Training Course on Cloud Data Platforms for Data Scientists (Unified Course)

Professional Development Training Course Outline

Category	Data Science	**Duration**	10 Days
Base Fee	$3,000 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

This comprehensive training course empowers data scientists with the critical skills to effectively leverage modern **cloud data platforms** for advanced **data analytics**, **machine learning**, and **business intelligence**. Participants will gain hands-on expertise in navigating, querying, and optimizing data workloads across the leading cloud data warehouses: **AWS Redshift**, **Azure Synapse Analytics**, and **GCP BigQuery**. This unified approach addresses the growing need for multi-cloud proficiency, enabling data professionals to build robust, scalable, and cost-efficient data solutions in today's dynamic **big data ecosystem**.

Training Course on Cloud Data Platforms for Data Scientists (Unified Course) is meticulously designed to bridge the gap between theoretical knowledge and practical application, focusing on real-world **case studies** and **hands-on labs**. Data scientists will learn to integrate diverse data sources, perform complex transformations, and build predictive models directly within these powerful cloud environments. Mastering these platforms is crucial for driving data-driven innovation, enhancing **data governance**, and achieving significant **performance optimization** in various industry verticals.

Programme Curriculum

Training Course on Cloud Data Platforms for Data Scientists (Unified Course)

Introduction

This comprehensive training course empowers data scientists with the critical skills to effectively leverage modern **cloud data platforms** for advanced **data analytics**, **machine learning**, and **business intelligence**. Participants will gain hands-on expertise in navigating, querying, and optimizing data workloads across the leading cloud data warehouses: **AWS Redshift**, **Azure Synapse Analytics**, and **GCP BigQuery**. This unified approach addresses the growing need for

multi-cloud proficiency, enabling data professionals to build robust, scalable, and cost-efficient data solutions in today's dynamic **big data ecosystem**.

Training Course on Cloud Data Platforms for Data Scientists (Unified Course) is meticulously designed to bridge the gap between theoretical knowledge and practical application, focusing on real-world **case studies** and **hands-on labs**. Data scientists will learn to integrate diverse data sources, perform complex transformations, and build predictive models directly within these powerful cloud environments. Mastering these platforms is crucial for driving data-driven innovation, enhancing **data governance**, and achieving significant **performance optimization** in various industry verticals.

Course Duration

10 days

Course Objectives

Upon completion of this training, participants will be able to:

1. Understand the core concepts and architectures of cloud-native data warehouses.
2. Proficiently manage, query, and optimize data within Amazon Redshift.
3. Leverage Azure Synapse for unified data integration, enterprise data warehousing, and big data analytics.
4. Effectively perform scalable data analysis and machine learning with Google Cloud BigQuery.
5. Implement strategies for seamless data movement and integration across AWS, Azure, and GCP.
6. Apply advanced techniques for query optimization and cost management across all three platforms.
7. Understand best practices for data security, access control, and compliance in cloud environments.
8. Connect cloud data platforms with popular data science tools like Python, R, and Jupyter notebooks.
9. Design and implement efficient Extract, Transform, Load (ETL) or Extract, Load, Transform (ELT) workflows.
10. Utilize in-database machine learning capabilities and platform integrations for predictive analytics.
11. Write complex SQL queries for sophisticated data analysis and aggregation.
12. Diagnose and resolve performance bottlenecks and data pipeline issues in cloud data platforms.
13. Design and deploy scalable, resilient, and cost-effective data architectures for data science workloads.

Organizational Benefits

- Equip teams with the capabilities to extract faster, deeper insights from massive datasets.
- Streamline data processing, reduce infrastructure overhead, and automate data workflows.
- Reduce cloud expenditure through optimized resource utilization and efficient query execution.
- Enable the organization to handle growing data volumes and complex analytical workloads with ease.

- Foster adaptability and resilience by empowering teams to operate across diverse cloud ecosystems.
- Implement robust data security measures and ensure compliance with regulatory standards.
- Empower data scientists to rapidly experiment with new models and analytical approaches.
- Stay ahead in the market by leveraging cutting-edge cloud technologies for superior data capabilities.

