VIT Vidyalankar Institute of Technology
Accredited A+ by NAAC

Version 06/24-7

Consolidated Academic Administration Plan for the Course

Big Data Analytics (Core) Sem. 7 – Program Computer Engineering

2024-25-Odd Semester

Faculty - Prof. Santosh Tamboli (Cluster Mentor) & Pankaj Vanwari

The academic resources available in VIT –

VMIS (ERP)	V-Refer and V-Live	VIT Library	VAC & MOOC Courses
Institute & Department Vision and Mission	Former IA question papers and solutions (prepared by faculty)	Former IA question papers solutions - hardcopy	Value Added Courses (VAC) are conducted throughout the semester & in the semester break - Enrol for the VACs
Program Educational Objectives (PEO)	MU end semester examination question papers and solutions (prepared by faculty)	MU end semester exam question paper & solutions - by faculty, hardcopy	
Program Specific Outcome (PSO)	Class notes and Digital Content for the subject (scanned / typed by faculty)	All text books, reference books, e -books mentioned in the syllabus & AAP	Online courses from NPTEL, Coursera etc. are pursued throughout the semester - Register for the course & get certified
Program Outcome (PO)	Comprehensive question bank, EQ, GQ, PPT, Class Test papers	Technical journals and magazines for reference	
Departmental Knowledge Map	Academic Administration Plan & Beyond Syllabus Activity report	VIT library is member of IIT Bombay Library	Watch former lectures captured in LMS at VIT

1.a Course Objectives (Write in detail – as per NBA guidelines)

Cognitive	What do you want students to know?	To know the basic concepts of networking and its topological design and mathematical functions like modulo arithmetic.
Affective	What do you want students to think / care about?	To understand how various supporting tools in providing assurance concerning privacy and integrity of information.
Behavioural	What do you want students to be able to do?	To provide skills to design security protocols for recognizing security problems and use of various cryptographic & security algorithms for real time applications.

Advice to Students:

Attend every class!!! Missing even one class can have a substantial effect on your ability to understand the course. Be prepared to think and concentrate, in the class and outside. I will try to make the class very interactive. Participate in the class discussions. Ask questions when you don't understand something. Keep up with the class readings. Start assignments and homework early. Meet me in office hour to discuss ideas, solutions or to check if, what you understand is correct.

The v-Refer Link

http://vidyalankarlive.com/vrefer/index.php/apps/files/?dir=/vRefer/CMPN/SEM%20VII/2024-25/BDA/PV&fileid=821880

https://teams.microsoft.com/l/team/19%3AbeIZSzL-N46Vx63b18M4IxzS7T2B_6keojrigkVfkI01%40thread.tacv2/conversations?groupId=c751066d-d065-4bba-8340-2351569e9da8&tenantId=c7b00d7f-ad99-442a-b12f-c2c912044fdc

Collaboration Policy:

We encourage discussion between students regarding the course material. However, no discussion of any sort is allowed with anyone on the assignment and homework for the class. If you find solution to some problems in a book or on the internet, you may use their idea for the solution; provided you acknowledge the source (name and page in the book or the website, if the idea is found on the internet). Even though you are allowed to use ideas from another source, you must write the solution in your own words. If you are unsure whether or not certain kinds of collaboration is possible please ask the teacher.

1.b Course Outcome (CO) Statements and Module-Wise Mapping (follow NBA guideline)

CO No.	Statements	Related Module/s
CO1	Understand the building blocks of Big Data Analytics	1
CO2	Apply fundamental enabling techniques like Hadoop and Map Reduce in solving real world problems	2
CO3	Understand different NoSQL systems and handles big data.	3
CO4	Apply advanced techniques for emerging applications like stream analytics.	4
CO5	Achieve adequate perspectives of big data analytics in various applications like recommender systems, social media applications etc..	5
CO6	Apply statistical computing techniques and graphics for analysing big data.	6

1.c Mapping of COs with POs (mark S: Strong, M: Moderate, W: Weak, Dash '–': not mapped)

(List of POs is available in V-refer)

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12
CO 1	S	M	-	-	M	-	-	-	-	-	-	1
CO 2	M	M	S	-	S	-	-	-	-	-	-	1
CO 3	-	M	W	-	S	-	-	-	-	-	-	2
CO 4	-	S	S	M	W	-	-	-	-	-	-	-
CO 5	M	S	M	S	M	-	-	-	S	-	-	2
CO 6	M	S	S	M	S	-	-	-	M	-	-	2

