Northeastern University

Course Syllabus

DATA MANAGEMENT and DATABASE DESIGN
INFO 6210
Lectures: Thursdays 5:00 – 7:00 pm PST (8:00 – 10:00 pm EST)
Office Hours: By appointment
September 5 to December 15
Bluejeans online meeting room
Meeting URL:
https://bluejeans.com/7795873366

Simon W. Wang
425.829.2503
w.wang@neu.edu
swpwang@uw.edu

COURSE DESCRIPTION

Database Design and Implementation introduces database development theory and skills, including database concepts, architecture, and deployment in a user-centered system context using SQL. This course presents many of the valuable skills required for managing, manipulating and analyzing large amounts of data, related to the relational data model, the database system lifecycle (including entity-relationship diagramming and normalization), SQL programming, security and privacy issues, as well as principles of transaction management. NoSQL is also explored.

Students will develop a project database on a business premise of their choosing as well as conduct analyses upon a set of sample data. This course also presents a foundation for more advanced study with database technology; each student will have the opportunity to gain a solid understanding of how database systems are deployed and used within the context of real-world challenges.

BOOKS

Connolly, T. M. & Begg, C. E. (2015)
Database Systems: A Practical Approach to Design, Implementation, and Management (6th Edition) Addison-Wesley Publishing, [ISBN-10: 0-13-294326-3]

The 4th Edition is also acceptable. Copy of Fourth Edition on .pdf is available at:
http://www.palinfonet.com/download/software2/DATABASE%20SYSTEMS.pdf

Dusan Petkovic (2016)
Microsoft SQL Server 2016: A Beginner's Guide, Sixth Edition
McGraw Hill, [ISBN: 978-1259641794]

These textbooks have been selected because of their breadth and depth of coverage of databases. They are well written and contain many examples. Students should find these books to be useful for several years to come.

Recorded lectures:
YouTube Channel for Database Design:
https://www.youtube.com/channel/UCP4n040ay46QjKYLmnBYCmw

YouTube Channel for SQL Development:
https://www.youtube.com/channel/UCwMn1c7Oq1VmW1t8gM7J5IA

YouTube Channel for Data Mining and Database Management:
https://www.youtube.com/channel/UCACs2TalmQuAa-M5042lBTg

Additional resources:

Elmasri, Ramez and Navathe, Shamkant (2011); *Fundamentals of Database Systems* (6th Edition) New York: Pearson [ISBN 10: 0-136-08620-9]
http://www.cse.hcmut.edu.vn/~ttqnguyet/CSDL/EbookDB.pdf

Halvorson, H.P. (2014). *Introduction to Database Systems.* (1st Edition) Porsgrunn, Norway: Faculty of Technology.
http://home.hit.no/~hansha/documents/database/training/Introduction%20to%20Database%20Systems/Introduction%20to%20Database%20Systems.pdf

Garcia-Molina, H., Ullman, J., & Widom, J. (2009); *DATABASE SYSTEMS The Complete Book* (2nd Edition) Upper Saddle River, NJ: Pearson Prentice Hall [ISBN 0-13-606701-8]
http://people.inf.elte.hu/nikovits/DB2/Ullman_The_Complete_Book.pdf

W3schools SQL Tutorial will be leveraged to develop basic SQL skills. This learning will be used to support more extensive SQL development enhancing the course objectives.

T-SQL Querying (Developer Reference) 1st Edition
https://www.amazon.com/T-SQL-Querying-Developer-Reference-Ben-Gan/dp/0735685045

SOFTWARE

SQL Server will be provided in a hosted environment. Students will need to download and install SQL Server Management Studio to connect to it. Students can also install the complete SQL Server software on their computers. The Developer Edition of SQL Server 2012/14/16/17 is recommended. In addition to SQL Server, MS Visio or MySQL Workbench is required.

Software can be downloaded from the Northeastern Dreamspark site. Details about the Dreamspark site is available at: http://www.northeastern.edu/its/dreamspark/ . Instructor will provide the URL to the hosted server in class.

LEARNING OBJECTIVES

Upon successfully completing the course, students will be able to conduct the following:

- describe the rationale for designing and deploying database management systems
- explain the differences between Relational Database Management Systems and NoSQL Database Management Systems
- communicate the various forms of data integrity (domain, entity and referential)
- define the process of developing a fully-normalized database design
- explain the structural components of databases (entities, attributes, data types & indexes)
- perform queries and analysis of data using SQL programming language
- articulate concepts of ACID properties and principles of transaction management
- describe legal and ethical issues related to data privacy and ownership

EVALUATION:

Assignments balance between theory and practice and between individual and group work.

Assessment	% Grade
2 Quizzes	30%
5 Lab Exercises	30%
Database project	40%

ATTENDANCE
This course will meet once a week on Thursday evenings. Your attendance is paramount to your success in this class. Contact the instructor if you have a question about the class attendance.

