AWS Notes

Chapter -1

Contents

IT Resources!!

IT (Information Technology) resources refer to the components, tools, and assets that organizations use to manage, store, process, transmit, and secure information. These resources are crucial for supporting the technology infrastructure of an organization and facilitating its day-to-day operations. IT resources can be broadly categorized into the following types:

1. **Hardware Resources**:
   - Servers: Physical or virtual machines that provide computing power for running applications and storing data.
   - Storage Devices: Devices such as hard drives, SSDs, and network-attached storage (NAS) for storing data.
   - Networking Equipment: Routers, switches, firewalls, and other devices that enable communication and data transfer within a network.

2. **Software Resources:**
   - Operating Systems: Software that manages hardware resources and provides a platform for running applications.
   - Application Software: Programs designed to perform specific tasks or functions, such as word processors, databases, and enterprise applications.
   - Middleware: Software that facilitates communication and data exchange between different applications and systems.

3. **Networking Resources:**
   - Network Infrastructure: The physical and logical components that make up a network, including routers, switches, cables, and protocols.
   - Internet Connectivity: Access to the internet through ISPs (Internet Service Providers) or other means.
   - Wireless Networks: Technologies like Wi-Fi that enable wireless communication and connectivity.

Problems with Traditional IT Approach

traditional IT approaches had several limitations and challenges. Here are some of the key problems associated with the **traditional IT model**

1. Infrastructure Costs: Traditional IT required organizations to invest heavily in physical hardware, including servers, storage devices, and networking equipment. These upfront costs could be significant and were often challenging to predict accurately**.**

2. Scalability Issues: Scaling up or down in traditional IT environments was often a time-consuming and expensive process. Organizations had to purchase additional hardware to accommodate increased workloads, and scaling down meant dealing with underutilized resources.

3. Resource Underutilization: Physical servers often operated at low levels of utilization, leading to inefficiencies. Organizations had to provision resources for peak demand, resulting in underutilized capacity during periods of lower demand.

4. Complexity of Management: Managing a complex IT infrastructure, including hardware procurement, installation, maintenance, and troubleshooting, required specialized skills and significant manpower. This complexity could lead to longer deployment times and increased risk of errors.

5. Long Deployment Cycles: Setting up and configuring physical servers and infrastructure components took time. This led to longer deployment cycles for new applications or updates, slowing down the pace of innovation and responsiveness to business needs.

6. Data Center Dependence: Traditional IT often relied on on-premises data centers, making organizations susceptible to issues like power outages, natural disasters, and other disruptions. Ensuring high availability and disaster recovery required additional investments and planning.

7. Limited Accessibility: Access to IT resources was often restricted to specific physical locations, making it challenging for remote or distributed teams to collaborate effectively. This limitation hindered the adoption of flexible work arrangements.

8. Security Concerns: Security was a major concern in traditional IT environments. Organizations had to implement and manage their own security measures, including firewalls, intrusion detection systems, and regular software updates. The responsibility for security fell largely on the organization itself.

What is Cloud Computing?

Cloud computing is a technology that allows users to access and use computing resources (such as servers, storage, databases, networking, software, analytics, and intelligence) over the internet, often referred to as "the cloud." Instead of owning and maintaining physical hardware or servers, users can rent or lease these resources from a cloud service provider.

The key characteristics of cloud computing include:

1.**On-Demand Self-Service**: Users can provision and manage computing resources as needed without requiring human intervention from the service provider.

2.**Broad Network Access**: Cloud services are accessible over the internet from various devices such as laptops, smartphones, and tablets.

3.**Resource Pooling**: Resources are pooled to serve multiple users, and different physical and virtual resources are dynamically assigned and reassigned according to demand.

4.**Rapid Elasticity**: Resources can be rapidly and elastically provisioned or released to scale up or down based on demand. This provides flexibility and cost efficiency.

5.**Measured Service**: Cloud computing resources are metered, and users are billed based on their usage. This pay-as-you-go model allows for cost optimization.

Types of Cloud Computing:

- Cloud Deployment
- Cloud Services

Cloud Deployment

1.**Public Cloud**: Cloud resources are owned and operated by a third-party cloud service provider and are made available to the general public.

2.**Private Cloud**: Cloud resources are used exclusively by a single organization. It can be managed internally or by a third-party provider.

3.**Hybrid Cloud:** Combines elements of both public and private clouds, allowing data and applications to be shared between them.

