

## **DEPARTMENT OF COMPUTER SCIENCE COURSE SYLLABUS**

**CS 63017/73017    BIG DATA MANAGEMENT    3 Credit Hours**

**Instructor's Name: Dr. Ruoming Jin**

### **Reference Books**

1. Hadoop: The Definitive Guide, Tom White, O'Reilly, 2015.
2. Hadoop In Action, Chuck Lam, Manning, 2010.
3. Data-Intensive Text Processing with MapReduce, Jimmy Lin and Chris Dyer, 2010.
4. Data Mining: Concepts and Techniques, Third Edition, by Jiawei Han et al., Morgan Kaufmann, 2000.

**Course Content:** Introduces computing platforms with focus on how to use them in processing, managing and analyzing massive datasets. Utilizes several key data processing tasks, including simple statistics, data aggregation, join processing, frequent pattern mining, data clustering, information retrieval, pagerank and massive graph analytics as the case study for large scale data processing.

**Prerequisites or co-requisites:** Graduate standing  
**Required, elective, or selected elective**

### **Goals:**

1. Understand the background of big data
2. Learn basic data mining and warehouse problems over big data
3. Understand MapReduce/Hadoop basics
4. Learn Hadoop and MapReduce programming
5. Learn machine learning problems over big data
6. Learn graph processing algorithms via MapReduce

### **Outcomes:**

1. Understanding of big data basics
2. Understanding of problems over big data
3. Be proficient with Hadoop and MapReduce programming
4. Use Hadoop and MapReduce programming to tackle big data problems
5. Develop programs to process big data
6. Collaborate with team members to complete a project related to big data

### **Topics to be Covered:**

1. Introduction to Big Data

2. Exploratory Data Analysis
3. Business Intelligence: OLAP, Data Warehouse, and Column Store
4. Frequent Pattern Mining
5. Intro to MapReduce/Hadoop
6. Hadoop Programming Tutorial & MapReduce Programming Patterns
7. Information Retrieval & MapReduce
8. Relational Database Operators & MapReduce
9. Machine Learning: Clustering (Unsupervised Learning)
10. Machine Learning: Classification (Supervised Learning)
11. Machine Learning & MapReduce
12. Graph Algorithms & MapReduce

**Abet Learning Outcomes:**

- Analyze a complex computing problem and to apply principles of computing and other relevant disciplines to identify solutions.
- Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline.
- Apply computer science theory and software development fundamentals to produce computing-based solutions.
- Communicate effectively in a variety of professional contexts.
- Recognize professional responsibilities and make informed judgments in computing practice based on legal and ethical principles.
- Function effectively as a member or leader of a team engaged in activities appropriate to the program's discipline.