

8장 정리 - 정규화

5-1 테이블

5-2 설계 규칙

제 1정규형

제 2정규형

제 3정규형

5-1 테이블

테이블 == 어떤 공통의 속성을 가진 것의 집합

- 데이터를 관리하는 그릇.
- 공통적인 요소의 집합
 - 테이블은 관련성 없는 것이 이것저것 모여서는 안된다.
- 테이블은 공통 속성을 가진 것의 집합이다.
- 고유한 기본키를 가진 공통점에 의해 정리된 것들의 집합

| 테이블명은 반드시 복수형이나 집합명사로 표현해야 한다.

5-2 설계 규칙

| 열이란 개체의 속성이다

열 == 인스턴스의 상태

- 테이블과 객체지향언어
 - 테이블 == 클래스
 - 행 == 인스턴스
 - 열 == 상태, 속성

| 현실 세계에 같은 사람은 2명 있지 않다.

기본키

- 테이블에는 기본키가 반드시 있어야한다.
 - 이것은 "한 개 테이블의 내용에는 중복 행을 허용하지 않는다"라는 의미가 된다.
- `hash` 과 유사.
- 기본키의 특성
 - 기본키는 중복되면 안된다. (유일성)
 - 기본키 할당은 관리의 기본.
 - 기본키 열로 `NULL` 은 안된다.
 - `NULL` : 값이 불명확하거나 정하는 것이 가능하지 않는 경우

정규화 (Normalization)

- 일반적으로 정규화란, 제대로 조직되지 않은 테이블들과 관계들을 작고 잘 조직된 테이블과 관계들로 나누는 것을 말한다.
- 정규화는 정제되지 않은 표를 관계형 데이터베이스에 어울리는 표로 만들어주는 것을 말한다.
- 정규화의 목적은 다음 두 가지이다.
 - 불필요한 데이터를 제거한다. (데이터 중복)
 - 데이터 저장을 논리적으로 한다.
 - 하나의 테이블에서의 삽입, 삭제, 변경이 정의된 관계들로 인하여 DB의 나머지 부분들로 전파되게 하는 것.
 - + 디스크 IO를 최적화한다

제 1정규형

제 1정규형

- 테이블 셀에 복합적인 값을 포함하지 않는다.
- 원자값 (Atomic Value)
 - 1차 정규형은 각 로우마다 컬럼의 값이 1개씩만 있어야 한다.
 - 스칼라값 이외의 값을 포함 X

제 2정규형

제 2정규형

- 모든 테이블의 모든 컬럼이 **완전 함수적 종속**을 만족하는 것
 - **부분함수 종속성**이 존재하지 않아야 한다.
 - 테이블의 기본키중에 중복키가 있다면 이것을 해결해야 한다는 의미
 - 문제
 - 2정규형이 지켜지지 않을 경우, 기본 키 이외의 행에 NULL 혹은 중복 값이 존재하면서 갱신 이상이 발생 가능
 - 중복된 정보 중 하나만 수정되고 다른 중복 정보는 수정되지 않으면 데이터베이스 내의 정보가 서로 모순될 가능성 존재

부분함수 종속성

- 기본키를 구성하는 열의 일부에만 **함수 종속성**이 존재하는 것

주문 테이블(제2정규형 아님)

| 고객기업 ID | 주문번호 | 주문접수일 | 고객기업명 | 고객기업 규모 |
|---------|------|------------|-------|---------|
| CA | ○001 | 2014/12/20 | A 상사 | 대규모 |
| CA | ○002 | 2014/12/21 | A 상사 | 대규모 |
| CA | ○001 | 2014/12/12 | B 건설 | 중규모 |
| CA | ○002 | 2014/12/25 | B 건설 | 중규모 |
| CB | ○003 | 2014/12/25 | B 건설 | 중규모 |
| CB | ○001 | 2014/12/1 | C 화학 | 소규모 |

- 1NF는 충족하고 있지만, 2NF는 충족하지 못하고 있다.
 - 기본키는 **고객기업 ID** 와 **주문번호** 의 조합이며, 이 조합으로 행을 유일하게 특정하는 것이 가능.
 - **고객기업ID** → **고객기업명** 과 **고객기업 규모**
 - **주문번호** -X→ **고객기업명** 과 **고객기업 규모**
 - 기본키를 구성하는 열(**고객기업 ID** , **주문번호**)의 일부(**고객기업 ID**) 만 **함수 종속성**이 존재할 경우 ⇒ **부분함수 종속성**
- 제 2 정규형 만족

고객기업 주문 테이블(제2정규형)

| 고객기업 ID | 주문번호 | 주문접수일 |
|---------|------|------------|
| CA | ○001 | 2014/12/20 |
| CA | ○002 | 2014/12/21 |
| CB | ○001 | 2014/12/12 |
| CB | ○002 | 2014/12/25 |
| CB | ○003 | 2014/12/25 |
| CC | ○001 | 2014/12/1 |

고객기업 테이블(제2정규형)

| 고객기업 ID | 고객기업명 | 고객기업 규모 |
|---------|-------|---------|
| CA | A 상사 | 대규모 |
| CB | B 건설 | 중규모 |
| CC | C 화학 | 소규모 |

제 3정규형

제 3정규형

- 추이함수 종속이 존재하면 안된다.
 - 기본키를 제외한 속성들 간의 이행적 함수 종속이 없어야 한다.
 - 우편번호를 적으면 주소가 자동적으로 결정되면 안된다.
 - 메일 id → 우편번호 → 주소 (x)

추이함수 종속

- 추이함수 종속이란 기본키에서 볼 때 2단계의 함수 종속이 존재한다는 의미이다.

표 8-14 고객기업 테이블(제3정규형 아님)

고객기업 테이블(제3정규형 아님)

| 고객기업 ID | 고객기업명 | 고객기업규모 | 업계코드 | 업계명 |
|---------|-------|--------|------|-----|
| CA | A상사 | 대규모 | D001 | 석유 |
| CB | B건설 | 중규모 | D002 | 건설 |
| CC | C화학 | 소규모 | D003 | 바이오 |

- {고객기업 ID} -> {업계코드} -> {업계명}

- 추이함수 종속이 존재할 경우 테이블 갱신 이상이 존재.

- 이행적 함수 종속성 → 일부 데이터를 변경하려 할 때 다른 데이터도 함께 변경되어야 하는 상황이 발생
- 테이블에 업계를 등록할 방법이 없다 → 원치 않는 추가 정보 없이 특정 데이터를 삽입하기 어려운 상황이 발생
- ✓ 테이블에는 **기업**이라는 단위의 집합을 반영할 수 있어도 **유통** 업계를 속한 기업과 거래하지 않은 이상 해당 업계 코드를 만들 수 없다.

고객기업 테이블(제3정규형)

| 고객기업 ID | 고객기업명 | 고객기업 규모 | 업계코드 (FK) |
|---------|-------|---------|-----------|
| CA | A상사 | 대규모 | D001 |
| CB | B건설 | 중규모 | D002 |
| CC | C화학 | 소규모 | D003 |
| CD | D상사 | 중규모 | D001 |
| CD | E섬유 | 대규모 | D003 |

업계 테이블(제3정규형)

| 업계코드 | 업계명 |
|------|-----|
| D001 | 석유 |
| D002 | 건설 |
| D003 | 바이오 |
| D004 | 시스템 |

부분함수 종속 과 **추이함수 종속** 을 생각하며 테이블을 설계하면 90%는 능숙하게 진행할 수 있다.