Capital University of Economics and Business
Overseas Chinese College
Course Syllabus

Year and Semester	2020 Fall (August 2, 2019 - December 18, 2020)
Course Name	Big Data
Course Code	MIS452
Course Type	☐ General Education (Required) ☐ General Education (Elective) ☐ Professional Course (Required) ☑ Professional Course (Elective) ☐ Basic Disciplinary Course
Course Credits	3
Course Hours	48
Prerequisites	Statistics, Database, SQL
Instructor	Leilei Zhu (Emma Zhu)
Contact Information	Office: C217 Tele: None Email: zhuleilei@cueb.edu.cn
Office Hour	TBA
Learning Centre	TBA
Grade/Section	2017IT/CFA
Course Time/Place	TBA;

Textbook

Viktor Mayer-Schönberger. *Big Data:A Revolution That Will Transform How We Live, Work, and Think, First Edition*. Zhejiang Renmin Press, ISBN: 978- 7- 213-05254-5

Reference Book

1. Viktor Mayer-Schönberger. *Delete: The Virtue of Forgetting in the Digital Age, First Edition*. Zhejiang Renmin Press, ISBN: 978- 7- 213-05251-4.

Course Description

This course will focus on the various aspects of big data. It contains: the detailed introduction of Hadoop, which includes two basic blocks of Hadoop system, Hadoop Distributed File System (HDFS) and Hive; how data is stored in HDFS; how can do some basic query and data analysis; the definition of Big Data and the features of big data time will be discussed in detail; several cases of applications of correlation in big data will be covered; a new concept Datafication will be defined and some examples will be followed; the values of big data, the implication and potential risks of big data will also be discussed in terms of several cases.

Student Learning Objectives

After completing this course, students will be able to:

- Obtain some knowledge of big data technology
- Gain a full understanding of big data time

Website Source

https://www.webopedia.com/TERM/B/big_data.html
http://spark.apache.org/

Teaching Methods

This course consists of lectures, discussions and student presentations. Students will be divided into small groups with a group leader helping others in the group. Students must be prepared to finish some small questions and small quizzes during the class.

Grade Criterion

Component	Weight	Description
Final Exam	20%	A cumulative final examination will be given based on all of the contents of the class. The exam paper may be composed of multiple-choice questions, short answer questions, essay questions, problems, and preparation of financial statements. Students should rely primarily on homework assignments to give them a sense of what they may see for material on exams.
Mid-Term Test	20%	A cumulative midterm test will be given based on all of the contents that have been taught in class. The test paper may be mainly composed of multiple-choice questions and it should be completed within 15 minutes in class.
Homework	15%	Most of the assigned homework is taken from the Exercises in the textbook. Assignments will be collected at the clearly stated date. Late assignments will not be accepted. The graded assignments will be kept by the tutor for reference and won't be returned to students.
Quizzes	15%	There will be at least 2 quizzes during the semester. Quizzes may or may not be announced in advance. It may also be used as a way to check the attendance. Quizzes will test your knowledge of both concepts and the application of those concepts.
Presentation	10%	The students will be divided into several groups to prepare a presentation. Each student is required to be involved in the presentation. The topics can be selected from the textbook or lectures. Each group need to finish a PPT related to the topic which is given and hand in the related resources to the teacher before the presentation.
Participation	10%	Individuals will be asked to participate individually in a question and answer at least 5 times during the semester. The performances should be counted in their participation.
Attendance	10%	Refer to attendance policy listed below

Total	100%	

Detailed Grade Computation

	Before Midterm	After Midterm
Attendance	5%	5%
Participation	5%	5%
Homework	5%	10%
Quizzes	5%	10%
Presentation		10%
Midterm test	20%	
Final exam		20%
Total	40%	60%

Grading Policy

A+ 97-100	A 93-96	A- 90–92	B+ 87-89	B 83-86	B- 80–82
C+ 75-79	C 70-74	C- 67–69	D+ 63–66	D 62-60	F 0- 59

Exam Schedule

Midterm Test: 8th week

Final Exam: December 23, 2020-January 1, 2021

Assessment of Student Performance

☞ ***Self-Study and Reading ability Practice***

Instructor will give out the chapters or the reference books to read and use class hours to have discussion; students should be able to show a proactive attitude and ability for self-study and reading. Knowledge and oral English will be elements of homework or presentation score.

