City University of Hong Kong
Course Syllabus

offered by Department of Information Systems
with effect from Semester A in 2022 / 2023

Part I Course Overview

Course Title:	Techniques for Big Data
Course Code:	IS3100
Course Duration:	One Semester
Credit Units:	3
Level:	B3
Proposed Area: *(for GE courses only)*	☐ Arts and Humanities ☐ Study of Societies, Social and Business Organisations ☐ Science and Technology
Medium of Instruction:	English
Medium of Assessment:	English
Prerequisites: *(Course Code and Title)*	IS3230 Java Programming for Business, or IS2240 Python Programming for Business, or CB2240 Introduction to Business Programming in Python, or Other programming courses
Precursors: *(Course Code and Title)*	Nil
Equivalent Courses: *(Course Code and Title)*	Nil
Exclusive Courses: *(Course Code and Title)*	Nil

Part II Course Details

1. Abstract

(A 150-word description about the course)

Nowadays, the rapid development in social media, smart devices, and the Internet of Things (IoT) resulted in an explosion of data available for analysis. The unprecedented abundance of data gives both technical challenges in processing and value creation opportunities in analytics, which together form the concept of Big Data.

In this course, students will learn to appreciate the superior value of data through analytics in different applications. To handle the large amount data for analysis, new technology solutions need to be adopted. This course introduces the practical skills to process and analyse data, which can facilitate more innovative data analysis and decision making applications.

The course will use Python and third party packages in data collection, preparation, preliminary modelling, and visualization. The course will also study issues related to Big Data storage and implementation using Hadoop, Spark and related technologies. Upon successful completion of this course, students should be able to better facilitate data-related applications and generate business value.

2. Course Intended Learning Outcomes (CILOs)

(CILOs state what the student is expected to be able to do at the end of the course according to a given standard of performance.)

No.	CILOs[#]	Weighting* (if applicable)	Discovery-enriched curriculum related learning outcomes (please tick where appropriate)		
			A1	*A2*	*A3*
1.	Develop understanding of the challenge brought by big data and the value of combining different data to develop innovative business solutions.	35%	✓		
2.	Acquire techniques on Python packages and the tools in the big data application stack, such as Hadoop, Spark, etc.	30%	✓	✓	
3.	Learn to build solutions with data mining algorithms by examples.	10%	✓	✓	✓
4.	Learn to build and run large-scale data processing demo programs.	25%	✓	✓	✓
		100%			

** If weighting is assigned to CILOs, they should add up to 100%.*

[#] Please specify the alignment of CILOs to the Gateway Education Programme Intended Learning outcomes (PILOs) in Section A of Annex.

A1: Attitude
Develop an attitude of discovery/innovation/creativity, as demonstrated by students possessing a strong sense of curiosity, asking questions actively, challenging assumptions or engaging in inquiry together with teachers.

A2: Ability
Develop the ability/skill needed to discover/innovate/create, as demonstrated by students possessing critical thinking skills to assess ideas, acquiring research skills, synthesizing knowledge across disciplines or applying academic knowledge to self-life problems.

A3: Accomplishments
Demonstrate accomplishment of discovery/innovation/creativity through producing /constructing creative works/new artefacts, effective solutions to real-life problems or new processes.

3. Teaching and Learning Activities (TLAs)

(TLAs designed to facilitate students' achievement of the CILOs.)

TLA	Brief Description	CILO No.				Hours/week (if applicable)
		1	2	3	4	
TLA1: Lecture	Introduce the conceptualization of Big Data. Introduce different applications in the Big Data area. Teach programming techniques and the MapReduce framework. Small Exercises on programming implementation and big data techniques.	✓	✓	✓		3 Hours/Week

4. Assessment Tasks/Activities (ATs)

(ATs are designed to assess how well the students achieve the CILOs.)

