Section 1 Intro

What is Cloud Computing?

• Cloud computing is the on-demand delivery of compute power, database storage, applications, and other IT resources

• Through a cloud services platform with pay-as-you-go pricing

• You can provision exactly the right type and size of computing resources you need

• You can access as many resources as you need, almost instantly

• Simple way to access servers, storage, databases and a set of application services

The Deployment Models of the Cloud?

- Private Cloud: Cloud services used by a single organization, not exposed to the public
- Public Cloud: Cloud resources owned and operated by a thirdparty cloud service provider delivered over the Internet
- Hybrid Cloud: Keep some servers on premises and extend some capabilities to the Cloud.

The Five Characteristics of Cloud Computing?

• On-demand self service

• Broad network access

• Multi-tenancy and resource pooling

• Rapid elasticity and scalability

• Measured service

Six Advantages of Cloud Computing

• Trade capital expense (CAPEX) for operational expense (OPEX)

• Benefit from massive economies of scale

• Stop guessing capacity

• Increase speed and agility

• Stop spending money running and maintaining data centers

• Go global in minutes

Problems solved by the Cloud

•Flexibility: change resource types when needed

•Cost-Effectiveness: pay as you go, for what you use

•Scalability: accommodate larger loads by making hardware stronger or adding additional nodes

• Elasticity: ability to scale out and scale-in when needed

• High-availability and fault-tolerance: build across data centers

• Agility: rapidly develop, test and launch software applications

Types of Cloud Computing

• Infrastructure as a Service (IaaS) : EC2

    • Provides networking, computers, data storage space

    • Highest level of flexibility

    • Easy parallel with traditional on-premises IT

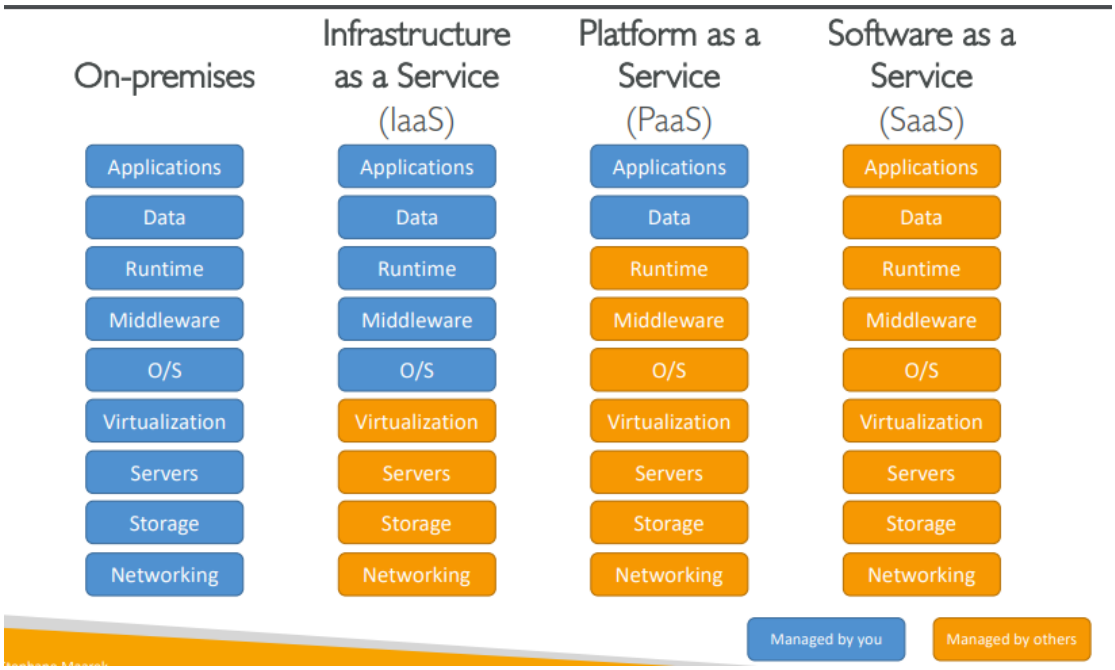
• Platform as a Service: Elastic Beanstalk

    • Removes the need for your organization to manage the underlying infrastructure

    • Focus on the deployment and management of your applications

• Software as a Service (SaaS): Rekognition for Machine Learning, Gmail

• Completed product that is run and managed by the service provider



Pricing of the Cloud? (pay-as-you-go pricing model)

• Compute: Pay for compute time

• Storage: Pay for data stored in the Cloud

• Data transfer OUT of the Cloud

• Data transfer IN is free

AWS Global Infrastructure?

- AWS Regions:
  - A region is a **cluster of data centers**
  - ends with a number (e.g. eu-west-1)
  - **Most AWS services are region-scoped**

- AWS Availability Zones

  • Each region has many availability zones (usually 3, min is 2, max is 6).

  •Each availability zone (AZ) is one or more discrete data centers with redundant power, networking, and connectivity

  •connected with high bandwidth, ultra-low latency connections

• AWS Points of Presence (Edge Locations)

  • Content is delivered to end users with lower latency

  • the places where data are cached to reduce latency

  • used by CloudFront to cache copies

How to choose an AWS Region?

• Compliance with data governance and legal requirements: data never leaves a region without your explicit permission

• Proximity to customers: reduced latency

• Available services within a Region: new services and new features aren't available in every Region

• Pricing: pricing varies region to region and is transparent in the service pricing page

Shared Responsibility?

- CUSTOMER = RESPONSIBILITY FOR THE SECURITY IN THE CLOUD
- AWS = RESPONSIBILITY FOR THE SECURITY OF THE CLOUD

AWS Acceptable Use Policy?

- No Illegal, Harmful, or Offensive Use or Content - No Security Violations - No Network Abuse - No E-Mail or Other Message Abuse

Section 2 IAM

IAM: Users & Groups?

• **IAM manage access to AWS services and resources securely**

• IAM = Identity and Access Management, **Global service**

• **Root account** created by default, shouldn't be used or shared

• **Users** are people within your organization, and can be grouped or mapped to a physical user, has a password for AWS Console

• **Groups** only contain users, not other groups

• Users don't have to belong to a group, and user can belong to multiple groups

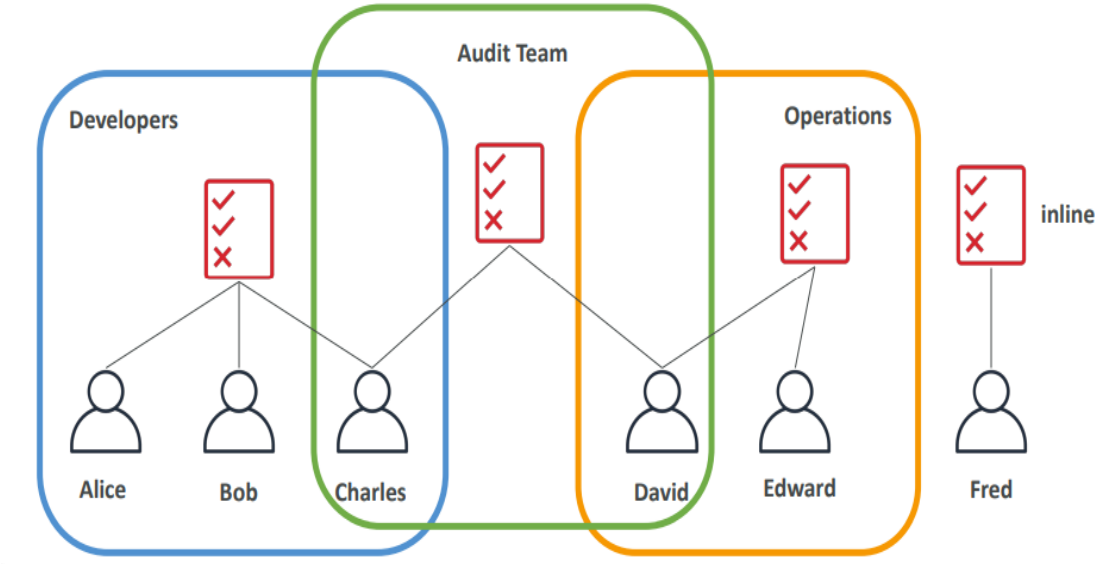
IAM: Permissions?

• **Users or Groups** can be assigned JSON documents called policies

• These policies define the **permissions** of the **users**

• In AWS you apply the **least privilege principle**: don't give more permissions than a user needs

IAM Policies inheritance?



IAM Policies Structure?

- Consists of
  - **Version:** policy language version, always include "2012-10-17"
  - **Id:** an identifier for the policy (optional)
  - **Statement:** one or more individual statements (required)
- Statements consists of
  - **Sid:** an identifier for the statement (optional)
  - **Effect:** whether the statement allows or denies access (Allow, Deny)
  - **Principal:** account/user/role to which this policy applied to
  - **Action:** list of actions this policy allows or denies
  - **Resource:** list of resources to which the actions applied to
  - **Condition:** conditions for when this policy is in effect (optional)

IAM – Password Policy?

• Strong passwords = higher security for your account

- Require length
- prevent-reuse
- force rotation

Multi Factor Authentication – MFA?

•You want to protect your Root Accounts and IAM users

•if a password is stolen or hacked, the account is not compromised

MFA devices options in AWS?

- Virtual MFA device**:** Support for multiple tokens on a single device
- Universal 2nd Factor (U2F) Security Key**:** Support for multiple root and IAM users using a single security key
- Hardware Key Fob MFA Device
- Hardware Key Fob MFA Device for AWS GovCloud (US)

How can users access AWS ?

To access AWS, you have three options:

• AWS Management Console (protected by password + MFA)

• AWS Command Line Interface (CLI): protected by access keys

• AWS Software Developer Kit (SDK) - for code: protected by access keys

• Access Keys are secret, just like a password. Don't share them

• AWS CLI: manage your AWS services using the command-line

• AWS SDK: manage your AWS services using a programming language

IAM Roles for Services?

• Some AWS service will need to perform actions on your behalf

• To do so, we will assign permissions to AWS services with IAM Roles

IAM Security Tools? (audit permission)

• **IAM Credentials Report (account-level) :** a report that lists all your account's users and the status of their various credentials

• **IAM Access Advisor (user-level)** : Access advisor shows the service permissions granted to a user and when those services were last accessed, You can use this information to revise your policies.

IAM Guidelines & Best Practices?

• Don't use the root account except for AWS account setup

• One physical user = One AWS user

• **Assign users to groups** and assign permissions to groups

• Create a **strong password policy**

• Use and enforce the use of Multi Factor Authentication (**MFA**)

• Create and use **Roles** for giving permissions to AWS services

• Use Access Keys for Programmatic Access (CLI / SDK)

• Audit permissions of your account with the IAM Credentials Report

• **Never share IAM users & Access Keys**

Shared Responsibility Model for IAM?

You

- Infrastructure (global network security)
- Configuration and vulnerability analysis
- Compliance validation

- Users, Groups, Roles, Policies management and monitoring
- Enable MFA on all accounts
- Rotate all your keys often
- Use IAM tools to apply appropriate permissions
- Analyze access patterns & review permissions

Section 3 EC2

Amazon EC2?

• EC2 = Elastic Compute Cloud = **Infrastructure as a Service**

• EC2 Instance: AMI (OS) + Instance Size (CPU + RAM) + Storage + security groups + EC2 User Data

• It mainly consists in the capability of :

  • Renting virtual machines (EC2) instance.

