Lingaya's Vidyapeeth

Deemed-to-be-University u/s 3 of UGC Act 1956, Government of India
NAAC ACCREDITED
Approved by MHRD / AICTE / PCI / BCI / COA / NCTE
Nachauli, Jasana Road, Faridabad- 121002 (Haryana)
Website: www.lingayasvidyapeeth.edu.in | Ph: 0129-2598200-05

COURSE PLAN & COURSE DATA SHEET

PROGRAM: BCA	DEGREE: UG
COURSE: Big Data Analytics	SEMESTER: VI CREDITS: 3
COURSE CODE: BCA-310 REGULATION: NA	COURSE TYPE: ELECTIVE
COURSE AREA/DOMAIN: IT	CONTACT HOURS: 3+1 (Tutorial) hours/Week.
CORRESPONDING LAB COURSE CODE (IF ANY): NA	LAB COURSE NAME (IF ANY): NA

PROGRAM EDUCATIONAL OBJECTIVES:

- Equip students with a strong foundation in statistics, computer science, and mathematics relevant to Big Data Analytics.
- Develop proficiency in managing, processing, and analyzing large datasets using advanced tools and technologies.
- Foster analytical thinking and problem-solving skills pertinent to extracting insights from complex data.
- Teach students the principles and applications of machine learning and artificial intelligence in the context of Big Data.
- Train students in effectively visualizing data and communicating their findings to both technical and non-technical audiences.
- Instill a strong understanding of ethical considerations, data privacy, and security issues in the handling and analysis of big data.
- Encourage research skills that enable students to contribute new knowledge or innovations in the field of Big Data Analytics.
- Provide practical experience through internships, projects, or collaborations with industry partners.
- Expose students to the applications of Big Data Analytics across various fields such as business, healthcare, finance, and more.
- Leadership and Teamwork: Cultivate leadership abilities and teamwork skills essential for working in diverse and interdisciplinary environments.
- Prepare students to adapt to rapidly evolving technologies and methodologies in Big Data.
- Promote a commitment to lifelong learning and continuous professional development in the field of Big Data Analytics.

SYLLABUS:

UNIT	DETAILS	HOURS
I	**INTRODUCTION TO BIG DATA:** Introduction – distributed file system – Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications. Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce, Challenges for processing big data, Using big data in businesses.	8
II	**INTRODUCTION TO HADOOP:** Introduction to Hadoop, why we use Hadoop, History of Hadoop, Use cases of Hadoop, Big Data – Apache Hadoop – Moving Data in and out of Hadoop – Understanding inputs and outputs of MapReduce - Data Serialization.	7
III		6

	HADOOP ARCHITECTURE: Hadoop Architecture, Hadoop Storage: HDFS, Common Hadoop Shell commands, Anatomy of File Write and Read., Name Node, Secondary Name Node, and Data Node, Hadoop MapReduce paradigm, Map and Reduce tasks, Job, Task trackers - Cluster Setup – SSH & Hadoop Configuration – HDFS Administering –Monitoring & Maintenance.	
IV	**HADOOP ECOSYSTEM AND YARN:** Learning MapReduce concepts and framework, Testing and Debugging Map Reduce Applications, Background of YARN; Hadoop YARN architecture; advantages of YARN, working with YARN, backward compatibility with YARN, YARN Commands, log management etc.	9
V	**HIVE AND HBASE:** Introduction to Hive and HBASE, HIVE: Architecture, Managing tables, data types, schemas, partitions, HBASE: Architecture, Schema design; Advance Indexing - PIG, Zookeeper - how it helps in monitoring a cluster, HBase uses Zookeeper and how to Build Applications with Zookeeper., HBASE commands, HIVE Vs RDMS, HBASE Vs RDMS.	12
	TOTAL HOURS	42

Teacher Centric Approach **TC1: Chalk and Talk, TC2: PPT, TC3: Video Lectures TC4: Blended learning**
Learner Centric Approach: **LC1: Assignment. LC2: Mini project. LC3: Quiz/Class test. LC 4: Seminar on recent trends.**

