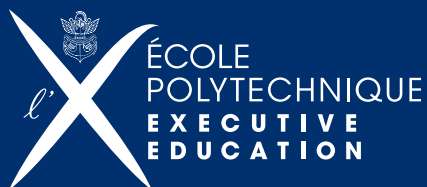


# DATA SCIENCE STARTER PROGRAM

2015



## A PROFESSIONAL TRAINING ON DATA SCIENCE AND BIGDATA, OFFERED BY ÉCOLE POLYTECHNIQUE

JOINTLY BY THE APPLIED MATHEMATICS AND INFORMATICS DEPARTMENTS

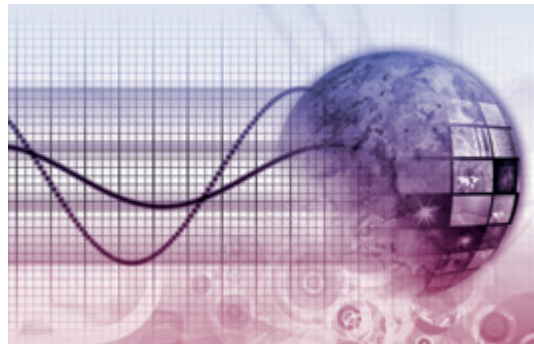
### TARGET AUDIENCE AND PREREQUISITE(S)

The proposed modules are suitable for anyone with some basic knowledge of Computer Science, Statistics and familiarity with computer programming. The program is designed for individuals (practitioners and researchers).

The concepts and training delivered in this program enable a sound understanding of the context and challenges of Big Data, a challenge that shapes the evolution of sciences and many business domains. The offered program is suitable to both early career professionals as well as senior managers that need an understanding of this challenging area and its applications.

### DATA SCIENCE STARTER PROGRAM

The training program aims at professionals and executives and covers taught modules, labs. It addresses state-of-the-art topics in Data Science and Big Data ranging from data collection, storage and processing to analytics and visualization, as well as a range of real-world applications and business/laboratory cases. This program is large-scope, and will cover the methods and tools with sufficient depth to allow the participants to really tackle big data problems.



## PROGRAM STRUCTURE

The training spans 120 hours taught (Friday and Saturday, in October, November and December). A typical day consists in a 3h course in the morning followed by a 3h labs (with hands on experience) and a one-hour conference by an invited expert from academia or industry. The courses cover to a sufficient degree the related to Data Science disciplines: Databases, Big Data tools, Data Preprocessing, Visualization, Data Analysis and Machine Learning. One fourth of the program is organized as a Data Camp in which the participants will work on real world datasets applying the full Data Science life cycle using the tools methods and knowledge they acquired in the courses.

### The program is organized as follows:

- Day 1. Data Science introduction. Big Data ecosystem, Data-project cycle/management, Computer architecture and introduction to distributed computing, Privacy issues.
- Days 2-4. Databases Big Data Tools. Databases: SQL and NoSQL, Distributed computing, Cloud computing, Map Reduce and Hadoop, HIVE/PIG, Spark.
- Days 5-7. Data preprocessing and Visualization. Data cleaning, Normalization, Feature selection and creation, Dimensionality Reduction, Data exploration, Visualization with R and Python, Browser based visualization, Mashup, Data Munging, Feature design & engineering.
- Day 8. Data Camp Part 1. Introduction to the data set and the objective of the camp. Exploration and feature selection/engineering.
- Days 9-12. Data Analysis and Machine Learning. Introduction to learning, Unsupervised learning, Supervised learning (Regression and feature selection, Logistic regression, Naive Bayes, KNN, SVM, Trees and Neural Nets), Learning evaluation, over fitting, Model selection, Ensemble methods.
- Days 13-14. Machine learning for Graph and Text data, with applications to Collaborative filtering, Web Recommendation and personalization, web advertising and marketing, social networks etc.
- Days 15-18. Data Camp Part 2: Students in small groups solve data challenges (different application domains) adopting the data science pipeline learned throughout the training. Data pre-processing, dimensionality reduction, bigdata technologies and Model learning, selection and ensemble will be employed.

## TEACHING STAFF

S. Gaiffas (X-CMAP), C. Giatsidis (X-LIX), B. Kegl (X, LAL), A. Papadopoulos (X, Aristotle U. of Thessaloniki) E. Le Pennec (X-CMAP), E. Matzner-Lober (X, U. Rennes) M. Vazirgiannis (X-LIX).

### More information available at:

<http://exed.polytechnique.edu/formations/11915/data-science-starter-program>

### Deadline for applications.

July 5, 2015. Submit your applications online at the above URL.

**CONTACT**

dssp@polytechnique.fr  
<http://exed.polytechnique.edu>

**PLACE**

École polytechnique

**PRICE**

7 500 euros

**DEADLINE OF APPLICATION**

July 5, 2015

**DATES OF PROGRAM**

From october 2, 2015  
to December 12, 2015



**ÉCOLE POLYTECHNIQUE**

91128 PALAISEAU CEDEX

[www.polytechnique.edu](http://www.polytechnique.edu)