



# Normalization

⚙ 상태	진행 중
📂 분류	Database
🕒 주차	1주차

## 정규화(Normalization)

- 정규화의 기본 목표
  - 중복된 데이터를 허용하지 않는 것
  - 중복된 데이터를 허용하지 않음으로써 이상 현상을 방지하여 데이터 무결성을 유지하기 위함

### 이상 현상

- 데이터 중복으로 인해 발생할 수 있음
- 갱신 이상
  - 중복된 데이터 중 일부를 갱신할 때 의도치 않은 데이터가 갱신됨으로써 생기는 데이터의 불일치
- 삽입 이상
  - 새 데이터를 삽입할 때 의도치 않은 데이터가 삽입됨으로써 생기는 데이터의 불일치
- 삭제 이상
  - 데이터를 삭제할 때 의도치 않은 데이터까지 삭제됨으로써 생기는 데이터의 불일치

### 제1 정규화

- 테이블의 컬럼이 원자값을 갖도록 테이블을 분해한다
  - 추가로, 유사한 속성이 중복, 반복되는 경우도 분해 대상이다
- ex) 고객취미 테이블
  - 추신수와 박세리는 제1 정규형을 만족하지 못하고 있기 때문에 분해

고객취미들(이름, 취미들)

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

고객취미(이름, 취미)

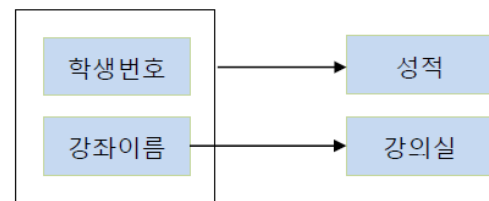
이름	취미
김연아	인터넷
추신수	영화
추신수	음악
박세리	음악
박세리	쇼핑
장미란	음악
박지성	게임

## 제2 정규화

- 제1 정규형을 만족하는 테이블에서 부분 함수 종속을 제거한다
  - 부분적 함수 종속 : 기본키의 부분집합이 결정자가 되는 것
- ex) 수강강좌 테이블
  - 수강강좌 테이블의 기본키는 (학생번호, 강좌이름)
  - (학생번호, 강좌이름)에 의해 성적을 결정할 수 있음
  - 그러나 강의실은 기본키의 부분집합인 (강좌이름)만으로 결정이 가능 → 부분 함수 종속 존재

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



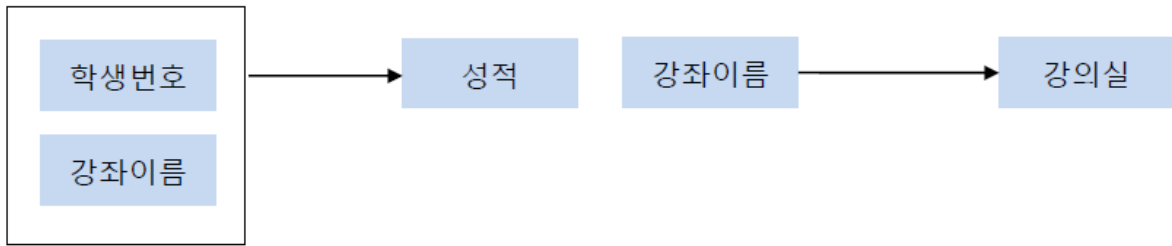
- 따라서, 아래와 같이 테이블을 분리하여 부분 함수 종속을 제거할 수 있음
  - 수강 테이블과 강의실 테이블 모두 제2 정규형을 만족

