CHAPTER - 1

INTRODUCTION

Cloud computing has revolutionized the way businesses manage IT infrastructure. In 2006, **Amazon Web Services (AWS)** pioneered this transformation by offering IT services in the form of web-based, on-demand resources. This innovation eliminated the need for heavy upfront investment in physical infrastructure, allowing organizations to pay only for what they use and scale as needed.

With cloud computing, companies no longer wait for weeks to deploy servers. Instead, they can instantly launch hundreds or even thousands of instances within minutes. This flexibility leads to increased speed, agility, and reduced operational overhead.

Today, AWS powers the infrastructure of hundreds of thousands of organizations across more than 190 countries. Its services provide a secure, scalable, and cost-effective platform for businesses, educational institutions, and government bodies.

As part of this virtual internship, I explored various core modules of AWS, gaining hands-on experience in its cloud ecosystem. The following is an overview of the modules covered during the program:

1.1 Cloud Concepts Overview

In this module, I learned to:

- Identify different types of cloud computing models (IaaS, PaaS, SaaS)
- Understand the advantages of adopting cloud computing
- Recognize AWS core services and service categories

1.2 Cloud Economics and Billing

This module covered:

- Basics of AWS pricing strategies
- Estimating costs using AWS Pricing Calculator
- Navigating the AWS Billing Dashboard

1.3 AWS Global Infrastructure Overview

Key learning points included:

- Understanding AWS Regions, Availability Zones, and Edge Locations
- Overview of global AWS service distribution
- Categorization of AWS services

1.4 AWS Cloud Security

Focus areas in this module were:

- The Shared Responsibility Model
- Working with IAM users, roles, and groups
- Managing security credentials and access policies

1.5 Networking and Content Delivery

This module taught me about:

- Core networking principles in the cloud
- Amazon VPC architecture
- Use of Amazon Route 53 and CloudFront for DNS and content delivery

1.6 Compute Services

Topics included:

- Overview of AWS compute services like EC2, Lambda, and Elastic Beanstalk
- Use cases for server-based vs. serverless computing
- Deploying applications with scalability and minimal configuration

1.7 Storage Services

Exploration of storage solutions involved:

- Amazon S3 for object storage
- Amazon EBS for block-level storage
- Amazon S3 Glacier for archival and backup needs

1.8 Database Services

In this module, I learned about:

- Amazon RDS and its managed relational database capabilities
- Amazon DynamoDB for NoSQL applications
- Use of Redshift for data warehousing and Aurora for high-performance workloads

1.9 Cloud Architecture

Key takeaways included:

- The AWS Well-Architected Framework and its six pillars
- Designing highly available and fault-tolerant systems
- Using AWS Trusted Advisor for architecture optimization

1.10 Auto Scaling and Monitoring

This final module covered:

- Real-time monitoring using Amazon CloudWatch
- Enabling Auto Scaling for applications
- Load balancing strategies for varying workloads

CHAPTER - 2
TECHNOLOGY

AWS, as a cloud computing platform, incorporates a wide range of technologies that enable scalable, secure, and flexible virtual infrastructure. To succeed in an AWS-oriented role, you'll need a blend of cloud expertise, problem-solving skills, and strong familiarity with AWS tools and best practices. Below are the key technologies and competencies that power AWS virtual deployments:

1. Compute Services (Amazon EC2 and AWS Lambda)

A foundational grasp of EC2 allows users to launch and manage virtual servers in the cloud. Lambda supports server less computing, enabling event-driven applications without provisioning infrastructure.

2. Storage Technologies (Amazon S3, EBS, Glacier)

Understanding how to use S3 for object storage, Elastic Block Store (EBS) for persistent volumes, and Glacier for archival solutions is crucial for managing data efficiently and cost-effectively.

3. Networking and Security (Amazon VPC, IAM)

Configuring **Virtual Private Clouds (VPCs)** helps create isolated environments for workloads. **IAM (Identity and Access Management)** governs access controls and user permissions across services—essential for security and compliance.

4. Deployment & Automation (CloudFormation, AWS CLI, Elastic Beanstalk)

AWS supports Infrastructure as Code through **CloudFormation**. Proficiency in using **AWS CLI** and **Elastic Beanstalk** enables quick deployment, updates, and automation of applications and services.

5. Monitoring & Logging (CloudWatch, CloudTrail)

Tools like CloudWatch help monitor metrics and logs in real time, while CloudTrail provides a detailed audit trail of user activity across your AWS account—critical for debugging and maintaining operational integrity.

6. Databases (Amazon RDS, DynamoDB)

Knowledge of **Amazon RDS** (relational databases) and **DynamoDB** (NoSQL) is important for building data-driven applications. Both offer high availability and automated scaling.

7. Encryption & Compliance

AWS supports data encryption at rest and in transit using services like **KMS (Key Management Service)** and **CloudHSM**. Understanding compliance frameworks (e.g., GDPR, HIPAA) can help organizations meet regulatory requirements.

8. Collaboration and DevOps Practices

Effective cloud professionals use DevOps tools like **CodeCommit, CodeDeploy**, and **CI/CD pipelines** to streamline collaboration. Strong communication fosters teamwork across cross-functional groups including developers, analysts, and security specialists.

