

Informatique pour le Big Data



Niveau
d'étude
BAC +5 /
master



ECTS
6 crédits



Composante
Faculté des
sciences

En bref

- › Langue(s) d'enseignement: Français
- › Ouvert aux étudiants en échange: Oui

Présentation

Description

Concepts et les enjeux du Big Data

Technologies du Big Data :

- Architecture et composants de la plateforme Hadoop.
- Modes de stockage (NoSQL, HDFS).
- Principes de fonctionnement de MapReduce.
- Programmation concurrente (générateurs, multi-threading, multi-processing, asynchronisme).
- Web services.
- Nasse de data visualisation.

Gérer les données structurées et non structurées :

- Principes de fonctionnement HDFS.
- Importer des données externes vers HDFS.
- Réaliser des requêtes SQL avec HIVE.

— Utiliser *PIG* pour traiter la donnée.

Les méthodes d'analyse des données pour le *Big Data* (en liaison avec *S3-UE1-DS*) :

- Les méthodes d'exploration.
- Segmentation et classification.
- Estimation et prédiction.
- Implémentation des modèles.
- Méthodes de réduction de dimensions.

Introduction au *MLOps* : Mise en production ou déploiement d'un modèle de *Machine Learning*.

- Aspects techniques d'un déploiement.
- Mise en pratique : déploiement d'une API Web avec *Flask*.
- Maintenance post-déploiement : logging, model drift, qualité des données.
- Prérequis d'un déploiement réussi : versionnage, tests unitaires, bonnes pratiques de code, etc.
- Vue d'ensemble des métiers et logiciels liés au *MLOps*.

Objectifs

- Comprendre les concepts et l'apport du *Big Data* par rapport aux enjeux métiers.
- Comprendre l'écosystème technologique nécessaire pour réaliser un projet *Big Data*.
- Acquérir les compétences techniques pour gérer des flux de données complexes, non structurés et massifs.
- Visualisation d'espaces latents, implémenter des modèles d'analyses statistiques pour répondre aux besoins métiers.
- Comprendre l'intérêt du déploiement d'un modèle de *Machine Learning*
- Avoir une vue d'ensemble des points d'attention d'une mise en production, avant, pendant et après sa réalisation
- Acquérir une compétence technique de base pour créer une API Web avec *Flask* ; mettre en pratique quelques aspects du *MLOps*.

Heures d'enseignement

CM	Cours magistral	24h
TP	Travaux pratique	32h

Pré-requis obligatoires

Notions et contenus : Bases de données relationnelles (S2-UE4-DS). Programmation (S1-UE5-DS). Statistique (S2-UE1-DS).

Compétences et capacités : Connaissances de base des modèles relationnels. Connaissances en statistique et des langages de programmation.

Infos pratiques

Lieu(x)

> Angers