



Big Data : mise en œuvre

Lien : <https://innov-systems.com/formation/big-data-mise-en-uvre>

 DURÉE
5 jours (35h)

 RÉFÉRENCE
BSI11

 CATÉGORIE
**Big Data :
fondamentaux et
architecture**

OBJECTIFS DE LA FORMATION

À l'issue de cette formation, vous serez capable de :

- ✓ Comprendre les spécificités propres aux projets Big Data
- ✓ Savoir piloter le cycle de vie des données (collecte, exploitation, analyse, visualisation)
- ✓ Mettre en œuvre une plateforme complète pour gérer vos données
- ✓ Travailler avec Hadoop, Spark et Talend pour la datavisualisation

POUR QUI ?

- ✓ Chefs de projet
- ✓ Data Scientists
- ✓ Data Analysts
- ✓ Développeurs
- ✓ Analystes et statisticien
- ✓ Toute personne en charge de la mise en oeuvre opérationnelle d'un projet Big Data en environnement Hadoop



Programme détaillé

1 / Introduction

- Les enjeux et perspectives du Big Data
- Marché data au niveau mondial
- Différents types de données
- Structures de données
- Sources de données
- Big Data : solutions et acteurs du marché
- Nouveaux métiers du Big Data

2 / Comprendre le Big Data

- Définitions et concepts clés
- Les dimensions en V du Big Data
- Les architectures Big Data orientées stockage, calcul, temps réel...
- Le cadre juridique des données
- Cycle de vie de la donnée
- La qualité des données
- Vue d'ensemble d'un processus de Data Science

3 / Projet Big Data : méthodologie et démarche de l'implémentation

- Les besoins métiers et objectifs d'un projet
- Spécificités d'un projet Big Data
- 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

4 / Technologies et outils associées au Big Data

- Architecture et composants de la plateforme Hadoop 2
- Les modes de stockage (NoSQL, HDFS)
- Fonctionnement de MapReduce et Yarn...
- Principales distributions Hadoop : Hortonworks, Cloudera, MapR...
- Les technologies émergentes : Spark, Storm, Machine Learning Azure...
- Installation d'une plateforme Hadoop
- Différentes des solutions de datavisualisation pour le Big Data (Talend, Tableau, Qlikview...)

5 / La collecte, le stockage des données et traitement des données

- Sources
- Analyser les caractéristiques d'un jeu de données
- Principes ETL (Extract Transform Load)
- Données structurées, semi structurées 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

6 / Outils et techniques pour l'analyse de données massives

- Différentes méthodes et techniques d'analyse
- Définition de l'analyse statistique
- Analyser un ensemble de données avec l'environnement Apache Spark
- L'Intelligence Artificielle dans l'analyse prédictive
- 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

7 / La visualisation des données (DATAVIZ)

- Enjeux et objectifs d'une bonne communication
- Parler les données, principes fondamentaux et objectifs de la dataviz
- Limites des outils et méthodes traditionnels (Excel, BusinessObjects, etc)
- 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
- Les bonnes pratiques de Dataviz
- Visualisation interactive

8 / Démonstration d'un environnement distribué HADOOP

- Les moteurs de recherche de grande envergure
- Présentation Lucene, Solr
- Mécanisme de traitement
- Gestion des journaux applicatifs
- Mécanismes et architecture Apache Flume
- La relation entre Cloud et Big Data
- Motivation entre clouds publics et privés
- Les clouds de stockage
- Se reconcentrer sur les problématiques métiers, grâce aux services managés

Approche pédagogique

- ✓ Support Ecrit et Projection
- ✓ Exposés Interactifs, Podcasts et Vidéos
- ✓ Brainstorming et Jeux de Rôle
- ✓ Cas Pratiques et Labs inclus pour leur impact opérationnel
- ✓ Test de Validation des Acquis des Connaissances

Prochaines dates programmées

 20 au 24 Juil. 2026

 Présentiel - Casablanca

 14 au 18 Sep. 2026

 Distanciel

 09 au 13 Nov. 2026

 Distanciel

 Autres dates possibles sur demande. Contactez-nous pour organiser une session intra-entreprise.

Réservation & Renseignements

 **Téléphone** : +212 522 247 210

 **Email** : contact@innov-systems.com

 **Web** : <https://www.innov-systems.com>