September 1, 2016

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

Prof. David Yarowsky

(600.315 and 600.415)

DATABASES
Instructor: Prof. David Varowsky
TA: 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

yarowsky@jhu.edu

410-316-3372
 HACKERMAN 324

Instructor: Prof. David Varowsky

600.315/415 - DATABASES
Required Textbook:


Other Potentially Useful Textbooks:

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

Recommended Textbook:
Course Requirements

- Class Participation: 5%
- Homeworks (4): 25%
- Midterm: 20%
- Final Exam: 25%
- Final Project: 25%

Homeworks will include paper-and-pencil exercises and MySQL implementation exercises.

The Final exam will be cumulative.
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.
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 1-3), unless group work is specified in writing. Slicing program output or results is prohibited. Please see your professors if there are any questions about what is permissible. 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. If you use reprints of assignments, assignments specific to the assignment. If you use in the class to be sure you understand what is required by the assignment. However, you may discuss assignment specific material with others in the class.

You must abide by JHU’s Ethics Code, available at http://jhunix.hcf.jhu.edu/~ethicsbd. University officials will suffer a serious course grade penalty in addition to being reported to the registrar if there are any questions about what is permissible. Students who cheat or falsify program output or results 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.
Homeworks in 600.415 will include 1 or more additional problems and the final project will include additional component(s) not required for 600.315.

- They will differ primarily in terms of assignments and grading.
- Exams will differ somewhat and will be graded on a different scale.
- Homeworks in 600.415 will share some lectures.

600.315 vs. 600.415
Databases (315/415, Fall) and Database Systems (316/416, Spring) are complementary courses and make a natural course sequence (see below).

• Databases (315/415, Fall) and Database Systems (316/416, Spring)

600.315/415 vs. 600.316/416:

In contrast, 316/416 will focus on:

• database internals and systems, including query and join processing,

• database programming languages, especially SQL and PL/SQL,

• object-oriented and XML-based data models and future directions,

• database programming languages, especially SQL and PL/SQL,

• formal database models, theory and foundations,

• how to design and use a database;

315/415 focuses on:

• low-level implementation of the concepts studied in the class.

• implement a database for a novel task (including practical execution of data mining and natural language interfaces).

315/415 focuses on:

• how to design and make a natural course sequence (see below).

• how to design and use a database;

316/416 will focus on:

• database internals and systems, including query and join processing;

• formal database models, theory and foundations;

• database programming languages, especially SQL and PL/SQL;

• object-oriented and XML-based data models and future directions;

• database programming languages, especially SQL and PL/SQL.

315/415 focuses on:

• low-level implementation of the concepts studied in the class.

• implement a database for a novel task (including practical execution of data mining and natural language interfaces).
The course project(s) will focus on database system internals and database architectures, streaming, and partitioning.
Can I take 316/416 as a stand-alone course without 315 or 415?

Yes, 316/416 does not have 315 or 415 as a formal prerequisite.

Graduate students who have prior database employment experience or prior course, or others who wish to begin directly with 416, should have some database experience before taking 316/416. However, you should have some database experience before taking 316/416. Therefore, anyone with a research focus in the database area should certainly begin directly with 416.

Yes, 316/416 does not have 315 or 415 as a formal prerequisite.
Can I take 315/316 or 415/416 as a 2-course sequence?

Yes.

Therewillbemodestoverlapofmaterial(10%)buttaughtviadifferent
perspectivesandemphasis,andwillserveasagoodrefresher.

Ifyouhavenotakenapriorcourseindatabasesandareinterestedin
boththetheory/applicationsandsystemsidesofthefield,thenthis
sequence makesa lot of sense and is encouraged.

However,ifyouhavealreadyhada priorcourseindatabases,oryou
intendtocontinueindatabasesystemsisresearch,thenyouarestronglyencour-
aged to take 316/416 and then another advanced follow-on course in
databases. The instructorswill work to make this a natural 2-course sequence.

Ifyouhavenottakenapriorcourseindatabases,andintendtopursuersystemresearch,
therewillbemodestoverlapofmaterial(10%)buttaughtvia
different
perspectivesandemphasis,andoftenwillserveasagoodrefresher.

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 catching up 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 a less difficult level, then this sequence could make sense.
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 will be able to select final projects of interest to them from a fairly diverse set of options.

For most projects, students will be required to populate and test their implemented database design with substantial quantities of real world data. Details will be provided in class. Students may work in teams of 1 or 2 people.

The final project submission, including a full database implementation system specification and design, will be due in early November, including a detailed project proposal. A project proposal will be due in early November, including a detailed project proposal.

In MySQL, will be due shortly after the end of classes in December. The final project submission, including a full database implementation, will be due in early November, including a detailed project proposal.

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.
Sample Final Project Domains (previous years)

- Stock market news and price correlations (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 Film 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)
Which country had the greatest number of earthquakes in 1996?

What was the magnitude of the most powerful earthquake in China?

What was the average magnitude of 1996 earthquakes in Asia?

Which country had the greatest number of earthquakes in 1996?

List the years in which there are at least two earthquakes of magnitude greater than 7 on the same continent.

Which country had the most powerful earthquake in 1994?

Which country had the most powerful earthquake in 1993?

Natural Language Interfaces to an Earthquake Database

Sample Final Project Domains (continued)
SEGMENT I - 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
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

SEGMENT 4 - Information Retrieval and Web Agents

The Impact of the WWW on Database Technology (and vice versa)

- Very Large Text Databases and Information Retrieval
- Multimedia Databases (images, sound, video, etc.)
- DNA and Human Genome Databases
- XML-based Data Models
- Spatial, Geometric, and Geographic Databases
- Natural Language Interfaces
- Data Warehousing
- Data Mining
- Decision Support Systems