AWS

Amazon Web Services (AWS) stands as the leading cloud service provider globally, offering a wide array of cloud computing services. It's the preferred choice for top companies like **Netflix, Airbnb, Spotify, and many more** due to its scalability, reliability, and extensive feature set.

AWS was started in 2006 with 3 types of services such as storage, computing, and messaging. After, it enhanced its network by providing all the required services based on the market trends.

AWS provides the **fundamental components** crucial for cloud computing such as EC2, S3, RDS, IAM.

EC2 provides scalable computing capabilities, **S3** offers storage for all kinds of files, **RDS** manages many kinds of databases, and **IAM** ensures secured access through addressing Authentication and Authorization. These components collectively empower the users to create and deploy various applications seamlessly.

EC2 stands for Elastic Compute Cloud. EC2 is an on-demand computing service on the AWS cloud platform. Under computing, it includes all the services a computing device can offer to you along with the flexibility of a virtual environment. It also allows the user to configure their instances as per their requirements i.e. allocate the RAM, ROM, and storage according to the need of the current task. Even the user can dismantle the virtual device once its task is completed and it is no more required. For providing, all these scalable resources AWS charges some bill amount at the end of every month, the bill amount is entirely dependent on your usage. EC2 allows you to rent virtual computers. The provision of servers on AWS Cloud is one of the easiest ways in EC2. EC2 has resizable capacity. EC2 offers security, reliability, high performance, and cost-effective infrastructure so as to meet the demanding business needs.

Amazon S3 is a Simple Storage Service in AWS that stores files of

different types like Photos, Audio, and Videos as Objects providing more scalability and security to. It allows the users to store and retrieve any amount of data at any point in time from anywhere on the web. It facilitates features such as extremely high availability, security, and simple connection to other AWS Services.

Amazon RDS or **Amazon Relational Database Service.** Amazon RDS is nothing but a cloud database, that typically runs on AWS or Amazon Web Services platform and access to the database is provided as-a-service. In simpler words we can state that RDS comes under PaaS i.e. (Platform as a service).

Benefits or Advantages

- Easier Management
- Higher scalability
- Available and durable in nature
- Faster and Securer
- Inexpensive

AWS Identity and Access Management (IAM) is a web service that helps you securely control access to AWS resources. With IAM, you can centrally manage permissions that control which AWS resources users can access. You use IAM to control who is authenticated (signed in) and authorized (has permissions) to use resources.

S3 (Simple Storage Service) is an object storage service suitable for storing various data types of files that can accessed through the internet.

EBS (Elastic Block storage) is a block-level storage attached to EC2 instances, offering persistent and high-performance storage for applications like databases. EBS provides the raw storage hardware helpful for I/O operations where as S3 comes with pre configured file system. For understaning think of S3 as a file storage system and EBS as a hard drive.

Auto Scaling is an aws service that provides dynamically adjusts, on running the number of EC2 instances based on traffic demand. For instance, during the high traffic periods, **Auto Scaling adds instances** , improving optimal performance as per the policies configuration. Conversely, while during low traffic, it will reduce the number of instances , optimizes the cost efficiency maintaining high availability.

Elastic Load balancer (ELB) is a service provided by AWS that helps in distribution of incoming traffic of the applications across multi targets such as EC2 instances, containers etc. in one or more Availability zones. It helps in improving fault tolerance and ensuring the utilization of resources, bringing high availability of the application by preventing a single node (instance) faulterance by improving application's resilience.

- **How Is Data Transfer handled in AWS?**

  The **data transfer** in AWS happens in between regions, within regions, and between the services. It is essential to consider that these data transfer comes with costs when designing the architectures. For example, transfer of the data between an EC2 instance and an S3 bucket within the same region is often free, but the transfer of data in between inter-region comes with charges.

Amazon VPC (Virutal Private Cloud) is an AWS service that helps the users to create isolated networks within AWS account through customizing IP address ranges and the defining their subnets. It helps in enhancing the security through controlling both the inbound and outbound of the traffic.

The **Domain Name System** is built using a distributed architecture. When the host needs to resolve the IP address of a domain name, the host device hands over this process to a DNS server. The DNS server finds the IP address of the domain name and returns it back to the host.

Amazon Route 53 is an aws service that offers DNS web services which are scalable. It helps in guranteeing dependable , low-latency routing to the AWS services through facilitating efficient translation of user-friendly domain names into IP addresses.

- **How Does AWS Handle Disaster Recovery and Backup?**

AWS comes up with various services for disaster recovery and backup. Amazon S3 service is the most perferable service for backup storage and centralized management. Additionally it supports in business continuity in the event of a disaster by replicating AWS workloads to on-premises.

AWS Elastic Beanstalk is a AWS managed service helps in providing simplifyed application's deployment and management through automatically handling the infrastructure provision. It allows the developers to focus completely on writing the code. For example, you only need to upload your code for deploying web application , Elastic Beanstalk will care of the rest of underlying infrastructures provisioning of EC2 instances and load balancing.

What is Cloud ---> Global network of servers

"The cloud" refers to servers that are accessed over the Internet, and the software and databases that run on those servers. Cloud servers are located in data centers all over the world. By using cloud computing, users and companies do not have to manage physical servers themselves or run software applications on their own machines.

Cloud Computing

Cloud computing is the on-demand delivery of various computing services over the internet. These services include servers, storage, databases, networking, software, analytics, and intelligence. Instead of owning and maintaining physical data centers and servers, businesses and individuals can access these resources as needed, paying only for

what they use.

Public cloud

Public clouds deliver resources (such as compute, storage, network, and applications) over the internet. They are owned and operated by third-party cloud service providers (e.g., Google Cloud, Microsoft Azure).

Accessibility: Accessible to anyone over the internet.

Private Cloud

Private clouds are built, run, and used by a single organization. They are typically located on-premises or in a dedicated data center.

Control: Greater control, customization, and data security.

Hybrid Cloud

Hybrid clouds combine at least one private computing environment (on-premises or private cloud) with one or more public clouds.

Hybrid Cloud is a mix of public and private cloud.

Why public cloud is so popular ?

the public cloud's flexibility, efficiency, and accessibility make it a popular choice for organizations across industries.

Why AWS ?

AWS's combination of services, scalability, reliability, and cost-effectiveness has made it the go-to choice for cloud computing across industries and company sizes. AWS is a public cloud platform.

EC2 Instance

An EC2 instance is essentially a virtual server in Amazon's Elastic

Compute Cloud (EC2). When you launch an EC2 instance, the instance type you specify determines the hardware available to your instance. Each instance type offers a different balance of compute, memory, network, and storage resources1. These instances allow you to run applications in a scalable and flexible manner within the AWS cloud.

Why EC2 ?

Amazon EC2 provides scalable computing capacity in the AWS cloud. Leveraging it enables organizations to develop and deploy applications faster, without needing to invest in hardware upfront. Users can launch virtual servers, configure security and networking, and manage resources from an intuitive dashboard1. It's a flexible and cost-effective way to run applications in the cloud.

Different types of EC2 instances

1. General Purpose Instances: These offer a balance of compute, memory, and networking resources. They're ideal for applications like web servers and code repositories.

Notable instance types: M7g, M6g, M5, and T31.

2. Compute Optimized Instances: Designed for compute-intensive workloads, these instances prioritize CPU performance.

Example: C6g1.

3. Memory-Optimized Instances: These are great for memory-intensive applications like databases and analytics.

Notable instance types: R6g, R5, and X2gd1.

4. Storage-Optimized Instances: Optimized for high-speed storage, they suit data-intensive tasks.

Example: I3en1.

5. Accelerated Computing Instances: These instances leverage GPUs or FPGAs for specialized workloads like machine learning and scientific simulations.

Notable instance types: P4, G4, and Inf11.

