INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal, Hyderabad - 500 043

COMPUTER SCIENCE AND ENGINEERING

COURSE DESCRIPTOR

Course Title	**BIG DATA AND BUSINESS ANALYTICS**				
Course Code	ACS012				
Programme	B. Tech				
Semester	SEVEN				
Course Type	Core				
Regulation	IARE - R16				
Course Structure	**Theory**			**Practical**	
	Lectures	**Tutorials**	**Credits**	**Laboratory**	**Credits**
	3	1	4	3	2
Chief Coordinator	Dr. M Madhu Bala, Professor				

I. COURSE OVERVIEW:

This course provides a clear understanding on concepts of sources of big data, characteristics, storing and processing components, and analytics applications. This course emphasizes on potential impact of big data challenges, open research issues, and various tools associated with it. This course includes the introduction and processing big data with an overview of Hadoop technology and its components such as pig, hive, etc.

II. COURSE PRE-REQUISITES:

Level	Course Code	Semester	Prerequisites	Credits
UG	ACS005	IV	Database Management System	**3**

III. MARKS DISTRIBUTION:

Subject	SEE Examination	CIA Examination	Total Marks
Big Data and Business Analytics	70 Marks	30 Marks	100

IV. DELIVERY / INSTRUCTIONAL METHODOLOGIES:

✔	Chalk & Talk	✔	Quiz	✔	Assignments	✘	MOOCs
✔	LCD / PPT	✔	Seminars	✘	Mini Project	✔	Videos
✘	Open Ended Experiments						

V. EVALUATION METHODOLOGY:

The course will be evaluated for a total of 100 marks, with 30 marks for Continuous Internal Assessment (CIA) and 70 marks for Semester End Examination (SEE). Out of 30 marks allotted for CIA during the semester, marks are awarded by taking average of two CIA examinations or the marks scored in the make-up examination.

Semester End Examination (SEE):

The SEE is conducted for 70 marks of 3 hours duration. The syllabus for the theory courses is divided into FIVE modules and each module carries equal weight age in terms of marks distribution. The question paper pattern is as follows. Two full questions with "either" or "choice" will be drawn from each module. Each question carries 14 marks. There could be a maximum of two subdivisions in a question.

The emphasis on the questions is broadly based on the following criteria:

50 %	To test the objectiveness of the concept
30 %	To test the analytical skill of the concept
20 %	To test the application skill of the concept

Continuous Internal Assessment (CIA):

CIA is conducted for a total of 30 marks (Table 1), with 25 marks for Continuous Internal Examination (CIE), 05 marks for Quiz/ Alternative Assessment Tool (AAT).

Table 1: Assessment pattern for CIA

Component	**Theory**		**Total Marks**
Type of Assessment	**CIE Exam**	**Quiz / AAT**	
CIA Marks	25	05	30

Continuous Internal Examination (CIE):

Two CIE exams shall be conducted at the end of the 8th and 16th week of the semester respectively. The CIE exam is conducted for 25 marks of 2 hours duration consisting of two parts. Part–A shall have five compulsory questions of one mark each. In part–B, four out of five questions have to be answered where, each question carries 5 marks. Marks are awarded by taking average of marks scored in two CIE exams

Quiz / Alternative Assessment Tool (AAT)

Two Quiz exams shall be online examination consisting of 25 multiple choice questions and are be answered by choosing the correct answer from a given set of choices (commonly four). Marks shall be awarded considering the average of two quizzes for every course. The AAT may include seminars, assignments, term paper, open ended experiments, five minutes video and MOOCs.

