USC Viterbi School of Engineering

DSCI 551: Foundations of Data Management
Units: 4

Term—Day—Time:
Spring 2026

32401: THH 201, 12-1:50pm, MW
32402: RTH 109, 3-6:20pm, Mon
32423: DEN, 3-6:20pm, Mon

Instructor: Wensheng Wu
Office Hours: 11am-12pm Tuesday, by appointment
Contact Info: wenshenw@usc.edu

TAs: To be announced on course web site.
Office Hours: TBD

A. Catalogue Course Description
Function and design of modern storage systems, including cloud; data management techniques; data modeling; network attached storage, clusters and data centers; relational databases; the map-reduce paradigm.

B. Expanded Course Description
This course aims to provide students with a comprehensive introduction to the landscape of modern data management, including SQL, NoSQL, and big data management. It is one of the core courses in the Data Science programs.

The course may be divided into **four** parts. (1) Fundamentals, covering data storage, file system, and file format (JSON, encoding); (2) SQL, covering conceptual data modeling, relational modeling, SQL, constraints, views, indexing (B+-tree), and query execution algorithms; (3) NoSQL, covering cloud-hosted JSON database (Firebase) & REST API, XML & XPath, and MongoDB & functions; (4) Big data system, covering distributed data storage (HDFS), map reduce (Hadoop), and Spark (DataFrame and RDD).

C. Recommended Preparation:
Solid background in computing at the level of DSCI 550. Basic understanding of operating systems and computer networking is desired. Familiar with Python programming (e.g., by taking DSCI 510 or equivalent).

D. Course Notes
The course will be run as lecture classes with required student participation. There are weekly readings and students are encouraged to complete the readings prior to the discussions in class. All course materials, including lecture notes, homework assignments, and project guideline will be posted online on the course Web site.

E. Technological Proficiency and Hardware/Software Required
Students are expected to have their own laptop or desktop computer where they can install and run software to complete the homework assignments and projects.

F. Recommended Readings and Supplementary Materials

- [**AA**] Remzi H. Arpaci-Dusseau and Andrea C. Arpaci-Dusseau. *Operating Systems: Three Easy Pieces*, 2015 (selected chapters only). Available free at: http://pages.cs.wisc.edu/~remzi/OSTEP/
- [**GUW**] Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom. Database Systems: The Complete Book (Second Edition), Prentice Hall, 2009 (selected chapters only, see schedule below). Book web site: http://infolab.stanford.edu/~ullman/dscb.html
- [SQL] Alan Beaulieu. Learning SQL: Generate, Manipulate, and Retrieve Data. 3rd Edition. O'Reilly, 2022. Freely accessible from USC library.
- [MongoDB] Kristina Chodorow. MongoDB: The Definitive Guide, 2nd Edition. O'Reilly, 2013. Freely accessible from USC library.
- [Hadoop] Tom White. Hadoop: The Definitive Guide. O'Reilly Media; 2010. Freely accessible from USC library.
- [Spark] Chambers, Bill ; Zaharia, Matei. Spark: Big Data Processing Made Simple. O'Reilly Media, 2018. Freely accessible from USC library.

  Note that the last four books are freely accessible from USC library. Links can be found above.

In addition to the textbook, students may be given additional reading materials such as research papers. Students are responsible for all reading assignments.

G. Grading Scheme

Homework Assignments: There will be 5 homework assignments. The assignments must be done individually. Each assignment is typically graded on a scale of 0-100 and the specific guideline will be provided for each assignment.

Exams: There will be three exams. Exams 1 and 2 will not be accumulative. Exam 3 will be comprehensive and emphasize the materials after the 2nd exam. All exams are closed-book and closed-notes, and in person (in classroom). Per the school policy, DEN students will take exams remotely at the **same** time as the on-campus students. Proctoring of DEN students will be conducted online (e.g., over Zoom with camera turned on).

Course project: Students are also expected to complete a term project on data management. Details will be announced separately.

Participation: On-campus students are required to attend the classes in person and participate in class discussions. Exceptions to attendance will be considered **only** for medical emergencies with doctor notes provided ahead of class meeting times.

