Infrastructure for massive data processing

ECTS
4 crédits

Hourly volume
61h

Introducing

Objectives

At the end of this module, the student will understand and be able to explain the concepts and techniques related to the main pillars that have to be managed by an IT service provider, in terms of:

- physical infrastructure (network, storage , computing) ;
- organizational and data management (allocation of storage , ...);
- computation services of such data (based on calculation models like map reduce, etc.).

The student will be able to:

1) With regard to physical infrastructures
- design and deploy a network architecture adapted to a big data oriented service, using advanced network technology (network virtualization, optimization protocols, etc.);
- dimension and deploy a physical storage infrastructure aimed at receiving massive amounts of data;
- assess and deploy the computing power required to process massive data, based on the latest technologies for processors, such as virtualization.

2) With regard to the organization and data management
- design and implement tools to organize data within the physical infrastructure;
- provide appropriate interfaces for access to such data;
- choose a data organization adapted to the constraints of treatment (offline versus real-time processing);

3) With regard to the data processing services
- provide facilities for analyzing data and extract value added information (e.g., learning, trends).

Necessary prerequisites

Networks
Operating systems
Databases
Algorithmic and programing

Practical info

Location(s)

Toulouse