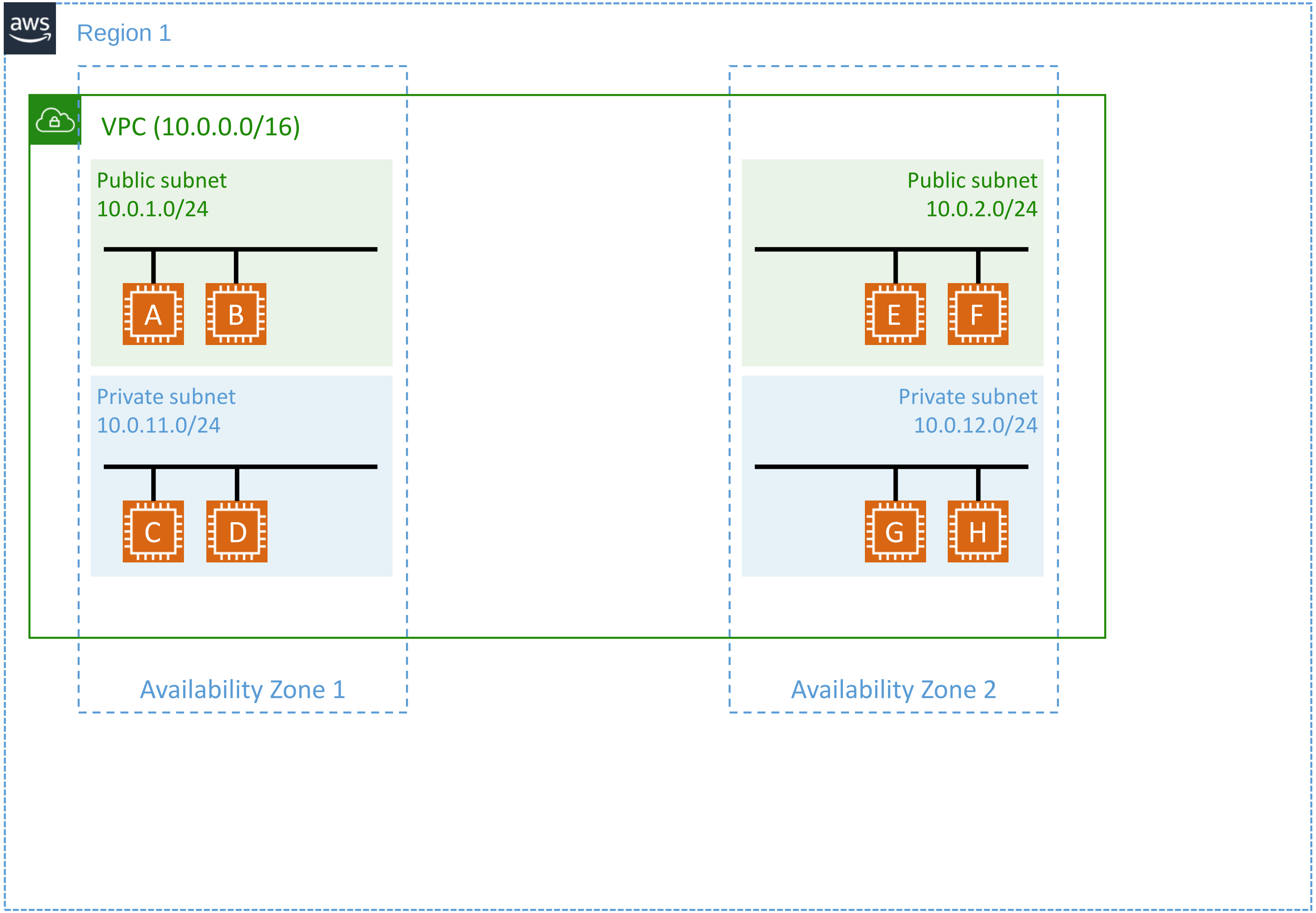
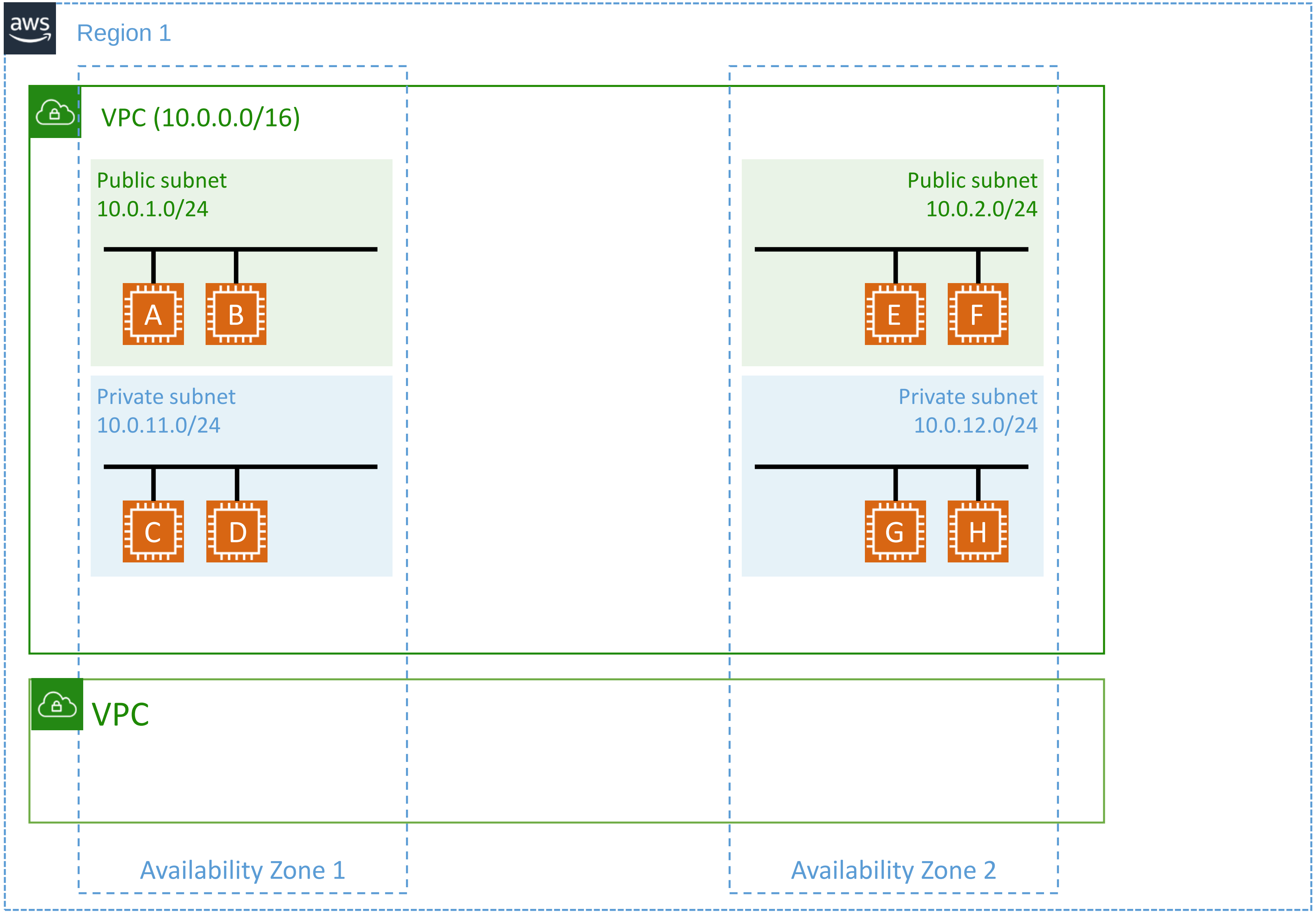
Networking- Compute (EC2) - Storage

