AWS solution architect

Associate Study metarial

Topic

- Introduction to AWS
- Amazon S3 and Glacier Storage
- Amazon EC2 and EBS
- Amazon VPC
- ELB, Amazon CloudWatch and Auto Scaling
- AWS Identity and Access Management(IAM)
- Amazon RDS
- SQS, SWF and SNS
- DNS and Amazon Route 53
- Amazon ElastiCache
- Additional Key Service
- Security on AWS
- AWS Risk and Compliance
- Architecture Best Practice
- Questions

Introduction to AWS

- Cloud Computing
- Advantage of Cloud Computing
- Six advantage of cloud computing
- Economies of Scale

Cloud computing

- Cloud computing is on-demand delivery of IT resource and application via the internet
- You don’t need to make large up-front investments in hardware and spend a lot of time managing that hardware
- easy way to access servers, storage, databases, and broad set of application services over the internet

Advantage of Could Computing

- Ability to reconfigure the computing environment quickly to adapt to changing business.
- Capacity can be automatically scaled up or down to meet fluctuating usage patterns.

Six advantage of cloud computing

Six Advantages

- Global in minutes
- Variable vs. capital expense
- Stop guessing capacity
- Economies of scale
- Focus on business differentiators
- Increase speed and agility

Variable vs. Capital Expense

- you can pay only when you consume computing resources and pay only for how much you consume.

Economies of Scale

- organization benefit from massive economies of scale.
- AWS can achieve higher economies of scale which translate into lower prices.

Stop Guessing Capacity

- they can access as much or as little as they need and scale up or down as required with only a few minutes' notice

Increase Speed and Agility

- In a cloud computing environment, new IT resources are one click away which allows organizations to reduce the time it takes to make those resources available to developers from weeks to just minutes.

Focus on Business Differentiators

- To focus on customers business priorities.
- Organizations can stop spending money on running and maintaining data center.
- to Focus on projects
- - Analyzing petabytes of data
- - Delivering video content,
- - Building application

Go Global in Minutes

- Organizations can easily deploy their applications to multiple locations around the world with just a few click.
- Allow organizations to provide redundancy across the globe and to deliver lower latency and better experiences to their customers at minimal cost.

Cloud Computing Deployment Model

- “all-in” cloud-based deployments and hybrid deployments.
- all-in-cloud-based application is fully deployed in the cloud, with all components of the application running
- Hybrid deployments is connects infrastructure and application between cloud-based resources and existing resources. -> Extended to cloud infra from existing data center.

AWS fundamentals

- on-demand delivery of IT resources via the internet on secure cloud services platform.
- Capacity exactly matches you need
- you pay only for what you use.
- economies of scale result in lower cost.
- service is provided by a vendor experienced in running large-scale networks.

Global Infrastructure

- Regions – separate geographic area
- Availability zones – isolated location in region.
- AWS enables the placement of resources and data in multiple locations.
- Resources aren’t replicated across regions unless organizations choose to do so.

Regions and Availability Zone

- The Region is one of physical location
- The Availability Zone is member of region
- Each AZ is separated, isolated with each AZ even though they located at same metropolitan.
- You can achieve high availability by deploying your application across multiple Availability Zones. Redundant instnaces for each tier.

AWS Cloud Computing Platform

Enterprise Application	Virtual Desktops		Sharing and Collaboration		
Platform Service	Database - Relational - NoSQL - Caching	Analytics - Hadoop - Real-time - Data warehouses - Data workflows	App Service - Queuing - Orchestratio n - App Streaming - Transcoding - Email - Search	Deployment and Management - Containers - Devops Toosl - Resources Templates - Usage Tracking - Monitoring and logs	Mobile Services - Identity - Syns - Mobile Analytics - Notification s
Foundation Service	Compute (VMs, Auto Scaling and load Balancing		Storage (Object, Block and archive)	Security and Access Control	Networking
Infrastructure	Regions	Availability Zone	Content Delivery Networks and Point of Presence		

Accessing the Platform

- AWS Management Console(WEB)
- AWS Command Line Interface(CLI)
- AWS Software Development Kits(SDK)

Compute and Networking Service

- Amazon Elastic Compute Cloud (EC2)
- AWS Lambda
- Auto Scaling
- Elastic Load Balancing
- AWS Elastic Beanstalk
- Amazon Virtual Private Cloud(VPC)
- AWS Direct connect
- Amazon Route 53

Storage and Content Delivery

- Amazon Simple Storage Service(S3)
- Amazon Glacier
- Amazon Elastic Block Store(EBS)
- AWS Storage Gateway
- Amazon CloudFront

Database Services

- Amazon Relational Database Service(RDS)
- Amazon DynamoDB
- Amazon RedShift
- Amazon ElastiCache

Management Tools

- Amazon CloudWatch
- AWS CloudFormation
- AWS CloudTrail
- AWS Config

Security and Identity

- AWS Identity and Access Management(IAM)
- AWS Key Management Service(KMS)
- AWS Directory Service
- AWS Certificate Manager
- AWS Web Application Firewall

Application Services

- Amazon API Gateway
- Amazon Elastic Transcoder
- Amazon Simple Notification Service
- Amazon Simple Email Service
- Amazon Simple Workflow Service
- Amazon Simple Queue Service

Amazon S3 and Glacier Storage

- Object Storage vs. Traditional Block and File Storage
- Buckets
  - S3 Object metadata , keys, URL and operation
- Durability and Availability
- Best practice for protect data
- Data Consistency
- Access Control
- Static website hosting
- S3 Advanced Features
- Prefixes and Delimiter
- Storage Classes
- Object Lifecycle Management
- Encryption
- Versioning
- MFA Delete
- Pre-signed URLs
- Multipart Upload
- Range GETs
- Cross Region Replication
- Logging
- Event Notifications
- Amazon Glacier

Object Storage versus Traditional Block and File Storage

- S3
  - cloud object storage
  - Independent of a server and is accessed over the internet
  - manage as object using an application program interface(API) build on standard HTTP verbs.
  - object contains both data and metadata
  - buckets are a simple flat folder with no file system hierarchy.
  - you cannot “mount” a bucket, “open” an object

Buckets

- Container(web folder) for objects(files) stored in Amazon S3
- your bucket names must be unique across all AWS account
- like DNS domain name
- bucket is create in a specific region that you choose
- object => entities or file store in buckets
- unlimited capacity (but one object can be 5TB)

S3 Object metadata

- name/value pair
- date last modified
- object size
- MD5 digest
- HTTP Content-Type
- Custom tag / data

S3 Object keys

- identified by a unique identifier called a key.
- key as a filename
- 1024 bytes of unicode UTF-8 characters.
- include embedded slashes,backslashes, dots, and dashes
- unique within single bucket

S3 Object URL

- amazon S3 object can be addressed to web service endpoint.
- There is no actual file and folder hierarchy.
- example :
  http://mybucket.s3.amazonaws.com/jack.doc

S3 Operations

- Create / delete a bucket
- Write an object
- Read an object
- Delete an object
- List keys in a bucekt
- REST interface
  - create : HTTP PUT (or sometimes post)
  - read : HTTP GET
  - delete : HTTP DELETE
  - update : HTTP POST (or sometimes put)

Durability and Availability

- Durability : Will my data still be there in the future?
- Availability : Can I access my data right now?
- 10,000 object save, 10,000,000 years later will lost one object (99.999999999%)
- 99.99% Availability
- multiple device and multiple facilities within a region.

Best practice for protect data

- use versioning
- cross-region replication
- MFA(Multi Factor Authentication) delete

Data Consistency

- replicate across multiple server and location within a region
- PUT : read-after-write
- PUT new data existing key, a subsequent GET might return the old data
- DELETE an object, a subsequent GET for that object might still read the delete object

Access Control

- S3 bucket Access control
  - be associated with different AWS account
- S3 bucket policies
  - CIDR Block / IP Address and time
- IAM policies
- Query-string authentication

Static website hosting

- Create a bucket with same name as the desired website hostname
- Upload the static files to the bucket
- Make all the files public(world readable)
- Enable static website hosting for the bucket. (include index , error)
- The website will now be available at the S3 website URL : <bucket-name>.s3-website-<AWS-region>.amazonaws.com
- Create friendly DNS name in your own domain for the website using DNS CNAME or ROUTE 53 alias

S3 Advanced Features

- Prefixes and Delimiters
- Storage Classes
- Object Lifecycle Management
- Encryption
  - SSE-S3 (AWS Managed Keys)
  - SSE-KMS(WAS KMS keys)
  - SSE-C(Customer Provisioned Keys)
  - Client-Side Encryption
- Versioning
- MFA Delete
- Pre-Signed URLs
- Multipart Upload
- Range GETs
- Cross-Region Replication
- Logging
- Event Notifications

Prefixes and Delimiter

- slash(/) and backslash(₩) are as delimiter.
- prefixes and delimiter are help to look easy as ordinary file system(folder-and-file structure)
- REST API, SDK, CLI and web all support the use of delimiters and prefixes.
- S3 is really not a file system (its real flat)

Storage Classes

- S3 Standard
  - General purpose use
- S3 Standard-Infrequent Access
  - designed for long-lived , less frequently accessed data
- S3 Reduced Redundancy Storage(RRS)
- Glacier
  - long-term backups, archive
  - optimized for infrequently accessed data
  - copy data to S3 RRS when request after 3 to 5 hours

Object Lifecycle Management

- Create(Hot) -> warm(less frequently) -> cold(long-term backup or archive) -> delete
- Store backup data initially in S3 Standard
- After 30 days, transition to Standard-IA
- After 90 days, transition to Glacier
- After 3 years, delete
- lifecycle can apply to all object in the bucket or only to objects specified by a prefix

Encryption

- SSE-S3 (AWS-Managed keys)
  - fully integrated “check-box-style” encryption solution where AWS handles the key management and key protection for S3
- SSE-KMS (AWS KMS keys)
  - fully integrated solution
  - separate permissions for using the master key
  - AWS KMS also provides auditing
- SSE-C (Customer-Provided keys)
  - maintain your own encryption keys
  - S3 will management for encrypt/decrypt
- Client-Side Encryption
  - Use an AWS KMS-managed customer master key
  - Use a client-side master key
  - handle by application before save and after read

Versioning

- protect data against accidental or malicious deletion
- preserve, retrieve and restore every version of every object stored in S3 bucket

MFA Delete

- data protection on top of bucket versioning
- MFA delete requires an authentication code(a temporary, one-time password) generated by a hardware or virtual Multi factor authentication device.

Pre-signed URLs

- Using their own credentials to grant time-limited permission to download object
- valid only for the specified duration
- prevent (protect against) “content scraping”

Multipart Upload

- Support uploading or copying large object
- ability to pause and resume
- initiation, uploading the parts, completion(or abort)
- recommend larger than 100Mbyte
- Must use this for larger than 5GB
- check SDK(manual) and CLI (automatic)

Range GETs

- Can download only portion
- range of bytes of the object
- dealing with large object on poor connectivity
- download only portion of a large Glacier backup

Cross Region Replication

- Asynchronously replicate all object from a region to another region.
- After set up cross-region replication -> any changes to the data ,metadata or ACLs on an object trigger a new replication to the destination bucket.
- Must enable versioning feature
- IAM policy to give S3 permission to replicate objects

Logging

- In order to tack request to your S3 Bucket
- content
  - Requestor account and IP address
  - Bucket name
  - Request time
  - Action (GET, PUT, LIST, and so forth)
  - Response status or error code

Event Notifications

- Object uploaded or stored in S3
- event notifications to set up triggers to perform actions

Amazon Glacier

- Archives
  - up to 40TB of data archive
  - Automatically encrypted
  - Immutable (cannot modified)
- Vaults
  - Containers of archives
- Vaults Locks
  - Write once ready many
  - once locked, no longer change
- Data Retrieval
  - 5% data retrieve is free each month from Glacier to S3
- Glacier vs S3
  - Glacier has generated by system key (as filename) -> not friendly

EC2 and EBS

- Compute Basics
  - Instance Types
  - Amazon Machine Images (AMIs)
- Securely Using an Instance
  - Addressing an Instance
  - Initial Access
  - Virtual Firewall Protection

- The Lifecycle of Instance
  - Launching
  - Bootstrapping
  - VM Import/Export
  - Instance Metadata
  - Managing Instance
  - Monitoring Instance
- Modifying an Instance
  - Instance type
  - Security groups
  - Termination Protection
- Options
- Instance Stores

EC2 – Compute Basics

- Instance Types
  - Virtual CPUs (vCPUs)
  - Memory
  - Storage (size and type)
  - Network performance
  - C4 : Compute Optimized (**C**pu)
  - R3 : Memory Optimized (**R**am)
  - i2 : Storage Optimized (**I**ops)
  - g2 : GPU-based instance (**G**pu)

Network Performance

- Some instance type support 10Gbps
- Enabling capability Single Root I/O Virtualization(SR-IOV)
- Available some instances launched in an VPC

Amazon Machine Images(AMIs)

- OS and its configuration
- Initial state of any patches
- Application or system software
- Based on x86 OSs, either Linux or Windows
- Published by AWS
- The AWS marketplace
- Generated from Existing Instances
- Uploaded Virtual Servers

Securely Using an Instance

- Addressing an Instance
  - Public Domain Name System(DNS) Name
    - Generated by AWS (Customer cannot change)
    - Persists during launch instance
  - Public IP
  - Private IP
    - Available in Amazon VPC
- Initial Access
  - Public key cryptography to encrypt and decrypt login
  - it can download first time when make it.
  - Amazon Linux user : ec2-user
  - Ubuntu ubunt : ubuntu
- Virtual Firewall Protection
  - Using Security group
  - SG is default deny all traffic
  - The rules are aggregated and all traffic allowed by each of indivisual groups is allowed
  - SG applied at the instance level

The Lifecycle of Instances

- Launching
  - Bootstrapping
    - some way to configure instances and install applications
    - Using called *"UserData"*
    - *"UserData"* can be shell script or windows batch
  - VM Import/Export
  - Instance Metadata
- Managing Instances
  - Tags are key/value pairs , associate with your instance or other service
- Monitoring Instances
- Modifying an Instance
  - Instance Type
    - stopped and change
  - Security Groups
    - can change SG in VPC, Cannot change EC2-Classic(without VPC)
- Termination Protection

Options

- Pricing Options
  - On-Demand Instances
  - Reserved Instances
  - Spot Instances
    - Recommend to OLAP
- Tenancy Options
  - shared tenancy
  - Dedicate instances
  - Dedicate host
- Placement Groups
  - A logical grouping of instances within a single AZ
  - For low network latency, High Network throughput
- Instance Stores
  - Temporary storage
  - Physically attached to the host computer
  - delete all data when stop or terminates

Amazon Elastic Block Store(EBS)

- Basic
  - Persistent block-level storage volume
- Types of Amazon EBS volumes
  - Magnetic Volumes
    - 1 GB to 1TB / 100 IOPS
  - General-Purpose SSD
    - 1 GB to 16 TB
    - Max 10,000 IOPS
    - billed based on the amount of data space provisioned
  - Provisioned IOPS SSD
    - very high performance random I/O
    - 4 GB to 16 TB
    - MAX 20,000 IOPS
    - billed based on size of volume and reserved IOPS
  - Amazon EBS-Optimized Instances
    - Dedicated capacity for Amazon EBS I/O
- Protecting Data
  - Backup/Recovery(snapshot)
    - Use S3 technology by AWS own , save to same region (Cannot move to another region)
  - Recovering Volumes
  - Encryption Options

EBS Comparison

Characteristic	General-Purpose SSD	Provisioned IOPS SSD	Magnetic
Use cases	- System boot volumes - Virtual desktops - Small-to-medium size databases - Development and test environments	- Critical business applications that required sustained IOPS - Performance or more than 10,000 IOPS or 160MB of throughput per volume - Large database	- Cold workloads Infrequently accessed - Lowest storage cost is important
Volume size	1GiB ~ 16TiB	4GiB ~ 16TiB	1GiB ~ 1TiB
Max throughput	160MB	320MB	40~90MB
IOPS Performance	Baseline performance of 3 IOPS/GiB (up to 10,000 IOPS)	Up to 20,000 IOPS	Avg. 100 IOPS

Amazon Virtual Private Cloud(VPC)

- Private network layer size of /16 ~ /28 (CIDR)
- Subnets
- Route Tables
- DHCP option sets
- Security Groups
- Network ACLs
- Internet Gateway
- Elastic IP Address
- Elastic Network Interfaces
- Endpoints
- Peering
- NAT instances and NAT gateway
- Virtual Private Gateway
- Customer Gateways
- Virtual Private Networks

VPC – Subnets

- First four and last one IP address is reserved for every subnet.
- Cannot span zones.
- By default, all subnet can communicate within VPC

Route Tables

- A route table is a logical construct a set of rules in VPC
- Permit instances within different subnets within a VPC to communicate with each other.
- Each subnet must be associated with a route table.
- Each route in a table specifies a dest CIDR and a target.

Internet Gateway

- Allows communicate between instances in VPC and the internet.
- VPC’s instances are only aware of their private IP address
- Attach an IGW -> Add rule for the internet -> Configure ACLs and SG rule
- Assign a public IP or EIP to EC2 instance

DHCP Option sets

- Provide configuration information to host on a TCP/IP
- Domain-name-servers(default to AmazonProvideDNS)
- Domain-name(default to domain name for your region)
- AmazonProvideDNS let enable communicate over IGW

Elastic IP Addresses(EIPs)

- Static, Public IP address in the pool for the region
- EIP for use within VPC then assign it to an instance.
- Specific region
- 1-to-1 relationship with Network Interface and EIPs.
- Can move EIPs to another instance (same region)
- remaining associated until release
- changes for EIPs allocated to your account

Elastic Network Interfaces(ENIs)

- ENIs is a virtual network interface within VPC.
- Only available within an Amazon VPC
- ENIs allow create a management network.
- Workloads/roles on distinct subnets or HA solution.

Endpoints

- Enables you to create private connection
- Create multiple endpoints for a single service
- Different route table to enforce different access policies
- identified by a prefix list
- Allow full access or create custom policy
-

Peering

- Each VPC can have got a connection even different AWS account within same region.
- Request/accept protocol.
- Same account, it’s identified by its VPC ID
- Different Account is AccountID and VPC ID.
- Do not support transitive routing.
- Cannot create a peering between same CIDR blocks.
- Cannot create peering between different region.

