

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

P.Likhitha Sri

204G1A3227



**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 **P. LIKHITHA SRI** bearing Roll Number **204G1A3227** 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: Anantapuramu

EXTERNAL EXAMINER

PREFACE

I had completed 10 weeks AI-ML Virtual Internship during May 2023 to July 2023 supported by AWS Academy, a Cloud Service Provider. The industry offers more than 200 services to the customers on that services top 10 services are Amazon EC2, Amazon RDS, Amazon S3, Amazon Aurora, Amazon DynamoDB, Amazon Lambda, Amazon VPC, Amazon CloudFront, AWS Elastic Beanstalk, Amazon EC2 Auto-Scaling. Edu skills is the modern and an innovative training institute provides high quality Computer Education to I.T aspirants from all sectors of society since 2012. Eduskills is an autonomous institution registered under the Govt. of Assam. Eduskills is member of ALMA International, ITRC, NISD, ATTEST. This internship project is a part of IV –Year B. Tech program which was conducted at Srinivasa Ramanujan Institute of Technology – Ananthapuramu.

AICTE has prepared a model curriculum with the help of prominent academicians of the country so that the country may produce competent employable graduates as per the needs of the industry. By doing this Internship, I gained knowledge on AWS foundation and AWS architecture modules to build my career in AWS Cloud based platform and Edu Skills also provided some different company placement drives to get employment opportunities in cloud-based companies.

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 **Dr. G. Hemanth Kumar Yadav, M.Tech,Ph.D, Associate Professor & Head, Department of 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/her for the stimulating support, constant encouragement and constructive criticism which have made possible to bring out this internship work.

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

I wish to convey my special thanks to **Dr. G. Balakrishna, 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 project 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.

Likhitha Sri P
(204G1A3227)

CONTENTS	Page No.
List of Figures	vii
List of Abbreviations	viii
Chapter 1 Introduction	1
Chapter 2 Technology	3
Chapter 3 Applications	5
Chapter 4 Modules Explanation	7
Chapter 5 Real time example	24
Chapter 6 Learning outcomes	25
Conclusion	26
Internship Certificate	27
Reference	28

LIST OF FIGURES

Fig No.	Description	Page No.
4.1	Types of cloud service models	7
4.2	AWS Billing Dashboard	8
4.3	AWS Global Infrastructure	9
4.4	AWS shared responsibility model	9
4.5	Content delivery	10
4.6	AWS Compute Services	11
4.7	AWS Storage Services	12
4.8	Amazon RDS	12
4.9	DynamoDB	13
4.10	Amazon Redshift	13
4.11	Amazon Aurora	13
4.12	Six Pillars of AWS	14
4.13	Amazon CloudWatch	15
4.14	Amazon ELB	15
4.15	Types of Machine Learning	17
4.16	Workflow of Amazon Pipeline	18
4.17	Overview of Amazon Forecast	19
4.18	NLP Use cases	22

LIST OF ABBREVIATIONS

AWS	Amazon Web Services
VPC	Virtual Private Cloud
RDS	Relational Database Service
ELB	Elastic Load Balancing
IaaS	Infrastructure as a Service
PaaS	Platform as a Service
SaaS	Software as a Service
JSON	JavaScript Object Notation
IAM	Identity and Access Management
EC2	Elastic Compute Cloud
EBS	Elastic Block Storage
S3	Simple Storage Service
NLP	Natural Language Processing
ECS	Elastic Container Service

CHAPTER 1

INTRODUCTION

In 2006, Amazon Web Services (AWS) began offering IT infrastructure services to businesses as web services—now commonly known as cloud computing. One of the key benefits of cloud computing is the opportunity to replace upfront capital infrastructure expenses with low variable costs that scale with your business. With the cloud, businesses no longer need to plan for and procure servers and other IT infrastructure weeks or months in advance. Instead, they can instantly spin up hundreds or thousands of servers in minutes and deliver results faster.

Today, AWS provides a highly reliable, scalable, low-cost infrastructure platform in the cloud that powers hundreds of thousands of businesses in 190 countries around the world. In this internship project we have completed some important modules related to AWS services.

1.1 Courses

These are the courses which we learned during the self-paced learning period.

