CS4224/CS5424 Lecture 0 Course Admin

CS4224/CS5424: Distributed Databases

- ## Lectures & Discussions
  - ▶ Time: Monday, 630 - 930pm
  - ▶ Location: i3 Auditorium (iCube Building)

- ## Lecturer: Chan Chee Yong
  - ▶ Office: #03-24, COM1 Building
  - ▶ Email: chancy@comp.nus.edu.sg

What's CS4224/CS5424?

This module studies the management of data in a distributed environment. It covers the fundamental principles of distributed data management and includes distribution design, data storage & indexing, distributed query processing and optimization, distributed transaction management, and replication.

Learning objectives:
- Theory: Understand how DDBMSs work
- Practice: Learn how to use DDBMSs

CS4224 vs CS5424

- Same lectures, assignments & projects
- Exam papers are different for CS4224 & CS5424; students will be graded differently

Prerequisite: CS3223

- Data storage

  ▶ Disk-based row store
  ▶ Sequential vs random I/O

- Indexing techniques (e.g., B^+-tree, Hash index)

- Query optimization & evaluation

  ▶ Enumeration of query plans
  ▶ Join algorithms (nested-loop, sort-merge, hash-based)

- Concurrency control

  ▶ Serializability theory
  ▶ Concurrency control protocols (e.g., strict 2PL)

- Crash recovery

  ▶ Log-based redo/undo recovery
  ▶ Write-ahead logging & Force-at-commit protocols

CS3223 Reference Books

- Raghu Ramakrishnan & Johannes Gehrke,
  Database Management Systems,
  McGraw-Hill, Third Edition, 2003.
  (QA76.9 Dbm.Ra 2003)

- Avi Silberschatz, Hank Korth & S. Sudarshan,
  Database Systems Concepts,
  McGraw-Hill, Seventh Edition, 2019.
  (QA76.9 D3.Sil 2020)

- Hector Garcia-Molina, Jeffrey Ullman, & Jennifer Widom,
  Database Systems: The Complete Book,
  Prentice Hall, Second Edition, 2009.
  (QA76.9 Dbm.Gar 2009)

CS4224/CS5424 Material

- ## Reference books
  - ▶ M. Tamer Özsu and Patrick Valduriez,
    Principles of distributed database systems (4th ed),
    Springer 2020.

    http://linc.nus.edu.sg/search/i?SEARCH=9783030262532&searchscope=16&SORT=A

  - ▶ M. Kleppmann,
    Designing Data-Intensive Applications,
    O'Reilly 2017.

    http://linc.nus.edu.sg/search/i?SEARCH=9781449373320&searchscope=16&SORT=A

- ## Papers
  - ▶ Download from **Files** > **Readings** on Canvas
  - ▶ https://canvas.nus.edu.sg/courses/45852/files

Workload

- Number of credits = 4
- Workload per week = 10 hours
  - ▶ 2 lecture hours
  - ▶ 1 tutorial hour
  - ▶ 3 assignment hours
  - ▶ 4 preparatory work hours
- Tutorial discussion sessions
  - ▶ Use for lectures & discussions
- Self-assessment quizzes
  - ▶ Assess your understanding of lecture material
  - ▶ Will not be graded; to be discussed during tutorials
  - ▶ Solutions will be released one week after quiz's release date

Module Assessment

Assessment Component	%
Individual written assignments	20
Team project	30
Closed-book final exam (with 1 sheet of notes)	50

Final Exam

- Date: December 4 (Monday, 5-7pm)

- Venue: To be confirmed

- Closed-book exam (allowed one single A4-sized, double-sided sheet of notes)

Assignments

- Four individual written assignments

Assignment	Tentative Due Date
1	Week 6
2	Week 8
3	Week 10
4	Week 13

- Penalty marks for late submissions

Number of Late Days	Penalty Marks
1	-1
2	-2
3	-3
≥ 4	Assignment will not be graded & will receive 0 marks

Late days are rounded up to the nearest integer. For example, a submission that is late by 1 hour will count as 1 day late.

Assignments (cont.)

- Assignments will be done using Examsoft's Examplify

- While you may discuss or clarify conceptual understanding of course material with your friends, you are to complete each assignment **on your own**
  - ▶ You are not allowed to collaborate with any friend or share answers with your friend

- Requests for assignment re-grading must be submitted to lecturer within one week of the announcement of the graded assignment marks

Examplify Software

- Student guide: https://wiki.nus.edu.sg/display/DA/Student

- If you're new to Examplify, you should watch CIT's briefing video titled Examplify Student Briefing at https://wiki.nus.edu.sg/display/DA/Examplify+Student+Briefing+Video

- The briefing slides are available at

  https://wiki.nus.edu.sg/display/DA/Examplify+Student+Briefing+Slides

- System requirements:

  https://wiki.nus.edu.sg/display/DA/Device+Minimum+System+Requirements

- Installation:

  https://wiki.nus.edu.sg/display/DA/Download+and+Install+Examplify

Examplify Software (cont.)

