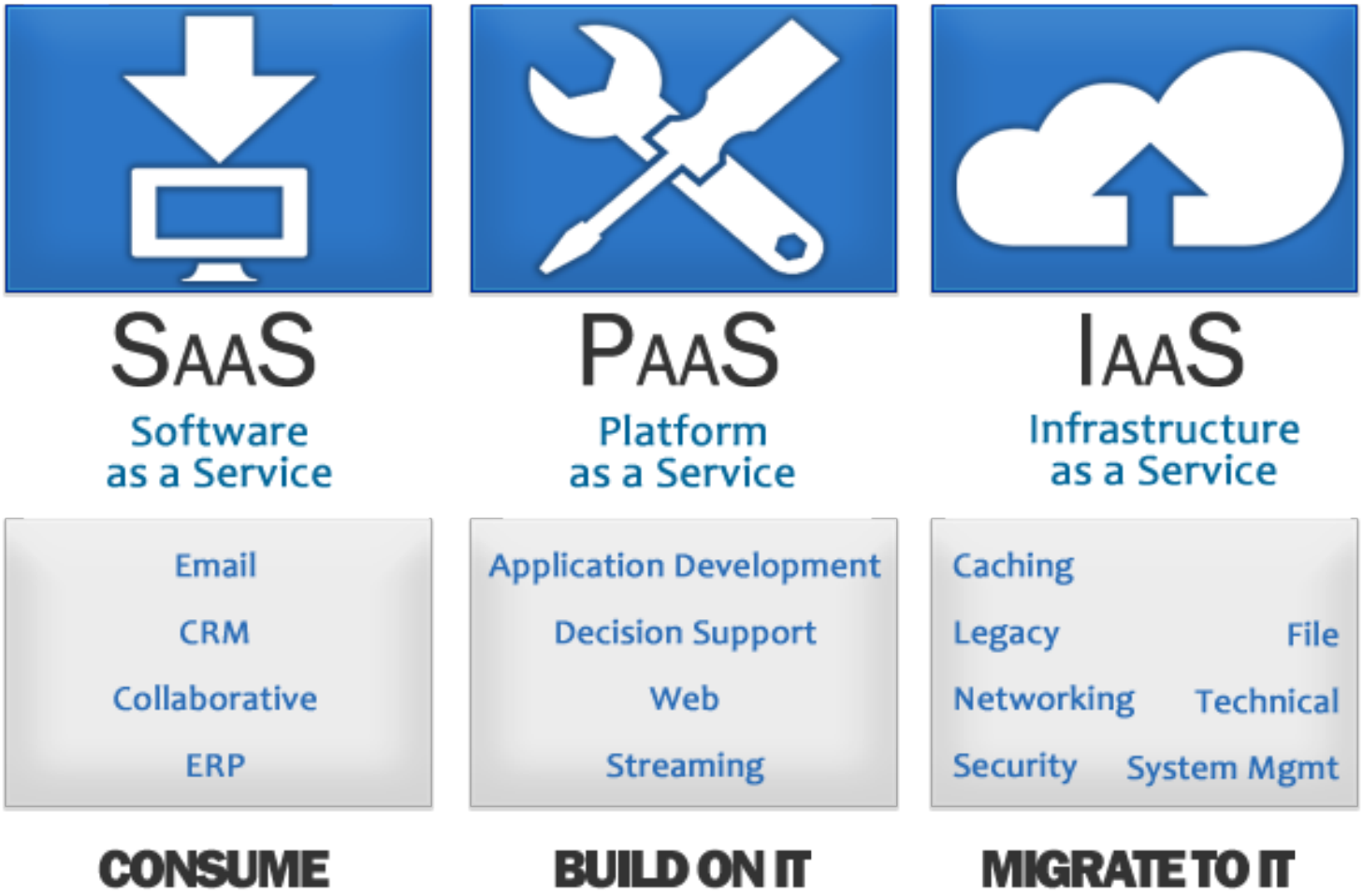
AWS Solution Architect Associate Exam

Topic	Answer
Exam Time:	80 Minutes
No. Questions:	60 Questions
Question Types:	Scenario and Multiple Choice
Passing Score:	~ 70%
Validity Period:	2 years
Renewal Exam:	1/2 price off

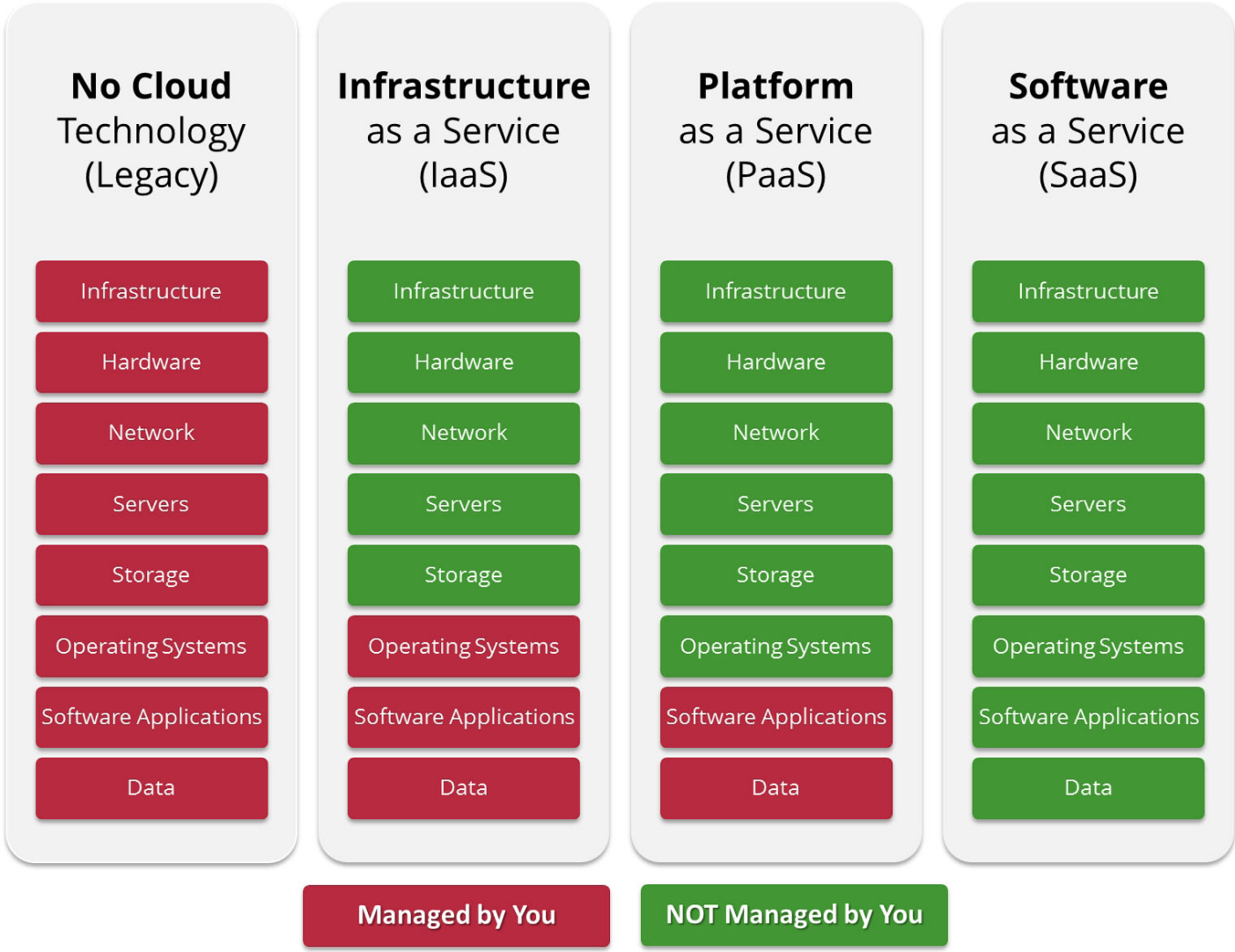
Cloud Models

How Cloud services were built/made



Cloud Models

On-premise v/s Managed service



Why Cloud services

On-demand self-service

Broad network access

Resource pooling

Pay as you use
Access via internet
Access from anywhere

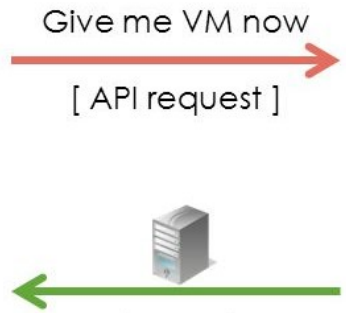
Rapid elasticity

Measured service

ON-DEMAND SELF-SERVICE

On demand service with few clicks/Api calls from anywhere via internet

Engineer



AWS Global Infrastructure

- Regions v/s Availability Zones:
  A Region is geographical area which consists of at least 2 Availability Zone's or AZ's. An AZ is simply a data centre.
- 16 Regions with 42 Availability Zones
- Projected to spin up 4 additional regions, with 9 additional AZ's over the next year
- Edge locations are CDN endpoints for CloudFront. Currently there are over 50 edge locations. Not the same as an AZ

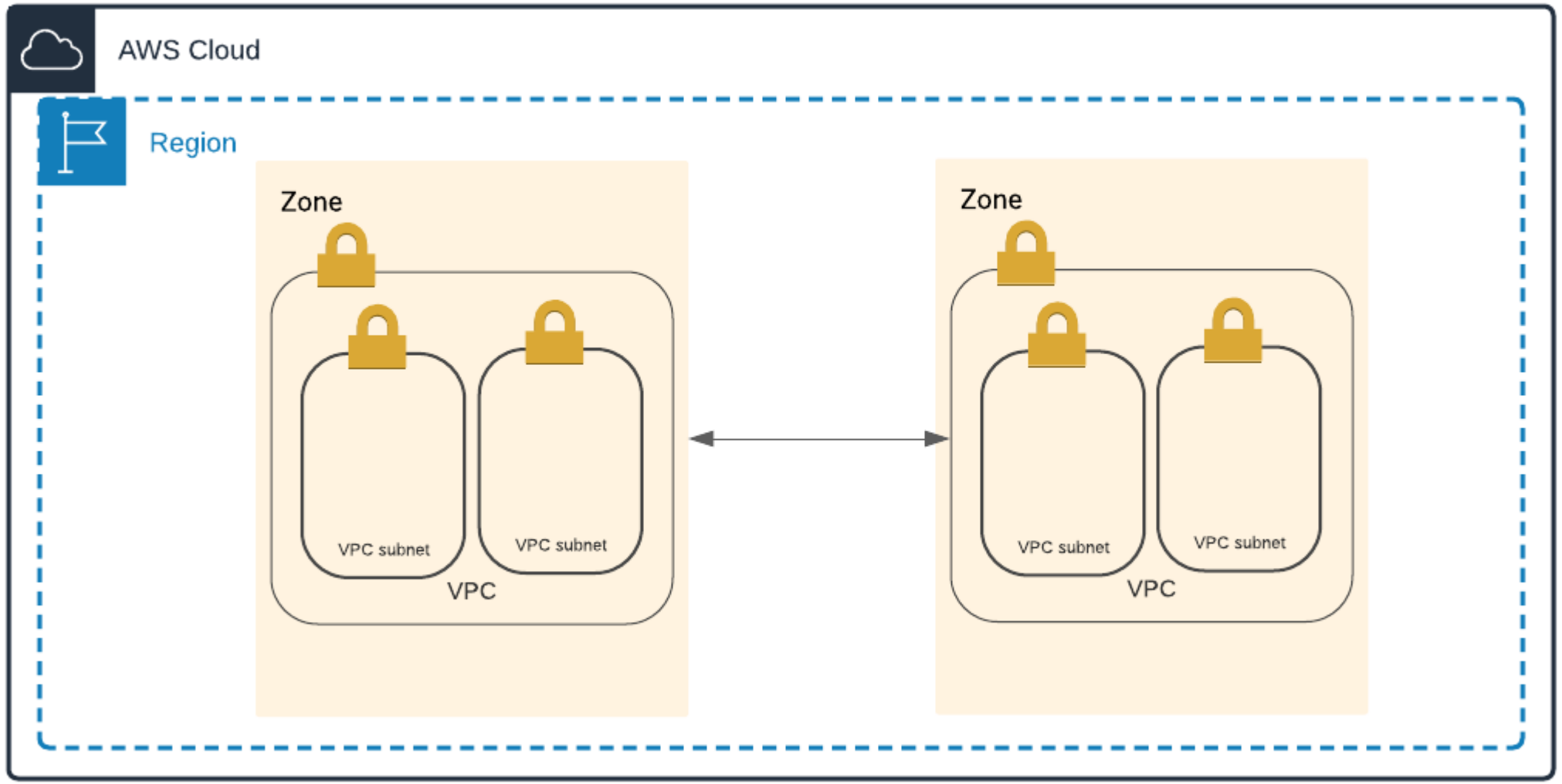
AWS IP address info

- Query meta-data:
  - curl http://169.254.169.254/latest/meta-data/
  - get http://169.254.169.254/latest/meta-data/
- AWS uses the first 4, and last IP addresses of a subnet:
  - x.x.x.0 - Network Address
  - x.x.x.1 - Gateway Address
  - x.x.x.2 - DNS Address
  - x.x.x.3 - Future Allocation Address
  - x.x.x.255 - Broadcast Address

AWS Resources

EC2, AMI & Instance type	Elastic Load balancer	Auto Scaling Group	IAM (Identity Access Management) AWS CLI
EBS, S3, S3 Glacier, EFS	AWS Networking	VPC, Security Group, VPC Peering	Cloud watch
Cloud Trail	SNS	Route53, CloudFront	Database (RDS, DynamoDB and Aurora)

AWS Regions and Zones



EC2 (Elastic Cloud Computing)

- Elastic Compute Cloud - Backbone of AWS, provides re-sizable compute capacity in the cloud
- Components of EC2
  - AMI
  - Instance type
  - VPC & userdata
  - EBS (Storage)
  - Security Group
  - Keypair

EC2 cont.

- AMI's are regional. You can only launch an AMI from the region in which it was stored
- AMI’s are template of configuration (OS, Apps etc.)
- You can copy AMI's to other regions using the console, CLI or Amazon EC2 API
- Hardware Virtual Machines (HVM) AMI's Available
- Paravirtual (PV) AMI's Available
- You can select an AMI based on:
  - Region
  - OS
  - Architecture (32 vs. 64 bit)
  - Launch Permissions
  - Storage for the root device (Instance Store Vs. EBS)

EC2 cont.