1.d Mapping of COs with PSOs (mark S: Strong, M: Moderate, W: Weak, Dash '–':not mapped)

	PSO 1	PSO 2	PSO 3	PSO 4
CO 1	M	W	W	-
CO 2	M	W	W	-
CO 3	M	W	W	-

CO 4	M	W	-	-
CO 5	S	W	M	-
CO6	S	W	W	-

1.e Teaching and Examination Scheme (As specified by the University) for the Course

Categories	Humanities and Social Sciences	Basic Science	Engineering Science	Professional Core	General Education	Professional Elective	Project/ Internship	Open Elective
Tick suitable category				√				

Subject Code	Subject Name	Teaching Scheme			Credits Assigned			
		Theory	Practical	Tutorial	Theory	TW/Practical	Tutorial	Total
CSC702	Big Data Analytics	3	-	-	3	-	-	3
CSL702	Big Data Analytics Lab	-	2	-	-	1	-	1

Subject Code	Subject Name	Examination Scheme					
		Theory			Practical/Oral		Total
		ISA	MSE	ESE	ISA	ESE	
CSC702	Big Data Analytics	20	30	50	-	-	-
CSL702	Big Data Analytics Lab	-	-	-	25	25	50

1.f Faculty-Wise Distribution of all Lecture-Practical-Tutorial Hours for the Course

Divisions	Lecture (Hrs.)	Practical (Hrs.)				Tutorial (Hrs.)			
		Batch 1	Batch 2	Batch 3	Batch 4	Batch 1	Batch 2	Batch 3	Batch 4
A	PV (3)	UMK(2)	PV (2)	PV (2)	PV (2)	-	-	-	-
B	PV (3)	UMK(2)	PV (2)	PV (2)	PV (2)	-	-	-	-

1.g Office Hours (Faculty will be available in office in this duration for solving students' query)

Division	Day	Time (at least 1 Hr. / Division)	Venue (Office Room No.)
A	Thursday	03:45 PM to 04:45 PM	M209
B	Friday	03:45 PM to 04:45 PM	M209

2.a Syllabus : Module Wise Teaching Hours and % Weightage in University Question Paper

Module No.	Module Title and Brief Details	Teaching Hrs. for each module	% Weightage in University Question Papers
1	**Introduction to Big Data and Hadoop:** 1.1 Introduction to Big Data - Big Data characteristics and Types of Big Data 1.2 Traditional vs. Big Data business approach 1.3 Case Study of Big Data Solutions 1.4 Concept of Hadoop, Core Hadoop Components; Hadoop Ecosystem	02	5%
Learning Outcome	Differentiate between traditional and big data business approach. Discuss various types of big data. Explain characteristics of big data.		
2	**Hadoop HDFS and MapReduce:** 2.1 Distributed File Systems: Physical Organization of Compute Nodes, Large Scale File-System Organization. 2.2 MapReduce: The Map Tasks, Grouping by Key, The Reduce Tasks, Combiners, Details of MapReduce Execution, Coping with Node Failures. 2.3 Algorithms Using MapReduce: Matrix-Vector Multiplication by MapReduce, Relational-Algebra Operations, Computing Selections by MapReduce, Computing Projections by MapReduce, Union, Intersection, and Difference by MapReduce 2.4 Hadoop Limitations	08	21%
Learning Outcome	Perform Matrix multiplication using MapReduce. Compute projections using MapReduce. Explain MapReduce with an example. Discuss the Distributed File systems.		
3	**NoSQL:** 3.1 Introduction to NoSQL, NoSQL Business Drivers 3.2 NoSQL Data Architecture Patterns: Key-value stores, Graph stores, Column family (Bigtable)stores, Document stores, Variations of NoSQL architectural patterns, NoSQL Case Study 3.3 NoSQL solution for big data, Understanding the types of big data problems; Analyzing big data with a shared-nothing architecture; Choosing distribution models: master-slave versus peer-to-peer; NoSQL systems to handle big data problems.	11	26%
Learning Outcome	Explain NoSQL in detail. Discuss a case study on NoSQL. Explain a NoSQL solution for big data. Analyze big data with shared-nothing architecture.		
4	**Mining Data Streams:** 4.1 The Stream Data Model: A Data-Stream-Management System, Examples of Stream Sources, Stream Queries, Issues in Stream Processing. 4.2 Sampling Data techniques in a Stream 4.3 Filtering Streams: Bloom Filter with Analysis. 4.4 Counting Distinct Elements in a Stream, CountDistinct Problem, Flajolet-Martin Algorithm, Combining Estimates, Space Requirements 4.5 Counting Ones in a Window: The Cost of Exact Counts, The Datar-Gionis-Indyk-Motwani Algorithm, Query Answering in the DGIM Algorithm, Decaying Windows.	11	28%
Learning Outcome	Give examples of stream sources. Explain the issues in stream processing. Discuss various Sampling Data techniques in a Stream. Explain Bloom filter in detail. Explain DGIM algorithm in detail.		
5	**Real-Time Big Data Models** 5.1 A Model for Recommendation Systems, Content-Based Recommendations, Collaborative Filtering 5.2 Case Study: Product Recommendation 5.3 Social Networks as Graphs, Clustering of Social-Network Graphs, Direct Discovery of Communities in a social graph	04	10%

