Diploma in Big Data Analytics

Syllabus

CENTRE FOR DISTANCE AND ONLINE EDUCATION

PROGRAMME CODE

2026 – 2027 ONWARDS

Centre for Distance and Online Education

BHARATHIAR UNIVERSITY

(A State University, Accredited with "A++" Grade by NAAC, Ranked 46th among Indian Universities by MHRD-NIRF)

Coimbatore - 641 046, Tamil Nadu, India

DIPLOMA IN BIG DATA ANALYTICS

SYLLABUS

(WITH EFFECT FROM 2026 ONWARDS)

PROGRAM CODE:

DEPARTMENT OF COMPUTER APPLICATION
CENTRE FOR DISTANCE AND ONLINE EDUCATION
Bharathiar University
(A State University, Accredited with "A++" Grade by NAAC and
46th Rank among Indian Universities by MHRD-NIRF)
Coimbatore 641 046, INDIA

BHARATHIAR UNIVERSITY : COIMBATORE 641046
DEPARTMENT OF COMPUTER APPLICATION

VISION

Bestow globally comparable quality education on youth, embodied with character building, to invoke the University's motto "Educate to Elevate" and uphold the secular ideals of the nation as envisioned by Mahakavi Subramania Bharathiar.

MISSION

- To be innovative and inclusive, committed to excelling in teaching, research and knowledge transfer and to serving the social, cultural and economic needs of the nation.
- To innovate and offer educational programmes in various disciplines through synergistic interaction with the industry and society.
- To impart knowledge and skills and to provide a learning environment to acquire attitudes to students and equip them to face the emerging challenges of the knowledge era.
- To provide equal opportunity to women students, differently abled, and minorities and prepare them to be equal partners in accomplishing the scientific and technological demands of the nation.
- To contribute to the advancement of knowledge through applied research leading to newer products and processes.
- To prepare the students to work for societal transformation with a commitment to justice and equality and emerge as job providers.
- To inculcate in students a global vision with skills of international competence.

TITLE OF THE PROGRAM

Diploma in Big Data Analytics

DURATION

- **Minimum Duration**: One year
- **Maximum Duration:** Two year (with provision for Re-registration if required)

CREDIT DISTRIBUTION

- **Total Credits**: 40 credits (as per UGC credit framework)

ELIGIBILITY

10 +2 (Higher Secondary) or equivalent from a recognized board unless otherwise specified.

MEDIUM OF INSTRUCTION

Medium of Instruction is **English** . However the learners are permitted to write the Continuous Internal Assessment and End Semester Examination in **English or Tamil**.

FEES STRUCTURE

Fee structure as approved by the Centre for Distance and Online Education, Bharathiar University, Coimbatore.

ATTENDANCE

- **Online Mode**:
    - Minimum 75% attendance in online sessions/webinars/tutorials recorded on LMS.
    - 120 hours of live online sessions are mandatory for 40 credit diploma programme (3 hours per credit).
- **ODL Mode**:
    - Participation of atleast Ten Personal Contact Programs (PCPs) with 60 hours of contact class (3 hours per credit) during one semester is mandatory.
    - A total Participation of atleast Twenty Personal Contact Programs (PCPs) with 120 hours of contact class (3 hours per credit) for entire programme is mandatory.

REQUIREMENT TO APPEAR FOR THE EXAMINATIONS

A candidate enrolled for the programme must have the minimum period of attendance in theory and practical classes prescribed to appear for the examinations with the fees prescribed and fulfills all other conditions stipulated from time to time.

CONDITIONS TO APPEAR FOR THE EXAMINATIONS

The learner appearing for the examination for the first time should register for all the course by remitting the fee

ODL

- The learners should specify the opted center of examinations in the examination application form.
- Learners will not be permitted to change their examination center.
- Learners shall collect their E-hall tickets from the website for the examination and get signature from the Chief Superintendent of the examinations centers opted, on
  submission of their identity cards during the three working days before the commencement of the examinations.

OL

- The learners shall collect their E-hall tickets from the website for the examination and write the examinations through online mode.
- The learners can download the question paper and scan & upload their answer script from their registered user id.

EVALUATION PATTERN

a. Distribution of Marks in Continuous Internal Assessments (CIA): 25% weightage

The following procedure shall be followed for awarding internal marks for theory courses.

- Continuous Internal Assessment (CIA) Test - 10 marks
- Assignment - 5 marks
- Seminar - 5 marks
- Participation - 5 marks

Total - 25 marks

1. **CIA Test and question pattern** :

   Two Continuous Internal Assessments (CIA-I & CIA II) must be conducted. Better of the TWO will be counted for Test Marks (10 Marks)

   - Section A – Objective Question 5*1 = 5 marks
   - Section B – Short Essay (2 out of 3) 2*5 = 10 marks
   - Section C – Essay Type (Either or type) 1*10 = 10 marks

2. **Assignments – I and II**. Better of the TWO will be counted for Assignment Marks (5 Marks)
3. **Seminar** will be conducted for 5 Marks.
4. **Participation** (which includes attendance) – 5 Marks

b. Project/Dissertation/Internship (4 Credits) – At the end of Second Semester

- Major project / Internship (100 marks)

  Internal – 50 marks & External – 50 Marks

c. End-Semester Examination (ESE): 75% weightage

ODL

- Examinations will be Conducted through offline at a designated physical location (University / Learner Support Centers (LSCs)) with invigilators and a fixed schedule.
- **Duration**: 3 hours, covering theory and applications.

OL

- Conducted through an online/proctored computer based examination (an assessment conducted using a computer or other digital device, replacing traditional pen-and-paper methods).
- **Duration**: 3 hours, covering theory and applications.
- **Examination Integrity:** Online proctoring, biometric verification, or designated centers.

