

PUBLIC

Directeurs SI – Responsables SI –
Chefs de projets – Architectes –
Consultants – Toute personne
amenée à participer à un projet Big
Data

PRÉ-REQUIS

Avoir des connaissances de base
des architectures techniques et du
fonctionnement d'un système de
gestion de base de données (SGBD)

NIVEAU D'EXPERTISE

Fondamentaux

LES POINTS FORTS

Docaposte Institute propose
plusieurs dispositifs pédagogiques
adaptés aux apprenants :

Formation en présentiel

En groupe (inter-entreprises ou
intra-entreprise)

En individuel (monitorat)

En journée ou en cours du soir (sur
demande spécifique)

Formation en distanciel

Distanciel synchrone

Distanciel asynchrone

MOYENS PÉDAGOGIQUES

- Apports des connaissances communes.
- Mises en situation sur le thème de la formation et des cas concrets.
- Méthodologie d'apprentissage attractive, interactive et participative.
- Equilibre théorie / pratique : 60 % / 40 %.
- Supports de cours fournis au format papier et/ou numérique.
- Ressources documentaires en ligne et références mises à disposition par le formateur.
- Pour les formations en présentiel dans les locaux mis à disposition, les apprenants sont accueillis dans une salle de cours équipée d'un réseau Wi-Fi, d'un tableau

Code
701984

Durée
**Sur mesure / 7
heures**

Tarif
Nous contacter

**Repas inclus (en présentiel)*

Objectifs pédagogiques

- Explorer les concepts clés du Big Data
- Comprendre les avantages et les limites du Big Data
- Comprendre les enjeux économiques du Big Data
- Connaître l'écosystème Big Data et comprendre les technologies impliquées
- Savoir anticiper son intégration dans les opérations informatiques de l'entreprise
- Tenir compte de la confidentialité et de la sécurité des données lors de l'extraction de Big Data

Programme de la formation

Présentation de l'Ecosystème Big Data

- Définition, enjeux et perspectives
- Qu'est-ce que le Big Data et ses quatre dimensions : volume, vitesse, variété, véracité
- Quels sont les enjeux sociétaux et économiques : exemples d'amélioration de la performance et des résultats de l'entreprise grâce au Big Data
- Mesurer l'importance du Big Data au sein d'une entreprise : les enjeux, la place du Big data dans le-commerce, le-marketing et le-réputation
- Réussir à extraire des données utiles
- Intégrer le Big Data aux données traditionnelles
- Intégrer les critères légaux et éthiques : que peut-on collecter, stocker et analyser
- Intégrer les impacts organisationnels : apparition de nouveaux rôles / métiers

Panorama des techniques liées au Big Data

- Acquisition de la donnée
- L'enchaînement des opérations.
- Le recueil des données : crawling, scraping
- La gestion de flux événementiel (Complex Event Processing, CEP)

blanc ou paperboard. Un ordinateur avec les logiciels appropriés est mis à disposition (le cas échéant).

SATISFACTION ET EVALUATION

- En amont de la formation
 - ▶ Recueil des besoins des apprenants afin de disposer des informations essentielles au bon déroulé de la formation (profil, niveau, attentes particulières...).
 - ▶ Auto-positionnement des apprenants afin de mesurer le niveau de départ.
- Tout au long de la formation
 - ▶ Évaluation continue des acquis avec des questions orales, des exercices, des QCM, des cas pratiques ou mises en situation...
- A la fin de la formation
 - ▶ Auto-positionnement des apprenants afin de mesurer l'acquisition des compétences.
 - ▶ Evaluation par le formateur des compétences acquises par les apprenants.
 - ▶ Questionnaire de satisfaction à chaud afin de recueillir la satisfaction des apprenants à l'issue de la formation.
 - ▶ Questionnaire de satisfaction à froid afin d'évaluer les apports ancrés de la formation et leurs mises en application au quotidien.

