Nagarjuna College of Engineering & Technology, Bengaluru

An Autonomous Institute, Affiliated to VTU Belagavi

Scheme & Syllabus
of 2022 Batch
VII Semester

As per the NEP 2020 Guidelines,
Choice-Based Credit System
&
Outcome-Based Education

CSE (Data Science)

w.e.f.
Academic Year 2025-2026

Vision:

To build a strong technical environment and foster leadership and problem-solving abilities in the domain of Data Science, creating professionals capable of addressing social and technical challenges.

Mission:

1. To equip and expose students with the latest tools and technologies.
2. To instill critical problem-solving capabilities, leadership qualities, research capabilities and to prepare them for global challenges.
3. To establish state-of-the-art laboratories and foster collaborations with leading industries in the field of Data Science.

PROGRAM OUTCOMES (POs): Graduates of the Computer Science andEngineering – Data Science Program will be able to achieve the following

PO1: Engineering Knowledge: Apply knowledge of mathematics, natural science, computing, engineering fundamentals and an engineering specialization as specified in Washington Accord Knowledge 1 (WK1) to Washington Accord Knowledge 4 (WK4) respectively to develop the solution of complex engineering problems.

PO2: Problem Analysis: Identify, formulate, review research literature and analyze complex engineering problems reaching substantiated conclusions with consideration for sustainable development. (WK1 to WK4)

PO3: Design/Development of Solutions: Design creative solutions for complex engineering problems and design/develop systems/components/processes to meet identified needs with consideration for the public health and safety, whole-life cost, net zero carbon, culture, society and environment as required. (WK5)

PO4: Conduct Investigations of Complex Problems: Conduct investigations of complex engineering problems using research-based knowledge including design of experiments, modelling, analysis & interpretation of data to provide valid conclusions. (WK8)

PO5: Engineering Tool Usage: Create, select and apply appropriate techniques, resources and modern engineering & IT tools, including prediction and modelling recognizing their limitations to solve complex engineering problems. (WK2 and WK6)

PO6: The Engineer and The World: Analyze and evaluate societal and environmental aspects while solving complex engineering problems for its impact on sustainability with reference to economy, health, safety, legal framework, culture and environment. (WK1, WK5, and WK7).

PO7: Ethics: Apply ethical principles and commit to professional ethics, human values, diversity and inclusion; adhere to national & international laws. (WK9)

PO8: Individual and Collaborative Team work: Function effectively as an individual, and as a member or leader in diverse/multi-disciplinary teams.

PO9: Communication: Communicate effectively and inclusively within the engineering community and society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations considering cultural, language, and learning differences

PO10: Project Management and Finance: Apply knowledge and understanding of engineering management principles and economic decision-making and apply these to one's own work, as a member and leader in a team, and to manage projects and in multidisciplinary environments.

PO11: Life-Long Learning: Recognize the need for, and have the preparation and ability for

i) Independent and life-long learning.
ii) Adaptability to new and emerging technologies and
iii) Critical thinking in the broadest context of technological change. (WK8)

Program Specific Outcome (PSO)

PSO1: Analyze complex computing problems and apply to derive appropriate solutions.

PSO2: Design, implement, and evaluate database-oriented, computing-based solutions that address a broad range of requirements in the field of Data Science.

PSO3: Communicate and work effectively within diverse teams and professional environments.

Program Educational Objectives (PEOs)

PEO1: To work as Data Scientist with an ability to solve wide range of computational problems.

PEO2: To work effectively in a diverse and multi-disciplinary field, as a team member or leader to solve the societal problems.

PEO3: Engage in self-directed and lifelong learning, continuously updating their skills by adapting emerging techniques, advancing in research and higher studies.

NAGARJUNA COLLEGE OF ENGINEERING & TECHNOLOGY, BENGALURU

B.E. in CSE (Data Science)

Scheme of Teaching and Examinations 2022

Outcome-Based Education (OBE) and Choice Based Credit System

(CBCS)(Effective from the academic year 2025 - 26)

VII SEMESTER

Sl. No	Course and Course Code		Course Title	Teaching Department (TD) and Question Paper Setting Board (PSB)	Teaching Hours /Week				Examination				Credits
					Theory Lectur	Tutorial	Practical /Drawing	Self -Study	Duration in hours	CIE Marks	SEE Marks	Total Marks	
					L	T	P	S					
1	IPCC	22CDI71	Advanced Data Warehousing with Snowflake	TD:CD PSB:CD	3	0	2	0	3	50	50	100	4
2	IPCC	22CDI72	Statistical Machine Learning for Data Science	TD:CD PSB:CD	3	0	2	0	3	50	50	100	4
3	PCC	22CDT73	Cryptography & Network Security	TD:CD PSB:CD	4	0	0	0	3	50	50	100	4
4	PEC	22XX74X	Professional Elective Course	TD:CD PSB:CD	3	0	0	0	3	50	50	100	3
5	OEC	22XX75X	Open Elective Course	TD:CD PSB:CD	3	0	0	0	3	50	50	100	3
6	PROJ	22CDP76	Major Project Phase – II	TD:CD PSB:CD	0	0	12	0	3	100	100	200	6
									Total	**350**	**350**	**700**	**24**

Professional Elective Course

22CDT741	Scalable Data System	22CDT743	Parallel Programming
22CDT742	Business Analytics	22CDT744	Deep Learning

Open Elective Course

22CDO751	Introduction to Data Science	22CDO753	Software Engineering
22CDO752	Introduction to Algorithms	22CDO754	Data Management

PCC: Professional Core Course, **PCCL**: Professional Core Course laboratory, **PEC**: Professional Elective Course, **OEC**: Open Elective Course PR: Project Work, **L:** Lecture, **T**: Tutorial, **P**: Practical **S= SDA**: Skill Development Activity, **CIE**: Continuous Internal Evaluation, **SEE**: Semester End Evaluation. **TD-** Teaching Department, **PSB**: Paper Setting department, **OEC**: Open Elective Course, **PEC**: Professional Elective Course. **PROJ**: Project work

Note: VII and VIII semesters of IV years of the program

(1) Institutions can swap the VII and VIII Semester Schemes of Teaching and Examinations to accommodate research internships/ industry internships after the VI semester**.**

(2) Credits earned for the courses of VII and VIII Semester Scheme of Teaching and Examinations shall be counted against the corresponding semesters whether the VII or VIII semesters is completed during the beginning of the IV year or the later part of IV years of the program.

Professional Elective Courses (PEC): A professional elective (PEC) course is intended to enhance the depth and breadth of educational experience in the Engineering and Technology curriculum. Multidisciplinary courses that are added supplement the latest trend and advanced technology in the selected stream of engineering. Each group will provide an option to select one course. The minimum number of students' strengths for offering professional electives is 10. However, this conditional shall not be applicable to cases where the admission to the program is less than 10.

Open Elective Courses:

Students belonging to a particular stream of Engineering and Technology are not entitled to the open electives offered by their parent Department. However, they can opt for an elective offered by other Departments, provided they satisfy the prerequisite condition if any. Registration to open electives shall be documented under the guidance of the Program Coordinator/ Advisor/Mentor. The minimum numbers of students' strength for offering Open Elective Course is 10. However, this condition shall not be applicable to class where the admission to the program is less than 10.

PROJECT WORK (21XXP75): The objective of the Project work is

(i) To encourage independent learning and the innovative attitude of the students.

(ii) To develop interactive attitude, communication skills, organization, time management, and presentation skills.

(iii) To impart flexibility and adaptability.

(iv) To inspire team working.

(v) To expand intellectual capacity, credibility, judgment and intuition.

(vi) To adhere to punctuality, setting and meeting deadlines.

(vii) To install responsibilities to oneself and others.

(viii) To train students to present the topic of project work in a seminar without any fear, face the audience confidently, enhance communication skills, involve in group discussion to present and exchange ideas.

CIE procedure for Project Work:

(1) Single discipline: The CIE marks shall be awarded by a committee consisting of the Head of the concerned Department and two senior faculty members of the Department, one of whom shall be the Guide.

The CIE marks awarded for the project work, shall be based on the evaluation of the project work Report, project presentation skill, and question and answer session in the ratio 50:25:25. The marks awarded for the project report shall be the same for all the batch mates.

ADVANCED DATA WAREHOUSING WITH SNOWFLAKE

Course Code	**22CDI71**	**CIE Marks**	**50**
Teaching Hours/Week (L:T:P: S)	**3:0:2:0**	**SEE Marks**	**50**
Total Hours of Pedagogy	**40 + 13 Lab Slots**	**Total Marks**	**100**
Credits	**04**	**Exam Hours**	**03**

Course Objectives:

This course will enable to,

1. Understand the fundamentals of cloud data warehousing and Snowflake's architecture, features, and benefits.
2. Learn data ingestion techniques, file formats, and automated data loading in Snowflake.
3. Apply Snowflake SQL for structured and semi-structured data analysis, including advanced querying.
4. Design and optimize data models, query performance, and cost management in Snowflake.
5. Integrate Snowflake with Python, Snowpark, and modern data pipelines for advanced analytics and real-world applications.

