Northeastern University

Course Syllabus

Course Information

Course Title: Advanced Database Management Systems
Number: DAMG 7275
Term and Year: Spring 2024
Credit Hour: 4
Course Format: On-Ground

Instructor Information

Full Name: Syed Farhan Mazhar
Email Address: s.mazhar@northeastern.edu

COURSE DESCRIPTION

The Advanced Database Management Systems course is an extension of the Data Management and Database Design course. It uses a data-centric approach to cover the concepts, theories, development and management of the architecture, technologies, security, and solutions relevant to working with large volumes of diversified data. Both the NoSQL and relational databases will be covered. This course presents many of the valuable knowledge and skills required for dealing with the data-related challenges.

LEARNING OBJECTIVES

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

- Understand and describe the Architecture of large-scale NoSQL and Relational Database Management Systems
- Design and implement Data Structure for NoSQL Databases based on the data usage pattern
- Implement and manage Data Movement, such as Transaction, Replication and Data Pipelines
- Survey major Data High Availability and Data Locality approaches
- Plan for Disaster Recovery and implement its solutions to meet the business requirements
- Architect and implement the Event-Driven Data Management
- Understand Data Governance and develop code to implement its solutions
- Explore and develop code to work with Data of Complex Relationships
- Write SQL commands to perform advanced data and table manipulation in the context of a prescribed business problem.
- Explain the basic concepts of security and the responsibilities of a database administrator.
- Write PL/SQL anonymous blocks, procedures, functions, triggers and packages to access and manipulate data.
- Create the back-end to a software application using functions, procedures, packages and triggers

PREREQUISITES

INFO 6210, INFO 5100, CSYE 6200, INFO 6205, or consent of the instructor.

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.

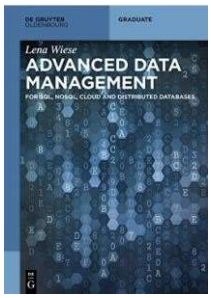
Text Book(s):

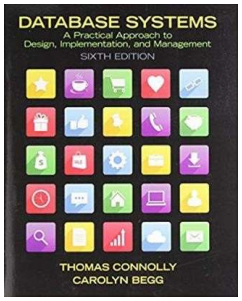
Casteel, Joan. 2013. Oracle 11g: PL/SQL Programming, 2nd Edition. Cengage Learning. ISBN-13: 9781133947363

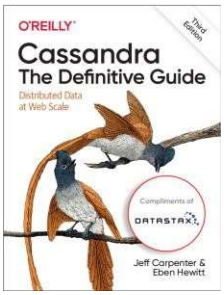
Jeff Carpenter and Eben Hewitt (2020)
Cassandra: The Definitive Guide, 3rd Edition
Published by O'Reilly Media, Inc.

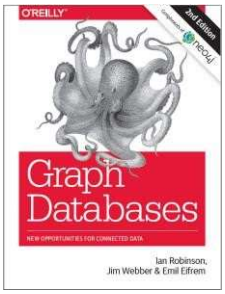
Lena Weise (2015)
Advanced Data Management: For SQL, NoSQL, Cloud and Distributed Databases
De Gruyter, [ASIN: B019LFN2MM]

Ian Robinson, Jim Webber, and Emil Eifrem (2015)
Graph Databases, 2nd Edition
Published by O'Reilly Media, Inc.









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.

Software: Students will need to download and install SQL Oracle database engine or SQL Server Management Studio to their local computers or in a cloud environment (Azure, Google, AWS etc.).
Entity-Relationship Diagram (ERD) tool of your choice is mandatory. Recommended ERD tools include draw.io, ERWin, and Microsoft Visio

EVALUATION:

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

Assessment	% Grade
Ind Lab Exercises	30%
Discussion	10%
Database project	30%
Final Exam	30%

ATTENDANCE

Your attendance is paramount to your success in this class. Contact the instructor if you have a question about the class attendance.

LATE WORK

All assignments must be submitted to the class Canvas site for the course on the due date before 11:59 pm. If an assignment is turned in late, 10% credit will be deducted from the total score for each day after the deadline. Assignments turned in more than one week late may 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 go to the link below and 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 honesty.

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	100-94
A-	93.99-90
B+	89.99-87
B	86.99-84
B-	83.99-80
C+	79.99-77
C	76.99-74
C-	73.99-70

Project

P1: Project Plan (5 points)
Select a topic, a data model and a target platform. The platform may be NoSQL, relational, or a combination of NoSQL and relational. Also determine the objectives or the scope of the project. For working with the database after its implementation, either develop an application and/or use tool(s).

* Rubric: Completeness 40%, Correctness 40%, Creativity 20%
* Individual submission is required.

