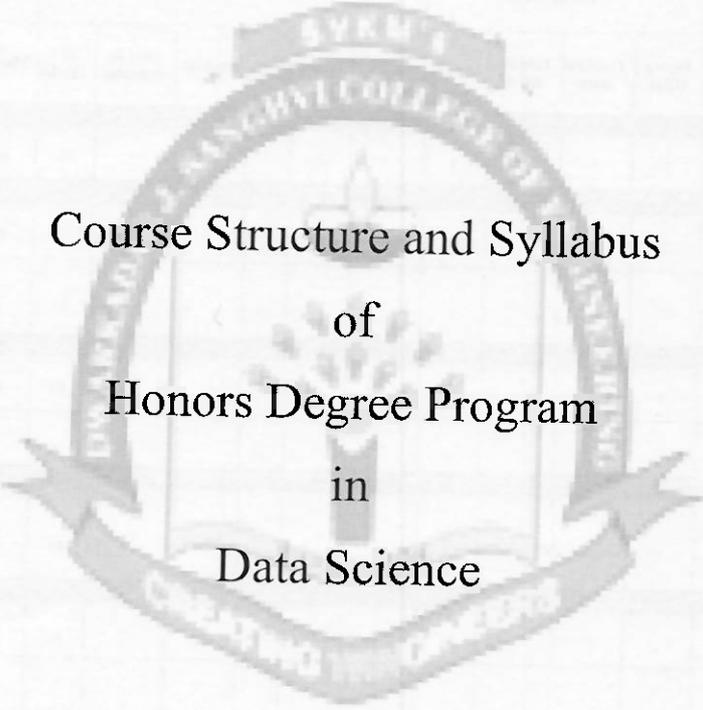




Shri Vile Parle Kelavani Mandal's

Dwarkadas J. Sanghvi College of Engineering

(Autonomous College Affiliated to the University of Mumbai)



Course Structure and Syllabus
of
Honors Degree Program
in
Data Science

Prepared by:- Board of Studies in Computer Engineering

Recommended by:- Academic Council of D. J. Sanghvi College of Engineering

Approved by:- Governing Body of D. J. Sanghvi College of Engineering

Revision: 1 (2024)

With effect from the Academic Year: 2024-2025

Scheme for Honors in Data Science (Computer Engineering) (Autonomous) (DJS23 Scheme)

Sr. No.	Course Code	Course	Teaching Scheme				Semester End Examination (SEE)						Continuous Assessment (CA)			Aggregate (A+B)	Credits Earned	
			Theory (Hrs)	Practical (Hrs)	Tutorial (Hrs)	Credits	Duration (Hrs)	Theory	Oral	Practical	Oral & Practical	SEE (Total)	Theory	Term Work Total	CA Total			
Semester III																		
1	DJS23CH1201	Statistics for Data Science	3	--	--	3	2	60	--	--	--	60	40	--	40	100	3	3
Semester IV																		
2	DJS23CH1251	Visualization in Data Science	3	--	--	3	2	60	--	--	--	60	40	--	40	100	3	4
	DJS23CH1251L	Visualization in Data Science Laboratory	--	2	--	1	--	--	25	--	--	25	--	25	25	50	1	
Semester V																		
3	DJS23CH1301	Graph Data Science	3	--	--	3	2	60	--	--	--	60	40	--	40	100	3	4
	DJS23CH1301L	Graph Data Science Laboratory	--	2	--	1	--	--	25	--	--	25	--	25	25	50	1	
Semester VI																		
4	DJS23CH1351	Essentials of Generative AI and Prompt Engineering	3	--	--	3	2	60	--	--	--	60	40	--	40	100	3	3
Semester VII																		
5	DJS23CH1401	Cloud Computing for Data Analysis	3	--	--	3	2	60	--	--	--	60	40	--	40	100	3	4
6	DJS23CH1401L	Cloud Computing for Data Analysis Laboratory	--	2	--	1	--	--	25	--	--	25	--	25	25	50	1	
Total			15	6	--	18	10	300	75	--	--	375	200	75	275	650	18	18

[Handwritten signature]

[Handwritten signature]



Continuous Assessment (A):

