

DB세미나 5장

논리적 설계, 정규화

논리적 설계: ER스키마를 관계 모델의 릴레이션으로 사상

단계	사상할 대상	구분
1단계	개체 타입	강한 개체 타입
2단계		약한 개체 타입
3단계	관계 타입	이진 1:1 관계 타입
4단계		이진 1:N 관계 타입
5단계		이진 NM 관계 타입
6단계		N진 관계 타입
7단계	속성	다중값 속성

단계 1: 정규 엔티티 타입과 단일 값 애트리뷰트

- ER 스키마의 각 정규 엔티티 타입 E에 대해 하나의 릴레이션 R을 생성한다. E에 있던 단순 애트리뷰트들을 릴레이션 R에 모두 포함.
- E에서 복합 애트리뷰트를 구성하는 단순 애트리뷰트들만 릴레이션 R에 포함시킨다.

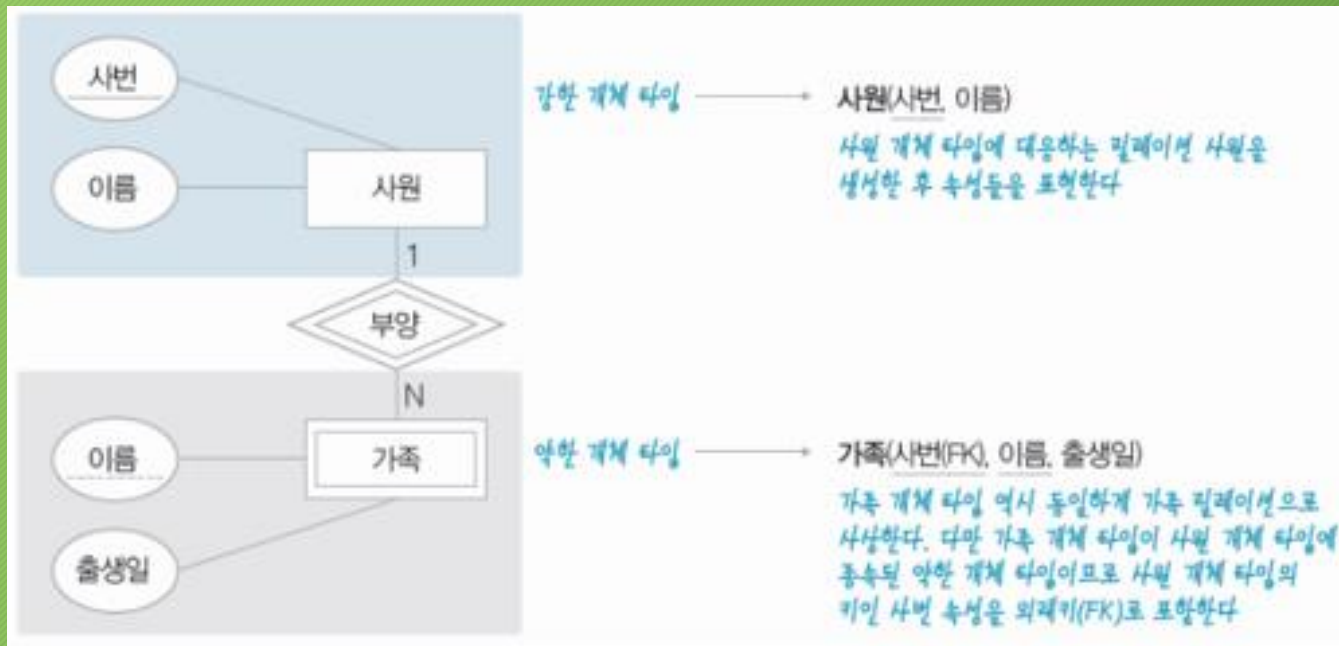


[그림 5.44] 강 엔티티 타입을 릴레이션으로 사상

단계 2: 약한 엔티티 타입과 단일 값 애트리뷰트

• 약한 개체 타입

약한 개체 타입에서 생성된 릴레이션은 자신의 키와 함께 강한 개체 타입의 키를 외래키로 사상해 기본키를 구성한다.