  • Storing data on virtual drives (EBS)

  • Distributing load across machines (ELB)

  • Scaling the services using an auto-scaling group (ASG)

EC2 sizing & configuration options*?*

• How much compute power & cores (CPU) , How much random-access memory (RAM)

• Operating System (OS): Linux, Windows or Mac OS

• How much storage space: Network-attached (EBS & EFS) , hardware (EC2 Instance Store)

• Firewall rules: security group

• Bootstrap script (configure at first launch): EC2 User Data

EC2 User Data*?*

• bootstrapping means launching commands when a machine starts

• That script is only run once at the instance first start

• The EC2 User Data Script runs with the root user

EC2 Instance Types – Overview*?*

• m5.2xlarge :

- m: instance class
- 5: generation (AWS improves them over time)
- 2xlarge: size within the instance class

• **EC2 Instance Types – General Purpose:**

- Great for a diversity of workloads such as web servers or code repositories
- Balance between: Compute , Memory , Networking

• **EC2 Instance Types – Compute Optimized:**

- Great for compute-intensive tasks that require high performance processors/**High performance computing (HPC)**

• **EC2 Instance Types – Memory Optimized:**

- Fast performance for workloads that process large data sets in memory

• **EC2 Instance Types – Storage Optimized**

- Great for storage-intensive tasks that require high, sequential read and write access to large data sets on local storage

Introduction to Security Groups? (stateful)

• Security Groups: network security or Firewall attached to the EC2 instance, Security groups are acting as a “firewall” on EC2 instances

• They control how traffic is allowed into or out of our EC2 Instances.

• Security groups only contain **allow** rules

• They regulate: Access to Ports , Authorised IP ranges – IPv4 and IPv6 , Control of inbound network (from other to the instance) , Control of outbound network (from the instance to other)

• Can be attached to multiple instances

• Locked down **to a region / VPC combination**

• All inbound traffic is **blocked** by default

• All outbound traffic is **authorised** by default

• **Return traffic is automatically allowed, no matter the rules (Stateful)**

• 22 = SSH (Secure Shell) - log into a Linux instance

• 21 = FTP (File Transfer Protocol) – upload files into a file share

• 22 = SFTP (Secure File Transfer Protocol) – upload files using SSH

• 80 = HTTP – access unsecured websites

• 443 = HTTPS – access secured websites

• 3389 = RDP (Remote Desktop Protocol) – log into a Windows instance

• EC2 Instance Connect **:** Connect to your EC2 instance within your browser , No need to use your key file that was downloaded , **Works only out-of-the-box with Amazon Linux 2 ,** Need to make sure the port 22 is still opened

Instance roles?

•Attach IAM role to the instance or link to IAM roles

•Use to configure aws access inside ec2

EC2 Instances Purchasing Options?

• **On-Demand Instances** – **short workload,** predictable pricing, pay by second (regular case) :

• Pay for what you use: Linux or Windows - **billing per second**, after the first minute , All other operating systems - **billing per hour**

• Recommended for **short-term and un-interrupted workloads**, where you can't predict how the application will behave

Capacity reservation (no discount):

• Reserve **On-Demand instances** capacity in a specific AZ for any duration

• No time commitment (create/cancel anytime), no billing discounts

•**Suitable for short-term, uninterrupted workloads that needs to be in a specific AZ**

• **Spot Instances** – **short workloads**, **cheap**, can lose instances **(less reliable**), **biggest discount**:

• Can get a discount of up to 90% compared to On-demand

• Instances that you can "lose" at any point of time if your max price is less than the current spot price

• The **MOST cost-efficient** instances in AWS

• Useful for workloads that are **resilient to failure** (Any distributed workloads, Workloads with a flexible start and end time)

• **Not suitable for critical jobs or databases**

•**EC2 Reserved Instances** big discount but need long-term commitment: 1 year or 3 years :

• You reserve a specific instance attributes (Instance Type, Region, Tenancy, OS)

• Reservation Period – 1 year (+discount) or 3 years (+++discount)

• Payment Options – No Upfront (+), Partial Upfront (++), All Upfront (+++)

• **Reserved Instance's Scope** – Regional or Zonal (reserve capacity in an AZ)

• **Convertible Reserved Instance** – long workloads with flexible instances:

• Can change the EC2 instance type, instance family, OS, scope and tenancy • Up to 66% discount

• **Savings Plans (1 & 3 years)** –commitment to an **amount of usage**, **long workload:**

• Get a discount based on long-term usage (up to 72% - same as RIs)

• Commit to a certain type of usage ($10/hour for 1 or 3 years)

• Locked to a specific instance family & AWS region (e.g., M5 in us-east-1)

• Flexible across: Instance Size (e.g., m5.xlarge, m5.2xlarge) , OS (e.g., Linux, Windows) , Tenancy (Host, Dedicated, Default)

• **EC2 Dedicated Hosts** (The most **expensive** option) :

• **A physical server** with EC2 instance capacity fully dedicated to your use

• Allows you address **compliance requirements** and use **your existing server- bound software licenses** (per-socket, per-core, pe—VM software licenses)

EC2 Dedicated Instances :

• **Instances run on hardware that's dedicated to you**

• May share hardware with other instances in same account

Shared Responsibility Model for EC2?

- Infrastructure (global network security)
- Isolation on physical hosts
- Replacing faulty hardware
- Compliance validation

- Security Groups rules
- Operating-system patches and updates
- Software and utilities installed on the EC2 instance
- IAM Roles assigned to EC2 & IAM user access management
- Data security on your instance

Section 4 EC2 Instance Storage Section

What's an EBS Volume?

• An EBS (Elastic Block Store) Volume is a **network drive** you can **attach to your instances while they run**

• **to attach an EBS to an EC2, they must be on the same AZ**

• It allows your instances to persist data, even after their termination

• **They can only be mounted to one instance at a time**

• They are bound to **a specific availability zone**

• It's a network drive (i.e. not a physical drive) **:**

  • It can be detached from an EC2 instance and attached to another one quickly

• It's locked to an Availability Zone (AZ)**:**

  • An EBS Volume in us-east-1a **cannot be attached** to us-east-1b

  • To move a volume across, you first need to **snapshot** it

• Have a provisioned capacity (size in GBs, and IOPS):

  • You get **billed** for all the **provisioned capacity , Storage & IOPS, Pay for snapshots, Pay for snapshot restore on Archive**

  • You can increase the capacity of the drive over time

EBS – Delete on Termination attribute?

• Controls the EBS behaviour when an **EC2 instance terminates:**

  • By default, **the root EBS volume is deleted** (attribute enabled)

  • By default, **any other attached EBS volume is not delete**d (attribute disabled)

• **Use case: preserve root volume when instance is terminated**

EBS Snapshots?

• Make a **backup** (snapshot) of your **EBS volume** at a point in time

• **Can copy snapshots across AZ or Region**

EBS Snapshots Features?

• **EBS Snapshot Archive:**

  • Move a Snapshot to an ”archive tier” that is 75% cheaper

  • Takes within 24 to 72 hours for restoring the archive

• **Recycle Bin for EBS Snapshots**

  • Setup rules to retain deleted snapshots so you can recover them after an **accidental deletion**

  • Specify retention (from 1 day to 1 year)

AMI Overview? (create ready-to-use EC2 instances with our customizations)

• AMI are a **customization** of an EC2 instance**:**

  • You add your own software, configuration, operating system, monitoring…

  • Faster boot / configuration time because all your software is **pre-packaged**

• AMI are built for a **specific region**

• You can **launch EC2 instances** from:

  • **A Public AMI: AWS provided**

  • **Your own AMI: you make and maintain them yourself**

  • **An AWS Marketplace AMI: an AMI someone else made (and potentially sells) third party's AMI**

EC2 Image Builder?

• Used to **automate** the creation of Virtual Machines or container images

• EC2 Image Builder: **automatically build, test and distribute AMIs**

• => Automate the creation, maintain, validate and test EC2 AMIs

• Free service (only pay for the underlying resources)

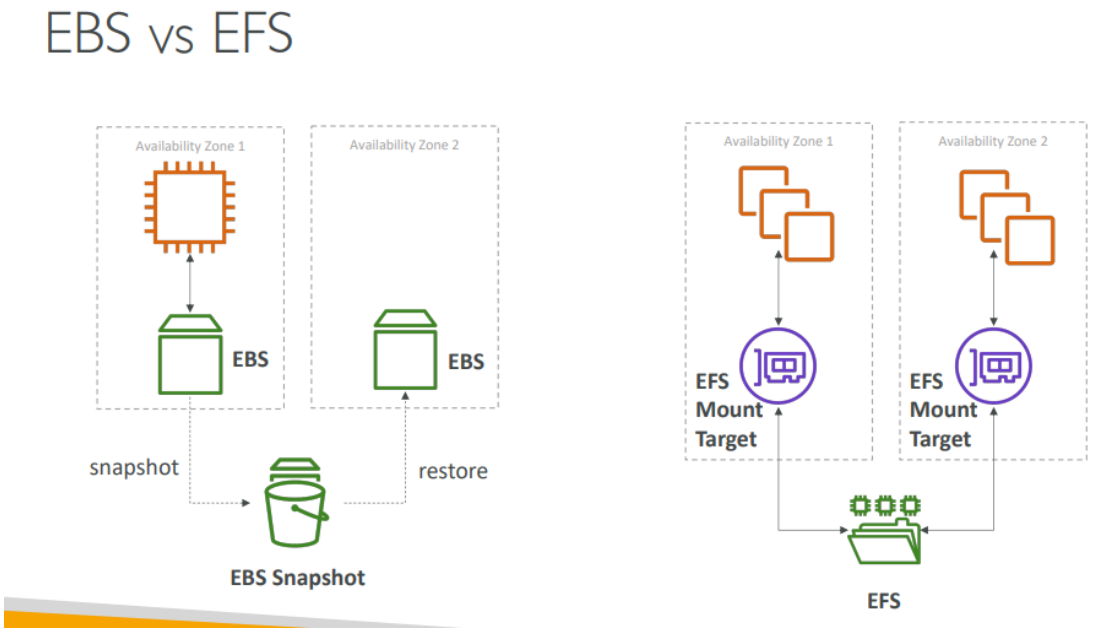
EFS?

• network file system, can be attached to 100s of instances in a region **,** shared network file system for your EC2 Instances

• Managed NFS (network file system) that can be mounted on 100s of EC2

• EFS works with **Linux** EC2 instances in **multi-AZ**

• Highly available, scalable, **expensive** (3x gp2), pay per use, no capacity planning You pay only for the storage you use



EFS Infrequent Access (EFS-IA)?

• **cost-optimized** storage class for infrequent accessed files

• Storage class that is cost-optimized for files not accessed every day

• Up to 92% lower cost compared to EFS Standard

• EFS will automatically move your files to EFS-IA based on the last time they were accessed

• EFS moves files to EFS-IA based on last time file is accessed (configurable using lifecyle Policy)

EC2 Instance Store?

• **network drives with good but "limited" performance**

• **If you need a high-performance hardware disk, use EC2 Instance Store**

• Better I/O performance

• **Lost if our instance is stopped / terminated (ephemeral)**

• Ephemeral storage, **not suitable for persistent data (short workload)**

•**Cannot attach/reattach to different EC2**

• Good for buffer / cache / scratch data / temporary content

• **Risk of data loss if hardware fails**

• **Backups and Replication are your responsibility**

Shared Responsibility Model for EC2 Storage

- Infrastructure
- Replication for data for EBS volumes & EFS drives
- Replacing faulty hardware
- Ensuring their employees cannot access your data

- Setting up backup / snapshot procedures
- Setting up data encryption
- Responsibility of any data on the drives
- Understanding the risk of using EC2 Instance Store

Amazon FSx – Overview?

- Launch **3rd** party high-performance file systems on AWS
- **Fully managed service**

Amazon FSx for Windows File Server?

- **A fully managed**, highly reliable, and scalable **Windows native shared file system**
- Built on Windows File Server
- Can be accessed from AWS or your on-premise infrastructure

Amazon FSx for Lustre?

- **A fully managed,** high-performance, scalable file storage for **High Performance Computing (HPC)**
- The name Lustre is derived from “Linux” and “cluster”

Section 5 Elastic Load Balancing & Auto Scaling Groups

Scalability & High Availability?

