INF 559: Introduction to Data Management
Units: 3

Spring, 2018
T,Th 3:30-4:50PM
Location: KAP 163

Instructor: Prof. Carl Kesselman
Office: Michelson 304
Office Hours: Tuesday 1:00PM-3:00PM
Contact Info: carl@isi.edu

TA: TBD
Office Hours: TBD
Contact Info: TBD

Classroom:

Instructor	**Carl Kesselman**
Office	**Michelson 210A**
Phone	**(310) 448-9338**
E-mail	**carl@isi.edu**
Office Hours	**Tuesday 1:00 PM to 3:00 PM**
Teaching Assistant:	
E-Mail:	
Office and Office Hours:	

Catalog 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.

Course Description

This course is one of the introductory courses in the Informatics program. It prepares the students with the fundamental knowledge on the data management. Such knowledge is critical for the students to succeed in more advanced data management courses in the program. It also exposes students to the cutting-edge data management concepts, systems, and techniques for managing large

scale of data, to ensure that students have adequate background for further exploring big data analytics in follow-up courses.

The course may be divided into three parts. (1) Fundamental of data management: data storage, file system, file format, relational data vs. semi-structured data such as XML and JSON, conceptual modeling, relational modeling, relational algebra, SQL, views, constraints, and query processing. (2) Big data analytics: NoSQL, key-value and document stores, cloud data storage, distributed file system, and; and (3) Advanced topics in data management: data warehousing, data cleaning, and data integration and metadata modeling.

The course will also provide students with hand-on experiences on RDBMS, e.g., PostgreSQL, cloud data storage, e.g., Amazon S3, and NoSQL databases such as Amazon DynamoDB.

Recommended Preparation

Basic understanding of engineering principles, including basic programming skills, knowledge of operating systems, networks, and databases; familiarity with the Python (preferably Java too) programming language is desired.

Students are expected to have their own laptop or desktop computer for doing class assignments.

Class participation

Students are expected to be in class on-time and participate in class discussions. There are weekly readings and students are encouraged to do the readings prior to the discussion in class. All of the course materials, including the readings, lecture slides, home works will be posted online

Grading

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 rubric for each assignment will be provided for the assignment.

Weekly quizzes: There will be weekly quizzes, typically based on the lectures in the past week.

Final Exam: There is a final exam at the end of the semester covering all of the material covered in the class.

Lab sessions: Hand-on exercises on database and big data software.

Student presentation: Students may also be expected to present research or application papers on subjects related to class materials.

Grade breakdown:

Homework	25%
Quizzes	20%
Midterm	25%
Labs	5%
Final	25%

Grading Policy

Homework assignments are due at 11:59pm on the due date and should be submitted in Blackboard. Late homework will be deducted 10% of its points for every 24 hours that it is late. No credit will be given after 72 hours of its due time.

Makeup for quizzes and exams are not permitted unless there are medical emergencies. Doctor notes are needed as proof. Typically no makeups will be given for situations such as interview, job fairs, etc. Students are responsible for scheduling to avoid conflicts with class meeting times and for any missing coursework due to these situations.

Homework, quizzes, and midterm exam regrading requests must be made within a week after the solutions have been posted. Grades are final after the regrading period.

Course Schedule

Date	Class	Assignment
Week 1 (1/8)	Introduction, Computer Systems, Notation and Models	
Week 2 (1/15)	Python Review	Homework 1
Week 3 (1/22)	Storage systems	
Week 4 (1/29)	File systems, network file systems	Homework 1 Due Homework 2 Assigned
Week 5 (2/5)	File formats, data encoding	
Week 6	Data modeling	Homework 2

(2/12)		
Week 7 (2/19)	Midterm, Relational databases	
Week 8 (2/26)	SQL	Homework 2 Due Homework 3 Assigned
Week 9 (3/5)	SQL continued, Constrains, normalization, data cleaning, ETL and integration	Homework 3 Due
Week 10 (3/12)	Spring Break	
Week 11 (3/19)	Data Warehousing	Homework 4 Assigned
Week 12 (3/26)	Cloud storage and NoSQL Databases	
Week 13 (4/2)	Metadata	Homework 4 Due Homework 5 Assigned
Week 14 (4/9)	Identifiers and Content Management	
Week 15 (4/16)	No class	
Week 16	Final Exam: Tuesday May 8th, 2-4PM	

Related Administrative Information

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. Website and contact information for DSP: http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html, (213) 740-0776 (Phone), (213) 740-6948 (TDD only), (213) 740-8216 (FAX) ability@usc.edu.

Emergency Preparedness/Course Continuity in a Crisis
In case of a declared emergency if travel to campus is not feasible, USC executive leadership will announce an electronic way for instructors to teach students in their residence halls or homes using a combination of Blackboard, teleconferencing, and other technologies.

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 Section 11, *Behavior Violating University Standards*https://scampus.usc.edu/1100-behavior-violating-university-standards-and-appropriate-sanctions/. 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/.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the *Office of Equity and Diversity* http://equity.usc.edu/ or to the *Department of Public Safety* http://capsnet.usc.edu/department/department-public-safety/online-forms/contact-us. This is important for the safety whole USC community. Another member of the university community – such as a friend, classmate, advisor, or faculty member – can help initiate the report, or can initiate the report on behalf of another person. *The Center for Women and Men* http://www.usc.edu/student-affairs/cwm/ provides 24/7 confidential support, and the sexual assault resource center webpage sarc@usc.edu describes reporting options and other resources.

Support Systems

A number of USC's schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the *American Language Institute* http://dornsife.usc.edu/ali, which sponsors courses and workshops specifically for international graduate students. *The Office of Disability Services and Programs* http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html provides certification for students with disabilities and helps arrange the relevant accommodations. If an officially declared emergency makes travel to campus infeasible, *USC Emergency Information* *http://emergency.usc.edu/*will provide safety and other updates, including ways in which instruction will be continued by means of blackboard, teleconferencing, and other technology.