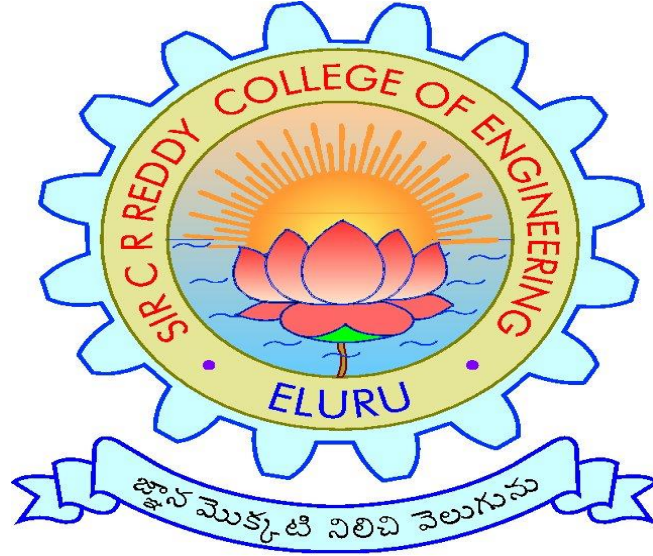


SIR C.R.REDDY COLLEGE OF ENGINEERING, ELURU
DEPARTMENT OF INFORMATION TECHNOLOGY
COURSE HANDOUT



SUBJECT: BIG DATA ANALYTICS
CLASS: III/IV B.Tech., II SEMESTER, A.Y.2022-23
INSTRUCTOR: Smt. B.Lalitha Bhavani

Course Handout Index

S.No	Description
1	College Vision & Mission
2	Department Vision & Mission
3	Program Educational Objectives (PEOs)
4	Program Outcomes (POs)
5	Program Specific Outcomes (PSOs)
6	JNTUK Academic Calendar
7	Department Academic Calendar
8	Course Description
9	Course Objectives
10	Course Outcomes
11	Lesson Plan
12	Evaluation Pattern
13	Timetable
14	Unit wise Questions

COLLEGE VISION

To emerge as a premier institution in the field of technical education and research in the state and as a home for holistic development of the students and contribute to the advancement of society and the region.

COLLEGE MISSION

To provide high quality technical education through a creative balance of academic and industry oriented learning; to create an inspiring environment of scholarship and research; to instill high levels of academic and professional discipline; and to establish standards that inculcate ethical and moral values that contribute to growth in career and development of society in general.

VISION OF THE DEPARTMENT

To be a premier Department in the region in the field of Information Technology through academic excellence and research that enable graduates to meet the challenges of industry and society

MISSION OF THE DEPARTMENT

- ❖ To Provide dynamic teaching-learning environment to make the students industry ready and advancement in career;
- ❖ To inculcate professional and leadership quality for better employability and entrepreneurship;
- ❖ To make high quality professional with moral and ethical values suitable for industry and society

PROGRAM EDUCATIONAL OBJECTIVES (PEOs)

PEO1: Solve real world problems through effective professional skills in Information Technology industry and academic research.

PEO2: Analyze and develop applications in Information Technology domain and adapt to changing technology trends with continuous learning.

PEO3: Practice the profession in society with ethical and moral values.

PROGRAM OUTCOMES (POs)

PO1: Engineering Knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO2: Problem Analysis: Identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using the first principles of mathematics, natural sciences, and engineering sciences.

PO3: Design/Development of Solutions: Design solutions for complex engineering problems and system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, society, and environmental considerations.

PO4: Conduct Investigations of Complex Problems: 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.

PO5: Modern Tool Usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO6: The Engineer and Society: 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.

PO7: Environment and Sustainability: Understand the impact of the professional engineering solutions in society and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO8: Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO9: Individual and Team Work: Function effectively as an individual, and as a member or leader in diverse teams, and in multi-disciplinary settings.

PO10: Communication: 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.

PO11: Project Management and Finance: 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 multi-disciplinary environments.

PO12: Life-long Learning: 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.

PROGRAM SPECIFIC OUTCOMES (PSOs)

PSO1: Design and develop software in the area of relevance under realistic constraints.

PSO2: Adapt new and fast emerging technologies in the field of Information Technology.

Website: www.jntuk.edu.in
Email: dap@jntuk.edu.in



Phone: 0884-2300991

Directorate of Academic Planning

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA
KAKINADA-533003, Andhra Pradesh, INDIA

(Established by AP Government Act No. 30 of 2008)

Lr. No. DAP/AC/III Year /B. Tech/B. Pharmacy/2022

Date 14.09.2022

Dr. KVSG Murali Krishna,
M.E, Ph.D.
Director, Academic Planning
JNTUK, Kakinada

To
All the Principals of Affiliated Colleges,
JNTUK, Kakinada.

