

287 Database Management with SQL

University of California, Davis
Graduate School of Management

Professor Yinghui (Catherine) Yang
yiyang@ucdavis.edu

Monday 6-9pm, 10/2, 10/9, 10/16, 10/23, 10/30, 11/6, 11/13, 11/20, 11/27, 12/4

Course Description

This course aims to provide a practical introduction to the fundamental principles of database management systems. After taking this course, students will be able to transform daily business activities into a database system from which information can be extracted, write SQL queries to extract information from the database, and understand some concepts of database marketing. Students will design and deploy a database solution using Microsoft SQL Server Express. A database containing purchase transactions will be available for students to design queries to answer business questions. Students will also learn how to connect databases to the data visualization tool Tableau and visualize database query results within Tableau.

Intended Audience and Prerequisites

The course is designed mainly for students who have no or very limited database background and are eager to learn how to set up his/her own database from scratch, write SQL queries to answer questions. No prerequisite is required for taking this course.

Software:

We will use the Microsoft SQL Server (the free Express edition) to build our databases and execute SQL queries. We will also use Tableau to visualize the results of the database queries. No programming skill is needed.

IMPORTANT: Every student **MUST** have a laptop with **Windows OS** before the third class. Microsoft SQL Server does not support Mac OS. Also, it's better to have a PC laptop running Windows instead of installing Windows on Mac which could have issues.

Textbook:

- Required:

SQL in 10 Minutes, Sams Teach Yourself (4th Edition)

By Ben Forta

Publisher: Sams Publishing

ISBN-10: 0672336073

ISBN-13: 978-0672336072

- Optional Book if you are interested:

Data Analysis Using SQL and Excel (2nd Edition) (More advanced SQL applications)

Gordon S. Linoff

Publisher: Wiley, December 14, 2015

ISBN-10: 111902143X

ISBN-13: 978-1119021438

For Tableau, it's more convenient to follow online videos provided by Tableau

(<https://www.tableau.com/learn/training>).

Grading:

| Components | Grades |
|--|------------|
| Class Participation | 6% |
| 7 Homework Assignments | 62% |
| Term Project (Phase 1 report: 5%; Meeting: 4%; Final report: 18%; Presentation/poster: 5%) | 32% |

Canvas: All materials I need to hand out to you will be distributed via Canvas. All deliverables need to be submitted via Canvas.

Homework Assignments: There are 7 homework assignments. Because the solution to the homework will be posted on Canvas or discussed in class on the due date, late homework will not be accepted.

Groups: I will post a sign-up sheet for you to self-select your group (in Homework 1). A group will work together on the term project.

Term Project

For the term project, you should select a comprehensive data set, develop a database for it, write SQL queries to answer questions based on the database, and visualize the data to show insights.

The project has 2 phases.

Phase 1: Choose your dataset, Design the database.

1, Pick a comprehensive dataset for your database. The data set you choose should be able to be mapped into at least 4 tables in the database. In your Phase 1 report, include the following information about the dataset:

- Where did you find the data? If you downloaded the data from the Internet, please provide the link to the dataset description. A site to find quality datasets is Kaggle.com.
- How many data files are available for download? What are the relationships among the data files?
- For each data file you plan to use to populate your database, list the attributes you plan to use, and include 3 sample data rows.
- List 5 meaningful questions you can answer based on the dataset.

2, Discuss how to convert your dataset to the tables in your database. If needed, you can add external data sources such as calendar and zipcode data. While you do not need to draw the actual Entity Relationship Diagram for your database design, you should use the principles of Entity Relationship Diagram to guide the process of creating well-structured tables. In your Phase 1 report, include the following information:

- The table structure: Name of the table, and the attributes in each table; identify the primary key attribute of each table. Describe how the tables are connected to each other (e.g. Foreign key + primary key pairs).
- Discuss the process of converting the original data files to the database tables. For example, did you convert each data file to a single database table? Did you break one data file into multiple tables in the database?

Phase 2: Implementation in Microsoft SQL Server, Project meeting with me, Final report and class presentation/poster.

1, Properly refine your phase 1 according to my suggestions. You are allowed to make other changes to phase 1 if desirable.

2, Implement the tables in Microsoft SQL Server.

3, Populate the tables with your dataset.

4, You are required to design queries to address 5 problems (or answer 5 questions). Some of the problems can be addressed with a single query, and some will need to be addressed with a series of queries. Implement all the queries in Microsoft SQL Server. You will be graded based on the complexity of the problems/queries, and the insights you can obtain from the queries.

5, Use Tableau to visualize some query results.

6, Meet with me to discuss your progress.

7, Submit final project report. The final report should properly integrate all the pieces you have done. In the report as well as the class presentation/poster, you should describe the business problem you are addressing, the challenges you encountered, the design, and the insights behind the queries.

8, Class presentation.

Class Schedule:

| Date | Topic |
|-------------|---|
| 1 (Oct. 2) | Introduction |
| 2 (Oct. 9) | Database Design 1 |
| 3 (Oct. 16) | Database Design 2 & SQL 1 |
| 4 (Oct. 23) | SQL 2 – multiple tables |
| 5 (Oct. 30) | SQL 3 – external tables (location&time) |
| 6 (Nov. 6) | SQL 4 – window functions |
| 7 (Nov. 13) | Database marketing via queries – RFM Segmentation |
| 8 (Nov. 20) | Database marketing via queries – Building Customer Signatures & Query Designer(video lecture, project meetings with me) |
| 9 (Nov. 27) | Data Visualization using Tableau (video lecture, project meetings with me) |
| 10 (Dec. 4) | Term project presentation |

The content of Lectures 8&9 is delivered via videos. During the class time of Lectures 8&9, I will meet with individual groups to discuss the details of the group project.

Due Dates:

| | Due Date |
|--|--------------------|
| Homework 1 (2 points) | Oct. 9 |
| Homework 2 (10 points) | Oct. 16 |
| Homework 3 (8 points) | Oct. 23 |
| Homework 4 (10 points) | Oct. 30 |
| Project Phase 1 (5 points) | Nov. 6 |
| Homework 5 (11 points) cover Lectures 5&6 | Nov. 13 |
| Homework 6 (10 points) cover Lecture 7 | Nov. 20 |
| Project meeting with me (4 points) | Nov. 20 or Nov. 27 |
| Homework 7 (11 points) cover Lectures 8,9 video lectures | Dec. 4 |
| Presentation (5 points) | Dec. 4 |
| Project Final Report (18 points) | Dec. 9 |

Code of Academic Conduct: <http://sja.ucdavis.edu/files/cac.pdf>