단계 3: 관계 타입의 사상(Mapping)

이진 관계 타입을 사상하는 방법은 네 가지가 있다.

- **방법1** 오른쪽 개체 타입 E2를 기준으로 관계 R을 표현한다.

E1(KA1, A2)

E2(KA2, A4, KA1)

- **방법2** 왼쪽 개체 타입 E1을 기준으로 관계 R을 표현한다.

E1(KA1, A2, KA2)

E2(KA2, A4)

- **방법3** 단일 릴레이션 ER로 모두 통합하여 관계 R을 표현한다.

ER(KA1, A2, KA2, A4)

- **방법4** 개체 타입 E1, E2와 관계 타입 R을 모두 독립된 릴레이션으로 표현한다.

E1(KA1, A2)

R(KA1, KA2)

E2(KA2, A4)

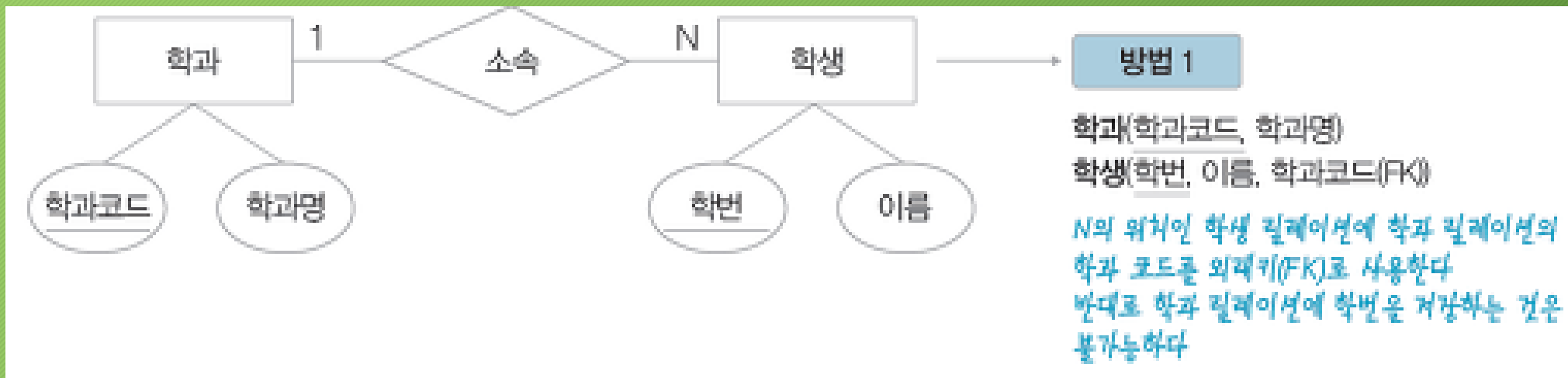
이진 1:1 관계 타입

- 이진 1:1 관계 타입의 경우 모든 유형으로 사상될 수 있다.



