Training Course on Data Warehousing and Data Lake Architecture for Analytics

Professional Development Training Course Outline

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

Course Introduction

In today's data-driven landscape, organizations are grappling with ever-increasing volumes and varieties of data. To transform this raw data into actionable insights, robust **Data Warehousing** and **Data Lake Architectures** are paramount. Training Course on Data Warehousing and Data Lake Architecture for Analytics provides an in-depth exploration of the fundamental principles and best practices for designing, implementing, and managing these critical data platforms. Participants will gain the expertise to leverage these architectures for advanced **Analytics**, **Business Intelligence**, and **Machine Learning**, enabling their organizations to make informed decisions and achieve a significant competitive advantage.

This program delves into the synergy between traditional data warehouses and modern data lakes, highlighting their respective strengths and optimal use cases. We will cover cutting-edge concepts like **Cloud-Native Data Platforms**, **Data Lakehouses**, **Real-time Data Processing**, and robust **Data Governance** strategies. By mastering these architectural patterns, professionals will be equipped to build scalable, flexible, and secure data ecosystems that fuel predictive analytics, operational efficiency, and transformative innovation across various industries.

Programme Curriculum

Training Course on Data Warehousing and Data Lake Architecture for Analytics

Introduction

In today's data-driven landscape, organizations are grappling with ever-increasing volumes and varieties of data. To transform this raw data into actionable insights, robust **Data Warehousing** and **Data Lake Architectures** are paramount. Training Course on Data Warehousing and Data Lake Architecture for Analytics provides an in-depth exploration of the fundamental principles and best practices for designing, implementing, and managing these critical data platforms. Participants will

gain the expertise to leverage these architectures for advanced **Analytics**, **Business Intelligence**, and **Machine Learning**, enabling their organizations to make informed decisions and achieve a significant competitive advantage.

This program delves into the synergy between traditional data warehouses and modern data lakes, highlighting their respective strengths and optimal use cases. We will cover cutting-edge concepts like **Cloud-Native Data Platforms**, **Data Lakehouses**, **Real-time Data Processing**, and robust **Data Governance** strategies. By mastering these architectural patterns, professionals will be equipped to build scalable, flexible, and secure data ecosystems that fuel predictive analytics, operational efficiency, and transformative innovation across various industries.

Course Duration

10 days

Course Objectives

1. Design and implement highly scalable and cost-effective data warehouses on leading cloud platforms (e.g., AWS Redshift, Snowflake, Google BigQuery, Azure Synapse Analytics).
2. Understand and apply best practices for building flexible, schema-on-read data lakes using object storage (e.g., AWS S3, Azure Data Lake Storage, Google Cloud Storage).
3. Explore the hybrid **Data Lakehouse** architecture to combine the flexibility of data lakes with the performance and governance of data warehouses.
4. Develop robust **Dimensional Modeling** techniques, including star and snowflake schemas, for efficient data warehousing.
5. Master data ingestion, transformation, and loading (ETL/ELT) processes for both batch and **Real-time Data Streaming** scenarios.
6. Implement comprehensive **Data Governance**, **Data Quality**, and **Data Security** frameworks for sensitive data within both environments.
7. Gain hands-on experience with key **Big Data** processing frameworks like Apache Spark, Hadoop, and Flink for large-scale data manipulation.
8. Prepare and optimize data within data warehouses and data lakes for advanced **Analytics**, **Business Intelligence (BI)**, and **Machine Learning (ML)** workloads.
9. Establish effective **Metadata Management** and data cataloging strategies for improved data discoverability and understanding.
10. Apply techniques for **Query Optimization**, indexing, and partitioning to ensure high performance for analytical queries.
11. Explore emerging architectural concepts like **Data Fabric** and **Data Mesh** for decentralized data management and accessibility.
12. Develop strategies for efficient **Data Lifecycle Management**, including data retention, archiving, and deletion policies.
13. Foster a culture of **Data Democratization** by providing accessible and trustworthy data to business users for self-service analytics.

Organizational Benefits

- Provide timely and accurate insights by building efficient data pipelines and analytical platforms.
- Streamline data management processes, reduce manual effort, and automate data workflows.
- Implement robust data governance and quality frameworks, leading to more reliable data for analytics.

