

## [종합시험] 데이터베이스

1. 파일 시스템을 이용하여 대용량 데이터를 관리하는 것은 많은 문제점들을 발생시킨다. 이들 문제점들에 대하여 설명하시오.
2. Relational data model의 세 가지 구성요소(components)인 relational data의 basic structure(용어 및 속성), 기본 연산자(basic operators), 그리고 referential integrity constraints에 대하여 가능한 상세히 설명하시오.
3. functional dependency의 정의와 용도에 대하여 설명하고 candidate key의 정의를 functional dependency를 이용하여 기술하여라.
4. lossless-join decomposition과 dependency-preserving decomposition을 예를 이용하여 설명하시오.
5. 다음 relation schema R과 functional dependency set F를 토대로 아래 질문에 대하여 각각 답하시오. (각 질문에 대한 답 작성 시 과정 설명도 포함되어야 함)

$R = (A, B, C, D, E)$

$F : A \rightarrow BC$

$CD \rightarrow E$

$B \rightarrow D$

$E \rightarrow A$

- (1) R의 모든 candidate key들을 구하고 두 개 이상의 애트리뷰트로 구성된 candidate key가 존재하는 경우 한 가지를 선택하여 candidate key임을 증명하여라.
  - (2) F의 canonical cover  $F_c$ 를 구하시오.
  - (3) R을 (A, B, E)와 (C, D, E)로 분해(decomposition)하는 경우 이러한 분해가 lossless-join decomposition인지를 판단하고 판단 결과를 증명하여라.
  - (4) (3)에서 제시된 분해가 dependency-preserving decomposition인지를 판단하고 판단 결과를 증명하여라.
6. 다음 테이블을 토대로 view의 예를 SQL로 정의하고 이러한 view를 이용하는 목적에 대하여 설명하시오.

account_number	branch_name	balance
A-101	Downtown	500
A-215	Mianus	700
A-102	Perryridge	400
A-305	Round Hill	350
A-201	Brighton	900
A-222	Redwood	700
A-217	Brighton	750

7. 다음 관계형 스키마(relational schema)를 토대로 아래의 질의어를 관계 대수(relational algebra)와 SQL로 각각 표현하시오.

- employee (person-name, street, city)
- works (person-name, company-name, salary)
- company (company-name, type)
- branch (company-name, branch-name, city)

- (1) Find the names and cities of residence of all employees who work for First Bank Corporation and earn more than \$100,000 per annum. (First Bank Corporation의 직원들 중 연봉이 \$100,000 이상인 직원들의 이름과 거주 도시를 찾아라.)
- (2) Find the name and average salary of the company whose the average salary is greater than \$50,000. (각 회사별 직원들의 연봉 평균이 \$50,000 이상인 회사와 그 평균 연봉을 찾아라.)
- (3) Find all companies located in every city in which the branches of Small Bank Corporation are located. (Small Bank Corporation의 지점이 위치한 모든 도시에 지점을 갖고 있는 회사의 이름을 찾아라.)
- (4) Find the number of branches of each company.(각 회사의 지점의 수를 찾아라.)
- (5) Find those companies whose employees earn a higher salary than the largest salary at First Bank Corporation. (First Bank Corporation의 최고 임금보다 많은 임금을 받는 직원들이 소속되어 있는 회사의 이름을 찾아라.)

8. 고려대학교에서 이루어지는 업무 규칙에 대한 아래 설명을 토대로 다음 질문에 답하시오.

고려대학교 정보대학 학사지원부에서는 개설과목과 학생, 그리고 학생들의 수강과목에 대한 다음과 같은 정보를 관리하고자 한다.

- 개설과목 정보에는 과목번호(primary key), 과목명, 담당교수명에 대한 정보들이 포함되며, 학생 정보에는 학번(primary key), 이름, 학과명 정보가 포함되어야 한다.
- 아울러 학생들이 수강하는 과목들에 대한 정보도 관리하여 한다.
- 학생들은 반드시 한 과목 이상 7 과목 이하의 과목을 수강하여야 하며 한 개설과목의 최대 수강 인원은 40명이다.
- 개설과목들 중에는 한 학생도 신청하지 않은 과목이 존재하는 경우도 있다.

- (1) 고려대학교 정보대학에서 이루어지는 업무를 관리하기 위한 데이터베이스를 설계하고자 한다. 이를 위한 E-R diagram를 작성하여라.(단, 아래 항목들을 포함하여야 함)
  - Entity Sets, Relationship Sets, Attributes, Cardinality Constraints(Cardinality Limit 포함), Participation Constraints
- (2) (1)의 E-R diagram으로 표현한 E-R modeling 결과를 relational schemas(tables)로 변환하여 나타내고 이를 SQL로 정의하여라.

