70-455 Modern Data Management (Spring 2013)

Lectures: Time & Place

Tue & Thu, 12:00 pm – 1:20 pm, PH A19C (Porter Hall basement)

Lecturer

Wolfgang Gatterbauer
Assistant Professor in Information Systems
Office: Room 354 (Posner Hall)
Email: gatt@cmu.edu, Web presence: http://gatterbauer.name

Class Admin

Online Class Management: We will use Blackboard (http://blackboard.cmu.edu) as official repository for grades and for submitting assignments. Yet, we will use Piazza (http://piazza.com/cmu/spring2013/70455) instead for everything else (lecture slides, current class calendar, readings, student's solutions to small exercises, etc.). Piazza allows a more flexible class interaction. The access code for Piazza will be distributed via Blackboard or email.

Support: You can contact me anytime by email. Yet I'd prefer if you ask questions of general interest on Piazza, first because one of your classmates may know the answer before I can respond, and second because the question and its answer may also be helpful for others.

Office Hours: Standing office hours are Mon 3:30 pm – 4:30 pm, @354 Posner. If you like to see me at other times, feel free to send me an email with alternative day suggestions.

Textbooks: Books 1 and 2 below are required, book 3 is optional but highly recommended:

1. "Learn Excel 2010 Expert Skills with The Smart Method" by Mike Smart.
   http://www.amazon.com/Learn-Excel-Expert-Skills-Method/dp/0955459982
2. "Modern Database Management (10th ed)" by Jeffrey Hoffer, R. Venkataraman, H. Topi.
   http://www.amazon.com/Modern-Database-Management-10th-Edition/dp/0136088392
   http://www.amazon.com/Modern-Database-Management-Jeffrey-Hoffer/dp/1408264315/
3. "The Say It With Charts Complete Toolkit" by Gene Zelazny.
   www.amazon.com/Say-Charts-Complete-Toolkit/dp/0071474706/

Equipment: There will be a large number of in-class, hands-on exercises throughout the course. Please bring your laptops to class and have Microsoft Excel, Microsoft PowerPoint, Mozilla Firefox browser, and a text processing system (e.g. Word) installed. Please also bring pen and paper for each class.

1. COURSE DESCRIPTION

The goal of this course is to learn how to manage data for making critical business decisions. The notion of "Data Management" here includes both the analysis of various sizes and types of data and their synthesis into fact-based, data-driven recommendations.

The course teaches the use of advanced functions in Excel (e.g., Pivot tables, lookup functions, array formulas), the abstraction and representation of business situations as entity relationship diagrams, the transformation of such diagrams into database schemata, and the use of SQL (Structured Query Language) to manipulate databases. The main focus will be on designing, building, and querying relational database systems. Why learn about databases? Databases are incredibly prevalent; most people every day if not every hour use their underlying technology. Databases reside behind a huge fraction of websites and are a crucial component of just about any software system or electronic device that maintains some amount of persistent information. In addition to persistence, database systems provide a number of other properties that make them exceptionally useful and convenient: reliability, efficiency, scalability, concurrency control, data abstractions, and SQL as high-level query language.

Learning Objectives

When you have successfully completed this course, you will be able to analyze data of varying sizes with varying tools and synthesize clear recommendations. In particular, you will be able to:

- use advanced data analysis functions in Excel,
- write SQL queries to retrieve information from a relational database,
- analyze and represent business situations as entity relationship diagrams,
- design a database schema from ER diagrams
- scope, setup, secure and administer a relational database management system
- apply basic query optimization and database transactions
- describe how to use data warehouses, OLAP, and reporting tools to create Decision Support Systems and Business Intelligence Systems

Prerequisites

The course assumes knowledge of an introductory programming class (e.g., 15-110) and basic knowledge of Microsoft Excel (e.g., absolute vs. relative references).

2. EVALUATION

You will acquire new skills in this class through a combination of hands-on work, readings, lectures, and exercises. Your evaluation will be based on your ability to demonstrate the skills being taught by applying them in tests, homework assignments, class preparations and discussions.

Midterm (15%) and Final (25%)	40%
Group Project	20%
Homework Assignments	20%
Class Preparation (10%) and Class Participation (10%)	20%

Midterm (15%) and Final (25%)

There will be two tests – a midterm and a final. The Midterm will be held in class and is 80min. The final exam will be 3h and held at a location and date scheduled by the HUB (TBD). Both tests will be closed book and cumulative. They may include material from any lectures, readings, and homework assignments covered up to the test date. The tests may also include a portion of exercises that need to be solved on a computer (TBD). Questions on exams are mostly short answer (several bullet points or answer in 3 or fewer sentences), but fill-in-the-blank, multiple-choice, or True/False (with a description of why one or the other is true) are also used. I rarely include essays, preferring to ask a large number and very comprehensive set of questions. This both rewards students with the broadest knowledge, and helps protect students who miss a concept here or there from suffering a huge point drop. Questions will test knowledge of cases, concepts, theory, terms, and technologies, and there must be a 'right' answer to a question. Part of the exam may include exercises to be solved on either your laptop or a lab computer.

