30.1.5

UNIVERSITY SCHOOL
OF
INFORMATION AND COMMUNICATION TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

PROGRAMME STRUCTURE

B.TECH. COMPUTER SCIENCE AND ENGINEERING
SPECIALIZATION IN DATA SCIENCE
2021-2025

30.1.5
08.09.23

GAUTAM BUDDHA UNIVERSITY
GAUTAM BUDH NAGAR, GREATER NOIDA, UP, INDIA

School of ICT
Gautam Buddha University
Greater Noida, (U.P.)

SEMESTER I

S.No.	Course Code	Course Name	L	T	P	Credits	Types
1	CS101	Fundamentals of Computer Programming	3	1	0	4	CC1 / FC
2	CD101	Fundamentals of Data Science and MS Excel	2	0	0	2	CC2 / FC
3	MA101	Engineering Mathematics-I	3	1	0	4	GE1
4	PH102	Engineering Physics	3	1	0	4	GE2
5	EC101	Basic Electronics Engineering	3	1	0	4	GE3 /FC
6	EN101	English Proficiency	2	0	0	2	OE1 / AECC
7	CE103	Engineering Graphics Lab	1	0	2	2	GE-L1
8	PH104	Engineering Physics Lab	0	0	2	1	GE-L2
9	CS181	Computer Programming Lab	0	0	2	1	CC-L1 / SEC
10	EC181	Basic Electronics Engineering Lab	0	0	2	1	GE-L3
11	GP	General Proficiency	Non Credit				
		Total Hours and Credits	**17**	**4**	**8**	**25**	

SEMESTER II

S.No.	Course Code	Course Name	L	T	P	Credits	Types
1	CD102	Introduction to Python	2	0	0	2	CC3 / FC
2	CD104	Computer Organistaion and Architecure	3	0	0	3	CC4 / SEC
3	MA102	Engineering Mathematics-II	3	1	0	4	GE4
4	EE102	Basic Electrical Engineering	3	1	0	4	GE5
5	ME101	Engineering Mechanics	3	1	0	4	GE6
6	ES101	Environmental Studies	3	1	0	4	OE2 / AECC
7	CD182	Python Programming Lab	0	0	2	1	CC-L2 / SEC
8	EE104	Basic Electrical Engineering Lab	0	0	2	1	GE-L4
9	ME102	Workshop Practice	1	0	2	2	GE-L5
10	GP	General Proficiency	Non Credit				
		Total Hours and Credits	**16**	**4**	**6**	**25**	

08/09/23

School of ICT
Gautam Buddha University
Greater Noida, (U.P.)

SEMESTER III

S.No.	Course Code	Course Name	L	T	P	Credits	Types
1	CD201	Internet Technology	3	0	0	3	CC5 / SEC
2	CD203	Operating Systems	3	0	0	3	CC6
3	CD205	Data Structure & Algorithms	3	0	0	3	CC7 / SEC
4	CD207	Optimization Problem for Data Science	3	0	0	3	CC8
5	CD209	Introduction to R Programming	3	0	0	3	CC9
6	MA201	Engineering Mathematics-III	3	1	0	4	GE7
7	CD281	R Programming Lab	0	0	3	2	CC-L3 / SEC
8	CD283	Data Structure & Algorithms Lab	0	0	3	2	CC-L4 / SEC
9	CD285	Internet Technology Lab	0	0	3	2	CC-L5 / SEC
10	GP	General Proficiency	Non Credit				
		Total Hours and Credits	**18**	**1**	**9**	**25**	

