Data Analytics

(40h / 5 Days)

Description:

This Data Analytics training program is designed to develop practical skills needed to perform complex data analysis using AWS services. The course begins with an introduction to core data analytics concepts and **AWS** tools like Redshift, Athena, and QuickSight, providing participants with a strong foundation in cloud-based data analysis. Trainees will learn how to manage data warehouses, create ETL pipelines, and efficiently query data, applying advanced SQL techniques and optimising performance for large datasets.

As participants progress, they will gain hands-on experience in data visualisation, using Amazon QuickSight to build interactive dashboards that communicate data-driven insights effectively. Additionally, the course introduces machine learning for data analytics, guiding participants through predictive analytics using **SageMaker**, and real-time data analytics with Kinesis. The capstone project consolidates all the learned skills,

challenging participants to design and implement a complete data analytics pipeline from data ingestion to insights visualisation.

Target Audience:

The target audience for the Data Analytics Training Program includes:

- **Data Analysts and Scientists**: Professionals who analyse large datasets to generate insights and are looking to enhance their skills with AWS tools.
- **Business Intelligence Professionals**: Individuals who work with data to inform strategic decision-making and want to leverage AWS for visualisation and data processing.
- **IT Professionals**: System administrators, software developers, and others in the IT field who need to manage or interact with data analytics platforms.
- **Aspiring Data Professionals**: Students or career changers interested in entering the data science or analytics fields.
- **Project Managers and Team Leads**: Managers overseeing data-driven projects who require a better understanding of data analytics processes and tools to improve project outcomes.
- **Marketing Analysts**: Marketing professionals who use data to track performance, predict trends, and enhance marketing strategies.
- **Finance Professionals**: Analysts in finance sectors looking to improve financial modelling, risk assessment, and customer data analysis with advanced analytics tools.

Training Expected Outcomes:

By the end of the training, participants will be able to:

1. **Conduct Comprehensive Data Analysis**: Effectively analyse data using AWS services, applying both basic and advanced analytical techniques.
2. **Utilise AWS Redshift and Athena**: Proficiently use these tools for data querying, management, and analytics.
3. **Create Data Visualisations**: Build insightful visualisations and dashboards using Amazon QuickSight.
4. **Implement Data Analytics Projects**: Successfully design and execute data analytics projects from start to finish.
5. **Solve Real-World Problems**: Apply data analytics skills to solve industry-specific problems, demonstrating readiness to tackle real business challenges.

Training Strategy

Module-Based Learning: Each module covers a specific aspect of data analytics, providing a comprehensive and structured learning experience.

Theoretical Lessons: These lessons offer foundational knowledge and theoretical concepts necessary for understanding data analytics principles and best practices.

Hands-On Labs: Practical application of concepts through labs using AWS services, enhancing learning through real-world simulations and tasks.

Assessments: Regular quizzes and a final project assess understanding and proficiency in applying data analytics techniques.

Real-World Case Studies: Insights into practical applications and challenges in data analytics through case studies, enhancing learning with real industry examples.

Course Modules

1. **Introduction to Data Analytics and AWS** – Learn data analytics fundamentals and AWS services for analytics.
2. **Data Warehousing with Amazon Redshift** – Design and manage data warehouses using Amazon Redshift.
3. **Data Ingestion and ETL Processes** – Set up and manage ETL pipelines with AWS Glue for data transformation.
4. **Advanced SQL Querying with Amazon Athena** – Perform complex SQL queries on large datasets using Athena.
5. **Data Visualisation with Amazon QuickSight** – Create insightful visualisations and dashboards using QuickSight.
6. **Machine Learning for Data Analytics** – Apply machine learning techniques to data analytics using AWS SageMaker.
7. **Real-Time Data Analytics with Amazon Kinesis** – Implement real-time data analytics pipelines using AWS Kinesis.
8. **Big Data Analytics with Amazon EMR** – Process big data using Hadoop and Spark on Amazon EMR.
9. **Advanced Analytical Techniques** – Explore advanced techniques like time-series forecasting and A/B testing.
10. **Capstone Project and Review** – Create a complete data analytics solution, from data ingestion to visualisation.

