Contenus

Code du Cours	Nom du Cours	Semestre du Cours	Cours Théoriques	Travaux Dirigés (TD)	Travaux Pratiques (TP)	Crédit du Cours	ECTS
INF 517	Genie Des Donnees	2	3	0	0	3	6

Cours Pré-Requis	
Conditions d'Admission au Cours	

Langue du Cours	Anglais
Type de Cours	Électif
Niveau du Cours	Master
Objectif du Cours	L'ingénierie des données est la discipline qui traite de la conception de systèmes et de l'utilisation de méthodes d'analyse pour l'acquisition, le stockage, la gestion, la sécurité et le traitement des données. Des systèmes de gestion des données sophistiqués sont nécessaires pour traiter les volumes considérables de « mégadonnées » disponibles. Ce cours constituera une introduction aux principes et pratiques de l'ingénierie des données et abordera les thèmes suivants : i. Le cycle de vie de l'ingénierie des données ii. Techniques de modélisation des données pour l'organisation et la gestion des données iii. Création de pipelines de données pour collecter, transformer, analyser et visualiser les données provenant de plusieurs systèmes sources. iv. Manipulation des données à l'aide de différents langages de requête v. Applications et algorithmes d'analyse de données vi. Ingénierie des types de données non traditionnels vii. Normes et qualité des données
Contenus	1. Introduction à l'ingénierie des données : concepts généraux 2. Technologies de stockage des données 3. Plateformes de données cloud (AWS/Azure/GCP) 4. Méthodes d'intégration des données et architectures de pipelines de données 5. Orchestration des flux de travail avec Apache Airflow 6. Transformation des données avec dbt (outil de construction de données) 7. Traitement par lots avec Spark 8. Principes fondamentaux du traitement de flux et Apache Kafka 9. Recherche et extraction d'informations : Elastic Search 10. Lac de données : architecture et principes 11. Maillage de données : architecture et principes 12. Gouvernance des données - 1 : Gestion des métadonnées 13. Gouvernance des données - 2 : Qualité et tests des données 14. Gouvernance des données - 2 : Traçabilité et observabilité des données
Ressources	1. Reis, J, Housley M, Fundamentals of Data Engineering: Plan and Build Robust Data Systems, 1st Edition, 2022, O'Reilly, 978-1098108304 2. Warren, J., & Marz, N. (2015). Big Data: Principles and best practices of scalable realtime data systems. Simon and Schuster. 3. Learning Spark: Lightning-Fast Big Data Analysis, by by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia. O'Reilly Media. Feb 2015 4. Hadoop: The Definitive Guide, by Tom White. O'Reilly Media. April 2015. (Fourth edition of the book at Amazon.com) 5. Gorelik, A. (2019). The enterprise big data lake: Delivering the promise of big data and data science. O'Reilly Media.

Intitulés des Sujets Théoriques

Semaine	Intitulés des Sujets
1	Génie des données -- concepts fondamentaux
2	Technologies de stockage des données
3	Plateformes de données cloud (AWS/Azure/GCP)
4	Méthodes d'intégration des données et architectures de pipelines de données

Semaine	Intitulés des Sujets
5	Orchestration des flux de travail avec Apache Airflow
6	Transformation des données avec dbt (outil de construction de données)
7	Traitement par lots avec Spark
8	Principes fondamentaux du traitement de flux et Apache Kafka
9	Recherche et extraction d'informations : Elastic Search
10	Lac de données : architecture et principes
11	Maillage de données : architecture et principes
12	Gouvernance des données - 1 : Gestion des métadonnées
13	Gouvernance des données - 2 : Qualité et tests des données
14	Gouvernance des données - 3 : Traçabilité et observabilité des données