

**INSTITUTE OF AERONAUTICAL ENGINEERING** 

(Autonomous) Dundigal, Hyderabad -500 043

# **INFORMATION TECHNOLOGY**

# **COURSE DESCRIPTOR**

Course Title	BIG DATA AND BUSINESS ANALYTICS LABORATORY					
Course Code	ACS11	1				
Programme	<b>B.Tech</b>					
Semester	VII	CSI	E/IT			
Course Type	Core					
Regulation	IARE - R16					
			Theory		Practic	al
Course Structure	Lectu	res	Tutorials	Credits	Laboratory	Credits
	3 2				2	
Chief Coordinator	Ms. S Swarajya Laxmi, Assistant professor					
Course Faculty	Mr. D l	Mr. D Rahul, Assistant professor				

### I. COURSE OVERVIEW:

Big data and Business Analytics Laboratory demonstrates distributed computing environment .It includes hands on experience installation process of VMWAre, Pig, Hive, program setting in three operating modes like standalone, Pseudo code distributed, fully distributed, implementing basic LINUX commands, HDFS file management, MapReduce functions, Pig commands and Hive operations.

### **II.** COURSE PRE-REQUISITES:

Level	Course Code	Semester	Prerequisites	Credits
UG	ACS104	III	Database Management Systems Lab	2

### **III. MARKS DISTRIBUTION:**

Subject	SEE Examination	CIA Examination	Total Marks
Big Data and Business Analytics Laboratory	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:

Each laboratory will be evaluated for a total of 100 marks consisting of 30 marks for internal assessment and 70 marks for semester end lab examination. Out of 30 marks of internal assessment, continuous lab assessment will be done for 20 marks for the day to day performance and 10 marks for the final internal lab assessment.

**Semester End Examination (SEE):** The semester end lab examination for 70 marks shall be conducted by two examiners, one of them being Internal Examiner and the other being External Examiner, both nominated by the Principal from the panel of experts recommended by Chairman, BOS.

The emphasis	on the ev	neriments is	broadly h	ased on t	he following	criteria
The emphasis	on the ex	perments is	Uluality t	Jaseu oli i	ne tonowing	cincila.

20 %	To test the preparedness for the experiment.
20 %	To test the performance in the laboratory.
20 %	To test the calculations and graphs related to the concern experiment.
20 %	To test the results and the error analysis of the experiment.
20 %	To test the subject knowledge through viva – voice.

### **Continuous Internal Assessment (CIA):**

CIA is conducted for a total of 30 marks (Table 1), with 20 marks for continuous lab assessment during day to day performance, 10 marks for final internal lab assessment.

Table 1:	Assessment	pattern	for	CIA
----------	------------	---------	-----	-----

Component	L		
Type of Assessment	Day to day performance	Final internal lab assessment	I otal Marks
CIA Marks	20	10	30

#### **Continuous Internal Examination (CIE):**

One CIE exams shall be conducted at the end of the 15<sup>th</sup> week of the semester. The CIE exam is conducted for 10 marks of 3 hours duration.

Preparation	Performance	Algorithm	Results and Error Analysis	Viva	Total
2	2	2	2	2	10

### VI. HOW PROGRAM OUTCOMES ARE ASSESSED:

	Program Outcomes (POs)	Strength	Proficiency assessed by
PO 1	Engineering knowledge: Apply the knowledge of	1	Experiments and Viva
	mathematics, science, engineering fundamentals, and an		voice
	engineering specialization to the solution of complex		
	engineering problems.		

	Program Outcomes (POs)	Strength	Proficiency assessed by
PO 2	<b>Problem analysis:</b> Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics natural sciences and engineering sciences	2	Experiments and Viva voice
PO 3	<b>Design/development of solutions:</b> Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.	3	Mini Project
PO 5	<b>Modern tool usage:</b> 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.	2	Mini Project

**3 = High; 2 = Medium; 1 = Low** 

### VII. HOW PROGRAM SPECIFIC OUTCOMES ARE ASSESSED:

Strength	Proficiency assessed by
l, analyze and 2	Videos
s related to	
b design, big	
analysis and	
g complexity.	
lity to apply 1	Mini Project
ware service	
programming	
ty service for	
The ability to -	-
onments, and	
ths to be an	
	Strength       I, analyze and s related to b design, big analysis and g complexity.     2       lity to apply ware service programming cy service for     1       The ability to onments, and ths to be an     -

