

Institute/Department	UNIVERSITY INSTITUTE OF ENGINEERING (UIE)	Program	Bachelor of Engineering - Computer Science & Engineering (CS201)
Master Subject Coordinator Name:	Urvashi	Master Subject Coordinator E-Code:	E12312
Course Name	Big Data Analytics	Course Code	21CSH-471

Lecture	Tutorial	Practical	Self Study	Credit	Subject Type
3	0	2	0	4.00	T

Course Type	Course Category	Mode of Assessment	Mode of Delivery
Major Elective	Graded (GR)	Hybrid	Hybrid (HYB)

Mission of the Department	M1: To provide practical knowledge using state-of-the-art technological support for the experiential learning of our students. M2: To provide industry recommended curriculum and transparent assessment for quality learning experiences. M3: To create global linkages for interdisciplinary collaborative learning and research. M4: To nurture advanced learning platform for research and innovation for students' profound future growth. M5: To inculcate leadership qualities and strong ethical values through value based education.
Vision of the Department	To be recognized as a leading Computer Science and Engineering department through effective teaching practices and excellence in research and innovation for creating competent professionals with ethics, values and entrepreneurial attitude to deliver service to society and to meet the current industry standards at the global level.

Program Educational Objectives(PEOs)

PEO1	Engage in successful careers in industry, academia, and public service, by applying the acquired knowledge of Science, Mathematics and Engineering, providing technical leadership for their business, profession and community.
PEO2	Establish themselves as entrepreneur, work in research and development organization and pursue higher education.
PEO3	Exhibit commitment and engage in lifelong learning for enhancing their professional and personal capabilities.

Program Specific OutComes(PSOs)

PSO1	Exhibit attitude for continuous learning and deliver efficient solutions for emerging challenges in the computation domain.
PSO2	Apply standard software engineering principles to develop viable solutions for Information Technology Enabled Services (ITES).

Program OutComes(POs)

PO1	Engineering knowledge: Apply the knowledge of Mathematics, Science, Engineering fundamentals and computer science fundamental and strategies which have the solution of complex computer science engineering problems.
PO2	Problem analysis: Identify, formulate, research literature, and analyze complex computer science engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
PO3	Design/development of solutions: Design solutions for complex database and software engineering problems and design system components or processes that meet the specified needs with appropriate considerations for the public health and safety, and the cultural, societal, and environmental considerations.
PO4	Conduct investigations of complex problems: Use research-based knowledge and research methods including design of software engineering & networking based 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 Computer science engineering and IT tools including prediction and modeling to complex database or software engineering activities with an understanding of the limitations.
PO6	The engineer and society: Apply reasoning informed by the contextual knowledge to assess Social, health, safety, legal and cultural issues and the consequent responsibilities relevant to the Professional Computer Science & Engineering practice.
PO7	Environment and sustainability: Understand the impact of the professional computer science and engineering solutions in social and environmental contexts, and demonstrate the knowledge of, and need for sustainable development goals.

PO8	Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of computer science engineering practice
PO9	Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
PO10	Communication: Communicate effectively on complex computer science engineering activities with the engineering community like CSI 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 computer science 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.
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 to technological change.

Text Books					
Sr No	Title of the Book	Author Name	Volume/Edition	Publish Hours	Years
1	Big Data Analytics with Spark	Mohammed Guller	1st edition	Apress	2015
2	Machine Learning	Tom Mitchell	3rd Edition	McGraw Hill	1997
3	Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business	Michael Minelli, Michehe Chambers	1st Edition	Ambiga Dhiraj, Wiely CIO Series	2013
4	Big Data Analytics: Disruptive Technologies for Changing the Game	Arvind Sathi	1st Edition	IBM Corporation	2012

Reference Books					
Sr No	Title of the Book	Author Name	Volume/Edition	Publish Hours	Years
1	Understanding Big data	Chris Eaton, Dirk deroos et al	2nd Edition	McGraw Hill	2012
2	Big Data Analytics with R and Hadoop	Vignesh Prajapati	1st edition	Packet Publishing	2013
3	Big Data and Business Analytics	JyLiebowitz	1st edition	CRC press	2013

Course OutCome	
SrNo	OutCome
CO1	Understand key concepts of Big Data, including the characteristics.
CO2	Master Big Data Architecture and Tools
CO3	Explore the Hadoop Ecosystem and Data Processing Models
CO4	Develop Data Science Skills and Tools
CO5	Implement Real-Time Data Analytics and Visualization

