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	VII	CSE | IT			
Course Type	Core				
Regulation	IARE - R16				
Course Structure	**Theory**			**Practical**	
	Lectures	**Tutorials**	**Credits**	**Laboratory**	**Credits**
	3	1	4	3	2
Chief Coordinator	Ms. G Sulakshana, Assistant Professor				
Course Faculty	Ms. S Swarajya Laxmi, Assistant Professor Ms. E Uma Shankari, Assistant Professor Ms. G Srilekha, Assistant Professor				

I. COURSE OVERVIEW:

The purpose of this course is to provide a clear understanding of the concepts that underlying fundamental concepts and tools in big data analytics with emphasis on their applications to computer science. In digital world, data are generated from various sources and the fast transition from digital technologies has led to growth of big data analytics. This course emphasizes to explore the 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, pig , hive.

II. COURSE PRE-REQUISITES:

Level	Course Code	Semester	Prerequisites	Credits
UG	ACS005	IV	Database Management Systems	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:

Theory Course:

Each theory 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 sessional examinations or the marks scored in the make-up examination conducted.

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 units and each unit carries equal weightage in terms of marks distribution. The question paper pattern is as follows.

Two full questions with ‗either' ‗or' choice will be drawn from each unit. Each question carries 14 marks. There could be a maximum of three sub divisions 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):

For each theory course the CIA shall be conducted by the faculty/teacher handling the course as given in Table-5. CIA is conducted for a total of 30 marks, with 25 marks for Continuous Internal Examination (CIE) and 05 marks for Quiz / Alternative Assessment Tool (AAT).

Table 1: Assessment pattern for CIA

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

Continuous Internal Examination (CIE):

Two CIE exams shall be conducted at the end of the 8th and 17th 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. The valuation and verification of answer scripts of CIE exams shall be completed within a week after the conduct of the Internal Examination.

Quiz / Alternative Assessment Tool (AAT)

Two Quiz exams shall be online examination consisting of 20 multiple choice questions and are be answered by choosing the correct answer from a given set of choices (commonly four). Such a question paper shall be useful in the testing of knowledge, skills, application, analysis, evaluation and understanding of the students. Marks shall be awarded considering the average of two quizzes for every course.

In order to encourage innovative methods while delivering a course, the faculty members have been encouraged to use the Alternative Assessment Tool (AAT) in place of two quizzes. This AAT enables faculty to design own assessment patterns during the CIA. However, the usage of AAT is completely optional. The AAT enhances the autonomy (freedom and flexibility) of individual faculty and enables them to create innovative pedagogical practices. If properly applied, the AAT converts the classroom into an effective learning centre. The AAT may include seminars, assignments, term paper, open ended experiments, micro- projects, five minutes video, MOOCs etc.

However, it is mandatory for a faculty to obtain prior permission from the concerned HOD and spell out the teaching/assessment pattern of the AAT prior to commencement of the classes.

.

VI. HOW PROGRAM OUTCOMES ARE ASSESSED:

	Program Outcomes (POs)	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	Assignments
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	Term paper
PO 4	**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	Seminars and Mini Project

PO 5	**Modern tool usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.	3	Seminars and Mini Project
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. doctoral level studies	2	Term paper

3 = High; 2 = Medium; 1 = Low

VII. HOW PROGRAM SPECIFIC OUTCOMES ARE ASSESSED:

Program Specific Outcomes (PSOs)		Strength	Proficiency assessed by
PSO 1	**Professional Skills:** The ability to understand, analyze and develop computer programs in the areas related to algorithms, system software, multimedia, web design, big data analytics, and networking for efficient design of computer-based systems of varying complexity.	3	Seminars and Mini project
PSO 2	**Problem-Solving Skills:** The ability to apply standard practices and strategies in software project development using open-ended programming environments to deliver a quality product for business success.	2	Seminars
PSO 3	**Successful Career and Entrepreneurship:** The ability to employ modern computer languages, environments, and platforms in creating innovative career paths to be an entrepreneur, and a zest for higher studies.	2	Mini project

3 = High; 2 = Medium; 1 = Low

VIII. COURSE OBJECTIVES :

The course should enable the students to:	
I	Optimize business decisions and create competitive advantage with Big data analytics.
II	Understand several key big data technologies used for storage, analysis and manipulation of data.
III	Recognize the key concepts of Hadoop framework, map reduce.
IV	Demonstrate the concepts in Hadoop for application development.

IX. COURSE OUTCOMES:

COs	Course Outcome	CLOs	Course Learning Outcome
CO 1	Understand the key issues in big data analytics and its associated applications in business analytics.	CLO 1	Understand what Big Data, importance and various sources of data.
		CLO 2	Describe the elements of big data-volume, variety, velocity and veracity.
		CLO 3	Understand the importance and challenges of big data.
		CLO 4	Define big data analytics advantages and its applications.
CO 2	Illustrate different types of big data technologies in	CLO 5	Define distributed and parallel computing for big data.

