

Big Data (DATA831_IDU)



Composante
Polytech
Annecy-
Chambéry

En bref

- > **Langues d'enseignement:** Français
- > **Méthodes d'enseignement:** En présence
- > **Ouvert aux étudiants en échange:** Oui

Présentation

Description

Avec l'avènement du Digital, les données sont de plus en plus omniprésentes et en nombre tel qu'il n'est plus possible d'envisager une analyse sur un ordinateur personnel. Le Big Data est né chez les géants de l'Internet (Google, Amazon, Facebook, Yahoo) et a donné lieu à de nombreux outils devenus Open Source comme Big Table, Hadoop, MongoDB pour n'en citer que quelques-uns. Le module Big Data présente les différentes approches existantes pour la prise en compte de données massives à savoir un traitement par batch (avec Hadoop et Map Reduce), un traitement live (avec Apache Spark) et finalement l'architecture lambda mixant les deux approches.

Objectifs

A l'issue du cours, l'étudiant sera capable :

- de comprendre les trois architectures et leur utilisation
- de mettre en oeuvre les trois architectures

Heures d'enseignement

CM	Cours Magistral	7,5h
TP	Travaux Pratiques	12h

Pré-requis obligatoires

- Systèmes distribués à large échelle (INFO 833)
 - Bases de données distribuées (INFO 834)
-

Plan du cours

1. Qu'est-ce que le Big Data ?
2. Quelles sont les caractéristiques du Big Data ?
 1. Volume
 2. Vitesse
 3. Variété
3. Du 3V au 5V
 1. Valeur
 2. Véracité
4. Processus de calcul
 1. Approche Map Reduce
 2. Approche Streaming
 3. Approche Lamba
1. Approche Map Reduce pour le traitement de données textuelles
2. Approche Streaming pour l'analyse de flux de données
3. Architecture Lambda pour l'analyse dans le domaine des réseaux sociaux

Libellé court : DATA831_IDU

Nature : MODULE

Infos pratiques

Lieux

➤ Annecy-le-Vieux (74)

Campus

› Anancy / campus d'Anancy-le-Vieux

Contacts

Responsable du cours

Flavien Vernier

☎ +33 4 50 09 65 90

✉ Flavien.Vernier@univ-savoie.fr