BeSA
Become a Solutions Architect

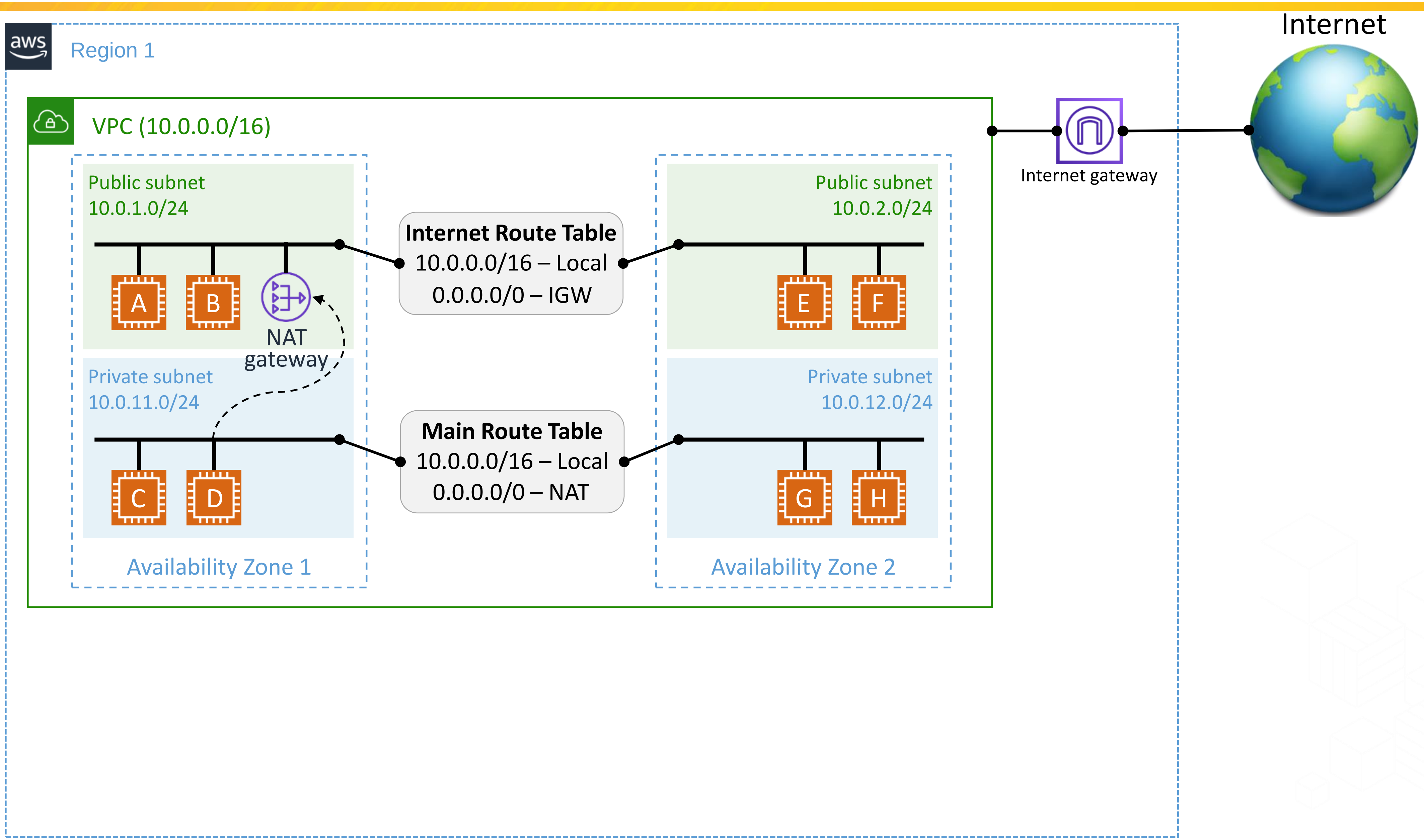
Amazon VPC


aws
Region 1
VPC (10.0.0.0/16)
Public subnet
10.0.1.0/24
A
B
Private subnet
10.0.11.0/24
C
D
Public subnet
10.0.2.0/24
E
F
Private subnet
10.0.12.0/24
G
H
Availability Zone 1
Availability Zone 2

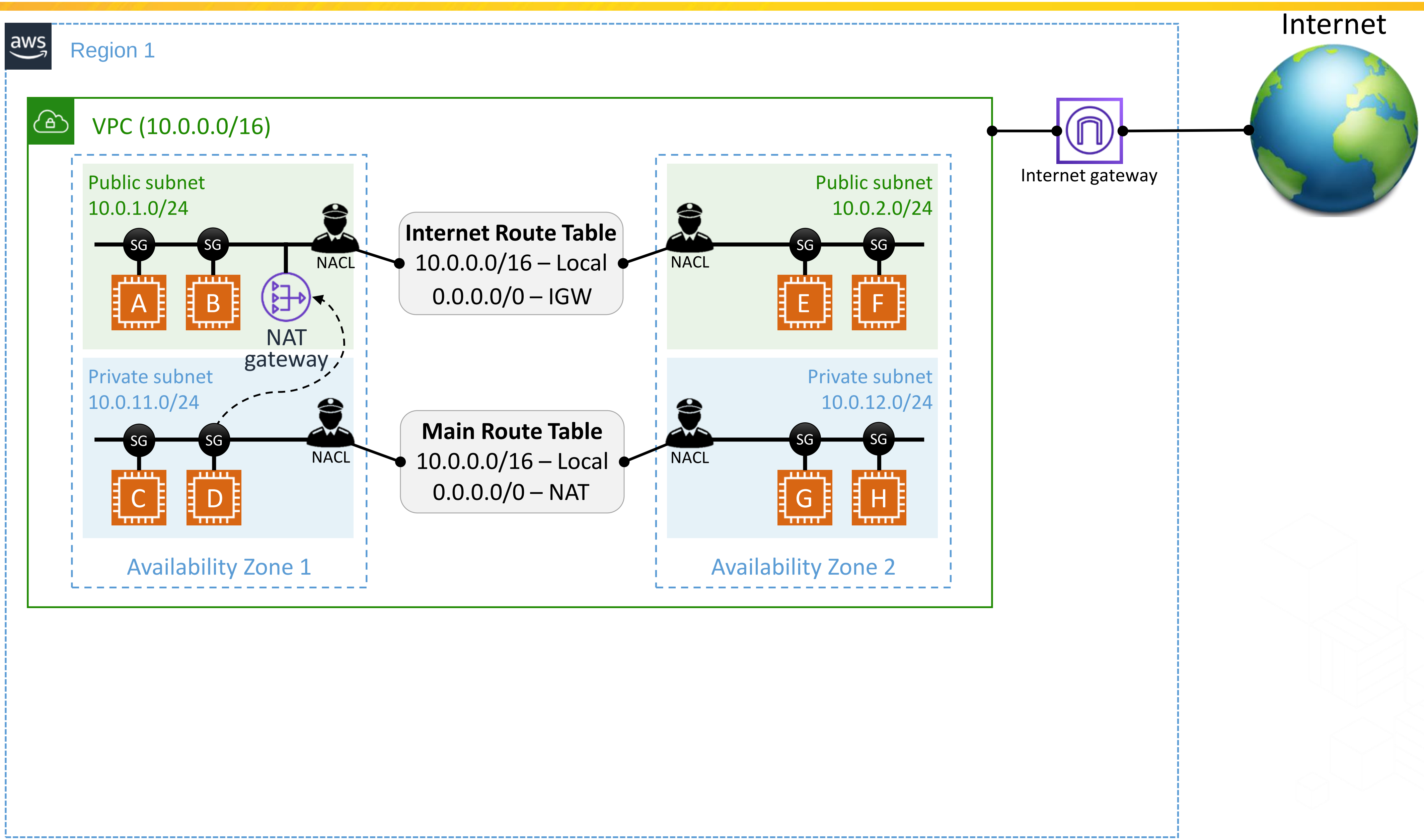
Amazon VPC


Region 1
VPC (10.0.0.0/16)
Public subnet
10.0.1.0/24
A
B
Private subnet
10.0.11.0/24
C
D
Public subnet
10.0.2.0/24
E
F
Private subnet
10.0.12.0/24
G
H
VPC
Availability Zone 1
Availability Zone 2

Amazon VPC


Internet
Region 1
VPC (10.0.0.0/16)
Internet gateway
Public subnet
10.0.1.0/24
A
B
NAT gateway
Private subnet
10.0.11.0/24
C
D
Availability Zone 1
Internet Route Table
10.0.0.0/16 – Local
0.0.0.0/0 – IGW
Main Route Table
10.0.0.0/16 – Local
0.0.0.0/0 – NAT
Public subnet
10.0.2.0/24
E
F
Private subnet
10.0.12.0/24
G
H
Availability Zone 2

Amazon VPC

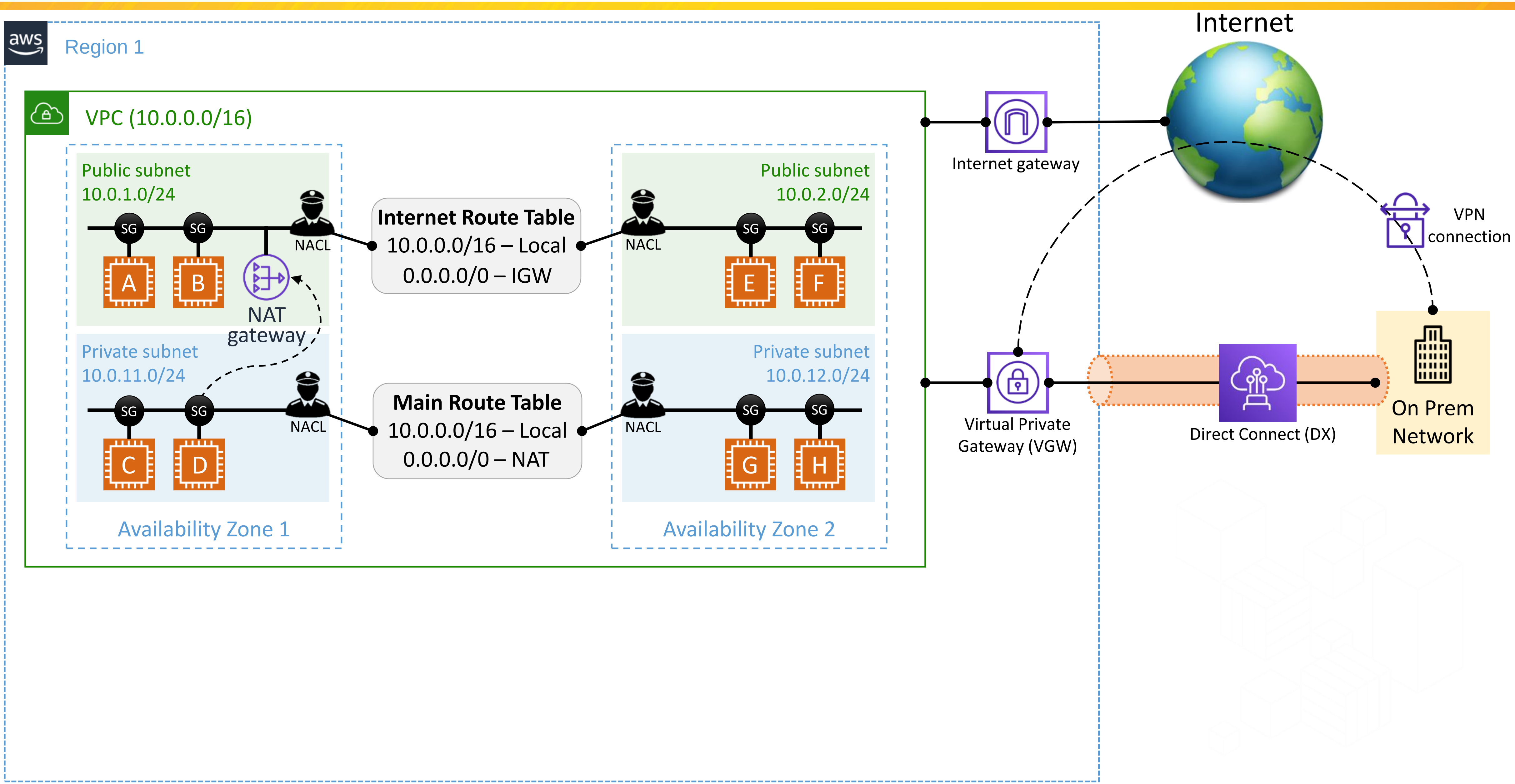


Internet
Region 1
VPC (10.0.0.0/16)
Internet gateway
Public subnet
10.0.1.0/24
SG
SG
A
B
NAT gateway
NACL
Internet Route Table
10.0.0.0/16 – Local
0.0.0.0/0 – IGW
Public subnet
10.0.2.0/24
NACL
SG
SG
E
F
Private subnet
10.0.11.0/24
SG
SG
C
D
NACL
Main Route Table
10.0.0.0/16 – Local
0.0.0.0/0 – NAT
Private subnet
10.0.12.0/24
NACL
SG
SG
G
H
Availability Zone 1
Availability Zone 2

