

Branch Name:	Computer Engineering/Information Technology
Program Code:	ET103 / ET105
Course Title:	Big Data Analytics [Specialization in Data Science]
Course Code:	2ET1030508T
Pre-requisite Course:	Should have knowledge of one Programming Language (Java preferably), Practice of SQL (queries and sub queries), exposure to Linux Environment.

Course Objective:

- 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. To understand working of internet and websites.

Teaching and Examination Scheme:

Credit	Teaching Scheme Hours per week				Evaluation Scheme (Marks)								Total (Marks)
					Theory (Marks)				Practical (Marks)				
	L	T	P	Total	SEE		CCA		SEE		CCA		
					Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	
3	3	-	-	3	24	60	16	40	-	-	-	-	100

Course Contents:

Sr. No	Topic	Total Hours	Weightage (%)
1	ESSENTIALS OF BIG DATA AND ANALYTICS: Data, Characteristics of data and Types of digital data, Sources of data, Working with unstructured data, Evolution and Definition of big data, Characteristics and Need of big data, Challenges of big data; Overview of business intelligence, Data science and Analytics, Meaning and Characteristics of big data analytics, Need of big data analytics, Classification of analytics, Challenges to big data analytics, Importance of big data analytics, Basic terminologies in big data environment.	09	20
2	HADOOP: Introducing Hadoop, Need of Hadoop, limitations of RDBMS, RDBMS versus Hadoop, Distributed computing challenges, History of Hadoop, Hadoop overview, use case of Hadoop, Hadoop distributors, HDFS (Hadoop Distributed File System), Processing data with Hadoop, managing resources and applications with Hadoop YARN (Yet another Resource Negotiator), Interacting with Hadoop Ecosystem.	09	20
3	MAPREDUCE PROGRAMMING: Introduction, Mapper, Reducer, Combiner, Partitioner, Searching, Sorting, Compression, Real time applications using MapReduce, Data serialization and working with common serialization formats, big data serialization formats	09	20
4	DATA ANALYTICS: Predictive Analytics (Regression, Decision Tree, Neural Networks) - Descriptive Analytics (Association Rules, Sequence Rules), Survival Analysis (Survival Analysis Measurements, Kaplan Meir Analysis, Parametric Survival Analysis) - Social Network Analytics (Social Network Learning-	06	15

	Relational Neighbor Classification).		
5	BIG DATA PRIVACY AND APPLICATIONS : Data Masking – Privately Identified Information (PII) -Privacy preservation in Big Data- Popular Big Data Techniques and tools- Map Reduce paradigm and the Hadoop system- Applications- Social Media Analytics- Recommender Systems- Fraud Detection.	07	15
6	DATA PRE-PROCESSING : Data pre-processing overview-Sampling- Missing Values -Outlier Detection and Treatment - Standardizing Data- Categorization - Weights of Evidence Coding -Variable Selection and Segmentation.	05	10

Reference Books:

- Bart Baesens," Analytics in a Big Data World: The Essential Guide to Data Science and its Applications", John Wiley & Sons, 2014
- Min Chen, Shiwen Mao, Yin Zhang, Victor CM Leung, Big Data: Related Technologies, Challenges and Future Prospects, Springer, 2014.
- Michael Minnelli, Michele Chambers, Ambiga Dhiraj , “Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends”, John Wiley & Sons, 2013
- Raj, Pethuru, “Handbook of Research on Cloud Infrastructures for Big Data Analytics”, IGI Global.

E-Resources (If any):

- "Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking" by Foster Provost and Tom Fawcett (2013)
- "Hadoop: The Definitive Guide" by Tom White (2015)
- "Big Data: Principles and Best Practices of Scalable Realtime Data Systems" by Nathan Marz and James Warren (2015)
- "Data Mining: Concepts and Techniques" by Jiawei Han, et al. (2016)
- "Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph" by David Loshin (2016)
- "Hands-On Big Data Analytics with Hadoop 3" by Hands-On Labs (2019)
- "Big Data and the Internet of Things" by John Bates (2020)

Course Outcome:

Co No.	Course Outcomes	Cognitive Level
CO 1	Identify the various sources of Big Data	Understand
CO 2	Analyze the Big Data framework like Hadoop and NOSQL to efficiently store and process Big Data to generate analytics	Understand Apply
CO 3	Design of Algorithms to solve Data Intensive Problems using Map Reduce Paradigm	Apply Remember
CO 4	Design methodologies to extract data from structured and un-structured data for analytics	Create
CO 5	Design and Implementation of Big Data Analytics using pig and spark to solve data intensive problems and to generate analytics	Create
CO 6	Understand Big Data and its analytics in the real world	Understand

Mapping of COs with POs & PSOs

Course Learning Outcomes	Program Outcomes (POs)												Program Outcomes (POs)	
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO 1	H	H	H		H							M	M	M
CO 2	H	H	H		H							M	M	M
CO 3						H		M						
CO 4	M	H	M	M		M	L	M						
CO 5	H	H	H		H							M	L	M
CO 6	H	H	H	M	H							M	L	M