Learning Outcome	Explain need of recommendation systems. Differentiate between Content-based and Collaborative filtering. Discuss a case study on product recommendation.		
6	**Data Analytics with R:** 6.1 Exploring Basic features of R, Exploring RGUI, Exploring RStudio, Handling, Basic Expressions in R, Variables in R, Working with Vectors, Storing and Calculating Values in R, Creating and using Objects, Interacting with users, Handling data in R workspace, Executing Scripts, Creating Plots, Accessing help and documentation in R 6.2 Reading datasets and Exporting data from R, Manipulating and Processing Data in R, Using functions instead of script, built-in functions in R 6.3 Data Visualization: Types, Applications	04	10%
Learning Outcome	Differentiate between Python and R language. Explain variables in R. How to handle data in R workspace. Explain applications of Data visualization.		
	Total	**39**	**100**

2.b Prerequisite Courses

No.	Semester	Name of the Course	Topic/s
1	4	Database Management System	Data Consistency
2	5	Data Warehousing and Mining	Data Analysis

2.c Relevance to Future Courses

No.	Semester	Name of the Course
1	8	Applied Data Science
2	8	Social Media Analytics

2.d

Identify real life scenarios/examples which uses the knowledge of the subject (Discussion on how to prepare examples and case studies e.g. "Boeing Plane": C Programming Language – Intro to Computer Science – Harvard's CS50 (2018) – Bing video)

Real Life Scenario	Concept Used
E-commerce and Retail	Recommendation system
Healthcare	Data Visualization
Social Media Platforms	Community detection algorithms

3 Past Results – Division-Wise

Details	Target – Dec 2024	Dec 2023	Dec 2022	Dec 2021
Course Passing % – Average of 2 Divisions	100	100	100	100

Marks Obtained by Course Topper (mark/80)	75	76	77	77

	Division A		Division B	
Year	Initials of Teacher	% Result	Initials of Teacher	% Result
Dec 2023	PJP	100	PJP	100
Dec 2022	PV	100	PJP	100
Dec 2021	PV	100	PV	100

4 All the Learning Resources – Books and E-Resources

4.a List of Text Books (T – Symbol for Text Books) to be Referred by Students

Sr. No	Text Book Titles	Author/s	Publisher	Edition	Module Nos.
1	Mining of Massive Datasets	Anand Rajaraman and Jeff Ullman	Cambridge University Press	2010	1,2,4,5
2	Hadoop in Practice	Alex Holmes	Manning Press	2012	1, 2
3	Making Sense of NoSQL	Dan McCreary and Ann Kelly	Manning Press	2014	3
4	Big Data Black Book	EMC Educational Services	Wiley	2020	6

4.b List of Reference Books (R – Symbol for Reference Books) to be Referred by Students

Sr. No	Reference Book Titles	Author/s	Publisher	Edition	Module Nos.
1	Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics	Bill Franks	Wiley	2012	1
2	Hadoop in Action	Chuck Lam	Dreamtech Press	2011	1, 2, 3
3	Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners	Jared Dean	Wiley	2014	5, 6
4	Data Mining: Concepts and Techniques	Jiawei Han and Micheline Kamber	Morgan Kaufmann	2010	4
5	Data Mining and Knowledge Discovery Handbook	Lior Rokach and Oded Maimon	Springer	2010	6
6	Hadoop The Definitive Guide	Tom White	O'Reilly	2018	2, 3

4.c List of E - Books (E – Symbol for E-Books) to be Referred by Students

Sr. No	E- Book Titles	Author/s	Publisher	Edition	Module Nos.
1	Big Data Analytics with R and Hadoop	Vignesh Prajapati	PACKT	2013	1, 2, 3

4.d Reading latest / top rated research papers (at least 5 papers)