- AWS Academy Cloud Foundations
- AWS Academy Machine Learning Foundations

1.1.1 AWS Academy Cloud Foundations

In this course we came across these modules

- Cloud Concepts Overview.
- Cloud Economics and Billing.
- AWS Global Infrastructure Overview
- AWS Cloud Security
- Networking and Content Delivery.
- Compute
- Storage
- Databases

- Cloud Architecture
- Auto Scaling and Monitoring

1.1.2 AWS Academy Machine Learning Foundations

In this course we came across these modules

- Welcome to AWS Academy Machine Learning Foundations.
- Introducing Machine Learning.
- Implementing a Machine Learning pipeline with Amazon Sage Maker.
- Introducing Forecasting.
- Introducing Computer Vision (CV).
- Introducing Natural Language Processing.

CHAPTER 2

TECHNOLOGY

Technologies required for AWS

The purpose of model explainability is to create an understandable solution which can communicate results of AI/ML technology. This field has been expressed as explainable artificial intelligence. Because AI/ML methods have increased in complexity to satisfy industry needs, the requirement for model explain ability has risen. When AI/ML solutions are launched into production within customer AWS environments, business leaders or AI/ML owners must trust non-human results that can directly impact business goals. Here are six skills needed to rock this in demand role.

1. 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.

2. 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.

3. 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.

4. 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

5. 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.

6. Communication

Designing great architecture doesn't mean a thing if you fail to explain your vision to software developers, managers and fellow architects. You will need to learn to explain your ideas via emails, documents and presentations in a way that convinces your team why your solution is the best. Work on your presentation skills, learn how to write concisely and use a diagramming tool to show complex environments. A single slide with a diagram is preferable to ten slides with twenty bullet points each.

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. 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.

5. Gaming:

AWS has been serving many gaming studios. Combining Amazon EC2 and S3 services with CloudFront 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 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.

CHAPTER 4

MODULES EXPLANATION

4.1 Course 1: AWS Academy Cloud Foundations

Module 1: Cloud Concepts Overview

This module deals with the 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.

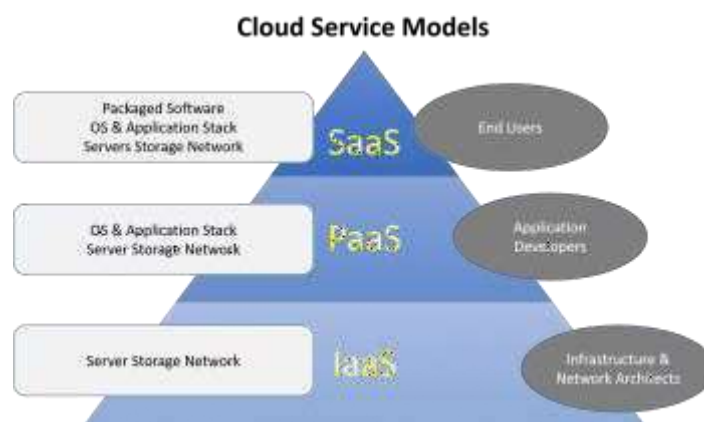


Fig. No. 4.1 Types of cloud service models

AWS is a secure cloud platform that offers a broad set of global cloud-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

This module describes about the three fundamental drivers 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 actually need. This utility-style pricing model includes: Pay for what you use, pay less when you reserve, Payless 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 AWS implementation.

