Semester	**TWO**
Paper Number	**7**
Paper Code	**MDTS 4213**
Paper Title	**Big Data Analytics**
No. of Credits	6
Course Description	CORE Composite Paper One Module Number of classes: Theory–4 per week Practical–3 per week
Course Objective	After completion of the course, a student is expected to have o Understanding of the challenges of computation related to big data. o Exposure to uses and misuses of big data. o Gaining knowledge about various computational platforms available for big data analytics. o Understanding the advantages and disadvantages of the big data analytics platforms, including the software frameworks. o Gaining knowledge about parallel computation in various big data analytics platforms. o Gaining hands-on experience in parallel computing with R and Python. o Knowledge to identify networks and study their essential properties through measures.
Syllabus	**Introduction:** Examples of big data in natural sciences, engineering, social sciences, industry, etc. Importance of analyzing big data. Use and misuse of big data. Limitations of the traditional computational platforms in the analysis of big data. **Scaling of big data analytics platforms:** Horizontal and vertical scaling, Peer-to-peer networks, Hadoop, Spark, Berkeley Data Analysis Stack(BDAS), High-Performance Computing (HPC) clusters, multi-core processors, Graphics Processing Unit (GPU), Field Programmable Gate Arrays (FPGA). **Distributed computing:** Importance of distributed computing for big data, Basic ideas of the communication systems for parallel computing in peer-to-peer networks (Message Passing Interface (MPI)), Hadoop (HDFS, YARN, Map Reduce), Spark, BDAS, (Tachyon + Mesos– an improvement over Spark due to more aggressive memory exploitation). Communication systems for vertical scaling – MPI for HPC and multicore processors; CUDA for GPUs, Hardware Descriptive Language (HDL) for FPGA. **Comparisons of different big data platforms**: communication mechanisms based on scalability, data I/O performance, **Network Analysis**: Examples, components, properties, and descriptive measures.

	Fault tolerance, real-time processing, data size supported, iterative task support. **Pseudocodes:** Illustrative examples of simple pseudocodes of the K-means algorithm in MapReduce, MPI, and GPU-based platforms.	
Practical	Based on the theory topics	
Reference List	1. Sourav Mazumder, Robin Singh Bhadoria and Ganesh Chandra Deka(2017), "Distributed Computing in Big Data Analytics", Springer. 2. Martin Van Steen and Andrew S Tanenbaums: Distributed Systems 3rd Edition (2017) 3. Singh, D. and G. K. Reddy(2014). A Survey on Platforms for Big Data Analytics, Journal of Big Data 1:8, 1–20	
Evaluation	Theory CIA: 10 End Sem Exam:50(25+25) Total:60	Practical Continuous Assessment Total:40
Paper Structure for End Semester Theory	Short questions:5marks each	Long questions:10 marks each
	2 out of 4	4 out of 6