Teaching-Learning Process (General Instructions)

These are sample Strategies; which teachers can use to accelerate the attainment of the various course outcomes.

1. Lecturer method (L) need not to be only a traditional lecture method, but alternative effective teaching methods could be adopted to attain the outcomes.
2. Use of Video/Animation to explain functioning of various concepts.
3. Encourage collaborative (Group Learning) Learning in the class.
4. Ask at least three HOT (Higher order Thinking) questions in the class, which promotes critical thinking.
5. Adopt Problem Based Learning (PBL), which fosters students 'Analytical skills, develop design thinking skills such as the ability to design, evaluate, generalize, and analyze information rather than simply recall it.
6. Introduce Topics in manifold representations.
7. Show the different ways to solve the same problem with different circuits/logic andencourage the students to come up with their own creative ways to solve them.
8. Discuss how every concept can be applied to the real world - and when that's possible, it helps improve the students' understanding.

Module – I: Introduction to Snowflake and Cloud Data Warehousing

- Introduction to Cloud Data Warehousing
- Snowflake Overview: Features & Benefits
- Snowflake Architecture: Storage, Compute, Services
- Editions and Pricing Model
- Accessing Snowflake via Web UI, CLI, and Snowsight
- Role-based Access Control (RBAC) Basics
- Understanding Warehouses and Databases

08 Hours

Module – II: Data Loading and File Formats

- Data Loading Techniques (Web UI, COPY INTO)
- File Formats: CSV, JSON, Parquet, Avro, XML
- Using Staging Areas (Internal & External)
- Creating and Managing File Formats
- Working with Snowpipe (Automated Data Loading)
- Validating and Troubleshooting Data Loads

08 Hours

Module – III: SQL in Snowflake for Data Analysis

- Snowflake SQL Basics
- Complex Queries: Joins, CTEs, Window Functions
- Semi-structured Data Querying: JSON, VARIANT
- Views and Materialized Views
- Time Travel and Fail-safe
- Working with Temporary and Transient Tables

08 Hours

Module – IV: Data Modeling and Performance Optimization

- Schema Design Approaches (Star, Snowflake)
- Clustering Keys and Partitioning Strategies
- Query Performance Tuning and Best Practices
- Query Profile and Execution Plan
- Data Sharing and Secure Data Exchange
- Cost Optimization Techniques

08 Hours

Module – V: Snowflake with Python and Advanced Analytics

- Snowpark for Python Overview
- Creating and Using UDFs & Stored Procedures
- Integrating Snowflake with Python (Connector)
- ML Model Inference in Snowflake
- Streamlit for Data Science Apps (Optional)
- Role of Snowflake in a Modern Data Pipeline

08 Hours

Teaching-Learning Process for all modules	Chalk and board, Active Learning, PPT Based presentation, Video

LIST OF EXPERIMENTS

Lab No.	Title	Objective
Lab 1	**Environment Initialization and Role Hierarchy Setup**	To create users, roles, warehouses, databases, and schemas. Implement basic RBAC model with role grants.
Lab 2	**Virtual Warehouse Management and Optimization**	To configure warehouses with scaling policies, monitor usage, and practice suspend/resume operations.
Lab 3	**Data Retention with Time Travel and Fail-safe**	To understand data recovery using Time Travel. Create and manage transient and temporary tables.
Lab 4	**User and Role Administration**	To create new users, assign custom roles, and grant/revoke permissions across Snowflake objects.
Lab 5	**Secure Data Loading from Internal Stage**	To load structured files (CSV/JSON) into tables using internal staging and COPY INTO command.
Lab 6	**Monitoring Query Performance and Session Activity**	To use Snowsight, QUERY_HISTORY, and WAREHOUSE_LOAD_HISTORY to analyze performance and usage.
Lab 7	**Automated Data Ingestion using Snowpipe**	To configure internal staging with Snowpipe for automatic file ingestion. Monitor load history.
Lab 8	**Secure Views and Data Protection Techniques**	To implement row access policies, masking policies, and secure views for controlled access.
Lab 9	**Data Analysis with SQL (Joins, Aggregation, Window Functions)**	To write analytical queries using joins, CTEs, aggregations, and advanced SQL features.
Lab 10	**Python-Snowflake Integration using Connector**	To connect to Snowflake using Python, execute SQL commands, and fetch query results programmatically.

Course Outcomes:

On completion of this course, the students will be able to,

CO1: Understand the architecture and core features of Snowflake.
CO2: Learn how to load, query, and analyze structured and semi-structured data.
CO3: Leverage Snowflake for scalable analytics and data science pipelines.
CO4: Perform transformations using SQL and Snowflake scripting.
CO5: Integrate Snowflake with external tools like Python, BI, and ML frameworks.

Assessment Details (both IAT and SEE)

Theory Component	IAT-1 after completion 45 to 50% Syllabus	25 Marks
	IAT-2 after completion 95 to 100% Syllabus	25 Marks
	Average of two IATs	**25 Marks**
	Total 25 Marks : Reduced to 15 Marks	
	CCE-1	25 Marks
	CCE-2	25 Marks
	Average of two CCEs	**25 Marks**
	Total 25 Marks : Reduced to 10 Marks	
Lab Component	Lab Record and execution of programs	**15 Marks**
	Lab Test at the end of 15th week	**10 Marks**
	Total	**25 Marks**
Grand Total of IAT Marks		**50 Marks**
Obtaining 40% of marks in both theory and lab component is essential for appearing for SEE		

Textbooks:

1. Joyce Avila, Snowflake: The Definitive Guide - Architecting, Designing, and Deploying on the Snowflake Data Cloud (Grayscale Indian Edition), 2022.
2. Serge Gershkovich, Kent Graziano, Data Modeling with Snowflake: A practical guide to accelerating Snowflake development using universal data modeling techniques, 2023.

STATISTICAL MACHINE LEARNING FOR DATA SCIENCE

Course Code	**22CDI72**	**CIE Marks**	**50**
Teaching Hours/Week (L:T:P: S)	**3:0:2:0**	**SEE Marks**	**50**
Total Hours of Pedagogy	**40 + 13 Lab Slots**	**Total Marks**	**100**
Credits	**04**	**Exam Hours**	**03**

Course Objective

1. Understand Exploratory Data Analysis
2. Explain Data and Sampling Distributions
3. To Analyse Statistical experiments and perform significance testing
4. To demonstrate how to perform regression analysis on the data
5. Explain Discriminant Analysis on the data.

Teaching-Learning Process (General Instructions)

These are sample Strategies; which teachers can use to accelerate the attainment of the various course outcomes.

1. Lecturer method (L) need not to be only a traditional lecture method, but alternative effective teaching methods could be adopted to attain the outcomes.
2. Use of Video/Animation to explain functioning of various concepts.
3. Encourage collaborative (Group Learning) Learning in the class.
4. Ask at least three HOT (Higher order Thinking) questions in the class, which promotes critical thinking.
5. Adopt Problem Based Learning (PBL), which fosters students 'Analytical skills, develop design thinking skills such as the ability to design, evaluate, generalize, and analyze information rather than simply recall it.
6. Introduce Topics in manifold representations.
7. Show the different ways to solve the same problem with different circuits/logic and encourage the students to come up with their own creative ways to solve them.
8. Discuss how every concept can be applied to the real world - and when that's possible, it helps improve the students' understanding.

Module - I

EXPLORATORY DATA ANALYSIS: Elements of Structured Data, Rectangular Data, Estimates of Location, Mean, Median and Robust Estimates, Example, Estimates of Variability-Standard Deviation, Exploring the Data Distribution-Frequency Tables and Histograms Correlation. **08 Hours**

Module - II

Random sampling and bias, selection bias, The Bootstrap, Confidence Intervals, Normal distribution, Binomial distribution, Poisson distribution. **08 Hours**

Module - III

Statistical Experiments and Significance Testing: A/B testing, hypothesis testing, resampling, statistical Significance & p-values, t-tests, multiple testing, degrees of freedom, Multi-arm bandit algorithm, power and sample size. **08 Hours**

Module - IV

Simple Linear Regression, Multiple Linear Regression, Factor variables in regression, interpreting the regression equation, Regression diagnostics, Polynomial and Spline Regression. **08 Hours**

Module - V

Discriminant Analysis: Covariance Matrix, Fisher's Linear discriminant, Logistic Regression-Generalized Linear Models, Interpreting the coefficients and odd ratios, Strategies for Imbalanced Data. **08 Hours**

Teaching-Learning Process for allmodules	Chalk and board, Active Learning, PPT Based presentation, Video