COs	Course Outcome	CLOs	Course Learning Outcome
	Hadoop parallel world.	CLO 6	Analyze the core components of Hadoop with basic commands.
		CLO 7	Explain the key features of Hadoop in processing big data.
		CLO 8	Understand Hadoop ecosystem with its animal planet.
CO 3	Interpret disparate data storing in Hadoop Distributed File Systems (HDFS).	CLO 9	Explain the basic terminology of Hadoop Distributed File Systems (HDFS).
		CLO 10	Describe in detail about Distributed file system.
		CLO 11	Understand the concept of Hadoop cluster architecture.
		CLO 12	Explain a file in HDFS and represent the anatomy of file read and write.
CO 4	Explore map reduce framework and optimize its jobs.	CLO 13	Understand Map Reduce and its qualities and retain advanced Map Reduce thoughts.
		CLO 14	Understand the architecture of Map Reduce framework.
		CLO 15	Demonstrate the techniques to optimize Map Reduce jobs.
		CLO 16	Understand the typical use occasions of input and output forms of Map Reduce.
CO 5	Explain the basic methodologies of pig and hive.	CLO 17	Demonstrate an ability to use frameworks like pig and hive to process Big Data and Analytics.
		CLO 18	Design the architecture of pig with its data types and operations.
		CLO 19	Explain the architecture of hive with different operations.
		CLO 20	Design and implement different technologies for processing big data in pig and hive.

X. COURSE LEARNING OUTCOMES (CLOs):

CLO Code	CLO's	At the end of the course, the student will have the ability to:	PO's Mapped	Strength of Mapping
ACS012.01	CLO 1	Understand what Big Data, importance and various sources of data.	PO 1	2
ACS012.02	CLO 2	Describe the elements of big data-volume, variety, velocity and veracity.	PO 1	2
ACS012.03	CLO 3	Understand the importance and challenges of big data.	PO 2	2
ACS012.04	CLO 4	Define big data analytics advantages and its applications.	PO 4	3
ACS012.05	CLO 5	Define distributed and parallel computing for big data.	PO 4	3
ACS012.06	CLO 6	Analyze the core components of Hadoop with basic commands.	PO 2	2
ACS012.07	CLO 7	Explain the key features of Hadoop in processing big data.	PO 5	3
ACS012.08	CLO 8	Understand Hadoop ecosystem with its animal planet.	PO 12	2
ACS012.09	CLO 9	Explain the basic terminology of Hadoop Distributed File Systems (HDFS).	PO 4	3
ACS012.10	CLO 10	Describe in detail about Distributed file system.	PO 2	2
ACS012.11	CLO 11	Understand the concept of Hadoop cluster architecture.	PO 5	3

CLO Code	CLO's	At the end of the course, the student will have the ability to:	PO's Mapped	Strength of Mapping
ACS012.12	CLO 12	Explain a file in HDFS and represent the anatomy of file read and write.	PO 4	3
ACS012.13	CLO 13	Understand Map Reduce and its qualities and retain advanced Map Reduce thoughts.	PO 5	3
ACS012.14	CLO 14	Understand the architecture of Map Reduce framework.	PO 2	2
ACS012.15	CLO 15	Demonstrate the techniques to optimize Map Reduce jobs.	PO 4	3
ACS012.16	CLO 16	Understand the typical use occasions of input and output forms of Map Reduce.	PO 5	3
ACS012.17	CLO 17	Demonstrate an ability to use frameworks like pig and hive to process Big Data and Analytics.	PO 2	2
ACS012.18	CLO 18	Design the architecture of pig with its data types and operations.	PO 4	3
ACS012.19	CLO 19	Explain the architecture of hive with different operations.	PO 1	2
ACS012.20	CLO 20	Design and implement different technologies for processing big data in pig and hive.	PO 4	3

3= High; 2 = Medium; 1 = Low

XI. MAPPING COURSE OUTCOMES LEADING TO THE ACHIEVEMENT OF PROGRAM OUTCOMES

Course Outcomes (COs)	Program Outcomes (POs)					Program Specific Outcomes (PSOs)		
	PO 1	PO 2	PO 4	PO 5	PO 12	PSO1	PSO2	PSO3
CO 1	2	2	3			3		
CO 2		2	3	3	2		2	
CO 3		2	3	3				
CO 4		2	3	3				
CO 5	2	2	3					

3= High; 2 = Medium; 1 = Low

XII. MAPPING COURSE LEARNING OUTCOMES LEADING TO THE ACHIEVEMENT OF PROGRAM OUTCOMES AND PROGRAM SPECIFIC OUTCOMES:

Course Learning Outcomes (CLOs)	Program Outcomes (POs)												Program Specific Outcomes (PSOs)		
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CLO 1	2														
CLO 2	2														
CLO 3		2													
CLO 4				3									3		

