

Course Code MCA2004	Big Data Analytics	Course Type LP	Credits 3
Prerequisite:	Data mining and Analysis		
Objectives			
<ul style="list-style-type: none">To introduce the fundamentals of various big data analysis techniquesTo analyze the big data analytic techniques for useful business applicationsTo perform map-reduce analytics using Hadoop and related tools.			
Expected Outcomes			
Students who complete this course will be able to <ul style="list-style-type: none">Analyze and interpret huge volume of data, create statistical models, identify insights that can lead to actionable resultsPerform map-reduce analytics using HadoopImplement using software tools such as R and Hadoop for big data analytics.			
Student Outcomes (SO):		a, b, i	
Unit No.	Unit Description	No of Hours	SO
1	Introduction to Big Data Analytics: Big Data Overview, State of practice in analytics, Role of Data Scientists, Examples of Big Data Analytics, Data Analytics Lifecycle.	5	a
2	Hadoop: Components of Hadoop, Analyzing Big data with Hadoop, Design of HDFS, Developing a Map reduce Application. Map Reduce: Distributed File System (DFS), Map Reduce, Algorithms using Map Reduce, Communication cost Model, Graph Model for Map Reduce Problem	6	a, b
3	Hadoop Environment: Setting up a Hadoop Cluster, Hadoop Configuration, Security in Hadoop, Administering Hadoop, Hadoop Benchmarks, Hadoop in the cloud.	5	a, b
4	Big Data Analytics Methods using R: Introduction to R-Attributes, R Graphical user interfaces, Data import and export, attribute and Data Types, Descriptive Statistics, Exploratory Data Analysis. Statistical methods for evaluation: Hypothesis Testing, Difference of Means, Wilcoxon Rank-Sum Test, Type I and Type II errors, power and sample size, ANOVA	9	b, i
5	Advanced Analytics - technologies and tools: Analytics for unstructured data, The Hadoop ecosystem – pig – Hive- HBase- Mahout- Introduction to NoSQL Guest Lecture on Contemporary Topics in big data analytics	5+2	b, i
	Total Lecture:	32	
Mode of Teaching and Learning: <i>Flipped Class Room, Activity Based Teaching/Learning, Digital/Computer based models, wherever possible to augment lecture for practice/tutorial and minimum 2 hours lectures by industry experts on contemporary topics.</i>			
Mode of Evaluation and Assessment: <i>The assessment and evaluation components may consist of unannounced open book</i>			

examinations, quizzes, student's portfolio generation and assessment, and any other innovative assessment practices followed by faculty, in addition to the Continuous Assessment Tests and Term End Examination.

Text Book(s):

1. Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data by EMC Education Services, John Wiley & Sons, 2015 publishing 2015.

Reference Book(s):

1. Anand Raja Raman and Jeffrey David Ullman, "Mining of Massive Datasets", Cambridge University Press, 2012.
2. Tom White "Hadoop: The Definitive Guide" Third Edition, O'reilly Media, 2012.
3. Jiawei Han, Micheline Kamber "Data Mining Concepts and Techniques", Third Edition, Elsevier, Reprinted 2011.

No.	Indicative List of Experiments	Applications Perspective & Tools	SO - i
1	Sports Analytics System: Sports teams are using data for tracking ticket sales and even for tracking team strategies.	Hadoop	
2	Health care Data Analytics: Hospitals are analyzing medical data and patient records to predict those patients that are likely to seek readmission within a few months of discharge. The hospital can then intervene in hopes of preventing another costly hospital stay.	Hadoop	
3	Marketing Data Analytics: Advertising and marketing agencies are tracking social media to understand responsiveness to campaigns, promotions, and other advertising mediums.	R & RStudio	
4	Customer Behavior Analysis: Consumer product companies and retail organizations are monitoring social media like Facebook and Twitter to get an unprecedented view into customer behavior, preferences, and product perception	R & RStudio	
5	Predictive analytics on Insurance, banking dataset and Recommendation system, Trend analytics and forecasting.	R & RStudio	
6	Counting No. of Words in the documents - Genome sequence analysis	Mapreduce	
7	Apply appropriate analytic techniques and tools to analyze big data, create statistical models, and identify insights that can lead to actionable results.	R & RStudio	
8	Select appropriate data visualizations to clearly communicate analytic insights to business sponsors and analytic audiences.	R & RStudio	
9	Use tools such as: R and RStudio, MapReduce/Hadoop, in-Database analytics, Window and MADlib functions	R & RStudio	

Recommendation by the Board of Studies on	June 25, 2018
Approval by Academic council on	July 18, 2018
Compiled by	Dr S Raju and Dr Pattabiraman V