Cloud Services

1.Infrastructure as a Service (IaaS): Provides virtualized computing resources over the internet. Users can rent virtual machines and storage.

2.Platform as a Service (PaaS): Offers a platform allowing customers to develop, run, and manage applications without dealing with the complexity of building and maintaining the underlying infrastructure.

3.Software as a Service (SaaS): Delivers software applications over the internet on a subscription basis. Users can access the software without worrying about installation or maintenance.

Advantages of Cloud Computing

Cloud computing offers numerous advantages for individuals, businesses, and organizations. Some of the key advantages include:

1. Cost-Efficiency:

- Pay-as-you-go pricing models allow users to pay only for the resources they use, reducing upfront capital expenses.
- Eliminates the need for purchasing and maintaining expensive hardware and infrastructure.
- Scalability enables businesses to easily adjust resources up or down based on demand.

2. Scalability:

- Cloud services can scale resources automatically to handle increased workloads or traffic spikes.
- Easily add or remove virtual machines, storage, and other resources as needed.

3. Accessibility:

- Cloud services are accessible from anywhere with an internet connection, promoting remote work and collaboration.
- Supports a wide range of devices, including smartphones, tablets, and laptops.

4. Reliability and Redundancy:

- Leading cloud providers offer high levels of uptime and reliability through redundant data centers and failover mechanisms.
- Data is often mirrored across multiple geographic locations, reducing the risk of data loss.

5. Flexibility:

- Cloud computing provides a wide range of services and deployment options, including Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).
- Users can choose the services that best fit their specific needs and customize them as required.

6. Security:

- Major cloud providers invest heavily in security measures, including data encryption, identity and access management, and threat detection.
- Cloud providers often have dedicated security teams and compliance certifications.

7. Automatic Updates and Maintenance:

- Cloud providers handle software updates, security patches, and infrastructure maintenance, reducing the burden on IT staff.
- This ensures that systems are kept up to date and secure.

8. Disaster Recovery:

- Cloud providers typically offer robust backup and disaster recovery solutions, allowing businesses to quickly recover data in case of a disaster or data loss.

Overall, cloud computing offers significant advantages in terms of cost savings, scalability, accessibility, security, and flexibility, making it an essential technology for many businesses and organizations.

AWS Notes

Chapter -2

Contents

Introduction to AWS

- Amazon launched AWS, a cloud computing platform to allow the different organizations to take advantage of reliable IT infrastructure.
- Based on the concept of **Pay-As-You-Go**, AWS provides the services to the customers.
- AWS offers a wide range of services and tools that can be easily combined to build and deploy a variety of applications, making it highly flexible.

History of AWS

2002: Amazon launched internally
2004: Simple Queue Service(SQS) launched
2006: Amazon officially Launched AWS (S3, EC2)

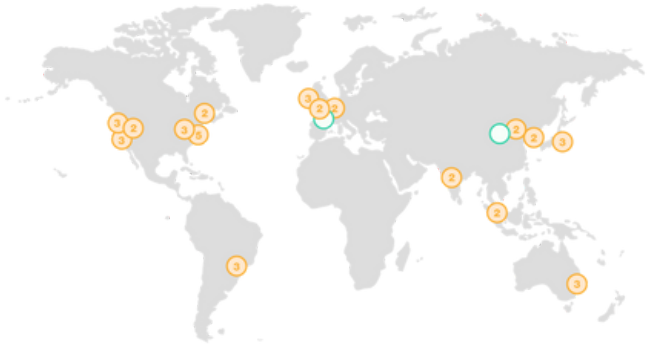
Features of AWS

- Flexibility (in customizing resources like CPU, RAM, Storage type..etc)
- Cost-effective (because of pay as you go model)
- Scalable and elastic (increase and decrease resources)
- Secure
- Experienced

AWS Global Infrastructure

The following are the components that make up the AWS infrastructure:

- Regions
- Availability Zones
- Edge locations
- Regional Edge Caches



Region

A Region is a cluster of Data Centers

How to choose Region

- Data governance and legal requirements
- Low latency
- New services and New features aren’t available in every Region
- pricing varies region to region

Availability Zone

- Availability Zone is a Single Data Center or a group of Data Centers in a region
- Each region has many availability zones (usually 3, min is 3, max is 6)
- Data Centers are separate from each other, so that they’re isolated from disasters

Edge Location

- Egde Location is the Data Center used to deliver content fast to users.
- Content is delivered to end users with lower latency
- Amazon has 550+ Points of Presence (550+ Edge Locations & 13 Regional Caches)

AWS Services

AWS provides wide range of cloud services that enable businesses and individuals to build, deploy, and manage applications and infrastructure in the cloud.