Lab Experiments	
1	A dataset contains the prices of houses in a city. Find the 25th and 75th percentiles and calculate the interquartile range (IQR). How does the IQR help in understanding the price variability?
2	You are given a dataset with categorical variables about customer satisfaction levels (Low, Medium, High) and whether customers made repeat purchases (Yes/No). Create visualizations such as bar plots or stacked bar charts to explore the relationship between satisfaction level and repeat purchases. What can you infer from the data?
3	A dataset contains information about car models, including the engine size (in Liters), fuel efficiency (miles per gallon), and car price. Use a pair plot or correlation matrix to explore the relationships between these variables. Which variables seem to have the strongest relationships, and what might be the practical significance of these findings?
4	You want to estimate the mean salary of software engineers in a country. You take 10 different random samples, each containing 50 engineers, and calculate the sample mean for each. Plot the distribution of these sample means. How does the Central Limit Theorem explain the shape of this sampling distribution, even if the underlying salary distribution is skewed?
5	A researcher conducts an experiment with a sample of 20 participants to determine if a new drug affects heart rate. The sample has a mean heart rate increase of 8 beats per minute and a standard deviation of 2 beats per minute. Perform a hypothesis test using the t-distribution to determine if the mean heart rate increase is significantly different from zero at the 5% significance level.
6	A company is testing two versions of a webpage (A and B) to determine which version leads to more sales. Version A was shown to 1,000 users and resulted in 120 sales. Version B was shown to 1,200 users and resulted in 150 sales. Perform an A/B test to determine if there is a statistically significant difference in the conversion rates between the two versions. Use a 5% significance level.
7	You are comparing the average daily sales between two stores. Store A has a mean daily sales value of $1,000 with a standard deviation of $100 over 30 days, and Store B has a mean daily sales value of $950 with a standard deviation of $120 over 30 days. Conduct a two-sample t-test to determine if there is a significant difference between the average sales of the two stores at the 5% significance level.
8	A company collects data on employees' salaries and records their education level as a categorical variable with three levels: "High School", "Bachelor's", and "Master's". Fit a multiple linear regression model to predict salary using education level (as a factor variable) and years of experience. Interpret the coefficients for the education levels in the regression model.
9	A hospital is using a Poisson regression model (a type of GLM) to predict the number of emergency room visits per week based on patient age and medical history. The model is given by: $\text{Log}(\lambda) = 2.5 - 0.03*\text{Age} + 0.5*\text{condition}$ where λ is the expected number of visits per week, **Age** is the patient's age, and **condition** is a binary variable (1 if the patient has a chronic condition, 0 otherwise). Interpret the coefficients of Age and condition. What is the expected number of visits per week for a 60-year-old patient with a chronic condition? How would the expected number of visits change if the patient did not have a chronic condition?
10	A bakery claims that its new cookie recipe is lower in calories compared to the old recipe, which had a mean calorie count of 200. You sample 40 new cookies and find a mean of 190 calories with a standard deviation of 15 calories. Perform a one-tailed t-test to determine if the new recipe has significantly fewer calories at a 5% significance level.

Course Outcomes:

Upon completion of this course, the students will be able to,

CO1: Analyse data sets using techniques to estimate variability, exploring distributions, and investigating relationships between variables.

CO2: Apply random sampling, confidence intervals, and recognize various data distributions on datasets.

CO3: Perform significance testing and identify statistical significance.

CO4: Apply regression analysis for prediction, interpret regression equations, and assess regression diagnostics.

CO5: Perform discriminant analysis on the varieties of datasets.

Assessment Details (both IAT and SEE)

Theory Component	IAT-1 after completion 45 to 50% Syllabus	25 Marks
	IAT-2 after completion 95 to 100% Syllabus	25 Marks
	Average of two IATs	**25 Marks**
	Total 25 Marks : Reduced to 15 Marks	
	CCE-1	25 Marks
	CCE-2	25 Marks
	Average of two CCEs	**25 Marks**
	Total 25 Marks : Reduced to 10 Marks	
Lab Component	Lab Record and execution of programs	**15 Marks**
	Lab Test at the end of 15th week	**10 Marks**
	Total	**25 Marks**
Grand Total of IAT Marks		**50 Marks**
Obtaining 40% of marks in both theory and lab component is essential for appearing for SEE		

Textbooks:

1. Peter Bruce, Andrew Bruce and Peter Gadeck, "Practical Statistics for Data Scientists", 2nd edition, O'Reilly Publications, 2020.

Web links and Video Lectures (e-Resources):

1. Statistical learning for Reliability Analysis: https://nptel.ac.in/courses/106105239
2. Engineering Statistics: https://nptel.ac.in/courses/127101233

CRYPTOGRAPHY & NETWORK SECURITY

Course Code	**22CDT73**	**CIE Marks**	**50**
Teaching Hours/Week (L:T:P: S)	**4:0:0:0**	**SEE Marks**	**50**
Total Hours of Pedagogy	**50**	**Total Marks**	**100**
Credits	**04**	**Exam Hours**	**03**

Course objectives:

1. Understand the basics of Cryptography concepts, Security and its principle.
2. To analyse different Cryptographic Algorithms.
3. To illustrate public and private key cryptography.
4. To understand the key distribution scenario and certification.
5. To understand approaches and techniques to build protection mechanism in order to secure computer networks

Teaching-Learning Process (General Instructions)

These are sample Strategies; which teachers can use to accelerate the attainment of the various course outcomes.

1. Lecturer method (L) need not to be only a traditional lecture method, but alternative effective teaching methods could be adopted to attain the outcomes.
2. Use of Video/Animation to explain functioning of various concepts.
3. Encourage collaborative (Group Learning) Learning in the class.
4. Ask at least three HOT (Higher order Thinking) questions in the class, which promotes critical thinking.
5. Adopt Problem Based Learning (PBL), which fosters students 'Analytical skills, develop design thinking skills such as the ability to design, evaluate, generalize, and analyze information rather than simply recall it.
6. Introduce Topics in manifold representations.
7. Show the different ways to solve the same problem with different circuits/logic andencourage the students to come up with their own creative ways to solve them.
8. Discuss how every concept can be applied to the real world - and when that's possible, it helps improve the students' understanding.

Module - I

A model for Network Security, Classical encryption techniques: Symmetric cipher model, Substitution ciphers-Caesar Cipher, Monoalphabetic Cipher, Playfair Cipher, Hill Cipher, Polyalphabetic Ciphers, One time pad, Steganography. Block Ciphers and Data Encryption Standards: Traditional Block Cipher structures, data Encryption Standard (DES), A DES Example, The strength of DES, Block cipher design Principles.

Chapter 1: 1.8 Chapter 3: 3.1, 3.2, 3.5 Chapter 4: 4.1, 4.2, 4.3, 4.4, 4.5 **10 Hours**

Module - II

Pseudorandom number Generators: Linear Congruential Generators, Blum Blum Shub Generator. Public key cryptography and RSA: Principles of public key cryptosystems-Public key cryptosystems, Applications for public key cryptosystems, Requirements for public key cryptography, Public key Cryptanalysis, The RSA algorithm: Description of the Algorithm, Computational aspects, The Security of RSA. Diffie-Hellman key exchange: The Algorithm, Key exchange Protocols, Man-in-the-middle Attack, Elliptic Curve Cryptography: Analog of Diffie-Hellman key Exchange, Elliptic Curve Encryption/Decryption, Security of Elliptic Curve Cryptography.

Chapter 8: 8.2 Chapter 9: 9.1, 9.2 Chapter 10: 10.1, 10.4 **10 Hours**

Module - III

Applications of Cryptographic Hash functions, Two simple Hash functions, Key management and distribution: Symmetric key distribution using symmetric encryption, Symmetric key distribution using asymmetric encryption, Distribution of public keys, X.509 Certificates, Public Key Infrastructures.

Chapter 11: 11.1, 11.2 Chapter 14: 14.1, 14.2, 14.3, 14.4, 14.5 **10 Hours**

Module - IV

User Authentication: Remote user authentication principles**,** Kerberos**,** Remote user authentication using asymmetric encryption. Web security consideration, Transport layer security. Email Threats and comprehensive email security, S/MIME, Pretty Good Privacy.

Chapter 15: 15.1, 15.3, 15.4 Chapter 17: 17.1, 17.2 Chapter 19: 19.3, 19.4, 19.5 **10 Hours**

Module - V

Domainkeys Identified Mail.

IP Security: IP Security overview, IP Security Policy, Encapsulating Security Payload, Combining security associations, Internet key exchange.

Chapter 19: 19.9 Chapter 20: 20.1, 20.2, 20.3, 20.4, 20.5 **10 Hours**

Teaching-Learning Process for all modules	Chalk and board, Active Learning, PPT Based presentation, Video

Course Outcomes:

On completion of this course, the students will be able to,

CO1: Explain the basic concepts of Cryptography and Security aspects.
CO2: Apply different Cryptographic Algorithms for different applications.
CO3: Analyze different methods for authentication and access control.
CO4: Describe key management, key distribution and Certificates.
CO5: Explain about Electronic mail and IP Security.

