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)
TEXTBOOK


3. *Data Mining -- Concepts and Techniques* by Jiawei Han and Micheline Kamber. Morgan Kaufmann Publishers.


5. *Social Network Data Analytics*, by Charu C. Aggarwal, Springer
# COURSE REQUIREMENTS

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project</td>
<td>50%</td>
</tr>
<tr>
<td>Final Exam</td>
<td>May</td>
</tr>
</tbody>
</table>

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 science application. ⇒ 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

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

- **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)
WELCOME TO DSAA 5012

Questions?

Skipping lectures, can be detrimental to your final grade!