湖北工業大學
HUBEI UNIVERSITY OF TECHNOLOGY

Course Title	Data Engineering Systems
Course Code	CMPT 3831
Semester	Summer 2026
Course Length	4 Weeks, 60 Contact Hours
Credits	4
Instructor	Wei Chao
Office	Science and Technology Building A403
Email	9030885373@qq.com
Prerequisite	CMPT 2141 Data Structures and Algorithms Proficiency in Python programming.

Course Description:

This course explores the technical foundations of modern data management, transitioning from the mathematical rigor of relational systems to the flexible architectures of non-relational environments. Students will master the complete data pipeline: programmatic acquisition via APIs, transformation using high-performance numerical libraries, and persistence in both SQL and NoSQL stores. The course emphasizes the application of data-first thinking, enabling students to build robust systems that can ingest, clean, and derive insights from diverse data formats.

Delivery Mode: In-Person, Face-to-Face Instruction

Course Goals:

Students who successfully complete this course will demonstrate competency in the following general education core goals:

- **Critical Thinking Skills** – Students will engage in analytical thinking, demonstrating the ability to critically evaluate, synthesize, and apply knowledge to complex problems, and construct well-reasoned solutions and arguments.
- **Independent Research and Inquiry** – Students will conduct independent research, utilizing academic resources to explore relevant topics, formulating research questions, analyzing data, and presenting findings in a coherent, scholarly manner.
- **Problem-Solving and Application** – Students will apply theoretical concepts and methodologies learned in the course to real-world problems, demonstrating the ability to develop practical solutions informed by academic inquiry.
- **Global and Cultural Awareness** – Students will gain awareness of the global and cultural contexts relevant to the course, appreciating diverse perspectives and considering the implications of their studies in a broader, international context.

Student Learning Outcomes:

Upon completion of this course, students will be able to:

- Architect and query relational databases using complex relational algebra and SQL;
- Implement high-efficiency data manipulation scripts using the Python scientific stack;
- Design and evaluate schema models for both structured (RDBMS) and semi-structured (JSON/Document) data;
- Deploy automated ETL (Extract, Transform, Load) workflows to handle heterogeneous data sources.

Textbooks/Supplies/Materials/Equipment/ Technology or Technical Requirements:

Textbooks:

- Wes McKinney. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Jupyter, 3rd Ed.* O'Reilly Media, Inc., 2022. (abbr. *WM*)
- Hector Garcia-Molina; Jeffrey D. Ullman; Jennifer Widom. *Database Systems: The Complete Book, 2nd Ed.* Pearson, 2012. (abbr. *GUW*)

Software Requirements (open-source and freely available):

- Python 3.9+ via Anaconda distribution (including Jupyter Notebook)
- Key libraries: pandas, NumPy, Matplotlib, and SQLAlchemy
- Relational database: PostgreSQL
- Non-relational database: MongoDB

Course Requirements:

Engineering Labs (30%): This component consists of four technical laboratories designed to bridge theoretical concepts with engineering reality.

- Lab 1 focuses on high-performance numerical computing and array manipulation using the Python scientific stack.
- Lab 2 requires students to build a robust ETL (Extract, Transform, Load) pipeline, handling messy real-world data from disparate formats like nested JSON and unstructured text.
- Lab 3 transitions into relational systems, where students must implement a complex schema in PostgreSQL and execute optimized queries involving multi-way joins and sub-selections.
- Lab 4 introduces modern NoSQL paradigms, requiring students to perform CRUD operations and aggregations on semi-structured document stores.

Theoretical Problem Sets (15%): Three written assignments covering formal logic, including relational algebra, functional dependencies, and normal form proofs.

Midterm Exam (25%): The midterm examination serves as a checkpoint for the first half of the semester. It is divided into two parts: a programming proficiency segment and a theoretical modeling segment. Students must demonstrate their ability to manipulate data structures efficiently in Python and transform a business problem into a valid Entity-Relationship (ER) diagram.

Capstone Project (30%): A final team project requiring the ingestion of a raw dataset, storage in a hybrid environment, and a full analytical report. Students must identify a raw, large-scale dataset (e.g., from public government APIs or financial repositories), architect a storage solution that may involve both relational and non-relational components, and perform a deep-dive analysis.

Assessments: Activity	**Percent Contribution**
Engineering Labs	30%
Theoretical Problem Sets	15%
Midterm Exam	25%
Capstone Project	30%

Grading:

Final grades will be based on the sum of all possible course points as noted above.

Grade	**Percentage of available points**
A	94-100
A-	90-93
B+	87-89
B	84-86
B-	80-83
C+	77-79
C	74-76
C-	70-73
D	64-69
D-	60-63
F	0-59

Course Schedule:
The schedule of activities is subject to change at the reasonable discretion of the instructor. Minor changes will be announced in class, major ones provided in writing.

