

Programme	BS Computational Statistics and Data Analytics	Course Code	CSTA-406	Credit Hours	3
Course Title	Big Data Analytics				
Course Introduction					
<p>Big Data Analytics offers a comprehensive exploration of modern data analytics techniques, focusing on the utilization of Hadoop ecosystem tools for large-scale data processing. This course equips students with essential skills in data management, analysis, and interpretation, preparing them for roles in industries reliant on big data insights. Through hands-on projects and theoretical foundations, students learn to harness the power of distributed computing and advanced analytics to extract valuable insights from massive datasets.</p>					
Learning Outcomes					
<p>By the end of this course, students will be able to:</p> <ol style="list-style-type: none"> 1. Understand the analytics of big data and its basic framework. 2. Learn about the Hadoop distributed file system and different types of interfaces. 3. Acquire knowledge about PIG and its related concepts. 					
Course Content				Assignments/Readings	
Week 1	Unit – I				
	Introduction to big data and Hadoop				
	Unit – II				
	Types of digital data, introduction to big data				
Week 2	Unit – III				
	big data analytics, history of Hadoop				
	Unit – IV				
	Apache Hadoop, analyzing data with unix tools				
Week 3	Unit – V				
	Analysing data with Hadoop, Hadoop streaming				
	Unit – VI				
	Hadoop echo system				
Week 4	Unit – VII				
	IBM big data strategy				
	Unit – VIII				
	Introduction to Infosphere				
Week 5	Unit – IX				
	BigInsights, and Big Sheets				
	Unit – X				
	HDFS (Hadoop distributed file system)				
Week 6	Unit – XI				
	The design of HDFS, HDFS concepts, command Line interface				
	Unit – XII				
	Hadoop file system interfaces, data flow, data ingest				

	with flume	
Week 7	Unit – XIII Scoop and Hadoop archives, Hadoop I/O: compression	
	Unit – XIV Serialization	
Week 8	Unit – XV Anatomy of a map reduce Job run	
	Unit – XVI Failures, Job scheduling, shuffle and sort	
Week 9	Unit – XVII Task execution	
	Unit – XVIII Map reduce types and formats	
Week 10	Unit – XIX Map reduce features	
	Unit – XX Hadoop Eco System	
Week 11	Unit – XXI Pig, introduction to PIG	
	Unit – XXII Execution modes of Pig	
Week 12	Unit –XXIII Comparison of Pig with databases, grunt	
	Unit – XXIV Pig latin, user defined functions, data processing operators	
Week 13	Unit – XXV Hive, hive Shell, hive services	
	Unit – XXVI Hive metastore	
Week 14	Unit – XXVII comparison with traditional databases	
	Unit – XXVIII hiveQL, tables, querying data	
Week 15	Unit – XXIX User defined functions	
	Unit – XXX Hbase, hbasics, concepts	
Week 16	Unit – XXXI Clients, example, hbase versus RDBMS. Big SQL	
	Unit – XXXII Review and Applications: Recap of key concepts and techniques	

Textbooks and Reading Material

Text Book

1. Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media.

Suggested Readings.

1. McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). *Big data: the management revolution*. Harvard business review. 90(10), 60-68.
2. Prajapati, V. (2013). *Big data analytics with R and Hadoop*. Packt Publishing Ltd.

Teaching Learning Strategies

Class Lecture method, which includes seminars, discussions, assignments and projects. (Audio-visual tools are used where necessary)

Assignments: Types and Number with Calendar

According to the choice of respective teacher.

Assessment

Sr. No.	Elements	Weightage	Details
1.	Midterm Assessment	35%	It takes place at the mid-point of the semester.
2.	Formative Assessment	25%	It is continuous assessment. It includes: Classroom participation, attendance, assignments, and presentations, homework, attitude and behavior, hands-on-activities, short tests, quizzes etc.
3.	Final Assessment	40%	It takes place at the end of the semester. It is mostly in the form of a test, but owing to the nature of the course the teacher may assess their students based on term paper, research proposal development, field work and report writing etc.