

SECTION 4.2 ERD와 정규화 과정

ERD(Entity, Relationship Diagram)

→ MySQL 등에서 볼 수 있음

- 데이터베이스를 구축할 때 가장 기초적인 빠대 역할을 하며, 릴레이션 간의 관계들을 정의한 것
- 요구 분석 사항에서 얻은 엔티티와 속성들의 관계를 그림으로 나타낸 개체-관계 모델

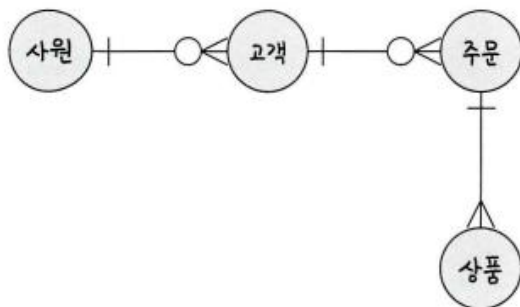
Ex)

요구 사항

- 영업사원은 0 ~ n명의 고객을 관리한다.
- 고객은 0 ~ n개의 주문을 넣을 수 있다.
- 주문에는 1 ~ n개의 상품이 들어간다.

정답

▼ 그림 4-17 승원 영업부서의 ERD

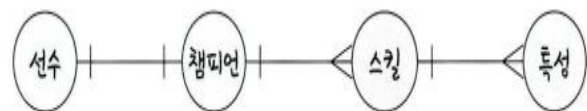


요구 사항

- 선수들은 1명의 챔피언을 고를 수 있다.
- 챔피언은 한 개 이상의 스킬을 갖는다.
- 스킬은 한 개 이상의 특성을 갖는다.

정답

▼ 그림 4-18 무무오브레전드의 ERD



1. ERD의 중요성

- 서비스를 구축할 때 가장 먼저 신경 써야 할 부분
- 관계형 구조로 표현할 수 있는 데이터를 구성하는 데 유용할 수 있지만
비정형 데이터를 충분히 표현할 수 없다는 단점

→ 각 데이터의 종류는 알 수 있음

2. 정규화 과정

- 정규화(Normalization)는 관계형 데이터베이스 설계에서 중복을 최소화하고 데이터 무결성을 보장하기 위해 데이터를 구조화하는 과정
- 정규형 원칙을 기반으로 정규형을 만들어가는 과정

- 데이터베이스 이상 현상

정규화되지 않은 테이블에서 릴레이션 조작 시 발생하는 문제

삽입: 특정 데이터를 저장하기 위해 불필요한 데이터까지 함께 저장해야 하는 문제

삭제: 특정 데이터를 삭제할 때 의도치 않게 다른 필요한 데이터까지 함께 삭제되는 문제

갱신: 중복된 데이터 중 일부만 변경되어 데이터 불일치가 발생하는 문제

Ex)

+어떤 테이블에 직원의 정보를 저장하는데, 특정 부서에 아직 소속되지 않은 직원의 데이터를 삽입할 때 부서 정보가 없어 삽입이 불가능한 경우

-부서 정보가 특정 직원에게만 연결돼 있고 그 직원이 퇴사하면서 부서 정보도 함께 삭제되는 경우

*부서 정보가 여러 직원에게 중복되어 저장되어 있고, 부서 정보를 수정할 때 일부 직원만 수정되어 일관성이 깨지는 경우

3. 정규형 원칙

같은 의미를 표현하는 릴레이션의 경우

- 좀 더 좋은 구조로 만들어야 함
- 자료의 중복성은 감소해야 함
- 독립적인 관계는 별개의 릴레이션으로 표현해야 함
- 각각의 릴레이션은 독립적인 표현이 가능해야 함

→ 12 관계를 갖지 않은 것은 별개의 릴레이션으로

4. 정규형 단계

정규화는 단계적으로 진행되며, 각 단계는 이전 단계의 조건을 모두 만족하면서 추가적인 조건을 충족시켜 나감

1) 제1 정규형

모든 속성이 원자값(Atomic Value)을 가지는 형태

- 테이블의 각 셀은 하나의 값만 가져야 함 *→ 스칼라 해야 함*
- 같은 유형의 데이터가 여러 속성(컬럼)으로 반복되지 않아야 함
- 각 레코드는 유일한 식별자인 기본 키를 가져야 함

▼ 그림 4-19 제1정규형

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



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

2) 제2 정규형

1NF를 만족하면서, 부분 함수적 종속(Partial Functional Dependency)이 없는 형태

- 기본키가 아닌 모든 속성이 기본키에 완전 함수 종속적이어야 함
- 릴레이션 분해시 동등한 릴레이션으로 분해해야 하고, 정보 손실이 발생하지 않는 무손실 분해가 돼야 함

▼ 그림 4-20 제2정규형

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



유저번호	유저ID
1	홍철
2	범석

유저ID	수강명	성취도
홍철	C++코딩테스트	90%
홍철	프런트특강	10%
범석	코드포스트강	7%
범석	DS특강	8%

애각

애각

종속적이지 X

3) 제3 정규형

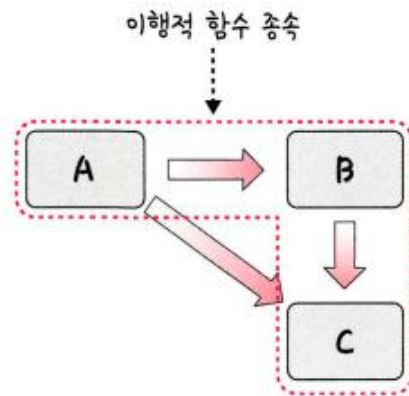
2NF를 만족하면서, 이행적 종속(Transitive Dependency)이 없는 형태

→ 이행적 함수 종속

; A->B와 B->C가 존재하면 논리적으로 A->C가 성립하는데, 이때 집합 C가 집합A에 이행적으로 함수 종속 되었다고 함.

