

The Hong Kong University of Science and Technology (Guangzhou)

UG Course Syllabus Template

[Course Title] Complex Data Management

[Course Code] DSAA 4032

[No. of Credits] 3

[Any pre-/co-requisites] DSAA 2031 Database Management Systems

Name: Yuyu Luo

Email: yuyuluo@hkust-gz.edu.cn

Office Hours: Wednesday 14:00-16:00, E2-615

Course Description

This course is an advanced course for understanding the diversity of data types (e.g., graph data, time series data, spatial data, video data) in data science: why they exist, how to store them, how to query them, and how to build scalable and efficient solutions for managing these data types. It will also introduce how to make predictions, especially for time-series data, which has important applications for financial data and health care data.

Intended Learning Outcomes (ILOs)

By the end of this course, students should be able to:

1. Develop the management methods for different data types.
2. Store different types of data.
3. Index different types of data.
4. Execute queries for different types of data.
5. Predict time-series data.

Weekly Schedule and Weekly ILOs

Week	Topics	Weekly ILOs
1	Course Introduction & Complex Data Overview	ILO1
2	Tabular Data and Text-to-SQL	ILO1, ILO4
3	Complex Table Structure Understanding	ILO1, ILO2, ILO4
4	Table & Text: Retrieval, Question Answering, and Reasoning	ILO1, ILO3, ILO4
5	Temporal and Time-Series Data Management	ILO1, ILO2, ILO3, ILO4
6	Time-Series Analytics and Prediction	ILO1, ILO4, ILO5
7	Spatial and Spatiotemporal Data Management	ILO1, ILO2, ILO3, ILO4
8	Graph Data Models, Storage, and Query Languages	ILO1, ILO2, ILO4
9	Graph Query Processing, Indexing, and Optimization	ILO1, ILO3, ILO4

10	High-Dimensional Vector Data and Similarity Search	ILO1, ILO2, ILO3, ILO4
11	Retrieval-Augmented Generation as a Data System: Cost Modeling and Optimization	ILO1, ILO2, ILO3, ILO4
12	Multi-Source Heterogeneous Data Management	ILO1, ILO2, ILO3, ILO4
13	Data Agents for Complex Data Management	ILO1, ILO4

Assessment and Grading

This course will be assessed using criterion-referencing and grades will not be assigned using a curve. Detailed rubrics for each assignment are provided below, outlining the criteria used for evaluation.

Assessments:

Assessment Task	Contribution to Overall Course grade (%)	Due date
Written Assignments	40%	See Canvas; throughout the term
Project Report	20%	TBD*
Final Examination	40%	TBD*

* Assessment marks for individual assessed tasks will be released within two weeks of the due date.

Mapping of Course ILOs to Assessment Tasks

Assessed Task	Mapped ILOs	Explanation
Written Assignments	ILO1-ILO5	The assignments assess students' understanding of complex-data characteristics and their ability to select appropriate representations, storage schemes, indexes, retrieval methods, queries, and time-series analysis techniques for specified problems.
Project Report	ILO1-ILO5	The project report assesses the ability to formulate a complex-data management problem, design an end-to-end solution, justify system choices, implement or prototype key components, evaluate results, and communicate limitations, provenance, and reproducibility.
Final Examination	ILO1-ILO5	The final examination evaluates integrated conceptual understanding and the ability to analyze, compare, and apply data models, storage and indexing methods, query and retrieval techniques, system trade-offs, and time-series prediction methods.

Grading Rubrics

This course utilizes an absolute grading system as follows.

A: 85–100; B: 70–<85; C: 55–<70; D: 40–<55; F: <40.

Subgrades (e.g., A-, B+, etc.) will be assigned for every 5 points within these ranges.

Final Grade Descriptors:

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance	Demonstrates comprehensive and integrated mastery of complex-data models and systems; selects and applies methods accurately; produces insightful analyses and well-justified designs; communicates work with strong rigor, reproducibility, and originality.
B	Good Performance	Shows solid understanding of the main topics and applies appropriate storage, indexing, query, retrieval, and analytics methods effectively, with generally sound analysis and only minor errors or omissions.
C	Satisfactory Performance	Demonstrates adequate understanding of core concepts and can solve familiar problems, but analysis, method selection, implementation, or evaluation may be incomplete or contain noticeable weaknesses.
D	Marginal Pass	Shows threshold knowledge and can perform basic tasks, but has substantial gaps in conceptual understanding, technical accuracy, system reasoning, or the ability to integrate methods across data types.
F	Fail	Does not demonstrate the minimum understanding or problem-solving ability required for the course; solutions are frequently incorrect, incomplete, unsupported, or fail to meet core assessment requirements.

Course AI Policy

Generative AI tools may be used as learning aids and, where permitted by an assessment brief, for brainstorming, language improvement, exploratory coding, and debugging. Students remain fully responsible for the correctness, originality, security, and academic integrity of all submitted work.

- Any permitted AI assistance used in assessed work must be disclosed. The disclosure should identify the tool, describe how it was used, and explain how the outputs were checked or revised.
- AI-generated claims, code, citations, data, or experimental results must be independently verified. Fabricated references, results, or evidence constitute academic misconduct.
- Uploading confidential, personal, licensed, or otherwise restricted course or project data to external AI services is prohibited unless explicit authorization has been obtained.
- The use of generative AI, internet access, and unauthorized software is prohibited in the final examination and in any assessment explicitly designated as closed-book or no-AI. Assessment-specific instructions supersede this general policy.

Communication and Feedback

- A. Assessment marks for individual assessed tasks (for example, homework) will be communicated within two weeks of submission in most cases.
- B. Feedback on assignments will include, for instance, strength and weakness, suggestion for further thoughts, and areas for improvement.
- C. Students who have further questions about the feedback including marks are expected to contact the instructor or the TA within five working days after the feedback is received.
- D. Before the grade is finalized after the final exam, tentative grades will be posted. If a student has any concern regarding the tentative grade, the student is expected to contact the instructor within the time interval posted together with the tentative grades.

Resubmission Policy

If a student has an unforeseen, uncontrollable, and unavoidable reason, resubmitting work or reassessment opportunity will be considered. In such a case, the student is expected to contact the instructor immediately, but no later than within five working days, for further details.

Required Texts and Materials

The lecture notes and required readings posted on Canvas will be self-contained. Students will use selected research papers, system documentation, datasets, and software tools relevant to relational, temporal, spatial, graph, vector, and heterogeneous data management. A laptop capable of running Python/Jupyter and course-provided database or data-system software is recommended.

Academic Integrity

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST(GZ)'s Academic Honor Code and to maintain the highest standards of academic integrity. The University has zero tolerance of academic misconduct. Please refer to Regulations for Academic Integrity and Student Conduct for the University's definition of plagiarism and ways to avoid cheating and plagiarism.