CPSC 304
Introduction to Database Systems

Instructor:
Rachel Pottinger (she/her)
(rap@cs.ubc.ca)

Course coordinator: Gale Chen (they/them)
(cpsc304-admin@cs.ubc.ca)

Beginning Housekeeping

- Introductions:
  - Me
  - Lecture TA(s)
- If you have questions at any time during lecture, you can use the “Live Q&A” feature on Piazza. The TAs will attempt to answer questions and/or alert me to the questions so that I can address them.
- Most information is stored on Canvas, including a preliminary version of these slides.
- More housekeeping to follow later once people have had a chance to find class.

Canvas

- Holds all of our course content
- Panapto live streams AND recordings for both sections should* be available there
- Your grades will be stored here
- Contains links to the other websites we use in this class (e.g., Piazza)
- Important: the tentative class schedule

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 are databases *interesting*?

- They're useful for jobs
- 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)
  - For more fun; try combining them!
  - We put the data in big data
  - *Everyone* has data!

Why use a database?

Suppose you are building a system to replace Workday Student. 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?

To replace Workday Student 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?
- How do you integrate data from different secondary school systems?

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)

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
- Some common DBMSs are Microsoft SQL Server, MySQL, MariaDB, and Oracle
- In this class, we will use Oracle as our DBMS

How does the DBMS fit in?

- The DBMS usually lives on a server (e.g., students.cs.ubc.ca)
- The client computer (e.g., user computer) will interact with an application.
- That application may be on the server or the client
- The application queries the DBMS, and answers are routed back to the application and the client

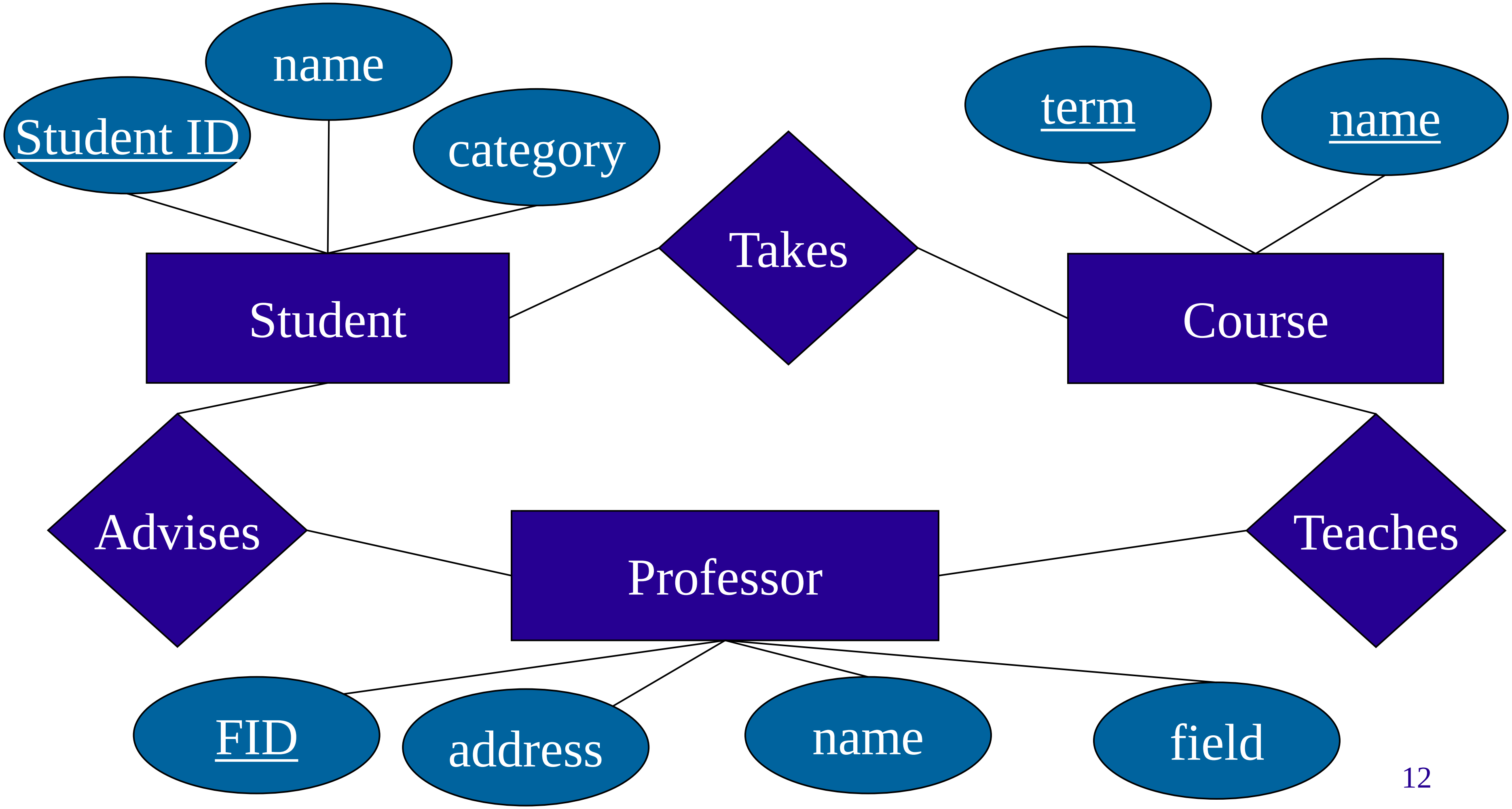
Great! What's left for us to do?

