

Hands-on course , 5  
day(s)  
Ref : BID

## Participants

Dataminers, statistical  
researchers, developers,  
project managers, business  
intelligence consultants.

## Pre-requisites

Basic knowledge of relational  
models, statistics, and  
programming languages.  
Basic knowledge of Business  
Intelligence concepts.

2018 Price : 3030€ excl.  
VAT

## Next sessions

### GENEVE

sep. 17 2018, dec. 03 2018

### LUXEMBOURG

sep. 17 2018, dec. 03 2018

# Big Data, practical methods and solutions for data analysis

*This course will enable you to understand the issues and benefits of Big Data as well as the technologies to implement it. You'll learn how to integrate massive volumes of structured and unstructured data via an ETL, then to analyze them using statistical models and dynamic dashboards.*

## OBJECTIVES

Understand the concepts and benefits of Big Data with respect to business challenges  
Understand the technological ecosystem needed to carry out a Big Data project  
Acquire the technical skills to manage massive, unstructured, complex data flows  
Implement statistical analysis models to address business needs  
Learn about a data visualization tool for reporting dynamic analyses

### 1) Understanding the concepts and challenges of Big Data

### 2) Big Data technologies

### 3) Managing structured and unstructured data

### 4) Data analysis methods for Big Data

### 5) Data visualization and concrete use cases

### 6) Conclusion

## Workshop

*Set up a Hadoop platform and its basic components, use an ETL to manage the data, create analysis modules and dashboards.*

## 1) Understanding the concepts and challenges of Big Data

- Origins and definition of Big Data: BI faced with the growth and diversity of data.
- Key figures in the international and French markets.
- The challenges of Big Data: ROI, organization, data privacy.
- An example of Big Data architecture.

## 2) Big Data technologies

- Description of the architecture and components of the Hadoop platform.
- Storage methods (NoSQL, HDFS).
- Operating principles of MapReduce.
- Overview of the main distributions on the market and additional tools (Hortonworks, Cloudera, MapR, Aster).
- Installing a Hadoop platform.
- Overview of Big Data specific technologies (Talend, Tableau, Qlikview, etc.).

### Exercise

*Installing a full Big Data platform via Cloudera and its components.*

## 3) Managing structured and unstructured data

- Operating principles of the Hadoop Distributed File System (HDFS).
- Importing outside data into HDFS.
- Creating SQL requests with HIVE.
- Using PIG to process the data.
- Using an ETL to industrialize the creation of massive data flows.
- Overview of Talend For Big Data.

### Exercise

*Implementing massive data flows*

## 4) Data analysis methods for Big Data

- Exploration methods.
- Segmentation and classification.
- Estimating and prediction.
- Implementing models.

### Exercise

*Setting up analyses with the software R.*

## 5) Data visualization and concrete use cases

- Offtheshelf visualization tools
- Report formatting methodology
- Benefits of Big Data for "Social Business".
- Measuring the reputation and fame of a brand.
- Measuring customer satisfaction and experience, optimizing the customer's path.

**Exercise**

*Installing and using a Data Visualization tool to create dynamic analyses, retrieving data from social media, and creating reputation analyses.*

**6) Conclusion**

- Takeaways.
- Summary of best practices.
- Bibliography.