

[DB 5강-정규화, 관계데이터연산, 관계대수]

[출제빈도 '중']

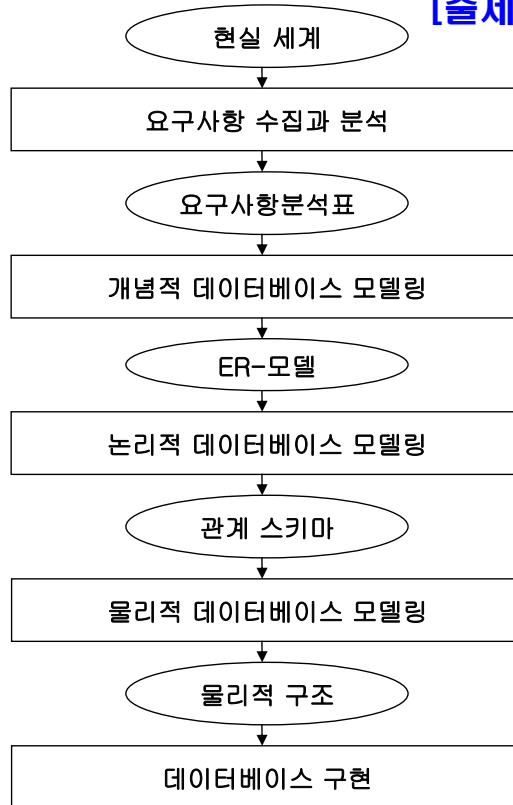
1. 관계형 데이터베이스 모델링

조직도/업무 분장표, 업무별 흐름도 ->
인터뷰 내용, 입출력 장표
기존 시스템 분석, 새로운 요구사항

실세계의 정보 구조의 모형을 변환하여 ->
일반화 시키는 단계

목표 **DBMS 이론** 적용, **정규화** ->

속성의 데이터타입과 사이즈 정의 ->
데이터 사용량 분석
사용자들의 업무 프로세스 분석
역정규화



[DB 5강-정규화, 관계데이터연산, 관계대수]

1. 정규화 (normalization) ★★

: 정규화를 하는 이유는 데이터의 **중복**을 방지하고 보다 효율적으로 데이터를 저장하기 위함.
(릴레이션 분리 -> 삽입, 삭제, 갱신 **이상**의 발생 가능성을 줄이는 것)

- 단점 : 연산 시간이 증가됨

2. 제1정규형 (1NF)

: 반복 되는 속성을 제거한 뒤 **모든 속성이 원자 도메인** 만으로 되어 있는 정규형

* 제1정규형에 위배되는 테이블 (중복 발생->공간 낭비, 이상 발생->무결성 위배 가능)

[회원]

회원번호	이름	주소	이메일	주민번호	수강과목	과목내용
K001	이상희	서울시	kkk	111	정보처리 컴활	A B
K002	김시현	광주시	ccc	222	사무자동화	C

[회원]

회원번호	이름	주소	이메일	주민번호
K001	이상희	서울시	kkk	111
K002	김시현	광주시	ccc	222

[수강]

수강과목	회원번호	과목내용
정보처리	K001	A
컴활	K001	B
사무자동화	K002	C

[DB 5강-정규화, 관계데이터연산, 관계대수]

3. 이상(anomaly)

: 릴레이션에서 일부 속성들의 종속으로 인해 데이터의 중복이 발생하여 테이블 조작 시 불일치가 발생하는 것

- (1) 갱신 이상 : 반복된 데이터 중에 일부만 수정하면 데이터의 불일치가 발생
-> 이상희라는 사람의 주소를 변경할 경우 모든 속성(칼럼)의 주소를 변경해야 한다.
만약 하나만 변경할 경우 데이터의 불일치가 발생한다.
- (2) 삽입 이상 : 불필요한 정보를 함께 저장하지 않고는 어떤 정보를 저장하는 것이 불가능
-> MOS라는 과목을 추가할 경우, 불필요한 회원정보까지 추가해야 한다.
- (3) 삭제 이상 : 유용한 정보를 함께 삭제하지 않고는 어떤 정보를 삭제하는 것이 불가능
-> 이상희라는 사람의 데이터를 삭제하고자 할 경우, 정보처리라는 과목까지 삭제되어 버린다.

