

4.2 ERD와 정규화 과정

- ERD(Entity Relationship Diagram)은 db를 구축할때 가장 기초적인 뼈대 역할을 하며, 릴레이션 간의 관계들을 정의한 것이다.
- 서비스를 구축한다면 가장 먼저 신경써야할 부분이며 이 부분을 신경쓰지 않고 서비스를 구축한다면 단단하지 않은 골조로 건물을 짓는것이다.

4.2.1 ERD의 중요성

- ERD는 시스템의 요구 사항을 기반으로 작성되며 이를 기반으로 db를 구축한다.
- db를 구축한 이후에도 디버깅 또는 비즈니스 프로세스 재설계가 필요한 경우에 설계도 역할을 담당하기도 한다.
- ERD는 관계형 구조로 표현할 수 있는 데이터를 구성하는데 유용하지만, 비정형 데이터를 충분히 표현할 수 없다.

비정형 데이터

- 비구조화된 데이터를 말하며, 미리 정의된 데이터 모델이 없거나 미리 정의된 방식으로 정리되지 않은 정보를 말한다.

4.2.2 예제로 배우는 ERD

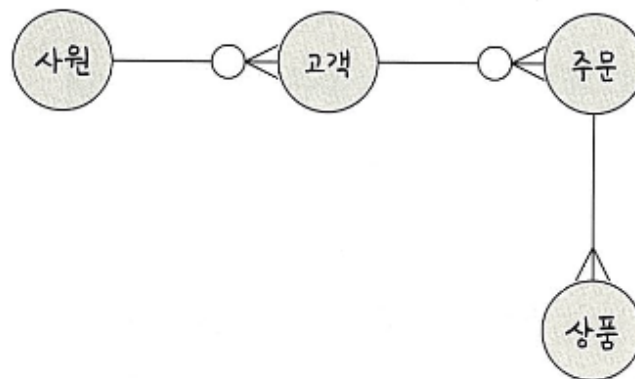
승원 영업부서의 ERD

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



무무오브레전드의 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



4.2.3 정규화 과정

- 정규화 과정은 릴레이션 간의 잘못된 종속 관계로 인한 db이상 현상이 일어나서 이를 해결하거나, 저장공간을 효율적으로 사용하기 위해 릴레이션을 여러 개로 분리하는 과정이다.
- db이상 현상이란 회원이 한 개의 등급을 가져야 하는데 세 개의 등급을 갖거나 삭제할 때 필요한 데이터가 같이 삭제되고, 데이터를 삽입해야하는데, 하나의 필드 값이 NULL이 되면 안되어서 어려운 현상을 말한다.
- - 정규화 과정은 정규형 원칙을 기반으로 정규형을 만들어나가는 과정이며 정규화 된 정도는 정규형으로 표현된다.
- 기본 정규형인 제1정규형, 제2정규형 제3정규형, 보이스/코드 정규형이 있고 고급 정규형인 제4정규형 제5정규형이 있다.

정규형원칙

- 정규형의 원칙이란 같은 의미를 표현하는 릴레이션이지만 좀 더 좋은 구조로 만들어야 하고, 자료의 중복성은 감소해야 하고, 독립적인 관계는 별개의 릴레이션으로 표현해야하며, 각각의 릴레이션은 독립적인 표현이 가능해야 하는 것을 말한다.

제 1 정규형

- 릴레이션의 모든 도메인이 더이상 분해될 수 없는 원자값만으로 구성되어야 한다.

- 릴레이션의 속성 값 중에서 한 개의 기본키에 대해 두개이상의 값을 가지는 반복 집합이 있어서는 안된다.
- 반복 집합이 있다면 제거해야한다.

▼ 그림 4-19 제1정규형

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



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

앞의 그림처럼 홍철이란 ID에 수강명이 {C++코딩테스트, 프런트특강}이 있었는데 이것을 나눠서 반복 집합을 제거하는 것을 볼 수 있습니다.

제2정규형

- 릴레이션이 제1정규형이며 부분함수의 종속성을 제거한 형태를 말한다.
- 부분 함수의 종속성 제거란 기본키가 아닌 모든 속성이 기본키에 완전 함수 종속적인 것을 말한다.

▼ 그림 4-20 제2정규형

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%
		홍철	프런트특강	10%
2	범석	범석	코드포스트특강	7%
		범석	DS특강	8%

앞의 그림을 보면 기본키인 {유저ID, 수강명}과 완전 종속된 유저번호 릴레이션과 '{유저ID, 수강명}'에 따른 성취도' 릴레이션으로 분리된 것을 볼 수 있습니다.

이때 주의할 점은 릴레이션을 분해할 때 동등한 릴레이션으로 분해해야 하고, 정보 손실이 발생하지 않는 무손실 분해로 분해되어야 한다는 것입니다.

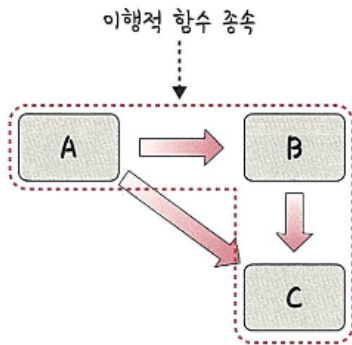
제3정규형

- 제2정규형이고 기본키가 아닌 모든 속성이 이행적 함수 종속을 만족하지 않는 상태를 말한다.

이행적 함수 종속

- 이행적 함수 종속이란 $a \rightarrow b \rightarrow c$ 가 존재하면 논리적으로 $a \rightarrow c$ 가 성립하는데, 이때 집합 c 가 집합 a 에 이행적으로 함수 종속이 되었다고 한다.

▼ 그림 4-21 이행적 함수 종속



예를 들어 무무쇼핑몰이 있다고 해봅시다. 유저ID와 등급, 할인율이 정해져 있는 테이블을 다음과 같이 분해하는 것을 말합니다.

▼ 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%

보이스/코드 정규형

- 보이스/코드 정규형은 제3정규형이고, 결정자가 후보키가 아닌 함수 종속 관계를 제거하여 릴레이션의 함수 종속 관계에서 모든 결정자가 후보키인 상태를 말한다.