Name of Paper	Name of Authors (Background)	Published in		Problem Statement
		Date	Journal	
A Thematic Travel Recommendation System Using an Augmented Big Data Analytical Model	Suriya Priya R. Asaithambi, Ramanathan Venkatraman and Sitalakshmi Venkatraman	7 February 2023	Technologies, MDPI	This paper proposes an end-to-end prototype Travel Recommendation System to evaluate their proposed big data-based recommendation model using augmented user-centric data analytics of various themes and arguments with relevant current environment details of potential destinations.
Large Language Models as Recommendation Systems in Museums	Georgios Trichopoulos, Markos Konstantakis, Georgios Alexandridis and George Caridakis	10 September 2023	Electronics, MDPI	This paper proposes the utilization of large language models as recommendation systems for museum visitors. Generative Pre-trained Transformer 4, a knowledge-based large language model is fine-tuned and turned into a context-aware recommendation system, adapting its suggestions based on user input and specific contextual factors such as location, time of visit, and other relevant parameters.
Movie Recommender System Using Parameter Tuning of User and Movie Neighbourhood via Co-Clustering	Sonu Airen and Jitendra Agrawal	2023	Procedia Computer Science, Elsevier	The focus of this paper is Partitional Weighted co-clustering for Movie Recommender System. The primary objective of this research article is to fine tune the parameters of user and movie neighborhoods by setting different values for row clusters number and column clutsers number parameters of co-clustering.
Social media content classification and community detection using deep learning and graph analytics	Mohsan Ali, Mehdi Hassan, Kashif Kifayat, Jin Young Kim, Saqib Hakak e, Muhammad Khurram Khan	2022	Technological Forecasting and Social Change, Elsevier	This study proposes novel deep learning and graph-based approaches to identifying hate content, followed by approaches to detecting communities and exploring social media to detect hate content. Twitter is used as a case study, and tweets are extracted and annotated by linguistic experts to develop a dataset for experimentation and validation.
A hybrid personality-aware	Sahraoui Dhelim,	2022	Journal of Ambient	In this paper, we study and compare four personality-aware

recommendation system based on personality traits and types models	Liming Chen, Nyothiri Aung, Wenyin Zhang and Huansheng Ning		Intelligence and Humanized Computing, Springer	recommendation systems based on different personality models, namely Big-Five traits model, Eysenck model and HEXACO model from the personality traits theory, and Myers–Briggs Type Indicator (MPTI) from the personality types of theory. They also propose a hybrid personality model for recommendation that takes advantage of the personality traits models, as well as the personality types models.

4.e Based on research paper an identify the current Problem statement

Problem Statement			Used in				
	Quiz	Assignment	Lab	Mini Project	Poster Presentation	Test	Any Other
Develop hybrid recommender built using various Spark models such as naïve Bayes classifier, trigonometric functions, deep learning convolutional neural network (CNN), time series, and NLP with sentiment scores using AFINN.		✓		✓			
Develop a recommendation system using LLMs. You can perform fine-tuning of GPT-4 through the incorporation of contextual information and user instructions during the training process.		✓		✓			
Develop a system to for content classification and community detection on any social media platform. LSTM-GRU model can be fine-tuned to classify hate content into six categories. The Girvan–Newman algorithm can be employed community detection.		✓		✓			

4.f Identify Companies / Industries which use the knowledge of the subject and thus may provide Internships and final Placements

Name of the Company	To be / Contacted for		
	Student Internship	Student Final Placement	Faculty Internship
JPMorgan Chase	√	√	
Capgemini	√	√	

4.g Identify suitable relevant TOP Guest Speakers from Industry (CS50 Lecture by Mark Zuckerberg - 7 December 2005 - YouTube)

Name of the Identified Guest Speaker	Designation	Name of the Company
Dr Anuradha Bhatia	Head - Data Analytics & AI	Standard Chartered
Mustafa Fatakdawala	Data Scientist	TCS

4.h Identify relevant Technical competitions to participate [Competitions -Paper Presentations, Projects, Hackathons, IVs etc..]