Remember, each instance type has specific vCPUs, memory, and

storage options, allowing you to choose the right fit for your workload.

Subnet -- A subnet is a range of IP addresses in your VPC. You can create AWS resources, such as EC2 instances, in specific subnets.

Subnet Types:

1. Public subnet – The subnet has a direct route to an internet gateway. Resources in a public subnet can access the public internet.

2. Private subnet – The subnet does not have a direct route to an internet gateway. Resources in a private subnet require a NAT device to access the public internet.

3. VPN-only subnet – The subnet has a route to a Site-to-Site VPN connection through a virtual private gateway. The subnet does not have a route to an internet gateway.

4. Isolated subnet – The subnet has no routes to destinations outside its VPC. Resources in an isolated subnet can only access or be accessed by other resources in the same VPC.

Putty -- Connect to a remote server

PuTTY is a terminal emulator application that can act as a client for the SSH, Telnet, rlogin, and raw TCP computing protocols. It is used to connect to a remote server, such as a Linux-based web server. PuTTY also comes with a secure FTP client (SFTP) called PSFTP, which allows you to upload and download files securely between computers.

Security Group: It performs the function of a virtual firewall, managing the inbound and outbound traffic for one or more Amazon EC2 instances or other AWS services within a VPC.

Inbound Rules: These outline the types of traffic that are permitted to use the resources. It serves as a virtual firewall, controlling the

traffic going in and coming out of a VPC for one or more Amazon EC2 instances or other AWS services.

Outbound Rules: These regulate the traffic that is permitted to depart from the resources. The destination for incoming traffic is dealt with by outbound rules. They may be forwarded to an alternative Security Group, a CIDR block, a single IPv4 or IPv6 address, or all three.

NACL -- Network Access Control List

NACL is defined by default for every Virtual Private Network(VPC). However, you can create custom NACL according to your requirements. These NACL define inbound and outbound rule for subnets present in VPC. These have almost the same function as Security Groups the only difference is NACL works in subnet levels and Security groups are defined for instances and other resources.

NACL is used to allow traffic or deny traffic whereas Security group is used to allow not able to deny.

VPC --- Virtual Private Cloud

Amazon Virtual Private Cloud is a service that allows its users to launch their virtual machines in a protected as well as isolated virtual environment defined by them. You have complete control over your VPC, from creation to customization and even deletion. It's applicable to organizations where the data is scattered and needs to be managed well. In other words, VPC enables us to select the virtual address of our private cloud and we can also define all the sub-constituents of the VPC like subnet, subnet mask, availability zone, etc on our own.

Amazon S3 - Amazon S3 is a Simple Storage Service in AWS that stores files of different types like Photos, Audio, and Videos as Objects providing more scalability and security to. It allows the users to store and retrieve any amount of data at any point in time from anywhere on

the web. It facilitates features such as extremely high availability, security, and simple connection to other AWS Services.

Amazon S3 is used for various purposes in the Cloud because of its robust features with scaling and Securing of data. It helps people with all kinds of use cases from fields such as Mobile/Web applications, Big data, Machine Learning and many more.

Amazon S3 Bucket -- Amazon S3 bucket is a fundamental Storage Container feature in AWS S3 Service. It provides a secure and scalable repository for storing of Objects such as Text data, Images, Audio and Video files over AWS Cloud. Each S3 bucket name should be named globally unique and should be configured with ACL (Access Control List).

Data, in S3, is stored in containers called buckets. Each bucket will have its own set of policies and configurations. This enables users to have more control over their data. Bucket Names must be unique. Can be thought of as a parent folder of data. There is a limit of 100 buckets per AWS account. But it can be increased if requested by AWS support.

S3 Versioning: Versioning means always keeping a record of previously uploaded files in S3. Points to Versioning are not enabled by default. Once enabled, it is enabled for all objects in a bucket. Versioning keeps all the copies of your file, so, it adds cost for storing multiple copies of your data. For example, 10 copies of a file of size 1GB will have you charged for using 10GBs for S3 space.

Bucket Policies: A document for verifying the access to S3 buckets from within your AWS account, controls which services and users have what kind of access to your S3 bucket. Each bucket has its own Bucket Policies.

Access control lists (ACLs): A document for verifying access to S3 buckets from outside your AWS account. An ACL is specific to each bucket.

How to access Amazon S3 bucket

You can work and access the Amazon S3 bucket by using any one of the following methods

AWS Management Console

AWS CLI Commands

Programming Scripts (Using boto3 library of Python)

Amazon S3 features

1. Scalability: S3 can handle any amount of data, and you can easily scale your storage needs up or down.

2. Data Availability: S3 ensures high availability, meaning your data is accessible from anywhere in the world.

3. Security: By default, S3 buckets and objects are private. You can configure access controls, encryption, and authentication to secure your data1.

4. Object-Based Storage: S3 stores data as objects within “buckets.” Each object can be up to 5 terabytes in size.

5. Metadata and Tags: You can append metadata tags to objects, making it easier to organize and manage your data.

6. Storage Classes: S3 offers different storage classes (e.g., Standard, Intelligent-Tiering, Glacier) to optimize cost and performance based on your use case.

7. Big Data Analytics: S3 integrates with analytics services like Amazon Athena and Amazon Redshift for data analysis.

8. Monitoring and Reporting: You can monitor storage usage, access patterns, and trends at the object and bucket levels.

9. Versioning: S3 supports data version control, allowing you to preserve and retrieve every version of an object.

10. S3 Batch Operations: Manage data at any scale by copying objects, modifying tags, or running custom business logic across objects1.

Route 53:

Amazon Route 53 is a highly available and scalable cloud Domain Name System (DNS) web service. It is designed for developers and corporations to route the end users to Internet applications by

translating human-readable names like www.karthik.com into the numeric IP addresses like 192.0.1.1 that computers use to connect. You cannot use Amazon Route 53 to connect your on-premises network with AWS Cloud.

Route53 also performs "Health Checks".

Amazon Route 53 health checks monitor the health of your resources, such as web servers and email servers. You can configure health checks to monitor specific endpoints, such as web servers, and set criteria for their health status

Functions Of Route53

If a web application requires a domain name, Route53 service helps to register the name for the website (i.e domain name).

Whenever a user enters the domain name, Route53 helps to connect the user to the website.

If any failure is detected at any level, it automatically routes the user to a healthy resource.

Amazon Route 53 is cost effective, secure and scalable.

Amazon Route 53 is flexible, highly available and reliable.

UAT

SDFC life cycle

rightsize,scaling

AWS Cloud Practitioner

What is a client-server model?

In computing, a client can be a web browser or desktop application that a person interacts with to make requests to computer servers. A server can be services, such as Amazon Elastic Compute Cloud (Amazon EC2) – a type of virtual server.

for example client go to coffee shop and requests for coffee and the maker is a server he gives the coffee / resource to the client.

Deployment models for cloud computing

The three cloud computing deployment models are cloud-based, on-premises, and hybrid.

1. In a **cloud-based deployment** model, you can migrate existing applications to the cloud, or you can design and build new applications in the cloud. You can build those applications on low-level infrastructure that requires your IT staff to manage them. Alternatively, you can build them using higher-level services that reduce the management, architecting, and scaling requirements of the core infrastructure.

2. **On-premises deployment** is also known as a private cloud deployment. In this model, resources are deployed on premises by using virtualization and resource management tools.

For example, you might have applications that run on technology that is fully kept in your on-premises data center. Though this model is much like legacy IT infrastructure, its incorporation of application management and virtualization technologies helps to increase resource utilization.

3. In a **hybrid deployment**, cloud-based resources are connected to on-premises infrastructure. You might want to use this approach in a number of situations. For example, you have legacy applications that are better maintained on premises, or government regulations require your business to keep certain records on premises.

