

# Python - BigData a Hadoop

Kód kurzu: PYTHON\_BIGDATA

Pojem Big Data sa momentálne odráža v mnohých oblastiach a sektoroch národného hospodárstva, a preto je nutné nielen získať teoretické vedomosti, ale aj praktické skúsenosti so spracovaním veľkého množstva dát. Kurz predstavuje vývoj Big Data v prostredí programovacieho jazyka Python. Python je najdôležitejším jazykom nielen pre vývojárov aplikácií, analytikov dát, ale predovšetkým pre profesionálov v oblasti Data Science.

## Pre koho je kurz určený

- Data Scientist, dátoví analytici, najmä v Big Data prostredí sú primárnym auditóriom pre tento intenzívny kurz
- Software vývojári, ktorí ovládajú jazyk Python aspoň na strednej až pokročilej úrovni, a ktorí majú za cieľ vytvárať data-intenzívne aplikácie pomocou enginu SPARK v prostredí Big Data (Cloud)
- Dátoví architekti

## Predpokladané vstupné znalosti

Základná znalosť jazyka Python na úrovni kurzu PYTHON\_INTRO

## Metódy výuky

Odborný výklad s praktickými ukážkami, cvičenia na počítačoch

## Študijné materiály

Tlačené prezentácie preberanej látky

## Osnova kurzu

Hadoop Distribuovaný Súborový Systém (HDFS) a Python

- Technológie MapReduce v Python
- Jazyk Pig a Python
- Spark/Yarn v Python
- Konfigurácia virtuálneho prostredia pre Spark
- Tvorba Batch a streamovanie aplikácií so Spark
- „Žonglovanie“ s dátami pomocou Spark
- Základy strojového učenia z dát pomocou Spark
- Streamovanie "živých" dát so Spark
- Vizualizácia dát
- Workflow Management v Python (Oozie)

### GOPAS Praha

Kodaňská 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