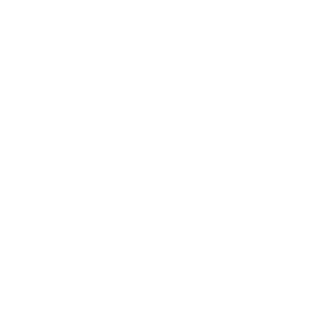
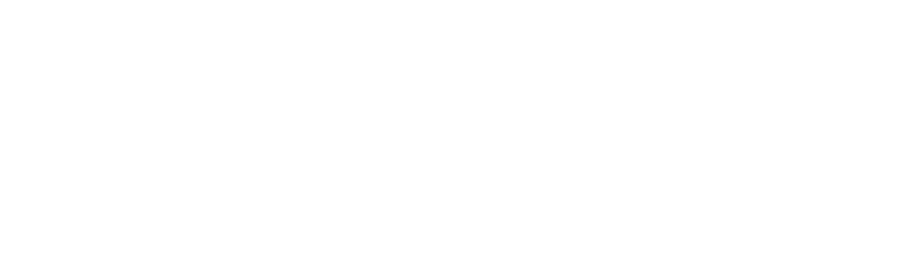
**Annexure ‘CD – 01’**

**FORMAT FOR COURSE CURRICULUM**

**Course Title:** **Fundamental of Big Data Analytics** 

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **L** | **T** | **P/S** | **SW/FW** | **TOTAL CREDIT UNITS** |
| 3 | 0 | 2 | 0 | 4 |

**Credit Units: 4**

**Course Level: UG**

**Course Code:443**

**Course Objectives:**

This course is designed for students who have no previous knowledge of data analytics but wish to acquire these skills. The students will learn how to analyze large data sets and identify patterns that will improve any company’s and organization decision-making process. After completing the course, they will be able to**.** 1 Capture, categorize, simplify, normalize and prepare data to be processed

2- Work with and analyze large data sets

3- Visually represent analysis’s conclusions to technical and non technical audiences

4 - Use the most common algorithms, to make sense of large amounts of data, which are applicable to most business and management problems

**Pre-requisites:**

**Knowledge of database and OOP concepts**

|  |  |
| --- | --- |
| **Course Contents/Syllabus:** | **Weightage (%)** |
| **Module I : Introduction to Bigdata and Hadoop** | **20%** |
| Understanding What is Bigdata, Analysing limitations and solutions of existing data analytics, Understanding Hadoop and its features, Hadoop Ecosystem, Understaing 2.X core components, 3.x core components, difference between 1.x,2,x and 3,x hadoop components, Performing Read and Write in Hadoop, Rack Awareness policy. |
| **Module II : Hadoop Architecture and HDFS** | **20%** |
| Hadoop 2.x cluster architecture, Federation Architecture, High Availability Architecture, Modes of hadoop cluster, Understading basis hadoop commands, Hadoop 2.x configuraton files and its parameters, mapreduce flow chart, Analysing different use cases of mapreduce programming, difference between traditional and mapreduce way, Execution flow of mapreduce applications, Input split in mapreduce, Combiner and Patitioners in mapreduce. Input Formats in Hadoop, Counters, Joins and Custom Input Formats in Mapreduce.  **.** |
| **Module III : Compute, Storage & Networking** | **20%** |
| Understanding the concepts of Pig and Pig latin, Analyzing datasets using pig,Understading hive, hive demand and its components, architecture and applications of Hive, Understanding Hbase and its architecture, Sqoop, Flume, Oozie, Zookeerper. |
| **Module IV: Security, Identify and Access Management (IAM)** | **20%** |
| Introduction: Sources, modes of availability, inaccuracies, and uses of data. Data Objects and Attributes: Descriptive Statistics; Visualization; and Data Similarity and Dissimilarity. Pre-processing of Data Cleaning for Missing and Noisy Data; Data Reduction – Discrete Wavelet Transform, Principal Component Analysis, Partial Least Square Method, Attribute Subset Selection; and Data Transformation and Discretization. |
| **Module V: Elasticity and Management Tools** | **20%** |
| . Machine Learning Techniques for data analysis, Introduction and Concepts of Supervised Learning with Regression and Classification techniques, Naïve Bayseian Classifier, K-means Clustering, Association Rules, Decision Trees, Time Series Analysis, Text Analytics |

**Student Learning Outcomes:**

1. Able to understand the importance of cloud computing

2. Design and development of cloud applications using public clouds

3. Able to understand the difference between the web services and web applications

4. Understanding of cloud service models IaaS, PaaS, SaaS

5. Able to understand that how virtualization enables cloud computing

**List of practical’s:**

1. Write a MapReduce script to count the occurrence of each word in a file.

2. Write a MapReduce script to find the max and min temperature recorded in last 50 years.

3. Write a MapReduce script to get the name of the customer, how many of transactions a customer has done and how

much amount of transactions a particular customer has made.

4. Analyze document database such as MongoDB using CRUD operations

5. Write a pig script to analyses the twitter data.

6. Write a hive script to analyses last 10 years of crime data.

**Pedagogy for Course Delivery:**

The class will be taught using theory and case studies of latest processors. Apart from assigning the casestudies, the course instructor will cover the quantitative approach for classification of modern processors.

**Assessment/ Examination Scheme:**

|  |  |  |
| --- | --- | --- |
| **Theory L/T (%)** | **Lab/Practical/Studio (%)** | **Total** |
| **70%** | **30%** | **100%** |

**Theory Assessment (L&T):**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Continuous Assessment/Internal Assessment** | | | | | **End Term Examination** |
| **Components (Drop down)** | **ATT** | **HA** | **VIVA** | **CT** |  |
| **Weightage (%)** | 5 | 8 | 7 | 10 | 70 |

**Lab/ Practical/ Studio Assessment:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Continuous Assessment/Internal Assessment 40** | | | | **End Term Examination** | | |
| **Components (Drop down** | **Attendance** | Lab Record | Performance | viva | Performance | viva | **total** |
| **Weightage (%)** | 5 | 10 | 15 | 5 | 30 | 30 | 60 |

**Text Reading:**

* **AWS Cloud Practitioner:**
* **Cloud Computing : A Practical Approach by Anthony T. Velte Toby J. Velte, Robert Elsenpeter, 2010 by The McGraw-Hill.**
* **Cloud Computing: SaaS, PaaS, IaaS, Virtualization and more. by Dr. Kris Jamsa.**

**References:**

* **Cloud Computing Bible by Barrie Sosinsky, Published by Wiley Publishing, 2011.**
* **Cloud Computing for Dummies by Judith Hurwitz, Robin Bloor, Marcia Kaufman, and Dr. Fern Halper, Wiley Publishing, 2010.**
* **Moving to The Cloud, Dinakar Sitaram, Elsevier, 2014.**
* **Cloud Computing Theory And Practice Danc.Marinercus, Elsevier, 2013.**

**Additional Reading:**

* **https://awsacademy.qwiklab.com**
* **https://aws.amazon.com**