SEMESTER IV

S.No.	Course Code	Course Name	L	T	P	Credits	Types
1	CD202	Software Engineering	3	0	0	3	CC10
2	CD204	Database Management System	3	0	0	3	CC11 / SEC
3	CD206	Java Programming	3	0	0	3	CC12
4	CD208	Artificial Intelligence	3	0	0	3	CC13
5	CD210	Theory of Automata	3	0	0	3	CC14
6	CD212	Data Handling and Visualization	3	1	0	4	CC15/SEC
7	CD282	Database Management System Lab	0	0	3	2	CC-L6 / SEC
8	CD284	Java Programming Lab	0	0	3	2	CC-L7 / SEC
9	CD286	Tableau Lab	0	0	3	2	CC-L8 / SEC
10	GP	General Proficiency	Non Credit				
		Total Hours and Credits	**18**	**1**	**9**	**25**	

8/9/23

School of ICT
Gautam Buddha University
Greater Noida, (U.P.)

SEMESTER V

S.No.	Course Code	Course Name	L	T	P	Credits	Types
1	CD301	Compiler Design	3	0	0	3	CC16 / AECC
2	CD303	Soft Computing Techniques	3	0	0	3	CC17
3	CD305	Analysis and Design of Algorithms	3	0	0	3	CC18
4	CD307	Big Data Analytics	3	0	0	3	CC19
5	CD309	Machine Learning	3	1	0	4	CC20 / SEC
6		Elective 1	3	0	0	3	E1 / DSE
7	CD381	Analysis and Design of Algorithms Lab	0	0	3	2	CC-L9 / SEC
8	CD383	Big Data Analytics Lab	0	0	3	2	CC-L10 / SEC
9	CD385	Machine Learning Lab using Python	0	0	3	2	CC-L11 / SEC
10	GP	General Proficiency	Non Credit				
		Total Hours and Credits	**18**	**1**	**9**	**25**	

SEMESTER VI

S.No.	Course Code	Course Name	L	T	P	Credits	Types
1	CD302	Web Development using PHP	3	0	0	3	CC21
2	CD304	Introduction to Statistical Learning	3	0	0	3	CC22
3	CD306	Operation Research in Data Science	3	1	0	4	CC23
4	CD308	Cloud Computing	3	0	0	3	CC24
5	CD310	Data Privacy and Database Security	3	0	0	3	CC25 / SEC
6		Elective 2	3	0	0	3	E2 / DSE
7	CD382	Web Development using PHP Lab	0	0	3	2	CC-L12
8	CD384	Statistical Learning Lab	0	0	3	2	CC-L13
9	CD386	Data Privacy and Database Security Lab	0	0	3	2	CC-L14
10	GP	General Proficiency	Non Credit				
		Total Hours and Credits	**18**	**1**	**9**	**25**	

Industrial Training will be done by candidate individually after third year during the summer break and it will be of minimum 4 weeks. It will be evaluated as per University Examination in VII semester.

08/04/23

School of ICT
Gautam Buddha University
Greater Noida, (U.P.)

SEMESTER VII

S.No.	Course Code	Course Name	L	T	P	Credits	Types
1	CD401	Parallel processing and CUDA programming	3	1	0	4	CC26
2	CD403	Cryptography and Network Security	3	0	0	3	CC27
3	CD405	Data Analytics using R	2	0	0	2	CC28 / SEC
4		Elective 3	3	0	0	3	E3 / DSE
5		Elective 4	3	0	0	3	E4 / DSE
6	CD481	Data Analytics using R Lab	0	0	3	2	CC-L15
7	CD491	Minor Project	0	0	10	5	MP1 / E
8	CD493	Industrial Traning	0	0	6	3	IT1 / E
9	GP	General Proficiency	Non Credit				
		Total Hours and Credits	**14**	**1**	**19**	**25**	

SEMESTER VIII

S.No.	Course Code	Course Name	L	T	P	Credits	Types
1	CD490	Seminar	0	0	3	2	S / E
2	CD492	Major Project	0	0	16	8	MP2 / E
3	CD494	Intenship	0	0	30	15	I / E
4	GP	General Proficiency	Non Credit				
		Total Hours and Credits	**0**	**0**	**49**	**25**	

GRAND TOTAL OF CREDITS = 200

In the **Seminar**, student need to study and present individually, on latest research paper of their specialized area and It will be evaluated as per University Examination Rules.

