An internship report

on

AI-ML Virtual Internship

Submitted in partial fulfillment of the requirements

for the award of the degree of

BACHELOR OF TECHNOLOGY

in

Computer Science and Engineering (Data Science)

by

G. USHODAYA   **204G1A3260**

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
(DATA SCIENCE)

SRINIVASA RAMANUJAN INSTITUTE OF TECHNOLOGY (AUTONOMOUS)

(Affiliated to JNTUA, accredited by NAAC with 'A' Grade, Approved by AICTE, New Delhi & Accredited by NBA (EEE, ECE & CSE))
Rotarypuram village, B K Samudram Mandal, Ananthapuramu-515701.

2023 - 2024

SRINIVASA RAMANUJAN INSTITUTE OF TECHNOLOGY (AUTONOMOUS)

**(Affiliated to JNTUA, accredited by NAAC with 'A' Grade, Approved by AICTE, New Delhi & Accredited by NBA (EEE, ECE & CSE))
Rotarypuram village, B K Samudram Mandal, Ananthapuramu-515701.**

Department of Computer Science & Engineering (Data Science)

Certificate

This is to certify that the internship report entitled AI-ML Virtual Internship is the bonafide work carried out by **G.Ushodaya** bearing Roll Number 204G1A3260 in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering (Data Science)** for three months from May 2023 to July 2023.

Internship Coordinator

Dr. G. Hemanth Kumar Yadav, M.Tech., Ph.D.,
Associate Professor

Head of the Department

Dr. P. Chitralingappa, M.Tech., Ph.D.,
Associate Professor

Date:

Place: Ananthapuramu

EXTERNAL EXAMINER

PREFACE

All India Council for Technical Education (AICTE) has initiated various activities for promoting industrial internship at the graduate level in technical institutes and EduSkills is a Non-profit organization which enables Industry 4.0 ready digital workforce in India. The vision of the organization is to fill the gap between Academic and Industry by ensuring world class curriculum access to the faculties and students. Formation of the All-India Council for Technical Education (AICTE) in1945 by the Government of India.

Purpose: With a vision to create an industry-ready work force who will eventually become leaders in emerging technologies, EduSkills & AICTE launches 'Virtual Internship' program on Cloud Technology, supported by Amazon Web Services (AWS). Demand for the cloud has shot through the roof since the beginning of the pandemic, as businesses try to build resiliency and AWS if the Pioneer.

Company's Mission Statement: The main mission of these initiatives is enhancement of the employability skills of the students passing out from Technical Institutions.

Business Activities:
- Contact Center.
- Supply Chain.
- Physical Stores.
- Communication APIs and SDKs.
- Secure Communications.
- Productivity Applications.

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of people who made it possible, whose constant guidance and encouragement crowned our efforts with success. It is a pleasant aspect that I have now the opportunity to express my gratitude for all of them.

It is with immense pleasure that I would like to express my indebted gratitude to my internship coordinator **Dr. G. Hemanth Kumar Yadav,** M.Tech, Ph.D, **Associate Professor & HOD, Computer Science and Engineering (AI & ML),** who has supported me a lot and encouraged me in every step of the internship work. I thank him for the stimulating support, constant encouragement and constructive criticism which have made possible to bring out this internship work.

I am very much thankful **Dr. P. Chitralingappa,** M.Tech, Ph.D, **Associate Professor & HOD, Computer Science and Engineering (Data Science),** for his kind support and for providing necessary facilities to carry out the work.

I wish to convey my special thanks to **Dr. G. Balakrishna,** M.Tech, Ph.D, **Principal** of **Srinivasa Ramanujan Institute of Technology** for giving the required information in doing my internship. Not to forget, I thank all other faculty and non- teaching staff, and my friends who had directly or indirectly helped and supported me in completing my internship in time.

I also express our sincere thanks to the Management for providing excellent facilities and support**.**

Finally, I wish to convey my gratitude to my family who fostered all the requirements and facilities that I need.

<div align="right">

G USHODAYA
(204G1A3260)

</div>

Contents   # Page No.

LISTOF FIGURES

LIST OF ABBREVIATIONS

CFM	Cloud Financial Management
AWS	Amazon Web Services
IAM	Identity Access Management
VPC	Virtual Private Cloud
CIDR	Classless Inter-Domain Routing
IaaS	Infrastructure as a Service
PaaS	Platform as a Service
SaaS	Software as a Service
S3	Simple Storage Service
RDS	Relational Database Service

CHAPTER 1

INTRODUCTION

The AWS Cloud Adoption Framework for Artificial Intelligence and Machine Learning is a starting and orientation point throughout your ML and AI journey. AI and ML have evolved from niche technologies to a powerful and broadly available business capability. ML is by now fuelling a new wave of information.

Iterating through our AI/ML transformation journey relies on your foundational capabilities to adopt AI/ML across business, people, governance, security, and operations. A capability is an organizational ability to use processes to deploy resources (such as people, technology, and other tangible or intangible assets) to achieve an outcome.

Business Perspective

There are different managements regarding business perspective those are Strategy, Product, Business, Portfolio, Innovation, Data Monetization, Strategic Partnership and Data Science.

Governance Perspective

People are one Cornerstone of AI/ML adoption, another cornerstone is managing, optimizing, and scaling the organizational AI/ML initiative. As much as the cloud offered new days of looking at operational expense (OPEX) versus capital expense (CAPEX), AI/ML requires adopters to look at cost structures in a different way while managing the organizational risks and opportunities that rise from that.

There are different foundation capabilities in Governance perspective those are Cloud Financial Management (CFM), Data Curation, Risk Management, Program and Project Management, Data Governance, Benefits Management, Application Portfolio Management. We need to consider that its fiendishly easy to build a first proof of concept (POC) in AI/ML, but both solutions will enable us a long term investment.