CHAPTER - 3
APPLICATIONS

Amazon Web Services (AWS) is quite possibly the most famous Cloud Computing platform embraced by many popular companies for various applications. As AWS has become universal, we must know where exactly we can use AWS services and what companies are using them. Here is the AWS applications list followed by a few AWS use cases.

1.Storage and Backup:

Storage and backup are important for any Cloud Computing service. AWS provides you with reliable storage services like **Amazon Simple Storage Service** to store large: scale data and backup services like AWS Backup to take backups of this data, which is stored in other AWS services. AWS stores the data in three different availability zones so that if one fails, you can still access your data. This makes AWS storage reliable and easily accessible. Therefore, companies with huge application data to store and backup securely can use AWS.

2.Big Data:

One of the biggest challenges faced by companies these days is **Big Data**. Companies are struggling to store their large amounts of data using traditional methods. With AWS Big Data storage services, they can manage to store their data even if the data limit increases unexpectedly as AWS provides virtually unlimited data storage with scale: in and scale: out options. AWS offers easy access and faster data retrieval as well. For data processing, it offers services like EMR, with which the companies can easily set up, operate, and scale their big data. Therefore, efficiently storing and managing Big Data is among the top AWS applications.

3. Enterprise IT:

AWS is a one: stop solution for any IT business. Many features of it such as secure storage, scalability, flexibility, and elasticity support companies to innovate faster than ever before. Using AWS for IT enterprises makes them profitable in terms of both money and time. As AWS maintains its ***cloud architecture***, it need not waste time and money on professionals to do the same.

4.Websites:

AWS offers a wide range of website hosting options to create the best website for customers. Its services like Amazon Light Sail have everything, such as a virtual machine, SSD: based storage, data transfer, DNS management, and a static IP, to launch a website in such a way that the user can manage the website easily. Amazon EC2, AWS Lambda, Elastic Load Balancing, AWS Amplify, Amazon S3, etc. also help users build reliable and scalable websites.

5.Gaming:

AWS has been serving many gaming studios. Combining Amazon EC2 and S3 services with **Cloud Front** enables gaming websites to deliver high: quality gaming experiences to their customers regardless of location.

6.Mobile Apps:

Mobile applications are embedded with day-to-day life. With AWS, you have the facility to create an app in your desired programming language. You can also keep up the applications that are consistently accessible and solid with high computer, storage, database, and application services.

CHAPTER - 4
Modules Explanation

4.1 AWS DATA ENGINEERING MODULES

Module 1: Welcome to AWS Academy Data Engineering

This module introduces learners to the AWS Academy Data Engineering course and provides an overview of what data engineering means in the modern cloud era. It explains the role of data engineers in organizations and how they help businesses make data-driven decisions. The module outlines the objectives, structure, and hands-on components of the course. Students learn about data pipelines—the backbone of data flow from source to analytics—and understand key pipeline stages such as ingestion, storage, processing, and analysis. It also introduces AWS cloud computing fundamentals, scalability, security, and reliability principles that are vital for building efficient data architectures. Learners are familiarized with core AWS services and the shared responsibility model. By the end of this module, students understand the overall goals of the course, what skills they will gain, and how AWS supports data-driven solutions.

Module 2: Data-Driven Organizations

This module explores the concept of data-driven organizations—companies that use data as a strategic asset to guide decision-making and innovation. Learners understand why businesses rely on accurate data for growth, forecasting, and competitiveness. The module explains how data is collected, processed, and transformed into insights that drive key performance indicators (KPIs). It introduces the collaboration between data engineers, analysts, and scientists in enabling business intelligence and analytics. The importance of data quality, governance, and democratization is discussed to ensure that trusted data reaches decision-makers. Case studies of successful data-driven organizations are shared to highlight best practices. By the end, students learn

how to align technical data processes with business goals and understand how a strong data culture supports organizational success.

Module 3: The Elements of Data

This module explains the fundamental characteristics of data, often described as the five Vs—Volume, Velocity, Variety, Veracity, and Value. Learners study how the scale, speed, and type of data affect pipeline design and storage choices. The module explores structured, semi-structured, and unstructured data, with examples like databases, JSON logs, and multimedia files. It also covers metadata, schemas, and the role of data profiling in improving quality. Students learn about data validation, cleaning, and transformation to ensure accuracy and reliability. File formats such as CSV, Parquet, and JSON are introduced, along with discussions on how they influence performance. Through this module, learners understand that good data is the foundation of every successful analytics or machine learning project, and they gain insight into the nature and lifecycle of data in real-world systems.

Module 4: Design Principles and Patterns for Data Pipelines

This module focuses on how to design efficient and scalable data pipelines using cloud architecture principles. It introduces the AWS Well-Architected Framework and the Analytics Lens, helping learners apply best practices for reliability, security, and cost optimization. Different pipeline architectures—batch, streaming, and hybrid—are discussed in detail. Students learn about modularity, fault tolerance, idempotency, and event-driven design patterns. The module highlights the importance of decoupling components and building reusable architecture templates. It also introduces AWS services such as Lambda, Glue, and Athena that support modern data engineering. Learners explore topics like schema evolution, orchestration, and logging. By the end, students can design robust, maintainable, and scalable pipelines that fit different data processing needs.

