

COSC6340: Database Systems

Instructor: Carlos Ordonez

1 Course information

Google newsgroup: "COSC6340-". Instructions on TA web page.
email: firstname AT uh edu (Start subject line with "COSC6340-")
office hours: posted on web page.

2 Prerequisites

There are no prereqs: the course is self-contained. No previous database knowledge is required. It is desirable students have basic CS background: discrete math, algorithm analysis, operating systems and "systems" programming.

3 Course contents

This is a graduate level course on database management systems (DBMSs). The textbook is [2], complemented by [3] and [1]. The course will require reading research papers available on the web.

Topics include the following. Fundamental theory: set theory, first order logic, relational model and relational algebra. Dependencies, normalization up to BCNF. Database design and software engineering: ER model, workflows. Query processing: SQL, SPJA queries, derived tables/view, pivoting, recursive queries. Internal subsystems of a DBMS: secondary storage, alternative storage architectures, buffer management, indexing data structures, concurrency control, transaction processing, query optimizer, fault tolerance: for transaction processing (recovery) and for long query processing (parallel, incremental). Advanced topics (they vary, optional): DBMS support for big data, advanced SQL (window functions, stored procedures, UDFs, triggers, cubes), normalization up to 5NF, reverse data engineering, security mechanisms, main memory processing, blockchain support, query languages beyond SQL.

4 Grading

- 80%: 2 programming projects.
- 20%: Midterm exam (around 10th week).

Project 1 will involve database normalization and transaction processing in a row DBMS. Project 2 will involve query processing, creating an SQL query generator to analyze a large database with many tables, in a columnar DBMS. DBMSs will work on Unix (Linux). Programming will be done primarily in SQL and Python. C++ will be optional. Programming project will be done in pairs (i.e. a team of 2 students). Programs will be carefully tested by TA for correctness and efficiency, with correctness being the most important requirement.

References

- [1] S. Abiteboul, R. Hull, and V. Vianu. *Foundations of Databases : The Logical Level*. Pearson Education POD, facsimile edition, 1994.
- [2] H. Garcia-Molina, J.D. Ullman, and J. Widom. *Database Systems: The Complete Book*. Prentice Hall, 1st edition, 2001.
- [3] J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, San Francisco, 2nd edition, 2006.