
PY401- Using Python for Big Data & Analytics

Course Duration: 3 Days;

WHAT YOU WILL LEARN

When it comes to data analytics or data science, its sheer complexity becomes a major concern. So much so that you might think you need a specialized programming language to handle such tasks. Of course, there are indeed programming languages that specialize in those fields. Yet, in most cases, data analytics professionals and data scientists prefer Python over these other languages.

Python has rapidly gained popularity in the IT community as a simple yet feature-rich language powering anything from simple web applications to the IoT, game development, and even artificial intelligence.

Big data and data analytics are another sector in which Python is currently making inroads. In this course, you will learn how Python is used in Big Data and Analytics.

AUDIENCE

- This course is for those who want to use Python for Big Data.

PREREQUISITES

- A basic understanding of Python language.
- A basic understanding of Data Analytics will be an added advantage.
- Basic Knowledge and skill in Web related technologies, especially HTML, XML, and JSON.
- Basic knowledge and skill in standard SQL statement will be added advantage.

METHODOLOGY

This program will be conducted with interactive lectures, PowerPoint presentation, discussion, and practical exercise.

COURSE OUTLINES

Module 1: Python refresher

- Development Environment
- Data types
- Conditionals
- Loops
- Sequences
- Mapping types
- Useful types from collections
- Program structure
- Functions, Modules, and packages

- Lambda functions
- Variable scope
- List comprehensions
- Generator expressions
- Creating modules
- Using the import statement
- Date Time module
- A Minimal Class in Python
- Attributes
- Methods
- The `__init__` method
- Static Methods
- Class Methods
- Properties - Getters and Setters

Module 2: Basic File I/O

- Open file for writing
- Reading Text Files

Module 3: Serializing data

- Reading/Writing CSV files
- Working with XML
- Reading/Writing JSON

Module 4: Getting Data from Web

- Fetching data from Restful API
- Reading data from Web Pages

Module 5: MySQL/MS SQL Database access

- Connecting to Server
- Connecting to a server
- Creating and executing a cursor
- Fetching data from Tables
- Parameterized statements

Module 6 – The NumPy Library

- Narray: The Heart of the Library
- Basic Operations
- Indexing, Slicing, and Iterating
- Conditions and Boolean Arrays
- Shape Manipulation
- Array Manipulation
- General Concepts
- Structured Arrays
- Reading and Writing Array Data on Files

Module 7 – The pandas Library—An Introduction

- The Python Data Analysis Library
- Testing Your pandas Installation
- Getting Started with pandas
- Introduction to pandas Data Structures
- Other Functionalities on Indexes
- Operations Between Data Structures
- Function Application and Mapping
- Sorting and Ranking
- "Not a Number" Data
- Hierarchical Indexing and Leveling

Module 8 – pandas: Reading and Writing Data

- I/O API Tools
- CSV and Textual Files
- Reading Data in CSV or Text Files
- Reading and Writing HTML Files
- Reading Data from XML
- Reading and Writing Data on Microsoft Excel Files
- JSON Data

Module 9 – pandas in Depth: Data Manipulation

- Data Preparation
- Concatenating
- Data Transformation
- Discretization and Binning
- Permutation
- String Manipulation
- Data Aggregation
- Group Iteration
- Advanced Data Aggregation

Module 10 – Data Visualization with matplotlib

- The matplotlib Library
- pyplot
- The Plotting Window
- Using the kwargs
- Adding Elements to the Chart
- Saving Your Charts
- Handling Date Values
- Chart Typology
- Line Charts
- Histograms
- Bar Charts
- Pie Charts
- Advanced Charts
- The mplot3d Toolkit
- Multi-Panel Plots

Module 11 – Better Visualization with Seaborn

- Using Seaborn with Matplotlib
- Customizing Seaborn Plots
- Color Palette
- Multiple plots with Seaborn
- Creating Different Types of Plots
- More Graphs in Seaborn