

## Course Syllabus

---

### **Advanced Database Management Systems**

DAMG 7275, CRN:

Lectures: Saturdays 4 pm to 7 pm Boston (1 pm - 4 pm Pacific)

Office Hours: TBD

September 4 to December 8

Zoom online meeting room

Meeting URL:

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

Simon W. Wang

425.829.2503

[w.wang@neu.edu](mailto:w.wang@neu.edu)

[swpwang@uw.edu](mailto:swpwang@uw.edu)

### **INSTRUCTOR BIOGRAPHY**

This is Wuping Simon Wang, your instructor. I go by Simon.

I did PHD coursework and earned my MS in Computer Science. I was certified by Microsoft and Oracle. In my career, I have worked as a Data Modeler, Software Development Engineer, Database Designer, Cloud Engineer, Production DBA, Software Technical Architect, or IT Director at places like Microsoft, Starbucks, AT&T Wireless, professional firms, and IBM.

I do data work in industry as a data and database professional specialized in both relational and NoSQL technologies in addition to teaching data management, data architecture, and data engineering graduate courses.

### **TEACHING ASSISTANT INFORMATION**

TBD

### **COURSE PREREQUISITES**

DAMG 6210, minimum B; or see the catalog.

### **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.

### **COURSE LEARNING OUTCOMES**

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

### **COURSE CONTENT**

Module 1 - Overview of Data Architecture, Data Engineering, and Data Analysis

- Module 2 - Data Pipelining and Cloud Data Platforms
- Module 3 - Multi-Model Data System Design and Architecture
- Module 4 - Columnar Data Model and Cassandra
- Module 5 - Document Data Model and Azure Multi-Model Cosmos DB
- Module 6 - Graph Data Model and Multi-Model ArangoDB
- Module 7 - Scripting and MS PowerShell
- Module 8 - XML and XML Database
- Module 9 - SQL Multi-Model and Advanced SQL Programming
- Module 10 - Performance Tuning, High Data Availability, and Disaster Recovery

## BOOKS

Lena Weise

*Advanced Data Management: For SQL, NoSQL, Cloud and Distributed Databases*  
De Gruyter, [ASIN: B019LFN2MM]

Connolly, T. M. & Begg, C. E.

*Database Systems: A Practical Approach to Design, Implementation, and Management* (6<sup>th</sup> Edition) Addison-Wesley Publishing, [ISBN-10: 0-13-294326-3]

The 4<sup>th</sup> or 5<sup>th</sup> Edition is also acceptable.

Jeff Carpenter and Eben Hewitt

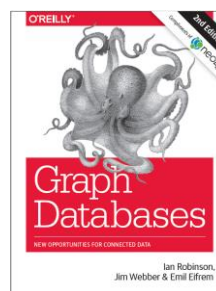
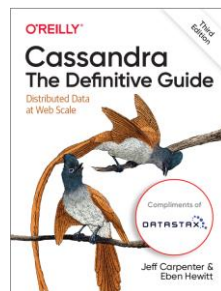
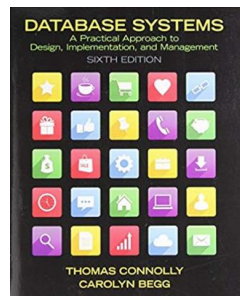
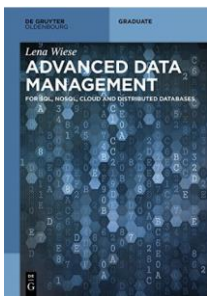
*Cassandra: The Definitive Guide, 3rd Edition*  
Published by O'Reilly Media, Inc.

Ian Robinson, Jim Webber, and Emil Eifrem

*Graph Databases, 2nd Edition*  
Published by O'Reilly Media, Inc.

Martin Kleppmann

*Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*  
1st Edition, Published by O'Reilly Media, Inc.



## Additional Course Materials

YouTube Channel for Database Design:

<https://www.youtube.com/@databasedesign8172/playlists>

YouTube Channel for SQL Development:

<https://www.youtube.com/@sqldevelopment7439/playlists>

YouTube Channel for Data Mining and Database Management:

<https://www.youtube.com/@dataarchitectureandmanagem2407/playlists>

## 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, GraphDB, 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, SPARQL, 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 and/or use tools to implement data engineering and data pipelining
- 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 and/or use tools to implement the data governance solutions
- 12) Write code to work with complex data, such as graphs

## Project

### P1: Project Plan (2 points)

Select a topic, data models, 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 (8 points)

Create ERDs (Entity Relationship Diagrams) 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 can help you document and communicate your design.

Each team will also designate a member to present the ERDs and the architecture diagram to the class.

- \* 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 and/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 is required.

