Data Lake Architecture Training Course

Professional Development Training Course Outline

Category	Business Intelligence	**Duration**	5 Days
Base Fee	$1,500 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

Data Lake Architecture has become a cornerstone in modern data-driven enterprises, enabling scalable data storage, real-time analytics, and advanced data engineering practices. This course provides a comprehensive, industry-aligned learning experience focused on cloud-based data lakes, big data frameworks, distributed storage systems, and modern data governance strategies. Participants will gain hands-on expertise in building secure, scalable, and high-performance data lake ecosystems using cutting-edge tools and technologies such as data ingestion pipelines, data cataloging, and metadata management.

With the rise of artificial intelligence, machine learning, and predictive analytics, organizations require robust data lake solutions to manage structured, semi-structured, and unstructured data efficiently. Data Lake Architecture Training Course equips professionals with in-demand skills in data architecture design, data integration, data security, and performance optimization. By leveraging real-world scenarios and case studies, learners will develop practical capabilities to implement modern data lake architectures that support digital transformation, business intelligence, and data democratization initiatives.

Programme Curriculum

Data Lake Architecture Training Course

Introduction

Data Lake Architecture has become a cornerstone in modern data-driven enterprises, enabling scalable data storage, real-time analytics, and advanced data engineering practices. This course provides a comprehensive, industry-aligned learning experience focused on cloud-based data lakes, big data frameworks, distributed storage systems, and modern data governance strategies. Participants will gain hands-on expertise in building secure, scalable, and high-performance data lake ecosystems using cutting-edge tools and technologies such as data ingestion pipelines, data cataloging, and metadata management.

With the rise of artificial intelligence, machine learning, and predictive analytics, organizations require robust data lake solutions to manage structured, semi-structured, and unstructured data efficiently. Data Lake Architecture Training Course equips professionals with in-demand skills in data architecture design, data integration, data security, and performance optimization. By leveraging real-world scenarios and case studies, learners will develop practical capabilities to implement modern data lake architectures that support digital transformation, business intelligence, and data democratization initiatives.

Course Objectives

1. Understand modern Data Lake Architecture and cloud-native data platforms
2. Design scalable and fault-tolerant data storage systems
3. Implement efficient data ingestion pipelines and ETL/ELT processes
4. Apply data governance, compliance, and data quality frameworks
5. Integrate big data technologies such as Hadoop and Spark
6. Optimize data lake performance using partitioning and indexing strategies
7. Develop real-time data processing and streaming solutions
8. Implement data security, encryption, and access control mechanisms
9. Build metadata management and data catalog solutions
10. Enable advanced analytics, AI, and machine learning workflows
11. Automate data pipelines using orchestration tools
12. Design hybrid and multi-cloud data lake architectures
13. Monitor and troubleshoot data lake environments effectively

Organizational Benefits

- Enhanced data-driven decision making and business intelligence
- Improved scalability and flexibility in handling large datasets
- Reduced data storage and processing costs
- Faster time-to-insight with real-time analytics capabilities
- Strengthened data governance and regulatory compliance
- Improved collaboration across data teams and business units
- Increased operational efficiency through automation
- Better support for AI and machine learning initiatives
- Centralized data repository for enterprise-wide access
- Improved data quality and consistency

Target Audiences

1. Data Engineers
2. Data Architects
3. Business Intelligence Professionals
4. Cloud Engineers
5. IT Managers
6. Database Administrators
7. Big Data Analysts
8. Software Developers

Course Duration: 5 days

Course Modules

Module 1: Introduction to Data Lake Architecture

- Fundamentals of data lakes and data warehousing concepts
- Key components of modern data lake ecosystems
- Differences between data lakes, warehouses, and lakehouses
- Industry trends and emerging technologies
- Benefits and challenges of implementing data lakes
- Case study: Enterprise transition from traditional data warehouse to data lake

Module 2: Data Ingestion and Integration

- Batch and real-time data ingestion techniques
- ETL vs ELT strategies in modern architectures
- Data integration from multiple sources
- API-based and streaming data ingestion
- Data pipeline automation tools
- Case study: Building a scalable ingestion pipeline for IoT data

Module 3: Storage and Data Management

- Distributed storage systems and object storage
- Data partitioning and indexing strategies
- Schema design for structured and unstructured data
- Data lifecycle management and archiving
- Cost optimization strategies
- Case study: Optimizing storage for high-volume transactional data

Module 4: Big Data Processing Frameworks

- Introduction to Hadoop and Spark ecosystems
- Batch vs stream processing models
- Data transformation and processing pipelines
- Performance tuning and optimization
- Integration with cloud platforms
- Case study: Processing large datasets using Apache Spark

Module 5: Data Governance and Security

- Data governance frameworks and policies
- Data quality management and validation
- Role-based access control and authentication
- Data encryption and privacy regulations
- Compliance with global standards
- Case study: Implementing governance in a financial institution

Module 6: Metadata Management and Cataloging

- Importance of metadata in data lakes
- Data catalog tools and techniques
- Data lineage and traceability
- Data discovery and indexing
- Integration with governance frameworks

- Case study: Building a centralized data catalog

Module 7: Advanced Analytics and Machine Learning Integration

- Enabling AI and ML workflows in data lakes
- Data preparation for analytics
- Integration with analytics tools
- Real-time analytics and predictive modeling
- Visualization and reporting tools
- Case study: Predictive analytics using data lake architecture

Module 8: Monitoring, Optimization, and Future Trends

- Monitoring tools and performance metrics
- Troubleshooting data pipeline issues
- Optimization of data workflows
- Hybrid and multi-cloud architectures
- Future trends in data lake technologies
- Case study: Scaling data lake infrastructure for global operations

Training Methodology

- Instructor-led interactive sessions with practical demonstrations
- Hands-on labs and real-world project implementation
- Case study analysis and group discussions
- Use of industry-standard tools and cloud platforms
- Continuous assessment through quizzes and assignments
- Collaborative learning and peer knowledge sharing
- Access to training materials and reference resources

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