Distribution of marks in the End-Semester Examination - Question Paper Pattern:

Section	Type of question	No. of questions	Marks	Total
A	Objective Question	10 (Compulsory)	1	10
B	Short Essay (300 words each)	3 out of 5	5	15
C	Essay Type (1000 words each)	5 Questions (Either or Type)	10	50
Total				**75**

d. Passing Criteria:

- Minimum 40% passing marks in each component
  (Internal – 10 marks out of 25 & External – 30 marks out of 75 separately).
- Minimum 40% passing marks in practical / project component
- Aggregate of 40% marks required for successful completion of the diploma programme.

e. Grading System : (As per 10-point UGC Credit Framework)

The following table gives the marks, grade points, letter, grades and classification to Indicate the performance of the candidate.

For the entire programme:

- CGPA = Sum of the multiplication of grade points by the credits of the entire programme / Sum of the credits of the courses for the entire programme.
  Cumulative Grade Point Average [CGPA) = $\sum n \sum i C_{ni} G_{ni} / \sum n \sum I C_{ni}$

RANGE	CGPA	Grade	Classification of Result
95-100	9.5-10.0	O+	**First Class- Exemplary***
90-94	9.0 and above but below 9.5	O	
85-89	8.5 and above but below 9.0	D++	**First Class with Distinction***
80-84	8.0 and above but below 8.5	D+	
75-79	7.5 and above but below 8.0	D	
70-74	7.0 and above but below 7.5	A++	**First Class**
65-69	6.5 and above but below 7.0	A+	
60-64	6.0 and above but below 6.5	A	
55-59	5.5 and above but below 6.0	B+	**Second Class**
50-54	5.0 and above but below 5.5	B	
45-49	4.5 and above but below 5.0	C+	**Third Class**
40-44	4.0 and above but below 4.5	C	
0-39	0.0 and above but below 4.0	U	**Re-appear**

a. A candidate who has passed all the courses in the first appearance within the prescribed duration of the diploma programmes and secured a CGPA of 9 to 10 and equivalent

grades "O" or "O+" in Core and Practical/Project courses shall be placed in the category of "**First Class – Exemplary**".

b. A candidate who has passed all the courses in the first appearance within the prescribed duration of the Diploma programmes and secured a CGPA of 7.5 to 9 and equivalent grades "D" or "D+" or "D++" in Core and Practical/Project courses shall be placed in the category of "**First Class with Distinction**".

c. A candidate who has passed all the courses of the Diploma programmes and secured a CGPA of 6 to 7.4 and equivalent grades "A" or "A+" or "A++" in Core and Practical/Project courses shall be declared to have passed in "**First Class**".

d. A candidate who has passed all the courses examination of the Diploma programmes and secured a CGPA of 5.0 to 5.9 and equivalent grades "B" or "B+" in Core and Practical/Project courses shall be declared to have passed in "**Second Class**".

e. A candidate who has passed all the courses examination of the Diploma programmes and secured a CGPA of 4.0 to 4.9 and equivalent grades "C" or "C+" in Core and Practical/Project courses shall be declared to have passed in "Second Class".

BHARATHIAR UNIVERSITY: COIMBATORE 641 046

Department of Computer Application
Diploma in Big Data Analytics
Centre or Distance and Online Education
(For the students admitted during the academic year 2026-2027 onwards)

PROGRAMME STRUCTURE

FIRST SEMESTER

Course Code	Title of the Course	Credits	Maximum Marks			
			Theory	CIA	ESE	Total
	INTRODUCTION TO BIG DATA	4	12	25	75	100
	BASICS OF DATA ANALYTICS	4	12	25	75	100
	DATA PROCESSING TOOL	4	12	25	75	100
	PROGRAMMING IN PYTHON	4	12	25	75	100
	PYTHON PROGRAMMING LAB	4	12	40	60	100
TOTAL		**20**	**60**	**140**	**360**	**500**

SECOND SEMESTER

Course Code	Title of the Course	Credits	Maximum Marks			
			Theory	CIA	ESE	Total
	DATABASE TECHNOLOGIES FOR BIG DATA	4	12	25	75	100
	DATA ENGINEERING	4	12	25	75	100
	CLOUD COMPUTING FOR BIG DATA	4	12	25	75	100
	FUNDAMENTALS OF MACHINE LEARNING	4	12	25	75	100
	MACHINE LEARNING LAB	4	12	40	60	100
TOTAL		**20**	**60**	**140**	**360**	**500**

OVERALL

SEMESTER	CREDIT	MARKS
FIRST	20	500
SECOND	20	500
TOTAL	**40**	**1000**

PROGRAM EDUCATIONAL OBJECTIVES (PEOs)

PROGRAM EDUCATIONAL OBJECTIVES (PEOs)	
PEO1	Build analytical thinking, problem-solving skills, and the ability to apply suitable tools, platforms, and algorithms to process large-scale data.
PEO2	Build analytical thinking, problem-solving skills, and the ability to apply suitable tools, platforms, and algorithms to process large-scale data.
PEO3	Gain skills in data visualization, statistical analysis, data interpretation, reporting, dashboard design, and effective communication.
PEO4	Understand ethical, social, and legal responsibilities in data usage and follow industry standards and performance evaluation methods.
PEO5	Apply Big Data analytics in real-world domains such as business, healthcare, finance, education, and government through projects, research, and innovation.