Course	Assessment Tools	Marks	Time (mins)
Theory	a. Term Test 1 (based on 40 % syllabus)	15	45
	b. Term Test 2 (on next 40 % syllabus)	15	45
	c. Assignment / course project / group discussion / presentation / quiz/ any other.	10	--
	Total marks (a + b + c)	40	--
Audit course	Performance in the assignments / quiz / power point presentation / poster presentation / group project / any other tool.	--	As applicable
Laboratory	Performance in the laboratory and documentation.	25	
Tutorial	Performance in each tutorial & / assignment.	25	
Laboratory & Tutorial	Performance in the laboratory and tutorial.	50	

The final certification and acceptance of term work will be subject to satisfactory performance upon fulfilling minimum passing criteria in the term work / completion of audit course.

Semester End Assessment (B):

Course	Assessment Tools	Marks	Time (hrs.)
Theory / * Computer based	Written paper based on the entire syllabus.	60	2
	* Computer based assessment in the college premises.		
Oral	Questions based on the entire syllabus.	25	As applicable
Practical	Performance of the practical assigned during the examination and the output / results obtained.	25	2
Oral & Practical	Project based courses - Performance of the practical assigned during the examination and the output / results obtained. Based on the practical performed during the examination and on the entire syllabus.	As per the scheme	2

Prepared by

Checked by

Head of the Department

Principal



Program: Computer Engineering
Honors in Data Science
Course Code: DJS23CH1201

S.Y B. Tech. Semester: III

Course: Statistics for Data Science

Pre-requisite: Knowledge of

1. Basics of probability

Objectives:

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.

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

1. Describe and summarize data using appropriate statistical measures.
2. Apply correlation analysis and regression to real-world problems.
3. Apply regression techniques to model and forecast time series data.
4. Apply nonparametric test and draw appropriate conclusions for data analysis.

Statistics for Data Science (DJS23CH1201)		
Unit	Description	Duration
1	Introduction to Data and Statistics: 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	08
2	Correlation Analysis: Types of correlation, Karl Pearson's Coefficient of Correlation, Coefficient of Determination, Rank Correlation Coefficient, Coefficient of Concurrent Deviation	06
3	Regression: 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	10

(Handwritten signatures)



	Logistic Regression: Modeling binary outcomes, Odds ratios and log odds, Interpreting the coefficients, Assessing model fit (deviance, likelihood ratio test)	
4	Hypothesis Testing in Regression: Significance testing for regression coefficients, t-tests and p-values, Confidence intervals for coefficients, F-test for overall model significance	05
5	Regression for Time Series Data: Autocorrelation and its impact on regression, Autoregressive models (AR), Moving average models (MA), Autoregressive integrated moving average models (ARIMA)	05
6	Non-Parametric Statistics: Nonparametric Methods, Sign Test, Wilcoxon Signed-Rank Test, MannWhitney-Wilcoxon Test, Kruskal Wallis Test, Spearman's Rank Correlation, Runs Test	08
	Total	42

Books Recommended:

Textbooks:

1. Dr. S. P. Gupta, "Statistical Methods", 46th Edition, S. Chand and Sons, 2021.
2. James T. McClave, P George Benson, Terry T Sincich, "Statistics for Business and Economics", 14th 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", 23rd Edition, S. Chand and Sons, 2004.

Prepared by

Checked by

Head of the Department

Principal



Program: Computer Engineering

S.Y B. Tech. Semester: IV

Honors in Data Science

Course: Visualization in Data Science (DJS23CH1251)

Course: Visualization in Data Science Laboratory (DJS23CH1251L)

Pre-requisite: Knowledge of

1. Basics of probability

Objectives:

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

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

1. Understand the fundamental design principles and different types of data visualization.
2. Identify both positive and negative impacts of data-informed decision across a variety of domains.
3. Apply the fundamental concepts of data visualization to define a project in your field of study.
4. Practice the core principles using widely available tools (e.g. Tableau).
5. 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.

