

VII Semester

BIG DATA ANALYTICS			
Course Code	21CS71	CIE Marks	50
Teaching Hours/Week (L:T:P: S)	3:0:0:0	SEE Marks	50
Total Hours of Pedagogy	40	Total Marks	100
Credits	03	Exam Hours	03
Course Learning Objectives: CLO 1. Understand fundamentals and applications of Big Data analytics CLO 2. Explore the Hadoop framework and Hadoop Distributed File system and essential Hadoop Tools CLO 3. Illustrate the concepts of NoSQL using MongoDB and Cassandra for Big Data CLO 4. Employ MapReduce programming model to process the big data CLO 5. Understand various machine learning algorithms for Big Data Analytics, Web Mining and Social Network Analysis.			
Teaching-Learning Process (General Instructions) These are sample Strategies, which teachers can use to accelerate the attainment of the various course outcomes. <div><div>1.</div><div>Lecturer method (L) does not mean only traditional lecture method, but different type of teaching methods may be adopted to develop the outcomes.</div></div> <div><div>2.</div><div>Show Video/animation films to explain functioning of various concepts.</div></div> <div><div>3.</div><div>Encourage collaborative (Group Learning) Learning in the class.</div></div> <div><div>4.</div><div>Ask at least three HOT (Higher order Thinking) questions in the class, which promotes critical thinking.</div></div> <div><div>5.</div><div>Adopt Problem Based Learning (PBL), which fosters students' Analytical skills, develop thinking skills such as the ability to evaluate, generalize, and analyze information rather than simply recall it.</div></div> <div><div>6.</div><div>Topics will be introduced in a multiple representation.</div></div> <div><div>7.</div><div>Show the different ways to solve the same problem and encourage the students to come up with their own creative ways to solve them.</div></div> <div><div>8.</div><div>Discuss how every concept can be applied to the real world - and when that's possible, it helps improve the students' understanding.</div></div>			
Module-1			
Introduction to Big Data Analytics: Big Data, Scalability and Parallel Processing, Designing Data Architecture, Data Sources, Quality, Pre-Processing and Storing, Data Storage and Analysis, Big Data Analytics Applications and Case Studies.			
Textbook 1: Chapter 1: 1.2 -1.7			
Teaching-Learning Process	Chalk and board https://www.youtube.com/watch?v=n_Krer6YWY4 https://onlinecourses.nptel.ac.in/noc20_cs92/preview		
Module-2			
Introduction to Hadoop (T1): Introduction, Hadoop and its Ecosystem, Hadoop Distributed File System, MapReduce Framework and Programming Model, Hadoop Yarn, Hadoop Ecosystem Tools.			
Hadoop Distributed File System Basics (T2): HDFS Design Features, Components, HDFS User Commands.			
Essential Hadoop Tools (T2): Using Apache Pig, Hive, Sqoop, Flume, Oozie, HBase.			
Textbook 1: Chapter 2 :2.1-2.6 Textbook 2: Chapter 3			

Textbook 2: Chapter 7 (except walk throughs)	
Teaching-Learning Process	<ol style="list-style-type: none"> 1. Chalk and Board 2. Laboratory Demonstration
Module-3	
NoSQL Big Data Management, MongoDB and Cassandra: Introduction, NoSQL Data Store, NoSQL Data Architecture Patterns, NoSQL to Manage Big Data, Shared-Nothing Architecture for Big Data Tasks, MongoDB, Databases, Cassandra Databases.	
Textbook 1: Chapter 3: 3.1-3.7	
Teaching-Learning Process	<ol style="list-style-type: none"> 1. Chalk and Board 2. Laboratory Demonstration https://www.youtube.com/watch?v=pWbMrx5rVBE
Module-4	
Introduction, MapReduce Map Tasks, Reduce Tasks and MapReduce Execution, Composing MapReduce for Calculations and Algorithms, Hive, HiveQL, Pig.	
Textbook 1: Chapter 4: 4.1-4.6	
Teaching-Learning Process	<ol style="list-style-type: none"> 1. Chalk and Board 2. Laboratory Demonstration
Module-5	
Machine Learning Algorithms for Big Data Analytics: Introduction, Estimating the relationships, Outliers, Variances, Probability Distributions, and Correlations, Regression analysis, Finding Similar Items, Similarity of Sets and Collaborative Filtering, Frequent Itemsets and Association Rule Mining.	
Text, Web Content, Link, and Social Network Analytics: Introduction, Text mining, Web Mining, Web Content and Web Usage Analytics, Page Rank, Structure of Web and analyzing a Web Graph, Social Network as Graphs and Social Network Analytics:	
Textbook 1: Chapter 6: 6.1 to 6.5 Textbook 1: Chapter 9: 9.1 to 9.5	
Teaching-Learning Process	<ol style="list-style-type: none"> 1. Chalk and Board 2. Laboratory Demonstration
Course outcome (Course Skill Set) At the end of the course the student will be able to: <ol style="list-style-type: none"> CO 1. Understand fundamentals and applications of Big Data analytics. CO 2. Investigate Hadoop framework, Hadoop Distributed File system and essential Hadoop tools. CO 3. Illustrate the concepts of NoSQL using MongoDB and Cassandra for Big Data. CO 4. Demonstrate the MapReduce programming model to process the big data along with Hadoop tools. CO 5. Apply Machine Learning algorithms for real world big data, web contents and Social Networks to provide analytics with relevant visualization tools. 	
Assessment Details (both CIE and SEE) The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 40% of the maximum marks (20 marks). A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/course if the student secures not less than 35% (18 Marks out of 50) in the semester-end examination (SEE), and a minimum of 40% (40 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together	

