	SCALABLE DATA SYSTEMS		Semester	7
	Course Code	BAD714A	CIE Marks	50
	Teaching Hours/Week (L:T:P: S)	3:0:0:0	SEE Marks	50
	Total Hours of Pedagogy	50	Total Marks	100
	Credits	04	Exam Hours	3
	Examination type (SEE)	Theory		

Course objectives:

1. To introduce distributed and cloud computing system models and enabling technologies.
2. To understand the architecture and components of scalable data processing using MapReduce and Spark.
3. To explore streaming platforms such as Kafka and their applications in real-time data systems.
4. To familiarize students with containerization and orchestration using Docker and Kubernetes.
5. To apply the knowledge of distributed systems for real-world scalable data processing solutions.

Teaching-Learning Process

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

1. Lecturer method (L) needs 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
9. Use any of these methods: Chalk and board, Active Learning, Case Studies

Module-1

Distributed System Models and Enabling Technologies: Scalable Computing Over the Internet, Technologies for Network-Based Systems, System Models for Distributed and Cloud Computing, Software Environments for Distributed Systems and Clouds, Performance, Security, and Energy Efficiency

Text book 1: Chapter 1 (1.1 to 1.5)

Module-2

	MapReduce Basics : Functional Programming , Mappers and Reducers , The Execution Framework , Partitioners and Combiners , The Distributed File System , Hadoop Cluster Architecture . **Basic MapReduce Algorithm Design:** Local Aggregation , Pairs and Stripes **Text book 2: Chapter 2(2.1 to 2.6), Chapter 3 (3.1 and 3.2)**
	Module-3
	Big Data Processing with Spark :Introduction to Data Analysis with Spark, Programming with RDDs, Working with Key/Value Pairs. **Text book 3: Chapter 1, 3 and 4**
	Module-4
	Introduction to Kafka: Messages and Batches , Schemas, Topics and Partitions , Producers and Consumers , Brokers and Clusters , Multiple Clusters, Benefits of Kafka, Use cases, Configuring Kafka, Programming with Kafka: Producers and Consumers. **Text book 4 : Chapter 1**
	Module-5
	Introduction : Need for Kubernetes, Container Technology, Introduction to Kubernetes. **Creating, running, and sharing a container image** :Building the container image, Running the container image, Exploring the inside of a running container , Stopping and removing a container , Pushing the image to an image registry, Setting up an alias and command-line completion for kubect. **Pods: running containers in Kubernetes** : Introducing pods , Creating a simple YAML descriptor for a pod, Organizing pods with labels. **Textbook 5:Part 1(1.1 to 1.3 and 2.1,2.2) Part 2(3.1 to 3.3)**

Course outcome

At the end of the course, the student will be able to :

1. Explain the Distributed Systems from A Network Perspective.
2. Apply MapReduce Algorithms for any given Problem.
3. Illustrate the Processing of Data in Spark Architecture.
4. Configure Kafka for Producer and Consumer Architecture.
5. Make use of Kubernetes for Building, Running, and Sharing a Container Image.

Assessment Details (both CIE and SEE)

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 40% of the maximum marks (20 marks out of 50) and for the SEE minimum passing mark is 35% of the maximum marks (18 out of 50 marks). A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/ course if the student secures a minimum of 40% (40 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together.

Continuous Internal Evaluation:

- For the Assignment component of the CIE, there are 25 marks and for the Internal Assessment Test component, there are 25 marks.
- The first test will be administered after 40-50% of the syllabus has been covered, and the second test will be administered after 85-90% of the syllabus has been covered
- Any two assignment methods mentioned in the 22OB2.4, if an assignment is project-based then only one assignment for the course shall be planned. The teacher should not conduct two assignments at the end of the semester if two assignments are planned.
- For the course, CIE marks will be based on a scaled-down sum of two tests and other methods of assessment.

Internal Assessment Test question paper is designed to attain the different levels of Bloom's taxonomy as per the outcome defined for the course.

Semester-End Examination:

Theory SEE will be conducted by University as per the scheduled timetable, with common question papers for the course (**duration 03 hours).**

1. The question paper will have ten questions. Each question is set for 20 marks.
2. There will be 2 questions from each module. Each of the two questions under a module (with a maximum of 3 sub-questions), **should have a mix of topics** under that module.
3. The students have to answer 5 full questions, selecting one full question from each module.
4. Marks scored shall be proportionally reduced to 50 marks

Suggested Learning Resources:

Textbooks

1. Kai Hwang, G. C. Fox, J.J. Dongarra "Distributed & Cloud Computing", Morgan Kauffman Publishers
2. Data-Intensive Text Processing, with Map Reduce, Jimmy Lin and Chris Dyer
3. Learning Spark, Holden Karau, et al., 1st Edition
4. Kafka: The Definitive Guide: Real-Time Data and Stream Processing at Scale
5. Kubernetes in Action, MARKO LUKŠA

Web links and Video Lectures (e-Resources):

https://www.youtube.com/watch?v=b-IvmXoO0bU
https://www.youtube.com/watch?v=R5D35zZa_vo
https://www.youtube.com/watch?v=FhdNV6704bU
https://www.youtube.com/watch?v=DzAco4Aq3mw

Activity Based Learning (Suggested Activities in Class)/ Practical Based learning

Assignment 1 (10 marks)

- Choose any **one scalable data technology** (Hadoop, Spark, Kafka, Kubernetes, Amazon Web Services (AWS) for Big Data, Google BigQuery or Dataflow). Write a report regarding What is the technology? Key features and architecture. Use cases in the real world. Advantages & limitations and a Simple diagram or workflow.