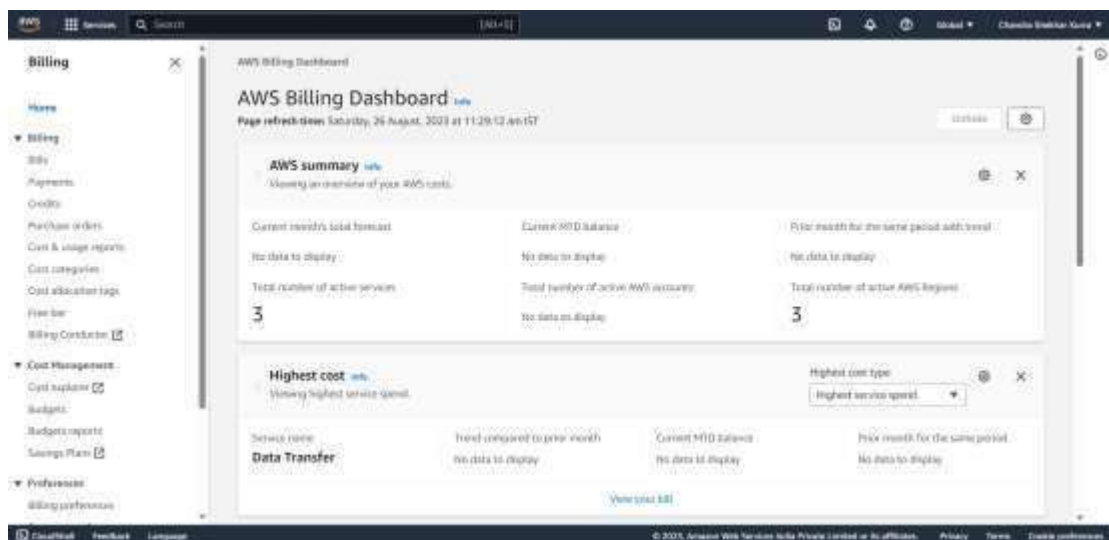


Fig. No. 4.2 AWS Billing Dashboard

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. 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.



Fig. No. 4.3 AWS Global Infrastructure

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.



Fig. No. 4.4 AWS shared responsibility model

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 run on compute 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. Amazon Virtual Private Cloud (VPC) is a service that lets you provision a logically isolated section of the AWS Cloud where you can launch your AWS resources. 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 reattached 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.

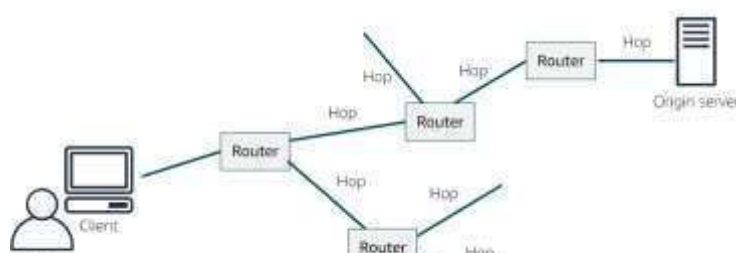


Fig. No. 4.5 Content delivery

Module 6: Compute

This module explains about the list of available of AWS compute services. Here are some major and important computing services and their summaries.



Fig. No. 4.6 AWS compute services

Amazon EC2: It 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 on-premises computing, 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: It is a zero-administration compute platform. AWS Lambda enables you to run code without provisioning or managing servers. You pay only for the compute time that is consumed. This serverless 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: It 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.

Module 7: Storage

This module explores about the list of available storage services in the AWS. Some of them are listed below.



Fig. No. 4.7 AWS storage services

Amazon EBS provides persistent block storage volumes for use with Amazon EC2 instances. 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.

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 re-upload the entire modified file. Amazon S3 stores data as objects within resources that are called buckets

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.

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. AmazonRDS is scalable for compute and storage, and automated redundancy and backup is available. Supported database engines include Amazon Aurora, PostgreSQL, MySQL, MariaDB, Oracle, and Microsoft SQL Server.



Fig. No. 4.8 Amazon RDS

DynamoDB is a fast and flexible NoSQL database service for all applications that need consistent, single-digit-millisecond latency at any scale. Amazon manages all the underlying data infrastructure for this service and redundantly stores data across multiple facilities in a native US Region as part of the fault-tolerant architecture. With DynamoDB, you can create tables and items. You can add items to a table. The system automatically partitions your data and has table storage to meet workload requirements



Fig. No. 4.9 DynamoDB

Amazon Redshift is a fast, fully managed data warehouse that makes it simple and cost-effective to analyse 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. No. 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. No. 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. No. 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 is operating 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. No. 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.

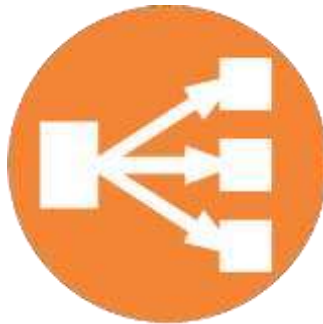


Fig. No. 4.14 Amazon ELB

4.2 Course 2: AWS Academy Machine Learning Foundations

Module 1: Welcome to AWS Academy Machine Learning Foundations

This module 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:

- Recognise 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 2: Introduction to 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 about 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 has three main types. The first type is supervised learning, where a model uses known inputs and outputs to generalize future outputs. The second type is unsupervised learning, where the model doesn't know inputs or outputs it finds patterns in the data without help. The third type is reinforcement learning, where the model interacts with its environment and learns to take actions that maximize rewards. It's important to know the different ML types because the type can guide you toward selecting algorithms that make sense for solving our problem.