Here are some of the main AWS services and categories:

- **Compute Services**
- **Storage Services**
- **Database Services**
- **Networking Services**
- **Security, Identity, and Compliance**
- **Analytics and Big Data**
- **Machine Learning and AI**
- **Containers and Orchestration**
- **Serverless and Application Integration**
- **Developer Tools**

Compute Services

- **Amazon EC2 (Elastic Compute Cloud)**: Provides scalable virtual servers (instances) that can run a wide range of operating systems and applications.
- **AWS Lambda**: A serverless compute service that allows you to run code in response to events without provisioning or managing servers.

Storage Services

- **Amazon S3 (Simple Storage Service):** Object storage service for storing and retrieving data, including files, documents, images, and backups.
- **Amazon EBS (Elastic Block Store):** Provides block-level storage volumes for use with EC2 instances.
- **Amazon Glacier:** A low-cost, long-term storage service designed for archiving and data backup.

.

Database Services

- **Amazon RDS (Relational Database Service):** Managed relational database service supporting various database engines like MySQL, PostgreSQL, Oracle, and SQL Server.
- **Amazon DynamoDB:** A fully managed NoSQL database service that offers fast and flexible data storage.
- **Amazon Redshift:** A fully managed data warehousing service designed for high-performance analytics.

Networking Services

- **Amazon VPC (Virtual Private Cloud):** Provides a private network within the AWS cloud, allowing you to isolate and secure resources.
- **Amazon Route 53:** A scalable domain name system (DNS) web service for routing domain traffic to AWS resources.
- **AWS Direct Connect:** Allows for dedicated network connections between on-premises data centers and AWS.

Security, Identity and Compliance

- **AWS Identity and Access Management (IAM):** Allows you to control access to AWS resources and services.
- **Amazon Cognito:** Provides user identity and access management for web and mobile applications.
- **AWS Key Management Service (KMS):** Manages encryption keys and provides secure key storage.

Analytics & Big Data

- **Amazon EMR (Elastic MapReduce):** A big data platform for processing and analyzing vast amounts of data.
- **Amazon Athena:** An interactive query service that makes it easy to analyze data stored in Amazon S3.
- **Amazon Kinesis:** A platform for real-time streaming data ingestion and processing.

Machine Learning & AI

- **Amazon SageMaker:** A fully managed service for building, training, and deploying machine learning models.
- **AWS Rekognition:** A service for image and video analysis using deep learning models.
- **Amazon Comprehend:** Natural language processing service for text analysis.

Containers & Orchestration

- **Amazon ECS (Elastic Container Service):** A fully managed container orchestration service.
- **Amazon EKS (Elastic Kubernetes Service):** Managed Kubernetes service for container orchestration.

Serverless & Application Integration

- **AWS Step Functions:** Coordinates serverless workflows and applications.
- **Amazon SQS (Simple Queue Service):** Fully managed message queuing service.
- **Amazon SNS (Simple Notification Service):** Pub/Sub messaging service for building distributed systems.

Developer Tools

- **AWS CodeCommit:** A source code repository service.
- **AWS CodeBuild:** A fully managed build service.
- **AWS CodeDeploy:** Automates code deployments to various compute services.

AWS Notes

Chapter -3

(EC2, S3, VPC, Lambda, Cloudwatch, Route 53)

Amazon EC2

EC2 is one of the most popular of AWS' offering
Infrastructure as a Service
It mainly consists in the capability of :

- Renting virtual machines (EC2)
- Storing data on virtual drives (EBS)
- Distributing load across machines (ELB)
- Scaling the services using an auto-scaling group (ASG)

EC2 sizing & configuration options

- Operating System (OS): Linux, Windows or Mac OS
- How much compute power & cores (CPU)
- How much random-access memory (RAM)
- How much storage space:
  Network-attached (EBS & EFS)
  hardware (EC2 Instance Store)
- Network card: speed of the card, Public IP address
- Firewall rules: security group
- Bootstrap script (configure at first launch): EC2 User Data

Types of Instances

Amazon EC2 (Elastic Compute Cloud) offers a wide range of instance types
Here are some of the common EC2 instance families

- General Purpose (T2, M5, M6g, etc.)
- Compute Optimized (C4, C5, C6g, etc.)
- Memory Optimized (R4, R5, R6g, etc.)
- Storage Optimized (I3, I4, D2, etc.)
- Accelerated Computing (P3, P4, G4, etc.)
- High Performance Computing (HPC, HPC6g)

General Purpose (T2, M5, M6g, etc.)
Great for a diversity of workloads such as web servers or code repositories
Balance between:
• Compute
• Memory
• Networking

Compute Optimized (C4, C5, C6g, etc.)