Lecture Plan Preview-Theory						
Unit No	LectureNo	ChapterName	Topic	Text/ Reference Books	Pedagogical Tool**	Mapped with CO Numer (s)
1	1	Understanding Big Data and the 5 V's	Introduction to Big Data and Its Characteristics	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO1

1	2	Understanding Big Data and the 5 V's	The 5 V's of Big Data: Understanding Volume, Velocity, and Variety	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO1
1	3	Understanding Big Data and the 5 V's	Veracity and Value: Trust and Insights in Big Data	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO2
1	4	Understanding Big Data and the 5 V's	Challenges and Opportunities in Big Data	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO1
1	5	Understanding Big Data and the 5 V's	Real-World Applications of Big Data	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	Case Study,PPT	CO2
1	6	Big Data Architecture	Understanding Big Data Architecture: Key Components and Layers	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	Case Study,PPT	CO2
1	7	Big Data Architecture	Data Ingestion and Storage in Big Data Architecture	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	Case Study,PPT	CO2
1	8	Big Data Architecture	Data Processing and Visualization in Big Data Architecture	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	Case Study,PPT	CO2
1	9	Big Data Architecture	Streaming Data in Big Data: Frameworks and Use Cases	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	Case Study,PPT	CO2
1	10	Big Data Architecture	Architectural Patterns in Big Data: Lambda, Kappa, and Hybrid Models	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO1
1	11	The Hadoop Ecosystem	Understanding the Hadoop Ecosystem: An Overview	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3
1	12	The Hadoop Ecosystem	HDFS (Hadoop Distributed File System): Architecture and Functionality	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3

1	13	The Hadoop Ecosystem	MapReduce Programming Model: Workflow and Applications	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3
1	14	The Hadoop Ecosystem	YARN: Resource Management in Hadoop	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3
1	15	The Hadoop Ecosystem	Advanced Tools in the Hadoop Ecosystem and Data Processing	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3
2	16	Big Data Frameworks	Big Data Frameworks: Hadoop and Apache Spark – Concepts and Comparison	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3
2	17	Big Data Frameworks	Introduction to NoSQL Databases: MongoDB, Cassandra, and HBase	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO2
2	18	Big Data Frameworks	Big Data Visualization Tools: Tableau, Power BI, and Apache Zeppelin	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO2
2	19	Big Data Frameworks	Real-Time Big Data Processing: Apache Storm and Flink	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Understanding Big data	PPT	CO3
2	20	Big Data Frameworks	Emerging Trends in Big Data Technologies	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Understanding Big data	PPT	CO3
2	21	Big SQL and NO SQL Databases	SQL vs. NoSQL: Differences, Features, and Use Cases	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3
2	22	Big SQL and NO SQL Databases	Introduction to Big SQL: Bridging SQL and Big Data	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,T-Machine Learning,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3
2	23	Big SQL and NO SQL Databases	Query Optimization Techniques in Big SQL	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,T-Machine Learning,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3

2	24	Big SQL and NO SQL Databases	Exploring NoSQL Database Types and Applications	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,T-Machine Learning,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3
2	25	Big SQL and NO SQL Databases	Advantages and Limitations of Big SQL and NoSQL Databases	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3
2	26	AI in Big Data	Introduction to IBM Watson: AI Capabilities and Big Data Synergy	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO4
2	27	AI in Big Data	Exploring Watson Services: Discovery, Studio, and Assistant	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO4
2	28	AI in Big Data	Integrating IBM Watson with Big Data Tools and Workflows	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,T-Machine Learning,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3
2	29	AI in Big Data	AI and Machine Learning Applications in Big Data: Natural Language Processing (NLP) and Beyond	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,T-Machine Learning,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO4
2	30	AI in Big Data	Sentiment Analysis and Predictive Analytics with Watson AI	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3
3	31	The Iterative Nature of Data Science Projects	Introduction to Data Science Projects: Stages and Lifecycle	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3
3	32	The Iterative Nature of Data Science Projects	Introduction to Data Science Projects: Stages and Lifecycle	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3
3	33	The Iterative Nature of Data Science Projects	Refinement and Deployment in Data Science Projects	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,T-Machine Learning,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO4
3	34	The Iterative Nature of Data Science Projects	Importance of Iteration: Continuous Improvement and Error Correction	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,T-Machine Learning,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO4