Training Program

Data Analytics

Training Objectives:

- Understand Data Analytics Fundamentals
- Utilise AWS for Data Analytics
- Perform Data Analysis and Visualization
- Implement Data Warehousing Solutions
- Apply Advanced Analytical Techniques
- Handle Real-World Data Analytics Projects

Time	*Modules*

5 Hours	**Module 1: Introduction to Data Analytics and AWS**
	Objective: Introduce trainees to the fundamental concepts of data analytics and familiarise them with the AWS platform and its key data analytics services.
	--- 	
	• Overview of Data Analytics
	  ◦ Key concepts and value of data analytics.
	  ◦ Differences between descriptive, predictive, and prescriptive analytics.
	• Data Analytics on AWS
	  ◦ Introduction to AWS analytics services such as Amazon Redshift, Amazon Athena, AWS Glue, and Amazon QuickSight.
	• Setting Up the Environment
	  ◦ How to set up an AWS account and navigate the AWS Management Console.
	  ◦ Initial configuration of AWS services for data analytics tasks.
	1h
	2h
	2h
	Lab: Setting up and configuring AWS environments specifically for data analytics tasks.
	Assessment: Quiz to test knowledge of basic data analytics

4 Hours	**Module 2: Data Warehousing with Amazon Redshift**
	Objective: Enable trainees to design, implement, and manage data warehouses using Amazon Redshift, focusing on architecture, performance optimization, and best practices.
	--- 	
	2h
	- Introduction to Data Warehousing
	  - Understanding data warehousing concepts, benefits, and architecture.
	  - Comparison of data lakes and data warehouses.
	2h
	- Using Amazon Redshift
	  - Deep dive into Amazon Redshift features, architecture, and its use in data warehousing.
	  - Best practices for data modelling and performance optimization.
	
	Lab: Creating and populating a data warehouse in Amazon Redshift, including data loading and query optimization.
	
	Assessment: Quiz on data warehousing principles and hands-on exercises in Redshift usage.

3 Hours	**Module 3: Data Ingestion and ETL Processes**
	Objective: Enable trainees to efficiently ingest and transform data using AWS services, emphasising the setup and management of ETL workflows with AWS Glue.
	--- 	
	1.5h
	- Data Ingestion Techniques
	  - Overview of data ingestion methods and tools available on AWS.
	  - Choosing the right technique based on data velocity and variety.
	1.5h
	- ETL with AWS Glue
	  - Introduction to AWS Glue, setting up crawlers, and creating ETL jobs.
	  - Managing and automating ETL workflows.
	
	Lab: Developing ETL pipelines using AWS Glue to process and transform data.
	
	Assessment: Project focused on implementing a complete ETL workflow from data ingestion to storage.

4 Hours	**Module 4: Advanced SQL Querying with Amazon Athena**
	Objective: Develop trainees abilities to perform complex SQL queries using Amazon Athena, focusing on optimising queries to handle large datasets effectively.
	--- 	
	2h
	- SQL Basics for Data Analysis
	  - Refreshing advanced SQL skills critical for data querying and analysis.
	2h
	- Using Amazon Athena
	  - Leveraging Athena for serverless SQL queries directly on data stored in Amazon S3.
	  - Best practices for schema design and query performance.
	Lab: Using Amazon Athena to perform complex SQL queries on large datasets.
	Assessment: Series of SQL challenges to solve real-world data querying problems using Athena.

4 Hours

Module 5: Data Visualization with Amazon QuickSight

Objective: Enable trainees to create insightful data visualisations and interactive dashboards using Amazon QuickSight, enhancing their ability to communicate data-driven insights.

2h

- **Basics of Data Visualization**
  - Principles of effective data visualisation.
  - Choosing the right chart types for various data analysis needs.

