Course: CSCI5570  
Course ID: 012233  
Eff Date: 2024-07-01  
Crse Status: Active  
Apprv. Status: Approved  

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.

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.

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.
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.

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.

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.

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

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:

<table>
<thead>
<tr>
<th>Type</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essays</td>
<td>20%</td>
</tr>
<tr>
<td>Essay test or exam</td>
<td>40%</td>
</tr>
<tr>
<td>Others</td>
<td>40%</td>
</tr>
</tbody>
</table>

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:

OFFERINGS
1. CSCI5570
   Acad Organization=CSEGv; Acad Career=RPG

COMPONENTS
LEN : 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

Additional Information

VTL-Onsite face-to-face hrs  0
VTL-Online synch. hrs       0
VTL-Online asynch. hrs      0

<END OF REPORT>