

Data Engineering

This course will take you through all essential aspects of modern Data Engineering. From foundational concepts and architecture design to workflow orchestration and real-time data processing, this bootcamp is designed to prepare you for the demands of today's data-driven landscape.

With hands-on labs and deep dives into tools like Snowflake, Apache Spark, dbt, and Airflow, you'll learn to build robust, scalable, and high-performance data pipelines. This edition also introduces real-time streaming and advanced visualization using modern BI tools like Tableau and Looker.

DAY 1 Data Engineering Foundations



- Core responsibilities of data engineers
- Data system architecture
- Emerging industry trends
- Key performance metrics

Data Modeling



- Relational vs dimensional modeling
- Star and snowflake schemas
- Normalization techniques
- Best practices in data design



Tools and technologies

- Snowflake deep dive
- Cloud data warehousing concepts
- Architecture and deployment models
- Apache Spark overview
- Distributed computing principles
- Performance optimization strategies

Development of the environment



- Snowflake configuration
- Account setup
- Security configurations
- dbt introduction
- Data transformation workflows
- Version control integration
- Hands-on lab: Initial data model creation

DAY 2 Data Processing and Workflow Management

Advanced Apache Spark



- DataFrame operations
- RDD transformations
- Performance tuning
- Memory management strategies

Workflow Orchestration



- Apache Airflow fundamentals
- DAG design
- Scheduling and monitoring
- Dependency management

Batch Processing Techniques



- Data ingestion patterns
- Spark-Snowflake integration
- Error handling
- Scalability considerations

Practical Pipeline Development



- End-to-end data pipeline construction
- Using Spark, dbt, Airflow
- Code review and optimization
- Performance benchmarking

Data Engineering



DAY 3 Advanced Topics and Capstone Project



- Streaming Data Processing
- Real-time data concepts
- Spark Streaming architecture
- Event-driven architectures
- Use case scenarios

Capstone Project



- Comprehensive data pipeline development
- Multi-tool integration
- Performance optimization
- Peer review and feedback
- Presentation of solutions



Data Visualization and Reporting

- Tool ecosystem (Tableau, Looker)
- Snowflake data extraction
- Visualization best practices
- Interactive dashboard design

Conclusion and Learning Paths



- Key learnings recap
- Industry certification paths
- Advanced learning resources



Learning outcomes

This training will equip Data Engineering professionals to design and manage scalable data systems, automate workflows, and transform raw data into actionable insights and enable your teams to optimize data management processes, improve data quality, accelerate analysis, and reduce operational costs.



Your profile

- Data Engineers
- Software researchers interested in data engineering and management systems.



Prerequisites

- Programming skills (1+ year experience)
- Basic SQL knowledge
- Comfort with command line
- Familiarity with Python is a plus