PROGRAMME OUTCOMES (POs)

PROGRAM OUTCOMES (POs)	
After successful completion of the program, students will be able to:	
PO1	Demonstrate strong knowledge in Big Data Analytics, Data Mining, Machine Learning, Statistical Modeling, and modern data engineering concepts.
PO2	Apply analytical, computational, and statistical skills using tools like Python, R, Hadoop, Spark, SQL, NoSQL, and cloud platforms to solve real-world problems.
PO3	Design data pipelines, scalable systems, dashboards, reports, and visualizations to support business intelligence and decision-making.
PO4	Work effectively in teams with strong communication, leadership, adaptability, and professional ethics while addressing privacy and AI bias issues.
PO5	Conduct research, case studies, and innovative projects while applying evaluation metrics, validation techniques, and industry best practices.

PROGRAM SPECIFIC OUTCOMES (PSOS)

PROGRAM SPECIFIC OUTCOMES (PSOS)	
After successful completion of the program, students will be able to:	
PSO1	Apply Big Data and Data Science techniques to analyze structured and unstructured data effectively.
PSO2	Select appropriate tools, platforms, and algorithms for data collection, processing, modeling, and deployment.
PSO3	Use visualization, storytelling, and reporting techniques to communicate analytical insights clearly.
PSO4	Understand end-to-end data pipelines, distributed systems, cloud architectures, and lifecycle management in analytics projects.
PSO5	Address industry challenges and societal impacts of data-driven systems through innovative and research-oriented solutions.

SEMESTER - I

SEMESTER I

DIPLOMA IN BIG DATA ANALYTICS

Course code		**TITLE OF THE COURSE**	**L**	**T**	**P**	**C**
Core paper		**INTRODUCTION TO BIG DATA**	**4**	**-**	**-**	**4**

Course Objectives

The main objectives of this course are:

1. Understand the concept of Big Data
2. Gain knowledge of Big Data technologies and platforms
3. Learn Big Data storage and processing techniques,
4. Understand the fundamentals of Data Analytics.
5. Explore real-world Big Data applications

Expected Course Outcomes:

On the successful completion of the course, student will be able to

1	Explain the definition, characteristics (5Vs), types, and sources of Big Data and differentiate Big Data from traditional data.	K1,K2
2	Describe the evolution of Big Data technologies, distributed systems, Hadoop architecture, and Hadoop ecosystem tools.	K2
3	Demonstrate understanding of Big Data storage (HDFS) and processing concepts including blocks, replication, Mapper, Reducer, Shuffle & Sort.	K2
4	Explain the Data Analytics lifecycle, types of analytics, preprocessing techniques, and basics of data visualization and Python for analytics.	K1,K2
5	Identify major applications of Big Data, discuss security/privacy issues, challenges, and future industry trends.	K2

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

UNIT 1	**Introduction to Big Data**	**04 hours**

Definition of Big Data-Characteristics of Big Data (5Vs – Volume, Velocity, Variety, Veracity, Value)-Traditional Data vs Big Data-Types of Data: Structured, Semi-Structured, Unstructured-Sources of Big Data (social media, Sensors, IoT, Transactions, Logs)-Importance & Need for Big Data.

UNIT 2	**Big Data Technologies & Platforms**	**04 hours**

History and Evolution of Big Data Technologies- Distributed Systems Basics- Introduction to Hadoop- Hadoop Components: HDFS, MapReduce- Overview of Hadoop Ecosystem (Hive, Pig, HBase, Sqoop, Flume)- Limitations of Traditional Databases in Big Data

UNIT 3	**Big Data Storage & Processing**	**04 hours**

Data Storage Concepts- Data Warehousing Basics- HDFS Architecture-Name Node, Data Node-Data Blocks, Replication- Concept of Data Processing- MapReduce Programming Model-Mapper, Reducer, Shuffle & Sort- Batch vs Real-Time Processing

UNIT 4	**Introduction to Data Analytics**	**04 hours**

Data Analytics Lifecycle-Types of Analytics: Descriptive, Diagnostic, Predictive, Prescriptive- Data Preprocessing: Cleaning, Transformation, Normalization- Basics of Data Visualization- Introduction to Python for Data Analytics (Concept Level)

UNIT 5	**Applications, Challenges & Future of Big Data**	**04 hours**

Applications of Big Data: Healthcare, Education, Banking & Finance, Social Media, E-Commerce, Government & Smart Cities- Security & Privacy Issues in Big Data- Challenges in Big Data

Implementation. Future Trends: Machine Learning, Artificial Intelligence, Cloud for Big Data, IoT & Big Data Integration.		
	Total Lecture hours	**20 hours**
Text Book(s)		
1	Big Data: Principles and Best Practices of Scalable Realtime Data Systems, Nathan Marz & James Warren, Manning Publications,2015	
2	Hadoop: The Definitive Guide, Tom White, O'Reilly Media,2015 (4th Edition)	
Reference Books		
1.Big Data and Analytics, Seema Acharya & Subhashini Chellappan, Wiley India,2015 2.Data Science and Big Data Analytics, EMC Education Services,Wiley,2015 3.Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman Cambridge University Press, Year of Publication: 2020 (3rd Edition)		
Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]		
Course Designed By: Dr.P.VIJAYAKUMAR		
Web link: https://www.ibm.com/topics/big-data https://www.oracle.com/big-data/what-is-big-data https://hadoop.apache.org/ https://pig.apache.org/		

Mapping with programme outcomes:

COS	PO1	PO2	PO3	PO4	PO5
CO1	S	M	S	L	L
CO2	S	S	M	L	L
CO3	S	S	S	M	L
CO4	M	S	M	L	L
CO5	M	M	L	L	S

S- Strong; M-Medium; L-Low

Course code		TITLE OF THE COURSE	L	T	P	C
Core paper		BASICS OF DATA ANALYTICS	4	-	-	4

Course Objectives

- Understand the **importance of data analytics** in the digital era and its applications across industries.
- Identify various **types and sources of data** and perform proper data collection methods.
- Apply **data cleaning, transformation, and normalization techniques**.
- Perform **Exploratory Data Analysis (EDA)** using descriptive statistics and visualization tools.
- Understand and apply **basic statistical analysis techniques** including hypothesis testing.
- Interpret analytical results for informed decision-making.
- Gain foundational knowledge of **regression, multivariate analysis, time series, Bayesian thinking, SVM, and rule-based analysis**.
- Develop analytical thinking skills for solving real-world problems.

