



**GITAM INSTITUTE OF MANAGEMENT (GIM)**  
**Gandhi Institute of Technology and Management (GITAM)**  
**(Declared as Deemed to be University u/s 3 of UGC Act. 1956)**  
**Visakhapatnam – 45.**

<b>Course Code: MAN 322</b>	<b>Course Title: BIG DATA ANALYTICS</b>	
<b>Semester: VI</b>	<b>Course Type: Core</b>	<b>Credits: 3</b>
<b>Home Programme(s): BBA (BA)</b>		<b>Batch: 2020-23</b>
<b>Course Leader: Dr. M. Kamakshaiah</b>		

#### Course description and learning outcomes

Big data is a term used to describe a massive amount of structured and unstructured data collected over the years from different sources. Analysis of such data may provide great insights for a business. However, traditional data management functions are not capable for handling such data and requires specialized tool. Hadoop is a popular platform for carrying out big data analytics

#### Learning Objectives

- To acquaint the students with the concepts of big data (I1)
- To provide hands on experience in working with hadoop (I2)

On successful completion of this course, students will be able to:

CO	Learning Outcome	Assessment
CO1	Understand the concept of big data and the process of big data analytics	A1,
CO2	Write programs in Hadoop Mapreduce	A1, A2, A3
CO3	Write programs in APACHE PIG	A1, A2, A3, A4
CO4	Work with APACHE Hive & HBase	A1, A2, A3, A4
CO5	Understand the components in APACHE Spark	A1, A2, A3, A4

#### Course outline and indicative content

##### UNIT-I: Introduction to Big Data (CO1, L1)

Concept, Features of big data, big data challenges, Hadoop and its features, Hadoop Ecosystem, Hadoop Components, Hadoop Architecture, Hadoop Cluster, Hadoop Storage: HDFS

##### UNIT-II: Hadoop Mapreduce (CO2, L1, L2)

Concept, YARN components, YARN architecture, YARN mapreduce application execution flow, YARN workflow, Java for Mapreduce programming; Mapreduce examples; Mapreduce for data analytics: analyzing numerical and categorical data sets; Mapreduce for statistical analysis; Hadoop streaming.

##### UNIT-III: Apache PIG (CO3, L2)

PIG Components & Execution, PIG data types, Data models in PIG, Programming in PIG.

##### UNIT-IV: APACHE HIVE & HBase (CO4, L2)

Introduction, Architecture and components, data types and data models, HIVE partitioning and bucketing, HIVE tables, HIVE QL: joining tables, dynamic partitioning. Introduction, Architecture and components, Run modes, configuration, data models, HIVE data loading techniques,

## UNIT – V: Introduction to Apache Spark

Interactive analysis, RDD programming; Spark SQL, Data sets and DataFrames; Basics of MLib and GraphX.

### Assessment methods

	Task	Task type	Task mode	Weightage (%)
A1	Quiz (2)	Individual		40
A2	Assignments / Lab Tasks / Written Test/ Coursera Groups* or Individual (3)	Individual / Group		30
A3	Project	Individual		20
A4	Presentation	Individual		10

### Mapping Cos – Blooms Levels – Assessment Tools

Knowledge dimension / Cognitive dimension	L1. Remember	L2. Understand	L3. Apply	L4. Analyze	L5. Evaluate	L6. Create
Factual Knowledge	CO1 (A1)					
Conceptual Knowledge	CO1, CO2 (A1, A2)	CO2, CO3, CO4, CO5 (A1, A2)	CO1, CO3, CO4, CO5 (A1, A2, A3)	CO2, CO3, CO4, CO5 (A1, A2, A3)		
Procedural Knowledge		CO3, CO4, CO5 (A1, A2)	CO2, CO3, CO4, CO5 (A1, A2, A3)	CO3, CO4, CO5 (A1, A2, A3)		CO2, CO3, CO4, CO5 (A1, A2, A3)
Meta Cognitive Knowledge						

### Learning and teaching activities

Classroom Lectures, Application cases, Demonstration, Lab Sessions

### Teaching and learning resources

Computer Lab, Hadoop Software, Textbooks, Ebooks, Reference Materials, Web resources

### CO PO Mapping

This is to map the level of relevance of the Course Outcome (CO) with Programme Outcome (PO).

0= No Relevance; 1= Low Relevance; 2= Medium Relevance; 3= High Relevance

CO PO Mapping	PO1	PO2	PO3	PO4	PO5	PO6	Sum
	CO1	3	0	0	0	0	0
CO2	1	3	3	3	3	3	16
CO3	0	3	3	0	0	3	9

<b>CO4</b>	<b>0</b>	<b>3</b>	<b>3</b>	<b>0</b>	<b>0</b>	<b>3</b>	<b>9</b>
<b>CO5</b>	<b>0</b>	<b>3</b>	<b>3</b>	<b>0</b>	<b>0</b>	<b>3</b>	<b>9</b>
<b>Target Level Max.</b>	<b>4</b>	<b>12</b>	<b>12</b>	<b>3</b>	<b>3</b>	<b>12</b>	<b>46</b>

### **Program Outcomes**

1	Ability to understand the business problem with their knowledge in different functional areas of management
2	Ability to work with structured, semi – structured and unstructured data.
3	Ability to use tools such as Java, Hadoop to solve business analytics problem
4	Ability to apply analytics techniques to analyze and interpret the data.
5	Ability to perform descriptive, predictive and prescriptive analytics.
6	Have necessary skills and understanding to take up advanced topics in the area of analytics and thus enhance their career.