UBC Department of Computer Science
CPSC 404: Advanced Relational Databases
(aka Advanced Database Systems)
Course Outline for September-December 2017
Last Update: September 7, 2017 @ 13:15

1. Course Description

This course is a continuation of CPSC 304. Whereas CPSC 304 focuses on *using* an RDBMS, CPSC 404 focuses on the *internals* and *performance issues* of an RDBMS. Topics include disks and the storage hierarchy, the I/O cost model, buffer pool management, disk scheduling, record layouts, metadata and system catalogs, tree-structured indexes, hash-structured indexes, query evaluation, query optimization, transaction processing, concurrency, and crash recovery.

2. Prerequisites

CPSC 304 and one of CPSC 213 or CPSC 261. Students without appropriate prerequisites will get a "missing prerequisite" letter near the start of the course. Watch for it in your e-mail, if you fall into this category. Follow the instructions in the letter to resolve it. For example, you may have taken a course equivalent to CPSC 213, 261, or 304 at another institution, and the department needs to become aware of this. If you do not have the prerequisites, you will be dropped from the course.

3. Instructor and TAs

Please be advised that because this is a large class, it is very difficult to respond to everyone's e-mail request. Please limit such e-mail to items of a personal or confidential nature (e.g., illness), and use the course bulletin board to post course-related questions … or ask during TA/instructor office hours (see below).

Instructor:	Dr. Ed Knorr	= knorr@cs.ubc.ca
TAs:	Hlib Naumenko	= naumenko@cs.ubc.ca
	Jan Pilzer	= pilzer@cs.ubc.ca
	Xing Zeng	= xingzeng@cs.ubc.ca

4. Lectures, Holidays, and Dates of Interest including Midterms

- **Lectures** take place on Tuesdays and Thursdays from 14:00-15:20 in DMP 110
- There are no classes on Tuesday, September 5, 2017 because it is **Imagine Day** at UBC, and almost all undergrad classes are cancelled. The first CPSC 304 lecture will be on Thursday, September 7.
- The **add/drop deadline** is Tuesday, September 19. You can drop the course on or before this date without getting a 'W' (for "withdrawal") on your official transcript.
- The **final withdrawal deadline** (i.e., the 'W' drop deadline) is Friday, October 13. Beyond this date, you will get a numeric grade for the course.
- The **holidays** for this term—when UBC is closed—are:

- Monday, October 9—Thanksgiving Day (no effect on our lectures)
- Monday, November 13—Remembrance Day holiday in lieu of Saturday, November 11 (no effect on our lectures)

- The **last day of classes** for UBC's Term 1 is Friday, December 1.
- The **final exam period** runs from Tuesday, December 5 to Wednesday, December 20.
- The midterms are tentatively scheduled for these dates:

Midterm #	**Date** (about one-third of our class will go to another room—a different third each time—details later)
1	Thursday, September 28 (6 lectures of coverage)
2	Tuesday, October 24 (6 lectures of coverage)
3	Keep both of these dates open since we're having trouble finding rooms on campus: Tuesday, November 14 (5 lectures of coverage, room available) and Thursday, November 16 (6 lectures of coverage (better); but so far, no room)

5. Labs/Tutorials?

There are no scheduled labs or tutorials for this course. The TAs and I will be holding office hours, in addition to monitoring the discussion board on Piazza. For SQL Server, you can use either a PC in the lab or your own machine to access the SQL Server production instance hosted by our department. You can work in pairs, and you will demonstrate your work to the TAs.

6. Office Hours for Instructor and TAs

All of the TAs will hold office hours. I will update Connect with the office hours, and I'll frequently show them at the start of class. Our office hours start during Week 2 (i.e., during the week of September 11). The department also keeps an online list of the TAs' office hours for the Demco Learning Centre (ICCS X150) and other TA-accessible rooms at: `https://my.cs.ubc.ca/students/ta-hours`.

