

1. 정규화(Normalization)

[정규화(Normalization)이란?]

정규화(Normalization)의 기본 목표는 테이블 간에 중복된 데이터를 허용하지 않는다는 것이다. 중복된 데이터를 허용하지 않음으로써 무결성(Integrity)을 유지할 수 있으며, DB의 저장 용량 역시 줄일 수 있다.

이러한 테이블을 분해하는 정규화 단계가 정의되어 있는데, 여기서 테이블을 어떻게 분해되는지에 따라 정규화 단계가 달라지는데, 각각의 정규화 단계에 대해 자세히 알아보도록 하자.

[제1 정규화]

제1 정규화란 테이블의 컬럼이 원자값(Atomic Value, 하나의 값)을 갖도록 테이블을 분해하는 것이다. 예를 들어 아래와 같은 고객 취미 테이블이 존재한다고 하자.

고객취미들(이름, 취미들)	
이름	취미들
김연아	인터넷
추신수	영화, 음악
박세리	음악, 쇼핑
장미란	음악
박지성	게임

위의 테이블에서 추신수와 박세리는 여러 개의 취미를 가지고 있기 때문에 제1 정규형을 만족하지 못하고 있다. 그렇기 때문에 이를 제1 정규화하여 분해할 수 있다. 제1 정규화를 진행한 테이블은 아래와 같다.

고객취미(이름, 취미)

이름	취미
김연아	인터넷
추신수	영화
추신수	음악
박세리	음악
박세리	쇼핑
장미란	음악
박지성	게임

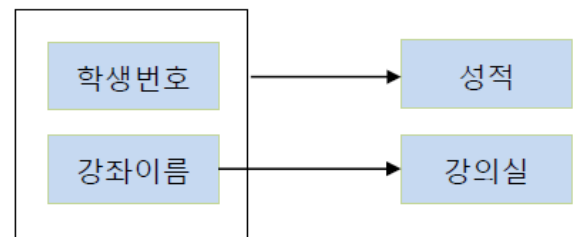
[제2 정규화]

제2 정규화란 제1 정규화를 진행한 테이블에 대해 완전 함수 종속을 만족하도록 테이블을 분해하는 것이다. 여기서 완전 함수 종속이라는 것은 **기본키의 부분집합이 결정자가 되어선 안된다는 것**을 의미한다.

예를 들어 아래와 같은 수강 강좌 테이블을 살펴보자.

수강강좌

학생번호	강좌이름	강의실	성적
501	데이터베이스	공학관 110	3.5
401	데이터베이스	공학관 110	4.0
402	스포츠경영학	체육관 103	3.5
502	자료구조	공학관 111	4.0
501	자료구조	공학관 111	3.5



이 테이블에서 기본키는 (학생번호, 강좌이름)으로 복합키이다. 그리고 (학생번호, 강좌이름)인 기본키는 성적을 결정하고 있다. (학생번호, 강좌이름) --> (성적)

그런데 여기서 강의실이라는 컬럼은 기본키의 부분집합인 강좌이름에 의해 결정될 수 있다. (강좌이름) --> (강의실)

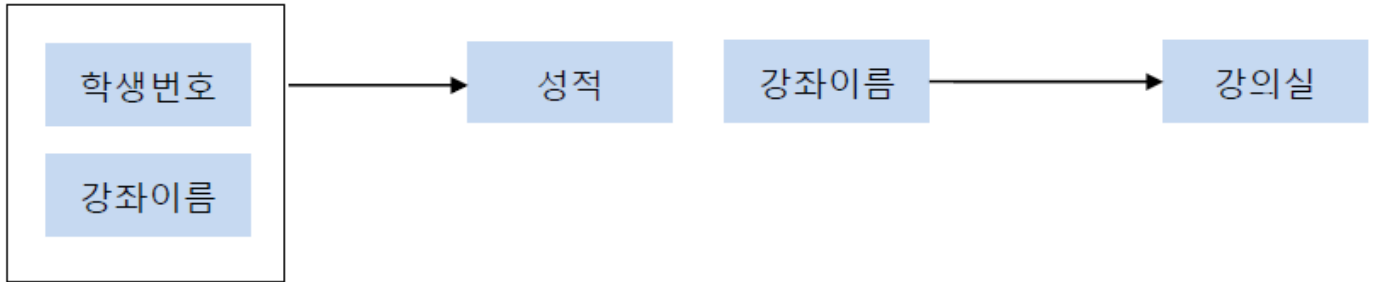
즉, 기본키(학생번호, 강좌이름)의 부분키인 강좌이름이 결정자이기 때문에 위의 테이블의 경우 다음과 같이 기존의 테이블에서 강의실을 분해하여 별도의 테이블로 관리하여 제2 정규형을 만족시킬 수 있다.

