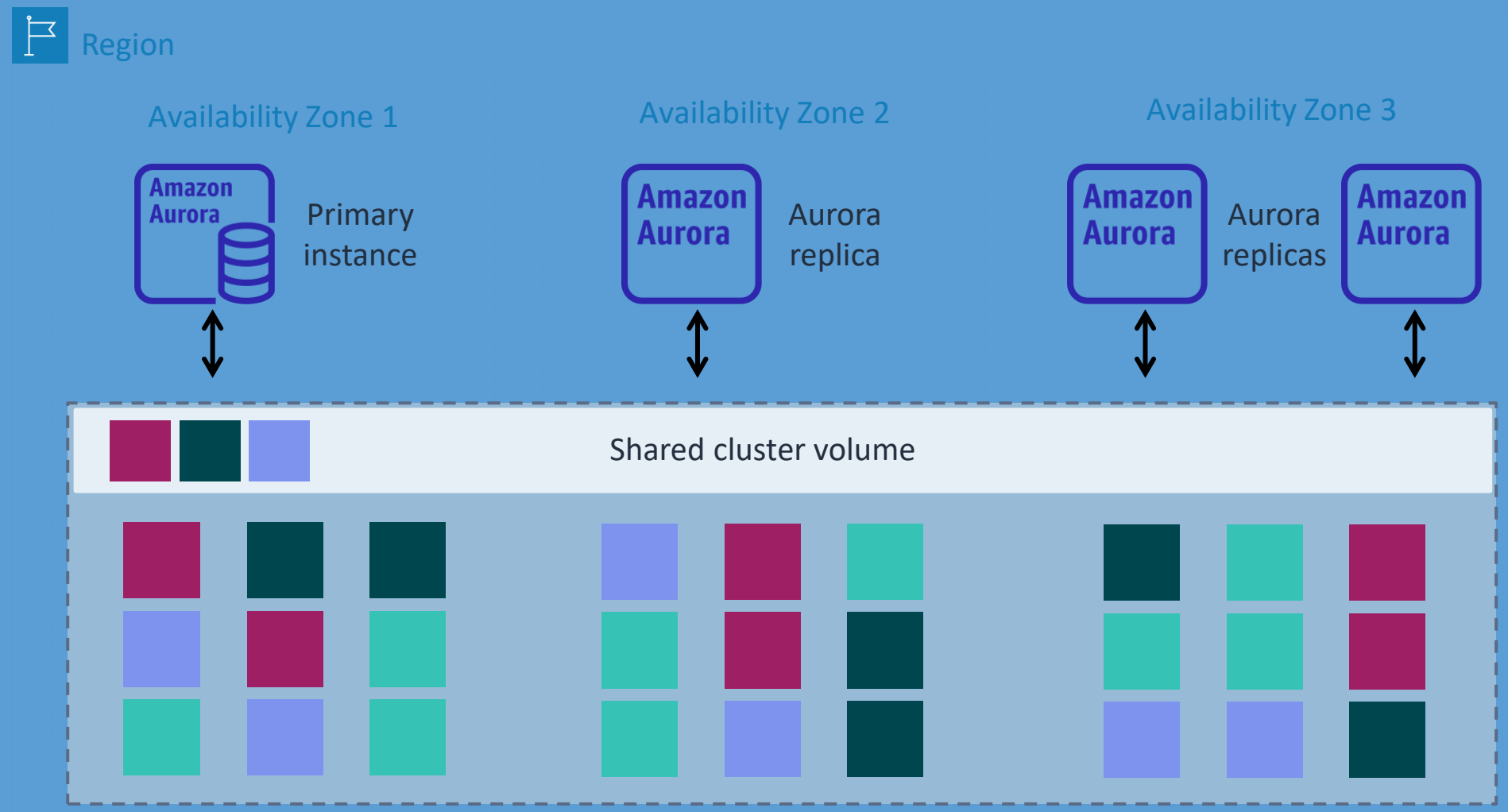
Fully
managed

Aurora DB clusters

- A DB cluster consists of one or more DB instances and a cluster volume.
- Primary instances perform read/write operations.
- Aurora replicas are read-only.
- A cluster volume is a virtual database storage volume that spans multiple Availability Zones.



Aurora storage and DB scaling



Aurora Serverless v2 for PostgreSQL and MySQL

Scaling configuration for Aurora that automatically scales capacity up or down based on your application's needs

Starts up on demand

Only pay for what you use

No application impact when scaling

Amazon DynamoDB

“How can we build a scalable key-value NoSQL database?”

DynamoDB

A fully managed NoSQL AWS database service

Performance at scale

No servers to manage

Enterprise ready

Key-value data

- Structured in simple key-value pairs with a flexible schema
- Ideal for uses where needed data can be mapped to a primary key
- Partitions data by key
- Delivers high-throughput, low-latency reads and writes

Database name: Gamers

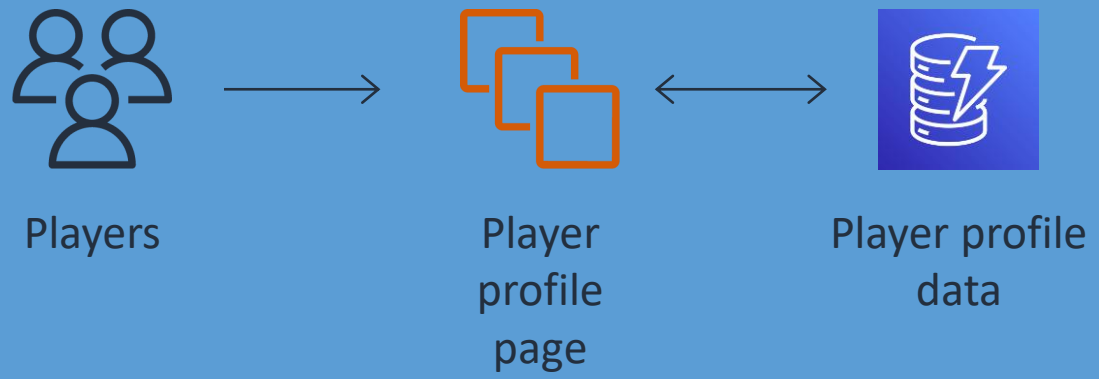
Primary Key

Attributes

UserId	GamerTag	Level	Points	High Score	Plays
107	“Hammer57”	21	4,050	483,610	1,722
285	“FluffyDuffy”	5	1,123	10,863	43

DynamoDB use case 1

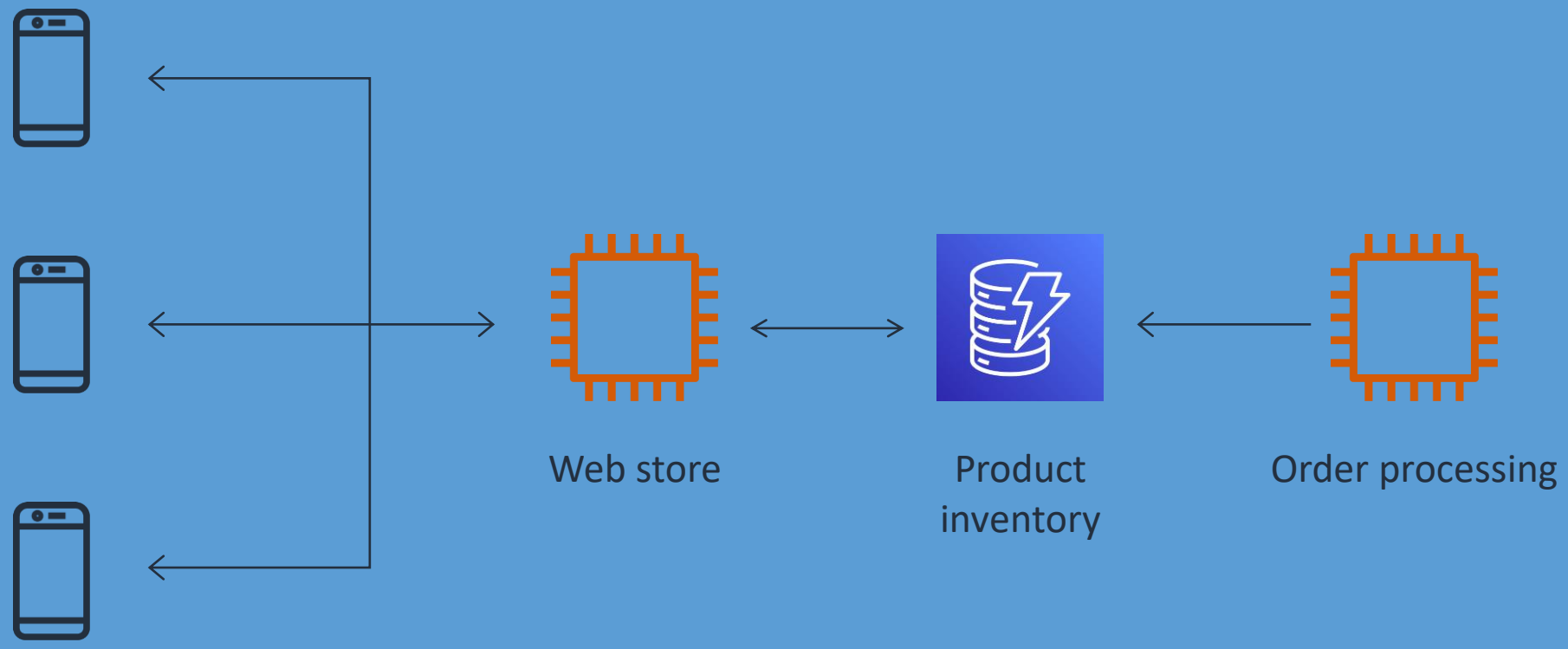
Player profile page



UserId	GamerTag	TopScore	MemberSince	SubscriptionType
101	“Hammer57”	5,842	“2021-09-15:17:24:31”	“Gold”
243	“FluffyDuffy”	1,024	“2021-10-22:23:18:01”	“Platinum”
623	“NewPlayer”	687	“2021-10-22:23:22:01”	“Free”

DynamoDB use case 2

ecommerce application



DynamoDB tables

- Mandatory key-value access pattern
- Partition key determines data distribution
- Sort key permits rich query capabilities

Composite primary key

Partition Key

Sort Key (Optional)

Table name: Actors

Item

Actor	Role	Year	Genre	Height
Tom Hanks	Chuck Noland	2000	Drama	
Tim Allen	Buzz Lightyear	1995		5’ 10”

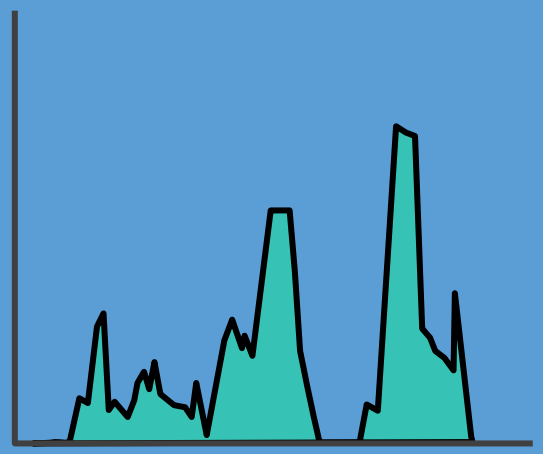
Attributes

DynamoDB capacity and scaling

- DynamoDB measures read capacity in read capacity units (RCUs).
  - Read requests for up to a 4 KB item
- DynamoDB measures write capacity in write capacity units (WCUs).
  - Number of write requests per second for up to a 1 KB item

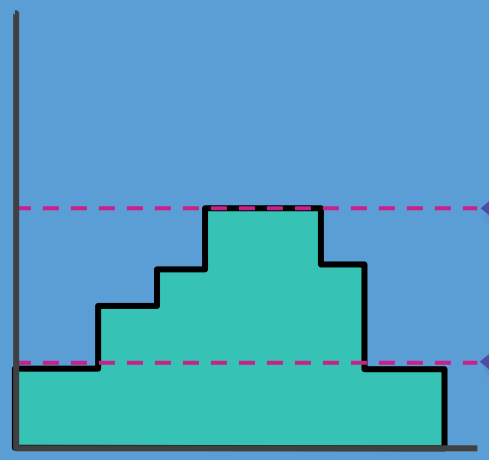
DynamoDB has two options for managing capacity:

On-Demand



Pay-per-request on reads and writes

Provisioned

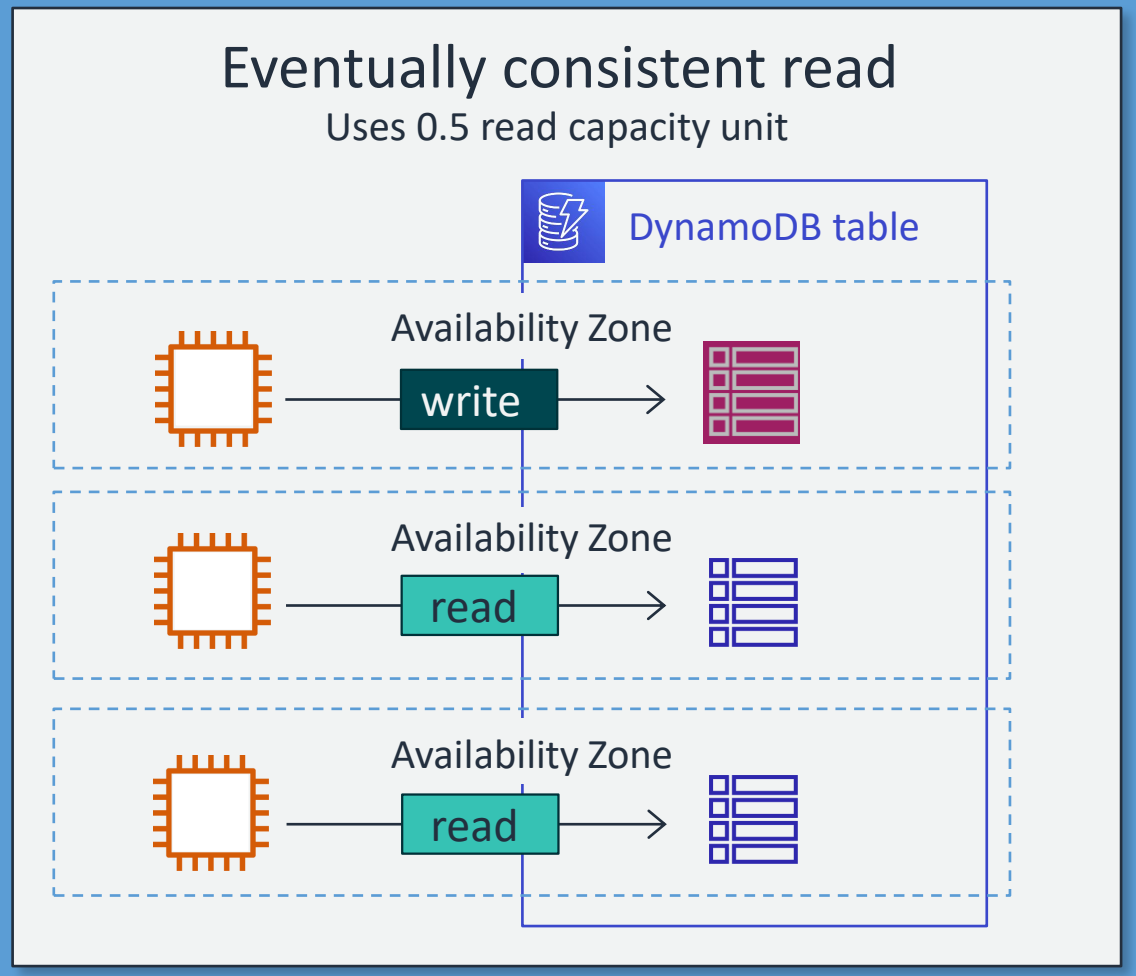


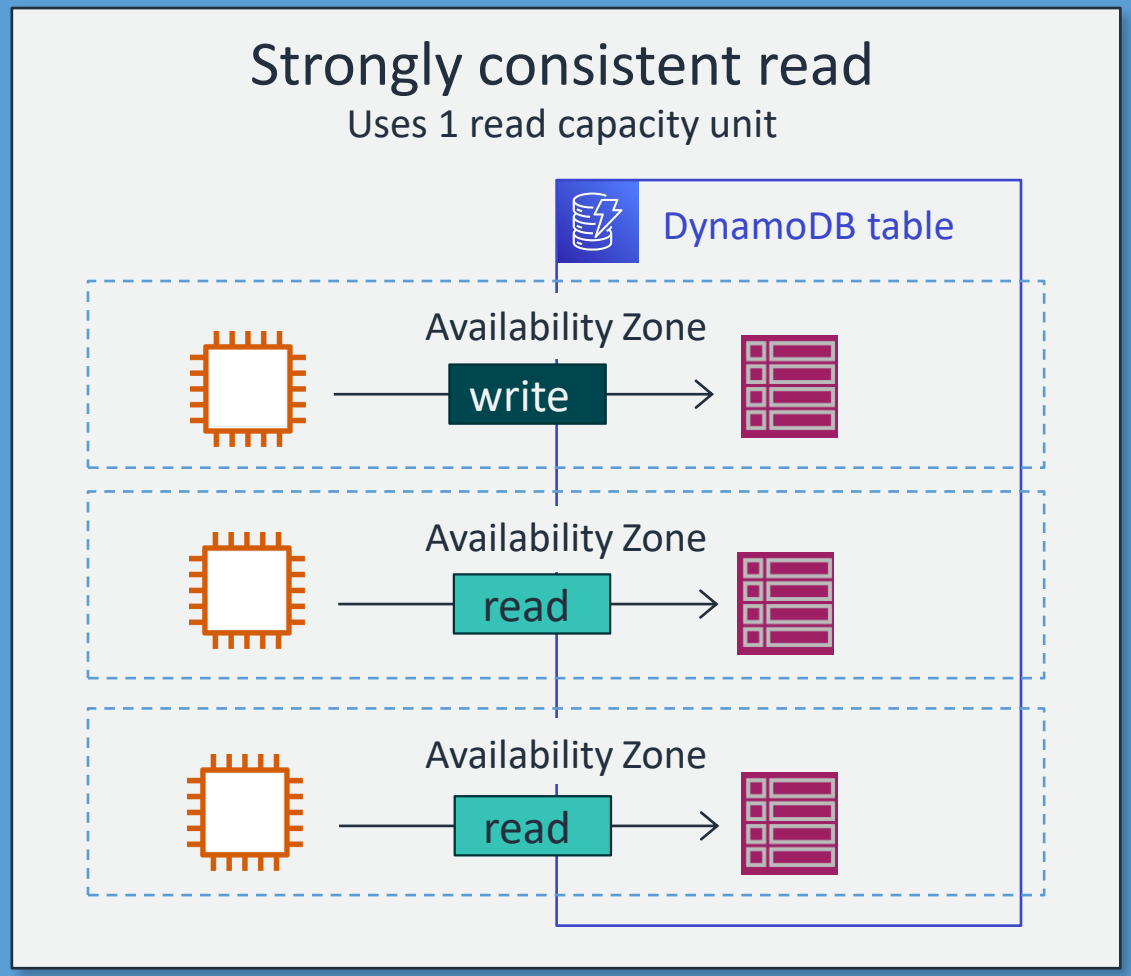
Set maximum RCUs and WCUs

Use auto scaling to adjust your provisioned capacity to match demand

DynamoDB consistency options

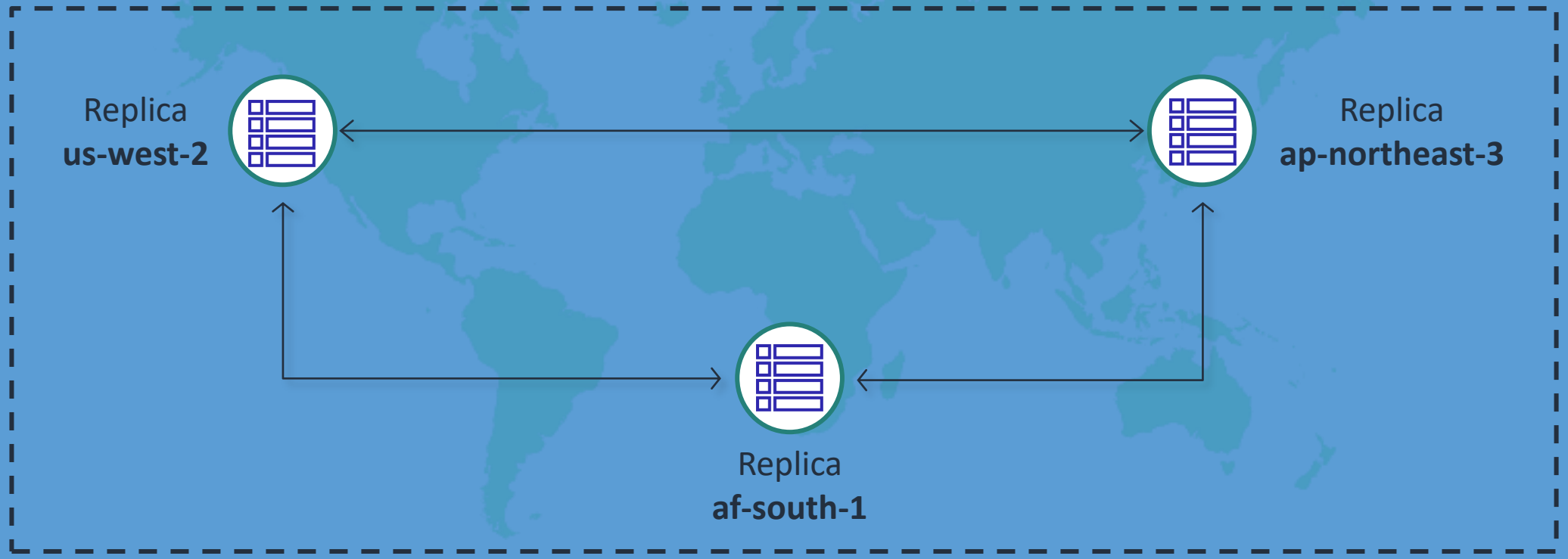
DynamoDB replicates table data across three Availability Zones in a Region usually within one second.





DynamoDB global tables

Global tables automate replication across Regions.



Database caching

“How can we cache databases in the cloud to maximize performance?”

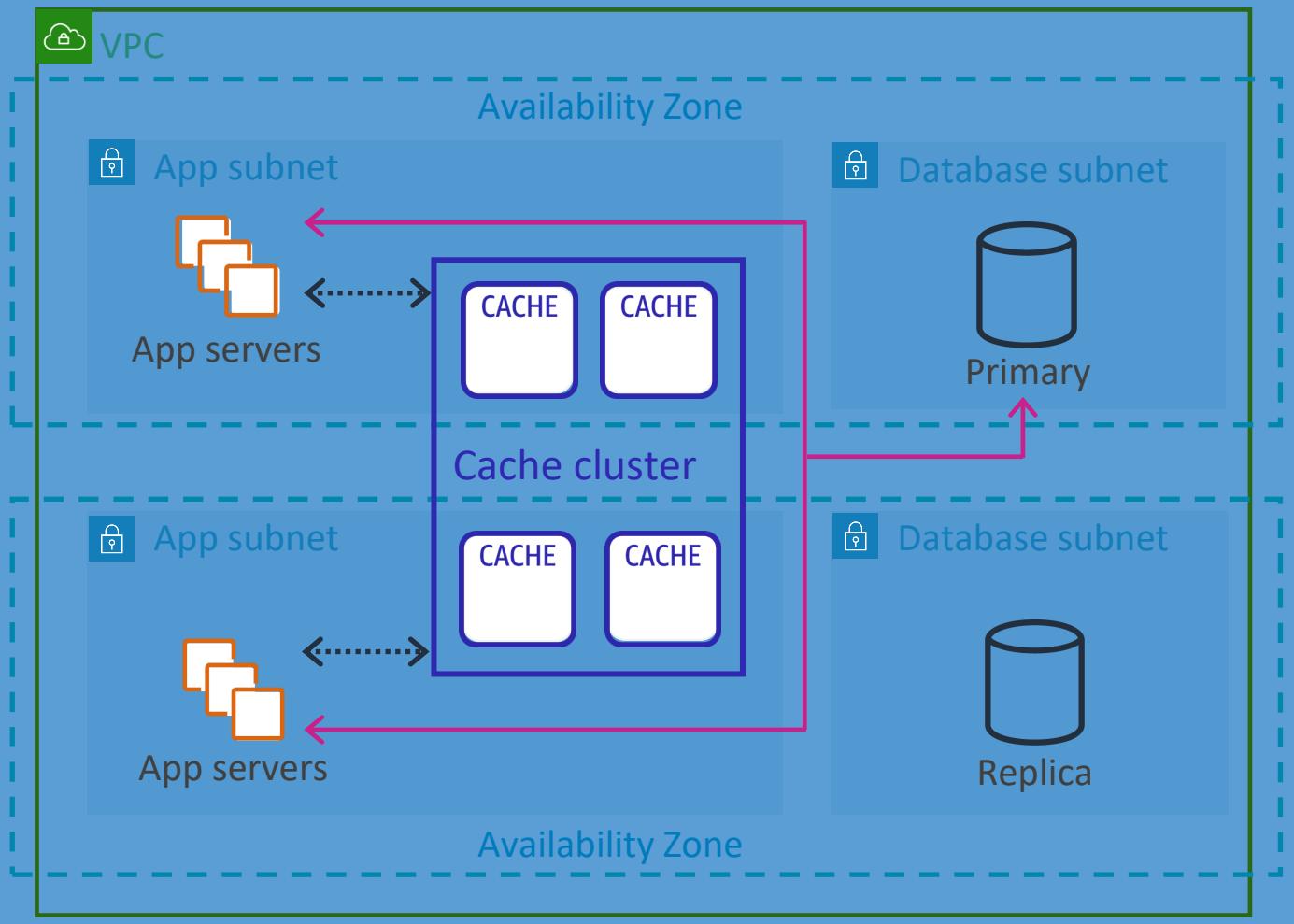
What should you cache?

Data that requires a slow and expensive query

Frequently accessed data

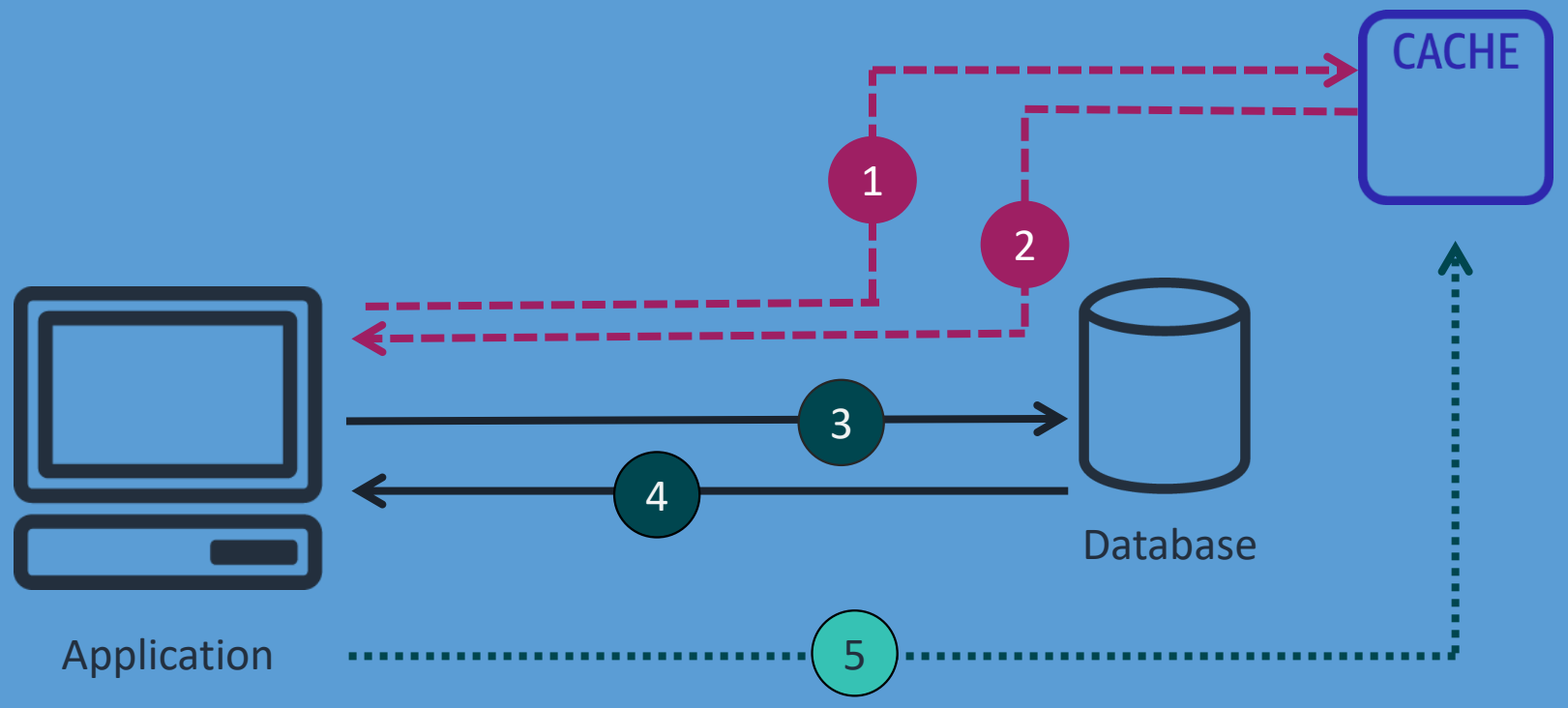
Information that is relatively static

Caching architecture



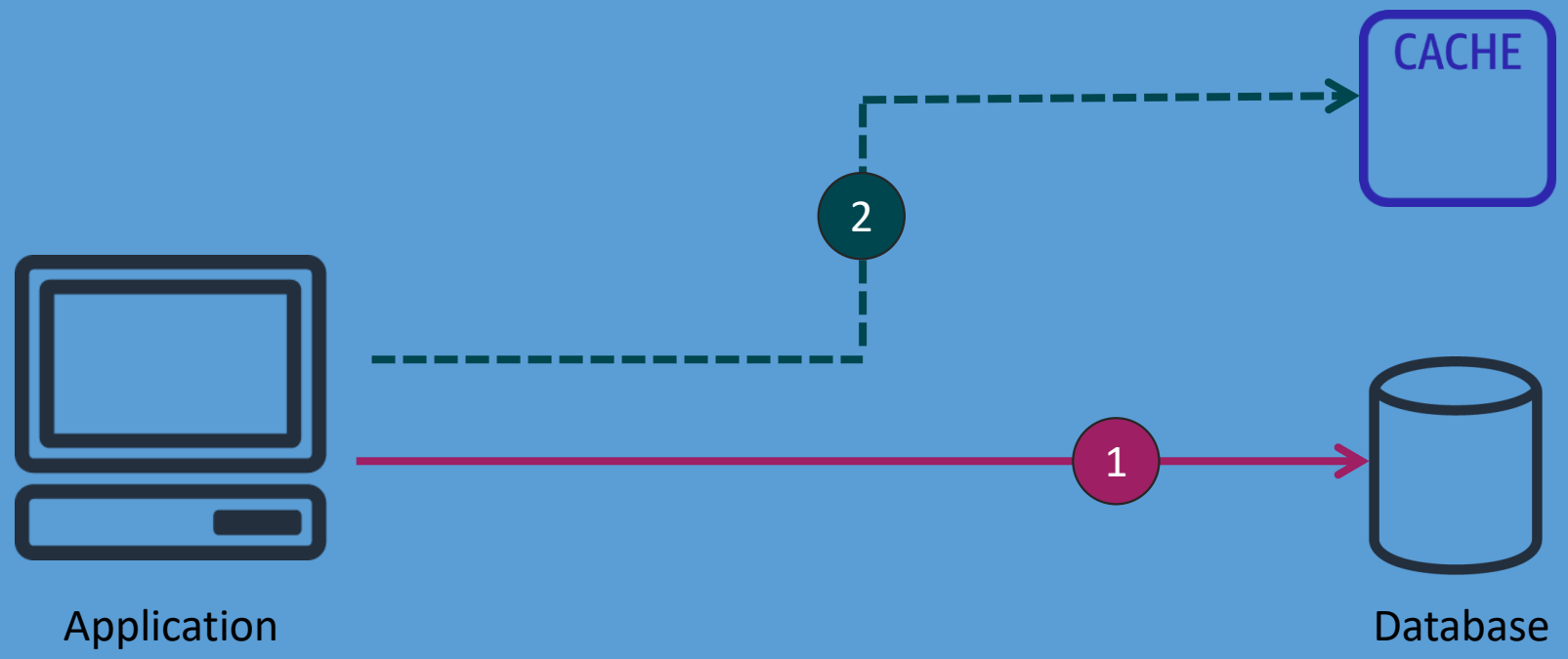
Common caching strategies – Lazy loading

1. Data request to the cache by the application
2. Cache miss
3. Missing data requested by the application from the database
4. Data returned from the database
5. Returned value written to the cache by the application



Common caching strategies – Write-through

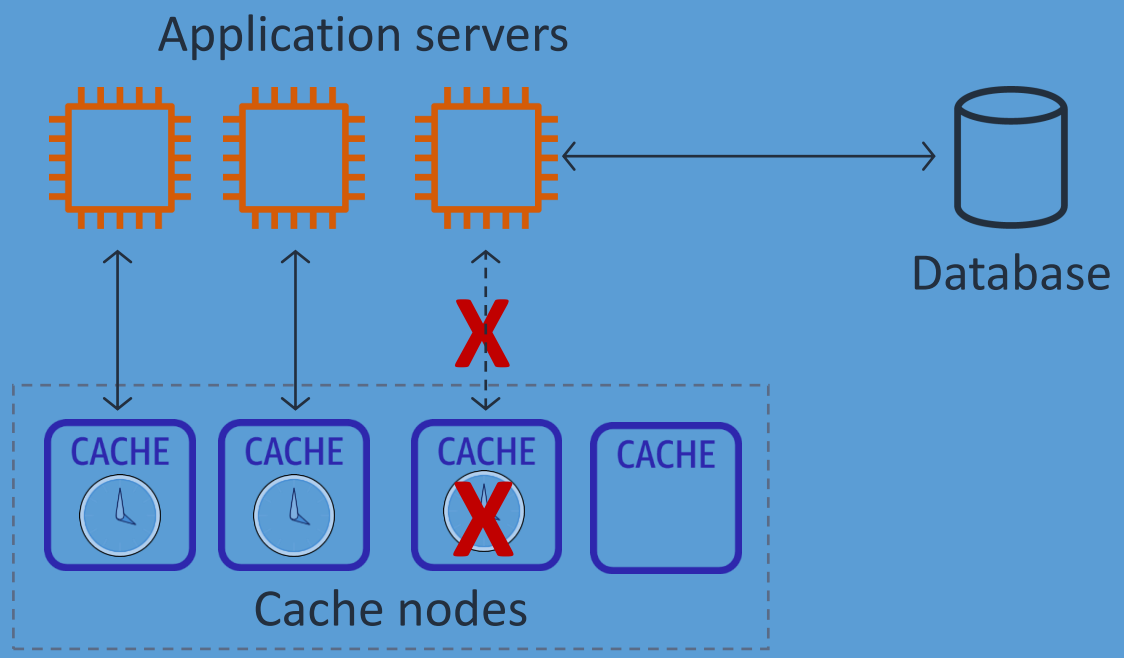
1. Application writes data to the database
2. Application also writes data to the cache



Managing your cache

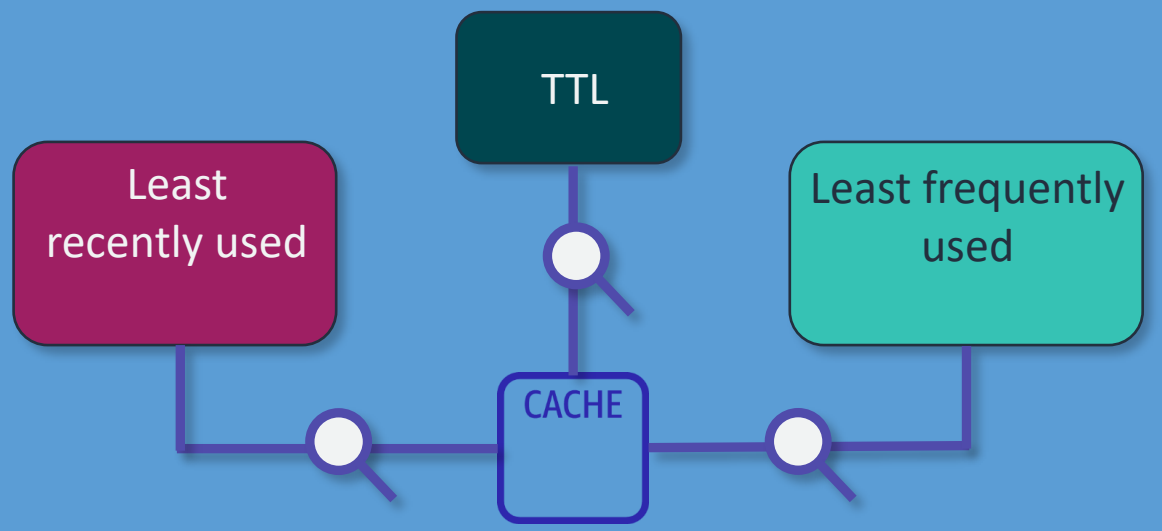
Cache validity

To minimize stale data, you can add a time to live (TTL) value to each application write.



Managing memory

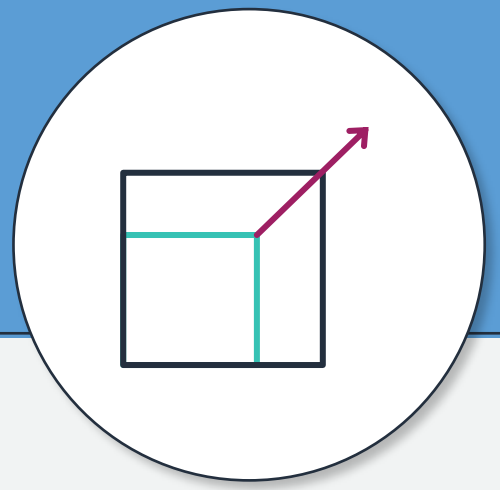
When your cache memory is full, your cache evicts data based on your selected eviction policy. Eviction policies can evaluate any combination of the following:



Amazon ElastiCache

Extreme performance

Fully managed



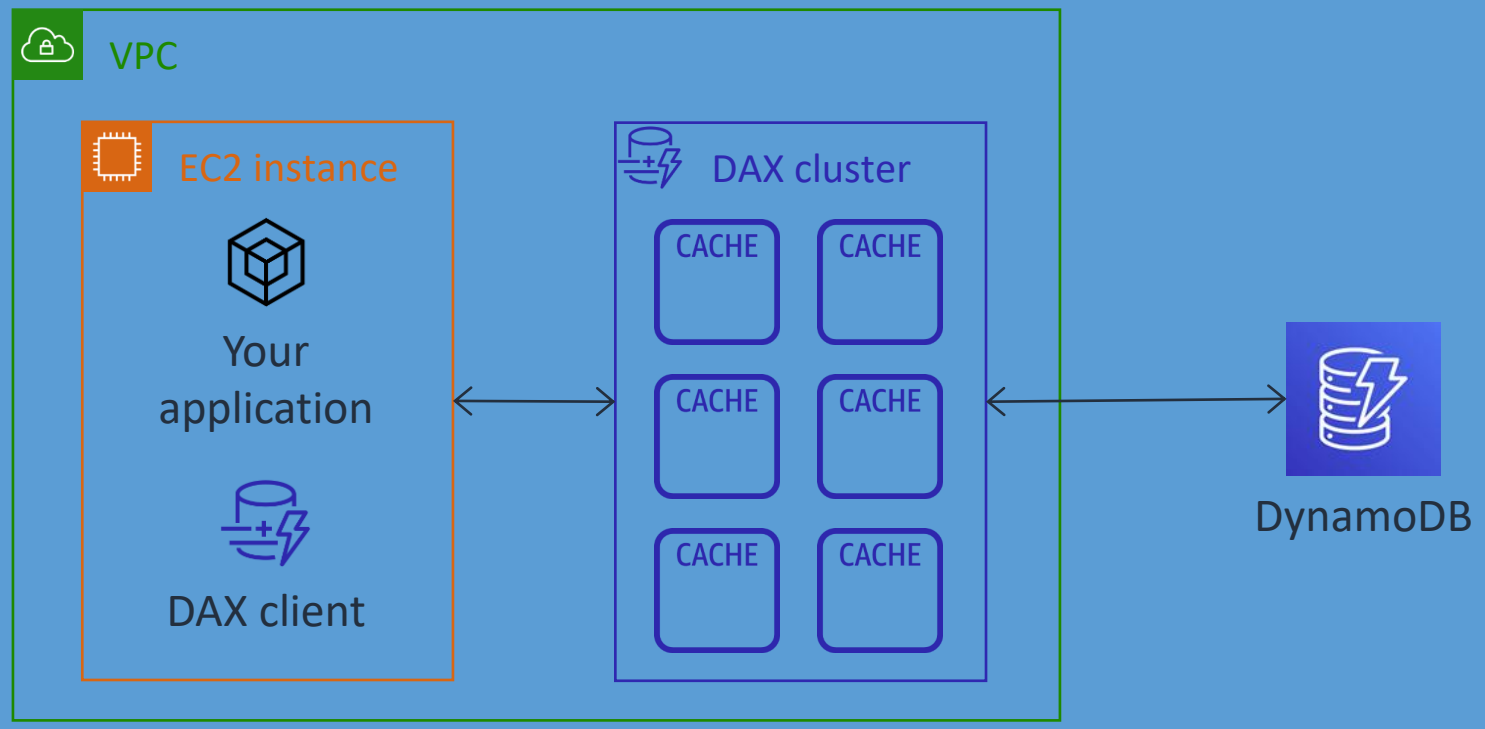
Easily scalable

ElastiCache engines

Feature	ElastiCache for Memcached	ElastiCache for Redis
Simple cache to offload database burden	Yes	Yes
Ability to scale horizontally for writes and storage	Yes	Yes (when using cluster mode)
Multi-AZ deployments	Yes	Yes
Multi-threaded performance	Yes	Not featured
Advanced data types	Not featured	Yes
Sorting and ranking data sets	Not featured	Yes
Publish and subscribe capability	Not featured	Yes
Backup and restore	Not featured	Yes

DynamoDB Accelerator (DAX)

- A fully managed, highly available cache for DynamoDB
- Can deliver microsecond response times
- Can scale to millions of read requests per second



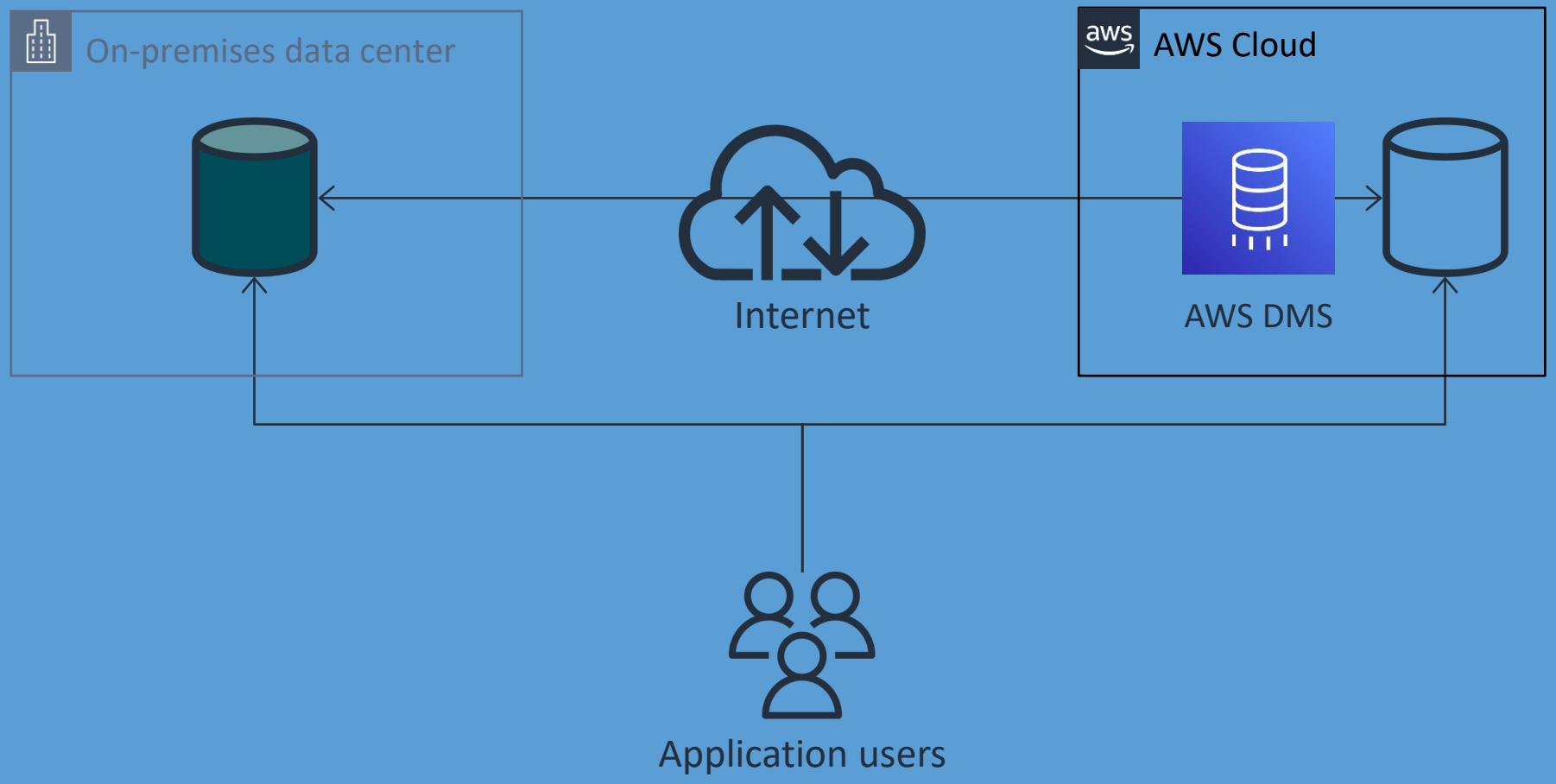
Database migration tools

“What tools are available for migrating an existing database to the AWS Cloud?”

AWS Database Migration Service

AWS Database Migration Service (AWS DMS)

- Heterogeneous database migrations
- Database consolidation
- Continuous data replication
- Can point to a database, Amazon S3, Snowball Edge, or other services

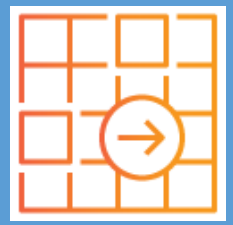


AWS Schema Conversion Tool

Source Databases

Oracle Database
Oracle Data Warehouse
Azure SQL
SQL Server
Teradata
IBM Netezza
Greenplum
HPE Vertica
MySQL and MariaDB
PostgreSQL
Aurora
IBM DB2 LUW
Apache Cassandra
SAP ASE

→



AWS Schema Conversion Tool (AWS SCT)

→

Target Databases on AWS

MySQL
PostgreSQL
Oracle
Amazon Redshift
DynamoDB
RDS for MySQL
Aurora MySQL
RDS for PostgreSQL
Aurora PostgreSQL

Review

Present solutions

Database Services Manager

Consider how you would answer the following:

- What are the AWS database solutions?
- How can we more efficiently manage our relational databases in the cloud?
- How can we build a scalable key-value NoSQL database?
- How can we cache databases in the cloud to maximize performance?
- What tools are available for migrating an existing database to the AWS Cloud?

