

# Python - Data Analysis II (Advanced Data Processing)

Course code: PYTHON\_DATAN2

During these 5 days, participants will have the opportunity to acquire basic and advanced knowledge in the field of data processing in Python using libraries such as numpy, pandas, modin, polars, vaex. It will be a combination of theoretical lectures and practical exercises to enable participants to effectively work with data and perform data analysis in Python. Interactive visualization of the analyzed data is an integral part of the training for each module taken over.

## Who is the course for

- Data Scientist, data analysts, especially in a Big Data environment, are the primary audience for this intensive course.
- Software developers who know the Python language at least at an intermediate and advanced level and who aim to create data-intensive applications using the SPARK engine in a Big Data (Cloud) environment.
- Data architects

## Required skills

- Knowledge of Python and Data Analysis at the PYTHON\_ADV and PYTHON\_DATAN course level

## Course outline

### Python Basics and Introduction to NumPy

- Introduction to Python as a programming language for data analysis
- Install and import the NumPy module
- Working with NumPy arrays and matrices
- Operations with NumPy arrays (addition, multiplication, indexing)
- NumPy statistical and mathematical functions

### Pandas - Data Manipulation

- Introduction to the Pandas library
- Loading and saving data in a Pandas DataFrame
- Work with data in DataFrame (selection, filtering, change)
- Group operations and data aggregation
- Merge and concatenate data frames
- Practical exercises with Pandas

### Parallel data processing with Modin

- Introduction to Modin - parallel data processing in Pandas
- Modin installation and configuration
- Speed comparison of Pandas and Modin on real data
- Practical exercises to optimize data processing

### Polars - Modern data processing

- Introduction to Polars - a modern library for data manipulation
- Comparison of Polars and Pandas
- Work with data in Polars DataFrame
- Analytical functions and SQL queries in Polars
- Practical exercises with Polars

### Vaex - Fast and efficient processing of big data

- Introduction to Vaex - a library for fast processing of big data
- Work with Vaex DataFrame
- Reading and writing large data files
- Performance optimization in Vaex
- Hands-on exercises on processing big data with Vaex

#### GOPAS Praha

Kodařská 1441/46  
101 00 Praha 10  
Tel.: +420 234 064 900-3  
[info@gopas.cz](mailto:info@gopas.cz)

#### GOPAS Brno

Nové sady 996/25  
602 00 Brno  
Tel.: +420 542 422 111  
[info@gopas.cz](mailto:info@gopas.cz)

#### GOPAS Bratislava

Dr. Vladimíra Clementisa 10  
Bratislava, 821 02  
Tel.: +421 248 282 701-2  
[info@gopas.sk](mailto:info@gopas.sk)



Copyright © 2020 GOPAS, a.s.,  
All rights reserved