Course code: Course Title	Course Structure			Pre-Requisite
SE414: Big Data Analytics	L	T	P	Database Management Systems
	3	1	0	

Course Objective: Understand the fundamentals of various big data analysis techniques, Hadoop structure, environment and framework.

S. NO	Course Outcomes (CO)
CO1	Understand basic concepts of Big Data, its challenges, and modern data analytic tools to analyze large datasets.
CO2	Apply data stream mining techniques such as filtering, sampling, and real-time analytics for applications like sentiment analysis and stock market predictions.
CO3	Implement MapReduce programs using Hadoop Distributed File System (HDFS) and analyze the execution of MapReduce jobs.
CO4	Configure and manage a Hadoop cluster by handling security, monitoring, maintenance, and cloud integration for efficient big data processing.
CO5	Demonstrate big data frameworks like Pig, Hive, HBase, and ZooKeeper to process, query, and visualize large-scale datasets effectively.

S. NO	Contents	Contact Hours
UNIT 1	**Introduction to Big Data:** Introduction to Big Data Platform – Challenges of Conventional Systems - Intelligent data analysis – Nature of Data - Analytic Processes and Tools - Analysis vs Reporting - Modern Data Analytic Tools - Statistical Concepts: Sampling Distributions - Re-Sampling - Statistical Inference - Prediction Error.	8
UNIT 2	**Mining Data Streams:** Introduction To Streams Concepts – Stream Data Model and Architecture - Stream Computing - Sampling Data in a Stream – Filtering Streams – Counting Distinct Elements in a Stream – Estimating Moments – Counting Oneness in a Window – Decaying Window - Real time Analytics Platform(RTAP) Applications - Case Studies - Real Time Sentiment Analysis, Stock Market Predictions.	8
UNIT 3	**Hadoop:** History of Hadoop- The Hadoop Distributed File System – Components of Hadoop- Analyzing the Data with Hadoop- Scaling Out- Hadoop Streaming- Design of HDFS-Java interfaces to HDFS- Basics- Developing a Map Reduce Application-How Map Reduce Works-Anatomy of a Map Reduce Job run-Failures-Job Scheduling-Shuffle and Sort – Task execution - Map Reduce Types and Formats- Map Reduce Features.	10
UNIT 4	**Hadoop Environment:** Setting up a Hadoop Cluster - Cluster specification - Cluster Setup and Installation - Hadoop Configuration-Security in Hadoop - Administering Hadoop – HDFS - Monitoring-Maintenance-Hadoop benchmarks- Hadoop in the cloud.	8
UNIT 5	**Frameworks:** Applications on Big Data Using Pig and Hive – Data processing operators in Pig – Hive services – HiveQL – Querying Data in Hive - fundamentals of HBase and ZooKeeper - IBM InfoSphere BigInsights and Streams. Visualizations - Visual data analysis techniques, interaction techniques; Systems and applications.	8
	TOTAL	**42**

REFERENCES

S.No.	Name of Books/Authors/Publishers	Year of Publication / Reprint
1	Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer, 2007.	2007
2	Tom White "Hadoop: The Definitive Guide", O'Reilly, 3rd Edition.	2012
3	Paul Zikopoulos, Chris Eaton, Dirk Deroos, Tom Deutsch, George Lapis, "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data", McGraw Hill Education, 1st Edition.	2017
4	Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman, "Mining of Massive Datasets", Cambridge University Press, 2nd Edition.	2014
5	Bill Franks, Thomas H. Davenport, "Taming the Big Data Tidal Wave", Wiley, 1st Edition.	2012
6	Glenn J. Myatt, "Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining", Wiley–Blackwell.	2006
7	Pete Warden, "Big Data Glossary: A Guide to the New Generation of Data Tools", Shroff/O'Reilly, 1st Edition.	2011
8	Jiawei Han, Micheline Kamber, "Data Mining: Concepts & Techniques", 2nd Edition.	2010
9	Da Ruan, Guoquing Chen, Etienne E. Kerre, Geert Wets, "Intelligent Data Mining", Springer.	2007
10	Paul Zikopoulos, Dirk deRoos, Krishnan Parasuraman, Thomas Deutsch, James Giles, David Corrigan, "Harness the Power of Big Data The IBM Big Data Platform", McGraw Hill Publications.	2012
11	Michael Minelli, Michele Chambers, Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's", Wiley.	2013