Great for compute-intensive tasks that require high performance processors
• Media transcoding
• High performance web servers
• High performance computing (HPC)
• Dedicated gaming servers

Memory Optimized (R4, R5, R6g, etc.)

Advantages: Memory-optimized instances are ideal for applications that require a large amount of RAM, such as in-memory databases, data caching, and analytics. They offer a high memory-to-CPU ratio.

Storage Optimized (I3, I4, D2, etc.)

Advantages: Storage-optimized instances are tailored for applications that require high disk I/O performance and large storage capacities, such as NoSQL databases, data warehousing, and big data processing.

Accelerated Computing (P3, P4, G4, etc.)

Advantages: These instances are equipped with specialized GPUs or FPGAs, making them well-suited for machine learning, deep learning, high-performance computing (HPC), and graphics-intensive applications.

High Performance Computing (HPC, HPC6g)

Advantages: These instances are designed for high-performance computing workloads, such as simulations, modeling, and scientific research. They offer low-latency networking and high CPU/GPU capabilities

Advantages of AWS EC2-Instances

- EC2 instances can be easily scaled up or down as per the requirement.
- EC2 instances are charged based on usage
- It can be easily deployed and managed using Amazon Web Services (AWS) management console, APIs, or CLI.
- It can be deployed in multiple availability zones to ensure high availability and data durability.
- It can be customized with different operating systems, applications, and network configurations.

INTRO TO

AMAZON S3

- Amazon S3 (Simple Storage Service) is a highly scalable and durable object storage service provided by Amazon Web Services (AWS).

- It is designed to store and retrieve large amounts of data and is widely used for various use cases across industries.

AMAZON S3 USE CASES

- Backup and storage
- Archive
- Application hosting
- Media hosting
- Data lakes & big data analytics
- Static website

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
- Creating Bucket follows Naming convention

AMAZON S3 - OBJECTS

- Object values are the content of the body:
  Max. Object Size is 5TB (5000GB)
- Objects (files) have a Key
- Metadata (list of text key / value pairs – system or user metadata)
- Tags (Unicode key / value pair – up to 10) – useful for security / lifecycle
- Version ID (if versioning is enabled)

AWS VPC (Virtual Private Cloud)

Amazon Virtual Private Cloud (Amazon VPC) is a service provided by Amazon Web Services (AWS)

AWS VPC provides a way to create a private, isolated network environment within the AWS Cloud, giving you control over your virtual networking environment and allowing you to build scalable, highly available applications.

It forms the foundation for deploying and connecting various AWS resources in a secure and controlled manner.

WHY VPC?

- Isolation and Customization
- Networking Components
- Security
- Connectivity
- Elastic Load Balancing (ELB)
- Hybrid Cloud Connectivity
- Scalability and High Availability
- Integration with Other AWS Services

Isolation and Customization

Isolation: With VPC, you can create isolated sections of the AWS Cloud, known as VPCs, where you can launch resources in a virtual network.

Customization: You have complete control over your VPC, including the selection of your own IP address range, creation of subnets, and configuration of route tables and network gateways.

Networking Components

Subnets: You can divide your VPC into multiple subnets, each associated with a specific availability zone. Subnets allow you to group resources and apply different network policies.

Route Tables: Each subnet in a VPC must be associated with a route table, which controls the traffic routing between subnets.

Internet Gateway: Provides a connection between your VPC and the internet, allowing resources in your VPC to communicate with the internet.

NAT Gateway/NAT Instance: Allows private subnets to initiate outbound traffic to the internet while preventing inbound traffic from reaching those resources directly.

Security

Security Groups: Act as a virtual firewall for your instances to control inbound and outbound traffic at the instance level.

Network Access Control Lists (NACLs): Operate at the subnet level and act as a firewall for controlling traffic in and out of one or more subnets.

Connectivity

VPC Peering: Allows you to connect one VPC with another via a direct network route.

