

**CPSC 304**

# **Introduction to Database Systems**

---

## Introduction

Textbook Reference

Database Management Systems: Chapter 1

Hassan Khosravi

Borrowing many slides from Rachel Pottinger

# Learning Goals for Chapter 1

---

- Define the term *database* and explain the purpose of having a database.
- Explain the high-level objectives of a *database management system* (DBMS), and explain how a DBMS relates to a database. List benefits that result from the usage of a DBMS.

# Why use a database?

---

Suppose you are building a system to store the information pertaining to a university from scratch. You have access to an operating system of your choice, but that's it. What do you need to figure out? What do you want your system to do and why?

# Why use a database?

---

Suppose you are building a system to store the information pertaining to the university.

## **You'll need to figure out:**

- How do we store the data? (file organization, etc.)
- How do we query the data? (write programs...)
- Make sure that updates don't mess things up?
- Access requirements (Provide different views on the data for registrar versus students)
- How do we deal with crashes?

# What is a database?

---

- A **database** is an organized collection of related data, usually stored on disk. It is typically:
  - Important data
  - Shared
  - Secured
  - Well-designed (minimal redundancy)
  - Variable size
- A DB typically models some real-world **enterprise**
  - Entities (e.g., students, courses)
  - Relationships (e.g., Ting got 95% in CPSC 221 )

Sounds like our UBC problem!

# Who watches the watchers?

---

- A ***Database Management System (DBMS)*** is a bunch of software designed to store and manage databases. It is used to:
  - Define, modify, and query a database
  - Control access
  - Permit concurrent access
  - Maintain integrity
  - Provide loading, backup, and recovery

# Great! What's left for us to do?

---

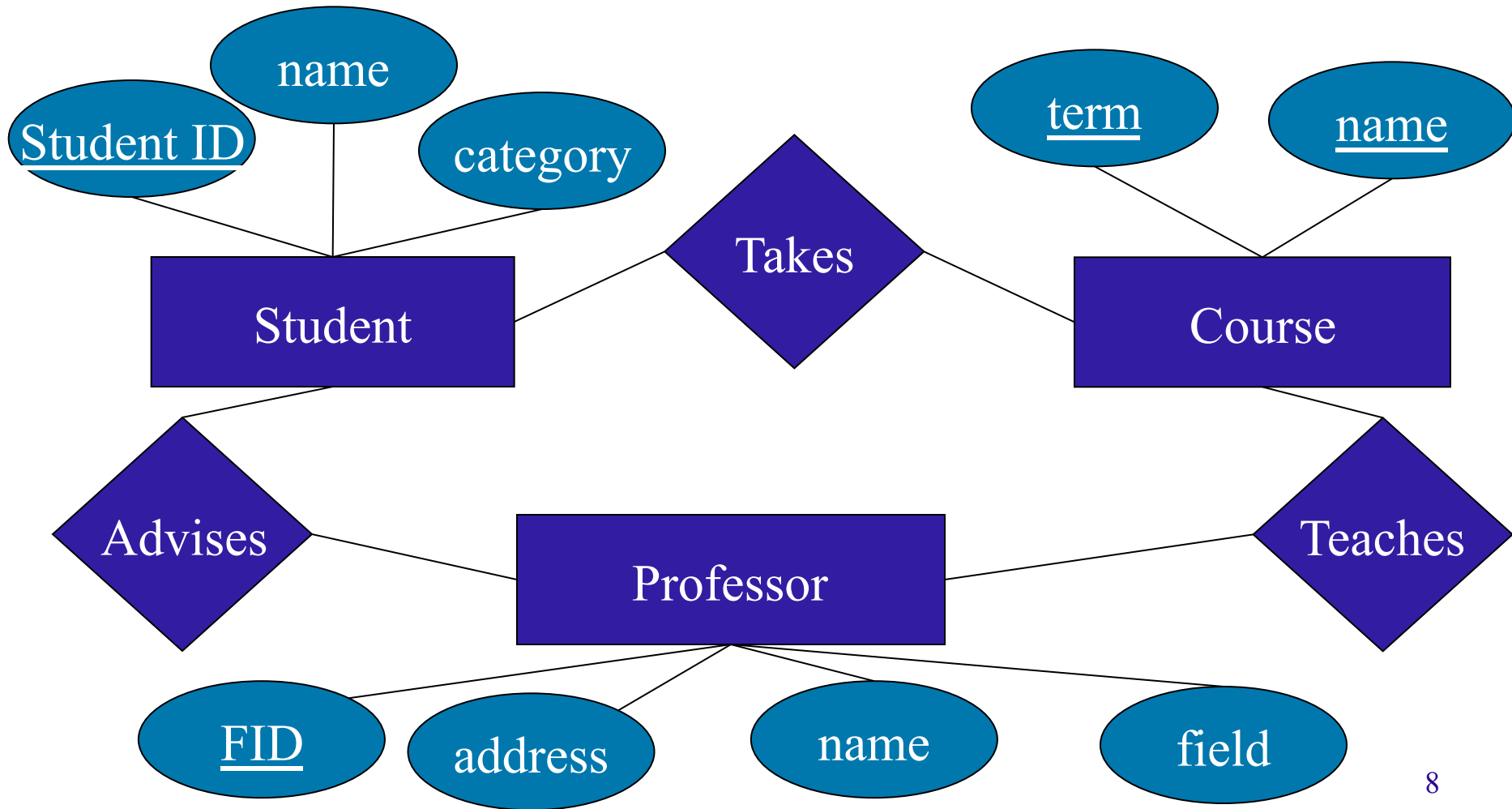
In this course we'll look at all of these ....