Name of the Relevant Technical Competition Identified to participate	Organized by	Date of the Event
HR Analytics	Analytics Vidhya	28-01-2019 10:00 AM to 31-12-2024 11:59 PM
Big Data and Artificial Intelligence Conference (BDA)	University of Hyderabad, Telangana, India	December 17-20, 2024

4.i Identify faculty in TOP schools / Universities who are teaching same / similar subject and develop rapport e.g. Exchange Lecture Material (Assignments / Tests / Project etc..), Joint Paper Publication

University	Name of the Course	Name of Faculty	Type of Collaboration		
			Exchange of Lecture Material	Joint Publication/ Research	Other
IIT Patna	Big Data Computing	Prof. Rajiv Misra			√

4.j Module Best Available in - Title best resource [from *4.a* to *4.d* in this AAP] & give details

Module No.	Title of the Module	Web Link	Mention the Tile			
			Journal	E-Journal	Magazine	Other Resource
1	Introduction to Big Data and Hadoop	https://towardsdatascience.com/ https://www.inderscience.com/jhome.php?jcode=ijbdi https://www.frontiersin.org/journals/big-data https://www.sciencedirect.com/journal/big-data-research https://journalofbigdata.springeropen.com/	Journal of Big Data	Frontiers in Big Data	Big Data Quarterly - The Big Data Magazine	
2	Hadoop HDFS and MapReduce	https://towardsdatascience.com/ https://www.inderscience.com/jhome.php?jcode=ijbdi https://www.frontiersin.org/journals/big-data https://www.sciencedirect.com/journal/big-data-research https://journalofbigdata.springeropen.com/	Big Data Research	Frontiers in Big Data	Big Data Quarterly - The Big Data Magazine	
3	NoSQL	https://towardsdatascience.com/ https://www.inderscience.com/jhome.php?jcode=ijbdi https://www.frontiersin.org/journals/big-data https://www.sciencedirect.com/journal/big-data-research	Journal of Big Data	Frontiers in Big Data	Big Data Quarterly - The Big Data Magazine	

		https://journalofbigdata.springeropen.com/				
4	Mining Data Streams	https://towardsdatascience.com/ https://www.inderscience.com/jhome.php?jcode=ijbdi https://www.frontiersin.org/journals/big-data https://www.sciencedirect.com/journal/big-data-research https://journalofbigdata.springeropen.com/	Big Data Research	Frontiers in Big Data	Big Data Quarterly - The Big Data Magazine	
5	Real-Time Big Data Models	https://towardsdatascience.com/ https://www.inderscience.com/jhome.php?jcode=ijbdi https://www.frontiersin.org/journals/big-data https://www.sciencedirect.com/journal/big-data-research https://journalofbigdata.springeropen.com/	Journal of Big Data	Frontiers in Big Data	Big Data Quarterly - The Big Data Magazine	
6	Data Analytics with R	https://towardsdatascience.com/ https://www.inderscience.com/jhome.php?jcode=ijbdi https://www.frontiersin.org/journals/big-data https://www.sciencedirect.com/journal/big-data-research https://journalofbigdata.springeropen.com/	Big Data Research	Frontiers in Big Data	Big Data Quarterly - The Big Data Magazine	

4.k **Referred to any top-rated university in that subject for content**

University	Name of the Course	Name of Faculty	Date of Delivery of the Course	Remarks
IIT Patna	Big Data Computing	Prof. Rajiv Misra	21 Aug 2023	

4.l **Faculty received any certification related to this subject. List of Certifications Identified / Done**

Course	Certifying Agency	No. of Hours	Level of the Course		Certification		Remarks
			Introductory	Advance Skill Development	Done on	Proposed to be on	
Introduction to R	Infosys Springboard	15	✓		11 Nov 2022		
Big Data Computing	NPTEL	8 weeks	✓			Nov 2024	

4.m **Completed subject wise/cluster wise training with cluster mentor.**
List of relevant Refresher Course Identified / Done

Course	Certifying Agency (As suggested by DAB/Cluster Mentor/Industry/ University other than MU)	Certification		Remarks
		Done on	Proposed to be on	
Pedagogy				
PBL	Introduction to R by Infosys Springboard	11 Nov 2022		

Sub. Content Training				

4.n Best Practices Identified and adopted

No.	Item	Best Practices Identified		
		Standford	Univ. 2	Univ. 3
1	Microsite	https://web.stanford.edu/class/cs246/	1	Microsite
2	Video Lectures	https://www.youtube.com/watch?v=jofiaetm5bY	2	Video Lectures
3	Assignments		3	Assignments
4	Mini Project	https://colab.research.google.com/drive/1ntuhHz0ohwg-0CQWFU1l2oqWB4gbE72T?usp=sharing	4	Mini Project
5	Assessment Metric		5	Assessment Metric
6	Quizzes		6	Quizzes
7	Labs/ Practical (PBL)		7	Labs/ Practical (PBL)
8	Tests		8	Tests
9	Peer Assessment		9	Peer Assessment
10	Any Other		10	Any Other

4.o Web Links for Online Notes/YouTube/VIT Digital Content/VIT Lecture Capture/NPTEL Videos

Students can view lectures by VIT professors, captured through LMS 'Lecture Capture' in VIT campus for previous years.