• Scalability means that an application / system can **handle greater loads by adapting.**

• **Scalability**: ability to accommodate a larger load by making the **hardware stronger (scale up),** or by **adding nodes (scale out)**

• There are two kinds of scalability:

  • **Vertical Scalability** : Vertical Scalability means increasing the **size** of the instance (= scale up / down)

  • Horizontal Scalability (= **elasticity**) **:** means increasing the **number** of instances / systems for your application(= scale out / in)

• High Availability usually goes hand in hand with **horizontal scaling**

• High availability means **running your application / system in at least 2 Availability Zones**

• High Availability: **Run instances for the same application across multi AZ**

• The goal of high availability is to **survive/thriving a data center loss** (disaster)

• **Elasticity**: once a system is **scalable**, elasticity means that there will be some "**auto-scaling**" so that the system can scale based on the load. This is "cloud-friendly": pay-per-use, match demand, optimize costs

• **Agility**: (not related to scalability - distractor) new IT resources are only a click away, which means that you **reduce the time to make those resources available to your developers from weeks to just minutes**

What is load balancing?

• Load balancers are servers that forward internet traffic to multiple servers (EC2 Instances) downstream.

• Distribute traffic across backend EC2 instances, can be Multi-AZ

• An ELB (Elastic Load Balancer) is a **managed load balancer**

Why use a load balancer?

• Spread load across multiple downstream instances

• Expose a single point of access (DNS) to your application

• Seamlessly handle failures of downstream instances

• Do regular health checks to your instances

• Provide SSL termination (HTTPS) for your websites

• High availability across zones

3 kinds of load balancers offered by AWS?

• **Application Load Balancer:**

  •HTTP/HTTPS/gPRC protocol (layer 7)

  •HTTP routing feature

  •Static DNS (url)

• **Network Load Balancer:**

  •TCP/UDP protocol (layer4)

  •High performance : million of request per seconds

  •Static ip

• **Gateway Load Balancer:**

  •GENEVE Protocol on ip packet (layer 3)

  •Route Traffic to Firewall That Manage on ec2 instance like instruction detection

What's an Auto Scaling Group?

• Implement **Elasticity** for your application, **across multiple AZ**

• In real-life, the load on your websites and application can change

• In the cloud, you can create and get rid of servers very quickly

• The goal of an Auto Scaling Group (ASG) is to:

  • **Scale out** (add EC2 instances) to match an increased load

  • **Scale in** (remove EC2 instances) to match a decreased load

  • Ensure we have a minimum and a maximum number of machines running

  • **Automatically register new instances to a load balancer**

  • **Replace unhealthy instances**

• **Cost Savings**: only run at an optimal capacity (principle of the cloud)

•**Three capacity: minimum/maximum/Actual Size or Desired Capacity**

Auto Scaling Groups – Scaling Strategies?

• **Manual Scaling**: Update the size of an ASG manually

• **Dynamic Scaling**: Respond to changing demand

  • **Simple / Step** Scaling **:** When a CloudWatch alarm is triggered (example CPU > 70%), then add 2 units

  • **Target Tracking Scaling:** Example: I want the average ASG CPU to stay at around 40%

  • **Scheduled Scaling :** Anticipate a scaling based on known usage patterns

• **Predictive Scaling:** Uses Machine Learning to predict future traffic ahead of time , Automatically provisions the right number of EC2 instances in advance

Section 6 Amazon S3

Amazon S3 – Buckets*?*

• Amazon S3 allows people to store **objects (files)** in **"buckets" (directories)**

• Buckets must have **a globally unique name** (across all regions all accounts)

• Buckets are defined at **the region level**

• **S3 looks like a global service** but buckets are **created in a region**

Amazon S3 – Objects*?*

• Objects (files) have a **Key**

• The **key is the FULL path**:

  • s3://my-bucket/**my_file.txt**

• The key is composed **of prefix + object name**

  • s3://my-bucket**/my_folder1/another_folder/my_file.txt**

• **Max. Object** Size is 5TB (5000GB)

• **Max upload size is 5GB** If uploading more than 5GB, must use **"multi-part upload"**

Amazon S3 – Security?

• S3 security: IAM policy, S3 Bucket Policy (public access), S3 Encryption

• **User-Based *:* IAM Policies** – which API calls should be allowed for a specific user from IAM

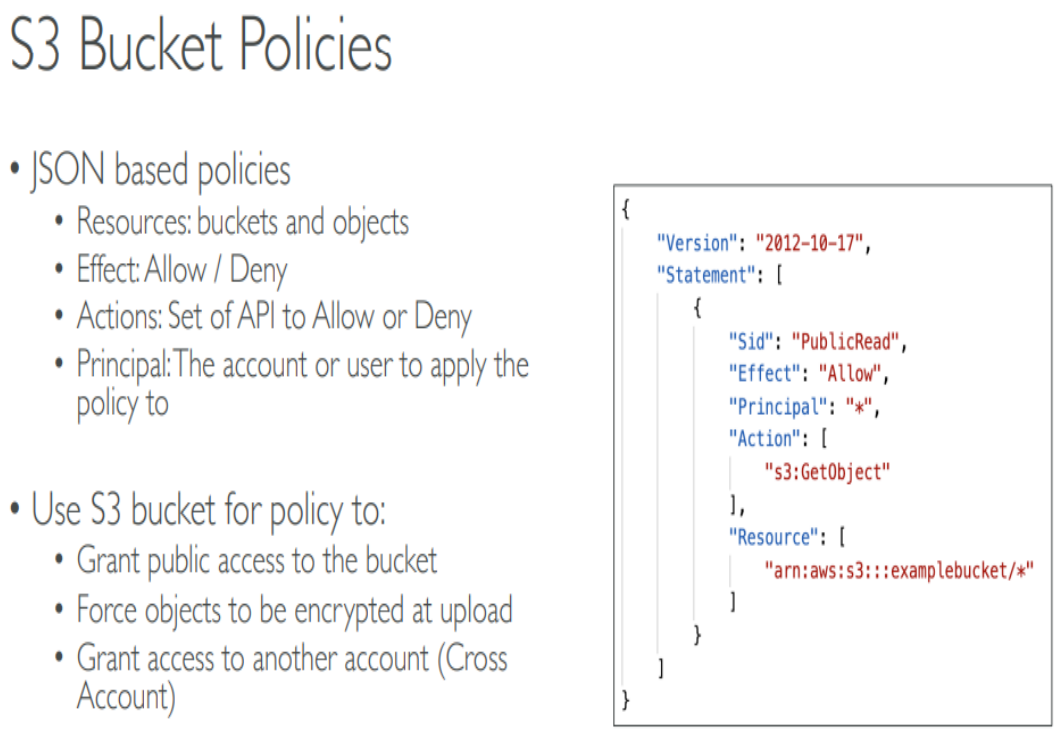
• **Resource-Based *:***

  • **Bucket Policies – bucket wide rules from the S3 console - allows cross account**

• Object Access Control List (ACL) – finer grain (can be disabled)

• Bucket Access Control List (ACL) – less common (can be disabled)

• **Encryption: encrypt objects in Amazon S3 using encryption keys**



Bucket settings for Block Public Access?

• **These settings were created to prevent company data leaks**

Amazon S3 – Static Website Hosting?

• S3 can host static websites and have them accessible on the Internet

• The website URL will be (**depending on the region**) :

http://bucket-name.s3-website-**aws-region**.amazonaws.com

• **If you get a 403 Forbidden error, make sure the bucket policy allows public reads**!

Amazon S3 -Versioning?

• It is enabled at the **bucket level**

• similar to Git

• It is best practice to version your buckets:

• Protect against **unintended deletes (ability to restore a version)**

• **Easy roll back to previous version**

Amazon S3 – Replication (CRR & SRR)?

• **Must enable Versioning** in source and destination buckets

• Disabled by default

• Copying is **asynchronous**

• CRR – compliance, lower latency access, **replication across accounts**

• SRR – log aggregation, live replication between production and test account

S3 Durability and Availability?

• **Durability:**

• High durability (99.999999999%, **11** 9's) of objects across multiple AZ

• If you store 10,000,000 objects with Amazon S3, you can on average expect to incur a loss of a single object once every 10,000 years

• **Same for all storage classes**

• **Availability:**

• Measures how readily available a service is

- **Varies depending on storage class**

- **Example: S3 standard has 99.99% availability = not available 53 minutes a year**

S3 Storage Classes?

•**S3 Standard – General Purpose:** 99.99% Availability ,Used for **frequently accessed data**

•**S3 Storage Classes – Infrequent Access:** For data that is **less frequently accessed**, but requires rapid **access when needed** ,Lower cost than S3 Standard

- **Amazon S3 Standard-Infrequent Access** (S3 Standard-IA): **Access instantly, Use cases: Disaster Recovery, backups**

- **Amazon S3 One Zone-Infrequent Access (S3 One Zone-IA): 99.5% availability**. **Access instantly**, **single AZ (Cheaper than standard IA)** , Use Cases: **Storing secondary backup copies of on-premise data, or data you can recreate**

•**Amazon S3 Glacier Storage Classes : Low-cost object storage** meant for archiving / backup , Pricing: **price for storage + object retrieval cost**

- **Amazon S3 Glacier Instant Retrieval**: **Millisecond retrieval ,** Minimum storage duration of 90 days

- **Amazon S3 Glacier Flexible Retrieval**: **retrieved in minutes,** Expedited (1 to 5 minutes), Standard (3 to 5 hours), Bulk (5 to 12 hours) it's free , Minimum storage duration of 90 days

- **Amazon S3 Glacier Deep Archive – for long term storage: Lowest cost storage class designed for long-term retention of data** , Standard (12 hours), Bulk (48 hours) , Minimum storage duration of 180 days

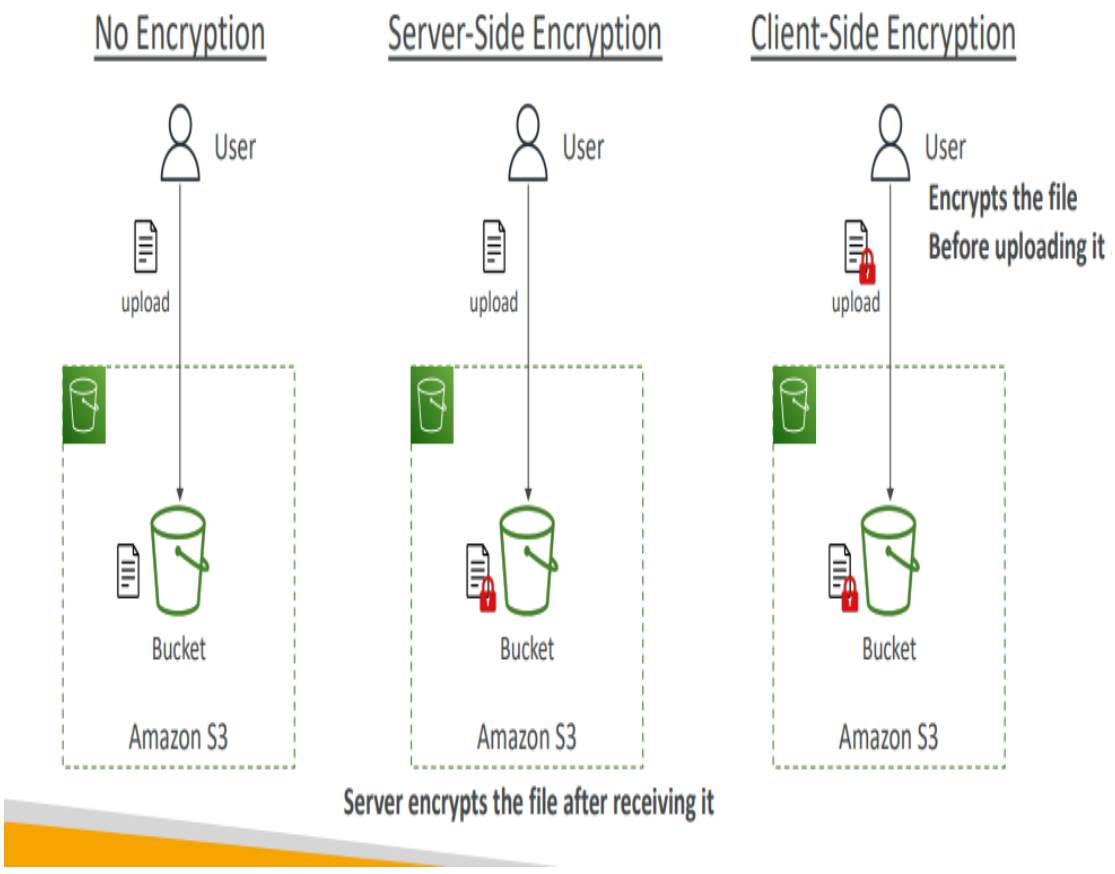
•**S3 Intelligent-Tiering**: **Automatically move objects between classes to save cost**

- **Can move between classes manually or using S3 Lifecycle configurations**

S3 Encryption?

