國立雲林科技大學資訊管理系

115 學年度第 1 學期博士班資格考 (Ph.D. Qualify Exam)

考試科目 (Subject)：資料庫系統 (Database Systems)

時數與方式 (Hour and type)：~~四小時 (4 hours)~~ 3.5小時 (3.5 hours)

Close book, 不可使用計算器 (Calculators not allowed)

Part 1: 14 points (Optional Questions: Answer 2 out of 3, 7 points each; if you answer more than 2 questions, only the first 2 answered questions will be graded.)

1. Discuss the two main types of constraints on specializations and generalizations in the EER model.
2. What is multivalued dependency? When does it arise?
3. What is meant by the completeness and soundness of Armstrong's inference rules?

Part 2: 36 points (Optional Questions: Answer 6 out of 9, 6 points each; if you answer more than 6 questions, only the first 6 answered questions will be graded.)

4. Please write out the full names of the following abbreviations: SQL, DML, SDL, VDL, ODMG, OQL. (1 point each)
5. What is the difference between controlled and uncontrolled redundancy? Illustrate with examples.
6. What is the difference between logical data independence and physical data independence? Which one is harder to achieve? Why?
7. Under what conditions can an attribute of a binary relationship type be migrated to become an attribute of one of the participating entity types?
8. Discuss user-defined and attribute-defined specializations, and identify the differences between the two.
9. Consider the following relations for a database that keeps track of student enrollment in courses and the books adopted for each course:

STUDENT (Ssn, Name, Major, Bdate)
COURSE (Course#, Cname, Dept)
ENROLL (Ssn, Course#, Quarter, Grade)
BOOK_ADOPTION (Course#, Quarter, Book_isbn)
TEXT (Book_isbn, Book_title, Publisher, Author)

(a) Identify the primary key for each relation, explicitly stating any underlying assumptions.

(b) Identify all foreign keys and illustrate the corresponding referential integrity constraints on the relational schema diagram.

10. Discuss how each of the following constructs is used in SQL, and discuss the various options for each construct. Specify what each construct is useful for. (2 points each)

(a) Nested queries

(b) Aggregate functions and grouping

(c) Views and their updatability

11. Consider the two tables T1 and T2 shown below. Show the results of the following operations: (2 points each)

(a) $T1 \bowtie_{(T1.P = T2.A)} T2$

(b) $T1 \cup T2$

(c) $T1 \bowtie_{(T1.P = T2.A \text{ AND } T1.R = T2.C)} T2$

TABLE T1			TABLE T2		
P	Q	R	A	B	C
10	a	5	10	b	6
15	b	8	25	c	3
25	a	6	10	b	5

12. Define first, second, and third normal forms when only primary keys are considered. Give the tests used in each case, and the corresponding remedy or normalization performed to achieve the normal form.

Part 3: 20 points (Optional Questions: Answer 2 out of 4 questions, 10 points each; if you answer more than 2 questions, only the first 2 answered questions will be graded.)

13. The B+-tree has become the dominant indexing structure in commercial database management systems (DBMSs). Explain why the B+-tree is generally preferred over hash indexing and the B-tree.

14. A company stores the following two relations:
    EMPLOYEE: 500,000 tuples
    DEPARTMENT: 500 tuples
    Both relations reside on disk. Compare Nested-Loop Join, Sort-Merge Join, and Hash Join in terms of their disk I/O behavior. Based on the sizes of the two relations, explain which join algorithm you would prefer and why.

15. SQL is the standard query language for relational databases. Write an SQL script to create a database with three tables (Customers, Orders, Products). Each table should have at least five columns, and implement a query to calculate the total order amount for each customer.

16. A query joins six relations with multiple selection predicates. Assume that the query optimizer has already estimated the cardinality of each intermediate relation. Answer the following questions:
    (a) Why does join ordering have a significant impact on query performance?
    (b) Why is exhaustive search generally impractical for queries involving a large number of joins?
    (c) Compare dynamic programming, greedy heuristics, and randomized optimization algorithms for determining join order. Discuss their trade-offs in terms of optimization time and the quality of the resulting execution plan.

Part 4: 30 points (Optional Questions: Answer 3 out of 6 questions, 10 points each; if you answer more than 3 questions, only the first 3 answered questions will be graded.)

17. Database security is a crucial aspect of data management. Describe the major threats to database security and their countermeasures and explain how the Role- Based Access Control (RBAC) model is implemented.

18. Explain why Write-Ahead Logging (WAL) is essential for database recovery. Discuss how REDO, UNDO, and checkpointing work together to recover a database after a system failure.

19. Explain the CAP theorem and why a distributed database cannot guarantee both consistency and availability when a network partition occurs. Compare CP and AP systems and give one practical example of each.

20. A multinational enterprise plans to deploy a distributed database across Asia, Europe, and North America. Design an appropriate data distribution strategy for this enterprise. Explain the trade-offs involved in your design and justify how the proposed strategy can provide both acceptable performance and reliability across geographically distributed sites.

21. Relational databases have dominated enterprise computing for decades. However, NoSQL databases have become increasingly important for large-scale and distributed applications. Compare relational databases and NoSQL databases from the following perspectives:
    Data model
    Scalability
    Transaction support
    Query flexibility
    Consistency
    Typical applications

22. A hospital intends to build an enterprise data warehouse integrating data from the following information systems:
    Hospital Information System (HIS)
    Laboratory Information System (LIS)
    Radiology Information System (RIS)
    Picture Archiving and Communication System (PACS)
    Electronic Medical Record (EMR)

Design an appropriate data warehouse architecture for this hospital. Your discussion should address:

    ETL processes
    Star schema
    Snowflake schema
    OLTP versus OLAP
    Materialized views