No.	Websites / Links	Module Nos.
1	https://onlinecourses.nptel.ac.in/noc23_cs112/preview	All
2	https://infyspringboard.onwingspan.com/	1, 2, 3, 4

4.p Recommended MOOC Courses like Coursera / NPTEL / MIT-OCW / edX/VAC etc.

Sr. No.	MOOC Course Link	Course conducted by – Person / University / Institute / Industry	Course Duration	Certificate (Y / N)
1	https://onlinecourses.nptel.ac.in/noc23_cs112/preview	NPTEL	8 weeks	N
2	https://infyspringboard.onwingspan.com/	cognitiveclass.ai	36 hrs	N
3	https://www.mygreatlearning.com/academy/learn-for-free/courses/mastering-big-data-analytics	mygreatlearning	19 hrs	N

5 Consolidated Course Lesson Plan

	From (date/month/year)	From (date/month/year)	Total Number of Weeks

Semester Duration	08/07/2024	18/10/2024	15

Week	Lecture no.	Module No.	Lecture Topics / MSE / BSA planned to be covered	Actual date of Completion (Handwritten)	COs Mapped	Recommended Prior Viewing / Reading	
						Lecture No. (on LMS)	Chapter No./ Books/ Web Site
1	1,2	1	**Introduction to Big Data and Hadoop** 1.1 Introduction to Big Data 1.2 Big Data characteristics, types of Big Data	--	CO1	--	R1, Chp 1, Pg. No. 4-27
	3	1	1.3 Traditional vs. Big Data business approach 1.4 Case Study of Big Data Solutions	--	CO1	--	R2, Chp 3, Pg. No. 37-60
2	4,5	1	1.5 Concept of Hadoop 1.6 Core Hadoop Components; Hadoop Ecosystem **Pop Quiz**	--	CO1	--	R1, Chp 1, Pg. No. 4-27
	6	2	**Hadoop HDFS and MapReduce** 2.1 Distributed File Systems: Physical Organization of Compute Nodes, Large-Scale File-System Organization	--	CO2	--	R7, Part II, Pg. No. 141-279
3	7,8	2	2.2 MapReduce: The Map Tasks, Grouping by Key, The Reduce Tasks, Combiners, Details of MapReduce Execution, Coping with Node Failures	--	CO2	--	R7, Part II, Pg. No. 141-279
	9	2	2.3 Algorithms Using MapReduce: Matrix-Vector Multiplication by MapReduce, Relational-Algebra Operations	--	CO2	--	R7, Part II, Pg. No. 141-279
4	10,11	2	Computing Selections by MapReduce, Computing Projections by MapReduce,	--	CO2	--	R7, Part II, Pg. No. 141-279

Week	Lecture no.	Module No.	Lecture Topics / MSE / BSA planned to be covered	Actual date of Completion (Handwritten)	COs Mapped	Recommended Prior Viewing / Reading	
						Lecture No. (on LMS)	Chapter No./ Books/ Web Site
			Union, Intersection, and Difference by MapReduce				
	12	2	2.4 Hadoop Limitations **Guest Lecture**	--	CO2	--	R7, Part II, Pg. No. 141-279
5	13,14	3	**NoSQL** 3.1 Introduction to NoSQL, NoSQL Business Drivers	--	CO3	--	Chp 4/ Pg. No. 62-95/ T3
	15	3	3.2 NoSQL Data Architecture Patterns: Key-value stores, Graph stores, Column family (Bigtable)stores, Document stores, Variations of NoSQL architectural patterns, NoSQL Case Study	--	CO3	--	Chp 4/ Pg. No. 62-95/ T3
6	--	--	MSE	--	--	--	--
	--	--	MSE	--	--	--	--
7	16,17	3	3.3 NoSQL solution for big data, Understanding the types of big data problems; Analyzing big data with a shared-nothing architecture; Choosing distribution models: master-slave versus peer-to-peer; NoSQL systems to handle big data problems.	--	CO3	--	Chp 4/ Pg. No. 62-95/ T3
	18	4	**Mining Data Streams:** 4.1 The Stream Data Model: A Data-Stream-Management System, Examples of Stream Sources, Stream Queries, Issues in Stream Processing	--	CO4	--	Chp 4/ Pg. 123/ T1