- Instance Type: Template for H/W configuration
  - D - Density
  - I - IOPS
  - R - RAM
  - T - Cheap General Purpos
  - M - Main General Purpose
  - C - Compute
  - G - Graphics

Family	Generation	Smallest	Largest
General Purpose (GP)	t2	t2.micro	t2.2xlarge
	m4	m4.large	m4.16xlarge
	m3	m3.medium	m3.2xlarge
Compute Optimized	c4	c4.large	c4.8xlarge
	c3	c3.large	c3.8xlarge
Memory Optimized	r3	r3.large	r3.8xlarge
	r4	r4.large	r4.16xlarge
	x1	x1.16xlarge	x1.32xlarge
Storage Optimized	i2	i2.xlarge	i2.8xlarge
	d2	d2.2xlarge	d2.8xlarge
Accelerated Computing	g2	g2.2xlarge	g2.8xlarge
	p2	p2.xlarge	P2.16xlarge

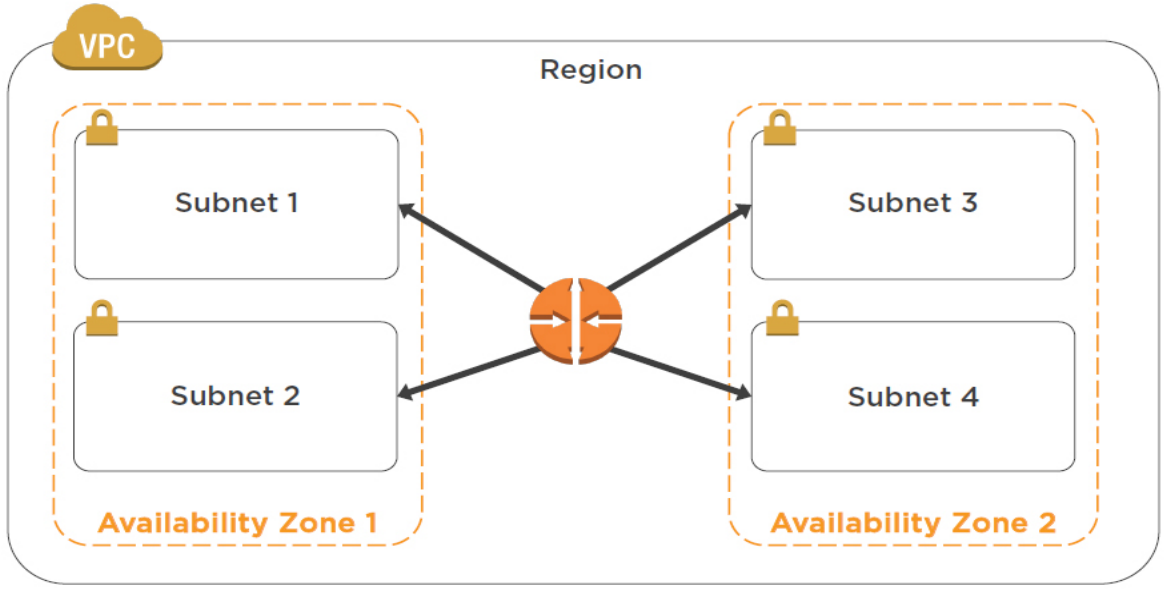
EC2 cont.

- EC2 Pricing Models & Instances:
    - On Demand:
        - Pay fixed rate by the hour with no commitment
        - Users that want the low cost and flexibility of EC2
        - Apps being developed or tested on EC2 for the first time
    - Reserved:
        - Provide capacity reservation and offer significant discount on the hourly charge for an instance (1-3 year terms)
        - Applications have steady state, or predictable usage
        - Apps that require reserved capacity
        - Users able to make upfront payments to reduce their total computing costs even further.
    - Spot:
        - Bid whatever price you want for instance capacity by the hour
        - When your bid price is greater than or equal to the spot price, your instance will boot
        - When the spot price is greater than your bid price, your instance will terminate with an hours notice.
        - If the spot instance is terminated by Amazon EC2, you will not be changed for a partial hour of usage
        - If you terminate the instance yourself you WILL be charged for any partial hours of usage.

.

EC2 cont.

- VPC(virtual private cloud): Select default VPC and respective zones (subnet)
- EBS (Elastic Block Storage): is persistent storage that can be used to procure storage to EC2 instances
- Security Groups: Act like virtual firewalls for the associated EC2 instance
- Keypair (.pem) is generated at the time of spinning ec2 instance.
  - Privacy Enhanced Mail (pem)
  - PuTTY Private Key Header (pp



Elastic block storage

- Elastic Block Storage is persistent storage that can be used to procure storage to EC2 instances. EBS are specific to ZONE
- You can NOT mount 1 EBS volume to multiple EC2 instances instead you must use EFS
- Default action for EBS volumes is for the root EBS volume to be deleted when the instance is terminated
- By default, ROOT volumes will be deleted on termination

Elastic block storage

- 3 Types of available EBS volumes can be provisioned and attached to an EC2 instance:General Purpose SSD (GP2):
  - General Purpose up to 10K IOPS
  - 99.999% availability
  - Ratio of 3 IOPS per GB with up to 10K IOPS and ability to burst
  - Up to 3K IOPS for short periods for volumes under 1GB
- Provisioned IOPS SSD (I01)
  - Designed for I/O intensive applications such as large relational or No-SQL DBs.
  - Use if need more than 10K IOPS
- Magnetic (Standard)
  - Lowest cost per GB
  - Ideal for workloads where data is accessed infrequently and apps where the lowest cost storage is important.
  - Ideal for fileservers

IAM (Identity Access Management)

- IAM is Global scope role
- Allows for centralized control and shared access to your AWS Account and/or AWS services
- By default when you create a user, they have NO permissions to do anything
- Root account has full admin access upon account creation
- Granular permission sets for AWS resources
- Multi-factor authentication support (Eg: Google Authenticator)

IAM (Identity Access Management)

- Access can be applied to:
  - Users - End users (people)
  - Groups - Collection of users under one set of permissions
  - Roles - Assigned to AWS resources, specifying what the resource (such as EC2) is allowed to access on another resource (S3)
  - Policies - Document that defines one or more permissions
    - Policies can be applied to users, groups and roles
    - You can assign up to 10 policies to a single group

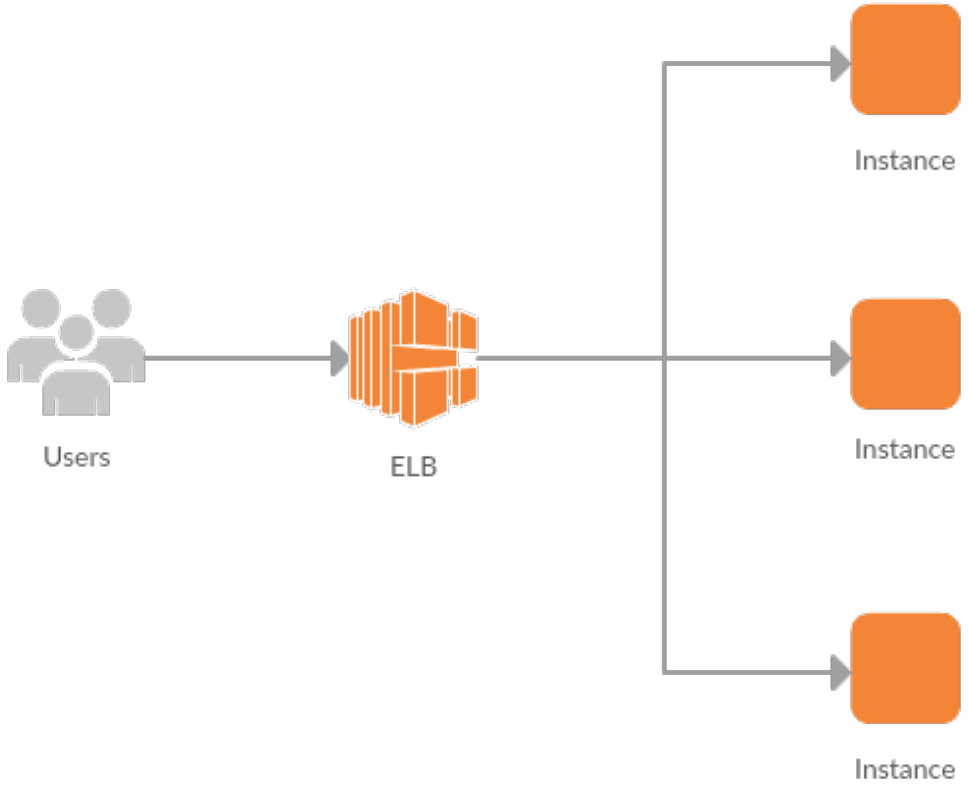
IAM (Identity Access Management)

- Two way to Access AWS resources:
  - Users log into the console, the user must have a username and password with account link
  - Access Keys/ Secret Access Keys to allow IAM users (or service accounts) to be used with AWS CLI or API calls
    - Access Keys can be retired, and new ones can be created in the event that secret access keys are lost
    - Access Key ID is equivalent to a user-name, Secret Access Key is equivalent to a password
- Policy:
  - AWS Managed
  - Customer Managed

Resource or Operation	Default Limit
Groups per account:	100
Instance profiles:	100
Roles:	250
Server Certificates:	20
Users:	5000
Number of policies allowed to attach to a single group:	10

ELB (Elastic Load Balancer)

- Elastic Load Balancer automatically distributes incoming application traffic across multiple applications, microservices, and containers hosted on Amazon EC2 instances.
- There are three types of Elastic Load Balancers available
  - Classic Load Balancer(CLB)
  - Application Load Balancer(ALB)
  - Network Load Balancer(NLB)



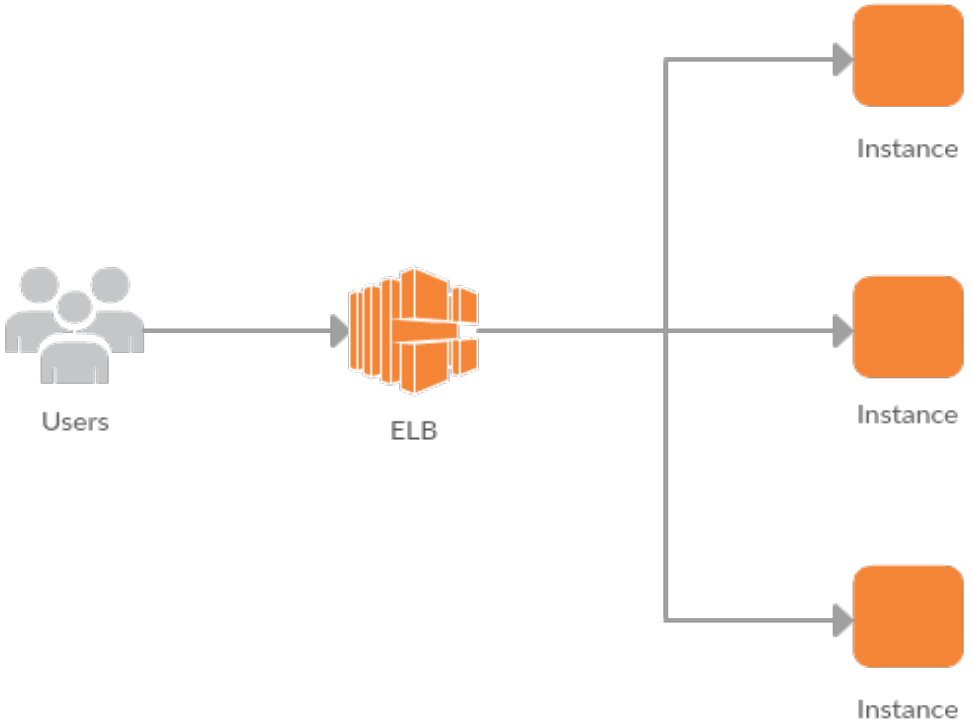
ELB (Elastic Load Balancer)

- Algorithms Used For Load Balancing
  - DNS Round Robin
    - Round Robin will first traverse the list of server in the group and then forward the client request to each server one by one and once it reaches the end of the list, it will again start from the top of the list
  - Least Connection
    - When a load balancer is configured to use the least connection algorithm, it selects a server with the least number of connections to serve the request
  - Source
    - This algorithm basically selects a server based on a hash of source IP requests e.g. Client's IP address

- Need For Load Balancing (forwarding rules)
  - HTTP
  - HTTPS
  - UDP
  - TCP

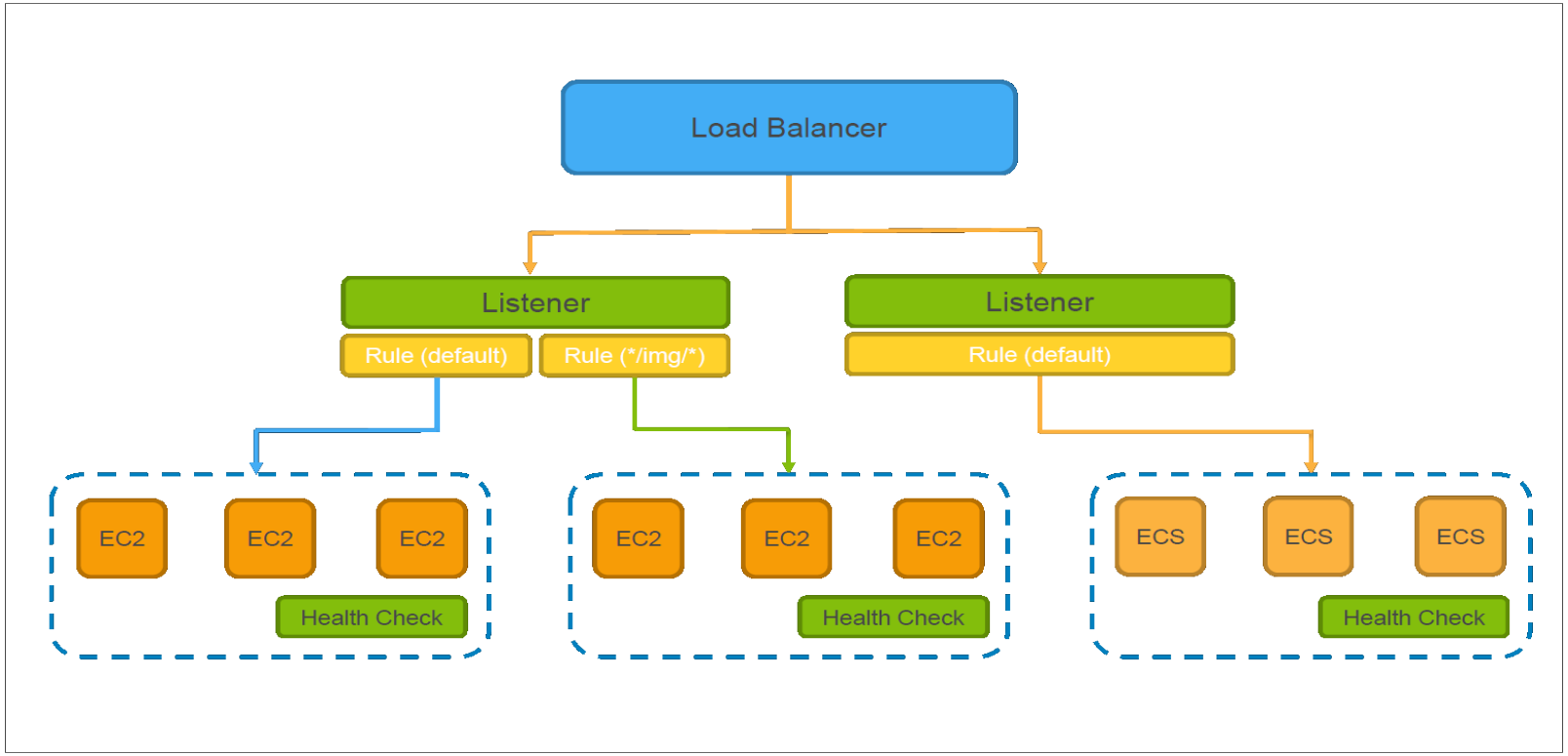
ELB (Elastic Load Balancer)