Module 5: Securing and Scaling the Data Pipeline

In this module, learners focus on how to build data pipelines that are secure, scalable, and reliable. It covers cloud security concepts, emphasizing the AWS shared responsibility model. Students learn how to implement access control using IAM, encrypt data in transit and at rest, and enforce compliance policies. Network security principles like VPC and private subnet design are introduced. The module then moves to scaling techniques—explaining horizontal and vertical scaling, auto-scaling, and elasticity. Learners understand how to balance performance and cost while managing large data workloads. Monitoring tools such as AWS CloudWatch and CloudTrail are introduced for auditing and performance tracking.

Module 6: Ingesting and Preparing Data

This module teaches how to ingest raw data from various sources and prepare it for analysis. Learners explore ETL (Extract, Transform, Load) and ELT (Extract, Load, Transform) workflows, understanding their differences and use cases. AWS Glue is introduced as a powerful tool for automated data ingestion and transformation. Students learn data cleaning techniques such as removing duplicates, handling null values, and normalizing formats. They also work on data validation, enrichment, and profiling to ensure high-quality datasets. Metadata management using Glue Data Catalog is covered. The module emphasizes how prepared data forms the foundation for reliable analytics and machine learning. Through hands-on labs, learners gain practical experience in transforming raw data into structured and usable information stored in Amazon S3 or databases.

Module 7: Ingesting by Batch or by Stream

This module explains two major data ingestion methods—batch and streaming. Batch ingestion handles data in scheduled intervals, while streaming ingestion captures and processes data in real time. Learners compare these approaches based on latency,

complexity, and scalability needs. The module introduces AWS services such as Glue for batch jobs and Kinesis or MSK (Kafka) for stream processing. Concepts like windowing, checkpointing, and delivery guarantees are explored in detail. Students learn how to handle high-velocity data from IoT devices or logs and integrate both batch and stream methods when needed. Real-time use cases such as fraud detection or live analytics are discussed. The module ends with practical demonstrations on setting up both ingestion types and understanding when to apply each strategy effectively.

Module 8: Storing and Organizing Data

This module focuses on how to store, structure, and manage data efficiently in the cloud. Learners explore different storage systems such as data lakes, data warehouses, and purpose-built databases. Amazon S3 is introduced as the foundation for building scalable data lakes, while Amazon Redshift is covered as a data warehousing solution for analytics. Students learn to differentiate between OLTP (transactional) and OLAP (analytical) systems. The importance of data partitioning, indexing, and compression is discussed to optimize cost and performance. The module covers lifecycle management, access control, and metadata cataloging. Learners also study the concept of lakehouse architecture and the integration of multiple storage layers. By the end, students can design and organize storage structures that support efficient querying and secure data management.

Module 9: Processing Big Data

This module introduces distributed data processing and big data frameworks. Learners understand the need for parallel computation when dealing with large datasets. The module covers Apache Hadoop, MapReduce, and Spark, explaining their architectures and key functions. Students use Amazon EMR to run scalable big data jobs and explore concepts like RDDs, DataFrames, and transformations in Spark. Techniques for job optimization, shuffling, and caching are discussed. Learners also explore tools

like Hive and Hudi for querying and incremental updates. Integrating EMR with S3 and Glue for end-to-end workflows is demonstrated. Monitoring and cost control for large-scale clusters are also emphasized. By completing this module, students gain the skills to process and analyze terabytes of data using distributed computing tools.

Module 10: Processing Data for Machine Learning

This module bridges data engineering and machine learning by focusing on preparing data for model training. Learners understand the difference between analytical and ML pipelines and explore preprocessing steps like feature extraction, normalization, and encoding. Feature engineering concepts such as one-hot encoding, scaling, and dealing with missing values are covered. Students learn to split data into training, validation, and test sets. The module introduces AWS SageMaker and how it integrates with other data services. Learners study feature stores, data drift detection, and reproducibility. Bias and fairness in datasets are also discussed. The importance of MLOps and automated retraining pipelines is highlighted. Through practical examples, students learn how clean, structured data enables more accurate and efficient ML models.

Module 11: Analyzing and Visualizing Data

This module teaches learners how to analyze and visualize data to uncover insights. It starts with SQL-based analytics using tools like Amazon Athena and Redshift for querying large datasets. Learners perform aggregation, filtering, and summarization to identify key trends. Visualization concepts are introduced using Amazon QuickSight, where students design dashboards and charts for decision support. The module explains how to select the right visualization type and how storytelling improves data communication. Best practices for performance, accessibility, and interactivity are also covered. Learners connect multiple data sources and build secure dashboards with role-based access.

Module 12: Automating the Pipeline

This module covers how to automate and orchestrate data workflows for efficiency and reliability. Learners understand why manual pipelines are error-prone and time-consuming. Tools such as AWS Step Functions, Apache Airflow, and AWS Glue Workflows are introduced for automation. The module teaches scheduling, dependency management, retries, and error handling. Students learn how to integrate CI/CD concepts and infrastructure-as-code (IaC) tools like AWS CloudFormation for deploying pipelines. Monitoring and alerting through CloudWatch are covered to ensure smooth operations. Automation in data validation, ingestion, and transformation is demonstrated through labs. By the end, learners can build fully automated, self-healing data pipelines that improve productivity, reduce downtime, and ensure consistent data delivery.

