



# FORMATION

## Big Data - Architecture et infrastructure

### En bref

**Durée:**

5 jours

**Audience:**

administrateurs systèmes, développeurs, architectes et chefs de projets techniques, consultants Business Intelligence.

**Pré-requis:**

avoir une connaissance pratique du Shell Linux, des bases du réseau, des bases de données relationnelles et des architectures réparties.

**Objectifs Pédagogiques:**

- Exploiter les architectures Big Data
- Concevoir l'intégration des flux de données

**Méthode Pédagogique:**

Alternance de cours magistraux (70%) et de travaux pratiques et cas d'usage (30%).

### Description

Le phénomène Big Data apparaît aujourd'hui comme le pétrole du 21ème siècle. Alimenté par l'utilisation massive des réseaux sociaux, le boom des objets connectés, l'adoption du quantified self et la transformation digitale des entreprises, le marché du big data ne cesse de croître. A l'horizon 2020, la valeur des données personnelles des consommateurs européens pourrait frôler les 1 000 milliards d'euros selon le Boston Consulting Group. Cette formation vous permettra d'envisager l'implémentation du big data dans votre entreprise.

### Plan de formation

*1ère journée :*

- I. Introduction au Big Data
  - A. Présentations formateur et participants
  - B. Principes et enjeux du Big Data
  - C. Les chiffres clés
  - D. Les enjeux (ROI/Confidentialité)
  - E. Concept d'architecture Big Data

*2ème journée :*

- II. Les technologies Big Data
  - A. Les outils du marché :
    1. MapR
    2. Les autres outils
      - a) Watson
      - b) Oracle Exalead
      - c) Microsoft
      - d) Teradata
      - e) Talend
      - f) Qlik Sense
      - g) AWS
      - h) ...
  - B. Les méthodes de stockage
  - C. Données chaudes
  - D. Données tièdes
  - E. Archivage



- III. Les langages:  
Python, Matlab, R, ...
- IV. Etat de l'art du Big Data
  - A. Les acteurs principaux
  - B. Les différents métiers du Big Data

### *3ème journée :*

- V. Architecture Hardware
  - 1. Stockage
  - 2. CPU/GPU
  - 3. Mémoire
  - 4. Réseau
  - 5. Systèmes distribués
- VI. Architecture Technique
  - A. La mise en qualité des données (Data Quality) : utilisation d'une exploitation d'un flux Open Data avec Talend Data Quality
  - B. Le stockage des données : Hadoop (HDFS), NoSQL (Cassandra, MongoDB)
  - C. La diffusion des données
    - 1. Streaming avec Apache Kafka, Amazon AWS Kinesis
  - D. Traitements en temps réel ou différé
    - 1. Apache Spark
    - 2. ELK
      - a) ElasticSearch
      - b) Logstash
      - c) Kibana
  - E. Les architectures réparties : clustering Hadoop
  - F. Supervision des plateformes : App Dynamix, Ambari

### *4ème journée :*

- VII. Architecture Applicative (étude de cas)
  - A. Les différentes étapes de gestion des données
  - B. L'ingestion depuis la source
  - C. La transformation



- D. La restitution / visualisation
- E. Analyse prédictive

## VIII. Atelier: Analyse client 360

*5ème journée :*

- IX. Applications Métiers (cas d'usages)
  - A. Text Analytics
  - B. Détection de fraude
  - C. Ciblage clientèle
  
- X. Risques du projet Big Data
  - A. Risques métiers liés au cadrage du scope
  - B. Risques stratégiques et manque de sponsoring
  - C. Les KPIs
  - D. La maturité des solutions du marché
  
- XI. Sécurité, éthique & enjeux juridiques
  - A. Assurer la protection des données
    - 1. L'anonymisation d'une donnée
    - 2. Le contrôle d'intégrité
    - 3. Le chiffrement d'une donnée
  - B. Qu'est-ce que la blockchain ?
  - C. Cas d'usage : Bitcoin
  
- XII. Conclusion (débrief global)