Shirpur Education Society's

R. C. Patel Institute of Technology, Shirpur

(An Autonomous Institute)

Course Structure and Syllabus

Honors Program in Data Science

Computer Engineering

With effect from Year 2025-26

Shahada Road, Near Nimzari Naka, Shirpur, Maharashtra 425405
Ph: 02563 259 802, Web: www.rcpit.ac.in

R. C. PATEL
INSTITUTE OF TECHNOLOGY
An Autonomous Institute

Department of Computer Engineering

Vision of the Institute

To become a leading Institute in Technical education fostering innovation, research, ethical values, and sustainable development for the betterment of society.

Mission of the Institute

To impart high quality Technical Education through:

M1. Innovative and Interactive learning process and high quality, globally recognized instructional programs.

M2. Fostering a collaborative scientific temper among students with ethical responsibility towards the society.

M3. Preparing students from diverse backgrounds to have aptitude for employment, entrepreneurship and research with a spirit of Professionalism.

M4. To contribute to nation's sustainable development.

Vision of the Department

To provide high quality Computer Engineering education with socio-moral values.

Mission of the Department

M1 To provide state-of-the-art ICT based teaching-learning process.

M2 To groom the students to become professionally sound computer engineers to meet emerging needs of industry and society.

M3 To make the students employable professionals by inculcating ethical values.

Program Educational Objectives (PEOs) of the Department

PEO1 Graduates will have technical proficiency, lifelong learning skills to become a professional, entrepreneur and leader.

PEO2 Graduates will function effectively in diverse cultural and professional environments, respecting societal perspectives and inclusive practices.

PEO3 To foster ethical and social values to be socially responsible human being.

R. C. PATEL
INSTITUTE OF TECHNOLOGY
An Autonomous Institute

Department of Computer Engineering

Program Specific Outcomes (PSOs)

PSO 1 Apply programming principles, algorithms, and data structures to design efficient software solutions and intelligent systems using structured, object-oriented, and emerging technologies**.**

PSO 2 Design, develop, and deploy responsive web and mobile applications integrated with databases and cloud platforms, leveraging modern frameworks and tools for digital transformation.

Honors Program in Data Science offered by Computer Engineering (w.e.f. 2025-26)

Sr. No.	Course Category	Course Code	Course Title	Teaching Scheme			Evaluation Scheme					Total	Credit	
							Continuous Assessment (CA)							
							TA	Term Test 1 (TT1)	Term Test 2 (TT2)	Average of (TT1 & TT2)	ESE			
				L	T	P	[A]			[B]	[C]	[A+B+C]		
Sem-IV														
1	H1	RCP23CH1251	Statistics for Data Science	3			25	15	15	15	60	100	3	3
Sem-V														
2	H1	RCP23CH1301	Visualization in Data Science	3			25	15	15	15	60	100	3	4
	H1	RCP23CH1301L	Visualization in Data Science Laboratory			2	25				25	50	1	
Sem-VI														
3	H1	RCP23CH1351	Graph Data Science	3			25	15	15	15	60	100	3	4
	H1	RCP23CH1351L	Graph Data Science Laboratory			2	25				25	50	1	
4	H1	RCP23CH1352	Essentials of Generative AI and Prompt Engineering	3			25	15	15	15	60	100	3	3
Sem-VII														
5	H1	RCP23CH1401	Cloud Computing for Data Analysis	3			25	15	15	15	60	100	3	4
	H1	RCP23CH1401L	Cloud Computing for Data Analysis Laboratory			2	25				25	50	1	
			Total	15		6	200			75	375	650	18	

Prepared by:
Ms. J. S. Sonawane

Checked by:
Dr. S. M. Pardeshi

Prof. Dr. R. B. Wagh
BOS Chairman

Prof. S. P. Shukla
C.O.E.

