

정규화(Normalization)

정의

- ✓ 테이블의 속성들이 상호 종속적인 관계를 갖는 특성을 이용하여 테이블을 무손실 분해하는 과정
- ✓ 정규화의 목적 : 가능한 한 중복을 제거하여 삽입, 삭제, 갱신 이상의 발생 가능성을 줄이는 것

- 데이터베이스 정규화 단계 < 도 / 부 / 이 / 결 / 다 / 조 >

단계	조건
1정규형(1NF)	도메인이 원자 값으로 구성
2정규형(2NF)	부분 함수 종속 제거(완전 함수적 종속 관계)
3정규형(3NF)	이행 함수 종속 제거
보이스-코드 정규형(BCNF)	결정자가 후보 키가 아닌 함수 종속 제거
4정규형(4NF)	다치(다중 값) 종속성 제거
5정규형(5NF)	조인 종속성 제거

[상세 설명]

1 정규형(1NF) - 도메인이 원자 값으로 구성의 의미는 말 그대로 테이블의 컬럼이 원자값(하나의 값)을 갖도록 테이블을 분해해야 한다는 것이다.

아래와 같은 취미 테이블이 있다고 했을 시

이름	취미
윤파고	게임
송이	영화,음악
너굴맨	게임,쇼핑
보로로	쇼핑

위의 테이블에서 송이와 너굴맨은 여러가지 취미를 가지고 있기 때문에 제 1정규형을 만족하지 못하고 있다. 그렇기에 이를 제 1정규화를 한다면 아래와 같이 분해할 수 있다.

이름	취미
윤파고	게임
송이	영화
송이	음악
너굴맨	게임
너굴맨	쇼핑
보로로	쇼핑

2 정규형(2NF) - 부분 함수 종속 제거(완전 함수적 종속 제거)란 제 1정규화를 진행한 테이블에 대해 완전 함수 종속을 만족할 수 있도록 테이블을 분해하는 것이다. 여기서 완전 함수 종속이라는 것은 기본키의 부분집합이 결정자가 되어서는 안된다는 것을 의미한다. 기본키와 부분집합의 결정자라는 내용이 무엇인지 모를 수 있으니 아래의 예시를 보면서 설명하도록 하겠다.

예를 들어 아래와 같은 수강생 테이블이 있다고 가정하자

학생번호	강좌 명	강의실	성적
1	컴퓨터공학	101	3.5
2	컴퓨터물리학	102	3.3
3	양자역학	103	4.0
4	자료구조	104	3.8
1	자료구조	104	3.5

기본 키란 관계형 데이터베이스에서 레코드의 식별자로 이용하기에 적합한 키를 말하는데, 해당 테이블에서 기본키는 '학생번호'와 '강좌 명'을 합친 복합키이다. (학생번호나 강좌 명 자체가 중복 값이 있기에 하나만 기본 키로 사용할 수 없다)

그리고 '학생번호'와 '강좌 명'으로 된 기본키는 '성적'을 결정하고 있다.

그런데 여기서 강의실이라는 컬럼은 기본키의 부분집합인 강좌이름에 의해 결정될 수 있다.

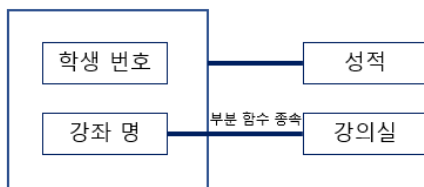
즉, 기본키('학생번호'+ '강좌 명')의 부분키인 강좌이름이 강의실이라는 컬럼값을 결정짓는 '결정자'이기 때문에 위의 테이블의 경우 아래와 같이 기존의 테이블에서 강의실을 분해하여 별도의 테이블로 관리하면 제 2 정규형을 만족시킬 수 있다.

학생번호	강좌 명	성적
1	컴퓨터공학	3.5
2	컴퓨터물리학	3.3
3	양자역학	4.0
4	자료구조	3.8
1	자료구조	3.5

강좌 명	강의실
컴퓨터공학	101
컴퓨터물리학	102
양자역학	103
자료구조	104

추가적으로 이를 그림으로 표현하면 아래와 같이 표현할 수 있다.

기본키 = 학생 번호 + 강좌 명

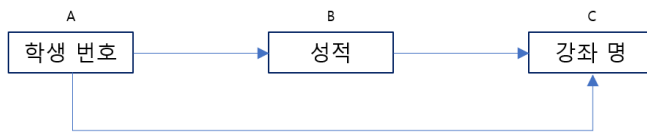


3 정규형(3NF) - 이행 함수 종속 제거란 제 2 정규화를 진행한 테이블에 대해 이행적 종속을 없애도록 테이블을 분해하는 것이다. 여기서 이행적 종속이라는 것은 $A \rightarrow B, B \rightarrow C$ 가 성립할 때 $A \rightarrow C$ 가 성립하는 것을 의미한다.