**Academic Calendar for III Year - B. Tech/B. Pharmacy for the AY 2022-23
(2020-21 Admitted Batch)**

I SEMESTER			
Description	From	To	Weeks
Community Service Project	15.07.2022	30.07.2022	2W
I Unit of Instruction	01.08.2022	24.09.2022	8W
I Mid Examinations	26.09.2022	01.10.2022	1W
II Unit of Instructions	03.10.2022	26.11.2022	8W
II Mid Examinations	28.11.2022	03.12.2022	1W
Preparation & Practicals	05.12.2022	10.12.2022	1W
End Examinations	12.12.2022	25.12.2022	2W
Commencement of II Semester Class Work	02.01.2023		
II SEMESTER			
I Unit of Instructions	02.01.2023	25.02.2023	8W
I Mid Examinations	27.02.2023	04.03.2023	1W
II Unit of Instructions	06.03.2023	29.04.2023	8W
II Mid Examinations	01.05.2023	06.05.2023	1W
Preparation & Practicals	08.05.2023	13.05.2023	1W
End Examinations	15.05.2023	27.05.2023	2W


* As per the APSICHE Guidelines Out of the Total 180 hours of Community Service Project leading to 4 Credits, two weeks will be offline and remaining project work can be done during the III-I semester weekends and holidays. The summer internship can be done in online cum offline during III-I and III-II semesters.


14/9/22

Director,
Academics & Planning, JNTUK
Director
Academic Planning
JNTUK Kakinada

Copy to the Secretary to the Hon'ble Vice Chancellor, JNTUK
Copy to Rector, Registrar, JNTUK
Copy to Director Academic Audit, JNTUK
Copy to Director of Evaluation, JNTUK

Department Academic Calendar

		Department of Information Technology III/IV B.Tech Academic Calendar for 2022-23																																	
2022-23	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M					
Jul 22					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Aug 22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31				
Sep 22				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30		
Oct 22					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Nov 22		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30				
Dec 22					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Jan 23	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31				
Feb 23			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28					
Mar 23			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31		
Apr 23					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
May 23	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31				
Jun 23				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30		

List of Holidays	Oct 9: Maulud Nabi	Mar 22: Ugadhi	Mid exams
July 10: Bakrid	Oct 24: Diwali	Mar 30: Srirama Navami	End Examinations
Aug 9: Moharum	Dec 25: Christmas	Apr 5: Babu Jagjivan Ram Jayanti	Commencement of Class work
Aug 15: Independence day	Jan 14-16: sankranti	Apr 7: Good Friday	Workshops
Aug 31: Ganesh Chaturdi	Jan 26: Republic Day	Apr 14: Ambekar Jayanthi	Department fest Elite
Oct 2: Gandhi Jayanthi	Feb 18: Sivaratri	Jun 29: Bakrid	
Oct 5: Vijayadasami	Mar 8: holi		
			H.O.D.
			Department of IT

Course Description

This course gives an overview of Big Data, i.e. storage, retrieval and processing of big data. In addition, it also focuses on the “technologies”, i.e., the tools/algorithms that are available for storage, processing of Big Data. It also helps a student to perform a variety of “analytics” on different data sets and to arrive at positive conclusions.

Course Objectives

The main objectives of this course are given below:

- To optimize business decisions and create competitive advantage with Big Data analytics
- To learn to analyze the big data using intelligent techniques
- To introduce programming tools PIG & HIVE in Hadoop ecosystem

Course Outcomes

Students are able to

CO No's	COs	Level
CO1	Understand Big Data and its analytics in the real world	L2
CO2	Apply various techniques for mining stream and Hadoop	L3
CO3	Analyze predictive analysis and hadoop concepts	L4
CO4	Design hadoop and Implementation of Big Data Analytics using pig, HIVE and spark to solve data intensive problems and to generate analytics	L5

S.No	Unit	Description	Teaching Aids	CO
1.	I	Introduction to big data: Introduction to Big Data Platform	BB	CO1
2.		Challenges of Conventional Systems	BB	CO1
3.		Intelligent data analysis, Nature of Data,	BB	CO1
4.		Analytic Processes and Tools	BB	CO1
5.		Analysis vs Reporting	BB	CO3
6.	II	Stream Processing: Introduction to Streams Concepts	PPT	CO1
7.				
8.		Stream Data Model and Architecture	BB	CO1
9.		Stream Computing	PPT	CO1
10.		Sampling Data in a Stream	PPT	CO2
11.		Filtering Streams	PPT	CO2
12.		Counting Distinct Elements in a Stream	PPT	CO2
13.		Estimating Moments	PPT	CO1
14.		Counting Oneness in a Window	BB	CO1
15.		Decaying Window	PPT	CO1
16.		Case Studies - Real Time Sentiment Analysis -	BB	CO3

