

94 Ph.D. Qualify Exam., Subject: Database Systems

Four hours (Part One and Two), Closed book.

Part One

1. 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) List the number of courses taken by all students named 'John Smith' in Winter 2000 (i.e., Quarter='W00'). (5%)
- (b) Produce a list of textbooks (include Course#, Book_ISBN, Book_Title) for courses offered by the 'MIS' department that have used more than two books. (5%)
- (c) List any department that has all its adopted books published by 'PH Publishing'. (5%)
2. What is a data warehouse? How does it differ from a database? (10%)
3. Consider the universal relation $R = \{A, B, C, D, E, F, G, H, I, J\}$ and the set of functional dependencies $F = \{\{A, B\} \rightarrow \{C\}, \{A\} \rightarrow \{D, E\}, \{B\} \rightarrow \{F\}, \{F\} \rightarrow \{G, H\}, \{D\} \rightarrow \{I, J\}\}$. What is the key for R? Decompose R into 2NF and then 3NF relations. (10%)
4. In the Enhanced Entity-Relationship data model, how does a category differ from a regular shared subclass? What is a category used for? Illustrate your answer with examples. (10%)
5. What is the difference between procedural and nonprocedural DMLs (data manipulation languages)? (5%)

第二部分：

1. How does clustering differ from classification? 5%
2. Compare the primary site method with the primary copy method for distributed concurrency control. How does the use of backup sites affect each? 5%
3. What are covert channels? Give an example of a covert channel. 5%
4. What is meant by the type orthogonality of OQL? 5%
5. Describe the write-ahead logging protocol. 5%
6. What is a schedule (history)? Define the concepts of recoverable, cascadeless, and strict schedules, and compare them in terms of their recoverability. 5%
7. Discuss the cost components for a cost function that is used to estimate query execution cost. Which cost components are used most often as the basis for cost functions? 5%
8. Between the properties of dependency preservation and losslessness, which one must definitely be satisfied? Why? 5%
9. What is the impedance mismatch problem? Which of the three programming approaches minimizes this problem? 5%
10. Define foreign key. What is this concept used for? 5%