

Formation 1 initiation big data (2 jours)

A qui s'adresse cette formation ?

- Débutant en informatique
- Jeunes diplômés
- Personne en reconversion professionnelle
- Consultants en Informatique Décisionnelle

Prérequis

- Aucun

Objectifs

- Comprendre les concepts de la Big Data
- Être capable d'identifier l'écosystème big data et comprendre les technologies associées
- Savoir anticiper la mise en place et l'intégration des activités informatiques de l'entreprise

Langue d'enseignement

- Français

Les points forts de la formation

- Les formateurs sont des experts en Big Data ils interviennent régulièrement en entreprise sur des sujets big data.

Contenus

- **Définition du Big Data selon les grands acteurs du marché**
 - Caractéristiques techniques des 3V de Gartner (Vélocité, Variété et Volume) et les variantes (Véracité, Valeur, Validité...)
 - Big Data : pourquoi maintenant ?
 - Traitement des données structurées, semi-structurées et déstructurées
 - Transformation des données massives en informations utiles et en valeur
 - Gestion des données en cycles, de l'acquisition à la gouvernance. Use Case et stratégies (GAFA (Google, Apple, Facebook, Amazon), exemples santé, grande distribution, bancassurance...)
 - Les grands acteurs et le marché du Big Data

- **Description des technologies de référence du Big Data**
 - NoSQL (Not Only SQL), les nouvelles compétences attendues (Python, R, Scala, Java ...)
 - Hadoop : un modèle d'open source du Big Data adopté par les grands acteurs de l'informatique (IBM, Oracle, Amazone, EMC, Google...)
 - Principaux composants d'Hadoop : HDFS (Hadoop Distributed File System), MapReduce, Pig, Flume, Zookeeper, H-BASE, Lucene, Hive, Cloudera, Oozie, Cassandra, Machine Learning...
 - Autres solutions : Docker, OpenStack, Elastic, Splunk...
 - Nouvelles architectures techniques pour traiter des données massives et non-structurées, en temps réel (SPARK)

- **Déploiement et utilisation du Big Data**
 - Description d'une plate-forme de Big Data et bonnes pratiques
 - Intégration des données et systèmes existants à la plate-forme Big Data
 - Acquisition et qualification des données ouvertes des organisations publiques et sociales (Open Data)
 - Traitement des données à la volée (Data Streaming)
 - Analyse de données (Data Analytics et Business Intelligence)
 - Présentation des informations (Data Visualisation)

- **Une journée découverte au cœur des écosystèmes Hadoop et Spark**
 - Introduction générale
 - Introduction à Apache Hadoop et l'écosystème Hadoop
 - ✓ Aperçu sur Apache Hadoop
 - ✓ Ingestion et stockage des données
 - ✓ Analyse des données et exploration
 - ✓ Autres outils de l'écosystème
 - Apache Hadoop et Stockage de fichiers
 - ✓ Composants d'un cluster Apache Hadoop
 - ✓ Architecture HDFS
 - Traitement distribué sur un cluster Apache Hadoop
 - ✓ Architecture Yarn
 - ✓ Prise en main sur Yarn
 - Les principales briques et leur rôle
 - PIG, Hive, Oozie, Flume, HDFS, MapReduce, Sqoop, Zookeeper, Hbase, Mahout, les connecteurs ...
 - SparkML, SparkQL, MLib, Kafka, Storm, Flink , Beam, ...
 - Démonstrations et discussions

Formation 2 Développeur Big Data (5 jours)

A qui s'adresse cette formation ?

- Jeunes diplômés
- Consultants en informatique décisionnelle
- Consultants en Mission
- Développeur java/python

Prérequis

- Maîtrise des bases du développement JAVA/PYTHON, SHELL
- Connaissance du SQL

Objectifs

- Apprendre à analyser des données non structurées
- Disposer d'une vision claire du Big Data, de ses enjeux et de son écosystème
- Maîtriser la boîte à outils technologiques que constitue Hadoop et son écosystème
- Appliquer les bonnes pratiques de développement dans un contexte agile

Langue d'enseignement

- Français

Les points forts de la formation

- Les consultants formateurs sont des experts et interviennent régulièrement en entreprise sur ces aspects.

Contenus

- Introduction
- Qu'est-ce que le big data ?
- Introduction à Hadoop
 - Atelier 1 HDFS blocs
 - Atelier 2 HDFS fichiers
 - Ateliers 3 HDFS java
- HDFS et MAPREDUCE
- YARN
- **Workshop** : Développement MapReduce avec java
- Hadoop comme une plate-forme
- Les distributions d'Hadoop
- Présentation de Sqoop

- Ateliers Sqoop 1
- Ateliers Sqoop 2
- Présentation de Hive
- **Workshop** : Sqoop et Hive
- Présentation de HBase
- Atelier HBase
- Présentation de Spark
- **Workshop** : Spark
- Spark Streaming
- Atelier Spark Streaming
- **Workshop** Architecture

Nos valeurs : L'humilité, L'honnêteté, La bienveillance, L'altruisme.

- L'humilité favorise l'esprit d'équipe, l'apprentissage et amélioration des compétences
- L'honnête : dans notre logique parole vaut contrat
- La bienveillance : l'écoute active et management participative
- L'altruisme : entraide en cas de difficulté le bien-être au travail est essentiel pour nous.
- La loyauté : fidélité de quelqu'un à l'égard de quelqu'un ou quelque chose, c'est aussi d'être fidèle à soi-même.