Assignment 2 (15 marks)

- Select **one real-world scenario**, such as:
  - Online shopping site tracking customer clicks
  - Social media platform processing video uploads
  - Real-time weather data processing
  - Sensor data from IoT devices in smart homes

  Design a **conceptual scalable data pipeline** for the scenario. The design should include:
  - Data source and ingestion method (e.g., REST API, sensors)
  - Data processing tool (e.g., Spark, MapReduce)
  - Real-time/streaming tool (e.g., Kafka, Flink)
  - Data storage options (e.g., HDFS, cloud storage)

PARALLEL COMPUTING		Semester	7
Course Code	**BCI714B**	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
Examination nature (SEE)	Theory		

Course objectives:

This course will enable to,

- Explore the need for parallel programming
- Explain how to parallelize on MIMD systems
- To Interpret how to apply MPI library and parallelize the suitable programs
- To analyze how to apply OpenMP pragma and directives to parallelize the suitable programs
- To demonstrate how to design CUDA program

Teaching-Learning Process (General Instructions)

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

1. Lecturer method (L) need not to be only traditional lecture methods, 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 Programming assignment, which fosters student's Analytical skills, develop design thinking skills such as the ability to design, evaluate, generalize, and analyze information rather than simply recall it.

MODULE-1

Introduction to parallel programming, Parallel hardware and parallel software – Classifications of parallel computers, SIMD systems, MIMD systems, Interconnection networks, Cache coherence, Shared-memory vs. distributed-memory, Coordinating the processes/threads, Shared-memory, Distributed-memory.

Textbook1: Chapter-1, Chapter 2

MODULE-2

GPU programming, Programming hybrid systems, MIMD systems, GPUs, Performance – Speedup and efficiency in MIMD systems, Amdahl's law, Scalability in MIMD systems, Taking timings of MIMD programs, GPU performance.

Textbook1: Chapter2

MODULE-3

Distributed memory programming with MPI – MPI functions, The trapezoidal rule in MPI, Dealing with I/O, Collective communication, MPI-derived datatypes, Performance evaluation of MPI programs, A parallel sorting algorithm.

Textbook1: Chapter 3

MODULE-4

Shared-memory programming with OpenMP – openmp pragmas and directives, The trapezoidal rule, Scope of variables, The reduction clause, loop carried dependency, scheduling, producers and consumers, Caches, cache coherence and false sharing in openmp, tasking, tasking, thread safety.

Textbook1: Chapter 5

MODULE-5
GPU programming with CUDA - GPUs and GPGPU, GPU architectures, Heterogeneous computing, Threads, blocks, and grids Nvidia compute capabilities and device architectures, Vector addition, Returning results from CUDA kernels, CUDA trapezoidal rule I, CUDA trapezoidal rule II: improving performance, CUDA trapezoidal rule III: blocks with more than one warp. **Textbook1: Chapter 6**

Course outcomes (Course Skill Set):

At the end of the course, the student will be able to:

- Explain the need for parallel programming.
- Illustrate parallelism in MIMD system.
- Make use of MPI library to parallelize the code to solve a problem.
- Interpret OpenMP pragma and directives to parallelize the code to solve the given problem.
- Demonstrate a CUDA program for the given problem.

Assessment Details (both CIE and SEE)

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 40% of the maximum marks (20 marks out of 50) and for the SEE minimum passing mark is 35% of the maximum marks (18 out of 50 marks). A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/ course if the student secures a minimum of 40% (40 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together.

Continuous Internal Evaluation:

- For the Assignment component of the CIE, there are 25 marks and for the Internal Assessment Test component, there are 25 marks.
- The first test will be administered after 40-50% of the syllabus has been covered, and the second test will be administered after 85-90% of the syllabus has been covered
- Any two assignment methods mentioned in the 22OB2.4, if an assignment is project-based then only one assignment for the course shall be planned. The teacher should not conduct two assignments at the end of the semester if two assignments are planned.
- For the course, CIE marks will be based on a scaled-down sum of two tests and other methods of assessment.

Internal Assessment Test question paper is designed to attain the different levels of Bloom's taxonomy as per the outcome defined for the course.

Semester-End Examination:

Theory SEE will be conducted by University as per the scheduled timetable, with common question papers for the course (**duration 03 hours).**

1. The question paper will have ten questions. Each question is set for 20 marks.
2. There will be 2 questions from each module. Each of the two questions under a module (with a maximum of 3 sub-questions), **should have a mix of topics** under that module.
3. The students have to answer 5 full questions, selecting one full question from each module.
4. Marks scored shall be proportionally reduced to 50 marks

Suggested Learning Resources:

Textbook:

1. Peter S Pacheco, Matthew Malensek – An Introduction to Parallel Programming, second edition, Morgan Kauffman, 2021.

Reference Books:

1. Michael J Quinn – Parallel Programming in C with MPI and OpenMp, McGrawHill, 2004.
2. Calvin Lin, Lawrence Snyder – Principles of Parallel Programming, Pearson
3. Barbara Chapman – Using OpenMP: Portable Shared Memory Parallel Programming, Scientific and Engineering Computation
4. 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

Activity Based Learning (Suggested Activities in Class)/ Practical Based learning

Programming Assignment - 25 marks

1. Write a OpenMP program to calculate n Fibonacci numbers using tasks.
2. Write a OpenMP program to find the prime numbers from 1 to n employing parallel for directive. Record both serial and parallel execution times.
3. Write a MPI Program to demonstration of MPI_Send and MPI_Recv.
4. Write a MPI program to demonstration of deadlock using point to point communication and avoidance of deadlock by altering the call sequence
5. Write a MPI Program to demonstration of Broadcast operation.
6. Write a MPI Program demonstration of MPI_Scatter and MPI_Gather
7. Write a MPI Program to demonstration of MPI_Reduce and MPI_Allreduce (MPI_MAX, MPI_MIN, MPI_SUM, MPI_PROD)

	DATA ENGINEERING AND MLOps		Semester	7
	Course Code	BAD714C	CIE Marks	50
	Teaching Hours/Week (L:T:P: S)	3:0:0:0	SEE Marks	50
	Total Hours of Pedagogy	50	Total Marks	100
	Credits	04	Exam Hours	3
	Examination type (SEE)	Theory		

Course objectives:

1. To introduce the concepts and lifecycle of Data Engineering.
2. To explore principles of data architecture and distributed systems.
3. To familiarize students with MLOps pipelines for scalable ML solutions.
4. To understand model deployment, CI/CD, monitoring, and feedback loops.
5. To ensure governance, reproducibility, and responsible AI compliance.

