भारतीय सूचना प्रौद्योगिकी संस्थान, नागपुर

Indian Institute of Information Technology, Nagpur

An Institution of National Importance by an Act of Parliament

Survey No. 140,141/1 Behind Br. Sheshrao Wankhade Shetkari Soot Girni, Village: Waranga, Po: Dongargaon (Butibori), District : Nagpur (Maharashtra) -441108

Website: www.iiitn.ac.in Email: director@iiitn.ac.in, registrar@iiitn.ac.in Phone: 9405215010

Department of Computer Science and Engineering

Teaching Plan

Course Name & Code	Big Data Analytics (CSL 444)		
Semester	VI		
Type	DC		
Academic Year	2024-2025		
Course Coordinator			
Credits - 4	L	T	P
	3	0	2

Lecture	Topic
	Module 1 Introduction
1	Introduction to course – Syllabus, outcomes, evaluation strategy, exam pattern, Introduction to Big Data Analytics
2-4	Challenges of Conventional Systems - Intelligent data analysis, Nature of Data, Analytic Processes and Tools, Analysis vs Reporting
	Module 2 Mining Data Streams
5-6	Introduction To Streams Concepts, Stream Data Model and Architecture ,Stream Computing ,Sampling Data in a Stream
7-9	Filtering Streams, Counting Distinct Elements in a Stream, Estimating Moments, Counting Oneness in a Window, Decaying Window, Real time Analytics Platform (RTAP) Applications
10-12	Case Studies, Real Time Sentiment Analysis, Stock Market Prediction
	SESSIONAL 1
	Module 3 Hadoop
13-15	History of Hadoop- the Hadoop Distributed File System, Components of Hadoop Analyzing the Data with Hadoop-, Scaling Out- Hadoop Streaming-
16-18	Design of HDFS, HDFS Basics- Developing a Map Reduce Application-How Map Reduce Works-Anatomy of a Map Reduce Job run-Failures-Job Scheduling
19-21	Shuffle and Sort, Task execution - Map Reduce

	Types and Formats- Map Reduce Features Hadoop environment
	Module 4 – Frameworks
22-23	Applications on Big Data Using Pig and Hive
24-25	Data processing operators in Pig, Hive services, HiveQL, Querying Data in Hive
26-27	fundamentals of HBase and ZooKeeper - IBM InfoSphere BigInsights and Streams.
	SESSIONAL 2
	Module 5 Predictive Analysis
28	Simple linear regression, Multiple linear regression
29-31	Interpretation 5 of regression coefficients.
32-34	Visualizations
35-37	Visual data analysis techniques
38-43	interaction techniques - Systems and applications
	END SEM EXAM

Evaluation Plan

Sr. No	Evaluation based on	Weightage	Dates
1.	Sessional exam 1	15 Marks	
2.	Sessional exam 2	15 Marks	
3.	End semester exam	60 Marks	
4.	Teacher's Assessment	10 Marks	

Total 100 will scale down to 75

List of Lab Activities

No.	Lab Activities
1	Understanding and working with Hadoop
2	Hadoop Map Reduce Program
3	Understanding and implementing Concept of spark
4	Word count Program in MapReduce
5	Structured Streaming Programming
6	Build a recommendation system using Spark and MLib
7	Handling and analysing databases like Cassandra, hive, and mongo db
8	Handling and analysing database mongo db
9	Data analytics algorithms implementation
10	Classification using Clustering Techniques
11	Build a Project with implementation

Lab Evaluation Plan

Sr. No	Evaluation based on	Weightage	Dates
1.	Lab exam 1	40 Marks	
2.	Lab exam 2	40 Marks	
3.	Lab Activities	20 Marks (10 * 10 =100 Marks+ Attendance 10 Marks, Scale down 110 Marks to 20 Marks)	

Total 100 will scale down to 25

Mrs. Roma Goel
Course Instructor

Dr. Kaushlendra Sharma
Course Coordinator