•S3 Glacier

- Automatic encryption enabled



Pricing?

- Storage
- Request & Data retrieval
- Data transfer out
- Management & analytics
- Replication
- S3 Object Lambda

Shared Responsibility Model for S3

- Infrastructure (global security, durability, availability, sustain concurrent loss of data in two facilities)
- Configuration and vulnerability analysis
- Compliance validation

- S3 Versioning
- S3 Bucket Policies
- S3 Replication Setup
- Logging and Monitoring
- S3 Storage Classes
- Data encryption at rest and in transit

AWS Snow Family*?*

• Highly-secure, portable devices to collect and process data at the edge, and migrate data into and out of AWS

• Snow Family: import data onto S3 through a physical device, edge computing

• AWS sends actual devices so the client can copy big data and send it back to AWS to upload to WAS

• **Data migration**: Snowcone, Snowball Edge ,Snowmobile

• **Edge computing**: Snowcone, Snowball Edge

Snowball Edge?

• Physical data transport solution: move **PBs** of data in or out of AWS

• Alternative to moving data over the network (and paying network fees)

• **Snowball Edge Storage Optimized** • 80 TB of HDD capacity for block volume and S3 compatible object storage

• **Snowball Edge Compute Optimized** • 42 TB of HDD capacity for block volume and S3 compatible object storage

• Storage Clustering Up to 15 nodes

AWS Snowcone?

• **Small**, portable computing, anywhere, rugged & secure, **withstands harsh environments**

• **8 TBs of usable storage**

• **Can be sent back to AWS offline, or connect it to internet and use AWS DataSync to send data**

AWS Snowmobile?

• **Transfer exabytes of data** (1 EB = 1,000 PB = 1,000,000 TBs)

• **Each Snowmobile has 100 PB of capacity (use multiple in parallel)**

• **Better than Snowball if you transfer more than 10 PB**

What is Edge Computing?

• **Process data** while it's being created on an **edge location** (A truck on the road, a ship on the sea, a mining station underground.)

• **Machines to do computing when a connection to AWS is not possible**

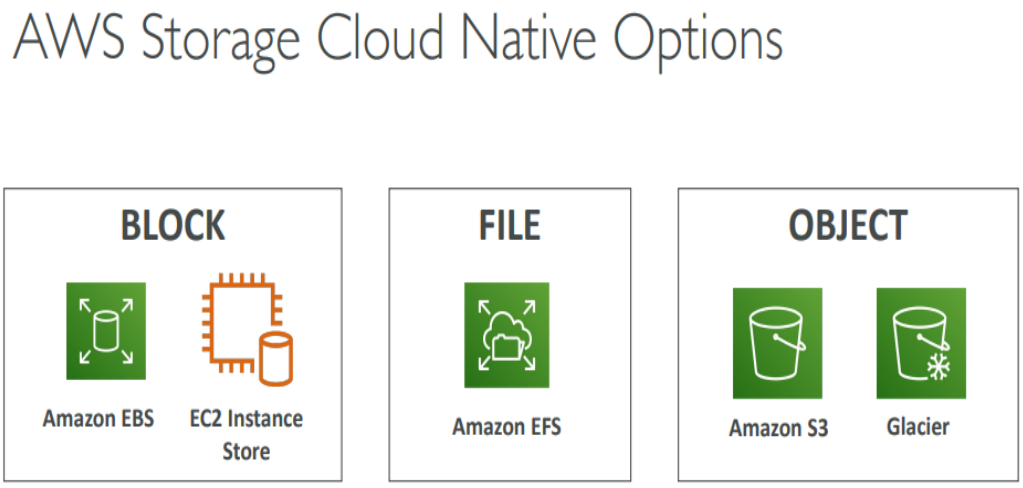
• These locations may **have Limited / no internet access , Limited / no easy access to computing power**

• We setup a **Snowball Edge / Snowcone** device to do edge computing

AWS OpsHub?

• to use Snow Family devices, you needed a CLI

• Today, you can use AWS OpsHub (a software you install on your computer / laptop) **to manage your Snow Family Device**



AWS Storage Gateway

• **Bridge between on-premise data and cloud data in S3**

• **Hybrid storage service to allow on- premises to seamlessly use the AWS Cloud , hybrid solution to extend on-premises storage to S3**

• **Connect on-premise to the cloud storage**

Section 7 Databases

Databases Intro*?*

• Storing data on disk (EFS, EBS, EC2 Instance Store, S3) can have its limits

• Sometimes, you want to store data in a database… You can structure the data , You build indexes to efficiently query / search through the data

Relational Databases?

• Can use the SQL language to perform queries / lookups

• Relational Databases – **OLTP**

NoSQL Databases?

• NoSQL databases are purpose built for specific data models and have flexible schemas for building modern applications.

• JSON is a common form of data that fits into a NoSQL model

Databases & Shared Responsibility on AWS

- AWS offers use to **manage** different databases
- **Benefits** include:
  - Quick Provisioning, High Availability, Vertical and Horizontal Scaling
  - Automated Backup & Restore, Operations, Upgrades
  - Operating System Patching is handled by AWS
  - Monitoring, alerting

- Note: many databases technologies could be run on EC2, but you must handle yourself the resiliency, backup, patching, high availability, fault tolerance, scaling…

AWS RDS Overview?

• RDS stands for **Relational Database Service**

• It's **a managed DB service for DB use SQL as a query language.**

• It allows you to create databases in the cloud that are managed by AWS • Postgres • MySQL • MariaDB • Oracle • Microsoft SQL Server • Aurora (AWS Proprietary database)

• **RDS has free tier**

RDS Deployments: Read Replicas, Multi-AZ?

• **Read Replicas: Scale the read workload** of your DB, **distribute read load between servers**, **scalability**

• **Multi-AZ**: **Failover in case of AZ** outage (high availability)

• **Multi-Region :** Disaster recovery in case **of region issue**, **Local performance for global reads , Replication cost**

Advantage over using RDS versus deploying DB on EC2?

• RDS is a managed service:

• **Automated provisioning, OS patching**

• Continuous backups and restore to specific timestamp (Point in Time Restore)!

• Monitoring dashboards

• Read replicas for improved read performance

• Multi AZ setup for DR (Disaster Recovery)

• Maintenance windows for upgrades

• Scaling capability (vertical and horizontal)

• Storage backed by EBS (gp2 or io1)

Amazon Aurora?

• Aurora is a **proprietary** technology from AWS (**not open sourced**)

• **PostgreSQL and MySQL are both supported as Aurora DB**

• **Not in the free tier**

• Aurora costs more than RDS (20% more) – but is more efficient

Amazon ElastiCache Overview?

• ElastiCache is to get **managed** Redis or Memcached

• Caches are **in-memory databases** with high performance, low latency

• **Helps reduce load off databases for read intensive workloads**

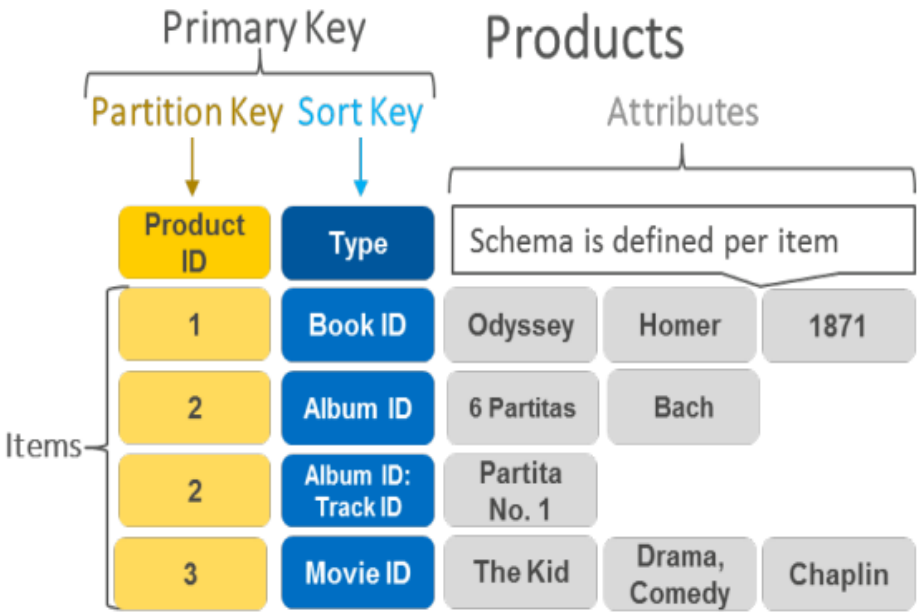
DynamoDB?

• **NoSQL database** - not a relational database

• **Single-digit millisecond latency – low latency retrieval**

• Scales to massive workloads, distributed "**serverless**" database

• DynamoDB is a key/value database



DynamoDB Accelerator – DAX?

• **Fully Managed in-memory cache for DynamoDB**

DynamoDB – Global Tables?

• Make a DynamoDB table accessible with **low latency in multiple-regions**

• **Active-Active replication**

Redshift Overview?

• Redshift is based on **PostgreSQL**, but it's **not used for OLTP**

• It's OLAP – online analytical processing (analytics and data **warehousing**)

• **Columnar** storage of data (instead of row based)

• Has SQL console

Amazon EMR?

• EMR stands for "Elastic MapReduce"

• EMR helps creating **Hadoop** clusters (Big Data) to analyze and process vast amount of data

Amazon Athena?

• **Serverless** query service to **analyze data stored in Amazon S3**

• **Uses standard SQL language to query the files**

• analyze data in S3 using serverless SQL, use Athena

Amazon QuickSight?

• **Serverless** machine learning-powered business intelligence service to **create interactive dashboards**

DocumentDB?

• Aurora is an “AWS-implementation” of PostgreSQL / MySQL … **DocumentDB is the same for MongoDB (which is a NoSQL database)**

Amazon Neptune?

• Fully managed **graph database**

• A popular graph dataset would be a social network

Amazon QLDB?

• QLDB stands for ”Quantum **Ledger** Database”

• A ledger is a **book recording financial transactions**

• **Used to review history of all the changes made to your application data over time**

• **Immutable system**: no entry can be removed or modified, cryptographically verifiable

• Like blockchain but **not decentralized**

Amazon Managed Blockchain?

• Blockchain makes it possible to build applications where multiple parties can execute transactions **without** the need for **a trusted, central authority.**

• Compatible with the frameworks **Hyperledger Fabric & Ethereum**

AWS Glue?

• **Managed extract, transform, and load (ETL) service, serverless**

• **Glue Data Catalog: catalog of datasets**

DMS – Database Migration Service

• **Database Migration Service ,Migrate this db to other db ,Support transporting different DB.**

Section 8 Other Compute Section

What is Docker?