Teaching-Learning Process

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

1. Lecturer method (L) needs 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
9. Use any of these methods: Chalk and board, Active Learning, Case Studies

Module-1

Data Engineering: Definition, The Data Engineering Lifecycle, Evolution of the Data Engineer, Data Engineering and Data Science, Data Engineering Skills and Activities, Data Maturity and the Data Engineer, The Background and Skills of a Data Engineer, Business Responsibilities, Technical Responsibilities, The Continuum of Data Engineering Roles, Data Engineers Inside an Organization ,

Internal-Facing Versus External-Facing Data Engineers, Data Engineers and Other Technical Roles, Data Engineers and Business Leadership.

Data Engineering Lifecycle: The Data Lifecycle Versus the Data Engineering Lifecycle, Generation: Source Systems, Major Undercurrents Across the Data Engineering Lifecycle

Textbook 1:Chapter 1 (1.1–1.5), Chapter 2 (2.1–2.4)

Module-2

Data Architecture: Enterprise Architecture Defined, Data Architecture Defined, "Good" Data Architecture, Principles of Good Data Architecture, Major Architecture Concepts, Domains and Services , Distributed Systems, Scalability, and Designing for Failure ,Tight Versus Loose Coupling: Tiers, Monoliths, and Microservices , User Access: Single Versus Multitenant , Event-Driven Architecture , Examples and Types of Data Architecture

Choosing Technologies Across the Data Engineering Lifecycle: Team Size and Capabilities, Speed to Market, Interoperability, Cost Optimization and Business Value, Total Cost of Ownership Total Opportunity Cost of Ownership, FinOps, Today Versus the Future: Immutable Versus Transitory Technologies: Hybrid Cloud, Multicloud , Decentralized: Blockchain and the Edge ,Monolith Versus Modular , Serverless Versus Servers, Server Versus Serverless evaluation

	Textbook 1:Chapter 3 (3.1–3.7), Chapter 4 (4.1–4.6)
	Module-3
	MLOps Challenges, MLOps to Mitigate Risk, Risk Assessment, Risk Mitigation, MLOps for Responsible AI,MLOps for Scale. Key MLOps Features: Model Development, Establishing Business Objectives, Data Sources and Exploratory Data Analysis, Feature Engineering and Selection, Training and Evaluation, Reproducibility, Responsible AI, Productionalization and Deployment, Model Deployment Types and Contents, Model Deployment Requirements, Monitoring Developing Models: Machine Learning Model, Required Components, Different ML Algorithms, Different MLOps Challenges, Data Exploration, Feature Engineering and Selection, Feature Engineering Techniques, How Feature Selection Impacts MLOps Strategy, Experimentation, Evaluating and Comparing Models, Choosing Evaluation Metrics, CrossChecking Model Behavior, Impact of Responsible AI on Modeling, Version Management and Reproducibility **Textbook 2: Chapter 1 (1.1–1.3), Chapter 2 (2.1–2.4)**
	Module-4
	Preparing for Production: Runtime Environments, Adaptation from Development to Production Environments, Data Access Before Validation and Launch to Production, Final Thoughts on Runtime Environments, Model Risk Evaluation, The Purpose of Model Validation, The Origins of ML Model Risk, Quality Assurance for Machine Learning. Deploying to Production: CI/CD Pipelines, Building ML Artifacts, The Testing Pipeline, Deployment Strategies, Categories of Model Deployment, Considerations When Sending Models to Production, Maintenance in Production, Containerization, Scaling Deployments, Requirements and Challenges. **Textbook 2:Chapter 3 (3.1–3.5), Chapter 4 (4.1–4.4)**
	Module-5 10 hours
	Monitoring and Feedback Loop: Models Be Retrained, Understanding Model Degradation, Ground Truth Evaluation, Input Drift Detection, Drift Detection in Practice, Example Causes of Data Drift, Input Drift Detection Techniques, The Feedback Loop, Logging, Model Evaluation, Online Evaluation Model Governance: Governance the Organization Needs, Matching Governance with Risk Level, Current Regulations Driving MLOps Governance, Pharmaceutical Regulation in the US: GxP Financial Model Risk Management Regulation, GDPR and CCPA Data Privacy Regulations, The New Wave of AI-Specific Regulation, The Emergence of Responsible AI, Key Elements of Responsible AI (Element 1 to Element 5), A Template for MLOps Governance (Step 1 to 8). **Textbook 2:Chapter 5 (5.1–5.4), Chapter 6 (6.1–6.3)**

Course outcome

At the end of the course, the student will be able to :

1. Explain Data Engineering and various roles.
2. Analyze various major architecture concepts of Data engineering.
3. Apply MLOps Features and analyze the challenges in developing and Deploying Machine Learning Models
4. Design CI/CD Pipelines for Deploying Machine Learning Models
5. Explain the need for model governance and MLOps governance.

Assessment Details (both CIE and SEE)

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 40% of the maximum marks (20 marks out of 50) and for the SEE minimum passing mark is 35% of the maximum marks (18 out of 50 marks). A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/ course if the student secures a minimum of 40% (40 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together.

Continuous Internal Evaluation:

- For the Assignment component of the CIE, there are 25 marks and for the Internal Assessment Test component, there are 25 marks.
- The first test will be administered after 40-50% of the syllabus has been covered, and the second test will be administered after 85-90% of the syllabus has been covered
- Any two assignment methods mentioned in the 22OB2.4, if an assignment is project-based then only one assignment for the course shall be planned. The teacher should not conduct two assignments at the end of the semester if two assignments are planned.
- For the course, CIE marks will be based on a scaled-down sum of two tests and other methods of assessment.

Internal Assessment Test question paper is designed to attain the different levels of Bloom's taxonomy as per the outcome defined for the course.