[회원]

회원번호	이름	주소	이메일	주민번호	수강과목	과목내용
K001	이상희	서울시	kkk	111	정보처리	A
K001	이상희	서울시	kkk	111	컴활	B
K002	김시현	광주시	ccc	222	사무자동화	C
K002	김시현	광주시	ccc	222	컴활	B
K002	김시현	광주시	ccc	222	워드	D

[DB 5강-정규화, 관계데이터연산, 관계대수]

4. 제2정규형 (2NF)

: 제1정규형이고, **부분함수적 종속을 제거**하여 **완전(충분한) 함수적 종속**을 만족하는 정규형.

1) 함수적 종속

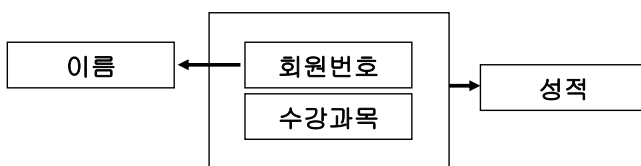
- 어떤 릴레이션 R에서 X와 Y를 각각 R의 애트리뷰트 집합의 부분 집합이라고 할 경우, 애트리뷰트 X의 값 각각에 대하여 시간에 관계없이 항상 애트리뷰트 Y의 값이 오직 하나만 연관되어 있을때 Y는 X에 함수 종속이라 하고, X를 결정자, Y를 종속자 한다.

- 표기법 : $X \rightarrow Y$
- 예) 회원번호 -> 이름

[회원]

회원번호	이름	주소	이메일
K001	이상희	서울시	kkk
K002	김시현	광주시	ccc

2) 부분 / 완전 함수적 종속



[성적]

회원번호	수강과목	성적	이름
K001	정보필기	80	이상희
K002	정보필기	90	김시현
K002	정보실기	90	김시현
K001	정보실기	100	이상희

완전 함수적 종속 : 회원번호, 수강과목 -> 성적
부분 함수적 종속 : 회원번호 -> 이름

이름 속성을 분리하면 성적 테이블은 완전함수적 종속 상태가 되므로 2정규형을 만족한다.

[DB 5강-정규화, 관계데이터연산, 관계대수]

5. 제3정규형 (3NF)

: 제2정규형이고, **이행적 함수적 종속 관계 제거**하여 **비이행적 함수적** 종속 관계를 만족하는 정규형

1) 이행적 함수적 종속

신청번호 → 회원번호, 회원번호 → 회원취미

: 신청번호 → 회원취미 는 이행적 함수적 종속

[신청]

신청번호	제품번호	회원번호	회원취미
1	A1111	kkk	독서
2	B2222	qqq	영화
3	C3333	www	음악
4	D4444	eee	축구

회원번호, 회원취미 속성을 분리하면 신청 테이블은 비이행적 함수적 종속 상태가 되므로 3정규형을 만족한다.

6. BCNF (Boyce/Codd Normal Form)

: 제3정규형이고, **결정자가 후보키가 아닌 함수 종속 제거** 모든 **결정자**가 **후보키**이어야 한다는 것

[수강]

학번	과목	교수
001	프로그래밍	A
001	자료구조	B
002	프로그래밍	A
002	자료구조	C
003	자료구조	C
003	프로그래밍	D

후보키 : (학번, 과목), (학번, 교수)

함수적 종속 : (학번, 과목) → 교수, 교수 → 과목

결정자 : (학번, 과목), 교수

교수 속성이 결정자이나 후보키가 아니므로 교수 속성을 분리하면 수강 테이블은 모든 결정자가 후보키이므로 BCNF를 만족한다.

[DB 5강-정규화, 관계데이터연산, 관계대수]

7. 제4정규형 (4NF)

: **다치 종속** 제거

- 키워드 : 다중값 종속(=다치 종속) 제거, A -->> B

- 다치 종속 :

먼저 함수적 종속의 개념을 아셔야 합니다.