4.2 AWS CLOUD FOUNDATIONS MODULES

Module 1: Cloud Concepts Overview

Different types of cloud computing models are Infrastructure as a service (IaaS), Platform as a service (PaaS), Software as a service (SaaS). There are three cloud deployment models: cloud, hybrid, and on premises or private cloud. Six advantages of cloud computing are Trade capital expense for variable expense, massive economies of scale, stop guessing capacity, increase speed and agility, stop spending money on running and maintaining data centers, Go global in minutes. AWS is a secure cloud platform that offers a broad set of global clouds: based products called services that are designed to work together. There are many categories of AWS services, and each category has many services to choose from. Choose a service based on your business goals and technology requirements. Cloud adoption is not instantaneous for most organizations and requires a thoughtful, deliberate strategy and alignment across the whole organization.

Module 2: Cloud Economics and Billing

There are three fundamental drives of cost with AWS: compute, storage, and outbound data transfer. These characteristics vary somewhat, depending on the AWS product and pricing model you choose. AWS offers a range of cloud computing services. For each service, you pay for exactly the number of resources that you need. This utility: style pricing model includes Pay for what you use, pay less when you reserve, pay less when you use more, pay even less as AWS grows. Total Cost of Ownership is a concept to help you understand and compare the costs that are associated with different deployments. AWS provides the AWS Pricing Calculator to assist you with the calculations that are needed to estimate cost savings. AWS Billing and Cost Management provides you with tools to help you access, understand, allocate, control, and optimize your AWS costs and usage. These tools include AWS Bills, AWS Cost Explorer, AWS Budgets, and AWS Cost and Usage Reports. Knowing and understanding your usage and costs will enable you to plan ahead and improve your implementations.

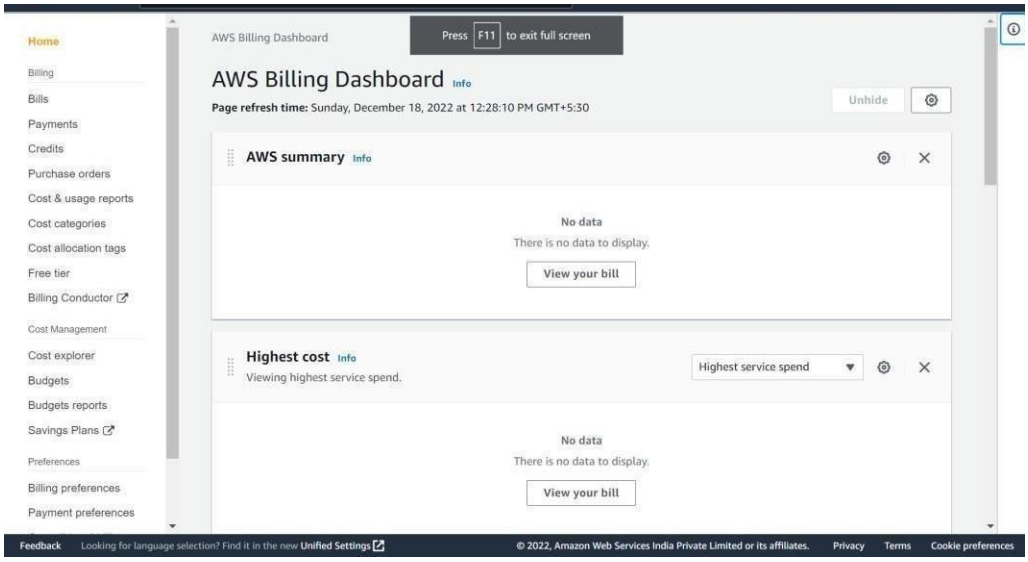


Fig: 4.1 AWS Billing Dashboard

Module 3: AWS Global Infrastructure Overview

The AWS Global Infrastructure is designed and built to deliver a flexible, reliable, scalable, and secure cloud computing environment with high: quality global network performance.

