Programming for Data Engineers 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

Programming for Data Engineers Training Course is meticulously designed to equip professionals with the skills required to manage, process, and analyze large-scale data efficiently. With the exponential growth of data in modern organizations, the demand for data engineers proficient in programming languages such as Python, SQL, and Scala is rapidly increasing. This course focuses on blending practical programming techniques with data engineering concepts to build scalable data pipelines, ensure data integrity, and optimize data-driven solutions for real-world applications. Participants will gain hands-on experience through interactive labs, real-time projects, and case studies, enabling them to solve complex data challenges effectively.

In addition to technical proficiency, this course emphasizes problem-solving, workflow automation, and optimization strategies for large datasets. Trainees will learn how to integrate programming skills with cloud platforms, data warehousing, and ETL frameworks, ensuring they can design robust and scalable data systems. The course leverages the latest industry-standard tools and best practices to provide a comprehensive understanding of modern data engineering processes. By the end of the training, participants will be capable of delivering end-to-end data solutions that drive organizational decision-making and innovation.

Programme Curriculum

Programming for Data Engineers Training Course

Introduction

Programming for Data Engineers Training Course is meticulously designed to equip professionals with the skills required to manage, process, and analyze large-scale data efficiently. With the exponential growth of data in modern organizations, the demand for data engineers proficient in programming languages such as Python, SQL, and Scala is rapidly increasing. This course focuses on blending practical programming techniques with

data engineering concepts to build scalable data pipelines, ensure data integrity, and optimize data-driven solutions for real-world applications. Participants will gain hands-on experience through interactive labs, real-time projects, and case studies, enabling them to solve complex data challenges effectively.

In addition to technical proficiency, this course emphasizes problem-solving, workflow automation, and optimization strategies for large datasets. Trainees will learn how to integrate programming skills with cloud platforms, data warehousing, and ETL frameworks, ensuring they can design robust and scalable data systems. The course leverages the latest industry-standard tools and best practices to provide a comprehensive understanding of modern data engineering processes. By the end of the training, participants will be capable of delivering end-to-end data solutions that drive organizational decision-making and innovation.

Course Objectives

1. Master Python, SQL, and Scala programming for data engineering applications
2. Design and implement efficient ETL pipelines for large-scale data
3. Develop data processing workflows with Apache Spark and Hadoop
4. Apply data modeling and schema design best practices
5. Integrate cloud-based platforms like AWS, Azure, and Google Cloud for data engineering
6. Ensure data quality, governance, and security in enterprise systems
7. Optimize data storage and retrieval for performance efficiency
8. Implement real-time data streaming solutions using Kafka and Spark Streaming
9. Perform automated testing and debugging of data pipelines
10. Leverage version control systems and collaborative programming techniques
11. Analyze big data using advanced algorithms and distributed computing
12. Understand emerging trends in AI-driven data engineering
13. Apply programming solutions to real-world business challenges through case studies

Organizational Benefits

- Accelerate data-driven decision-making within the organization
- Improve efficiency in handling large datasets and workflows
- Reduce operational costs through optimized data pipelines
- Ensure compliance with data governance and security standards
- Enhance cross-team collaboration in data projects
- Facilitate adoption of cloud technologies for scalable solutions
- Promote a culture of innovation through automated data processing
- Reduce errors in ETL processes with standardized programming practices
- Enable real-time analytics and faster business insights
- Strengthen competitive advantage with advanced data engineering skills

Target Audiences

1. Data Engineers seeking to enhance programming skills
2. Software Developers transitioning into data engineering
3. Business Intelligence Analysts aiming for advanced data processing knowledge
4. Database Administrators focusing on big data solutions
5. Cloud Architects integrating data pipelines with cloud infrastructure
6. Data Scientists requiring pipeline and data engineering expertise
7. IT Managers overseeing data-driven projects
8. Students and professionals aspiring to enter the data engineering field

Course Duration: 5 days

Course Modules

Module 1: Introduction to Programming for Data Engineering

- Overview of data engineering and its role in modern organizations
- Programming essentials: Python, SQL, and Scala
- Data types, structures, and object-oriented programming concepts
- Hands-on exercises: Writing your first data scripts
- Real-world case study: Automating batch data processing
- Practical assignment: Build a simple ETL pipeline

Module 2: SQL for Data Engineers

- Advanced SQL queries for data extraction and transformation
- Joins, subqueries, window functions, and CTEs
- Optimizing SQL queries for large datasets
- Hands-on exercises with relational databases
- Real-world case study: Building a data mart for sales analytics
- Practical assignment: Implement complex queries for reporting

Module 3: Python for Data Engineering

- Python libraries for data manipulation: pandas, numpy, and pyarrow
- Writing reusable functions and scripts for data processing
- Handling missing data, transformations, and cleaning
- Hands-on exercises: Automating file ingestion and transformation
- Real-world case study: Developing a customer data pipeline
- Practical assignment: Create a Python ETL script

Module 4: Data Modeling and Schema Design

- Fundamentals of relational and non-relational databases
- Designing normalized and denormalized schemas
- Best practices for scalable and maintainable data models
- Hands-on exercises: Data modeling for analytics
- Real-world case study: Designing a warehouse for e-commerce data
- Practical assignment: Implement a star schema model

Module 5: ETL Pipelines and Workflow Automation

- Introduction to ETL frameworks and best practices
- Scheduling workflows using Apache Airflow
- Automating data ingestion and processing
- Hands-on exercises: Building a complete ETL pipeline
- Real-world case study: Automating payroll data processing
- Practical assignment: Create an automated ETL workflow

Module 6: Big Data Frameworks: Spark and Hadoop

- Introduction to distributed computing concepts

- Hands-on with Apache Spark and Hadoop ecosystems
- Processing batch and real-time data with Spark
- Optimizing Spark jobs for performance
- Real-world case study: Processing social media data at scale
- Practical assignment: Build a Spark-based data pipeline

Module 7: Real-Time Data Streaming

- Understanding data streaming architectures
- Kafka and Spark Streaming for real-time analytics
- Hands-on exercises: Implementing streaming pipelines
- Monitoring and debugging streaming data flows
- Real-world case study: Real-time fraud detection in banking
- Practical assignment: Create a real-time data dashboard

Module 8: Cloud Integration and Deployment

- Overview of cloud platforms for data engineering
- AWS, Azure, and Google Cloud services for data pipelines
- Deploying ETL pipelines and managing cloud storage
- Security, permissions, and cost optimization in the cloud
- Real-world case study: Cloud-based data warehouse deployment
- Practical assignment: Migrate an ETL workflow to cloud infrastructure

Training Methodology

- Interactive lectures with live programming demonstrations
- Hands-on labs and exercises to reinforce learning
- Real-world case studies for practical understanding
- Group discussions and collaborative problem-solving sessions
- Continuous assessment through assignments and quizzes
- Instructor-led feedback and one-on-one mentoring

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