CHAPTER 2

TECHNOLOGY

Cloud architects are projected to be the second most in-demand tech job in 2021. This role is responsible for designing and developing advanced cloud-based solutions for organizations migrating their existing workloads and infrastructure to the AWS cloud. Using AWS, cloud architects have limitless virtual resources, which can be quickly provisioned and disposed. It can be overwhelming; there are many services you need to become familiar with and, on top of that, having infrastructure and data in the cloud can become a security nightmare if not handled correctly. Here are seven skills cloud architects need to rock this in demand role.

Java, Python or C#

Most architects have a software development background. An efficient AWS architect should be able to write code in Java, Python, C# or any other of the programming languages which have an official AWS SDK. Understanding programming in general is important for creating viable, logical solutions that would work as intended. And a good architect can use programming to quickly create a proof of concept or demo to show a point or investigate how to use the latest and greatest technologies.

Networking

It's hard to create a secure, scalable cloud-based solution without understanding networking. DNS, TCP/IP, HTTP, CDN and VPN are only a few of the terms you want to make yourself familiar with. That doesn't mean you need to know the port that you need to open for SSH access (although it helps). As an architect you're expected to be able to use services such as Route 53 (DNS), CloudFront (CDN) and Virtual Private Cloud (VPC) to design your cloud networking using public and private subnets, internet access and VPC peering.

Data storage fundamentals

Every software architect needs to know and understand how and when to use databases.

In AWS, where you have many data storage options available, you need to be able to know when to use each. From simple, yet powerful, bucket storage using S3 to Relational Database. Service (RDS) and all the way to full-fledged Hadoop clusters, you'll need to compare different capabilities, performance and price, and choose the best way to store some or all of your company's data.

Security foundations

From securing access to your AWS account to securing access to your data, AWS has several services and guidelines created specifically to help you make sure only authorized code and people are allowed to perform specific tasks. You will also need to learn about Identity and Access Management (IAM), a service that will help you define which services and users can access which resources. Learn how to secure your networks using Security Groups and Access Control Lists.

AWS service selection

Cloud architecture involves front-end and back-end technologies backed by components provided by a cloud vendor, in many cases Amazon. Good cloud architects should know what services are at their disposal and have a good understanding of the services relevant to the organization. This is no easy task since Amazon tends to release new services throughout the year. Knowing what to focus on is crucial. Basic services every AWS architect should know include SQS (simple queuing), SNS (notifications) and RDS (Relational Database Service). Knowledge of more specific ones—such as one of the AWS IoT related services—is extra helpful.

Cloud-specific patterns and technologies

Once you move code into the cloud some rules change. Scalability, availability and recovery become easy—as long as you design your workloads and harness AWS infrastructure correctly. Using messages, storing state in the right place and handling failures correctly are a big part of creating scalable and cost-effective applications. On top of that you'll need to choose between single applications to services or go serverless using AWS Lambda all of which could provide cost-effective and performant solutions to make them.

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**. The 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. Social Networking

Social networking is essential for businesses in the present-day scenario where Digital Marketing is key, and it is easier with AWS. Companies can connect with customers and

stakeholders and communicate through social networking sites and develop their business. Services like AWS social networking engine, which is powered by TurnKey GNU/Linux (HVM) AMI stack, are used for performance and scalability to help companies build a suitable social networking site and gain profits.

5. 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 compute, storage, database, and application services. You can take advantage of **AWS auto-scaling** and managed relational database service for the better performance of your apps.

6. Websites

AWS offers a wide range of website hosting options to create the best website for customers. Its services like Amazon LightSail 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.

7. 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.

CHAPTER 4
MODULE EXPLANATION

Module-1: Cloud Concepts Overview

The module describes about on-demand delivery of compute power, database, storage, applications, and other IT resources via the internet with pay-as-you-go pricing. Cloud computing enables you to stop thinking of infrastructure as hardware, and instead think of (and use) it as software.

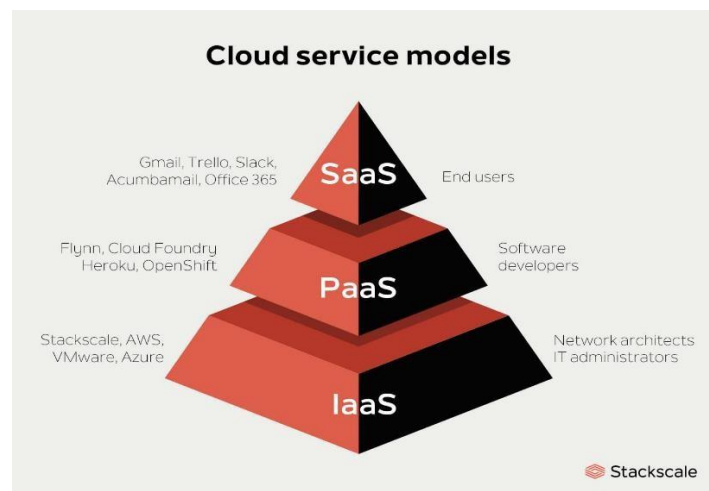Cloud service models vary on how much control you have over IT resources.

Fig 4.1: Cloud Service Model

Cloud computing deployment models

- Cloud

- Hybrid

- On-premise (Private Cloud)•

Advantages of Cloud Computing:

- Pay only for the resources you consume (variable cost vs upfront capital expenditure)
- Economies of scale achieved by aggregate of all users
- Scaling on demand
- Speed and flexibility - changes are software level, not hardware like traditional computing
- Lower overhead due to not maintaining hardware and data centers
- Data centers are global, like a company's customer base

Module-2: Cloud Economics and Billing

Three Fundamental Cost Drivers with AWS

- Compute - charged by use time, varies by instance

- Storage - charged per GB

- Data Transfer - outbound transfers are aggregated and charged per GB, inbound transfers and data transfers between services in the same AWS Region typically have no charge

- Paying for AWS Pay for what you use

- Reserve and save up to 75% versus On-Demand

  i. All Upfront Reserved Instance (AURI) -> Large Discount

  ii. Partial Upfront Reserved Instance (PURI) -> Lower Discount

  iii. No Upfront Payments Reserved Instance (NURI) -> Smallest Discount

- Scale and save as usage increases

  i.   Tiered pricing for services like S3, EBS, EFS

- Custom Pricing

  i.   Meet varying needs through custom pricing.

  ii.   Available for high-volume projects with unique requirements.

