



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

## Formation Big Data - Architectures *Panorama des solutions*

N° ACTIVITÉ : 11 92 18558 92

TÉLÉPHONE : 01 85 77 07 07

E-MAIL : inscription@hubformation.com

### Objectifs

| Identifier les concepts essentiels du Big Data et les technologies implémentées  
| Savoir analyser les difficultés propres à un projet Big Data, les freins et les apports, tant sur les aspects techniques que sur ceux liés à la gestion du projet

### Public

| Chefs de projets, architectes, développeurs, data scientists et toute personne souhaitant connaître les outils et solutions pour concevoir et mettre en oeuvre une architecture Big Data

### Prérequis

| Posséder une bonne culture générale des systèmes d'information

### Programme de la formation

#### Introduction

| L'essentiel du Big Data : calcul distribué, données non structurées  
| Besoins fonctionnels et caractéristiques techniques des projets  
| La valorisation des données  
| Le positionnement respectif des technologies de Cloud, Big Data et NoSQL et les liens, implications  
| Concepts clés : ETL, Extract Transform Load, CAP, 3V, 4V, données non structurées, prédictif, Machine Learning  
| Exemple d'application : Amazon Rekognition, Polly, EMR  
| L'écosystème du Big Data : les acteurs, les produits, état de l'art  
| Cycle de vie des projets Big Data  
| Emergence de nouveaux métiers : Data scientists, Data labs, Hadoop scientists, CDO, ...  
| Rôle de la DSI dans la démarche Big Data  
| Gouvernance des données : importance de la qualité des données, fiabilité, durée de validité, sécurité des données  
| Aspects législatifs : sur le stockage, la conservation de données, etc ... sur les traitements, la commercialisation des données, des résultats

#### Stockage distribué

| Caractéristiques NoSQL  
| Les différents modes et formats de stockage  
| Les types de bases de données : clé/valeur, document, colonne, graphe  
| Besoin de distribution  
| Définition de la notion d'élasticité  
| Principe du stockage réparti  
| Définitions : réplication, sharding, gossip, hachage  
| Systèmes de fichiers distribués : GFS, HDFS, Ceph  
| Les bases de données : Redis, Cassandra, DynamoDB, Accumulo, HBase, MongoDB, BigTable, Neo4j, ...  
| Données structurées et non structurées, documents, images, fichiers XML, JSON, CSV, ...  
| Moteurs de recherche

Référence	BIG-ARC
Durée	2 jours (14h)
Tarif	1 690 €HT

### SESSIONS PROGRAMMÉES

#### A DISTANCE (FRA)

du 1er au 2 juillet 2025  
du 30 sept. au 1er octobre 2025  
du 18 au 19 novembre 2025

#### PARIS

du 1er au 2 juillet 2025  
du 30 sept. au 1er octobre 2025  
du 18 au 19 novembre 2025

[VOIR TOUTES LES DATES](#)

- | Principe de fonctionnement
- | Méthodes d'indexation
- | Recherche dans les bases de volumes importants
- | Présentation d'Elasticsearch et SolR
- | Principe du schemaless, schéma de stockage, clé de distribution, clé de hachage

### Calcul et restitution, intégration

- | Différentes solutions : calculs en mode batch, ou en temps réel, sur des flux de données ou des données statiques
- | Les produits : langage de calculs statistiques, R Statistics Language, sas, RStudio
- | Outils de visualisation : Tableau, QlikView
- | Ponts entre les outils statistiques et les bases Big Data
- | Outils de calcul sur des volumes importants : Kafka/Spark Streaming/Storm en temps réel, Hadoop/Spark en mode batch
- | Zoom sur Hadoop : complémentarité de HDFS et YARN
- | Restitution et analyse : logstash, kibana, elk, zeppelin
- | Principe de map/reduce et exemples d'implémentations, langages et sur-couches
- | Présentation de Pig pour la conception de tâches map/reduce sur une grappe Hadoop

## Méthode pédagogique

Chaque participant travaille sur un poste informatique qui lui est dédié. Un support de cours lui est remis soit en début soit en fin de cours. La théorie est complétée par des cas pratiques ou exercices corrigés et discutés avec le formateur. Le formateur projette une présentation pour animer la formation et reste disponible pour répondre à toutes les questions.

## 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.