Grade breakdown:

Homework	30%
Exam 1	15%
Exam 2	18%
Exam 3	20%
Course project	15%

Participation   2%
_____
Total   100%

Letter grades will range from A through F. The following are the cut-offs:

[93, 100] = A   [73, 76) = C
[90, 93) = A-   [70, 73) = C-
[87, 90) = B+   [67, 70) = D+
[83, 87) = B   [63, 67) = D
[80, 83) = B-   [60, 63) = D-
[77, 80) = C+   Below 60 is an F

Note that this is an absolute grading (no curving will be applied). Grades are not negotiable. Note that every point counts and no rounds up will be performed. Requests for rounding up, asking for special treatments, etc. will be ignored and will be subject to a minimum penalty of 10% deduction of the grade.

H. Grading Policy
Your coursework (including homework assignments and project deliverables) is due at 11:59pm on the due date and should be submitted on the course Web site as announced. **No late submissions will be accepted**. You are responsible for making sure you have stable network connection for the submission and submitting the correct files for your coursework.

Makeup for exams and extension of coursework deadlines may be considered only when there are documented medical emergencies. Doctor notes are needed as proof, signed and with contact information of doctors. Any requests after the exam time and homework deadline will not be considered. **One-week** in advance notices are required for scheduling a makeup of exam. **48** hours in advance notice are required for the extension of homework deadline. No makeups will be given for situations such as interview, job fairs, family emergencies, and other personal matters. Students are responsible for scheduling to avoid conflicts with class meeting times and for any missing coursework under these situations.

Regrading requests must be made (by emailing TAs or following the instructions posted by TAs) within one week after the solutions or grades have been posted, unless announced otherwise. Grades are final after the regrading period.

I. Course Schedule: A Weekly Breakdown (may be revised when the course progresses)
Note for Monday only section, due to holidays, the lectures may be adjusted and/or merged to the meetings before or after the holidays.

Week	Topic	Readings	Homework/Project
1 (1/12)	• Course Introduction	• [AA] Chapter 2 (optional) • [AA] Chapter 4 (optional)	
2 (1/19)	*No class on Monday, 1/19, MLK holiday*	• [AA] Chapter 37 (storage system)	HW1 out

	• *NoSQL 1: Firebase & JSON*		
3 (1/26)	• Storage System • File System	• [AA] Chapter 39 • [AA] Chapter 40	HW1 due HW2 out
4 (2/2)	• File Format • XML & XPath	• [GVW] Sec. 11.1-3, 12.1	HW2 due
5 (2/9)	• Review • Exam 1 (2/9 for Monday only section; 2/11 for MW section)		
6 (2/16)	*No class on 2/16, Monday, President's day* • Data Modeling (ER & relational) • SQL	• [GUW] Sec. 4.1-4.6, 2.1-2.1 • [GUW] Sec. 2.3, 6.1-6.5 [SQL] Chapters 3-6	Project proposal due
7 (2/23)	• SQL • Constraints & views	• [GUW] Sec. 2.3, 6.1-6.5 • [SQL] Chapters 7-10	HW3 out
8 (3/2)	• *NoSQL 2: MongoDB* • Data organization & external sorting	• [MongoDB] Parts 1 and 2 • [GUW] Sec. 7.1-7.2, 8,1, 8.3	HW3 due HW4 out
9 (3/9)	• Indexing (B+-tree) • Query execution	• [GUW] Sec. 13.5, 13.7 • [GUW] Sec. 14.1-14.2, • [GUW] Chapter 15	HW4 due Project midterm report due
10 (3/16)	*No class. Spring recess.*		
11 (3/23)	• Review • Exam 2 (3/23 for Monday only section; 3/25 for MW section)		
12 (3/30)	• Hadoop HDFS	• K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop distributed file system," in Mass Storage Systems and Technologies (MSST), 2010 IEEE 26th Symposium on, 2010, pp. 1-10. • [Hadoop] Chapter 3	
13 (4/6)	• Hadoop MapReduce	• J. Dean and S. Ghemawat, MapReduce: simplified data processing on large clusters," Communications of the ACM, vol. 51, pp. 107-113, 2008. • R. Cattell, "Scalable SQL and NoSQL data stores,"	HW5 out