Semester-End Examination:

Theory SEE will be conducted by University as per the scheduled timetable, with common question papers for the course (**duration 03 hours).**

1. The question paper will have ten questions. Each question is set for 20 marks.
2. There will be 2 questions from each module. Each of the two questions under a module (with a maximum of 3 sub-questions), **should have a mix of topics** under that module.
3. The students have to answer 5 full questions, selecting one full question from each module.
4. Marks scored shall be proportionally reduced to 50 marks

Suggested Learning Resources:

Textbooks:

1. Joe Reis, Matt Housley, Fundamentals of Data Engineering, O'Reilly, 2022
2. Mark Treveil & Dataiku Team, Introducing MLOps, O'Reilly, 2020

Web links and Video Lectures (e-Resources):

https://www.ibm.com/think/topics/data-engineering
https://martinfowler.com/articles/microservices.html
https://www.coursera.org/specializations/mlops

Activity Based Learning (Suggested Activities in Class)/ Practical Based learning

- Assignment 1 (15 marks) : Select a simple machine learning use case (e.g., house price prediction, customer churn prediction, or fraud detection). Design an MLOps pipeline that includes
  a. Problem statement and business objective
  b. Data sources and exploratory data analysis summary
  c. Feature engineering and selection approach
  d. Model training and evaluation plan
  e. Reproducibility and version control strategy
  Draw a flowchart or block diagram representing the MLOps pipeline
- Assignment 2(10 marks): Choose any one open-source MLOps tool (like MLflow, Kubeflow, or TFX). Study how the tool supports
  a. CI/CD pipelines
  b. Model testing & validation

c. Deployment strategies.

Write a report on tool overview and core components, CI/CD support, One deployment strategy explained in context, Advantages & challenges of using this tool in a real project.

BIG DATA ANALYTICS		Semester	7
Course Code	**BCS714D**	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	3
Examination nature (SEE)	Theory		

Course objectives:

1. To implement MapReduce programs for processing big data.
2. To realize storage and processing of big data using MongoDB, Pig, Hive and Spark.
3. To analyze big data using machine learning techniques.

Teaching-Learning Process (General Instructions)

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

1. Lecturer method (L) needs 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. Discuss how every concept can be applied to the real world - and when that's possible, it helps improve the students' understanding.
6. Use any of these methods: Chalk and board, Active Learning, Case Studies.

MODULE-1

Classification of data, Characteristics, Evolution and definition of Big data, What is Big data, Why Big data, Traditional Business Intelligence Vs Big Data,Typical data warehouse and Hadoop environment.

Big Data Analytics: What is Big data Analytics, Classification of Analytics, Importance of Big Data Analytics, Technologies used in Big data Environments, Few Top Analytical Tools , NoSQL, Hadoop.

TB1: Ch 1: 1.1, Ch2: 2.1-2.5,2.7,2.9-2.11, Ch3: 3.2,3.5,3.8,3.12, Ch4: 4.1,4.2

MODULE-2

Introduction to Hadoop: Introducing hadoop, Why hadoop, Why not RDBMS, RDBMS Vs Hadoop, History of Hadoop, Hadoop overview, Use case of Hadoop, HDFS (Hadoop Distributed File System),Processing data with Hadoop, Managing resources and applications with Hadoop YARN(Yet Another Resource Negotiator).

Introduction to Map Reduce Programming: Introduction, Mapper, Reducer, Combiner, Partitioner, Searching, Sorting, Compression.

TB1: Ch 5: 5.1-,5.8, 5.10-5.12, Ch 8: 8.1 - 8.8

MODULE-3

Introduction to MongoDB: What is MongoDB, Why MongoDB, Terms used in RDBMS and MongoDB, Data Types in MongoDB, MongoDB Query Language.

TB1: Ch 6: 6.1-6.5

MODULE-4

Introduction to Hive: What is Hive, Hive Architecture, Hive data types, Hive file formats, Hive Query Language (HQL), RC File implementation, User Defined Function (UDF).

Introduction to Pig: What is Pig, Anatomy of Pig, Pig on Hadoop, Pig Philosophy, Use case for Pig, Pig Latin Overview, Data types in Pig, Running Pig, Execution Modes of Pig, HDFS Commands, Relational Operators, Eval Function, Complex Data Types, Piggy Bank, User Defined Function, Pig Vs Hive.

TB1: Ch 9: 9.1-9.6,9.8, Ch 10: 10.1 - 10.15, 10.22

MODULE-5

Spark and Big Data Analytics: Spark, Introduction to Data Analysis with Spark.

Text, Web Content and Link Analytics: Introduction, Text Mining, Web Mining, Web Content and Web Usage Analytics, Page Rank, Structure of Web and Analyzing a Web Graph.
TB2: Ch5: 5.2,5.3, Ch 9: 9.1-9.4

Course outcomes (Course Skill Set):
At the end of the course, the student will be able to:

- Illustrate Big Data concepts, tools and applications.
- Develop programs using HADOOP framework.
- Use Hadoop Cluster to deploy Map Reduce jobs, PIG,HIVE and Spark programs.
- Analyze the given data set to identify deep insights.

Assessment Details (both CIE and SEE)

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 40% of the maximum marks (20 marks out of 50) and for the SEE minimum passing mark is 35% of the maximum marks (18 out of 50 marks). A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/ course if the student secures a minimum of 40% (40 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together.

Continuous Internal Evaluation:

For the Assignment component of the CIE, there are 25 marks and for the Internal Assessment Test component, there are 25 marks.

The first test will be administered after 40-50% of the syllabus has been covered, and the second test will be administered after 85-90% of the syllabus has been covered

Any two assignment methods mentioned in the 22OB2.4, if an assignment is project-based then only one assignment for the course shall be planned. The teacher should not conduct two assignments at the end of the semester if two assignments are planned.

For the course, CIE marks will be based on a scaled-down sum of two tests and other methods of assessment.

Internal Assessment Test question paper is designed to attain the different levels of Bloom's taxonomy as per the outcome defined for the course.

Semester-End Examination:

- Theory SEE will be conducted by University as per the scheduled timetable, with common question papers for the course (duration 03 hours).
- The question paper will have ten questions. Each question is set for 20 marks.
- There will be 2 questions from each module. Each of the two questions under a module (with a maximum of 3 sub-questions), should have a mix of topics under that module.
- The students have to answer 5 full questions, selecting one full question from each module.
- Marks scored shall be proportionally reduced to 50 marks.