Assessment Details (both IAT and SEE)

Theory Component	IAT-1 after completion 45 to 50% Syllabus	25 Marks
	IAT-2 after completion 95 to 100% Syllabus	25 Marks
	Average of two IATs	**25 Marks**
	CCE-1	25 Marks
	CCE-2	25 Marks
	Average of two CCEs	**25 Marks**
Grand Total of IAT Marks (min marks 20 / 50)		**50 Marks**
SEE conducted for 100 and scaled down to 50 **(min marks 18/50)**		**50 Marks**
IAT + SEE (min marks 40)		**100 Marks**

Text Books:

1. William stallings, "Cryptography and Network Security", Pearson Publication, Seventh Edition.

Reference Books:

1. Keith M Martin, "Everyday Cryptography", Oxford University Press.
2. V.K Pachghare, "Cryptography and Network Security", PHI, 2nd Edition

SCALABLE DATA SYSTEM			
Course Code	**22CDT741**	**CIE Marks**	**50**
Teaching Hours/Week (L:T:P: S)	**3:0:0:0**	**SEE Marks**	**50**
Total Hours of Pedagogy	**40**	**Total Marks**	**100**
Credits	**03**	**Exam Hours**	**03**

Course Objectives:

This course will enable to,

CLO1: Understand the principles and components of scalable data systems and distributed computing.

CLO2: Analyze and apply appropriate NoSQL data models for large-scale data storage.

CLO3: Design and implement batch and real-time big data processing systems using industry-standard tools.

CLO4: Evaluate the suitability of different data architectures (data lakes, lakehouses) for enterprise-scale needs.

CLO5: Construct and manage data pipelines for end-to-end data engineering workflows.

Teaching-Learning Process (General Instructions)

These are sample Strategies; which teachers can use to accelerate the attainment of the various course outcomes.

1. Lecturer method (L) need not to be only a traditional lecture method, but alternative effective teaching methods could be adopted to attain the outcomes.
2. Use of Video/Animation to explain functioning of various concepts.
3. Encourage collaborative (Group Learning) Learning in the class.
4. Ask at least three HOT (Higher order Thinking) questions in the class, which promotes critical thinking.
5. Adopt Problem Based Learning (PBL), which fosters students 'Analytical skills, develop design thinking skills such as the ability to design, evaluate, generalize, and analyze information rather than simply recall it.
6. Introduce Topics in manifold representations.
7. Show the different ways to solve the same problem with different circuits/logic andencourage the students to come up with their own creative ways to solve them.
8. Discuss how every concept can be applied to the real world - and when that's possible, it helps improve the students' understanding.

Module - I

Introduction to Scalable Systems and Distributed Computing – Characteristics of scalable systems: horizontal vs vertical scaling, Overview of distributed computing and CAP theorem, Consistency, availability, and partition tolerance, Brewer's theorem; Introduction to data partitioning and replication strategies.

08 Hours

Assignments:

- Analyze and compare CAP properties for Google Docs, Amazon DynamoDB, and traditional RDBMS.
- Implement a basic consistent hashing simulation in Python or Java.

Case Study:

Netflix Microservices Architecture: How Netflix handles scalability using distributed services.

Module - II

NoSQL Databases and Data Modeling – Introduction to NoSQL databases: key-value, document, column-family, and graph stores; Data modeling strategies in NoSQL systems; Comparison with relational databases; MongoDB, Cassandra, Redis, and Neo4j overview; Query patterns, indexing, and performance tuning. **08 Hours**

Assignments:

- Design a NoSQL schema for a real-time chat application using MongoDB or Cassandra.

- Compare query performance of MongoDB and PostgreSQL for a document-heavy dataset.

Case Study:
Facebook Messenger and Cassandra: Why Facebook chose Cassandra for message storage and delivery.

Module - III

Big Data Processing with Hadoop Ecosystem – Hadoop Distributed File System (HDFS); MapReduce programming model; Introduction to YARN, Hive, Pig, and HBase; Data ingestion tools: Sqoop, Flume; Batch vs stream processing. **08 Hours**

Assignments:
- Write a MapReduce program to analyze server logs (e.g., count 404 errors).
- Use Hive to process and summarize a dataset (e.g., NYC taxi trip data).

Case Study:
Yahoo and Hadoop: How Yahoo scaled web indexing using the Hadoop ecosystem.

Module - IV

Real-Time Data Streaming and Processing – Stream processing vs batch processing; Apache Kafka architecture and message queues; Apache Storm, Apache Flink, Apache Spark Streaming; Real-time analytics and event-driven architectures; Windowing, watermarking, and stateful processing. **08 Hours**

Assignments:
- Set up a Kafka producer-consumer system for Twitter stream ingestion.
- Build a Spark Streaming application to compute rolling averages over time.

Case Study:
Uber's Real-Time Analytics Platform: How Uber uses Apache Flink and Kafka for dynamic pricing.

Module - V

Data Lake Architecture and Cloud-Native Data Engineering - Data lake vs data warehouse; Lakehouse architecture (Delta Lake, Apache Iceberg); Cloud-native data platforms: AWS (S3, EMR, Athena), Google BigQuery, Azure Synapse; Orchestration with Apache Airflow, dbt basics; Data governance, quality, and cataloging tools. **08 Hours**

Assignments:
- Design a data lake architecture on AWS for a retail business.
- Build a data pipeline using Airflow to extract, transform, and load weather data into BigQuery.

Case Study:
Spotify's Data Lake Strategy: Managing terabytes of event data with data lakes and pipelines

Teaching-Learning Process for allmodules	Chalk and board, Active Learning, PPT Based presentation, Video, Case study.

Course Outcomes:
On completion of this course, the students will be able to,

CO1: Explain the fundamentals of scalable and distributed data systems.
CO2: Apply NoSQL databases and design data models for high-volume data applications.
CO3: Implement big data workflows using Hadoop, Hive, and MapReduce.
CO4: Develop real-time stream processing solutions using Kafka and Spark.
CO5: Design cloud-native data lakes and orchestrate pipelines using tools like Airflow.

Assessment Details (both IAT and SEE)

Theory Component	IAT-1 after completion 45 to 50% Syllabus	25 Marks
	IAT-2 after completion 95 to 100% Syllabus	25 Marks
	Average of two IATs	**25 Marks**
	CCE-1	25 Marks
	CCE-2	25 Marks
	Average of two CCEs	**25 Marks**
Grand Total of IAT Marks (min marks 20 / 50)		**50 Marks**
SEE conducted for 100 and scaled down to 50 **(min marks 18/50)**		**50 Marks**
IAT + SEE (min marks 40)		**100 Marks**

Textbooks:

1. Nathan Marz and James Warren ‑Big Data: Principles and Best Practices of Scalable Real-Time Data Systems, Manning Publications.
2. Martin Kleppmann ‑Designing Data-Intensive Applications, O'Reilly Media.
3. Pramod J. Sadalage and Martin Fowler - NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence, Addison-Wesley

Reference Books:

1. Tyler Akidau, Slava Chernyak, and Reuven Lax‑Streaming Systems: The What, Where, When, and How of Large-Scale Data Processing, O'Reilly Media.
2. Hasso Plattner and Alexander Zeier ‑Cloud Data Management, Springer.
3. Tom White ‑Hadoop: The Definitive Guide, O'Reilly Media
4. Bill Inmon, Mary Levins - The Data Lakehouse: Fundamentals and Best Practices, Technics Publications

Web links and Video Lectures (e-Resources):

1. Scalable Data Science : https://onlinecourses.nptel.ac.in/noc20_cs61/preview

BUSINESS ANALYTICS

Course Code	**22CDT742**	**CIE Marks**	**50**
Teaching Hours/Week (L:T:P: S)	**3:0:0:0**	**SEE Marks**	**50**
Total Hours of Pedagogy	**40**	**Total Marks**	**100**
Credits	**03**	**Exam Hours**	**03**

Course Learning objectives:

1. Explain the Business Intelligence, Analytics and Decision Support system
2. List the technologies for Decision making, Automated decision systems
3. Explain sentiment analysis techniques
4. Illustrate Multi-criteria Decision making systems, predictive modelling techniques

Teaching-Learning Process (General Instructions)

These are sample Strategies; which teachers can use to accelerate the attainment of the various course outcomes.

1. Lecturer method (L) need not to be only a traditional lecture method, but alternative effective teaching methods could be adopted to attain the outcomes.
2. Use of Video/Animation to explain functioning of various concepts.
3. Encourage collaborative (Group Learning) Learning in the class.
4. Ask at least three HOT (Higher order Thinking) questions in the class, which promotes critical thinking.
5. Adopt Problem Based Learning (PBL), which fosters students 'Analytical skills, develop design thinking skills such as the ability to design, evaluate, generalize, and analyze information rather than simply recall it.
6. Introduce Topics in manifold representations.
7. Show the different ways to solve the same problem with different circuits/logic and encourage the students to come up with their own creative ways to solve them.
8. Discuss how every concept can be applied to the real world - and when that's possible, it helps improve the students' understanding.

Module – I: An Overview of Business Intelligence, Analytics, and Decision Support

Information Systems Support for Decision Making, An Early Framework for Computerized Decision Support, The Concept of Decision Support Systems, A Framework for Business Intelligence, Business Analytics Overview, Brief Introduction to Big Data Analytics.

