O'REILLY®

# AWS Core Architecture Concepts

# Amazon Web Services: Architect Associate Certification

# Core Architecture Concepts

# What we will cover:

- Fundamentals of AWS: architecture, terminology and concepts
- Virtual Private Cloud (VPC): networking services
- Amazon Elastic Compute Cloud (EC2): instance deployment and configuration
- Storage solutions: Elastic Block Storage (EBS) and snapshot management
- Simple Storage Service (S3): Object storage
  - S3 Glacier: Archive storage

# Steps for AWS Certification Success

- Think like a Cloud Architect
- Architects “build”  (i.e. design)  “construct
- Architects propose solutions based on existing building blocks
- The Associated Architect is based on common sense
- Every question is a “situation”
  - Current or Proposed
- The correct answer is the best answer based on the suggested answers to multiple choice questions

# Documentation

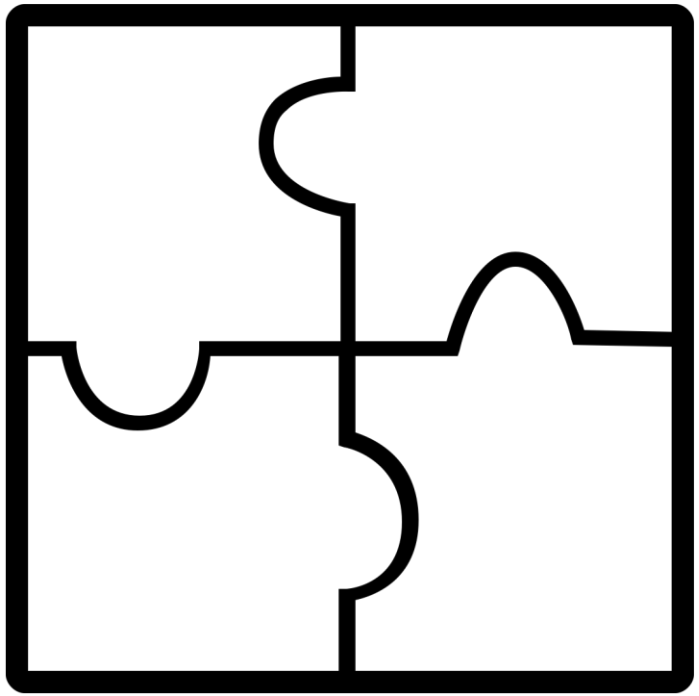
- AWS Certified Solutions Architect – Sample Questions
- AWS Certified Solutions Architect - Associate
- AWS Certified Solutions Architect – Study Guide

# Solutions Architect Documentation

- AWS Quick Starts
  - https://aws.amazon.com/quickstart/
- Self Paced Labs
  - https://aws.amazon.com/training/self-paced-labs/?nc2=h_l2_tr
- AWS Documentation
  - https://aws.amazon.com/documentation/
- AWS Discussion Forums
  - https://forums.aws.amazon.com/index.jspa?nc2=h_l2_su

# Certification Study Guide Example:

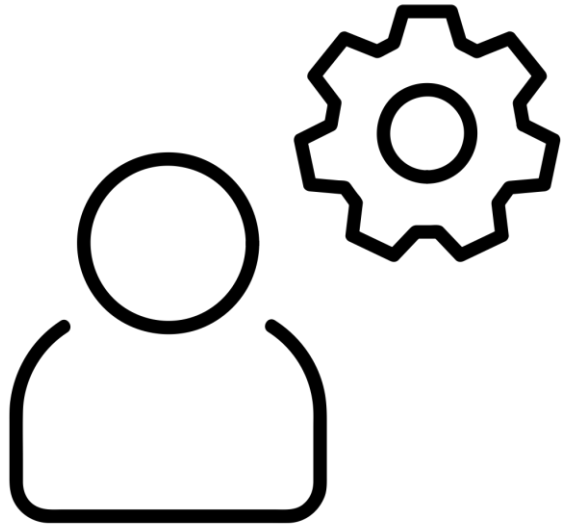
- AWS Component
  - Regions and Zones
- Read FYI, documentation, or watch video
  - https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Concepts.RegionsAndAvailabilityZones.html
  - https://docs.aws.amazon.com/acm/latest/userguide/acm-regions.html
- Setup component
  - Open EC2, CloudFront, and S3 to see the change from Region to Global location (Edge)



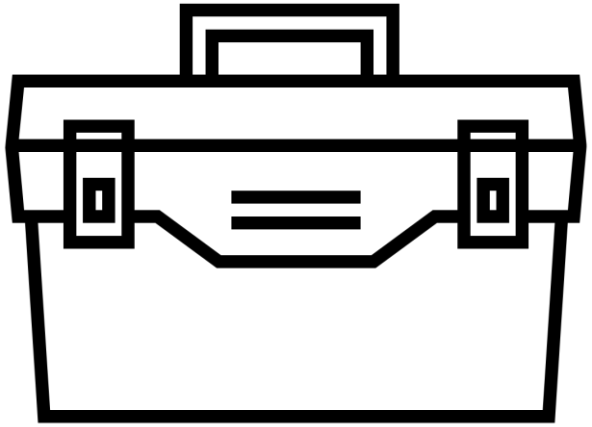
# Core Architecture Concepts

# AWS Cloud Services

- AWS Administration – Management portal
- Compute Services – Elastic compute cloud
- Networking Services – Virtual private cloud
- Auto Scaling – Scale EC2 capacity automatically
- Elastic Load Balancing – Distribute application traffic across EC2 instances or containers
- Elastic Block Storage – Virtual hard drives
- S3 – Durable, scalable object storage
- S3 Glacier – Long-term data archiving



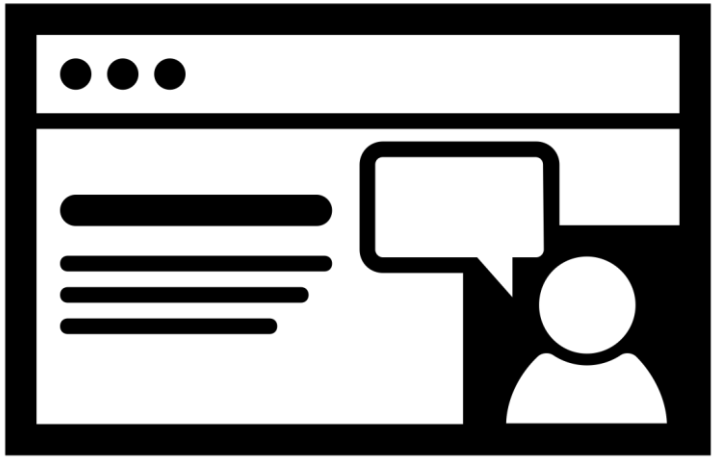
# Cloud Services



- Managed services: AWS does most of the setup
- Less managed services: You can do whatever you want
  - You to do most of the setup, management, and monitoring (VPC, EC2)
- The reality – there are no completely unmanaged services at AWS

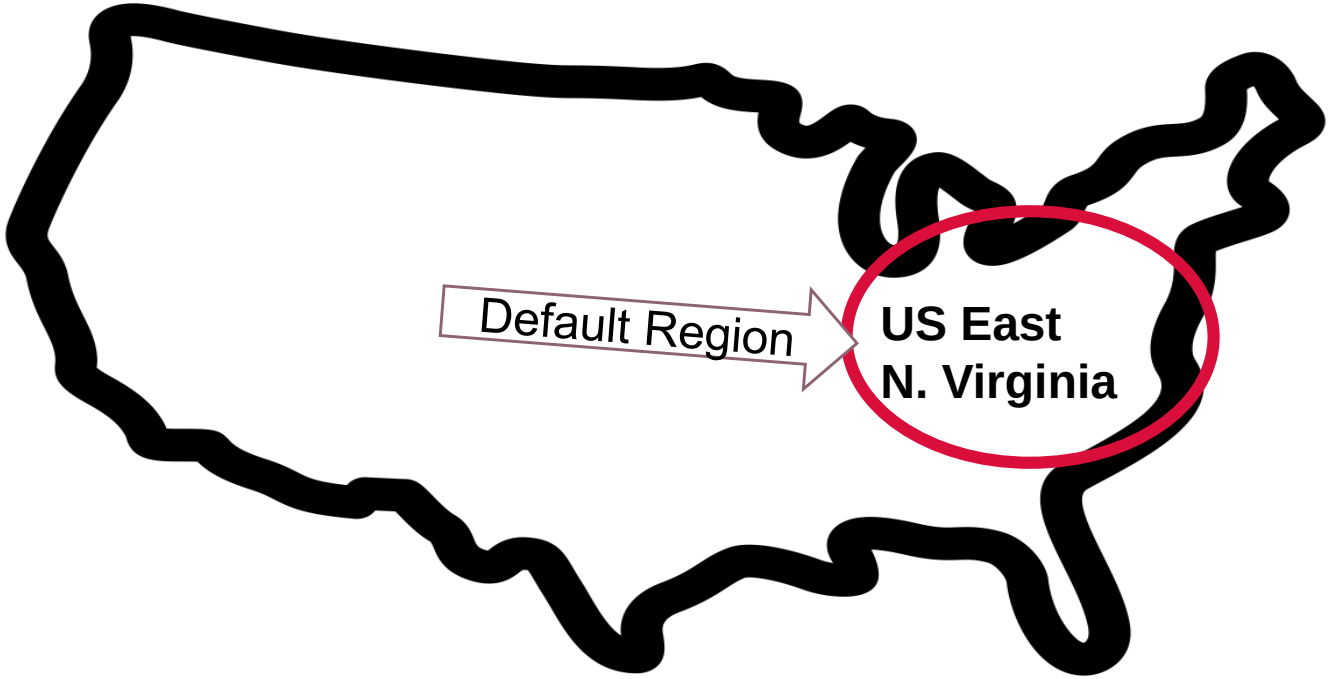
# What’s the least managed AWS Service?

# Please open your Case Study……

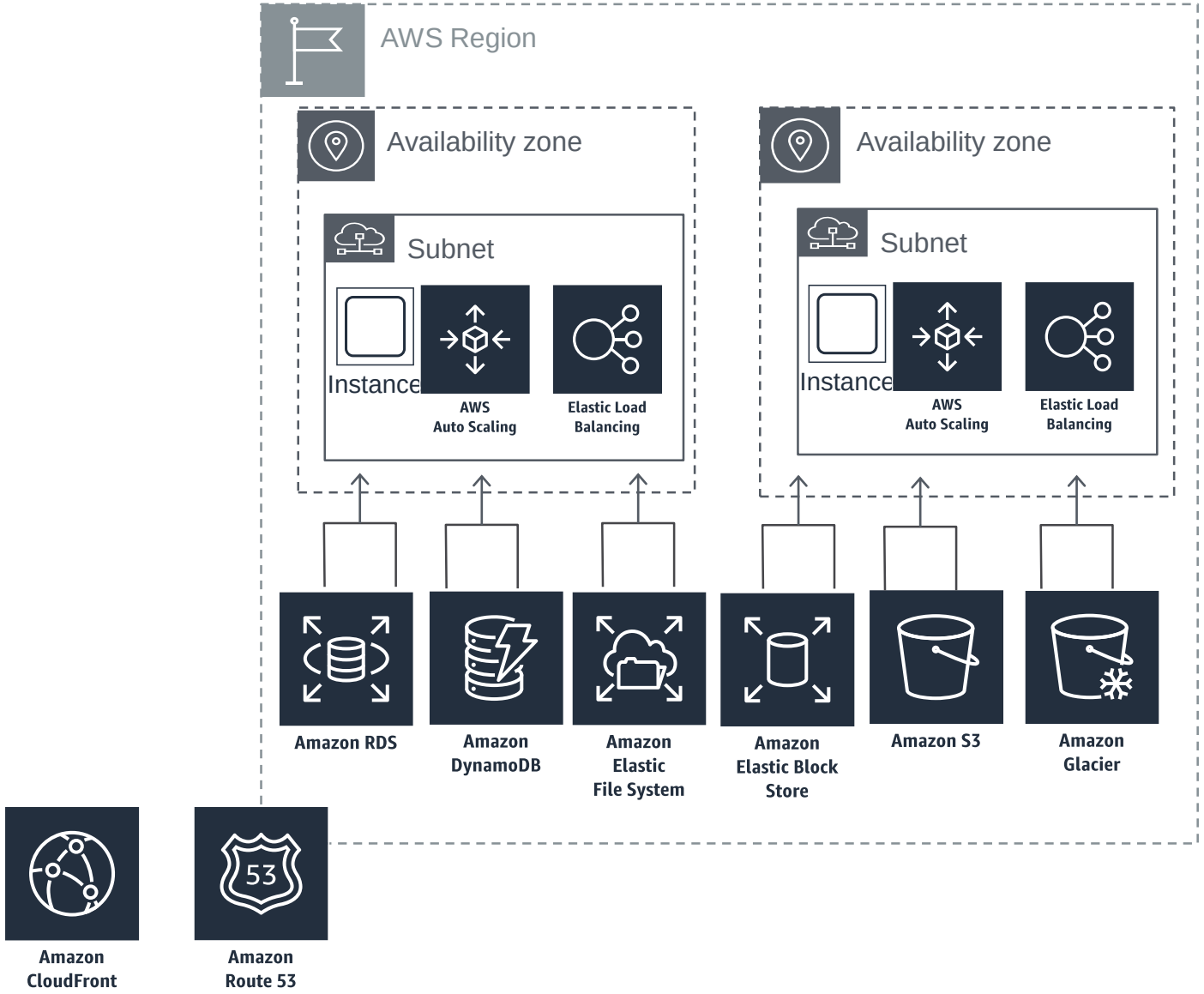


# **Exercise:** AWS Management Console

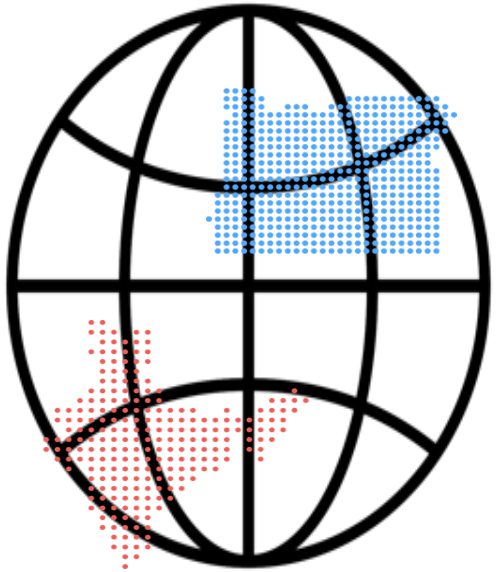
# AWS Regions



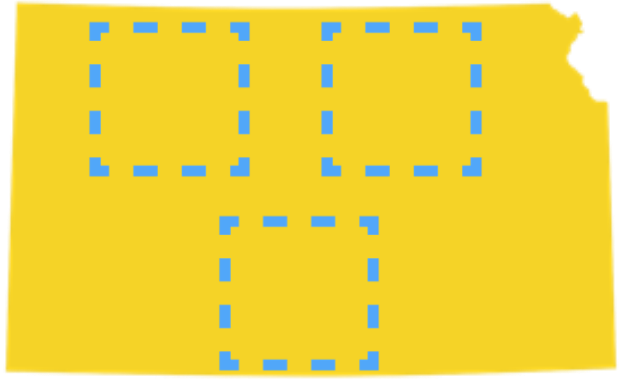
# Regions, Zones, and Supporting Services



# AWS Regions



Regions start off independent

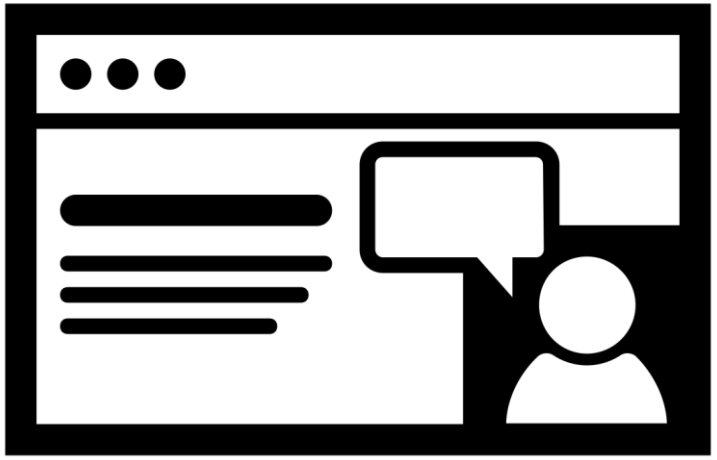


Regions have (multiple) Availability Zones

Data transfer charges apply within and across regions

Resources are usually not automatically replicated between regions

What AWS region would you suggest for Terra Firma?



