Bachelor of Computer Applications

(Data Science)

Semester–III

Course Code	Course Title	L	T	P	Credits
BCA-112A	Data Structures using C++ (Pre-requisite: C Programming)	3	0	0	3
		CIE	SEE		Total
		40	60		100

Course Outcomes (COs): At the end of this course, students will be able to	
BCA-112A.1	Understand the fundamentals of Object-Oriented Programming (OOP) in C++
BCA-112A.2	Explore the Data Structures concepts and operations of Arrays and Strings
BCA-112A.3	Examine the features and operations of Stack and Queue data structure.
BCA-112A.4	Explore the implementation and operations of linked lists applying searching and sorting algorithms.

Course Outcomes (CO) to Program Outcomes (PO) mapping (scale 1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
BCA-112A.1	3	-	-	-	-	-	-	-
BCA-112A.2	3	2	-	-	-	-	-	2
BCA-112A.3	3	2	-	2	-	-	-	2
BCA-112A.4	3	2	2	2	-	-	-	2

Instructions for Paper Setter: The examiner will set 9 questions asking two questions from each unit and one compulsory question. The compulsory question (Question No. 1) will consist at least 4 parts covering entire syllabus. The examinee will be required to attempt 5 questions, selecting one question from each unit and the compulsory question. All questions will carry equal marks.

UNIT-I **Contact Hours: 10**

Object-Oriented Programming concepts

Object-Oriented Programming: Object-Oriented programming features and benefits. Object-Oriented features of C++, Class and Objects, Data Hiding & Encapsulation, Structures, Data members and Member functions, Scope resolution operator and its significance, Accessing Members of Class and Structure, Passing and Returning Objects to Functions.

UNIT-II **Contact Hours: 11**

Introduction to Data Structures and Arrays

Basic data Structure: Introduction to Data Structures, algorithms and Pseudo code, Types of Data Structures, Relationship among data, data structures, and algorithms, Abstract Data Types, Analysis of Algorithms, asymptotic notations, asymptotic analysis.

Arrays, Representation of Arrays in Memory, Operations on Array (Traversing, Insertion, Deletion), Strings

and its Representation in Memory, Operations on Strings (length, reverse).

UNIT-III **Contact Hours: 12**

Stacks and Queues

Stacks: Introduction, Representation of Stacks Using Sequential Organization (Arrays), Applications of Stack, Expression Evaluation and Conversion, Polish notation and expression conversion, Processing of Function Calls, Reversing a String with a Stack, Recursion. Stack Abstract Data Type

Queues: Concept of Queues, Realization of Queues Using Arrays, Circular Queue, Multi-queues, Dequeue, Priority Queue, Applications of Queues, Queue as Abstract Data Type

UNIT-IV **Contact Hours: 12**

Linked Lists, Searching and Sorting techniques

Linked Lists: Introduction, Types (singly, double, circular, circular double) and Operations (Insertion, Deletion, Traversal, Searching, Sorting), Applications.

Searching and Sorting: Search Techniques-Linear Search, Binary Search, Sorting Techniques- Selection Sort, Bubble Sort, Insertion Sort, Merge Sort, Quick Sort, Comparison of Sorting Methods

Text Books:

1. Michael T. Goodrich, R. Tamassia and Mount, *Data structures and Algorithms in C++,* John Wiley and Sons.
2. Mark Allen Weiss, *Data structures and Algorithm Analysis in C++,* Pearson Education.
3. Robert L. Kruse and A.J. Ryba, *Data Structures and Program Design in C++,* Prentice Hall, Inc., NJ
4. Herbert Scildt, *C++ The Complete Reference*, Tata McGraw-Hill

Reference Books:

1. Narasimha Karamanchi, *Data Structures and Algorithms Made Easy,* CareerMonk Publications
2. Adam Drozdek, *Data Structures and Algorithms in C++,* Course Technology.
3. Balaguruswami, E., *Object Oriented Programming In C++,* Tata McGraw-Hill.