08 Hours

Module – II : Decision Making

Introduction and Definitions, Phases of the Decision, Making Process, The Intelligence Phase, Design Phase, Choice Phase, Implementation Phase, Decision Support Systems Capabilities, Decision Support Systems Classification, Decision Support Systems Components**.**

08 Hours

Module – III: Neural Networks and Sentiment Analysis

Basic Concepts of Neural Networks, Developing Neural Network-Based Systems, Illuminating the Black Box of ANN with Sensitivity, Support Vector Machines, A Process Based Approach to the Use of SVM, Nearest Neighbor Method for Prediction, Sentiment Analysis Overview, Sentiment Analysis Applications, Sentiment Analysis Process, Sentiment Analysis, Speech Analytics.

08 Hours

Module – IV: Model-Based Decision Making

Decision Support Systems modeling, Structure of mathematical models for decision support, Certainty, Uncertainty, and Risk, Decision modeling with spreadsheets, Mathematical programming optimization, Decision Analysis with Decision Tables and Decision Trees, Multi-Criteria Decision Making With Pairwise Comparisons.

08 Hours

Module – V: Automated Decision Systems and Expert Systems

Automated Decision Systems, The Artificial Intelligence field, Basic concepts of Expert Systems, Applications of Expert Systems, Structure of Expert Systems, Knowledge Engineering, Development of Expert Systems.

08 Hours

Teaching-Learning Process for all modules	Chalk and talk method / PowerPoint Presentation

Course Outcomes:

On completion of this course, the students will be able to,

CO1: Able to analyze Business Intelligence, Analytics and Decision Support.
CO2: Explain the technologies for Decision making.
CO3: Apply predictive modelling techniques (can be attained through assignment or CIE).
CO4: Apply sentiment analysis techniques (can be attained through assignment or CIE).

Assessment Details (both IAT and SEE)

Theory Component	IAT-1 after completion 45 to 50% Syllabus	25 Marks
	IAT-2 after completion 95 to 100% Syllabus	25 Marks
	Average of two IATs	**25 Marks**
	CCE-1	25 Marks
	CCE-2	25 Marks
	Average of two CCEs	**25 Marks**
Grand Total of IAT Marks (min marks 20 / 50)		**50 Marks**
SEE conducted for 100 and scaled down to 50 **(min marks 18/50)**		**50 Marks**
IAT + SEE (min marks 40)		**100 Marks**

Textbooks:

1. Ramesh Sharda, Dursun Delen, EfraimTurban, J.E.Aronson,Ting-Peng Liang, David King, "Business Intelligence and Analytics: System for Decision Support", 10th Edition, Pearson Global Edition, 2013.

Reference Books:

1. Data Analytics: The Ultimate Beginner's Guide to Data Analytics Paperback – 12 November 2017 by Edward Mize.

Web links and Video Lectures (e-Resources):

1. https://www.youtube.com/watch?v=zbcCdoHeS4w

PARALLEL PROGRAMMING

Course Code	**22CDT743**	**CIE Marks**	**50**
Teaching Hours/Week (L:T:P: S)	**3:0:0:0**	**SEE Marks**	**50**
Total Hours of Pedagogy	**40**	**Total Marks**	**100**
Credits	**03**	**Exam Hours**	**03**

Course Objectives:

This course will enable to,

1. To understand the fundamental concepts of parallel processing and its necessity in modern computing.
2. To explore various parallel architectures and models including shared and distributed memory systems.
3. To develop skills in designing and implementing parallel algorithms.
4. To learn parallel programming constructs using tools such as OpenMP and MPI.
5. To analyze synchronization techniques and load balancing strategies in parallel environments.
6. To apply parallel processing techniques in real-world applications including scientific computing and big data frameworks.

Teaching-Learning Process (General Instructions)

These are sample Strategies; which teachers can use to accelerate the attainment of the various course outcomes.

1. Lecturer method (L) need not to be only a traditional lecture method, but alternative effective teaching methods could be adopted to attain the outcomes.
2. Use of Video/Animation to explain functioning of various concepts.
3. Encourage collaborative (Group Learning) Learning in the class.
4. Ask at least three HOT (Higher order Thinking) questions in the class, which promotes critical thinking.
5. Adopt Problem Based Learning (PBL), which fosters students 'Analytical skills, develop design thinking skills such as the ability to design, evaluate, generalize, and analyze information rather than simply recall it.
6. Introduce Topics in manifold representations.
7. Show the different ways to solve the same problem with different circuits/logic and encourage the students to come up with their own creative ways to solve them.
8. Discuss how every concept can be applied to the real world - and when that's possible, it helps improve the students' understanding.

Module - I

Introduction to parallel programming – Definition and need for parallel processing ; Classifications of parallel computers, Flynn's taxonomy (SISD, SIMD, MISD, MIMD), Interconnection networks, Cache coherence, Shared-memory vs. distributed-memory, Coordinating the processes/threads, Parallel programming models (shared memory, message passing); Performance metrics: Speedup, Efficiency, Scalability. **08 Hours**

Assignment:

- Report comparing Flynn's taxonomy with real-world CPU/GPU examples
- Implement sequential vs. parallel addition of arrays using OpenMP

Case Study:

- Amdahl's Law and its implications
- Google Data Center Architecture

Module - II

Parallel Architectures – Multicore processors, SMP, NUMA systems, Interconnection networks: bus, crossbar, multistage, Cache coherence and memory consistency, Examples: Intel Xeon, NVIDIA CUDA GPUs. **08 Hours**

Assignment:

- Draw and explain interconnection network for 8-core SMP
- Compare UMA vs. NUMA architecture

Case Study:

- NVIDIA CUDA GPU architecture
- Thread scheduling in Intel multi-core processors

Module - III

Parallel Algorithms and Programming – Principles of parallel algorithm design, Decomposition and dependency analysis, Task and data parallelism, Programming tools: OpenMP, MPI basics, Parallel sorting, matrix multiplication. **08 Hours**

Assignment:

- Parallel matrix multiplication using OpenMP/MPI
- Parallel recursive algorithm implementation

Case Study:

- MPI in high-performance clusters
- Parallel Merge Sort analysis

Module - IV

Synchronization and Load Balancing – Critical sections, race conditions, Locks, semaphores, barriers, Deadlock and prevention, Load balancing strategies, Granularity of tasks. **08 Hours**

Assignment:

- Producer-consumer simulation using POSIX threads
- OpenMP scheduling for load balancing

Case Study:

- Load balancing in Hadoop MapReduce
- Thread scheduling in Linux

Module - V

Applications of Parallel Processing - Applications in AI/ML, bioinformatics, cryptography, Big Data frameworks: Hadoop, Spark, Scientific computing and weather forecasting, Cloud and edge computing, Future trends: Quantum, neuromorphic computing. **08 Hours**

Assignment:

- Report on parallelism in TensorFlow/PyTorch
- Spark RDD parallel word count program

Case Study:

- Genome sequencing in cloud clusters
- Deep learning training on GPUs

Teaching-Learning Process for all modules	Chalk and board, Active Learning, PPT Based presentation, Video

Course Outcomes:

On completion of this course, the students will be able to,

CO1: Understand different models of parallel computation and performance metrics.
CO2: Analyze and design parallel architectures and memory models.
CO3: Implement parallel algorithms using suitable programming frameworks.
CO4: Handle synchronization, data sharing, and load balancing issues in parallel systems.
CO5: Apply parallel processing concepts to real-world applications in science and industry.

Assessment Details (both IAT and SEE)

Theory Component	IAT-1 after completion 45 to 50% Syllabus	25 Marks
	IAT-2 after completion 95 to 100% Syllabus	25 Marks
	Average of two IATs	**25 Marks**
	CCE-1	25 Marks
	CCE-2	25 Marks
	Average of two CCEs	**25 Marks**
Grand Total of IAT Marks (min marks 20 / 50)		**50 Marks**
SEE conducted for 100 and scaled down to 50 **(min marks 18/50)**		**50 Marks**
IAT + SEE (min marks 40)		**100 Marks**

Textbooks:

1. Peter S Pacheco, Matthew Malensek – An Introduction to Parallel Programming, second edition, Morgan Kauffman.
2. Michael J Quinn – Parallel Programming in C with MPI and OpenMp, McGrawHill.

Reference Books:

1. Calvin Lin, Lawrence Snyder – Principles of Parallel Programming, Pearson.
2. Barbara Chapman – Using OpenMP: Portable Shared Memory Parallel Programming, Scientific and Engineering Computation.
3. William Gropp, Ewing Lusk – Using MPI:Portable Parallel Programing, Third edition, Scientific and Engineering Computation

Web links and Video Lectures (e-Resources):

1. Introduction to parallel programming: https://nptel.ac.in/courses/106102163

DEEP LEARNING

Course Code	**22CDT744**	**CIE Marks**	**50**
Teaching Hours/Week (L:T:P: S)	**3:0:0:0**	**SEE Marks**	**50**
Total Hours of Pedagogy	**40**	**Total Marks**	**100**
Credits	**03**	**Exam Hours**	**03**

Course Objectives:

This course will enable to,

1. Understand the fundamentals of deep learning.
2. Understanding the working of Convolutional Neural Networks and RNN in decision making.
3. Illustrate the strength and weaknesses of many popular deep learning approaches.
4. Introduce major deep learning algorithms, the problem settings, and their applications to solve real world problems

Teaching-Learning Process (General Instructions)

These are sample Strategies; which teachers can use to accelerate the attainment of the various course outcomes.

