

Academic Org: Div of Computer Science & Engg – Subject: Computer Science

**Course:** CSCI5570      **Course ID:** 012233      **Eff Date:** 2022-07-01      **Crse Status:** Active      **Apprv. Status:** Approved      **【Course Rev】**  
Large Scale Data Processing Systems 大規模數據處理系統

This course introduces contemporary systems for large scale data processing. Topics to be covered include, but are not limited to: (1) advanced database systems (including distributed, parallel, columnar, in-memory systems, etc., for both OTLP and OLAP applications); (2) NoSQL and NewSQL systems; (3) distributed data stores; (4) big data analysis systems; (5) graph processing systems; (6) stream processing systems; and (7) data visualization. Advanced algorithms for data analytics (e.g., distributed machine learning algorithms, streaming algorithms, etc.) that are implemented using the systems introduced in the course will also be discussed.

本科介紹當代系統的大規模數據處理。主題包括但不限於：（1）高級數據庫系統（包括分佈式的、平行的、柱狀的、內存系統等，為OTLP和OLA所應用）；（2）NoSQL和NewSQL系統；（3）分佈式數據存儲；（4）大數據分析系統；（5）圖處理系統；（6）流處理系統；和（7）數據可視化。在課程中介紹正在使用該系統執行作數據分析的高級算法（例如，分佈式機器學習算法，流算法等）也將被討論。

**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

**Topics:**

**COURSE OUTCOMES**

**Learning Outcomes:**

At the end of the course of studies, students will have acquired the ability to

1. understand the key concepts in the design and development of systems for large scale data processing;
2. understand the key ideas in the design and implementation of contemporary large scale systems for processing different types of data, and their applications;
3. analyze the strengths and limitations of various contemporary systems for large scale data processing;
4. master the basic skills and techniques for processing different types of data using systems introduced in the course.

**Course Syllabus:**

This course introduces contemporary systems for large scale data processing. Topics to be covered include, but are not limited to: (1) advanced database systems (including distributed, parallel, columnar, in-memory systems, etc., for both OTLP and OLAP applications); (2) NoSQL and NewSQL systems; (3) distributed data stores; (4) big data analysis systems; (5) graph processing systems; (6) stream processing systems; and (7) data visualization. Advanced algorithms for data analytics (e.g., distributed machine learning algorithms, streaming algorithms, etc.) that are implemented using the systems introduced in the course will also be discussed.

**Assessment Type:**

Essays	: 20%
Essay test or exam	: 40%
Others	: 40%

**Feedback for Evaluation:**

1. Assignments, course project and essays
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. Andrew S. Tanenbaum, Maarten van Steen: Distributed systems - principles and paradigms (2. ed.). Pearson Education 2007, ISBN 978-0-13-239227-3.
2. Jerome Saltzer, M. Frans Kaashoek: Principles of Computer System Design: An Introduction (1. ed.). Morgan Kaufmann 2009, ISBN: 978-01-2374957-4.
3. Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom: Database systems - the complete book (2. ed.). Pearson Education 2009, ISBN 978-0-13-187325-4.
4. M. Tamer Özsu, Patrick Valduriez: Principles of Distributed Database Systems (3. ed.). Springer 2011, ISBN 978-1-4419-8833-1.

**OFFERINGS**

1. CSCI5570 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. CSCI5570 **Enrollment Requirement Group:**  
For students in MSc Computer Science or MPhil-PhD Computer Science & Engineering or UG Computer Science or UG Computer Engineering

**CAF**

---

---

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