Security Groups

- Control inbound and outbound traffic.
- Create up to 500 SG for each VPC.
- 50 Inbound Rule, 50 Outbound rule for one SG
- Associate up to five SG with one Network Interface
- Specify allow rule, not deny (can setup deny rule in network ACLs)
- Can separate inbound and outbound
- By default, no inbound allow , all outbound allow.

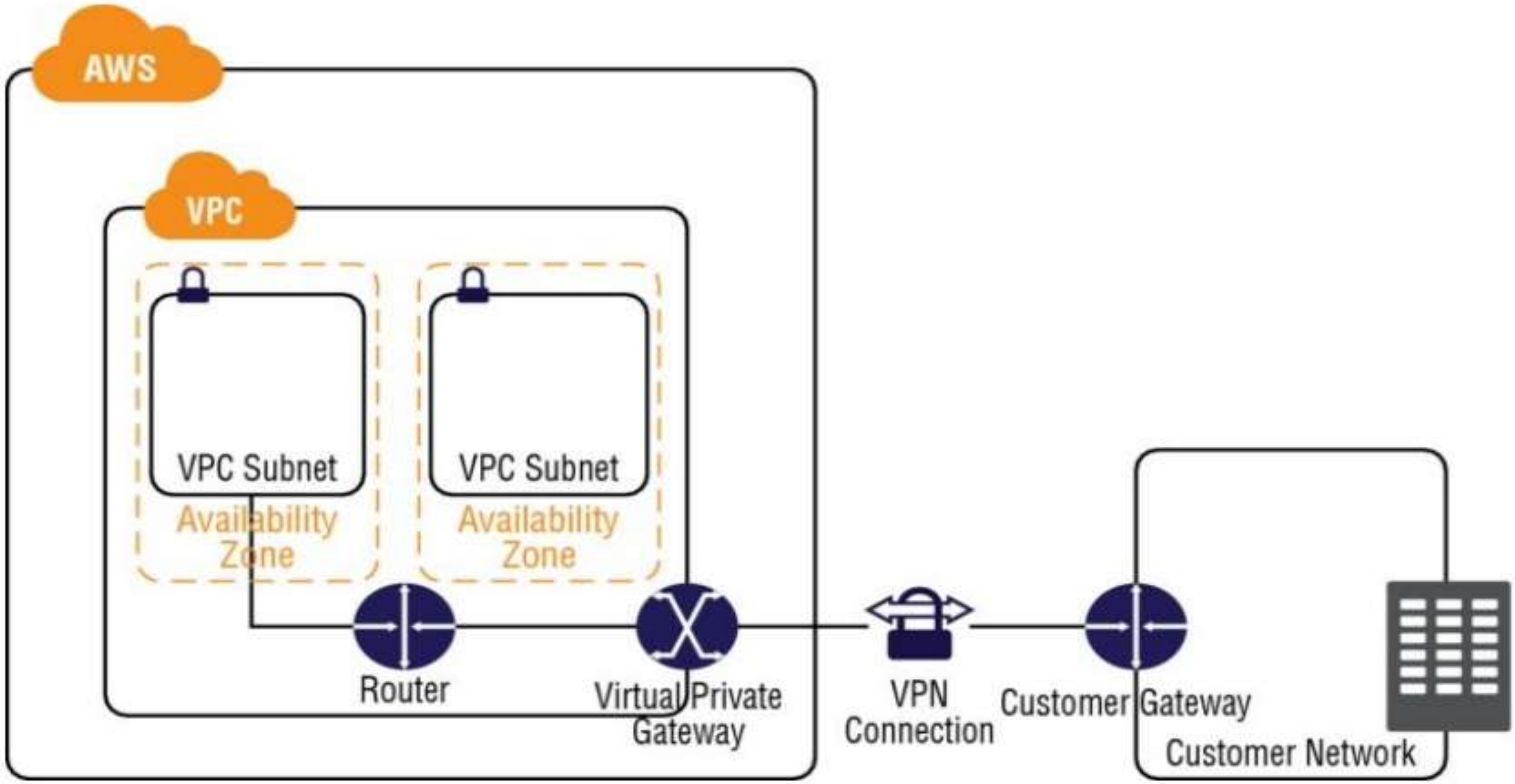
Network Access Control Lists(ACLs)

- Stateless firewall on a subnet level.
- Evaluates in order (starting with the lowest numbered rules)
- Support allow and deny rules
- stateless

NAT instance / gateway

- NAT instance
  - Use amzn-ami-vpc-nat AMI
  - Create SG for NAT
  - NAT AMI as in instance in a public subnet
  - **Source/Destination check attribute of the NAT**
  - Associate route table with private subnet and NAT instance
  - EIP associate it with the NAT instance
- NAT Gateway
  - NAT gateway is designed to operate just like NAT instance

VPGs, CGWs, VPNs



VPGs, CGWs, VPNs

- VPC supports multiple CGWs
- VPN connection to a single VPG(many-to-one design)
- CGW IP addresses must be unique within the region.
- VPG is the AWS end of the VPN tunnel
- CGW is a H/W or S/W on the customer side
- VGPs support BGP and Static routing
- VPN connection consists of two tunnel for HA to the VPC

ELB, CloudWatch, Auto Scaling

- Elastic Load Balancing
  - Support HTTP/S , SSL, TCP
  - CNAME of DNS (internet-facing , internal)
  - Integrate Auto Scaling
  - Distribute traffic
  - SSL Termination
- CloudWatch
- Auto Scaling

Types of Load Balancers

- Internet-Facing Load Balancers
  - Takes request from client over the internet and distributes
  - Only support IPv4 VPC (EC2-Classic is support IPv4 and IPv6)
- Internal Load Balancers
  - Load Balancing between tires of the application
- HTTPS Load Balancers
  - Dose not support *Server Name Indication(SNI)*
  - avoid SSL via *Subject Alternative Name(SAN) ->* for multi host

Listeners

- HTTP
- HTTPS
- TCP
- SSL

Configuring Elastic Load Balancing

- Idle Connection Timeout
  - By Default, 60sec for both connections.
  - enable keep-alive on server when use HTTP/S listener
- Cross-Zone Load Balancing
  - load balancing across AZ
- Connection Draining
  - Should enable connection draining in order to stop sending traffic to unhealthy instance
  - Can set up to 3,600sec from 1 (default : 300sec)
- Proxy Protocol
  - Add source / destination IP and port number into header
- Sticky Sessions
  - if enable -> LB to bind a user's session to a specific instance
- Health Checks
  - test to instances behind ELB
  - InService is can to serve
  - OutOfService is cannot to serve

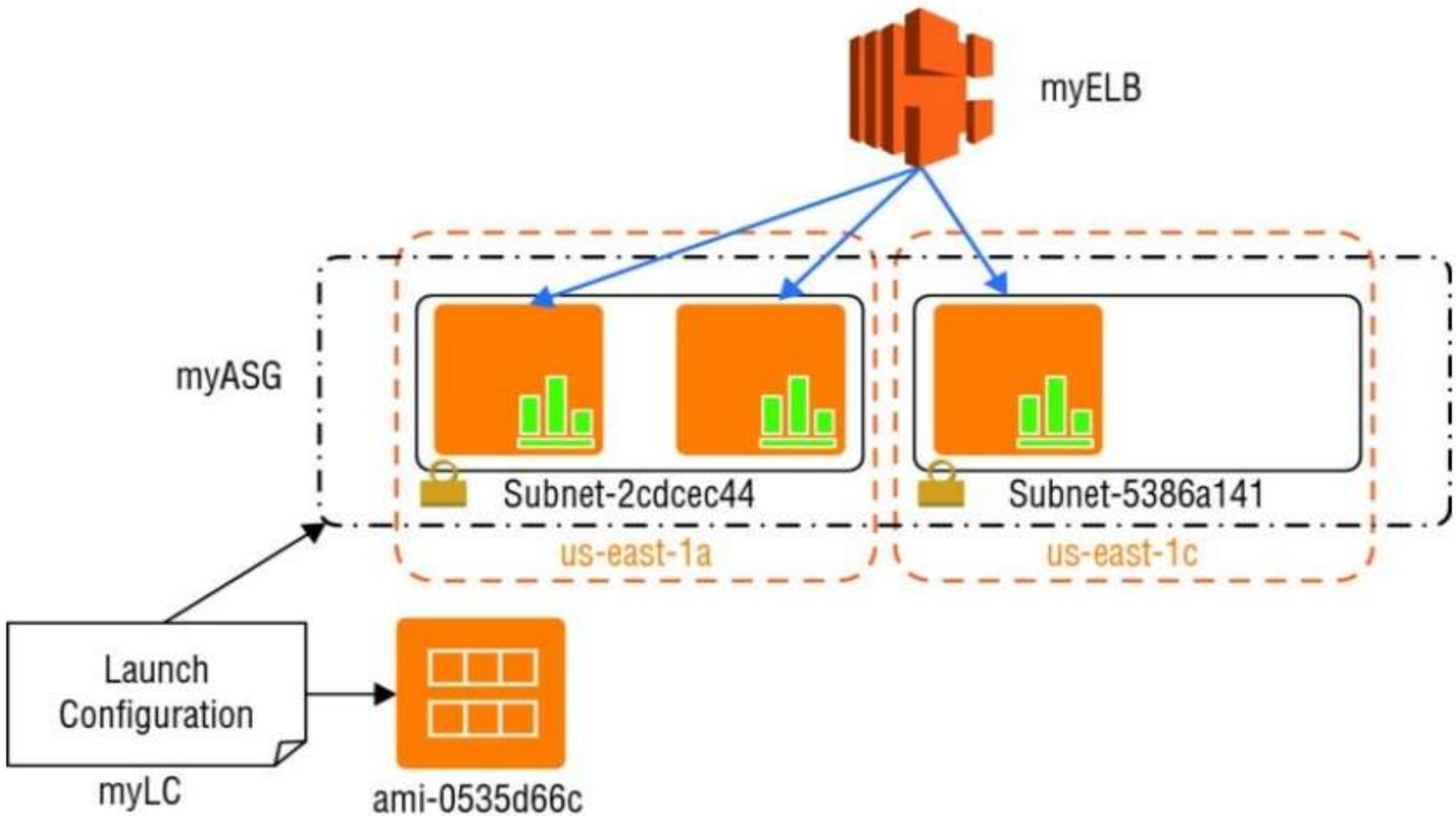
Amazon CloudWatch

- Monitor AWS resource and application in real time.
- Supports multiple action
- Basic Monitoring
  – Every 5min , limited number of preselected metric
- Detailed Monitoring
  – Every 5min, aggregate across AZ within region
- Cannot monitor inside instance's memory or cpu usage and application specific log or threshold
- A CloudWatch Logs agent is provide an automated to send log data to CloudWatch logs for AWS EC2 instances running. (Linux or Ubuntu)
- limited to 5,000 alarm per AWS account
- metrics data is retained for two weeks by default
- Metrics data can be store to S3 or Glacier

Auto Scaling

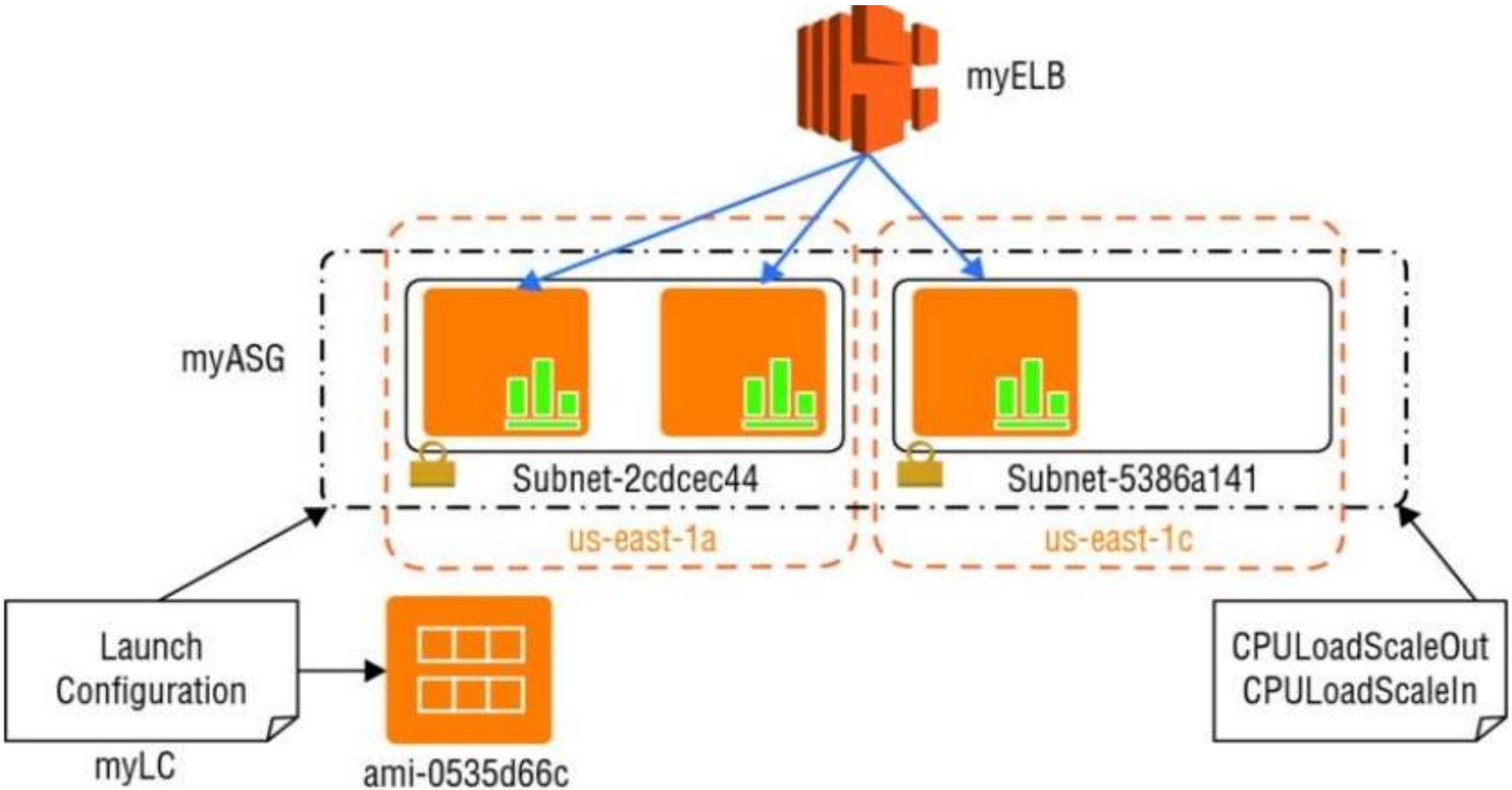
- Automatically change to scale depend on user's define
- Auto Scaling Plans
  - Manual
  - Scheduled
  - Dynamic
- Auto Scaling Components
  - Launch Configuration
    - Required define AMI, instance type
    - Optionally define SG, keypair
    - Limited 100 per region
  - Auto Scaling Group
    - minimum, maximum, desire capacity
  - Scaling Policy
    - Associate CloudWatch alarm and Scaling Policies
    - Define CPU load or Memory usage

Auto Scaling



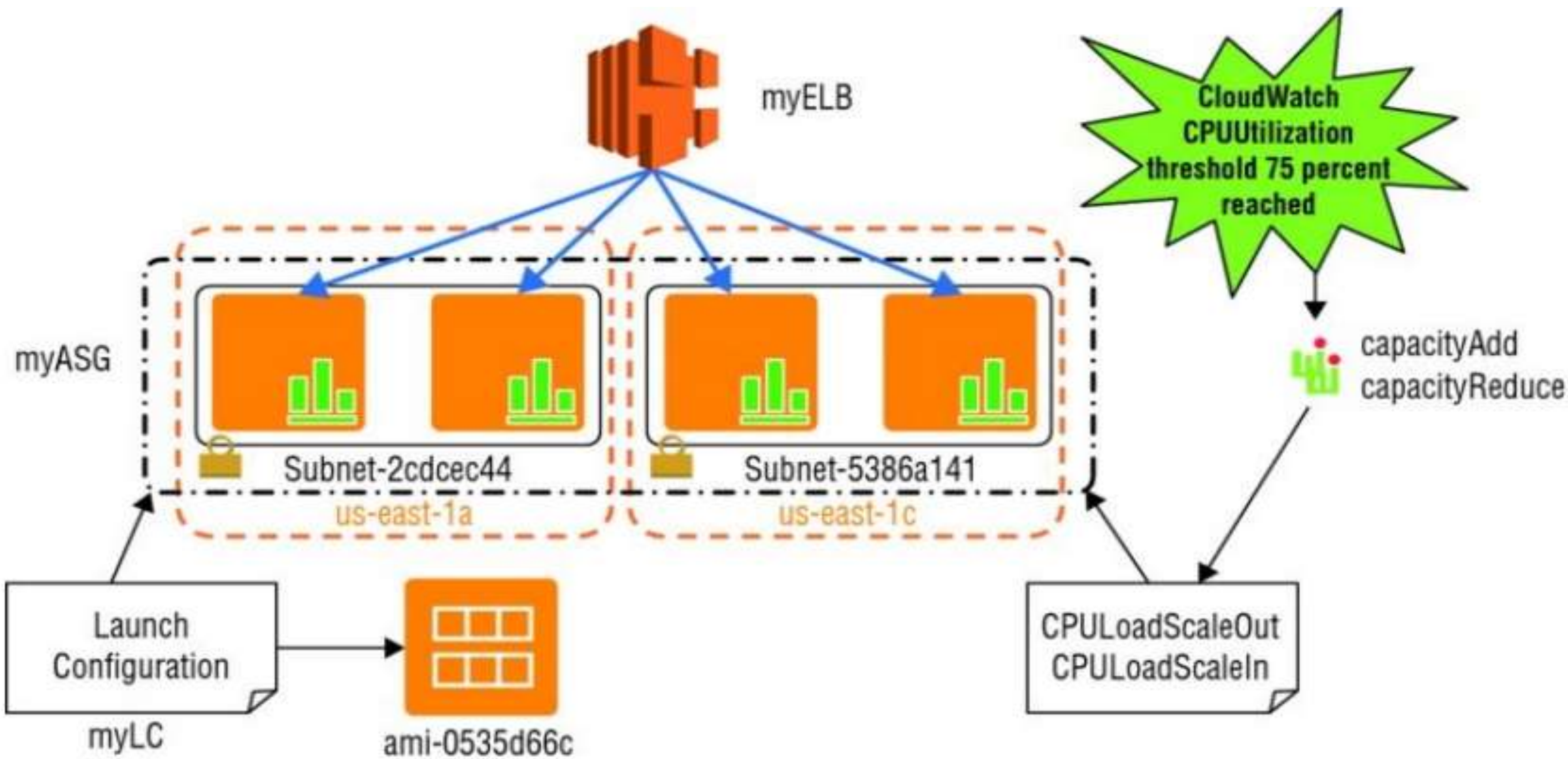
Auto Scaling group behind an Elastic Load Balancing load balancer

