



# Big Data for All

Duration 1 day(s) (INTRO-BIGDATA)

Presentation of the Big DATA and get out of the buzz word

## Description

This training presents the different key stages of a Big Data project within the company.

## Goals

The objective is to define a common vocabulary on Big Data to the different interlocutors. A second objective is to present both organizational and technical challenges on Big Data projects.

### *Public*

Managers, Projects Director, Technical, Domain Experts

### *Prerequisites*

Nil

### *Structure*

70% Theory, 30% Practice

## Program

# Introduction

- Origin of Big Data
- Areas of application
- Success stories

# A Big Data Project

- Attention, it's not just a technical problem
- Let's create our first big data application
- Let's understand the different key steps
  - Acquisition
  - Cleaning
  - Consolidation
  - Aggregation
  - Storage
  - Analysis
  - Extraction
  - Publication

## Storing the data

- the different storage tools
- The NoSQL revolution, the different families

## Notebooks

# Data management

- Definition of a data
- Data processing in batch mode (Hadoop and spark)
- Streaming, kafka and flink
- The Datalake
  - The stakes and the different zones
  - Storage formats
  - Lambda and kappa architecture

# Big Data in the company

- Malfunctions
- The failure of exploratory approaches
- Governance and security
- Successfully adopt the datalake