- Classic Load Balancer (CLB)
  - Basic form of the load balancer, that was used initially for classic EC2 instances
  - Operated on connection level as well as request level
  - It does not support features like host-based routing or path-based routing
  - Once the load balancer is configured, it balances load across the servers regardless of what is present



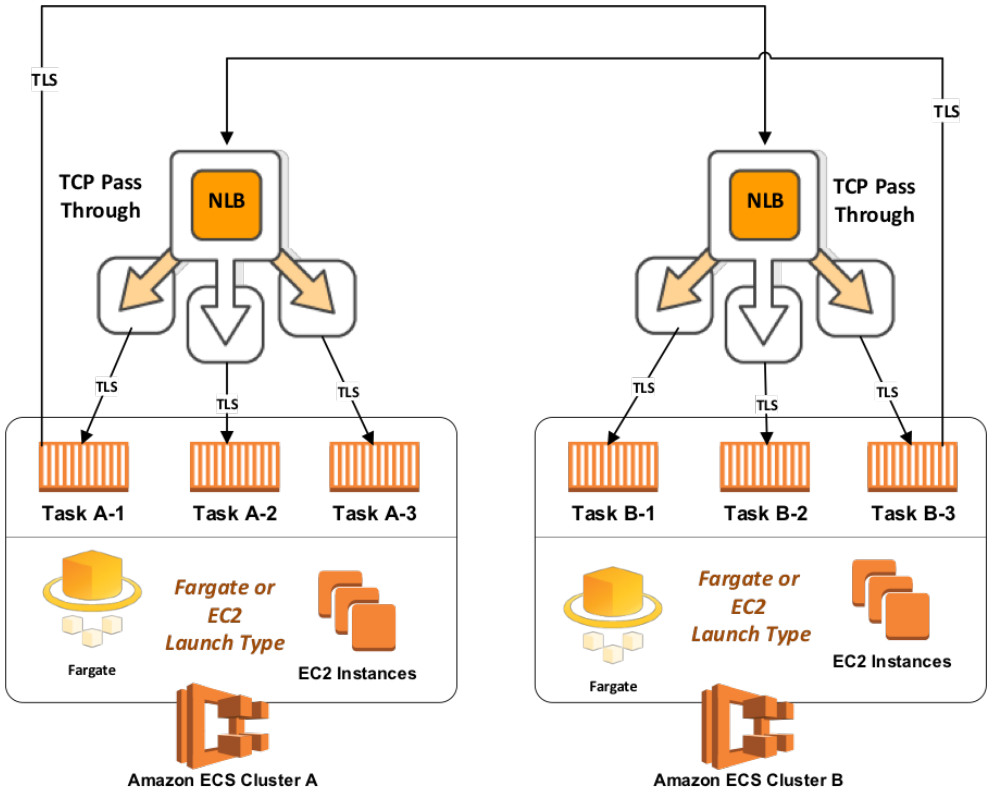
ELB (Elastic Load Balancer)

- Application Load Balancer (ALB)
  - Used when there are HTTP and HTTPS traffic routing
  - This load balancer works at the Application layer of the OSI Model
  - It provides advanced routing features such as host-based and path-based routing.
  - It also works well with containers and microservices



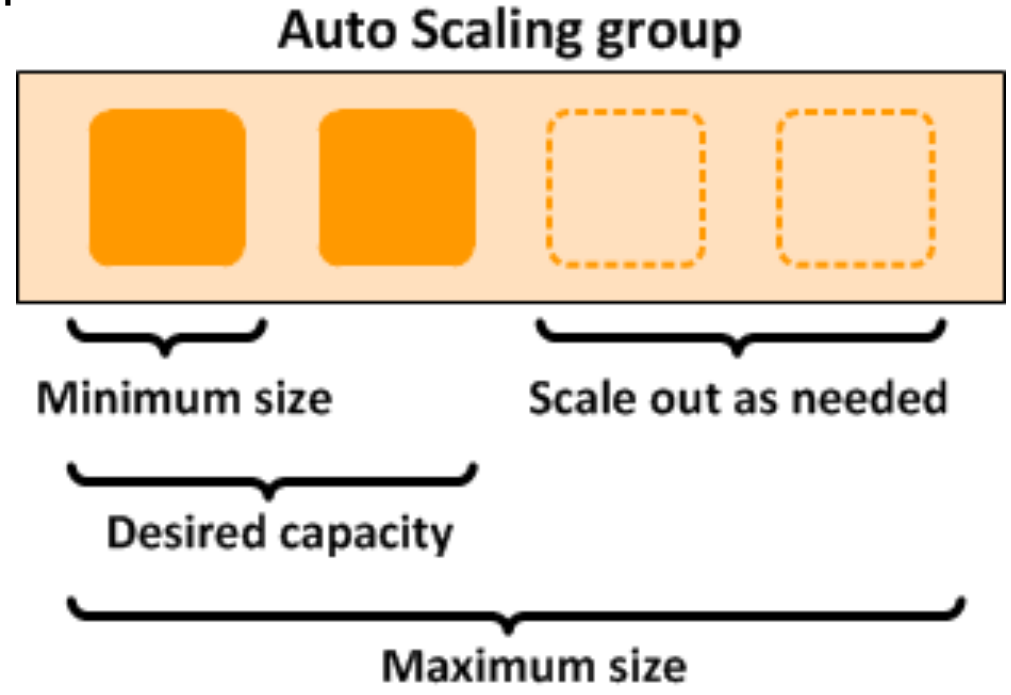
NLB (Network Load Balancer)

- Network Load Balancer (NLB)
  - Network Load Balancer works at layer 4 of the OSI Model(connection level)
  - Mainly used for load balancing TCP traffic
  - This Load Balancer can handle millions of traffic and is best suited for maintaining low latencies



ASG (Auto Scaling Group)

- Amazon EC2 Auto Scaling helps you ensure that you have the correct number of Amazon EC2 instances available to handle the load for your application.
- Automatic scaling for AWS resources to quickly have resource allocated for customer traffic
- Optimize for availability and cost



ASG (Auto Scaling Group)

- Auto Scaling Components

Groups

Logical unit for the purposes of scaling and management
Specify its minimum, maximum, and, desired number of EC2 instances

Launch Configuration

Launch configuration as a configuration template for its EC2 instances
Specify AMI ID, instance type, key pair, security groups, and block device mapping for your instances

Scaling Options

Scale IN and OUT from the capacity. Uses Dynamic and Predictive scaling
Dynamic Scaling to add and remove capacity for resources to maintain resource utilization at the specified target value

Predictive Scaling to forecast your future load demands by analyzing your historical records for a metric.

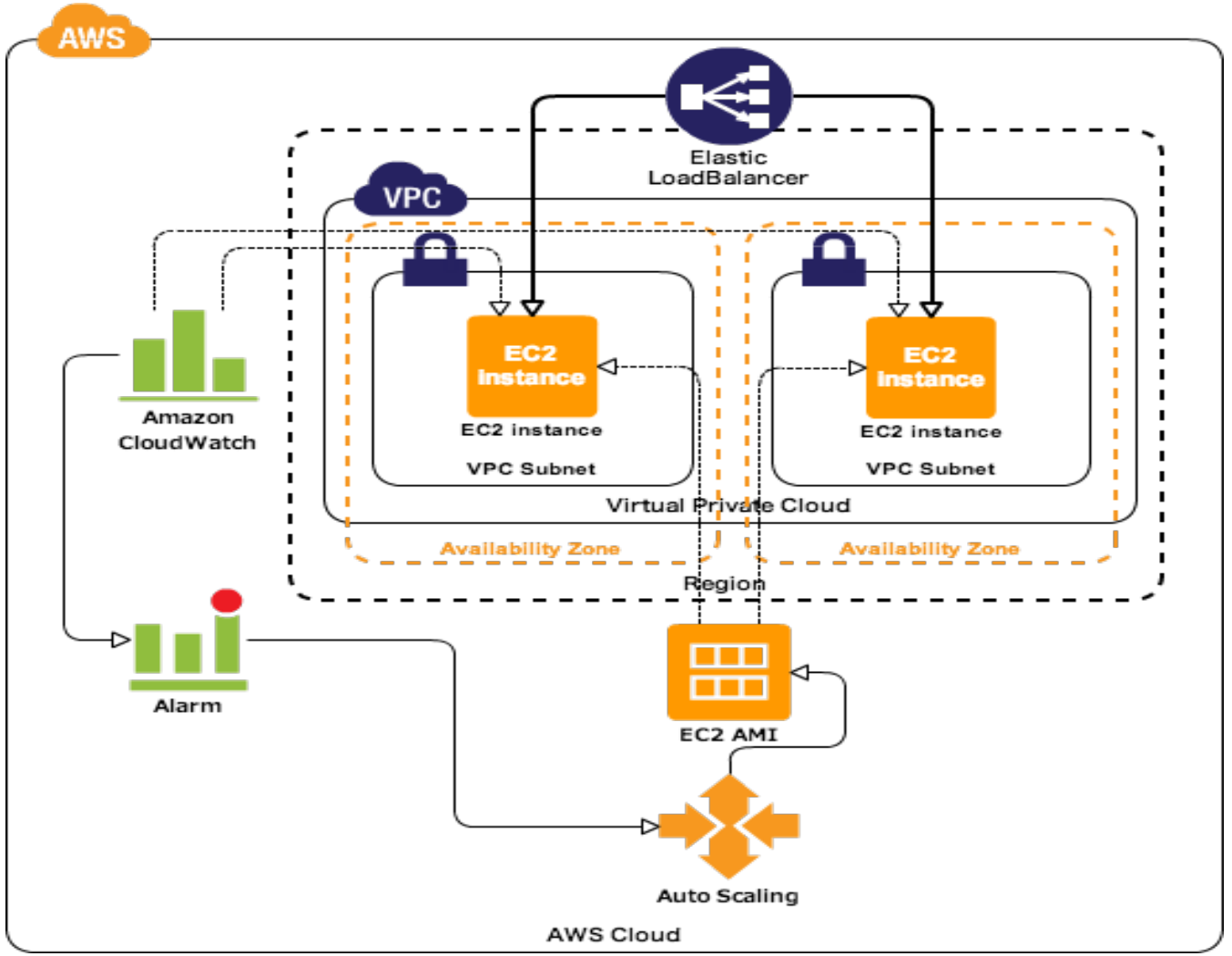
ASG (Auto Scaling Group)

- AWS resources to configure Auto Scaling Group

  - AMI (custom Image)
  - Launch Configuration (Template for EC2 instances)
  - Auto Scaling Group policy (Group Capacity)
  - Scaling policies (IN and OUT)
  - Cloud watch (Metrics)
  - Simple Notification Service (Alerting)

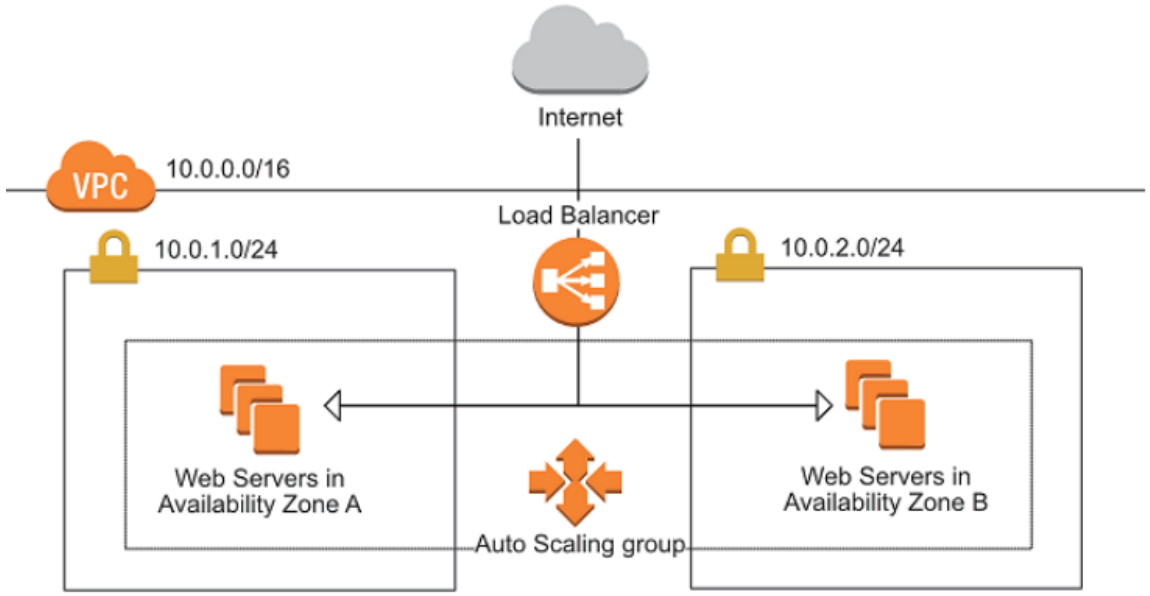
ASG (Auto Scaling Group)

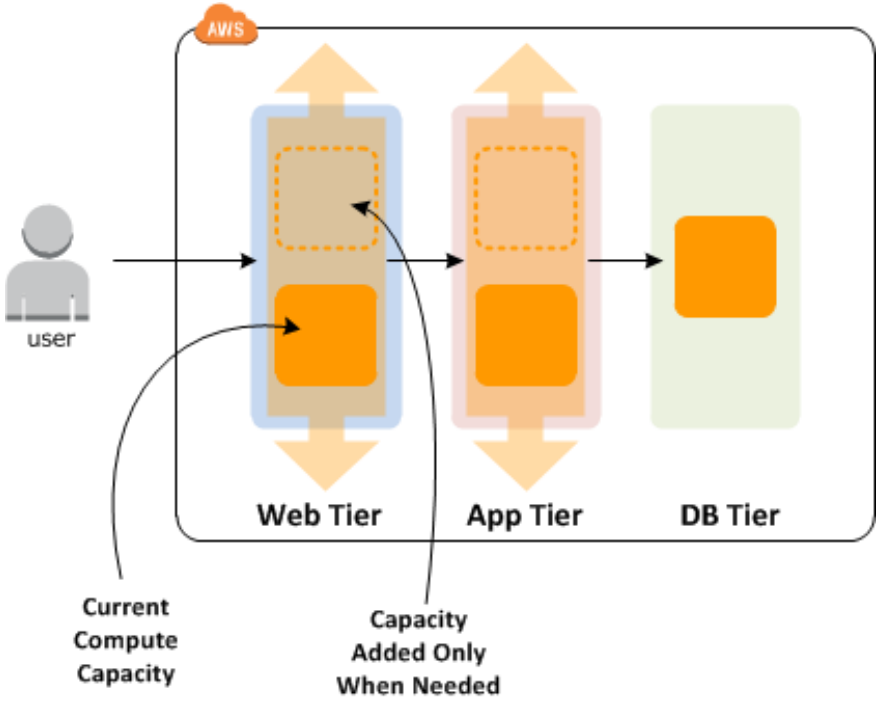
- Workflow Auto Scaling Group



ASG (Auto Scaling Group)