• Docker is a **software development platform** to **deploy apps**

• Apps are packaged in containers that can be run on any OS

• Docker images are stored in Docker Repositories

• Private: Amazon ECR (Elastic Container Registry)

ECS?

• ECS = Elastic Container Service

• Launch Docker containers on AWS

• You must provision & maintain the infrastructure (the EC2 instances)

Fargate?

• Launch Docker containers on AWS

• Serverless offering

• You do not provision the infrastructure (no EC2 instances to manage) – simpler!

ECR?

• Elastic Container Registry

• Private Docker Registry on AWS

• This is where you store your Docker images so they can be run by ECS or Fargate

What's serverless?

• Initially... Serverless == **FaaS** (Function as a Service)

• Serverless does not mean there are **no servers**… it means you **just don't manage / provision / see them**

Why AWS Lambda?

• **Lambda is Serverless**, **Function as a Servic**e, seamless scaling, reactive

• Virtual functions – no servers to manage!

• **Limited by time - short executions**

• **Run on-demand**

• **Scaling is automated!**

• Lambda Billing: **By the time run x by the RAM provisioned , By the number of invocations (Pay per request and compute time)**

• **Event-Driven: functions get invoked by AWS when needed**

Amazon API Gateway?

• **API Gateway: expose Lambda functions as HTTP API**

• **Fully managed service for developers to easily create, publish, maintain, monitor, and secure APIs**

• Serverless and scalable

AWS Batch?

• Batch: **run batch jobs on AWS across managed EC2 instance**s

• Efficiently run 100,000s of computing batch jobs on AWS

• A "batch" **job is a job with a start and an end (opposed to continuous)**

• **Batch will dynamically launch EC2 instances or Spot Instances to run many task**

• **Batch jobs are defined as Docker images and run on ECS**

Amazon Lightsail?

• **Great for people with little cloud experience!**

• ***Simple web applications (has templates for LAMP, Nginx, MEAN, Node.js…)***

• Has high availability but **no auto-scaling**, limited AWS integrations

• **Low & predictable pricing**

Section 8 Deploying and Managing Infrastructure at Scale

What is CloudFormation?

• **Infrastructure as Cod**e, works with almost all of AWS resources

• **Use to create resources**

• **Define code in YAML file (called templates)**

• **Repeat across Regions & Accounts**

• **Free to use, the user pays for the infra**

• **CloudFormation Stack Designer** We can see the relations between the components

AWS Cloud Development Kit (CDK)?

• **Define your cloud infrastructure using a familiar language**

• The code is "compiled" into a **CloudFormation template (JSON/YAML**)

• You can therefore deploy infrastructure and application runtime code together

• Convert programming languages code to CloudFormation

• **defines your cloud application resources using familiar programming languages.**

AWS Elastic Beanstalk Overview?

• Beanstalk = **Platform as a Service (PaaS)**

• **Beanstalk is free but you pay for the underlying instances**

• Elastic Beanstalk is **a developer centric view of deploying an application on AWS**

• **Helps developers run code faster by provisioning required infra (EC2/ELB/ASG…)**

• **Just the application code is the responsibility of the developer**

• **Elastic Beanstalk – Health Monitoring :** Health agent pushes metrics to CloudWatch , Checks for app health, publishes health events

AWS CodeDeploy?

• **CodeDeploy (hybrid):** deploy & upgrade any application onto servers

• Servers / Instances must be provisioned and configured ahead of time with the CodeDeploy Agent

AWS CodeCommit?

• Before pushing the application code to servers, it needs to be stored somewhere

• **Developers usually store code in a repository, using the Git technology**

• **Makes it easy to collaborate with others on code**

• **The code changes are automatically versioned**

AWS CodeBuild?

• Code building service in the cloud

• **CodeBuild: Build & test code in AWS**

• **Compiles source code, run tests, and produces packages that are ready to be deployed (by CodeDeploy for example)**

AWS CodePipeline?

• **Orchestrate** the different **steps** to have the code **automatically pushed to production**

- Code => Build => Test => Provision => Deploy
- **Basis for CICD (Continuous Integration & Continuous Delivery)**

AWS CodeArtifact?

• Storing and retrieving these **dependencies** is called **artifact management**

• CodeArtifact is a secure, scalable, and cost-effective artifact management for software development

• **CodeArtifact: Store software packages / dependencies on AWS**

AWS CodeStar?

• Unified **UI** to easily manage software development activities **in one place and allowing developers to do CICD and code**

• "**Quick way**" to get started to correctly set-up CodeCommit, CodePipeline, CodeBuild, CodeDeploy, Elastic Beanstalk, EC2, etc…

AWS Cloud9?

• **AWS Cloud9 is a cloud IDE (Integrated Development Environment) for writing, running and debugging code**

AWS Systems Manager (SSM)?

• **Helps you manage your EC2 and On-Premises systems at scale (fleet)**

• **Run commands across an entire fleet of servers**

• Patching automation for enhanced compliance

• To do this, all instances (EC2/on-premise) must have **SSM agent installed)**

Systems Manager – SSM Session Manager?

• **Allows you to start a secure shell on your EC2 and on-premises servers without SSH**

AWS OpsWorks?

• **Chef & Puppet** help you perform server configuration automatically, or repetitive actions

Section 9 Global Infrastructure

Why make a global application?

• A global application is an application deployed in **multiple geographies**

• **Decreased Latency**

• **Disaster Recovery (DR)**

• **Attack protection**: distributed global infrastructure is harder to attack

Global AWS Infrastructure?

• **Regions**: For deploying applications and infrastructure

• **Availability Zones**: Made of multiple data centers

• **Edge Locations (Points of Presence): for content delivery as close as possible to users**

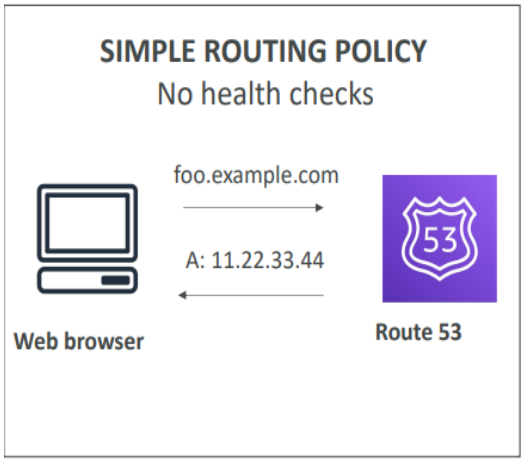
Amazon Route 53 Overview?

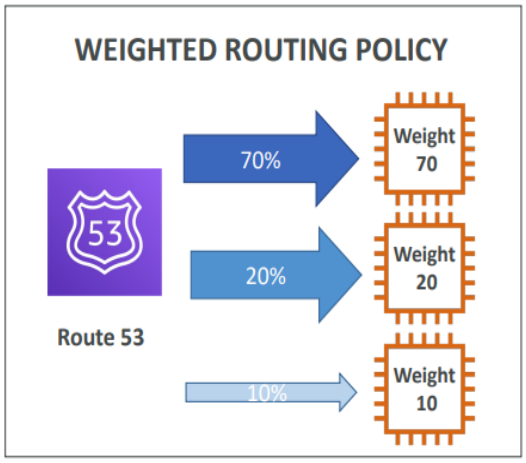
• **Great to route users to the closest deployment with least latency**

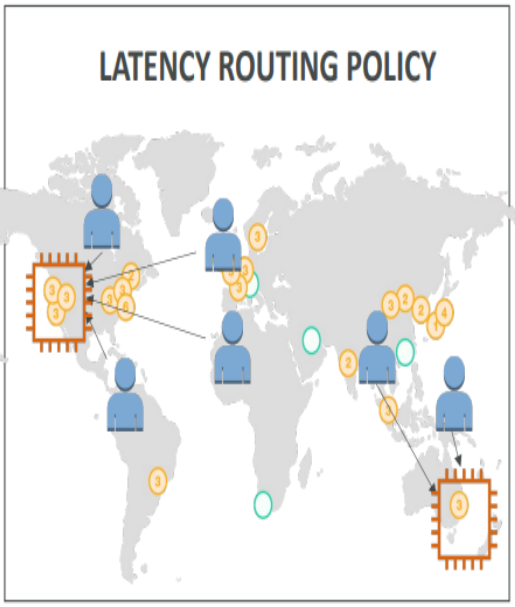
• Great for disaster recovery strategies

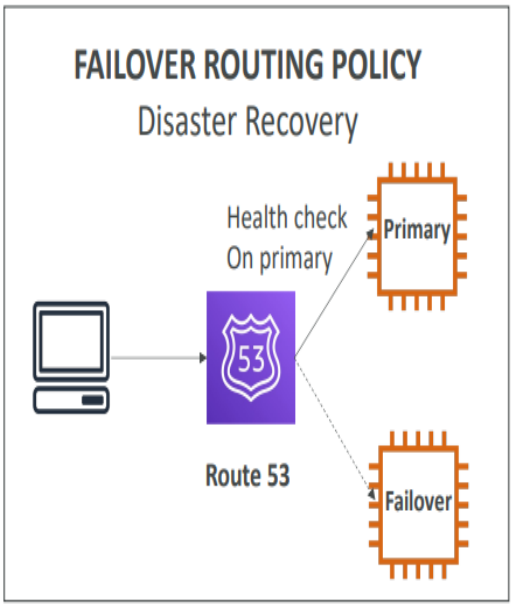
• **Route53 is a Managed DNS (Domain Name System)**

Route 53 Routing Policies?









Amazon CloudFront?

- **Content Delivery Network (CDN)**
- **For distributing files and caching them at the edge**
- Improves read performance, content is cached at the edge
- **DDoS protection** (because worldwide), integration with Shield, AWS Web Application Firewall
- Serves static content
- Automatic replication to serve static content everywhere

CloudFront – Origins?

- S3 bucket Enhanced security with CloudFront Origin Access Control (OAC)
- Custom Origin (HTTP)_ Application Load Balancer ,EC2 instance, Any HTTP backend you want

S3 Transfer Acceleration?

• Increase transfer speed by transferring file to an AWS **edge location** which will forward the data to the S3 bucket in the target region

•fast, easy, and secure transfers of files over long distances between your client and an S3 bucket.

AWS Global Accelerator?

• Improve global application availability and performance using the **AWS global network**

•No caching (unlike CloudFront)

• Leverage the AWS internal network to optimize the route to your application (60% improvement)

•direct users to the nearest AWS Region that contains an endpoint for an application

• 2 Anycast IP are created for your application and traffic is sent through Edge Locations

•AWS Global Accelerator provides **you with a set of static IP** addresses that can map to multiple application endpoints across AWS Regions, to improve redundancy.

AWS Outposts?

• **Hybrid Cloud:** businesses that keep an on - premises infrastructure alongside a cloud infrastructure

•**AWS Outposts are "server racks"** that offers the same AWS infrastructure, services, APIs & tools **to build your own applications on - premises just as in the** cloud and run AWS infrastructure and services on-premises

• **AWS will setup and manage "Outposts Racks"** within your on - premises infrastructure and you can start leveraging AWS services on-premises

• You are responsible for the Outposts Rack physical security

AWS WaveLength?