Total Cost of Ownership

The financial estimate to help identify direct and indirect costs of a system. To Compare the costs of running an entire infrastructure environment or specific workload on-premises versus on AWS, budget and build the business case for moving to the cloud

Billing

AWS Organizations: An account management service that enables you to consolidate multiple AWS accounts into an organization that you create and centrally manage. AWS Organizations includes account management and consolidated billing capabilities that enable you to better meet the budgetary, security, and compliance needs of a business.

Module-3: AWS Global Infrastructure Overview

AWS Global Infrastructure

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.

AWS Regions

An AWS Region is a geographical area.

- Data replication across Regions is controlled by you.
- Communication between Regions uses AWS backbone network infrastructure.
- Each Region provides full redundancy and connectivity to the network.
- A Region typically consists of two or more Availability Zones

AWS Storage Services

- Amazon Simple Storage Services (S3) - Object storage service that offers industry-leading scalability, data availability, security, and performance.
- Amazon Elastic Block Storage (EBS) - An easy to use, high performance block storage service designed for use with Amazon Elastic Compute Cloud (EC2) for both throughput and transaction intensive workloads at any scale.
- Amazon Elastic File System (EFS) - A simple, scalable, fully managed elastic NFS file system for use with AWS Cloud services and on-premises resources.
- Amazon Simple Storage Service Glacier - A secure, durable, and extremely low-cost Amazon S3 cloud storage classes for data archiving and long-term backup.

AWS Compute Services

- Amazon EC2 - A web service that provides secure, resizable compute capacity in the cloud.
- Amazon EC2 Auto Scaling - Helps to maintain application availability and allows you to automatically add or remove EC2 instances according to conditions you define.
- Amazon Elastic Container Services (ECS) - A fully managed container orchestration service.
- Amazon EC2 Container Registry (ECR) - A fully-managed Docker container registry that makes it easy for developers to store, manage, and deploy Docker container images.
- AWS Elastic Beanstalk - An easy-to-use service for deploying and scaling web applications and services developed with Java, .NET, PHP, Node.js, Python, Ruby, Go, and Docker on familiar servers such as Apache, Nginx, Passenger, and IIS.
- AWS Lambda - Lets you run code without provisioning or managing servers.
- Amazon Elastic Kubernetes Services (EKS) - A fully managed Kubernetes service.
- AWS Fargate - A serverless compute engine for containers that works with both Amazon Elastic Container Service (ECS) and Amazon Elastic Kubernetes Service (EKS).
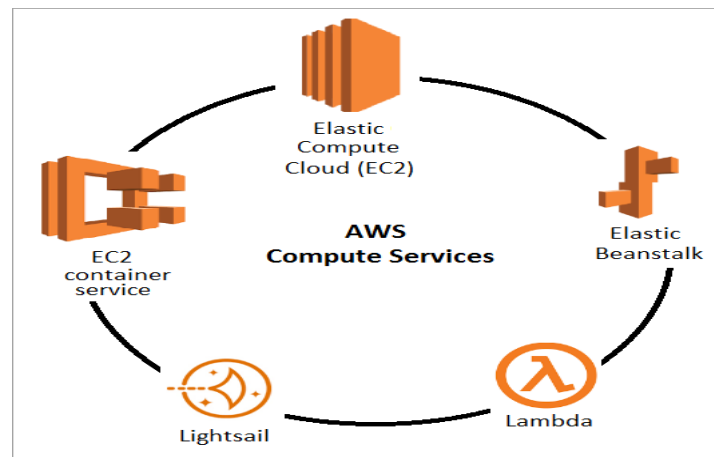
Fig 4.2: AWS Compute Services

Module-4: AWS Cloud Security

AWS Shared Responsibility Model

AWS security is divided by part of the cloud: customers are responsible for security **in** the cloud, AWS is responsible for security **of** the cloud.

Customer Security

- Amazon Elastic Compute Cloud (Amazon EC2) instance operating system - Including patching, maintenance
- Applications - Passwords, role-based access, etc.
- Security group configuration
- OS or host-based firewalls - Including intrusion detection or prevention systems
- Network configurations
- Account management - Login and permission settings for each user

AWS Security

- Physical security of data centers - Controlled, need-based access
- Hardware and software infrastructure - Storage decommissioning, host operating system (OS) access logging, and auditing
- Network infrastructure - Intrusion detection
- Virtualization infrastructure - Instance isolation

Identity and Access Management (IAM):

AWS Identity and Access Management (IAM) is a web service that enables Amazon Web Services (AWS) customers to manage users and user permissions in AWS. With IAM, you can centrally manage users, security credentials such as access keys, and permissions that control which AWS resources users can access.

AWS Identity and Access Management (IAM) can be used to:

- **Manage IAM Users and their access:** You can create Users and assign them individual security credentials (access keys, passwords, and multi-factor authentication devices). You can manage permissions to control which operations a User can perform.
- **Manage IAM Roles and their permissions:** An IAM Role is similar to a User, in that it is an AWS identity with permission policies that determine what the identity can and cannot do in AWS. However, instead of being uniquely associated with one person, a Role is intended to be assumable by anyone who needs it.
- **Manage federated users and their permissions:** You can enable identity federation to allow existing users in your enterprise to access the AWS Management Console, to call AWS APIs and to access resources, without the need to create an IAM User for each identity.

Module-5: Networking and Content Delivery

Networking Basics

Network: Two or more machines that are connected together in order to communicate. A network can be divided into subnets and networking requires a networking device such as a router or a switch.

IP Address: A unique numerical label assigned to each device connected to a computer network. IPv4 defines an IP address as a 32-bit number, but because of the growth of the Internet IPv6 was created, using 128 bits for the IP address.