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

1. Conceptually model the concepts
2. *Logically* model concepts in a database
3. Make our design efficient
4. Decide which users will ask which queries
5. Data warehousing, OLAP (**O**n**l**ine **A**nalytical **P**rocessing)
6. Data mining
7. Create an application and enjoy!

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:

Student	Course	Term
Ying	CPSC 304	Winter 1, 2024
Andrew	CPSC 221	Summer 2, 2006

Separates logical and physical views of the data.

3. 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”

4. 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, 2017.

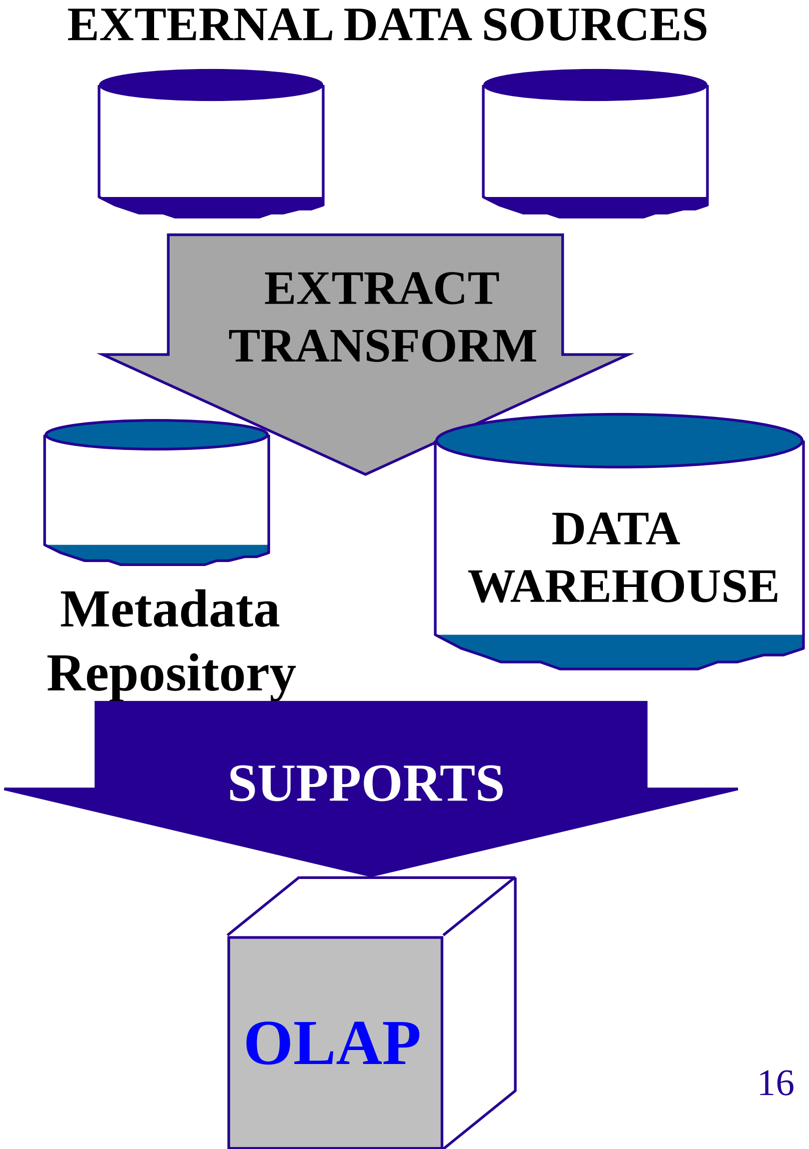
```
select  E.name
from Enroll E
where E.course=“CPSC 304” and
          E.term=“Winter Term 2, 2017”
```

