



# ORGANISME DE FORMATION AUX TECHNOLOGIES ET METIERS DE L'INFORMATIQUE

## Formation Big Data - Mise en oeuvre pratique d'une solution complète d'analyse des données

*Collecte et analyse de données massives*

N° ACTIVITÉ : 11 92 18558 92

TÉLÉPHONE : 01 85 77 07 07

E-MAIL : inscription@hubformation.com

Comprendre les besoins métier pour livrer aux décideurs des indicateurs fiables et pertinents, tel est le rôle attendu des spécialistes Big Data. Précisément conçu pour leur apprendre à mettre en oeuvre une solution de Big Data en environnement Hadoop, solution phare pour les traitements Big Data, ce programme reprend le cheminement logique d'un projet d'analyse de données. De leur collecte initiale à la mise en place de solutions de stockage spécifiques permettant d'organiser un très grand volume d'information, à la réalisation de scripts Pig et Hive qui, convertis en tâches MapReduce, permettent d'agréger et de filtrer les données pour finalement les analyser, tous les aspects seront abordés.

### Objectifs

- | mettre en oeuvre des analyses Big Data
- | Démontrer le cadre juridique du stockage et de l'analyse de données
- | utiliser des outils de collecte Open Source
- | choisir la bonne solution de stockage de données au regard des spécificités d'un projet (OLAP, NoSQL, graph)
- | explorer la boîte à outils technologique que constitue Hadoop et son écosystème
- | utiliser chaque brique (MapReduce, HIVE, SPARK...)
- | analyser les résultats et la signification des données extraites

### Public

- | 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

### Prérequis

- | Avoir suivi la formation Big Data - Les fondamentaux de l'analyse de données (BD540)
- | La connaissance d'un langage de programmation et du langage SQL est un plus pour suivre cette formation

### Programme de la formation

#### Introduction

- | Objectifs
- | Schématisation du projet
- | Écosystème et stack technologique
- | Résultats attendus

#### Ingestion de données massives

- | Description
- | Caractéristiques clés des outils d'ingestion
- | Solutions (packagées, cloud computing et Hadoop)

Référence	BIG550
Durée	4 jours (28h)
Tarif	2 750 €HT

### SESSIONS PROGRAMMÉES

#### A DISTANCE (FRA)

- du 30 juin au 3 juillet 2025
- du 15 au 18 septembre 2025
- du 3 au 6 novembre 2025

#### PARIS

- du 30 juin au 3 juillet 2025
- du 15 au 18 septembre 2025
- du 3 au 6 novembre 2025

[VOIR TOUTES LES DATES](#)

- | Focus Apache NIFI
- | Focus Apache KAFKA
- | Ingestion de données en streaming NIFI on KAFKA
- | Réalisation d'un workflow NIFI d'ingestion de donnée streaming dans HDFS

### **Traitement de données Big Data en batch**

- | Diagramme de fonctionnement
- | Solutions logicielles associées (MapReduce, Hive, Spark, ...)
- | Big Data Batch scripting
- | Data Warehousing Big Data
- | Création de tables partitionnées, clusterisées et/ou transactionnelles
- | Écriture des scripts d'updates des différentes tables
- | Analytics Big Data

### **Traitement avancé Big Data : l'apprentissage machine**

- | L'écosystème Spark
- | Schéma général de création d'un modèle de ML
- | Création d'un modèle de clusterisation de données
- | Création d'un modèle d'analyse prédictive supervisé
- | Enrichissement d'un DWH avec Spark
- | Sauvegarde d'un modèle d'apprentissage machine
- | Application d'un modèle ML

### **Stockage de données distribuées**

- | Principes des bases de données distribuées
- | Solutions (NoSQL, NewSQL)
- | Création d'une base de données distribuée
- | Ingestion de données dans une base de données distribuées
- | Interrogation d'une base de données distribuées

### **Automatisation de chaîne de traitement Batch**

- | L'orchestrateur Oozie
- | Ordonnancement de scripts HIVE
- | Combinaison avec des scripts SPARK
- | Création d'un workflow Oozie complet

### **Traitement de données massives en flux (streaming)**

- | Principe de fonctionnement
- | Solutions logicielles
- | Le prétraitement de données en streaming
- | L'inscription de streams à un Hub Streaming
- | La consommation de streams auprès d'un Hub
- | Le traitement avancé de données en flux (machine learning)

### **Mise en oeuvre dans une architecture Big Data**

- | Les approches standards
- | Réalisation d'une solution complète de traitement de données type Lambda
- | Réalisation d'une solution complète de traitement de données type Kappa

## **Méthode pédagogique**

Une formation très opérationnelle durant laquelle s'alternent les phases d'apports théoriques, d'échanges, de partage d'expériences. Les participants réalisent sur un cas concret une chaîne de traitement Big Data (Batch et Streaming). Ils travaillent dans un environnement Cloud sur une des distributions Hadoop les plus utilisées. Les contenus digitaux mis à disposition des stagiaires avant et après la formation renforcent l'efficacité pédagogique du programme et garantissent un bénéfice durable de l'action de formation.

## **Méthode d'évaluation**

Tout au long de la formation, les exercices et mises en situation permettent de valider et contrôler les acquis du stagiaire. En fin de formation, le stagiaire complète un QCM d'auto-évaluation.

---

## **Suivre cette formation à distance**

Voici les prérequis techniques pour pouvoir suivre le cours à distance :

- | Un ordinateur avec webcam, micro, haut-parleur et un navigateur (de préférence Chrome ou Firefox). Un casque n'est pas nécessaire suivant l'environnement.
- | Une connexion Internet de type ADSL ou supérieure. Attention, une connexion Internet ne permettant pas, par exemple, de recevoir la télévision par Internet, ne sera pas suffisante, cela engendrera des déconnexions intempestives du stagiaire et dérangera toute la classe.
- | Privilégier une connexion filaire plutôt que le Wifi.
- | Avoir accès au poste depuis lequel vous suivrez le cours à distance au moins 2 jours avant la formation pour effectuer les tests de connexion préalables.
- | Votre numéro de téléphone portable (pour l'envoi du mot de passe d'accès aux supports de cours et pour une messagerie instantanée autre que celle intégrée à la classe virtuelle).
- | Selon la formation, une configuration spécifique de votre machine peut être attendue, merci de nous contacter.
- | Pour les formations incluant le passage d'une certification la dernière journée, un voucher vous est fourni pour passer l'examen en ligne.
- | Pour les formations logiciel (Adobe, Microsoft Office...), il est nécessaire d'avoir le logiciel installé sur votre machine, nous ne fournissons pas de licence ou de version test.
- | Horaires identiques au présentiel.

---

## Accessibilité



Les sessions de formation se déroulent sur des sites différents selon les villes ou les dates, merci de nous contacter pour vérifier l'accessibilité aux personnes à mobilité réduite.  
Pour tout besoin spécifique (vue, audition...), veuillez nous contacter au 01 85 77 07 07.