Auto Scaling



Auto Scaling group with policy

Auto Scaling

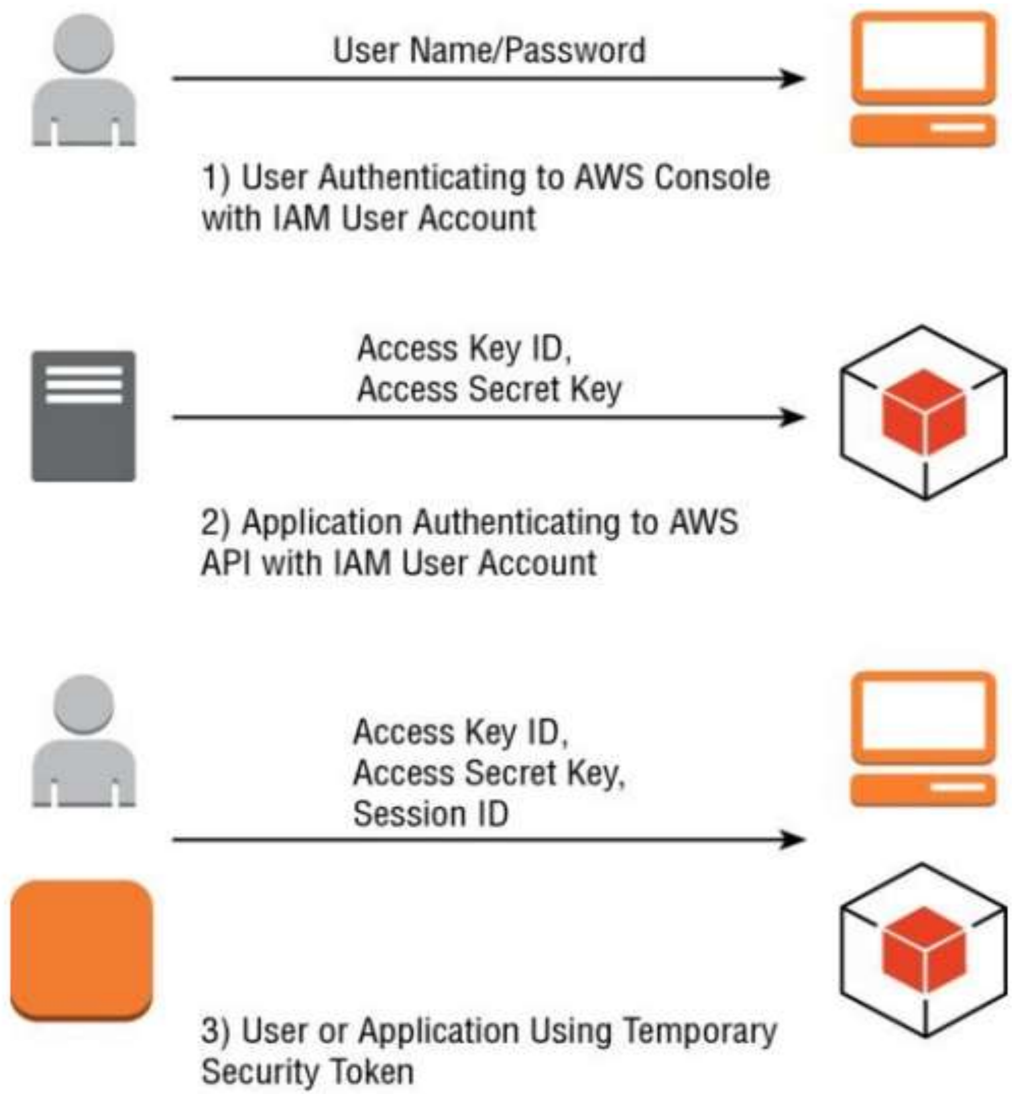


Amazon CloudWatch alarm triggering scaling out

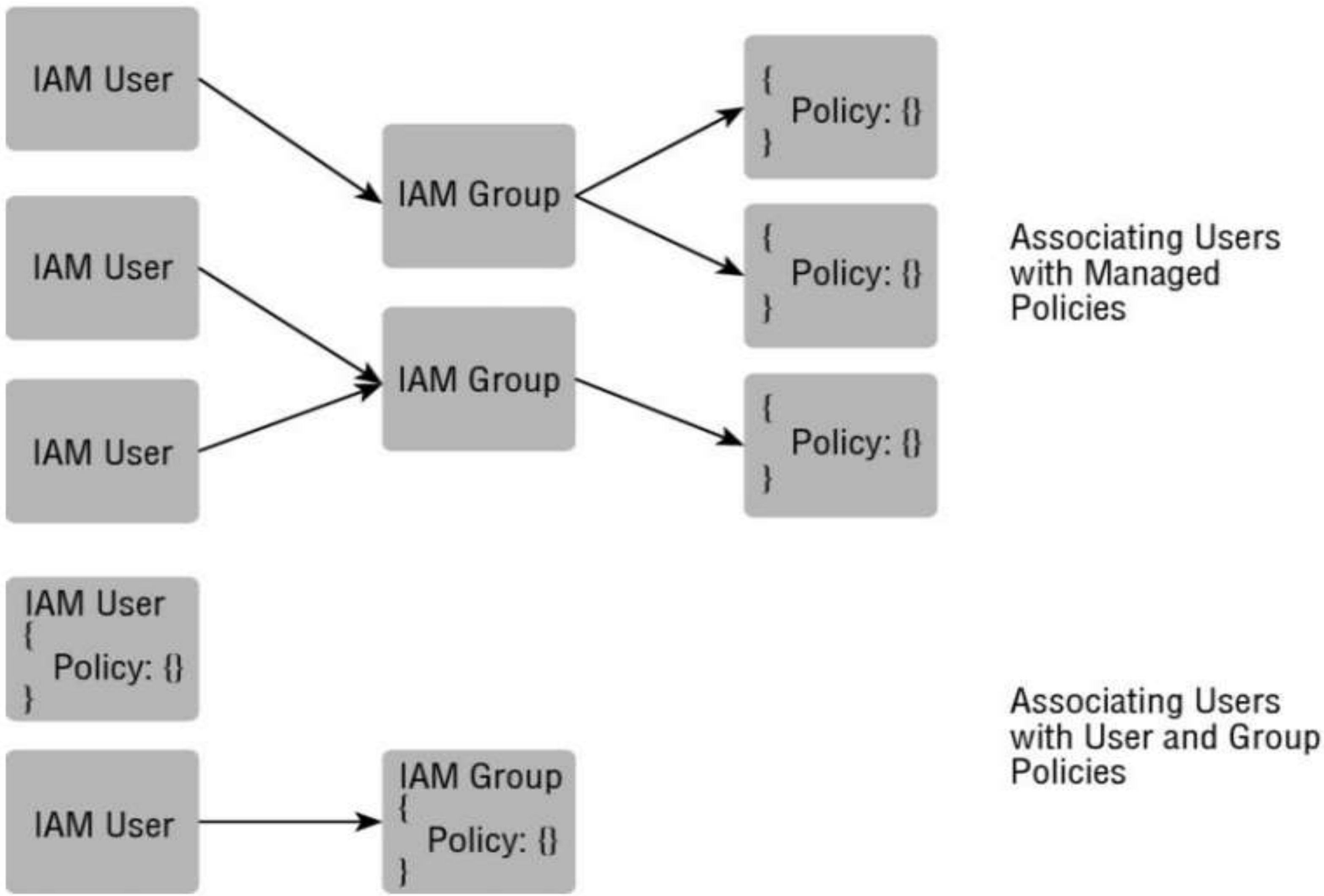
IAM

- Principals
  - Root User
    - Same as Unix root or Windows Administrator account
    - make in first time when user registration
    - Access and Control everything
  - IAM Users
    - Individual people or application
- Roles/Temporary Security Tokens
  - Roles : Specific user to specific requirements
  - Temporary security token's lifetime is 15MIN to 36H
  - Cross-account Access : grant with another AWS account
  - Federation : authenticated by external system.
    - ODIC, Active Directory, or LDAP
- Authentication
  - User Name/Password
  - Access key – Combination of an access key ID(20char) and access secret key (40char)
  - Access key/ Session Token – operates under an assumed role
- Authorization
  - Policies
    - Effect : Allow or Deny
    - Service : Service name
    - Resource : ARN
    - action and Condition
- Other Key Features
  - MFA Rotating Keys

IAM


User Name/Password
1) User Authenticating to AWS Console with IAM User Account
Access Key ID,
Access Secret Key
2) Application Authenticating to AWS API with IAM User Account
Access Key ID,
Access Secret Key,
Session ID
3) User or Application Using Temporary Security Token

IAM



Associating IAM users with policies

Databases and AWS

- Relational Databases
  - Amazon RDS
    - connected by endpoint (same as S3)
- Data Warehouses
  - Amazon RedShift
- NoSQL
  - Amazon DynamoDB

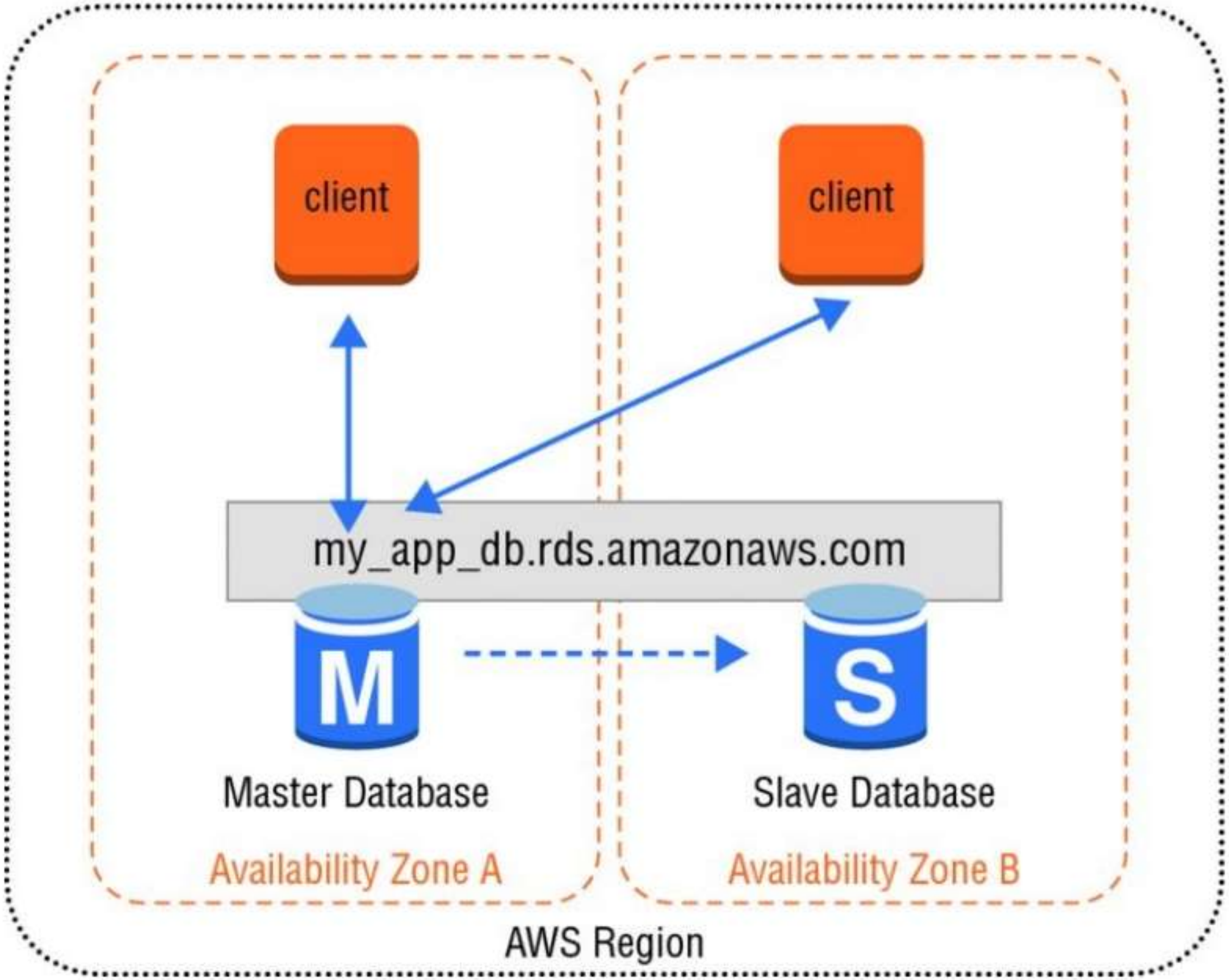
Amazon RDS

- Connected by hostname string(endpoint)
- Does not support console login
- Database (DB) Instances
  - Isolated database environment deployed in your private network segments in the cloud
- Benefits
  - safe, very consistent deployment and operational model.
  - no need to handle OS
- Engines
  - mysql, mariadb, postgresql, MSSQLServer, Oracle, Amazon Aurora
  - License model
    - BYOL : Oracle Standard Edition, Enterprise Edition, All of SQLServer and enterprise
    - Include : Oracle Standard Edition, SQLServer Express, Web, Standard
- Storage Options
  - Magnetic, General Purpose SSD, Provisioned SSD (Same as EBS)

Amazon RDS

- Backup and Recovery
  - Automated Backup
    - backup of DB Instance (not database)
    - By default, one day of backup (period up to max 35days)
  - Manual Snapshot
    - Manual DB snapshots are kept until explicitly delete
  - Restore
    - Create new DB instance from automatically backup or manual snapshot
- High Availability with Multi-AZ
  - using replication
  - Loss of Availability , network connectivity
  - Compute or storage failure
  - Same endpoint (DNS name) change (CNAME)

Amazon RDS



Multi-AZ Amazon RDS architecture

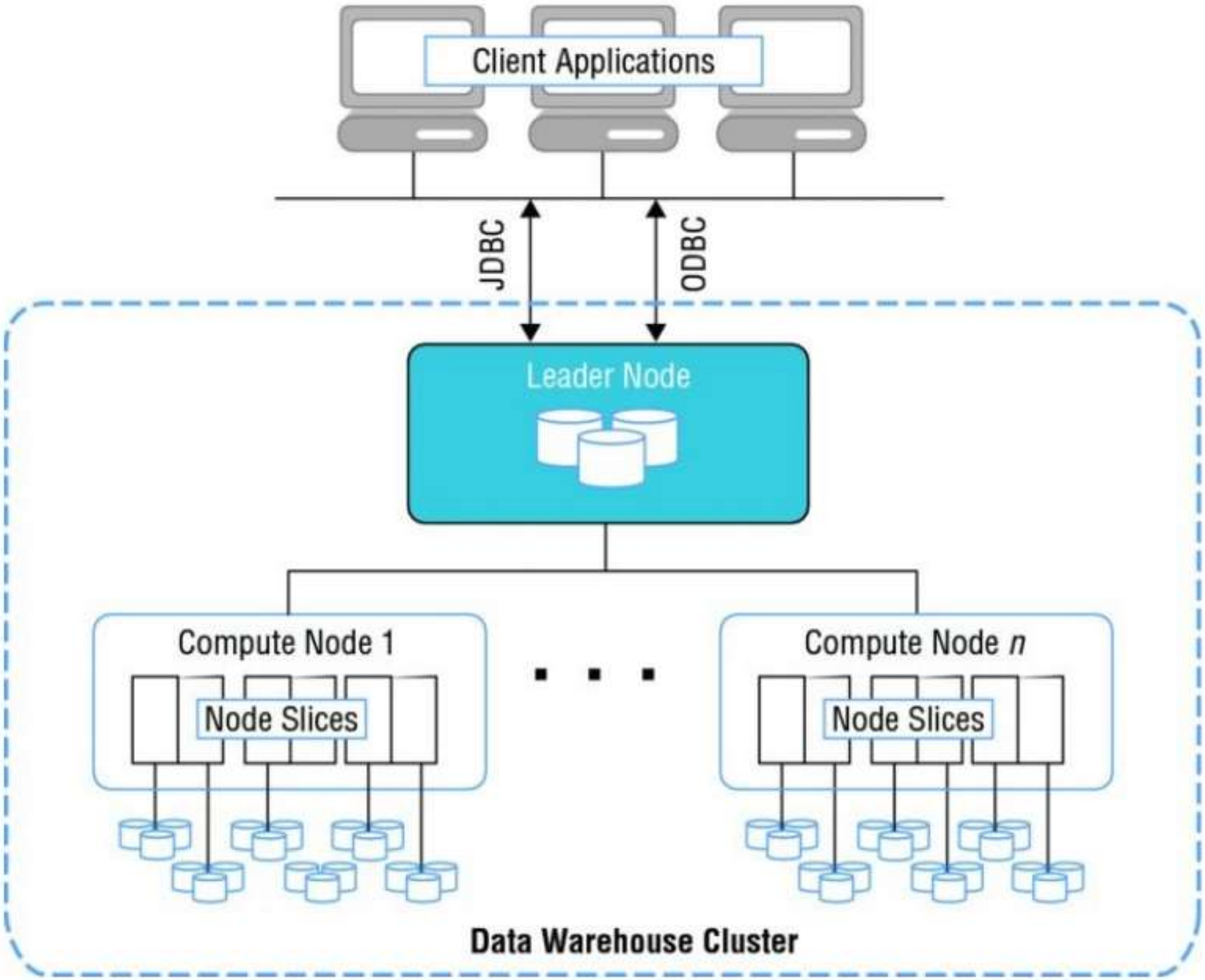
Amazon RDS

- Scaling Up and Out
  - Vertical Scalability
    - Next maintenance window or begin immediately
    - Select more higher DB instance class
    - automates the migration process
  - Horizontal Scalability with Partitioning
    - requires additional logic in the application layer.
    - NoSQL database are designed to scale horizontally
  - Horizontal Scalability with Read Replicas
    - Create a special type of DB Instance, called a read replica
    - cross-region read replicas to serve read traffic from a region closest
- Security
  - IAM policy
  - Deploy to VPC
  - Network ACLs within VPC
  - Using Security Groups for restrict inbound source IP
  - Encryption : SSL, TDE(Transparent Data Encryption), KMS