### 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

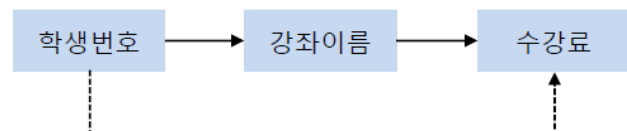


## 제3 정규화

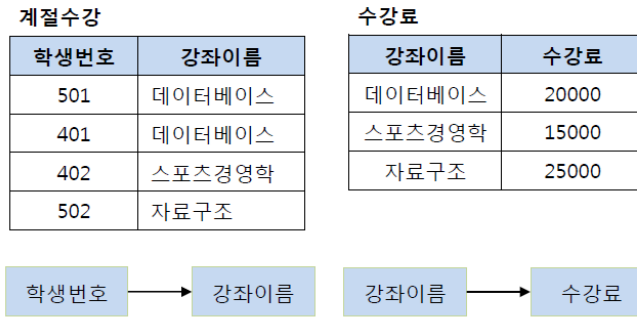
- 제2 정규형을 만족하는 테이블에서 **이행적 종속**을 제거한다
  - 이행적 종속 : 테이블에 존재하는 필드의 부분 집합을 X와 Y라고 함
  - X의 한 값이 Y에 속한 하나의 값에만 매핑이 될 경우 Y는 X에 함수 종속적이다 라고 하며, 이를  $X \rightarrow Y$ 로 표기
    - 이때 X를 결정자, Y를 종속자 라고 한다
  - $A \rightarrow B, B \rightarrow C$ 가 성립할 때  $A \rightarrow C$ 가 성립되는 관계를 이행적 종속이라고 함
- ex) 계절학기 테이블
  - 학생번호 속성은 강좌이름을 결정하고, 강좌이름 속성은 수강료를 결정

### 계절학기

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



- 학생번호가 수강료를 결정하게 되므로 제3 정규형을 만족시키기 위해 테이블 분리가 필요

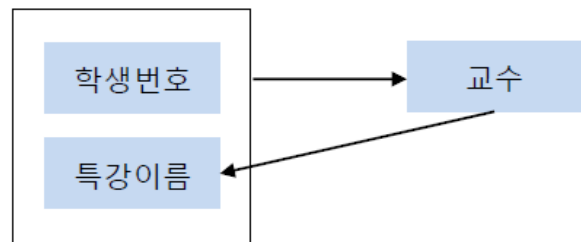


## BCNF 정규화

- 제3 정규형을 만족하는 테이블에서 모든 결정자가 후보키가 되도록 테이블을 분리
  - 후보키 : 유일성과 최소성을 모두 만족하는 속성들의 집합
  - 유일성 : 하나의 키 값으로 튜플을 식별할 수 있는 성질
  - 최소성 : 키를 구성하는 속성들 중 꼭 필요한 최소한의 속성들로만 키를 구성하는 성
- 대부분의 릴레이션은 BCNF 정규형까지 만족하면 실질적인 이상현상이 없어지기 때문에 그 이상 정규화를 하지 않음
- ex) 특강수강 테이블
  - 아래 테이블의 기본키는 (학생번호, 특강이름)
  - 기본키는 교수를 결정하고 부분 함수 종속, 이행적 종속이 없으므로 제3 정규형 만족
  - 후보키가 아닌 교수 속성이 특강이름을 결정

특강수강

학생번호	특강이름	교수
501	소셜네트워크	김교수
401	소셜네트워크	김교수
402	인간과 동물	송교수
502	창업전략	박교수
501	창업전략	홍교수



- 교수가 특강이름을 결정하는데 교수는 후보키가 아니므로 BCNF 정규형을 만족하기 위해서는 테이블 분리가 필요

특강신청

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

특강교수

특강이름	교수
소셜네트워크	김교수
인간과 동물	송교수
창업전략	박교수
창업전략	홍교수



## 제4 정규화

- BCNF 정규형을 만족하는 테이블에서 다치 종속을 제거
  - 다치 종속이란 하나의 릴레이션에서 속성이 1:N 관계로 대응하는 것
    - $X \rightarrow Y$ 일 때 하나의 X값에 여러 개의 Y값이 존재하면 다치 종속성을 가진다 하고  $X \twoheadrightarrow Y$ 로 표시
- ex) 학생 테이블
  - 왼쪽 릴레이션은 제1 정규형을 만족하지 않기 때문에 오른쪽과 같이 테이블을 나눠준다
  - 오른쪽 릴레이션은 BCNF 정규형을 만족한다
    - $이름 \rightarrow 학과$ ,  $이름 \rightarrow 동아리$  모두 1:N 관계이지만 학과와 동아리 사이엔 아무런 관련이 없다

이름	학과	동아리	이름	학과	동아리
다한	컴퓨터공학과	밴드부 토론동아리	다한	컴퓨터공학과	밴드부
			다한	컴퓨터공학과	토론동아리
원기	디자인과 컴퓨터공학과	밴드부 축구부	원기	디자인과	밴드부
			원기	디자인과	축구부
정은	기계공학과 컴퓨터공학과	밴드부	원기	컴퓨터공학과	밴드부
			원기	컴퓨터공학과	축구부
			정은	기계공학과	밴드부
			정은	컴퓨터공학과	밴드부