Module review

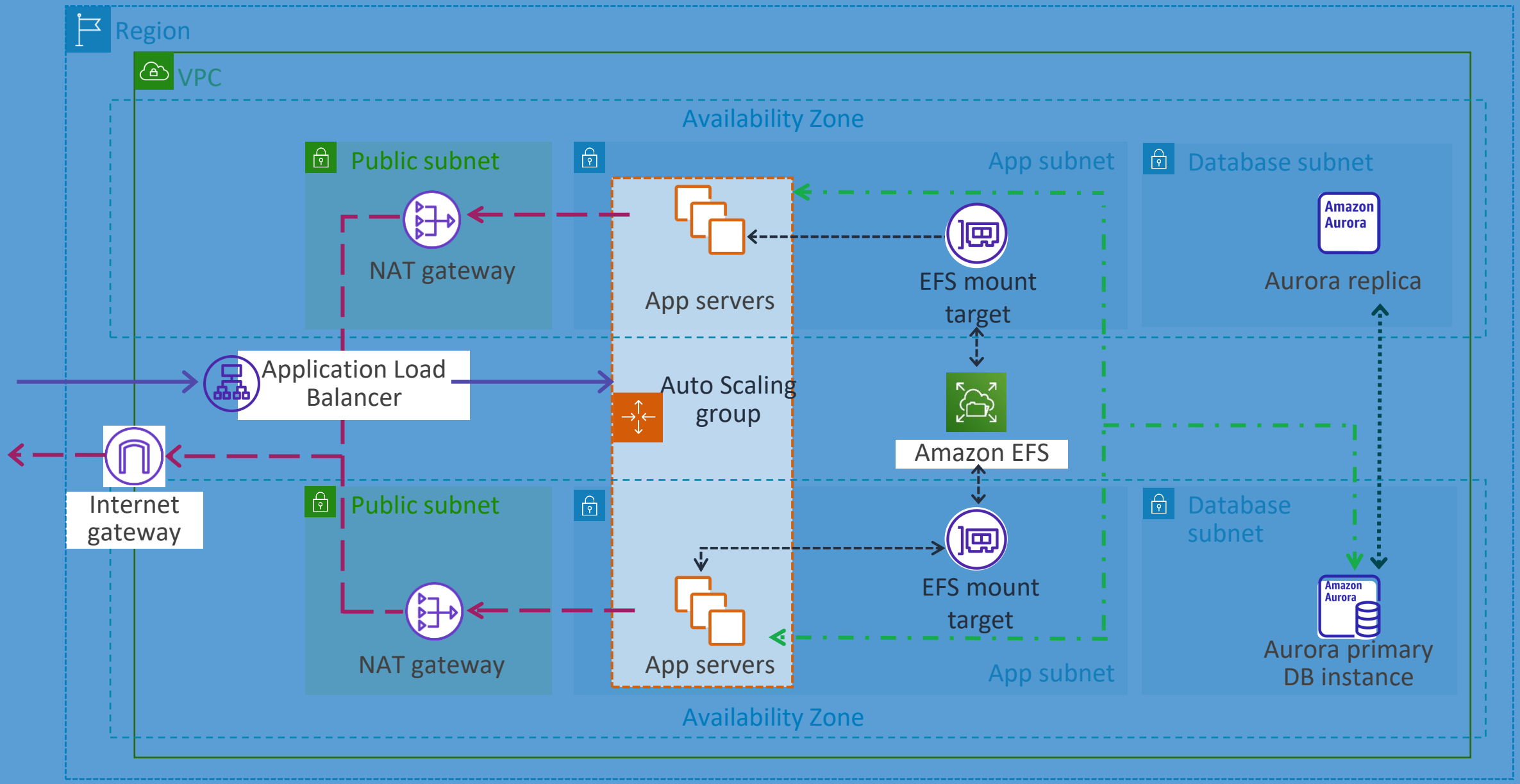
In this module you learned about:

- ✓ Database services
- ✓ Amazon RDS
- ✓ Amazon DynamoDB
- ✓ Database caching
- ✓ Database migration tools

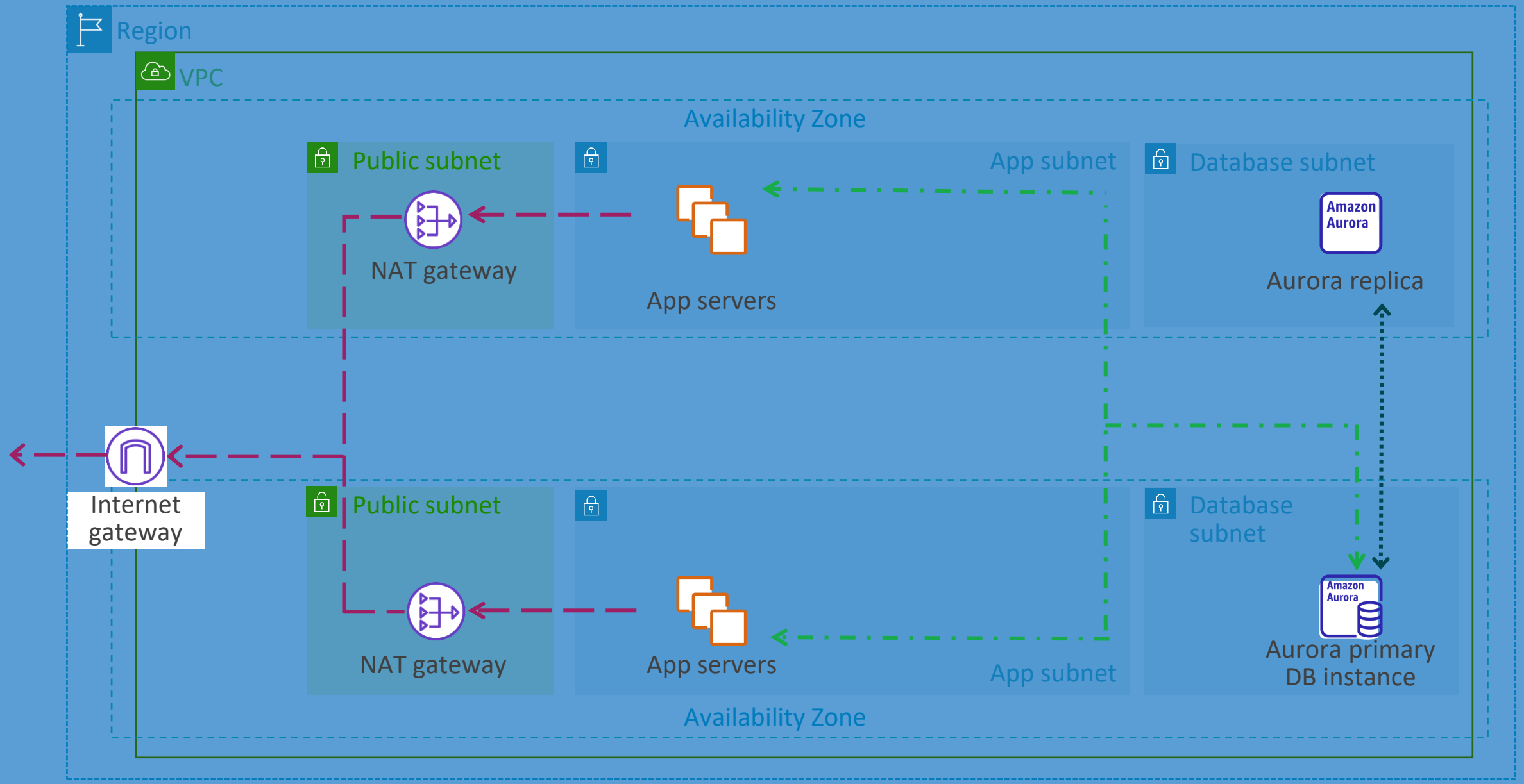
Next, you will review:

- Capstone check-in
- Knowledge check
- Lab introduction

Capstone architecture



Capstone architecture check-in



Knowledge check

Knowledge check question 1

What is a benefit of using Amazon RDS in a Multi-AZ configuration?

A	It delivers two live copies of the database running concurrently.
B	It provides automatic failover across Availability Zones.
C	It provides automatic cross-Region replication.
D	It eliminates the need for read replicas.

Knowledge check question 1 and answer

What is a benefit of using Amazon RDS in a Multi-AZ configuration?

A	It delivers two live copies of the database running concurrently.
B correct	It provides automatic failover across Availability Zones.
C	It provides automatic cross-Region replication.
D	It eliminates the need for read replicas.

Knowledge check question 2

What type of ElastiCache installation offers sorting and ranking capabilities for data sets?

A	ElastiCache for Redis
B	DAX
C	Lazy loading
D	ElastiCache for Memcached

Knowledge check question 2 and answer

What type of ElastiCache installation offers sorting and ranking capabilities for data sets?

A correct	ElastiCache for Redis
B	DAX
C	Lazy loading
D	ElastiCache for Memcached

Knowledge check question 3

Which of the following is true regarding DynamoDB global tables?

A	Tables are updated manually or through automation tools.
B	Only two tables are active at one time.
C	You can select different instance sizes to adjust performance.
D	Tables can be in different AWS Regions.

Knowledge check question 3 and answer

Which of the following is true regarding DynamoDB global tables?

A	Tables are updated manually or through automation tools.
B	Only two tables are active at one time.
C	You can select different instance sizes to adjust performance.
D correct	Tables can be in different AWS Regions.

Knowledge check question 4

Which of the following is true regarding an Aurora database?

A	Nine copies of the data are stored across three Availability Zones.
B	Aurora has a limit of five replicas.
C	Aurora is compatible with MySQL or PostgreSQL.
D	Multi-AZ deployments are not required for high availability.

Knowledge check question 4 and answer

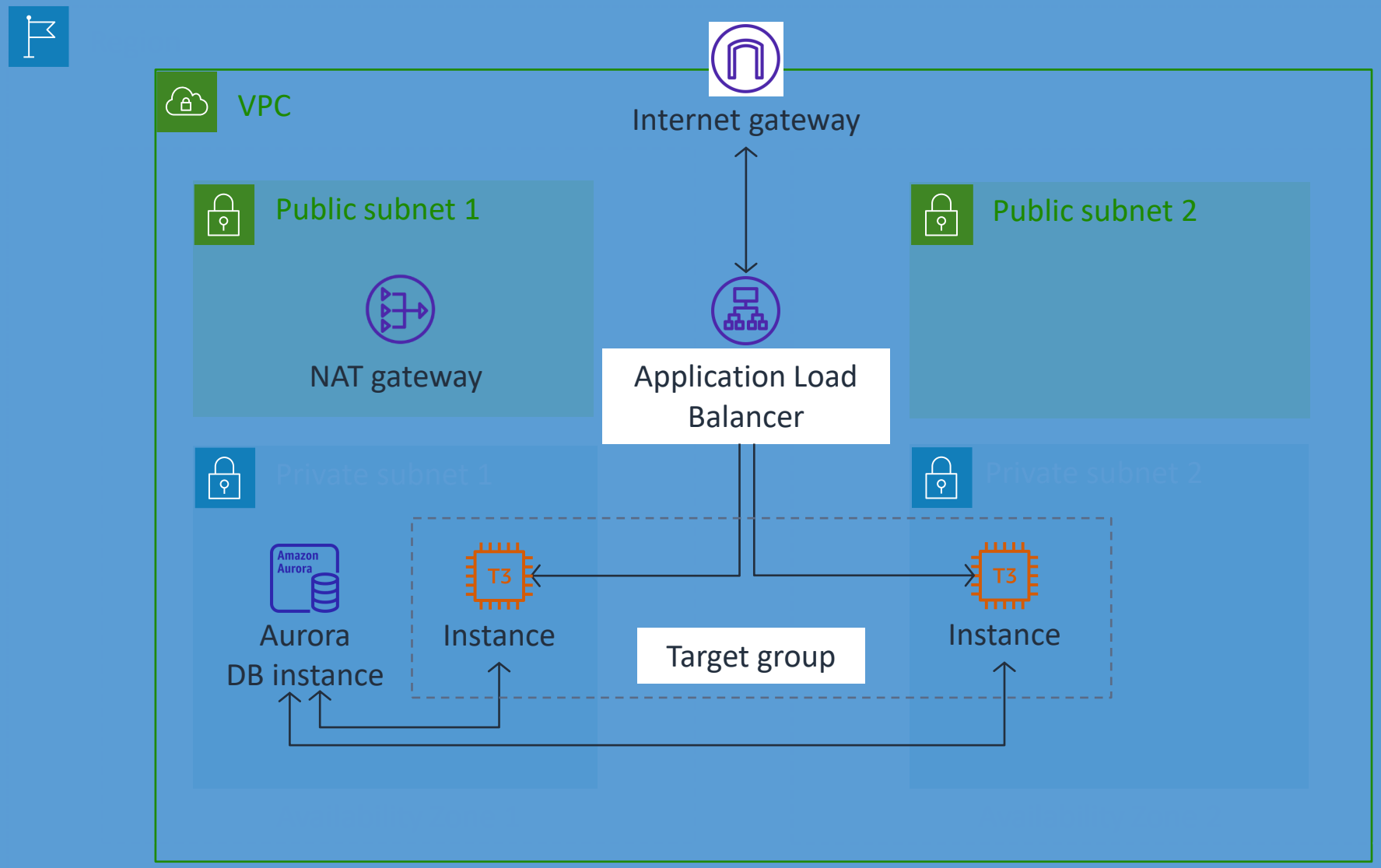
Which of the following is true regarding an Aurora database?

A	Nine copies of the data are stored across three Availability Zones.
B	Aurora has a limit of five replicas.
C correct	Aurora is compatible with MySQL or PostgreSQL.
D	Multi-AZ deployments are not required for high availability.

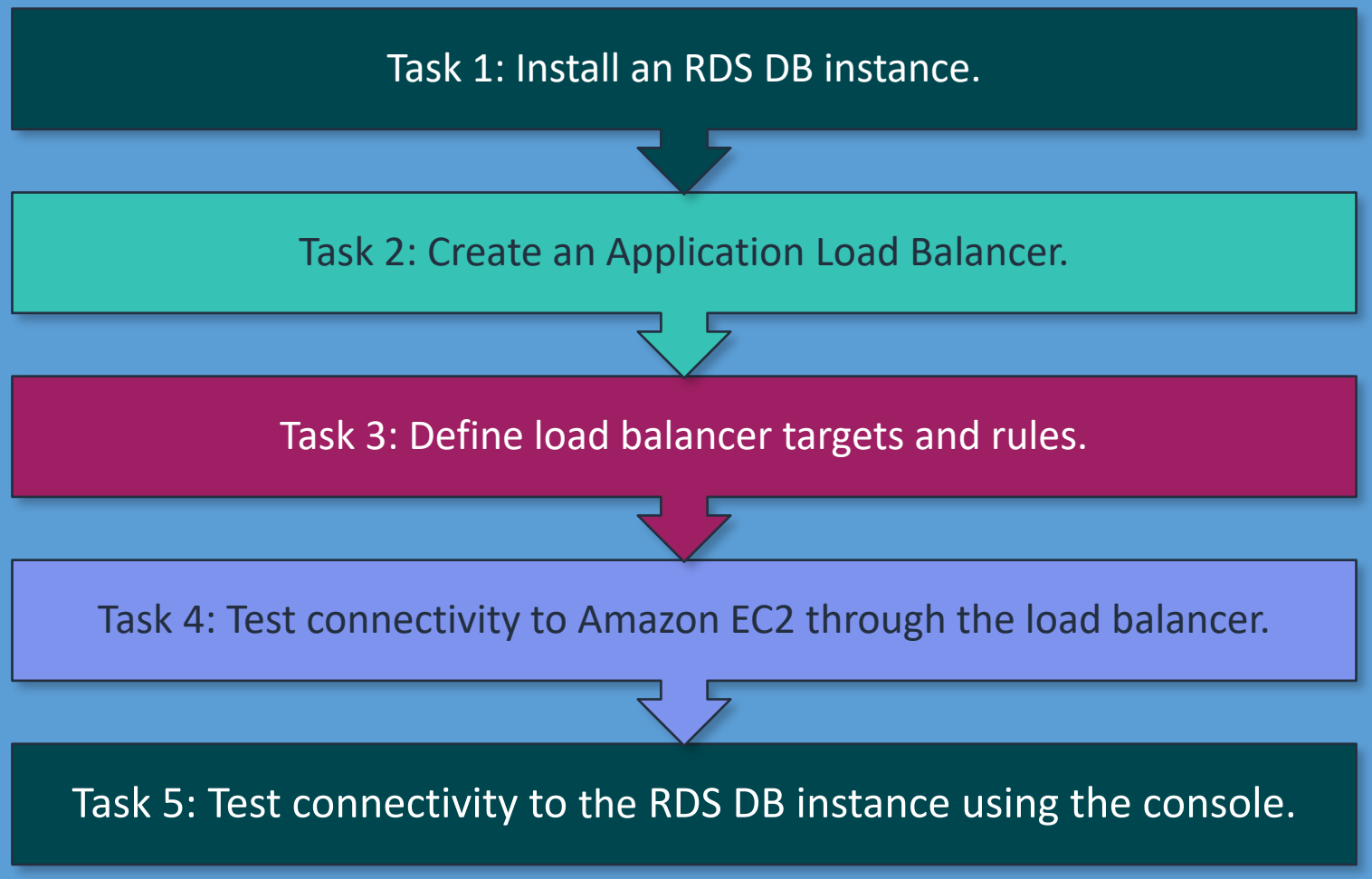
Lab 3:

Create a database layer in your Amazon VPC infrastructure

Lab 3 diagram



Lab tasks


Task 1: Install an RDS DB instance.
Task 2: Create an Application Load Balancer.
Task 3: Define load balancer targets and rules.
Task 4: Test connectivity to Amazon EC2 through the load balancer.
Task 5: Test connectivity to the RDS DB instance using the console.

End of Module 6

AWS

Monitoring and Scaling

Lab 4

Question

What factors impact your decisions for making your workloads scalable?

A. Cost

B. Usage patterns

C. Expected growth

D. Criticality of workload

E. All of the above

Module overview

- Business requests
- Monitoring
- Alarms and events
- Load balancing
- Auto scaling
- Present solutions
- Capstone check-in
- Knowledge check
- Lab 4: Configure high availability in your Amazon VPC

Business requests

Operations Manager

The operations manager needs to know:

- What tools and services are available to monitor and log activity in my AWS accounts?
- How can we set thresholds and be alerted to changes in our infrastructure?
- How do we add high availability to our Amazon EC2 workloads and distribute traffic across multiple targets?
- How can we dynamically increase and decrease capacity to meet changing demand?

Monitoring

“What tools and services are available to monitor and log activity in our AWS accounts?”

Reasons for monitoring

Operational health

Get operational visibility and insight.

Application performance

Collect data at every layer of the performance stack.

Resource utilization

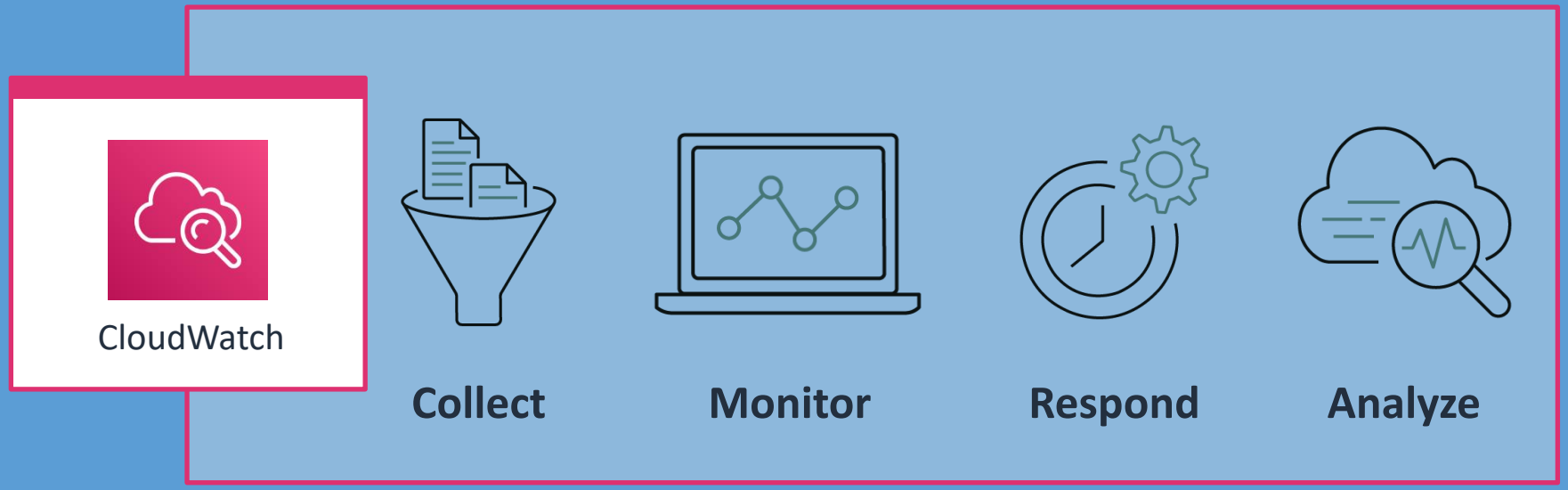
Improve resource optimization.

Security auditing

Automate and manage evidence collection, security, and integrity.

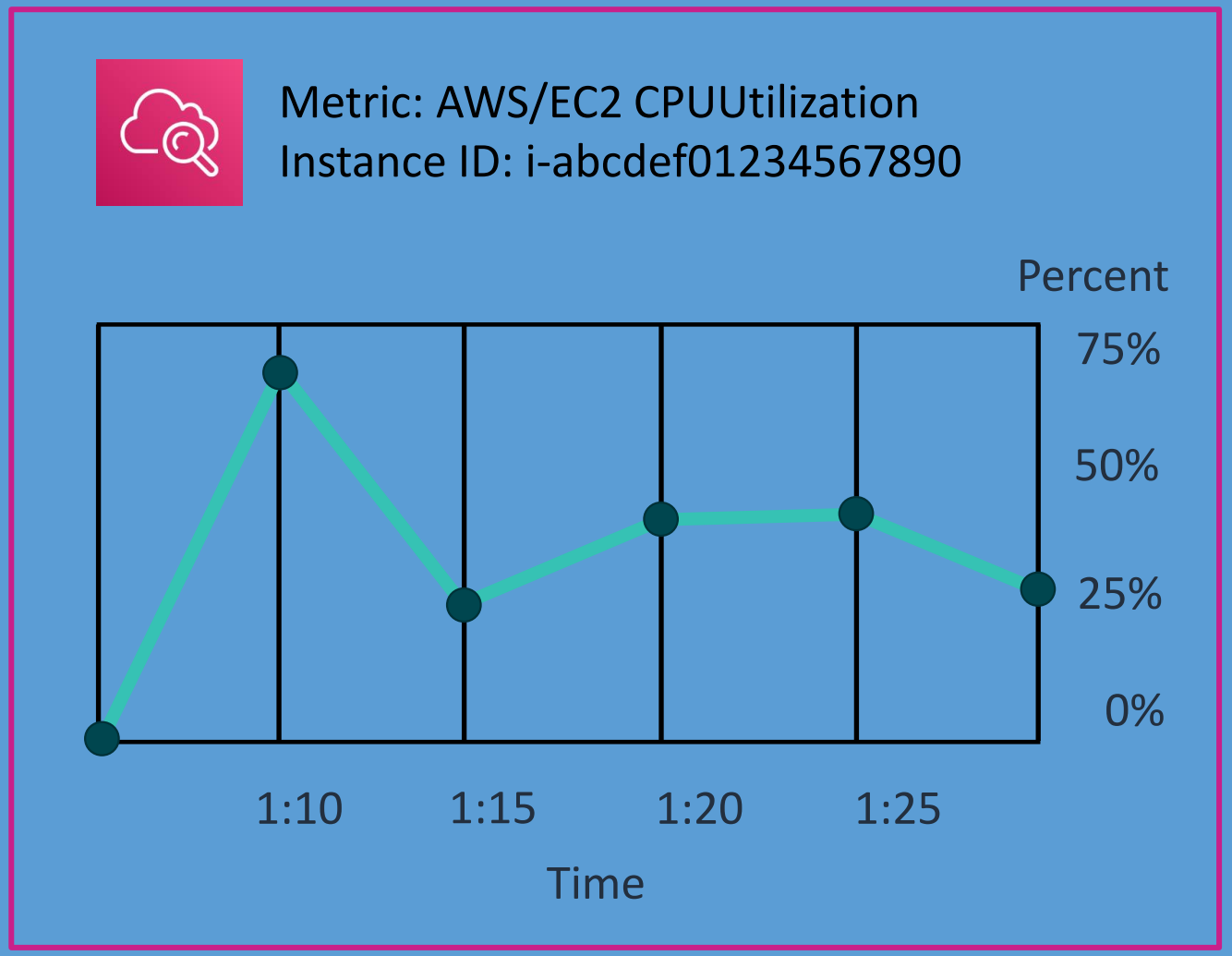
Amazon CloudWatch

- Collect near real-time metrics and logs.
- Access monitoring data in one place.
- Create alarms and send notifications.
- Initiate changes to resource capacity based on rules.
- Create and view dashboards.



CloudWatch metrics

- Metrics are data about system performance.
- CloudWatch ingests and tracks metrics so you can search and visualize data.



Types of logs

Amazon CloudWatch Logs

Monitor apps using log data, store, and access log files.

AWS CloudTrail

Track user activity and API usage.

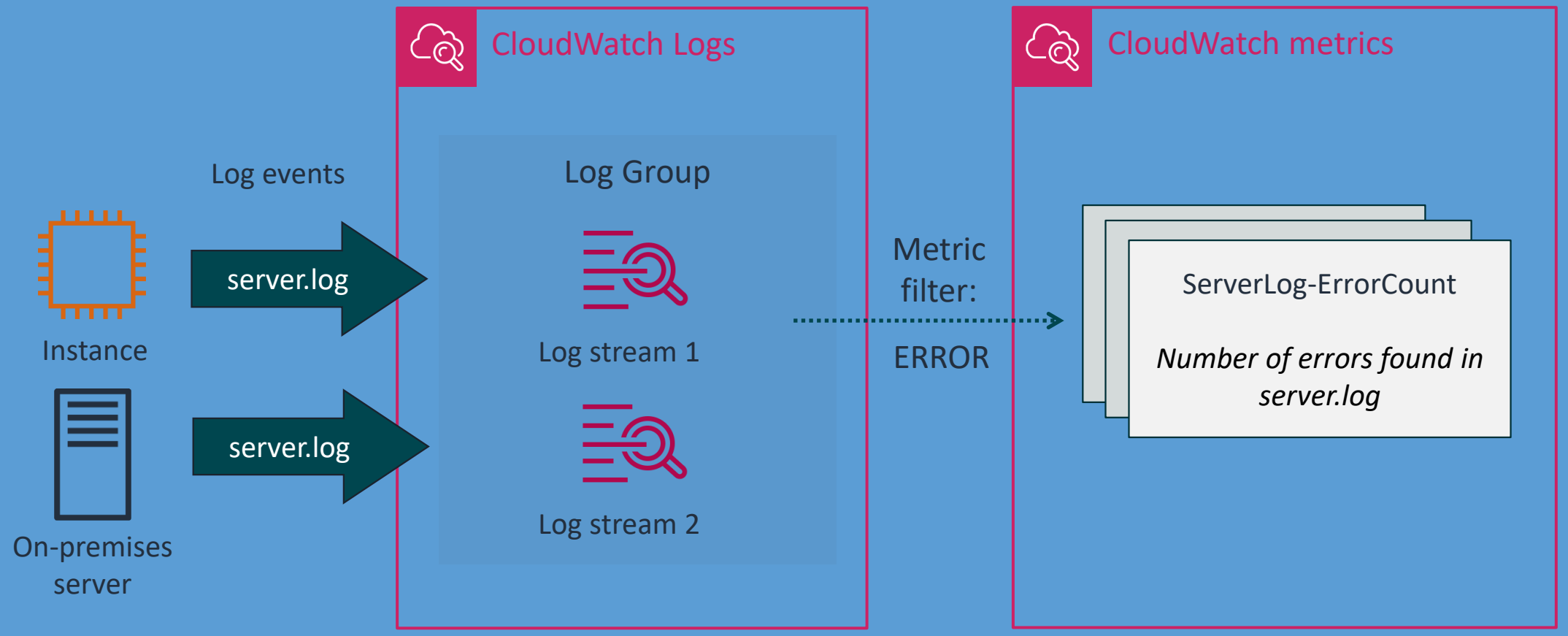
VPC Flow Logs

Capture information about IP traffic to and from network interfaces.

Custom logs

Store custom logs generated from your application instances.

CloudWatch Logs example



AWS CloudTrail

- Log and monitor account activity across your AWS infrastructure.
- Record API call interactions for most AWS services.
- Automatically push logs to Amazon S3.

CloudTrail helps you understand events in your accounts.

Who shut down a specific instance

What activities were denied due to lack of permissions

Who changed a security group configuration

Example: CloudTrail log (1 of 4)

```
{
    "Records": [{
        "eventVersion": "1.0",
        "userIdentity": {
            "type": "IAMUser",
            "principalId": "EX_PRINCIPAL_ID",
            "arn": "arn:aws:iam::123456789012:user/Alice",
            "accountId": "123456789012",
            "accessKeyId": "EXAMPLE_KEY_ID",
            "userName": "Alice"
        },
```

Who made the request?

Example: CloudTrail log (2 of 4)

```
"requestParameters": {
        "instancesSet": {
            "items": [{
                "instanceId": "i-abcdefg01234567890"
            }]
        },
        "force": false
    },
```

What was the focus of the request?

Example: CloudTrail log (3 of 4)

```
"eventTime": "2018-03-06T21:01:59Z",
        "eventSource": "ec2.amazonaws.com",
        "eventName": "StopInstances",
        "awsRegion": "us-west-2",
        "sourceIPAddress": "205.251.233.176",
        "userAgent": "ec2-api-tools 1.6.12.2",
```

When did the request occur?

What was the API call?

Where did it occur?

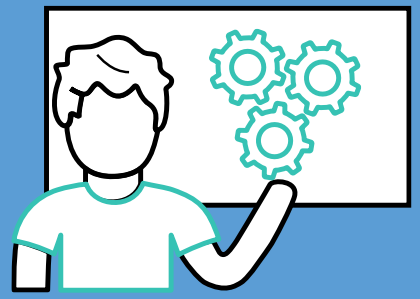
Example: CloudTrail log (4 of 4)

```
"responseElements": {
    "instancesSet": {
        "items": [{
            "instanceId": "i-abcdefg01234567890",
            "currentState": {
                "code": 64,
                "name": "stopping"
            },
            "previousState": {
                "code": 16,
                "name": "running"
            }
```

What was the response?

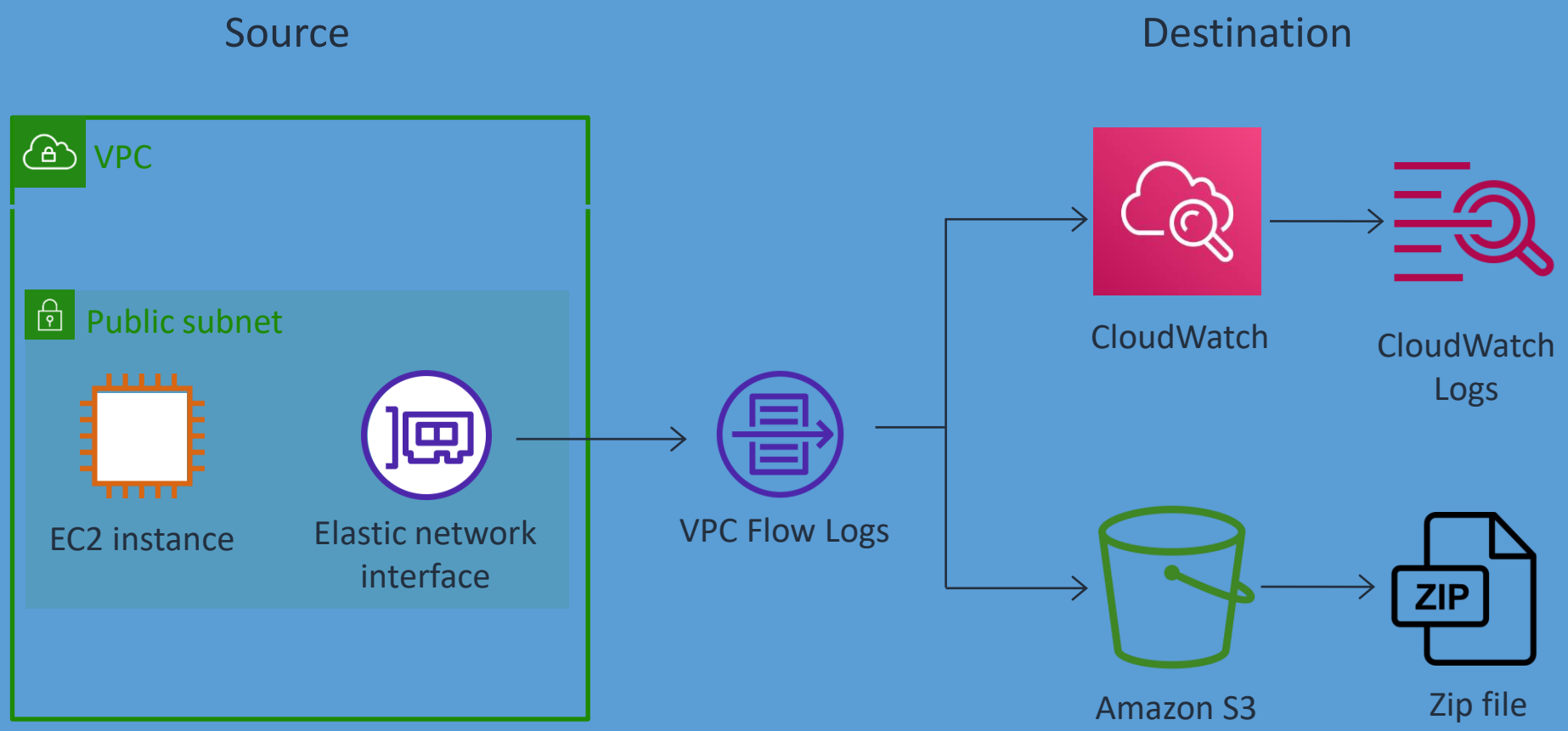
Demonstration:

CloudTrail



VPC Flow Logs

Capture IP traffic information going to and from VPC network interfaces.



Contents of a flow log record

VPC Flow Logs

AWS account

Elastic network interface ID

Time in Unix seconds; number of packets and bytes transferred

Action taken based on security group or network ACL

Version	2
Account ID	123456789010
Interface ID	eni-04b10a1942977452f
Source address	172.168.1.2
Destination address	32.68.32.56
Source port	36490
Destination port	443
Protocol	6
Packets	77
Bytes	5040
Start	1560385064
End	1560385070
Action	ACCEPT
Log status	OK

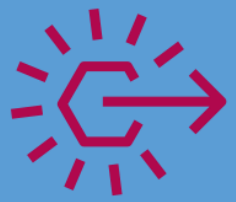
Alarms and events

“How can we set thresholds and be alerted to changes in our infrastructure?”

CloudWatch alarms

1 Identify the CloudWatch metric.

2 Create your alarms based on metrics.



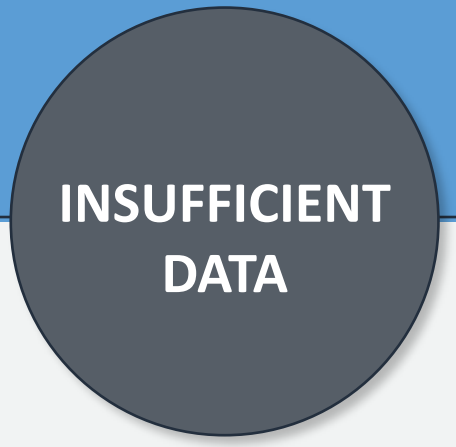
3 Define the actions to take when your metric's threshold is exceeded.

Alarm states

Test a selected metric against a specific threshold value. ALARM is not necessarily an emergency condition.

Threshold not exceeded

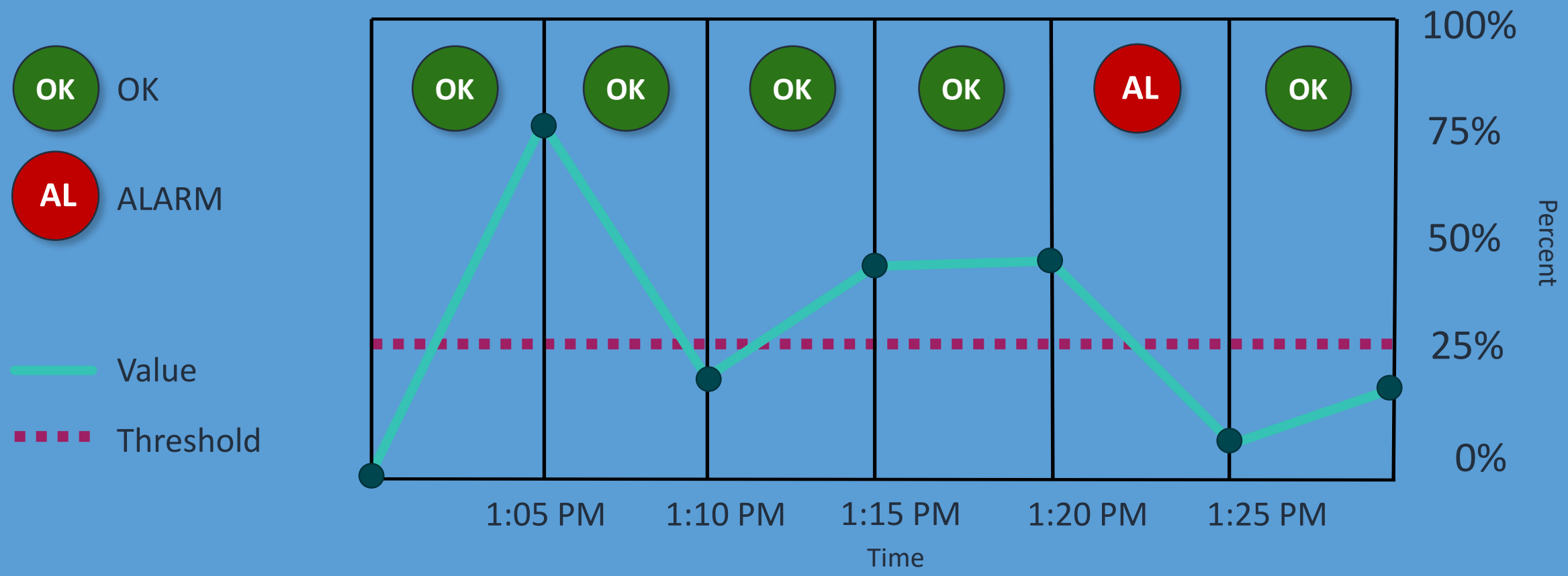
Threshold exceeded



Not enough information

Alarm components

Metric name	Instance ID	Statistic	Period	Datapoints to alarm
CPUUtilization	i-abcdef012345	Average	5 minutes	2 of 2



Amazon EventBridge

Amazon CloudWatch Events is now part of Amazon EventBridge.

EventBridge can:

- Send messages to respond to the environment.
- Activate functions or initiate actions.
- Capture state information.

Monitoring

Event-driven architectures

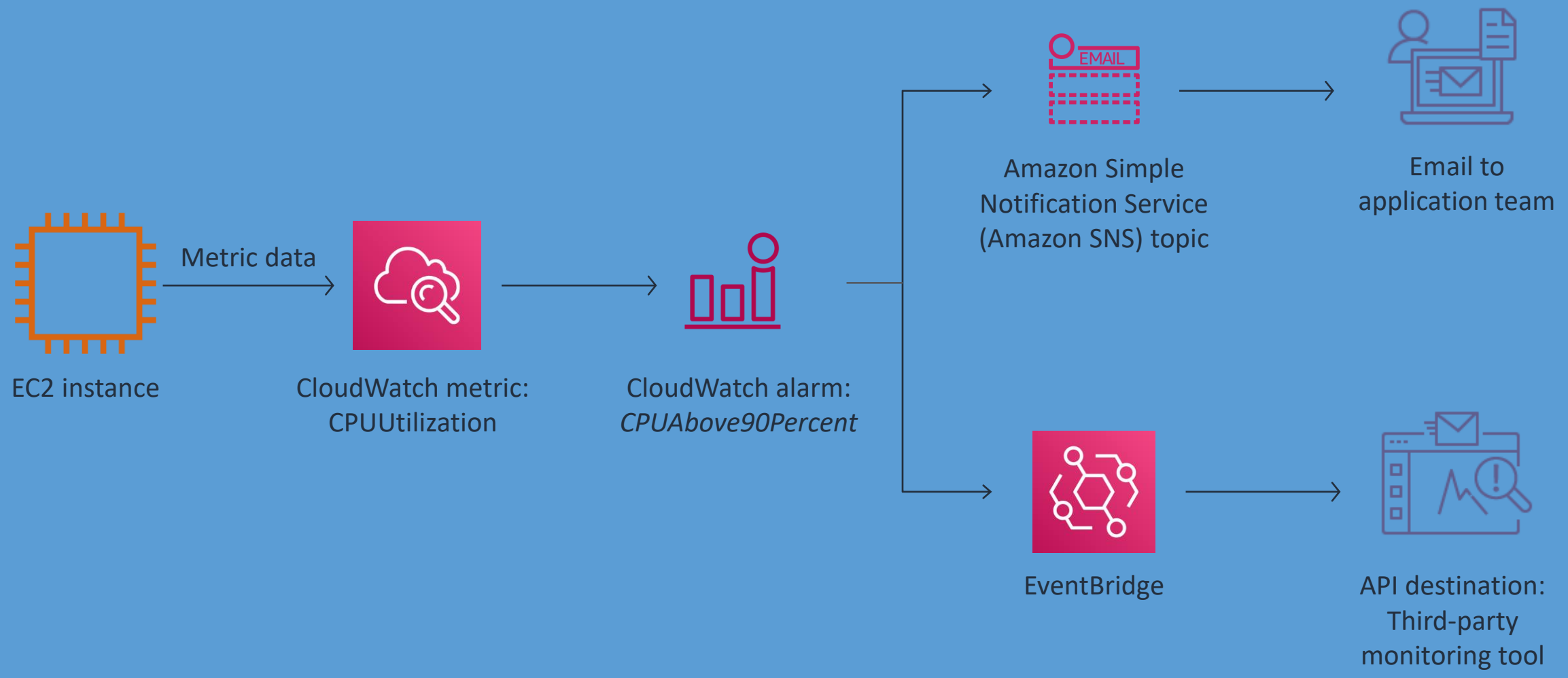
CloudWatch

Monitor AWS services, or your applications and systems.

EventBridge

Manage events and alarms. Initiate an automated workflow.

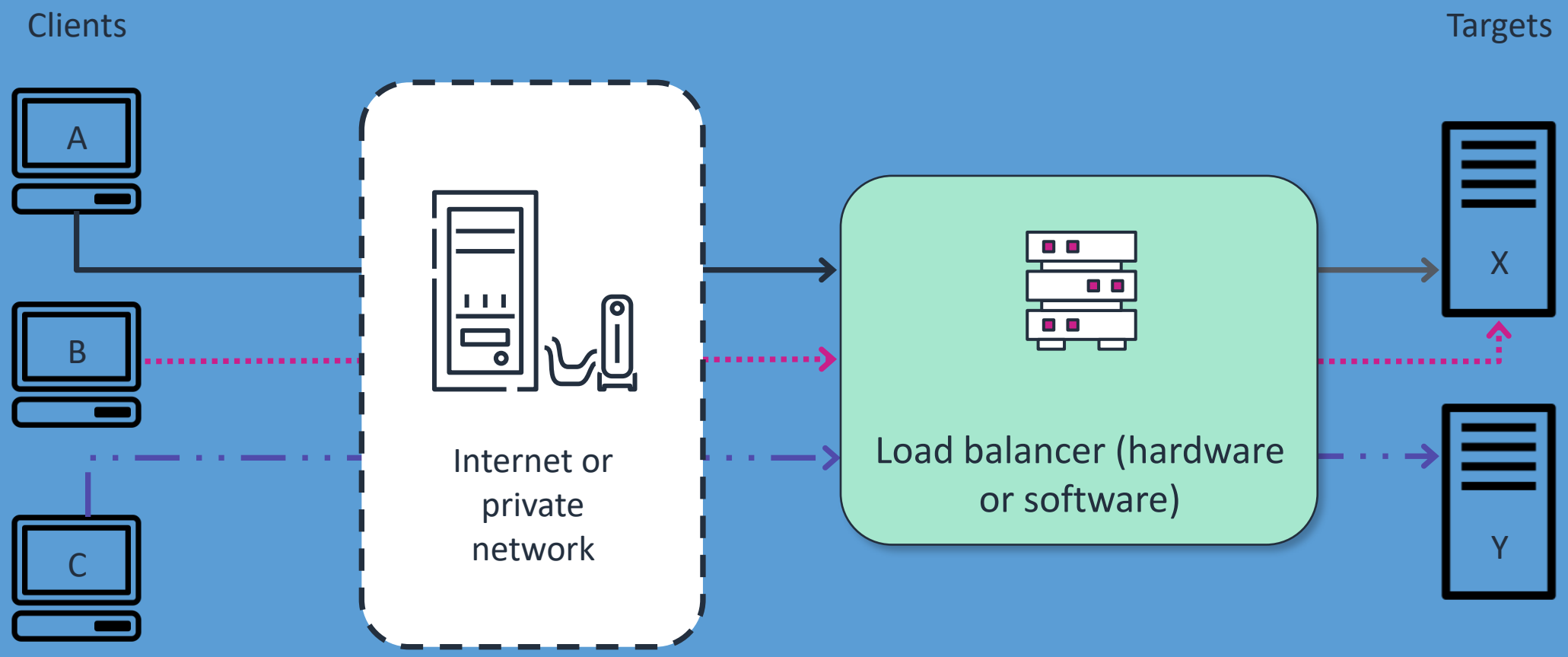
Example: CloudWatch alarm automated response



Load balancing

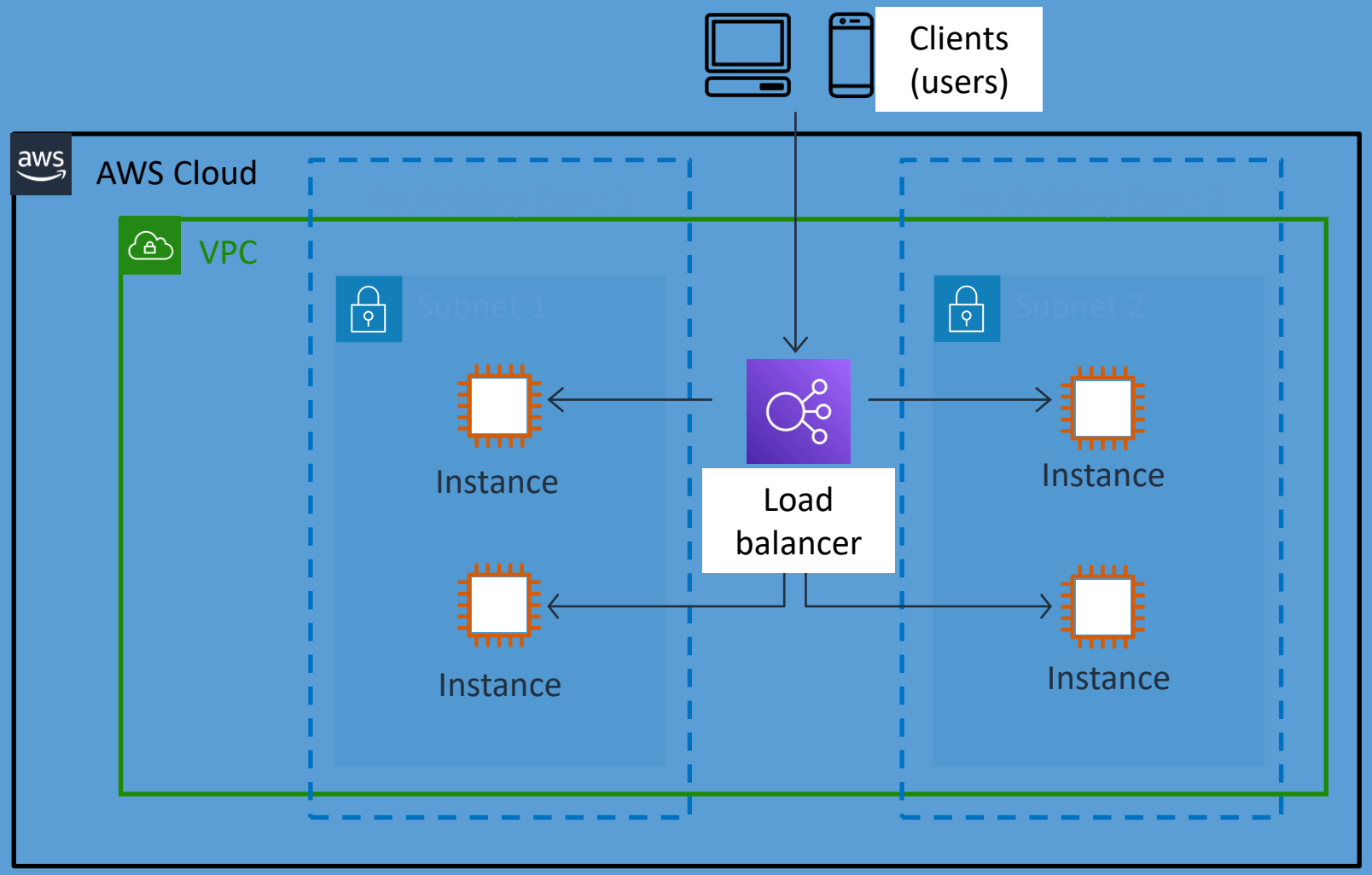
“How do we add high availability to our Amazon EC2 workloads and distribute traffic across multiple targets?”

