

**City University of Hong Kong  
Information on a Course  
offered by Department of Computer Science  
with effect from Semester A in 2015 / 2016**

---

---

**Part I**

**Course Title:** Data Engineering

**Course Code:** CS5481

**Course Duration:** One Semester

**Credit Units:** 3

**Level:** P5

**Medium of Instruction:** 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 Aims**

This course has four objectives. First, it aims to enhance students' understanding of relational databases, especially on relational algebra and SQL. Second, it introduces the object-relational data model. Third, it discusses data indexing techniques, query algorithms and optimization, transactions and concurrency control. Last, it introduces important issues about data storage, query processing and transaction management in parallel, distributed and real-time databases.

## Course Intended Learning Outcomes (CILOs)

*Upon successful completion of this course, students should be able to:*

No.	CILOs	Weighting (if applicable)
1.	Develop in-depth understanding of relational databases and proficiency in using SQL for data retrieval;	15%
2.	Explain the main concepts of the object-relational data model and contrast different data models;	10%
3.	Apply data indexing and query algorithms on query optimization;	30%
4.	Describe the issues specific to efficient transaction execution;	20%
5.	Describe data storage techniques, query and transaction processing strategies for parallel, distributed and real-time databases.	25%

## Teaching and Learning Activities (TLAs)

*(Indicative of likely activities and tasks designed to facilitate students' achievement of the CILOs. Final details will be provided to students in their first week of attendance in this course)*

Teaching pattern:

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

CILO No.	TLAs	Hours/week (if applicable)
CILO 1	<b>Lectures:</b> The instructor will explain the concepts in detail. Examples and demonstrations with real systems will be used to supplement the explanation.	2 weeks
CILO 2		1.5 week
CILO 3		4 weeks
CILO 4	<b>Tutorials:</b> Some problem cases will be given to students. Students will be required to present and explain her/his solution to the problem case. The instructor will guide the discussion.	2.5 weeks
CILO 5		3 weeks

## Assessment Tasks/Activities

*(Indicative of likely activities and tasks designed to assess how well the students achieve the CILOs. Final details will be provided to students in their first week of attendance in this course)*

CILO No.	Type of Assessment Tasks/Activities	Weighting (if applicable)	Remarks
CILO 1 – 2	<b>Assignment 1:</b> The questions in this assignment will require students' to use SQL and relational algebra to retrieve data from a database; test their understanding on main concepts of relational and object-relational data models.	8%	
CILO 1 – 3	<b>Mid-term quiz:</b> In addition to questions that serve similar purpose as those appeared in Assignment 1, other questions in the mid-term quiz require students to judge an evaluation plan and estimate its cost.	10%	
CILO 3 – 4	<b>Assignment 2:</b> Students will be given practical scenario and are required to apply data indexing and various query algorithms to design a	12%	

	number of evaluation plans for cost-based optimization. Other questions in this assignment will request students to generate possible schedules for processing concurrent transactions.		
CILO 1 – 5	<b>Examination:</b> In addition to questions that serve similar purpose as those appeared in the coursework, other questions in the examination will test students' knowledge in analysing and evaluating various data storage techniques, query and transaction processing strategies for parallel, distributed and real-time databases.	70%	

**Grading of Student Achievement:** Refer to Grading of Courses in the Academic Regulations for Taught Postgraduate Degrees.

*Examination duration:* 2 hours

*Percentage of coursework, examination, etc.:* 30% CW; 70% Exam

*Grading pattern:* Standard (A+AA-...F)

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

### Part III

#### Keyword Syllabus

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

#### Syllabus

1. Introduction  
Relational concepts: Row, column, primary key, foreign key. SQL. Disk organization.
2. Relational algebra  
Selection. Projection. Join. Aggregation.
3. Object-relational data model  
Advanced data structures. Inheritance. Data definition language. Extended SQL.
4. Data indexing  
Dense/sparse primary/non-primary index. B+-tree.
5. Query algorithm.

- External sort. Block nested loop. Indexed nested loop. Sort merge join. Hash join.
6. Query optimization.  
Query cost analysis. Pipelining. Cost-based optimization. Histogram. Sampling.
  7. Transactions and concurrency control  
Transaction concepts: ACID properties and serializability. Two-phase locking. Deadlocks.
  8. Distributed database  
Horizontal/vertical partitioning. Distributed concurrency control. Semi-join.
  9. Parallel database  
Data partitioning: round robin, hash partition, range partition. Inter-query parallelism. Intra-query parallelism.
  10. Real-time database  
Transaction deadlines. Priority inversion. Real-time concurrency control protocols. Temporal consistency. Data deadlines.

### **Recommended Reading**

#### **Text(s)**

##### Essential Text

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

##### Supplementary Reading

*Hector Garcia-Molina, Jeffrey D. Ullman and Jennifer Widom. Database Systems The Complete Book. 2<sup>nd</sup> Ed. Pearson New International Edition (2013) (latest edition)*

*Elmasri R. and Navathe S.B. Database Systems: Models, Languages, Design and Application Programming. 6<sup>th</sup> Ed. Pearson (2011) (latest edition)*