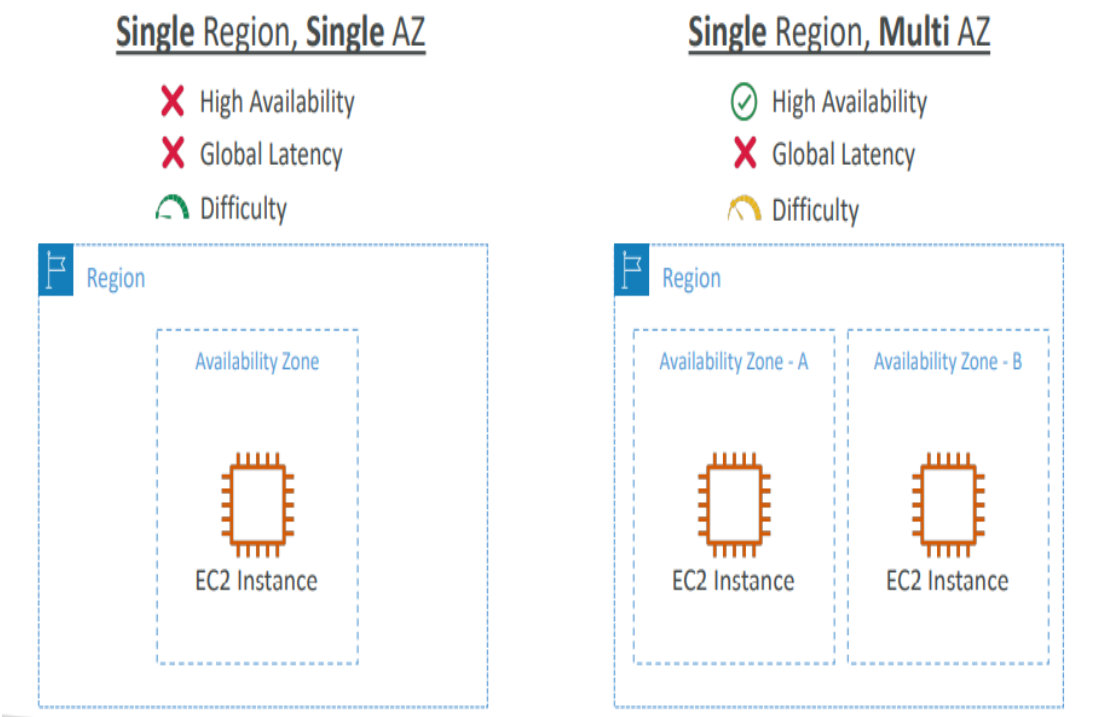
• WaveLength Zones are infrastructure deployments embedded within the telecommunications providers' datacenters at the edge of the **5G** networks

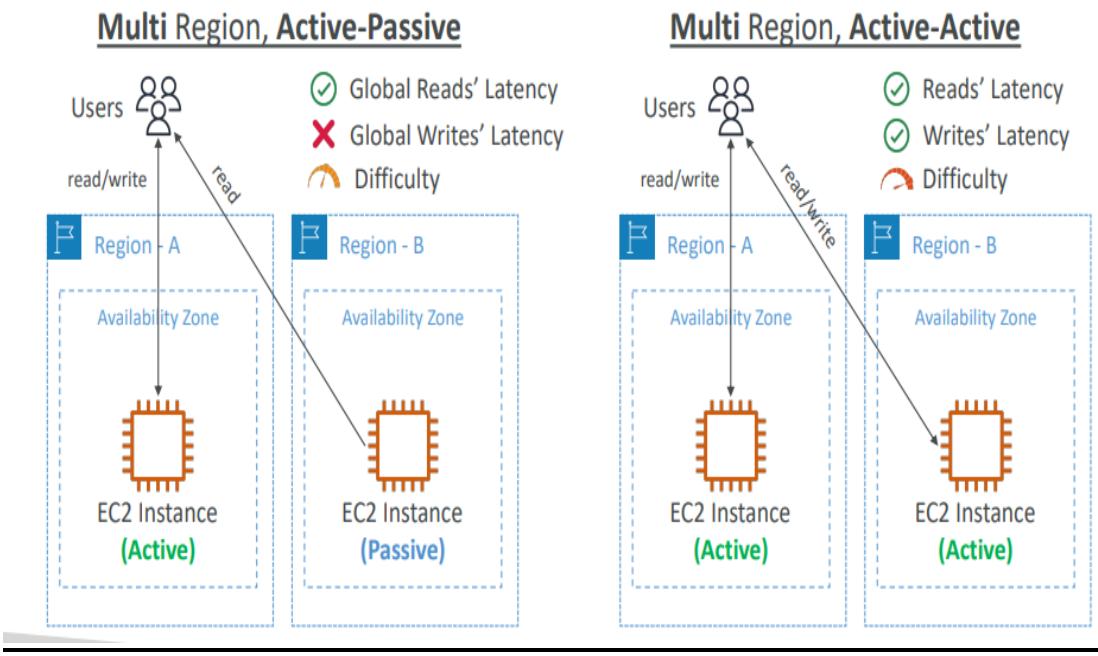
AWS Local Zones?

• **Places** AWS compute, storage, database, and other **selected AWS services closer to end users to run latency-sensitive applications**

• Extend your VPC to more locations – "**Extension of an AWS Region**" **, not enabled by default**

Global Applications Architecture?




Multi Region, Active-Passive
Users
read/write
read
Global Reads' Latency
Global Writes' Latency
Difficulty
Region - A
Region - B
Availability Zone
Availability Zone
EC2 Instance
(Active)
EC2 Instance
(Passive)
Multi Region, Active-Active
Users
read/write
read/write
Reads' Latency
Writes' Latency
Difficulty
Region - A
Region - B
Availability Zone
Availability Zone
EC2 Instance
(Active)
EC2 Instance
(Active)

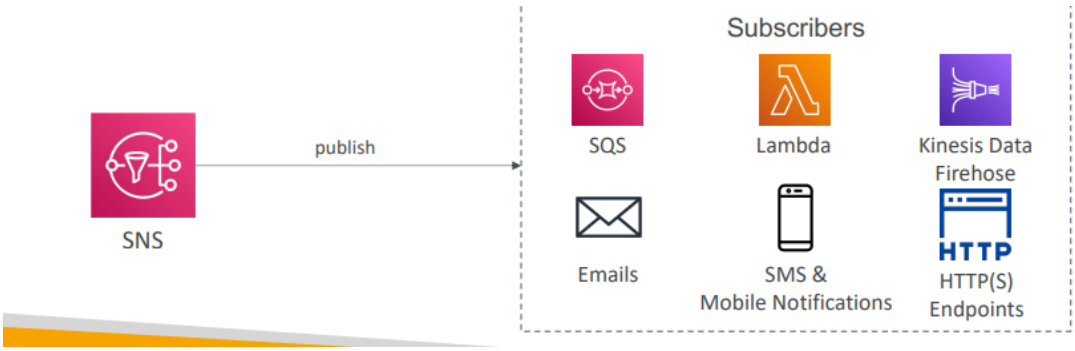
Cloud Integration Section

Amazon SQS – Standard Queue?

- Simple queue service
- **Fully managed queue serverless**
- **pull-based**
- use to **decouple** applications
- Scales from **1 message per second to 10,000s** per second
- Default retention of messages: **4 days, maximum of 14 days**
- **No limit to how many messages can be in the queue**
- Messages are **deleted after they're read by consumers**

Amazon SNS?

- **Pub/Sub**
- use to **decouple** applications
- Simple **notification service**
- **Subscribers get all mesages**
- **Send message to topics**
- **push-based**
- **Subscribers: Email, Lambda, SQS, HTTP, Mobile…**
- **Multiple Subscribers, send all messages to all of them**
- **No message retention**
- The "**event publishers**" only sends message to one SNS topic
- Each subscriber to the topic will get all the messages



Amazon Kinesis?

- **Realtime big data streaming**
- Managed service to collect, process, and analyze **real-time streaming data at any scal**e

Amazon MQ?

• **SQS, SNS are "cloud-native" services: proprietary protocols from AWS**

• Traditional applications running from on-premises may use open protocols such as: MQTT, AMQP, STOMP, Openwire, WSS

• When **migrating** to the cloud, instead of **re-engineering the application to use SQS and SNS, we can use Amazon MQ**

• Amazon MQ is a **managed message broker service** for **Apache MQ and RabbitMQ in the cloud** (MQTT, AMQP.. protocols)

• Amazon MQ doesn't "**scale**" as much as SQS / SNS

• Amazon MQ runs on servers, can run in Multi-AZ with failover

• Amazon MQ has both queue feature (~SQS) and topic features (~SNS)

Cloud Monitoring Section

Amazon CloudWatch Metrics?

•Amazon CloudWatch is a web service that enables you to **monitor** and manage various metrics and configure **alarm actions based on data from those metrics**. CloudWatch uses **metrics** to represent the data points for your resources. AWS services send metrics to CloudWatch. CloudWatch then uses these metrics to create **graphs automatically that show how performance has changed over time.**

• Metrics: **monitor** the **performance** of AWS **services** and billing metrics

• CloudWatch provides **metrics for every services in AWS**

• Metric is a variable to **monitor** (CPUUtilization, NetworkIn…)

• Metrics have **timestamps**

•Users can use the metrics to **calculate statistics and view graphs in the Cloudwatch dashboard**

•Important Metrics :

- **EC2 instances**: CPU Utilization, Status Checks, Network (not RAM):
  - **Default metrics every 5 minutes**
  - Option for Detailed Monitoring ($$$): **metrics every 1** minute
- **EBS volumes:** Disk Read/Writes
- S3 buckets: BucketSizeBytes, NumberOfObjects, AllRequests
- **Billing Alarm :Total Estimated Charge (only in us-east-1)**
- Service Limits: how much you've been using a service API
- **Custom metrics: push your own metrics**

Amazon CloudWatch Alarms?

• With CloudWatch, you can create **alarms** that automatically perform actions if the value of your metric has gone above or below a predefined threshold. You could create a CloudWatch alarm that automatically stops an Amazon EC2 instance when the CPU utilization percentage has remained below a certain threshold for a specified period. When configuring the alarm, you can specify to receive a notification whenever this alarm is triggered

• **Alarms are used to trigger notifications for any metric when a metric reaches a threshold**

• **Automatically initiate actions on user's behalf**

• Alarms action is automate notification, perform EC2 action, notify to SNS based on metric

• **Example: create a billing alarm on the CloudWatch Billing metric**

Amazon CloudWatch Logs?

• Logs: **collect log files from EC2 instances, servers, Lambda functions…**

• **By default, no logs from your EC2 instance will go to CloudWatch ,** You need to run a **CloudWatch agent on EC2 to push the log files you want**

• **CloudWatch has a log agent on EC2 to collect logs**

• CloudWatch log agents: **on EC2 machines or on-premises servers**

• **Enables real-time monitoring of logs**

• **The CloudWatch log agent can be setup on-premises too**

Amazon EventBridge (CloudWatch Events)?

• Events (or EventBridge): **react to events** in AWS, or trigger **a rule on a schedule**

• **Schedule**: Cron jobs (scheduled scripts)

• Event Pattern: **Event rules to react to a service doing something**

• Events have **rules (for example, trigger when someone logs in). React to events in AWS, configurable, can trigger on schedule**

AWS CloudTrail?

• Provides governance, compliance and audit for your AWS Account

• **Get an history of events / API calls made within your AWS Account by: • Console • SDK • CLI • AWS Services**

• A trail can be applied to **All Regions (default) or a single Region**.

• **CloudTrail: audit/records API calls made within your AWS account**

• **If a resource is deleted in AWS, investigate CloudTrail first!**

• **CloudTrail Insights**: Within CloudTrail, you can also enable CloudTrail Insights. This optional feature allows **CloudTrail to automatically detect unusual API activities in your AWS accoun**t. For example, CloudTrail Insights might detect that a higher number of Amazon EC2 instances than usual have recently been launched in your account. You can then review the full event details to determine which actions you need to take next.

AWS X-Ray?

• **Debugging**: one big monolith "easy", **distributed services "hard**"

• X-Ray: **trace requests** made through your **distributed applications**

• **Mainly to identify performance issues in multiple services model (microservices, SOA…)**

• **Distributed Tracing**

• **Troubleshooting performance**

• **Understand dependencies in a microservice architecture**

•**helps developers analyze and debug production as well as distributed applications**

Amazon CodeGuru?