Load balancers



Elastic Load Balancing (ELB)

- Automatically distributes traffic across multiple targets
- Provides high availability
- Incorporates security features
- Performs health checks



ELB load balancer types

Application Load Balancer

HTTP and HTTPS

Flexible application management

Advanced load balancing of traffic

Operates at the application layer (Layer 7)

Network Load Balancer

TCP and UDP

Extreme performance and static IP

Load balancing of TCP traffic

Operates at the transport layer (Layer 4)

Gateway Load Balancer

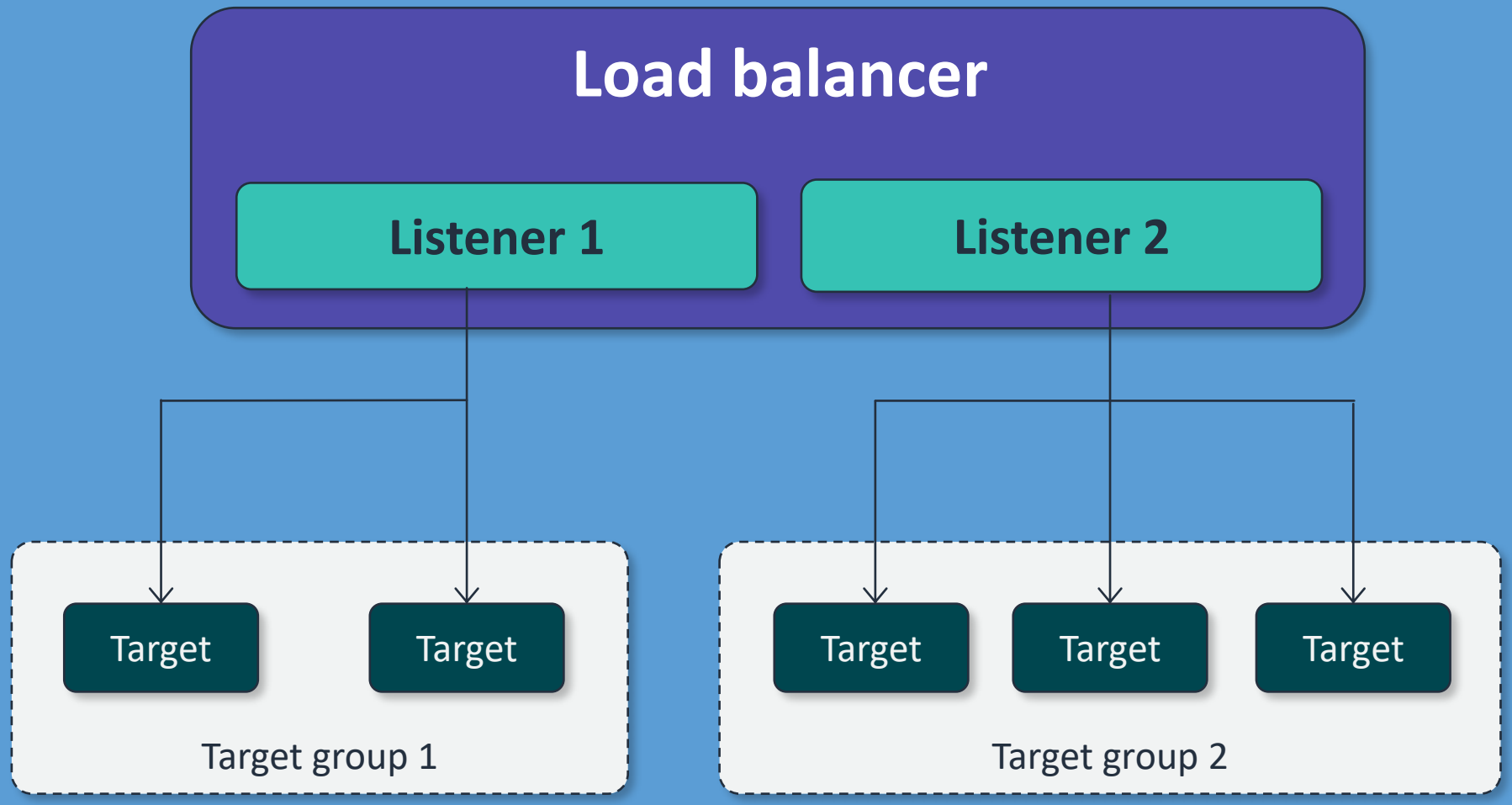
IP

Flexible application management

Advanced load balancing of traffic

Operates at the network layer (Layer 3)

ELB load balancer components



ELB common features

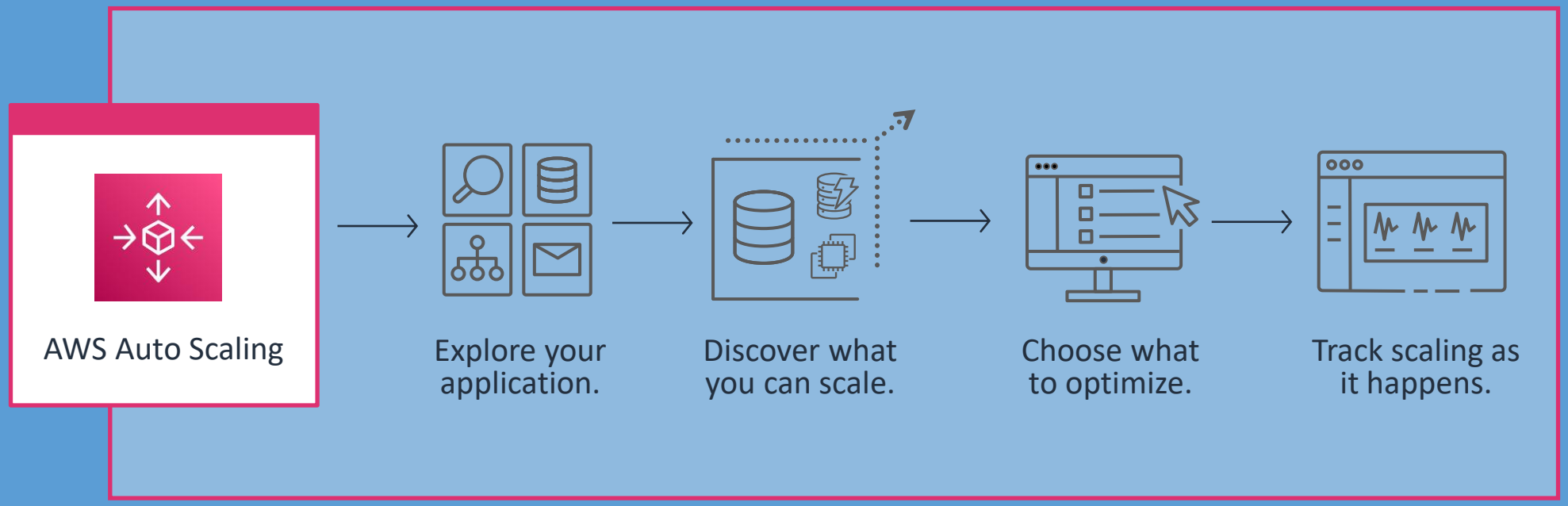
Feature	Application Load Balancer	Network Load Balancer	Gateway Load Balancer
Health checks	Yes	Yes	Yes
CloudWatch metrics	Yes	Yes	Yes
Logging	Yes	Yes	Yes
SSL offloading	Yes	Yes	
Connection draining	Yes	Yes	Yes
Preserve source IP address	Yes	Yes	Yes
Static IP address	**	Yes	
Lambda functions as a target	Yes		
Redirects	Yes		
Fixed-response actions	Yes		

Auto scaling

“How can we dynamically increase and decrease capacity to meet changing demand?”

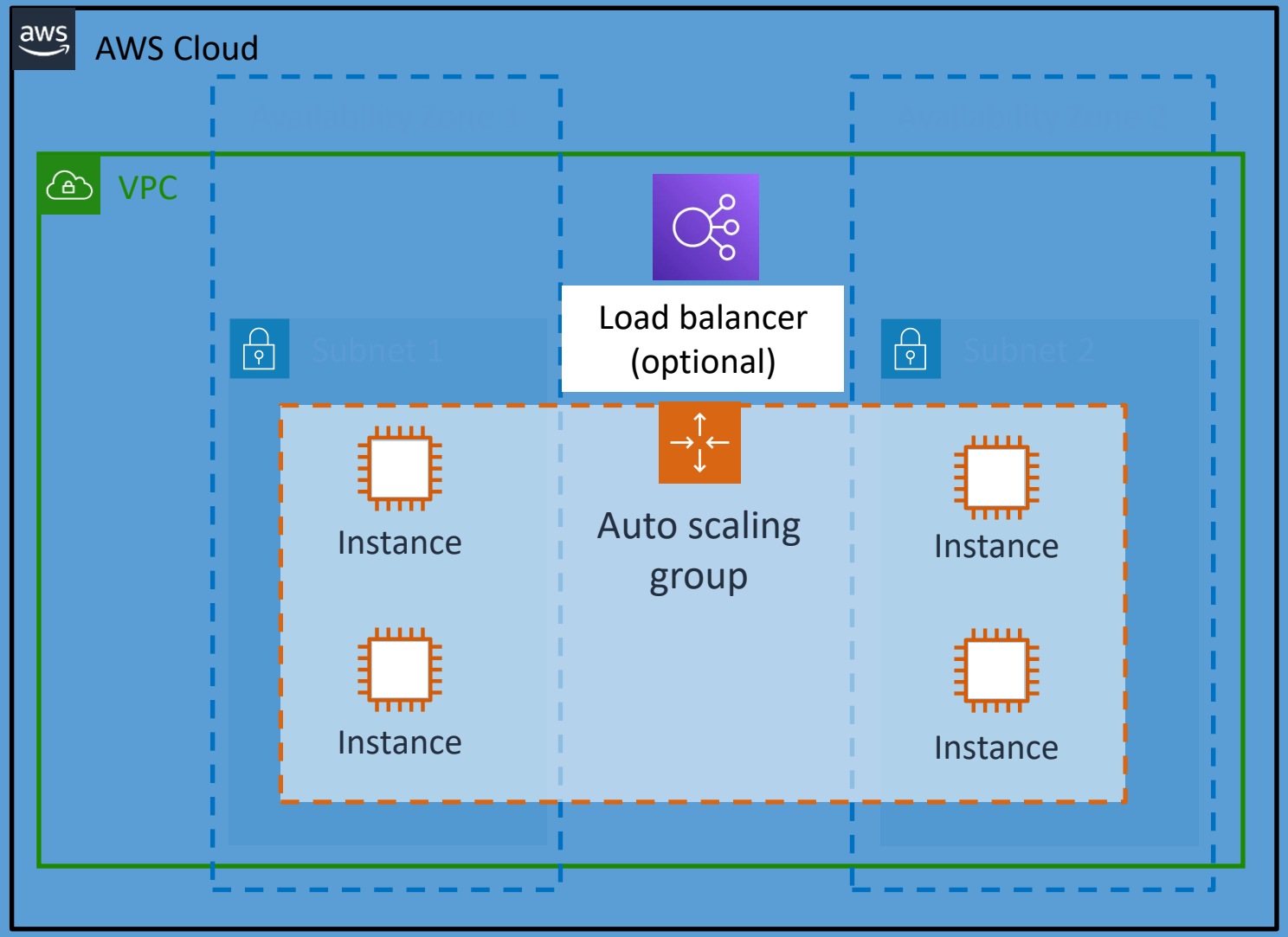
AWS Auto Scaling

Provides application scaling for multiple resources across services, in short intervals

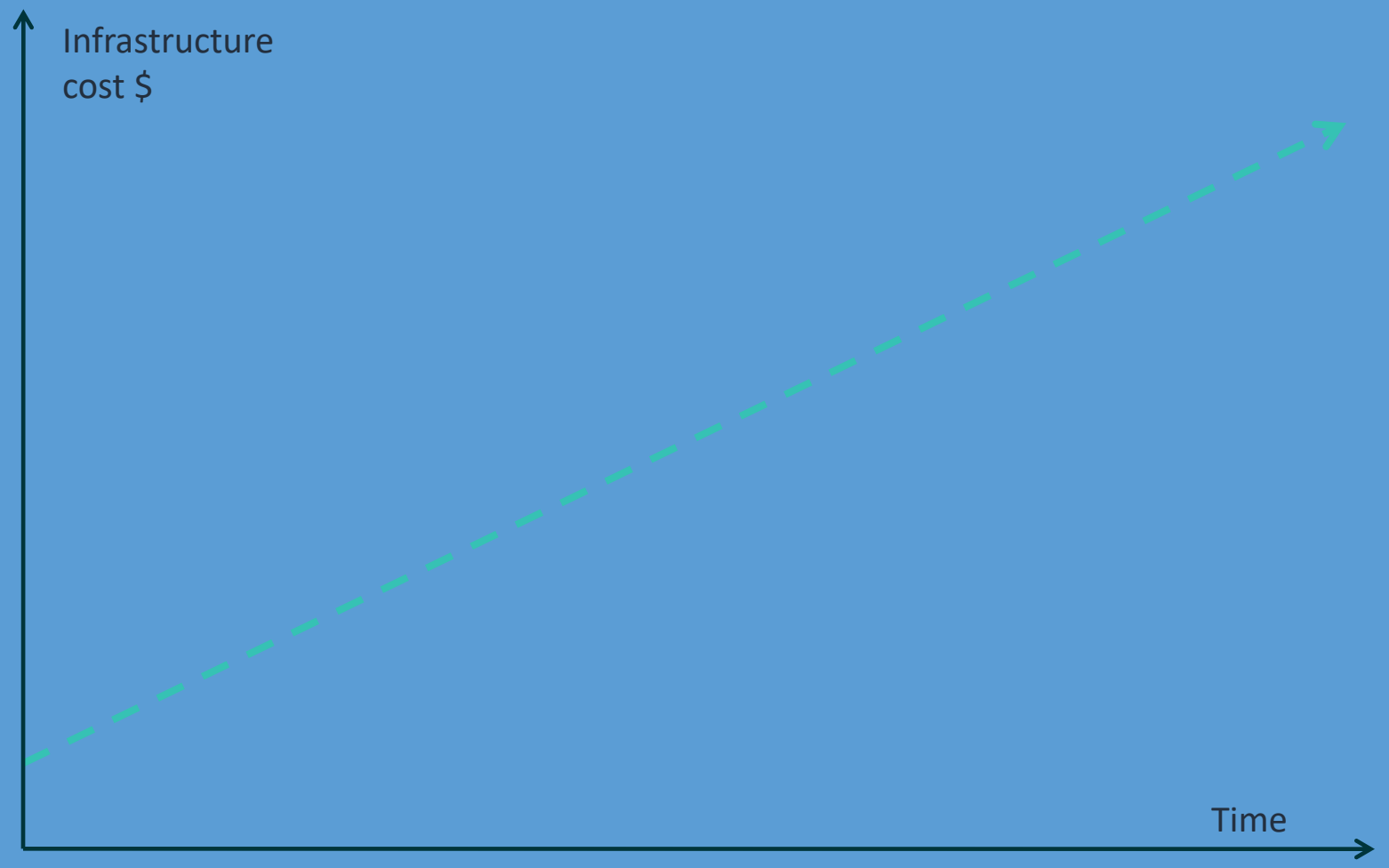


Amazon EC2 Auto Scaling

- Helps you control EC2 instances available to handle the load for your application
- Launches or terminates your AWS resources based on specified conditions
- Registers new instances with load balancers, when specified



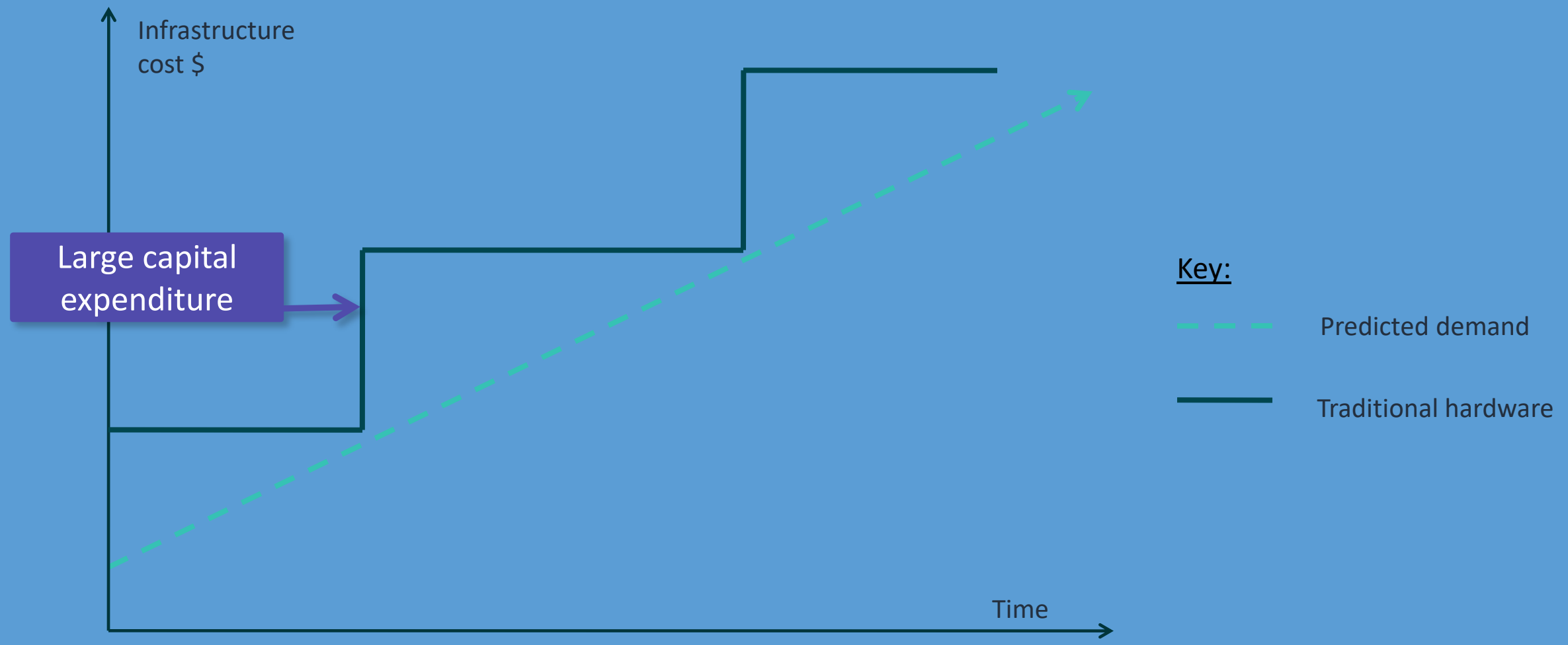
Elasticity: Scaling in and out (1 of 6)



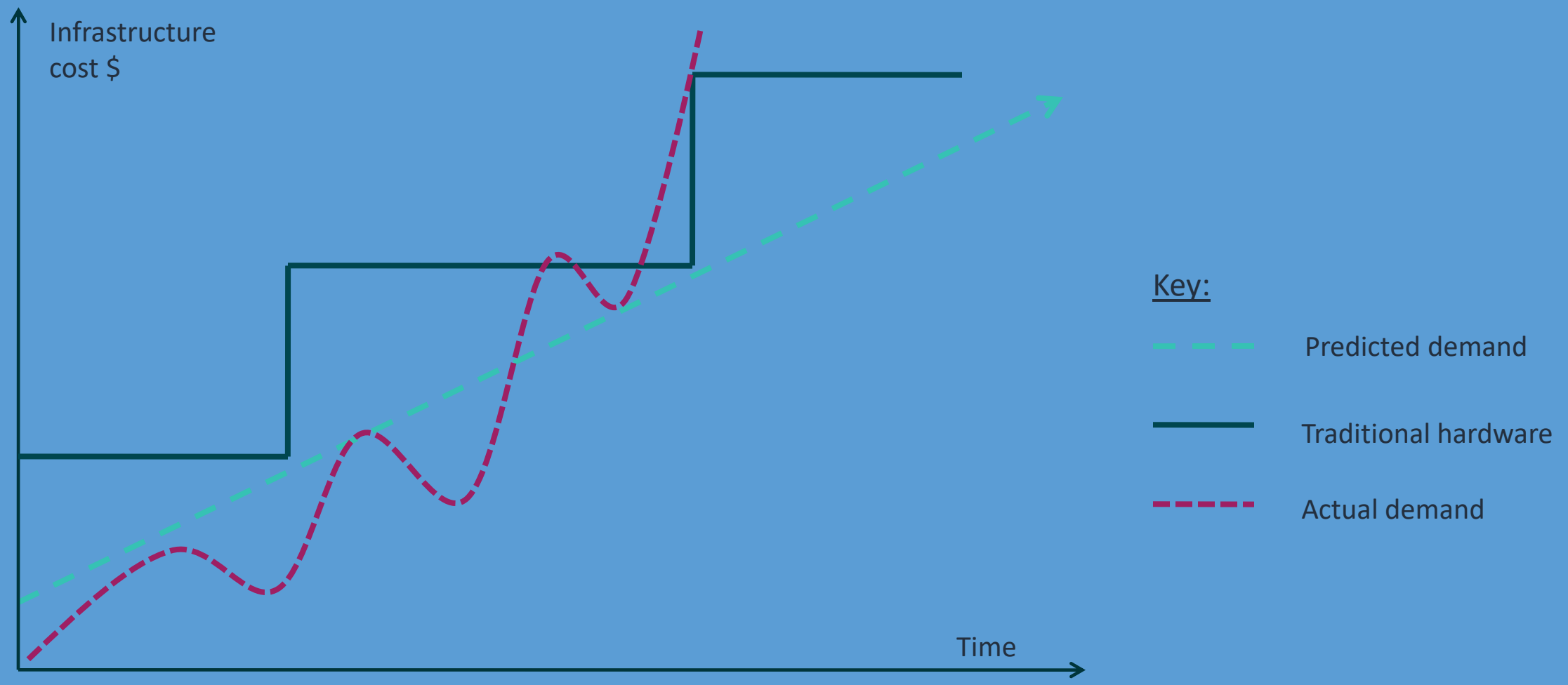
Key:

Predicted demand

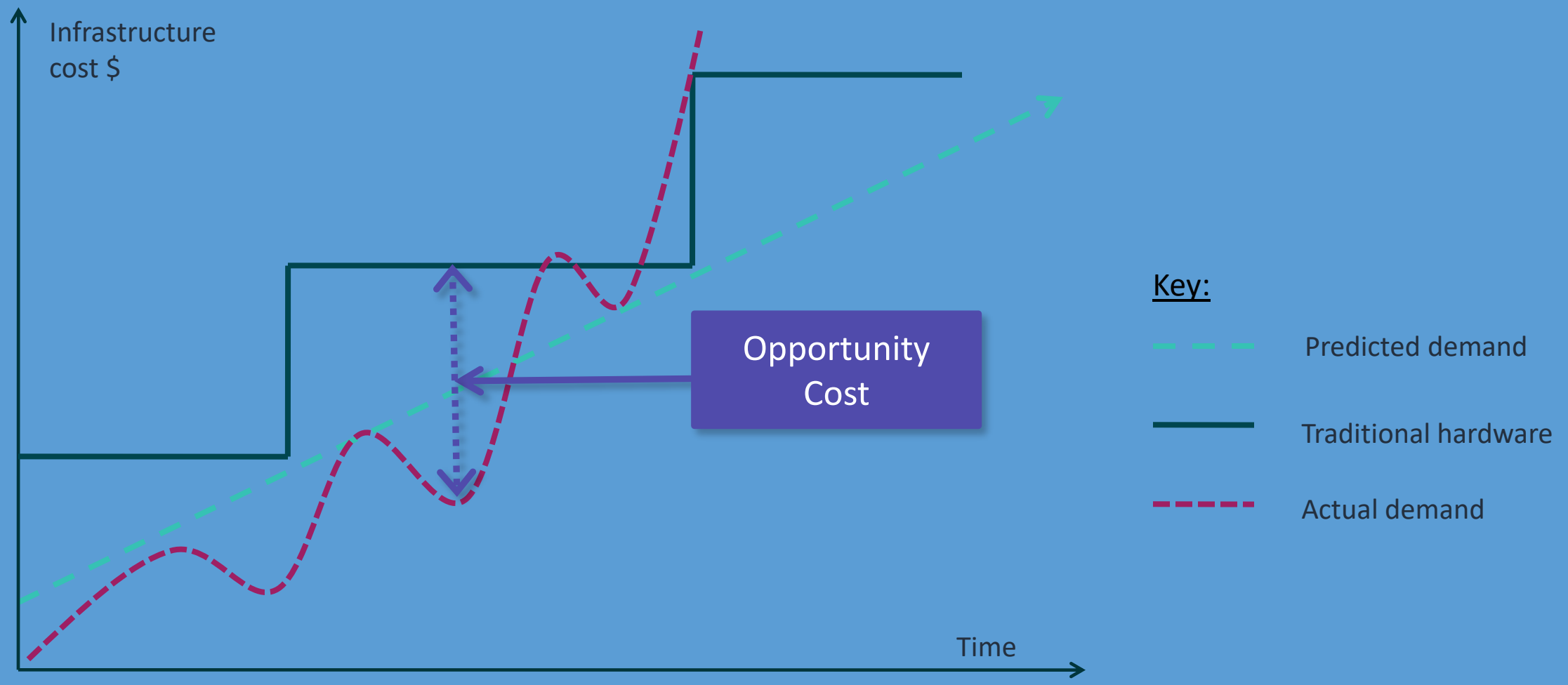
Elasticity: Scaling in and out (2 of 6)



Elasticity: Scaling in and out (3 of 6)

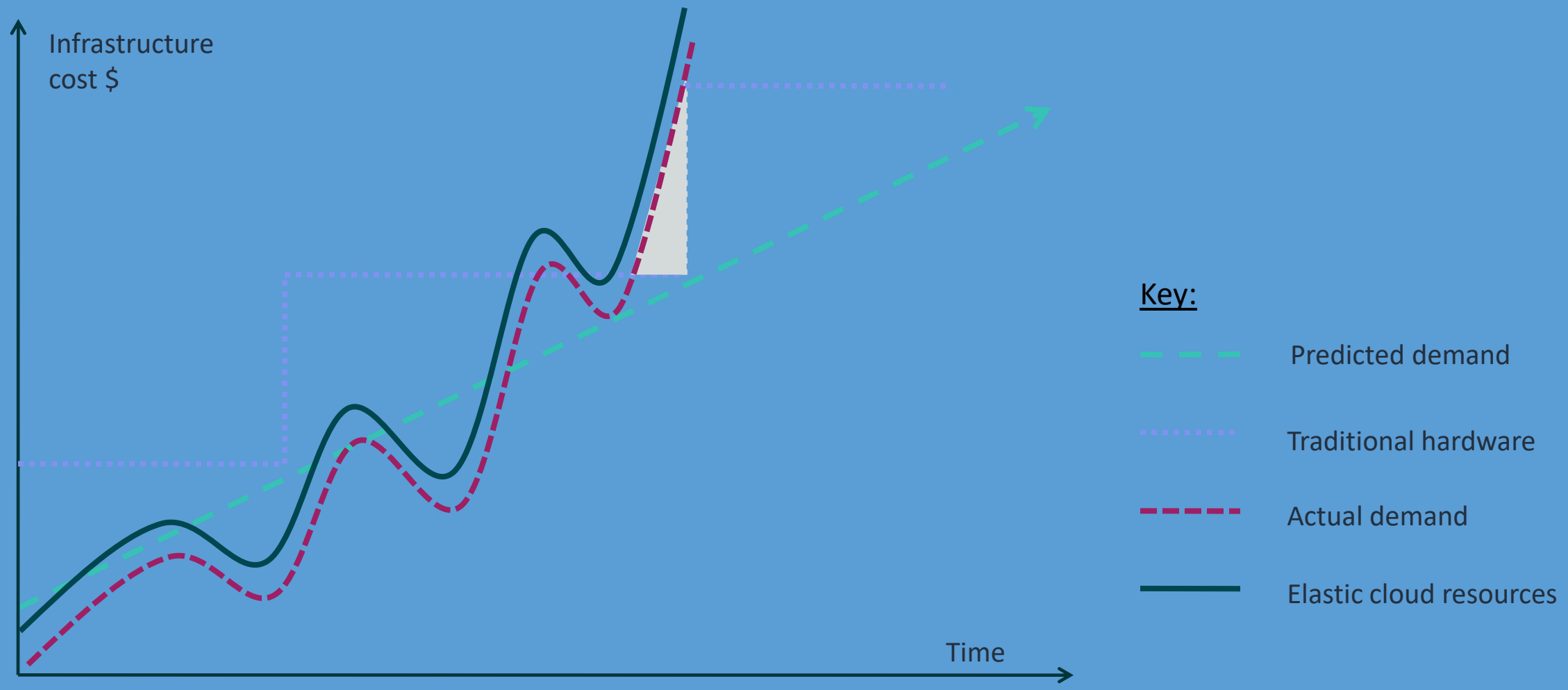


Elasticity: Scaling in and out (4 of 6)

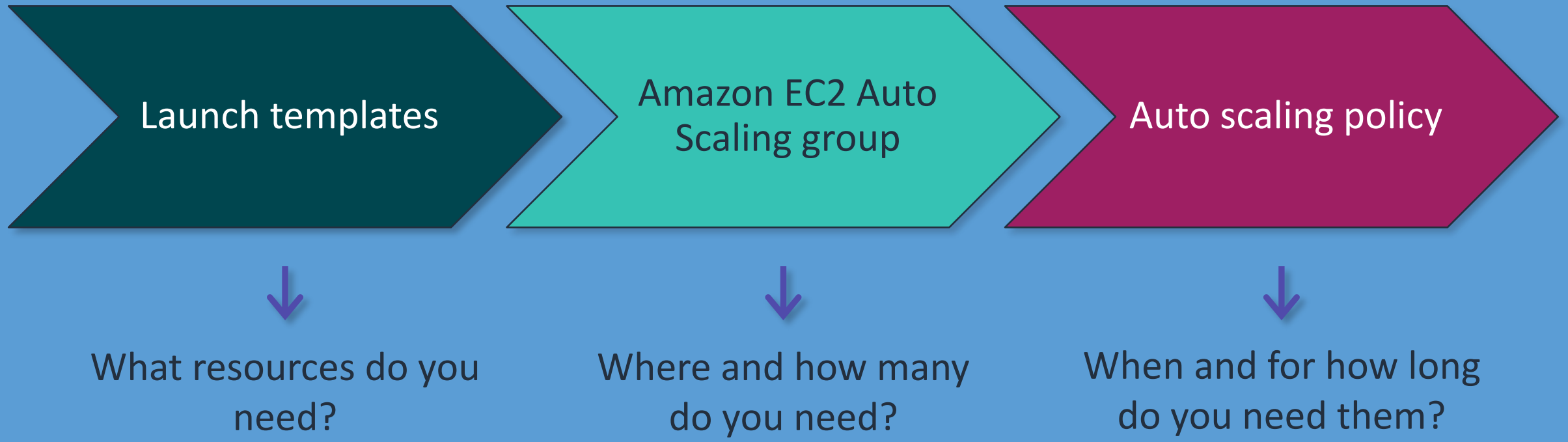


Elasticity: Scaling in and out (5 of 6)

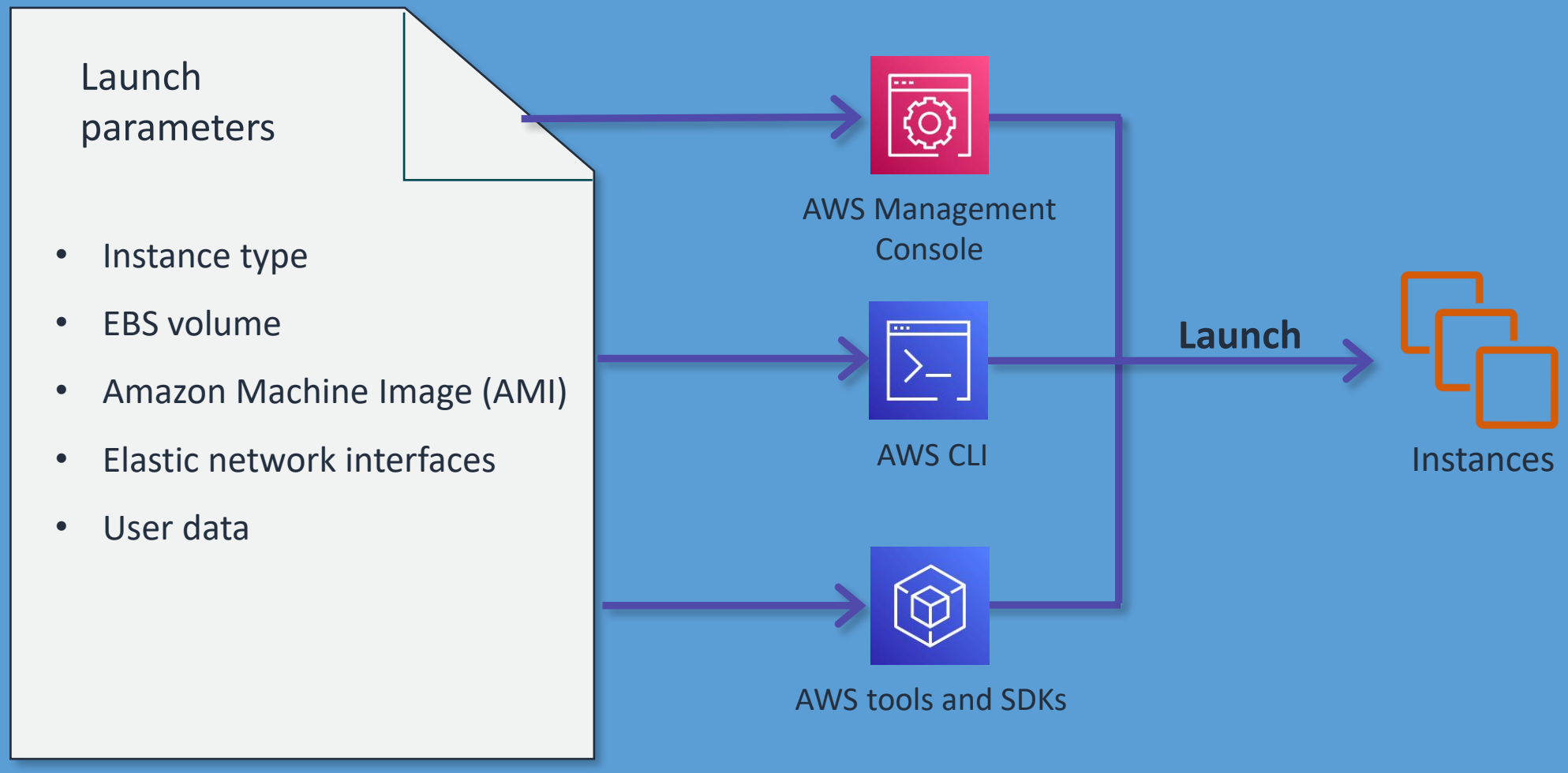
Elasticity: Scaling in and out (6 of 6)



Amazon EC2 Auto Scaling components

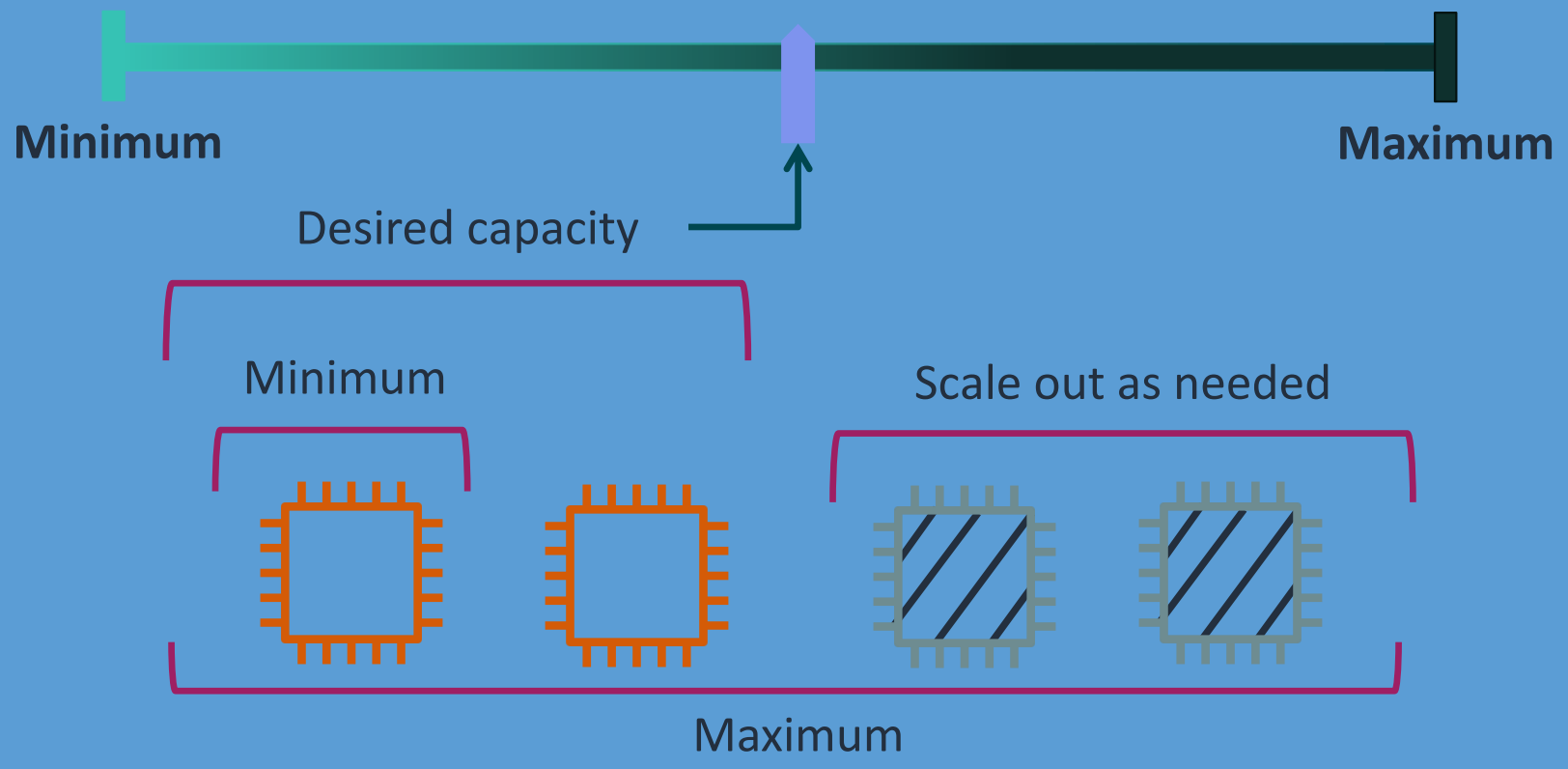


Launch template



Group capacity

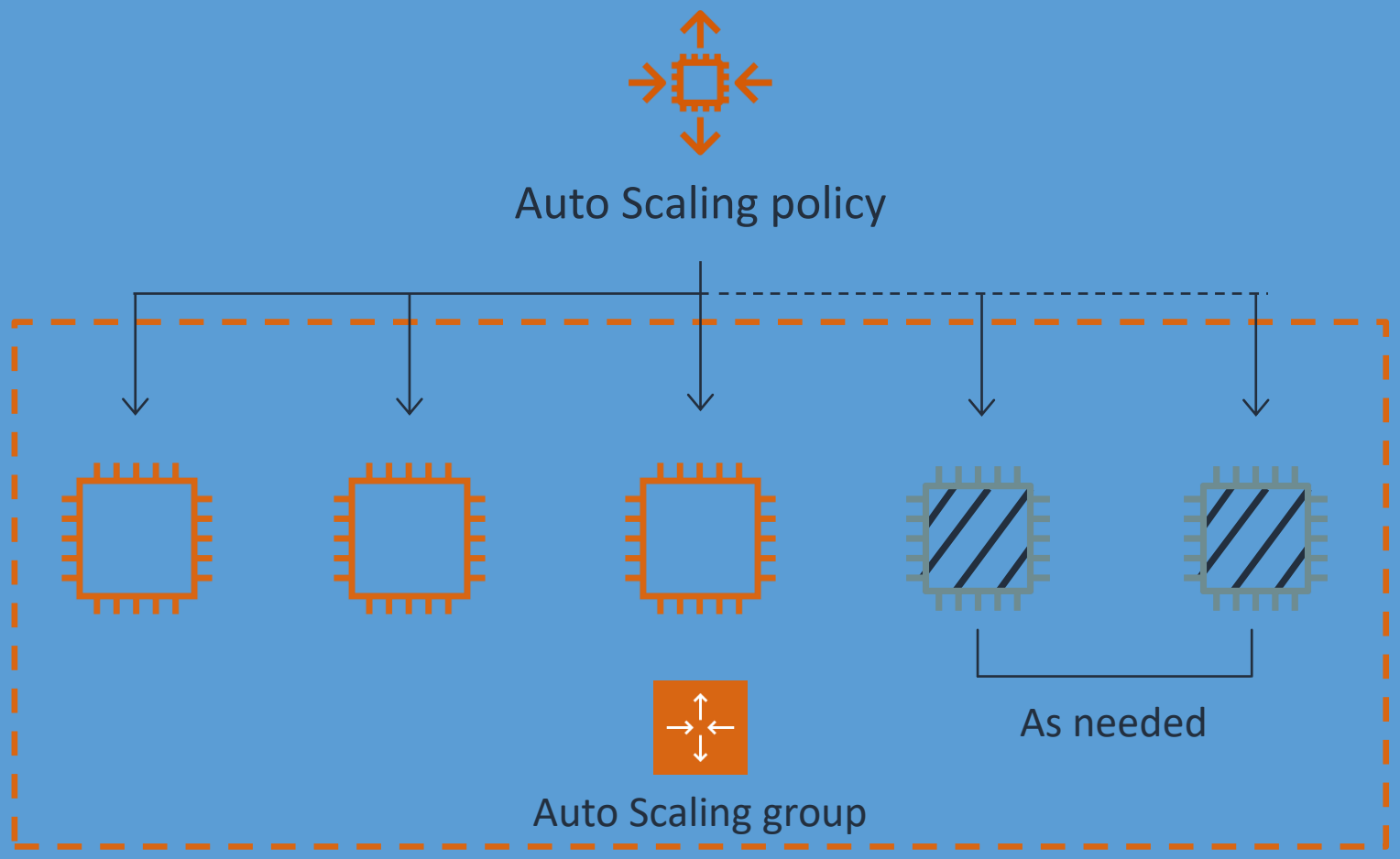
- Choose the VPC and subnets for your Amazon EC2 Auto Scaling group.
- Set minimum and maximum number of instances allowed.
- Launch or terminate instances to meet capacity demands.



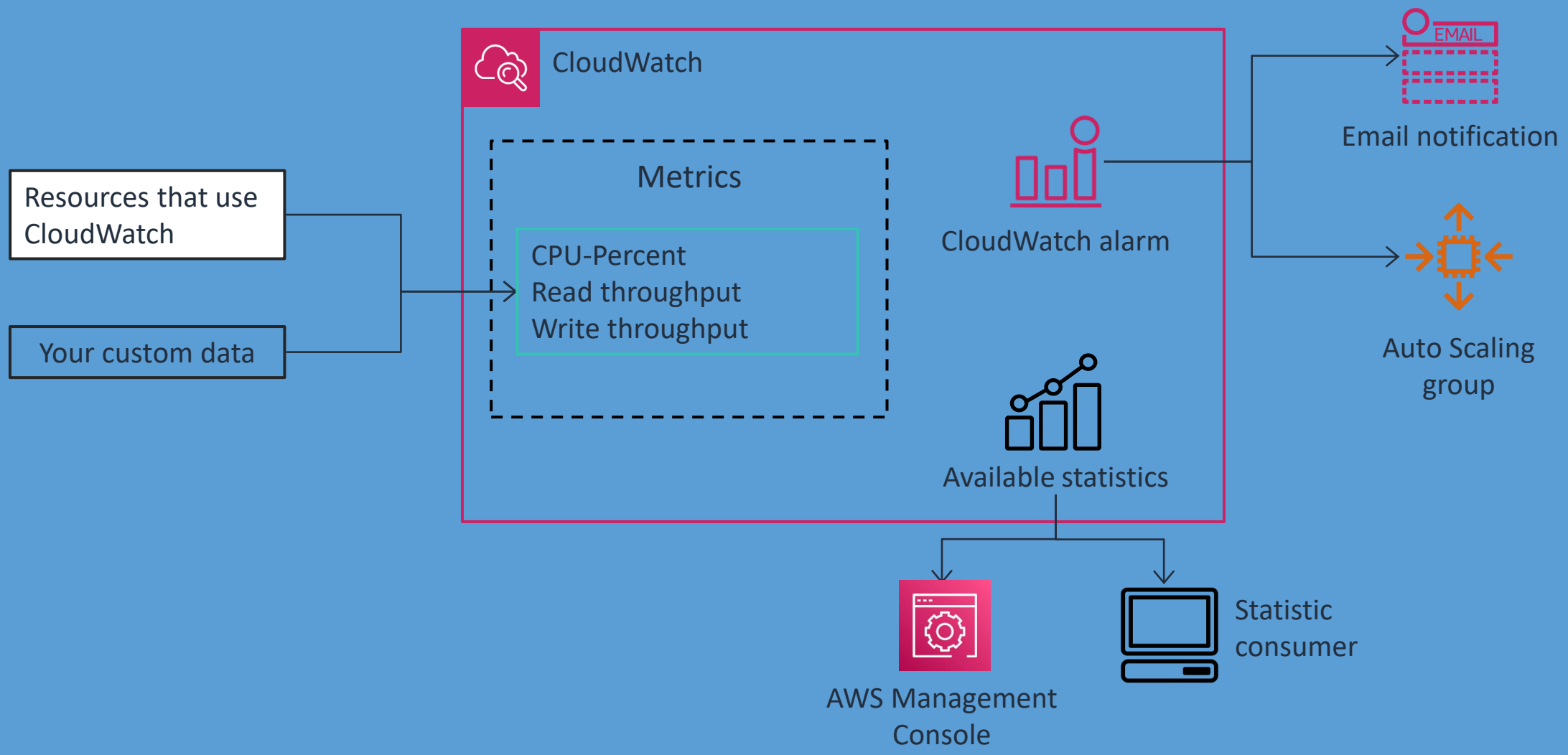
Invoke Amazon EC2 Auto Scaling

Invoke scaling with:

- Health status checks
- CloudWatch alarms
- Schedules
- Manual scaling



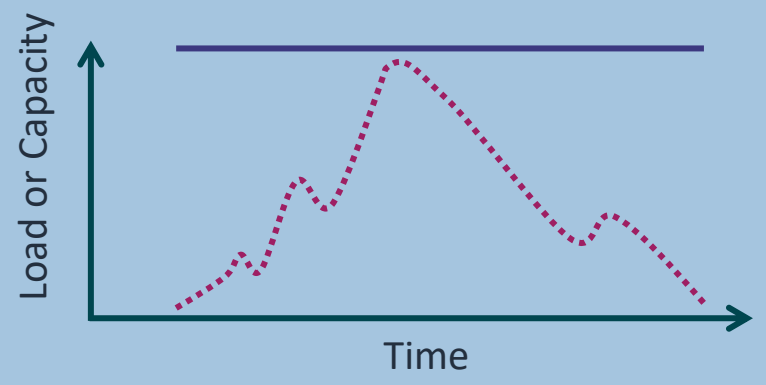
Invoke scaling with CloudWatch alarms



Ways to scale with EC2 Auto Scaling

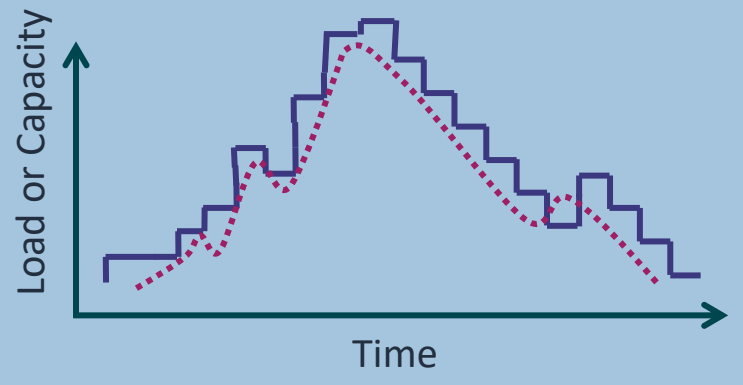
Scheduled

For predictable workloads



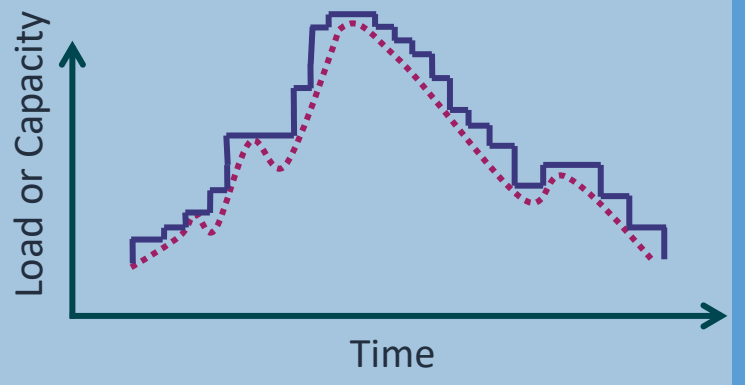
Dynamic

For general scaling



Predictive

Easiest to use

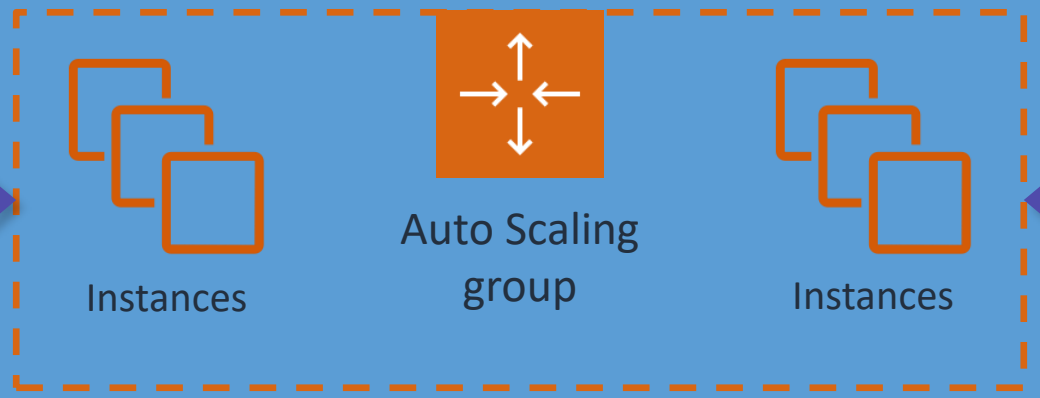


Provisioned capacity

Actual capacity demand

No need to manually adjust rules

Optimize cost with EC2 Auto Scaling



On-Demand Instances

Spiky workloads, to define needs

Savings Plans

Committed and steady-state usage

Spot Instances

Fault-tolerant, flexible, stateless workloads

Review

Present solutions

Operations Manager

Consider how you would answer the following:

- What tools and services are available to monitor and log activity in our AWS accounts?
- How can we set thresholds and be alerted to changes in our infrastructure?
- How do we add high availability to our Amazon EC2 workloads and distribute traffic across multiple targets?
- How can we dynamically increase and decrease capacity to meet changing demand?