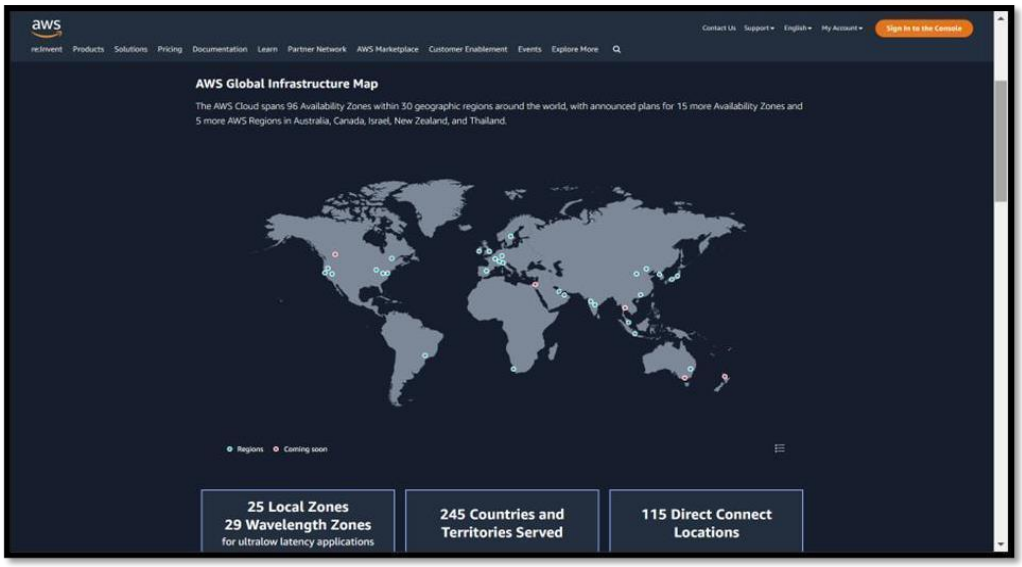


Fig: 4.2 AWS Global Infrastructure Map

The AWS Cloud infrastructure is built around regions. AWS has 22 Regions worldwide. An AWS Region is a physical geographical location with one or more Availability Zones. Availability Zones in turn consist of one or more data centers.

Region: Region is a physical location around the world where we cluster data centers.

Availability Zones: An Availability Zone (AZ) is one or more discrete data centers with redundant power, networking, and connectivity in an AWS Region. AZs give customers the ability to operate production applications and databases that are more highly available, fault tolerant, and scalable than would be possible from a single data center.

Edge locations: Edge locations are AWS data centers designed to deliver services with the lowest latency possible.

Module 4: AWS Cloud Security

Security and compliance are a shared responsibility between AWS and the customer. This shared responsibility model is designed to help relieve the customer's operational burden. AWS responsibility: AWS operates, manages, and controls the components from the software virtualization layer down to the physical security of the facilities where AWS services operate. AWS is responsible for protecting the infrastructure that runs all the services that are offered in the AWS Cloud.

Customer responsibility: The customer is responsible for the encryption of data at rest and data in transit. The customer should also ensure that the network is configured for security and that security credentials and logins are managed safely. Additionally, the customer is responsible for the configuration of security groups and the configuration of the operating system that runs on computer instances that they launch. IAM policies are constructed with JavaScript Object Notation and define permissions. IAM policies can be attached to any IAM entity. Entities are IAM users, IAM groups, and IAM roles. An IAM user provides a way for a person, application, or service to authenticate to AWS. An IAM group is a simple way to attach the same policies to multiple users.

Module 5: Networking and Content Delivery

A computer network is two or more client machines that are connected together to share resources. A network can be logically partitioned into subnets. Networking requires a networking device (such as a router or switch) to connect all the clients together and enable communication between them. An elastic network interface is a virtual network interface that you can attach or detach from an instance in a VPC.

A network interface's attributes follow it when it is attached to another instance. When you move a network interface from one instance to another, network traffic is redirected to the new instance. An internet gateway is a scalable, redundant, and highly available VPC component that allows communication between instances in your VPC and the internet.

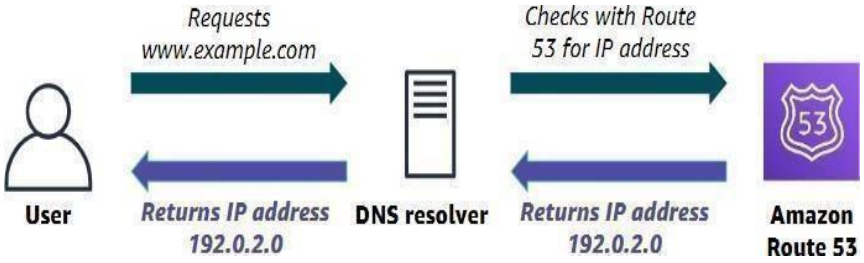
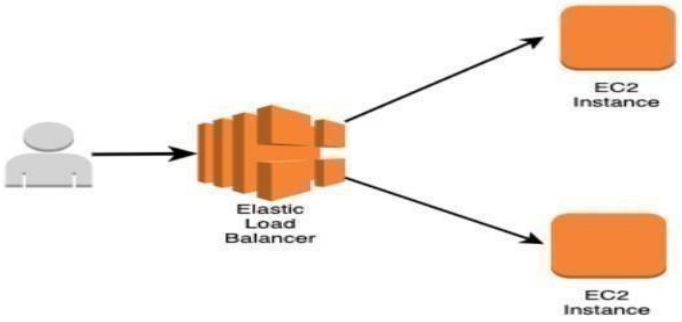


Fig: 4.3 Content Delivery

Module 6: Compute

Amazon EC2 provides virtual machines, and you can think of it as infrastructure as a service (IaaS). Iaas services provide flexibility and leave many of the server management responsibilities to you. You choose the operating system, and you also choose the size and resource capabilities of the servers that you launch. For IT professionals who have experience using computing on premises, virtual machines are a familiar concept. Amazon EC2 was one of the first AWS services, and it remains one of the most popular services. **AWS Lambda** is a zero: administration computing platform. AWS Lambda enables you to run code without provisioning or managing



servers. You pay only for the computer time that is used. This server less technology concept is relatively new to many IT professionals. However, it is becoming more popular because it supports cloud: native architectures, which enable massive scalability at a lower cost than running servers 24/7 to support the same workloads.