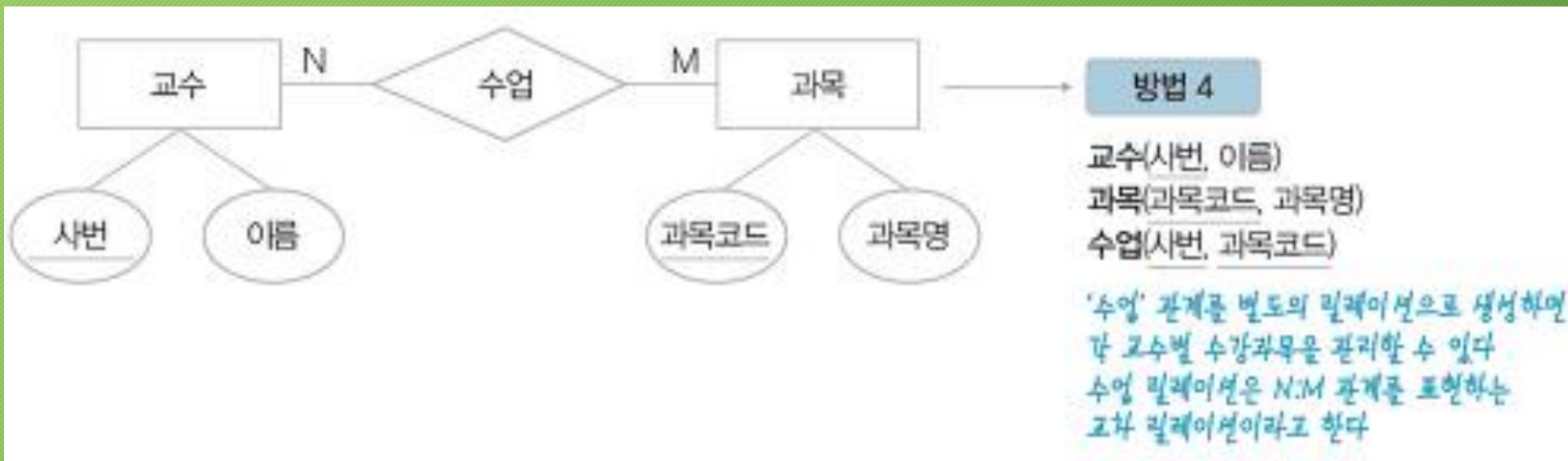
이진 1:N 관계 타입

- 이진 1:N 관계 타입의 경우 N의 위치에 따라 방법1 또는 방법 2의 유형으로 사상된다.



이진 N:M 관계 타입

- 이진 N:M 관계 타입은 방법 4의 유형으로 사상된다.



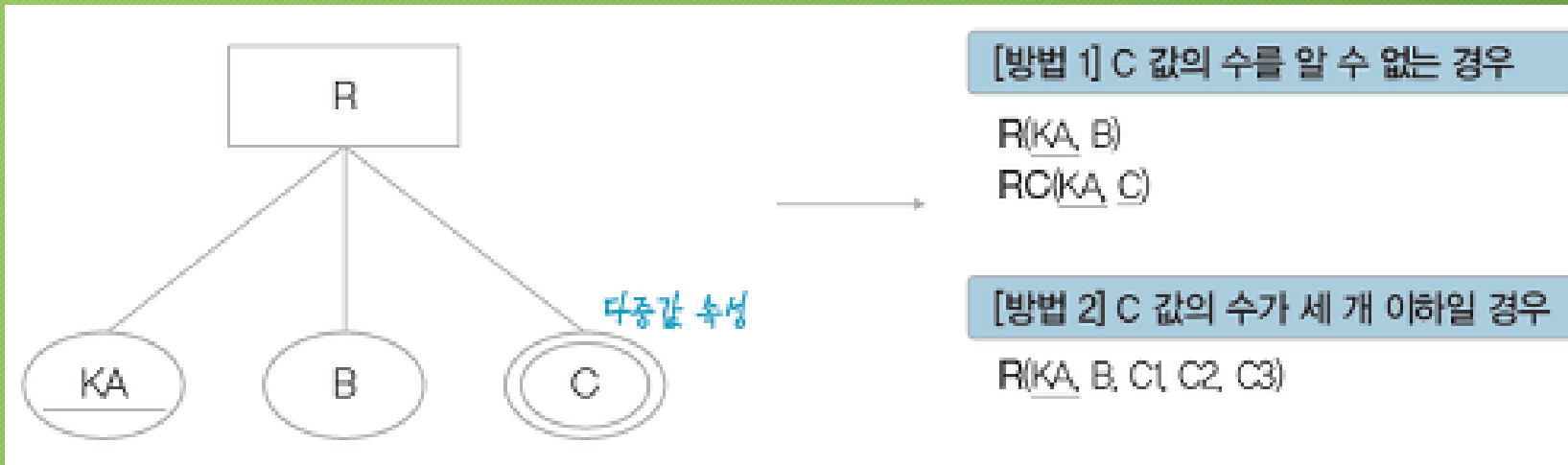
N진 관계 타입

- ER 모델의 차수가 3 이상인 다진 관계 타입의 경우 방법 4의 유형으로 사상된다.
- 그림에서 수강릴레이션의 호실은 기본키가 되면 안된다. 왜그럴까?



