Big Data Processing and Analytics:
Course Introduction

Instructor: Dr. Mehmet S. Aktaş

About the Course

2018 – Spring Semester – BLM-4821

- Instructor: Dr. Mehmet S. Aktaş
- Thursday 14:00-17:00, D011

Course Logistics

- **Textbooks:**
  - 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

- **Course materials**
  - You should check the website regularly

Logistics: Communication

- **For e-mailing us, always use:**
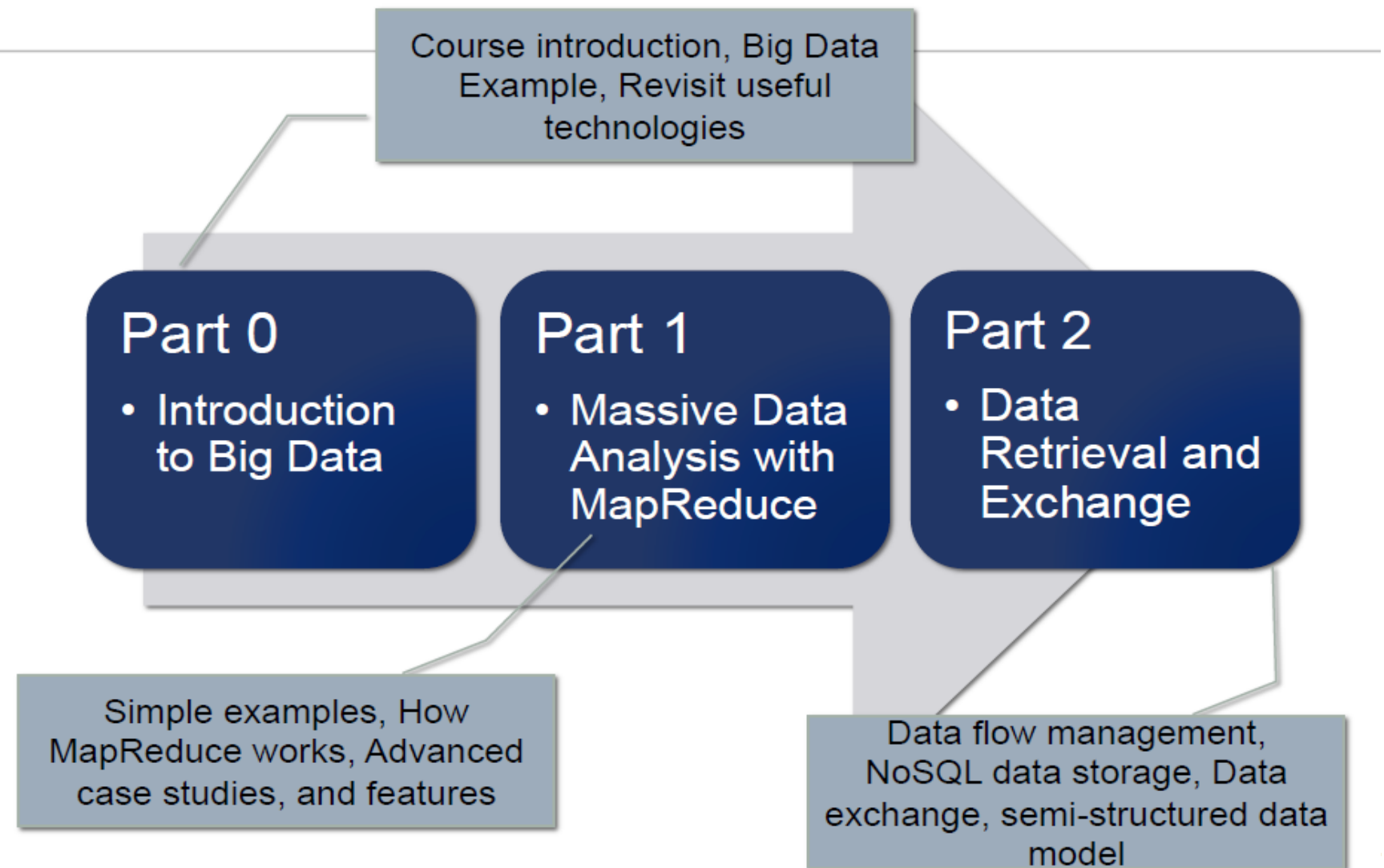 - ytu.ce.blm4821@gmail.com