1. Conceptually model the concepts
2. *Logically* model concepts in a database
3. Decide which users will ask which **queries**
4. Make our design efficient
5. Create an application and enjoy!
6. Data warehousing, OLAP
7. Semi-structured data (XML)
8. Data mining

# 1. Conceptual Modeling

---

We'll look at Entity-Relationship (ER) Diagrams:





## 2. Logical modeling

---

- Data model : a collection of tools for describing
  - data , data relationships, semantics, constraints
- We'll use the **Relational Model**

**Students Table:**

<b>Student</b>	<b>Course</b>	<b>Term</b>
Ying	CPSC 304	Winter 2, 2006
Andrew	CPSC 221	Summer 2, 2006

### 3. Decide on queries:

---

- We'll mainly use Structured Query Language (SQL):

Example: Find all the students who have taken CPSC 304 in Winter Term 2, 2011.

**select** E.name

**from** Enroll E

**where** E.course="CPSC 304" and

E.term="Winter Term 2, 2011"

- A declarative language – the query processor figures out how to answer the query efficiently
- We'll also look at Relational Algebra and Datalog

## 4. Make our design efficient

---

- We'll learn how to find a design that
  - captures all the information we need to store
  - takes up less space
  - It is easy to maintain
- We'll look at different “Normal Forms”