Prof. Dr. P. J. Deore
Dean Academics/Dy. Director

Prof. Dr. J. B. Patil
Director

R. C. Patel Institute of Technology, Shirpur
SES
SHIRPUR EDUCATION SOCIETY
Autonomous

Program: Computer Engineering	S.Y. B.Tech.	Semester: IV
Statistics for Data Science (RCP23CH1251)		

Prerequisite: Knowledge of Basics of probability.

Course Objective(s):

1. To provide students with a solid understanding of statistical concepts and techniques essential for data science.
2. To develop students' ability to apply statistical methods to real-world data sets and draw meaningful insights.
3. To equip students with the skills to use statistical software tools for data analysis and interpretation.
4. To foster critical thinking and problem-solving skills in the context of statistical analysis for data science.

Course Outcomes:

On completion of the course, the learner will be able to:

COs	Course Outcomes	Blooms Level	Blooms Description
CO1	Summarize data using appropriate statistical measures	L2	Understand
CO2	Apply correlation analysis and regression to real-world problems.	L3	Apply
CO3	Apply regression techniques to model and forecast time series data.	L3	Apply
CO4	Apply nonparametric test and draw appropriate conclusions for data analysis.	L3	Apply

Statistics for Data Science (RCP23CH1251)
Course Contents

Unit-I Introduction to Data and Statistics 08 Hrs.

Elements, Variables, and Observations, Scales of Measurement, Qualitative and Quantitative Data, Cross-Sectional and Time Series Data, Descriptive Statistics, Statistical Inference, Summarizing Qualitative and Quantitative Data using Tables and Graphs, Cross Tabulations and Scatter Diagram, Measures of Central Tendency, Measures of Dispersion, Skewness, Moments and Kurtosis

Unit-II Correlation Analysis 06 Hrs.

Types of correlation, Karl Pearson's Coefficient of Correlation, Coefficient of Determination, Rank Correlation Coefficient, Coefficient of Concurrent Deviation

Unit-III Regression 08 Hrs.

Simple Linear Regression: Concept of linear relationship between variables Least squares method, Estimating the coefficients (slope and intercept), Assessing the goodness of fit (R-squared), Interpretation of coefficients

Multiple Linear Regression: Extending simple linear regression to multiple predictors, Estimating the coefficients, Interpreting the coefficients, Assessing the model fit (adjusted R-squared), Multicollinearity and its implications, Heteroscedaticity.

Logistic Regression: Modeling binary outcomes, Odds ratios and log odds, Interpreting the coefficients, Assessing model fit (deviance, likelihood ratio test).

Unit-IV Hypothesis Testing in Regression 04 Hrs.

Significance testing for regression coefficients, t-tests and p-values, Confidence intervals for coefficients, F-test for overall model significance.

Unit-V Regression for Time Series Data 05 Hrs.

Autocorrelation and its impact on regression, Autoregressive models (AR), Moving average models (MA), Autoregressive integrated moving average models (ARIMA).

Unit-VI Non-Parametric Statistics 08 Hrs.

Nonparametric Methods, Sign Test, Wilcoxon Signed-Rank Test, MannWhitney-Wilcoxon Test, Kruskal Wallis Test, Spearman's Rank Correlation, Runs Test.

Text Books:

1. Dr. S. P. Gupta, "Statistical Methods", 46^{th} Edition, S. Chand and Sons, 2021.
2. James T. McClave, P. George Benson, Terry T Sincich, "Statistics for Business and Economics", 14^{th} Edition, Pearson, 2021.

Reference Books:

1. Maurits Kaptein, Edwin van den Heuvel, "Statistics for Data Scientists: An Introduction to Probability, Statistics, and Data Analysis", Springer, 2022.
2. Peter Bruce, Andrew Bruce, "Practical Statistics for Data Scientists", O'Reilly, 2017.
3. C.B. Gupta & Vijay Gupta, "An Introduction to Statistical Methods", 23^{rd} Edition, S. Chand and Sons, 2004.

Program: Computer Engineering	T.Y. B.Tech.	Semester: V
Visualization in Data Science (RCP23CH1301)		
Visualization in Data Science Laboratory (RCP23CH1301L)		

Prerequisite: Knowledge of Basics of probability.