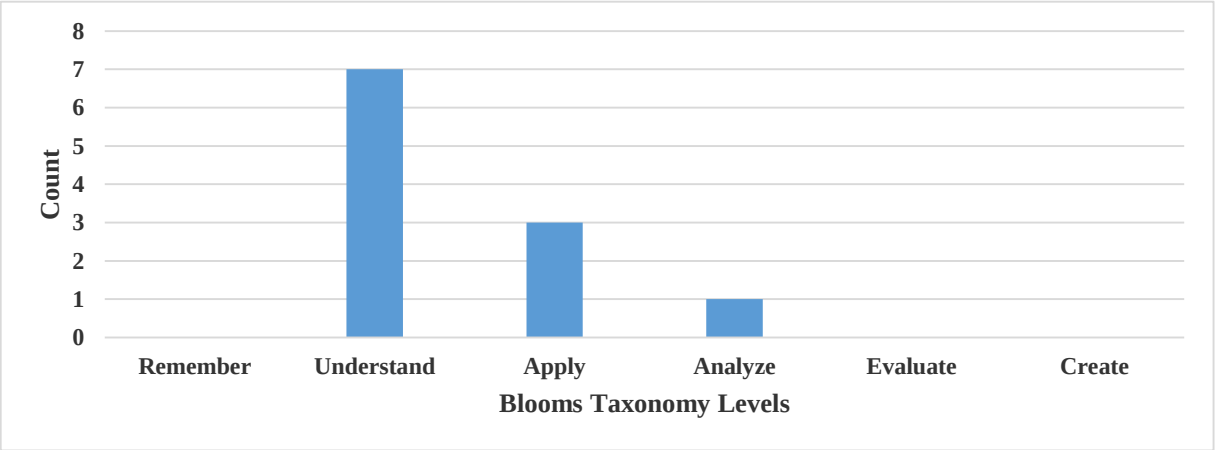
VI. COURSE OBJECTIVES:

Students will try to learn:	
I	The scope and essentiality of Big Data and Business Analytics.
II	The technologies used to store, manage, and analyze big data in a Hadoop ecosystem.
III	The techniques and principles in big data analytics with scalability and streaming capability.
IV	The hypothesis on the optimized business decisions in solving complex real-world problems.

VII. COURSE OUTCOMES:

After successful completion of this course, students will be able to:		
Course Outcomes		**Knowledge Level (Bloom's Taxonomy)**
CO 1	Explain the evolution of big data with its characteristics and challenges with traditional business intelligence.	Understand
CO 2	Compare big data analysis and analytics in optimizing the business decisions.	Understand
CO 3	Classify the key issues and applications in intelligent business and scientific computing.	Understand
CO 4	Explain the big data technologies used to process and querying the bigdata in Hadoop, MapReduce, Pig and Hive.	Understand
CO 5	Make use of appropriate components for processing, scheduling and knowledge extraction from large volumes in distributed Hadoop Ecosystem.	Apply
CO 6	Translate the data from traditional file system to HDFS for analyzing big data in Hadoop ecosystem.	Understand
CO 7	Develop a Map Reduce application for optimizing the jobs.	Apply
CO 8	Develop applications for handling huge volume of data using Pig Latin.	Apply
CO 9	Explain the importance of bigdata framework HIVE and its built-in functions, data types and services like DDL.	Understand
CO 10	Demonstrate business models and scientific computing paradigms, and tools for big data analytics.	Understand
CO 11	Categorize Hadoop components for developing real time big data analytics in various applications like recommender systems, social media applications etc.	Analyze

COURSE KNOWLEDGE COMPETEMCY LEVEL



VIII. HOW PROGRAM OUTCOMES ARE ASSESSED:

PO No	Program Outcomes	Strength	Proficiency Assessed by
PO 1	**Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.	**2**	CIE/Quiz/AAT
PO 2	**Problem analysis**: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences	**2**	CIE/Quiz/AAT
PO 3	**Conduct Investigations of Complex Problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.	**3**	CIE/Quiz/AAT
PO 4	**Modern Tool Usage:** Create, select, and apply appropriate techniques, resources, and modern Engineering and IT tools including prediction and modelling to complex Engineering activities with an understanding of the limitations.	**3**	CIE/Quiz/AAT
PO 5	**Individual and Teamwork:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings	**3**	CIE/Quiz/AAT
PO 6	**Communication:** Communicate effectively on complex Engineering activities with the Engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.	-	-
PO 7	**Life-Long Learning:** Recognize the need for and having the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.	-	-
PO 8	**Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the Engineering practice.	-	-
PO 9	**Individual and Teamwork:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.	-	-
PO 10	**Communication:** Communicate effectively on complex Engineering activities with the Engineering community and with	-	-

	society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions**.**		
PO 11	**Project Management and Finance:** Demonstrate knowledge and understanding of the Engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary Environments**.**	-	-
PO 12	**Life - Long Learning:** Recognize the need for and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change**.**	-	-

IX. HOW PROGRAM SPECIFIC OUTCOMES ARE ASSESSED:

Program Specific Outcomes		Strength	Proficiency Assessed by
PSO 1	Understand, design and analyse computer programs in the areas related to **Algorithms, System Software, Web design, Bigdata, Artificial Intelligence, Machine Learning and Networking.**	3	Research papers/ Group discussion/ Short term courses
PSO 2	Focus on improving **software reliability, network security and information retrieval systems.**	2	Research papers/ Industry exposure
PSO 3	Make use of **modern computer tools** for creating innovative career paths, to be an entrepreneur and desire for higher studies.	3	Research papers/ Group discussion/ Short term courses

