



ISE 558: Data Management for Analytics
4 Units

Day/Time: Tuesday/Thursday 4:00PM – 5:50PM

Location: THH 301

Instructor: Carl F Kesselman

Office: MCB 304

Office Hours:

See Piazza for weekly updates and CP availability

Contact Info:

Please use Piazza for all course communications

Email: carl@isi.edu

Course Producers:

Shree Vaishnavi Bacha (sbacha@usc.edu)

Xinrui Li (xli71479@usc.edu)

Catalog Description

Introduction to technologies and disciplines used to collect, integrate, clean, and otherwise prepare data into a well-structured, controlled, documented, and understood analytic-ready dataset.

Learning Objectives and Outcomes

- Develop an advanced level of proficiency with all the activities associated with managing data with an emphasis on data preparation for analytics.
- Primary focus areas of the class are developing proficiency in the use of Structured Query Language (SQL), data modeling, data warehousing, big data and cloud data engineering, and preparing and structuring data for subsequent analytical modeling.

Class Delivery Mode: This class will be conducted in a fully in-person mode. Lectures will be broadcast and recorded using Zoom to accommodate students who are ill and those who wish to re-view portions of the lecture after the class. The mid-term and final exams must be taken in person.

Prerequisite(s): None

Recommended Preparation: Python programming on the level of ISE 150.

Course Notes: All course materials (PowerPoints, assigned readings, etc.) will be distributed via Blackboard.

Technological Proficiency and Hardware/Software Required

The SQL and ETL modules will make significant use of the SQL database language. We will be using a modern cloud-based database/data warehouse environment from Snowflake for this purpose. Students will be provided with free accounts at the beginning of the SQL module. The Semi-Structured data, ETL, and Data Preparation for

Analytics modules will also involve use of the Python programming language and its two key data science packages (NumPy and Pandas) in the Jupyter Notebook development environment. These are all open source and can be downloaded by the student for no cost.

Required Readings and Supplementary Materials

There are no mandatory texts for this class. It is intended that the PowerPoint materials presented in class cover all the content required. However, students are encouraged to consult the texts that the material is based on for clarification and elaboration.

The material in the course is primarily drawn from the following texts:

- Beaulieu, *Learning SQL: Generate, Manipulate, and Retrieve Data*, 3rd Edition, O'Reilly, 2020, (LSQL)
- Kimball & Ross, *The Data Warehouse Toolkit*, Wiley, 2013 (DWT)
- Ponniah, *Data Modeling Fundamentals*, Wiley, 2007 (DMF)
- Adamson, *Star Schema, The Complete Reference*, Mc-Graw-Hill, 2010 (SS)
- Svolba, *Data Preparation for Analytics Using SAS*, SAS Press Series, 2006 (DPA)
- Deka, *NoSQL: Database for Storage and Retrieval of Data in the Cloud*, CRC Press, 2017 (NoSQL) ISBN: 9781498784368 (NoSQL)
- Kimball, *The Data Warehouse ETL Toolkit*, Wiley, 2004, (DWETL)
- Additional resources to be announced.

Grading Breakdown

Grading will be based on four primary components:

- 10-12 homework assignments (approx. one per week) - 50% of final grade
- Mid-term exam (in class) – 20% of final grade
- Final exam - 30% of final grade

Mid-term and final exams will be held in-class and cannot be taken remotely. They will be closed book with one page of notes permitted.

Grading Scale

Course final grades will be determined using the following scale

A	95-100
A-	90-94
B+	87-89
B	83-86
B-	80-82
C+	77-79
C	73-76
C-	70-72
D+	67-69
D	63-66
D-	60-62
F	59 and below

Borderline averages between two letter grades may be rounded up based on class engagement at the instructor's discretion.

Assignment Submission Policy, Timelines, and Rules for Submission

- Assignments will be posted on Blackboard and submitted for grading on GradeScope (student instructions will be provided)
- Homework assignments will generally be posted on or shortly after the last class of each week be due one week after posted.
- Late submissions are accepted for 48 hours after the due date and will incur a 10% penalty.
- No submissions will be accepted after 48 hours past the due date and assignments not submitted will receive 0 credit.
- The lowest homework grade for the semester will be dropped from the final grade computation.
- Regrade requests are accepted through Gradescope for one week after grades are published.