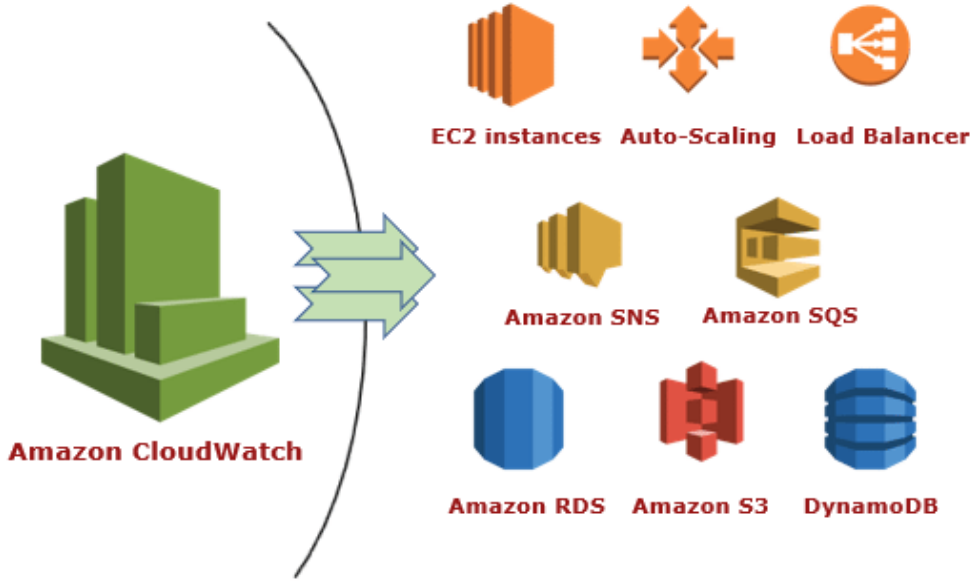
- ASG three tier with ELB





Cloud Watch

- Monitoring is crucial to make sure that all the services are running smoothly and efficiently
- Provides real-time monitoring of AWS resources and customer applications running on Amazon infrastructure



Resources Monitored By CloudWatch

Cloud Watch

- Enables robust monitoring of resources like :
  - Virtual instances hosted in Amazon EC2
  - Databases located in Amazon RDS
  - Data stored in Amazon S3
  - Elastic Load Balancer
  - Auto-Scaling Groups
  - Other resources
- Monitors, stores and provides access to system and application log files
- Provides a catalog of standard reports that you can use to analyze trends and monitor system performance
- Provides various alert capabilities, including rules and triggers high resolutions alarms and sends notifications
- Collects and provides a real-time presentation of operational data in form of key metrics like CPU utilization, disk storage etc.

Cloud Watch

- Amazon CloudWatch operates there are certain primary area

  - **Metrics**
    - Metrics represents a time-ordered set of data points that are published to CloudWatch
    - Metrics are uniquely defined by a name, a namespace, and zero or more dimensions

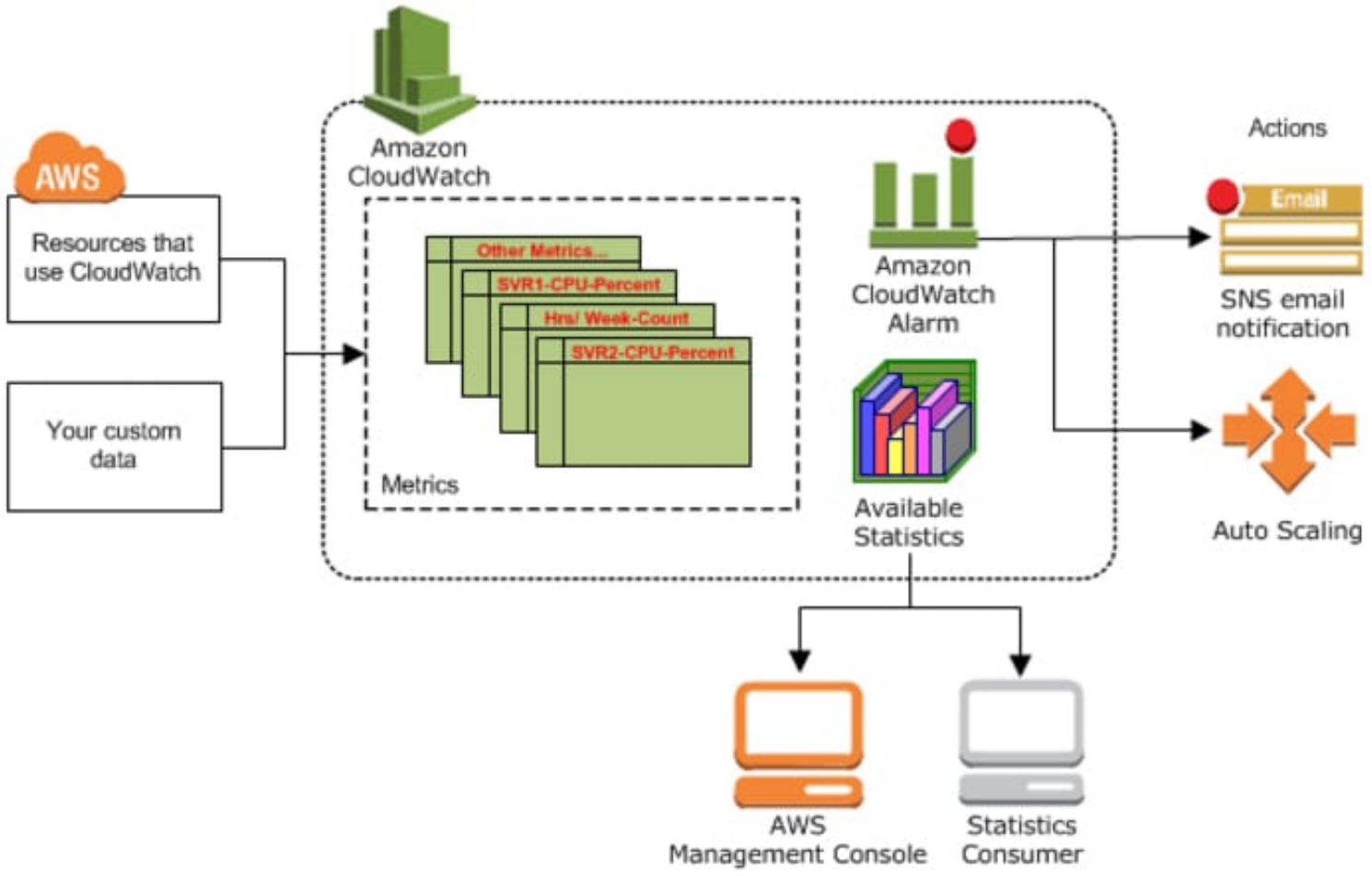
  - **Dimensions**
    - A dimension is a name/value pair that uniquely identifies a metric
    - Dimensions can be considered as categories of characteristics that describe a metric

  - **Statistics**
    - Statistics are metric data aggregations over specified periods of time
    - Few available statistics are maximum, minimum, sum, average and sample count

  - **Alarm**
    - An alarm can be used to automatically initiate actions on your behalf
    - The action is a simply a notification that is sent to Amazon SNS topic

Cloud Watch

- Amazon CloudWatch

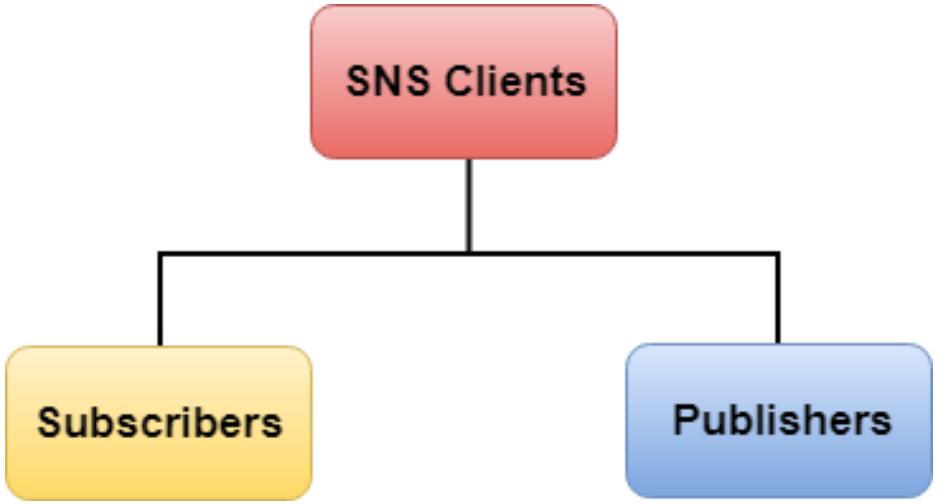


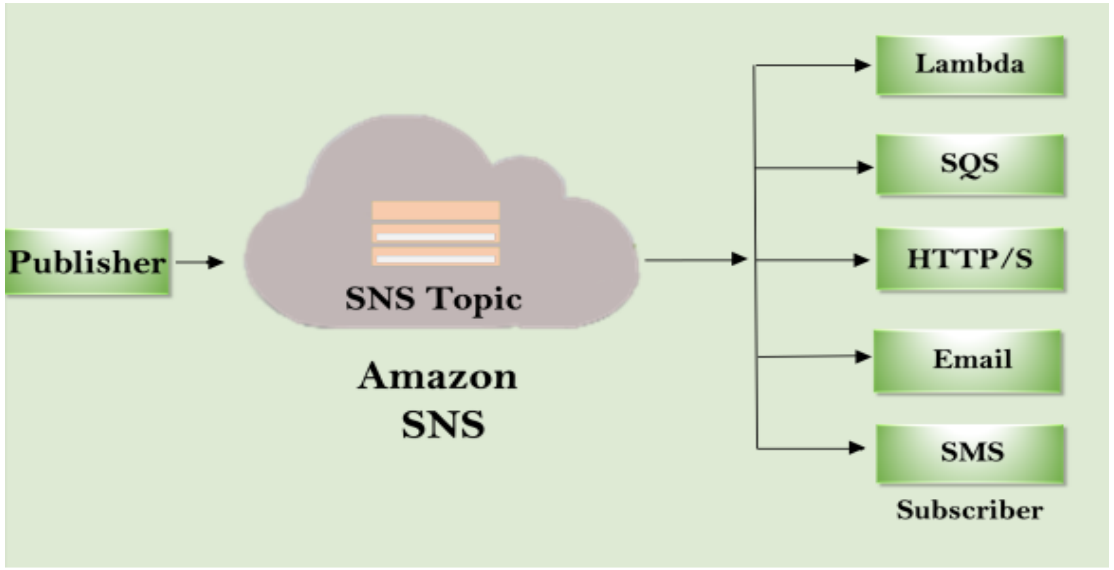
SNS (Simple Notification Service)

- SNS stands for Simple Notification Service
- It is a web service which makes it easy to set up, operate, and send a notification from the cloud
- It is a way of sending messages. When you are using AutoScaling, it triggers an SNS service which will email you that "your EC2 instance is growing”
- Amazon SNS allows you to group multiple recipients using topics
- Topic is a logical access point that sends the identical copies of the same message to the subscribe recipients.
- Amazon SNS supports multiple endpoint types.
- Once you publish the message to the topic, SNS delivers the formatted copies of your message to the subscribers.
- To prevent the loss of data, all messages published to SNS are stored redundantly across multiple availability zones.

SNS (Simple Notification Service)

- SNS subscribers and endpoints





Publishers
Publishers are also known as producers that produce and send the message to the SNS which is a logical access point

Subscribers
Subscribers such as web servers, email addresses receive the message or notification from the SNS

SNS (Simple Notification Service)

- SNS benefits

  - **Instantaneous delivery**
    SNS is based on push-based delivery, SNS is pushed once you publish the message in a topic and the message is delivered to multiple subscribers
  - **Flexible**
    SNS supports multiple endpoint types such as email, SMS, Lambda, Amazon SQS, HTTP, etc
  - **Inexpensive**
    SNS service is quite inexpensive as it is based on pay-as-you-go model
  - **Ease of use**
    SNS service is very simple to use as Web-based AWS Management Console offers the simplicity of the point-and-click interface.
  - **Simple Architecture**
    SNS is used to simplify the messaging architecture by offloading the message filtering logic from the subscribers and message routing logic from the publishers

Storage service (S3)

- S3: Simple Storage Service
  - Object based storage only for files, can not install OS or applications
  - Files can be between 1 byte and 5TB, and has no storage limit. Objects are stored flatly in buckets, Folders don't really exist,
  - S3 buckets must be globally unique and stores data in alphabetical order
  - S3 URL structures are region/amazon.aws.com/bucketname (https://s3-eu-west-1.amazonaws.com/myawsbucket)
  - Amazon guarantees 99.99% availability for the S3 platform
  - Amazon guarantees 99.999999999% durability for S3 information (11 x 9's)
  - Versioning is available but must be enabled. It is off by default
  - Offers encryption, and allows you to secure the data using ACLs
  - S3 charges for storage, requests, and data transfer
  - When you upload a file to S3, by default it is set private
  - You can transfer files up to 5GB using PUT requests
  - You can setup access control to control your buckets access by using bucket policies or ACLs
  - S3 buckets can be configured to create access logs which logs all requests to the S3 bucket

Storage service (S3)

- Versioning and Cross-Region Replication (CRR):
  - Supports versioning but once Versioning is turned on, it can not be turned off, it can only be suspended
  - With versioning enabled, if you delete a file, S3 creates a delete marker for that file, which tells the console to not display the file any longer
  - In order to restore a deleted file you simply delete the delete marker file, and the file will then be displayed again in the bucket
  - Versioning does NOT support de-duplication or any similar technology currently
  - Cross Region Replication (CRR) has to be enabled on both the source and destination buckets in the selected regions
  - Destination bucket must be created and again globally unique
  - CRR does NOT replicate existing objects, only future objects meaning that only objects stored post turning the feature on will be replicated
  - Versioning integrates with life-cycle management and also supports MFA delete capability

Storage service (S3)

- Life Cycle Management:
  - Rules can be set to move objects to either separate storage tiers or delete them all together
  - A rule can be applied to either the entire bucket or a single 'folder' in a bucket
  - Can be applied to current version and previous versions
  - Calculates based on UPLOAD date not Action data
  - Transition from STD to IA storage class requires MINIMUM of 30 days. You can not select or set any data range less than 30 days
- Static Web hosting
  - Supports Web hosting
  - Provides endpoint to access website
  - Bucket should be public in-order to host website

Storage service (EFS)

