PVP 23

PRASAD V. POTLURI SIDDHARTHA INSTITUTE OF TECHNOLOGY
(Autonomous)
Kanuru, Vijayawada-520007

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE)
II B. Tech – II Semester

Data Engineering Lab

Course Code	23DS3451	**Year**	II	**Semester**	II
Course Category	PCC Lab	**Branch**	CSE(DATA SCIENCE)	**Course Type**	Practical
Credits	1.5	**L-T-P**	0-0-3	**Prerequisites**	Python Programming
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	Demonstrate proficiency in reading, writing, and parsing data from various file formats, including text, CSV, JSON, XML, and binary files, and submit comprehensive documentation.	L2
CO2	Apply Python libraries and techniques, such as Regular Expressions, NumPy, and Pandas, to manipulate and analyze structured and unstructured data for effective data preprocessing.	L3
CO3	Analyze and evaluate data workflows by integrating Python with databases to perform CRUD operations and leveraging tools like Apache NiFi for data ingestion and transformation.	L4
CO4	Design and evaluate end-to-end data processing pipelines that incorporate file handling, data parsing, analysis, and storage, using relevant Python libraries and data engineering tools.	L5

Contribution of Course Outcomes towards achievement of Program Outcome & Strength of correlation (3: High, 2: Medium, 1: Low)

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**	**PSO1**	**PSO2**
CO1	2									2				
CO2	3													

CO3		3									2		
CO4				3							2		

Syllabus		
EXP. NO	**CONTENTS**	**Mapped CO**
1	Reading & Writing Files (a) Reading and writing text files. (b) Write a Python program to write a CSV file. (c) Write a program to write an XML file. (d) Write a program to write & read a JSON file. (e) Reading and writing binary files.	CO1, CO2, CO3, CO4
2	Parsing Files (a) Parsing text files in Python. (b) Parsing CSV files. (c) Parsing HTML files. (d) XML. (e) JSON.	CO1, CO2, CO3, CO4
3	Write a Python program for searching, splitting, and replacing strings based on pattern matching using Regular Expressions.	CO1, CO2, CO3, CO4
4	Write a program to create NumPy arrays of different shapes and from different sources, reshape and slice arrays, add array indexes, and apply arithmetic, logic, and aggregation functions to some or all array elements.	CO1, CO2, CO3, CO4
5	Write a program to use the pandas data structures: i. Single level & hierarchical indexing ii. Handling missing data iii. Reading data & writing data iv. Arithmetic and Boolean operations on columns & tables v. Database type operations & plotting	CO1, CO2, CO3, CO4
6	Write a Python program to connect to a database & perform CRUD operations.	CO1, CO2, CO3, CO4
7	Write a Python program to create multi-dimensional NumPy arrays & operations.	CO1, CO2, CO3, CO4
8	Create a Python MongoDB client using the Python module pymongo. Using a collection object practice functions for inserting, searching, removing, updating, replacing, and aggregating documents, as well as for creating indexes.	CO1, CO2, CO3, CO4
9	Introduction to DE tools: (a) Apache NiFi installation & configuration (b) Simple operations on NiFi	CO1, CO2, CO3, CO4

10	**Capstone project:** Build a complete data processing and management pipeline. The system will integrate reading, writing, parsing, analyzing, and storing data across various formats while leveraging regular expressions, NumPy, Pandas, databases, and Apache NiFi.	CO1, CO2, CO3, CO4
Learning Resources		
Text Books		
1.Fundamentals of Data Engineering, Joe Reis, Matt Housley, Inc., June 2022, O'Reilly Media, ISBN: 9781098108304		
References		
1. Data Engineering with Python, Packt Publishing Paul Crickard, October 2020.		