; 키가 아닌 속성이 다른 키가 아닌 속성에 종속되는 경우

▼ 그림 4-21 이행적 함수 종속



▼ 그림 4-22 제3정규형

유저ID	등급	할인율
홍철	플래티넘	30%
범수	다이아	50%
가영	마스터	70%

유저ID	등급
홍철	플래티넘
범수	다이아
가영	마스터

등급	할인율
플래티넘	30%
다이아	50%
마스터	70%

4) 보이스/코드 정규형(BCNF)

3NF를 만족하면서, 모든 결정자(Determinant)가 후보 키(Candidate Key)인 형태

→ 결정자

; 함수 종속 관계에서 특정 종속자(dependent)를 결정짓는 요소

; X->Y 일 때 'X'는 결정자, 'Y'는 종속자

Ex) 예시 요구사항

각 수강명에 대해 한 학생은 오직 한 강사의 강의만 수강

각 강사는 한 수강명만 담당

한 수강명은 여러 강사가 담당 가능

▼ 그림 4-25 보이스/코드 정규형을 만족한 릴레이션

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



학번	강사
12010	큰돌
12010	재엽
12011	큰돌
12011	가영

수강명	강사
코딩테스트	큰돌
MEVN	재엽
MEVN	가영
롤	범석

5. 정규화 예제

0) 비정형 데이터

고객취미들(이름, 취미들)

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

- 추신수와 박세리는 여러 개의 취미를 가지고 있기 때문에 제1 정규형을 만족하지 못하고 있음
- 제1 정규화하여 분해

1) 1 NF

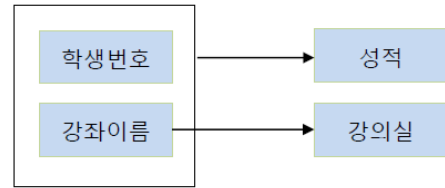
고객취미(이름, 취미)

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

2) 2 NF

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



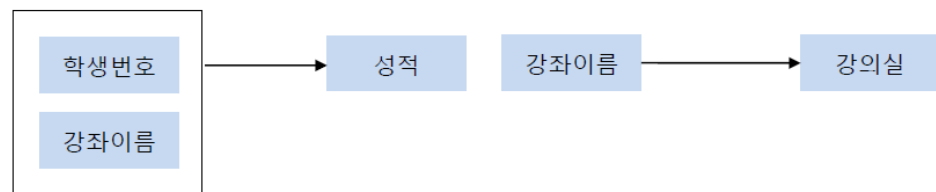
- 이 테이블에서 기본키는 (학생번호, 강좌이름)으로 복합키이며, 성적을 결정하고 있음
(학생번호, 강좌이름) --> (성적)
- 강의실이라는 컬럼은 기본키의 부분집합인 강좌이름에 의해 결정될 수 있음
(강좌이름) --> (강의실)
- 즉, 기본키(학생번호, 강좌이름)의 부분키인 강좌이름이 결정자이기 때문에 다음과 같이 기존의 테이블에서 강의실을 분해하여 별도의 테이블로 관리하여 제2 정규형을 만족시킬 수 있음

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

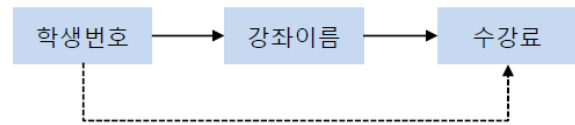
강좌이름	강의실
데이터베이스	공학관 110
스포츠경영학	체육관 103
자료구조	공학관 111



3) 3 NF

계절학기

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



- 학생 번호는 강좌 이름을 결정하고 있고, 강좌 이름은 수강료를 결정하고 있음
- 따라서 이를 (학생 번호, 강좌 이름) 테이블과 (강좌 이름, 수강료) 테이블로 분해해야 함

계절수강

학생번호	강좌이름
501	데이터베이스
401	데이터베이스
402	스포츠경영학
502	자료구조

수강료

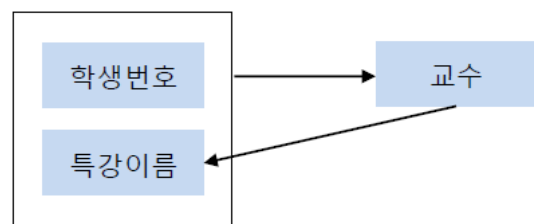
강좌이름	수강료
데이터베이스	20000
스포츠경영학	15000
자료구조	25000



4) BCNF

특강수강

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



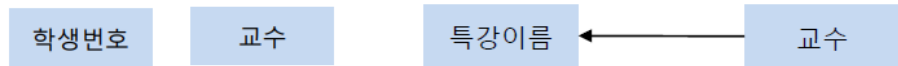
- 기본키는 (학생번호, 특강이름)이고 교수를 결정하고 있음
- 교수는 특강 이름을 결정하고 있음
 - ⇒ 교수가 특강이름을 결정하는 결정자이지만, 후보키가 아님
- BCNF 정규화를 만족시키기 위해서 위의 테이블을 분해해야 함

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



▶▶ 예상 질문

1. ERD와 이를 작성하는 목적과 장점에 대해 설명해주세요.
2. 데이터베이스 이상 현상에 대해 설명해주세요.
3. 정규화와 비정규화에 대해 설명하고, 각각의 장단점을 비교해주세요.
4. 정규화의 필요성과 각 단계별 특징을 설명해주세요.