1. Lecturer method (L) need not to be only a traditional lecture method, but alternative effective teaching methods could be adopted to attain the outcomes.
2. Use of Video/Animation to explain functioning of various concepts.
3. Encourage collaborative (Group Learning) Learning in the class.
4. Ask at least three HOT (Higher order Thinking) questions in the class, which promotes critical thinking.
5. Adopt Problem Based Learning (PBL), which fosters students 'Analytical skills, develop design thinking skills such as the ability to design, evaluate, generalize, and analyze information rather than simply recall it.
6. Introduce Topics in manifold representations.
7. Show the different ways to solve the same problem with different circuits/logic and encourage the students to come up with their own creative ways to solve them.
8. Discuss how every concept can be applied to the real world - and when that's possible, it helps improve the students' understanding.

Module - I

Introduction: What is a Neural Network?, The Human Brain, Models of a Neuron, Neural Networks Viewed As Directed Graphs, Feedback, Network Architectures, Rosenblatt's Perceptron: Introduction, Perceptron, The Perceptron Convergence Theorem, Relation Between the Perceptron and Bayes Classifier for a Gaussian Environment

08 Hours

Module - II

Multilayer Perceptrons: Introduction, Batch Learning and On-Line Learning, The Back-Propagation Algorithm, XOR Problem, Heuristics for Making the Back- Propagation Algorithm Perform Better, Back Propagation and Differentiation.

08 Hours

Module - III

Regularization for Deep Learning: Parameter Norm Penalties - L2 Parameter Regularization, Dataset Augmentation, Semi-Supervised Learning. **Optimization for Training Deep Models**: Challenges in Neural Network Optimization – III Conditioning, Local Minima, Plateaus, Saddle Points and Other Flat Regions.

08 Hours

Module - IV

Convolution neural networks: The Convolution Operation, Motivation, Pooling, Convolution and Pooling as an Infinitely Strong Prior, Variants of the Basic Convolution Function, Structured Outputs, Data Types, Efficient Convolution Algorithms, Convolutional Networks and the History of Deep Learning.

08 Hours

Module - V

Sequence Modeling: Recurrent and Recursive Nets: Unfolding Computational Graphs, Recurrent Neural Networks, Bidirectional RNNs, Encoder-Decoder Sequence-to- Sequence Architectures, Deep Recurrent Networks, Recursive Neural Networks, The Long Short-Term Memory and Other Gated RNNs.

08 Hours

Teaching-Learning Process for all modules	Chalk and board, Active Learning, PPT Based presentation, Video

Course Outcomes:

On completion of this course, the students will be able to,

CO1: Analyze and interpret the concepts of neural networks relating to artificial intelligence.
CO2: Illustrate the learning processes and their statistical properties.
CO3: Design deep learning models using regularization and convolutional operations.
CO4: Analyze sequential data to build recurrent and recursive models.
CO5: Develop and analyze the applications using Auto encoders

Assessment Details (both IAT and SEE)

Theory Component	IAT-1 after completion 45 to 50% Syllabus	25 Marks
	IAT-2 after completion 95 to 100% Syllabus	25 Marks
	Average of two IATs	**25 Marks**
	CCE-1	25 Marks
	CCE-2	25 Marks
	Average of two CCEs	**25 Marks**
Grand Total of IAT Marks (min marks 20 / 50)		**50 Marks**
SEE conducted for 100 and scaled down to 50 **(min marks 18/50)**		**50 Marks**
IAT + SEE (min marks 40)		**100 Marks**

Textbooks:

1. Simon Haykin, Neural networks and Learning Machines, Third Edition, Pearson, 2016
2. Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning, MIT Press, 2016. https://www.deeplearningbook.org/lecture_slides.html

Reference Books:

1. Tom M Mitchell, ‚Machine Learning', First Edition, McGraw Hill Education, 2013
2. Yegnanarayana, B, Artificial Neural Networks', PHI Learning Pvt. Ltd, 2009.
3. Satish Kumar, ‚Neural Networks: A Classroom Approach', Tata McGraw-Hill Education, 2004.
4. Christopher Bishop, ‚Pattern Recognition and Machine Learning' 2e, Springer, 2006.

Web links and Video Lectures (e-Resources):

1. https://www.youtube.com/watch?v=VyWAvY2CF9c
2. https://www.youtube.com/watch?v=7sB052Pz0sQ
3. https://www.youtube.com/watch?v=Mubj_fqiAv8
4. https://www.coursera.org/learn/neural-networks-deep-learning
5. https://onlinecourses.nptel.ac.in/noc20_cs62/preview

INTRODUCTION TO DATA SCIENCE

Course Code	**22CDO751**	**CIE Marks**	**50**
Teaching Hours/Week (L:T:P: S)	**3:0:0:0**	**SEE Marks**	**50**
Total Hours of Pedagogy	**40**	**Total Marks**	**100**
Credits	**03**	**Exam Hours**	**03**

Course Objectives:

This course will enable students to,

1. Provide a strong foundation for data science and application areas related to it.
2. Learn the process of working with data on large scale.
3. Explore the concepts of Data Processing.
4. Learn basic concepts of Machine Learning.
5. Prepare students for advanced courses in Data Science

Teaching-Learning Process (General Instructions)

These are sample Strategies; which teachers can use to accelerate the attainment of the various course outcomes.

1. Lecturer method (L) need not to be only a traditional lecture method, but alternative effective teaching methods could be adopted to attain the outcomes.
2. Use of Video/Animation to explain functioning of various concepts.
3. Encourage collaborative (Group Learning) Learning in the class.
4. Ask at least three HOT (Higher order Thinking) questions in the class, which promotes critical thinking.
5. Adopt Problem Based Learning (PBL), which fosters students 'Analytical skills, develop design thinking skills such as the ability to design, evaluate, generalize, and analyze information rather than simply recall it.
6. Introduce Topics in manifold representations.
7. Show the different ways to solve the same problem with different circuits/logic andencourage the students to come up with their own creative ways to solve them.
8. Discuss how every concept can be applied to the real world - and when that's possible, it helps improve the students' understanding.

Module – I

Introduction to Data Science: Importance of data Science-Need for Data Science, What is Data Science? Data Science Process, prerequisites for data science, Components of Data Science, Tools and Skills needed.

Statistics: Data Types, Variable Types, Statistics, Sampling Techniques.

[Text Book1:1(1.1,1.2,1.3,1.5,1.6,1.7),2(2.1,2.2,2.3,2.4,2.5)]. **08 Hours**

Module – II

Probability: Probability Theory, Probability types, Probability Distribution Functions, Bayes Theorem.

Data Modeling and Analytics: Data Science Methodology-Analytics for data science, Example of Data Analytics, Data Analytics Life Cycle-Data Discovery, Data Preparation, Model Planning, Model Building, Communicate Results, Operationalization.

[Text Book1:2(2.6,2.7,2.8,2.9),4(4.1,4.2,4.3). **08 Hours**

Module – III

Machine learning –Designing a Learning System, Perspective and Issues in Machine Learning, Supervised learning, Unsupervised learning, Semi- supervised learning, Reinforcement Learning, Role of Machine Learning in Data Science, Data Science vs Machine Learning.

[Text Book2:1.2(1.2.1,1.2.2,1.2.3,1.2.4,1.2.5),

https://www.geeksforgeeks.org/data-science-vs-machine- learning

https://www.zucisystems.com/blog/what-is-the-role-of-machine-learning-in-data-science

08 Hours

Module – IV

Databases for Data Science – SQL-for Data Science, Basic Statistics with SQL, Data Wrangling, Filtering, Joins, Aggregation, Advanced NoSQL for Data Science, Document Databases for Data Science,

Wide Column Databases for Data Science, Graph Databases for Data Science.
TextBook1:3.1(3.1.1,3.1.2,3.1.3), 3.2(3.2.1,3.2.2,3.2.3,3.2.4). **08 Hours**

Module – V

Data Analytics and Text Mining: What is Text Mining?, Process of Text Mining, Difference between Text Mining and Data Mining, Major Text Mining Areas, Text Analytics, Text Analytics Steps, Basic Text Analytics Steps.

Introduction to NLP: Introduction, Major Components of NLP, Stages of NLP, Statistical Processing of NLP, Applications of NLP.

Text Book1:6.1(6.1.1,6.1.1.1,6.1.1.2,6.1.1.3),6.2(6.2.1,6.2.2),6.3(6.3.1,6.3.2,6.3.3,6.3.4) **08 Hours**

Teaching-Learning Process for allmodules	Chalk and talk method / PowerPoint Presentation

Course Outcomes:

On completion of this course, the students will be able to,

CO1: Apply the fundamental concepts of data science.

