

# ERD와 정규화 과정

≡ 태그

데이터베이스 - ERD와 정규화 과정

## ERD(Entity Relationship Diagram)

엔티티 간의 관계를 다이어그램으로 표현한 것이며 데이터베이스 구축의 설계도 역할을 한다. 관계형 구조로 표현할 수 있는 데이터를 구성하는데는 유용하지만 비정형 데이터를 충분히 표현할 수 없다는 단점도 있다.

## 데이터베이스 정규화

### 데이터베이스 정규화

잘못된 릴레이션 간의 종속관계는 데이터베이스 이상현상을 발생시킬 수 있다. 데이터베이스 정규화는 이런 이상 현상을 해결하는 과정으로 관계형 데이터베이스에서 데이터 모델의 중복을 최소화하고 데이터의 일관성, 유연성을 확보하기 위한 목적으로 데이터를 분해하는 과정을 말한다. 테이블 간의 중복된 데이터를 허용하지 않음으로써 데이터 무결성을 유지할 수 있고 DB의 저장용량을 줄일 수 있다.

데이터베이스 정규화 과정은 정규형 원칙을 기반으로 정규형을 만들어가는 과정이며 정규화된 정도는 정규형(NF, Normal Form)으로 표현한다. 기본정규형인 제1,2,3 정규형과 BCNF 그리고 고급 정규형인 제4,5 정규형이 있다.

### 이상현상(Anomaly)

삭제이상 : 데이터 삭제 시 의도하지 않은 다른 정보까지 연쇄적으로 삭제되는 현상.

삽입이상 : 데이터 삽입 시 특정 데이터가 존재하지 않아 데이터베이스에 데이터를 삽입할 수 없는 현상.

수정이상 : 데이터 수정 시 중복된 데이터 중 일부 데이터만 수정되어 데이터 불일치가 일어나는 현상.

## 정규형 원칙

1. 같은 의미를 표현하는 릴레이션이지만 더 좋은 구조로 구성
2. 자료의 중복성 감소
3. 독립적인 관계는 별개의 릴레이션으로 표현
4. 각 릴레이션은 독립적 표현이 가능

## 제1정규형

테이블의 모든 도메인이 더 이상 분해될 수 없는 원자 값 만으로 구성되어야 한다.

유저번호	유저ID	수강명	성취도
1	홍철	{C++코딩테스트, 프런트특강}	{90%, 10%}
2	범석	{코드포스특강, DS특강}	{7%, 8%}



유저번호	유저ID	수강명	성취도
1	홍철	C++코딩테스트	90%
1	홍철	프런트특강	10%
2	범석	코드포스특강	7%
2	범석	DS특강	8%

## 제2정규형

테이블이 제1정규형을 만족하며 부분 함수의 종속성을 제거한 상태를 말한다. 부분 함수의 종속성 제거란 기본키가 아닌 모든 속성이 기본키에 완전 함수 종속적인 것을 말한다.

유저번호	유저ID	수강명	성취도
1	홍철	C++코딩테스트	90%
1	홍철	프런트특강	10%
2	범석	코드포스트특강	7%
2	범석	DS특강	8%



유저번호	유저ID	유저ID	수강명	성취도
1	홍철	홍철	C++코딩테스트	90%
2	범석	홍철	프런트특강	10%
		범석	코드포스트특강	7%
		범석	DS특강	8%

## 함수 종속성

함수 종속성은 어떤 속성 A의 값을 알면 다른 속성 B의 값이 유일하게 정해지는 관계를 종속성이라고 한다.

- A->B로 표기하며 A를 B의 결정자(Determinant)라고 한다.
- A->B이면 A는 B를 결정한다(Determine) 한다고 하고, B는 A에 종속한다(Dependent)라고 한다.

