Big Data – Les fondamentaux de l'analyse des données

INFORMATIONS GÉNÉRALES

Type de formation : Formation continue

Domaine : IA, Big Data et Bases de données

Filière : Big Data

Rubrique : Fondamenteaux

Éligible au CPF : Non

Action collective : Non

PRÉSENTATION

Objectifs & compétences

- Comprendre le rôle stratégique de la gestion des données pour l'entreprise
- Identifier ce qu'est la donnée, et en quoi consiste le fait d'assurer la qualité de données
- Synthétiser le cycle de vie de la donnée
- Assurer l'alignement des usages métiers avec le cycle de vie de la donnée
- Découvrir les bonnes pratiques en matière de contrôle de qualité des données
- Assurer la mise en œuvre de la gouvernance de la donnée

Public visé

MOA, Chef de projet, Urbaniste fonctionnel, Responsable de domaine, Analystes, Développeurs, Data Miners … Futurs Data Scientists, Data Analysts et Data Stewards

Pré-requis

Si aucune connaissance technique particulière n'est nécessaire, il est toutefois recommandé d'avoir suivi le module «Big Data - Enjeux et perspectives » (BD004) pour suivre cette formation dans des conditions optimales

PROGRAMME

INTRODUCTION
Les origines du Big Data
La donnée en tant que matière première
La connaissance de la question Big Data, Données, qualité et stratégie d'entreprise
Problématiques d'alignement de la qualité de la donnée avec les usages métiers
Les différentes sources de données de l'entreprise, de l'Internet, des objets connectés
Les différentes formes d'exploitation de données
Système d'information opérationnel
Système d'information décisionnel
Big Data et smart Data

LA COLLECTE DE DONNÉES
Où et comment collecter des données ?
Les sources de données, les API, les fournisseurs, les agrégateurs…
Les principaux outils de collecte et de traitement de l'information (ETL)
Les particularités de la collecte des données semi-structurées et non-structurées
LE STOCKAGE DES DONNÉES
Les différentes formes de stockage des données : rappel de l'architecture relationnelle de stockage des données transactionnelles (SGBD/R) et multidimensionnelles (OLAP)
Prise en main d'une base de données OLAP
Les nouvelles formes de stockage des données
Compréhension, positionnement et comparaison : Bases NoSQL, Hadoop, Spark, Bases de données graph…
Panorama des bases de données NoSQL
Particularités liées au stockage des données non-structurées
Comment transformer des données non structurées en données structurées
L'ÉCOSYSTÈME HADOOP

Code de formation : BD007

€ Tarifs

Prix public : 2390 €

Tarif & financement :
Nous vous accompagnons pour trouver la meilleure solution de financement parmi les suivantes :
Le plan de développement des compétences de votre entreprise : rapprochez-vous de votre service RH.
Le dispositif FNE-Formation.
L'OPCO (opérateurs de compétences) de votre entreprise.
France Travail: sous réserve de l'acceptation de votre dossier par votre conseiller Pôle Emploi.
CPF -MonCompteFormation
Contactez nous pour plus d'information : contact@aston-institut.com

Lieux & Horaires

Campus : Ensemble des sites

Durée : 21 heures

Délai d'accès : Jusqu'a 8 jours avant le début de la formation, sous condition d'un dossier d'insciption complet

Distanciel possible : Oui

Prochaines sessions

Cliquez sur la date choisie pour vous inscrire :

- 23 / 06 / 2025
  : Ensemble des sites
  ✓ : Distanciel possible
  : 21 heures
  : 3 jours
- 08 / 09 / 2025
  : Ensemble des sites
  ✓ : Distanciel possible
  : 21 heures
  : 3 jours
- 13 / 10 / 2025
  : Ensemble des sites
  ✓ : Distanciel possible
  : 21 heures
  : 3 jours
- 17 / 11 / 2025
  : Ensemble des sites
  ✓ : Distanciel possible
  : 21 heures

Informations non contractuelles - PDF généré le 12/11/2026

Présentation des principaux modules de la distribution
Apache Hadoop
Présentation et comparaison des principales distributions commerciales (Cloudera, Hortonworks...)
L'infrastructure matérielle et logicielle nécessaire au fonctionnement de Hadoop
Serveur local ou cloud
Les concepts de base de l'architecture Hadoop: Data Node, Name Node, Job Tracker, Task Tracker
Présentation de HDFS (Système de gestion des fichiers de Hadoop)
Présentation de MapReduce (Outil de traitement de Hadoop)
Les commandes exécutées au travers de PIG
Présentation de HIVE pour transformer du SQL en MapReduce
L'ANALYSE DE DONNÉES
Comment requêter les données ?
Analyser et comprendre la signification des données extraites
Particularités liées à l'analyse des données non structurées
Analyse prédictive : transformer des données du passé en prévisions pour le futur
Calculer des tendances
Machine Learning : les bases de l'apprentissage machine
Deep Learning : notions de base de l'analyse future automatisée de données non structurées
TRANSFORMER LES DONNÉES EN DÉCISIONS
Comprendre les besoins et les attentes des utilisateurs business
Traduire les demandes des utilisateurs en requêtes
Évaluer et vérifier la qualité des données extraites en fonction des résultats obtenus
Définir un indice de confiance permettant d'échanger avec les utilisateurs business

: 3 jours

MODALITÉS

Modalités

Modalités : en présentiel, distanciel ou mixte . Toutes les formations sont en présentiel par défaut mais les salles sont équipées pour faire de l'hybride. – Horaires de 9H à 12H30 et de 14H à 17H30 soit 7H – Intra et Inter entreprise.
Pédagogie : essentiellement participative et ludique, centrée sur l'expérience, l'immersion et la mise en pratique. Alternance d'apports théoriques et d'outils pratiques.
Ressources techniques et pédagogiques : Support de formation au format PDF ou PPT Ordinateur, vidéoprojecteur, Tableau blanc, Visioconférence : Cisco Webex / Teams / Zoom.
Pendant la formation : mises en situation, autodiagnostics, travail individuel ou en sous-groupe sur des cas réels.

Méthode

Fin de formation : entretien individuel.
Satisfaction des participants : questionnaire de satisfaction réalisé en fin de formation.
Assiduité : certificat de réalisation.
Validations des acquis : grille d'evalution des acquis établie par le formateur en fin de formation.

Informations non contractuelles - PDF généré le 12/11/2026