Suggested Learning Resources:

Books:

1. Seema Acharya and Subhashini Chellappan "Big data and Analytics", Wiley India Publishers, 2nd Edition, 2019.
2. Rajkamal and Preeti Saxena, "Big Data Analytics, Introduction to Hadoop, Spark and Machine Learning", McGraw Hill Publication, 2019.

Reference Books:

1. Adam Shook and Donald Mine, "MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems" - O'Reilly 2012

2. Tom White, "Hadoop: The Definitive Guide" 4th Edition, O'reilly Media, 2015.
3. Thomas Erl, Wajid Khattak, and Paul Buhler, Big Data Fundamentals: Concepts, Drivers & Techniques, Pearson India Education Service Pvt. Ltd., 1st Edition, 2016
4. John D. Kelleher, Brian Mac Namee, Aoife D'Arcy -Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, MIT Press 2020, 2nd Edition

Web links and Video Lectures (e-Resources):

- https://www.youtube.com/watch?v=bAyrObl7TYE&list=PLEiEAq2VkUUJqp1k-g5W1mo37urJQOdCZ
- https://www.youtube.com/watch?v=VmO0QgPCbZY&list=PLEiEAq2VkUUJqp1kg5W1mo37urJQOdCZ&index=4
- https://www.youtube.com/watch?v=GG-VRm6XnNk https://www.youtube.com/watch?v=JglO2Nv_92A

Activity Based Learning (Suggested Activities in Class)/ Practical Based learning

1. Implement MongoDB based application to store big data for data processing and analyzing the results [15 marks]
2. Install Hadoop and Implement the following file management such as Adding files and directories, Retrieving files, Deleting files and directories and execute Map- Reduce based programs.[10]

Introduction to DBMS		Semester	7
Course Code	**BCS755A**	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	3
Examination nature (SEE)	Theory		

Course objectives:

- To Provide a strong foundation in database concepts, technology, and practice.
- To Practice SQL programming through a variety of database problems.
- To Understand the relational database design principles.
- To Demonstrate the use of concurrency in database.
- To Design and build database applications for real world problems.

Teaching-Learning Process (General Instructions)

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

- Lecturer method (L) needs not to be only a traditional lecture method, but alternative effective teaching methods could be adopted to attain the outcomes.
- Use of Video/Animation to explain functioning of various concepts.
- Encourage collaborative (Group Learning) Learning in the class.
- Ask at least three HOT (Higher order Thinking) questions in the class, which promotes critical thinking.
- Discuss how every concept can be applied to the real world - and when that's possible, it helps improve the students' understanding.
- Use any of these methods: Chalk and board, Active Learning, Case Studies.

MODULE-1

Introduction to Databases: Introduction, Characteristics of database approach, Advantages of using the DBMS approach, History of database applications.

Overview of Database Languages and Architectures: Data Models, Schemas, and Instances. Three schema architecture and data independence, database languages, and interfaces, The Database System environment.

Textbook 1:Ch 1.1 to 1.8, 2.1 to 2.6

MODULE-2

Conceptual Data Modeling using Entities and Relationships: Entity types, Entity sets and structural constraints, Weak entity types, ER diagrams, Specialization and Generalization.

Mapping Conceptual Design into a Logical Design: Relational Database Design using ER-to-Relational mapping

Textbook 1: Ch 3.1 to 3.10, 9.1 & 9.2

MODULE-3

Relational Model: Relational Model Concepts, Relational Model Constraints and relational database schemas, Update operations, transactions, and dealing with constraint violations.

Relational Algebra: Unary and Binary relational operations, additional relational operations (aggregate, grouping, etc.) Examples of Queries in relational algebra.

Textbook 1: Ch 5.1 to 5.3, Ch 8.1 to 8.5

MODULE-4

SQL: SQL data definition and data types, Schema change statements in SQL, specifying constraints in SQL, retrieval queries in SQL, INSERT, DELETE, and UPDATE statements in SQL, Additional features of SQL

Normalization: Database Design Theory – Introduction to Normalization using Functional and Multivalued Dependencies: Informal design guidelines for relation schema, Functional Dependencies, Normal Forms based on Primary Keys, Second and Third Normal Forms, Boyce-Codd Normal Form, Multivalued Dependency and Fourth Normal Form, Join Dependencies and Fifth Normal Form.

Textbook 1: Ch 6.1 to 6.5,14.1 to 14.7

MODULE-5

SQL: Advanced Queries: More complex SQL retrieval queries, Specifying constraints as assertions and action triggers, Views in SQL.

Concurrency Control in Databases: Two-phase locking techniques for Concurrency control, Concurrency control based on Timestamp ordering, Multiversion Concurrency control techniques, Validation Concurrency control techniques, Granularity of Data items and Multiple Granularity Locking.

Textbook 1: Ch 7.1 to 7.3, 21.1 to 21.5

Course outcomes (Course Skill Set):
At the end of the course, the student will be able to:

- Demonstrate the basic elements of a database management system.
- Design entity relationship and convert entity relationship diagrams into RDBMS.
- Use Structured Query Language (SQL) for database manipulation.
- Apply normalization to increase the efficiency of database design.
- Illustrate the concepts of concurrency control techniques.

Assessment Details (both CIE and SEE)

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 40% of the maximum marks (20 marks out of 50) and for the SEE minimum passing mark is 35% of the maximum marks (18 out of 50 marks). A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/ course if the student secures a minimum of 40% (40 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together.

Continuous Internal Evaluation:

For the Assignment component of the CIE, there are 25 marks and for the Internal Assessment Test component, there are 25 marks.

The first test will be administered after 40-50% of the syllabus has been covered, and the second test will be administered after 85-90% of the syllabus has been covered

Any two assignment methods mentioned in the 22OB2.4, if an assignment is project-based then only one assignment for the course shall be planned. The teacher should not conduct two assignments at the end of the semester if two assignments are planned.

For the course, CIE marks will be based on a scaled-down sum of two tests and other methods of assessment.