VPN Connections: You can establish secure connections between your on-premises data centers and your VPC using Virtual Private Network (VPN) connections.

Elastic Load Balancing (ELB)

Application Load Balancer (ALB) and Network Load Balancer (NLB): Can be deployed within a VPC to distribute incoming traffic across multiple instances.

Hybrid Cloud Connectivity

AWS Direct Connect: Establishes dedicated network connections from your on-premises data centers to AWS.

AWS VPN: Allows you to connect your on-premises network to your VPC over the internet.

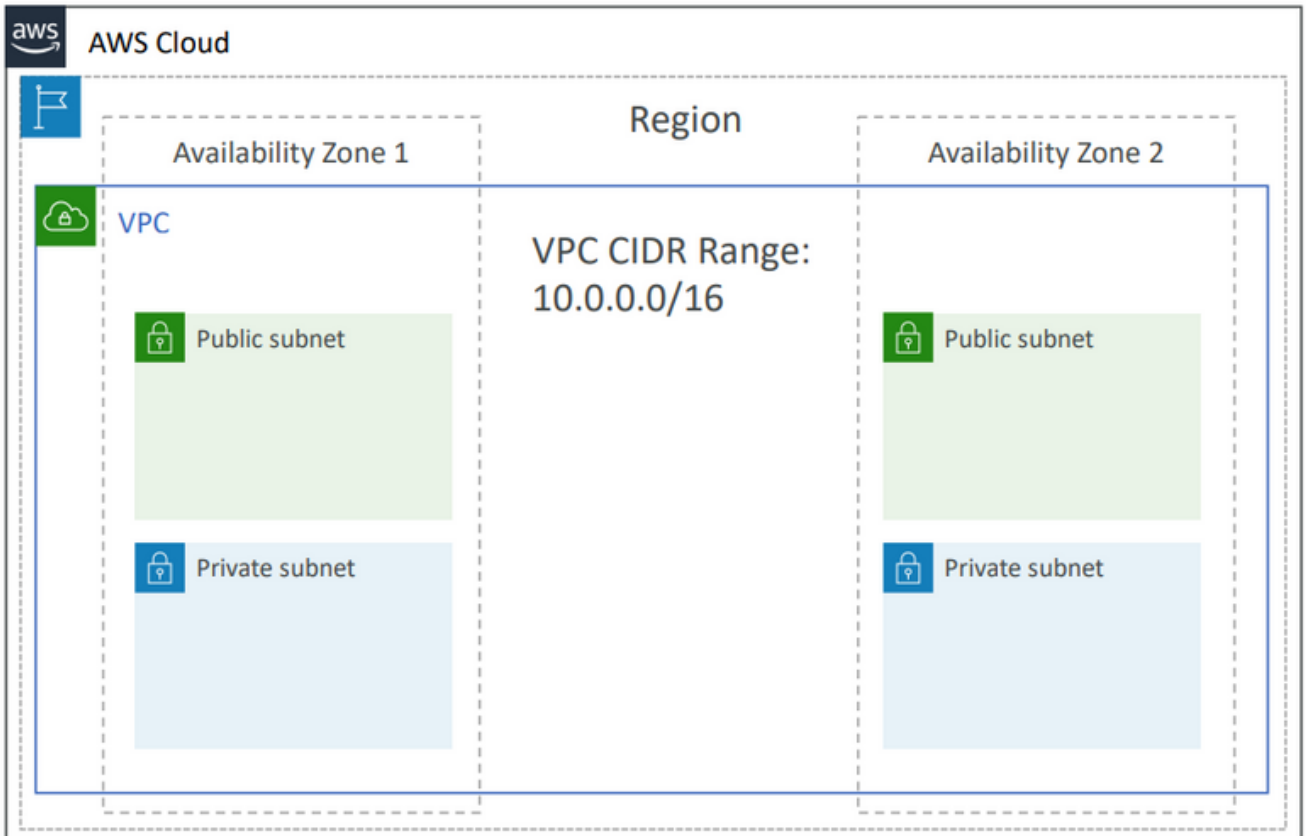
Scalability and High Availability

Auto Scaling: Allows you to automatically scale your Amazon EC2 instances based on conditions you define.

Multi-AZ Deployments: Resources can be deployed across multiple availability zones for improved fault tolerance.

Integration with Other AWS Services

VPC integrates with various AWS services, such as Amazon S3, Amazon RDS, AWS Lambda, etc., allowing you to build complex and scalable architectures.



