

Big Data



En bref

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

Présentation

Description

Le but de ce cours est de comprendre les enjeux du "Big Data" et de construire et manipuler des bases de données massives dans un environnement NoSQL (Not Only SQL). Nous allons ainsi travailler sur des données réelles à l'aide de bases de données orientées graphes dans l'environnement neo4j. Il s'agit d'une technologie qui est basée sur la théorie des graphes et qui permet de stocker et relier l'information dans la base de données et ensuite de faire des requêtes dans un langage qui s'appelle Cypher. Cette technologie passe à l'échelle au sens où l'on peut utiliser les mêmes outils sur sa machine ou sur des clusters de serveurs selon le volume de données à traiter.

Objectifs

L'objectif principal porte sur la maîtrise d'une base de données dans un environnement Big Data. Ce cours est à la fois théorique, en particulier pour construire la base de données à partir de la structuration d'un graphe, et pratique car il s'appuie sur la programmation d'un moteur de recommandation se basant sur un apprentissage statistique directement réalisé sur les données pour pouvoir faire du ciblage et de la recommandation personnalisés.

Heures d'enseignement

CM	Cours Magistral	12h
TD	Travaux Dirigés	7,5h
TP	Travaux Pratiques	8h

Pré-requis obligatoires

Langage SQL.

Plan du cours

- 1) Théorie des graphes.
 - 2) neo4j et les requêtes Cypher.
 - 3) Mesure de similarité
 - 3) Construction de la base de données
 - 4) Parcours de la base et valorisation des données
 - 5) Moteur de recommandation sur des données réelles
-

Compétences visées

Savoir manipuler et valoriser des données massives.

Apprendre neo4j et le langage Cypher.

Construire un moteur de recommandation sur des données réelles.

Utiliser une technologie à la frontière entre l'informatique et la théorie des graphes.

Bibliographie

Graph Algorithms: Practical Examples in Apache Spark and Neo4j By Mark Needham & Amy Hodler By O'Reilly Media 300 pages (téléchargeable sur le site <https://neo4j.com/graph-algorithms-book/>).

Infos pratiques

Contacts

Responsable du cours

Laurent Vuillon

☎ +33 4 79 75 87 33

✉ Laurent.Vuillon@univ-savoie.fr

Lieux

➤ Le Bourget-du-Lac (73)

Campus

➤ Le Bourget-du-Lac / campus Savoie Technolac