수강

학생번호	강좌이름	성적
501	데이터베이스	3.5
401	데이터베이스	4.0
402	스포츠경영학	3.5
502	자료구조	4.0
501	자료구조	3.5

강의실

강좌이름	강의실
데이터베이스	공학관 110
스포츠경영학	체육관 103
자료구조	공학관 111

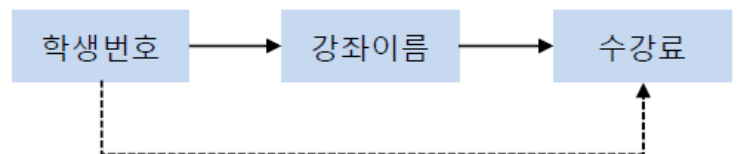


[제3 정규화]

제3 정규화란 제2 정규화를 진행한 테이블에 대해 이행적 종속을 없애도록 테이블을 분해하는 것이다. 여기서 이행적 종속이라는 것은 $A \rightarrow B$, $B \rightarrow C$ 가 성립할 때 $A \rightarrow C$ 가 성립되는 것을 의미한다. 예를 들어 아래와 같은 계열 학기 테이블을 살펴보자.

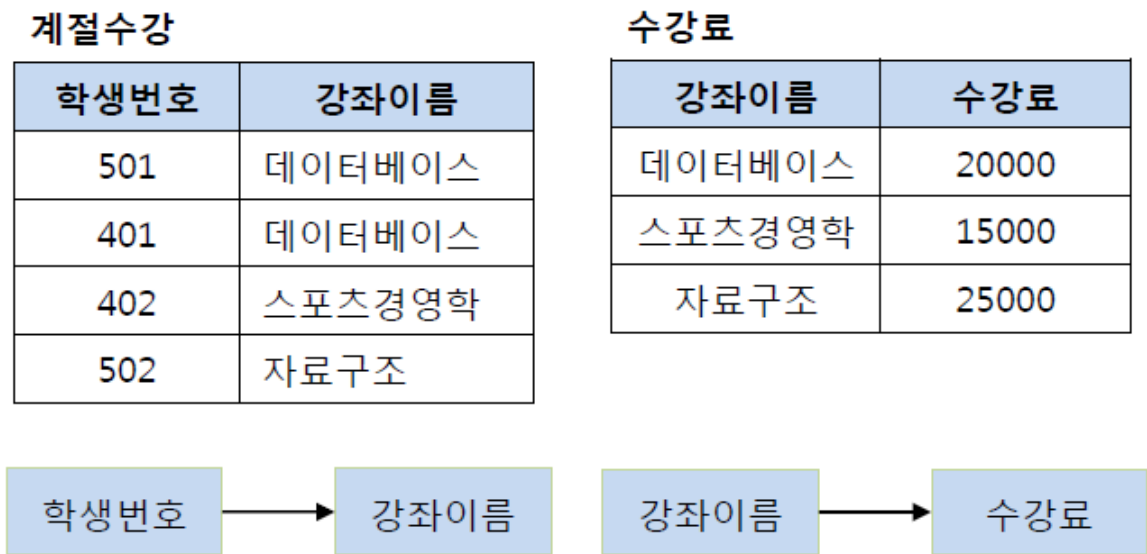
계절학기

학생번호	강좌이름	수강료
501	데이터베이스	20000
401	데이터베이스	20000
402	스포츠경영학	15000
502	자료구조	25000



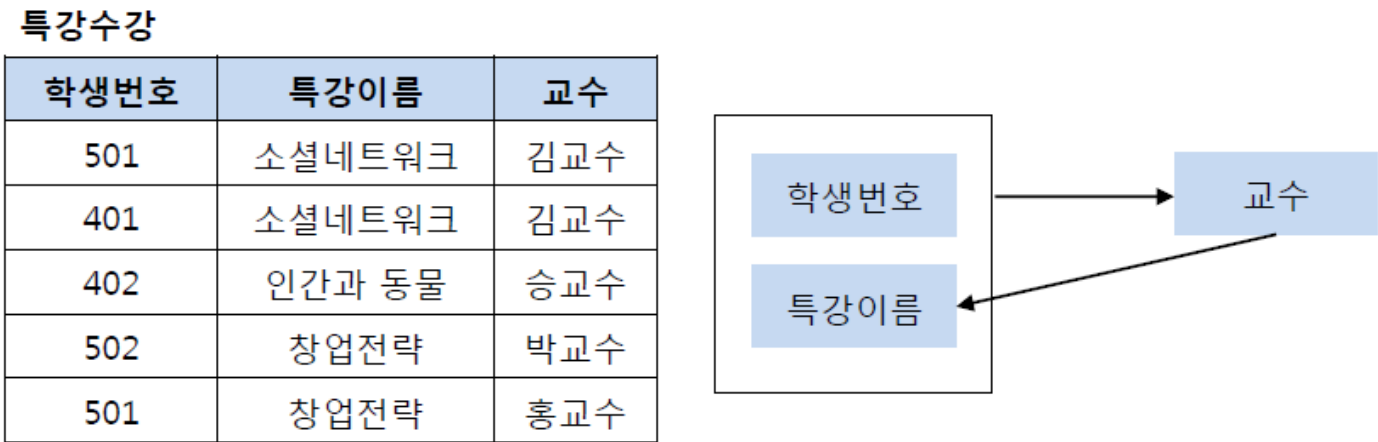
기존의 테이블에서 학생 번호는 강좌 이름을 결정하고 있고, 강좌 이름은 수강료를 결정하고 있다. 그렇기 때문에 이를 (학생 번호, 강좌 이름) 테이블과 (강좌 이름, 수강료) 테이블로 분해해야 한다. 이행적 종속을 제거하는 이유는 비교적 간단하다. 