3	35	The Iterative Nature of Data Science Projects	Tools Supporting Iteration in Data Science: Notebooks, Version Control, and CI/CD	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,T-Machine Learning,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3
3	36	Notebooks in Data Science	Introduction to Data Science Notebooks: Features and Workflow	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO2
3	37	Notebooks in Data Science	Programming Languages for Data Science: Python and R	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,T-Machine Learning,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3
3	38	Notebooks in Data Science	Exploring Notebook Functionality: Code Cells, Markdown, and Widgets	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO2
3	39	Notebooks in Data Science	Extending Notebook Capabilities: Extensions and Tools	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO2
3	40	Notebooks in Data Science	Integrating Notebooks with Git and Data Science Workflows	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,T-Machine Learning,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3
3	41	Notebooks and Data Science tools in Big Data	Major Data Science Notebooks: Jupyter Notebook, Google Colab, and Zeppelin	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,T-Machine Learning,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO4
3	42	Notebooks and Data Science tools in Big Data	Getting Started with Jupyter Notebook: Installation and Basics	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,T-Machine Learning,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO4
3	43	Notebooks and Data Science tools in Big Data	Advanced Features of Jupyter: Extensions, Widgets, and Multilingual Support	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,T-Machine Learning,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO4
3	44	Notebooks and Data Science tools in Big Data	Introduction to Tableau: Building Dashboards for Data Insights	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,T-Machine Learning,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO2

3	45	Notebooks and Data Science tools in Big Data	Collaboration and Presentation Tools for Data Insights	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic	PPT	CO3
---	----	--	--	---	-----	-----

Lecture Plan Preview-Practical

Unit No	ExperimentNo	Experiment Name	Text/ Reference Books	Pedagogical Tool**	Mapped with CO Numer(s)
1	1	Write a Program to Analyze a dataset to identify the 5 V's of Big Data.	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic	PPT	CO1
1	2	Write a Program to store and retrieve data in Hadoop HDFS.	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic	PPT	CO2
1	3	Write a program to use Spark and Apache Hive to query structured data stored in HDFS.	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic	PPT	CO3
1	4	Write a program to create an interactive dashboard to visualize insights from a dataset.	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic	PPT	CO3
2	5	Write a program to understand the working of Tableau and PowerBI.	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic	PPT	CO5
2	6	Write a program to evaluate the performance of Big SQL and No SQL database for the same dataset.	,T-Big Data Analytics with Spark,T-Big Data Analytics: Disruptive,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic	PPT	CO1
2	7	Write a program to evaluate the performance of Watson AI with Big Data.	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO2
3	8	Write a program for building and iterating a Predictive Model: A Hands-On Approach to the Data Science Lifecycle	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic	PPT	CO3
3	9	Write a program to explore the capabilities of Jupyter Notebook for a sample data	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic,R-Understanding Big data	PPT	CO3
3	10	Write a program to compare Google Colab and Apache Zeppelin for a sample data analysis project	,T-Big Data Analytics with Spark,T-Big Data, Big Analytics: Emerg,R-Big Data Analytics with R and ,R-Big Data and Business Analytic	PPT	CO4

Assessment Model

Sr No	Assessment Name	Exam Name	Max Marks
1	Hybrid Course All	Practical Evaluations	40
2	Hybrid Course All	End Term Hybrid Theory	60

3	Hybrid Course All	Attendance Marks	2
4	Hybrid Course All	Surprise Test	12
5	Hybrid Course All	Practical MST	10
6	Hybrid Course All	Practical Worksheet/Projects 1	30
7	Hybrid Course All	Practical Worksheet/Projects 2	30
8	Hybrid Course All	Practical Worksheet/Projects 3	30
9	Hybrid Course All	Practical Worksheet/Projects 4	30
10	Hybrid Course All	Practical Worksheet/Projects 5	30
11	Hybrid Course All	Practical Worksheet/Projects 6	30
12	Hybrid Course All	Practical Worksheet/Projects 7	30
13	Hybrid Course All	Practical Worksheet/Projects 8	30
14	Hybrid Course All	Practical Worksheet/Projects 9	30
15	Hybrid Course All	Practical Worksheet/Projects 10	30
16	Hybrid Course All	Quiz	4
17	Hybrid Course All	Assignment/PBL	10
18	Hybrid Course All	MST-1 Hybrid	20
19	Hybrid Course All	MST-2 Hybrid	20