Classless Inter-Domain Routing (CIDR): A method for allocating IP addresses and IP routing. CIDR notation is a compact representation of an IP address and its associated routing prefix. The notation is constructed from an IP address, a slash ('/') character, and an integer. The integer is the count of leading 1 bits in the subnet mask. Larger values here indicate smaller networks. The maximum size of the network is given by the number of addresses that are possible with the remaining, least-significant bits below the prefix.

Example: The IPv4 block 192.168.100.0/22 represents the 1024 IPv4 addresses from 192.168.100.0 to 192.168.103.255.

Open Systems Interconnection (OSI) Model: A conceptual model that characterizes and standardizes the communication functions of a computing system without regard to its underlying internal structure and technology. Its goal is the interoperability of diverse communication systems with standard communication protocols. The model partitions a communication system into abstraction layers.

Internet Gateway

A scalable, redundant, and highly available VPC component that allows communication between instances in your VPC and the public internet. An internet gateway serves two purposes:

- Provide a target in your VPC route tables for internet traffic
- Perform network address translation for instances that were assigned public IPv4 addresses.
- To make a subnet public, you attach an internet gateway to your VPC and add a route entry to the route table associated with the subnet.
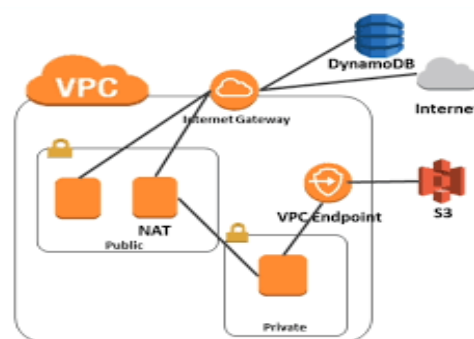

Fig 4.3: Internet Gateway

Module-6: Compute

Docker is a software platform that enables you to build, test, and deploy applications quickly. Containers are created from a template called an image.

Amazon Pricing Models

On-Demand Instances

- Pay by the hour
- No long-term commitments.
- Eligible for the AWS Free Tier.

Dedicated Instances

- Instances that run in a VPC on hardware that is dedicated to a single customer.

Spot Instances

- Instances run as long as they are available and your bid is above the Spot Instance price.
- They can be interrupted by AWS with a 2-minute notification
- Interruption options include terminated, stopped or hibernated
- Good choice when you have flexibility in when your applications can run.

Reserved Instances

- Full, partial, or no upfront payment for instance you reserve
- Discount on hourly charge for that instance
- 1-year or 3-year term

Scheduled Reserved Instances

- Purchase a capacity reservation that is always available on a recurring schedule you specify 1-year term.
- Per second billing available for On-Demand Instances, Reserved Instances, and Spot• Instances that run Amazon Linux or Ubuntu.

Module-7: Storage

Amazon Elastic Block Store (EBS)

Amazon Elastic Block Store (EBS) is an easy to use, high performance block storage service designed for use with Amazon Elastic Compute Cloud (EC2) for both throughput and transaction intensive workloads at any scale. With **block storage**, files are split into evenly sized blocks of data, each with its own address but with no additional information (metadata) to provide more context for what that block of data is. **Object storage**, by contrast, doesn't split files up into raw blocks of data. Instead, entire clumps of data are stored in, yes, an object that contains the data, metadata, and the unique identifier. With block storage you can update a single block without having to update the entire file like in object storage.

Amazon EBS enables you to create individual storage volumes and attach them to an Amazon EC2 instance:

- Offers block-level storage
- HDD and SSD available
- Volumes are automatically replicated within its Availability Zone
- It can be backed up automatically to Amazon S3 through snapshots
- Designed for resiliency - Annual Failure Rate (AFR) is between 0.1% and 1%
- Uses include:
  - Boot volumes and storage for (Amazon EC2) instances
  - Data storage with a file system
  - Database hosts
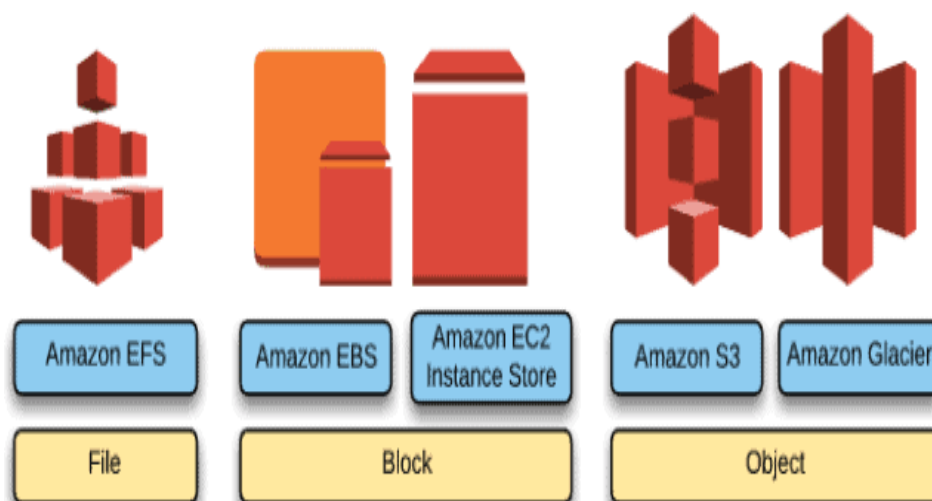 - Enterprise applications

Fig 4.3: AWS Storage Services

Module-8: Databases

Amazon Relational Database Service (RDS)

Amazon Relational Database Service (Amazon RDS) makes it easy to set up, operate, and scale a relational database in the cloud. It provides cost-efficient and resizable capacity while automating time-consuming administration tasks such as hardware provisioning, database setup, patching and backups. RDS provides you with six familiar database engines to choose from: Amazon Aurora, Oracle, Microsoft SQL Server, PostgreSQL, MySQL and MariaDB.