LC5: Group Task.	LC6: Others

DETAILED SESSION PLAN

Lecture session/ Number	Topics to be covered	CO addressed	Teacher Centric Approach	Learner Centric Approach	References	Relevance with POs and PSOs
1	INTRODUCTION TO BIG DATA : Introduction – distributed file system	1	TC1, TC2	LC1,LC2,LC3	T1/T2/R1	2
2	Big Data and its importance, Four Vs, Drivers for Big data	1	TC1, TC2	LC1,LC2,LC3	T1/T2/R1	3
3	Big data analytics, Big data applications	1	TC1, TC2	LC1,LC2,LC3	T1/T2/R1	3
4	Algorithms using map reduce	1	TC1, TC2	LC1,LC2,LC3	T1/T2/R1	2
5	Matrix-Vector Multiplication by Map Reduce	1	TC1, TC2	LC1,LC2,LC3	T1/T2/R1	3
6	Challenges for processing big data	1	TC1, TC2	LC1,LC2,LC3	T1/T2/R1	3
7	Using big data in businesses	1	TC1, TC2	LC1,LC2,LC3	T1/T2/R1	2
8	Doubt Class	1	TC1, TC2	LC1,LC2,LC3	R1/R2/R3	3
9	INTRODUCTION TO HADOOP: Introduction to Hadoop	2	TC1, TC2	LC1,LC2,LC3	T1/T2/R1	3
10	why we use Hadoop, History of Hadoop	2	TC1, TC2	LC1,LC2,LC3	T1/T2/R1	3
11	Use cases of Hadoop, Big Data – Apache Hadoop	2	TC1, TC2	LC1,LC2,LC3	T1/T2/R1	2
12	Moving Data in and out of Hadoop	2	TC1, TC2	LC1,LC2,LC3	T1/T2/R1	3

13	Understanding inputs and outputs of MapReduce	2	TC1, TC2	LC3	T1/T2/R1	2
14	Data Serialization	2	TC1, TC2	LC1,LC2,LC3	T1/T2/R1	3
15	Doubt Class	2	TC1, TC2	LC1,LC2,LC3	R1/R2/R3	3
16	HADOOP ARCHITECTURE: Hadoop Architecture, Hadoop Storage	3	TC1, TC2	LC1,LC2,LC3	T1/T2/R1	3
17	HDFS, Common Hadoop Shell commands , Anatomy of File Write and Read.	3	TC1, TC2	LC1,LC2,LC3	T1/T2/R1	2
18	Name Node, Secondary Name Node, and Data Node, Hadoop MapReduce paradigm	3	TC1, TC2	LC1,LC2,LC3	T1/T2/R1	2
19	Map and Reduce tasks, Job, Task trackers - Cluster Setup	3	TC1, TC2	LC3	R1/R2/R3	2
20	SSH & Hadoop Configuration – HDFS Administering – Monitoring & Maintenance	3	TC1, TC2	LC1,LC2,LC3	R1/R2/R3	3
21	Doubt Class	3	TC1, TC2	LC1,LC2,LC3	R1/R2/R3	3
22	HADOOP ECOSYSTEM AND YARN	4	TC1, TC2	LC1,LC2,LC3	R1/R2/R3	2

23	Learning MapReduce concepts and framework	4	TC1, TC2	LC1,LC2,LC3	R1/R2/R3	3
24	Testing and Debugging Map Reduce Applications	4	TC1, TC2	LC1,LC2,LC3,LC4	R1/R2/R3	2
25	Background of YARN	4	TC1, TC2	LC1,LC2,LC3	R1/R2/R3	3
26	Hadoop YARN architecture	4	TC1, TC2	LC1,LC2,LC3	R1/R2/R3	2
27	advantages of YARN, working with YARN, backward compatibility with YARN	4	TC1, TC2	LC1,LC2,LC3	R1/R2/R3	3
28	YARN Commands	4	TC1, TC2	LC3	R1/R2/R3	2
29	Doubt Class	4	TC1, TC2	LC1,LC2,LC3	R1/R2/R3	2
30	log management	4	TC1, TC2	LC1,LC2,LC3	R1/R2/R3	2
31	HIVE AND HBASE	5	TC1, TC2	LC1,LC2,LC3	R1/R2/R3	3
32	Introduction to Hive and HBASE	5	TC1, TC2	LC1,LC2,LC3	R1/R2/R3	3
33	HIVE: Architecture, Managing tables, data types, schemas, partitions	5	TC1, TC2	LC1,LC2,LC3	R1/R2/R3	2
34	HBASE: Architecture, Schema design	5	TC1, TC2	LC1,LC2,LC3,LC4	R1/R2/R3	3
35	Advance Indexing - PIG, Zookeeper - how it helps in monitoring a cluster	5	TC1, TC2	LC1,LC2,LC3	R1/R2/R3	2