Amazon Redshift

- Designed for OLAP
- Based on postgresql
- Clusters and Nodes
  - The client working on leader node
  - Compute nodes are transparent to external application
  - Table Design
- Table Design
  - Data Types
  - Compression Encoding
  - Distribution Strategy
    - Even, Key, All
  - Sort key
    - compound, interleaved
- Loading Data
- Querying Data
- Snapshots
- Security

Amazon RDS



Client Applications
JDBC
ODBC
Leader Node
Compute Node 1
Node Slices
Compute Node n
Node Slices
Data Warehouse Cluster

Amazon DynanmoDB

- Full managed NoSQL Service
- Data Model
  - tables, item, attribute
  - Data Types
    - Scalar Data Types
      - String, Number, Binary, Boolean, Null
    - Set Data Types
      - String Set, Number Set, Binary Set
    - Document Data Types
      - List, Map
    - Primary key
      - Partition key , Partition and Sort key
- Secondary Indexes
  - Global Secondary Index – create/delete in at anytime
  - Local Secondary Index – only create at create table
- Writing and Reading Data
  - writing Items
    - create, update, delete
  - Reading Items
    - GetItem

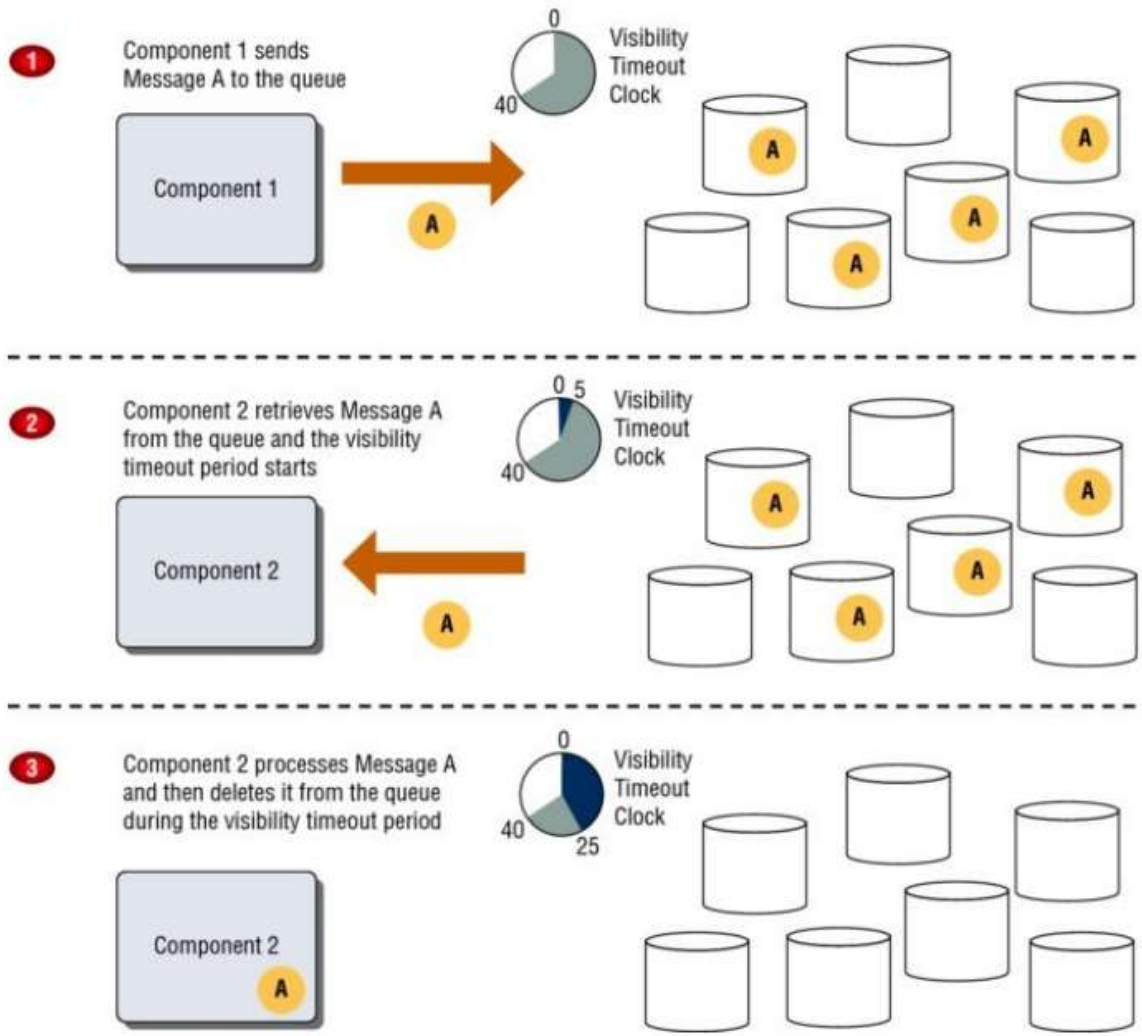
Amazon DynamoDB

- Eventual Consistency
  - Eventually Consistence Reads
    - fast, but might not reflect the result
  - Strongly Consistence Reads
    - can be slow, but exact
- Batch Operations
  - Searching Items
    - Query
    - Scan
- Scaling and Partitioning
  - Provisioned by user predict capacity
  - Can be split partition, but cannot merge again
- Security
  - Using IAM policies and role / user

Amazon Simple Queue Service

- SQS is buffer between application and backend
- Message delivery to multiple readers
- Does not guarantee FIFO

Amazon Simple Queue Service



Message Lifecycle

Amazon Simple Queue Service

- Valid properties are message ID and body
- retention period
  - default : 4days
  - max : 14days
- Delay Queues and Visibility Timeouts
  - Delay queue timeout is max 15Min
  - Visibility Timeout is up to 12hour
  - up to 120,000 messages in flight
- Queue Operations, Unique IDs, and Meta Data
  - Messages are identified via a globally unique ID
- Queue and Message Identifiers
  - queue URLs, Message IDs, receipt handles
  - Maximum length of message handle is 1,024 char

Amazon Simple Queue Service

- Message Attribute
  - Structured metadata items
    - timestamp
    - geographical data
    - Signatures
- Long Polling
  - in order to prevent looping for check message
  - WaitTimeSecond argument to ReceiveMessage up to 20 Sec.
- Dead Letter Queues
  - Move unsuccessfully processed message to another queue
  - sideline and isolate
- Access Control
  - Using IAM role
  - Grant another AWS account access to queue, (period) or deny

Amazon Simple Workflow Service

- Workflows
  - Distributed , asynchronous application
  - Run as asynchronously among in multiple devices
  - Sequential and parallel processing
- Actors
  - starter, Decider, Activity worker
- Tasks
  - Activity tasks, AWS lambda tasks, decision tasks
- Task Lists
  - Task associated with a workflow and routing task
- Long Polling
- Object Identifiers
  - workflow type , activity type => domain, name, version
  - Decision task, activity task => unique task token
  - workflow => domain, workflow ID, run ID (run ID is return value)
- Workflow Execution Closure
  - completed , canceled. failed, timed out
- Lifecycle of a Workflow Execution

Amazon Simple Notification Service

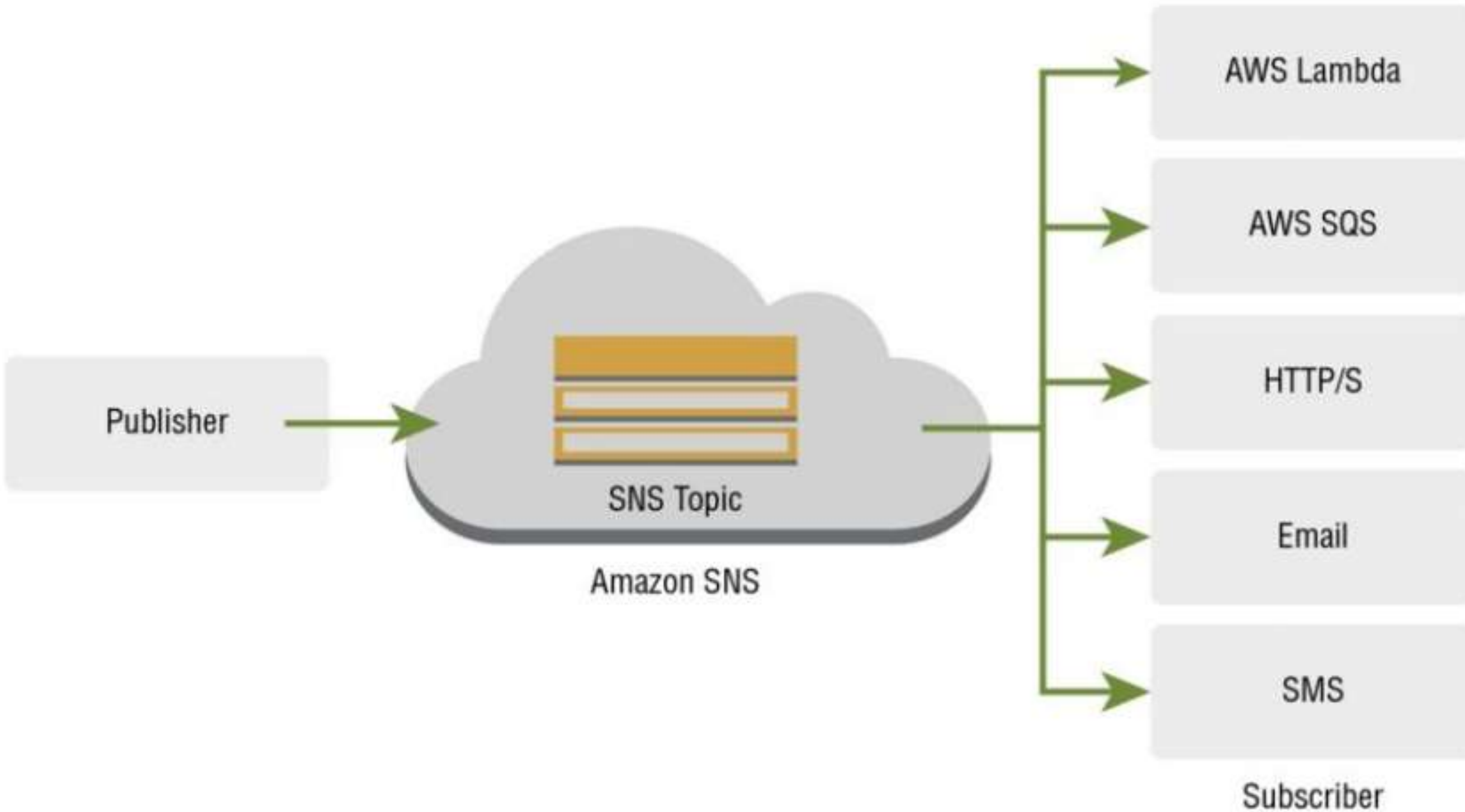


Diagram of topic delivery

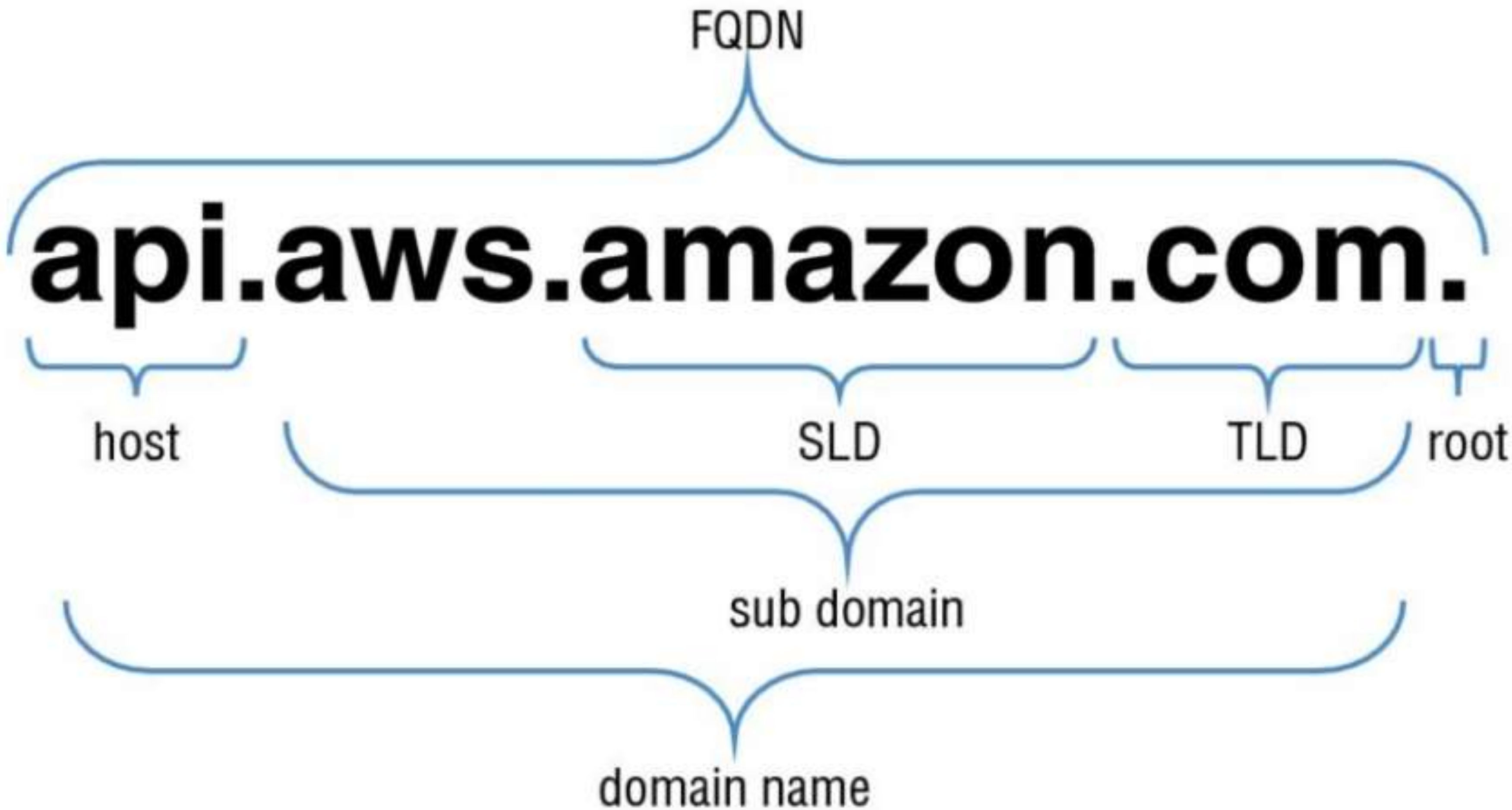
Amazon Simple Notification Service

- Mobile and enterprise messaging web service
- publish-subscribe(pub-sub) messaging paradigm
- send to AWS lambda, SQS, HTTP/S, Email, SMS
- Common Amazon SNS Scenarios
  - Fanout
    - send topic then and then replicated and pushed to multiple SQS queues, HTTP endpoint, or email
    - Allow parallel asynchronous processing
  - Application and System Alerts
  - Push Email and Text messaging
  - Mobile Push Notification

DNS and Route53

- Domain Name System(DNS)
  - TLDs
    - farthest portion to the right(as separated by a dot)
  - Domain Names
    - Human-friendly name
  - IP Addresses
    - Consist of four sets of numbers separated by dot
  - Hosts
    - refer to separate computers or service
  - Subdomains
    - method of subdividing the domain itself
  - Fully Qualified Domain Name(FQDN)

DNS and Route53



FQDN components

DNS and Route3

- Name Servers
  - Zone Files
    - Mapped plain text file between domain names and IP addresses
  - Top-Level Domain(TLD) Name Registrars
- Steps Involved in DNS(Server) Resolution
  - TLD Servers
  - Domain-Level Name Servers
  - Resolving Name Servers
  - More About Zone Files

DNS and Route 53

- Record Types
  - SOA (Start of Authority) Record
    - Mandatory in all zone file and identifies the base DNS (single)
  - A and AAA
    - map a host to IP Address (A : IPv4, AAAA: IPv6)
  - CNAME (Canonical Name)
    - Alias name to another name
  - MX (Mail Exchange)
    - define the mail server
  - NS (Name Server)
    - used by TLD server
  - PTR (Pointer)
    - reverse of an A record
  - SPF (Sender Policy Framework)
    - used by mail server to combat span
    - SPF record with the IP address of mail server
  - TXT (Text)
    - Description
  - SRV (Service)

Route53

- Domain Registration
- Domain Name System Service
  - Route 53 to route internet traffic to CloudFront, S3, ELB
- Hosted Zones
  - Private hosted zone -> Amazon VPC
  - Public hosted zone -> Internet facing
- Support Record type
  - A, AAAA, CNAME, MX, NS, PTR, SPF ,SRV, TXT, Routing Policies
    - Routing policies : weighted , latency-based, failover, geolocation

