

This module teaches students concepts, techniques and technologies in handling and analyzing big data sets, i.e., data sets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. Common sources and domains of big data include ubiquitous information-sensing mobile devices, web and software logs, financial trading transactions, large-scale e-commerce, RFID and wireless sensor networks, etc. Conceptual big data topics covered include big data infrastructure, analytics scalability and processes, etc. Technical frameworks for big data handling and analytics taught include Hadoop, HBase, Cassandra, MapReduce, Dynamo, R, in-database analytics, mining of data streams, etc.