Day	**Time**	**Who**	**Location**
Mondays	14:00-15:30	Xing	TBA
Tuesdays	TBA	TBA	TBA
Wednesdays	TBA	TBA	TBA
Thursdays	TBA	TBA	TBA
Fridays	14:00-15:30	Xing	TBA

7. Textbook (Required)

Raghu Ramakrishnan and Johannes Gehrke. *Database Management Systems*, 3rd Edition, McGraw-Hill, 2003.

This is the same book we use for CPSC 304. We will be making good use of it for most of the topics in this course. In particular, we will study substantial portions of Chapters 8 (useful background information), 9 (some), 10-14, and 16-18. You will probably find the book to be of value in your computing career; so, if you don't already have it, it's worth buying and keeping. You can get it at the UBC Bookstore, the Discount Textbooks store at University Village (upstairs, near McDonalds), and various online bookstores. It's pricy (probably over $200 at the

UBC Bookstore plus 5% GST, and maybe $10-15 cheaper at Discount Textbooks); therefore, you may want to get a used copy either from the above sources or via private sales at UBC (e.g., check the bulletin boards on some of the walls in the CPSC department for student-posted ads). The UBC Bookstore arranges its books by the author's last name; therefore, remember the name "Ramakrishnan" for the textbook, and the name "Hayden/McNeil" for the (optional) carbonless copy notebook.

Copies of the textbook are on reserve for CPSC 304 and 404 for short-term loan at the Ike Barber Library Reserve Room (1 copy for 2-hour loan, available 7 days a week including evenings) and at the CS Reading Room (i.e., the CPSC library located across the atrium from the main office on the 2nd floor of the ICCS building—3 copies, daytime hours, and all students can use the Reading Room). The respective links are here:

- Ike Barber Reserves: `http://search.library.ubc.ca` (enter keywords "`ramakrishnan gehrke`" to narrow the search)
- CS Reading Room: `https://www.cs.ubc.ca/our-department/facilities/reading-room/course-reserves`

8. Pre-Class Exercises (you supply the paper) and In-Class Exercises (we supply the paper)

We will do a bunch of pre-class and in-class exercises. Both of these will be handed in—for participation points based on effort and reasonable progress. You won't get your papers back. If you would like to keep an electronic copy for reference purposes, you can simply take a photo of your paper before handing it in. If you would like to have a *paper* copy of it (but you don't have to), you should get a carbonless copy lab notebook (e.g., top page is white; the bottom page is either yellow, pink, or white) to use for the pre-class exercises. Turn in the bottom copy, unless it's very unreadable, and only then turn in the top copy. Be sure to clearly print your student ID number on your work because the TAs will be entering your points by student ID number. If you have a carbonless copy notebook from a previous course at UBC, then that same book will be fine. Otherwise, if you want a paper copy of your work, purchase a copy of the very popular "Hayden/McNeil" lab notebook (used by several Chemistry courses) from the UBC Bookstore (about $20 plus 5% GST) or from Discount Textbooks.

Pre-Class Exercises: You may wish to wait for a week before deciding whether or not you want to buy the notebook; many students will be OK without it. The questions and solutions are online.

In-Class Exercises: The in-class exercises can be done in pairs: you won't get your work back; but we'll post all the questions and the sample solutions, usually within 24 hours.

You will hand in a copy of your pre-class and in-class work at the end of class, by the door, as you leave DMP 110. TAs will be awarding participation points for the exercises on a 0-2 point scale for effort and reasonable progress. Points for in-class exercises can only be earned during class. Points for pre-class exercises can be awarded for a missed class (e.g., if you're sick) during the following class. In the latter case, *clearly* indicate that it is the pre-reading exercise for the *prior* class. You must have your student number on each page that you hand in. A few times per term, I will upload the pre-class, in-class, and clicker grades to Connect. I'll announce this on Piazza. **We'll take the best 90% of these pre-class and in-class exercises to allow for the odd missed class.**

9. iClickers (Required)

This course will use iClickers. If you don't have an iClicker handheld device from a previous UBC course, you should get one from the UBC Bookstore, the Discount Textbook store, or from a reliable, second-hand source at UBC. You'll need to register the clicker ID on Connect sometime during the term to map your clicker ID to your student number (to properly credit your marks), unless you've already done this for another course on Connect. A few times per term, I will upload the clicker grades to Connect—and announce this on Piazza. If the sticker on the back of your clicker has rubbed off and you don't know your clicker number, come and see me during office hours and we can figure it out by setting up a mobile receiver in my office. Alternatively, visit CTLT in the basement of the Ike Barber building, and they can do the same for you.

