

Introduction au Big Data et Machine Learning

Détails

- **Code** : DB-BIGML
- **Durée** : 2 jours (14 heures)

Public

- Consultants
- Consultants IT
- Data Analysts
- Data Scientist

Pré-requis

Objectifs

- Présenter les enjeux de l'analyse des données numériques à grande échelle.
- Comprendre le concept de valorisation de la donnée.
- Maîtriser le vocabulaire lié au big data et à la data science
- Comprendre la problématique de traitement de données massives à des fins d'analyse.
- Identifier les risques et opportunités du Big Data pour les individus (vie privée, protection des données personnelles...).

Programme

Comprendre les concepts et les enjeux du Big Data

- Découvrir les concepts du Big Data
- Définition d'un projet Big Data
- Big data en entreprise
 - Principes de fonctionnement
 - Les secteurs impactés
 - Comment se reconvertir dans le big data ?
 - La gouvernance des données
- Les spécificités d'un projet big data
- Les enjeux du Big Data : organisation, confidentialité des données, GPDR

Les technologies du Big Data

- Business Intelligence vs Big Data
- Architecture et composants de la plateforme Hadoop
- Les modes de stockage
 - Le stockage fichiers : Hadoop Distributed File System (HDFS)
 - Stockage colonne, graph, clé/valeur et document :
- Paradigme MapReduce
- Spark : framework de calcul distribué in memory
- Les architectures big data
- Les technologies spécifiques pour le Big Data (Tableau, Talend, Qlikview ...)

Atelier

- Installation d'une plateforme Big Data Hadoop
- Implémentation de flux de données massives

Data Science : Méthodes d'analyse et traitements des données pour le Big Data

- Les bases de l'apprentissage Machine (Machine Learning)
- Prétraitement des données
- La réduction de dimensions
- Les types des modèles du machine learning : Clustering, Classification, Regression
- L'évaluation des modèles
- Les méthodes d'ensemble
- Le langage Python pour le Big Data et le Machine Learning : Pandas, Scipy, Scikit-learn, TensorFlow
- Communiquer ses résultats à l'aide de représentations graphiques lisibles et pertinentes

Ateliers Données

- Visualisation des données
- Exploration et sélection/ingénierie des variables
- Application des méthodes d'apprentissage automatique
- Sélection de modèles et méthodes d'ensemble

Modalités

- **Type d'action** :Acquisition des connaissances
- **Moyens de la formation** :Formation présentielle – 1 poste par stagiaire – 1 vidéo projecteur – Support de cours fourni à chaque stagiaire
- **Modalités pédagogiques** :Exposés – Cas pratiques – Synthèse
- **Validation** :Exercices de validation – Attestation de stages