P2: Design (5 points)
Create an ERD (Entity Relationship Diagram) to depict the database design. Also prepare an architecture diagram displaying how different components of the project will eventually work together to accomplish the project objectives. A design document could help you document and communicate your design.

Each team will also designate a member to present the ERD and the architecture diagram to the class. The duration of the presentation will be 3 - 5 minutes.

* Rubric: Completeness 40%, Correctness 40%, Creativity 20%
* Individual submission is required.

P3: Implementation (10 points)
Use tools and/or write code to implement your database and other components contained in your architecture diagram. An example of implementing other components is configuring a tool and/or writing some code to implement a data pipeline to do the initial data load into your project database. The submission needs to include:

1) A brief description of the implementation process

2) Screenshots which capture key steps in implementing the database and other architecture components

3) Code if you'll write some code

Submit a pdf file containing the above items. If you'll write code, please also submit the code in its original format.

* Rubric: Completeness 40%, Correctness 40%, Creativity 20%
* Individual submission is required.

P4: Work with the database Data (5 points)
Develop an application and/or use tool(s) to maintain data stored in the project database after its implementation. The minimum requirement for this assignment is to implement the ongoing data refresh. This can be done by using some code or a tool.

You don't need to create any report or visualization in this assignment. Reports and visualizations are included in P5.

Submit a pdf file containing the description of the data maintenance process and key-step screenshots of the process. If you'll write code, please also submit the code in its original format.

* Rubric: Completeness 40%, Correctness 40%, Creativity 20%
* Individual submission is required.

P5: Presentation (5 points)
Prepare a slide deck to showcase the project and use it for the presentation. A live demo during the presentation is fine. Make sure all of the project deliverables are up to date and in sync. Create reports and visualizations to demonstrate how the database data can be used for reporting and analysis.

Submit all deliverables of the project for this assignment.

Note: Only one member of a team needs to submit the presentation materials. No individual submission.

Subjects

1) Survey and describe the logical and physical architecture of the large-scale NoSQL, NewSQL and SQL databases
2) Cassandra, Azure NoSQL Cosmos DB, ArangoDB, and Relational databases will be covered
3) Design, implement and compare different data structures based on the data usage pattern
4) Write SQL, CQL, AQL and other code to work with different types of data
5) Investigate the major data movement technologies and identify a use case for each option
6) Write code to implement data engineering
7) Architect and deploy the Data High Availability and Data Locality solutions
8) Plan for disaster recovery and implement the disaster recovery solutions
9) Design the Event-Driven Data Management and write code to implement it
10) Survey and describe the data governance solutions
11) Write code to implement the data governance solutions
12) Write code to work with complex data

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 9 - 15
Course Introduction
Review of Data Hierarchy and Aggregation

Reading
- chapter 1: “Background” (Advanced Data Management)

Assignment
Lab 1 (Data Hierarchy, Data Aggregation and Reporting)

Week Two: January 16 - 22
NewSQL
Azure SQL Database
Data Pipeline
Use SQL Trigger, MERGE, and SQL Server Job to Build Data Pipelines

Reading
- chapter 2: “Relational Database Management Systems” (Advanced Data Management)
- chapter 3: “New Requirements, Not only SQL and the cloud” (Advanced Data Management)

Lambda Architecture and Databricks for Data Engineering
https://www.youtube.com/watch?v=jQC0ndhBiWg&list=PL-zncNSJGgbi81X-5pTaDdoj3r6U_XHci&index=9

Azure Databricks Training Series
https://databricks.com/p/thank-you/webinar-azure-databricks-3-part-training-series-139977

Lambda Architecture in the Cloud with Azure Databricks
https://databricks.com/session/lambda-architecture-in-the-cloud-with-azure-databricks

Assignment
P1
Lab 2 (Data Pipelines)

Week Three: January 23 - 29
Multi-Model Database Management Systems
NoSQL Database Design
Architecture Diagram

Data Model Comparison
https://www.youtube.com/watch?v=8rZn3y3uHhY&list=PL-zncNSJGgbjuUF-OyzCR6eKs4o_Fq7zj&index=1

Multi-Model Databases
https://www.youtube.com/watch?v=Qc9Kx4SuH_8&list=PL-zncNSJGgbjuUF-OyzCR6eKs4o_Fq7zj&index=2

Modern data modeling: Multi-Model approach using ArangoDB database
https://www.youtube.com/watch?v=on1l2pEEWnw&list=PL0tn-TSss6NUe2QeS5xSOe_f58Koj_9qw&index=31

Reading
- chapter 7: “Column Stores” (Advanced Data Management)

Week Four: January 30 - February 5
Cassandra Database (Columnar Data Model)
Data Modeling for Cassandra Database
CQL (Cassandra Query Language)