# **Exercise:** Regions

# Choosing a Region

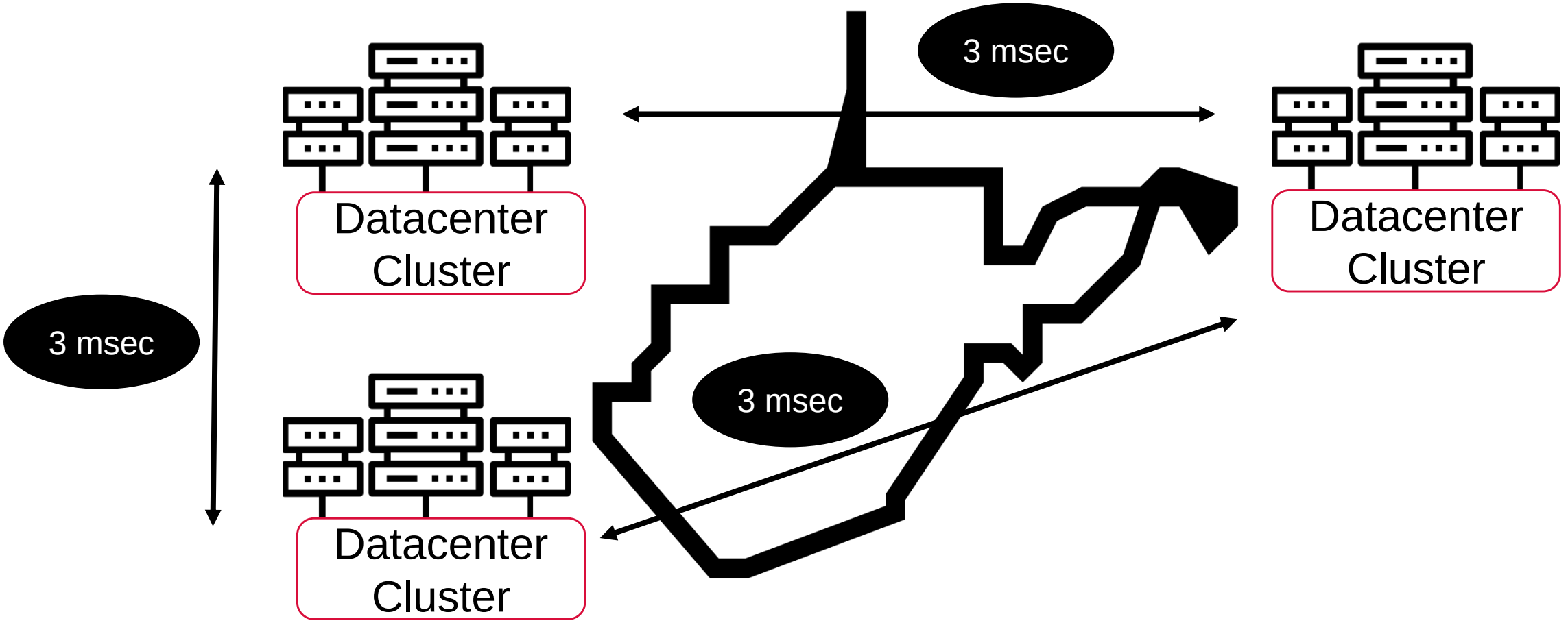
Latency – to on-premise customer location

Costs are different per Region, and AZ

Features are different per region

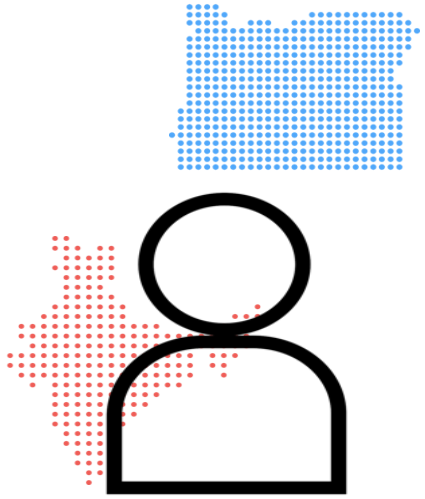
Compliance: Industry, country, and business

# Availability Zones (AZ)



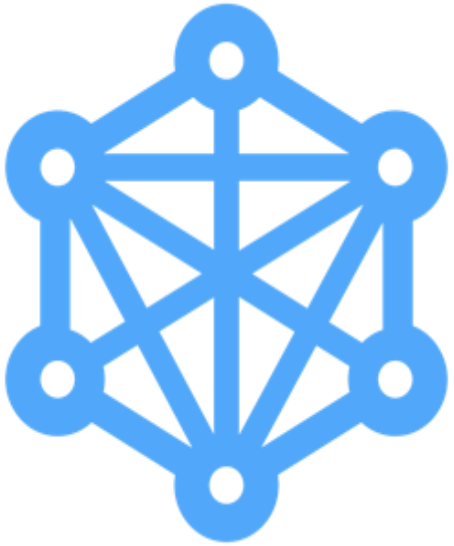
# Availability Zones (AZ)

Designed as an independent failure zone



AWS account has access to multiple regions and AZ's

Data transfer charges for outbound AZ traffic



Redundantly connected with multiple Tier-1 transit private fiber connections

# Workload Considerations

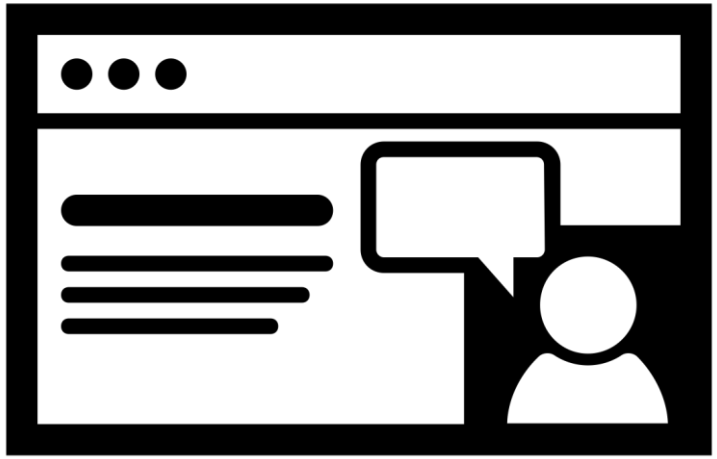
Select region matching compliance needs

Use multiple AZ’s for application failover

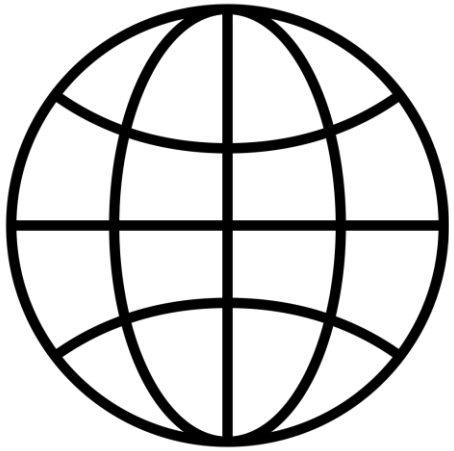
How many AZ’s should Terra Firma use for production?

# Single or Multi-AZ Design ?

| Single - AZ | Multi - AZ |
| --- | --- |
| No potential recovery when disaster happens<br>▪No high availability<br>▪No automatic failover<br>▪All regions have at least 2 availability zones | Better high availability design options<br>Scalability across AZ's provides HA<br>Load balancing (ELB) can balance across availability zones<br>Use Route 53 (DNS) Global accelerator to balance across multiple AWS regions |



# Exercise: Availability Zones



# Edge Locations

# Edge Locations

Provides a closer entry point to AWS through CloudFront

client

dns

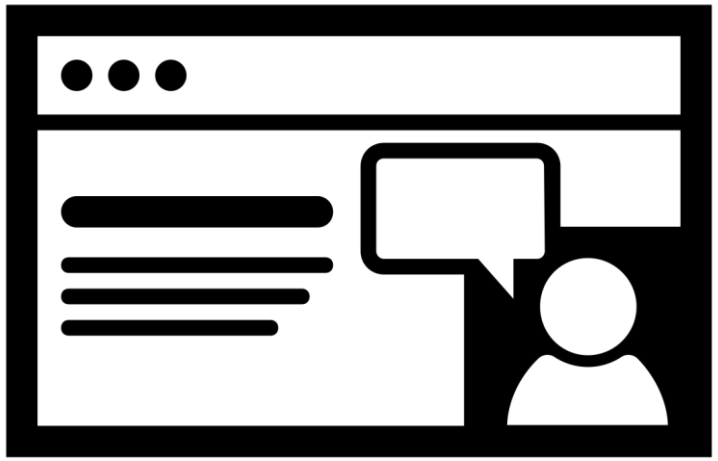
data

POPs in over 55 cities across 24 countries

- Edge services (Global)
  - Route 53
  - CloudFront
    - Regional edge cache locations (for larger AWS regions)
  - Redundant connections to multiple 3rd party communication services at each AWS edge location

# AWS Resource Locations

Resources are either Global, Region specific, or associated to an Availability Zone



# Exercise: Edge Locations

# Virtual Private Cloud

# What’s a VPC?

- Networking layer at AWS
- A logical and isolated data-center (virtual private cloud)
- Launch EC2 instances and various AWS resources into your virtual network (software data-center)
- Logically isolated from all other virtual networks hosted in the AWS cloud
- Networking platforms: EC2 - VPC and EC2 - Classic
- EC2 - Classic is not available for new customers (since December 2013)

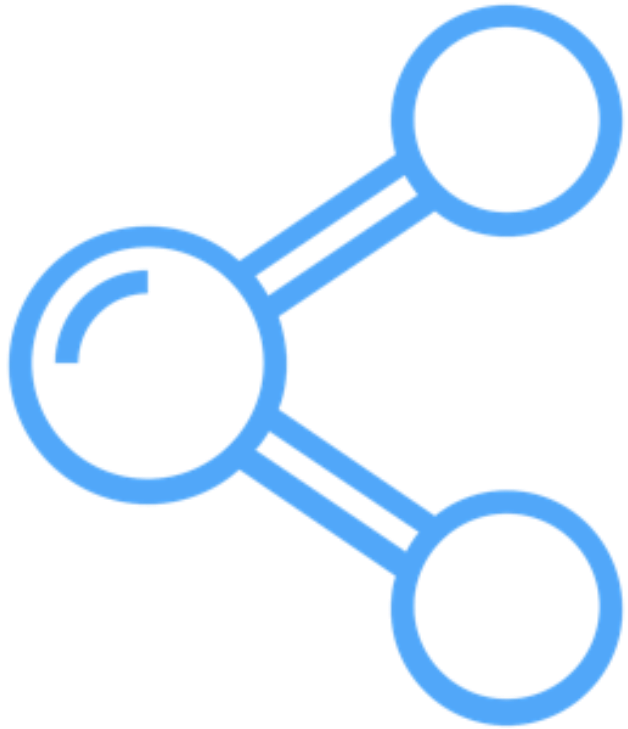
# Network Platforms

- EC2 – Classic
  - The original network infrastructure for EC2 instances
  - Instances run in a single flat network that you share with other customers
  - Doesn’t support enhanced networking, multiple IP addresses, changing security groups, etc.
- EC2 – VPC

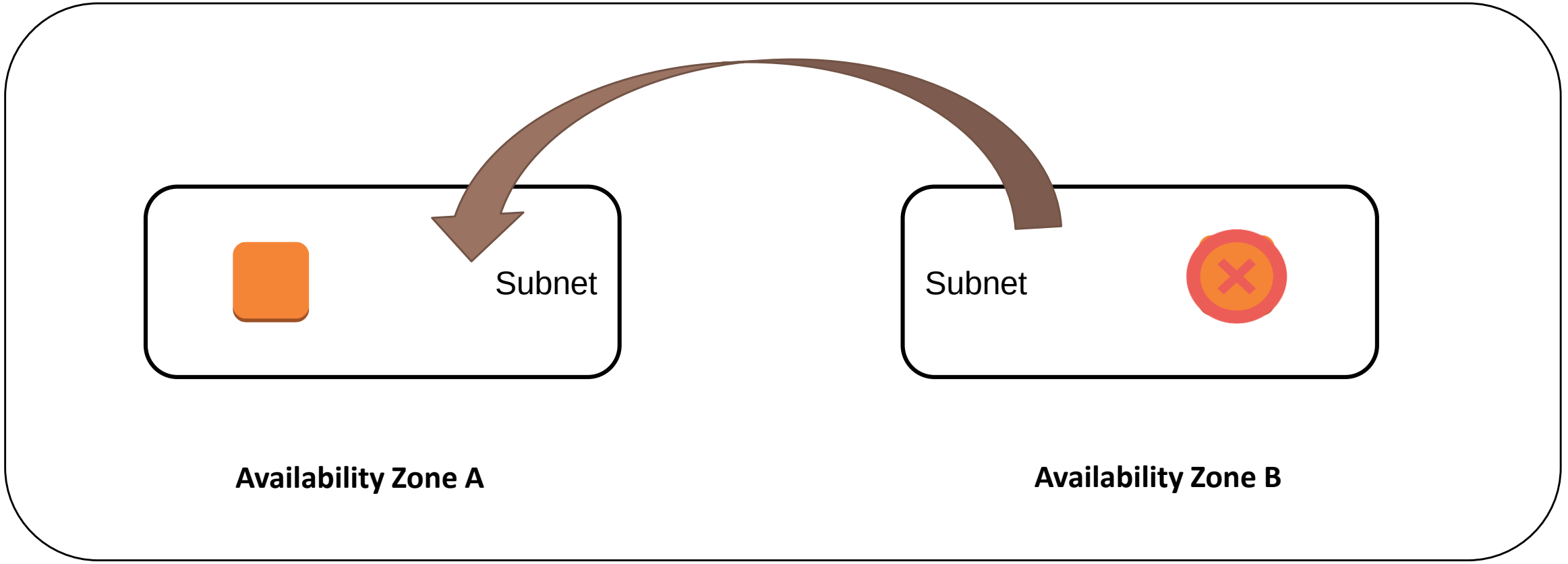
  Instances run in a virtual private cloud that is logically isolated to your AWS account

# Creating a New VPC

- When a VPC is created, it spans all the availability zones within the region
- Subnets can be created in each availability zone
  - Each subnet is defined by a CIDR block which is a subset of the VPC CIDR block
- Each subnet is assigned the default route table that enables local routing throughout the VPC