		Stock Market Predictions.		
17.	III	Introduction to Hadoop: History of Hadoop,	BB	CO1
18.		Hadoop Distributed File System	BB	CO1
19.		Components of Hadoop Analysing the Data with Hadoop	BB	CO1
20.		Scaling Out	BB	CO1
21.		Hadoop Streaming	BB/PPT	CO2
22.		Design of HDFS	BB	CO1
23.		Java interfaces to HDFS Basics	BB/PPT	CO1
24.		Developing a Map Reduce Application	BB/PPT	CO2
25.		How Map Reduce Works	PPT	CO1
26.		Anatomy of a Map Reduce Job run,Failures	BB	CO2
27.		Job Scheduling	BB	CO2
28.		Shuffle and Sort	BB	CO3
29.		Task execution	PPT	CO4
30.		Map Reduce Types and Formats	BB	CO1
31.	Map Reduce Features Hadoop environment	BB	CO1	
32.	IV	Frameworks: Applications on Big, HiveQL	BB	CO1
33.		Hive Data processing operators in Pig	BB/PPT	CO4
34.		Data Using Pig and, Hive services	BB	CO4
35.		Querying Data in Hive	BB/PPT	CO4
36.		fundamentals of HBase and ZooKeeper	BB/PPT	CO1
37.	V	Predictive Analytics	BB	CO3
38.		Simple linear regression	BB	CO3
39.		Multiple linear regression	BB/PPT	CO3
40.		Interpretation of regression coefficients	BB/PPT	CO3
41.		Visualizations	BB/PPT	CO1

42.		Visual data analysis techniques	BB	CO3
43.		interaction techniques, , Systems and application	BB	CO3
Total Classes			43	10

Reference Books: Links:

1. Hadoop: <http://hadoop.apache.org/>
2. Hive: <https://cwiki.apache.org/confluence/display/Hive/Home>
3. Piglatin: <http://pig.apache.org/docs/r0.7.0/tutorial.html>

Text Books:
1. Tom White, “Hadoop: The Definitive Guide”, Third Edition, O’reilly Media, Fourth Edition, 2015.
2. Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos, “Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data”, McGrawHill Publishing, 2012.
3. Anand Rajaraman and Jeffrey David Ullman, “Mining of Massive Datasets”, CUP, 2012
References Books:
1. Bill Franks, “Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics”, John Wiley& sons, 2012.
2. Paul Zikopoulos, DirkdeRoos, Krishnan Parasuraman, Thomas Deutsch, James Giles, David Corrigan, “Harness the Power of Big Data: The IBM Big Data Platform”, Tata McGraw Hill Publications, 2012.
3. Arshdeep Bahga and Vijay Madisetti, “Big Data Science & Analytics: A Hands On Approach “, VPT, 2016.
4. Bart Baesens, “Analytics in a Big Data World: The Essential Guide to Data Science and its Applications (WILEY Big Data Series)”, John Wiley & Sons, 2014. Software

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	2	2												
CO2	2	2	2									2		2
CO3	2	2	2											2
CO4	2	2	2									2		2
Avg	2	2	2									2		2

Evaluation Pattern

S. No	Components	Internal	External	Total
1	Theory	25	75	100
2	Engineering Graphics/Design/Drawing	25	75	100
3	Practical	20	30	50
4	Mini Project/Internship/Industrial Training/ Skill Development programmes/Research Project	-	50	50
5	Project Work – Part I	20	30	50
5	Project Work – Part II	60	90	150

Marks Range Theory (Max – 100)	Marks Range Lab (Max – 75)	Letter Grade	Level	Grade Point
≥ 90	≥ 67	O	Outstanding	10
≥80 to <90	≥60 to <67	S	Excellent	9
≥70 to <80	≥52 to <60	A	Very Good	8
≥60 to <70	≥45 to <52	B	Good	7
≥50 to <60	≥37 to <45	C	Fair	6
≥40 to <50	≥30 to <37	D	Satisfactory	5
<40	<30	F	Fail	0
			Absent	0

Day/Ti me	09.00- 09.50	09.50- 10.40	11.00- 11.50	11.50- 12.40	01.40- 02.30	02.30- 03.20	03.20- 04.10	04.10- 05.00
Mon		BDA(B)				BDA(A)		
Tue	BDA(A)		BDA(B)		BDA LAB (B)			
Wed	BDA(B)					BDA(A)		
Thu				BDA(A)	BDA LAB (A)			
Fri		BDA(A)		BDA(B)				
Sat				BDA(B)				

COMPUTER NETWORKS UNIT WISE Important Questions

Unit-1

1. Explain the importance of bigdata
2. Describe patterns for bigdata development
3. compare the behavior of data in warehouse and data in hadoop
4. Explain Challenges of Conventional Systems

Unit-2

1. Explain Stream Data Model
2. Discuss Stream Computing In detail
3. Explain Counting Distinct Elements in a stream
4. Anlayze Real time Sentiment Analysis
5. Analyze stock market prediction

Unit-3

1. Compare sql databases with hadoop
2. write a counting words program in hadoop
3. Explain the building blocks of hadoop
4. I llustrate job traker and task tracker with a program
5. Identify name node,data node and secondary name node in a hadoop program.
6. Explain Hadoop compents
7. Explain Hadoop architecture

Unit-4

1. Explain pig
2. Analyze Big data using pig and hive
3. Explain Fundamentals ofHBase and ZooKeeper
4. Explain Hive services

Unit-5

1. Compare simple and multiple linear regression
2. Discuss interpretation of regression
3. Discuss Visual Data Analysis techniques
4. Explain systems and application