

Course Code	Course Title	L	T	P	C
BACSE207	Data Engineering	3	1	0	4
Pre-requisite	NIL	Syllabus Version			
		1			
Course Objectives					
1. To cultivate a comprehensive understanding of modern data engineering principles and practices. 2. To learn the fundamental concepts of scalable data storage, data generation, ingestion, and processing techniques. 3. To explore tools for developing, monitoring, and managing data pipelines, and to implement industry-standard frameworks for real-world challenges.					
Course Outcomes					
1. Build a comprehensive understanding of data engineering concepts, architectural patterns, and data storage mechanisms. 2. Develop skills to ingest data from diverse sources and implement data quality frameworks to monitor and enhance data integrity. 3. Specialize in constructing robust data pipeline frameworks optimized for scalability and performance. 4. Design and implement efficient data models and deploy data-serving solutions tailored for analytics applications. 5. Leverage industry-standard tools and security protocols to orchestrate modern data engineering workflows in real-world applications.					
Module:1	Foundations of Data Engineering	8 hours			
Data Engineering Definition - Data Engineering and Data Science - Data Engineering Skills and Activities - Data Engineering Lifecycle - Designing Good Data Architecture - Principles of Good Data Architecture - Major Architecture Concepts, Types of Data Architecture - Data Generation in Source Systems – Files and Unstructured Data, APIs - Application databases - Data Storage Systems - Data Engineering Storage Abstractions.					
Module:2	Data Ingestion and Data Quality Management	10 hours			
Data Ingestion – Ingestion Phase - Batch Ingestion - Message and Stream Ingestion – Ways to Ingest Data - Data Quality – Data Quality Metrics – Designing and Building Data Catalog - Data Quality for Stream Processing – Normalizing Data: Handling Heterogeneous Data Sources - Schema Checking and Type Coercion - Syntactic Versus Semantic Ambiguity in Data - Measuring and Maintaining High Data Reliability at Ingestion- Prioritizing Data Governance and Compliance.					
Module:3	Data Pipelines and Workflow Orchestration	10 hours			

Data Pipelines – Modern Data Infrastructure - Common Data Pipeline Patterns (ETL and ELT) – Orchestrating Pipelines – Data Validation in Pipelines – Maintaining Pipelines – Measuring and Monitoring Pipeline Performance.		
Module:4	Data Modelling and Serving Data	9 hours
Queries - Query Optimizer – Queries on Streaming Data – Data Modeling – Normalization - Techniques for Modeling Batch Analytical Data and Streaming Data - Transformations – Batch Transformations – Stream Transformations and Processing – Serving Data for Analytics, Machine Learning, and Reverse ETL.		
Module:5	Modern Data Engineering and Security	6 hours
Security and Privacy - People, Processes, and Technology – Cloud Scale Data – “Enterprisey” Data Engineering - Modern Data Stack – Live Data Stack – Fusion of Data with Applications - Cloud Networking.		
Module:6	Contemporary Topics	2 hours
Total Lecture Hours:		45 hours
Tutorial Hours:		15 hours
Text Book(s)		
Joe Reis, Matt Housley, " Fundamentals of Data Engineering ", O Reilly Media Inc, 1 st Edition, 2022 James Densmore, " Data Pipelines Pocket Reference ", O Reilly Media Inc, 1 st Edition, 2021		
Reference Books		
Moses, B., Gavish, L., & Vorwerck, M, " Data Quality Fundamentals: A Practitioner's Guide to Building Trustworthy Data Pipelines ", O Reilly Media Inc, 1 st Edition, 2022 Gareth Eager, " Data Engineering with AWS: Acquire the skills to design and build AWS-based data transformation pipelines like a pro ", Packt Publishing , 2 nd Edition, 2023 Andreas Kretz, " The Data Engineering ", Cookbook, 1 st Edition, 2019		
Mode of Evaluation :Continuous Assessment Test, Digital Assignment, Quiz, Final Assessment Test, Seminar, Group Discussion, Presentaion		
Recommended by Board of Studies :		02-06-2025
Approved by Academic Council : No. 78		12-06-2025