Unmanaged Challenges

- Server maintenance and energy footprint
- Software installation and patches
- Database backups and high availability
- Limits on scalability
- Data security
- Operating system (OS) installation and patches

RDS is a managed service that sets up and operates a relational database in the cloud. AWS Manages:

- OS installation and patches
- Database software installation and patches
- Database backups
- High availability
- Scaling
- Power and racking and stacking servers
- Server maintenance

Amazon DynamoDB

- Fast and flexible NoSQL database service for any scale.
- NoSQL database tables with no limits
- Virtually unlimited storage
- Items can have differing attributes
- Low-latency queries
- Scalable read/write throughput with no limits
- Supports document and key-value store models.
- Replicates your tables automatically across your choice of AWS Regions
- Works well for mobile, web, gaming, AdTech, and Internet of Things (IoT) applications
- Provides consistent, single-digit millisecond latency at any scale

Module-9: Cloud Architecture

AWS Well-Architecture Framework

- A guide for designing infrastructures that are, secure, high-performing, resilient, and efficient
- A consistent approach to evaluating and implementing cloud architectures
- A way to provide best practices that were developed through lessons learned by reviewing customer architectures
- There are 5 pillars to the Well-Architected Framework: Operational Excellence, Security, Reliability, Performance Efficiency, and Cost Optimization
- The AWS Well-Architected Tool helps you to implement the Well-Architected Framework

Fig 4.5: AWS Well-Architecture Framework

Reliability and Availability

Reliability

- A measure of your system's ability to provide functionality when desired by the user
- System includes all system components: hardware, firmware, and software
- Probability that your entire system will function as intended for a specified period
- Mean time between failures (MTBF) = total time in service/number of failures

Metrics

- Mean Time to Failure (MTTF)
- Mean Time to Repair (MTTR)
- Mean Time Between Failures (MTBF) = MTTF + MTTR

Availability

- Normal operation time / total time
- A percentage of uptime (for example, 99.9 percent) over time (for example, 1 year)
- Number of 9s – Five 9s means 99.999 percent availability

High Availability

- System can withstand some measure of degradation while still remaining available
- Downtime is minimized

Module-10: Automatic Scaling and Monitoring

Elastic Load Balancing

Elastic Load Balancing distributes incoming application or network traffic across multiple targets in a single Availability Zone or across multiple Availability Zones. It also scales your load balancer as traffic to your application changes over time. Monitoring is done via Amazon CloudWatch, access logs, and AWS CloudTrail logs.

Amazon CloudWatch

Monitors: AWS resources and applications that run on AWS

Collects and tracks: Standard and custom metrics

Alarms: Send notifications to an Amazon SNS topic and perform Amazon EC2 Auto Scaling or Amazon EC2 actions

- Create alarms based on:
o Static threshold
o Anomaly detection
o Metric math expression
- Specify:
o Name space
o Metric
o Statistic
o Period
o Conditions
o Additional configuration
o Actions

Amazon EC2 Auto Scaling

- Monitors your applications and automatically adjusts capacity to maintain steady, predictable performance at the lowest possible cost
- Provides a simple, powerful user interface that enables you to build scaling plans for resources
- Helps you maintain application availability
- Enables you to automatically add or remove EC2 instances according to conditions that you define
- Detects impaired EC2 instances and unhealthy applications, and replaces the instances without your intervention
- Provides several scaling options: Manual, scheduled, dynamic or on-demand, and predictive
- An Auto Scaling group is a collection of EC2 instances that are treated as a logical grouping for the purposes of automatic scaling and management.
- Scale out (launch instances), Scale in (terminate instances)

Module-11: Welcome to AWS Academy Machine Learning Foundations

It is helpful to understand the prerequisites of this course. Its preferable for attendees to have some general IT Knowledge. The foundational computer literacy skills that you need to be successful include basic computer concepts, email, file management, and a good understanding of the internet. You also should have completed the AWS academy Cloud Foundations course, intermediate skills with python programming, and general knowledge of applied statistics. General business knowledge is important, including insight into how information technology is used in business. Communication skills, leadership abilities, and a customer service orientation are also important skill sets. To achieve success in this course, you also should have:

- A general familiarity with cloud computing concepts
- A working knowledge of distributed systems
- Familiarity with general networking concepts
- A working knowledge of multi-tier architectures.

In this course, you will learn how to describe machine learning (ML), which include how to:

- Recognize how machine Learning and deep learning are part of artificial intelligence
- Describe artificial intelligence and machine leaning terminology
- Identify how machine learning can be used to solve a business problem
- Describe the machine learning process
- List the tools available to data scientists
- Identify when to use machine learning instead of traditional software development methods.

Module-12: Introduction to Machine Learning

What is machine learning?

Machine learning is a subset of AI, which is a broad branch of computer science for building machines that can do humans tasks. Deep learning itself a subdomain of machine learning of machine learning. To understand where these ideas fit together, you will learn abount each field.

Deep learning represents a significant leap forward in the capabilities for AI and ML. The theory behind deep learning was created from how the human brain works. An artificial neural network (ANN) is inspired from the biological neurons in the brain, although the implementation is different.

Artificial neurons have one or more inputs and a single output. These neurons fire (or activate their outputs), which are based on a transformation of the inputs. A neural network is composed of layers of these artificial neurons, with connections between the layers. Typically, a network has input, output, and hidden layers.

Machine Learning Process

After you formulate the problem, you move to the data preparation and preprocessing phase. In this phase, you will extract data from one or more data sources. These data sources might have differences in data or types that must be reconciled to form a single cohesive view of your data. You must visualize your data and use statistics to determine whether the data is consistent and can be used for machine learning. You will look at some data sources later in the course.

In the example data, you have four columns that contain data that was assembled from three data sources. The sources had slightly different ways of representing data, and the results are shown in the table. In ML problems, columns represents features, and rows represent instances. You can see some issues with the data in some of the instances.