- 위의 오른쪽 예시에서 원기가 동아리를 하나 더 가입하면 오른쪽과 같이 2개의 튜플이 추가되어야 하는 문제점이 생긴다

이름	학과	동아리
다한	컴퓨터공학과	밴드부
다한	컴퓨터공학과	토론동아리
원기	디자인과	밴드부
원기	디자인과	축구부
원기	디자인과	탁구부

이름	학과	동아리
원기	컴퓨터공학과	밴드부
원기	컴퓨터공학과	축구부
원기	컴퓨터공학과	탁구부
정은	기계공학과	밴드부
정은	컴퓨터공학과	밴드부

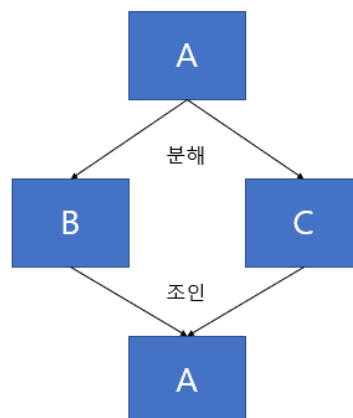
- 따라서 아래와 같이 분해
  - 여전히 다치 종속성을 가지지만, 2개 이상의 컬럼이 하나의 컬럼에 다치 종속되지는 않아 제4 정규형을 만족하게 된다

이름	학과
다한	컴퓨터공학과
원기	디자인과
원기	컴퓨터공학과
정은	기계공학과
정은	컴퓨터공학과

이름	동아리
다한	밴드부
다한	토론동아리
원기	밴드부
원기	축구부
정은	밴드부

## 제5 정규화

- 제4 정규형을 만족하는 테이블에서 **조인 종속**을 제거
  - 조인 종속은 하나의 릴레이션을 여러 개의 릴레이션으로 무손실 분해 후 다시 결합할 수 있으면 조인 종속이라고 한다
  - 아래 예시와 같이 A 릴레이션을 B와 C로 분해했다가 조인했을 때 A가 된다면 A는 조인 종속성이 존재



## 정규화의 문제점

- 정규화는 데이터 조회 시 조인을 유발하기 때문에 CPU와 메모리를 많이 사용한다
- 따라서 이 문제를 반정규화를 통해 해결

## 반정규화

---

- 데이터베이스의 성능 향상을 위해 데이터 중복을 허용하여 조인을 줄이는 성능 향상 방법
- 조회 속도를 향상시키지만 데이터 모델의 유연성은 낮아진다

### 반정규화 장점

- 다음과 같은 경우에서 반정규화를 고려
  - 정규화에 충실해 종속성과 활용성은 향상되었지만 수행 속도가 느려진 경우
  - 다량의 범위를 자주 처리해야하는 경우
  - 특정 범위의 데이터만 자주 처리하는 경우
  - 요약, 집계 정보가 자주 요구되는 경우
- 정규화가 끝난 후 적용

### 반정규화 절차

- 대상 조사 및 검토
  - 데이터 처리 범위, 통계성 등을 확인해서 반정규화 대상을 조사
- 다른 방법 검토
  - 반정규화 수행 전 다른 방법이 있는지 검토
  - 클러스터링, 뷰, 인덱스 튜닝, 파티션 등
    - 클러스터링 : 인덱스 정보를 저장할 때 물리적으로 정렬 후 저장
    - 파티션 : 논리적으로 하나의 테이블이지만 여러 데이터 파일에 분산되어 저장
- 반정규화 수행
  - 테이블, 속성, 관계 등을 반정규화

### 테이블 반정규화

- 테이블 병합
  - 업무 특성상 조인이 필요한 경우가 많아 테이블 통합이 성능상 유리할 때 고려
- 테이블 분할
  - 테이블 수직 분할
    - 엔터티의 일부 속성을 별도의 엔터티로 분할
    - 자주 사용하는 속성이 아니거나 대부분의 값이 NULL일 때 고려

- 테이블 수평 분할
  - 특정 기준으로 엔터티의 인스턴스를 별도의 엔터티로 분할
  - ex) 주문년도에 따라 데이터를 물리적으로 분리 (2021, 2022년)
- 테이블 추가
  - 중복 테이블 추가
  - 통계 테이블 추가
    - 통계치를 미리 계산하여 테이블 생성
  - 이력 테이블 추가
  - 부분 테이블 추가

## 속성(컬럼) 반정규화

- 중복 컬럼 추가
  - 조인 연산 횟수가 많을 경우
- 파생 컬럼 추가
  - 계산 값을 미리 추가하는 방식
  - ex) 프로모션 적용 할인가
- 이력테이블 컬럼 추가

## 관계 반정규화

- 조인 연산 횟수가 많아 중복 관계를 추가하는 것이 성능상 유리할 때 사용
- 데이터 무결성을 깨뜨릴 위험성이 없이 데이터 처리 성능 향상이 가능한 반정규화 기법임