Python for the Data Scientist

Duration 3 day(s) (PYTHON-DATA-SCIENCE)

Discovering the Python Environment for Data Science

Description

Python is an interpreted, multi-paradigm and multiplatform programming language. It is often appreciated, at first, for its clear syntax allowing an easy initiation to the basic concepts of programming. The strength of Python resides in these numerous specialized libraries developed by a community of contributors very active in various fields such as web development, scientific numerical computation, education, software development or business applications. Python is particularly used in the field of data science thanks to the power of its Numpy package (Numerical Python) and to the simplicity of its syntax which allows, for example, to communicate easily with big data (Apache Spark with PySpark) or deep learning (TensorFlow) environnements.

Goals

- Reviewing the basics of Python
- Mastering the basic data science packages
- Introduction to big data processing with PySpark
- Introduction to machine learning with Scikit-learn and deep learning with Tensorflow/Keras

Public

Analysts, Developers, Architects, Data Scientist

Prerequisites

Knowledge in programming

Structure

50% Theory, 50% Practice

Program

Python for data analysis

- Tools to code in Python (Anaconda, Jupyter)
- Data science ecosystem presentation
- Reminder of the basics of the language

Python and data

- The arrays with Numpy
- Structured data with Pandas
- Read/write data
- Data operations
- Statistics
- Preparation/cleaning data for machine learning

Data visualization with Python

- Building graphics with Matplotlib
- Advanced graphics with Seaborn
- Interactive graphics with Bokeh

Other tools

- Extract web data with BeautifulSoup
- Text analysis, text mining with NLTK

Introduction to the learning machine

- Preparation of data
- Supervised/unsupervised learning with Scikit-Learn
- Deep learning with Tensorflow/Keras

Introduction to big data processing

- Using Spark with pyspark
- Machine learning with spark.ml