Security Group vs. Network ACL

Security Group	Network ACL
Applied at Instance (ENI) Level	Applied at Subnet Level
Stateful - Response is always allowed	Stateless - Request and Response both have to be allowed
Default Rules (For Default SG) - All inbound is allowed from the same SG - All outbound is Allowed Default Rules (For a new SG) - All Inbound is Deny - All outbound in Allowed	Default Rules (For Default NACL) - All inbound is Allowed - All outbound is Allowed Default Rules (For a new NACL) - All inbound is Deny - All outbound is Deny
1 Instance can have many SG assigned	1 Subnet can have only 1 NACL
Only allow statements	Allow and Deny both statements
Order is not important	Order is important (lower order rule is applied first)
Source - IP / IP Range / Port / SG-<xxxxxxx>	Source - IP / Port / IP Range

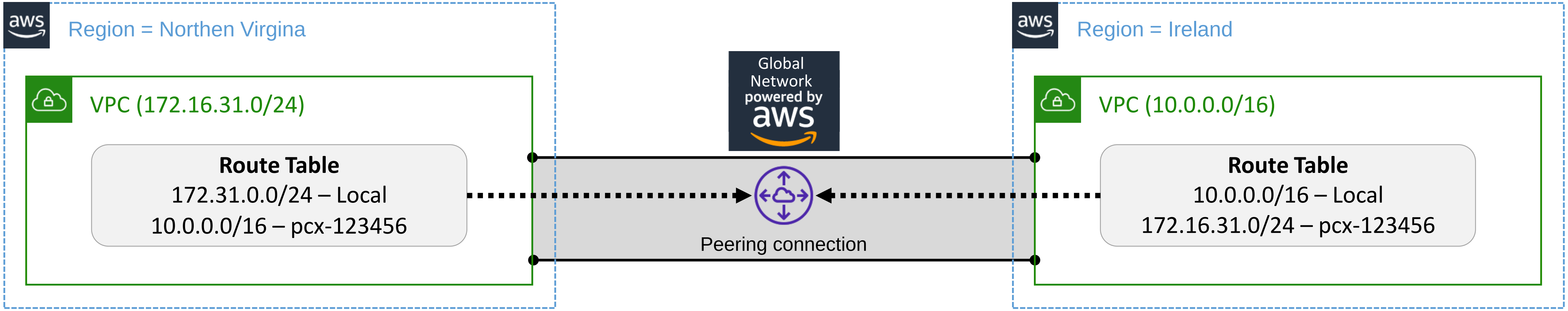
Amazon VPC


Internet
Region 1
VPC (10.0.0.0/16)
Public subnet
10.0.1.0/24
SG
SG
NACL
A
B
NAT
gateway
Private subnet
10.0.11.0/24
SG
SG
NACL
C
D
Availability Zone 1
Internet Route Table
10.0.0.0/16 – Local
0.0.0.0/0 – IGW
Main Route Table
10.0.0.0/16 – Local
0.0.0.0/0 – NAT
Public subnet
10.0.2.0/24
NACL
SG
SG
E
F
Private subnet
10.0.12.0/24
NACL
SG
SG
G
H
Availability Zone 2
Internet gateway
VPN
connection
Virtual Private
Gateway (VGW)
Direct Connect (DX)
On Prem
Network

Site-to-Site VPN vs. Direct Connect

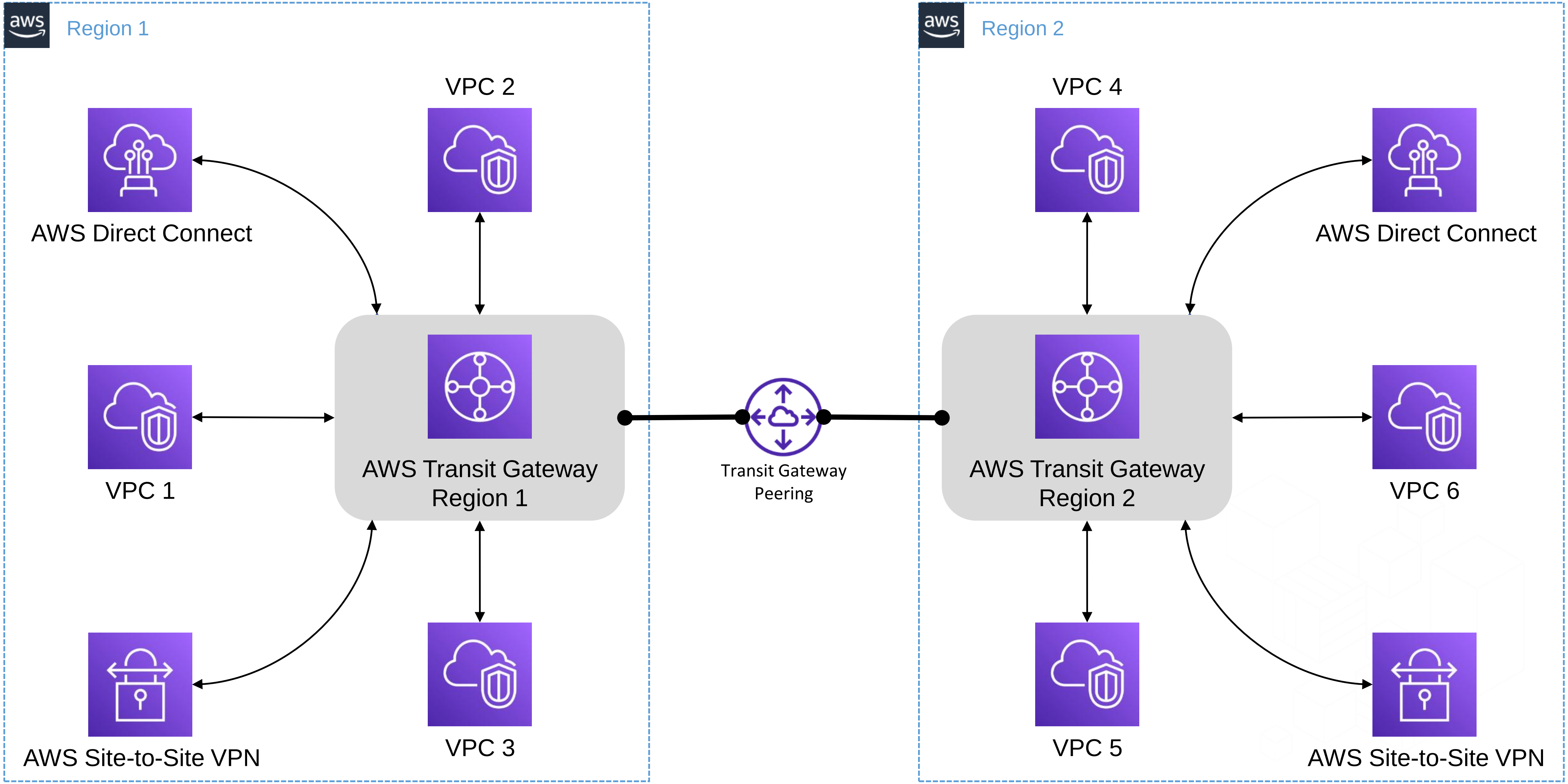
	Site-to-Site VPN	Direct Connect
Use case	Connecting remote networks to AWS VPC which doesn't require heavy data transfer or doesn't require a consistent connection	Connecting remote networks to AWS VPC which require heavy data transfer or require a consistent connection
Choose when…	Cost is important	Predictable performance is important
Supported speed	1.25 Gbps per tunnel	1 / 10 / 100 Gbps (sub 1 Gbps connections may be available from some service providers)
How it works?	Establishes a tunnel over existing internet connection	Establishes a connectivity over a dedicated network. Doesn't use Internet
High Availability	Highly available on AWS side (VGW is deployed across 2 AZs)	Single connection. No high availability by default
Encryption	Uses IPSec	Not encrypted by default
Time to establish	Can be setup in few minutes in a self-service fashion	Requires a Service Provider, may take few hours/days to get established
Cost dimension	Per connection hour and data transfer out	Variable port fees and data transfer out

VPC Peering



- VPC Peering is established in a 1:1 fashion and is not transitive.
- You can setup cross-region, cross-account peering.
- Two step process – Request Peering and Accept Peering.
- Route Table entries direct the traffic.
- Peering connection uses private IP Address, traffic always stays on the global AWS backbone.
- There is no charge to create a VPC peering connection but there is a charge for data transfer across peering connections.

