**17IT4604A BIGDATA**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Course**  **Category:** | Programme Elective-II | | | | | | | | | **Credits:** | | | | | | 3 | |
| **CourseType:** | Theory | | | | | | | | | **Lecture-Tutorial-**  **Practice:** | | | | | | 3-0-3 | |
| **Prerequisites:** | 17IT3502- Data Warehousing And Mining | | | | | | | | | **Continuous Evaluation:** | | | | | | 30 | |
|  | | | | | | | | | | **Semester EndEvaluation:** | | | | | | 70 | |
| **Total Marks:** | | | | | | 100 | |
|  | | | | | | | | | | | | | | | | | |
| **Course**  **Outcomes** | Upon successful completion of the course, the student will be able to: | | | | | | | | | | | | | | | | |
| CO1 | Analyze Hadoop Architecture—Name Node, Big Data Lifecycle. | | | | | | | | | | | | | | | |
| CO2 | Master the concepts of Hadoop Distributed File System. | | | | | | | | | | | | | | | |
| CO3 | Acquire knowledge on Map Reduce Framework. | | | | | | | | | | | | | | | |
| CO4 | Apply Pig and Hive concepts for Data Processing. | | | | | | | | | | | | | | | |
| **Contribution of**  **Course Outcomes towards achievement of Program Outcomes**  **(L-Low, H- High)** |  | PO  1 | PO  2 | PO  3 | PO  4 | PO  5 | PO  6 | PO  7 | PO  8 | | PO  9 | PO  10 | PO  11 | PO12 | PSO1 | | PSO2 |
| CO1 | M |  |  | L | M |  |  |  | |  |  |  |  |  | |  |
| CO2 | L |  |  |  | M |  |  |  | |  |  |  |  |  | |  |
| CO3 | M |  |  |  | M |  |  |  | |  |  |  |  |  | |  |
| CO4 | M |  |  | L | H |  |  |  | |  |  |  |  |  | |  |
| **Course Content** | **UNIT I**  **Introduction to Big Data:**  Big Data-definition, Characteristics of Big Data (Volume, Variety, Velocity), Data in the Warehouse and Data in Hadoop, Why is Big Data Important? Patterns for Big Data Development.  **Introduction to Hadoop:**  Data, Data Storage and Analysis, Comparison with Other Systems: RDBMS, Grid Computing, Volunteer Computing, A Brief History of Hadoop, Apache Hadoop and the Hadoop Ecosystem, Hadoop Releases. | | | | | | | | | | | | | | | | |
| **UNIT II**  **Hadoop Distributed File System:** The Design of HDFS, HDFS Concepts, Blocks, Namenodes and Datanodes, Basic Filesystem Operations, Hadoop Filesystems, Interfaces, The Java Interface, Reading Data from a HadoopURL, Data Flow, Anatomy of a FileRead, Anatomy of a FileWrite, Coherency Model. | | | | | | | | | | | | | | | | |
| **UNIT III**  **Map Reduce**–A Weather Dataset, Data Format, Analyzing the Data with Unix Tools, Analyzing the Data with Hadoop, Map and Reduce, Java Map Reduce, Scaling Out, Hadoop Streaming, Hadoop Pipes.  **Pig**-Installation and Running of Pig, Execution Types, Running Pig Programs, Pig Latin Editors, Comparison with databases, Pig Latin, Functions, Data Processing Operators. | | | | | | | | | | | | | | | | |
|  | **UNITIV:**  **Hive**-Installing Hive, An Example, Running Hive, Comparison with Traditional Databases, HiveQL, Tables, Querying Data. | | | | | | | | | | | | | | | | |
| **Textbooks and**  **Reference books** | **Text Book(s):**  [1].Dirk deRoos, Chris Eaton, George Lapis, Paul Zikopoulos, Tom Deutsch , “Understanding Big DataAnalyticsforEnterprise ClassHadoop and StreamingData”, 1st Edition, TMH,2012.  [2].TomWhite, Hadoop,“ The Definitive Guide”, 3rd Edition, O’Reilly Publications, 2012  **Reference Books:**  [1].Michael Berthold, DavidJ. Hand, “Intelligent Data Analysis”, Springer, 2007**.**  [2].David Loshin, "Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools,Techniques,NoSQL,and Graph”, MorganKaufmann Publishers, 2013  [3].Hadoop in Practice by AlexHolmes, MANNING  [4].Hadoop in Action by ChuckLam, MANNING | | | | | | | | | | | | | | | | |
| **E-resources and**  **Other digital materials** | [1].Big Data Use cases for Beginners | Real Life Case Studies | Success Stories <https://www.youtube.com/watch?v=HHR0-iJp2sM>  [2]. Alexey Grishchenko, Hadoop vs MPP, <https://0x0fff.com/hadoop-vs-mpp/>  [3].Random notes on bigdata- SlideShare: Available www.slideshare.net/yiranpang/random-notes-on-big-data-26439474 | | | | | | | | | | | | | | | | |