# VPC Design: Best Practice



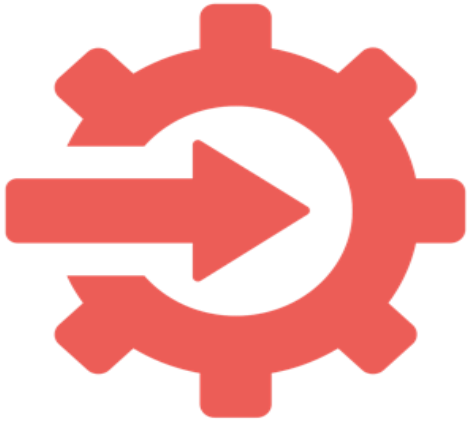
# VPC Design Decisions

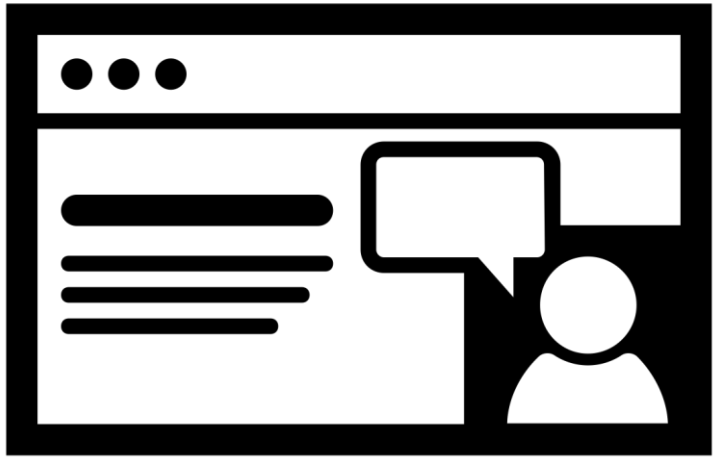
- IP address range
- EC2 instance placement
- Subnets
- Route tables
- External network connections
- Security settings –  per instance
- Security settings – per subnet

# VPC Core Components

- Subnets
- Route tables
- Security groups (SG)
- Network access control list (NACLs)
- Internet gateway (IGW)

- Virtual private gateway (VGW)
- Private endpoints
- Peering connections
- NAT gateway services (Instance / service)
- Transit gateway





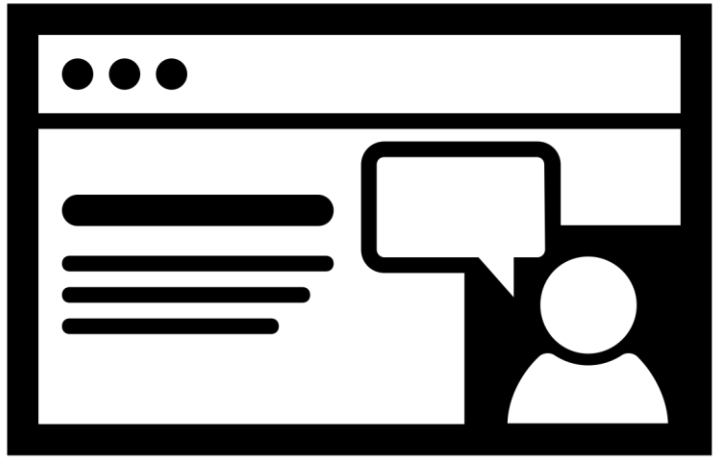
# Exercise: Create a Custom VPC

How many VPC’s should Terra Firma consider using for production?

# The Default VPC

- /20 CIDR Block is assigned by default
- An internet gateway is connected to the default VPC
- Default route table sends internet traffic to the internet gateway
- Default security group
- Default network access control list
- Default subnets
- Instances are assigned both a private and public IPv4 address

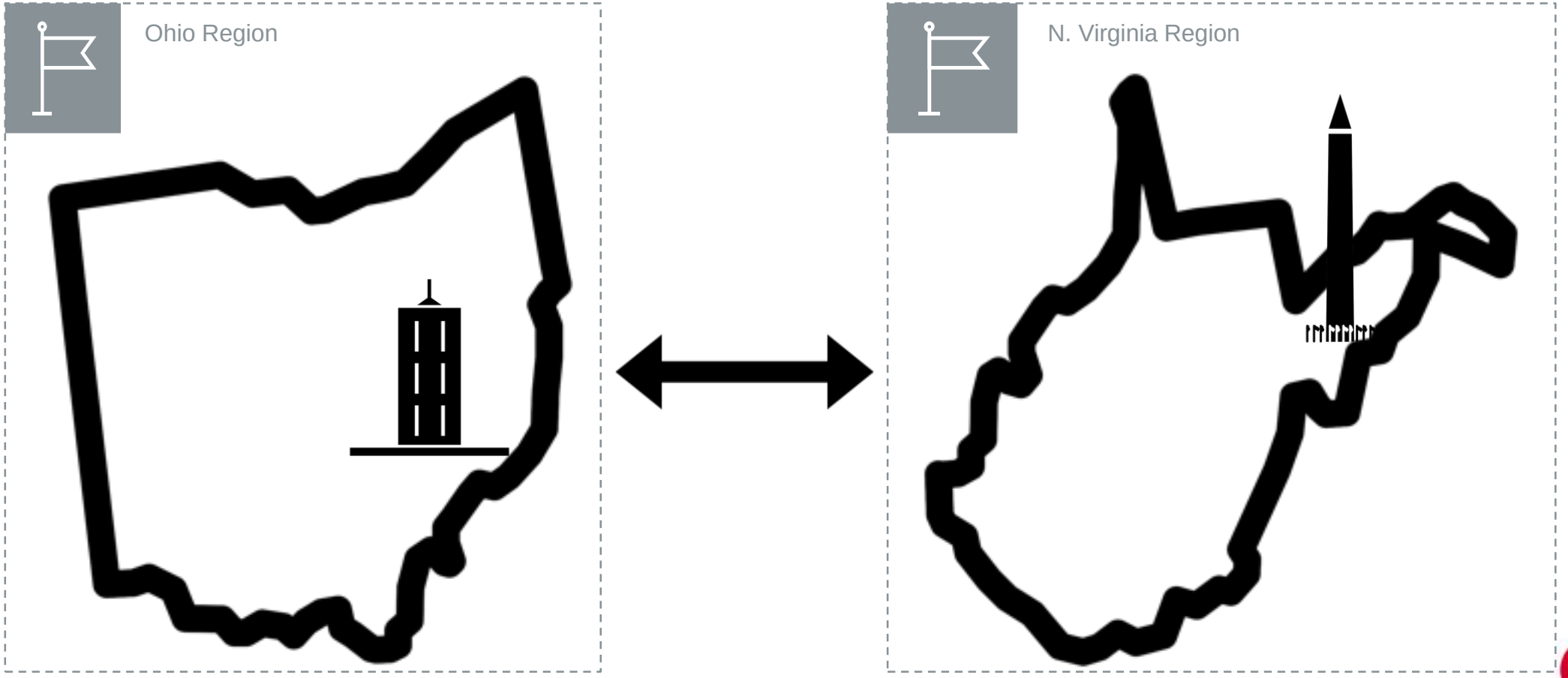
Should Terra Firma just use the Default VPC?



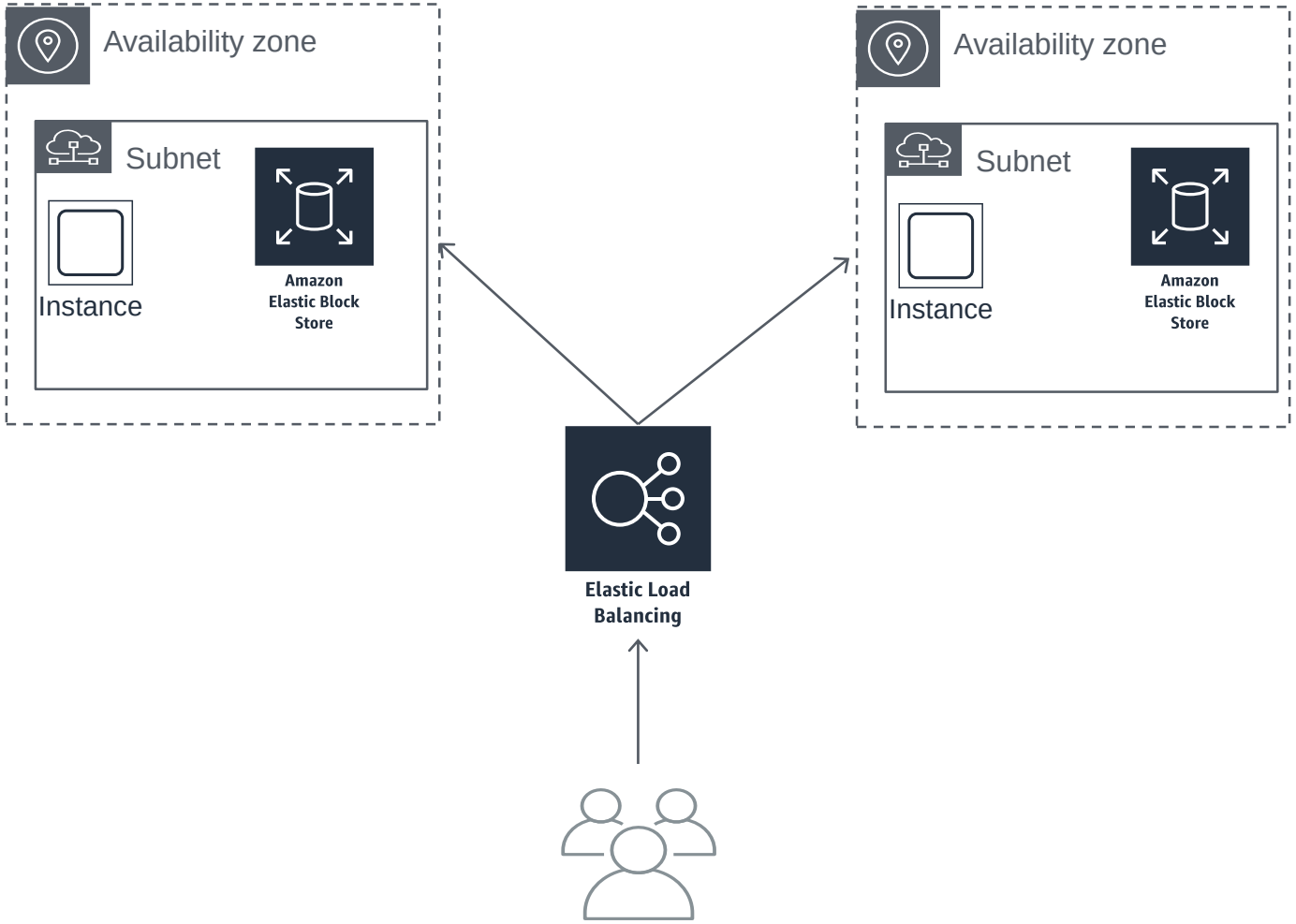
# Exercise: The Default VPC

# VPC Design

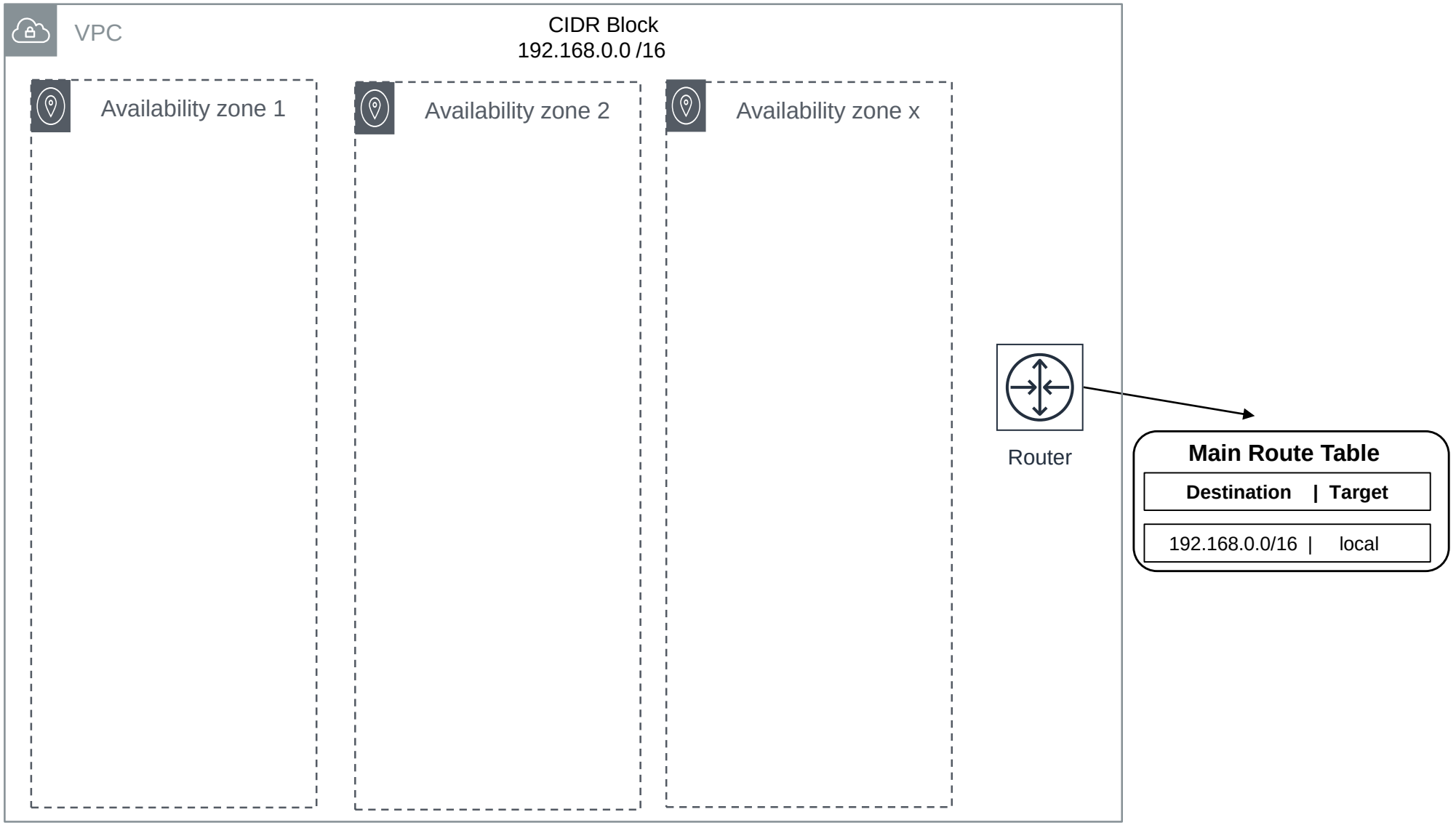
# Designing for HA and failover



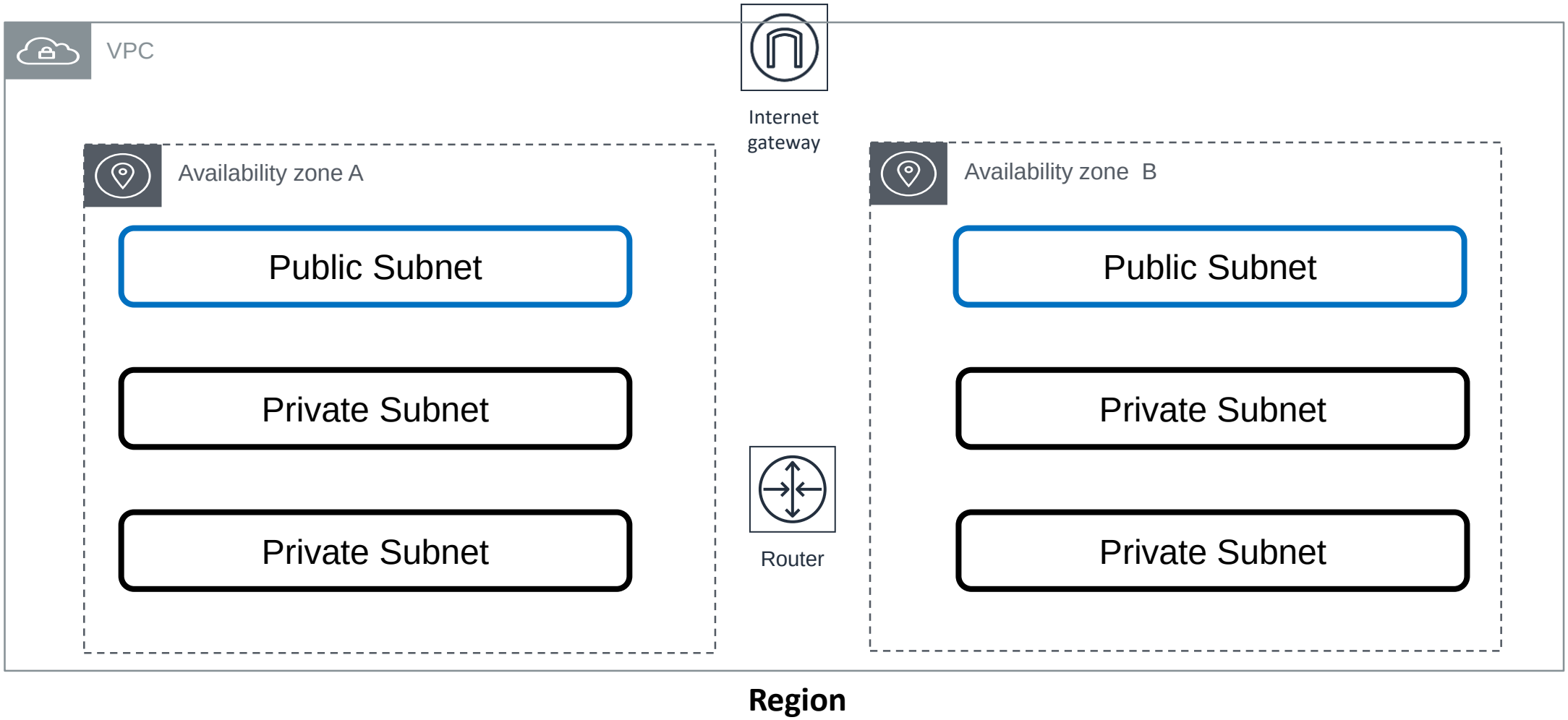
# Availability zones provide failover possibilities