Course Communications

- All materials will be uploaded to Blackboard
- Assignments will be submitted through Gradescope
- We will use Piazza as the primary communications mechanism
 - Class announcements will be posted there, and we request that any questions you have be posted there so that other students can benefit from your question and responses from the instructors, TAs, and hopefully other students
 - Students who actively post responses to questions MAY receive extra credit (which could result in an increase by one letter grade in borderline cases)
- I will periodically post "discussion questions" on Piazza. Class engagement credit can be earned by participating in these online discussions

Course Schedule: A Weekly Breakdown

Week	W/E	Topics/Daily Activities	Assignments	References
1	1/12	Module 1. Introduction	SQL and Python tutorials assigned	
2	1/19	Module 2. Relational Database Design (Data Modeling)	SQL tutorials due Module 2A HW assigned	DMF Chapters 3-6
3	1/26		Module 2A HW due Module 2B HW assigned	
4	2/2	Module 3. Relational Database Programming (SQL)	Module 2B HW due Module 3A HW assigned	LSQL Chapters 3-10
5	2/9		Module 3A HW due Module 3B HW assigned	
6	2/16	Module 4. Semi-Structured Data (JSON and XML)	Python tutorials due Module 3B HW due Module 4 HW assigned	NoSQL Chapters 3-5
7	2/22	Mid-Term (in-class)	Module 4 HW due	
8	3/1	Module 5. Data Warehousing	Module 5A HW assigned	DWT Chapters 2-3
9	3/8		Module 5A HW due Module 5B HW assigned	SS Chapters 2,3, 6-11
10	3/22	Module 6. ETL, Data Pipelines, and Workflow Management	Module 5B HW due Module 6A HW assigned	DWETL Chapters 5-6
11	3/29		Module 6A HW due Module 6B HW assigned	
12	4/5	Module 7. Big Data Technologies and Cloud Data Engineering	Module 6B HW due Module 7 HW assigned	Lecture handouts
13	4/12	Module 8 – Data Preparation for Analytics	Module 7 HW due Module 8A HW assigned	DPA Chapters 7-12
14	4/19		Module 8A HW due Module 8B HW assigned	
15	4/26	Review/final exam preparation	Module 8B HW due	
		Final exam May 7, 2:00PM – 4:00PM		

Notes:

- This is a “generic” schedule that will be tailored to the dates of the specific semester being offered. The syllabus for the specific semester will be distributed shortly before the first class.

Statement on Academic Conduct and Support Systems

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, 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 of the 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 <http://sarc.usc.edu> describes reporting options and other resources.

Support Systems:

Student Health Counseling Services - (213) 740-7711 – 24/7 on call
engemannshc.usc.edu/counseling

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call
suicidepreventionlifeline.org

Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-4900 – 24/7 on call
engemannshc.usc.edu/rsvp

Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office of Equity and Diversity (OED) | Title IX - (213) 740-5086
equity.usc.edu, titleix.usc.edu

Information about how to get help or help a survivor of harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants. The university prohibits discrimination or harassment based on the following protected characteristics: race, color, national origin, ancestry, religion, sex, gender, gender identity, gender expression, sexual orientation, age, physical disability, medical condition, mental disability, marital status, pregnancy, veteran status, genetic information, and any other characteristic which may be specified in applicable laws and governmental regulations.

Bias Assessment Response and Support - (213) 740-2421
studentaffairs.usc.edu/bias-assessment-response-support

Avenue to report incidents of bias, hate crimes, and microaggressions for appropriate investigation and response.

The Office of Disability Services and Programs - (213) 740-0776

dsp.usc.edu

Support and accommodations for students with disabilities. Services include assistance in providing readers/notetakers/interpreters, special accommodations for test taking needs, assistance with architectural barriers, assistive technology, and support for individual needs.

USC Support and Advocacy - (213) 821-4710

studentaffairs.usc.edu/ssa

Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity at USC - (213) 740-2101

diversity.usc.edu

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

dps.usc.edu, emergency.usc.edu

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) 442-120 – 24/7 on call

dps.usc.edu

Non-emergency assistance or information.