**3 = High; 2 = Medium; 1 = Low** 

# VIII. COURSE OBJECTIVES (COs):

The co	The course should enable the students to:					
Ι	Optimize business decisions and create competitive advantage with Big Data analytics.					
II	Practice java concepts required for developing map reduce programs.					
III	Impart the architectural concepts of Hadoop and introducing map reduce paradigm.					
IV	Practice programming tools PIG and HIVE in Hadoop eco system.					
V	Implement best practices for Hadoop development.					

# IX. COURSE LEARNING OUTCOMES (CLOs):

CLO Code	CLO's	At the end of the course, the student will	PO's Mapped	Strength of
		have the ability to:		Mapping
ACS111.01	CLO 1	Understand the installation of VMWare	PO1, PO3	2
ACS111.02	CLO 2	Understand and apply the Perform	PO1, PO3	3
		setting up and Installing Hadoop in its		
		three operating modes.		
ACS111.03	CLO 3	Implementing the basic commands of	PO2	2
		LINUX Operating System		
ACS111.04	CLO 4	Implement the file management tasks in	PO2, PO3,	3
		Hadoop.	PO5	

CLO Code	CLO's	At the end of the course, the student will	PO's Mapped	Strength of
		have the ability to:		Mapping
ACS111.05	CLO 5	Understand Map Reduce Paradigm.	PO2	2
ACS111.06	CLO 6	Apply Map Reduce program that mines	PO3, PO5	3
		weather data.		
ACS111.07	CLO 7	Implement matrix multiplication with	PO2, PO5	2
		Hadoop Map Reduce		
ACS111.08	CLO 8	Apply Map Reduce program that makes	PO1, PO2	2
		the dataset to be compressed.		
ACS111.09	CLO 9	Understand the installation of PIG.	PO1	2
ACS111.10	CLO 10	Understand Pig Latin scripts sort, group, PO1, PO2		2
		join, project, and filter your data.		
ACS111.11	CLO 11	Implement the Pig Latin scripts in two	PO2	2
		different modes		
ACS111.12	CLO 12	Understand the installation of HIVE	PO1, PO5	2
ACS111.13	CLO 13	Apply Hive to create, alter, and drop	PO5	3
		databases, tables, views, functions, and		
		indexes.		

**3** = High; **2** = Medium; **1** = Low

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

Course Learning	Program Outcomes (POs)									Program Specific Outcomes (PSOs)					
(CLOs)	PO1	PO2	PO3	PO4	PO5	PO6	<b>PO7</b>	<b>PO8</b>	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CLO 1	1		3										2		
CLO 2	2		3										2	1	
CLO 3		2													
CLO 4		2	2		2								2	2	
CLO 5		2													
CLO 6			2		3								1	1	
CLO 7		2			1								1	2	
CLO 8	2	1											2	1	
CLO 9	2														
CLO 10	1	3											1		
CLO 11		2													
CLO 12	1				3								1	1	
CLO 13					3								2	1	

**3** = High; **2** = Medium; **1** = Low

### XI. ASSESSMENT METHODOLOGIES – DIRECT

CIE Exams	PO1, PO2 PO3, PO5	SEE Exams	PO1, PO2 PO3, PO5	Assignments	-	Seminars	-
Laboratory Practices	PO1, PO2 PO3, PO5	Student Viva	PO1,PO2	Mini Project	PO3,PO5	Certification	-

### XII. ASSESSMENT METHODOLOGIES - INDIRECT

~	Early Semester Feedback	~	End Semester OBE Feedback
×	Assessment of Mini Projects by Experts		

