

# Infrastructures pour le traitement de données massives

 **ECTS**  
4 crédits

 **Volume horaire**  
61h

## Présentation

---

### Objectifs

A la fin de ce module, l'étudiant devra avoir compris et pourra expliquer les concepts et techniques liés aux principaux piliers que doit gérer un fournisseur de services IT, en termes :

- d'infrastructures physiques incluant le Cloud (réseau, stockage, puissance de calcul) ;
- d'organisation et gestion de données (répartition du stockage, ...)
- de services de traitement de ces données (basés sur des modèles de calcul de type map reduce, etc.).

L'étudiant devra être capable de :

- 1) concernant les infrastructures physiques
  - de concevoir et déployer une architecture adaptée à une offre de service orientée big data, à l'aide de technologies réseaux évoluées (virtualisation de réseau, optimisation de protocoles, etc) ;
  - de dimensionner et déployer une infrastructure physique ou virtuelle de stockage destinée à recevoir des quantités massives de données et/ou exécuter des applications à base de services;
  - d'estimer et déployer la puissance de calcul nécessaire au traitement de données massives, en s'appuyant sur les technologies récentes des processeurs, telles que la virtualisation.
- 2) concernant l'organisation et de la gestion de données
  - de concevoir et déployer des outils permettant

d'organiser l'ensemble des données au sein de l'infrastructure physique ;  
- de choisir une organisation des données adaptée aux contraintes des traitements (offline versus real-time processing) ;

3) concernant les services de traitement de ces données  
- d'offrir des moyens d'analyser ces données afin d'en extraire de la valeur ajoutée (connaissances, tendances, prédiction, etc.)

---

### Pré-requis nécessaires

Réseaux  
Système d'exploitation  
Bases de données  
Algorithmique et programmation

---

## Infos pratiques

### Lieu(x)

 Toulouse