- Enable the adoption of Machine Learning and Artificial Intelligence initiatives by providing well-structured and accessible data.
- Leverage cloud-native solutions and efficient architectural patterns to reduce data storage and processing expenses.
- Implement strong security measures and compliance protocols to protect sensitive organizational data.
- Adapt quickly to changing business requirements by building flexible and scalable data architectures.
- Utilize advanced analytics capabilities to uncover new opportunities, optimize customer experiences, and differentiate in the market.

Target Audience

1. Data Architects
2. Data Engineers
3. BI Developers/Analysts
4. Cloud Architects
5. Database Administrators
6. Data Scientists
7. IT Managers/Leaders
8. Solution Architects

Course Outine

Module 1: Introduction to Data Warehousing & Data Lakes

- Defining Data Warehouses: Purpose, characteristics, and historical evolution.
- Understanding Data Lakes: Benefits, challenges, and "schema-on-read" principle.
- Key Differences & Synergy: Data Warehouses vs. Data Lakes and when to use each.
- Introduction to the Data Lakehouse Concept: Bridging the gap.
- **Case Study:** A retail company's journey from traditional reporting to a combined data strategy for omnichannel analytics.

Module 2: Core Concepts of Dimensional Modeling

- Star Schema Design: Fact tables, dimension tables, and primary/foreign keys.
- Snowflake Schema: Normalization and its implications for query performance.
- Conformed Dimensions and Degenerate Dimensions: Ensuring data consistency.
- Slowly Changing Dimensions (SCD Type 1, 2, 3): Handling evolving master data.
- **Case Study:** Designing a dimensional model for a financial institution to analyze customer transactions and account balances.

Module 3: Data Warehousing Architecture & Design Principles

- Layered Architecture: Staging, data warehouse, and data mart layers.
- Data Integration Strategies: ETL (Extract, Transform, Load) vs. ELT (Extract, Load, Transform).
- Data Partitioning and Indexing: Optimizing data retrieval.
- Physical Data Model Design: Storage considerations and performance tuning.
- **Case Study:** Re-architecting a large-scale enterprise data warehouse for improved performance and scalability using cloud-native services.

Module 4: Data Lake Architecture & Best Practices

- Data Lake Components: Storage (e.g., S3, ADLS), processing, and cataloging.
- Data Ingestion Patterns: Batch, streaming, and real-time ingestion.
- Data Zones (Raw, Curated, Consumption): Organizing data within the data lake.
- File Formats for Analytics: Parquet, ORC, Avro for optimal storage and query.
- **Case Study:** A media company building a data lake to store user clickstream data and enable personalized content recommendations.

Module 5: Cloud-Native Data Warehousing Platforms

- AWS Redshift: Architecture, best practices, and cost optimization.
- Snowflake: Unique architecture, features, and use cases.
- Google BigQuery: Serverless analytics warehouse, pricing, and scalability.
- Azure Synapse Analytics: Unified analytics platform features.
- **Case Study:** Migrating an on-premise data warehouse to a cloud-native platform to reduce operational overhead and improve elasticity.

Module 6: Data Ingestion and ETL/ELT Processes

- Data Source Connectivity: Integrating with various relational and NoSQL databases.
- Batch Processing Tools: Apache NiFi, AWS Glue, Azure Data Factory, etc.
- Streaming Data Ingestion: Apache Kafka, AWS Kinesis, Azure Event Hubs.
- Data Transformation Techniques: Cleaning, enriching, and standardizing data.
- **Case Study:** Implementing a real-time data ingestion pipeline for IoT sensor data into a data lake for immediate anomaly detection.

Module 7: Data Governance and Data Quality

- Data Governance Frameworks: Policies, standards, and roles.
- Data Quality Dimensions: Accuracy, completeness, consistency, timeliness.
- Data Profiling and Cleansing Techniques: Identifying and resolving data issues.
- Metadata Management and Data Catalogs: Improving data discoverability and understanding.
- **Case Study:** A financial services firm implementing data governance to ensure regulatory compliance and improve data trust.