Course Objective(s):

1. Gaining a basic understanding of the subject (e.g., factual knowledge, methods, principles, generalizations, theories)
2. Learning to analyze and critically evaluate ideas, arguments, and points of view.
3. Learning appropriate methods for collecting, analyzing, and interpreting numerical information
4. Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course.

Course Outcomes:

On completion of the course, the learner will be able to:

COs	Course Outcomes	Blooms Level	Blooms Description
CO1	Explain the fundamental design principles and different types of data visualization.	L2	Understand
CO2	Analyze the positive and negative impacts of data-informed decision-making across various domains.	L4	Analyze
CO3	Apply the fundamental concepts of data visualization to define a project in your field of study.	L3	Apply
CO4	Apply core data visualization principles using widely available tools (e.g., Tableau)	L3	Apply
CO5	Demonstrate the best practice that presents your story in the process of creating data visualization including connecting to different data sources, assessing to the quality of the data, and converting raw data into data visualizations that provide actionable information.	L3	Apply

Visualization in Data Science (RCP23CH1301)
Course Contents

Unit-I Introduction to Data Visualization — 08 Hrs.

- Importance of Data Visualization
- History of Data Visualization
- Basic Principles and Concepts
- Tools and Technologies Overview
- Exploring Data with R and Python
- Introduction to R for Data Visualization
- Introduction to Python (matplotlib, seaborn)
- Data Cleaning and Preparation

Unit-II Advanced Visualization Technique — 06 Hrs.

- Advanced Charts: Histograms, Heatmaps, Box Plots
- Multivariate Data Visualization
- Geospatial Data Visualization

Unit-III Interactive Dashboards and Storytelling — 08 Hrs.

- Principles of Dashboard Design
- Tools for Creating Dashboards (Tableau, Power BI)
- Data Storytelling Techniques
- Data Sources and Data Wrangling
- Finding and Collecting Data
- APIs and Web Scraping
- Data Wrangling Techniques

Unit-IV Visual Perception and Design Principles — 04 Hrs.

- Understanding Human Perception in Visualization
- Principles of Design in Data Visualization
- Color Theory and Its Application in Visualizations

Unit-V Network and Graph Visualization — 05 Hrs.

- Principles of Network Visualization
- Tools and Techniques for Graph Visualization
- Time Series Data Visualization
- Techniques for Visualizing Time Series Data

- Tools and Libraries (e.g., matplotlib, Plotly)

Unit-VI Data Ethics and Privacy 08 Hrs.

- Ethical Considerations in Data Visualization
- Privacy and Data Protection
- Misleading Visualizations and Avoiding Bias
- Evaluating and Improving Visualizations
- Critiquing and Improving Visualizations
- Best Practices and Common Mistakes
- User Experience in Visualization
- Special Topics in Data Visualization
- Visualization for Different Domains (Healthcare, Finance, etc.)
- Emerging Trends and Technologies

Visualization in Data Science Laboratory (RCP23CH1301L)

List of Laboratory Experiments

Suggested Experiments:(Any 10)

1. Basic visualizations using Excel and Google Charts.
2. Cleaning and visualizing a dataset.
3. Advanced visualizations with Tableau and Python.
4. Creating an interactive dashboard.
5. Data collection and wrangling in Python.
6. Designing effective visualizations using color and design principles.
7. Case studies on ethical issues in data visualization.
8. Evaluating visualizations for ethical concerns.
9. Network visualization using Gephi.
10. Time series visualization in Python.
11. Peer review of visualizations And Improving a given set of visualizations
12. Mini project :Domain-specific visualization project(Healthcare, Finance, etc.)

Any other experiment based on syllabus may be included which would help the learner to understand topic/concept.

Oral examination will be based on the entire syllabus including, the practicals performed during laboratory sessions.

Text Books:

1. "The Visual Display of Quantitative Information" by Edward Tufte.
2. "Storytelling with Data" by Cole Nussbaumer Knaflic.

