Office Hours: Instructor - Wed 3-4, Thursday after class and by appointment.

Classroom: Hackerman B17

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

Prof. David Yarovsky

Instructor: Prof. David Yarovsky

TA: Shyam Yarovsky@jhu.edu

Yarovsky@jhu.edu

410-316-5372

Hackerman 324

Shyam Malone 255

Hackerman B17

Other Potentially Useful Textbooks:
- S. Feuerstein, B. Pribyl, D. Russell, Oracle PL/SQL Programming, O'Reilly
- L. Wall, S. Potter and R. Schwartz, Programming Perl, O'Reilly
- L. Wall, S. Potter and R. Schwartz, Programming Perl, O'Reilly
- A. Silberschatz, H. Korth and S. Sudarshan, Database System Concepts
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.

**Academichonestyisrequiredinallworkyousubmittobegraded.**

You must solve all homework and programming assignments entirely on your own (Homeworks 1-3), unless group work is specified in writing. Group work is permissible (Homework 4, Project). This means you must not show your program output or results to 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. If you are uncertain about what is permissible or what is required by the assignment, please see your professor. Plagiarism program output or results is prohibited. Submitting program output or results is prohibited.

If you use source code from another source, include a full reference to that effect in your submission. You may discuss assignment specifications with others in the class to be sure you understand what is required. However, you may not discuss assignment specifications with others in the class to be sure you understand what is required. You must not discuss assignment specifications with others in the class. You must not use source code from sources other than your text.

**You must not show your program output or results to others in the class.**

If you have any questions about what is permissible, please see your professor. 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.
600.315 vs. 600.415

...and upperclass students are encouraged to enroll.

Nevertheless, 600.415 should be manageable by advanced undergraduates.

Exams will differ somewhat and will be graded on a different scale.

Exams in 600.315.

The final project will include additional component(s) not required for 600.315.

Homeworks in 600.415 will include 1 or more additional problems and will differ primarily in terms of assignments and grading.

They will differ primarily in terms of assignments and grading.

600.315/415 will be share common lectures.
Databases (315/415, Fall) and Database Systems (316/416, Spring) are complementary courses and make a natural course sequence (see below).

- In contrast, 316/416 will focus on:
  - Indexing, the organization, estimation and optimization of database internals and systems, including query and join processing,
  - Database internals and systems, including practical execution of a database for a novel task (including practical execution of 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),
  - Formal database models, theory and foundations,
  - How to design and use a database.

315/415 focuses on:
- How to design and make a natural course sequence (see below).
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 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.

Over, either through prior employment or via a prior course, you should have some database experience before taking 316/416, however.
Can I take 315/316 or 415/416 as a 2-course sequence?

Yes.

The instructors will work to make this a natural 2-course sequence. There will be modest overlap of material (~10%) but taught via different perspectives and emphases, and will serve as a good refresher.

However, if you have already had a prior course in databases, or intend to continue in database systems research, then you are strongly encouraged to take 316/416 and then another advanced follow-on course in database/systems taught by Professors Ahmad or Burns.

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.

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

Yes.

Therewillbemodestoverlapofmaterial(10%)buttaughtviadifferentperspectivesandemphasis,andallofserveasagoodrefresher.

Ifyouhavenottakenapriorcourseindatabasesandareinterestedinboththetheory/applicationsandsystemsidesofthefield,thenthissequencemakesalotsenseandisencouraged.

However,ifyouhavealreadyhadapriorcourseindatabases,orintendtocontinueindatabasesystemsresearch,thenyouarestronglyencouragedtotake316/416andthenanotheradvancedfollow-oncourseindatabase/systems taughtbyProfessorsAhmadorBurns.

Ifyouhavenottakenapriorcourseindatabasesandareinterestedinboththetheory/applicationsandsystemsidesofthefield,thenthissequencemakesalotsenseandisencouraged.

However,ifyouhavealreadyhadapriorcourseindatabases,orintendtocontinueindatabasesystemsresearch,thenyouarestronglyencouragedtotake316/416andthenanotheradvancedfollow-oncourseindatabase/systems taughtbyProfessorsAhmadorBurns.
Can I take 315/416 as a sequence?

Yes, 416 does not require 415 as a prerequisite, but you should have

meet the expectations of the 416 instructor: done well in 315 and be prepared to do some background catching up to
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.
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.
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 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 (continued)

Which country had the greatest number of earthquakes in 1996?
What was the magnitude of the most powerful earthquake in China?
What was the magnitude of the most powerful earthquake in Asia?
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?

```
SELECT Countryname
FROM Quake
WHERE Year = 2013
FROM Quake
(SELECT MAX(magnitude)
WHERE magnitude IN
FROM Quake
WHERE Quake
SELECT Countryname
WHERE Year = 2013

SELECT Countryname
FROM Quake
WHERE Year = 2013
FROM Quake
(SELECT MAX(magnitude)
WHERE magnitude IN
FROM Quake
WHERE Quake
SELECT Countryname
WHERE Year = 2013
```

Natural Language Interfaces to an Earthquake Database
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 1 - SURVEY OF DATA MODELS
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 2 - Database Design and Implementation
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

Emerging Technologies and Applications

- Decisionsupport systems
- Datamining
- Datawarehousing
- Natural language interfaces
- Spatial, geometric and geographic databases
- XML-based data models
- Multimedia Databases (image, sound, video, etc.)
- DNA and Human Genome databases
- Very large text databases and information retrieval
- The impact of the WWW on database technology (and v.v.)