CO2: Evaluate the data analysis techniques for applications handling large data and Demonstrate the data science process.

CO3: Analyze the concept of machine learning used in the data science process.

CO4: Demonstrate and present the inference using various tools.

CO5: Analyze to think through the ethics surrounding privacy, data sharing.

Assessment Details (both IAT and SEE)

Theory Component	IAT-1 after completion 45 to 50% Syllabus	25 Marks
	IAT-2 after completion 95 to 100% Syllabus	25 Marks
	Average of two IATs	**25 Marks**
	CCE-1	25 Marks
	CCE-2	25 Marks
	Average of two CCEs	**25 Marks**
Grand Total of IAT Marks (min marks 20 / 50)		**50 Marks**
SEE conducted for 100 and scaled down to 50 **(min marks 18/50)**		**50 Marks**
IAT + SEE (min marks 40)		**100 Marks**

Text Books:

1. Fundamentals of Data Science**,** Sanjeev J. Wagh, Manisha S. Bhende, and Anuradha D. Thakare, First edition published 2022 by CRC Press.
2. Machine Learning, Tom Mitchell, McGraw Hill, 1997.

E-Resources:

1. https://www.geeksforgeeks.org/data-science-vs-machine-learning
2. **https://www.zucisystems.com/blog/what-is-the-role-of-machine-learning-in-data-science**

INTRODUCTION TO ALGORITHMS

Course Code	**22CDO752**	**CIE Marks**	**50**
Teaching Hours/Week (L:T:P: S)	**3:0:0:0**	**SEE Marks**	**50**
Total Hours of Pedagogy	**40**	**Total Marks**	**100**
Credits	**03**	**Exam Hours**	**03**

Course Objectives:

This course will enable students to,

CLO1. Explain the methods of analysing the algorithms and to analyze performance of algorithms.
CLO2. State algorithm‘s efficiencies using asymptotic notations.
CLO3. Solve problems using algorithm design methods such as the brute force method, greedy method, divide and conquer, decrease and conquer, transform and conquer, dynamic programming, backtracking and branch and bound.
CLO 4. Choose the appropriate data structure and algorithm design method for a specified application.
CLO 5. Introduce P and NP classes.

Teaching-Learning Process (General Instructions)

These are sample Strategies; which teachers can use to accelerate the attainment of the various course outcomes.

1. Lecturer method (L) need not to be only a traditional lecture method, but alternative effective teaching methods could be adopted to attain the outcomes.
2. Use of Video/Animation to explain functioning of various concepts.
3. Encourage collaborative (Group Learning) Learning in the class.
4. Ask at least three HOT (Higher order Thinking) questions in the class, which promotes critical thinking.
5. Adopt Problem Based Learning (PBL), which fosters students ‘Analytical skills, develop design thinking skills such as the ability to design, evaluate, generalize, and analyze information rather than simply recall it.
6. Introduce Topics in manifold representations.
7. Show the different ways to solve the same problem with different circuits/logic andencourage the students to come up with their own creative ways to solve them.
8. Discuss how every concept can be applied to the real world - and when that's possible, it helps improve the students' understanding.

Module – I

Introduction: What is an Algorithm?, Fundamentals of Algorithm problem solving, Fundamentals of analysis of algorithm efficiency, Asymptotic notations and basic efficiency classes, Mathematical analysis of Recursive and Non Recursive Algorithms.

Brute Force Approaches: Introduction, Selection Sort, Bubble Sort, Sequential Search, Brute force String Matching. **08 Hours**

Module – II

Divide and Conquer: General method, Recurrence equation for divide and conquer, solving it using Master‘s theorem. , Divide and Conquer algorithms and complexity Analysis - Finding the maximum & minimum, Binary search, Merge sort, Quick sort.

Greedy Method-The General Method, Minimum-cost Spanning Tree, Prims Algorithm, Kruskal’s Algorithm, Single Source Shortest Path. **08 Hours**

Module – III

Decrease and Conquer Approach: Introduction, Insertion sort, Depth First Search, Breadth First Search, Topological Sorting.

Transform and Conquer Approach: Introduction,2-3 Trees, Heaps and Heap Sort. **08 Hours**

Module – IV

Dynamic Programming: General method, Warshall’s Algorithm, Floyd’s Algorithm for all pair shortest path, Travelling Salesman Problem, Computing Binomial Coefficient.

Space-Time Tradeoffs: Introduction, Sorting by Counting, Input Enhancement in String Matching-Harspool‘s algorithm. **08 Hours**

Module – V

Limitations of algorithmic power: Decision Trees, P, NP, and NP-Complete Problems. **Backtracking**: N-Queens problem, Sum of subsets problem.
Branch and Bound: Assignment Problem, Knapsack problem
Hashing-Open Hashing, Closed Hashing. **08 Hours**

Teaching-Learning Process for allmodules	Chalk & board, Active Learning, MOOC, Problem based Learning.

Course Outcomes:
On completion of this course, the students will be able to,

CO1: Apply and identify asymptotic notations and basic efficiency classes.
CO2: Solve problems using various techniques like greed and divide and conquer.
CO3: Compute problems using various techniques like decrease-and-conquer and transfer-and-conquer.
CO4: Interpret different algorithms like TSP, Floyd's, etc. to solve real-world problems.
CO5: Design and develop solutions for n-Queens problem, Subset-sum problem, Assignment problem, Knapsack problem etc.

Assessment Details (both IAT and SEE)

Theory Component	IAT-1 after completion 45 to 50% Syllabus	25 Marks
	IAT-2 after completion 95 to 100% Syllabus	25 Marks
	Average of two IATs	**25 Marks**
	CCE-1	25 Marks
	CCE-2	25 Marks
	Average of two CCEs	**25 Marks**
Grand Total of IAT Marks (min marks 20 / 50)		**50 Marks**
SEE conducted for 100 and scaled down to 50 **(min marks 18/50)**		**50 Marks**
IAT + SEE (min marks 40)		**100 Marks**

Text Books:

1. Introduction to the Design and Analysis of Algorithms, Anany Levitin: 2nd Edition, 2009. Pearson.
2. Computer Algorithms/C++, Ellis Horowitz, SatrajSahni and Rajasekaran, 2nd Edition, 2014, Universities Press.

Reference books:

1. Introduction to Algorithms, Thomas H. Cormen, Charles E. Leiserson, Ronal L. Rivest, Clifford Stein, 3rd Edition, PHI.
2. Design and Analysis of Algorithms, S. Sridhar, Oxford (Higher Education)

Web links and Video Lectures (e-Resources):

1. http://elearning.vtu.ac.in/econtent/courses/video/CSE/06CS43.html.
2. https://nptel.ac.in/courses/106/101/106101060/
3. http://elearning.vtu.ac.in/econtent/courses/video/FEP/ADA.html.
4. http://cse01-iiith.vlabs.ac.in/.
5. http://openclassroom.stanford.edu/MainFolder/CoursePage.php?course= Intro To Algorithms.

SOFTWARE ENGINEERING

Course Code	**22CDO753**	**CIE Marks**	**50**
Teaching Hours/Week (L:T:P: S)	**3:0:0:0**	**SEE Marks**	**50**
Total Hours of Pedagogy	**40**	**Total Marks**	**100**
Credits	**03**	**Exam Hours**	**03**

Course Objectives:

This course will enable students to,

- Recollect Software process models and compare their applicability
- Acquire Software Requirement Analysis and Specification
- Acquire Systematic software design procedure for Object Oriented and Real Time software
- Interpret how to develop and test a software application/product
- Cognize software cost estimation techniques and to know project management

Teaching-Learning Process (General Instructions)

These are sample Strategies; which teachers can use to accelerate the attainment of the various course outcomes.

1. Lecturer method (L) need not to be only a traditional lecture method, but alternative effective teaching methods could be adopted to attain the outcomes.
2. Use of Video/Animation to explain functioning of various concepts.
3. Encourage collaborative (Group Learning) Learning in the class.
4. Ask at least three HOT (Higher order Thinking) questions in the class, which promotes critical thinking.
5. Adopt Problem Based Learning (PBL), which fosters students 'Analytical skills, develop design thinking skills such as the ability to design, evaluate, generalize, and analyze information rather than simply recall it.
6. Introduce Topics in manifold representations.
7. Show the different ways to solve the same problem with different circuits/logic andencourage the students to come up with their own creative ways to solve them.
8. Discuss how every concept can be applied to the real world - and when that's possible, it helps improve the students' understanding.