X. MAPPING OF EACH CO WITH PO(s), PSO(s):

Course Outcomes	Program Outcomes												Program Specific Outcomes		
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3
CO 1	√														
CO 2	√												√		
CO 3	√	√	√												
CO 4	√	√	√										√		
CO 5	√	√	√										√		
CO 6	√	√													
CO 7	√	√	√	√	√								√	√	√
CO 8	√	√	√	√	√								√	√	√
CO 9	√	√	√										√		
CO 10				√											
CO 11				√									√	√	√

XI. JUSTIFICATIONS FOR CO - (PO, PSO) MAPPING - DIRECT:

Course Outcomes	POs / PSOs	Justification for mapping (Students will be able to)	No. of key competencies
CO 1	**PO 1**	Explain the evolution of big data with its characteristics and challenges by applying **computer science methodologies**	**1**
CO 2	**PO 1**	Compare big data analysis and analytics in optimizing business decisions by using the **mathematical principles** and **computer science methodologies**.	**2**
	PSO 1	Understand the differences between analysis and analytics in the areas related to **Algorithms, Bigdata, Artificial Intelligence, Machine Learning and Networking.**	**4**
CO 3	**PO 1**	Understand the key issues and applications in intelligent business by applying **mathematical principles** and **computer science methodologies**	**2**
	PO 2	Understand the key issues in **problems identification and formulation, data collection, model translation, validation, interpretation of results and documentation** in optimizing business decisions.	**6**
	PO 3	Classify the key issues in terms of **defining various problems**, **customer and user needs**, **cost effective** and **creative solutions, design process, economic context and management techniques.**	**7**
CO 4	**PO 1**	Explain the big data technologies used to process and querying the bigdata by applying **mathematical principles** and **computer science methodologies**	**2**
	PO 2	Understand the **problem** and **develop** solutions using big data technologies and **document** the **results for interpretation**	**4**
	PO 3	Identify the appropriate technology like pig, hive etc. suitable for **various problems**, by understanding **customer and user needs**, with **cost effective** and **creative solutions** by **managing the design process, knowledge on economic context, management techniques.**	**7**
	PSO 1	Explain the big data technologies used to process and querying the bigdata in the areas related to **Algorithms, Bigdata, Artificial Intelligence, Machine Learning and Networking.**	**4**
CO 5	**PO 1**	Select appropriate Hadoop component for mining large volumes of data in Hadoop using **computer science methodologies**	**1**
	PO 2	Make use of Hadoop components on huge volume of **information and data collected** from various sources and perform **model translation and validation**	**3**
	PO4	Make use of Hadoop components for developing applications based on **technical literature and quality issues. Identify, classify and describe the performance of systems through analytical methods and techniques**.	**3**
	PSO 1	Make use of Hadoop components on huge volume data used to develop analytical solutions related to **Bigdata, Artificial Intelligence, Machine Learning and Networking.**	**4**
CO 6	**PO 1**	Translate the data from traditional file system to HDFS for analyzing big data in Hadoop ecosystem using the **mathematical principles** and **computer science methodologies**	**2**

	PO 2	Translation of data structure from traditional to HDFS includes volume of **information and data, file structure translation methods, validation and solution development** with proper **documentation.**	**6**
CO 7	**PO 1**	Develop a Map Reduce application for optimizing the jobs by applying **mathematical and scientific principles by integrating computer science knowledge**.	**3**
	PO 2	Develop a Map Reduce application for optimizing the jobs for specific **problems** by including volume of **information and data, file structure translation, validation and solution development** with proper **documentation.**	**6**
	PO 3	Develop a Map Reduce application by **investigating and defining various problems**, understanding **customer and user needs**, with **cost effective** and **creative solutions** by **managing the design process, knowledge on economic context, management techniques.**	**7**
	PO 4	Develop MapReduce applications with **laboratory skills, technical literature and quality issues. Identify, classify and describe the performance of systems through analytical methods and techniques**.	**5**
	PO 5	Make use of **library resources** for optimizing the jobs using Map Reduce application.	**1**
	PSO 1	Develop a Map Reduce application for optimizing the jobs related to **Algorithms, Bigdata, Artificial Intelligence, Machine Learning and Networking.**	**4**
	PSO 2	Develop a Map Reduce application with major focus on improving software reliability, network security **and information retrieval systems.**	**1**
	PSO 3	Develop a Map Reduce application by using modern computer tools for creating **innovative career paths, to be an** entrepreneur **and desire for higher studies.**	**2**
CO 8	**PO 1**	Understand the pig latin functionalities for handling huge volume of data by applying **mathematical and scientific principles by integrating computer science knowledge**.	**3**
	PO 2	Identify pig latin functions for big data applications related to specific **problems** by including huge volume of **information and data collection, file structure translation, validation and solution development** with proper **documentation.**	**6**
	PO 3	Develop a Pig Latin application by **investigating and defining various problems**, understanding **customer and user needs**, with **cost effective** and **creative solutions** by **managing the design process, knowledge on economic context, management techniques.**	**7**
	PO 4	Develop applications for handling huge volume of data using Pig Latin by working on **laboratory skills, technical literature and quality issues. Identify, classify and describe the performance of systems in computer software by applying quantitative methods through analytical methods and techniques**.	**8**
	PO 5	Develop **software applications** for handling huge volume of data using Pig Latin.	**1**
	PSO 1	Develop Pig Latin big data applications for specific problems by including huge volume of data and related to **Algorithms, Bigdata, Artificial Intelligence, Machine Learning and Networking.**	**4**

