

COURSE PLAN

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Department** | **:** | Data Science Engineering And Computer Applications | | | | | |
| **Course Name & code** | **:** | DSE-3264 & Big Data Analytics Laboratory | | | | | |
| **Semester & branch** | **:** | VI Sem & BTech Data Science & Engineering | | | | | |
| **Name of the faculty** | **:** | Rashmi Laxmikant Malghan , Nandini Kumari | | | | | |
| **No of contact hours/week:** | |  | **L** | **T** | **P** | **C** |  |
|  | |  | 0 | 0 | 3 | 1 |  |

Course Outcomes (COs)

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***At the end of this course, the student should be able to:*** | **No. of Contact**  **Hours** | **Marks** |
| CO1: |  |  |  |
| CO2: |  |  |  |
| CO3: |  |  |  |
| CO4: |  |  |  |
| CO5: |  |  |  |
|  | **Total** | 36 | 100 |

***(Page 1 of 4)***

**MIT/GEN/F-01/R2**

Assessment Plan

|  |  |
| --- | --- |
| **1. Continuous Evaluation** | 60% |
| * Evaluations will be conducted every week, total marks = 60 marks * Project will evaluated for 20 Marks, * (10\*2)= 20 Marks for 2 viva, * Weekly Evaluation = 20 Marks. | |
| **2. Lab Examination** | 40% |
| * Examination of 3 hours duration (Max. Marks: 40) Performance Analysis : 15 marks, Program execution :25 marks. | |

Lesson Plan

|  |  |  |
| --- | --- | --- |
| **L. No.** | **Topics** | **Course**  **Outcome Addressed** |
| **L1** | 1. Implement the following file management tasks in Hadoop: 2. Creation of directories 3. Loading of new Files 4. Retrieving the data from the file 5. Renaming, updating and Deleting files      1. Interacting with HDFS using command line interface to understand the basic working structure of Hadoop cluster. 2. Use web based tools to monitor your Hadoop setup. |  |
| **L2** | 1. Implement the Map Reduce Program to find mutual friends between each individual and store them (Similar case study as used by Facebook), 2. Develop a Map Reduce program to analyze the given Earthquake data and generate statistics report based on the region 3. Identify the maximum earthquake occurrence region. 4. List the different region names |  |
| **L3** | Write a program to understand the map Reduce algorithm to online music listeners Sound.XYZ or Hangama website or any other website) to note done the following :     1. Maximum Time the track was downloaded. 2. Count of Unique Listeners 3. Number of Listeners usage of same track 4. Maximum Time the track was shared with others. |  |
| **L4** | 1. Implement a MapReduce program to identify and analyze the tags specified (Genre, Adventure, Comedy, Romance, Action etc) using movie lens data find the Low, High and average for any 5 columns from dataset. 2. Create a MapReduce program to analyze Titanic ship data and perform the following: 3. Display the Gender wise count of passengers travelling in Titanic ship 4. To display the age wise count of passengers with different class labels as a) Children b) Adult c) Senior Citizens. 5. Identify the Death rate and survival rate based on Gender. |  |
| **L5** | 1. Write a program to compute the average, minimum and maximum recorded temperature by year wise using the PIG Latin commands (PIG script). Hint: Temperature datasheet yearly collected through sensors 2. Develop a program to perform the following using the sales order dataset 3. To count the maximum, minimum number of units sold region wise (East, West and Central region) 4. Number of Unique Items sold along with representative name |  |
| **L6** | 1. Develop a java application to fetch the sum of employee salaries department wise using Spark 2. Use Spark Scala to find the tuples based on class sensitivity by creating an object, copying it, and testing for equality. 3. Demonstrate the basic math operation like addition, multiplication , subtraction and division and perform aggregate functions (approx\_count\_distinct(), collect\_list(), collect\_set(), avg(), count(), countDistinct()) by utilizing Spark Scala |  |
| **L7** | 1. Demonstrate the built in functions using Scala by creating an array, perform following functions:   i) Non-Parameterized Function  ii) Parameterized Function  iii) Recursion Function  iv) Function Parameter with Default Value  v) Anonymous Functions  Vi) Functions with map method  vii) Currying Functions  viii) Nested Functions   1. i) Use spark Scala to demonstrate the add & multiplication of two numbers using Currying function.   ii) Execute Scala program of Partial function using Case statement by creating array.  iii) Create Scala program to demonstrate the function overloading by changing the number of parameters  iv) Scala program of Multiple Nested Function |  |
| **L8** | Use HBASE to perform put(), get(), scan() Command in HBase in shell and the java application and execute the below mentioned list  1) Write Data to HBase Table: Shell  2) Read Data from HBase Table: Shell  3) Write Data to HBase Table: JAVA API  4) Read Data from HBase Table: JAVA API |  |
| **L9** | 1. Create a student database considering the attributes as student id, Student name, Grade of Previous semester and perform the basic database (Using HiveQL operations) like 1) Create , Update, Delete, Drop operations 2) Fetch the data in both ascending and descending order 3) Count and display the number of students based on “Grade wise” . 2. Using the Food dataset execute the following queries using HiveQL: 3. Compute the total content of calories in the food. 4. Sort the data based on attribute of “Rating” and perform the basic operations of database with views, indexes and functions. 5. List down and count the different food labels (Protein, Vitamin, Fiber, Fat, Sodium and Sugar) with inclusion of “Manufacture” attribute. |  |
| **L10** | 1. Create table “Hospital Employee” with 6 attributes such as Employee Name, ID, Address, Department, Salary, Duty Shifts 2. Load data into employees table from Hospital Employees.txt file. 3. Display the list of corresponding employee id, name and their address. 4. Create a new table department with the attributes Eno and Ename. Now display the department id in which each employee works along with the employee id and their age. |  |
| **L11** | 1. Develop the predictive analysis algorithms and utilizing R perform any task for wine dataset. 2. Find which attribute has more correlation with quality of wine. 3. Identify the premium quality and non-premium quality wine based on its maximum and minimum number respectively. 4. Compute the mean value of pH 5. Understand, relate and create basic commands of MongoDB to perform query ,Consider the dataset Vehicle\_registration ,   https://data.gov.in/node/2865481/datastore/export/json The dataset consists of total vehicle registered in India state wise. Write query for the following   1. How many vehicles have been registered in state Andhra 2. Display the number of buses have been registered in state Karnataka? 3. List the different category of vehicles based on region wise. |  |
| **L12** | Examination |  |