Reference Books:

1. Edward Tufte, "The Visual Display of Quantitative Information", 2^{nd} Edition, Read: pp. 53–77 & 107–121.
2. William S. Cleveland, "The Elements of Graphing Data", 1985. Read: pp. 24–55 & 68–88.
3. Alberto Cairo, "The Truthful Art: Data, Charts, and Maps for Communication". Read: pp. 41–65 & 121–149.

Online Resources:

1. Designing Great Visualizations: https://www.tableau.com/sites/default/files/media/designinggreat-visualizations.pdf One Dataset, Visualized 25 Ways: https://flowingdata.com/2017/01/24/one-datasetvisualized-25- ways/ Tableau tutorial videos Getting started (25 mins): https://www.tableau.com/learn/tutorials/ondemand/gettingstart 5bc7a876b95af77 22b08fff9224 & Practice]
2. Tableau how-to videos: https://public.tableau.com/enus/s/resources (31 mins)

Program: Computer Engineering	T.Y. B.Tech.	Semester: VI
Graph Data Science (RCP23CH1351)		
Graph Data Science Laboratory (RCP23CH1351L)		

Prerequisite: Database Management Systems.

Course Objective(s):

1. Analyze real-world problems and design graph-based solutions.
2. Evaluate graph databases and construct efficient data models.
3. Develop an applications using Neo4j and graph database concepts

Course Outcomes:

On completion of the course, the learner will be able to:

COs	Course Outcomes	Blooms Level	Blooms Description
CO1	Solve fundamental graph theory problems including shortest paths and cycles.	L3	Apply
CO2	Compare graph databases with relational/NoSQL alternatives.	L5	Evaluate
CO3	Create efficient graph data models while avoiding common pitfalls.	L6	Create
CO4	Apply graph database concepts using Neo4j and Cypher.	L3	Apply

Graph Data Science (RCP23CH1351)
Course Contents

Unit-I Introduction to Graph 05 Hrs.

Definitions and examples, Three puzzles, Paths and cycles, Connectivity, Eulerian graphs, Hamiltonian graphs, shortest path, Chinese postman problem, traveling salesman problem, trees, properties of trees.

Unit-II Introduction Graph databases 07 Hrs.

A High-Level View of the Graph Space, Graph Databases, Graph Compute Engines, The Power of Graph Databases, Performance, Flexibility, Agility, Options for Storing Connected Data, Relational Databases Lack Relationships, NOSQL Databases Also Lack Relationships, Graph databases embraces relationship

Unit-III Data Modelling with Graphs 08 Hrs.

Models and Goals, The Labelled Property Graph Mode Querying Graphs, A Comparison of Relational and Graph Modelling, Cross-Domain Models, Common Modelling Pitfalls, Identifying Nodes and Relationships, Avoiding Anti-Patterns

Unit-IV Building a Graph Database Application 07 Hrs.

Data Modelling, Application Architecture, Testing, Capacity Planning, Importing and Bulk Loading Data.

Unit-V Graphs in the Real World 07 Hrs.

Why Organizations Choose Graph Databases, Common Use Cases, Real-World Examples, Authorization and Access Control, Geospatial and Logistics, Graph Database Internals, Native Graph Processing, Native Graph Storage Programmatic APIs, Kernel API, Core API, Traversa Framework, Non-functional Characteristics.

Unit-VI Neo4j 05 Hrs.

Neo4j – About, Neo4j – Installation, Neo4j – Browser Neo4j - Query Language (Cypher), Neo4j - Create a Node Neo4j - Create a Relationship, Neo4j - Create an Index Neo4j - Create a Constraint, Neo4j - Select Data with MATCH, Neo4j - Import Data from CSV, Neo4j - Drop an Index, Neo4j - Drop a Constraint, Neo4j - Delete a Node, Neo4j - Delete a Relationship.

