Course Code	Course Name	Credit
CSC702	Big Data Analysis	03

Prerequisite: Database, Data mining.	
Course Objectives: The course aims:	
1	To provide an overview of the big data platforms, its use cases and Hadoop ecosystem.
2	To introduce programming skills to build simple solutions using big data technologies such as MapReduce, Scripting for No SQL and R
3	To learn the fundamental techniques and principles in achieving big data analytics with scalability and streaming capability.
4	To enable students to have skills that will help them to solve complex real-world problems for decision support.
Course Outcomes:	
1	Understand the building blocks of Big Data Analytics.
2	Apply fundamental enabling techniques like Hadoop and MapReduce in solving real world problems.
3	Understand different NoSQL systems and how it handles big data.
4	Apply advanced techniques for emerging applications like stream analytics.
5	Achieve adequate perspectives of big data analytics in various applications like recommender systems, social media applications, etc.
6	Apply statistical computing techniques and graphics for analyzing big data.

Module		Detailed Content	Hours
1		**Introduction to Big Data and Hadoop**	2
	1.1	Introduction to Big Data - Big Data characteristics and Types of Big Data	
	1.2	Traditional vs. Big Data business approach	
	1.3	Case Study of Big Data Solutions	
	1.4	Concept of Hadoop, Core Hadoop Components; Hadoop Ecosystem	
2		**Hadoop HDFS and MapReduce**	8
	2.1	Distributed File Systems: Physical Organization of Compute Nodes, Large-Scale File-System Organization.	
	2.2	MapReduce: The Map Tasks, Grouping by Key, The Reduce Tasks, Combiners, Details of MapReduce Execution, Coping With Node Failures.	
	2.3	Algorithms Using MapReduce: Matrix-Vector Multiplication by MapReduce, Relational-Algebra Operations, Computing Selections by MapReduce, Computing Projections by MapReduce, Union ,Intersection, and Difference by MapReduce	

	2.4	Hadoop Limitations	
3		**NoSQL**	10
	3.1	Introduction to NoSQL, NoSQL Business Drivers	
	3.2	NoSQL Data Architecture Patterns: Key-value stores, Graph stores, Column family (Bigtable)stores, Document stores, Variations of NoSQL architectural patterns, NoSQL Case Study	
	3.3	NoSQL solution for big data, Understanding the types of big data problems; Analyzing big data with a shared-nothing architecture; Choosing distribution models: master-slave versus peer-to-peer; NoSQL systems to handle big data problems.	
4		**Mining Data Streams**	11
	4.1	The Stream Data Model: A Data-Stream-Management System, Examples of Stream Sources, Stream Queries, Issues in Stream Processing.	
	4.2	Sampling Data techniques in a Stream	
	4.3	Filtering Streams: Bloom Filter with Analysis.	
	4.4	Counting Distinct Elements in a Stream,Count-Distinct Problem, Flajolet-Martin Algorithm, Combining Estimates, Space Requirements	
	4.5	Counting Ones in a Window: The Cost of Exact Counts,The Datar-Gionis-Indyk-Motwani Algorithm, Query Answering in theDGIM Algorithm, Decaying Windows.	
5		**Real-Time Big Data Models**	4
	5.1	A Model for Recommendation Systems, Content-Based Recommendations, Collaborative Filtering	
	5.2	Case Study: Product Recommendation	
	5.3	Social Networks as Graphs, Clustering of Social-Network Graphs, Direct Discovery of Communities in a social graph	
6		**Data Analytics with R**	4
	6.1	Exploring Basic features of R, Exploring RGUI, Exploring RStudio, Handling Basic Expressions in R, Variables in R, Working with Vectors, Storing and Calculating Values in R, Creating and using Objects, Interacting with users, Handling data in R workspace, Executing Scripts, Creating Plots, Accessing help and documentation in R	
	6.2	Reading datasets and Exporting data from R, Manipulating and Processing Data in R, Using functions instead of script, built-in functions in R	
	6.3	Data Visualization: Types, Applications	

Textbooks:	
1	Cre Anand Rajaraman and Jeff Ullman ―Mining of Massive Datasets‖, Cambridge UniversityPress
2	Alex Holmes ―Hadoop in Practice‖, Manning Press, Dreamtech Press.
3	Dan Mcary and Ann Kelly ―Making Sense of NoSQL‖ – A guide for managers and the rest of us, Manning Press.
4	DT Editorial Services, “Big Data Black Book”, Dreamtech Press
5	EMC Education Services,”Data Science and Big Data Analytics”,Wiley