Expected Course Outcomes:

On the successful completion of the course, student will be able to

1	Explain the fundamentals of data analytics and its role in business, healthcare, education, and governance.	K2
2	Collect, clean, and prepare data from multiple sources for analysis.	K3
3	Perform exploratory data analysis using descriptive statistics and basic visualization techniques.	K3
4	Apply fundamental statistical methods including hypothesis testing (Chi-square, t-test, ANOVA, correlation).	K4
5	Demonstrate understanding of basic data analysis methods such as regression, multivariate analysis, time series, and rule-based approaches.	K2/K3

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

UNIT 1	Introduction to Data Analytics	04 hours

Need for data analytics in the digital era- Types of data- Data analytics vs reporting-Stages in data analytics process-Data Analytics Types- Applications of data analytics in business, healthcare, education, and governance.

UNIT 2	Data Collection and Data Preparation	04 hours

Sources of data: databases, files, web data - Methods of data collection- Data quality issues- Data cleaning concepts- Data transformation and simple normalization-Importance of data preparation in analytics.

UNIT 3	Exploratory Data Analysis and Statistics	04 hours

Introduction to Exploratory Data Analysis (EDA)- Descriptive Statistics-Mean, median, mode, Range and standard deviation - Basic data summarization- Data visualization concepts- Bar chart, pie chart, line chart, histogram- Interpretation of statistical and graphical results.

UNIT 4	Fundamentals of Statistical Analysis	04 hours

Basic analysis techniques- Statistical hypothesis generation and testing- Chi-Square test- T-Test, Analysis of variance- Correlation analysis- Basic interpretation of analytical results.

UNIT 5	Introduction to Data Analysis Methods	04 hours

Introduction to data analysis techniques- Basic idea of regression and its uses- Introduction to multivariate data analysis- Concept of probabilistic (Bayesian) thinking- Overview of support vector machines - Basics of time series data and simple trends- Introduction to rule-based analysis.		
	Total Lecture hours	**20 hours**
Text Book(s)		
1	**Data Analytics and Decision Making,**Ali AbdulHussein, University of Windsor **2022** (Copyright © 2022)	
2	**Data Analytics Made Accessible,** Dr. Anil Maheshwari, McGraw-Hill Education (India) / Amazon Publishing,**2017** (original edition)	
Reference Books		
1.Practical Data Analysis by Hector Cuesta, Christian Hofmann, and Benjamin Cook, Publishing,2013		
Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]		
Course Designed By: Dr.P.VIJAYAKUMAR		
Web link: https://nptel.ac.in/courses https://www.ibm.com/topics/data-analytics https://grow.google/certificates/data-analytics https://www.khanacademy.org/math/statistics-probability https://www.coursera.org/browse/data-science/data-analysis		

Mapping with programme outcomes:

COS	PO1	PO2	PO3	PO4	PO5
CO1	S	M	L	L	L
CO2	S	S	M	M	M
CO3	S	S	M	S	M
CO4	S	S	M	S	M
CO5	S	S	S	M	S

S- Strong; M-Medium; L-Low

Course code		TITLE OF THE COURSE	L	T	P	C
Core paper		DATAPROCESSING TOOLS	4	-	-	4

Course Objectives

- To understand the fundamentals of Big Data and Hadoop ecosystem.
- To learn Pig Latin scripting for processing large datasets.
- To perform data transformations using Apache Pig.
- To understand Hive data warehouse concepts and HiveQL queries.
- To analyze and process structured data using Hive for business applications.

Expected Course Outcomes:

On the successful completion of the course, student will be able to

1	Explain Big Data concepts and Hadoop architecture	K2
2	Write and execute basic Pig Latin scripts for data processing.	K3
3	Apply advanced Pig operations including joins and UDFs.	K3/K4
4	Create and manage databases and tables using HiveQL.	K3
5	Analyze large datasets using Hive queries and compare Pig and Hive tools.	K4

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

UNIT 1	Introduction to Big Data & Hadoop	04 hours

Need for Big Data- Characteristics of Big Data (5Vs)- Introduction to Hadoop Architecture- HDFS (Hadoop Distributed File System)-MapReduce Overview-Hadoop Ecosystem Components

UNIT 2	Introduction to Apache Pig	04 hours

Introduction to Apache Software Foundation-Overview of Apache Pig-Pig Architecture- Pig Latin Basics- Data Types in Pig-Loading and Storing Data-Simple Operators (FILTER, FOREACH, GROUP, JOIN)

UNIT 3	Advanced Pig Concepts	04 hours

Relational Operators-User Defined Functions (UDFs)-Debugging Pig Scripts-Performance Tuning Running Pig in Local and MapReduce Mode

UNIT 4	Introduction to Apache Hive	04 hours

Overview of Apache Hive-Hive Architecture-Hive Data Types- HiveQL Basics- Creating Databases and Tables- Managed vs External Tables-Partitioning and Bucketing

UNIT 5	Advanced Hive & Integration	04 hours

Hive Queries (SELECT, WHERE, GROUP BY, JOIN)-Views and Indexing-Hive Functions-Integration of Hive with Hadoop- Comparison: Pig vs Hive- Practical Applications in Industry

	Total Lecture hours	20 hours

Text Book(s)

1. Alan Gates (2011) – Programming Pig, O'Reilly Media.
2. Edward Capriolo, Dean Wampler & Jason Rutherglen (2012) – Programming Hive, O'Reilly Media.