Machine Learning tools Overview

Jupyter Notebook is an open-source web application that enables you to create and share documents that contain live code, equation, visualizations, and narrative text. Uses include data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.
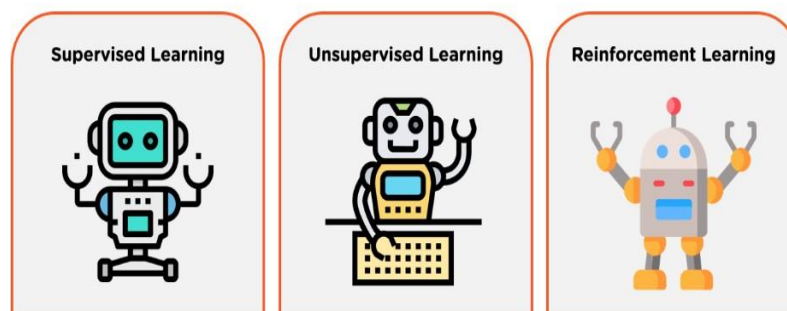

Fig 4.6: Types of Machine Learning

Module-13: Implementing a Machine Learning Pipeline with Amazon Sagemaker

The machine learning pipeline is the focus of this module. The diagram shows which sections of this module cover each stage in the pipeline

Collecting and securing data

You can obtain data from several places.

- Private data is data that you (or your customers) have in various existing systems.
- Commercial data is data that a commercial entity collected and made available
- Open-source data comprises many different open-source datasets that range from scientific information to movie reviews.

Evaluating your data

Evaluating data involves a comprehensive process of assessing its quality, accuracy, and relevance to determine its suitability for analysis or decision-making. This essential step ensures that the results and insights derived from the data are reliable and meaningful. To evaluate data effectively, several considerations must be taken into account. First, understanding the source of the data is crucial – knowing how and by whom the data was collected can shed light on potential biases or limitations. Next, data integrity needs to be checked, including identifying missing values and duplicates. This helps ensure the data is complete and accurate. Cross-referencing the data with trusted sources and verifying its consistency across various formats and time periods is vital to ensure its accuracy.

Feature Engineering

Feature engineering is the process of crafting and selecting relevant input variables, known as features, to enhance the performance of machine learning models. These features are derived from raw data and play a pivotal role in helping models understand underlying patterns and relationships. By refining features, we can extract more valuable information from the data, ultimately improving the model's ability to make accurate predictions or classifications.

Training

Training data is the foundational set of examples used to teach a machine learning model how to make predictions or classifications. It consists of input data points paired with their corresponding output labels or target values. The model learns patterns, relationships, and features from this data, allowing it to generalize and make accurate predictions on new, unseen data. During the training process, the model adjusts its internal parameters to minimize the difference between its predictions and the actual target values in the training data. The quality and representativeness of the training data directly impact the model's performance, as it forms the basis for the model's understanding of the problem and its ability to generalize to new situations.
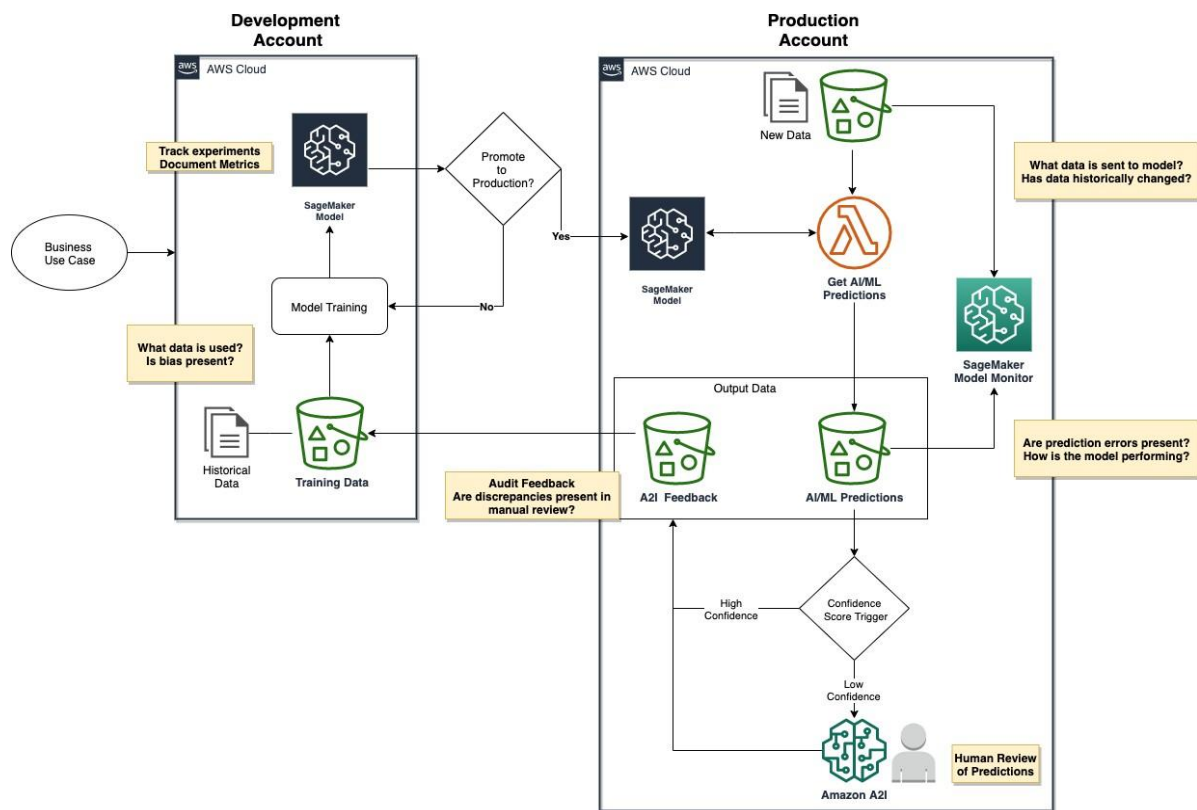
Fig 4.7: AI/ML lifecycle with Amazon A2I and SageMaker Model Monitor

Module-14: Introducing Forecasting

Forecasting Overview
Forecasting Overview:

Forecasting is like predicting the future using historical data. It's used to make educated guesses about what might happen next.

Time Series Data:

Data collected over time is often used for forecasting. For example, tracking sales each month or temperature each day.

Types of Forecasts:
- Short-term forecasts predict what will happen soon.
- Long-term forecasts try to predict events far into the future.

