

W21 COMP 8157: Advanced Database Topics

Lecture 00: Course Outline and Introduction

Dr. Kalyani Selvarajah
kalyanis@uwindsor.ca

School of Computer Science
University of Windsor
Windsor, Ontario, Canada

January 11, 2021



Today's Lecture

- Course Overview
- Introduction to Databases

Why DBMS course?

- Very widely used.
- Part of many software solutions.
- DB expertise is a career asset.
- More Money



Background

- **Fundamental DBMS:**
 - 3-level architecture,
 - introduction to file structures: B-trees, B+ Trees and Hashed files,
 - relational model,
 - relational algebra and calculus,
 - Query languages SQL, PL/SQL
 - database design with Normalization Theory



Course Instructor

Welcome to COMP 8157 Advanced Database Topics Course!

I am Kalyani Selvarajah, former PhD student at University of Windsor and Currently working as Assistant Professor. I did my undergrad and Masters studies from University of Peradeniya, Sri Lanka.

My research interests are Social Network Analysis and Knowledge graphs using AI and Deep learning techniques.



kalyanis@uwindsor.ca



Office hours: Every Monday from 4:00 pm to 6:00 pm on MS Teams

Other times can be arranged by appointment (meaning if you are unable to make my office hours to discuss a concept in a given week, you are encouraged to send me an e-mail asking for a different time. Please include a list of proposed times since I do not have access to your schedule.)

GAs



Head GA: Shaon Bhatta Shuvo

A Ph.D. student in the School of Computer Science and demonstrated teaching experience in different universities in Bangladesh and India. He is skilled in various fields of Artificial Intelligence, especially Machine Learning and Deep Learning. His research interests include Social Network Analysis, Agent-based Modeling, Natural Language Processing, Computer Vision etc. Besides, he has a keen interest in sports, mainly Cricket, Soccer etc.

Office hours: Fridays 2:00 pm to 3:00 pm
Email: shuvos@uwindsor.ca



Yicheng Lu

A thesis-based master student. The main research field is Graph Embedding. Graph embedding is a series of machine learning technologies to capture the graph topology, vertex-to-vertex relationship, and other relevant information about graphs, subgraphs, and vertices.

Office hours: Tuesdays 2:00 pm to 3:00 pm
Email: lu16a@uwindsor.ca

Course Information

This course will introduce students to advanced topics in database design and information retrieval. Topics covered may include DBMS three-schema level architectures, data models (e.g., relational, object-oriented model), query languages (e.g., Oracle SQL, PL/SQL), file organization and indexing, transaction management, concurrency control, security and recovery procedures, information retrieval on the internet, and other advanced topics (e.g., online analytical processing - OLAP, data warehouses and data mining).

#01 Course Name



COMP-8157-01/02
Advanced Database
Topics

#02 Prerequisite



None

#03 Lectures & Lab



Monday(I) or Wednesday (II)
8:30AM to 11:20AM
@ BB virtual classroom

#04 Course Website



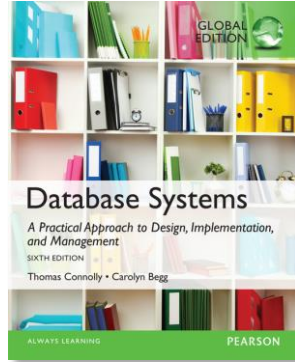
[https://blackboard.uwin
dsor.ca](https://blackboard.uwin
dsor.ca)

Learning Outcome

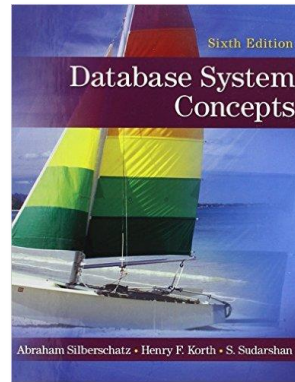
- At the end of this course, successful students are expected to know and be able to:
 - Demonstrate skills of analysis and design of large-scale database systems by using modern development tools.
 - Design databases for real life applications.
 - Understand the various issues related to the development, performance and reliability of largescale and distributed database systems.
 - Understand database technologies used in large-scale applications such as Google search Engines.
 - Understand Big Data analytics, Transaction management, and Failure recovery.
 - Research and apply the latest database management strategies (e.g. from ACM SIGMOD).
 - Evaluate and compare different data models, as well as select and implement the most appropriate models.
 - Recognize and discuss the importance of protecting data against component faults and from unauthorized access.
 - Describe and explain how other components of an application will interface with the database.