Module 8: Data Security in Data Warehouses & Data Lakes

- Authentication and Authorization: Access control mechanisms.
- Data Encryption: At rest and in transit.
- Row-Level and Column-Level Security: Granular data access control.
- Auditing and Logging: Tracking data access and changes.
- **Case Study:** Securing sensitive customer data in a cloud data lake environment to meet strict privacy regulations.

Module 9: Querying and Performance Optimization

- SQL Optimization Techniques: Efficient query writing and indexing.
- Distributed Query Engines: Presto, Trino, Apache Hive for data lakes.
- Workload Management and Concurrency: Handling multiple users and queries.
- Caching Strategies: Improving query response times.
- **Case Study:** Optimizing query performance in a large-scale data warehouse for complex analytical reports, reducing execution time by 50%.

Module 10: Advanced Analytics and Machine Learning Integration

- Data Preparation for ML: Feature engineering, data scaling, and handling missing values.
- Integrating with ML Platforms: Connecting data lakes/warehouses to TensorFlow, PyTorch, Databricks.
- Batch vs. Real-time ML Inference: Using data platforms for model deployment.
- Data Versioning and Lineage for ML Reproducibility.
- **Case Study:** Leveraging a data lake to train machine learning models for customer churn prediction in a telecom company.

Module 11: Data Lakehouse Best Practices

- Delta Lake, Apache Iceberg, Apache Hudi: Open table formats for data lakes.
- ACID Transactions on Data Lakes: Ensuring data consistency.
- Schema Evolution: Managing changes in data schema over time.
- Unified Data Access: SQL, Python, R access on a single data store.
- **Case Study:** Implementing a Delta Lake on an existing data lake to enable reliable data updates and streaming analytics for a logistics company.

Module 12: Data Fabric and Data Mesh

- Data Fabric Concepts: Unified data access and virtualized data layers.
- Data Mesh Principles: Decentralized data ownership and domain-driven data products.
- Comparing Data Fabric and Data Mesh: Use cases and implementation considerations.
- Governance and Interoperability in Distributed Data Architectures.
- **Case Study:** A large enterprise exploring Data Mesh to break down data silos and empower individual business units with data ownership.

Module 13: Real-time Data Warehousing and Streaming Analytics

- Real-time Data Sources: IoT, social media feeds, financial transactions.
- Stream Processing Engines: Apache Flink, Spark Streaming.
- Lambda and Kappa Architectures: Designing for real-time and batch processing.
- Low-Latency Data Storage Solutions.
- **Case Study:** Building a real-time analytics dashboard for a ride-sharing company to monitor driver and passenger activity.

Module 14: Data Lifecycle Management and Archiving

- Data Retention Policies: Legal and business requirements.
- Data Archiving Strategies: Cost-effective storage for historical data.
- Data Deletion and Compliance: GDPR, CCPA considerations.
- Data Tiering: Moving data between hot, warm, and cold storage.
- **Case Study:** Developing a data lifecycle management strategy for a healthcare provider to comply with patient data retention regulations.

Module 15: Future Trends and Emerging Technologies

- Data Virtualization: Accessing data without physical consolidation.
- Graph Databases for Analytics: Analyzing relationships in data.
- AI-Powered Data Management: Automation and intelligent data discovery.
- Quantum Computing and its potential impact on data processing.
- **Case Study:** A research institution exploring the use of graph databases for complex scientific data analysis.

Training Methodology

This training course employs a blended learning approach, combining:

- **Interactive Lectures:** Engaging presentations of core concepts and design principles.
- **Hands-on Labs & Exercises:** Practical application of learned skills using industry-standard tools and cloud platforms.
- **Real-world Case Studies:** In-depth analysis of successful implementations and challenges.
- **Group Discussions:** Fostering knowledge sharing and problem-solving among participants.
- **Q&A Sessions:** Addressing specific queries and reinforcing understanding.
- **Project-Based Learning:** Participants will work on a capstone project to design a data warehousing/data lake solution for a hypothetical scenario.
- **Best Practices and Pitfalls:** Highlighting common mistakes and how to avoid them.

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