The **Internship** in Industry will be done by candidate individually during the 8th semester and it will be for a minimum of 4 (-6) months. It will be evaluated as per University Examination Rules.

Minor and Major Project will be in a group and It will be evaluated as per University Examination Rules.

USICT will provide a mentor/supervisor for industrial training, seminar, internship, minor and major projects.

School of ICT
Gautam Buddha University
Greater Noida, (U.P.)

ELECTIVES FROM DCSE

S.No.	Course Code	Course Name	L	T	P	Credits	Types
1	CD311	Pattern Recognition	3	0	0	3	E1
2	CD313	Data Mining	3	0	0	3	E1
3	CD315	Data Science Life Cycle	3	0	0	3	E1
4	CD317	Data Storage Technologies and Networking	3	0	0	3	E1
5	CD319	Internet of Things	3	0	0	3	E1
6	CD312	Deep Learning	3	0	0	3	E2
7	CD314	Big Data Platforms	3	0	0	3	E2
8	CD316	Research Techniques for Data Science	3	0	0	3	E2
9	CD318	High Performance Computing	3	0	0	3	E2
10	CD320	Information Retrieval Systems	3	0	0	3	E2
11	CD407	Business Intelligence	3	0	0	3	E3
12	CD409	Computer Vision with Machine Learning	3	0	0	3	E3
13	CD411	Digital Image Processing	3	0	0	3	E3
14	CD413	Mobile and Wireless Network Security	3	0	0	3	E3
15	CD415	Quantam Computing	3	0	0	3	E3
16	CD417	SAS Programming	3	0	0	3	E4
17	CD419	Biomedical Image and signal processing	3	0	0	3	E4
18	CD421	AI Enabled Data Science	3	0	0	3	E4
19	CD423	Web Analytics	3	0	0	3	E4
20	CD425	Social Media Analytics and Techniques	3	0	0	3	E4

CD Computer Science & Engineering / Data Science for Course Code

C Core Course from USICT for course type

GE General Elective from related discipline of other Deptt./School

GE L General Elective Lab from related discipline of other Deptt./School

OE Open Elective from other discipline of other Deptt./School

AECC Ability Enhancement Compulsary Course

DSE Discipline Specific Course

SEC Skill Enhancement Course

E Elective from USICT

CC-L Core Course Lab from USICT

IT1 Industrial Training

MP Minor / Major Project

S Seminar

I Internship

28/09/23

School of ICT
Gautam Buddha University
Greater Noida, (U.P.)

BIG DATA ANALYTICS			
Course Code:	CD307	Course Credits:	3
Course Category:	CC	Course (U / P)	U
Course Year (U / P):	3U	Course Semester (U / P):	6U
No. of Lectures + Tutorials (Hrs/Week):	03 + 00	Mid Sem. Exam Hours:	1.5
Total No. of Lectures (L + T):	45 + 00	End Sem. Exam Hours:	3

COURSE OBJECTIVES
1. Understand the Big Data Platform and its Use cases
2. Provide an overview of Apache Hadoop
3. Provide HDFS Concepts and Interfacing with HDFS
4. Understand Map Reduce Jobs
5. Apply analytics on Structured, Unstructured Data. Exposure to Data Analytics with R.

COURSE OUTCOMES
At the end of the course the students should be able to:
1. Identify Big Data and its Business Implications Access and Process Data on Distributed File System
2. List the components of Hadoop and Hadoop Eco-System
3. Manage Job Execution in Hadoop Environment
4. Develop Big Data Solutions using Hadoop Eco System
5. Analyze Infosphere Big Insights Big Data Recommendations

UNIT I INTRODUCTION

Introduction to big data : Introduction to Big Data Platform, Challenges of Conventional Systems, Intelligent data analysis, Nature of Data, Analytic Processes and Tools, Analysis vs Reporting.