## **Course Schedule/Topics Covered**

**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: September 4 - 8**

#### **Class One:**

#### **Course Introduction**

#### **Review of Data Hierarchy and Aggregation**

#### **Reading**

- chapter 1: "Background" (Advanced Data Management)

#### **Assignment**

Team Formation and Project Planning

### **Week Two: September 9 - 15**

#### **Class Two:**

#### **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=iQC0ndhBiWg&list=PL-zncNSJGgbi81X-5pTaDdoj3r6U\\_XHci&index=9](https://www.youtube.com/watch?v=iQC0ndhBiWg&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

Lab 1 (Data Hierarchy, Data Aggregation and Reporting)

## Week Three: September 16 - 22

### Class Three:

**Multi-Model Database Management Systems**

**NoSQL Database Design**

**Architecture Diagram**

Data Model Comparison

[https://www.youtube.com/watch?v=8rZn3y3uHhY&list=PL-zncNSJGgbi81X-5pTaDdoj3r6U\\_XHci&index=1](https://www.youtube.com/watch?v=8rZn3y3uHhY&list=PL-zncNSJGgbi81X-5pTaDdoj3r6U_XHci&index=1)

Multi-Model Databases

[https://www.youtube.com/watch?v=Qc9Kx4SuH\\_8&list=PL-zncNSJGgbi81X-5pTaDdoj3r6U\\_XHci&index=2](https://www.youtube.com/watch?v=Qc9Kx4SuH_8&list=PL-zncNSJGgbi81X-5pTaDdoj3r6U_XHci&index=2)

Modern data modeling: Multi-Model approach using ArangoDB database

[https://www.youtube.com/watch?v=on1I2pEEWnw&list=PL0tn-TSss6NUe2QeS5xSOe\\_f58Koj\\_9qw&index=31](https://www.youtube.com/watch?v=on1I2pEEWnw&list=PL0tn-TSss6NUe2QeS5xSOe_f58Koj_9qw&index=31)

## Reading

- chapter 2: “Data Models and Query Languages” (Designing Data-Intensive Applications)

## Assignment

P1

## Week Four: September 23 - 29

### Class Four:

**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-zncNSJGgbi81X-5pTaDdoj3r6U\\_XHci&index=9](https://www.youtube.com/watch?v=Q1hhtNrTDrM&list=PL-zncNSJGgbi81X-5pTaDdoj3r6U_XHci&index=9)

## Reading

- chapter 7: “Column Stores” (Advanced Data Management)
- chapter 4: “The Cassandra Query Language” (Cassandra: The Definitive Guide)
- chapter 5: “Data Modeling” (Cassandra: The Definitive Guide)

## Assignment

Lab 2 (Data Pipelines)

## Week Five: September 30 - October 6

### Class Five:

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-zncNSJGqbhTsQTQ2VJgUs325j-9-grz&index=20>

Introduction to SQL for Cosmos DB SQL API

<https://www.youtube.com/watch?v=Q1hhtNrTDrM&list=PL-zncNSJGqbqohO7bTb3ZiXmx8i6DxzE1&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)

## Week Six: October 7 - 13

### Class Six:

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](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-zncNSJGqbgBO5uREx42morYDiyYfiEf&index=11>

### Reading

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

### Assignment

Lab 3 (Data Structures for Document Database)

P2

## Week Seven: October 14 - 20

### Class Seven:

Property Graphs vs Knowledge Graphs

SPARQL

GraphDB

Graph Storage Models and Graph Algorithms

<https://www.youtube.com/watch?v=pfjou9NzHg&list=PL-zncNSJGqbgBO5uREx42morYDiyYfiEf&index=4>

## Week Eight: October 21 - 27

### Class Eight:

**Quiz 1, in class**

**Attendance is mandatory**

## Week Nine: October 28 - November 3

### Class Nine:

How to Implement Schema for ArangoDB

<https://www.youtube.com/watch?v=sblVYvI7dxQ&list=PL-zncNSJGgbgBO5uREx42morYDiyYfiEf&index=14>

Hacktoberfest 2020 - Intro to Knowledge Graphs & reKnowledge

[https://www.youtube.com/watch?v=ZZt6xBmltz4&list=PL0tn-TSss6NUe2QeS5xSOe\\_f58Koj\\_9qw&index=5](https://www.youtube.com/watch?v=ZZt6xBmltz4&list=PL0tn-TSss6NUe2QeS5xSOe_f58Koj_9qw&index=5)

Graph Analytics with ArangoDB

<https://www.youtube.com/watch?v=fpQ75pfIE0>

### Reading

- chapter 4: "Encoding and Evolution" (Designing Data-Intensive Applications)

### Assignment

Lab 4 (Graphs)

## Week Ten: November 4 - 10

### Class Ten:

PowerShell

Riak Database (Key-Value Data Model)

Redis Database (Key-Value Data Model)

XML Database and XML

SQL CURSOR

SQL and Graphs

Use Iterative Approach and CURSOR for processing graphs

### 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)

## Week Eleven: November 11 - 17

### Class Eleven:

SQL Node Table, Edge Table and MATCH()

Dynamic SQL

New Database Technologies

Data High Availability, Data Locality and Disaster Recovery

Why Graph Model Has Better Query Performance Than Relational Model

[https://www.youtube.com/watch?v=sbQVG-2-i70&list=PL-zncNSJGgbjuUF-OyzCR6eKs4o\\_Fq7zi&index=11](https://www.youtube.com/watch?v=sbQVG-2-i70&list=PL-zncNSJGgbjuUF-OyzCR6eKs4o_Fq7zi&index=11)

