



ISE 559: Introduction to Data Management
3 Units

Day/Time: TBA

Location: TBA

Instructor: Bruce Wilcox

Office: GER 203

Office Hours: TBA

Contact Info: brucewil@usc.edu

Teaching Assistants:

TBA

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, and DataMart implementation (data integration, cleaning, reduction, and enhancement).
- Gain an understanding of cost performance alternative for data storage and data management.
- Understand and apply data management techniques to improve the quality and correctness of data driven decisions based on analytic results.

Class Delivery Mode: This class will be conducted in a hybrid mode. While the primary mode is in-person lectures, it is possible to take this class remotely via Zoom. However, the mid-term and final exams must be taken in person.

Prerequisite(s): None

Recommended Preparation: ISE 225 (Engineering Statistics I) or equivalent, working knowledge of Python.

Course Notes

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

Technological Proficiency and Hardware/Software Required

This class will use the SQLite relational database management system and the SQLiteStudio developer environment for SQL programming exercises. This is a free open-source system that will be installed on the student's computer (www.sqlitestudio.pl). It will also involve some limited use of the Python programming language and two of its key data science packages (NumPy and Pandas). 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)
- Additional resources to be announced

Course Overview and Schedule

The course is structured into modules that correspond to the lifecycle of data starting from the source data structures and formats through the process of extracting the data and consolidating it into data warehouses, and creating analytic DataMarts suitable for subsequent analytics activities. Readings will be assigned prior to each module.

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 (covering Modules 1 – 4)
- Final exam - 30% of final grade

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

Up to three points may be added to the overall grade based on class engagement.

Assignment Submission Policy

Assignments will be posted on Blackboard and submitted for grading on GradeScope (student instructions will be provided)

Timeline and Rules for Submission

Assignments will be accepted late for two days with a 10% penalty and will not be accepted beyond that time. The lowest homework grade will be dropped.

Course Schedule: A Weekly Breakdown

Week	W/E	Topics/Daily Activities	Assignments	References
1		Module 1. Data Structures Overview Files, structured data, semi-structured, performance metrics and analysis	Module 1 HW assigned	
2		Module 2. Relational Database Design RDMBS overview, data modeling, entity-relationship diagrams	Module 1 HW due Module 2A HW assigned	DMF
3			Module 2A HW due Module 2B HW assigned	
4			Module 2B HW due Module 2C HW assigned	
5		Module 3. Relational Database Programming Relational algebra, SQLite, SQL/DQL	Module 2C HW due	LSQL
6			Module 3 HW assigned	
7		Mid-Term (in-class)	Module 3 HW due	
8		Module 4. Data Warehousing Architecture, dimensional data modeling, star schema development, SQL/DDDL & DML	Module 4A HW assigned	DWT, SS
9			Module 4A HW due Module 4B HW assigned	
10			Module 4B HW due Module 4C HW assigned	
11		Module 5 – Data Preparation for Analytics Python/Pandas, ETL programming, analytic dataset structures, data quality assessment and cleansing, dimensionality reduction, feature engineering.	Module 4C HW due Module 5A HW assigned	DPA
12			Module 5A HW due Module 5B HW assigned	TBA
13		Review/final exam preparation	Module 5B HW due	
		Final Exam (per USC published schedule)		

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.
- This schedule is based on a 13-week semester (which is generally the shortest semester schedule at USC). Semesters with additional weeks will have expanded materials on module 5.

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/ssu

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.