What is Cloud Computing?

On-demand delivery of IT resources and applications through the internet with pay-as-you-go pricing.

Amazon Elastic Compute Cloud (Amazon EC2) provides secure, resizable compute capacity in the cloud as Amazon EC2 instances.

You can provision and launch an Amazon EC2 instance within minutes.

You can stop using it when you have finished running a workload.

You pay only for the compute time you use when an instance is running, not when it is stopped or terminated.

You can save costs by paying only for server capacity that you need or want.

How Amazon EC2 works ----> Launch, Connect, Use

Amazon EC2 instance types are optimized for different tasks. When selecting an instance type, consider the specific needs of your workloads and applications. This might include requirements for compute, memory, or storage capabilities.

- General Purpose Instances
- Compute Optimized Instances
- Memory Optimized instances
- Accelerating Computing Instances
- Storage Optimized Instances

General purpose instances provide a balance of compute, memory, and networking resources. You can use them for a variety of workloads, such as:

application servers

gaming servers

backend servers for enterprise applications

small and medium databases

Compute optimized instances are ideal for compute-bound applications that benefit from high-performance processors. Like general purpose instances, you can use compute optimized instances for workloads such as web, application, and gaming servers.

Memory optimized instances are designed to deliver fast performance for workloads that process large datasets in memory. In computing,

memory is a temporary storage area. It holds all the data and instructions that a central processing unit (CPU) needs to be able to complete actions. Before a computer program or application is able to run, it is loaded from storage into memory.

Accelerated computing instances use hardware accelerators, or coprocessors, to perform some functions more efficiently than is possible in software running on CPUs. Examples of these functions include floating-point number calculations, graphics processing, and data pattern matching.

Storage optimized instances are designed for workloads that require high, sequential read and write access to large datasets on local storage. Examples of workloads suitable for storage optimized instances include distributed file systems, data warehousing applications, and high-frequency online transaction processing (OLTP) systems.

Amazon EC2 pricing

With Amazon EC2, you pay only for the compute time that you use. Amazon EC2 offers a variety of pricing options for different use cases.

- On-Demand
- Reserved Instances
- EC2 Instance Savings Plan
- Spot Instances
- Dedicated Hosts

On-Demand Instances are ideal for short-term, irregular workloads that cannot be interrupted. No upfront costs or minimum contracts apply. The instances run continuously until you stop them, and you pay for only the compute time you use.

Reserved Instances are a billing discount applied to the use of On-

Demand Instances in your account. There are two available types of Reserved Instances:

- Standard Reserved Instances
- Convertible Reserved Instances

You can purchase Standard Reserved and Convertible Reserved Instances for a 1-year or 3-year term. You realize greater cost savings with the 3-year option.

Standard Reserved Instances: This option is a good fit if you know the EC2 instance type and size you need for your steady-state applications and in which AWS Region you plan to run them.

Convertible Reserved Instances: If you need to run your EC2 instances in different Availability Zones or different instance types, then Convertible Reserved Instances might be right for you.

EC2 Instance Savings Plans reduce your EC2 instance costs when you make an hourly spend commitment to an instance family and Region for a 1-year or 3-year term. This term commitment results in savings of up to 72 percent compared to On-Demand rates.

Spot Instances are ideal for workloads with flexible start and end times, or that can withstand interruptions. Spot Instances use unused Amazon EC2 computing capacity and offer you cost savings at up to 90% off of On-Demand prices.

Dedicated Hosts are physical servers with Amazon EC2 instance capacity that is fully dedicated to your use.

You can use your existing per-socket, per-core, or per-VM software licenses to help maintain license compliance. You can purchase On-Demand Dedicated Hosts and Dedicated Hosts Reservations. Of all the Amazon EC2 options that were covered, **Dedicated Hosts are the most expensive.**

Scalability involves beginning with only the resources you need and designing your architecture to automatically respond to changing demand by scaling out or in. As a result, you pay for only the resources you use. You don’t have to worry about a lack of computing capacity to meet your customers’ needs.

The AWS service that provides this functionality for Amazon EC2 instances is **Amazon EC2 Auto Scaling.**

Amazon EC2 Auto Scaling enables you to automatically add or remove Amazon EC2 instances in response to changing application demand. By automatically scaling your instances in and out as needed, you can maintain a greater sense of application availability.

Within Amazon EC2 Auto Scaling, you can use two approaches: dynamic scaling and predictive scaling.

Dynamic scaling responds to changing demand.

Predictive scaling automatically schedules the right number of Amazon EC2 instances based on predicted demand.

Scalability refers to the capacity of a software solution to manage rising workloads. In simple terms, "it is the ability of a system to readily add extra processing resources to handle the increased loads".

Scaling Amazon EC2 means you start with the resources you require at the time of starting your service and build your architecture to automatically scale in or out, in response to the changing demand. As a result, you only pay for the resources you really utilize.

When you create an **Auto Scaling group**, you can set the minimum number of Amazon EC2 instances. The minimum capacity is the number of Amazon EC2 instances that launch immediately after you have created the Auto Scaling group. In this example, the Auto Scaling group has a minimum capacity of one Amazon EC2 instance.

Amazon EC2 Auto Scaling uses Amazon EC2 instances, you pay for only the instances you use, when you use them. You now have a cost-effective architecture that provides the best customer experience while reducing expenses.

When you have multiple EC2 instances all running the same program, to serve the same purpose, and a request comes in, how does that request know which EC2 instance to go to? How can you ensure there's an even distribution of workload across EC2 instances? So not just one is backed up while the others are idle sitting by. You need a way to

route requests to instances to process that request. What you need to solve this is called **load balancing.**

A **load balancer** is an application that takes in requests and routes them to the instances to be processed. Now, there are many off the shelf load balancers that work great on AWS.

A load balancer acts as a single point of contact for all incoming web traffic to your Auto Scaling group.

Elastic Load Balancing is the AWS service that automatically distributes incoming application traffic across multiple resources, such as Amazon EC2 instances.

Elastic Load Balancing and Amazon EC2 Auto Scaling are separate services, they work together to help ensure that applications running in Amazon EC2 can provide high performance and availability.

Tightly Coupled Architecture is where if a single component fails or changes, it causes issues for other components or even the whole system. For example, if we have Application A and it is sending messages directly to Application B, if Application B has a failure and cannot accept those messages, Application A will begin to see errors as well. This is a tightly coupled architecture.

"To help maintain application availability when a single component fails, you can design your application through a microservices approach"

We introduced a **Message queue**. Messages are sent into the queue by Application A and they are processed by Application B. If Application B fails, Application A doesn't experience any disruption. Messages being sent can still be sent to the queue and will remain there until they are eventually processed. This is **Loosely coupled**. This is what we strive to achieve with architectures on AWS. And this brings me to two AWS services that can assist in this regard. Amazon Simple Queue Service or **SQS** and Amazon Simple Notification Service or **SNS**.

Amazon Simple Notification Service (Amazon SNS) is a publish/subscribe service. Using Amazon SNS topics, a publisher publishes messages to subscribers. This is similar to the coffee shop;

the cashier provides coffee orders to the barista who makes the drinks.

In Amazon SNS, subscribers can be web servers, email addresses, AWS Lambda functions, or several other options.

Amazon Simple Queue Service (Amazon SQS) is a message queuing service.

Using Amazon SQS, you can send, store, and receive messages between software components, without losing messages or requiring other services to be available. In Amazon SQS, an application sends messages into a queue. A user or service retrieves a message from the queue, processes it, and then deletes it from the queue.

Serverless Computing

If you have applications that you want to run in Amazon EC2, you must do the following:

*Provision instances (virtual servers).

*Upload your code.

*Continue to manage the instances while your application is running.

