

1.	Title of the course	Big Data Systems Laboratory
2.	Course number	CS601P
3.	Structure of credits	0-0-3-2
4.	Offered to	PG
5.	New course/modification to	Modification To CS6191/21
6.	To be offered by	Department of Computer Science and Engineering
7.	To take effect from	July 2022
8.	Prerequisite	Nil
9.	Course Objective(s): To give a hands-on experience on designing and building of distributed computing and storage systems for big data; To give an exposure to different commercial big data hosting services and infrastructures through case study.	
10.	Course Content: Distributed computing: implementing distributed coordination framework using Apache Zookeeper, managing and implementing Hadoop and Spark services, Hadoop distributed file system (HDFS) and developing application frameworks that use HDFS; Design and implement Ceph storage; Hands-on experience in implementing OpenStack infrastructure; Design and use Apache Spark; Case study on understanding the commercial frameworks such as Amazon elastic compute cloud.	
11.	Textbook(s): 1. Marinescu D C, <i>Cloud Computing: Theory and Practice</i> , 1st Edition, Morgan Kaufman (2019). 2. White T, <i>Hadoop: The Definitive Guide</i> , 1st Edition, O Reilly (2012).	
12.	Reference(s): 1. Chambers B and Zaharia M, <i>Spark: The Definitive Guide</i> , 1st Edition, O Reilly (2018). 2. Solberg M and Silverman B, <i>Open Stack for Architects</i> , 2nd Edition, Packt (2018).	