Reference Books

1.Tom White (2015, 4th Edition) – Hadoop: The Definitive Guide, O'Reilly Media. 2.Seema Acharya & Subhashini Chellappan (2015) – Big Data and Analytics, Wiley Publications.
Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]
Course Designed By: Dr.P.VIJAYAKUMAR
Web link: https://www.oreilly.com/library/view/hadoop-the-definitive/9781491901687/?utm_source=chatgpt.com https://www.oreilly.com/library/view/programming-hive/9781449326944/?utm_source=chatgpt.com https://www.oreilly.com/library/view/programming-pig/9781449317881/?utm_source=chatgpt.com

Mapping with programme outcomes:

COS	PO1	PO2	PO3	PO4	PO5
CO1	S	M	L	L	M
CO2	S	S	M	S	S
CO3	S	S	M	S	S
CO4	S	S	M	S	S
CO5	S	S	S	M	S

S- Strong; M-Medium; L-Low

Course code		TITLE OF THE COURSE	L	T	P	C
Core paper		PROGRAMMING IN PYTHON	4	-	-	4

Course Objectives

1. Introduce Python programming fundamentals required for data analytics.
2. Develop skills in handling, cleaning, and processing data using Python libraries.
3. Teach data visualization concepts using Python tools.
4. Enable students to perform basic statistical analysis and exploratory data analysis (EDA).
5. Provide hands-on practice in applying Python for real-world data analytics tasks

Expected Course Outcomes:

On the successful completion of the course, student will be able to

1	Understand Python fundamentals and write basic programs.	K1,K2
2	Apply Python data structures for data manipulation.	K2
3	Use NumPy for numerical and scientific data operations.	K2,K3
4	Handle real datasets using Pandas for analysis.	K2,K3
5	Create effective data visualizations using Matplotlib & Seaborn.	K3

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

UNIT 1	**Introduction to Programming and Python Basics**	**04 hours**

Concept of programming and problem-solving; role of programming in daily life; introduction to Python and its applications in education, business, and data handling; features of Python; installing and running Python; understanding the Python interpreter; basic syntax rules; keywords and identifiers; writing and executing simple Python programs.

UNIT 2	**Data Types, Variables, and Operators**	**04 hours**

Understanding variables and assignment; basic data types (integers, floating-point numbers, strings, and Boolean values); type conversion; input and output statements; arithmetic, relational, logical, and assignment operators; precedence of operators.

UNIT 3	**Control Structures and Decision Making**	**04 hours**

Concept of flow control; conditional statements (if, if-else, nested if, elif); looping structures (while loop and for loop); use of break and continue; indentation and logical flow in Python.

UNIT 4	**Data Structures, Functions, and File Handling (Basic Level)**	**04 hours**

Introduction to basic data structures: lists, tuples, dictionaries, and sets; basic operations on lists and dictionaries; accessing and updating data; concept of functions; defining and calling functions; parameters and return values; importance of modular programming; introduction to file handling; reading from and writing to text files; simple case-oriented examples such as student lists, product-price records, reusable utility functions, and basic record storage.

UNIT 5	**Data Science Concepts using Python and Excel**	**04 hours**

Overview of data science and its applications; types of data (numerical and categorical); basic data analysis concepts; introduction to spreadsheets for data handling; understanding rows, columns, and datasets in Excel; basic Excel operations for data analysis (sorting, filtering, simple formulas, charts); using Python for basic data analysis tasks such as reading data, computing totals, averages, and simple summaries; comparison of Python and Excel for data handling; simple case studies such as student performance analysis and basic sales data analysis.

	Total Lecture hours	**20 hours**

Text Book(s)

1. Python Programming - Using Problem Solving Approach 2e, Reema Thareja, 2017 (Oxford University Press).
Reference Books
1. Introduction to Computing Using Python 2e, Ljubomir Perkovic, John Wiley & Sons, 2015
Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]
https://www.w3schools.com/python https://pandas.pydata.org/docs/ https://support.microsoft.com/excel https://nptel.ac.in/courses
Course Designed By: Dr.P.VIJAYAKUMAR

Mapping with programme outcomes:

COS	PO1	PO2	PO3	PO4	PO5
CO1	S	M	L	M	M
CO2	S	S	M	S	M
CO3	S	S	S	S	M
CO4	S	S	S	S	S
CO5	S	S	M	S	S

S- Strong; M-Medium; L-Low

PYTHON PROGRAMMING LAB

List of Programs

1. Program to find the sum of two numbers Input two numbers and print the sum.
2. Program to check whether a number is even or odd Use conditional statements.
3. Program to find the largest of three numbers Use if–elif–else.
4. Program to generate Fibonacci series Display first n Fibonacci numbers.
5. Program to check whether a string is palindrome Example: madam → palindrome.
6. Program to count vowels and consonants in a string
7. Program to find factorial of a number using loop
8. Program to calculate simple interest and compound interest
9. Program to read N numbers and find the largest element in a list
10. Program to sort a list in ascending order Use built-in sort() or custom loop.
11. Program to create a dictionary and perform CRUD operations Add, update, delete, search keys.
12. Program to read a CSV file and display its contents using Python (Pandas) Useful for data analytics lab.

