# Advanced Python Training for Data Scientists

## A 2-3 Day Intensive Course

### Target Audience

Data scientists with basic Python knowledge seeking to enhance their big data processing and software engineering skills.

### Schedule Overview

Day 1: Advanced Python Fundamentals & Big Data Basics

* 09:00 - 10:30: Advanced Python Features
* Decorators, generators, context managers
* Type hints and advanced OOP concepts
* 10:30 - 11:00: Coffee Break
* 11:00 - 12:30: Memory Management & Optimization
* Memory profiling
* Code optimization techniques
* 12:30 - 13:30: Lunch Break
* 13:30 - 15:00: Big Data Processing Fundamentals
* Parallel processing with multiprocessing
* Dask fundamentals
* 15:00 - 15:30: Coffee Break
* 15:30 - 17:00: Practical Workshop
* Implementing parallel data processing pipelines

Day 2: Advanced Data Processing & Best Practices

* 09:00 - 10:30: Advanced Pandas & NumPy
* Memory-efficient operations
* Chunked data processing
* 10:30 - 11:00: Coffee Break
* 11:00 - 12:30: Software Engineering Best Practices
* Unit testing
* Code documentation
* 12:30 - 13:30: Lunch Break
* 13:30 - 15:00: Version Control & CI/CD
* Advanced Git workflows
* GitHub Actions
* 15:00 - 15:30: Coffee Break
* 15:30 - 17:00: Project Structure & Packaging
* Creating maintainable Python packages
* Distribution best practices

Day 3 (Optional): Advanced Topics & Project Work

* 09:00 - 10:30: Cloud Computing Integration
* AWS/GCP basics for data processing
* 10:30 - 11:00: Coffee Break
* 11:00 - 12:30: Performance Monitoring
* Logging and monitoring
* Debugging techniques
* 12:30 - 13:30: Lunch Break
* 13:30 - 15:00: Practical Project
* Real-world data processing challenge
* 15:00 - 15:30: Coffee Break
* 15:30 - 17:00: Project Presentations & Wrap-up

### Learning Outcomes

* Master advanced Python features for efficient coding
* Implement scalable data processing solutions
* Apply software engineering best practices
* Create maintainable and documented code
* Deploy data processing pipelines effectively

### Prerequisites

* Basic Python programming knowledge
* Familiarity with Pandas and NumPy
* Basic understanding of version control (Git)

### Materials

* Jupyter notebooks with exercises
* Sample datasets
* Reference documentation
* Code templates

---