UNIT II DATA STREAMS

Mining data streams : Introduction To Streams Concepts, Stream Data Model and Architecture, Stream Computing, Sampling Data in a Stream, Filtering Streams, Counting Distinct Elements in a Stream , Estimating Moments, Counting Oneness in a Window, Decaying Window, Real time Analytics Platform(RTAP) Applications - Case Studies - Real Time Sentiment Analysis- Stock Market Predictions.

UNIT III HADOOP

Hadoop: History of Hadoop, the Hadoop Distributed File System, Components of Hadoop Analysing the Data with Hadoop- Scaling Out- Hadoop Streaming- Design of HDFS-Java interfaces to HDFS Basics, Developing a Map Reduce Application-How Map Reduce Works, Anatomy of a Map Reduce Job run, Failures, Job Scheduling-Shuffle and Sort – Task execution, Map Reduce Types and Formats- Map Reduce Features Hadoop environment.

UNIT IV DATA PROCESSING

School of ICT
Gautam Buddha University
Greater Noida, (U.P.)

Frameworks: Applications on Big Data Using Pig and Hive, Data processing operators in Pig, Hive services, HiveQL, Querying Data in Hive, fundamentals of HBase and ZooKeeper, IBM InfoSphere BigInsights and Streams.

UNIT V DATA ANALYTICS TECHNIQUE

Predictive Analytics- Simple linear regression, Multiple linear regression., Interpretation 5 of regression coefficients. Visualizations, Visual data analysis techniques, interaction techniques, Systems and applications.

Text Books:

1. Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer, 2007.

2. Tom White "Hadoop: The Definitive Guide" Third Edition, O'reilly Media, 2012.

3. Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos, "Understanding Big Data:

4. Analytics for Enterprise Class Hadoop and Streaming Data", McGrawHill Publishing, 2012.

School of ICT
Gautam Buddha University
Greater Noida, (U.P.)

BIG DATA ANALYTICS LAB

Course Code	**CD383**	**Course Credit**	**02**
Course Category	**CC**	**Course(U/P)**	**U**
Course year(U/P)	**3U**	**Course Semester**	**5U**
No of Lectures + Tutorials(Hrs./Week)	**03+00**	**Mid Semester Exam Hours:**	
Total no of Lectures(L+T)	**10+00**	**End Term Exam Hours:**	**03**
COURSE OBJECTIVES			
1. To study the basic technologies that forms the foundations of Big Data.			
2. To study the programming aspects of cloud computing with a view to rapid prototyping of complex applications.			
3. To understand the specialized aspects of big data including big data application, and big data analytics.			
4. To study different types Case studies on the current research and applications of the Hadoop and big data in industry			
5. Understand the HIVE database and Tables			
COURSE OUTCOMES			
1.At the end of the course the student should be able to understand the :			
2.Student must be Able to understand the building blocks of Big Data			
3.Student must be able to articulate the programming aspects of cloud computing(map Reduce etc)			
4.Student must be able to understand the specialized aspects of big data with the help of different big data applications			
5. Student must be able to represent the analytical aspects of Big Data			

LIST OF EXPERIMENTS:

1. Installation of VMWare to setup the Hadoop environment and its ecosystems.
2. Perform setting up and Installing Hadoop in its three operating modes. i. Standalone. ii. Pseudo distributed. iii. Fully distributed.
3. Use web based tools to monitor your Hadoop setup.
4. Implementing the basic commands of LINUX Operating System – File/Directory creation, deletion, update operations
5. Implement the following file management tasks in Hadoop: i. Adding files and directories ii. Retrieving files iii. Deleting files Hint: A typical Hadoop workflow creates data files (such as log files) elsewhere and copies them into HDFS using one of the above command line utilities.
6. Run a basic word count Map Reduce program to understand Map Reduce Paradigm.
7. Write a Map Reduce program that mines weather data. Hint: Weather sensors collecting data every hour at many locations across the globe gather a large volume of log data, which is a good candidate for analysis with Map Reduce, since it is semi structured and record-oriented.
8. Implement matrix multiplication with Hadoop Map Reduce
9. Installation of PIG.