DATABASE PROJECT
Students will form teams of 3 and develop a relational database based on reading and class lectures. The project will have the following deliverables:

Deliverable	%	Points	Week
P1. Team Formation, DB Topic and Objectives	5	5	2
P2. Database Design, Initial ERD	10	10	5
P3. Final ERD	5	5	8
P4. Database Implementation	10	10	11
P5. Presentation	10	10	14

The rubrics for the project grading is Completeness 40%, Correctness 40%, and Creativity 20%.

Team Formation, Database Topic and Mission Statement/Objectives
Form a team of three members. Each team will collaborate to decide a database topic. The database topics may be like Book Store, University Registration, etc. Each team will also establish the mission statement and identify the mission objectives that the database will accomplish. The mission objectives may be like Book Sale, Inventory Control, etc.

* Individual submission is required.

Database Design and Initial ERD
Based on reading and class lectures, each team will create an initial Entity-Relationship diagram (ERD) that depicts a database for a real or fictitious business. This database will allow for data collection, processing, and reporting for a particular organization. It is strongly suggested that each team model a database for a type of organization that they have relatively deep understanding---such as the current or previous work experience or perhaps a personal hobby. In the past, students have created databases to capture data about video rental stores, bike repair shops, beer tasting/review professionals, athletic leagues, and airlines. Students are encouraged to use their imagination!

Each team will submit an ERD for the database of their choosing. The target for the initial ERD is 12 entities or more. In addition to the ERD, students should submit a database design document containing the description of the business problems being addressed by their database, list all entities and how they are related to each other, and key design decisions.

For the part of the business problems being addressed, this section could be similar to the mission objective document completed earlier. Additionally and more importantly, this document should contain your team's key database design decisions, such as why an entity is included and how that entity is related to other entities.

* Entity-Relationship Diagramming tool, Microsoft Visio or MySQL Workbench can be downloaded for free.
* Individual submission is required.

Final ERD
Based on the instructor's feedback of the initial ERD, each team will make improvements to the initial ERD. Most likely, these changes will be in regards to further 'normalization' of the database entities, reducing redundant data, and recognizing additional entities.

In addition to submitting a fine-tuned ERD, each team will also submit a brief description identifying the changes made to the initial ERD. It is also important to update the design document to reflect the new design changes. Resubmission of the updated design document is not required at this time.

* Individual submission is required.

Database Implementation

Each team will submit the 'SQL code' to implement the database design as well as enter a minimal amount of data (at least 10 rows for each table) using the SQL INSERT scripts, Data Import Wizard, and/or stored procedures. Specific objects to be reflected in the code include the database, tables, data types, primary and foreign keys, and views. Each team is expected to create at least 2 views (often used for reporting purposes).

The implementation must include at least two of the following three items:

- Table-level CHECK Constraints
- Computed Columns based on a function
- Column Data Encryption

* Individual submission is required.

Project Presentation

Each team will present the database design project to the class. The presentation should include the following items.

1) A Power Point slide deck, containing highlights, to showcase the project
2) The design document
3) The final ERD
4) The SQL DDL statements for implementing the database
5) At least two views for reporting purposes and the SQL DDL statements used to create them
6) At least two reports (Using PowerBI is required. Other data mining tools could also be used.)
7) Audio/Video presentation if preferred but not required
8) Only one member of team needs to submit the presentation materials

LATE WORK

All assignments must be submitted to the class Blackboard site for the course on the due date before 11:59 pm. If you turn in an assignment late, 10% credit will be deducted from the total score for each day after the deadline. Assignments turned in more than one week late will not receive credit. In the case of unexpected events, you must contact the instructor before the assignment due date in order to receive a grace period.

ACADEMIC HONESTY & PLAGIARISM

It is contrary to justice, academic integrity, and to the spirit of intellectual inquiry to submit another's statements or ideas of work as one's own. To do so is plagiarism or cheating, offenses punishable under the University's disciplinary system. Because these offenses undercut the distinctive moral and intellectual character of the University, we take them very seriously.

Proper acknowledgment of another's ideas, whether by direct quotation or paraphrase, is expected. In particular, if any written or electronic source is consulted and material is used from that source, directly or indirectly, the source should be identified by author, title, and page number, or by website and date accessed. Any doubts about what constitutes "use" should be addressed to the instructor.

GRADING CRITERIA

Work in this course will be graded to criteria. In other words, you won't be graded on a curve. Each assignment is designed to test your achievement against one or more of the learning objectives. Different assignments emphasize different learning objectives. The meanings of grades are described below:

Letter	Percent
A	100-97
A-	96-90
B+	89-87
B	86-84
B-	83-80
C+	79-77
C	76-74
C-	73-70

Class Schedule

Important Note: Changes may occur to the syllabus at the instructor's discretion. When changes are made, students will be notified via Blackboard and/or in-class announcement.

Week One: September 5 - 9

General Database Purpose and Development History

Presenting an overview of the entire course as well as an introduction to the reasons behind the growth of database management systems (DBMS). It explores the history of database use and the mistakes and dead-ends of the past to present the student with a context in which to develop criteria for judging database design and effectiveness.