Purpose of the Course

- ## Big Data Processing Knowledge:
  - To introduce you basic big data processing concepts

- ## Computer Science Skills:
  - To do literature survey, read research papers, present a paper in the class

Course Plan

- I will give the basics on big data processing
  - Overview big data processing concepts
  - Discuss research related issues

- You are expected to
  - Attend the class (70% attendance required)
  - Understand the basic big data mining concepts
  - Read research papers, present a paper in the class
  - Read the studied chapters and do the homework(s)/project(s)

What will we learn?

Course introduction, Big Data Example, Revisit useful technologies

Part 0
- Introduction to Big Data

Part 1
- Massive Data Analysis with MapReduce

Part 2
- Data Retrieval and Exchange

Simple examples, How MapReduce works, Advanced case studies, and features

Data flow management, NoSQL data storage, Data exchange, semi-structured data model
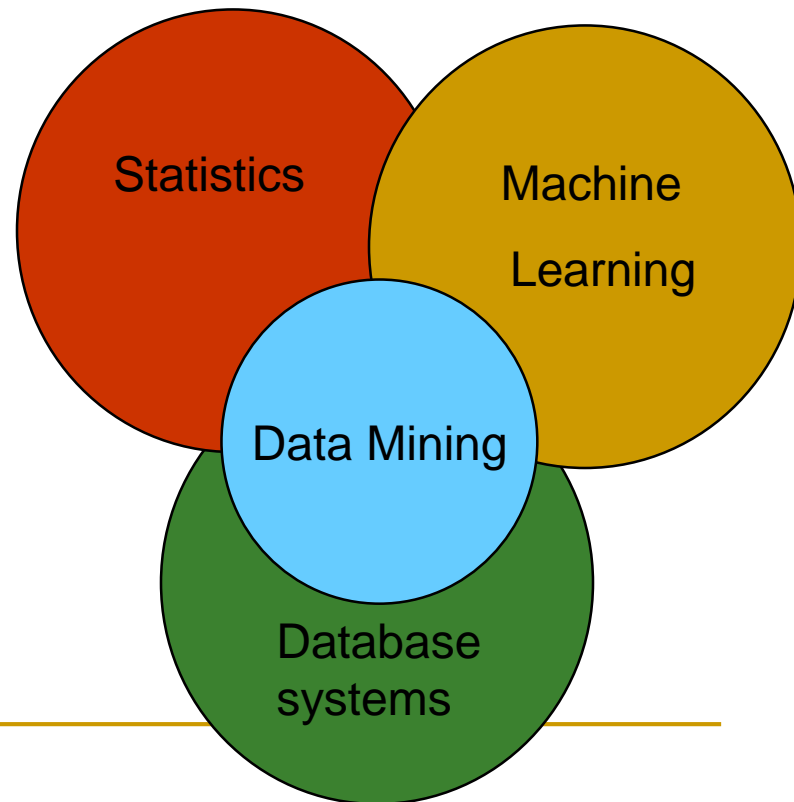
This Class: BLM-4821

- **This class overlaps with machine learning, statistics, artificial intelligence, databases but more stress on**
  - **Scalability** (big data)
  - **Algorithms**
  - **Computing architectures**
  - Automation for handling **large data**

Course Outline (tentative)

- Introduction to Big Data
- Map Reduce
- Big Data Topics
  - Finding Similar Items
  - Link Analysis
  - Frequent Itemsets
  - Recommendation Systems

Grading Policy (tentative)

- Attendance : 5%
- Midterm : 30%
- Term Project: 25%
- Final : 40%

Course Schedule (tentative)

- **February, March**
  - Overview of big data processing concepts
  - Midterm exam
- **April**
  - Big data processing concepts
  - Class term project presentations
  - Last day for classes
- **May**
  - Class term project presentations
  - Final exam