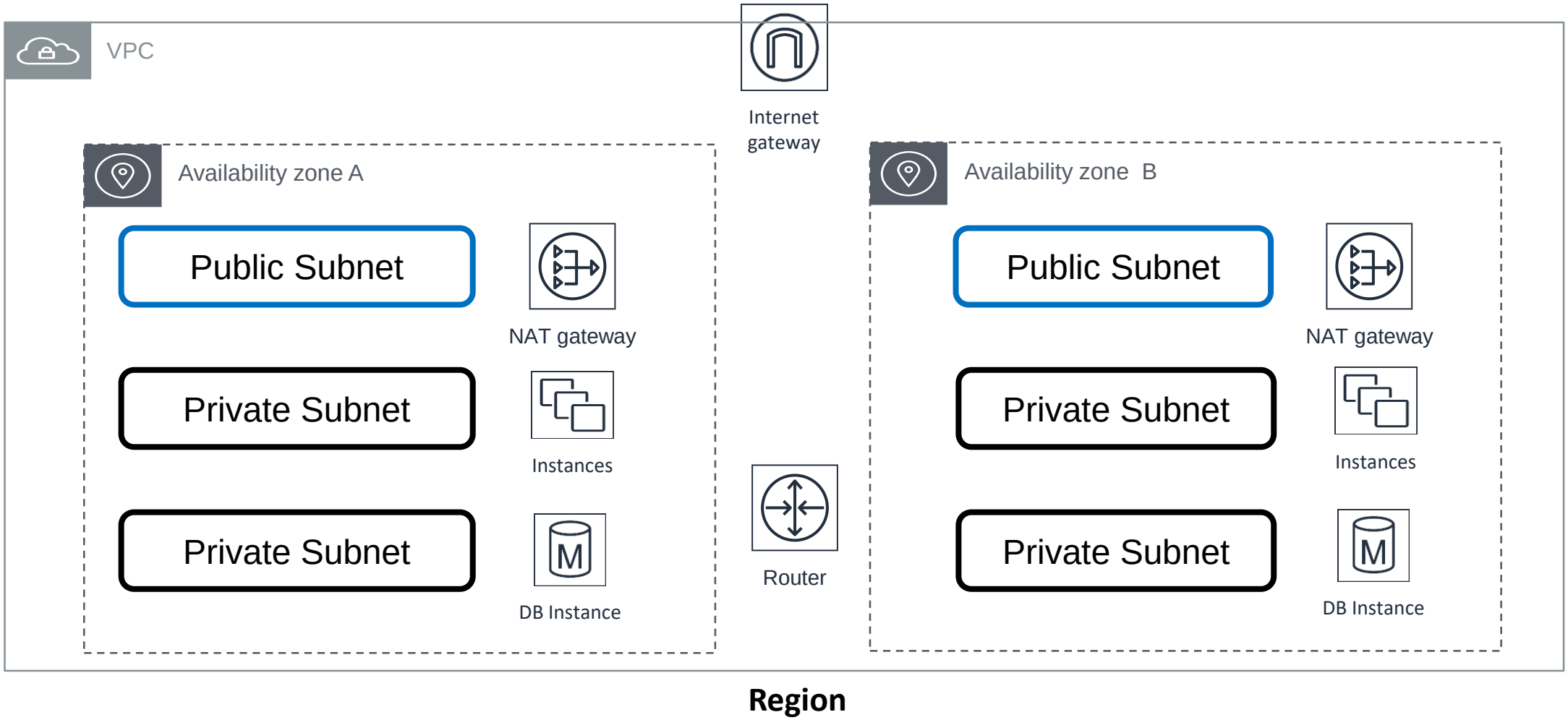
# VPC’s have Multiple AZ’s



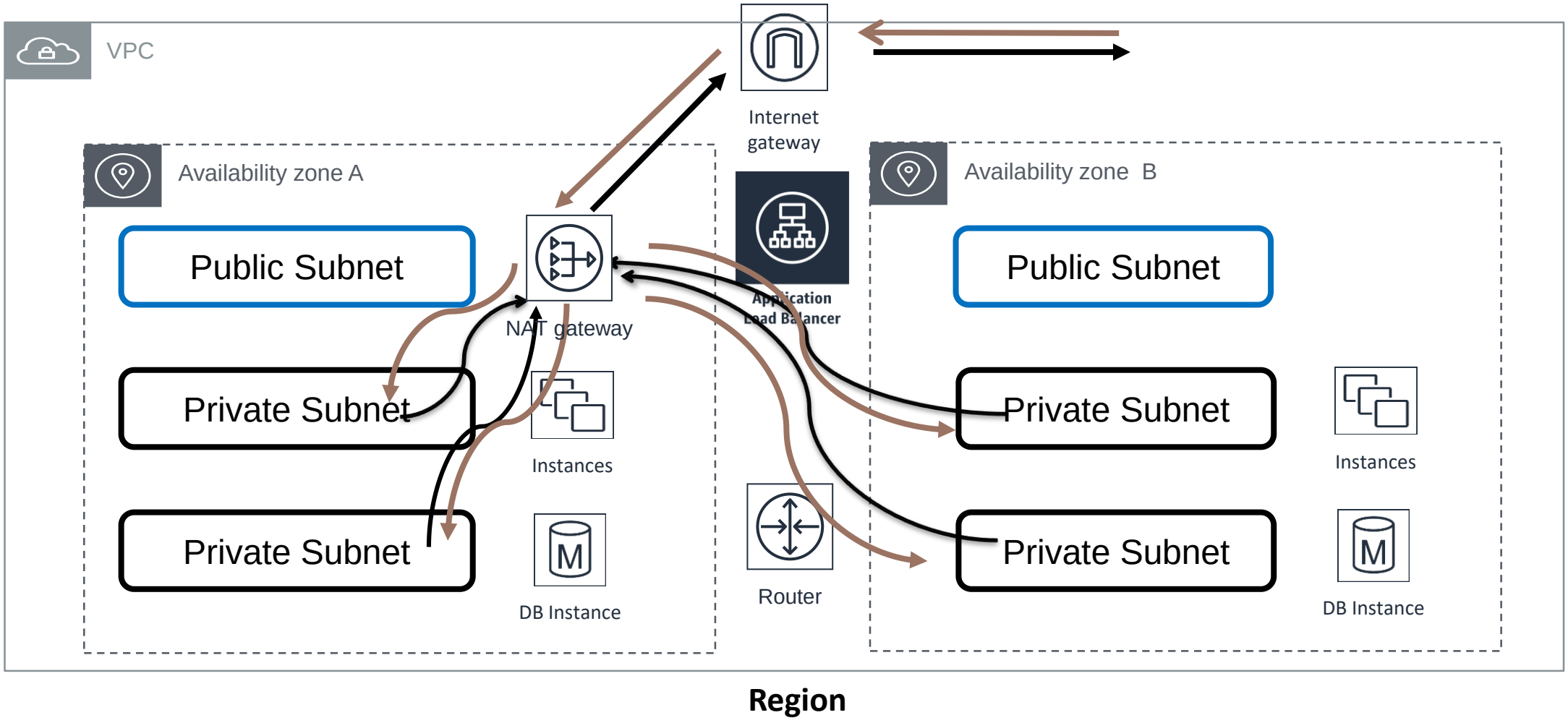
# Two Tier App


VPC
Internet gateway
Availability zone A
Public Subnet
Private Subnet
Private Subnet
Router
Availability zone B
Public Subnet
Private Subnet
Private Subnet
Region

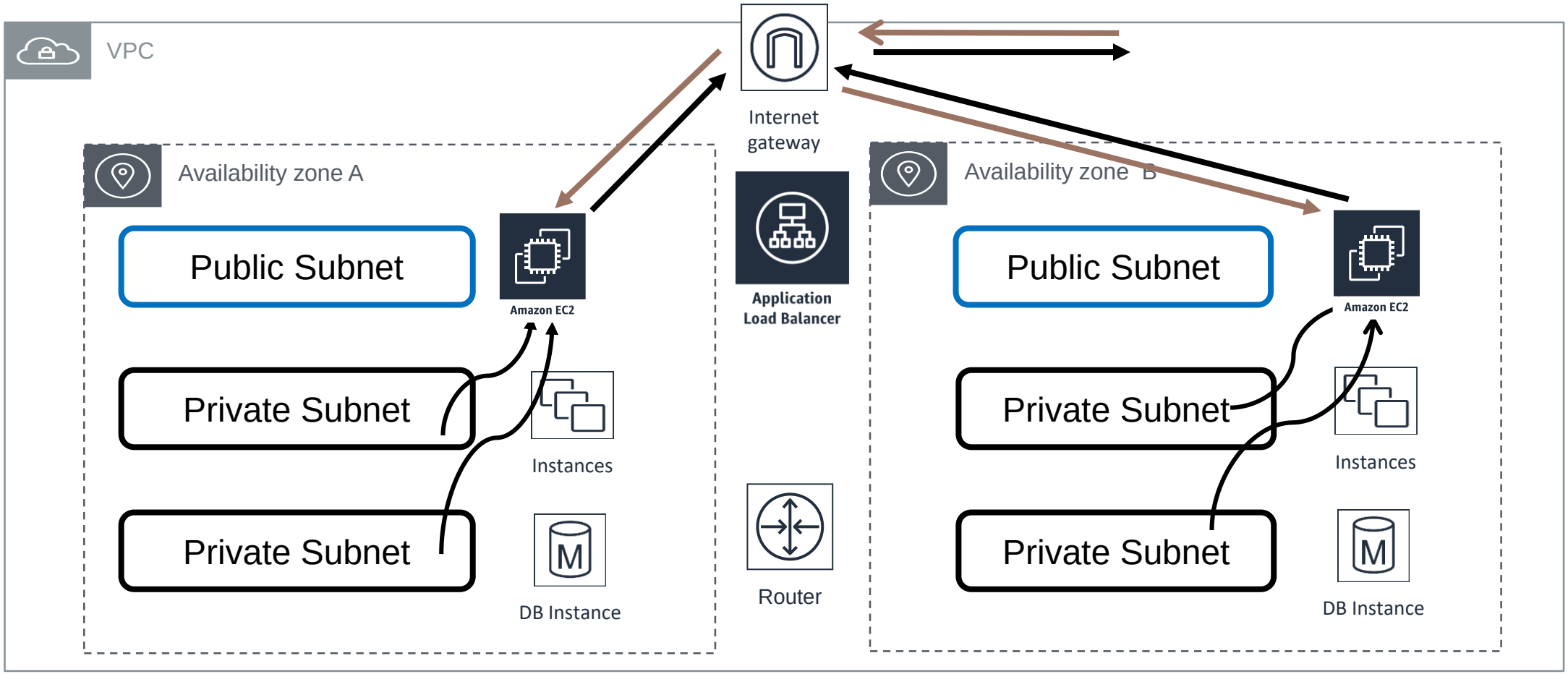
# Two Tier App

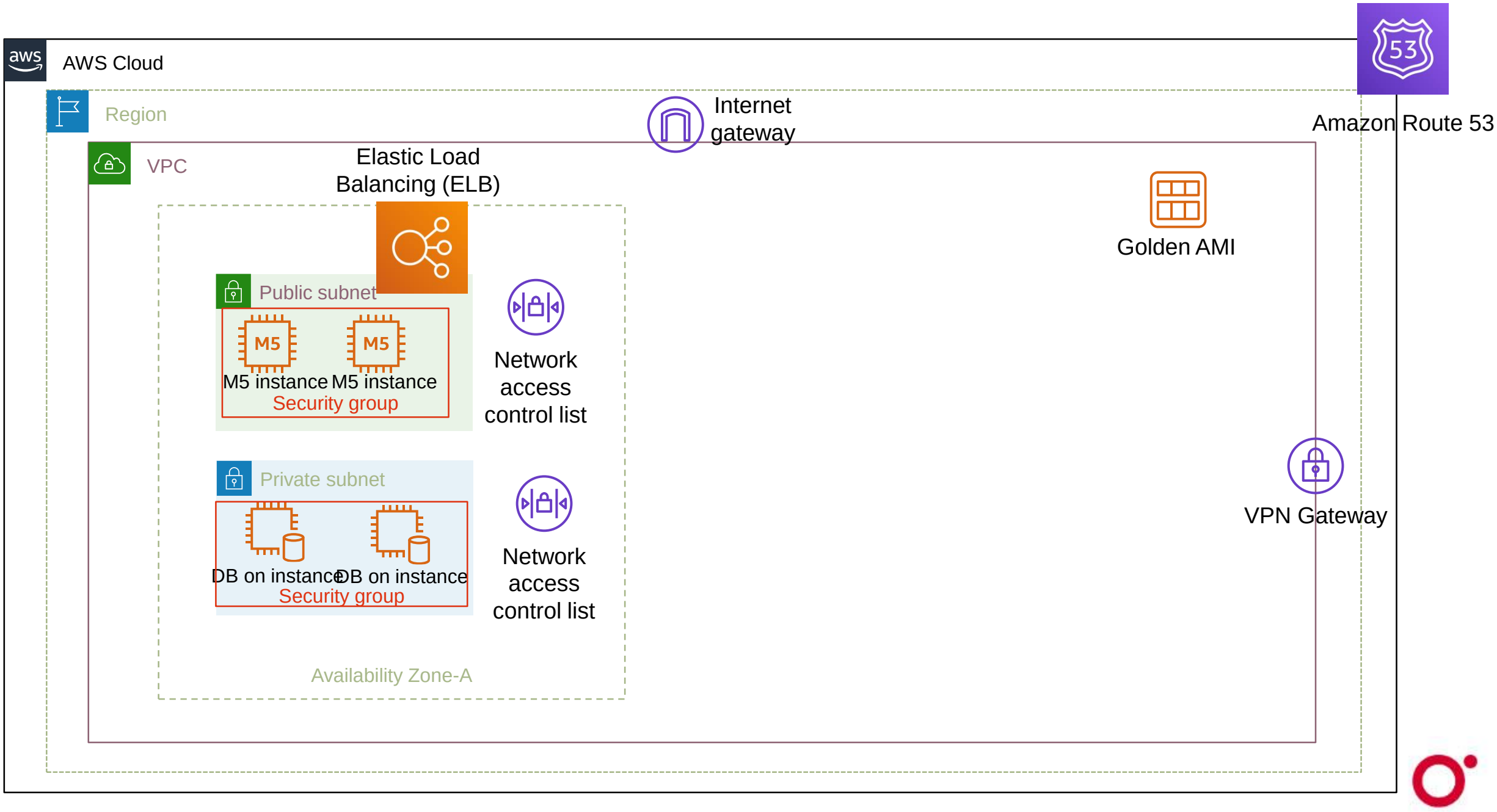

VPC
Internet gateway
Availability zone A
Public Subnet
NAT gateway
Private Subnet
Instances
Private Subnet
DB Instance
Router
Availability zone B
Public Subnet
NAT gateway
Private Subnet
Instances
Private Subnet
DB Instance
Region