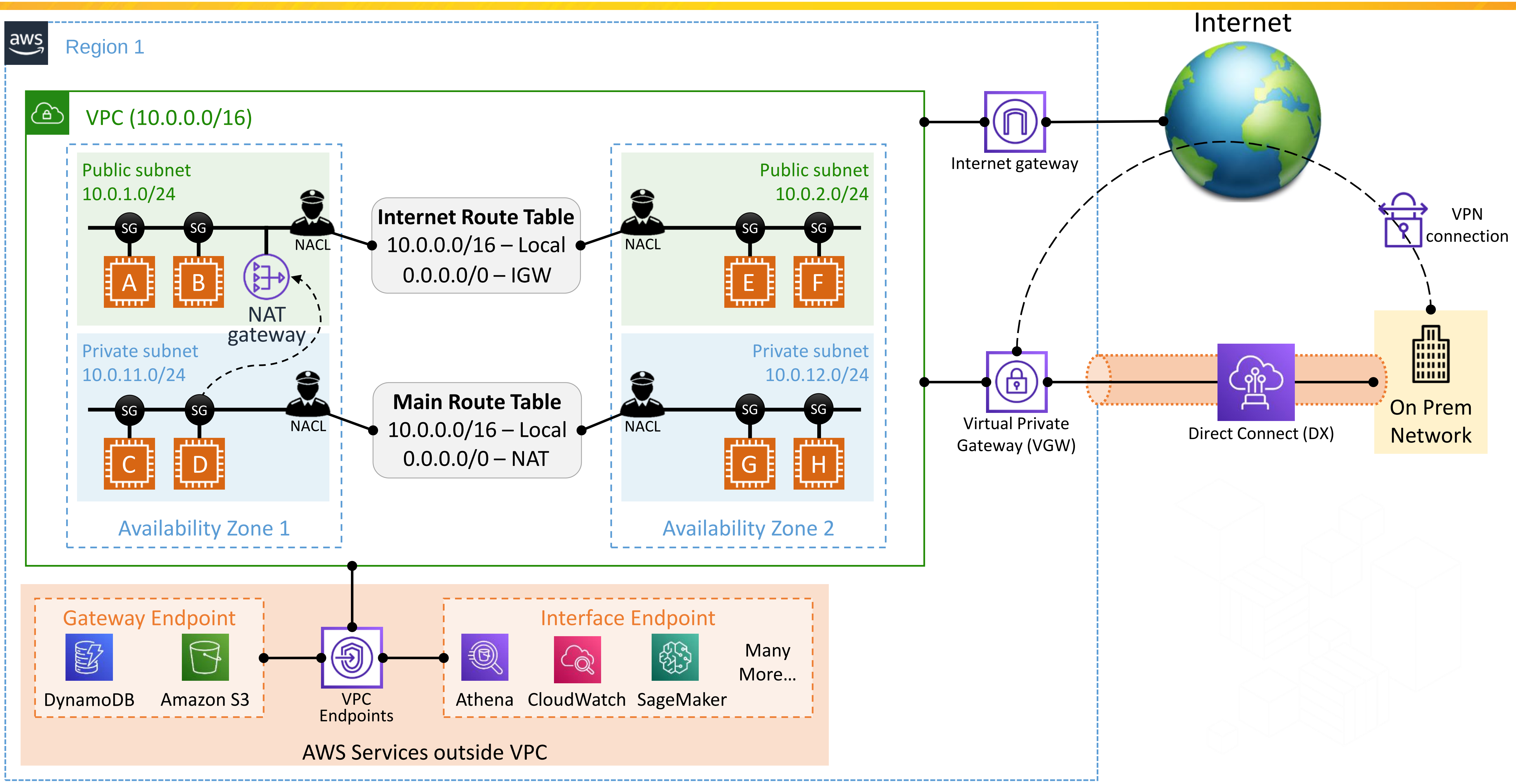
Transit Gateway



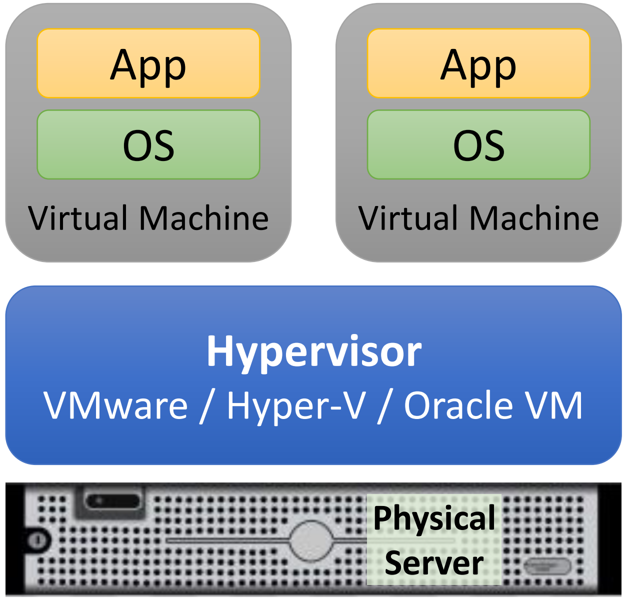
VPC Endpoints

	Gateway Endpoint	Interface Endpoint
Used for	Private connectivity to Amazon S3 and Amazon DynamoDB	Private connectivity to 100+ AWS Services (including Amazon S3)
How it works?	An entry for prefix list (IP addresses) for supported services is added in to the routing table	An ENI(s) is provisioned in the selected subnet(s) which serves as an entry point for traffic destined to a supported service. (powered by AWS PrivateLink)
Provisioned at	VPC Level then entry added to Route Table	Subnet Level (no entry required in Route Table)
Security	Through VPC Endpoint Policy	Through Security Group

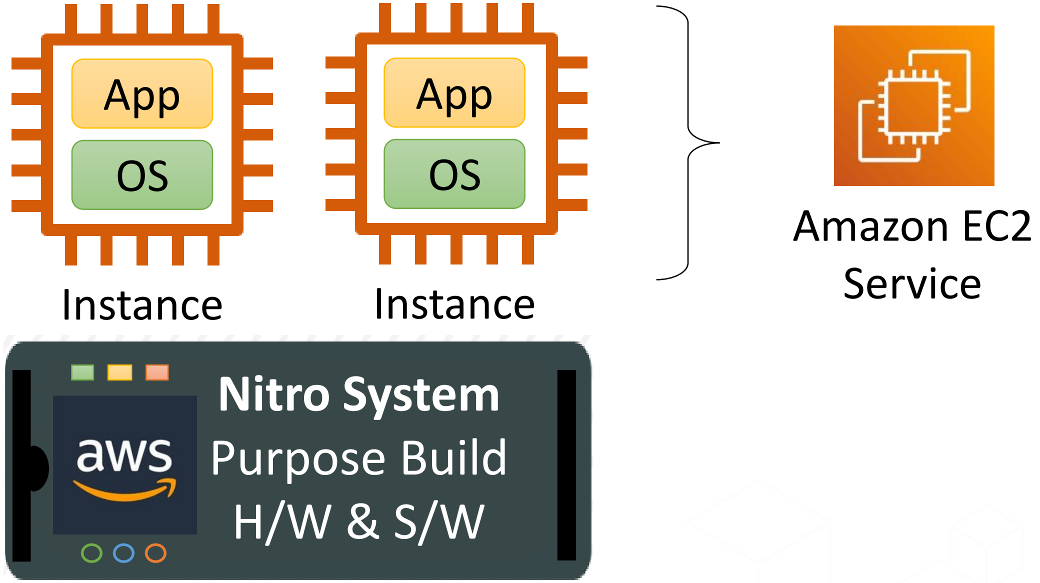
Amazon VPC


Internet
aws
Region 1
VPC (10.0.0.0/16)
Public subnet
10.0.1.0/24
SG
SG
NACL
A
B
NAT gateway
Private subnet
10.0.11.0/24
SG
SG
NACL
C
D
Availability Zone 1
Internet Route Table
10.0.0.0/16 – Local
0.0.0.0/0 – IGW
Main Route Table
10.0.0.0/16 – Local
0.0.0.0/0 – NAT
Public subnet
10.0.2.0/24
NACL
SG
SG
E
F
Private subnet
10.0.12.0/24
NACL
SG
SG
G
H
Availability Zone 2
Internet gateway
VPN connection
Virtual Private Gateway (VGW)
Direct Connect (DX)
On Prem Network
Gateway Endpoint
DynamoDB
Amazon S3
VPC Endpoints
Interface Endpoint
Athena
CloudWatch
SageMaker
Many More...
AWS Services outside VPC

Amazon Elastic Compute Cloud (EC2) – Virtual Server in Cloud



Virtualization Approach

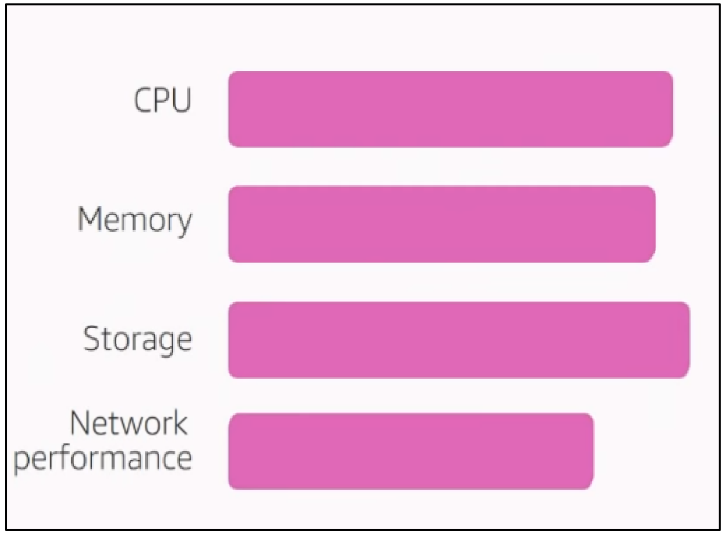


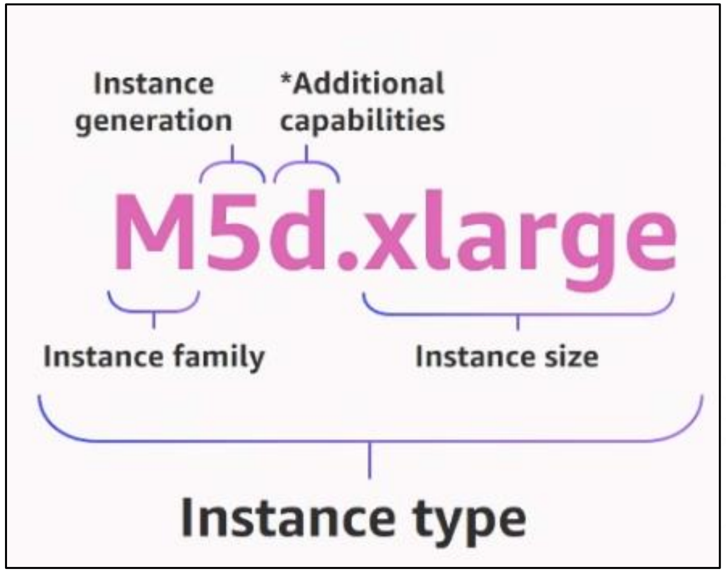
AWS Approach

Provisioning an EC2 Instance

- AMI – Amazon Machine Image
  - Template of common OS images
    - Quick Start
    - My AMI
    - Marketplace
    - Community