AWS Elastic Beanstalk provides a platform as a service (PaaS). It facilitates the quick deployment of applications that you create by providing all the application

services that you need. AWS manages the OS, the application server, and the other infrastructure components so that you can focus on developing your application code.

Fig: 4.4 AWS EC2, AWS Lambda, AWS Elastic Beanstalk

Module 7: Storage

Amazon EBS provides persistent block storage volumes for use with Amazon EC2instances. Persistent storage is any data storage device that retains data after power to that device is shut off. It is also sometimes called non: volatile storage.

Fig: 4.5 Amazon EBS

Amazon S3 is object: level storage, which means that if you want to change a part of a file, you must make the change and then reupload the entire modified file. Amazon S3 stores data as objects within resources that are called buckets

Fig: 4.6 Amazon S3

Amazon S3 Glacier is a secure, durable, and extremely low: cost cloud storage

service for data archiving and long: term backup. Data that is stored in Amazon S3 Glacier can take several hours to retrieve, which is why it works well for archiving.

Amazon S3 Glacier

Fig: 4.7 Amazon S3 Glacier

Module 8: Database

Amazon RDS is a web service that makes it easy to set up, operate, and scale a relational database in the cloud. It provides cost: efficient and resizable capacity while managing time: consuming database administration tasks so you can focus on your applications and your business. Features include that it is a managed service, and that it can be accessed via the console, AWS Command Line Interface (AWS CLI), or application programming interface (API) calls. Amazon RDS is scalable for computer and storage, and automated redundancy and backup are available. Supported database engines include Amazon Aurora, PostgreSQL, MySQL, MariaDB, Oracle, and Microsoft SQL Server.

Amazon Relational Database Service (Amazon RDS)

Fig: 4.8 Amazon RDS

Amazon DynamoDB

Fig :4.9 Amazon DynamoDB

Amazon Redshift is a fast, fully managed data warehouse that makes it simple and cost: effective to analyze all your data by using standard SQL and your existing business intelligence (BI) tools. Here is a look at Amazon Redshift and how you can use it for analytic applications.

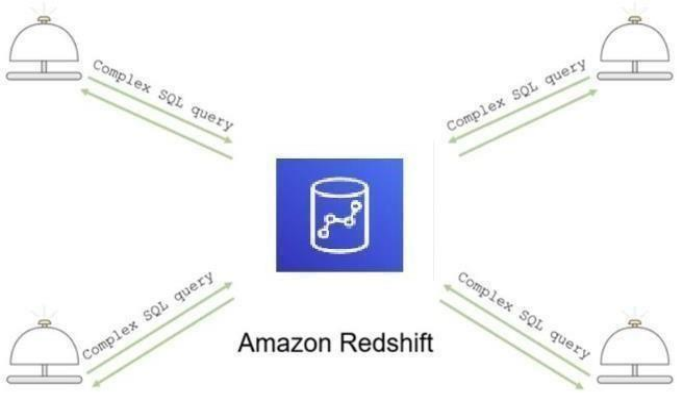


Fig: 4.10 Amazon Redshift

Amazon Aurora is a MySQL: and PostgreSQL: compatible relational database that is built for the cloud. It combines the performance and availability of high: end commercial databases with the simplicity and cost: effectiveness of open: source databases. Using Amazon Aurora can reduce your database costs while improving the reliability and availability of the database.

Fig: 4.11 Amazon Aurora

Module 9: Cloud Architecture

The AWS Well: Architected Framework is organized into six pillars: operational excellence, security, reliability, performance efficiency, cost optimization, and sustainability. The first five pillars have been part of the framework since the framework's introduction in 2015. The sustainability pillar was added as the sixth pillar in 2021 to help organizations learn how to minimize the environmental impacts of running cloud workloads.

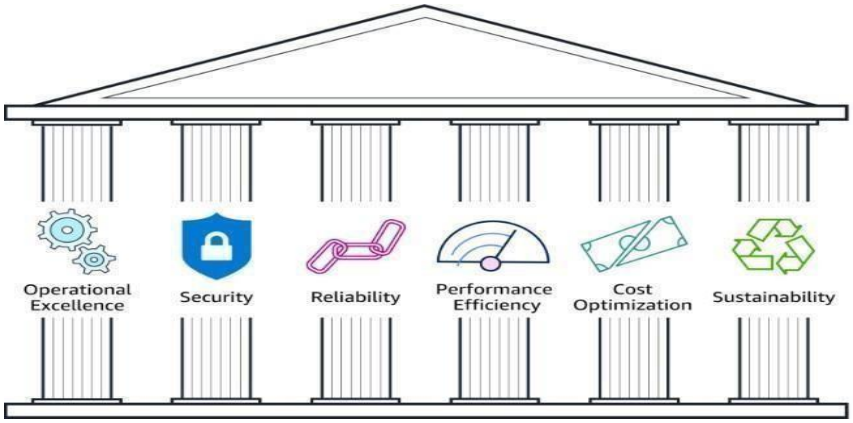


Fig: 4.12 Six Pillars of AWS

Reliability is a measure of your system's ability to provide functionality when desired by the user. Because "everything fails, all the time," you should think of reliability in statistical terms. Reliability is the probability that an entire system will function as intended for a specified period. Note that a system includes all system components, such as hardware, firmware, and software. Failure of system components

impacts the availability of the system. **Availability** is the percentage of time that a system operates normally or correctly performing the operations expected of it (or normal operation time over total time). Availability is reduced anytime the application is not operating normally, including both scheduled and unscheduled interruptions. **AWS Trusted Advisor** is an online tool that provides real: time guidance to help you provision your resources by following AWS best practices. AWS Trusted Advisor looks at your entire AWS environment and gives you real: time recommendations in five categories. You can use AWS Trusted Advisor to help you optimize your AWS environment as soon as you start implementing your architecture designs.

