Data Management and Database Design
INFO 6210 SEC 09

Instructor: Manuel Montrond / m.montrond@northeastern.edu
Office Hours: Before / After Class/By Appointment

SCHEDULED MEETING TIMES:
Time: 6:00pm – 9:30pm
Days: Wednesday
Location: Behrakis Health Science Cntr 310
Date Range: Jan 06, 2020 - Apr 25, 2020

COURSE DESCRIPTION

This course introduces database development theory and skills, including data modeling concepts, architecture, and implementation using Structured Query Language (SQL). It presents valuable skills from translating requirements into conceptual, logical and physical models. Students will learn how to implement optimized relational databases using SQL programming, learn principles of transaction management, and database security concerns. NoSQL and operationalization of machine learning models are also explored. Students will develop a database project on a business premise of their choosing. 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

Jeffrey A. Hoffer, V.Ramesh, Heikki Topi
Modern Database Management Hoffer Database Management, Tenth Edition
Prentice Hall, [ISBN-13: 978-0-13-608839-4]

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

Additional resources:

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

Students will need to download and install SQL Server database engine and SQL Server Management Studio to their local computers or in a cloud environment (Google, AWS etc). The Developer Edition of SQL Server 2012/14/16/17 is recommended. SQL Server 2017 is preferred. In addition to SQL Server, Microsoft Visio/Toad Data Modeler or any other Data Modeling tool is required.

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
- Translate business requirements to conceptual, logical, and physical model
- 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
Mid-Term	25%
Final	30%
Participation	5%
Homework	5%
Database project	35%

ATTENDANCE

This course will meet once a week on Wednesday evenings. Your attendance is paramount to your success in this class. Contact the instructor if you have a question about the class attendance. Please review the important note on attendance below:

Attendance Policy

> The Information Systems Department has a strict class attendance policy. Students who miss two or more classes will automatically receive one letter grade lower in their final grade. Students who miss three classes will receive an automatic F for the class. No exceptions are allowed for this rule.

PARTICIPATION

Participation is an important portion of your grade and should be reflected in class and online. Attendance does not equal participation.

DATABASE PROJECT

Students will form teams of three (or four max) and develop a relational database based on reading and class lectures. The project will have the following deliverables:

Deliverable	% of grade
P1. Topic and Objectives	3
P2. Database Design, Initial ERD	7
P3. Final ERD	5
P4. Database Implementation	10
P5. Presentation	10
Total Project	35%

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

P1. Topic and Objectives

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.

* One submission per team

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 an 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 10 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 like 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 Toad Data Modeler can be downloaded for free.

* One submission per team

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 regarding 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.

* One submission per team

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 the following database objects:

- at least 3 stored procedures containing input parameters output parameters or result set returned
- at least 2 views (often used for reporting purposes).
- At least 1 trigger
- The implementation must include at three of the following three items:
    - Table-level CHECK Constraints
    - Computed Columns based on a function
    - Column Data Encryption
    - Non-clustered indexes
- A simple Power BI/Tableau report/ or a GUI (a plus, not 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 documents
3) The final ERD
4) Snippets of SQL DDL statements for implementing the database
5) At least one report (using PowerBI/SSRS (not required). Other data mining tools could also be used.)

*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. 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	100-93
A-	92-90
B+	89-87
B	86-83
B-	82-80
C+	79-77
C	76-73
C-	72-70

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 1

Course Overview
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: “Database Environment and Development Process” (Hoffer, Ramesh, & Topi)

Week Two

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 2: “Modeling Data in the Organization” (Hoffer, Ramesh, & Topi)
- chapter 3: “The enhanced E-R Model” (Hoffer, Ramesh, & Topi)

Week Three

Conceptual and Logical Database Design
Data Normalization
Processes for conducting the conceptual and logical database design are discussed. This lesson presents the Normalization process for fine-tuning and validating the database

Reading

- chapter 4: “Logical Database Design and the Relational Model” (Hoffer, Ramesh, & Topi)
- chapter 3: “SQL Server Management Studio” (Petkovic)

Week Four

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

Reading

- chapter 5: “Physical Database Design and Performance” (Hoffer, Ramesh, & Topi)
- chapter 4: “SQL Components” (Petkovic)

Week Five

Introduction to SQL

This module explores the different components involved in creating SQL to implement relational databases including data definition language and an introduction to data manipulation language.

