



TECHNOLOGIES INFORMATIQUES DU BIG DATA

COMPUTER-BASED TECHNOLOGY FOR BIG DATA

Lecturers: Stéphane DERRODE, Alexandre SAIDI, Daniel MULLER, Mohsen

| Lecturers : 6 | TC : 0.0 | PW : 0.0 | Autonomy : 0.0 | Study : 14 | 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, has led to a significant production of data in recent years. Scientific applications are now facing new problems related to the storage and use of these large volumes of data. The problem is much the same for the management of data collected by social networks, this time with the objective of commercial optimization.

The proposed teaching will allow students to discover 3 major technologies emblematic of big-data processing

Keywords : Big Data, NoSQL, MongoDB, Hadoop, Spark, python

Programme

- 3 sessions of 2 hours each on MongoDB, Hadoop and Spark.
- 3 sessions of 4 hours of practical works on MongoDB, Hadoop and Spark.
- 1 practical work session of 2 hours on Spark MLlib.

Learning outcomes

- - Know how to manipulate No-SQL databases with MongoDB
- - Know how to write a map-reduce algorithm with Hadoop with Python, in an HDFS storage environment.
- - Know how to write a Spark algorithm with Python, in an HDFS storage environment.

Independent study

Objectifs : The practical works are essentially in autonomy. Additional work is required to write the reports.

Méthodes :

Core texts

Assessment

The average of marks obtained on the reports of 3 practical works.