• **An ML-powered service** for automated **code reviews** and **application performance recommendations**

• Provides two functionalities

  • **CodeGuru Reviewer**: **automated code reviews for static code analysis (development)**

  • **CodeGuru Profiler: visibility/recommendations about application performance during runtime (production)**

•**Automated code review**

•**performance recommendations**

AWS Status - Service Health Dashboard?

• **Shows all regions**, **all services health**

• Service Health Dashboard: **status of all AWS services across all regions**

• **Monitor health of all services across regions**

AWS Personal Health Dashboard?

• AWS Personal Health Dashboard provides **alerts and remediation guidanc**e when **AWS is experiencing** events **that may impact you.**

• Personal Health Dashboard: **AWS events that impact your infrastructure**

•**Provides alerts and guide to remedy the issues** when there are problems with AWS **may affect the user**

•**provides alerts and remediation guidance when AWS is experiencing events that may impact user**

VPC Section

VPC & Subnets Primer*?*

• VPC -Virtual Private Cloud: **private network to deploy your resources (Bound to region)**

• a logically isolated section of AWS, where you can launch AWS resources in a private network

• Subnets allow you to partition your network inside your VPC (Availability Zone)

• A public subnet is a subnet that is accessible from the internet

• A private subnet is a subnet that is not accessible from the internet

• To define access to the internet and between subnets, we use Route Tables

Internet Gateway & NAT Gateways*?*

• **Internet Gateway**: at the **VPC level**, provide Internet Access

• Internet Gateways helps our VPC instances connect with the internet

• Public Subnets have a route to the internet gateway

• **NAT Gateways** (AWS-managed) & NAT Instances (self-managed) **allow your instances in your Private Subnets to access the internet while remaining private**

Network ACL & Security Groups*?*

• NACL (Network ACL):

- **A firewall which controls traffic from and to subnet**
- Are attached at the **Subnet level**
- Can contain rules for IP only
- can have ALLOW/DENY rules **Stateless**

• **Security Groups**

  • A firewall that controls traffic to and from **an ENI / an EC2 Instance**

  • Can only have ALLOW rule **Stateful**

VPC Flow Logs?

- **Provides info about IP traffic in and out of interfaces**
- **VPC Flow Logs: network traffic logs**

VPC Peering?

- **Connect two VPC**
- **Only work with two VPC**
- VPC Peering: Connect two VPC with non overlapping IP ranges, nontransitive
- Connect two VPC, privately using AWS' network
- VPC Peering connection **is not transitive** (must be established for each VPC that need to communicate with one another)

VPC Endpoints?

• **Endpoints allow you to connect** to **AWS Services** using a **private network instead of the publi**c www network

• VPC Endpoints: **Provide private access to AWS Services within VPC**

• **Connect to AWS services using a private network**

• VPC Endpoint **Gateway**: S3 & DynamoDB

• VPC Endpoint **Interface**: the rest

AWS PrivateLink (VPC Endpoint Services)?

• **PrivateLink Privately connect to a service in a 3rd party VPC**

Site to Site VPN & Direct Connect?

- **Site to Site VPN**
    - Connect an on-premises VPN to AWS
    - **The connection is automatically encrypted**
    - **Goes over the public internet**
    - **On-premises**: must use a **Customer Gateway** (CGW)
    - **AWS**: must use a Virtual **Private Gateway** (VGW)
- **Direct Connect (DX)**
- Establish a **physical connection between on-premises and AWS**
- **The connection is private, secure and fast**
- **Goes over a private network**
- **Takes at least a month to establish**

AWS Client VPN?

- Client VPN: **OpenVPN connection** from your **computer into your VPC**
- Connect from **your computer using OpenVPN to your private network in AWS and on-premises**
- Goes over **public Internet**

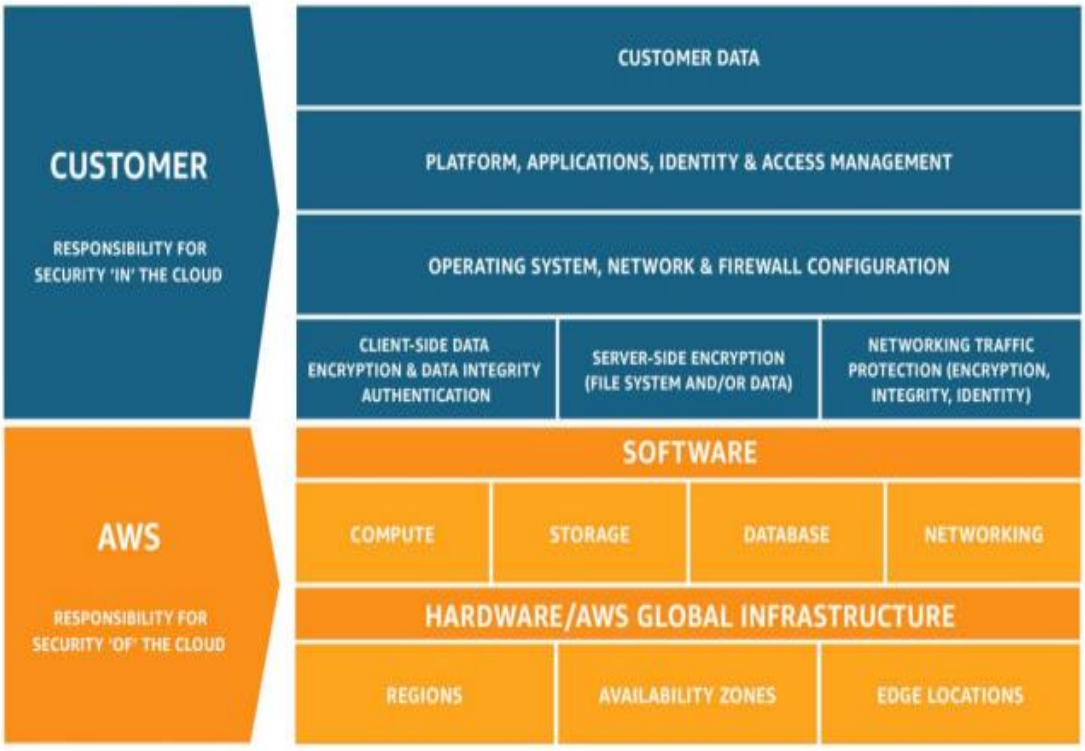
Transit Gateway?

- **For having transitive peering between thousands of VPC and on - premises, hub-and-spoke (star) connection**

Security & Compliance Section

AWS Shared Responsibility Model*?*

- **AWS responsibility - Security of the Cloud**
    - **Protecting infrastructure** (hardware, software, facilities, and networking) that runs all the AWS services
    - **Managed services** like S3, DynamoDB, RDS, etc.
- **Customer responsibility - Security in the Cloud**
    - **For EC2 instance, customer is responsible for management of the guest OS (including security patches and updates), firewall & network configuration, IAM**
    - **Encrypting application data**
- **Shared controls:**
    - **Patch Management**, **Configuration Management, Awareness & Training**



DDOS Protection on AWS?

- AWS Shield standard is available for all users, free of charge
- AWS Shield advanced: premium service with support

AWS WAF?

- protect your web applications or APIs against common **web exploits and bots**
- monitor web requests that are forwarded to Amazon CloudFront distributions or an Application Load Balancer
- block common attack patterns, such as SQL injection or cross-site scripting
- **pricing is based on how many rules you deploy and how many web requests your application receives**
- **AWS WAF: Filter specific requests based on rules**
- **AWS WAF can be deployed on Amazon CloudFront, the Application Load Balancer (ALB), Amazon API Gateway, and AWS AppSync**

Penetration Testing on AWS Cloud?

• AWS customers are welcome to carry out security assessments or penetration tests against their AWS infrastructure without prior approval **for 8 services**

• **Prohibited Activities**

• DNS zone walking via Amazon Route 53 Hosted Zones

• Denial of Service (DoS), Distributed Denial of Service (DDoS), Simulated DoS, Simulated DDoS

• Port flooding

• Protocol flooding

• Request flooding

Data at rest vs. Data in transit?

• **At rest**: **data stored** or **archived on a device On a hard disk, on a RDS instance, in S3 Glacier Deep Archive, etc.**

• **In transit (in motion): data being moved from one location to another , Means data transferred on the network**

AWS KMS (Key Management Service)?

- KMS: Encryption keys **managed by AWS**
- **Encryption Opt-in:**
  - **EBS** volumes: encrypt volumes
  - **S3** buckets: Server-side encryption of objects
  - **Redshift** database: encryption of data
  - RDS database: encryption of data
  - EFS drives: encryption of data
- **Encryption Automatically enabled:**
  - **CloudTrail Logs**
  - **S3 Glacier**
  - **Storage Gateway**

CloudHSM?

- AWS provides hardware
- User manage keys
- CloudHSM: Hardware encryption, we manage encryption keys

AWS Certificate Manager (ACM)?

• Let's you easily provision, manage, and deploy **SSL/TLS Certificates**

AWS Secrets Manager?

- Encrypts secrets at rest using keys that customer owns and store in KMS
- **Secrets are encrypted using KMS**
- **Automatic secret rotation (for example, rotate database password)**
- **To store secrets**

AWS Artifact (not really a service)?

• Portal that provides customers with on-demand access to AWS **compliance documentation and AWS agreements**

• Artifact Reports

• Artifact Agreements

Amazon GuardDuty?

• Intelligent Threat discovery to Protect AWS Account

- **Threat discovery using ML**
- Can protect against **CryptoCurrency** attacks (has a dedicated "finding" for it)

Amazon Inspector?

• Automated Security Assessments

- vulnerability management service that continuously scans your AWS workloads for vulnerabilities and unintended network exposure
- **Only for EC2 and container-related resources (ECR)**

AWS Config?

• Helps with auditing and recording compliance of your AWS resources

• Helps record configurations and changes over time

- **assess, audit, and evaluate the configurations of your AWS resources to check for compliance**

AWS Macie?

• machine learning and pattern matching to discover and protect your **sensitive data** in AWS

AWS Security Hub?

• **Central security tool** to manage security across **several AWS accounts** and automate security checks

- Centralized tool to manage security **across AWS accounts**

Amazon Detective?

• Amazon Detective analyzes, investigates, and quickly identifies **the root cause** of security issues or suspicious activities

AWS Abuse?

• Report suspected AWS resources used for abusive or illegal purposes

• Abusive & prohibited behaviors are:

- Spam
- Port scanning
- DoS or DDoS attacks
- Intrusion attempts
- Hosting objectionable or copyrighted content
- Distributing malware

Root user privileges?

- **Lock away your AWS account root user access keys!**
- Actions that can be performed only by the root user
    - **Change account settings (account name, email address, root user password, root user access keys) • View certain tax invoices**
    - **Close your AWS account**
    - Restore IAM user permissions
    - **Change or cancel your AWS Support plan**
    - **Register as a seller in the Reserved Instance Marketplace**
    - Configure an Amazon S3 bucket to enable MFA
    - Edit or delete an Amazon S3 bucket policy that includes an invalid VPC ID or VPC endpoint ID
    - Sign up for GovCloud

Machine Learning Section

Amazon Rekognition*?*

• Find objects, people, text, scenes in **images and videos using ML**

Amazon Transcribe*?*

• Automatically convert **speech to text**

Amazon Polly*?*

• **Turn text into lifelike speech** using deep learning

Amazon Translate?

• **Natural and accurate language translation**

Amazon Lex & Connect*?*

• **Amazon Lex: (same technology that powers Alexa**)

- **Speech to text**
- **NLP**
- **Chatbot**

• **Amazon Connect:**

• **Receive calls, create contact flows, cloud-based virtual contact center**

- **Build a contact center with Connect**

Amazon Comprehend?

• **For Natural Language Processing – NLP**

Amazon SageMaker*?*

• **Fully managed service for developers / data scientists to build ML models**

- **Build ML models**

Amazon Forecast?