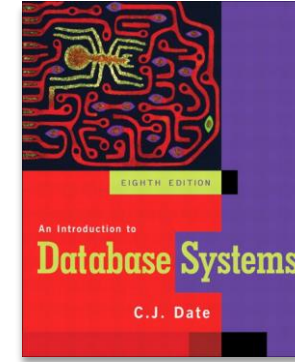
Recommended References



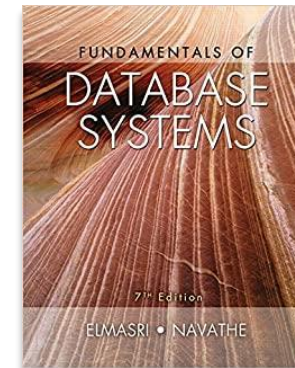
Database systems: a practical approach to design, implementation, and management (6th Edition). T. Connolly and C. Begg.



Database System Concepts (6th Edition). A. Silberschatz, H. F. Korth, and S. Sudarshan



An Introduction to Database Systems (8th Edition). C.J. Date

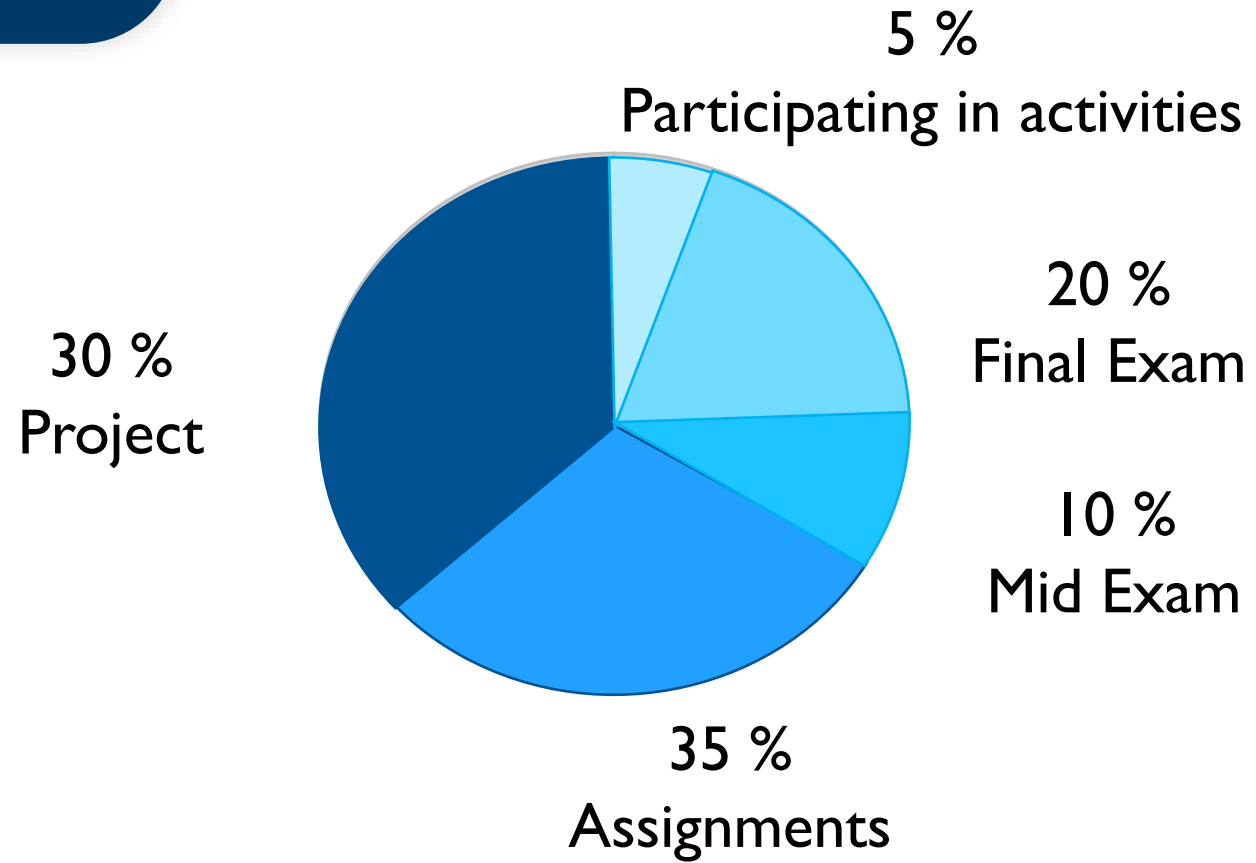


Fundamentals of Database Systems. R. Elmasri & S. B. Navathe

Course Schedule: (**schedule may vary):