- Instance Type
  - Performance Characteristics
  - Optimized for different workloads
  - Elastic – Can be changed later
  - Region Specific





User Data

- Customize your instance at launch

```
#!/bin/bash
yum update -y
yum install httpd -y
echo "<html><body><center><h1>Welcome to AWS. Here is my web  page!
</h1></center></body></html>" > /var/www/html/index.html
systemctl start httpd
systemctl enable httpd
```

EC2 Purchase Options

Highest Price | Lower Price | Lowest Price

On Demand

- Pay by second – Amazon Linux Ubuntu and Windows
- Pay by hours – Other Operating Systems
- Unpredictable workload

Reserved Instance

- 1 Year or 3 Year commitment
- Long running and predictable workload
- Reserve Region / AZ / Instance Type / OS

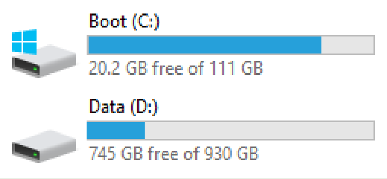
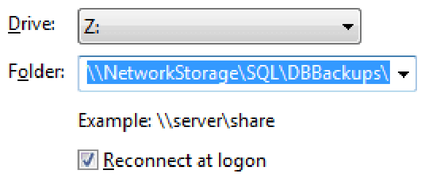
Savings Plan (RI 2.0)

- Commitment of a fixed payment
- Flexible in terms of reservation

Spot Instance

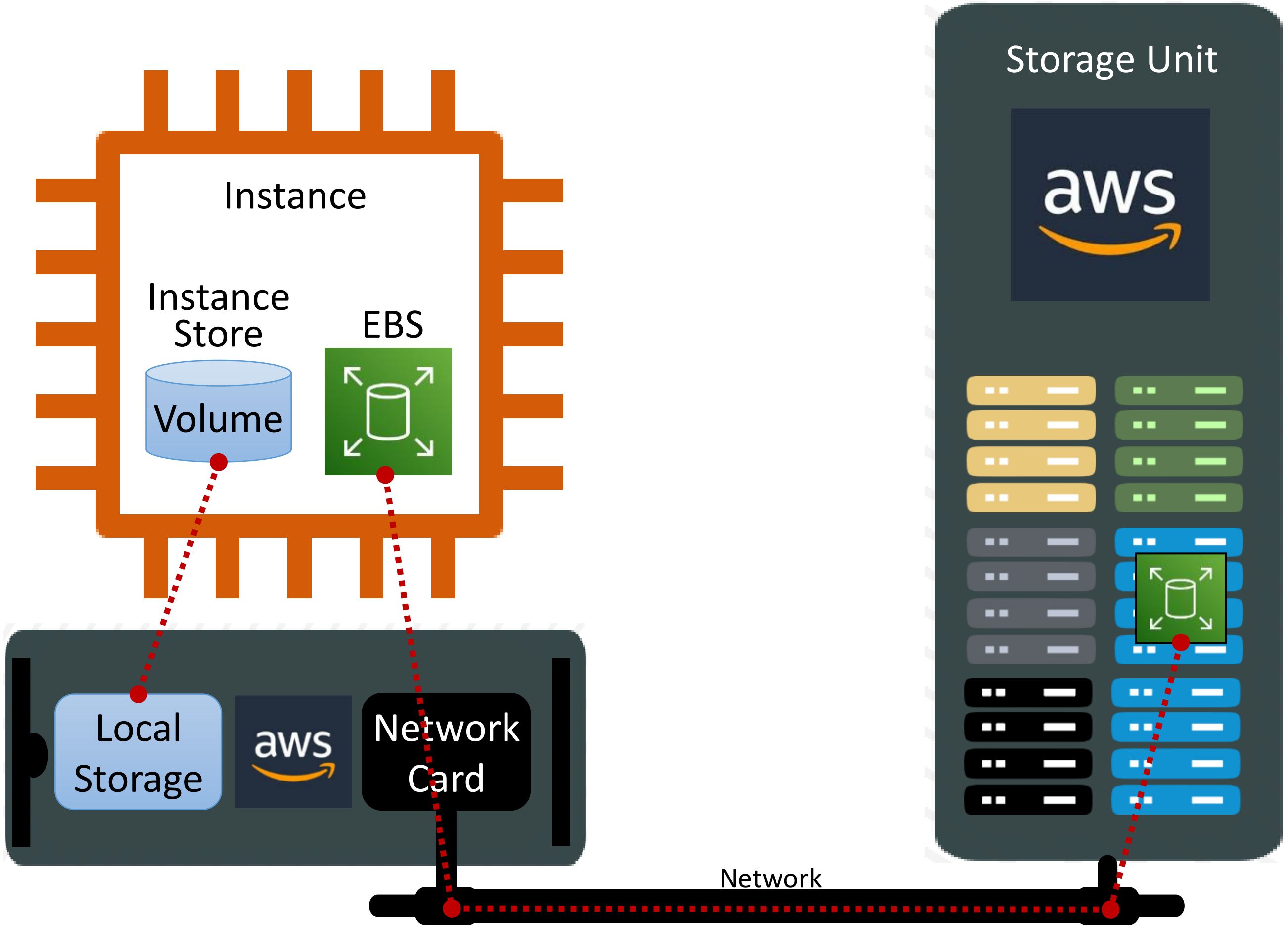
- Unused capacity of AWS infrastructure
- Supply and demand based pricing

Storage Types

	Block Storage	File Storage	Object Storage
Unit of Transaction	Blocks	Files	Objects (files with metadata)
Example	Laptop Disk  Boot (C:) 20.2 GB free of 111 GB Data (D:) 745 GB free of 930 GB	Windows Share  Drive: Z: Folder: \\NetworkStorage\SQL\DBBackups\ Example: \\server\share ☑ Reconnect at logon	OneDrive / Google Drive / Dropbox OneDrive Google Drive Dropbox
How can you update?	You can directly update the file	You can directly update the file	You cannot update the object directly. You create a new version of the object and replace the existing one or keep multiple versions of the same object.
Protocols	SCSI, Fiber Channel, SATA	SMB, CIFS, NFS	REST/SOAP over HTTP/HTTPs
Support for metadata	No metadata support it stores only file system attributes	No metadata support it stores only file system attributes	Supports custom metadata
AWS Services	Amazon EBS Amazon Instance Store	Amazon EFS Amazon FSX	Amazon S3 Amazon Glacier

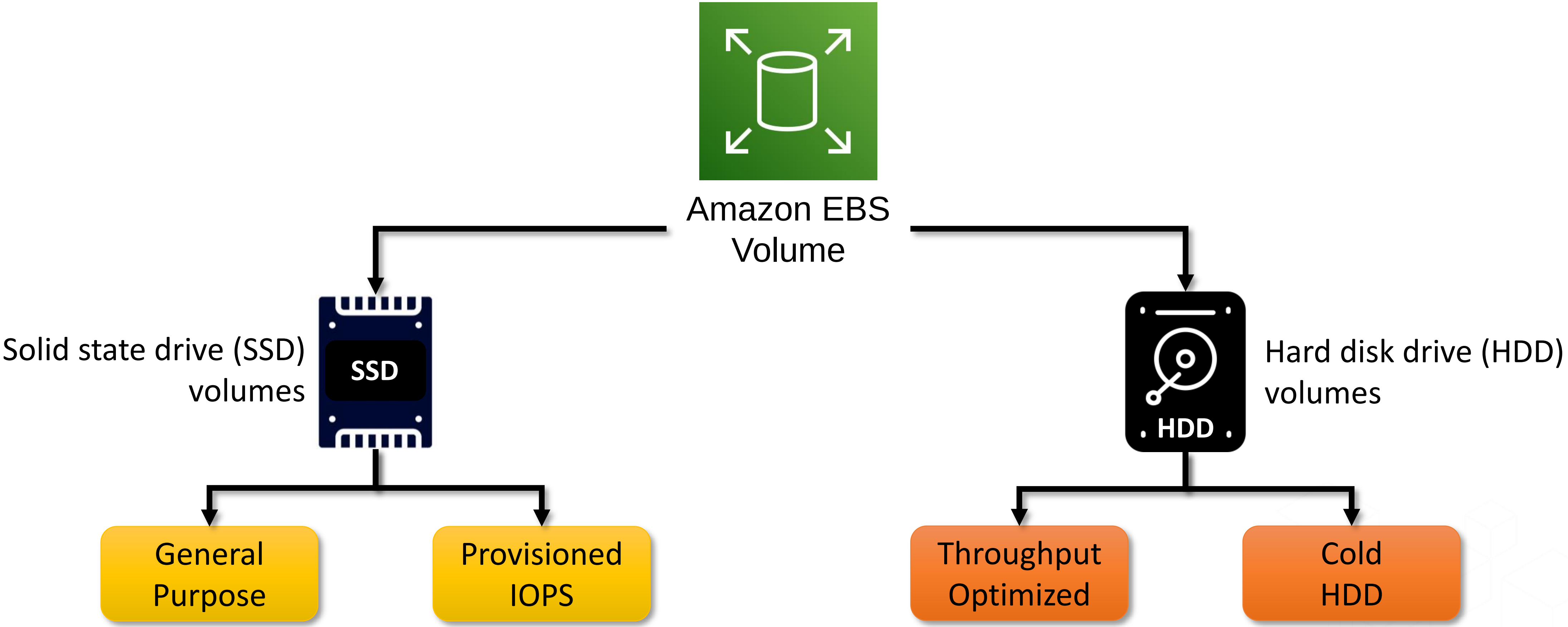
EC2 Storage Options – Block Storage

- Instance Store
  - Local Storage
  - Ephemeral
  - Limited Size
  - Not available on all instance types
  - Use case
    - Swap space
    - Temp. Storage