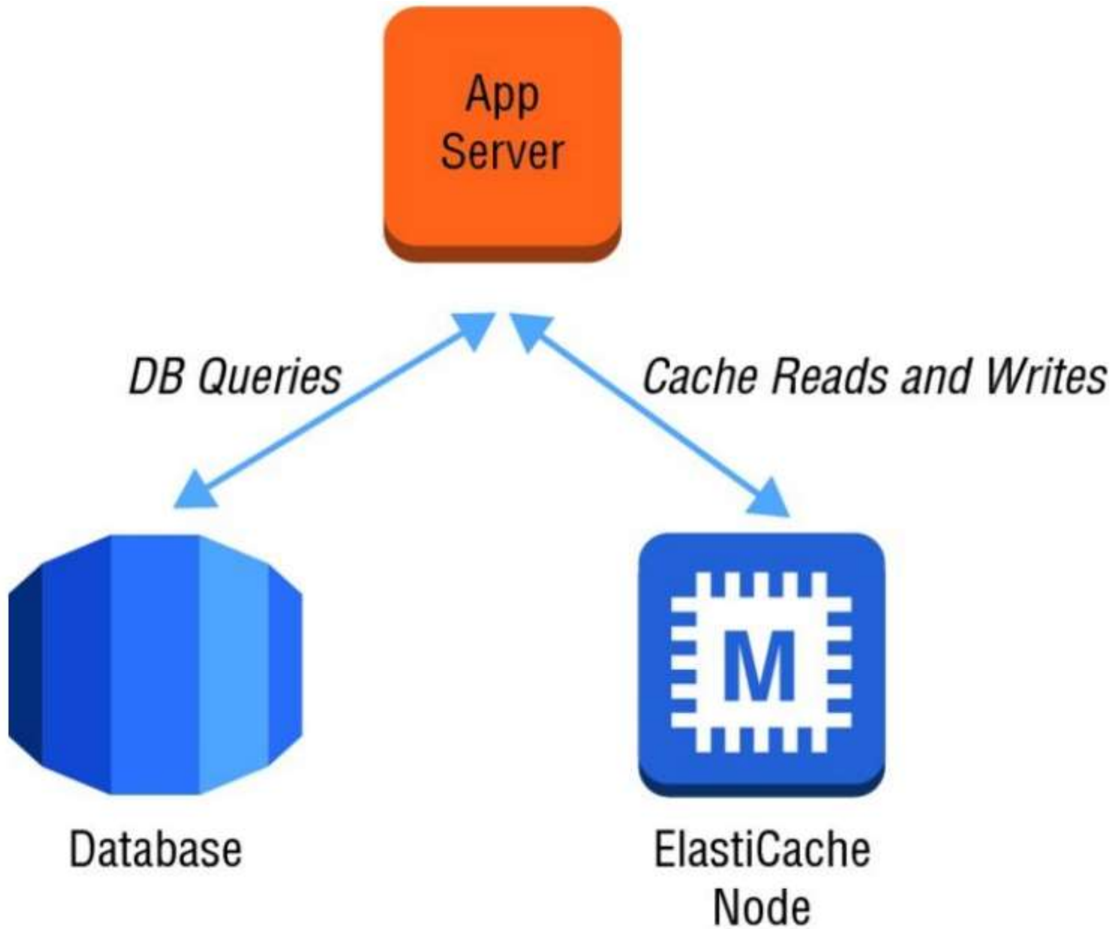
Route53

- Routing Policies
  - Weighted
    - Associated with single DNS and EC2 instances or ELB
  - Latency-based
  - Failover
    - Active – Passive failover (route53 automatically health check)
  - Geolocation

Amazon ElastiCache

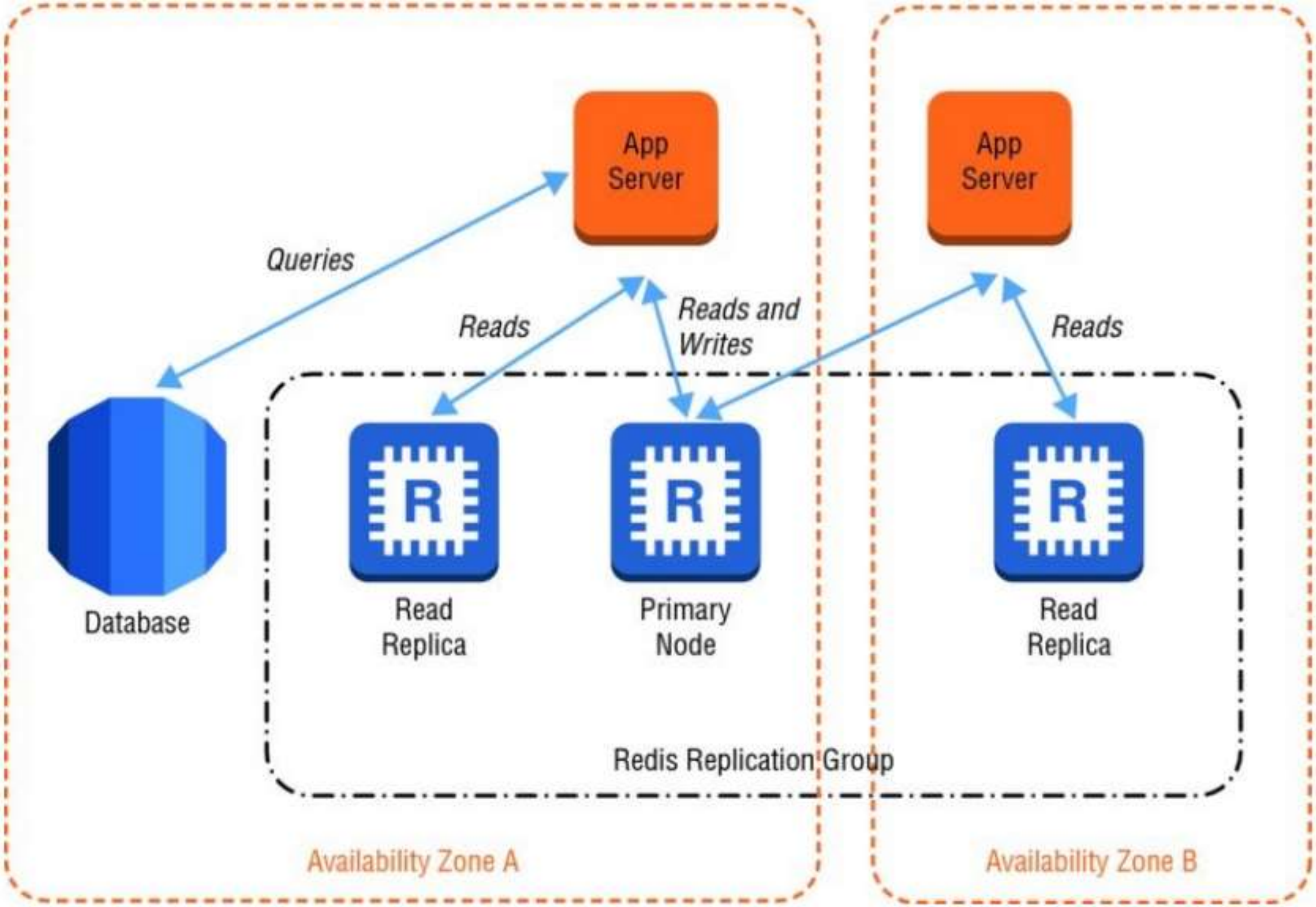
- Data Access Patterns
- Cache Engines
  - Memcached
  - Radis
- Nodes and Clusters
  - A single Memcached cluster can be have up to 20 nodes.
  - Redis is always single; multiple clusters can be grouped into replication group
- Memcached Auto Discovery
  - Auto discovery with the provided client library
- Scaling
  - Horizontal
  - Vertical
- Replication and Multi-AZ
- Multi-AZ Replication Groups
- Backup and Recovery
- Access Control

Amazon ElastiCache



Common caching architecture

Amazon ElastiCache



Redis replication group

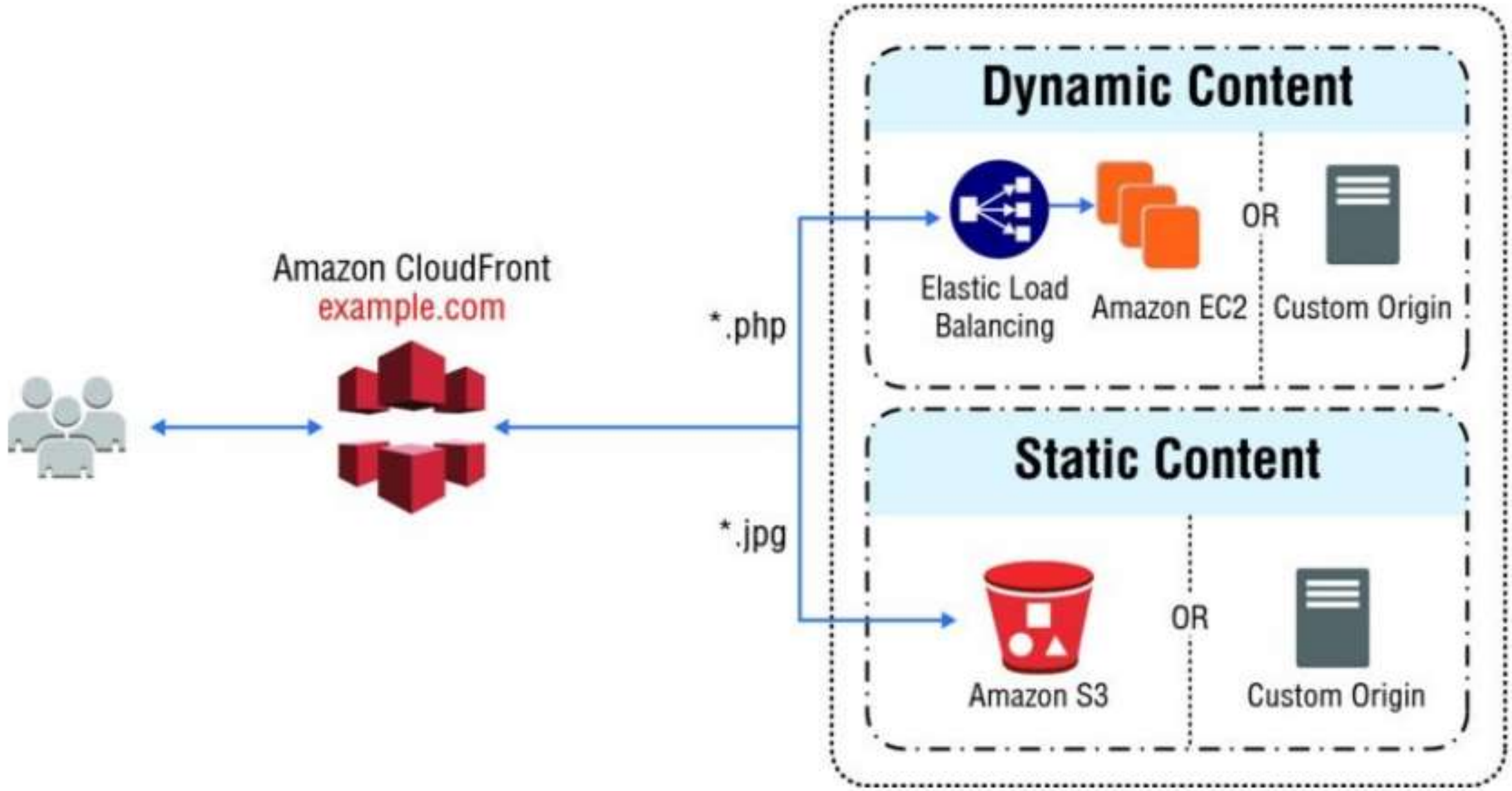
Additional key features

- Amazon CloudFront
- AWS Storage Gateway
- AWS Key Management Service (KMS) and AWS CloudHSM
- AWS CloudTrail
- Amazon Kinesis
- Amazon Elastic MapReduce(Amazon EMR)
- AWS Data Pipeline
- AWS Import/Export
- AWS OpsWorks
- AWS CloudFormation
- AWS Trusted Advisor
- AWS Config

Storage and Content Delivery

- Amazon CloudFront
  - CDN Service
  - Optimized to S3, S3 Static website, ELB, EC2
  - Integrated with on-premises server with Route53
  - Support HTTP/S, RTMP
  - cached objects are expire after 24H
  - Whole website
  - Private Content
    - Signed URL
    - Singed Cookie
    - Origin Access Identities(OAI)

Amazon CloudFront



Delivering static and dynamic content

AWS Storage Gateway

- Software appliance (VM Image)
- support iSCSI device as storage
- Server Side Encryption (SSE)
- Mount on-premises
- Gateway-Cached volumes
  - Expend storage capacity into S3
  - Maximum volume size : 32TB
  - Single gateway support 32Volumes (1PB)
- Gateway-Stored volumes
  - Asynchronously backup to S3 from on-premises
  - Backed up in EBS snapshot
  - Maximum volume size : 16TB
  - Single gateway support 32 Volumes (512TB)
- Gateway-Virtual Tape Libraries
  - up to 1,500 tapes (1PB)

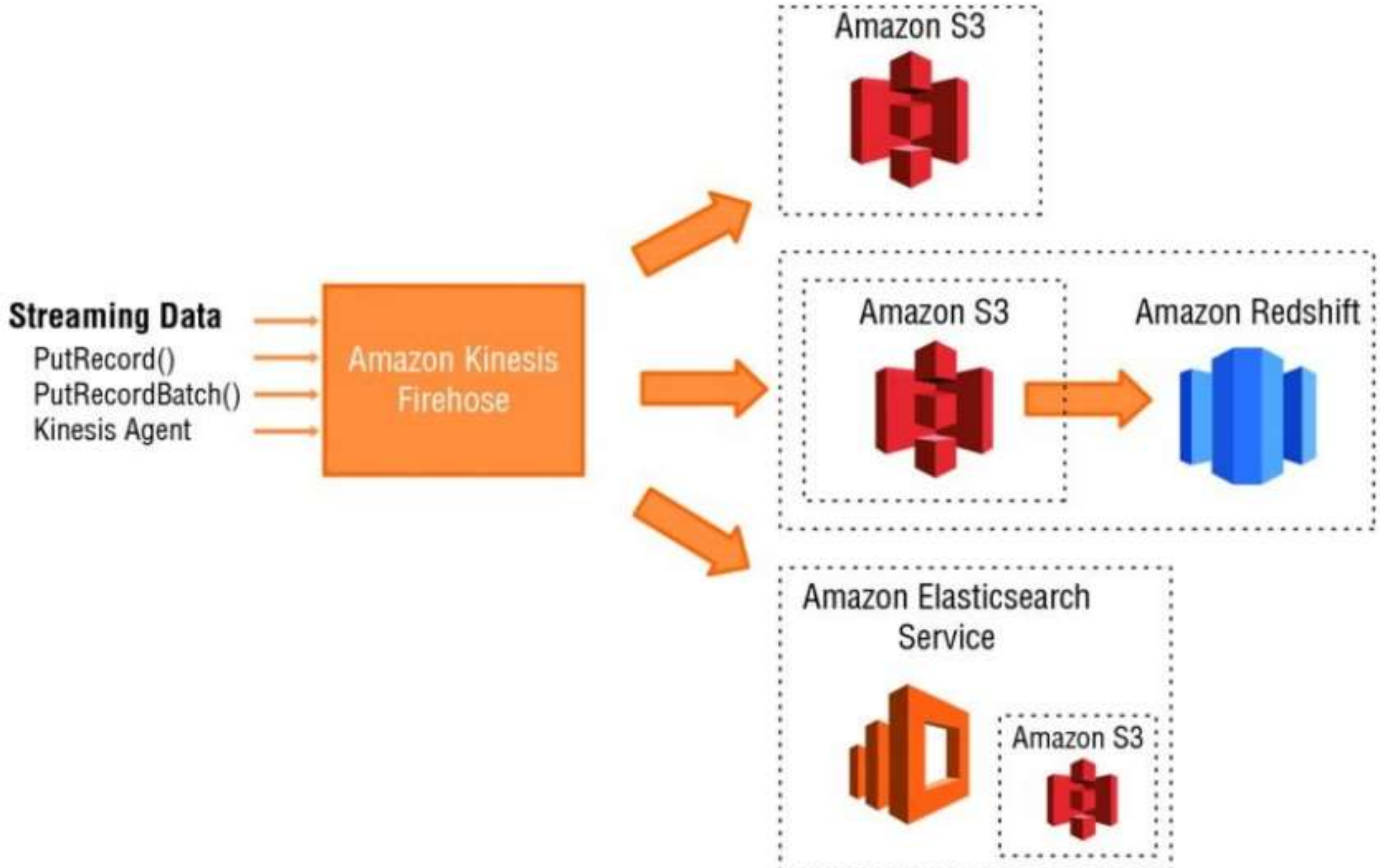
Security

- AWS Directory Service(Enterprise Edition)
  - Managed Microsoft Active Directory
    - Setup trust with existing AD
  - Simple AD
    - Provide by samba4
    - Daily automated snapshot, enable point-in-time recovery
  - AD Connector
    - Proxy connector for on-premises AD
- AWS Key Management Service(KMS)
  - generate, store, enable/disable, delete symmetric keys
  - Customer Managed Keys
  - Data Keys
  - Envelope Encryption
  - Encryption Context
- AWS CloudHSM
  - Cryptographic key storage by Hardware Security Module
- AWS Cloud Trail

Analytics

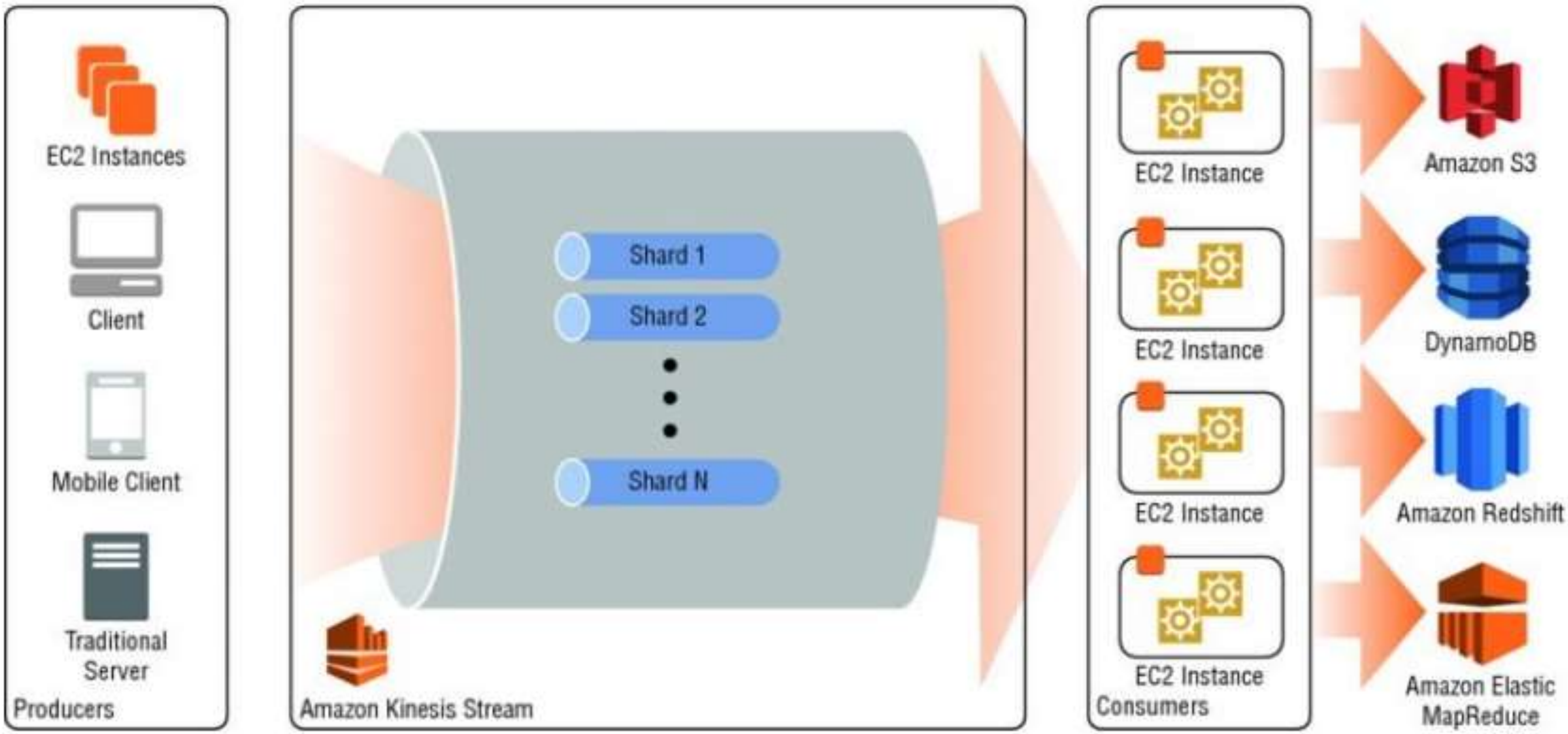
- Amazon Kinesis
  - Firehose
    - Receive streams save to S3, Redshift, Elastic Search
  - Streams
    - Collect and process large streams of data recode
    - Data Ingestion
    - Real-Time Processing
  - Analytics