School of ICT
Gautam Buddha University
Greater Noida, (U.P.)

10. Write Pig Latin scripts sort, group, join, project, and filter your data.
11. a. Run the Pig Latin Scripts to find Word Count b. Run the Pig Latin Scripts to find a max temp for each and every year
12. Installation of HIVE.
13. Use Hive to create, alter, and drop databases, tables, views, functions, and indexes

School of ICT
Gautam Buddha University
Greater Noida, (U.P.)

Parallel Processing and CUDA Programming			
Course Code:	CD401	Course Credits:	4
Course Category:CC	CC	Course (U / P)	
Course Year (U / P):U	4U	Course Semester (U / P):	7U
No. of Lectures + Tutorials H rs/Week):	3+ 01	id Sem. Exam Hours:	1
Total No. of Lectures (L + T):30	45+ 15	End Sem. Exam Hours:	3

COURSE OBJECTIVES
1. Define terminology commonly used in parallel computing, such as efficiency and speedup.
2. Describe common GPU architectures and programming models
3. Implement efficient algorithms for common application kernels, such as matrix multiplication
4. Given a problem, develop an efficient parallel algorithm to solve it.
5. Given a problem, implement an efficient and correct code to solve it, analyze its performance, and gi onvincing written and oral presentations explaining the achievements

COURSE OUTCOMES
At the end of the course the students should be able to:
1.Understand the distributed and parallel computing systems
2. Familiar with parallel and distributed languages MPI, Pthread, OpenMP, and CUDA
3. Design parallel and distributed algorithms using these parallel languages
4. Writing Parallel Programs.
5 Able to meaure performance metrics

Unit-1 Introduction

Introduction to Parallel Computers ,Message-Passing Computing and Programming, Multithread Programming , Parallel Programming Paradigms, Parallel Architecture, Parallel Architecture (case studies)

Unit-2

CUDA Programming , OpenMP Programming , Embarrassingly Parallel Computation, GPU-Compute Architecture, CUDA, Memory organization in CUDA Multi- CoreCPU programming, MPI, PVM, Performance evaluation and scalability

Unit-3

Partitioning and Divide-and-Conquer Strategies , Pipelined Computation ,Synchronous Computations, Pipelining and Throughput Latency and Latency hiding

Unit-4

Synchronous Computations,Load Balancing and Termination Detection,Distributed Shard Memory, Synchronization: Memory Consistency, Barriers (local versus global), Atomics, Memory fence. Prefix sum, Reduction. Programs for concurrent Data Structures such as Worklists, Linked-lists. Synchronization across CPU and GPU

Unit-5

School of ICT
Gautam Buddha University
Greater Noida, (U.P.)

Sorting Algorithms, Min/Max, Sum Searching, Merging, Sorting, Prefix operationsN- body problems, Matrix operations

REFERENCE BOOKS:

1. David Kirk and Wen-mei Hwu, Programming Massively Parallel Processors: A Hands-On Approach, 2nd Edition, Publisher: Morgan Kaufman, 2012, ISBN: 9780124159921.
2. Shane Cook, CUDA Programming: A Developer's Guide to Parallel Computing with GPUs, Morgan Kaufman; 2012 (ISBN: 978-0124159334
3. An Introduction to Parallel Algorithms by Joseph Jaja (Addison-Wesley Professional)
4. Introduction to Parallel Computing by Ananth Grama, George Karypis, Vipin Kumar and Anshul Gupta (Pearson)
5. Parallel Programaming in C with MPI and openMP by Michael J Quinn (McGraw H

School of ICT
Gautam Buddha University
Greater Noida, (U.P.)