City University of Hong Kong
Course Syllabus

offered by Department of Computer Science
with effect from Semester A 2017/18

Part I Course Overview

Course Title:	Data Engineering
Course Code:	CS5481
Course Duration:	One semester
Credit Units:	3 credits
Level:	P5
Medium of Instruction:	English
Medium of Assessment:	English
Prerequisites: *(Course Code and Title)*	Nil
Precursors: *(Course Code and Title)*	Nil
Equivalent Courses: *(Course Code and Title)*	Nil
Exclusive Courses: *(Course Code and Title)*	Nil

Part II Course Details

1. Abstract

This course has three objectives. First, it aims to enhance students' understanding of relational databases, especially on relational algebra and SQL. Second, it describes a number of advanced database topics including data indexing techniques, query optimization, transaction processing and concurrency control. These are important to the performance and correctness of a database system. Last, it discusses important issues about data storage, query processing and transaction management for parallel, distributed and real-time databases. All these are important in supporting sophisticated systems or applications.

2. Course Intended Learning Outcomes (CILOs)

(CILOs state what the student is expected to be able to do at the end of the course according to a given standard of performance.)

No.	CILOs	Weighting (if applicable)	Discovery-enriched curriculum related learning outcomes (please tick where appropriate)		
			A1	*A2*	*A3*
1.	Develop in-depth understanding of relational databases and proficiency in using SQL/relational algebra for data retrieval.	15%		✓	
2.	Apply data indexing techniques and query algorithms to design efficient evaluation plans.	35%	✓	✓	
3.	Describe issues specific to correct and efficient transaction execution.	15%		✓	
4.	Discuss data storage techniques, query and transaction processing algorithms for parallel, distributed and real-time databases.	35%	✓	✓	
		100%			

A1: Attitude
Develop an attitude of discovery/innovation/creativity, as demonstrated by students possessing a strong sense of curiosity, asking questions actively, challenging assumptions or engaging in inquiry together with teachers.

A2: Ability
Develop the ability/skill needed to discover/innovate/create, as demonstrated by students possessing critical thinking skills to assess ideas, acquiring research skills, synthesizing knowledge across disciplines or applying academic knowledge to self-life problems.

A3: *Accomplishments*
Demonstrate accomplishment of discovery/innovation/creativity through producing /constructing creative works/new artefacts, effective solutions to real-life problems or new processes.

3. Teaching and Learning Activities (TLAs)

(TLAs designed to facilitate students' achievement of the CILOs.)

Teaching pattern:
Suggested lecture/tutorial/laboratory mix: 2 hrs. lecture; 1 hr. tutorial.

TLA	Brief Description	CILO No.				Hours/week (if applicable)
		1	2	3	4	
Lectures	Explain the concepts in detail.	✓	✓	✓	✓	2 hrs/wk
Tutorials	Require students to apply knowledge learnt in the lectures to present and explain her/his solutions to given problems.	✓	✓	✓	✓	1 hr/wk
Individual assignments	Require students to independently work on two assignments. Each assignment contains questions designed to help students apply techniques / algorithms to solve practical problems.	✓	✓	✓	✓	

4. Assessment Tasks/Activities (ATs)

(ATs are designed to assess how well the students achieve the CILOs.)

Assessment Tasks/Activities	CILO No.				Weighting	Remarks
	1	2	3	4		
Continuous Assessment: 30%						
Assignments	✓	✓	✓	✓	18%	
Mid-term quiz	✓	✓			12%	
Examination^: 70% (duration: 2 hours)						
					100%	

^ For a student to pass the course, at least 30% of the maximum mark for the examination must be obtained.

5. Assessment Rubrics

(Grading of student achievements is based on student performance in assessment tasks/activities with the following rubrics.)

Assessment Task	Criterion	Excellent (A+, A, A-)	Good (B+, B, B-)	Fair (C+, C, C-)	Marginal (D)	Failure (F)
Assignment 1 & 2	1.1 Ability to apply data indexing techniques and query algorithms on practical scenario to design evaluation plans. 1.2 Ability to compare and minimize cost of evaluation plans. 1.3 Ability to apply algorithms to ensure database correctness in the presence of concurrent transaction executions. 1.4 Ability to apply algorithms to process query in parallel and distributed databases.	High	Significant	Moderate	Basic	Inadequate
Mid-term quiz	2.1 Ability to demonstrate a good understanding of data indexing. 2.2 Ability to apply different algorithms to process database operations. 2.3 Ability to evaluate the cost of a query.	High	Significant	Moderate	Basic	Inadequate
Examination	3.1-3 Same as 2.1-3 3.4 Ability to apply algorithms to exploit parallelism in processing query in parallel and distributed databases. 3.5 Ability to demonstrate a good understanding on basic concepts of transaction processing and concurrency control. 3.6 Ability to process transactions in distributed and real-time databases.	High	Significant	Moderate	Basic	Inadequate

Part III Other Information (more details can be provided separately in the teaching plan)

1. Keyword Syllabus

(An indication of the key topics of the course.)

Relational algebra. SQL. Data indexing. Query algorithm and optimization. Transactions and concurrency control. Distributed database. Parallel database. Real-time database.

Syllabus

1. Introduction
   Relational concepts. SQL. Storage.

2. Relational algebra
   Selection. Projection. Join. Aggregation.

3. Data indexing
   Dense/sparse primary/non-primary index. B+-tree. Hashing.

4. Query algorithm.
   Selection. External sorting. Join: Block nested loop, Indexed nested loop, Sort merge join, Hash join.

5. Query optimization.
   Query cost analysis. Materialization and pipelining. Histogram and size estimation. Cost-based optimization and heuristics.

6. Transactions and concurrency control
   ACID properties. Serializability. Concurrency control. Two-phase locking. Deadlock prevention, detection and recovery.

7. Distributed database
   Horizontal/vertical partitioning. Distributed concurrency control. Semi-join.

8. Parallel database
   Data partitioning: round robin, hash partition, range partition. Inter-query parallelism. Intra-query parallelism. Intra-operation parallelism. Inter-operation parallelism.

9. Real-time database
   Transaction correctness. Deadlines. Priority assignment. Real-time concurrency control. Temporal consistency.

2. Reading List

2.1 Compulsory Readings

(Compulsory readings can include books, book chapters, or journal/magazine articles. There are also collections of e-books, e-journals available from the CityU Library.)

1.	*Silberschatz A., Korth H.F. and Sudarshan S. Database System Concepts. 6th Ed. McGraw Hill (2011) (latest edition)*

2.2 Additional Readings

(Additional references for students to learn to expand their knowledge about the subject.)

1.	*Hector Garcia-Molina, Jeffrey D. Ullman and Jennifer Widom. Database Systems The Complete Book. 2nd Ed. Pearson New International Edition (2013) (latest edition)*
2.	*Elmasri R. and Navathe S.B. Database Systems: Models, Languages, Design and Application Programming. 6th Ed. Pearson (2011) (latest edition)*