- CIT will email you once your Examplify account for the module has been setup & ready for use

- Troubleshooting guide:

  https://wiki.nus.edu.sg/display/DA/General+Support+and+Troubleshooting

- Online service desk

  https://wiki.nus.edu.sg/display/cit/Get+help#Gethelp-OnlineServiceDesks

- If you've questions on Examplify, email citbox25@nus.edu.sg

Team Project

I hear and I forget. I see and I remember. I do and I understand. - Confucius

- **Objective**: Hands-on experience with modern distributed database systems
- **Tasks**:
  - ▶ Application development & benchmarking with distributed database systems
  - ▶ Project focuses on back-end implementation (without frontend development)
- **Deliverables**
  - ▶ Written report & Source code
  - ▶ Deliverables are due on November 9 (Thursday, 11:59pm)
  - ▶ Project evaluation during week 13

Team Project (cont.)

- **Project Environment**
  - ▶ **SoC's Compute Cluster (running Ubuntu Linux)**
    - ★ Compute Cluster Guide

      https://dochub.comp.nus.edu.sg/cf/services/compute-cluster

- **Project Teams**
  - ▶ Team of 5 students
  - ▶ Register your project team (name & NUS userid of team members) at **Canvas > Quizzes > Project Team Registration**
    - ★ Registration will open on August 23
    - ★ Registration deadline: **August 31 (Thursday, 11:59pm)**
  - ▶ Each team can consist of CS4224 & CS5424 students
  - ▶ A team with fewer than 5 members may be split or expanded with additional student(s)
  - ▶ A student not in any registered team will be assigned to a random team

Schedule

Week	Topic
1	Introduction
2	Database Design
3	Storage & Indexing
4	Commit Protocols
5	Concurrency Control
6	Data Replication
-	Recess Week
7	Consensus Protocol
8	Data Consistency
9	Query Processing
10	Query Optimization
11	Case Study
12	Case Study
13	Project Evaluation

Course Policies

- Students are responsible for the following:
  - ▶ Attending lectures & discussions
  - ▶ Studying referenced materials
  - ▶ Checking Canvas & emails for course-related announcements
    - ★ You should enable announcement notifications on Canvas: Account > Notifications

- For clarifications on course material, the best way is to post questions on **Canvas Discussions**. You can also email the lecturer to arrange for consultation sessions.

Plagiarism

- **Zero-tolerance for plagiarism/cheating**
- Students will be reported to University for disciplinary action for academic offence
- Resources:
  - ▶ http://www.nus.edu.sg/celc/programmes/plagiarism.html
  - ▶ https://libguides.nus.edu.sg/copyright_essentials_teaching_learning_research/plagiarism
  - ▶ https://www.comp.nus.edu.sg/cug/plagiarism/

Plagiarism (cont.)

All students share the responsibility for upholding the academic standards and reputation of the University. Academic honesty is a prerequisite condition in the pursuit and acquisition of knowledge. Academic dishonesty is any misrepresentation with the intent to deceive or failure to acknowledge the source or falsification of information or inaccuracy of statements or cheating at examinations/tests or inappropriate use of resources. There are many forms of academic dishonesty and plagiarism is one of them. **Plagiarism is generally defined as the practice of taking someone else's work or ideas and passing them off as one's own** *(The New Oxford Dictionary of English). The University does not condone plagiarism.*

Plagiarism (cont.)

Students should adopt this rule - **You have the obligation to make clear to the assessor which is your own work, and which is the work of others.** *Otherwise, your assessor is entitled to assume that everything being presented for assessment is being presented as entirely your own work. This is a minimum standard.*

Unintentional Plagiarism

- A student may not knowingly intend to plagiarise, but that should not be used as an excuse for plagiarism. **Students should seek clarification from their instructors or supervisors if they are unsure whether or not they are plagiarising the work of another person.**

- You also need to protect your own work from being plagiarized. If your work is carelessly accessible by others and is being plagiarized by others, you are responsible too.

  - Example: You should not store your own work in a publicly accessible GitHub repo

Guidelines on the Use of AI Tools

- Don't use AI to plagiarize
  - For example, processing an original source not created by yourself to plagiarize it (e.g., using an AI paraphrasing tool to disguise someone else's original work, or even the output of an AI tool, and then presenting the final output as your own work or idea).

- Remember that just because there are legitimate uses for AI tools in your academic work, it does not mean you should resort to them at every turn, especially if you are still learning the subject matter. By jumping straight to using the tools, you may end up missing an opportunity to learn the subject matter for yourself. Furthermore, if you don't already have the subject matter knowledge yourself, you might not even be able to tell if the output is accurate or relevant.

- Reference: https://libguides.nus.edu.sg/new2nus/acadintegrity#s-lib-ctab-22144949-5

- For CS4224/CS5424, you are not allowed to use AI tools for assignments, project, or exam.
  - If you believe that you've a legitimate use of AI tools for the above, please first consult the lecturer

Copyrighted Course Materials

`https://libguides.nus.edu.sg/copyright_essentials_teaching_learning_research/plagiarism`

The slides, tutorial handouts and other teaching materials uploaded to Canvas are copyrighted and owned by NUS and/or its teaching staff. You are not to: sell, upload, share, or distribute NUS slides, tutorial handouts and other teaching materials on commercial platforms. Doing so is a violation of NUS's policies (e.g. NUS Student Code of Conduct) and Copyright Laws and may subject you to disciplinary action.

Reminder on **Canvas > Files**:

The materials and their content are confidential information and copyright materials of the rights holder. The materials are provided to you solely for your personal use in advancement of education, training and research purposes. All rights are expressly reserved by the rights holder. You may not, except with express written permission from the rights holder, distribute or exploit the content, commercially or otherwise. You may not transmit the materials or store them in any other website or other form of electronic retrieval system.