36	HBase uses Zookeeper and how to Build Applications with Zookeeper	5	TC1, TC2	LC1,LC2,LC3	R1/R2/R3	2
37	HBASE commands	5	TC1, TC2	LC1,LC2,LC3	T1/T2/R1	2
38	HIVE Vs RDMS	5	TC1, TC2	LC3	T1/T2/R1	3
39	HBASE Vs RDMS.	5	TC1, TC2	LC1,LC2,LC3	T1/T2/R1	2
40	Doubt Class	5	TC1, TC2	LC1,LC2,LC3	T1/T2/R1	3
41	Quiz - 1		TC1, TC2	LC1,LC2,LC3	T1/T2/R1	3
42	Quiz - 2		TC1, TC2	LC1,LC2,LC3	T1/T2/R1	2

TEXT/REFERENCE BOOKS:

T/R	
1	Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, "Professional Hadoop Solutions", Wiley, ISBN: 9788126551071, 2015.
2	Chris Eaton, Dirk deroos et al. , "Understanding Big data ", McGraw Hill, 2012.
3	Tom White, "HADOOP: The definitive Guide" , O Reilly 2012. 6 IT2015 SRM(E&T)
4	Vignesh Prajapati, "Big Data Analytics with R and Haoop", Packet Publishing 2013.
5	Tom Plunkett, Brian Macdonald et al, "Oracle Big Data Handbook", Oracle Press, 2014.
6	Jy Liebowitz, "Big Data and Business analytics, CRC press, 2013.

 # WEB SOURCE REFERENCES (W):

1	https://www.javatpoint.com/software-engineering-tutorial
2	https://www.tutorialspoint.com/software_engineering/index.htm
3	W3schools.com
4	http://www.bigdatauniversity.com/

COURSE PRE-REQUISITES: Data Structure, Electronics and Mechanics Concepts

C.CODE	COURSE NAME	DESCRIPTION	SEM
BCA-310	Big Data Analytics	This course involves examining large, complex data sets to uncover hidden patterns, unknown	VI

		correlations, market trends, customer preferences, and other useful business information.	

COURSE OBJECTIVES:

1	Identify Big Data and its Business Implications.
2	List the components of Hadoop and Hadoop Eco-System
3	Access and Process Data on Distributed File System
4	Manage Job Execution in Hadoop Environment
5	Develop Big Data Solutions using Hadoop Eco System

COURSE OUTCOMES:

S.NO	DESCRIPTION	PO(1..12) MAPPING	PSO(1..3) MAPPING
CO1	Gain expertise in utilizing advanced data analysis tools and software for handling, processing, and interpreting large data sets.	PO1, PO2, PO5, PO9	PSO1
CO2	Acquire a thorough understanding of key big data concepts, architectures, and technologies including Hadoop, Spark, and NoSQL databases.	PO3, PO4, PO5, PO6	PSO3
CO3	Develop the ability to effectively visualize data and interpret analytical results for informed decision-making.	PO3, PO4, PO5, PO6	PSO3
CO4	Learn to apply machine learning algorithms and predictive modeling techniques to big data for actionable insights.	PO10, PO11	PSO2, PSO3
CO5	Understand the ethical and legal implications of data analytics, including data privacy, security, and responsible use of data.	PO1, PO2, PO5, PO9	PSO1
COURSE OVERALL PO/PSO MAPPING: 2/2			

COURSE OUTCOMES VS POs MAPPING (DETAILED; HIGH:3; MEDIUM:2; LOW:1**):**

S.NO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	1	1	1								2		2		
CO2	1	1	1								1		1	2	
CO3	2	1									1			1	1
CO4	1	1									1		1		1
CO5	1	1									1				1
CO1	1	1	1								2		2		

** For Entire Course, PO & PSO Mapping*

POs & PSO REFERENCE:

PO 1	Apply the knowledge of mathematics, science, engineering and Application fundamentals, and an engineering and Application specialization to the solution of complex engineering problems.	PO7	Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.	PSO1	To equip the students with theoretical and implementation knowledgebase in all the latest areas of Computer Science & Engineering for a successful career in software industries,

					pursuing higher studies, or entrepreneurial establishments.
PO 2	Identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.	PO8	Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.	PSO2	To nurture the students with the critical thinking abilities for better decision making by offering them a socially acceptable solutions to real life problems through computing paradigm.
PO 3	Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.	PO9	Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.	PSO3	To nurture the students with the comprehensive analytical and design abilities by offering them techno-commercially feasible solutions of real business problems through computing.
PO 4	Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.	PO10	Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.		
PO 5	Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.	PO11	Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.		
PO 6	Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.	PO12	Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.		