The term “**serverless**” means that your code runs on servers, but you do not need to provision or manage these servers. With serverless computing, you can focus more on innovating new products and features instead of maintaining servers.

An AWS service for serverless computing is **AWS Lambda**.

AWS Lambda is a service that lets you run code without needing to provision or manage servers.

While using AWS Lambda, "you pay only for the compute time that you consume. Charges apply only when your code is running". You can also run code for virtually any type of application or backend service, all with zero administration.

How AWS Lambda works

***You upload your code to Lambda.

*You set your code to trigger from an event source, such as AWS services, mobile applications, or HTTP endpoints.

*Lambda runs your code only when triggered.

*You pay only for the compute time that you use. In the previous example of resizing images, you would pay only for the compute time that you use when uploading new images. Uploading the images triggers Lambda to run code for the image resizing function.

"In AWS, you can also build and run containerized applications".

Containers provide you with a standard way to package your application's code and dependencies into a single object. You can also use containers for processes and workflows in which there are essential requirements for security, reliability, and scalability.

-> One host with multiple containers

Example of a container that includes apps, bins/libs, an operating system, and a server.

-> Tens of Hosts with Hundreds of Containers

Example of scaling up to tens of hosts with hundreds of containers.

Amazon Elastic Container Service (Amazon ECS) is a highly scalable, high-performance container management system that enables you to run and scale containerized applications on AWS.

Amazon ECS **supports Docker containers**. Docker is a software platform that enables you to build, test, and deploy applications quickly. With Amazon ECS, you can use API calls to launch and stop Docker-enabled applications.

Amazon Elastic Kubernetes Service (Amazon EKS) is a fully managed service that you can use to run Kubernetes on AWS.

Kubernetes is open-source software that enables you to deploy and manage containerized applications at scale. A large community of volunteers maintains Kubernetes, and AWS actively works together with the Kubernetes community. As new features and functionalities release for Kubernetes applications, you can easily apply these updates to your applications managed by Amazon EKS.

AWS Fargate is a serverless compute engine for containers. It works with both Amazon ECS and Amazon EKS.

When using AWS Fargate, you do not need to provision or manage servers. AWS Fargate manages your server infrastructure for you. You can focus more on innovating and developing your applications, and you pay only for the resources that are required to run your containers.

The main **difference between** these two serverless compute platforms is that **AWS Fargate** takes care of the underlying VMs, networking, and other resources you need to run containers using ECS or EKS, whereas **AWS Lambda** lets you run standalone, stateless functions without having to consider any of the infrastructure whatsoever.

REGION

Amazon cloud computing resources are hosted in multiple locations world-wide. These locations are composed of **AWS Regions**, Availability Zones, and Local Zones. "Each AWS Region is a separate geographic area". Each AWS Region has multiple, isolated locations known as Availability Zones. Each AWS Region is designed to be isolated from the other AWS Regions.

Four Business Factors before choosing Regions

Compliance: Depending on your company and location, you might need to run your data out of specific areas. For example, if your company requires all of its data to reside within the boundaries of the

UK, you would choose the London Region.

Not all companies have location-specific data regulations, so you might need to focus more on the other three factors.

Proximity: Selecting a Region that is close to your customers will help you to get content to them faster. For example, your company is based in Washington, DC, and many of your customers live in Singapore. You might consider running your infrastructure in the Northern Virginia Region to be close to company headquarters, and run your applications from the Singapore Region.

Features Availability: Sometimes, the closest Region might not have all the features that you want to offer to customers. AWS is frequently innovating by creating new services and expanding on features within existing services. However, making new services available around the world sometimes requires AWS to build out physical hardware one Region at a time.

Pricing: Suppose that you are considering running applications in both the United States and Brazil. The way Brazil's tax structure is set up, it might cost 50% more to run the same workload out of the São Paulo Region compared to the Oregon Region. You will learn in more detail that several factors determine pricing, but for now know that the cost of services can vary from Region to Region.

"AWS has very transparent Granular Pricing"

"Each Region has different Price Sheet"

Availability Zone is a single data center or a group of data centers within a Region. Availability Zones are located tens of miles apart from each other. This is close enough to have low latency (the time between when content requested and received) between Availability Zones. However, if a disaster occurs in one part of the Region, they are distant enough to reduce the chance that multiple Availability Zones are affected.

if you have customers in Mumbai who need access to your data, but the data is hosted out of the Tokyo Region, rather than having all the Mumbai-based customers, send requests all the way to Tokyo, to access the data, just place a copy locally or cache a copy in Mumbai.

Caching copies of data closer to the customers all around the world uses the concept of **content delivery networks, or CDNs**.

CDNs are commonly used, and on AWS, we call our CDN **Amazon CloudFront**. **Amazon CloudFront** is a service that helps deliver data, video, applications, and APIs to customers around the world with low latency and high transfer speeds. Amazon CloudFront uses what are called Edge locations, all around the world, to help accelerate communication with users, no matter where they are.

Edge locations are separate from Regions, so you can push content from inside a Region to a collection of Edge locations around the world, in order to accelerate communication and content delivery. AWS Edge locations, also run more than just CloudFront. They run a domain name service, or DNS, known as **Amazon Route 53**, helping direct customers to the correct web locations with reliably low latency.

AWS Outposts, where AWS will basically install a fully operational mini Region, right inside your own data center. That's owned and operated by AWS, using 100% of AWS functionality, but isolated within your own building.

Edge locations

An edge location is a site that Amazon CloudFront uses to store cached copies of your content closer to your customers for faster delivery.

- Regions are Geographically isolated areas.
- Regions contains Availability Zones.
- Edge Locations run Amazon CloudFront.

AWS Management Console is a web-based interface for accessing and managing AWS services. You can quickly access recently used

services and search for other services by name, keyword, or acronym. The console includes wizards and automated workflows that can simplify the process of completing tasks.

You can also use the AWS Console mobile application to perform tasks such as monitoring resources, viewing alarms, and accessing billing information. Multiple identities can stay logged into the AWS Console mobile app at the same time.

To save time when making API requests, you can use the **AWS Command Line Interface (AWS CLI).** AWS CLI enables you to control multiple AWS services directly from the command line within one tool. AWS CLI is available for users on Windows, macOS, and Linux.

By using AWS CLI, you can automate the actions that your services and applications perform through scripts. For example, you can use commands to launch an Amazon EC2 instance, connect an Amazon EC2 instance to a specific Auto Scaling group, and more.

AWS Elastic Beanstalk

With AWS Elastic Beanstalk, you provide code and configuration settings, and Elastic Beanstalk deploys the resources necessary to perform the following tasks:

Adjust capacity

Load balancing

Automatic scaling

Application health monitoring

AWS CloudFormation, you can treat your infrastructure as code. This means that you can build an environment by writing lines of code instead of using the AWS Management Console to individually provision resources.

AWS CloudFormation provisions your resources in a safe, repeatable manner, enabling you to frequently build your infrastructure and applications without having to perform manual actions. It determines the right operations to perform when managing your stack and rolls back

changes automatically if it detects errors.

Amazon CloudFront is a content delivery service. It uses a network of edge locations to cache content and deliver content to customers all over the world. When content is cached, it is stored locally as a copy. This content might be video files, photos, webpages, and so on.

AWS Outposts is a service that enables you to run infrastructure in a hybrid cloud approach.

AWS Fargate is a serverless compute engine for containers.

Amazon Simple Queue Service (Amazon SQS) is a service that enables you to send, store, and receive messages between software components through a queue.

Amazon Virtual Private Cloud (Amazon VPC)

A VPC, or Virtual Private Cloud, is essentially your own private network in AWS. A VPC allows you to define your private IP range for your AWS resources, and you place things like EC2 instances and ELBs inside of your VPC.

