



PYTHON TRAINING

Data Engineering With Python

Develop expert-level data engineering skills using Python. Build scalable data pipelines, automate workflows, and process big data effectively.

DURATION

2 hours

LEVEL

Intermediate

FORMAT

Instructor-Led

- Globally Recognized Certification

- Expert-Led Hands-On Training

- Flexible Scheduling

Computer Pride | 1st Floor, ICEA Building, Kenyatta Avenue, Nairobi | www.computer-pride.co.ke

COURSE OVERVIEW

Unlock the power of Python to build robust and scalable data pipelines with our Data Engineering With Python course. This intermediate-level program is meticulously designed for aspiring and current data professionals looking to master the foundational principles and practical applications of data engineering using one of the most versatile programming languages. You will delve into essential concepts such as data collection, transformation, storage, and orchestration, gaining hands-on experience with industry-standard Python libraries and tools.

This course emphasizes practical, real-world scenarios, enabling you to design, implement, and maintain efficient data solutions. You will learn to work with various data sources, perform complex data manipulations, ensure data quality, and automate workflows, preparing you to tackle the challenges of modern data ecosystems. By the end of this program, you will possess the critical skills needed to contribute significantly to data-driven projects, enhancing your career prospects in the rapidly evolving field of data engineering and contributing to the growing demand for skilled professionals in Kenya and beyond.

Duration	2 hours
Level	Intermediate
Vendor	Python

COURSE CURRICULUM

01 Module 1: Foundations of Data Engineering with Python

- Introduction to Data Engineering Concepts and Ecosystem
- Python Refresher for Data Professionals (Advanced Features)
- Working with Data Structures: Pandas DataFrames and Series
- Data I/O with Python (CSV, JSON, XML, Excel)
- Introduction to Command-Line Tools for Data Operations

02 Module 2: Data Ingestion and Acquisition

- Connecting to Relational Databases (SQLAlchemy, psycopg2)
- API Integration for Data Retrieval (Requests library)
- Web Scraping Fundamentals (Beautiful Soup, Scrapy overview)
- Handling Large Datasets and Streaming Data Concepts
- Introduction to Message Queues (e.g., Kafka/RabbitMQ concepts with Python)

03 Module 3: Data Transformation and Cleaning

- Advanced Pandas Techniques for Data Manipulation
- Data Cleaning and Validation Strategies (Missing Values, Duplicates, Outliers)
- Data Type Conversions and Schema Enforcement
- Implementing Data Quality Checks and Error Handling
- Introduction to PySpark for Distributed Data Processing (conceptual)

04 Module 4: Data Storage and Warehousing Principles

- Overview of Data Lakes vs. Data Warehouses
- Working with Cloud Storage Services (AWS S3/Google Cloud Storage with boto3/google-cloud-storage)
- Introduction to Data Modeling for Analytics (Star/Snowflake Schema)
- Optimizing Data Storage Formats (Parquet, ORC)
- Interacting with NoSQL Databases (e.g., MongoDB with PyMongo)

05 Module 5: Workflow Orchestration and Monitoring

- Introduction to Workflow Orchestration Tools (Apache Airflow concepts)
- Building Directed Acyclic Graphs (DAGs) with Python
- Scheduling and Automating Data Pipelines
- Monitoring and Logging Best Practices for Data Workflows
- Error Handling and Retries in Production Pipelines

06 Module 6: Capstone Project and Best Practices

- Designing an End-to-End Data Pipeline Project
- Implementing a Mini Data Engineering Solution using Learned Tools
- Code Version Control with Git for Data Projects
- Testing Strategies for Data Pipelines
- Deployment Considerations and Best Practices for Production Environments

CAREER PATHWAYS

Graduates of this program are prepared for the following roles:

- > **Junior Data Engineer: Specializing in building and maintaining ETL/ELT pipelines using Python and related frameworks.**
- > **Data Pipeline Developer: Focused on designing, implementing, and optimizing data ingestion and transformation workflows.**
- > **Data Infrastructure Support Analyst: Providing technical support and ensuring the smooth operation of data systems built with Python technologies.**

FREQUENTLY ASKED QUESTIONS

Q: What specific Python libraries and tools will I master to build data pipelines effectively?

This course focuses on practical application, so you'll gain proficiency with essential Python libraries such as Pandas for data manipulation, Requests for API interaction, SQLAlchemy for database connectivity, and potentially boto3 for cloud storage. You will also get an introduction to concepts behind workflow orchestration tools like Apache Airflow for scheduling and managing your data pipelines, ensuring you're equipped with a robust toolkit for modern data engineering.

Q: What are the prerequisites for this 'Data Engineering With Python' course, particularly regarding Python proficiency?

This is an intermediate-level course, so a solid foundational understanding of Python programming is essential. You should be comfortable with Python syntax, data types, control flow, functions, and object-oriented programming concepts. While we'll touch upon some advanced Python features relevant to data, the course assumes you can write functional Python code and understand basic debugging. Familiarity with SQL concepts is also highly recommended.

Q: How will this course help me transition into a dedicated Data Engineer role, specifically focusing on data pipeline development?

This course is designed to equip you with the practical skills needed for data pipeline development. You will learn to collect data from diverse sources, perform complex transformations, manage data storage solutions, and orchestrate automated workflows. The curriculum includes a capstone project that allows you to build an end-to-end data pipeline, providing

tangible experience and a portfolio piece that directly demonstrates your capability to design, implement, and maintain robust data engineering solutions for real-world scenarios.

ABOUT COMPUTER PRIDE



East Africa's Premier IT Training Center

With over 15 years of excellence, Computer Pride delivers globally recognized certification programs from leading technology vendors. Located on Kenyatta Avenue in Nairobi's CBD, our state-of-the-art facilities and expert instructors have empowered thousands of professionals across East Africa.

15+

Years of Excellence

10,000+

Professionals Trained

50+

Vendor Certifications

READY TO ENROL?

Take the next step in your technology career.

Location

1st Flr, ICEA Building
Kenyatta Ave, Nairobi

Email

info@computer-pride.co.ke

Phone

+254 705 900 900

www.computer-pride.co.ke