- Elastic File System (EFS):
  - File storage service for EC2 instances
  - EFS storage capacity is elastic, growing and shrinking automatically as you add and remove files
  - Think NFS, only without a set storage limit
  - Supports NFSv4, and you only pay for the storage you use, billing rate is 30 cents per GB
  - Can scale to exabytes and also support thousands of concurrent NFS connections
  - Data is stored across multiple AZ within a region
  - Block based storage.
  - Can be shared with multiple instances
  - Read after Write Consistency

Storage service (Glacier)

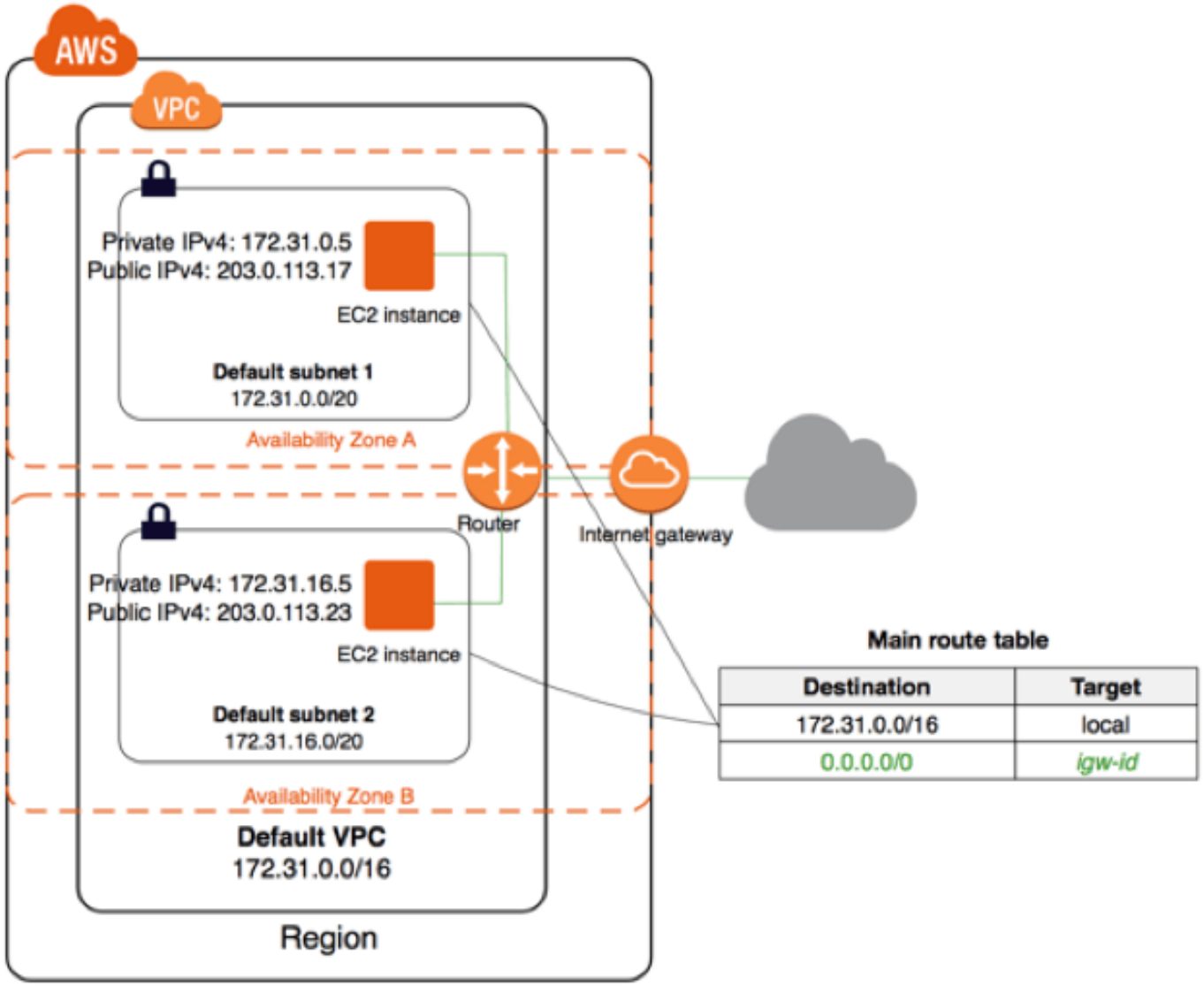
- S3 Glacier:
  - Data is stored in Amazon S3 Glacier in "archives"
  - A single archive can be as large as 40 terabytes
  - Amazon S3 Glacier uses "vaults" as containers to store archives
  - Amazon S3 Glacier provides three retrieval features
    - Expedited : retrievals are typically available within 1 – 5 minutes
    - Standard: archives typically become accessible within 3 – 5 hours
    - Bulk retrievals

VPC (Virtual Private Cloud)

- Amazon Virtual Private Cloud (Amazon VPC) provides a logically isolated area of the AWS cloud where you can launch AWS resources in a virtual network that you define
- You have complete control over your virtual networking environment, including a selection of your IP address range, the creation of subnets, and configuration of route tables and network gateways
- You can easily customize the network configuration for your Amazon Virtual Private Cloud. For example, you can create a public-facing subnet for web servers and private-facing subnet for backend system
- You can provide multiple layers of security, including security groups and network access control lists, to help control access to Amazon EC2 instances in each subnet

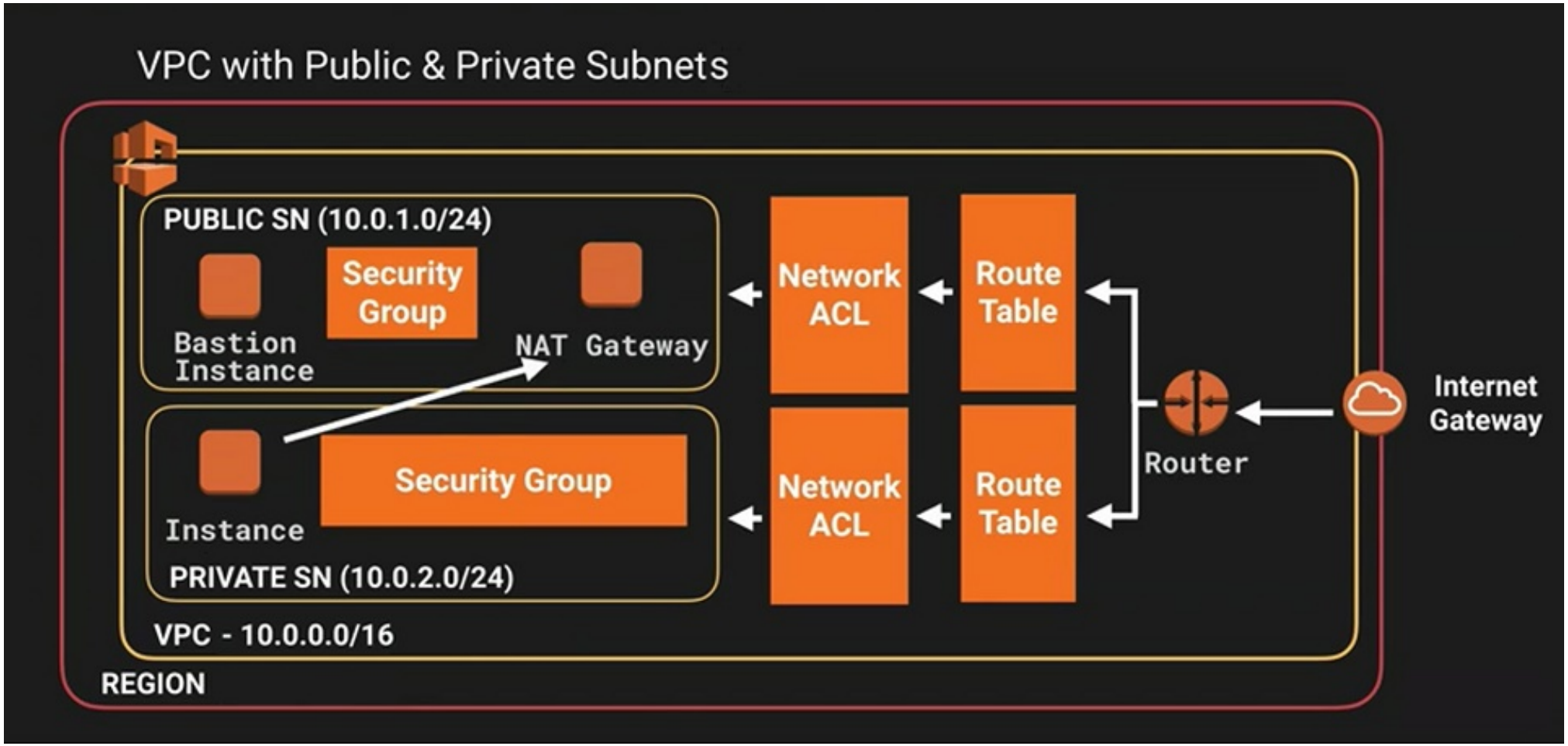
VPC (Virtual Private Cloud)

- AWS Default VPC



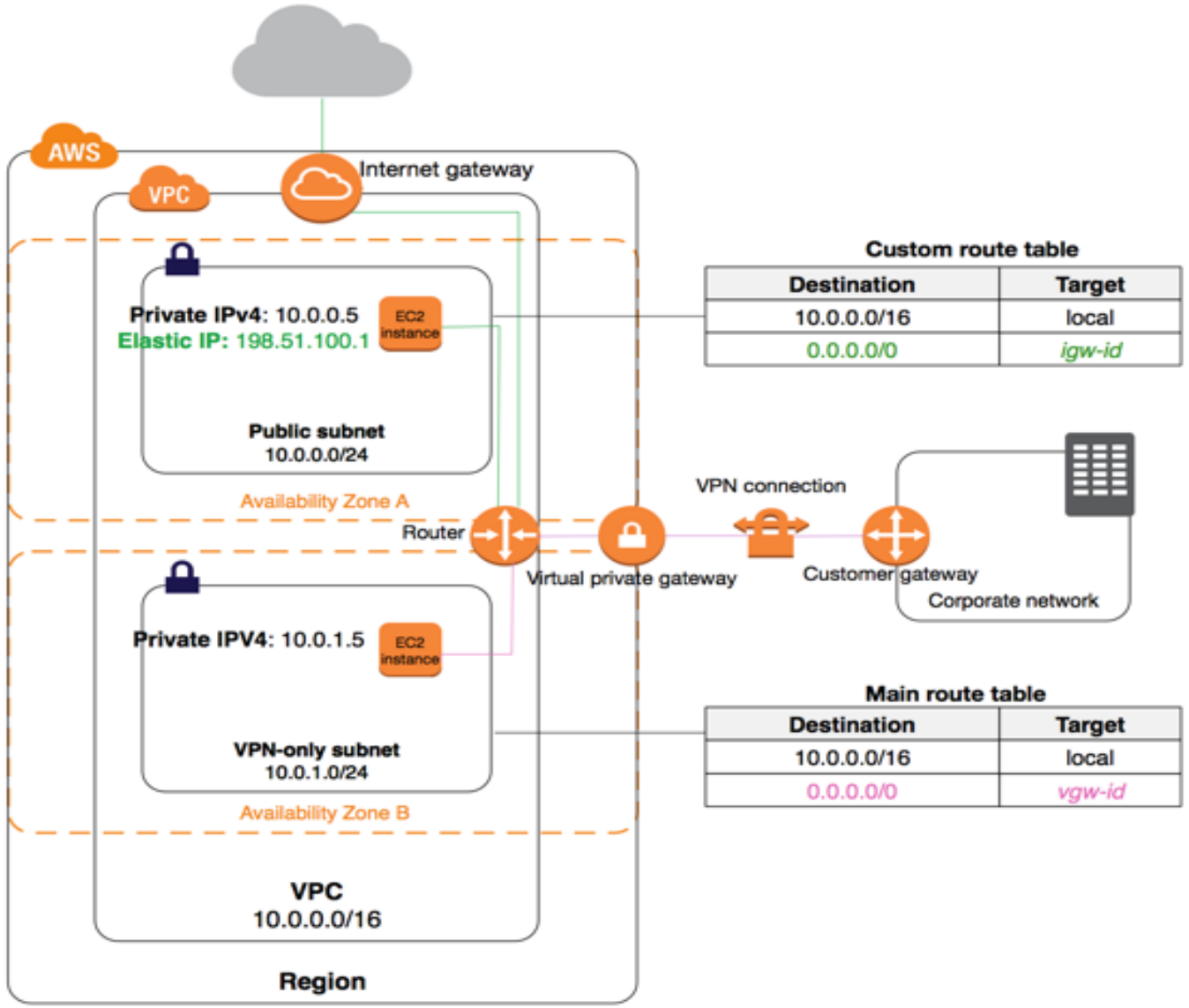
VPC (Virtual Private Cloud)

- Custom VPC architecture



VPC (Virtual Private Cloud)

- Custom VPC architecture



VPC (Virtual Private Cloud)

- AWS VPC Guidelines
  - Allowed up to 5 VPCs in each AWS region by default
  - All subnets in default VPC have an Internet gateway attached
  - Multiple IGW's can be created, but only a single IGW can be attached to a VPC.. No exceptions
  - Again, You can only have 1 Internet gateway per VPC
  - Each EC2 instance has both a public and private IP address
  - Subnets are always mapped to a single AZ
  - Subnets can not be mapped to multiple AZ's
  - /16 is the largest CIDR block available when provisioning an IP space for a VPC
  - /28 is the smallest CIDR block available when provisioning an IP space for a VPC
  - 169.254.169.253 - Amazon DNS

VPC (Virtual Private Cloud)

- AWS VPC Guidelines
  - Amazon uses 3 of the available IP addresses in a newly created subnet
    - x.x.x.0 - Always subnet network address and is never usable
    - x.x.x.1 - Reserved by AWS for the VPC router
    - x.x.x.2 - Reserved by AWS for subnet DNS
    - x.x.x.3 - Reserved by AWS for future use
    - x.x.x.255 - Always subnet broadcast address and is never usable
  - By default when you create a VPC, a default main routing table automatically gets created as well
  - Ranges are reserved for private subnet:
    - 10.0.0.0 - 10.255.255.255 (10/8 prefix)
    - 172.16.0.0 - 172.31.255.255 (172.16/12 prefix)
    - 192.168.0.0 - 192.168.255.255 (192.168/16 prefix)

VPC (Virtual Private Cloud)

- Route Tables
  - Your VPC has an implicit router.
  - Your VPC automatically comes with a main route table that you can modify.
  - You can create additional custom route tables for your VPC.
  - Each subnet must be associated with a route table, which controls the routing for the subnet
- Two types
  - Main Route table
  - Custom Route table

VPC (Virtual Private Cloud)

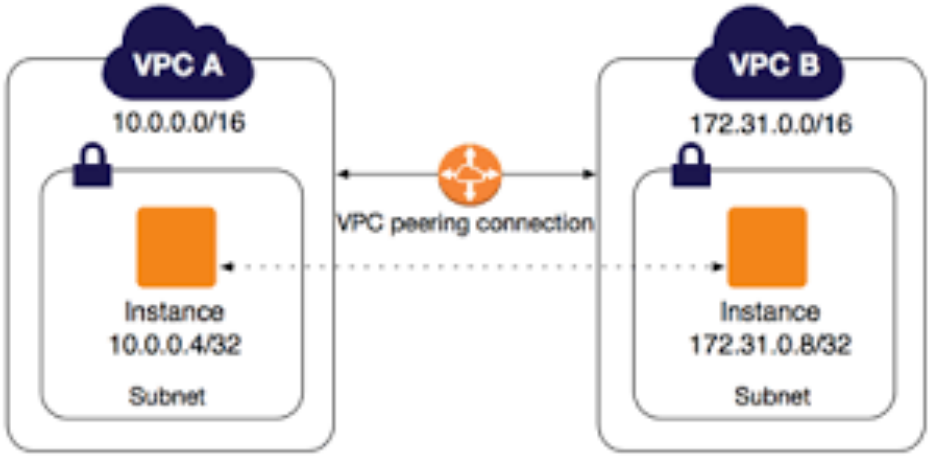
- IGW (Internet Gateway)
  - Internet Gateway is a very important component that allows your instance to connect to the internet
  - Used for two way communication
  - One VPC Once IGW
  - 5 VPC per region
- NGW (NAT Gateway)
  - NAT stands for Network Address Translation.
  - EC2 instance in a private subnet can access the internet via NGW
  - In real time, NAT Gateways are highly used than NAT instances as NAT instances are an individual EC2 instances, and NAT Gateways are highly available across multiple availability zones, and they are not on a single EC2 instance.

