601.315/415/615 - DATABASES
Fall 2020 Syllabus

Instructor: Prof. David Yarowsky
Hackerman 324G
410-516-5372
yarowsky@jhu.edu

Head CA: Ankur Kejriwal
akejriw2@jhu.edu

Meeting Time: Tu,Th: 3:00-4:15 PM (Baltimore Time)
and other interactive sessions to accommodate remote time zones

Classroom: Online (details distributed to registered students and by request to instructor)

Office Hours: TBA and by appointment. zoom details will be provided
CAs - TBA and by appointment. zoom details will be provided.

Required Textbook:

- A. Silberschatz, H. Korth and S. Sudarshan, *Database System Concepts*, 7th Edition, McGraw Hill, 2019, ISBN: 978-0-07-802215-9 or 6th Edition, McGraw Hill, 2010, ISBN: 978-0-07-352332-3.

601.315/415/615 will explore both formal and practical issues in databases. Hands-on database design and implementation using the MySQL DBMS will be an important component of the course.

Course Requirements: Final grades will be based on the following (subject to change):

Class Participation:	5%
Homeworks (4):	28%
Midterm:	20%
Written Final Exam:	10%
Brief Oral Exam (S/U):	7%
Final Project:	30%

Lateness Policy:

One homework assignment may be handed in up to 5 days late without penalty. No other late homeworks will be accepted. Final projects handed in late will receive a penalty of 10% for every day late.

601.315 vs. 601.415/615:

601.315/415/615 will be share common lectures. They will differ primarily in terms of assignments and grading. Homeworks in 601.415/615 will include 1 or more additional problems and the final project will include additional component(s) not required for 601.315. Exams will differ somewhat and will be graded on a different scale. Nevertheless, 601.415/615 should be manageable by advanced undergraduates and upperclass students are encouraged to enroll.

601.315/415/615 vs. 601.316/416/616:

Databases (315/415/615, Fall) and *Database Systems* (316/416/616, Spring) are complementary courses and make a natural course sequence (see below). 315/415/615 focuses on how to design and use a database; formal database models, theory and foundations; database programming languages, especially SQL and PL/SQL; object-oriented and XML-based data models and future directions (including data mining and natural language interfaces). The final project will be application-focused (e.g. how to design an implement a database for a novel task) including practical execution of the concepts studied in the class. In contrast, 316/416/616 will focus on database internals and systems, including query and join processing, indexing, file organization, estimation and optimization, as well as database architectures, streaming and partitioning. The course project(s) will focus on database system internals and their development.

Can I take 316/416/616 as a stand-alone course without 315 or 415/615?

Yes, 316/416/616 does not have 315 or 415/615 as a formal prerequisite. You should have some database experience before taking 316/416/616, however, either through prior employment or via a prior course.

Graduate students who have prior database employment experience or have taken a prior course in database systems are normally expected to begin directly with 416/616.

Anyone with a research focus in the databases area should certainly begin directly with 416/616.

Can I take 315/316 or 415-615/416/616 as a 2-course sequence?

Yes. There will be modest overlap of material (10%) but taught via different perspectives and emphasis, and will serve as a good refresher.

If you have not taken a prior course in databases and are interested in both the theory/applications and systems sides of the field, then this sequence makes a lot of sense and is encouraged.

The instructors will work to make this a natural 2-course sequence.

Can I take 315/416/616 as a sequence?

Yes, 416/616 does not require 415/615 as a prerequisite, but you should have done well in 315 and be prepared to do some background catchup to meet the expectations of the 416/616 instructor.

Can I take 415/316 as a sequence?

Yes, if you are an undergraduate and would like to continue focusing on database systems and database systems internals but a less difficult level, then this sequence could make sense.

Final Exam and Brief Oral Exam:

To better accommodate the unique Fall 2020 virtual class format and COVID-related stresses, the final exam will be substantially shorter than the previous standard 3-hour exams, and worth only 10% of the final grade rather than the previous 25%. The exam will be cumulative, but focus on the post-midterm components of the class.