릴레이션에서 아이디 속성이 있고, 주민번호 속성이 있다고 할때, 아이디 속성을 알면 주민번호 속성을 알 수 있습니다.(아이디 속성은 주민번호 속성을 결정하는 결정자입니다.)

이 때 아이디 → 주민번호 로 표현하고 함수적 종속이라고 합니다. 여기서 주의할 것은 1:1 로 대응되어야 합니다.

(하나의 아이디 값이 2개 이상의 주민번호 속성을 결정하면 안됩니다.) 그럼 다치 종속은 설명이 쉽습니다. 1:

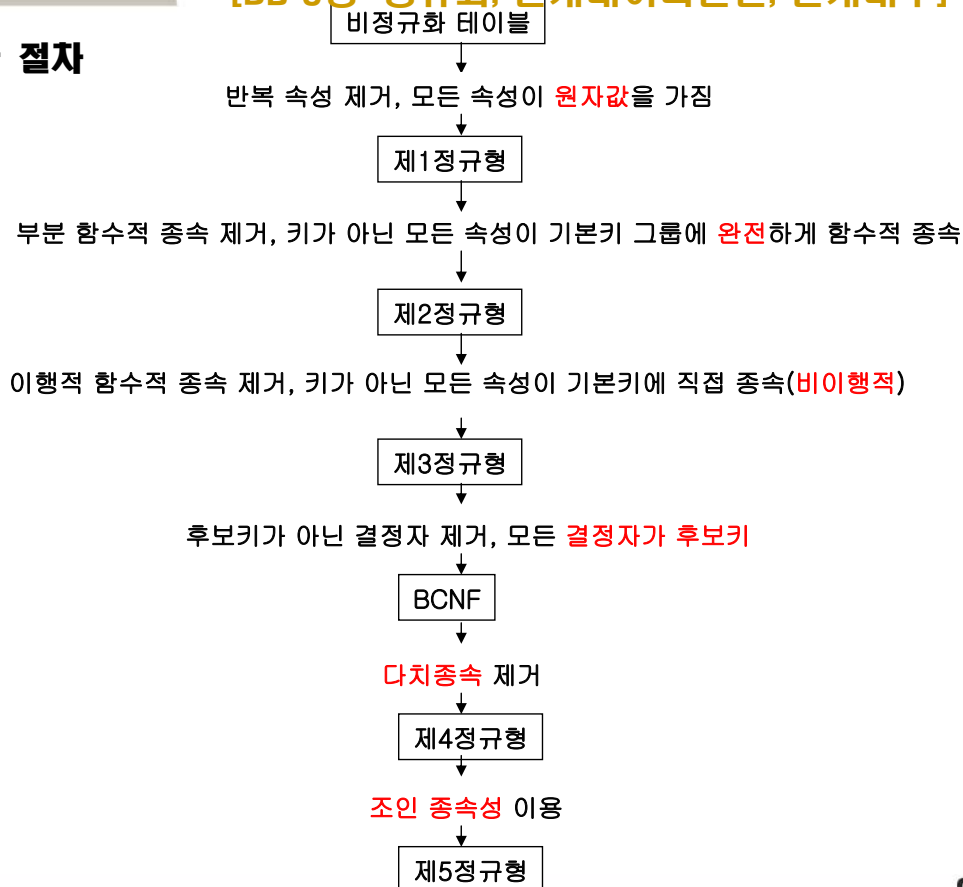
다 대응을 생각하시면 됩니다. 릴레이션에서 아이디 속성이 있고, 수강과목 속성이 있다면 하나의 아이디는 여러개 과목을 수강할 수 있으므로 아이디와 수강과목 속성은 함수적 종속에서 다치 종속이라고 합니다. 표현은 아이디 →> 수강과목 입니다. 이런 다치 종속은 이상이 발생할 수 있으므로 무손실 분해(정규화)되어야 합니다. 이것을 4정규형이라고 합니다.

