UE SPECIALITE
Coef : 4,5/12

Maxime Guériau

Ingénierie des données

Big Data

En quoi consiste cette matière ?

01. Lien du wiki

https://wiki.insa-rouen.fr/doku.php?id=insa:iti:maquette:semestre_8:big_data

02. Pré-requis

- Bases de données 1
- Bases de données 2 (option)
- Systèmes d'Exploitation
- Réseaux (option)

03. Résumé

Ce cours a pour objectif principal de développer les compétences indispensables permettant de manipuler de gigantesques volumes et fréquences de données sur des architectures distribuées.
Objectifs du cours :

- Comprendre les enjeux et les problématiques associées au stockage et au traitement des mégadonnées
- Comprendre les limitations des systèmes de gestion de bases de données (relationnelles) classiques et les solutions plus adaptées aux mégadonnées
- Connaître le fonctionnement de l'algorithme MapReduce
- Apprendre à utiliser l'écosystème Apache Hadoop
- Apprendre à gérer des flux en streaming avec Spark
- S'initier au calcul distribué avec Spark
- Découvrir le stockage temps-réel avec les technologies NoSQL (en utilisant MongoDB)
- Étudier les architectures distribuées (RMI) et la mise à disposition de données à grande échelle (REST).

04. Les applications futures

Applications liées aux jobs de Data Analyst, Data Scientists (pré-traitement, traitement, stockage, visualisation, etc.). Conception d'architectures distribuées, interfaçage logiciel, applications en streaming, etc.

LES CONSEILS DU PROF

Il y a un fort contraste entre les cours très théoriques et assez haut niveau (et en anglais) et les TPs très appliqués même si assez peu poussés techniquement. Le plus difficile est d'arriver à faire le lien entre les applications et les connaissances mais aussi d'arriver à jongler entre les différentes technologies. Il ne faut pas hésiter à prendre un peu de recul pendant les TPs pour essayer de retrouver la vision haut niveau de ce qu'on essaie de vous faire faire.

LES RETOURS DES ÉLÈVES

"Il y a beaucoup de choses à apprendre mais c'est une matière très intéressante pour la culture informatique. Plus facile si on a fait réseaux."

"De nombreuses choses sont abordées. À noter que c'est un cours qui a de fortes chances d'être fait en anglais"

"Un cours d'introduction qui fait un tour d'horizon donc rien de très approfondi, les TP sont basiques mais on comprend un peu mieux comment fonctionnent les outils de Big Data. Le QCM de l'évaluation finale est dur mais le TP est identique à ceux faits en cours. "