		ACM SIGMOD Record, vol. 39, pp. 12-27, 2011. • [Hadoop] Chapter 2	
14 (4/13)	• Apache Spark	• Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing, Matei Zaharia, et. al., NSDI, 2012. • Zaharia, Matei and Chowdhury, Mosharaf and Franklin, Michael J. and Shenker, Scott and Stoica, Ion. Spark: cluster computing with working sets. HotCloud, 2010. • [Spark] Chapter 12	HW5 due
15 (4/20)	• Project in-class demo		
16 (4/27)	• Exam 3 (4/27 for Monday only section; 4/29 for MW section)		
Final week	• Project final report due on 5/8, Friday		

Statement on Academic Conduct and Support Systems

Academic Integrity:
The University of Southern California is a learning community committed to developing successful scholars and researchers dedicated to the pursuit of knowledge and the dissemination of ideas. Academic misconduct, which includes any act of dishonesty in the production or submission of academic work, comprises the integrity of the person who commits the act and can impugn the perceived integrity of the entire university community. It stands in opposition to the university's mission to research, educate, and contribute productively to our community and the world.

All students are expected to submit assignments that represent their own original work, and that have been prepared specifically for the course or section for which they have been submitted. You may not submit work written by others or "recycle" work prepared for other courses without obtaining written permission from the instructor(s).

Other violations of academic integrity include, but are not limited to, cheating, plagiarism, fabrication (e.g., falsifying data), collusion, knowingly assisting others in acts of academic dishonesty, and any act that gains or is intended to gain an unfair academic advantage.

The impact of academic dishonesty is far-reaching and is considered a serious offense against the university. All incidences of academic misconduct will be reported to the Office of Academic Integrity and could result in outcomes such as failure on the assignment, failure in the course, suspension, or even expulsion from the university.

For more information about academic integrity see the student handbook or the Office of Academic Integrity's website, and university policies on Research and Scholarship Misconduct.

Please ask your instructor if you are unsure what constitutes unauthorized assistance on an exam or assignment, or what information requires citation and/or attribution.

Students and Disability Accommodations:

USC welcomes students with disabilities into all of the University's educational programs. The Office of Student Accessibility Services (OSAS) is responsible for the determination of appropriate accommodations for students who encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each course. The LOA must be given to each course instructor by the student and followed up with a discussion. This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at osas.usc.edu. You may contact OSAS at (213) 740-0776 or via email at osasfrontdesk@usc.edu.

Support Systems:

Counseling and Mental Health - *(213) 740-9355 – 24/7 on call*
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

988 Suicide and Crisis Lifeline - *988 for both calls and text messages – 24/7 on call*
The 988 Suicide and Crisis Lifeline (formerly known as the National Suicide Prevention Lifeline) provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week, across the United States. The Lifeline is comprised of a national network of over 200 local crisis centers, combining custom local care and resources with national standards and best practices. The new, shorter phone number makes it easier for people to remember and access mental health crisis services (though the previous 1 (800) 273-8255 number will continue to function indefinitely) and represents a continued commitment to those in crisis.

Relationship and Sexual Violence Prevention Services (RSVP) - *(213) 740-9355(WELL) – 24/7 on call*
Free and confidential therapy services, workshops, and training for situations related to gender- and power-based harm (including sexual assault, intimate partner violence, and stalking).

Office for Equity, Equal Opportunity, and Title IX (EEO-TIX) - *(213) 740-5086*
Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

Reporting Incidents of Bias or Harassment - *(213) 740-5086 or (213) 821-8298*
Avenue to report incidents of bias, hate crimes, and microaggressions to the Office for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

The Office of Student Accessibility Services (OSAS) - *(213) 740-0776*
OSAS ensures equal access for students with disabilities through providing academic accommodations and auxiliary aids in accordance with federal laws and university policy.

USC Campus Support and Intervention - *(213) 740-0411*
Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity, Equity and Inclusion - *(213) 740-2101*
Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - *UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call*
Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - *UPC: (213) 740-6000, HSC: (323) 442-1200 – 24/7 on call*
Non-emergency assistance or information.

Office of the Ombuds - *(213) 821-9556 (UPC) / (323-442-0382 (HSC)*
A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.

Occupational Therapy Faculty Practice - *(323) 442-2850 or* otfp@med.usc.edu
Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.