

# AIBT1 – Big Data

From the Advanced Master AIBT

(Artificial Intelligence and Business Transformation)



## Highlights

- Data warehousing and visualisation
- Data quality management
- Introduction to Big Data processing
- Practical courses

Sources of data are heterogeneous, flow at different speeds and vary in volume. Defining a strong data integration framework is key to report on data quality and to efficiently explore and visualize their contents. Big Data can be characterized by the 3 Vs of Volume, Velocity and Variety and often requires dedicated computing solutions, which will be explored.

## Prerequisites

- General knowledge on computer science.
- Work experience of minimum 3 years.

## Key elements

Dates:

**4 November - 12 December 2019**

Duration: **56 hours, 8 days**

For whom:

**recent graduates, jobseekers and experienced employees**

Location:

**ISAE-SUPAERO, Toulouse**

Course fees: **4 000 €**

Language: **English**

## Learning objectives

After completing this course, participants will be able to:

- Explain the key components of ETL-based data warehousing;
- Set up indicators on data quality and management;
- Perform a simple data visualization task;
- Implement the distribution of simple operations via the Map/Reduce principle in Spark;
- Connect on a cloud computing engine (e.g. Google Cloud Platform) and launch a simple task;
- Deploy a Docker container.

## Practical information and registration

Natalia Perthuis - 05 61 33 80 47 – [info.exed@isae-supero.fr](mailto:info.exed@isae-supero.fr)

# AIBT1 – Big Data

From the Advanced Master AIBT

(Artificial Intelligence and Business Transformation)



## Course Content

### AIBT102 – Data integration and exploration (28h):

#### Data Warehousing:

- History and recent evolutions
- Extract-Transform-Load process
- Architecture
- Key functions
- Layers

#### Data quality:

- Indicators
- improvement
- assurance
- control

#### Data visualisation:

- visual perception
- effective graphical display
- tools

### AIBT103 – Big data processing (28h):

#### Distributed computing with Spark:

- History
- MapReduce paradigm
- Hadoop Stack
- Hadoop Distributed File System
- MLlib Machine Learning library

#### Virtualization and cloud computing:

- Different approaches to virtualization
- Economical models
- Technical benefits (snapshots, dynamic deployment and migration, failover...)
- cloud engines (principles, deployment examples, node choices)

#### Docker:

- History,
- Fundamental differences w.r.t. virtualization
- Docker components
- Tools