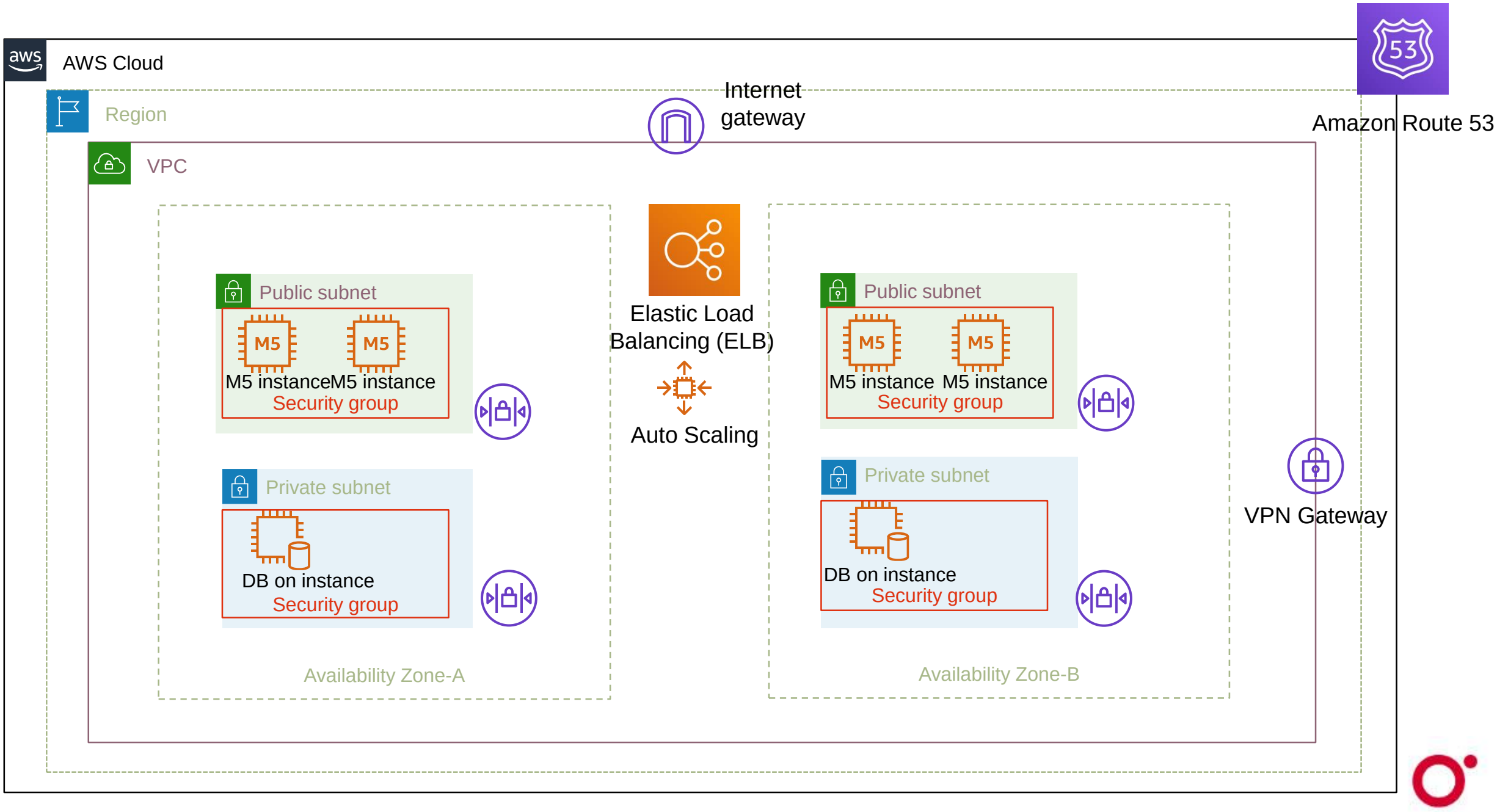
# Updates using NAT gateway service



# Updates using NAT instance




AWS Cloud
Region
Internet gateway
Amazon Route 53
VPC
Elastic Load Balancing (ELB)
Golden AMI
Public subnet
M5
M5
M5 instance
M5 instance
Security group
Network access control list
Private subnet
DB on instance
DB on instance
Security group
Network access control list
VPN Gateway
Availability Zone-A


AWS Cloud
Region
Internet gateway
Amazon Route 53
VPC
Public subnet
M5
M5
M5 instance
M5 instance
Security group
Elastic Load Balancing (ELB)
Auto Scaling
Public subnet
M5
M5
M5 instance
M5 instance
Security group
Private subnet
DB on instance
Security group
Private subnet
DB on instance
Security group
VPN Gateway
Availability Zone-A
Availability Zone-B

# Subnets and Addressing

# Subnets

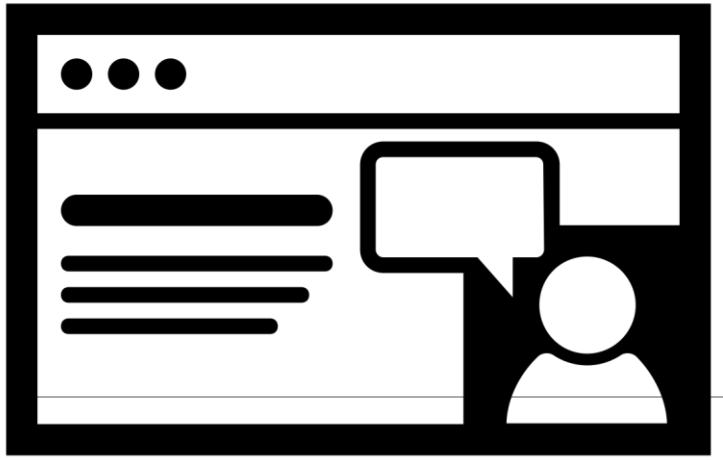
- Public or private subnets can be created in each availability zone
- Subnets cannot span across multiple availability zones
- If a subnet has traffic routed to an internet gateway it is defined as a public subnet
- Instances in a public subnet must have a public IP address, or an Elastic IP address to be able to communicate with the internet gateway
- A subnet that doesn’t route to an internet gateway is a private subnet

# Reserved Addresses

- The first four IP addresses and the last IP address in each subnet CIDR block are not available for use.
- Example: In a subnet with CIDR block 10.0.0.0/24, the following IP addresses are reserved:
  - 10.0.0.0: Network address
  - 10.0.0.1: Reserved for the VPC router (AWS)
  - 10.0.0.2: The IP address of the AWS DNS server is always the base of the VPC network range + 2
- 10.0.0.3: Reserved for future AWS use
- 10.0.0.255: Network broadcast address for the subnet
- Broadcasts are not supported across the VPC

# Public IPv4 Addresses

- A subnet attribute determines whether network interfaces within a subnet automatically receive a public IPv4 address
- Public IP addresses are assigned from AWS's pool of public IP addresses
  - These addresses are assigned and managed by AWS
  - When public IP addresses are released, they are added back into the common AWS pool



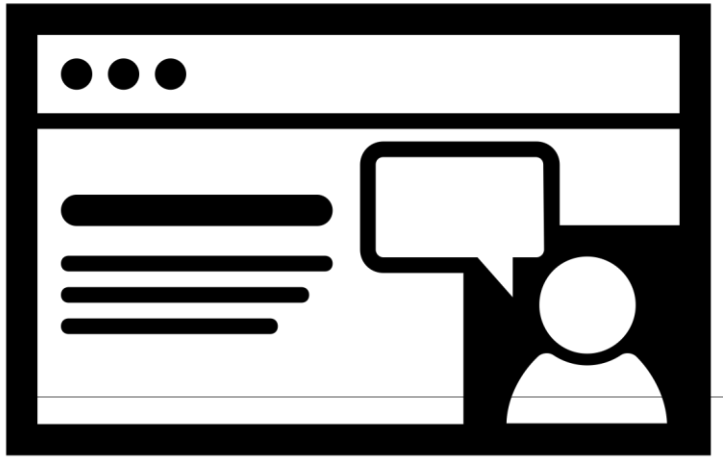
# Exercise: Create Subnets

How many subnets should Terra Firma use?

# Elastic IP Addresses (EIPs)

- An elastic IP address is a static public IP address
- Elastic IP addresses are assigned to your account
- An EIP is first allocated for use within a VPC; then assigned to a specific instance
- EIPs are specific to the region they are created in; they cannot be moved to a different region
- EIPs can be moved from one instance to another instance within the same VPC, or a different VPC within the same region

Do web servers need public IP addresses?



# Exercise: Order Elastic IP

# Route Tables

# Route Tables

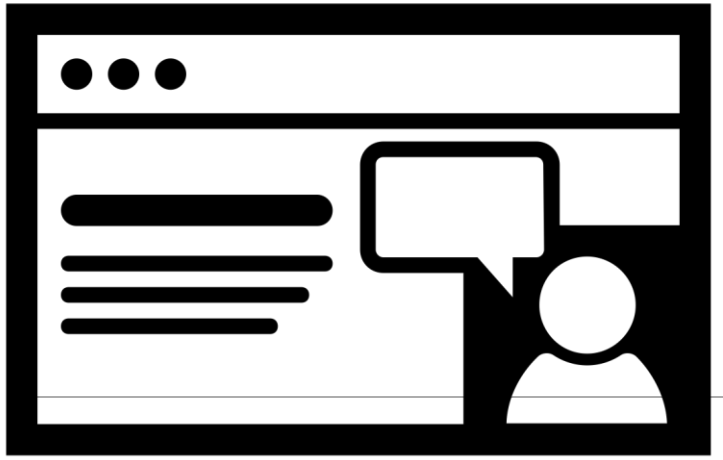
- Each route table contains a default route called the “local route”
  - This enables local communication within the VPC
- Each subnet is automatically associated with the master route table assigned to the VPC
- Each subnet must be associated with a route table
- Subnet traffic patterns are defined with a route table; all traffic leaving a subnet is processed by the route table to determine the destination of the traffic.
- Additional route table entries allow VPC traffic to connect to the Internet gateway (IGW), a Virtual private gateway (VPG), or to a NAT service

# NAT Services

# NAT Gateway Service

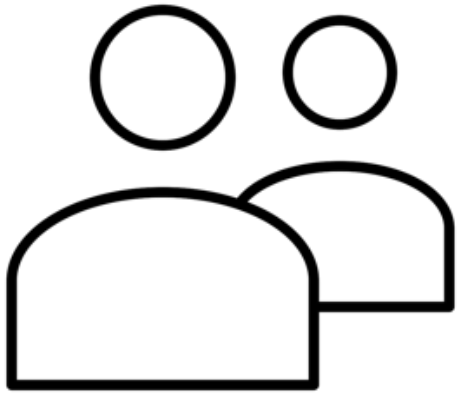
- The NAT gateway service accepts traffic from instances hosted on a private subnet
  - Translate the source IP address to the elastic IP address of the NAT gateway service
  - Forward the traffic request to the IGW
  - Returns incoming traffic to the private instance that made the request
- NAT gateway creation steps:
- Order a NAT gateway service
- Associate an EIP with the NAT gateway service

Are NAT services required for Terra Firma?



# Exercise: Order NAT Gateway Service

# Security Groups



# Security Group Details

- Security groups are defined as “virtual firewall’ protecting EC2 instance’s inbound and outbound traffic
- Security groups contain rules that control the inbound and outbound traffic to an instance
- Each instance launched into a VPC can have up to 5 security groups
- Each SG can have 50 inbound / outbound rules
- Each VPC can have up to 500 Security Groups
- When security groups are created, they are linked to a VPC and Ec2 instance

# Security Group Rules

- Rules apply to either inbound traffic (ingress) or outbound (egress) traffic
  - Allow rules can be specified
  - Deny rules can’t be specified
- Inbound rules – the source of the traffic, and the destination port or port range
  - Any protocol that is defined with a standard protocol and number is supported
- Outbound rules – the destination for the traffic and the destination port or security group
- Separate rules can be defined for both inbound and outbound traffic

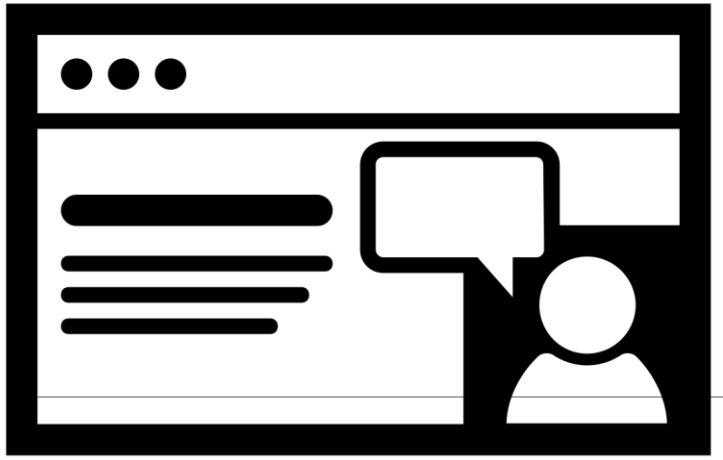
| Inbound | | | |
|---|---|---|---|
| **Source** | **Protocol** | **Port Range** | **Comments** |
| The security group ID (sg-*xxxxxxxx*) | All | All | Allow inbound traffic from instances assigned to the same security group. |
| **Outbound** | | | |
| **Destination** | **Protocol** | **Port Range** | **Comments** |
| 0.0.0.0/0 | All | All | Allow all outbound IPv4 traffic. |
| ::/0 | All | All | Allow all outbound IPv6 traffic. This rule is added by default if you create a VPC with an IPv6 CIDR block or if you associate an IPv6 CIDR block with your existing VPC. |

