

# Python - BigData and Hadoop

Course code: PYTHON\_BIGDATA

The concept of Big Data is now reflected in many areas and sectors of the national economy and therefore it is necessary not only to gain theoretical knowledge but also practical experience in processing large amounts of data.

## Who the course is for:

- Data Scientist, data analysts, especially in the Big Data environment, is the primary auditor for this intensive course.
- Software developers who are familiar with Python at least at intermediate and advanced levels and aim to create data-intensive allocations using the SPARK Big Data (Cloud) engine.
- Data architects

## Required skills:

- Basic knowledge of Python at the PYTHON\_INTRO level

## Teaching methods:

- Professional explanation with practical samples and examples.

## Teaching materials:

- Powerpoint handouts and module printouts.

## Course syllabus:

Hadoop Distributed File System (HDFS) and Python

- MapReduce technology in Python
- Pig and Python
- Spark / Yarn in Python
- Configure the virtual environment for Spark
- Creating Batch and Streaming Applications with Spark
- Juggle 's data using Spark
- Fundamentals of machine learning from data using Spark
- Streaming "live" data with Spark
- Data visualization
- Workflow Management in Python (Oozie)

**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)

 **GOPAS**®

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