Module review

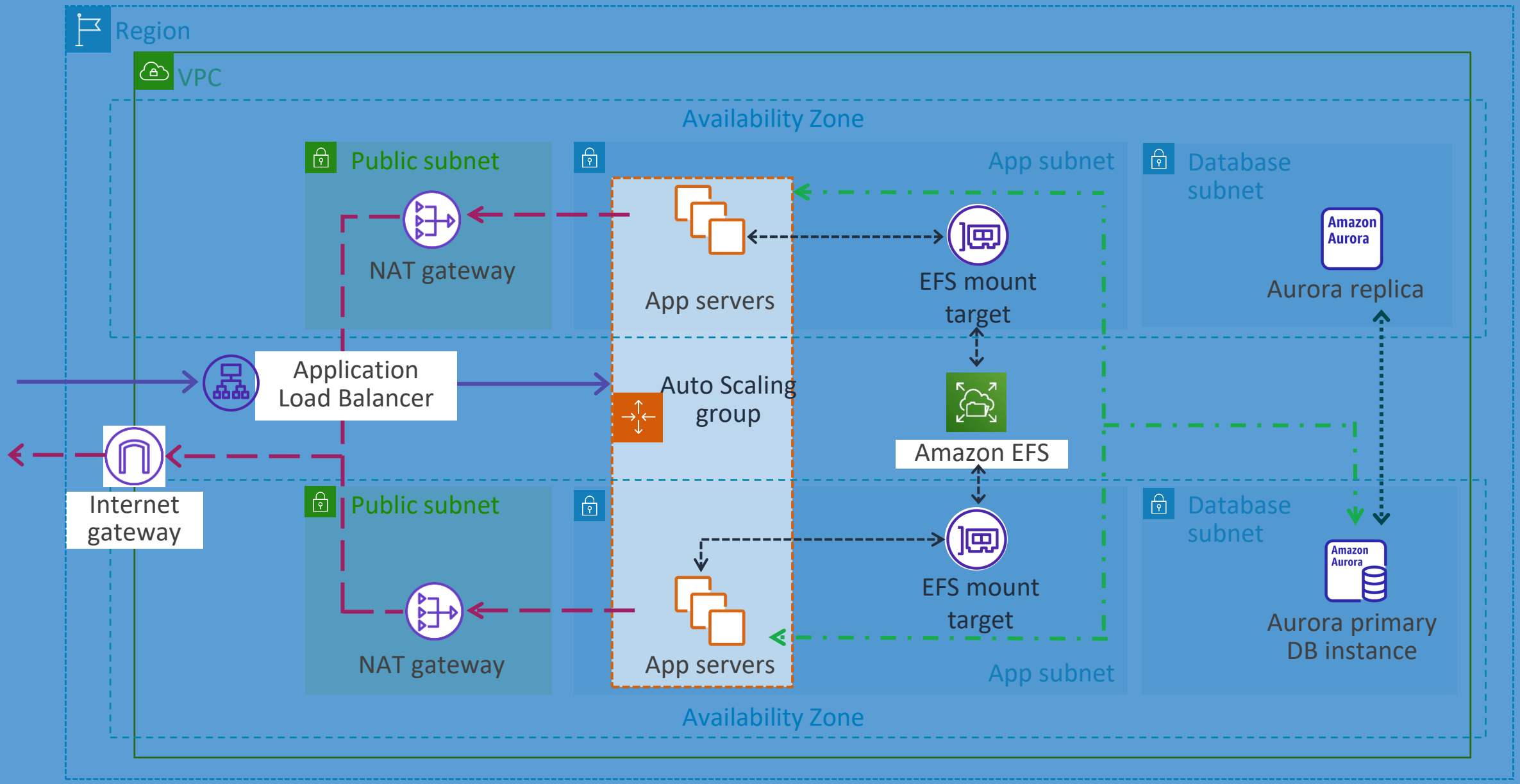
In this module you learned about:

- ✓ Monitoring
- ✓ Alarms and events
- ✓ Load balancing
- ✓ Auto scaling

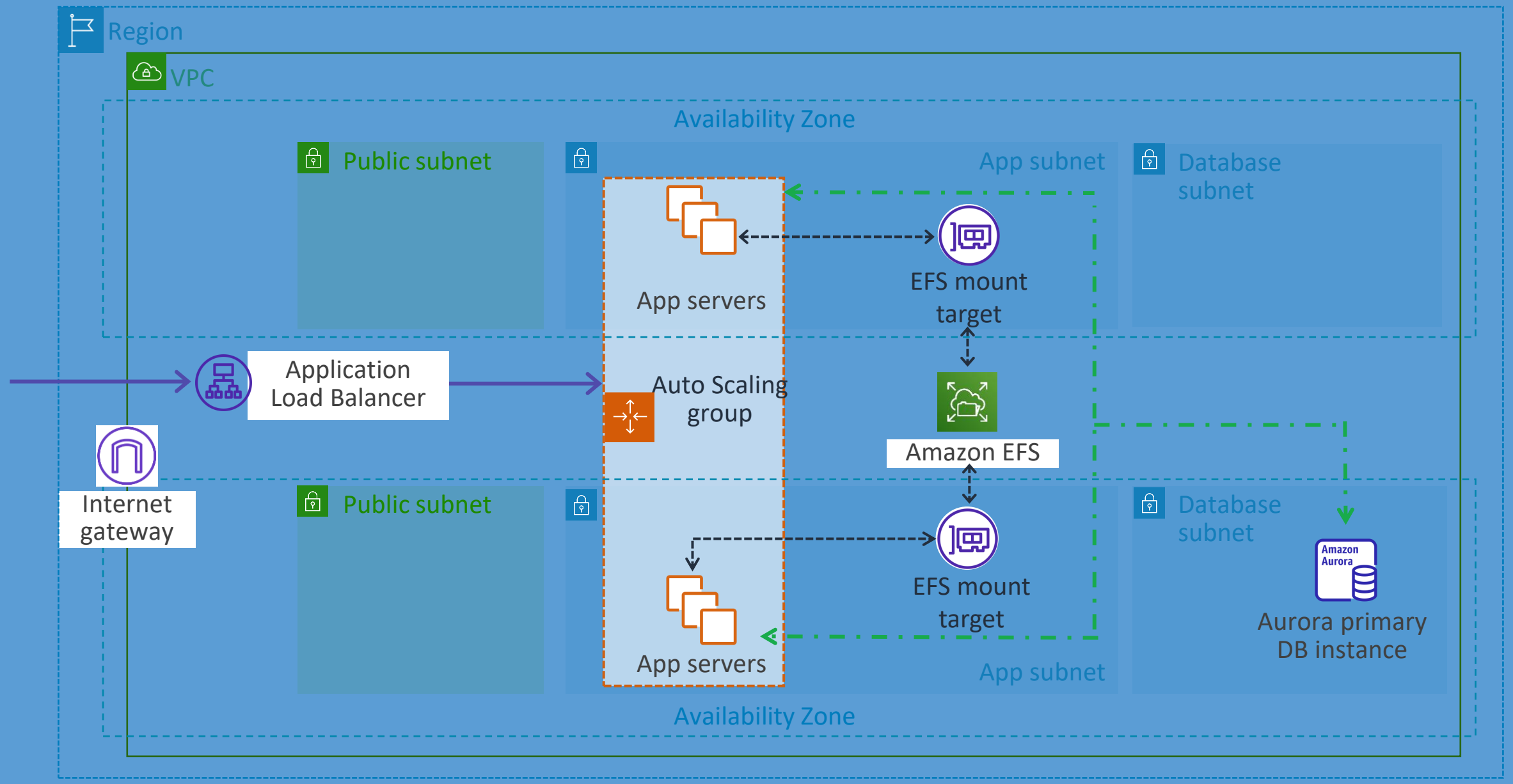
Next, you will review:

- Capstone check-in
- Knowledge check
- Lab introduction

Capstone architecture



Capstone architecture check-in



Knowledge check

Knowledge check question 1

Which of these is a valid target for an Application Load Balancer?

A	Amazon EC2 instance
B	An Availability Zone
C	An Amazon S3 bucket
D	VPN connection

Knowledge check question 1 and answer

Which of these is a valid target for an Application Load Balancer?

A correct	Amazon EC2 instance
B	An Availability Zone
C	An Amazon S3 bucket
D	VPN connection

Knowledge check question 2

You have an application with unpredictable traffic patterns that runs on at least two instances. You want the CPU utilization to stay at about 75 percent. Which Amazon EC2 Auto Scaling strategy should you choose?

A	Scheduled
B	Dynamic
C	Predictive
D	Manual

Knowledge check question 2 and answer

You have an application with unpredictable traffic patterns that runs on at least two instances. You want the CPU utilization to stay at about 75 percent. Which Amazon EC2 Auto Scaling strategy should you choose?

A	Scheduled
B correct	Dynamic
C	Predictive
D	Manual

Knowledge check question 3

What service can invoke actions based on data from account resources and supported third-party management services?

A	CloudWatch Logs
B	EventBridge
C	CloudTrail
D	Amazon EC2 Auto Scaling

Knowledge check question 3 and answer

What service can invoke actions based on data from account resources and supported third-party management services?

A	CloudWatch Logs
B correct	EventBridge
C	CloudTrail
D	Amazon EC2 Auto Scaling

Knowledge check question 4

Which of the following are valid alarm states in CloudWatch? (Select TWO.)

A	READY
B	ALERT
C	ALARM
D	INSUFFICIENT_DATA
E	FAILED

Knowledge check question 4 and answer

Which of the following are valid alarm states in CloudWatch? (Select TWO.)

A	READY
B	ALERT
C correct	ALARM
D correct	INSUFFICIENT_DATA
E	FAILED

Knowledge check question 5

Which of the following are use cases for CloudTrail data? (Select TWO.)

A	Provide real-time observability of AWS resources.
B	Store log data as a record of account usage.
C	Log events for a particular service or application.
D	Capture root login failures.
E	Collect metric data measuring CPU utilization.

Knowledge check question 5 and answer

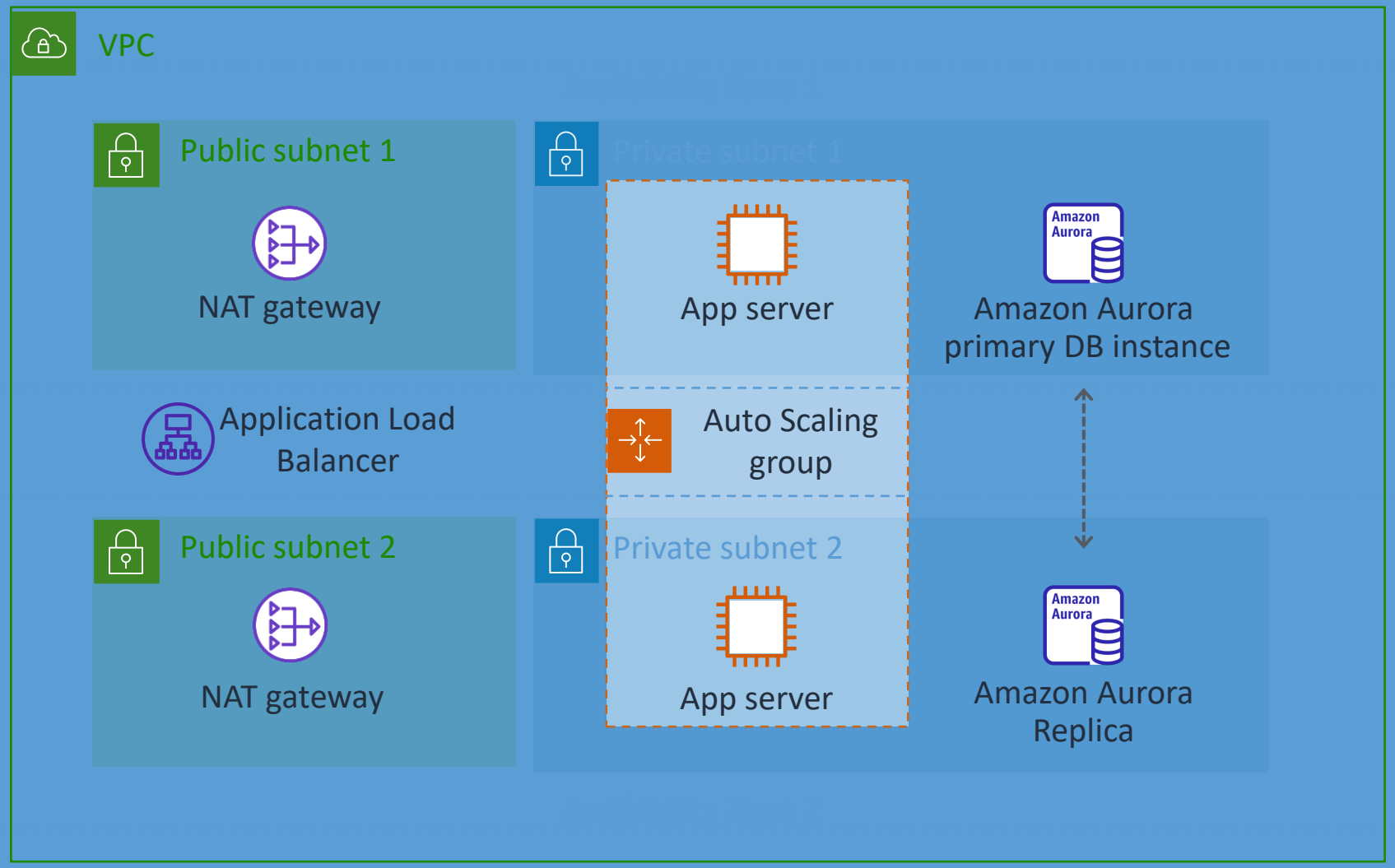
Which of the following are use cases for CloudTrail data? (Select TWO.)

A	Provide real-time observability of AWS resources.
B correct	Store log data as a record of account usage.
C	Log events for a particular service or application.
D correct	Capture root login failures.
E	Collect metric data measuring CPU utilization.

Lab 4:

Configure high availability in your Amazon VPC

Lab 4 diagram



Lab tasks

Task 1: Inspect your existing lab environment.

Task 2: Create a launch template.

Task 3: Create an Auto Scaling group.

Task 4: Test your application.

Task 5: Test high availability of your application tier.

Task 6: Configure high availability of the database tier.

Task 7: Make the NAT gateway highly available.

Task 8: Test the high availability of the Aurora database.

AWS

Automation

Question

Have you used any of these resource provisioning tools?

A. AWS CloudFormation

B. AWS Elastic Beanstalk

C. Others (for example, Terraform or OpenStack Heat)

D. Not yet

Module overview

- Business requests
- AWS CloudFormation
- Infrastructure management
- Present solutions
- Knowledge check

Business requests

Chief Technology Officer

The chief technology officer wants to know:

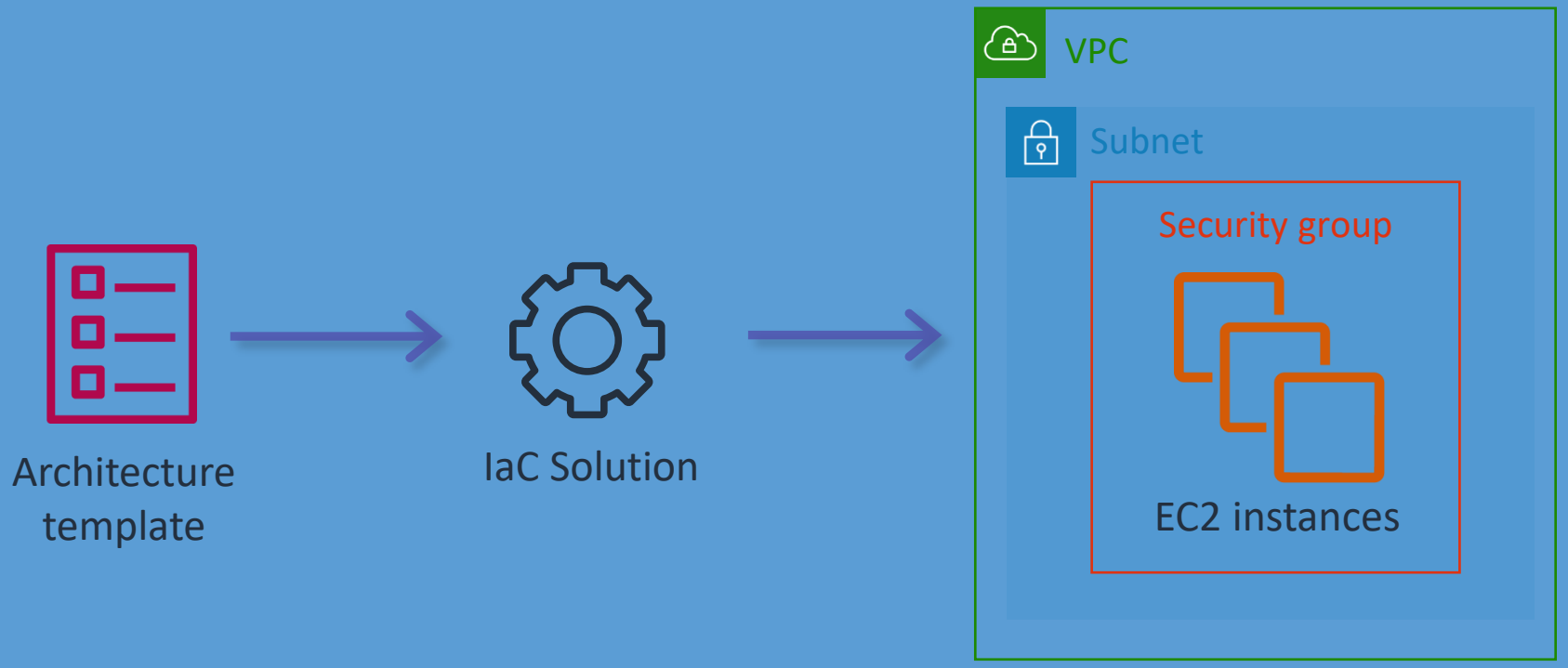
- How can we simplify our cloud infrastructure build?
- How can we deploy, maintain, and scale applications in the cloud?

AWS CloudFormation

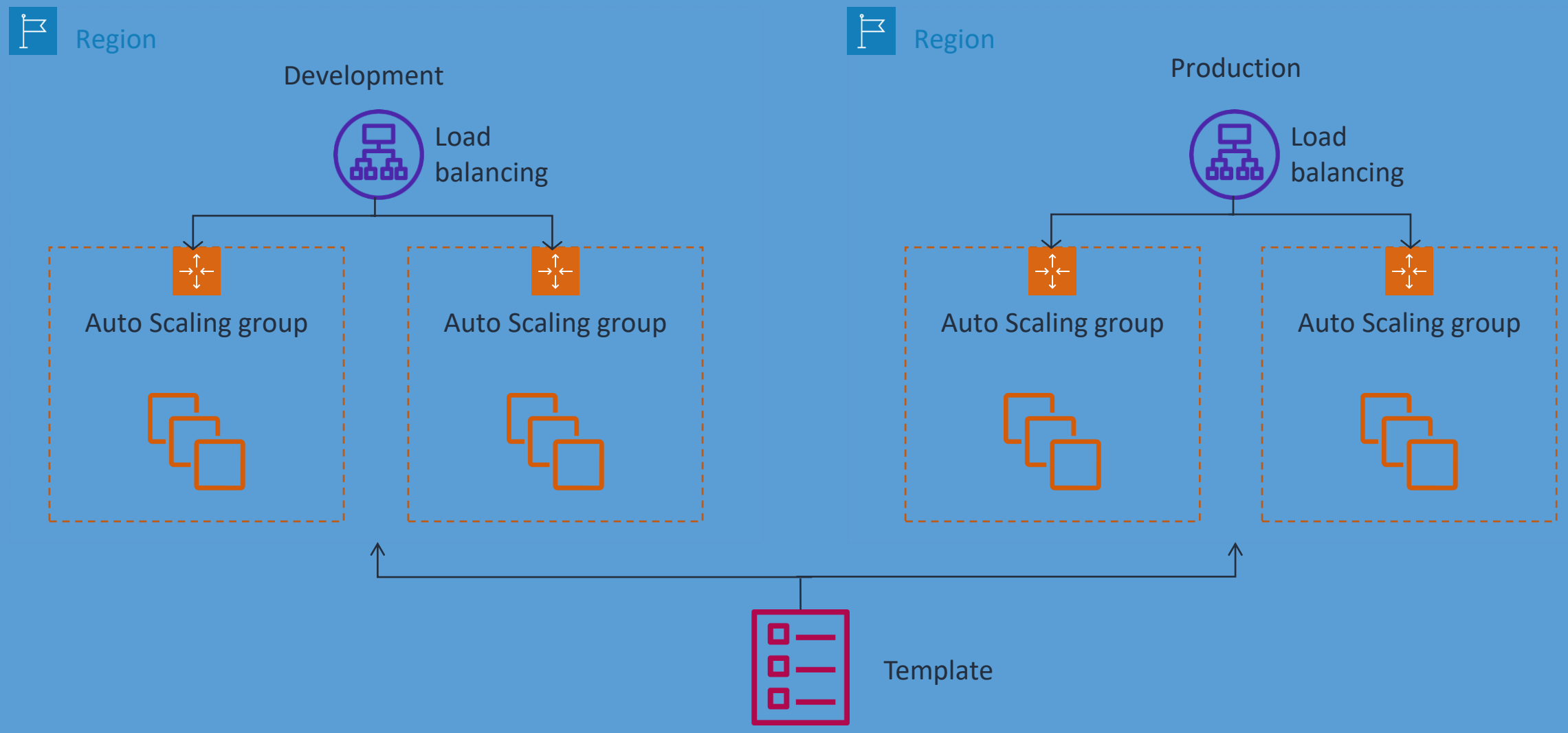
“How can we simplify our cloud infrastructure build?”

Infrastructure as code (IaC)

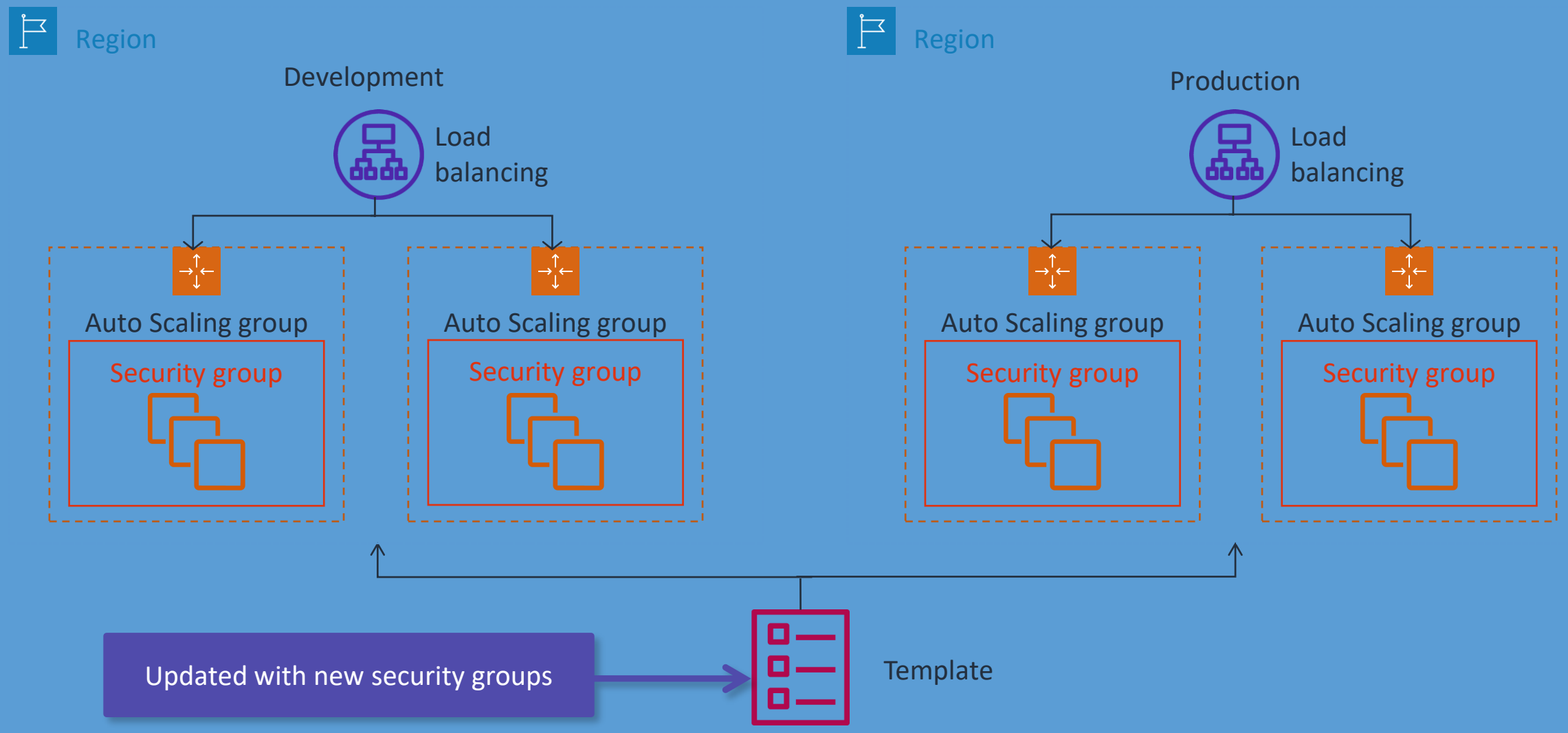
- Replicate, redeploy, and repurpose.
- Control versioning on infrastructure and applications.
- Detect drift.
- Roll back the service to the last good state.



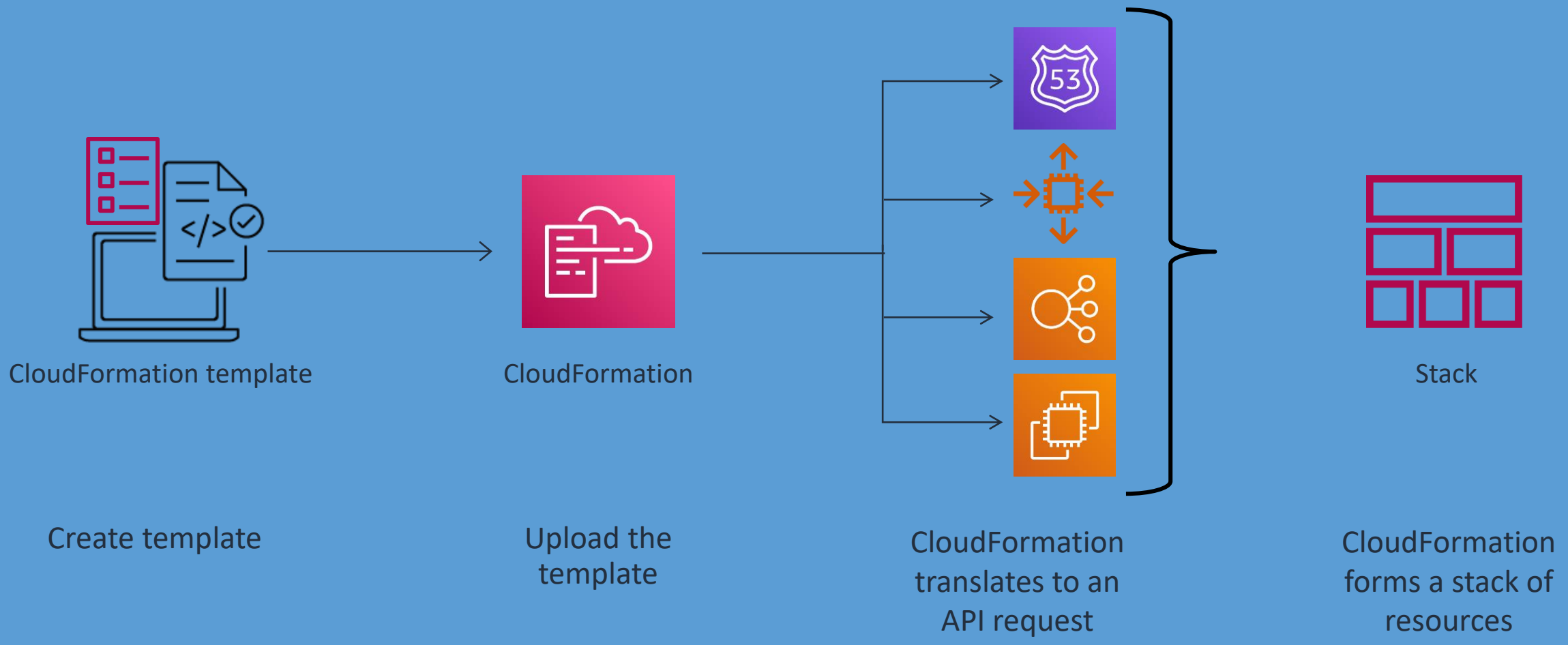
Benefits of IaC – Reusability



Benefits of IaC – Updates

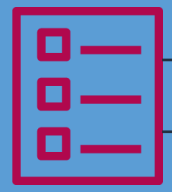


AWS CloudFormation



Understanding CloudFormation

- Written as JSON or YAML
- Describes the resources to be created or modified
- Treated as source code:
  - Code reviewed
  - Version controlled



JSON

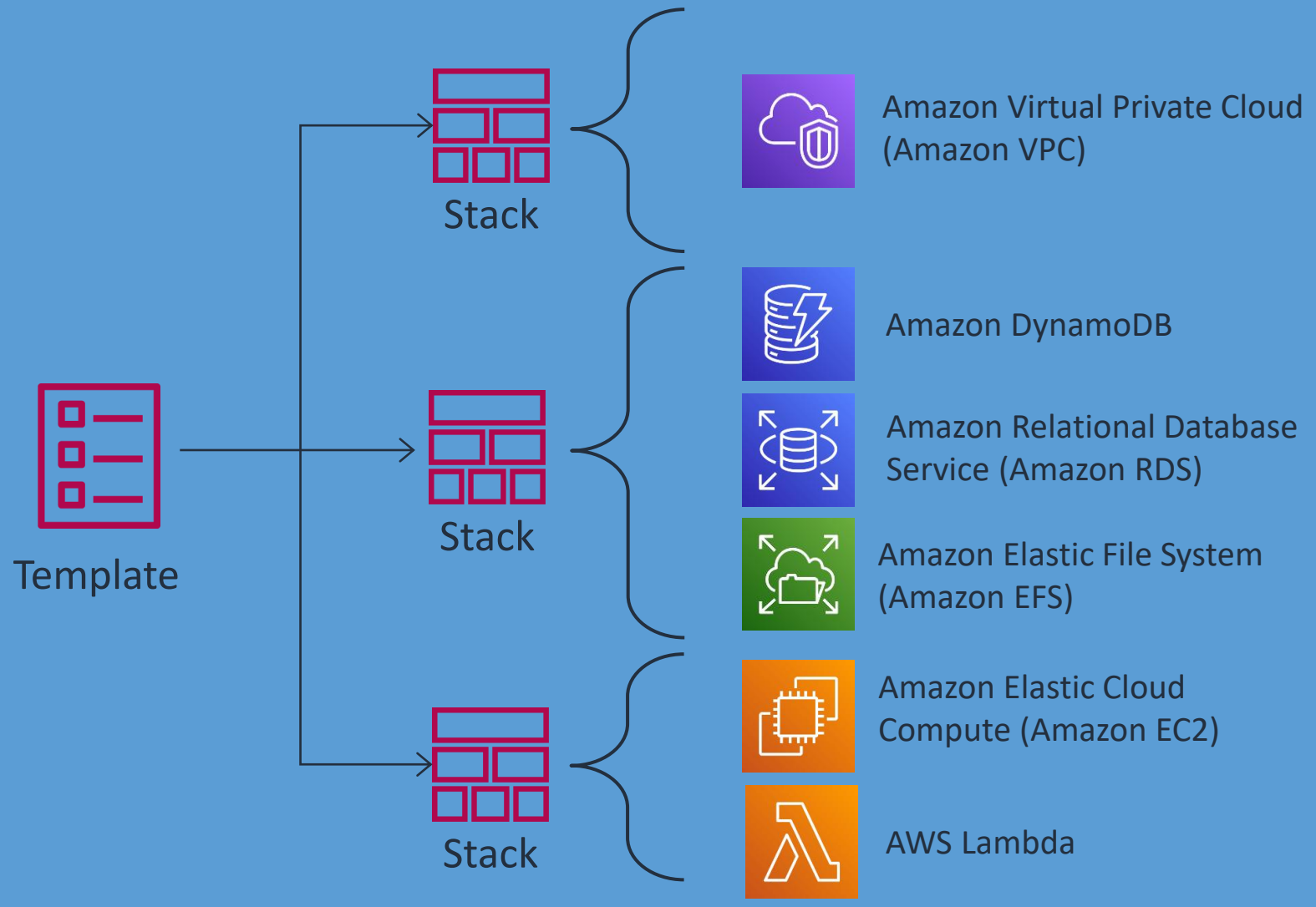
```
{
  "Resources" : {
    "HelloBucket" : {
      "Type" : "AWS::S3::Bucket"
    }
  }
}
```

YAML

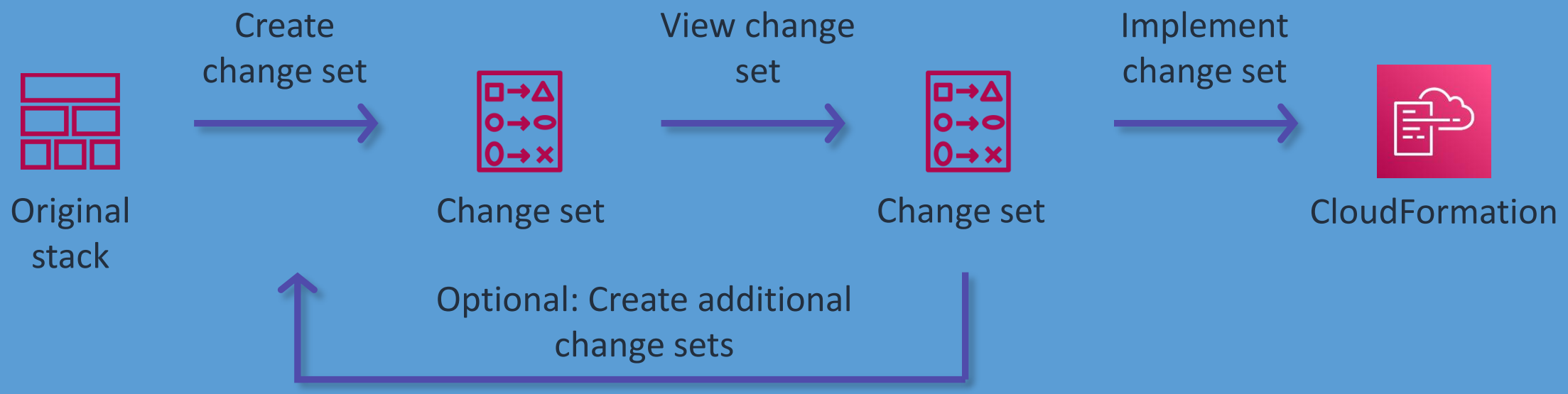
```
Resources:
 HelloBucket:
  Type: AWS::S3::Bucket
```

Stacks

- A collection of AWS resources managed as a single unit
- Can deploy and delete resources as a unit
- Can update resources and settings on running stacks
- Supports nested stacks and cross-stack references



Change sets


Create change set
View change set
Implement change set
Original stack
Change set
Change set
CloudFormation
Optional: Create additional change sets

Template anatomy

```
{
    "AWSTemplateFormatVersion": "version date",
    "Description": "JSON string",
    "Metadata": {
        template metadata
    },
    "Parameters": {
        set of parameters
    },
    "Mappings": {
        set of mappings
    },
    "Conditions": {
        set of conditions
    },
    "Transform": {
        set of transforms
    },
    "Resources": {
        set of resources
    },
    "Outputs": {
        set of outputs
    }
}
```

Required

Parameters

```
AWSTemplateFormatVersion: "2010-09-09"

Parameters:
  EnvType:
    Description: Environment type.
    Default: test
    Type: String
    AllowedValues: [prod, dev, test]
    ConstraintDescription: must specify prod, dev, or test.

Mappings:
  RegionMap:
    us-east-1:
      AMI: "ami-0ff8a91507f77f867"

Conditions:
  CreateProdResources: !Equals [!Ref EnvType, prod]
  CreateDevResources: !Equals [!Ref EnvType, "dev"]

Resources:
```

Conditions

```
AWSTemplateFormatVersion: "2010-09-09"

Parameters:
  EnvType:
    Description: Environment type.
    Default: test
    Type: String
    AllowedValues: [prod, dev, test]
    ConstraintDescription: must specify prod, dev, or test.

Mappings:
  RegionMap:
    us-east-1:
      AMI: "ami-0ff8a91507f77f867"

Conditions:
  CreateProdResources: !Equals [!Ref EnvType, prod]
  CreateDevResources: !Equals [!Ref EnvType, "dev"]

Resources:
```

Resources

```
Resources:
  EC2Instance:
    Type: "AWS::EC2::Instance"
    Properties:
      ImageId: !FindInMap [RegionMap, !Ref "AWS::Region", AMI]
      InstanceType: !If [CreateProdResources, c1.xlarge, !If [CreateDevResources,
m1.large, m1.small]]
  MountPoint:
    Type: "AWS::EC2::VolumeAttachment"
    Condition: CreateProdResources
    Properties:
      InstanceId: !Ref EC2Instance
      VolumeId: !Ref NewVolume
      Device: /dev/sdh
  NewVolume:
    Type: "AWS::EC2::Volume"
    Condition: CreateProdResources
    Properties:
      Size: 100
      AvailabilityZone: !GetAtt EC2Instance.AvailabilityZone
```

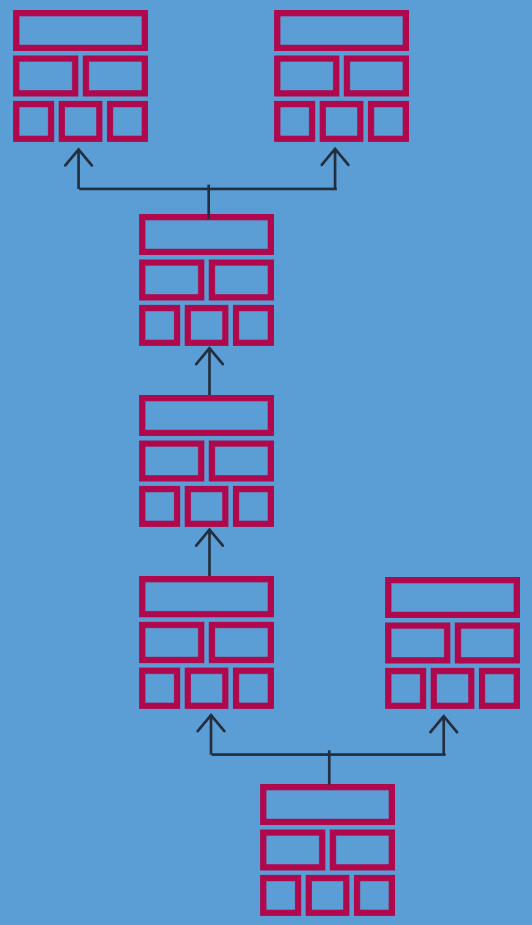
Outputs

```
  MountPoint:
      Type: "AWS::EC2::VolumeAttachment"
      Condition: CreateProdResources
      Properties:
        InstanceId: !Ref EC2Instance
        VolumeId: !Ref NewVolume
        Device: /dev/sdh
    NewVolume:
      Type: "AWS::EC2::Volume"
      Condition: CreateProdResources
      Properties:
        Size: 100
        AvailabilityZone: !GetAtt EC2Instance.AvailabilityZone

  Outputs:
      InstanceID:
        Condition: CreateProdResources
        Description: The instance ID
        Value: !Ref EC2Instance
```

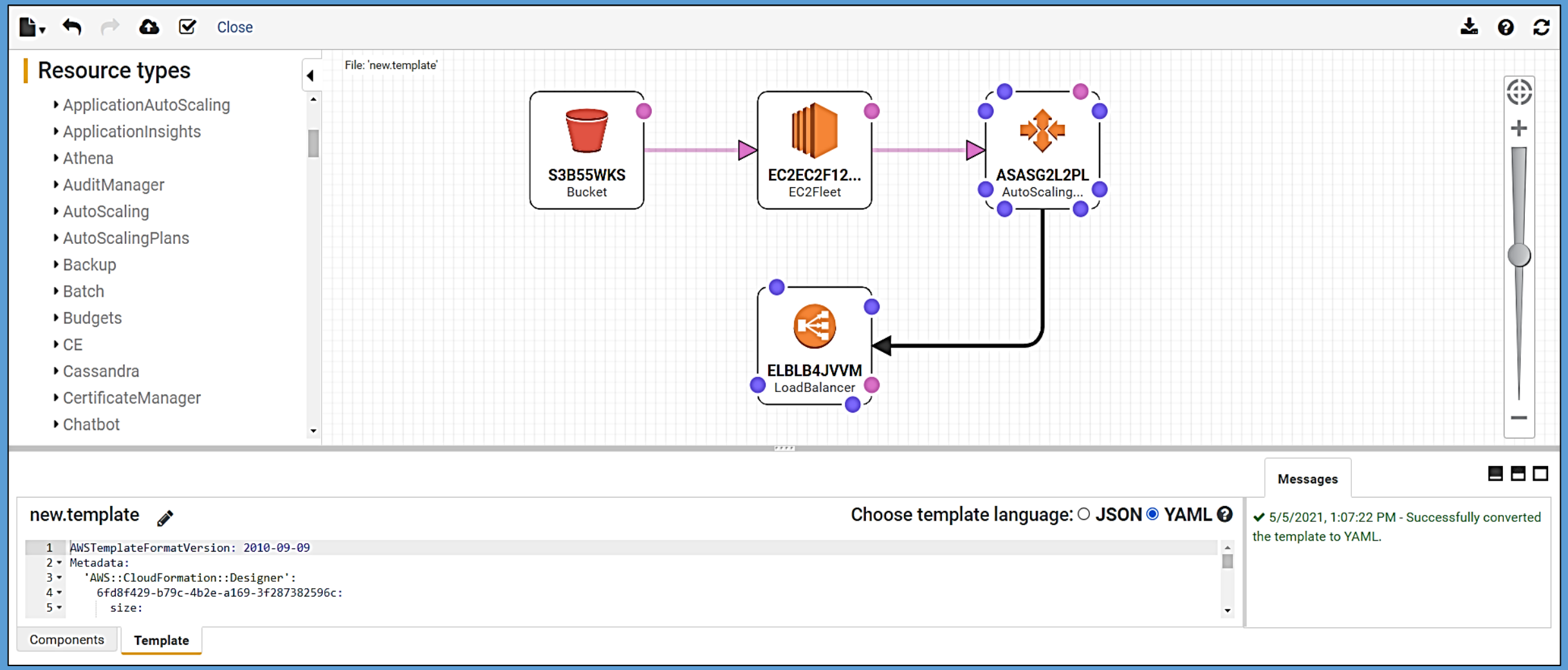
Using multiple templates

A layered architecture

Frontend	Web interface, admin interface, analytics dashboard
Backend	Customers, campaigns, products, analytics, marketing collateral
Shared	Databases, common monitoring or alarms, subnets, security groups
Base network	VPCs, internet gateways, Virtual Private Networks (VPNs), NAT gateways
Identity	Identity and Access Management users (IAM) users, groups, roles



AWS CloudFormation Designer



Infrastructure management

“How can we deploy, maintain, and scale applications in the cloud?”

Infrastructure tools

Deployments

Management

AWS Elastic Beanstalk

AWS Solutions Library

AWS Cloud Development Kit (AWS CDK)

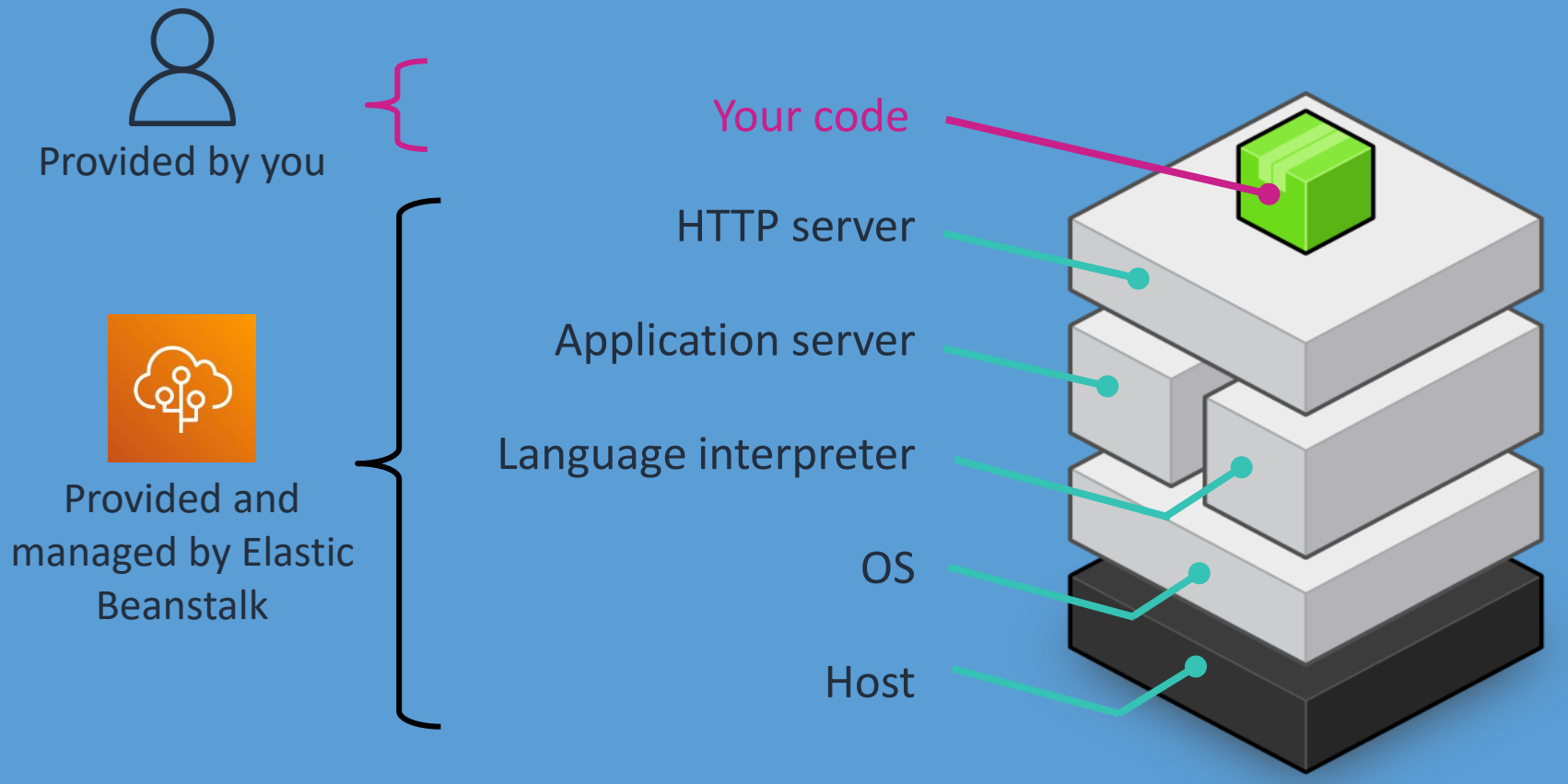
AWS CloudFormation

AWS Systems Manager

Convenience ⟷ Control

AWS Elastic Beanstalk

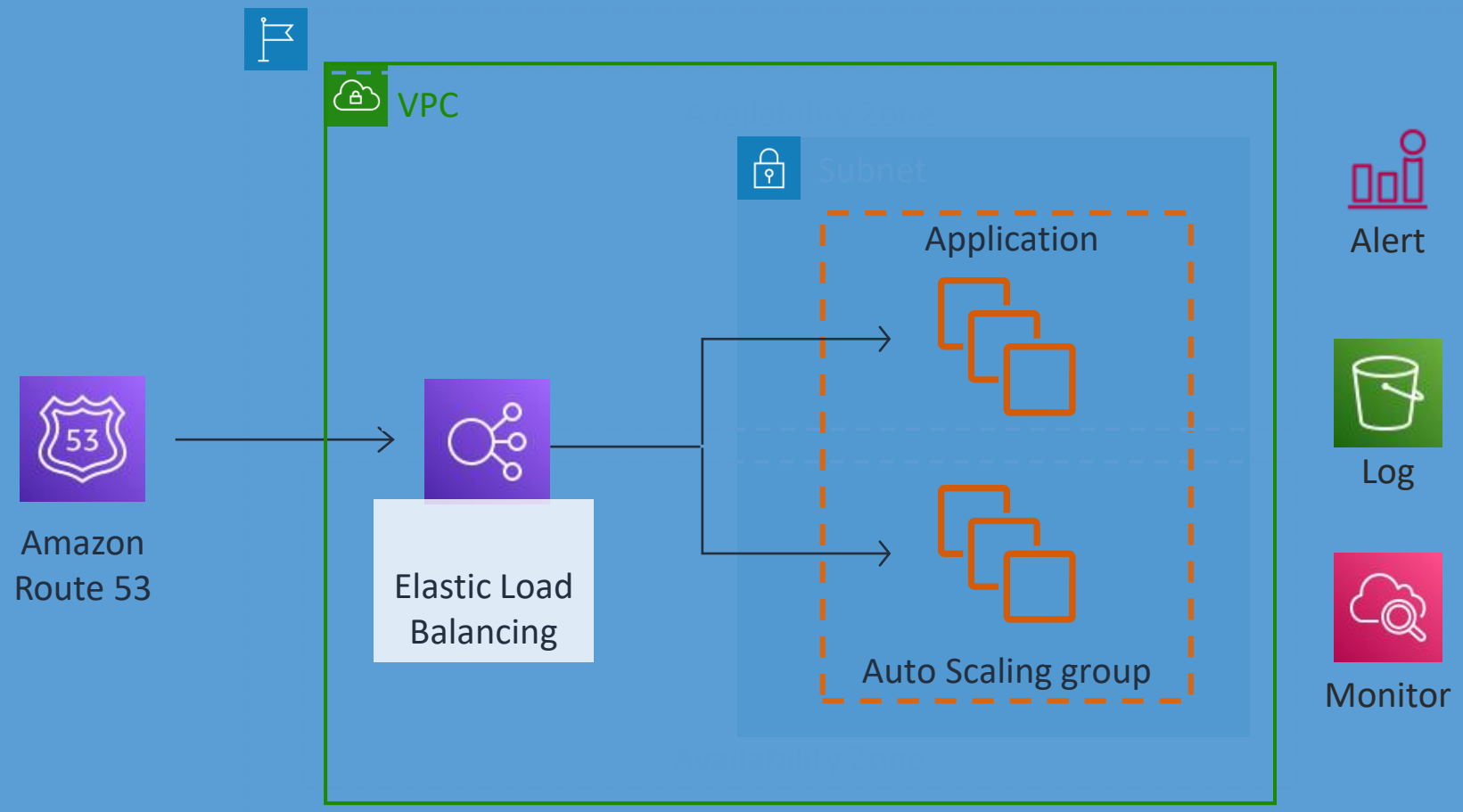
- Provisions and operates the infrastructure
- Manages the application stack for you
- Shows everything that is created
- Automatically scales your application up and down