Course Code	Course Title	L	T	P	Credits
BCA-271A	Data Structures Lab (Pre-requisite: C++)	0	0	4	2
		CIE	SEE		Total
		50	50		100

Course Outcomes (COs): At the end of this course, students will be able to

BCA-271A.1	Write programs in C++ implementing various operations on Array
BCA-271A.2	Implement various operations on Linked lists.
BCA-271A.3	Write programs implementing various operations on Stack and Queue
BCA-271A.4	Implement various Searching and Sorting techniques in C++.

Course Outcomes (CO) to Program Outcomes (PO) mapping (scale 1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
BCA-271A.1	3	2	-	2	-	-	-	2
BCA-271A.2	3	2	2	2	-	-	-	2
BCA-271A.3	3	2	2	2	-	-	-	2
BCA-271A.4	3	2	2	2	-	-	-	2

List of Experiments

No.	Experiment Detail
1.	Write a program to remove the duplicates from an array.
2.	Write a program that prints a table indicating the number of occurrences of each alphabet in the text entered.
3.	Write a program that uses functions to perform the following operations on an array (a) Creation (b) Insertion (c) Deletion (d) Traversal.
4.	Write a program to implement simple Stack using array to perform the following operations: (a) Create (b) Peek (c) Push (d) Pop (e) Traverse
5.	Write a menu driven program that implements singly linked list for the following operations: (b) Creation (b) Insertion (c) Deletion (d) Traversal.
6.	Write a menu driven program that implements doubly linked list for the following operations: Create, Display, Insert, Delete, Search,
7.	Write a program to implement simple Stack using Link list to perform the following operations: (a) Create (b) Peek (c) Push (d) Pop (e) Traverse
8.	Write a program to implement simple Queue using array to perform the following operations: (a) Create (b) Enqueue (c) Dequeue (d) Traverse
9.	Write a program to implement circular Queue using array to perform the following operations: (a) Create (b) Enqueue (c) Dequeue (d) Traverse
10.	Write a program to search an elements in an array using linear search technique.
11.	Write a program to search an elements in an array using binary search technique.
12.	Write a program to sort the elements using Insertion Sort algorithm.

13.	Write a program to sort the elements using Bubble Sort algorithm.
14.	Write a program to sort the elements using Quick Sort algorithm.
15.	Write a program to sort the elements using Merge Sort algorithm.
16.	Write a program to sort the elements using Selection Sort algorithm.

Text Books:

1. Michael T. Goodrich, R. Tamassia and Mount, *Data structures and Algorithms in C++*, John Wiley and Sons.
2. Mark Allen Weiss, *Data structures and Algorithm Analysis in C++,* Pearson Education.
3. Robert L. Kruse and A.J. Ryba, *Data Structures and Program Design in C++*, Prentice Hall, Inc., NJ
4. Herbert Scildt, *C++ The Complete Reference*, Tata McGraw-Hill

Reference Books:

1. Narasimha Karamanchi, *Data Structures and Algorithms Made Easy,* CareerMonk Publications
2. Adam Drozdek, *Data Structures and Algorithms in C++*, Course Technology.
3. Balaguruswami, E., *Object Oriented Programming In C++*, Tata McGraw-Hill.

Course Code	Course Title	L	T	P	Credits
BCA-203A	Database Management System (Pre-requisite: None)	3	0	0	3
		CIE	SEE		Total
		40	60		100

Course Outcomes (COs): At the end of this course, students will be able to	
BCA-203A.1	Understand the fundamentals of DBMS, data models, schema architecture, and ER/EER design.
BCA-203A.2	Apply SQL and PL/SQL concepts to create, query, and manage relational databases.
BCA-203A.3	Normalize database schemas using functional dependencies and various normal forms.
BCA-203A.4	Analyze transaction processing, concurrency control, and data warehousing concepts.

