



Northeastern University

Course Syllabus

DATA MANAGEMENT and DATABASE DESIGN

DAMG 6210

Live Sessions: Saturdays 9:30 am – 11:30 am Pacific (12:30 – 2:30 pm Boston)

Office Hours: TBD

January 18 to May 7

Zoom online meeting room

Meeting URL:

<https://northeastern.zoom.us/j/5715124862?pwd=YU1PcG9ZaTdicGFqWDBOUUdORjFIUT09>

Simon W. Wang

425.829.2503

w.wang@neu.edu

swpwang@uw.edu

COURSE DESCRIPTION

The Data Management and Database Design course introduces the database theory and skills, including the concepts, architecture, and development in a data-centric system context. This course presents the valuable knowledge and skills related to managing, manipulating and analyzing large amounts of data. Both the relational and NoSQL databases are explored. Students will develop a project database on a business premise of their choosing as well as conduct analysis upon a set of sample data.

This course not only prepares students to pursue the career opportunities but also presents a foundation for the more advanced study related to the database discipline.

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 or 5th Edition is also acceptable.

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 in coverage of data and 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:

[W3schools SQL Tutorial](#) will be leveraged to develop basic SQL skills. This learning can also be used to support the 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

The class SQL Server will be provided in a hosted environment. Students will need to have access to other software according to the following table. The Developer Edition of SQL Server 2019 and Visio Professional 2016 are recommended as optional options.

Function	Requirement	Apple	Windows
Entity Relationship Diagram	mandatory	MySQL Workbench, or Lucidchart	Visio Professional, MySQL Workbench, or Lucidchart
MongoDB Server	mandatory	MongoDB Atlas in cloud	MongoDB Atlas in cloud
MongoDB Client	mandatory	MongoDB Compass Community	MongoDB Compass Community
MongoDB Programming	optional	Node.js	Node.js
SQL Programming - Server	mandatory	Class SQL Server	Class SQL Server
SQL Programming - IDE	mandatory	dBeaver	SQL Server Management Studio, or dBeaver
Data Mining Tool	mandatory	Tableau, or others	PowerBI, Tableau, or others

Important Notes:

* IDE (Integrated Development Environment)

*** The minimum requirement for Entity Relationship Diagram is Lucidchart**

*** The minimum requirement for SQL Programming is dBeaver**

* Please watch the videos in the YouTube Channels listed in the Recorded Lectures section for help to get started with the software

* If you want a more powerful tool and want to learn more, please explore other options listed in the above table.

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

2 Quizzes

% Grade

30%

5 Lab Exercises	30%
Database project	40%

ATTENDANCE

This course will meet once a week on Saturday. 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 up to 5 and develop a relational database based on reading and class lectures. The project will have the following deliverables:

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

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 up to five members. Team members will collaborate to decide a database topic, mission statement, and mission objectives that the database will accomplish.

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 the video rental stores, bike repair shops, beer tasting/review professionals, athletic leagues, and airlines. Students are encouraged to use their imagination!

The mission objectives may be like Book Sale, Inventory Control, etc., for a Book Store database.

* Individual submission is required.

Database Design and Initial ERD

Based on the reading and class lectures, each team will create an initial Entity-Relationship diagram (ERD) that depicts the database topic chosen in the first project assignment and is normalized to the Third Normal Form. This database will allow for data collection, processing, and reporting. 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, listing all entities and how they are related to each other, and explaining the key design decisions. The business problems being addressed could be similar to the mission statement completed earlier. An example of the key database design decisions is why an entity is included in the database.

* 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 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 the appropriate 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 based on a function
- 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 up-to-date 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 a data mining tool, such as PowerBI and/or Tableau, is required)
- 7) Audio/Video presentation if preferred but not required
- 8) Only one member of a team needs to submit the presentation materials

LATE WORK

All assignments must be submitted to the **class Canvas** 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

Occurrences of academic dishonesty, such as submitting work that is not the student's own, will be dealt with according to the NEU's and COE's policies on the academic dishonesty. **Students who allow their work to be copied will be treated the same as those who copy it.**

Please read what constitutes the academic dishonesty and how the University will respond to such incidents:

<http://www.northeastern.edu/osccr/academic-integrity-policy/>

Academic integrity is important for two reasons. First, independent and original work ensures that students derive the most from their educational experience and the pursuit of knowledge. Second, academic dishonesty violates the most fundamental values of an intellectual community and depreciates the achievements of the entire University. It is the student's responsibility to know and follow the codes of academic dishonesty.

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 Canvas and/or in-class announcement.

Week One: January 18 -23

General Database Purpose and Development History

Presents 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) - **Optional**
- chapter 4: “The Relational model (1980 - present)” (Connolly& Begg)

Week Two: January 24 -30

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 Development Lifecycle” (Connolly& Begg)
- chapter 12: “Entity–Relationship Modeling” (Connolly& Begg)
- chapter 13: “Enhanced Entity–Relationship Modeling” (Connolly& Begg)

Assignment

P1 Due 2/3

Week Three: January 31 - February 6

Conceptual and Logical Database Design

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

Normalization

This lesson presents the Normalization process for fine-tuning and validating the database design.

