# Cover Page

# 23CD22P1 - DATA ENGINEERING LAB

(CSE(DS))

|  |  |  |  |
| --- | --- | --- | --- |
| Course Category: | Professional Core | Credits: | 1.5 |
| Course Type: | Practical | Lecture-Tutorial-Practical: | 0-0-3 |
| Prerequisite: | Familiarity with SQL and **NoSQL databases**  Understanding of **data warehousing concepts** and **ETL process.** | Sessional Evaluation:  Univ. Exam Evaluation:  Total Marks: | 30  70  100 |
| Objectives: | **Upon successful completion of the course, the students will be able to:** | | |
| * The main objective of this course is to teach how build data engineering infrastructure and data pipelines. | | |

|  |  |  |
| --- | --- | --- |
| Course Outcomes | **Upon successful completion of the course, the students will be able to:** | |
| CO1 | Analyze and troubleshoot data engineering problems using systematic approaches. (L5) |
| CO2 | Work effectively in teams to solve data engineering challenges and deliver projects on time. (L6) |
| Course Content | Experiments   1. Installing and configuring Apache NiFi, Apache Airflow 2. Installing and configuring Elasticsearch, Kibana, PostgreSQL, pgAdmin 3. Reading and Writing files    1. Reading and writing files in Python    2. Processing files in Airflow    3. NiFi processors for handling files    4. Reading and writing data to databases in Python    5. Databases in Airflow    6. Database processors in NiFi 4. Working with Databases    1. Inserting and extracting relational data in Python    2. Inserting and extracting NoSQL database data in Python    3. Building database pipelines in Airflow    4. Building database pipelines in NiFi 5. Cleaning, Transforming and Enriching Data    1. Performing exploratory data analysis in Python    2. Handling common data issues using pandas    3. Cleaning data using Airflow 6. Building the Data Pipeline 7. Building a Kibana Dash Board 8. Perform the following operations    1. Staging and validating data    2. Building idempotent data pipelines    3. Building atomic data pipelines 9. Version Control with the NiFi Registry    1. Installing and configuring the NiFi Registry    2. Using the Registry in NiFi    3. Versioning your data pipelines    4. Using git-persistence with the NiFi Registry 10. Monitoring Data Pipelines     1. Monitoring NiFi in the GUI     2. Monitoring NiFi using processors     3. Monitoring NiFi with Python and the REST API 11. Deploying Data Pipelines     1. Finalizing your data pipelines for production     2. Using the NiFi variable registry     3. Deploying your data pipelines 12. Building a Production Data Pipeline     1. Creating a test and production environment     2. Building a production data pipeline     3. Deploying a data pipeline in production | |
| Text Books &  References  Books | **REFERENCE BOOKS:**   1. Paul Crickard , Data Engineering with Python,Packt Publishing, October 2020. | |