All students are required to take the midterm and final exams at the scheduled time and place. If an emergency or significant extenuating circumstances prevent you from doing so you must contact me as soon as it is practical to do so and make alternative arrangements. Emergencies will be accommodated at my discretion and I will require documented proof of the situation. You should expect that make-up examinations will be different and more difficult than the original examination.

Group Project (20%)

The group project is a comprehensive assignment that should help you put together what you learned in the class and get a well-rounded understanding of how ER diagrams, logical database design and SQL all fit together. Each team identifies a real-world business problem and provides an information systems solution. There will be 2 intermediate deliverables before the final deliverable with a project report and a formal presentation. Project deliverables are due in class on their respective due dates. Groups can be 2 or 3 students and are assembled by the students themselves on Piazza. Details on the project and phases will be available on Piazza in time. The late policy for homeworks also applies to project deliverables.

Homework Assignments (20%)

There will be 5 homework assignments made available at Piazza over the course of the semester. These assignments are designed to provide you with hands-on experience designing, implementing, and working with databases and business intelligence tools. You may complete the assignments individually or with one other student. Two-person teams should submit a single write-up that identifies both students who worked on the assignment. You may work with different partners for different assignments, but only one partner on each assignment. You will be allowed to drop your lowest homework grade; thus, your overall homework grade will thus be proportional to the average of your four highest homework scores. Homeworks will be posted at least one week before they are due.

Late policy: Homework assignments are due the day of the deadline at 11:59pm and must be submitted via Blackboard and time stamped. I will generally accept late homework up to 3 days after their due date. You will be assessed a 33% penalty for each day that the homework is late. The 33% penalty will be imposed starting at 0:01am the day after the assignment was due. I reserve the option to disallow late homework for assignments that require me (or students) to post or otherwise present a solution to the class shortly after the due date for the assignment. Otherwise, the late homework policy will be strictly enforced.

Class Preparation (10%) and Class Participation (10%)

Preparation: As part of the preparation for classes, you will complete short exercises and post the results on Piazza. Most often, this exercise is very short and will require a bit of reflection on the reading or other preparation required before each class. I will generally incorporate your responses from Piazza into the day's lecture so all postings must be done no later than 2h before the lecture. Solutions to exercises submitted after those times are not counted. I will monitor who posts thoughtful responses by the deadline (and who does not) and may respond at times. All responses are visible to the whole class. Detailed instructions will be found in Piazza over the course of the semester (Friday for Tuesday lecture, Tuesday for Thursday lecture).

Participation: I expect you to attend each class and actively participate in discussion and in-class exercises. Your class participation grade is a combination of objective (attendance/ frequency of contribution) and subjective (involvement in class/ quality of contribution). I follow an active learning approach requiring 1) high personal introspection/ reflection/ learning via outside class readings/ preparations, and 2) high interaction during class where students share their insights. In general, the quality of your contributions to the class is much more important than the quantity. I will be happy to let you know how you are doing with participation if you stop by my office to discuss the matter, but your participation grade is assigned at my sole discretion and is completely nonnegotiable. If you find that your participation grade to date is below where you would like it to be I will be happy to work with you to figure out how to raise it for the remainder of the course.

Re-grade Policy

If you believe that your homework, final project, or exam has been incorrectly graded, please feel free to explain why you believe your answer or solution is correct. If after discussing it with me you still believe that your answer is correct, you can submit a brief written request to me to re-grade the assignment. Your request must explain why you believe your answer is correct and include the original assignment and my feedback. Upon receiving this request I will re-grade the entire assignment, not just the part in question. There is no guarantee that your grade will increase, as I may discover errors in the assignment that I missed the first time through.

I will only re-grade assignments where the correctness of an answer is in dispute. Emotional appeals for a better grade are not grounds for a re-grade request.

3. CLASS ETIQUETTE

In the interest of providing a comfortable environment for learning, I ask that you observe the following points of etiquette.