Internal Assessment Test question paper is designed to attain the different levels of Bloom's taxonomy as per the outcome defined for the course.

Semester-End Examination:

- Theory SEE will be conducted by University as per the scheduled timetable, with common question papers for the course (duration 03 hours).
- The question paper will have ten questions. Each question is set for 20 marks.
- There will be 2 questions from each module. Each of the two questions under a module (with a maximum of 3 sub-questions), should have a mix of topics under that module.
- The students have to answer 5 full questions, selecting one full question from each module.
- Marks scored shall be proportionally reduced to 50 marks.

Suggested Learning Resources:

Text Books:

1. Fundamentals of Database Systems, Ramez Elmasri and Shamkant B. Navathe, 7th Edition, 2017, Pearson.

Reference Books:

1. Database management systems, Ramakrishnan, and Gehrke, 3rd Edition, 2014, McGraw Hill

Activity Based Learning (Suggested Activities in Class)/ Practical Based learning

Course Project (25 marks)

- For any problem selected
    - Develop the application having at least five tables & domain areas shall include health care, agriculture & so on.

	Introduction to Algorithms		Semester	7
	Course Code	BCS755B	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
	Examination type (SEE)	Theory		

Course objectives:

- To learn the methods for analyzing algorithms and evaluating their performance.
- To demonstrate the efficiency of algorithms using asymptotic notations.
- To solve problems using various algorithm design methods, including brute force, greedy, divide and conquer, decrease and conquer, transform and conquer, dynamic programming, backtracking, and branch and bound.
- To learn the concepts of P and NP complexity 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) does not mean only the traditional lecture method, but different types of teaching methods may be adopted to achieve the outcomes.
2. Utilize video/animation films to illustrate the functioning of various concepts.
3. Promote collaborative learning (Group Learning) in the class.
4. Pose at least three HOT (Higher Order Thinking) questions in the class to stimulate critical thinking.
5. Incorporate Problem-Based Learning (PBL) to foster students' analytical skills and develop their ability to evaluate, generalize, and analyze information rather than merely recalling it.
6. Introduce topics through multiple representations.
7. Demonstrate various ways to solve the same problem and encourage students to devise their own creative solutions.
8. Discuss the real-world applications of every concept to enhance students' comprehension.

Module-1

INTRODUCTION: What is an Algorithm?, Fundamentals of Algorithmic Problem Solving, Important problem Types, Fundamental Data Structures, Analysis Framework, Asymptotic Notations and Basic Efficiency Classes, ,Analysis Framework, Asymptotic Notations and Basic Efficiency Classes,

Chapter 1 (Sections 1.1 to 1.4), Chapter 2 (2.1, 2.2)

Module-2

FUNDAMENTALS OF THE ANALYSIS OF ALGORITHM EFFICIENCY: Mathematical Analysis of Non-recursive Algorithms, Mathematical Analysis of Recursive Algorithms.

BRUTE FORCE APPROACHES: Selection Sort and Bubble Sort, Sequential Search and Brute Force String Matching.

Chapter 2(Sections 2.3,2.4), Chapter 3(Section 3.1,3.2)

	Module-3
	Exhaustive Search (Travelling Salesman problem and Knapsack Problem). **Depth First search and Breadth First search.** **DECREASE-AND-CONQUER:** Insertion Sort, Topological Sorting. **DIVIDE AND CONQUER:** Merge Sort, Binary Tree Traversals. **Chapter 3(3.4,3.5), Chapter 4 (Sections 4.1,4.2), Chapter 5 (Section 5.1,5.3)**
	Module-4
	TRANSFORM-AND-CONQUER: Balanced Search Trees (AVL Trees), Heaps and Heapsort. **SPACE-TIME TRADEOFFS:** Sorting by Counting: Comparison counting sort, Input Enhancement in String Matching: Horspool's Algorithm, Hashing. **Chapter 6 (Sections 6.3,6.4), Chapter 7 (Sections 7.1,7.2, 7.3)**
	Module-5
	DYNAMIC PROGRAMMING: Three basic examples, The Knapsack Problem and Memory Functions. **THE GREEDY METHOD:** Kruskal's Algorithm, Dijkstra's Algorithm, Huffman Trees and Codes. **Chapter 8 (Sections 8.1,8.2), Chapter 9 (Sections 9.2,9.3,9.4)**

Course outcome (Course Skill Set)

At the end of the course, the student will be able to:

1. Explain the algorithm design steps and computational problem types.
2. Apply the asymptotic notational method to analyze the performance of the algorithms in terms of time complexity.
3. Demonstrate divide & conquer approaches and decrease & conquer approaches to solve computational problems.
4. Make use of the transform & conquer design approach to solve the given real-world or complex computational problems.
5. Apply greedy and dynamic programming methods to solve graph & string-based computational problems.

Assessment Details (both CIE and SEE)

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 40% of the maximum marks (20 marks out of 50) and for the SEE minimum passing mark is 35% of the maximum marks (18 out of 50 marks). A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/ course if the student secures a minimum of 40% (40 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together.

Continuous Internal Evaluation:

- For the Assignment component of the CIE, there are 25 marks and for the Internal Assessment Test component, there are 25 marks.
- The first test will be administered after 40-50% of the syllabus has been covered, and the second test will be administered after 85-90% of the syllabus has been covered
- Any two assignment methods mentioned in the 22OB2.4, if an assignment is project-based then only one assignment for the course shall be planned. The teacher should not conduct two assignments at the end of the semester if two assignments are planned.
- For the course, CIE marks will be based on a scaled-down sum of two tests and other methods of assessment.

Internal Assessment Test question paper is designed to attain the different levels of Bloom's taxonomy as per the outcome defined for the course.