ACCOMPAGNEMENT FORMATION À DISTANCE

En cas de nécessité, une assistance technique et pédagogique est joignable entre 8h30 et 18h (jours ouvrés):

- par téléphone : 01 83 10 10 10
- par mail : care-formation@lefebvre-dalloz.fr

Une réponse immédiate est apportée ; si besoin, le demandeur est mis en relation avec un expert dans un délai maximum de 48h.

- L'indexation du flux entrant
- L'intégration avec les anciennes données
- La qualité des données : un cinquième V ?
- Les différents types de traitement : recherche, apprentissage (machine learning, transactionnel, data mining)
- D'autres modèles d'enchaînement : Amazon, e-Santé
- Un ou plusieurs gisements de données ? De Hadoop à l'in-memory
- De l'analyse de tonalité à la découverte de connaissances.

Stockage de donnée du Big Data

- Analyser les caractéristiques de vos données
- Sélectionner les sources de données à analyser
- Définir le rôle et les caractéristiques des bases NoSQL
- Présenter un entrepôt Big Data
- Modèles de données : valeur clé, graph, document, famille de colonnes
- Système de fichiers distribués Hadoop (HDFS)
- Un éventail de bases de données : HBase, Cassandra, BigTable, DynamoDB, MongoDB, Redis, Riak, Neo4J
- Solution de recherche : Elastic Search
- Comment bien choisir un entrepot de données
- Choisir un entrepôt de données en fonction des caractéristiques de vos données
- Injecter du code dans les données, mettre en œuvre des solutions de stockage des données multilingues
- Choisir un entrepôt de données capable de s'aligner sur les objectifs de l'entreprise

Traitement du Big Data

- Intégrer différents entrepôts de données digitales
- Mapper les données avec le framework de programmation, se connecter aux données et les extraire de l'entrepôt de stockage, transformer les données à traiter
- Fractionner les données pour Hadoop MapReduce

Présentation d'autres Framework

- SPARK: l'alternative à Hadoop MapReduce
- KAFKA: le middleware de message distribué

Mise en œuvre et élaboration d'une stratégie dédiée au Big Data

- Définir les besoins en matière de Big Data
- Atteindre les objectifs grâce à la pertinence des données
- Évaluer les différents outils du marché dédiés au Big Data
- Répondre aux attentes du personnel de l'entreprise



Une méthode analytique innovante

- Identifier l'importance des traitements métier
- Cerner le problème
- Choisir les bons outils
- Obtenir des résultats exploitables

Mettre en œuvre une solution Big Data

- Bien choisir les fournisseurs et options d'hébergement
- Trouver le juste équilibre entre les coûts engendrés et la valeur apportée à l'entreprise
- Garder une longueur d'avance

Mesurer les enjeux de la protection des données

- Intégrer le droit comme outil de valorisation des bases de données
- Gérer la paternité des fichiers et des données
- Assurer la sécurité informatique des fichiers de données
- Assurer la sécurité juridique des fichiers de données : Cloud Computing
- Étude de cas : analyse de la protection des données de l'entreprise
- Mesurer les intérêts du big data et ses interactions avec le droit
- Les données issues de l'open data : droit sur les données, obligations et coût
- La réutilisation des données de l'open data : la licence
- Étude de cas : analyse de la licence proposée par ETATLAB
- Identifier les spécificités des données à caractère personnel Concevoir des systèmes d'information et des traitements conformes (privacy by design) via l'identification des critères d'une collecte et d'un traitement légal des données
- Appliquer les exigences de la qualité Informatique et libertés : proportionnalité, conservation, interconnexion, archivage
- L'information des personnes et les limites du détournement de finalité
- Assurer le droit des personnes sur leurs données
- Mesurer les risques posés par une exploitation non conforme à la loi Informatique et libertés
- Étude de cas : analyse des dernières sanctions de la CNIL

A noter

...

En amont et en aval de la formation, le positionnement pédagogique sera effectué à l'aide d'un questionnaire d'auto-positionnement.