- Attendance: Much of what will be presented and discussed in class is available only, or primarily, in class. Hence, regular attendance is crucial to your success. I expect you to arrive to class on time. Coming in late disrupts and distracts the rest of the class. Likewise, I expect you to stay until the end of the class. It is not appropriate to wander in and out of the classroom during lecture. If you absolutely need to leave early then please see me before class to explain the reason and sit near the door to minimize the disruption that your departure will have on the rest of the class.
- Active Participation: Participate in class actively and come to class prepared to enter the discussion – to ask questions and provide information that will further your colleagues', and my understanding of the topic. Do not limit your role to that of student, but expand it to include teacher, trainer, guide and friend. You should think of the classroom as a laboratory in which you can test your ability to convince your peers of the correctness of your approach to complex problems and of your ability to achieve the desired results through the use of that approach. Ask questions whenever you have problem in understanding any concept. Responding to questions gives me an opportunity to either explain in more depth or offer an alternative explanation. Your questions may help you and other students in the class to understand the material more clearly. Your questions and comments in the class will be part of your grade for class participation, because all of these things contribute to the learning of class. Outside of class, you can post articles and links to news that relates to our subject to Piazza, and provide your perspectives on interesting new developments.
- Respect: Be respectful of other members of the class. We will spend time exploring ideas, expressing opinions, and trying to work through interesting problems in class that don't necessarily have clear-cut answers. Expressing strong opinions is fine but please avoid personal attacks during discussion.

- Laptops: There will be a large number of in-class, hands-on exercises throughout the course, so it is essential to bring your laptops to class. As for uses of notebooks following activities unrelated to class: I really think it is a poor use of your tuition dollars to spend time in class tracking your portfolio, surfing, checking Facebook updates, playing solitaire, etc. It also makes me a less effective teacher; like most other faculty, I lecture better if I can see your reaction to what I am saying. However, I strongly believe in your autonomy to make your own decision how to spend your time. I make two exceptions to this autonomy: 1) you need to be able to follow the class content. I may cold call students to see if they are following the class; and 2) whatever you do in class, it cannot be a distraction to other people in the class.
- Cell phones: Turn your phone off during class. Having your phone ring in class will result in a "0" grade for that day's participation. If unusual circumstances absolutely require you to keep your phone on during a class, see me before class to explain your situation.
- Nameplates: I will distribute nameplates at first class. Please bring nameplates to every class and put it in front of you. If you loose it, please make a new one (I will post a template in Word on Piazza). If I never see you in class (either because you don't show up, or show up but never make a comment, or show up and make a comment but I don't recognize you, cannot add to your class participation. Nameplates are a simple way to avoid the latter point.
- Recording: No student may record or tape any classroom activity without the express written consent of the lecturer. If a student believes that he/she is disabled and needs to record or tape classroom activities, he/she should contact the Office of Disability Resources to request an appropriate accommodation.
- Feedback: I take all student feedback very seriously and there will be ample possibilities to reach out to me with your feedback, e.g., through early course evaluations, anonymous postings on Piazza, or speaking up in class. Since we are covering a broad range of topics, the transition from topic to topic can be confusing. Don't hesitate to speak up and let me know so that I can be more responsive to your concerns. If you have any questions, don't hesitate to contact me. Constructive feedback that helps me make the class of more value to you will never be to any disadvantage for you. Very much in contrary! Feedback allows me to improve this course and how it is taught and will therefore help you improve your learning. And please note that you can always write anonymous postings to me via Piazza.
- Slides: I will post the slides for each class on Piazza after we discussed them in class.
- Preliminary Class Calendar: A detailed and preliminary schedule of lecture topics, homework and project due dates is available at the end of this document. This calendar is not set in stone and designed for change. I plan to make frequent checks on the pacing of the course and intend to make regular adjustments if necessary. If I find it necessary to modify the schedule of lectures I will post an updated calendar on Piazza as well as announcing the changes in class, but will not update this document. Thus, whenever Piazza has a different topic listed for a given day than this schedule, then Piazza beats this PDF document (compare with the date at the bottom of this document). I do not anticipate changing the dates of midterm, and when assignments and project phases are due.

4. COLLABORATION AND ACADEMIC INTEGRITY:

Statement on Academic Integrity

The university's policies on academic integrity govern the class. These policies are available at: http://www.studentaffairs.cmu.edu/acad_integ/acad_integ_text.html

Collaboration with Other Students in Class

In general, the purpose of assignments and exercises is to force you to think deeply through the material presented and apply it to solve specific problems, or to apply the tools, models, and techniques covered in class, and also to prepare you for the exams. You will be best served (also with regard to the exams) if you do this work by yourself or in close collaboration with your teammates. Yet the ability to collaborate with others to solve problems and produce items of value is a tremendously important skill. And often you learn a lot by discussing and bouncing ideas back and forth with your classmates.

To that end, I encourage you to collaborate with other students in the class in discussing reading and lecture material and evaluating your ideas and concepts. Also, most of the assignments and class preparations are explicitly collaborative. However, everything that you turn in to be graded must be your (or your team's) own work. On some homework assignments I may encourage more or less collaboration depending on the nature of the assignment. I will try to make the boundaries of appropriate collaboration very clear for each assignment.