Course Outcomes (CO) to Program Outcomes (PO) mapping (scale 1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
BCA-203A.1	3	-	-	-	-	-	-	-
BCA-203A.2	3	2	2	2	-	-	-	-
BCA-203A.3	3	2	2	2	-	-	-	2
BCA-203A.4	3	3	-	2	-	2	2	3

Instructions for Paper Setter: The examiner will set 9 questions asking two questions from each unit and one compulsory question. The compulsory question (Question No. 1) will consist at least 4 parts covering entire syllabus. The examinee will be required to attempt 5 questions, selecting one question from each unit and the compulsory question. All questions will carry equal marks.

UNIT-I **Contact Hours: 11**

Database and Database Management System

Introduction and need of Database and Database Management System (DBMS). Characteristics of DBMS, Database Users, Data Models, Schema, and Instances. Three-Schema architecture and data Independence.

Database Design using ER model: Entity types, Entity set, attributes and Keys. Relationship types. Expressing M:N relation. Enhanced Entity Relationship Model (EER): Specialization, Generalization, Attribute inheritance and Aggregation

UNIT-II **Contact Hours: 12**

Relational model and Relational Algebra

Relational model concepts. Introduction to SQL, Types of SQL commands (DDL, DML, and DCL etc),SQL operators and their procedure, views and indexes. Queries and sub queries, Group by and Having clause, Aggregate functions and Constraints. PL/SQL: Architecture of PL/SQL, Basic Elements of PL/SQL.

Relational Algebra: Introduction of Relational Algebra, Selection and projection. Renaming, Joins, Unions, Intersection, Cartesian product and Division.

UNIT-III **Contact Hours: 10**

Normalization

Need for Normalization, anomalies (insert, delete and update), Functional Dependencies, Minimal set of

Functional Dependencies, Normal Forms: 1NF, 2NF, 3NF. Higher Level Normal Forms: Boyce/ Codd Normal Form, multi-valued dependency (MVD), Fourth Normal Form, Join dependencies and Fifth Normal Form.

UNIT-IV **Contact Hours: 12**

Transaction processing, Overview of data mining and warehousing

Transaction Processing Concepts: Transaction and Schedules, transaction properties, concurrent execution of transaction, Conflict and View serializability, testing for serializability, concepts in Recoverable and Cascadeless schedules.Concurrency Control Techniques: Lock based protocols, Two phase locking technique, time stamp based protocols, validation based protocols.

Overview of data mining technology, Introduction of data warehousing, Characteristics of data warehouses.

Text Books:

1. Ramez Elmasri and Shamkant B. Navathe, *Fundamentals of Database Systems*, Pearson
2. Abraham Silberschatz, Henry F. Korth, and S. Sudarshan, *Database System Concepts*, McGraw Hill
3. Ivan Bayross, *SQL, PL/SQL the Programming Language of Oracle*, BPB

Reference Books:

1. Raghu Ramakrishnan and Johannes Gehrke, *Database Management Systems*, McGraw Hill
2. Markus Winand, *SQL Performance Explained*, Self Publishing

Course Code	Course Title	L	T	P	Credits
BCA-273A	Database Management System Lab (Pre-requisite: DBMS, PL/SQL)	0	0	4	2
		CIE	SEE		Total
		50	50		100

Course Outcomes (COs): At the end of this course, students will be able to

BCA-273A.1	Understand the fundamentals ER design.
BCA-273A.2	Apply SQL and PL/SQL concepts to create, query, and manage relational databases.
BCA-273A.3	Perform pattern matching and other operations
BCA-273A.4	Write PL/SQL programs using selection, loop, exception and Normalization.