8. 제5정규형 (5NF)

: **조인 종속성** 이용

[DB 5강-정규화, 관계데이터연산, 관계대수]

9. 정규화 절차



[DB 5강-정규화, 관계데이터연산, 관계대수]

10. 정규화 특징

- 예를 들어 현재 테이블이 3정규형 상태라면 1,2 정규형은 자동으로 만족한다.
- 정규형들은 차수가 높아질수록(제1정규형→제5정규형)만족시켜야 할 제약조건이 증가된다.
- 정규화는 논리적 처리 및 품질에 큰 영향을 미친다.
- 정규화의 목적은 논리적 데이터베이스 구조상에 있어 삽입, 수정, 그리고 삭제 결과 생기는 이상현상(anomaly)을 제거하는데 있다.
- 레코드들의 관련 속성들 간의 종속성을 최소화하기 위한 구성 기법이다
- 정규화가 잘못되면 데이터의 불필요한 중복을 야기하여 릴레이션 조작시 문제를 일으킨다.
- 정규화되지 못한 릴레이션의 조작시 발생하는 이상(anomaly) 현상의 근본적인 원인은 여러 가지 종류의 사실들이 하나의 릴레이션에 표현되기 때문이다.

[DB 5강-정규화, 관계데이터연산, 관계대수]

1. 관계 데이터 연산 ★★☆☆☆

관계 대수	관계 해석
1. 절차적 언어(절차 중심) - 원하는 정보를 '어떻게' 유도하는가를 연산자와 연산규칙 이용하여 기술 2. 분류 : 순수 관계 연산자, 일반 집합 연산자 3. SQL의 이론적인 기초	1. 비절차적 언어(결과 중심) - 원하는 정보가 '무엇'이라는 것만 정의 2. 분류 : 튜플 관계 해석과 도메인 관계 해석

- 기본적으로 관계 해석과 관계 대수는 관계 데이터베이스를 처리하는 기능과 능력면에서 동등하다.
- 관계 해석으로 표현한 식은 관계대수로 표현할 수 있다.

[DB 5강-정규화, 관계데이터연산, 관계대수]

1. 관계 대수 종류 ★★☆☆☆

- 순수 관계 연산자 : SELECT, PROJECT, JOIN, DIVISION
- 일반 집합 연산자 : 합집합, 교집합, 차집합, 카티션프로덕트(곱하기 연산)

2. SELECT (σ)

- 릴레이션에서 주어진 조건을 만족하는 튜플들을 검색하는 것으로 기호는 그리스 문자의 **σ** 그마(σ)를 이용한다. (**행, 수평적 연산**)

- 형식 : σ 조건 (R)

수강생

이름	과목	주소	수강료
김길현	정보	남구	100
이상인	정보	서구	120
남기욱	정보	서구	100
최영희	컴활	중구	80
김상현	워드	북구	50



이름	과목	주소	수강료
김길현	정보	남구	100

σ 이름 = '김길현' (**수강생**)

[DB 5강-정규화, 관계데이터연산, 관계대수]

3. PROJECT (π)

- 릴레이션에서 주어진 조건을 만족하는 **속성**들을 검색하는 것으로, 기호는 그리스 문자의 **파이**(π)를 이용한다. (**열**, 수직적 연산)

- 형식 : π 속성 (R)