Outside Resources

In general, I encourage students in the class to make use of resources available outside of the assigned readings. These resources include, but are not limited to, web sites, articles, books, online discussion groups, Google searches, etc. There is a tremendous amount of information available on the web and elsewhere. Make use of it. The only two exceptions are 1) work (in whole or in part) that has been completed by other students in this or previous years for the same or substantially the same assignment, and 2) internet materials directly related to a case/problem set unless explicitly authorized by the instructor. If you choose to make use of somebody else's work you must provide appropriate attribution for the work and add a significant contribution of your own to the original work. Although I encourage you to synthesize and build on what you have found elsewhere (with appropriate citations and with the two exceptions above) when completing homework assignments or solving technical problems, everything that you turn in to be graded must be your (or your team's) own work.

If you have questions regarding how to appropriately attribute work that you have built on or incorporated into your own, or what constitutes an acceptable amount of extension of the prior work, please ask me in class, send me an email, or come see me to discuss an appropriate course of action before submitting your assignment.

Some Don'ts summarized:

- Do not copy all or part of another student's work (with or without "permission").
- Do not allow another student to copy your work.
- Do not ask another person to write all or part of an assignment for you.
- Do not use material without explicit quotation and/or citation.
- Do not consult or submit work (in whole or in part) that has been completed by other students in this or previous years for the same or substantially the same assignment.
- Do not use material directly related to a case/problem set unless explicitly authorized by the instructor.
- Do not submit the same, or similar, piece of work for two or more courses without the explicit approval of the two or more instructors involved.

The midterm and final exam will be closed book. This holds even if part of the exams may be to be completed on either your laptop or a lab computer. During the exams, any student who either receives or knowingly gives assistance or information concerning the examination will be in violation of the policy on individual work. The violation of the policy on individual work is a serious offense, and suitable consequences include grade reduction, an F grade, a transcript notation, delay of graduation, or expulsion from CMU.

Resolving Ethical Dilemmas

I recognize that most ethical dilemmas do not necessarily present obvious and clearcut right and wrong alternatives. If you find yourself wondering whether a particular course of action will violate the academic integrity policy I suggest that you use the following guidelines:

- Ask yourself whether you would be embarrassed or concerned if the instructor, your peers, or a possible recruiter found that you had completed your assignment in this way (in which case you should probably not do it…).
- If still in doubt, come talk with me to discuss whether I am likely to consider the course of action a violation of the academic integrity policies.

ACKNOWLEDGEMENTS

While the individual topics and organization of this class are new, some of the content of this course is building upon previous iterations of similar courses taught by Professors Anjana Susarla, Bob Monroe, and Dan Suciu. I appreciate their help and permission to build on their great work. I also thank Marsha Lovett from the Eberly Center, Bob Monroe, and Mike Trick for ideas concerning curriculum design.

5. PRELIMINARY CLASS CALENDAR

Week	Tuesday	Thursday
1	L1 (Jan 15) *Course introduction & Excel 1: Excel as database*	L2 (Jan 17) *Excel 2: Advanced Formulas*
2	L3 (Jan 22) *Excel 3: Pivot tables*	L4 (Jan 24) *Excel 4: Array formulas and Macros*
3	L5 (Jan 29) *Synthesis 1*	L6 (Jan 31) *Synthesis 2*
4	L7 (Feb 5) *SQL 1* *HW 1 due*	L8 (Feb 7) *SQL 2*
5	L9 (Feb 12) *SQL 3*	L10 (Feb 14) *SQL 4*
6	L11 (Feb 19) *Data Modeling 1* *HW 2 due*	L12 (Feb 21) *Data Modeling 2*
7	L13 (Feb 26) *Data Modeling 3*	L14 (Feb 28) *Data Modeling 4* *Project phase 1*
8	L15 (March 5) *Review for Midterm*	L16 (March 7) *Midterm*
9	(March 12) Spring break	(March 14) Spring break
10	(March 19) Spring break	L17 (March 21) *Data Modeling 5*
11	L18 (March 26) *Data Modeling 6*	L19 (March 28) *Advanced DB Management 1*
12	L20 (April 2) *Advanced DB Management 2* *HW3 due*	L21 (April 4) *Advanced DB Management 3*
13	L22 (April 9) *Advanced DB Management 4*	L23 (April 11) *Advanced DB Management 5* *Project phase 2*
14	L24 (April 16) *Advanced DB Management 6* *HW4 due*	L25 (April 18) *TBD*
15	L26 (April 23) *Project presentations*	L27 (April 25) *TBD*
16	L28 (April 30) *Probabilistic Databases* *HW 5 due*	L29 (May 2) *Review for Exam*
17	*Final exam TBD: May 9th – 10th*	