CMPT 3831 Schedule			
Lecture	Topic	Readings	Activities
L1	The Data Ecosystem: Evolution and Overview	*GUW* 1.1; *WM* 1.1	
L2	Development Environment: IPython and Jupyter	*WM* 2.1 - 2.2	
L3	Python Structures: Lists, Dictionaries, and Files	*WM* 3.1, 3.3	
L4	NumPy I: The ndarray and Vectorized Arithmetic	*WM* 4.1	Lab 1
L5	NumPy II: Array Programming and Logic	*WM* 4.3 - 4.4	
L6	pandas I: Series, DataFrames, and Indexing	*WM* 5.1 - 5.2	PS 1
L7	pandas II: Descriptive Statistics and Summaries	*WM* 5.3	
L8	Data Ingestion: Text Formats, JSON, and Web APIs	*WM* 6.1, 6.3	Lab 2
L9	Data Preparation: Handling Missing Values	*WM* 7.1	
L10	Data Transformation and String Manipulation	*WM* 7.2, 7.4	
L11	Data Wrangling: Joins, Merges, and Reshaping	*WM* 8.2 - 8.3	
L12	Midterm Review & Discussion		
L13	**Midterm Examination**		Exam
L14	Relational Model: Relations and Schemas	*GUW* 2.1 - 2.3	
L15	Relational Algebra: Operations on Bags	*GUW* 2.4, 5.1	
L16	The Entity/Relationship (ER) Model	*GUW* 4.1 - 4.3	Lab 3

L17	ER Diagrams to Relational Designs	*GUW* 4.5	
L18	Functional Dependencies and Design Theory	*GUW* 3.1 - 3.2	PS 2
L19	Normalization: 3NF and BCNF Decomposition	*GUW* 3.3 - 3.5	
L20	SQL I: Basic Queries and Modifications	*GUW* 6.1, 6.5	
L21	SQL II: Multi-relation Joins and Subqueries	*GUW* 6.2 - 6.3	Lab 4
L22	NoSQL: Semistructured Data and XML/JSON	*GUW* 11.1 - 11.2	
L23	Document Databases: MongoDB Operations	Handout	PS 3
L24	Data Aggregation and Group Operations	*WM* 10.1 - 10.2	
L25	**Final Project Presentation**		Project

Accommodation Statement:

Academic accommodations may be made for any student who notifies the instructor of the need for an accommodation. It is imperative that you take the initiative to bring such needs to the instructor's attention, as he/she is not legally permitted to inquire. Students who may require assistance in emergency evacuations should contact the instructor as to the most appropriate procedures to follow.

Academic Integrity Statement

Each student is expected to maintain the highest standards of honesty and integrity in academic and professional matters. The University reserves the right to take disciplinary action, up to and including dismissal, against any student who is found guilty of academic dishonesty or otherwise fails to meet the standards. Any student judged to have engaged in academic dishonesty in coursework may receive a reduced or failing grade for the work in question and/or for the course.

Academic dishonesty includes, but is not limited to, dishonesty in quizzes, tests, or assignments; claiming credit for work not done or done by others; hindering the academic work of other students; misrepresenting academic or professional qualifications within or without the University; and nondisclosure or misrepresentation in filling out applications or other University records.

Other Items:

Attendance and Expectations

All students are required to attend every class, except in cases of illness, serious family concerns, or other major problems. We expect that students will arrive on time, be prepared to listen and participate as appropriate, and stay for the duration of a meeting rather than drift in or out casually. In short, we anticipate that students will show professors and fellow students maximum consideration by minimizing the disturbances that cause interruptions in the learning process. This means that punctuality is a must, that cellular phones be turned off, and that courtesy is the guiding principle in all exchanges among students and faculty. You will be responsible for the materials and ideas presented in the lecture.

Assignment Due Dates

All written assignments must be turned in at the time specified. Late assignments will not be accepted unless prior information has been obtained from the instructor. If you

believe you have extenuating circumstances, please contact the instructor as soon as possible.

Make-Up Work
The instructor will not provide students with class information or make-up assignments/quizzes/exams missed due to an unexcused absence. Absences will be excused and assignments/quizzes/exams may be made up only with written documentation of an authorized absence. Every effort should be made to avoid scheduling appointments during class. An excused student is responsible for requesting any missed information from the instructor and setting up any necessary appointments outside of class.

Access, Special Needs, and Disabilities
Please notify the instructor at the start of the semester if you have any documented disabilities, a medical issue, or any special circumstances that require attention, and the school will be happy to assist.