PROGRAM ELECTIVE - I

FUNDAMENTALS OF BIG DATA ANALYTICS

Course Code	19IT4501F	Year	III	Semester	I
Course Category	PE	Branch	IT	Course Type	Theory
Credits	3	L-T-P	3-0-0	Prerequisites	Data Mining
Continuous Internal Evaluation :	30	Semester End Evaluation:	70	Total Marks:	100

Course Outcomes		Blooms Taxonomy Level
Upon successful completion of the course, the student will be able to		
CO1	Understand the concepts of Hadoop, HDFS, Map Reduce, Hadoop I/O, Cassandra and Spark operations for analytics of bigdata.	L2
CO2	Apply the knowledge of Hadoop distributed file system, Cassandra and Spark for solving real time problems	L3
CO3	Identify the appropriate concepts of big data to solve a given application.	L2
CO4	Devise solutions for real world use cases using appropriate big data concepts.	L4

Contribution of Course Outcomes towards achievement of Program Outcomes & Strength of correlations (H:High, M: Medium, L:Low)														
	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO 6**	**PO7**	**PO8**	**PO9**	**PO1 0**	**PO1 1**	**PO1 2**	**PSO 1**	**PSO 2**
CO 1	3												1	
CO 2	3				2								1	
CO 3	3	2			2								1	
CO 4	3		3		2								1	

Syllabus		
Unit No	**Contents**	**Mapped CO**
I	Types of Digital Data: Classification of Digital Data. Introduction to Big Data: Characteristic of Data, Evolution of Big Data, Definition of Big Data, Challenges with Big Data, What is Big Data?. Big Data Analytics: Where do we Begin?, What is Big Data Analytics?, What Big Data Analytics isn't?, Classification ofAnalytics, Terminologies Used in Big Data Environments. The Big Data Technology Landscape: NoSQL,	**CO1 CO2**
II	Introduction to Cassandra: Apache Cassandra – An Introduction, Features of Cassandra, CQL Data Types, CQLSH, Keyspaces, CRUD, Collections, Using a Counter, Time to Live, Alter Commands, Import and Export..	**CO1 CO2 CO3**
III	Hadoop Overview, HDFS (Hadoop Distributed File System), Processing Data with Hadoop, Managing Resources and Applications with Hadoop YARN (Yet another Resource Negotiator). Introduction to MAPREDUCE Programming: Introduction, Mapper, Red ucer, Combiner, Partitioner, Searching, Sorting, Compression.	**CO1 CO2 CO3**
IV	Introduction to Data Analysis with Spark: What is a Apache Spark, A unified Spark, Who uses Spark and for what?, A Brief Histor yof Spark, Spark version and releases, Storage layers for Spark. Programming with RDDs: RDD Basics, Creating RDDs, RDD Operations, Passing functions to Spark, Common Transformationsand Actions, Persistence. Spark SQL: Linking with Spark SQL, Using Spark SQL in Applications, Loading and Saving Data, JDBC /ODBC Server, User-defined functions, Spark SQL Performance.	**CO1 CO2 CO3**
V	Use case Study: Recommendation Systems: Introduction, A Model for Recommendation Systems, Collaborative Filtering System and Content Based Recommendations.	**CO1 CO2 CO3 CO4**

Learning Resources
Text Books
1. Big Data and Analytics, Seema Acharya, SubhashiniChellappan, First Edition,Wiley,2015 2. Learning Spark Lightning-Fast Big Data Analysis, Andy Konwinski, Holden Karau, MateiZaharia, Patrick Wendell , First Edition, O'Reilly, 2015 3. Big Data Analytics, Radha Shankarmani, M Vijaya Lakshmi, Second Edition, Wiley, 2017
References

1. Tom White, Hadoop: The Definitive Guide, Fourth Edition, O'Reilly, 2015
2. Hrushikesha Mohanty, Prachet Bhuyan, Deepak Chenthati Editors Big Data A PremierSpringer Volume 11
3. Using Flume: Flexible, Scalable, and Reliable Data Streaming by Hari Shreedharan

E- Resources and other Digital Material

1. https://www.coursera.org/courses?query=introduction%20to%20big%20data%20anal ytics
2. https://www.edx.org/learn/big-data
3. https://swayam.gov.in/nd1_noc20_cs46/