

Academic Org: Div of Computer Science & Engg – Subject: Courses offered by Fac of Erg

Course: ENGG5108 **Course ID:** 012049 **Eff Date:** 2023-07-01 **Crse Status:** Active **Apprv. Status:** Approved **[Course Rev]**
Big Data Analytics 大數據分析學

This course aims at teaching students the state-of-the-art big data analytics, including techniques, software, applications, and perspectives with massive data. The class will cover, but not be limited to, the following topics: advanced techniques in distributed file systems such as Google File System, Hadoop Distributed File System, CloudStore, and map-reduce technology; similarity search techniques for big data such as minhash, locality-sensitive hashing; specialized processing and algorithms for data streams; big data search and query technology; managing advertising and recommendation systems for Web applications. The applications may involve business applications such as online marketing, computational advertising, location-based services, social networks, recommender systems, healthcare services, or other scientific applications.

本科旨在教導學生最先進的針對大數據的分析，包括技術、軟件、應用和遠景。本科內容將包括，但不限於以下內容：分佈式文件系統的高等技術如谷歌文件系統，Hadoop文件系統，CloudStore和map-reduce；大數據的相似搜索技術，如最小哈希，局部敏感哈希等；針對數據流的專門處理方法和算法；大數據的搜索和查詢技術；互聯網應用中的廣告管理和推薦系統。本科涉及的應用程序可能包括商業應用程序，如網絡營銷、計算廣告、基於位置的服務、社交網絡、推薦系統、醫療保健服務或其它科學領域的應用。

Grade Descriptor:

A

EXCELLENT – exceptionally good performance and far exceeding expectation in all or most of the course learning outcomes; demonstration of superior understanding of the subject matter, the ability to analyze problems and apply extensive knowledge, and skillful use of concepts and materials to derive proper solutions.

有關等級說明的資料，請參閱英文版本。

B

GOOD – good performance in all course learning outcomes and exceeding expectation in some of them; demonstration of good understanding of the subject matter and the ability to use proper concepts and materials to solve most of the problems encountered.

有關等級說明的資料，請參閱英文版本。

C

FAIR – adequate performance and meeting expectation in all course learning outcomes; demonstration of adequate understanding of the subject matter and the ability to solve simple problems.

有關等級說明的資料，請參閱英文版本。

D

MARGINAL – performance barely meets the expectation in the essential course learning outcomes; demonstration of partial understanding of the subject matter and the ability to solve simple problems.

有關等級說明的資料，請參閱英文版本。

F

FAILURE – performance does not meet the expectation in the essential course learning outcomes; demonstration of serious deficiencies and the need to retake the course.

有關等級說明的資料，請參閱英文版本。

Equivalent Offering:

Units:	3 (Min) / 3 (Max) / 3 (Acad Progress)
Grading Basis:	Graded
Repeat for Credit:	N
Multiple Enroll:	N
Course Attributes:	MSc Computer Science MPhil-PhD Computer Sci & Erg MPhil-PhD Electronic Erg MPhil-PhD Info Engineering MPhil-PhD Mechan & Auto Erg MPhil-PhD System Erg & Erg Mgt MPhil-PhD Information Engineering MPhil-PhD Biomedical Engineering

Topics:

COURSE OUTCOMES

Learning Outcomes:

At the end of the course of studies, students will have acquired the ability to
1. understand the key issues on big data and the associated applications in intelligent business and scientific computing.

2. acquire fundamental enabling techniques and scalable algorithms in big data analytics.
3. interpret business models and scientific computing paradigms, and apply software tools for big data analytics.
4. achieve adequate perspectives of big data analytics in marketing, financial services, health services, social networking, astrophysics exploration, and environmental sensor applications, etc.

Course Syllabus:

This course aims at teaching students the state-of-the-art big data analytics, including techniques, software, applications, and perspectives with massive data. The class will cover, but not be limited to, the following topics: advanced techniques in distributed file systems such as Google File System, Hadoop Distributed File System, CloudStore, and map-reduce technology; similarity search techniques for big data such as minhash, locality-sensitive hashing; specialized processing and algorithms for data streams; big data search and query technology; managing advertising and recommendation systems for Web applications. The applications may involve business applications such as online marketing, computational advertising, location-based services, social networks, recommender systems, healthcare services, or other scientific applications.

Assessment Type:

Essay test or exam	: 40%
Lab reports	: 30%
Others	: 20%
Presentation	: 10%

Feedback for Evaluation:

1. Quiz and examinations
2. Course evaluation and questionnaire
3. Question-and-answer sessions during class
4. Student consultation during office hours or online

Required Readings:

To be provided by course teacher.

Recommended Readings:

1. Frank J. Ohlhorst. Big Data Analytics: Turning Big Data into Big Money. Wiley. 2012.
2. Toby Segar. Programming Collective Intelligence: Building Smart Web 2.0 Applications. O'Reilly Media, 2007.
3. Paul Zikopoulos, Chris Eaton, Paul Zikopoulos, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw-Hill Osborne Media, 2011.
4. Anand Rajaraman and Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press, 2011.

1. ENGG5108

Acad Organization=CSEGV; Acad Career=RPG

COMPONENTS

LEC : Size=30; Final Exam=Y; Contact=3
TUT : Size=30; Final Exam=N; Contact=1

ENROLMENT REQUIREMENTS

1. ENGG5108

Enrollment Requirement Group:

For students in MSc Computer Science or MPhil-PhD programmes under Faculty of Engineering or UG Computer Science
or UG Computer Engineering;
Not for students who have taken CMSC5741 or CSCI5510 or SEEM5020

New Enrollment Requirement(s):

Exclusion = SEEM5020

CAF

< E N D O F R E P O R T >