Graph Data Science Laboratory (RCP23CH1351L)

List of Laboratory Experiments

Suggested Experiments:(Any 10)

1. Install and set up a graph database system (e.g., Neo4j) on a local machine and familiarize yourself with the graph database environment, including the query language (Cypher) and browser interface.
2. Design a data model using the labeled property graph model for a specific domain (e.g., social network, e-commerce).
3. Implement the data model in the graph database and populate it with sample data.
4. Perform basic graph queries using Cypher to retrieve nodes, relationships, and their properties.
5. Explore different query patterns, such as finding paths, filtering nodes, and ordering results.
6. Learn and practice essential administrative tasks, such as managing users, roles, and access control.
7. Perform backup and restore operations to ensure data integrity.
8. Import data from external sources (e.g., CSV files) into the graph database.
9. Export graph data to different formats for analysis or sharing.
10. Apply graph algorithms to analyze and extract insights from your graph data.
11. Explore and navigate the graph visually to gain a better understanding of its structure and relationships.
12. Choose a specific real-world use case (e.g., recommendation systems, fraud detection) and apply graph database techniques to solve the problem.

Any other experiment based on syllabus may be included which would help the learner to understand topic/concept.

Oral examination will be based on the entire syllabus including, the practicals performed during laboratory sessions.

Text Books:

1. Dave Bechberger, Josh Perryman, "Graph Databases in Action", 1st Edition, Manning Publications, 2020.
2. Ian Robinson, Jim Webber & Emil Eifrem, "Graph Databases", 2nd Edition, O'Reilly, 201

3. Robin J. Wilson, "Introduction to Graph Theory", 4^{th} Edition.

4. Dr. Jim Webber and Rik Van Bruggen, "Graph Databases for Dummies, Neo4j Special Edition", John Wiley and Sons, 2020.

Reference Books:

1. Ian Robinson, Jim Webber & Emil Eifrem, "Graph Databases: New Opportunities for Connected Data", 2^{nd} Edition, O'Reilly, 2015.

2. Tjortjis Christos, "Graph Databases: Applications on Social Media Analytics and Smart Cities", 1^{st} Edition, CRC Press, 2023.

3. Aleksa Vukotic, Nicki Watt, Tareq Abedrabbo, Dominic Fox, and Jonas Partner, "Neo4j in Action", Manning Publication, 2014.

Program: Computer Engineering	T.Y. B.Tech.	Semester: VI
Essentials of Generative AI and Prompt Engineering (RCP23CH1352)		

Prerequisites: Artificial Intelligence, Data Mining, Introduction to Machine Learning.

Course Objective(s):

1. To equip students with the skills to understand and apply generative AI models and prompt engineering techniquesto create and optimize AI-generated content across diverse applications.

Course Outcomes:

On completion of the course, the learner will be able to:

COs	Course Outcomes	Blooms Level	Blooms Description
CO1	Apply generative AI techniques for real-world problem-solving.	L3	Apply
CO2	Design and refine effective prompts to generate high-quality outputs from AI models.	L6	Create

Essentials of Generative AI and Prompt Engineering (RCP23CH1352) Course Contents

Unit-I Introduction to Generative Models 07 Hrs.

Definition and purpose of generative models, Difference between generative and predictive (discriminative) models, Types of Data Generated by AI, Text, images, and sound generation, Examples of popular generative tools and applications, How Generative Models Work, Ethical Considerations in Generative AI, Responsible use of generative AI: copyright, misinformation, and bias

Unit-II Prompt Engineering 08 Hrs.

Definition of prompts and prompt engineering, Basic Principles of Writing Prompts, Characteristics of a good prompt: clarity, detail, and specificity, Types of Prompts and Their Effects, Instructional prompts, questions, and open-ended prompts, Applications of Prompt Engineering, Real-life applications: content creation, summaries, creative writing

Unit-III Generative Models 08 Hrs.

Generative Adversarial Networks (GANs), Language models (e.g., GPT, BERT) and diffusion models, Overview of encoder-decoder frameworks and transformers

Unit-IV Generative AI in Text and Image Creation 08 Hrs.