# Default Security Group

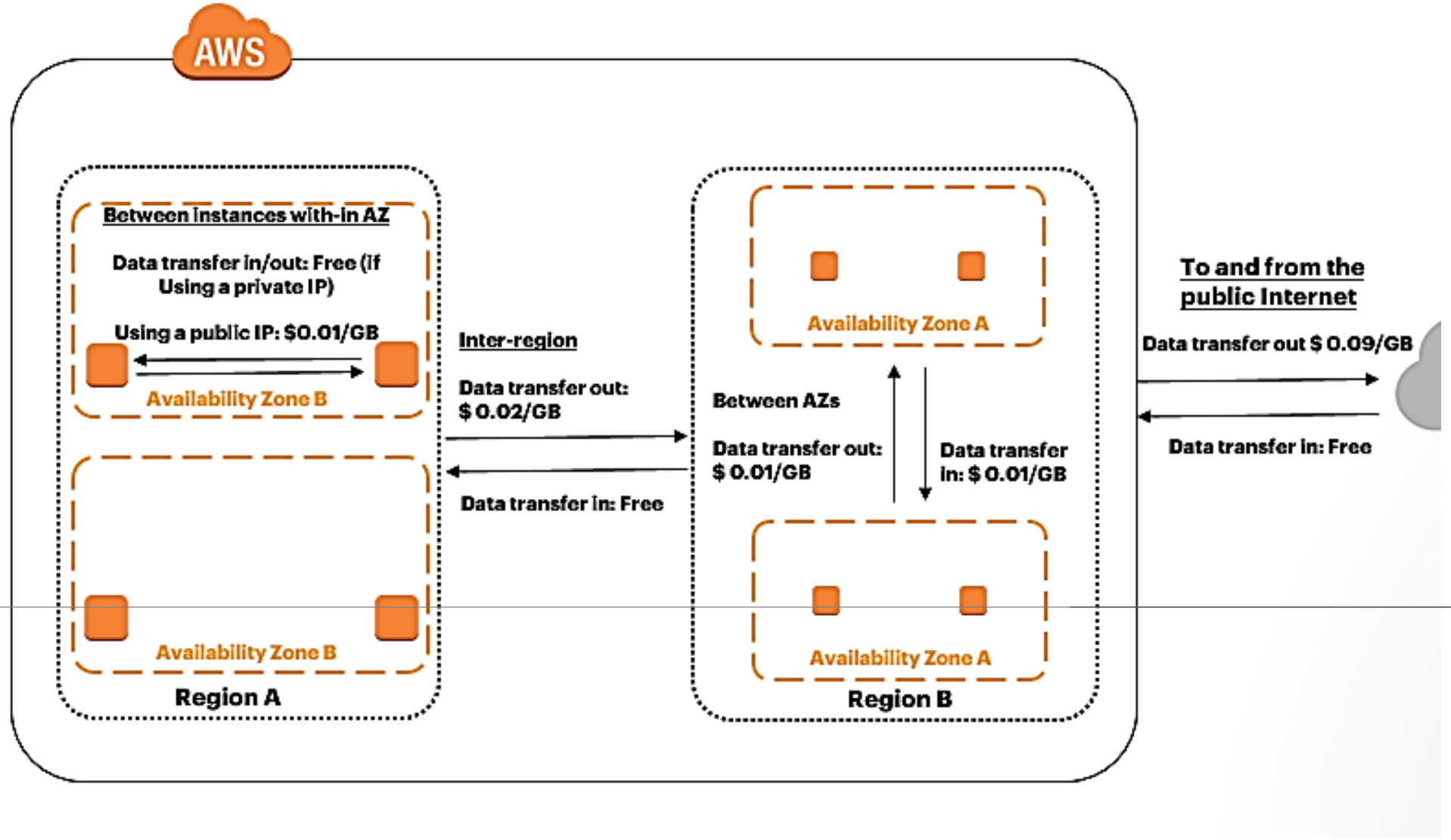
- Each EC2 Instance created in a VPC is associated with a default security group
- However you can change the association, or the default security group

# Security Group Operation

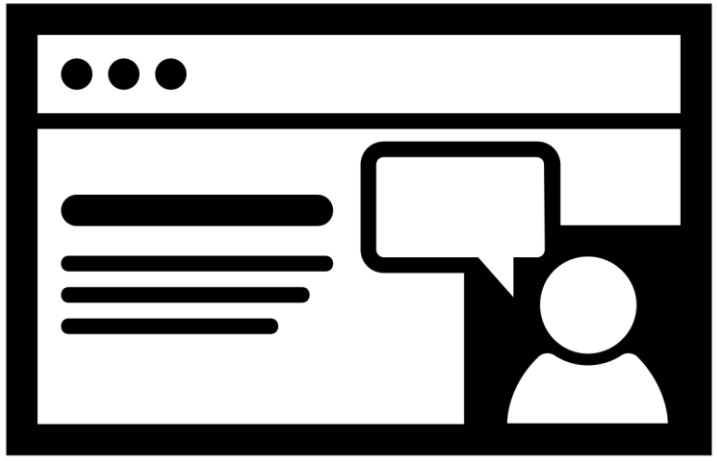
- Security groups are **stateful** – if a request is made to an instance inbound, the response traffic for that request is allowed out
- Responses to allowed inbound traffic can flow out
- Traffic can be restricted by IP protocol, service port, and source / destination IP address or CIDR block



# Exercise: Create Security Groups

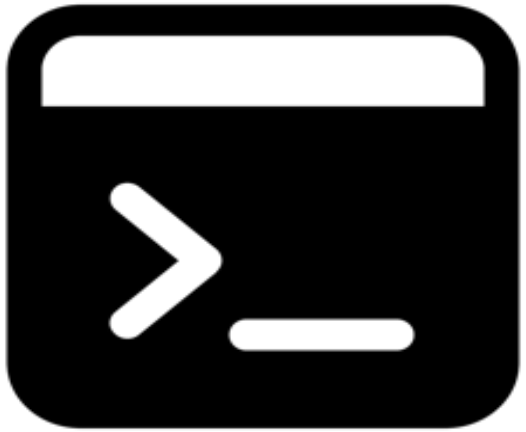

AWS
Between instances with-in AZ
Data transfer in/out: Free (If Using a private IP)
Using a public IP: $0.01/GB
Availability Zone B
Availability Zone B
Region A
Inter-region
Data transfer out: $ 0.02/GB
Data transfer in: Free
Availability Zone A
Between AZs
Data transfer out: $ 0.01/GB
Data transfer in: $ 0.01/GB
Availability Zone A
Region B
To and from the public Internet
Data transfer out $ 0.09/GB
Data transfer in: Free

# AWS Costs



# Exercise: Simple Pricing Calculator

# Accessing AWS Cloud Services



- Access to all AWS services is by firing a specific API call
- Application programming interface (API)
- Common Access Methods
  - The AWS management console
  - AWS command line interface (CLI) Windows, Mac, and Linux
  - AWS Tools for Windows PowerShell
  - AWS Software development kits (SDK)

# AWS CLI examples

Describe existing EC2 instances:

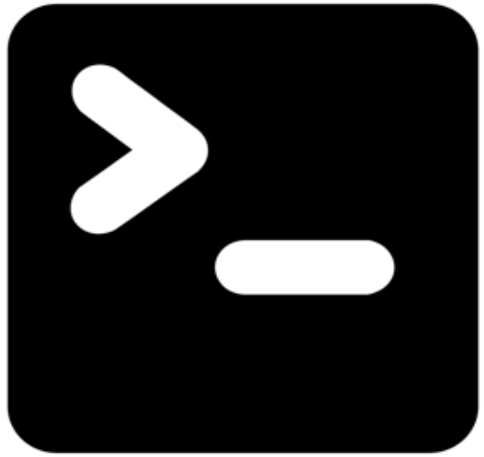
**$ aws ec2 describe-instances**

Start an EC2 Instance:

**$ aws ec2 start-instances --instance-ids i-1348636c**

Get help for a service:

**$ aws autoscaling help**



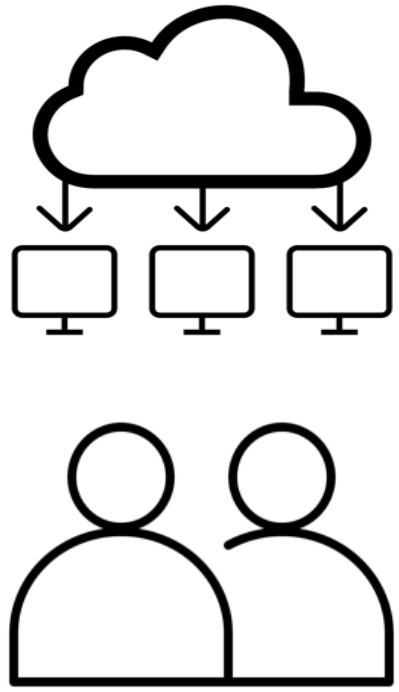
# PowerShell examples

**Launch an EC2 instance:**

```
New-EC2Instance -ImageId ami-c49c0dac -MinCount 1 -MaxCount 1   -
KeyName myPSKeyPair -SecurityGroupId sg-5d293231
- InstanceType m1.small -SubnetId subnet-d60013bf
```

**Create a security group:**

```
New-EC2SecurityGroup -VpcId "vpc-da0013b3" -GroupName
"myPSSecurityGroup" -GroupDescription "EC2-VPC Admin access"
```



# NACLs

# Network ACLs

- NACLs are an **optional security control** for subnets

  NACLs act as an “subnet firewall” for controlling traffic in and out of each subnet

  The default network ACL for a VPC allows all inbound and outbound IPv4 traffic

  A subnet can be associated with only one network ACL at a time

  A network ACL can be associated with multiple subnets

# Network ACL Rules

- Inbound Rule
  - Allow or deny for the specified traffic pattern
- Outbound Rule
  - Allow or deny for the specified traffic pattern

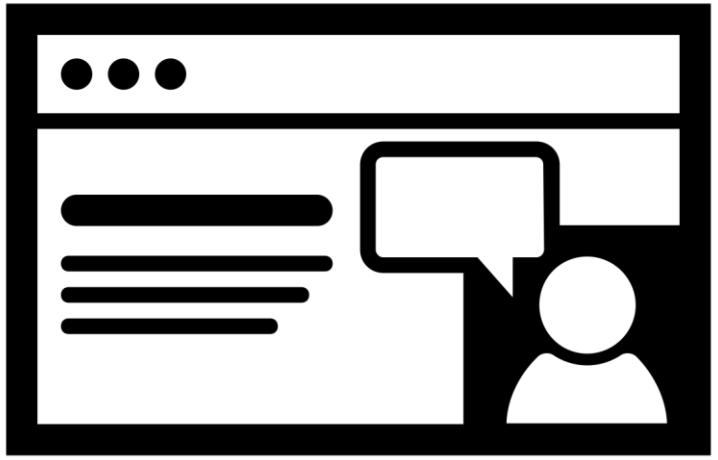
Each subnet within a VPC must be associated with a network ACL

# Network ACL Operation

- NACL rules are defined as **stateless**
- Rules are evaluated in order until a match is found
- Evaluation starts with the lowest numbered rule to determine if traffic is allowed in or out of the subnet associated with the network ACL
- Best practice: Create rules in multiples of 10, so adding new rules doesn’t cause problems in the future

# Security Groups vs NACLs

| Security Groups | NACLs |
| --- | --- |
| ▪Operates at the instance level | ▪Operates at the subnet level |
| ▪Allow rules only supported | ▪Allow and deny rules supported |
| ▪Stateful: return traffic is automatically allowed | ▪Stateless: return traffic must be explicitly allowed by a rule(s) |
| ▪All rules are processed before traffic decisions are made | ▪Rules are processed in numerical order before traffic decisions are made |
| ▪Applied to the selected instance elastic network adapter | ▪Applied to the subnet |



# Exercise: Configure Network ACLs

# VPC Options

# VPC Endpoints

- A private gateway connection between a VPC and S3 storage, or Dynamo DB table
- A private interface connection between an AWS service
- **Endpoint Creation Steps:**
- Specify the VPC
- Select S3 bucket, DynamoDB table, or AWS resource
- Define the IAM policy
- Update the route table

# Peering VPC’s

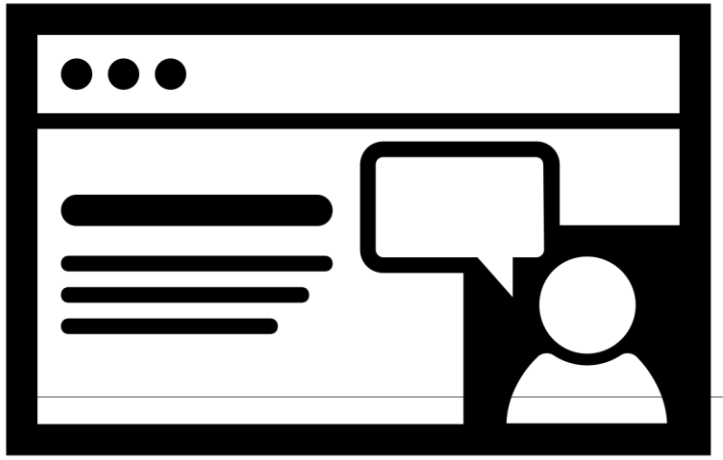
- Networking connection between two VPC’s
- Peer your VPC’s or between other account holders VPC’s using a private IP address
- Peering is a one-to-one relationship
- Peering connections are not transitive
- CIDR blocks can’t overlap in a peering relationship
- Peering connections can be created between VPCs in the same region
- Peering connections can be created between VPCs in different regions

# VPC Sharing

- Can replace VPC peering, accounts that will be sharing resources are either the owner or a participant.
- The owner creates and manages the VPC resources using the AWS Resource Access Manager
- Owners cannot modify or delete participant resources
- The participant in a shared VPC retains responsibility for the creation, management, and deletion of their resources which could be instances, RDS, and load balancer’s
- No peering required

# VPC Flow Logs

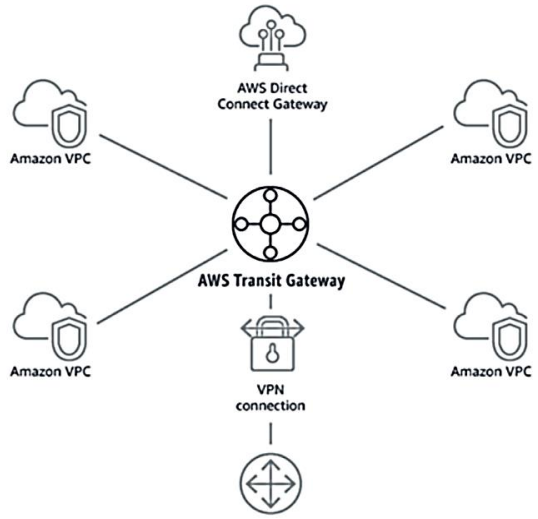
- Flow logs can be created for a VPC, a subnet, or a network interface
- Logs IP traffic to and from network interfaces in a VPC (accepted / rejected)
- Each NIC has a unique log stream
- Flow log data is published to a log group stored as a CloudWatch log group, or S3 Bucket
- Does not capture DNS, license, metadata, or default VPC router traffic



# Exercise: Enable Flow Logs

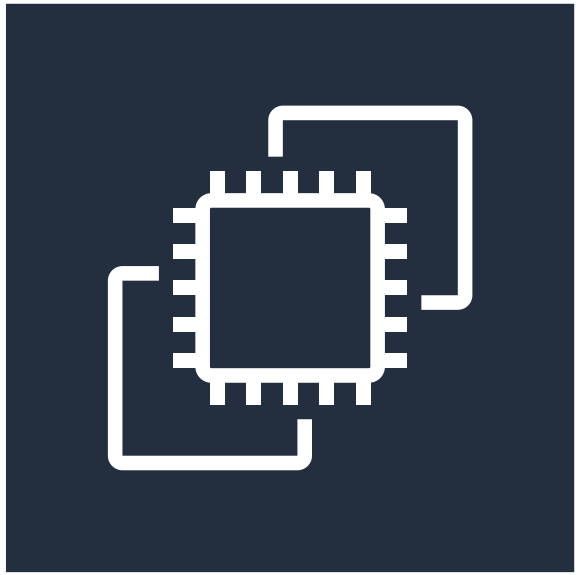
# Transit Gateway

- For larger companies with many VPCs, provisioned across many AWS accounts and regions
- Transit gateway allows you to connect a single gateway device that routes communicates to the networks that are connected to the transit gateway using a hub and spoke model
- Any VPC connected to the transit gateway is automatically routed to connected VPCs, direct connect gateways, and customer gateway routes
- Each transit gateway can connect to 5000 VPCs
- Network traffic can burst up to 50 Gbps



# EC2 Instances

# EC2 Instances



Amazon EC2

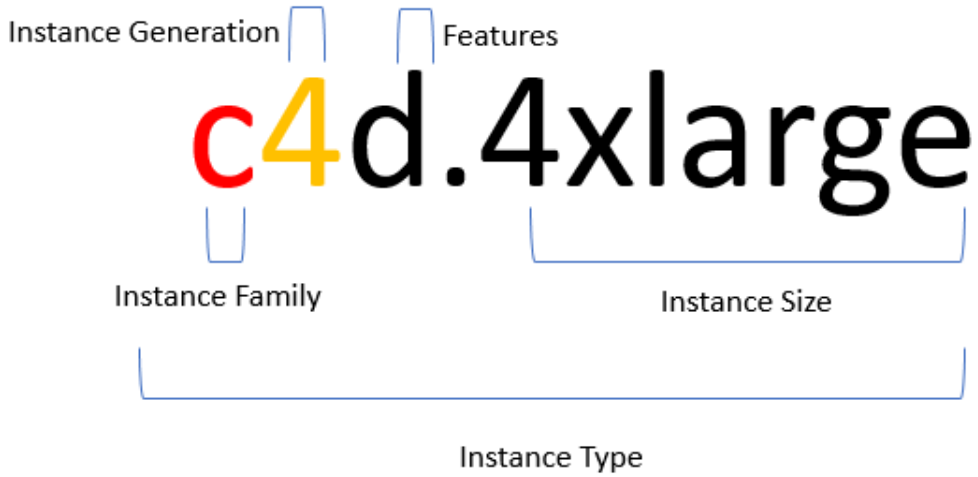
Virtual servers are called EC2 instances