Continuous Internal Evaluation:

Three Unit Tests each of **20 Marks (duration 01 hour)**

1. First test at the end of 5th week of the semester
2. Second test at the end of the 10th week of the semester
3. Third test at the end of the 15th week of the semester

Two assignments each of **10 Marks**

4. First assignment at the end of 4th week of the semester
5. Second assignment at the end of 9th week of the semester

Group discussion/Seminar/quiz any one of three suitably planned to attain the COs and POs for **20 Marks (duration 01 hours)**

6. At the end of the 13th week of the semester

The sum of three tests, two assignments, and quiz/seminar/group discussion will be out of 100 marks and will be **scaled down to 50 marks**

(to have less stressed CIE, the portion of the syllabus should not be common /repeated for any of the methods of the CIE. Each method of CIE should have a different syllabus portion of the course).

CIE methods /question paper has to be designed to attain the different levels of Bloom's taxonomy as per the outcome defined for the course.

Semester End Examination:

Theory SEE will be conducted by University as per the scheduled timetable, with common question papers for the subject (**duration 03 hours**)

1. The question paper will have ten questions. Each question is set for 20 marks. Marks scored shall be proportionally reduced to 50 marks
2. There will be 2 questions from each module. Each of the two questions under a module (with a maximum of 3 sub-questions), **should have a mix of topics** under that module.

The students have to answer 5 full questions, selecting one full question from each module

Suggested Learning Resources:**Textbooks**

1. Raj Kamal and Preeti Saxena, "Big Data Analytics Introduction to Hadoop, Spark, and Machine-Learning", McGraw Hill Education, 2018 ISBN: 9789353164966, 9353164966
2. Douglas Eadline, "Hadoop 2 Quick-Start Guide: Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem", 1 stEdition, Pearson Education, 2016. ISBN13: 978-9332570351

Reference Books

1. Tom White, "Hadoop: The Definitive Guide", 4 th Edition, O'Reilly Media, 2015.ISBN-13: 978-9352130672
2. Boris Lublinsky, Kevin T Smith, Alexey Yakubovich, "Professional Hadoop Solutions", 1 stEdition, Wrox Press, 2014ISBN-13: 978-8126551071
3. Eric Sammer, "Hadoop Operations: A Guide for Developers and Administrators",1 stEdition, O'Reilly Media, 2012.ISBN-13: 978-9350239261
4. ArshdeepBahga, Vijay Madiseti, "Big Data Analytics: A Hands-On Approach", 1st Edition, VPT Publications, 2018. ISBN-13: 978-0996025577

Weblinks and Video Lectures (e-Resources):

1. https://www.youtube.com/watch?v=n_Krer6YWY4
2. https://onlinecourses.nptel.ac.in/noc20_cs92/preview
3. <https://www.digimat.in/nptel/courses/video/106104189/L01.html>

4. https://web2.qatar.cmu.edu/~mhhammou/15440-f19/recitations/Project4_Handout.pdf

Activity Based Learning (Suggested Activities in Class)/ Practical Based learning

Mini Project Topics for Practical Based Learning :Search Engine Optimization, Social Media Reputation Monitoring, Equity Research, Detection of Global Suicide rate, Find the Percentage of Pollution in India, Analyze crime rate in India, Health Status Prediction, Anomaly Detection in cloud server, Tourist Behaviour Analysis, BusBest Not limited to above topics