Amazon Kinesis



Amazon Kinesis Firehose

Amazon Kinesis



Amazon Kinesis Streams

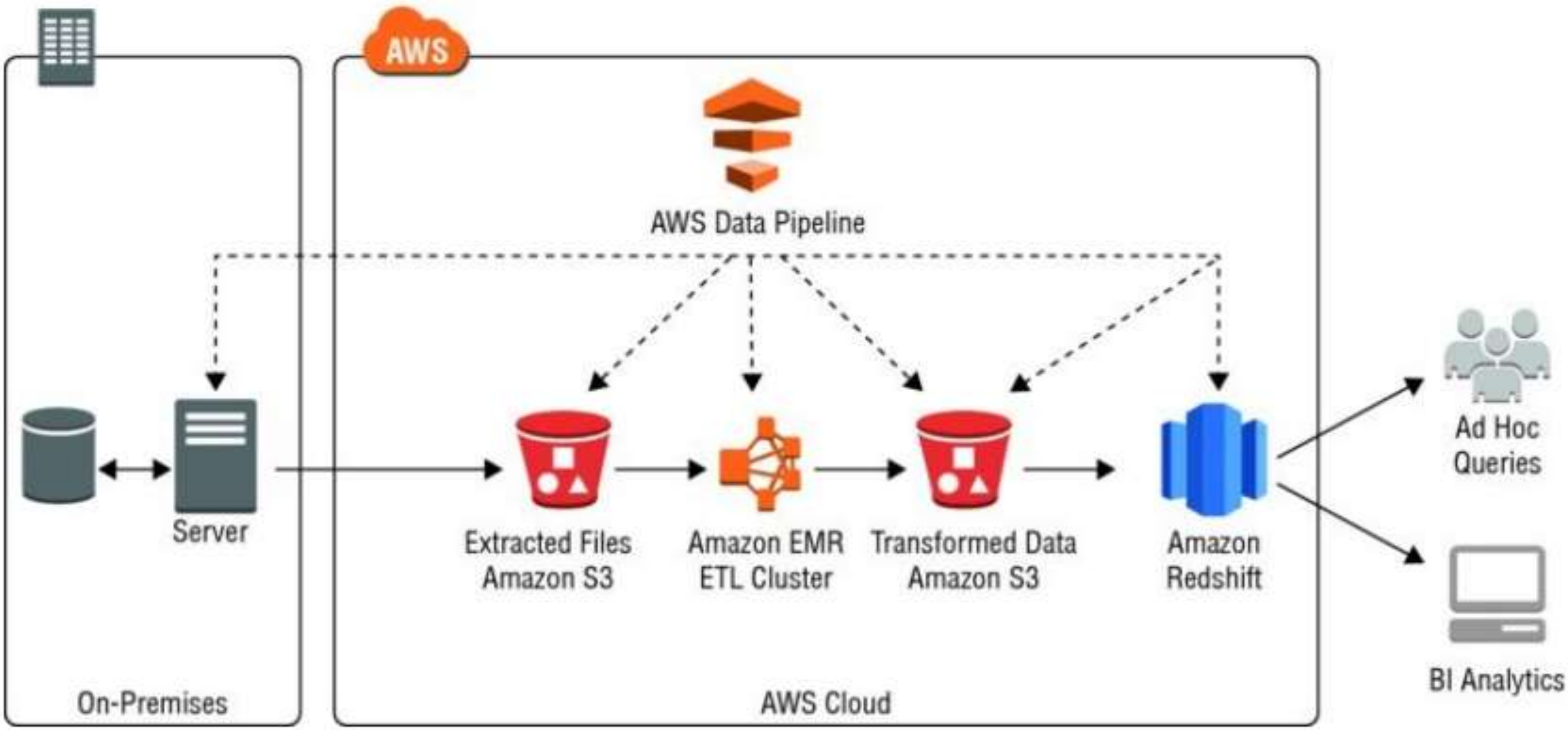
Amazon Elastic MapReduce(EMR)

- HDFS
  - Instance storage (not persist)
  - EBS for HDFS (persist)
- EMR File System (EMRFS)
  - Storing data to S3
  - Suite for transient clusters

Amazon Data Pipeline

- Move data between AWS compute and storage (on-premises)
- Read/Write from S3, MySQL, RedShift
- Used for virtual ETL process

Amazon Data Pipeline



Example pipeline

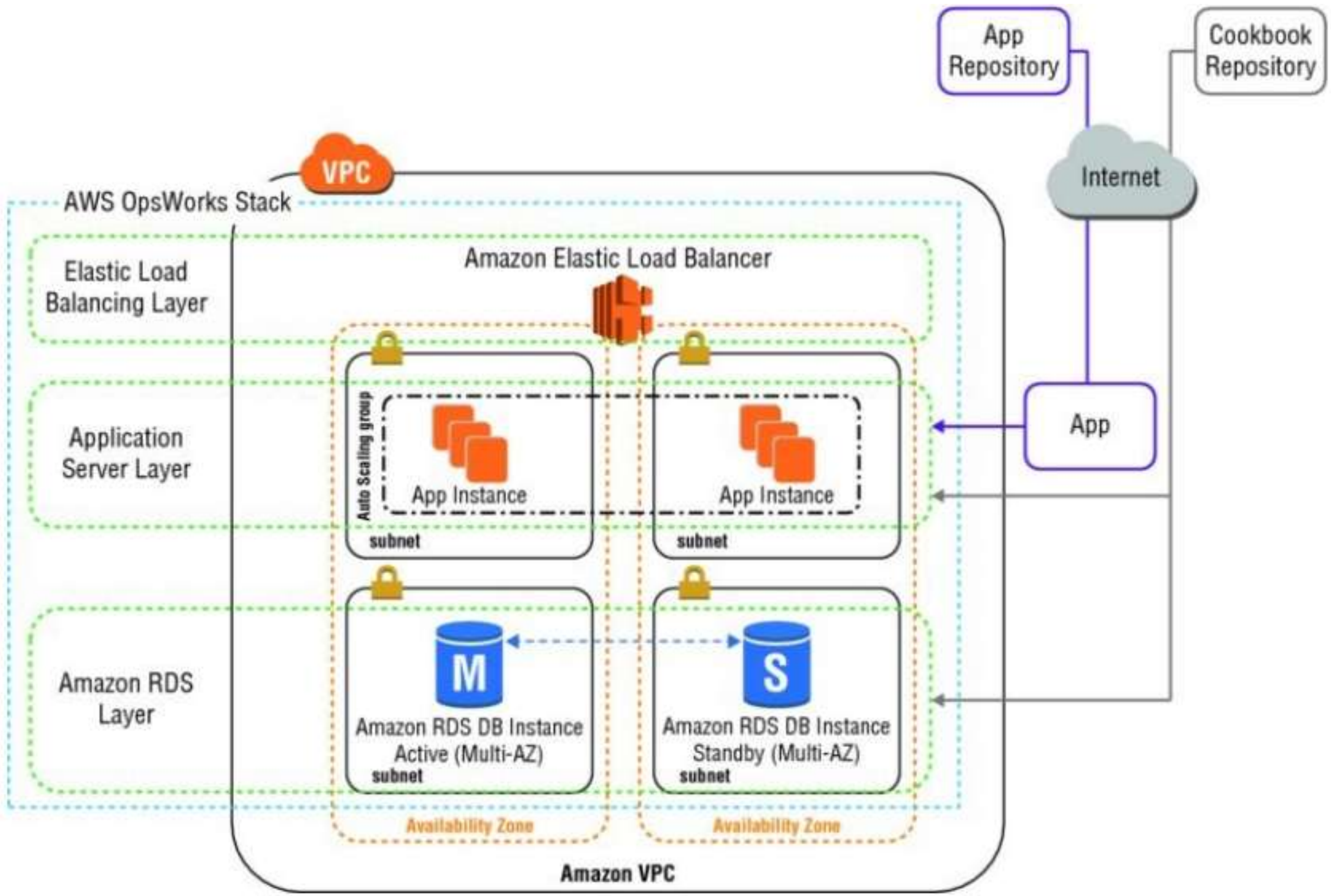
AWS Import/Export

- AWS Snowball
  - shippable storage appliances
  - 50TB and 80TB
  - Encryption is enforced
- AWS Import/Export Disk
  - Using Amazon Internal network
  - Import data into Glacier, EBS, S3
  - Export data from S3
  - Optional encryption
  - 16TB limit

DevOps

- AWS OpsWorks
  - Configuration management service (using Chef)
  - Layers depend on Chef recipes to handle task
  - Set of lifecycle (automatically run receipts)
  - Sends all of resource metric to CloudWatch

DevOps



Simple application server stack with AWS OpsWorks

AWS CloudFormation

- Model and set up your AWS resources
- Use Case
  - Quickly launch new test environments
  - Reliably Replicate Configuration Between Environments
  - Launch Application in New AWS Region

AWS Elastic Beanstalk

- PasS
- Managed all environments for application
- Automatically handled the deployments
  - Capacity Provisioning
  - Load Balancing
  - Auto Scaling
  - Monitoring

AWS Trusted Advisor

- inspects environments and make recommendation
- categories
  - Cost optimization
  - Security
  - Fault Tolerance
  - Performance improvement
- Server Limits
- Security Groups-Specific Ports Unrestricted
- IAM Use
- MFA on Root Account

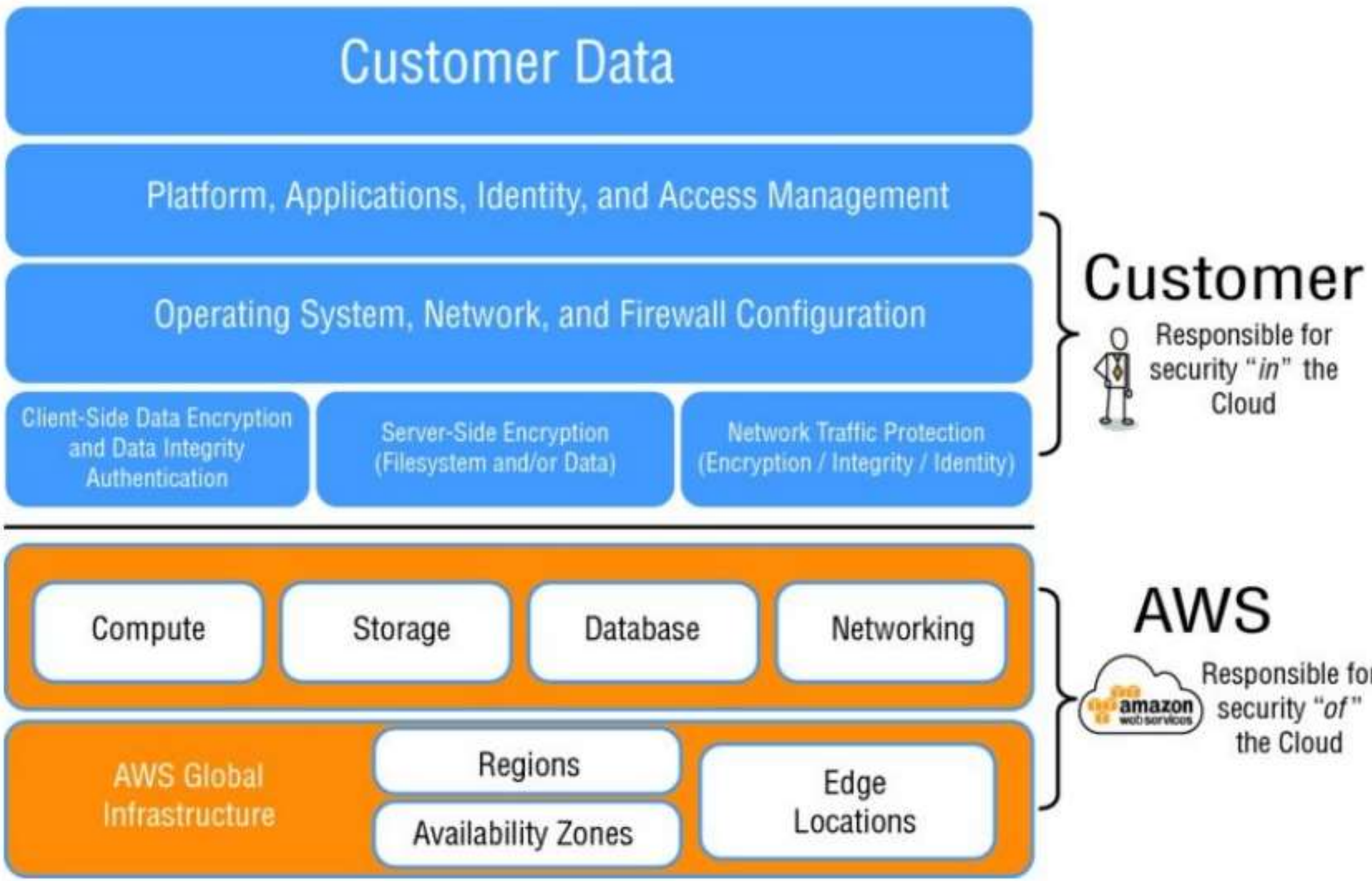
AWS Config

- Use Cases
  - Discovery
    - record current configuration
  - Change Management
    - notified all configuration change
  - Continuous Audit and Compliance
    - visibility configuration
    - evaluating relevant configuration
  - Troubleshooting
  - Security and Incident Analysis
    - Examine the configuration
- Key Features

Security for AWS

- Shared Responsibility Model
- Compliance Program
- AWS Global Infrastructure Security
- AWS Account Security
- AWS Cloud Service Specific Security

Security for AWS



Shared Responsibility Model

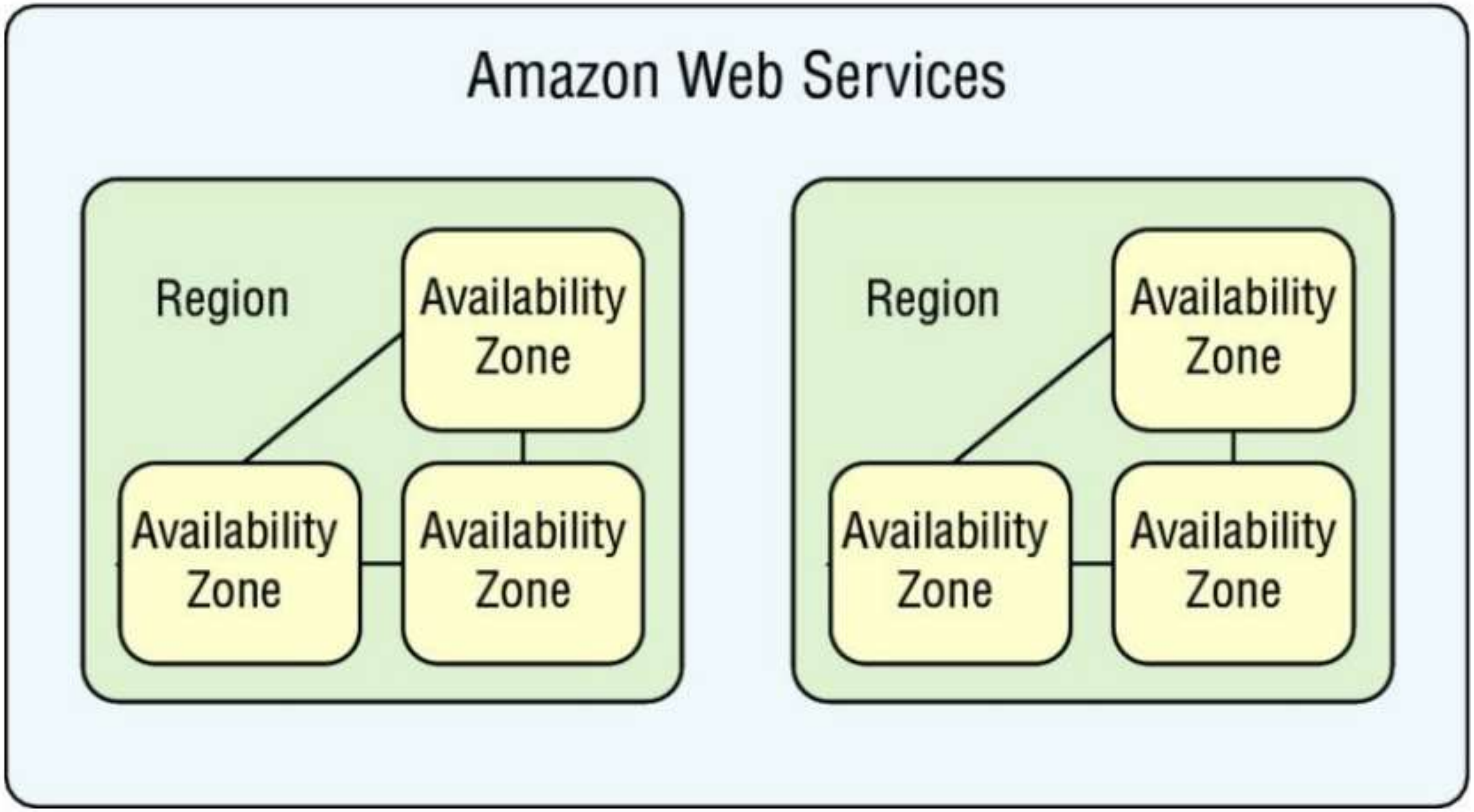
AWS Compliance Program

- SOC 1,2,3
- SSAE 16
- ISAE 3024
- SAS 70
- PCI SDD level1
- ISO 27001
- CJIS
- CSA
- PERPA
- HIPAA
- MPAA
- FISMA
- FedRAMP