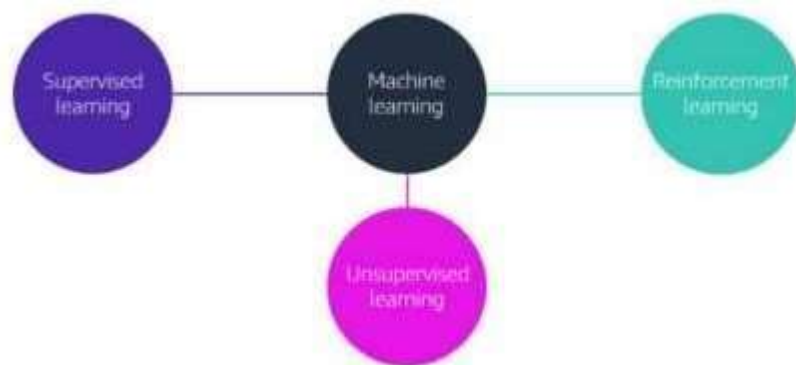


Fig. No. 4.15 Types of Machine Learning

Module 3: Implementing a Machine Learning Pipeline with Amazon Sagemaker

This module introduces and describes a typical process for handling a machine learning problem. A machine learning pipeline can be applied to many machine learning problems. This module focuses on supervised learning, but the process that you learn in this module can be adapted to other types of machine learning.

The diagram shows which sections of this module cover each stage in the pipeline

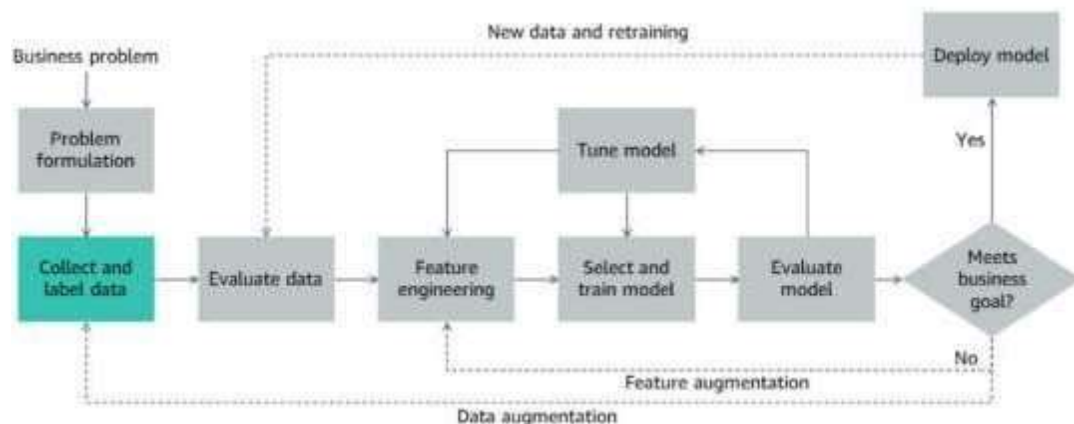


Fig. No. 4.16 Workflow of Amazon pipeline

1. 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.

2. 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.

3. 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.

4. 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.

Module-14: Introducing Forecasting

Forecasting is an important area of machine learning. It is important because so many opportunities for predicting future outcomes are based on historical data. Many of these opportunities involve a time component. Although the time component adds more information, it also makes time series problems more difficult to handle than other types of predictions.

Time series data is captured in chronological sequence over a defined period of time. Introducing time into a machine learning model has a positive impact because the model can derive meaning from change in the data points over time. Time series data tends to be correlated, which means that a dependency exists between data points.

