



DAMG 7245: Big-Data Systems and Intelligence Analytics **[SPRING 2025]**

Course Information

Course Title: Big-Data Systems and Intelligence Analytics

Course Number: DAMG 7245

Term and Year: Spring 2025

Credit Hour: 4

Course Format: Traditional

Oakland, CA	Silicon Valley, CA
CRN: 40292 Location: TBD Time: R 10:35 am - 1:55 pm PT	CRN: 40482 Location: TBD Time: S 1:30 pm - 4:50 pm PT

Instructor Information

Full Name: Raja Alomari, PhD

Email Address: r.alomari@northeastern.edu

Office Hours: Wed 4-5 PM (Link: [Meeting Link - MS Teams](#)) (Jan 6, 2024 - April 26, 2024).

Appointment: (Link: [Book time with Dr. Alomari](#)) (Jan 6, 2024 - April 26, 2024).

Teaching Assistant Information

Oakland, CA	Silicon Valley, CA
TA: TBD Location: TBD TA Office Hours: TBD Meeting Link: TBD Meeting ID: TBD Passcode: TBD	TA: TBD Location: TBD TA Office Hours: TBD Meeting Link: TBD Meeting ID: TBD Passcode: TBD

Course Prerequisites

DAMG 6105 with a minimum grade of B or INFO 6105 with a minimum grade of B

Course Description

Big-Data Systems and Intelligence Analytics Offers students an opportunity to learn a hands-on approach to understanding how large-scale data sets are processed and how data science algorithms are adopted in the industry through case studies and labs. This project-based course builds on INFO 7390 and focuses on enabling students with tools and frameworks primarily to build end-to-end applications. The course is divided into three parts: building the data pipeline for data science, implementing data science algorithms, and scaling and deploying data science algorithms.

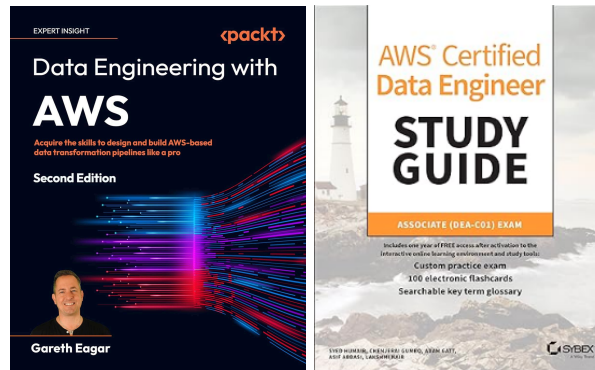
Course Learning Outcomes

- Data Pipeline Overview: Students will design and analyze data pipelines, including ETL vs. ELT, integration, and storage, while evaluating on-prem vs. cloud solutions and addressing data governance and cost considerations.
- Data Sources and Formats: Students will compare schema-based and schema-less data sources, and assess formats like CSV, JSON, and Parquet for different use cases.
- Data Storage Solutions: Students will differentiate between data warehouses, databases, and data lakes, exploring AWS solutions such as Redshift, DynamoDB, and S3.
- Feature Engineering: Students will apply techniques for data cleansing, transformation, encoding, and feature extraction, handling various data types and addressing feature scaling and normalization.
- Serverless Data Pipeline: Students will build a serverless data pipeline with AWS Glue, Athena, and QuickSight, focusing on data ingestion, storage, and automation.
- Machine Learning Pipeline: Students will create a machine learning pipeline using AWS SageMaker, covering model training, evaluation, deployment, and MLOps principles.
- Serverless ML Pipeline: Students will design a serverless ML pipeline with AWS services, including SageMaker and Athena, and explore data science using Jupyter notebooks.
- Data Governance and Anonymization: Students will understand data governance and anonymization, focusing on PII handling and compliance with GDPR and CCPA/CPRA.
- Real-Time Processing: Students will develop real-time processing pipelines with AWS services, exploring Kafka and Kinesis, and managing low-latency and autoscaling.
- On-Prem Data & ML Pipeline - Orchestration: Students will implement orchestration strategies using DAGs and Jupyter notebooks with Anaconda.
- Job Roles Overview: Students will learn the roles and responsibilities of Data Engineers, Feature Engineers, Data Scientists, MLOps Engineers, and BI Analysts.
- Orchestration Technologies: Students will explore and compare orchestration tools like Airflow, Dagster, and Argo for managing data and ML workflows.