SEMESTER - II

Course code		TITLE OF THE COURSE	L	T	P	C
Core paper		DATABASE TECHNOLOGIES FOR BIG DATA	4	-	-	4

Course Objectives

Provide basic knowledge of database systems and Big Data concepts.
Introduce relational databases and SQL fundamentals.
Familiarize students with NoSQL and distributed database systems.
Develop understanding of cloud database technologies.
Create awareness of data security, scalability, and modern database trends.

Expected Course Outcomes:

On the successful completion of the course, student will be able to

1	Explain fundamentals of databases and Big Data systems	K2
2	Describe relational databases and basic SQL concepts	K2
3	Explain NoSQL databases and distributed database concepts	K2
4	Describe cloud databases and scalability concepts	K2
5	Explain data security, privacy, and future trends in Big Data databases	K2

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

UNIT 1	**INTRODUCTION TO DATABASES AND BIG DATA**	**04 hours**

Basics of Database Systems- Traditional Databases vs Big Data Systems- Relational Database Concepts (Tables, Keys, Constraints)- Overview of Big Data (5Vs)- Limitations of Traditional RDBMS in Big Data- Applications of Big Data Databases

UNIT 2	**RELATIONAL DATABASES AND SQL**	**04 hours**

Structure of RDBMS-SQL Basics: SELECT, INSERT, UPDATE, DELETE (Conceptual)-Joins and Aggregation (Concept only)-Transactions and ACID properties-Indexing and Performance Basics Introduction to Data Warehousing

UNIT 3	**NoSQL DATABASES**	**04 hours**

Introduction to NoSQL-Need for NoSQL in Big Data-Types of NoSQL Databases:Key-Value Stores Document Databases, Column-Family Databases,Graph Databases. CAP Theorem (Conceptual)-Comparison of SQL and NoSQL

UNIT 4	**DISTRIBUTED DATABASE SYSTEMS**	**04 hours**

Concept of Distributed Databases-Data Partitioning and Replication- Hadoop Distributed File System (HDFS – overview) -Introduction to Apache HBase and MongoDB (conceptual)- Scalability and Fault Tolerance

UNIT 5	**CLOUD DATABASES AND DATA MANAGEMENT**	**04 hours**

Introduction to Cloud Databases-Database as a Service (DBaaS)-Overview of AWS RDS, Google Cloud SQL, Azure SQL (conceptual)-Data Security and Privacy in Big Data-Data Backup and Recovery-Future Trends in Big Data Database Technologies

	Total Lecture hours	**20 hours**

Text Book(s)

1 2	**Database System Concepts,** Abraham Silberschatz, Henry F. Korth, S. Sudarshan, McGraw-Hill Education,7th Edition,2019

3	**NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence,**Pramod J. Sadalage & Martin Fowler,Addison-Wesley Professional,2012 **Hadoop: The Definitive Guide,**Tom White,O'Reilly Media,4th Edition,2015

Reference Books

1. **Designing Data-Intensive Applications,**Martin Kleppmann,O'Reilly Media,2017
2. **MongoDB: The Definitive Guide,**Shannon Bradshaw, Kristina Chodorow,O'Reilly Media 3rd Edition,2019
3. **Cloud Computing: Concepts, Technology & Architecture,**Thomas Erl,Pearson,2013

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

https://docs.cloud.google.com/sql/docs?utm_source=chatgpt.com
https://spark.apache.org/documentation.html?utm_source=chatgpt.com
https://docs.cloud.google.com/sql/docs/sqlserver?utm_source=chatgpt.com
https://github.com/mongodb/docs?utm_source=chatgpt.com
https://spark.apache.org/?utm_source=chatgpt.com

Course Designed By: Dr.P.VIJAYAKUMAR

Mapping with programme outcomes:

COS	PO1	PO2	PO3	PO4	PO5
CO1	S	L	L	L	L
CO2	S	M	M	L	L
CO3	S	M	S	M	L
CO4	S	M	M	M	M
CO5	M	L	L	L	S

S- Strong; M-Medium; L-Low

Course code		TITLE OF THE COURSE	L	T	P	C
Core paper		DATA ENGINEERING	4	-	-	4

Course Objectives

Provide fundamental knowledge of Data Engineering concepts, roles, and data lifecycle.
Introduce data storage systems including relational, NoSQL, data warehouses, and cloud storage.
Develop understanding of ETL processes, data preprocessing, and data pipelines.
Familiarize students with Big Data technologies and cloud-based data platforms.
Create awareness about data governance, privacy, security, and real-world applications.

Expected Course Outcomes:

On the successful completion of the course, student will be able to

1	Explain the fundamentals of Data Engineering and data lifecycle concepts.	K2
2	Describe data storage systems including SQL, NoSQL, data warehouses, and cloud storage.	K2
3	Explain ETL processes, data cleaning, and basic data pipeline workflows	K2
4	Describe Big Data technologies, distributed systems, and cloud platforms.	K2
5	Explain data governance, privacy principles, and applications of Data Engineering in various domains.	K2

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

UNIT 1	INTRODUCTION TO DATA ENGINEERING	04 hours

Meaning and importance of Data Engineering-Role of Data Engineer in data-driven organizations Difference between Data Science, Data Analytics, and Data Engineering-Types of data: structured, semi-structured, unstructured-Overview of data lifecycle=Introduction to databases and data storage systems-Applications of Data Engineering in business, healthcare, banking, education, and governance

UNIT 2	DATA COLLECTION AND DATA STORAGE	04 hours

Data sources: files, databases, APIs, sensors, web data-Introduction to relational databases (concept of tables, rows, columns, keys)-Basics of SQL (SELECT, INSERT, UPDATE, DELETE – concept only)-Introduction to NoSQL databases (MongoDB concept)-Data warehouses and data lakes (conceptual understanding)-Cloud storage basics (AWS S3, Google Cloud Storage – overview)