	PSO 2	Develop big data applications using Pig Latin for specific problems with a major focus on improving software reliability, network security **and information retrieval systems.**	**1**
	PSO 3	**Develop applications** by using **modern computer tools** related to Pig Latin for **creating innovative career paths, to be an entrepreneur and desire for higher studies.**	**3**
CO 9	**PO 1**	Understand the importance of big data framework HIVE by using **computer science methodologies, mathematical** and **scientific principles.**	**3**
	PO 2	Demonstrate the HIVE functions and services for specific **problems** by including huge volume of **information and data collection, file structure translation, validation and solution development** with proper **documentation.**	**6**
	PO 3	Explain the HIVE application process by including v**arious problems**, **customer and user needs**, with **cost effective** and **creative solutions** by **managing the design process, knowledge on economic context, management techniques.**	**7**
	PSO 1	Explain the HIVE features and services for analyzing programs in the areas related to **Algorithms, Bigdata, Artificial Intelligence, Machine Learning and Networking.**	**5**
CO 10	**PO 4**	Make use of tools for **designing business models** through **laboratory skills, technical literature, technical uncertainty and quality issues. Identify, classify and describe the performance of systems in computer software by applying quantitative methods through analytical methods and techniques**.	**8**
CO 11	**PO 4**	Develop real time applications through **laboratory skills** and considering **technical literature, technical uncertainty and quality issues. Identify, classify and describe the performance of systems in computer software by applying quantitative methods through analytical methods and techniques**.	**9**
	PSO 1	Categorize various Hadoop components in the areas related to **Algorithms, Bigdata, Artificial Intelligence, Machine Learning.**	**4**
	PSO 2	Develop applications using Hadoop ecosystem with a major focus on improving software reliability, network security **and information retrieval systems.**	**1**
	PSO 3	**Develop applications** by using Hadoop **modern computer tools** for creating **innovative career paths, to be an entrepreneur and desire for higher studies.**	**3**

XII. TOTAL COUNT OF KEY COMPETENCIES FOR CO – (PO, PSO) MAPPING

Course Outcomes	**Program Outcomes/Number of Key Competencies Matched**												**PSO/ No. of key competencies**		
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**1**	**2**	**3**
	3	**10**	**10**	**11**	**1**	**5**	**3**	**3**	**12**	**5**	**12**	**12**	**6**	**2**	**2**
CO 1	1														
CO 2	2												4		

CO 3	2	6	7												
CO 4	2	4	7										4		
CO 5	1	3	3												
CO 6	2	6													
CO 7	3	6	7	5	1								4	1	2
CO 8	3	6	7	8	1								4	1	2
CO 9	3	6	7										5		
CO 10				8											
CO 11				9									4	3	3

XIII. PERCENTAGE OF KEY COMPETENCIES FOR CO – (PO, PSO):

Course Outcomes	Program Outcomes / Number of Vital Features												PSO/ No. of Vital Features		
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3
	3	10	10	11	1	5	3	3	12	5	12	12	6	2	1
CO 1	33.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CO 2	66.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	66.7	0.0	0.0
CO 3	66.7	60.0	70.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CO 4	66.7	40.0	70.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	66.7	0.0	0.0
CO 5	33.3	30.0	30.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CO 6	66.7	60.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CO 7	100.0	60.0	70.0	45.5	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	66.7	33.3	66.7
CO 8	100.0	60.0	70.0	72.7	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	66.7	33.3	66.7
CO 9	100.0	60.0	70.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	66.7	0.0	0.0
CO 10	0.0	0.0	0.0	72.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CO 11	0.0	0.0	0.0	81.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	66.7	100.0	66.7

XIV. COURSE ARTICULATION MATRIX (CO - PO/PSO MAPPING)

COs and POs and COs and PSOs on the scale of 0 to 3, **0** being **no correlation**, **1** being the **low correlation**, **2** being **medium correlation** and **3** being **high correlation**.