Reading

- chapter 6: “Introduction to SQL” (Hoffer, Ramesh, & Topi) from page 243-261
- chapter 5: “Data Definition Language” (Petkovic)
- chapter 7: “Modification of a Table's Contents” (Petkovic)

Week Six

Single Table processing
Using and Defining Views
This procedure explores SQL techniques to retrieve data from the database, Boolean operators, and aggregate functions. It also covers additional objects using to query the database such as views, and synonyms.

Reading

- chapter 6: “Introduction to SQL” (Hoffer, Ramesh, & Topi) from page 261 – end of chapter
- chapter 6: “Queries: SELECT Statement” (Petkovic) from page 193 -230

Week Seven

Multiple Table Processing: JOINS
We learn the concepts of retrieving and manipulating data from multiple tables using JOINS.

- chapter 6: “Queries: JOIN Operator” (Petkovic) from page 240 -255
- chapter 7: “Advanced SQL” (Hoffer, Ramesh, & Topi) page 289- 298

Week Eight

Queries
This module builds on previous modules to solidify query writing by leveraging various types of of Boolean operations, temporary tables, views, and other database objects.

Reading:

- Preparation for mid-term (all previous chapters)

Mid -Term

Week Nine

Multiple Table Processing: Subquery
We learn the concepts of retrieving and manipulating data from multiple tables using SQL concepts such as creating noncorrelated and corelated subquery

Reading

- chapter 6: “Queries: Correlated Subqueries” (Petkovic) from page 255 to the end of the chapter
- chapter 7: “Advanced SQL” (Hoffer, Ramesh, & Topi) from page 289 -310

Week Ten

Persistent Stored Module: Stored Procedure and User Defined Functions
We will cover creation of different types of routines: stored procedures, and user defined functions.

Reading

- chapter 7: “Advanced SQL” (Hoffer, Ramesh, & Topi) from page 320 -end of chapter
- chapter 8: “Stored Procedures and User-Defined Functions” (Petkovic)

Week Eleven

Transaction Management / ACID Properties & Datawarehouse
This module describes the principles of managing data consistency and integrity while processing transactions. The concepts of ACID properties are also introduced. We will also go cover the main concepts for creating a Data Warehouse and multi-dimensional modeling.
Explicit Transaction
Error Handling

Reading

- ACID Property: https://en.wikipedia.org/wiki/ACID_(computer_science)
- Concurrency control https://en.wikipedia.org/wiki/Concurrency_control
- chapter 13: “Concurrency Control” (Petkovic)

Week Twelve

Introduction to Data warehouse and multi-dimensional modeling. We will cover three different types of architectures and go over normalization and denormalization in the context of data marts.
Advanced SQL technique for data querying
CTE, RANK, window functions

Reading

- chapter 22: “Business Intelligence: An Introduction” (Petkovic)
- chapter 9: “Data Warehousing” (Hoffer, Ramesh, & Topi)

Week Thirteen

Databases Administration and Performance Tuning. We will cover tools and techniques for administration of modern database systems, and emphasize the security aspect of database administration

Reading

- chapter 11: “Database Administration” (Hoffer, Ramesh, & Topi)

Week Fourteen

Introduction to NoSQL and Big Data. Operationalization of machine learning models in SQL Server 2017 and above. We will used python 3.5 to integrate machine learning with the database

Final Exam
Reading:

- https://en.wikipedia.org/wiki/NoSQL
- https://www.thoughtworks.com/insights/blog/nosql-databases-overview
- https://en.wikipedia.org/wiki/Big_data
- https://docs.microsoft.com/en-us/sql/advanced-analytics/what-is-sql-server-machine-learning?view=sql-server-2017

Week Fifteen

Project Final presentation