- Elastic Block Store (EBS)
  - Over Network
  - Persistent Storage
  - 64 TB Max Size
  - Choice of volume types
  - EBS Optimized instance type*
  - Use case
    - Any block storage need (OS/DB/Log)

Amazon EBS Volume Types



Amazon EBS Volume

Solid state drive (SSD) volumes

SSD

Hard disk drive (HDD) volumes

HDD

General Purpose

- GP2
- GP3

Provisioned IOPS

- io1
- io2
- io2 Block Express

Throughput Optimized

Cold HDD

Amazon EBS Volume

- Specific to a AZ
- Choose based on performance/price
- Can be expanded (can’t shrink)
- Can be changed while instance is running

- Supports Snapshot
- Snapshots can be copied to another Region or account
- Supports encryption
- Attached to Single EC2 Instance*

Amazon EBS Snapshot

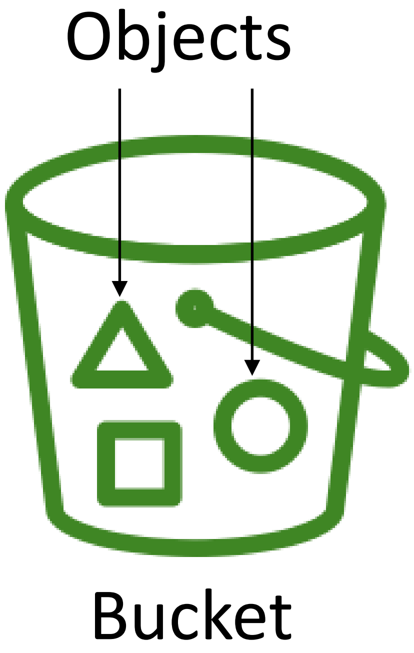
- EBS Snapshots are a point-in-time copy of your data.

- It can be used to enable disaster recovery, migrate data across regions and accounts, and improve backup compliance.

- The snapshots are automatically saved to Amazon S3 for long-term retention.

- Amazon EBS Snapshots are incremental, storing only the changes since the last snapshot.

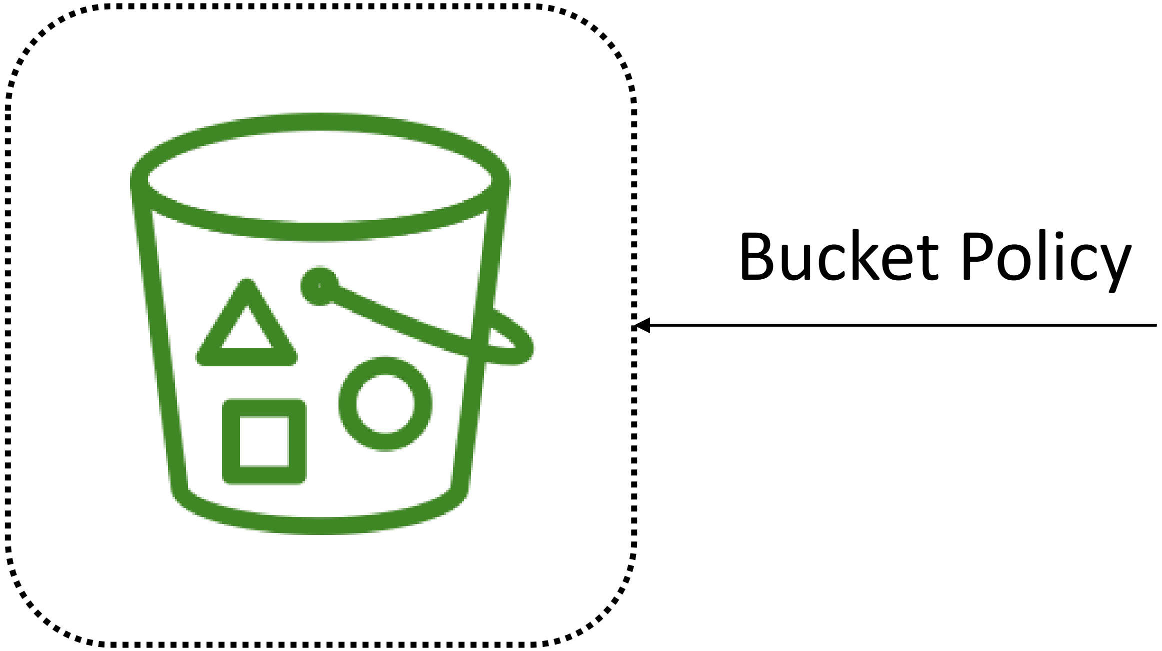
Buckets and Objects



- Buckets are containers for data stored in S3.
- Objects are the fundamental entities stored in Amazon S3.
- S3 Name Space is global, Buckets are regional.
- Durability – 99.999999999 %
- Availability – 99.9x %
- Max object size 5 terabytes

Bucket Policy

- The bucket policy, written in JSON, provides access to the objects stored in the bucket.



Bucket Policy

```
{
  "Id": "Policy1675378587932",
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Stmt1675378555005",
      "Action": [
        “<API Request>"
      ],
      "Effect": “<Allow/Deny>",
      "Resource": “<Bucket-ARN>",
      "Principal": “<Identity>"
    }
  ]
}
```

