JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY ANANTAPUR
(Established by Govt. of A.P., ACT No.30 of 2008)
ANANTHAPURAMU – 515 002 (A.P) INDIA

MASTER OF COMPUTER APPLICATIONS

Course Code	BIG DATA TECHNOLOGIES LABORATORY	L	T	P	C
21F00307		0	1	2	2
	Semester	III			

Course Objectives:

- Apply quantitative modeling and data analysis techniques to the solution of real-world business problems, communicate findings, and effectively present results using data visualization techniques.
- Apply principles of Data Science to the analysis of business problems.

Course Outcomes (CO):

- Understand and implement the basics of data structures like Linked list, stack, queue, set and map in Java.
- Demonstrate the knowledge of big data analytics and implement different file management task in Hadoop.
- Understand Map Reduce Paradigm and develop data applications using variety of systems.
- Analyze and perform different operations on data using Pig Latin scripts.
- Illustrate and apply different operations on relations and databases using Hive.

List of Experiments:

Week 1: Hadoop Installation on a)Single Node and SPARK Installation, Launch a cloud instance for AWS instance on Centos 7

Week 2: Design a distributed application using MapReduce which processes a log file of a system. List out the users who have logged for maximum period on the system. Use simple log file from the Internet and process it using a pseudo distribution mode on Hadoop platform.

Week 3: Design and develop a distributed application to find the coolest/hottest year from the available weather data. Use weather data from the Internet and process it using MapReduce.

Week 4: Write an application using HBase and HiveQL for flight information system which will include 1) Creating, Dropping, and altering Database tables, 2) Creating an external Hive table to connect to the HBase for Customer Information Table, 3) Load table with data, insert new values and field in the table, Join tables with Hive, 4) Create index on Flight information Table, and 5) Find the average departure delay per day in 2008.

Week 5: Display the hierarchical structure of your data by generating Trees, graphs and network visualization. Install and Run Pig then write Pig Latin scripts to sort, group,join,project and filter the data. Install and Run Hive then use Hive to Create, alter and drop databases, tables, views, functions and Indexes.

Week 6: Input file contains a series of tweets made by few people. Do a word count on the text object value Hint: Json Parsing in python – this sample snippet can be used within Map to read the JSON

Week 7: Reading different types of data sets (.txt, .csv) from web and disk and writing in file in specific disk location. And Reading Excel,XML data sheets in R. Using with and without R objects on console, mathematical functions on console create R objects for calculator application and save in a specified location in disk.

Write an R script to find basic descriptive statistics using summary,str, quartile unction on mtcars& cars datasets and to find subset of dataset by using subset (),aggregate () functions on dataset.

Week 8:

Implementing data visualization using R : Find the data distributions using box and scatter plot, Find the outliers using plot and Plot the histogram, bar chart and pie chart on sample data.