Semester	**TWO**
Paper Number	**7**
Paper Code	**MDTS 4213**
Paper Title	**Big Data Analytics**
No. of Credits	6
Course Description	CORE Composite PaperOne 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 Gaining wholesome knowledge about various computational platforms available for big data analytics. o Understanding the advantages and disadvantages of the big data analytics platforms, including the softwareframeworks. o Gaining wholesome knowledge about parallel computation in various big data analytics platforms. o Gaining hands-on experience in parallel computing with R and Python. o Gaining hands-on experience in cloud computing.
Syllabus	**Introduction:** Examples of big data in natural sciences, engineering, social media, industry, etc. Importance of analysingbig data. Limitations of the traditional computational platforms in analysis of big data. **Scaling of big data analytics platforms:** Horizontal and vertical scaling, Peer-to-peer networks, Hadoop, Spark, BerkeleyData Analysis Stack (BDAS), High Performance Computing (HPC) clusters, multicore 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 forparallel computing in peer-to-peer networks (Message Passing Interface (MPI)), Hadoop (HDFS, YARN, Map Reduce), Spark, BDAS, (Tachyon+Mesos– improvement over Spark due to more aggressive memory exploitation). Communicationsystems 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,

	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 basedplatforms. **Cloud Computing**: Introduction, Characteristics, Cloud Delivery Models, Cloud Deployment Models, Cloud platforms, Case study on AWS Services, Model Deployment using Flask APIs on AWS, Amazon SageMaker	
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: 30 End Sem Viva: 10 Total: 40
Paper Structure for End Semester Theory	Short questions: 5 marks each	Long questions: 10 marks each
	2 out of 4	4 out of 6