수강생

이름	과목	주소	수강료
김길현	정보	남구	100
이상인	정보	서구	120
남기욱	정보	서구	100
최영희	컴활	중구	80
김상현	워드	북구	50



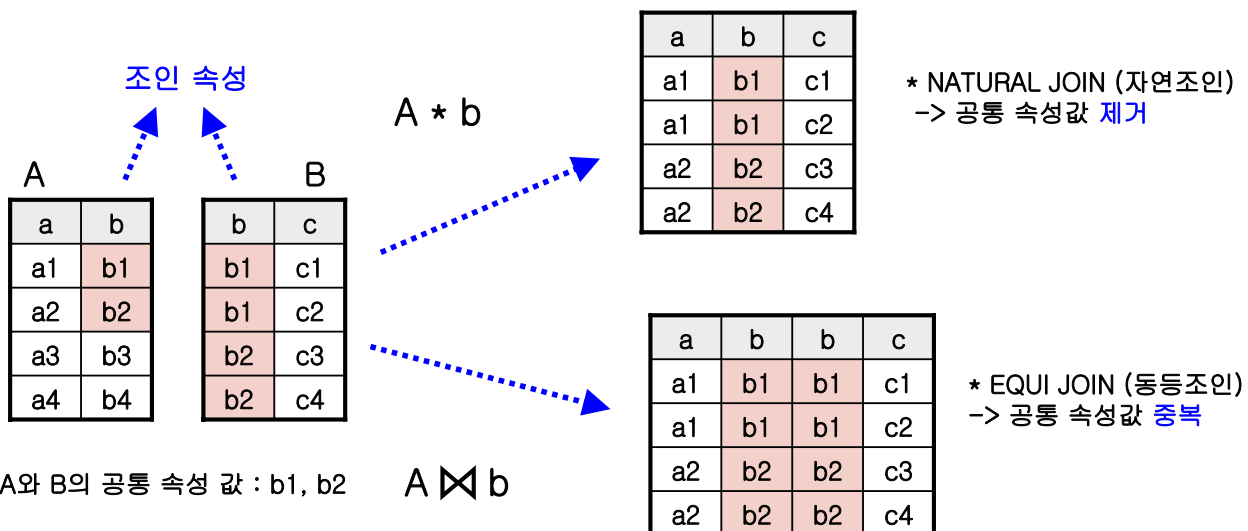
이름
김길현
이상인
남기욱
최영희
김상현

π 이름 (수강생)

[DB 5강-정규화, 관계데이터연산, 관계대수]

4. JOIN

- 두 개의 릴레이션 A와 B에서 공통된 속성을 연결하는 것이다.

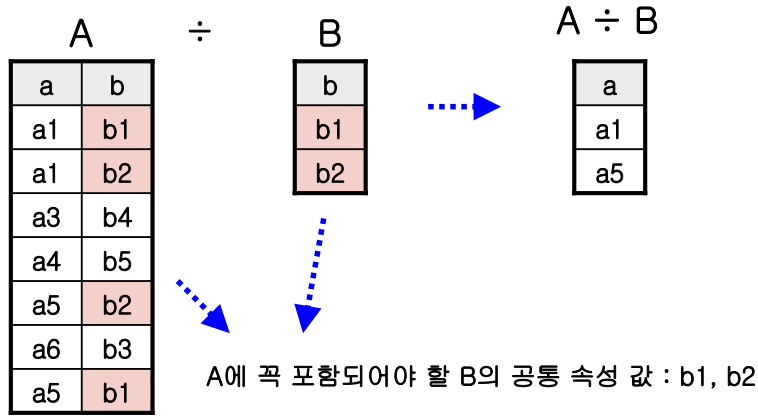


A와 B의 공통 속성 값 : b1, b2

[DB 5강-정규화, 관계데이터연산, 관계대수]

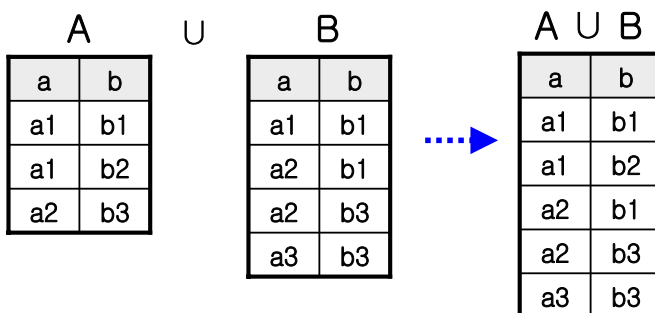
5. DIVISION (÷)