AWS Global Infrastructure Security

- Physical and Environmental Security
  - Fire Detection and suppression
  - Power
  - Climate and Temperature
  - Management (monitoring)
  - Storage device decommissioning
- Business Continuity Management
  - Availability
  - Incident Response
  - Communication
- Network Security
- Network Monitoring and Protection
  - DDoS
  - Man in the middle (MITM)
  - IP Spoofing
  - Port Scanning

Amazon web service regions



AWS Account Security Features

- AWS Credentials

type	Use	Description
Password	AWS root account or IAM user account login	String character
Multi Factor Authentication (MFA)	AWS root account or IAM user account login	A six-digit, single-use code
Access keys	SDK, CLI, REST/Query API	Access key ID + secret access key (15 minute)
Key Pairs	SSH login CloudFront-signed URL	
X.509 Certification	SOAP request to APIs SSL Server for HTTPS	

AWS CloudTrail

- The name of the API
- The identity of the caller
- The time of the API call
- The request parameters
- The response elements returned by the AWS Cloud service

AWS Cloud Service-Specific Security

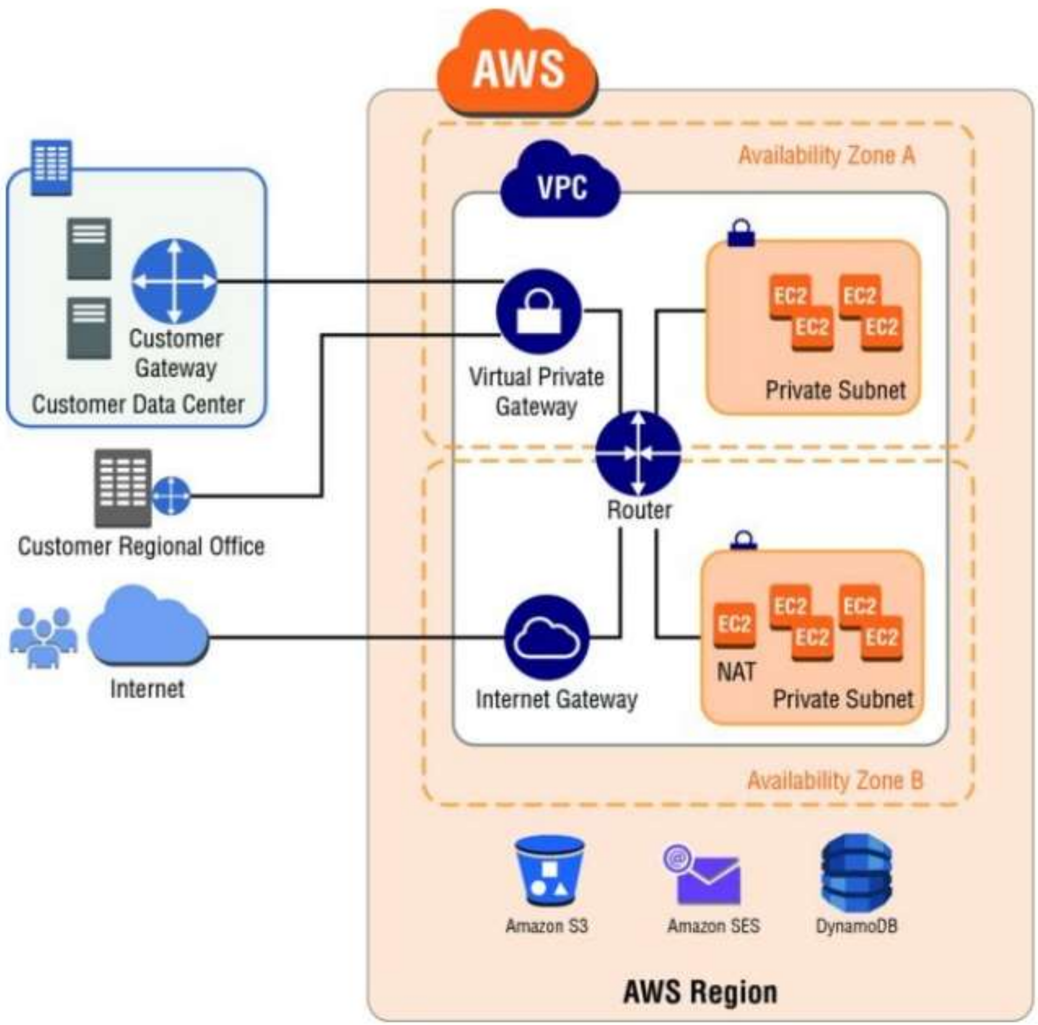
- Amazon Elastic Compute Cloud(EC2) Security
- Multiple Levels of Security
  - The Hypervisor
  - Instance Isolation
  - Host Operating System
  - Guest Operating System
- Firewall
  - By default, all inbound traffic is deny
- API Access
  - Using Signed secret access key
- Amazon EBS Security
  - replication is stored within same AZ
  - No automatic backup
  - share volume : create new volume -> restore from snapshot-> share new volume
  - Encrypt EBS and snapshot with AES-256

AWS Cloud-Service Specific Security

- Elastic Load Balancing Security
- Amazon Virtual Private Cloud Security
  - Security Groups
  - Network ACLs
  - Routing Tables
  - External gateway
- API Access
- Subnets and Route Tables
  - Protect network from MAC and ARP spoofing
- Firewall(Security Groups)

Isolate network

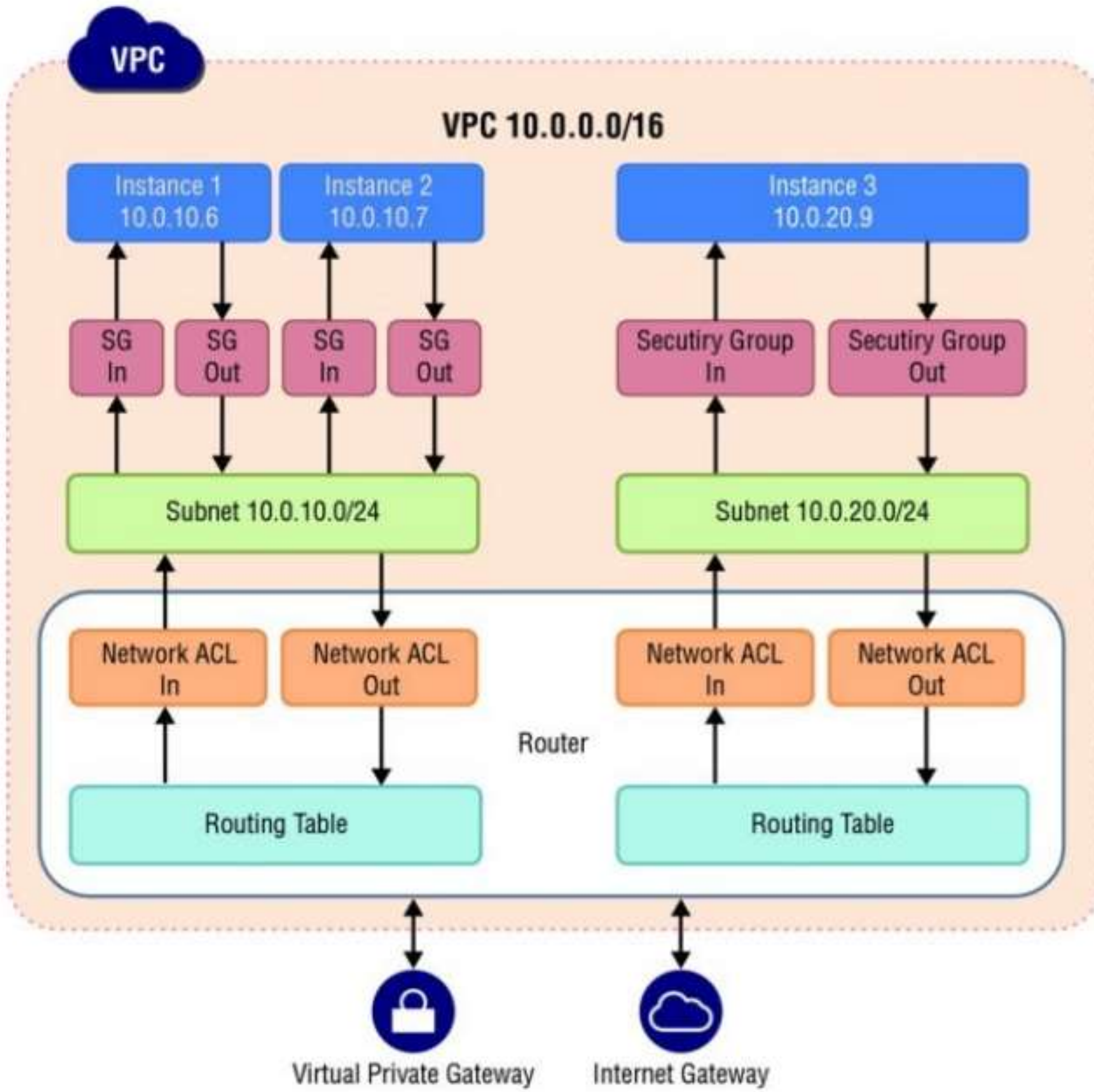
VPC network architecture



AWS Cloud-Service Specific Security

- Network ACLs
  - Stateless filter for subnet within VPC
  - Allow or Deny (protocol, port, src/dest IP addr
- Virtual Private Gateway
- Internet Gateway
- Dedicated Instances
- Amazon Clouldfront
  - requires authenticated / authorized user
  - only accessible via SSL-enabled endpoint

Flexible network architecture



AWS Cloud-Service Specific Security

- Amazon Simple Storage Service Security
- Data Access => By Default, only owner can access
- IAM Polices => assign to user / role
- ACLs => grant permission of bucket or object another AWS account (not specific users)
- Bucket Policies => allow / deny permission within single bucket
- Query String Authentication => date , Boolean, IP Address, String Conditions
- Data Transfer => SSL
- Data Storage => SSE(AES-265) or Client encryption before upload
- Access Logs =>
- Cross-Origin Resource Sharing
- Glacier Security
- Data Transfer
- Data Retrieval
- Data Storage => AES-256
- Data Access

AWS Storage Gateway Security

- Data Transfer => SSL (Asynchronously)
- Data Storage => symmetric key 256bit (encrypted)

AWS Cloud-Service Specific Security

- Amazon DynamoDB Security
  - IAM policy can restrict access to individual items in a table, attributes
- Amazon Relational Database Service (Security)
  - Access Control
  - Network Isolation => Private subnet
  - Encryption
    - SSL Connection
    - TDE for SQL Server
    - Oracle native network encryption
  - Automated Backups and DB Snapshots
    - Retention period : 5Min ~ 35Days
    - DB Instance snapshot
  - DB Instance Replication
    - Multi-AZ deployment
    - Automatically failover
  - Automatic Software Patching

AWS Cloud-Service Specific Security

- Amazon Redshift Security
  - Cluster Access
  - Data Backups
    - User-defined-period (1 ~ 35 days)
  - Data Encryption
    - four-tier, key based architecture
    - master -> cluster -> database -> data encryption key
    - S3 (SSE) encryption
  - Data Audit Logging
  - Automatic Software Patching
  - SSL Connections

Application Service

- Amazon Simple Queue Service
  - Granted based on an AWS Account or user create with AWS IAM
  - SSL-Encrypted endpoint
  - User can encrypt data before upload
- Amazon Simple Notification Service

Application Service

- Amazon EMR
  - master : allow SSH , communicate with slave
  - slave : only allow communicate with master
- Amazon Kinesis
  - SSL-encrypted endpoint
  - IAM role
  - AWS credentials

Application Service

- AWS Identity and Access Management(IAM)
  - Role
    - temporary security credentials
    - 12H is default expiration
- Federated users
  - assign permission to MS AD, LDAP, Kerberos
- SAML 2.0
  - Facebook, Amazon, Google
- Cross-Account Access
  - Using roles help to ensure (provided temporary and only as needed)
- Applications Running on EC2 Instances That Need to access AWS Resource
  - it need security credentials
  - consists with Access Key ID and Secret Access Key

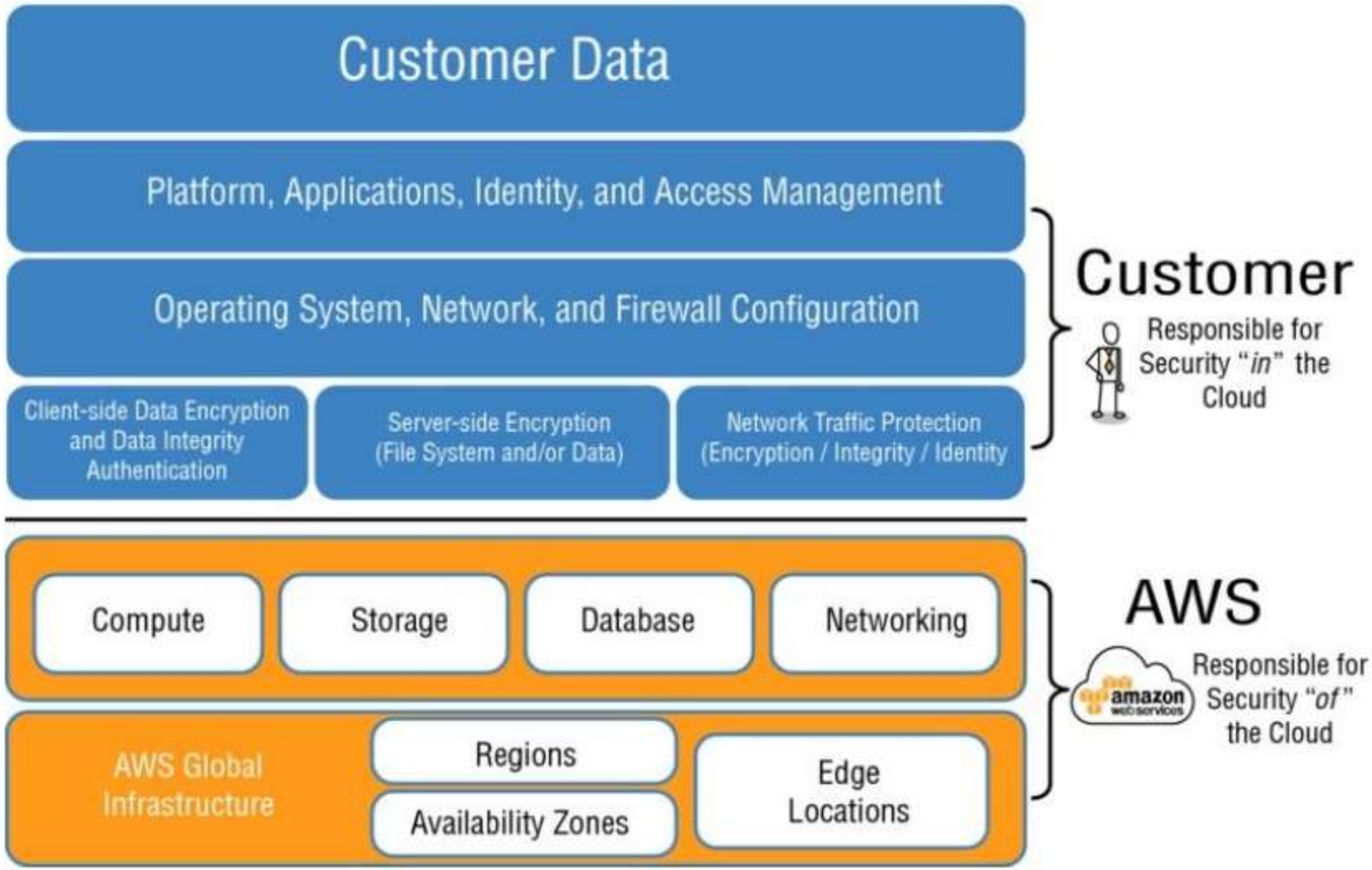
Amazon Cognito Security

- Identity and sync service for mobile/web base application
- cannot be reused after expire
- no need to create individual AWS accout
- Store data locally

Applications

- Amazon WorkSpaces Security
  - PC-over-IP(PCoIP)
  - MFA upon Sign-in
  - RADIUS authentication
    - PAP
    - CHAP
    - MS-CHAP 1, 2
  - Provided EBS and automatically backed up twice a day on S3
  - workspace can sync with your MAC or PC

AWS Risk and Compliance



Shared Responsibility Model

AWS Risk and Compliance

- Shared Responsibility Model
- Strong Compliance Governance
  - Document all compliance requirements(AWS)
  - Compliance requirements(organization)
  - Identify and document control(3rd party)
  - Verify all control objective
- Evaluating and Integrating AWS Controls
  - Specific Control Definition
  - General Control Standard Compliance
- AWS Global Regions
  - Consists by multiple Availability Zone