Week	Lecture no.	Module No.	Lecture Topics / MSE / BSA planned to be covered	Actual date of Completion (Handwritten)	COs Mapped	Recommended Prior Viewing / Reading	
						Lecture No. (on LMS)	Chapter No./ Books/ Web Site
8	19,20	4	4.2 Sampling Data techniques in a Stream	--	CO4	--	Chp 4/ Pg 127/ T1
	21	4	4.3 Filtering Streams: Bloom Filter with Analysis	--	CO4	--	Chp 4/ Pg 130/ T1
9	22,23	4	4.4 Counting Distinct Elements in a Stream, Count-Distinct Problem, Flajolet-Martin Algorithm, Combining Estimates, Space Requirements	--	CO4	--	Chp 4/ Pg 133/ T1
	24	4	4.5 Counting Frequent Items in a Stream, Sampling Methods for Streams, Frequent Itemsets in Decaying Windows	--	CO4	--	Chp 6/ Pg 191-224/ T1
10	25,26	4	4.6 Counting Ones in a Window: The Cost of Exact Counts, The Datar-Gionis-Indyk-Motwani Algorithm,	--	CO4	--	Chp 4/ Pg 142-149/ T1
	27	4	Query Answering in the DGIM Algorithm, Decaying Windows	--	CO4	--	Chp 3/ Pg87-92/ T1
11	28,29	5	**Real-Time Big Data Models** 5.1 A Model for Recommendation Systems, Content-Based Recommendations, Collaborative Filtering	--	CO 5	--	Chp 7/ Pg 249-255/ T1
	30	5	5.2 Case Study: Product Recommendation	--	CO5	--	Chp 7/ Pg 256-265/ T1

Week	Lecture no.	Module No.	Lecture Topics / MSE / BSA planned to be covered	Actual date of Completion (Handwritten)	COs Mapped	Recommended Prior Viewing / Reading	
						Lecture No. (on LMS)	Chapter No./ Books/ Web Site
12	31,32	5	5.3 Social Networks as Graphs, Clustering of Social-Network Graphs,	--	CO5	--	Chp 7/ Pg 256-265/ T1
	33	5	Direct Discovery of Communities in a social graph	--	CO5	--	Chp 5/ Pg 154-186/ T1
13	34,35	6	**Data Analytics with R** 6.1 Exploring Basic features of R, Exploring RGUI, Exploring RStudio, Handling Basic Expressions in R, Variables in R, Working with Vectors, Storing and Calculating Values in R,	--	CO6	--	Chp 5/ Pg 154-186/ T1
	36	6	Creating and using Objects, Interacting with users, Handling data in R workspace, Executing Scripts, Creating Plots, Accessing help and documentation in R		CO6	--	Chp 9/ Pg 292-311/ T1
14	37,38	6	6.2 Reading datasets and Exporting data from R, Manipulating and Processing Data in R, Using functions instead of script, built-in functions in R		CO6	--	Chp 10/ Pg 325-337/ T1
	39	6	6.3 Data Visualization: Types, Applications		CO6	--	Chp 10/ Pg 338/ T1
15							

6 Rubric for Grading and Marking of Term Work (inform students at the beginning of semester)

Theory (ISA=20)			Practical (ISA= 25)			Total
Class Participation	Certification	10 Assignments	Lab Participation	Lab work	Mini Project	
05	05	10	05	10	10	45

7 Assignments / Tutorials Details

Assignment/ Tutorial No.	Title of the Assignments / Tutorials	CO Map	Assignment/ Tutorials given to Students on	Week of Submission
1	Hadoop Ecosystem and Case Study	CO1	Week 2	Week 3
2	Pop Quiz on Hadoop components	CO1	Week 4	Week 4
3	Take Home Test on designing map reduce algorithms	CO2	Week 5	Week 6
4	Video Quiz on NoSQL	CO3	Week 6	Week 6
5	Poster Presentation on latest NoSQL platforms	CO3	Week 7	Week 9
6	Open Book Test on Stream Analysis	CO4	Week 8	Week 8
7	Pop Quiz on Prerequisites from Data Mining	CO5	Week 9	Week 9
8	Summary of landmark papers	All	Week 10	Week 11
9	Technical Paper Review	All	Week 11	Week 13
10	Student Seminar	All	Week 12	Week 13
11	Certification	All	22/07/2024	Week 14