아래의 예시를 보도록 하자

학생번호	강좌 명	수강료
1	컴퓨터공학	10000
2	컴퓨터공학	10000
3	양자역학	20000
4	자료구조	40000

해당 테이블에서 학생 번호는 강좌 명을 결정하고 있고, 강좌 명은 수강료를 결정하고 있다. $A \rightarrow B \rightarrow C$ 가 성립하고 있는 것이다.



따라서 이를 각각 (학생번호, 강좌명) 레이블과 (강좌명, 수강료) 레이블로 분할을 해주어야 하며, 3정규형을 진행했을 시 레이블은 아래와 같다.

학생번호	강좌 명
1	컴퓨터공학
2	컴퓨터공학
3	양자역학
4	자료구조

강좌 명	수강료
컴퓨터공학	10000
양자역학	20000
자료구조	40000

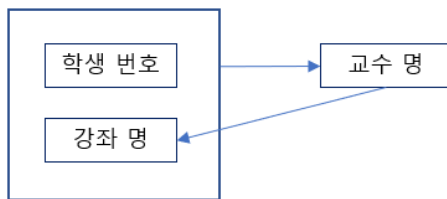
3 정규형이 필요한 이유는 만약 1번 학생이 수강하려는 강좌를 양자역학으로 바꾸려고 하는 경우를 생각해보면 된다. 이행적 종속이 존재하는 상태에서는 1번 학생은 양자역학을 10000원 내고 수업을 들을 수 있다. 물론 강좌 이름에 맞게 수강료를 다시 변경하면 되지만, 이러한 번거로움을 해결하기 위해 3 정규화를 진행하는 것이다.

따라서 학생 번호를 통해 강좌 명을 참조하고, 강좌 명이 수강료를 참조하도록 레이블을 분해해야 한다.

보이스-코드 정규형(BCNF) - 결정자가 후보키가 아닌 함수의 종속을 제거한다는 말의 의미는 테이블의 모든 결정자가 후보키가 되도록 테이블을 분해한다는 것을 의미한다. 아래의 예시를 보도록 하자.

학생번호	강좌 명	교수 명
1	컴퓨터공학	윤교수
2	컴퓨터공학	윤교수
3	디지털공학	오교수
4	양자역학	박교수
1	양자역학	최교수

위의 강좌 신청 테이블에서 기본키는 (학생번호, 강좌 명)이다. 그리고 기본키는 교수 명을 결정하고 있다. 그런데 여기서 교수 명은 특강이름을 결정짓고 있기도 하다. 이를 그림으로 확인하면 아래와 같다.



그런데 이렇게 됐을 시 문제는 교수 명이 특강이름을 결정하는 결정자이지만, 후보키는 아니라는 점이다. 그렇기 때문에 BCNF 정규화를 통해 위의 테이블을 분해해야 하며, BCNF를 진행하고 나면, 아래와 같이 테이블을 분해할 수 있다.

학생번호	교수 명
1	윤교수
2	윤교수
3	오교수
4	박교수
1	최교수

강좌 명	교수 명
컴퓨터공학	윤교수
디지털공학	오교수
양자역학	박교수
양자역학	최교수

4정규형과 5정규형은 아래 정도의 문구만 알고 있으면 됩니다. 구체적인 개념을 묻기에는 너무 깊어서 딱 문구 정도만 시험에서 다루고 있습니다.

4 정규형 - 다치(다중 값) 종속성 제거

5 정규형 - 조인 종속성 제거

1.4.4.2 반정규화(Denormalization) = 비정규화 = 역정규화

정의

- ✓ 시스템의 성능을 향상하고 개발 및 운영의 편의성 등을 높이기위해 정규화된 데이터 모델을 의도적으로 통합, 중복, 분리하여 정규화 원칙을 위배하는 행위
- ✓ 정규화된 엔티티, 속성, 관계를 시스템의 성능 향상과 개발 운영의 단순화를 위해 중복, 통합, 분리 등을 수행하는 데이터 모델링 기법
- ✓ 반정규화를 수행하면 시스템의 성능이 향상되고 관리 효율성은 증가하지만, 데이터의 일관성 및 정합성이 저하될 수 있음
- ✓ 과도한 반정규화는 오히려 성능을 저하시킬 수 있음

- 반 정규화 방법

테이블 통합(병합)	- 두 개의 테이블이 조인되어 사용되는 경우가 많을 경우 성능 향상을 위해 하나의 테이블로 만들어 사용하는 것
테이블 분할	- 테이블의 수직 또는 수평으로 분할하는 것 - 수직 분할 : 속성 기준으로 테이블을 분할 - 수평 분할 : 행 단위 기준으로 테이블을 분할
중복 테이블 추가	- 작업의 효율성을 향상시키기 위해 테이블을 추가하는 것
중복 속성 추가	- 조인해서 데이터를 처리할 때 데이터를 조회하는 경로를 단축하기 위해 자주 사용하는 속성을 하나 더 추가하는 것 - 중복 속성을 추가하면 데이터의 무결성 확보가 어렵고, 디스크 공간이 추가로 필요함