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

5. Data warehousing, OLAP

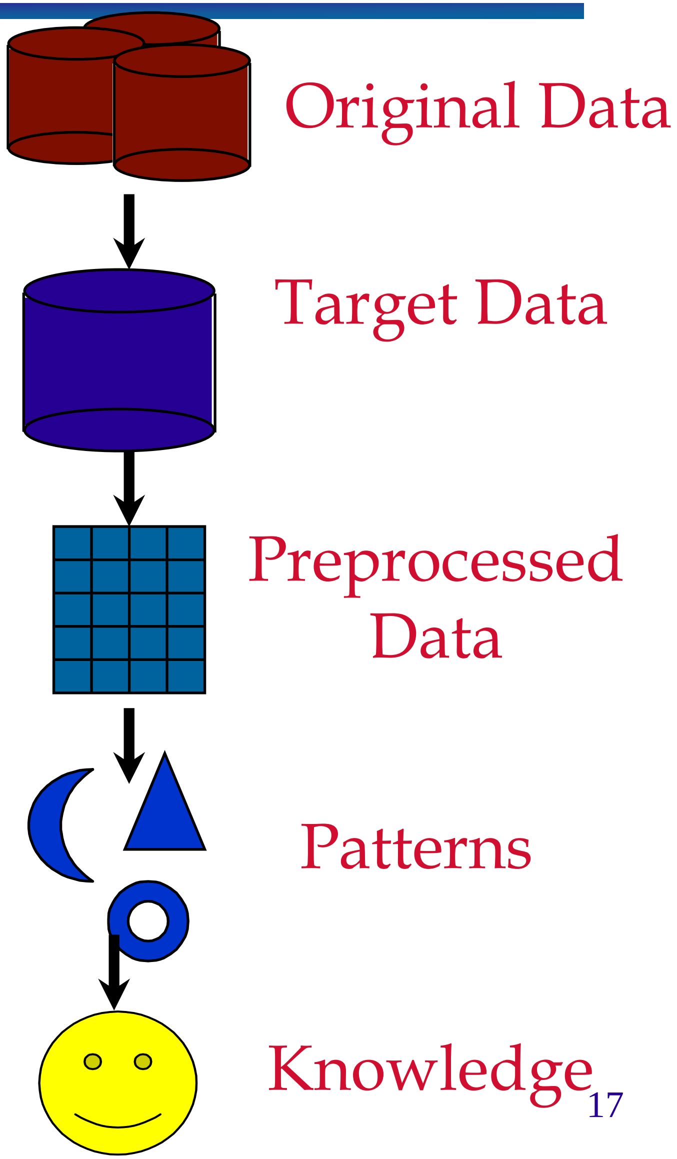
- Increasingly, organizations are analyzing current and historical data to identify useful patterns and support business strategies.

- The emphasis is on analysis of complex queries on very large datasets created by integrating data from across all parts of an enterprise.



6. 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



7. Create an application

- Use a programming language to interface with database in a group of 3
- We’ll support:
  - Departmental servers
  - Programming an application in one of:
    - The web & PHP
    - The web & JavaScript
- You can use other database/language/server combos (e.g., Java/JDBC) but the TAs won’t be able to help
- Project description document is on Canvas. Be sure to read it to get a feel for what we want.

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

COVID-19: Why do I care?

- A lot of people seem to think (or wish) that COVID-19 is over.
- It isn't.
- Additionally, people with severe medical conditions can have very bad outcomes and also not respond to vaccines or be able to take vaccines.
- Please protect yourself and those around you who may have a medical condition or live with someone who does by taking precautions, including getting vaccinated and wearing a mask indoors.

My reason being so cautious

I note that I can keep my distance from you.
Others in class will have similar concerns
but not the same advantages.

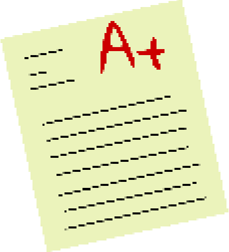
COVID-19 safety in this class

- If you have not yet had a chance to get vaccinated against Covid-19, vaccines are available to you, free: http://www.vch.ca/covid-19/covid-19-vaccine#clinics
- There will be extra space for exams to allow for adequate distancing
- If you have doubts about whether to attend class or not, please, please, please stay home.
- Consider wearing a mask, especially if you’re concerned that you may be sick.