Weeks	Topics	Important Dates	Deadlines
1	Course Outline and Introduction		
2	Introduction to Big Data	Jan 20th ADD/Drop	Assignment 1
3	Introduction to Distributed DBMS		Assignment 2, P1: Project Proposal
4	Distributed DBMS Design	Feb 3rd Final Drop	Assignment 3
5	Distributed DBMS Transparency		P2:Work Distribution
6	Reading Week (No Classes)		
7	File organization and indexing		
8	Transaction Management		Assignment 4, P3: Milestone I
9	Concurrency Control		Mid Exam
10	Deadlock and Recovery	Mar 17th VW	Assignment 5
11	Security and Administration		Assignment 6, P4: Milestone II
12	Data Mining/Data Streaming		Assignment 7, P6: Final Report P5: Project Presentation
13	Courses wrap up		

Course Evaluation



Assignments

- LinkedIn Learning Materials.
 - A1: Introduction to Nosql
 - A2: MongoDB
 - A3: Kafka
 - A4: Hadoop
 - A5: Spark
 - A6: Graph database
 - A7: Data Visualization
- Assignment grading:
 - Complete the exercises which we provide and submit the results/answers as GA requested in the lab.
 - Submit the certificate

Group Project

The goal of this project is to develop knowledge on advanced database technology. You can propose a project that you think is interesting and valuable and goes in depth on database implementation on various aspects within the field of databases.

#01 Form Teams

 Jan 18, 2021

#02 Proposal

 Jan 31, 2021

#03 P2: Work Distribution

 Feb 7, 2021


#05 Milestone I

 Mar 7, 2021

#06 Milestone II

 Mar 21, 2021

#07 Demo &
Presentation

 Mar 29/31, 2021

#08 Final Report

 Apr 7, 2021

Forming Teams

- The project needs to be developed by a group of **4 students** (Mandatory). A group of 5 members is acceptable in an unexpected scenario.
- All members of a group will receive the same grade on group work.
 - Therefore, it is in your interest to choose other group members who have the same goal in the class as you do.
 - We will closely monitor the contribution of each member to their group project.

Proposal

You can propose a project that you think is interesting and valuable and goes in depth on database implementation on various aspects within the field of databases.

- Focus on value to the final outcome.
- Be realistic - Think of a project you can complete (you will have limited timeline).

Suggestion: You can create a database model based on the latest database management strategies such as Distributed Databases, Data Mining, Graph Databases, Timeseries Databases, Cloud Databases and Privacy. You are welcome to choose other than the above list of topics.

However, it should not be a topic related to simple database management system such as library management system and Hospital management system. It is not an undergrad project or fundamental database.

Proposal

You will produce the following two deliverables, and submit them on BB:

1. a single page document (absolutely no more than 5 min read), and
2. a URL of a video presentation (use of 1-4 slides) for 3 minutes. (use OneDrive and share it)

You should include:

- (i) a description of the problem,
- (ii) the motivation for the problem (e.g., why is the problem interesting, why is it challenging, who will benefit from a solution to the problem, etc.),
- (iii) your initial ideas on how to solve the problem, and
- (iv) a brief discussion of previous work related to this problem if exists.

Communication

MS Teams

BB

Uwindsor Email

You can also reach the instructor through MS Teams (highly recommended) or email.

*** I will respond back to you at the earliest possibility. If you don't hear back from instructor or GA/TAs within 48 hours, you can send a follow up message or email.

Repository

GitHub

<https://www.youtube.com/watch?v=nI5VdsVI0FM>

You can reach our GA/TAs for further help.

Your project should be stored and maintained in a GitHub Repository.

Missed Test Make-up Policy

For missing a mid-exam, no make-ups will be considered. If a student misses the mid exam for a valid reason, the corresponding mark will be carried over to the final exam.