Block Diagram of VPC

IP Address

An IP address, or Internet Protocol address, is a numerical label assigned to each device participating in a computer network that uses the Internet Protocol for communication.

It serves two main purposes:
host or network interface identification and location addressing.

IPv4 vs. IPv6

IPv4 (Internet Protocol version 4):
This is the most widely used IP version. It consists of four sets of numbers separated by dots,
for example, 192.168.0.1.

IPv6 (Internet Protocol version 6):
This newer version was introduced to address the exhaustion of IPv4 addresses. IPv6 uses a longer address format, composed of eight groups of four hexadecimal digits, separated by colons, like 2001:0db8:85a3:0000:0000:8a2e:0370:7334.

Public vs. Private IP Addresses

Public IP Address:
This is the address assigned to a device by the Internet Service Provider (ISP) and is visible on the Internet.

Private IP Address:
These are used within a private network and are not directly accessible from the Internet. Common private IP address ranges include 192.168.x.x, 172.16.x.x to 172.31.x.x, and 10.x.x.x.

Static vs. Dynamic IP Addresses

Static IP Address:
A fixed IP address that doesn't change. It's manually configured and is often used for servers and network devices that need a consistent address.

Dynamic IP Address:
An IP address that is automatically assigned by a DHCP (Dynamic Host Configuration Protocol) server. This is more common for devices in home networks.

DHCP (Dynamic Host Configuration Protocol)

DHCP is a network protocol used to automatically assign IP addresses and other network configuration information to devices on a network.

Subnetting

Subnetting is the practice of dividing a network into sub-networks to improve performance and security.

IP Address Classes (0nly IPv4)

IP addresses are classified into five classes: A, B, C, D, and E. These classes are based on the range of IP addresses they include and the number of hosts they can support. Here's a brief overview of each class:

Class A (1.0.0.0 to 126.255.255.255)

- The first octet represents the network address, and the remaining three octets represent host addresses.
- Supports a large number of hosts (2^24 - 2 hosts).
- Typically used for large networks.

Class B (128.0.0.0 to 191.255.255.255)

- The first two octets represent the network address, and the remaining two octets represent host addresses.
- Supports a moderate number of hosts (2^16 - 2 hosts).
- Suitable for medium-sized networks.

Class C (192.0.0.0 to 223.255.255.255)

- The first three octets represent the network address, and the last octet represents host addresses.
- Supports a smaller number of hosts (2^8 - 2 hosts).
- Commonly used for small networks.

Class D (224.0.0.0 to 239.255.255.255)

Reserved for multicast groups.
Intended for group communication.

Class E (240.0.0.0 to 255.255.255.255)

Reserved for experimental purposes.

IP Addresses in AWS

IPv4 – Internet Protocol version 4 (4.3 Billion Addresses)
• Public IPv4 – can be used on the Internet
• EC2 instance gets a new a public IP address every time you stop then start it (default)
• Private IPv4 – can be used on private networks (LAN) such as internal AWS
networking (e.g., 192.168.1.1)
• Private IPv4 is fixed for EC2 Instances even if you start/stop them

IPv6 – Internet Protocol version 6

• Every IP address is public (no private range)
• Example: 2001:db8:3333:4444:cccc:dddd:eeee:ffff

Overview

• VPC – Virtual Private Cloud
• Subnets – Tied to an AZ, network partition of the VPC
• Internet Gateway – at the VPC level, provide Internet Access
• NAT Gateway / Instances – give internet access to private subnets
• NACL – Stateless, subnet rules for inbound and outbound
• Security Groups – Stateful, operate at the EC2 instance level or ENI
• VPC Peering – Connect two VPC with non overlapping IP ranges, nontransitive
• Elastic IP –fixed public IPv4, ongoing cost if not in-use
•VPC Endpoints – Provide private access to AWS Services within VPC
• PrivateLink – Privately connect to a service in a 3rd party VPC
• VPC Flow Logs – network traffic logs
• Site to Site VPN – VPN over public internet between on-premises DC and AWS
• Client VPN – OpenVPN connection from your computer into your VPC
• Direct Connect – direct private connection to AWS
• Transit Gateway – Connect thousands of VPC and on-premises networks together

What is AWS Lambda?