***(Page 2 of 4)***

**MIT/GEN/F-01/R2**

References:

|  |  |
| --- | --- |
| 1. | Vignesh Prajapathi, *Big Data Analytics with R and Hadoop*, Packt Publishing, 2013. |
| 2. | Holden Karau,  [Andy Konwinski](https://www.amazon.com/s/ref=dp_byline_sr_book_2?ie=UTF8&text=Andy+Konwinski&search-alias=books&field-author=Andy+Konwinski&sort=relevancerank) , [Patrick Wendell](https://www.amazon.com/s/ref=dp_byline_sr_book_3?ie=UTF8&text=Patrick+Wendell&search-alias=books&field-author=Patrick+Wendell&sort=relevancerank), [Matei Zaharia](https://www.amazon.com/Matei-Zaharia/e/B00CYNEOLC/ref=dp_byline_cont_book_4), *Learning Spark: Lightning-Fast Big Data Analysis,* 1st Edition, O’Reilley Media Inc, 2015. |
| 3. | Jay Liebowitz, ―Big Data And Business Analytics Laboratory, CRC Press. |
| 4. | Michael Minnelli, Michele Chambers, *Big Data Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses*, Wiley India Pvt. Ltd., 2013. |
| 5. | Big data Analytics- Introduction to Hadoop, Spark and Machine Learning by Raj Kamal & Preeti Saxena |
| 6. | Big Data Analytics for Healthcare - Jimeng Sun & Chandan K. Reddy |
| 7. | Arvind Sathi, Big Data Analytics, MC Press, LLC, 2012. |

***(Page 3 of 4)***

**MIT/GEN/F-01/R2**

|  |  |
| --- | --- |
|  |  |
| 6. | Click or tap here to enter text. |
| 7. | Click or tap here to enter text. |

|  |  |
| --- | --- |
| **Submitted by:** | Rashmi Laxmikant Malghan, Nandini Kumari |

(Signature of the faculty)

|  |  |
| --- | --- |
| **Date:** | 20-01-2023 |

|  |  |
| --- | --- |
| **Approved by:** | Dr. Radhika M Pai |

(Signature of HOD)

|  |  |
| --- | --- |
| **Date:** |  |

FACULTY MEMBERS TEACHING THE COURSE (IF MULTIPLE SECTIONS EXIST):

|  |  |  |  |
| --- | --- | --- | --- |
| **FACULTY** | **SECTION** | **FACULTY** | **SECTION** |
| Dr. Nandini Kumari | A | Dr. Rashmi Laxmikant Malghan | B |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

\*\*\*\*\*\*\*\*\*

***(Page 4 of 4)***

**MIT/GEN/F-01/R2**