Course Outcomes (CO) to Program Outcomes (PO) mapping (scale 1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
BCA-273A.1	3	-	-	-	-	-	-	-
BCA-273A.2	3	2	3	2	-	2	2	-
BCA-273A.3	3	-	-	-	-	-	-	-
BCA-273A.4	3	2	3	2	-	-	2	2

List of Experiments

No.	Experiment Detail
1.	Design an ER diagram for a college database with students, departments, and courses.
2.	Convert the ER diagram into relational schema and identify primary, foreign keys.
3.	Create tables in a relational DBMS using SQL based on the ER schema.
4.	Execute basic SQL commands: DDL, DML, DCL on a sample database
5.	Demonstrate the use of Data Constraints.
6.	Write SQL queries using WHERE, BETWEEN, LIKE, IN, and ORDER BY.
7.	Use aggregate functions with GROUP BY and HAVING clauses.
8.	Demonstrate pattern matching and range searching functions.
9.	Demonstrate the use of ORACLE operators • Arithmetic operators • Comparison operators • Logical operators
10.	Demonstrate the use of DUAL table
11.	Demonstrate the use of INDEX in DBMS
12.	Perform different types of joins (INNER, OUTER).
13.	Write PL/SQL programs using variables, IF-THEN, loops, and exceptions.
14.	Identify anomalies and perform step-by-step normalization (1NF to 3NF) on a sample unnormalized relation.

15.	Demonstrate the concept of locking using a small transaction simulation.

Text Books:

1. Ramez Elmasri and Shamkant B. Navathe, *Fundamentals of Database Systems*, Pearson
2. Abraham Silberschatz, Henry F. Korth, and S. Sudarshan, *Database System Concepts*, McGraw Hill
3. Ivan Bayross, *SQL, PL/SQL the Programming Language of Oracle,* BPB

Reference Books:

1. Raghu Ramakrishnan and Johannes Gehrke, *Database Management Systems*, McGraw Hill
2. Markus Winand, *SQL Performance Explained*, Self Publishing

Course Code	Course Title	L	T	P	Credits
BCA-221A	Introduction to Data Science (Pre-requisite: None)	3	0	0	3
		CIE	SEE		Total
		40	60		100

Course Outcomes (COs): At the end of this course, students will be able to	
BCA-221A.1	Understand the scope, evolution, and real-world applications of data science, and explain the phases of a typical data science project.
BCA-221A.2	Identify and evaluate data sources, collection methods, and data quality issues.
BCA-221A.3	Explain data preprocessing, integration techniques, and storage models.
BCA-221A.4	Apply data analysis concepts and principles of effective data visualization.

Course Outcomes (CO) to Program Outcomes (PO) mapping (scale 1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
BCA-221A.1	3	-	-	-	-	-	-	-
BCA-221A.2	3	2	-	-	-	-	-	2
BCA-221A.3	3	2	-	2	-	-	-	2
BCA-221A.4	3	3	3	2	-	2	-	2

Instructions for Paper Setter: The examiner will set 9 questions asking two questions from each unit and one compulsory question. The compulsory question (Question No. 1) will consist at least 4 parts covering entire syllabus. The examinee will be required to attempt 5 questions, selecting one question from each unit and the compulsory question. All questions will carry equal marks.

UNIT-I **Contact Hours: 12**

Introduction to Data Science and Analysis

Definition and scope of Data Science, Historical evolution of the field, Importance of data-driven decision-making, Applications of Data Science**,** Healthcare (predicting diseases, patient data analysis) ,Finance (fraud detection, credit scoring), Phases of a Data Science Project: Business ,Understanding Data Collection, Data Preparation, exploratory Data Analysis

UNIT-II **Contact Hours: 11**

Data Collection Methods, Data Sources and Tools

Data Collection Methods (Surveys, Web Scraping, Sensors, Logs, APIs), Types of Data Sources (Databases, Data Warehouses, Open Data Repositories), Internal vs. External Data Sources, Metadata and Data Provenance, Data Quality Issues (Accuracy, Completeness, Consistency, Timeliness, Validity), Data Governance and Stewardship, Introduction to Big Data Ecosystems (Hadoop, Spark – Conceptual)

UNIT-III **Contact Hours: 12**

Data pre-processing techniques and Data storage models

Data Cleaning (Missing Data, Outliers, Noise), Data Transformation (Normalization, Aggregation, Discretization), Data Integration and Fusion (Schema Integration, Redundancy Analysis), Data Reduction Techniques (Sampling, Dimensionality Reduction – Conceptual), Data Storage Models (Relational, NoSQL,