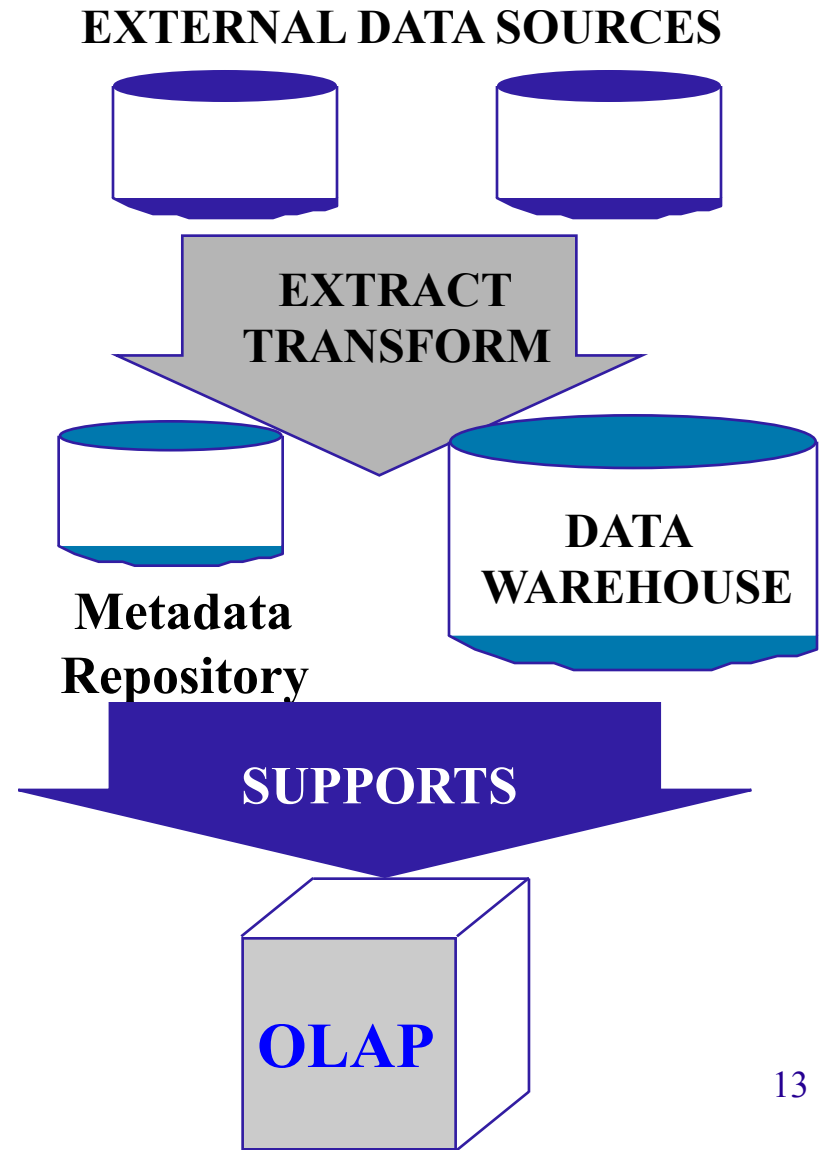
## 5. Create an application

---

- Use a programming language to interface with database
- We'll focus on
  - Java & JDBC *or*
  - the web and PHP

## 6. Data warehousing, OLAP

- Increasingly, organizations are analyzing current and historical data to identify useful patterns and support business strategies.
- The emphasis is on analysis of complex queries on very large datasets created by integrating data from across all parts of an enterprise.



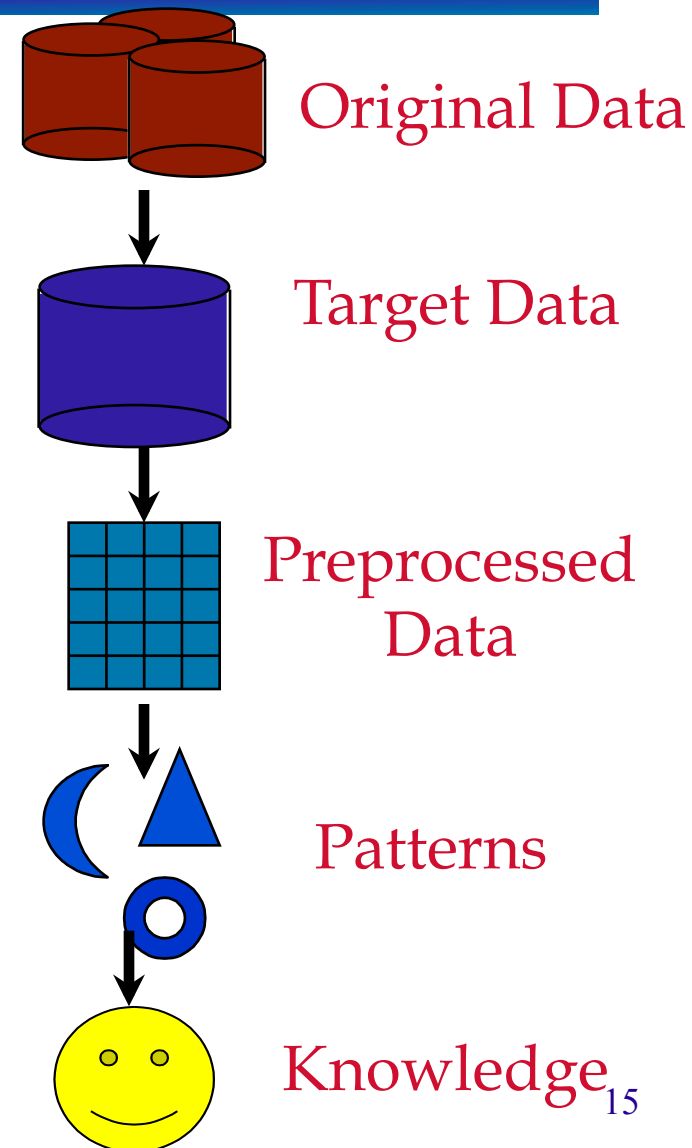
# 7- Semi-Structured Data (XML)