COs VS POs MAPPING JUSTIFICATION:

S.NO	PO/PSO MAPPED	LEVEL OF MAPPING	JUSTIFICATION

CO1	2/1	1	It ensures students acquire basic knowledge and understanding of Big Data Analytics, laying the groundwork for further study and application in this field.
CO2	3/3	2	It ensures intermediate proficiency, equipping students with essential skills and understanding necessary for practical applications and further advanced study.
CO3	3/3	2	It ensures students develop intermediate skills in data handling, analysis techniques, and application of theoretical concepts to real-world problems.
CO4	2/3	1	It indicates a basic understanding of Big Data Analytics, focusing on fundamental concepts and introductory skills, essential for building a strong foundational knowledge base.
CO5	2/1	2	It justifies that students will demonstrate intermediate proficiency in Big Data Analytics, showcasing practical application skills and a foundational understanding of key concepts and technologies.

GAPS IN THE SYLLABUS - TO MEET INDUSTRY/PROFESSION REQUIREMENTS, POs & PSOs:

SNO	DESCRIPTION	PROPOSED ACTIONS
1	Gap in Real-World Data Handling: The syllabus may lack extensive practical exposure to real-world, large-scale datasets.	Integrate more industry-based projects and case studies involving complex, real-world data.
2	Insufficient Focus on Emerging Technologies: The current syllabus might not adequately cover emerging tools and technologies in Big Data Analytics.	Update the curriculum regularly to include training in cutting-edge tools and methodologies.
3	Lack of Soft Skills Development: There's often a gap in developing soft skills crucial for the industry, like communication and teamwork.	Incorporate modules focused on developing presentation, communication, and collaborative skills.
4	Limited Industry Interaction: Students may not get enough exposure to industry professionals and real-world scenarios.	Facilitate guest lectures, internships, and mentorship programs with industry

		experts.
5	Inadequate Emphasis on Data Ethics and Privacy: Syllabus might not fully address the growing concerns around data ethics and privacy in the industry.	Include comprehensive modules on ethical data handling, privacy laws, and security protocols in the curriculum. .

PROPOSED ACTIONS: TOPICS BEYOND SYLLABUS/ASSIGNMENT/INDUSTRY VISIT/GUEST LECTURER/NPTEL ETC

 # TOPICS BEYOND SYLLABUS/ADVANCED TOPICS/DESIGN:

1	Quantum Computing and Big Data
2	Advanced Machine Learning Techniques
3	Real-Time Analytics and Stream Processing
4	Big Data in IoT and Edge Computing
5	Ethical AI and Responsible Data Usage
6	Advanced Visualization Techniques
7	Blockchain for Data Security and Integrity

DELIVERY/INSTRUCTIONAL METHODOLOGIES:

☐ CHALK & TALK	☐ STUD. ASSIGNMENT	☐ WEB RESOURCES	☐ NPTEL/OTHERS
☐ LCD/SMART BOARDS	☐ STUD. SEMINARS	☐ ADD-ON COURSES	☐ WEBNIARS

ASSESSMENT METHODOLOGIES-DIRECT

☐ ASSIGNMENTS	☐ STUD. SEMINARS	☐ TESTS/MODEL EXAMS	☐ UNIV. EXAMINATION
☐ STUD. LAB PRACTICES	☐ STUD. VIVA	☐ MINI/MAJOR PROJECTS	☐ CERTIFICATIONS
☐ ADD-ON COURSES	☐ OTHERS		

ASSESSMENT METHODOLOGIES-INDIRECT

☐ ASSESSMENT OF COURSE OUTCOMES (BY FEEDBACK, ONCE)	☐ STUDENT FEEDBACK ON FACULTY (TWICE)
☐ ASSESSMENT OF MINI/MAJOR PROJECTS BY EXT. EXPERTS	☐ OTHERS

 # INNOVATIONS IN TEACHING/LEARNING/EVALUATION PROCESSES:

1. Implementing virtual labs where students can engage in simulated big data projects, allowing them to apply concepts in a controlled, interactive environment. This hands-on approach enhances practical understanding and skills.
2. Introducing gamification in course content, where complex concepts are taught through interactive games and challenges. This approach can increase student engagement and make learning more enjoyable and effective.
3. Encouraging students to work on collaborative projects and participate in peer review sessions. This method promotes teamwork, enhances critical thinking, and provides diverse perspectives on solving big data problems.

4. Adopting a flipped classroom model where theoretical content is learned outside the classroom through videos and readings, and classroom time is devoted to discussions, problem-solving, and practical applications. This approach encourages active learning and better concept retention.

5. Utilizing AI to analyze student performance and create personalized learning paths. This technology can recommend resources, adjust difficulty levels, and provide tailored feedback, ensuring that each student's unique learning needs are addressed.

Prepared by
Dr. Tapsi Nagpal

Approved by
A.Dean &HOD

 # Additionally, the details to be compiled separately by the Departmental Coordinator for the entire Department.