Module 10: Auto Scaling and Monitoring

Amazon CloudWatch is a monitoring and observability service that is built for DevOps engineers, developers, site reliability engineers (SRE), and IT managers. CloudWatch monitors your AWS resources (and the applications that you run on AWS) in real time. You can use CloudWatch to collect and track metrics, which are variables that you can measure for your resources and applications.

Fig: 4.13 Amazon CloudWatch

Elastic Load Balancing distributes incoming application or network traffic across multiple targets (such as Amazon EC2 instances, containers, IP addresses, and Lambda functions) in one or more Availability Zones. Elastic Load Balancing offers several monitoring tools for continuous monitoring and logging for auditing and analytics.

CHAPTER - 5

Real Time Example

Netflix – Powering Global Streaming with AWS

Netflix, the world's leading video streaming platform, serves over 260 million subscribers in more than 190 countries. To deliver high-quality video content to users worldwide, Netflix relies heavily on **Amazon Web Services (AWS)** for nearly all its infrastructure needs.

Why AWS for Netflix?

Netflix faced the challenge of delivering seamless video experiences globally while handling unpredictable spikes in traffic (e.g., new releases, live events). To meet this demand, they needed a scalable, reliable, and low-latency cloud platform.

AWS Services Used by Netflix:

- **Amazon EC2** – Hosts Netflix's streaming, encoding, and recommendation engines with flexibility to scale resources based on demand.
- **Amazon S3** – Stores video content, backup data, and user metadata with durability and availability across regions.
- **Amazon CloudFront** – Distributes content globally through edge locations, ensuring fast video delivery and minimal buffering.
- **AWS Lambda** – Automates backend tasks such as transcoding logs, analytics triggers, and dynamic scaling.
- **Amazon RDS & DynamoDB** – Manages both structured and unstructured data related to user profiles, billing, and watch history.

Impact of AWS on Netflix Operations:

- Netflix handles **petabytes of data daily** with AWS's robust storage and analytics systems.

- It achieves **99.99% uptime** and quick content delivery across continents.
- Netflix engineers can **deploy new features instantly** without disrupting service.
- With **AWS Auto Scaling and Load Balancing**, Netflix manages spikes in traffic during peak hours or series launches effortlessly.

By building on AWS, Netflix maintains high performance, global scalability, and real-time analytics, all while focusing on user experience and innovation rather than infrastructure maintenance.

Zomato – Scaling Food Delivery with AWS Cloud

Zomato, one of India's largest food delivery platforms, relies on **Amazon Web Services (AWS)** to handle millions of transactions, manage real-time restaurant data, and deliver seamless user experiences across the country.

With a user base spanning across hundreds of cities and thousands of partner restaurants, Zomato required a cloud platform that could provide high scalability,

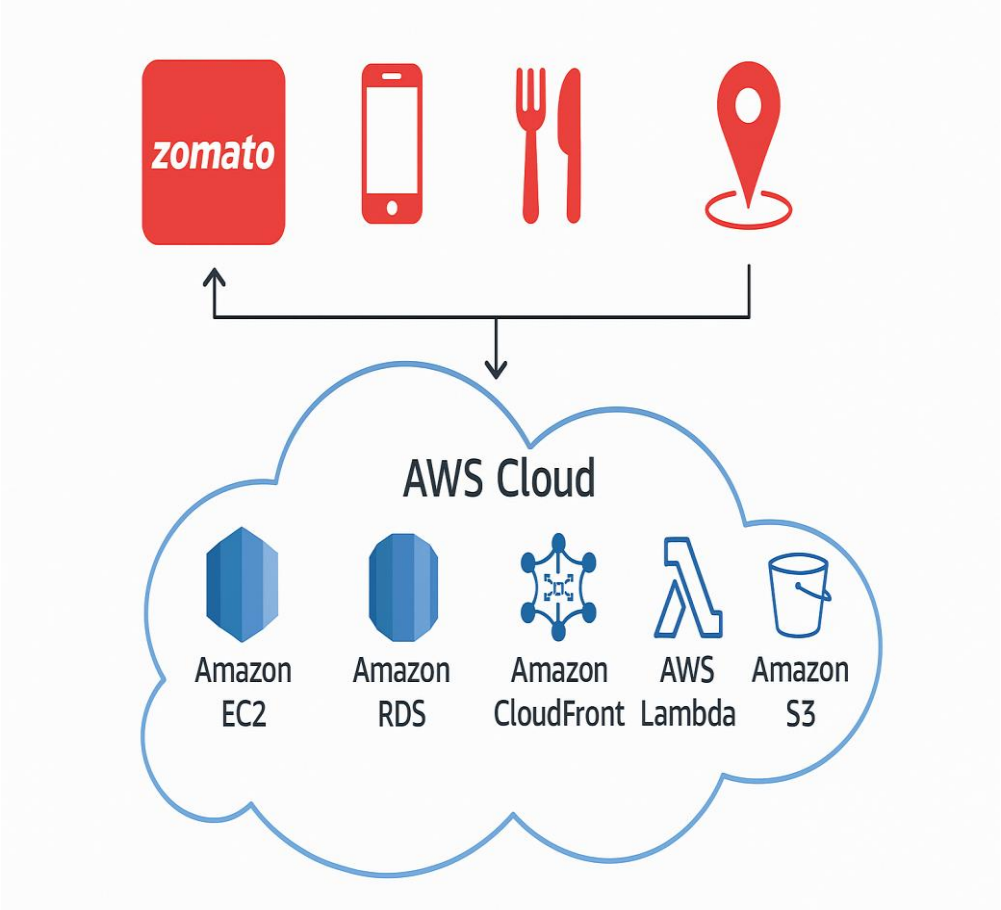
uptime, security, and performance. AWS offered a complete solution with services such as:

- **Amazon EC2** for scalable compute instances to host the core application backend.
- **Amazon RDS** for managing relational databases with high availability and automated backup.
- **Amazon CloudFront** for low-latency delivery of images, menus, and dynamic content across mobile and web platforms.
- **AWS Lambda** to run serverless functions like payment processing, notifications, and real-time order updates.
- **Amazon S3** for storing restaurant images, customer receipts, and delivery logs securely.

Zomato also uses **Elastic Load Balancing (ELB)** and **Auto Scaling** to manage traffic surges during peak hours like lunch and dinner time. For user data security and compliance, they implement strict IAM roles, encrypted data storage, and regular audits using AWS CloudTrail and AWS Config.

By leveraging AWS, Zomato can scale instantly based on demand, provide personalized experiences, and maintain operational efficiency. Their cloud-native architecture supports fast deployment of new features, reduces downtime, and ensures customer satisfaction even during high-load scenarios like national festivals and cricket

match days. This real-world use case clearly demonstrates how AWS can support the growth, security, and agility of a high-traffic, customer-facing application in a competitive industry.



CHAPTER - 6

Learning outcomes

During the course of this virtual internship on **Amazon Web Services (AWS)**, I acquired valuable technical and professional skills. The following are the key learning outcomes from this program:

➢Gain a deep understanding of cloud computing principles, including its benefits, deployment models (public, private, hybrid), and service models (IaaS, PaaS, SaaS).

➢ Understand how to create, configure, and manage AWS resources using the AWS Management Console.

➢ Learn how to design and implement cloud architectures that are scalable, resilient, and cost-effective, following AWS best practices.

➢ Data Engineering Concepts: Developed a solid understanding of core data engineering concepts like ETL processes, big data processing, and real-time data streaming, using AWS services such as AWS Glue, Amazon Kinesis, and Amazon EMR.

➢ Hands-on Experience with AWS Tools: Acquired practical experience in deploying and managing cloud-based data pipelines, from data ingestion to transformation, using services like Amazon Redshift, Amazon Athena, and AWS Lambda for automation.

➢ Understood how AWS integrates with machine learning tools, facilitating data preparation, transformation, and training using Amazon SageMaker, and applied these concepts to real-world datasets.

➢ Explore emerging technologies in the cloud space, such as artificial intelligence (AI), machine learning (ML), Internet of Things (IoT), and edge computing, and how they integrate with AWS services.

Conclusion

In conclusion, AWS Data Engineering Virtual internship has been a transformative experience, offering not only theoretical insights but also hands-on practice with industry-standard tools and technologies. Over the course of the internship, I gained a deep understanding of data engineering fundamentals, particularly in leveraging AWS services like Amazon S3, Redshift, Glue, and Lambda to create, maintain, and optimize data pipelines.

One of the major highlights of the internship was the opportunity to work with real-world datasets, which enhanced my understanding of the complexities involved in data ingestion, transformation, and storage at scale. Through projects like building ETL pipelines and designing efficient data architectures, I learned how to optimize workflows for performance, reliability, and cost-effectiveness.

By the end of the internship, I had significantly improved my technical proficiency with AWS and cloud computing, developed a more strategic approach to data engineering, and gained confidence in handling large, complex datasets. This experience has been instrumental in solidifying my interest in data engineering and has equipped me with the skills and mindset needed to excel in future projects and roles within the field. It equipped me with the skills and confidence to contribute meaningfully to future data-driven projects and provided a strong foundation for continued learning and growth in the field of cloud computing and data engineering.

INTERNSHIP CERTIFICATE

REFERENCES

- https://awsacademy.instructure.com/courses/133060/modules
- https://awsacademy.instructure.com/courses/18381
- https://aws.amazon.com/education/awseducate/
- https://aws.amazon.com/student-hub/
- https://aws.amazon.com/events/online-tech-talks/
- https://www.aws.training/Labs

.