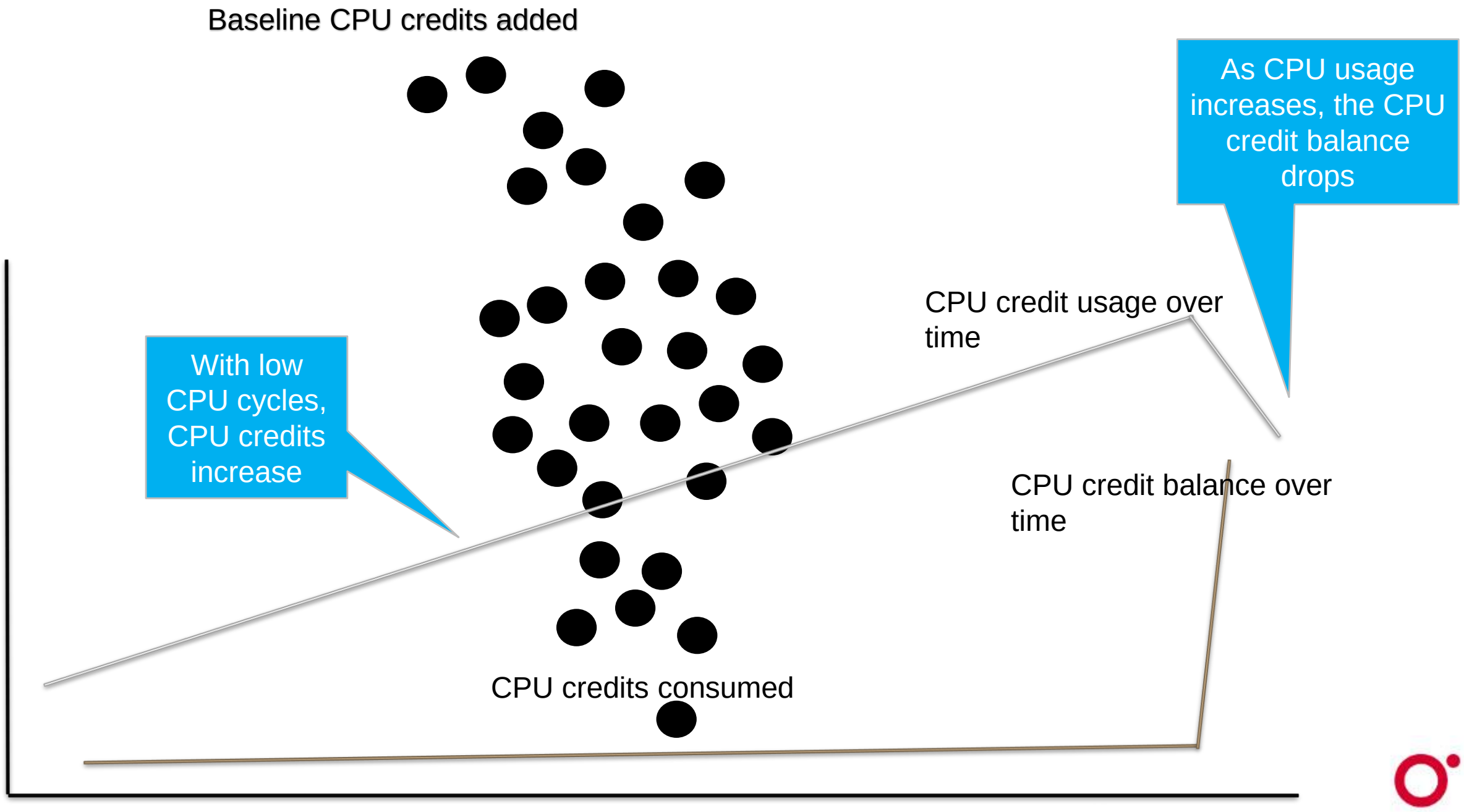
- Instance types – vCPU's, memory, storage (type and size), network speed
- Enhanced storage
- Enhanced networking

Performance builds

| | | |
|---|---|---|
| Compute | c4 | Extreme processing |
| Memory | r3 | Memory intense |
| Storage | i2 | Fast SSD storage |
| GPU | g2 | Graphic workloads |

# Instance families at AWS




Baseline CPU credits added
As CPU usage increases, the CPU credit balance drops
CPU credit usage over time
With low CPU cycles, CPU credits increase
CPU credit balance over time
CPU credits consumed

What type of instances should be considered for the human resources software?

Compute, Memory, or Storage performance builds?

# Amazon Machine Images

- AMI - Amazon Machine Images
- Defines initial s/w installed on instance when launched
  - O/S, state, system software
  - Launch parameters
- AMI Types
  - Published – AWS Marketplace
  - Published by AWS – Linux and Window versions / variants
  - Server migration service for on premise VMs to AMI’s
  - Generate from existing instance

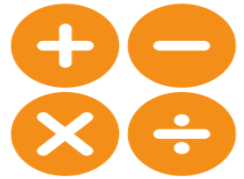
# Golden Image Maintenance

- Customize an EC2 instance and save configuration as an AMI
  - Launch (many) instances from customized AMI
- Update golden image
  - Launch (many) instances from modified AMI

# EC2 Pricing Options

- On-Demand – Billed by the second (for Linux instances)
- Reserved – All upfront, No upfront, Partial upfront (1, and 3 year), Convertible
- Scheduled – Monday, Wednesday, Friday 1-7PM
- Capacity reservations per AZ
- Spot Instances
  - Spot price (2-minute CW alarm)
  - Hibernate – until spot price decreases
  - Up to 6 Hour guarantee
- Spot Fleets
  - A mixture of spot, on-demand, RI, and spot pools

# Pricing Scenarios



Reserved Reserved Reserved
Scheduled Scheduled Scheduled

+

On-Demand On-Demand On-Demand

Reserved Reserved Reserved

+

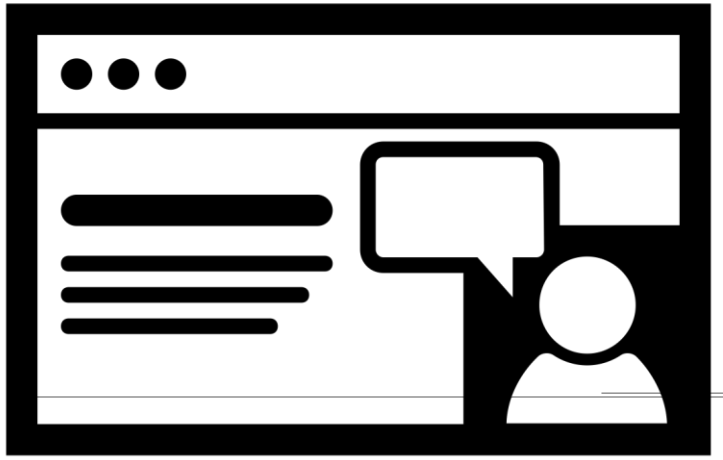
Spot Spot Spot

Reserved Reserved Reserved

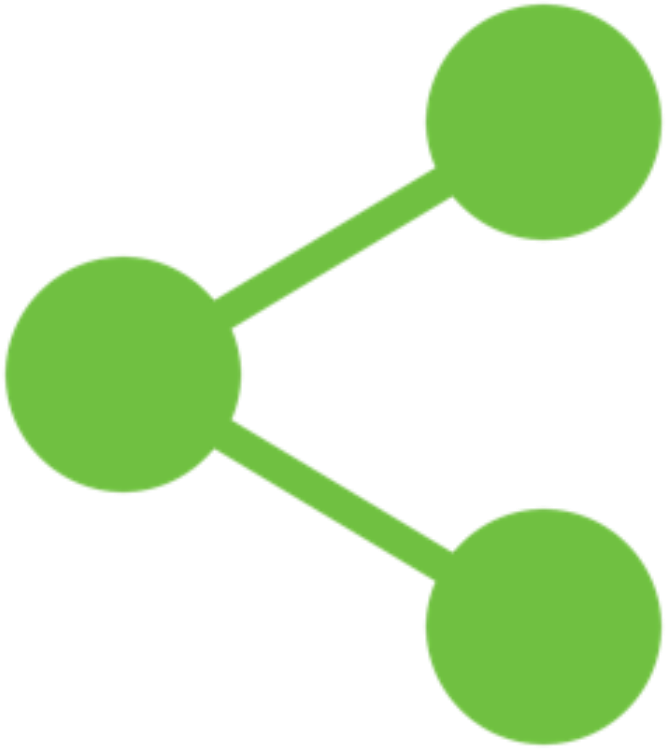
+

On-Demand Spot Fleet Spot Fleet

What type of purchasing option should be considered for the human resources SQL database instances?



# Exercise: Order an EC2 Instance

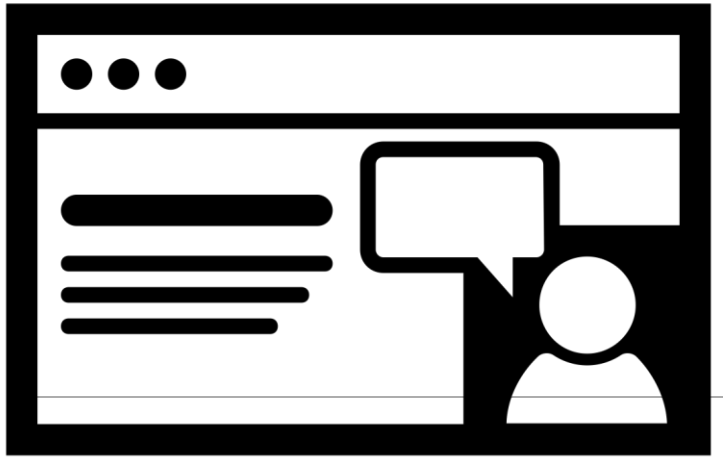


# EC2 Tenancy Options

- Shared tenancy (Default)
  - VPC can be set to dedicated tenancy
  - Dedicated instances – no sharing
- Dedicated Host – Whole host CPU core control
- Bare Metal
- Placement groups – instances on the same subnet within a single AZ

# EC2 Admin Tasks

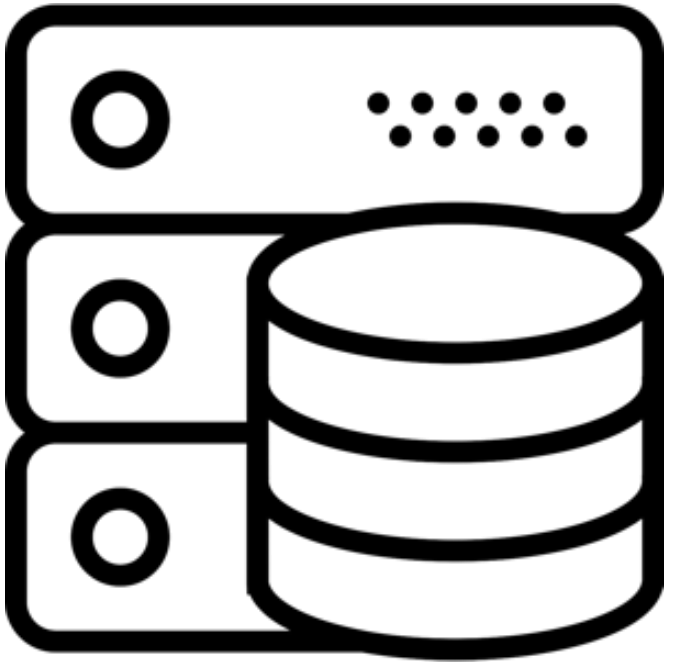
- Initial Logon
  - Public / Private key pair
  - Windows instances – decrypt p/w with private key
  - Linux instances – Private key is used to login via SSH
- Instance Lifecycle
  - Bootstrapping initial launch – User Data
  - Managing instances – Tagging
  - Monitoring instances – CloudWatch
- Modifying an Instance
  - Change instance type – Turn Off / Change instance type / Turn on  (New billing cycle)

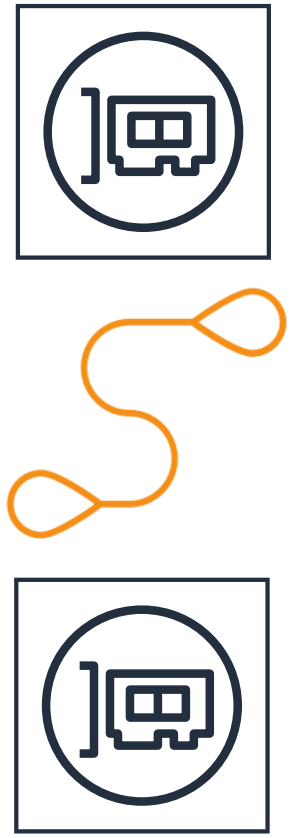


# Exercise: EC2 Administration

# EC2 Instances Stores

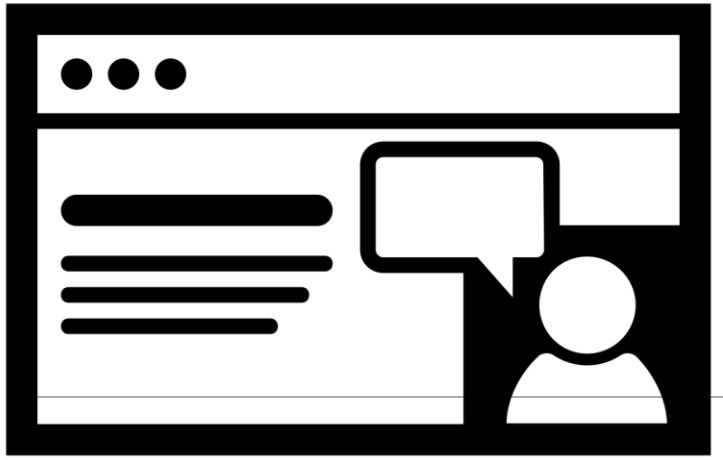
- Local disks attached to the bare metal server that hosts your instance(s)
- Called “Ephemeral storage”
- Temporary storage – buffers , cache, etc.
- Up to 24 TB depending on instance type
- Deleted when instance is stopped, or fails





# Elastic Network Interfaces (ENIs)

- Virtual network interface that can be attached to an instance within a VPC
- Each ENI can have one public IP address and multiple private IP addresses
- ENI’s once created are associated with a subnet, then instance
- Use case: Management networks, multi-homed instances, or Virtual appliances



# Exercise: Add Elastic Network Interface Card

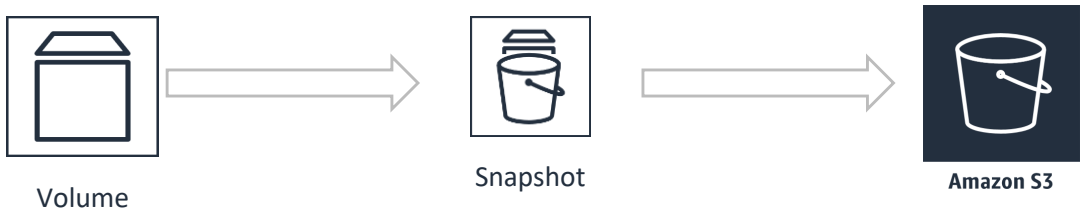
# Elastic Block Storage

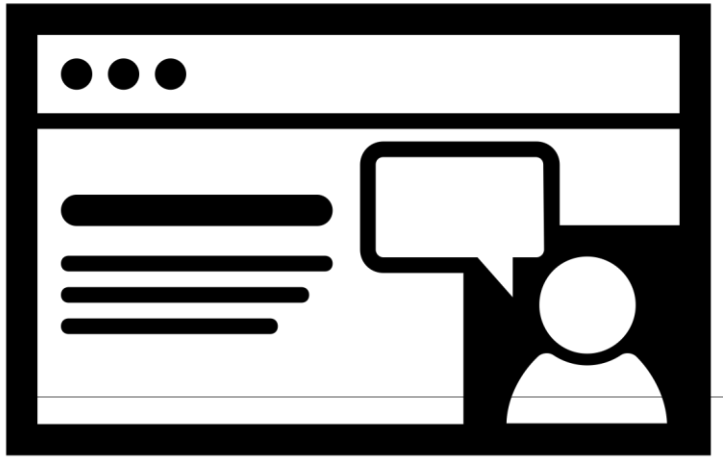
# Elastic Block Storage (EBS)