- 나누어지는 릴레이션인 A는 릴레이션 B의 모든 내용을 포함한 것이 결과 릴레이션이 된다

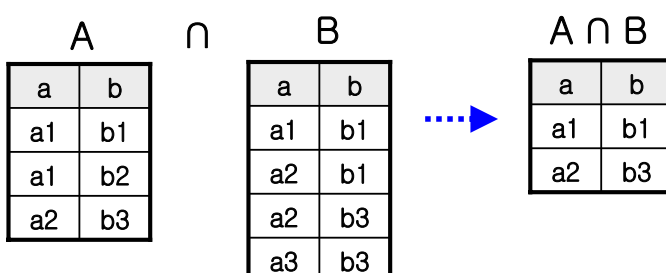


[DB 5강-정규화, 관계데이터연산, 관계대수]

6. 합집합 (U) : 릴레이션 A 또는 B에 속하는 튜플들로 구성된 릴레이션이다. (UNION)




7. 교집합 (∩) : 릴레이션 A 와 B에 공통적으로 속하는 튜플들로 구성된 릴레이션이다. (INTERSECTION)



관계 대수 연산자 중 합집합, 교집합, 차집합 연산은 이항 연산으로서 연산에 참가하는 두 개의 릴레이션은 차수와 도메인이 같아야 연산을 수행할 수 있다.

[DB 5강-정규화, 관계데이터연산, 관계대수]

8. 차집합 (-) : 릴레이션 A에만 있고 B에는 없는 튜플들로 구성된 릴레이션이다. (DIFFERENCE)

A	-	B		A - B	B - A																												
<table><tr><th>a</th><th>b</th></tr><tr><td>a1</td><td>b1</td></tr><tr><td>a1</td><td>b2</td></tr><tr><td>a2</td><td>b3</td></tr></table>	a	b	a1	b1	a1	b2	a2	b3		<table><tr><th>a</th><th>b</th></tr><tr><td>a1</td><td>b1</td></tr><tr><td>a2</td><td>b1</td></tr><tr><td>a2</td><td>b3</td></tr><tr><td>a3</td><td>b3</td></tr></table>	a	b	a1	b1	a2	b1	a2	b3	a3	b3		<table><tr><th>a</th><th>b</th></tr><tr><td>a1</td><td>b2</td></tr></table>	a	b	a1	b2	<table><tr><th>a</th><th>b</th></tr><tr><td>a2</td><td>b1</td></tr><tr><td>a3</td><td>b3</td></tr></table>	a	b	a2	b1	a3	b3
a	b																																
a1	b1																																
a1	b2																																
a2	b3																																
a	b																																
a1	b1																																
a2	b1																																
a2	b3																																
a3	b3																																
a	b																																
a1	b2																																
a	b																																
a2	b1																																
a3	b3																																

9. 카티션 프로덕트(cartesian product) (X)

- A에 속한 각 튜플 a에 대하여 B에 속한 튜플 b를 모두 접속시킨 튜플들(a b)로 구성된 릴레이션이다.

A

a
a1
a2

×

B

b
b1
b2

→

A × B

A.a	B.b
a1	b1
a1	b2
a2	b1
a2	b2

릴레이션 A 튜플 수 : 2

릴레이션 B 튜플 수 : 2

결과 튜플 수 : 2 × 2 = 4

[DB 5강-정규화, 관계데이터연산, 관계대수]

1. 조건을 만족하는 릴레이션의 수평적 부분집합으로 구성하며, 연산자의 기호는 그리스 문자 시그마(σ)를 사용하는 관계대수 연산자는?

- 가. Select 연산자 나. Project 연산자
다. Join 연산자 라. Division 연산자

2. 관계 데이터베이스에 적용할 순수 관계 연산자로 거리가 먼 것은?

- 가. 링크(Link) 나. 셀렉트(Select)
다. 디비전(Division) 라. 프로젝트(Project)

3. 정규화의 의미로 틀린 것은?