2h

- **Creating Dashboards with Amazon QuickSight**
  - Comprehensive guide to Amazon QuickSight features, creating dashboards, and sharing insights.
  - Advanced techniques like calculated fields, embedding, and dashboard interactivity.

Lab: Designing and building interactive dashboards using Amazon QuickSight to visualise data insights.

Assessment: Participants create a dashboard for a given dataset and present their insights

4 Hours	**Module 6: Machine Learning for Data Analytics**
	Objective: Introduce machine learning concepts and techniques within the context of data analytics, teaching participants to apply these using Amazon SageMaker for predictive analytics.
	--- 	
	1h
	- Introduction to Machine Learning in Analytics
	  - How machine learning can be applied to enhance data analytics.
	  - Overview of machine learning processes: from data preparation to model evaluation.
	3h
	- Using Amazon SageMaker
	  - Introduction to Amazon SageMaker capabilities for building, training, and deploying machine learning models.
	  - Exploring built-in algorithms and Jupyter notebook integration.
	
	Lab: Building basic predictive models using Amazon SageMaker, focusing on integration with AWS data services.
	
	Assessment: Applying predictive analytics on a provided dataset

4 Hours
 	
1h
 	
3h
	Module 7: Real-Time Data Analytics with Amazon Kinesis
	
	Objective: Provide trainees with the knowledge to build and manage real-time data analytics pipelines using Amazon Kinesis, focusing on stream processing and real-time data decision-making.
	
	--- 	
	
	- Introduction to Real-Time Data Analytics
	  - Importance and applications of real-time data analytics.
	  - Understanding the architecture and components of Amazon Kinesis.
	- Using Amazon Kinesis for Data Streams
	  - Setting up Kinesis Data Streams for real-time data ingestion.
	  - Processing data in real-time using Kinesis Data Analytics.
	
	Lab: Configuring Amazon Kinesis to collect, process, and analyze real-time data streams.
	
	Assessment: Creating a real-time dashboard that reflects streaming data analytics.

4 Hours	**Module 8: Big Data Analytics with Amazon EMR**
	Objective: Instruct trainees on utilising Amazon EMR for big data processing, covering setup, management, and application of big data technologies like Hadoop and Spark.
	--- 	
	2h
	- Introduction to Big Data on AWS
	  - Overview of big data concepts and the role of Amazon EMR in handling big data.
	2h
	- Using Amazon EMR
	  - Configuring and launching an EMR cluster.
	  - Running big data processing jobs using popular frameworks like Hadoop and Spark.
	
	Lab: Executing a big data processing task using Spark on EMR to derive insights from large datasets.
	
	Assessment: Analysis of job performance and optimization of the EMR environment.

4 Hours
 	
1.5h
 	
2.5h
	Module 9: Advanced Analytical Techniques (4 hours)
	Objective: Expand trainees expertise in advanced analytics techniques such as time-series forecasting, segmentation analysis, and A/B testing, using appropriate AWS tools.
	--- 	
	• Advanced Analytics
	  ◦ Techniques such as time series forecasting, segmentation analysis, and A/B testing.
	• Integrative Analytics with AWS Services
	  ◦ Combining AWS services like AWS Lambda and Amazon SageMaker to enhance analytics workflows.
	
	Lab: Implementing a complex analytical solution that uses multiple AWS services.
	
	Assessment: Comprehensive case study requiring the application of multiple advanced analytical techniques.

4 Hours	**Module 10: Capstone Project and Review**
	Objective: Consolidate all learned skills through a comprehensive capstone project that involves creating a complete data analytics solution from data ingestion to insights presentation, culminating in a review and feedback session to ensure mastery of the material.
	--- 	
	• Capstone Project
	  ◦ Participants undertake a comprehensive project that involves setting up a complete data analytics pipeline using AWS.
	• Review and Feedback
	  ◦ In-depth review of project work with feedback from instructors.
	  ◦ Sharing projects with peer reviews to gain insights and alternative approaches.
	
	Assessment:
	
	● Final presentation of the capstone project to demonstrate the skills and knowledge acquired throughout the training.