3IF2080 - Algorithmes pour systèmes distribués

Responsables : **Francesca BUGIOTTI**

Langues d'enseignement : **FRANCAIS**

Campus où le cours est proposé : **CAMPUS DE PARIS - SACLAY**

Nombre d'heures d'études élèves (HEE) : **40**

Nombre d'heures présentielles d'enseignement (HPE) : **21**

Année académique : **2024-2025**

Niveau avancé : **non**

Présentation, objectifs généraux du cours :

Nombreuses applications traitent et stockent des masses de données pour produire des analyses efficaces et pertinents.

Pour exploiter ces données et utiliser efficacement les systèmes de stockage et d'analyse il faut comprendre comment développer des façon efficace sur des systèmes distribués.

Dans ce contexte ce cours a comme objectif de décrire comment traiter des données, hétérogènes, volumineuses, et complexes en utilisant des méthodes, algorithmes et surtout des systèmes distribués.

Chaque challenge et défi théorique sera analysé en relation avec une technologie "big data" et un framework innovant comme Docker, Kubernetes, et Kubeflow.

Période(s) du cours (n° de séquence ou hors séquence) :

SM11

Prérequis :

aucun

Plan détaillé du cours (contenu) :

Chapitre 1. Introduction

- Concepts de distribution
- Modèles de distribution

Chapitre 2. Distribution des données

- Algorithmes distribués et parallèles sur les données
- Architectures à services

Chapitre 3. Les environnements de programmation parallèle

- Spark
- Docker
- Kubernetes
- Kubeflow

Chapitre 4. Analyse de plusieurs environnements Cloud

- Analyser les aspects spécifiques de différents environnements Cloud

Chapitre 5. Réalisation du projet d'optimisation des performances en collaboration avec l'entreprise de référence.

Déroulement, organisation du cours :

Introduction. Cours magistral : 1,5h

Distribution des données . Cours magistral : 1,5h

Les environnements de programmation parallèle. Cours magistral : 3h

Techniques avancés pour l'analyse distribuée. Cours magistral : 3h

Comparer des environnements Cloud. Cours magistral : 3h

14 créneaux d'1h30 : 8 cours, 1 présentation des projets, 4 suivi projet, 1 discussion du projet.

Organisation de l'évaluation :

L'évaluation se fera sur la qualité du travail personnel fait sur le projet, sur l'investissement pendant les séances de suivi de projet, ainsi que sur la soutenance.

Acquis d'apprentissage visés dans le cours :

A l'issue de ce cours, l'élève aura enrichi ses connaissances concernant :

1. Distribution des données
2. Algorithmes distribués
3. Optimisation du calcul distribué
4. Outils pour l'analyse distribuée
5. Architecture à services
6. Algorithmes parallèles pour graphes
7. Analyse de modèles de distribution d'Hadoop, Docker et Kubernetes
8. Workflows et leur application

En plus l'élève prendra part au développement d'un projet pour l'optimisation d'une application distribuée en collaboration dans un vrai environnement cloud avec une entreprise de référence.

Description des compétences acquises à l'issue du cours

⋮

C2 Develop in-depth skills in an engineering field and a family of professions.

- C2.5 Carry out a global project in a profession via the validation of the specific skills of a business sector (or equivalent if mobility).

C6 Be operational, responsible, and innovative in the digital world.

- C6.1 Solve a problem numerically.
- C6.2 Design software.

C6.3 Process Data.

C8 Lead a team, manage a project.

- C8.2 Demonstrate leadership capacity, both individually and collectively within the team, on a complex project.
- C8.3 Contribute to the development of team members' skills through the acquisition and appropriation of external resources.

Bibliographie :

Managing Kubernetes: Operating Kubernetes Clusters in the Real World - Brendan Burns, Craig Tracey

Docker: up & Running - Sean p. Kane, Karl Matthias