CSCI 550: Advanced Data Stores

Pre-requisites: CSCI 485 or 585
Time: Fall 2020, Mon/Wed 12-1:50
Instructor: Professor Shahram Ghandeharizadeh, shahram@usc.edu, 213-740-4781
Office: SAL 208, Mon 3-4:30, Thu 1-2:30

Introduction and Purposes
Past decade has witnessed a flurry of new architectures, consistency techniques, data models, and query languages for data stores. These are in support of alternative workloads and application use cases. CSCI 550 introduces students to both the applications and fundamentals of these novel data stores. We focus on the NoSQL data model, SQL data stores that support machine learning (ML), cache augmented data store architecture and its data model, how it provides persistence by complementing both SQL and NoSQL data stores, consistency guarantees, alternative cache replacement techniques, blockchain, and scalability. In addition, we investigate biologically inspired data stores as an emerging area of research.

Purpose: Advanced data stores are designed to provide high performance, scale both vertically and horizontally, and continue operation in the presence of node failures. We exercise these concepts in the context of our hands-on project while implementing concepts such as atomicity, isolation, durability, and strong consistency. One lecture each week discusses designs and implementation details.

Course Requirements
Student must have enrolled in an introductory course on database management systems (USC’s CSCI 485 or 585) and be familiar with SQL query language, query optimization techniques, transaction processing systems and ACID properties, concurrency control and crash recovery protocols, magnetic disk and SSD, block-based index structures such as hash and B+-tree.

Students must have strong programming skills and be familiar with the c and Java programming languages, Github, and Eclipse IDE.

Course 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, Monday, October 5, 2020: 25%
   Exam 2, Monday, November 23, 2020: 25%
   Class project is cumulative and due Dec 4 (scheduled finals): 30%
   Class participation: 20%

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.
Weeks 1 and 2 (Classes begin August 24): Introductions


Suggested Reading:


Weeks 3 and 4: Parallelism and Load Balancing


Week 5: Data Stores and ML, Elasticity


Week 6: Hybrid OLTP and OLAP, Data Skew

Week 7: Consistency & Start Class Project
- Students start to form a 2 to 3 member team.

Weeks 8 and 9: BG and YCSB Benchmarks
- BG Client for MySQL. BG Client for MySQL and memcached.
- YCSB Client for MySQL. YCSB Client for MySQL and memcached.

Suggested Reading:

Weeks 10 and 11: Data Reorganization

Suggested Reading:

Week 12: Blockchain

Suggested Reading:

**Weeks 13: Memory Management**

*Suggested Reading:*

**Time permitting: Log Structured Data Stores**

*Suggested Reading:*

**Time permitting: DNA Storage**

*Suggested Reading:*

**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/

**USC Viterbi School of Engineering Honor code**
Engineering enables and empowers our ambitions and is integral to our identities. In the Viterbi community, accountability is reflected in all our endeavors.

- Engineering + Integrity.
- Engineering + Responsibility.
- Engineering + Community.

Think good. Do better. Be great.

These are the pillars we stand upon as we address the challenges of society and enrich lives.

**Academic Conduct:**
Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in Scampus in Part B, Section 11, “Behavior Violating University Standards” policy.usc.edu/scampus-part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in Scampus and university policies on scientific misconduct, http://policy.usc.edu/scientific-misconduct.

**Support Systems:**
*Student Counseling Services (SCS) – (213) 740-7711 – 24/7 on call*
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention. engemannshc.usc.edu/counseling

*National Suicide Prevention Lifeline – 1 (800) 273-8255*
Provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week. www.suicidepreventionlifeline.org

*Relationship and Sexual Violence Prevention Services (RSVP) – (213) 740-4900 – 24/7 on call*
Free and confidential therapy services, workshops, and training for situations related to gender-based harm. engemannshc.usc.edu/rsvp

*Sexual Assault Resource Center*
For more information about how to get help or help a survivor, rights, reporting options, and additional resources, visit the website: sarc.usc.edu

*Office of Equity and Diversity (OED)/Title IX Compliance – (213) 740-5086*
Works with faculty, staff, visitors, applicants, and students around issues of protected class. equity.usc.edu

*Bias Assessment Response and Support*
Incidents of bias, hate crimes and microaggressions need to be reported allowing for appropriate investigation and response. studentaffairs.usc.edu/bias-assessment-response-support
The Office of Disability Services and Programs
Provides certification for students with disabilities and helps arrange relevant accommodations. dsp.usc.edu

Student Support and Advocacy – (213) 821-4710
Assists students and families in resolving complex issues adversely affecting their success as a student EX: personal, financial, and academic. studentaffairs.usc.edu/ssa

Diversity at USC
Information on events, programs and training, the Diversity Task Force (including representatives for each school), chronology, participation, and various resources for students. diversity.usc.edu

USC Emergency Information
Provides safety and other updates, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible. emergency.usc.edu

USC Department of Public Safety – UPC: (213) 740-4321 – HSC: (323) 442-1000 – 24-hour emergency or to report a crime.
Provides overall safety to USC community.