VPC (Virtual Private Cloud)

- Elastic IP address
  - An Elastic IP address is a static, public IPv4 address designed for dynamic cloud computing
  - Your Elastic IP addresses remain associated with your AWS account until you explicitly release them
  - Limited to five Elastic IP addresses; to help conserve them, you can use a NAT device
  - An Elastic IP address is accessed through the Internet gateway of a VPC

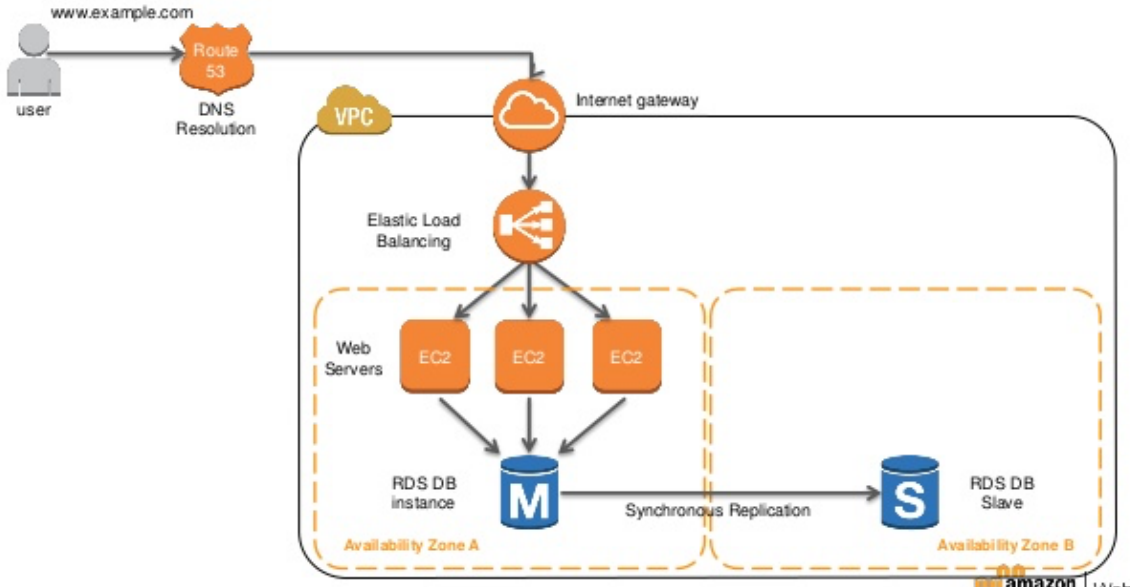
VPC (Virtual Private Cloud)

- VPC Peering:
  - Connection between two VPCs that enables you to route traffic between them using private IP addresses via a direct network route
  - You can create VPC peering connections between your own VPCs or with a VPC in another account within a SINGLE REGION
  - AWS uses existing infrastructure of a VPC to create a VPC peering connection. It is not a gateway nor a VPN, and does not rely on separate hardware
  - There is NO single point of failure for communication nor any bandwidth bottleneck
  - There is no transitive peering between VPC peers (Can't go through 1 VPC to get to another)
  - Be mindful of IPs in each VPC, if multiple VPCs have the same IP blocks, they will not be able to communicate
  - You can peer VPC's with other AWS accou



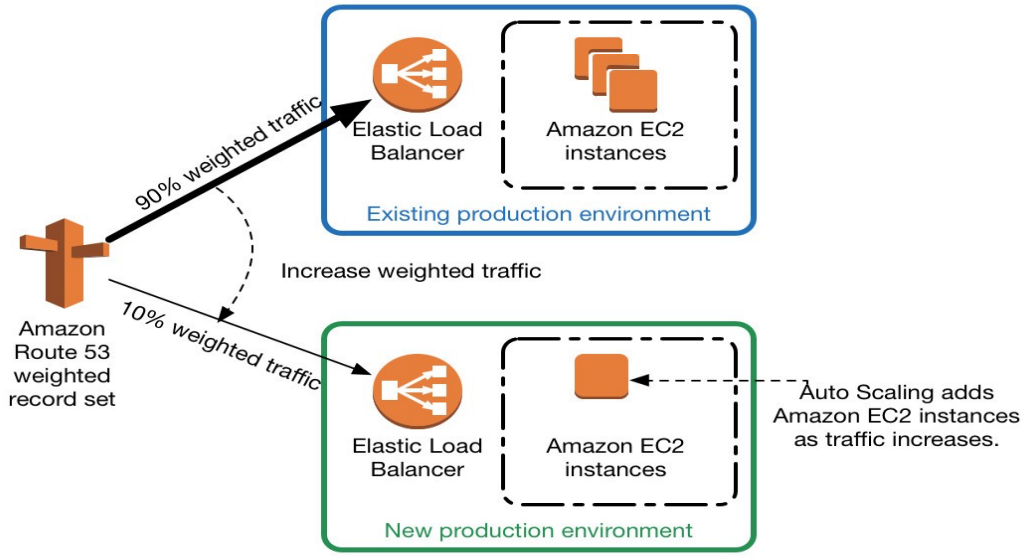
Route53

- Amazon Route 53 is a highly available and scalable cloud Domain Name System (DNS) web service
- ELBs do not have a pre-defined IPv4 address. You resolve them using a DNS name
- Aliases map AWS resources to zone records Alias records you are not charged for, CNAME records you are charged for
- Always chose an alias record, over a CNAME record, as alias records are free, and can be mapped to a domain apex record where CNAMES cannot
- Limit of 50 Domain Names can be managed in Route53. This limit can be raised by support
- Route 53 Routing Policies:
  - Simple
  - Weighted
  - Latency
  - Fail-over
  - Geo-Location



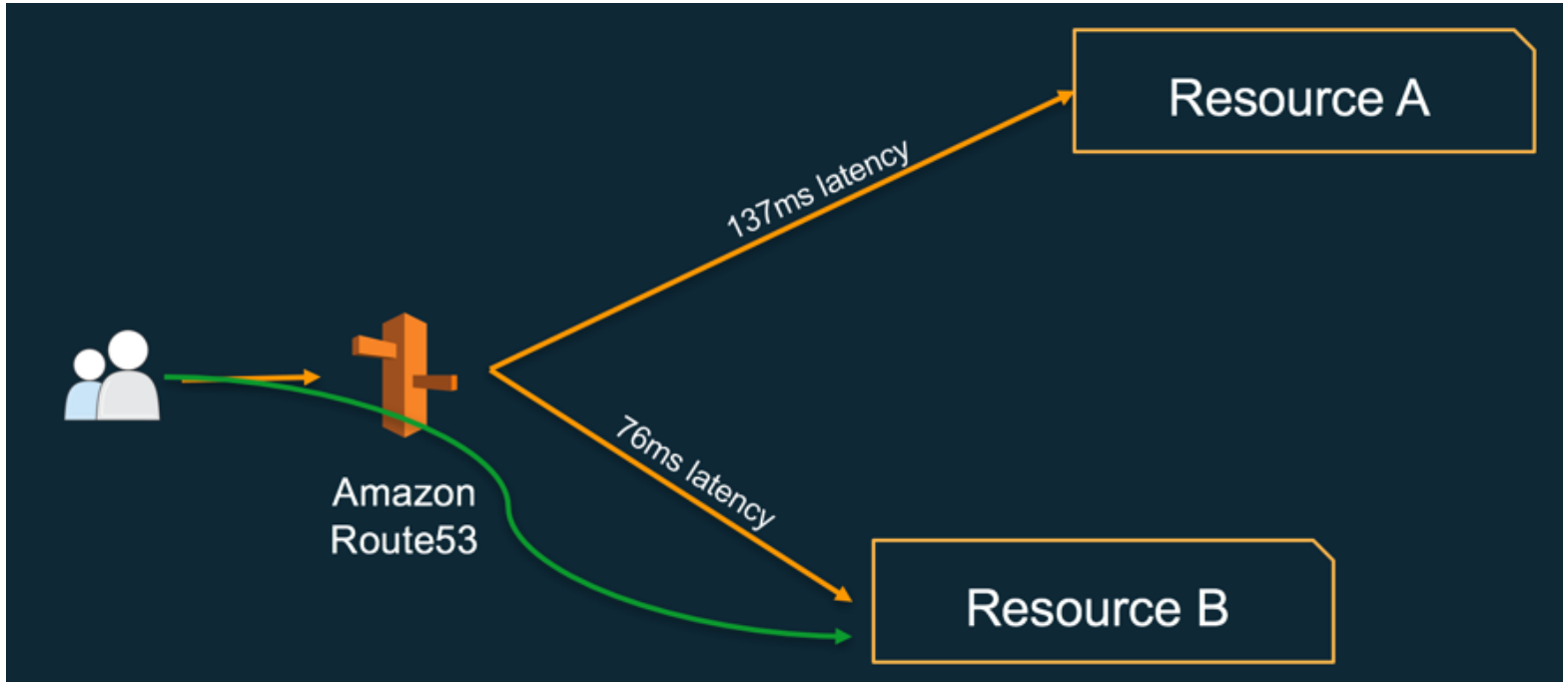
Route53

- Simple Routing Policy
  - Default routing policy when you create a new record set
  - Most common when you have a single resource that performs given function for your domain
  - Route53 will respond to DNS queries that are only in the record set.
  - No Intelligence is built into the response
- Weighted Routing Policy
  - Weighted Let you split traffic based on different weights defined
  - 1 AZ can be set to 90%, and another can be set to 10% for example



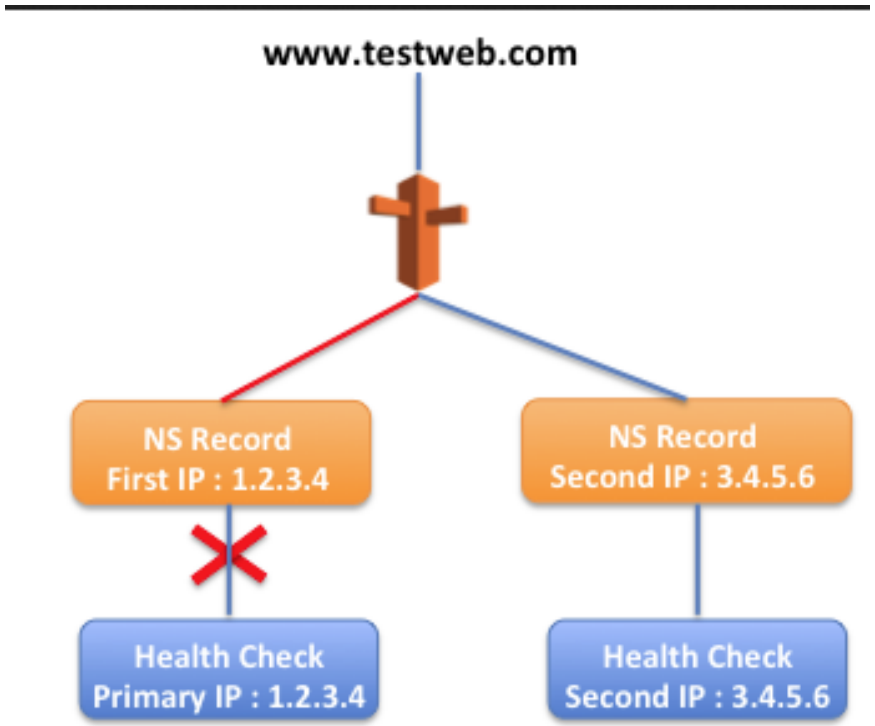
Route53

- Latency Routing Policy
  - Allows you to route your traffic based on the lowest network latency for your end user. (Which region will give them the fastest response time)
  - Create a latency resource record set in each region that hosts your website
  - When Route53 receives a query for your site, it selects the latency resource for the region that gives the user the lowest latency



