|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Course Code | 21CSE222T | Course Name | BIG DATA TOOLS AND TECHNIQUES | Course Category | E | Professional Elective | L | T | P | C |
| 3 | 0 | 0 | 3 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Pre-requisite Courses** |  | | **Co-requisite Courses** | *Nil* | | **Progressive Courses** |  |
| **Course Offering Department** | | *Computer Science and Engineering* | | | **Data Book / Codes/Standards** | *Nil* | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | **Course Learning Rationale (CLR):** | | *The purpose of learning this course is to:* | | **CLR-1 :** | *Gain knowledge about the various tools and techniques used in big data analytics* | | | **CLR-2 :** | *Learn the fundamentals of Hadoop and the related technologies* | | | **CLR-3 :** | *Understand the basics of development of applications using MapReduce, HDFS, YARN* | | | **CLR-4 :** | *Learn the basics of Pig, Hive and Sqoop* | | | **CLR-5 :** | *Learn the basics of Apache Spark, Flink and understand the importance of NoSQL databases* | | | **CLR-6 :** | *Learn about Enterprise Data Science and data visualization tools* | | | | |  | **Program Learning Outcomes (PO)** | | | | | | | | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | |
| Engineering Knowledge | Problem Analysis | Design & Development | Analysis, Design, Research | Modern Tool Usage | Society & Culture | Environment & Sustainability | Ethics | Individual & Team Work | Communication | Project Mgt. & Finance | Life Long Learning | PSO - 1 | PSO - 2 | PSO – 3 | |
| **Course Learning Outcomes (CO):** | | *At the end of this course, learners will be able to:* |
| CO-1: | *Use the various tools and techniques in big data analytics* | |  |  |  |  | *1* |  |  |  |  |  |  |  |  |  |  |
| CO-2: | *Apply Hadoop and related technologies to big data analytics* | |  |  |  |  | *1* |  |  |  |  |  |  |  | *1* |  |  |
| CO-3: | *Apply MapReduce, HDFS and YARN develop big data applications* | |  |  |  |  | *1* |  |  |  |  |  |  |  | *1* |  |  |
| CO-4: | *Develop applications using Pig, Hive and Sqoop* | |  |  |  |  | *1* |  |  |  |  |  |  |  | *1* |  |  |
| CO-5: | *Apply Apache Spark and Flink to applications and understand the importance of NoSQL databases* | |  |  |  |  | *1* |  |  |  |  |  |  |  | *1* |  |  |
| CO-6: | *Understand the applications of Enterprise Data Science and data visualization tools* | |  |  |  |  | *1* |  |  |  |  |  |  |  |  |  |  |

|  |
| --- |
| **Unit1:** *Overview of Big Data Analytics-Introduction to data analytics and big data-Big data mining-Technical elements of the Big Data platform, Analytics Toolkit, Components of the analytics toolkit -Distributed and Parallel Computing for Big Data-Cloud computing and Big Data-In-Memory Computing Technology for Big Data-Fundamentals of Hadoop-Hadoop Ecosystem-The core modules of Hadoop-Introduction to Hadoop MapReduce-Introduction to Hadoop YARN.***.9hrs** |
| ***Unit 2:*** *MapReduce-Analyzing data with Unix tools and Hadoop-Scaling Out – Data Flow, Combiner Functions-Hadoop Streaming-HDFS-Hadoop filesystems-Java Interface to Hadoop-YARN-Job Scheduling-Hadoop I/O-Data Integrity-Compression-Serialization-File based Data Structures-Developing a MapReduce Application.* ***9hrs*** |
| Unit 3: *Setting up a Hadoop Cluster-Cluster specification and setup-Hadoop configuration-YARN configuration-Introduction to Pig-Installing and running pig-Basics Pig Latin -Example Programs-Introduction to Hive-Installing and running Hive-Introduction to HiveQL-Create-Drop-Alter-order by-Group by-Joins-Introduction to Zookeeper-Installing and running Zookeeper-Creating different types of Znodes-Flume Architecture-Introduction to Sqoop.***9hrs** |
| *Unit 4:Introducing Oozie-Apache Spark-Limitations of Hadoop and overcoming the Limitations-Core components and architecture of Spark-Introduction to Apache Flink-Installing Flink-Batch analytics using Flink-Big Data Mining with NoSQL-Why NoSQL?-NoSQL databases-Introduction to MongoDB,-Basi queries in MongoDB-Introduction to Cassandra.***9hrs** |
| Unit 5: *Enterprise Data Science Overview-Data Science Solutions in the enterprise-Enterprise data science – Machine Learning and AI-Enterprise Infrastructure Solutions-Visualizing Big Data-Using Python and R for visualization-Big Data Visualization Tools-Data Visualization with Tableau-Case Studies: Hadoop-Case Studies: Spark-Case Studies: NoSQL.* ***9hrs*** |

|  |  |  |
| --- | --- | --- |
| **Learning Resources** | 1. Herbert Schildt, ‘C++ - T Bjarne Stroustrup, ‘The C++ Programming Language’, Addison Wesley, 2000. 2. The Complete Reference’, Tata McGraw Hill, 1997. 3. Herbert Schildt, ―Java The complete reference‖, 8th Edition, McGraw Hill Education, 2011 | 1. Kris Jasma, ‘Java Programming – A Complete Reference’, Galgotia publication, 1994. 2. Cay S. Horstmann, Gary cornell, ―Core Java Volume –I Fundamentals‖, 9th Edition, Prentice Hall, 2013. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Bloom’s**  **Level of Thinking** | **CLA – 1**  **(50%)** | **CLA – 2 (10%)** | **Final exam**  **(40% Weightage)** |
| **Theory** | **Theory** | **Theory** |
| **Level 1** | **Remember** | *50%* | *50%* | *50%* |
| **Level 2** | **Understand** | *50%* | *50%* | *50%* |
| **Level 3** | **Apply** | *-* | **-** | - |
| **Level 4** | **Analyze** | - | ***-*** | *-* |
| **Level 5** | **Evaluate** | **-** | **-** | **-** |
| **Level 6** | **Create** | ***-*** | ***-*** | ***-*** |
|  | Total | 100 % | 100 % | 100 % |

|  |  |  |
| --- | --- | --- |
| **Course Designers** | | |
| Experts from Industry | Experts from Higher Technical Institutions |  |
| 1.Dr.R. SivaKumar,Sr. Consultant,rsivakoumar@gmail.com A2O Integrated services Pvt., Ltd., Chennai | 1. Dr.S Muthurajkumar, Asst. Professor, Department of Computer Technology, muthuraj@annauniv.edu, MIT Campus, Anna University, Chromepet, Chennai-600044. | *Mrs.S.Sindhu, AP, DSBS* |