---

- Paradigm Shift on the Web and databases
  - We'll learn what semi-structured data means and why they are becoming more common and important.
  - We'll talk about the difference between XML and HTML.
  - We'll briefly talk about how you can query semi-structured data.

# 8. Knowledge Discovery and Data Mining

- *Knowledge discovery and data mining* is the exploration and analysis of large quantities of data in order to discover valid, novel, potentially useful, and ultimately understandable patterns in data.
- The challenge of extracting knowledge from data draws upon research in **statistics**, **databases**, **pattern recognition**, **machine learning**, **data visualization**, **optimization**, and **high-performance computing**



# Course Learning Outcomes

---

At the end of the course you will be able to :

- describe how relational databases store and retrieve information
- develop a database that satisfies the needs of a small enterprise using the principles of relational database design
- express data queries using formal database languages like relational algebra, tuple and domain relational calculus
- express data queries using SQL
- develop a complete data-centric application with transactions and user interface.
- explain how user programs interact with a database management system
- identify the goals of data warehousing and Justify the use of a DW for decision-making
- explain the general steps involved in knowledge discovery from data (e.g., data acquisition, cleaning, integration, selection, transformation, mining)



# What's left for CPSC 404 then?

---



- How does a DBMS perform
  - Data storage
  - Data independence and efficient access.
  - Uniform data administration (centralized control).
  - Query evaluation and optimization
  - Transaction processing and concurrent access
  - Recovery from crashes

# Transaction processing and concurrent access (moved to 404)

The screenshot shows a Google Docs interface for a document titled "The Stress of Perfection", saved on February 11, 2008 at 4:13 PM by Ryan McCallum. The document is open in the "Edit" mode, with a toolbar showing various editing options like bold, italic, underline, and text color. The document content is a multi-paragraph text about planning a "Snow Daze" event. The text is heavily annotated with comments and revisions, indicating concurrent editing. Comments are shown in yellow boxes with the commenter's name and timestamp. Revisions are shown in pink boxes with the word "Revised" and the name of the person who made the change. The text includes several paragraphs, some of which are highlighted in yellow, and a final paragraph that is partially visible. The interface also shows a "Save" button and a "Share" button.

The Stress of Perfection  
saved on February 11, 2008 4:13 PM by Ryan McCallum

File Edit Insert Revisions Edit HTML Preview Print Email Share Publish

"No, I don't think it would be hard to plan," said Junior Mark Fennig. "Because you have other people to help you think of ideas of what to do for the Snow Daze week, it's not just you by yourself." This is alot of quotes back to back I would try to distribute them better in the story. -Caitlin Werder 2/8/08 1:01 PM

To me, it doesn't really matter if someone doesn't think the week would be hard to plan. If you do a great job of SHOWING the work that's being done, people will understand. Tell the story of the preparation, don't just get people's opinions about the preparation. Be a storyteller. -Ryan McCallum 2/11/08 4:12 PM

Everyone has their own opinion on how hard it would be to plan such a big event like Snow Daze. But most people don't understand that it is very stressful and hard to plan. Show me, don't tell me. -Ryan McCallum 2/11/08 4:13 PM

Between planning dress-up days, the theme, the music, AAA activities, and school activities the volunteers have a lot to think about.

