Submitted by: Minahil Umar
Phase N0 7

Cloud computing :Cloud computing is the delivery of computing services over the internet (“the cloud”) instead of using your own local computer or servers.
Types : ("In IaaS, you rent infrastructure like servers and databases to store and run your app.
In PaaS, you rent a platform to deploy your app without managing the infrastructure.
In SaaS, you use a fully functional software available on the internet according to your need.")

What is AWS?

> **AWS (Amazon Web Services) is a cloud computing platform by Amazon that lets you use services like servers (EC2), storage (S3), and databases (RDS) over the internet, without buying physical hardware.**

What is AWS used for?

- **Hosting websites & apps**
- **Storing data/files**
- **Running databases**
- **Machine Learning**
- **Backup and Disaster Recovery**
- **CI/CD Pipelines**
- **Big Data analytics**
- **IoT applications**

AWS Global Infrastructure

Term	Meaning
Region	**Regions are geographical areas where AWS offers cloud services. Each region consists of multiple isolated Availability Zones (AZs). This**

	means AWS is available in that area and you can choose to deploy your services there.
Availability Zone (AZ)	**Availability Zones (AZs) are separate physical data centers within a region. They are designed for fault isolation but are connected through low-latency, high-throughput, and redundant networking to support high availability.**
Edge Location	**Edge Locations are separate data centers used primarily for content delivery (via Amazon CloudFront). These are closer to end-users than AZs and Regions, and are used to cache and deliver content (like images, videos, APIs) and services with low latency** **.**

What is Amazon EMR?

Amazon EMR is a cloud-based big data platform by AWS that allows you to process, analyze, and transform massive amounts of data using open-source big data frameworks like:

- **Apache Spark**
- **Apache Hadoop**
- **Apache Hive**
- **Presto**
- **HBase**
- **Flink**

On-Premises vs ☁️ Cloud Computing

Feature	On-Premises	Cloud Computing (like AWS, Azure, GCP)
Location	Servers and hardware are on your own site	Hosted by a cloud provider
Ownership	You buy and maintain your own infrastructure	You rent resources (pay-as-you-go)
Upfront Cost	High — buy servers, storage, cooling, power, etc.	Low — no upfront cost; only pay for what you use
Maintenance	You are responsible for hardware, updates, security	Cloud provider handles all maintenance
Scalability	Hard — need to buy more hardware	Easy — scale up/down in minutes
Deployment Time	Slow — weeks/months to set up	Fast — minutes/hours
Availability	Depends on your setup	Highly available (using multiple regions and zones)
Disaster Recovery	You must build and maintain backups	Built-in tools for backups and failover
Security Control	Full control	Shared responsibility model (cloud provider + you)
Customization	Full — you control all hardware/software	Limited to services offered by cloud provider

Internet Needed?	**No (can be isolated)**	**Yes, required for access**

Amazon S3 is like a giant cloud-based drive, and inside it:

- **You create Buckets (like top-level folders)**
- **Inside each Bucket, you can store files (called objects → each object has key which has its full path.)**

💡 Technically, S3 does not have real folders, but it pretends to, using file paths (prefixes). So it looks like folders and subfolders.

🖥️ What is Amazon EC2 (Elastic Compute Cloud)?

Amazon EC2 is a service that lets you run virtual computers (servers) in the cloud.

> **✅ It's like renting a computer in AWS's data center, where you can install your own operating system, software, databases, websites, etc.**

HDFS vs Amazon S3

- **HDFS (Hadoop Distributed File System)**
  ➤ Used to store files in Hadoop clusters.
  ➤ Files are split into blocks and stored across multiple servers close to the compute.
  ➤ Designed for fast data processing, not for internet access.
- **Amazon S3 (Simple Storage Service)**
  ➤ Stores data in the form of objects (files + metadata).
  ➤ Can be accessed from anywhere via the internet, even outside AWS.
  ➤ Designed for scalable, durable, and highly available storage.

Elastic IP (EIP)

- **Elastic IP is a static IP address that you can assign to your EC2 instance.**
- **It's static, meaning it doesn't change, even if your EC2 instance restarts.**

- If you need a permanent IP to connect to your EC2, you use an Elastic IP.

What is Amazon EBS?

Amazon EBS is like a virtual hard drive for your EC2 instance.

- **When you launch an EC2 (virtual computer in the cloud), it needs storage—just like your laptop needs a hard disk.**
- **That storage is provided by EBS.**

Feature	Amazon EBS	Amazon S3
Type	Block Storage	Object Storage
Use With	EC2 only	Used globally, web/mobile apps
Access	Mounted to EC2 only	Accessible over the internet
Example	Like your hard disk	Like Dropbox or Google Drive

🖥️ **EC2 needs a disk ➜ Use EBS**

🌍 **Need to store/share files anywhere ➜ Use S3**

🌐 What is VPC?

VPC stands for Virtual Private Cloud.

It’s your own private network inside AWS where you can launch and control your resources (like EC2, RDS, etc.) securely and privately — just like a private data center.

Component	Meaning (Simple)
Subnets	Divide your VPC into smaller sections — like rooms in your house
Route Tables	Rules for how traffic moves between subnets or to the internet
Internet Gateway (IGW)	Lets your VPC talk to the internet
NAT Gateway	Lets private instances access the internet outbound only (like updates)
Security Groups	Acts like a firewall — controls what comes in and out of your EC2 instance
VPC Peering	Connect two VPCs so they can communicate (like a bridge between two homes)

✅ When to Use What?

Situation	Use This
Store images, backups, logs	✅ S3
Store EC2 OS & DB data	✅ EBS

Share files between EC2s	**✅ EFS**
Control your private cloud infra	**✅ VPC**

✅ AWS Glue — *ETL Tool (Extract, Transform, Load)*

Purpose:
Used to prepare and move data from one place to another by:

- **Extracting data (from databases, S3, etc.)**
- **Transforming it (cleaning, converting formats)**
- **Loading it into another storage or data warehouse (like Redshift)**

Key Points:

- **Serverless – you don't manage servers.**
- **Written in PySpark (Python + Spark).**
- **Has Crawlers: scan data and create a catalog (like an index of your data).**
- **Has Jobs: actually transform and move the data.**

✅ AWS Redshift — *Data Warehouse*

Purpose:
Used to store and analyze large amounts of structured data using SQL queries.

Key Points:

- **Good for complex analytics on big datasets.**
- **Columnar storage – optimized for reading large data (not writing).**
- **Integrated with tools like QuickSight, Tableau, or DBeaver.**