0 ≤ C ≤ 5%– No correlation; **2** – 40 % < C < 60%– Moderate.

1 < C ≤ 40%– Low / Slight; **3** – 60% ≤ C < 100%– Substantial / High

Course Outcomes	Program Outcomes												Program Specific Outcomes		
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3
CO 1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CO 2	3	0	0	0	0	0	0	0	0	0	0	0	3	0	0
CO 3	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0
CO 4	3	2	3	0	0	0	0	0	0	0	0	0	3	0	0
CO 5	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
CO 6	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0
CO 7	3	3	3	2	3	0	0	0	0	0	0	0	3	1	3
CO 8	3	3	3	3	3	0	0	0	0	0	0	0	3	1	3
CO 9	3	3	3	0	0	0	0	0	0	0	0	0	3	0	0
CO 10	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0
CO 11	0	0	0	3	0	0	0	0	0	0	0	0	3	3	3
TOTAL	23	18	16	11	6	0	0	0	0	0	0	0	1.5	1.5	1.5
AVERAGE	2.5	2.5	2.6	2.7	3.0	0	0	0	0	0	0	0	3.0	2.5	3.0

XV. ASSESSMENT METHODOLOGIES – DIRECT

CIE Exams	PO 1,PO 2, PO 3,PO 4, PO 5	SEE Exams	PO 1,PO 2, PO 3,PO 4, PO 5	Assignments	PO 1	Seminars	PO 2
Laboratory Practices	-	Student Viva	-	Mini Project	-	Certification	-
Term Paper	-						

XVI. ASSESSMENT METHODOLOGIES - INDIRECT

✔	Early Semester Feedback	✔	End Semester OBE Feedback
✗	Assessment of Mini Projects by Experts		

XVII. SYLLABUS

UNIT-I	INTRODUCTION TO BIG DATA
Introduction to Big data: Characteristics of Data, Evolution of Big Data, Definition of Big Data, Challenges with Big Data, Traditional Business Intelligence (BI) versus Big Data. Big data analytics: Classification of Analytics, Importance and challenges facing big data, Terminologies Used in Big Data Environments, The Big Data Technology Landscape.	

UNIT-II	INTRODUCTION TO HADOOP
Introducing Hadoop, RDBMS versus Hadoop, Distributed Computing Challenges, History and overview of Hadoop, Use Case of Hadoop, Hadoop Distributors, Processing Data with Hadoop, Interacting with Hadoop Ecosystem	
UNIT-III	**THE HADOOP DISTRIBUTED FILESYSTEM**
Hadoop Distributed File System (HDFS): The Design of HDFS, HDFS Concepts, Basic Filesystem Operations, Hadoop Filesystems. The Java Interface- Reading Data from a Hadoop URL, Reading Data Using the Filesystem API, Writing Data. Data Flow- Anatomy of a File Read, Anatomy of a File Write, Limitations.	
UNIT -IV	**UNDERSTANDING MAP REDUCE FUNDAMENTALS**
Map Reduce Framework: Exploring the features of Map Reduce, Working of MapReduce, Exploring Map and Reduce Functions, Techniques to optimize MapReduce jobs, Uses of MapReduce. Controlling MapReduce Execution with Input Format, Reading Data with custom Record Reader, - Reader, Writer, Combiner, Partitioners, MapReduce Phases, Developing simple MapReduce Application.	
UNIT -V	**INTRODUCTION TO PIG AND HIVE**
Introducing Pig: Pig architecture, Benefits, Installing Pig, Properties of Pig, Running Pig, Getting started with Pig Latin, Working with operators in Pig, Working with functions in Pig. Introducing Hive: Getting started with Hive, Hive Services, Data types in Hive, Built-in functions in Hive, Hive DDL.	
Text Books:	
1. Seema Acharya, Subhashini Chellappan, ―Big Data and Analytics, Wiley Publications, 2nd Edition,2014DT Editorial Services, ―Big Data, Dream Tech Press, 2nd Edition, 2015. 2. Tom White, ―Hadoop: The Definitive Guide, O'Reilly, 3rd Edition, 2012. 3. Black Book Big Data, dreamtech publications, 1st Edition, 2017	
Reference Books:	
1. Michael Minelli, Michele Chambers, Ambiga Dhiraj, ―Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business, Wiley CIO Series, 1st Edition, 2013. 2. Rajiv Sabherwal, Irma Becerra- Fernandez, ―Business Intelligence –Practice, Technologies and Management, John Wiley, 1st Edition, 2011. 3. Arvind Sathi, ―Big Data Analytics: Disruptive Technologies for Changing the Game, IBM Corporation, 1st Edition, 2012.	