UNIT 3	DATA PROCESSING AND ETL	04 hours

Concept of ETL (Extract, Transform, Load)- Data cleaning and preprocessing basics- Handling missing data and data quality issues-Batch processing vs real-time processing- Introduction to Big Data tools (Hadoop, Spark – overview)-Data pipeline concept and workflow

UNIT 4	BIG DATA AND CLOUD DATA ENGINEERING	04 hours

Introduction to Big Data (5Vs concept)- Distributed computing basics- Overview of Hadoop ecosystem- Introduction to Apache Spark-Cloud-based data platforms (AWS, Azure, GCP – overview)-Scalability and data security basics

UNIT 5	DATA PIPELINES, GOVERNANCE AND APPLICATIONS	04 hours

Designing simple data pipelines- Data governance and data privacy principles- Introduction to data security and compliance-Monitoring and maintaining data systems- Case studies of data engineering in fintech, healthcare, e-commerce, IoT- Future trends in Data Engineering

	Total Lecture hours	20 hours

Text Book(s)	
1	**Fundamentals of Data Engineering,** Joe Reis & Matt Housley, O'Reilly Media, 2022 (First Edition)
2	**Designing Data-Intensive Applications,** Martin Kleppmann, O'Reilly Media, 2017
3	**Data Engineering with Python,** Paul Crickard, Packt Publishing,2020

Reference Books

1. Hadoop: The Definitive Guide, Tom White,O'Reilly Media, 2015
2. Spark: The Definitive Guide,Bill Chambers & Matei Zaharia,O'Reilly Media, 2018
3. Cloud Computing: Concepts, Technology & Architecture,Thomas Erl, Pearson Publication 2013 (1st Edition)

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

https://www.databricks.com/glossary/data-engineering
https://www.w3schools.com/sql
https://hadoop.apache.org
https://spark.apache.org/docs/latest
https://aws.amazon.com/big-data/datalakes-and-analytics

Course Designed By: Dr.P.VIJAYAKUMAR

Mapping with programme outcomes:

COS	PO1	PO2	PO3	PO4	PO5
CO1	**S**	**L**	**L**	**L**	**L**
CO2	**S**	**S**	**L**	**L**	**L**
CO3	**S**	**S**	**M**	**L**	**L**
CO4	**S**	**S**	**M**	**L**	**L**
CO5	**M**	**S**	**M**	**S**	**S**

S- Strong; M-Medium; L-Low

Course code		TITLE OF THE COURSE	L	T	P	C
Core paper		CLOUD COMPUTING FOR BIG DATA	4	-	-	4

Course Objectives

1. To introduce the fundamentals of cloud computing concepts and service models.
2. To understand virtualization, cloud deployment architectures, and cloud infrastructure.
3. To learn Big Data concepts, Hadoop ecosystem components, and cloud storage.
4. To explore cloud platforms and services for Big Data processing.
5. To equip students with knowledge of cloud security, privacy, and cloud management tools.

Expected Course Outcomes:

On the successful completion of the course, student will be able to

1	Explain cloud computing concepts, characteristics, and service models.	K2
2	Demonstrate virtualization techniques and cloud deployment architectures.	K2
3	Describe Big Data characteristics and explain Hadoop ecosystem components.	K2
4	Apply cloud storage and cloud-based Big Data processing tools.	K3
5	Discuss cloud security, privacy issues, and cloud resource management.	K2

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

UNIT 1	INTRODUCTION TO CLOUD COMPUTING	04 hours

Definition and Evolution of Cloud Computing-Characteristics of Cloud (On-demand service, Resource pooling, Elasticity)-Cloud Service Models:IaaS (Infrastructure as a Service)-PaaS (Platform as a Service)-SaaS (Software as a Service)-Cloud Deployment Models:Public loud Private Cloud-Hybrid Cloud-Community Cloud-Benefits and Challenges of Cloud-Computing-Cloud Architecture Overview

UNIT 2	VIRTUALIZATION & CLOUD INFRASTRUCTURE	04 hours

Virtualization concepts-Types of Virtualization:Hardware Virtualization-OS-Level-Virtualization-Server & Storage Virtualization-Hypervisors: Type I and Type II-Virtual Machine Lifecycle-Cloud Infrastructure Components:Compute-Storage-Network-Containers and Docker Basics

UNIT 3	INTRODUCTION TO BIG DATA & HADOOP	04 hours

What is Big Data?-Characteristics of Big Data (3Vs / 5Vs)-Big Data Challenges-Hadoop Architecture-HDFS – Hadoop Distributed File System-MapReduce Programming Model-YARN Architecture-Hadoop Use Cases

UNIT 4	CLOUD PLATFORMS FOR BIG DATA	04 hours

Cloud Storage Systems: Amazon S3-Google Cloud Storage-Azure Blob Storage-Big Data Services on Cloud:Amazon EMR-Google Dataproc-Azure HDInsight-NoSQL Databases on Cloud: MongoDB Atlas- Cassandra-HBase-Distributed Computing Concepts-Cloud-based Data Analytics Tools

UNIT 5	CLOUD SECURITY, MANAGEMENT & APPLICATIONS	04 hours

Cloud Security Fundamentals-Security Threats in Cloud-Identity and Access Management (IAM)-Data Protection & Encryption in Cloud-Service Level Agreements (SLAs)-Cloud Monitoring & Management Tools-CloudWatch-Stackdriver-Azure Monitor-Applications of Cloud for Big Data-IoT Analytics-Real-time data processing-Machine Learning on Cloud

	Total Lecture hours	20 hours

Text Book(s)