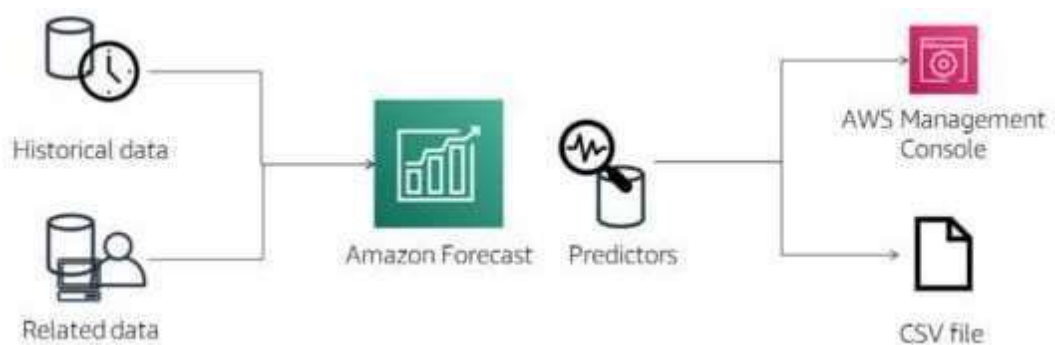


Fig. No. 4.17 Overview of amazon forecast

The overall process for working with Amazon Forecast is to import historical and related data. Amazon Forecast inspects the data, identifies key data, and selects an appropriate algorithm. It uses the algorithm to train and optimize a custom model and produce a predictor. You create forecasts by applying the predictor to your dataset. Then, you can either retrieve these forecasts in the AWS console, or export the forecasts as comma-delimited files. You can also use an application programming interface (API) and command line interface (CLI) commands to create and retrieve forecasts.

Amazon Forecast predictors use an algorithm to train a model. They then use the model to make a forecast by using an input dataset group. To help you get started, Amazon Forecast provides the following predefined algorithms:

5. ARIMA
6. DeepAR
7. ETS
8. NPTS
9. Prophet

You can also use the AutoML feature, which tries all the algorithms to see which one is the best at predicting data.

Module-15: Introducing Computer Vision (CV)

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 analyse 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 6: Introducing 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, analyse, 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 analyse 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. No. 4.18 NLP use Cases

Module 7: Course Wrap-up

In this course, you learned how to describe machine learning. This included learning how to:

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

You also learned how to implement a machine learning pipeline. This included how to:

- Formulate a problem from a business request.
- Obtain and secure data for machine learning.
- Build a Jupyter notebook by using Amazon SageMaker.
- Outline the process for evaluating data.
- Explain why data needs to be pre-processed.
- Use open-source tools to examine and preprocess data.

Finally, you learned how to use managed Amazon ML services for forecasting, computer vision, and natural language processing. You are now able to:

- Describe the business problems that Amazon Forecast solves
- Describe the challenges of working with time series data
- List the steps that are required to forecast by using Amazon Forecast
- Describe the computer vision use cases.
- Describe the managed Amazon ML services for image and video analysis.
- Use Amazon Rekognition to perform facial detection.
- Describe the natural language processing (NLP) use cases that are solved by using managed Amazon ML services.
- Describe the managed Amazon ML services available for NLP.

CHAPTER 5

REAL TIME EXAMPLE

Formula 1 uses AWS:

Formula 1, the pinnacle of motorsport, has harnessed the capabilities of Amazon Web Services (AWS) to revolutionize the way teams approach racing. Through a strategic partnership with AWS, Formula 1 has unlocked a new era of data-driven excellence. Teams now leverage AWS's cutting-edge cloud computing and data analytics technologies to gain a competitive edge on and off the track. The partnership has transformed the sport into a realm where real-time insights and predictive analytics guide split-second decisions, optimizing everything from race strategies to car performance.

The collaboration between Formula 1 and AWS showcases the marriage of technology and athleticism. With an array of sensors capturing intricate details from race cars and circuits, AWS's computational prowess analyzes this torrent of data, equipping teams with predictive models that inform pit-stop decisions, tire choices, and car adjustments. This dynamic synthesis also extends to fan engagement, with AWS's visualizations enhancing broadcasts with captivating statistics and real-time race analysis. Ultimately, the Formula 1-AWS alliance underscores how innovation in cloud computing and data analytics can push the boundaries of sports, creating a fusion where data becomes a decisive element in the quest for victory.

CHAPTER 6

LEARNING OUTCOMES

- Gain an overall understanding of basic cloud computing concepts
- Become familiar with AWS core services and tools.
- Learn the architectural principles of the AWS Cloud.
- Understand and be able to explain AWS Cloud security and compliance measures.
- Understand the AWS budget and pricing philosophy.
- Engage in hands-on practice to hone key skills.
- Learn the knowledge and skills required to take the AWS Certified Cloud Practitioner Exam.

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, I 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. At the end of this course, 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.

INTERNSHIP CERTIFICATE



REFERENCES

- [1] <https://awsacademy.instructure.com/courses/46904>
- [2] <https://awsacademy.instructure.com/courses/47864>