Students who miss the Final Exam for a valid reason (as per the Senate bylaws) which is given to the instructor before (if possible) or within 48 hours after the exam date, will be allowed to take a makeup exam. The Registrar will assign a date and time for writing a make-up Final Examination. Missing the Final Exam without a valid reason will result in a score of zero.

Late Assignment Policy

Late assignments due to a valid medical reason that has been accepted will be provided a deadline extension at the instructor's discretion.

If you are in need of a medically related extension – you must make arrangements in advance of the original due date/time.

If not, the students will be penalized 5% per day to a maximum of 50%, including weekends and holidays, for overdue assignments.

Policy on Cheating

The instructor and teaching assistants for COMP-8157 will put a great deal of effort into helping students to understand and to learn the material in the course. However, they will not tolerate any form of cheating.

The instructor and teaching assistants will report any suspicion of cheating to the Director of the School of Computer Science. Please read the link carefully for further information: <http://www.uwindsor.ca/academic-integrity/>

A plagiarism-detection software may be used for some or all of the student homework assignments and for your project report in this course, as the case may be, at the instructor's discretion as deemed necessary.

Academic integrity

You are expected to do all of your work in any of the labs, exams and project individually, without the help of others.

In cases of academic misconduct as defined in Senate Bylaw 31, instructor is required to report misconduct for necessary disciplinary action.

In such cases, you will get zero points for that lab, project and/or question in exam. This will be irrespective of who cheated from whom. You are responsible to protect your work from others.

Plagiarism includes, but is not limited to:

- a) Copying material, for example, from the Internet, or purchasing material and submitting it as one's own.
- b) Paraphrasing (changing some words) the ideas and concepts of others without proper referencing.
- c) Using passage(s) of any length from published or unpublished work of others without placing the passage(s) in quotation marks (or using indentation for long quotation(s)) and acknowledging their source.
- d) Submitting work completely or largely identical to that of other students, unless group work and joint submissions are explicitly permitted by the instructor."

Academic integrity (Cont.)

Remember availability of information on the Internet or you are being able to Google it, doesn't automatically makes it copyright free. You still need to check the terms of agreement (or end user agreement or something similar) for that site. Google or other search engines simply provide it to you but may not own the copyrights.

Please include the following declaration with EACH submission (Assignment, project, quiz, exam). In the absence of such declaration, your submission will NOT be evaluated.

I, _____, verify that the submitted work is my own, original work, that all sources are cited accurately, and that I have not submitted any portion of this work for any other university course.

Signature

Date

Student Responsibilities

- To be polite in all dealings with the professor, the GA/TAs, and the other students.
- To connect with online class on time and ready to participate in the learning process.
- If you miss any announcement, it is your responsibility to catch up on instruction you have missed.
- To ensure that you do not plagiarize in any assignment.
- To ensure to submit all assignments (including project) on time.