Reading

- chapter 1: "Introduction to Databases" (Connolly& Begg)
- chapter 4: "The Relational model (1980 - present)" (Connolly& Begg)

Week Two: September 10 - 16

Entity–Relationship Modeling

Basic concepts of diagramming business objects are presented in this lesson; how to identify entities, attributes, relationships and cardinality. Lecture includes Primary Keys and Foreign Keys that align with business rules.

Reading

- chapter 10: "Database System Lifecycle" (Connolly& Begg)
- chapter 11: "Entity–Relationship Modeling" (Connolly& Begg)
- chapter 12: "Enhanced Entity–Relationship Modeling" (Connolly& Begg)

Assignment

P1

Week Three: September 17 - 23

Conceptual and Logical Database Design

Processes for conducting the conceptual and logical database design are discussed.

Introduction to Structured Query Language (SQL)

Introduction to the standard database language, SQL, is presented.

Reading

- chapter 6: "SQL: Data Manipulation (DML)" (Connolly& Begg)
- chapter 16: "Methodology: Conceptual Database Design" (Connolly& Begg)
- chapter 17: "Methodology: Logical Database Design for Relational Model" (Connolly& Begg)

Assignment
Lab 1

Week Four: September 24 - 30
Normalization and Database Constraints

This lesson presents the Normalization process for fine-tuning and validating the database design. Database integrity constraints (domain, entity and referential) are also discussed.

SQL Concepts

SELECT

Reading

- chapter 13: "Normalization" (Connolly& Begg)
- chapter 3: "SQL Server Management Studio" (Petkovic)

Week Five: October 1 - 7
Physical Database Design

This module explores the database design process which adapts to a database management system for implementation.

SQL Concepts

Aggregate Functions and GROUP BY; JOIN

Reading

- chapter 6: "Queries" (Petkovic)
- chapter 18: "Physical Database Design for Relational Databases" (Connolly& Begg)

Assignment
P2

Week Six: October 8 - 14

Database Objects

Additional objects that are common in nearly all databases, including stored procedures, functions and indices are explored. Also, discussion will include enterprise-class database systems which increase in size and scale.

SQL Concepts

Subquery; CTE

Reading

- chapter 7: "SQL: Data Definition (DDL)" (Connolly& Begg)
- chapter 10: "Indices" (Petkovic)

Assignment
Lab 2

Week Seven: October 15 - 21
Database Design Review
Week seven is for review and exploration of the initial ERDs submitted by each student group.

SQL Concepts
CASE and RANK

Assignment
Lab 3

Week Eight: October 22 - 28
Transaction Management / ACID Properties
Transaction Management describes the principles of managing data consistency and integrity while processing transactions. The concepts of ACID properties are also introduced.

SQL Concepts
Explicit Transaction

Reading
- chapter 21: "Transaction Management (ACID Properties)" (Connolly& Begg)
- chapter 13 "Concurrency Control" (Petkovic)

Assignment
Quiz 1
P3

Week Nine: October 29 - November 4
Database Security
Data and database security is explored in this module.

SQL Concepts
DDL; Recursive Processing

Reading
- chapter 12: "Security System of the Database Engine" (Petkovic)
- chapter 20: "Security and Administration" (Connolly& Begg)
- chapter 21: "Professional, Legal, and Ethical Issues" (Connolly& Begg)
- chapter 5: "Data Definition Language" (Petkovic)
- chapter 11: "Views" (Petkovic)

Assignment
Lab 4

Week Ten: October 5 - 11
Database Constraints and Business Rules
Discussion will include the need for specifying constraints that reflect business rules unique to the organization using the database. Defining, documenting as well as coding the restrictions required of the data is presented.

SQL Concepts

Stored Procedures and Functions; APPLY

Assignment
Lab 5

Week Eleven: October 12 - 18
Data Warehousing Concepts: Design
Discussion will include the history and evolution of data warehousing, its main concepts and the competitive advantages that businesses realize after implementing data warehouse. An overview of data warehousing components and the process of construction are presented. ETL tools are presented.

Online Analytical Processing and Data Mining Concepts
Aspects of mining data from large-scale databases are analyzed with a focus on Multidimensional Data Model and OLAP Tools. Data mining tools are discussed.

SQL Concepts
MERGE; PIVOT

Reading
- chapter 32: "Data Warehousing Concepts" (Connolly& Begg)
- chapter 33: "Data Warehouse Design" (Connolly& Begg)
- chapter 34: "Data Mining" (Connolly& Begg)

Assignment
P4

Week Twelve: November 19 - 25
No class

Week Thirteen: November 26 - December 2
Database Administration Concepts I
Aspects of administering large-scale databases are analyzed with a focus on challenges of production operations. Topics include disaster recovery, maintenance, high-availability and scalability.

Reading
- chapter 19: "Methodology: Monitoring & Tuning the Operational System" (Connolly& Begg)

Assignment
Quiz 2

Week Fourteen: December 3 - 9
Database Administration Concepts II
Aspects of administering large-scale databases are analyzed with a focus on challenges of production operations. Topics include monitoring, troubleshooting and optimization.

Reading
- supplemental reading TBD

Assignment
Team Project Presentations

Week Fifteen: December 10 - 15