Processing time series data

Processing time series data involves handling and analyzing data that is collected and ordered over time. Time series data often comes from sources like sensors, stock prices, weather records, and more. The process includes several steps to understand patterns and make predictions: Initially, data is collected with timestamps, creating a chronological sequence. The first step is to clean the data by addressing missing values, outliers, and noise. This ensures accurate analysis. Exploratory analysis follows, where visualizations like line plots, histograms, and seasonal decomposition are used to identify trends, seasonality, and cyclic patterns.

Feature engineering is essential to extract relevant information from the data. This includes creating lag features, rolling averages, and other transformations that capture time-related patterns. Depending on the problem, data might need to be resampled to a lower or higher frequency for analysis convenience.

Using Amazon Forecast

Using Amazon Forecast simplifies the process of building accurate time series forecasting models. It leverages advanced machine learning techniques to make predictions based on historical data. With Amazon Forecast, users can easily prepare their data, automatically handle missing values and outliers, and select the best performing algorithms tailored to their specific data patterns. The service also takes care of model training, optimization, and hyperparameter tuning, reducing the need for manual intervention. By providing a simple API interface, Amazon Forecast enables developers to integrate the forecasting capabilities seamlessly into their applications and workflows. Whether for demand forecasting, inventory planning, or resource allocation, Amazon Forecast offers a scalable and efficient solution for generating accurate predictions, allowing businesses to make better informed decisions.

Module-15: Introducing Computer Vision (CV)

Introducing computer vision

Computer vision is an interdisciplinary field that bridges the gap between computers and visual information. It's a subset of artificial intelligence (AI) that equips machines with the ability to interpret and understand visual data, including images and videos, much like the human visual system. By enabling computers to "see" and process visual information, computer vision aims to replicate human perception and enable machines to make intelligent decisions based on what they observe.

Visual Data Interpretation:

Computer vision systems use a combination of mathematics, algorithms, and machine learning techniques to analyze and extract meaningful insights from visual data. They can identify objects, recognize patterns, measure distances, and even understand the context of a scene. These systems process pixels and convert them into actionable information, unlocking a world of possibilities across various domains.

Applications and Impact:

The applications of computer vision are diverse and far-reaching. In healthcare, it aids in diagnosing diseases from medical images; in automotive, it powers self-driving cars by recognizing road signs and obstacles; in agriculture, it assists in monitoring crop health; in retail, it enables facial recognition for personalized shopping experiences; and in security, it identifies suspicious activities from surveillance footage. The potential impact of computer vision spans industries, transforming how we interact with technology and the world around us.

Challenges and Advances:

Despite its impressive capabilities, computer vision faces challenges such as handling diverse lighting conditions, complex scenes, and occlusions. However, recent advances in deep learning, neural networks, and large datasets have propelled the field forward. Convolutional Neural Networks (CNNs) and techniques like transfer learning have revolutionized image classification, object detection, and semantic segmentation.

Future Directions:

As technology evolves, computer vision continues to push boundaries. The integration of computer vision with other AI technologies like natural language processing and robotics opens doors to even more sophisticated applications. With the growth of smart devices and the Internet of Things (IoT), computer vision is poised to transform how we interact with our surroundings and how machines interact with us, driving innovation and reshaping various industries in the process.

Module-16: Introducing Natural Language Processing

Overview of natural language processing

Natural Language Processing (NLP) is a branch of artificial intelligence that focuses on enabling computers to understand, interpret, and generate human language in a way that is both meaningful and valuable. NLP seeks to bridge the gap between human communication and computer processing, allowing machines to process, analyze, and generate text or speech in a manner that resembles human understanding. NLP encompasses a wide range of tasks and techniques, including:

- **Text Analysis**: NLP techniques help analyze and extract insights from text data. This includes tasks like sentiment analysis, text classification, and named entity recognition.

- **Language Generation**: NLP enables computers to generate human-like language, which finds applications in chatbots, text summarization, and content creation.

- **Machine Translation**: NLP plays a crucial role in translating text from one language to another, enabling communication across linguistic barriers.

- **Speech Recognition**: NLP algorithms convert spoken language into text, enabling voice assistants and transcription services.

- **Question Answering**: NLP models can process questions and provide relevant answers based on large amounts of textual information.

- **Language Understanding**: NLP allows computers to comprehend and respond to user queries, making human-computer interaction more intuitive.

- **Sentiment Analysis**: NLP can determine the sentiment or emotional tone expressed in text, helping businesses understand public opinion.
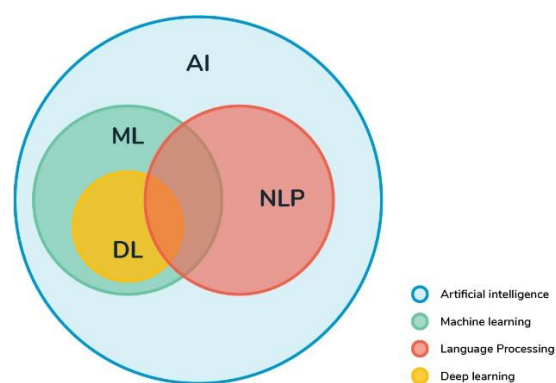

Fig 4.8: Types of Models

Module-17: Course Wrap-up

The AWS course is a comprehensive and immersive learning experience designed to provide participants with a solid foundation in Amazon Web Services (AWS) and its myriad offerings. Throughout the course, participants are introduced to the core concepts of cloud computing and the essential components that make up AWS's robust infrastructure. The journey begins with an exploration of AWS's global presence, understanding how regions, availability zones, and edge locations facilitate optimal performance and redundancy.

The course delves into a diverse array of AWS services, covering compute, storage, databases, networking, security, and more. Participants gain hands-on experience with Amazon EC2 instances, learning how to provision and manage virtual servers. They also delve into serverless computing using AWS Lambda, mastering the art of running code without worrying about server management. Storage solutions like Amazon S3 and EBS are demystified, enabling participants to effectively store and retrieve data while understanding the nuances of data durability and availability.

