BIG DATA ANAYTICS

(Professional Elective –IV)

Course Code	20IT4702E	**Year**	IV	**Semester**	I
Course Category	PE -IV	**Branch**	IT	**Course Type**	Theory
Credits	3	**L-T-P**	3-0-0	**Prerequisites**	DBMS, Data Mining
Continuous Internal Evaluation :	30	**Semester End Evaluation:**	70	**Total Marks:**	100

Course Outcomes		**Blooms Taxonomy Level**
Upon Successful completion of course, the student will be able to		
CO1	Understand the concepts of Hadoop, Cassandra, Pig and Hive.	L2
CO2	Apply the knowledge of Hadoop and Cassandra for solving real time problems	L3
CO3	Use the concepts Pig and Hive for big data analysis	L3
CO4	Analyze the appropriate concepts of bigdata to solve a given application.	L4

Contribution of Course Outcomes towards achievement of Program Outcomes & Strength of Correlations (H:High,M:Medium,L:Low)

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**	**PSO1**	**PSO2**
CO1	3												3	
CO2	3	3	3										3	
CO3	3		3										3	
CO4	3	3											3	

Syllabus		
Unit No	**Contents**	**Mapped CO**
I	**Types of Digital Data**: Classification of Digital Data. Introduction to BigData: Characteristic of Data, Evolution of BigData, Definition of Big Data, Challenges with Big Data, What is BigData?. Big Data Analytics: Where dowe Begin? What is BigData Analytics?, What Big Data Analytics isn't?, Classification of Analytics, Terminologies Used in Big Data Environments. The BigData Technology Landscape: NoSQL	**CO1**
II	**Introduction to Cassandra**: Apache Cassandra – An Introduction Features of Cassandra, CQL Data Types, CQLSH, Key spaces, CRUD ,Collections, Using a Counter, Time to Live, Alter Commands, Import and Export.	**CO1 CO2 CO4**

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, Reducer, Combiner, Partitioner, Searching, Sorting, Compression.	**CO1** **CO2** **CO4**
IV	**Introduction to Hive**: Introduction – Architecture - Data Types - File Formats - Hive Query Language Statements – Partitions – Bucketing – Views - Sub-Query – Joins – Aggregations - Group by and Having - RCFile Implementation - Hive User Defined Function - Serialization and Deserialization.	**CO1** **CO3** **CO4**
V	**Pig**: Introduction - Anatomy – Features – Philosophy - Use Case for Pig - Pig Latin Overview - Pig Primitive Data Types - Running Pig - Execution Modes of Pig - HDFS Commands - Relational Operators - Eval Function - Complex Data Types - Piggy Bank - User-Defined Functions - Word Count Example using Pig.	**CO1** **CO3** **CO4**

Learning Resources

Text Books

1. Big Data and Analytics, Seema Acharya, Subhashini Chellappan ,First Edition,Wiley,2015

References

1. Tom White, Hadoop: The Definitive Guide, FourthEdition,O'Reilly,2015
2. Hrushikesha Mohanty, Prachet Bhuyan, Deepak Chenthati Editors Big Data A PremierSpringer Volume 11
3. Learning Spark Lightning-Fast Big Data Analysis, Andy Konwinski, Holden Karau, MateiZaharia, Patrick Wendell , First Edition, O'Reilly, 2015
4. Big Data Analytics, Radha Shankarmani, M VijayaLakshmi, Second Edition, Wiley, 2017

E- Resources and other Digital Material

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