

## IDS 521-Fall 2015 Syllabus

### Advanced Database Management

Dr. Ali Tafti

Classroom: DH 220  
Class Time: 3:00 – 5:45 pm Mondays

Office Hours: W 1:30 – 3 pm

**Contact:**

E-Mail: [atafti@uic.edu](mailto:atafti@uic.edu)  
Office: 2403 University Hall

#### **Course Overview and Objectives:**

In this course, we will focus on using relational database management systems (RDBMS) and Data Warehousing (DW) systems for analysis and reporting, with a special focus on advanced topics in Structured Query Language (SQL) and dimensional modeling. We will begin with an intense review of RDBMS basics and common patterns in SQL queries.

This course assumes prior familiarity with SQL. If you are coming into this course as a SQL novice, the course will give you a very steep introduction to SQL, which you will need to supplement with additional readings and exercises. If you already have working knowledge of SQL, this course will give you some conceptual tools to write more efficient SQL queries, advanced querying tools, and a more systematic approach to difficult database queries. For example, practitioners with working knowledge of SQL sometimes encounter problems with null values (three-valued logic), complex joins, and aggregate functions. You can tackle such challenges with some knowledge of relational theory and thinking in terms of sets. Students will practice with numerous SQL examples and reinforce SQL skills in two in-class exams.

We will then turn our focus upon Data Warehouses, which serve as a foundation for Business Intelligence (BI) systems. Data Warehouses have very different goals and design practices than databases in operational systems. Data Warehouses enable analysts and managers to monitor all aspects of firm performance, and to facilitate analytical decision-making. Without properly designed Data Warehouses, analysts must patch together datasets from a variety of sources in an ad hoc fashion. As a result, they often end up making bad decisions based on untimely or unreliable data. Properly designed Data Warehouses enable decisions based on easily accessible, timely, adaptable, secure, and credible data.

Pending the progress of the class in grasping the core concepts, we will also gain practice with leading commercial tools for designing and querying databases and data warehouses – primarily using MySQL and the suite of BI tools on Microsoft SQL Server. We will apply techniques of dimensional modeling in specific business scenarios. If time permits, the course may include an overview of new paradigms of information architecture for Big Data analytics, and how data warehousing practices can best align with new developments in NoSQL, Hadoop/Map-Reduce, and other new Big Data paradigms of database architecture beyond relational databases.

#### **Prerequisites**

You are expected to have working familiarity with the design and querying of Relational Databases Management Systems (RDBMS); in particular, Entity Relationship (ER) models and Structured Query Language (SQL). We will schedule some intensive review of the prerequisite topics into class sessions as needed, and provide some links and references to additional tutorials such as on Lynda.com. Students who are not well versed on

prerequisite topics in RDBMS or SQL will need to invest extra time.

### **Textbook Resources:**

We will be drawing from a number of textbook resources in this course. Exam material will focus on your lecture notes and required readings. Lecture notes may include some of the key ideas from the optional readings.

### ***REQUIRED***

*The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, by Ralph Kimball and Margy Ross, 3rd edition (publication date: July 1, 2013), John Wiley & Sons (ISBN: 978-1118530801). The Kindle e-book version is available from Amazon.

### ***RECOMMENDED for SQL Advanced Topics***

*Joe Celko's SQL for Smarties, 4<sup>th</sup> edition*; by Joe Celko, 4th Ed. (2011), Elsevier (ISBN: 978-0-12-382022-8). **Available free online from the UIC library:**  
<http://www.sciencedirect.com.proxy.cc.uic.edu/science/book/9780123820228>

### ***RECOMMENDED for SQL Basic Topics***

*Database Management Systems*, 3<sup>rd</sup> ed. by Raghu Ramakrishnan and Johannes Gehrke, McGraw-Hill, 2002. **NOTE: A custom version of this book (print and digital) will be made available to you at a reduced price.**

*Discovering SQL : A Hands-on Guide for Beginners*, by Alex Kriegl. Hoboken, NJ, USA: Wrox, 2011. ProQuest ebrary. Web. 13 January 2015. **Available free online from the UIC library:**  
<http://site.ebrary.com.proxy.cc.uic.edu/lib/uic/reader.action?ppg=256&docID=10513926&tm=1421170699065>

*Learning MySQL*, by Seyed M.M. (Saeied) Tahaghoghi and Hugh Williams, 1<sup>st</sup> edition (November 21, 2006), O'Reilly Media (ISBN: 978-0596008642). **Available free online from the UIC library:**  
<http://proquestcombo.safaribooksonline.com.proxy.cc.uic.edu/book/databases/mysql/0596008643>

### ***OPTIONAL: Other resources we will draw from in class:***

*SQL Cookbook*, by Anthony Molinaro, (published in 2006), O'Reilly Media, 1<sup>st</sup> edition (publication date: December 23, 2005), O'Reilly Media (ISBN: 978-0596009762).