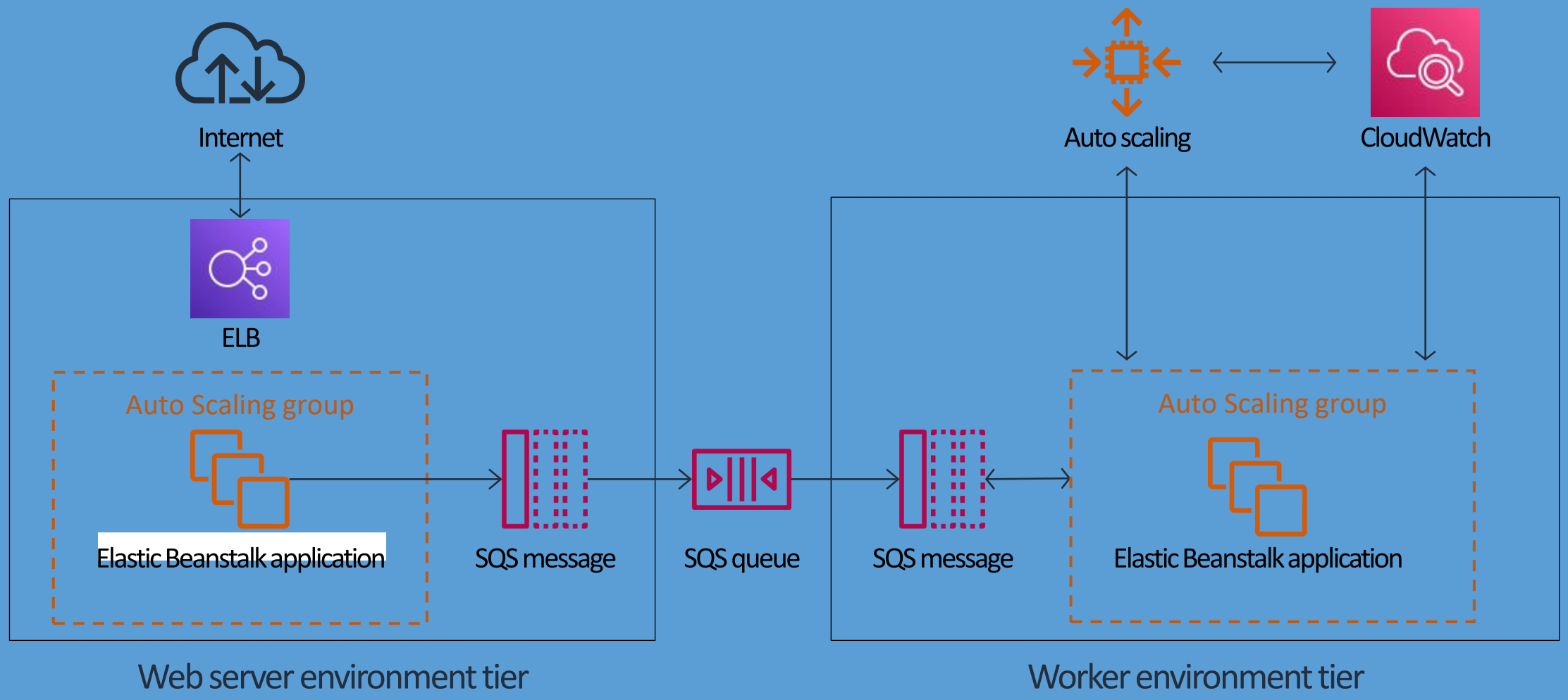
Elastic Beanstalk web server environment

http://[yourapp].elasticbeanstalk.com

- Provisions the necessary AWS resources
- Provides a unique domain name, or use your own
- Supports an EC2 instance or multiple instances with load balancing and auto scaling

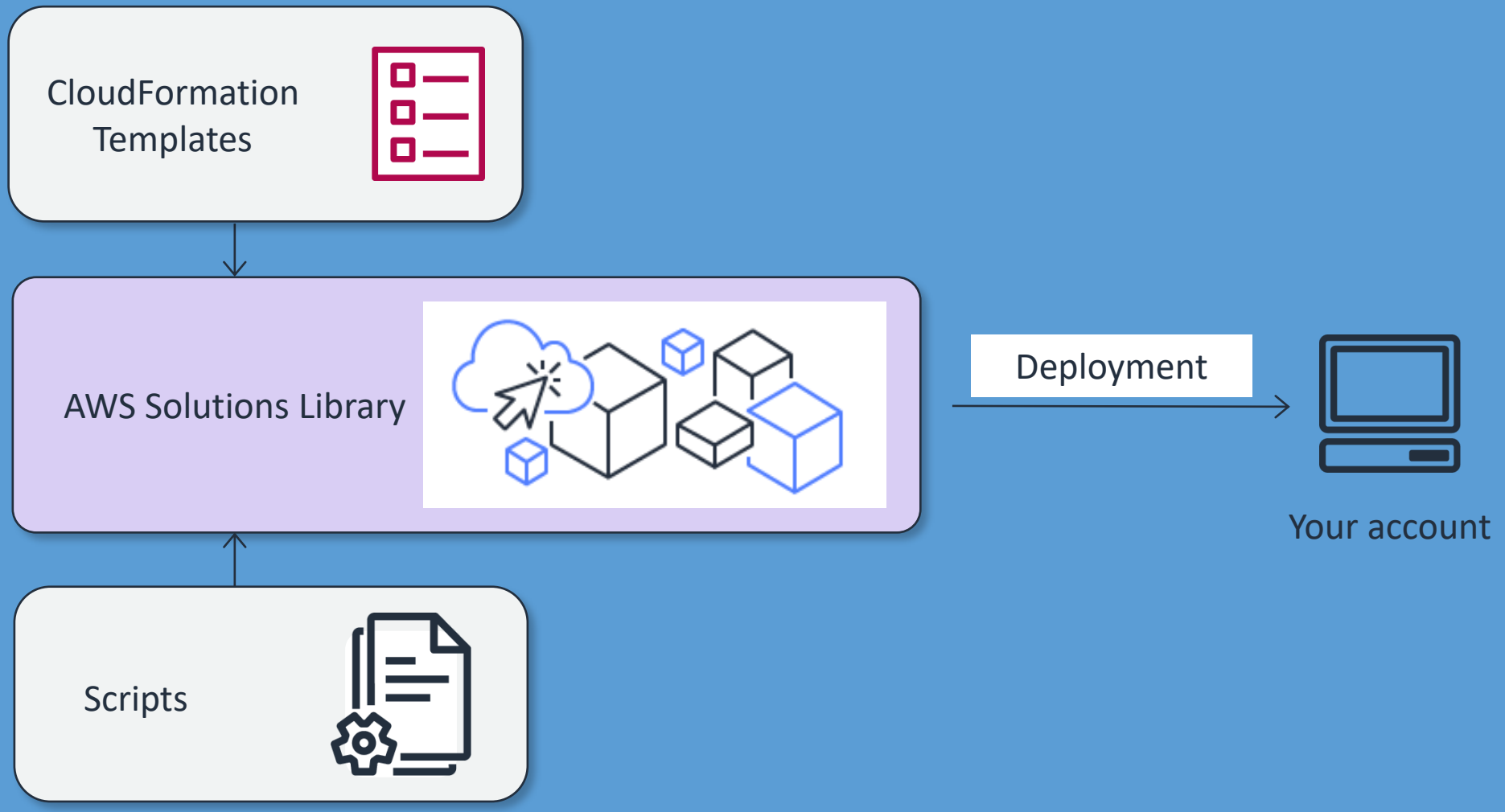


Elastic Beanstalk worker environment



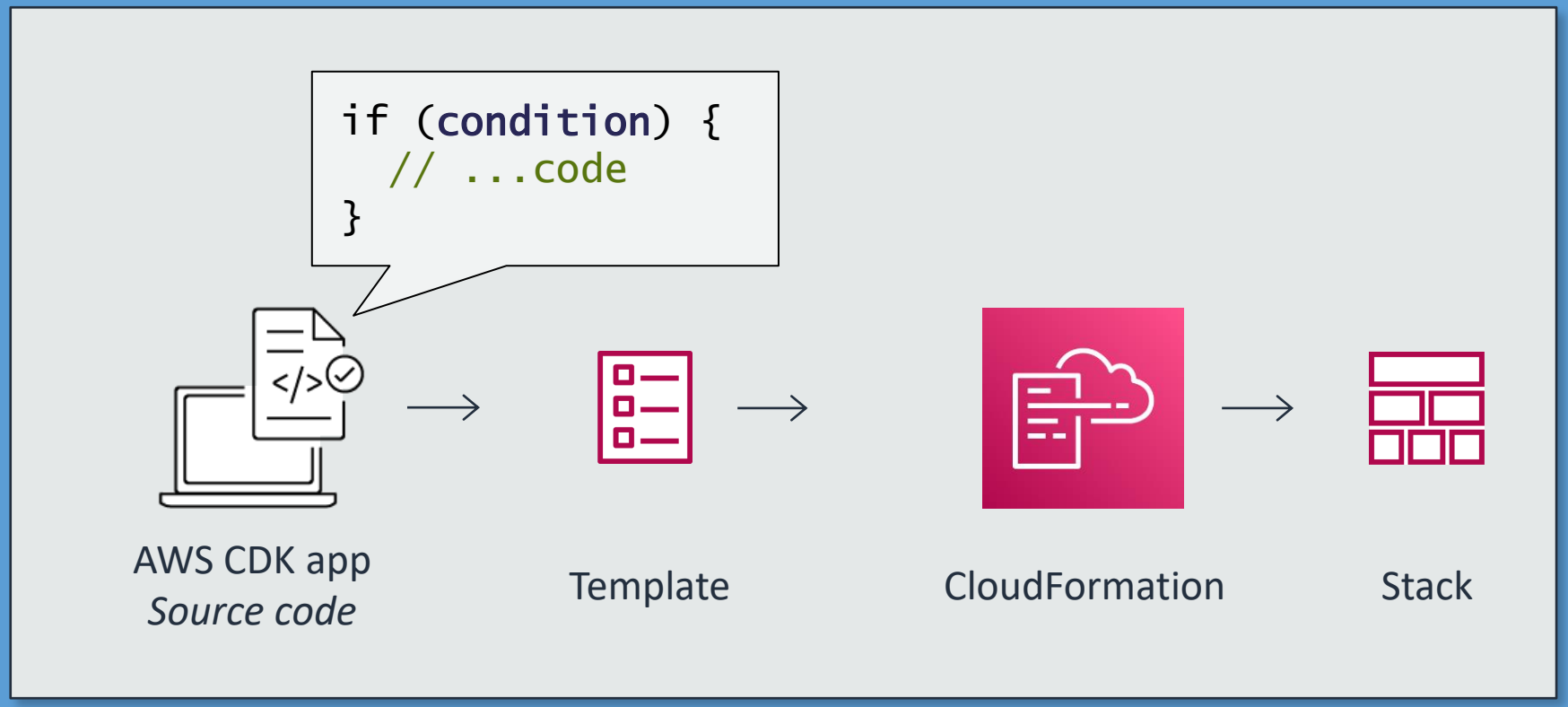
AWS Solutions Library

- Prebuilt reference architectures
- Deployment accelerator
- Solutions approved by AWS architects



AWS CDK

- Uses any supported language to generate templates
- Supports autocomplete and inline documentation
- Has proven defaults and reusable classes
- Provisions multiple environments



AWS Systems Manager

Provisioning and entitlement

Configuration management

Operations and compliance management

Monitoring

Review

Present solutions

Chief Technology Officer

Consider how you can answer the following:

- How can we simplify our cloud infrastructure build?
- How can we deploy, maintain, and scale applications in the cloud?

Module review

In this module you learned about:

- ✓ Amazon CloudFormation
- ✓ Infrastructure management

Next, you will review:

Knowledge check

Knowledge check

Knowledge check question 1

What is a CloudFormation stack?

A	All of the provisioned resources defined in a CloudFormation template
B	All of the resources identified as drifted in a CloudFormation template
C	A condition when resources are added on top of each other
D	The properties of a single resource

Knowledge check question 1 and answer

What is a CloudFormation stack?

A correct	All of the provisioned resources defined in a CloudFormation template
B	All of the resources identified as drifted in a CloudFormation template
C	A condition when resources are added on top of each other
D	The properties of a single resource

Knowledge check question 2

Which of the following are benefits of using AWS CDK with CloudFormation? (Select TWO.)

A	Developers can use common programming languages.
B	Bulk discounts are automatically applied to resource usage.
C	Developers can call preconfigured resources with proven defaults.
D	Components are limited to a single user.
E	Using AWS CDK does not require an AWS account or credentials.

Knowledge check question 2 and answer

Which of the following are benefits of using AWS CDK with CloudFormation? (Select TWO.)

A correct	Developers can use common programming languages.
B	Bulk discounts are automatically applied to resource usage.
C correct	Developers can call preconfigured resources with proven defaults.
D	Components are limited to a single user.
E	Using AWS CDK does not require an AWS account or credentials.

AWS

Containers

Question

What percentage of your workloads run on containers?

A. Less than 10 percent

B. 10–50 percent

C. More than 50 percent

D. I’m not sure

Module overview

- Business requests
- Microservices
- Containers
- Container services
- Present solutions
- Knowledge check

Business Requirements

Compute Operations Manager

The compute operations manager wants to know:

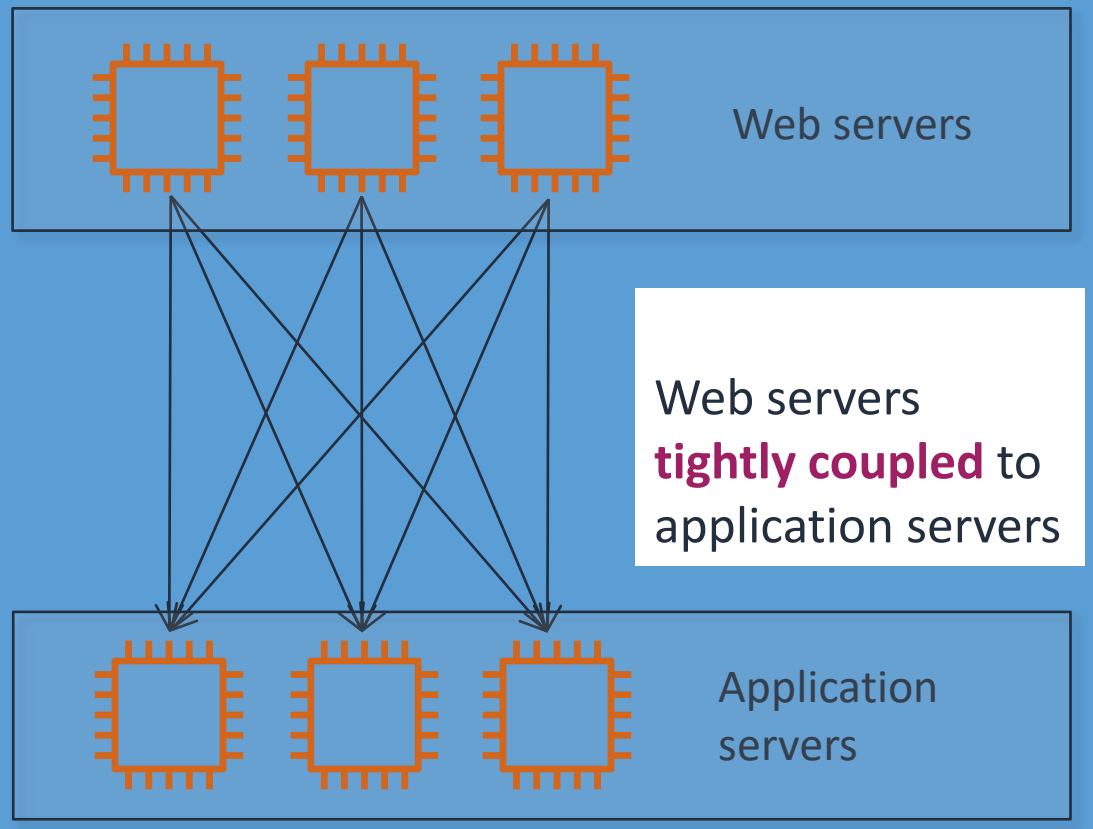
- How can we make components of our applications more independent so changes in one service will not affect any other?
- What are the benefits of using containers for our compute needs?
- What options do we have for managing containerized applications in the cloud?

Microservices

“How can we make components of our applications more independent so changes in one service will not affect any other?”

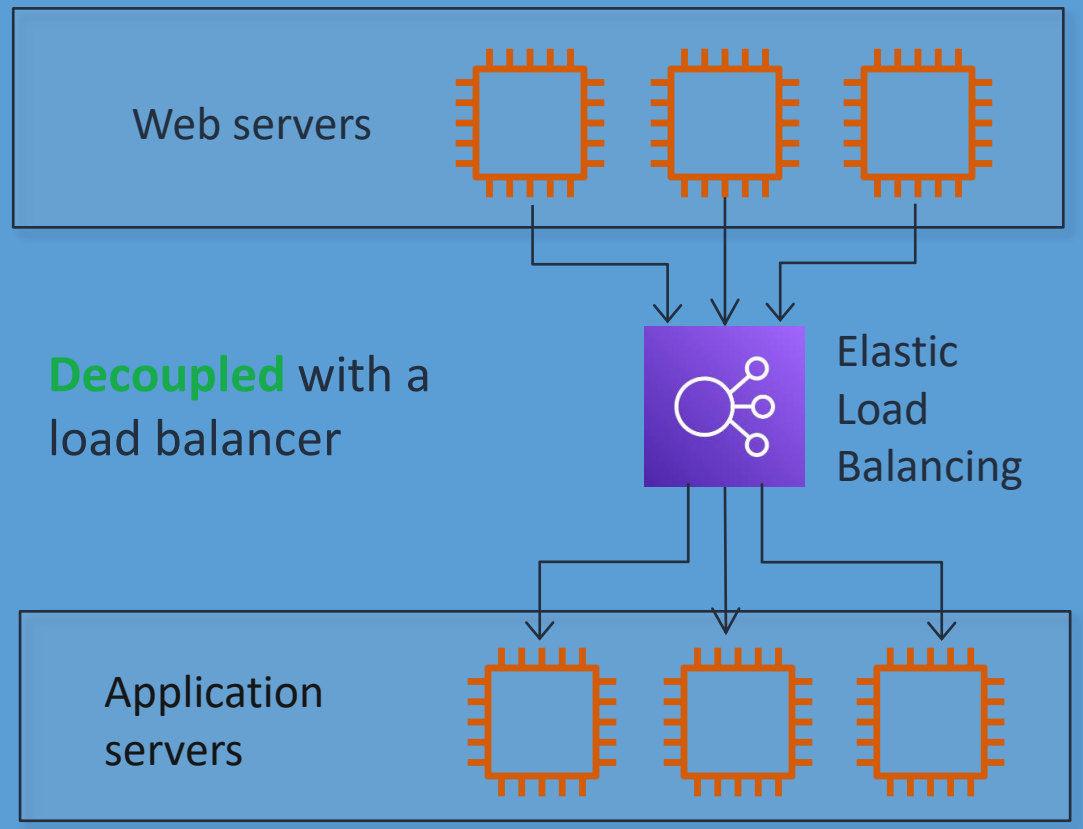
Loose coupling

Anti-pattern



Web servers **tightly coupled** to application servers

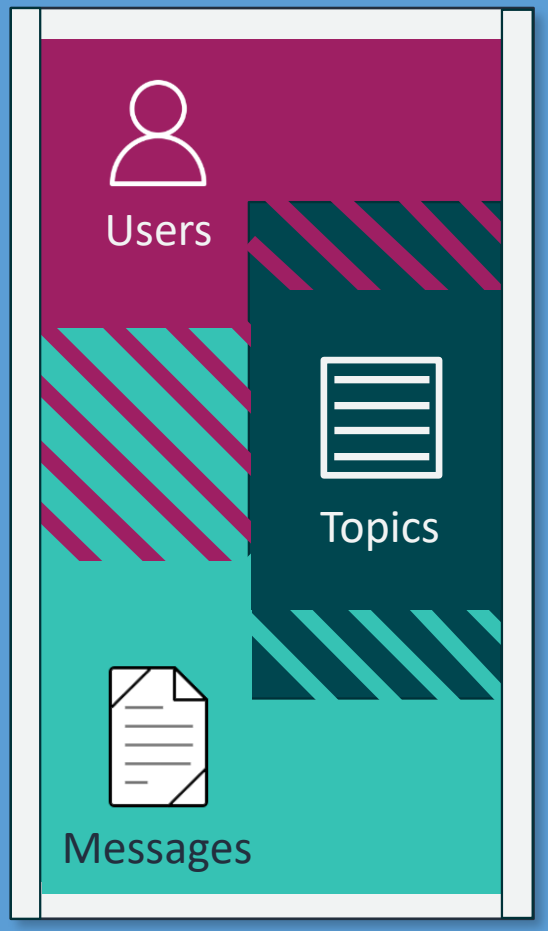
Best practice



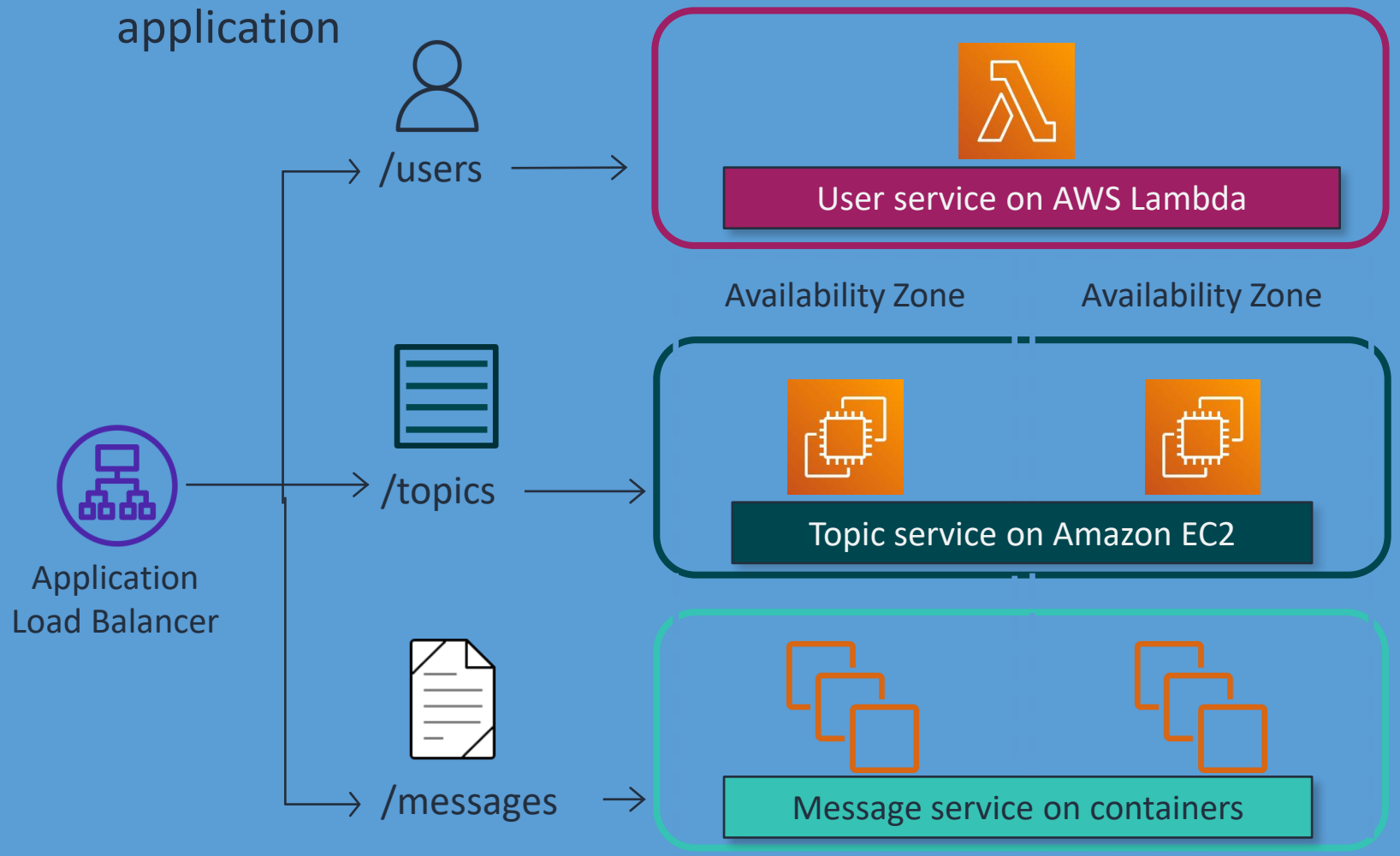
Decoupled with a load balancer

Microservices

Monolithic forum application



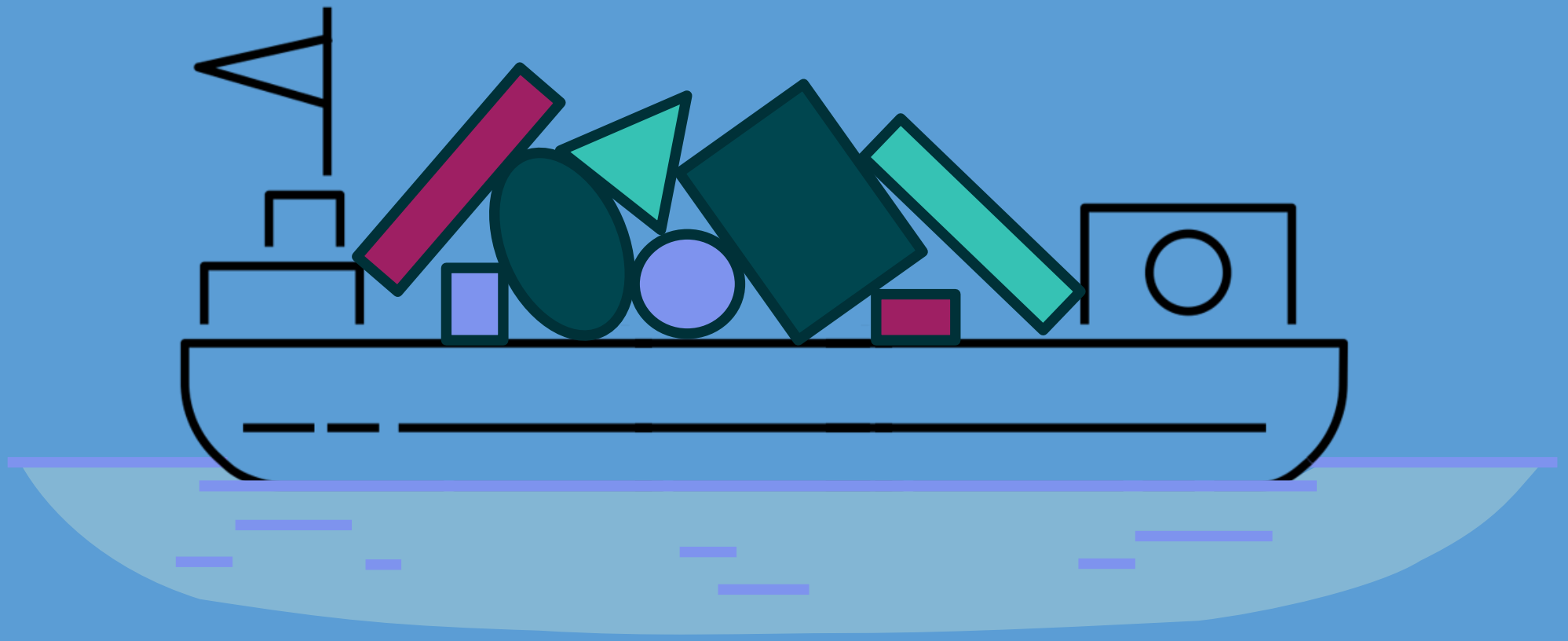
Microservice forum application



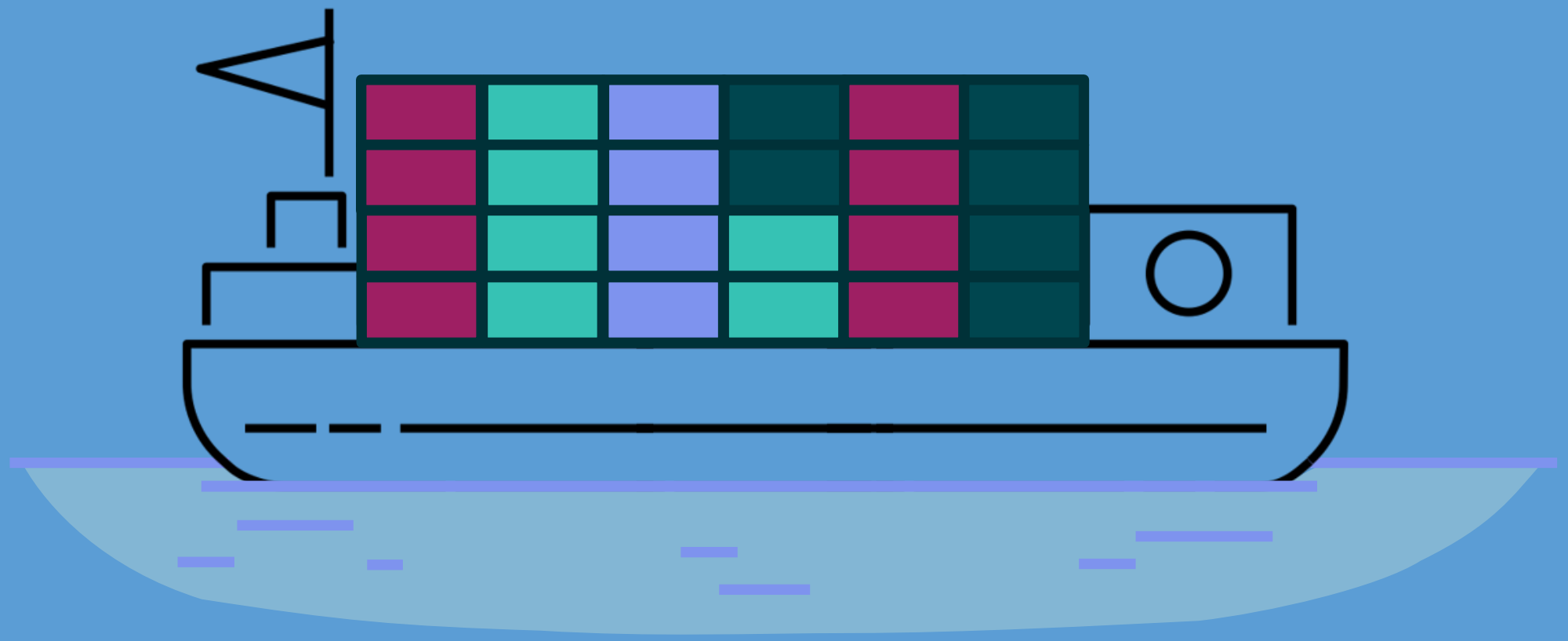
Containers

“What are the benefits of using containers for our compute needs?”

Shipping before standardization



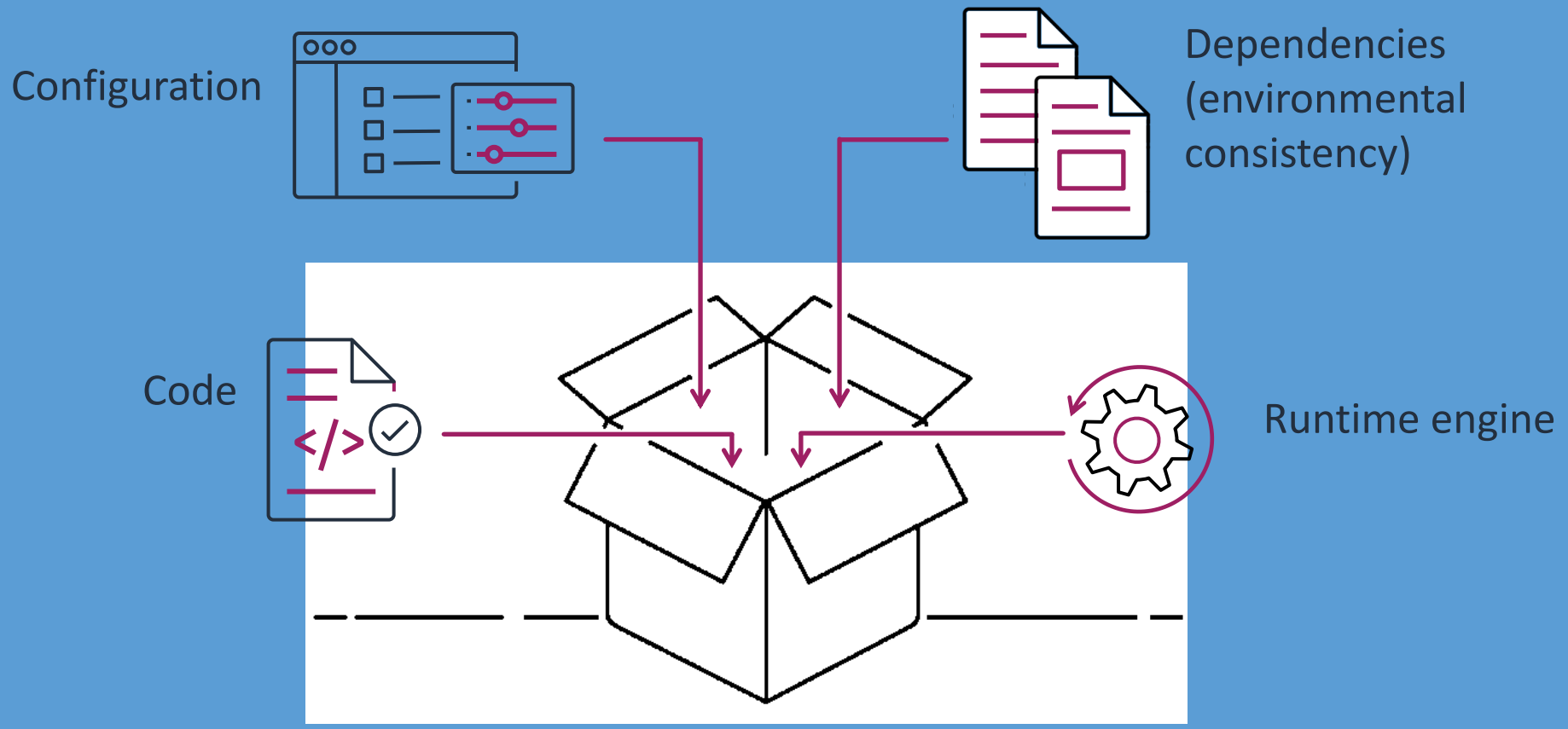
Standardized unit of storage



Containers

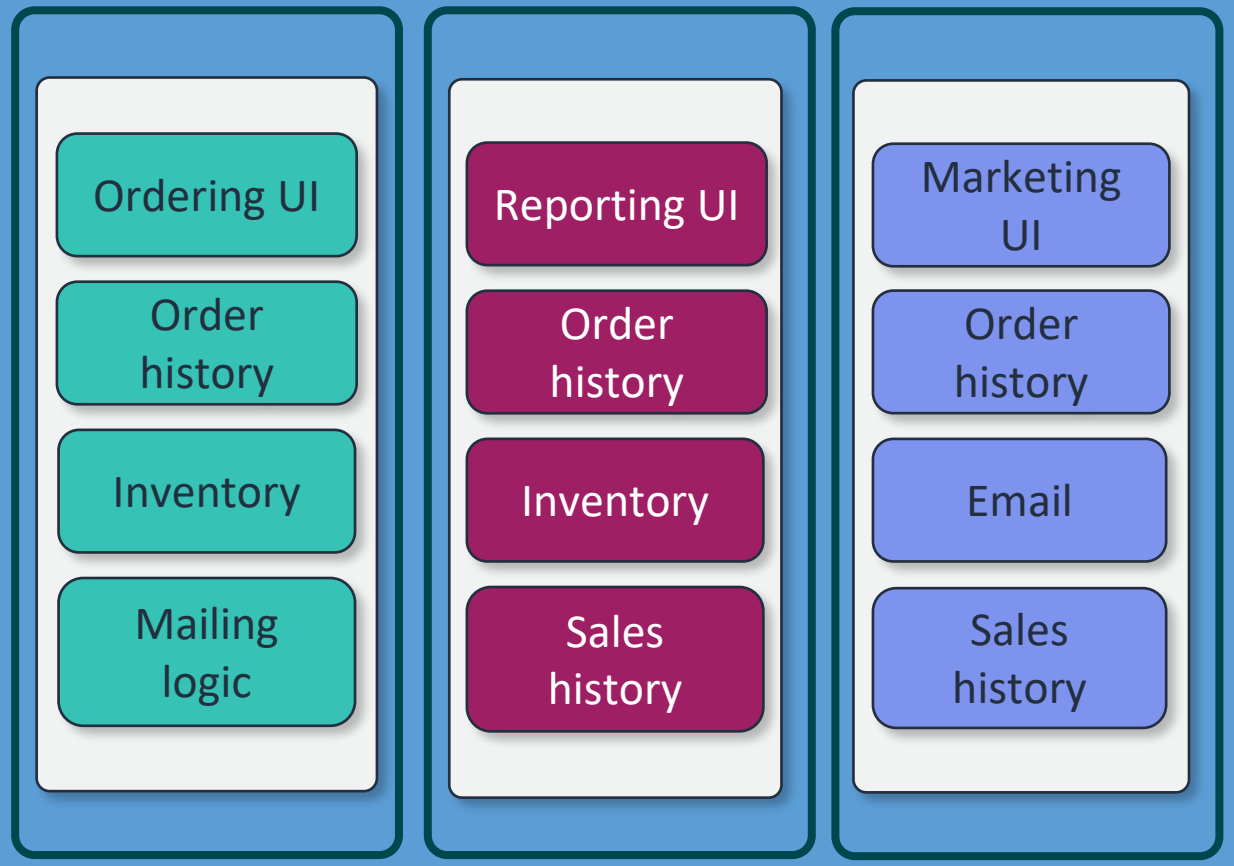
Containers are:

- Repeatable
- Self-contained environments
- Faster to spin up and down than VMs
- Portable
- Scalable

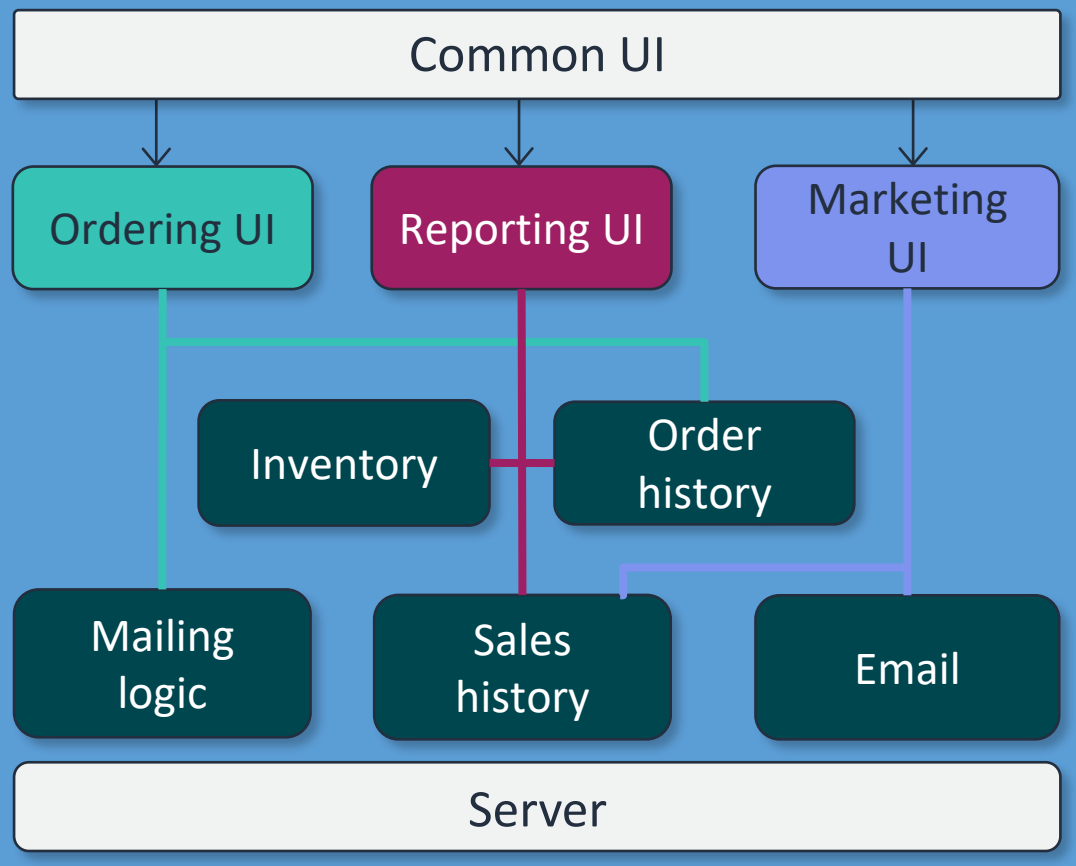


Containers and microservices

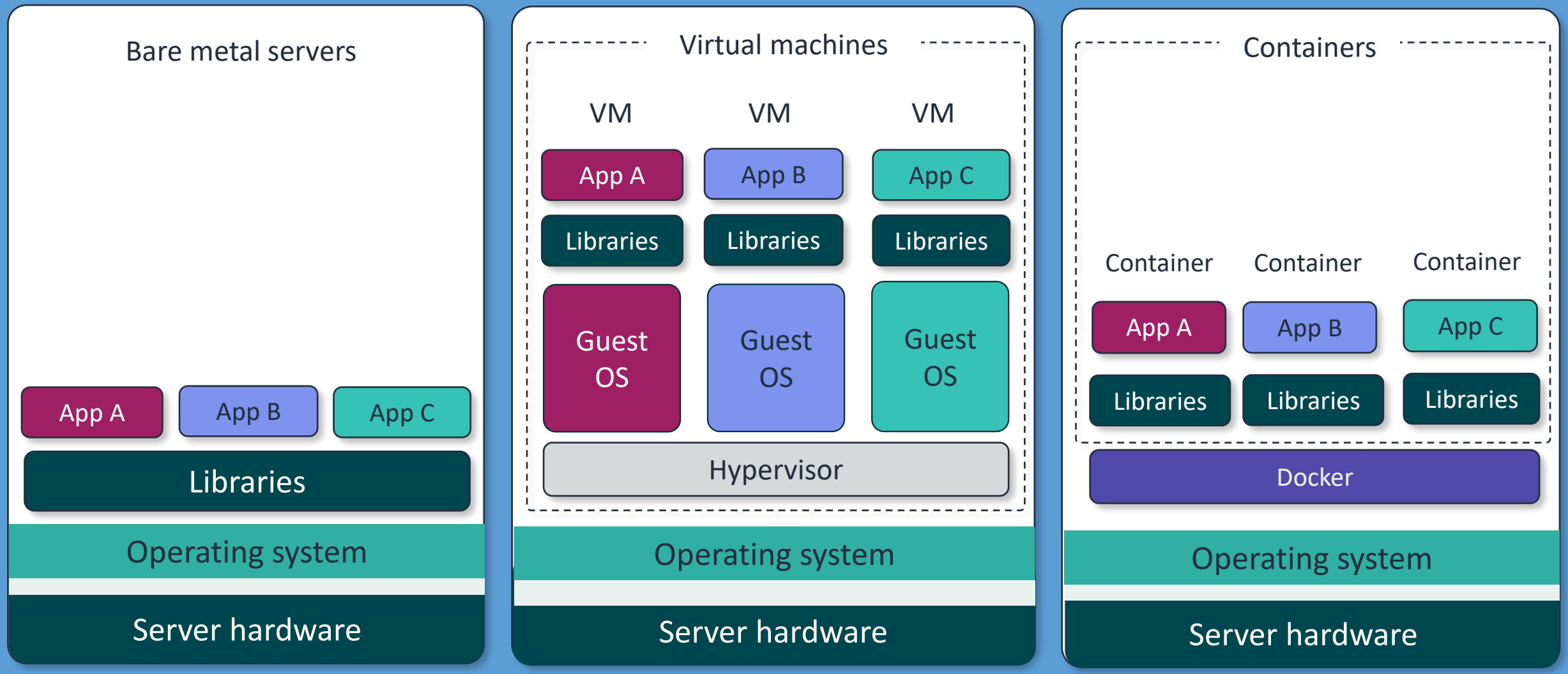
Monolithic order application



Microservice order application



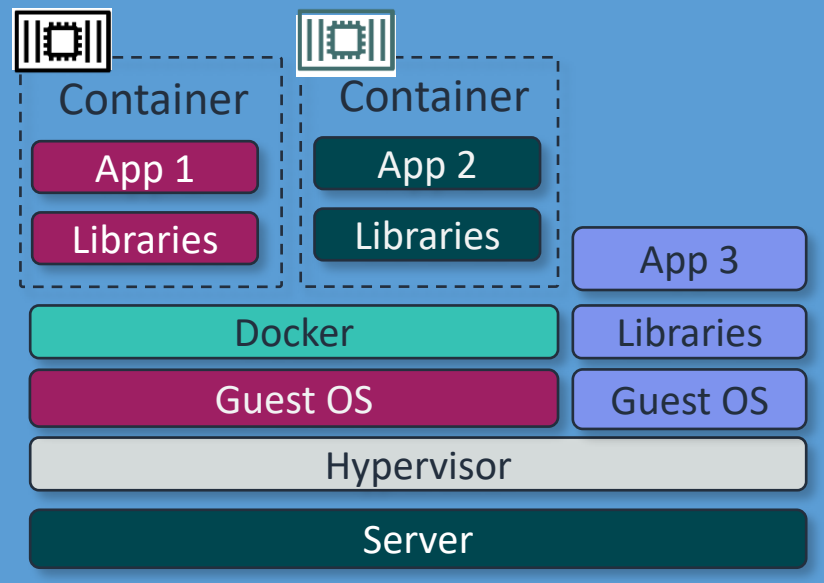
Levels of abstraction and virtualization



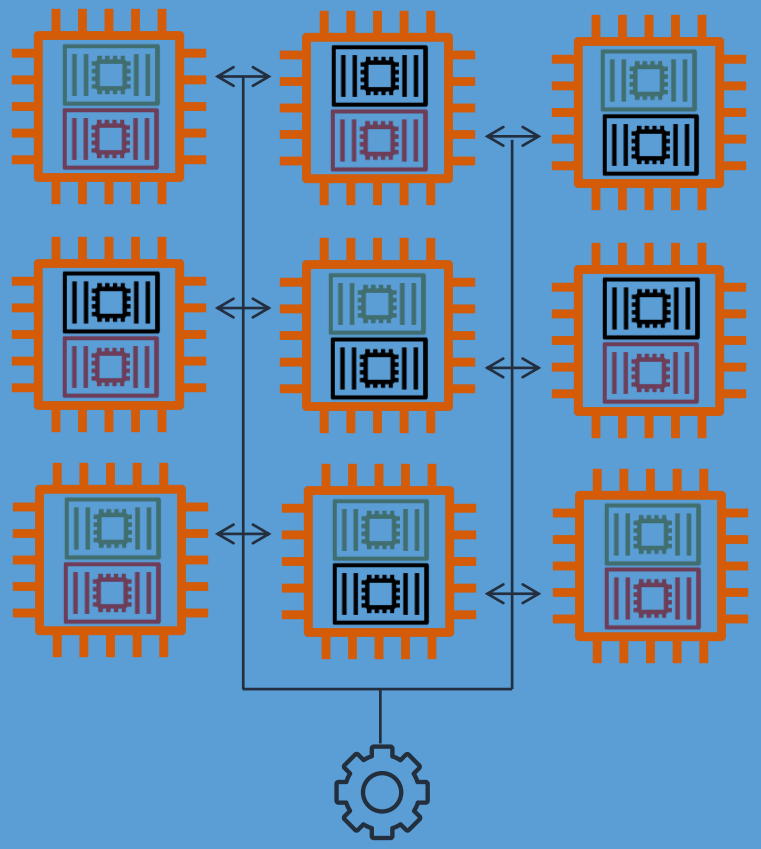
Containers on AWS

- Running containers directly on Amazon EC2 requires you to manage scaling, connectivity, and maintenance.
- Using an orchestration tool helps manage:
  - Scheduling
  - Placement
  - Networking
  - Monitoring

Containers on Amazon EC2



Containers with an orchestration tool



Container orchestration tool

Container services

“What options do we have for managing containerized applications in the cloud?”