단계 4: 다치 애트리뷰트

- 속성의 사상 시 다중값 속성의 경우 직접 사상할 수 없다.
- 다중값 속성의 개수를 알 수 없는 경우 새로운 릴레이션을 생성한다.
- 다중값 속성의 개수가 적고 제한 가능한 경우, 속성을 릴레이션에 같이 포함한다.



릴레이션 정규화

- 정규화는 보다 좋은 데이터 설계로 이끄는 상향식 방법론이다.
- 수정이상, 삽입이상, 삭제이상을 총칭하여 갱신이상이라고 한다.

- 정보의 중복

정보의 중복은 동일 정보를 두 곳 이상에 저장하는 것을 말한다.

- 수정이상

반복된 데이터 중에 일부만 수정하면 데이터의 불일치가 발생한다.

- 삽입이상

불필요한 정보를 함께 저장하지 않고는 어떤 정보를 저장하는 것이 불가능하다.

- 삭제이상

유용한 정보를 함께 삭제하지 않고는 어떤 정보를 삭제하는 것이 불가능하다.

※ 이 갱신이상들을 최소화 하기위해 있는 것이 릴레이션 정규화이다!

관계 데이터베이스 설계 비공식 지침

- 지침 1: 이해하기 쉽고 명확한 스키마를 만들라.

갱신이상 발생하지 않도록, 여러 엔티티 타입이나 관계 타입에 속한 애트리뷰트들을 하나에 포함 X

- 지침 2: 널값을 피하라

값이 널일 가능성이 높은 애트리뷰트를 릴레이션에 포함시키지 않는다.

- 지침 3: 가짜 튜플이 생기지 않도록 하라.

기본키 또는 외래 키로 사용되는 애트리뷰트들 간에 동등 조건으로 조인할 수 있도록 설계.

- 지침 4: 스키마를 정제하라.

더 완전하게 무결성 제약조건들을 고려함으로써 설계를 정제해야 한다.

함수적 종속성 (1)

- 결정자

어떤 애트리뷰트의 값은 다른 애트리뷰트의 값을 고유하게 결정할 수 있다.

결정자는 주어진 릴레이션에서 다른 애트리뷰트(또는 애트리뷰트들의 집합)를 고유하게 결정하는 하나 이상의 애트리뷰트를 의미

예로 들어서 사원 릴레이션에서 사원 번호는 사원 이름, 주소, 전화번호 등의 결정자이다. 또한 부서번호는 부서이름의 결정자이다.

결정자를 'A->B' 라고 표기하고, 이를 "A가 B를 결정한다(A는 B의 결정자이다)" 라고 말한다.

유니크한 사원번호는 이름, 주소, 전화번호를 결정해준다.

(사원 번호만 있으면 언제든지 정해진 이름 주소 등을 뽑아낼수 있죠? 그래서 결정자라고 합니다.)

