DSAA 5012
Advanced Database Management for Data Science

COURSE INFORMATION

COURSE INFORMATION

INSTRUCTOR

Lei CHEN Room 3561 2358-6980 leichen@cse.ust.hk

COURSE SCHEDULE

Lecture L1 Tue, Thur 4:30PM - 5:50PM @Zoom

Office hrs. by appointment

COURSE WEB SITE

https://www.cse.ust.hk/~leichen/courses/DSAA5012/ (can be accessed from Canvas)

COURSE TEXTBOOK

TEXTBOOK

1. *Database System Concepts, A. Silberschatz, H. Korth, and S. Sudarshan.*
2. *Database Management Systems, Raghu Ramakrishnan and Johannes Gehrke.*
3. *Data Mining -- Concepts and Techniques by Jiawei Han and Micheline Kamber. Morgan Kaufmann Publishers.*
4. *Community Detection and Mining in Social Media, by Lei Tang and Huan Liu. Morgan & Claypool Publishers.*
5. *Social Network Data Analytics, by Charu C. Aggarwal, Springer*

COURSE REQUIREMENTS

Requirement		Value
Project		50%
Final Exam	May	50%

The final exam are open book, but only course material (i.e., textbook, lecture notes, tutorial notes and lab notes) are permitted. The final exam is cumulative with emphasis on the post-midterm material.

COURSE OBJECTIVES

Understand how a database management system for data science applications.

1. An understanding of the concepts and techniques used by a database management system to manage data

2. Experience in designing, implementing and querying a database for a small data scapplication. $\Longrightarrow$ project

EXPECTED COURSE OUTCOMES

After completing this course you are expected to be able to:

1. Explain important database management system concepts including
   - principles of database systems
   - data models
   - logical and physical database design
   - query languages and query processing
   - database services (e.g., concurrency, crash recovery and integrity).

2. Apply database theories to practical data science applications.

3. Analyze a real-life problem, design a database and implement a computer-based system using mysql.

SYLLABUS

Lecture Topics	Lectures
Database Management Systems	1
Entity-Relationship (E-R) Model and Database Design	2
Relational Algebra	1
Structured Query Language (SQL)	3
Relational Database Design	2
Storage and File Structure	1
Indexing	3
Query Processing	3
Query Optimization	2
Transactions	1
Concurrency Control	2
Recovery System	1
NoSQL Databases, graph and uncertain databases	3

→ how to design & query a database (user level)

→ how a DBMS works internally (system level)

COURSE SCHEDULE

See Course Schedule web page for a detailed course schedule.

COURSE PROJECT

Project Overview

Please write a proposal including the following items.

A specific topic (or title) for this project

Type of this project: Survey/Research

Student ID, Student name

A brief description about this project (about 1000~2000 words)

A list of papers to be read in this project

Final Report

Content Survey Type (check ACM Computing Survey)

about 5,000~10,000 words

Research Type (check any Research papers listed below)

about 5,000~10,000 words

Guideline Write a normal report (e.g., Introduction, Related Work, Algorithm, Conclusion, References, ...)

DOWNLOAD AND INSTALL REQUIRED SOFTWARE

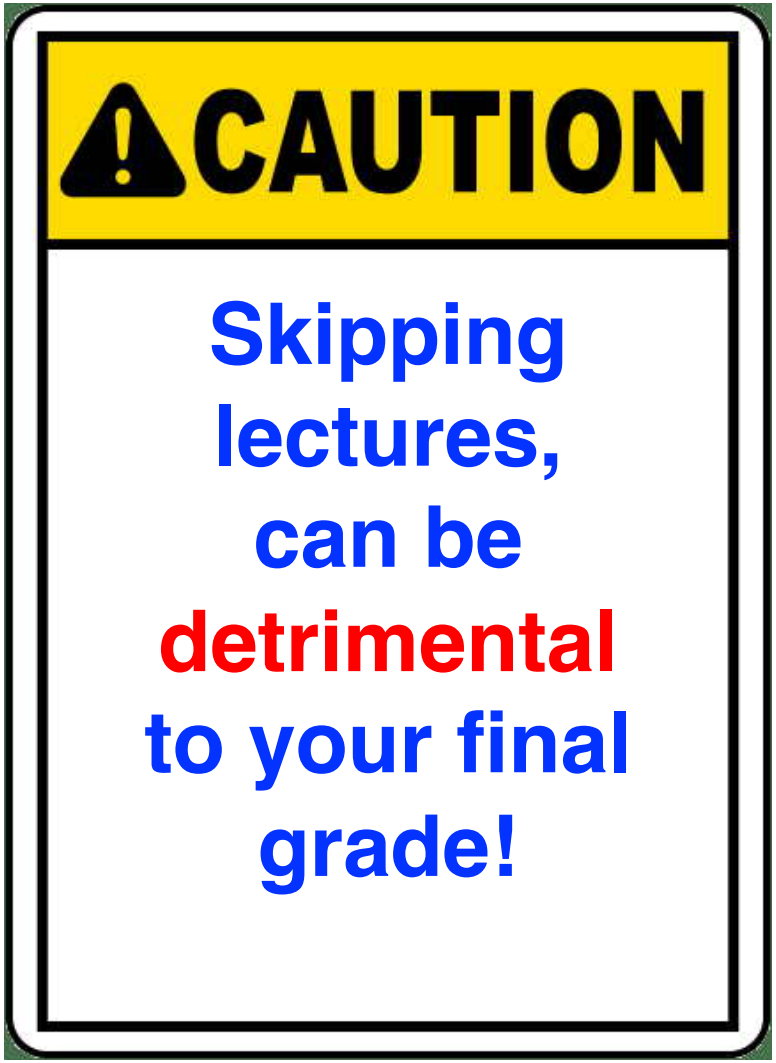
- Pulse Secure (VPN software)

  http://itsc.ust.hk/apps/vpn/

- Oracle SQL Developer (Windows / MacOS / Linux)

  https://www.oracle.com/tools/downloads/sqldev-downloads.html

WELCOME TO DSAA 5012



Questions?