San José State University
Department of Computer Science

CS/SE 157B Database Management Systems II

Section 3
Spring 2018

Course and contact information

Instructor: Ron Mak

Office Location: ENG 250

Email: ron.mak@sjsu.edu

Website: http://www.cs.sjsu.edu/~mak/

Office Hours: TuTh 3:00 - 4:00 PM

Class Days/Time: TuTh 9:00 - 10:15 AM

Classroom: MH 233

Prerequisites: CS 157A (with a grade of "C-" or better); Computer Science, Applied and Computational Math, or Software Engineering Majors only; or instructor consent. *The department strictly enforces course prerequisites.*

Course format

This course will be taught primarily face-to-face instruction. Course materials, syllabus, assignments, grading criteria, exams, and other information will be posted on the SJSU Canvas course site at http://sjsu.instructure.com/ You are responsible to check Canvas regularly for class work and exams. You also can find Canvas video tutorials and documentations at http://ges.sjsu.edu/canvas-students

Faculty web page and MySJSU messaging

Course materials such as syllabus, handouts, notes, assignment instructions, etc. can be found on my faculty web page at http://www.sjsu.edu/people/firstname.lastname and/or on Canvas Learning Management System course login website at http://sjsu.instructure.com. You are responsible for regularly checking with the messaging system through MySJSU at http://my.sjsu.edu to learn of any updates.

Course website

Course materials such as syllabus, handouts, notes, assignment instructions, etc. can be found on the class web page at http://www.cs.sjsu.edu/~mak/CS157B/.

Course catalog description

Survey course. Object-oriented data model, definition language, query language. Object relational database systems. Database trends like active, temporal, multimedia, deductive databases. Web database topics, namely, architectures, introduction to interface languages. Team projects.

Instructor's description

This course will concentrate on practical aspects of database management systems as currently practiced by industry:

- relational and NoSQL databases for web applications
- data warehousing and online analytical processing (OLAP)
- XML data processing
- data virtualization and data federation servers

The emphasis will be on acquiring *job skills that employers want* for new hires in data management. Gain confidence and proficiency in developing enterprise-class transactional and analytical applications that use back-end data management. The course will also introduce some key data science topics, including Big Data, cloud computing, and data mining.

This is a major programming class and a Computer Science Department *deep course***.**

Course learning outcomes

Upon successful completion of this course, students will be able to:

CLO 1: Incorporate relational and NoSQL databases into web applications.
CLO 2: Design and deploy operational databases for OLTP (online transaction processing).
CLO 3: Design and deploy analytical databases for OLAP (online analytical processing).
CLO 4: Implement dimensional modeling, star schemas, and ETL (extract-transform-load) for a data warehouse.
CLO 5: Understand semi-structured data and know how to process XML data.
CLO 6: Understand basic concepts of data virtualization and know how to program a commercial data federation server.
CLO 7: Understand some key concepts of data science.

You will develop the *critical job skill* of working in a small project team.

Recommended texts

Title:	**Database Systems: Introduction to Databases and Data Warehouses**
Authors:	Nenad Jukic, Susan Vrbsky, and Svetlozar Nestorov
Publisher:	Prospect Press, 2017
Paperback:	978-1-943153-19-0 available from Redshelf.com
eTextbook:	978-1-943153-18-3 available from Redshelf.com and VitalSource.com *These are much less expensive versions of the textbook originally published in 2014 by Pearson. However, diagrams are no longer in color. Pearson ISBN: 978-0-13-257567-6*
Title:	**The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence, 2nd edition**
Authors:	Ralph Kimball and Margy Ross
Publisher:	Wiley, 2015
ISBN:	978-1119216315
Title:	**Database Systems:** **The Complete Book, 2nd edition**
Authors:	Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Windom
Publisher:	Pearson Prentice Hall, 2009
ISBN:	978-0-13-187325-4

Software to install

For relational databases and PHP, download and install XAMPP (https://www.apachefriends.org/index.html), which will install and configure on your Mac, Windows, or Linux platform the following software packages:

- PHP
- Apache web server
- MariaDB database server (compatible with MySQL)

You can download and install these packages separately, but then you'll have the hassle of configuring them to work nicely together.

For NoSQL and the MEAN stack, download and install:

- Node.js: https://nodejs.org/en/
- Express: download and install using the node package manager (npm)
- MongoDB: https://docs.mongodb.com/master/administration/install-community/

Later in the semester, you will also download and install the Cisco Information Server (CIS) Studio application.

Course requirements and assignments

You will form project teams of three students each. *Team membership is mandatory for this class.* The teams will last throughout the semester. Once the teams are formed, you will not be allowed to move from one team to another, so form your teams wisely!

Student teams will work on a series of short project assignments during the semester. Each project will each take one to three weeks. *Each student on a team will receive the same score for each team project assignment.*

Each project will develop a working application that incorporates some assigned data management technologies (NoSQL, data warehousing, data virtualization, etc.) using data that the team can generate or download from the Internet. Each project will include a short report that describes the application, its data, and how it uses the technologies.

Each team will submit its projects into Canvas, where the rubric for scoring each project will be displayed.

Each project will be worth up to 100 points. Late projects will lose 20 points and an additional 20 points for each 24 hours after the due date.

This is a challenging course that will demand much of your time and effort throughout the semester.

