Manipal School of Information Sciences (MSIS)
Manipal Academy of Higher Education, Manipal
Master of Engineering - ME (Big Data Analytics)

Course File

Course Name   :   Architecture of Big Data Systems

Course Code   :   BDA 5102

Academic Year   :   2024 – 2025

Semester   :   I

Name of the Course Coordinator   :   Mr. DEEPAK RAO B

Name of the Program Coordinator   :   Dr. PRATHVIRAJ N

Signature of Program Coordinator **with Date**	**Signature of Course Coordinator** **with Date**

Program Education Objectives (PEOs)

The overall objectives of the Learning Outcomes-based Curriculum Framework (LOCF) for **ME (Big Data Analytics)**, program are as follows.

PEO No.	Education Objective
PEO 1	Develop in depth understanding of the key technologies in data engineering, data science and business analytics.
PEO 2	Practice problem analysis and decision-making using machine learning techniques.
PEO 3	Gain practical, hands-on experience with statistics, programming languages and big data tools through coursework and applied research experiences.

Program Outcomes (POs)

By the end of the postgraduate program in Big Data Analytics, graduates will be able to:

PO1	Independently carry out research /investigation and development work to solve practical problems.
PO2	Write and present a substantial technical report/document.

PO3	Demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program.
PO4	Develop and implement big data analysis strategies based on theoretical principles, ethical considerations, and detailed knowledge of the underlying data.
PO5	Demonstrate knowledge of the underlying principles and evaluation methods for analyzing data for decision-making.

1. Course Plan

1.1   Primary Information

Course Name	:	Architecture of Big Data Systems
L-T-P-C	:	3-0-0-3
Contact Hours	:	36 Hours
Pre-requisite	:	Programming with Python or Java

1.2   Course Outcomes (COs), Program outcomes (POs) and Bloom's Taxonomy Mapping

CO	At the end of this course, the student should be able to:	No. of Contact Hours	Program Outcomes (PO's)	BL
CO1	**Apply** various techniques to examine different types of data and understand lambda architecture.	12	PO3	3
CO2	**Apply** different tools and frameworks of Hadoop eco-system	9	PO4	4
CO3	**Apply** Spark engine to process real-time data.	9	PO4	4
CO4	**Design** applications to handle batch and streaming data using Hadoop and Spark tools.	6	PO5	6

1.3   Assessment Plan

2. Components	Mid Semester	Flexible Assessments (2 – 3 in number)	End semester/ Makeup examination
Duration	90 minutes	To be decided by the faculty.	180 minutes
Weightage	30%	20%	50%
Typology of questions	Applying; Analyzing and Evaluating	Applying; Analyzing. Evaluating.	Applying; Analyzing; Evaluating.
Pattern	Answer all 5 questions of 10 marks each. Each question may have 2 to 3 parts of 3/4/5/6/7 marks.	**Quiz / Test:** 10% **Assignment 1:** (Big Data and Lambda Architecture, 5% weightage) **Assignment 2:** (Spark Streaming, 5% weightage)	Answer all 10 full questions of 10 marks each. Each question may have 2 to 3 parts of 3/4/5/6/7 marks.
Schedule	As per academic calendar.	**Quiz / Test:** September 2024 **Assignment 1:** October 2024 **Assignment 2:** November 2024	As per academic calendar.

		Introduction – Big Data, Lambda Architecture, Batch Processing, Spark		Comprehensive examination covering the full syllabus. Students are expected to answer all questions.
Topics covered				

Components	Internal Test 1	Internal Test 2	Flexible Assessments (2 – 3 in number)	End semester/ Makeup examination
Duration	90 minutes	90 minutes	To be decided by the faculty.	180 minutes
Weightage	0.2	0.2	0.1	0.5
Typology of questions	Applying; Analyzing.	Applying; Analyzing.	Applying; Analyzing.	Applying; Analyzing; Evaluating.
Pattern	Answer all 5 questions of 10 marks each. Each question may have 2 to 3 parts of 3/4/5/6/7 marks.	Answer all 5 questions of 10 marks each. Each question may have 2 to 3 parts of 3/4/5/6/7 marks.	**Assignment:** (Big Data and Lambda Architecture, Hadoop ad Spark Assignments)	Answer all 10 full questions of 10 marks each. Each question may have 2 to 3 parts of 3/4/5/6/7 marks.

Schedule	As per academic calendar.	As per academic calendar.	**Assignment submission:** November 2023	As per academic calendar.
Topics covered	Introduction – Big Data, Lambda Architecture, Batch Processing	Hadoop eco system, Realtime data processing using Spark.		Comprehensive examination covering the full syllabus. Students are expected to answer all questions.