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 pen and paper and consider downloading iClicker Cloud. Pre-lecture notes available on the course website before the first lecture in that unit.
  - Project
    - In groups of 3: start thinking of groups now!
  - Tutorials: reflective exercises
    - Being in the same tutorial as your project partners will be helpful (but not mandatory).
  - Exercises for you to work through on your own

Planned Assessments



Syllabus quiz		1%
“In-Class” Exercises – attendance not required		4%
Clickers – attendance required		4%
Project		27%
Milestone 0: Form a Team	0%	
Milestone 1: Proposal and ER Diagrams	4%	
Milestone 2: Logical Design	7%	
Milestone 3: Project Check-In	3%	
Milestone 4: Implementation	2%	
Milestone 5: Demo	8%	
Milestone 6: Individual and Peer Evals	3%	
Midterm– October 22		25%
Final– Date TBD		39%

Minimum Passing Criteria

- Pass a combination of the midterm the final exam (as weighted on previous slide) **and**
- Achieve an overall grade of 50% or better.

- If you fail the final exam/midterm requirement, you will receive as your course grade the lower of the normally computed grade and 45%.

“In-Class” Exercises

- To help you practice the concepts, we will have “in-class” exercises that you will complete
- These exercises will be due one day after the last class to start the exercise at 10PM
    - If you always turn it in by 10PM the day after it’s been started in your class, you’ll always be on time.
- Graded based on effort and completeness (does not necessarily have to be correct). We will take the best 80% of days to allow for the odd missed class.
- We will release the solution to the exercise after the deadline

Clickers

- To make class more interactive and more tuned to how students are doing, we will use iClicker Cloud (it is free for UBC students at least this term): https://lthub.ubc.ca/guides/iclicker-cloud-student-guide/
- Clickers will be graded for participation, but we only require that you submit 70% of the questions for 100% of the marks.
  - You must be in class to be sure that clickers are working right: The streaming can be delayed by up to two minutes, so clicking remotely may not work
  - We are not responsible if clickers do not work
  - We do not usually excuse you from clicker questions, but feel free to tell us if there are extenuating circumstances.

Exams

- There will be one midterm held at 6pm on October 22
- The final exam covers more topics than the midterm
- Do not book travel plans until the final exam date has been announced (usually in mid to late October)
- We will use the “10% rule” for the midterm: If your grade on the final is X, and X is more than 10% higher than your midterm grade, we will count your midterm grade as X-10.

Tutorials

- Attendance at tutorials is completely optional
- Tutorial material can be found on Canvas
- It’s a great way to practice class material
- Tutorials will often start with a small lecture by your TAs. Depending on the tutorial, it can then become a mini-office hour or more of a lab

Slides

- A set of slides for each topic will be posted on Canvas prior to that topic's introduction in class
  - These slides will NOT have the clicker questions in it
  - We'll work on the clicker questions together in class
- After each class, I'll post the slides with the clicker questions in it so you can refer back to them if need be

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

- Most of the course material will be posted on Canvas but we will be using Piazza for our discussion board
- Readings are at the bottom of the title slide for a new set of slides.
- You are responsible for the material in the readings even if not explicitly covered in class.
  - Note that pre-lecture slides are subject to change and are only posted to help you prepare for the lecture.
  - Slides will be posted on Canvas
- **Piazza is the fastest way to get help!**
  - **Search for your question prior to posting**

Communication

- If you have questions of a general nature, please post on Piazza.
- If you have specific concerns, then please email the course coordinator (cpsc304-admin@cs.ub.ca) or me or talk to me after class!
- Office hours TBA
- **PLEASE don't message us using Canvas.**

  The Canvas messaging system is not a communication avenue we use and messages sent here may not be seen/responded to until much later.

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.

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.

TODOs

- Run get-acct (get your CS account set up if you haven't already done so)
- Sign up for the Piazza class via the Piazza link in the Canvas sidebar
  - Make sure your account is afflilated with a UBC email address (e.g., ugrad.cs.ubc.ca, students.cs.ubc.ca etc.)
- Do the syllabus quiz on Canvas
- Register for iClicker (make sure you are registering in the correct section)
  - Section 101: https://join.iclicker.com/IUPB
  - Section 103: https://join.iclicker.com/MSUN

Ask Me Things!

- For example:
  - What are your concerns?
  - Is anything unclear?
  - Is there something you want to talk about?
- Other topics welcome!
  Disclaimer: **Registration is controlled solely by the CS department and I have no information about it whatsoever**. Waitlists for CPSC 304 have traditionally cleared in the past, but there’s no way for us to make a prediction.
  - https://www.cs.ubc.ca/students/undergrad/courses/waitlists