A networking service that you can use to establish boundaries around your AWS resources is Amazon Virtual Private Cloud (Amazon VPC) (opens in a new tab).

Within a virtual private cloud (VPC), you can organize your resources into **subnets**. A subnet is a section of a VPC that can contain resources such as Amazon EC2 instances.

Internet gateway

To allow public traffic from the internet to access your VPC, you attach an internet gateway to the VPC. Without an internet gateway, no one can access the resources within your VPC.

Virtual Private Gateway

To access private resources in a VPC, you can use a virtual private gateway.

A virtual private gateway enables you to establish a virtual private network (VPN) connection between your VPC and a private network, such as an on-premises data center or internal corporate network. A virtual private gateway allows traffic into the VPC only if it is coming from an approved network.

AWS Direct Connect

AWS Direct Connect is a service that lets you to establish a dedicated private connection between your data center and a VPC.

The private connection that AWS Direct Connect provides helps you to reduce network costs and increase the amount of bandwidth that can travel through your network.

Subnets

A subnet is a section of a VPC in which you can group resources based on security or operational needs. Subnets can be public or private.

Public Subnet

Public subnets contain resources that need to be accessible by the public, such as an online store’s website.

Private Subnet

Private subnets contain resources that should be accessible only through your private network, such as a database that contains customers’ personal information and order histories.

Network Access Control Lists (Nacls) - Stateless

A network ACL is a virtual firewall that controls inbound and outbound traffic at the subnet level.

It means the network ACL has a particular list of who can allow into subnet. If someone requests for some data it will be sent on a packet then the packet is on the Network Access Control List it will allows to

the subnet or else it will not allows the packet.

When we send a packet from one instance to another instance, It will go to another instance after it will returning the own instance it will checks again it will checks again it will be on the list or not. This is why NACL is **Stateless.**

Security Groups - Stateful

Every EC2 instance, when it's launched, automatically comes with a security group. And by default, the security group does not allow any traffic into the instance at all. All ports are blocked; all IP addresses sending packets are blocked. That's very secure, but perhaps not very useful.

If you want an instance to actually accept traffic from the outside, then you need to modify the security group to accept a specific type or traffic.

Security groups, by default, allow all return traffic. So they don't have to check a list to see if they're allowed out. Instead, they automatically allow the return traffic to pass by no matter what.

The security group recognizes the packet from before. So it doesn't need to check to see if it's allowed in. Come on home.

Security Group Vs Network ACLs

The key difference between a security group and a network ACL is the security group is stateful, meaning, as we talked about, it has some kind of a memory when it comes to who to allow in or out, and the network ACL is stateless, which remembers nothing and checks every single packet that crosses its border regardless of any circumstances.

DNS (Domain Name System)

DNS, or the Domain Name System, translates human readable domain names (for example, www.amazon.com) to machine readable IP addresses (for example, 192.0.2.44).

Amazon Route 53

Amazon Route 53(opens in a new tab) is a DNS web service. It gives developers and businesses a reliable way to route end users to internet applications hosted in AWS.

Instance Stores - Temporary data bcz when the instance is stoppped or terminated and in the instance store it will be removed all the data.

Block-level storage volumes behave like physical hard drives.

An instance store provides temporary block-level storage for an Amazon EC2 instance. An instance store is disk storage that is physically attached to the host computer for an EC2 instance, and therefore has the same lifespan as the instance. When the instance is terminated, you lose any data in the instance store.

AWS recommends instance stores for use cases that involve temporary data that you do not need in the long term.

That's why Amazon introduces the Amazon EBS

Amazon EBS (Elastic Bulk Store) - An Amazon EBS volume stores data in a single Availability Zone.

Amazon Elastic Block Store (Amazon EBS) is a service that provides block-level storage volumes that you can use with Amazon EC2 instances. If you stop or terminate an Amazon EC2 instance, all the data on the attached EBS volume remains available.

To create an EBS volume, you define the configuration (such as volume size and type) and provision it. After you create an EBS volume, it can attach to an Amazon EC2 instance.

Amazon S3 (Simple Storage Service)

Amazon Simple Storage Service is a service that provides object-level storage. Amazon S3 stores data as objects in buckets.

You can upload any type of file to Amazon S3, such as images, videos, text files, and so on. For example, you might use Amazon S3 to store backup files, media files for a website, or archived documents. Amazon S3 offers unlimited storage space. The maximum file size for an object in Amazon S3 is 5 TB.

In the regional object storage corner, weighing in at unlimited storage, with individual objects at 5,000 gigabytes in size, they specialize in write once/read many, they are 99 .999 999 999% durable, they are Amazon Simple Storage Service!

Amazon S3 Storage Classes

With Amazon S3, you pay only for what you use. You can choose from a range of storage classes(opens in a new tab) to select a fit for your business and cost needs. When selecting an Amazon S3 storage class, consider these two factors:

- How often you plan to retrieve your data
- How available you need your data to be

1. S3 Standard - Amazon S3 costs more for infrequently accessed data and archival storage

- Designed for frequently accessed data
- Stores data in a minimum of three Availability Zones

2. S3 Standard Infrequent Access - same as s3 standard but the difference is low storage price and higher retrival price

- Ideal for infrequently accessed data
- Similar to Amazon S3 Standard but has a lower storage price and higher retrieval price

3. S3 One Zone Infrequent Access - compared to s3 standard and s3 standard infrequent access, the s3 one zone infrequent access

has only one availabilty zone

- Stores data in a single Availability Zone
- Has a lower storage price than Amazon S3 Standard-IA

4. S3 Intelligent-Tiering - If you haven't used any object in last consecutive 30 days, Amazon S3 automatically moves it to infrequent access tier, if you use the object in the infrequent access tier then amazon S3 moves it to frequent access tier (Amazon S3 Standard)

- Ideal for data with unknown or changing access patterns
- Requires a small monthly monitoring and automation fee per object

5. S3 Glacier instant retrieval

- Works well for archived data that requires immediate access
- Can retrieve objects within a few milliseconds

6. S3 Glacier Flexible Retrieval

- Low-cost storage designed for data archiving
- Able to retrieve objects within a few minutes to hours

7. S3 Glacier Deep Archive

- Lowest-cost object storage class ideal for archiving
- Able to retrieve objects within 12 hours

8. S3 Outposts - Amazon S3 Outposts delivers object storage to your on-premises AWS Outposts environment.

- Creates S3 buckets on Amazon S3 Outposts
- Makes it easier to retrieve, store, and access data on AWS Outposts

File Storage - Compared to block storage and object storage, file storage is ideal for use cases in which a large number of services and resources need to access the same data at the same time.

Amazon Elastic File System (EFS)

Amazon Elastic File System (Amazon EFS) is a scalable file system used with AWS Cloud services and on-premises resources. As you add and remove files, Amazon EFS grows and shrinks automatically. It can scale on demand to petabytes without disrupting applications.

Relational Database - relational database, data is stored in a way that relates it to other pieces of data.

Relational databases use structured query language (SQL) to store and query data. This approach allows data to be stored in an easily understandable, consistent, and scalable way.

Amazon RDS (Relational Database Service) - Amazon Relational Database Service (Amazon RDS)(opens in a new tab) is a service that enables you to run relational databases in the AWS Cloud.

Amazon RDS Database Engines - Amazon RDS is available on six database engines, which optimize for memory, performance, or input/output (I/O). Supported database engines include:

Amazon Aurora

PostgreSQL

MySQL

MariaDB

Oracle Database

Microsoft SQL Server

Amazon Aurora

Amazon Aurora is an enterprise-class relational database. It is compatible with MySQL and PostgreSQL relational databases. It is up to five times faster than standard MySQL databases and up to three times faster than standard PostgreSQL databases.