XVIII. COURSE PLAN:

The course plan is meant as a guideline. Probably there may be changes.

Lecture No	Topics to be covered	CO	Reference
1	Define big data and its importance.	CO 1	T1:2.3
2-3	Describe the elements of big data-volume, variety, velocity and veracity	CO 1	T1:2.1, 2.5
4-5	Understand the life cycle of big data	CO 1 CO 2	T1:2.4
6-7	Define the importance and challenges of big data.	CO 1,CO 2	T1:2.5 – 2.6 R2:21.51
8	Understand Traditional Vs Big Data Business Approach	CO 1,CO 2	T1:2.9
9-10	Classify the Big data analytics - Classification of Analytics	CO 1,CO 4	T1:3.1 R2:21.51

Lecture No	Topics to be covered	CO	Reference
11	Importance and challenges facing big data,	CO 4	T1:3.7 -3.8
12-14	Explain the terminologies Used in Big Data Environments	CO 4	T1:3.12 R2:21.55
15	Explain the Big Data Technology Landscape with Hadoop ecosystem.	CO 4	T1:4.1 – 4.2 R2:21.58
16	Understand the core components of Hadoop-big data.	CO 4	T2:26.16 R2:21.61
17-18	Outline Hadoop ecosystem and Computing Challenges, RDBMS versus Hadoop	CO 5,CO 6	T1:5.1 – 5.5 R2:21.24
19	Recall the history and overview of Hadoop	CO 5,CO 6	T1:5.5 R2:21.29
20	Demonstrate the real time use case in Hadoop	CO 4	T1:5.6 – 5.7 R2:21.31
21-22	Explain Hadoop Distributors and processing Data with Hadoop	CO 4,CO 5	T1:5.8 R2:21.33
23	Summarize the other components in Hadoop Interacting in Hadoop Ecosystem	CO 4,CO 6	T1:5.9
24	Explain the Design concepts of HDFS	CO 5	T1:5.11 R2:21.64
25	Find differences between Basic Filesystem Operations and Hadoop Filesystems.	CO 4,CO 6	T1:5.10-5.13 T2:3
26-27	Explain the Java Interface for Reading Data from a Hadoop URL Using the Filesystem API	CO 4,CO 6	T2:3
28-29	Explain Writing Data and Data Flow- Anatomy of a File Read, Anatomy of a File Write, Limitations	CO 4,CO 6	T1:5.10 T2:3
30-31	Explore the features of MapReduce and Map and Reduce Functions	CO 4,CO 7	T1:8.1-8.3 T2:8
32	Outline the techniques to optimize MapReduce jobs and uses	CO 4, CO 7	T2:27.8
33-35	Illustrate the controlling MapReduce Execution with Input Format	CO 4,CO 7, CO 10	T2:7
36-37	Explain the reading Data with custom Record Reader, - Reader, Writer, Combiner, Practitioners, MapReduce Phases	CO 7	T1:8.2 – 8.3
38	Develop a simple MapReduce Application	CO 7,CO 10, CO 11	T1:8.4 – 8.8
39	Explain Pig architecture	CO 8,CO 9	T1:10.1-10.6
40-41	Summarize Installation process of Pig along with Properties and getting started with Pig Latin,	CO 8,CO 9	T2:11
42	Develop applications by working with operators in Pig, Working with functions in Pig.	CO 8,CO 9, CO 10,CO 11	T1:10.7-10.12
43	Explain the Hive component and Hive Services	CO 9	T1:9.1-9.2 T2:12
44-45	Demonstrate Hive Data types, Built-in functions and Hive DDL.	CO 9,CO 10, CO 11	T1:9.3-9.8

Prepared by:
Dr. M Madhu Bala, Professor **HOD, CSE**