Office Hours: Instructor - Tuesday/Thursday after class and by appointment.

Classroom: Hackerman B17

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

Instructor: Prof. David Varowsky

TA: Geet Sawhney

Varowsky@jhu.edu

410-516-5372

Hackerman 324G

geet.sawhney@jhu.edu

Malone Lab

601.315/415/615/615 - DATABASES
Required Textbook:


Other Potentially Useful Textbooks:


Required Textbook:

Textbooks
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 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 (Homework 3, Project). This means you must not show your program output or results 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.

Each homework (Homeworks 1/2/4) unless group work is specified in writing must be submitted individually.

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. Falsifying program output or results is prohibited.

Computer Science Academic Integrity Code
Graduates and upperclass students are encouraged to enroll.

Nevertheless, 601.415/615 should be manageable by advanced under-

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

For 601.315.

Homeworks in 601.415/615 will include 1 or more additional problems

They will differ primarily in terms of assignments and grading.

601.315/615 vs. 601.415/615 will be share common lectures.
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:
  - database programming languages, especially SQL and PL/SQL;
  - formal database models, theory and foundations;
  - how to design and use a database

316/416/616 focuses on:
- database internals and systems, including query and join processing, including practical execution of the concepts studied in the class.
- indexing, the organization, estimation and optimization of database internals and systems, including query and join processing.

In contrast, 316/416/616 will focus on:
- database internals and systems, including query and join processing, including practical execution of the concepts studied in the class.
- indexing, the organization, estimation and optimization.
- database programming languages, especially SQL and PL/SQL;
- formal database models, theory and foundations;
- how to design and use a database

(see below).
database architectures, streaming and partitioning.

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

- Database architectures, streaming and partitioning.
Can I take 316/416/616 as a stand-alone course without a research focus in the databases area? 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. However, either through prior employment or via a prior course, you should have some database experience before taking 316/416/616.

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

...
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 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.

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

Students may work in teams of 1 or 2 people.

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.

The final project submission, including a full database implementation in MySQL, will be due shortly after the end of classes in December.
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 Factbook)

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

Sample Final Project Domains (Previous Years)
<table>
<thead>
<tr>
<th>Question</th>
<th>SQL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which country had the greatest number of earthquakes in 2017?</td>
<td><code>SELECT Countryname FROM Quake WHERE Year = 2017</code></td>
</tr>
<tr>
<td>What was the magnitude of the most powerful earthquake in China?</td>
<td><code>SELECT magnitude FROM Quake WHERE Year = 2017 AND Countryname = 'China'</code></td>
</tr>
<tr>
<td>What was the average magnitude of 2017 earthquakes in Asia?</td>
<td><code>SELECT AVG(magnitude) FROM Quake WHERE Year = 2017 AND Continent = 'Asia'</code></td>
</tr>
<tr>
<td>Which earthquake had the greatest number of earthquakes on the same continent?</td>
<td><code>SELECT magnitude FROM Quake WHERE Year = 2017 AND Continent = 'Asia'</code></td>
</tr>
<tr>
<td>List the years in which there were at least two earthquakes of magnitude greater than 7</td>
<td><code>SELECT Year FROM Quake GROUP BY Year HAVING COUNT(*) &gt;= 2</code></td>
</tr>
<tr>
<td>Which country had the most powerful earthquake in 2017?</td>
<td><code>SELECT Countryname FROM Quake WHERE Year = 2017 AND magnitude = (SELECT MAX(magnitude) FROM Quake WHERE Year = 2017)</code></td>
</tr>
</tbody>
</table>

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:
- PL/SQL and stored procedures
- Embedded SQL (in a host language like C or Perl)
- MySQL (a detailed exploration)
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:

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