PRASAD V. POTLURI SIDDHARTHA INSTITUTE OF TECHNOLOGY
(Autonomous)
KANURU, VIJAYAWADA
CSE (DATA SCIENCE)
II B. Tech – II Semester
Data Engineering

Course Code	23DS3401	Year	II	Semester	II
Course Category	PCC	Branch	CSE (DATA SCIENCE)	Course Type	Theory
Credits	3	L-T-P	3-0-0	Prerequisites	Introduction to Data Science
Continuous Internal Evaluation	30	Semester End Examination	70	Total Marks	100

Course Outcomes		
Upon successful completion of the course, the student will be able to:		
CO1	Describe the fundamental concepts of Data Engineering	L2
CO2	Apply Data Engineering Concepts to develop Data Architecture, Data Generations	L3
CO3	Apply Storage, Ingestion, Modeling and Transformation for Data Engineering	L3
CO4	Analyze the data engineering concepts and measure the performance	L4

Contribution of Course Outcomes towards achievement of Program Outcomes & Strength of correlations (3: Substantial, 2: Moderate, 1: Slight)													
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	2												2	
CO2	2												2	
CO3	3												3	
CO4		3										2		2

Syllabus		
Unit No.	**CONTENTS**	**Mapped CO**
I	**Introduction to Data Engineering:** Definition, Data Engineering Life Cycle, Evolution of Data Engineer, Data Engineering Versus Data Science. **Data Engineering Skills and Activities:** Data Maturity, Data Maturity Model, Skills of a Data Engineer, Business Responsibilities, Technical Responsibilities, Data Engineers and Other Technical Roles.	**CO1**
II	**Data Engineering Life Cycle:** Data Life Cycle Versus Data Engineering Life Cycle, Generation: Source System, Storage, Ingestion, Transformation, Serving Data. **Major undercurrents across the Data Engineering Life Cycle:** Security, Data Management, DataOps, Data Architecture, Orchestration, Software Engineering.	**CO1**
III	**Designing Good Data Architecture:** Enterprise Architecture, Data Architecture, Principles of Good Data Architecture, Major Architecture Concepts, Examples and Types of Data Architecture.	**CO1, CO2, CO4**
IV	**Data Generation in Source Systems:** Sources of Data, Files and Unstructured Data, APIs, Application Databases (OLTP), OLAP, Change Data Capture, Logs, Database Logs, CRUD. **Storage:** Raw Ingredients of Data Storage, Data Storage Systems, Data Engineering Storage Abstractions, Data warehouse, Data Lake, Data Lakehouse.	**CO1, CO3, CO4**
V	**Ingestion:** Data Ingestion, Key Engineering considerations for the Ingestion Phase, Batch Ingestion Considerations, Message and Stream Ingestion Considerations, Ways to Ingest Data.	**CO1, CO3, CO4.**

Learning Resources
Text Books
1.Fundamentals of Data Engineering, Joe Reis, Matt Housley, Inc., First Edition, June 2022, O'Reilly Media, ISBN: 9781098108304
Reference Books 1. Data Engineering with Python, Paul Crickard, October 2020, Packt Publishing. 2. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Ralph Kimball, Margy Ross, 3rd Edition, 2013, Wiley. 3. Data Pipelines Pocket Reference: Moving and Processing Data for Analytics, James Densmore, 1st Edition,2021, O'Reilly Media.
E-Resources & other digital material

1) https://in.video.search.yahoo.com/search/video?fr=mcafee&p=yoututbe+videos+on+data+engineering&type=E211IN826G0#action=view&id=5&vid=83299bf8ce6d2d14eacd02d82655b918