Target Audience

1. Data Scientists
2. Data Analysts
3. Machine Learning Engineers
4. BI Developers
5. Data Engineers.
6. Solution Architects.
7. Database Administrators
8. IT Professionals

Course Outline

Module 1: Introduction to Cloud Data Platforms for Data Science

- Understanding the evolution of data platforms: On-premise to Cloud.
- Key characteristics and benefits of cloud data warehousing for data scientists.
- Overview of AWS, Azure, and GCP data ecosystem landscapes.
- Challenges and opportunities in a multi-cloud data strategy.
- **Case Study:** Analyzing how a retail company migrated from on-premise data marts to a hybrid cloud data platform for real-time analytics.

Module 2: Core Concepts of Cloud Data Warehousing

- Differentiating between Data Warehouses, Data Lakes, and Data Lakehouses.
- Columnar vs. Row-Oriented Storage and their impact on analytical queries.
- Massively Parallel Processing (MPP) architectures in cloud data warehouses.
- Data Partitioning, Clustering, and Indexing for performance.
- **Case Study:** Examining a financial institution's decision to implement a data lakehouse architecture for both structured and unstructured data analysis.

Module 3: AWS Redshift Fundamentals for Data Scientists

- Redshift Architecture: Clusters, Nodes, Slices, and Leader Node.
- Data Loading strategies: COPY command from S3, DynamoDB, EMR.
- Querying data with SQL and understanding DISTKEY and SORTKEY.
- Introduction to Redshift Spectrum for querying data in S3.
- **Case Study:** Optimizing a marketing campaign's analytics by migrating large customer datasets to Redshift and leveraging Redshift Spectrum for ad-hoc queries on raw logs.

Module 4: Advanced AWS Redshift for Data Scientists

- Workload Management (WLM) and Query Queues for performance tuning.
- Materialized Views and their application in accelerating complex queries.
- Vacuuming and Analyzing tables for optimal performance.
- Security features: IAM integration, VPC, encryption.

- **Case Study:** Improving report generation time by 70% for a logistics company using Redshift Materialized Views and fine-tuned WLM.

Module 5: Azure Synapse Analytics Overview

- Unified Analytics Platform: Dedicated SQL Pools, Serverless SQL Pools, Spark Pools.
- Data Ingestion using Azure Data Factory and PolyBase.
- Exploring data with Serverless SQL Pools over Azure Data Lake Storage Gen2.
- Introduction to Synapse Studio for development and monitoring.
- **Case Study:** A healthcare provider building a unified patient data platform on Azure Synapse, integrating diverse data sources using Data Factory.

Module 6: Deep Dive into Azure Synapse Analytics for Data Scientists

- Optimizing Dedicated SQL Pool performance: Distribution, Indexing, Partitioning.
- Using Apache Spark Pools for big data processing and machine learning.
- Integration with Azure Machine Learning and Azure Databricks.
- Data Security and Compliance in Azure Synapse.
- **Case Study:** Developing a fraud detection model using Azure Synapse Spark Pools and integrating with Azure ML for deployment in a banking scenario.

Module 7: GCP BigQuery Essentials for Data Scientists

- BigQuery Architecture: Serverless, Scalable, and Cost-Effective.
- Loading data into BigQuery: Batch (Cloud Storage) and Streaming Inserts.
- Standard SQL querying and BigQuery specific functions.
- Understanding BigQuery Pricing: Storage and Querying costs.
- **Case Study:** A media company analyzing petabytes of user interaction data in real-time with BigQuery for personalized content recommendations.

Module 8: Advanced GCP BigQuery for Data Scientists

- Partitioning and Clustering for query optimization and cost reduction.
- BigQuery ML: In-database machine learning capabilities.
- External Tables and Federated Queries for data access outside BigQuery.
- Data Security, Access Control, and Data Loss Prevention (DLP).
- **Case Study:** Building a customer churn prediction model using BigQuery ML on historical customer data for a telecommunications company.

Module 9: Cross-Cloud Data Integration Strategies

- Data migration patterns: Lift-and-shift, gradual migration.
- Data replication and synchronization across cloud platforms.
- Tools and services for inter-cloud data transfer (e.g., AWS DataSync, Azure Data Box, GCP Transfer Service).
- Developing a multi-cloud data strategy for resilience and vendor lock-in avoidance.
- **Case Study:** An e-commerce business using a multi-cloud strategy to manage peak load analytics and ensure business continuity across AWS and GCP.

Module 10: Data Ingestion and ETL/ELT Pipelines

- Designing robust data ingestion pipelines: Batch vs. Streaming.
- Utilizing cloud-native ETL tools: AWS Glue, Azure Data Factory, GCP Dataflow.
- Implementing ELT patterns for faster data loading and transformation.