☞ ***Homework***

Students should finish their homework by themselves. Copying from others will be treated as cheating and the homework scores will be lowered. Students should hand in all assignments on time. Late assignments will be accepted at the discretion of the instructor (i.e., when the student was ill or had an excused absence). Late assignments without reasonable proof will be reduced in score by 50%.

☞ ***Attendance***

Because the course covers a great deal of material, attending every class session is very important for performing well.

- Being late for 15 minutes or more is considered an absence.
- Five hours or above of unexcused absences will result in the lower level of the final grade by one grade band (e.g. from C – to D +). Any excused absence must be discussed directly with the teacher.
- Absence which is more than 1/3 of the total teaching hours will cause an F (a failing grade) directly.

but students are welcome to continue attending classes.

- An incomplete grade (I) will be considered in case of medical or family emergencies.

☞ ***Participation***

- Students should participate in classes actively. Half of participation grade is determined by their presentation in class. They are encouraged to ask questions relevant to the subject and express their own opinions. Every student should respect the ideas, opinions, and questions of their classmates.
- Students should also use office hours to ask questions or talk with the instructor for good communication and effective learning.
- Frequent visiting the instructor and chatting in English during office hours is highly recommended.
- Any misbehavior and non-class related activities in class will result in the lower level of the participation grade, including ringing cell phones.
- All above behaviors will be solely evaluated by the instructor for scoring.

☞ ***Textbook***

Students must bring the textbook to class.

Topical Course Outline

Week Index	Topics
Week 1	**Syllabus & Orientation** **Chapter 1 Brief Introduction of Big Data** Section 1.1: Basic concepts Section 1.2:official definition of big data Section 1.3: Features of big data
Week 2	**Chapter 1 Brief Introduction of Big Data** Section 1.4: Challenges and opportunities Section 1.5: Applications of big data analysis
Week 3	**Chapter 1 Brief Introduction of Big Data** Section 1.5: Platforms of big data analytics **Chapter2 Brief Introduction of Hadoop** Section 2.1: What is Hadoop?
Week 4	**Chapter2 Brief Introduction of Hadoop** Section 2.2: Why Hadoop? Section 2.3: The Hadoop ecosystem
Week 5	**Chapter2 Brief Introduction of Hadoop** Section 2.4: Introduction of HDFS Section 2.5: Features of HDFS
Week 6	**Chapter2 Brief Introduction of Hadoop** Section 2.6: How HDFS works Section 2.7: Introduction of MapReduce
Week 7	**Chapter2 Brief Introduction of Hadoop** Section 2.8: What is Hive? Section 2.9: Applications of Hive

Week 8	**Midterm Examination**
Week 9	**Chapter 3 Correlation** Section 3.1: Predictions and Predilections Section 3.2: Illusions and Illuminations
Week 10	**Chapter 3 Correlation** Section 3.3: Man and Manhole Section 3.4: The End of Theory?
Week 11	**Chapter 4 Datafication** Section 4.1: Qualifying the World Section 4.2: When Words, locations and interactions become data
Week 12	**Chapter 4 Datafication** Section 4.3: The Datafication of Everything Section 4.4: More examples and cases
Week 13	**Chapter 5 Some Algorithms** Section 5.1: Machine learning algorithms Section 5.2: Linear regression
Week 14	**Chapter 5 Some Algorithms** Section 5.3: K-nearest neighbors for classification Section 5.4: K-means for clustering
Week 15	Review and Presentations
Week 16	Review and Presentations

Note: *Some chapters or sections may leave for self-study, this is the students' duty to learn and understand, they may also be included in the quizzes or exams Final exam is in term of presentations.*
A review in Chinese may be held during L.C. and O.H. in the semester.

Teacher's Office Hour

- The instructor's office hour is shown in the front of the office door.
- Students are suggested to use the instructor's office hour and learning center to ask questions or talk with the instructor once at least per week for good communication and effective learning, which is recorded in the students' participation.
- The time can be scheduled by instructors or students, or both.

Cheating and Plagiarism

Cheating is not tolerated. Any student caught cheating on a quiz; test or exam will be given a mark of zero (0) for the particular work. At the beginning of the semester the definition of plagiarism will be carefully explained, when any thoughts or writings of another person are used, they must be clearly identified (usually one uses quotation marks) and the source notes. **If any student is caught cheating on any homework assignment, the highest score the student can earn in that course is a "C".**

Important Dates

TBA

Note: *This syllabus is tentative and may be changed or modified throughout the semester. All students will be notified and a new syllabus will be given.*

Instructor: **Emma Zhu** **Department Head:** **Jingning Li**