Traitement de Données Massives

Infos pratiques

- > ECTS : 4.5
- > Nombre d'heures : 36.0
- > Niveau d'étude : BAC +5
- > Période de l'année : Enseignement neuvième semestre
- > Méthodes d'enseignement : En présence
- > Forme d'enseignement : Cours magistral et Travaux dirigés
- > Composante : Sciences économiques, gestion, mathématiques et informatique

Objectifs

Ce cours vise à enseigner aux étudiants les fondements de différents types de bases de données pour le Big Data. On étudiera les bases de données basées sur le SQL, le NoSQL et le NewSQL. Les notions apprises seront mises en pratique dans des travaux dirigés et petits projets où les étudiants devront manipuler différents types de bases de données afin de les comparer.

Approche pédagogique et plan de cours.

- Présentation d'architectures et de différents frameworks permettant le traitement de Big Data.
- Les paradigmes de programmation MapReduce et Spark.
- Présentation des modèles de données avancés et différentes technologies que leur sont associées.
- Les notions acquises en cours sont appliquées en Travaux Dirigés et petit-projets.

Évaluation

Session 1 : Évaluation continue (cf. règle par défaut de la section « Modalités spécifiques » des M3C spécifiques)

Session 2 : Règle par défaut décrite dans la section « Modalités de contrôle et examens / Modalités spécifiques »

Pré-requis nécessaires

- Connaissance des principes de la programmation fonctionnelle et distribuée.
- Bases de données avancées.
- Indexation de données.

Compétences visées

- Compréhension des différents modèles de stockage et gestion de données.
- Connaître les différentes technologies que leur sont associées.
- Pouvoir choisir le système de gestion de données le plus adapté à un problème donné.

Bibliographie

- Guy Harrison. Next Generation Databases: NoSQL, NewSQL, and Big Data. Apress 2015.
- Nick Dimiduk and Amandeep Khurana. 'HBase in action'. Manning Publications, 2013.
- Rudi Bruchez, Les bases de données NoSQL comprendre et mettre en œuvre. Eyrolles 2013.
- Matei Zaharia, Bill Chambers. Spark: The Definitive Guide. Big Data Processing Made Simple. O'Reilly Media, February 2018.
- Aaron Ploetz, Devram Kandhare, Sudarshan Kadambi , Xun (Brian) Wu. Seven NoSQL Databases in a Week: Get up and running with the fundamentals and functionalities of seven of the most popular NoSQL databases, Packt publishing, March 29, 2018.
- Articles de recherche du domaine