Amazon Aurora helps to reduce your database costs by reducing unnecessary input/output (I/O) operations, while ensuring that your database resources remain reliable and available.

Non Relational Databases

nonrelational database, you create tables. A table is a place where you can store and query data.

Nonrelational databases are sometimes referred to as **"NoSQL databases"** because they use structures other than rows and columns to organize data. One type of structural approach for nonrelational databases is key-value pairs. With key-value pairs, data is organized into items (keys), and items have attributes (values). You can think of attributes as being different features of your data.

Amazon DynamoDB

Amazon DynamoDB is a key-value database service. It delivers single-digit millisecond performance at any scale.

DynamoDB is serverless, which means that you do not have to provision, patch, or manage servers.

As the size of your database shrinks or grows, DynamoDB automatically scales to adjust for changes in capacity while maintaining consistent performance. - **Auto Scaling**

Amazon Redshift - Amazon Redshift is designed for data analytics and

business intelligence. It allows you to access and analyze data without complex configurations.

Amazon Redshift is a data warehousing service that you can use for big data analytics. It offers the ability to collect data from many sources and helps you to understand relationships and trends across your data.

Amazon DMS (Database Migration Service)

AWS Database Migration Service (AWS DMS)(opens in a new tab) enables you to migrate relational databases, nonrelational databases, and other types of data stores.

With AWS DMS, you move data between a source database and a target database. The source and target databases(opens in a new tab) can be of the same type or different types. During the migration, your source database remains operational, reducing downtime for any applications that rely on the database.

Amazon DocumentDB - Amazon DocumentDB is a document database service that supports MongoDB workloads. (MongoDB is a document database program.)

Amazon Neptune - Amazon Neptune is a graph database service.

You can use Amazon Neptune to build and run applications that work with highly connected datasets, such as recommendation engines, fraud detection, and knowledge graphs.

Amazon Quantum Ledger Database (QLDB) - Amazon Quantum Ledger Database (Amazon QLDB) is a ledger database service.

You can use Amazon QLDB to review a complete history of all the changes that have been made to your application data.

Amazon Managed Blockchaon - Amazon Managed Blockchain is a service that you can use to create and manage blockchain networks

with open-source frameworks.

Amazon ElastiCache - Amazon ElastiCache(opens in a new tab) is a service that adds caching layers on top of your databases to help improve the read times of common requests.

It supports two types of data stores: Redis and Memcached.

Amazon DynamoDB Accelarator (DAX) - Amazon DynamoDB Accelerator (DAX)(opens in a new tab) is an in-memory cache for DynamoDB.

It helps improve response times from single-digit milliseconds to microseconds.

AWS Shared Responsibility Model - AWS is responsible for some parts of your environment and you (the customer) are responsible for other parts. This concept is known as the shared responsibility model. AWS is responsible for security of the cloud and you are responsible for security in the cloud.

The shared responsibility model divides into customer responsibilities (security in the cloud) and AWS responsibilities (Security of the cloud).

Customers: Security in the cloud

Customers are responsible for the security of everything that they create and put in the AWS Cloud.

When using AWS services, you, the customer, maintain complete control over your content. You are responsible for managing security requirements for your content, including which content you choose to store on AWS, which AWS services you use, and who has access to that content. You also control how access rights are granted, managed, and revoked.

AWS: Security of the cloud: AWS is responsible for security of the cloud.

AWS operates, manages, and controls the components at all layers of

infrastructure. This includes areas such as the host operating system, the virtualization layer, and even the physical security of the data centers from which services operate.

AWS is responsible for protecting the global infrastructure that runs all of the services offered in the AWS Cloud.

Amazon Root User: When you first create an AWS account, you begin with an identity known as the root user.

AWS IAM: AWS Identity and Access Management (IAM)

AWS Identity and Access Management (IAM) enables you to manage access to AWS services and resources securely.

IAM gives you the flexibility to configure access based on your company's specific operational and security needs. You do this by using a combination of IAM features, which are explored in detail in this lesson:

IAM users, groups, and roles

IAM policies

Multi-factor authentication

IAM policies enable you to customize users' levels of access to resources. For example, you can allow users to access all of the Amazon S3 buckets within your AWS account, or only a specific bucket.

Follow the Security Principle when granting permissions

Principle of Least Privilege - A user is granted access only to what they need.

AWS Artifact: AWS Artifact is a service that provides on-demand access to AWS security and compliance reports and select online agreements. AWS Artifact consists of two main sections: AWS Artifact Agreements and AWS Artifact Reports.

AWS Artifact Agreements: In AWS Artifact Agreements, you can review, accept, and manage agreements for an individual account and for all your accounts in AWS Organizations. Different types of agreements are offered to address the needs of customers who are subject to specific regulations, such as the Health Insurance Portability and Accountability Act (HIPAA).

AWS Artifact Reports: AWS Artifact Reports provide compliance reports from third-party auditors. These auditors have tested and verified that AWS is compliant with a variety of global, regional, and industry-specific security standards and regulations. AWS Artifact Reports remains up to date with the latest reports released. You can provide the AWS audit artifacts to your auditors or regulators as evidence of AWS security controls.

Customer Compliance Center: The Customer Compliance Center contains resources to help you learn more about AWS compliance.

In the Customer Compliance Center, you can read customer compliance stories to discover how companies in regulated industries have solved various compliance, governance, and audit challenges.

Denial-of-Service(DoS): A denial-of-service (DoS) attack is a deliberate attempt to make a website or application unavailable to users.

DDoS (Distibuted denial-of-service): In a distributed denial-of-service (DDoS) attack, multiple sources are used to start an attack that aims to make a website or application unavailable. This can come from a group of attackers, or even a single attacker. The single attacker can use multiple infected computers (also known as “bots”) to send excessive traffic to a website or application.

To help minimize the effect of DoS and DDoS attacks on your applications, you can use AWS Shield.

AWS Shield: AWS Shield is a service that protects applications against DDoS attacks. AWS Shield provides two levels of protection: Standard and Advanced.

In this AWS Shield there are two categories they are:

- **AWS Shield Standard** - AWS Shield Standard automatically protects all AWS customers at no cost. It protects your AWS resources from the most common, frequently occurring types of DDoS attacks.
- **AWS Shield Advanced** - AWS Shield Advanced is a paid service that provides detailed attack diagnostics and the ability to detect and mitigate sophisticated DDoS attacks.

Encryption: Securing a message or data in a wat that only authorized parties can access it.

AWS Key Management Service (KMS) - AWS Key Management Service (AWS KMS) enables you to perform encryption operations through the use of cryptographic keys. A cryptographic key is a random string of digits used for locking (encrypting) and unlocking (decrypting) data. You can use AWS KMS to create, manage, and use cryptographic keys. You can also control the use of keys across a wide range of services and in your applications.

With AWS KMS, you can choose the specific levels of access control that you need for your keys. For example, you can specify which IAM users and roles are able to manage keys. Alternatively, you can temporarily disable keys so that they are no longer in use by anyone. Your keys never leave AWS KMS, and you are always in control of them.

AWS WAF - AWS WAF is a web application firewall that lets you monitor network requests that come into your web applications.

AWS WAF works together with Amazon CloudFront and an Application Load Balancer. Recall the network access control lists that you learned

about in an earlier module. AWS WAF works in a similar way to block or allow traffic. However, it does this by using a web access control list (ACL)(opens in a new tab) to protect your AWS resources.

Amazon Inspector: Amazon Inspector helps to improve the security and compliance of applications by running automated security assessments. It checks applications for security vulnerabilities and deviations from security best practices, such as open access to Amazon EC2 instances and installations of vulnerable software versions.

After Amazon Inspector has performed an assessment, it provides you with a list of security findings. The list prioritizes by severity level, including a detailed description of each security issue and a recommendation for how to fix it.