The university's syllabus policies:

- University Syllabus Policy S16-9 at http://www.sjsu.edu/senate/docs/S16-9.pdf.
- Office of Graduate and Undergraduate Programs' Syllabus Information web page at http://www.sjsu.edu/gup/syllabusinfo/

The university's credit hour requirement:

"Success in this course is based on the expectation that students will spend, for each unit of credit, a minimum of 45 hours over the length of the course (normally 3 hours per unit per week with 1 of the hours used for lecture) for instruction or preparation/studying or course related activities including but not limited to internships, labs, clinical practica. Other course structures will have equivalent workload expectations as described in the syllabus."

Exams

The midterm and final examinations will be closed book. Instant messaging, e-mails, texting, tweeting, file sharing, or any other forms of communication with anyone else during the exams will be strictly forbidden.

There can be no make-up midterm examination unless there is a documented medical emergency. Make-up final examinations are available only under conditions dictated by University regulations.

Postmortem report

At the end of the semester, each student must also submit into Canvas in a short (1- or 2-page) **individual postmortem report** that includes:

- A brief description of what you learned in the course.
- An assessment of your accomplishments for your project team on the assignments and the compiler project.
- An assessment of each of your other project team members.

Only the instructor will see these reports. How your teammates evaluate you will affect your class grade.

Class grade

Your individual final class grade will be weighted as follows:

65%	Project assignments*
15%	Midterm exam**
20%	Final exam**

** project team scores*
*** individual scores*

During the semester, you can keep track of your individual progress in Canvas. Each project assignment and exam will be scored (given points) but not assigned a letter grade. The average score can be seen in Canvas after each assignment and exam.

At the end of the semester, all the students will be ranked in the order of their weighted class scores. Students with scores near the median score will be assigned the B grade. Higher and lower grades, respectively, will then be assigned above and below the median.

Your final class grade can be adjusted up or down depending on your level and quality of participation on your project team as determined by your teammates' assessments of your performance.

All students have the right, within a reasonable time, to know their academic scores, to review their grade-dependent work, and to be provided with explanations for the determination of their course grades." See University Policy F13-1 at http://www.sjsu.edu/senate/docs/F13-1.pdf for more details.

Classroom protocol

It is very important for each student to attend classes and to participate. Cell phones in silent mode, please.

University policies

Per University Policy S16-9, university-wide policy information relevant to all courses, such as academic integrity, accommodations, etc. will be available on Office of Graduate and Undergraduate Programs' Syllabus Information web page at http://www.sjsu.edu/gup/syllabusinfo/"

CS/SE 157B Database Management Systems II

Section 3
Spring 2018

Course schedule

Subject to change with fair notice.

Week	Date	Topics
1	Jan 25	Overview of the course Data modeling *Form programming teams*
2	Jan 30 Feb 1	Conceptual models and entity-relationship diagrams Logical models and relational schemas Mapping ER diagrams to relational schemas Physical data models phpMyAdmin
3	Feb 6 Feb 8	MySQL Workbench Update anomalies Functional dependencies and normalization Entity and referential integrity constraints
4	Feb 13 Feb 15	Relational databases and web applications Basic HTML and PHP
5	Feb 20 Feb 22	SQL injection attacks PHP prepared statements Object-relational mapping (ORM) PHP Data Objects (PDO)
6	Feb 27 Mar 1	Data warehousing Dimensional modeling and star schemas Dimension tables and fact tables
7	Mar 6 Mar 8	Operational databases and online transaction processing (OLTP) Analytical databases and online analytical processing (OLAP) Extract-transform-load (ETL) Content management and WordPress
8	Mar 13 Mar 15	*Midcourse review* ***Midterm exam Thursday, March 15***
9	Mar 20 Mar 22	NoSQL databases and web applications MongoDB Documents and collections CAP theorem vs. ACID
	Mar 26-30	Spring break

Week	Date	Topics
10	Apr 3 Apr 5	The Express server-side framework Database CRUD actions and HTTP verbs The REST API and RESTful web services
11	Apr 10 Apr 12	Semi-structured data and XML Oxygen XML Editor XPath and XQuery FLWOR expressions
12	Apr 17 Apr 19	Data virtualization The Cisco Information Server
13	Apr 24 Apr 26	Query optimization Database failure and recovery RAID
14	May 1 May 3	Distributed databases Object databases Cloud computing
15	May 8 May 10	Data science Data mining Big Data, Hadoop, and MapReduce *Course review*
Final Exam	Wednesday May 16	Time: 7:15 – 9:30 **AM** Room: MH 233

9	Apr 4	Operational vs. analytical databases Data warehousing Extract-transform-load (ETL) Data marts Dimensional modeling and star schemas Fact tables and dimension tables	7
10	Apr 11	Types of fact tables Fact table granularity Slowly changing dimensions Data warehouse architectures Online transaction processing (OLTP) Online analytical processing (OLAP) Business intelligence (BI) tools	8, 9
11	Apr 18	Data virtualization Cisco Information Server (CIS) Semi-structured data XML XML Schema	Appendix H
12	Apr 25	XPath and XQuery FLWOR expressions NoSQL databases	Appendix I
13	May 2	Object databases Distributed databases Cloud computing Data mining Big data MapReduce	Appendix E Appendix F Appendix G Appendix J
14	May 9	Project presentations	
15	May 16	Project presentations	
	May 17	***Final projects due Wednesday, May 17***	
Final	May 23	***Final exam Tuesday, May 23*** 5:15 - 7:30 PM, DMH 227	