Course Learning Outcomes (CLOs)	Program Outcomes (POs)												Program Specific Outcomes (PSOs)		
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CLO 5				3									3		
CLO 6		2													
CLO 7					3										
CLO 8												2			2
CLO 9				3											
CLO 10		2													
CLO 11					3										
CLO 12				3										2	
CLO 13					3										
CLO 14		2													
CLO 15				3									3		
CLO 16					3										
CLO 17		2													
CLO 18				3									3		
CLO 19	2														
CLO 20				3										3	

3 = High; 2 = Medium; 1 = Low

XIII. ASSESSMENT METHODOLOGIES – DIRECT

CIE Exams	PO1, PO2, PO4, PO5	SEE Exams	PO1, PO2, PO4, PO5	Assignments	-
Seminars	PO4,PO5 PSO1,PSO2	Laboratory Practices	-	Student Viva	-
Mini Project	PO4, PO5 PSO1,PSO3	Certification	-	Term Paper	PO2, PO12

XIV. ASSESSMENT METHODOLOGIES - INDIRECT

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

XV. 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 Map Reduce, Exploring Map and Reduce Functions, Techniques to optimize Map Reduce jobs, Uses of Map Reduce. Controlling MapReduce Execution with InputFormat, Reading Data with custom RecordReader,-Reader, Writer, Combiner, Partitioners, Map Reduce 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.	

XVI. COURSE PLAN:

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

Lecture No	Topics to be covered	Course Learning Outcomes (CLOs)	Reference
1	Define big data and its importance.	CLO 1	T2:26.3
2-3	Describe the elements of big data-volume, variety, velocity and veracity	CLO 1	R2:21.48

Lecture No	Topics to be covered	Course Learning Outcomes (CLOs)	Reference
4-5	Understand the life cycle of big data	CLO 2	T2:26.6 R2:21.50
6-7	Define the importance and challenges of big data.	CLO 2	T2:26.7 R2:21.51
8	Understand Traditional Vs Big Data Business Approach	CLO 3	T2:26.8
9-10	Describe big data analytics applications.	CLO 3	T2:26.7 R2:21.51
11	Define distributed and parallel computing for big data.	CLO 4	T2:26.10
12-14	Describe the overview of Hadoop.	CLO 5	T2:26.14 R2:21.55
15	Define the history of Hadoop with an divide and conquer philosophy.	CLO 5	T2:26.15 R2:21.58
16	Understand the core components of Hadoop-big data .	CLO 6	T2:26.16 R2:21.61
17-18	Understand the features of Hadoop	CLO 6	T2:25.12 R2:21.24
19	Define Hadoop eco system.	CLO 7	T2:25.16 R2:21.29
20	Define the Introduction to Hadoop Distributed File Systems (HDFS).	CLO 8	T2:25.14 R2:21.31
21-22	Understand the HDFS architecture.	CLO 9	T2:25.14 R2:21.33
23	Understand the connecting and extracting data from Distributed file system.	CLO 9	R2:21.33
24	Describe the Block replication.	CLO 10	T2:27.2 R2:21.64
25	Understand the Hadoop cluster architecture.	CLO 10	T2:27.2
26	Creating the typical workflow in HDFS and Map Reduce components of Hadoop map reduce jobs	CLO 11	T2:27.2 R2:21.67
27	Describe A file in HDFS distributing data processing.	CLO 11	T2:27.2
28	Understand the Hadoop limitations.	CLO 12	T2:27.3 R2:21.71
29	Understanding Map Reduce Fundamentals**.**	CLO 13	T2:27.4 R2:21.68
30	Define the building blocks of Hadoop map reduce	CLO 13	T2:27.7 R2:21.74
31	Describe working of Map Reduce engine	CLO 14	T2:27.12 R2:21.75
32-35	Understand the techniques to optimize Map Reduce jobs	CLO 14	T2:27.8 R2:21.72
36	Define Map Reduce phases	CLO 15	T2:27.8 R2:21.73
37	Understand the input and output forms of Map Reduce	CLO 16	T2:27.14 R2:21.78
38	Describe the Java API's to Map Reduce.	CLO 16	T2:27.19 R2:21.814
39	Describe the Real-time architecture of Map Reduce	CLO 17	T2:27.12 R2:21.82
40	Understand the PIG architecture.	CLO 17	T2:27.18 R2:21.82
41	Understand the commands of running pig.	CLO 18	T2:27.2 R2:21.64
42	Implementation of PIG scripts in processing big data analytics.	CLO 18	T2:27.2

Lecture No	Topics to be covered	Course Learning Outcomes (CLOs)	Reference
43	Define the Architecture of hive.	CLO 19	T2:27.2 R2:21.67
44-45	Describe the use Hive to create, alter, and drop databases.	CLO 20	T2:27.2

Prepared by:
Ms. G Sulakshana, Assistant Professor **HOD, CSE**