- **Build forecasting models**

Amazon Kendra?

Fully managed **document search service** powered by Machine Learning

Amazon Personalize?

• Fully managed ML-service to **build apps with real-time personalized recommendations**

- **Recommender system based on user's data**

Amazon Textract

• Automatically **extracts text, handwriting, and data from any scanned documents using AI and ML**

- **Extract text from scanned documents**

Account Management, Billing & Support Section

AWS Organizations?

• Global service

• Allows to **manage multiple AWS accounts**

• The main account is the master account

• Restrict account privileges using Service Control Policies (SCP)

Service Control Policies (SCP)?

- Applied at organization account or account level
- Whitelist and blacklist IAM account
- Does not apply to the master account
- Applies to all users and Roles, including root

AWS Organization – Consolidated Billing?

• **Combined Usage** – combine the usage across all AWS accounts in the AWS Organization to **share the volume pricing**, Reserved Instances and Savings Plans discounts

• **One Bill – get one bill for all AWS Accounts in the AWS Organization**

• **Combined usage for better pricing**

• A simple account with 1 bill

AWS Control Tower?

• Easy way to set up and govern a secure and **compliant multi-account** AWS environment based on best practices

Pricing Models in AWS?

• **Pay as you go**: pay for what you use, remain agile, responsive, meet scale demands

• **Save when you reserve**: minimize risks, predictably manage budgets, comply with long-terms requirements

• **Pay less by using more: volume-based discounts**

• **Pay less as AWS grows**

Free services in AWS?

• IAM • VPC • Consolidated Billing • Elastic Beanstalk • CloudFormation • Auto Scaling Groups

free tier in AWS?

• EC2 t2.micro instance for a year • S3, EBS, ELB, AWS Data transfer

Compute Pricing – EC2

• Only charged for what you use

• Number of instances

•Instance configuration: • Physical capacity • Region • OS and software • Instance type • Instance size

Compute Pricing – Lambda & ECS

• Lambda: • Pay per call • Pay per duration

• ECS: • EC2 Launch Type Model: No additional fees, you pay for AWS resources stored and created in your application

• Fargate : • Fargate Launch Type Model: Pay for vCPU and memory resources allocated to your applications in your container s

Storage Pricing – S3

• Number and size of objects

• Number and type of requests

• Data transfer OUT of the S3 region

• S3 Transfer Acceleration

• Lifecycle transitions

Storage Pricing – EBS?

• Volume type

• Storage volume in GB per month provisioned

• IOPS

• Snapshots

• Data transfer: • Outbound

Content Delivery – CloudFront?

• Pricing is different across different geographic regions

Savings Plan?

•Commit a certain $ amount per hour for 1 or 3 years

• Easiest way to setup long-term commitments on AWS

• **EC2 Savings Plan**

  • Commit to usage of individual instance families in a region

• **Compute Savings Plan**

  • Regardless of Family, Region, size, OS, tenancy, compute options

AWS Compute Optimizer?

• Reduce costs and improve performance by recommending optimal AWS resources for your workloads

AWS Pricing Calculator?

• Estimate the cost for your solution architecture

Cost Allocation Tags?

• Use cost allocation **tags to track your AWS costs** on a detailed level

• Cost Allocation Tags: tag resources to create detailed reports

Tagging and Resource Groups?

• Tags are used for organizing resources

Cost and Usage Reports?

• Cost and Usage Reports: most comprehensive billing dataset

Cost Explorer

• Forecast usage up to 12 months based on previous usage

Billing Alarms in CloudWatch?

• Billing data metric is stored in CloudWatch us-east1

AWS Budgets?

• Create budget and send alarms when costs exceeds the budget

Trusted Advisor?

• Analyze your AWS accounts and provides recommendation on 5 categories: **• Cost optimization • Performance • Security • Fault tolerance • Service limits**

Trusted Advisor – Support Plans?

- **Basic Support: free**
    - **Customer Service & Communities**
    - **AWS Trusted Advisor**
    - **AWS Personal Health Dashboard –**
- **AWS Developer Support Plan**
    - **Business hours email access**
- **AWS Business Support Plan**
    - **Trusted Advisor – Full set of checks + API access**
    - **24x7 phone, email, and chat access to Cloud Support Engineers**
- ***AWS Enterprise On-Ramp Support Plan***
    - **Concierge Support Team (for billing and account best practices)**
    - **Access to a pool of Technical Account Managers (TAM)**
    - **Business-critical system down: < 30 minutes**
- **AWS Enterprise Support Plan**
    - **Business-critical system down: < 15 minutes**

*Advanced Identity Section (*Single Sign-On)

AWS STS (SecurityToken Service)?

• Enables you to create **temporary, limited- privileges credentials** to access your AWS resources

Amazon Cognito (simplified)*?*

• **Identity for your Web and Mobile applications users *(login* with*),*** A company just created a new mobile application and wants to add a simple and secure user sign-up, sign-in

What is Microsoft Active Directory (AD)?

• **Directory Services – integrate Microsoft Active Directory in AWS,** A company would like to use their on-premises Microsoft Active Directory to connect to its AWS resources.

IAM Identity Center?

• **One login (single sign-on) for all your AWS accounts in AWS**

Other AWS Services section

Amazon WorkSpaces*?*

• Great to eliminate management of on-premise **VDI** (Virtual Desktop Infrastructure)

Amazon AppStream 2.0

• **Desktop Application Streaming Service within a web browser**

• The application is delivered from **within a web browser**

Amazon Sumerian*?*

•Create and run virtual reality (VR), augmented reality (AR), and 3D applications • Can be used to quickly create 3D models with animations

AWS IoT Core*?*

• AWS IoT Core allows you to easily connect IoT devices to the AWS Cloud

Amazon Elastic Transcoder

• Elastic Transcoder is used **to convert media files stored in S3 into media files in the formats required by consumer playback devices** (phones etc..)

AWS AppSync*?*

• Makes use of **GraphQL**

AWS Amplify*?*

• A set of tools and services that helps you develop and deploy scalable **full stack web and mobile applications**

AWS Device Farm*?*

• Fully-managed service that **tests your web and mobile apps against desktop browsers, real mobile devices, and tablets**

AWS Backup*?*

• Fully-managed service to centrally manage and **automate backups across AWS services**

Disaster Recovery Strategies*?*

• **Backup and Restore *cheap***

•Pilot Light

•Warm Standby

•**Multi-Site / Hot-Site *expensive***

AWS Elastic Disaster Recovery (DRS)?

• Quickly and easily recover your physical, virtual, and cloud-based servers into AWS

AWS DataSync*?*

• **Move large amount of data from on-premises to AWS**

AWS Application Discovery Service*?*

• **Plan migration** projects by **gathering information about on-premises** data centers

AWS Application Migration Service (MGN)*?*

• Lift-and-shift (rehost) solution which **simplify migrating applications to AWS**

AWS Fault Injection Simulator (FIS)?

• A fully managed service for running **fault injection experiments on AWS workloads**

• Based on **Chaos Engineering**

AWS Step Functions?

• Build serverless visual **workflow to orchestrate your Lambda functions**

AWS Ground Station*?*

• Fully managed service that lets you control **sattelite** communications, process data, and scale your satellite operations

Amazon Pinpoint?

• Scalable 2-way (outbound/inbound) **marketing communications service**

AWS Architecting & Ecosystem Section

Well Architected Framework 6 Pillars*?*

• 1) Operational Excellence • 2) Security • 3) Reliability • 4) Performance Efficiency • 5) Cost Optimization • 6) Sustainability

1) Operational Excellence

• Includes the ability to run and **monitor** systems to deliver business value and to continually improve supporting processes and procedures

• **Design Principles**

- **Perform operations as code** - Infrastructure as code
- **Annotate documentation** - Automate the creation of annotated documentation after every build
- **Make frequent, small, reversible changes** - So that in case of any failure, you can reverse it • Refine operations procedures frequently - And ensure that team members are familiar with it
- **Anticipate failure**
- **Learn from all operational failures**

2) Security

• Includes the ability to **protect information, systems, and assets while** delivering business value through risk assessments and mitigation strategies

• Design Principles

- **Implement a strong identity foundation -** Centralize privilege management and reduce (or even eliminate) reliance on long-term credentials - Principle of least privilege - IAM
- **Enable traceability** - Integrate logs and metrics with systems to automatically respond and take action
- **Apply security at all layers** - Like edge network, VPC, subnet, load balancer, every instance, operating system, and application
- **Automate security best practices**

- **Protect data in transit and at rest** - Encryption, tokenization, and access control
- **Keep people away from data** - Reduce or eliminate the need for direct access or manual processing of data
- **Prepare for security events** - Run incident response simulations and use tools with automation to increase your speed for detection, investigation, and recovery
- **Shared Responsibility Model**

3) **Reliability**

• Ability of a system to **recover from infrastructure or service disruptions**, dynamically acquire computing resources to meet demand, and mitigate disruptions such as misconfigurations or transient network issues

• Design Principles

- **Test recovery procedures** - Use automation to simulate different failures or to recreate scenarios that led to failures before
- **Automatically recover from failure** - Anticipate and remediate failures before they occur
- **Scale horizontally to increase aggregate system availability** - Distribute requests across multiple, smaller resources to ensure that they don't share a common point of failure
- **Stop guessing capacity** - Maintain the optimal level to satisfy demand without over or under provisioning - Use Auto Scaling
- Manage change in automation - Use automation to make changes to infrastructure

4) Performance Efficiency

• Includes the ability to use **computing resources efficiently to meet system requirement**s, and to maintain that efficiency as demand changes and technologies evolve

• Design Principles

  • Democratize advanced technologies - Advance technologies become services and hence you can focus more on product development

  • Go global in minutes - Easy deployment in multiple regions

  • Use serverless architectures - Avoid burden of managing servers

  • Experiment more often - Easy to carry out comparative testing

  • Mechanical sympathy - Be aware of all AWS services

5) Cost Optimization

• Includes the ability **to run systems to deliver business value at the lowest price point**

• Design Principles

  • Adopt a consumption mode - Pay only for what you use

  • Measure overall efficiency - Use CloudWatch

  • Stop spending money on data center operations - AWS does the infrastructure part and enables customer to focus on organization projects

  • Analyze and attribute expenditure - Accurate identification of system usage and costs, helps measure return on investment (ROI) - Make sure to use tags

  • Use managed and application level services to reduce cost of ownership

6) Sustainability

• The sustainability pillar focuses **on minimizing the environmental impacts of running cloud workloads.**

• Design Principles

  • Understand your impact – establish performance indicators, evaluate improvements

  • Establish sustainability goals – Set long-term goals for each workload, model return on investment (ROI)

  • Maximize utilization – Right size each workload to maximize the energy efficiency of the underlying hardware and minimize idle resources.

  • Anticipate and adopt new, more efficient hardware and software offerings – and design for flexibility to adopt new technologies over time.

  • Use managed services – Shared services reduce the amount of infrastructure; Managed services help automate sustainability best practices as moving infrequent accessed data to cold storage and adjusting compute capacity.

  • Reduce the downstream impact of your cloud workloads

AWS Right Sizing

• Right sizing is the process of **matching instance types and sizes to your workload performance and capacity requirements at the lowest possible cost**

• Scaling up is easy so always start small

AWS Marketplace

• Digital catalog with thousands of software listings from independent **software vendors (3rd party)**

AWS Professional Services & Partner Network?

• The AWS Professional Services organization is a global team of experts

• **APN Technology Partners**: providing hardware, connectivity, and software

• **APN Consulting Partners**: professional services firm to help build on AWS

• **APN Training Partners**: find who can help you learn AWS

AWS IQ?

• **Quickly find professional help for your AWS projects**

AWS re:Post?

• AWS-**managed Q&A service**

• **Questions from AWS Premium Support customers that do not receive a response from the community are passed on to AWS Support engineers**