AWS Risk and Compliance

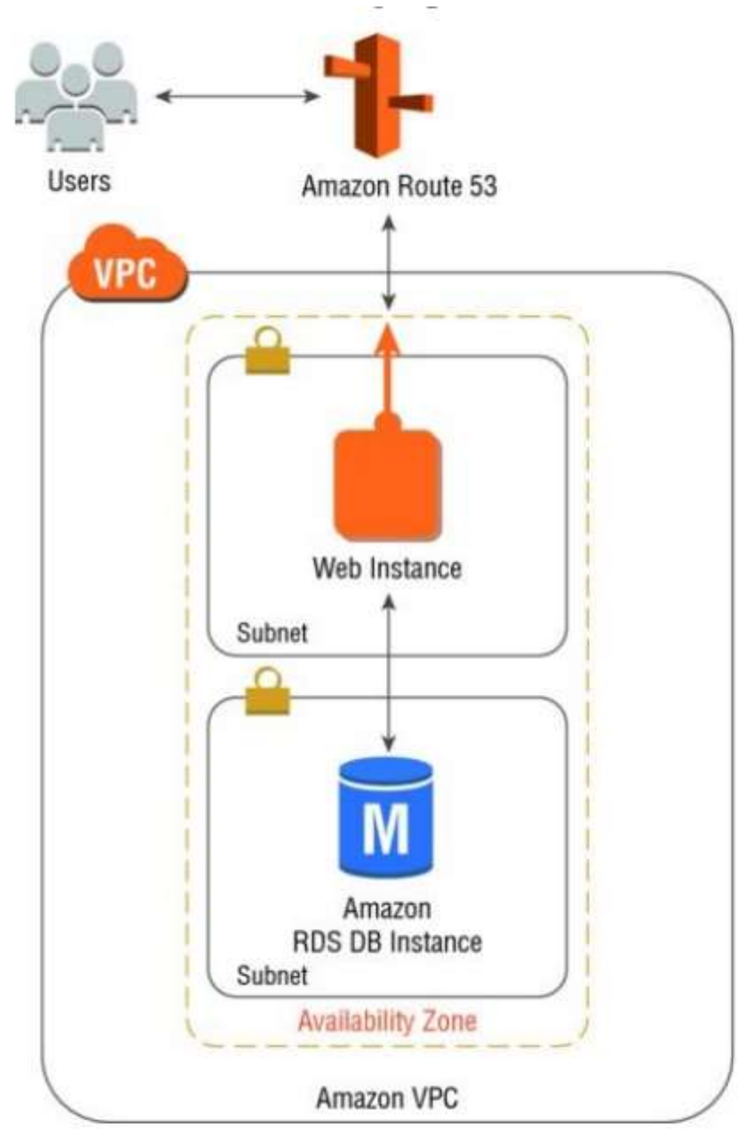
- Risk Management
- Revaluates the business risk plan at least twice a year
- Customer can request to vulnerability scans on own environments.
- AWS control environment is People, Technology, Processes

Best Practices

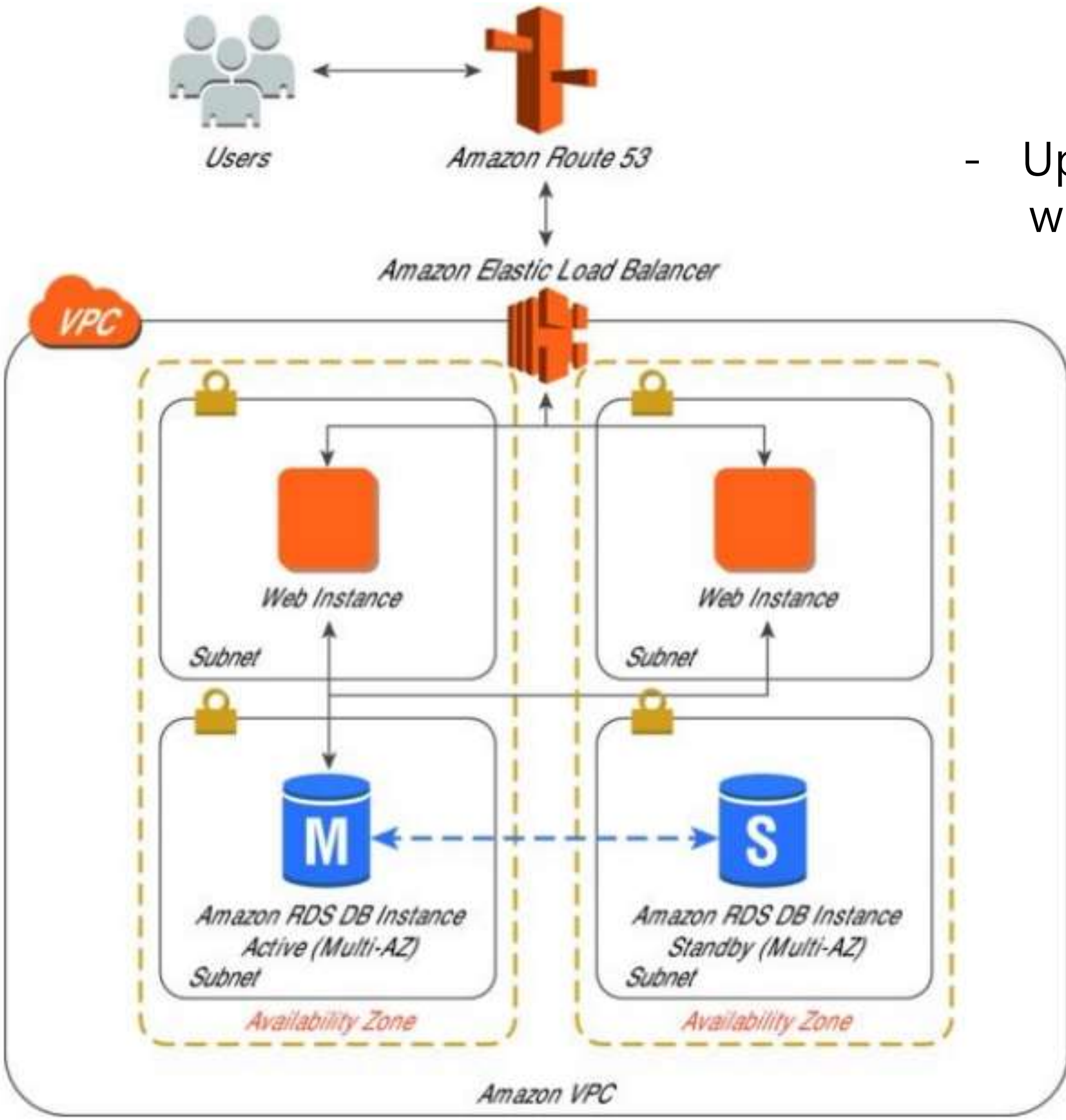
- Design for failure and nothing will fail
- Implement elasticity
- Leverage different storage options
- Builds security in every layer
- Think parallel
- Loose coupling sets you free
- Don’t fear constraints

Best Practices

- No redundancy
- No Failover
- Has Single point of failure (SPOF)



Best Practices

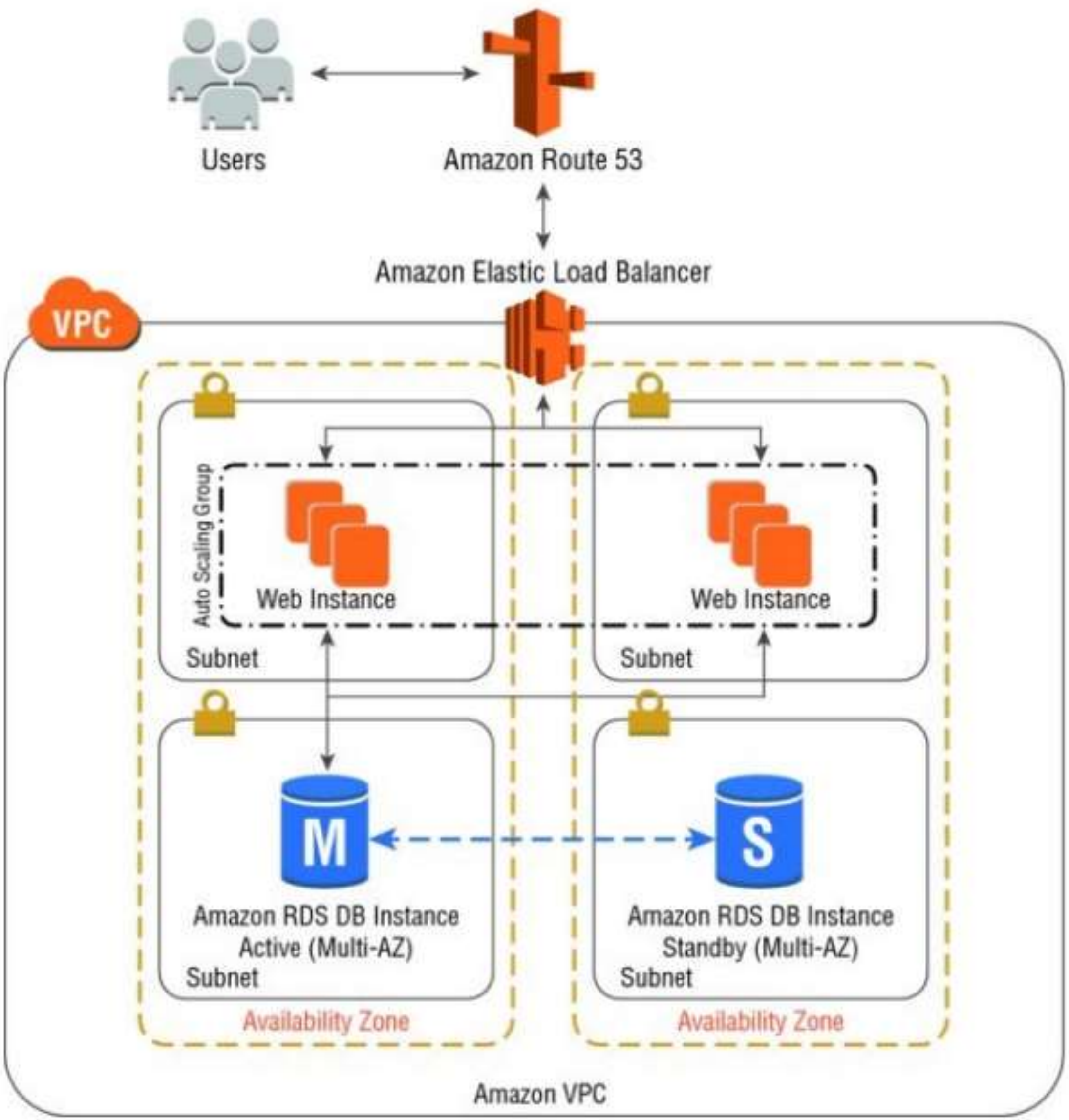


- Updated web application architecture with redundancy

Best Practices

- Implement Elasticity
- Scaling Vertically
  - Increase in the specification of an individual resource
- Scaling Horizontally
  - Increase number of resources
  - Stateless Application
  - Stateless Components
  - Stateful Components
- Deployment Automation
  - Automate Your Infrastructure
  - Bootstrap Your Infrastructure

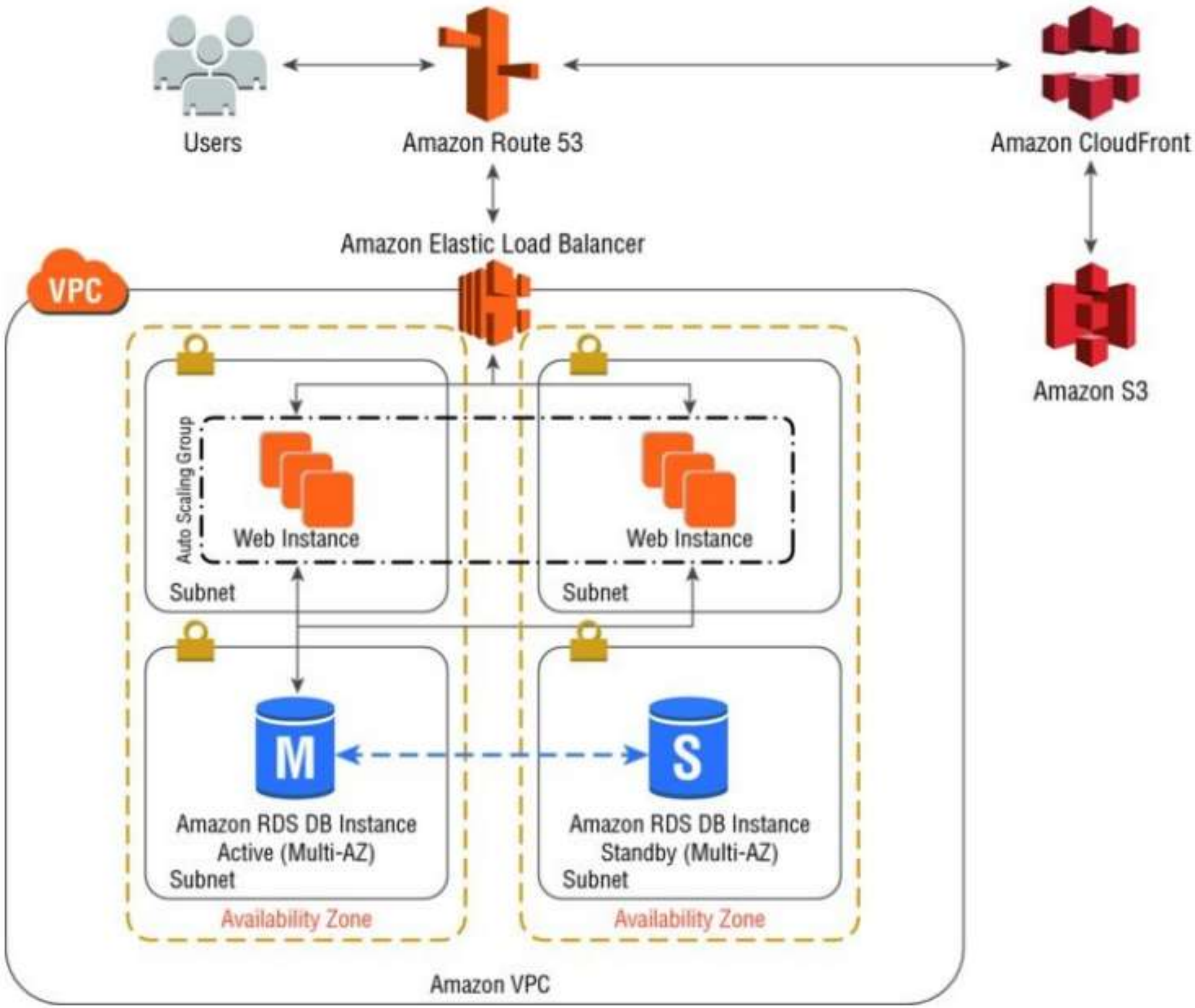
web application architecture with auto scaling



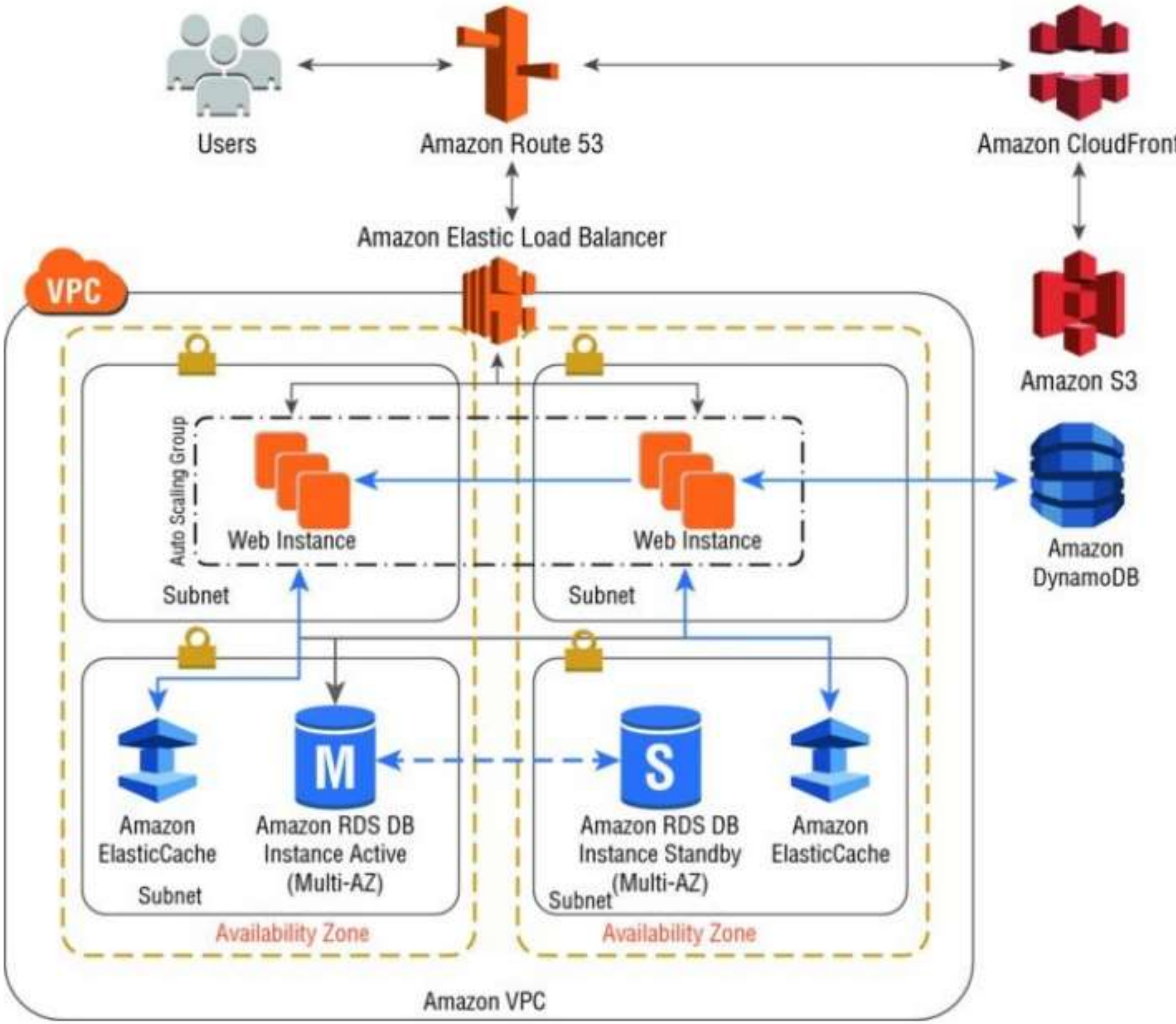
Storage Scenarios and AWS Storage Options

Sample Scenarios	Storage Option
Web application need Large-scale storage capacity and performance	S3
-or-	
Need cloud storage with high data durability to support backup and active archives for disaster recovery	
Require cloud storage for data archiving and long-term backup	Glaicer
Require a CDN to deliver entire website, (inc Dynamic , static, streaming, interactive content)	CloudFront
Fast and flexible NoSQL	DynamoDB
Reliable block storage, run mission-critical application	EBS
RDMS	RDS
OLAP	Redshift
Redis cluster or memcached cluster	ElastiCache
Shared between more than one EC2 instnaces	Elastic File system

Web Application architecture with S3 and CloudFront



Web Application architecture with ElastiCache and DynamoDB



Build Security in Every Layer

- Use AWS Features for Defense in Depth
  - Network Level, VPC topology, subnet, Security groups, routing table, WAF(web application firewall) , access control using IAM
- Offload Security Responsibility to AWS
  - Reduce Privileged Access
    - Using IAM roles to grant permissions
  - Security as Code
    - Using CloudFormation for Golden Environment
  - Real-time Auditing