Running containers on AWS

Registry

Pulls container image from the registry

Orchestration tool

Deploys container to hosting resource

Container hosting

Amazon Elastic Container Registry (Amazon ECR)

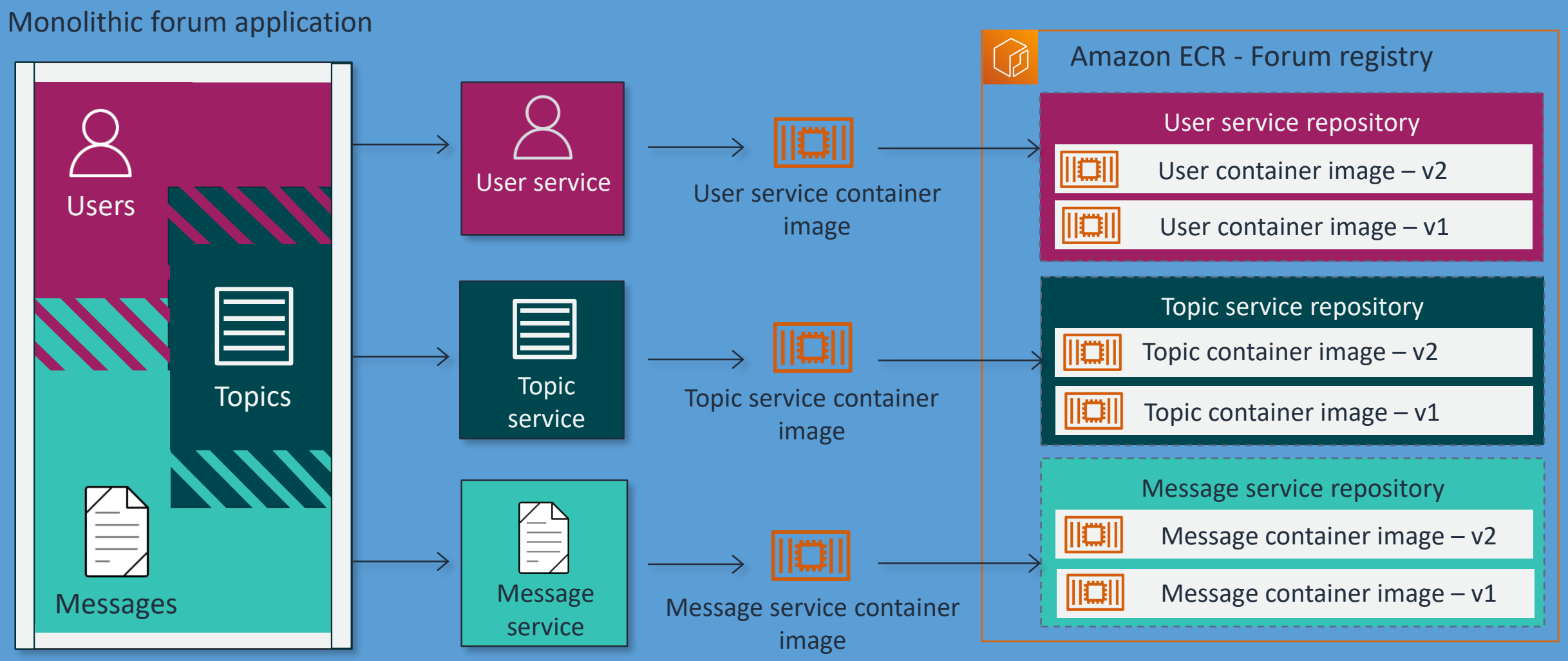
Amazon Elastic Kubernetes Service (Amazon EKS)

Amazon Elastic Container Service (Amazon ECS)

AWS Fargate

Amazon EC2

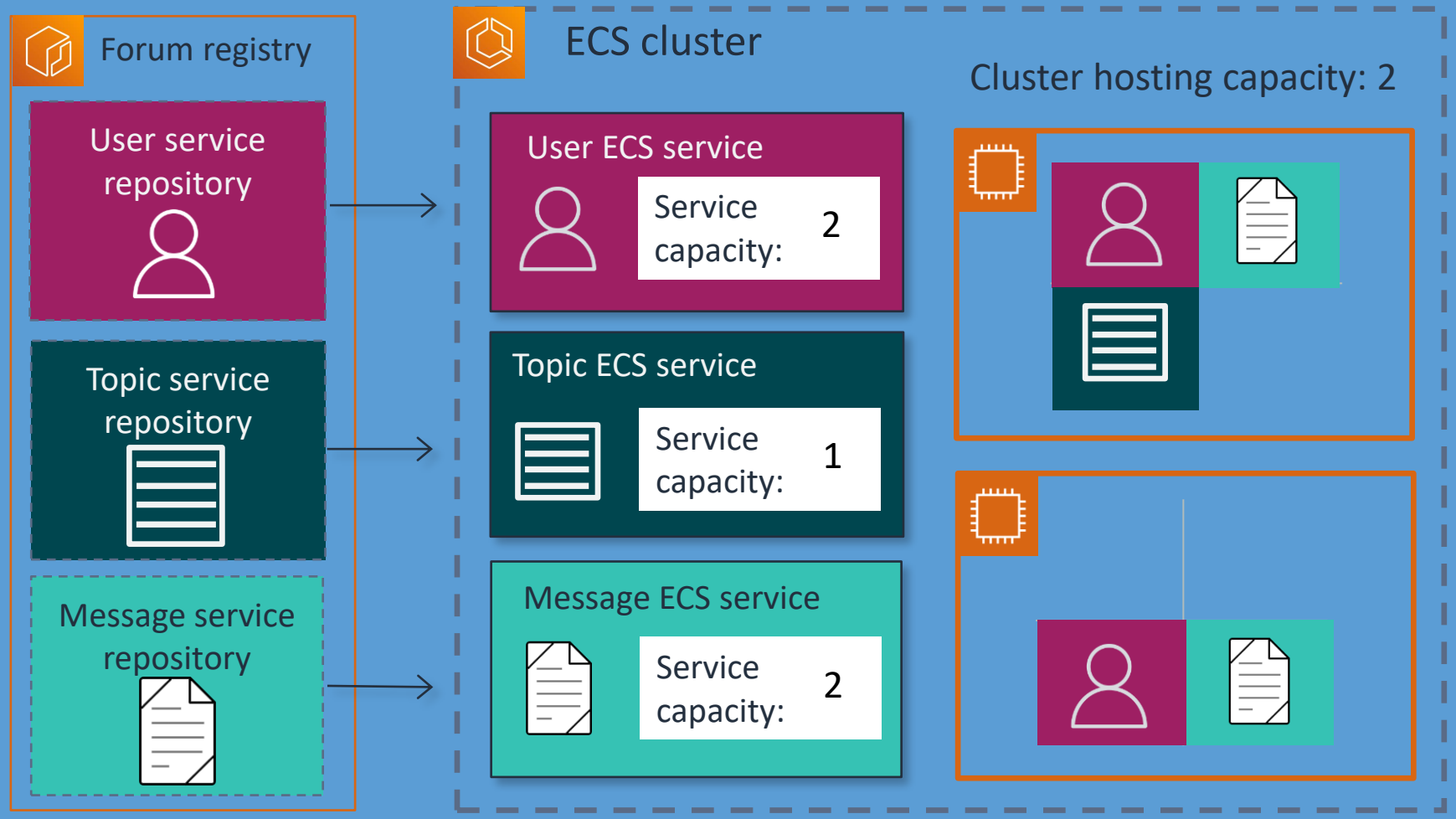
Amazon ECR


Monolithic forum application
Users
Topics
Messages
User service
Topic service
Message service
User service container image
Topic service container image
Message service container image
Amazon ECR - Forum registry
User service repository
User container image – v2
User container image – v1
Topic service repository
Topic container image – v2
Topic container image – v1
Message service repository
Message container image – v2
Message container image – v1

Amazon ECS orchestration

Managed container orchestration service tightly integrated with AWS

- Pulls images from your repositories
- ECS services scale service capacity by managing container count
- ECS clusters scale hosting capacity



Amazon ECS features

Amazon ECS

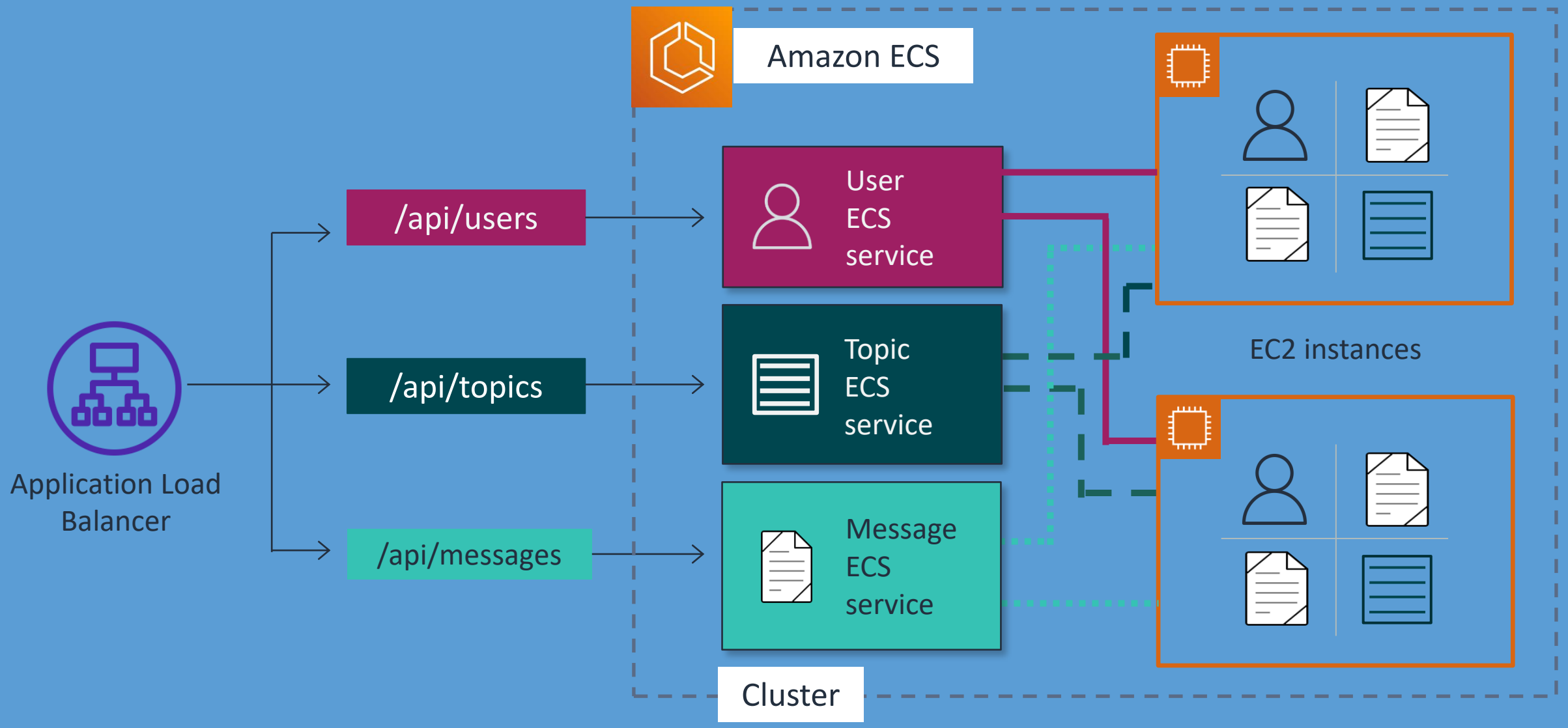
Fully managed

Service discovery

AWS integrations

Works with common development workflows

Monolithic to container-based microservices



Amazon EKS

kubernetes

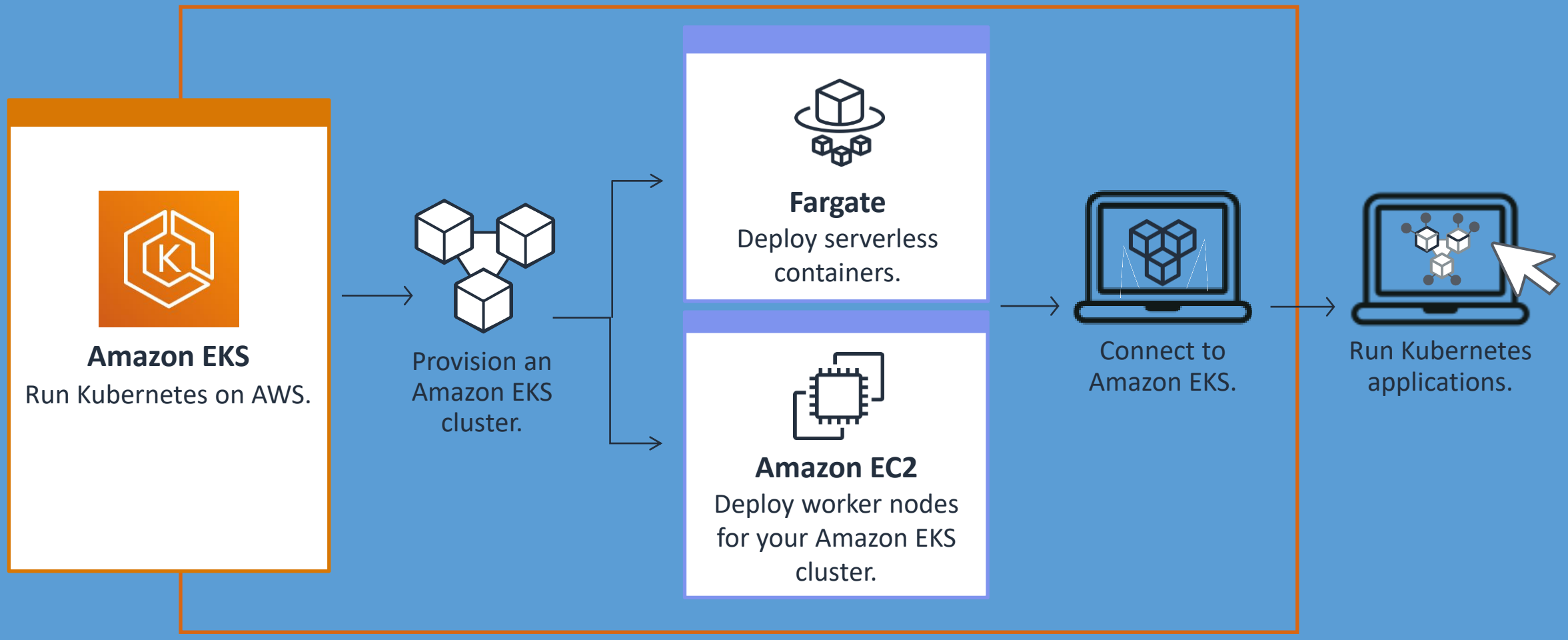
Run applications at scale

Seamlessly move applications

Run anywhere

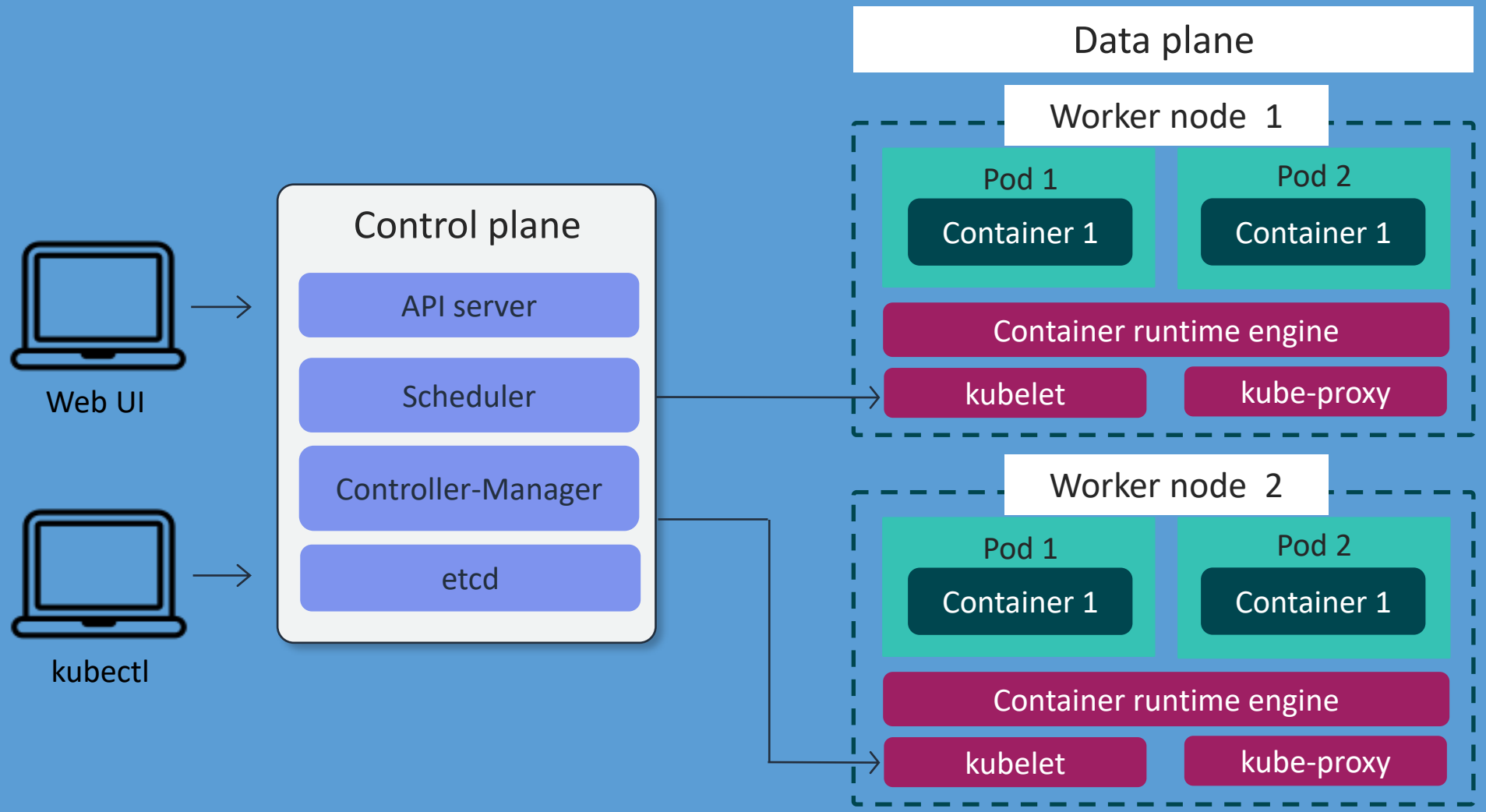
Add new functionality

Amazon EKS solutions



Kubernetes architecture

A Kubernetes cluster is a set of worker machines, called nodes, that run containerized applications.



AWS Fargate

Build a container image.

Amazon ECS

or

Amazon EKS

Define the images and resources needed for your app.

AWS Fargate

Launch containers, and Fargate manages all of the underlying container infrastructure.

Launch containers

Fargate runs your containers for you.

Manage containers

Amazon ECS scales your applications and manages your containers for availability.

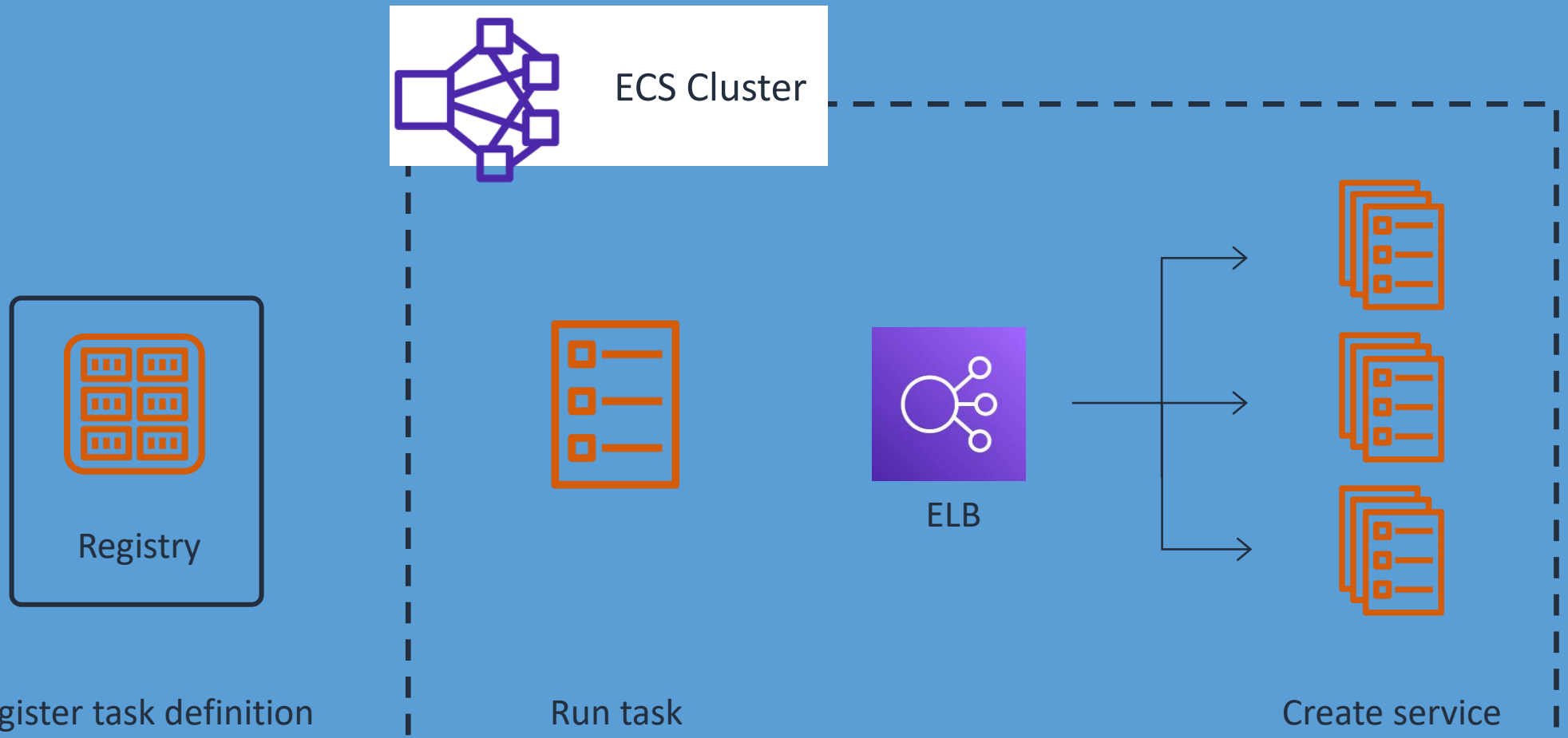
Choosing AWS container services

Least effort ⟷ Most effort

Choose your orchestration service.	Amazon ECS	Amazon EKS
Choose your hosting type.	AWS Fargate	Amazon EC2

Fargate compute constructs

No need to provision, configure, or scale clusters of VMs to run containers



Compute operational models

Least effort

Most effort

Compute service	AWS manages:	You manage:	
Lambda Serverless functions	Datasource integrations Physical hardware, software, networking, and facilities Provisioning	Application code	
Fargate Serverless containers	Container orchestration, provisioning, cluster scaling Physical hardware, host OS or kernel, networking, and facilities	Application code Datasource integrations Security configuration	Security updates Network configuration Management tasks
Amazon ECS and Amazon EKS Container management as a service	Container orchestration control plane Physical hardware, software, networking, and facilities	Application code Datasource integrations Work clusters	Security configuration and updates, network configuration, firewall, management tasks
Amazon EC2 Infrastructure as a service	Physical hardware, software, networking, and facilities	Application code Datasource integrations Scaling Security configuration	Security updates Network configuration Provisioning, managing, scaling, and patching

Amazon EKS container options

	On premises			In the cloud	
	EKS Distro	EKS Anywhere	EKS + AWS Outposts	EKS + EC2	EKS + Fargate
Control plane	Customer	Customer	aws	aws	aws
Compute	Customer	Customer	aws	aws	aws
Data plane	Customer	Customer	Customer	Customer	aws
Support	Community	aws	aws	aws	aws

Least → **You manage** → **Most**

Amazon ECS container options

	On premises		In the cloud	
	ECS Anywhere	ECS + Outposts	ECS + EC2	ECS + Fargate
Control plane	Customer	aws	aws	aws
Compute	Customer	aws	aws	aws
Data plane	Customer	Customer	Customer	aws
Support	aws	aws	aws	aws

Least → **You manage** → **Most**

Review

Present solutions

Compute Operations Manager

Consider how you would answer the following:

- How can we make components of our applications more independent so changes in one service will not affect any other?
- What are the benefits of using containers for our compute needs?
- What options do we have for managing containerized applications in the cloud?

Module review

In this module you learned about:

- ✓ Microservices
- ✓ Containers
- ✓ Container services

Next, you will review:

Knowledge check

Knowledge check

Knowledge check question 1

Which of the following are characteristics of microservices? (Select TWO.)

A	Loosely coupled
B	Redundant
C	Autonomous and independent
D	Tightly integrated
E	Interdependent components

Knowledge check question 1 and answer

Which of the following are characteristics of microservices? (Select TWO.)

A correct	Loosely coupled
B	Redundant
C correct	Autonomous and independent
D	Tightly integrated and dependent
E	Interdependent components

Knowledge check question 2

Which of the following are characteristics of containers? (Select TWO.)

A	Portable and scalable
B	Requires a hypervisor
C	Automatic
D	Repeatable
E	Each requires its own operating system

Knowledge check question 2 and answer

Which of the following are characteristics of containers? (Select TWO.)

A correct	Portable and scalable
B	Requires a hypervisor
C	Automatic
D correct	Repeatable
E	Each requires its own operating system

Knowledge check question 3

Containers in Amazon ECS are logically organized in:

A	A cluster
B	Pods
C	EBS volumes
D	Amazon S3

Knowledge check question 3 and answer

Containers in Amazon ECS are logically organized in:

A correct	A cluster
B	Pods
C	EBS volumes
D	Amazon S3

Knowledge check question 4

Why would you choose to deploy your containers to AWS Fargate over Amazon EC2?

A	To take control of your infrastructure
B	To avoid manual infrastructure updates
C	To optimize price for a large load
D	To manage your own patches and updates

Knowledge check question 4 and answer

Why would you choose to deploy your containers to AWS Fargate over Amazon EC2?

A	To take control of your infrastructure
B correct	To avoid manual infrastructure updates
C	To optimize price for a large load
D	To manage your own patches and updates

AWS

Networking 2

Question

How many VPCs does your organization use?

A. <20

B. 20 to 100

C. >100

D. I'm not sure

Module overview

- Business requests
- VPC endpoints
- VPC peering
- Hybrid networking
- AWS Transit Gateway
- Present solutions
- Knowledge check

Business requests

Network Engineer

The network engineer needs to know:

- What can we do to keep our connections to AWS services private?
- How can we privately route traffic between our VPCs?
- What are our options to connect our on-premises network to the AWS Cloud?
- Which services can reduce the number of route tables we need to manage our global network?

VPC endpoints

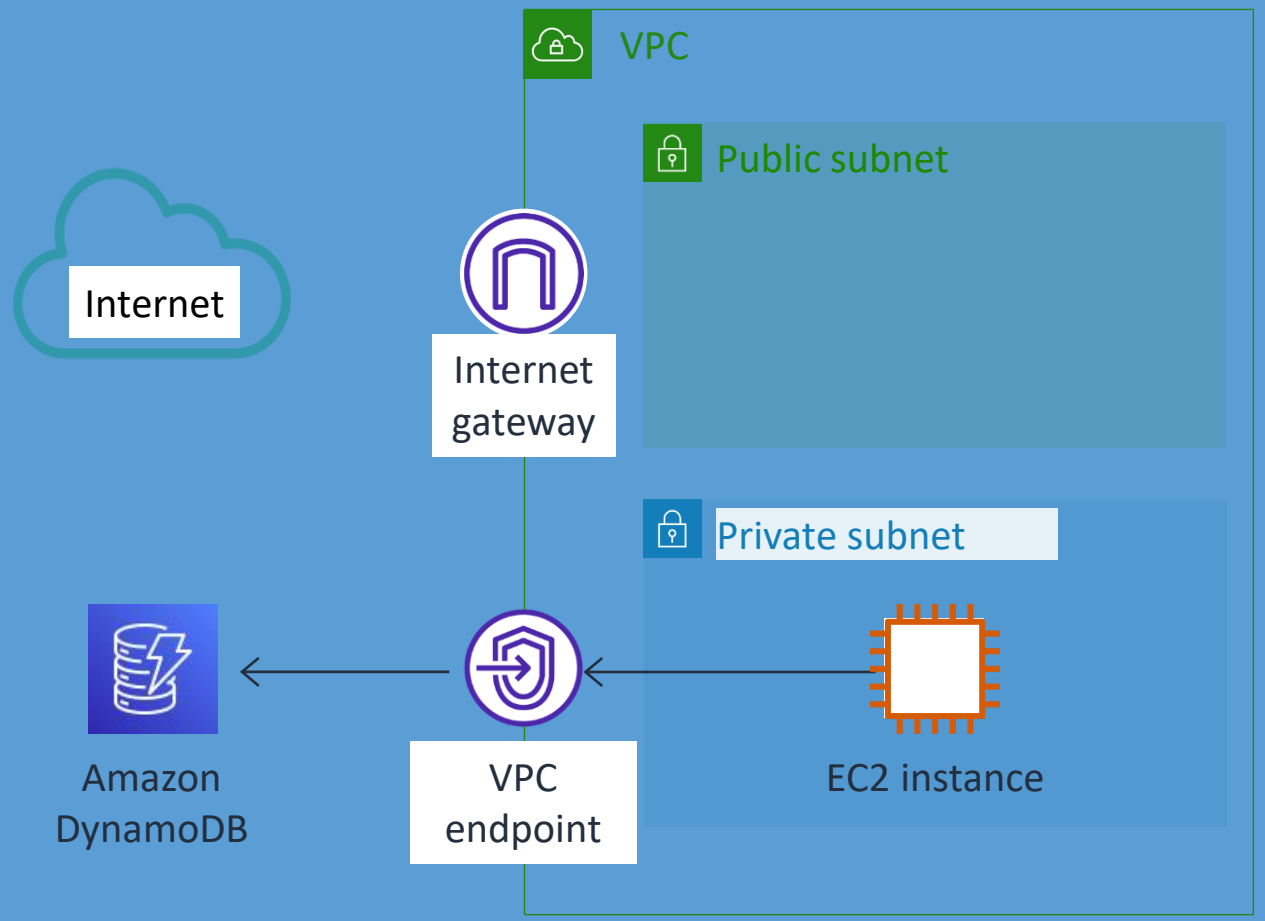
“What can we do to keep our connections to AWS services private?”

VPC endpoints

Access AWS services without an internet gateway, NAT gateway, or public IP address.

VPC endpoints are:

- Horizontally scaled
- Redundant
- Highly available



Gateway and interface VPC endpoints

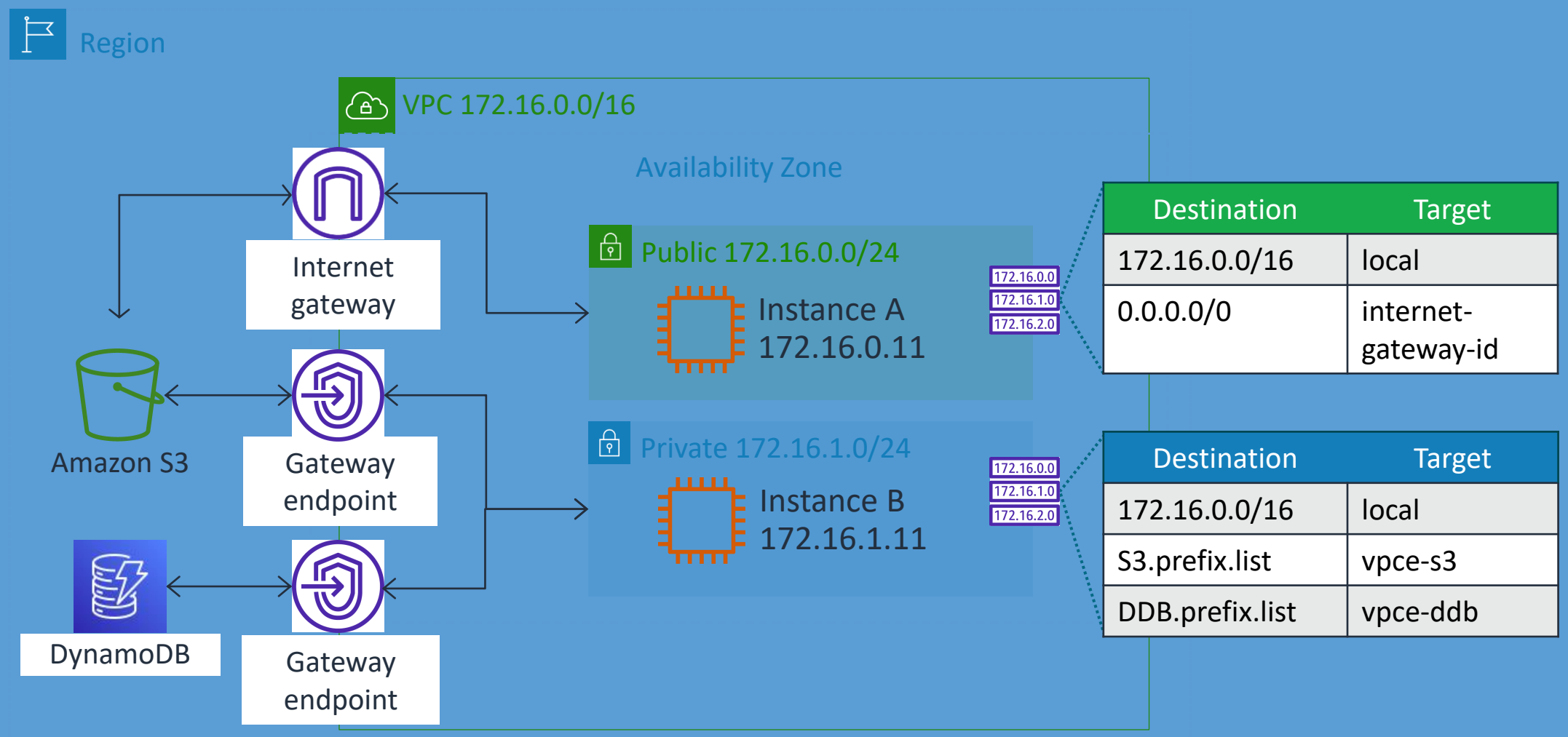
Gateway endpoint

- Target specified in route table
- Supports the following services:
  - Amazon Simple Storage Service (Amazon S3)
  - Amazon DynamoDB

Interface endpoint

- Elastic network interface with a private IP address
- Supports more services than gateway endpoints
- Powered by AWS PrivateLink

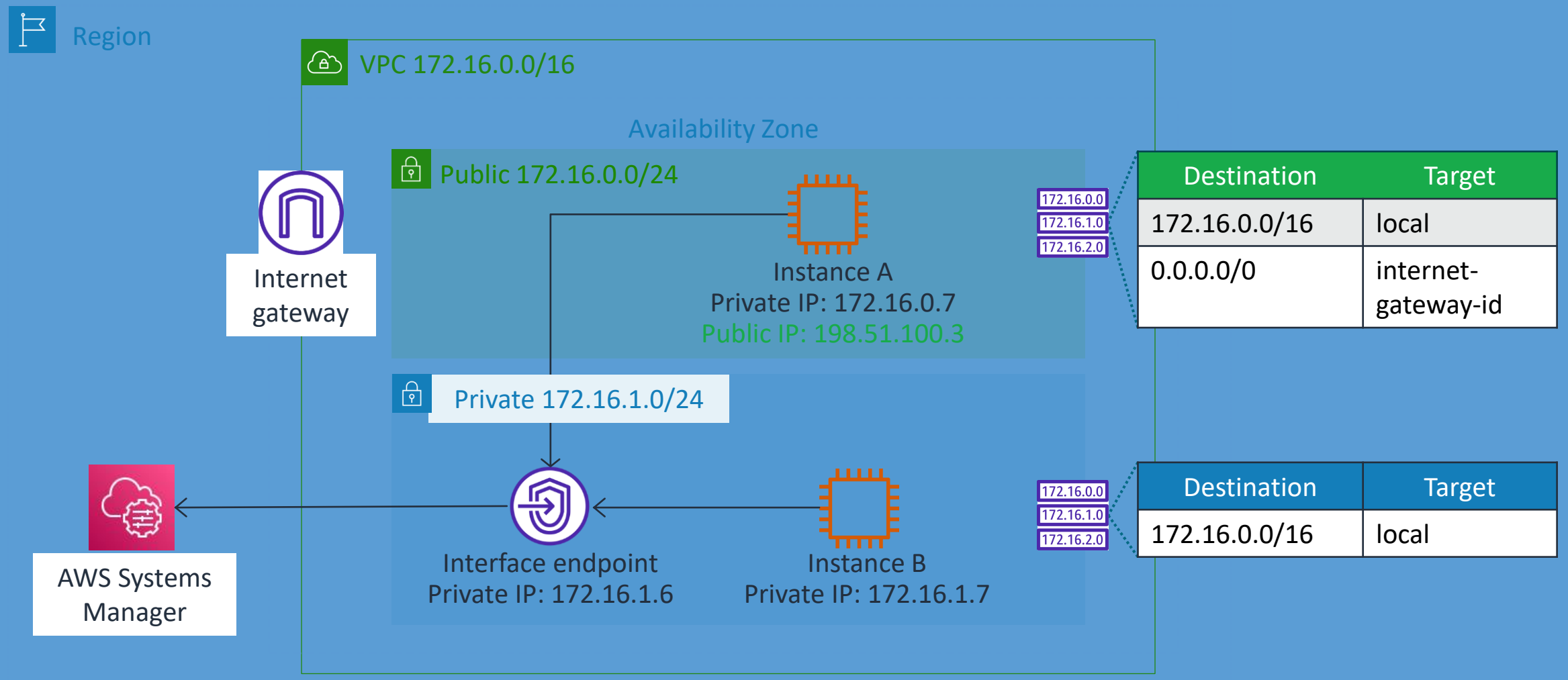
Gateway VPC endpoints



Destination	Target
172.16.0.0/16	local
0.0.0.0/0	internet-gateway-id

Destination	Target
172.16.0.0/16	local
S3.prefix.list	vpce-s3
DDB.prefix.list	vpce-ddb

Interface VPC endpoints



VPC peering

“How can we privately route traffic between our VPCs?“

VPC peering

VPC peering connects networks between two VPCs.

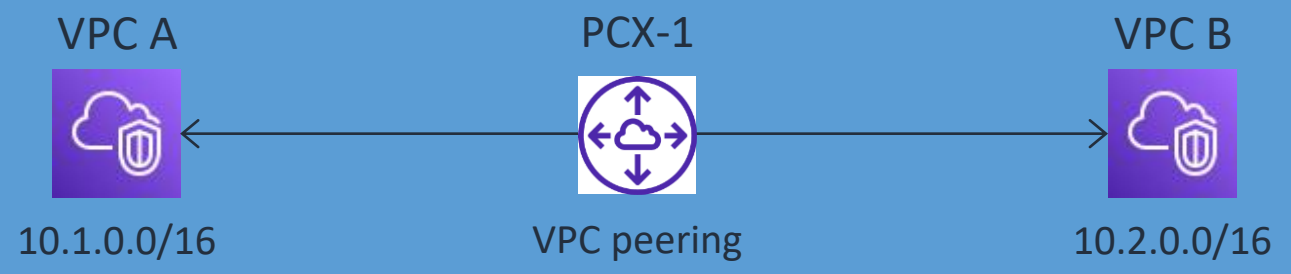
- Intra-region and inter-region support
- Cross-account support

Route Table: VPC A

Destination	Target
10.1.0.0/16	local
10.2.0.0/16	PCX-1

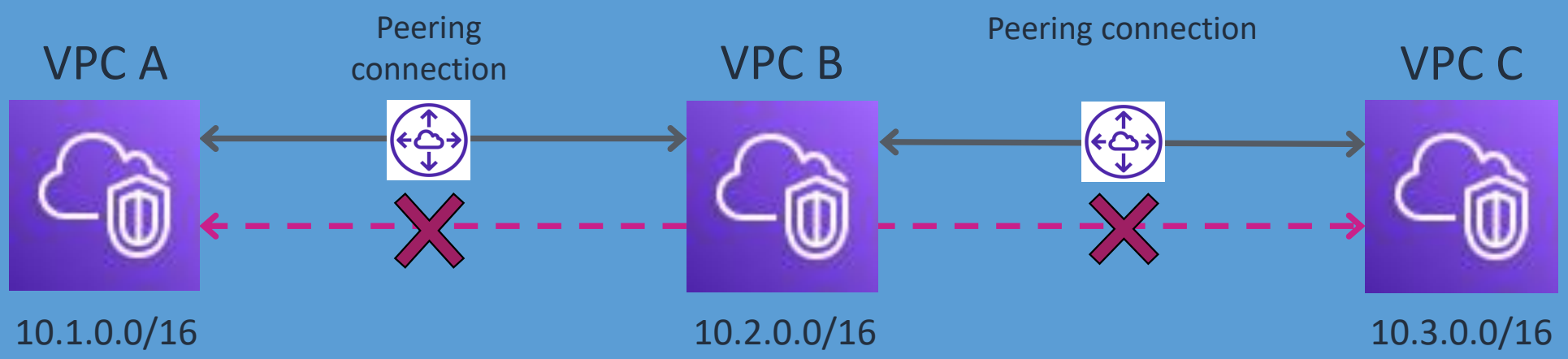
Route Table: VPC B

Destination	Target
10.2.0.0/16	local
10.1.0.0/16	PCX-1



Note: IP spaces cannot overlap

Multiple VPC peering connections



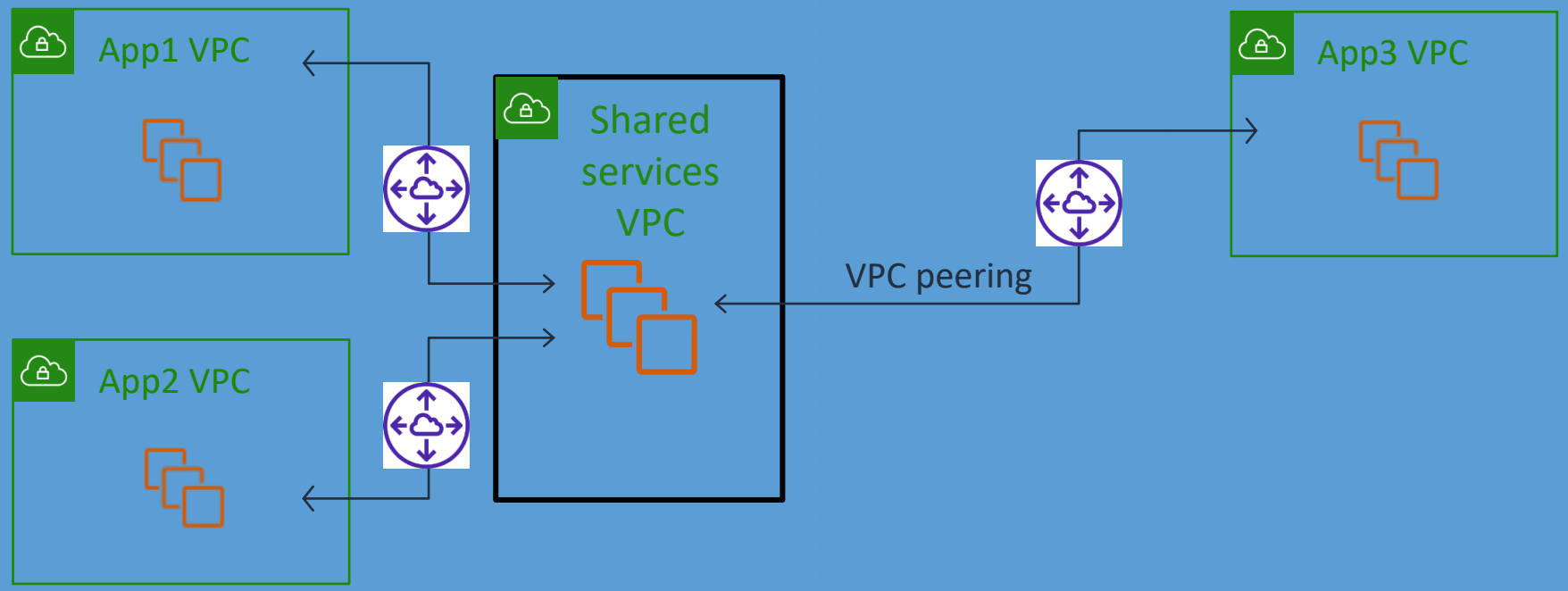
Note: No transitive peering relationships

Benefits of VPC peering

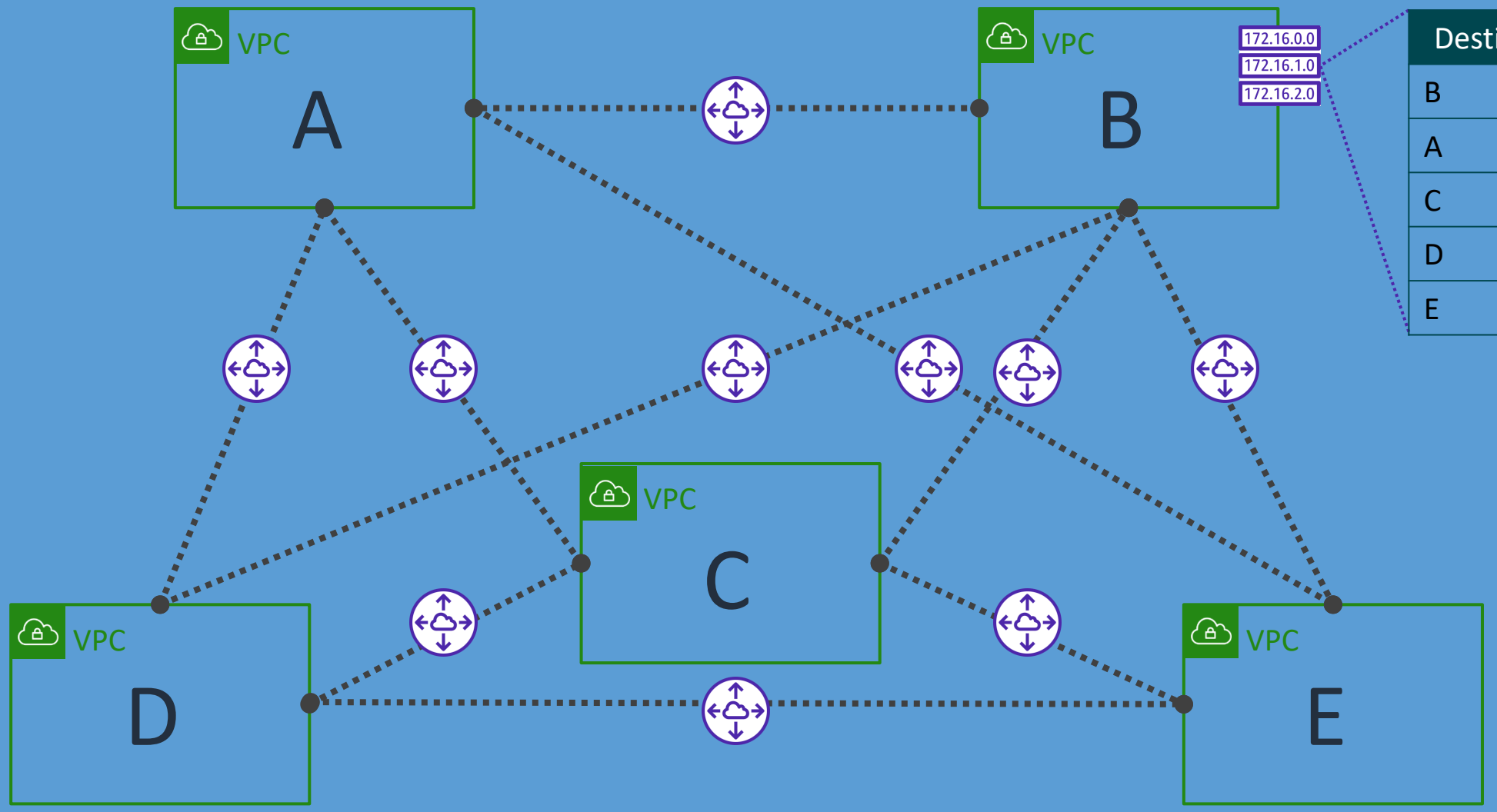
- Bypasses the internet gateway or virtual private gateway
- Provides highly available connections—no single point of failure
- Avoids bandwidth bottlenecks
- Uses private IP addresses to direct traffic between VPCs

Example: VPC peering for shared services

- App VPCs have no peering with each other.
- You cannot use the shared services VPC as a transit point between app VPCs.



Example: Full mesh VPC peering



Destination	Target
B	Local
A	PCX-1
C	PCX-2
D	PCX-3
E	PCX-4

Number of peering connections for a full mesh

$$\frac{n(n - 1)}{2}$$

Example

$$\frac{10(10 - 1)}{2} = 45$$

Example (cont.)

$$\frac{100(100 - 1)}{2} = 4{,}950$$

What is the problem?

Static routes per Amazon VPC route table

100

Amazon VPC peering connections per Amazon VPC

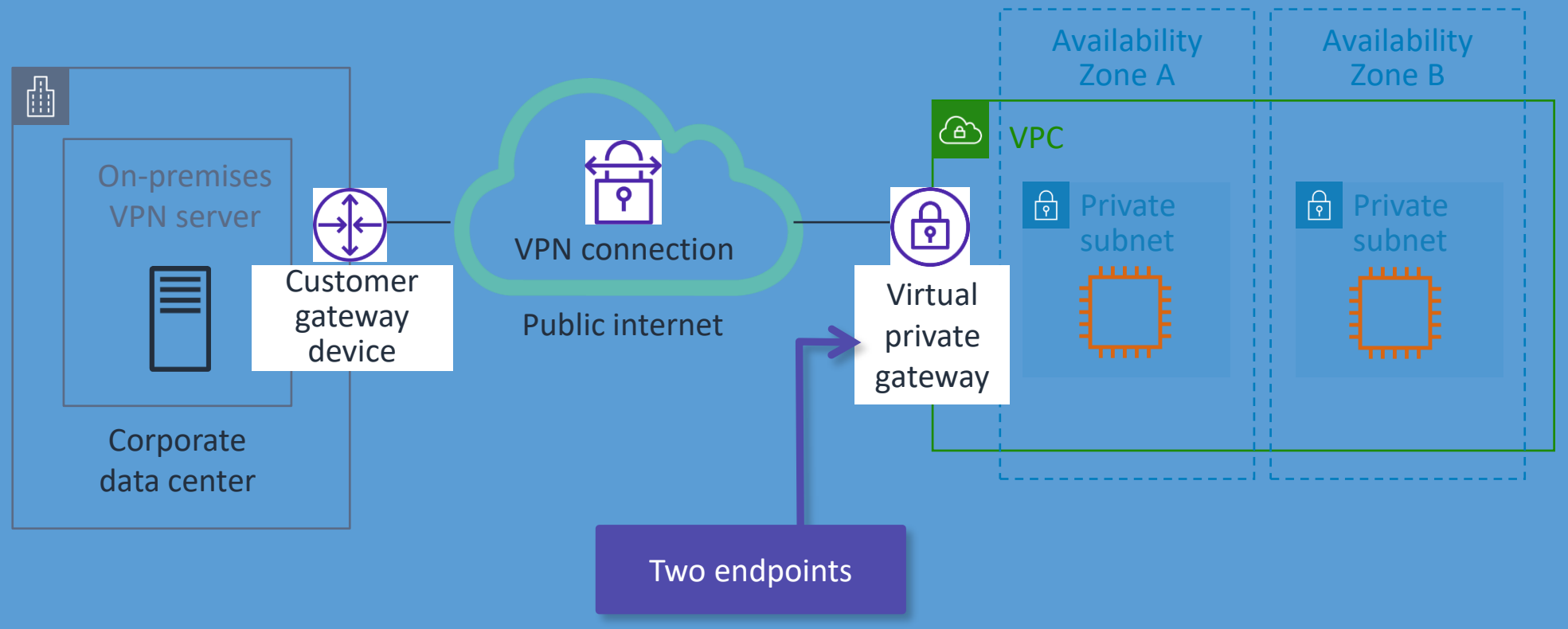
125

Hybrid networking

“What are our options to connect our on-premises network to the AWS Cloud?”