Text Generation with AI: How AI generates text responses to prompts,

Examples of applications: story generation, chatbot conversations, AI tools for text generation

Image Generation with AI: Basics of text-to-image AI models,

Examples of image generation applications: art, design, marketing, AI tools for image generation

Evaluating AI Outputs: Evaluating the quality and relevance of AIgenerated text or images, Adjusting prompts for better results

Unit-V Prompt Engineering Techniques and Approaches 08 Hrs.

Text-to-Text Prompt Techniques, Interview Pattern Approach, Chain-ofThought Approach, Tree-of-Thought Approach, Few-Shot Approach. Writing Effective Prompts for Various Purposes, Crafting prompts for specific tasks: summarization, image creation, Q&A, Emerging applications in entertainment, education, and business.

Text Books:

1. David Foster, "Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play", O'Reilly Media, 2019.
2. Ian Goodfellow, Yoshua Bengio, and Aaron Courville, "Deep Learning", MIT Press, 2016.
3. Rajalingappaa Shanmugamani, "Deep Learning for Computer Vision", Packt Publishing, 2018.
4. James Phoenix and Mike Taylor, "Prompt Engineering for Generative AI", O'Reilly Media, Inc., 2024.
5. Rafael Valle, "Hands-On Generative Adversarial Networks with Keras: Build Advanced GAN Architectures Using Keras and TensorFlow", Packt Publishing, 2019.

Reference Books:

1. Stuart J. Russell and Peter Norvig, "Artificial Intelligence: A Modern Approach", 3^{rd} Edition, Pearson Education, 2010.
2. Daniel Jurafsky and James H. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech", Pearson Publication, 2014.
3. Alexander Clark, Chris Fox, and Shalom Lappin (Editors), "The Handbook of Computational Linguistics and Natural Language Processing", July 2010.
4. Navveen Balani, "Prompt Engineering: Unlocking Generative AI – Ethical Creative AI for All", Amazon Digital Services LLC – KDP, 2023.

Program: Computer Engineering	Final Year B.Tech.	Semester: VII
Cloud Computing for Data Analysis (RCP23CH1401)		
Cloud Computing for Data Analysis Laboratory (RCP23CH1401L)		

Prerequisites: Data Science fundamentals, Programming in Python/R, Basic understanding of databases.

Course Objective(s):

1. To provide students with a comprehensive understanding of cloud computing platforms and services for data analysis.
2. To develop skills in designing and implementing cloud-based data processing pipelines.
3. To train students in deploying and scaling data analysis solutions in cloud environments.
4. To equip students with knowledge to evaluate and select appropriate cloud services for specific data analysis requirements.
5. To familiarize students with best practices for security, cost management, and performance optimization in cloud data solutions.

Course Outcomes:

On completion of the course, the learner will be able to:

COs	Course Outcomes	Blooms Level	Blooms Description
CO1	Design appropriate cloud-based solutions for various data analysis scenarios.	L6	Create
CO2	Design scalable data processing pipelines in cloud environments.	L6	Create
CO3	Apply cloud-native technologies to optimize data analysis workflows.	L3	Apply
CO4	Apply security best practices for cloud-based data solutions.	L3	Apply
CO5	Evaluate cloud service providers and their offerings for specific data analysis needs.	L5	Evaluate
CO6	Evaluate cloud resource utilization for cost-effectiveness and performance	L5	Evaluate

Cloud Computing for Data Analysis (RCP23CH1401) Course Contents

Unit-I Introduction to Cloud Computing for Data Analysis 06 Hrs.

Evolution of Cloud Computing for Data Science, Key cloud service models (IaaS, PaaS, SaaS) for data analysis, Major cloud providers and their data analytics offerings, Advantages and challenges of cloud-based data analysis,Hybrid and multi-cloud strategies for data workloads, Cost models and TCO analysis for data analytics in the cloud.

Unit-II Cloud Storage Solutions for Data Science 07 Hrs.

Types of cloud storage (object, file, block) for data analytics, Data lakes and data warehouses in the cloud, Cloud-native databases (SQL, NoSQL, NewSQL), Data migration strategies and tools, Data governance and catalog services, Performance optimization for data access, Data persistence patterns.