Visualization in Data Science (DJS23CH1251)		
Unit	Description	Duration
1	Introduction to Data Visualization 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	08
2	Advanced Visualization Techniques - Advanced Charts: Histograms, Heatmaps, Box Plots - Multivariate Data Visualization - Geospatial Data Visualization	06



3	Interactive Dashboards and Storytelling - 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	10
4	Visual Perception and Design Principles - Understanding Human Perception in Visualization - Principles of Design in Data Visualization - Color Theory and Its Application in Visualizations	04
5	Network and Graph Visualization - 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)	06
6	Data Ethics and Privacy - 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	08
	Total	42

Books Recommended:

Textbooks:

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

Reference Books:

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



Online material :

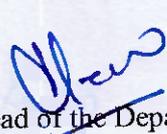
- Designing Great Visualizations:
<https://www.tableau.com/sites/default/files/media/designinggreat-visualizations.pdf>
- One Dataset, Visualized 25 Ways: <https://flowingdata.com/2017/01/24/one-datasetvisualized-25-ways/>
- Tableau tutorial videos Getting started (25 mins):
<https://www.tableau.com/learn/tutorials/ondemand/gettingstarted?playlist=554268&signin=4bec05bc7a876b95af7722b08fff9224>
& Practice]
- Tableau how-to videos: <https://public.tableau.com/enus/s/resources> (31 mins)

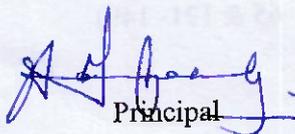
Experiment list:

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.)	


Prepared by


Checked by


Head of the Department


Principal



Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
 (Autonomous College Affiliated to the University of Mumbai)
 NAAC Accredited with "A" Grade (CGPA : 3.18)



Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
 (Autonomous College Affiliated to the University of Mumbai)
 NAAC Accredited with "A" Grade (CGPA : 3.18)



Program: Computer Engineering

T.Y B.Tech.

Semester: V

Course: Graph Data Science (DJS23CH1301)

Course: Graph Data Science Laboratory (DJS23CH1301L)

Pre-requisite:

1. Database Management Systems

Objectives:

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

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

1. Solve fundamental graph theory problems including shortest paths and cycles.
2. Compare graph databases with relational/NoSQL alternatives.
3. Create efficient graph data models while avoiding common pitfalls.
4. Implement graph database applications using Neo4j and Cypher.

Graph Data Science (DJS23CH1301)		
Unit	Description	Duration
1	Introduction to Graph 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	05
2	Introduction Graph databases 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	08
3	Data Modelling with Graphs 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	08
4	Building a Graph Database Application Data Modelling, Application Architecture, Testing, Capacity Planning, Importing and Bulk Loading Data	08
5	Graphs in the Real World 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	08



6	Neo4j 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	05
	Total	42

Graph Data Science Laboratory (DJS23CH1301L)

Exp.	Suggested experiments
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.

Batchwise laboratory work of minimum 8 experiments from the above suggested list or any other experiment based on syllabus will be included, which would help the learner to apply the concept learnt.

Books Recommended:

Text Books:

1. Dave Bechberger, Josh Perryman, "Graph Databases in Action", 1st Edition, Manning Publications, 2020.



Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



2. Ian Robinson, Jim Webber & Emil Eifrem, "Graph Databases", 2nd Edition, O'Reilly, 2015
3. Robin J. Wilson, "Introduction to Graph Theory", 4th 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", 2nd Edition, O'Reilly, 2015.
2. Tjortjis Christos, "Graph Databases: Applications on Social Media Analytics and Smart Cities", 1st Edition, CRC Press, 2023.
3. Aleksa Vukotic, Nicki Watt, Tareq Abedrabbo, Dominic Fox, and Jonas Partner, "Neo4j in Action", Manning Publication, 2014.

Prepared by

Checked by

Head of the Department

Principal



Program: Computer Engineering

T.Y B.Tech. Semester: VI

Course: Essentials of Generative AI and Prompt Engineering (DJS23CH1351)

Course:

Pre-requisite:

1. Artificial Intelligence, Data Mining, Introduction to Machine Learning

Objectives:

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

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

1. Understand and apply generative AI techniques for real-world problem-solving.
2. Design and refine effective prompts to generate high-quality outputs from AI models.