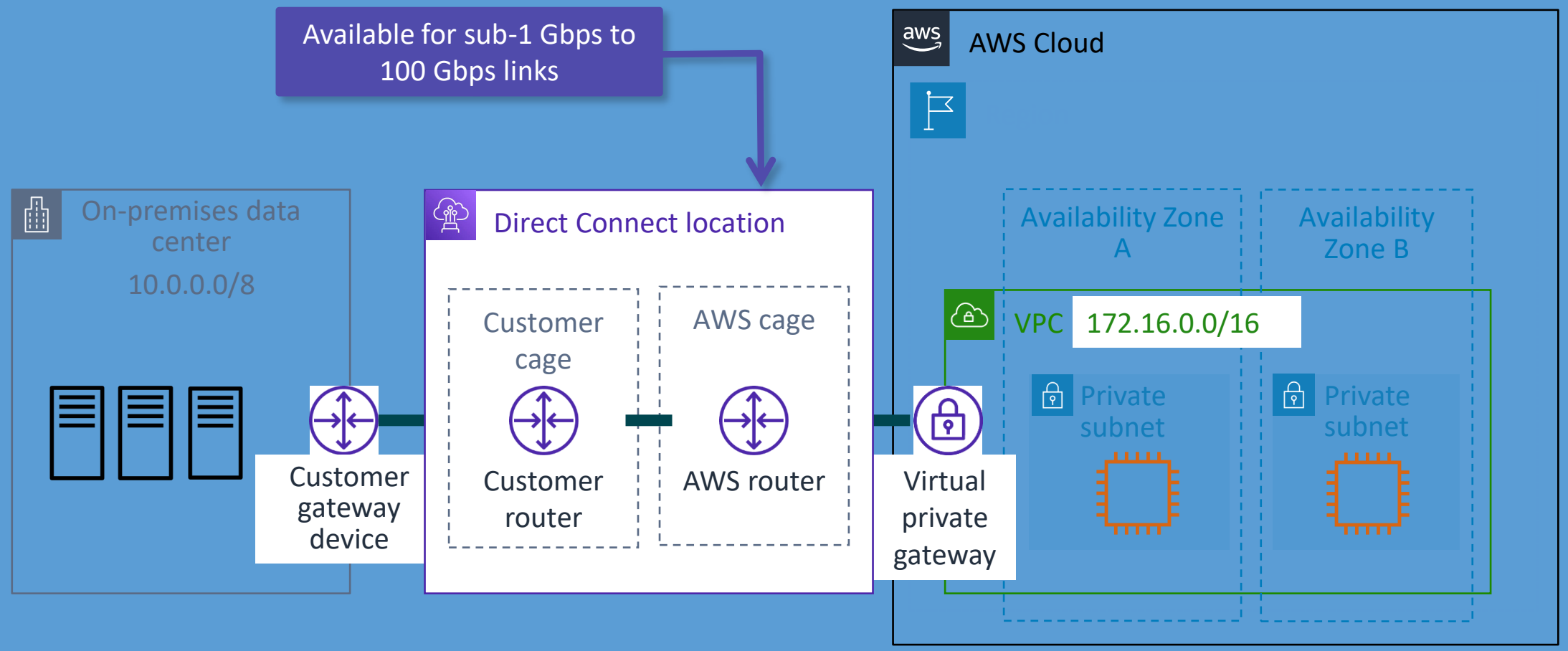
AWS Site-to-Site VPN

- Managed connection
- Static or dynamic VPN



AWS Direct Connect

Create a fiber link from your data center to your AWS resources.



Direct Connect and AWS Site-to-Site VPN pricing

Direct Connect

- Capacity (Mbps)
- Port hours
  - Time that a port is provisioned for your use in the data center
- Data transfer out (DTO)
  - Measured per gigabyte (GB)

Site-to-Site VPN

- Connection fee (per hour)
- Data transfer out (DTO)
  - Measured per gigabyte (GB)
  - First 100 GB are at no charge

Choosing AWS VPN or Direct Connect

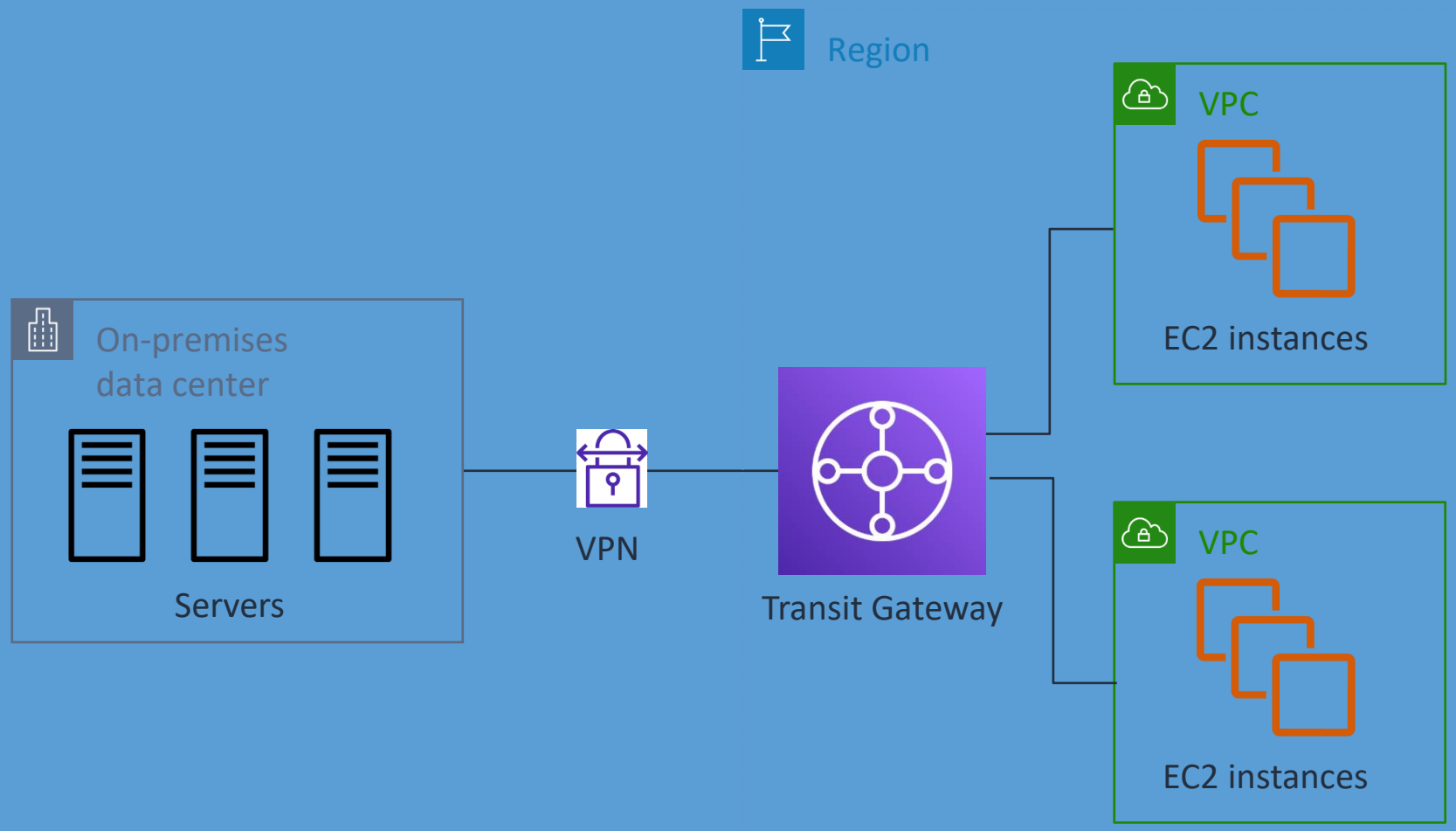
AWS Site-to-Site VPN	Direct Connect
Limited to 1.25 Gbps connection maximum	Sub-1, 1, 10, or 100 Gbps connection options
Faster to configure than Direct Connect	Requires special agreements and physical cabling to the data center
Don’t have to pay for inactive connections	Pay for port hours whether the connection is active or not
Encrypted in transit by default, but travels over public internet	Not encrypted by default, but it’s a private, dedicated connection

AWS Transit Gateway

“Which services can reduce the number of route tables we need to manage our global network?”

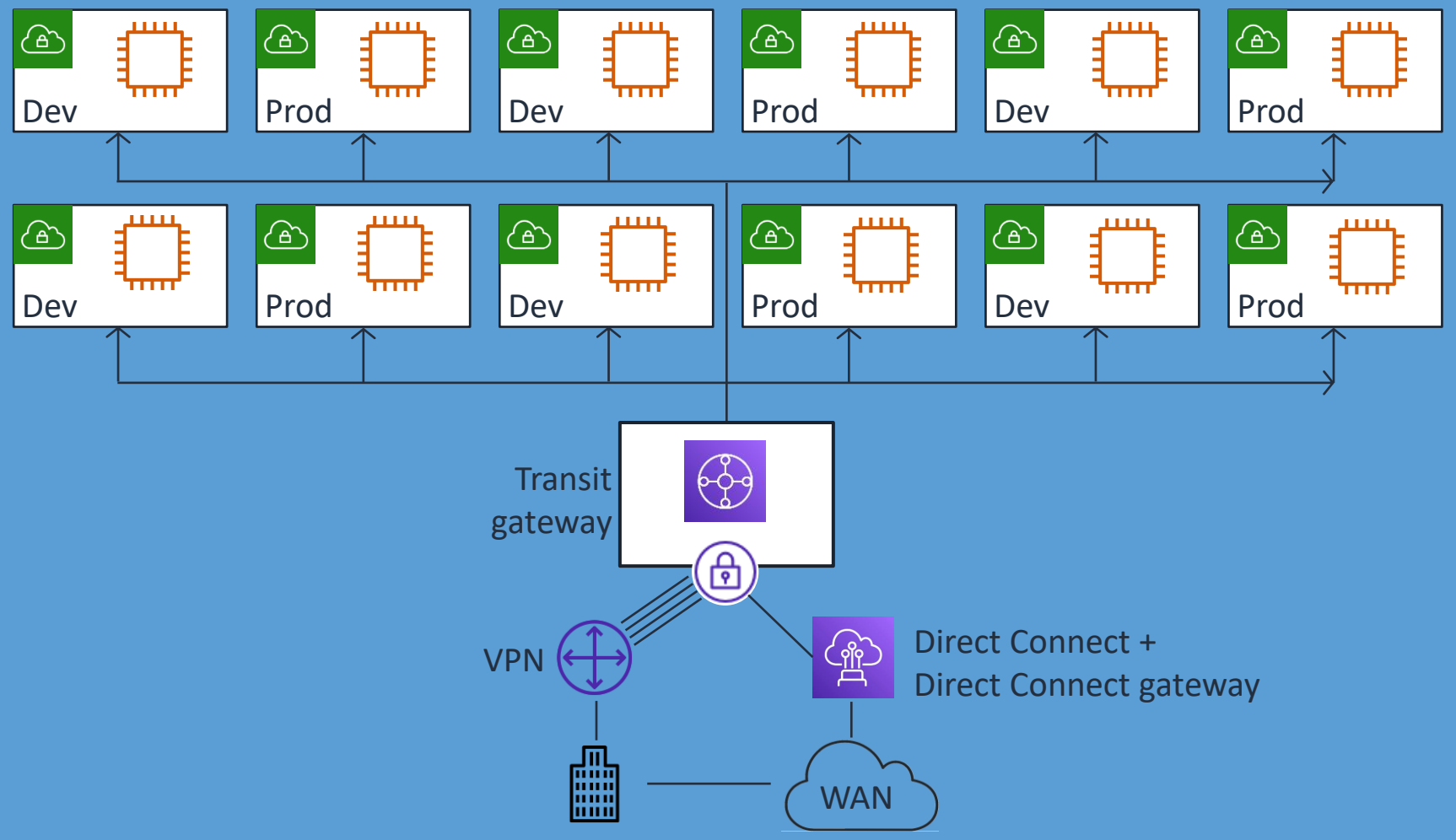
Transit Gateway

- Connects up to 5,000 VPCs and on-premises environments
- Acts as a hub for all traffic to flow through
- Allows multicast and inter-Region peering

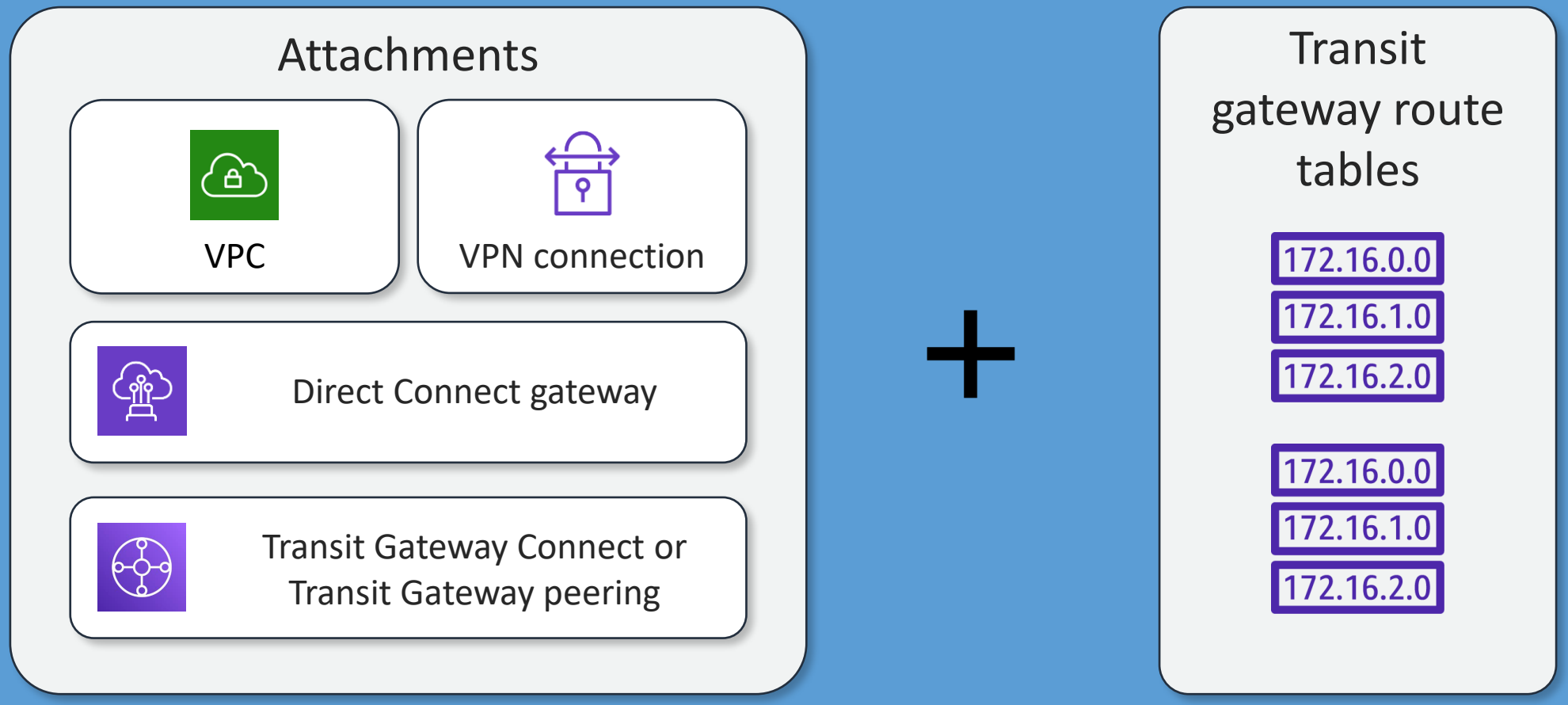


Scaling your network with Transit Gateway

- Attachment-based
- Flexible routing and segmentation
- Simplified connections
- Highly available and scalable

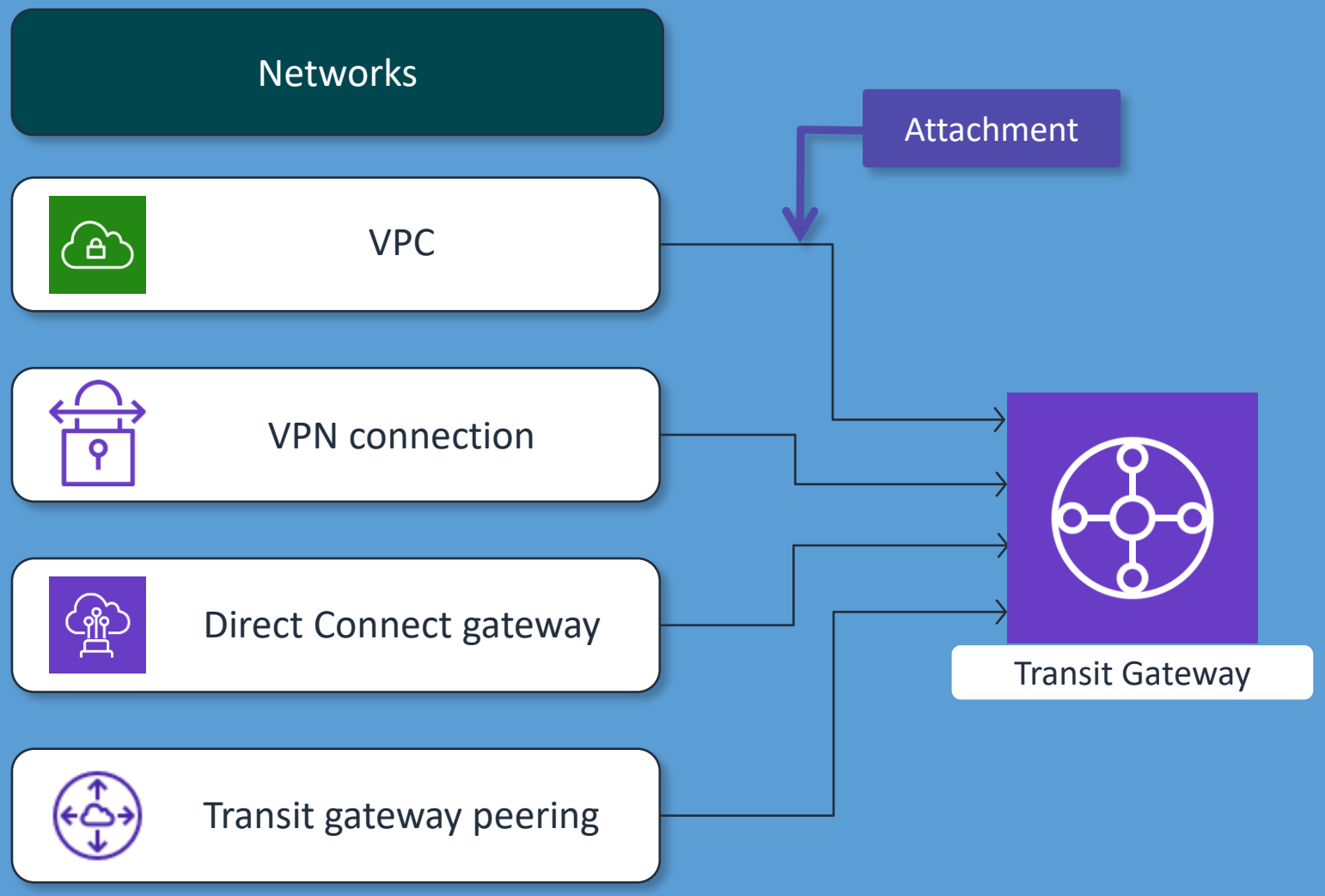


Transit Gateway components

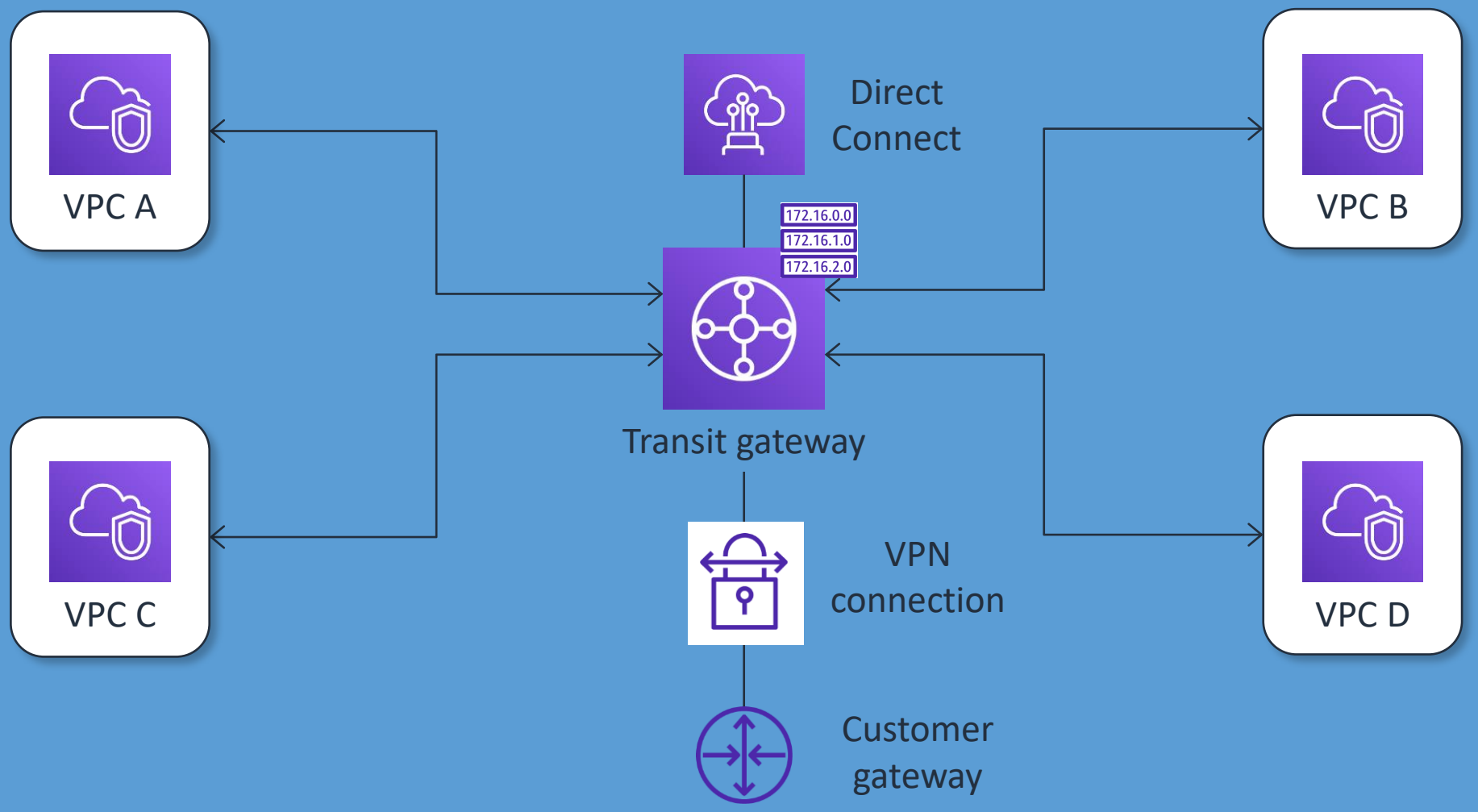


Transit Gateway setup

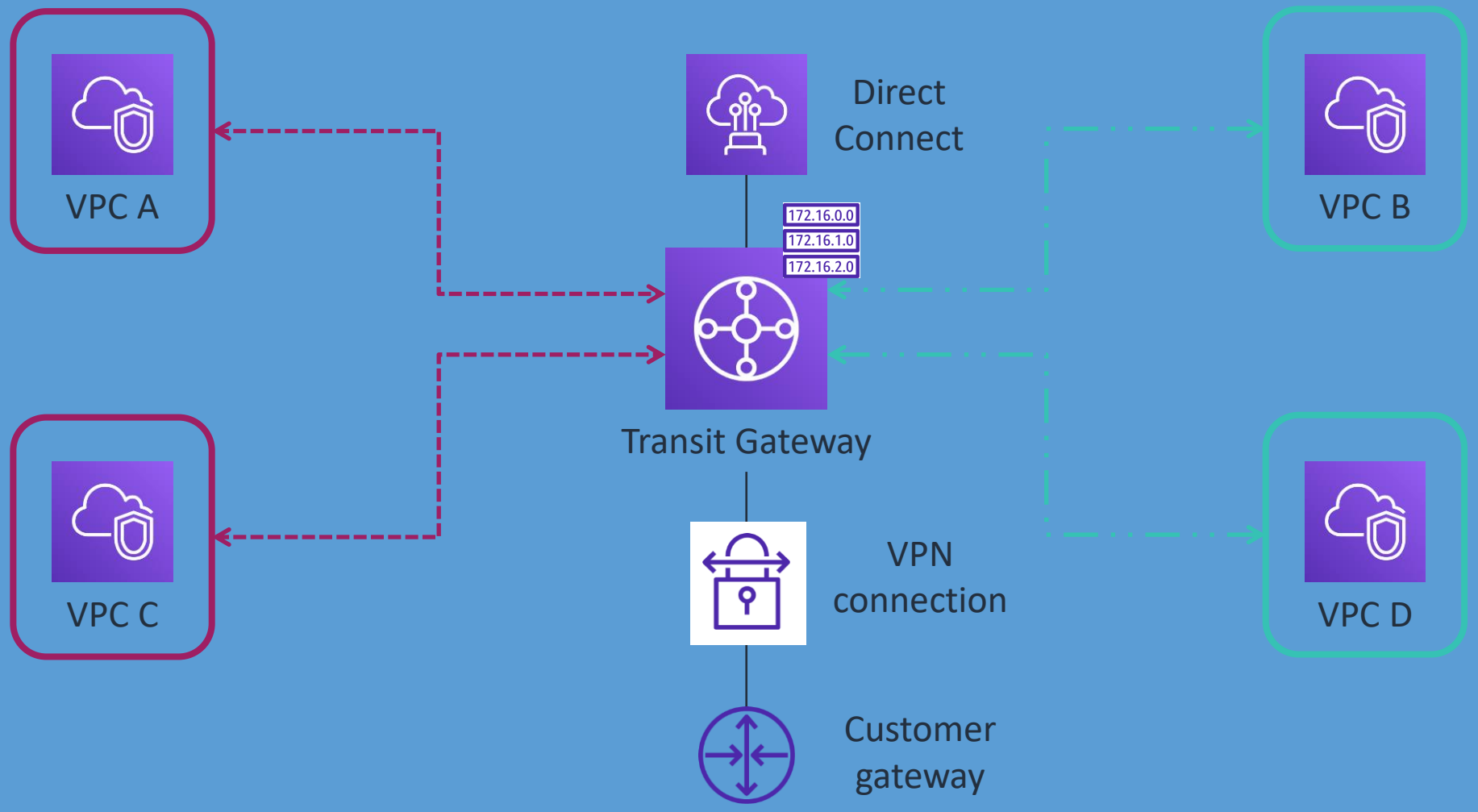
- Attach VPCs, VPN, Direct Connect gateway, and transit gateway peering connections.
- Network attachments must be in the same Region as the transit gateway.



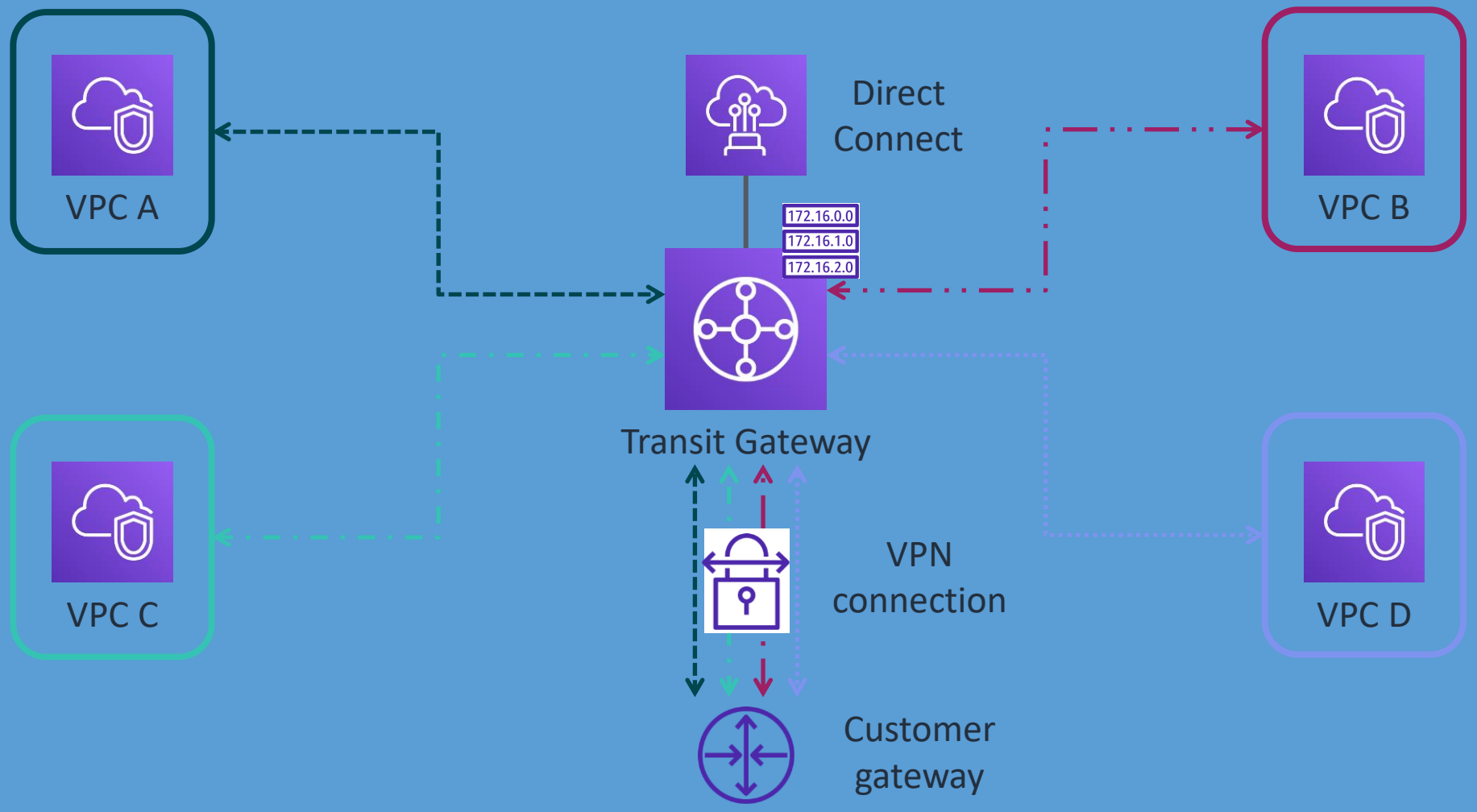
Full connectivity



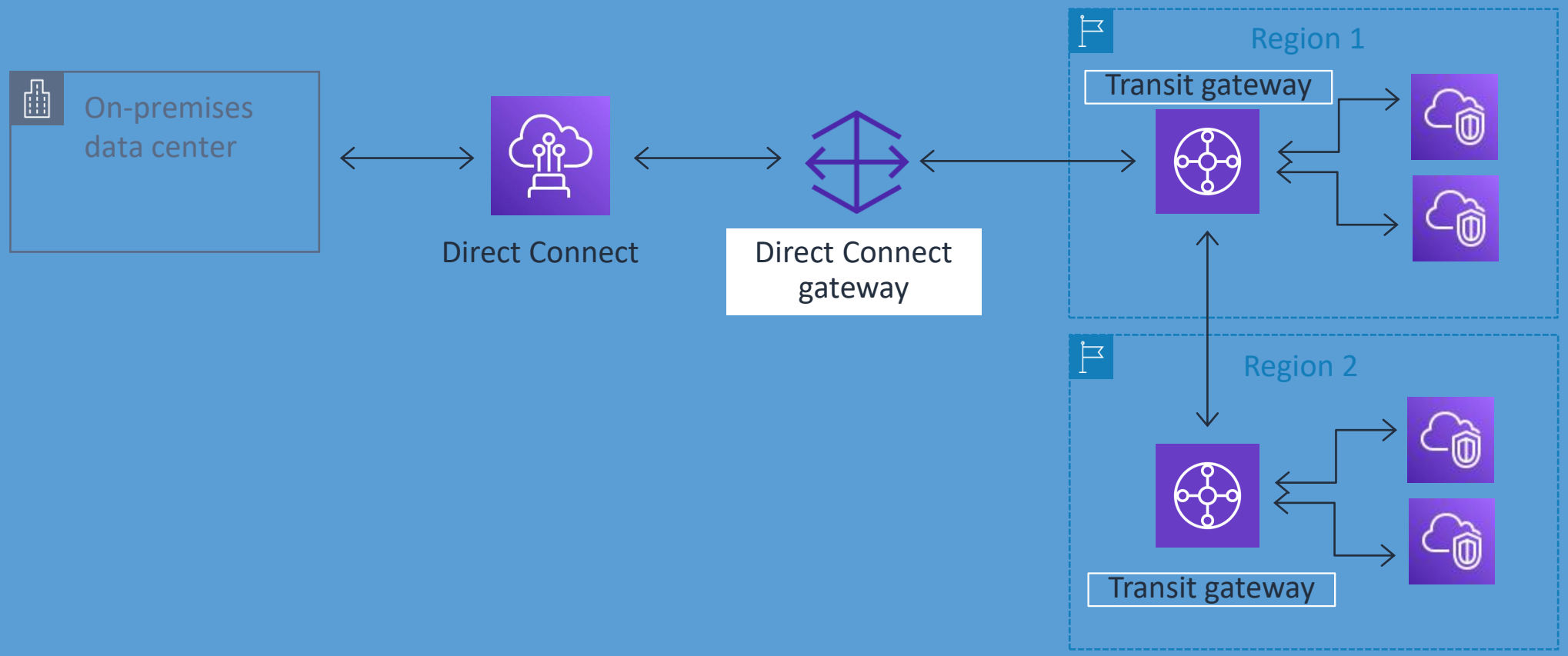
Partial connectivity



Isolation with full access from a VPN



Example global network architecture



Review

Present solutions

Network Engineer

Consider how you would answer the following:

- What can we do to keep our connections to AWS services private?
- How can we privately route traffic between our VPCs?
- What are our options to connect our on-premises network to the AWS Cloud?
- Which services can reduce the number of route tables we need to manage our global network?

Module review

In this module you learned about:

- ✓ VPC endpoints
- ✓ VPC peering
- ✓ Hybrid networking
- ✓ Transit Gateway

Next, you will review:

Knowledge check

Knowledge check

Knowledge check question 1

What is a connection to a transit gateway called?

A	VPN
B	Attachment
C	Route
D	VPC

Knowledge check question 1 and answer

What is a connection to a transit gateway called?

A	VPN
B correct	Attachment
C	Route
D	VPC

Knowledge check question 2

What are the components of an AWS Site-to-Site VPN connection? (Select TWO.)

A	Customer gateway device
B	Interface endpoint
C	Virtual private gateway
D	VPC peering connection
E	Gateway endpoint

Knowledge check question 2 and answer

What are the components of an AWS Site-to-Site VPN connection? (Select TWO.)

A correct	Customer gateway device
B	Interface endpoint
C correct	Virtual private gateway
D	VPC peering connection
E	Gateway endpoint

Knowledge check question 3

What is true of VPC peering connections? (Select TWO.)

A	Connections are one-to-many.
B	Connections are one-to-one.
C	Connections require a transit gateway.
D	Connections can span accounts.
E	Connections are transitive.

Knowledge check question 3 and answer

What is true of VPC peering connections? (Select TWO.)

A	Connections are one-to-many.
B correct	Connections are one-to-one.
C	Connections require a transit gateway.
D correct	Connections can span accounts.
E	Connections are transitive.

AWS

Serverless

Lab 5

Question

Which of the following best describes your familiarity with serverless architectures?

A. I have built solutions using serverless architectures.
B. I understand serverless architectures, but have not used them.
C. I know a little bit about serverless architectures.
D. I am not familiar with serverless architectures.

Module overview

- Business request
- What is serverless?
- Amazon API Gateway
- Amazon Simple Queue Service (Amazon SQS)
- Amazon Simple Notification Service (Amazon SNS)
- Amazon Kinesis
- AWS Step Functions
- Present solutions
- Knowledge check
- Lab 5: Build a serverless architecture

Business Requirements

Application Development Manager

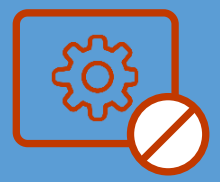
The application development manager wants to know:

- How can we reduce operational overhead and optimize our resource costs?
- What is a secure way to provide APIs that use our backend services?
- How do we create a message queue for reliable service-to-service communication?
- How can we give our applications the ability to send push notifications?
- How do we ingest streaming data to power our real-time applications?
- What is an easy way to orchestrate multi-step workflows?

What is serverless?

“How can we reduce operational overhead and optimize our resource costs?”

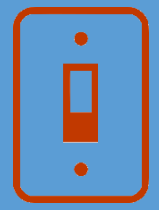
What is serverless?



No infrastructure to provision or manage

Scales automatically by unit of consumption

Consumption-based pricing



Built-in security; highly available compute

AWS serverless portfolio

Compute

AWS Lambda

AWS Fargate

API proxy

Amazon API Gateway

Storage

Amazon Simple Storage Service (Amazon S3)

Database

Amazon DynamoDB

Amazon Aurora Serverless

Authentication

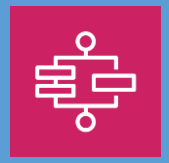
Amazon Cognito

Interprocess messaging

Amazon Simple Notification Service (Amazon SNS)

Amazon Simple Queue Service (Amazon SQS)

Orchestration



AWS Step Functions

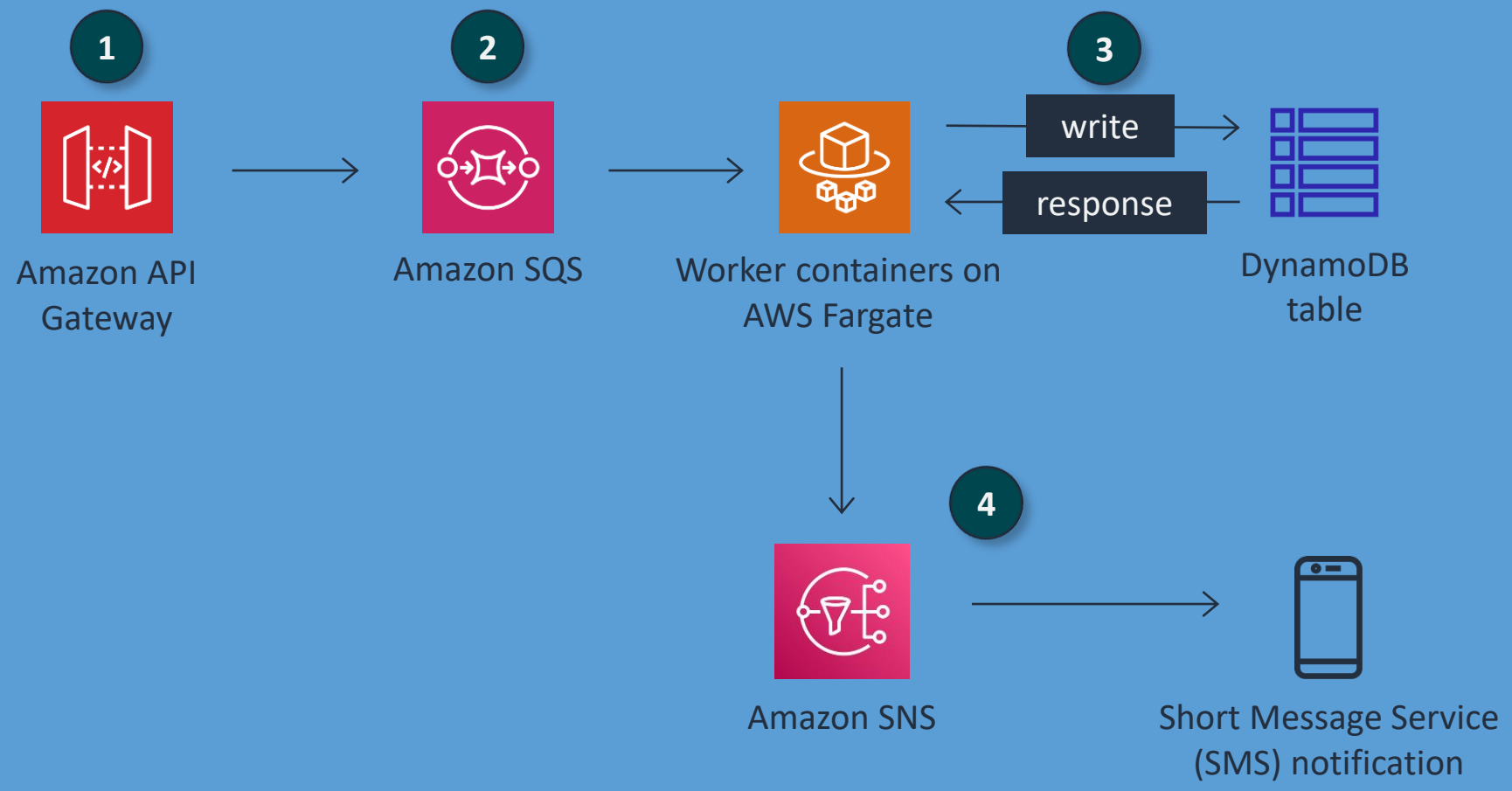
Analytics

Amazon Kinesis

Amazon Athena

Example serverless architecture

1. POST request received
2. Request goes to a message queue to await processing by a worker service
3. Worker service processes message and writes it to Amazon DynamoDB
4. Prompts the notification service to send an SMS notice to subscribed users

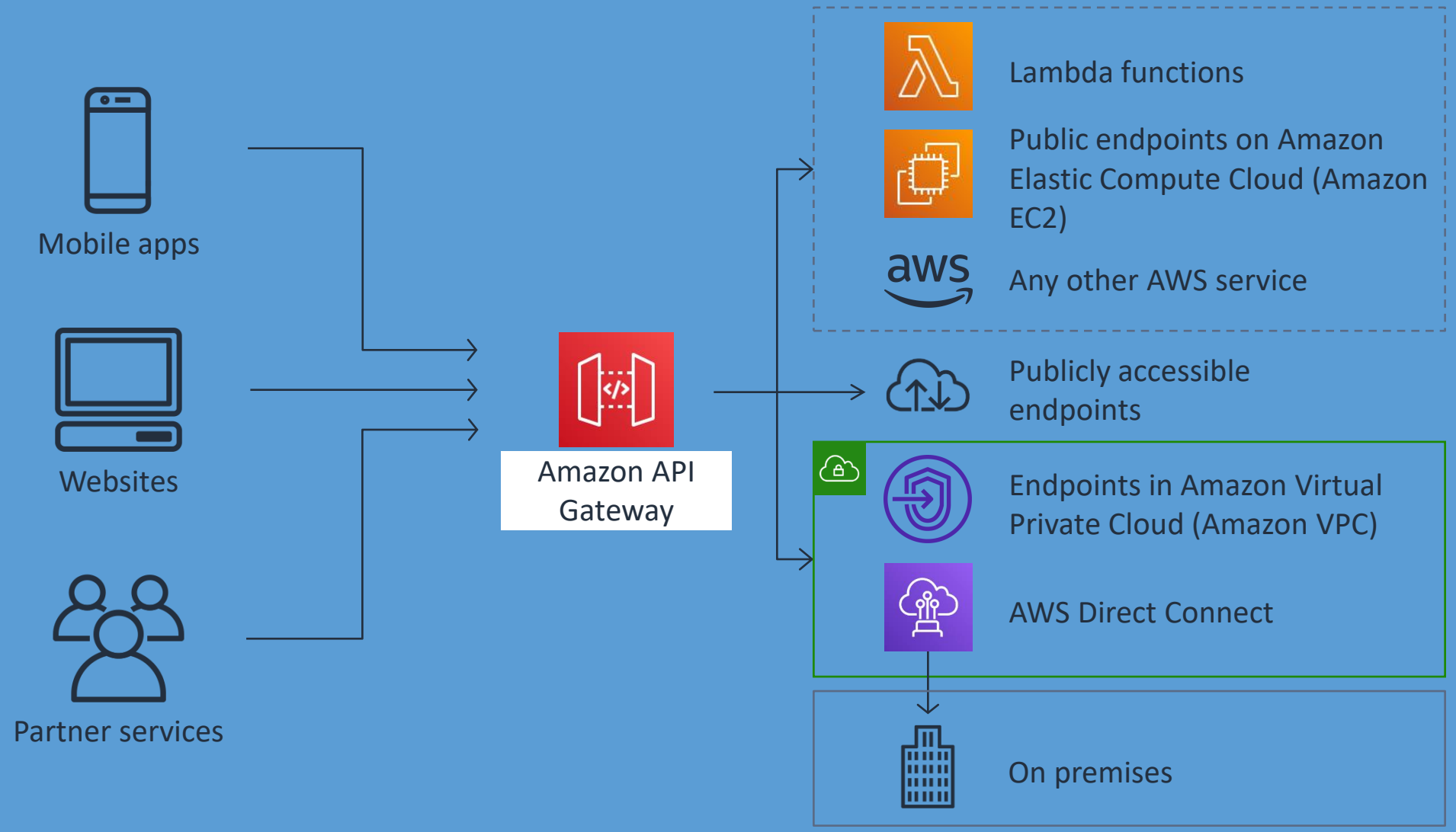


Amazon API Gateway

“What is a secure way to provide APIs that use our backend services?”

API Gateway

- Create an entry point for your applications.
- Process thousands of concurrent API calls.
- Choose internet facing or internal only.



API Gateway features

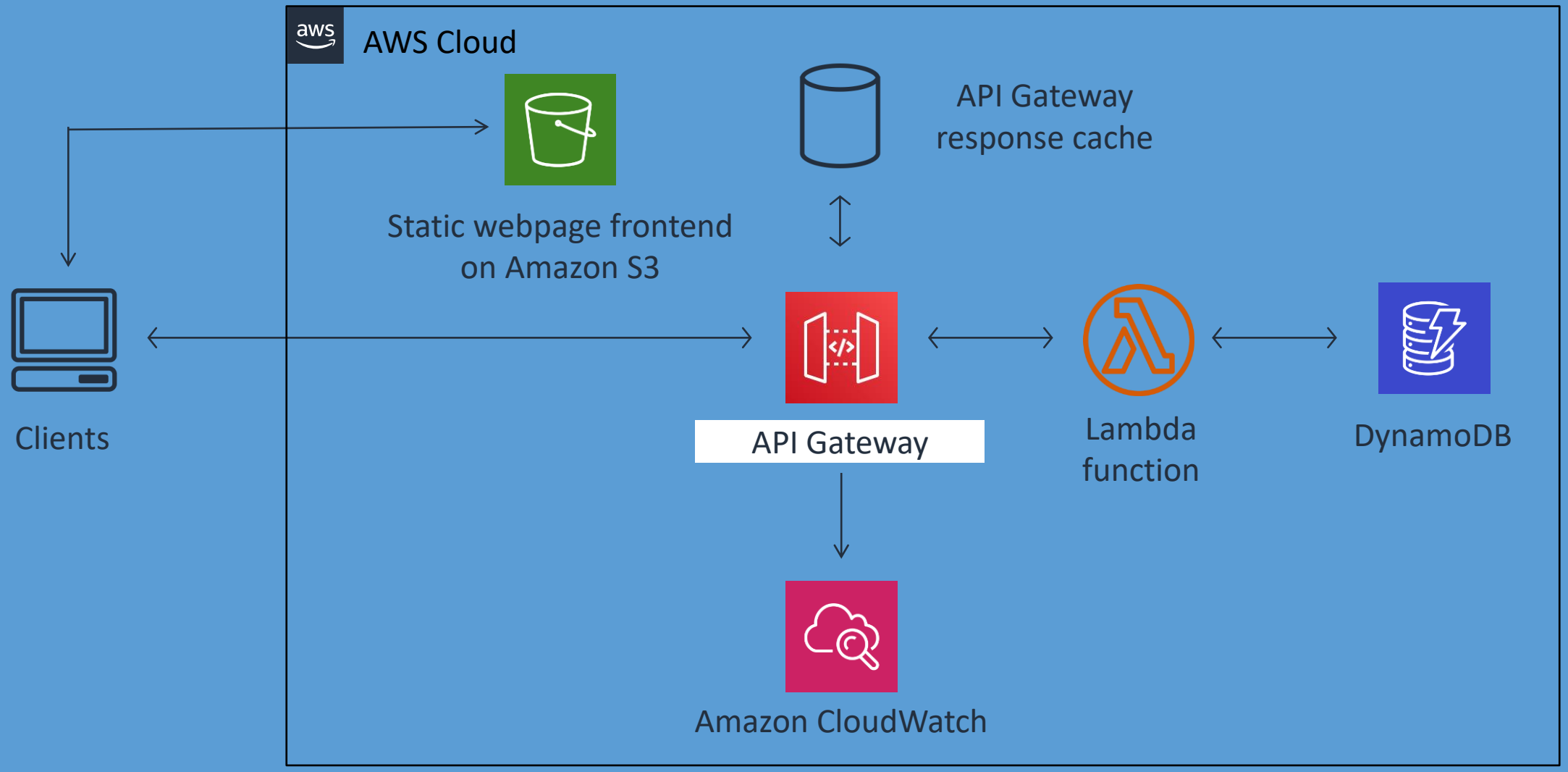
Creates a unified API frontend for multiple microservices

Provides distributed denial of service (DDoS) protection and throttling for your backend

Authenticates and authorizes requests to a backend

Throttles, meters, and monetizes API usage by third-party developers

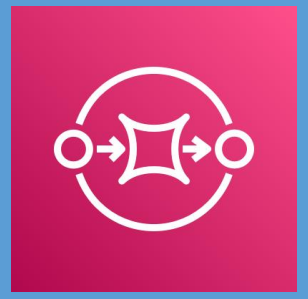
API Gateway sample architecture



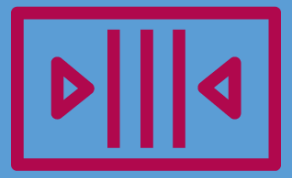
Amazon SQS

“How do we create a message queue for reliable service-to-service communication?”