## 부분함수 종속

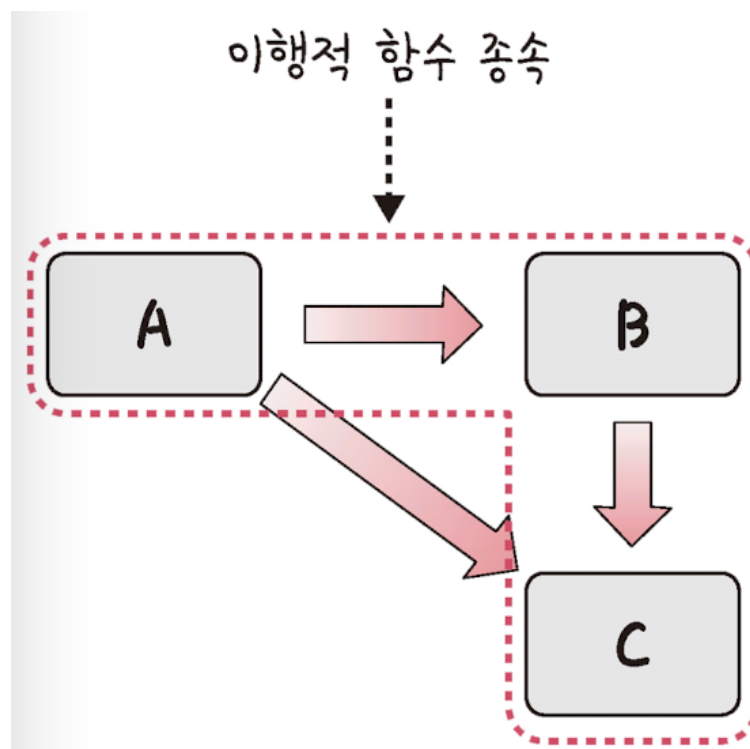
기본키가 2개 이상의 컬럼으로 이루어진 경우에 발생하며, 기본키의 부분집합에 종속되는 관계를 말한다.

### 완전 함수 종속

기본키가 2개 이상의 컬럼으로 이루어진 경우, 기본키의 부분집합이 결정자가 되어서는 안되는 관계를 말한다.

## 제3정규형

제2정규형을 만족하고 기본키가 아닌 모든 속성이 이행적 함수 종속을 만족하지 않는 상태를 말한다.



### 이행적 함수 종속

$A \rightarrow B$ 와  $B \rightarrow C$ 가 존재한다고 할때 논리적으로  $A \rightarrow C$  관계가 성립하는데, 이 때 집합 C가 집합 A에 이행적으로 함수 종속되었다고 한다.

## BCNF 정규형

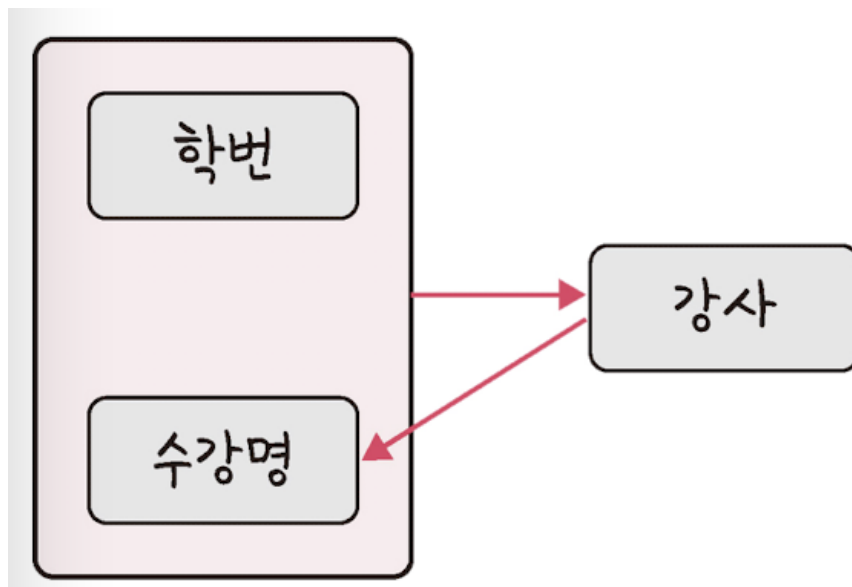
보이스/코드정규형은 제3정규형이고, 결정자가 후보키가 아닌 함수 종속 관계를 제거하여 테이블의 함수 종속 관계에서 모든 결정자가 후보키인 상태를 말한다.

요구 사항은 다음과 같다고 해봅시다.

