Cloud Computing for Large-Scale Data Analysis (AWS, Azure, GCP) Training Course

Professional Development Training Course Outline

Category	Research and Data Analysis	**Duration**	5 Days
Base Fee	$1,500 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

Cloud computing has revolutionized how organizations manage, store, and analyze large-scale data. As the demand for scalable, secure, and efficient data processing grows, the integration of platforms such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) becomes essential. This course empowers participants to harness cloud infrastructure to perform complex data analysis, implement distributed computing techniques, and optimize big data workflows with modern cloud-native tools.

Cloud Computing for Large-Scale Data Analysis Training Course is designed to deliver hands-on experience in deploying cloud-based data solutions using leading platforms. Whether it's real-time data ingestion, machine learning integration, or automated pipeline development, learners will gain critical skills needed to drive innovation and digital transformation in data-driven industries.

Programme Curriculum

Cloud Computing for Large-Scale Data Analysis Training Course

Introduction

Cloud computing has revolutionized how organizations manage, store, and analyze large-scale data. As the demand for scalable, secure, and efficient data processing grows, the integration of platforms such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) becomes essential. This course empowers participants to harness cloud infrastructure to perform complex data analysis, implement distributed computing techniques, and optimize big data workflows with modern cloud-native tools.

Cloud Computing for Large-Scale Data Analysis Training Course is designed to deliver hands-on experience in deploying cloud-based data solutions using leading platforms. Whether it's real-time data ingestion, machine learning integration, or automated pipeline development, learners will gain critical skills needed to drive innovation and digital transformation in

data-driven industries.

Course Objectives

1. Understand the fundamentals of cloud computing in the context of big data analytics
2. Explore key differences between AWS, Azure, and GCP for data workloads
3. Deploy scalable storage and compute resources using cloud-native tools
4. Implement distributed data processing with Hadoop and Spark on the cloud
5. Optimize data pipelines for performance and cost-efficiency
6. Utilize cloud-based databases (Redshift, BigQuery, Azure Synapse) for analytics
7. Configure and manage cloud-based ETL workflows
8. Analyze real-time and batch data using managed services
9. Integrate AI and ML models into cloud analytics pipelines
10. Ensure data security, governance, and compliance on cloud platforms
11. Use Kubernetes and Docker for cloud-based data containerization
12. Monitor and troubleshoot large-scale analytics systems in the cloud
13. Plan and architect end-to-end cloud solutions for enterprise-level data projects

Target Audience

1. Data Scientists
2. Cloud Engineers
3. IT Managers
4. Business Intelligence Analysts
5. DevOps Professionals
6. System Architects
7. AI/ML Engineers
8. Graduate Students in Data Analytics

Course Duration: 5 days

Course Modules

Module 1: Introduction to Cloud Computing for Data Analysis

- Overview of cloud ecosystems (AWS, Azure, GCP)
- Benefits and limitations of cloud computing for big data
- Key service models (IaaS, PaaS, SaaS)
- Storage and compute services overview
- Role of virtualization and containers
- **Case Study:** Migrating an on-premise data warehouse to AWS

Module 2: Cloud Storage and Compute Essentials

- Cloud storage types: Object, Blob, File
- Compute options: EC2, Azure VMs, GCP Compute Engine
- Resource provisioning and scaling
- Security and IAM configuration
- Cost management and budgeting
- **Case Study:** Choosing the right compute/storage strategy for e-commerce data

Module 3: Distributed Data Processing with Hadoop and Spark

- Installing and configuring Hadoop/Spark on cloud platforms
- Spark vs. Hadoop: Use cases and performance

- Running MapReduce jobs on the cloud
- Integrating with S3, Blob Storage, or GCS
- Cluster management with EMR, Dataproc, and HDInsight
- **Case Study:** Real-time sentiment analysis using Spark on GCP Dataproc

Module 4: Cloud-Based ETL and Data Pipelines

- Building ETL workflows with AWS Glue, Azure Data Factory, GCP Dataflow
- Data ingestion tools: Kinesis, Event Hubs, Pub/Sub
- Automating data transformation
- Handling unstructured data
- Orchestrating workflows with Apache Airflow and cloud-native tools
- **Case Study:** Streaming IoT sensor data into BigQuery using GCP pipelines

Module 5: Cloud Databases and Data Warehousing

- Introduction to cloud-native databases (RDS, Cosmos DB, Cloud SQL)
- Data warehousing: Redshift, BigQuery, Synapse Analytics
- Query optimization techniques
- Connecting BI tools (Tableau, Power BI) to cloud databases
- Backup, replication, and disaster recovery
- **Case Study:** Implementing a centralized cloud data warehouse for a logistics firm

Module 6: Integrating Machine Learning in Cloud Analytics

- ML model deployment using SageMaker, Azure ML, Vertex AI
- Training models on cloud GPUs and TPUs
- Integrating ML into ETL pipelines
- Monitoring and retraining models
- Ethical AI and compliance in cloud platforms
- **Case Study:** Predicting customer churn using Azure ML integrated with Data Factory

Module 7: Securing Data in the Cloud

- Identity and Access Management (IAM)
- Encryption at rest and in transit
- Compliance standards (GDPR, HIPAA, SOC2)
- Monitoring access and audit logs
- Risk assessment and threat detection tools
- **Case Study:** Implementing a secure data lake architecture for a healthcare provider

Module 8: Cloud Architecture, Monitoring, and Optimization

- Designing cloud-native data solutions
- High availability and fault tolerance
- Cost analysis and billing optimization
- Logging and monitoring with CloudWatch, Azure Monitor, Stackdriver
- Performance tuning for large-scale analytics
- **Case Study:** Building a cost-efficient multi-cloud architecture for real-time analytics

Training Methodology

- Instructor-led virtual training with cloud lab environments
- Hands-on exercises using AWS, Azure, and GCP consoles
- Real-world case studies and industry scenarios
- Capstone project and group presentations

- Weekly assessments and practical quizzes

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	27 Nov 2026	Online	$1,000
23 Nov 2026	27 Nov 2026	Physical	$1,500
23 Nov 2026	27 Nov 2026	Physical	$3,000
23 Nov 2026	27 Nov 2026	Physical	$3,000
23 Nov 2026	27 Nov 2026	Physical	$6,000
23 Nov 2026	27 Nov 2026	Physical	$4,500
23 Nov 2026	27 Nov 2026	Physical	$6,000
23 Nov 2026	27 Nov 2026	Physical	$6,000

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

Register Online Now

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