

[Formation] Big Data – Enjeux stratégiques et défis technologiques

OBJECTIFS

- Prendre en compte l'environnement économique et les obstacles non techniques en place dans un projet data/IA.
- 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

- Caractéristiques du big data (les trois "V" : volume, vitesse, variété)
- Exemples de projets big data

Écosystème économique du big data

- Data et sources de création de valeur
- Renforcement de business models et nouveaux business models
- Shared/Open Data: Gadget ou opportunité
- Valoriser la donnée (Data)

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

- RGPD et donnée à caractère personnel
- Enjeux pour l'entreprise et les citoyens
- Obligations d'un responsable de traitement
- Respecter les droits des personnes concernées par les données personnelles
- Rôle et pouvoirs de la CNIL et des autres autorités de protection des données de l'UE



DATES ET LIEUX

Du 25/03/2024 au 26/03/2024 à Paris
Du 09/09/2024 au 10/09/2024 à Paris

PUBLIC / PREREQUIS

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

Techniciens ou ingénieurs spécialistes des sciences de la donnée, décideurs et ingénieurs ayant des connaissances des systèmes de gestion de bases de données et des méthodes d'apprentissage automatique

Des connaissances en systèmes de gestion de bases de données et/ou en méthodes d'apprentissage automatique sont souhaitables afin de tirer pleinement profit de la formation.

COORDINATEURS

Bruno DEFUDE

Enseignant-chercheur 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.

MODALITES PEDAGOGIQUES

Des exemples illustrent les concepts

- 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

- Principe de fonctionnement de MapReduce
- Exemples en MapReduce
- Écosystème Hadoop (implémentation de

MapReduce par Apache)

- HDFS et Hadoop
- Exemples de programmes en Hadoop
- Écrire 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

théoriques.