Cloud Storage), Data Lifecycle Management (DLM**)**

UNIT-IV **Contact Hours: 10**

Descriptive Analytics and Data visualization techniques

Descriptive Analytics, Inferential Analytics, Exploratory Data Analysis (EDA), Types of Variables, Scales of Measurement, Central Tendency (Mean, Median, Mode), Dispersion (Variance, Standard Deviation), Data Distributions, Principles of Data Visualization, Types of Charts (Bar, Line, Histogram, Scatter, Heatmap), Color Theory and Perception, Misleading Visualizations, Dashboard Design, Storytelling with Data, Data Ethics in Visualization

Text Books:

1. Foster Provost & Tom Fawcett, *Data Science for Business*, O'Reilly Media
2. EMC Education Services, *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*, Wiley
3. Cathy O'Neil & Rachel Schutt, *Doing Data Science*, O'Reilly Media
4. David Stephenson, *Data Science for Executives: Leveraging Machine Intelligence to Drive Business ROI*, O'Reilly Media

Reference Books:

1. Kelleher, John D., & Brendan Tierney, *Data Science*, The MIT Press Essential Knowledge Series
2. Thomas H. Davenport & D.J. Patil, *Keeping Up with the Quants: Your Guide to Understanding and Using Analytics*, Harvard Business Review Press
3. V. K. Jain, *Data Science and Analytics*, Khanna Book Publishing
4. Davy Cielen, Arno D. B. Meysman & Mohamed Ali, *Introducing Data Science: Big Data, Machine Learning, and More*, Manning Publications

Course Code	Course Title	L	T	P	Credits
BCA-223A	Fundamentals of R Programming (Pre-requisite: None)	3	0	0	3
		CIE	SEE		Total
		40	60		100

Course Outcomes (COs): At the end of this course, students will be able to

BCA-223A.1	Describe the features of R Programming.
BCA-223A.2	Use the various data structures in R.
BCA-223A.3	Apply data frames, control statements and functions for the simulation.
BCA-223A.4	Use R for data entry, management, and basic analysis, Tables, charts, Graphics and Statistics

Course Outcomes (CO) to Program Outcomes (PO) mapping (scale 1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
BCA-223A.1	3	-	-	-	-	-	-	-
BCA-223A.2	3	-	2	-	-	-	-	2
BCA-223A.3	3	2	3	2	-	-	-	2
BCA-223A.4	3	2	3	2	-	2	-	3

Instructions for Paper Setter: The examiner will set 9 questions asking two questions from each unit and one compulsory question. The compulsory question (Question No. 1) will consist at least 4 parts covering entire syllabus. The examinee will be required to attempt 5 questions, selecting one question from each unit and the compulsory question. All questions will carry equal marks.

UNIT-I Contact Hours: 11

Basics of R programming and Data structures

Introduction to Programing language, Overview f R Language and its Applications, Why to use R language, Difference between Python And R Language, introduction of IDE. Basic Data Types: numeric, character, logical, Integer, complex, Variables and Strings.
R-Operators: Arithmetic Operators, Logical Operators, Relational Operators, Assignment Operators.
Data Structures: Vectors, Lists, Data frames , Matrices, Arrays, Factors, Tibbles, Help Function NA and Null values, Using All and Any.

UNIT-II Contact Hours: 12

Control statements, String, Vectors, and Matrices

Conditional statements: If Statement, If-Else Statement, Looping: While, Do-While, Functions and Recursions in R.
String in R: Various Types of Strings, Adding and Deleting Data into String. String Manipulation Functions.
Vectors in R: Types of Vectors, Operations on Vectors.
Matrices: Creating Matrices, Matrix Operations, Matrix Functions, Add, update and Delete Rows and Columns. Vector/Matrices Distinction, Introduction to array.