Amazon S3 Storage Pricing

Compute
No. of requests

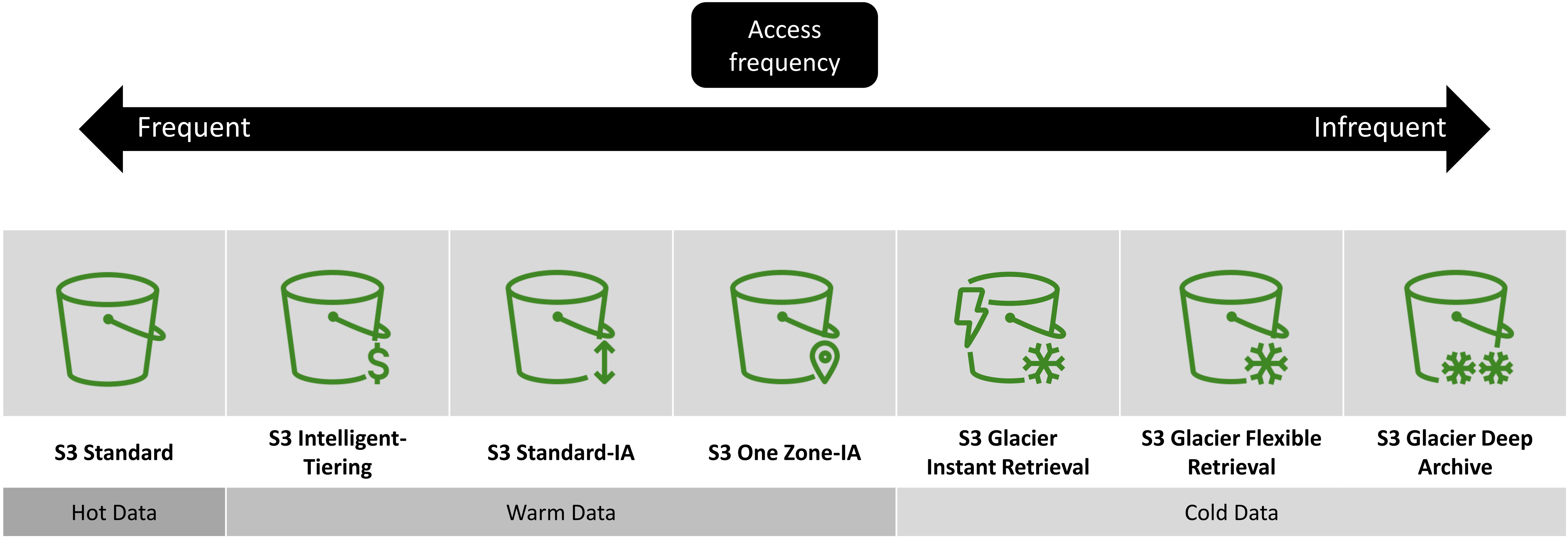
Storage
Space Used

Data Transfer
Download

Other
Management

AWS Pricing Calculator

Amazon S3 Storage Class



Access frequency

Frequent

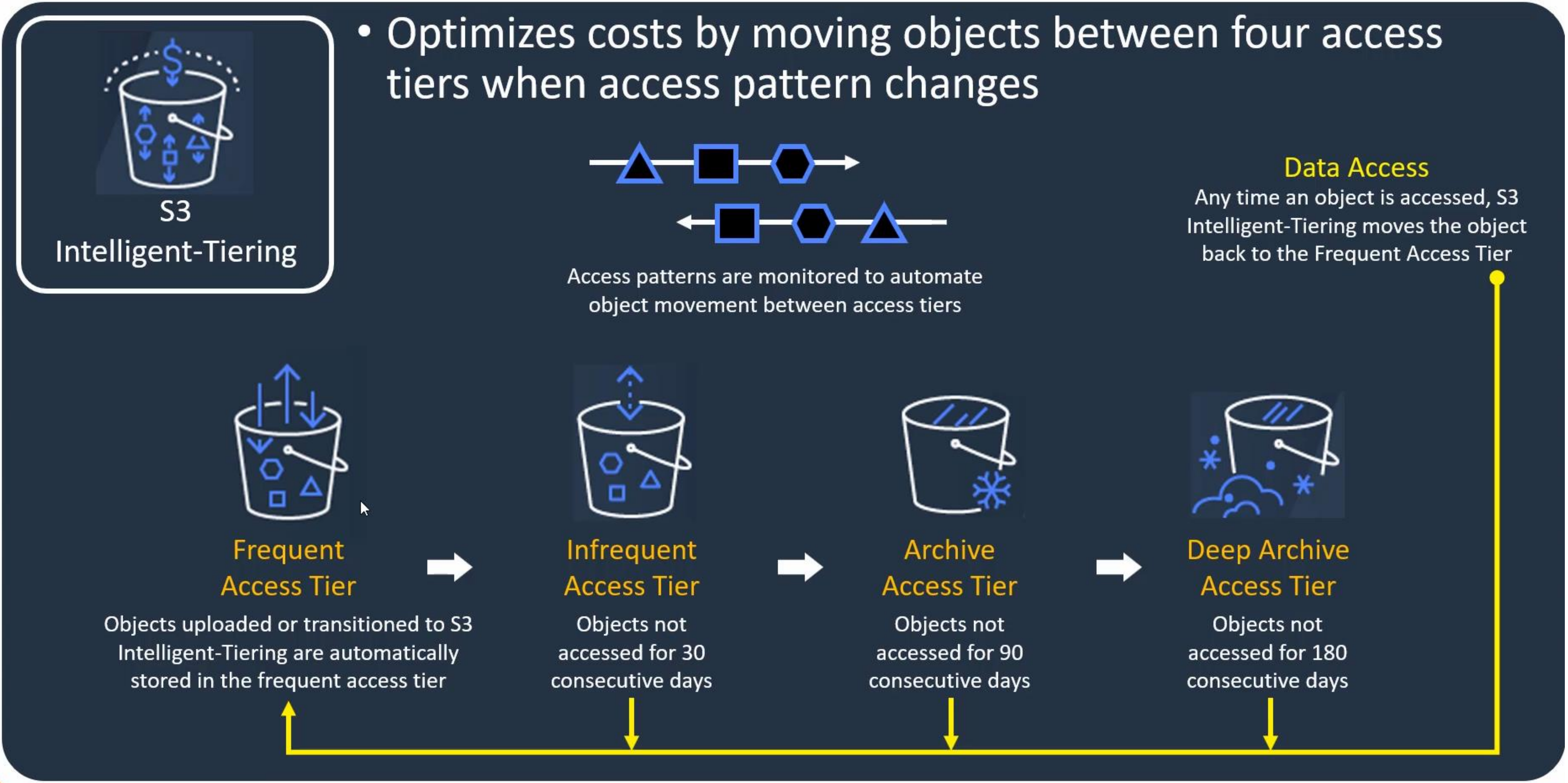
Infrequent

S3 Standard	S3 Intelligent-Tiering	S3 Standard-IA	S3 One Zone-IA	S3 Glacier Instant Retrieval	S3 Glacier Flexible Retrieval	S3 Glacier Deep Archive
Hot Data	Warm Data			Cold Data		

Amazon S3 Storage Class

Storage class	Designed for
Standard	Frequently accessed data (more than once a month) with milliseconds access
Intelligent-Tiering	Data with changing or unknown access patterns
Standard-IA	Infrequently accessed data (once a month) with milliseconds access
One Zone-IA	Re-creatable, infrequently accessed data (once a month) stored in a single Availability Zone with milliseconds access
Glacier Instant Retrieval	Long-lived archive data accessed once a quarter with instant retrieval in milliseconds
Glacier Flexible Retrieval (formerly Glacier)	Long-lived archive data accessed once a year with retrieval of minutes to hours
Glacier Deep Archive	Long-lived archive data accessed less than once a year with retrieval of hours
Reduced redundancy	Noncritical, frequently accessed data with milliseconds access (not recommended as S3 Standard is more cost effective)

Amazon S3 Intelligent-Tiering



S3 Intelligent-Tiering

- Optimizes costs by moving objects between four access tiers when access pattern changes

Access patterns are monitored to automate object movement between access tiers

Data Access

Any time an object is accessed, S3 Intelligent-Tiering moves the object back to the Frequent Access Tier

Frequent Access Tier

Objects uploaded or transitioned to S3 Intelligent-Tiering are automatically stored in the frequent access tier

Infrequent Access Tier

Objects not accessed for 30 consecutive days

Archive Access Tier

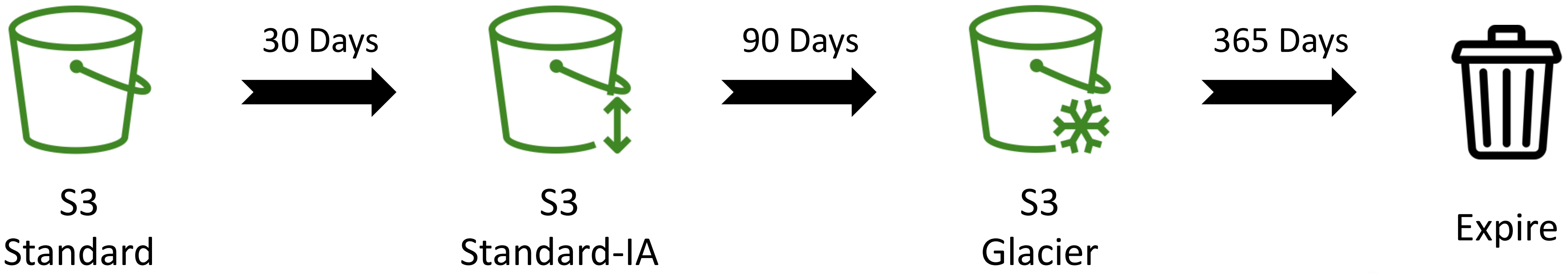
Objects not accessed for 90 consecutive days

Deep Archive Access Tier

Objects not accessed for 180 consecutive days

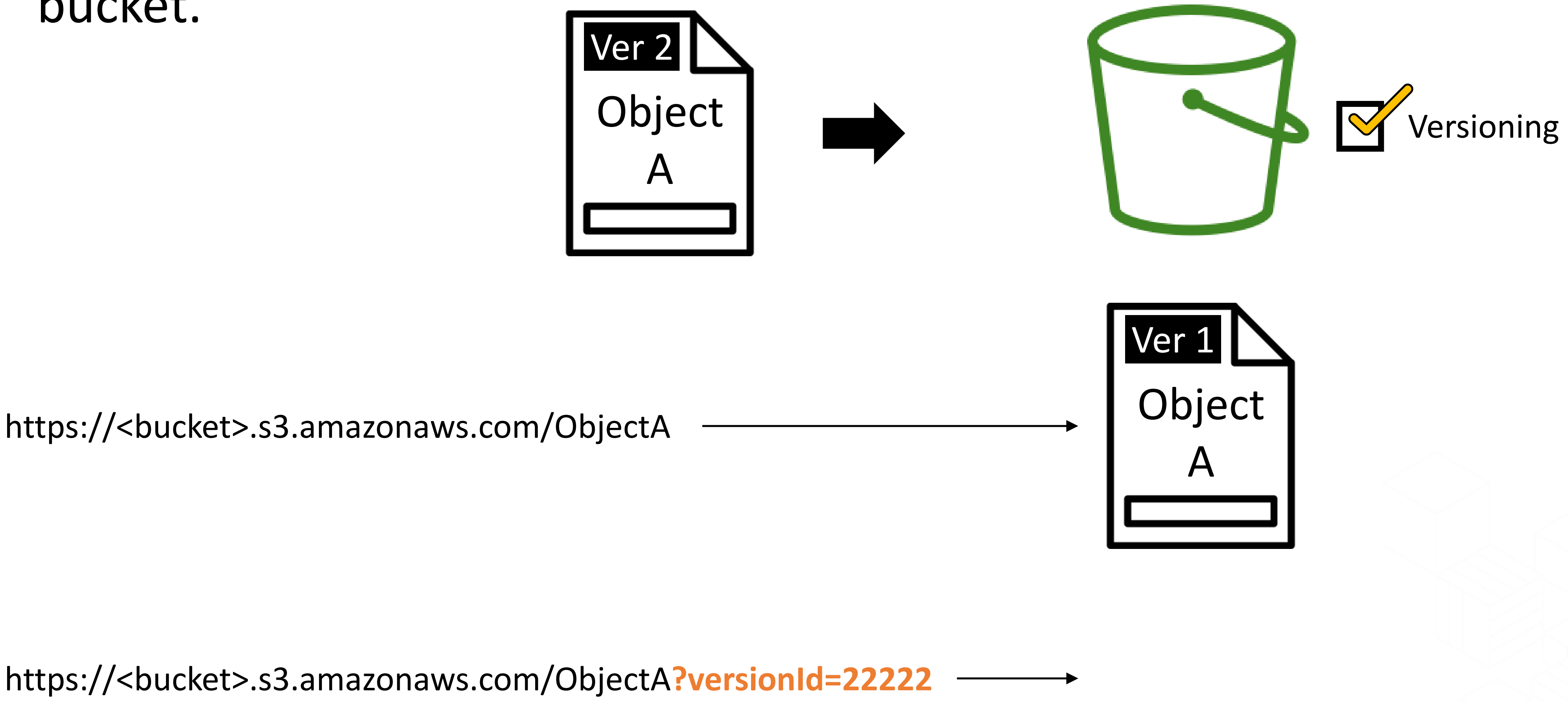
Amazon S3 Life Cycle Policies

- Storing objects cost effectively throughout their lifecycle



Bucket Versioning

- Versioning is a means of keeping multiple variants of an object in the same bucket.