AWS Lambda is a serverless computing service that allows you to run programs without having to worry about provisioning, maintaining, or waiting for servers to be built

AWS Lambda Features

Serverless: Run code without provisioning or maintaining a server.
Automatic Scaling: Scale your applications automatically as per the workload.
Pay per use: Billed per millisecond of use.
Consistency: Performance consistency is achieved by selecting the right memory size for the function.
Language support: AWS Lambda supports multiple programming languages.

Event Source

AWS Lambda Event Source is any AWS Service, custom application, a stream of data or queue which triggers a Lambda function to run.

Lambda Event Sources Type

Push-based model

- Synchronous
- Asynchronous

Pull-based model

- Stream
- Queue

Push-based model:

Other service directly triggers Lambda and tells it to activate when something happens.

- Synchronous: Lamba returns a response back to the event source.
- Asynchronous: For asynchronous invocations, Lambda does handle retries. After Lambda is invoked by the source, it will first place the event into a queue and immediately sends a success response back to the source.

Pull-based model:

Information is flowing through a stream or put in a queue that Lambda periodically polls, and it then goes into action when certain events are detected .

- Stream: Lambda will stop polling while it retries the message.
- Queue: Lambda will return the message to the queue if the invocation fails or times out but will also keep retrying until it’s either successful or expires.

Access Permissions

Security is crucial when dealing with Lambda because of its power to run code and take actions that affect other AWS services.
Two types of security permissions are needed for Lambda:
Invocation permissions and Execution roles.

- Invocation permissions are only needed for push event sources and are granted through an IAM resource policy automatically created when configuring an AWS service as an event source.
- Execution roles grant Lambda permissions to interact with other AWS services. They include an IAM policy defining what Lambda is allowed to do and a Trust policy allowing Lambda to assume the execution role and perform actions on the other service.

Adding the execution role to your Lambda function is the
final step in granting permissions and policies.

What is
Cloud watch?

- Cloud monitoring is a process of continuously observing and collecting data about the performance, health, and security of your AWS resources and applications.

- AWS provides a service called Amazon CloudWatch to facilitate cloud monitoring.

- CloudWatch enables you to gain insights into your applications, infrastructure, and overall AWS environment by collecting and tracking metrics, logs, and events.

- CloudWatch provides metrics for every services in AWS
- Metric is a variable to monitor (CPUUtilization, NetworkIn…)
- Metrics have timestamps

WHY Cloud Watch?

- Performance Optimization
- Proactive Issue Detection
- Resource Scaling
- Troubleshooting and Debugging
- Cost Management
- Security Assurance
- Compliance Requirements

Metrics and Alarms

Metrics: CloudWatch collects and stores data in the form of metrics. Metrics represent various aspects of your AWS resources, such as CPU utilization, network traffic, and more.

Alarms: You can set up alarms based on thresholds for these metrics. Alarms notify you when certain conditions are met, allowing you to take proactive actions before issues escalate.

Logs

CloudWatch Logs enables you to collect, store, and analyze log data from your applications and AWS resources. This is crucial for troubleshooting, debugging, and understanding the behavior of your applications.

Events

CloudWatch Events allows you to respond to changes in your AWS resources. You can set up rules that trigger automated actions in response to events, helping you to automate workflows and respond to changes in your environment.

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: Total Estimated Charge (only in us-east-1)

Route 53

Amazon Route 53 is a scalable and highly available Domain Name System (DNS) web service offered by Amazon Web Services (AWS).

It provides domain registration, DNS routing, and health checking of resources within your infrastructure.

DNS is a system that translates human-readable domain names (like www.example.com) into IP addresses that computers use to identify each other on the network.

Route 53 Overview

Route 53 offers a variety of domain registration, management, and DNS services, making it easier for businesses and developers to route end-user traffic to internet applications.

Domain Registration

- Route 53 allows you to register new domain names, or you can transfer existing domain names from other registrars to Route 53.
- It supports various top-level domains (TLDs) such as .com, .net, .org, and country-code TLDs.

DNS Management

- Route 53 provides a scalable and highly available DNS infrastructure to route end-user requests to globally distributed AWS resources.
- It supports the creation and management of various DNS record types, including A (IPv4 address), AAAA (IPv6 address), CNAME (canonical name), MX (mail exchange), TXT (text), and more.

Traffic Routing and Health Checking

- Route 53 allows you to configure health checks for your resources, and it can automatically route traffic away from unhealthy resources to healthy ones.
- Traffic can be routed based on various policies, such as simple round-robin, weighted round-robin, latency-based routing, geolocation-based routing, and failover.