



FC9BD02 1 700 € 2 jour(s)

Big Data : Enjeux stratégiques et défis technologiques

OBJECTIFS

- Prendre en compte l'environnement économique en place dans un projet big data. Identifier la donnée à caractère personnel.
- Tenir compte des enjeux et des modalités de la conformité au regard de la législation européenne et française.
- Intégrer les techniques, qu'il s'agisse du traitement de requêtes ou de l'analyse sur des données très volumineuses (« data analytics ») dans la proposition de solutions big data.

PROGRAMME

Introduction

- Big Data : pourquoi s'y intéresser, pourquoi maintenant
- Caractéristiques du big data (les trois V : volume, vitesse, variété)
- Exemples de projets big data

Ecosystème économique du Big Data

- Comment lier la data aux sources de création de valeur de l'entreprise
- Renforcement de business models et nouveaux business models
- Shared/open data : gadget ou opportunité
- Valoriser la donnée (data) au-delà des questions techniques

Introduction aux aspects juridiques : Règlement Général sur la Protection des Données (RGPD) et Big Data

- RGPD et notion de donnée à caractère personnel
- Quels sont les enjeux pour l'entreprise et les citoyens
- Quelles sont les obligations d'un responsable de traitement
- Comment respecter les droits des personnes concernées par les données personnelles
- Quel est le rôle de la CNIL et des autres autorités de protection des données de l'UE ? Quels sont leurs pouvoirs ?
- Comment piloter la conformité, en particulier l'obligation de responsabilité

Interroger des données très volumineuses

- Limites des SGBD relationnels et SGBD parallèles
- Modèle de programmation parallèle sur les données : MapReduce



DATES ET LIEUX

Nous contacter pour les sessions à venir

PUBLIC / PREREQUIS

Décideurs et ingénieurs désireux d'appréhender les enjeux du big data et comprendre sa mise en œuvre.

Une connaissance des systèmes de gestion de bases de données et/ou des méthodes d'apprentissage automatique permet de tirer un meilleur profit de la formation.

COORDINATEURS

Bruno DEFUDE

Professeur à Télécom SudParis, il est spécialiste de la gestion de données et a notamment travaillé sur la distribution des données à grande échelle. Il travaille aujourd'hui sur la gestion de données dans le Cloud.

Claire LEVALLOIS-BARTH

Maître de conférences en droit à Télécom Paris, coordinatrice de la chaire « Valeurs et politiques des informations personnelles ». Elle est spécialiste du droit des nouvelles technologies et plus particulièrement de la protection des données à caractère personnel (Privacy).

MODALITES PEDAGOGIQUES

Des exemples illustrent les concepts théoriques.

- Principe de fonctionnement de MapReduce
- Exemples en MapReduce
- Présentation de l'éco-système Hadoop (implémentation de MapReduce par Apache)
- HDFS et Hadoop
- Exemples de programmes en Hadoop
- Ecrire des programmes Hadoop plus facilement : Hive et Pig
- Limites de MapReduce

Interroger des flux en temps réel

- Problématique de traitement des flux de données
- Gérer des flux de données avec Apache Kafka
- Processeurs de flux de données
- Présentation de Storm (Yahoo)
- Exemples en Storm

Analyser des données très volumineuses

- Principes du machine learning
- Présentation de Mahout, librairie d'algorithmes de machine learning d'Apache
- Spark, un modèle de programmation parallèle adapté au machine learning
- Exemples d'un système de recommandation en Mahout

Visualisation des données

Synthèse et conclusion