Required Tools and Course Textbooks.

The nature of this course requires multiple textbooks. The following list includes recommended reading resources, which are not mandatory.

- Data Engineering with AWS: Acquire the skills to design and build AWS-based data transformation pipelines like a pro
- AWS Certified Data Engineer Study Guide: Associate (DEA-C01) Exam



Tools: The following access and tools will be required to perform your assignments:

- Access to a Linux system (using University credentials).
- Access to AWS, Azure, or GCP. Signup for free access if you do not have one.

Course Schedule/Topics Covered.

The following is the tentative schedule for this course. Please note that the provided date in each row is for the start of the week and not the actual class meeting date.

Order	Week	Module	Topics	Activities & Quizzes
1 + 2	Jan 6 Jan 13	M1	Course overview Data pipeline overview: <ul style="list-style-type: none"> • data sources, ingest, ETL vs ELT, integration, data quality, data store, lineage, streaming vs batching, on-prem vs cloud, orchestration and workflow, Consumers of data pipelines, query engines, data lake, data governance, retention, monitoring, size and cost considerations. 	Hands-on: Load a dataset, compute aggregates, and store in desired format. Coding assignment on GitHub Classroom. Requires github account. Assignment M1-1 out Quiz M1-1
3	Jan 20	M2	Data Sources and Formats: <ul style="list-style-type: none"> • Schema vs schema-less. • Selected formats overview: csv, tsv, Json, protobuf, parquet, xml, etc. Data Store: <ul style="list-style-type: none"> • Data Warehouse vs Database vs Data lake. 	Hands-on: Load datasets in at least three distinct formats, compute aggregates, and store output in one destination under a unified schema. Coding assignment on GitHub Classroom.

			<ul style="list-style-type: none"> Structured, semi-structured, unstructured data. <p>Tech overview: AWS Redshift, AWS Data lake, AWS DynamoDB, AWS RDS, AWS Aurora, S3.</p>	<p>Requires github account.</p> <p>Assignment M1-1 due Assignment M2-1 out</p> <p>Quiz M2-1</p>
4 + 5	Jan 27 Feb 3	M3	<p>Feature Engineering:</p> <ul style="list-style-type: none"> Data cleansing, transformation, encoding, one-hot-encoding, feature extraction, feature selection, dimensionality reduction, missing data, feature normalization and scaling, Numerical, categorical, time-series, and text features. <p>Tech overview: Amazon Feature Store.</p>	<p>Hands-on: Deploy M2-1 codebase on an EC2 instance and schedule the pipelines with cron jobs.</p> <p>Submission: screen recording + in class demo.</p> <p>Requires AWS account (free).</p> <p>Assignment M2-1 due Assignment M3-1 out</p> <p>Quiz M3-1</p>
6	Feb 10	M4	<p>Case study: Serverless Data pipeline with focus on AWS.</p> <ul style="list-style-type: none"> Overview of Serverless pipelines. AWS Glue, Workflow, Ingest, Store, crawler, Athena, QuickSight. 	<p>Hands-on: Build your data pipeline using AWS Glue serverless services (via AWS console).</p> <p>Submission: screen recording + in class demo.</p> <p>Requires AWS account (free).</p> <p>Assignment M3-1 due Assignment M4-1 out</p> <p>Quiz M4-1</p>