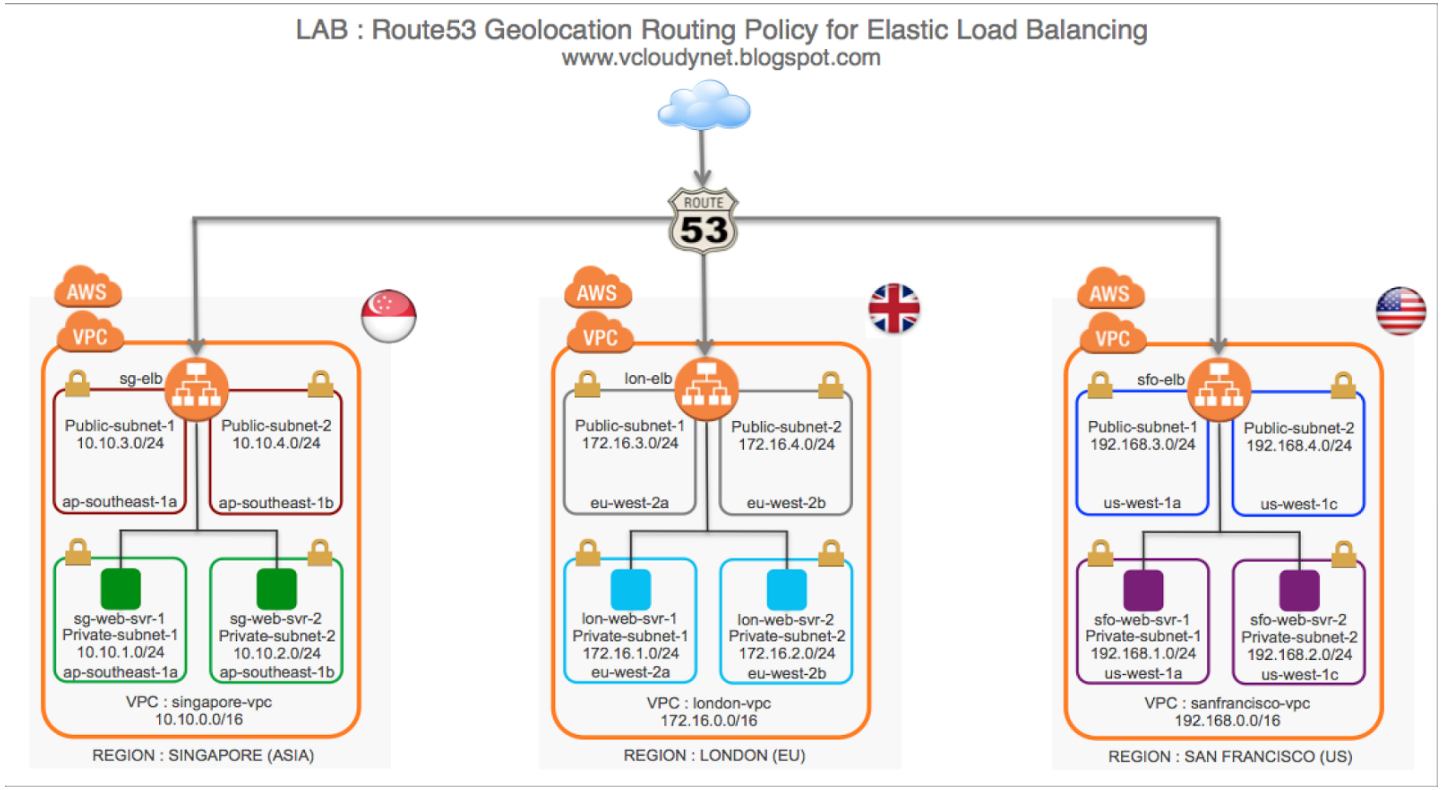
Route53

- Failover Routing Policy
  - Used when you want to create an active/passive set up
  - Route53 will monitor the health of your primary site using a health check
  - Health check monitors the health of your endpoints



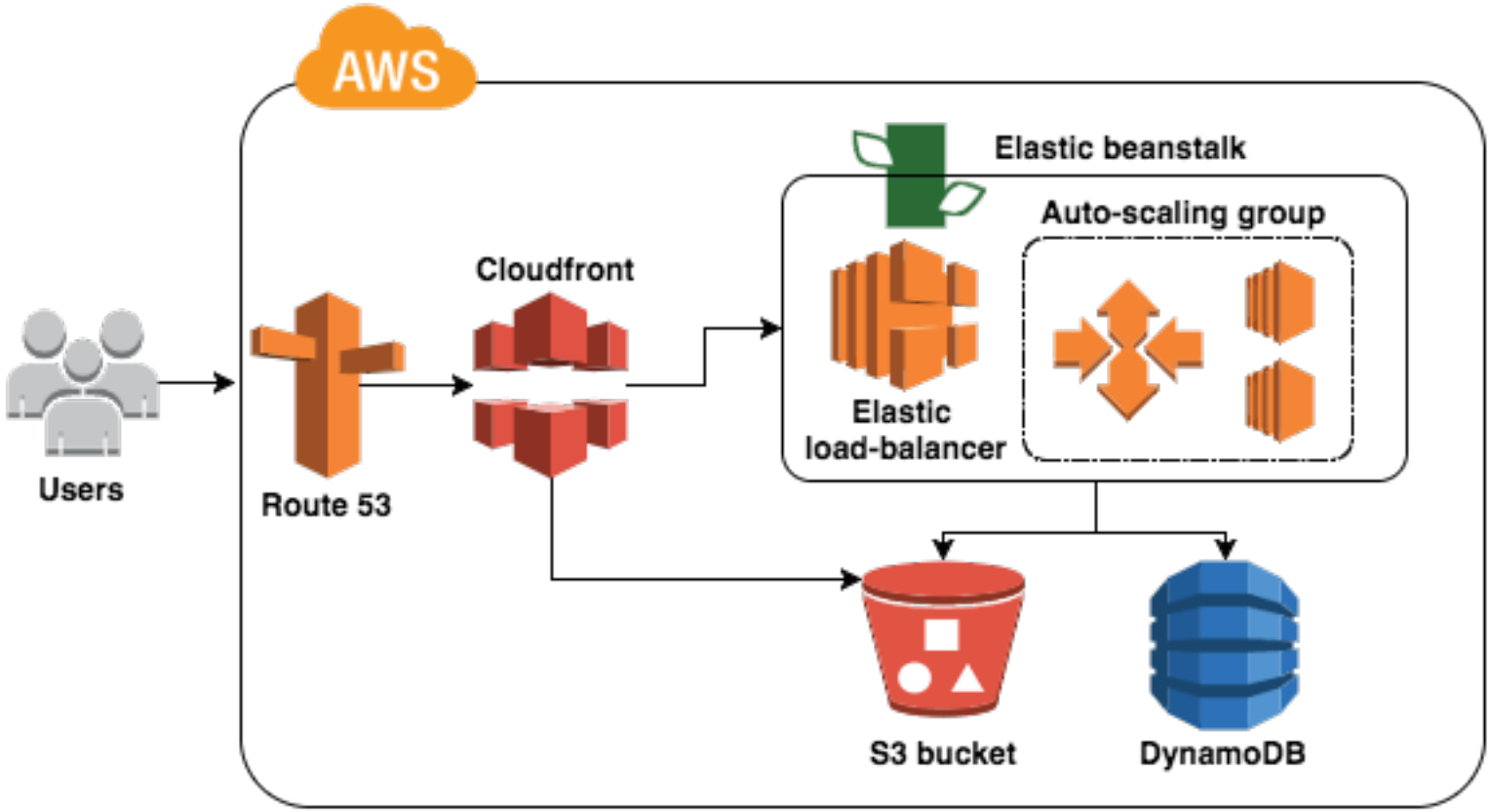
Route53

- Geo-Location Routing Policy
  - Lets you choose where your traffic will be sent based on the geographic location of your users
  - Good if you want all queries from Europe to be routed to a fleet of EC2 instances in one of the EU regions
  - Servers in these locations could have all prices and language set to EU standards for example



CloudFront (CDN)

- CDN (Content Delivery Network)
  - CloudFront is a content delivery network (CDN) service that accelerates delivery of your websites, APIs, video content or other web assets.

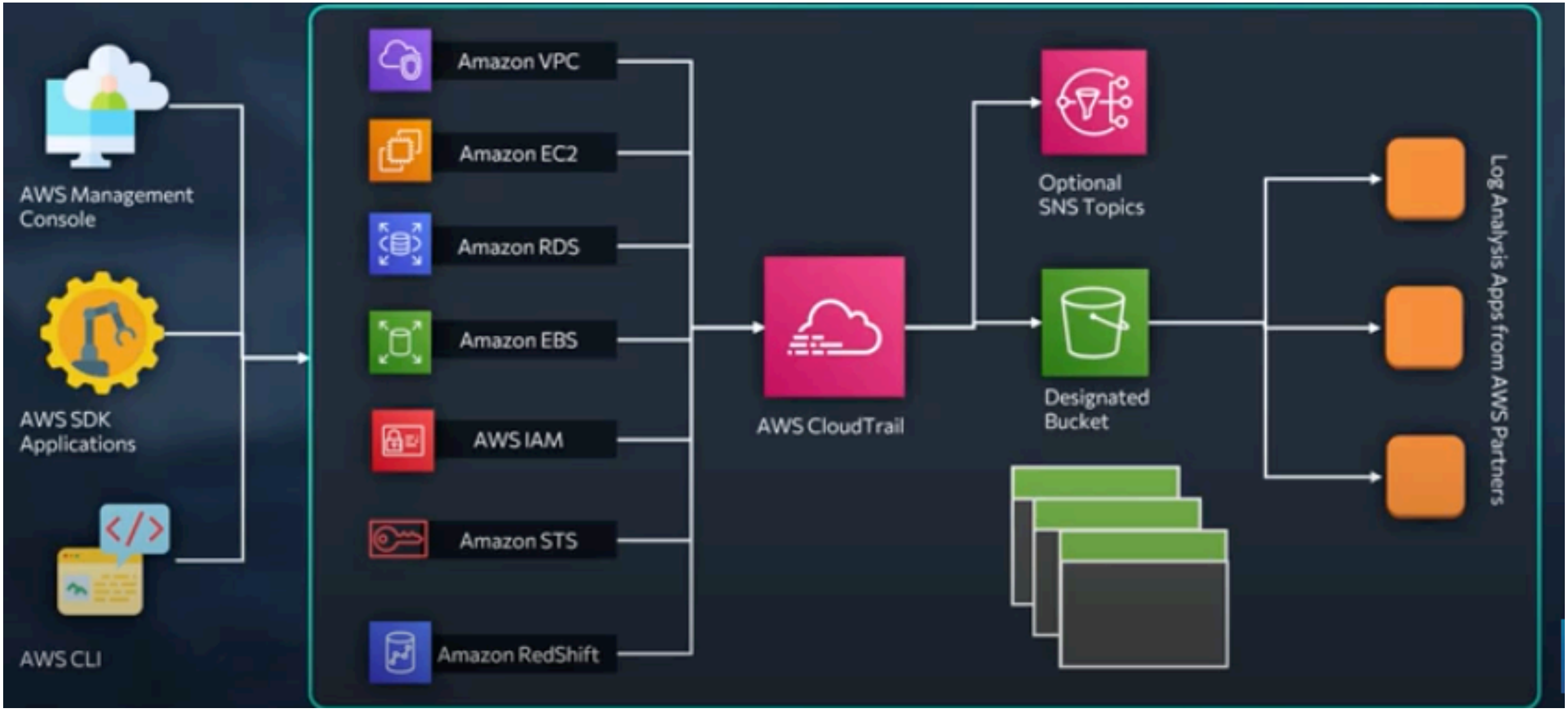


CloudFront (CDN)

- CDN (Content Delivery Network)
  - Edge Location is the location where content will be cached, separate from an AWS Region/AZ
  - Origin is the origin of all files, can be S3, EC2 instance, a ELB, or Route53
  - RTMP - (Real-Time Messaging Protocol) used for streaming media typically
  - Objects are cached for the life of the TTL (24 hours by default)
  - You can assign Web Application Firewall rules to your distributions
  - Edge locations can be R/W and will accept a PUT request on an edge location, which then will replicate the file back to the origin

Resource or Operation	Default Limit
Data transfer rate per distribution:	40 Gbps
Requests per second per distribution:	100,000
Web distributions per account:	200
RTMP distributions per account:	100
Alternate domain names (CNAMEs) per distribution:	100

CloudTrail


AWS Management Console
AWS SDK Applications
AWS CLI
Amazon VPC
Amazon EC2
Amazon RDS
Amazon EBS
AWS IAM
Amazon STS
Amazon RedShift
AWS CloudTrail
Optional SNS Topics
Designated Bucket
Log Analysis Apps from AWS Partners

