Course Code	Course Name	Teaching Scheme (Hrs/week)			Credits Assigned			
		L	**T**	**P**	**L**	**T**	**P**	**Total**
ITC802	Big Data Analytics	**4**	**-**	**-**	**4**	**-**	**-**	**4**
		Examination Scheme						
		ISE	**MSE**		**ESE**			
		10	**30**		**100 (60% Weightage)**			

Pre-requisite Course Codes		
After successful completion of the course, student will be able to:		
Course Outcomes	CO1	Analyze the key issues in big data management and its associated applications in intelligent business and scientific computing
	CO2	Experiment with fundamental enabling techniques and scalable algorithms like Hadoop, Map Reduce and NO SQL in big data analytics
	CO3	Interpret business models and scientific computing paradigms
	CO4	Apply software tools for big data analytics
	CO5	Apply big data analytics in various applications like recommender systems, social media applications etc

Module No.	Topics	Ref.	Hrs.
1	**Introduction to Big data** Introduction to Big Data, Big Data characteristics, types of Big Data, Traditional vs. Big Data business approach, Case Study of Big Data Solutions.	1	03
2	**Introduction to Hadoop** What is Hadoop? Core Hadoop Components ; Hadoop Ecosystem; PhysicalArchitecture; Hadoop limitations.	1,2	02
3	**NoSQL** 1. What is NoSQL? NoSQL business drivers ;NoSQL case studies; 2. NoSQL data architecture patterns: Key-value stores, Graph stores, Column family(Big table)stores, Document stores, Variations of NoSQL architectural patterns; Using No SQL to manage big data: What is a big data NoSQL solution? Understanding the types of big data problems; Analyzing	3	04

	big data with a shared-nothing architecture; Choosing distribution models: master-slave versus peer-to-peer; Four ways that NoSQL systems handle big data problems.		
4	**Map Reduce and the New Software Stack** **Distributed File Systems :** Physical Organization of Compute Nodes, Large-Scale File-System Organization. **Map Reduce:** The Map Tasks, Grouping by Key, The Reduce Tasks, Combiners, Details of Map Reduce Execution, Coping With Node Failures. **Algorithms Using Map Reduce**: Matrix-Vector Multiplication by Map Reduce ,Relational-Algebra Operations, Computing Selections by Map Reduce, Computing Projections by Map Reduce, Union, Intersection, and Difference by Map Reduce, Computing Natural Join by Map Reduce, Grouping and Aggregation by Map Reduce, Matrix Multiplication, Matrix Multiplication with One Map reduce step.	1,4	06
5	**Finding Similar Items** Applications of Near-Neighbor Search, Jaccard Similarity of Sets, Similarity of Documents, Collaborative Filter in gasa Similar-Sets Problem . **Distance Measures:** Definition of a Distance Measure, Euclidean Distances, Jaccard Distance, Cosine Distance, Edit Distance, Hamming Distance.	1,5	03
6	**Mining Data Streams** **The Stream Data Model**: A Data-Stream-Management System, Examples of Stream Sources, Stream Querie, Issues in Stream Processing. **Sampling Data in a Stream**: Obtaining a Representative Sample,The General Sampling Problem, Varying the Sample Size. **Filtering Streams**: The Bloom Filter, Analysis. **Counting Distinct Elements in a Stream** The Count-Distinct Problem, The Flajolet-Martin Algorithm, Combining Estimates, Space Requirements . **Counting One sin a Window**: The Cost of Exact Counts, The Datar-Gionis-Indyk-	5,6	
7	**Link Analysis** Page Rank Definition, Structure of the web, dead ends, Using	5,6,7	05

	Page rank in a search engine, Efficient computation of Page Rank: Page Rank Iteration Using Map Reduce, Use of Combiners to Consolidate the Result Vector. Topic sensitive Page Rank, link Spam, Hubs and Authorities.		
8	**Frequent Item sets** Handling Larger Data sets in Main Memory Algorithm of Park, Chen and Yu, The Multi stage Algorithm, The Multihash Algorithm. **The SON Algorithm and Map Reduce** **Counting Frequent Items in a Stream** Sampling methods for streams, frequent item sets in Decaying window	5,6,7	05
9	**Clustering** CURE Algorithm, Stream-Computing, A Stream-Clustering Algorithm, Initializing &Merging Buckets, Answering Queries	5,6,7	05
10	**Recommendation Systems** A Model for RecommendationSystems,Content-BasedRecommendations,Collaborative Filtering	6,7	04
11	**Mining Social- Network Graphs** Social Networks as Graphs, Clustering of Social-Network Graphs, Direct Discovery of Communities, SimRank, Counting triangles	6,7	05
	Total hours of instructions		48

References:

1. Anand Raja Raman and Jeff Ullman“ *Mining of Massive Datasets*”, Cambridge University Press,
2. Alex Holmes “*Hadoop in Practice”,* Manning Press, Dreamtech Press.
3. Dan McCreary and Ann Kelly“ *Making Sense of No SQL”– A guide for managers and the rest of us*, Manning Press
4. Bill Franks**, “***Taming The Big Data Tidal Wave: Finding Opportunities In Huge*
5. *Data Streams With Advanced Analytics*”, Wiley
6. Judith Hurwitz, Alan Nugent, Dr. Fern Halper,Marcia Kaufman,“ *Big Data for Dummies***”,** Wiley India
7. Michael Minelli, Michele Chambers, Ambiga Dhiraj, “*Big Data Big Analytics: Emerging Business Intelligence And Analytic Trends For Today's Businesses ”,*Wiley India.
8. Paul Zikopoulos, Chris Eaton**, “***Understanding Big Data: Analytics for Enterprise*
9. *Class Hadoop and Streaming Data’***,** Mc Graw Hill Education.