Semester-End Examination:

Theory SEE will be conducted by University as per the scheduled timetable, with common question papers for the course (**duration 03 hours).**

1. The question paper will have ten questions. Each question is set for 20 marks.
2. There will be 2 questions from each module. Each of the two questions under a module (with a maximum of 3 sub-questions), **should have a mix of topics** under that module.
3. The students have to answer 5 full questions, selecting one full question from each module.
4. Marks scored shall be proportionally reduced to 50 marks

Suggested Learning Resources:

Textbooks

1. Introduction to the Design and Analysis of Algorithms, By Anany Levitin, 3rd Edition (Indian), 2017, Pearson.

Reference books

1. Computer Algorithms/C++, Ellis Horowitz, SatrajSahni and Rajasekaran, 2nd Edition, 2014, Universities Press.
2. Introduction to Algorithms, Thomas H. Cormen, Charles E. Leiserson, Ronal L. Rivest, Clifford Stein, 3rd Edition, PHI.
3. Design and Analysis of Algorithms, S. Sridhar, Oxford (Higher Education)

Web links and Video Lectures (e-Resources):

- Design and Analysis of Algorithms: https://nptel.ac.in/courses/106/101/106101060/

Activity Based Learning (Suggested Activities in Class)/ Practical Based learning

1. Problem Solving - Competitive programming (Hacker Rank/ Hacker Earth / Leetcode) – 10 Marks
2. Problem solving (Numerical examples) related to different algorithms – 15 Marks

	SOFTWARE ENGINEERING		Semester	7
	Course Code	**BCS755C**	CIE Marks	50
	Teaching Hours/Week (L:T:P: S)	3:0:0:0	SEE Marks	50
	Total Hours of Pedagogy	50	Total Marks	100
	Credits	04	Exam Hours	3
	Examination type (SEE)	Theory		

Course objectives:

To understand foundational principles and the evolving nature of software engineering.
- To learn various software process models and their practical applications.
- To acquire skills in gathering, modeling, and validating software requirements.
- To apply Agile methodologies and understand core software engineering practices.
- To build a foundation for software design, testing, and quality assurance.

Teaching-Learning Process

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

1. Lecturer method (L) needs 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
9. Use any of these methods: Chalk and board, Active Learning, Case Studies

Module-1

Software and Software Engineering: The nature of Software, The unique nature of WebApps, Software Engineering, The software Process, Software Engineering Practice, Software Myths.

Process Models: A generic process model, Process assessment and improvement, Prescriptive process models: Waterfall model, Incremental process models, Evolutionary process models, Concurrent models, Specialized process models. Unified Process , Personal and Team process models

Textbook 1: Chapter 1: 1.1 to 1.6, Chapter 2: 2.1 to 2.5

Module-2

	Understanding Requirements: Requirements Engineering, Establishing the ground work, Eliciting Requirements, Developing use cases, Building the requirements model, Negotiating Requirements, Validating Requirements. **Requirements Modeling Scenarios, Information and Analysis classes**: Requirement Analysis, Scenario based modeling, UML models that supplement the Use Case, Data modeling Concepts, Class-Based Modeling. Requirement Modeling Strategies : Flow oriented Modeling , Behavioral Modeling. **Textbook 1: Chapter 5: 5.1 to 5.7, Chapter 6: 6.1 to 6.5, Chapter 7: 7.1 to 7.3**
	Module-3
	Agile Development: What is Agility?, Agility and the cost of change. What is an agile Process?, Extreme Programming (XP), Other Agile Process Models, A tool set for Agile process . **Principles that guide practice**: Software Engineering Knowledge, Core principles, Principles that guide each framework activity. **Textbook 1: Chapter 3: 3.1 to 3.6, Chapter 4: 4.1 to 4.3**
	Module-4
	Software Design: Design within the context of software engineering, Design process and quality, Design concepts: abstraction, modularity, architecture, patterns. **Architectural Design**: Architectural styles and patterns, reference architectures, component-level design, designing class-based components, conducting component-level design, design for reuse. **Textbook 1:Chapter 8: 8.1–8.6, Chapter 9: 9.1–9.5**
	Module-5
	Software Testing: Introduction to software testing, Strategic approach, Test strategies for conventional and object-oriented software, Validation testing, System testing, White-box and Black-box testing, Basis Path Testing, Control structure testing. **Software Quality:** Concepts of quality, Software quality assurance, Reviews, Software reliability and metrics. **Textbook 1: Chapter 14: Sections 14.1 to 14.5,Chapter 15: Sections 15.1 to 15.5, Chapter 19: Sections 19.1 to 19.5**

Course outcome

At the end of the course, the student will be able to :

1. Explain the software nature, engineering practices, myths, and software process models.
2. Apply requirements engineering, elicitation, modeling, and validation in software development.
3. Demonstrate agile principles, practices, and tools for software development agility.
4. Apply design concepts, process, and architecture for quality software development.
5. Explain software testing strategies and quality assurance for reliable software.

Assessment Details (both CIE and SEE)

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 40% of the maximum marks (20 marks out of 50) and for the SEE minimum passing mark is 35% of the maximum marks (18 out of 50 marks). A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/ course if the student secures a minimum of 40% (40 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together.

Continuous Internal Evaluation:

- For the Assignment component of the CIE, there are 25 marks and for the Internal Assessment Test component, there are 25 marks.
- The first test will be administered after 40-50% of the syllabus has been covered, and the second test will be administered after 85-90% of the syllabus has been covered
- Any two assignment methods mentioned in the 22OB2.4, if an assignment is project-based then only one assignment for the course shall be planned. The teacher should not conduct two assignments at the end of the semester if two assignments are planned.
- For the course, CIE marks will be based on a scaled-down sum of two tests and other methods of assessment.

Internal Assessment Test question paper is designed to attain the different levels of Bloom's taxonomy as per the outcome defined for the course.

Semester-End Examination:

Theory SEE will be conducted by University as per the scheduled timetable, with common question papers for the course (**duration 03 hours).**

1. The question paper will have ten questions. Each question is set for 20 marks.
2. There will be 2 questions from each module. Each of the two questions under a module (with a maximum of 3 sub-questions), **should have a mix of topics** under that module.
3. The students have to answer 5 full questions, selecting one full question from each module.
4. Marks scored shall be proportionally reduced to 50 marks

Suggested Learning Resources:

Textbook

Roger S. Pressman: Software Engineering – A Practitioner's Approach, 7th Edition, Tata McGraw Hill, 2010.

