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 Lab

--- --- ---

Course Code : BDA 5152

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 Signature of Course Coordinator

--- ---

with Date 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 Bid Data Systems Lab

--- --- ---

L-T-P-C : 0-0-3-1

Contact Hours : 36 Hours

Pre-requisite : Programming with Python or Java

1.2 Course Outcomes (COs)

CO At the end of this course, the student should be able No. of Contact Program Outcomes BL

--- --- --- --- ---

to: Hours (PO's)

Use data extraction tools to ingest various types of data

--- --- --- --- ---

CO1 into big data systems. 6 PO3 3

CO2 Experiment with different tools and frameworks of Hadoop eco-system. 12 PO4 4

CO3 Experiment with Spark Engine to process real-time data. 12 PO4 4

CO4 Design applications to handle batch and streaming data using Hadoop and Spark tools. 6 PO5 6

1.3 Assessment Plan

Components	Lab Test	Flexible Assessments	End semester/ Makeup examination
	(4- 5 in number)		
Duration	90 minutes	To be decided by the faculty.	180 minutes
Weightage	0.3	0.2	0.5
Typology of questions	Applying; Analyzing. Evaluating.	Applying; Analyzing.	Applying; Analyzing; Evaluating.
Pattern	Answer all the questions. Maximum marks 30.	Assignment: Data Extraction, Batch processing and handling real time	Answer all the questions. Maximum marks 50.
Schedule	As per academic calendar.	Assignment submission: November 2024	As per academic calendar.
Topics covered	HDFS, SQOOP, HIVE, Map-Reduce program, Spark, Data Frames and Data Streaming		Comprehensive examination covering the full syllabus.

1.4 Lesson Plan

L. No.	TOPICS	Course Outcome Addressed
L0	Course delivery plan, Course assessment plan, Course outcomes, Program outcomes, CO-PO mapping, reference books	---
Lab1	Installing and configuring MySQL	CO1

Lab2 Write script to handle data in MySQL, Shell commands for HDFS CO1

Lab3 Data ingestion using SQOOP CO2

Lab4 Data Analysis using HIVE CO2

Lab5 Introduction to Map-Reduce program CO2

Lab6 Experiment with Map-Reduce programs CO2

IT1 Internal lab test CO1, CO2

Lab7 Introduction to Spark Program CO3

Lab8 Data handling with Spark RDD CO3

Lab9 Data Analysis with Spark Data Frames CO3

Lab10 Data Streaming applications with Spark Structured Streams CO3

Lab11 Develop batch processing application CO4

Lab12 Develop real time processing application CO4

1.5 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.

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

- 1. Web Resources: Blog, Online tools and cloud resources.

- 2. Journal Articles.

1.7 Course Timetable

st 1 Semester Big Data Analytics Room: LG1 LH 3 Lab: Data Science Lab

--- --- --- --- --- --- --- --- ---

9-10 10-11 11-12 12-1 1-2 2-3 3-4 4-5

MON

TUE

WED ABD LAB

THU

FRI

SAT

1.8 Assessment Plan

Cos Marks & weightage

--- --- --- --- --- ---

CO No. CO Name Lab Test (Max. 30) Assignment (Max. 20) End Semester (Max. 50) CO wise Weightage

CO1 Use data extraction tools to ingest various types of data into big data systems. 5 5 5 0.15

CO2 Experiment with different tools and frameworks of Hadoop eco-system. 10 5 15 0.3

CO3 Experiment with Spark Engine to process real-time data. 10 5 20 0.35

CO4 Design applications to handle batch and streaming data using Hadoop and Spark tools. 5 5 10 0..2

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 lab subject is conducted for a maximum of 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.

1.9 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.)

--- --- --- --- --- ---

- Performance is measured using sessional attainment level.

1 Sessional 0.3 2 - Reference: question paper and answer scheme.

- Each test is assessed for a maximum of 30 marks.

2 Assignments 0.2 - - Performance is measured using assignments/quiz attainment level.

- Assignments/quiz are evaluated for a maximum of 20 marks.

ESE 0.5 1 - - Performance is measured using ESE attainment level.

3 Reference: question paper and answer scheme.

- ESE is assessed for a maximum of 50 mark.

1.10 Course Articulation Matrix

CO PO1 PO2 PO3 PO4 PO5

--- --- --- --- --- ---

CO1 Y

CO2 Y

CO3 Y

--- --- --- ---

CO4 Y

Average Articulation Level * * *

Note: Enter correlation levels 1, 2 or 3 as defined below for both CO-PO mapping:

- 1: Slight (Low), 2: Moderate (Medium), 3: Substantial (High)

- If there is no correlation, apply "-"