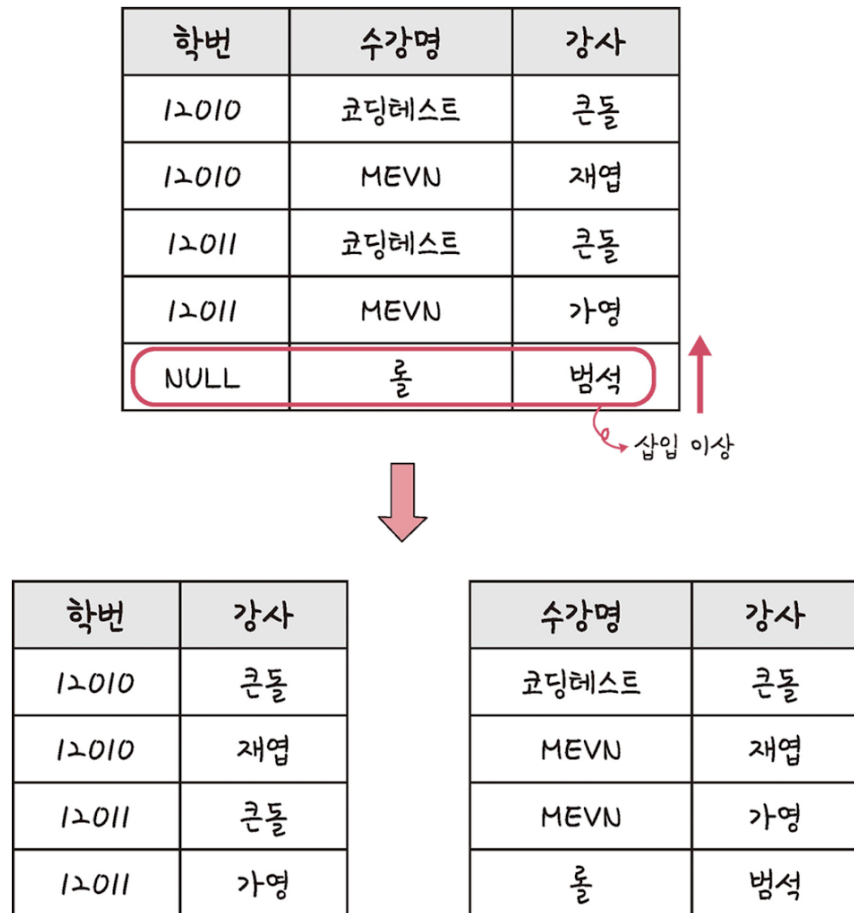
- 각 수강명에 대해 한 학생은 오직 한 강사의 강의만 수강한다.
- 각 강사는 한 수강명만 담당한다.
- 한 수강명은 여러 강사가 담당할 수 있다.

학번	수강명	강사
12010	코딩테스트	큰돌
12010	MEVN	재엽
12011	코딩테스트	큰돌
12011	MEVN	가영
NULL	롤	범석

↑  
삽입 이상



위 그림처럼 강사 속성은 후보키가 아니지만 결정자 역할을 하고 있으므로 테이블에서 이 강사속성을 아래처럼 나누어야 한다.



## 결정자

함수 종속 관계에서 특정 종속자를 결정짓는 요소,  $X \rightarrow Y$  일 때 X는 결정자, Y는 종속자이다.

## 정규화의 장단점

### [정규화의 장점]

- 데이터베이스의 이상현상을 제거할 수 있다.
- 데이터의 무결성 유지

- 데이터베이스 저장공간의 효율적 사용

#### [정규화의 단점]

- 테이블의 분해로 인해 데이터 조회 시 JOIN연산이 증가로 인한 질의 응답시간 저하

## 데이터 무결성

데이터의 정확성과 일관성을 유지하고 보증하는 것.

## 데이터베이스 반정규화

데이터베이스의 성능향상을 위해 데이터의 중복을 허용하고 조인을 줄이는 데이터베이스 성능 향상 기법을 말한다. 정규화된 데이터 모델을 통합, 중복하여 의도적으로 정규화 원칙을 위배한다.

## 반정규화의 장단점

#### [반정규화 장점]

- 빠른 데이터 조회 가능
- 데이터 조회 쿼리가 간단해진다.

#### [반정규화 단점]

- 데이터 갱신이나 삽입비용이 높다.
- 데이터 일관성이 깨질 수 있다.
- 데이터 저장공간의 사용이 비효율적이다.