MongoDB Aggregation Pipeline

Introduction to the MongoDB Aggregation Pipeline is presented.

Introduction to Structured Query Language (SQL)

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

Reading

- chapter 6: “SQL: Data Manipulation (DML)” (Connolly& Begg) - *Optional*
- chapter 14: “Normalization” (Connolly& Begg)
- chapter 16: “Methodology: Conceptual Database Design” (Connolly& Begg)
- chapter 17: “Methodology: Logical Database Design for Relational Model” (Connolly& Begg)

Assignment

Lab 1 Due 2/10

Week Four: February 7 - 13

Database Constraints

Database integrity constraints (domain, entity and referential) are discussed.

SQL Concepts

SELECT

Reading

- chapter 3: “SQL Server Management Studio” (Petkovic) - *Optional*

Week Five: February 14 – 20

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) - *Optional*

Assignment

P2 Due 2/24

Week Six: February 21 - 27

Database Objects

Database objects that are common in nearly all databases, such as indices, are explored.

SQL Concepts

Subquery; CTE

Reading

- chapter 7: “SQL: Data Definition (DDL)” (Connolly& Begg) - *Optional*
- chapter 10: “Indices” (Petkovic)

Assignment

Lab 2 Due 3/3

Week Seven: February 28 - March 6

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 Due 3/9 (Wednesday)

Week Eight: March 7 - 13

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 22: “Transaction Management (ACID Properties)” (Connolly& Begg)
- chapter 13 “Concurrency Control” (Petkovic)

Assignment

Quiz 1, Saturday, 3/12, 12:30 pm Boston time (9:30 am Pacific time)

Attendance is mandatory

Week Nine: March 14 - 20

Spring Break, no class

Week Ten: March 21 - 27

Database Security

Data and database security is explored in this module.

SQL Concepts

DDL; Recursive Processing; Graph

Reading

- chapter 12: “Security System of the Database Engine” (Petkovic)
- chapter 20: “Security and Administration” (Connolly& Begg) - *Optional*
- chapter 21: “Professional, Legal, and Ethical Issues” (Connolly& Begg)
- chapter 5: “Data Definition Language” (Petkovic)

- chapter 11: “Views” (Petkovic)

Assignment

Lab 4 Due 3/31

Week Eleven: March 28 – April 3

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 Due 4/7

P3 Due 4/7

Week Twelve: April 4 - 10

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. Data mining tools are discussed.

SQL Concepts

PIVOT

Reading

- chapter 31: “Data Warehousing Concepts” (Connolly& Begg)
- chapter 32: “Data Warehouse Design” (Connolly& Begg)
- chapter 34: “Data Mining” (Connolly& Begg) - *Optional*

Assignment

P4 Due 4/14

Week Thirteen: April 11 - 17

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, Saturday, 4/16, 12:30 pm Boston time (9:30 am Pacific time)

Attendance is mandatory

Week Fourteen: April 18 - 24**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 Saturday, 4/23, 11:30 am Boston time (8:30 am Pacific time)

Attendance is mandatory

Week Fifteen: April 25 - May 1

TBD

Student Accommodations

Northeastern University and the Disability Resource Center (DRC) are committed to providing disability services that enable students who qualify under Section 504 of the Rehabilitation Act and the Americans with Disabilities Act Amendments Act (ADAAA) to participate fully in the activities of the university. To receive accommodations through the DRC, students must provide appropriate documentation that demonstrates a current substantially limiting disability.

For more information, visit <http://www.northeastern.edu/drc/getting-started-with-the-drc/>.

Library Services

The Northeastern University Library is at the hub of campus intellectual life. Resources include over 900,000 print volumes, 206,500 e-books, and 70,225 electronic journals.

For more information and for Education specific resources, visit <http://subjectguides.lib.neu.edu/edresearch>.

Diversity and Inclusion

Northeastern University is committed to equal opportunity, affirmative action, diversity and social justice while building a climate of inclusion on and beyond campus. In the classroom, member of the University community work to cultivate an inclusive environment that denounces discrimination through innovation, collaboration and an awareness of global perspectives on social justice.

Please visit <http://www.northeastern.edu/oidi/> for complete information on Diversity and Inclusion

TITLE IX

Title IX of the Education Amendments of 1972 protects individuals from sex or gender-based discrimination, including discrimination based on gender-identity, in educational programs and activities that receive federal financial assistance.

Northeastern's Title IX Policy prohibits Prohibited Offenses, which are defined as sexual harassment, sexual assault, relationship or domestic violence, and stalking. The Title IX Policy applies to the entire community, including male, female, transgender students, faculty and staff.

In case of an emergency, please call 911.

Please visit www.northeastern.edu/titleix for a complete list of reporting options and resources both on- and off-campus.