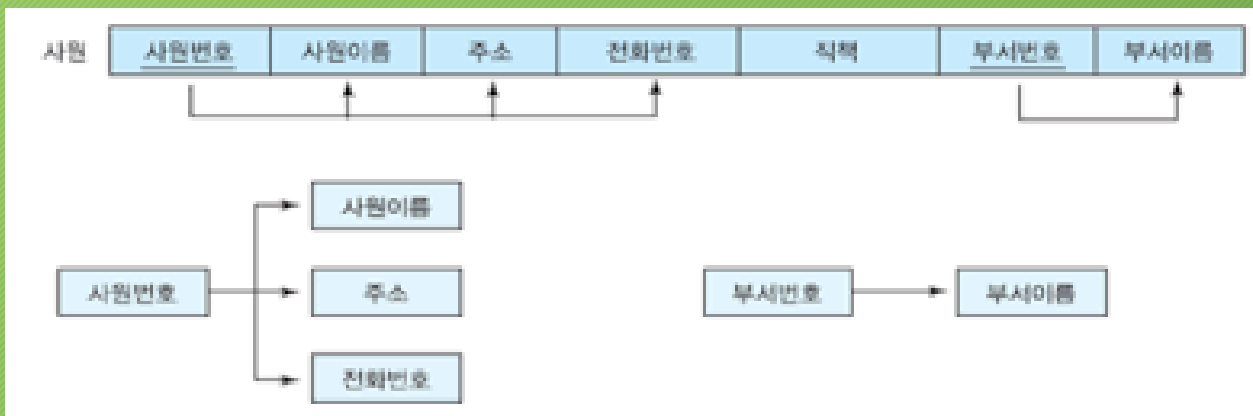
함수적 종속성 (2)

- 함수적 종속성

앞에서 말했던 직원번호는 이름 주소의 결정자 라고 했죠? 반대로 이름 주소가 직원번호에 **함수적으로 종속**한다고 합니다.

- 완전 함수적 종속성

임의의 릴레이션 R에서 속성 또는 속성들의 집합 X에 대해 Y가 함수적으로 종속되면서, X의 부분 집합에 대하여서는 함수적으로 종속하지 않은 경우, Y는 X에 대하여 완전 함수적 종속성을 갖는다고 한다.



★애트리뷰트가 릴레이션의 모든 결정자에 대해 종속하고 있으면 **완전 함수적 종속성**

☆결정자들의 부분집합에 대해 종속하고 있으면 **부분 함수적 종속성**

함수적 종속성 (3)

- 이행적 함수적 종속성

$A \rightarrow B$ AND $B \rightarrow C$ 면 $A \rightarrow C$ 이다.

삼단논법 그대로입니다.

B가 A에 종속이고 C가 B에 종속이면 C는 A에 종속인것입니다.

릴레이션 분해

- 정보의 손실

원래의 릴레이션을 분해한 후에 생성된 릴레이션들을 조인한 결과에 들어 있는 정보가 원래의 릴레이션에 들어있는 정보보다 적거나 많은 것을 모두 포함한다.

- 무손실 분해

무손실 분해는 이러한 정보의 손실이 전혀 없는 분해를 말한다.

정규형을 들어가기 앞서.

- 정규화 하기전의 모든 릴레이션은 삽입, 삭제, 수정 이상이 있습니다.
(1정규형 빼고요.)
- 각각의 1,2,3,BCNF 정규화를 하기전에 삽입, 삭제, 수정이상을 찾아 봅시다.

제1정규형

- 한 릴레이션 R이 제1정규형을 만족할 필요충분조건은 릴레이션 R의 모든 애트리뷰트가 원자값만을 갖는다는 것이다.
- 릴레이션의 모든 애트리뷰트에 반복 그룹이 나타나지 않으면 제1정규형을 만족한다. 반복그룹은 한 개의 기본 키값에 대해 두개 이상의 값을 가질 수 있는 애트리뷰트를 의미