Time plays a huge role in the planning as well. New paragraph-Kelly Kiernan 2/10/08 9:22 PM

"I think it would be hard because you have to get kids together to plan it that are usually in other activities and you have to get the time to plan it," said Spanish Teacher Eivnck. You need the full name, bolded. -Owen Tierney 2/10/08 4:32 PM

As the end of Snow daze approaches, NHS gets ready for the final event, the dance. This is the event that everyone looks forward to and will be the deciding factor of whether the week was a success. That is an opinion, that everyone looks forward to the dance. try not to put in your opinion.-Kelly I agree with Kelly. Some people hate dances. -Ryan McCallum 2/11/08 4:13 PM

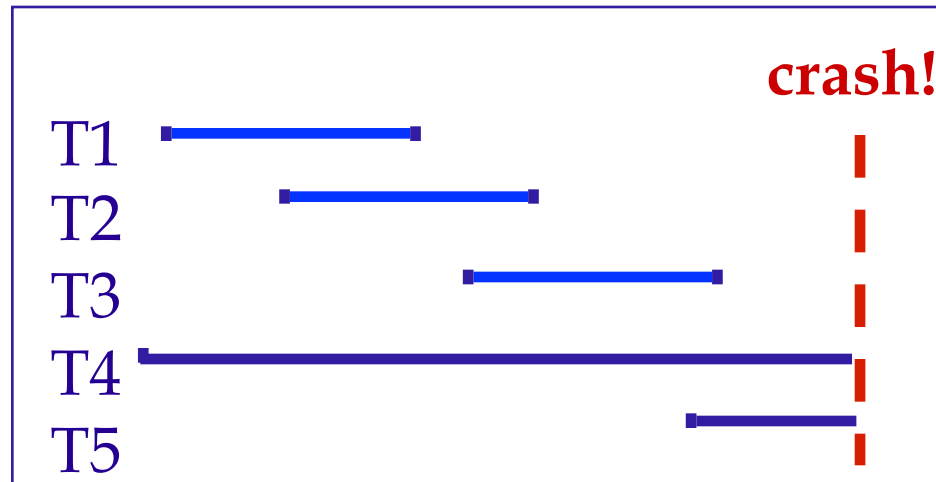
Kiernan 2/10/08 9:24 PM

No matter what the outcome is one thing is certain, a lot of hard work has been put into the week. maybe you could add some more about what snow daze is going to be like. Maybe what the theme and plans are for snow daze. You could also explain who is in charge of planning the snowdaze and what kind of stuff they have to go through. -Caitlin Werder 2/8/08 1:03 PM

First of all, your normal writing should not be Bold. Second off, if you're starting out new as a writer, remember to keep you ENTIRE story in Quote, Transition, Quote, Transition pattern, except for the intro and the ending. And remember to watch for ANY opinion. -Owen Tierney 2/10/08 4:32 PM

- Imagine thousands of people working on the same file!
- You will learn how DBMS handles transactions and concurrent access.

# Recovery from crashes (Moved to 404)



- Assume that T4 is a transaction that is transferring \$1,000,000,000 from an account to another.
- What if DBMS stops running during the process? What can go wrong?

# Why are databases *interesting*?

---



- DBMS encompasses most of CS
  - OS, languages, theory, AI, multimedia, logic,...
- Datasets are increasing in diversity and volume.
  - Digital libraries,
  - Interactive video
  - Human Genome project...
  - Amount doubles every 18 months (since 1990's)
  - XML
  - For more fun; try combining them!
  - *Everyone* has data!

# What data do *you* have?

---

# Here is some data I have:

---

- Books & papers I've read
- Addresses
- Job search data
- Experiments I've run
- Grades
- CDs, DVDs, and books I own
- Financial records
- Photos
- Etc.

# The primary goal of this course:

---

- To help you manage data using a database management system
- You'll learn the most by *doing*. To help, we'll provide:
  - In class exercises – bring clickers and pen and paper. Pre-lecture notes available on the course website by 10 pm the night before the first lecture in that unit.
  - Project – of your own choosing
    - In groups of 4: start thinking of groups now!
  - Tutorials: worked examples, reflective exercises, project work
    - Wednesdays and Fridays
    - Due to space limitations, attend your own tutorial slot. Being in the same tutorial as your project partners will be helpful.
    - No new material, two different locations
    - Tutorial assignments are due at the **beginning of your next tutorial**.
  - Exercises for you to work through on your own
- To encourage doing exercises, at least one question per exam will be isomorphic to a tutorial, homework or exercise problem

# PeerWise

---

- Generating a question requires students to think carefully about the topics of the course and how they relate to the learning outcomes.
- Writing questions focuses attention on the learning outcomes and makes teaching and learning goals more apparent to students.
- The act of creating plausible distracters (multiple-choice alternatives) requires students to consider misconceptions.
- Explanations require students to express their understanding of a topic with as much clarity as possible.
- Answering questions in a drill and practice fashion reinforces learning, and incorporates elements of self-assessment.