1	Cloud Computing: Concepts, Technology & Architecture Authors: Thomas Erl, Zaigham Mahmood, Ricardo Puttini,Pearson,2013
2	Big Data: Principles and Best Practices of Scalable Real-Time Data Systems Authors: Nathan Marz, James Warren,Manning Publication,2015
3	Hadoop: The Definitive Guide Author: Tom White,O'reilly media,2015

Reference Books

1. Cloud Computing: A Practical Approach Authors: Anthony T. Velte, Toby J. Velte, Robert Elsenpeter,McGraw-Hill 2009
2. Mastering Cloud Computing Authors: Rajkumar Buyya, Christian Vecchiola, S. Thamarai Selvi,Mc Graw-Hill,2013
3. Big Data Fundamentals: Concepts, Drivers & Techniques Authors: Thomas Erl, Wajid Khattak, Paul Buhler,Pearson, 2015
4. NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence Authors: Pramod J. Sadalage, Martin Fowler, Addison-Wesley Professional,2012

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

https://www.nist.gov
https://docs.aws.amazon.com
https://learn.microsoft.com/azure
https://cloud.google.com/docs
https://hadoop.apache.org
https://www.mongodb.com/docs/atlas
https://cassandra.apache.org
https://hbase.apache.org

Course Designed By: Dr.P.VIJAYAKUMAR

Mapping with programme outcomes:

COS	PO1	PO2	PO3	PO4	PO5
CO1	S	M	L	**L**	**L**
CO2	**S**	**S**	**M**	**L**	**L**
CO3	**S**	**S**	**M**	**M**	**L**
CO4	**M**	**S**	**S**	**M**	**L**
CO5	**M**	**S**	**M**	**M**	**L**

S- Strong; M-Medium; L-Low

Course code		TITLE OF THE COURSE	L	T	P	C
Core paper		FUNDAMENTALS OF MACHNE LEARNING	4	-	-	4

Course Objectives

- Provide foundational understanding of Machine Learning concepts, evolution, and real-world relevance.
- Develop conceptual knowledge about data types, preprocessing, and the learning process in ML systems.
- Introduce different types of machine learning techniques and their practical applications.
- Familiarize students with fundamental ML models and their selection based on problem context.
- Enable students to understand ML project lifecycle, evaluation techniques, and societal impact of ML

Expected Course Outcomes:

On the successful completion of the course, student will be able to

1	Explain fundamentals, evolution, applications, and ethics of ML	K2
2	Understand and describe role of data and preprocessing	K2
3	Differentiate types of machine learning approaches	K2
4	Describe basic ML models and their selection	K2
5	Explain ML lifecycle, evaluation concepts, and case studies	K2

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

UNIT 1	INTRODUCTION TO MACHINE LEARNING	04 hours

Meaning and importance of Machine Learning (ML); evolution of ML; difference between traditional programming and machine learning; artificial intelligence vs machine learning vs deep learning (conceptual); real-life applications of ML in education, healthcare, banking, agriculture, marketing, and governance; types of ML problems; ethical considerations and societal impact of ML; limitations of ML systems.

UNIT 2	DATA AND LEARNING PROCESS	04 hours

Role of data in machine learning; types of data (structured, semi-structured, unstructured); features and labels; training data, testing data, and validation data (conceptual understanding); data quality issues; introduction to datasets and examples from daily life; basic idea of data preprocessing; importance of domain knowledge in learning systems.

UNIT 3	TYPES OF MACHINE LEARNING	04 hours

Supervised learning – concept, workflow, and examples (classification and regression explained using real-world cases); unsupervised learning – concept and use cases (clustering and association explained intuitively); semi-supervised learning and reinforcement learning – overview and practical relevance; comparison of learning types; application-oriented illustrations.

UNIT 4	MACHINE LEARNING FUNDAMENTALS	04 hours

Overview of simple ML models at a conceptual level: linear regression, logistic regression, decision trees, k-nearest neighbours, naïve Bayes, and clustering models; understanding model behaviour using examples and diagrams; strengths and weaknesses of each model; selection of models based on problem context; interpretability and explainability of ML models.

UNIT 5	MACHINE LEARNING APPLICATIONS AND EVALUATION	04 hours

Steps in a machine learning project lifecycle; concept of model evaluation (accuracy, error, and performance measures explained intuitively); bias, variance, and overfitting (conceptual); real-world ML case studies in business, social sciences, healthcare, and smart systems; future trends in machine learning.	
	Total Lecture hours **20 hours**

Text Book(s)

1	Machine Learning for Absolute Beginners, Oliver Theobald, Scatterplot Press, 2017

Reference Books

1. Machine Learning Basics, Tom M. Mitchell, McGraw-Hill Education, 1997.

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

https://www.ibm.com/topics/machine-learning
https://developers.google.com/machine-learning/crash-course
https://scikit-learn.org/stable/tutorial/index.html
https://towardsdatascience.com/model-evaluation
https://www.unesco.org/en/artificial-intelligence

Course Designed By: Dr.P.VIJAYAKUMAR

Mapping with programme outcomes:

COS	PO1	PO2	PO3	PO4	PO5
CO1	S	**L**	L	M	**L**
CO2	**S**	**M**	**L**	**L**	**L**
CO3	**S**	**M**	**L**	**L**	**L**
CO4	**S**	**M**	**M**	**L**	**L**
CO5	**M**	**M**	**M**	**M**	**S**

S- Strong; M-Medium; L-Low

Machine Learning Lab Program

1. Simple Linear Regression
2. Logistic Regression
3. Decision Tree Classifier
4. K-Nearest Neighbours
5. Naïve Bayes Classifier
6. K-Means Clustering
7. Train-Test Split Example
8. Model Accuracy Calculation
9. Bias-Variance Demonstration
10. Simple ML Project Workflow