

## Programme de formation Big Data Mise en oeuvre

---

### ● Objectifs

Notre formation Big Data Mise en oeuvre couvre l'ensemble des processus nécessaires pour exploiter les ensembles de données massives. Vous découvrirez les différentes technologies pour collecter, stocker, traiter, analyser et présenter la « data », nouvelle ressource stratégique pour les entreprises. Lors des nombreux ateliers pratiques de la formation, vous apprendrez à mettre en oeuvre une plateforme complète pour gérer vos données. Travaillez avec Hadoop, Spark ou encore Talend pour la datavisualisation et soyez en mesure de sélectionner les bons outils Big Data selon vos besoins. Une introduction aux concepts d'apprentissage automatique (Machine Learning, Deep Learning) sera également présentée durant la formation.

---

### ● Pré requis

Connaissances sur les bases de données, en statistiques et développement (Java et SQL idéalement). Une familiarité avec les concepts BI. Il est également recommandé d'avoir suivi la formation Big Data Etat de l'art ou de posséder des connaissances équivalentes.

### ● Durée

5 jours

### ● Public

Administrateurs, Analystes, Chefs-de-projet, Consultants, Développeurs

---

### ● Plan de formation

#### Introduction à la formation Big Data Mise en oeuvre

Présentation générale et objectifs de cette formation Big Data  
Comprendre les enjeux et perspectives du Big Data  
Chiffres clés en France et à l'international  
La donnée en tant que matière première : « data is the new oil » ?  
Différents types, structures et sources de données  
Panorama des solutions et acteurs du marché : l'écosystème Big Data  
Les nouveaux métiers de la data  
Atelier : Présentation du déroulé de la formation et échanges sur les différents contextes des stagiaires.

#### Rappels des fondamentaux du Big Data

Big Data, Data Science, Machine Learning, etc

: définitions et concepts clés

Caractéristiques des données dites massives (les dimensions en V du Big Data)

Les architectures Big Data orientées stockage, calcul, temps réel...

Enjeux juridiques et éthiques liés à la collecte, au traitement et à l'analyse de données

Cycle de vie de la donnée

Assurer la qualité des données

Créer de la valeur à partir des données : vue d'ensemble d'un processus de Data Science

Atelier : Discussions sur les potentiels usages des technologies de Big Data suivant les secteurs d'activité.

#### Méthodologie pour l'implémentation d'un projet Big Data

Définir les besoins métiers et objectifs d'un projet

Spécificités d'un projet Big Data : incertitude,

budget, délais, gestion organisationnelle...  
Sélectionner les bons outils  
Poser les bonnes questions  
Obtenir les données  
Explorer et prétraiter les données  
Analyser les données : planifier et construire un modèle viable  
Communiquer les résultats  
Transformer les résultats en décisions et en actions  
Exemples de cas pratiques : Présentation du schéma d'architecture d'un projet Big Data et élaboration d'une première feuille de route.

## **Outils et technologies Big Data : démarrer avec Hadoop**

L'écosystème Apache Hadoop et les principales distributions (Hortonworks, Cloudera...)  
Clusters Hadoop, HDFS, YARN et MapReduce : comprendre les principes de fonctionnement  
Introduction aux composants HBase, Pig et Hive, Spark...  
Cas d'utilisation et exemples de mises en oeuvre de la plateforme Hadoop.  
Exemples de cas pratiques : Installation et configuration d'une plateforme complète Big Data. Prise en main des concepts de base de l'architecture Hadoop (NameNode, DataNode, Replications...).

## **Collecte, stockage et traitement des données**

Connaître les différentes sources de données (internes/externes, publiques/privées)  
Analyser les caractéristiques d'un jeu de données  
Principes ETL (Extract Transform Load) : présentation et prise en main de Talend  
Gérer les spécificités des données semi et non-structurées  
Rappels des principes du stockage distribué  
Intérêt et vue d'ensemble des bases de données NoSQL (Cassandra, Neo4j, MongoDB...)  
Prise en main d'HDFS et du modèle MapReduce  
Utiliser les fonctionnalités de Pig et Hive pour requêter et traiter les données  
ElasticSearch pour l'indexation et la recherche de données

Exemples de cas pratiques : Mises en pratiques des concepts, prise en main des différents outils pour importer, stocker et manipuler des ensembles de données.

## **Analyse des données issues du Big Data**

Vue d'ensemble des différentes méthodes et techniques d'analyse  
Notions d'analyse statistique et prédictive, l'impact de l'augmentation des volumes de données  
Analyser un ensemble de données avec l'environnement Apache Spark  
Comprendre les apports de la convergence entre Big Data et intelligence artificielle  
Introduction aux principes de Machine Learning  
Les différentes familles de modèles : régression, classification, clustering, renforcement learning, NLP  
Technique de feature engineering pour préparer les données  
Sélectionner, entraîner et tester un algorithme de Machine Learning  
Principes du Deep Learning : apprentissage profond et réseaux neuronaux  
Exemples de cas pratiques : Présentation de l'environnement Spark. Mise en oeuvre d'analyses de données avec des techniques de Machine Learning (langage R ou Python).

## **Datavisualisation : apprendre à communiquer sur les données**

Les enjeux et objectifs d'une bonne communication  
Principes fondamentaux de la représentation visuelle  
La Datavisualisation tout au long du processus Big Data  
Techniques et outils principaux pour la visualisation de données (Tableau, Qlik, etc)  
Sélectionner une solution de datavisualisation selon l'objectif recherché  
Connaître les bonnes pratiques de Dataviz : utilisation des couleurs, graphiques, tableaux de bord, infographies, notebooks...  
Visualisation interactive  
Exemple de cas pratique : Utilisation d'une solution de datavisualisation pour réaliser une présentation des résultats d'analyse.