Assessment Tasks/Activities	CILO No.				Weighting*	Remarks[#]
	1	2	3	4		
Continuous Assessment: 50%						
AT1: Continuous Assessment Students are encouraged to discuss and reflect on the materials covered in the lecture and attempt extra laboratory exercises.	✓	✓			10%	
AT2: Project & Assignments Assignments will be given to assess the students' capability to creatively apply the skills to solve problems in big data analytics. Group project will be conducted to give students more chance to deepen their understanding of the technique concepts and the programming skills introduced during the lectures and the tutorials.	✓	✓	✓	✓	40%	
Examination: 50% (duration: one 2-hour exam)						
AT3: Examination This will assess both the conceptual understanding and the modelling skills using one or more small case studies. Students will be assessed via the examination on their understanding of concepts learned in class and reading materials, and their ability to apply subject-related knowledge.	✓	✓	✓	✓	50%	
** The weightings should add up to 100%.*					100%	

[#] Remark: Students must pass BOTH coursework and examination in order to get an overall pass in this course.

5. Assessment Rubrics

(Grading of student achievements is based on student performance in assessment tasks/activities with the following rubrics.)

Assessment Task (AT)	**Criterion**	**Excellent (A+, A, A-)**	**Good (B+, B, B-)**	**Fair (C+, C, C-)**	**Marginal (D)**	**Failure (F)**
AT1: Continuous Assessment	CILO 1-2 Demonstrate evidence of active learning through participating in the class discussion, asking critical questions and completing extra-credit activities.	High	Significant	Moderate	Basic	Not even reaching marginal levels
AT2: Project and Assignments	CILO 1 Ability to discover and describe the big data challenges and solutions.	High	Significant	Moderate	Basic	Not even reaching marginal levels
	CILO 4 Capability to run large-scale data processing applications.	High	Significant	Moderate	Basic	Not even reaching marginal levels
	CILO 2-3 Capability to accomplish computation and machine learning tasks with big data tools.	High	Significant	Moderate	Basic	Not even reaching marginal levels
	CILO 1-2 Ability to communicate system and algorithm design with oral and written forms.	High	Significant	Moderate	Basic	Not even reaching marginal levels
AT3: Examination	CILO 1 Understanding of big data opportunities and big data challenges.	High	Significant	Moderate	Basic	Not even reaching marginal levels
	CILO 2-4 Capability to accomplish computation tasks and machine learning with big data tools.	High	Significant	Moderate	Basic	Not even reaching marginal levels
	CILO 1-2 Ability to communicate system and algorithm design.	High	Significant	Moderate	Basic	Not even reaching marginal levels

Part III Other Information (more details can be provided separately in the teaching plan)

1. Keyword Syllabus

(An indication of the key topics of the course.)

Big Data Concept; Big Data Applications; Python; Data Analysis Process; Data Collection; Data Modelling; Big Data Infrastructure; Distributed File System; Parallel Data Processing.

2. Reading List

2.1 Compulsory Readings

(Compulsory readings can include books, book chapters, or journal/magazine articles. There are also collections of e-books, e-journals available from the CityU Library.)

1.	Wes McKinney, Python for Data Analysis, Wes McKinney, 2e, O'Reilly, 2017.
2.	Alberto Boschetti, Luca Massaron, Python Data Science Essentials, 2e, Packet Publishing, 2016.
3.	Benjamin Bengfort & Jenny Kim, Data Analytics with Hadoop, O'Reilly, 2016.
4.	Dirk Deroos et al., Hadoop for Dummies, Dreamtech Press, 2014.
5.	Bart Baesens, Analytics in a Big Data World: The Essential Guide to Data Science and its Applications, Wiley, 2014.

2.2 Additional Readings

(Additional references for students to learn to expand their knowledge about the subject.)

1.	Erik Brynjolfsson et al., The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies, W. W. Norton & Company, 2014.
2.	John W. Foreman, Data Smart: Using Data Science to Transform Information into Insight, Wiley, 2013.

2.3 Online Resources

Readings will be augmented by pertinent journal/newspaper/magazine articles.