

Subject Code	Subject Name	Credits
26CS513	DATA ENGINEERING AND AI INFRASTRUCTURE	4

Course Objectives:

1. Understand the fundamentals of data engineering and its role in AI systems.
2. Analyse data storage, data pipelines, and distributed processing techniques.
3. Apply AI infrastructure technologies for model development and deployment.
4. Evaluate scalable and secure infrastructure solutions for modern AI applications.

Learning Objectives:

After completion of the course the students will be able to:

1. Students will be able to design data pipelines and manage data infrastructure for AI applications.
2. Students will be able to apply data engineering tools and platforms for building scalable AI systems.

Unit I - Introduction to Data Engineering (12 Hrs.)

Role of Data Engineers in AI Systems-Data Sources and Data Collection Methods-Structured, Semi-structured and Unstructured Data-Data Pipelines and Data Lifecycle-Overview of Big Data Technologies- framework: Apache Hadoop

Unit II - Data Storage and Data Management (12 Hrs.)

Data Warehousing Concepts-Data Lakes and Data Lake Architecture-Distributed Storage Systems-Data Integration and ETL (Extract, Transform, Load)-Data Quality and Data Governance-Metadata Management- platform: Apache Hive

Unit III - Data Processing for AI Systems (12 Hrs.)

Batch Processing and Stream Processing-Distributed Data Processing-Data Transformation and Feature Engineering-Data Preprocessing for Machine Learning-Data Pipelines for AI Models- processing engine: Apache Spark

UNIT IV - AI Infrastructure and Deployment (12 Hrs.)

Infrastructure Requirements for AI Systems-Cloud Computing for AI-GPU and TPU Computing for AI Workloads-Containerization and Microservices Architecture-Model Deployment and Monitoring- framework: TensorFlow

UNIT V - Advanced AI Infrastructure and Future Trends (12 Hrs.)

Scalable AI Infrastructure-MLOps (Machine Learning Operations)-Data Security and Privacy in AI Systems-Edge Computing and AI at the Edge-Ethical and Responsible AI Infrastructure-Future Trends in Data Engineering and AI Systems

Reference Books:

1. Data Engineering for AI/ML Pipelines,Venkata Karthik Penikalapati, Mitesh Mangaonkar
BPB Publications
2. Data Engineering for AI,Sundeeep Goud Katta, Lav Kumar,BPB Publications
3. Data Engineering for Large Models (Open Online Book)
4. https://datascale-ai.github.io/data_engineering_book