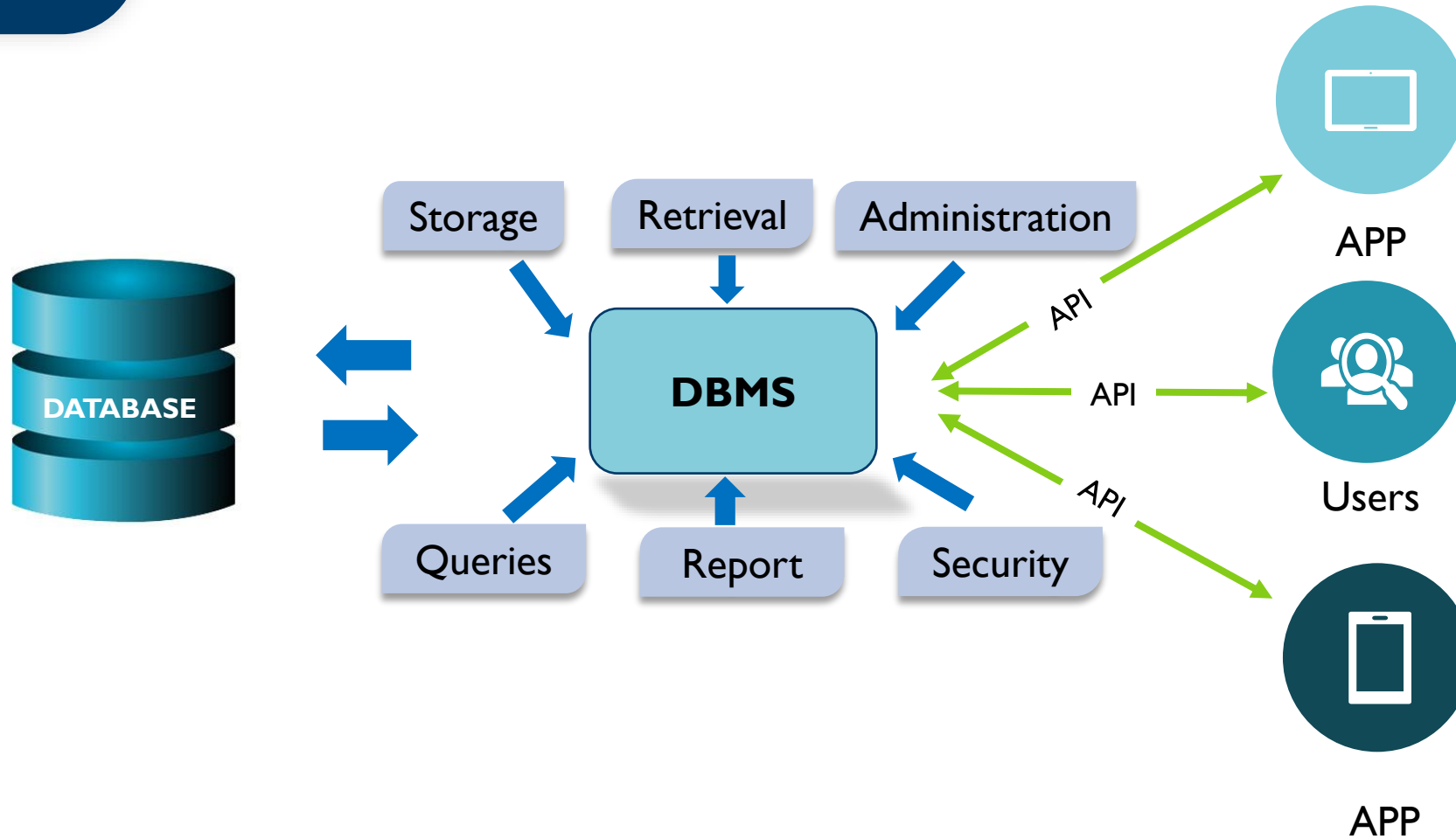
Final Exam

- MCQ
- Take home exam: Written long-form examination on the mandatory readings and topics discussed in class.

Any Question?

Introduction to Database

DBMS



History of Databases



IDS

Honeywell - The first network DBMS.
Charles Bachman

1960

1970

IDMS

Cullinet – by **Charles Bachman**- primarily a network model (CODASYL) database management system for mainframes.

IMS

IBM - to keep track of the supplies and parts inventory for the Saturn V and Apollo space exploration projects.

Relational Model

IBM– by **Ted Codd**- high level of abstraction of databases to avoid all this sort of maintenance burden on humans.

System R, INGRES, Oracle

OBJECT-ORIENTED DATABASES

In 1983

DB2

IBM– “**SEQUEL**” becomes the standard (SQL).- by **Donald D. Chamberlin and Raymond F. Boyce**

1980

1990

SQL SERVER

MySQL

PostgreSQL

SQLite

NoSQL

MongoDB

Cassandra

HBase

Oracle NoSQL

Data Warehouses

Distributed / Shared-Nothing

Netezza, ParAccel

MonetDb, Greenplum

Datallegro, Vertica

2000

2010

NewSQL

a relational database with the scalable properties of NoSQL

VoltDB, Clustrix,

MemSQL, Spanner

Hybrid Systems

Execute fast OLTP like a NewSQL system while also executing complex OLAP queries like a data warehouse system.

Hyper, nappy, JustOne

Cloud System

First database-as-a-service (DBaaS) offerings were "containerized" versions of existing DBMSs.

RedShift, Aurora, from Amazon, Snowflake, Xeround

2020

The World's Largest Data Center



References

- <https://en.wikipedia.org/>
- Pavlo, Andrew, and Matthew Aslett. "What's really new with NewSQL?." ACM Sigmod Record 45.2 (2016): 45-55.

Next Lecture

- Introduction to Big Data