## XIII. SYLLABUS

	LIST OF EXPERIMENTS							
Week-1	INSTALL VMWARE							
Installatio	Installation of VMWare to setup the Hadoop environment and its ecosystems.							
Week-2	HADOOP MODES							
a. Per i. ii. iii. b. Us	rform setting up and Installing Hadoop in its three operating modes. Standalone. Pseudo distributed. Fully distributed. e web based tools to monitor your Hadoop setup.							
Week-3	USING LINUX OPERATING SYSTEM							
Implemen update op	ting the basic commands of LINUX Operating System – File/Directory creation, deletion, erations.							
Week-4	FILE MANAGEMENT IN HADOOP							
a. Impler i. Add ii. Re iii. De Hint: A ty HDFS usi	<ul> <li>a. Implement the following file management tasks in Hadoop:</li> <li>i. Adding files and directories</li> <li>ii. Retrieving files</li> <li>iii. Deleting files</li> <li>Hint: A typical Hadoop workflow creates data files (such as log files) elsewhere and copies them into</li> <li>HDFS using one of the above command line utilities.</li> </ul>							
Run a bas	ic word count Map Reduce program to understand Map Reduce Paradigm.							
Week-6	MAPREDUCE PROGRAM 2							
Write a Mathematical Write a Mathematical Write a Mathematical American Ame	Iap Reduce program that mines weather data. Hint: Weather sensors collecting data every hour ocations across the globe gather a large volume of log data, which is a good candidate for vith Map Reduce, since it is semi structured and record-oriented.							
Week-7	MAPREDUCE PROGRAM 3							
Wook-8	PIC LATIN LANCUACE - PIC							
Installation	Installation of PIG.							
Week-9	PIG COMMANDS							
Write Pig	Write Pig Latin scripts sort, group, join, project, and filter your data.							
Week-10	PIG LATIN MODES, PROGRAMS							
a. Run th b. Run th	<ul> <li>a. Run the Pig Latin Scripts to find Word Count.</li> <li>b. Run the Pig Latin Scripts to find a max temp for each and every year.</li> </ul>							

Week-11	HIVE

Installation of HIVE.

### Week-12 HIVE OPERATIONS

Use Hive to create, alter, and drop databases, tables, views, functions, and indexes.

#### **Text 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, 1<sup>st</sup> Edition, 2013.
- 2. Tom White, "Hadoop: The Definitive Guide", O'Reilly, 3<sup>rd</sup> Edition, 2012.
- Rajiv Sabherwal, Irma Becerra- Fernandez, "Business Intelligence Practice, Technologies and Management", John Wiley, 1<sup>st</sup> Edition, 2011.

#### **Reference Books:**

1. Jay Liebowitz, "Big Data and Business Analytics Laboratory", CRC Press.

### **XIV. COURSE PLAN:**

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

Week	Topics to be covered	Course Learning Outcomes	Reference
No.		(CLOs)	
1	Install VMware	CLO9	T2:1
2	Hadoop Modes	CLO1, CLO2	T1:2.6-2.9
3	Using Linux Operating System	CLO1, CLO2	T1:4.1-4.2
4	File Management In Hadoop	CLO1, CLO2	T1:2.1-2.4
			T1:2.6-2.9
5	Mapreduce Program 1	CLO2, CLO4	T1:4.2-4.3
6	Mapreduce Program 2	CLO1, CLO2,CLO4	T1:4.4-4.7
7	Mapreduce Program 3	CLO2, CLO4,CLO5	T1:5.6- 5.10
8	Mapreduce Program 4	CLO2, CLO4	T1:5.6- 5.10
9	Pig Latin Language - Pig	CLO2, CLO4	T1:5.6- 5.10
10	Pig Commands	CLO10,CLO11	T2:1, 2
11	Pig Latin Modes, Pig Program	CLO10,CLO11	T2:2
12	Hive	CL011,CL012,CL013,CL014	T2:10
13	Hive Operations	CL011,CL012,CL013,CL014	T2:10

### XV. GAPS IN THE SYLLABUS - TO MEET INDUSTRY / PROFESSION REQUIREMENTS:

S No	Description	Proposed actions	Relevance with POs	Relevance with PSOs
1	Implementation of application that stores big data in MongoDB	Seminars	PO1,PO2,PO5	PSO 1
2	Implementation of application that stores big data in R language	Seminars / NPTEL	PO2,PO5	PSO 3

#### Prepared by:

Ms. S Swarajya Laxmi, Assistant Professor