Essentials of Generative AI and Prompt Engineering (DJS23CH1351)		
Unit	Description	Duration
1	Introduction to Generative Models 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	08
2	Prompt Engineering 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	08
3	Generative Models Generative Adversarial Networks (GANs), Language models (e.g., GPT, BERT) and diffusion models, Overview of encoder-decoder frameworks and transformers	08
4	Generative AI in Text and Image Creation 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 AI-generated text or images, Adjusting prompts for better results	10
5	Prompt Engineering Techniques and Approaches Text-to-Text Prompt Techniques, Interview Pattern Approach, Chain-of-Thought Approach, Tree-of-Thought Approach, Few-Shot Approach	08



Shri. Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



	Writing Effective Prompts for Various Purposes, Crafting prompts for specific tasks: summarization, image creation, Q&A, Emerging applications in entertainment, education, and business	
	Total	42

Books Recommended:

Text Books:

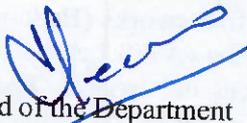
1. David Foster, Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play, O'Reilly Media, 2019
2. Goodfellow, I., Bengio, Y., and Courville, A., Deep Learning, MIT Press, 2016.
3. Rajalingappaa Shanmugamani, "Deep Learning for Computer Vision" Packt Publishing, 2018
4. James Phoenix, 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

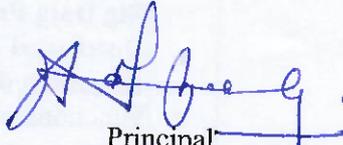
Reference Books:

1. Stuart J. Russell and Peter Norvig, "Artificial Intelligence A Modern Approach" 3rd Edition, Pearson Education 2010
2. Daniel Jurafsky, James H. Martin—Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech, Pearson Publication, 2014.
3. Alexander Clark (Editor), Chris Fox (Editor), Shalom Lappin (Editor) "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


Prepared by


Checked by


Head of the Department


Principal



Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
 (Autonomous College Affiliated to the University of Mumbai)
 NAAC Accredited with "A" Grade (CGPA : 3.18)



Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
 (Autonomous College Affiliated to the University of Mumbai)
 NAAC Accredited with "A" Grade (CGPA : 3.18)



Program: Computer Engineering **Final Year B.Tech.** **Semester: VII**
Course: Cloud Computing for Data Analysis (DJS23CH1401)
Course: Cloud Computing for Data Analysis Laboratory (DJS23CH1401L)
Pre-requisite: Knowledge of

1. Data Science fundamentals
2. Programming in Python/R
3. Basic understanding of databases

Objectives:

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

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

1. Select and deploy appropriate cloud services for various data analysis scenarios
2. Design and implement scalable data processing pipelines in cloud environments
3. Apply cloud-native technologies to optimize data analysis workflows
4. Implement security best practices for cloud-based data solutions
5. Evaluate cloud service providers and their offerings for specific data analysis needs
6. Optimize cloud resource utilization for cost-effectiveness and performance

Cloud Computing for Data Analysis (DJS23CH1401)		
Unit	Description	Duration
1	Introduction to Cloud Computing for Data Analysis 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	06
2	Cloud Storage Solutions for Data Science 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	08
3	Big Data Processing in the Cloud 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	08



	patterns for cloud data services, Performance monitoring and optimization	
4	Cloud-Based Machine Learning and AI Services 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	08
5	Data Visualization and Business Intelligence in the Cloud 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	06
6	Security, Compliance, and Best Practices 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 optimization strategies, Sustainability considerations in cloud computing	06
	Total	42

Cloud Computing for Data Analysis Laboratory (DJS23CH1401L)

Exp.	Suggested experiments
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.)



Batchwise laboratory work of **minimum 8** experiments from the above suggested list or any other experiment based on syllabus will be included, which would help the learner to apply the concepts learned.

Books Recommended:

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," 2nd 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," 2nd Edition, O'Reilly Media, 2023
5. Valentin, Jonas and Patel, Janani, "Practical MLOps: Cloud Deployment and Automation for Machine Learning," O'Reilly Media, 2023

Prepared by

Checked by

Head of the Department

Principal