Formation Introduction au Big Data

Objectifs : Comprendre le concept du Big Data
Être capable d'identifier l'écosystème et comprendre les technologies associées
Savoir anticiper son intégration dans les activités informatiques de l'entreprise

Compétences visées : - Avoir une introduction dédiée à la compréhension du Big Data
- Comprendre et appréhender le potentiel de rupture qu'offre la Big Data
- Connaître les méthodes et outils pour gérer efficacement un projet Big Data
- Cerner les fondamentaux du Big Data pour optimiser son Business
- Disposer d'une démonstration des principales technologies pour en comprendre les potentialités et les enjeux

Durée : 2 jours (14 heures)

Public : Data miners, data scientists, développeurs, chefs de projet, consultants en informatique décisionnelle, statisticiens ...

Méthode pédagogique : Pédagogie active mêlant exposés, exercices et applications pratiques

Modalités d'évaluation : Un formulaire d'auto-évaluation proposé en amont de la formation nous permettra d'évaluer votre niveau et de recueillir vos attentes. Ce même formulaire soumis en aval de la formation fournira une appréciation de votre progression.
Des exercices pratiques seront proposés à la fin de chaque séquence pédagogique pour l'évaluation des acquis.
En fin de formation, vous serez amené(e) à renseigner un questionnaire d'évaluation à chaud.
Une attestation de formation vous sera adressée à l'issue de la session.
Trois mois après votre formation, vous recevrez par email un formulaire d'évaluation à froid sur l'utilisation des acquis de la formation.

Accessibilité : Vous souhaitez suivre notre formation Formation par ville et êtes en situation de handicap ? Merci de nous contacter afin que nous puissions envisager les adaptations nécessaires et vous garantir de bonnes conditions d'apprentissage

Tarif : 1200 € HT (-10% pour 2 inscrits, -20% dès 3 inscrits)

Nos prochaines sessions

Distance
du 24 au 25 juin 2024
du 28 au 29 novembre 2024

Lyon
du 19 au 20 septembre 2024

Paris
du 10 au 11 juin 2024

du 4 au 5 novembre 2024

Toulouse
du 21 au 22 mai 2024
du 7 au 8 octobre 2024

Programme :

- Définition commune 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) et 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)

- Relation entre Big Data et Cloud DaaS (Data as a Service)

- Technologies associées au Cloud (datacenters, stockage, virtualisation, Grid, OS d'orchestration d'Openstack, réseaux...)
- Emergence des solutions Big Data proposées en mode Cloud DaaS (Data as a service)

- 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 Visualization)

- L'apport du NoSQL

- Description des principales familles de bases de données NoSQL
- Zoom sur MongoDb
- Zoom sur Cassandra
- Zoom sur Neo4j

- Une journée au cœur des écosystèmes Hadoop et Spark

- Introduction générale
- Les principales briques et leur rôle
- PIG, Hive, Oozie, Flume, HDFS, MapReduce, Sqoop, Zookeeper, Hbase, Mahout, les connecteurs, …
- SparkML, SparkQL, MLlib, Kafka, Storm, Flink , Beam, …
- Démonstrations et discussions

Date de dernière modification : 14 novembre 2023