- Orchestrating data pipelines with Apache Airflow or cloud equivalents.
- **Case Study:** Automating financial report generation by building an ELT pipeline using AWS Glue and Redshift for a large enterprise.

Module 11: Advanced SQL for Cloud Data Platforms

- Window Functions for complex analytical queries.
- Common Table Expressions (CTEs) for query modularity and readability.
- JSON and Semi-structured data querying in all three platforms.
- Performance tuning SQL queries: Execution plans, indexing.
- **Case Study:** A ride-sharing company using advanced SQL queries with window functions in BigQuery to analyze driver efficiency and passenger wait times.

Module 12: Data Governance and Security in the Cloud

- Identity and Access Management (IAM) best practices across AWS, Azure, GCP.
- Data encryption at rest and in transit.
- Auditing and monitoring data access and usage.
- Compliance considerations: GDPR, HIPAA, CCPA.
- **Case Study:** Implementing a robust data governance framework for sensitive customer data on Azure Synapse to meet regulatory compliance standards.

Module 13: Integrating with Data Science Ecosystems

- Connecting cloud data platforms to Python (Pandas, Dask), R, and Jupyter Notebooks.
- Leveraging cloud SDKs for programmatic data access.
- Using BI tools (Tableau, Power BI, Looker) with cloud data warehouses.
- Version control for data science projects and notebooks.
- **Case Study:** A research firm integrating Redshift with Python for complex statistical modeling and then visualizing results in Tableau.

Module 14: Performance Optimization and Cost Management

- Monitoring cloud data platform usage and performance metrics.
- Strategies for query cost optimization: caching, materialized views, proper table design.
- Resource scaling and autoscaling configurations.
- Best practices for minimizing cloud expenses without compromising performance.
- **Case Study:** A media streaming service reducing its Redshift costs by 30% through effective WLM, query optimization, and rightsizing clusters.

Module 15: Future Trends and Advanced Topics

- Serverless data processing beyond SQL: AWS Lambda, Azure Functions, Cloud Functions.
- Data Mesh and Data Fabric architectures.
- The rise of Generative AI and its impact on data platforms.
- Ethical considerations in data science and cloud computing.
- **Case Study:** Exploring how a smart city initiative is planning to integrate real-time IoT data streams into BigQuery and apply AI for urban planning.

Training Methodology

This training course employs a highly interactive and practical methodology to ensure maximum knowledge retention and skill development. It combines:

- **Instructor-Led Sessions**: Engaging lectures, discussions, and Q&A sessions.
- **Hands-on Labs**: Extensive practical exercises using real cloud environments (AWS, Azure, GCP) to reinforce concepts.
- **Live Demos**: Step-by-step demonstrations of key functionalities and best practices.
- **Real-world Case Studies**: Analysis of industry examples to illustrate practical applications and problem-solving.
- **Group Activities & Discussions**: Collaborative learning and knowledge sharing among participants.
- **Best Practices and Troubleshooting**: Insights into common challenges and effective solutions.
- **Q&A and Interactive Problem Solving**: Dedicated time for addressing participant queries and tackling real-world scenarios.

Register as a group from 3 participants for a Discount

Send us an email: info@fineskilltrainingcenter.org or call *+254769199797*

Certification

Upon successful completion of this training, participants will be issued with a globally- recognized certificate.

Tailor-Made Course

We also offer tailor-made courses based on your needs.

Key Notes

a. The participant must be conversant with English.

b. Upon completion of training the participant will be issued with an Authorized Training Certificate

c. Course duration is flexible and the contents can be modified to fit any number of days.

d. The course fee includes facilitation training materials, 2 coffee breaks, buffet lunch and A Certificate upon successful completion of Training.

e. One-year post-training support Consultation and Coaching provided after the course.

f. Payment should be done at least a week before commence of the training, to Fineskill Training Center account, as indicated in the invoice so as to enable us prepare better for you.

Upcoming Scheduled Sessions

Start Date	End Date	Location	Price
23 Nov 2026	04 Dec 2026	Online	$2,000
23 Nov 2026	04 Dec 2026	Physical	$3,000
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$9,000
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$12,000

Ready to enroll or request a customized program? Book online or contact us:

Register Online Now

Email: info@fineskilltrainingcenter.com | Website: www.fineskilltrainingcenter.com