

DÉFIS INFORMATIQUES DU BIG DATA

COMPUTING CHALLENGES OF BIG-DATA

Responsable(s): Stéphane DERRODE, Alexandre SAIDI, Céline HARTWEG-HELBERT

| Cours : 16.0 | TD : 0.0 | TP : 0.0 | Autonomie : 0.0 | BE : 12.0 | Projet : 0.0 | Langue du cours : FR

Objectifs de la formation

Dans de nombreux domaines scientifiques, tels que la biologie ou les sciences de l'environnement, l'évolution rapide des instruments scientifiques ainsi que le recours intensif à la simulation informatique ont conduit, ces dernières années, à une production importante de données. Les applications scientifiques sont confrontées à de nouveaux problèmes liés essentiellement au stockage et à l'exploitation de ces données.

L'enseignement proposé permettra de découvrir les grandes problématiques soulevées par l'émergence de ces flots de données (stockage, interrogation, analyse et visualisation), et d'aborder certaines solutions technologiques actuellement proposées. Les questionnements d'ordre éthique et juridique que soulèvent la

Mots-clés : Big Data, Data scientist, NoSQL, Hadoop, Big Data analytics, Open Data, Web des données

Programme

- Big Data : une introduction aux enjeux, perspectives et applications
- Problématique des grandes bases de données (NoSql, NewSql)
- Big Data et modèle économique : le cas de l'intermédiation
- Open Data : données publiques ouvertes
- Big-Data Analytics : les bases de l'analyse de grands volumes de données
- Représentation et visualisation des données
- Trois BEs sur la visualisation, sur Apache/Hadoop et sur le web des données (SparQL).

Compétences

- A l'issue de ce MOS, l'étudiant sera capable : - De connaître les enjeux, les opportunités et les problèmes éthiques soulevés par le big-data. - De réaliser des programmes Hadoop/Map-Reduce simples pour exploiter des données réparties. - De manipuler des bases NoSql à l'aide d'un SGBD moderne (p. ex. Mongo-Db).

Travail en autonomie

Objectifs : Rédaction d'une synthèse bibliographique sur un thème du big data, par groupe de 3 à 6 étudiants

Méthodes :

Bibliographie

Philippe Lacomme, *BASES DE DONNÉES NOSQL ET BIG DATA : CONCEVOIR DES BASES DE DONNÉES POUR LE BIG DATA : COURS ET TRAVAUX PRATIQUES*, Ellipses, 2014
Viktor Mayer-Schönberger *BIG DATA : LA RÉVOLUTION DES DONNÉES EST EN MARCHÉ*, R. Laffont, 2014
Srinath Perera *HADOOP MAPREDUCE COOKBOOK : RECIPES FOR ANALYZING LARGE AND COMPLEX DATASETS WITH HADOOP MAPREDUCE*, 2013

Contrôle des connaissances

Note = 50% savoir + 50% savoir-faire
Note de savoir = 100% examen terminal
Note de savoir-faire = 50% synthèse bibliographique + 50% CR de BE