SQLI INSTITUT

400 FORMATIONS
POUR FAIRE LA DIFFÉRENCE

BIG DATA – LES FONDAMENTAUX DE L'ANALYSE DES DONNÉES

CODE STAGE : BD007

OBJECTIFS

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 ?uvre de la gouvernance de la donnée

DURÉE

3 jours

PUBLIC

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
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