Analysis of Assignment / Tutorial Questions and Related Resources

Assignment / Tutorial No.	Week No.	Type* (√)			Module No.	Based on #			Question Type (√)	
		R	PQ	OBT		Text Book	Reference Book	Other Learning Resource	MU EQ	Thought Provoking
1	2	√			1	√			√	
2	4		√		1		√		√	
3	5	√			2	√			√	
4	6		√		3			√	√	
5	7	√			3			√		√
6	8			√	4	√			√	
7	9		√		5	√			√	
8	10	√			All			√		√
9	11	√			All			√		√
10	12	√			All			√		√
11	3	√			All			√	√	

* Tick (√) the Type of the Assignment: Regular (R); Pop Quiz (PQ) ; Open Book Test for TE/BE/ME (OBT)
# Write number for text book, reference book, other learning resource from this AAP – *from Points 4.a to 4.d*

8 In Semester Assessment (ISE) / Other Class Test / Open Book Test (OBT)/Take Home Test (THT) Details

Tests	Test Dates	Module No.	CO Map	MSE Question Paper Pattern	Policy
MSE1	-	1 and 2	CO1 and CO2	Q1. 3 questions of 5 marks (attempt any 2) Q.2 and Q.3. 2 questions of 10 marks (attempt any 1)	40:30:30
MSE2	-	3, 4 and 5	CO3, CO4 and CO5	Q1. 3 questions of 5 marks (attempt any 2) Q.2 and Q.3. 2 questions of 10 marks (attempt any 1)	40:30:30
Pop Quiz	Week 4 and 9	1 and 5	CO1 and CO5		
Open Book Test	Week 8	4	CO4		
Take Home Test	Week 5	2	CO2		
Class tests / prelims					

Class tests / prelims					
Any other test/exams					

*** Failures of IA test (IA1+IA2) shall appear for IA test in the next semester. There is no provision for re-test in the same semester.**

9.a Practical Activities

Practical No.	Module No.	Title of the **Experiments**	Type of Experiment		Topics to be highlighted	CO Map
			PBL	Newly Added		
1	1	Set up a Hadoop cluster and verify its functionality.			Hadoop	CO1
2	2	Create directories and files in HDFS and perform read/write operations			HDFS	CO2
3	2	Develop and execute a simple MapReduce program to determine average sales of each product category.		√	MapReduce	CO2
4	3	To install and configure MongoDB to execute NoSQL commands			NoSQL	CO3
5	4	Data Stream Algorithms Implementation (any one): DGIM/ Bloom Filter / Flajolet Martin			Mining Data Streams	CO4
6	6	Clustering algorithm (eg. CURE)		√	Any Lang	CO5
7	6	Social Network Analysis	√		R	CO6
8	4 to 6	Mini Project: One real life large data application to be implemented.	√		Mini Project	All

10 Beyond Syllabus Activities for Gap Mitigation

No.	Type of the Activity	Activities	Number of beneficiaries	Other Details – guest profile, feedback, mark sheet, report
1	**Experiential learning/Interaction with Outside World**	1- Guest Lectures by Industry Expert	NA	
		2- Workshops	NA	
		3- Mini Project	Y	
		4- Industrial Visit	NA	
		5- Any other activity	NA	

2	**Collaborative & Group Activity**	6- Poster Presentation	Y	
		7- Minute Papers	Y	
		8- Students Seminars	Y	
		9- Students Debates	NA	
		10- Panel Discussion / Mock GD	NA	
		11- Mock Interview	NA	
		12- Any other activity	NA	
3	**Co-Curricular Activity**	13- Informative videos (NPTEL/Youtube /TEDx/ MIT OW/edX)	Y	
		14- Lecture Capture Usage	Y	
		15- Any other activity	NA	
4	**Tests & Assessments**	16- Class Tests/ Weekly Tests	NA	
		17- Pop Quiz	Y	
		18- Mobile App Based Quiz	NA	
		19- Open Book Test	NA	
		20- Take Home Test	Y	
		21- Any other activity	NA	

11 AAP/ Lecture Guide

No.	Programme	Course	Uploaded on V-refer	Date
1	CMPN	BDA	Yes	29/07/2024
2				
3				

*** Do not delete any activity. Give details for planned events. Write 'NA' for activity Not Planned.**

Consolidated Academic Administration Plan Prepared by (mention all theory teaching faculty names with signature)

Please write below your name and sign with date of the external cluster mentor meeting

Pankaj Vanwari	Faculty 2	Faculty 3

External Industry Mentor	External Academic Mentor	Prof. Santosh Tamboli VIT Cluster Mentor	Dr. Sachin Bojewar CMPN HOD