Module – I

Introduction to Software Engineering & Software Process Model: what is Software Engineering (SE), Difference between SE and System Engineering. The evolving role of software – the changing nature of software- Life cycle models [Text Book-1]

Software Process Model: - Water fall, Incremental, Spiral, Evolutionary, Prototyping Concurrent development – Specialized Process Models: Component-Based Development, The Formal Methods Model, and Aspect-Oriented Software Development [Text Book-2] **08 Hours**

Module – II

Software Requirement Analysis and Specifications: Functional and Non-Functional, User, System – Requirement, Interface specification, software requirement documents.[Text Book-1]

Requirement Engineering Process: Feasibility Studies, Requirements Elicitation and analysis, Requirement Validation and Requirement management. [Text Book-1]

System Model: Context Model, Behavioral model, Data Model, Object Model, Structured Model [Text Book-1] **08 Hours**

Module – III

Software Design: Architectural design, Architectural Design Document, Client Server Architecture Distributed Object Architecture. [Text Book-1]

Object Oriented Design: Object Oriented Design Process, Design Evolution, [Text Book-1]

Real time Software Design: System Design, Real time Operating System, Monitoring and Control System and Data Acquisition System [Text Book-1] **08 Hours**

Module – IV

Software Development and Testing: Rapid Software Development-Agile Methods, Extreme Programming, Rapid Application Development, [Text Book-1]

Software Reuse: Reuse landscape, Design Pattern, Application system Reuse [Text Book-1]
Verification and Validation; Planning Verification & Validation, Software Inspection, Verification and formal Methods. [Text Book-1]
Software Testing: Approaches of Software Testing, Software Testing Strategies, Test Strategies for Object Oriented Software-Unit Testing, Integration Testing [Text Book-2] **08 Hours**

Module – V

Software Cost Estimation and Project Management: Software cost estimation - COCOMO model – Estimation Techniques, Project Duration and Staffing, [Text Book-1]
Quality management: Quality Assurance and Standard, Quality Planning and Quality Control [Text Book-1]
Configuration Management: Configuration Management Planning, Change Management, Version and Release Management [Text Book-1]
Emerging Technology: Security Concepts, Security Risk Management. [Text Book-1]. **08 Hours**

Teaching-Learning Process for allmodules	Chalk and talk method / PowerPoint Presentation

Course Outcomes:
On completion of this course, the students will be able to,
CO1: Identify and apply Software life cycle and process models to compare their applicability.
CO2: Analyze the types of requirements and summarize Requirement Engineering for various System models.
CO3: Design data, functional and behavioural model for any given software requirement.
CO4: Apply appropriate techniques and Test the software application/product for a given problem.
CO5: Comprehend concepts of software quality assurance and software configuration management.

Assessment Details (both IAT and SEE)

Theory Component	IAT-1 after completion 45 to 50% Syllabus	25 Marks
	IAT-2 after completion 95 to 100% Syllabus	25 Marks
	Average of two IATs	**25 Marks**
	CCE-1	25 Marks
	CCE-2	25 Marks
	Average of two CCEs	**25 Marks**
Grand Total of IAT Marks (min marks 20 / 50)		**50 Marks**
SEE conducted for 100 and scaled down to 50 **(min marks 18/50)**		**50 Marks**
IAT + SEE (min marks 40)		**100 Marks**

Text Books:

1. Ian Sommerville,"Software Engineering", 8th Edition, ISBN-10-9332582696, ISBN-13-978-9332582699,Pearson Education (24 May 2017).
2. Rogar Pressman, "Software Engineering and Application", 7th Edition, McGraw Hill Education Publication, 2009, ISBN-13:9789339212087.

Reference books:

1. Pankaj Jalote, "Software Engineering, A Precise Approach", Wiley India, 2010,ISBN: 9788126523115
2. Pfleeger and Lawrance, "Software Engineering: Theory and Practice" Pearson Education, 2 nd Edition, 2001.
3. Stephan Schach, "Software Engineering", Tata McGraw Hill, 2007.
4. Rajib Mall, "Fundamentals of Software Engineering", 3 rd Edition, PHI Learning Private Limited, 2009, .ISBN-10-9788120338197,ISBN-13-978-8120338197.
5. Kelkar S.A., "Software Engineering", ISBN 10: 8120332725, ISBN 13: 9788120332720, Publisher: Prentice-Hall of India Pvt. Ltd, 2007.

E-Resources:
1. https://www.pearson.com/us/higher-education/product/Sommerville-Software-Engineering- 9th-Edition/9780137035151.html.
2. https://www.abebooks.com/9788120332720/Software-Engineering-Kelkar-S-A- 8120332725/plp
3. https://www.wileyindia.com/pankaj-jalote-s-software-engineering-a-precise-approach.html

DATA MANAGEMENT

Course Code	**22CDO754**	**CIE Marks**	**50**
Teaching Hours/Week (L:T:P: S)	**3:0:0:0**	**SEE Marks**	**50**
Total Hours of Pedagogy	**40**	**Total Marks**	**100**
Credits	**03**	**Exam Hours**	**03**

Course Objectives:

This course will enable to,

1. Fundamentals of data collection **and** data quality management.
2. Managing big data and understand the architecture behind data storage solutions.
3. Principles of data privacy and compliance.
4. Basics of machine learning, gaining an understanding of how it can be used to automate data analysis and make predictions.

Teaching-Learning Process (General Instructions)

These are sample Strategies; which teachers can use to accelerate the attainment of the various course outcomes.

1. Lecturer method (L) need not to be only a traditional lecture method, but alternative effective teaching methods could be adopted to attain the outcomes.
2. Use of Video/Animation to explain functioning of various concepts.
3. Encourage collaborative (Group Learning) Learning in the class.
4. Ask at least three HOT (Higher order Thinking) questions in the class, which promotes critical thinking.
5. Adopt Problem Based Learning (PBL), which fosters students 'Analytical skills, develop design thinking skills such as the ability to design, evaluate, generalize, and analyze information rather than simply recall it.
6. Introduce Topics in manifold representations.
7. Show the different ways to solve the same problem with different circuits/logic andencourage the students to come up with their own creative ways to solve them.
8. Discuss how every concept can be applied to the real world - and when that's possible, it helps improve the students' understanding.

Module – I: overview of collecting data

What is Data? Importance of Data in Decision-Making-Types of Data-Data Sources-Surveys and Questionnaires-Web Scraping and APIs-Sensor Data Collection-Common Tools for Data Collection-Data Acquisition Best Practices-Data Quality Dimensions-Data Preprocessing and Cleaning-Data Profiling and Validation Tools

08 Hours

Module – II: Storing Data

Overview: Storing Data ,Importance of Data Storage, Data Storage Formats, Data Models, Relational Databases (SQL) Document (NoSQL) Databases, Data Warehousing, What Is Big Data?, Hadoop Ecosystem ,Distributed Data Storage ,Data Lakes vs. Data Warehouses, Data Storage in the Cloud-Data Storage Tools and Frameworks.

08 Hours

Module – III: Securing data

Overview: Securing Data, Data Security vs. Data Privacy, Security Threats and Vulnerabilities, Encryption and Access Control, Data Classifications, GDPR, CCPA, and Other Data Privacy Regulations, Data Anonymization and Pseudonymization, Compliance Best Practices, Data Governance Framework , Data Stewardship and Data Ownership, Data Cataloging and Metadata Management ,Secure Communication Protocols, Data Encryption Methods, Secure Storage Solutions

08 Hours

Module – IV: Using data

Overview: Data Analysis and Exploration , Data Analysis vs. Data Analytics, Exploratory Data Analysis (EDA),Statistical Analysis and Hypothesis Testing, Common Tools Used in Data Analysis and

Exploration, Importance of Data Visualization, Types of Data Visualization, Data Visualization Tools, Machine Learning Concepts, Supervised and Unsupervised Learning, Machine Learning Tools.

08 Hours

Module – V: Data Integration ,Warehousing And Governance

Data integration- Overview of ETL tools (e.g., Informatica, Talend), Data Warehousing- Design, implementation, and maintenance, Data Governance, Data Quality, Data Security.

08 Hours

Teaching-Learning Process for allmodules	Chalk and talk method / PowerPoint Presentation

Course Outcomes:

On completion of this course, the students will be able to,

CO1: Describe the fundamentals of data, types, sources, and data collection techniques to support informed decision-making.

CO2: Explain various data storage models, tools, and architectures, including big data, SQL/NoSQL databases, and cloud storage systems.

CO3: Explain various data storage models, tools, and architectures, including big data, SQL/NoSQL databases, and cloud storage systems

CO4: Perform data analysis using EDA, statistical techniques, and visualize data insights using suitable tools and basic machine learning techniques.

CO5: Demonstrate data integration using ETL tools and implement data governance principles to maintain data quality and consistency

Assessment Details (both IAT and SEE)

Theory Component	IAT-1 after completion 45 to 50% Syllabus	25 Marks
	IAT-2 after completion 95 to 100% Syllabus	25 Marks
	Average of two IATs	**25 Marks**
	CCE-1	25 Marks
	CCE-2	25 Marks
	Average of two CCEs	**25 Marks**
Grand Total of IAT Marks (min marks 20 / 50)		**50 Marks**
SEE conducted for 100 and scaled down to 50 **(min marks 18/50)**		**50 Marks**
IAT + SEE (min marks 40)		**100 Marks**

Text Books:

1. "Database Systems: The Complete Book" by Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Wisdom

Reference books:

1. "Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program" by John Ladley:

Web links and Video Lectures (e-Resources):

1. https://www.coursera.org/learn/introduction-to-data-management