In addition, unique for this year, all students will take a brief (approximately 15 minute) 1-on-1 oral exam scheduled at a convenient time for you near the time of the final exam, so single preparation can be used for both. A key goal of the exam will be to confirm your mastery overall course material via an independent medium, and the majority of the content will cover questions that you've already answered on previous homeworks and exams, asking you to explain your reasoning etc. The oral exam will be graded S/U (Satisfactory/Unsatisfactory) and students passing the exam will receive full credit for its 7% of the grade. Students not passing the exam will have a 2nd opportunity for a 1-on-1 follow-on oral exam with the instructor.

Final Projects:

Students will be able to select final projects of interest to them from a fairly diverse set of options. Details will be provided in class. Students may work in teams of 1 or 2 people. A project proposal will be due in early November, including a detailed system specification and design. The final project submission, including a full database implementation in MySQL, will be due shortly after the end of classes in December. For most projects, students will be required to populate and test their implemented database design with substantial quantities of *real world data* extracted from the web or other online sources.

Computer Science Academic Integrity Code:

Academic honesty is required in all work you submit to be graded. You must solve all homework and programming assignments entirely on your own, unless group work is specified in writing. This means you must not show your program code, problem solutions, or work to other students. However, you may discuss assignment specifications with others in the class to be sure you understand what is required by the assignment. If you use fragments of source code from sources other than your text (such as on-line resources), you must put a reference to that effect in your homework submission. Falsifying program output or results is prohibited. Please see your professor if there are any questions about what is permissible. Students who cheat will suffer a serious course grade penalty in addition to being reported to university officials. You must abide by JHU's Ethics Code, available at http://jhunix.hcf.jhu.edu/~ethicsbd.

Solutions to Previous Exams and Homeworks:

A copy of the previous year's midterm (and one other midterm) and their solutions will be explicitly distributed to students for practice and guidance regarding expectations, and students are encouraged to study using them. Likewise, Homework 4 is composed of questions given on previous final exams, and is intended as preparation for the final exam, with sample solutions given **after** HW4 is submitted but before the final exam.

With the above exceptions, students are explicitly forbidden from looking at or using other 601.315/415/615 exams, homeworks and/or sample solutions.

Preliminary Class Schedule (subject to change):

Date	Topic	KS 7e	KS 6e	KS 5e	KS 4e
Tu. 9/1	Introduction	–	–	–	–
Th. 9/3	Overview of databases and data modeling	1	1	1	1
Tu. 9/8	Entity-Relationship data model	6	7	6	2
Th. 9/10	Database design principles	6	2	6	2
Tu. 9/15	Relational data model	2	3	2	3
Th. 9/17	Relational algebra	2	6	2	3
Tu. 9/22	Relational algebra and relational calculus	2	6	2,5	3
Th. 9/24	SQL	3	3	3	4
Tu. 9/29	SQL (continued)	4	4	4	4
Th. 10/1	Advanced SQL	5	5	4	4
Tu. 10/6	QBE (Query by Example), Views	27	C1	5	5
Th. 10/8	Relational database design, integrity constraints	7	8	7	7
Tu. 10/13	Relational database design, normalization	7	8	7	7
Th. 10/15	Query processing and optimization	15-16	12-13	14	14
Tu. 10/20	Embedded SQL; PL-SQL/stored procedures	hnd	5	hnd	hnd
Th. 10/22	Application design and development	9	9	8	
Tu. 10/27	Transactions and database recovery	17-19	14,16	17-19	17-19
Th. 10/29	MIDTERM	–	–	–	–
Tu. 11/3	TBA	hnd	hnd	hnd	hnd
Th. 11/5	Distributed databases	20-22	19	22	19
Tu. 11/10	Database security	9	9	8	6
Th. 11/12	Object-oriented databases	29	22	9	8,9
Tu. 11/17	Data warehousing, data mining, multimedia databases	10-11	25-26	18,24	22,23
Th. 11/19	nosql, datalog and XML data model	hnd,30	hnd	hnd	hnd
Tu. 11/24	THANKSGIVING BREAK				
Th. 11/26	THANKSGIVING BREAK				
Tu. 12/1	Natural language interfaces, Natural language databases	hnd	hnd	hnd	hnd
Th. 12/3	WWW-based technologies/interfaces; future directions	26	21	10	22
Tu. 12/8	Future directions	hnd	hnd	hnd	hnd

KS = Korth and Silberschatz

Details regarding final exam scheduling and oral exam scheduling will be distributed later in the semester. We will work to accommodate students in remote time zones.