- 가. 함수적 종속성 등의 종속성 이론을 이용하여 잘못 설계된 관계형 스키마를 더 작은 속성의 세트로 쪼개어 바람직한 스키마로 만들어 가는 과정이다.
나. 좋은 데이터베이스 스키마를 생성해 내고 불필요한 데이터의 중복을 방지하여 정보 검색을 용이하게 해준다.
다. 정규형에는 제1정규형, 제2정규형, 제3정규형, BCNF형, 제4정규형, 제5정규형 등이 있다.
라. 어떠한 Relation 구조가 바람직한 것인지, 바람직하지 못한 Relation을 어떻게 합쳐야 하는지에 관한 구체적인 판단 기준을 제공한다.

4. 다음 관계 대수 문장의 의미는?

Π 이름(σ 학과='전산'(교수))

- 가. 전산학과 교수들의 이름을 검색하시오
나. 전산학과 교수들의 이름 테이블을 삭제하시오
다. 전산학과 교수들의 이름을 삭제하시오
라. 전산학과 교수들의 이름을 삽입하시오

5. 릴레이션 R에는 10개의 튜플이 있고, 다른 릴레이션 S에는 5개의 튜플이 있을 때, 두 개의 릴레이션 R과 S의 교차곱(Cartesian Product) 연산을 수행한 후의 튜플의 수는?

- 가. 15개 나. 50개 다. 10개 라. 2개

6. 테이블에서 특정 속성에 해당하는 열을 선택하는데 사용되며 결과로는 릴레이션의 수직적 부분 집합에 해당하는 관계 대수 연산자는?

- 가. project 연산자
나. join 연산자
다. division 연산자
라. select 연산자

[정답] 1.가 2.가 3.라 4.가 5.나 6.가

[DB 5강-정규화, 관계데이터연산, 관계대수]

7. 관계형 데이터베이스의 릴레이션을 조작할 때 발생하는 이상(Anomaly) 현상에 관한 설명으로 적절하지 않은 것은?

- 가. 데이터의 종속으로 인해 발생하는 이상 현상에는 삭제 이상, 삽입 이상, 갱신 이상이 있다.
- 나. 릴레이션의 한 튜플을 삭제함으로써 연쇄 삭제로 인해 정보의 손실을 발생시키는 현상이 삭제 이상이다.
- 다. 데이터를 삽입할 때 불필요한 데이터가 함께 삽입되는 현상을 삽입 이상이라 한다.
- 라. 튜플 중에서 일부 속성을 갱신함으로써 정보의 모순성이 발생하는 현상이 갱신 이상이다.

8. 키가 아닌 모든 속성이 기본 키(Primary key)에 충분한 함수적 종속을 만족하는 정규형은?

- 가. 1NF 나. 2NF 다. 3NF 라. 4NF

9. 제3정규형에서 보이스코드 정규형(BCNF)으로 정규화하기 위한 작업은?

- 가. 원자값이 아닌 도메인을 분해
- 나. 부분 함수 종속 제거
- 다. 이행 함수 종속 제거
- 라. 결정자가 후보키가 아닌 함수 종속 제거

10. 어떤 릴레이션 R에 존재하는 모든 조인 종속성이 릴레이션 R의 후보키를 통해서만 성립된다. 이 릴레이션 R은 어떤 정규형의 릴레이션인가?

- 가. 제3정규형
- 나. 보이스-코드 정규형
- 다. 제4정규형
- 라. 제5정규형

11. 관계 데이터베이스에 있어서 관계 대수 연산이 아닌 것은?

- 가. 디비전(division) 나. 프로젝션(projection)
- 다. 조인(join) 라. 포크(fork)

12. 관계 해석(relational calculus)에 대한 설명으로 옳지 않은 것은?

- 가. 관계 해석으로 질의어를 표현한다.
- 나. 원하는 릴레이션을 정의하는 방법을 제공하며, 비절차적인 언어이다.
- 다. 튜플 관계 해석과 도메인 관계 해석이 있다.
- 라. 릴레이션 조작을 위한 연산의 집합이다.

[정답] 7.가 8.나 9.라 10.라 11.라 12.라

[DB 5강-정규화, 관계데이터연산, 관계대수]

13. 어떤 릴레이션에 속한 모든 도메인이 원자값(atomic value)만으로 되어 있는 릴레이션을 무엇이라고 하는가?