10. Additional Reference Material (Optional)

If you want additional reference material, then any book on database systems that has been published in the past 10-20 years should be fine, providing it deals with relational database systems and addresses most of the topics described on this outline. One such book—I will use some material from it in class—is the following (there is a copy on reserve in the CS Reading Room):

> Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom. *Database Systems: The Complete Book*, Prentice-Hall, 2009 (but 2002's edition is fine).
>
> Or equivalently, for most of our purposes (the relevant subset of the above book is): Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom. *Database System Implementation*, Prentice-Hall, 2000.

For those of you planning to work in the database area, a great book on general database administration tasks is:

> Mullins, Craig S. *Database Administration: The Complete Guide to DBA Practices and Procedures*. Second Edition, Addison-Wesley, 2013.

Craig has had a long career of outstanding work in database systems and administration. I also recommend the following thick book (especially if you'll be using IBM's DB2 in your career):

> Mullins, Craig S. *DB2 Developer's Guide*. Sixth Edition, IBM Press, Pearson, 2012.

11. Use of Connect (Blackboard) Course Management System

I will bring paper copies of the lecture slides to class for the first lecture; but, after that, you should log on to Connect and either print off or download the lecture slides.

Lecture notes, pre-class readings, pre-class exercises, and exam practice questions (with full solutions) are available on Connect all term. Please download the current lecture slides before coming to class. If any changes are required, I will try to make the updates available by 6 PM the evening before the lecture. Unless your memory is great, you'll probably need to take your own

supplementary notes in class. Please be advised that you are responsible for all material presented in the lectures and on Connect (including the corresponding textbook readings, pre-class examples, and practice questions).

UBC's ITServices group hosts Connect. If you don't already have a Campus-Wide Login (CWL) account, you should visit `www.cwl.ubc.ca` to get one. All students registered in CPSC 404, and who have a CWL ID, will automatically have their CWL ID linked to the CPSC 404 Connect pages. If you are taking other courses that use Connect, then your CWL ID will also be linked to those courses.

12. Piazza's Discussion Board

Piazza will host our course bulletin board (also called a "discussion board"), and it is ***required reading*** for this course. You should read it at least once per day. In the past, we've used Connect's discussion board, but an overwhelming majority of students requested a move to Piazza. UBC Connect, however, will be used for course grades, sample exams, sample solutions, pre-class readings, pre-class exercises, tutorials, and almost everything else.

Piazza can provide quick and efficient help from classmates, TAs, and myself. Rather than e-mailing questions to the teaching staff, I encourage you to post your questions on Piazza. If you encounter any problems, or if you have feedback for the Piazza developers, e-mail them at `team@piazza.com`.

Piazza registration instructions are found on Connect (see the CPSC 404 home page for detailed instructions). If you are registering for Piazza with your CPSC department e-mail address of the form `xxxx@ugrad.cs.ubc.ca` or `yyyy@cs.ubc.ca`, then you don't have to do anything else. However, if you *not* using a CPSC department e-mail address or if you are using a pseudonym (i.e., a fake name, but please *don't* pick the name of a celebrity, sports star, other student, etc.), then you need to fill out the short, Piazza pseudonym survey (it takes about one minute to fill out), which is hosted at UBC and is found at:

`https://survey.ubc.ca/s/cpsc404_2017WT1_piazza`

You'll find our class's Piazza discussion board at:
`https://piazza.com/ubc.ca/winterterm12017/cpsc404`

The access code is found on Connect, and will also be mentioned on Day 1 in class.

