Infrastructure for cloud and big data

ECTS
3 crédits

Hourly volume
38h

Introducing

Description

This lecture aims at the introduction of software technologies which allow the deployment and execution of treatments in infrastructures based on clusters of machines. Such infrastructures are widely used in the domain of cloud computing where datacenters allow hosting externalized services, of big data and machine learning for the treatment and exploitation of large data sets.

The first part of the lecture introduces concepts and tools associated with cloud computing such as virtualization (KVM), containers (Docker), the administration of virtualized clusters (OpenStack, Kubertenes) and the main services provided by cloud operators (AWS).
The second part introduces concepts and tools associated with big data, such as the parallel processing of large data sets (Hadoop, Spark) and the real-time processing of data (Spark-streaming, Storm).

Objectives

At the end of this module, the student will have understood and be able to explain (main concepts):

- General concepts of cloud and big data computing infrastructures
- Principles of virtualized infrastructures
- Cloud services
- Tools associated with cloud infrastructures
- Principles of big data computing platforms (map-reduce, stream processing)
- Big data treatment environments (Hadoop, Spark, Storm)

At the end of this module, the student should be able to:

- Use virtualization platforms
- Use cloud platforms
- Program big data applications
- Execute big data applications in a computing infrastructure

Necessary prerequisites

Algorithmic, Java programming, Linux environment handling (shell commands)

Évaluation

L'évaluation des acquis d'apprentissage est réalisée en continu tout le long du semestre. En fonction des enseignements, elle peut prendre différentes formes : examen écrit, oral, compte-rendu, rapport écrit, évaluation par les pairs...

Practical info

Location(s)

Toulouse