- 가. 제 1 정규형(1NF) 나. 제 2 정규형(2NF)
- 다. BCNF 라. 제 4 정규형(4NF)

14. 정규화의 목적으로 거리가 먼 것은?

- 가. 삽입, 삭제, 갱신 이상의 발생을 방지한다.
- 나. 데이터의 중복성을 최소화 한다.
- 다. 효율적으로 데이터를 조작할 수 있다.
- 라. 릴레이션을 분해하여 연산시간을 감소시킨다.

15. 다음 중 SQL의 합집합 연산이 제대로 수행되는 경우는?

- 가. 두 테이블의 속성 개수가 같고, 대응되는 각 속성들의 도메인이 같으나 속성명들이 다른 경우
- 나. 두 테이블의 속성 개수가 같고, 대응되는 각 속성들의 도메인이 같으나 속성들의 도메인이 다른 경우
- 다. 두 테이블의 속성 개수가 다르나, 대응되는 각 속성들의 도메인이 같으며 속성명들이 같은 경우
- 라. 두 테이블의 속성 개수가 같으나 대응되는 각 속성들의 도메인이 다르고 속성명들이 다른 경우

16. 관계 해석(Relational Calculus)을 옳게 설명한 것은?

- 가. 연산들의 절차(sequence)를 사용하여 데이터를 가져온다.
- 나. 계산 수식을 사용하여 어떤 데이터를 가져올지 명시한다.
- 다. 기본적인 연산자로 UNION, INTERSECTION, DIFFERENCE를 사용한다.
- 라. 전체 관계를 조작하는데 사용되는 연산들의 집합이다.

17. 관계 대수(Relational Algebra)의 연산 중에서 두 릴레이션(Relation)의 교차 곱을 수행하기 때문에 두 릴레이션의 공통 튜플 수와 관계가 없는 것은?

- 가. UNION 나. INTERSECTION
- 다. DIFFERENCE 라. CARTESIAN PRODUCT

[정답] 13.가 14.라 15.가 16.나 17.라

[DB 5강-정규화, 관계데이터연산, 관계대수]

18. 다음의 조건을 모두 만족하는 정규형은?

- 모든 도메인은 원자 값이고, 기본 키가 아닌 모든 속성들이 기본 키에 대해 완전 함수 종속적이며, 이행적 함수 종속 관계는 제거되었다.

- 가. 제 1 정규형 나. 제 2 정규형
다. 제 3 정규형 라. 제 1 정규형과 제 2 정규형

[정답] 18.다

[DB 5강-정규화, 관계데이터연산, 관계대수]

[산-08년9월][산-08년5월]

19. 관계대수의 프로젝트 연산의 연산자 기호는?

- 가. π 나. \cap
다. \div 라. \cup

[기-08년5월]

20. 관계데이터베이스의 정규화에 대한 설명으로 옳지 않은 것은?

- 가. 정규화를 거치지 않으면 여러 가지 상이한 종류의 정보를 하나의 릴레이션으로 표현하여 그 릴레이션을 조작할 때 이상(Anomaly) 현상이 발생할 수 있다.
나. 정규화의 목적은 각 릴레이션에 분산된 종속성을 하나의 릴레이션에 통합하는 것이다.
다. 이상(Anomaly) 현상은 데이터들 간에 존재하는 함수종속이 하나의 원인이 될 수 있다.
라. 정규화가 잘못되면 데이터의 불필요한 중복이 야기되어 릴레이션을 조작할 때 문제가 발생할 수 있다.

[기-08년9월]

21. 정규화 과정 중 1NF에서 2NF가 되기 위한 조건은?

- 가. 1NF를 만족하고 모든 도메인이 원자 값이어야 한다.
나. 1NF를 만족하고 키가 아닌 모든 애트리뷰트들이 기본키에 이행적으로 함수 종속되지 않아야 한다.
다. 1NF를 만족하고 다치 종속이 제거되어야 한다.
라. 1NF를 만족하고 키가 아닌 모든 속성이 기본키에 완전 함수적 종속되어야 한다.

[정답] 19.가 20.나 21.라