2.1   Lesson Plan

L. No.	TOPICS	Course Outcome Addressed
L0	Course delivery plan, Course assessment plan, Course outcomes, Program outcomes, CO-PO mapping, reference books	---
L1	Definition of Big Data and its sources	CO1
L2	Characteristics of Big Data	CO1
L3	Challenges of Big Data	CO1
L4	Drawbacks of traditional data handling systems	CO1
L5	Introduction to Lambda architecture	CO1
L6	Requirements and responsibilities of batch layer	CO1
L7	Requirements and responsibilities of service layer	CO1

L8	Requirements and responsibilities of speed layer	CO1
L9	Introduction to Hadoop	CO1
L10	Architecture of HDFS	CO1
L11	Name Node and Data Nodes	CO1
L12	How to ensure high availability of data and services	CO1
L13	Introduction to YARN	CO2
L14	Components of YARN	CO2
L15	Introduction to Hadoop map-reduce	CO2
L16	Responsibilities of Map task	CO2
L17	Responsibilities of Reduce task	CO2
L18	Types of failures in map reduce jobs	CO2
L19	Recovery from failures	CO2
L20	Writing map reduce programs	CO2
L21	Writing map reduce programs	CO2
L22	Introduction to Spark Engine	CO3
L23	Difference between Spark and Hadoop	CO3
L24	Architecture of Spark	CO3
	Mid Semester Evaluation	CO1, CO2, CO3
L25	Different components and responsibilities of Spark	CO3
L26	Different stages of running jobs in Spark	CO3
L27	Actions and Transformations	CO3

L28	RDDs and Data Frames	CO3
L29	Introduction to Data Streaming	CO3
L30	Different models in Data Streaming	CO3
L31	Challenges in data streaming	CO4
L32	Developing programs for real time data handling	CO4
L33	Developing programs for real time data handling	CO4
L34	Develop programs for machine learning	CO4
L35	Developing programs using Hadoop tools	CO4
L36	Develop programs using Spark	CO4

2.2   References

1. Big Data: Principles and best practices of scalable real-time data systems - Nathan Marz and James Warren. Manning Publisher
2. Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale – Tom White, O'Reilly Publication 4th Edition.
3. Spark: The Definitive Guide: Big Data Processing Made Simple – Bill Chambers, Matei Zaharia, O'Reilly Publication 1st Edition.

2.3   Other Resources (Online, Text, Multimedia, etc.)

1. Web Resources: Blog, Online tools and cloud resources.
2. Journal Articles.

2.4   Course Timetable

1st Semester BDA			Room: LG1 LH8		Lab: Data Science		
	9 – 10	10 - 11	11 – 12	12	2 - 3	3 - 4	4 - 5
MON							
TUE						ABD	
WED							
THU						ABD	
FRI							
SAT		ABD					

2.5   Assessment Plan

Cos		Marks & Weightage			
CO No.	CO Name	IT-1 (Max. 50)	Assignment (Max. 10)	End Semester (Max. 100)	CO wise Weightage
CO1	**Apply** various techniques to examine different types of data and understand lambda architecture.	20	5	30	**0.32**
CO2	**Apply** different tools and frameworks of Hadoop eco-system	20	10	30	**0.35**

		Apply Spark engine to process real-time data.	10	5	20	0.21
CO3		**Apply** Spark engine to process real-time data.	10	5	20	**0.21**
CO4		**Design** applications to handle batch and streaming data using Hadoop and Spark tools.	-		20	**0.12**
		Marks (weightage)	**0.3**	**0.2**	**0.5**	**1.0**

- In-semester Assessment is considered as the Internal Assessment (IA) in each subject for 50 marks, which includes the performances in class / tutorial participation, assignment work, lab work, class tests, mid-term tests, quizzes etc.
- End-semester examination (ESE) for each theory subject is conducted for a maximum of 100 and the same will be scaled down to 50.
- End-semester mark for a maximum of 50 and IA marks for a maximum of 50 are added for a maximum of 100 marks to decide upon the grade in a subject.

2.6   Assessment Details

The assessment tools to be used for the Current Academic Year (CAY) are as follows:

SI. No.	Tools (TLP)	Weightage	Frequency	Details of Measurement (Weightage/Rubrics/Duration, etc.)
1	Mid Semester	0.3	1	• Performance is measured using sessional attainment level. • Reference: question paper and answer scheme.

2	Assignments	0.2	1	• Performance is measured using assignments/quiz attainment level. • Assignments/quiz are evaluated for a maximum of 20 marks.
				• Mid semester is assessed for a maximum of 50 marks and scaled down to 30 marks
3	ESE	0.5	1	• Performance is measured using ESE attainment level. • Reference: question paper and answer scheme. • ESE is assessed for a maximum of 100 marks and scaled down to 50 marks.

2.7   Course Articulation Matrix

CO	PO1	PO2	PO3	PO4	PO5
CO1			Y		
CO2				Y	
CO3				Y	
CO4					Y
Average Articulation Level			*	*	*