The course takes a deep dive into AWS databases, addressing both relational and NoSQL databases like Amazon RDS and DynamoDB. Networking concepts are elucidated, with Amazon VPC providing a platform to create isolated and customizable network environments. The importance of security is highlighted through Identity and Access Management (IAM), ensuring that resources are accessed only by authorized individuals or systems.

The course equips participants with essential skills in monitoring and management, introducing them to Amazon CloudWatch for resource tracking and AWS Trusted Advisor for optimizing their cloud environment. Scalability and high availability are explored, demonstrating how services like Auto Scaling and Elastic Load Balancing contribute to efficient and reliable cloud architectures.

Throughout the course, practical scenarios and real-world examples illustrate how AWS services can be applied across various industries. By the course's conclusion, participants have a comprehensive understanding of AWS's capabilities and a solid foundation for designing, deploying, and managing cloud solutions that drive innovation and efficiency in their respective fields.

CHAPTER 5

Real Time Examples

Amazon Web Services (AWS) is widely utilized across industries, offering real-time solutions to address various challenges and opportunities. In e-commerce, powerhouse Amazon.com relies on AWS to manage its high-volume website traffic, ensuring seamless shopping experiences even during peak periods. In the media and entertainment realm, streaming giant Netflix uses AWS for content delivery, enabling millions of users to stream movies and shows without interruption. The gaming industry benefits from AWS as well, with Epic Games utilizing its infrastructure to host online gaming sessions and handle real-time interactions in games like Fortnite.

Healthcare leverages AWS for efficient data processing and analysis, as demonstrated by GE Healthcare, which employs the platform to manage and interpret medical images. In finance, NASDAQ employs AWS to power its trading platforms, ensuring low-latency transactions and seamless stock trading experiences. IoT is another domain where AWS shines; Philips Hue's smart lighting system connects to AWS, allowing users to remotely control their lighting setups. In manufacturing, Siemens employs AWS to collect and analyse data from industrial equipment, optimizing operations and maintenance procedures.

Industries such as agriculture witness AWS's impact through companies like John Deere, which employs the platform to analyse data from farm equipment sensors, aiding farmers in data-driven decision-making. Even energy behemoth BP utilizes AWS for complex simulations in oil and gas exploration, improving reservoir analysis and strategic choices. In the travel and hospitality sector, Airbnb relies on AWS's cloud infrastructure to manage user data, bookings, and customer interactions. These examples underscore AWS's role as a transformative force, offering dynamic and adaptable solutions that cater to the diverse needs of modern businesses across various domains.

CHAPTER 6
LEARNING OUTCOMES

By the end of this we will be able to know about:

1. **Foundational Understanding of AWS**: Participants will grasp the fundamental concepts of cloud computing, gaining insight into the core components of AWS's architecture, such as regions, availability zones, and edge locations. This foundational knowledge sets the stage for deeper exploration of AWS services.

2. **Service Proficiency**: Through hands-on experience and practical exercises, participants will become proficient in using a variety of AWS services. They'll learn to provision and manage virtual servers using Amazon EC2, store and retrieve data efficiently with Amazon S3 and EBS, and implement serverless computing using AWS Lambda.

3. **Database Expertise**: Participants will gain an understanding of AWS's database offerings, both relational and NoSQL. They'll be able to configure and manage databases such as Amazon RDS and DynamoDB, empowering them to make informed decisions about choosing the right database solution for various applications.

4. **Networking and Security Mastery**: The course empowers participants to create isolated network environments using Amazon VPC and manage access to resources through IAM. This networking and security proficiency ensures data protection and optimal network configurations.

5. **Efficient Monitoring and Management**: Participants will learn to monitor and manage their AWS resources effectively using tools like Amazon CloudWatch. They'll also understand how to optimize their cloud environment with AWS Trusted Advisor's actionable insights.

6. **Scalability and High Availability**: Participants will grasp the concepts of scalability and high availability and learn how to design and implement architectures that ensure the reliability and flexibility of applications using services like Auto Scaling and Elastic Load Balancing.

7. **Real-World Applications**: By working through practical scenarios and real-world examples, participants will be prepared to apply their AWS skills across various industries. They'll be able to identify opportunities to leverage AWS services to drive innovation and efficiency within their organizations.

CONCLUSION

In conclusion, the AWS course provides a comprehensive and invaluable journey into the realm of Amazon Web Services, equipping participants with the essential knowledge and practical skills required to harness the power of cloud computing. Throughout the course, participants have gained insights into cloud concepts, explored a wide array of AWS services, and acquired the proficiency to design, deploy, and manage robust cloud architectures. As they emerge from this course, participants are not only equipped to address real-world challenges with AWS solutions but also prepared to pursue further certifications and specialization paths within the dynamic AWS ecosystem. With the ability to leverage AWS's scalability, security, and innovation, course graduates are well-positioned to contribute to their organizations' success, drive digital transformation, and seize the boundless opportunities presented by cloud computing in the modern technological landscape.

CERTIFICATE

N·E·A·T
प्रौद्योगिकी के लिए राष्ट्रीय शैक्षणिक सहयोग
National Educational Alliance for Technology

AICTE

अखिल भारतीय तकनीकी शिक्षा परिषद्
All India Council for Technical Education

EduSkills®
Nation Building Through Skills

AICTE - EduSkills
2023
VIRTUAL INTERNSHIP

Certificate of Virtual Internship

This is to certify that

USHODAYA GUNDA

Srinivasa Ramanujan Institute of Technology

has successfully completed 10 weeks

AI-ML Virtual Internship

during May - July 2023

Supported By aws academy

Shri Buddha Chandrasekhar
Chief Coordinating Officer (CCO)
NEAT Cell, AICTE

Dr. Satya Ranjan Biswal
Chief Technology Officer (CTO)
EduSkills

Certificate ID :923bd8f1cd5897f0a7cc2922036c4265
Student ID :STU620f971b430f81645188891

REFERENCES

[1] http://awsacademy.instructor.com

[2] https://internship.aicte-india.org