예를 들어 501번 학생이 수강하는 강좌가 스포츠경영학으로 변경되었다고 하자. 이행적 종속이 존재한다면 501번의 학생은 스포츠경영학이라는 수업을 20000원이라는 수강료로 듣게 된다. 물론 강좌 이름에 맞게 수강료를 다시 변경할 수 있지만, 이러한 번거로움을 해결하기 위해 제3 정규화를 하는 것이다.

즉, 학생 번호를 통해 강좌 이름을 참조하고, 강좌 이름으로 수강료를 참조하도록 테이블을 분해해야 하며 그 결과는 다음의 그림과 같다.



[BCNF 정규화]

BCNF 정규화란 제3 정규화를 진행한 테이블에 대해 모든 결정자가 후보키가 되도록 테이블을 분해하는 것이다. 예를 들어 다음과 같은 특강수강 테이블이 존재한다고 하자.



특강수강 테이블에서 기본키는 (학생번호, 특강이름)이다. 그리고 기본키 (학생번호, 특강이름)는 교수를 결정하고 있다. 또한 여기서 교수는 특강이름을 결정하고 있다. 그런데 문제는 교수가 특강이름을 결정하는 결정자이지만, 후보키가 아니라는 점이다. 그렇기 때문에 BCNF 정규화를 만족시키기 위해서 위의 테이블을 분해해야 하는데, 다음과 같이 특강신청 테이블과 특강교수 테이블로 분해할 수 있다.

특강신청

학생번호	교수
501	김교수
401	김교수
402	승교수
502	박교수
501	홍교수

특강교수

특강이름	교수
소셜네트워크	김교수
인간과 동물	승교수
창업전략	박교수
창업전략	홍교수

학생번호

교수

특강이름

교수