<고객테이블 - 1>

고객번호	고객명	취미
1	아무개	(영화, 여행)
2	홍길동	등산

반복그룹(repeating group)
→ 다치 애트리뷰트

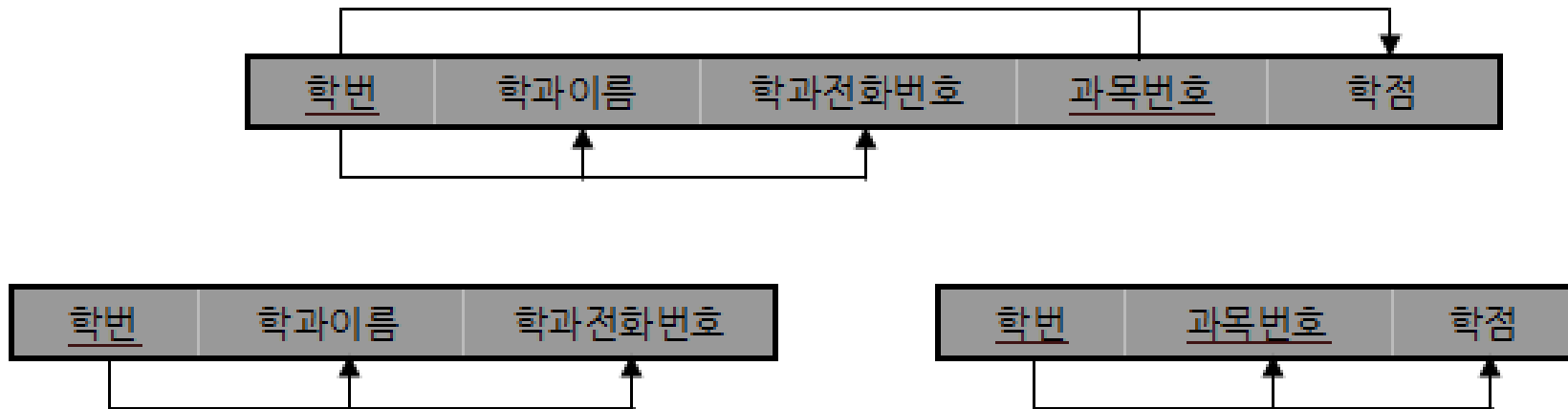
반복그룹 제거

<고객테이블 - 2>

고객번호	고객명	취미
1	아무개	영화
1	아무개	여행
2	홍길동	등산

제2정규형

- 부분 함수적 종속성을 제거
- 제2정규형을 만족할 조건은 릴레이션 R이 제1정규형을 만족하고, 어떤 후보키에도 속하지 않는 모든 애트리뷰트들이 R의 기본 키에 완전하게 함수적으로 종속하는 것.



[부분 함수적 종속성이 존재하지 않도록 분해된 두 릴레이션]

제3정규형

- 키가 아닌 모든 애트리뷰트가 릴레이션 R의 기본 키에 이행적으로 종속하지 않는 것이다.

• 이행 함수 종속성을 가지는 릴레이션

사번	성명	기숙사	임대료
9201	박성식	다산관	200,000원
9202	오기자	퇴계관	300,000원
9203	이기도	퇴계관	300,000원
9204	홍민화	다산관	200,000원
9205	홍숙희	다산관	200,000원

사번 → 성명
사번 → 기숙사
사번 → 임대료
기숙사 → 임대료

• 제3정규형으로 변환된 릴레이션

사번	성명	기숙사
9201	박성식	다산관
9202	오기자	퇴계관
9203	이기도	퇴계관
9204	홍민화	다산관
9205	홍숙희	다산관

기숙사	임대료
다산관	200,000원
퇴계관	300,000원

BCNF 정규형

- 모든 결정자가 후보 키 이어야함. (간단하죠?)

보이스/코드 정규화 과정

수강(결정자가 키가 아닌 속성인 릴레이션)

학 번	과목명	교수명
100	전자계산기구조	이문세
100	데이터베이스	김연아
100	운영체제	강승범
101	데이터베이스	김연아
101	운영체제	전혜영

지도교수 (모든 결정자가 후보키인 BCNF 릴레이션)

학 번	교수명
100	이문세
100	김연아
100	강승범
101	김연아
101	전혜영

후보키가 아닌 결정자를 제거함

교수(모든 결정자가 후보키인 BCNF 릴레이션)

교수명	과목명
이문세	전자계산기구조
김연아	데이터베이스
강승범	운영체제
전혜영	운영체제

모든 정규화 요약