# PeerWise

---

- Check the course website for details of how you would be assessed on PeerWise
- **1.5% Authoring and answering questions**
  - There will be a total of 3 calls for authoring and answering questions (check the course website for dates).
  - At least author 1 and answer 15 questions per call
- **1.5% Active Participation**
  - Based on “Reputation Score” and “Answer Score”.
- **Up to 1% bonus mark**
- **Penalties**
  - Students who post question(s) that are well below the quality expected of them will receive a penalty, and can only earn as much as 50% of the grade associated with PeerWise.

# Planned Assessments

---

- Clickers 3% (1 point participation + 1 point right answer)
- Tutorials 9% (start on Friday)
- PeerWise 3% + 1% bonus
- Project 20%
- Midterm 1 15% during lecture on May 25
- Midterm 2 15% during lecture on June 8
- Final 35% To be scheduled
  
- Piazza 1% actively answering questions on Piazza
  - Enrollment code (cpsc304)
  
- You must pass the final to pass the course
- This grading scheme is subject to change

# The Collaboration/Cheating Policy:

---

In general, you can collaborate as much as you want with whomever you want on turned-in work, with three restrictions:

- You must acknowledge everyone you collaborated with
  - You may not take any record away from the collaboration.
  - You must spend at least an hour after collaborating and before working on your submission doing mindless activities
- 
- The exceptions are:
  - On the project you may collaborate with your group.
  - Collaboration with the instructor and TAs is excluded from the above rules.
  - You may *not* collaborate on the midterms, the final, unless explicitly stated.
  - Follow the spirit of the rule and use common sense

# Miscellaneous

---

- Readings are at the bottom of the title sheet for a new set of slides and on the course website.
- You are responsible for the material in the readings even if not explicitly covered in class.
- No handouts – pre-lecture on a unit will be released by 10 pm the night before the first lecture in that unit.
  - Note that pre-lecture slides are subject to change and are only posted to help you prepare for the lecture.

# A word on using your computer during class...

---

- Multitasking on your computer during 304 lectures:
  - A. Helps me to learn 304 materials because I'm not bored and has no effect on other people
  - B. Helps me to learn 304 better, but distracts the people around me
  - C. Makes me learn worse, but has no effect on other people
  - D. Makes me learn worse, and distracts the people around me

# Students' use of laptops in class lowers grades: Canadian study

---

- “all the participants used laptops to take notes during a lecture on meteorology. But half were also asked to complete a series of unrelated tasks on their computers when they felt they could spare some time. Those tasks — which included online searches for information — were meant to mimic what distracted students might do during class.”
- The students who were asked to multitask averaged 11% lower on the exam
- Their neighbors also did significantly worse than the average.

# Where to get more information

---

- CPSC 304 webpage: <http://www.ugrad.cs.ubc.ca/~cs304/2016S1>

- Course staff:

- Instructor: Hassan Khosravi

Office hours:

Wed 2:20-3:00 in ICCS 241

Fri 2:20-3:00 in ICCS 241

By appointment

TAs info on course website

Mostly answering questions on Piazza, but you can email them to set up an appointment.

- Book: Data Management Systems: Ramakrishnan and Gehrke

# Anonymous Feedback

---

- You can use my personal webpage to provide anonymous feedback.
- My main intention is to provide you with the opportunity to express what you like or dislike about the course anonymously.
- Constructive comments and feedback of all kinds are welcome!
- It would be really helpful if you could mention the name of the course that you're taking with me.
- If appropriate, I might post your comment or a paraphrase of it on the discussion boards and respond to it.



# Learning Goals for Chapter 1 Revisited

---

- Define the term *database* and explain the purpose of having a database.
- Explain the high-level objectives of a *database management system* (DBMS), and explain how a DBMS relates to a database. List benefits that result from the usage of a DBMS.



# Our 6-Week Journey

---

We will learn

- that databases are great things
- how to conceptually model them in ER diagrams
- how to logically model them in the relational model
- how to normalize them
- how to query them (many ways)
- about warehouses and data mining and their use cases.
- How to use semi-structured data such as XML