|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Subject Code** | **Subject Title** | **Credit** | **Lecture** | **Tutorial** | **Practical** | **Type** |
|  | Big Data and Hadoop programming | **5** | **5** | **0** | **0** | **Theory** |
| **Introduction :**  To Provides an introduction to fundamental concepts of Big Data and Hadoop programming language and software environment for Data Analytics. | | | | | | |  |
| Course Outcome:   |  |  |  | | --- | --- | --- | | CO1 | : | To understand the basic concepts of Big Data and its Applications. | | CO2 | : | To understand the basic concepts of Big Data Analytics and its Approaches. | | CO3 | : | To learn Background And Fundamentals concepts of Hadoop | | CO4 | : | Understand the Data serialization and Map Reduces Concepts in Hadoop | | CO5 | : | To familiarize the Approaches to Handle and work with the Hive . |   **Unit 1**  Introduction to Big Data: From the Business Perspective - Characteristics of Big Data - Data in the Warehouse and Data in Hadoop- Big Data Important- Big Data Solution - Big Data Use Cases: Patterns for Big Data Deployment- IT for IT Log Analytics- The Fraud Detection Pattern - The Social Media Pattern - The Call Center Mantra- Risk: Patterns for Modeling and Management- Big Data and the Energy Sector.  **Unit 2**  Introduction to Big Data Analytics - Key Concepts - Big Data overview: Data Structures of Big Data- Analyst Perspective on Data Repositories-State of the practice in analytics: Business Intelligence versus Data Science-Current Analytical Architecture-Drivers of Big Data-Emerging Big Data Ecosystem and a New Approach to Analytics - Key roles for the new Big Data ecosystem- The Data Scientist- Examples of Big Data analytics  **Unit 3**  Background And Fundamentals: What is Hadoop- Running Hadoop - Data Logistics :Key elements of ingress and egress - Moving data into Hadoop: Pushing system log messages into HDFS with Flume-An automated mechanism to copy files into HDFS -Moving data out of Hadoop: Automated file copying from HDFS-Using Sqoop to export data to My SQL.  **Unit 4**  Data serialization—working with text and beyond -Understanding inputs and outputs in MapReduce-Processing common serialization formats-Big data serialization formats-Custom file formats-Big Data Patterns: Applying MapReduce patterns to big data-Joining-Sorting-Sampling.  **Unit 5**  Hive fundamentals-Installing Hadoop - Testing Hadoop- Installing Hive- Starting Hive- Data Types and File Formats- Collection Data Types- Text File Encoding of Data Values- HiveQL: Data Definition- HiveQL: Data Manipulation- HiveQL: Queries. | | | | | | |  |
|  | | | | | | |  |
|  | | | | | | |  |
| **Reference :** | | | | | | |  |

1. Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data, Paul C. Zikopoulos,Chris Eaton,DirkdeRoos,Thomas Deutsch,George Lapis ,McGraw Hill Publication, 2012
2. Data Science & Big Data Analytics, Discovering, Analyzing, Visualizing and Presenting Data ,John Wiley & Sons, Second 2015
3. Programming Hive ,Edward Capriolo, Dean Wampler, and Jason Rutherglen, O’Reilly MediaFirst Edition 2012

**Mapping of Course Outcome and Program Outcome:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Course Outcome** | **Program Outcome** | | | | | | | |
| **PO1** | **PO2** | **PO3** | **PO4** | **PO5** | **PO6** | **PO7** | **PO8** |
| **CO1** | L | H | -- | H | M | - | M | -- |
| **CO2** | - | M | L | - | M | -- | H | M |
| **CO3** | M | - | H | M | - | H | - | L |
| **CO4** | - | H | - | H | L | - | H | - |
| **CO5** | H | - | M | - | H | M | - | H |