Unit-III Big Data Processing in the Cloud 07 Hrs.

Distributed computing frameworks (Hadoop, Spark) in cloud environments, Serverless data processing (AWS Lambda, Azure Functions, Google Cloud Functions), Stream processing services (Kafka, Kinesis, Pub/Sub), ETL/ELT pipelines in the cloud, Batch vs. real-time processing trade-offs, Integration patterns for cloud data services, Performance monitoring and optimization.

Unit-IV Cloud-Based Machine Learning and AI Services 07 Hrs.

Machine learning platforms in the cloud (SageMaker, Azure ML, Vertex AI), MLOps in cloud environments, Pre-built AI services and APIs, GPU and TPU computing resources, Model training, deployment, and serving, AutoML capabilities, Edge computing for ML inference.

Unit-V Data Visualization and Business Intelligence in the Cloud 06 Hrs.

Cloud-based BI platforms and services, Interactive dashboards and reporting tools, Real-time visualization techniques, Embedding analytics in applications, Multi-device visualization strategies, Collaborative analytics environments, Accessibility and performance considerations .

Unit-VI Security, Compliance, and Best Practices 06 Hrs.

Data security in cloud environments, Privacy regulations and compliance (GDPR, HIPAA, etc.), Identity and access management for data resources, Encryption strategies for data at rest and in transit, Monitoring and logging for data access, Disaster recovery and business continuity, Cost optimiza[illegible] strategies, Sustainability considerations in cloud computing.

Cloud Computing for Data Analysis Laboratory (RCP23CH1401L)

List of Laboratory Experiments

Suggested Experiments:(Any 08)

1. Set up a cloud environment for data analysis using a major cloud provider (AWS/Azure/GCP)
2. Implement data storage solutions using cloud-native storage services and databases
3. Design and deploy a data pipeline for batch processing of large datasets
4. Implement real-time data streaming and processing using cloud services
5. Build a serverless data processing workflow for specific analysis tasks
6. Train and deploy a machine learning model using cloud ML services
7. Create interactive dashboards and visualizations using cloud BI tools
8. Implement data security measures including encryption and access controls
9. Build a multi-region data analysis solution with redundancy
10. Design and implement a cost-optimized data analytics architecture
11. Implement monitoring and logging for a cloud-based data solution
12. Mini-project: End-to-end cloud data analysis solution for a specific domain (healthcare, finance, etc.)

Any other experiment based on syllabus may be included which would help the learner to understand topic/concept.

Oral examination will be based on the entire syllabus including, the practicals performed during laboratory sessions.

Text Books:

1. Marinescu, Dan C., "Cloud Computing: Theory and Practice," 3rd Edition, Morgan Kaufmann, 2023.
2. Erl, Thomas, et al., "Cloud Computing Design Patterns,"2nd Edition, Prentice Hall, 2023.
3. Kleppmann, Martin, "Designing Data-Intensive Applications in the Cloud," O'Reilly Media, 2022.

4. Warden, Pete and Bradski, Gary, "Machine Learning in the Cloud: Build and Deploy ML Solutions with AWS, GCP, and Azure," O'Reilly Media, 2023.

Reference Books:

1. Sarkar, Arshdeep and Bandyopadhyay, Arunava, "Hands-On Data Science on AWS," 2^{nd} Edition, Packt Publishing, 2023.

2. Kroonenburg, Noah, "Machine Learning Design Patterns for Cloud Platforms," O'Reilly Media, 2024.

3. Kochhar, Prabhdeep Singh, "Data Science in the Cloud with Microsoft Azure Machine Learning and Python," Microsoft Press, 2023.

4. Fregly, Chris and Barth, Antje, "Data Science on AWS: Implementing End-to-End, Continuous AI and Machine Learning Pipelines," 2^{nd} Edition, O'Reilly Media, 2023.

5. Valentin, Jonas and Patel, Janani, "Practical MLOps: Cloud Deployment and Automation for Machine Learning," O'Reilly Media, 2023.