Web links and Video Lectures (e-Resources):

https://www.geeksforgeeks.org/software-engineering/software-engineering/

Activity-Based Learning (Suggested Activities in Class)/Practical-Based learning

- Course project (Group of two students): Simulation that covers all the phases of SDLC - 25 marks

	Introduction to Machine Learning		Semester	7
	Course Code	BCI755D	CIE Marks	50
	Teaching Hours/Week (L:T:P: S)	3:0:0:0	SEE Marks	50
	Total Hours of Pedagogy	50	Total Marks	100
	Credits	04	Exam Hours	3
	Examination type (SEE)	Theory		

Course objectives:

1. To introduce the fundamental concepts and techniques of machine learning.
2. To understand various types of machine learning techniques and challenges that are faced in real-world applications.
3. To evaluate machine learning models for different types of problems.
4. To be familiar with learning approaches such as regression, decision trees, clustering, and neural networks.

Teaching-Learning Process

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

1. Lecturer method (L) needs 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
9. Use any of these methods: Chalk and board, Active Learning, Case Studies

Module-1

Introduction: Need for Machine Learning, Machine Learning Explained, Machine Learning in Relation to other Fields, Types of Machine Learning, Challenges of Machine Learning, Machine Learning Process, Machine Learning Applications.
Understanding Data – 1: Introduction, Big Data Analysis Framework, Descriptive Statistics, Univariate Data Analysis and Visualization.
Chapter-1, 2 (2.1-2.5)

Module-2

Understanding Data – 2: Bivariate Data and Multivariate Data, Multivariate Statistics, Essential Mathematics for Multivariate Data, Feature Engineering and Dimensionality Reduction Techniques.
Basic Learning Theory: Design of Learning System, Introduction to Concept of Learning, Modelling in Machine Learning.
Chapter-2 (2.6-2.8, 2.10), Chapter-3 (3.3, 3.4, 3.6)

Module-3

Similarity-based Learning: Nearest-Neighbor Learning, Weighted K-Nearest-Neighbor Algorithm, Nearest Centroid Classifier, Locally Weighted Regression (LWR).
Regression Analysis: Introduction to Regression, Introduction to Linear Regression, Validation of Regression Methods, Multiple Linear Regression, Polynomial Regression, Logistic Regression.
Chapter-4 (4.2-4.5), Chapter-5 (5.1-5.7)

Module-4

	Decision Tree Learning: Introduction to Decision Tree Learning Model, Decision Tree Induction Algorithms, Validation and pruning of decision trees. **Bayesian Learning:** Introduction to Probability-based Learning, Fundamentals of Bayes Theorem, Classification Using Bayes Model- Naive Bayes , Naïve Bayes Algorithm for Continuous Attributes. **Chapter-6 (6.1-6.3) Chapter-8 (8.1-8.3.1, 8.4)**
	Module-5
	Artificial Neural Networks: Introduction, Artificial Neurons, Perceptron and Learning Theory, Types of Artificial Neural Networks, Popular Applications of Artificial Neural Networks, Advantages and Disadvantages of ANN. **Clustering Algorithms:** Introduction to Clustering Approaches, Proximity Measures, Hierarchical Clustering Algorithms, Partitional Clustering Algorithm, Density-based Methods **Chapter-10 (10.1,10.3-10.5, 10.9,10.10), Chapter -13 (13.1-13.5)**

Course outcome

At the end of the course, the student will be able to :

1. Explain the machine learning techniques, their types and data analysis framework.
2. Apply mathematical concepts for feature engineering and perform dimensionality reduction to enhance model performance.
3. Develop similarity-based learning models and regression models for solving classification and prediction tasks.
4. Develop probabilistic learning models and neural network models using perceptrons and multilayer architectures.
5. Utilize clustering algorithms to identify patterns in data .

Assessment Details (both CIE and SEE)

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 40% of the maximum marks (20 marks out of 50) and for the SEE minimum passing mark is 35% of the maximum marks (18 out of 50 marks). A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/ course if the student secures a minimum of 40% (40 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together.

Continuous Internal Evaluation:

- For the Assignment component of the CIE, there are 25 marks and for the Internal Assessment Test component, there are 25 marks.
- The first test will be administered after 40-50% of the syllabus has been covered, and the second test will be administered after 85-90% of the syllabus has been covered
- Any two assignment methods mentioned in the 22OB2.4, if an assignment is project-based then only one assignment for the course shall be planned. The teacher should not conduct two assignments at the end of the semester if two assignments are planned.
- For the course, CIE marks will be based on a scaled-down sum of two tests and other methods of assessment.

Internal Assessment Test question paper is designed to attain the different levels of Bloom's taxonomy as per the outcome defined for the course.

Semester-End Examination:

Theory SEE will be conducted by University as per the scheduled timetable, with common question papers for the course (**duration 03 hours).**

1. The question paper will have ten questions. Each question is set for 20 marks.
2. There will be 2 questions from each module. Each of the two questions under a module (with a maximum of 3 sub-questions), **should have a mix of topics** under that module.
3. The students have to answer 5 full questions, selecting one full question from each module.
4. Marks scored shall be proportionally reduced to 50 marks

Suggested Learning Resources:

Books

1. S Sridhar, M Vijayalakshmi, "Machine Learning", OXFORD University Press 2021, First Edition.

Web links and Video Lectures (e-Resources):

- https://www.universitiespress.com/resources?id=9789393330697
- https://www.drssridhar.com/?page_id=1053
- Machine Learning Tutorials: https://www.geeksforgeeks.org/machine-learning/
- Machine Learning Tutorials: https://www.tutorialspoint.com/machine_learning/index.htm
- Python for Machine Learning: https://www.w3schools.com/python/python_ml_getting_started.asp
- Introduction to Machine Learning: https://onlinecourses.nptel.ac.in/noc22_cs29/preview

Activity Based Learning (Suggested Activities in Class)/ Practical Based learning

- Course project that inputs a messy data set, pre-processes it , builds model and evaluates the performance . (15 (experiment)+10 (report) marks)