CSCI 685: Advanced Topics in Database Management Systems

Pre-requisites: CSCI 485 or 585
Time: Fall 2013, Monday and Wednesday 10 to 11:20
Instructor: Professor Shahram Ghandeharizadeh, shahram@usc.edu, 213-740-4781
Office: SAL 208
TA: Jason Yap, jyap@usc.edu, SAL 102

Introduction and Purposes

This course introduces students to advanced topics in the area of database management systems. Topics include: query optimization techniques, cache management, parallel and distributed database systems, extendible storage management techniques, data availability techniques, data mining and knowledge discovery, stream data management, multidimensional data analysis, decision support, spatial index structures, expedited software development using ORMs, peer-to-peer and sensor data management, continuous media (audio and video clips). It covers both the standard techniques beyond the undergraduate level, and some advanced topics of active research. It assumes student familiarity with a conceptual data modeling tool such as Entity-Relationship (ER) data model, a logical data model such as the relational data model, SQL as a commercial query language, normal forms and logical data design, physical characteristics of storage devices such as magnetic disks and memory, physical design of a database using persistent data structures such as B+-tree and Hash indexes, concurrency control and crash recovery techniques.

Course Requirements and Grades

There are no required text books. The reading material is based on recently published technical papers available via the ACM/IEEE/Springer digital libraries. All USC students have automatic access to these digital archives.

Grading breakdown

- Midterm 1, Oct 7, 2013: 30%
- Midterm 2, Dec 4, 2013: 30%
- Class project: 30%
- Homework assignments: 10%

Course Readings/Class Sessions

The reading material for the course is organized chronologically and based on a specific theme. This material is tentative and might be fine tuned using different publications. We will make adjustments to the list as new manuscripts (currently under review) are accepted for publication.

Week 1: Introduction and review of transactions, concurrency control and crash recovery protocols. This topic is covered by most introductory textbooks in the area of databases, e.g., Chapters 14, 15 and 16 of Database Systems Concepts, sixth edition, by Silberschatz, Korth and Sudarshan.

Suggested Reading:

1. P. Bailis and A. Ghodsi. Eventual Consistency Today: Limitations, Extensions, and Beyond. CACM, Vol. 56, No. 5, May 2013.

Weeks 2-4: Alternative Architectures

2. S. Ghandeharizadeh and J. Yap. Cache Augmented Database Management Systems. In ACM SIGMOD Workshop on Databases and Social Networks, June 2013.
3. S. Barahmand, S. Ghandeharizadeh and J. Yap. A Comparison of Two Physical Data Designs for Interactive Social Networking Actions. CIKM, Oct 2013.

4. M. Stonebraker and R. Cattell. 10 Rules for Scalable Performance in 'Simple Operation' Datastores, CACM, Vol. 54, No. 6, June 2011.

Suggested Reading:
5. S. Barahmand and S. Ghandeharizadeh. Expedited Benchmarking of Social Networking Actions with Faster Data Loading Techniques. CIKM, Oct 2013.

Week 5: Programming database management systems
6. P. Bernstein et. al. Incremental Mapping Compilation in an Object-to-Relational Mapping System. ACM SIGMOD 2013.

7. J. Dean and S. Ghemawat. MapReduce: Simplified Data Processing on Large Clusters. In Communications of the ACM, Vol. 51, No. 1, 2008.

Suggested Reading:
8. M. Schatz and B. Langmead. The DNA Data Deluge. IEEE Spectrum, July 2013.

Week 6: Data availability techniques
9. D. Patterson, G. Gibson, and R. Katz. A Case for Redundant Arrays of Inexpensive Disks (RAID). ACM SIGMOD 1988.

Week 7, 8: Advanced index structures
10. Guttman. R-Trees: A Dynamic Index Structure for Spatial Searching. SIGMOD 1984.

11. P. E. O'Neil, and D. Quass. Improved Query Performance with Variant Indexes. SIGMOD 1997.

Week 9: Performance Modeling
12. S. Currim, et. al. DBMS Metrology: Measuring Query Time, SIGMOD 2013.

Suggested Reading:
13. S. Barahmand and S. Ghandeharizadeh. BG: A Social Networking Benchmark. CIDR, Jan 2013.
14. P. Meenan. How Fast is Your Website? CACM, Vol.56, No. 4, April 2013.

Weeks 10, 11: Multi-core CPUs
15. H. Jung et. al. A Scalable Lock Manager for Multicores. SIGMOD 2013.

16. T. Horikawa. Latch-Free Data Structures for DBMS. SIGMOD 2013.

Suggested Reading:
17. S. Bahra. Nonblocking Algorithms and Scalable Multicore Programming. CACM, Vol. 56, No. 7, July 2013.

Week 12: Beyond SQL and relational data model

18. M. Seltzer. Beyond Relational Databases. Communications of the ACM, July 2008, Vol. 51, No. 7.

Week 13-14: Data Center Scale Data Stores

19. J. Baker et. al. MegaStore: Providing Scalable, Highly Available Storage for Interactive Services. CIDR, Jan 2011.

Suggested Reading:

20. F. Chang et al. Bigtable: A Distributed Storage System for Structured Data. In OSDI 2006.
21. R. Nishtala, et. al. Scaling Memcache at Facebook. NSID, April 2013.
22. D. DeWitt et al. The Gamma Database Machine Project. IEEE Transactions on Knowledge and Data Engineering, Vol. 2, 1990.
23. J. Corbett et al. Spanner: Google's Globally Distributed Database. OSDI 2012.

Week 15: Flash Memory

24. M. Bender et al. Don't Thrash: How to Cache Your Hash on Flash. VLDB 2012.

Suggested Reading:

25. L. Bouganim et al. UFLIP: Understanding Flash IO Patterns. CIDR 2009.
26. A. Leventhal. A File System All Its own. CACM, Vol. 56, No. 5, May 2013.

Class Project

The class project requires students to investigate an innovative concept. This might be either conceptual or involve a prototyping effort. With the first, a student (team) explores design alternatives underlying the concept. They start by identifying the broader topic. After surveying the relevant literature, they produce a survey like paper that may include a discussion of future extensions. With a prototyping effort, a student (team) investigates implementation of a design. Such implementations may employ a relational storage manager such as Berkeley DB. Given the explosion of open software releases in the area of databases, there are many opportunities for such projects. In each case, an outstanding project may develop novel extensions that result in a publication.

Statement for Students with Disabilities

Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me (or to TA) as early in the semester as possible. DSP is located in STU 301 and is open 8:30 a.m.–5:00 p.m., Monday through Friday. The phone number for DSP is (213) 740-0776.

Statement on Academic Integrity

USC seeks to maintain an optimal learning environment. General principles of academic honesty include the concept of respect for the intellectual property of others, the expectation that individual work will be submitted unless otherwise allowed by an instructor, and the obligations both to protect one's own academic work from misuse by others as well as to avoid using another's work as one's own. All students are expected to understand and abide by these principles. Scampus, the Student Guidebook, contains the Student Conduct Code in Section

11.00, while the recommended sanctions are located in Appendix A: http://www.usc.edu/dept/publications/SCAMPUS/gov/

.

Students will be referred to the Office of Student Judicial Affairs and Community Standards for further review, should there be any suspicion of academic dishonesty. The Review process can be found at: http://www.usc.edu/student-affairs/SJACS/