Amazon GuardDuty: Amazon GuardDuty(opens in a new tab) is a service that provides intelligent threat detection for your AWS infrastructure and resources. It identifies threats by continuously monitoring the network activity and account behavior within your AWS environment.

After you have enabled GuardDuty for your AWS account, GuardDuty begins monitoring your network and account activity. You do not have to deploy or manage any additional security software. GuardDuty then continuously analyzes data from multiple AWS sources, including VPC Flow Logs and DNS logs.

If **GuardDuty detects any threats**, you can review detailed findings about them from the AWS Management Console. Findings include recommended steps for remediation.

Amazon CloudWatch - 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.

CloudWatch Alarms - With CloudWatch, you can create alarms(opens in a new tab) that automatically perform actions if the value of your metric has gone above or below a predefined threshold.

For example, suppose that your company's developers use Amazon EC2 instances for application development or testing purposes. If the developers occasionally forget to stop the instances, the instances will continue to run and incur charges.

In this scenario we can create a cloudWatch Alarm that automatically stops amazon EC2 instance

AWS CloudTrail - AWS CloudTrail records API calls for your account. The recorded information includes the identity of the API caller, the time of the API call, the source IP address of the API caller, and more. You can think of CloudTrail as a "trail" of breadcrumbs (or a log of actions) that someone has left behind them.

AWS Trusted Advisor - AWS Trusted Advisor(opens in a new tab) is a web service that inspects your AWS environment and provides real-time recommendations in accordance with AWS best practices.

Trusted Advisor compares its findings to AWS best practices in five categories: cost optimization, performance, security, fault tolerance, and service limits. For the checks in each category, Trusted Advisor offers a list of recommended actions and additional resources to learn more about AWS best practices.

When you access the Trusted Advisor dashboard on the AWS Management Console, you can review completed checks for cost optimization, performance, security, fault tolerance, and service limits.

For each category:

•The green check indicates the number of items for which it detected no

problems.

•The orange triangle represents the number of recommended investigations.

•The red circle represents the number of recommended actions.

AWS Support Plans: 4

- **Basic**
- **Developer**
- **Bussiness**
- **Enterprise**

Basic: Basic Support is free for all AWS customers. It includes access to whitepapers, documentation, and support communities. With Basic Support, you can also contact AWS for billing questions and service limit increases.

Developer: Customers in the Developer Support plan have access to features such as:

- Best practice guidance
- Client-side diagnostic tools

Building-block architecture support, which consists of guidance for how to use AWS offerings, features, and services together

Bussiness: Customers with a Business Support plan have access to additional features, including:

- Use-case guidance to identify AWS offerings, features, and services that can best support your specific needs
- All AWS Trusted Advisor checks
- Limited support for third-party software, such as common operating systems and application stack components

Enterprise: customers with Enterprise Support have access to:

- A designated Technical Account Manager to provide proactive

guidance and coordinate access to programs and AWS experts

- A Concierge support team for billing and account assistance
- Operations Reviews and tools to monitor health
- Training and Game Days to drive innovation
- Tools to monitor costs and performance through Trusted Advisor and Health API/Dashboard

AWS MarketPlace: AWS Marketplace(opens in a new tab) is a digital catalog that includes thousands of software listings from independent software vendors. You can use AWS Marketplace to find, test, and buy software that runs on AWS.

For each listing in AWS Marketplace, you can access detailed information on pricing options, available support, and reviews from other AWS customers.

AWS Marketplace offers products in several categories, such as Infrastructure Software, DevOps, Data Products, Professional Services, Business Applications, Machine Learning, Industries, and Internet of Things (IoT).

When Migration what we have to be give the impotance and then AWS introduces the AWS Cloud Adaption Framework (CAF) in that there is a six core perspectives they are : Business, People, Governance perspectives these focuses on Business capabilities whereas Platform, Security and Operations Perspectives are focusess on technical capabilities.

Business Perspective:

Use the Business Perspective to create a strong business case for cloud adoption and prioritize cloud adoption initiatives. Ensure that your business strategies and goals align with your IT strategies and goals.

Common roles in the Business Perspective include:

Business managers

Finance managers

Budget owners

Strategy stakeholders

People Perspective:

Use the People Perspective to evaluate organizational structures and roles, new skill and process requirements, and identify gaps. This helps prioritize training, staffing, and organizational changes.

Common roles in the People Perspective include:

Human resources

Staffing

People managers

Governance Perspective:

Use the Governance Perspective to understand how to update the staff skills and processes necessary to ensure business governance in the cloud. Manage and measure cloud investments to evaluate business outcomes.

Common roles in the Governance Perspective include:

Chief Information Officer (CIO)

Program managers

Enterprise architects

Business analysts

Portfolio managers

Platform Perspective:

The Platform Perspective includes principles and patterns for implementing new solutions on the cloud, and migrating on-premises

workloads to the cloud.

Common roles in the Platform Perspective include:

Chief Technology Officer (CTO)

IT managers

Solutions architects

Security Perspective:

The Security Perspective ensures that the organization meets security objectives for visibility, auditability, control, and agility.

Use the AWS CAF to structure the selection and implementation of security controls that meet the organization's needs.

Common roles in the Security Perspective include:

Chief Information Security Officer (CISO)

IT security managers

IT security analysts

Operations Perspective:

The Operations Perspective helps you to enable, run, use, operate, and recover IT workloads to the level agreed upon with your business stakeholders.

Common roles in the Operations Perspective include:

IT Operations Managers

IT Support Managers

AWS Migration Strategies:

- **Rehosting:** Rehosting also known as "lift-and-shift" involves moving applications without changes. In the scenario of a large legacy migration, in which the company is looking to implement its

migration and scale quickly to meet a business case, the majority of applications are rehosted.

- **Replatforming:** Replatforming, also known as “lift, tinker, and shift,” involves making a few cloud optimizations to realize a tangible benefit. Optimization is achieved without changing the core architecture of the application. for example, you could take your MYSQL database and replatform into RDS MYSQL, without any code changes to all.
- **Refactoring/re-architecting:** Refactoring (also known as re-architecting) involves reimagining how an application is architected and developed by using cloud-native features. Refactoring is driven by a strong business need to add features, scale, or performance that would otherwise be difficult to achieve in the application’s existing environment.
- **Repurchasing:** This strategy involves moving to a different product, commonly seen as a shift to a SaaS platform. Repurchasing involves moving from a traditional license to a software-as-a-service model. For example, a business might choose to implement the repurchasing strategy by migrating from a customer relationship management (CRM) system to Salesforce.com.
- **Retaining:** Retaining consists of keeping applications that are critical for the business in the source environment. This might include applications that require major refactoring before they can be migrated, or, work that can be postponed until a later time.
- **Retiring:** Retiring is the process of removing applications that are no longer needed. Using the AWS migration plan as the opportunity to actually end-of-life these applications can save significant cost and effort for your team. Sometimes you just have to turn off the lights.

AWS Snow Family: The AWS Snow Family is a collection of physical devices that help to physically transport up to exabytes of data into and out of AWS.

1. **AWS SNOWCONE:** AWS Snowcone(opens in a new tab) is a small, rugged, and secure edge computing and data transfer

device. It features 2 CPUs, 4 GB of memory, and up to 14 TB of usable storage.