Introduction to Cassandra Query Language (CQL)
https://www.youtube.com/watch?v=Q1hhtNrTDrM&list=PL-zncNSJGgbgohO7bTb3ZiXmx8i6DxzE1&index=9

Reading
- chapter 4: “The Cassandra Query Language” (Cassandra: The Definitive Guide)
- chapter 5: “Data Modeling” (Cassandra: The Definitive Guide)

Assignment
P2

Week Five: February 6 - 12
Cosmos DB SQL API Database (Document Data Model)
Data Modeling for Cosmos DB SQL API Database
SQL and JSON
Import Data into Cosmos DB

Data Engineering on Azure
https://www.youtube.com/watch?v=T37mHRwzPXE&list=PL-zncNSJGgbhTsQTQ2VJgUs325j-9-qrz&index=20

Introduction to SQL for Cosmos DB SQL API
https://www.youtube.com/watch?v=Q1hhtNrTDrM&list=PL-zncNSJGgbgohO7bTb3ZiXmx8i6DxzE1&index=9

Reading
- chapter 6: “Key-valued Stores and Document Databases” (Advanced Data Management)
- chapter 10: “Distributed Database Systems” (Advanced Data Management)
- chapter 11: “Data Fragmentation” (Advanced Data Management)

Assignment
P2 Presentation (in class)
Lab 3 (Data Structures for Document Database)

Week Six: February 13 - 19
ArangoDB (Graph/Document/Key-Value Models)
Data Modeling for ArangoDB Database

ArangoDB-GraphCourse_Beginners.pdf:
https://northeastern.instructure.com/courses/64307/files/7144123?module_item_id=5261134

Introduction to ArangoDB Query Language (AQL)
https://www.youtube.com/watch?v=abfBMCZOv7k&list=PL-zncNSJGgbgBO5uREx42morYDiyYfiEf&index=11

How to Implement Schema for ArangoDB
https://www.youtube.com/watch?v=sbIVYvl7dxQ&list=PL-zncNSJGgbgBO5uREx42morYDiyYfiEf&index=14

Reading

- chapter 4: “Graph Databases” (Advanced Data Management)
- chapter 3: “Data Modeling with Graphs” (Graph Databases)

Week Seven: February 20 - 26
Property Graphs vs Knowledge Graphs

Graph Storage Models and Graph Algorithms
https://www.youtube.com/watch?v=pfbjou9NzHg&list=PL-zncNSJGgbgBO5uREx42morYDiyYfiEf&index=4

Hacktoberfest 2020 - Intro to Knowledge Graphs & reKnowledge
https://www.youtube.com/watch?v=ZZt6xBmltz4&list=PL0tn-TSss6NUe2QeS5xSOe_f58Koj_9qw&index=5

Graph Analytics with ArangoDB
https://www.youtube.com/watch?v=fpQ75pfFiE0

Assignment
Lab 4 (Graphs)

Week Eight: February 27 - March 5
Quiz 1, in class
Attendance is mandatory

Week Nine: March 6 - 12
Spring Break, no class

Week Ten: March 13 - 19
PowerShell
Riak Database (Key-Value Data Model)
Redis Database (Key-Value Data Model)
XML Database and XML

Reading

- chapter 5: “XML Databases” (Advanced Data Management)
- chapter 24: “Distributed DBMSs – Concepts and Design” (Database Systems: A Practical Approach to Design, Implementation, and Management)

Assignment
P3

Week Eleven: March 20 - 26
SQL CURSOR
SQL and Graphs
Use Iterative Approach and CURSOR for processing graphs
SQL Node Table, Edge Table and MATCH()
Dynamic SQL

Why Graph Model Has Better Query Performance Than Relational Model
https://www.youtube.com/watch?v=sbQVG-2-i70&list=PL-zncNSJGgbjuUF-OyzCR6eKs4o_Fq7zj&index=11

Reading
Graph processing with SQL Server and Azure SQL Database:
https://docs.microsoft.com/en-us/sql/relational-databases/graphs/sql-graph-overview?view=sql-server-ver15

Assignment
P4
Lab 5 (Dynamic SQL and SQL Graphs)

Week Twelve: March 27 - April 2
New Database Technologies
Data High Availability, Data Locality and Disaster Recovery

Reading

- chapter 12: “Replication and Synchronization” (Advanced Data Management)
- chapter 14: “Further Database Technologies” (Advanced Data Management)
- chapter 15: “Concluding Remarks” (Advanced Data Management)
- chapter 25: “Distributed DBMSs – Advanced Concepts” (Database Systems: A Practical Approach to Design, Implementation, and Management)

Week Thirteen: April 3 - 9
Quiz 2, in class
Attendance is mandatory

Week Fourteen: April 10 - 16
Project Presentations, in class
Attendance is mandatory

Assignment
P5

Week Fifteen: April 17 - 23
T