13. Lecture Content vs. Pre-Reading

To make classroom time more productive (e.g., to give us more time in solving problems), lecture material that is found in the pre-readings will probably not be repeated in class. If you do not understand something, please ask during TA or instructor office hours, or on the bulletin board, or just ask your friends. If you do not do the pre-readings or pre-class exercises, then you will probably not be very productive in class. The section numbers of the pre-reading material in the textbook, and a list of pre-class exercises, will be indicated on Connect. I'll try to post these at least a full day before the next lecture. Lecture slides will be there, too. The pre-class material will be relatively simple, and we will build on it during class. (Many mobile devices, including

Kindle, don't use page numbers for downloaded books because of different screen/page sizes; that's why we're using section numbers.)

14. Database Systems that We Will Study

CPSC 304 used Oracle and Java/JDBC (and optionally some of you may have used MySQL and either Java/JDBC or PHP). However, in CPSC 404 this term, we will use: (a) IBM's DB2 for case studies in class, and (b) Microsoft's SQL Server for hands-on activities and exercises (including a demo for a TA). Thus, between CPSC 304 and 404, you will be exposed to several of the major commercial RDBMSs.

15. Problems and Technical Support

Problems with the CPSC 404 course contents (e.g., lecture, textbook, exercises, Connect content) can be posted on the bulletin board, *but please check to make sure that your question hasn't already been asked or answered.* The TAs and instructor will be monitoring and responding to questions frequently; but, other students are also encouraged to respond to postings. Problems with Connect itself (i.e., other than with the content of the CPSC 404 Web pages) should be directed to `help AT itservices.ubc.ca`. Problems with undergrad accounts and servers should be directed to `help AT ugrad.cs.ubc.ca`.

16. Worked Examples, Exercises, Assignments, and Midterms

Because we have regular pre-class readings, pre-class exercises, and in-class exercises, there are no formal homework assignments. However, there will be a lab exercise involving Microsoft SQL Server and there will be marks for completing it and giving the TAs a demo. You will get points for participating, answering a bunch of short written questions, performing an in-lab demo, and answering a few oral questions from the TAs.

There will be 3 midterms—based on lecture content and practice exercises both inside and outside of class. Online, there are lots of sample questions from old assignments and exams, with lots of detailed solutions. By practicing with a wide variety of worked examples, you will be prepared to solve a wide range of related problems. If you want even more practice questions, check the textbook; a solutions guide is available on Connect for the odd-numbered questions.

Note: On Connect, there are also questions and answers for some different topics than are covered this term. If we haven't covered the topic in class, then don't worry: you can safely ignore such questions.

The midterms will be held during class time, on the tentative dates listed in the table in Section 4 above. All exams will be closed book with no notes; however, a simple, non-programmable, non-communicating calculator will be permitted. The midterms will not be cumulative. So, for example, Midterm #2 will not include lecture material that was covered on Midterm #1 *unless* the new material explicitly re-uses older material in an important way. For example, Sort-Merge Join calculations use important principles from External Mergesort; so, a lot of the External Mergesort principles re-appear in a new unit, and are therefore fair game. Also, indexing principles and complexity are discussed throughout the course; so, those would also be fair game.

The final exam will cover the whole course, with a goal of providing reasonably-balanced coverage of topics from the whole term. There will be a slight emphasis on material between Midterm #3 and the final exam, since that material will not have contributed to any midterm.

If you cannot write an exam (e.g., due to illness), you should obtain suitable documentation (e.g., doctor's note) and inform the instructor ASAP—*not* after you've written the exam or gotten your mark back. Students with a valid excuse and documentation will have their final exam grade applied to the missing midterm. Note: Depending on the number of students who miss an exam, and at the instructor's discretion, there may be a make-up exam scheduled during a subsequent lecture period (but obviously in a different room than our lecture room). In this case, you'll need to get any missed lecture notes from a fellow student.