2. **AWS SNOWBALL:** AWS Snowball has two devices:
   ->**Snowball Edge Storage Optimized** devices are well suited for large-scale data migrations and recurring transfer workflows, in addition to local computing with higher capacity needs.
   Storage: 80 TB of hard disk drive (HDD) capacity for block volumes and Amazon S3 compatible object storage, and 1 TB of SATA solid state drive (SSD) for block volumes.
   Compute: 40 vCPUs, and 80 GiB of memory to support Amazon EC2 sbe1 instances (equivalent to C5).
   ->**Snowball Edge Compute Optimized** provides powerful computing resources for use cases such as machine learning, full motion video analysis, analytics, and local computing stacks.
   Storage: 80-TB usable HDD capacity for Amazon S3 compatible object storage or Amazon EBS compatible block volumes and 28 TB of usable NVMe SSD capacity for Amazon EBS compatible block volumes.
   Compute: 104 vCPUs, 416 GiB of memory, and an optional NVIDIA Tesla V100 GPU. Devices run Amazon EC2 sbe-c and sbe-g instances, which are equivalent to C5, M5a, G3, and P3 instances.
3. **AWS SNOWMOBILE:** AWS Snowmobile(opens in a new tab) is an exabyte-scale data transfer service used to move large amounts of data to AWS. You can transfer up to 100 petabytes of data per Snowmobile, a 45-foot long ruggedized shipping container, pulled by a semi trailer truck.

Amazon SageMaker: Quickly build, train, and deploy machine learning models at scale.

Amazon Augmented AI (Amazon A2I): Amazon A2I, provide a machine learning platform that any business can build upon without needing PhD level expertise in-house.

Amazon Lex: Amazon Lex is a service provided by AWS for building conversational interfaces into any application using voice and text.

Serverless Applications: AWS, serverless refers to applications that don't require you to provision, maintain, or administer servers. You don't need to worry about fault tolerance or availability. AWS handles these capabilities for you.

AWS Lambda is an example of a service that you can use to run serverless applications. If you design your architecture to trigger Lambda functions to run your code, you can bypass the need to manage a fleet of servers.

Artificial Intelligence: AWS offers a variety of services powered by artificial intelligence (AI).

For example, you can perform the following tasks:

Convert speech to text with Amazon Transcribe.

Discover patterns in text with Amazon Comprehend.

Identify potentially fraudulent online activities with Amazon Fraud Detector.

Build voice and text chatbots with Amazon Lex.

Machine Learning: Traditional machine learning (ML) development is complex, expensive, time consuming, and error prone. AWS offers Amazon SageMaker to remove the difficult work from the process and empower you to build, train, and deploy ML models quickly.

You can use ML to analyze data, solve complex problems, and predict outcomes before they happen.

AWS Well-Architected Framework:

The AWS Well-Architected Framework(opens in a new tab) helps you understand how to design and operate reliable, secure, efficient, and cost-effective systems in the AWS Cloud. It provides a way for you to consistently measure your architecture against best practices and design principles and identify areas for improvement.

The well-architected framework is based on six pillars:

1. **Operational excellence**: Operational excellence is the ability to run and monitor systems to deliver business value and to continually improve supporting processes and procedures. Design principles for operational excellence in the cloud include performing operations as code, annotating documentation, anticipating failure, and frequently making small, reversible changes.
2. **Security:** The Security pillar is the ability to protect information, systems, and assets while delivering business value through risk assessments and mitigation strategies.
3. **Reliability**: Reliability is the ability of a system to do the following: Reliability includes testing recovery procedures, scaling horizontally to increase aggregate system availability, and automatically recovering from failure.
4. **Performance efficiency**: Performance efficiency is the ability to use computing resources efficiently to meet system requirements and to maintain that efficiency as demand changes and technologies evolve. Evaluating the performance efficiency of your architecture includes experimenting more often, using serverless architectures, and designing systems to be able to go global in minutes.
5. **Cost optimization:**Cost optimization is the ability to run systems to deliver business value at the lowest price point. Cost optimization includes adopting a consumption model, analyzing and attributing expenditure, and using managed services to reduce the cost of ownership.
6. **Sustainability:**Sustainability is the ability to continually improve sustainability impacts by reducing energy consumption and increasing efficiency across all components of a workload by maximizing the benefits from the provisioned resources and minimizing the total resources required.

Six Benefits of AWS:

1. Trade upfront expense for variable expense.

2. Benefit from massive economies of scale.
3. Stop guessing capacity.
4. Increase speed and agility.
5. Stop spending money running and maintaining data centers.
6. Go global in minutes.

QA

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.

Amazon Lex: Building interfaces - Build Chatbots

Web service's performance globally - Amazon CloudFront

AWS Compute optimizer - get recomendations of what you use

AWS WaveLength - Deliver ultra low latency applications for 5G devices.

AWS Well Architected Framework

AWS Cloud Farmation Framework

AWS Shared Responsibility - Hardware and Infrastucture

Largest Discount - Spot Instances - Can't tolerate Interuptions

SQL Injection - AWS WAF - Web Application Firewall

AWS OpSworks - Automate Operations with chef and puppet

AWS TimeStream - Fast, scalable and serverless time-series database

Amazon Comprehend - Derive and understand valuable insights from text within the documents.

Amazon Transcribe - Automatically convert speech to text

Limit Human Error - Automatic

AWS Code Deploy - Code Deploy is a deployment service that

automates application deployments to amazon EC2 instances, on-premises instances, serverless lambda functions or amazon ecs services

AWS Responsible - Physical server - Hardware and Infrastructure

SCALABILITY - ability of a software system to process higher amount of workload on its current hardware resources (scale up) or on current and additional hardware resources (scale out) without application service interruption

ELASTICITY - ability of the hardware layer below (usually cloud infrastructure) to increase or shrink the amount of the physical resources offered by that hardware layer to the software layer above. The increase / decrease is triggered by business rules defined in advance (usually related to application's demands). The increase / decrease happens on the fly without physical service interruption.

Six Pillars of Well-Architected Framework

1. Operational Excellence - ability to support development and run workloads effectively and continuosly improve supporting processes and procedures to deliver business value

2. Security - ability to protect data, systems, and assets to take advantage of cloud technologies to improve your security

3. Reliability - ability of a workload to perform its intended function correctly and consistently when it’s expected to. This includes the ability to operate and test the workload through its total lifecycle.

4. Performance Efficiency - ability to use computing resources efficiently to meet system requirements, and to maintain that efficiency as demand changes and technologies evolve.

5. Cost Optimization - ability to run systems to deliver business value at the lowest price point

6. Sustainability - The discipline of sustainability addresses the long-term environmental, economic, and societal impact of your business activities.

AWS Elastic Load Balancer (ELB) - Distributes traffic evenly across multiple healthy instances

AWS Control Tower - Set-up and govern a secure, multi-account AWS environment

AWS AppRunner - Deploy containerized web applications and APIs at scale

AWS Amplify - Build full-stack web and mobile apps in hours

AWS CodeStar - Quickly develop, build and deploy applications on AWS, easily manage your software development activities in one place

AWS CloudHSM - AWS CloudHSM is a cryptographic service that combines the benefits of the AWS cloud with the security of hardware security modules. HSMs are specialized security devices that process cryptographic operations and provide secure storage for cryptographic keys.

AWS Cloud9 - It is a cloud IDE for writing, running and debugging code

AWS AppSync - GraphQL and Pub/Sub APIs (whenever we see these two directly goes to AWS AppSync)

AWS CodePipeline - AWS CodePipeline is a continuous delivery service that enables you to model, visualize, and automate the steps required to release your software

AWS DirectConnect - Whenever we see "Dedicated network Connection" then directly go for AWS Direct Connect

AWS EC2 - Compute as a service

AWS EKS - Kubernetes as a service

Route53 - Domain Name System as a service

DNS keeps a lot of records and there are lot of servers that basically maps this domain name to the IP address.

A **load balancer** distributes workloads across multiple compute resources, such as virtual servers. Using a load balancer increases the availability and fault tolerance of your applications.

A **NAT gateway** is a Network Address Translation (NAT) service. You can use a NAT gateway so that instances in a private subnet can connect to services outside your VPC but external services cannot initiate a connection with those instances.