*Delivering Business Intelligence Solutions with Microsoft SQL Server 2012*, by Brian Larson, Third edition (2012), McGraw Hill (ISBN: 978-0-07-175938-0).

### **Software:**

The first two assignments will require MySQL, which is an open source database available for free download. Students may use their RDBMS of choice for the project. At least one of the homework assignments in this course will require Microsoft SQL Server 2012 BI Analysis Services, and each student will be given a temporary license for this software through the *Microsoft Development Network Academic Alliance* (MSDNAA).

## **Office Hours and Contact Information**

Please feel free to walk in during my posted office hours. If you need to see me outside of office hours, please email me to schedule an office appointment or phone call. I usually check my email at least once per day, and typically respond within 48 hours. During vacation, conference travel, or weekend hours, responses may take longer.

### **Part 1: Relational Database Management Systems: SQL Review and RDBMS foundations**

Topics
<ol style="list-style-type: none"> <li>1) Relational Database Management System (RDBMS) foundations</li> <li>2) Normal Forms</li> <li>3) Relational Theory</li> <li>4) Thinking in Sets</li> <li>5) Declarative programming</li> <li>6) SQL Recipes</li> </ol>

### **Part 2: Advanced topics in SQL and RDBMS**

Topics
<ol style="list-style-type: none"> <li>1) Null-value and duplicate row complications</li> <li>2) Optimizing SQL queries</li> <li>3) Indexes</li> <li>4) Trees and Hierarchies</li> <li>5) SQL for analytic reporting: Descriptive Statistics and Aggregate Functions</li> </ol>

### **Part 3: Data Warehousing: Patterns in Dimensional Modeling**

Topics	Business Process Examples
<ol style="list-style-type: none"> <li>1) DW/BI Lifecycle Overview</li> <li>2) Overview of Dimensional Modeling</li> <li>3) Slowly changing dimensions</li> <li>4) Enterprise Bus Matrix</li> <li>5) Inventory modeling patterns</li> <li>6) Hierarchies and multi-dimensional attributes</li> <li>7) Bridge tables, outriggers, and step dimensions</li> </ol>	<ol style="list-style-type: none"> <li>1) Retail Sales</li> <li>2) Inventory Management</li> <li>3) Procurement</li> <li>4) Order Management</li> <li>5) Accounting</li> <li>6) Human Resources Management</li> <li>7) Customer Relationship Management</li> </ol>

**Course Website:** <https://uic.blackboard.com>

Course materials will be posted on Blackboard. It is your responsibility to check the Blackboard on a regular basis so that you would not miss any important information.

**Performance Evaluation Summary:**

Your final grade for this class will be based on the following point distribution:

Activity	Team/individual	Points
In-class participation quality and attendance	Individual	10%
Homework exercises	Team	15%
Exam 1	Individual	10%
Exam 2	Individual	30%
Online quizzes	Individual	10%
Final Presentation and Project Report	Team	25%
<b>TOTAL*</b>	Individual	<b>100%</b>

\* Bonus points towards in-class participation or exams may be awarded for outstanding test questions written by students.

**In Class Participation Quality and Attendance (10%)**

In order to make the learning process more effective and interesting, it is important that you come prepared for each class and participate in all class activities. During the semester, we will have several in-class case discussions and exercises. You are highly encouraged to share your experience, and your thoughts on topics being discussed. If you do miss a class, you will be fully responsible to get the notes and assignment from a peer in your class or from the course website. Attendance will be taken in class and will be considered in the final grade.

**Use of computers in class:** I will encourage use of laptops or tablet computers in regular class sessions as long as *the use is related to what we're covering in that session*. I will require laptops to be closed and put away during specific segments of lecture, guest lectures and some class exercises. If you abuse this policy, I reserve the right to consider this in your final participation grade and also to revoke these privileges. The use of mobile phones or smart phones will not be permitted at all.

**Team Homework Exercises (15%)**

There will be three team homework exercises this semester. Assignments focus on using database tools to conduct summary and analysis of large datasets. The first pair of assignments will focus on RDBMS and SQL skills. The first will involve basic SQL exercises, and the second will be an advanced SQL project. The third assignment will focus on dimensional modeling design, and analyzing data within an OLAP cube or dimensional model. These homework assignments are designed to help prepare you to complete the final project; although you can begin working on the project before completing the assignments. You may choose to be in different groups for any of the homework exercises.