9 B+ 인덱스와 Bitmap 인덱스 기법 각각에 대하여 각 기법을 효율적으로 적용할 수 있는 경우를 기술하고 Bitmap 인덱스를 이용한 질의처리 과정을 예(예제 테이블과 Bitmap index 예, SQL 예)를 들어 설명하시오.

10. Query processing에서 사용되는 두 가지 질의어 수행(query evaluation) 방법에 대하여 설명하시오.

11. 휴리스틱 최적화 절차(heuristic optimization steps)에서 사용하는 휴리스틱의 예를 두 가지만 기술하시오.[5]

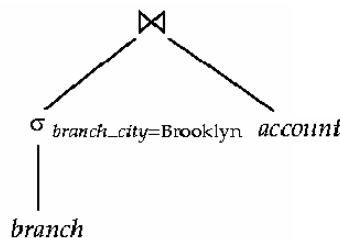
12. 릴레이션 branch, account 이 다음과 같이 주어졌을 때 다음 질문에 대하여 답하시오.

branch ( brach_name, branch_city, assets )	- branch 릴레이션은 10,000 tuples로 구성되어 있다. - branch 릴레이션에 있는 tuple들은 한 block에 25개가 저장된다. - $V(\text{branch\_city}, \text{branch}) = 100$
account ( account_number, branch_name, balance )	- account 릴레이션은 30,000 tuples로 구성되어 있다. - account 릴레이션에 있는 tuple들은 한 block에 30개가 저장된다.

(1) 비용-기반(cost\_based) 질의처리 과정을 상세히 설명하시오.

(2) 다음과 같은 logical query execution plan과 가정이 주어졌을 때 이 질의어를 처리하는데 소요되는 총비용을 계산하시오.(계산 과정도 기술하기 바랍니다.)

<ul style="list-style-type: none"> <li>- select 연산은 linear-select, join 연산은 block nested-loop join 연산으로 수행한다.</li> <li>- 비용은 메모리로 입력(INPUT)된 block 수만을 이용하여 계산한다.</li> <li>- 메모리의 크기는 각 릴레이션의 block 하나씩만을 수용할 수 있는 크기로 가정한다.</li> <li>- select 결과 또는 조인 결과를 디스크에 기록하는 비용은 고려하지 않는다.</li> <li>- 조인 연산에 왼쪽 입력 테이블이 outer table로 가정한다.</li> </ul>
--



13. 트랜잭션의 네 가지 속성에 대하여 설명하시오.

14. conflict equivalent와 conflict serializability의 정의를 기술하고 이를 토대로 conflict serializable schedule의 예를 제시한 후 이를 증명하시오.

15. Recoverable schedule에 대하여 기술하고 recoverable schedule이 필요한 이유에 대하여 예를 이용하여 설명하시오.

16. 다음 질문에 답하시오.

(1) recovery가 불가능한 schedule의 예와 recoverable schedule의 정의를 기술하시오.

(2) (1)에서 제시한 schedule의 예를 토대로 strict 2-Phase Locking protocol이 recoverable schedule을 보장함을 증명하시오.

17. 다음 두 트랜잭션내의 연산 수행 순서는 시간 순서에 따라 수행된다. 이를 토대로 다음 질문에 답하시오.(그림은 (2)번 문제에서만 사용)

(1) strict two-phase locking protocol과 rigorous two-phase locking protocol을 비교 설명하시오. [

(2) 그림에 제시된 두 트랜잭션이 strict two-phase locking protocol에 의해 수행되는 과정을 설명하시오.

time	$T_1$	$T_2$
	read(A) write(A)	
		read(A) write(A)
	read(B) write(B)	
		read(B) write(B)

18. 다음 질문들에 대하여 답하시오.

(1) Write-ahead Log 규칙에 대하여 설명하시오.

(2) Log buffer의 로그 레코드들이 stable Log 저장소(예, 디스크)로 옮겨지는 상황들에 대하여 설명하시오.

(3) checkpoint 이용 목적에 대하여 설명하시오.

(4) 오른쪽에 제시되어 있는 로그 정보를 토대로 이루어지는 recovery 과정에 대하여 설명하시오.

```

< T0 start >
< T0, A, 1000, 950 >
< T1 start >
< T1, B, 2000, 2500 >
< checkpoint { T0, T1 } >
< T1, C, 700, 600 >
< T1, C, 700 >
< T1, B, 2000 >
< T1 abort >
< T0, B, 2000, 2050 >
< T0 commit >
< T2 start >
< T2, A, 950, 900 >
< T2, B, 2050, 2100 >
< T3 start >
< T3, C, 700, 650 >
< T3, D, 1000, 1050 >
< T3 commit >
★ End of Log at crash
  
```