UNIT-III Contact Hours: 10

List and Data Frames

List: Creating List, Operations on List, Accessing List Data, Functions on List, Recursive List, Various Operations in R using List, Matrix, Array.
Data Frames: Creating and Manipulating Data Frames, Operations on Data Frames in form of a Matrix, Subsetting and Indexing, Data Manipulation with "dplyrin". Directly read Data-File

UNIT-IV **Contact Hours: 12**

Data manipulation, visualization and Statistical methods

Tables: Creating Table, Operation on Tables. Factors: Creating Factor, Functions used with Factors, operations on Factors.
Charts & Graphs: How to create PIE Chart, Bar Charts, Bar Plot, Line Graphs, Scatterplots, Histogram and Box Plot .
Statistics: Means, Median & Mode, Linear Regression, Multiple Regression, Logistic Regression

Text Books:

1. Crawley, M. J., *The R book*. John Wiley & Sons.
2. Gardener, M., *Beginning R: The statistical programming language*. John Wiley & Sons.
3. Aphalo, P. J., *Learn R: as a language*. Chapman and Hall/CRC.

Reference Books:

1. Jared P. Lander, *R for Everyone: Advanced Analytics and Graphics*, Pearson Education.
2. Michael J. Crawley, *The R Book*, John Wiley & Sons
3. Brian S. Everitt, Toresten Hothorn, *A Handbook of Statistical Analyses using R*, CRC Press

Course Code	Course Title	L	T	P	Credits
BCA-291A	R Programming Lab (Pre-requisite: R language)	0	0	4	2
		CIE	SEE		Total
		50	50		100

Course Outcomes (COs): At the end of this course, students will be able to

BCA-291A.1	Understand the installation process and environment setup of R and RStudio
BCA-291A.2	Apply basic R programming constructs such as data types, lists, vectors, and operators
BCA-291A.3	Develop programs using control structures like loops and conditional statements in R
BCA-291A.4	Analyze and manipulate complex data structures like lists, vectors, and matrices &Perform basic statistical analysis and data manipulation using R and its packages

Course Outcomes (CO) to Program Outcomes (PO) mapping (scale 1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
BCA-291A.1	3	-	2	-	-	-	-	-
BCA-291A.2	3	-	2	-	-	-	-	-
BCA-291A.3	3	2	3	2	-	-	-	2
BCA-291A.4	3	3	3	3	-	2	2	2

List of Experiments

No.	Experiment Detail
1.	Write steps for the Installation of R and RStudio
2.	Write an R program to create list containing strings, numbers, vectors and logical value.
3.	Design a program to solve quadratic equation in R language.
4.	Calculate the sum of first 100 natural numbers by using the concept of loops .
5.	Write a program to calculate Even or Odd number by using the concept of control statement.
6.	Write an R program to find all elements of a given list that are not in another list.
7.	Write a program to convert list to vector, merging list.
8.	Write an R program to add 10 to each element of the first vector in the given list.
9.	Write a program for: Create a variable outside a function & use it inside the function. Create a variable inside a function & use it outside the function.
10.	Create a program to implement various operation on matrix.
11.	Write an R program to create data set & find mean, mode and median data set
12.	Write an R program to read and load data from larger data set.
13.	Write a program to create sample data in R and perform data manipulation
14.	Write a program to manipulate data with dplyr package and data.table package

15.	Write a program to create histogram, piechart & barchart from csu file.

Text Books:

1. Crawley, M. J., *The R book*. John Wiley & Sons.
2. Gardener, M., *Beginning R: The statistical programming language*. John Wiley & Sons.
3. Aphalo, P. J., *Learn R: as a language*. Chapman and Hall/CRC.

Reference Books:

1. Jared P. Lander, *R for Everyone: Advanced Analytics and Graphics*, Pearson Education.
2. Michael J. Crawley, *The R Book*, John Wiley & Sons
3. Brian S. Everitt, Toresten Hothorn, *A Handbook of Statistical Analyses using R*, CRC Press