**Two exams (10% and 30%)**

There will be two proctored in-class exams designed to test your grasp of the concepts covered in the textbook readings as well as in the lectures. The second exam will be comprehensive. Please check the schedule on Blackboard for exam dates.

### **Online quizzes (10%)**

You will take up to four online quizzes via Blackboard. The quizzes will typically be in multiple-choice format, and they will be designed to test your comprehension of the reading assignments. These are open-book quizzes.

### **Final Presentation and Project Report (team) (25%)**

- Teams of 3 or 4 students will present a database or data warehouse design to model a specific business process. The solution will involve a detailed discussion of the business problem and how it compares and contrasts to cases covered in the course texts.
- A presentation session is scheduled for the last week(s) of class.
- Your team will submit a final professional-quality report at the end of the semester. Guidelines for this report will be given later in the semester.

### **Academic Integrity**

Academic integrity is a cornerstone of the intellectual life of a university student. Consistent with University policy, violations of academic integrity will be a serious offense. Additionally, citing your references and identifying the sources for all direct quotations are required for your projects or written work submitted online. Students are expected to respect and act in accordance with the UIC Student Honor Code. Violations of the student code, including but not limited to assignments, exams, or classroom conduct, will be referred to the Office of the Dean of Students. **The range of disciplinary actions for academic dishonesty can include an automatic F in the course and expulsion from the graduate program.** *Please note: Software tools can automatically detect and alert us to possible instances of plagiarism or copying. Software can detect similarities to text in any Internet sources, published materials, as well as unusually high levels of similarity between student submissions. Once the software raises an alert, we can manually review passages flagged as similar and then take further action if necessary.*

### **Special Accommodations**

If you have any condition, such as a physical or learning disability, which will make it difficult for you to carry out the work as we have outlined it or which will require academic accommodations, please notify Dr. Tafti during the *first two weeks* of the course, or as soon as your condition becomes known to you, and we will strive to accommodate you. For observance of religious holidays that require you to miss a class, you need to notify the professor by the tenth day of the semester of the date when you will be absent, unless the religious holiday is observed on or before the tenth day of the semester. In such cases, you shall notify the professor at least five days in advance of the date when you will be absent.

**Tentative Schedule of Activities: Updated August 18, 2015**

**NOTE:** Below is a **tentative** schedule. The course schedule may be modified during the semester to adjust for our pace of progress. \*\*\*Please check Blackboard for schedule updates, readings, and exact assignment due dates. This schedule may be modified throughout the semester. The due dates posted on Blackboard with each assignment or quiz take precedence over this document, whenever the posted due dates differ.

Session	Tentative Dates	Topic	Project Assignments	Quizzes
1	Aug 24	Course Introduction, Relational Database Management System (RDBMS) foundations		
2	Aug 31	Normal Forms, Relational Theory		
3	Sep 7	Labor Day Holiday – class does not meet		Quiz 1
4	Sep 14	Thinking in Sets for SQL; Review of database design exercises	Hwk 1: Basic SQL Exercise	
5	Sep 21	Practical SQL topics: Descriptive statistics and aggregate functions, Null-value and duplicate row complications		
7	Sep 28	Advanced SQL topics: SQL Trees and Hierarchies; Optimizing SQL queries, and Indexes		Quiz 2
8	Oct 5	Review of Exam topics; Part 1 of Research tutorial: Using database tools to analyze Twitter and financial markets		
9	Oct 12	<b>In-class Exam 1.</b>		
10	Oct 19	Part 2 of Research tutorial MySQL demo with indexes Introduction to Data Warehousing		
11	Oct 26	Ch. 1 DW/BI & Dim. Modeling Primer Ch. 3 Retail Sales: Dimensions and Fact Tables	Hwk 2: Advanced SQL Exercise	
12	Nov 2	Ch. 4 Inventory: Fact table types, Enterprise bus matrix Ch. 5 Procurement: Slowly changing dimensions		Quiz 3

13	Nov 9	Ch. 6 Order Management: Accumulating snapshot modeling patterns Ch. 7: Accounting: Modeling hierarchies; multi-dimensional attributes		
13	Nov 16	Ch. 8 – 9: Customer relationship management (CRM) and Human resources management (HRM): Bridge tables, outriggers, and step dimensions		Quiz 4
14	Nov 23	<i>[placeholder for guest lecture]</i>	Hwk 3: Dimensional modeling exercise	
15	Nov 30	<b>Group project presentations</b>		
16	Dec 7	<b>In-class Exam 2 (comprehensive)</b>		
	Dec 11	<b>Final project reports are due</b>		