September 3, 2019

Yarowsky@jhu.edu
Johns Hopkins University
Department of Computer Science

Prof. David Yarowsky

(601.315/415/615)

Databases
Instructor: Prof. David Yarowsky
TA: Noam Finkelstein and TBA

Office Hours: Instructor - Tuesday/Thursday after class and by appointment. TA - TBA, special review sections, and by appointment.

Classroom: Hackerman B17
Meeting Time: Tu, Th: 3:00-4:15 PM

Hackerman 324
Malone Lab

yarowsky@jhu.edu
noam@jhu.edu

410-516-5372
601.315/415/615 - DATABASES
Required Textbook:


Other Potentially Useful Textbooks:

- S. Feuerstein, B. Pirby, D. Russell, *Oracle PL/SQL Programming*, O'Reilly
- Other

Reprinted Textbook:
Course Requirements

- Homeworks will include paper-and-pencil exercises and MySQL implementation exercises.
- The Final exam will be cumulative.

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Project</td>
<td>25%</td>
</tr>
<tr>
<td>Final Exam</td>
<td>25%</td>
</tr>
<tr>
<td>Midterm</td>
<td>20%</td>
</tr>
<tr>
<td>Homeworks (4)</td>
<td>25%</td>
</tr>
<tr>
<td>Class Participation</td>
<td>5%</td>
</tr>
</tbody>
</table>
One homework assignment may be handed in up to 5 days late without penalty.

Final projects handed in late will receive a penalty of 10% for every day late.

No other late homeworks will be accepted.

Lateness Policy
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 (Homeworks 1/2/4), unless group work is specified in writing (Homework 3, Project).** 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 midterms (and one other midterm) and their solutions will be explicitly distributed to students for practice and guidance regarding expectations. Students are encouraged to study using them. Likewise, Homework 4 is composed of questions given on previous 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.
Graduates and upperclass students are encouraged to enroll.

Nevertheless, 601.415/615 should be manageable by advanced undergraduates.

- Exams will differ somewhat and will be graded on a different scale.
- Homeworks in 601.415/615 will include 1 or more additional problems.
- They will differ primarily in terms of assignments and grading.
- 601.315 vs. 601.415/615 will be share common lectures.

601.315 vs. 601.415/615
Databases (315/415/615, Fall) and Database Systems (316/416/616, Spring) are complementary courses and make a natural course sequence.

- 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;
  - indexing data mining and natural language interfaces.

- 316/416/616 will focus on:
  - database internals and systems, including query and join processing,
  - database internals and systems, including query and join processing,
  - database internals and systems, including query and join processing,
  - database internals and systems, including query and join processing.

In contrast, 316/416/616 will focus on:

- The final project will be application-focused (e.g., how to design an implementation a database for a novel task) including practical execution of the concepts studied in the class.

(see below).
The course project(s) will focus on database system internals and development.

- database architectures, streaming, and partitioning.
Can I take 316/416/616 as a stand-alone course without any prerequisite? Yes, 316/416/616 does not have 315/415/615 as a formal prerequisite.

- 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.
- Anyone with a research focus in the databases area should certainly begin directly with 416. However, either through prior employment or via a prior course, you should have some database experience before taking 316/416/616.
Can I take 315/316 or 415/416 or 616/616 as a 2-course sequence?

Yes.

The instructors will work to make this a natural 2-course sequence. However, if you have already had a prior course in databases or in-

- There will be modest overlap of material (10%) but taught via different perspectives and emphases, 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.
- If you have already had a prior course in databases and/or storage systems, transaction processing and/or storage systems, then you are strongly encouraged to take 316/416/616 and then another advanced follow-on course in databases systems research, then you are strongly encouraged to continue in databases systems research.

Yes.
Can I take 315/416 as a sequence?

- Yes, 416 does not require 415 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 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 at a less difficult level, then this sequence could make sense.
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 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 world wide web or other online sources.
Stock market news and price correlation (data mining)

Internet proxy server database

Human genome databases

Bibliographic database for medical robotics

Astronomical and pharmaceutical databases for research support

Representations of acoustic data for speech recognition

Fantasy hockey league

Connecticut volunteer emergency rescue organization

JHU Fencing club and Anime club

Olympic sports data

Movie industry data (directors, producers, actors, films, etc.)

World geography and population data (from CIA World Fact Book)

Used car information (by model and year, from Edmunds)

Sample Final Project Domains (Previous Years)
Natural Language Interfaces to an Earthquake Database

Sample Final Project Domains (continued)

- Which country had the most powerful earthquake in 2017?
- Which country had the greatest number of earthquakes in 2017?
- What was the average magnitude of 2017 earthquakes in Asia?
- What was the magnitude of the most powerful earthquake in China?
- List the years in which there are at least two earthquakes of magnitude greater than 7 on the same continent.
SEGMENT 1 - SURVEY OF DATA MODELS

- Object-Oriented models
- Relational query languages: SQL, QBE (Query-by-Example)
- Formal representations: Relational algebra and calculus
- Relational model
- Entity-Relationship model (formal conceptual framework)
- Network and Hierarchical models (of historical interest)
SEGMENT 2 - Database Design and Implementation

Formal Analysis:
- Integrity constraints
- Domain constraints
- Triggers
- Functional dependencies
- Normalization

Practical Database Implementation:
- MySQL (a detailed exploration)
- Embedded SQL (in a host language like C or Perl)
- PL/SQl and stored procedures
SEGMENT 3 - Database System Internals

- Distributed databases
- Parallel databases
- Database system architectures
- Database security
- Recovery systems
- Transaction processing
- Query optimization
- Query processing
SEGMENT 4 - Emerging Technologies and Applications

- Decision Support Systems
- Data Mining
- Data Warehousing
- Natural Language Interfaces
- Very Large Text Databases and Information Retrieval
- Multimedia Databases (image, sound, video, etc.)
- XML-based data models
- Spatial, Geometric and Geographic Databases
- XML-based data models
- DNA and Human Genome Databases
- The impact of the WWW on database technology (and vice versa.)