BLM-4821 Big Data Processing and Analytics Course

Syllabus

Course Description:	Modern scientific instruments and Internet-scale applications generate voluminous data pertaining to vital signs, weather phenomena, social networks that connect millions of users, and the origins of distant planets. Data produced in these settings hold the promise to significantly advanced knowledge. This course covers fundamental issues in Big Data. The course examines issues related to data organization, storage, retrieval, analysis, and knowledge discovery at scale. This will include topics such as large-scale data analysis, data storage systems, self-descriptive data representations, semi-structured data models.
Text Books:	1) Anand Rajaraman, Jure Leskovec, and Jeffrey Ullman, "Mining of Massive Datasets", Cambridge University Press, 2012 (2nd Edition) 2) Tom White, "Hadoop: The Definitive Guide", O'Reilly Media, 2015 (3rd Edition) 3) Donald Miner, Adam Shook, "MapReduce Design Patterns", 2012
Lecture Hours and Location:	Thursday, 14:00 - 17:00 AM - D011
Course Topics (tentative):	- Distributed File Systems - MapReduce - Link Analysis algorithms - Similarity measure algorithms - Advanced programming patterns in MapReduce - Data Filtering algorithms - NoSQL databases - Scalable data retrieval - Key-value storage systems
Grading (tentative):	* Attendance: 5% * Reading Assignments: 5% * Homeworks: 10% (Programming Assignments) * Midterm Exam: 20% * Term Project: 20% * Final: 40%
Academic Integrity:	Cheating and disruptive talking in class will not be tolerated. Any kind of cheating and plagiarism will be severely penalized. On the other hand, asking and responding to questions in class is strongly encouraged.