7	Feb 17	M5-1	Machine Learning Pipeline: <ul style="list-style-type: none"> Overview of ML pipelines. Overview, Model Exploration, training, evaluation, model selection, and fine-tuning. 	Hands-on: Use Sagemaker to train, validate and test a model of your choice. Show full ML pipeline flow. Submission: Screen recording + in-class demo. Requires AWS account (free). Assignment M4-1 due Assignment M5-1 out
8	Feb 24		Midterm Exam	Optional project proposals (due before final). 25% Extra Credit.
9	Mar 3		Spring Break	
10	Mar 10	M5-2	Machine Learning Pipeline - Cont'd: <ul style="list-style-type: none"> Model deployment, model aging, training vs validation vs testing, monitoring, and feedback loop. MLOps. Batch vs Real Time. Tech overview: Building ML pipeline using Sagemaker.	Quiz M5-1
11	Mar 17	M5-3	Case study: Serverless Machine Learning pipeline: <ul style="list-style-type: none"> AWS Sagemaker, Workflow, Ingest, Store, Athena, QuickSight. Intro to data science with serverless Jupyter notebook. - In-class Demos for Assignments.	Hands-on: Expand on assignment M5-1 to show model deployment, live inference, and model monitoring leveraging AWS Sagemaker. Submission: Screen recording + in class demo. Requires AWS account (free). Assignment M5-1 due Assignment M5-2 Quiz M5-2

12	Mar 24	M6	<p>Microservices Architecture:</p> <ul style="list-style-type: none"> • Overview of Microservices Architecture. • Monolithic vs Microservices. • Challenges and benefits of MS Architecture. • Main AWS services to build MS Architecture. 	Quiz M6-1
13	Mar 31	M7	<p>Data governance & anonymization:</p> <ul style="list-style-type: none"> • PII and data access in production for ML pipelines. • GDPR, CCPA/CPRA implications on MLOps and data scientists. <p>Case study: AWS Data Lake access control (lake formation).</p>	<p>Write a report (3 to 10 pages with text font of 12 points) summarizing the requirements and comparing GDPR with CCPA/CPRA.</p> <p>Submission: pdf.</p> <p>Assignment M5-2 due Assignment M7-1 out</p> <p>Quiz M7-1</p>
14	April 7	M8	<p>Real time processing:</p> <ul style="list-style-type: none"> • Streaming (Kafka, Kinesis), Low-latency, message size, events, streaming, near-real time, SLA, SLO, SLI. Autoscaling. Sharding. <p>Tech overview: Kafka, Kinesis streaming, Kinesis firehose, Lambda, EventBridge (EventBus) SQS.</p>	<p>Hands-on: Build a realtime pipeline to ingest realtimes data using AWS serverless services leveraging at least one AWS Kinesis stream, one kinesis firehose, one lambda function, and an object store.</p> <p>Submission: Screen recording + in-class demo.</p> <p>Requires AWS account (free).</p> <p>Assignment M7-1 due Assignment M8-1 due</p> <p>Quiz M8-1</p>

15	April 14		Demo Day	Demo all assignments that have not been presented. Demo Projects.
16	April 21		Final Exam	
<p>Note: If the lecture falls on a holiday or canceled for any reason, the topic moves directly to the week after.</p> <p>Note: Quizzes use Canvas (Quiz functionality). Quizzes will be scheduled outside the classroom.</p>				

Assignment Grading

- Assignments: 25%
- Quizzes: 15%
- Midterm: 25%
- Final Exam: 35%
- Optional Extra Credit project 25%.

Grading Scale

92 - 100% A	86 - 88% B+	76 - 78% C+	Below 70% F
	82 - 85% B	72 - 75% C	
89 - 91% A-	79 - 81% B-	70 - 71% C-	

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) 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. **No late work will be accepted.** Each student is responsible for proper planning and submitting on time.

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

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/uHCS>.

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

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

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.

Syllabus Changes: The instructor reserves the right to modify the syllabus, including course content, assignments, grading criteria, and schedule, as deemed necessary to enhance the learning experience or respond to unforeseen circumstances. Any changes will be communicated to students in a timely manner through official channels. It is the responsibility of the students to stay informed about such updates. Latest syllabus is posted under the Syllabus section in Canvas.