



DÉFIS INFORMATIQUES DU BIG DATA

COMPUTING CHALLENGES OF BIG-DATA

Lecturers: Stéphane DERRODE, Alexandre SAIDI, Céline HARTWEG-HELBERT

| Lecturers : 16.0 | TC : 0.0 | PW : 0.0 | Autonomy : 0.0 | Study : 12.0 | Project : 0.0 | Language : FR

Objectives

In many scientific fields, such as biology or environmental sciences, the rapid evolution of scientific instruments as well as the intensive use of computer simulation have led, in the last few years, to an important production of data. Scientific applications are confronted with new problems mainly related to the storage and exploitation of these data.

The teaching will allow us to discover the major problems raised by the emergence of these data flows (storage, interrogation, analysis and visualization), and to approach certain technological solutions currently proposed. The ethical and legal issues raised by the collection and exploitation of these data will also be examined.

Keywords : Big Data, Data scientist, NoSQL, Hadoop, Big Data analytics, Open Data, Linked open data

Programme

- Big Data: an introduction to the issues, perspectives and applications
- The problem of large databases (NoSql, NewSql)
- Big Data and business model: the case of intermediation
- Open Data: open public data
- Big-Data Analytics: the basics of analyzing large volumes of data
- Data representation and visualization
- Three BEs on visualization, on Apache/Hadoop and on the web of data (SparQL).

Learning outcomes

- Upon completion of this MOS, students will be able to: - To know the issues, opportunities and ethical problems raised by big-data. - Create simple Hadoop/Map-Reduce programs to exploit distributed data. - Manipulate NoSql databases using a modern DBMS (e.g. Mongo-Db).

Independent study

Objectifs : Writing of a review on a big data theme, by groups of 3 to 6 students

Méthodes :

Core texts

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

Assessment

Grade = 50% knowledge + 50% know-how
Knowledge grade = 100% final exam
Know-how mark = 50% bibliographic synthesis + 50% report on BE