Course Code	Course Name				
MCAL31	Big Data Analytics and Visualization Lab				
Contact Hours (Per Week)	Credits Assigned	Examination Scheme (Marks)			
		Term Work	Practical	Oral	Total
02	01	50	30	20	100

Pre-requisite:
Lab Course Objectives: Course aim to

Sr. No.	Course Objective
1	Understand Various Components of Hadoop for instance Hadoop2.x, HDFS, Map Reduce
2	Understand and gain knowledge of NoSQL DB and Data Modelling Concept
3	Teach Hadoop Ecosystem Projects Hive and Pig and its Programming Modules.
4	Learn Functional programming in spark and execute and create spark applications.
5	Teach Data Visualization and its importance using Tableau

Lab Course Outcomes (CO): On successful completion of course learner/student will be able to

Sr. No.	Course Outcome	Bloom Level
CO1	Demonstrate HDFS Commands in Hadoop	Understanding
CO2	Apply Map Reduce Programming Paradigm to solve the algorithmic problems	Applying
CO3	Build No SQL Database and Query it Using Mongo DB	Applying
CO4	Analyze the Data Using Hadoop Ecosystem Projects: Hive and Pig	Analyzing
CO5	Explain RDD and Data Frame Creation in Apache Spark	Evaluating
CO6	Create various Visualizations using Tableau.	Creating

Course Contents:

Module No.	Detailed Contents	Hrs.	CO No.	Ref No.
1	Set up and Configuration Hadoop Using Cloudera / in Linux environment Creating a HDFS System with minimum 1 Name Node and 1 Data Nodes HDFS Commands **Self-Learning Topics**: Creating a HDFS System with minimum 1 Name Node and multiple Data Nodes	2	1	1,2
2	Map Reduce Programming Examples Word Count. Union, Intersection Matrix Multiplication **Self-Learning Topics**: Natural Join Programming Example	4	2	1,3
3	Mongo DB: Installation and Creation of database and Collection CRUD Document: Insert, Query, Update and Delete Document. **Self-Learning Topics:** HBASE Commands	4	3	4

Module No.	Detailed Contents	Hrs.	CO No.	Ref No.
4	Hive: Introduction Creation of Database and Table, Hive Partition, Hive Built in Function and Operators, Hive View and, HiveQL – where, order by, group by, join **Self-Learning Topics**: Configure Hive Metastore to MySQL	4	4	5,6
5	Pig: Pig Latin Basic Pig Shell, Pig Data Types, Creating a Pig Data Model, Reading and Storing Data, Pig Operations **Self-Learning Topics:** Writing UDF (user-defined functions) in Apache Pig	4	4	5,6
6	Spark: RDD, Actions and Transformation on RDD , Ways to Create -file, data in memory, other RDD. Lazy Execution, Persisting RDD **Self-Learning Topics:** Machine Learning Algorithms using pySpark	4	5	5,7,8
7	Visualization: Connect to data, Build Charts and Analyze Data, Create Dashboard, Create Stories using Tableau **Self-Learning Topics:** Forecasting and trend analysis using Tableau	4	6	9

Assessment:

Term Work: Will be based on Continuous Assessment

- Laboratory work will be based on the syllabus with minimum 10 experiments. The experiments should be completed in the allotted time duration.
    - Experiments 40 marks
    - Attendance 10 marks
- Practical will be evaluated by the subject teacher and documented according to a rubric

End Semester Practical Examination:

Practical and oral examination will be based on suggested practical list and entire syllabus.

Reference Books:

Reference No	Reference Name
1	Tom White, "HADOOP: The definitive Guide" O Reilly 2012, Third Edition, ISBN: 978-1-449-31152-0
2.	Rohit Menon, "Cloudera Administration Handbook" Released July 2014 Publisher(s): Packt Publishing, ISBN: 9781783558964
3	Chuck Lam, "Hadoop in Action", Dreamtech Press 2016, First Edition ,ISBN:13 9788177228137
4	Kyle Banker, "MongoDB in Action" December 2011 ISBN 9781935182870
5	Shiva Achari," Hadoop Essential " PACKT Publications, ISBN 978-1-78439-668-8
6	RadhaShankarmani and M. Vijayalakshmi ,"Big Data Analytics "Wiley Textbook Series, Second Edition, ISBN 9788126565757
7	Jeffrey Aven,"Apache Spark in 24 Hours" Sam's Publication, First Edition, ISBN: **0672338513**
8	Bill Chambers and MateiZaharia,"Spark: The Definitive Guide: Big Data Processing Made Simple "O'Reilly Media; First edition, ISBN-10: 1491912219;
9	James D. Miller," Big Data Visualization" PACKT Publications. ISBN-10: 1785281941

Web References:

Reference No	Reference Name
1	https://hadoop.apache.org/docs/stable3/hadoop-project-dist/hadoop-hdfs/HdfsUserGuide.html
2	https://shorturl.at/4ij9O
3	https://www.mongodb.com/try/download/community
4	https://www.mongodb.com/docs/manual/crud
5	https://hive.apache.org/
6	https://pig.apache.org/
7	https://spark.apache.org/documentation.html
8	https://help.tableau.com/current/pro/desktop/en-us/default.htm

Suggested list of experiments:

Practical No	Problem Statement
1	**HDFS:** List of Commands (ls, mkdir, touchz, copy from local/put, copy to local/get, move from local, viewing file content(cat, head, tail),cp, rmr, du, dus, stat)
2	**Map Reduce:** **1.** Write a program in Map Reduce for WordCount operation. **2.** Write a program in Map Reduce for Union operation. **3.** Write a program in Map Reduce for Intersection operation. **4.** Write a program in Map Reduce for Matrix Multiplication
3	**MongoDB :** 1. Installation 2. Sample Database Creation 3. Query the Sample Database using MongoDB querying commands 4. Create Collection 5. Insert Document 6. Query Document 7. Delete Document 8. Indexing
4	**Hive:** 1. Hive Data Types 2. Create Database & Table in Hive 3. Hive Partitioning 4. Hive Built-In Operators 5. Hive Built-In Functions 6. Hive Views **7.** HiveQL : Select Where, Select OrderBy, Select GroupBy, Select Joins
5	**Pig:** 1. Pig Latin Basic 2. Pig Data Types, 3. Download the data 4. Create your Script 5. Save and Execute the Script 6. Pig Operations : Diagnostic Operators, Grouping and Joining, Combining & Splitting, Filtering, Sorting
6	**Spark**: 1. Downloading Data Set and Processing it Spark 2. Word Count in Apache Spark.
7	**Visualization using Tableau:** Tableau: Tool Overview, Importing Data, Analyzing with Charts, Creating Dashboards, Working with maps, Telling Stories with tableau.