

Analyse de données massives

Description

Étude des enjeux et des défis liés à la collecte, au stockage et à l'analyse des données massives.

Contextes d'application et perspectives.

Le cours mettra l'accent sur l'étude de cas utilisant des données issues de domaines variés (réseaux sociaux, sécurité, génomique, etc.).

Objectifs

Ce cours permettra aux étudiants de :

- se familiariser avec l'analyse de données massives,
- connaître les enjeux et les difficultés particulières au traitement des données massives,
- découvrir et comprendre les différentes approches algorithmiques permettant de traiter ces données,
- connaître les principaux outils logiciels du domaine et leur cadre d'utilisation,
- approfondir leur connaissance de l'état de l'art sur un thème d'application de leur choix (santé, sécurité, réseaux sociaux, génomique, marketing, etc.).

Contenu du cours

Introduction	Éthique et sécurité
Problèmes spécifiques	Données massives connectées
Collecte et stockage des données massives	Framework Apache : Hadoop, MapReduce, Spark, Storm...
Manipulation des données massives	Bases de données non relationnelles

Déroulement du cours

- Une séance de 3 heures chaque semaine
- Présentations des étudiants : articles scientifiques, démonstrations techniques, projets
- Présentations de chercheurs/spécialistes invités

Evaluation

Analyse et présentation d'articles scientifiques	30%	Présentation technique	20%
Quiz	10%	Projet	40%

L'analyse et la présentation d'articles scientifiques, la présentation technique sont des travaux individuels.

Le projet doit être réalisé par équipe de 2 à 3 étudiants. Les étudiants devront constituer un dossier projet sur un thème de leur choix qui pourra être lié à leur sujet de recherche. Le sujet de chaque projet sera défini en accord avec la professeure.