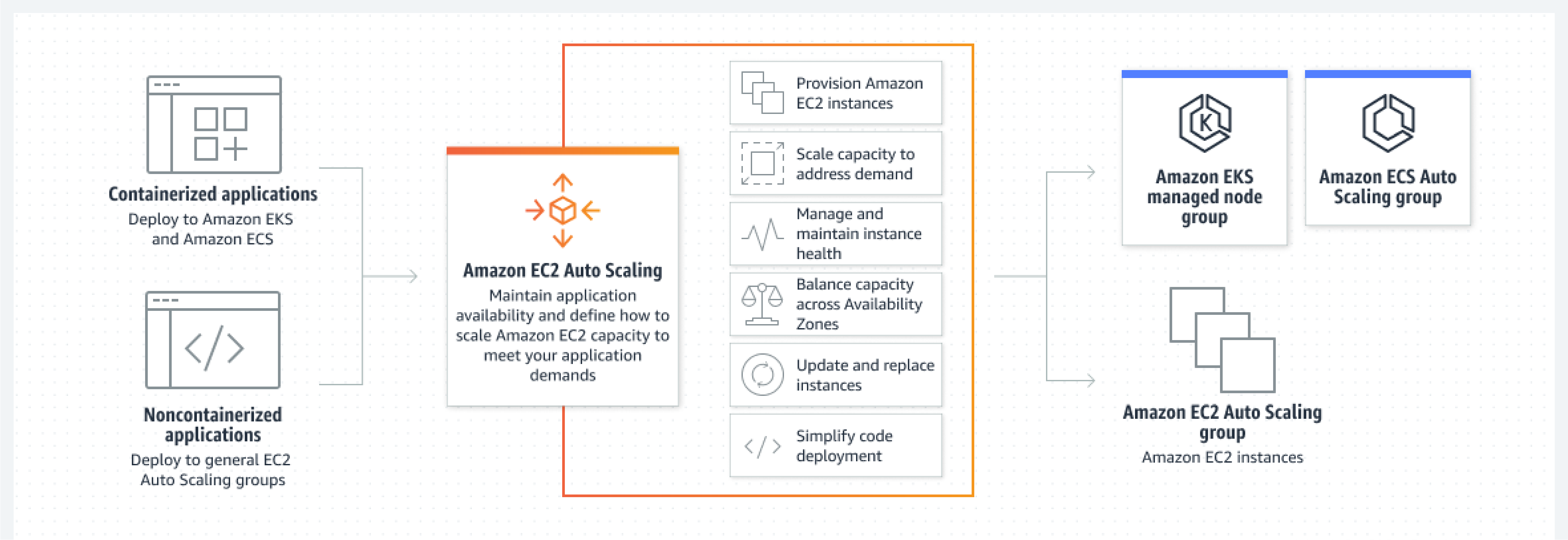
https://<bucket>.s3.amazonaws.com/ObjectA

https://<bucket>.s3.amazonaws.com/ObjectA?versionId=22222

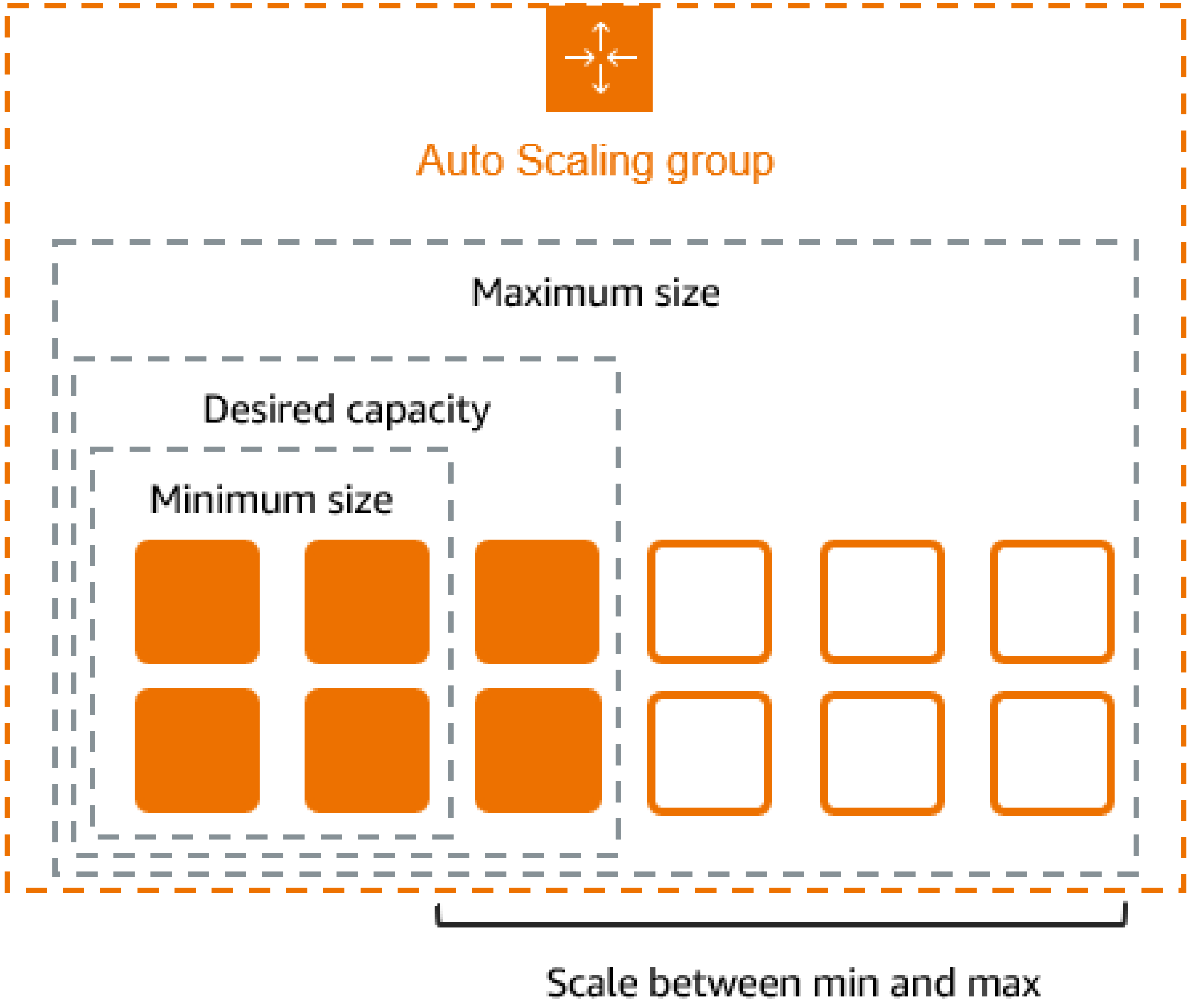
Replication Rules

- Replication enables automatic, asynchronous copying of objects across Amazon S3 buckets.
- It requires versioning to be enabled.
- You can replicate data from one source bucket to multiple destination buckets.
- Supports cross-account and cross-region replication.

EC2 Auto Scaling feature


Containerized applications
Deploy to Amazon EKS and Amazon ECS
Noncontainerized applications
Deploy to general EC2 Auto Scaling groups
Amazon EC2 Auto Scaling
Maintain application availability and define how to scale Amazon EC2 capacity to meet your application demands
Provision Amazon EC2 instances
Scale capacity to address demand
Manage and maintain instance health
Balance capacity across Availability Zones
Update and replace instances
Simplify code deployment
Amazon EKS managed node group
Amazon ECS Auto Scaling group
Amazon EC2 Auto Scaling group
Amazon EC2 instances

EC2 Auto Scaling Group



Question 1

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

Which actions should be taken to allow the EC2 instances to download the needed patches? (Select TWO.)

A) Configure a NAT gateway in a public subnet.

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

C) Assign Elastic IP addresses to the EC2 instances.

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

E) Configure a NAT instance in a private subnet

Question 1

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

Which actions should be taken to allow the EC2 instances to download the needed patches? (Select TWO.)

A) Configure a NAT gateway in a public subnet.

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

C) Assign Elastic IP addresses to the EC2 instances.

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

E) Configure a NAT instance in a private subnet

Question 2

A solutions architect wants to design a solution to save costs for Amazon EC2 instances that do not need to run during a 2-week company shutdown. The applications running on the EC2 instances store data in instance memory that must be present when the instances resume operation.

Which approach should the solutions architect recommend to shut down and resume the EC2 instances?

A) Modify the application to store the data on instance store volumes. Reattach the volumes while restarting them.

B) Snapshot the EC2 instances before stopping them. Restore the snapshot after restarting the instances.

C) Run the applications on EC2 instances enabled for hibernation. Hibernate the instances before the 2-week company shutdown.

D) Note the Availability Zone for each EC2 instance before stopping it. Restart the instances in the same Availability Zones after the 2-week company shutdown.

Question 2

A solutions architect wants to design a solution to save costs for Amazon EC2 instances that do not need to run during a 2-week company shutdown. The applications running on the EC2 instances store data in instance memory that must be present when the instances resume operation.

Which approach should the solutions architect recommend to shut down and resume the EC2 instances?

A) Modify the application to store the data on instance store volumes. Reattach the volumes while restarting them.

B) Snapshot the EC2 instances before stopping them. Restore the snapshot after restarting the instances.

C) Run the applications on EC2 instances enabled for hibernation. Hibernate the instances before the 2-week company shutdown.

D) Note the Availability Zone for each EC2 instance before stopping it. Restart the instances in the same Availability Zones after the 2-week company shutdown.

Question 3

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

Which approach will meet these requirements?

A) Deploy an Application Load Balancer configured with a listener for the private IP address and register the primary EC2 instance with the load balancer. Upon failure, de-register the instance and register the standby EC2 instance.

B) Configure a custom DHCP option set. Configure DHCP to assign the same private IP address to the standby EC2 instance when the primary EC2 instance fails.

C) Attach a secondary elastic network interface to the EC2 instance configured with the private IP address. Move the network interface to the standby EC2 instance if the primary EC2 instance becomes unreachable.

D) Associate an Elastic IP address with the network interface of the primary EC2 instance. Disassociate the Elastic IP from the primary instance upon failure and associate it with a standby EC2 instance.

Question 3

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

Which approach will meet these requirements?

A) Deploy an Application Load Balancer configured with a listener for the private IP address and register the primary EC2 instance with the load balancer. Upon failure, de-register the instance and register the standby EC2 instance.

B) Configure a custom DHCP option set. Configure DHCP to assign the same private IP address to the standby EC2 instance when the primary EC2 instance fails.

C) Attach a secondary elastic network interface to the EC2 instance configured with the private IP address. Move the network interface to the standby EC2 instance if the primary EC2 instance becomes unreachable.

D) Associate an Elastic IP address with the network interface of the primary EC2 instance. Disassociate the Elastic IP from the primary instance upon failure and associate it with a standby EC2 instance.