



## TECHNOLOGIES INFORMATIQUES DU BIG DATA

### COMPUTER-BASED TECHNOLOGY FOR BIG DATA

Responsable(s): Stéphane DERRODE, Alexandre SAIDI, Daniel MULLER, Mohsen

| Cours : 6 | TD : 0.0 | TP : 0.0 | Autonomie : 0.0 | BE : 14 | Projet : 0.0 | Langue du cours : FR

#### Objectifs de la formation

---

Dans de nombreux domaines scientifiques, tels que la biologie ou les sciences de l'environnement, l'évolution rapide des instruments scientifiques, ainsi que le recours intensif à la simulation informatique, ont conduit à une production importante de données ces dernières années. Les applications scientifiques sont aujourd'hui confrontées à de nouveaux problèmes liés au stockage et à l'exploitation de ces grands volumes de données. La problématique est sensiblement la même dans le cadre de la gestion des données collectées par les réseaux sociaux, avec cette fois-ci des objectifs d'optimisation commerciale.

**Mots-clés :** Big Data, NoSQL, MongoDB, Hadoop, Spark, python

---

#### Programme

- 3 séances de cours de 2h portant sur MongoDB, Hadoop et Spark.
- 3 séances de TP de 4h sur MongoDB, Hadoop et Spark.
- 1 séance de TP de 2h sur Spark MLlib.

#### Compétences

- - Savoir manipuler des bases de données No-SQL avec MongoDB
- - Savoir écrire un algorithme map-reduce avec Hadoop avec Python, dans un environnement de stockage de type HDFS
- - Savoir écrire un algorithme en langage Spark avec Python, dans un environnement de stockage de type HDFS

#### Travail en autonomie

**Objectifs :** Les TPs sont pour l'essentiel en autonomie. Un travail complémentaire est requis pour rédiger les compte-rendus.

**Méthodes :**

#### Bibliographie

#### Contrôle des connaissances

La moyenne de notes obtenues sur les compte-rendus de 3 TP.