## 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>

### 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)
- chapter 12: “The Future of Data Systems” (Designing Data-Intensive Applications)

### Assignment

P3

Lab 5 (Dynamic SQL and SQL Graphs)

## Week Twelve: November 18 - 24

### Class Twelve:

**Quiz 2, in class**

**Attendance is mandatory**

### Assignment

P4

## Week Thirteen: November 25 - December 1

**No Class (Fall Break)**

## Week Fourteen: December 2 - 8

### Class Thirteen:

**Project Presentations, in class**

**Attendance is mandatory**

### Assignment

P5

## EVALUATION

Course work balances between theory and practice. The grading components are listed below.

Assessment	% Grade
2 Quizzes	30%
5 Labs	40%
Project	30%

## GRADING SCALE

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

### Attendance/Late Work Policy

#### Attendance Policy

Students registered in MGEN courses (INFO, CSYE, and DAMG) are allowed **a maximum of 2 absences per course, with 3 or more absences resulting in an automatic 'F' for that course.** Students are expected to inform their instructors of any absences in advance of the class; if a student is sick long-term or experiences a medical issue that prevents class attendance, it is strongly encouraged that they speak with their Academic Advisor ([coe-mgen-gradadvising@northeastern.edu](mailto:coe-mgen-gradadvising@northeastern.edu)) to learn more about the Medical Leave of Absence. Should a student anticipate being unable to attend 3 or more classes, they should discuss their situation with their Academic Advisor to explore other types of leave in accordance with the University's academic and global entry expectations. International students should review the Office of Global Services webpage to understand their visa compliance requirements.

Teaching Assistants (TAs) or Instructional Assistants (IAs) will be present at each class to collect student attendance.

#### Late Work Policy

Students must submit assignments by the deadline in the time zone noted in the syllabus. Students must communicate with the faculty prior to the deadline if they anticipate work will be submitted late. Work submitted late without prior communication with faculty will not be graded.

### End-of-Course Evaluation Surveys

Your feedback regarding your educational experience in this class is particularly important to the College of Engineering. Your comments will make a difference in the future planning and presentation of our curriculum.

At the end of this course, please take the time to complete the evaluation survey at <https://neu.evaluationkit.com>.

Your survey responses are **completely anonymous and confidential**. For courses 6 weeks in length or shorter, surveys will be open one week prior to the end of the courses; for courses greater than 6 weeks in length, surveys will be open for two weeks. An email will be sent to your Northeastern University Mail account notifying you when surveys are available.

### Academic Integrity

A commitment to the principles of academic integrity is essential to the mission of Northeastern University. The promotion of independent and original scholarship ensures that students derive the most from their educational experience and their pursuit of knowledge. Academic dishonesty violates the most fundamental values of an intellectual community and undermines the achievements of the entire University.

As members of the academic community, students must become familiar with their rights and responsibilities. In each course, they are responsible for knowing the requirements and restrictions regarding research and writing, examinations of whatever kind, collaborative work, the use of study aids, the appropriateness of assistance, and other issues. Students are responsible for learning the conventions of documentation and acknowledgment of sources in their fields. Northeastern University expects students to complete all examinations, tests, papers, creative projects, and assignments of any kind according to the highest ethical standards, as set forth either explicitly or implicitly in this Code or by the direction of instructors.

Go to <http://www.northeastern.edu/osccr/academic-integrity-policy/> to access the full academic integrity policy.

### **MGEN Student Feedback**

Students who would like to provide the MGEN unit with anonymous feedback on this particular course, Teaching Assistants, Instructional Assistants, professors, or to provide general feedback regarding their program, may do so using this survey: [https://neu.co1.qualtrics.com/jfe/form/SV\\_cTIAbH7ZRaaW0Ki](https://neu.co1.qualtrics.com/jfe/form/SV_cTIAbH7ZRaaW0Ki)

### **University Health and Counseling Services**

As a student enrolled in this course, you are fully responsible for assignments, work, and course materials as outlined in this syllabus and in the classroom. Over the course of the semester if you experience any health issues, please contact UHCS.

For more information, visit <https://www.northeastern.edu/uahcs>.

### **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 <https://drc.sites.northeastern.edu>.

### **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 <https://library.northeastern.edu>  
Network Campus Library Services: [Northeastern University Library Global Campus Portals](#)

### **24/7 Canvas Technical Help**

For immediate technical support for Canvas, call 617-373-4357 or email [help@northeastern.edu](mailto:help@northeastern.edu)  
Canvas Student Resources: <https://canvas.northeastern.edu/student-resources/>

For assistance with my Northeastern e-mail, and basic technical support:

Visit ITS at <https://its.northeastern.edu>

Email: [help@northeastern.edu](mailto:help@northeastern.edu)

ITS Customer Service Desk: 617-373-4357

### **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, members 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 <https://www.northeastern.edu/ouec> for a complete list of reporting options and resources both on- and off-campus.