The final exam will be written on a date in December to be determined by the University. If you cannot write the final exam, then you must provide documentation (e.g., a doctor's note) to your home Faculty office. For example, if you are in Science, then it will be the Faculty of Science; if you are in Engineering, then it will be the Faculty of Applied Science; and similarly for Arts, Commerce, etc. You will write an exam at a future date determined by the University: usually this is during the deferred exam period in the summer (typically late July or early August), or during Term 2's final exam period in April 2018. Important: *Please note that your Faculty may not permit you to write a deferred final exam, even if you have a doctor's note, if your term work is incomplete* (e.g., missing midterms, lots of missing pre-class and in-class exercises, absent for a lot of classes).

17. Tentative Grading Scheme

To pass the course, you must obtain a 50% overall mark, and pass the final exam. The final exam will cover the whole course. In keeping with department policies, students who do not pass the final exam (or don't come very close to passing the final exam) will receive the minimum of 45% and their overall failing grade.

Final grades will be tentatively calculated as follows:

- 25% for the participation component and term work:
    - 6% for the clickers (half for participation, half for the right answers)—but we'll only count your best 90% of these marks (to allow for a forgotten clicker, dead battery, missed class, doctor or dentist appointment, sickness, funeral, "I missed the bus", "I slept in", "There was an accident on Broadway", etc.) If you are sick for 3 or more *consecutive* classes, let me know via e-mail, and I can make an adjustment.
    - 12% for the in-class and pre-class exercises (participation will be based on a 2-point scale: 2 for full effort, 1 for partial effort, and 0 for little to no effort)—but, just like for the clickers, we'll only count your best 90% of the marks. If you are sick for 3 or more *consecutive* classes, let me know via e-mail, and I can make an adjustment.
    - 7% for the SQL Server activities/exercises.
- 37.5% for the 3 midterms (one-third for each)
- 37.5% for the final exam

Grading errors on the midterms must be reported within 4 days of when the papers are initially returned in class (the TAs and I are more likely to remember your paper). Please forward grade disputes to me, and not to the TAs—although the TAs will probably be able to answer questions about the markers' comments. We'll post the solutions within a day of returning the midterms in class.

18. Course Topics

Introduction	Course overview, DB-related tasks, self-managing DBMSs, definitions of DB2 database objects
Storage (Chapter 9)	Storage/memory hierarchy, disks, access times, I/O cost model (number of page I/Os), page layouts, metadata (system catalogs, SQL), buffer management
Tree-Structured Indexes (Chapter 10 and some of Chapter 8)	Introduction to indexing; clustering; B+-trees for searching, inserting, updating, and deleting; I/O costs; complexity analysis
Hash-Structured Indexes (Chapter 11)	Extendible hashing, linear hashing; analysis
External Sorting (Chapter 13)	External sorting strategies and analysis, including cylindrification and large block sizes
Query Evaluation (Chapters 12 & 14)	Relational algebra, SQL, join algorithms, buffer sizes, aggregation (group-by operations), access paths, index usage
Query Optimization (more from Chapters 12 & 14)	Strategies for efficiently executing a database query; the role of the optimizer in DB performance; pipelining
Transaction Management and Concurrency Control (Chapters 16 & 17)	ACID Properties, Concurrency, Serializability, Isolation Levels, Deadlock, Deadlock Detection and Prevention, Concurrency Control with and without Locking (e.g., 2PL, Strict 2PL, Multiple-Granularity Locks, Timestamps, Optimistic)
Logging and Crash Recovery (Chapter 18)	Backup and Recovery (Application- and System-Level), ARIES Algorithm, Logging, Checkpoints

19. Learning Goals

The course-level and topic-level learning goals are provided on Connect. The lecture slides will also contain the topic-level learning goals. These are useful for self-evaluation and when studying for exams. They also give you (and future employers) a more detailed view of the course's contents than a typical calendar entry or a list of topics would.

20. Carl Wieman Science Education Initiative (CWSEI)

UBC's CWSEI (http://www.cwsei.ubc.ca) is aimed at improving the quality of teaching and learning in science courses at UBC and (through disseminated research) around the world. In our course, we will try to adopt some best practices from evidence-based research in education (e.g., well-defined learning goals, deliberate practice, pre-reading, pre-class exercises, clickers, in-class exercises, surveys).