Course Code	Course/Subject Name	Credits
ADC702	**Big Data Analytics**	**3**

Course Outcomes:	
1	Understand the key issues in big data management and its associated applications for business decisions and strategy.
2	Develop problem solving and critical thinking skills in fundamental enabling techniques like Hadoop, Map reduce and NoSQL in big data analytics.
3	Collect, manage, store, query and analyze various forms of Big Data.
4	Interpret business models and scientific computing paradigms, and apply software tools for big data analytics.
5	Adapt adequate perspectives of big data analytics in various applications like recommender systems, social media applications etc.
6	Solve Complex real world problems in various applications like recommender systems, social media applications, health and medical systems, etc.

Prerequisites: Database Management System.

Module		Content	Hrs.
0		**Prerequisites:** Data Mining, database Systems, Algorithms	
1		**Introduction to Big Data**	**4**
	1.1	Introduction to Big Data, Big Data characteristics, types of Big Data, Traditional vs. Big Data business approach, Big Data Challenges, Examples of Big Data in Real Life, Big Data Applications { **5 mark questions can come from this chapter}**	
2		**Introduction to Big Data Frameworks: Hadoop, NOSQL**	**7**
	2.1	What is Hadoop? Core Hadoop Components; Hadoop Ecosystem; **Overview of**: Apache Spark, Pig, Hive, Hbase, Sqoop **Introduction to NoSQL,** CAP Theorem, BASE characteristics for Databases; NoSQL data architecture patterns: Key-value stores, Graph stores, Column family (Bigtable) stores, Document stores **{just overview of NoSQL and description of the patterns}**	
3		**MapReduce Paradigm**	**5**
	3.1	**Introduction to Map Reduce:** The Map Tasks, Grouping by Key, The Reduce Tasks, Combiners, Partitioners **Algorithms Using MapReduce:** ● Matrix-Vector Multiplication ● Relational-Algebra Operations - Computing Selections, Projections, Union, Intersection, and Difference ● Database operations - Computing Natural Join, Group By and Aggregation ● Matrix Multiplication with two and One MapReduce Steps. **Illustrating benefits of MapReduce** Real life examples of databases and applications	
4		**Mining Big Data Streams**	**9**
	4.1	The Stream Data Model: A Data Stream Management System, Examples of Stream Sources, Stream Queries, Issues in Stream Processing. **1.** Sampling Data in a Stream : Sampling Techniques. (**Reservoir, Biased and Concise)** 2. Filtering Streams: The Bloom Filter; 3. Counting Distinct Elements in a Stream: The Flajolet-Martin Algorithm 4. Counting Ones in a Window: The Datar-Gionis-Indyk- Motwani (DGIM) Algorithm **{ problems can be asked for either Bloom's filter, FM algorithm DGIM Algorithm**}	
5		**Big Data Algorithms**	**6**
	5.1	**Frequent Itemset Mining** : Overview of Market basket analysis, using main memory to count frequent itemsets, Cost of counting pairs. Apriori. Handling Larger Datasets in Main Memory Basic Algorithm of Park, Chen, and Yu. The SON Algorithm and MapReduce. **{PCY algorithm can be asked with numericals, memory diagrams}** **Finding Similar items:** Distance Measures: Euclidean Distances, Jaccard Distance, Cosine Distance, Edit Distance, Hamming Distance. Jaccard Similarity of Sets, Similarity of Documents, Shingling of documents **{**problems on distance measures} **Clustering Algorithms:** CURE Algorithm.	

6	**6.1**	**Big Data Analytics Applications** **Link Analysis :** Early Search Engines, Spam PageRank, Definition, Structure of the web, dead ends, Spider traps, Using Page rank in a search engine, Efficient computation of Page Rank using matrices. MapReduce and PageRank, link Spam and spam Farm, HITS Algorithm. **Mining Social- Network Graphs** : Social Networks as Graphs, Types , Clustering of Social Network Graphs, Girvan newman method, Direct Discovery of Communities CPM method **{problems for Page Rank, HITS, Girvan Newmann, CPM can be asked, Link farm spam farm with figures, }** **Recommendation Engine**s: A Model for Recommendation Systems, Content-Based Recommendations, Collaborative Filtering**.{just overview with applications can be asked}**	**8**

End Semester Theory Examination:	
1	Question paper will be of 60 marks
2	Question paper will have a total of five questions
3	All questions have equal weightage and carry 20 marks each
4	Any three questions out of five needs to be solved.