



PYTHON TRAINING

Python For Data Engineering

Master Python for data engineering, building robust ETL pipelines and scalable data warehouses.

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

This intensive 2-hour intermediate course, "Python For Data Engineering," is meticulously designed for IT professionals and aspiring data engineers looking to leverage Python's powerful ecosystem for robust data solutions. You'll dive deep into core data engineering principles, mastering how to efficiently extract, transform, and load (ETL) data using Python's essential libraries like Pandas and understanding fundamental database connectors. The course focuses on practical, hands-on techniques to build the foundational skills necessary for handling real-world data challenges and preparing data for analytics.

By completing this focused program, you will gain the expertise to automate data workflows, ensure data quality, and optimize data processing tasks. This training equips you with the crucial Pythonic tools to construct scalable and maintainable data preparation scripts, significantly boosting your ability to contribute to data-driven decision-making processes within any organization. It's an essential stepping stone for anyone aiming to solidify their role in the rapidly evolving data landscape, making data accessible and reliable.

Duration	2 hours
Level	Intermediate
Vendor	Python

COURSE CURRICULUM

01 Module 1: Introduction to Python for Data Engineering

- The role of Python in modern data engineering
- Overview of ETL/ELT concepts and their importance
- Setting up a productive Python environment for data tasks (virtual environments, key packages)
- Understanding common data formats (CSV, JSON, XML)

02 Module 2: Data Extraction and Transformation with Pandas

- Efficiently reading data from various sources (files, simple APIs)
- Introduction to the Pandas library for data manipulation
- Data cleaning techniques: handling missing values, duplicates, and inconsistencies
- Essential data transformation operations: filtering, aggregation, merging, and reshaping dataframes

03 Module 3: Data Loading and Storage Fundamentals

- Writing processed data to different file formats (CSV, Parquet, Feather)
- Basic interaction with relational databases using Python (SQLAlchemy introduction)
- Understanding data storage considerations and best practices for structured data

04 Module 4: Building and Automating Simple Data Flows

- Structuring Python scripts for data engineering projects
- Automating repetitive data tasks with functions and basic scheduling concepts
- Implementing robust error handling and logging in data processing scripts
- Introduction to modular programming for data pipeline components

CAREER PATHWAYS

Graduates of this program are prepared for the following roles:

- > **Junior Data Engineer specializing in Python-based ETL processes**
- > **Data Analyst enhancing data manipulation and workflow automation capabilities**
- > **ETL Developer focused on modernizing data integration with Python scripting**

FREQUENTLY ASKED QUESTIONS

Q: What specific Python libraries and tools will I become proficient in through this course for data engineering tasks?

This course primarily focuses on mastering essential data manipulation with the Pandas library, alongside foundational Python programming for scripting and automation. You'll also gain exposure to standard library modules crucial for file I/O, error handling, and basic database interaction concepts through examples like SQLAlchemy.

Q: What are the recommended prerequisites in terms of Python knowledge before enrolling in this 'Intermediate' level course?

To maximize your learning from this intermediate course, it is highly recommended that you possess a solid understanding of fundamental Python programming concepts. This includes familiarity with data types, control flow (loops, conditionals), functions, basic file operations, and core data structures such as lists, dictionaries, and sets. Basic SQL knowledge is beneficial but not strictly required.

Q: How will this 2-hour course help me practically contribute to real-world data pipeline projects immediately after completion?

Upon completing this intensive 2-hour course, you will be equipped to write robust Python scripts for data extraction from common sources, implement efficient data cleaning and transformation logic using Pandas, and develop basic automation for data loading tasks. You'll be able to contribute effectively to smaller ETL processes and understand the core components required for more complex data pipelines, making you an immediate asset to data engineering initiatives.

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