Amazon Simple Queue Service (Amazon SQS)



Amazon SQS



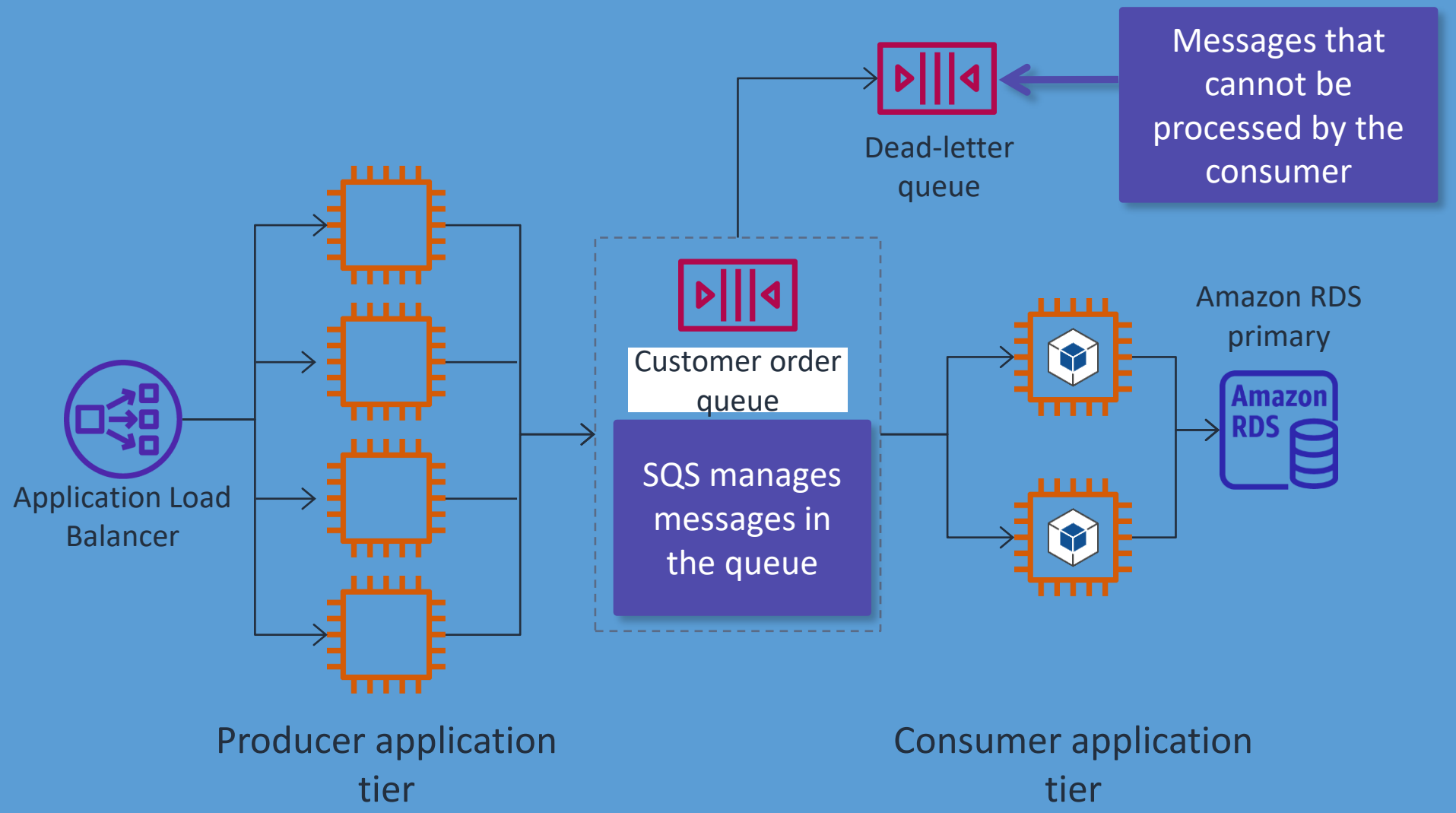
Fully managed message queueing service

Stores messages until they are processed and deleted

Acts as a buffer between senders and receivers

Loose coupling with Amazon SQS

- Loosely couples application components
- Uses asynchronous processing
- Creates tolerance for failed steps
- Absorbs demand spikes

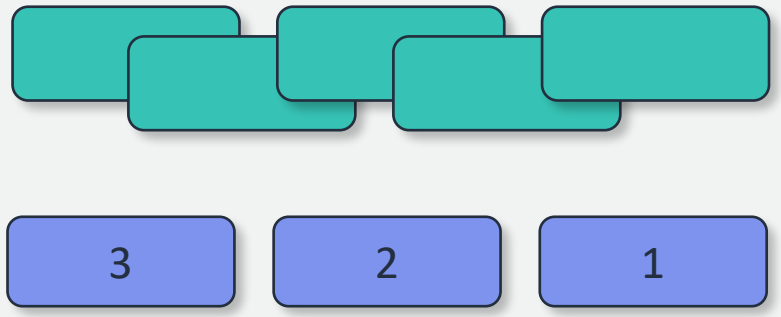


Amazon SQS use cases

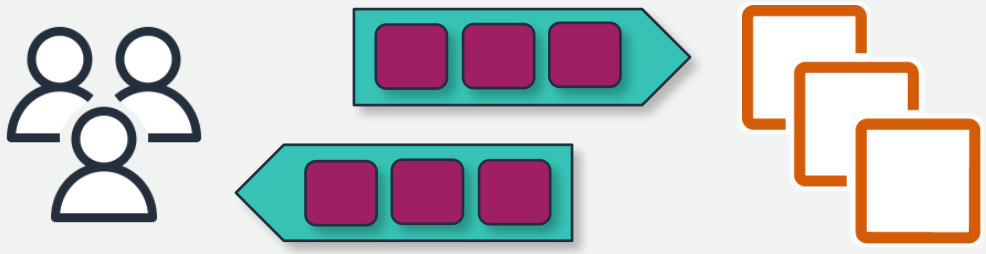
Work queues

Standard

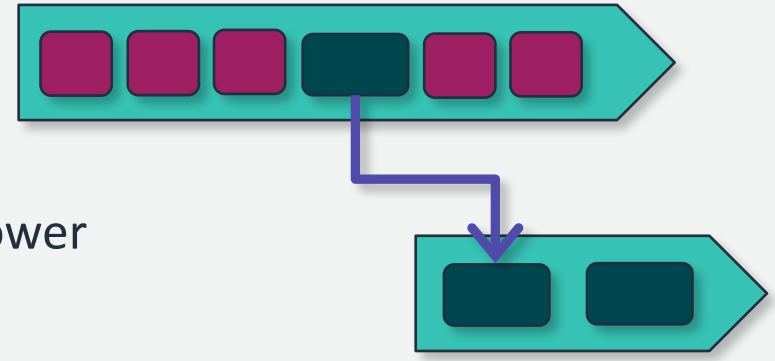
FIFO



Buffering and batch operations



Request offloading

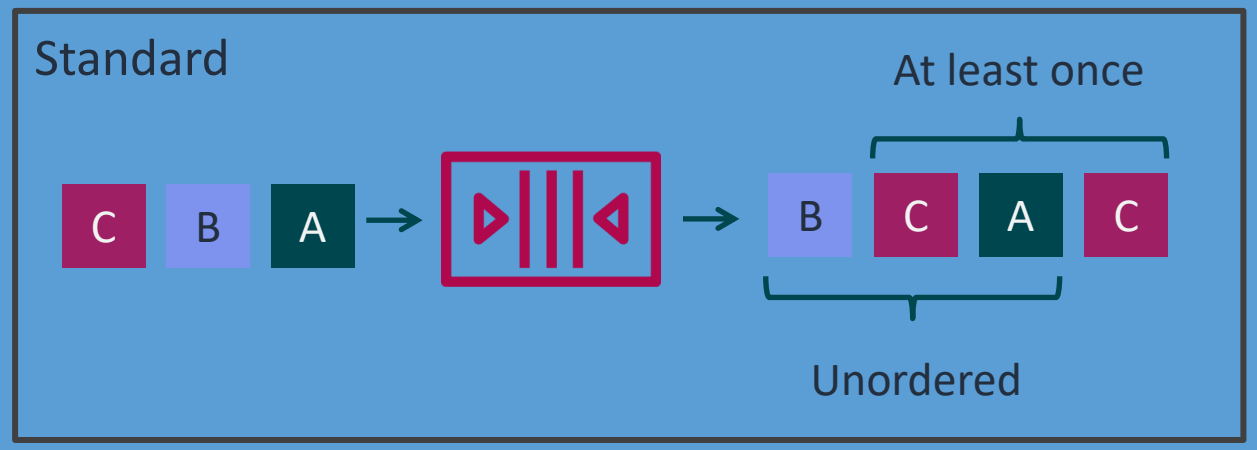


Enqueue slower requests.

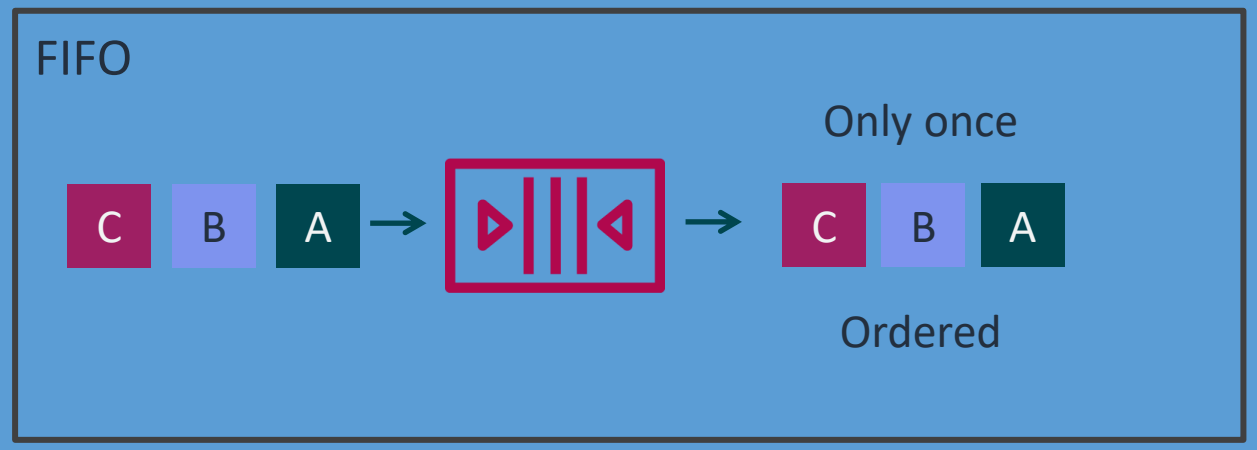
Auto scaling

Add another process to scale up the rate of messages.

SQS queue types



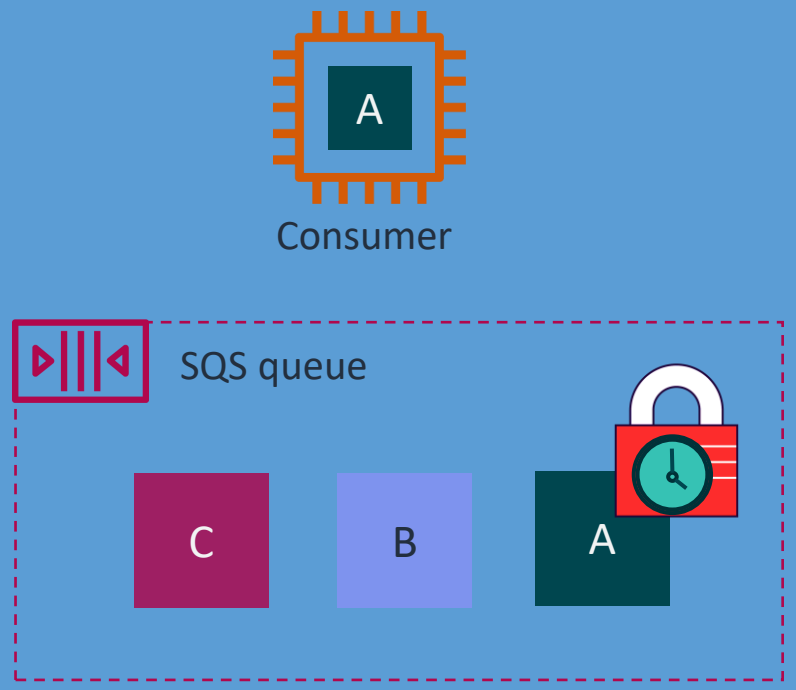
Nearly unlimited API calls per second



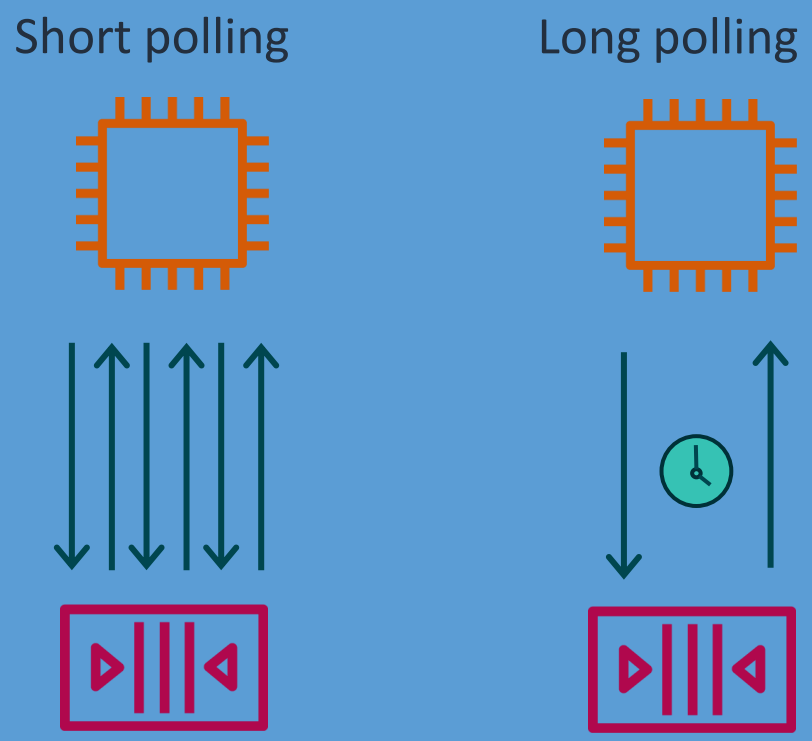
Limited API calls per second

Optimizing your Amazon SQS queue configurations

Tune your visibility timeout



Choose the right polling type



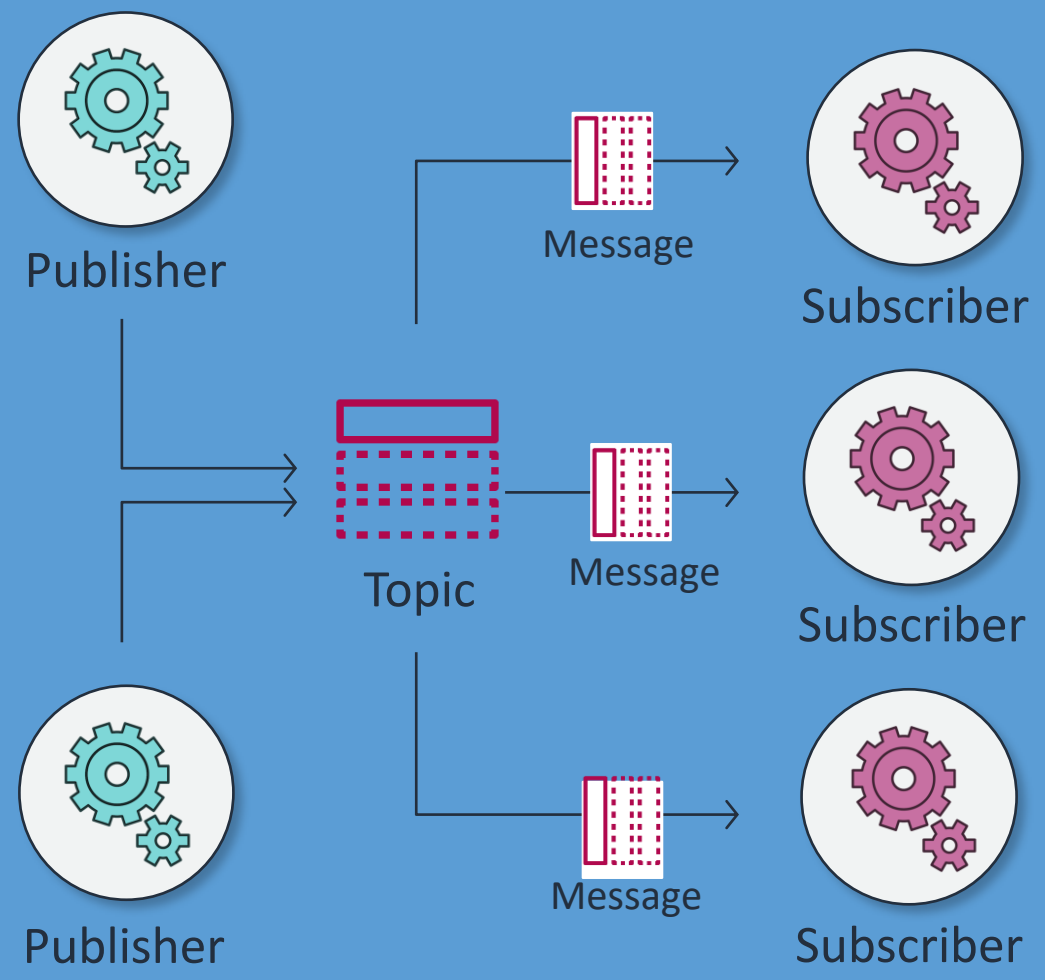
When to use message queues

✓ Service-to-service communication	✗ Selecting specific messages
✓ Asynchronous work items	✗ Large messages
✓ State change notifications	

Amazon SNS

“How can I give our applications the ability to send push notifications?”

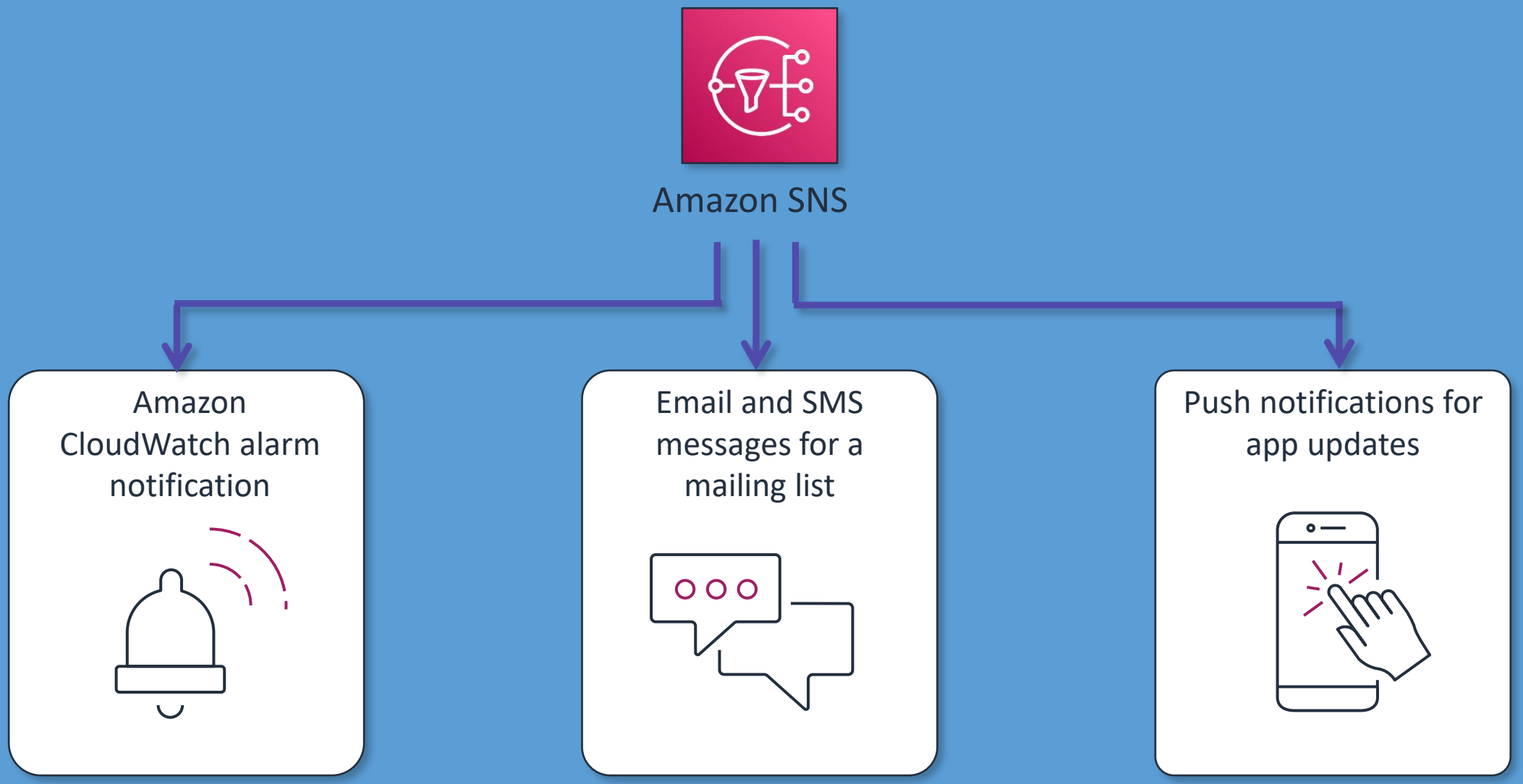
Amazon Simple Notification Service (Amazon SNS)



Types of subscribers

- Email/Email-JSON
- Mobile text messaging (SMS)
- Mobile push notification
- HTTP/HTTPS
- AWS Lambda
- Amazon SQS
- Kinesis Data Firehose

Use cases for Amazon SNS

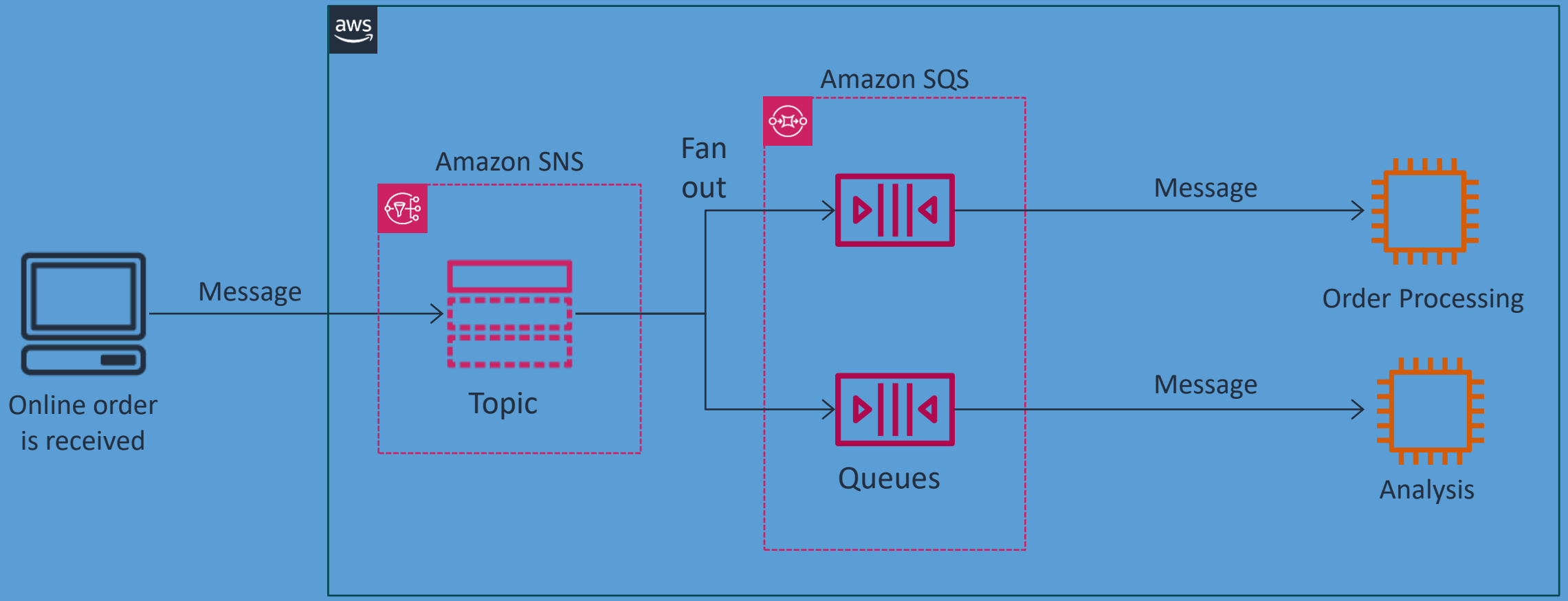


Characteristics of Amazon SNS

- Single published message
- No recall options
- HTTP or HTTPS retry
- Standard or FIFO topics

Amazon SNS publish to multiple SQS queues

Architecture example



Amazon SNS and Amazon SQS

Features	Amazon SNS	Amazon SQS
Message persistence	No	Yes
Delivery mechanism	Push (passive)	Poll (active)
Producer and consumer	Publisher and subscriber	Send or receive
Distribution model	One to many	One to one

Amazon Kinesis

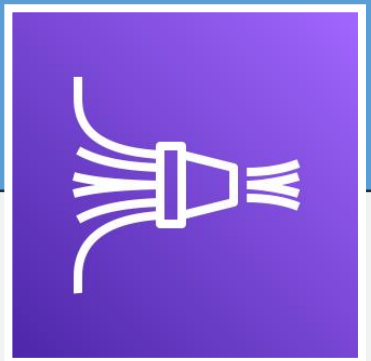
“How do we ingest streaming data to power our real-time applications?”

Kinesis for data collection and analysis

Amazon Kinesis Data Streams

Collect and store data streams for analytics.

Amazon Kinesis Data Firehose

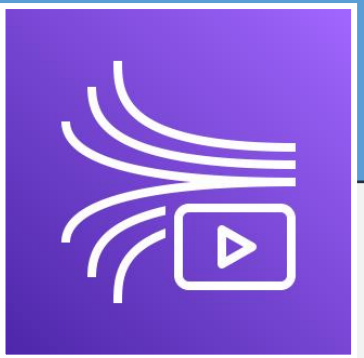


Load data streams into AWS data stores.

Amazon Kinesis Data Analytics

Analyze data streams with SQL or Apache Flink.

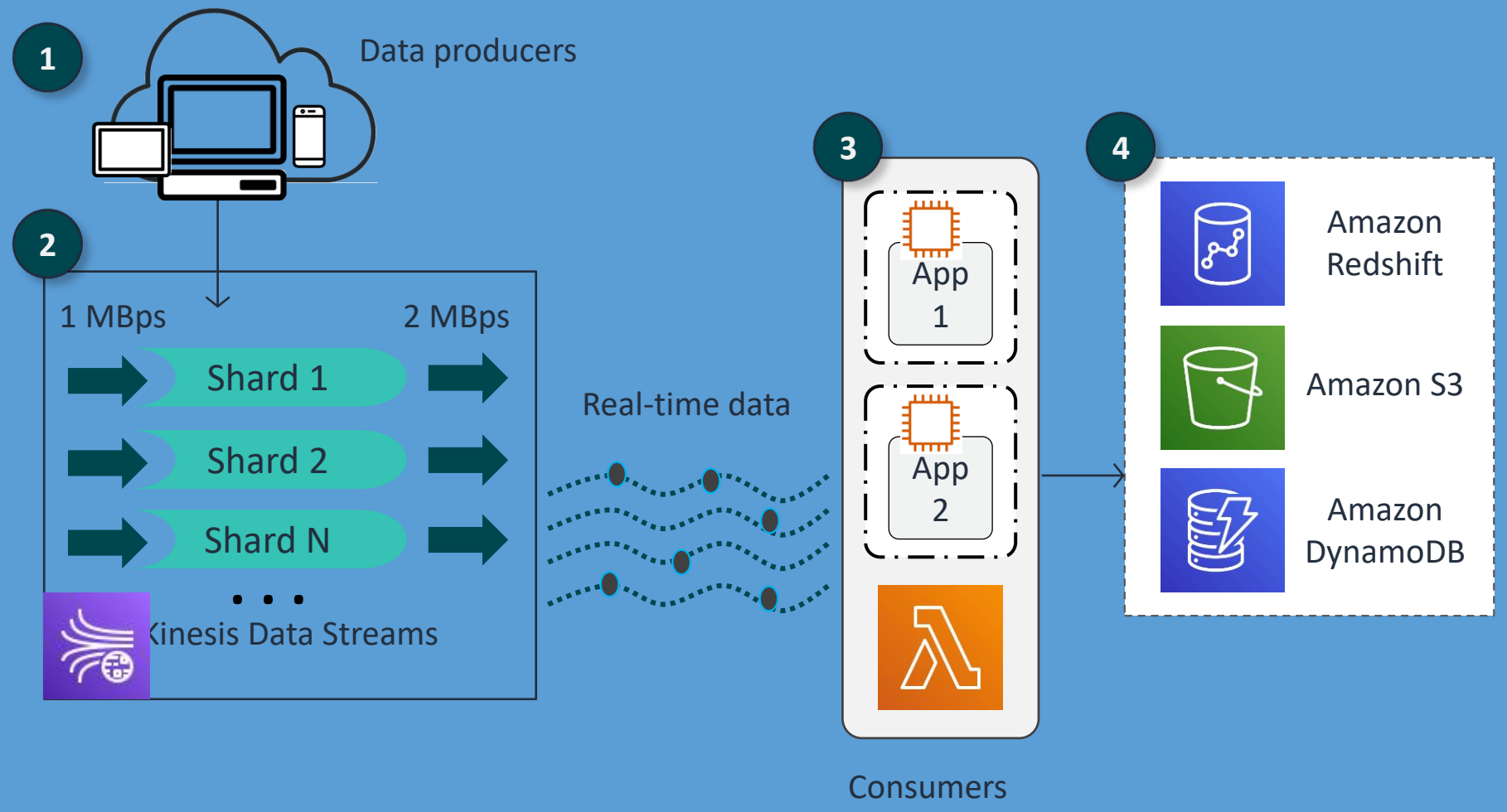
Amazon Kinesis Video Streams



Collect and store video streams for analytics.

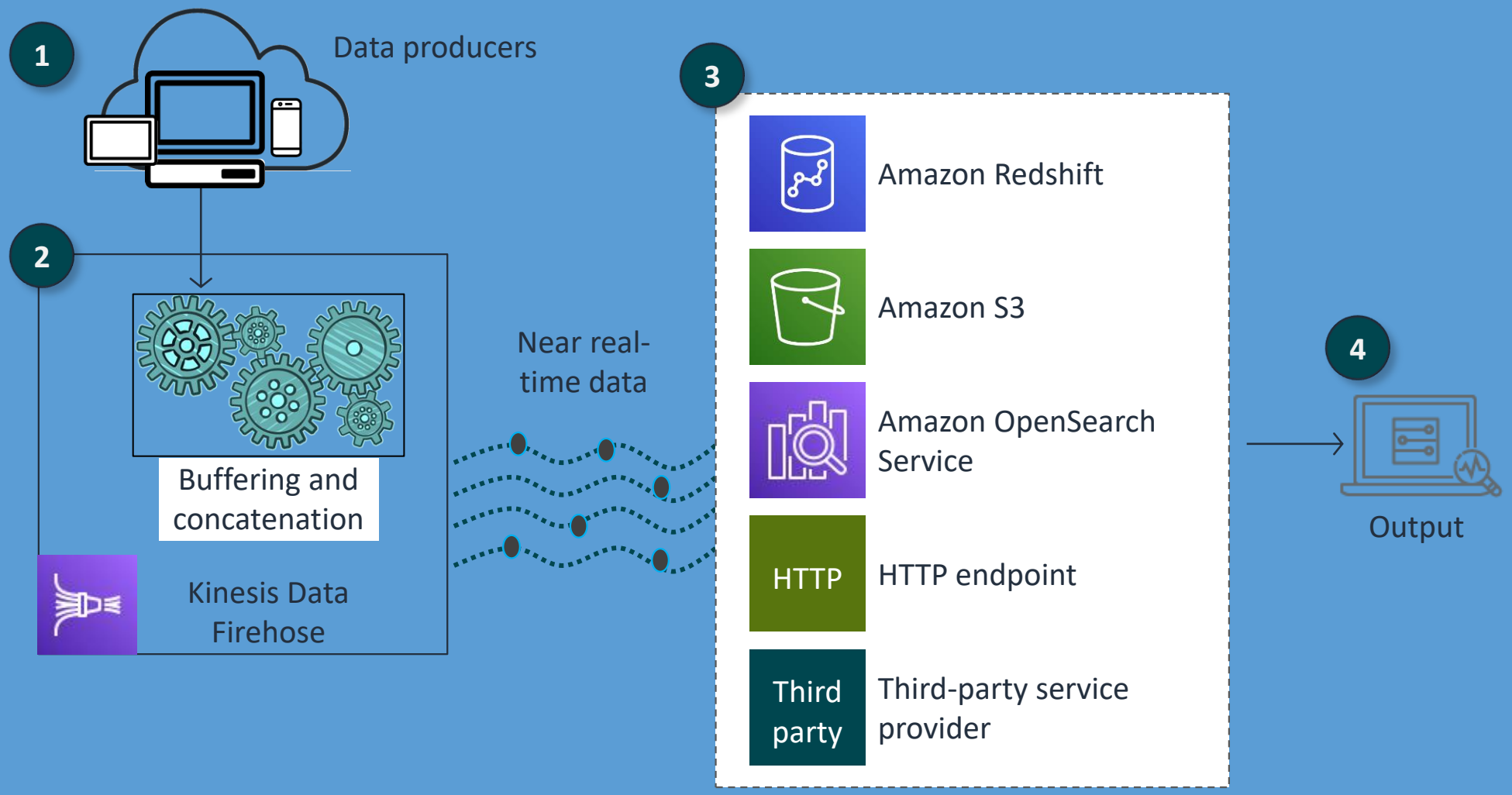
Kinesis Data Streams overview

1. Producers put data records into Kinesis Data Streams.
2. Shards hold real-time, sequenced data.
3. Consumers read from shards and process data.
4. Output can be stored using AWS services.



Kinesis Data Firehose overview

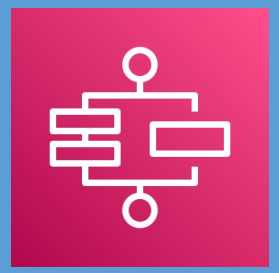
1. Data producers send data.
2. Data can be batched and compressed before loading it into AWS.
3. Kinesis Data Firehose writes to the destination.
4. Streaming data is processed using analytics and business intelligence.



AWS Step Functions

“What is an easy way to orchestrate multi-step workflows?”

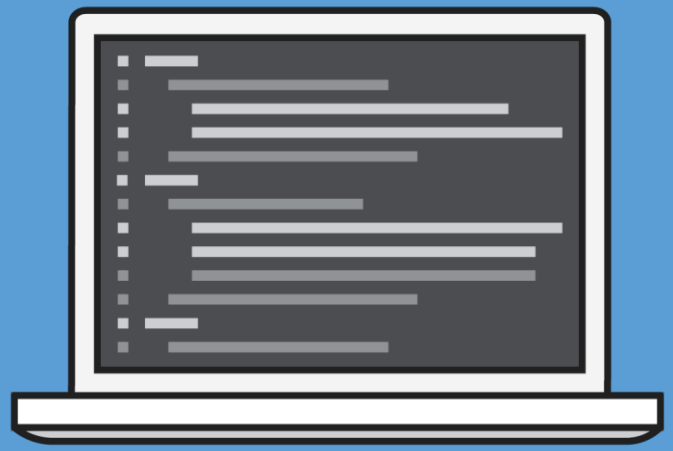
Step Functions



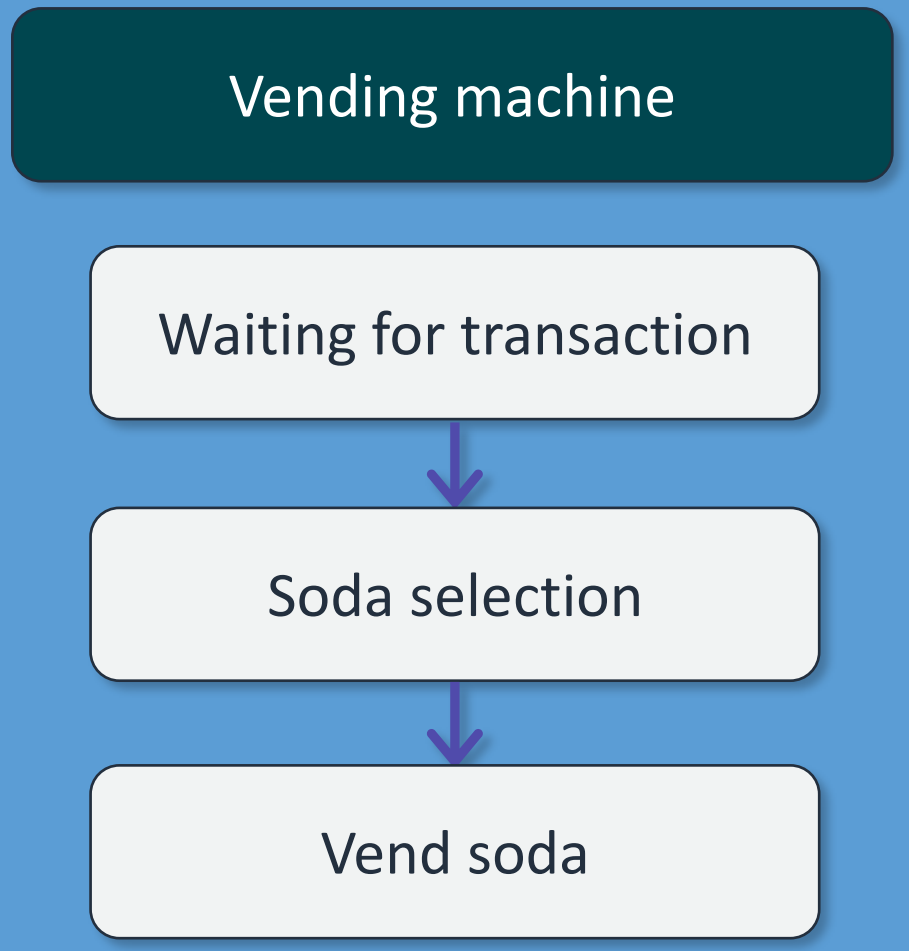
Step Functions

- Coordinates microservices using visual workflows
- Permits you to step through the functions of your application
- Automatically initiates and tracks each step
- Provides simple error catching and logging if a step fails

Step Functions: State machine



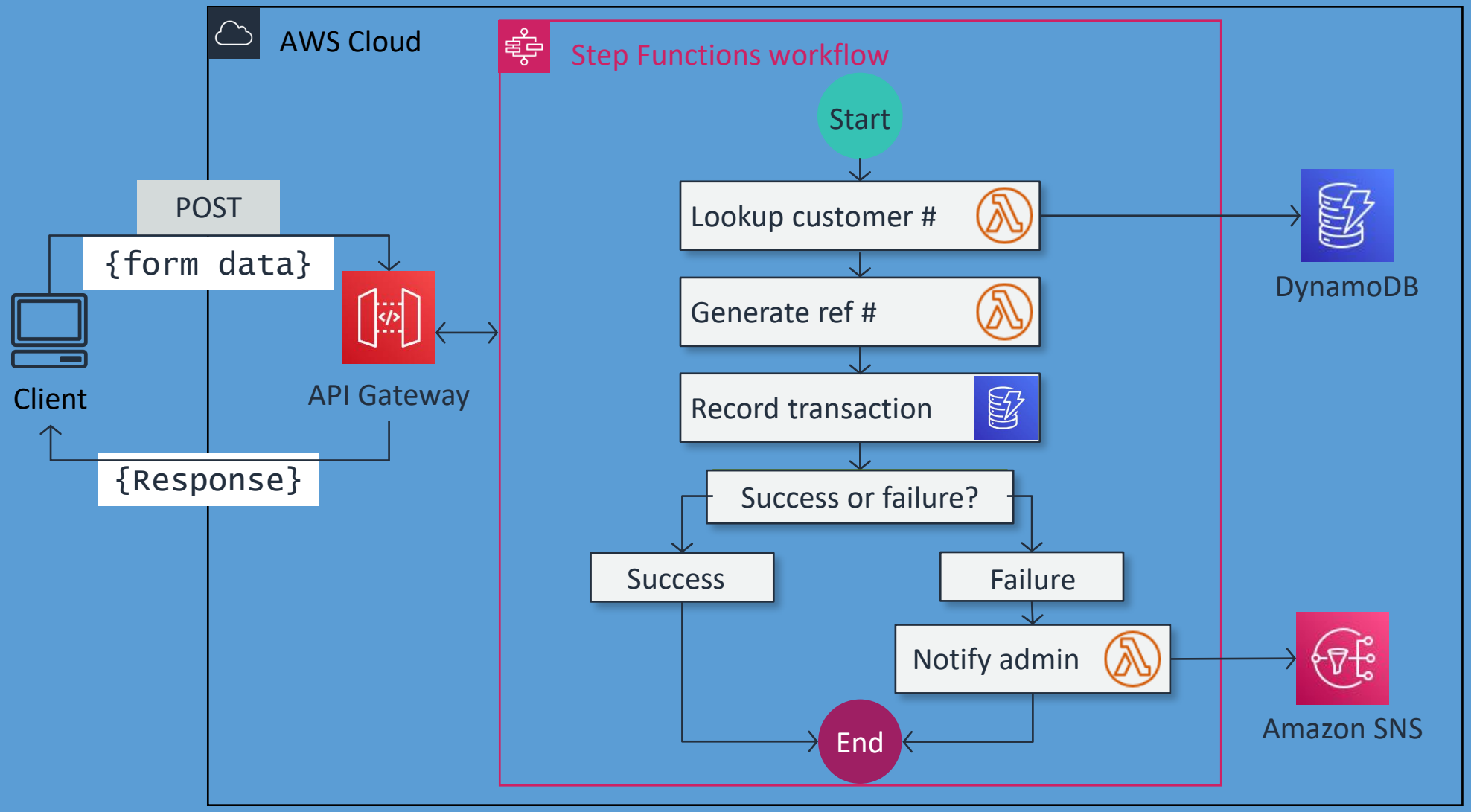
A *state machine* is an object that has a set number of operating conditions that depend on its previous condition to determine output.



Orchestration of complex distributed workflows

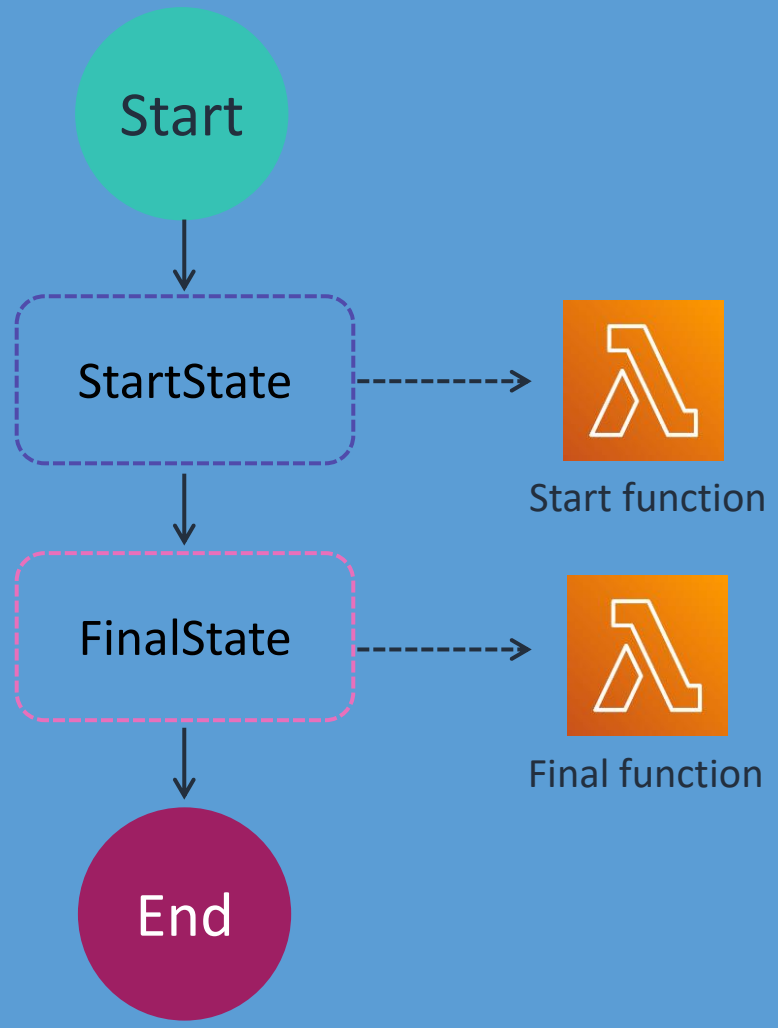
Step Functions supports the following state types:

- Task
- Choice
- Fail or Succeed
- Pass
- Wait
- Parallel
- Map



Amazon States Language

```
{
  "Comment": "An example of the ASL.",
  "StartAt": "StartState",
  "States": {
    "StartState": {
      "Type": "Task",
      "Resource": "arn:aws:lambda:us-east…",
      "Next": "FinalState"
    },
    "FinalState": {
      "Type": "Task",
      "Resource": "arn:aws:lambda:us-east…",
      "End": true
    }
  }
}
```



Review

Present solutions

Application Development Manager

Consider how you would answer the following:

- How can we reduce operational overhead and optimize our resource costs?
- What is a secure way to provide APIs that use our backend services?
- How do we create a message queue for reliable service-to-service communication?
- How can I give our applications the ability to send push notifications?
- How do we ingest streaming data to power our real-time applications?
- What is an easy way to orchestrate multi-step workflows?

Module review

In this module you learned:

- ✓ What is serverless?
- ✓ API Gateway
- ✓ Amazon SQS
- ✓ Amazon SNS
- ✓ Amazon Kinesis
- ✓ AWS Step Functions

Next, you will review:

- Knowledge check
- Lab introduction

Knowledge check

Knowledge check question 1

Which type of Amazon SQS queue provides at-least-once delivery?

A	FIFO queue
B	Standard queue
C	Dead-letter queue
D	Long polling

Knowledge check question 1 and answer

Which type of Amazon SQS queue provides at-least-once delivery?

A	FIFO queue
B correct	Standard queue
C	Dead-letter queue
D	Long polling

Knowledge check question 2

What is an advantage of long polling compared to short polling?

A	Long polling provides an immediate response from a `ReceiveMessage` call.
B	Long polling is more stable when using a single thread to poll multiple queues.
C	Long polling reduces the cost of using Amazon SQS by reducing the number of empty responses and false empty responses.
D	Long polling reduces cost by only sampling a subset of Amazon SQS servers.

Knowledge check question 2 and answer

What is an advantage of long polling compared to short polling?

A	Long polling provides an immediate response from a `ReceiveMessage` call.
B	Long polling is more stable when using a single thread to poll multiple queues.
C correct	Long polling reduces the cost of using Amazon SQS by reducing the number of empty responses and false empty responses.
D	Long polling reduces cost by only sampling a subset of Amazon SQS servers.

Knowledge check question 3

What is a feature of Amazon SNS?

A	Amazon SNS exchanges messages through a polling model.
B	Amazon SNS can send messages to decoupled components of a distributed application that do not process the same amount of work simultaneously.
C	Amazon SNS can push messages to multiple subscribers.
D	Amazon SNS keeps messages persistent.

Knowledge check question 3 and answer

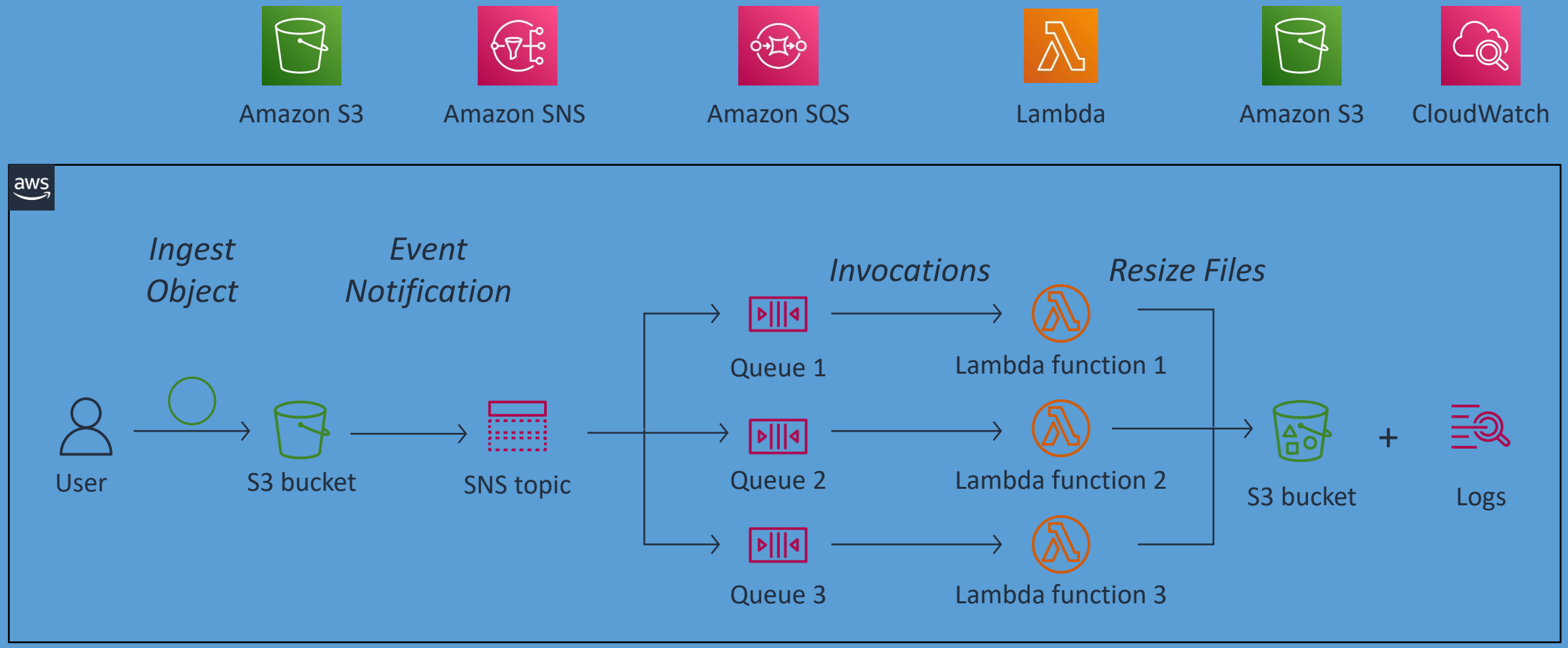
What is a feature of Amazon SNS?

A	Amazon SNS exchanges messages through a polling model.
B	Amazon SNS can send messages to decoupled components of a distributed application that do not process the same amount of work simultaneously.
C correct	Amazon SNS can push messages to multiple subscribers.
D	Amazon SNS keeps messages persistent.

Lab 5:

Build a serverless architecture

Lab 5 diagram



Lab tasks

Task 1: Create an Amazon SNS topic.

Task 2: Create three Amazon SQS queues and subscribe to the SNS topic.

Task 3: Create an Amazon S3 Event Notification to SNS.

Task 4: Create and configure three Lambda functions.

Task 5: Upload an object to the Amazon S3 bucket.

Task 6: Validate the processed file.