Content

Course Code	Course Name	Semester	Theory	Practice	Lab	Credit	ECTS
INF 517	Data Engineering	2	3	0	0	3	6

Prerequisites	
Admission Requirements	

Language of Instruction	English
Course Type	Elective
Course Level	Masters Degree
Objective	Data engineering is the discipline concerned with the design of systems and use of analysis methods for the acquisition, storage, management, security, and processing of data. Rich data management schemes are needed to handle the sizeable “Big Data” that is available for processing. This class will be a foundational course in Data Engineering principles and practices and will consists of following headlines: i. The data engineering lifecycle ii. Data modelling techniques for organizing and managing data iii. Building data pipelines to collect, transform, analyse, and visualize data from multiple source systems. iv. Manipulate the data with different query languages v. Data analytics application and algorithms vi. Engineering non-traditional data types vii. Data standards and data quality
Content	1. Introduction to Data Engineering: General Concepts 2. Data Storage Technologies 3. Cloud Data Platforms (AWS/Azure/GCP) 4. Data Integration Methods & Data Pipeline Architectures 5. Workflow Orchestration with Apache Airflow 6. Data Transformation with dbt (data build tool) 7. Batch Processing with Spark 8. Stream Processing Fundamentals & Apache Kafka 9. Search and Information Retrieval: Elastic Search 10. Data Lakehouse: Architecture and Principles 11. Data Mesh : Architecture and Principles. 12. Data Governance - 1: Metadata Management 13. Data Governance - 2: Data Quality and Testing 14. Data Governance - 2: Data Lineage and Observability
References	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.

Theory Topics

Week	Weekly Contents
1	Data Engineering: General Concepts

Week	Weekly Contents
2	Data Storage Technologies
3	Data Integration Methods & Data Pipeline Architectures
4	Data Integration Methods & Data Pipeline Architectures
5	Workflow Orchestration with Apache Airflow
6	Data Transformation with dbt (data build tool)
7	Batch Processing with Spark
8	Stream Processing Fundamentals & Apache Kafka
9	Search and Information Retrieval: Elastic Search
10	Data Lakehouse: Architecture and Principles
11	Data Mesh : Architecture and Principles.
12	Data Governance - 1: Metadata Management
13	Data Governance - 2: Data Quality and Testing
14	Data Governance - 2: Data Lineage and Observability