

---

# **DATABASE SYSTEMS**

**(600.315 and 600.415)**

**Prof. David Yarowsky**

Department of Computer Science  
Johns Hopkins University  
yarowsky@cs.jhu.edu

September 3, 2009



---

## **600.315/415 - DATABASE SYSTEMS**

**Instructor:** Prof. David Yarowsky      **TAs:** TBA

CSEB 324

410-516-5372

yarowsky@gmail.com

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

**Classroom:** CSEB B17

**Office Hours:** Instructor - Wed 3-4, Tuesday/Thursday after class and by appointment  
TA - TBA, special review sections, and by appointment.



---

## Textbooks

### Required Textbook:

- A. Silberschatz, H. Korth and S. Sudarshan, *Database System Concepts*, 5th Edition, McGraw Hill, 2005. ISBN: 0-07-295886-3 (earlier editions acceptable).

### Other Potentially Useful Textbooks:

- R. Ramakrishnan and J. Gehrke, *Database Management Systems*, 2nd Edition. McGraw Hill, 2000. ISBN: 0-07-232206-3.
- S. Feuerstein, B. Pribyl, D. Russell, *Oracle PL/Sql Programming*, O'Reilly & Associates, ISBN: 1-565-92335-9 (well-written PL/SQL guide)
- L. Wall, S. Potter and R. Schwartz, *Programming Perl*, O'Reilly. ISBN: 1-56592-149-6 (strongly recommended for many final projects).



---

## 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.



---

## 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 1-3), unless group work is specified in writing (Homework 4, 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>.



---

## **600.315 vs. 600.415**

- 600.315/415 will be share common lectures.
- They will differ primarily in terms of assignments and grading.
- 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.
- Exams will differ somewhat and will be graded on a different scale.
- Nevertheless, 600.415 should be manageable by advanced undergraduates and upperclass students are encouraged to enroll.



---

## 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 world wide web or other online sources.



---

## **Sample Final Project Domains (previous years)**

- Used car information (by model and year, from Edmunds)
- World geography and population data (from CIA world fact book)
- Movie industry data (directors, producers, actors, films, etc.)
- Olympic sports data
- JHU Fencing club and Anime film club
- Connecticut volunteer emergency rescue organization
- Fantasy hockey league
- Representations of acoustic data for speech recognition
- Astronomical and pharmaceutical databases for research support
- Bibliographic database for medical robotics
- Human genome databases
- Internet proxy server database
- Stock market news and price correlations (data mining)



---

## Sample Final Project Domains (continued)

- Natural language interfaces to an earthquake database
  - 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?
  - 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 magnitude IN
    ( SELECT MAX (magnitude)
      FROM Quake
    WHERE Year = 1994 )
```



---

## **SEGMENT 1 - Survey of Data Models**

- Network and Hierarchal models (of historical interest)
- Entity-Relationship model (formal conceptual framework)
- Relational model
  - Formal representations: relational algebra and calculus
  - Relational query languages: SQL, QBE (Query-by-Example)
- Object-Oriented models



---

## 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**

- Physical organization (storage and file structure)
- Data structures for efficient access
- Query processing
- Query optimization
- Transaction processing
- Concurrency control
- Recovery systems
- Database security
- Database system architectures
- Parallel databases
- Distributed databases



---

## **SEGMENT 4 - Emerging Technologies and Applications**

- Decision support systems
- Data mining
- Data warehousing
- Natural language interfaces
- Spatial, geometric and geographic databases
- xml-based data models
- DNA and Human Genome databases
- Multimedia Databases (image, sound, video, etc.)
- Very large text databases and information retrieval
- The impact of the WWW on database technology (and v.v.)  
⇒ 600.466 - Information Retrieval and Web Agents