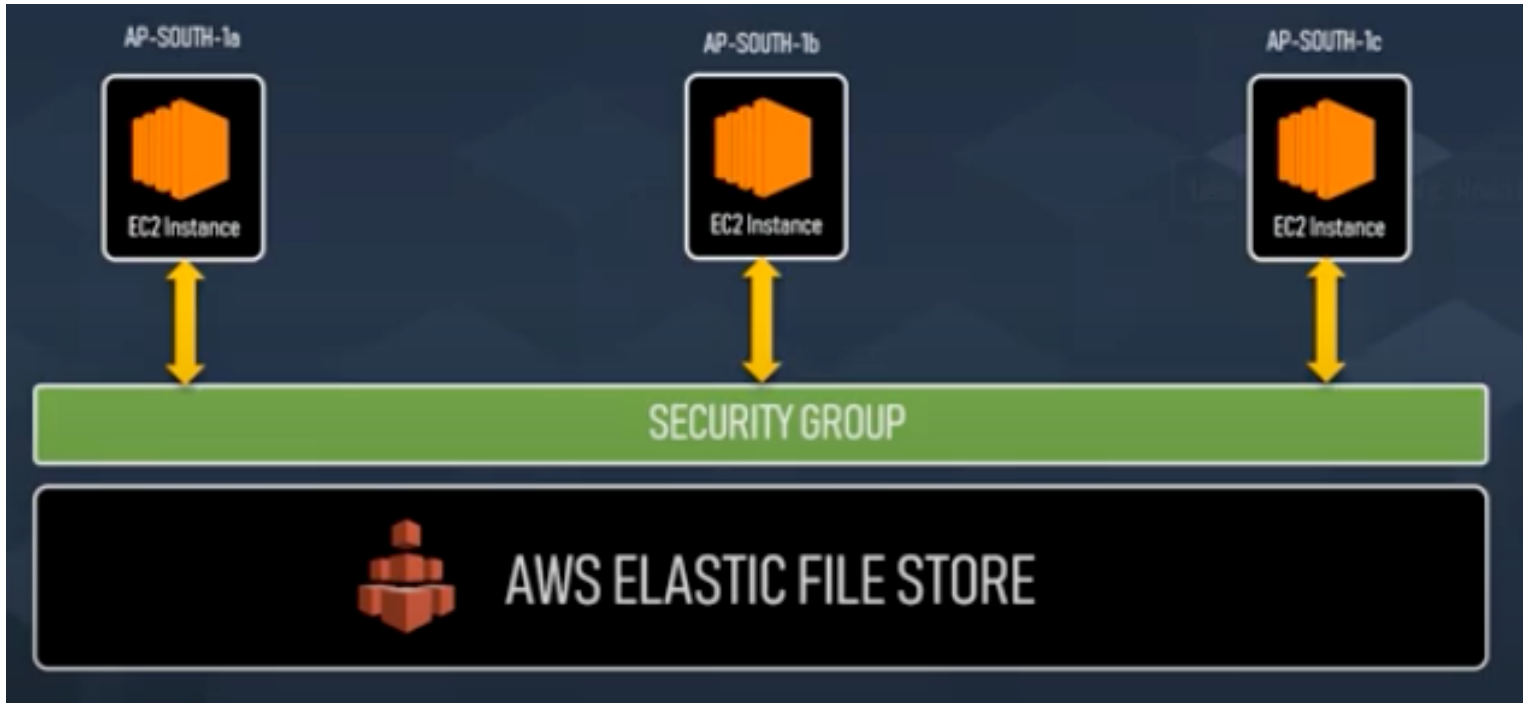
Cloud Trail

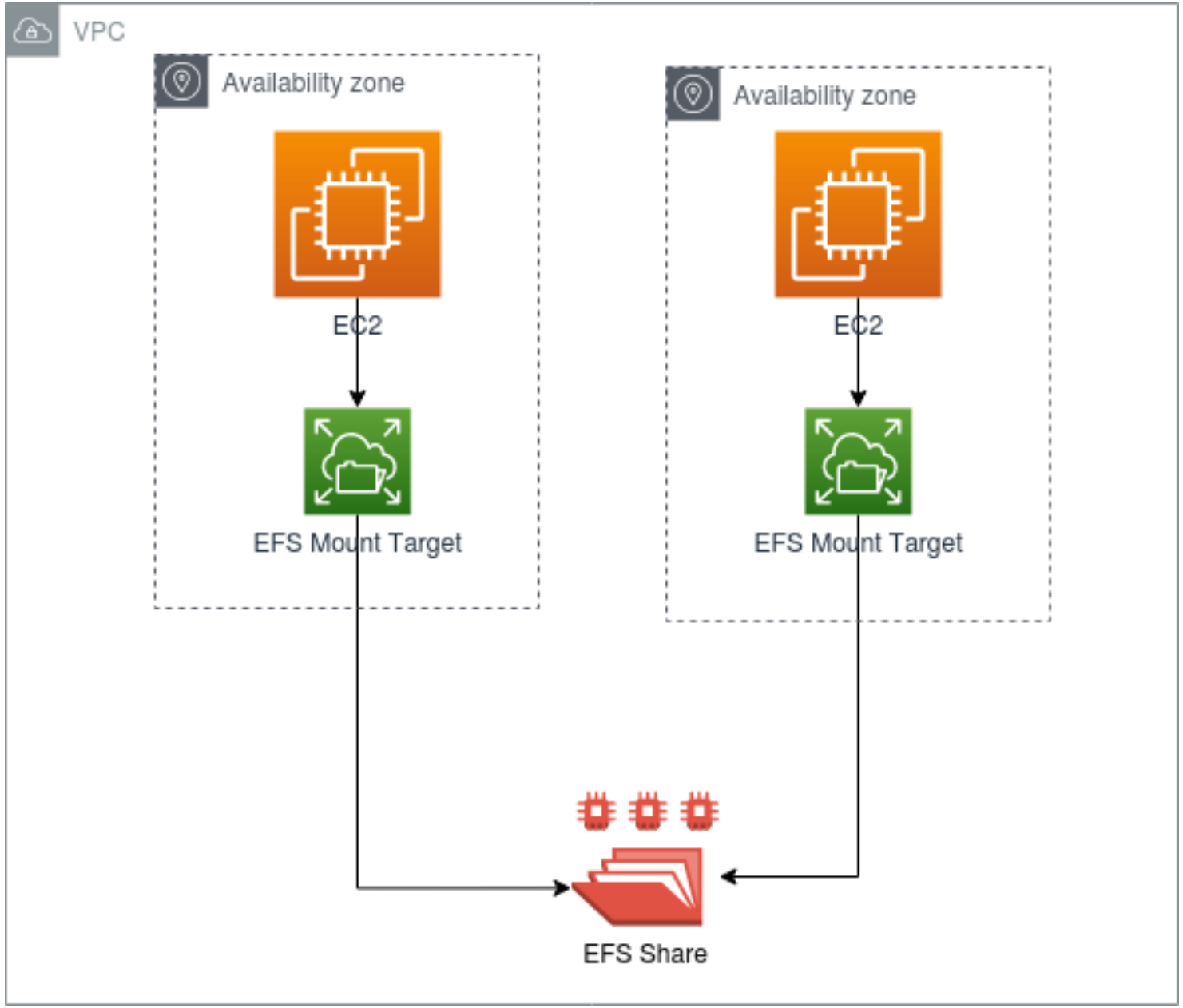
- Cloud Trail

  - AWS CloudTrail is a web service that records AWS API calls for your account and delivers log files to you
  - Provides way for customers to audit access to what people are doing on the platform in your account
  - CloudTrail can also be configured to support security information (SIEM) and event management platforms and and resource management
  - Configured to aggregate log files across multiple accounts so that log files are delivered to a single S3 bucket

EFS (Elastic File Storage)

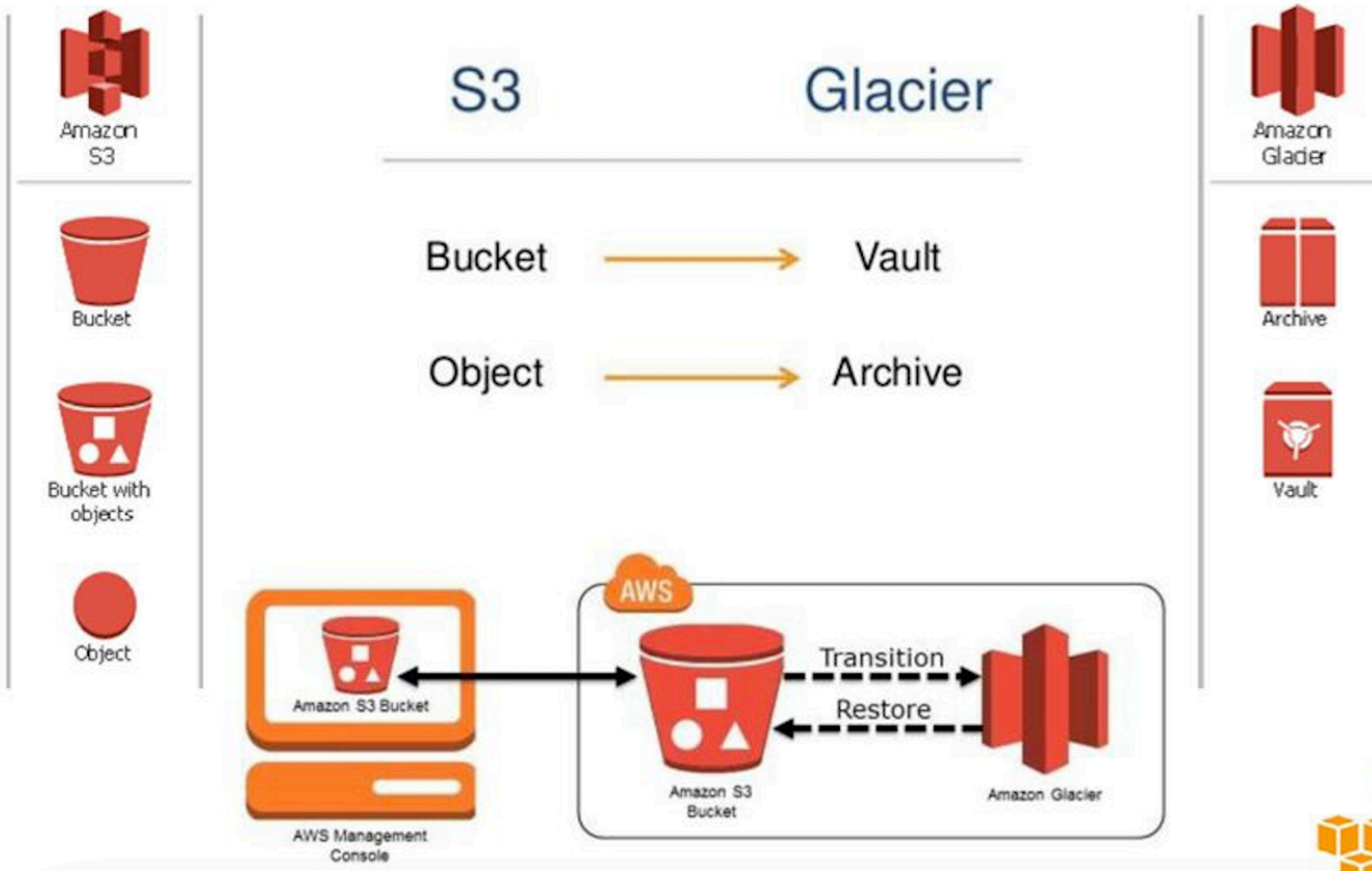
- Simple, scalable, fully managed NFS
- Build to scale on demand to petabytes without disrupting applications
- EFS is designed to provide massively parallel shared access to multiple ec2 instances
- EFS is regional scope service




VPC
Availability zone
EC2
EFS Mount Target
Availability zone
EC2
EFS Mount Target
EFS Share

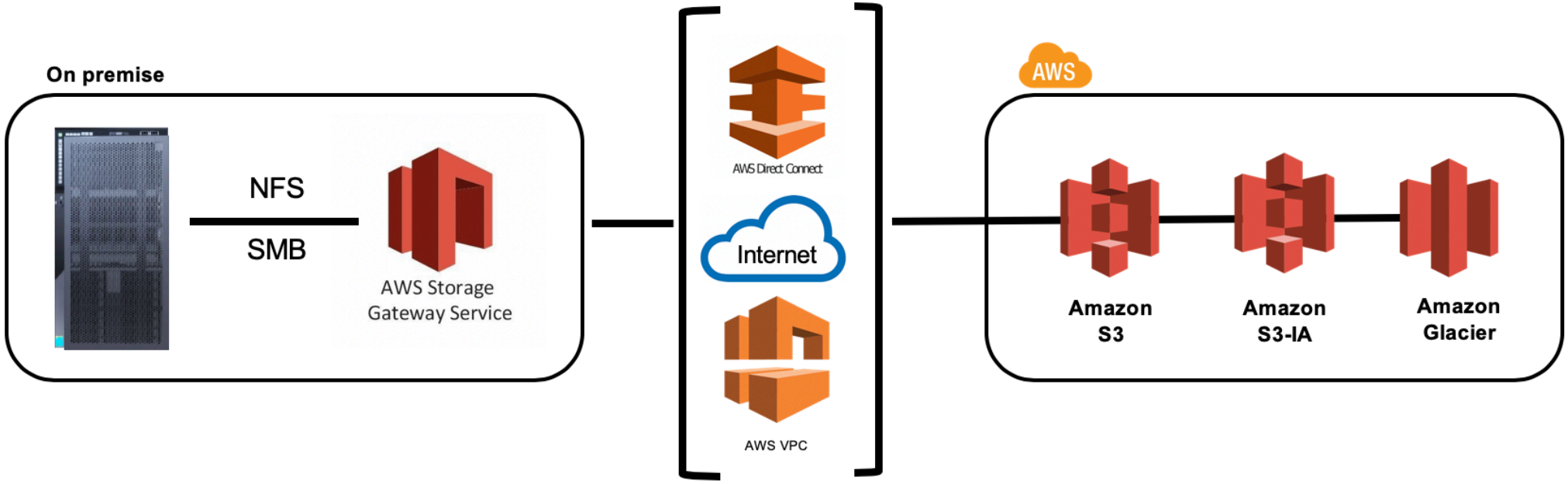
S3 Glacier

- Storage class to keep data backup or archive
- Stores in format of vault
- Extremely low cost archive storage service
- Provides high-durability storage
- Store for as little as $0.01 per GB per month
- Allows you to retrieve data within 3-5 hours



Storage Gateway

AWS Storage Gateway File Gateway Architecture



AWS on-premise be:

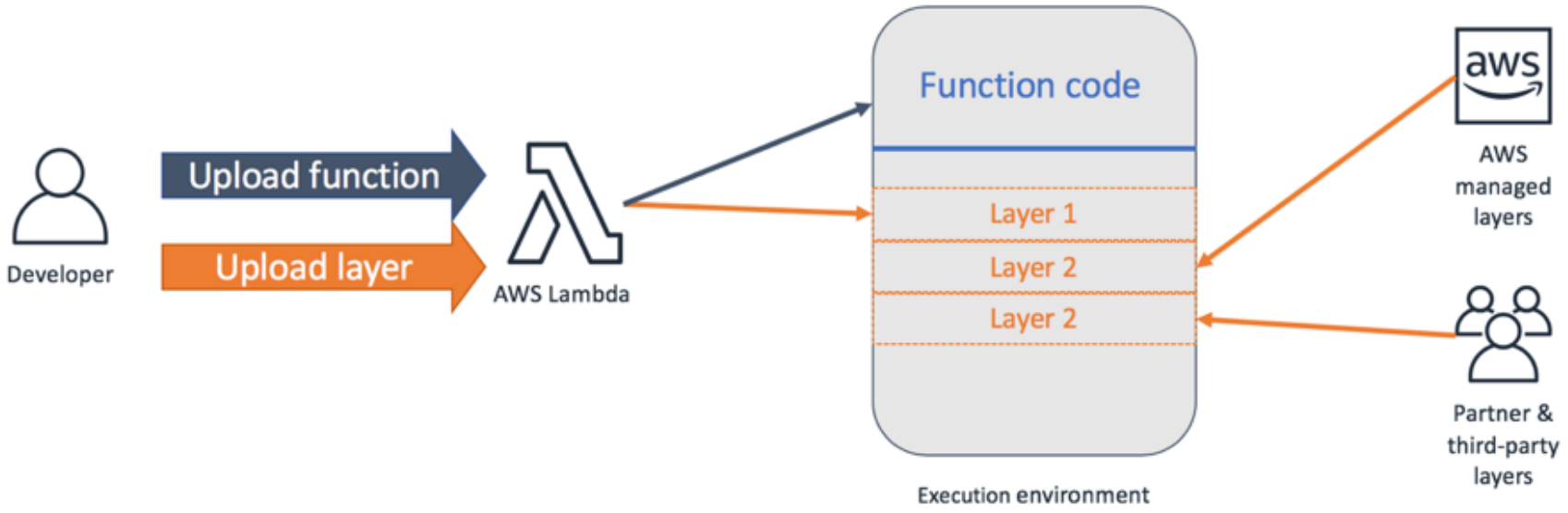
- Virtualized on VMware or Hyper-V
- Hardware Appliance

Features

- Transfer between on-premise and S3 via HTTPS
- Multi-part upload and local cache

Lambda

- AWS Lambda is an Amazon serverless computing system that runs code and automatically manages the underlying computing resources.
- **Advantages:**
  - It doesn't require the user to manage any servers.
  - It empowers the user to easily scale.
  - It's affordable.
- **What Can You Build with AWS Lambda?**
  - Data processing
  - Real-time file processing
  - Data validation
  - Filtering
  - Sorting
  - Real-time stream processing
  - 3rd-party API requests



Database (RDS)

- RDS (Relational Database Service)
  - Amazon RDS makes it easy to set up, operate, and scale a relational database in the cloud
  - Traditional relational databases that include tables, rows, fields
  - On-Line Transaction Processing (OLTP) type DB
  - You scale your DB by taking a snapshot and doing a restore to a larger sized tier
  - Encryption is done using the AWS Key Management Service (KMS)
  - To use RDS encryption, create a new DB instance with encryption enabled and migrate your data to it
  - Encrypting an existing DB instance is not supported
  - Supported RDS Platforms:
    - MS SQL Server
    - Oracle
    - MySQL Server
    - PostgreSQL
    - Aurora
    - MariaDB

Database (RDS)

- RDS (Relational Database Service)
  - Multi-AZ:
    - Allows you to have an exact copy of your production database in another AZ
    - AWS handles the replication for you, so when your prod database is written to, the write will automatically be synchronized to the stand-by DB
    - In a fail-over scenario, the same DNS name is used to connect to the secondary instance, There is no need to reconfigure your application
    - Multi AZ configurations are used for HA/DR only, and is not used for improving performance
    - To scale for performance you need to set up read replicas
    - Available for SQL Server, Oracle, MySQL, PostGreSQL, and Aurora