- Persistent Block Storage
  - Each EBS volume replicated within its AZ location
  - Single EBS volume attached to one instance
  - Multiple EBS volumes can be attached to one instances
- Magnetic Volume – 1 GB to 16 TB
  - Min: 100  Max: 2000
  - Throughput Optimized (500) / Cold storage (Min:100, Max: 20000)
- General Purpose SSD – 1 GB to 16 TB
  - ( 3 IOPS per GB) burstable to 10,000 IOPS
- Provisioned IOPS SSD  4GB – to 16 TB
  - Minimum  100 IOPS, Max: 64000 IOPS

# Protecting EBS Volumes

- Backup / Recovery snapshots
  - Snapshot
    - Point in time
    - Stores in S3 in AWS “Controlled storage”
- Create a Volume from a snapshot
- Increase the size of an existing EBS volume
- Detach, and re-attach existing volumes
- EBS volumes can be encrypted – KMS service handles key management





# Exercise: Create EBS Volumes and Snapshots

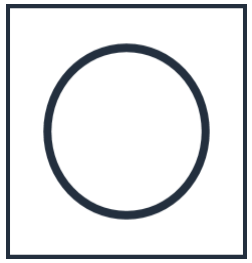
# Amazon S3

# What is S3 Storage ?

- Simple Storage Service
  - Secure, durable and scalable
- Object Storage – Cloud object storage
  - Pay only for the storage you use
  - Each object contains data and metadata
- Accessed over the Internet
- Private endpoint from a subnet hosted in a VPC
- Data is managed as an object using API calls and HTTP verbs (PUT,GET)
- Native interface to S3 using a Restful API (HTTP or HTTPS methods)

# S3 Buckets

Bucket



Object

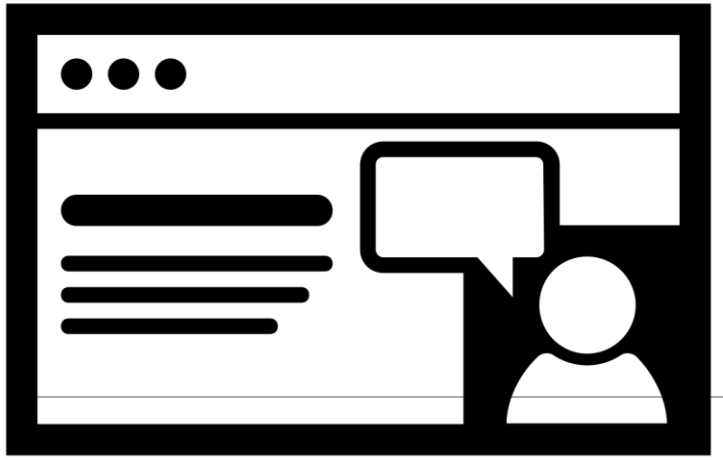
- Objects are stored in containers called buckets
  - Buckets are top-level management components
- Bucket names are global, must be unique across all AWS accounts
- Each object is identified, and accessed using a specified unique key
- Each bucket can be divided into folders (delimiters) \
  - Each bucket can hold an unlimited number of objects
  - You can’t mount a bucket, install software, open files, host a database
- Highly durable, scalable object store optimized for Reads

# S3 FYI

- S3 can store any type of data
  - Up to 5 TB max of single object
- Each object has a unique key
  - Key = filename
  - Must be unique within each bucket
  - Multi-part upload for objects greater than 5 GB
  - Bucket contents can be copied to buckets in other regions (additional costs)
- Metadata describes the data
  - System metadata – AWS  date, size, content-type
  - User metadata – tags specified only at the time the object is created

# S3 Storage Classes

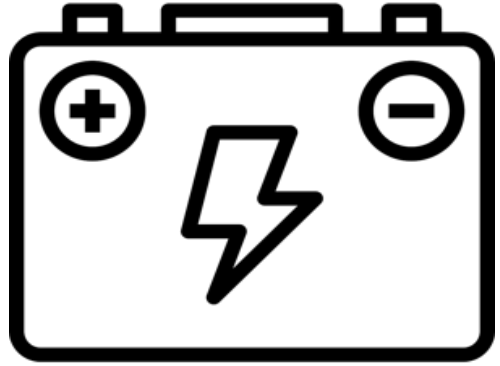
- S3 Standard – no minimum
- S3 Intelligent-tiering – monitor and move
- S3 Standard-1A – min 30 days
- S3 One Zone-1A – one AZ – 30 days
- S3 Glacier – 90 days
- S3 Glacier Deep Archive – 180 days



# Exercise: Create S3 Bucket

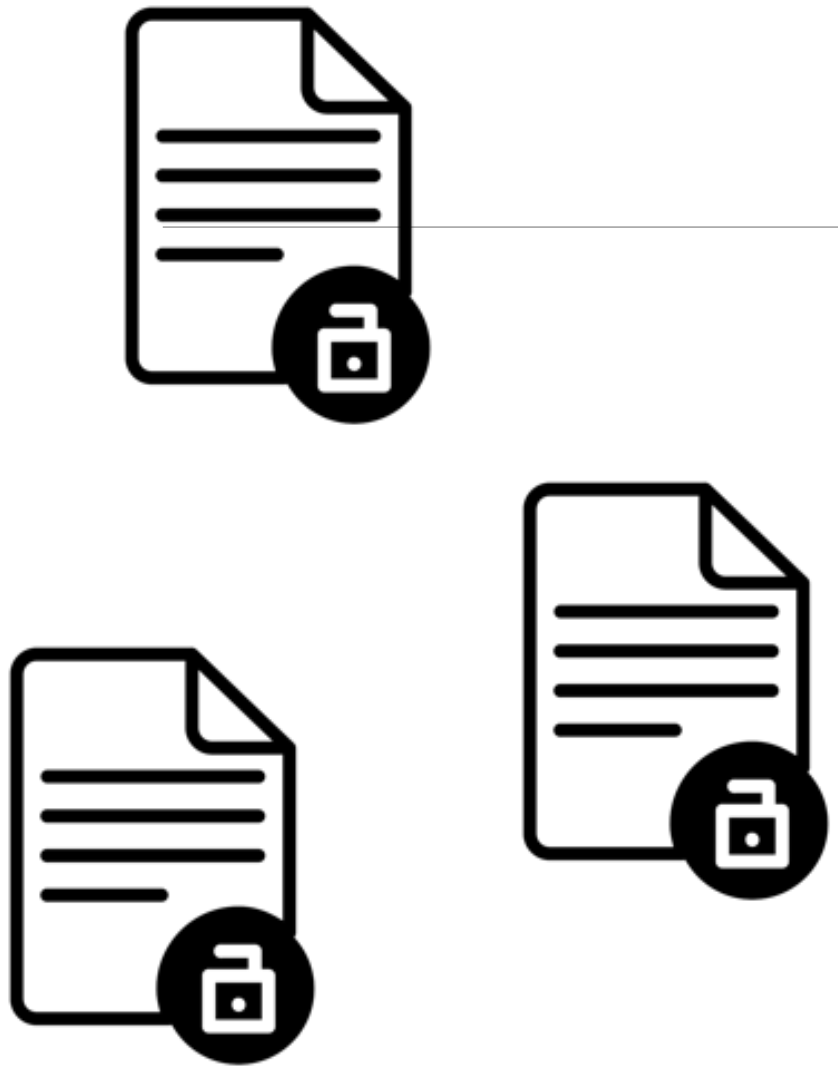
# S3 Durability

- Stored in multiple devices in multiple facilities, within a region
  - Designed to sustain concurrent loss of two facilities without loss of data
- Standard
  - 11 9's durability
  - 4 9's availability
  - Over a given year
- Standard 1-A
  - 4 9's durability



# S3 Consistency

- Objects are eventually consistent
- Multiple copies means replicated storage
- PUT’s to new objects – read after write consistency
- PUT’s to existing object – eventual consistency



# S3 Object Replication

- Cross-Region Replication
  - Asynchronous replication from source bucket in one region to bucket in another region.
  - Helps move data closer to end-users
  - Compliance / additional durability

# Access Control

- Only owner has access by default
  - Private by default
- Coarse grained – S3 acl
  - Read / Write / Full Control at object level
- Fine-grained – bucket policies
  - Associated with the bucket / not an IAM security principal
  - Can specify access from where, who can access, and what time  of day
- IAM polices can also be created for control
- Can be associated with different AWS accounts

# Managed Encryption

**AWS KMS** – Managed service allow you to generate, store, enable / disable and delete symmetric keys

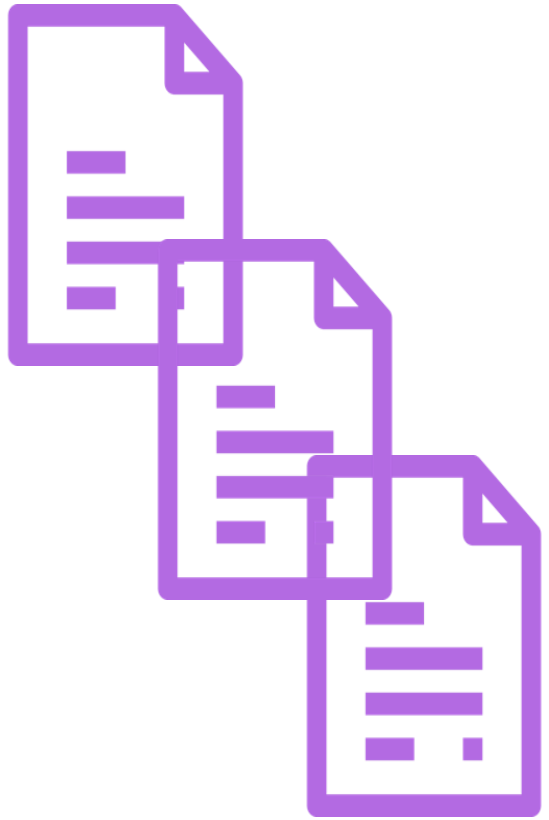
- Customer managed keys – Each CMS is per customer and is used to encrypt and decrypt data
- Data keys – Used to encrypt data objects within data storage

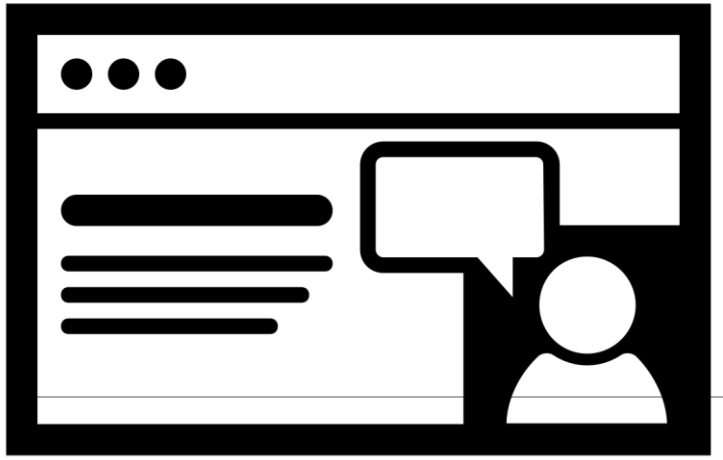
**AWS Cloud HSM** – Secure your cryptographic keys using Hardware Security Modules

- Recommendation is to use two HSM’s configured in a highly available configuration

# Versions and Lifecycle Management

- Multiple copies of each object in the bucket
  - Preserve, retrieve, and restore every version of every object
  - Enabled at the bucket level
  - Can be suspended but not disabled
- Lifecycle Management
  - Example: S3 to Glacier then delete





# Exercise: Lifecycle Management

How would a lifecycle rule help manage office records moved to S3 cloud storage?

# S3 Notifications

S3 server-access logs track requests to S3 bucket

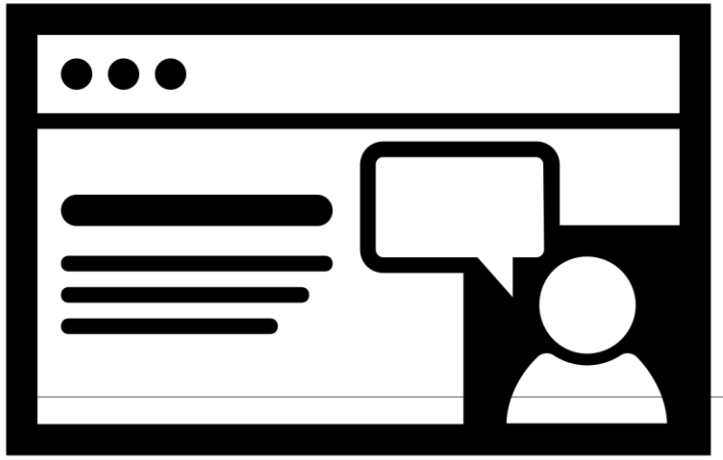
- Account name and IP address
- Bucket name
- Request time
- Action ( GET PUT LIST)
- Response or error code

Event Notifications

- Response to objects uploaded to S3
- Monitored at the bucket level

Object creation, removal triggers response

- Simple notification service, Simple queue service, transcoding, Lambda



# Exercise: S3 Administration

# S3 Glacier Storage

- Low cost archival storage
  - Data is stored in archives (Up to 40 TB)
  - Unlimited # of archives
  - Automatically encrypted

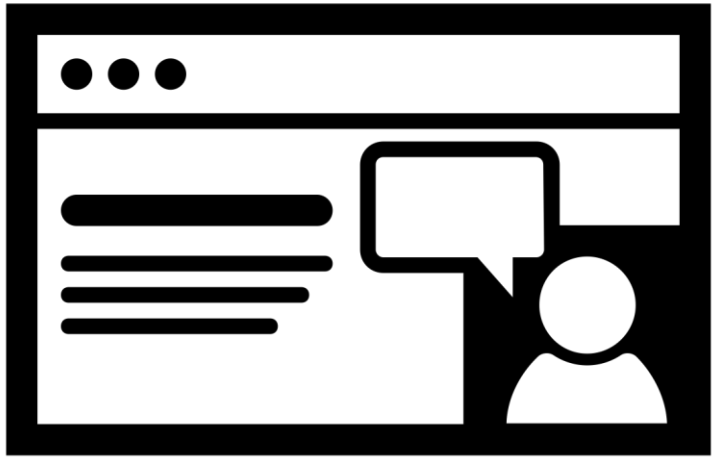
- S3 – 5 TB Object  size limit
- Glacier – 40 TB archives
- Glacier – Encrypted by default
- Glacier – Archive IDs
- S3 – Friendly names

What cloud storage option would you choose for archived records?  S3 storage or Glacier?

# S3 Glacier Vaults

Vault

- Archives are held in containers called vaults
- Each account can have up to 1,000 vaults
- Compliance controls per vault with a vault lock policy (WORM)
- Retrieval policy to control data access



# Exercise: S3 Glacier Vaults

# Core: What We Covered

- Fundamentals of AWS architecture, terminology and concepts
- Virtual Private Cloud (VPC) networking
- Amazon Elastic Compute Cloud (EC2) Instance deployment and configuration
- Storage solutions including Elastic Block Storage (EBS), and snapshot management
- Simple Storage Service (S3)
- S3 Glacier storage